WILEY

Encyclopedia of Global Environmental Change

egec

Editor-in-Chief
Ted Munn

Encyclopedia of
Global Environmental Change

egec

The volumes in this Encyclopedia are:

Volume One
The Earth system: physical and chemical dimensions of global environmental change

Volume Two
The Earth system: biological and ecological dimensions of global environmental change

Volume Three
Causes and consequences of global environmental change

Volume Four
Responding to global environmental change

Volume Five
Social and economic dimensions of global environmental change

Encyclopedia of
Global Environmental Change

Editor-in-Chief

Ted Munn

Institute for Environmental Sciences, University of Toronto, Canada

1

egec

The Earth system: physical and chemical dimensions of global environmental change

Volume Editors

Michael C MacCracken

Lawrence Livermore National Laboratory, California, USA

and

John S Perry

formerly, National Research Council, USA

JOHN WILEY & SONS, LTD

Copyright © 2002 John Wiley & Sons, Ltd, Chichester
Baffins Lane, Chichester
West Sussex PO19 1UD, UK

National: 01243 779777
International: (• 44) 1243 779777
e-mail (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on http://www.wiley.co.uk
or http://www.wiley.com

Copyright Acknowledgments

A number of articles in the Encyclopedia of Global Environmental Change have been written by government employees in the United Kingdom, Canada and the United States of America. Please contact the publisher for information on the copyright status of such works, if required. In general, Crown copyright material has been reproduced with the permission of the Controller of Her Majesty's Stationery Office. Works written by US government employees and classified as US Government Works are in the public domain in the United States of America.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except under the terms of the Copyright Designs and Patents Act 1988 or under the terms of a license issued by the Copyright Licensing Agency, 90 Tottenham Court Road, London, W1P 0LP, UK, without the prior permission in writing of the publisher.

Other Wiley Editorial Offices

John Wiley & Sons Inc., 605 Third Avenue,
New York, NY 10158-0012, USA

Wiley-VCH Verlag GmbH, Pappelallee 3,
D-69469 Weinheim, Germany

Jacaranda Wiley Ltd, 33 Park Road, Milton,
Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01,
Jin Xing Distripark, Singapore 129809

John Wiley & Sons (Canada) Ltd, 22 Worcester Road,
Rexdale, Ontario M9W 1L1, Canada

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library.

ISBN: 0 471 97796 9

Typeset in 10pt Times by Laser Words Private Limited, Chennai, India
Printed and bound in Great Britain by Antony Rowe, Chippenham, Wiltshire

Editorial Board

Editor-in-Chief

Ted Munn
Institute for Environmental Studies, University of Toronto, Canada

Volume Editors

Volume One ***The Earth system: physical and chemical dimensions of global environmental change***
Dr Michael C MacCracken, *Lawrence Livermore National Laboratory, California, USA* and
Dr John S Perry, *formerly, National Research Council USA*

Volume Two ***The Earth system: biological and ecological dimensions of global environmental change***
Professor Harold A Mooney, *Stanford University, Stanford, USA* and
Dr Josep G Canadell, *GCTE/IGBP, CSIRO Sustainable Ecosystems, Canberra, Australia*

Volume Three ***Causes and consequences of global environmental change***
Professor Ian Douglas, *University of Manchester, Manchester, UK*

Volume Four ***Responding to global environmental change***
Dr Mostafa K Tolba, *International Center for Environment and Development, Cairo, Egypt*

Volume Five ***Social and economic dimensions of global environmental change***
Mr Peter Timmerman, *IFIAS, Toronto, Canada*

International Advisory Board

Dr Joe T Baker
Commissioner for the Environment, ACT, Australia

Professor Francesco di Castri
CNRS, France

Professor Paul Crutzen
Max Planck Institute for Chemistry, Germany

Professor Eckart Ehlers
IHDP/IGU Germany

Professor José Goldemberg
Universida de São Paulo, Brazil

Dr Robert Goodland
World Bank, USA

Professor Hartmut Grassl
WCRP, Switzerland

Professor Ronald Hill
University of Hong Kong, Hong Kong

Dr Yu A Izrael
Institute of Global Climate & Technology, Russia

Professor Roger E Kasperson
Clark University, USA

Professor Peter Liss
University of East Anglia, UK

Professor Jane Lubchenco
Oregon State University, USA

Mr Jeffrey A McNeely
IUCN, Switzerland

Professor Thomas R Odhiambo
Hon. President, African Academy of Sciences and
Managing Trustee, RANDFORUM, Kenya

Sir Ghillean T Prance
University of Reading, UK

Professor Steve Rayner
Columbia University, USA

Contents

Preface to the Encyclopedia of Global Environmental Change xi

Preface to Volume One xvii

The Earth System 1
Earth System Processes 13
Earth System History 31
Earth Observing Systems 61
The Global Temperature Record 82
Models of the Earth System 99
Model Simulations of Present and Historical Climates 114
Projection of Future Changes in Climate 126
Depletion of Stratospheric Ozone 140
International Organizations in the Earth Sciences 156

Acoustic Thermometry 161
ACSYS (Arctic Climate System Study) 161
Aerosols, Effects on the Climate 162
Aerosols, Polar Stratospheric Cloud 167
Aerosols, Stratosphere 169
Aerosols, Troposphere 172
Agassiz, Louis 175
Air Pressure 176
Air Quality, Global 177
Albedo 182
AMIP (Atmospheric Model Intercomparison Project) 183
Antarctica 184
Anthropocene 189
Anticyclone 191
Arctic Air Quality 191
Arctic Climate 193
Arctic Ocean 199
Arctic Oscillation 201
ARM (Atmospheric Radiation Measurement) Program 203
Arrhenius, Svante 204
Asteroids and Comets, Effects on Earth 204
Atlantic Ocean 211
Atmospheric Angular Momentum and Earth Rotation 211
Atmospheric Composition, Past 213
Atmospheric Composition, Present 216
Atmospheric Electricity 218
Atmospheric Motions 221
Atmospheric Structure 243

Bjerknes, Jacob 245
Bolin, Bert 245
Broecker, Wallace S 246
Budyko, Mikhail Ivanovich 248

Carbon Dioxide Concentration and Climate Over Geological Times 249
Carbon Dioxide, Recent Atmospheric Trends 254
Carbon Monoxide 261
CEOS (Committee on Earth Observation Satellites) 261
Chandler Wobble 262
Chaos and Predictability 263
Charney, Jule Gregory 266
Chlorofluorocarbons (CFCs) 267
CLIC (Climate and Cryosphere) 269
CLIMAP (Climate: Long-range Investigation, Mapping, and Prediction) 269
Climate 270
Climate Agenda 270
Climate Analogues 271
Climate Change 271
Climate Change, Abrupt 272
Climate Change, Detection and Attribution 278
Climate Feedbacks 283
Climate Model Simulations of the Geological Past 296
Climate Sensitivity 301
Climatology 308
CLIVAR (CLImate VARiability and Predictability) 311
Cloud–Radiation Interactions 312
Clouds 316
Continental Drift 321
Convection 325
Coriolis Effect 326
Cretaceous 329
Crutzen, Paul J 330
Cryosphere 330

Dansgaard–Oescheger Cycles 332
Deserts 332
Dimethylsulfide (DMS) 343
Downscaling 346
Dust 347

Eemian 352
El Niño 353
El Niño and La Niña: Causes and Global Consequences 353
El Niño/Southern Oscillation (ENSO) 370
El Viejo 370
Energy Balance and Climate 371
Energy Balance Climate Models 376
EOS (Earth Observing System) 382
Equilibrium Response 382

Feedbacks, Chemistry – Climate Interactions 384
Fingerprinting 388
Franklin, Benjamin 390
Fronts 391

GAIM (Global Analysis, Interpretation and Modeling) Program 392
GARP (Global Atmospheric Research Program) 392
GAW (Global Atmosphere Watch) 393
GCDIS (Global Change Data and Information System) 393
GCOS (Global Climate Observing System) 394
General Circulation Models (GCMs) 394
Geological Cycling 397
Geomorphology 402
Geothermal Heat 402
GEWEX (Global Energy and Water Cycle Experiment) 402
Glaciers 404
Global Plate Tectonics 410
Global Warming Potential (GWP) 411
GOALS (Global Ocean–Atmosphere–Land System) 411
Greenhouse Effect 413
Greenland 413
Greenland Ice Sheet 419
Ground Temperature 422

Hadley Circulation 427
Halocarbons 427
Hare, Kenneth 428
Heinrich (H-) Events 429
Holocene 431
Houghton, John Theodore 432
Humidity 432
Hurricanes, Typhoons and other Tropical Storms – Descriptive Overview 433
Hurricanes, Typhoons and other Tropical Storms – Dynamics and Intensity 439
Hydrofluorocarbons 447
Hydrogen Peroxide Trends in Greenland Glaciers 447
Hydrologic Cycle 450
Hydrology 464

IAHS (International Association of Hydrological Sciences) 466
IAMAS (International Association of Meteorology and Atmospheric Sciences) 466
IAPSO (International Association for the Physical Sciences of the Oceans) 467
Icebergs 467
IGAC (International Global Atmospheric Chemistry) 467
IGY (International Geophysical Year) 468
IHP (International Hydrological Program) 468
Imbrie, John 469
Infrared Radiation 470
Intertropical Convergence Zone (ITCZ) 476
Ionosphere 476
IPCC (Intergovernmental Panel on Climate Change) 477
IRI (International Research Institute for Climate Prediction) 477
ISCCP (International Satellite Cloud Climatology Project) 478
ISLSCP (International Satellite Land Surface Climatology Project) 478
Isostasy 479
IUGG (International Union of Geodesy and Geophysics) 479

Jet Stream 481
JGOFS (Joint Global Ocean Flux Study) 483
Junge Layer 483

Keeling, Charles David 484
Kondratyev, Kirill Yakovlevich 485

La Niña 487
Lamb, Hubert H 487
Land Cover and Climate 488
Land Surface 493
Lapse Rate 499
Last Glacial Maximum 500
Latent Heat 500
Lifetime (of a Gas) 501
Lightning 501
Lightning and Atmospheric Electricity 502
Limnology 503
Lithosphere 503
Little Ice Age 504
Lorenz, Edward N 509
Lovelock, James 510

Madden–Julian Oscillation 511
Malone, Thomas F 511
Manabe, Syukuro 512
Maunder Minimum 514
Medieval Climatic Optimum 514

Mesosphere 516
Mesozoic 516
Meteorology 517
Methane 517
Methane Clathrates 518
Methyl Bromide 520
Milankovitch, Milutin 522
Modeling Regional Climate Change 523
Molina, Mario J 533
Monitoring Systems, Global Geophysical 534
Monsoons 539
Mountain Climates 540
MSU (Microwave Sounding Unit) 541
Munk, Walter 541
Munn, Robert Edward (Ted) 542

Natural Climate Variability 544
Natural Records of Climate Change 550
Nitrous Oxide 554
North Atlantic Oscillation 555

Ocean Circulation 557
Ocean Conveyor Belt 579
Ocean Drilling Program 581
Ocean Observing Techniques 581
Oceanography 584
Oeschger, Hans A 585
Orbital Variations 586
Ozone Depletion Potential (ODP) 590
Ozone Hole 590

Pacific–Decadal Oscillation 592
Pacific–North American (PNA) Teleconnection 594
PAGES (Past Global Changes) 596
Paleoclimatology 596
Paleozoic 597
Parameterization 598
Perfluorocarbons (PFCs) 598
Permafrost 598
Photochemical Reactions 601
Planetary Boundary Layer 603
Plate Tectonics 605
Pleistocene 607
Polynyas 607
Prediction in the Earth Sciences 607

Quasi–Biennial Oscillation (QBO) 611
Quasi–Decadal Oscillation 613
Quaternary 615

Radiative Forcing 616
Radionuclides, Cosmogenic 618
Radiosondes 619
Rain 622
Residence Time (of an Atom, Molecule or Particle) 622
Revelle, Roger Randall Dougan 623
Richardson, Lewis Fry 624
Roberts, Walter Orr 625
Rossby, Carl-Gustaf 626
Rossby Waves 626
Rowland, F Sherwood 627
Runoff 628

Salinity 629
Salinity Patterns in the Ocean 629
SCOR (Scientific Committee on Oceanic Research) 640
Sea Ice 640
Sea Level 645
Sea Surface Temperature 650
Sensible Heat 656
Snow 656
Soil Moisture 658
Solar Irradiance and Climate 659
Solar Variability, Long-term 666
Southern Ocean 668
Southern Oscillation 672
SPARC (Stratospheric Processes and their Role in Climate) 672
Storm Surge 673
Stratosphere 674
Stratosphere, Chemistry 675
Stratosphere, Ozone Trends 682
Stratosphere, Temperature and Circulation 697
Sub-grid Processes 704
Sunspots 704
Suomi, Verner Edward 707
Sverdrup, Harald Ulrik 708

Temperature 709
Thermohaline Circulation 710
Tides, Atmospheric 710
Tides, Oceanic 710
TOGA (Tropical Ocean Global Atmosphere) 713
Tornadoes 713
Trade Winds 715
Transient Response 715
TRMM (Tropical Rainfall Measuring Mission) 716
Tropopause 717
Troposphere 717
Troposphere, Ozone Chemistry 718
Tropospheric Temperature 720
Tsunamis, Causes and Consequences 725
Tunguska Phenomenon 730

UARS (Upper Atmosphere Research Satellite) 732
Ultraviolet Radiation 732

Volcanic Eruption, El Chichon 736
Volcanic Eruption, Krakatau 736
Volcanic Eruption, Mt. Pinatubo 737
Volcanic Eruption, Tambora 737
Volcanic Eruptions 738
Vostok, Subglacial Lake 744

Walker Circulation 749
Water Vapor: Distribution and Trends 750
WCC (World Climate Conferences 1979 and 1990) 752
WCP (World Climate Programme) 752
WCRP (World Climate Research Programme) 753
Weather 754
Weathering 755
Wegener, Alfred 755
WGNE (Working Group on Numerical Experimentation) 756
Wind Chill 756
WMO (World Meteorological Organization) 757
WOCE (World Ocean Circulation Experiment) 759

Younger Dryas 761

Alphabetical List of Articles 763

Preface to the Encyclopedia of Global Environmental Change

THE ENVIRONMENT

The word *environment*, whose dictionary meaning is simply that which surrounds, has in the last few decades become a buzzword , encompassing an exceedingly diverse array of elements and social issues. Taking the original meaning as a point of departure, it is clear that we humans depend totally on the environment provided by planet Earth for the food we eat, the water we drink, and the air we breathe. Thus changes in this environment must be of vital concern. Will Earth continue to sustain humans in a way that also encourages the flourishing of the other living things with whom we share the planet? This question has loomed ever larger as it has become more evident that human activities have been inducing major changes in all of the compartments of the global environment. We have converted forests and savannas to farms and cities; we have exhumed ancient treasures of fuels and minerals; we have used the rivers and winds as convenient sewers; and we have released entirely new chemical compounds and organisms into the environment. In the 1960s, the scientific community began to use the word *environment* in this new non-specialist sense. Soon too, Departments of the Environment were created by many governments, and new scientific journals began publication while others were re-named. For example, the *International Journal of Air Pollution* became *Atmospheric Environment*! In the ensuing decades, the world community has come to see the environment in many different ways, as a life-support system, as a fragile sphere hanging in space, as a problem, a threat and a home.

GLOBAL ENVIRONMENTAL CHANGE

A broader and deeper understanding of the global aspects of environmental concerns emerged in the 1970s and 1980s, and a new phrase *global environmental change* acquired popular currency. Paleoresearch had revealed that environmental change was far from new, and by no means the sole result of the actions of heedless humans. Since the planet s formation, virtually every element of its environment has been undergoing massive changes on all space and time scales. Oxygen waxed and carbon dioxide waned in the atmosphere. Continents moved about the planet s surface like scum on a soup kettle. Great ice sheets grew and shrank. Above all, the force of life (the biosphere) emerged as a dominant driver of planetary change.

However, another vital insight began to emerge about 1980: the inescapably interlinked nature of these many environmental changes. On the very longest time scales, continental drift moved lands into different climates – but also changed the climate of the globe itself. Photosynthetic life changed the atmosphere – but also made possible more advanced life forms that could take advantage of the new environment. On shorter time scales, atmosphere and ocean often interact to produce the massive changes in the Southern Pacific that we term El Niño and La Niña, whose consequences extend across the planet, and profoundly affect even our socioeconomic systems. Indeed, we have come to realize that human-induced perturbations in the environment are becoming increasingly large, and are potentially coming to dominate the natural workings of the complex and interdependent global system that sustains life on Earth. Humans and their global environment are no longer independent; they are ever-increasingly becoming interdependent components of a single global system.

Thus, the term *global environmental change* has come to encompass a full range of globally significant issues relating to both natural and human-induced changes in Earth s environment, as well as their socioeconomic drivers. This implicitly includes concerns for the capacity of the Earth to sustain life that have motivated the development of studies of global change and sustainable development in the last few decades. Analyses of global environmental change therefore demand input from the social sciences as well as the natural sciences (and indeed also from the engineering and health sciences) – necessitating an inescapably interdisciplinary approach.

Scientists from many disciplines have been attracted to this growing field of global environmental change. This is particularly noticeable in the biological sciences through the encouragement of IGBP (International Geosphere–Biosphere Programme), which invites ecologists to expand their field of vision from the plot and landscape scales to the regional, continental and global ones, and to interact with scientists from other disciplines in exploring environmental change at these larger scales. Indeed this trend has encouraged publishers and scientific societies to introduce new journal titles, and these publications from all accounts appear to be flourishing.

At the same time, human social, economic, and cultural systems are rapidly changing under the influence of growing globalization. In the economic sphere, for example, today s discourse centers on multinational corporations, the Global

Economy, and the Globalization of Trade. Thus, the agendas of most of the physical, life, and social sciences increasingly focus on global-scale changes.

THE EVER-CHANGING ENVIRONMENT

Since the launching of Earth-viewing satellites, we have been able to see a constantly updated moving picture of our planet. Great weather systems sweep around the globe, waves of green and brown ebb and flow over the continents with the seasons, as do white waves of ice and snow. Nevertheless, a remarkable feature of the Earth system has been its relative stability. Since the dawn of life, the planet s environment has remained within a range of conditions that has supported life. Moreover, with some notable and perhaps cautionary exceptions, changes from year to year, decade to decade, millennium to millennium, have been modest, particularly during the last 10 000 years. Many features of human society exhibit similar behavior. Civilizations and cultures evolve slowly for the most part in response to environmental, economic, and social driving forces over their long lives. Great cities like Rome endure through the ages. A citizen of ancient Babylon probably would have little difficulty understanding the politics of modern Toronto.

But today we appear to be entering an era of change unprecedented since Babylon was a cluster of mud huts. There are many reasons to believe that changes greater than humankind has experienced in its history are in progress and are likely to accelerate.

Virtually every measure of human society – numbers of people and automobiles, airplanes, energy consumption, generation of waste – is increasing exponentially. While the values of a few indicators are leveling off, the magnitudes of annual increases of others remain immense.

Major changes are becoming evident in many critical elements of the environment – increasing carbon dioxide concentrations, stratospheric ozone depletion (not to mention the stratospheric Antarctic ozone hole and the possibility of a similar Arctic event), rising sea level, declining productivity of soils, widespread collapses of fisheries, dramatic declines in biodiversity, destruction of tropical forests, etc. Within living memory, the countryside around large cities has been swallowed up by suburban developments and highways. Humans are stepping on the gas pedal of the planet s environment, and perhaps recklessly breaking the survivable speed limit of global change.

Paleoclimatological studies have shown that the stable environment that we take for granted has not always prevailed. Indeed, the relatively brief period in which human civilization has developed is somewhat unusual in its equable stability. Neolithic hunters in Europe some 13 000 years ago saw their climate shift from temperate to glacial in a single short lifetime. The evolution of human society is punctuated by wars, pestilence, and technological revolutions. Major – perhaps even catastrophic – change could occur in the future, because we know it has occurred in the past.

Rapidly advancing understanding of both natural and human systems – and above all the ability to translate that understanding into quantitative models – has enabled us to explore the future of our global society and its global climate with unprecedented realism. Although prediction of the single future path that we will follow is inherently unpredictable, it is possible to map a broad range of future environmental trajectories that we might take, each completely consistent with our understanding of how the system works. Such scenario-building exercises amply confirm our concerns that the changes of the 21st Century could be far greater than experienced in the last several millennia. Business-as-usual for human society appears to imply business-as-highly-unusual for the global environment.

THE INTERLOCKING BIOGEOPHYSICAL – SOCIOECONOMIC SYSTEMS

Recognition came in the 1970s that many of the environmental issues are inter-connected through the biogeochemical cycling of trace substances, especially carbon, sulfur, nitrogen and phosphorus. In fact, in a prescient statement on the main environmental research priority of the 1980s, Mostafa Tolba (then Executive Director of the United Nations Environment Programme) and Gilbert White (then President of the Scientific Committee on Problems of the Environment) drew the attention of both the scientific and the science-policy communities to the need to understand the major global biogeochemical cycles in order to maintain the global life support systems in a healthy state (Tolba and White, 1979). Quoting from that statement,

> We draw attention to the fundamental scientific importance of understanding the biogeochemical cycles which link and unify the major chemical and biological processes of the Earth s surface and atmosphere. The results will have practical significance for all of us who inhabit an Earth with limited resources and who, by our actions, increasingly affect the quality of the human environment.

So it is that many of the global environmental issues – acid rain, stratospheric ozone depletion, climate change, nitrogen over-fertilization – are inter-related through the global biogeochemical cycles.

One of the interesting results of the study of the Earth s history has been the discovery of the global teleconnections of the Earth system, with some major climatic shifts occurring simultaneously in the two hemispheres. In an analogous way, there has been recognition for a long time that human social, economic, and cultural systems are globally inter-related. That these are connected in turn to the

Earth system was implicitly recognized as early as the 1972 Stockholm Conference on the Human Environment.

To manage the human responses to this enormously varied but at least moderately coupled world system in an era of increasing global change through the diverse array of local, national and international organizations is indeed a daunting challenge. An essential first step is to describe past and present states, then to explain the various phenomena observed, and finally to develop predictive models (or at least a range of scenarios) describing the future behavior of this total Socioeconomic–Cultural–Environmental system. In this process, environmental scientists must learn how to assimilate better the new information constantly being received, although uncertainties, often large, will continue to exist. Within this mix, rational and effective policies must be developed that will balance the risks and costs of global environmental change in an adaptive way. Only then will we be able to even formulate, much less start to implement, rational and effective policies that cope effectively with global change. The prospect of change tempts us to think in terms of winners and losers. However, such analyses often do not play out in simple ways. For example, a modest increase in rainfall would cause farmers to rejoice while vacationers would despair. However, greater farm production can lead to lower commodity prices, thereby reducing farm incomes while making vacation food purchases less expensive. Human society is pretty well adapted to the present environment, so change is necessarily a challenge.

In the longer-term, much depends on the ability of societies to respond, to adapt. Societies will differ in the resources – natural, human, and technological – that are available to them. They will also differ in terms of the values and priorities they attach to physical, social, and environmental goals, and in the social and political mechanisms that they employ to seek these goals. These differences in the human world of generations yet unborn may be as great – and as significant – as the changes in the global environment. A major challenge for our times is to develop frameworks for understanding complex interdisciplinary issues of this complexity.

No discussion of change and the future can be complete without consideration of risk. Projections of the future, however imaginative or soundly based, necessarily center on plausible, surprise-free scenarios – population will increase, economies will advance, climate will change – all typically slowly and smoothly. However, such projections made at the beginning of the 20th Century would have missed two World Wars, the automobile, aviation, space travel, television, and McDonald s in Beijing. By definition, genuine surprises cannot be predicted. However, an understanding of the impacts of past surprises may help us to make our society and our world somewhat more robust in the face of the unknown surprises that await us.

THE HUMAN DIMENSIONS OF GLOBAL ENVIRONMENTAL CHANGE

The human dimension of global change has been defined as the ways in which individuals and societies contribute to global environmental change, are influenced by it, and adapt to it. Many empirical studies have been undertaken that describe human activities in physical terms, based on various kinds of indicators. Rates of deforestation, urbanization, and changing levels of emissions of greenhouse gases are only some measurable contributors to environmental change. But the dynamics of human activities and global change are much more complex, and they reflect the complexities of the human–nature relationship.

While the sheer burden of human activity on the planet is important, we know that the major forces at work are human beings operating together through political systems, corporations, interest groups, and beliefs that sway whole peoples. These raise questions about the nature of choice, about the hopes, dreams, and frustrations that impel people forward, or block them. At the moment, for instance, the dream of the good life – which has been a critical element of every religious tradition in the world – is increasingly being defined in terms of material possessions and powerful images that are shown on global media. Can this particular version of human happiness be sustained on a limited planet? Some people say that it can; others say that it cannot.

The central questions about the role of human beings in global environmental change revolve around social, cultural, economic, ethical, and even religious issues. These are becoming more and more pressing, and more and more foundational as human beings deliberately or inadvertently modify more and more of the planet. It is also obvious that in this generation the modification of organisms and ecosystems may well extend to the modification of human beings themselves. Among the fundamental questions are: What motivates us towards saving or harming the environment? How do we see ourselves with regard to nature? What is our responsibility to this and future generations? Who do we think we are, and who would we like to become? Appropriately enough, Volume Five of this Encyclopedia wrestles, in many different voices, with these ultimate questions that remain intimately linked with the sweep of physical, chemical, biological, geographical, and institutional changes documented and discussed throughout the earlier four volumes.

The Brundtland philosophy urges us not to reduce options for future generations. Implementation of this idea is, however, difficult. For example, it often seems more compelling to alleviate current poverty than to protect the environment and renewable resources for future generations. Many scientists agree that new approaches are needed to meld the social with the natural sciences in the policy arena. Some of the new methodologies that go beyond the physical sciences

are described in Volumes Four and Five of this Encyclopedia, e.g., post-normal science, integrated environmental assessment, and the precautionary principle.

In a commentary in which he described the 21st Century as the century of the environment, Edward O. Wilson began by referring to the growing *consilience* (the interlocking of causal explanations across disciplines) so that the interfaces between disciplines become as important as the disciplines themselves. He then stated that this interlocking amongst the natural sciences will in the 21st Century also touch the borders of the social sciences and humanities. In the environmental context, environmental scientists in diverse specialities, including human ecology, are more precisely defining the arena in which that species arose, and those parts that must be sustained for human survival (Wilson, 1998).

This is a major challenge for environmental scientists. Already through DNA techniques, for example, it is possible to trace back connections amongst prehistoric peoples through the last Ice Age to modern times.

THIS *ENCYCLOPEDIA OF GLOBAL ENVIRONMENTAL CHANGE*

More than two million papers are published every year in science and medicine, a twentyfold increase since 1940 that taxes the resources of concerned citizens, scientists, university departments, research institutes and libraries. At the same time, many policy-motivated organizations find it difficult to draw together the necessary expertise for resolving the newly emerging environmental issues. The scientific literature relating to the environment is burgeoning. However, research syntheses are in general still scattered across a broad spectrum of journals and books, and information on global environmental change is not readily available in an inter-related way. In particular, it is quite uncommon for contributions from the natural sciences to appear in the same journal or workshop proceedings as contributions from the social sciences and the humanities.

This *Encyclopedia of Global Environmental Change* is a comprehensive and integrated reference in the broad area of global environmental change that will be conveniently accessible and productively usable by students, managers, administrators, legislators, and concerned citizens. The Encyclopedia consists of five volumes of interrelated material:

Volume 1 The Earth system: physical and chemical dimensions of global environmental change (Michael C MacCracken and John S Perry, editors)

Volume 2 The Earth system: biological and ecological dimensions of global environmental change (Harold A Mooney and Josep G Canadell, editors)

Volume 3 Causes and consequences of global environmental change (Ian Douglas, editor)

Volume 4 Responding to global environmental change (Mostafa K Tolba, editor)

Volume 5 Social and economic dimensions of global environmental change (Peter Timmerman, editor)

The first four volumes cover the broad issues of the science and politics of global environmental change. Volume Five adds an often understated but extremely important aspect, linking as it does, global environmental change to the socioeconomic, cultural and ethical dimensions of human societies. Here will be found a rich panoply of writings by people who are not natural scientists but who have thought deeply about the environment. It places global environmental change in a refreshing historical, sociological and cultural context. In many contributions, the time horizons of most interest are the last hundred and the next hundred years. However, some contributions dip backward millions of years. Throughout, the emphasis is upon the dynamics of the various processes discussed – how and why did they change? A second recurrent theme is the interconnection of processes and changes – What produces change? What is impacted by change? Finally, we attempt to deal even-handedly with natural and human-induced change, and with impacts on both the natural world and human society. From the numerous diverse articles in the Encyclopedia, we believe that the user can obtain a coherent picture of this complex and dynamic system of which we all are a part.

To assist in promoting this coherence, each volume begins with a group of extended essays on major topics that embrace the field covered in that volume. These are intended to provide an introduction to the topic, a convenient road map through cross-references for exploring the Encyclopedia. Then there follows, in alphabetical order, shorter articles on a variety of scientific topics, descriptions of scientific programs, definitions, acronyms, and biographies of leading contributors to the field – from Charles Darwin, through the Russian ecologist Vernadski, to the three most recent environmental Nobel Laureates – Crutzen, Rowland and Molina. Indeed, these definitions, biographies and acronym definitions are, we believe, a uniquely valuable feature of the Encyclopedia. In the case of acronyms of international organizations and programs, it is no exaggeration to state that a young environmental scientist requires not only a good understanding of science, but also a good knowledge of acronyms, if they are to follow the discussions taking place at many international meetings! Also included in the alphabetic listings are abundant cross-references to related topics in the same or other volumes.

The substantive scientific articles that comprise the meat of the Encyclopedia are original contributions by active scientists from around the world, and thus

represent authoritative and up-to-date summaries of the state of current knowledge, direct from the producers of this knowledge. A number of these articles break new ground in synthesizing and summarizing our understanding from novel viewpoints and in unconventional ways. Thus, readers will find a wide variety of styles and approaches within the articles, reflecting in a unique way the rich diversity of today s world science. The articles have also had the benefit of careful reviews, particularly by our Editors.

The scientific essays and some of the program descriptions begin with a few italicized paragraphs written for non-specialists. These are not intended to be abstracts of the paper to follow, but rather are aimed at providing the reader with an introduction into why the topic is important and where it fits into the broader aspects of global change – a kind of encouragement to read on. Reading on should not be too difficult a task, since most of the scientific essays are written at the level of journals such as *Scienti c American* and *AMBIO* that are intended for the non-specialist. The Encyclopedia will be a valuable source of information for everyone with a general interest or a need-to-know in the various environmental fields (the natural sciences, socio-economics, engineering, the health sciences, and policy analyses) particularly as they relate to global-scale environmental change, its drivers, and its consequences. It is also expected that among the audiences for each volume will be practitioners and researchers in the fields covered by the other four volumes. We believe that this rather unusual – indeed unique – Encyclopedia will be used in a variety of ways.

1. Some people will employ it simply as a convenient source of information on specific topics – What has been happening recently to the sea ice in the Arctic Ocean? Who was Roger Revelle? What is soil mineralization? What is ICSU or IGBP? What is the Kyoto Protocol? What is deep ecology?
2. Serious students in the environmental sciences may begin by reading one or all of the introductory essays, which present highly compressed crash courses in a range of central topics.
3. We believe that the substantive specialized articles that constitute the bodies of the volumes are productively and enjoyably readable in their own right.
4. Finally, a case has been made by one of the volume editors (Peter Timmerman) for encyclopedia browsing as an enjoyable and productive pastime (*see* Encyclopedias: Compendia of Global Knowledge, Volume 5).

We believe that these five volumes are in the tradition of the human aspiration towards the compilation of global knowledge that sparked the preparation of the first encyclopedias in the 18th Century.

Our Editors have had the benefit of a marvellous postgraduate education in the course of the Encyclopedia s evolution! We are immensely grateful to the many, many authors in our virtual university who have contributed their wisdom to this project.

ACKNOWLEDGMENT

Many helpful comments and draft paragraphs were received from the Volume Editors during the preparation of this Preface.

REFERENCES

Tolba, M K and White, Gilbert F (1979) Global Life Support Systems, SCOPE Newsletter, No. 7, October. 2 pp.

Wilson, Edward O (1998) Integrated science and the coming century of the environment, SCIENCE 279, 2048–2049.

Ted Munn
Editor-in-Chief

Preface to Volume One

Volume One of this Encyclopedia deals with the physical and chemical dimensions of the Earth system, including for example the atmosphere, the oceans, the cryosphere, and those aspects of the land surface particularly relevant to interactions with other components of the Earth system. These unique characteristics of our planet provide the foundation for the living and human worlds that are treated in other volumes of this Encyclopedia.

This volume begins with a group of extended review essays, followed by a large number of shorter articles. Together, these articles cover various aspects of the history, current state, and possible future states of the Earth system, including the interactions among the components of this system. Special attention is given to historical trends in various environmental indicators, both in the past century and extending back through geological time. There are also articles on what we have learned, and the tools used to gain knowledge about the functioning of this complex system, from field programs to model simulations. Although interactions of the environment with human activities are for the most part treated in other volumes, some of the effects of human intervention in environmental change, for example climate change and stratospheric ozone depletion, are included here. Because of the special importance of organized international observing, data management, and research programs in the Earth sciences, there are also numerous articles on relevant international institutions and programs. Finally, science is created by scientists, and we include brief biographies of a selection of leading individuals. Throughout, we have attempted to focus on the most dynamic aspects of the system, on the factors and processes that produce change, and on the programs and individual scientists most concerned with measuring and understanding change.

We have endeavored to construct a volume broad enough in scope to illuminate every corner of the relevant geophysical sciences, while remaining concise and readily usable by the non-specialist. The articles in the volume have for the most part been contributed by active researchers deeply expert in their fields. Other informational material has been assembled and summarized from a variety of public sources. Thus, we believe that this volume, like the Encyclopedia as a whole, represents a uniquely valuable source of focused, timely, and authoritative information relating to the issues of global environmental change.

Assembling this volume has been an immensely rewarding experience for us, refreshing and expanding our knowledge and understanding. We believe that the result will similarly expand the knowledge and understanding of its readers. We hope that its users will not simply open its pages for a definition or a number, but will take the time to read the extended essays and follow the network of cross-references that mirror the interactions of the planet s components. We believe that the users of this and other volumes of the Encyclopedia will share in our gratitude to the numerous contributors to its contents, and we hope that this unique compendium will to some extent help to illuminate the common future of our species and our planet.

Michael C MacCracken
John S Perry
Editors of Volume One

The Earth System

JOHN S PERRY[1] AND MICHAEL C MacCRACKEN[2]

[1]*Alexandria, VA, USA*
[2]*Lawrence Livermore National Laboratory, Livermore, CA, USA*

Our planet is, as far as we know, unique in its diversity, complexity, capacity to support life, and never-ending change. The very earth beneath our feet, which we perceive as solid, is on geological time scales a boiling uid bearing the wandering continents like foam on a cauldron. The uid envelope in which we live the atmosphere and ocean is in constant motion, driven by the evolving ow of energy from our star, the Sun. The benign climate that nurtures life is governed by the atmosphere s composition, which itself was shaped by life. Earth is a restless planet.

Today, with burgeoning human populations and increasingly complex societies and economies, the consequences of environmental change are growing. Moreover, humanity itself has become a force for change comparable in magnitude to the geological forces of the past. Virtually every characteristic of every corner of the global environment bears the ngerprints of humankind.

This volume is predicated on the concept that the past, present, and future of our global environment can only be fully understood by considering the planet as an integrated whole: the Earth System. However, as in all complex puzzles, this understanding can best be built one piece at a time. Hence, this volume presents a disaggregated assembly of focused articles that provide entry points for the student and reader into the maze of phenomena, processes, and linkages that drive the great environmental machine in which we live. It strives to paint the giant canvas of our planet through a patchwork of brushstrokes by many hands in the hope that like the Earth it depicts the whole will be greater than the aggregate of its parts.

For most of the brief tenure of our species on this planet, we humans have experienced our environment only as a patchwork of snapshots. The world of the Neolithic hunter or the Mesopotamian farmer was circumscribed by his limited experience in his short life. Beyond his hunting territory or his farm, the Earth was a frightening mystery. As our societies and technologies developed, humans moved farther a eld, venturing across oceans and deserts and gradually acquiring some notion of a wider world. Eventually, the planet was circumnavigated, both poles were explored, and virtually every patch of its surface felt human footsteps. The number of individual snapshots in our album of the Earth increased immeasurably, but still largely remained an unconnected assemblage of individual observations.

ONE EARTH, ONE SYSTEM

Today, however, as we begin a new millennium, our view of the Earth is radically different. Past explorations revealed to us the surface of the globe in its in nite variety. A few seminal thinkers, such as the pioneering meteorologist Hadley, drew on fragmentary observations to synthesize larger pictures, for example of the circulation of the global atmosphere. However, our giant leap into space in the concluding decades of the last century revealed, for the rst time, the reality of our world. We could for the rst time see our planet as a single entity, a unique oasis in the inhospitable vastness of space. In the words of MacLeish (1968) (see Figure 1):

> To see the Earth as it truly is, small and blue and beautiful in that eternal silence where it oats, is to see ourselves as riders on the Earth together, brothers on that bright loveliness in the eternal cold brothers who know now they are truly brothers.

This growing intuitive, in some respects spiritual, view of the Earth as a uni ed whole was paralleled in the world of science. Many threads of understanding converged on the concept that, despite its variegated complexity, the Earth s past, present, and future could be fully understood only if all the diverse elements of our planetary environment were considered as a uni ed system. As our horizons in time and space expand, these interactions assume greater and

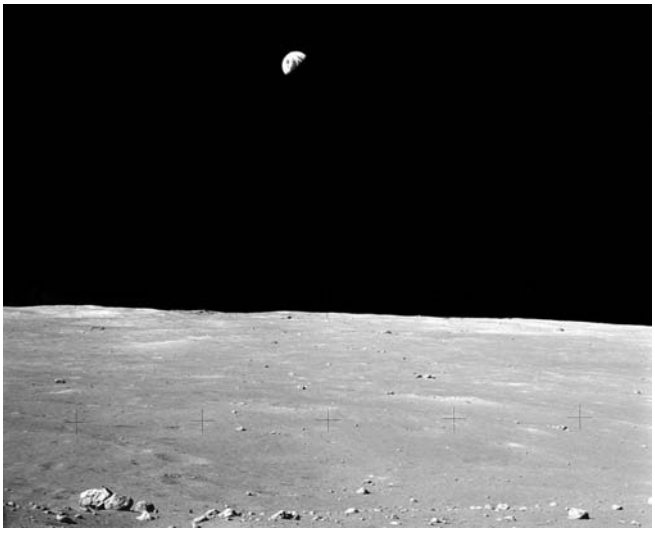


Figure 1 Our Earth, a lone oasis of air, water, and life in the cold desert of space, as seen in December 1968 by the first Apollo mission to orbit the Moon. (*Source* NASA)

greater importance, and the behavior of virtually any single component can be understood only as part of the system as a whole (*see also* **The Human Dimensions of Global Change**, Volume 5).

INSIGHTS FROM THE PAST

One impetus to this view came from study of the Earth s past. As evidence accumulated from geological records, deep-sea sediments, ice cores and other sources of indirect evidence, it became evident that all components of the natural world have been constantly changing and interacting since the planet s birth. The early Earth exuded gases, including water vapor, to form the planet s fluid envelope of air and water. The moving plates of its surface formed continents and mountains that shaped the wind and ocean margins and basins that held the ocean. These provided the cradle for the most powerful force of change – life. Living organisms found ways to harness the Sun s energy to power photosynthetic chemistry that slowly changed the planet s surface, atmosphere, and climate into the equable regime that we know.

EARTH AS GAIA

Indeed, as has been pointed out by Lovelock (1972), an observer from a distant planet could readily detect that Earth is a living planet: the atmospheric abundance of oxygen and paucity of carbon dioxide could not be supported for long by an inert body governed by inorganic chemistry alone. Today s life depends vitally on an oxygen-rich environment created by life itself. Thus, the Earth System inescapably includes the living system. The essential unity of the physical and living systems of the planet in quasi-equilibrium is comparable to the complex web of interactions that preserve the equilibrium of a living organism, a phenomenon known in biology as *homeostasis*.

Whatever the processes, the Earth has, in fact, somehow long maintained a remarkably stable and remarkably temperate environment for life. Indeed, Lovelock (1972) speculated it is illuminating to think of the Earth as a single living creature, managing its interacting systems so as to preserve the conditions for its continued welfare. To emphasize its uniqueness and individuality, he named this hypothetical planetary creature Gaia, after the ancient Greek mother-goddess of the Earth. Lovelock s insight, founded equally on science, poetry, and perhaps a touch of mysticism, has provided a powerfully inspiring vision of the complex and ever-changing system of which we are a part (*see* **Gaia**, Volume 2; **Gaia Hypothesis**, Volume 5).

IMPLICATIONS FOR SCIENCE AND UNDERSTANDING

In more practical terms, the vision of an intricately interconnected, ever-changing world has important implications for science s quest for understanding. If the world were static – if ocean currents, snow and ice boundaries, and vegetation were exactly the same from season to season and from year to year – then we could meaningfully study each component separately and at our leisure. We could combine observations of ocean currents taken centuries ago with those taken yesterday, without taking account of changing winds or rain or river flows. Oceanographers would have no need to learn the history of the ocean s waters and currents from geologists and biologists, nor provide geologists and biologists with insights on how the oceans might have shaped sediments and climate.

The luxury of such splendid isolation is denied to science by the hard facts of our planet s history. The bottom waters of the ocean slowly upwelling to the surface today gained their temperature and physical characteristics from the climate, precipitation, and river flow of a millennium past. The resources of coal and oil upon which our civilization depends are fossil remnants of long-vanished plants that grew on wandering continents under climates markedly different from ours. Thus, the fundamental task of observing our environment can only be accomplished in the framework of an ever-changing Earth System.

An enormous gulf lies, however, between recognizing a system s components and characteristics, and understanding how it works, not to mention predicting its future behavior. How can science come to grips with the complexity of Earth? To say that something is complex and interconnected is to say that it can be considered in many parts, each of which may hopefully be studied individually. Thus, just as an automobile mechanic focuses successively

on one component after another until the engine runs smoothly, scientists focus on individual components of the system – atmosphere, oceans, ice and snow, rivers and lakes, living ecosystems, and so on. Each of these systems presents its own cascade of complexity, but within a delimited and hopefully more tractable scope. No one can hope to understand the Earth as a whole, but it may be possible to understand a raindrop.

This *reductionist approach* is, at once, science s most powerful weapon, and science s most crippling handicap. Deep understanding of tightly delimited systems and processes can be developed only by ruthlessly discarding all that lies outside the spotlight of inquiry, and by organizing scientific effort in terms of rigorous scientific disciplines. A lifetime may not be enough to reach full understanding of a single microbe, or for that matter, a single thunderstorm. But, through the focused efforts of many lifetimes, a well-stocked library of knowledge can be built up. The challenge then is to transform this untidy pile of building materials stored in a host of disciplinary warehouses into a coherent edifice of understanding, to move from piecemeal insights about individual components and processes to a coherent predictive vision of the system as a whole.

Today s science is guided by a vision of the Earth System as an intricate machine composed of numerous components, each potentially a subject for focused research and steadily improving understanding. A schematic diagram of such a framework is presented in Figure 2.

An illustrative example of the interactions between components of the system is provided by the El Niño/Southern Oscillation phenomenon. Meteorologists had long known about long-period swings in pressure across the Pacific. Oceanographers had long known about temperature and circulation patterns in the tropics. Fishermen had long known about irregular and sudden warmings off the coasts of Ecuador and Peru. Field observing programs, powerfully aided by satellite observations, permitted researchers to gain insights into the ocean s effects on tropical weather systems, and the atmosphere s influence on the oceans.

But none of these insights by itself led to clear understanding and useful prediction. Eventually, union of observations, analyses, and insights from atmosphere and ocean, not only in the tropical Pacific, but from all around the globe, led to numerical models and, for the first time, scientifically based and highly successful seasonal predictions. It was found that the essence of the phenomenon, and the essential basis for prediction, lay in the interaction between ocean and atmosphere, not simply in processes and events within either medium.

GLOBAL CHANGE

Although the record of the past is a history of constant change in all aspects of our planetary environment, the last decades of the 20th century brought a new focus on environmental change.

First, advances in our knowledge of the Earth s past brought a new realization of its variability, and of the possibility that these fluctuations seen in the past might reoccur. While in the 19th century Agassiz revealed the outlines of a past ice age, research in the 20th century showed a series of glaciations interrupted by brief interglacial respites much like our own. Indeed, some raised the possibility that the balmy interglacial in which we live, the *Holocene*, might be coming to its natural end after some 10 millennia. The knowledge of past changes, and the possibility of future changes, naturally heightened the need to understand the processes of change itself.

Our knowledge of the present similarly advanced. Observing networks for weather prediction had begun in the 19th century, and grown with the steady expansion of world populations and trade. During World War II and the following years, exciting new tools for probing the atmosphere, ocean, and solid earth emerged, culminating in the deployment of powerful observing systems on Earth-orbiting satellites. For the first time, the planet could be observed as a whole through networks of well-calibrated instruments. Increasingly, these measurements became unified through internationally coordinated observing systems. The resulting databases provided a powerful foundation for research and immeasurably advanced understanding of the Earth System.

At the same time, there was a growing perception that burgeoning human society was becoming ever more sensitive to environmental change. For example, as agriculture spread across the American plains, tornadoes that once stirred grasslands and stampeded buffaloes now tore through airports and trailer parks. As coastal settlements crowded ever closer to the water s edge, storms and storm surges posed ever-greater threats to life and property. As populations increased, and possibilities for migration decreased, variations in rainfall produced human and ecological tragedy in climatically marginal regions such as the African Sahel. Accelerating globalization held economies and food supplies everywhere hostage to environmental changes in distant breadbaskets. Environmental change was no longer an academic curiosity for scholars, but a burning concern for all humankind.

HUMANITY'S HEAVY FOOTPRINTS ON THE EARTH

Perhaps the most pressing and prominent concern for environmental change, however, was the growing realization that human activities were impacting the global environment as powerfully as the geological forces of the past, a prospect foreseen a generation earlier by Vernadsky (1926) and Lotka (1924).

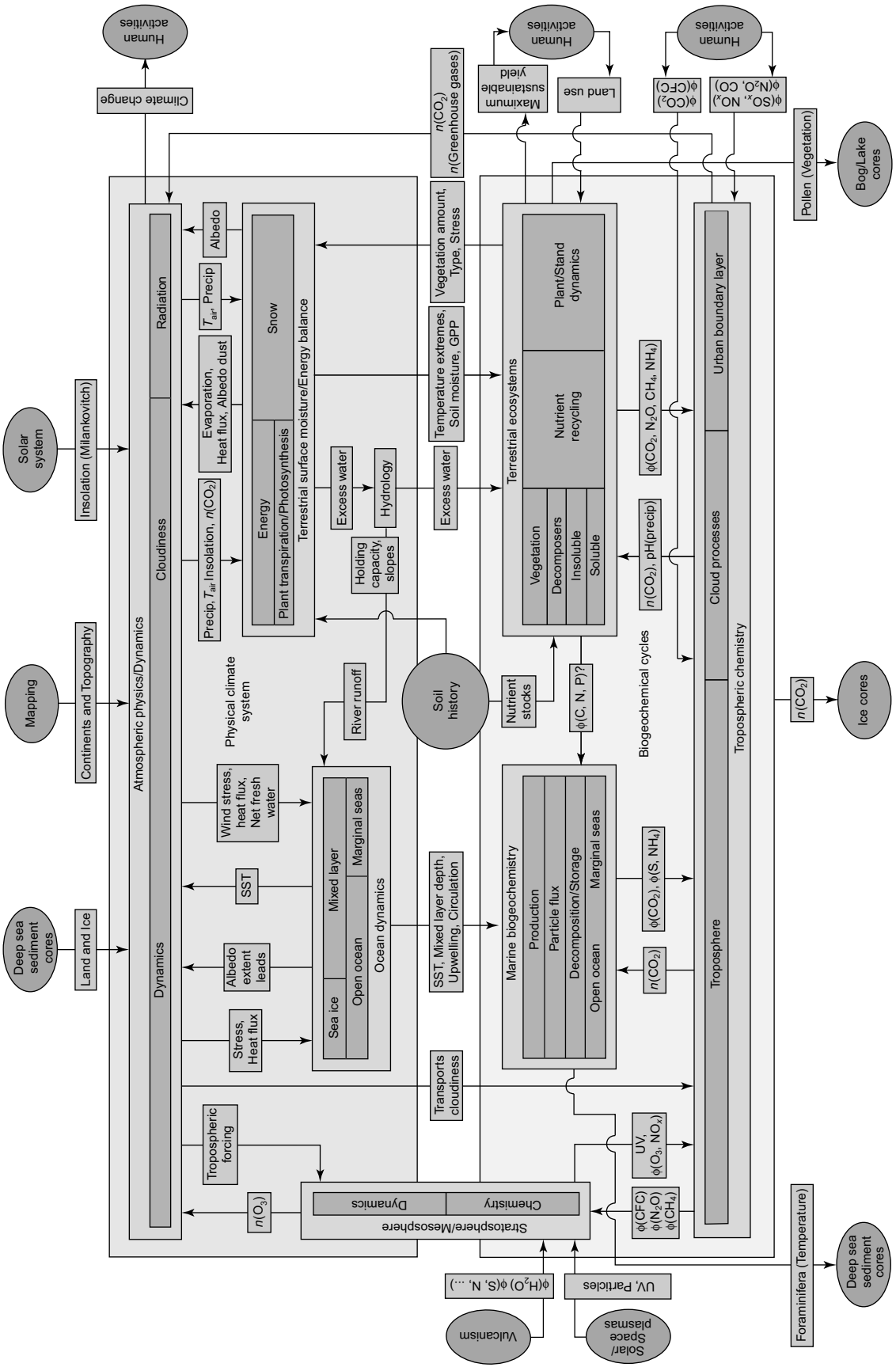


Figure 2 Conceptual model of Earth system processes operating on timescales of decades to centuries. In the figure, $n \bullet x \bullet$ refers to number density and $\phi \bullet x \bullet y \bullet$ indicates functional dependence. (*Source:* NASA, 1988)

In the 19th century, the mechanisms of the natural greenhouse effect that maintains the Earth s equable climate were largely elucidated. The climate to which we are accustomed is maintained by surprisingly small concentrations of a few key gases in the atmosphere – mainly water vapor, carbon dioxide, ozone, and methane – that absorb and radiate thermal energy. It was clear that burning of fossil fuels – coal, oil, gas – released carbon atoms (that had been entombed in eons past) in geological formations. By transforming forests into farms, pastures, and cities, we release additional quantities of carbon dioxide, and we have altered virtually every factor that determines climate, such as the absorption of solar energy, the fluxes of water and energy, the runoff of water into rivers and streams, and the roughness of the surface as encountered by the wind. Numerical models indicate, for example, that removing the forests of the Amazon would reduce precipitation to the point where reforestation would be unlikely.

By the end of the 20th century, quantitative estimates of possible effects of these atmospheric changes on climate had been made. Concerns lay dormant, however, until detailed measurements in mid-20th century showed incontrovertibly that concentrations of key greenhouse gases, principally carbon dioxide, were indeed increasing alarmingly. Data on past concentrations of carbon dioxide obtained from ice cores confirmed that atmospheric concentrations had increased markedly since pre-industrial times. Projections of the impacts of continued increases on climate confirmed the estimates made a half century earlier, heightening apprehension.

At the same time, exquisitely sensitive techniques for measuring trace constituents in the atmosphere were developed. Lovelock found that human-made gases such as the chlorofluorocarbons (CFCs) were detectable throughout the globe, even in the center of the oceans. Questions were also raised about the possible impact of emissions from the burgeoning fleets of aircraft thronging the stratosphere, particularly the fuel-hungry supersonic aircraft then being proposed. These concerns prompted greatly enhanced research efforts in global atmospheric chemistry.

In parallel, changes in stratospheric ozone were noted, and linked to emissions of synthetic chemicals, primarily the CFCs employed in refrigeration and many industrial processes. Discovery of the virtual disappearance of stratospheric ozone over Antarctica in the early Southern Hemisphere spring helped to focus international attention, leading to international agreements to control emissions of harmful chemicals. Fortuitously, a sound scientific basis for these agreements had been laid by almost two decades of intensive research.

Paralleling these changes in the physical environment, life scientists documented massive changes in the living world – increasing extinction of species, introduction of alien species into sensitive environments, changes and losses in natural habitats, and degradation of soils. These issues are treated in Volumes 2 and 3 of this encyclopedia.

Thus, in the final quarter of the 20th century, the concepts of environmental change, and humankind s role in driving change, were high on both the scientific and policy agendas. Often equated to the single issue of climate change, the phenomenon of accelerating change in all components of the global environment is, in fact, a far broader and even more complex challenge to science and society – a challenge that has been addressed by national and international programs focused on global change. This growing context of awareness and concern for the entire spectrum of change in our planetary environment prompts and shapes this *Encyclopedia of Global Environmental Change*.

COMPONENTS OF THE EARTH SYSTEM

This encyclopedia, as a whole, is designed to offer the concerned citizen, student, or general reader, an introduction to the immense range of topics relevant to the issues of global environmental change in all its aspects. The present volume focuses on the physical Earth System – the atmosphere, oceans, solid earth, and the space environment in which our planet is immersed. While striving to present the basic knowledge base needed for understanding, our emphasis is on the elements of the system that exhibit the most marked changes or exert the most significant influence upon change – particularly change on a global scale.

As suggested above, the most prominent concerns at the start of the 21st century center on global climate, and hence much of the volume will deal with the elements of the global climate system. However, the concept of global environmental change, in truth, encompasses a far wider range of environmental factors important to human life, notably changes in the biological and chemical environment. These are in large part treated in the second volume.

In introducing this volume, we have highlighted the need to study the Earth as an integrated system, in order to understand, and hopefully predict, its changes. However, as we noted above, this broad foundation of knowledge is built on a multitude of deep foundations of closely delimited study. Similarly, the reader seeking understanding must find points of entry that address specific questions.

Thus, the articles in this volume dedicated to the notion of an integrated system are highly specific and narrowly focused. They have been designed to provide in-depth information on individual topics, while hopefully recognizing throughout that the elements described are components of an integrated system. In this introductory essay, we attempt to present a rough sketch of the physical Earth. Our subject, and thus the canvas upon which we must paint, is dauntingly large. Hence, we must use the largest brushes and the broadest strokes while striving to produce a reasonable likeness of our planet, leaving to the expert articles of

the volume s body the task of completing the portrait with realism and fidelity.

How can we best disaggregate the intricate machine in which we live? What are the leading figures that must appear on our sketchpad? On the broadest scale of abstraction, the following components provide a basic framework for understanding, and a context for the specifically focused articles that follow.

Our Planet

The ball of rock upon which we live is believed to have coalesced from the debris of our solar system about 4.5 billion years ago. Its raw materials slowly sorted themselves out into the ingredients we know today. The planet is, to a good approximation, a ball a little over 40 000 km in circumference, slightly bulging at the equator because of centrifugal forces arising from its rotation. It lies almost 150 million km away from its sun. On average, its parent star bathes it in a stream of energy at an average rate of about 1366 W m^{-2}, but the energy received at the surface varies markedly over the sphere as a function of latitude, season, and cloud cover. The Earth s mildly elliptical orbit currently takes it closest to the Sun in January and farthest away in July. However, because the Earth s axis of rotation is tilted with reference to the plane of its orbit, the hemispheres receive seasonally varying amounts of radiation. Over much longer times, the parameters of the Earth s orbit slowly change in response to gravitational forces from our sister planets.

Internally, the primordial ingredients have separated into a dense metallic core surrounded by lighter viscous layers surmounted by the crustal plates upon which we live. Slow convective processes bring new materials to the surface, and suck existing crust downward, while the plates themselves migrate over the planet s surface like patches of ice in a lake. Earthquakes and volcanoes mark the boundaries at which the plates collide, which for the most part form a continuous network of mid-ocean ridges spanning the globe. This slowly seething mass receives a steady rain of debris from space, occasionally including comets and minor asteroids. One of the latter is hypothesized to have exterminated the dinosaurs some 65 000 000 years ago; one of the former flattened a Siberian forest in the 20th century.

Inherent in this brutally compressed description is an image of ceaseless change. Over geological time, the continental rafts upon which land-borne life evolved have moved from tropics to poles and split into continental islands, carrying with them both their characteristic geology and their life forms. The slow oscillations of Earth s orbit, and the inclination of its axis, redistribute solar radiation over the seasons and latitudes. On relatively short time scales, its daily rotation rate fluctuates detectably due to friction with the ever-changing atmosphere.

Atmosphere

The lighter ingredients of the primordial Earth became its fluid envelope – its ocean and atmosphere. The composition of the atmosphere evolved markedly over time due to continued emissions, chemical reactions, escape of lighter constituents into space, and finally by the emergence of photosynthetic life in the last 600 million years or so. The atmosphere is today composed primarily of nitrogen and oxygen, with lesser but vital concentrations of water vapor, carbon dioxide, and other trace gases.

On a global scale, the atmosphere is frighteningly thin: thinner than an onion skin over the planet s surface, half of the atmosphere s mass lies within 5.5 km of the surface. Nevertheless, it plays a special role in Earth s story, for it is the atmosphere that provides the principal fuels for life, carbon and oxygen, and maintains the equable climate indispensable to life. A tenuous layer of ozone shields the surface from the Sun s deadly streams of ultraviolet radiation. Without an atmosphere, Earth would suffer a climate as inhospitable as that of the Moon, with an average temperature some 34 °C colder than we now enjoy, and daily temperature swings of hundreds of degrees. In the greenhouse provided by a wispy blanket of gases such as water vapor, carbon dioxide, methane, ozone, and the industrially produced CFC s, much of the heat radiation emitted by the surface is absorbed by the atmosphere. The atmosphere then emits heat radiation back downward to warm the surface along with the heat absorbed by solar radiation. As a result, the Earth s surface is much warmer than that of a naked planet.

The atmosphere, moreover, serves as the working fluid in a giant planetary heat engine whose furnace lies in the tropics facing the Sun and whose radiator is in the polar regions inclined toward space. Due to the planet s rotation, air and the energy it carries, cannot move directly from equator to pole. Instead, the flow sorts itself out into a triad of circulation cells, complicated by moving turbulent eddies. In this complex circulation, energy derived from the Sun drives atmospheric motions that ultimately transport the atmosphere s share of energy from the tropics to the Arctic and Antarctic. En route, this turbulent flow gives rise to a menagerie of weather systems.

Rather arbitrarily, we characterize the conditions brought to us by the atmosphere as weather or climate – the former referring to the day-to-day sequence of sunshine and storm, the latter to the more or less steady annual cycle of the statistical properties of the weather (e.g., average temperature, frequency of precipitation, etc.). Viewed from the perspective of Earth s history, however, we see that the atmosphere supports a continuum of change on all time scales.

The atmosphere is more than simply a heat-driven machine, however. It transports a host of substances across the face of the globe, and provides a cauldron for their transformation. Its turbulent motions carry water from its

reservoir in the oceans, produce clouds and precipitation that intercept and re ect solar radiation, transport and release heat, and provide rain and snow for the thirsty continents. The atmosphere also carries a host of particulate and gaseous materials Saharan dust, hydrocarbons and gases emitted by vegetation, pollution from human industry and transportation all the while transforming them into acid deposition, smog, and sometimes essential nutrients for life. These mostly invisible particles also play a crucial role in the formation of clouds and rain by providing essential nuclei for condensation.

Thus, the atmosphere is without doubt the most dynamic element of the Earth System. It is also the most vulnerable to human in uences, primarily because of its thin and tenuous nature. In terms of sheer mass alone, human emissions into the atmosphere are no longer negligible in uences on this wispy lm of air that clings to our rock in space. Moreover, our equable climate depends crucially on the particular concentrations of a small group of radiatively active greenhouse gases. Directly and indirectly, human activities are markedly in uencing all of these. By burning coal and oil, we return to the atmosphere, over a few centuries, the carbon that was entombed by biological processes over hundreds of millions of years. Industrial and agricultural emissions of gases such as the CFC s and methyl bromide participate in photochemical and radiative processes, and catalytically destroy ozone in the stratosphere, where its presence is important both in climate and in its role as a shield against ultraviolet radiation. On local and regional scales, emissions of hydrocarbons, nitrous oxides, sulfur dioxide and a host of other chemical species contribute to a diverse array of environmental impacts such as increasing near-surface smog episodes that are damaging both to health and agriculture, and lead to acid deposition that damages vulnerable lakes.

These fundamental changes in the all-enveloping and ever-changing planetary atmosphere in which we swim demonstrate vividly the shortcomings of MacLeish s (1968) insight: we humans are no longer simply riders on this planet we are its drivers.

Ocean

Today, about 70% of our planet s surface is covered by water (H_2O) the vast world ocean. The molecules of H_2O of which they are comprised are thought to have been outgassed from the primordial Earth, and retained by its gravity. As noted above, the basins in which the ocean now resides, averaging about 4 km in depth, are products of the tectonic movements that have shaped Earth s crust. Thus, the boundaries of the oceans have constantly changed over time as the continents have drifted, and the ocean basins are bisected by a global system of mid-ocean ridges marking the rifting of the continental plates.

Given humanity s propensity to settle near the water s edge, sea level is a continuing concern. Global sea level is determined rst by the topography of the basins and their margins how deep is the bathtub, how high its rim? On a dynamic Earth, many regions show long-term rises or falls in local sea level. For example, Scandinavia and parts of eastern North America are still rising as they recover from the weight of the great glaciers that last covered much of these regions over 10 000 years ago, and locally perceived changes in sea level re ect these motions. Sea level also depends on the varying amount of water in the oceans. During the great glaciations, for example, much of the Earth s water resided on land as mountains of ice, and sea level was about 150 m lower. Further, the great currents of the ocean are associated with strong gradients in sea level. At the beginning of the 21st century, concerns regarding global sea level center on projected rises resulting from melting of land-borne ice and, more speculatively, increases in the ow of Antarctic ice into the world ocean.

The chemistry of the oceans is also shaped by their solid earth containers. Runoff from the continents carries sediments and dissolved salts into the oceans, feeding the descending plate boundaries and keeping the ocean salty. The presence of salt means that sea water s density is determined by both salinity and temperature.

Earth s domination by the ocean has many consequences. First, water is a unique substance. Chemically, it serves as a solvent and reaction medium for innumerable substances. Thus, the ocean annually absorbs and re-emits vast quantities of atmospheric gases, notably carbon dioxide, through a web of life processes. Fluxes of carbon dioxide into and out of the ocean play a major role in the global carbon cycle that fuels life, and in determining atmospheric concentrations of carbon dioxide.

The ocean absorbs, transports, and re-emits remarkable quantities of heat, particularly as it undergoes phase transformations between ice and liquid, or between liquid and vapor. Approximately one-half the heat transported from the tropics to the poles is moved by the ocean. A number of its physical characteristics crucially determine important aspects of our environment. Because fresh water reaches its maximum density at about 4 °C above its freezing point, a stable top layer of water rapidly cools in winter, and ice can form and oat in lakes. The ocean s salt lowers this temperature of maximum density, and sea water thus cools steadily to a freezing point at about • 1.8 °C. Finally, the sea surface freezes in a slurry of crystals that reject much of the water s salt, providing an additional source of salty, dense water that sinks to the ocean depths.

The ocean s structure is as complex as the atmosphere s. Over most of its extent, a layer averaging about 100 m thick is churned by wind and waves, altered by daily and seasonal heating and cooling, and lightened by freshwater rain and runoff. This mixed layer forms an ever-changing lid over

cooler and more dense waters that only slowly mix with the surface waters, except in areas where the winds blow the surface mixed layer away. Much of the ocean s activity, particularly its biological activity, takes place in coastal regions where rivers and upwelling from the deeper ocean provide essential nutrients.

Today s oceans participate in a gigantic planet-wide circulation of water driven by heating and cooling, density differences due to varying salinity, and the force of the winds, a circulation that Wallace Broecker of the Lamont-Doherty Earth Observatory has termed the *conveyor belt*. This globe-encircling stream has important consequences for climate. The northward-flowing surface current of the western North Atlantic, for example, transports a major fraction of the heat deposited in the tropics by incoming solar radiation to warm Europe, especially in winter. Descending currents in the North Atlantic and the Southern Ocean carry dissolved atmospheric gases deep into the ocean s bottom water, which may not again reach the surface in this millenium. At its northern boundary in the Arctic, the continent-girdled Arctic Ocean is almost entirely covered by multi-year ice, with seasonally varying ice packs extending southwards. Slow circulations carry Arctic waters and ice southward. In the southern polar regions, the giant continent of Antarctica is girdled again by ice whose extent varies markedly over the seasons.

To land-dwellers, the most significant oceanic influences are felt through the atmosphere. On the largest scale, the capacity of the oceans to absorb, transport, and liberate heat, greatly tempers the climates of the continents, particularly regions downwind from the major oceans. Comparison of the climates of the US Pacific Northwest, the Dakotas, Labrador, the British Isles, and Siberia forcibly demonstrates the ocean s influence; all are at about the same latitude, yet their climates differ greatly. It is not surprising, therefore, that shorter-term fluctuations in the ocean have large impacts on global climate. Prominent among these are the irregularly periodic swings in atmospheric circulation and ocean temperature known as the Southern Oscillation, and most vividly evident as the El Niño (often termed together as ENSO). This rapid and massive warming of the eastern tropical Pacific produces torrential rains along the South American coast, and a worldwide chain of climatic anomalies. Arising as it does from interactions between ocean and atmosphere, its understanding and prediction have been a difficult challenge for science. However, observation and research in the latter decades of the 20th century yielded successful predictive models.

Cryosphere

Liquid water dominates the Earth s surface; but frozen water also plays a crucial role in the planet s workings. Unlike other substances found on Earth, ice and snow exist relatively close to their melting point and frequently change phase from solid to liquid and back again, producing dramatic changes in the landscape. Snow and ice march with the seasons across continents and oceans, making the domain of frozen water – the *cryosphere* – a dynamic component of the Earth System.

In terms of spatial extent, seasonal snow cover is the largest single component of the cryosphere. In winter, snow covers about 47 million km^2, principally in the Northern Hemisphere. Due to its influence on energy and moisture budgets, snow cover is an important variable influencing climate. It produces large differences between summer and winter land surface reflectance, or *albedo*, since snow reflects up to 90% of incoming solar energy whereas soil or vegetation may reflect only 10–20%. Snow cover also absorbs much heat as it melts, providing significant thermal inertia in high latitudes.

The largest quasi-permanent accumulations of ice at present are the ice sheets that cover Greenland and Antarctica. The latter, for example, covers about 12 million km^2, and reaches over 4 km in thickness. In Greenland and East Antarctica, the ice rests on solid ground above sea level. On the other hand, most of the West Antarctic Ice Sheet is grounded below sea level. These ice masses result from the slow accumulation of snow in regions of perpetual cold. The resulting domes of ice slowly flow at speeds up to a few km $year^{-1}$ into ice streams and, in Antarctica, into immense thick ice sheets extending into the surrounding ocean. These sheets calve into floating icebergs – sometimes the size of a small country – that populate the high latitudes of the oceans.

Other long-standing ice masses reside in mountain glaciers. Here also, ice accumulations flow down through valley glaciers sometimes many kilometers in length, and carrying large amounts of rocky debris. If they reach the sea, as in Greenland, they provide a source of icebergs.

Large areas of the northern oceans are covered by seasonally varying amounts of sea ice that forms in winter and largely melts back in summer. As ice forms, it rejects salts and entrains air, becomes lighter, and thus forms a thickening cap on the ocean that forms an insulating layer between ocean and atmosphere At the same time, the cold, salty, and dense waters produced by freezing sink to fill the deep ocean and form an important link in the global ocean conveyor system.

In essence, the polar ice caps of our planet thus constitute seasonally expanding and contracting continents of ice that rank among the most dynamic elements of the Earth System. Each winter, for example, the sea-ice around Antarctica grows to roughly twice the size of the continent. In contrast, Arctic sea ice is largely confined to the landlocked Arctic basin, where much of it endures and thickens over several years.

The Earth s ice continents have waxed and waned even more markedly on geological time scales. Ever since the continents drifted into their present northern cluster, great glaciations have swept over much of the planet with a periodicity apparently linked to the Milankovitch orbital cycles. Cores obtained from glacial ice, land deposits, and deep-sea sediments provide records of these glacial advances, and the comparatively brief intervening interglacial periods, such as the Holocene, in which we live, and the preceding Eemian about 125 000 years ago. At the peak of the last glacial advance, about 20 000 years ago, ice masses thousands of meters thick extended deep into Europe and the northern tier of US states, forcing atmospheric circulation into patterns very different from today, and depressing the Earth s very crust. Much of the planet s water was locked up in ice, leaving the ocean about 150 m lower and exposing land bridges between continents.

Today, permanent ice has retreated to mountains and the poles, where it presumably will contribute to seed the next glacial advance. Even in these strongholds, however, ice is not safe from the influence of humankind. The warming climate due to increasing greenhouse gases threatens to shrink the extent of glaciers, permafrost and sea ice. Moreover, increasing amounts of pollutants such as persistent organic compounds, heavy metals, and even radioactive species are transported by winds, currents, and rivers from lower latitudes into polar regions.

Thus, the world of solid water is, in truth, no less fluid and dynamic than the realms of gas and liquid, nor is it immune from the heavy footsteps of human society.

Hydrosphere

Air, ocean, and ice on this particular planet share one vital constituent: water. Earth has rightly been termed the water planet, because of the central role of this unique substance in the workings of its physical system and of the life that it supports.

Water is such a ubiquitous substance that its remarkable properties and myriad uses are easy to overlook. It dissolves innumerable substances, and forms a medium for the chemistry that shapes our world and its life – water erodes our monuments and pumps through our own veins. Life mostly lives in the temperature band between water s freezing and boiling points. Indeed, water is the only common substance found naturally in solid, liquid, and gas forms. Water vapor is a key architect of the natural greenhouse effect that maintains Earth s climate. The large amounts of energy absorbed and released in its heating, cooling, and changes of phase, permit it to transport heat efficiently and make it a powerful buffer for heating and cooling. The Earth as we know it is, in great measure, a product of this single substance that defines the hydrosphere – the water-inhabited region extending from the deepest ocean to the tops of the highest clouds.

The oceans cover about 70% of the planet s surface and contain about 97% of the planet s water. The remaining 3% occurs as fresh waters, three-quarters of which are locked up in the cryosphere. Most of the remaining fresh water is groundwater held in soils and rocks; less than 1% of it occurs in lakes and rivers. Water vapor in the atmosphere is negligible in mass, but is vital to the hydrologic cycle that supports our environment and life.

The hydrologic cycle transfers water from the oceans through evaporation into the atmosphere, thence to the continents, and finally, over and beneath the land surface, back to the oceans. Overland, the path of the water involves a complex set of processes: precipitation, evaporation, interception, transpiration, infiltration, percolation, and runoff. Some of these processes extend throughout the entire hydrosphere, the layers of the atmosphere, and the crust within which water may be found.

About 30% of the solar energy that reaches the Earth from the Sun is expended on evaporating water from the oceans. This condenses into clouds, rain, snow, and dew. Water fuels storms, and by freezing into ice crystals, is responsible for separating electrical charge, and producing lightning. Rain irrigates the land, replenishes subterranean aquifers, chemically weathers rocks, erodes and shapes landscapes, nourishes life, and fills the rivers, which carry dissolved chemicals and sediments back into the oceans. The green mantle of life on our planet acts as a buffer and regulatory valve, capturing water and gradually releasing it to runoff and returning it into the atmosphere by evapotranspiration.

Water also plays a vital role in the geochemical cycles such as the carbon cycle. Calcium is weathered from continental rocks and is then returned to the oceans, where it combines to form calcium carbonates (such minerals constitute the shells of marine life). Eventually the carbonates are deposited on the seafloor and form limestones. Some of these carbonate rocks are later dragged deep into the Earth s interior and melted, producing carbon dioxide that can vent back into the atmosphere, mostly through volcanoes. These cycles of water, carbon dioxide, and oxygen through physical and biological systems are fundamental in maintaining the balance of the planetary environment.

Human activities influence this cycle in both obvious and subtle ways. By replacing forests with pastures and pavements, we speed the flow of runoff into rivers and streams, with resulting erosion and flooding. Waste products leach into aquifers, while exploitation of ground water drains lakes and wetlands. Emissions such as sulfur dioxide change air chemistry, producing acidic rain and deposition. Changes in atmospheric aerosols due to land use and industrial activity alter the processes of cloud formation and precipitation, changing rainfall and cloud reflectivity. Humans play an increasing role in

the processes of the hydrosphere, both as actors and victims. *See also* Circulating Freshwater: Crucial Link between Climate, Land, Ecosystems, and Humanity, Volume 5.

BIOSPHERE

From the standpoint of our species, the most important layer of our planet s shell is the biosphere – the realm in which life is possible. But, as we have seen, life permeates all elements of our global environment; life crucially depends on its environment; and life powerfully shapes that environment. A planet different from ours would have no life; a planet without life would be far different from Earth. The realm of life, the biosphere, is treated in Volume 2 of this encyclopedia; however, let there be no doubt that life is an integral part of this living planet. As Vernadsky (1926) realized, Living matter is no accidental creation, but forms an essential part of the structure of the Earth .

CONCLUSION

Throughout this whirlwind tour of our planet, several overarching themes have emerged.

Diversity and Complexity

At their root, all the workings of the global environment are governed by a small number of fundamental laws of physics and chemistry. However, the concrete manifestations of these laws are infinitely complex and diverse. A lifetime may not be enough to understand fully a brief storm, let alone longer-term glacial cycling. Somehow we must piece together an understanding of our planet from fragmentary, crude, and imperfect insights into the infinite variety of our world. We are like students rummaging through the pages written in a gigantic dictionary of some foreign tongue – the words may be important, but many of them and the grammar that interconnects them remain unknown and undefined.

Interconnection and Interaction

We have seen that it is virtually impossible to discuss any component of our planet without speaking of other elements. The winds are driven by heating and cooling, but the oceans are the main reservoirs of heat. Snow and ice are creatures of the weather, but also powerfully influence global climate. The land shapes the chemistry and motions of air and ocean, but winds, currents, rain, and chemistry over time shape the land. Above all, life depends on air, soil, and water, but life is in itself among the most powerful agents of change.

Dynamics and Change

Since its beginnings, Earth has been a restless place. The planet has seen a variety of atmospheres, climates far warmer than today s, great glacial periods, and ever-wandering continents. On all time scales from a morning to an eon, Earth changes, and all the components of its planetary machine adjust to new patterns. Thus, study of our planet can never be a static science. Dynamics and change are intrinsic to Earth science, not optional.

With complexity, interconnection, and change comes chaos. As Lorenz elegantly demonstrated, even the simplest system embodying these characteristics has a distinct limit to its predictability because any error, however small, in any component grows in magnitude with time and effects a chain of consequent changes in other components of the system. Within broad existential constraints, every page of Earth s history is thus only one of many possible pages that could have been written by the hand of time. Nevertheless, the constraints imposed on the Earth system by the physical laws governing conservation of mass, momentum, energy, and species do create a limiting envelope that constrains the system s changes. These limits can be used to help project how the range of conditions is likely to change in response to various influences. Looking to the future, we can thus map out only the broadest features of the landscape through which we and our planet will blaze a single trail.

Inherent in this existential reality is the possibility of surprise. We cannot assume that the pace of change will always be measured and smooth. Quite to the contrary, the record of the past shows innumerable examples of rapid, sometimes catastrophic change. Our Neolithic ancestors in Europe, for example, saw their climate abruptly shift from increasing warmth to near-glacial cold in perhaps a decade. We cannot assume that our path along the road of time will never stray into a pothole or two.

Human Influence and Impact

Permeating this diverse, complex, interlinked, dynamic, and chaotic system is a relatively new source of change. Prehistoric hunters decimated large herbivores, set fires to create savannahs, and markedly changed the landscapes of the continents. Agriculture transformed still more of the planet s surface, and significantly altered flows of water, sediments, and chemicals. Finally, industry mobilized vast quantities of fuel and materials for human purposes, dispersing wastes into earth, water, and atmosphere, influencing chemical and biological processes from the depths of the ocean to the thinnest reaches of the atmosphere – and even threatening the climate itself. Truly, as Vernadsky foresaw, humankind has become a new geological force. See Volume 3 for many examples of human impacts on the environment.

THE EARTH SYSTEM – OUR COMMON HOME

Thus, our planet is composed of many complex parts; these parts interact in complex ways; they all continually change in response to each other; and they are all being increasingly disturbed by human actions. Conceptually, climate represents the average behavior of a global heat engine that is driven by the Sun, utilizing air and water as working fluids, with strong influences from natural and human activities. Forecasting weather beyond a few days requires atmospheric observations over the entire Earth. Predicting a season ahead requires knowledge of the world s oceans. Assessing a century s change in climate requires projecting the development of human society and economy. We, therefore, have no choice but to understand and learn to deal with Earth in all its parts and with all its passengers as a system – a regularly interacting or interdependent group of items forming a unified whole.

The only truly comprehensive encyclopedia of this changing Earth is the Earth itself. This volume comprises simply a collection of snapshots of the planet taken with diverse lenses, centers of attention, and points of view. We hope, however, that these will provide useful stepping-stones toward an understanding of the complex dynamic system of which we are an integral and increasingly important part. As Martin Luther King said, We are caught in an inescapable network of mutuality, tied in a single garment of destiny. Whatever affects one directly, affects all indirectly . Increasingly true of the human condition, this mutuality has always been true of the Earth System in which humanity exists.

REFERENCES

Lotka, A J (1924) *Elements of Physical Biology*, Williams and Wilkins, Baltimore, MD, reprinted 1956, Dover, New York.

Lovelock, J (1972) Gaia as seen Through the Atmosphere, *Atmos. Environ.*, 6, 579–580.

MacLeish, A (1968) *The New York Times*, December 25th.

NASA (1988) *Report of the Earth System Sciences Committee*, NASA Advisory Council, National Aeronautics and Space Administration, Washington, DC.

Vernadsky, V I (1926) *The Biosphere*, reprinted by Synergetic Press, Oracle, AZ, 1986.

13

Earth System Processes

KEVIN TRENBERTH

National Center for Atmospheric Research/Climate Analysis Section, Boulder, CO, USA

We experience weather every day in all its incredible variety. Most of the time it is familiar, yet it never repeats exactly. We also experience the changing seasons and associated kinds of weather. In summer, fine sunny days may be interrupted by outbreaks of rain and thunderstorms. Outside of the tropics as winter approaches, the days get shorter, it gets colder and the weather typically fluctuates from warmer and fine spells to cooler and rainy or maybe snowy conditions. In the tropics, the seasonal variations are more often experienced as monsoonal fluctuations between a wet season and a somewhat longer dry season. These seasonal changes are the largest climate changes we experience at any given location. Because they arise in a well understood way from the regular orbit of the Earth around the Sun, we expect them, and we look forward to them. We plan summer vacations and winter ski trips accordingly. Farmers plan their crops and harvests around the seasonal cycle. By comparison, variations in the average weather from one year to the next are quite modest, and longerterm changes in climate occurring over decades or human lifetimes may be even smaller. Nevertheless, these variations can be very disruptive and costly if we do not expect them. Climate changes, lasting decades to millennia, have occurred in the past as a result of various natural influences. Interannual variations are also an important ingredient of climate and can arise through, for example, interactions between the atmosphere and the oceans, as is the case with El Niño. This article discusses weather and climate variations in the context of the Earth system as a whole, and provides a basis for understanding the reasons why climate may vary, and how these variations may be manifested in terms of weather. The main focus of this article is on the atmosphere, which is the most variable component of the Earth system; it is also where we live and it provides the air we breathe. But the atmosphere interacts with the oceans, the land surface and its vegetation, and with the other components of the climate system, so those too are important, even from this perspective. Their role in climate is also addressed in this article. Like the oceans, the atmosphere is a global commons (Soroos, 1997) (see ***Commons, Tragedy of the***, *Volume 5). It is globally connected and air that is over one nation can easily lie over another on the next continent a day later. Recent attempts by manned balloons to circumnavigate the globe have dramatically shown how the winds can carry a balloon half way around the world in a week or less and that air currents can often take the balloon in unwanted directions. So the atmosphere belongs to no one nation; rather all nations may use it for their own purposes (such as discharging pollution into it) and thus it is also subject to abuse and the phenomenon known as the "tragedy of the commons" (Hardin, 1968) in which the best interests of an individual or individual nation may conflict with the health of the commons itself. Human influences on climate change, often referred to as "global warming," are therefore also discussed. The Earth System includes many other important processes and phenomena that are covered elsewhere in this Encyclopedia. Even in the atmosphere alone, environmental problems include ozone depletion and the "ozone hole," acid rain, air quality and pollution, and, a few decades ago, radioactivity and atomic bomb test debris. Other problems exist in the oceans and on land, such as biodiversity, deforestation, desertification, exploitation of water resources and fisheries, and so on. Many of these environmental problems can be exacerbated by climate change, so that it is the intersection of these problems and climate change that will make for major challenges in the years ahead.*

THE EARTH AND CLIMATE SYSTEM

Our planet orbits the Sun once per year at an average distance of 1.50×10^{11} m. Looking directly at the Sun, the Earth receives an average radiation of 1368 W m^{-2} at this distance. This value is referred to as the total solar irradiance; this value used to be called the "solar constant" even though it does vary by small amounts with

the sunspot cycle and related changes on the Sun (*see* Solar Irradiance and Climate, Volume 1). The Earth s shape is close to that of an oblate spheroid, with an average radius of 6371 km, but it varies from 6378 km at the equator to 6357 km at the poles. The Earth s daily rotation is about an axis with a current tilt, relative to the ecliptic plane, of 23.5°. The Earth s annual orbit around the Sun is slightly elliptical, presently bringing the Earth closest to the Sun on January 3rd (called *perihelion*). The rotation gives rise to the seasons because the Northern Hemisphere (NH) points more toward the Sun in June while it is the Southern Hemisphere s (SH) turn in late December.

The Earth also turns on its axis once per day with an angular velocity of $7{\bullet}292 \times 10^{-5}\,s^{-1}$. This rotation gives us the day–night cycle. A consequence of the Earth s roughly spherical shape and the rotation is that the average solar radiation received at the top of the Earth s atmosphere is the solar constant divided by 4, which is the ratio of the Earth s surface area ($4\pi a^2$, where a is the mean radius) to that of the cross-section (πa^2). On timescales of tens of thousands of years, the Earth s orbit slowly changes, the shape of the orbit is altered, the tilt changes, and the Earth precesses on its axis like a rotating top, all of which combine to alter the latitudinal and seasonal distribution of solar radiation received by the Earth (*see* Orbital Variations, Volume 1). These variations have been associated with the Earth s glacial cycling (*see* Milankovitch, Milutin, Volume 1).

The rotation of the Earth provides a centrifugal force outward, away from the axis of rotation. This force is greatest on the equator at the Earth s surface. Meanwhile, the equatorial bulge in the Earth s shape, which presumably arises from the rotation, provides a greater gravitational attraction than at high latitudes. Thus, it is not a coincidence that the effective net gravity at the surface, which is a combination of the actual gravitational attraction and the centrifugal force, is almost constant and at right angles to the surface, thereby allowing the Earth to be treated as a sphere for many purposes.

The Earth System can be altered by effects or influences from outside the planet, usually regarded as externally imposed. Most important are the Sun and its output, the Earth s rotation rate, Sun–Earth geometry and the slowly changing orbit, the physical make up of the Earth system such as the distribution of land and ocean, the geographic features on the land, the ocean bottom topography and basin configurations, and the mass and basic composition of the atmosphere and ocean. These components determine the mean climate, which can be affected by variations due to these natural causes. For example, a change in the average net radiation at the top of the atmosphere due to perturbations in the incident solar radiation or the emergent infrared radiation leads to a change in heating. Changes in the net incident radiation energy at the Earth s surface can occur from the changes internal to the Sun or, for example, from changes in atmospheric composition such as may arise from natural events like volcanoes, which can create a cloud of debris that blocks a portion of the incoming solar radiation. Other forcings that might be regarded as external include those arising as a result of human activities.

The internal interactive components in the climate system (Figure 1) include the atmosphere, the oceans, sea ice, the land, snow cover, land ice, and fresh water reservoirs. The greatest variations in the composition of the atmosphere involve water in various phases, which include water vapor, clouds of liquid water, ice crystal clouds, and rain, snow, and hail (*see* Hydrologic Cycle, Volume 1). However, other constituents of the atmosphere and the oceans can also change, thereby bringing considerations of atmospheric chemistry, marine biogeochemistry, and land surface exchanges into climate change.

The atmosphere is the most volatile part of the climate system; for example, winds in jet streams at about 10 km (32800 feet) altitude often exceed $50\,m\,s^{-1}$ (112 miles per hour) and sometimes even double these values. Changes in weather can occur in just a few hours. The atmosphere is quite a thin envelope around the planet, with 90% of its mass of $5{\bullet}1 \times 10^{18}$ kg within about 16 km (roughly 10 miles) of the surface (about one four hundredth the radius of the Earth). This relative shallowness of the Earth s atmosphere often allows the atmospheric motions to be considered as occurring at the Earth s surface. For example, satellite photographs of Earth appear to show clouds hugging the surface.

The atmosphere is composed of several fairly distinct layers. Immediately above the surface is the troposphere, which extends upwards to about 10 km in the extratropics and 16 km in the tropics. About 80–90% of the atmosphere is contained in the troposphere. It is defined as the layer where the temperature generally decreases with altitude and is the region where most of what we call weather occurs. The troposphere is the region where vertical air movements occur and produce clouds and precipitation in rising air and clear skies in gently subsiding air.

The second layer is the stratosphere, which extends to about 50 km in altitude. Because this is a region where temperatures generally increase with height, it is very stable, resists vertical motions, and is highly stratified (as the name implies). Many jet aircraft fly in the lower stratosphere in the relatively non-turbulent conditions above tropospheric weather systems. In this region, oxygen molecules, consisting of two oxygen atoms, are broken apart by intense solar radiation in the ultraviolet and participate in a process to form ozone molecules, in which three atoms of oxygen are combined. So the main ozone layer is contained in the stratosphere (*see* Stratosphere, Chemistry, Volume 1). Its presence shields the surface from harmful ultraviolet rays, making it possible for life to exist (*see* Ultraviolet Radiation, Volume 1).

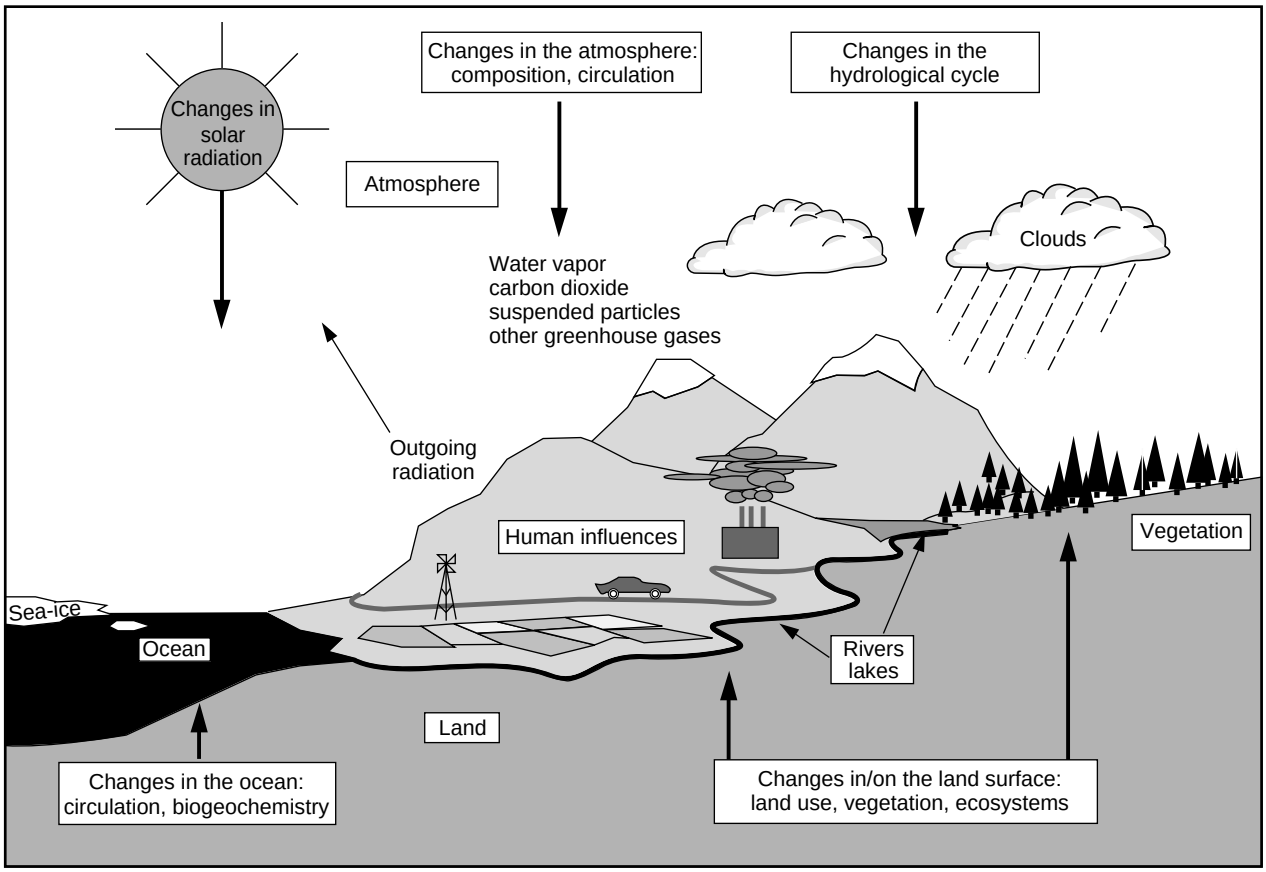


Figure 1 Schematic view of the components of the global climate system (bold), their processes and interactions (thin arrows) and some aspects that may change (bold arrows). (Adapted from Trenberth *et al.,* 1996)

Moving upward, the third atmospheric layer is the mesosphere, which extends to roughly 80 km with temperatures decreasing with altitude. Above the mesosphere is the thermosphere, in which the temperature again increases with height. The thermosphere contains only 0.01% of the atmosphere and it is in this region that atmospheric gases are ionized by energetic and short ultraviolet solar radiation, so that it includes most of the ionosphere. The ionosphere is particularly important because it influences radiowave transmission.

The other main fluid component of the climate system is the oceans. The oceans contain slightly more than 97% of the roughly $1{\bullet}3 \times 10^9\,\mathrm{km}^3$ of water on Earth. They cover 70.8% of the surface, although with a much greater fraction in the SH (80.9% of the area) than the NH (60.7%), which has substantial implications for climate. Ocean currents can be $\bullet\ 1\,\mathrm{m\,s}^{-1}$ in strong currents like the Gulf Stream, but are more typically a few $\mathrm{cm\,s}^{-1}$ at the ocean s surface. The average depth of the ocean is 3795 m. The oceans are stratified opposite to the atmosphere, with warmest waters near the surface. The cold deep abyssal ocean turns over only very slowly on time scales of hundreds to thousands of years.

Other major components of the climate system include sea ice, the land and its features (including the vegetation, albedo (reflective character), biomass, and ecosystems), snow cover, land ice (including the semi-permanent ice sheets of Antarctica and Greenland and glaciers), and rivers, lakes and surface and subsurface water. The components in the climate system (Figure 1) are shown together with the main interactions and sources of climate change.

WEATHER AND CLIMATE

For the Earth, on an annual mean basis, the excess of incoming solar radiation over outgoing longwave radiation in the tropics, and the deficit at mid to high latitudes (Figure 2), sets up an equator-to-pole temperature gradient that results, with the Earth s rotation, in a broad band of westerlies in each hemisphere in the troposphere. Embedded within the mid-latitude westerlies are large-scale weather systems which, along with the ocean, act to transport heat polewards to achieve an overall energy balance, as described below.

In the atmosphere, phenomena and events are loosely divided into the realms of weather and climate . The

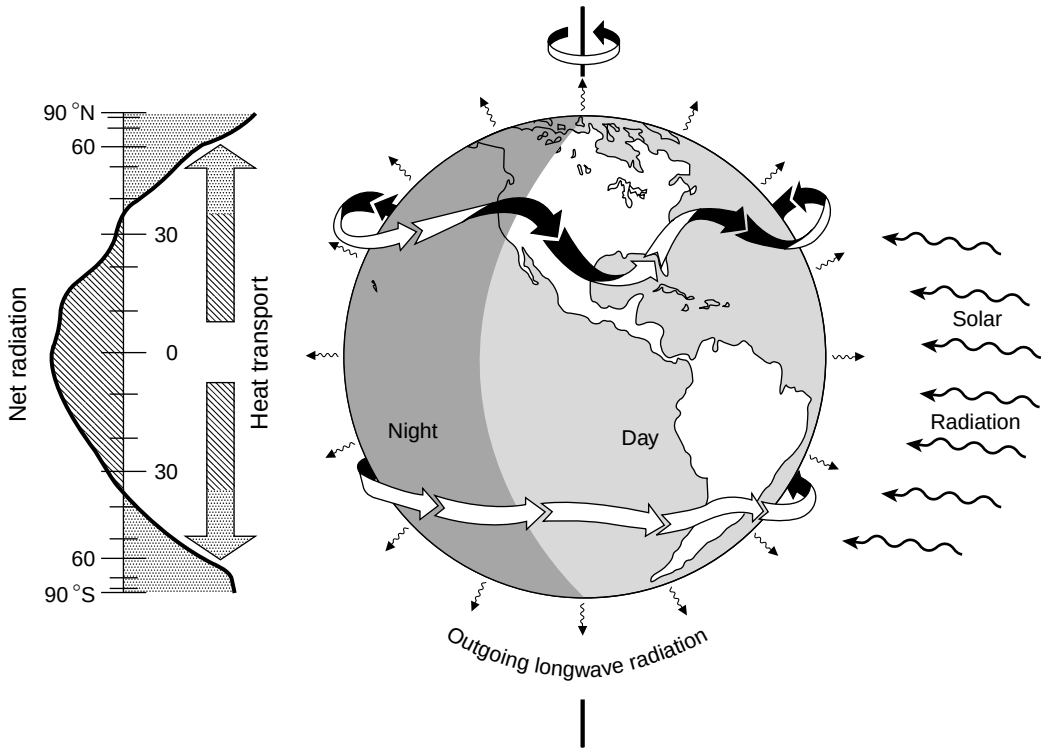


Figure 2 The incoming solar radiation (right) illuminates only part of the Earth while the outgoing longwave radiation is distributed more evenly. On an annual mean basis, the result is an excess of absorbed solar radiation over the outgoing longwave radiation in the tropics, while there is a deficit at middle to high latitudes (left), so that there is a requirement for a poleward heat transport in each hemisphere (broad arrows) by the atmosphere and the oceans. The radiation distribution results in warm conditions in the tropics but cold at high latitudes, and the temperature contrast results in a broad band of westerlies in the extratropics of each hemisphere in which there is an embedded jet stream (shown by the ribbon arrows) at about 10 km above the Earth's surface. The flow of the jet stream over the different underlying surface (ocean, land, mountains) produces waves in the atmosphere and geographic spatial structure to climate. [The excess of net radiation at the equator is 68 W m^{-2} and the deficit peaks at -100 W m^{-2} at the South Pole and -125 W m^{-2} at the North Pole; from Trenberth and Solomon, 1994]. (From Trenberth *et al.*, 1996)

large fluctuations occurring in the atmosphere from hour-to-hour and day-to-day constitute the weather. Weather is described by such elements as temperature, air pressure, humidity, cloudiness, precipitation of various kinds and winds. Weather occurs as a wide variety of phenomena ranging from small cumulus clouds to giant thunderstorms, from clear skies to extensive cloud decks, from gentle breezes to gales, from small wind gusts to tornadoes, from frost to heat waves and from snow flurries to torrential rain. Many such phenomena occur as part of much larger-scale organized weather systems that consist, in middle latitudes, of cyclones (low pressure areas or systems) and anticyclones (high pressure systems) and their associated warm and cold fronts (*see* Atmospheric Motions, Volume 1). Tropical storms (referred to as hurricanes if exceeding certain intensity (64 knots or 74 miles per hour) in many regions or typhoons in the western North Pacific) are organized, large-scale systems of intense low pressure that occur in low latitudes (*see* Hurricanes, Typhoons and other Tropical Storms - Dynamics and Intensity, Volume 1). Weather systems migrate, develop, evolve, mature, and decay over periods of days to weeks and constitute a form of atmospheric turbulence. These weather systems arise mainly from atmospheric instabilities driven by heating patterns from the Sun and their evolution is governed by nonlinear chaotic dynamics, so that they are not predictable in an individual deterministic sense beyond about two weeks into the future (*see* Chaos and Predictability, Volume 1). Examples of weather systems are the cyclones and anticyclones and associated cold and warm fronts, which arise from the equator-to-pole temperature differences (Figure 2) and thus have their origin in the Sun–Earth geometry and the distribution of solar heating. The atmosphere responds by continual attempts to reduce those temperature gradients by producing, in the NH, southerly winds to carry warm air polewards and cold northerly winds to reduce temperatures in lower latitudes, while in the SH, the southerlies are cold and the northerly winds are warm. Another example is convection, which gives

rise to the clouds and thunderstorms, driven by solar heating at the Earth s surface, that produce buoyant thermals, which rise, expand and cool, and produce rain along the way.

Climate is usually defined to be average weather and thus is thought of as the prevailing weather, which includes not just average conditions but also the range of variations. It is often described in terms of the mean and other statistical quantities that measure the variability. Climate extends over a period of time and possibly over a certain geographical region. Climate involves variations in which the atmosphere is influenced by and interacts with other parts of the climate system and the external forcings (Figure 1).

It is the sum of many weather phenomena that determines how the large-scale general circulation of the atmosphere works (i.e., the average three dimensional structure of atmospheric motion); and it is the circulation that essentially determines climate. This intimate link between weather and climate provides a basis for understanding how weather events may change as the climate changes. There are many very different weather phenomena that can take place under an unchanging climate, so a wide range of conditions occurs naturally. Consequently, even with a modest change in climate, many if not most of the same weather phenomena will still occur after the change.

This latter point can be illustrated by consideration of the mean and natural variability, and the effects of a change in the mean while leaving the variability unchanged. Many atmospheric variables, such as temperature, follow a normal (bell-shaped) frequency distribution quite closely. An example for temperature is shown in Figure 3 for a simple case where the bell-shaped distribution of anomalies of temperature, defined as departures from the mean annual cycle, is shifted to correspond to a warmer climate. For any change in mean climate, there is likely to be an amplified change in extremes. If the mean temperature is 15 °C and the standard deviation is 5 °C, 95% of the values fall within plus or minus two standard deviations, or between 5–25 °C, on a given day. Then, if the mean temperature is increased by 2.5 °C but with the same variability, there is only a small change in occurrence of temperatures near the mean. The biggest percentage change is for the extremes: the frequency of occurrence of temperatures above 25 °C increases by well over 100% while similar decreases occur

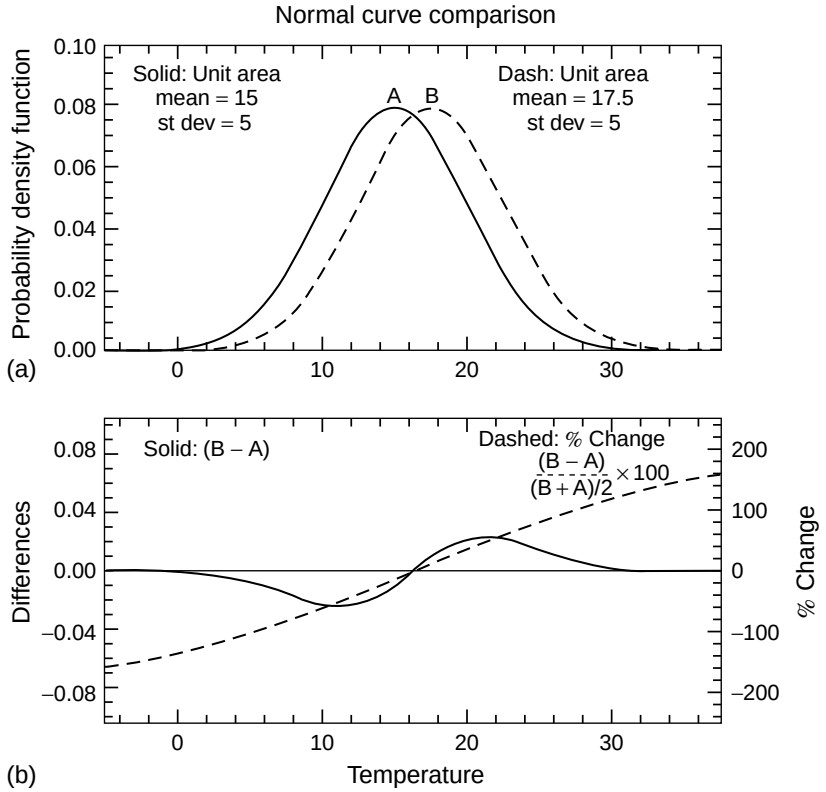


Figure 3 The two bell curves in (a) show the distribution of a climate variable (temperature in this case) before and after a climate change. The units are given in degrees Celsius. In this example, the climate change is assumed to shift the mean temperature by half a standard deviation (or 2.5 °C) but with no change in the spread of the curve (or equivalently the standard deviation, which is set to 5 °C). In (b), the differences in frequency of occurrence that result are shown as actual values (left axis) and as percentages (right axis)

for temperatures below 5 °C. The wide range of natural variability associated with day-to-day weather is the reason that we are unlikely to notice small climate changes except for the extremes.

Changes in any of the climate system components, whether internal and thus a part of the system, or from the external forcings, cause the climate to vary or to change. Thus climate can vary because of alterations in the internal exchanges of energy or in the internal dynamics of the climate system. Examples are El Niño and La Niña events, which arise from natural coupled interactions between the atmosphere and the ocean centered in the tropical Pacific. As such they are a part of the year-to-year climate and they lead to large and important systematic variations in weather patterns (events such as floods and droughts) throughout the world. Often, however, climate is taken to refer to much longer time scales – the average statistics over a 30-year period is a widespread and long-standing working definition. On these longer time scales, ENSO events vanish from mean statistics but become strongly evident in the measures of variability, such as the extremes. However, the mean climate is also influenced by the variability. These considerations become very important in the development of models of the climate system designed to serve as tools to simulate and project climate change.

THE DRIVING FORCES OF CLIMATE

The Global Energy Balance

The shortwave radiation from the Sun provides the source of energy that drives the climate. Much of this energy is in the visible part of the electromagnetic spectrum, although some incoming radiation extends beyond the red part of the spectrum into the infrared and some extends beyond the violet into the ultraviolet. As noted earlier, because of the roughly spherical shape of the Earth, at any one time half the Earth is in night (Figure 2). Because solar radiation is coming from just one direction, the average amount of energy incident at the top of the atmosphere is one quarter of the solar constant, or about 342 $W m^{-2}$. About 31% of this energy is scattered or reflected back to space by molecules, tiny airborne particles (known as aerosols) and clouds in the atmosphere, and by the Earth s surface. This leaves about 235 $W m^{-2}$ on average to warm the Earth s surface and atmosphere (Figure 4).

To balance the incoming energy, the Earth itself must radiate, on average, the same amount of energy back to space (Figure 4). It does this by emitting thermal longwave radiation in the infrared part of the spectrum. The amount of thermal radiation emitted by a warm surface depends on its temperature and on how absorbing it is. For

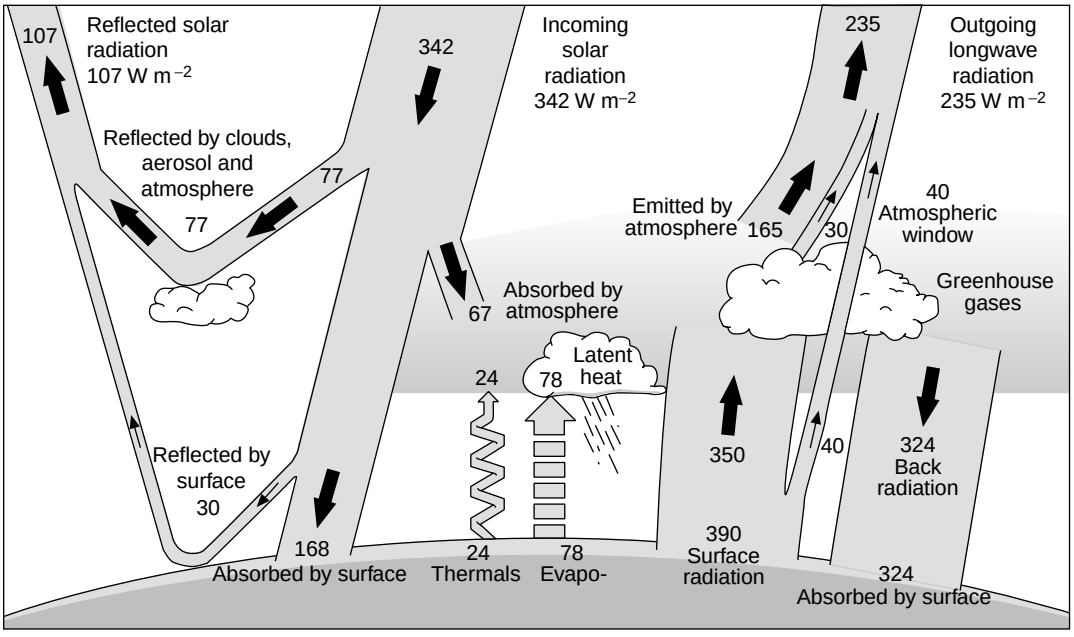


Figure 4 The Earth's radiation balance. The net incoming solar radiation of about 342 $W m^{-2}$ is partially reflected by clouds and the atmosphere, or at the surface, but 49% is absorbed by the surface. Some of that heat is returned to the atmosphere as sensible heat and most as evapotranspiration that warms the atmosphere by release of latent heat during precipitation. The rest of the energy absorbed at the surface is radiated upward as thermal infrared radiation, most of which is absorbed by the atmosphere and re-emitted both up and down, producing a greenhouse effect because the radiation lost to space comes from cloud tops and parts of the atmosphere that are much colder than the surface. (From Kiehl and Trenberth, 1997)

a completely absorbing surface to emit 235 W m^{-2} of thermal radiation, it would have a temperature of about $-19\,^{\circ}$C. This is much colder than the conditions that actually exist near the Earth s surface, where the annual average global mean temperature is about 14 °C. However, because the temperature in the troposphere falls off quite rapidly with height, a temperature of $-19\,^{\circ}$C is reached typically at an altitude of about 5 km above the surface in mid-latitudes. This provides a clue about the role of the atmosphere in making the surface climate warmer and thereby hospitable (*see* Energy Balance and Climate, Volume 1).

The Greenhouse Effect

Some of the infrared radiation leaving the atmosphere originates at or near the Earth s surface and is transmitted relatively unimpeded through the atmosphere; this is the radiation from areas where there are no clouds and which is emitted in the part of the spectrum known as the atmospheric window (Figure 4). The bulk of the radiation emitted from the surface, however, is intercepted and re-emitted both up and down. The emissions to space occur either from the tops of clouds at different atmospheric levels (which are almost always colder than the surface), or by gases present in the atmosphere that absorb and emit infrared radiation. Most of the atmosphere consists of nitrogen and oxygen (composing about 99% of dry air), which are transparent to infrared radiation. It is the water vapor, which varies in amount from 0 to about 3%, carbon dioxide and some other minor gases present in the atmosphere in much smaller quantities that absorb some of the thermal radiation leaving the surface and re-emit radiation from much higher and colder levels out to space. These radiatively active gases are known as greenhouse gases because they act as a partial blanket for the thermal radiation from the surface and enable the surface to be substantially warmer than it would otherwise be, analogous to the effects of a greenhouse. (While a real greenhouse does work this way, the main heat retention in a greenhouse comes through protection from the wind.) This blanketing is known as the natural greenhouse effect. In the current climate under clear sky conditions, water vapor is estimated to account for about 60% of the greenhouse effect, carbon dioxide 26%, ozone 8% and other gases 6% (Kiehl and Trenberth, 1997) (*see* Greenhouse Effect, Volume 1).

Effects of Clouds

Clouds also absorb and emit thermal radiation; as a result, they have a blanketing effect similar to that of the greenhouse gases. However, clouds are also bright reflectors of solar radiation and thus also act to cool the surface. While on average there is strong cancellation between the two opposing effects of shortwave and longwave cloud heating, the net global effect of clouds in our current climate, as determined by space-based measurements, is a small cooling of the surface. A key issue is how clouds will change as climate changes. This issue is complicated by the fact that clouds are also strongly influenced by particulate pollution, which tends to lead to much smaller cloud droplets, and thus makes clouds brighter and more reflective of solar radiation. These effects may also influence the timing and amounts of precipitation. In addition, if cloud altitude changes, the climate can be affected. If cloud tops get higher, the radiation to space from clouds is at a colder temperature and so this reduction in the loss of the energy from the Earth produces a warming influence. However, more extensive low clouds would be likely to produce cooling because of their high reflectivity and so their greater influence on solar radiation.

The Hydrological Cycle

The Earth contains roughly $1{\bullet}3 \times 10^9$ km^3 of water. Slightly more than 97% of the water is in the oceans and is therefore salty. The fresh water in rivers, lakes, glaciers, and underground aquifers make up roughly 36×10^6 km^3 of water. About 220 000 km^3 of fresh water is in the lakes and rivers and 12 000 km^3 is in the atmosphere. Of the fresh water, 28×10^6 km^3 is locked up frozen in ice sheets, ice caps and glaciers. Most of the ice is contained in the Antarctic ice sheet which, if melted, would increase sea level by about 65 m. By contrast, Greenland contains the equivalent of about 7 m of sea level and the other glaciers and ice caps contain about 0.5 m (*see* Glaciers, Volume 1). Most of the remaining 8×10^6 km^3 of fresh water is stored underground as ground water (*see* Hydrologic Cycle, Volume 1).

The hydrological cycle involves the transfer of water from the oceans to the atmosphere, to the land and back to the oceans, both on top of and beneath the land surface. Water is evaporated from the ocean surface, and as water vapor it is typically transported thousands of kilometers before it is taken up in clouds and weather systems and is precipitated out as rain, snow, hail or some other frozen pellet back to the Earth s surface. Over land, some precipitation infiltrates or percolates into soils and some runs off into streams and rivers. Ponds and lakes or other surface water may evaporate moisture into the atmosphere, and can also freeze so that water can become locked up for a while. The surface water weathers rocks and erodes the landscape, and replenishes subterranean aquifers. Over land, plants transpire moisture into the atmosphere.

A schematic view of the cycling of water through the climate system is given in Figure 5. This figure not only shows the main water reservoirs and the amounts in each in units of 10^3 km^3, but also gives the volumetric flows between ocean, land and atmosphere (based on results from Trenberth and Guillemot, 1998). In units of 10^3 km^3 per

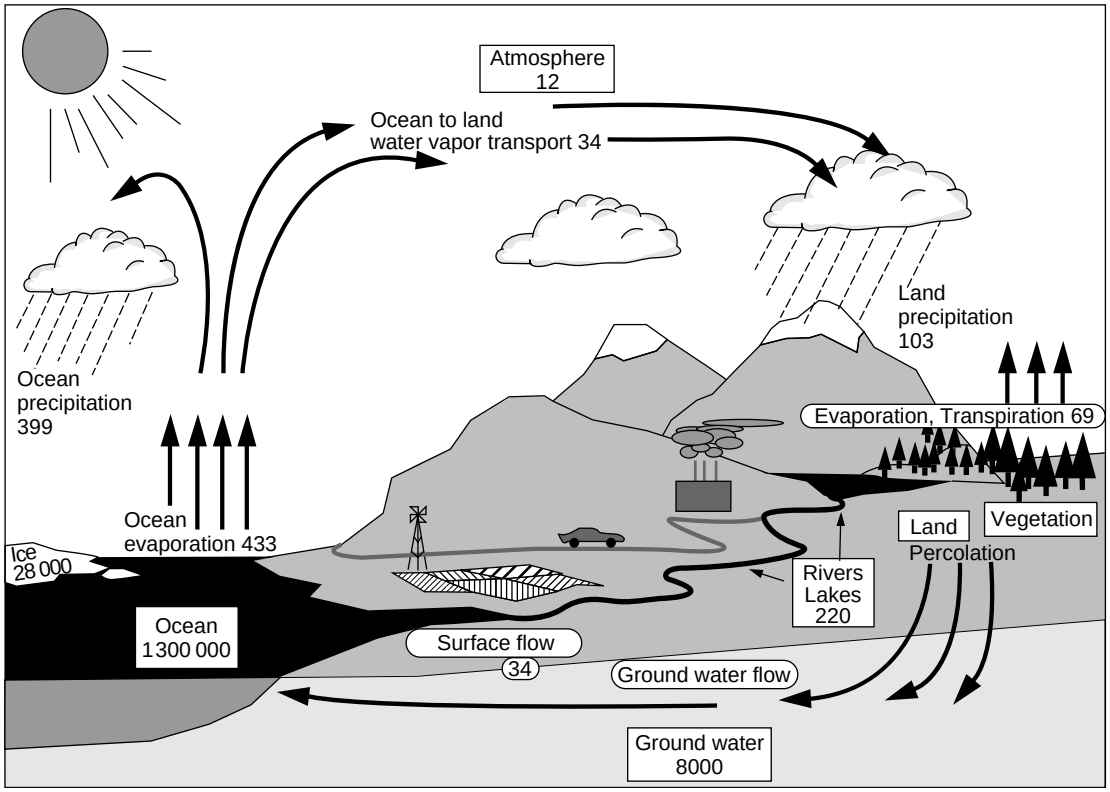


Figure 5 The hydrological cycle, showing the reservoirs of water in square boxes and amounts in units of 10^3 km^3. The figure also shows the flows of water between reservoirs (in oval boxes) in units of 10^3 km^3 year^{-1}

year, evaporation over the oceans (433) exceeds precipitation (399), leaving a net of 34 units of moisture transported onto land as water vapor. On average, this flow must be balanced by a return flow over and beneath the ground through river and stream flows, and subsurface ground water flow. Consequently, precipitation over land exceeds evapotranspiration by this same amount (34). Not surprisingly, perhaps, the average precipitation rate over the oceans exceeds that over land by 72% (allowing also for the differences in areas). It has been estimated (Trenberth, 1998) that on average over 80% of the moisture precipitated out comes from locations over 1000 km distant, highlighting the important role of the winds in moving moisture around.

The Role of Oceans

The oceans cover 70% of the Earth s surface and through their fluid motions, their high heat capacity, and their ecosystems play a central role in shaping the Earth s climate and its variability. The most important characteristic of the oceans is that they are wet and, while obvious, this is sometimes overlooked. Water vapor, evaporated from the ocean surface, provides latent heat energy to the atmosphere during the precipitation process. Wind blowing on the sea surface drives the large-scale ocean circulation in its upper layers. The ocean currents carry heat and salt along with the fresh water around the globe (*see* Ocean Circulation, Volume 1; Salinity Patterns in the Ocean, Volume 1). The oceans therefore store heat, absorbed at the surface, for varying durations and release it in different places, thereby ameliorating temperature changes over nearby land and contributing substantially to variability of climate on many time scales. Additionally, the ocean thermohaline circulation (*see* Thermohaline Circulation, Volume 1), which is the circulation driven by changes in sea water density arising from temperature (thermal) or salt (haline) effects, allows water from the surface to be carried into the deep ocean, where it is isolated from atmospheric influence and hence it may sequester heat for periods of a thousand years or more. The oceans also absorb carbon dioxide and other gases and exchange them with the atmosphere in ways that change with ocean circulation and climate change. In addition, it is likely that marine biotic responses to climate change will result in subsequent changes that may have further ramifications.

The Role of Land

The heat penetration into land occurs mainly through conduction, except where water plays a role, so that heat penetration is limited and slow. Temperature profiles taken

from bore holes into land or ice caps therefore provide a blurry coarse estimate of temperatures in years long past (see Pollack *et al.*, 1998 and Ground Temperature, Volume 1). Consequently, surface air temperature changes over land occur much faster and are much larger than over the oceans for the same heating and, because we live on land, this directly affects human activities.

The land surface encompasses an enormous variety of topographical features and soils, differing slopes (which influence runoff and radiation received) and water capacity. The highly heterogeneous vegetative cover is a mixture of natural and managed ecosystems that vary on very small spatial scales. Changes in soil moisture affect the disposition of heat at the surface and whether it results in increases in air temperature or increased evaporation of moisture. The latter is complicated by the presence of plants, which can act to pump moisture out of the root zone into the leaves, where it can be released into the atmosphere as the plant participates in photosynthesis; a process called *transpiration*. The behavior of land ecosystems can be greatly influenced by changes in atmospheric composition and climate. The availability of surface water and the use of the Sun s energy in photosynthesis and transpiration in plants influence the uptake of carbon dioxide from the atmosphere as plants transform the carbon and water into usable food. Changes in vegetation alter how much sunlight is reflected and how rough the surface is in creating drag on the winds, and the land surface and its ecosystems play an important role in the carbon cycle and fluxes of water vapor and other trace gases.

The Role of Ice

Major ice sheets, like those over Antarctica and Greenland, have a large heat capacity but, like land, the penetration of heat occurs primarily through conduction so that the mass experiencing temperature changes from year to year is small. Temperature profiles can be taken directly from boreholes into ice (see Dahl-Jensen *et al.*, 1998 for temperatures directly from boreholes in the Greenland ice sheet). The temperature profiles suggest that the upward heat flow from the underlying ground is about $51\,\mathrm{mW\,m^{-2}}$, which is very small compared to the various components of the Earth s energy balance. On century timescales, however, the ice sheet heat capacity becomes important. Unlike land, the ice can melt, which has major consequences on longer timescales through changes in sea level.

A major concern has been the possible instability of the West Antarctic Ice Sheet (WAIS) because it is partly grounded below sea level. At the present rate of accumulation, the time needed to restore the ice is estimated as over 10 000 years (Oppenheimer, 1998) so that changes are very slow and occur on millennial time scales unless an instability arises (*see also* Antarctica, Volume 1). If warming alters the grounding of the ice sheet, making it float, it becomes more vulnerable to rapid disintegration, ultimately leading to a rise in sea level of 4–6 m. The present assessment by Oppenheimer is that the risk of substantial change in WAIS contributing to major sea level rise in the 21st century is small, but the risk increases in future centuries as human-induced global climate change progresses. However, there is concern that, after a certain critical point is reached, such changes may be irreversible and unstoppable once begun.

Sea ice is an active component of the climate system and varies greatly in areal extent with the seasons, but only at higher latitudes (*see* Sea Ice, Volume 1). In the Arctic where sea ice is confined by the surrounding continents, mean sea ice thickness is 3–4 m thick and multi-year ice can be present. Around Antarctica the sea ice is unimpeded and spreads out extensively, but as a result the mean thickness is typically 1–2 m.

The Role of Heat Storage

The different components of the climate system contribute on different timescales to climate variations and change. The atmosphere and oceans are fluid systems and can move heat around through convection and advection (in which the heat is carried by the currents, whether small-scale short-lived eddies or large-scale atmospheric jet streams or ocean currents). Changes in phase of water, from ice to liquid to water vapor, affect the storage of heat. However, even ignoring these complexities, many facets of the climate can be deduced simply by considering the heat capacity of the different components of the climate system. The total heat capacity considers the mass involved as well as its capacity for holding heat, as measured by the specific heat of each substance.

The atmosphere does not have much capability to store heat. The heat capacity of the global atmosphere corresponds to that of only a 3.2 m layer of the ocean. However, the depth of ocean actively involved in climate is much greater than that. The specific heat of dry land is roughly a factor of 4.5 less than that of sea water (for moist land the factor is probably closer to 2). Moreover, heat penetration into land is limited by the low thermal conductivity of the land surface; as a result only the top two meters or so of the land typically play an active role in heat storage and release (e.g., as the depth for most of the variations over annual time scales). Accordingly, land plays a much smaller role than the ocean in the storage of heat and in providing a memory for the climate system. Similarly, the ice sheets and glaciers do not play a strong role, while sea ice is important where it forms.

The seasonal variations in heating penetrate into the ocean through a combination of radiation, convective overturning (in which cooled surface waters sink while warmer

more buoyant waters below rise) and mechanical stirring by winds. These processes mix heat through what is called the *mixed layer*, which, on average, involves about the upper 90 m of ocean. The thermal inertia of the 90 m layer can add a delay of about 6 years to the temperature response to an instantaneous change (this time corresponds to an exponential time constant in which there is a 63% response toward a new equilibrium value following an abrupt change). As a result, actual changes in climate tend to be gradual. With its mean depth of about 3800 m, the total ocean would add a delay of 230 years to the response if rapidly mixed. However, mixing is not a rapid process for most of the ocean so that in reality the response depends on the rate of ventilation of water between the well-mixed upper layers of the ocean and the deeper, more isolated layers that are separated by the thermocline (ocean layer, below the mixed layer, exhibiting a strong vertical temperature gradient). The rate of such mixing is not well established and varies greatly geographically. An overall estimate of the delay in surface temperature response caused by the oceans is 10–100 years. The slowest response should be in high latitudes where deep mixing and convection occur, and the fastest response is expected in the tropics. Consequently, the oceans are a great moderating effect on climate changes, especially changes such as those involved with the annual cycle of the seasons.

Generally, the observed variability of temperatures over land is a factor of two to six greater than that over the oceans. At high latitudes over land in winter there is often a strong surface temperature inversion whose strength is very sensitive to the amount of stirring in the atmosphere (Trenberth, 1993). Such wintertime inversions are significantly affected by human activities; for instance an urban heat island effect exceeding 10 °C has been observed during strong surface inversion conditions in Fairbanks, Alaska. Strong surface temperature inversions over mid-latitude continents also occur in winter. In contrast, over the oceans, surface fluxes of heat into the atmosphere keep the air temperature within a narrow range. Thus it is not surprising that over land, month-to-month persistence in surface temperature anomalies is greatest near bodies of water. Consequently, it is clear that for a given heating perturbation, the response over land should be much greater than over the oceans; the atmospheric winds are the reason why the observed factor is only in the two to six range.

A further example of the role of the oceans in moderating temperature variations is the contrast in the mean annual cycle of surface temperature between the NH (60.7% water) and SH (80.9% water) (Figure 6). The amplitude of the 12-month cycle between 40 and 60 °latitude ranges from • 3 °C in the SH to ∼12 °C in the NH. Similarly, in mid-latitudes from 22.5–67.5 °latitude, the average lag in temperature response relative to peak solar radiation is 32.9 days in the NH versus 43.5 days in the SH (Trenberth, 1983),

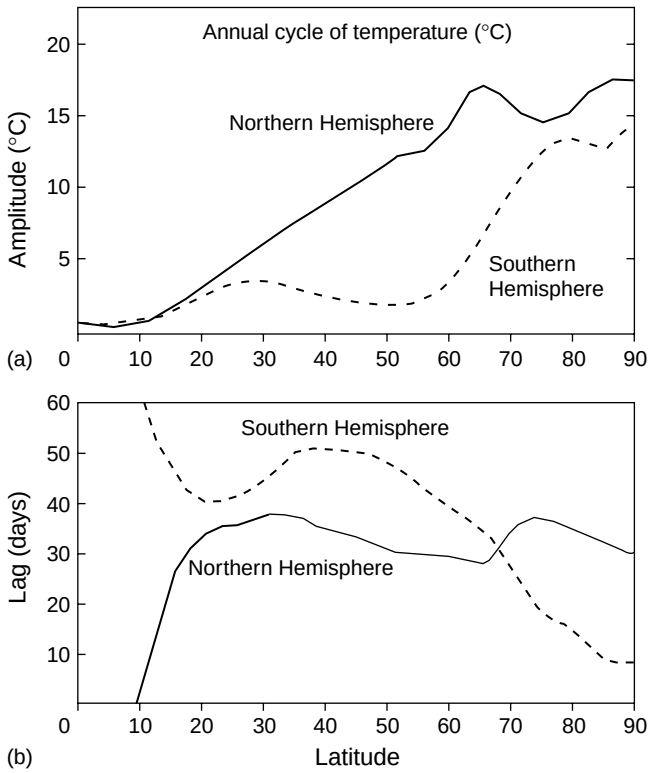


Figure 6 Amplitude of the annual cycle in zonal mean surface temperatures in °C as a function of latitude (a), and the lag of the temperature response behind the Sun in days (b). (From Trenberth, 1983)

again reflecting the difference in thermal inertia. Even the sea surface temperatures (SSTs) in the two hemispheres undergo quite different amplitudes in the annual cycle, and average SSTs are higher in the NH versus the SH at each latitude because of the land distribution and the winds that blow from land to ocean.

Atmosphere–Ocean Interaction: El Niño

Understanding the climate system becomes more complex as the components interact. El Niño events are a striking example of a phenomenon that would not occur without interactions between the atmosphere and ocean. El Niño events involve a warming of the surface waters of the tropical Pacific Ocean (*see* El Niño and La Niña: Causes and Global Consequences, Volume 1). Ocean warming takes place from the International Dateline to the west coast of South America and results in changes in the local and regional ecology. Historically, El Niño events have occurred about every 3–7 years and alternated with the opposite phases of below average temperatures in the tropical Pacific, dubbed La Niña. In the atmosphere, a pattern of change called the Southern Oscillation is closely linked with these ocean changes, so that scientists refer to the total phenomenon as ENSO (*see* El Niño/Southern

Oscillation (ENSO), Volume 1). Then El Niño is the warm phase of ENSO and La Niña is the cold phase.

The strong SST gradient from the warm pool in the western tropical Pacific to the cold tongue in the eastern equatorial Pacific is maintained by the westward-flowing trade winds, which drive the surface ocean currents and determine the pattern of upwelling of cold nutrient-rich waters in the east. Because of the Earth s rotation, easterly winds along the equator deflect currents to the right in the NH and to the left in the SH and thus away from the equator, creating upwelling along the equator. Low sea-level pressures are set up over the warmer waters while higher pressures occur over the cooler regions in the tropics and subtropics. The moisture-laden winds tend to blow toward low pressure so that the air converges, resulting in organized patterns of heavy rainfall and a large-scale overturning along the equator called the Walker Circulation (*see* Walker Circulation, Volume 1). Because convection and thunderstorms preferentially occur over warmer waters, the pattern of SSTs determines the distribution of rainfall in the tropics, and this in turn determines the atmospheric heating patterns through the release of latent heat. The heating drives the large-scale monsoonal-type circulations in the tropics, and consequently determines the winds.

If the Pacific trade winds relax, the ocean currents and upwelling change, causing temperatures to increase in the east, which decreases the surface pressure and temperature gradients along the equator, and so reduces the winds further. This positive feedback leads to the El Niño warming persisting for a year or so, but the ocean changes also sow the seeds of the event s demise. The changes in the ocean currents and internal waves in the ocean lead to a progression of colder waters from the west that may terminate the El Niño and lead to the cold phase La Niña in the tropical Pacific. The El Niño develops as a coupled ocean–atmosphere phenomenon and, because the amount of warm water in the tropics is redistributed, depleted and restored during an ENSO cycle, a major part of the onset and evolution of the events is determined by the history of what has occurred one to two years previously. This means that the future evolution is potentially predictable for several seasons in advance.

The changes in atmospheric circulation are not confined to the tropics but extend globally and influence the jet streams and storm tracks in mid-latitudes. For El Niño conditions, higher than normal sea level pressures over Australia, Indonesia, Southeast Asia, and the Philippines signal drier conditions or even droughts. Dry conditions also prevail for Hawaii, parts of Africa, and extend to the northeast part of Brazil and to Colombia. Intense rains prevail over the central and eastern Pacific, along the west coast of South America and over parts of South America near Uruguay and southern parts of the United States in winter.

OBSERVED CLIMATE CHANGE

Global aspects

We usually view the climate system from our perspective as individuals on the surface of the Earth. Just a few decades ago, the only way a global perspective could be obtained was by collecting observations of the atmosphere and the Earth s surface made at points over the globe and analyzing them to form maps of various fields, such as temperatures. The observational coverage has increased over time, but even today it is not truly global because of the sparse observations over the southern oceans and the polar regions. Earth observing satellites have changed this, as two polar-orbiting sun-synchronous satellites can provide global coverage in about six hours, and geostationary satellites provide images of huge portions of the Earth almost continuously (*see* Earth Observing Systems, Volume 1). The satellites have revealed the myriad cloud types and patterns of infinite variety, and glimpses of the surface underneath. Over land, changes in vegetation and snow cover with the seasons and from year to year can be seen. However, it is difficult to see below the ocean surface and observations in the ocean domain remain few and far between.

In the more distant past, instrumental observations are not available at all and the nature of the weather and climate has to be estimated from proxy indicators (*see* Natural Records of Climate Change, Volume 1). These are organisms that are known to be sensitive to changes in temperature and precipitation, such as trees (through the width and composition of annual tree rings), corals (through annual layers in coral colonies), glaciers (through annual layers of snow and ice), and deposits of small organisms and pollen from plants or land dust on the bottoms of lakes and in marine sediments in oceans. Such fossil indicators can give some estimate of past climate, although the geographical coverage diminishes rapidly the further back in time we go because the most recent ice age obliterates the evidence of the earlier ones.

Even for instrumental observations, the long time-series of high quality observations needed to discern small changes are often compromised by spurious effects, and special care is required in interpretation. Most observations have been made for other purposes, such as weather forecasting, and therefore typically suffer from changes in instrumentation, instrument exposure, measurement techniques, station location and observation times, and in the general environment (such as the building of a city around the measurement location) and there have been major changes in distribution and numbers of observations. Adjustments must be devised to take into account all these influences in estimating the real changes that have occurred.

Analysis of observations of surface temperature show that there has been a global mean warming of about 0.7 °C

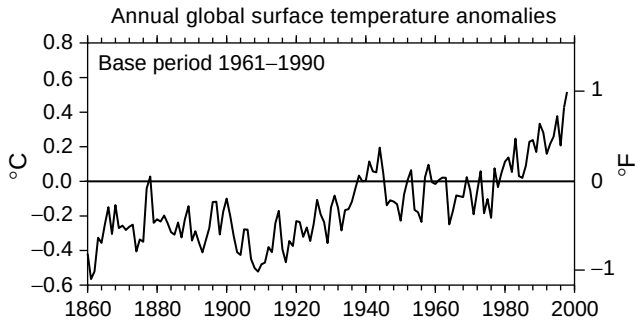


Figure 7 Global annual mean temperatures from 1860–1998 as departures from the 1961–1990 means. (Courtesy Jim Hurrell, adapted from data provided by Hadley Centre, UKMO, and Climate Research Unit, University East Anglia)

over the past one hundred years (*see* The Global Temperature Record, Volume 1); see Figure 7 for the instrumental record of global mean temperatures. The warming became noticeable from the 1920s to the 1940s, leveled off from the 1950s to the 1970s and took off again in the late 1970s. The calendar year 1998 was by far the warmest on record, exceeding the previous record held by 1997. The 1990s were the warmest decade on record. Information from paleodata further indicates that these years are the warmest in at least the past 1000 years, which is as far back as a hemispheric estimate of temperatures can be made (Mann *et al.*, 1999 and *see* Little Ice Age, Volume 1; Medieval Climatic Optimum, Volume 1). The melting of glaciers over most of the world and rising sea levels confirm the reality of the global temperature increases. The observed trend for a larger increase in minimum than maximum temperatures is linked to the associated increases in low cloud amount and aerosol as well as the enhanced greenhouse effect (Dai *et al.*, 1998). There is good evidence for decadal changes in the atmospheric circulation and some evidence for ocean changes. Changes in precipitation and other components of the hydrological cycle vary considerably geographically. Changes in climate variability and extremes are beginning to emerge, but global patterns are not yet apparent.

Changes in climate have occurred in the distant past as the distribution of continents and their landscapes have changed, as the so-called Milankovitch changes in the orbit of the Earth and the Earth s tilt relative to the ecliptic plane have varied the insolation received on Earth, and as the composition of the atmosphere has changed, all through natural processes (*see* Earth System History, Volume 1; Orbital Variations, Volume 1). Recent evidence from ice cores drilled through the Greenland ice sheet (e.g., Bond *et al.*, 1997) have indicated that changes in climate may often have been quite rapid and large, and not associated with any known external forcings (*see* Climate Change, Abrupt, Volume 1). Understanding the spatial scales of this variability and the processes and mechanisms involved is very important as it seems quite possible that strong non-linearities may be involved that result in large changes from relatively small perturbations by provoking positively reinforcing feedback processes in the internal climate system. Changes in the thermohaline circulation in the Atlantic Ocean are one way such abrupt changes might be realized (*see* Thermohaline Circulation, Volume 1). An important question therefore is whether there might be prospects for major surprises as the climate changes.

Spatial Structure of Climate and Climate Change

Because of the land/ocean contrasts and obstacles such as mountain ranges, the mid-latitude westerly winds and the embedded jet stream in each hemisphere contain planetary-scale waves (Figure 2). These waves are usually geographically anchored but can change with time as heating patterns change in the atmosphere. A consequence is that anomalies in climate on a seasonal time scales typically occur over large geographic regions with surface temperatures both above and below normal in different places. The same is true even on much longer time scales. Figure 8 shows that the recent warming has been largest over most of the northern continents, much less over the eastern half of the US, and with cooling over the North and South Pacific and North Atlantic (Trenberth and Hurrell, 1994; Hurrell, 1996; Trenberth and Hoar, 1996, 1997). These changes are known to be dominant in the northern winter and associated with changes in the atmospheric circulation and influences such as El Niño. On a year by year basis, extensive regions of both above and below normal temperatures are the rule, not the exception, as should clearly be expected from the wave motions in the atmosphere.

Temperatures vary enormously on all space and time scales. At individual locations there is a large diurnal cycle, e.g., at Boulder, Colorado (at 40 °N in the central US) from $-7\,^{\circ}$C to $+7\,^{\circ}$C in early January and 15 °C to 30 °C in mid-July. There is also a large annual cycle in daily-average temperature, in this case a 23 °C range. The standard deviation is a measure of the variability. A rule of thumb is that 95% of values fall within ± 2 times the standard deviation and 68% of the values fall within ± 1 standard deviation of the mean. The standard deviation of daily temperature anomalies at Boulder is 4.5 °C while for monthly means, the value drops to 2.1 °C as the vagaries of day to day weather are averaged out. Note that for the daily values, in this case, 71% fall within ± 1 standard deviation of the mean and there are some extreme values well outside ± 2 standard deviations. Both situations arise from the annual cycle, which exhibits much less variability in summer and much greater variability in winter. Spatial averages also average out the pluses and minuses, as illustrated in Figure 9, which presents the

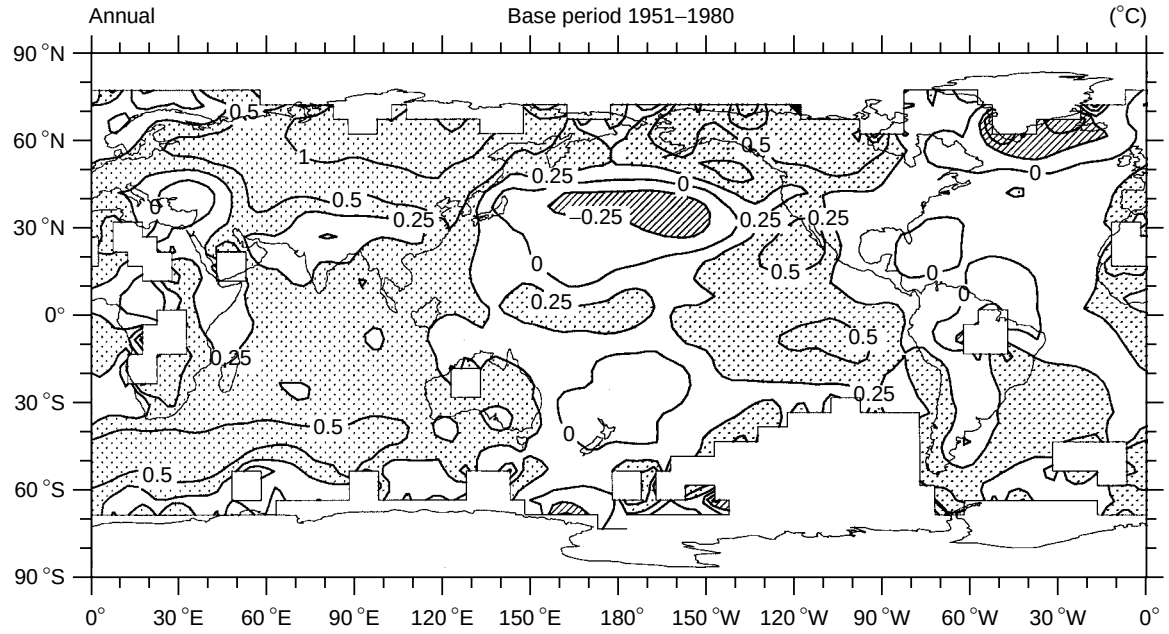


Figure 8 Annual mean surface temperature anomalies for the period 1981–1997 relative to a baseline set of temperatures for 1951–80 . (Courtesy J. Hurrell, adapted from data provided by Hadley Centre, UKMO, and CRU, University East Anglia)

monthly temperature anomalies for Boulder, Colorado, the US, and the globe. For the US as a whole the standard deviation is 1.2 °C and for the globe it is 0.24 °C. Only in the latter is the global warming really apparent. It becomes even clearer in yearly averages (Figure 7).

Figure 9 shows that small trends in global mean values are not readily perceived in regional or local values, simply because of the large natural weather-related variability; this has implications for predictability. The latter depends on the size of the signal from some climate forcing versus the noise of natural variability. As indicated above, the noise (variability) is large locally but can be reduced by spatial and temporal averaging. Consequently, random weather variations mean that local predictions for a month are less reliably determined than averages over large areas and over longer times.

ANTHROPOGENIC CLIMATE CHANGE

Human Influences

Climate can vary for multiple reasons and, in particular, human activities can lead to changes in several ways. There have been major changes in land use over the past two centuries. Conversion of forest to cropland, in particular, has led to a higher albedo in places such as the eastern and central US and changes in evapotranspiration, both of which have probably cooled the region, in summer by perhaps 1 °C and especially in autumn by more than 2 °C, although global effects are less clear (Bonan, 1997, 1999). In cities, the building of concrete jungles allows heat to be soaked up and stored during the day and released at night, moderating nighttime temperatures and contributing to an urban heat island. Space heating also contributes to this effect. Urbanization changes also affect the runoff of water, leading to drier conditions unless compensated by the cooling influences of water usage and irrigation. However, these influences, while causing real changes in urban areas, are quite local. Widespread irrigation on farms can have more regional effects, and so management and storage of water in general is important.

Combustion of fossil fuels not only generates carbon dioxide and heat, but also generates particulate pollution (e.g., soot, smoke) as well as gaseous pollution that can become particulates (e.g., sulfur dioxide, nitrogen dioxide; which get oxidized to form tiny sulfate and nitrate particles) (*see* Aerosols, Troposphere, Volume 1). Other gases, such as carbon monoxide, are also formed in burning and, as a result, the composition of the atmosphere is changing. Several other gases, notably methane, nitrous oxide, the chlorofluorocarbons (CFCs) and tropospheric ozone are also observed to have increased from human activities (especially from biomass burning, landfills, rice paddies, agriculture, animal husbandry, fossil fuel use, leaky fuel lines, and industry), and these are all greenhouse gases. However, the observed decreases in lower stratospheric ozone since the 1970s (*see* Stratosphere, Ozone Trends, Volume 1), caused principally by human-introduced CFCs and halons, contribute to a small cooling in that region (Intergovernmental Panel on Climate Change (IPCC), 1994).

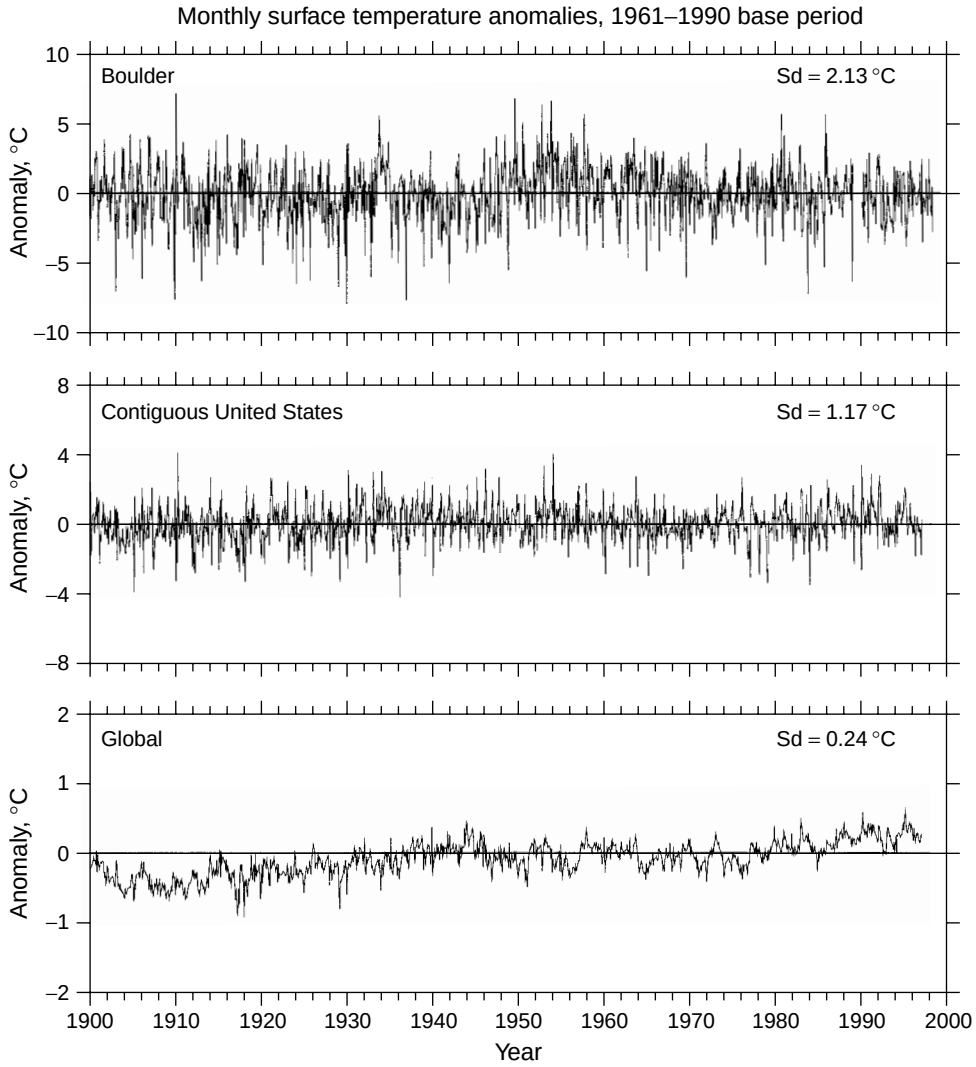


Figure 9 Monthly temperature anomalies for Boulder, Colorado, the US and the globe. The standard deviations (Sd) are listed as 2.13 °C for Boulder, 1.17 °C for the US and 0.24 °C for the globe. The anomalies, defined as departures from the mean annual cycle, are relative to the 1961–90 base period

The Enhanced Greenhouse Effect

The amount of carbon dioxide in the atmosphere has increased by more than 30% over the past two centuries since the beginning of the Industrial Revolution, an increase that is known to be due largely to combustion of fossil fuels and the removal of forests (*see* Carbon Dioxide, Recent Atmospheric Trends, Volume 1). Most of this increase has occurred since World War II. Because carbon dioxide is recycled through the atmosphere many times before it is finally removed from the active atmosphere-ocean-vegetation resevoirs, the excess concentration has a lifetime exceeding a hundred years. As a result, continuing emissions will lead to a continuing build-up in the atmospheric concentrations. In the absence of controls, future projections are that the rate of increase in carbon dioxide emissions (*see* Trends in Global Emmisions: Carbon, Sulfur, and Nitrogen, Volume 3) may accelerate and concentrations could double their pre-industrial values within the next sixty or so years (IPCC, 1994). Effects of implementing the Kyoto Protocol of 1997 (*see* United Nations Framework Convention on Climate Change and Kyoto Protocol, Volume 4) are quite uncertain but may delay doubling of pre-industrial values by perhaps 15 years (Wigley, 1998), and thus would buy time but would not solve the problem.

Effects of Aerosols

Human activities also affect the amount of aerosol in the atmosphere, which influences climate in other ways. The main direct effect of aerosols is the scattering of some solar radiation back to space; which tends to cool the Earth s

surface. Aerosols can also influence the radiation budget by directly absorbing solar radiation leading to local heating of the atmosphere and, to a lesser extent, by absorbing and emitting thermal radiation. A further influence of aerosols is that many of them act as nuclei on which cloud droplets condense. A changed concentration therefore tends to affect the number and size of droplets in a cloud and hence alters the reflection and the absorption of solar radiation by the cloud. Recent evidence highlights the possible importance of this effect, although the magnitude is very uncertain (Hansen *et al.*, 1997 and *see* Aerosols, Effects on the Climate, Volume 1).

Aerosols occur in the atmosphere from natural causes; for instance, they are blown off the surface of deserts or dry regions. The eruption of Mt. Pinatubo in the Philippines in June 1991 added considerable amounts of aerosol to the stratosphere (*see* Aerosols, Stratosphere, Volume 1; Volcanic Eruption, Mt. Pinatubo, Volume 1), which for about two years, scattered solar radiation leading to a loss of radiation at the surface and a cooling there. Human activities contribute to aerosol particle formation mainly through injection of sulfur dioxide into the atmosphere (which contributes to acid rain) particularly from power stations and through biomass burning.

Because aerosols resulting from human activities typically remain in the atmosphere for only a few days, they tend to be concentrated near their sources, such as near and downwind of industrial regions. The cooling therefore exhibits a very strong regional pattern, and the presence of aerosols adds further complexity to possible climate change as it can help mask, at least temporarily, global warming arising from increased greenhouse gases. However, the aerosol effects do not cancel the global-scale effects of the much longer-lived greenhouse gases, and significant climate changes can still result.

CLIMATIC RESPONSE

Feedbacks

We use the global warming as a specific example of climate change to highlight the issues of determining a climatic response to a particular climate forcing. Determining the climatic response to a change in the radiative forcing is complicated by feedbacks. Some of these can amplify the original warming (positive feedback) while others serve to reduce it (negative feedback) (*see* Climate Feedbacks, Volume 1).

If, for instance, the amount of carbon dioxide in the atmosphere were suddenly doubled, but with other things remaining the same, the outgoing long-wave radiation would be reduced by about $4\,\mathrm{W\,m^{-2}}$ and this energy would be trapped in the surface–atmosphere system. To restore the radiative balance, the surface–atmosphere system must warm up. In the absence of other changes, the warming at the surface and throughout the troposphere would be about $1.2\,^{\circ}\mathrm{C}$. In reality, many other factors will change, and various feedbacks come into play, so that the best estimate of the average global warming for doubled carbon dioxide is $2.5\,^{\circ}\mathrm{C}$ (IPCC, 1990, 1995). In other words the net effect of the feedbacks is positive and roughly doubles the response otherwise expected.

Increased heating therefore leads naturally to expectations for increases in global mean temperatures (often mistakenly thought of as global warming), but other changes in weather are also important. Increases in greenhouse gases in the atmosphere produce global warming through an increase in downwelling infrared radiation, and thus not only increase surface temperatures but also enhance the hydrological cycle because much of the heating at the surface goes into additional evaporation of surface moisture. Global temperature increases signify that the water-holding capacity of the atmosphere increases and, together with enhanced evaporation, this means that the actual atmospheric moisture increases (*see* Water Vapor: Distribution and Trends, Volume 1), as is observed to be happening in many places (Trenberth, 1998). It follows that naturally occurring droughts are likely to be exacerbated by enhanced drying. Thus droughts, such as those set up by El Niño, are likely to start sooner, plants will wilt sooner, and the droughts may become more extensive and last longer with global warming. Once the land is dry, then all the solar radiation goes into raising temperature, bringing on sweltering heat waves.

Further, globally there must be an increase in precipitation to balance the enhanced evaporation. The presence of increased moisture in the atmosphere implies stronger moisture flow converging into all precipitating weather systems, whether they are thunderstorms, or extratropical rain or snowstorms. This leads to the expectation of enhanced rainfall or snowfall events, which is also observed to be happening (Karl and Knight, 1998; Trenberth, 1998).

The main positive feedback thus comes from water vapor as the amount of water vapor in the atmosphere increases as the Earth warms and, because water vapor is an important greenhouse gas, it amplifies the warming. However, increases in cloud may act either to amplify the warming through the greenhouse effect of clouds or reduce it by the increase in albedo; which effect dominates depends on the height and type of clouds, and varies greatly with geographic location and time of year (*see* Cloud- Radiation Interactions, Volume 1). Ice-albedo feedback probably leads to amplification of temperature changes in high latitudes. It arises because decreases in sea ice and snow cover, which have high albedo, decrease the radiation reflected back to space and thus produces warming that may further decrease the sea ice and snow cover extent. However, increased open water may lead to more

evaporation and atmospheric water vapor, thereby increasing fog and low cloud amount, offsetting the change in surface albedo.

Other more complicated feedbacks may involve the atmosphere and ocean, e.g., cold waters off of the western coasts of continents (such as California or Peru) encourage development of extensive low stratocumulus cloud decks which block the Sun and this helps keep the ocean cold. A warming of the waters, such as during El Niño, could eliminate the cloud deck and lead to further sea surface warming because of an increase in solar radiation. El Niño itself involves strong positive and negative feedbacks that are the essence of this natural phenomenon. Warming of the oceans from increased carbon dioxide may diminish the ability of the oceans to continue to take up carbon dioxide, thereby enhancing the rate of change.

A number of feedbacks involve the biosphere and are especially important when considering details of the carbon cycle and the impacts of climate change. For example, in wetland soils rich in organic matter, such as peat, the presence of water limits oxygen access and creates anaerobic microbial decay, which releases methane into the atmosphere. If climate change acts to dry the soils, aerobic microbial activity develops and produces enhanced rates of decay, which releases large amounts of carbon dioxide into the atmosphere. Similarly, in areas of permafrost, natural gas hydrates (*see* Methane Clathrates, Volume 1), which are a crystalline form of water and mostly methane, can be destabilized by permafrost melting, releasing methane or carbon dioxide into the atmosphere, depending on the nature of the associated microbial activity that develops. Empirical evidence from ice cores reveals very strong parallel changes in temperatures, carbon dioxide and methane over the past 400 000 years through the major glacials and interglacials, suggesting that atmospheric concentrations of these greenhouse gases somehow responded to the climate changes underway and reinforced the changes through positive feedbacks. Much remains to be learned about these feedbacks and their possible influences on predictions of future carbon dioxide concentrations and climate.

Modeling of Climate

To quantify the response of the climate system to changes in forcing it is essential to account for all the complex interactions and feedbacks among the climate system components and this is done using numerical models of the climate system based upon well established physical principles. Global climate models include representations of all processes indicated in Figure 1. With comprehensive climate models, experiments can be run with and without increases in greenhouse gases and also other influences, such as changes in aerosols. The best models encapsulate the current understanding of the physical processes involved in the climate system, the interactions, and the performance of the system as a whole. They have been extensively tested and evaluated using observations (*see* Model Simulations of Present and Historical Climates, Volume 1; Projection of Future Changes in Climate, Volume 1). While models are exceedingly useful tools for carrying out numerical climate experiments, they are not perfect, and so have to be used carefully (Trenberth, 1997).

Attribution of Climate Change

It is desirable to examine and evaluate the past observational record by running models forced with realistic radiative forcing, although the latter is not that well known. It is in this way that models can be further tested and it may be possible to attribute the observed changes to particular changes in forcing, such as from volcanic or solar origins, and to achieve detection of the effects of human activities and specifically the effects from increases in aerosols and greenhouse gases.

Examining observed changes and comparing with the signature provide by models may enable attribution of the changes to the changes in atmospheric composition, and in 1995 the Intergovernmental Panel on Climate Change (IPCC) assessment concluded that the balance of evidence suggests that there is a discernible human influence on global climate . Since then the evidence has become stronger (*see* Climate Change, Detection and Attribution, Volume 1).

For the observational temperature record (Figure 7), the best evidence compiled to date suggests that solar variability has played a small role but has contributed to some of the warming of the twentieth century – perhaps 0.2 °C – up to about 1950 (Cubasch *et al.*, 1997). Changes in aerosols in the atmosphere, both from volcanic eruptions, from increased visible pollutants and their effects on clouds have also contributed to reduced warming, perhaps by a couple of tenths of a degree C (Hansen *et al.*, 1993). Some year-to-year fluctuations may have arisen from volcanic debris, such as the temporary cooling in 1991 and 1992 following the eruption of Mount Pinatubo in June 1991 (Hansen *et al.*, 1997). Heavy industrialization following World War II may have contributed to the plateau in global temperatures from 1950–1970 or so. Natural fluctuations arising from interactions between the atmosphere and the oceans have probably also contributed to decadal fluctuations of perhaps as much as 0.1 °C in global mean temperatures, and El Niños contribute to the interannual variations, typically of one or two tenths degree C warming. It is only after the late 1970s that global warming from increases in greenhouse gases has probably emerged as a clear signal in global temperatures.

Projection of Climate Change

When a model is employed for projecting changes in climate it is first run for many simulated decades without any changes in external forcing in the system. The quality of the simulation can then be assessed by comparing the mean, the annual cycle and the variability statistics on different time scales with observations of the climate. In this way the model is evaluated (*see* Model Simulations of Present and Historical Climates, Volume 1). The model is then run with changes in external forcing, such as with a possible future profile of greenhouse gas concentrations. The differences between the climate statistics in the two simulations provide an estimate of the accompanying climate change. However, definitive projections of possible local climate changes, which are most needed for assessing impacts, are the most challenging to do with any certainty.

Projections have been made of future global warming effects based upon model results to the year 2100 (*see* Projection of Future Changes in Climate, Volume 1). Because the actions of humans are not predictable in any deterministic sense, future projections necessarily contain a what if emissions scenario. In addition, for a given scenario, the rate of temperature increase depends on the model and features such as how clouds are depicted, so that a range of possible outcomes exists. The IPCC projections (IPCC, 1996) for a mid-range emissions scenario in which carbon dioxide concentrations approximately double 1990 values by the year 2100 produces global mean temperature increases ranging from 1.3 to 2.9 °C above 1990 values with a best estimate of about 2 °C. However, somewhat larger temperature increases have been projected in the IPCC s Third Assessment Report (IPCC, 2001) as a result of projections that sulfur dioxide emissions are likely to be controlled. Note that while these projections include crude estimates of the effects of sulfate aerosol they deliberately omit other possible human influences such as changes in land use. A major concern is that the projected rates of climate change will exceed anything seen in nature in the past 10 000 years.

REFERENCES

Bonan, G B (1997) Effects of Land Use on the Climate of the United States, *Clim. Change*, 37, 449–486.

Bonan, G B (1999) Frost Followed the Plow: Impacts of Deforestation on the Climate of the United States, *Ecol. Appl.*, 9(4), 1305–1315.

Bond, G, Showers, W, Cheseby, M, Lotti, R, Almasi, P, deMenocal, P, Priore, P, Cullen, H, Hajdas, I, and Bonani, G (1997) A Pervasive Millennial-scale Cycle in North Atlantic Holocene and Glacial Climates, *Science*, 278, 1257–1266.

Cubasch, U, Voss, R, Hegerl, G C, Waszkewitz, J, and Crowley, T J (1997) Simulation of the Influence of Solar Radiation Variations on the Global Climate with an Ocean-atmosphere General Circulation Model, *Climate Dyn.*, 13, 757–767.

Dahl-Jensen, D, Mosegaard, K, Gundestrup, N, Clow, G D, Johnsen, S J, Hansen, A W, and Balling, N (1998) Past Temperatures Directly from the Greenland Ice Sheet, *Science*, 282, 268–271.

Hansen, J, Sato, M, Lacis, A, and Ruedy, R (1997) The Missing Climate Forcing, *Philos. Trans. R. Soc. London*, 352, 231–240.

Hansen, J, Sato, M, Ruedy, R, Lacis, A, Asamoah, K, Borenstein, S, Brown, E, Cairns, B, Caliri, G, Campbell, M, Curran, B, de Castro, S, Druyan, L, Fox, M, Johnson, C, Lerner, J, McCormick, M P, Miller, R, Minnis, P, Morrison, A, Pandolfo, L, Ramberran, I, Zaucker, F, Robinson, M, Russell, P, Shah, K, Stone, P, Tegen, I, Thomason, L, Wilder, J, and Wilson, H (1996) A Pinatubo Climate Modeling Investigation, in *The Mount Pinatubo Eruption: Effects on the Atmosphere and Climate*, eds G Fiocco, D Fua, and G Visconti, NATO ASI Series, Springer-Verlag. Heidelberg, Vol. I, 42, 233–272.

Hansen, J, Lacis, A, Ruedy, R, Sato, M, and Wilson, H (1993) How Sensitive is the World s Climate? *Natl. Geogr. Res. Exploration*, 9(2) 142–158.

Hardin, G (1968) The Tragedy of the Commons, *Science*, 162, 1243–1248.

IPCC (1990) *Climate Change: the IPCC Scienti c Assessment*, eds J T Houghton, G J Jenkins, and J J Ephraums, Cambridge University Press, Cambridge, 1–365.

IPCC (1994) *Climate Change 1994. Radiative Forcing of Climate Change and an Evaluation of the IPCC IS92 Emission Scenarios*, eds J T Houghton, L G Meira Filho, J Bruce, H Lee, B A Callander, E Haites, N Harris, and K Maskell, Cambridge University Press, Cambridge, 1–339.

IPCC (1996) *Climate Change (1995): The Science of Climate Change*, eds J T Houghton, F G Meira Filho, B A Callander, N Harris, A Kattenberg, and K Maskell, Cambridge University Press, Cambridge, UK, 1–572.

IPCC (2001) *Climate Change (2001): The Scienti c Basis*, eds J T Houghton, Y Ding, D J Griggs, M Noguer, P J van der Linden, X Dai, K Maskell, and C A Johnson, Cambridge University Press, Cambridge, 1–881.

Karl, T R and Knight, R W (1998) Secular Trends of Precipitation Amount, Frequency and Intensity in the USA, *Bull. Am. Meteorol. Soc.*, 79, 231–242.

Mann, M E, Bradley, R S, and Hughes, M K (1999) Northern Hemisphere Temperatures During the Past Millennium: Inferences, Uncertainties, and Limitations, *Geophys. Res. Lett.*, 26, 759–762.

Kiehl, J T and Trenberth, K E (1997) Earth s Annual Global Mean Energy Budget, *Bull. Am. Meteorol. Soc.*, 78, 197–208.

Oppenheimer, M (1998) Global Warming and the Stability of the West Antarctic Ice Sheet, *Nature*, 393, 325–332.

Pollack, H N, Huang, S, and Shen, P Y (1998) Climate Change Record in Subsurface Temperatures: a Global Perspective, *Science*, 282, 279–281.

Soroos, M S (1997) *The Endangered Atmosphere*, University of South Carolina Press, Columbia, SC, 1–339.

Trenberth, K E (1983) What are the Seasons? *Bull. Am. Meteorol. Soc.*, 64, 1276–1282.

Trenberth, K E (1993) Northern Hemisphere Climate Change: Physical Processes and Observed Changes, in *Earth System*

Responses to Global Change: Contrasts between North and South America, eds H A Mooney, E R Fuentes, and B I Kronberg, Academic Press, New York, 35–59.

Trenberth, K E (1998) Atmospheric Moisture Residence Times and Cycling: Implications for Rainfall Rates With Climate Change, *Clim. Change*, **39**, 667–694.

Trenberth, K E and Hoar, T J (1996) The 1990–1995 El Niño-Southern Oscillation Event: Longest on Record, *Geophys. Res. Lett.*, **23**, 57–60.

Trenberth, K E and Hoar, T J (1997) El Niño and Climate Change, *Geophys. Res. Lett.*, **24**, 3057–3060.

Trenberth, K E and Hurrell, J W (1994) Decadal Atmosphere–ocean Variations in the Pacific, *Climate Dyn.*, **9**, 303–319.

Trenberth, K E and Solomon, A (1994) The Global Heat Balance: Heat Transports in the Atmosphere and Ocean, *Climate Dyn.*, **10**, 107–134.

Trenberth, K E and Guillemot, C J (1998) Evaluation of the Atmospheric Moisture and Hydrogical Cycle in the NCEP/NCAR Reanalyzes, *Climate Dyn.*, **14**, 213–231.

Trenberth, K E, Houghton, J T, and Meira Filho, L G (1996) The Climate System: an Overview, Chapter 1 of *Climate Change* 1995. *The science of Climate Change*. Contribution of WG 1 to the Second Assessment Report of the Intergovernmental Panel on Climate Change, eds J T Houghton, L G Meira Filho, B Callander, N Harris, A Kattenberg, and K Maskell, Cambridge University Press, Cambridge, 51–64.

Wigley, T M (1998) The Kyoto Protocol: CO_2, CH_4 and Climate Implications, *Geophys. Res. Lett.*, **25**, 2285–2288.

Earth System History

W RICHARD PELTIER

University of Toronto, Toronto, Canada

In our collective effort to better understand the changes in the global environment that humankind has begun to induce, the end state of which we are as yet unable to clearly discern, it is vital that we establish an appropriate context in which to appreciate the magnitude of our in uence. This requires an appreciation of the history of Earth's evolution and knowledge of the processes that have controlled it. Only by re ection upon the nature and rates of environmental changes that have occurred naturally, in the absence of human in uence, since the planet rst formed, will we be able to appreciate how truly unprecedented are the changes we seem to have initiated. This article is therefore directed towards providing a review of existing knowledge of the evolution of the Earth system, not only of the variability of surface climate through time, but also of the changing properties of the surface itself and of the solar stimulus through which climate is most signi cantly controlled. Of fundamental concern in this discussion will be the issue of the way in which carbon dioxide, a primary greenhouse gas, has covaried with the dynamical state of solid Earth itself and the evidence from a variety of geological archives as to how surface climate responded to these variations. The existing lack of a consensus explanation of the co-evolution of continental ice volume and atmospheric carbon dioxide concentration constitutes one of the most signi cant outstanding problems in paleoclimatology.

EARTH FORMATION AND THERMOCHEMICAL HISTORY: GEODYNAMIC CONTROLS

The fact that Earth formed approximately 4.56 billion years ago (Ga) by accretion of dust and planitesimal sized objects out of the primordial solar nebula is no longer disputed. Detailed simulations of the accretion process strongly suggest that this occurred on a time scale of approximately 100 million years, rapidly enough that internal temperatures were raised sufficiently close to (or above) the melting point that gravitational differentiation of the elemental composition could occur. This composition is well approximated by that of those meteorites called type II carbonaceous chondrites, which continue to infall to the Earth s surface as subtle reminders of the mechanism through which the planet was born. Following this relatively rapid process of Earth formation, the internal structure of the planet was very similar to that which exists at present and which has come to be extremely well understood on the basis of detailed interpretations of the elastic waves that are generated by the occurrence of earthquakes. This structure (see Figure 1) consists of a primarily iron core containing a small quantity of a light alloying element such as silicon, the innermost part of which is solid and the outer part of which is liquid, surrounded by a solid iron-magnesium silicate mantle and an overlying crust. During the entire process of Earth formation, which occurred simultaneously with the formation of the other planets (both terrestrial and gas giant planets) in our solar system, the Sun itself was already burning brightly. It had formed through gravitational collapse of most of the hydrogen gas that comprised the original nebula, most of the rest having coalesced to form the gas giant Jupiter and Saturn. The Sun has existed for approximately 4.6 billion years. In what follows we will have reason to understand that a main sequence star such as our Sun burns ever more brightly as it ages, and the hydrogen gas in its core is thereby transformed into helium.

The Earth s surface, the stage upon which modern global change is occurring, has always been an extraordinarily unstable platform. This is extremely well documented on the basis of the geological investigations that have been conducted over the past two centuries. That the Earth s surface is in a state of high agitation is clear on the basis of observations of the process of continental drift and plate tectonics (*see* Plate Tectonics, Volume 1). Figure 2 illustrates the current state of surface kinematic motions of the twelve plates, into which the surface of the Earth may be divided. In the interiors of these plates the motion of the surface crust is primarily that of a rigid body with the motion being characterized by a constant angular velocity around a fixed pole of rotation. In a frame of reference in

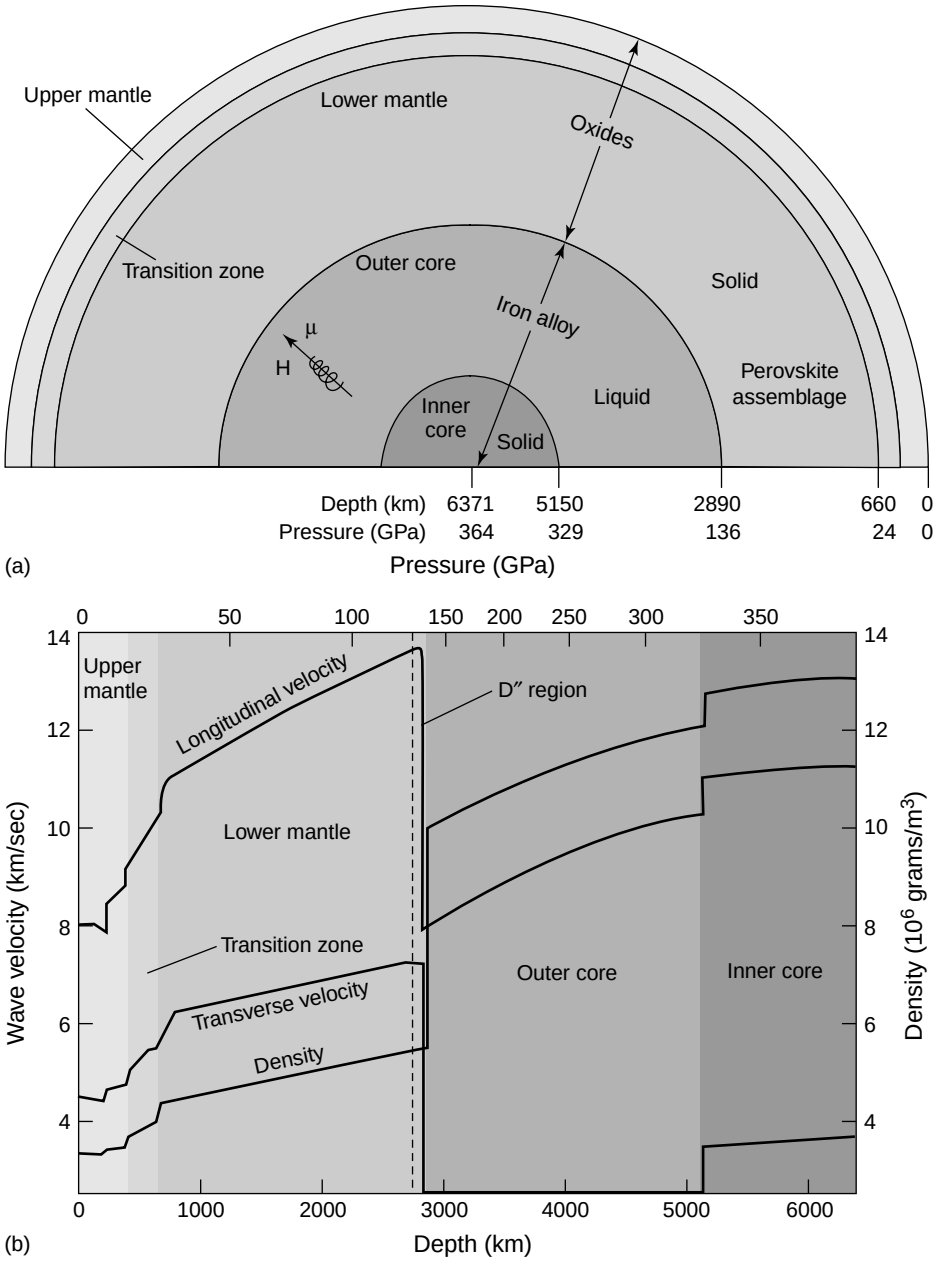


Figure 1 Cross sections of the Earth showing its properties. (a) Seismologically determined regions and pressures as a function of depth (100 Gpa equals 1 megabar). (b) Average elastic parameters as a function of depth. The P-wave (longitudinal) velocity V_p, S-wave (transverse) velocity V_s and density ρ are determined on the basis of seismological analysis. (Reprinted with permission from Lay *et al.* (1990). Copyright 1990, American Institute of Physics)

which the collectivity of plates exhibits no net rotation, the relative velocities of the plates vary between approximately 10 cm year^{-1} and 2 cm year^{-1} with an average value of approximately 4 cm year^{-1}. This kinematic depiction of the process of surface plate tectonics was originally inferred on the basis of geological information, specifically from the alternately normal and reversed polarization of the magnetic anomalies of known age which are observed to run parallel to mid-ocean ridges. However, within the past two decades it has been confirmed in detail using the space geodetic technique of very long baseline (radio) interferometry (VLBI) in which arrays of radio telescopes are employed to accurately measure the relative motion of points on the Earth s surface.

A major accomplishment of the past several decades of research in the area of solid Earth geophysics has been to establish that it is through the process of thermal convection in Earth s mantle that the surface plate tectonic

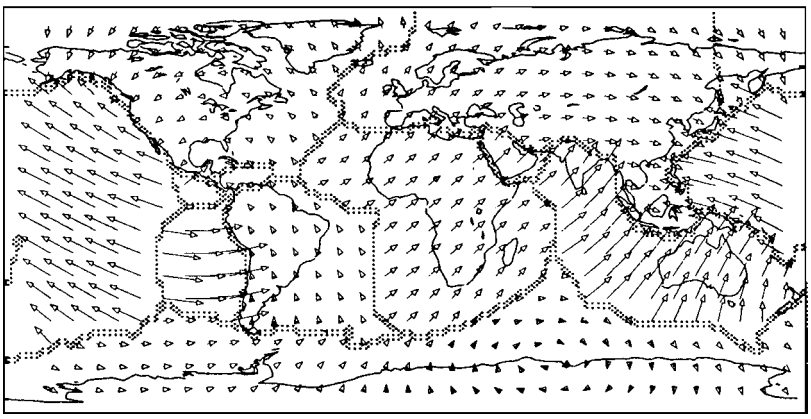


Figure 2 Tangential velocities of material at the Earth's surface in the no net rotation frame of reference based upon a standard geologically derived surface plate reconstruction. Regions of convergence correspond to deep ocean trenches (subduction zones) and regions of divergence correspond to mid-ocean ridges

process described above is maintained. Figure 3 illustrates this physical process through a sequence of depictions of the internal temperature field of the mantle, from a spherical axially symmetric model of the convection process that span a simulated time interval of several 100 million years. Clearly evident on the basis of this simulation is the occurrence of avalanche like events during which intense downwellings of cold material occur across the Spinel post Spinel phase transformation at 660 km depth in the Earth that marks the boundary between the lower mantle and the overlying transition zone (see Figure 1). One of the most recent discoveries concerning the physics of the mantle convection process concerns this influence of the endothermic transformation upon the circulation of the mantle, which drives the motion of the surface plates. This influence is such as to cause the circulation to develop a highly intermittent form in which relatively long periods of quiescence, during which the circulation is significantly layered, are followed by relatively brief periods of highly energetic activity that are triggered by the avalanche-like process and during which the flow is negligibly layered.

Figure 4 presents an early histogram of the number of radiometrically dated rock samples from the surface of the continents as a function of sample age. This figure strongly suggests that the real Earth has undergone a process of thermal and chemical evolution, which is characterized by a degree of intermittency that is very similar to that exhibited by the numerical model from which snapshots of the internal mantle temperature field are shown in Figure 3. The histogram in Figure 4 is therefore indicative of the temporal variability of the process through which continental crust is created by the irreversible differentiation of mantle material. This process involves melting and volcanic outgassing and leads to the creation of continental crust that is strongly enriched in the large ionic radii lithophile elements, included among which are the radioactive elements uranium, thorium and potassium. The episodicity evidenced in this histogram, of the time dependence of the process of crust formation (and thus volcanic out-gassing), is to be interpreted as being controlled by or at least closely connected to, the so-called Wilson cycle of supercontinent creation and destruction. This cycle refers to the fact, now well established on the basis of paleomagnetic and geological evidence that, at various times in the past, all or most of the continental fragments were collected together to form so-called supercontinents. Referring to the conventional geological time scale shown on Figure 5, the most recent epoch during which this occurred was during the Carboniferous period approximately 300 million years ago (Ma) when the supercontinent of Pangea was located near the south magnetic pole (shown later in Figure 7). Prior to this, in the so-called Neoproterozoic period, which lasted from approximately 1 Ga until 545 Ma, the supercontinents of Rodinia and Pannotia (shown later in Figure 13) were apparently in place at a similar south magnetic polar location, at 600 Ma, in the case of Pannotia, and near the beginning of the Neoproterozoic in the case of Rodinia. Inspection of the histogram shown on Figure 5 will show that the times during which supercontinents had formed coincide with the times during which new continental crust is being formed least efficiently. Likewise it is apparently during times when the supercontinents are being actively rifted and when, presumably, volcanic out-gassing is especially intense, that new continental crust is being formed most efficiently. It is interesting to note that the time scale of several hundred million years which separates the episodes of supercontinent aggregation is very similar to the time scale between distinct episodes of avalanche activity determined on the basis of the previously discussed numerical model of the mantle convection process in which

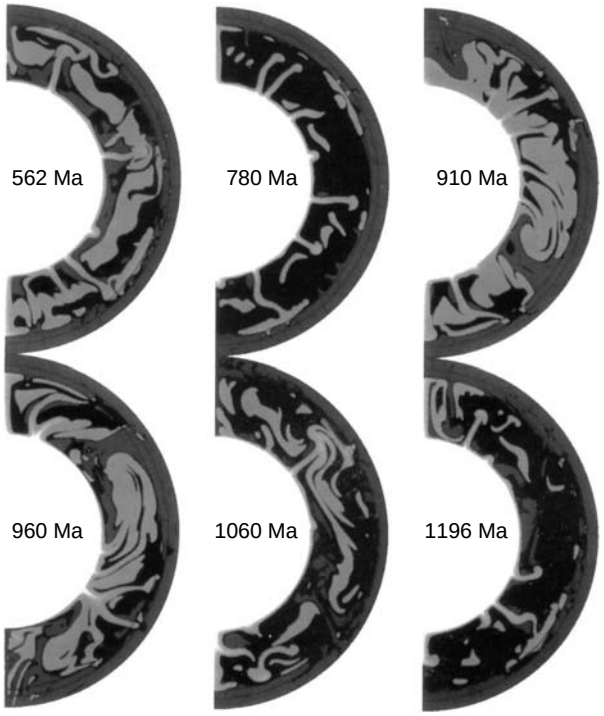


Figure 3 Internal mantle temperature fields from an axi-symmetric numerical model of the mantle convection process. The lighter colors denote hot upwelling material that originates through instability at the core-mantle boundary whereas the darker regions correspond to cold downwelling material. The six frames cover a range of (simulated) time of 634 million years during which period there occurs an interval in which intense avalanches of cold material descend across the 660 km depth seismic discontinuity. The 660 km phase change interface, which is due to the Spinel–Perovskite and Magnesiowustite phase transformation, is endothermic and so is able to initially inhibit the flow of cold downwelling material across it. Eventually, however, the thermal boundary layer caused by this inhibition of radial flow becomes convectively unstable and an avalanche occurs

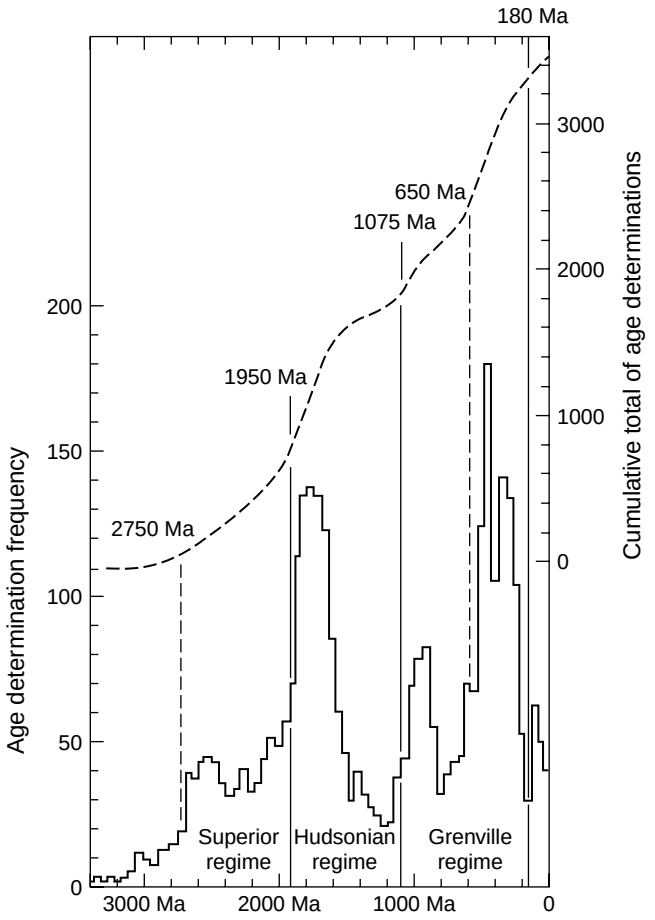


Figure 4 Age histogram of continental rocks showing the number of dated samples as a function of sample age. (Reprinted with permission from Dearnly (1965). Copyright 1965 Macmillan Magazines)

the endothermic phase transformation at 660 km depth plays an important role.

The possible importance of this line of argument to the understanding of ultra low frequency climate variability will be clear on the basis of Figure 6, which shows the reconstructions of atmospheric carbon dioxide through time that have been produced by Robert A Berner of Yale University (*see* Carbon Dioxide Concentration and Climate Over Geological Times, Volume 1), reconstructions that he has designated GEOCARB1 and GEOCARB2. In these best currently available reconstructions of atmospheric carbon dioxide concentration since the end of the Neoproterozoic epoch at 545 Ma, it will be observed that carbon dioxide concentration rose dramatically during the breakup of Pangea subsequent to 300 Ma and also during the breakup of Pannotia subsequent to 600 Ma. Because we expect volcanic out-gassing to be especially intense during an episode of supercontinent rifting and because carbon dioxide is a primary output of surface volcanism, we may usefully understand the paleo-carbon dioxide reconstruction of Berner as being deeply linked to the Wilson cycle of supercontinent creation and destruction. To the extent that this connection is as secure as appears to be the case, we may furthermore link episodes of supercontinent rifting to episodes (*see* Plate Tectonics, Volume 1), during which the mantle convective circulation is relatively layered. Similarly, episodes of supercontinent formation may be linked to episodes of occurence of avalanche events during which initially dispersed continental fragments are drawn together at the stagnation point at the Earth s surface, which is located immediately above the avalanche of material that is occurring across the 660 km depth interface. To the extent that the GEOCARB reconstructions of atmospheric carbon dioxide concentration are reasonable facsimiles of reality, it therefore seems clear that profound geodynamic controls actively contribute to the determination of the climatic state

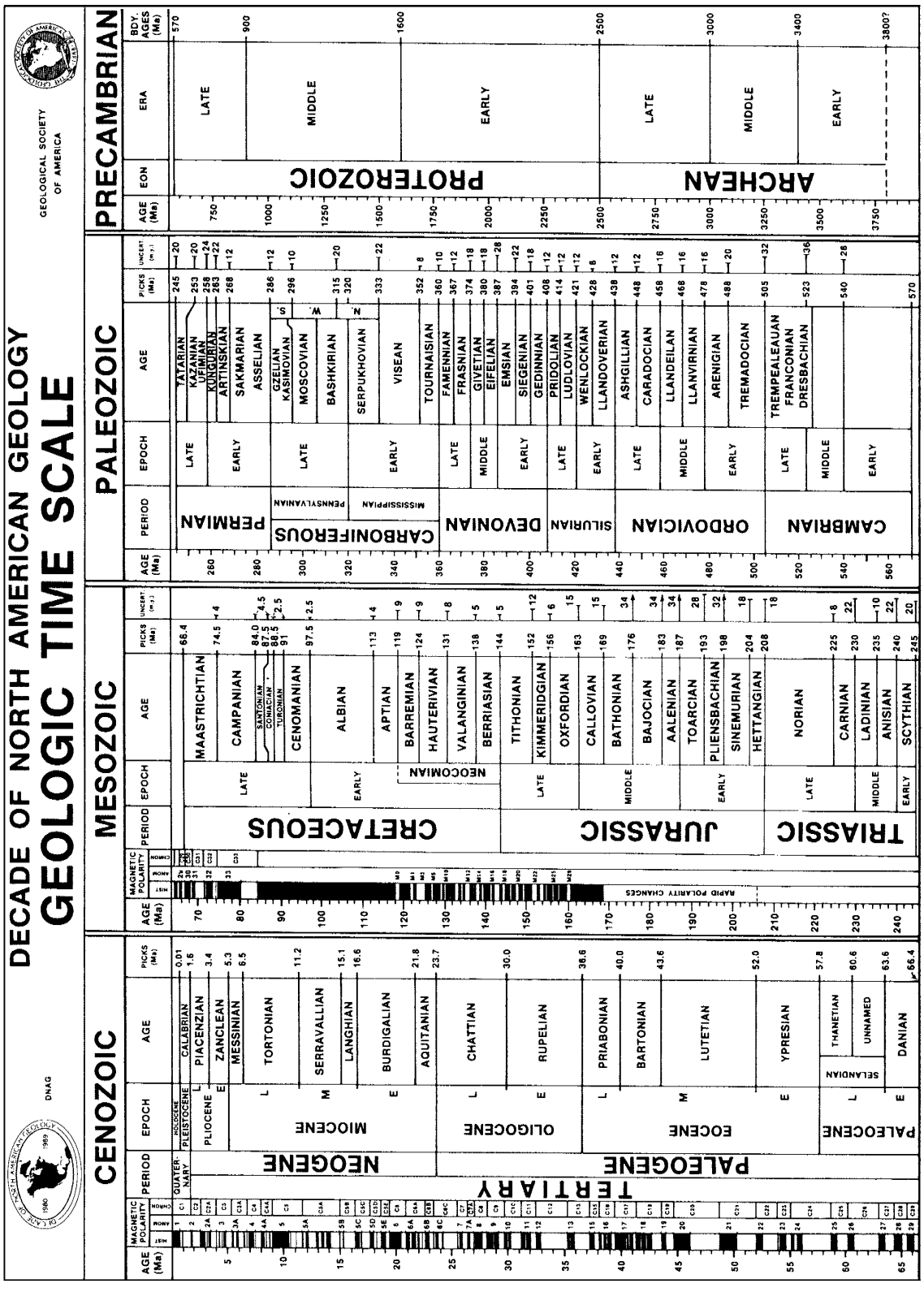


Figure 5 A detailed chart depicting the geological time scale of the Earth, and the names of periods, epochs and ages

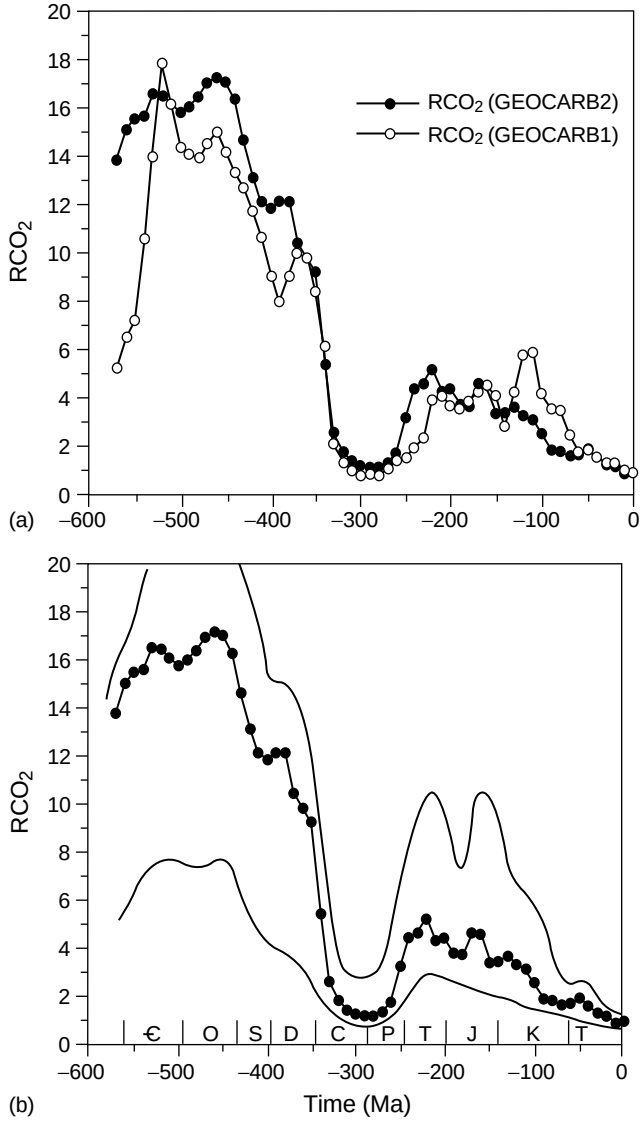


Figure 6 (a) Comparison of the GEOCARB1 and GEOCARB2 reconstructions of the concentration of carbon dioxide in the atmosphere over the past 600 Ma since the end of the Neoproterozoic period. (b) Error bars on the GEOCARB2 reconstruction (from Berner, 1994). RCO_2 is the ratio of the CO_2 concentration to the preindustrial value. (Reprinted with permission from the *American Journal of Science*)

found at Earth s surface at any particular epoch in the past. We may usefully investigate this hypothesis by focusing firstly upon those (relatively infrequent) times in the past, which are well defined in the GEOCARB reconstruction, when carbon dioxide concentrations are believed to have been relatively low. These being the late Pleistocene epoch (900 thousand years ago (ka) to present), the Carboniferous epoch (∼300 Ma) and perhaps also the Neoproterozoic (∼600 Ma). It will prove most illuminating to focus upon the two oldest epochs first.

ICE AGE CLIMATES OF THE CARBONIFEROUS AND NEOPROTEROZOIC: CONSEQUENCE OF A WEAK SUN AND LOW ATMOSPHERIC CARBON DIOXIDE

During the Carboniferous epoch centered upon approximately 300 Ma, the supercontinent of Pangea was centered over the south geomagnetic pole in the configuration shown on Figure 7. In attempting to understand the surface climate state that existed on the Pangean supercontinent, it is critical to appreciate that the solar luminosity at that time would have been lower than present by about 3%. This is indicated in Figure 8, which displays the results of two different theoretical models of solar evolution that differ only modestly from one another. Also indicated on Figure 7 by the star symbols are the locations on Pangea where there is clear geological evidence that the surface was heavily glaciated at that time. In order to demonstrate at least a first order understanding of why the Carboniferous supercontinent should have been glaciated, a simple climate model that includes a detailed description of neither the atmosphere nor the oceans can nonetheless be useful. Consider the surface climate as being controlled by a simple balance of energies that may be described schematically as follows

$$C \times [\text{time rate of change of surface temperature}] = Q_o(1-\alpha) - \varepsilon\sigma T_s^4 - \text{divergence of latitudinal heat transport} \quad (1)$$

In Equation (1), $Q_o = 1370\,\mathrm{W\,m^{-2}}$ is the solar constant, $\sigma = 5.67 \times 10^{-8}\,\mathrm{W\,m^{-2}\,K^{-4}}$ is the Stefan–Boltzman constant, T_s is the surface temperature in degrees Kelvin, α is the surface albedo which depends upon surface type (i.e., land, sea, sea ice, land ice, etc.), and ε is the infrared (IR) emissivity. The latitudinal heat transport in the actual climate system is effected by the circulations of the atmosphere and oceans. At the level of the energy balance description, which is the basis of Equation (1), this may be parameterized in terms of the mid-latitude value of the pole-to-equator temperature gradient. The parameter C in Equation (1) is the surface heat capacity which may be assumed to be a function of latitude and longitude in such a way as to differentiate continents from oceans and ice covered from non-ice covered regions. In ice covered regions the surface albedo is typically of order 0.8 whereas under ice free conditions it is typically of order 0.3 or somewhat lower.

We may augment the model of surface temperature change embodied in Equation (1) by connecting it to a model of global glaciology that may be described in a similarly schematic fashion:

$$[\text{time rate of change of continental ice sheet thickness}] = G - \text{divergence of ice flux} \quad (2)$$

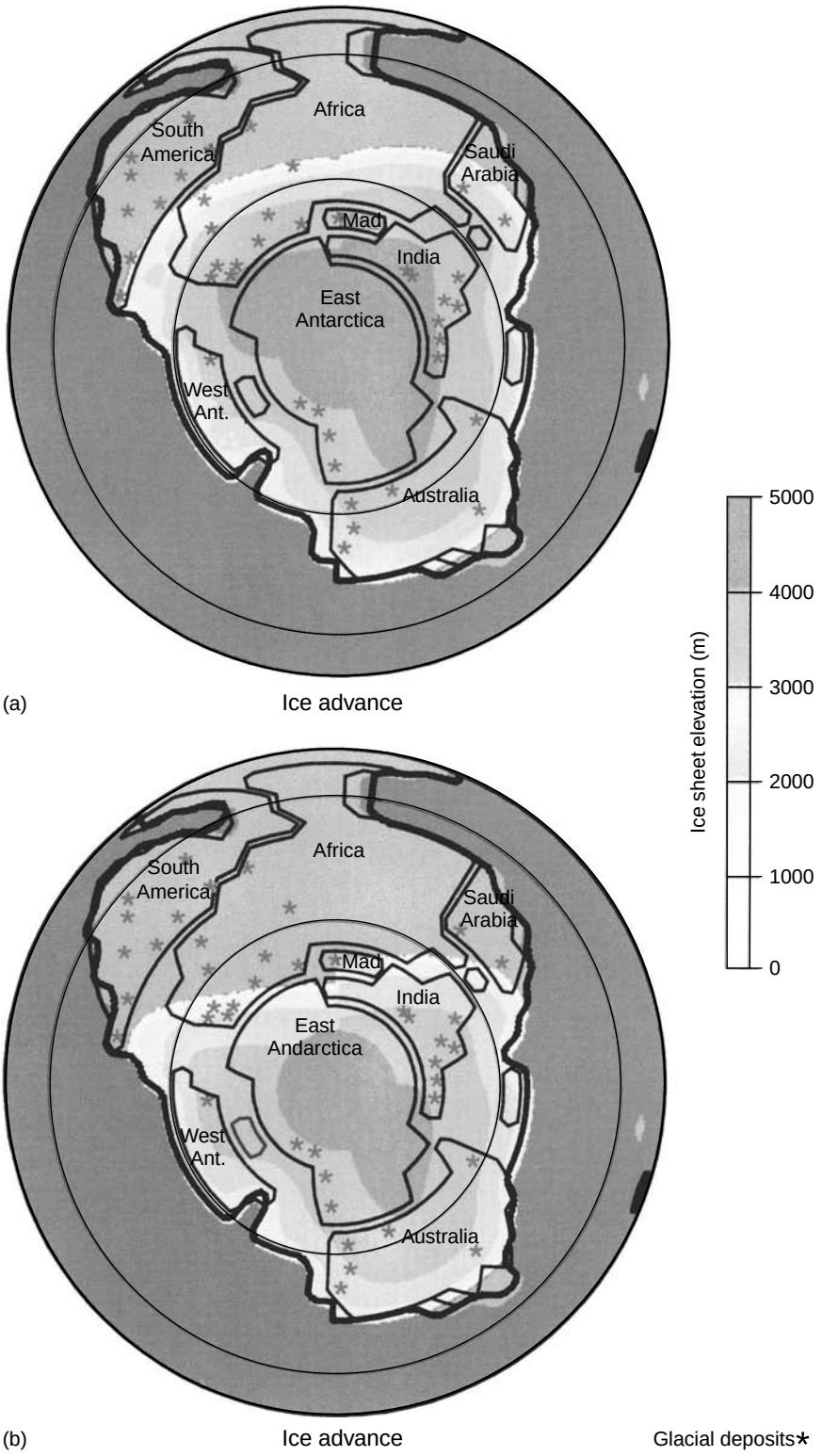


Figure 7 Reconstruction of the supercontinent of Pangea that existed around the south magnetic pole during the Carboniferous period approximately 300 Ma. The locations on each of the named continental fragments of which Pangea was constructed, where there exist indications of heavy continental glaciation, are indicated by stars. Theoretical predictions of the extent of maximum and minimum glaciation calculated by the climate model described in the text are superimposed upon the two maps of Pangea. (From Hyde *et al.* (1999). Reprinted with permission of Springer-Verlag)

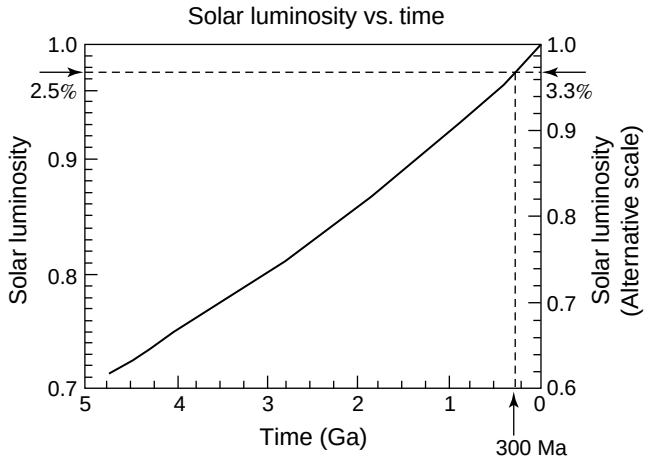


Figure 8 Solar luminosity as a function of time for two different models of the evolution of a main sequence star having properties equal to those of our Sun (from Endal *et al.*, 1981)

In this equation, G is the so-called mass balance function, which is positive over that portion of a continental ice sheet that is experiencing a net accumulation of mass (typically in the interior region) and negative around the periphery where net ablation occurs. These two governing equations of the ice sheet coupled energy balance climate model are strongly coupled by virtue of the fact that the latitude and longitude dependent precipitation rate depends upon temperature as do the rates of accumulation and ablation.

These processes are further strongly linked through the glacial isostatic adjustment (GIA) effect whereby the surface of the solid Earth is depressed by the weight of an ice load. Because of the fact that the temperature in the atmosphere decreases strongly as a function of height, and because the height with respect to sea level of the ice sheet surface is strongly influenced by the GIA effect, this process has an important influence on surface climate and thus the glaciation–deglaciation process. In simple climate models of the kind being discussed, the GIA process may be adequately described (schematically) by the relationship:

$$[\text{time rate of change of bedrock elevation } h'] = \frac{(h' - h'_{\text{o}})}{\tau} + \frac{(p_I/p_m)H}{\tau} \tag{3}$$

in which h'_{o} is the initial longitude and latitude dependent bedrock topography, τ is the relaxation time constant, ρ_I and ρ_m are the densities of ice and the Earth's mantle, respectively, and H is ice thickness.

The final ingredient that is in general required to activate a model of this kind consists of the effective time variations of the solar luminosity that are caused by time variations in the geometry of Earth's orbit around the Sun. If we define by Q_{s} and Q_{w} the summer and winter (half year) effective solar constants, then detailed mathematical analysis shows that the deviations of these quantities from their present values are simply:

$$\Delta Q_{\text{s}} = \left(\frac{\partial S_{\text{s}}}{\partial \varepsilon}\right) \Delta\varepsilon - m\Delta(e \sin s) \tag{4}$$

$$\Delta Q_{\text{w}} = \left(\frac{\partial S_{\text{w}}}{\partial \varepsilon}\right) \Delta\varepsilon + m\Delta(e \sin s) \tag{5}$$

in which ε and e (see Figure 9 for illustration) are, respectively, the obliquity of the spin axis of the planet with respect to the normal to the plane of the ecliptic, and the orbital eccentricity (*see* Climate Model Simulations of the Geological Past, Volume 1). The parameter m is:

$$m = \frac{2TQ\cos\theta}{\pi^2(1-e^2)^{1/2}} \tag{6}$$

in which T is the duration of the tropical year, and θ is the latitude of a point on Earth s surface. In Equations (4) and (5), s is the angle of perihelion and S_{s} and S_{w} are as follows:

$$S_{\text{s}} = \frac{QT}{2\pi(1-e^2)^{1/2}} [f_{\text{o}} + \cos\varepsilon \sin\sigma] \tag{7}$$

$$S_{\text{w}} = \frac{QT}{2\pi(1-e^2)^{1/2}} [f_{\text{o}} - \sin\varepsilon \sin\sigma] \tag{8}$$

Given a detailed reconstruction of the orbit of the Earth through time from the solution of the gravitational n-body problem for the planetary system as a whole (see e.g., Laskar, 1988; Quinn *et al.*, 1991), we have detailed knowledge of $\varepsilon(t)$ and $e(t)$ and so may compute the variations of solar insolation reaching Earth s surface due to changes in the geometry of its orbit. Although we are now able to accurately compute these variations over the past several million years of Earth history, we cannot obtain an accurate model of this contribution to solar forcing for the Carboniferous epoch. On such time scales the influence of deterministic chaos becomes clearly evident when integrating the gravitational n-body problem backwards in time from present values of the ephemeris. What we can do,

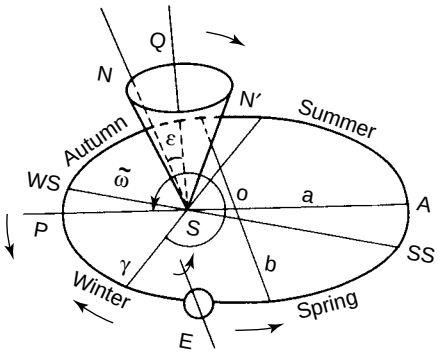


Figure 9 Orbital geometry of the Earth around the Sun (*see* **Orbital Variations**, Volume 1)

however, is to make the not unreasonable assumption that the dominant amplitudes and time scales of variations in ε and e were not then substantially different from present and include the influence of these variations on the predicted Pangean climate, recognizing that the phase of the variations so predicted will be meaningless.

Figure 10 shows the calculated variation of the quasi-equilibrium ice volume as a function of the particular level of atmospheric carbon dioxide concentration assumed in integrating the above described ice sheet coupled energy balance model of planetary climate. At $1 \times CO_2$, the ice volume equilibrates at a relatively high level, varying from approximately $10^8\,km^3 - 1.5 \times 10^8\,km^3$. The maximum and minimum spatial extents of the calculated Pangean ice sheet are shown in parts (a) and (b) of Figure 7, respectively. These extremes are observed to nicely cover the surface area within which evidence exists that intense glaciation occurred. Figure 10 also shows ice volume predictions for atmospheric carbon dioxide concentrations of both $2\times$ and $3\times$ the modern (preindustrial) value. In the former case the quasi-equilibrium ice volume calculated by the model drops to about half of that delivered by the $1 \times CO_2$ model whereas, in the latter case, glaciation is predicted not to occur at all. Inspection of the GEOCARB model inference of the concentration of carbon dioxide in the atmosphere at 300 Ma, shown previously in Figure 6, demonstrates the model to suggest that this level was tightly constrained to a concentration near the present value. This is entirely consistent with the result of the ice sheet coupled energy balance climate model which, in the case of $1 \times CO_2$, accurately reproduces the observed extent of glaciation on the super continent of Pangea. Figure 11 demonstrates that the success of the model is equally strongly dependent upon the assumed 3% reduction in solar luminosity. When the

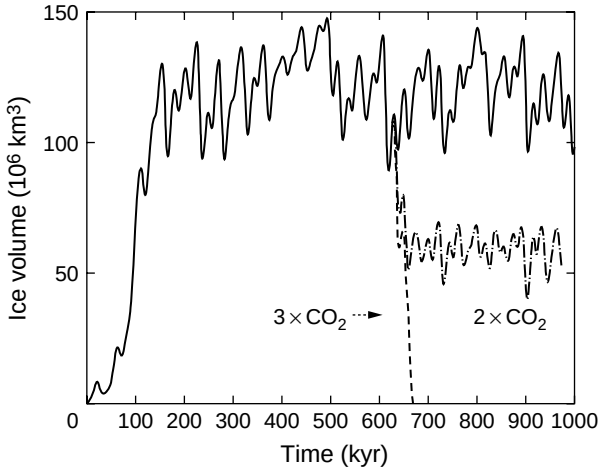


Figure 10 Ice volume verses time on the Pangean supercontinent for three different levels of atmospheric carbon dioxide concentration (from Hyde *et al.,* 1999. Reprinted with permission from Springer-verlag)

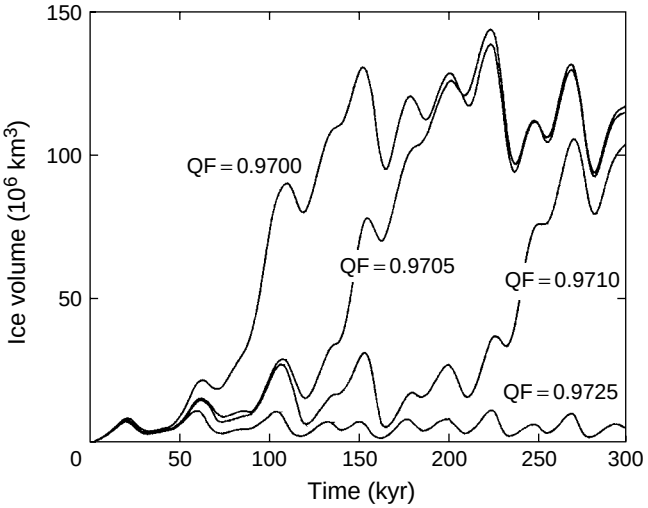


Figure 11 Ice volume vs. time for Pangea, assuming modern values of the atmospheric carbon dioxide concentration and for several values of the reduction in solar luminosity (from Hyde *et al.,* 1999. Reprinted with permission from Springer-verlag)

parameter QF = 1, the assumed solar luminosity is equal to its modern value; QF = 0.97 thus corresponds to a 3% reduction. Inspection of the Figure shows that when QF = 0.9725, so that the assumed reduction of the solar constant is only 2.75%, then no significant glaciation occurs at all.

Reaching still further back in the history of Earth system evolution, into the Neoproterozoic era which extended from ∼1 Ga−∼545 Ma, we enter the range of time within which the most extensive episodes of glaciation in Earth history are believed to have occurred. These glaciations, called the Marinoan and Veranger glaciations, are estimated to have occurred at some time within the intervals from ∼760 to 700 Ma and from ∼620−580 Ma, respectively, although the accuracy with which the times of occurrence of these events have been dated and their durations determined is still a subject of active debate. It is nevertheless known that intense continental glaciation extended to the equator during at least the Veranger episode and that these glaciations were marked in the geological record by highly significant changes in strontium, sulfur and carbon isotopes.

This epoch of Earth history was also a critical interval in biological evolution as it was during the Neoproterozoic epoch that the first appearance of metazoans (multi-celled animals) occurred. This critical event may well have preceded the earliest Marinoan phase of deep glaciation. It has been suggested that the Earth s surface may well have been entirely ice covered during the deepest Veranger phase of glaciation and this has led to the widespread use of the *Snowball Earth* appellation to describe the climatic state that would have existed at this time (Hoffman *et al.*, 1993). The issue as to whether an entirely ice covered Earth, with thick continental ice sheets on the continents

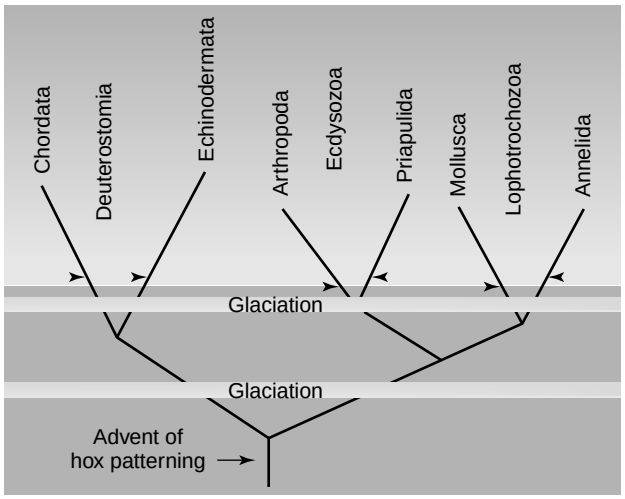


Figure 12 This diagram depicts the temporal association between the two phases of glaciation that occurred in the late Neoproterozoic (the Sturtian and Veranger events) and the onset of the Cambrian period at the beginning of which there began an intense acceleration in the process of biological evolution (from Runnegar, 2000. Reprinted with permission from *Nature*. Copyright 2000 Macmillan Magazines)

and sea ice covering the oceans, is required to satisfy the geochemical and geological data remains a subject of debate. If the Earth were entirely ice covered, there clearly exists a potentially serious problem for biological evolution because it is rather unclear how metazoans could have survived to serve as the required substrate for the extremely rapid proliferation and diversification of life that began at the outset of the Cambrian epoch (545 Ma). This problem is described graphically in Figure 12. Because the first appearance of metazoans preceded the episode of deepest glaciation and because it has been suggested on the basis of genetic clock analyses that these multi-celled animals were the ancestors of the Ediacaran fauna whose fossil remains first appear following the Varanger ice age, it seems evident that the metazoans were not only not eliminated by the snowball event, but that their rate of evolution may in fact have been accelerated because of it.

We may investigate the climate of the late Neoproterozoic ice age by applying the same ice sheet coupled climate model as was successfully employed to understand the Carboniferous glaciation of Pangea (see Hyde *et al.*, 2000 for detailed discussion). Figure 13 shows the south (magnetic) polar supercontinent of Pannotia, which was in place during the late stage of the Neoproterozoic. Based upon the results shown previously on Figure 8, it is evident that at 600 Ma, during the Varanger phase of glaciation, solar luminosity would have been reduced by approximately 6% from its modern level. Based on the results of the GEOCARB model shown on Figure 6, which shows large differences between the GEOCARB1 and GEOCARB2 reconstructions for this period, the atmospheric carbon dioxide level at that time is highly uncertain. We may nevertheless employ the model to ask whether a snowball state of complete glaciation could have occurred, and if so, at what level of atmospheric carbon dioxide concentration such conditions might arise. Results of an initial investigation of this kind are shown on Figure 14, part (a) of which shows land ice volume

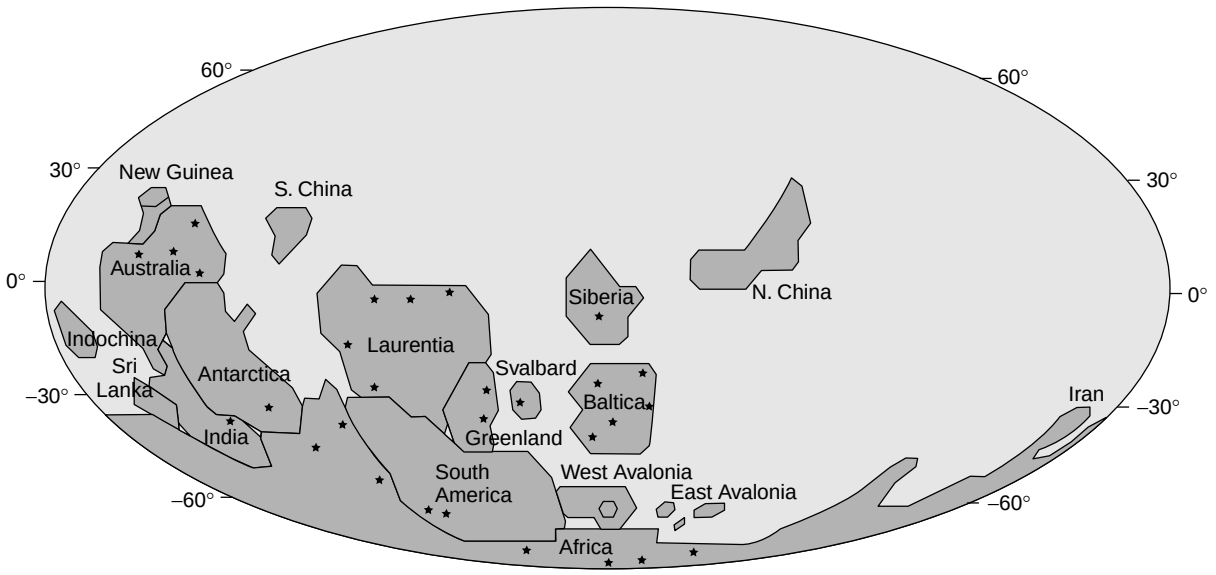


Figure 13 An approximate reconstruction of the supercontinent of Rodinia during the Veranger glaciation at 600 Ma. The stars on the individual continental fragments represent locations from which evidence exists of intense continental glaciation having occurred at this time (from Hyde *et al.*, 2000. Reprinted with permission from *Nature*. Copyright 2000, Macmillan Magazines)

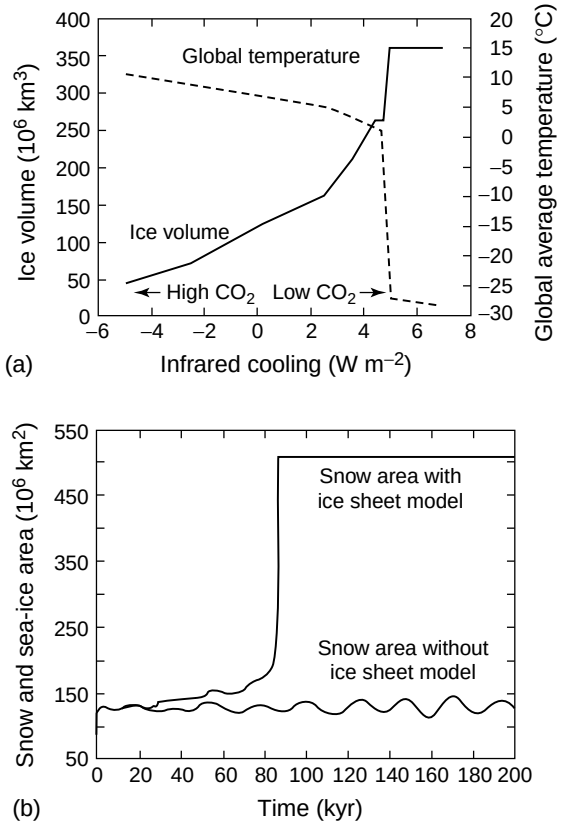


Figure 14 (a) Land ice volume and mean surface temperature on Rodinia as a function of the assumed degrees of IR cooling due to the decrease in atmospheric carbon dioxide concentration; (b) the impact on the evolution of the area of Rodinia covered by snow of excluding the ice-dynamics component of the model (from Hyde *et al.*, 2000. Reprinted with permission from *Nature*. Copyright 2000, Macmillan Magazines)

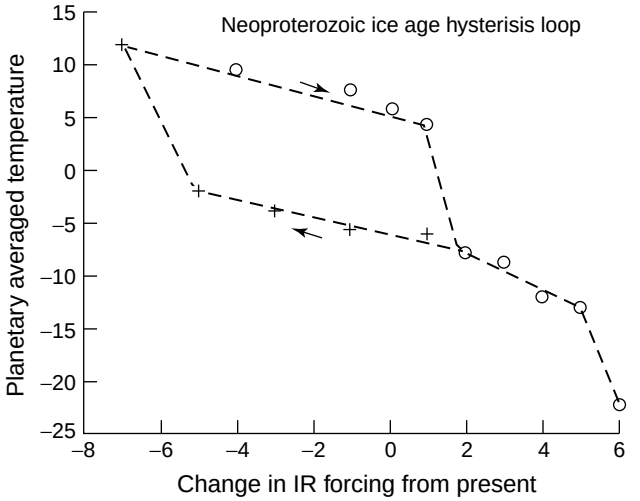


Figure 15 Mean surface temperature of the Earth as a function of the degree of IR cooling caused by varying the concentration of carbon dioxide in the atmosphere. Note the hysteresis involved in determining the thermal state of the system on cooling out of a hot state, as opposed to warming out of a cold state. On the cold branch of the hysteresis loop the solutions are characterized by states of heavy glaciation that include, however, a significant area of open water at the equator which could serve as an equatorial refugium in which complex marine life could have survived (from Crowley *et al.*, 2001)

and globally averaged surface temperature as a function of the degree of cooling (in $\mathrm{W\,m^{-2}}$ by infrared radiation (IR)) associated with the level of carbon dioxide reduction. Results indicate that at a level of IR cooling of $\sim 5\,\mathrm{W\,m^{-2}}$ the model exhibits a bifurcation into a state of total ice cover in which the mean surface temperature falls to a value near $-30\,^{\circ}\mathrm{C}$ (the average surface temperature at present is approximately $14.5\,^{\circ}\mathrm{C}$). Part (b) of the same figure demonstrates the extremely important role played in this model by the explicitly described continental ice sheet dynamics. In the absence of the physics associated with the accumulation and flow of the continental ice sheets, the bifurcation into the state of complete continental ice cover (represented here by the area of the surface that is covered by snow) does not occur even though the carbon dioxide level has been reduced to approximately half of its preindustrial value.

A somewhat more complete classification of the solution space of the ice sheet coupled energy balance climate model is shown in Figure 15, where the quasi-equilibrium planetary average temperature is plotted as a function of the degree of IR cooling in $\mathrm{W\,m^{-2}}$. Clearly evident on this diagram is that, in the range between -5 and $+2\,\mathrm{W\,m^{-2}}$ of IR cooling, there exist two possible quasi-equilibria, one of which corresponds to a relatively high mean surface temperature while the other corresponds to a relatively low mean surface temperature. The cold branch is a branch that I refer to as consisting of *oasis* solutions. These are solutions in which, although all of the continents are covered by thick continental ice sheets and sea ice covers all of the high latitude and some of the low latitude oceans, there nevertheless exists an equatorially confined open water refugium within which metazoans could clearly have survived. The hard snowball state occurs only for levels of IR cooling in excess of $6\,\mathrm{W\,m^{-2}}$. At lower levels of IR cooling, the issue of whether the solution lies on the hot branch or the cold oasis branch of the hysteresis loop is determined by the initial condition, which is perturbed by the change in carbon dioxide concentration. If the system begins to exit a cold state because carbon dioxide is increasing, say due to the influence of volcanic out-gassing, then one remains on the cold oasis branch until a critical degree of warming is applied. On the contrary, if one is cooling from a warm state because the carbon dioxide level is decreasing, then one remains on the warm branch of the hysteresis loop until a critical degree of cooling is applied.

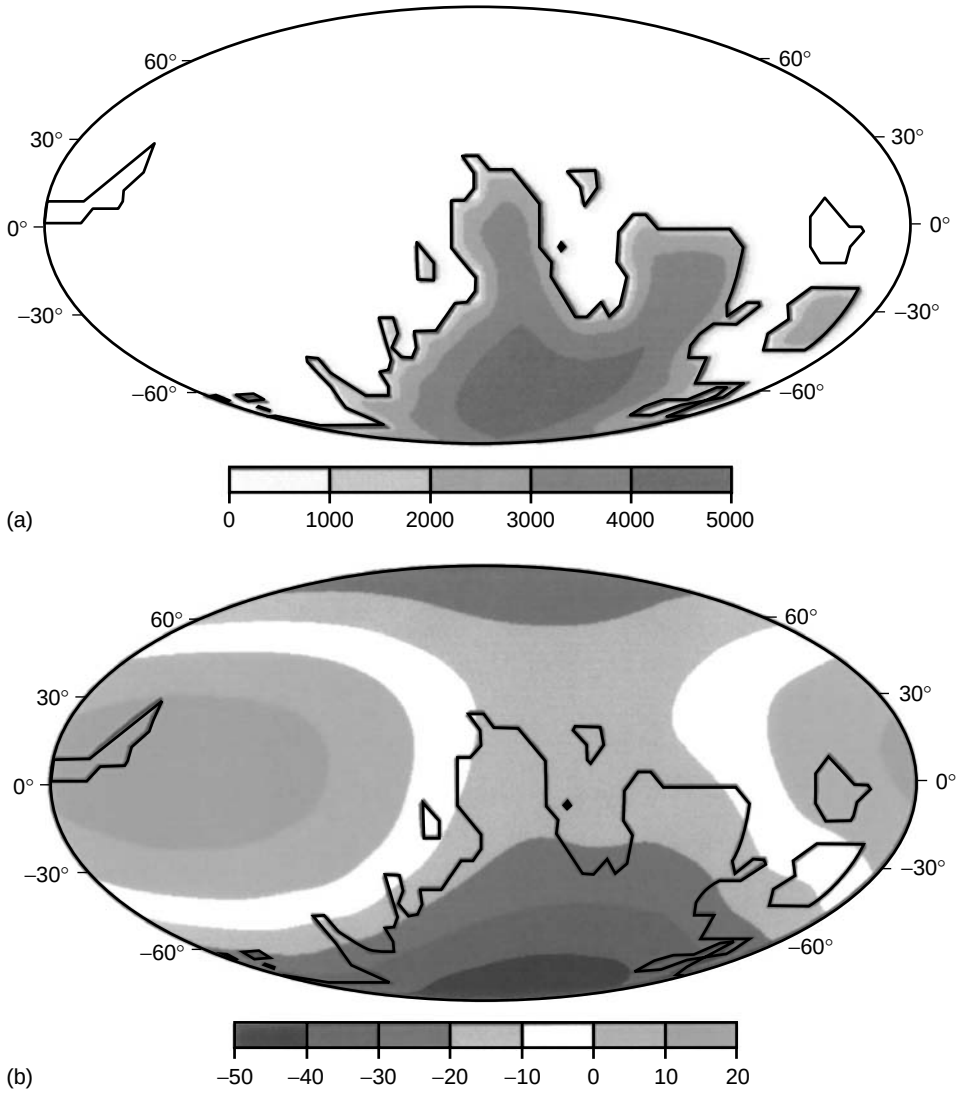


Figure 16 A typical oasis solution calculated by the University of Toronto intermediate complexity climate model for the case of the Late Neoproterozoic, with −6% solar forcing and the CO_2 concentration set at 75% of its preindustrial value. Panel (a) shows the elevation of the ice sheets (m); Panel (b) shows mean annual temperature (°C) (from Crowley *et al.*, 2001)

Figure 16 shows a typical oasis solution, in terms of mean annual temperature, for late Neoproterozoic climate with the solar constant reduced by 6% and the carbon dioxide level at three quarters of its preindustrial value. Although there exist continents near the equator that are covered by thick ice sheets, there is nevertheless a large area of the equatorial ocean which is free of sea ice.

It is important to remember that the simple climate model that is represented by the schematic Equations (1–3) is one in which many climatologically significant processes are not represented at all. These unrepresented processes include all impacts associated with ocean dynamics, both those due to the wind driven circulation and those due to the thermohaline circulation (THC). Also unrepresented are all influences upon radiative transfer processes associated with the presence of clouds in the atmosphere and with a changing water vapor concentration in the air caused by evaporation from the sea surface. Furthermore, the model contains only a crude thermodynamic treatment of the formation and movement of sea ice. Perhaps most importantly it contains no explicit treatment of the transports of heat and momentum in the atmosphere by the movement of the air. Although some of these processes are represented in parameterized form in the simple ice sheet coupled energy balance climate model, the parameterizations have been determined by tuning the model to the modern climate. Because the climate of the late

Neoproterozoic is profoundly different from that of the present day, one must be cautious in accepting the results delivered by a model of this type until they are confirmed by application of a more complete model of the climate system.

Modern atmosphere–ocean general circulation models are fully capable of being employed to provide the required tests of the simulations of such models of intermediate complexity. The GENESIS2 model (see Thompson and Pollard, 1995) has already, in fact, been applied in an initial test of the existence of the refugium solutions revealed by the simpler model (Hyde *et al.*, 2000). Although this test did not include any of the influences of ocean dynamics, it did include diurnal solar forcing, interactive clouds, semi-Lagrangian water vapor transport and a six-layer sea ice model. In this model, simulations with 0.5 and 1.0 times the preindustrial carbon dioxide levels quickly delivered oceans that were entirely ice covered whereas a doubled carbon dioxide experiment resulted in ice free tropical oceans, although the model continued to drift to lower temperatures. At a level of 2.5 times the preindustrial atmospheric carbon dioxide concentration, the solution was characterized by an apparently stable, equatorial open water refugium that covered the entire equatorial belt. A much more complete model of the climate system than GENESIS2 is clearly required to incorporate the full influence of ocean dynamics on the climate of the late Neoproterozoic.

The Climate System Model (CSM) of the US National Center for Atmospheric Research has been employed to perform this additional test and results are shown in Figure 17. They characterize an apparently stable solution at $1 \times CO_2$ and 6% solar constant reduction in terms of surface air temperature, sea surface temperature (SST), precipitation and sea ice extent. The solution is clearly characterized by an extensive equatorial open water refugium. The zonally averaged overturning stream function characteristic of this solution, shown in Figure 18, is representative of the THC. The results demonstrate that the deep circulation is rather weak in this apparently stable state, which is consistent with the anoxia of the oceans that was apparently characteristic of the Veranger glacial epoch. The fact that this preliminary oasis solution has been obtained with the atmospheric carbon dioxide concentration at the preindustrial level, whereas the ocean dynamics free GENESIS2 model delivers a hard snowball climate under these conditions, is likely to be indicative of the profound impact of the dynamics of the oceans upon late Neoproterozoic climate. This is believed to be the case because ocean dynamics is acting as a negative feedback, which tends to inhibit the formation of sea ice at the equator. Although the computation of climate states such as this using fully coupled atmosphere–ocean GCMs is very expensive in terms of the computational resources required, the cost effectiveness of this technology continues to improve. The near future of such investigations will therefore include intensive analyses of the solution space of these most complex characterizations of ancient climate regimes.

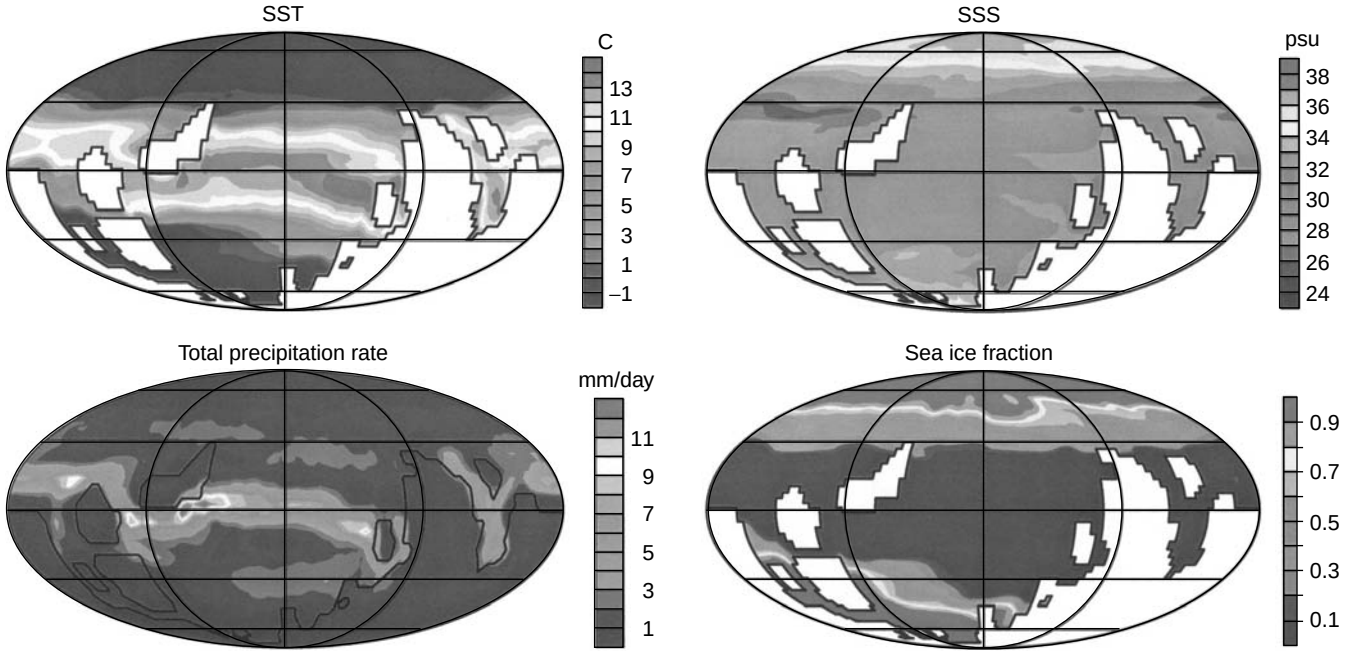


Figure 17 Characteristics of the climate of the Neoproterozoic snowball epoch as determined using the fully coupled atmosphere–ocean general circulation model of the US National Center for Atmospheric Research. The individual plates of the figure display SST, sea surface salinity (SSS), precipitation rate and sea ice fraction

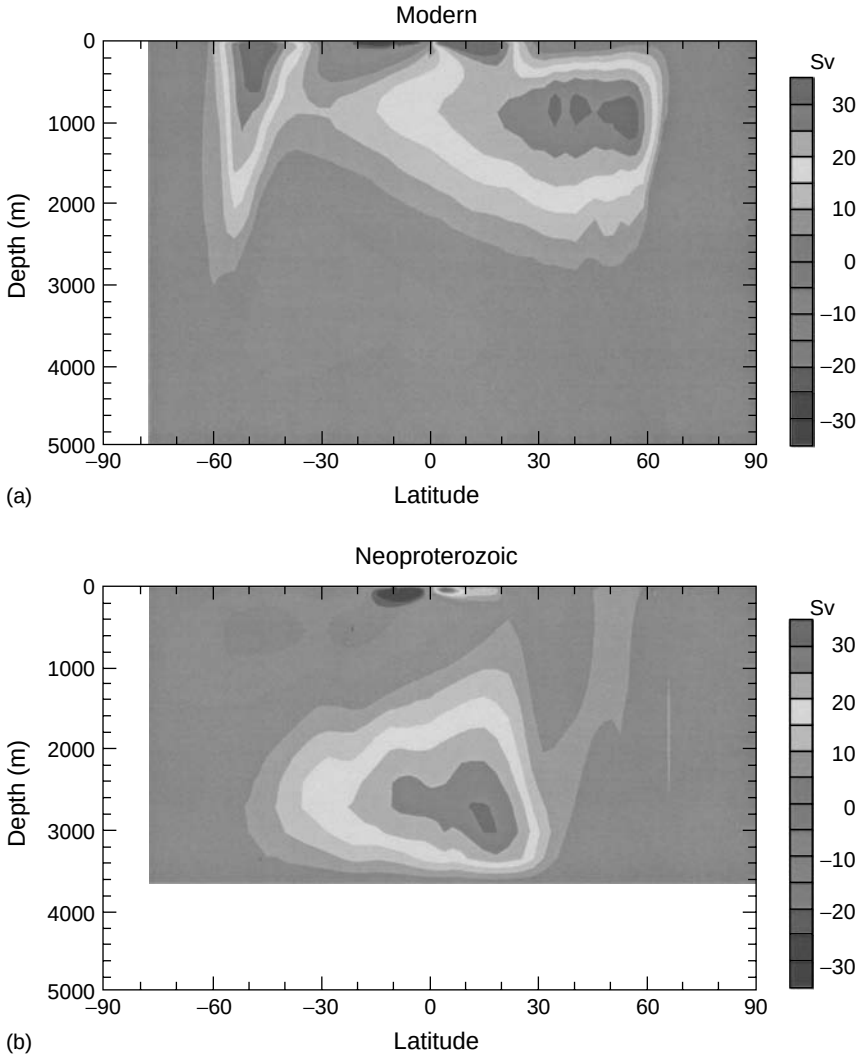


Figure 18 Shows zonal average of the meridional overturning streamfunction in Sverdrups (1 Sv = $10^6\,m^3\,s^{-1}$) for the global ocean for both the (a) Modern and (b) Neoproterozoic climates. For the purpose of the latter simulation the oceans were assumed to be of a uniform depth of 4 km

THE LATE PLEISTOCENE ICE AGE: CARBON DIOXIDE COVARIATION IN ORBITALLY FORCED GLACIAL CYCLES

Although the base of geological data that may be invoked to constrain the ancient glacial climates of the Carboniferous and Neoproterozoic epochs is rather modest, the same is not the case for the Pleistocene epoch. This epoch extended from approximately 2.4 Ma until the onset of the current Holocene period that began at approximately 10 ka. Because the oldest ocean floor on the planet is everywhere less than approximately 150 Ma in age, geological data from the ocean floors could not be invoked to constrain the characteristics of the Carboniferous and earlier glacial epochs, as it is entirely absent. During the Pleistocene epoch, however, such data are extremely abundant and furthermore may be placed accurately on a calendar year time scale using both direct and indirect methods of dating. The most important data from the ocean floors consist of mass spectrometric measurements of the stable isotopes of oxygen taken up by foraminifera extracted from deep sea sedimentary cores. The isotopic ratio, generally expressed as $\delta^{18}O$, is defined as follows (Equation 9):

$$\delta^{18}O = \frac{[^{18}O]}{[^{16}O]} - \text{SMOW} \tag{9}$$

in which $[^{18}O]$ and $[^{16}O]$ are, respectively, the concentrations of the heavier and less abundant and lighter and more abundant isotopes of oxygen measured on the test of

a particular species of foraminifera, whereas SMOW represents the same ratio measured in standard mean ocean water. The utility of the $\delta^{18}O$ measurement as a proxy indicator of climate variability, when measured in a deep sea sedimentary core, arises because the climate related processes of evaporation and precipitation are processes that fractionate mass. When water vapor is produced by evaporation from the surface of the sea, the influence of isotopic fractionation causes the vapor to be isotopically lighter (richer in ^{16}O) than the mean molecular weight of the oceanic reservoir from which it was extracted. The more water that is irreversibly extracted from this reservoir and locked into ice sheets on the continents, the heavier isotopically becomes the ocean that is left behind. While they live, the foraminifera, both benthic species living at the ocean bottom and planktonic species living near the surface, record the characteristic $\delta^{18}O$ of the ocean reservoir. When the foraminifera die, their shells fall to the ocean floor, younger overlying older as this detritus accumulates. An excellent proxy for the amount of land ice that existed on the continents as a function of time can thus be reconstructed by measuring the $\delta^{18}O$ of foraminifera extracted from progressively greater depth in a given sedimentary core. Although the isotopic ratio is also sensitive to temperature, the temperature contamination of the ice volume signal is relatively modest when based on measurements from a benthic species, which by definition inhabit the abyssal ocean.

An example of one such set of measurements from Ocean Drilling Project (ODP) core 677, which was drilled from the Panama Basin off the west coast of the northern part of South America, is shown in Figure 19(a). The

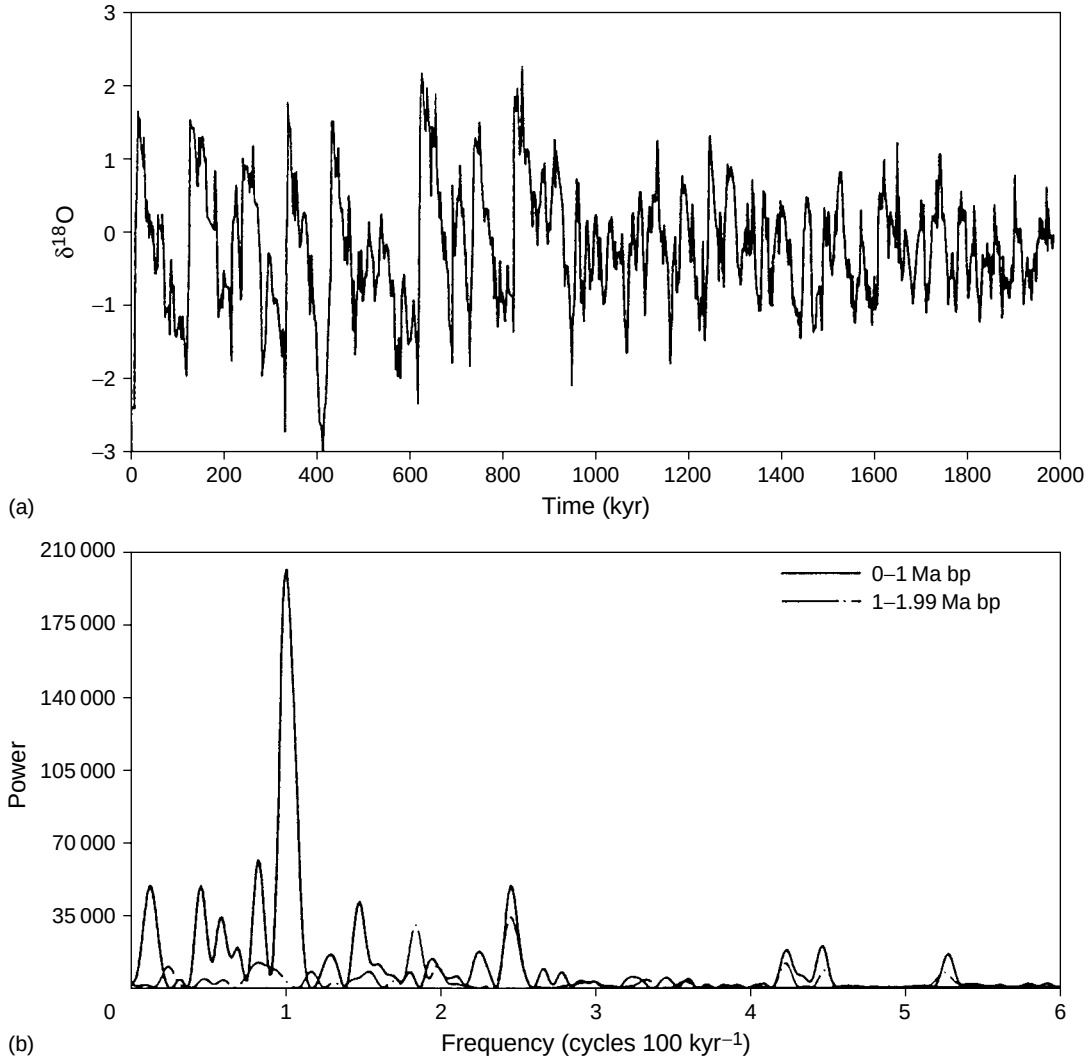


Figure 19 (a) The $\delta^{18}O$ record from the Panama Basin ODP 677 deep sea sedimentary core on an orbitally tuned time scale. (b) Power spectra of the most recent and immediately preceding million year segments (Ma BP – millions of years ago, before present) of the record in part (a)

mass spectrometrically measured $\delta^{18}O$ data are plotted in Figure 19(a) on a calendar age time scale rather than as a function of depth, the depth versus age relation having been determined by tuning to the orbital insolation anomaly time series for Northern Hemisphere summer delivered by Equation (4), when the orbital reconstruction of Laskar (1988) is employed to define the time series of $\varepsilon(t)$ and $e(t)$. The most striking demonstration of the validity of this method of determining a calendar year chronology for the isotopic record from a deep sea sedimentary core was provided by employing it on the ODP 677 data set and showing that if the method were correct, then there would have to be a significant error in the previously assumed age of 730 000 years for the most recent Brunhes–Matuyama reversal in the direction of the Earth s magnetic field. Subsequent to the calculation by orbital tuning that the age of this reversal was actually 780 000 years (Shackleton *et al.*, 1990), the new age was confirmed by applying the $^{40}Ar/^{39}Ar$ dating technique to a suite of volcanic rocks from Hawaii in which the reversal was recorded.

Inspection of the $\delta^{18}O$ data on the orbitally tuned time scale immediately reveals that throughout the latter half of the Pleistocene epoch, beginning about 900 000 years ago, continental ice volume has varied in an oscillatory fashion with a dominant characteristic time scale very close to 100 000 years. This is confirmed by the power spectral analyses of the $\delta^{18}O$ record shown in part (b) of Figure 19. Whereas the power spectrum of the earlier 1 Ma of the record has essentially no power at this period, during the most recent 1 Ma the variance in the time series is strongly focused at this time scale. Also evident in both of these power spectral partitions, however, is the spectral line at a period of 41 000 years which is the dominant time scale on which the orbital obliquity ε varies, as well as the triplet of spectral lines at periods near 19 000, 21 000 and 23 000 years (*see* Orbital Variations, Volume 1). The origin of this triplet of spectral lines may be understood by inspection of Equation (4), which shows that the variation of orbital eccentricity e, which has a dominant period of 100 000 years, acts so as to modulate a variation of insolation that would otherwise occur at the precession period of 21 000 years. The Fourier decomposition of the low frequency eccentricity modulation of the higher frequency precessional singlet gives rise to the splitting of the singlet into the multiplet seen in this spectrum on Figure 19(b) (to use the language of spectroscopy). The power spectrum of the time series of orbital insolation itself contains effectively no power at the period of the eccentricity modulation of the precessional effect, as is clearly evident by inspection of Equation (4). The important climatological issue that is posed by the proxy ice volume time series from ODP core 677 therefore concerns the nature of the nonlinear effects that must be involved in transforming the orbital climatological forcing into a climate system response that is dominated by variance at a period that is not present in the power spectrum of the excitation.

In order to address this important issue, one additional piece of information is required concerning the way in which other climatologically important Earth system properties vary during a typical 100 000 year late Pleistocene ice age cycle. To this end Figure 20 shows a sequence of such records from both ice cores ($\delta^{18}O$ from Camp Century Greenland, deuterium from Vostok, Antarctica, carbon dioxide from Vostok, Antarctica, and dust from Vostok, Antarctica), deep sea sedimentary cores ($\delta^{18}O$ from a stack of records from the SPECMAP suite of deep sea sedimentary cores), and lake cores (a paleo-precipitation record based upon a palynology based reconstruction from La Grande Pile). These records all encompass the most recent 100 000 year ice age cycle and demonstrate that these various climate proxy data sets all exhibit a high degree of correlation with the ice volume reconstruction based on deep sea sediment. Of greatest interest from the current perspective, however, is the strong correlation between the record of carbon dioxide concentration in the atmosphere measured on air bubbles in the Vostok ice core and the proxy record of continental ice volume. This correlation is clearly such that when ice volume is greatest, carbon dioxide concentration is lowest and vice versa, once more suggesting that atmospheric carbon dioxide concentration is playing a critical role in the process of continental glaciation and deglaciation. A critically important question is whether the apparent covariation of carbon dioxide concentration with ice volume occurs because carbon dioxide is a prime mover of the ice age cycle or whether it is simply a, perhaps weak, amplifier of the cycle. Note that the net variation of the carbon dioxide concentration through a typical cycle is approximately 80 parts per million by volume (ppmv) which is to be compared to the mean pre-industrial level of 280 ppmv so that in the state of deepest glaciation the atmospheric carbon dioxide concentration is reduced by almost 30%, not an insignificant depression as we will see.

We may address the issue of the mechanics of the glaciation–deglaciation process by once more invoking the simple ice sheet coupled energy balance climate model that is embodied in the schematic Equations (1–3) (*see also* Climate Model Simulations of the Geological Past, Volume 1). In comparison with the integrations of this model that were employed to investigate the Carboniferous and Neoproterozoic glaciations, for the purpose of which the atmospheric carbon dioxide level was held fixed, however, we will force the model for present purposes with both orbital insolation variations and with the variation of atmospheric carbon dioxide concentration. For the purpose of these integrations, furthermore, the initially non-glaciated surface topography and land–sea distribution of the planet are taken equal to their modern values, the spatial distribution of the precipitation rate is taken equal to its modern

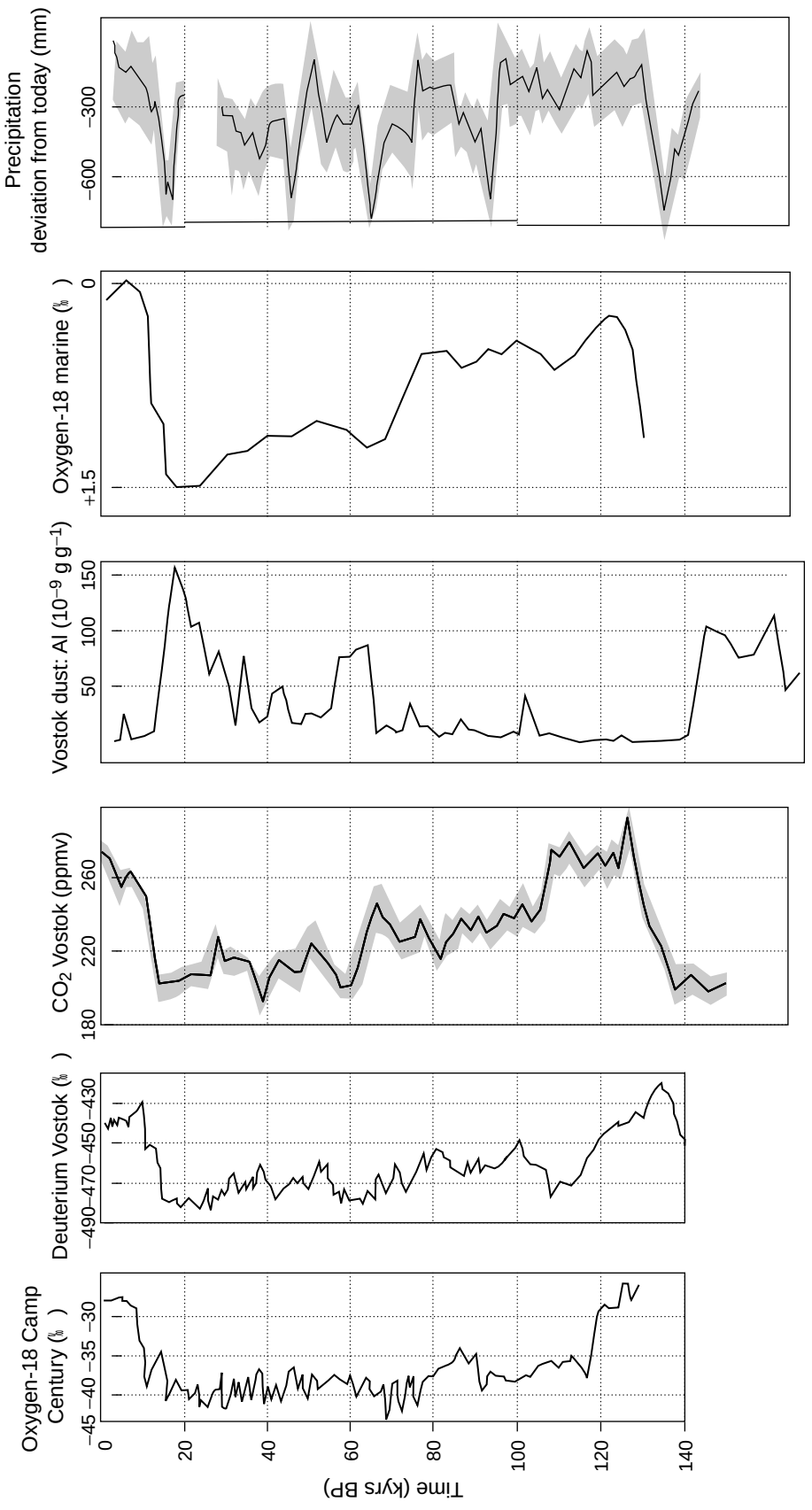


Figure 20 Six proxy climate time series for the most recent 100 kyr ice age cycle. This graphic was produced by the international Past Global Changes (PAGES) core project of the International Geosphere–Biosphere Programme using data from sources indicated

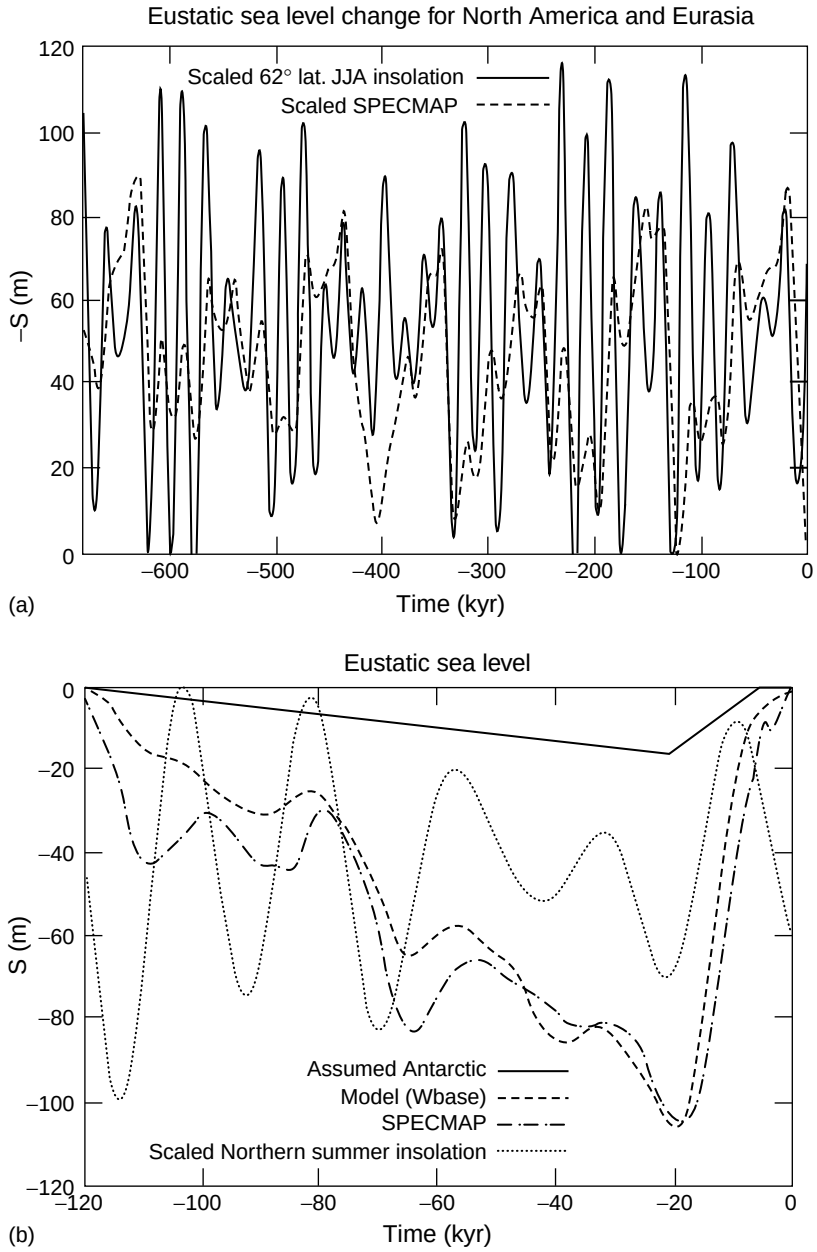


Figure 21 (a) June-July-August (JJA) orbital insolation anomaly at 62 °N latitude (scaled) and the SPECMAP $\delta^{18}O$ record derived on the basis of deep sea sediments which is scaled to the contribution to eustatic sea level history from the deglaciation of North America (NA) and Eurasia (EA). (b) Eustatic sea level history predicted by the ice sheet coupled climate model (Wbase) compared with the last glacial cycle from the SPECMAP record and to (scaled) Northern Hemisphere summer insolation

pattern although the rate is modulated by the surface air temperature in order to account in an approximate way for the fact that the hydrological cycle is less intense under colder conditions.

Figure 21 shows, in plate (a), the (scaled) June, July, August (JJA) insolation anomaly used to force the model together with the SPECMAP proxy for land ice volume over the period subsequent to about 650 ka. Plate (b) of the same Figure shows the same information concerning the orbital climate forcing and the $\delta^{18}O$ proxy for ice volume response together with the output ice volume response from the orbitally forced ice sheet coupled climate model over the time interval from the Eemean interglacial period at 120 ka (oxygen isotope stage (OIS) 5e) to the present. Included in the ice volume response curve is an assumed 15 m eustatic sea level variation derivative of (primarily West)

Antarctica. Intercomparison of the proxy ice volume record and the model simulated record shows that the theoretical model result mimics the observations extremely well, being characterized by a relatively slow glaciation phase lasting approximately 90 thousand years (kyrs) and a relatively rapid deglaciation phase lasting approximately 10 kyrs. The canonical 100 kyr ice age cycle, which has the character of a nonlinear relaxation oscillation, is therefore well explained by a model incorporating a relatively simple set of key processes.

Figure 22 illustrates the detailed structure of the Laurentide ice sheet that is calculated by the model at Last Glacial Maximum (LGM) at 21 ka, both in terms of the surface topography of the ice complex and in terms of the basal temperature distribution. These results demonstrate that the model not only delivers the correct time dependence of the growth and decay of the mass of glacial ice, but also the distinct Cordilleran, Laurentian and Innuitian ice sheet components. Inspection of the temperature distribution at the base of the ice sheet shows that the LGM Laurentide and Innuitian ice sheets are expected to have been frozen to the bed whereas the Cordilleran ice sheet may have been near the melting point at its base.

A fundamental question that clearly arises concerning this theory of the 100 kyr ice age cycle concerns the physical processes that drive the sharp terminations which characterize the deglaciation phase of the cycle. Detailed analyses demonstrate that there are three interlinked contributions, which together control the occurrence of these events. Firstly, there is the strength of the orbital insolation forcing, which is modulated at the period of 100 kyrs by the variation of orbital eccentricity. Secondly, there is the temporal variation of the atmospheric carbon dioxide concentration, which, as shown previously in Figure 20, is itself characterized by a slow decrease during glaciation and a rapid rise during deglaciation. Finally, there is the physics of the process of glacial isostatic adjustment (GIA), which introduces a strong positive feedback on the process of deglaciation by virtue of the enhancement of the temperature in the ablation zone of the ice sheet as it melts back into the deep isostatic depression created by the LGM surface load.

All three of these ingredients appear to be required in order for the ice sheet coupled climate model to deliver an acceptable synthetic ice volume history. Because the carbon dioxide forcing is the only one of these three ingredients that is not explicitly modeled but rather is assumed known on the basis of the measurement of carbon dioxide from air bubbles in the Vostok ice core, and because the carbon dioxide record so precisely contains the 100 kyr sawtooth cycle itself, it seems clear that we will not be able to claim to fully understand this intense climate oscillation until we are able to fully explain why atmospheric carbon dioxide concentration varies in this way. At present there

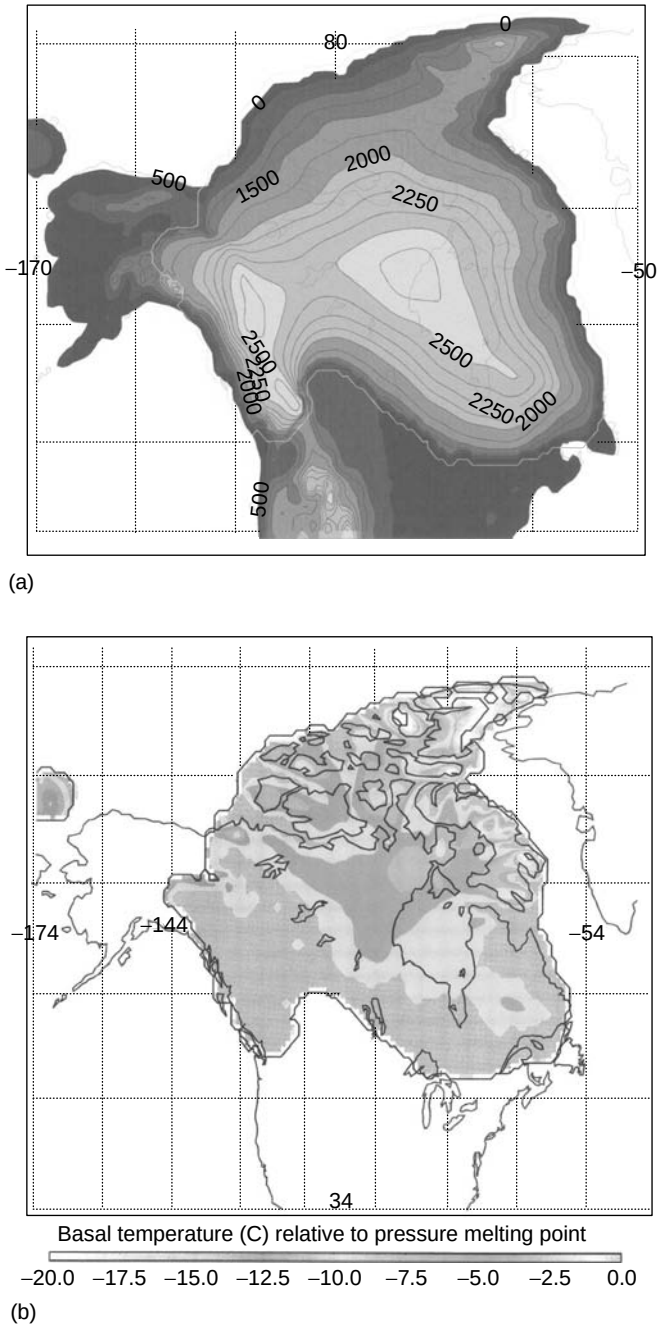


Figure 22 (a) Elevation with respect to sea level at LGM of the Laurentide ice sheet over North America as simulated by the ice sheet coupled climate model described in the text; (b) temperatures at the bedrock ice sheet interface for the Laurentide ice sheet simulated by the climate model (from Tarasov *et al.*, 1999)

is no generally accepted explanation although there does exist a number of potentially explanatory hypotheses. These inevitably invoke some combination of a biological pump and a solubility pump. The latter relies simply upon the

fact that the equilibrium concentration of carbon dioxide dissolved in the surface waters of the oceans depends upon temperature such that the lower the temperature the greater the amount of carbon dioxide that will be drawn down from the atmosphere into the oceans. As the system cools during glaciation we must therefore expect the atmospheric carbon dioxide concentration to fall for this reason alone; this fall of the CO_2 concentration would continue so long as deep water continued to form and the thermohaline circulation (THC) remained active.

Although no fully satisfactory demonstration has been provided that this influence could not be dominant in

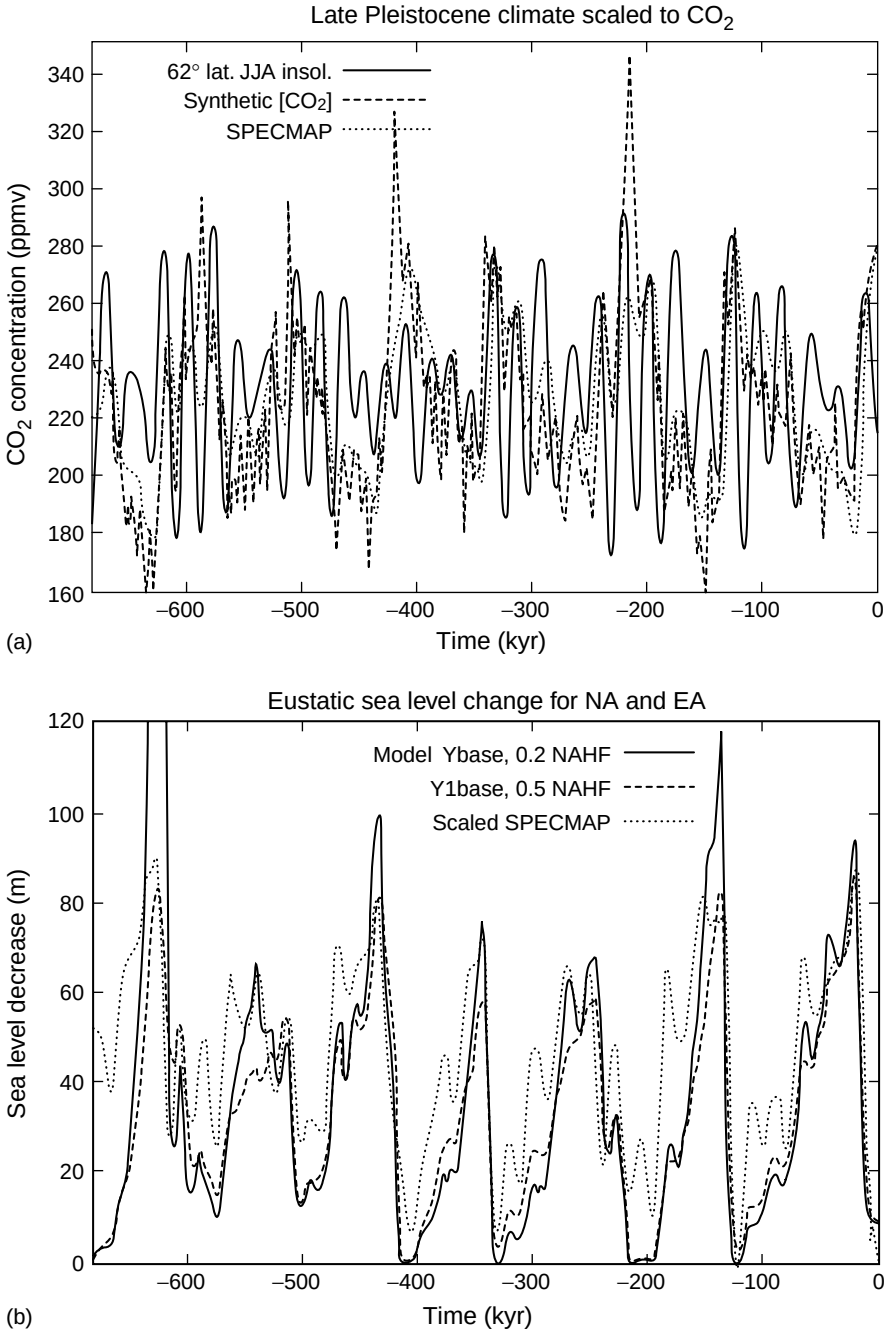


Figure 23 (a) JJA insolation and carbon dioxide forcing compared to the SPECMAP ice volume proxy for the most recent 650 kyr of Earth history; (b) Ice volume histories (represented in terms of decrease in eustatic sea level for North America (NA) and Eurasia (EA)) simulated using the ice sheet coupled climate model for two versions of the model which differ as to the amount of additional forcing due to variation through time of the strength of the thermohaline circulation. (Ybase and Y1base), both of which are compared to the SPECMAP ice volume proxy

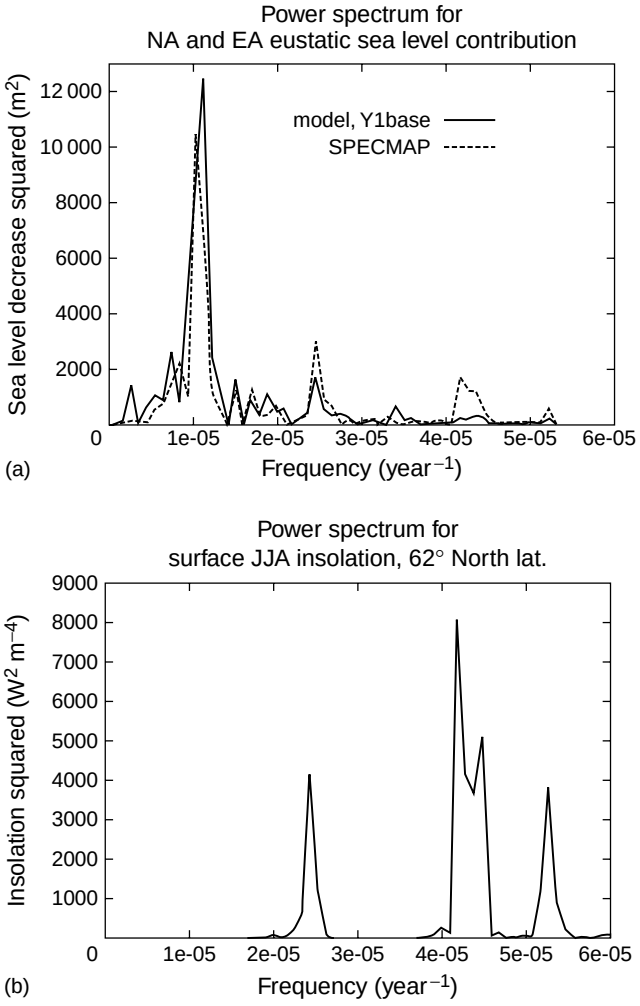


Figure 24 (a) Comparison of the power spectra of the Y1base model of ice volume variability from Figure 23 with the power spectrum of the SPECMAP record; (b) power spectrum of the JJA insolation anomaly employed to force the Y1base model

explaining the 100 kyr cycle in the CO_2 concentration, it is generally believed that the biological pump is the primary mechanism involved in the fall of the CO_2 concentration under glacial conditions. One hypothesis as to how this might work involves the iron fertilization of those regions of the present ocean in which biological productivity is low because of the low level of this element. Inspection of Figure 20 shows that glacial conditions are characterized by strongly enhanced levels of terrigenous dust in the atmosphere. It has been suggested that the iron content of this dust might act so as to dramatically increase net ocean productivity during glacial times, thereby explaining the carbon dioxide cycle. The existing lack of a consensus explanation of the coevolution of continental ice volume and atmospheric carbon dioxide concentration constitutes one of the most significant outstanding problems in paleoclimatology. It is conceivable that these pumps could operate asymmetrically with solubility dominant during glaciation, an asymmetry that could explain the recent results of Shackleton (2000).

In spite of this outstanding problem, we may further probe the ability of the above described model of the 100 kyr ice age cycle by employing a longer input record of the atmospheric CO_2 concentration, four ice age cycles of such data now existing from Vostok, in conjunction with a longer input record of the astronomical forcing. Figure 23(a) shows both these inputs to the model, together with the SPECMAP ice volume proxy, while Figure 23(b) shows two different versions of the ice volume simulation (in terms of the implied eustatic sea level variation) along with a scaled version of the SPECMAP record. These two versions of the model of the 100 kyr cycle include two different levels of the atmospheric heating and cooling that is assumed to occur due to time variations in the strength of the thermohaline circulation as the Earth system undergoes transitions from glacial to interglacial conditions. This additional thermal influence has very little effect on the glaciation and deglaciation of the North American continent. The same is not true for the northwestern European ice complex, which lies downstream in the atmosphere from the main regions of deepwater formation in the Greenland, Iceland and Norwegian (GIN) Seas. Figure 24(a) shows the power spectrum of one of the synthetic ice volume records from Figure 23 together with the power spectrum of the SPECMAP time series itself. Comparison of these two spectra demonstrates that the model appears to successfully explain the dominance of variability on the 100 kyr time scale even though the orbital forcing has no power at all at this period.

MILLENNIUM TIME SCALE CLIMATE VARIABILITY UNDER GLACIAL CONDITIONS: DANSGAARD–OESCHGER OSCILLATIONS, HEINRICH EVENTS AND THE BÖLLING–ALLERØD YOUNGER–DRYAS HOLOCENE TRANSITION

Although the simple ice sheet coupled climate model employed as a vehicle for understanding episodes of glaciation in both the distant and more recent past has proven to be very effective for this purpose, it fails entirely when confronted with observations of ice age climate variability on time scales that are sub-orbital. The primary source of observational data that constrain the nature of climate variability on the millennium time scale is once more the isotopic records of the $\delta^{18}O$ ratio, but now measured in ice core ice rather than the tests of foraminifera from deep sea sedimentary cores. Figure 25 shows the $\delta^{18}O$ record from the European Greenland Ice Project (GRIP) ice core that

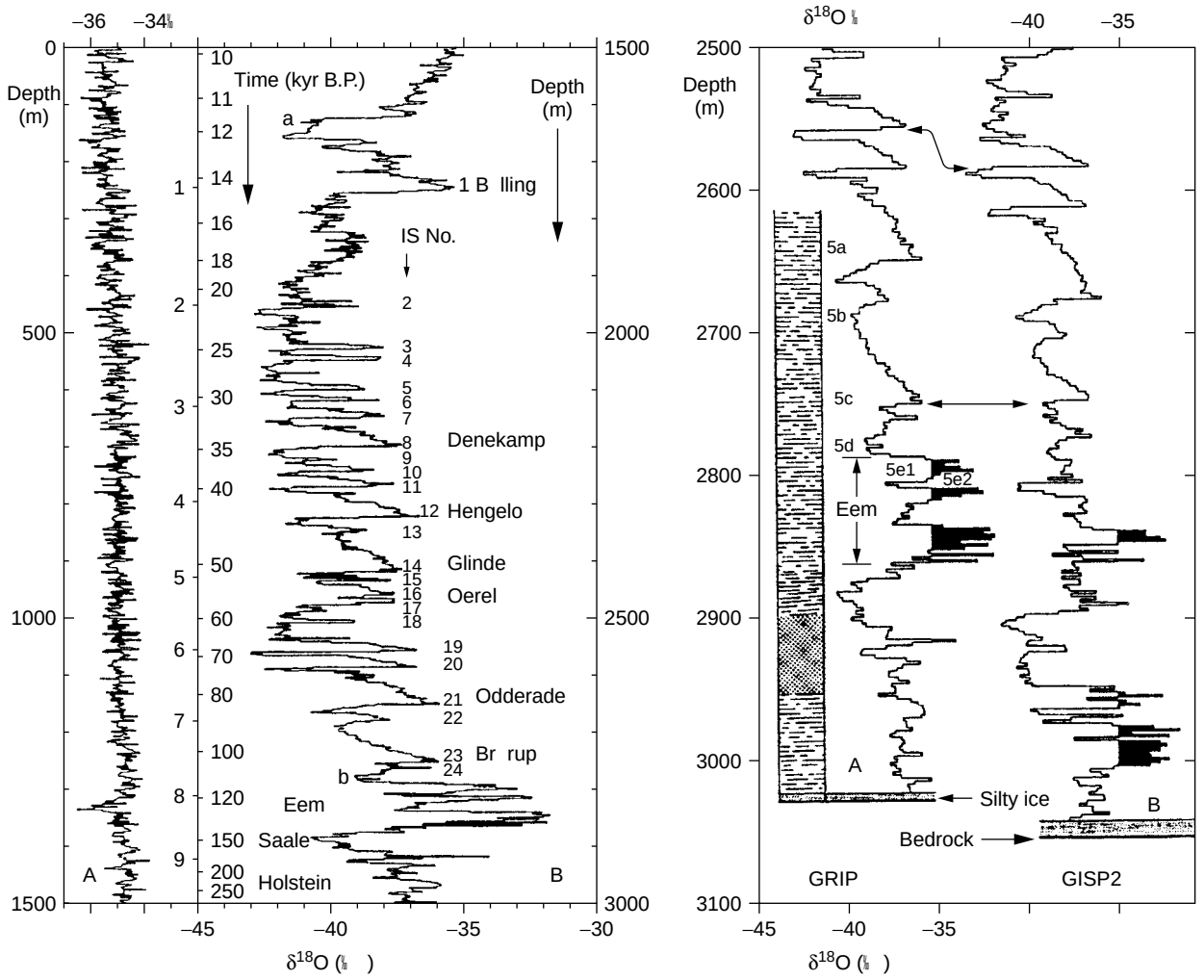


Figure 25 The $\delta^{18}O$ record from the deep ice core drilled at Summit, Greenland by the European GRIP project. The Eemian segment of this record is also compared to the equivalent segment from the American GISP2 record obtained from a deep ice core drilled at a nearby location (from Johnsen *et al.*, 1994)

was drilled at Summit, Greenland. This record is essentially identical to the American Greenland Ice Sheet Project 2 (GISP2) record except in the OIS 5 section, which is also shown on this Figure for both cores. When measured in ice sheet ice, the $\delta^{18}O$ climate proxy is indicative of the temperature of the air from which the precipitation was derived that formed the ice. Inspection of this Figure, in which the data are shown both as a function of depth in the core and as a function of time in calendar years, demonstrates that the temperature of the air over Greenland has been rather stable during the Holocene epoch, except perhaps for the event which occurred near 8.2 ka, whereas during marine OIS 3, which extends from approximately 30 ka–~65 ka, the record is characterized by the occurrence of large amplitude millennium time scale variability. The most recent calibrations of the $\delta^{18}O$ air temperature thermometer suggest that the peak to peak variations in air temperature during these oscillations are on the order of at least 10 °C (*see* Climate Change, Abrupt, Volume 1). That this millennium time scale variability is recorded not only in Greenland ice but also in the oceans is clear on the basis of the oxygen isotopic and other records shown in Figure 26. All these records are from a deep sea sedimentary core (MD 2024) that was drilled on Orphon Knoll, a site located at ~50 °N latitude and 45 °W longitude, to the southeast of the Labrador Sea. This is a site at which the deep sea sedimentary isotopic record samples the Western Boundary Under Current. This deep current of the thermohaline circulation is one in which deep water flows southwards. The Orphon Knoll site is especially important as it lies in the path of the iceberg discharges that occurred episodically through Hudson Strait due to calving instabilities on the eastern flank of the Laurentide ice sheet. These discharge events, called Heinrich events (*see* Heinrich (H-) Events, Volume 1), are clearly marked in Figure 26 by the symbols H0–H6 which correspond to

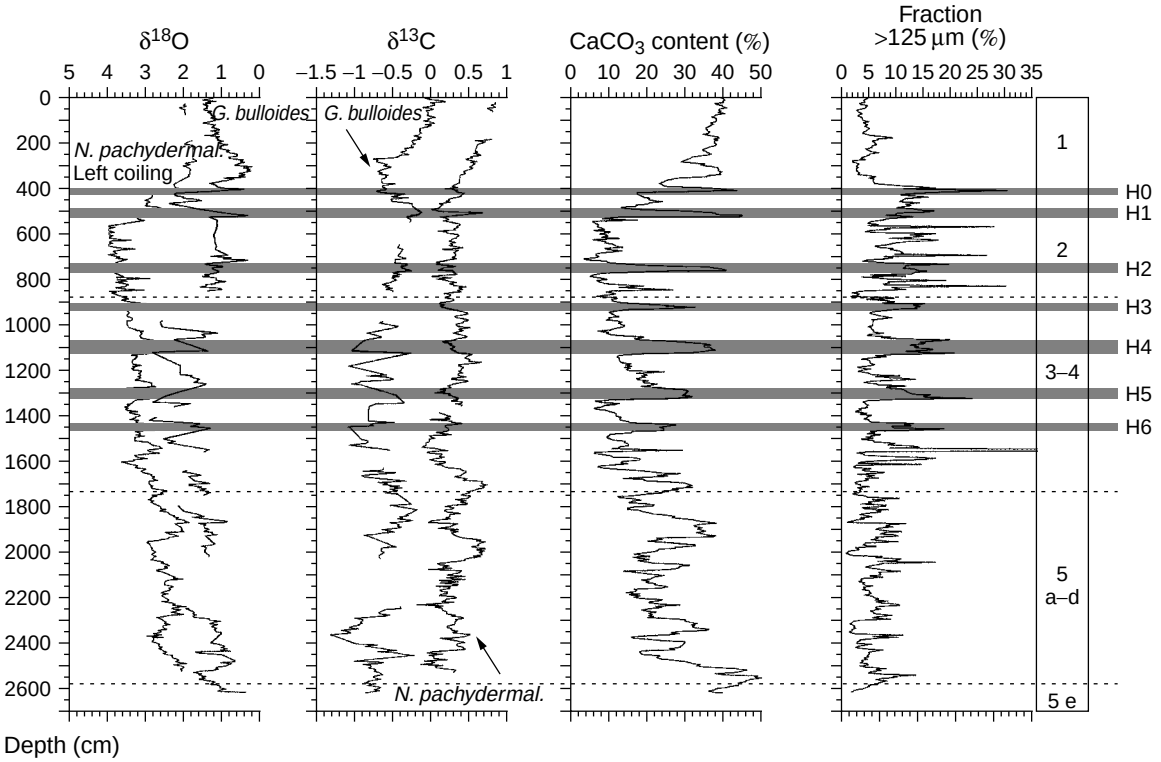


Figure 26 Heinrich event data from the deep sea sedimentary core MD-2024 from the Orphan Knoll in terms of $\delta^{18}O$, $\delta^{13}C$, $CaCO_3$ content and size fraction >125 μm data. The horizontal bars denote the times of occurrence of Heinrich events H0–H6 (from Hillaire-Marcel and Bilodeau, 2000)

horizons in the sedimentary core in which the concentration of ice rafted debris is exceptionally high. This can be seen clearly in the final part of this figure, which shows the fraction of particles of size greater than 125 μm as a function of depth in the core. Detailed intercomparison of the ice core (Figure 25) and sedimentary core (Figure 26) records shows that Heinrich events labeled H3–H6 appear to precede the most pronounced rapid warming events recorded in the European GRIP ice core; these warming events are named Denekamp, Hengelo, Glinde and Oerel, respectively. These Heinrich events therefore appear to coincide with cold phases of the millennium time scale variability, which end in a much more rapid episode of warming. A single millennium time scale oscillation therefore appears to have the same relaxation oscillation form as does the previously discussed 100 kyr cycle itself, in that the cooling phase is relatively long lived whereas the warm phase is extremely abrupt. Of particular note is the fact that the most intense warming phases of the millennium time scale variability, which follow the individual Heinrich events, are themselves followed by a sequence of progressively weaker millennium time scale oscillations. From the cold ambient climate of the glacial period to a relatively warm interstadial climate (the individual interstadials that occur during marine OIS 3 are denoted IS on Figure 25). Because this pattern is repeated several times during OIS 3, it has come to be referred to as the Bond cycle after its discoverer, Gerard Bond of the Lamont–Doherty Earth Observatory of Columbia University. The basic millennium time scale variability is referred to as the Dansgaard–Oeschger (D–O) oscillation and was first discovered by the persons for whom the oscillation is named on the basis of analyses of oxygen isotopic data derived from the earliest Greenland ice cores (*see* Dansgaard- Oescheger Cycles, Volume 1). It is clearly an extremely interesting question in environmental physics as to the reason for the existence of this complex, large amplitude, sub-orbital millennium time scale phenomenology.

The origins of these rapid oscillations between cold glacial climate and warm interstadial conditions appear to involve an interaction between the large continental ice sheets that surrounded the North Atlantic basin under glacial conditions, as illustrated in Figure 27, and the thermohaline circulation of the ocean(s) itself. Although the basic mechanism of oscillatory behavior has yet to be demonstrated in a complete coupled atmosphere–ocean–sea ice–land surface processes model, several models of intermediate complexity have shown behavior that is consistent with the primary ice core and deep sea sedimentary core observations. The basic geometry of one such intermediate complexity model, shown on Figure 28, consists of a series of three north south oriented two-dimensional (latitude–depth)

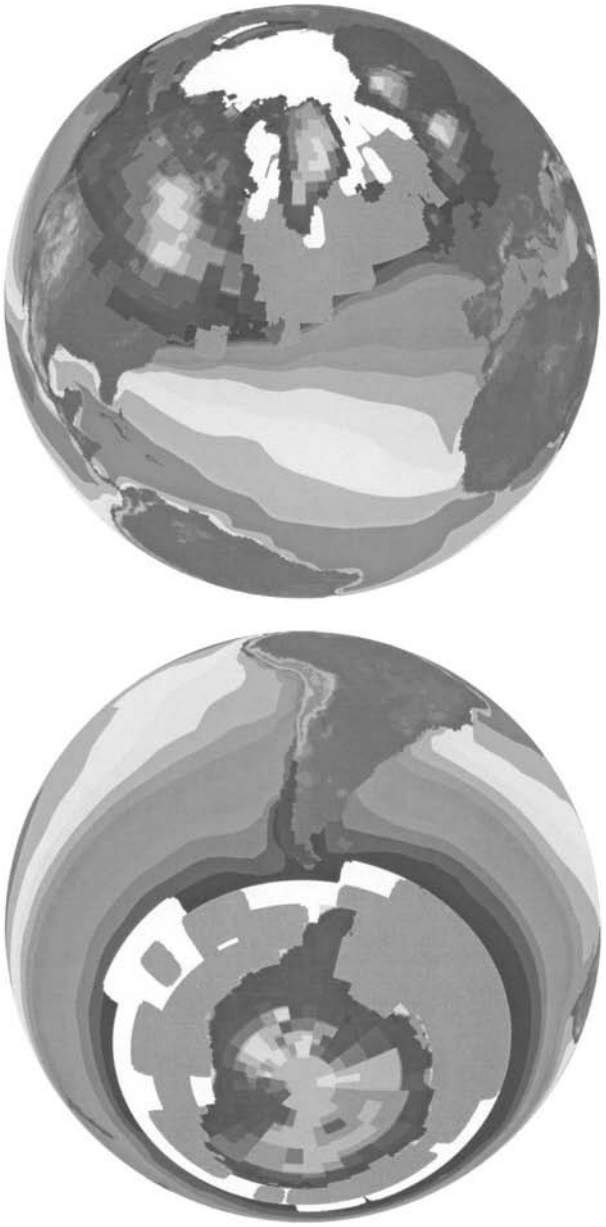


Figure 27 Northern and Southern Hemisphere land ice and sea ice extents at LGM approximately 21 000 calendar years ago. Also shown, in various shadings, are the sea surface temperatures (lighter shadings are warmer). Northern Hemisphere summer sea ice is shown as white whereas the Northern Hemisphere winter sea ice extent is shown as a grey extension of the winter distribution

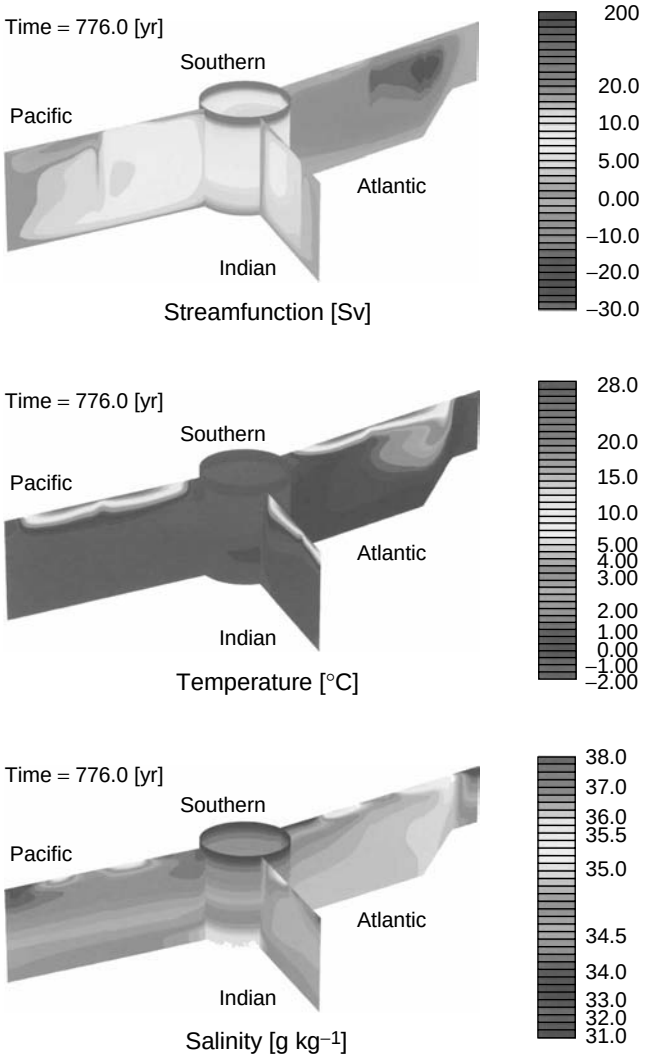


Figure 28 Flywheel geometry of a simple model of the THC of the global ocean showing examples of each of the three 2-dimensional dynamical fields (stream function, temperature and salinity) which the model explicitly predicts as a function of time. The units of stream function are Sverdrups (1 Sv = 10^6 m^3 sec^{-1})

ocean basins arranged in a flywheel pattern, and interconnected through a fourth two-dimensional basin that is of conical form and situated so as to represent the circumpolar Southern Ocean. This model of the global thermohaline circulation is coupled to the same energy balance model of the atmosphere employed as one element of the ice sheet coupled climate model described by schematic Equation (2). Because the thermohaline circulations in the individual basins of the model are described as hydrostatic, the deep-water formation process must be parameterized, as is the case in a conventional ocean general circulation model. The surface boundary conditions imposed on the individual basins consist of specified precipitation–evaporation fluxes (P–E = freshwater flux) and a restoring boundary condition on temperature that continuously nudges the surface temperature towards that calculated by the energy balance climate model. It is the mixed nature of these boundary conditions that allows the model to exhibit a mode of dynamical behavior that may, when the P–E boundary condition is appropriately selected, mimic extremely well the behavior implied by the Summit, Greenland isotopic observations.

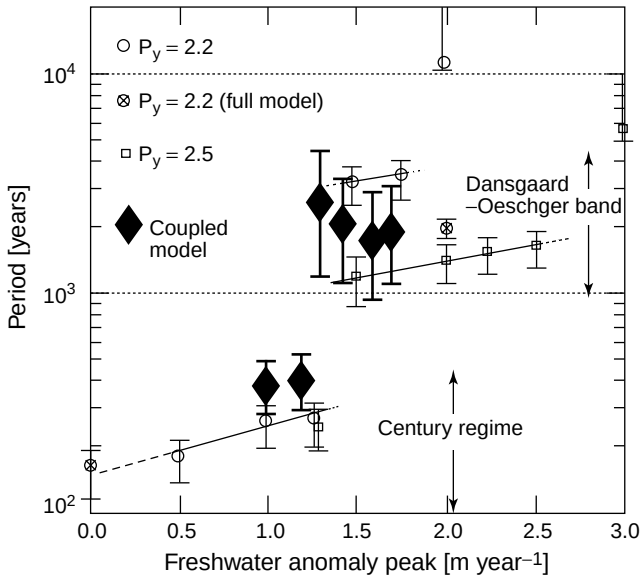


Figure 29 Bifurcation diagram for the multiple 2-dimensional basin model of the THC showing the characteristic period of the temporal variability of the THC in the Atlantic basin of the model as a function of the magnitude of the anomalous freshwater forcing that is applied as a Gaussian function of latitude to the northern part of the Atlantic basin. Note that as the peak anomaly is increased above $\sim$1.25 m year^{-1} the system undergoes sharp transition into a new state characterized by a variability time scale of 1000–3000 years (from Sakai and Peltier, 1997)

Figure 29 presents a bifurcation diagram for this model of the thermohaline circulation (THC) of the oceans. The diagram plots a measure of the dominant period of the temporal variability in the strength of the THC in the Atlantic basin as a function of the strength of the anomaly in P–E that is applied as a function of Gaussian form to the northern part of the basin in which North Atlantic Deep Water (NADW) forms under present-day climate conditions. Inspection of this diagram demonstrates that as the strength of the applied anomaly increases from zero (this corresponding to the present P–E distribution), the characteristic period is initially short and of order a century or so. However, there is a critical strength of the anomaly which, if exceeded, results in a bifurcation into a state of THC activity in which the characteristic period is of order 1–3 millennia. This result is quite precisely the period of the oscillatory behavior of the THC that would be necessary to explain the δ^{18}O oscillations which dominate the proxy climate record from the Greenland ice core during marine OIS 3.

Insight into the change of physical behavior of the simulated ocean circulation that is associated with the bifurcation is provided by Figures 30 and 31. The first of these Figures shows a series of 20 kyr simulations of the meridional heat transport in the Atlantic Basin of the model as a function of the strength of the anomaly in P–E that is applied to the northern part of the basin over the region where deep water forms. Inspection of this sequence of time series shows that, as the maximum amplitude of the P–E anomaly exceeds a value of approximately 1.15 m year^{-1}, the heat transport changes from a state of quasi-steady variation in which weak variability exists with a characteristic period of centuries, to a state of large amplitude strongly oscillatory behavior on the millennium time scale. For a modest post-bifurcation amplitude of the P–E anomaly, the system is primarily in a state characterized by strong meridional heat transport. As the maximum magnitude of the anomaly increases, however, the millennium time scale behavior is transformed such that the oscillations consist of 1–3 millennium-spaced excursions from a state of null heat transport into a state of strong heat transport (note that negative heat transport in this model corresponds to transport from the south to the north). For a sufficiently large applied perturbation, the THC dies and the northward heat transport vanishes entirely except for weak small amplitude fluctuations.

Figure 31 describes the impact of the bifurcation in terms of the output of North Atlantic surface air temperature delivered by the energy balance component of the coupled model. Comparison with the results on Figure 30, which shows the average air temperature estimates for both the summer and winter seasons, demonstrates that, when the northward heat transport is low, the surface air temperature is low. This is an entirely expected consequence of the fact that northward heat transport ceases when North Atlantic deep water (NADW) ceases to form. In the present climate; the formation of NADW during winter in the Greenland, Iceland and Norwegian seas heats the atmosphere at an annually averaged rate of approximately 100 W m^{-2} (all primarily during the winter season). This heating contributes in a telling way to moderation of the climate of northwestern Europe. It is important to note that the amplitude of the millennium time scale variations of North Atlantic air temperature simulated by this model are on the order of 6 °C. This is of the same order as those which have been inferred on the basis of the GRIP and GISP2 δ^{18}O measurements at Summit, Greenland. The hypothesis that the millennium time scale variability observed in these ice cores is caused by millennium time scale variability in the THC of the oceans would therefore appear to be well justified.

Seen from a mathematical perspective, the transition that occurs in the nature of model solutions as the perturbation to the P–E distribution increases in strength, wherein the solution changes its character from quasi-steady to intensely oscillatory, is termed a Hopf bifurcation. Such occurrences in nonlinear dynamical systems are extremely common in circumstances in which multiple equilibria or multiple quasi-equilibria exist. In the present circumstance, for weak P–E perturbation the North Atlantic THC is strong and

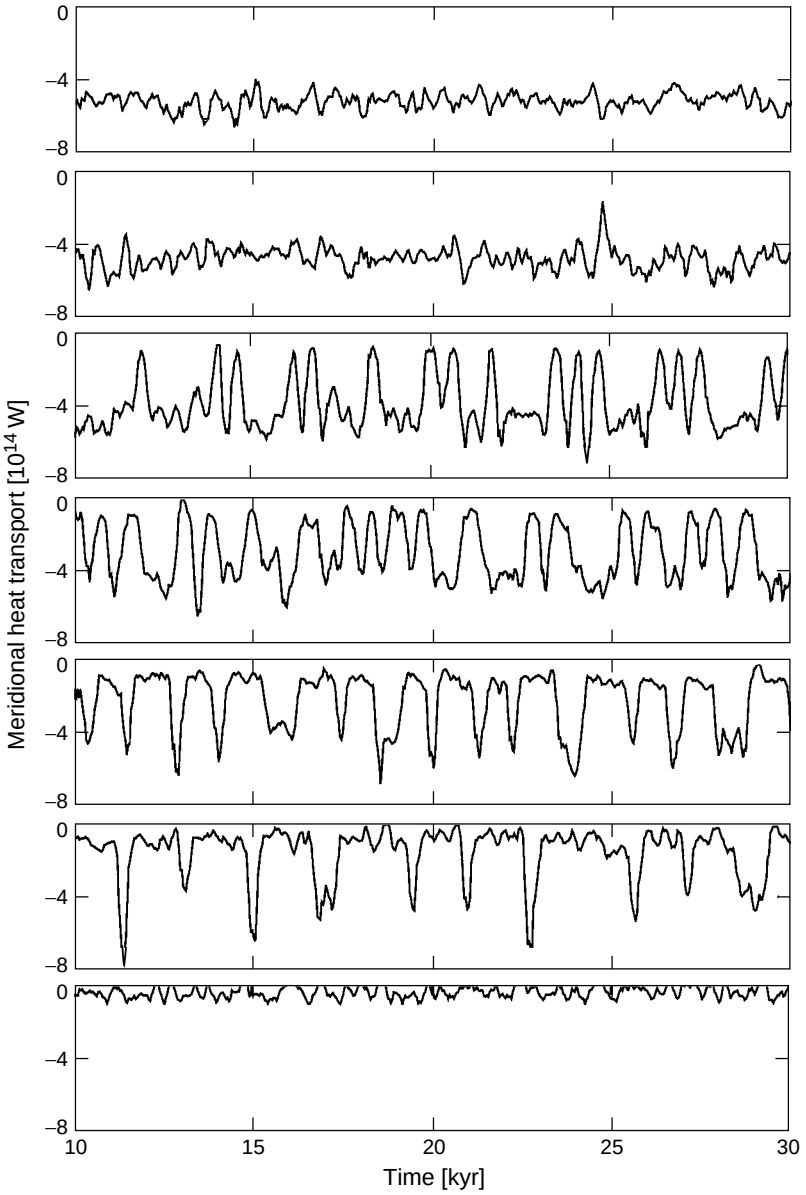


Figure 30 Northward heat flux in the Atlantic basin of the coupled model for a range of values of the peak freshwater anomaly applied to the northern part of the North Atlantic basin

essentially steady, consistent with the Holocene segment of the GRIP ice core shown on Figure 25. For a strong positive P−E perturbation NADW ceases to form and the THC stops entirely! In the intermediate range, a Hopf bifurcation occurs that induces oscillatory behavior on the millennium time scale. This oscillator may legitimately be referred to as a salt-oscillator, but it is of a kind that is quite different from that previously suggested by Professor Wallace Broecker of Columbia University. This is because the basic oscillation does not require covariation in P−E, which would presumably be connected with time dependence of the iceberg flux contribution to effective P−E. Rather it occurs under conditions of constant P−E. An explicit depiction of the covariation of sea surface salinity (SSS) during the Hopf bifurcation induced millennium time scale variability in this model is provided in Figure 32, which demonstrates that the deep circulation is shut down when the high latitude SSS is low. That full glacial THC conditions do indeed correspond to a state of strongly depressed SSS in the North Atlantic was originally demonstrated by Dr J-C Duplessy of the Gif-sur-Yvette Laboratory at Saclay, France.

Although the basic origins of the millennium time scale climate variability during OIS 3 is therefore very well explained by the existence of a Hopf bifurcation induced oscillation of THC strength, which occurs in the absence of

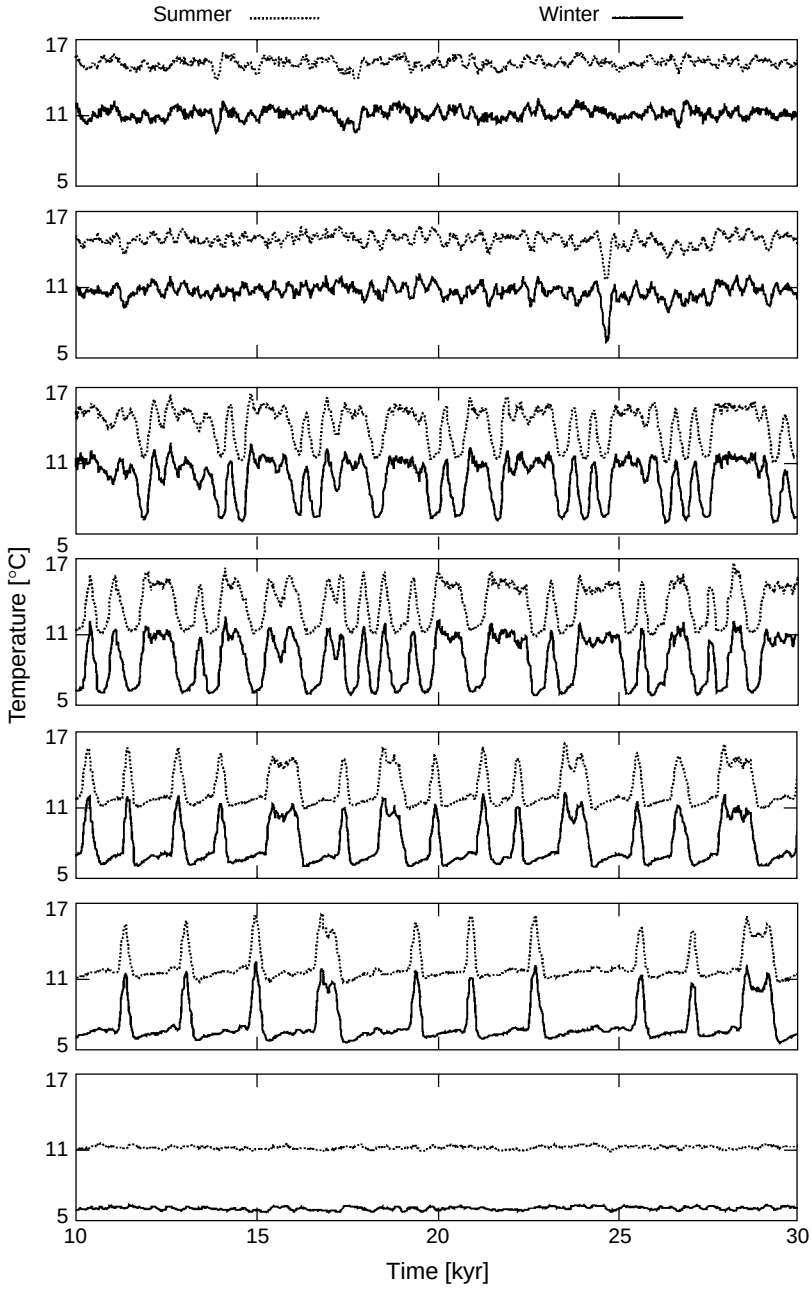


Figure 31 Same as Figure 30 but for the summer and winter air temperature time series

time dependence of the boundary condition on P−E, it is also clear that this boundary condition could not have been constant under glacial conditions. Because effective P−E includes not only the atmospherically derived precipitation and the evaporation from the sea surface, but also the freshwater added to the surface ocean by the melting of the icebergs, it is clear that effective P−E must have varied considerably due to the variability of the iceberg flux that is documented in the episodic nature of the occurrence of individual Heinrich events. It seems reasonable to suggest that the episodic Heinrich activity be viewed as a stochastic contribution to the net freshwater forcing of the surface of the North Atlantic Ocean. It remains unclear as to the extent to which the episodic nature of this activity is phase entrained to the underlying millennium time scale D−O oscillation that appears to be so well explained by the Hopf bifurcation. However, it would not be unexpected that the calving instability on the eastern flank of the Laurentide ice sheet might phase lock to some degree onto the primary metronome provided by the D−O oscillation. It is the interplay of these influences that appears to be responsible for the Bond cycle.

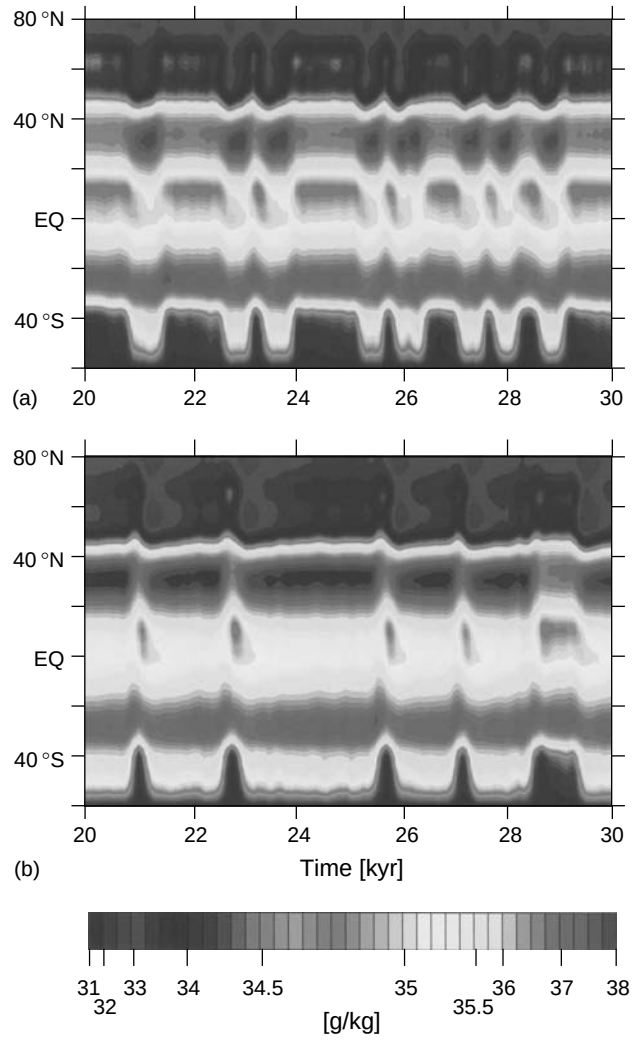


Figure 32 Latitude verses time variations in the SSS of the North Atlantic basin of the climate model for a series of D–O oscillations

SUMMARY: LESSONS IN CLIMATOLOGY FROM THE GEOLOGICAL RECORD OF EARTH SYSTEM HISTORY

It should be abundantly clear on the basis of the previous discussion of the history of Earth evolution, from the Neoproterozoic through to the Cenozoic epoch, that this most recent billion years is a period during which Earth has experienced a sequence of three episodes of significant cooling associated with the appearance of massive continent scale ice sheets. These cold phases of evolution have occupied no more than approximately 20% of this period, however, the remainder of which was characterized by climates considerably warmer than present. A primary determinant of climatic state during this period appears to have been the concentration of carbon dioxide in the atmosphere of the planet. Also contributing significantly to the long time scale evolution of surface climate has been the variation of solar luminosity and the variation of what one might refer to as polar continentality. It appears that the latter influence has perhaps been insufficiently appreciated in the past.

Even less well understood has been the apparently strong control upon the atmospheric carbon dioxide concentration that has been exerted by the so-called Wilson cycle of supercontinent creation and destruction. In particular it appears that it is during episodes of supercontinent break up that massive volcanic out-gassing leads to a significant rise of the atmospheric CO_2 concentration. Furthermore, it would appear that during the breakup phase, the mantle convective circulation may be most significantly layered by the influence of the endothermic phase transformation at 660 km depth. In contrast, continental fragments may be brought together to form a supercontinent by the occurrence of internal mantle avalanches that develop through episodic convective destabilization of the internal thermal boundary layer that forms across the 660 km discontinuity during periods of strongly layered flow. This demonstrates that, on the longest time scales, the evolution of Earth s surface climate is tightly linked to surface plate tectonics and thus to the mantle convection process.

During the Pleistocene epoch, the link between episodes of deepest glaciation and the concentration of carbon dioxide in the atmosphere is equally apparent in the strong correlation between the $\delta^{18}O$ inferred 100 kyr cycle of continental ice volume variability and the covariation of the CO_2 concentration. The same simple ice sheet coupled climate model that serves as an excellent basis for understanding the Carboniferous and Neoproterozoic glacial epochs, also explains the way in which the variations in the effective intensity of the Sun, caused by the changing geometry of Earth s orbit, is able to cause quasi periodic variations of continental ice volume at 100 kyr period. This model is successful, however, only as a consequence of the interplay between the astronomical forcing, the action of the glacial isostatic adjustment process in amplifying the deglaciation process once ice sheets become sufficiently large, and the fact that the CO_2 concentration itself varies on the 100 kyr time scale. One of the most outstanding problems in paleoclimatology is to provide a detailed understanding of the reason why the CO_2 concentration covaries with ice volume. It has even been suggested in the recent literature that the CO_2 concentration may lead ice volume at the period of 100 kyr and therefore should most properly be viewed as a prime mover of the 100 kyr ice age cycle rather than simply being an amplifier (Shackleton, 2000). What we can safely say on the basis of the analyses that have been performed with the ice sheet coupled climate model, however, is that in the absence of the additional forcing of ice volume by the variation of the CO_2 concentration the model could not explain this primary mode of late Pleistocene climate variability.

With each distinct 100 kyr cycle of late Pleistocene glaciation, during the equivalent within each cycle of OIS 3, there has also existed strong variability in surface climate on sub-orbital time scales of millennia. This higher frequency mode of climate change appears to be very tightly linked to variations in the strength of the THC. This component of the general circulation of the oceans has apparently been rather steady throughout the last 10 000 years of the Holocene epoch. However, during OIS 3 there was strong evidence that the strength of the overturning circulation in the Atlantic Basin exhibited quasi-periodic oscillations between a state of strong overturning and a state of essential quiescence. A range of different models of the THC have now been shown to exhibit such oscillatory solutions when they are subjected to sufficiently intense freshwater forcing over the region in which deepwater would otherwise form when this component of effective P−E is unperturbed from its modern distribution. The transition into the oscillatory mode of behavior in these models occurs via a classical Hopf bifurcation. When the excess freshwater forcing exceeds the critical value but is less than would be required to arrest the THC entirely, these models exhibit spontaneous oscillatory behavior even when the boundary condition on P−E is kept constant. This regime appears to provide the explanation of the millennium time scale D−O oscillation. The Heinrich events, which occur episodically during the cold phase of one of the fundamental D−O oscillations, are associated with massive iceberg calving events that appear to be triggered by instability of the eastern flank of the Laurentide ice sheet and repeat approximately every 10 kyr. Following each Heinrich event induced cooling of North Atlantic climate state, a sequence of D−O oscillations occurs, with decreasing amplitude which together form a Bond cycle of oscillations. One explanation of the Bond cycle, which fits well with the Hopf bifurcation explanation of the more fundamental D−O oscillation, is that it arises as a consequence of stochastic forcing of the D−O phenomenon due to the temporal variability of the iceberg flux.

In the course of the foregoing discussion of climate system history, I have not mentioned at all the Heinrich events labeled H0 and H1 on Figure 26. They immediately follow and precede, respectively, the Younger Dryas (Y−D) climatic cooling event, which interrupted the general trend to warmer climate that began subsequent to the last glacial maximum (LGM) as the Northern Hemisphere continents began to deglaciate. H1 appears to coincide with the B lling warm period, which was accompanied by a massive discharge of freshwater to the global ocean, a discharge which is extremely well documented by what has been called meltwater pulse 1a that is so clearly registered in the coral based and uranium−thorium (U−Th) dated relative sea level history at the island of Barbados in the Caribbean Sea that has been recovered by Richard Fairbanks of the Lamont−Doherty Earth Observatory of Columbia University. This meltwater event appears to have been derived from the Laurentide ice sheet, which was centered on Hudson Bay at LGM. The pulse could have been derivative of an outburst flood of water from the base of the ice sheet. Although this event was sufficient to cause the THC to decrease in intensity, the shutdown associated with Y−D cooling may have been caused by additional freshwater loading due to the meltback of the Fennoscandian ice complex. The Y−D event is therefore an example of a millennium time scale cooling event that is known to have been induced by the shutdown of the THC in the Atlantic Basin caused by an increase in high latitude freshwater forcing, the shutdown lasted approximately 1 kyr and the circulation appears to have recovered spontaneously after this length of time once the anomalous meltwater forcing had ceased to operate.

One clear message that is delivered by these analyses of the rapid climate changes that have occurred in the past is one that has a strong bearing on what we might expect concerning future global changes in the Earth System. Probably the most important characteristic of Earth s climate system is that it is intensely nonlinear. It is because of its intrinsic nonlinearity that the climate system may evolve in such a way as to provide surprises because of the occurrence of bifurcations through which the system may execute transitions between dramatically different modes of behavior once a particular threshold is exceeded. If the greenhouse gas warmed world into which the system appears to be evolving should cross such a threshold, say through an increase of P−E over the North Atlantic which was sufficient to cause THC collapse, we could very well be surprised to discover the imperfect degree to which our current models allow us to know what these thresholds might be. Analyses of Earth System History clearly demonstrate that such thresholds do exist and that, once they are crossed, significant changes do occur.

REFERENCES

Berner, R A (1994) GEOCARBII: A Revised Model of Atmospheric CO_2 over Phanerozoic Time, *Am. J. Sci.*, 294, 56−91.

Crowley, T J, Hyde, W T, and Peltier, W R (2001) CO_2 Levels Required for Deglaciation of a Near Snowball Earth, *Geophys. Res. Lett.*, 28, 283−286.

Dearnly, R (1965) Orogenic Fold Belts and Continental Drift, *Nature*, 206, 1083−1087.

Endal, A S and Sofia, S (1981) Rotation of Solar Type Stars: I Evolutionary Models for the Spin-down of the Sun, *Astrophys. J.*, 243, 625−640.

Hillaire-Marcel, C and Bilodeau, G (2000) Instabilities in the Labrador Sea Watermass Structure During the Last Climate Cycle, *Can. J. Earth Sci.*, 37, 195−805.

Hoffman, P F, Kaufman, A J, Halverson, G P, and Schrag, D P (1998) A Neoproterozoic Snowball Earth, *Science*, 281, 1342−1346.

Hyde, W T, Crowley, T J, Baum, S K, and Peltier, W R (2000) Neoproterozoic Snowball Earth Simulations with a Coupled Climate/Ice-sheet Model, *Nature*, 405, 425–429.

Hyde, W T, Crowley, T J, Tarasov, L, and Peltier, W R (1999) The Pangean Ice-age: Studies with a Coupled Climate–ice Sheet Model, *Clim. Dyn.*, 15, 619–629.

Johnsen, S J, Clausen, H B, Dansgaard, W, Gundestrup, N, Hammer, C U, and Tuber, H (1994) The Eem Stable Isotope Record Along the GRIP Ice Core, and its Interpretation, *Quaternary Research*, 43, 117–124.

Laskar, J (1988) Secular Evolution of the Solar System over 10 Million Years, *Astron. Astrophys.*, 198, 341–362.

Lay, T, Ahrens, T J, Olson, P, Smyth, J, and Loper, D (1990) Studies of the Earth s Deep Interior: Goals and Trends. *Phys. Today*, 63(10): 44–52.

Quinn, T R, Tremaine, S, and Duncan, M (1991) A Three Million Year Integration of the Earth s Orbit, *Astron. J.*, 101, 2287–2305.

Runnegar, B (2000) Loophole for Snowball Earth, *Nature*, 405, 403–404.

Sakai, K and Peltier, W R (1997) Dansgaard–Oeschger Oscillations in a Coupled Atmosphere–Ocean Climate Model, *J. Clim.*, 10, 949–970.

Shackleton, N J (2000) The 100 000-year Ice-age Cycle Identified and Found to lag Temperature, Carbon Dioxide and Orbital Eccentricity, *Science*, 289, 1897–1902.

Shackleton, N, Berger, A, and Peltier, W R (1990) An Alternative Astronomical Calibration of the Lower Pleistocene Time Scale based upon ODP site 677, *Trans. R. S. Edinburgh*.

Tarasov, L and Peltier, W R (1999) Impact of Thermomechanical Ice Sheet Coupling on a Model of the 100 kyr Ice Age Cycle, *J. Geophys. Res.*, 104, 9517–9545.

Thompson, S L and Pollard, D A (1995) A Global Climate Model (GENESIS) with a Land-surface Transfer Scheme (LSX). Part I: Present Climate Simulation, *J. Clim.*, 8, 732–761.

Earth Observing Systems

SUSHEL UNNINAYAR AND ROBERT A SCHIFFER

Columbia, MD, USA

In the 20th century, we saw dramatic advances in our understanding of how the Earth System functions, much of which may be attributed to substantial improvements in the technology used for Earth observing and modeling. We have also come to realize that the planet faces the potential of rapid environmental changes, including global and regional climate change, rising sea level, deforestation, deserti cation and land degradation, ozone depletion, acid rain, reduction in biodiversity, and depletion of natural resources such as fresh water and ecosystem diversity. Such changes would have a profound impact on all nations, yet important scienti c questions remain unanswered. Many processes controlling the Earth's global climate and regional environments have yet to be fully observed. The quantitative detection of subtle changes in the global system continues to be problematic, especially the attribution of cause. Earth observing systems must simultaneously address the needs for long-term observational continuity with stringent delimits on inter-annual and inter-decadal accuracy, while concurrently assimilating new methods and observing instruments to improve our understanding and modeling of detailed processes better than before. The transition of proven research observing instruments to operational systems is critical to the evolution of Earth observing. As a result of decades of development in Earth observing technology, we may now be on the brink of being able to observe nearly all aspects of the Earth System.

Neither a surface-based and/or in situ *observing system nor a space-based system has the capacity to singularly observe the Earth System. The scienti c and technological challenge is to determine the optimal mix of both, that maximizes their unique capabilities while applying innovative methods to compensate for or reduce the uncertainties and errors inherent in each. Importantly, better cross-connections and links between observing systems are needed, as inter-disciplinary approaches are developed to address the broad range of pressing environmental issues and applications related to the management of natural resources.*

KEY EARTH SYSTEM PARAMETERS AND VARIABLES

The global Earth System may be defined as comprising the atmosphere (troposphere, stratosphere and mesosphere), the hydrosphere (oceans and other important components of the hydrological cycle such as lakes, rivers, and subterranean waters), the land surface and biosphere (soils, vegetation cover, continental fauna and flora, and the fauna and flora of the oceans), the cryosphere (the ice fields of Greenland and the Antarctic, other continental glaciers, snow fields and sea ice), and the geosphere/lithosphere (continents and their topography, ocean bottom structure, the Earth s crust and molten core, tectonic motions and volcanic and earthquake processes, among others). Recent additions to this list are the chemosphere (chemical constituents and exchanges), and the technosphere (technology, anthropogenic forcing, land use changes, deforestation, resettlement programs, etc.). The latter two are increasingly becoming major forcing components of the Earth System. They are a direct consequence of the rapid escalation of human activity over the past century. One could call this the homo-sphere, and this age the Anthropocene, a term reportedly coined by Paul Crutzen (*see* Anthropocene, Volume 1).

Observing the Earth System requires the definition of a key set of parameters or variables that collectively describe its dynamic character. It is sometimes mathematically convenient to consider internal state variables and forcing or feedback variables for each component of the Earth System, as detailed in Table 1. A state variable, as the name implies, would provide monitoring information on the basic internal structure and state of a system or subsystem. A forcing or feedback variable would comprise a parameter that could change the state of a system or subsystem. Consonant with such an approach, the set of variables identified in Table 1 could be considered a minimum set to adequately describe the state and behavior of the physical Earth System when considering time scales ranging from days to years and decades. For longer time scales, processes and interactions within the solid Earth

Table 1 Summary of state and forcing/feedback variables required to observe the major components of Earth System[a]

State variables (Sv)	External forcing or feedback variables (Fv)
(1) Atmosphere	
Wind (*I*/*s*)	SST (*I*/*S*)
U/A temperature (*I*/*S*)	Land surface soil moisture/temperature (*I*/*S*)
Surface air temperature (*I*/*s*)	Land surface structure and topography (*I*/*S*)
Sea level pressure (*I*)	Land surface vegetation (*I*/*S*)
U/A water vapor (*I*/*S*)	GHGs, ozone and chemistry, aerosols (*i*/*S*)
Surface air humidity/wv (*I*/*s*)	Evaporation and evapotranspiration (*i*/*s*)
Precipitation (*I*/*S*)	Snow/ice cover (*I*/*S*)
Clouds (*i*/*S*)	SW and LW radiation budget – surface (*i*/*s*)
Liquid water content (*i*/*S*)	Solar Irradiance and SW/LW radiation budget (*S*)
(2) Ocean	
Upper ocean currents (*I*/*s*)	Ocean surface wind & wind stress (*i*/*S*)
Surface ocean temperature (*I*/*S*)	Incoming surface shortwave radiation (*i*/*s*)
Sea level/surface topography (*I*/*S*)	Downwelling longwave radiation (*i*/*s*)
Upper ocean surface salinity (*I*/*s*)	Surface air temperature/humidity (*I*/*s*)
Sea ice (*I*/*S*)	Precipitation (fresh water/salinity flux) (*i*/*s*)
Mid and deep ocean currents (*i*)	Fresh water flux from rivers & ice melt (*i*/*s*)
Sub-surface thermal structure (*I*)	Evaporation (*i*/*s*)
Sub-surface salinity structure (*I*)	Geothermal heat flux–ocean bottom (*i*)
Ocean biomass/phytoplankton (*i*/*S*)	Organic & inorganic effluents (into ocean) (*i*/*s*)
(3) Land and Water (Non-ocean)	
Topography/elevation (*I*/*S*)	Incoming shortwave radiation (*I*/*s*)
Land cover (*I*/*S*)	Net downwelling longwave radiation (*i*/*s*)
Soil moisture/wetness (*I*/*s*)	Surface winds (*I*)
Soil structure/type (*I*/*s*)	Surface air temperature and humidity (*I*/*s*)
Vegetation/biomass vigor (*I*/*S*)	Evaporation and evapotranspiration (*i*/*s*)
Water runoff (*I*/*s*)	Precipitation (*I*/*S*)
Surface ground temperature (*I*/*S*)	Land use and land use practices (*I*/*s*)
Snow/ice cover (*I*/*S*)	Deforestation (*i*/*s*)
Sub-surface temp and moisture (*I*/*s*)	Human impacts – land degradation (*i*/*s*)
Soil C, N, P, nutrients (*I*)	Erosion, sediment transport (*i*/*s*)
Necromass (plant litter) (*i*)	Fire occurrence (*I*/*S*)
Sub-surface biome/vigor (*i*)	Volcanic effects (on surface) (*I*/*s*)
Land use (*I*/*s*)	Biodiversity (*i*/*s*)
Ground water (and subterranean flow) (*i*)	Chemical (fertilizer/pesticide and gas exchange) (*i*)
Lakes and reservoirs (*I*/*S*)	Waste disposal and other contaminants (*i*)
Rivers and river flow (*I*/*s*)	Earthquakes, tectonic motions (*I*/*S*)
Glaciers and ice sheets (*I*/*S*)	Nutrients and soil microbial activity (*i*)
Water-turbidity, N, P, dissolved O (*I*/*s*)	Coastal zones/margins (*I*/*S*)

Abbreviations: SST, sea surface temperature; U/A, upper air; N, nitrogen; P, phosphorous; O, oxygen; C, carbon.

[a] In parenthesis, *I*(or *S*) denotes measurements that can be made by *in situ* (or observing systems for space-based) instruments and that operational or systematic (research) observing systems/networks and international programs exist. An *i*(or *s*) denotes that the *in situ* (or space-based) measurements made are restricted in space and/or time coverage, or that operational monitoring and data exchange systems do not exist, or that the measurements made are insufficient in accuracy or precision. In some cases, observing systems for the parameters flagged by an *i*(*s*) may be scheduled for significant improvements in the near future.

need to be taken into account, as well as solar, orbital and planetary gravitational interactions. It is underscored that the distinction between state and forcing variables is, to an extent, an artifact of the partitioning of the Earth System into land, atmosphere and ocean. In doing so, a state variable for one component could be, and often is, a forcing or feedback variable for another component. However, a partitioning of the Earth System becomes useful, if not necessary, to enable observing systems and models to be constructed. Table 1 contains a summary of the parameters that need to be monitored to document the state of the total Earth System. The table may be elaborated into one of substantially greater complexity if detailed processes are taken into account. Here, we apply a large- to global-scale filter to reduce the number of macro variables needed to define the Earth System from the perspective of

a particular time window of research and application. The table may be considered fairly comprehensive, but by no means exhaustive. It is presented to provide a framework for Earth observations.

In Table 1, emphasis is placed on geophysical parameters, and a selection of atmospheric chemistry and land–ocean biospheric parameters that are involved in the interaction between the atmosphere and the land and ocean. The parameters include large-scale estimates of vegetation cover and ocean biological productivity, both of which are important aspects of the global carbon cycle, as well as parameters that capture hydrological processes that comprise the global water cycle. They are needed because they represent sub-components of the Earth System that contain inherent memory as far as the time scales of atmospheric processes are concerned, and therefore extend the prediction time range for weather, climate and their impacts on the environment. Certain aspects of the Earth System, such as soil and water nutrients, and chemical discharge, are included due to their importance in assessments of anthropogenic impact on the environment and ecosystems.

CURRENT REQUIREMENTS AND OBSERVING ISSUES

The Earth observing system that is required for the 21st century needs to be an integrated system comprising substantial space-based assets such as remote sensing satellites, surface-based remote sensing instruments such as radar and lidar, and *in situ* instrument networks. Each set of observations should provide a three-dimensional snapshot of each component (or sub-component) of the total system. Repeated observations would provide information on the time evolution (and history) and change of the system. The space and time resolutions required are often determined on the basis of the dominant spatial and temporal features that need to be described quantitatively, and how fast these features change. In principle, however, it is not sufficient to monitor only the dominant scales, because a large number of interactive processes occur on rather small space and time scales – even down to 1 s and 1 m or less. At present, it is not impractical to continually observe the global system at these scales, nor can the data be effectively used. It is customary to obtain very high resolution data sets during special observing experiments of limited duration, in order to understand and model processes, while operating global observing systems (GOSs) at much lower time and space resolutions. These global observations need to remain accurate over time in order to meet monitoring, diagnostic analysis and modeling needs.

The accuracy, precision, and absolute calibration over the observation periods required for climate change and global change purposes are considerably more stringent than those for real time or near real time applications. Though there are notable exceptions, at present few operational observing systems (and observing practices) maintain accuracy and calibration over long periods of time. Most operational observing systems and networks were established for relatively short-term operational applications dominated by the annual solar cycle within which temperatures, for example, can vary between a monthly average of -20 to $+40\,^{\circ}$C, vegetation cover (green leaf area) from 0 to 100%, rainfall from 0 to 500 mm (or much more) between seasons, and so forth depending on location. (In some locations, even the diurnal variation could be very high; for example, up to $50\,^{\circ}$C between day and night in summer in Colorado, USA, or in the Sahara, in continental Africa.) That is, the intra-annual signal is relatively large, and thereby relatively easy to monitor. In contrast, the inter-annual to decadal change signals may be very small and pose significant problems as regards their monitoring and detection. The change in global average temperature, which rose $0.6 \pm 0.2\,^{\circ}$C during the 20th century, illustrates the aforementioned differences in the signal associated with seasonal change and long-term climate change. Correspondingly, observing technology and data processing methods need to be capable of addressing the large differences implied in signal to noise ratio. Evaluating the causes of change and the regional impacts of predicted global change are equally, if not more, complicated and troublesome. Observations must be able to initialize and validate the numerical models used to simulate the Earth System and project future changes.

The introduction of errors and discontinuities in the measurement of absolute values of change on a global or large-scale basis (which could also be national and regional) arise for several reasons, not all of which are electronic sensor related. For surface or near surface and *in situ* measurements, the station or observing network s resolution and coverage on a global basis are usually insufficient. The term *in situ* is commonly used in reference to direct instrumental measurements, in contrast to remotely sensed observations. Surface-based station networks are, by and large, *in situ* stations making observations of, for example, air temperature, humidity, precipitation, winds and pressure. However, surface-based radar stations (precipitation, and wind shear) are not *in situ* stations but rather remote sensing platforms. Instruments measuring soil moisture and temperature at the root zone, the subterranean water table depth, and ocean temperature and salinity at 1000 m depth are examples of *in situ* measurements that are not surface-based. Generally, major problems exist in monitoring oceanic areas (which cover approximately 70% of the Earth s surface) by surface-based or *in situ* station networks. Thus, for example, measuring precipitation globally and observing and quantifying the components of the global hydrological cycle (this includes cloud and convective processes, etc.) are almost impossible with traditional observing systems;

satellite technology will be indispensable. Problems are encountered even over land surface areas on account of changes (often undocumented) in instrument calibration, instrument relocation, and observing practices. Moreover, most land stations are located near populated or cultivated areas, leaving mountainous areas, wetlands, swamps, forests and other remote or inaccessible areas poorly sampled. While having the capacity to provide global coverage, existing space-based observing systems suffer from many problems such as: inadequate time resolution, uncertain vertical resolution, and uncertain or unknown calibration and accuracy between years. These problems can significantly complicate the monitoring and detection of inter-annual (variability) and decadal change. Complicating matters further, accurate sensor calibration is necessary although not entirely sufficient; calibration accuracy and time continuity must be ensured for the end product, namely, the physical, geophysical or chemical or biological parameter, to be accurately measured and monitored.

Notwithstanding various problems with the international implementation of observing systems and networks, significant advances have been made in deploying *in situ* observing networks. Examples include the tropical atmosphere–ocean buoy array deployed in the Pacific during the 10 year World Climate Research Program s Tropical Ocean Global Atmosphere (TOGA) experiment that focused on understanding the processes involved in the El-Niño/Southern Oscillation (ENSO). Another example is the ocean drifting profiler buoys used during the World Ocean Circulation Experiment (WOCE). Following the experience gained during TOGA and WOCE, plans are in place to deploy up to 3000 drifting/profiling buoys to cover all Earth s oceans on a continuing, operational basis. Substantial strides have also been made in establishing networks of instruments to measure the fluxes of carbon dioxide (CO_2), methane (CH_4), water vapor and other gases between the land surface and the atmosphere as a part of investigations into the carbon cycle. Improvements have also been made in instruments and land networks to measure surface ozone and ultraviolet (UV) radiation, and geophysical and biogeochemical exchange parameters in the ocean.

In recent years, space-based technology has also advanced to a point that we are able to accurately observe and globally sense nearly every aspect of the Earth System and to understand the processes that are central to determining the Earth s climate. Active and passive remote sensing systems spanning a wide range of spatial, spectral and temporal scales contribute to capturing and documenting the variability of the Earth s atmosphere, the oceans and continental land surface processes. On-board sensor calibration has taken a major step forwards with the new millennium technology satellite technology. Concurrent, multi-sensor observations of clouds, aerosols and other atmospheric constituents at unprecedented detail permit adaptive atmospheric corrections to be applied so that a desired signal such as ocean temperature or ocean color or vegetation can be obtained with greatly improved accuracy and resolution. Space-based platforms, with their unique capacity to observe the Earth on a global basis, complement surface and *in situ* measurements.

Together with advances in computing and information systems technology, modern data assimilation techniques and prediction models, a powerful combination of tools is available for Earth science research, the adaptive management of natural resources, and the mitigation of the impacts of natural hazards.

SPACE/TIME RESOLUTIONS AND ACCURACY

At any one point in time, global scale observational requirements are often described either in terms of what is possible with the technology at hand, or by the spatial and temporal resolutions of the best available mathematical/numerical or conceptual models of the Earth System. In parallel, local to regional scale requirements have evolved in response to specific (and sometimes widespread) user application sectors such as agriculture, forestry, water resources management, and flood forecasting. Thus, to an extent, the specification of requirements has been a somewhat arbitrary exercise based on a combination of local needs, regional or global scale feasibility, modeling technology, and scientific understanding. It is also a complex exercise in the current setting of inter-disciplinary research. That is, requirements for the observation of a particular parameter can be drastically different between different scientific disciplines and user applications sectors. Requirements also can change substantially with time and improved understanding of physical processes. A case in point is measurement of precipitation, temperature and wind, all commonly known variables. Prior to the 1980s, synoptic meteorological requirements called for observations every 3 h at a spatial resolution of about 500 km. Five hundred kilometers was also, coincidentally, the best resolution of operational weather forecast models at the time. The spatial resolution requirement was refined to between 50 and 100 km in the late 1990s due to higher resolution weather and climate models. This was possible because of advances in computational speeds, data storage technology and data communications bandwidth. However, the measurement of precipitation for hydrological applications dealing with water runoff, river flows, ground water recharge, and water resource management among others, requires observations at much higher resolutions, typically better than 1 km.

The specification of required accuracies can also be made in one of several different ways, depending on how the observations are used. For land cover and vegetation

mapping purposes, a spatial accuracy of 10% (by area) or even lower (15%) might suffice for broad categorization purposes, and for the study of relationships between biomes and average climate and soil conditions. However, for the study of deforestation rates, much better accuracy is needed because the rate of deforestation is a second order term requiring the calculation of inter-annual differences over time between differences over time within a year. Similarly, the accuracy usually specified for the measurement of upper atmospheric winds, based on the limits of the capabilities of rawinsondes and surface-based instruments, may have been sufficient for operational weather forecasting in the 1980s and hazard warnings, but could produce unacceptable errors in second order calculations for divergence fields, momentum fluxes and surface ocean wind stress.

As alluded to before, time scales also enter into the determination of required measurement accuracies. The measuring system should be able to resolve the rate of change of the parameter in question. Moreover, the accuracy of an instrument measuring temperature at a single point would not necessarily be applicable to an accuracy with which global average temperature is measured to detect change over 100 years. That is the accuracy with which the same single point represents an area surrounding the point (often as large as 250 000 km^2) must also be known or estimated before an accuracy measure for a network can be determined. As an example, global temperature change is estimated to be approximately of the order of 0.55 °C over the last 100 years. This would suggest that the combined global network of land stations, ships, ocean buoys and satellites (when applicable), when averaged together, should be accurate to of order 0.005 °C so that the magnitude of the global change signal can be identified over a relatively limited observational period. That is, it would be desirable if the combined uncertainty of all measurements taken by various platforms and over the globe over one year did not exceed the magnitude of the expected signal. Certain types of random errors can be reduced by data averaging over time and space. However, systematic errors, instrument drift, changes in observing practices and instruments, as well as changes in the environment of the instrument cannot easily be removed without an exact knowledge of the performance of the instrument, its environment, and its representativeness over time. From this example, it should be clear that the specification of accuracies, and even to some extent, space/time resolutions, is not a simple undertaking.

In the case of global climate change, the measurement accuracy of the composite GOS can only be vaguely known in absolute terms. Credibility is given to the global change signal by the fact that all sub-sets of data from the existing (composite) observing system, whether it draws from land stations alone, ocean stations alone, or ships, etc., seem to show similar trends over the period of observational record. Some controversy and debate continues over explaining the differences between satellite observations over the past 20 years and observations of surface stations, with the satellite systems showing differences in magnitude and, sometimes, even sign (*see* The Global Temperature Record, Volume 1; Tropospheric Temperature, Volume 1).

Due to such a disparate range in the types of observing requirements that need to be specified and the various theoretical and empirical means by which requirements can be determined, this overview merely presents specifications for Earth observing as a consolidation of a somewhat broad range of applications-dependent space and time resolutions, and approximate accuracy requirements. They are presented to give the reader a sense of the order of magnitude requirements for the measurement of Earth System parameters. It should be noted that the specification of individual sensor accuracy is a simpler matter, and depends on the electronic characteristics of the instrument alone, an accuracy that can be tested in a laboratory. In contrast, the accuracy of an Earth observing network is much more difficult to verify and validate, especially on a global scale. Various sampling methods and observing system simulation experiments are used to develop observational requirements. None of them is necessarily universally applicable. That is, their results would, *de facto*, depend on the particular application for which the observation is used.

Notwithstanding the above commentary, the requirements as specified by various disciplines would provide observational data that capture the dominant processes and the structure of the most important phenomena within a component of the Earth System. Summaries of observational requirements in terms of space and time resolutions and accuracies for the variables and parameters identified in Table 1 are presented in Tables 2 and 3 for state and forcing or feedback variables, respectively.

We want to reiterate that, by and large, these requirements are not met by existing observing systems, particularly on a global scale. Surface-based and *in situ* observing networks generally suffer from a lack of global coverage on account of the world s oceans. The accuracies possible with *in situ* instruments and those of remote sensing instruments can also be somewhat different. In instances where a parameter is observed using more than one observing technology, the two instrument types may see slightly different properties. For example, satellite measurements of SST actually measure sea surface skin temperature, whereas the majority of *in situ* SST measurements are at 1–10 m in depth, depending on the intake manifold depth of ocean vessels making the measurement. Operational satellite platforms typically suffer from insufficient time resolution or coarse resolution observations or both. The newest systems

Table 2 Summary of approximate, though representative, requirements for Earth System state variables and parameters[a]

State variables (Sv)	Horizontal Resolution	Vertical Resolution	Time Resolution	Accuracy/Units
(1) Atmosphere				
Wind	50–100 km	0.1–0.5 km	3 h	2–5 m/s
U/A temperature	50–100 km	0.1–0.5 km	3 h	0.5–1.0 K
Surface air temperature	50–100 km	N/A	3 h	0.5–1.0 K
Sea level pressure	50–100 km	N/A	3 h	0.5 hPa
U/A relative humidity/wv	50–100 km	0.1–0.5 km	3 h	5%
Surface relative humidity/wv	50–100 km	N/A	3 h	1–5%
Precipitation (liquid/solid)	50–100 km	N/A	3 h	0.1 mm/5%
Clouds	50–100 km	By type	3 h	10% cover
Liquid water content	50–100 km	Column total	3 h	5% ($kg\ m^{-2}$)
(2) Ocean				
Upper ocean currents	50–500 km	N/A	1 week–1 mo	$cm\ s^{-1}$
Upper ocean temperature	50–500 km	20 m	15 d–1 mo	0.3–0.5 K
Sea level/surface topography	100–1000 km	N/A	1 mo–1 year	mm–cm
Upper ocean salinity	50–500 km	20 m	1 week–1 mo	Parts per trillion (ppt)
Sea ice	25–100 km	10 cm	1 d–1 mo	2%–10%
Mid/deep ocean currents	100–1000 km	0.1–1 km	3 mo–10 years	$cm\ s^{-1}$ (tbd)
Mid/deep ocean thermal structure	100–1000 km	0.1–1 km	3 mo–10 years	0.5 K (tbd)
Mid/deep ocean salinity structure	100–1000 km	0.1–1 km	3 mo–10 years	ppt (tbd)
Ocean biomass/ phytoplankton	100–500 km	0.01–1 km	3 mo–10 years	10% (tbd)
(3) Land and Water (Non-ocean)				
Topography/elevation	1–1000 km	1 cm–1 m	1–10 years	1 cm–1 m
Land cover	1 m–100 km	N/A	1–5 years	5%
Surface soil moisture/ wetness	1–100 km	10 cm deep	1 d–3 mo	$0.02\ m^3\ m^{-3}$
Soil structure/type	1–10 000 ha	30 cm, 1 m	1–10 years	5%
Vegetation (biomass above ground)	1–100 km	1 m (tbd)	1 wk–10 years	5–10% ($kg\ ha^{-1}$)
Water runoff	1–100 km	N/A	1 h–1 mo	5%
Surface ground temperature	1–100 km	10 cm deep	3 h–1 wk	0.5–1 K
Snow/ice cover/depth	1–100 km	Depth	1 d–1 mo	10%
Sub-surface temp and moisture	1–100 km	20 cm, 1 m	1 d–3 mo	0.5 K/5% (cm)
Soil C, N, P, nutrients	1–10 000 ha	30 cm, 1 m	1–10 years	5% ($g\ m^{-2}$)
Necromass	1–100 km	N/A	1 year	10% ($g\ m^{-2}$)
Sub-surface biomass	1–100 km	N/A	1 year	10% ($g\ m^{-2}$)
Land use	1 m–100 km	N/A	1–5 years	%–m^2 (tbd)
Ground water (and subterranean flow)	100 km (tbd)	Depth	3 mo–1 year	5% ($cm\ s^{-1}$)
Lakes and reservoirs	1–100 km	Actual depth	1 week–1 mo	5% (s^{-1})
Rivers and river flow/discharge	1–100 km	N/A	1 h–1 mo	5% ($m^3\ s^{-}$)
Glaciers and ice sheets	1–100 km	Depth	1 year–10 years	$10\ kgm\ m^{-2}\ year^{-1}$
Water-turbidity, N, P, dissolved O	1–100 km	1 m	3 mo–1 year	%, ppt/ppm

Abbreviations: ppm, parts per million; WA, upper air; WV, water vapor; N, nitrogen; P, phosphorous; O, oxygen; C, carbon; d, days; mo, month.

[a] Exact accuracy and resolution vary with application and the inherent time/space scales of the involved Earth processes. Lower resolutions typically reflect large to global scale specifications often defined by operational models of the Earth System component(s). The high-resolution end of the range specified usually represents the space or time scale required for local-to-regional applications. Research projects usually require even higher space/time resolutions and better accuracy.

Table 3 Description: as in Table 2, but for forcing and feedback variables

External forcing or feedback variables (Fv)	Horizontal Resolution	Vertical Resolution	Time Resolution	Accuracy/Units
(1) Atmosphere				
SST	50–500 km	N/A	15 d–1 mo	0.3–0.5 K
Land surface-ground temperature	50–100 km	N/A	3 h–1 d	0.5–1 K
Land surface structure and topography	50–100 km	1 cm–1 m	1 d–1 year	1 cm–1 m
Land surface vegetation	50–100 km	1 m (tbd)	1 wk–1 mo	5–10%
GHGs, ozone and chemistry, aerosols	50–100 km	Variable	3 h–1 year	10% (Various)
Soil moisture/temperature at depth	50–100 km	10 cm	3 h–1 d	5%/0.5 K
Snow/ice cover/type	50–100 km	N/A	1 d–1 mo	10% (by area)
SW and LW radiation budget-surface	50–100 km	N/A	3 h–1 wk	2–10 $W m^{-2}$
Solar irradiance, SW/LW radiation budget	50–100 km	N/A	3 h–1 wk	2–10 $W m^{-2}$
(2) Ocean				
Ocean surface wind and wind stress	100–500 km	N/A	1 d–1 mo	2 $m s^{-1}$, 0.01 $N m^{-2}$
Incoming/downwelling SW/LW radiation	100–500 km	N/A	1 d–1 mo	5–10% ($W m^{-2}$)
Downwelling longwave radiation	100–500 km	N/A	1 d–1 mo	5–10% ($W m^{-2}$)
Surface air temperature/humidity	100–500 km	N/A	1 d–1 mo	0.5 K/5%
Precipitation/evaporation (salinity flux)	100–500 km	N/A	1 d–1 mo	5% (0.1 mm)
Fresh water flux from rivers and ice melt	100–500 km	N/A	1 d–1 mo	10%
Ocean boundaries and bottom topography	100–500 km	Tbd	1–10 years	10 m
Geothermal heat flux – ocean bottom	1000 km (tbd)	N/A	10 years	tbd ($W m^{-2}$)
Organic and inorganic effluents (into ocean)	Variable	N/A	1 year	Various
(3) Land and Water (Non-ocean)				
Incoming shortwave radiation	50–100 km	N/A	1 h–1 d	1–5% ($W m^{-2}$)
Net downwelling longwave radiation	50–100 km	N/A	1 h–1 d	1–5% ($W m^{-2}$)
Surface winds	50–100 km	N/A	1 h–1 d	2 $m s^{-1}$
Surface air temperature and humidity	50–100 km	N/A	1 h–1 d	0.5 K/1–5%
Evaporation and evapotranspiration	50–100 km	N/A	1 h–1 d	0.1 mm/5%
Precipitation (liquid/solid)	50–100 km	N/A	1 h–1 d	0.1 mm/5%
Land use and land use practices	50–100 km	N/A	1–5 years	Various
Deforestation	50–100 km	N/A	1 m–1 year	5%
Human impacts – land degradation	50–100 km	N/A	1 year–10 year	5%
Erosion, sediment transport	50–100 km	N/A	1 year–5 years	5–10% (tbd)
Fire occurrence	50–100 km	N/A	1 d–1 year	Various
Volcanic effects (on surface)	1–100 km	N/A	1 d–1 year	Various
Biodiversity	1–100 km	N/A	1–10 years	Various
Chemical (fertilizer, pesticide) infusion	Variable	N/A	1 mo–1 year	Various
Waste disposal and other contaminants	Variable	N/A	1 mo–1 year	Various
Earthquakes, tectonic motions	1–100 km	1 cm	1 d–10 year	Various
Soil microbial activity	1–100 km	N/A	1 week–1 year	Various
Coastal zones/margins	1–100 km	N/A	1 week–1 year	Various

Abbreviations: SW short wave; Lw long wave; d, days; mo, months.

(both *in situ* and satellite) attempt to overcome some of these problems.

IN SITU OBSERVING SYSTEMS, NETWORKS AND PROGRAMS

Observing systems and networks are usually developed and deployed to meet certain objectives pertaining to either the need to characterize a particular feature of the Earth System or to predict a future state of a component (or components) of the combined, integrated Earth System. Thus, the prediction time scale or time window will determine the number of parameters that must be observed. For time scales of hours to a few days, the primary feature that is changeable is weather, for which one needs to observe only the atmosphere in some detail. All other parameters may be specified as forcing functions. As the prediction time scale is increased to months, seasons, years and decades, many of the parameters specified as unchanging forcing functions must also be monitored, and their dynamics incorporated into Earth System models. Moreover, human dimensions and environmental concerns demand that a host of observations be made supporting a range of applications sectors. Over the years, a large number of programs have been implemented both within countries and internationally through the coordinated, cooperative efforts of all countries by mutual agreement.

The basic state variables required to characterize the structure and variability of the atmosphere, identified in Table 1, include vertical profiles of temperature, water vapor or humidity, and wind. Weather and climate parameters, such as surface air temperature and humidity, clouds and precipitation, surface pressure, and solar radiation, are also needed. For time scales beyond a few days, seasonal to inter-annual climate variability and prediction applications, observations are also required of the sub-surface layers of the land and ocean boundaries. For longer-term diagnostics and prediction, more complex observations are necessary of the ocean circulation and the land surface (including vegetation) and ground hydrology and the mid and deep ocean circulation. Current modeling technology involves the time integration of land–atmosphere–ocean coupled models. These models need observational data for initialization, validation, and for improvements to the parameterization of processes. Climate and global change issues demand observations of greenhouse gases, ozone and ozone precursor chemical species together with better characterizations of land use and land cover. The major *in situ* global and/or international observing systems and programs that monitor these variables are described in the following sections.

World Weather Watch (WWW): the earliest weather-related instrumental observations may be traced back to the mid-17th century, with the invention of the barometer by Torricelli in 1644. Precipitation was, of course, measured for several centuries before that. The first attempt (1780) to establish a worldwide network of weather stations, reporting three observations per day, is attributed to Karl Theodore, Prince Elector of the Palentine (Germany), an ardent supporter of the arts and sciences. The first modern day effort to internationally coordinate meteorological observations from networks of stations in all countries was made through the creation of the World Meteorological Organization (WMO) in 1950 following the World Meteorological Convention, adopted at the Twelfth Conference of Directors of the International Meteorological Organization (IMO) in 1947. The WMO commenced operations as the successor to IMO in 1951 and, later that year, was established as a specialized agency of the United Nations (UN) by agreement between the UN and the WMO. The WWW, coordinated by the WMO, is arguably one of the oldest and best established global systems for the observation, collection, and international exchange of meteorological and climate data from surface-based, *in situ*, and satellite-based observing systems. The Fourth Congress of the WMO formally initiated the WWW in June 1963 to collect and analyze weather and environmental information on a worldwide scale through a fully integrated system. The WWW comprises observing systems on land, sea, in the air and in outer space, and advanced data communications and data processing systems operated by the member states of the WMO (*see* Radiosondes, Volume 1). Three primary systems constitute the WWW: The Global Observing System (GOS); the Global Data Processing System (GDPS); and the Global Telecommunications System (GTS). The GOS includes about 9500 surface synoptic land stations, 700 upper-air land-based stations, 7500 mobile ship stations, 3000 aircraft stations, 250 ocean buoys, 200 background air pollution monitoring stations, and several polar orbiting and geostationary satellites providing meteorological and other environmental observations. The GDPS and GTS components operate through a system of three World Meteorological Centers, 26 Regional Specialized Meteorological Centers, 30 Regional Telecommunications Hubs, and about 190 National Meteorological Centers, linked together by a data transmission network.

WWW- U/A-Synoptic: *in situ* atmospheric upper air (U/A) temperature, moisture, wind, and pressure (replacing height). Currently, about 700 U/A synoptic sounding stations profile the atmosphere on an operational global basis coordinated through cooperative agreements between countries under the WWW of the WMO. This network has formed the backbone of WMO s operational system. To meet current requirements, WMO s Commission for Basic Systems calls for eight soundings per day at a horizontal resolution of approximately 100 km. Typically, however, observations are made once or twice per day. In recent years, the number of stations and/or the number

of soundings per day have decreased even in developed countries. Conventional radiosondes and rawinsondes are also beginning to be replaced in some regions by global positioning system (GPS)-Sondes that are more expensive. The WWW also includes operational geostationary and polar orbiting satellites. The geostationary satellites (approximately six are required for global coverage) carry an imager (typically one visible and four infrared (IR) channels) and a sounder with more spectral bands than the imager to desire atmospheric profiles of temperature and humidity.

WWW- Surface-Synoptic: surface land–ocean *in situ* air temperature, precipitation, moisture, winds, pressure, clouds, solar radiation. Approximately 9500 land-based surface synoptic stations spread around the globe make four to eight observations per day, under the auspices of the WWW of WMO. Over the ocean surface, a sub-set of the above observations is made by approximately 7500 ship and 250 buoys. Ship (of opportunity) observations are made mostly under the auspices of the WMO/WWW Voluntary Ocean Observing System. The 250 buoys are implemented jointly by WMO and the International Oceanographic Commission (IOC) as a part of the Global Ocean Observing System (GOOS) program. Several land stations also measure solar radiation (sunshine duration) and some (relatively few) measure spectral solar radiation. Some stations also measure pan evaporation. Ship based precipitation measurements are prone to error and not considered reliable.

WMO- Global Climate Observing System (GCOS): climate variables. In response to concerns about climate and global change, several internationally coordinated observing systems were established in the 1990s. Under the GCOS, a sub-set of the WMO/GOS has been designated as reference climate station network. In 1995, a network of about 148 U/A reference stations was established. Soon after, in 1996, a network of approximately 900 surface reference climate stations was also established to measure temperature, precipitation, moisture, pressure, wind, clouds and solar radiation. The surface and U/A climate reference station networks represent a first step towards establishing an integrated GCOS. The GCOS program plans to expand this observation program to include key measurements that capture the climate of the oceans and the land surface and vegetation as well. GCOS also draws data from space-based platforms such as meteorological, geostationary and polar orbiting satellites. However, baseline reference platforms have not been established yet for space-based measurements. The objectives of GCOS are to establish a global climate observing strategy based on existing operational and systematic research observing systems that would provide climate-quality data with long-term continuity for climate research and applications. The term climate-quality refers primarily to the need for tighter accuracy, calibration and time-continuity standards than those typically specified for shorter-term operational applications (*see* GCOS (Global Climate Observing System), Volume 1).

IOC/WMO GOOS: upper ocean surface and sub-surface temperature, currents, salinity. The GOOS, declared an operational system in 1999, is based on a long history of long-term monitoring of ocean conditions, coordinated by the IOC of the United Nations Education, Science and Cultural Organization (UNESCO) jointly with WMO, in order to facilitate the provision of detailed now-casts and forecasts of ocean conditions for the benefit of coastal states and national and international marine users. About 7500 ships are a part of the WMO/WWW Voluntary Observing Ship (VOS) program, measuring surface meteorological parameters as well as surface ocean parameters such as temperature. Designated ships routinely deploy expendable bathy-thermographs (XBTs, which are instrument packages that drop downward through the ocean reporting data back) to obtain ocean temperature profile data. Some also deploy expendable conductivity–temperature depth probes to obtain temperature and salinity profile data. The operational ocean observing system includes networks of moored and drifting buoys under WMO/IOC GOOS and WMO/WWW. About 300 drogued (to 15 m depth) drifting buoys measure ocean currents in the Pacific. Other aspects of GOOS implemented through national and regional efforts in pilot projects are: (1) the Euro GOOS Association comprising 22 operational agencies from 14 countries to implement six pilot projects in the Baltic Sea, Arctic Ocean, Mediterranean Sea, Black Sea, North West Shelf and the Atlantic Ocean; (2) the tropical Pacific – the TOGA program and the TOGA automated observations array of moorings for ENSO predictions, led by the US; (3) the PIRATA array in the tropical Atlantic, led by Brazil; and (4) the five coastal GOOS projects being developed in the US. Under the auspices of GOOS and the Integrated Global Observing Strategy (IGOS), the implementation of the first operational array of upper ocean sensors, Array for Real-time Geostrophic Oceanography (ARGO), is underway. The goal of the international ARGO program is to operate 3000 free-drifting ocean profiling floats, to cover the global oceans for the first time, by year 2004. The floats are being contributed and deployed by Australia, Canada, France, Germany, Japan, South Korea, the UK and the US, all of which have already begun deployment or have plans to do so. Other countries are expected to join in the deployment of the floats. ARGO measures temperature and salinity profiles and drift (i.e., ocean currents) in all of the Earth s oceans. The float drifts at about 2 km in depth and cycles up to the surface approximately every 10 days to transmit the data via satellite links. ARGO represents the oceanic counterpart to the instrumented weather balloons used to make measurements of the atmosphere. A spacing of around 3° or approximately

350 km at the equator is planned. ARGO originated as a research instrument, the Autonomous Lagrangian Circulation Explorer, deployed at around 1000 m depth, during several oceanographic experiments, including the WOCE. Other instruments, such as acoustic current doppler profilers (ADCP) to measure ocean currents, are also used operationally and in research mode, although on a somewhat limited, regional basis. For example, TOGA and WOCE obtained ADCP data during hydrographic surveys on research oceanographic ships and a few selected VOS lines (*see* Global Ocean Observing System (GOOS), Volume 2).

WMO- GAW (Global Atmosphere Watch): ozone, CO_2, CH_4, chlorofluorocarbons (CFCs), precipitation chemistry and environmental pollution. Environmental pollution concerns led to the establishment, in 1989, of the GAW by the WMO. The WMO/GAW is seen as an umbrella program for the Background Air Pollution Monitoring Network (BaPMON) of the UN Environment Programme (UNEP) that was implemented by the WMO in 1968, and the Global Ozone Observing System (GO_3OS), established in 1957 by the WMO. The BaPMON network consists of about 200 stations worldwide. The measurements taken and the number of stations include (approximately): 160 that measure precipitation chemistry; 95 that measure atmospheric turbidity; 80 that measure suspended particulate matter; 38 that measure CO_2 concentrations; 26 that measure surface ozone; 9 that measure CH_4 concentrations; and about 5 that measure CFC concentrations. The GO_3OS has a worldwide network of approximately 150 monitoring stations. These ground-based stations have been complemented by remote sensing instruments on both operational and research satellite systems. The number of measured parameters varies, sometimes substantially, from station to station, but the core program calls for observation of a fairly broad range of parameters such as: CO_2; CFCs and their substitute compounds identified in the Montreal Protocol; CH_4; nitrous oxide; ozone (surface, total column, vertical profile); radiation (including UV-B); atmospheric turbidity; total aerosol load; water vapor; chemical composition of rainfall and snow; reactive gas species (sulfur dioxide (SO_2), nitrogen oxide, etc.); particulate concentration and chemical composition; and the concentration of some radionuclides. Data are stored at a number of WMO centers operated by member states on behalf of the organization (*see* GAW (Global Atmosphere Watch), Volume 1).

IOC- Global Ocean Sea Level Observing System (GLOSS): ocean sea level (*in situ*). GLOSS is an international network of sea level measuring stations, coordinated by the IOC, to monitor sea level, a parameter considered to be a good indicator of upper ocean heat content. GLOSS comprises about 300 high quality tide gauges/stations spaced approximately 1000 km apart along coastlines and ocean islands. GLOSS operates a set of tide gauges as a part of its monitoring activities. A sub-set of these stations (approximately 30–40) is also connected as a network called the global geodetic reference system established by the International Earth Rotation Service. The development of new geodetic techniques based on very long baseline interferometry, the GPS, and absolute gravity measurements has created the opportunity to link these tidal gauges to a highly accurate global reference system. GLOSS is coordinated by the IOC through the permanent service for mean sea level (PSMSL), Bidston Observatory, UK. The PSMSL also archives historical data from about 1400 tide gauge stations worldwide, some with data records beginning about 1880.

GTOS (Global Terrestrial Observing System): terrestrial land/water ecosystems, biodiversity, pollution and toxicity. The GTOS was established in 1996 by five international organizations: the Food and Agricultural Organization (FAO) of the UN, the International Commission on Science (ICSU), UNESCO, UNEP and the WMO. The GTOS secretariat, located at FAO headquarters in Rome, Italy, provides for the day-to-day operations of the program, including liaison with GCOS and GOOS. Unlike the GOSs that exist for climate and for the oceans, there is no single organization that can provide comprehensive information (or the means for gaining access to it) on land and water resources, biodiversity and pollution impacts. The central mission of GTOS is to address this problem by linking existing networks and terrestrial observing systems to provide policy makers, resource managers and researchers with access to the data needed to detect, quantify, locate, understand and warn of changes (especially reductions) in the capacity of terrestrial ecosystems to support sustainable development. This is achieved by focusing on five issues of global concern: changes in land quality; availability of freshwater resources; loss of biodiversity; pollution and toxicity; and climate change.

ICSU- World Glacier Monitoring Service (WGMS): glaciers and ice sheets. Over 750 glaciers in 21 countries are monitored by the WGMS, which is based at the Swiss Federal Institute of Technology (ETH) in Zurich, Switzerland. The WGMS, a cooperative effort between ICSU, UNEP, UNESCO, WMO and ETH, was established in 1986, combining the ICSU/Permanent Service on Fluctuations of Glaciers, established in 1967, and the ICSU/World Glacier Inventory, established in 1976.

Global Environmental Monitoring System (GEMS): air quality, water quality, food. Under UNEP s Earth-watch program, GEMS was established in 1975, involving over 30 international monitoring projects implemented by FAO, UNEP, UNESCO, World Health Organization (WHO) and WMO. GEMS concentrates primarily on five areas: climate; trans-boundary pollution; terrestrial renewable resources; oceans; and the health consequences

of pollution. GEMS/Air (urban air quality monitoring project, WHO/UNEP) was initiated in 1973 by the WHO; the program became a part of GEMS/UNEP in 1975. The objective of the program is to monitor SO_2 and suspended particulate matter in urban areas; the network includes approximately 170 monitoring stations in 80 urban sites in over 47 countries. Most cities have established three monitoring sites: one in an industrial zone, one in a commercial zone, and one in a residential area. Data are managed by the US Environmental Protection Agency s (EPA s) laboratory at Research Triangle Park, North Carolina. GEMS/Water (Assessment of fresh water quality, UNEP/WHO/Canada) was initiated in 1976 as a freshwater monitoring network, operated by WHO and UNEP with support from UNESCO and WMO. The US EPA provides quality control support to the program. Eight laboratories, covering over 40 countries, participate in the AQC program. Assessments include: toxic chemicals, nutrients and pollutants discharged from major river basins to the world s oceans and inland seas. A computerized database is maintained at the WHO Collaborating Center on Surface and Ground Water Quality, at Canada s National Water Research Institute. GEMS/Water publishes global water quality data every five years, as well as biennial reports on special global water quality issues. GEMS/Food (food contamination and monitoring project), WHO/UNEP/FAO, was initiated in 1976 to provide reliable and internationally comparable information to governments, the CODEX Alimentarius Commission, other relevant institutions, and the general public, on contaminants in food. Over 40 countries participate by providing data on levels and trends of chemical contaminants in food, including polychlorinated biphenyls (PCBs), lead and cadmium, organophosphorous pesticides such as dichlorodiphenyltrichloroethane, aldrin, dieldrin, and aflotoxins. Microbial contamination of food is not covered by the project. Data are available for the period beginning in 1977 from the WHO. Collaborating institutions for quality assurance include the International Agency for Research on Cancer in Lyon, France for aflotoxins; the Ministry of Agriculture, Fisheries and Food, UK, for metals (cadmium, mercury and lead); and the National Food Administration, Sweden, for organochlorine compounds (organochlorine pesticides and PCBs). The Isototopes-In-Precipitation Network, International Atomic Energy Agency (IAEA)/WMO, was initiated in 1958 by the IAEA in collaboration with the WMO. In 1977, it was extended to fit within the framework of UNEP s GEMS. In 1988, over 80 network stations were in operation, while another 82 nationally implemented stations contributed data. Approximately 50% of the collected precipitation samples are analyzed in the IAEA laboratories in Vienna, Austria (*see* UNESCO (United Nations Educational Scienti c and Cultural Organization) and the Environment, Volume 4).

National observing systems and networks: all parameters and variables. National observing networks and programs make more measurements than those internationally exchanged. For parameters that are internationally exchanged, observational network density within a country is usually much higher than the data exchanged globally to meet international agreements. Depending on the policies adopted by individual countries, national data may or may not be available for international access. By and large, most scientific programs call for the free and open access and exchange of data for research purposes. Such a policy has been in place under the auspices of the ICSU World Data Centers, and most internationally coordinated programs under the UN agencies. The extent to which the parameters identified in Table 1 are regularly measured on a national basis varies widely between countries.

SPACE-BASED OBSERVING SYSTEMS AND PROGRAMS

All satellite systems possess certain similar characteristics: they carry instruments which sense electromagnetic radiation in spectral bands that are related, via retrieval algorithms, to some property of the Earth s atmosphere, the ocean or the land surface. The precise technology used, spectral discrimination, number of channels and view angles differ, but they possess common objectives in terms of what they wish to see from the standpoint of applications and research. Most satellites carry passive sensors that detect the natural radiation reflected or emitted by an object. Increasingly, however, new generation satellites carry active sensors such as synthetic aperture radar (SAR), microwave radar and laser scatterometers. In recent years, there have been well articulated moves towards the establishment of an IGOS that would encourage partnerships between countries to coordinate their efforts so as to comprehensively observe the global Earth System. Operational weather satellites are internationally coordinated by WWW of the WMO. Research and experimental Earth observing satellites are coordinated by the Committee on Earth Observing Satellites (CEOS), and the IGOS-partners. Interdisciplinary programs such as the International Geosphere–Biosphere Programmes and the World Climate Research Programmes among others, rely on data from both operational and research satellite systems. Several international observing system programs have also been established in recent years that endeavor to coordinate both satellite and *in situ* measurements to meet the requirements of specific research and applications objectives, notably: GCOS; GOOS; and GTOS. The lead agencies for their coordination are, respectively, the WMO in

Geneva, Switzerland, the IOC in Paris, France, and the FAO in Rome, Italy.

The parameters that are measured by space-based remote sensing satellites are identified in Tables 1, 2, and 3 by an *S*.

Operational Polar Orbiting Environmental Satellites (POES)

The first meteorological satellite observations of the Earth began with the launch by NASA (US National Aeronautics and Space Administration) of the first Television and IR Observation Satellite (TIROS-I) on the 1st of April 1960. This satellite demonstrated the feasibility of observing the Earth s cloud cover and weather by means of slow-scan television cameras in an Earth-orbiting, spin stabilized satellite.

TIROS-II was launched on the 23rd of November 1960, to demonstrate, in addition to the wide- and narrow-angle television cameras, an experimental five-channel scanning IR radiometer and a 2-channel non-scanning IR device. These instruments measured the thermal energy of both the Earth s surface and the atmosphere in order to provide data on the planet s heat balance and add a new dimension to the understanding of weather. A series of these TIROS satellites, developed by NASA and operated by the National Oceanic and Atmospheric Administration (NOAA), was launched throughout the 1960s. The NOAA/TIROS satellites continuously orbit the Earth from North to South Pole and back (hence the name polar orbiting) at an altitude of approximately 470 nautical miles (870.44 km or 540.86 statute miles).

The NOAA/TIROS series (and other similar satellites) basically carry scanning radiometers as imaging and radiometric devices and sounders to obtain information on the vertical profiles of selected parameters. Key instrument packages include: (1) the advanced very high resolution scanning radiometer (AVHRR), currently a five-channel cross-track scanning instrument, providing image and radiometric data in the visible, near-IR, and far-IR portions of the spectrum, is used to observe clouds, land-water boundaries, snow and ice, water vapor, temperature of clouds, and land and sea surface; (2) the TIROS operational vertical sounder, a three-part TIROS system to measure temperature profiles of the Earth s atmosphere from the surface to about 10 hPa, water content of the Earth s atmosphere, and total column ozone content of the atmosphere; (3) the high resolution IR sounder (HIRS), a 20-channel, step-scanned visible and IR spectrometer used to produce tropospheric temperature and moisture profiles; (4) the stratospheric sounding unit, a three-channel, pulse modulated, step-scanned, far-IR spectrometer used to produce temperature profiles of the stratosphere; and (5) the microwave sounding unit (MSU), a four-channel, step-scanned spectrometer with a response in the 60-Ghz oxygen band, used to produce temperature profiles in the atmosphere in the presence of clouds. The satellite also carries a space environment monitor to measure energetic particles emitted by the Sun. Earth radiation budget experiment (ERBE) radiometers, flown on NOAA 9, 10 and later satellites, were designed to measure all radiation striking and leaving the Earth. The present two-POES system will continue through 2002 with AVHRR-3 and HIRS-3 as core instruments. NOAA-L, the most recent in the series of NOAA/POES satellites, was launched in September 2000. The current NOAA/POES series will be operational for the next 12 years, after which it will be replaced by the National Polar Orbiting Environmental Satellite series, jointly sponsored by the US NASA, NOAA, and the Department of Defense.

Other countries with operational polar orbiting satellites include the Russian Federation and China. Since the 1970s, the Russian Federation has operated a polar orbiting series of satellites based on the meteorological (METEOR) system spacecraft for the exploration and monitoring of the Earth s natural resources. The METEOR series was started in 1985 and continues to be operational. NASA s total ozone mapping spectrometer, and the scanner for radiation budget, among other instruments, will be integrated on board the next in the METEOR series of satellites (METEOR-III), possibly in 2001. China operates a Fengyun-1 (FY-1) polar orbiting satellite series. The first of the FY-1 series was launched on the 7th of September 1988. An improved version, particularly in terms of IR images, was launched in September 1990, followed by FY-C in the late 1990s. The main instrument on the first satellites (FY-1-A, B) was the Chinese advanced very high resolution radiometer, a visible and IR scanning radiometer with five channels and a resolution of 1.1 km at the sub-satellite point. FY-1C has a 10-channel multispectral visible and IR scanning radiometer (resolution: 1.25 to 5 km) to observe clouds, the Earth s surface, middle and upper atmospheric water vapor, SST, snow/ice, soil moisture, and ocean color.

Operational Geostationary Environmental Satellites (GOES)

Geostationary satellites observe the Earth from above the equator, and maintain their position relative to a location at the surface of the Earth at an altitude of 35 790 km or 22 240 miles. The satellite travels around the Earth in the same direction as the rotation of the Earth with an orbital period equal to that of the rotation of the Earth (23 h, 56 minutes, 04.09 s). The US GOES satellites, developed by NASA and operated by the NOAA, are positioned longitudinally around 75 °W and 135 °W. A worldwide network of operational geostationary meteorological satellites provides

visible and IR images of the Earth s surface and atmosphere. The satellite systems include: the GOES satellites operated by the US; METEOSAT (launched by the European Space Agency (ESA), and operated by the European Meteorological Satellite Organisation); the Japanese geostationary meteorological satellite; and the geostationary satellites operated by India and China (the Fengyun-2 series). GOES carries three major sensor systems: (1) a multispectral imaging instrument capable of simultaneously sweeping one visible and four IR channels in a north to south swath across an east to west path, providing full disk imagery once every 30 minutes; (2) a sounder with more spectral bands than the imager, to produce high quality profiles of temperature and moisture, capable of stepping one visible and eight IR channels in a north to south swath across an east to west path; (3) the space environment monitor to measure the condition of the Earth s magnetic field, solar activity and radiation around the spacecraft. The spacecraft also carries a data collection system and a search and rescue transponder. Other satellites are comparably instrumented.

Earth Resources Mapping and Surface Characterization Satellites

Landsat (USA): Landsat is a series of satellites designed to gather data on the Earth s resources including: land use inventories; geological and mineralogical exploration; crop and forestry assessment; and cartography. The first Landsat-1 was developed and launched by NASA in 1972, following the early success of NASA s TIROS and Nimbus series, to expand the possibilities for Earth remote sensing from space in response to not only geologists and biologists but also a broad community of farmers, miners, timber and water resource managers, and engineers. The multispectral scanner (MSS) proved to be so valuable that a version of it has been flown on each of the first five Landsat missions. With the 1982 launch of Landsat-4, the thematic mapper (TM) was introduced. TM was a significant improvement over the MSS, providing greater resolution in the visible and near-IR regions (30 m versus 80 m resolution), and three additional spectral bands. The Landsat-3 version of the MSS was also flown to provide data continuity. Lansat-6 failed to reach orbit in 1993, but Landsat-5 continued to provide coverage well beyond its design lifetime. Landsat-7 was launched in 1999 to provide continuity and enhanced coverage to characterize, monitor, manage, explore, and observe the Earth s land surface. Landsat-7 s payload consists of the enhanced thematic mapper plus (ETM+). New features on Landsat-7 include a panchromatic band with 15-m spatial resolution, a thermal IR channel with 60-m resolution and both high and low gain, and on-board radiometric calibration. ETM+ still provides continuity with the TM that flew on Landsat-4 and Landsat-5.

SPOT- Image (France-CNES [Centre National d'Etudes Spatiales]): the SPOT satellite Earth observation system (EOS) was designed by the CNES, France, and developed with the participation of Sweden and Belgium. The system comprises a series of spacecraft plus ground facilities for satellite control and programming, image production and distribution. SPOT is sun-synchronous (descending node is at 10.30, local time), in near-polar orbit, with a repeat time-cycle duration: 26 days, 369 orbital revolutions per cycle, 822 km altitude, 98° inclination, and $14 + 5/26$ revolutions per day. The satellite s payload comprises two identical high resolution visible (HRV) imaging instruments that can operate simultaneously or individually in panchromatic and multispectral modes. The position of each HRV entrance mirror can be commanded by ground control to observe a region of interest not necessarily vertically beneath the satellite. Each HRV offers an oblique viewing capability. The viewing angle is adjustable through $\pm 27^\circ$ relative to the vertical. In panchromatic mode, imaging is performed in a single spectral band corresponding to the visible part of the spectrum without the blue, covering $0.51-0.73\,\mu m$. The panchromatic mode supplies black/white images at 10 m pixel resolution, primarily for applications calling for fine geometric detail. The multispectral mode images in three spectral bands: XS1: $0.5-0.59\,\mu m$ (green); XS2: $0.61-0.68\,\mu m$ (red); and XS3: $0.79-0.98\,\mu m$ (near IR). By combining data recorded in these channels, color composite images can be produced at 20 m pixel resolution. Different land surface features (e.g., vegetation, fresh snow, white limestone, dry ground, water, sand, etc.) are distinguished due to their different spectral signatures in these three bands. SPOT 1 was launched on the 22nd of February 1986, and withdrawn from active service on 31 December 1990. SPOT 2 was launched on the 22nd of January 1990 and is still operational. SPOT 3 was launched on the 26th of September 1993 but malfunctioned on November 14th, 1996. Since 1998, SPOT 4, with an increased life expectancy of five years, has been fully operational, and brings enhanced capability over its predecessors with the inclusion of a short-wave IR (SWIR) band offering better discrimination of crops and plant cover. With Spot 4, Spot 1 and Spot 2, the Spot program now comprises three fully operational satellites in orbit. Spot 5, scheduled for launch in 2001, will ensure continuity of service and further improved performance.

ERS (Earth Resources Satellite) Series (ESA): ERS-1 was launched by the European Space Agency in 1991, followed by ERS-2 in 1995. The series concentrates on global and regional environmental issues, making use of microwave techniques that enable a range of measurements to be made of land, sea and ice surfaces independent of cloud cover. The core instruments on ERS-1 comprise an active microwave instrument combining the functions of a SAR and a wind scatterometer. The SAR operates in

image mode for the acquisition of wide-swath, all weather images over the oceans, polar regions, coastal zones and land. In wave mode, the SAR produces imagettes (5 km^2) at regular intervals, for the derivation of the length and direction of ocean waves. The wind scatterometer uses three antennae for the generation of sea surface wind speeds and direction. The radar altimeter provides accurate measurements of sea surface elevation, significant wave heights, various ice parameters and an estimate of sea surface wind speed. ERS also carries an along track scanning radiometer (ATSR), which combines an IR radiometer and a microwave sounder for the measurement of SST, cloud top temperature, cloud cover and atmospheric water vapor content. ERS-1 is in a near-polar orbit at a mean altitude of 780 km with an instrument payload comprising active and passive microwave sensors, and a thermal IR radiometer (inclination: 98.52°; repeat cycles: three-day, 35-day, 176-day; orbits per day: 14.3; period: about 100 minutes). ERS-2 is a more sophisticated version of ERS-1, and includes, in addition to ERS-1 instruments, the global ozone monitoring experiment instrument for atmospheric chemistry measurements and a microwave sounder.

RADARSAT (Canada): RADARSAT-1, launched in 1995, is an advanced Earth observation satellite to monitor environmental change and support resource sustainability. The payload comprises a Synthetic Aperture Radar (SAR), a powerful microwave instrument that will transmit and receive signals to see through clouds, haze, smoke, and darkness, and obtain high quality images of the Earth in all weather at any time of the day or night. Using a single frequency, C-Band (5.3 GHz SAR), the RADARSAT has the unique ability to shape and steer its radar beam over a 500 km range. User defined beam selections can have image swaths from 100–500 km, with resolutions from 10–100 m, respectively. Incidence angles range from less than 20° to more than 50°. Repeat time is 24 days. The satellite provides complete daily coverage of the Arctic, views of any part of Canada every three days, and achieves complete coverage at equatorial latitudes every six days using a 500 km wide swath. The orbit is sun-synchronous with over-passes at the same local time. The descending equatorial crossing is at 0600, which is a dawn–dusk orbit, meaning that the satellite is rarely in eclipse and is able to acquire data at any time. RADARSAT-2, with a high resolution, multi-polarization SAR, is proposed for launch in 2001, and RADARSAT-3 in 2004.

IRS (Indian Remote Sensing Satellite): The second generation Earth observing satellites, IRS-1C/1D, were launched on the 28th of December 1996 (1C; Russian Molniya launcher) and the 29th of September 1997 (1D; Indian Polar Satellite Launch Vehicle), respectively. The imaging sensor suite on 1C/1D comprise: a multispectral linear imaging self-scanner (LISS-III) in visible, near-IR spectral bands with spatial resolution of around 23 m and a SWIR band with a resolution of around 70 m; a panchromatic camera with a resolution of around 6 m with across track stereo viewing capability; and a wide-field sensor in visible and near-IR region with a spatial resolution of 188 m and a wide swath of around 800 km. Because of the wide swath, the camera provides a repetitive coverage every five days.

Research and Experimental Satellites – Next Generation Instruments

Substantially improved satellite instruments have been under development for the past decade. They address several of the deficiencies found in the older technologies used in the current generation of operational satellites. The sensors typically have a broader spectral range, better spectral resolution, more flexibility in terms of view angles, and much better spatial resolutions. Importantly, they have much better on board calibration systems. Most of the key Earth System variables identified in Table 1 are or will soon be measured or estimated by advanced satellite observing instruments.

The sections that follow describe a representative selection of new-millennium satellite instruments, several of which have recently been launched.

UARS (Upper Atmospheric Research Satellite): atmospheric ozone depletion chemistry, trace gases, water vapor, solar variations. Launched by NASA in 1991, UARS was the first satellite dedicated to increasing understanding of the chemistry and dynamics of the Earth s stratosphere and mesosphere. UARS confirmed the role of CFCs in ozone depletion and clarified chemical processes that cause the Antarctic ozone hole. UARS made the first measurements of several key stratospheric gases: chlorine nitrate ($CLONO_2$), chlorine monoxide (CLO), hydrogen fluoride (HF), hydrogen chloride (HCL), and dinitrogen pentoxide (N_2O_5). These measurements, along with the measurements of nitric oxide (NO), nitrogen dioxide (NO_2), and nitric acid (HNO_3), provided a comprehensive picture of stratospheric photochemical processes involved in ozone depletion. UARS provided the first comprehensive mapping of volcanic aerosol layers and documented the slow increase in aerosol amounts following the eruption of Mt Pinatubo in the Philippines. The satellite also provided the first global picture of atmospheric tides, a clearer understanding of solar UV radiation variations, insight into the relationship between upper tropospheric water vapor and climate, and multiyear information on the transport of trace gases into the stratosphere. UARS was designed to make observations of the upper atmosphere for 18 months. More than seven years after launch, UARS continued to make observations with 8 of 10 instruments and continues to collect data. The UARS instruments include: chemistry: cryogenic limb array etalon spectrometer, halogen occultation experiment, improved stratospheric and mesospheric sounder,

microwave limb sounder (MLS); wind measurements: high resolution doppler imager, wind imaging interferometer; solar measurements: active cavity radiometer irradiance monitor (ACRIM-III), solar/stellar irradiance comparison experiment, solar UV spectral irradiance monitor; energetic particle measurements: particle environment monitor.

TOPEX/Poseidon (Ocean Topography Experiment): sea surface height, winds, waves. Jointly developed by NASA and CNES (France), the satellite, carrying dual and single frequency altimeters, was launched in 1992 to accurately measure the altitude of the satellite above the sea surface, sea surface wind speed and wave energy. The dual-frequency altimeter corrects for ionospheric effects. A GPS receiver is one of the three systems used to precisely track the satellite orbit. A laser retroreflector, used with ground-based lasers, measures satellite position and verifies the satellite altimeter s height measurements. A microwave radiometer measures atmospheric water vapor content and is used to correct altimeter measurements for pulse delay caused by water vapor. Sea surface height is measured to within 4.3 cm. The triple tracking system determines the satellite s location to within 2.8 cm. The distance between the ocean surface and the satellite is measured with an accuracy of 3.2 cm. The follow-on to TOPEX/Poseidon is Jason-1, scheduled for launch in 2001.

SeaWifs (Sea-viewing Wide Field-of-view Sensor): ocean color/biomass and atmospheric aerosols. SeaWifs was launched in 1994 and resumes the measurement of ocean color (phytoplankton) made from 1978 to 1986 by the coastal zone color scanner (CZCS) satellite. Aerosols directly affect the radiative transfer in the atmosphere, and therefore the radiation received by a satellite from the surface of the Earth (land and ocean). When retrieving surface or near surface signals from sensor-measured radiances at satellite altitude, the atmospheric effect must be removed. This atmospheric correction removes more than 90% of the observed radiance at visible wavelengths measured at the top of the atmosphere. The SeaWifs satellite s atmospheric correction algorithm uses radiances measured at two near-IR wavelengths, 765 nm and 865 nm, at which the ocean appears black due to strong absorption by water, to estimate the aerosol optical properties and extrapolate these into the visible. This enables SeaWifs observations to be used to derive global ocean color and ocean bio-optical properties. Aerosol optical properties, particularly optical depth or thickness, is an important byproduct of the SeaWifs atmospheric correction.

TRMM (Tropical Rainfall Measuring Mission): rainfall, clouds, SST, radiation, lightning. In November 1997, TRMM was launched as a collaborative venture between NASA (US) and the National Space Development Agency (NASDA) of Japan, with five instruments: The first spaceborne precipitation radar (PR), the TRMM microwave imager (TMI), a visible and IR scanner (VIRS), a cloud and Earth radiant energy system (CERES), and a lightning imaging sensor. The PR and the microwave radiometer measure the vertical distribution of precipitation over the tropics between $\pm 35^\circ$ in latitude, from the surface to about 20 km; horizontal resolution: 4 km; swath width: 220 km. Rain rates down to about $0.7\,\mathrm{mm\,h^{-1}}$ can be detected. TRMM will reduce the uncertainty in global tropical rainfall estimates to about 10%, an 80% improvement over the current uncertainty of about 50%. The TMI has a full suite of channels ranging from 10.7−85 GHz. and is the first satellite sensor capable of accurately measuring SST through clouds. When rain is not present, attenuation at 10.7 GHz is small and 97% of the sea surface radiation reaches the top of the atmosphere. The high frequency channels (19−37 GHz) can precisely estimate the 3% attenuation due to oxygen, water vapor and clouds. The 37 GHz channels are very sensitive to rain and are used to determine when rain is in the radiometer s field of view. At frequencies below about 12 GHz, microwaves penetrate clouds with little attenuation, providing a clear view of the sea surface under all weather conditions except rain. At these frequencies, aerosols have no effect. The first satellite microwave radiometers operating at such low frequencies were successfully demonstrated on Nimbus-7 and Seasat in 1978, but they were limited by poor calibration systems. The VIRS senses radiation in five spectral regions ranging from the visible to IR (0.63 to 12 μm). It serves as a rainfall indicator, and a transfer standard between TRMM measurements and those made routinely on the operational POES and GOES satellites (*see* TRMM (Tropical Rainfall Measuring Mission), Volume 1).

ACRIM-III: solar irradiance. The ACRIM-III mission (small satellite dedicated mission), called ACRIMsat, launched in 1999, continues the series of satellite data monitoring solar irradiance. Heritage: ACRIM-1 (on NASA s SMM) (solar maximum mission), ACRIM-II (on UARS). ACRIM-III instrument is a miniaturized version of the technology flown successfully on SMM, Spacelab-1, and Atlas missions. The satellite is spin stabilized with the instrument s solar viewing axis aligned with the Sun within 0.25° using magnetic torquers. The Sun is a variable star, and luminosity has been found to vary by about 0.1% over a solar cycle in phase with the level of solar magnetic activity. Sustained change in total solar irradiance (TSI) of as little as a few tenths of 1% per century could be a primary causal factor of climate change. A precision TSI database, with resolution adequate to relate centuries of systematic TSI variation to climate change, must be compiled from the results of many flight experiments. With a nominal lifetime of five years per experiment, their contiguous results must be related with the maximum precision accessible to current technology, on the order of 10 ppm/year (e.g., 0.1% per century). This far exceeds the capabilities of current ambient temperature flight instrumentation to define the absolute uncertainty of

the TSI (>1000 ppm) and even that of cryogenic instrumentation currently under development (>100 ppm). Modeling TSI using ground-based observations of proxy solar emission features is orders of magnitude less precise. The single approach capable of providing the required precision for the long-term TSI database with current measurement technology employs an overlap strategy, in which successive ambient temperature TSI satellite experiments (each with a nominal lifetime of five years) are compared in flight, transferring their operational precision to the data base. The current generation of ambient temperature ACRIM flight instrumentation has demonstrated a capability of providing annual precision of approximately 5 ppm of TSI. A second phase ACRIM instrument is planned for launch in late 2001 to combine the TSI database maintenance function with a new technology development initiative designed to develop a prototype instrument for both total and limited spectral solar monitoring on an operational basis in the National POESS (NPOESS) program.

EOS Satellite Series

NASA s EOS is a series of polar orbiting and mid-inclination satellites specifically developed to monitor nearly all aspects of the Earth System with an unprecedented accuracy, spectral discrimination, and spatial resolution. The first major satellite of the new millennium is EOS-Terra (AM) with a sun-synchronous polar orbit, descending southward across the equator in the morning. The second will be EOS-Aqua (PM) to be launched into a sun-synchronous polar orbit ascending northward across the equator in the afternoon, followed by EOS-CHEM. The key instrument developed to fly on Terra and Aqua is the moderate resolution imaging spectroradiometer (MODIS). MODIS is comprehensive in that it will continue to take measurements in spectral regions that have been and are currently being measured by other satellites, or heritage instruments. Therefore, MODIS will extend the time series of data sets taken by such instruments as AVHRR used for meteorology and monitoring SST, sea ice, and vegetation; and the CZCS used to monitor ocean biomass and ocean circulation patterns (*see* EOS (Earth Observing System), Volume 1).

The EOS satellite sensors have unprecedented onboard calibration systems, enabling a characterization of their performance throughout the lifetime of each satellite s mission, and the correction of errors introduced into the data due to system degradation. Calibration is a set of operations or processes that are used to determine the relationship between satellite instrument output (i.e., digital counts) and corresponding known values of a standard. Pre-flight calibration and characterization includes radiometric and geometric calibration. Radiometric calibration involves determining the relationship between instrument output and radiant input while the instrument views a calibrated radiant source. For the solar spectrum (400–2500 nm), the calibrated radiant source is an integrating sphere that has been calibrated using standards. For thermal IR (above 2500 nm), the calibrated source is a variable temperature blackbody. Pre-flight geometric calibration involves the determination of the detailed spatial response of each instrument band with respect to the nominal instrument telescope pointing direction. Geometric calibration is necessary for converting the instrument radiometric data into images with known relationships to observed Earth targets. Instruments will also be radiometrically and geometrically calibrated on orbit using solar diffusers, blackbodies, space viewports, and in the case of MODIS, a monochromatic source known as the spectroradiometric calibration assembly. EOS and other satellite instruments will also use the Moon as a common, stable, on-orbit radiance reference target source. The accuracy and precision of Earth remote sensing data sets produced by different instruments are only achieved by the consistent use of common on-orbit calibration sources and measurement methodologies. EOS-Terra will perform an on-orbit, pitch-based, calibration attitude maneuver to enable the instruments to view deep space and the Moon. The deep space view provides CERES and ASTER (advanced spaceborne thermal emission and reflection radiometer) an accurate determination of the DC offsets in their thermal bands, and it provides MODIS the opportunity to characterize the dependence of the thermal IR reflectance on the angle of incidence of the scan mirror. The same pitch-based maneuver also provides EOS instruments the opportunity to view the Moon at a lunar phase of 22°. The Moon affords the ability to monitor, among others, radiometric responsivity and stability in the visible, near-IR, and SWIR wavelengths. Validation of satellite measurements is carried out through various means, such as cross-comparing the satellite measurements with aircraft observations.

EOS- Terra: with an AM equatorial crossing, EOS-Terra was launched on December 18th 1999 by NASA. Terra houses five instruments for simultaneous, geolocated measurements and for the intercomparison of new measurement techniques. For example, MODIS s detailed internal calibrations are used, through simultaneous geolocated measurements, to help in the calibration of ASTER. The MODIS and ASTER high-resolution multi-channel observations of clouds are used by the CERES s low resolution radiative flux measurements and by the measurement of pollution in the troposphere (MOPITT) instrument to determine the location of clouds as well as their distribution and properties. The multi-angle imaging spectrometer s (MISR s) multi-angle measurements determine the angular reflectance properties of land surface features, aerosols, and clouds, all of which are used in the data analysis of MODIS, ASTER and CERES. Vegetation properties are derived from both MODIS and MISR data. Aerosol

properties are measured by MODIS using its wide spectral range and 1–2 day single view coverage, and also independently by MISR using its multi-angle data, narrow spectral range, and 2–9 day coverage. Water vapor is derived independently from MODIS measurements of reflected near IR sunlight and emitted terrestrial IR radiation. Emitted smoke particles from biomass burnings and forest fires are observed by MISR and MODIS, and emitted trace gases (carbon monoxide (CO) and CH_4) observed by MOPITT. Terra provides, for the first time, simultaneous measurements of clouds, aerosols, atmospheric trace gases, land and ocean surface properties, as well as Earth radiation budget parameters. The first CERES instrument was flown on TRMM and will also provide continuity to ERBE and pre-ERBE Earth radiation budget measurements. ASTER, developed and built by the Ministry of International Trade and Industry (MITI) of Japan, has 14 spectral bands in the visible, near-IR, SWIR and thermal IR wavelengths, and spatial resolutions ranging from about 15–90 m. MISSR (multi-angle imaging spectroradiometer), a new instrument, views the Earth in four color bands (blue: 446 nm; green: 558 nm; red: 672 nm; and near-IR: 866 nm) and from nine view angles with nine cameras pointed in different directions. MODIS views the Earth every 1–2 days, making observations in 36 co-registered spectral bands at moderate resolution (0.25–1 km). MOPITT, developed at the University of Toronto, Canada, employs gas correlation spectroscopy to measure upwelling and reflected radiation in three absorption bands of CO and CH_4. EOS-Terra instruments are individually described in more detail in the following sections to provide the reader with a sense for the latest available technology in satellite Earth observing technology. With a development time of approximately 10 years, the EOS series of NASA s satellites represents an unprecedented effort in space-based remote sensing of the Earth System as a whole. It is noteworthy that NASA s EOS has involved a large number of partnerships between NASA and the space agencies of many different countries. The enterprise represents a significant achievement in international collaboration. The selection of the EOS series for expanded description in this paper should not detract the reader from important developments in other countries that launch Earth observing satellites. They are also mentioned in this paper, even though not elaborated on.

EOS- Terra/CERES. CERES is based on NASA s ERBE, which used three satellites to provide global energy budget measurements from 1984–1990. The first CERES instrument was flown on TRMM. CERES provides continuity to ERBE and pre-ERBE measurements. CERES is a broadband scanning thermistor bolometer package with extremely high radiometric measurement precision and accuracy. Terra/CERES carries two identical instruments: one operating in a cross-track mode and the other in a biaxial scan mode. The cross-track mode essentially continues the measurements of the ERBE and TRMM missions, while the biaxial mode provides new angular flux information that improves the accuracy of angular models used to derive the Earth s radiation balance. Each CERES instrument has three channels – a shortwave channel for measuring reflected sunlight, a longwave channel for measuring Earth emitted thermal radiation in the 8–12 μm window region, and a total channel for total radiation. On board, calibration hardware includes a solar diffuser, a tungsten lamp system with a stability monitor, and a pair of blackbody sources. Cold space and internal calibration looks are performed during each normal Earth scan. Both CERES scanners operate continuously throughout the day and night portions of an orbit. In the cross-track scan mode, calibration occurs biweekly. In the biaxial mode calibrations also occur biweekly, and sun-avoidance short scans occur twice per orbit. CERES will be used to: study cloud radiative forcings and feedbacks; develop an observational baseline for clear-sky radiative fluxes; determine radiant input to atmospheric and oceanic energetics models; validate general circulation models; and enhance extended-range weather predictions.

EOS- Terra/MODIS (Moderate-resolution Imaging Spectroradiometer): land and ocean surface temperature, primary productivity, land surface cover (including vegetation), ocean color (sediment, phytoplankton), clouds, aerosols, water vapor, temperature profiles, and fire. MODIS views the entire Earth s surface every 1–2 days, making observations in 36 co-registered spectral bands at moderate resolution (0.25–1 km). MODIS is a whiskbroom scanning imaging radiometer consisting of a cross-track scan mirror, collecting optics, and a set of linear arrays with spectral interference filters located in four focal planes. With a viewing swath width of 2330 km (the field of view sweeps $\pm 55^\circ$ cross-track) MODIS provides high radiometric resolution images of daylight-reflected solar radiation and day/night thermal emissions over all regions of the globe. Spatial resolution varies from 0.25–1 km at nadir: 250 m (2 bands), 500 m (5 bands), 100 m (29 bands) at nadir. The broad spectral coverage of the instrument (0.4–14.4 μm) is divided into 36 bands of various bandwidths optimized for imaging specific surface and atmospheric features: 21 bands within 0.4–3.0 μm; 15 within 3–14.5 μm. Polarization sensitivity: 2% from 0.43–2.2 μm and $\pm 45^\circ$ scan. Absolute irradiance accuracy: 5% for <3 μm and 1% for >3 μm. The observational requirements also lead to a need for very high radiometric sensitivity, precise spectral band and geometric registration, and high calibration accuracy and precision. The MODIS instrument has one of the most comprehensive onboard calibration subsystems ever flown on a remote sensing instrument. Calibration hardware includes a solar diffuser stability monitor, a spectroradiometric calibration assembly, a plate-type black body, and a space view port.

EOS- Terra/ASTER: spectral radiances and reflectances of the Earth s surface; surface temperatures and emissivities; digital elevation from stereo images; surface composition and vegetation; cloud, sea ice, and polar ice; volcanoes and other natural hazards. Developed and built by MITI-Japan, ASTER has a broad spectral coverage and high spectral resolution with 14 spectral bands covering visible and near-IR (VNIR), SWIR and thermal IR (TIR) wavelengths. Spatial resolution is high, ranging from about 15–90 m. ASTER, with three distinct telescope subsystems, is a high spatial, spectral, and radiometric resolution, 14-band, imaging radiometer. Spectral separation is accomplished through discrete bandpass filters and dichroics. The VNIR subsystem operates in three visible and near IR bands with 15 m resolution and 60 km swath width. It consists of two telescopes, one that looks backward (along track) and one that looks at nadir, which together produce same-orbit stereo images. The telescope pair is pointable cross-track over a $\pm 24^{\circ}$ angle to increase the revisit frequency of any given Earth location. Light from either of two onboard halogen lamps will be used periodically for calibration. The SWIR subsystem operates in six SWIR channels with a 30 m resolution and 60 km swath width. It contains a pointing mirror that can point $\pm 8.54^{\circ}$ from nadir to allow coverage of any point over the spacecraft s 16 day cycle. The mirror is periodically used to direct light from either of two onboard calibration lamps into the subsystem s fixed, aspheric, refracting telescope. The TIR subsystem operates in five TIR channels with 90 m resolution and 60 km swath width. It contains a scan mirror used for both scanning and pointing up to $\pm 8.54^{\circ}$ from nadir. As with the SWIR, this mirror is also used periodically to view the onboard blackbody for calibration. Light from the TIR scan mirror is reflected into a Newtonian catadioptric telescope system with an aspheric primary mirror and lens for aberration correction. ASTER instruments operate for a limited time during the day and night portions of an orbit. The full configuration (all bands plus stereo) collects data for an average of eight minutes per orbit. Reduced configurations (limited bands, different gains, etc.) can be implemented as requested by investigators. ASTER is the highest spatial resolution instrument on EOS-Terra, and the only one that does not acquire data continuously. Monitoring applications products include: active volcanoes, crops and crop stress, cloud morphology and physical properties, wetlands evaluations, thermal pollution, coral reef degradation, and surface heat balance.

EOS- Terra/MISSR: albedo, clouds, aerosols, land surface vegetation structure, land cover. MISSR is a new instrument on board EOS-Tera which views the Earth in each of four color bands (446 nm – blue, 558 nm – green, 672 nm – red and 866 nm – near-IR) and from nine view angles with nine cameras pointed in different directions. Spatial resolution is about 0.275–1.1 km, with swaths of 360 km. The nine angles are: 70.5, 60.0, 45.6, 26.1° in the forward and aft directions, and 0.0°. MISSR is the first instrument to carry instruments making simultaneous observations at different angles from the same platform. Different combinations of look angles and spectral band are optimized to meet different scientific objectives. Over a period of seven minutes, a 360 km wide swath of the Earth comes into view at all nine angles. Highly accurate absolute and relative radiometric calibration is achieved onboard using deployable solar diffuser plates and several types of photodiodes. Onboard calibration is complemented by *in situ* field instruments (such as PARABOLA-III), which automatically scan the sky and ground at many different angles, and a multi-angle aircraft camera (AirMISR). Global coverage is acquired every nine days at the equator. MISSR will monitor the monthly, seasonal and long-term trends in the amount and type of atmospheric aerosol particles (natural and human produced); amount, type and heights of clouds; and the distribution of land surface cover, including vegetation canopy structure.

EOS-Terra/MOPITT: CO and CH_4 profiles. MOPITT, developed at the University of Toronto, Canada, is a scanning radiometer employing gas correlation spectroscopy to measure upwelling and reflected IR radiance in three absorption bands of CO and CH_4. The instrument modulates sample gas density by changing the length or the pressure of the gas sample in the optical path of the instrument, with detectors at 2.3, 2.4, and 4.7 μm. CO concentrations are obtained using pressure and length modulation with three independent pieces of information represented by values on seven pressure levels as well as CO and CH_4 columns. Spatial resolution: 22 km at nadir. Swath width: 640 km. MOPITT operates continuously, providing data on both day and night portions of an orbit. Calibration using onboard blackbodies and a space look occurs monthly and provides calibration at an elevated blackbody temperature. MOPITT data is used to: measure and model CO and CH_4 in the troposphere; obtain CO profiles with a resolution of 22 km horizontally and 3 km vertically, with an accuracy of 10%; measure the CH_4 column in the troposphere with a resolution of 22 km and a precision of better than 1%; and generate global maps of CO and CH_4 distribution, and provide increased knowledge of tropospheric chemistry. Heritage: AVHRR, HIRS, Landsat TM, Nimbus-7 Color Scanner.

EOS- Aqua (PM crossing): scheduled for launch in July 2001, Aqua will monitor the Earth System at a PM equatorial crossing time, complimenting EOS-Terra (AM crossing). Aqua will measure: clouds and aerosols; land surface temperature (visible conditions); SST (all weather); atmospheric temperature and humidity; snow cover and sea ice; global biological productivity; and atmosphere–land–ocean interaction. The instrument suite on board Aqua includes: MODIS and CERES (also on Terra), The advanced scanning microwave radiometer (AMSR-E), atmospheric IR sounder

(AIRS), AMSU-A, a humidity sounder for Brazil (HSB). MODIS and CERES on Aqua provide complementary and higher diurnal time resolution with the identical instruments on Terra (AM crossing). AIRS will fly on Aqua with two operational microwave sounders, the AMSU and HSB. The AMSU and HSB measurements will be analyzed jointly to filter out the effects of clouds from the IR data in order to derive clear column air temperature profiles with very high vertical resolution and accuracy, plus high accuracy SSTs. The HSB microwave radiometer is being developed by Brazil.

EOS- Aqua/AMSR-E - NASDA/Japan and NASA: rain, clouds, water vapor, SST, surface winds, snow/ice, soil wetness. The AMSR, developed by NASDA, Japan, and NASA, will measure rain rates over both land and ocean, as well as cloud water, water vapor, sea surface winds, SST, ice, snow, and soil moisture (surface wetness). Of particular importance is an external calibration design, which has proved suitable in other satellite microwave instrumentation for the long-term monitoring of subtle changes in temperature and other variables. AMSR-E will have an offset reflector, 1.6 m in diameter, and rotating drum at 40 rotations per minute. Multiple feedhorns (six) will cover six bands from 6.9–89 Ghz with 0.3–1.1 K radiometric sensitivity; vertical and horizontal polarization. Accuracy: 1 K or better. Swath: 1445 km. Resolution: 6×4 km (89.0 Ghz), 14×8 km (36.5 Ghz), 32×18 km (23.8 Ghz), 27×16 km (18.7 Ghz), 51×29 km (10.65 Ghz), 75×43 km (6.925 Ghz). Over the ocean, AMSR-E microwave frequencies can probe through smaller cloud particles to measure the microwave emissions from larger raindrops; sensitivity to ocean rainfall rates as high as $50\,\mathrm{mm\,h^{-1}}$ will be possible. Over the ocean, AMSR-E will also produce SST through most types of cloud cover, supplementing IR measurements of SST that are restricted to cloud free areas. The marked contrast between the microwave emissions of sea ice and water will be used to obtain observations of ice concentration in both polar regions. Over land, AMSR-E can measure the scattering effects of large ice particles that later melt to form raindrops. Though less direct, these measurements can be converted to rainfall rates using cloud models. Low frequency microwave (6.925 Ghz) will be used to measure soil wetness in areas where not too much vegetation is present. AMSR-E will provide the most useful data yet to assess how well low frequency microwave observations can measure surface soil wetness.

EOS- Aqua/AIRS (AMSU-A, HSB): temperature/humidity profiles, clouds, day and night land/ocean surface skin temperature and outgoing longwave radiation flux. NASA s AIRS will fly on Aqua with two operational microwave sounders: AMSU-A and HSB. AMSU-A and HSB measurements will be analyzed jointly to filter out the effects of clouds from the IR data in order to derive clear column air temperature profiles with very high vertical resolution and accuracy, plus high accuracy surface temperatures. Together these instruments constitute an advanced sounding system relative to the HIRS/MSU system currently operating on NOAA satellites. AIRS was designed to meet the NOAA requirement of a HIRS to fly on future operational weather satellites. The AIRS IR sounder measures the Earth s outgoing radiation at 0.4–1.0 μm and 3.7–15.4 μm, measuring simultaneously in over 2300 spectral channels. The high spectral resolution enables the separation of the contribution of unwanted spectral emissions and, in particular, provides spectrally clean super windows, which are ideal for surface observations. The AMSU-A microwave radiometer provides atmospheric temperature measurements from the surface to 40 km. Its primary objective is to obtain profiles of stratospheric temperature and to provide a cloud filtering capability for tropospheric observations. The HSB microwave radiometer, developed by the Brazilian National Institute for Space Studies, is designed to obtain profiles of atmospheric water vapor/humidity and to detect precipitation under clouds with 15 km (nadir) resolution. The AMSU-A and HSB have a total of 19 channels; 15 are assigned to AMSU-A, each having a 3.3° beamwidth, and four are assigned to HSB each having a 1.1° beamwidth. Channels 3–14 on AMSU-A are situated on the low frequency side of the oxygen resonance band (50–60 Ghz) and are used for temperature soundings. Successive channels in this band are situated at frequencies with increasing opacity, therefore responding to radiation from increasing altitude. Channel one, located on the first (weak) water vapor resonance line, is used to obtain estimates of total column water vapor in the atmosphere. Channel two, at 31 Ghz, is used to indicate the presence of rain. Channel 15 on AMSU-A, at 89 Ghz, is used to indicate precipitation using the fact that at 89 Ghz, ice more strongly scatters radiation than it absorbs or emits. Channels 16–19 are located on the wings of the strongly opaque water vapor resonance lines at 183.3 Ghz. Successive channels in this group have decreasing opacity and consequently their data correspond to humidities at decreasing altitudes.

EOS-Aura (Previously called CHEM): atmospheric chemical composition. Scheduled for launch in June 2003, Aura will comprehensively measure atmospheric chemistry. Instruments include the MLS, and the high resolution dynamic limb sounder (HRDLS). MLS measures lower stratospheric temperature and concentrations of water, ozone, ClO, bromine oxide, HCl, OH, HNO_3, hydrogen cyanide, and nitrous oxide (N_2O) for their effects on (and diagnosis of) ozone depletion, transformation of greenhouse gases, and radiative forcing of climate change. HRDLS is an IR limb sounder to sound the upper troposphere, stratosphere, and mesosphere to determine: temperature; concentrations of ozone, water, CH_4, N_2O, NO_2, HNO_3, N_2O_5,

CFC11, CFC12, $ClONO_2$; aerosols; and polar stratospheric clouds and cloud tops.

Various other specialized satellites are also planned under NASA s Earth Science Enterprise (ESE) program, and the counterpart programs of Europe, Japan, and other countries. NASA/ESE examples include: the Vegetation Canopy Lidar (VCL) to measure heights of the vegetation canopy, the EOS–Ice Sheet Altimetry Mission (ICESat), the Solar Radiation and Climate Experiment, CloudSat to understand the role of optically thick clouds on the Earth s radiation budget (jointly with Canada), PCASSO-CENA (jointly with France) to measure the vertical distribution of clouds and aerosols. The descriptions of some of these follow.

EOS- ICESat (rst precision altimetry mission): the ice, cloud and land elevation satellite; measurement of changes in topography and mass balance of polar ice sheets. Scheduled for launch in December 2001, the EOS-ICES will carry the geoscience laser altimetry system and a GPS orbit determination system, to measure along-track ice sheet and land topography to an absolute accuracy of 10 cm or better, within a footprint of 70 m or less. The mission will also provide cloud profile information with reduced vertical information. A repeat mission, focused on the primary science goal of measuring changes in the topography and mass balance of polar ice sheets is in planning for the 2010 time frame.

SeaWinds (surface winds and wind stress over the ocean surface): scheduled for launch on Japan s ADEOS-II in November 2001, SeaWinds builds on the heritage of the NASA Scatterometer (NSCAT) instrument flown aboard the Japanese ADEOS-I (Advanced Earth Observing Satellite) in 1996–1997. As an interim measure, NASA flew QuickScat in June 1999, to further improve and test the active scatterometer technique. SeaWinds is a dual, conically scanning pencil-beam radar operating at 13.4 Ghz. SeaWinds vector wind measurements are obtained with a 25 km resolution across nearly the entire swath of about 1800 km, and cover more than 90% of the global, ice-free oceans daily. Significant ground processing is required to extract backscatter cross-section and vector wind measurements from raw SeaWinds radar measurements. Concurrently, the US Navy is preparing an experimental satellite mission (CORIOLIS) to test the passive microwave polarimetry-radiometer technique for vector wind finding applications.

VCL: the full realization of the scientific potential from instruments on board EOS-Terra and EOS-Aqua as regards land cover/global productivity, short-term climate modeling, and natural hazards studies, will require global topography at digital terrain elevation data Level-2 resolution (30 m spatial, and 16 m vertical). Interferometric SAR data or the conventional photogrammetric data sets are only relative in their measurement of surface elevation. Direct measurements are essential to control the vertical dimension of the topographic image. The shuttle radar topographic mission has been flown to address some of these needs, in particular for accurately mapping the Arctic and the Antarctic ice surfaces. The multi-beam laser altimeter (MBLA) instrument on board the VCL will provide a common global reference frame that is a direct measurement rather than one that is inferred. Within the waveform from MBLA, the first return above a threshold represents the top of the canopy, and the last return represents the ground return. Canopy height is calculated from the difference between the two. VCL is scheduled to be launched in the 2004 or later time frame.

International Developments in Satellite Observing Technology

The Earth Observing System satellites developed by Japan is an example of the development of satellite technology internationally. The ADEOS-II, scheduled for launch around 2001 will replace ADEOS-1, launched in August 1996. The ADEOS satellite series were designed with a comprehensive suite of instruments to monitor the whole Earth System. They include: an ocean color and temperature scanner, developed by NASDA to obtain ocean color and SST data; an advanced visible and near-IR radiometer, developed by NASDA, an optical sensor for the observation of the solar light reflected by the land and coastal zones in the visible and near-IR regions (field of view: 80 km). The instrument will also be useful to monitor deforestation, desertification, and for land use and resource exploration; an improved limb atmospheric spectrometer developed by the Environment Agency (EA) of Japan to monitor high altitude stratospheric ozone layer above the Earth; a retroreflector in space, developed by EA (Japan), is used for Earth–satellite–Earth laser long-path absorption measurements. Ozone, CFC-12, CO_2, CH_4 and others will be measured using pulsed visible and IR lasers; an interferometric monitor for greenhouse gases, developed by the MITI, to monitor the density profile of greenhouse gases such as CO_2, water vapor, CH_4, nitrogen monoxide, and ozone; NSCAT, to measure wind speed and direction over the ocean without being affected by weather conditions. NSCAT data will include backscatter maps, ocean wind vectors, spatially averaged wind field maps, and temporally averaged wind field maps; polarization and directionality of the Earth s reflectances, developed by the CNES of France, to observe the polarization and directional/spectral characteristics of solar light reflected by clouds, oceans, land surface, etc. ADEOS-I successfully operated for about one year before encountering problems with its solar array power supply – it ceased operation on 30 June 1997.

Programs are also planned by the ESA for the next generation of environmental satellites (ENVISAT) that will expand the capabilities of the ERS series of satellites. ENVISAT-1, will carry an advanced SAR, a medium resolution imaging spectrometer, a Michelson interferometer for passive atmospheric soundings, a global ozone

monitoring by occultation of stars instrument, a scanning imaging absorption spectrometer for atmospheric chartography, a radar altimeter (RA-2), a K-band passive microwave radiometer, an advanced along-track scanning radiometer, a doppler orbitography and radio positioning integrated by satellite system, and a laser retro-reflector.

Similarly, the Indian Space Research Organization plans to expand the capabilities of the IRS series of satellites. Future IRS satellites planned include, among others: IRS-P4 OCEANSAT-1, with an ocean color monitor (OCM) with eight narrow spectral bands with a 350 m resolution and a 1500 km swath, and a multi-frequency scanning microwave radiometer operating in four frequencies to provide valuable ocean–surface related observation capability; and the IRS-P5 (CARTOSAT-1) with an OCM with eight narrow spectral bands, 350 m resolution, and a 1500 km swath. Details on other satellite programs are generally accessible through the Internet.

CONCLUSION

Earth observing in the future is likely to be a seamless blend of surface-based, *in situ*, and space-based systems and networks, each with its own unique and characteristic merits and demerits. Some parameters will be measured by both systems, while others will rely exclusively on one or the other to maximize performance of the sensors used while recognizing their physical limits. More and more, the sharp differences between raw observations and analysis are disappearing through computer models of the Earth System (or its components) that are used for data assimilation. They produce data fields needed for the initialization of prediction models as well as retrospective analyses. Importantly, they produce quantitative information for a range of parameters and products that can either not be observed directly or are derived from more than one observation. Energy, momentum, heat, and moisture exchange fluxes belong in this category. There are certain drawbacks that need to be recognized in applying model-dependent data assimilation systems to observed data. The resulting data or analysis field is no longer a direct measurement of an Earth System parameter, but rather has a component that depends on the mathematical or numerical model used at the time. That is, the observation could be prone to unknown model dependent analysis error, especially for the detection of change over time. To minimize the risk arising from such a procedure, it would be necessary to maintain the raw observations, especially those directly measuring a parameter, so that observing systems may be considered baseline or reference systems. Such a concept has been initiated for climate through the definition, identification, and initial implementation of climate reference networks. It would be useful to have counterpart space-based reference observational sensors. In the case of satellites, this would entail carrying a specified set of instruments on all future satellites that would provide continuity to older measurements alongside new sensors with improved technology and accuracy and calibration.

Another emerging concept is that of a sensor web. The sensor web, proposed as one of the centerpieces of a new NASA vision of the future for Earth System science, could consist of a reconfigurable, smart constellation of microsatellites that would adjust their data collection based on emerging events on Earth. Such a web of satellites could affordably provide the robustness that an operational system requires by allowing several small, separate platforms working together to accomplish what was formerly thought to need large, multi-instrument platforms. Scientists and other users could, in principle, have access to any on-orbit sensor(s) and be able to direct and control those sensors in the same manner as we access information on the Internet today. Such a constellation of smart satellites could, in principle, be able, on command, to provide the high-resolution data needed to cover energetically active events, especially hurricanes, volcanic eruptions, large forest fires, and floods, among others, and provide the information required for the mitigation of the impact of natural disasters. The concept will be tested in the EOS-Era. The EOS Landsat, Terra, Aqua, and Aura satellites will be flown in tandem with Cloudsat, and PCASSO in two separate constellations to create an *observatorium*, in the words of Ghassem Asrar, associate administrator for NASA s ESE. Future satellite systems are also likely to use new technologies such as: micro-electromechanical systems including micro-gyroscopes; sub-millimeter solid thrusters, measuring millimeters in diameter, to maintain attitude control for microsatellites; optical communications devices using laser transmitters to beam data at rates exceeding 1 Gigabit sec^{-1}; monolithic microwave integrated circuits; miniature digital active pixel sensors; and diffractive optics.

FURTHER READING

Unninayar, S and Schiffer, R A (1997) *In Situ Observations for the Global Observing Systems: A Compendium of Requirements and Systems*, NASA, NP-1997(01)-002-GSFC, 1–286, (http://www.earth.nasa.gov/visions/in situ/).

CEOS (1998) *Resources in Earth Observation* (CD-ROM and WWW), Production Editors, Centre Nationale d Etudes Spatiales, ed (Version 1) Commonwealth Scientific & Industrial Research Organization (CSIRO), Australia, (http://ceos.cnes.fr:8100/cdrom-98/astart.htm).

NASA (1999) *1999 EOS Reference Handbook: a Guide to NASA's Earth Science Enterprise and the Earth Observing System*, eds M D King and Greenstone, NASA NP-1999-08-134-GSFC, 1–361, (http://eos.nasa.gov) (http://eospso.gsfc.nasa.gov/eos_homepage/misc_html/refbook.html).

NASA (1999) *EOS Science Plan: the State of Science in the EOS Program*, eds M D King, R Greenstone, and W Bandeen, NASA NP-1998-12-069-GSFC, 1–397, (http://eos.nasa.gov) (http://eospso.gsfc.nasa.gov/sci_plan/chapters.html).

The Global Temperature Record

PHILIP D JONES

University of East Anglia, Norwich, UK

Global mean surface temperature is the most commonly used measure of the state of the climate system. Periods in the past are generally named according to their relative global mean temperature, e.g., the ice ages in relation to past and present interglacials. Variations of global mean temperature are also the measure against which we assess the sensitivity of the climate system to external factors thought to be of importance, such as changes in greenhouse gases, sulphate aerosols, solar output and the frequency of explosive volcanic eruptions. One of the main foci of climatology is quantifying the response of the climate system, as measured by global mean temperature, to all possible external forcings. Studies of this kind are essential if we are to predict future climatic change.

Local surface temperatures are one of the most important environmental indicators. Together with precipitation, seasonal variations determine the characteristic oral and faunal variability over terrestrial areas. All human life resides at the surface and the air temperature (AT) can determine events and moods, through the changes of temperature that occur in response to the seasons.

The record of past variations in surface temperature is extensive but of variable quality. Most is known about the instrumental era (1850 onwards), but even here, measuring sites are irregularly spaced, so that even now about 15- 20% of the surface is not being monitored. Measuring temperature in a consistent manner is harder than it may appear. Changes in thermometers, their housing and location, the time of day that readings are taken, and the environment around the instrument can all impair the record. All measuring seeks homogeneous values that are solely caused by the vagaries of weather and climate.

This article addresses the quality of the available records and what they tell us about the last 150 years. Since 1850, the global surface temperature average has risen by about 0.6°C, with the four warmest years of the record occurring in the 1990s: 1998 (0.57°C above the 1961- 1990 average), 1997 (0.43), 1995 (0.39) and 1990 (0.35). Recent record global averages may have caught the headlines, but the individual station and marine records tell us about the spatial patterns of the changes, the relative warming of night compared with daytime values, all factors that need explanation, either through modeling or empirical studies. Successful prediction, through general circulation models (GCMs), requires that we must also be able to adequately explain the spatial and temporal patterns of the last 150 years.

With both current assessments and modeling explanations we must realize that the last 150 years is a relatively short period in Earth's history. To place the period in a longer-term context we must resort to less well de ned (proxy) measures of past temperature change, from information held in written historical archives or in natural archives such as tree rings, ice cores, corals and varves (lake sediments). Proxy measures are both less reliable and less complete, most natural archives being biased somewhat to the growing season rather than the calendar year, and restricted spatially in some regions. The last 1000 years is the most relevant to the last 150 years and to the next century. Proxy evidence is becoming more extensive and analyses of this resource for the last 1000 years indicate that the 20th century was the warmest and the 17th the coldest. Individually, 1998 was the warmest year, with 1601 the coldest, being about 1.0°C colder than the 1961- 1990 average.

QUALITY OF BASIC TEMPERATURE DATA

In searching for trends or changes in observational data, it is essential that all series be as internally consistent, or homogeneous, as possible. In climatology, a temperature time series is homogeneous if the variations exhibited by the series are solely the result of the vagaries of the weather and climate. Numerous non-climatic factors can influence

the basic data, causing erroneous conclusions to be drawn concerning the course of temperature change. Before any analyses can be undertaken, it is vital, therefore, that all data are tested for homogeneity. Causes of inhomogeneity are varied and sometimes surprising. For temperature data measured at the Earth s surface, they are best discussed by dividing the data into its two components, land-based and marine records.

Land

Observations made at weather or climate stations are subject to inhomogeneities due to:

- changes in instrumentation, exposure and measurement technique;
- changes in station location (both position and elevation);
- changes in observation times and the method used to calculate monthly averages;
- changes in the environment of the station, particularly with regard to urbanization, that might affect the representativeness of local temperature records.

Few temperature series have followed the same protocols for their entire history. Since the late 19th century, the accepted exposure of thermometers has been in louvred (Stevenson) screens so that the sensors are not affected by direct sunlight. All measurements prior to the use of screens must be treated with caution. The common practice before screens was to attach the thermometer to a wall or building away from direct sunlight (e.g., north facing in the Northern Hemisphere) but direct sunlight can strike such exposed thermometers in summer, particularly in more poleward locations. Overhanging eaves are also a problem as are potential heat sources from inside the building.

In the last 20 years, automation has led to the use of electrical resistance as opposed to mercury or alcohol thermometers, thereby introducing different response times. Simultaneous measurements of the new and old systems are essential for calibration, but these are rarely adequately carried out. It is also important to note that technological improvements may be introduced for many reasons but they do not always lead to more accurate or more homogeneous measurements.

Few observation sites have remained exactly the same over time. Many sites were moved to airport locations in the 1940s and 1950s and were then regularly moved as the airport developed. Again, parallel measurements are essential to correct early readings to the new locations. Station history information is vital, but details of moves may have been lost or not been recorded because it was not considered important at the time.

Over the course of the last 150 years, many schools of thought have existed about when to take measurements and how to calculate the average daily temperature and hence the monthly and annual averages. Practices now vary between and within countries and have varied over time at almost all sites. In the early 20th century, efforts were made to measure the true daily mean temperature (the average of readings every hour) by a combination of a few (generally three or four) readings at fixed and convenient times. In most English speaking countries, the accepted method for the daily temperature average became the mean of the daily maximum and minimum temperature, the development of a thermometer to achieve these two measurements coming around 1860. Most other regions generally used measurements made at fixed hours. With hindsight, it would have been best for all sites to have maintained the same observation times and practices. Changes in specified observation times, the switch in many countries to the use of the simpler maximum and minimum method, and changes to the times of reading the maximum and minimum thermometers have all led to inhomogeneities in the records. All can be overcome with care and with good station history information. For many sites, however, this information is lacking or was not considered important enough to record.

The final problem, changes in the station environment, is generally beyond the control of the observer. Sites which were small towns in the 19th century are now large cities and the measured temperatures are likely to be warmer than without the urban growth. The problem is a real change in climate, although not all urban sites are affected. Some, because of locations in parks or on coasts, are not significantly affected. Urbanization is considered an inhomogeneity in the sense that the measured temperatures from affected sites are now only representative of very localized areas. This problem is the most serious of the four because environmental changes most often produce gradual warming (irrigation of agricultural fields or creation of a golf course are examples that can create cooling), whereas the other three generally induce abrupt jumps in station time series.

Correction of the temperature time series for available stations (about 2500 series are available, digitally, in total, varying in length from between 200+ and 40 years) is a necessary but laborious task, often made difficult or near impossible by the lack of adequate information. Numerous subjective and relatively objective criteria have been developed for testing monthly average data (see recent review by Peterson *et al.*, 1998a). All rely on intercomparison of neighbouring time series, assuming that the closer the sites, the more the records should agree. The urbanization question can be addressed only if suitable rural stations are available. If all sites are cities/large towns, their records may be similarly affected. While there is no definitive basis for establishing accuracy, all the necessary adjustments are made according to station

history information and experience and so are not arbitrary. Adjustments are generally applied on a monthly basis and must be considered only first approximations if daily data are used.

Marine

Historical temperature data over marine regions are largely derived from *in situ* measurements of sea surface temperature (SST) and marine air temperature (MAT) taken by ships and buoys (*see* Sea Surface Temperature, Volume 1). To be of use, each measurement must be accompanied by a location value. Incorrectly located data must be discarded. Up to 15% of the data have been found to be mislocated (e.g., SST/MAT values occurring over land) and can be easily discarded. The major problem arises with the data located in the wrong ocean, generally from coding mistakes with hemispheres (north for south, etc.). Homogeneity problems relate to changes in instrumentation, exposure and measurement technique. For SST data, the changes in the sampling method from uninsulated canvas buckets (generally prior to the early 1940s) to engine intake measurements (early 1940s onwards) causes a rise in SST values of 0.3–0.7 °C. For MAT data, the average height of ships decks above the ocean surface has increased during the 20th century, but the more serious problem is the contamination of daytime MAT through solar heating of the ships infrastructure, rendering only the nighttime MAT (NMAT) data of value.

In combining marine data with land-based surface temperatures to provide a more complete coverage of the Earth s surface, SST anomalies are used in preference to NMAT because they are considered more reliable, principally because there are at least twice as many observations due to the rejection of all daytime MAT values. The basic assumption in their use is that, on monthly and longer timescales, an anomaly value of SST will be approximately equal to that of the air above.

Correction of the SST data for the change from canvas buckets to engine intakes is achieved using a physical-empirical model to estimate the amount of cooling of sea water that occurs in buckets of various designs, depending upon the time between sampling and measurement and the ambient weather conditions (Folland and Parker, 1995). Correction factors for the bucket SST values for all years before 1942 have been estimated using this model. Corrections, which are greater in winter, are tuned using an additional assumption that SST annual cycle amplitudes have not experienced a long-term trend. This gives a sampling time of 3–4 minutes approximately what is generally believed to have occurred.

Since the marine and land components are independent, the two records can be used to evaluate each other after correcting both. The components have been shown to agree well with one another on an hemispheric basis by Parker *et al.* (1994) and using island and coastal data (Folland and Salinger, 1995).

AGGREGATION OF THE HOMOGENEOUS DATA

The basic homogeneous data are irregularly located over the Earth s surface. For ease of use they must be combined into a regular grid, generally based on latitude and longitude. The land and marine components are first developed separately and then amalgamated.

Land

Due to differing station elevations and national practices with regard to the calculations of monthly mean temperatures, interpolation to a regular grid is much more easily achieved by converting all the monthly data to anomalies or departures from a common period. The period with the best available data is 1961–1990. Many station series are not complete, even for this period, so, in practice, each station is required to have 20 years of data in this period to be included in compilations (see Jones, 1994 for more discussion).

The station anomaly values are then simply averaged together within each $5 \times 5^{\circ}$ grid-box. Resulting grid-box time series will thus differ in the number of stations used, both between grid boxes and through time. Individual grid-box time series may, therefore, have differences in variance with time due to changes in station numbers. The implications of this may affect some analyses and will be highlighted later. Several different variants of this basic approach to aggregation have been introduced. Although the results differ slightly, differences in the estimated trends of global averages over the 1880–1990 period differ by only a few hundredths of a $^{\circ}$C 100 years^{-1} (Peterson *et al.*, 1998b).

Marine

For SST, the raw data must be aggregated in a different manner. First, an average field for the 1961–1990 period is derived for each $1 \times 1^{\circ}$ square of ocean for each five day period (pentad) and day of the year. This background climatology is then smoothed spatially and across pentads/days to reduce errors in regions with few observations during the 30 years. Additionally, satellite estimates of SST from 1979 onwards are used to help infill data voids and very poorly sampled (by *in situ* measurements) regions. Grid-box anomalies, from the daily climatology, for each month from 1856 are produced by trimmed averaging of all 1° anomalies for each of the pentads that make up each month. The trimming reduces the effect of outliers. The

final dataset is checked for outliers by comparing neighbouring 5° grid-box values.

Combining the Land and Marine Components

After processing, both components are available on the same 5° grid-box basis. The two data-sets are merged in the simplest manner. Anomaly values are taken where available from each component. For grid boxes where both components are available, the combined value is a weighted one, the weights determined by the fractions of land and ocean in the box. Since the land component for oceanic islands is potentially more reliable than a few surrounding SST values, the land fraction is assumed to be at least 25% for each of the island and coastal boxes. A corresponding condition is applied to the ocean fraction of boxes where this is less than 25%.

Hemispheric and Global Anomaly Time Series

Calculation of hemispheric average values is achieved by weighting all available grid boxes for each month. With latitude/longitude grid boxes, the weights used are normally the cosine values of the mid-point latitude value of each grid box to account for the convergence of longitude lines with latitude. Improvements in coverage mean that estimates improve in accuracy through time (see the following section on the accuracy of hemispheric and global series).

Figure 1 shows the two hemispheric time series by season and year. The global annual series exhibits a warming of 0.57 °C over the 1861–1997 period. Table 1 gives monthly linear trend values, estimated by least squares, for each hemisphere and for the globe calculated over the 1861–1997 and 1901–1997 periods. The extreme warm years of 1998 (and 1999) are omitted in the calculation of trends. Calculation of temperature trends by least squares regression is just one means of estimation. All methods have some shortcomings as the series shown in Figure 1 are only poorly approximated by a linear trend. Warming is slightly, but not significantly greater in the Southern Hemisphere (SH) compared with the Northern Hemisphere (NH). Warming in all months for all three regions is highly statistically significant (better than the 99% significance level), even allowing for the greater uncertainties earlier in the series (see the following section on the accuracy of hemispheric and global series). Warming is least in the NH summer months, perhaps because the warmth of the 1860s and 1870s may reflect exposure problems of thermometers

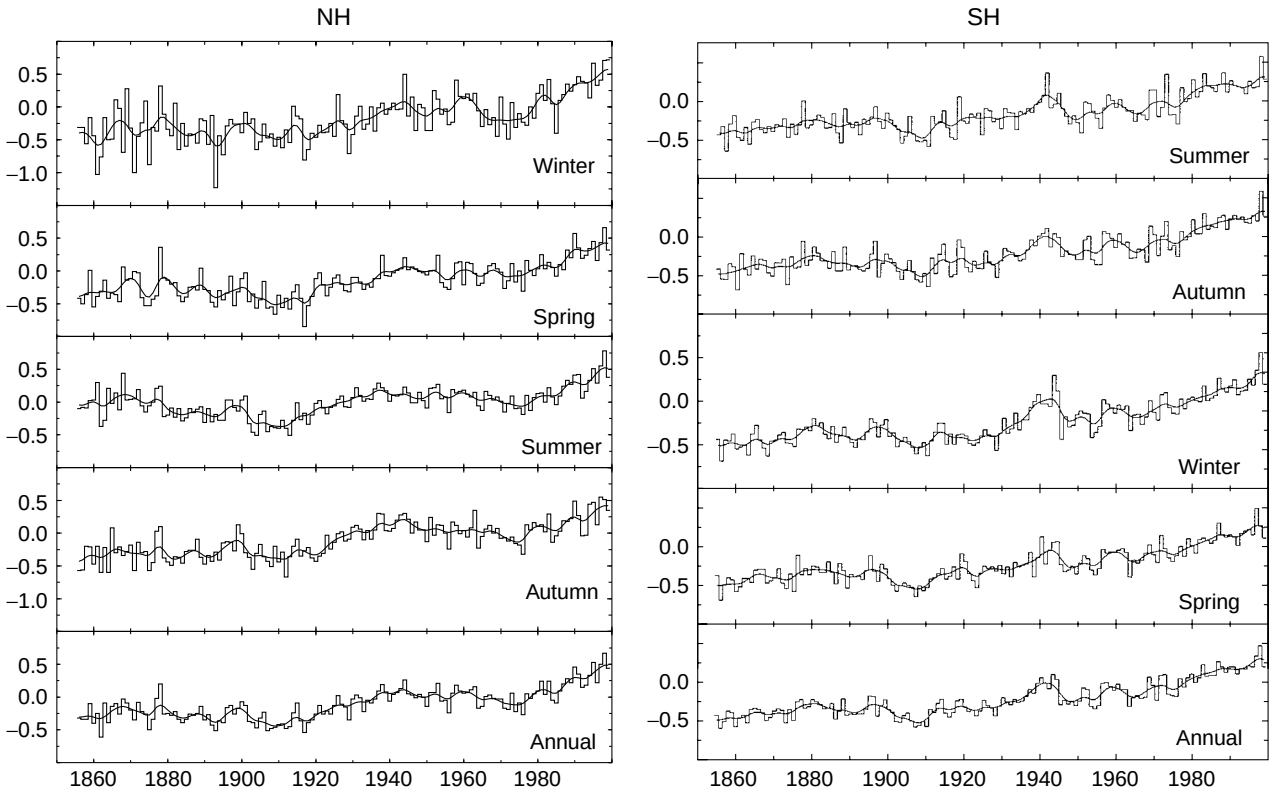


Figure 1 Hemispheric temperature averages by season and year for the period 1856–1999, relative to 1961–1990. Seasons are those routinely used in climatology, winter is the December, January and February average in the NH and the June, July and August average in the SH. The smooth line on these and subsequent time series in this section is a data adaptive filter (10 year Gaussian) that highlights variations on the decadal timescale

Table 1 Temperature change (°C) explained by the linear trend over two periods

	1861–1997			1901–1997		
	NH	SH	GL	NH	SH	GL
January	0.94	0.32	0.63	0.61	0.42	0.52
February	0.92	0.37	0.65	1.01	0.41	0.71
March	1.04	0.39	0.71	0.90	0.44	0.67
April	0.69	0.32	0.51	0.78	0.37	0.67
May	0.65	0.68	0.66	0.72	0.66	0.69
June	0.33	0.57	0.45	0.56	0.43	0.49
July	0.19	0.49	0.34	0.41	0.46	0.44
August	0.34	0.44	0.39	0.41	0.61	0.51
September	0.42	0.30	0.36	0.37	0.46	0.42
October	0.87	0.35	0.61	0.42	0.52	0.47
November	1.07	0.35	0.71	0.52	0.59	0.55
December	1.05	0.25	0.65	0.82	0.51	0.66
Year	0.71	0.41	0.56	0.63	0.49	0.56

in mid- to high-latitude land areas (see earlier section on homogeneity over land).

Differences between the seasons are much greater in the NH compared with the SH. The NH also shows greater year to year variability in winter than in summer. The greater variability in all seasons in both hemispheres before 1881 is principally due to the sparser data available during these years and represents greater uncertainty in the combined data set. This feature is more marked in the NH and occurs predominantly over land areas.

The seasonal and annual hemispheric series show extremely warm and cool years relative to the underlying trend (the smooth curve). Many of these anomalously warm and cool years are coincident between hemispheres suggesting a response to common forcing. Warm years are generally related to El Niño events and the cold years to La Niña events or to the effects of major explosive volcanic eruptions. The effects of volcanoes tend to give the NH series a slight negative skew in the year to year variability about the smoothed line.

Both hemispheres clearly show two periods of warming, 1920–1940 especially in the NH and since the mid-1970s. Slight cooling occurred between the two periods, albeit barely perceptible in the SH. In the global annual series (Figure 2), 12 of the 14 warmest years occured between 1986 and 1999. The warmest year of the entire series on a global basis occurred in 1998, 0.57°C above the 1961–1990 average. The four warmest years of the series occured in the 1990s with 1997 the second warmest (0.43), 1995 (0.39) and 1990 (0.35), the third and fourth.

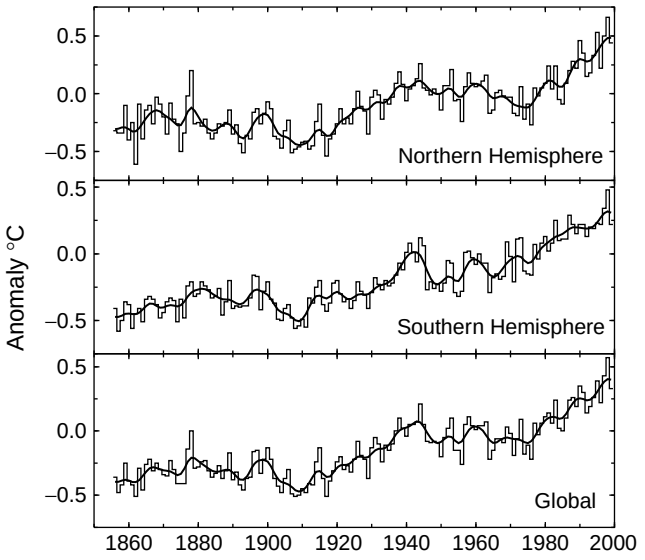


Figure 2 Anomalies of hemispheric and global temperature averages on the annual timescale, relative to 1961–1990

Accuracy of Hemispheric and Global Series

The series in Figures 1 and 2 and the underlying individual grid boxes are subject to two sources of errors: bad measurements and most importantly, changes in the availability and density of the raw data. Errors of the latter type are referred to as sampling errors. For the whole time period, data from the land and marine components have been evaluated for measurement errors. Extreme values exceeding three standard deviations from the 1961–1990 statistics are all checked for spatial consistency. Despite these checks, it is likely that erroneous values slip through, though any effect is considered to be essentially random and will be totally insignificant in the large-scale averages.

Sampling errors are much more important and occur because the availability and density of the data used in the compilation of the surface temperature field are ever changing. For most of the post-1860 period, apart from the two World Wars and since about 1980, the changes should have led to an improvement in the accuracy of

large-scale averages through the use of more complete data fields. Estimates of accuracy need to take into account both changes in the available network of data through time and the areas (particularly the higher latitude Southern Oceans) that are always missing.

Standard errors of regional and hemispheric estimates depend upon two factors: the locations and the standard errors (SE) of grid boxes with data. In the most comprehensive study of the issue, Jones *et al.* (1997) show that the errors are a function of four factors, the temporal variability of each grid-box temperature series, the spatial variability of temperature within each grid box, the number of stations available in each grid box, and the variability of temperature between grid boxes.

Temperature is more variable from one year to the next over continental interiors and polar regions compared with oceanic and coastal regions. To achieve the same accuracy in a grid-box average over Siberia requires more station series than the same sized box over, for example, Ireland. Spatial variability between station time series is also greater in continental and polar regions compared with oceanic and coastal areas. The fourth factor relates to the lack of statistical independence between neighbouring grid-box time series. This interdependence between the time series leads to the concept of the effective number of independent series. This is referred to in statistical terms as the effective number of spatial degrees of freedom. Since the 1950s, about 80% of the 2592 5° grid boxes have anomaly data, yet the number of independent values are only of the order 40–60 for annual averages (i.e., only this number are statistically independent of each other).

All of the factors, except the number of stations per grid box, depend upon timescale, so must be estimated for the time period over which the estimates of accuracy are required. The first two factors are relatively smooth variables when mapped, so this property can be used to estimate the values of these variables in unsampled regions (Jones *et al.*, 1997).

Figure 3 shows SEs calculated on the decadal time scale for hemispheric and global temperature time series. The greater percentage of missing regions in the SH gives higher SEs. Values decrease with time since 1860, with typical values after 1951 for the globe of 0.048 °C when calculating decadal averages. Values are only slightly larger for the interannual timescale (0.054 °C). Standard errors are about twice the post-1950 values for the two hemispheres during the 1860s and 1870s. The larger errors in the 19th century and particularly before 1881 are reflected in the greater variability of the seasonal series in Figure 1.

Standard errors on the interannual timescale may be used to determine whether one year is different in a statistically significant sense from another year. The year 1998, for example, was so extreme that it can be said with high statistical confidence to be significantly warmer than the next warmest year of 1997. The year 1998 also broke most of the monthly records with the first eight months of the year being the warmest such months in the global series.

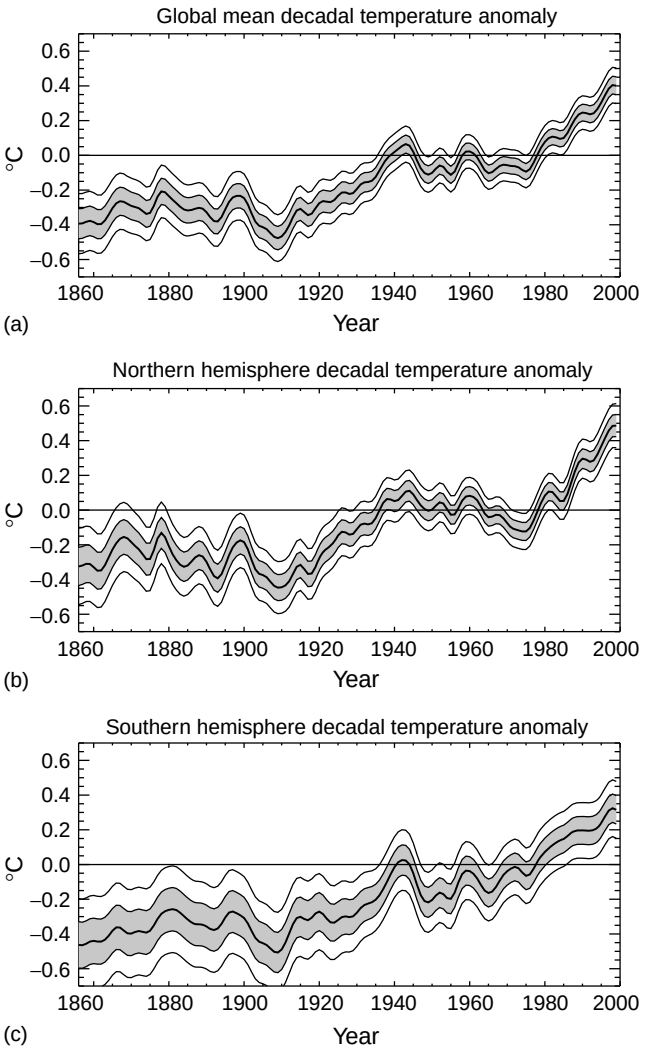


Figure 3 One and two SEs on the hemispheric and global temperature series for the decadal timescale. The shaded areas highlight ±1 SE

Consideration of these errors in the calculation of trends given in Table 1 enables these additional uncertainties to be added to the statistical uncertainties from estimating a linear trend. Combined, the effect of both types of errors creates uncertainties that are less than ±0•05 °C (monthly) and ±0•03 °C (annual) for the temperature changes given in Table 1.

ANALYSES OF THE SURFACE TEMPERATURE RECORD

The record of surface temperature has been extensively analyzed, particularly over the past 10–15 years, in review

papers and international reviews (e.g., those for the Intergovernmental Panel on Climate Change (IPCC) (see for e.g., Folland *et al.*, 1990, 1992; and Nicholls *et al.*, 1996 and earlier references therein). In this section, the following aspects will be addressed:

- patterns of recent warming;
- trends in areas affected by monthly extremes of temperature;
- trends in continental scale temperatures;
- trends in maximum and minimum temperatures;
- daily temperature extremes in long European series;
- the last 150 years in the context of the second millennium.

Global scale temperatures have recently also been calculated for the lower part of the troposphere. Although we might, *a priori*, expect some differences with surface values, much has been written about the comparisons made.

Patterns of Recent Warming

Figures 1–3 clearly indicate that most of the warming this century occurred in two distinct periods. The periods differ slightly between the hemispheres and among the seasons, but the periods from about 1920–1945 and since 1975 stand out. Figure 4 shows the trend of temperature change (calculated by linear trends for each grid box), by season and annually, for the two 20 year periods, 1925–1944 (a) and 1978–1997 (b). The annual global temperature increases during the two periods were 0.37 and 0.32 °C, respectively, with the second period 0.27 °C warmer than the first. Despite these periods exhibiting strong warming in the global series, not every region warms. Statistically significant cooling (not shown) occurs in a few areas in some seasons. There are marked seasonal patterns and the term *global warming* should only be used to refer to warming in the global average. It clearly does not mean warming everywhere on the globe.

In both periods, the regions with the greatest warmings and coolings occur over the northern continents, and, with the exception of a few isolated grid boxes, during the December January February (DJF), March April May (MAM) and September October November (SON) seasons. The largest warmings/coolings may not always be the most statistically significant because of the high year to year variability of continental areas. Small trends over some oceanic areas are significant because of low variability from year to year. Warming clearly dominates and the greatest magnitude values during the DJF and MAM seasons are over the high latitudes of the northern continents.

The 1978–1997 period, despite a lower warming in the global average, has a much larger area with statistically significant warming compared with 1925–1944. Some of this is related to better coverage recently, especially in the tropics. The improvements in coverage between the two periods are apparent, with shipping routes surrounded by data voids in the 1925–1944 period. The pattern of recent warming (1978–1997) is strongest over northern Asia, especially northern Siberia. Warming is also evident over much of the Pacific basin, western parts of the USA, western Europe, southeastern Brazil and parts of southern Africa. Seasonally, the annual pattern is dominated by the DJF and MAM seasons, especially north of 30 °N. Despite this, the season with the greatest spatially significant warming is June July August (JJA), because of lower levels of year to year variability. The area with greatest cooling occurs in SON along the Rocky mountain chain.

The 1925–1944 warming was strongest over northern North America and is clearly evident over the northern Atlantic, parts of the western Pacific and central Asia. The largest coolings occur over northern Asia in SON and much of Europe in DJF. This latter area highlights a problem with trends over a short period of time such as 20 years. European winters were anomalously cold during the early 1940s. Changing the period from 1925–1944 to 1921–1940 would change the slight cooling in this region to slight warming.

Both sets of maps illustrate large differences in trends between apparently close regions. This high degree of spatial variability is partly real, but part of it is likely due to data problems. Measurement errors and inadequacies with homogeneity are factors, but the main problems relate to regions with a few missing years. Grid boxes required 10 of the 20 years to calculate a trend and each season (year) required 2 (8) months of data present.

The choice of the two periods may be somewhat arbitrary, but all other choices would have similar failings. Mapping longer trends over the whole century would have placed great emphasis on the decision of how many years are necessary to constitute a trend, creating situations in which maps would have sparser coverage than for 1925–1944. The best way to view the data is to look at running trends as part of a video sequence, considering 10- or 20-year averaging intervals.

Trends in Areas Affected by Monthly Extremes of Temperature

The previous section showed that regions with the strongest temperature trends are not always those which are the most statistically significant. A more useful measure of temperature change, from an impact point of view, would be one that treats all areas equally by removing the effects of different magnitudes of year to year variability. Taking this approach, the rarity of extreme monthly temperatures can be directly compared between regions. An appropriate transformation uses the gamma distribution (in simple terms

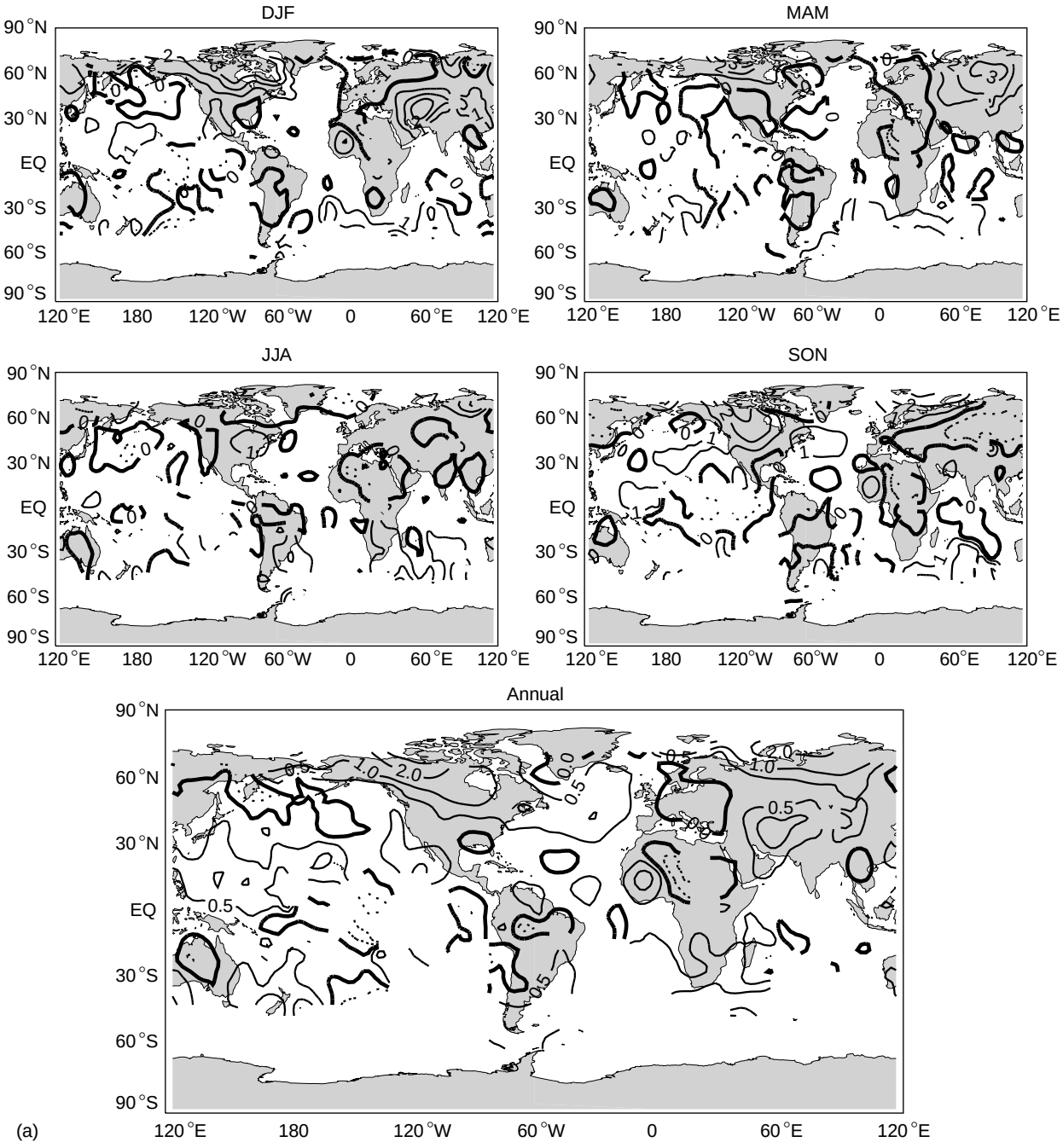


Figure 4 Trend of temperature on a seasonal and annual basis for two 20 year periods, (a) 1925–1944 and (b) 1978–1997. Regions with significant warming and cooling are shown in Jones *et al.* (1999)

this can be considered as a skewed version of a normal distribution). Each grid-box time series, on a monthly basis, is transformed from one based on anomalies with respect to 1961–1990 to one based on percentiles with respect to the same period (see for e.g., Jones *et al.*, 1998a). Using a normal distribution (i.e., simply dividing the grid-box anomaly values by the standard deviation calculated over the 1961–1990 period) works almost as well, but a gamma distribution gives a better representation because monthly temperatures are significantly negatively skewed in many regions of the world.

Figure 5(a) compares the anomaly and percentile method for displaying annual temperatures for 1998. The zero anomaly and the 50% contour are essentially the same in

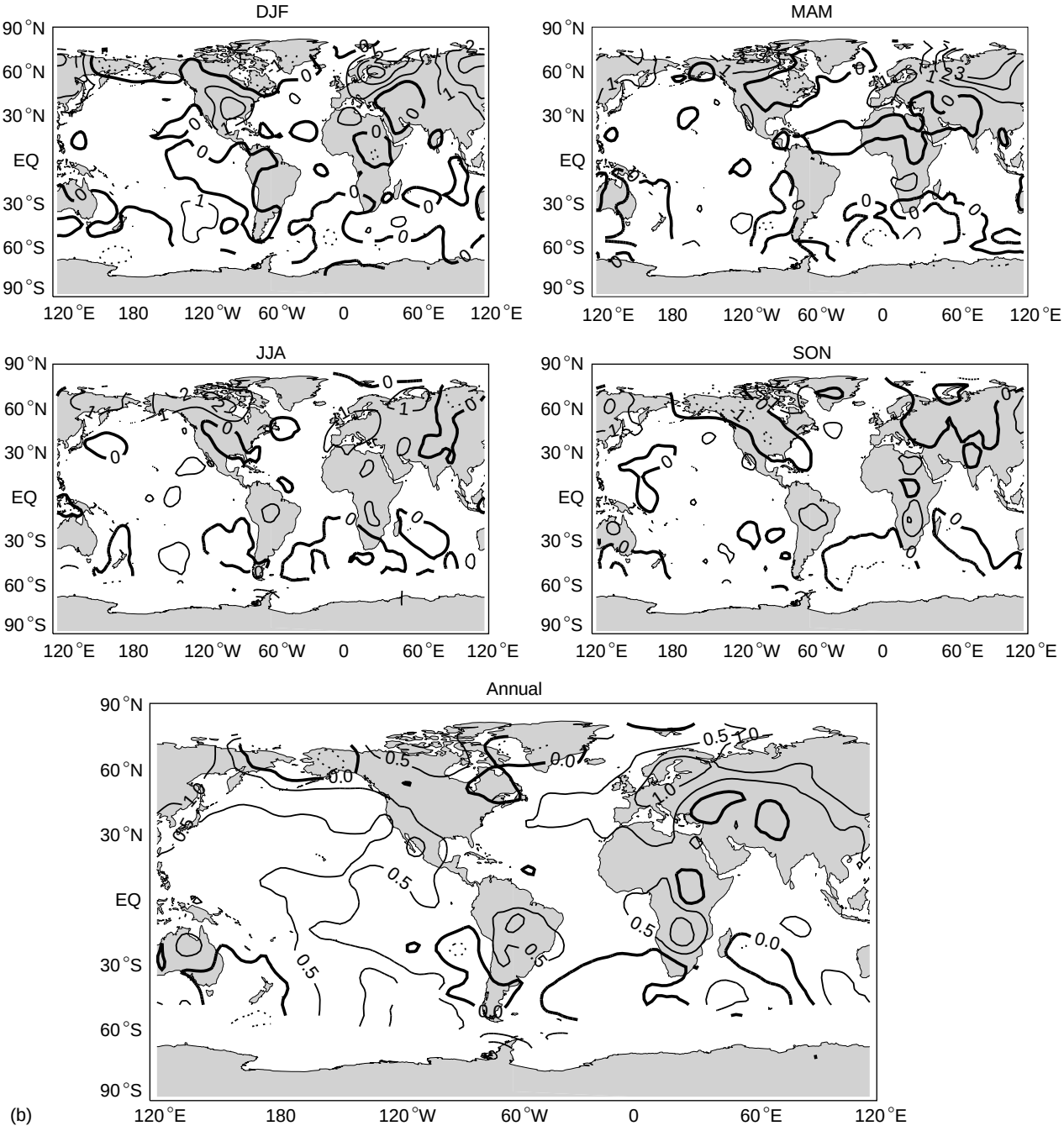


Figure 4 (*Continued*)

both plots. The percentile map indicates extreme annual temperatures over many tropical and oceanic regions that might not warrant a second glance in anomaly form.

For this year, the area in excess of the 90th percentile covers a large proportion of the globe. Figure 5(b) shows the percentage of the analyzed area of the globe with annual temperatures greater than the 90th and less than the 10th percentile since 1900. An increase in the percentage of the analyzed area with warm extremes is evident, but by far the greatest change is a reduction in the percentage of the analyzed area with cold extremes. Some caution should, however, be exercised when interpreting these results. Large changes in coverage have occurred over the 20th century (see Figure 4a) so some of the changes may be due to areas entering the analysis through time. This same problem affects the hemispheric and global anomaly

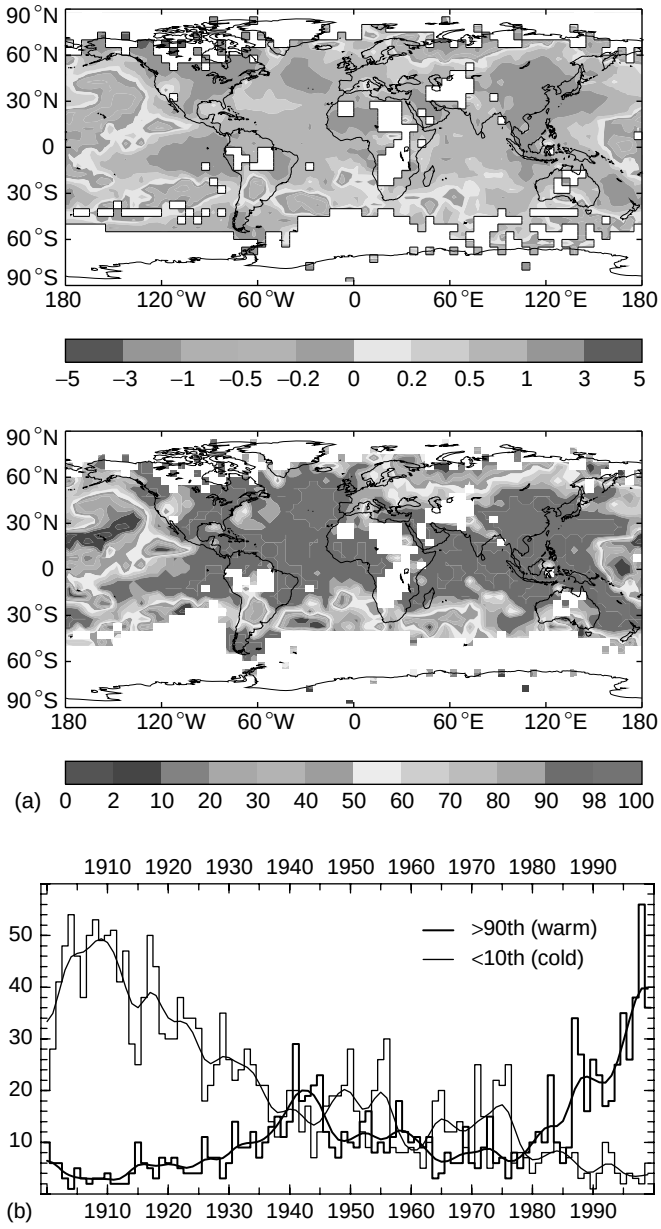


Figure 5 (a) Surface temperatures for 1998, relative to the 1961–1990 average, expressed as anomalies and as percentiles. The percentiles were defined by fitting gamma distributions to the 1961–1990 annual deviations relative to the 1961–1990 average, for all $5 \times 5^{\circ}$ boxes with at least 21 years of annual data in this period. (b) Percentage of the monitored area of the globe, for each year 1900–1999, with annual surface temperatures above the 90th percentile and below the 10th percentile

series in Figures 1 and 2, where the implicit assumption is made that the average of the unsampled regions is the same as the average of the sampled regions. Coverage changes are minimal since 1951 and reanalysis, using only those regions with data for the whole century, produces similar trends to those seen in Figure 5(b). Another problem is that during the early decades of this century grid-box values were based on fewer stations or few individual SST observations, possibly producing more extreme values in the percentile sense. Again, however, changes in the density of observations per grid box are minimal since 1951.

Trends in Continental Scale Temperatures

Another way of looking at grid-box temperature data is to derive continental-scale time series. Figure 6 shows eight regions on an annual basis, each plotted as anomalies from the 1961–1990 period. Each series has been calculated using all $5 \times 5^{\circ}$ land grid boxes in the areas defined in the plots. All eight series show long-term warming but by varying amounts. Europe and South America show a gradual warming over the 20th century, with North America, Africa and the Arctic indicating a strong warming during the first part of the century, cooling over the 1940–1975 period, followed by renewed warming. Asia and Australia show warming principally over the last 25 years. In some regions, temperatures during the 1980s and 1990s are about $1\,^{\circ}$C warmer than those prevailing during the 19th century. Only in Asia and Africa, however, are the extreme individual years clearly evident during the late 1980s and 1990s.

The series also show marked differences in year to year variability. This is partly a function of region size, but also reflects the underlying spatial variability of year to year temperatures. Greater spatial variability is evident over the polar regions and the interior parts of North America and Eurasia. Some regions (the Arctic, Africa and South America) show increased variability during the 19th century. This is not real, but due to the sparser availability of data during these years.

Trends in Maximum and Minimum Temperatures

Up to now, all the surface temperature analyses have used monthly-mean temperatures. This situation principally arises from the widespread availability of this variable. Recently, more extensive data-sets of monthly-mean maximum and minimum temperatures and their difference, the diurnal temperature range (DTR) have become available. The most extensive analysis has been undertaken by Easterling *et al.* (1997) using over 3000 station time series that encompass 54% of the world s land areas, covering the period 1950–1993. Despite the number of stations, the area covered is less than for mean temperatures. Homogeneity problems were assessed but they are more difficult to address with this type of data, requiring much more station history information than is generally available. This study confirmed some earlier analyses and showed that over the 44 years, minimum temperatures warmed by $0.79\,^{\circ}$C (linear

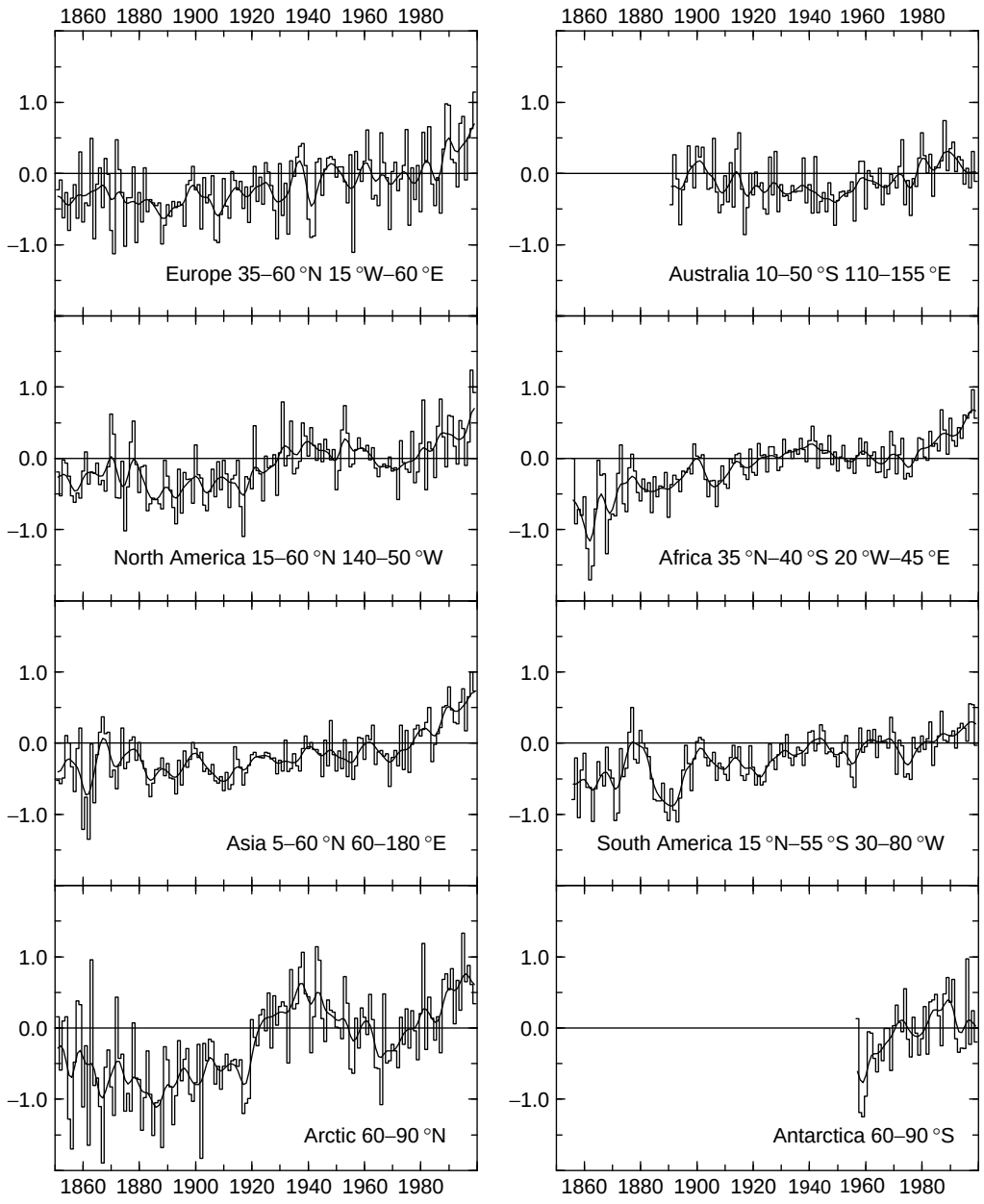


Figure 6 Continental scale temperature anomaly series on an annual basis relative to 1961–1990. The continents are defined in each of the plots. The series have been produced using only land data from the $5 \times 5^{\circ}$ grid-box dataset

trend calculated using the average global series), maximums warmed by 0.35 °C and the DTR decreased by 0.35 °C. The study also attempted to eliminate any possible influence of urbanization effects by using only non-urban stations in the analysis. Whether these trends are part of longer term changes is open to question. Longer series, back to the turn of the century, are only readily available for the United States. Analyses there shows little change over the first half of the 20th century.

Figure 7 shows the annual trends for each 5° grid box for 1950–1993, expressing the trends as °C per century. Minimum temperatures decrease in only a few areas while maximums show decreases over large areas, notably eastern Canada, southern United States, parts of Europe and much of China and Japan. The DTR shows decreases almost everywhere except Arctic Canada and some Pacific Islands. As with Figure 4, there are again large differences between neighbouring grid boxes which reflect both the short period of analysis, some missing years, and more probably, some errant station values in a few series where an individual grid-box series may be made up from a single station. This problem would be reduced if the trends were estimated from

the percentile values of the previous section, but the units would no longer be the easily understood °C.

Daily Temperature Extremes in Long European Series

The last two sections considered extremes on a monthly basis, but the public perception of climate change is often gleaned from daily extremes. A few extreme cold days in a row in a region might lead to questioning of the long-term warming, whereas a heat wave might produce the opposite reaction. In the context of the global warming issue, daily data should be relatively unimportant as detection of change and attribution to human influences is primarily focussed on underlying trends on decadal timescales. In the public and political worlds, however, changes in the frequencies of extremes are how global warming is often perceived.

In examining these data, daily temperature series present greater problems with respect to homogeneity than monthly data. Many of the homogeneity issues discussed earlier have a much greater influence on daily than on monthly values. Site and observation time changes are particularly important and may not be correctable in all cases. Few

long daily series are, therefore, totally homogeneous and the availability of series in digital form is generally restricted to the last 50 years. The longest daily series are available in Europe and a few have had extensive homogeneity analyses undertaken, enabling some confidence to be placed in derived results. The longest currently available daily record (1772 to the present) is that for Central England, which represents the average temperature for three sites (Parker *et al.*, 1992).

The public perception of extremes might be cold winter days or hot summer days, but this definition only considers the extreme seasons and necessitates a somewhat arbitrary definition of a threshold for cold and hot. A cold day threshold of 0 °C clearly has important consequences, but a warm day threshold would clearly be somewhat arbitrary. A better and universally applicable means of defining extremes is to let the data decide the threshold (Jones *et al.*, 1998a). Here an annual cycle of temperature on a daily basis is defined for the period 1961–1990.

With the smooth curve defined, the daily temperature anomalies for the 30 January 1st s for 1961–1990 are fitted to a gamma distribution, as described earlier. The gamma distribution is used because these daily anomalies also have a tendency to be negatively skewed. This distribution enables each January 1st value for the whole series to be transformed to a percentile value. Similar analyses are undertaken for the other days of the year. It is now a simple matter to count the number of days above the 90th (warm) and below the 10th (cold) percentile in a calendar year or in a season.

Figure 8 shows the annual counts of warm and cold days for the Central England series for 1772–1998. As would be expected, there is a strong correlation with annual mean temperatures. The counts show an increase in warm days over the present century, but the largest trend is in the reduction of cold days. The method is insensitive to the choice of base period in that another choice would produce similar trends, but centered around a different base value. Analyses for other long daily series in Europe (St. Petersburg, Stockholm, Uppsala, Brussels, Milan, Padua and Cadiz) show similar trends to the Central England series.

The last three sections have all looked at extremes in different ways, but all yield similar conclusions. The warming of the present century is mostly manifest, not by increases in warm extremes, but by reductions in cold extremes. These sorts of extremes often pass by unnoticed by the public. Only in sectors where cold temperatures are important have their effects been noticed. For example, pesticide use has increased because insect killing frosts are less severe, the skiing season is shorter, and freeze-up periods of some major rivers are reduced.

The Last 150 years in the Context of the Present Millennium

The global average surface temperature has clearly risen over the last 150 years (Figure 2 and Table 1), but what significance does this have when compared with changes over longer periods of time? The last 1000 years is the pre-instrumental period about which we know the most, although it must be remembered that such knowledge is of considerably poorer quality than for the post-1850 period. The last 500–1000 years is also the most important when

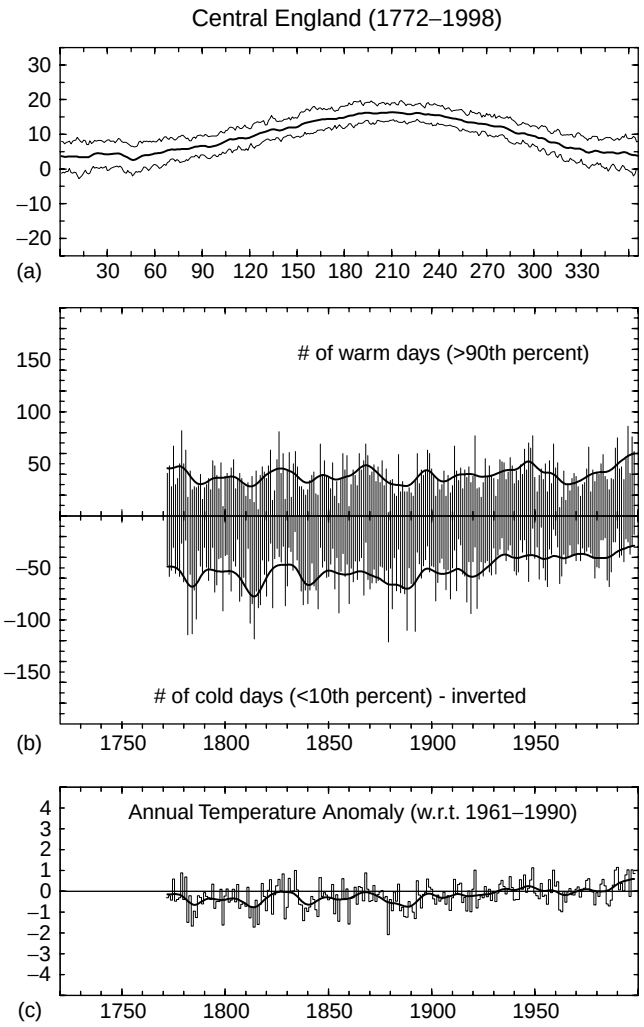


Figure 8 (a) Annual cycle of daily Central England temperatures (CET), 1961–1990. The two outer thinner lines represent the 90th (warm) and 10th (cold) percentiles. (b) Counts of days each year, 1772–1998 in CET above/below the 90th/10th percentile thresholds. (c) Annual temperature anomalies from the 1961–1990 base

considering attribution of recent changes to human influences. Earlier millennia are less important because they are known to have been influenced by changes in solar irradiance caused by changes in the Earth s orbit around the Sun (the Milankovitch effect, *see* Orbital Variations, Volume 1).

Information about the millennium comes from a variety of high-frequency proxy climatic sources (early instrumental records back to the late 17th century, written historical records, tree-ring densities and widths, ice cores, corals and some highly resolved lake and marine sediments). All these sources are discussed in other sections (but see also Jones *et al.*, 1998b and Mann *et al.*, 1998; Natural Records of Climate Change, Volume 1). Uncertainties in

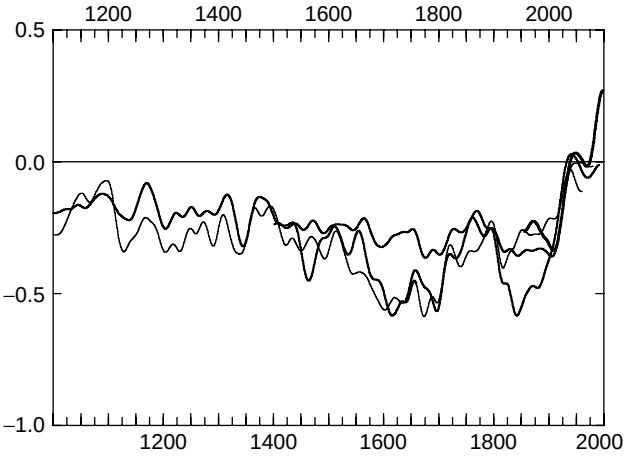


Figure 9 NH temperature reconstructions from palaeoclimatic sources. The three series are Mann *et al.* (1998) (thick), Briffa (2000) (medium) and Jones *et al.* (1998b) (thin). All three annual resolved reconstructions have been smoothed with a 50-year Gaussian filter. The fourth (thickest) line is the short annual instrumental record also smoothed in a similar manner. All series are plotted as departures from the 1961–1990 average

this information are considerably greater because evidence is restricted to limited areas where these written and natural archives survive, and more importantly, all proxy records are only imperfect records of past temperature change.

During the last five years, the number of regions from which records have been developed has dramatically increased. Figure 9 compares three different reconstructions of NH temperatures for part of the last millennium. The reconstructions are all of different seasons, so on the basis of the instrumental record would be expected to differ. The source details are given in the figure caption and the annual instrumental record is superimposed. The most striking feature of the multiproxy averages is the warming over the 20th century, for both its magnitude and duration. The agreement with the instrumental records over the last 150 years should be taken as real in that all three have, to some extent, been calibrated against instrumental data, either locally or as a whole. Agreement is poorer during the 19th century, accentuated partly by the instrumental series being annual and two of the compilations being most responsive to summer conditions. If the proxy records have any validity, then the 20th century is both the warmest of the millennium and the warming during it has been unprecedented. Four recent years (1998, 1997, 1995 and 1990) are the warmest in the instrumental series and are the warmest since 1400 and probably since 1000. The coolest year of the millennium was 1601, about 1.0 °C below the 1961–1990 average. If SEs could be assigned to these series, as was done for the instrumental series, they would be considerably larger, so it is probably very prudent to consider only the longer (decadal and century) timescale changes.

Numerous palaeoclimatic studies have considered the last millennium and two unusual periods, the *Little Ice Age* (*see* Little Ice Age, Volume 1) and the *Medieval Warm Period* (*see* Medieval Climatic Optimum, Volume 1) are often discussed. The three curves in Figure 9 all raise questions about whether, on hemispheric scales, these two periods were universally colder and warmer, respectively, than the present. This questioning arises from the significant advances in palaeoclimatology that have been made over the past 15–20 years. The series used in Figure 9 contain from ten to several hundred diverse proxy climatic series, considerably more than previously analyzed. The coldest century of the millennium, the 17th was on average only 0.5–0.8 °C below the 1961–1990 base. Centuries in the first half of the millennium were milder but still cooler than the 20th century. The differences between these later studies and earlier work is that they are based on considerable amounts of diverse (both in terms of variety and location) proxy data than used earlier. Earlier studies were restricted in location, the terms themselves being defined by early European/North Atlantic analyses, and often used more slowly responding proxies (such as glacial advances and tree lines), where the influence of temperature was generally inferential and qualitative rather than of the quantitative nature recently achieved.

ABSOLUTE VERSUS ANOMALY TEMPERATURES

Up to now, all analyses have been quoted as °C with respect to the period 1961–1990. Analyses with respect to other base periods would be expected simply to change the reference level, offsetting the results by the difference between 1961–1990 and the new period. Random differences would be expected because of missing values in both reference periods but the course of change on the decadal timescale would be unchanged. The use of anomaly values overcomes problems of using absolute temperatures discussed earlier, i.e., elevation, observation times, the varying methods used to calculate monthly-mean temperatures and screen design. Although anomaly values are all that are required, the question often asked is what is the absolute temperature of the surface of the Earth?

Recent advances in data availability and interpolation methods now enable this question to be answered. The means by which this has been achieved are given in detail in Jones *et al.* (1999). Separate analyses are undertaken for four regions:

- land areas except Antarctica;
- ocean areas except the Arctic;
- the Arctic north of 60 °N;
- the Antarctic south of 60 °S.

Land Areas Except Antarctica

For these regions, New *et al.* (1999) have developed a climatology for the period 1961–1990 using 12 092 station estimates of mean temperature. To attempt to minimize problems with the definition of mean temperature, the average of 1961–1990 values of average maximum and minimum temperatures were used wherever possible. These station averages were then interpolated as a function of latitude, longitude and elevation. The use of elevation as an explicit variable enables inclusion of the effect of spatially varying temperature lapse rates in different mountainous regions. The accuracy of interpolation was evaluated by omitting a subset of the data and comparing the results interpolated to these locations. Errors of estimation, even in data sparse regions, were between 0.5 and 1.0 °C. Values for every 1° square containing any land were then calculated using elevation data from a five minute digital elevation model (DEM). The value for each 1° square was the average of all of the 120 five-minute pixels that are land.

Ocean Areas Except the Arctic

In the anomaly dataset, SST anomalies were used because of their superior accuracy compared with MAT data. For the climatology, however, it is preferable to use MAT values as these will combine much better with the land temperatures around coasts and islands than an SST climatology. The MAT climatology, however, will be seriously affected by the solar heating on board ships (see earlier). To overcome this, an air minus sea temperature difference climatology was used, also for 1961–1990, developed using only NMAT and SST values from one hour before sunset to one hour after local sunrise. The AT climatology is the sum of the air-sea and the SST climatology.

Arctic North of 60 °N

This region is treated separately because of the lack of measurements over the sea ice regions. From the late 1930s, measurements have been made on ice islands (principally Russian), these being augmented dramatically since 1979 by automated measurements from the International Arctic Buoy Program. Although much of the buoy data comes from the 1979 to 1997 period, the network provides an extensive array from which to derive a climatology for 1979–1997. The early Russian data are particularly useful in assessing the quality (and adjusting) the buoy data during summers, when there are serious homogeneity problems with thermister housing due to the effects of direct solar radiation. The climatology for the Arctic was an amalgamation of the buoy data with poleward extensions of the land and ocean components already discussed. This aspect of the work presents challenging problems for analysts in

deriving consistent values and is discussed in detail in Jones *et al.* (1999).

Antarctica South of 60°S

This aspect of the climatology is the weakest in terms of accuracy simply because of the lack of data availability. There are less than 30 stations on the continent recording enough temperature data to derive 1961–1990 averages. These are not well located in that over half are located on the Antarctic Peninsula and five of these on one island, King George V Island. Away from the land, there are no measurements made over the sea ice regions and oceanic data are virtually negligible south of 60°S. Automatic weather station (AWS) data augment manned station records in some regions, but record lengths are short. The climatology developed for this region is, therefore, speculative with large amounts of extrapolation. It is, however, the best that can be done with available data.

Combination of the Four Climatologies

Combining the areas produces an average global annual temperature of 14.0°C for the 1961–1990 period, with the NH warmer than the SH (14.6 versus 13.4°C). Due to the problems of developing values over the Antarctic and Arctic, it is difficult to estimate a SE for these values. The 14.0°C global average is believed to be within ±0•5°C of the true value (Jones *et al.*, 1999). The magnitude of the hemispheric seasonal cycle is 13.1°C in the NH and 5.7°C in the SH. In the NH, the warmest and coldest months are July and January as expected. In the SH, January is warmest but August rather than July is the coldest. As would be expected, the global seasonal cycle is dominated by the NH.

With improvements in satellite technology, climatologies for all four areas will be improved. Over both the Arctic and, especially the Antarctic, improvements are necessary but at the present, consistent long-term monitoring of surface temperatures over snow and ice are several years away. It will then take several years to develop a new climatology.

EXPECTED IMPROVEMENTS IN THE NEXT DECADE

Improvements in both the quality and numbers of surface station data are constantly being sought. Only a small fraction (~10%) of surface data measured over land is internationally exchanged. Concerted international efforts are currently underway to improve both the land and marine components. Over land, nearly 1000 sites have been designated as the Global Climate Observing System (GCOS) Surface Network (GSN). Most National Hydrometeorological Services (NHMSs) (through The World Meteorological Organization, WMO) have agreed to attempt to maintain these stations and, where possible, the environments around them, well into the 21st century. The selection of sites aimed to utilize the best stations, improving spatial coverage in recent years, but acceding to Members who have budgetary constraints. Once the network is fully operational, it will improve the quality of the data, particularly for recent years. At present, all climatic data are essentially second-hand weather data in that most NHMSs are only interested in weather forecasting and climate is a minor consideration. GCOS aims to rectify this, bringing the importance of the data for climatic purposes to the fore. If successful, GCOS will provide a network of good stations, enabling researchers in the future to concentrate more on what the data is telling us, rather than worrying about long-term homogeneity.

Improvements are also being undertaken over marine areas. Efforts are underway to digitize much of the ships log material that are known to be available in some regions, principally for the two World War periods and the 19th century (*see* GCOS (Global Climate Observing System), Volume 1). Once available, the new data should significantly improve analyses over much of the Pacific before and during World War II and in the SH.

Since 1979, satellite estimates of SST can be used to infill gaps in shipborne measurements. Absolute values measured from space differ by 1–2°C from ship measurements, but the patterns of variation between the two observing systems agree well. Combining the two involves blending complete satellite patterns to ship measurements using the latter as the anchor for the absolute values. Modifications of this basic technique are undergoing trials for land regions. Results are encouraging for all land surfaces except snow. The challenge for the future will be to blend satellite and surface data over snow and sea ice surfaces.

The past should also not be ignored. Dramatic improvements in our understanding of the last 1000 years have recently been made. These have been achieved through both basic government funding for science and by the increased collaboration of scientists working in different palaeoclimatic disciplines. The National Geophysical Data Center in Boulder, CO, serves as a repository for much of the paleoclimatic data.

See also: Climate Change, Detection and Attribution, Volume 1; Tropospheric Temperature, Volume 1; Water Vapor: Distribution and Trends, Volume 1.

ACKNOWLEDGMENTS

P D J acknowledges the support of the US Department of Energy (Grant No. DE-FE02-98ER62601).

REFERENCES

Briffa, K R (2000) Annual Climate Variability in the Holocene: Interpreting the Message of Ancient Trees, *Quatern. Sci. Rev.*, 19, 87–105.

Easterling, D R, Horton, B, Jones, P D, Peterson, T C, Karl, T R, Parker, D E, Salinger, M J, Razuvayev, V, Plummer, N, Jamason, P, and Folland, C K (1997) Maximum and Minimum Temperature Trends for the Globe, *Science*, 277, 364–367.

Folland, C K and Parker, D E (1995) Corrections of Instrumental Biases in Historical Sea Surface Temperature Data, *Q. J. R. Meteorol. Soc.*, 121, 319–367.

Folland, C K and Salinger, M J (1995) Surface Temperature Trends in New Zealand and the Surrounding Ocean, *Int. J. Climatol.*, 15, 1195–1216.

Folland, C K, Karl, T R, and Vinnikov, K (1990) Observed Climate Variations and Change, in *Climate Change. The IPCC Scienti c Assessment*, eds J T Houghton, G J Jenkins, and J J Ephraums, Cambridge University Press, Cambridge, 195–238.

Folland, C K, Karl, T R, Nicholls, N, Nyenzi, B S, Parker, D E, and Vinnikov, K (1992) Observed Climate Variability and Change, in *The Supplementary Report to the IPCC Scienti c Assessment*, eds J T Houghton, B A Callendar, and S K Varney, Cambridge University Press, Cambridge, 135–170.

Jones, P D (1994) Hemispheric Surface Air Temperature Variations: a Reanalysis and an Update to 1993, *J. Clim.*, 7, 1794–1802.

Jones, P D, Osborn, T J, and Briffa, K R (1997) Estimating Sampling Errors in Large-scale Temperature Averages, *J. Clim.*, 10, 2548–2568.

Jones, P D, Horton, B, Folland, C K, Hulme, M, Parker, D E, and Basnett, T A (1998a) The Use of Indices to Identify Changes in Climatic Extremes, *Clim. Change*, 42, 131–149.

Jones, P D, Briffa, K R, Barnett, T P, and Tett, S F B (1998b) High-Resolution Palaeoclimatic Records for the Last Millennium: Interpretation, Integration and Comparison with General Circulation Model Control Run Temperatures, *Holocene*, 8, 455–471.

Jones, P D, New, M, Parker, D E, Martin, S, and Rigor, I G (1999) Surface Air Temperature and its Changes Over the Past 150 years, *Rev. Geophys.*, 37, 173–199.

Mann, M E, Bradley, R S, and Hughes, M K (1998) Global-scale Temperature Patterns and Climate Forcing over the Past Six Centuries, *Nature*, 392, 779–787.

New, M, Hulme, M, and Jones, P D (1999) Representing 20th Century Space-time Climate Variability, 1: Development of a 1961–90 Mean Monthly Terrestrial Climatology, *J. Clim.*, 12, 829–856.

Nicholls, N, Gruza, G V, Jouzel, J, Karl, T R, Ogallo, L A, and Parker, D E (1996) Observed Climate Variability and Change, in *The IPCC Second Scienti c Assessment*, eds J T Houghton, L G Meira Filho, B A Callander, N Harris, A Kattenberg, and K Maskell, Cambridge University Press, Cambridge, 133–192.

Parker, D E, Jones, P D, Bevan, A, and Folland, C K (1994) Interdecadal Changes of Surface Temperature since the 19th Century, *J. Geophys. Res.*, 99, 14 373–14 399.

Parker, D E, Legg, T P, and Folland, C K (1992) A New Daily Central England Temperature Series, *Int. J. Clim.*, 12, 317–342.

Peterson, T C, Easterling, D R, Karl, T R, Groisman, P Y, Nicholls, N, Plummer, N, Torok, S, Auer, I, Bohm, R, Gullett, D, Vincent, L, Heino, R, Tuomenvirta, H, Mestre, O, Szentimerey, T, Salinger, M J, F rland, E J, Hanssen-Bauer, I, Alexandersson, H, Jones, P D, and Parker, D E (1998a) Homogeneity Adjustments of *In-situ* Atmospheric Climate Data: a Review, *Int. J. Climatol.*, 18, 1495–1517.

Peterson, T C, Karl, T R, Jamason, P F, Knight, R, and Easterling, D R (1998b) The First Difference Method: Maximising Station Density for the Calculation of Long-term Global Temperature Change, *J. Geophys. Res.*, 103, 25 967–25 974.

Models of the Earth System

MICHAEL C MacCRACKEN

Lawrence Livermore National Laboratory, Livermore, CA, USA

In Earth system sciences, the term model refers, broadly interpreted, to any conceptualization of an object or system, whether mechanical or mental, physical or theoretical. Thus, a model can be a reduced-size replica of a river system, or it can be a conceptual model of the origin and evolution of the universe; it can be a representation of future climatic conditions based on analogues drawn from the past, or it can be a mathematical representation of all the processes that govern how the atmosphere, oceans (e.g., radiation or convection), and land surface work. Models are often used to connect together what is known about a system and to gain insight about how a system has worked in the past, is working now, or may work in the future. What is most important in choosing a type (or types) of model to use to gain insight is to make sure that it is well-suited to the issue being investigated: models that are too simple can leave out important processes that may only come into play over long periods, whereas models that are too complex can prevent the gaining of insight, or even provide misleading insights, because they hide key interactions in a milieu of extraneous detail.

For understanding the Earth system, a variety of types of models are being used, ranging from insights gained from study of past climates to computer-based models that represent the Earth system in terms of numbers (hence the term numerical models). Because of the scope and complexity of the Earth system and the need to represent the wide range of Earth system components and processes, a hierarchy of numerical models has been developed. Such numerical models have become so pervasive that they are often simply referred to as models. Some of these models focus on particular processes in the atmosphere and oceans (e.g., radiation or convection), some models focus on particular components of the system (e.g., the atmosphere or oceans), some models focus on particular sets of interactions within the Earth system (e.g., the carbon cycle or atmospheric chemistry), and some models focus on the overall climate system. By selecting carefully, and cross-checking across models of varying sophistication and comprehensiveness, signi cant insights have been gained about the workings of the Earth system.

INTRODUCTION

For a scientist seeking to understand the workings of the very large, complex, and intercoupled Earth system, finding a way to represent the system and its behavior using a model of that system can provide many benefits. Among the reasons models are constructed are to accomplish the following.

- *Diagnosis of prevailing conditions and empirical (experiment-based) prediction*: Models are used to provide a framework for organizing and inter-relating observations of the Earth system. As an example, databases that are constructed to store the diverse sets of observations of winds, temperatures and humidity taken at the surface, by balloons and by satellites, serve as a record for how the atmosphere is behaving. By analyzing the time evolution of such records, diagnostic studies can provide insights about how, for example, changes in particular places relate to changes over larger spatial scales. Relationships derived from this type of study, which uses the time and space history of system behavior as the model, benefit the public every day, for example, in the weather forecasts of what will happen at particular locations.
- *Explanation of causal connections*: Having available a representation of the Earth system, whether as a physical or mathematical model, allows investigation of how particular outcomes or events are caused. Although it is possible to find correlations in observational data sets, determining if there is a connecting mechanism is essential to the gaining of understanding and the development of predictive capability. For example, ocean sediment records indicate that the growth and decay of continental ice sheets over many tens of thousands of years correlate well with variations in the amount of solar radiation reaching high latitudes of the Northern

Hemisphere that are caused by cyclic variations in the Earth s orbit around the Sun. Models are being used to determine how such seemingly small solar variations over hundred-thousand year periods can cause ice-age cycling when seemingly larger variations in solar radiation over tens of thousands of years are apparently not causing such large climatic shifts. Without a model, there is no way of testing what mechanisms might explain such apparent paradoxes.

- *Prediction and projection of future conditions*: While we might want to know what the future brings, there is no way to run the full Earth forward in time – one can only live what is happening. At least some types of models, however, can be used to explore what the future might bring. Weather forecast models start with present observations and, based various on experience and mathematics, attempt to foretell how the existing weather will evolve. Other types of models, checked for relative validity against past behavior, can be used to try to look ahead months to centuries, being careful to recognize the limits of the results. Without models, whether simple or complex, the future would just come as a complete surprise.
- *Management and policy analysis*: Now that human activities have become significant enough to affect the Earth system (or at least parts of it), models can be used to explore what the consequences of various actions might be. Often, a range of possible future conditions is investigated by considering a range of scenarios for the future (*see* Prediction in the Earth Sciences, Volume 1). Such if-then experiments with these models thereby allow consideration of the potential costs and benefits of particular choices and policies, and what changes in their intensity or timing might mean, all without having to actually impose the policies or experience the outcomes. Of course, it always must be remembered that the experiments are being done using models rather than evolving in the real world, a situation that emphasizes the importance of making good choices in the selection, testing, and use of models.

With this perspective on the purposes of having models, the following sections explore the possible types of models, the approaches to numerical modeling (which is a particularly flexible and useful way of constructing models), and approaches to testing models. In that the results of models are described throughout this encyclopedia, the detailed discussion of models, their testing, and their application are treated in other articles.

GENERAL APPROACHES TO CREATING A MODEL OF THE EARTH SYSTEM

There are a variety of possible approaches to creating a model of the Earth system, each relying on creating a particular framework for encapsulating and representing what we do and do not understand about the system. Some of the approaches are physical and others mathematical; some are focused on particular processes and others on the entire Earth system. Wide ranges of factors are important to consider in selecting among the various approaches. For example, a key aspect in choosing a modeling approach is not only to create a model that can be observed, but also to create a model that can be manipulated to see how it might respond to various types of changes. For some types of models, a physical change, such as warmer temperatures, can actually be imposed; for other types of models, thought (or *gedanken*) experiments must be conducted by investigating the potential effects of a hypothetical change using mathematical or theoretical approaches. A model must also be practical, a constraint that introduces consideration of issues of size and type of problem and the resources required to use the model into the selection process. A model must also be understandable (meaning we must be able to use the model to investigate cause–effect relationships). Models are simply approximations of the real world, and so it is vital that a model be able to provide insights into how things are working in ways that can be applied to understanding not just the model, but the real world.

A number of possible ways of constructing models, broadly conceived, have been used to help provide insights about the Earth system. While each of these approaches has its strengths and shortcomings, by choosing an approach best suited to the problem being investigated, scientists have been able to use these models in actual and *gedanken* experiments to better understand how the Earth system actually works. Such experiments generally consider the present world as the control – or baseline state – and then use the various types of situations, whether past states of the system or experimentally altered states, as the perturbed situation to be compared to the control situation. For this to work, of course, there must be either the possibility of changing the conditions of the model, or having available a historical record of how conditions have changed in the past.

With this brief context, a brief overview of approaches that have been used to create models of the Earth system, or at least of some of its parts.

Laboratory Models and Field Experiments

A traditional approach to trying to understand a problem is to construct a replica, to some scale, of the process or system of interest. The Earth sciences are rich with examples of such laboratory models, ranging from chambers in which the absorption of radiation by various gases can be measured or clouds can be created, to rotating containers of fluids designed to look at how heating creates the large scale circulation of the atmosphere. A particularly interesting example has been Biosphere 2, located in Oracle,

AZ, which is one of the largest enclosed laboratories in the world. Covering about 1.27 hectares and enclosing a volume of about 200 000 m^3, Biosphere 2 is being used to study the response of several different biomes to controlled changes in atmospheric composition, temperature, and other variables.

Moving out of the laboratory, instrumentation of the atmosphere or ocean as part of field experiments has provided extensive data sets of how parts of the Earth system behave, thus using a part of the system as a model of how processes function throughout the whole system. For example, the wind drag of the surface and the effects of clouds on radiation can be studied by instrumenting a relatively small set of representative areas and then extrapolating to the world. Alternatively, the whole atmosphere at one time can be used to draw inferences about how the system will work at other times or with other conditions (e.g., global scale observations of the evolution of the weather have provided insights about warm and cold fronts, storms and droughts).

Special events also provide interesting opportunities. For example, how global average temperature changes after solar radiation is reduced by a volcanic eruption gives an indication of how the climate might respond to a change in the amount of available energy, and an eclipse of the Sun, which blocks out solar radiation reaching the stratospheric ozone layer, provides an opportunity to record the dependence of stratospheric chemistry on ultraviolet (UV) radiation. Indeed, so much can be learned from studying special locations and situations that there are major international cooperative programs of many types aimed at understanding the Earth s water cycle, atmospheric chemistry, and much more.

Although much can be learned from such approaches, the Earth system is so complex, with scales of processes extending from the microscopic, when dealing with the physics of clouds and aerosols, to the global, when dealing with the circulation of the oceans and atmosphere, that scientists have been unable to construct a sufficiently comprehensive model (or set of field experiments) that can be manipulated in appropriate ways to determine how human activities might affect the future. Thus, such models have important limitations when trying to understand the functioning of the entire Earth system, although they do provide a very useful approach for understanding particular processes or places.

Earth System History as a Model

One way to overcome the limitation of present conditions not providing enough of a range of examples for study is to use the Earth s past behavior as the model. In that the Earth is some 4.6 billion years old and geological records indicate that the Earth has experienced a wide range of climates, it can be argued that the Earth has experienced virtually all of the imaginable conditions that can occur (*see* Earth System History, Volume 1). The key shortcoming, of course, is that scientists were not there to keep accurate records of what did occur (and, most importantly, why the changes occurred). The challenge accepted by paleoclimatologists, acting as forensic scientists for the planet, has been to use the evidence of past conditions that remains in order to figure out what did happen, and why, because if we don t know why, it is hard to understand whether such events might occur in the future and it is difficult to infer what lessons they have to offer (*see* Natural Records of Climate Change, Volume 1).

All kinds of evidence have been found and are being used to generate information on past changes. Cores drilled in ice cores have found layers of ice and trapped air that are over 400 000 years old, providing a very useful record of past temperatures and atmospheric composition. The widths of tree rings are being used to get a sense of climatic conditions during the growing season (and even through the year), and scientists have now been able to overlap from tree to tree in a number of locations and reconstruct climatic conditions back several thousand years. The variations of the Earth s magnetic field captured as molten lava solidified are used to date the movement of the continents and their past latitudinal positions. The types of minerals deposited are being used to indicate where they were formed and, for example, the intensity of volcanic activity, characteristics of atmospheric composition, and, by use of radioactive isotopes, the age of the sample. The skeletons of very small animals in the oceans accumulate on the ocean floors, and their type and location, are providing information about the temperatures of the upper and deep ocean environments where they lived. An immense and growing number of such proxy indicators are being employed to gain insights into the Earth s past (*see* Natural Records of Climate Change, Volume 1).

In these ways and more, paleoclimatic science has been able to document the rich history of changes, from periods when the Earth was extensively glaciated to times when the polar regions were apparently warm enough for palm trees to grow. What we now have, albeit with many gaps and blurred images, is a rich tapestry of inferences about the types of conditions that occurred in the past. What has also become clear, however, is that the types of modifications that human activities are now forcing on the Earth system are quite unlike, in magnitude and rate, any transitions that have occurred in the past. As a consequence, while very useful for improving the Earth system, the paleoclimatic record provides only a limited basis for constructing analogues about how the future may evolve (see US/USSR, 1990 and Climate Analogues, Volume 1).

Other Celestial Bodies as Models

While the Earth is, to the best of our knowledge, unique, the rich diversity of other celestial bodies provides a very diverse set of examples of how planetary environments work and function. For example, we have learned that the surface of Venus is much hotter than the Earth. While this might not seem surprising because Venus is closer to the Sun than the Earth, this is actually not the main reason why the surface of Venus is so hot. What observations indicate is that, because of its bright clouds (which is why Venus is so visible in the night sky), Venus actually absorbs less solar energy per unit area than the Earth. Instead, Venus is warm because its atmosphere creates a greenhouse effect that is many times stronger than the Earth s. Looking at Mars, observations indicate that its surface is colder than the Earth s, again not surprising in that Mars is farther from the Sun. However, because the atmosphere of Mars contains many times more carbon dioxide than the Earth s atmosphere, the study of Mars makes clearer how important water vapor is as a greenhouse gas for the Earth.

Of even more relevance have been the lessons learned about how stars behave. Not having a long record of measurements of how solar output may have varied, astrophysicists have used information about sun-like stars to get a sense of the evolution of stars and how the intensity of the Sun s energy output has likely varied over the Earth s history. It is such studies, using other stars as models, that have provided indications that our Sun is becoming brighter over time, and was a considerably dimmer in much earlier times (raising a question about whether the Earth was ever ice-covered, and, if not, why).

The Recent Past as a Model for the Future

The similarity of the weather from one day to the next and of the climate from one year to the next can lead to a notion that the recent past might be a very good model for the future. For example, those living in tropical and subtropical areas often advertise the persistence of their weather as a reason to come vacation and walk along their beaches. Indeed, an important test of the skill of weather forecast models has been to determine if they can show more skill than persistence, and if predictions of seasonal climate fluctuations are skillful relative to climatology, which is simply the average of the past.

For the case of climate change, using the recent past as a model might suggest using the trend that has been observed over the 20th century as a basis for predictions that the next century will experience the same amount of change. When actually trying to use the past as a model for the future, a number of problems arise, however. First, how far back into the past should be used to generate a trend line? That we have warm winters and cold winters intermittently, and seemingly at random, suggests that we need to go back at least a couple of decades to try to get a smooth trend. But once we go back a few decades, we start to find that there have also been multi-decadal fluctuations, with warm conditions in the 1940s followed by cooler conditions in the 1960s (at least in the Northern Hemisphere) and then warmer conditions thereafter. If we go back further, we start to find multi-centennial fluctuations and then, even further back, we find cycles of glacial advance and retreat. In addition, it is virtually certain that human activities that affect atmospheric composition will be changing in the future, rapidly creating new types of conditions.

While surely important to consider as one means of gaining insight about the future, extrapolating from the past into the future raises a great many complications, and ignores the insights that can be gained by understanding how the Earth system works – whether of the cold front approaching on a spring day or the warming likely to be caused by the changing of atmospheric composition.

Mathematical Analysis

In addition to direct experimentation, a traditional approach to understanding the workings of a system is to formulate a set of mathematical equations (e.g., algebraic or differential equations) that describe its behavior, and then to simply solve the equations for the unknown term as a function of other terms that are known or can be measured. For example, measurements led to development of the Stefan-Boltzman law, which uses a proportionality constant (conventionally termed sigma) to relate the amount of infrared heat radiated by a body (such as the Earth) to the fourth power of the average temperature (T) of the body. Because the amount of heat radiated by the Earth equals the amount absorbed, assuming conditions have prevailed long enough for the climate to reach a stable state, the equations can be used to determine the Earth s average temperature if the amount of solar energy absorbed is known. The amount of solar energy absorbed is quite well established to be the intensity of solar radiation ($S \cdot 1368\,W\,m^{\cdot 2}$) divided by four to spread it around the sphere and multiplied by the co-albedo (which is the fraction absorbed, being 1 minus the amount reflected, which is about 30%). Solving for the temperature and filling in the values gives an average planetary radiating temperature of 255 K (or about $\cdot 18\,^{\circ}C$). This temperature is not the temperature at the surface, but is instead the average temperature about 5–6 km above the surface, which is the altitude that, because of the greenhouse gases, is the average height from which radiation is emitted to space (*see* Energy Balance and Climate, Volume 1). This analysis also starts to hint at the difficulty of this approach; namely that, while much can be learned from mathematical analyses of some aspects of the Earth system, there are in reality so many equations and relationships that

it is not possible to solve them analytically except for very special limiting cases. Hence, this mathematical approach is useful for gaining broad insights, but not adequate for understanding the full system without introducing significant complexities.

Numerical Models

To overcome the difficulty of not being able to make a physical model of the system in the laboratory, and the difficulty of not being able to solve the sets of equations representing the global system, resort is often made to creating a model designed to simulate (or represent) the Earth using a computer. In these models, the spatial domain is divided into grid cells that, like tiles, cover the domain (e.g., cover the Earth). Grid cells are then stacked up to represent the various layers of the atmosphere or ocean. Numbers are then used to represent conditions at particular points, and equations are used to represent how the values of various parameters in each grid cell change with time. Because the equations are often set up for a finite grid of cells, and exchanges between these grid cells are based on differences in their conditions, the resulting numerical models are also sometimes called finite difference models. In addition, because the total set of equations often numbers in the thousands to millions, such models are also often referred to as computer models because only a powerful computer can keep track of all the numbers and relationships.

One of the first efforts to actually construct such a model of the atmosphere was made by Lewis Fry Richardson. Lacking a computer, but having much time while serving as an ambulance driver during World War I, he tried to actually demonstrate that a numerical model of the European weather could be constructed by doing all the calculations himself by hand. While the effort failed to yield a successful forecast, it did set the stage for improved approaches several decades later. For further information on this effort (*see* Richardson, Lewis Fry, Volume 1).

With these sets of equations representing conditions in many grid cells over a wide domain, one does not really try to solve the equations in an algebraic sense to derive a parametric relationship that can be evaluated by just filling in the numbers. Rather, with equations defining relationships for particular parts of the system, these numerical models seek to use the equations and the numbers representing the system with the expectation that the modeled system will evolve in essentially the same way that the real system evolves. That is, an equation is not solved that gives, for example, the strength of the jet stream in an analytic sense; instead the numerical model calculates or simulates the changes that lead to the evolution of the atmosphere (and jet stream) in a series of small time steps. The key point is that features like the jet stream emerge from the calculations just as they emerge from the motions of the real atmosphere; such features are not imposed explicitly or independently in the model s structure. The sequence of the computed atmospheric changes, if the model is a good one, will evolve in a manner that cannot be distinguished, in a statistical sense, from how the jet stream (and everything else) is changing in the real world.

Numerical models now exist for a wide variety of systems, from chemistry in the atmosphere to how the oceans and atmosphere move and interact. These models are designed to incorporate the quantitative relationships that we know are applicable, and as such, they potentially provide an objective, quantitative representation of all that we know. What is of most value about numerical models is that experiments can be done with them by simply changing numbers or relationships – not having to actually change the Earth or atmosphere.

Although a wide range of types of models have been developed and tested, as highlighted more extensively in the next section, there are also limits to the utility of numerical models. These arise, for example, because the Earth is an exceedingly complex system. As a result, models cannot represent every possible process that plays a role (from the microscopic to the planetary). In addition, even the largest computers are too limited at present to allow representation of all scales that are recognized to be important. In addition, the Earth system is not a fully deterministic system; just as for the real world, there is a range of possible outcomes from a model (*see* Chaos and Predictability, Volume 1). Thus, while models can provide very useful indications about how the system works and, of particular importance, how the Earth may be affected by changes in various conditions such as atmospheric composition, such models are not really the Earth, and so there are necessarily uncertainties involved in applying them and interpreting their results (e.g., *see* Model Simulations of Present and Historical Climates, Volume 1).

APPROACHES TO NUMERICAL MODELING OF THE EARTH SYSTEM

While the development of numerical models of the Earth system provides an important tool that can be probed and experimented with to learn about how the system functions, having available only a model as complex as the entire Earth system would simply provide a new challenge to understanding in that all the same complexities would be present. Although we would be able to make changes, such as in atmospheric composition, that we could not make in the real world, and this would help give a better sense of the sensitivity of the Earth system to changes and perturbations, the complexity of the Earth system would still be present, and would likely cloud our ability to deduce the most important couplings and relationships. Thus, just as in many other fields, the key to enhancing understanding is best

promoted by having available an array of approaches, each suitable to particular issues or parts of the problem.

In the field of Earth system studies, this is accomplished by construction of a hierarchy of numerical models for simulating the processes critical to particular issues. The hierarchy is such that it ranges from simple, conceptual models to fully comprehensive models based, for example, on the choice of the number of spatial dimensions treated, the detail of the time calculation, and the complexity and comprehensiveness of the processes that are represented. In seeking to understand a particular process or aspect of the Earth system, looking at how a process works in a simple model can lay the basis for understanding how the process may play a role when intercoupled with other processes, and finding an unexplained variation in a comprehensive model can be explored using simple models to determine which factors are playing a dominant role.

Throughout this Encyclopedia and throughout the field of Earth system studies, the term model is used in many ways, sometimes rather seamlessly moving across the following hierarchy of numerical models. The examples below provide a sampling of the various possibilities along the way.

Zero-dimension Models

For some questions about the Earth system, what is most important is not the spatial variation, but the detailed interactions happening at a particular location or within a particular parcel of air, land, or ocean. A good example is the investigation of basic mechanisms relating to stratospheric chemistry. The air in the stratosphere is quite layered because temperature increases with height in the stratosphere, creating quite stable conditions where chemistry can act in the absence of major influences by transport from other layers of the atmosphere. In addition, for each layer, north–south transport is relatively slow compared to the rate of chemical reactions. Thus, it is useful to assume that air within a given layer at a particular place reacts chemically without mixing significantly. Under such circumstances, and for certain types of questions, a model can be constructed that devotes attention to the hundreds of chemical reactions that are occurring, while assuming that there will not be significant influences from transport of air from outside the box being looked at. It was in this way that some of the initial studies were done about how various chemicals called chlorofluorocarbons (CFCs) (like Freon) might catalyze the depletion of stratospheric ozone (*see* Depletion of Stratospheric Ozone, Volume 1). While providing much useful insight, an obvious limitation is that over time the air does actually mix, and changes in ozone at different levels of the atmosphere will obviously affect the amount of solar UV radiation getting through to that layer. Thus, for more integrated estimates, additional aspects of the Earth system must be considered.

Other examples of zero-dimension models include: (a) early models of photochemical smog, which treated an appropriate set of chemical reactions in a volume of air moving from the city center to outlying areas; (b) models of areas of land on which various plants and trees develop; and (c) models of cloud microphysics in which vapor condenses and droplets combine to create cloudy air and then precipitation. For all such models, focusing on the details of mechanisms while ignoring the effects of how the environment is affected by spatial variations and transport of materials allows inquiry without unnecessary complexity.

Vertical-dimension Models

The thickness of the Earth s atmosphere – at least that part of it critical to the climate system – extends upward for only about 50 km. While the surface climate might be roughly the same for two locations about 50 km apart, the differences in conditions going by this distance from the surface are very dramatic, with temperature, composition, and pressure changing most markedly. For this reason, it is very useful to use vertical models to study certain aspects of the Earth system. There have been three major types of such models, each of which provides particular types of insight.

Radiation Transport Models

Radiation transport models are used to study the fluxes of solar and infrared radiation through the atmosphere. These models are designed to treat the penetration, reflection, and absorption of solar radiation as it enters the atmosphere, strikes clouds, is partially absorbed by atmospheric water vapor and other gases and aerosols, reaches the surface, and is reflected to space. The energy absorbed is then radiated as infrared radiation upward from the surface and upward and downward from the atmospheric gases and clouds. Svante Arrhenius did a mathematical analysis of this type in 1896 (*see* Arrhenius, Svante, Volume 1), in which he calculated the climate s response to a hypothetical doubling of the atmospheric carbon dioxide (CO_2) concentration. The calculation indicated that the global average surface temperature would rise by 4–6 °C. Since that time, very detailed radiation transport models have been developed that treat in detail the spectral dependencies of the various gases (*see* Infrared Radiation, Volume 1). When detailed information is used to define temperatures and gas composition based on observations at a particular location, the model results can be carefully checked against observations. Such models can also be tested against laboratory experiments where radiation at particular wavelengths is passed through a gas of a given concentration. Thus, such models can be carefully validated, and they provide significant information about the relative roles of various greenhouse gases in the

energy balance of the atmosphere (Ellingson and Fouquart, 1991). Simplified versions can also be constructed that can be used as components in other models that represent more processes and simultaneously treat more locations.

Radiative–Convective Models (RCMs)

Radiative–convective models are models of the Earth s temperature that start at the surface and extend up into the stratosphere. These models start from the radiation models just described (note that Regional Climate Models also, confusingly, have the same acronym). However, because the radiation models do not account for all of the necessary processes controlling the addition and removal of energy and its movement up and down through the atmosphere, they cannot be used to calculate temperature change. To account for the warming of the surface and then the upward movement of energy, a representation is included in RCMs of the process of convection that accounts for vertical motions of the air and causes, for example, the formation of clouds. Although convection is often represented in a somewhat *ad hoc* fashion, the effect of convection is to redistribute energy upward through the troposphere. This has the effect of creating and sustaining the observed atmospheric lapse rate (which is the vertical decrease in temperature with altitude, and is observed to be about $\bullet$ 6$\bullet$5 $^{\circ}$C km$^{\bullet 1}$). The first of these models was developed by Syukuro Manabe and Richard Wetherald at the NOAA Geophysical Fluid Dynamics Laboratory (*see* Manabe, Syukuro, Volume 1). The model was used in 1967 to simulate the vertical variation of atmospheric temperature. Through its representation of convective adjustment, the model quite accurately represented surface, tropospheric, and stratospheric temperatures (Manabe and Wetherald, 1967). The model was readily solved via computer, and could be used for many types of experiments. For example, with this model, a doubling of the atmospheric CO_2 concentrations was calculated to cause a global average warming of about 2.4 $^{\circ}$C.

Another well-known application of such models was to estimate the response of surface temperature if a major nuclear war were to inject very large amounts of smoke into the upper troposphere (Turco *et al.*, 1983). These simulations made clear that if solar radiation was absorbed in the atmospheric layers above the greenhouse gases (particularly at levels above the atmospheric water vapor), the greenhouse effect would be disrupted. As a result, for a global smoke cover, the land surface temperature would drop toward the radiating temperature of the Earth (so toward $\bullet$ 18 $^{\circ}$C), which would freeze the entire surface of the Earth. While an interesting insight, what quickly became obvious was that the model was treating neither the large heat capacity of the oceans (which it would take a long time to cool) nor the many areas of the Earth that would not be covered with smoke were a postulated war to take place. Thus, while a starting point, more needed to be done to get a handle on that issue (see Pittock *et al.*, 1986).

Overall, the strength of RCMs is that they can quite accurately represent atmospheric radiation. The key shortcomings are that such atmospheric features as clouds have to be very simply represented (even specified) and that there is no representation of the horizontal distribution of climate from north to south or east to west.

Atmospheric Chemistry Models

Atmospheric chemistry models that treat only the vertical variations in atmospheric composition have also proven to be very useful. These models add chemistry to the set of processes treated in RCMs. While about 99% of the atmosphere is made up of nitrogen and oxygen molecules, the rest of the atmosphere is made up of hundreds of other species. Although present in very low concentrations, these species can be highly reactive. The chemistry that goes on controls, for example, how much ozone is present (e.g., *see* Stratosphere, Chemistry, Volume 1). This in turn affects other parts of the system, because when present in the stratosphere, ozone protects the surface from high-energy UV radiation, and, when present in the boundary layer, ozone contributes to air pollution and affects the chemistry of the rain that falls (so controlling acid deposition and deposition of various nutrients). These models have been used, for example, to study the effects of various human activities on the stratospheric ozone layer, including the effects of CFCs, nuclear war, rocket launches, and so on (e.g., *see* Depletion of Stratospheric Ozone, Volume 1). Such models are very useful at getting a first estimate of how important various perturbations might be, but must be complemented by studies that can also treat the variations with latitude of the solar radiation and temperatures that strongly affect chemical reaction rates.

Horizontal-dimension Models

In contrast to models that focus on variations in the vertical dimension, a class of models, most notably those referred to as energy balance models, focuses on representing the processes that create variations across the surface of the Earth in the horizontal dimension. These models represent and attempt to explain, for example, the geographic variations in surface temperature. Such models thus seek to include the processes that make the equatorial regions warm and polar regions cold. While at least some of the processes creating these differences do depend on the vertical structure of the atmosphere, these models represent (or parameterize) the effects of the complexities of atmospheric processes in highly simplified ways as a way of reducing the computational burden, thereby allowing the models to run for very long periods of time. This ability to simulate very long periods in turn allows study of problems dealing with the long-term variation of the climate in response

to perturbations (e.g., the slow increase in the output of the Sun over the Earth s history, the effect of changing the intensity of the greenhouse effect, etc.). There are two main variants of such energy balance models: energy balance models that consider only latitudinal variation, and energy balance models that also treat the longitudinal dimension.

Energy Balance Models

Energy balance models that treat only the latitudinal variation were first developed independently in the late 1960s by Mikhail Budyko (*see* Budyko, Mikhail Ivanovich, Volume 1) and by William Sellers. This type of model is based on the physical requirement that energy be conserved (that is, sources minus sinks of energy determine the amount of energy accumulated at any given point). In implementing this fundamental conservation law, the early models calculated the surface temperature for various latitude bands based on the influences of solar and infrared radiation, evaporation, heat transport to the air, and a parameterization of the effects of the north–south transport of energy. Energy transport, including both sensible heat (i.e., transport of warm or cool air) and latent heat (i.e., transport of water vapor that could condense and give off heat in the process of precipitation) was represented in a combined way using an approximation drawn from the treatment of heat in a solid – namely a diffusion based algorithm with the diffusion coefficient determined based on fitting the model results to observations. The models typically included specifications for cloud cover, related infrared radiation to temperature in a parametric way, and accounted for the albedo effect of snow and other factors by changing surface reflectivity based on temperature. A surface heat capacity was created to account for the effects of oceanic and subsurface heat storage. Among the most interesting studies with these energy balance models are studies of whether glaciation exceeding a certain amount (i.e., reaching equatorward of a certain latitude) might trigger a transition to a snow-covered Earth and whether, if this occurred, such a situation would be stable and the Earth would remain frozen. That this might be possible was one of the spurs to development and application of models that represented the various controlling processes in more detail. For example, Hyde *et al.* (2000) extend this work with higher dimensional models (*see* Earth System History, Volume 1).

Over the past several decades, representation of the processes in energy balance models has continued to improve (e.g., including the effects of ocean currents and ocean overturning) (*see* Energy Balance Climate Models, Volume 1). Because it is very computer intensive to simulate the climatic effects of every emissions scenario that is suggested, such energy balance models are now used by the Intergovernmental Panel on Climate Change (IPCC, 2001) for this purpose. In using the models in this way, the parameters of the energy balance model are first calibrated to the set of simulations carried out with the full global models described below, and then the energy balance models are used, in essence, as smart interpolation tools to calculate the effects on the climate of a much wider set of emission scenarios than can be treated by the full global models (IPCC, 2001). In addition, because these models can be run in minutes on computers, they can be used for a wide variety of studies of the effects of various changes in factors that may affect the climate.

Two-dimensional Energy Balance Models

Latitude–longitude energy balance models represent not only the large horizontal temperature gradients from pole to equator, but also the significant variations in climate in the east–west direction caused by the land–ocean distribution. These models still greatly simplify the representation of processes in the vertical dimension, but do attempt to represent the time-averaged effects of atmospheric and oceanic processes that transport heat and water vapor north–south and east–west. These transports have been approximated in various ways, in some models by using an eddy diffusion approximation and in other models using an approach based on the wave patterns evident in the global atmospheric circulation. Among the interesting studies with these two-dimensional energy balance models have been investigations of how variations and fluctuations in the transport of energy (e.g., based on observed fluctuations in the atmosphere) might affect the global distribution of temperature, and study of how the distribution of continents and other large-scale features of the Earth might have contributed to past changes in the climate. More recently, models of this general type have been used with some success to forecast the evolution of the El Niño/Southern Oscillation (ENSO).

While horizontal dimension models are generally able to represent the broad scale horizontal distribution of climate (which is to be expected in that they are fundamentally calibrated to do so), what becomes apparent is that it is very difficult to build confidence in highly simplified representations of many of the processes occurring in the vertical. Therefore, while very useful for many exploratory studies, studies with energy balance models need to be complemented with studies using models that more fully represent various processes determining the vertical structure of the atmosphere and oceans.

Land Surface Models

Land surface models are another type of horizontal dimension model. These models have been developed to represent the processes determining vegetation, water resources, and other features of the land surface. In these models, atmospheric conditions, including temperature and precipitation, are imposed based on observations or outputs from other models. Representations of various landscape types,

from forests to deserts, are included to varying degrees of complexity. Some biogeographic models use empirically derived relationships about the type of vegetation typically present within certain temperature and precipitation ranges while others use parameterizations that actually follow the biological and biochemical growth of plants of various types. Biochemical processes are often treated in order to address such issues as carbon storage and the effects of atmospheric deposition. Such models often treat hydrologic processes, including flow of waters, accumulation and melting of snow, soil moisture, runoff, and other factors controlling the timing and amount of water resources. Conditions leading to fires and other disturbances can also be calculated. Comparisons among these types of models have been undertaken as a means of gaining insight about the various simulations of the potential consequences of climate change. A number of groups report significant progress in representing continental scale vegetation distributions (e.g., see Vegetation/Ecosystem Modeling and Analysis Project Members, 1995; Neilson *et al.*, 1998).

Zonal Models

As a step toward modeling the full system, zonal models represent both altitude and the north–south dimension (roughly equivalent to representing the complexity of an orange by treating one section). Such models are able to represent the variation and complexity of processes in the vertical from the surface to, typically, the stratosphere (with a vertical resolution of one to a few kilometers), and to represent latitudinal variations from the poles to the equator (with a resolution, typically, of about 10° of latitude, or about 1100 km). In constructing such models, the conservation laws for mass (of air, species, water, etc.) and of energy are generally quite well represented, whereas the processes determining the wind fields and their effects are parameterized based on observations, or simplifications of the laws for fluid flow and conservation of momentum. Such models have been constructed for both the climate system (atmosphere, oceans, land surface) and to represent stratospheric chemistry in that both, to a large degree, vary to first order primarily with latitude and height rather than with longitude.

Zonal Climate Models

Zonal climate models were first developed in the late 1960s and used to study issues ranging from the causes of ice ages to the potential effects of volcanoes, changes in atmospheric composition, and smoke from nuclear war on the climate (MacCracken and Ghan, 1988). These models have generally focused most on treatment of atmospheric radiation and the factors that affect it (including changes in atmospheric composition), and on surface processes that affect the storage of energy and surface temperature. Because this type of model does not fully calculate atmospheric and oceanic circulation, these models can typically be used to simulate periods from decades to many thousands of years. Present uses for such models, which typically take tens to hundreds of hours of computer time, include study of glacial–interglacial cycling (*see* Climate Model Simulations of the Geological Past, Volume 1) and simulating the potential consequences of changes in atmospheric composition (e.g., Prinn *et al.*, 1999).

Zonal Chemistry Models

Zonal chemistry models have proven very useful tools for studying atmospheric chemistry since first being developed in the early 1980s. This type of model typically extends from the surface up about 40–50 km and represents the global scale chemistry of both the stratosphere and troposphere. Large sets of species and chemical interactions are also included in these models. Such models have been used, for example, to simulate the effects of CFCs on the stratospheric ozone layer (*see* Depletion of Stratospheric Ozone, Volume 1), the processes governing the Antarctic ozone hole, and the effects of volcanic eruptions on stratospheric chemistry. Prinn *et al.* (1999) also describe how zonal chemistry and climate models are being coupled to models of the global economy.

Although zonal climate and chemistry models represent many important processes in impressive detail, they are unable to represent the geographical distribution of the climate, which is essential for many types of studies about potential climate impacts. In addition, the simplified treatments of horizontal transport processes, which are likely to be affected by the very perturbations that these models are often used to study, create uncertainties about the latitudinal details of the simulations.

Three-dimensional Global Models

While it might seem obvious that it is essential to treat the full three-dimensional complexity of the atmosphere–ocean–land system, this has proven to be a very significant challenge in practice because the world is so extensive and involves so many different processes. To represent the full range of processes and domains, a grid is created that covers the Earth from deep in the ocean to the surface and then well up into the stratosphere. In the atmosphere, for example, a grid resolution of 4° of latitude by 5° of longitude (which would provide, on average, information every 400 km around the Earth, or roughly one point of information for an area the size of the state of Colorado in the US) requires a total of over 3200 grid cells spread out to cover the Earth s surface. To adequately resolve the variations in temperature, clouds, and chemistry in the vertical typically requires a resolution

of 2–3 km up to an altitude of about 50 km, so a total of about 20 levels, creating a total of roughly 64 000 grid cells in the atmosphere. To represent the more detailed aspects of the weather such as storm fronts and regional variations in mountainous regions, much finer resolution is needed, thereby requiring use of many times this number of grid boxes.

The conditions in each of the grid boxes and the exchanges between the grid boxes are calculated using approximations of the fundamental conservation laws of mass, energy, and momentum (or motion). As an example, for each grid box in the atmosphere, these equations include terms to treat a wide range of processes and characteristics (*see* Climate Feedbacks, Volume 1; Earth System Processes, Volume 1), including, for example:

- solar radiation (with an appropriate set of spectral intervals covering UV, visible, and near infrared wavelengths and a range of absorbers and scatterers of radiation including gases, aerosols, and clouds);
- infrared radiation (with an appropriate set of spectral intervals to allow treatment of the various absorbing gases such as water vapor, CO_2, ozone, etc.);
- vertical transport (with treatments of both large-scale upward motions and the smaller scale convective motions in thermals and clouds, which can themselves create clouds);
- horizontal transport (with treatments of both the large scale features and the disruptions caused by the presence of mountain ranges);
- hydrologic interactions (with treatments of the processes governing the sources, movement, and removal of water and its transition from vapor to liquid to solid form);
- composition and chemistry (with treatments of processes that add, move, transform, and remove substances including ozone, carbon dioxide, aerosols, and, to adequately treat chemistry and climate feedbacks, many other species);
- land surface exchanges and processes (with treatments of reflection, absorption and emission of radiation, evaporation, thermal heating of the atmosphere, emission and deposition of various chemical species, land-ocean differences, mountain ranges, etc.).

Treatment of the oceans requires a similar set of grid boxes and processes. Because oceans are more layered than the atmosphere, layers must be thinner; also because oceanic eddies and motions (like the Gulf Stream) are smaller than in the atmosphere, horizontal resolution must be finer. Oceanic processes also introduce new types of terms, including salinity, biochemical processes, sea ice formation, and more.

For the land surface, spatial resolution is a particularly important consideration. While it may be adequate for grid cells in the atmosphere to be the size of a US state (e.g., Colorado in the early global models), this is generally insufficiently fine scale to represent surface conditions (e.g., in Colorado from the edge of the Great Plains to the middle of the Rocky Mountains and all the vegetation and climatic zones in between). Were the number of grid points not a limitation, the desired resolution for the surface might be 10 km or less. In practice, however, the best resolution being achieved in global climate models that can be run in reasonable time on available supercomputers is typically 10–30 times this size. Treatment of the land grid points also requires treatment of additional processes, including soil moisture, runoff, snow and ice build-up and retreat, growth and decay of vegetation, uptake, release, and transformation of various chemicals, etc.

An additional challenge is to represent the evolution of all of the processes and domains for times from days to millennia. This requires keeping close track of the fluxes of substance, energy, and momentum from one grid box to another at each increment in time. Because the time step must be such that transport of a quantity from one box to another does not pass through a box before being recorded (e.g., for the wind to carry moisture in one side of a grid box and out the other), time steps for the calculations must be limited. This has typically meant that the time steps of the calculations can only be fractions of an hour, even with approximations that filter out sound waves and other high velocity waves that are solutions to the set of flow equations, but not important in determining the climate. This can mean that models must take a few tens of thousands of time steps each year, or a few million time steps to simulate a century.

As a result of this computational burden, approximations have to be made, even when using the fastest supercomputers in the world. Given available computer resources and the urgency of the calculation, each modeling group has made what it deems to be the optimal approximations and simplifications regarding spatial resolution, the set of processes, and the lengths of the simulations. So that simulations can be completed in reasonable times, the choice of approximations is usually such that the models run of order a few hundred to a few thousand times as fast as real time. Thus, a few century simulation might take a few months of time to complete.

Several types of models have emerged, each aimed at addressing various types of problems.

General Circulation Models (GCMs)

General circulation models have been constructed that focus on the dynamics and movement of the atmosphere GCMs (AGCMs) or the ocean GCMs (OGCMs); see Nihoul (1985), Washington and Parkinson (1986), Schlesinger (1988), Mote and O Neill (2000); and www-pcmdi.llnl.gov, and General Circulation Models (GCMs), Volume 1. For

the atmosphere, simplified models of this type (called quasi-geostrophic models) were first developed in the late 1940s and early 1950s and were used in early attempts to forecast the weather. These models were followed in the early 1960s by AGCMs that were able to simulate weather-like features over periods of seasons to years and, when initialized by observed weather conditions, were used to forecast the evolution of the weather out to a few days. Over the past several decades, significant improvements have been made and AGCMs are now used to assimilate the variety of atmospheric observations taken at the surface, aloft, and from satellites and to forecast the weather out to approximately a week with useful skill. Because AGCMs used for weather forecasting typically only have to simulate for weeks of time, computer resources are typically used to refine resolution and to run ensembles of simulations (in an attempt to overcome the chaotic nature of the weather). For climate studies, however, AGCMs are typically used to simulate periods of many decades to many centuries. Initial studies in the 1970s, for example, evaluated the effects of the climate of a doubling of the atmospheric CO_2 concentration. AGCMs, now more fully developed, serve as the atmospheric component of Earth system models (ESMs) (see section on ESMs herein).

Similarly, OGCMs have evolved from their initial development in the 1960s and are now able to simulate many of ocean currents and features of the global ocean. Because forecasting of changes in the oceans requires having knowledge of the current state of the oceans, extensive efforts are being made using buoys, satellites, and other means to develop data sets on the conditions of the ocean surface and deeper waters. These models now also serve as the oceanic component of ESMs (see section on ESMs herein).

Global Chemical Models

Global chemical (or chemistry) models are sometimes also called global biogeochemical models because they treat the global cycles of various species that interact with the biosphere (*see* Biogeochemical Cycle, Volume 2). These models are developed based on representations of the processes affecting composition of the atmosphere and the oceans and the various cycles of particular compounds such as carbon, ozone, etc. Such models generally focus on treating the atmospheric chemistry and surface processes that emit, transform, or sequester various species and, rather than calculating the wind fields, use wind fields derived from observations or simulations with AGCMs. Global chemical models are typically used to study the build-up of carbon in the atmosphere and its uptake by surface vegetation, the chemistry of the troposphere (*see* Troposphere, Ozone Chemistry, Volume 1; Aerosols, Troposphere, Volume 1; Carbon Cycle, Volume 2), and the characteristics and concentrations of atmospheric aerosols.

Earth System Models (ESMs)

Earth system models are designed to provide full treatment of the Earth system, in a combined and interactive way (Ojima, 1992). These models are often referred to as global climate models (for which, confusingly, GCM is also the acronym). ESMs are just starting to be developed and tested that treat atmospheric dynamics and chemistry, oceanic transport and the carbon cycle, and the land surface with its ecological and snow–ice dynamics. A key challenge is in deciding which processes to include, and whether the set is sufficiently complete for the problem being investigated. For example, an initial study by the Hadley Centre, which is a climate modeling center established by the UK Meteorological Office, has connected atmosphere, ocean, and land/vegetation models and been able to simulate the pre-industrial CO_2 concentration and climate without arbitrarily fixing key parameters (Cox *et al.*, 2000). This model has been used to simulate the effects on climate and vegetation of the human-induced buildup of CO_2 in the atmosphere. Among the interesting features of the simulation is the projected drying up of the Amazon region in the mid-21st century, a change that kills off the Amazon rain forest, adding further CO_2 to the atmosphere and further warming the climate. While a very interesting simulation, the key issue is whether this initial attempt to construct a true Earth system model includes all of the ecological processes that might be involved, including allowing for the various communities of vegetation that might develop.

Regional Models

Because of the limitations of current supercomputers, those developing current versions of global GCMs or ESMs are currently unable to treat spatial resolutions of less than a few hundred kilometers. However, it is quite clear from climate observations that important variations occur over much smaller scales. For example, in some regions along the west coast of the Americas, the terrain goes from ocean to coastal plain to mountain to valley to desert, for example, all in the space of roughly 100 kilometers. For this reason, GCMs and ESMs provide only the coarsest indication of the local climatic conditions and how these conditions may change if the climate changes and regional models have been developed. Regional models, as their name implies, treat only particular regions of the globe, but do so with finer spatial resolution, typically in the tens of kilometers. Such models are of most value in representing regions with very sharp topographic variations (e.g., western North America, the Alps), particular features of the climate system (e.g., large storm systems such as hurricanes and typhoons), or locations where large changes are occurring (e.g., the tropical Pacific Ocean that experiences the ENSO cycle, and the Arctic Ocean with its variable sea ice). In these models, the effects of the rest of the world are typically handled by

imposing boundary conditions that are generally taken from coarser grid global simulations (hence, regional models are sometimes nested within global models). Even by treating only parts of the globe and thereby greatly limiting the number of different locations for which calculations must be done, the increased resolution and necessarily shorter time step generally limit the lengths of the simulated periods in these calculations to days to years rather than centuries.

In spite of the limitations in coverage and length of simulation, regional models have proven to be very useful. In weather forecasting, fine-scale regional models can be driven by coarser-scale global scale weather models and, particularly for intense weather situations, generate much-improved estimates, for example, of rainfall distribution in regions of complex terrain. For climate studies, regional models are starting to be used to study the spatial distribution of changes in precipitation, being able to better represent, for example, the accumulation and melting of snow in mountainous areas (*see* Modeling Regional Climate Change, Volume 1).

APPROACHES TO TESTING THE PERFORMANCE OF ESMs

While models, particularly numerical models, may be quite comprehensive in terms of the set of processes and components that are represented, models are simply representations of the actual Earth system. With the Earth system being governed not only by a complex set of physical and chemical relationships, but also by biological relationships and human participation, the Earth system is arguably the most complex and challenging system under scientific study (particularly if the urgency and relevance of enhancing understanding is included). Therefore, one cannot expect any of the models to be perfect replicas of the system, and an important task is to determine how much confidence can be placed in such models. This is particularly important because models might exhibit unusual, even chaotic, behavior as a result of their own structure, which may not include the full complement of stabilizing processes or may introduce thresholds or nonlinearities not present in the actual Earth system.

While it is easy to suggest that models should be tested (and not be used unless they match some specified set of observations), there are quite a number of complications. First, processes range from the microphysical to global, and span times from less than a second to many millions of years, and all the processes and their interactions need to be treated under a wide range of conditions. Second, we only have the data that we have – while we can use a number of techniques to reconstruct some aspects about what happened in the past, even today, much less a century or a millennium ago, the data sets being collected are quite limited and are not always the data sets that are needed. Third, we have learned from study of chaotic systems (*see* Chaos and Predictability, Volume 1) that the Earth system, or at least major parts of it, is following only one of what could be many paths that would be consistent with the underlying physical laws. It is a little like what happens in considering the history of the world: while the overall growth and advancement of the world population is proceeding (and this is what models of the world would tend to represent as the major course of history), there are many possible paths, some involving major disruptions such as wars, that can alter what happens in particular places and at particular times along the way. For this reason, models can provide general indications, but not detailed predictions (*see* Prediction in the Earth Sciences, Volume 1). Fourth, a key issue is developing confidence in the ability of the models to project into the future rather than to simply reproduce past occurrences. Because humankind is perturbing the Earth in unique ways for which there are no comparable analogues in the past (*see* Climate Analogues, Volume 1 and the Anthropocene, Volume 1), there is no perfectly analogous test case. As the philosopher Karl Popper has noted, the complexity and irreproducibility of the Earth system make it impossible to actually validate a model – about all that can be attempted is to invalidate the model, and then, once this is done, to improve the model and try the process again with an ever widening set of tests. Thus, while success with some tests can add to confidence in the model, there can never be complete assurance that the models will be correct, only that they have not yet been proven wrong.

The types of tests that have been proposed to test global scale models span a wide range of space and time scales, as well as covering a wide set of the processes included in the models (e.g., Schlesinger, 1991). As an example, the following types of tests are typically used to test the ability of ESMs to simulate the climate.

Testing Representations of Individual Processes

To the extent possible, model developers test the representations of each process introduced into the global models against relevant observational and experimental data. For example, radiation modules of global models, which generally approximate the full spectral dependence of radiation in terms of a few to a few dozen wavelength bands, can be tested against laboratory experiments and against detailed spectrally resolved models. Similarly, the treatment of surface drag can be tested against wind flow experiments across lakes and fields, and so on. The particular challenge in global modeling arises because these representations of particular processes must apply across spatial scales of modest-sized US states or even countries. Thus, representing the effects of clouds, which typically have dimensions up to a few kilometers, using variables calculated on grid sizes of hundreds of kilometers requires

a number of simplifying assumptions and approximations (*see* Cloud- Radiation Interactions, Volume 1). While the empirical relationships that have typically been developed are valid under many conditions, there is no guarantee that they will remain valid as the climate changes.

Testing the Model's Ability to Simulate the Weather

Although climate models are designed to represent the time-average conditions that make up the climate, they do so by trying to simulate the evolution of the system in terms of its weather, i.e., the models simulate the development of weather systems and storms, heat waves and snowfalls, and day and night variations. Two types of tests can be done. First, if the atmospheric and surface components of the model are initialized with the observed weather variables, then the model can be used as a weather forecasting model and its skill can be evaluated. For theoretical reasons, relating to the chaotic nature of the atmosphere, forecasting of the precise state of the atmosphere should be expected to diverge from the behavior of the real atmosphere over periods of one to three weeks, or so. Models developed to date show reasonable skill out to several days to a week, depending on scale of motion. While much success has been shown, more should be possible. The second type of test is to compare the weather statistics generated by the model, with the observations describing the real world. The very wide range of weather conditions around the world is a particularly severe test of the performance of model processes; for example, a cloud or convection parameterization must work from the cold of the Arctic to the heat of the equatorial regions. Again, while there is much agreement with the observed evolution of the weather, there are also systematic differences that have been identified and further improvement should be possible.

Testing the Model's Ability to Simulate the Cycle of the Seasons

In that there are very large seasonal variations in the Earth s climate, testing to determine if a model can represent the annual cycle of the seasons poses a stringent test of model parameterizations. Wintertime conditions, for example, test particularly well how the oceans and atmosphere move energy from warm to cold conditions, whereas summertime conditions test how the model stores heat and cools the surface by radiation and convection. As for the other tests, models generally do fairly well (Gates *et al.*, 1999), on average, in simulating the range of seasonal variations and their variation with latitude. However, shortcomings remain. For example, among the features not yet adequately represented is the location of the intertropical convergence zone that determines the position of equatorial rain belts. Again however, improvements are occurring as greater model resolution and improved treatment of processes become possible.

Testing the Model's Ability to Represent Interannual Fluctuations

Although summer is always warmer than winter, not all summers or winters are alike. Thus, an important test of models is their ability to represent the year to year and decade to decade fluctuations of the climate. The largest contributor to these fluctuations is ENSO, which is most evident over the Pacific Ocean, but which has influences across much of the world (*see* El Niño and La Niña: Causes and Global Consequences, Volume 1). The best of the ESMs are starting to be able to simulate El Niño-like features, and the focus is now moving on to test how well models can represent other such variations (e.g., the North Atlantic Oscillation). Major international model intercomparison projects are currently underway to test model performance on these time scales (Gates *et al.*, 1999; Lambert and Boer, 2001; and www-pcmdi.llnl.gov). Again, the indications are that current models have both strengths and weaknesses, but are improving.

Testing the Model's Ability to Represent the Climate of the Recent Past

With the intense interest in projecting potential climate change during the 21st century, an important test of climate models is their ability to simulate the climatic changes and variations over the past 100 years (or more generally over the past 1000 years). A key challenge in attempting such simulations is that we have only limited records of what happened over the recent past (e.g., we have only very limited information about deep ocean temperatures or above ground air temperatures back more than 50 years), and even less about some of the factors that may have caused climate change over this period (e.g., changes in solar radiation, volcanic eruptions, etc.). In spite of these limitations, model simulations are demonstrating some skill at representing variations over these periods, albeit with significant questions remaining (*see* Climate Change, Detection and Attribution, Volume 1; Model Simulations of Present and Historical Climates, Volume 1).

Testing the Model's Ability to Simulate Glacial Cycling

While we expect the future climate to be warmer than at present, determining how well models can simulate climates that are much colder provides important information about how sensitive the climate is to changes in radiation of the type believed to have been the major driver of glacial cycling. However, limitations in the data and limitations in computer resources have limited use of this type of

testing to either simple models that attempt to simulate the actual cycling, or full models that test model simulations against particular periods in the past rather than against the periods of change. For example, results from the Paleoclimatic Model Intercomparison Project also indicates that models show some limited skill in simulating the climatic conditions prevailing 6000 and 20 000 years ago, when the climate was slightly warmer and significantly colder, respectively (see Joussaume and Taylor, 2000). As explained in Climate Model Simulations of the Geological Past, Volume 1, tests with simplified climate models suggest that, in addition to the normal feedback mechanisms treated in most GCMs, additional feedbacks involving isostatic variations, changes in vegetation, and others also seem to be important in determining long-term climatic behavior.

Testing the Model's Ability to Simulate Warm Climates of the Geological Past

One must go back many millions of years to find test-case climatic conditions from the geological past that were as warm as the Earth might become over the next few centuries, if emissions from fossil fuel combustion continue to grow as rapidly as they have in the past. A key limitation in using such conditions as analogues for the future, however, is that they occurred gradually rather than rapidly. Test cases for models include the Pliocene of about 3 million years ago, which was a few degrees warmer than at present, and the Cretaceous of greater than 65 million years ago when dinosaurs roamed the Earth, which was so warm that tropical vegetation was apparently present in near polar latitudes. Attempts to simulate the very warm Cretaceous using models have not yet been successful – basically, the climate models, responsive as they are to changes in greenhouse gas concentrations, are not able to find a way to transport enough heat to high latitudes to make them much warmer while not significantly warming the equatorial regions (a condition suggested by the geological evidence). Indeed, what is particularly troubling about the various paleoclimatic simulations is that virtually all of the model simulations underestimate the response of the climate system, suggesting that the models are still missing amplifying feedback mechanisms that operate over the long-term.

While models have yet to succeed in satisfactorily representing all of these conditions, the numerical models do provide the most rigorous, quantitative basis for understanding the Earth system, keeping track of the myriad of processes that interact around the world and on many different time scales to create the climate that we now experience. Confidence in our understanding arises, for example, from the ability of the models to represent reasonably well the stability of the climate during historical times (i.e., the model climates are roughly stable in the absence of human-induced forcing), but confidence is limited by an apparent conflicting tendency of models to lean toward overestimating climate change during the 20th century while simultaneously leaning toward underestimating the long-term changes in the climate. In coming years, with better data sets for the present, better reconstructions of past climates, improving model representations, and faster computers, significantly better models are likely to emerge as a result of efforts to test the Earth system and model simulations against an even wider set of conditions.

SUMMARY

A diversity of modeling approaches, ranging from physical to numerical, is being used to improve understanding of the behavior of the Earth system, providing a rich appreciation of both the complexity and the sensitivity of the Earth system. This article has provided an overview of the various types of models and how they are tested. Other articles in this volume describe the application of these models to a wide range of ways to explore and encapsulate understanding of the Earth system, to interpret past and project future conditions, and to test scenarios of what could happen and how changes in policies may affect the outcome (e.g., *see* Projection of Future Changes in Climate, Volume 1).

REFERENCES

Cox, P M, Betts, R A, Jones, C D, Spall, S A, and Totterdell, I J (2000) Acceleration of Global Warming by Carbon Cycle Feedbacks in a 3D Coupled Model, *Nature*, 408, 184–187.

Ellingson, R and Fouquart, Y (1991) The Intercomparison of Radiation Codes in Climate Models (ICRCCM): An Overview, *J. Geophys. Res.*, 96, 8926–8929.

Gates, W L, Boyle, J S, Covey, C, Dease, C G, Doutriax, C M, Drach, R S, Fiorino, M, Gleckler, P J, Hnilo, J J, Marlais, S M, Phillips, T J, Potter, G L, Santer, B D, Sperber, K R, Taylor, K E, and Williams, D N (1999) An Overview of the Results of the Atmospheric Model Intercomparison Project (AMIP), *Bull. Am. Meteorol. Soc.*, 80, 29–56.

Hyde, W T, Crowley, T J, Baum, S K, and Peltier, W R (2000) Neoproterozoic Snowball Earth Simulations with a Coupled Climate/Ice-Sheet Model, *Nature*, 405, 425–429.

IPCC (Intergovernmental Panel on Climate Change) (2001) *Climate Change 2000: The Scienti c Basis*, Working Group I Report, Cambridge University Press, Cambridge UK, 1–881.

Joussaume, S and Taylor, K E (2000) The Paleoclimate Modeling Intercomparison Project, in *PMIP, 2000: Paleoclimate Modeling Intercomparison Project (PMIP)*, Proceedings of the Third PMIP Workshop, Canada, 4–8 October 1999, ed P Braconnot, World Climate Research Program Report WCRP-111, WMO/TD-No 1007, 9–24, see also http://www-pcmdi.llnl.gov/pmip/).

Lambert, S J and Boer, G J (2001) CMIP1 Evaluation and Intercomparison of Coupled Climate Models, *Clim. Dyn.*, 17, 83–106.

MacCracken, M and Ghan, S (1988) Design and Use of Zonally-averaged Models, in *Physically-based Modeling and*

Simulation of Climate and Climate Change (two volumes), ed M Schlesinger, NATO Advanced Science Institute, Kluwer Academic Publishers, Dordrecht, 755–809.

Manabe, S and Wetherald, R (1967) Thermal Equilibrium of the Atmosphere with a Given Distribution of Relative Humidity, *J. Atmos. Sci.*, 24, 241–259.

Mote, P and O Neill, A, eds (2000) *Numerical Modeling of the Global Atmosphere in the Climate System*, NATO Advanced Science Institute, Kluwer Academic Publishers, Dordrecht, 1–517.

Neilson, R P, Prentice, I C, Smith, B, Kittel, T, and Viner, D (1998) Simulated Changes in Vegetation Distribution under Global Warming, in *The Regional Impacts of Climate Change: an Assessment of Vulnerability*, eds R T Watson, M C Zinowera, R H Moss, and D J Dokken, Cambridge University Press, Cambridge, 439–456.

Nihoul, J, ed (1985) *Coupled Ocean- Atmosphere Models*, Elsevier, Amsterdam, 1–767.

Ojima, D, ed (1992) *Modeling the Earth System*, University Corporation for Atmospheric Research/Office for Interdisciplinary Earth Studies, Boulder, CO, 1–488.

Pittock, A B, Ackerman, T, Crutzen, P, MacCracken, M, Shapiro, C, and Turco, R (1986) *The Environmental Consequences of Nuclear War: Volume 1. Physical and Atmospheric Effects*, SCOPE Volume 28, John Wiley & Sons, Chichester, 1–359.

Prinn, R, Jacoby, H, Sokolov, A, Wang, C, Xiao, X, Yang, Z, Eckeus, R, Stone, P, Ellerman, D, Melillo, J, Fitzmaurice, J, Kicklighter, D, Holian, G, and Liu, Y (1999) Integrated Global System Model for Climate Policy Assessment: Feedbacks and Sensitivity Studies, *Clim. Change*, 41, 469–546.

Schlesinger, M, ed (1988) *Physically-Based Modeling and Simulation of Climate and Climate Change* (two volumes), NATO Advanced Science Institute, Kluwer, Dordrecht, 1–1084.

Schlesinger, M, ed (1991) *Greenhouse-gas-induced Climatic Change: A Critical Appraisal of Simulations and Observations*, Elsevier, Amsterdam, 1–615.

Turco, R P, Toon, O B, Ackerman, T P, Pollack, J B, and Sagan, C (1983) Nuclear Winter: Global Consequences of Multiple Nuclear Explosions, *Science*, 222, 1283–1297.

US/USSR (1990) *Prospects for Future Climate: a Special US/USSR Report on Climate and Climate Change*, eds M C MacCracken, M I Budyko, A D Hecht, and Y A Izrael, Lewis Publishers, Chelsea, MI, 1–270.

VEMAP (Vegetation/Ecosystem Modeling and Analysis Project) Members (1995) Comparing Biogeography and Biogeochemistry Models in a Continental-Scale Study of Terrestrial Ecosystem Responses to Climate Change and CO_2 Doubling, *Global Biogeochem. Cycles*, 9, 407–437.

Washington, W and Parkinson, C (1986) *An Introduction to Three-Dimensional Climate Modeling*, University Science Books, Oxford University Press, Oxford, 1–421.

Model Simulations of Present and Historical Climates

CURT COVEY

Lawrence Livermore National Laboratory, Livermore, CA, USA

Theoretical models of climate are used to study the mechanisms of climate change, to help reconstruct climates in the distant past, and to predict future changes in climate such as human-induced global warming. In evaluating the credibility of these models, a natural step is to investigate how well or how poorly they simulate the climate of the present day and the (relatively) well documented climatic record of the recent past. Comparison of model simulations with observations is an ongoing activity that began with development of the first global climate models in the 1970s. Since that time, climate models – like their close relatives, weather prediction models – have slowly but steadily improved.

In the first climate simulations, merely achieving qualitative agreement with the observed large-scale patterns of atmospheric temperature, pressure, rainfall and winds was considered a notable accomplishment. More recent models also simulate clouds, sea ice, ocean circulation and many other climate-relevant phenomena with an increasing degree of fidelity. They also mimic features of observed natural climate variations such as El Niño. Any complex 3-D model of the global climate system will contain simplifications and flaws, so one must ask what level of accuracy is required for a model to be usable. Answering this question (which of course depends on the particular use being made of the model) has proved elusive, but progress has been made in quantifying model errors as well as differences among models.

Evidence of climatic changes over the course of the 20th century implies that the 'present climate' is a moving target. Accordingly, the time evolution as well as the equilibrium state of climate model simulations is studied by comparing it with the observed record. Modern global climate models can reproduce the global mean surface warming of the 20th century – about 0.6°C – provided they include human-produced greenhouse gases (GHGs) and airborne particles (aerosols). The magnitudes of the aerosol effect and other climate forcing factors, however, are highly uncertain. It is thus a matter of controversy how much confidence should be placed in a climate model because of a successful 20th century simulation. An even greater challenge to models is simulation of the pre-20th century climate, in which both the forcing and the observed climate evolution are extremely uncertain.

EVALUATING THE RELIABILITY OF CLIMATE MODELS

A variety of simulation models have been used to investigate factors controlling the climate (*see* **Models of the Earth System**, Volume 1). Some relatively simple models, encoded in as few as 100 lines of computer language and executing quickly on a personal computer, calculate little more than the average temperature of the Earth's surface (*see* **Energy Balance Climate Models**, Volume 1). At the other extreme, 3-D coupled ocean–atmosphere general circulation models (Coupled GCMs) rank among the most demanding users of supercomputer time (*see* **General Circulation Models (GCMs)**, Volume 1; **Projection of Future Changes in Climate**, Volume 1). These models simulate the evolution of global climate including temperature, pressure, clouds and winds throughout the atmosphere; ocean temperature, salinity and circulation; sea ice concentration and thickness; and land surface properties such as snow cover, soil moisture and river runoff. They are the models most often used to investigate human-induced climate changes such as global warming.

Coupled GCMs simulate individual weather systems (as do their close relatives, numerical weather prediction models), producing, for example, day-to-day rainfall and snowfall at thousands of points on the Earth's surface. Such details are believed to be inherently unpredictable beyond a few weeks, so model output is often averaged over each

month of simulated time before it is archived for climate studies. Even so, a modern coupled GCM simulating the development of climate over a century or two can readily produce hundreds of Gigabytes of data encompassing dozens of variables, each recorded at tens of thousands of grid points in 3-D space. Given, for example, assumptions about future increases in atmospheric carbon dioxide (CO_2), such a model will make definite projections of global warming including regional details.

How much of this staggering volume of data is reliable enough to be useful? An obvious first step in answering this question is to ask models to simulate the present-day climate. Success in doing so, while not guaranteeing a model s ability to project future conditions, would seem to be a necessary condition for model credibility. (Even this innocuous-sounding statement begs the question of how to define success, i.e., how much accuracy in a present day simulation is required to accurately simulate future changes in climate.) The present contribution discusses results from model simulations of the present climate and the recent past, recent being defined as extending back in time about 1000 years. Climate model simulation of the more distant past is discussed elsewhere (*see* Climate Model Simulations of the Geological Past, Volume 1).

The emphasis in this contribution will be on coupled GCMs. The extensive output of these models allows a detailed comparison of simulations with observations – so detailed that some discrepancies are inevitable. The question then arises as to what level of agreement between a model and observations is required for the model to be pronounced reliable. The answer, of course, depends on the particular use one contemplates for the model. An error of a degree or two in North Pacific sea surface temperature, for example, could prove disastrous in assessing the state of Alaskan fisheries, but it might not be significant in predicting the globally averaged temperature of the Earth s surface. It must be confessed, however, that even for the latter purpose there is no agreed standard that separates acceptable from unacceptable model performance.

Accordingly, no attempt will be made here to reach a final conclusion about model reliability. The consensus among climate scientists, in the words of one official assessment, is that current models are now able to simulate many aspects of the observed climate with a useful level of skill (Gates *et al.*, 1996). By examining the model-observation comparisons in the following sections, readers can draw their own conclusions.

EQUILIBRIUM CLIMATE STATES

A natural starting point in evaluating a coupled GCMs performance is to examine the long-term average climate it simulates in a control run, with external forcing factors such as the Sun s energy output and the atmospheric concentration of carbon dioxide being kept constant at today s levels. Control run simulations are typically compared to observations of the climate averaged over recent decades. Because the recent observational record includes the response of the climate to both natural perturbations (such as volcanic eruptions and varying solar output) and human-induced effects (most notably the production of GHGs), one cannot expect a model control run to precisely match the observations. Nevertheless, a considerable degree of agreement is expected and achieved.

For example, Figures 1 and 2 compare simulated near-surface temperature and precipitation – taken from one recently developed coupled GCM – with the best available observations. The simulated and observed temperatures are remarkably similar, especially considering that the model uses physical conservation laws without direct knowledge of the observed present day climate. In contrast, conventional wisdom in the early days of climate modeling held that coupled GCMs could not accurately reproduce the observed climate without including observed surface temperatures as input and assuming artificial sources of energy at the Earth s surface to force the models into agreement with the observations. This technique, known euphemistically as flux adjustment or flux correction, is no longer employed in many coupled GCMs.

Of course any model that attempts to simulate global climate in detail will have errors to some degree. From Figures 1 and 2 several discrepancies between the model and observations are apparent. The simulated near-surface temperature is generally too low. The simulated precipitation is more problematic: the global mean error is of the order of 10%, and the spatial pattern over the Earth s surface, while recognizably similar to the observations in many respects, includes several areas of disagreement. One artifact occurs in the central and eastern tropical Pacific Ocean (about $90-180\,^{\circ}$W longitude). Here the simulated precipitation shows a minimum along the Equator, flanked by symmetric maxima at $10\,^{\circ}$N and $10\,^{\circ}$S latitudes. Observations in this region reveal only one precipitation maximum that is oriented east to west, the Intertropical Convergence Zone at $10\,^{\circ}$N. South of the Equator a band of maximum precipitation runs from northwest to southeast in the observations, but this so-called Southwest Pacific Convergence Zone does not appear in the model simulation. Precipitation has been a notoriously stubborn problem in coupled GCMs, especially in the tropics, as one might expect from the large difference between the scales of relevant processes (clouds and convection, ranging from micrometers to kilometers) and the scales explicitly resolved by the atmospheric component of global models (typically in excess of 1 km vertically and 100 km horizontally).

Despite such difficulties, the accuracy of coupled GCM control runs is impressive. Areas of agreement with observations may be found not only near the surface

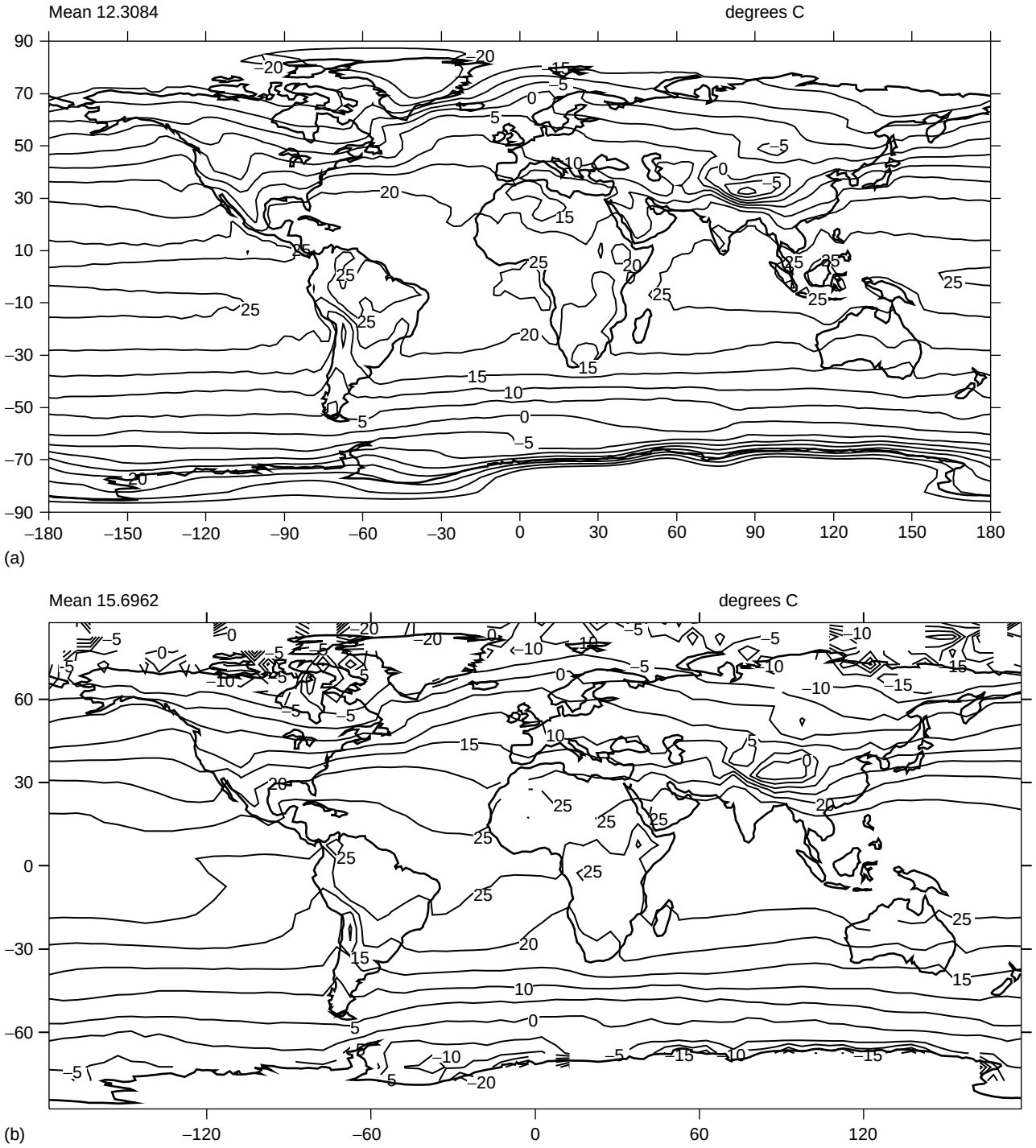


Figure 1 Annual mean near-surface air temperature: (a) as simulated by the Parallel Climate Model (Washington *et al.*, 2000); and (b) as observed by Jones *et al.* (1999). (For this figure, model output was obtained from the Coupled Model Intercomparison Project database (http://www-pcmdi.llnl.gov/cmip) and observed data were obtained from the PCMDI observational data system (http://www-pcmdi.llnl.gov/obs/pods/mo/obs.mo.source.htm))

but also in the overlying atmosphere, in sea ice regions near the poles, and within the ocean. Simulation of mid-atmospheric temperatures and winds is generally the most accurate aspect of GCM performance, presumably due to the primary attention modelers have given to meteorological applications of atmospheric thermodynamics and fluid dynamics. The stratosphere, a more recent focus of GCM development, is not as well simulated. Overly low polar

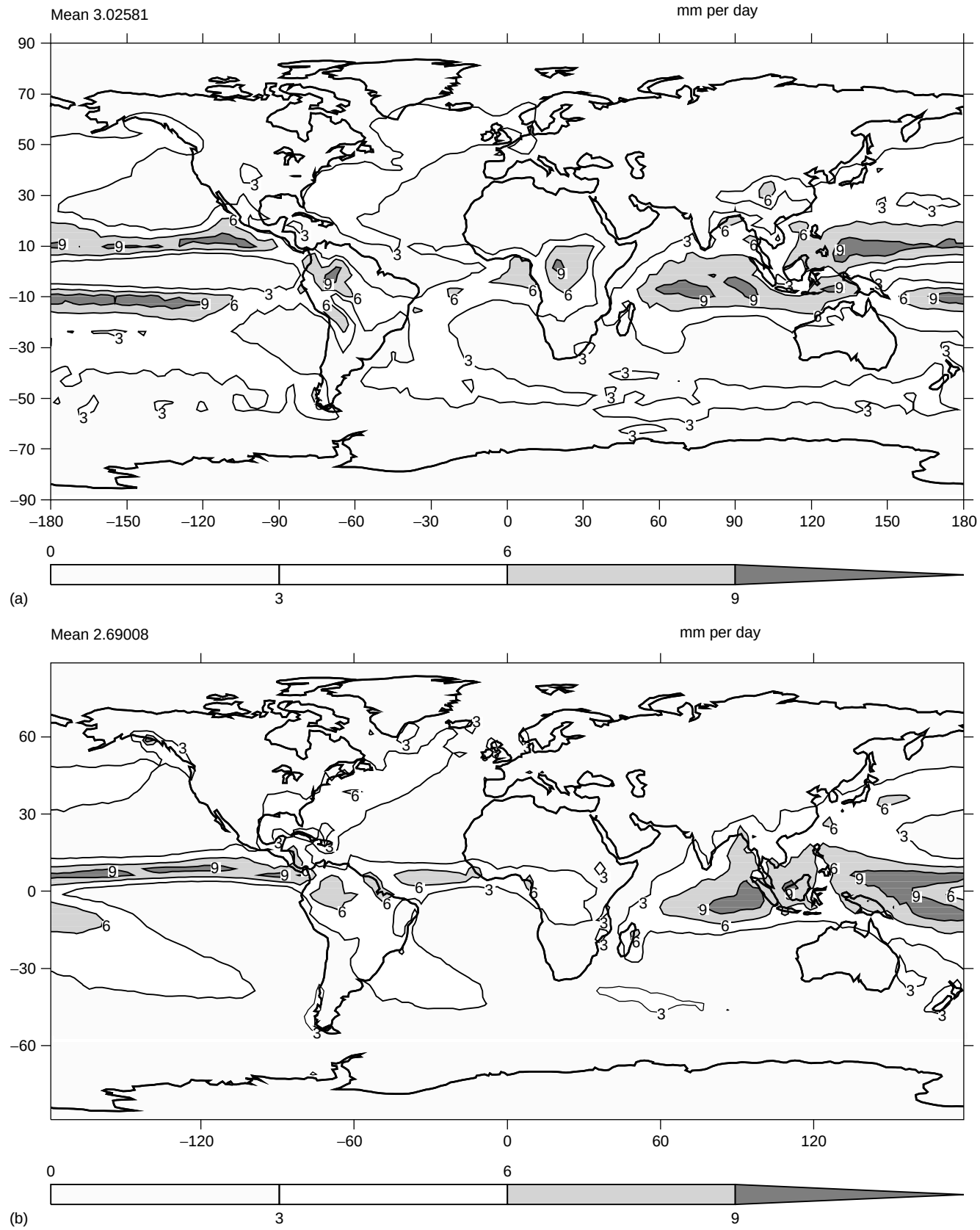


Figure 2 As in Figure 1 for annual mean precipitation (with observations from Xie and Arkin, 1997)

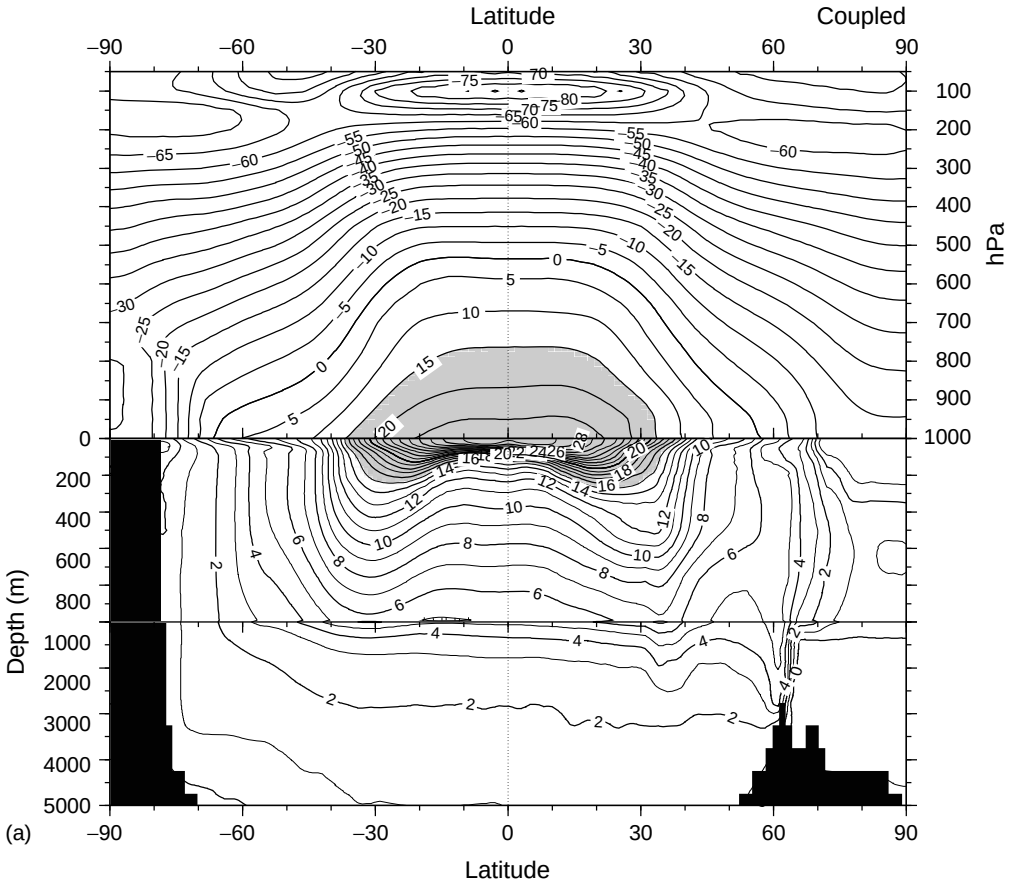


Figure 3 Longitude average (zonal mean) of annual mean temperature simulated in the atmosphere and the ocean by a coupled GCM: (a) after 50 simulated years without flux adjustment and; (b) in its initial state, which was chosen to closely match observations. The vertical coordinate in the atmosphere is pressure in hectopascals, and the vertical coordinate in the ocean is depth in meters. From Guilyardi and Madec (1997). (Reproduced by permission of Springer–Verlag Publishers)

temperatures and misplaced jet stream circulations are common deficiencies in model stratospheres. Nevertheless, GCMs succeed in reproducing the large-scale structure of stratospheric temperature and wind, producing a 3-D pattern of these quantities that is recognizably similar to observations. Analogous comments apply to ocean temperatures, salinity and currents. Modern coupled GCMs simulate temperatures and currents at and below the sea surface with sufficient fidelity to reproduce the observed magnitude and direction of large-scale ocean heat transport. This success, in turn, has reduced the need for flux adjustments at the air-sea interface.

To exemplify the state of both atmospheric and oceanic modeling, Figure 3 shows a cross-section of temperatures from the ocean bottom to the stratosphere, both as observed and as simulated by a recently developed coupled GCM. The similarity between the model and the observations is remarkable, considering that the model was allowed full opportunity to wander away from its observation-based initial conditions for 50 simulated years without flux adjustments. For example, the sharp decline of ocean temperatures in the first 1000 m of depth between latitudes 40 °S and 40 °N (the so-called ocean thermocline) was maintained during the control run. On the other hand, a detailed comparison of the changes during the 50-year simulation reveals a warming of most of the ocean (by 0.5–1 °C) and the atmosphere (by 1–2 °C generally and by 4 °C or more in the Northern Hemisphere stratosphere). This shift of the simulated climate away from observations – often called climate drift – is a continuing problem, particularly with models that do not use flux adjustment.

Climate models simulate the cycle of the seasons as well as the annual mean climate. Figure 4 shows both model simulations and observations of the amplitude of the cycle, which is measured by the difference between July and January values of near-surface air temperature over land. The solid line in the figure represents the average of results from eight coupled GCMs that do not use flux adjustment; dashed lines indicate the range defined by adding and

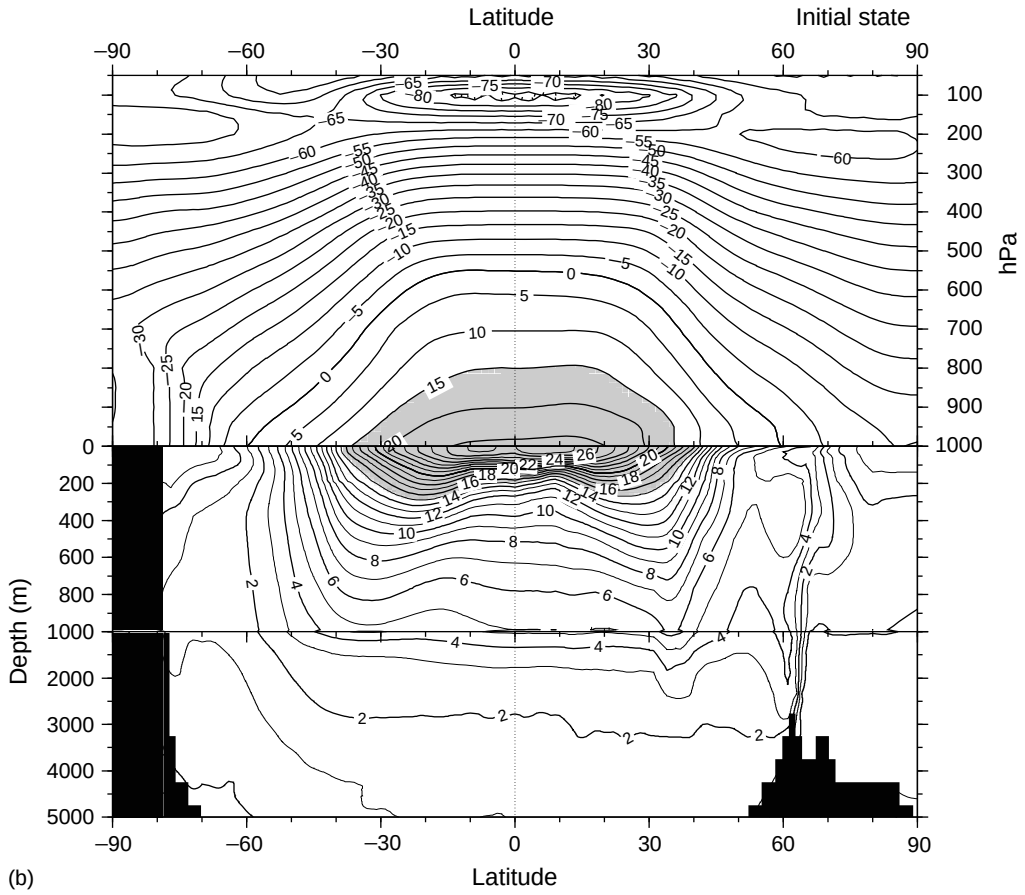


Figure 3 (*Continued*)

subtracting one standard deviation taken over the different model results. At most latitudes the average model simulation nearly coincides with the observational data (shown as points). The standard deviation of model results is typically about 20% of the average value, indicating substantial differences among the models. Nevertheless, the extent to which the models on average match the observations is striking, since the models shown (unlike their flux-adjusted counterparts) make no explicit assumptions designed to force them into agreement with the present day climate.

The most troubling deficiencies in coupled GCMs involve water vapor and clouds, the behavior of which account for most of the uncertainty in predicting climate changes such as human-induced global warming, and air-sea exchanges of heat and momentum, by which the atmosphere and ocean interact to determine the behavior of the climate system. These phenomena have been notoriously difficult to simulate even to a rough degree of accuracy. Examination of several models in the early 1990s, for example, revealed that many underestimated cloud coverage by nearly a factor of two over large parts of the world, and that about half erroneously calculated not only the magnitude but also the direction of large-scale ocean heat transport implied by air-sea exchanges. A few years later, however, the same examination of then-current models showed a noticeable reduction in the most egregious mistakes. The overall picture that emerges from such investigations is one of continuing serious problems but steady incremental improvements in model performance.

The gradual improvement of coupled GCMs is well illustrated by comparison of the model exhibited in Figures 1 and 2 with a much earlier version. The current model (Washington *et al.*, 2000) takes advantage of modern computer technology to simulate the ocean at roughly 50 km horizontal resolution using 32 separate vertical levels, permitting a detailed regional simulation of currents, temperature and salinity. In the earlier version (Washington *et al.*, 1980), the ocean component of the model used only four vertical levels and a horizontal resolution of about 500 km. The developers assessment at that time was that even though many first-order aspects of the climate system are simulated, improvements were still needed. The crude resolution appeared to induce a spurious numerical diffusion of heat in the model ocean, resulting in overly low temperatures in the tropics and overly high temperatures at higher latitudes (a problem typical of coupled GCMs at that time).

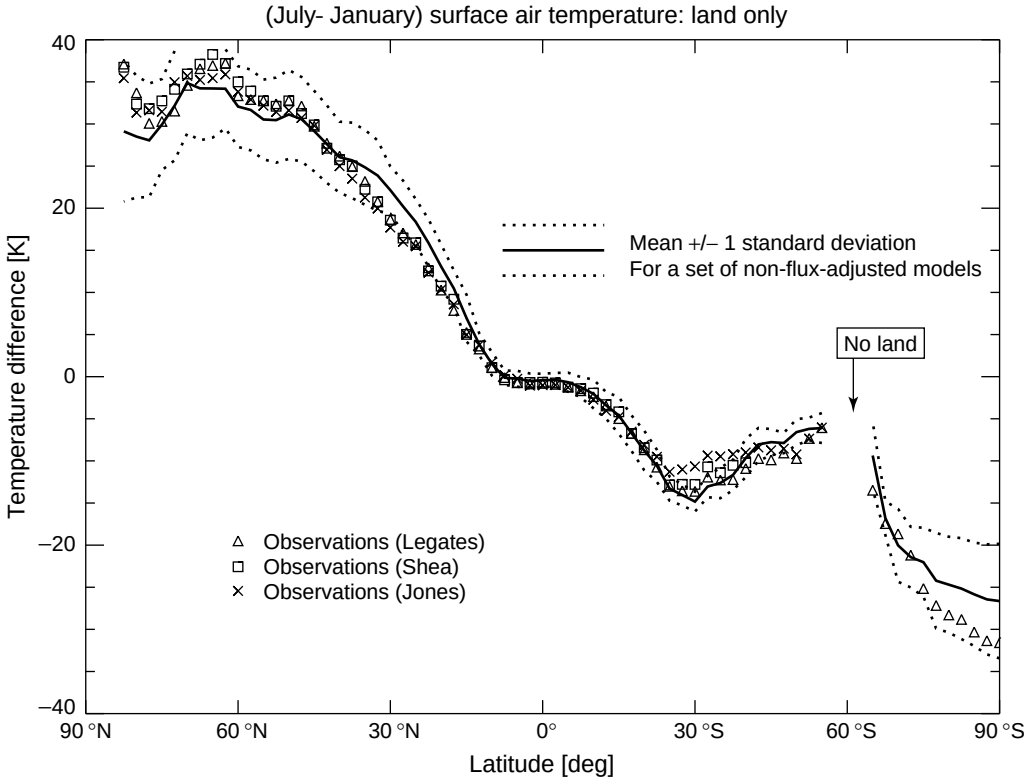


Figure 4 The seasonal cycle of near-surface air temperature over land, averaged over longitude and expressed as the difference between July and January mean temperatures. Points indicate observed values from three different datasets (identified in the graphic). The solid line is an average over results from eight coupled GCMs that do not use flux adjustment, and the dashed lines indicate the range implied by the standard deviation over these eight models. (Data from Figure 6(a) of Covey *et al.*, 2000. Adapted by permission of Springer–Verlag Publishers)

The model s western tropical Pacific warm pool region was too cool by about 5 °C. The absence of key features such as the warm pool in model simulations led many observational specialists to the opinion that detailed model-data comparison would be unproductive. In coupled GCM papers published during this era, evaluations of control run simulations tended to be rather perfunctory and to claim, at best, a qualitative similarity between the model and observations.

In contrast, the control run simulations of modern coupled GCMs are sufficiently Earth-like to compare quantitatively with observations. For example, examining the dozen or so coupled models in current use, one finds that the space-time pattern correlation between simulations and observations is around 50% for precipitation and greater than 95% for near-surface air temperature (on a scale in which 0% indicates no correlation and 100% indicates perfect correlation). The correlation values for precipitation indicate continuing model deficiencies, but even in this case model improvements may be limited by uncertainties in the observations as much as by problems with the models themselves. This conclusion follows from the relatively low (80%) correlations between different observational datasets for precipitation.

NATURAL CLIMATE VARIABILITY

In addition to the long-term mean climate, confidence in models depends on their ability to simulate the natural year-to-year fluctuations that result from the interactions of the ocean and atmosphere (*see* Natural Climate Variability, Volume 1). The most prominent of these inter-annual variations is the El Niño/Southern Oscillation (ENSO) (*see* El Niño/Southern Oscillation (ENSO), Volume 1). ENSO simulations from present versions of global GCMs suffer from a number of deficiencies including, typically, an underestimate of this phenomenon s amplitude and a misplacement of the region of maximum temperature variability. Global GCMs also show only modest skill at best at forecasting the detailed evolution in time of individual ENSO events (Landsea and Knaff, 2000). Despite these problems, the ability of the models to even partly simulate ENSO came as a pleasant surprise in the late 1980s. At that time conventional wisdom held that the models crude horizontal resolution in the tropical Pacific precluded simulation of anything resembling ENSO.

Since then, many coupled GCMs have been designed with enhanced resolution in this region, allowing improved

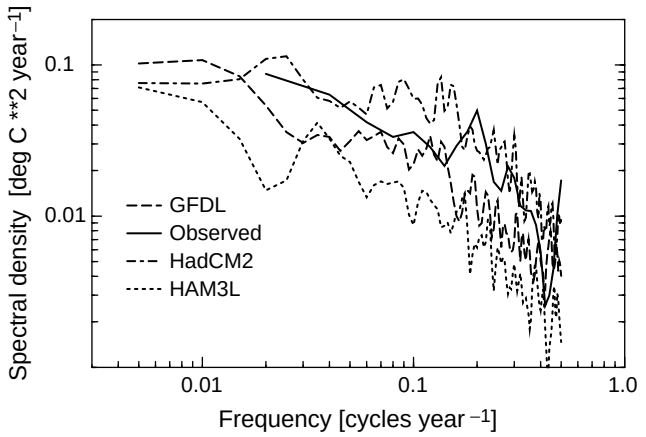


Figure 5 Intensity of global and annual mean near-surface air temperature variations as simulated by three different coupled GCMs (from the Geophysical Fluid Dynamics Laboratory (GFDL) in the US, the Hadley Center for Climate Research and Prediction (HadCM2) in the UK, and the Max Planck Institute for Meteorology in Hamburg, Germany (HAM3L)). The simulated variability is compared with observations by Jones *et al.* (1999). For both models and observations the intensity of variations is plotted as a function of their frequency in time. (From Stouffer *et al.* (2000). Redrawn by permission of the American Meteorological Society)

ENSO simulation. The models have also succeeded in reproducing observed aspects of other modes of climate variability, such as monsoons and the North Atlantic Oscillation. To summarize the state of the art, Figure 5 compares the global near-surface temperature variability of three coupled GCMs with observations over a range of time scales, from frequencies of 0.02–0.5 cycles per year (i.e., periods from 2–50 years). Over this range, the models and the observations have the same magnitude of variations.

Of course a plot that displays global mean results on logarithmic scales can hide a multitude of errors. One indication of the problematic aspects of model-simulated variability is provided by Figure 6, which shows results for land and ocean areas separately. This figure presents the near-surface temperature variability of 16 different coupled GCMs in the form of histograms. It appears that most of the models underestimate variability over the oceans but overestimate it over land. The ocean-area underestimate might be expected from continuing problems with ENSO simulation (although it seems that other factors are involved as well). Over land, the overestimate of variability seems worse in models having less sophisticated parameterizations of land surface processes.

One complication in interpreting the variability of model control runs is that, by assumption, such runs do not include the effects of changing the degree of climate forcing. Changes in forcing can be induced by human activities such as the burning of fossil fuels, which add heat-trapping carbon dioxide to the atmosphere. Changes in forcing can also be induced by purely natural phenomena such as volcanic eruptions or changes in solar brightness. Volcanic eruptions add airborne particles (aerosols) to the upper atmosphere, where they reflect sunlight back to space. Changes in solar brightness affect the amount of solar energy reaching Earth; such changes have been minor (no more than 0.1%) during the few years they have been measured, but they may be greater over time scales of several decades or more. All such phenomena presumably add to the climate variability of the real world. In the following two sections their effect on climate model simulations is considered.

THE 20th CENTURY'S CLIMATE

Globally averaged near-surface air temperature increased over the course of the 20th century by about 0.6 °C (*see* The

Global Temperature Record, Volume 1). This amount, while seemingly tiny, is confirmed by the agreement of independent land-based and ocean-based measurements, and it appears to be much greater than global mean temperature changes over the preceding 900 years. (Note, however, that for estimates of pre-20th century temperatures the margins of error – to the extent they are known at all – are distressingly large.) Climate models are able to simulate the observed increase of global mean surface temperature over the course of the 20th century. They can do so, however, only if they assume changes in climate forcing due to human production of both GHGs (mainly carbon dioxide) and aerosols. Many if not all climate scientists have concluded from these simulations that human activities are altering the climate enough to make a detectable difference in the observational record (*see* Climate Change, Detection and Attribution, Volume 1).

The matter of human-produced aerosols is the most problematic in this tangle of issues (*see* Aerosols, Effects on the Climate, Volume 1). Burning fossil fuels release not only carbon dioxide – which together with water vapor is the main product of the chemical reactions involved in combustion – but also a number of trace pollutants including aerosols. The aerosols include sulfate particles that scatter sunlight more than they absorb it. At the moment the scientific consensus is that the net effect of the aerosols is to cool the Earth by reflecting sunlight back to space, counteracting some, but not all of the greenhouse warming effect of carbon dioxide and related gases. Because of the relatively short lifetime of human-produced aerosols in the atmosphere, their concentration and thus their climate forcing varies substantially over different locations (unlike carbon dioxide, which is relatively long lived and well mixed in the atmosphere). Complicating matters further is the presence of a number of other natural and human-induced perturbations to the climate, including land cover changes, volcanic eruptions, ill-defined changes in solar brightness, ozone depletion, a variety of additional GHGs and aerosols in addition to carbon dioxide and sulfate particles, and above all the possibility of indirect climate forcing by aerosol-induced changes in cloud properties.

Although it seems highly likely that the sum of human-induced greenhouse and aerosol forcing amounts to heating rather than cooling, data on aerosol amounts and properties are not sufficient to accurately determine the amount of heating even in the global average. Model simulations of the climate of the 20th century use plausible values as input and their output shows fair agreement with the observational record. Figure 7 shows a typical set of calculations by a coupled GCM. The global mean near-surface temperature increase simulated by the model over the entire 20th century is a close match to the observed value. In the mid 20th century, however, the observations show a pause in global warming that the model

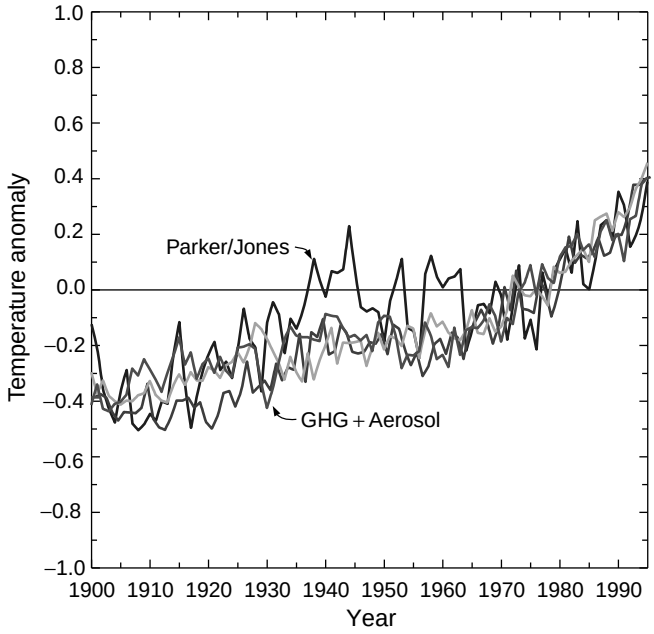


Figure 7 Twentieth century time series of global and annual mean near-surface air temperature, expressed as a deviation from the average temperature around mid-century. The curve marked Parker/Jones gives observations from Jones *et al.* (1999). The remaining (GHG • Aerosol) curves give simulated values from a coupled GCM including anthropogenic GHGs and aerosols; the simulations producing these three curves differed only in the initial conditions supplied to the GCM. (From Boer *et al.* (2000). Reproduced by permission of Springer–Verlag Publishers)

does not reproduce. This discrepancy, which is common in model simulations of 20th century climate, may be due to climate forcing that is present in the real world but missing in the model run, to deficiencies in the model itself, or even to inherent limits to the predictability of the climate system due to chaos. The last possibility is illustrated by the somewhat different results obtained by three model runs begun with slightly different initial conditions (all within the realm of possibility for the climate of the beginning of the 20th century, when the state of the oceans in particular was poorly characterized). An expanded ensemble of model runs might include one simulated temperature curve that matches the observations much better than others. Such a result has occurred for at least one coupled GCM (Delworth and Knutson, 2000). Other models obtain better agreement with early to mid-century global mean temperatures if they include plausible variations in solar energy output (e.g., Stott *et al.*, 2000).

Regardless of the details, model runs including both human-produced GHGs and aerosols tend to give more accurate reproductions of the 20th century climate record than those that omit them. This conclusion follows from

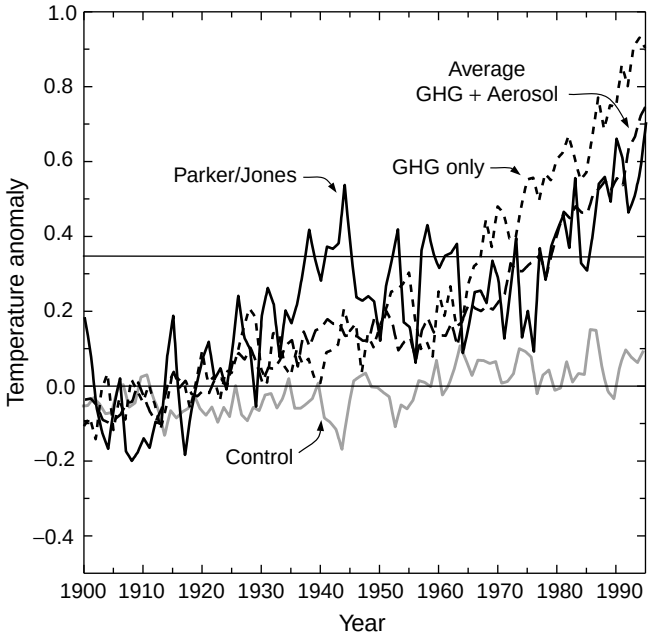


Figure 8 Additional 20th century time series of global and annual mean near-surface air temperature, expressed as a deviation from the average temperature around mid-century. Again the curve marked Parker/Jones gives observations from Jones *et al.* (1999). The long-dashed line (labeled Average GHG • Aerosol) gives the average of the three model-simulated GHG • Aerosol time series shown in Figure 7. The short-dashed line (labeled GHG only) gives results from a simulation in which anthropogenic GHGs were included, but aerosols were not included. The line labeled Control gives results from a model control run in which neither anthropogenic GHGs nor aerosols were included; the thin horizontal line at about 0.35 °C indicates three times the standard deviation of the control run. (From Boer *et al.* (2000). Reproduced by permission of Springer–Verlag Publishers)

Figure 8, which compares the average of the three model runs in Figure 7 with analogous runs in which aerosols are omitted (leaving the GHGs only) and in which no changes in climate forcing at all are included (the control run). For the last third of the 20th century, agreement is excellent between observations and the model simulation that includes both GHGs and aerosols. The control run, in contrast, produces more or less steady temperatures in obvious disagreement with observations, while the run that omits aerosols produces too much warming. Results such as these have been obtained from a variety of climate models and have led many to conclude that a human fingerprint is apparent in the observational record of climate. One cautionary note is that the spatial patterns of simulated global warming (omitted by the global averaging that produces the curves in Figures 7 and 8) are rather sensitive to the details of the assumed climate forcing.

THE PRE-INDUSTRIAL PERIOD

Finally, one may consider the performance of climate models in simulating the history of climate over a period of several centuries. This evaluation is made difficult by the paucity of observational data for both climate forcing and temperature response. In recent years, however, compilations of indirect or proxy data indicative of surface temperature – sometimes referred to as high resolution paleodata – and somewhat more extensive datasets on forcing have provided enough confidence to allow model simulations to begin (*see* Natural Records of Climate Change, Volume 1; Little Ice Age, Volume 1; Medieval Climatic Optimum, Volume 1). Figure 9 shows the results of one such experiment. To save computer time, this study employed a simple energy balance ocean – atmosphere model that was calibrated to have the same response to climate forcing as a typical coupled GCM. Forced by a plausible scenario of pre-industrial natural climate forcing, the model simulates Northern Hemisphere temperatures in agreement with the observed gradual decline over most of the millennium (called the Little Ice Age). When the simulation is extended through the 19th and 20th centuries with forcing similar to that used to produce Figures 7 and 8, results analogous to those obtained with coupled GCMs emerge: a general agreement with the overall temperature rise during the period, but disagreement in detail.

With multi-century climate simulations in their infancy, it is difficult to make firm statements about model performance in such experiments. Future studies will no doubt include coupled GCMs, allowing among other things a more complete assessment of the role of internally generated long-term oscillations of the climate system. The simplified nature of energy balance climate models rules out many of these modes of variability. One conclusion from the study that produced Figure 9 was that about half of climate variability in the pre-industrial period was generated by externally varying natural forcing from the Sun and volcanoes. Given the uncertainty in the forcing, one could imagine a coupled GCM study using less external forcing but achieving equally good agreement with the observed variability.

CONCLUSIONS

Coupled GCMs retain enough problematic features to keep climate scientists busy for the indefinite future, but when viewed over the past few decades they show a steadily improving agreement between their control run simulations and recent observations. Much of this improvement is attributable to higher resolution in the models finite difference grids (or numerical equivalents such as a greater number of basis functions in spectral-space models). Higher resolution is made possible by the tremendous increase in

computer power and storage capacity that has occurred since climate models were first developed. Making productive use of this computer power is a nontrivial task, however, especially in parallel processing machines. The challenge has been met by a variety of research groups, often with the additional benefit of entraining mathematicians and computer scientists into global change research.

An equally praiseworthy accomplishment of model developers has been and continues to be increasingly sophisticated sub-grid scale parameterizations (*see* Sub-grid Processes, Volume 1; Earth System Processes, Volume 1). These are required by even the highest-resolution GCMs in order to account for spatial scales not explicitly resolved by the models. For the atmosphere, parameterizations describing clouds and convection have evolved from simple algorithms – which do little more than prevent unphysical instabilities from persisting – to extremely complex accounting schemes incorporating a wealth of empirical data. An analogous increase in sophistication has occurred in the treatment of land surface processes. Similar progress has also taken place in the oceanic components of coupled models, many of which now employ parameterizations of eddies operating at scales down to a few kilometers (where most of the kinetic energy of ocean circulation resides).

Coupled GCMs, together with simpler energy balance models, are increasingly used not only for purely scientific research but also in the policy-making arena, most notably in deciding what if anything to do about human-induced global warming. Controversy exists about how far the models predictions can be trusted. On occasion the discussion sounds sterile, like an argument over whether a glass of water is half empty or half full, but continued debate is healthy and inevitable considering the stakes involved in adapting or responding to global change. At the same time, both scientists and policy-makers are taking the models seriously. Much of this confidence arises from the fidelity with which the models simulate the present day climate as well as that of the past 100–1000 years.

ACKNOWLEDGMENTS

I thank the editors of this encyclopedia and a large number of additional colleagues, especially my fellow authors in the Intergovernmental Panel on Climate Change (IPCC), for extensive discussions of climate model evaluation. This work was performed under auspices of the US Department of Energy by University of California Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.

See also: Climate Model Simulations of the Geological Past, Volume 1; Modeling Regional Climate Change, Volume 1; Projection of Future Changes in Climate, Volume 1.

REFERENCES

Bell, J, Duffy, P B, Covey, C, and Sloan, L (2000) Comparison of Temperature Variability in Observations and 16 Climate Model Simulations, *Geophys. Res. Lett.*, 27, 261–264.

Boer, G J, Flato, G, Reader, M C, and Ramsden, D (2000) A Transient Climate Change Simulation with Greenhouse Gas and Aerosol Forcing: Experimental Design and Comparison with the Instrumental Record for the 20th Century, *Clim. Dyn.*, 16, 405–425.

Covey, C, Abe-Ouchi, A, Boer, G J, Boville, B A, Cubasch, U, Fairhead, L, Flato, G M, Gordon, H, Guilyardi, E, Jiang, X, Johns, T C, Le Treut, H, Madec, G, Meehl, G A, Miller, R, Noda, A, Power, S B, Roeckner, E, Russell, G, Schneider, E K, Stouffer, R J, Terray, L, and von Storch, J-S (2000) The Seasonal Cycle in Coupled Ocean-Atmosphere General Circulation Models, *Clim. Dyn.*, 16, 775–787.

Delworth, T L and Knutson, T R (2000) Simulation of the Early 20th Century Global Warming, *Science*, 287, 2246–2250.

Gates, W L, Henderson-Sellers, A, Boer, G J, Folland, C K, Kitoh, A, McAvaney, B, Semazzi, F, Smith, N, Weaver, A J, and Zeng, Q-C (1996) Climate Models – Evaluation, in *Climate Change 1995: The Science of Climate Change*, eds J T Houghton, L G Meira Filho, B A Callander, N Harris, A Kattenburg, and K Maskell, Cambridge University Press, 229–284.

Guilyardi, E and Madec, G (1997) Performance of the OPA/ARPEGE-T21 Global Ocean-Atmosphere Coupled Model, *Clim. Dyn.*, 13, 149–165.

Jones, P D, New, M, Parker, D E, Martin, S, and Rigor, I G (1999) Surface Air Temperature and its Changes Over the Past 150 Years, *Rev. Geophys.*, 37, 173–199.

Landsea, C W and Knaff, J A (2000) How Much Skill was There in Forecasting the Very Strong 1997–1998 El Niño? *Bull. Am. Meteorl. Soc.*, 81, 2107–2120.

Stott, P A, Tett, S F B, Jones, G S, Allen, M R, Mitchell, J F B, and Jenkins, G J (2000) External Control of 20th Century Temperature by Natural and Anthropogenic Forcings, *Science*, 290, 2133–2137.

Stouffer, R J, Hegerl, G, and Tett, S (2000) A Comparison of Surface Air Temperature Variability in Three 1000-year Coupled Ocean-Atmosphere Model Intergations, *J. Clim.*, 13, 513–537.

Washington, W M, Semtner, Jr, A J, Meehl, G A, Knight, D J, and Mayer, T A (1980) A General Circulation Experiment with a Coupled Atmosphere, Ocean and Sea Ice Model, *J. Phys. Oceanogr.*, 10, 1887–1908.

Washington, W M, Weatherly, J W, Meehl, G A, Semtner, Jr, A J, Bettge, T W, Craig, A P, Strand, Jr, W G, Arblaster, J, Wayland, V B, James, R, and Zhang, Y (2000) Parallel Climate Model (PCM) Control and Transient Simulations, *Clim. Dyn.*, 16, 755–774.

Xie, P and Arkin, P (1997) Global Precipitation: A 17-year Monthly Analysis Based on Gauge Observations, Satellite Estimates, and Numerical Model Outputs, *Bull. Am. Meteorl. Soc.*, 78, 2539–2558.

FURTHER READING

Covey, C, AchutaRao, K M, Lambert, S J, and Taylor, K E (2000) *Intercomparison of Present and Future Climates Simulated by Coupled Ocean-Atmosphere GCMs*, PCMDI Report No. 66 (UCRL-ID-140325; available on the Web at http://www-pcmdi.llnl.gov/pcmdi/pubs/pdf/report66.

Crowley, T J (2000) Causes of Climate Change Over the Past 1000 Years, *Science*, 289, 270–277.

Gates, W L, Boyle, J, Covey, C, Dease, C, Doutriaux, C, Drach, R, Fiorino, M, Gleckler, P, Hnilo, J, Marlais, S, Phillips, T, Potter, G, Santer, B D, Sperber, K R, Taylor, K, and Williams, D (1999) An Overview of the Atmospheric Model Intercomparison Project (AMIP). *Bull. Am. Meteorl. Soc.*, 80, 29–55.

McAvaney, B, Covey, C, Joussaume, S, Kattsov, V, Kitoh, A, Ogana, W, Pitman, A J, Weaver, A J, Wood, R A, and Zhao, Z-C (2001) Climate Models – Evaluation, in *Climate Change 2001: The Scienti c Basis*, eds J T Houghton, Y Ding, D J Griggs, M Noguer, P J van der Linden, X Dai, K Maskell, and C A Johnson, Cambridge University Press, Cambridge, 1–881.

Projection of Future Changes in Climate

JERRY D MAHLMAN AND RONALD J STOUFFER

Princeton University, Princeton, NJ, USA

This article summarizes the basic science of human-caused climate warming, the role of climate data in documenting long-term climate changes, the role of climate models in evaluating climate warming to date, and model projections of expected climate changes over the next century. These projections assume that substantial reductions of fossil fuel use will not have occurred.

Such projections of climate changes include further global-mean surface- air temperature increases of roughly 3.0°C by the year 2100. While regional details are more uncertain, some substantial effects are rather well understood. Land is expected to warm more than ocean. Northern Hemisphere high latitudes are expected to warm more than the lower latitudes. The hydrological cycle is expected to become more intense (more oods and more droughts). Vertical mixing of temperature and trace constituents in the ocean is likely to become more inhibited. Summertime soil moisture in continental interiors is likely to decrease substantially. Summertime heat index in moist subtropical areas is likely to increase by about 50% more than the actual temperature increases. Intense tropical cyclones are expected to become more intense, on the average. Finally, sea level is expected to rise by roughly 15- 70 cm over this century, with steady rises for many centuries thereafter.

AN OVERVIEW OF THE SCIENCE OF HUMAN-CAUSED CLIMATE WARMING

Historical Setting

Much of the material in the first three major sections is adapted from Mahlman (1998). Since the pioneering work of Arrhenius (1896), the possibility of a net warming of the global climate due to an increase in atmospheric carbon dioxide (CO_2) concentration produced by the burning of fossil fuel has been recognized. The subject matured scientifically with the publication by Manabe and Wetherald (1967) of the first fully self-consistent model calculation of this so-called greenhouse warming effect. They used a simple one-dimensional (altitude only) model of the global atmosphere. In the three decades since, a tremendous amount of observational, theoretical, and modeling research worldwide has been directed at the climate system and possible changes in it due to human activity.

Some Fundamental Aspects of Greenhouse Warming Science

The Earth is strongly heated every day by incoming radiation from the Sun. This heating is offset by an equally strong flux of infrared radiation leaving the planet. Interestingly, if Earth were without any atmosphere, and if its surface reflectivity did not change, the global-mean surface–air temperature would be roughly 33°C colder than it is today. This large difference is due to the strong atmospheric absorption of infrared radiation emitted from the Earth s surface, and a portion of this absorbed energy being re-emitted back to the surface. The major gaseous atmospheric infrared absorbers are water vapor, CO_2, and ozone; in popular language, they are called greenhouse gases. This strong infrared absorption and strong downward atmospheric emission effect is extremely robust. It is readily measured in the laboratory and is straightforwardly measured from Earth-orbiting satellites. Simply put, adding CO_2 and other greenhouse gases to the atmosphere adds yet another infrared blanket to the planet and, thus, directly acts to warm the Earth s atmosphere.

The basic story of human-induced greenhouse warming remains simple. Increased infrared absorptivity due to increasing atmospheric concentrations of CO_2 and other greenhouse gases produces a net heating effect at the Earth s surface, due mainly to increased downward infrared radiation. The effect is not dissimilar to the observable suppression of nighttime cooling when there is cloud cover or when the air is very humid.

It is an inescapable scientific conclusion that added CO_2, water vapor, and other greenhouse gases provide a measurable direct addition to the atmospheric trapping of infrared radiation leaving the surface of our planet. However, a naive comparison of the relative greenhouse efficiencies of water

vapor and CO_2 quickly becomes problematic because water vapor exists in the climate system almost exclusively as a feedback gas. That is, the warming due to increased CO_2 concentrations produces a feedback that increases the water vapor amount. All models and observations currently indicate that, as climate warms or cools, the observed and calculated global-mean relative humidity of water vapor remains roughly constant as the climate changes, whereas the water vapor mixing ratio changes considerably. Relative humidity is the ratio (in percentage) of the actual amount of water vapor in the air to the maximum possible amount (saturated). The saturation vapor pressure, determined from the Clausius–Claperon equation of classical thermodynamics, is a strong exponential function of temperature, roughly doubling for each 10 °C increase in temperature. Water vapor mixing ratio is the mass of water vapor of air divided by the mass of dry air; following an air parcel, it is generally conserved for a few days when no condensation is present. As climate warms due to an increasing atmospheric CO_2 concentration, the holding capacity of atmospheric water vapor increases exponentially. Thus, the increase in water vapor provides a powerful positive feedback mechanism – that is, a process that acts to amplify the original warming caused by increasing atmospheric CO_2 levels. With this major positive feedback, the modeled climate sensitivity increases by about a factor of three, to approximately 3 °C warming for a doubling of the atmospheric CO_2 concentration ($2 \times CO_2$) over its preindustrial level.

The term climate sensitivity typically refers to the level of equilibrium global-mean surface–air temperature increase that the climate system would experience in response to a doubling of CO_2 (*see* Climate Sensitivity, Volume 1). Each model has its own climate sensitivity, almost guaranteed to be somewhat different from the unknown value for the real world. Currently, observational evidence remains roughly consistent with the modeling results that project a strong positive water vapor mixing ratio feedback under approximate constancy of relative humidity as the climate changes (Oort and Liu, 1993; Sun and Held, 1996). The long-term quality of water vapor data in the upper troposphere, however, is not particularly good, thus making it difficult to evaluate model calculations of the important water vapor feedback process in that layer of the atmosphere.

An additional, but smaller, positive climate feedback is the relationship between ice and snow (or its absence) at the Earth s surface and its reflectivity (albedo) of solar radiation. In essence, if ice or snow cover melts, the surface left exposed (ground, vegetation, or water) is generally less reflective of incoming solar radiation. This leads to more absorption of the solar radiation, thus more warming, which then leads to less ice or snow, and so on.

Inclusion of these snow and ice albedo feedback processes in mathematical models of the climate further amplifies the calculated warming response of the climate to increased atmospheric concentrations of CO_2 and other infrared absorbing gases; these feedback processes would also amplify any calculated cooling effects. Other kinds of feedbacks, both positive and negative, result from interaction of land surface properties (e.g., changes of vegetation that lead to albedo and evaporation changes) with climate warming/cooling mechanisms or from changes in CO_2 uptake by the biosphere (*see* Climate Feedbacks, Volume 1).

The major source of uncertainty in determining the climatic response to any atmospheric radiative perturbation arises from the impact of clouds on the radiative balance of the climate system. Clouds are effective absorbers and reflectors of solar (visible plus ultraviolet) and infrared radiation. Their net effect is to cool the planet, but the effect is very small relative to the 33 °C atmosphere/no-atmosphere difference noted above. However, for predicting smaller human-caused climate changes, the effect of clouds becomes crucially important. An increase in thick low clouds mainly acts to reflect more solar radiation and thus would provide a negative feedback to global warming. An increase in thin high clouds mainly adds to the absorption of infrared radiation trying to escape the planet and would thus provide a positive feedback. A relatively small change in cloud microphysical and optical properties could make the sign of the cloud feedback go either way. Which of these changes, if any, would dominate in an increasing CO_2 world? We are not sure. Our inability to answer this question with confidence is the major source of uncertainty in today s projections of how the climate would respond to increasing infrared-absorbing gases. Furthermore, it is not likely this cloud–radiation uncertainty will be sharply reduced within the next 10 years (*see* Cloud- Radiation Interactions, Volume 1).

Although shortcomings in the treatment of clouds dominate the climate modeling uncertainty, other key processes are also in need of improved understanding and modeling capability. A key example is the effect on climate response of human-produced airborne particles (aerosols), which are composed mostly of sulfate (produced by oxidation of the sulfur in fossil fuels) and carbon (from open fires) (*see* Aerosols, Troposphere, Volume 1). Sulfate aerosols strongly reflect solar radiation, producing a cooling effect, whereas carbonaceous aerosols strongly absorb solar radiation, producing a net heating effect. Efforts to reduce the current uncertainty of aerosol effects are limited by inadequate measurements. Even more uncertain are the so-called indirect effects of atmospheric aerosols. By indirect effect we mean the uncertain role the presence of these aerosols plays in the determination of cloud amounts and the optical properties of a cloud s water droplets and ice crystals (*see* Aerosols, Effects on the Climate, Volume 1).

Another key uncertainty lies in modeling the response of the ocean to changed atmospheric greenhouse gas

concentrations. Uncertainties in the oceanic heat uptake rate affect the calculated rate of response of the climate on time scales greater than a decade. The uncertain ocean response also includes the possibility of changed ocean circulation, a potential major factor in shaping regional climate changes.

A frequently overlooked aspect of the human-caused greenhouse-warming phenomenon is its fundamentally long time scales. The current rate of adding to the atmospheric CO_2 amount is slightly more than half a percent per year, and total CO_2 emissions are expected to increase over the next half century (*see* Carbon Dioxide, Recent Atmospheric Trends, Volume 1). The time required for atmospheric CO_2 concentrations to approach twice pre-industrial levels ($2 \times CO_2$) is roughly a half-century to a century from now, a process well underway (CO_2 concentrations are now over 30% higher than pre-industrial levels). Because of the large heat capacity of the oceans, the atmospheric temperature is not expected to respond quickly to the higher CO_2 concentrations (*see* Carbon Cycle, Volume 2). This oceanic heat capacity effect can produce delays in the realized warming on time scales ranging from decades to centuries. Moreover, the deep ocean carries over a thousand years of temperature memory. Thus, after CO_2 concentrations stop increasing, it is expected that it will take many centuries for the atmosphere and ocean to stop warming.

This large ocean-induced resistance to changes in the climate is also a big factor at the other end of the problem. What if we have produced a climate regime we do not like and want our normal one back? The current evaluation of the net atmospheric lifetime of fossil fuel-produced CO_2 is about three-quarters of a century. The net CO_2 atmospheric lifetime is expected to become considerably longer as more CO_2 is added, and the fast-removal CO_2 reservoirs (upper ocean and land surface) become increasingly burdened with CO_2 (Houghton *et al.*, 1992). Thus, a natural draw down of the enhanced atmospheric CO_2 concentration is expected to require multiple centuries. Also, the warmed ocean would add an even longer time to give up its accumulated heat after the atmospheric CO_2 concentrations had begun to return toward appreciably lower levels.

THE ROLE OF CLIMATE MODELS

Over the past four decades, a quiet revolution has fundamentally changed the way that much of the research in climate science works. Earlier, the controlling science paradigm was the interchange between theory and observation concerning the structure and behavior of the climate system. Today, much climate research is driven by the interactions among theory, observation, and modeling. Modeling involves computer-based simulations of various phenomena, based on numerical solutions of the known equations governing the phenomena under investigation. This combined approach (theory, observation, modeling) is now widespread in the physical sciences. It is significant that mathematical modeling of weather and climate literally pioneered this now-widespread new approach to scientific research in the 1950s and 1960s.

Mathematical models of climate can range from simple descriptions of simple processes to comprehensive simulations of the entire dauntingly complex climate system (*see* Models of the Earth System, Volume 1). Models of the coupled atmosphere–ocean–ice–land climate system now lie close to the most complex limit of such models. This very complexity can lead to highly divergent human reactions to them, varying from deeply skeptical to totally uncritical. The truth is far from either of these unscientific characterizations.

Newcomers to the greenhouse-warming problem tend to be unaware of the long and rich history of mathematical modeling of the atmosphere and the ocean. In the late 1940s and early 1950s, simple mathematical models were created to attack the weather forecasting problem. More advanced models were built in the late 1950s and early 1960s (Phillips, 1956; Smagorinsky, 1963) because of a strong research interest in understanding and predicting the circulation of the atmosphere. Shortly thereafter, the first model bearing a strong resemblance to today s atmospheric models was created (Smagorinsky *et al.*, 1965) (*see* General Circulation Models (GCMs), Volume 1). That early model, as well as all of today s models, solved the equations of classical physics relevant for the interacting atmosphere–ocean–ice–land system. These equations are conservation of momentum (Newton s second law of motion), conservation of energy (first law of thermodynamics), and conservation of matter (air, water, chemicals, etc., can be blown around by wind or currents, undergo phase changes, transferred across boundaries, or converted chemically, but the number of atoms of each kind remains unchanged).

The modeling approach thus provides high potential for fundamental tests of applications of these theoretical first principles. Such modeling appears deceptively simple. These equations are taught in high school physics. There are some daunting challenges, however. When atmosphere, land, ocean, and ice processes are coupled and applied to moving (and deforming) fluids such as air and water, the model equations form complex systems that are intrinsically nonlinear and can thus exhibit surprisingly counterintuitive behaviors. Moreover, their solution in a climate model requires a reasonably fine-scale grid of computational points all over the atmosphere–ocean–ice–land system, ideally, well more than a million such points. In addition, important small-scale processes such as atmospheric moist convection (e.g., thunderstorms) and turbulent dissipation (smaller-scale mixing of heat, momentum, etc.) remain formidably difficult to incorporate on a first-principles basis. Often, such processes are included statistically (parameterized) through use of field observations.

Finally, no meaningful steady-state (time-independent) solutions of the governing equations can solve directly for the average climate. In effect, the average climate in such a model must be described as a statistical equilibrium state of an unstable system that exhibits important natural variability on time scales of hours (thunderstorms), days (weather systems), weeks to months (planetary-scale waves/jet-stream meanders), years (El Niño), and decades to centuries (ocean circulation variations and glacial ice fluctuations). A complete model must calculate all of these time-varying processes directly. Clearly, models of such a large, complex, and time-varying system are intrinsically computer intensive. Fortunately, today s supercomputers are nearly 10 000 times faster than those of 30 years ago. Because of today s widespread availability of relatively inexpensive computer power, the number of fully coupled atmosphere–ocean–ice–land climate models in the world has increased from a few in the early 1980s to roughly 10 independently conceived models today. Roughly another 20 models are essentially based on these 10 models.

Over the last half century, use of these kinds of physically based mathematical models has resulted in major improvements in the science of weather forecasting. Sharp forecasting skill improvements have been achieved through finding the useful short-term predictability in a fundamentally chaotic system (which means that the details of weather variations become essentially unpredictable after a sufficient lapse of time, say a couple of weeks) (Lorenz, 1963) (*see* Chaos and Predictability, Volume 1).

Recently, it has become clear that slower variations of the coupled atmosphere–ocean–ice–land system provide potential for finding useful predictability on time scales longer than the couple of weeks characteristic of individual weather systems. The most visible example is the realization that El Niño events, which produce warming in the tropical eastern Pacific Ocean, may be partially predictable a year or so in advance under certain circumstances (Cane *et al.*, 1986) (*see* El Niño and La Niña: Causes and Global Consequences, Volume 1). The existence of such a predictable spot of warm eastern Pacific Ocean water provides the basis for a nearby improvement of the prediction of seasonal weather anomalies (e.g., a wetter than normal California winter).

The existence of limited extended-range predictive potential in the climate system leads to obvious questions about such models validity for projecting systematic changes in the statistical equilibrium climate (say a 20-year running average) resulting from the inexorably increasing concentrations of infrared-active gases currently underway. Projections for greenhouse warming scenarios depend on a number of physical processes that are subtle, complex, and not very important to weather prediction. However, people outside the climate field have frequently said that climate models are ill posed and irrelevant because they attempt to forecast climate behavior that is well beyond the limits of deterministic predictability. Thus, they assert that, if one cannot predict weather more than a week or two in advance, the climate change problem is basically impossible. Such a statement is scientifically incorrect. The weather prediction problem is essentially an initial-value problem, in which the predictability of interesting details (i.e., weather) is fundamentally limited by uncertainty in initial conditions, model errors, and instabilities in the atmosphere itself. In contrast, climate change projections are actually time-varying boundary value problems, where the objective is to determine the changes in average conditions (including the average features of the evolution toward the new equilibrium) as the planet is heated or cooled by newly added processes (e.g., increased concentrations of CO_2 or sulfate particles).

The differences between weather and climate models are further instructive when one considers how their strengths and weaknesses are evaluated. Thanks to massive amounts of weather and climate data, both kinds of models can be evaluated by careful comparison with data from the real world. In practice, however, the approaches to improving these superficially similar models are very different. The skill of weather forecasting models is evaluated by comparing model-based forecasts, which are started up from real data on a given day, with what actually happens hours to weeks later. Interestingly, one of the key problems with such weather models is that, in longer-range forecasts (days to weeks), they can easily reject their initial conditions by drifting toward a model climate that is quite different from that of the real data that was used to start up the detailed forecast calculation. In effect, such a drifting weather forecast model is deficient in the climate that it would produce if released from the constraints of its starting data.

In sharp contrast, a climate model has the responsibility of properly simulating the time-averaged climate for today s conditions. In this case, the focus of the scientific inquiry is quite different from weather forecasting. Here, attention is directed toward proper simulation of the statistics of climate, such as the average daily and seasonal temperature cycles forced by the variations in incoming solar radiation, the average number and intensity of extratropical cyclones, the locations of deserts and rainy areas, the strength and location of jet streams and planetary waves, the fidelity of El Niño simulation, the location and characteristics of clouds and water vapor, the strength and location of ocean currents, the average magnitude and location of snow accumulation and snow melt, and, finally, the amplitudes and time scales of the natural variability of all of these on a wide range of time scales (days to centuries) (*see* Model Simulations of Present and Historical Climates, Volume 1).

Achieving all of this in a climate model is a daunting task because of the enormous wealth of phenomena in the climate system, many of which exist on scales smaller than can be resolved by current models. This leads to the need for judicious science-based or *ad hoc* tuning and/or adjustment of various poorly defined processes (such as cloud properties, or the fluxes of heat between atmosphere and ocean) to improve the model s agreement with observed climate statistics. Such tunings and adjustments are widespread, especially for the global-mean radiative balance, and are often done to ensure that the model agrees with the various time-mean features of the climate. If this is not done, a coupled model started up with today s climate conditions will tend to drift toward a less realistic climate.

These practices have been criticized as evidence that climate models have no credibility for addressing the greenhouse-warming problem. Interestingly, such tunings and adjustments (or lack thereof) may have little to do with the ability of a model to reduce its fundamental uncertainty in projecting human-caused climate change. Recall that the key uncertainties highlighted above (water vapor, cloud, and ice-albedo feedbacks) revolve around how such properties might change when greenhouse gases are added to the atmosphere. This is a set of challenging modeling problems that cannot be evaded by judicious model tunings or adjustments. Likely to prove much more fruitful in the long run would be improved fundamental modeling of the key processes that govern the most important climate feedbacks as the atmospheric CO_2 concentration increases (e.g., clouds, water vapor, ice, ocean circulation). Many of these modeling issues are addressed in the major scientific assessments prepared by the Intergovernmental Panel on Climate Change (IPCC) (Houghton *et al.*, 1990, 1995, 2001).

Thus, climate models are imperfect tools for making projections of climate change. Does this mean we should shift our focus to other tools? Definitely not! Unfortunately, there exist no good alternatives. Statistically based models that use historical data are possible alternatives, but they are of marginal validity, mainly because the recent Earth has apparently never experienced the rate of global-mean warming expected to result from the current run up of atmospheric greenhouse gas concentrations. In this sense, the large, but very slow, global-mean climate excursions of the past geological epochs are instructive, but they are far from definitive as guidelines or analogs for the next century.

The above considerations make it clear that there is no competitive alternative to coupled climate models for projecting future climate states and how they might unfold. The physically based climate models have the huge scientific advantage of being fundamentally grounded in known theory that can be evaluated against all available observations. There are indeed reasons to be skeptical of the ability of such models to make quantitatively accurate projections of the future climate states that will result from various hypothesized future greenhouse gas concentration scenarios. Fortunately, the weak points of such climate models can be analyzed, evaluated, and improved with properly focused, process-oriented measurements, complemented by well-posed numerical experiments with various formulations of the climate models. In short, the use of such climate models allows a systematic approach to close the gap between theory and observations of the climate system. No alternative approach comes close.

THE ROLE OF CLIMATE DATA

The availability of climate data in many forms is crucial in the quest to understand, simulate, and predict the behavior of the climate system, how it has changed to date, and how it might change in the future. Such data provide the basics for our characterizations of the time-averaged climate states of various statistics of temperature, pressure, wind, water-vapor amounts, cloudiness, and precipitation as a function of geographical location, time, and altitude. Most importantly, such data provide invaluable information on the natural variability of climate, ranging from seasons to decades.

These data sets have empowered important direct insights on how the climate system works. For example, the observed average daily and seasonal ranges of mean temperature provide important evaluations of our theoretical understanding of how the climate changes in response to changed radiative circumstances (e.g., day to night, summer to winter). On much longer time scales, indirect inferences (or proxy measures such as ice core or tree ring data) provide valuable information on how ice ages and warm epochs appear to depend sensitively on subtle changes to the heating of Earth due to seemingly small variations in the precession of Earth s orientation toward the Sun and in Earth s elliptical orbit around the Sun. Interestingly, the onset of ice ages and their terminations appear to respond more sensitively to these small solar heating changes than are calculated by our current climate models. For example, the ice core records show that ice-age atmospheric CO_2 concentrations decline very slowly as the climate cools, seemingly, a positive ocean–biological feedback effect that is slow enough that we would not expect it to be of significant importance over the next century. Unfortunately, proxy observations of prehistoric climates are still ambiguous enough that they do not justify any confident conclusions that our current climate model projections may be underestimating or overestimating the century-scale global temperature increase due to added greenhouse gases.

Recently, climate data have been the focus of a substantial shift in emphasis. As the climate models have become more mature, we have gained more confidence in their

ability to simulate a future climate change by evaluating their capabilities to diagnose the presence of human-caused climate change already present in the climate record to date. These accelerated efforts have already yielded important insights about how the climate system appears to be responding to added greenhouse gases plus sulfate and carbon particles over the past century or so, as will be discussed later.

Most importantly, however, recent research has revealed important deficiencies in our capability to diagnose such responses accurately. The best models tend to agree in the broadest sense, but can differ substantially when very specific and regional processes are examined. The available data sets have deficiencies that can inhibit the capability to diagnose long-term changes in the climate, particularly so in undersampled regions. Fortunately, our measurements of the global trends in greenhouse gas concentrations and their associated direct infrared radiative effects are very well known. Other potential perturbations of the climate balance, such as effects of sulfate and carbon particles, and changes in heat energy from the Sun, are much more poorly known.

An additional complication arises in our efforts to determine how the climate responds to added greenhouse gases and aerosols. The climate system is naturally noisy, and thus tends to obscure its systematic changes until such changes are larger in magnitude than the background climate noise. By climate noise, we mean the natural fluctuations in, say, the atmospheric temperature even when no external effects are influencing the climate system. For example, the global-mean surface temperature of a climate model or the real world can change by a couple of tenths of a degree Celsius over a 30-year period just due to the interactions between the atmosphere and ocean. If one is looking at a magnitude of expected systematic change of a few tenths of a degree Celsius over a comparable period, it is not surprising that the conclusion can be a bit blurred due to the natural fluctuations. This is why very long climate records are so invaluable in the efforts to understand and diagnose climate change.

Fortunately, careful examination of the characteristics of the systematic changes in climate models, in the presence of their own internally generated climate noise, can also give sharpened insights into how climate change in the real world might be expected to emerge and slowly become ever clearer as time progresses. As the climate-warming signal continues to grow, the interpretive confusion caused by natural climate fluctuations is expected to diminish. In addition, such models can help sharpen our understanding of what capabilities are required to measure and monitor the actual evolving systematic changes in the climate. Without improvements in our current climate measurement systems, our ability to detect climate change and to attribute cause and effect to such changes will remain insufficient.

MODEL HINDCASTS OF THE 20TH CENTURY WARMING

In the previous two sections, we argued that combined use of climate models and climate data is an inseparable requirement for understanding climate change. This strong interdependence is quite evident in the challenge of understanding the climate warming observed over the 20th century.

It is now well known that the global mean surface–air temperature has increased considerably, roughly 0.6 °C over the 20th century, with a substantial amount of that warming occurring over the past couple of decades (Nicholls *et al.*, 1996). Is this observed warming due to natural climate variability, volcanoes, or changes in the Sun s radiant energy? Or is it due to the CO_2 and other greenhouse gases that we have added to the atmosphere? (*see* Climate Change, Detection and Attribution, Volume 1) Why should we be so focused on what has already happened?

The answers to these questions are fundamental to evaluating the credibility of the projections we will present in the next section on expected future climate warming, the warming that we expect to be considerably larger than that measured over the 20th century. An evaluation of the past data from a modeling and theory perspective is the only direct way we currently have of evaluating the skill, or lack thereof, of model-based projections of future climate warming. In recognition of this, the climate community has developed estimates of the past radiative forcings. Because some of the 20th century climate forcings were not directly measured, there is important uncertainty associated with these estimates. Mitchell *et al.* (1995) were the first to perform this type of integration using a physically comprehensive climate model. They used best-available estimates of the greenhouse and sulfate aerosol forcing of climate change. Their experiment has become a standard that a number of groups have followed, including our own Geophysical Fluid Dynamics Laboratory (GFDL) group. This forcing scenario, while highly simplified, captures the essence of much of the known changes in radiative forcing of climate over the past century. This scenario purposely ignored the forcings due to volcanic aerosols, changes in the Sun s luminosity, and ozone depletion. Recently, several groups have made integrations that included some or all of these other forcings (Santer *et al.*, 1996; Tett *et al.*, 1999). However, the message below is not noticeably changed by these presumably more realistic forcing integrations.

For sake of simplicity, we present only the results from the GFDL coupled climate model. Figure 1 shows the Jones *et al.* (1999) compilation of observed global-mean surface–air temperature (solid line) from 1860 through 1999 and an ensemble of six (GFDL) model runs (Delworth and Knutson, 2000; using the Mitchell *et al.*, 1995 scenarios described above). This figure reveals that the apparent magnitudes of the observed natural variability are

comparable with the model-simulated magnitude of natural climate variability. Also, the model captures the run up of global-mean temperatures in the 20th century rather well. Experiment three of the six-member ensemble even captures the observed increase of temperatures from 1915 to 1945. This result suggests that the observed 1915–1945 increase may be attributable to a combination of greenhouse gas increases and to a particular form of natural variability operative during that period, namely, the presence of a natural-variability-induced warming in high northern latitudes in the observations and in ensemble member number three (Knutson *et al.*, 1999). Other factors not included in the model experiments, such as total solar irradiance variations, may also have contributed to the 1915–1945 warming period.

This line of reasoning is quite plausible, although it falls a bit short of the definitive proof about global warming that some would require of the 20th century information. The picture gets a bit fuzzier when one considers the uncertainties in the IPCC greenhouse gas plus sulfate particle scenarios used in these calculations. However, we argue that this retrospective exercise serves as a good test for projections over the next century, when one considers that there is still a complete lack of a viable natural-variability-only hypothesis to explain the observed 20th century warming.

We further argue that the additional model results of 20th century warming from Mitchell *et al.* (1995) and Haywood *et al.* (1997) provide a further check on the credibility of these models for use in projections of future climates. It must be cautioned, however, that these 20th century tests

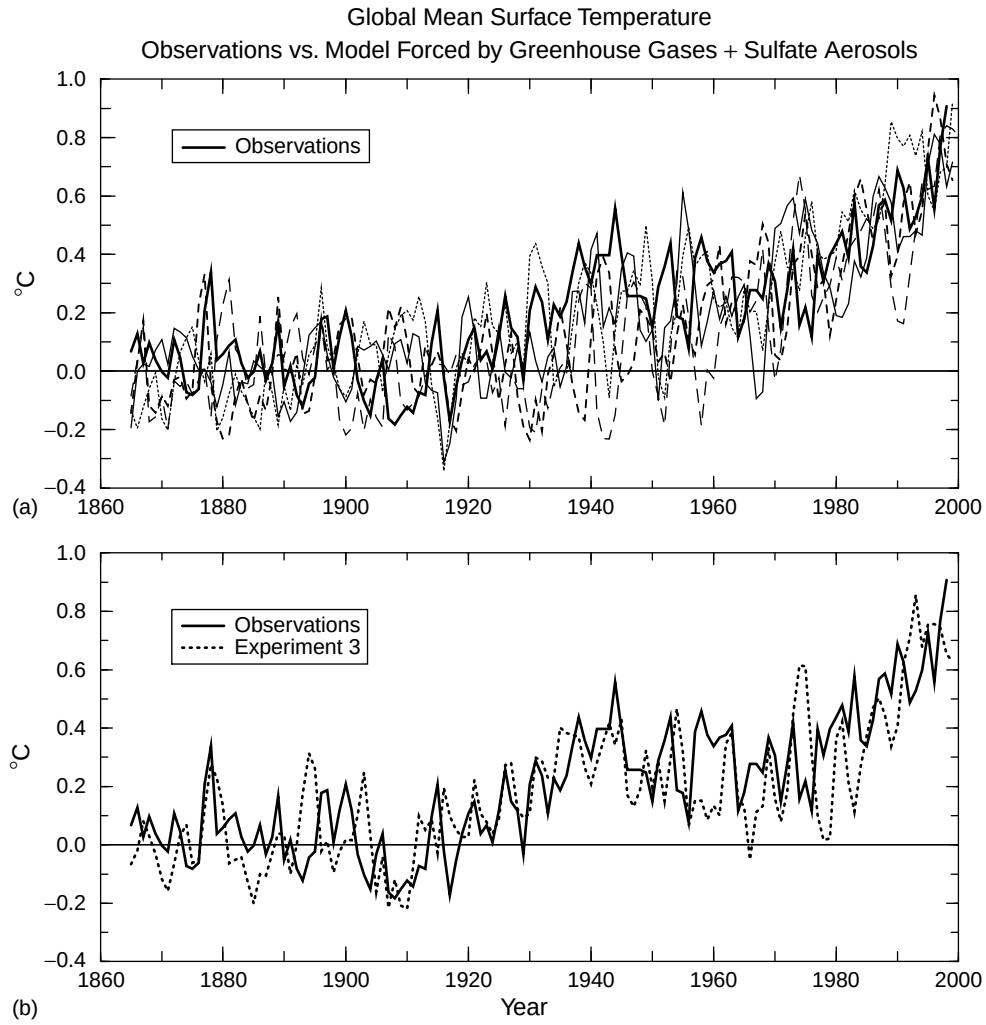


Figure 1 Time series of global mean surface temperature, 1865 to present day. The thick line in both panels are the observations. The thin lines are the various members of five model integrations. Four model integrations are shown in (a). The time series from experiment 3 is shown in (b). Over land, surface–air temperature is used to compute the global mean. Over the oceans, sea surface temperature is used. Units are deviations from the 1880–1920 average. In constructing the global averages, the model output is sampled only at times and locations where observations are available

do not tightly constrain model projections of future climate warming. It is still possible, for example, to imagine that the world s best model-based evaluations of the expected 3 °C global-mean climate warming due to $2 \times CO_2$ could be high or low by as much as 1.5 °C. We do estimate, however, that the likelihood of model errors of this magnitude is now less than 10% (Mahlman, 1997). The main basis for this assertion is that we currently lack either observational insights or published self-consistent theoretical arguments that can produce departures of the models warming levels as much as 1.5 °C from the average warming levels obtained by the most tested and evaluated models.

We thus assert that future climate change projections using such models are likely to yield results that will resemble what will happen in the real world with sufficient credibility, that the global warming issue will eventually be taken very seriously.

MODEL PROJECTIONS FOR THE 21ST CENTURY

To project changes in climate over the 21st century, we need to project both the changes in the radiative forcing of the planet and the climate response to those changes. The changes in the future radiative forcing are controlled by the emissions of greenhouse gases, sulfate aerosols, and other radiatively active substances. The future emissions of these substances depend on global population growth, technological developments, CO_2 emissions mitigation decisions, economic changes, and a host of other societal decisions. These factors all have a great deal of uncertainty associated with them. In 1992, the IPCC developed a set of emission scenarios by considering reasonable estimates of many of these factors. Of the 1992 emission scenarios, the IS92a scenario (Houghton *et al.*, 1992) has become a *de-facto* standard used by the modeling community to investigate future climate changes. In 2000, the IPCC developed a new set of emission scenarios, published in the so-called Special Report on Emission Scenarios (SRES) scenarios (Nakicenovic *et al.*, 2000). The SRES scenarios are grouped into four families where each family has an associated story line for the various assumed factors that affect the emission rates. The IS92a scenario, and many of the new SRES scenarios, produce increases in radiative forcing during the 21st century that lie between those of an effective 1% $year^{-1}$ atmospheric CO_2 increase and an effective 0.5% $year^{-1}$ CO_2 increase.

In the discussion below, the model projections shown are based on models forced using either the IPCC SRES A2 or B2 scenarios, or an idealized 1% $year^{-1}$ scenario leading to doubling ($2 \times CO_2$) or quadrupling ($4 \times CO_2$) of the atmospheric CO_2 concentration relative to pre-industrial levels, then held constant thereafter. For the remainder of this section, we purposely ignore the uncertainty associated with the forcing. Although the uncertainty in the future radiative forcing remains substantial for longer time scales, alternative forcing scenarios can be usefully explored by interpolating global-average changes logarithmically between the $2 \times CO_2$ and $4 \times CO_2$ model experiments. The time-dependent model response uncertainty for a given scenario arises from uncertainties associated with estimates of the actual climate response, which are mainly related to the changes in radiative properties associated with clouds (both in their distribution and changes in their direct radiative properties), and with uncertainties associated with the oceanic heat uptake rate.

Oceanic heat uptake rate is important for determining the climate changes during the time when the radiative forcing is changing. After the radiative forcing becomes steady and as the ocean temperature comes progressively nearer to equilibrium with that forcing, the oceanic heat uptake rate becomes less and less important in determining the final climate change. During the time when the radiative forcing is changing in time, a model with a relatively large (small) oceanic heat uptake rate will have a smaller (larger) surface–air temperature response and a larger (smaller) amount of sea level rise (considering only thermal expansion) in comparison with the typical climate model, all other factors being equal. Because of their large heat capacity, the oceans tend to delay the surface–air warming caused by the increases in the radiative forcing. In addition, the strength of the various feedbacks (water vapor, snow–ice albedo, etc.) is dependent on the magnitude of the surface temperature changes (Hansen *et al.*, 1984). Therefore, the still uncertain oceanic heat uptake rate will play an important role in determining the magnitude of the time evolving climate warming over the 21st century.

As the climate responds to the increasing atmospheric concentrations of greenhouse gases, different parts of the climate system warm at different rates. This is due to the differing thermal heat capacities characterizing the climate system. The land surface and atmosphere have a relatively small heat capacity and a response time scale of about 1 month. The uppermost part of the ocean (the so-called oceanic mixed layer) has a moderate heat capacity and a response time scale of about 10 years, while in the deepest parts of the ocean, the full warming response requires roughly 1000 years.

To illustrate the various responses of the climate system to increasing greenhouse gases, we will show only the surface–air temperature response and the sea-level response. The initial surface–air temperature response has a time scale of about 10 years, indicating that it is closely related to the response time of the upper part of the ocean (mixed layer).

In the model simulations, the calculated global-mean surface–air temperature rises by about 2.25 °C between the years 2000 and 2100 when using the SRES B2 scenario

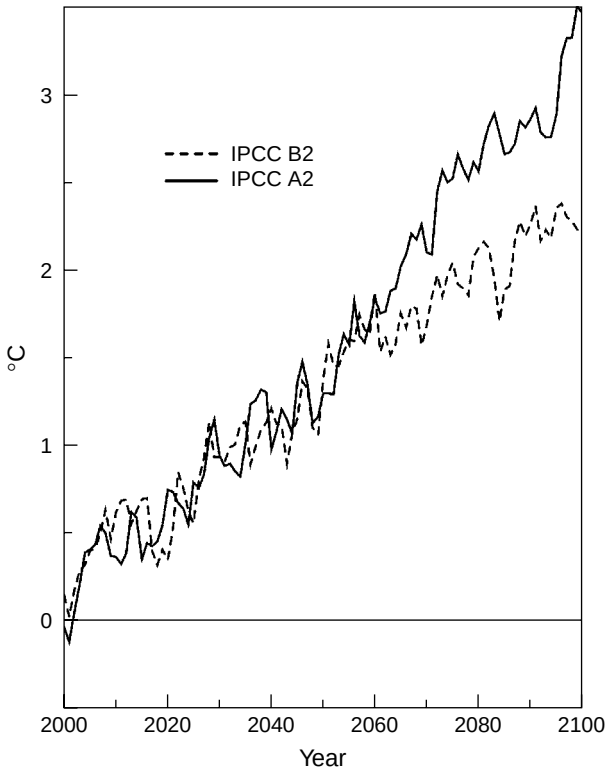


Figure 2 GFDL model calculations of global mean surface–air temperatures for two IPCC SRES radiative forcing scenarios: A2, an essentially business as usual scenario and B2, a middle of the road CO_2 emissions mitigation scenario. Temperatures are expressed as a departure from the temperature typical for the period around the year 2000

in the GFDL medium-resolution R30 (2.25° latitude by 3.75° longitude) coupled model; the calculated 21st century warming is 3.5 °C for the A2 scenario (Figure 2). The magnitude of this response is similar to that found in other coupled models (Houghton *et al.*, 2001). This calculated future warming is much larger than the roughly 0.6 °C warming observed and calculated for the past century (see the previous section). Figure 3 indicates that the higher Northern Hemisphere latitudes warm more than the lower latitudes. The land areas warm more than the adjacent oceans. There are minima in the warming in the North Atlantic Ocean near Greenland/Iceland and in the Circum-Antarctic Ocean. These general characteristics can be found in both the SRES A2 and B2 distributions (Figure 3).

The exact details (i.e., the specific amount) of the expected warming remain rather uncertain. Over the continents, the surface–air temperature warming is influenced by local changes of aerosol emissions, clouds, vegetation, etc. However, the gross patterns, as noted in the preceding paragraph, are not as uncertain and tend to be more related to changes in the global-mean surface–air temperature rather than the local differences in the forcings.

As the climate warms, an intensified hydrological cycle results, i.e., there is more global-mean precipitation and evaporation. This is the result of the warmer atmosphere being able to hold more water vapor. More water vapor in the atmosphere results in increased poleward water vapor transport, so that high-latitude precipitation increases substantially. Over most mid-continental areas, the warmer temperatures cause the soil to become drier in the summer, while the soil becomes wetter in winter due to increased precipitation.

The warming of the atmosphere makes the modeled oceans more stable at all latitudes. In low latitudes, the ocean is made more stable due to the strong heating of the upper layers. In high latitudes, where much of the vertical mixing in the ocean takes place, the precipitation-caused increase in lower-density fresh water at the ocean s surface also makes the ocean more stable. This increased stability in the world s ocean further inhibits vertical mixing, which can act to change the ocean circulation. The warming of the atmosphere and sea surface also allows more latent-heat energy to be available in smaller-scale storms. Studies have indicated that, in the warmer world, the strongest hurricanes and typhoons would tend to get even stronger (Knutson *et al.*, 1998). Finally, surface–air temperatures are expected to warm more during nighttime than in daytime (Houghton *et al.*, 1995).

These changes in the physical climate system will lead to changes in other systems as well. For example, the modeled heat index (the apparent temperature due to a combination of sensible temperature and moisture felt by the human body) greatly increases in the warm, humid regions of the planet (Delworth *et al.*, 1999). The combination of heat and moisture increases makes the summertime heat index increases in the moister climate roughly 50% larger than for just the sensible temperature increases.

The final part of this discussion centers on how the climate may be expected to change over the 21st century and centuries beyond. Due to the long response times in the deep ocean, the warming climate is still well short of equilibrium with today s radiative forcing. This means that, without any further increase in the radiative forcing, Earth s climate would continue to warm over the next few centuries as the entire ocean slowly warms to reach equilibrium with the warmer atmosphere. This illustrative calculation of future warming, which assumes a situation that is effectively impossible because it would allow for no changes from today s radiative forcing due to added greenhouse gases and aerosols, is defined here as the absolute future warming commitment beyond the 0.6 °C that has already been experienced. There are corresponding future commitments for all other climate variables as well. Hansen *et al.* (1984) concluded that the magnitude of the absolute future warming commitment beyond 1984 is about 1 °C. This is significantly larger than the

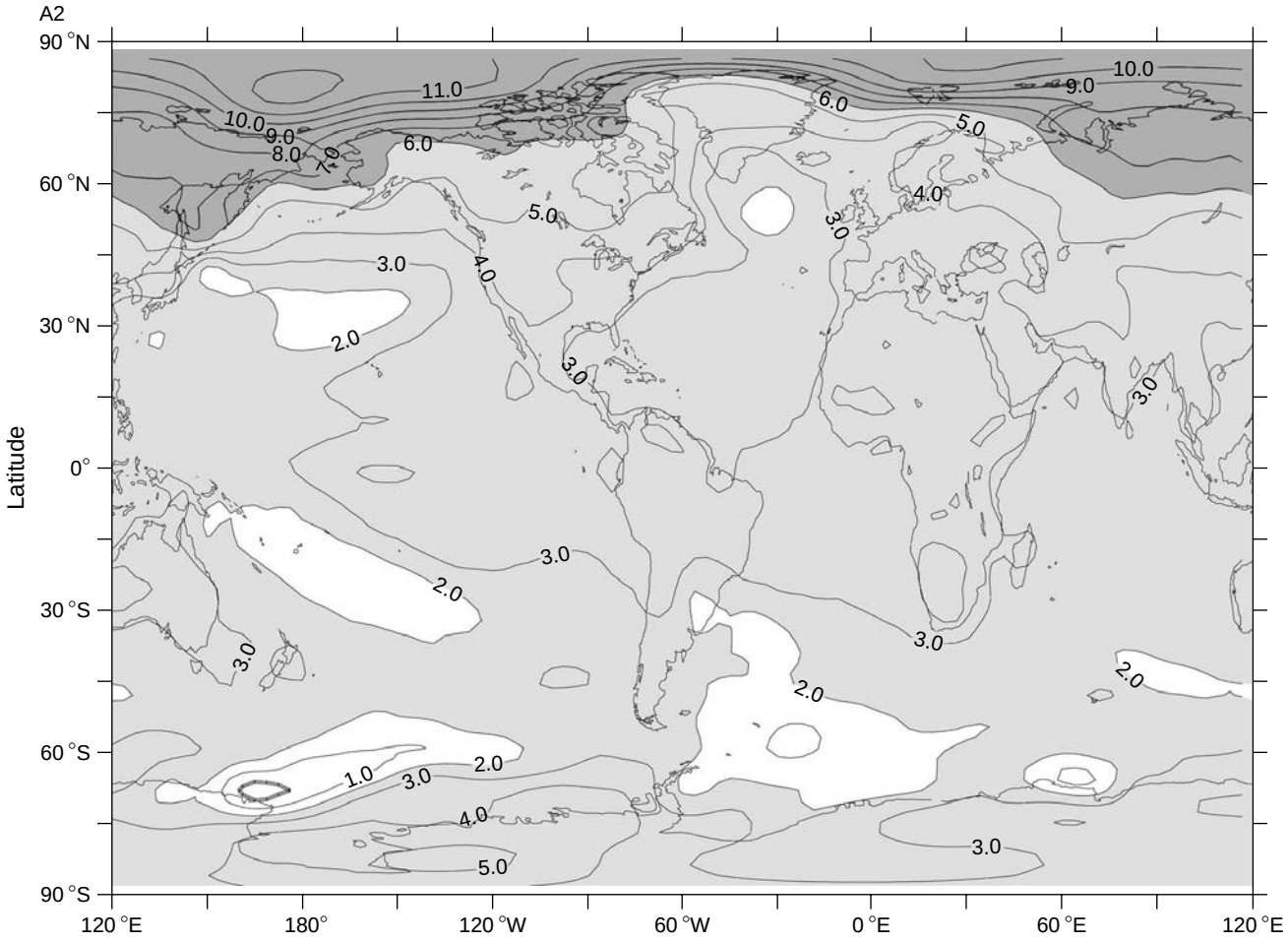


Figure 3 GFDL model-calculated regional distribution of surface–air temperature changes over the 21st century using the IPCC SR A2 and B2 scenarios described in Figure 2. The light shading range is from 2 to 6 °C. The dark shading covers areas greater than 6 °C. The contour interval is 1 °C

observed warming during the 20th century, roughly 0.6 °C. Remember that these absolute future warming commitment calculations implicitly assume very small future emissions, a nearly impossible scenario, but they serve to illustrate the major societal challenges required to keep future warming amounts from being much larger than the Earth has experienced to date.

As mentioned earlier, sea level rise has a very long response time scale due to its close relationship to the deep-ocean temperature. During the 21st century, sea level is projected to rise by 9–88 cm, due mainly to the thermal expansion of the sea water as it slowly warms (Houghton *et al.*, 2001) (see also Figure 4). Generally for a given mass of water, warmer water occupies a larger volume than colder water. This thermal expansion of sea water will probably be the largest contributor to sea level rise over the next century, but we cannot discount the possibility that melting land ice could also contribute substantially (*see* Sea Level, Volume 1).

It is unclear what the net effect on sea level will be from changes in the large Antarctic and Greenland ice sheets. The general warming may cause parts of the ice sheets to melt at the lower elevations. However, most of the ice sheet is well above sea level, where the initial likely effect of the warming may be to increase snowfall. Therefore, the net effect of changes in Antarctic and Greenland ice volume on sea level could be small over the 21st century. Beyond 2100, however, the story begins to change. At some point in the future, if the climate continues to warm, substantial parts of these great ice sheets could melt. The magnitude of the sea level rise estimated from melting all of the Greenland ice sheet is about 7 m, while the West Antarctic ice sheet would contribute about 4–5 m.

Finally, it is useful to ask what happens to the climate after the radiative forcing becomes constant (i.e., after atmospheric concentrations of human-produced radiatively active substances stop changing). Typically, in stabilization

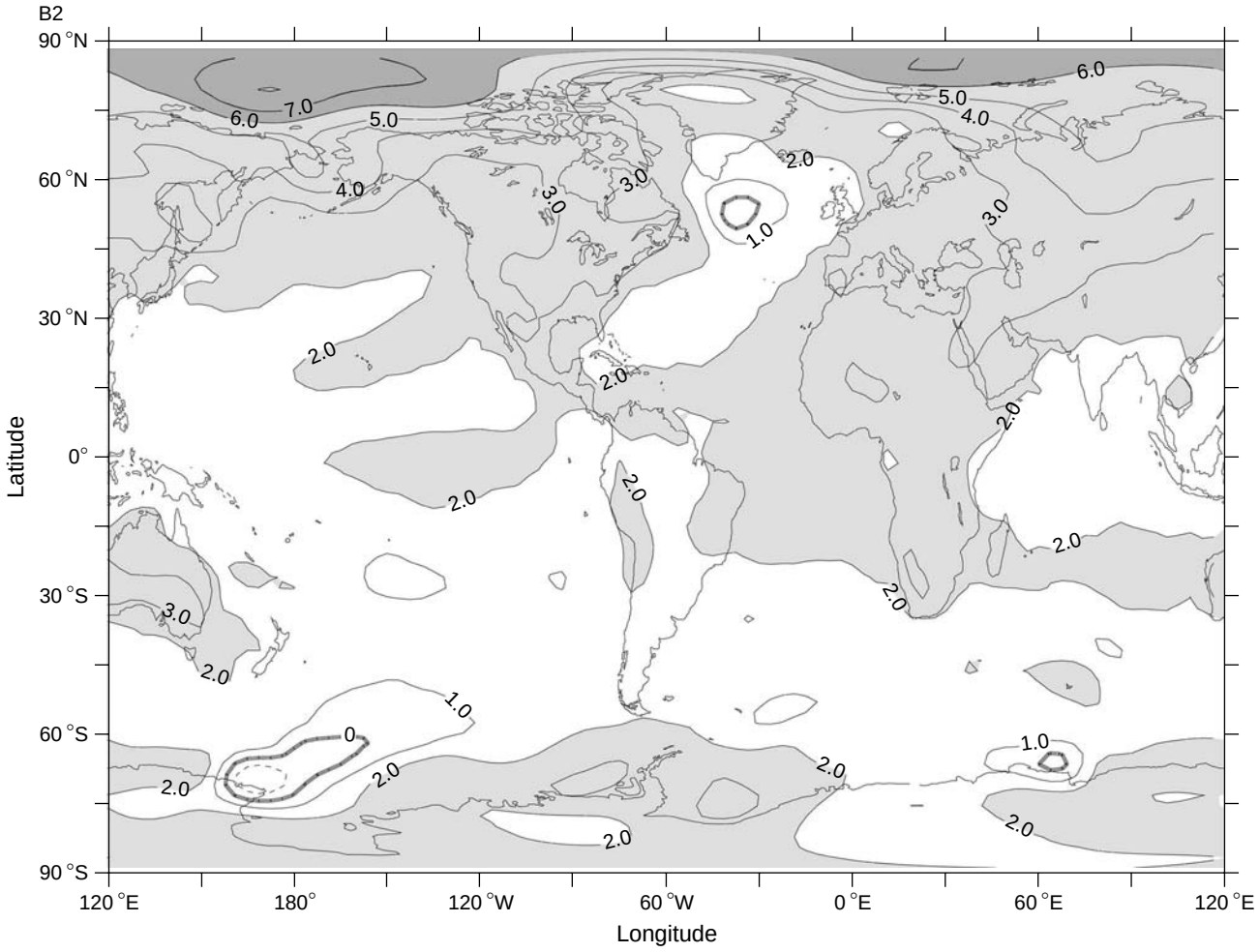


Figure 3 (*Continued*)

integrations, the atmospheric CO_2 concentration increases up to a certain value (e.g., $2 \times CO_2$) and then is hypothesized to remain constant for the remainder of the integration. Because this type of integration involves integrating the model for very long time periods (at least several centuries), only a few integrations have been performed using coupled climate models. Furthermore, no standard emission scenarios have been used for forcing these model runs and most have used idealized stabilization values ($2 \times CO_2$, $3 \times CO_2$, or $4 \times CO_2$, for example). In these integrations, simple CO_2 changes are used to represent the radiative forcing changes of all of the greenhouse gases and their aerosol offsets (equivalent CO_2). Model experiments, where the coupled system is allowed time to reach equilibrium with the radiative forcing, clearly show the response times of its various components (Manabe and Stouffer, 1994).

Even after the radiative forcing becomes constant, the calculated surface–air temperatures continue to increase for many centuries (Figure 4). The rate of warming after stabilization is relatively small (• 0•3 °C century^{-1}). However, the total warming after the radiative forcing stabilizes can be significant (• 1 °C) because the warming continues for many centuries.

The slow rate of surface–air temperature increase occurs as the heat anomaly slowly penetrates to depth in the ocean. The rate of penetration is dependent on a model s simulation of ocean overturning and treatment of small-scale vertical mixing processes. It appears that much of the modeled difference in the oceanic heat uptake are due to differing treatments of ocean mixing, thus explaining much of the difference in the modeled rates of surface warming.

In these stabilization experiments, sea level continues to rise for more than 1000 years after the radiative forcing becomes constant (just due to thermal expansion of sea water). The total sea level rise computed in this and other models is quite large, nearly 2 m for $2 \times CO_2$ and nearly 4 m for $4 \times CO_2$. The potentially large additional effects of melting ice sheets have not been included here because of the large uncertainty in their modeled simulations.

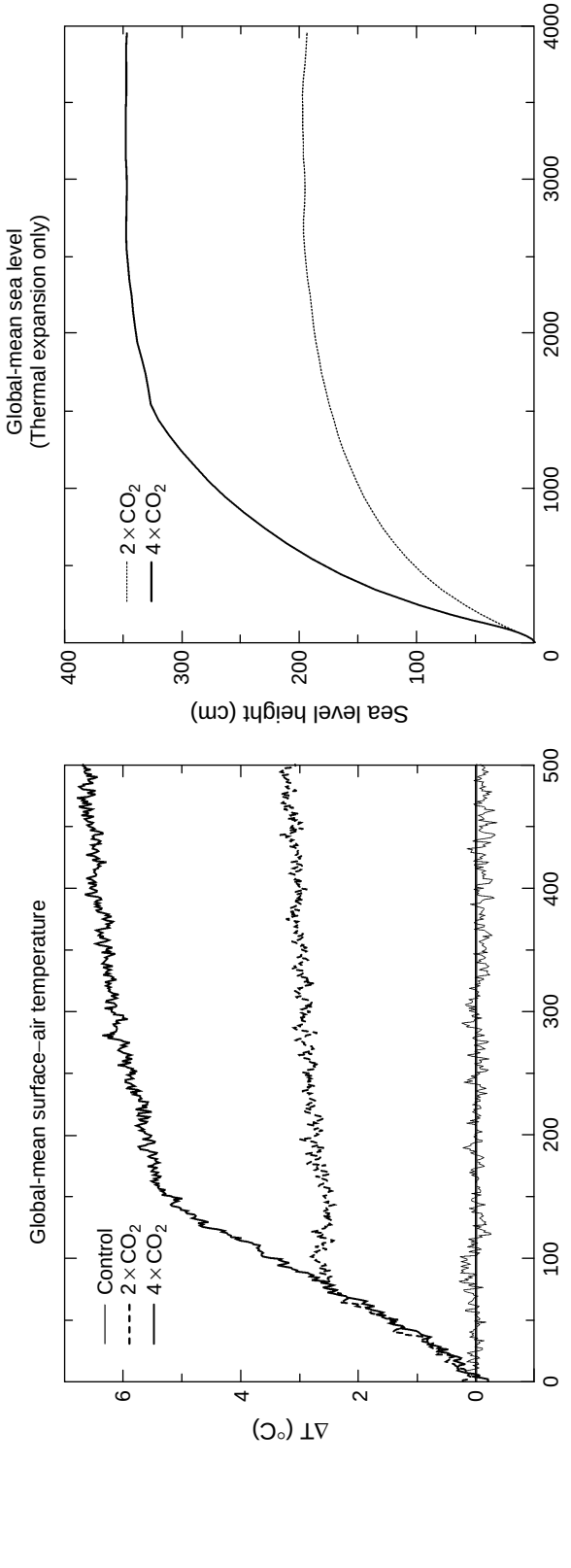


Figure 4 Three GFDL model experiments using idealized forcings. The first (thin solid line) assumes no forcing changes for a model that is in equilibrium. The second (dotted lines) assumes a 1% per year increase of atmospheric CO_2 up to a doubling over pre-industrial levels. The third (thick solid lines) assumes a 1% per year increase of CO_2 up to a quadrupling of atmospheric CO_2. (a) Global-mean surface–air temperature. (b) Global-mean sea level (assuming sea level rise is due to thermal expansion only)

The totality of all the observations of the climate system, the variety of coupled climate models used, and the still-notable lack of viable alternative hypotheses to the model calculations, all lead to a simple conclusion. The warming of the climate due to human-emitted greenhouse gases and aerosols is a very real phenomenon that will not go away. The specific details of the expected climate changes remain elusive. However, the major ways that the climate is expected to change are rather well framed out. A significant amount of climate warming is already committed in the 20th and 21st centuries. Very large climate changes can still be prevented if global society does eventually decide to reduce emissions of CO_2 and other greenhouse gases substantially. Such decisions are fundamentally social and political, thus lying well beyond the realm of the climate-change science presented here.

ACKNOWLEDGMENTS

The authors are indebted to Syukuro Manabe because of his inspirational leadership and mentoring to both of us on this still daunting challenge of climate change. We are grateful to Anthony Broccoli, Steven Carson, Thomas Knutson, Michael MacCracken, Ted Munn, John Perry, and Betty Williams for their valuable improvements to the manuscript.

REFERENCES

Arrhenius, S (1896) On the Influence of Carbonic Acid in the Air Upon the Temperature of the Ground, *London Edinburgh Dublin Philos. Mag. J. Sci.*, 41, 237–276.

Cane, M, Zebiak, S E, and Dolan, S C (1986) Experimental Forecast of El Niño, *Nature*, 321, 827–832.

Delworth, T L, Mahlman, J D, and Knutson, T R (1999) Changes in Heat Index Associated with CO_2-induced Global Warming, *Clim. Change*, 43, 369–386.

Delworth, T L and Knutson, T R (2000) Simulation of Early 20th Century Global Warming, *Science*, 287, 2246–2250.

Hansen, J, Lacis, A, Rind, D, Russell, G, Stone, P, Fung, I, Ruedy, R, and Lerner, J (1984) Climate Sensitivity: Analysis of Feedback Mechanisms, Climate Processes and Climate Sensitivity, *Geophys. Monogr. Am. Geophys. Union*, 19, 130–164.

Haywood, J M, Stouffer, R J, Wetherald, R T, Manabe, S, and Ramaswamy, V (1997) Transient Response of a Coupled Model to Estimated Changes in Greenhouse Gas and Sulfate Concentrations, *Geophys. Res. Lett.*, 24, 1335–1338.

Houghton, J T, Jenkins, G J, and Ephraums, J J, eds (1990) *Climate Change: The IPCC Scienti c Assessment Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, 1–364.

Houghton, T R, Callander, B A, and Varney, S K, eds (1992) *Climate Change 1992: The Supplementary Report to the IPCC Scienti c Assessment*, Cambridge University Press, Cambridge, 1–200.

Houghton, J T, Meira Filho, L G, Callender, B A, Harris, N, Kathenberg, A, and Mackell, K, eds (1995) *Climate Change 1995: The Science of Climate Change*, Cambridge University Press, Cambridge, 1–572.

Houghton, J T, Ding, Y, Griggs, D J, Noguer, M, van der Linden, P J, Dai, X, Maskell, K, and Johnson, C A, eds (2001) *Climate Change 2001: The Scienti c Basis*, Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, 1–881.

Jones, P D, New, M, Parker, D E, Martin, S, and Rigor, I G (1999) Surface Air Temperature and its Changes Over the Past 150 Years, *Rev. Geophys.*, 37, 173–199.

Knutson, T R, Tuleya, R E, and Kurihara, Y (1998) Simulated Increase of Hurricane Intensities in a CO_2-warmed Climate, *Science*, 279, 1018–1020.

Knutson, T R, Delworth, T L, Dixon, K W, and Stouffer, R J (1999) Model Assessment of Regional Surface Temperature Trends (1949–1997), *J. Geophys. Res.*, 104(D24), 30 981–30 996.

Lorenz, E M (1963) Deterministic Non-periodic Flow, *J. Atmos. Sci.*, 20, 130–141.

Mahlman, J D (1997) Uncertainties in Projections of Human-caused Climate Warming, *Science*, 278, 1416–1417.

Mahlman, J D (1998) Science and Non-science Concerning Human-caused Climate Warming, *Annu. Rev. Energy Environ.*, 23, 83–105.

Manabe, S and Wetherald, R T (1967) Thermal Equilibrium of the Atmosphere with a Given Distribution of Relative Humidity, *J. Atmos. Sci.*, 24, 241–259.

Manabe, S and Stouffer, R (1994) Multiple-century Response of a Coupled Ocean–Atmosphere Model and Increase of Atmospheric Carbon Dioxide, *J. Clim.*, 7, 5–23.

Mitchell, J F B, Johns, T C, Gregory, J M, and Tett, S F B (1995) Climate Response to Increasing Levels of Greenhouse Gases and Sulphate Aerosols, *Nature*, 376, 501–504.

Nakicenovic, N, Alcamo, J, Davis, G, de Vries, D, Fenhann, J, Gaffin, S, Gregory, K, Gruber, A, Jung, T Y, Kram, T, La Rovere, E L, Michaelis Mori, S, Morita, T, Pepper, W, Pitcher, H, Price, L, Raihi, K, Roehrl, A, Rogner, H-H, Sankovski, A, Schlesinger, M, Shula, P, Smith, S, Swart, R, van Rooijen, S, Victor, N, and Dadi, Z (2000) *IPCC Special Report on Emission Scenarios*, Cambridge University Press, Cambridge, 1–599.

Nicholls, N, Gruza, G V, Jousel, J, Karl, T R, Ogallo, L A, and Parker, D E (1996) Observed Climate Variability and Change, in *Climate Change 1995: The Science of Climate Change*, Cambridge University Press, Cambridge, 1–572.

Oort, A H and Liu, H (1993) Upper Air Temperature Trends Over the Globe, *J. Clim.*, 6, 292–307.

Phillips, N A (1956) The General Circulation of the Atmosphere: A Numerical Experiment, *Q. J. R. Meteorol. Soc.*, 82, 123–164.

Santer, B D, Taylor, K E, Wigley, T M L, Jones, P D, Karoly, D J, Mitchell, J F B, Oort, A H, Penner, J E, Ramaswamy, V, Schwarzkopf, M D, Stouffer, R J, and Tett, S (1996) A Search For Human Influences on the Thermal Structure of the Atmosphere, *Nature*, 382, 39–45.

Smagorinsky, J (1963) General Circulation Experiments with the Primitive Equations: 1. The Basic Experiment, *Mon. Weather Rev.*, 41, 99–164.

Smagorinsky, J, Manabe, S, and Holloway, Jr, J L (1965) Numerical Results From a Nine-level General Circulation Model of the Atmosphere, *Mon. Weather Rev.*, 43, 727–768.

Sun, D Z and Held, I M (1996) A Comparison of Observed Relationships Between Interannual Variations of Water Vapor and Temperature, *J. Clim.*, 9, 665–675.

Tett, S F, Stone, P A, Allen, M R, Ingram, W J, and Mitchell, J F B (1999) Causes of 20th Century Temperature Change Near the Earth s Surface, *Nature*, 399, 569–572.

Depletion of Stratospheric Ozone

DONALD J WUEBBLES

University of Illinois at Urbana-Champaign, Urbana, IL, USA

Human activities have had a devastating effect on the concentration and distribution of stratospheric ozone over the last few decades. The distribution of ozone that is naturally formed in the stratosphere is often referred to as the ozone layer. This layer protects life on Earth from harmful levels of solar ultraviolet (UV) radiation. Therefore, a decrease in the amount of ozone would allow an increase in the amount of the UV radiation from the Sun to reach the Earth's surface. Atmospheric measurements have shown that the depleted levels of ozone have indeed increased the amount of UV radiation at the surface. Corresponding to this increase in UV radiation are projected signi cant impacts on ecosystems and human health, including increases in incidences of skin cancer, eye cataracts, damage to genetic DNA, and suppression of the ef ciency of the immune system. It is the concerns about increased biologically harmful levels of UV from the decreasing levels of ozone that have largely been the driver for policy actions to protect the ozone layer.

The amounts of ozone in the stratosphere vary naturally throughout the year as a result of production and destruction processes, and as a result of winds and other transport processes that move the ozone molecules around the planet. In addition, changes in ozone occur associated with changes in the solar radiation reaching the Earth during the 11-year solar cycle and with various events such as large explosive volcanic eruptions. However, natural causes cannot explain the large observed decrease in ozone over recent decades. The signi cant global decrease in stratospheric ozone since the 1970s is well correlated with increasing amounts of chlorine (Cl) and bromine (Br) in the stratosphere. The sources of this Cl and Br are chloro uorocarbons (CFCs) and other halocarbons produced industrially for a variety of uses such as refrigerants in refrigerators, air conditioners and large chillers, as propellants for aerosol cans, as blowing agents for making plastic foams, and as solvents for dry-cleaning and for degreasing of materials. Atmospheric measurements have clearly corroborated theoretical studies showing that the Cl and Br released from the destruction of these halocarbons in the stratosphere are reacting to destroy ozone.

Ozone is also a greenhouse gas, having a large infrared absorption band in the atmospheric window at 9.6 μm. It is the balance between the solar and infrared radiative processes that determines the net effect of ozone on climate. Decreases in ozone in the stratosphere above about 30 km (roughly 18 miles above the Earth's surface) tend to increase the surface temperature as a result of the increased absorption of solar radiation, effectively increasing the solar energy that warms the Earth's surface. Below about 30 km, decreases in ozone tend to cool the surface temperature, because the infrared greenhouse effect dominates in this region. Scienti c analyses have shown that the decrease in ozone over recent decades has had a cooling effect, counteracting a fraction of the warming effect over this time from increasing concentrations of carbon dioxide and other greenhouse gases.

Over coming decades, the stratospheric ozone layer should gradually recover as a result of international actions taken to control the production and use of CFCs and other halocarbons under the Montreal Protocol. While ozone concentrations slowly recover, scientists and policymakers will need to work together to ensure that new problems do not develop as a result of the introduction of new chemicals into the atmosphere. Attention will also need to be devoted to interactions with industry and governments to ensure that the potential introduction of commercial supersonic aircraft, which will have cruise altitudes in the stratosphere, does not have a signi cant impact on ozone. Deeper understanding is also needed about the potential effects on the ozone layer of various natural events, such as volcanic eruptions and solar events, and possible human activities, including nuclear explosions.

FORMATION AND DESTRUCTION OF OZONE

The stratospheric ozone layer is an important natural feature of the atmosphere. For example, the formation of the ozone layer is generally believed to have played an important role in the development of life on land. The accumulation of oxygen molecules in the atmosphere as a result of the emergence of life in the ocean allowed for the production of ozone. Gradually the increasing levels of ozone led to the formation of the stratosphere, a region of the upper atmosphere where temperature increases with altitude largely as a result of the absorption of solar radiation by ozone. The resulting screening of lethal levels of solar UV radiation by the ozone layer is thought to have been important to allowing life to migrate from the oceans onto land.

Ozone (O_3) is composed of three oxygen atoms and is a gas at atmospheric pressures and temperatures. Although only about 15% of the atmospheric mass is in the stratosphere, approximately 90% of the atmospheric ozone is in the stratosphere. Most of the remaining ozone is in the troposphere, the lower region of the atmosphere extending from the Earth s surface up to roughly 10 km at mid-latitudes and 16 km in the tropics. Figure 1 shows a typical vertical ozone distribution for mid-latitudes. At these latitudes, the peak concentration of ozone occurs at altitudes between 20 and 30 km. At high latitudes, the peak occurs at lower altitudes, largely as the result of atmospheric transport processes and the lower height of the tropopause (the transition region between the troposphere and stratosphere).

How is Stratospheric Ozone Produced?

Without human intervention, the stratospheric ozone layer would be produced and destroyed through natural processes. In 1930, Sydney Chapman of Oxford University developed a photochemical theory for the formation and destruction of ozone based on a scheme involving only oxygen. Various experiments since then have verified that the production of ozone occurs as a result of the chemical reactions in his basic scheme.

Production of ozone in the stratosphere results primarily from photodissociation of oxygen (O_2) molecules (*see* Photochemical Reactions, Volume 1). The breaking of the molecular bond by high-energy solar photons at wavelengths less than 242 nm results in oxygen atoms that generally react rapidly with an oxygen molecule to form ozone. The sequence of reactions to form ozone is represented as:

$$O_2 + h\nu \ (\lambda < 242\,\text{nm}) \longrightarrow O + O \quad (1)$$

$$O + O_2 + M \longrightarrow O_3 + M \quad (2)$$

where $h\nu$ represents a photon, λ is wavelength, and M is a third body, normally a nitrogen (N_2) or oxygen (O_2) molecule, the primary components of air (*see* Stratosphere, Chemistry, Volume 1).

Because the atmosphere is uniformly filled with oxygen molecules, most of the ozone is generated where there is a balance between the decreasing atmospheric density with altitude and where there is available UV solar radiation in the wavelength region less than 242 nm. This primarily occurs in the upper stratosphere.

The high-energy solar radiation needed to produce ozone is largely absorbed in the stratosphere, resulting in the production of the ozone layer. Too little UV radiation reaches the troposphere for it to be a major cause of ozone production in this region. In contrast, the smaller amount of ozone in the troposphere is largely formed through a series of chemical reactions, generally referred to as smog reactions. The transport of stratospheric ozone down into the troposphere is also important to the budget of tropospheric ozone.

Although the primary production of ozone occurs in the tropics and primarily above 30 km altitude, the largest concentrations of ozone occur at lower altitudes, as shown in Figure 1, and at higher latitudes. The combined importance of chemical and transport processes in determining the ozone distribution is demonstrated in Figure 2, which shows the vertically-integrated total column of ozone as a function of latitude and season, based on ground-based and satellite observations. The minimum in total ozone actually occurs in the tropics, while the maximum amounts in total ozone occur in late winter and early spring in both hemispheres. The ozone peak is closer to the pole in the Northern Hemisphere than the peak in the Southern Hemisphere, largely due to the differences in winds and other processes affecting the transport of ozone within the two hemispheres. Also seen in the Antarctic region is the large ozone destruction in their springtime, termed the *Antarctic ozone hole*, to be discussed later. Variations in total ozone with longitude are generally small in comparison with variations with latitude; however, recent decreases in ozone during the Northern Hemisphere winter and early spring associated with ozone hole type processes have occurred at specific longitudes where the polar winter vortex was occurring in that hemisphere.

The total column amount of ozone in Figure 2 is reported in Dobson units, which were named after George Dobson of Oxford University, a pioneer in measuring atmospheric ozone. One Dobson unit equals one milliatmosphere–cm. The average amount of ozone around the globe is 300 Dobson units. Although ozone plays a vital role in the atmosphere, the molecules of ozone are exceedingly rare, accounting for fewer than 10 of every million molecules of air. If all of the ozone in the troposphere

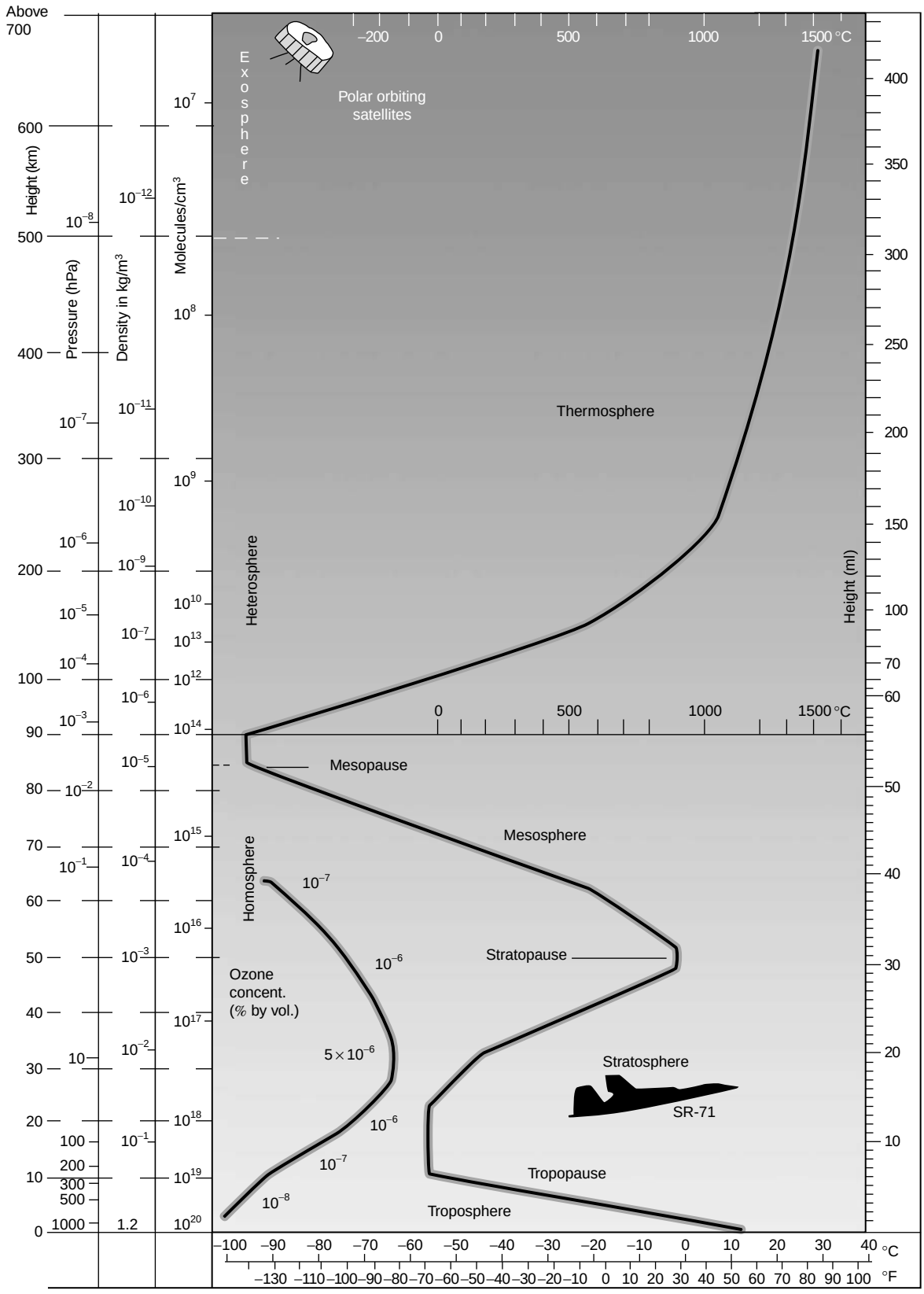


Figure 1 The typical variation in temperature and concentration of ozone with altitude in the troposphere and stratosphere over mid-latitudes in the Northern Hemisphere

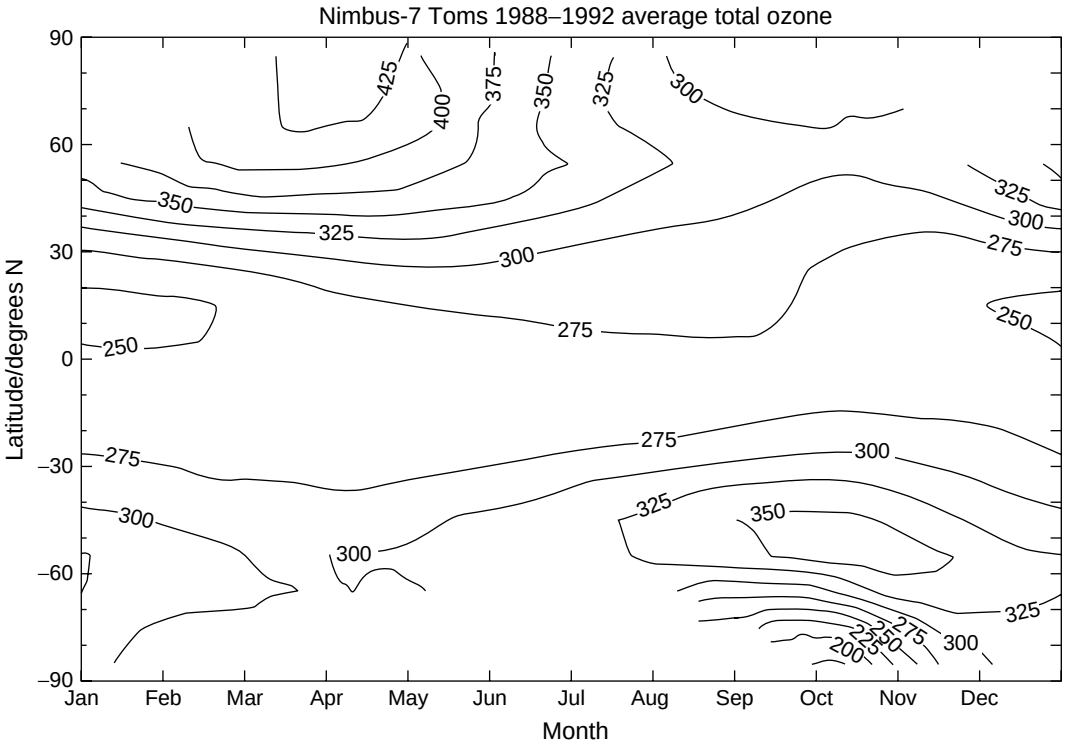


Figure 2 Integrated column amount of ozone, referred to as the total ozone column, in Dobson units, as a function of latitude and season, based on observations taken by the total ozone mapping spectrometer (TOMS) satellite (McPeters *et al.*, 1996)

and stratosphere were brought to sea level and spread evenly around the globe, it would fit in a layer only about 3 mm thick.

How is Ozone Destroyed?

Without natural ozone destruction, most of the oxygen in the stratosphere, and perhaps throughout the atmosphere, would eventually be converted to ozone. Such concentrations of ozone would be intolerable to many forms of life on Earth, both because of the direct toxicity of ozone and because no UV radiation would reach the Earth s surface.

Although Chapman s oxygen-based chemical scheme did not capture the reactions actually responsible for the primary destruction of ozone, he did recognize that ozone and atomic oxygen rapidly interchange with each other, while the sum of the two is linked to much slower chemical processes. This led to the conceptual picture of thinking of ozone destruction in terms of destruction of the odd-oxygen.

Ozone photodissociates at UV and visible wavelengths to produce O and O_2. However, because the oxygen atom will generally react to reform ozone, this mechanism produces no net change in the amount of odd-oxygen.

The destruction of ozone and odd-oxygen in the stratosphere occurs mainly through catalytic reactions with other gases. In the stratosphere, important catalysts for ozone destruction include nitric oxide (NO), hydroxyl (OH), chlorine (Cl) and bromine (Br). Such gases can be directly substituted for X in the following catalytic mechanism:

$$X + O_3 \longrightarrow XO + O_2 \quad (3)$$

$$XO + O \longrightarrow X + O_2 \quad (4)$$

These reactions result in the net reaction of $O + O_3$ being converted to two oxygen molecules. Because gas X is recycled through these reactions, it continues to destroy ozone until some other reaction converts X to a less reactive form, such as nitric acid or hydrogen chloride. In this way, a single Cl atom can lead to the net destruction of thousands of ozone molecules before it becomes tied up. Because of such cyclic reaction mechanisms, relatively small concentrations of reactive Cl and other catalytic species in the stratosphere can have a significant impact on the amount and distribution of stratospheric ozone (*see* Stratosphere, Chemistry, Volume 1).

In addition to the mechanism described by Equations (3) and (4), there are a number of other catalytic mechanisms also affecting ozone. These must all be considered in evaluating the effects of human-related emissions on ozone. In addition, some gases, like iodine (I), that are not present in significant amounts in the current atmosphere, are now known to react catalytically with ozone,

and require careful analysis to prevent future ozone destruction if chemicals containing these atoms are considered for use.

The total amount of ozone in the stratosphere would remain fairly constant (relative to well-recognized seasonal variations) were there no change in the destruction rate and the transport of ozone out of the stratosphere. However, increasing inputs of Cl and Br into the stratosphere over the past few decades have changed the balance. As an analogy, the production and destruction of ozone is akin to the faucet and drain in a tub of water. As long as the destruction is constant (i.e., the size of the drain remains the same), the ozone amount (i.e., the level of the water in the tub) will remain relatively constant. However, the increasing amount of Cl and Br act, in effect, to increase the size of the drain in the tub. As a result, the water level, or the amount of ozone, will decrease.

ATMOSPHERIC GASES AFFECTING OZONE

Without human intervention, concentrations of gases important to concerns about stratospheric ozone would be expected to change slowly, as indicated by ice core measurements of gases like methane (CH_4) for concentrations prior to the 20th century. These same ice cores show that there were very low amounts, if any at all, of the Cl-containing CFCs and Br-containing halons in the atmosphere before the 20th century. Since the beginning of the industrial age, however, emissions associated with human activities have risen rapidly. Agriculture, industry, waste disposal, deforestation, and fossil fuel use have been producing increasing amounts of carbon dioxide (CO_2), CH_4, nitrous oxide (N_2O), CFCs and other important gases. Due to increasing emissions, atmospheric levels of these gases have been increasing at an unprecedented rate, raising concerns regarding the impacts of these gases on ozone and climate. While all of these gases have had some impact on stratospheric ozone, some like the CFCs and other halocarbons have been particularly effective at depleting ozone.

The Growth in Halocarbon Concentrations

Several halocarbons have been largely responsible for stratospheric ozone loss over recent decades. Halocarbons containing Cl and/or Br are of particular concern with regard to destruction of stratospheric ozone (*see* Halocarbons, Volume 1). Amongst the most potent halocarbons in the current atmosphere are the CFCs CFC-11 ($CFCl_3$) and CFC-12 (CF_2Cl_2) (*see* Chloro uorocarbons (CFCs), Volume 1). Because of their additional importance as greenhouse gases that can affect climate, a number of halocarbons are of particular concern. With the exception of the naturally occurring fractions of methyl chloride (CH_3Cl) and methyl bromide (CH_3Br), all of the halocarbons in the atmosphere are thought to be a result of human activities. Their inertness and long lifetimes in the atmosphere have made them attractive chemicals for use as propellants, refrigerants, fire retardants, and other industrial applications.

All of the CFCs, halons and other halocarbons that affect stratospheric concentrations of Cl and Br have the common feature of not being soluble to any significant degree in water. On the other hand, sea salt and other Cl-containing emissions at the Earth s surface, such as the emissions of Cl gas from swimming pools or HCl from a variety of sources, are generally water-soluble; this Cl tends to be rapidly removed from the atmosphere by precipitation. Atmospheric measurements have confirmed that concentrations of the water-soluble chlorides decrease rapidly with altitude in the troposphere. Volcanic eruptions sometimes contain Cl, generally as HCl, that could be directly injected into the stratosphere; however, it appears that this Cl is removed before it reaches the stratosphere by processes associated with such eruptions. Sulfur emissions in explosive volcanic eruptions can and do affect ozone through heterogeneous reactions that occur on the resulting sulfuric aerosols.

All of the fully halogenated CFCs have long atmospheric lifetimes. The lifetime of CFC-11 is about 50 years, while the lifetime of CFC-12 is about 102 years. The long atmospheric lifetimes of the CFCs help make these compounds of particular concern to ozone. As first pointed out by F Sherwood Rowland and Mario Molina of the University of California at Irvine (*see* Molina, Mario J, Volume 1; Rowland, F Sherwood, Volume 1), the long lifetimes of CFCs mean that these molecules will not be destroyed until they reach the stratosphere, generally by photodissociation, thus releasing their Cl right where it can most readily affect the ozone concentration.

Measurements of CFCs and other compounds by the US National Oceanic and Atmospheric Administration s (NOAA) Climate Monitoring Diagnostic Laboratory at sites throughout the world are shown in Figure 3. CFC-11 and CFC-12 have the highest atmospheric concentrations, at 0.26 and 0.53 parts per billion by volume (ppbv), i.e., the number of molecules per one billion molecules of air, respectively. The tropospheric concentrations of both of these gases were increasing at about 4% year^{-1} in the early 1990s, but the rates have now slowed appreciably. The concentrations of CFC-11 and several other halocarbons controlled under the Montreal Protocol have already started to decline. The use of these compounds has diminished greatly, with all but essential applications being banned by the beginning of 1996. The atmospheric concentrations of several other halocarbons have, until recently, been growing at even faster rates than CFC-11 and CFC-12. For example, the concentration of CFC-113 ($C_2F_3Cl_3$) was increasing about 10% year^{-1} in the early 1990s, but its rate of rise

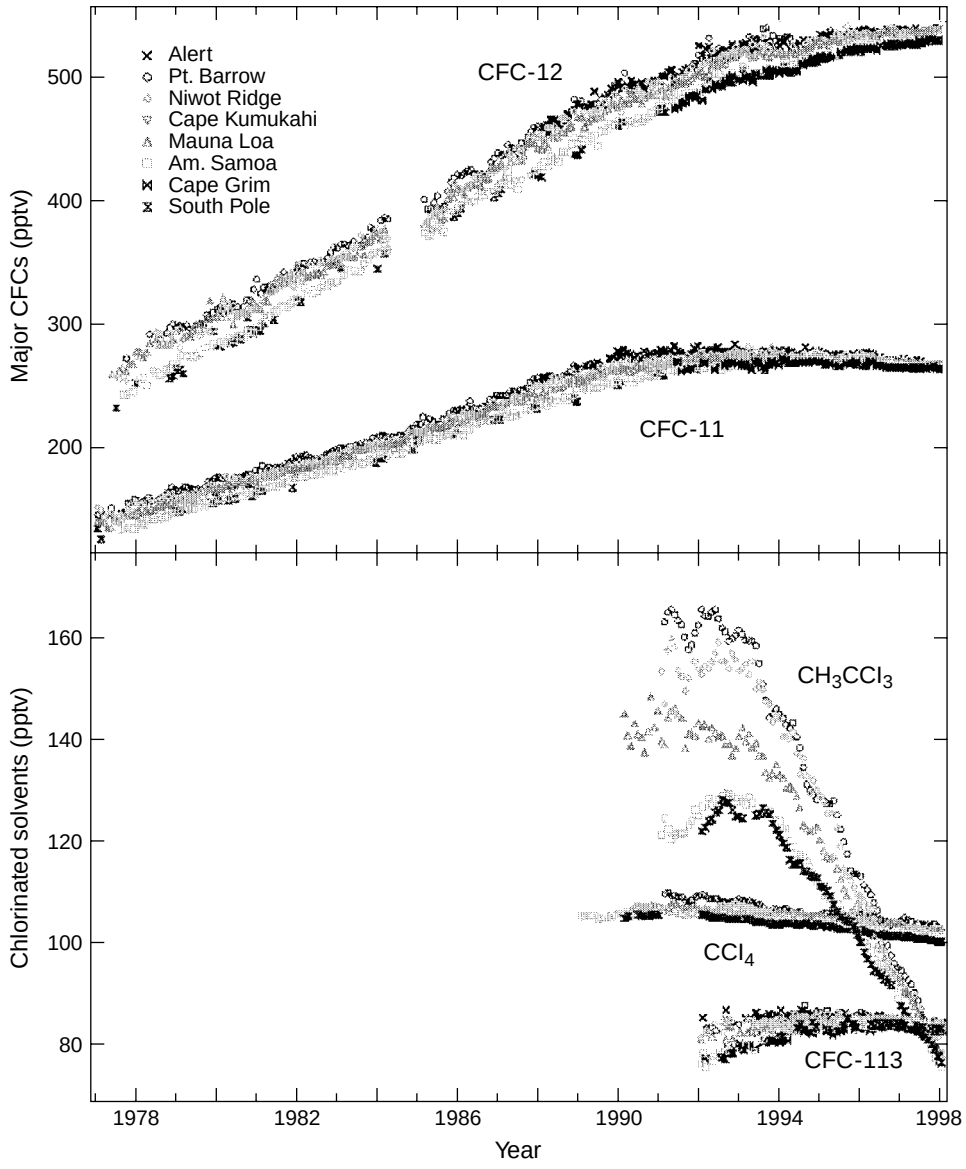


Figure 3 Changes over time of measured CFC concentrations (in parts per trillion by volume, pptv), showing a sharp rise over last few decades, and then the beginning of a decline in recent years. Measurements were taken by the NOAA Climate Monitoring and Diagnostics Laboratory (J Elkins, private communication, 2000)

has also slowed greatly; its current concentration is about 0.08 ppbv.

Abundances of the compound methyl chloroform (i.e., 1,1,1-trichloroethane (CH_3CCl_3)) have declined appreciably. Although its five year lifetime is much shorter than the lifetimes of CFCs, methyl chloroform has been controlled under the Montreal Protocol because the large amounts of this compound being used as a solvent and cleaning agent still resulted in a significant contribution to the stratospheric Cl concentration.

The atmospheric lifetimes of the hydrochlorofluorocarbons (HCFCs), hydrofluorocarbons (HFCs), and other halocarbons containing hydrogen tend to be much shorter than the CFCs. Several of these compounds are being used as replacements for CFCs and halons in a variety of uses. A few were already in use before the controls on CFCs were put into place. Due to their shorter lifetimes, smaller amounts of these replacement compounds are reaching the stratosphere and, as a result, their Cl and Br have less effect on ozone than the CFCs. HCFC-22 (CHF_2Cl), a refrigerant often found in home air conditioners, has gained increased use as a replacement for CFCs and its concentration has been increasing at about 5–6% $year^{-1}$ since 1995. The measured concentrations of several other replacement compounds are also increasing.

Br is more effective at destroying ozone than Cl. Studies suggest that its overall effectiveness in the stratosphere is roughly 50 times that of Cl. Therefore, rapid increases in atmospheric concentrations of Br-containing halons, most notably Halon-1301 (CF_3Br) and Halon-1211 (CF_2ClBr), have caused concern. Despite their control under the Montreal Protocol, the concentrations of several halons, including H-1211 and H-1301, continue to increase. Primary destruction of these compounds occurs through photolysis, resulting in long atmospheric lifetimes (65 years for Ha-1301; 20 years for Ha-1211). These compounds, however, currently have small atmospheric concentrations, about 4 pptv (parts per trillion by volume) or less. As a result, the impact of Br on the current atmosphere is smaller than that of Cl.

Largely as a result of the dissociation of CFCs, halons, and human-related emissions of other halocarbons, there are now over 3.5 ppbv of reactive Cl and over 20 pptv of reactive Br in the stratosphere as compared to less than 1 ppbv of reactive Cl and less than 7 pptv of reactive Br estimated as being in the pre-industrial stratosphere. The increase in stratospheric Cl during the 1980s and 1990s has been measured directly by instruments on satellites, balloons and aircraft, and has been shown to directly correlate to what would have been expected from the corresponding increase in concentrations of CFCs and other halocarbons at the ground.

HUMAN-RELATED EFFECTS ON OZONE

A Historical Perspective

Before examining the changes in ozone concentrations over recent decades, it is useful to consider how the understanding of the atmosphere has changed over the last 50 years. Through the work of George Dobson and others, the annual variation in the distribution of ozone for selected latitudes was beginning to be understood by the early 1950s. This knowledge was largely gained through the use of the Dobson spectrophotometer, an instrument still in use, which measures the total ozone column based on the extent of UV radiation absorbed at selected wavelengths. The measurements of solar spectra clearly indicated that ozone was important to life on Earth because of its absorption of UV light. However, the understanding of the chemistry affecting stratospheric ozone in the 1950s had not evolved from the oxygen-only hypothesis of Sydney Chapman in 1930.

By 1950, measurements of several major gases, including CO_2, water (H_2O), CH_4, and N_2O, were being made. The roles of these gases in atmospheric chemistry and climate, however, were not very well understood. Also, the presence of CFCs and other halocarbons in the atmosphere remained unknown until the first measurements of their concentration were made by a medical doctor from the UK, James Lovelock, in the early 1970s (*see* Lovelock, James, Volume 1). Until the 1970s, CFCs were thought to be chemically inert in the atmosphere.

In the 1950s, it began to be recognized that hydrogen oxides likely play a role in the destruction of ozone. Then in 1970, Paul Crutzen (*see* Crutzen, Paul J, Volume 1) suggested that nitrogen oxides were important in the destruction of ozone. Through this and many later studies, scientists were discovering that the stratosphere, and the atmosphere in general, was strongly affected by atmospheric chemical processes. A few years after Crutzen s landmark paper, Richard Stolarski and Ralph Cicerone, then at the University of Michigan, suggested that the Cl catalytic cycle could be important to ozone if a sufficient source of Cl in the stratosphere could be found. In 1974, Rowland and Molina determined that CFCs could provide the source of Cl. Soon after, it was determined that Br could also catalytically destroy ozone. In 1986, Susan Solomon, Rolando Garcia, F Sherwood Rowland and Donald Wuebbles suggested that heterogeneous reactions of Cl molecules on particles could be important to the chemistry affecting ozone. Since then it has been discovered that heterogenous reactions have had a significant effect on the observed trends in stratospheric ozone over recent decades.

In 1967, the first space-based spectral measurements of ozone were made by the Orbiting Geophysical Observatory (IV). Since the initial success of this satellite, numerous satellites have been launched for the purpose of observing the distribution of ozone around the globe and its changing concentration over time. Some of these satellite instruments, like the TOMS, measure total column ozone, while others, like the Solar Backscatter UV instrument and several instruments on the upper atmosphere research satellite (UARS), measure the vertical distribution of ozone. The UARS satellite also has been providing important measurements of the solar flux and many of the gases and particles affecting stratospheric ozone since it was launched in 1991.

To understand the underlying importance of the chemical and physical processes affecting ozone, the measurements from ground-based, aircraft, balloon and satellite instruments are coupled with and compared to analyses from computer-based models of the relevant atmospheric processes. These models attempt to derive the changing distributions of the gases and aerosols affecting ozone. Such models generally represent over 50 chemical species and more than 150 chemical reactions that have affected or can affect ozone. These models allow scientists to better understand the complex interactions that control past, present, and future concentrations of stratospheric ozone.

In the late 1970s and early 1980s, scientists projected that stratospheric ozone would decline gradually as a result of the increasing atmospheric burden of CFCs and other halocarbons. The discovery of the Antarctic ozone hole in 1985 demonstrated that the decrease in ozone was actually

occurring at a much faster rate than the earlier studies had projected.

Stratospheric Ozone Trends

As mentioned earlier, the connection between potential environmental effects on ozone and manmade CFCs was first pointed out by Molina and Rowland in 1974, when they suggested that Cl from these compounds could destroy stratospheric ozone. Research findings since then have continued to support the projections of the effects of these compounds on the global distribution of ozone. In addition, it is recognized that other gases containing Cl and Br, which is even more reactive with O_3 than Cl, are also affecting ozone. To a lesser extent, increasing concentrations of other gases like CO_2, CH_4, and N_2O, are also involved in chemical reactions that are affecting stratospheric ozone.

Measurements of ozone by satellite and ground-based measurements over the last several decades indicate that stratospheric ozone levels have decreased (*see* Stratosphere, Ozone Trends, Volume 1). Atmospheric ozone has decreased globally by more than 5% since 1970. Figure 4 shows satellite and ground-based measurements of the change in total ozone column from 60 °S to 60 °N (the latitude extent is limited primarily by the amount of satellite data available at higher latitudes, particularly in the winter). Detailed satellite measurements began in late 1978, while a reasonably representative global network of ground-based stations has been in operation since the late 1960s. Seasonal variations and the effects of the variations in the solar flux due to a 11-year solar cycle, which produce about a 1.2–1.5% variation from solar minimum to solar maximum, have been removed from the data shown. The ground-based data in Figure 4 show a decrease in ozone of over 4% since 1978. Model calculations extending the estimated ozone amounts back to the late 1960s indicate there has likely been an additional 1% decrease in total column ozone. Because the largest changes in ozone have occurred at higher latitudes as a result of the ozone hole heterogeneous chemistry

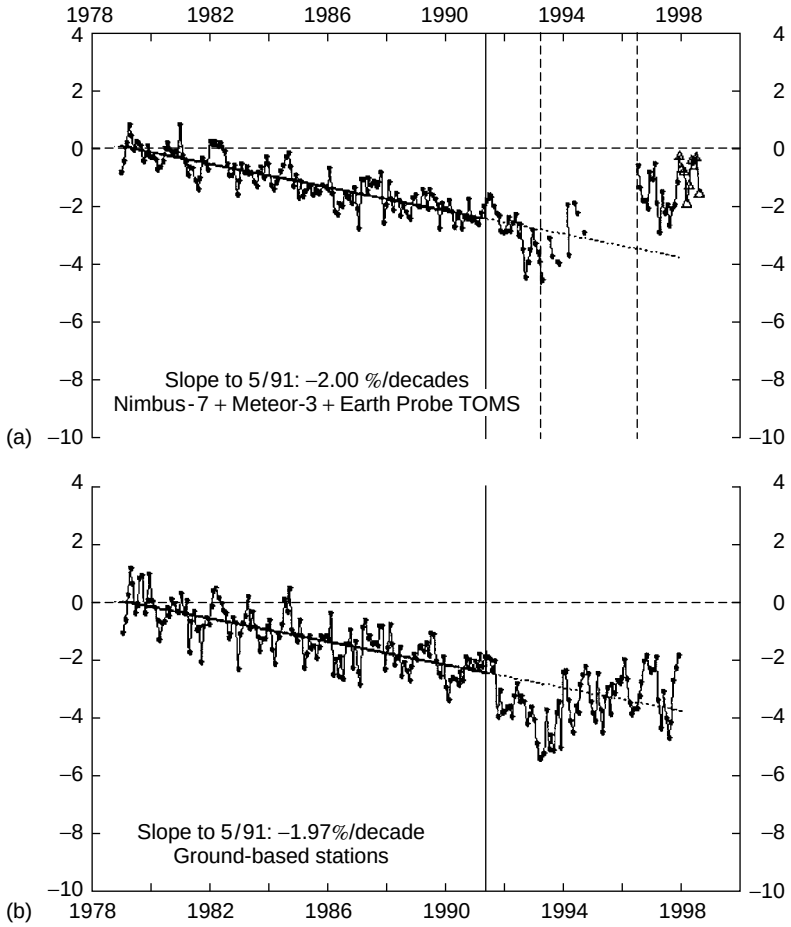


Figure 4 Deviations in total ozone with time relative to January 1979 from various ground-based (b) and satellite measurements (a). The data are area weighted over 60 °South to 60 °North. (Based on World Meteorological Organization, 1999)

(to be discussed later) the decrease in ozone in Figure 4 would be larger yet if the change in ozone at higher latitudes of both hemispheres were also included.

The trends in total column ozone in Figure 4 from the satellite data agree well with the trends determined from the ground-based measurements through the mid-1990s. However, the satellite data in the late-1990s show much smaller changes in ozone. This discrepancy appears to result from difficulties in calibrating the satellite data in the late-1990s after there was a change in the satellite being used to take the measurements. More recent analyses, using improved calibration of the irradiances measured by the satellite instrument, are now in better agreement with the ground-based measurements.

The decrease in ozone in Figure 4 appears to reach its maximum decrease in the early 1990s. However, the large decrease during the early 1990s was in part a response of ozone to the large amount of sulfate aerosols in the stratosphere resulting from the Mt Pinatubo volcanic eruption in June 1991 (*see* Volcanic Eruption, Mt. Pinatubo, Volume 1). These aerosols acted as heterogeneous reaction sites for reactions involving nitrogen, Cl and Br in the stratosphere, allowing for larger destruction of ozone than would be expected in the stratosphere with normal levels of sulfate aerosols. With the removal of the Pinatubo-induced sulfate aerosols from the stratosphere through transport processes, there has been a slight recovery of stratospheric ozone since the mid-1990s. However, the total column ozone from 60 °S to 60 °N is still less than it was pre-Pinatubo. Data for 1999 and early 2000 not shown in Figure 4 continue to show further decline in total column ozone, suggesting that the most significant decrease in ozone may not yet have been reached.

Figure 5 shows the seasonal variations of the trends in total ozone since 1979 (to January 1998) as a function of latitude. The changes in ozone are shown as a percentage per decade. The data in Figure 5 show generally good agreement between ground-based measurements and satellite measurements from the TOMS instrument, although some differences exist in the Southern Hemisphere, particularly at latitudes where there are few ground-based measurement sites. Significant decreases in total ozone are found in both the Northern and the Southern Hemisphere at the middle and high latitudes, with no significant change in the equatorial region.

Surface measurements indicate that the total ozone column at mid-latitudes in the Northern Hemisphere has decreased by 1.3% decade^{-1} in the summer months and 2.7% decade^{-1} in the winter months from 1969 to 1999. Much larger decreases in total ozone are found at latitudes greater than 60°, particularly in the Southern Hemisphere. The largest changes in total column ozone at high latitudes are found in the springtime of each hemisphere.

Satellite and ozonesonde (i.e., made by balloon) data sets indicate that ozone is decreasing particularly rapidly in the lower stratosphere, accounting for a major fraction of the trend in total ozone, although there is also significant ozone destruction occurring in the upper stratosphere. Figure 6 shows the annually averaged change in ozone with latitude and altitude for 1979 through 1996 from the Stratospheric Aerosol and Gas Experiment (SAGE) satellite observations. Available ground-based and balloon-based measurements show similar trends. Large percentage decreases in ozone, approaching 1% year^{-1}, are found in the upper stratosphere. The data at these altitudes are well described as a linear decline modulated by the 11-year solar cycle variation that has been subtracted from the analysis shown in Figure 6. The maximum annual decline of $-0.8 \pm 0.2\%$ year^{-1} (or about 8% decade^{-1}) occurs at 40 km altitude, with strong symmetry between the Northern and Southern Hemispheres. Measurements of relevant atmospheric gases as well as studies modeling the chemistry and physics of the stratosphere indicate that the long-term changes in ozone in the upper stratosphere are well explained by the effects of the gas-phase Cl catalytic cycles discussed earlier.

However, there is much less ozone in the upper stratosphere than in the lower stratosphere, where smaller percentage changes actually imply larger numbers of molecules destroyed. Maximum decreases in lower stratospheric ozone occur at about 15 km altitude, reaching $7.3 \pm 4.6\%$ decade^{-1} at mid-latitudes. However, there is strong variability with location. For the balloon-borne ozonesonde instrument data, 10 stations from 36 °N to 59 °N show maximum annual trends in ozone of -2 to -12% per decade for 15 to 20 km altitude. The largest decreases in ozone tend to be found in the winter and early springtime.

For mid-latitudes, between the two areas of peak change in the lower and upper stratosphere, the minimum trend is $-2 \pm 1.8\%$ decade^{-1} at 30 km, which is just statistically significant at the 95% confidence level. Thus, at northern mid-latitudes, a statistically significant negative trend in ozone exists at all altitudes between 12 and 45 km.

The Antarctic Ozone Hole

Beginning in the late 1970s, a special phenomenon began to occur in the springtime over Antarctica, referred to as the Antarctic ozone hole. A large decrease in total ozone, now over 60% relative to pre-hole levels, has been observed in the springtime (September to November) over Antarctica. Joseph Farman and colleagues first documented this rapid springtime decrease in Antarctic ozone over their British Antarctic Survey station at Halley Bay, Antarctica. These analyses attracted the attention of the scientific community, which soon found that decreases in the total ozone column were greater than 50% compared with historical values observed by both ground-based and satellite techniques.

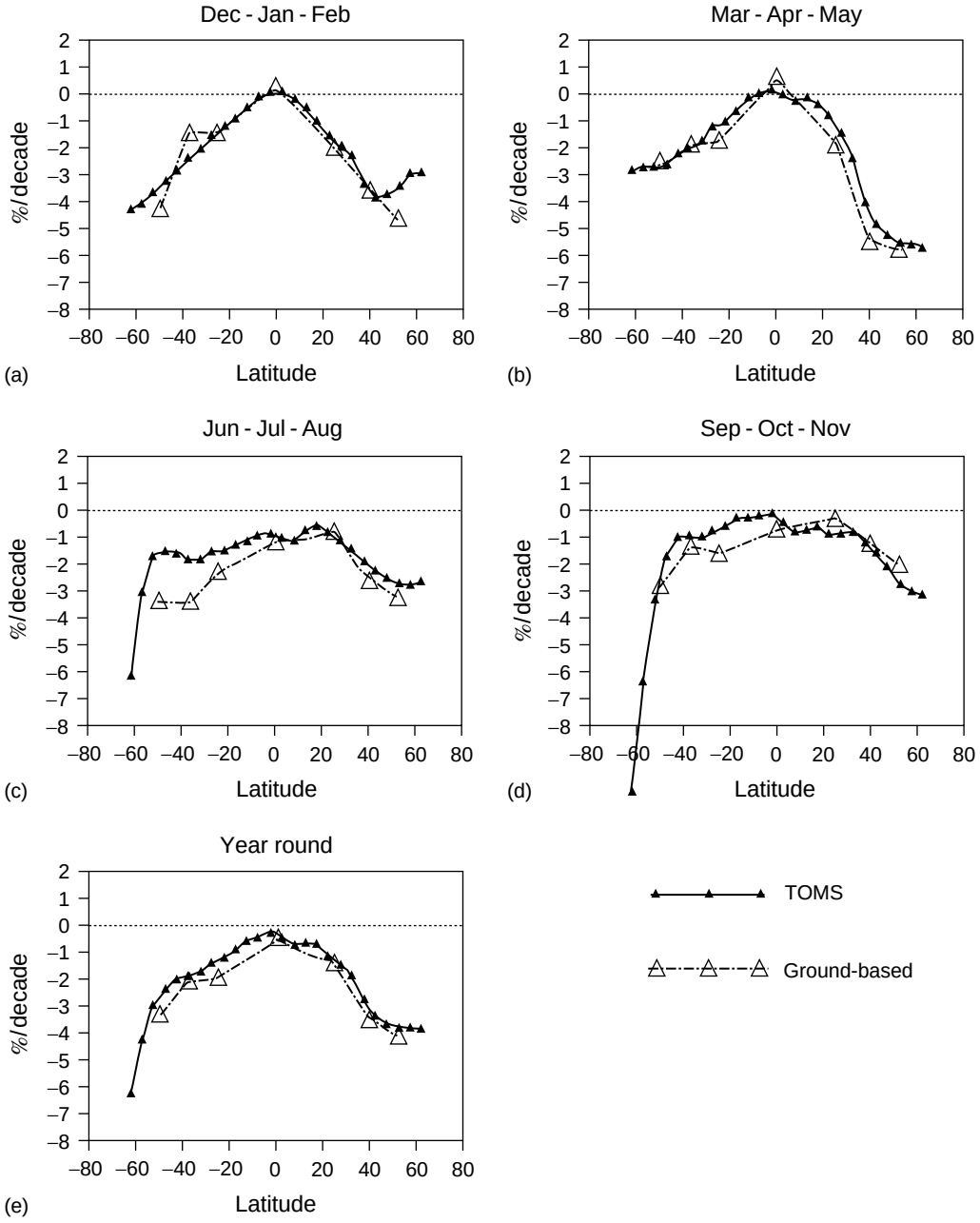


Figure 5 Change in total ozone with latitude (negative is Southern Hemisphere, positive is Northern Hemisphere) and season from January 1979 to January 1998 for satellite TOMS measurements and to May 1997 for ground-based Dobson measurements. (Based on the World Meteorological Organization, 1999)

Measurements made in 1987 indicated that more than 95% of the ozone over Antarctica at altitudes from 13–22 km had disappeared during September and October. More recent measurements have found 60–90% decreases in monthly average ozone amounts between 12 and 20 km for the September to November time period, relative to pre-ozone-hole levels, with essentially all the ozone destroyed at times. Since 1992, the Antarctic ozone holes have been the biggest (in areal extent) and the deepest (in terms of minimum amounts of ozone overhead), with ozone being locally depleted by more than 99% between about 14–19 km in October. There is no indication, as of mid-2000, that the magnitude of the Antarctic ozone hole has peaked or begun to decline.

At the time of the original 1985 paper by Farman and colleagues, there was no expectation that such a phenomenon was occurring. The traditional gas-phase approach to analyzing the expected ozone destruction from CFCs and other

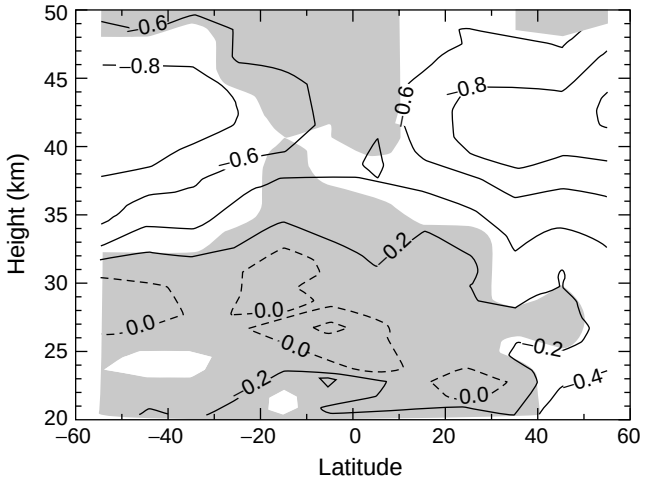


Figure 6 Annually-averaged change in O_3 with latitude and altitude from SAGE satellite observations (1979–1996). The shaded areas indicate where the trends do not differ from zero within the 95% confidence limits. (Based on World Meteorological Organization, 1999)

halocarbon emissions of Cl and Br projected ozone destruction in the upper stratosphere, but suggested little response of ozone levels in the lower stratosphere. As a result of the Farman paper, a number of hypotheses arose, each attempting to explain the large ozone destruction in the springtime over Antarctica. The Farman paper had proposed that the Cl catalytic cycle might explain the observed ozone decrease, but this did not match the expected ozone decrease possible from the reactive Cl available at the high latitudes. Other hypotheses suggested that a change in atmospheric downward transport processes in the polar region or a cyclic variation in natural production of nitrogen oxides might explain the ozone decrease. However, later measurements over Antarctica demonstrated that the concepts proposed in the 1986 paper by Susan Solomon of the NOAA in Boulder, CO, and co-authors, Rolando Garcia of the National Center for Atmospheric Research, also in Boulder, F Sherwood Rowland of the University of California, Irvine, and Donald Wuebbles of Lawrence Livermore National Laboratory, provided the basic explanation for the observed ozone decrease. An expedition organized by several US agencies to Antarctica in 1987, led by Susan Solomon, was able to take sufficient measurements to substantiate the basic premise and demonstrate the clear connection of Cl chemistry in explaining the observed ozone destruction.

According to this explanation, the air over the Antarctic becomes extremely cold during the winter as a result of the lack of sunlight over the polar region and because of greatly reduced mixing of the lower stratospheric air in the south polar region with warmer air outside this region. During the winter, a circumpolar vortex, also called the polar winter vortex, forms, which isolates the air in the polar region from that outside of the region as a result of a stratospheric jet of wind circulating around the world between approximately 50 °S and 65 °S. The extremely cold temperatures inside the vortex lead to the formation of clouds in the lower stratosphere (from roughly 12–22 km), called polar stratospheric clouds. The Solomon *et al.*, paper proposed that heterogeneous reactions occur on these particles that convert less reactive forms of Cl to much more reactive ones. When daylight starts to occur over Antarctica in the early spring, this Cl is available to react with and destroy ozone. Br compounds and nitrogen oxides can also react heterogeneously on the particles of these clouds. The ozone destruction continues until the polar vortex breaks up, usually in November. While further analyses have substantiated the basic concepts of the Solomon hypothesis, laboratory and atmospheric measurements have further refined understanding of the nature of the particles in the polar stratospheric clouds and their reactive properties and of the Cl and Br chemistry that leads to the ozone destruction.

Arctic Ozone Hole?

In the late 1980s, it was generally thought that the Arctic lower stratosphere did not get cold enough to lead to decreases in ozone during the winter and spring like those found in the Antarctic. This was because the polar vortex is not generally as strong in the Northern Hemisphere, and, although polar stratospheric clouds would form, they would be unlikely to last long enough for extensive decreases in ozone to occur. However, as of mid-2000, extremely low ozone concentrations have been found in the Arctic polar vortex in the late winter/spring during eight of the last 11 years. Ozone losses average about 100 Dobson units, with extremes exceeding about 200 Dobson units below the 1964–1976 average total ozone column amounts. Lower-stratospheric temperatures in the Arctic vortex have been sufficiently low during those years to lead to ozone destruction processes similar to those found in the Antarctic ozone hole. As with Antarctica, large increases in concentrations in Cl and O have been measured in the regions where the large ozone destruction is occurring.

Explaining the Observed Ozone Trends

Atmospheric measurements of relevant stratospheric gases that react with ozone have confirmed the role being played by increasing concentrations of Cl and Br. Further, computer-based models of atmospheric chemical and physical processes are able to largely explain the globally observed trends in ozone over the last few decades. Although the increase in Cl and Br largely explain the observed decrease in ozone, the trends could not be fully explained without also accounting for the natural effects of

volcanic eruptions and solar flux variations and the changes in climate affecting stratospheric chemistry and transport processes.

The modeling studies have clearly confirmed the original hypothesis put forth in 1974 that emissions of CFCs would lead to a significant reduction of upper stratospheric ozone, with the maximum decline due to gas phase Cl chemistry expected near 40 km latitude. There is good agreement between the observed and modeled long-term trends in ozone at this altitude. The models reveal a long-term decline in ozone throughout the upper stratosphere that is driven by the accumulation of human-related Cl. In addition, modeled and measured increases in upper stratospheric HCl and hydrogen fluoride (HF) are in excellent agreement with the increases expected as a result of the increases in atmospheric concentrations of CFCs and other halocarbons.

Modeling studies can also largely explain the observed changes in ozone in the lower stratosphere, demonstrating the effects of heterogeneous chemistry involving reactions on sulfate particles outside the polar winter vortices, and the importance of heterogeneous reactions on polar stratospheric cloud particles within the polar winter vortex in each hemisphere. However, some uncertainties remain in fully explaining the observed trends in lower stratospheric ozone.

CONSEQUENCES OF OZONE DEPLETION

UV and Ozone

While the ozone in the troposphere and stratosphere is chemically identical, the ozone in each region has very different effects on life at the Earth s surface. Stratospheric ozone plays a beneficial role by absorbing solar UV radiation, preventing biologically harmful levels of UV radiation from reaching the Earth s surface (*see* Ultraviolet Radiation, Volume 1). Concern about increased UV radiation from ozone loss has been the driver for policy actions to protect the ozone layer. On the other hand, direct contact with ozone is toxic to many life forms on Earth. Surface air pollution, which is often associated with high levels of ozone, and transport of stratospheric air to the ground, with associated stratospheric levels of ozone, have been known to have negative impacts on human health, animals and plants, including some agricultural crops.

Ozone and molecular oxygen ensure that little UV at wavelengths less than 280 nm reaches the Earth s surface. Molecular oxygen absorbs effectively at wavelengths less than 240 nm while ozone absorbs well from 240 nm to wavelengths greater than 300 nm. With normal ozone levels, by the time solar radiation at 278 nm reaches the ground, its intensity has been reduced by a factor of about 10^{17} relative to its intensity at the top of the atmosphere. At 300 nm, the intensity at the ground is about 1% of its value at the top of the atmosphere.

While most life on the Earth s surface co-exists with the existence of UV radiation (in fact, some is needed for the production of vitamin D in humans), too much radiation can be damaging. However, not all of the radiation is equally damaging. Longer wavelengths of solar radiation in the UV, at wavelengths from 315–400 nm, called UV-A, can cause some biological damage. However, UV-A also stimulates repair mechanisms, so that effects in most people are minimal as long as changes in UV-A are not large. UV-A is only weakly absorbed by ozone. In contrast, solar radiation at shorter wavelengths, termed UV-B (280–315 nm) and UV-C (100–280 nm), can cause severe biological damage at sufficient exposure. The large concern about decreases in the amount of stratospheric ozone is because of the resulting increase in UV-B radiation.

Ground-based spectrophotometer measurements of total ozone actually measure the UV solar radiation reaching the ground at selected wavelengths. Therefore, the measured decrease in ozone at many of the stations around the world is a direct indication that there has also been a corresponding increase in the UV radiation reaching the ground. In addition, special instruments that measure either the integrated change in UV-B or the response with wavelength throughout the UV spectrum are being used at a number of locations around the globe. These instruments show that the inverse relationship between changes in ozone and UV-B radiation is well established by both theoretical analyses and observations. Figure 7 shows this inverse relationship in measurements from several sites. As an example, an instrument in Melbourne, Australia measured a sharp increase in UV-B at the same time as a 10% decrease in ozone was being measured. Longer-term increases in UV-B have also clearly occurred over the last decade at clean air sites in the Swiss Alps.

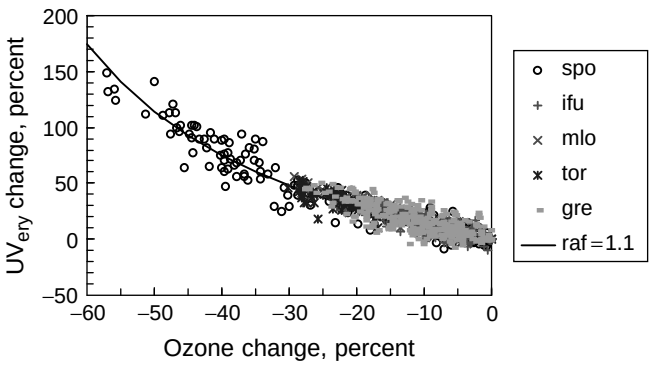


Figure 7 Dependence of erythemal UV radiation at the Earth's surface on stratospheric ozone, measured on cloud-free days at various locations, at fixed solar zenith angles. Legend: South Pole (spo); Lauder, New Zealand (ifu); Mauna Loa, Hawaii (mlo); Toronto, Canada (tor) and Garmisch, Germany (gre). Solid curve (raf) shows predicted relationship from a radiative transfer model of the atmosphere. (Based on Madronich *et al.*, 1998)

A number of studies have shown that sufficient exposure to UV-B, even at pre-ozone destruction levels, can lead to increased incidences of skin cancers and eye disorders, and a reduction in disease immunity, in humans, as well as damage to plants and animals. These effects can be exacerbated by the increase in UV-B at the ground resulting from ozone depletion.

As with other organisms, the human skin has developed various defense mechanisms to deal with exposure to sunlight. The skin adapts to UV exposure by thickening the outer layer, the epidermis, and by developing pigmentation that serves to protect the deeper, more vulnerable, residing dividing cells. Molecular damage is dealt with in the body through repair or replacement. When these natural safeguards are overcome or fail, which can happen with sufficient UV exposure, then serious problems can result. Because such safeguards do not protect the human eye as well, it is particularly sensitive to exposure. Various eye problems, such as the formation of cataracts (a clouding of the lens of the eye) can result. A 10% increase in UV-B is projected to cause a 6–8% increase in the incidence of cataracts. In the US, cataracts are the third largest cause of blindness, and there are about one million cataract operations in the US each year.

Amongst the most serious concerns from UV exposure are several types of skin cancer. Scientific consensus indicates that UV helps cause two types of non-melanoma skin cancer (squamous-cell carcinoma and basal-cell carcinoma) and may be linked through epidemiological evidence to some forms of malignant melanoma. The available data suggest that the incidence of non-melanoma skin cancers increase about 2% for every 1% loss in average stratospheric ozone concentration. In the US, with ozone down as much as 7% in the summertime compared to 30 years ago, there is expected to be a steady increase in squamous and basal cell cancers in the ensuing years, leading eventually to as many as 100 000 extra cases per year compared to the 1960s (there were about 700 000 reported cases in the US in 1994). The largest impacts are expected to occur at the lower latitudes and where there is a high incidence of sunny days. The largest incidence tends to occur amongst fair-skinned people of European origin. While these cancers are treatable, the added medical expenses are still large.

Malignant melanoma is a relatively rare but frequently lethal form of skin cancer. In the US, there are roughly 30 000 cases of melanomas diagnosed each year, with roughly 20% associated fatalities of all cases diagnosed. The link between UV exposure and malignant melanoma is more tenuous than for the non-melanoma skin cancers. Epidemiological evidence suggests that melanoma may be related to a history of sporadic over-exposure of UV radiation. Light skinned persons are 7–10 times more likely to contract malignant melanomas compared to dark-skinned people. A 10% increase in UV-B radiation is projected to cause a 3.5–9% increase in cases of malignant melanoma. At the current rate of increase of incidence of such cases in the US, it has been projected that about 10% of all cancer deaths in the US will be due to melanoma by 2010.

Experimental evidence also suggests that increased exposure to UV-B decreases the effectiveness of the immune system and can aggravate the spread of infectious diseases. There is also concern that UV-B radiation may lower the effectiveness of preventive vaccinations, such as that used against tetanus. There is little known about the current effects on human immunology from the decrease in stratospheric ozone.

Plants and animals are also affected by increases in UV-B. However, knowledge on the effects of increasing UV-B on ecosystems is quite limited. The most studied have been the effects of UV-B exposure on agricultural crops. Soybeans appear to be quite sensitive to increased levels of UV-B. In one study, the equivalent increase in UV-B to a 25% decrease in ozone reduced yields of soybeans by 25%. Marine life, such as zooplankton, phytoplankton and fish larvae are also quite sensitive to increased levels of UV-B. One study estimated that a 25% reduction in ozone would lead to a 35% reduction in primary productivity of phytoplankton near the water surface and a 10% reduction throughout the upper layer of the ocean.

Effects on Climate

A change in the distribution of ozone can affect climate in several ways. The importance of ozone as an absorber of solar radiation, particularly at UV wavelengths, results in a reduction in the amount of solar radiation reaching the Earth s surface. In addition, ozone absorbs solar radiation at visible wavelengths. Ozone is also a greenhouse gas, with several infrared absorption bands in the atmospheric window, where greenhouse gases other than water vapor can have their largest effect on climate. Ozone has a particularly large band at 9.6 μm in the infrared. It is the balance between the solar and infrared radiative processes that determines the net effect of ozone on climate.

The increase in temperature with altitude in the stratosphere, as seen in Figure 1, is explained by the absorption of solar radiation by ozone. The peak absorption defines the top of the stratosphere. In terms of the solar absorption by ozone, this is the altitude at which there is the right balance between the decreasing pressure and density of the atmosphere with altitude and the large concentration of ozone found in the stratosphere. As the available UV radiation decreases with altitude, so does the temperature, even though the amount of ozone remains large throughout the stratosphere.

Increases in stratospheric ozone above 30 km tend to decrease the Earth s surface temperature as a result of the increased absorption of solar radiation, effectively decreasing the solar energy that would otherwise warm the Earth s

surface. Below 30 km, increases in ozone tend to increase the surface temperature, and the infrared greenhouse effect dominates in this region. The large decrease in stratospheric ozone, particularly in the lower stratosphere, over the last several decades has resulted in a net cooling influence on climate, counteracting some of the warming trend over that time period from increasing concentration of CO_2, CH_4, and other greenhouse gases. While ozone has not been a driver of climate change in the way that the change in concentration of CO_2 has been, it is still an important contributor.

Policy Considerations to Protect Ozone

The recognition of the harmful effects of Cl and Br on ozone spawned international action to restrict the production and use of CFCs and halons in order to protect stratospheric ozone. These actions included the 1987 Montreal Protocol on Substances that Deplete the Ozone Layer, the subsequent 1990 London Amendment, the 1992 Copenhagen Amendment, and the 1997 Montreal Amendment. In the Montreal Protocol and its amendments, there is a distinction made between control measures to be taken in developed and developing countries. These agreements initially called for reduction of CFC consumption in developed countries. A November 1992 meeting called by the United Nations Environment Programme and held in Copenhagen resulted in substantial modifications to the protocol because of the large observed decrease in ozone. The Amendment called for the phase-out of CFCs, carbon tetrachloride (CCl_4), and methyl chloroform (CH_3CCl_3) by 1996 in developed countries. As a result of this Amendment, for example, the US, through the Clean Air Act, halted production and import of these chemicals. Production of these compounds is to be totally phased out in developing countries by 2006, while production of halons in developed countries was stopped in 1994. Human-related production and emissions of methyl bromide were not to increase after 1994 in developed countries, and are supposed to slowly decline, with total elimination by 2005. Production and emissions of HCFCs, many of which have been used as replacements for the CFCs but which still contain Cl that can destroy ozone, are to be phased out in the developed countries by 2030.

One of the criteria for the acceptability or unacceptability of a chemical under the Montreal Protocol and the US Clean Air Act is its ozone depletion potential (ODP), a concept developed by Donald Wuebbles, then of Lawrence Livermore National Laboratory of the University of California. This concept provides a relative measure of the effect on ozone from emission of a chemical per unit mass emitted relative to the effect on ozone from the same mass of CFC-11, one of the chemicals of most concern to ozone. By definition, CFC-11 has an ODP of 1.0. Based on their effects on ozone other CFCs also have large ODPs. Most of the compounds being eliminated have ODPs greater than 0.1, although compounds are now being regulated with ODPs smaller than this.

Further policy considerations to protect stratospheric ozone are likely to be primarily aimed at examining potential replacement compounds for CFCs, halons, and HCFCs. A number of the new generation of replacement compounds still contain Cl and/or Br, and in some cases, iodine (I), and could therefore still impact ozone. However, these chemicals are also being chosen in part because they have short atmospheric lifetimes as well as characteristics that make them desirable for a given application. The short atmospheric lifetimes result in more destruction of the chemical in the lower troposphere and therefore a smaller amount of the Cl, Br, or I reaches the stratosphere. As a result, these short-lived chemicals generally have smaller ODPs, making them less likely to affect ozone and more acceptable in terms of their environmental damage.

THE FUTURE OF STRATOSPHERIC OZONE CONCENTRATIONS

Projected Trends in Ozone

Worldwide compliance with the international agreements to protect ozone is resulting in significant reductions in the emissions of the CFCs, halons, and other halocarbons having the largest effects on ozone, As a result, levels of stratospheric ozone should slowly begin to recover over the coming decades as the reactive Cl and Br in the stratosphere declines. This recovery will be gradual, primarily because of the long time it takes for CFCs and halons to be removed from the atmosphere.

Projected changes in globally averaged total ozone based on several different assumptions about future emissions of CFCs, halons and their replacements are shown in Figure 8. These analyses are based on the Cl and Br available to affect ozone, and do not account for other factors that may affect future trends in ozone, as discussed below. With the original Montreal Protocol provisions, the ozone reduction would have continued to grow, although not as rapidly as if there had been no protocol. It is important to note that only under the London and Copenhagen Amendments provisions, calling for the complete phase-out of CFCs, halons and other halocarbons, is the ozone reduction trend reversed. The largest ozone reductions are predicted to have been reached in about 1998–2000. After this, the ozone layer is predicted to gradually recover, although it is not until about 2040 (Copenhagen Amendments) and about 2050 (London Amendments) that the 1980 level is again obtained. Thus, it will be roughly the middle of this century before the Cl and Br in the stratosphere return to levels corresponding to the beginning of the Antarctic ozone hole. Without any control measures, and assuming

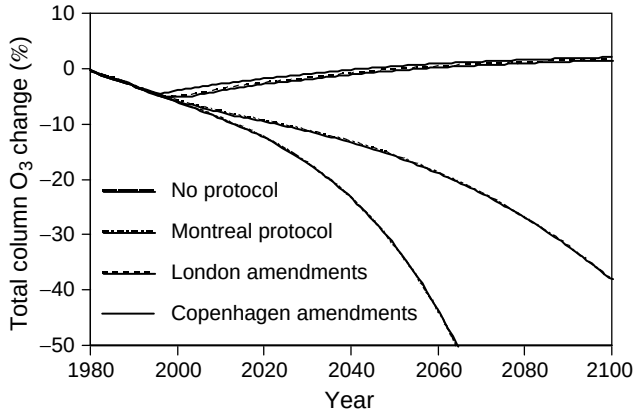


Figure 8 Estimated total column ozone change for the No Protocol, Montreal Protocol, London Amendment and Copenhagen Amendment scenarios. Based on analyses by D Wuebbles and A Jain using their model for equivalent Cl loading at the University of Illinois. Similar analyses can be found in the World Meteorological Organization (1999)

unfettered growth, the global mean total column of ozone would have decreased by about 30% by 2050, and would have continued to decrease with time.

The black line in Figure 8 corresponds to the observed changes in ozone relative to 1980 up to the mid-1990s, and follows the changes in ozone expected if the latest version of the Montreal Protocol is adhered to. However, uncertainty about other changes occurring in the background atmosphere could affect the timing of the return to natural ozone levels, possibly increasing the time by several decades.

Recent studies with complex computer-based models of the global atmosphere suggest that the ozone recovery will likely be slower than the analyses based on the Cl and Br loading in the stratosphere. These studies show that the future behavior and recovery of stratospheric ozone will also be affected by changes in the concentrations of other atmospheric gases like CO_2, CH_4, NO_2, and sulfate aerosols. Changes in the Earth s climate, as projected to result from the effects of increasing concentrations of greenhouse gases, are also likely to affect the recovery of ozone by changing the winds and other transport processes controlling the movement of ozone and other constituents in the stratosphere, by affecting the amount of tropospheric water vapor transported to the stratosphere, and by potentially affecting the tropopause, which is the transition region and barrier for transport between the stratosphere and troposphere. The expected cooling of the stratosphere, due to increasing concentrations of CO_2, could also lead to increased amounts of polar stratospheric clouds and extension of the effectiveness of ozone hole type processes in both the Arctic and Antarctic.

These additional factors, in combination with possible changes in atmospheric variability, suggest that the exact timing of the recovery of ozone, or even knowing when the recovery will begin, is thus uncertain. Some modeling studies suggest that the recovery of ozone could take several decades longer than the simpler scaling presented in Figure 8 (e.g., see the international assessment of ozone published by the World Meteorological Organization, 1998). Other studies are suggesting that the recovery could extend beyond 2100, without the ozone layer ever really returning to the distribution of ozone that existed before human intervention.

THE POTENTIAL FOR NEW IMPACTS ON OZONE

As mentioned above, the projected changes in climate could result in important impacts on stratospheric ozone during the 21st century. The increasing temperatures in the troposphere could change the processes influencing the transport of ozone and constituents in the stratosphere. In addition, changes in the transport of air across the tropopause could result in increased water vapor in the stratosphere, which, in turn, could change the chemistry controlling ozone in the stratosphere. The projected increase in the concentration of CO_2 has a radiative cooling effect on the stratosphere. The resulting cooler temperatures have direct effects on the temperature dependence of the chemistry affecting ozone in the stratosphere, and indirect effects through the increased probability for sufficiently cool temperatures to form polar stratospheric clouds and resulting ozone hole effects on ozone. Projected increases in emissions and atmospheric concentrations of CH_4 and NO_2, also important as greenhouse gases, could affect the balance of ozone destruction from hydrogen oxides and nitrogen oxides in the stratosphere.

Variations in the radiation from the Sun can also affect ozone concentrations. The integrated ozone column varies by about 1.2−1.5% during the 11-year cycle due to solar effects on the photodissociation of oxygen and resulting production of ozone in the stratosphere. Longer-term changes in the solar flux can also occur and could affect future levels of ozone. The current limitations in understanding the processes affecting the output of the Sun limit our ability to predict future changes in the solar flux.

As demonstrated by the 1991 eruption of Mt Pinatubo, large explosive volcanic eruptions can emit significant amounts of sulfur into the stratosphere. This sulfur is rapidly transformed to sulfuric acid aerosols, which can then serve as sites for heterogeneous reactions. Such reactions can convert nitrogen oxides to less reactive nitric acid and convert less reactive hydrochloric acid (HCl) and Cl nitrate ($ClONO_2$) to more reactive forms of Cl. For a few years, until transport processes from the stratosphere remove the sulfuric particles, there can be increased destruction of stratospheric ozone. Such eruptions are likely in the future, and could create further uncertainities in prediction of future changes in ozone resulting from human activities.

The exhaust of solid-fuel rockets can emit gaseous HCl and aluminum oxide particles directly into the stratosphere, leading to potentially large destruction of stratospheric ozone. However, current projections of the frequency of space shuttle flights and other solid-fuel rockets, such as the Titan IV and the Ariane, do not suggest any significant problem with respect to ozone depletion.

A large fraction of the jet aircraft in the current worldwide commercial fleet spend a significant amount of their time at the cruise phase of their flights in the upper troposphere and lower levels of the stratosphere. Projected increases in the number of subsonic aircraft and the tendency towards larger aircraft will mean more of these aircraft will be spending a significant amount of flight time in the lower stratosphere. In addition, there are continuing discussions about the eventual development of a fleet of commercial aircraft that would fly at supersonic speeds (only a few of the existing Concorde supersonic aircraft, which flies at Mach 1.6, were ever built). Emissions of nitrogen oxides and water vapor from aircraft are of particular concern with respect to ozone depletion. Current aircraft fleets are large enough to have a small effect on stratospheric ozone, and research groups continue to evaluate the potential effects of projected fleets of aircraft on ozone over the coming decades. Of particular concern are potential fleets of supersonic aircraft that would fly at higher altitudes, closer to the peak levels of stratospheric ozone. In addition, assessments are being carried out to quantify possible climatic effects created by aircraft emissions (IPCC, 1999).

Atmospheric nuclear explosions dissociate N_2 gas, leading to the production of nitrogen oxides that can destroy ozone. Nuclear explosions near the ground with an explosive power larger than about 1 megaton can loft these nitrogen oxides directly into the stratosphere. Atmospheric nuclear tests in the early 1960s appear to have affected ozone during that period, but such tests are now largely banned under international agreements. However, the devastating effects of a nuclear war would be further exacerbated by the potential for a substantial decrease in the protective ozone layer.

CONCLUSIONS

Human activities have seriously affected levels of stratospheric ozone over the last few decades. The effects that have occurred are now largely understood, and international policy has been put in place to correct the impacts on ozone caused by CFCs, halons and other halocarbons. However, there is the potential for significant perturbations to continue into the future. As a result, the need for research to understand the processes influencing stratospheric ozone and to measure the changes occurring in the atmosphere continues to be an important priority to the benefit of the life on our planet.

See also: Aerosols, Stratosphere, Volume 1; Chloro u-orocarbons (CFCs), Volume 1; Halocarbons, Volume 1; Ozone Depletion Potential (ODP), Volume 1; Solar Variability, Long-term, Volume 1; Stratosphere, Ozone Trends, Volume 1; Stratosphere, Temperature and Circulation, Volume 1; Troposphere, Ozone Chemistry, Volume 1; Industrial Responses to Stratospheric Ozone Depletion and Lessons for Global Climate Change, Volume 4; Ozone Layer: Vienna Convention and the Montreal Protocol, Volume 4; Tradable Permits for Limiting Stratospheric Ozone Depletion, Volume 4.

REFERENCES

Intergovernmental Panel on Climate Change (IPCC) (1999) *Aviation and the Global Atmosphere*, eds J E Penner, D H Lister, D J Griggs, D J Dokken, and M McFarland, Cambridge University Press, Cambridge, 1–373.

Madronich, S, McKenzie, R L, Bjorn, L O, and Caldwell, M M (1998) Changes in Biologically Active Ultraviolet Radiation Reaching the Earth s Surface, *J. Photochem. Photobiol. B*, 46, 5–19.

McPeters, R D, Hollandsworth, S M, Flynn, L E, Herman, J R, and Seftor, C J (1996) Long-term Ozone Trends Derived from the 16-year Combined Nimbus 7/Meteor 3 TOMS Version 7 Record, *Geophys. Res. Lett.*, 23, 3699–3702.

World Meteorological Organization (1999) *Scienti c Assessment of Ozone Depletion: 1998*, WMO, PO Box 2300, Geneva, Switzerland, ISBN: 92-807-1722-7.

FURTHER READING

Brasseur, G, Orlando, J J, and Tyndall, G S, eds (1999) *Atmospheric Chemistry and Global Change*, Oxford University Press, NY.

Makhijani, A and Gurney, K R (1995) *Mending the Ozone Hole: Science, Technology, and Policy*, MIT Press, Cambridge, MA, 1–360.

International Organizations in the Earth Sciences

JOHN S PERRY
Alexandria, VA, USA

By their very nature, environmental changes are often regional or global in their genesis and impact. Yet, human affairs are primarily organized in terms of political jurisdictions from counties to nations that have little relation to the processes of nature. Sulfur emissions from a power plant in Britain do not stop at the English Channel, or at the Swedish border.

Moreover, the branches of human knowledge and activity that may be brought to bear on environmental problems are similarly divided within traditional boundaries that do not correspond readily to real world problems. Universities maintain chemistry departments and law schools; governments have ministries of commerce and transportation; but there is no Department of Acid Rain.

International and interdisciplinary organizations have therefore become ever more essential, and have in fact become ever more effective in advancing our understanding of the global environment and developing sound policies for its protection. For these reasons, the increasing prominence of global-scale environmental issues has motivated increasing emphasis on activities that transcend political and disciplinary boundaries. Depletion of stratospheric ozone was a scienti c problem too large for a college chemistry department; controlling it by worldwide limits on chloro uorocarbon production was a problem too large for any individual nation. Yet effective global action was somehow agreed and acted upon through an international process based on an interdisciplinary scienti c consensus.

In volumes four and ve of this Encyclopedia, the social, political and policy dimensions of international organization for environmental issues will be addressed. Here, we will focus on the nuts and bolts of international organizations involved in encouraging and summarizing global environmental research.

INTRODUCTION

International endeavors in science have a long and highly productive history. Scientists from Aristotle to Einstein have developed many means to communicate observations, theories, insights, and hypotheses among themselves. In the 18th century, to gain large-scale benefit from the collective results of the many individual scientists, learned societies such as the Royal Society in Britain took on roles as national and to a growing extent international clearinghouses for scientific information. In addition, in the 19th century practical applications such as weather forecasting produced further demands for international exchange of data, and organized networks for weather observations were set up. In the late 19th century, growing interest in polar exploration led to organization of the First International Polar Year of coordinated observation and research in 1882–1883. In the 20th century, the advance of global communications and commerce led to the formation of operationally oriented international organizations such as the International Meteorological Organization.

DEVELOPMENT OF INTERNATIONAL ORGANIZATION

The emerging pattern of international cooperation in environmental science was cruelly shattered by World War II. As the planet emerged from that conflict, new tensions and new opportunities arose. The international schism between East and West loomed as a barrier to addressing global scale problems. But at the same time, satellites presented an opportunity to observe the whole Earth for the first time without regard to national boundaries or conflicts. These considerations led in 1958–1959 to the International Geophysical Year (IGY) (*see* IGY (International Geophysical Year), Volume 1), an ambitious attempt to transcend national boundaries and conflicts to study the whole Earth using all available resources of human intellect and technology.

Universally judged to be a resounding success, the IGY provided inspiration for US President John F Kennedy to call upon the United Nations (UN) to utilize the resources of space technology to address the problems of weather and

climate. The resulting Global Atmospheric Research Program (GARP) (*see* GARP (Global Atmospheric Research Program), Volume 1) built upon the organizational pattern of the IGY, again with great success, and laid the foundation for today s organized international activities in weather, climate, and global environmental change.

How are such enterprises conceived, organized, managed, and executed in today s world? Anyone seeking to understand the international framework is confronted with a bewildering smorgasbord of programs, projects, and institutions, each identified by an acronym apparently chosen more for euphony than for meaning. The quest for understanding should start with recognition of what international organizations can reasonably be expected to do. One taxonomy of functions has been presented by Kay and Jacobson (1984):

1. identifying problems;
2. coordinating monitoring and evaluation activities;
3. data gathering and information collection;
4. risk estimation and impact assessment;
5. information exchange and dissemination;
6. facilitation and coordination of national and international programs;
7. normative pronouncements;
8. standard setting and rule making;
9. supervision of norms, standards, and rules;
10. direct operational activity.

In recent years, there has in addition been increased emphasis on capacity building, i.e., enhancing the human and technological resources of countries to address their own environmental problems and to participate in regional and global activities.

INTERGOVERNMENTAL ORGANIZATIONS

The activities behind the acronyms may usefully be classified along several dimensions. The primary distinction is between intergovernmental and non-governmental organizations. The former group is epitomized by the organizations of the UN family; the UN and its institutional bodies, and the specialized agencies that work under its aegis. These organizations are primarily funded by dues, contributions, and project funds from national governments; governed by standing groups of governmental representatives from their constituent nations; and supported by permanent secretariats staffed by international civil servants. For example, the programs of the World Meteorological Organization (WMO) are overseen by the World Meteorological Congress, an assembly of some 185 nations that meets every four years, and an Executive Council that meets annually. WMO s activities are carried out by a couple of hundred staff members based in Geneva, Switzerland (*see* WMO (World Meteorological Organization), Volume 1).

Many of these organizations are closely linked with kindred agencies within their constituent nations. Thus, for example, the Food and Agriculture Organization communicates with and responds to the agriculture ministries of its member nations. Others, such as the United Nations Educational, Scientific, and Cultural Organization (UNESCO), have extremely broad mandates that do not correspond to single bureaucracies in member nations (*see* UNESCO (United Nations Educational Scienti c and Cultural Organization) and the Environment, Volume 4). Moreover, organization within individual nations for various environmental areas sometimes reflects unique political and historical patterns. For example, meteorology is found in departments of environment, agriculture, transportation, and defense in various countries, and is sometimes split between operational and pure research bodies in an unpredictable and confusing pattern. Nevertheless, it is helpful to think of the intergovernmental international bodies as extensions of, and creatures of, their corresponding national agencies.

The intergovernmental organizations carry out an exceedingly broad and diverse array of activities far too numerous to inventory in a brief encyclopedia entry. They convene workshops and symposia to exchange information. They recommend standards to the nations for weather observations, allocation of radio frequencies, protection of intellectual property, and coordination of a host of other topics. They often provide the secretariats for international agreements. On occasion, they perform very significant operational roles ranging from control of locust eruptions to provision of equipment and training. Perhaps their most important functions, however, are simply their ability to place issues of global importance on the agendas of all their constituent nations, and to give voice to the concerns of all nations, large and small, rich or poor. The international organizations have little or no power to command action by sovereign nations, but they have a highly potent power to command attention from their governmental bureaucracies.

A relatively new species of *ad hoc* international organization has arisen in recent years. The prime example is the Intergovernmental Panel on Climate Change (IPCC) (*see* Intergovernmental Panel on Climate Change (IPCC): an Historical Review, Volume 4), which was established under the aegis of the WMO and United Nations Environmental Programme (UNEP) to provide coordinated, authoritative assessments of human influence on global climate. These assessments are prepared in order to inform the work of the intergovernmental Framework Convention on Climate Change (*see* United Nations Framework Convention on Climate Change and Kyoto Protocol, Volume 4). Although consisting of scientists nominated by governments, the IPCC nevertheless operates independently to provide unbiased assessments of current scientific understanding of human impacts on global climate.

NON-GOVERNMENTAL ORGANIZATIONS

The second principal stream of international activity is led by non-governmental organizations. These diverse organizations carry out the day to day work of scientific communication and exchange that makes science (particularly environmental science) an efficient and productive global endeavor.

Loose associations of scientists in different countries have existed for many years. Beginning in the 19th century, national academies of science or their equivalents were formed, and formal societies or unions began to develop in various disciplines, as scientists began to realize the benefits of exchanging data and ideas. In 1931, this growing structure was brought together in the International Council of Scientific Unions (ICSU), now renamed the International Council for Science (*see* ICSU (International Council for Science), Volume 4). Currently, this comprises 98 national member organizations (academies of science or the equivalent), 26 disciplinary scientific unions, and 28 associated scientific organizations. Although the boundary between governmental and non-governmental science is blurred in many countries, the key feature of ICSU and its family of activities is its governance by scientists acting in their individual capacities rather than representatives of political bodies. This independence gives ICSU a unique degree of freedom, independence, and authority.

Among the unions operating under the aegis of ICSU, the International Union of Geodesy and Geophysics (IUGG) is particularly relevant to global environmental change. The IUGG comprises a family of semi-autonomous international associations whose fields of activity span the Earth sciences:

1. the International Association of Geodesy (IAG);
2. the International Association of Geomagnetism and Aeronomy (IAGA);
3. the International Association of Hydrological Sciences (IAHS) (*see* IAHS (International Association of Hydrological Sciences), Volume 1);
4. the International Association of Meteorology and Atmospheric Sciences (IAMAS) (*see* IAMAS (International Association of Meteorology and Atmospheric Sciences), Volume 1);
5. the International Association for the Physical Sciences of the Ocean (IAPSO) (*see* IAPSO (International Association for the Physical Sciences of the Oceans), Volume 1);
6. the International Association of Seismology and Physics of the Earth s Interior (IASPEI);
7. the International Association of Volcanology and Chemistry of the Earth s Interior (IAVCEI).

Each of these maintains a substructure of specialized committees or commissions, convenes dozens of international symposia and workshops each year, and holds periodic international assemblies of the association as a whole. For example, IAMAS has commissions focusing on atmospheric chemistry and global pollution, climate, and ozone among its 10 specialized commissions.

Much of ICSU s work is carried out through a network of interdisciplinary scientific and special committees focused on important topics that transcend disciplinary specialities. Among the most relevant to global environmental change are:

1. Scientific Committee on Antarctic Research (SCAR);
2. Programme on Capacity Building in Science (PCBS);
3. Committee on Data for Science and Technology (CODATA);
4. Scientific Committee on Problem of the Environment (SCOPE) (*see* SCOPE (Scienti c Committee on Problems of the Environment), Volume 4);
5. Scientific Committee for the International Geosphere–Biosphere Programme (SC–IGBP) (*see* IGBP (International Geosphere- Biosphere Programme), Volume 2);
6. Special Committee for the International Decade for Natural Disaster Reduction (SC-IDNDR) (*see* IDNDR (The International Decade for Natural Disaster Reduction), Volume 4);
7. Scientific Committee on Oceanic Research (SCOR) (*see* SCOR (Scienti c Committee on Oceanic Research), Volume 1);
8. Scientific Committee on Solar–Terrestrial Physics (SCOSTEP);
9. Committee on Space Research (COSPAR);
10. Scientific Committee on Water Research (SCOWAR).

Other important projects are carried out jointly with intergovernmental or non-governmental bodies, e.g., the World Climate Research Programme (WCRP), which is implemented jointly with the WMO and the Intergovernmental Oceanographic Commission (IOC) of UNESCO (*see* WCRP (World Climate Research Programme), Volume 1). The very broadly interdisciplinary IGBP, which works closely with the WCRP, the International Human Dimensions Programme (IHDP), the SCOR, and the Global Change System for Analysis Research and Training (START) regional networking system, has focused on developing detailed plans for the conduct of science in its eleven component program elements. These are composed of eight broadly discipline oriented projects, covering such topics as atmospheric science, terrestrial ecology, oceanography, hydrology, and links between the natural and the social sciences. Three framework activities on data, modeling, and regional research, facilitate incorporating scientific results into a holistic picture.

ICSU also provides a home for important international data acquisition, management, and archiving services relating to such parameters as sea level, glaciers, Earth rotation,

and so on. Typically, secretariats or project offices for these activities are established under the aegis of some host institution, often with major support from the host country. For example, the IPCC secretariat is housed by the WMO; ICSU s IGBP occupies a floor of the Royal Swedish Academy of Sciences; and ICSU itself inhabits a gracious townhouse in Paris provided by France.

Focused projects organized and coordinated through international organizations play an especially important role in environmental science. The observational data required to study phenomena such as the El Niño or pollutant transport cannot be gathered by one nation alone. Moreover, the attendant data processing, archiving, and research often demand resources greatly exceeding those of any single nation. It has therefore become common for dozens of nations to collaborate very actively over extended periods in scientific programs. During one 4-month field experiment of the Tropical Ocean Global Atmosphere (TOGA) program (*see* TOGA (Tropical Ocean Global Atmosphere), Volume 1), for example, nearly 1200 people from more than 20 nations conducted more than 700 days of ship operations, launched nearly 12 000 rawinsondes, completed 125 aircraft flights, and maintained continuous operation of 30 moored instrument systems. The International Satellite Cloud Climatology Program currently collects data from European, Japanese, and US satellites, calibrates the data in France, and analyses and archives it in the US (*see* ISCCP (International Satellite Cloud Climatology Project), Volume 1). In global scale environmental science, international organization is neither a luxury nor an afterthought; it is becoming the primary mode of performing effective research.

Numerous other international bodies relevant to environmental issues, many of which are scientific associates of ICSU, merit mention. For example, the International Institute for Applied Systems Analysis (IIASA) in Austria, governed and funded through member organizations in a group of countries, carries out significant innovative interdisciplinary research in demographics, energy systems and environmental impacts (*see* IIASA (International Institute for Applied Systems Analysis), Volume 4). The Third World Academy of Sciences (TWAS) supports scientific excellence and research in developing countries. The World Wildlife Fund (WWF) has a diverse constituency of both governmental and non-governmental partners, and carries out a broad range of conservation activities worldwide.

A full inventory of international organizations of all kinds relevant to global environmental change would require an encyclopedia of its own. Indeed, the complex and often overlapping networks of institutions, organizations, and programs are better suited to the hypertext of the World Wide Web than to the printed page. This complexity is, however, neither accidental nor perverse. Rather, it reflects the very real complexity and interconnection of both the natural and human worlds, and the need to balance disciplinary depth and rigor against the challenge of real world problems. Indeed, a former president of ICSU, Sir John Kendrew, once remarked that the distinguishing characteristic of 20th century science was the willingness to address messy, poorly defined (but important) problems that do not fall neatly within national or disciplinary boundaries. The overarching problem of global environmental change is the archetype of such problems.

See also: Scienti c Responses in an Era of Global Environmental Change, Volume 4).

ACCOMPLISHMENTS

The international organizations dealing with the environment should therefore be assessed in terms of their effectiveness in meeting these challenges, not in terms of any preconceived notions of bureaucratic neatness. From this standpoint, the international organizations have many successes to their credit. A very few examples drawn from their rich array of achievements are illustrative:

1. UNEP s orchestration of international agreements to reduce pollution of the Mediterranean.
2. The Global Atmospheric Research Programme of WMO, ICSU, and IOC that made today s long-range weather prediction possible.
3. The TOGA program, again a WMO/ICSU/IOC effort that elucidated the El Niño/Southern Oscillation phenomenon and developed seasonal/interannual forecasting capabilities.
4. The international assessments of ozone depletion that laid the foundation for control of ozone depleting chemicals, and thus averted what might well have been a global environmental disaster.
5. The comprehensive and authoritative interdisciplinary syntheses, assessments, and studies of natural and human made environmental changes and the effects of these changes on people carried out by ISCU s SCOPE.
6. The IPCC that is similarly laying a firm foundation for international efforts to limit and cope with damaging human impacts on global climate.

CONCLUSION

Ultimately, the work of the world is carried out by individual scientists and citizens, with resources primarily mobilized by the governments and corporations that levy taxes and render bills. However, the legal, political, and disciplinary boundaries that permit scientists and citizens to focus effectively on their individual tasks can at the same time hinder coordinated global action on problems of importance to humanity as a whole. International organizations of all kinds play a vital role in fostering communication and coordination across

these barriers, in providing a framework for planning and executing collaborative activities, and in providing a common agreed base of understanding on which the nations can build sound policies for securing humanity s future on our precious and unique planet.

REFERENCE

Kay, D A and Jacobson, H K (1984) *Environmental Protection - The International Dimension*, Allanheld, Osmun, Totowa, NJ.

FURTHER READING

Perry, J S (1983) International Organizations and Climate Change, in *World Climate Change - The Role of International Law and Institutions*, ed V P Nanda, Westview Press, Boulder, CO.

Perry, J S (1984) The GARP, in *Scienti c and Technological Cooperation among Industrialized Countries*, National Academy of Sciences, Washington, DC.

Perry, J S (1986) Managing the World Environment, *Environment*, 28(1), 10–40.

A

Abrupt Climate Change

see Cloud–Radiation Interactions (**Volume 1**)

Acoustic Thermometry

Acoustic thermometry is a method for measuring the average ocean temperature over large areas. This is particularly appropriate when trying to see if the climate is changing. The method is based on the fact that sound travels faster in warm water than in cold water. Thus, the travel time of an acoustic pulse can be used to determine the mean temperature of the intervening water between source and receiver. For a typical measurement designed to be taken over a distance of several thousand kilometers, there are many tens of possible transmission paths, and associated arrival times, for a single emitted sound pulse. Early arrival times indicate travel of the sound along steep rays oscillating between the surface and the bottom of the ocean, such that their travel time is affected by the temperature profile in the entire water column. Late arrival times indicate sound paths that hug a sound axis that exists typically in cold waters at 1 km depth. Thus, each path of a sound pulse weights the water column in a different way. The set of arrival times can be analyzed to estimate the vertical temperature profile and the horizontally averaged heat content between source and receiver. The basin-scale averages suppress the relatively large fluctuations associated with ocean weather that can hide the slowly changing climate signal.

This technique has proven useful in polar regions that are ice covered during winter. A 3000 km transmission path across the North Pole shows a 0.4 °C warming relative to average conditions over the past few decades. An experiment in the western Mediterranean has given accurate measurements of seasonal heat content. Transmissions in the western Pacific from a source on Pioneer Seamount off California to various sensor arrays near Hawaii and Christmas Island have been compared to seasonal and year-to-year changes in sea level (which is also affected by ocean temperatures) determined by satellite altimetry, and it has been shown that these two methods together lead to accurate estimates of heat storage in the ocean.

See also: Munk, Walter, Volume 1.

FURTHER READING

Munk, W, Worcester, P, and Wunsch, C (1995) *Ocean Acoustic Tomography*, Cambridge University Press, Cambridge.

WALTER MUNK *USA*

ACSYS (Arctic Climate System Study)

ACSYS, a component of the World Climate Research Programme, concentrates on the understanding of Arctic Ocean variability and change including sea ice processes. It will be expanded into the Climate and Cryosphere Initiative (CLIC) investigating the role of the entire cryosphere for global climate and as an early indicator of change.

The scientific goal of ACSYS is to ascertain the role of the Arctic in global climate:

- global consequences of natural or man-made change in the Arctic climate system;
- sensitivity of the Arctic climate system to increased greenhouse gases.

The project includes several subprograms.

ARCTIC OCEAN CIRCULATION PROGRAM

This provides the information necessary to realistically represent the role of the Arctic Ocean in the global climate

system in models of global climate. The Arctic Ocean Hydrographic Survey collects high-quality hydrographic data on physical and chemical properties of the Arctic Ocean. Arctic Ocean Shelf Studies are aimed at understanding shelf processes. The Arctic Ocean Variability Project assesses the variability of the circulation and density structure of the Arctic Ocean. The historical Arctic Ocean Climate Database Project builds a digital hydrographic database for the Arctic Ocean.

ARCTIC SEA ICE PROGRAM

Arctic Ocean pack ice has dramatic effects on the physical characteristics of the ocean surface, and plays a role in the global heat balance and thermohaline circulation. The program seeks to assemble a climatological archive of pack ice in both the Arctic and Antarctic seas, and to study the interaction of polar pack ice with other elements of the global climate system.

ARCTIC ATMOSPHERE PROGRAM

This program includes field campaigns, re-analysis of data, polar clouds and radiation research, *in situ* and satellite measurements, and improved modeling.

HYDROLOGICAL CYCLE IN THE ARCTIC REGION

This program includes compilation of an Arctic hydrological database, development of hydrological models, and subprojects on solid precipitation and development of an Arctic precipitation data archive.

ACSYS MODELING PROGRAM

Models will be optimized for the Arctic region, parameterizations will be developed, and numerical experiments performed.

DATA MANAGEMENT AND INFORMATION PANEL

This group develops a strategic approach to data management and information in ACSYS.

Information is available from the International ACSYS/CLIC Project Office, The Polar Environmental Center, N-9296 Troms , Norway.

JOHN S PERRY *USA*

Aerosols, Effects on the Climate

Joyce E Penner
University of Michigan, Ann Arbor, MI, USA

Aerosols are small particles suspended in the atmosphere. While technically the term can refer to both particles and cloud droplets, it is common to use this term to only refer to particles which are smaller than about 10 μm*. Aerosols are important because they are able to alter climate in two important ways. First, they scatter and absorb radiation in the atmosphere and, second, they change the microphysical structure and possibly the lifetime and extent of clouds within the atmosphere. Radiation in the atmosphere is divided into solar and thermal radiation. The former has wavelengths ranging from 0 to* ∼*4* μm*, while the latter has wavelengths ranging from 4- 20* μm*. Because most aerosols are smaller than 10* μm *and because aerosol particles interact most strongly with wavelengths near their size, most of the scattering and absorption of radiation by aerosols takes place in the solar spectrum. An exception to this rule applies to particles with diameters greater than about 2* μm *which consist primarily of dust particles and sea salt particles as well as compounds which condense onto or react with these aerosols. The scattering of solar radiation acts to cool the planet, while absorption of solar radiation by aerosols warms the air directly instead of allowing sunlight to be absorbed by the surface of the Earth. When sunlight is absorbed by the Earth's surface, the heat is then transferred back to the atmosphere either in the form of thermal (heat) radiation, by conduction of heat, or by evaporation of water which, when it condenses, releases the heat back to the atmosphere. Large particles in the troposphere (diameter (d)* $>$ *2* μm*) will absorb the thermal radiation. Absorption of this radiation by these particles tends to warm the planet.*

As a result of their abundance and size, cloud droplets both scatter solar radiation and absorb thermal radiation. Warm liquid water clouds at low altitudes tend to mainly scatter solar radiation, while ice clouds at high altitude mainly absorb thermal radiation. Thus, increases in the former act to mainly cool the planet while increases in the latter tend to warm it. Aerosols can impact clouds by acting as cloud condensation nuclei (CCN) (i.e., the particles on which water vapor condenses as cloud drops form). Increases in aerosols will tend to increase the number concentrations of warm liquid water cloud droplets (and decrease their size). Because increases in the number concentration of cloud droplets in warm clouds may greatly increase their scattering of solar radiation (while having little effect on their interaction with thermal radiation as

long as the total amount of water in the cloud remains constant), the primary effect of increases in aerosols is to tend to cool the planet. Little is known about how aerosols may impact ice clouds.

A secondary effect of increases in aerosols might occur if changes in droplet concentrations alter the lifetime and extent of clouds. Cloud lifetime might be increased because smaller droplets may require longer times to develop into precipitating clouds, resulting in more extensive cloud cover and/or deeper clouds. The issue is complex because the absorption of solar radiation by clouds may lead to dynamical responses in a cloud that either extend its vertical structure and lifetime or reduce it.

Changes in aerosol concentrations over time have the potential to greatly impact the radiation balance and therefore climate. Because aerosols primarily tend to cool the climate, if aerosol emissions decrease (as, for example, if controls limit future emissions of SO_2), warming by greenhouse gases may be expected to accelerate.

SOURCES OF AEROSOLS

Particles in the atmosphere are most often composed of sulfate, nitrate, ammonium, organics, silicates and other compounds associated with soils and dust, sea salt and compounds associated with the sea salt aerosol, soot or black carbon and trace metals (the total trace metal mass is small and is therefore not described here due to its very small climatic influence). The sources of these aerosol components vary significantly with region and time of year. This, together with the fact that the aerosol lifetime is relatively short (about five days), implies that the concentrations of the components also vary significantly with location and time.

Aerosol sources may be divided into gas phase sources and primary particulate sources (*see* Aerosols, Troposphere, Volume 1). *Gas phase sources* may either directly condense to the particulate phase, or they may be oxidized in the atmosphere through photochemical reactions to produce gases that may condense to the particulate phase. An example of the former is ammonia (NH_3) while the latter includes sulfur dioxide (SO_2) or dimethylsulfide (DMS), which may oxidize to form sulfate (H_2SO_4) (*see* Dimethylsul de (DMS), Volume 1) and nitrogen oxides (NO_x), which is converted to nitric acid (HNO_3). H_2SO_4 has low volatility and will partition almost wholly to the particulate phase while NH_3 and HNO_3 are both semi-volatile and their partitioning into the particle phase depends on local conditions. Organic gases can also oxidize to form either semi-volatile compounds or compounds with low volatility. The sources of gas phase precursors are summarized in Table 1.

Primary particulate sources are emitted as particles at or near the source. They include smoke from fires, dust uplifted by wind, sea salt from direct injection and bursting of bubbles, and biogenic particles (fungi, spores, epicuticular plant waxes etc.). Fires mainly originate from anthropogenic activity (burning for agriculture, to clear land, burning of fossil fuels). Dust outbreaks have also been associated with anthropogenic activity, since agricultural processing of the land can lead to more frequent dust outbreaks and because some desertification is associated with land use changes (*see* Dust, Volume 1). Tegen and Fung (1995) estimated that about 50% of their total dust source strength was associated with land use changes.

Table 2 presents NH and SH average sources for primary particulates according to aerosol size ($d < 2\,\mu m$ and $2 < d < 20\,\mu m$). The fine particle fraction of the atmospheric aerosol (defined as that portion with diameter less than $2\,\mu m$) has been studied the most because it generally contains the largest number fraction and is generally the most important fraction for determining the direct radiative forcing. The composition and number concentration of the sub-micron fraction is also important for determining the indirect effects of aerosols. However, because both the small and large fractions act both as available surfaces for the reaction of gas-phase aerosol precursors and as condensation sites for aerosol precursors, sources for the entire aerosol size spectrum must be quantified in order to quantify the sub-micron component. Also, in some circumstances, the available surface area for large particles can control the supersaturation maximum reached within an updraft as clouds form so that the abundance of large, highly soluble aerosols such as sea salt can partly determine droplet cloud droplet concentrations in clouds. Furthermore, in regions with high concentrations of dust and sea salt particles, the large particle component can be important to direct forcing. Thus, the budget for all sizes of aerosols needs quantification in order to determine the magnitude of the forcing of aerosols on climate.

If one assumes that the airborne portion of HNO_3 is in equilibrium with the sulfate aerosol and NH_3 concentrations, one can derive estimates of the total aerosol source strength from the gas phase aerosol precursor emissions in Table 1. These are presented in Table 3.

EFFECTS OF AEROSOLS ON CLIMATE

Because the sources of anthropogenic aerosols are a relatively large fraction of the total source strength, it is important to try to understand the possible impact of these recently increasing aerosol concentrations on past and future climate change (*see* Climate Change, Detection and Attribution, Volume 1). One method for estimating their possible climate impact is to estimate the radiative forcing associated with the anthropogenic fraction of aerosol. *Radiative forcing* is the change in net incoming radiation

Table 1 Annual source strength for emissions of gas phase aerosol precursors (Tg N or S or C year^{-1})

	NH	SH	Global	Range
NO_x (as Tg N year^{-1})				
Fossil fuel	19.9	1.1	21.0	15–27
Aircraft	0.54	0.04	0.58	0.4–0.9
Biomass burning	3.3	3.1	6.4	5.0–10.0
Soils	3.5	2.0	5.5	3–12
Agricultural soils			2.2	0–4
Natural soils			3.2	3–8
Lightning	3.7	2.4	7.0	2–12
NH_3 (as Tg N year^{-1})				
Domestic animals	17.5	4.1	21.6	10–30
Agriculture	11.5	1.1	12.6	6–18
Human	2.3	0.3	2.6	1.3–3.9
Biomass burning	3.5	2.2	5.7	3–8
Fossil fuel and industry	0.29	0.01	0.3	0.1–0.5
Natural soils	1.4	1.1	2.4	1–10
Wild animals	0.10	0.02	0.1	0–1
Oceans	3.6	4.5	8.2	3–16
SO_2 (as Tg S year^{-1})				
Fossil fuel and industry	68	8	76	65–85
Aircraft (1992)	0.04	0.002	0.06	0.03–1.0
Biomass burning	1.2	1.0	2.2	1.5–4.0
Volcanoes[a]	6.3	3.0	9.3	6–15
DMS or H_2S (as Tg S year^{-1})				
Oceans	11	13	24	16–32
Land biota and soils	0.6	0.4	1.0	0.4–5.6
Volatile organic emissions (as Tg C year^{-1})				
Anthropogenic	1.04	9	109	70–150
Terpenes	67	60	127	90–180

[a] Volcanic emissions are annually averaged emissions to the troposphere. Explosive volcanoes may emit much larger amounts to the stratosphere.
NH, Northern Hemispheric; SH, Southern Hemispheric.

at the top of the troposphere caused by a specific change in composition (*see* Radiative Forcing, Volume 1). It gives a first-order measure of the long-term steady state response of the climate because the global average change in surface temperature is approximately linearly related to the radiative forcing. Thus,

$$\Delta T_s \approx \lambda \Delta F \quad (1)$$

where λ is defined as the climate sensitivity (K (W m^{-2})$^{-1}$) and ΔF is the radiative forcing. The climate sensitivity depends primarily on the long-term changes in water vapor (the main greenhouse gas in the atmosphere), clouds, sea-ice cover and snow. The magnitude of λ has been estimated from climate models to range from 0.2 to 1.1 K (W m^{-2})$^{-1}$. This range of variability results from different representations of physical processes within climate models.

It is important to note that radiative forcing is only an approximate measure of climate effects. It relies on the concept that the surface temperature change is effectively communicated throughout the troposphere by convective and radiative processes. However, because emissions into specific regions can alter the vertical profile of local temperature as well as the local cloud fraction and water concentrations, the concept is only approximately valid. Nevertheless, it provides a first-order estimate of the relative impacts of different emissions. Radiative forcing by aerosols may be compared to the forcing from greenhouse gases and from changes in solar radiation over the past 100 years. These latter forcings were estimated as $\sim 2.5\,\mathrm{W\,m^{-2}}$ and $-0.3\,\mathrm{W\,m^{-2}}$, respectively, in the recent Intergovernmental Panel on Climate Change scientific assessments (IPCC, 1996, 2001).

DIRECT EFFECTS OF ANTHROPOGENIC AEROSOLS

Recent estimates for the direct forcing by anthropogenic aerosols are between -0.3 to $-0.9\,\mathrm{W\,m^{-2}}$ for sulfate aerosols, between -0.15 and $-0.25\,\mathrm{W\,m^{-2}}$ for biomass

Table 2 Primary particle emissions (Tg year^{-1})

	NH	SH	Global	Low	High
Organic matter (0–2μm)					
Biomass burning	28	26	54	45	100
Fossil fuel	28	0.4	28	10	30
Biogenic (0–2 μm)			50	10	100
Black carbon (0–2μm)					
Biomass burning	2.9	2.7	5.6	5	10
Fossil fuel	6.5	0.1	6.6	5	10
Aircraft	0.005	0.0004	0.006		
Industrial dust, etc.			100	40	130
Sea salt					
$d < 1\,\mu m$	23	31	54	18	100
$d = 1–16\,\mu m$	1420	1870	3290	1000	6000
Total	1440	1900	3340	1000	6000
Mineral (soil) dust					
<1 μm	90	17	110		
1–2 μm	240	50	290		
2–20 μm	1470	282	1750		
Total	1800	349	2150	1000	5000

Table 3 Estimates for secondary aerosol sources (Tg year^{-1})

	NH	SH	Global	Low	High
Sulfate (as NH_4HSO_4)					
Anthropogenic	116.0	16.5	132.6	80	175
Biogenic	20	30.0	50	30	100
Volcanic[a]	21.6	13.0	35.5	10	70
Nitrate (as NO_3)					
Anthropogenic	12	0.9	14	7	20
Natural	2.2	1.7	3.9	1.0	8
Organic compounds					
Anthropogenic VOC[b]	0.55	0.05	0.6	0.3	2.0
Biogenic VOC	8.2	7.4	15.6	8	40

[a] Volcanic emissions are annually averaged emissions to the troposphere. Explosive volcanoes may emit much larger amounts to the stratosphere.
[b] Volatile organic carbon (VOC).

aerosols (from biomass fuel use and land clearing), and between $+0.15$ and $+0.2\,W\,m^{-2}$ for fossil fuel and industrial sources of organic and black carbon aerosols. Figure 1 shows the spatial distribution of the estimated forcing of anthopogenic sulfur and carbonaceous aerosols (Penner *et al.*, 1998). Estimates of climate forcing from anthropogenic nitrates and anthropogenic ammonia are small compared to forcing by sulfates. Anthropogenic sources of dust aerosols are estimated to have a global forcing of $+0.09\,W\,m^{-2}$ according to Tegen *et al.* (1996). Sea salt, while considered natural, appears to be the most important aerosol component with a total radiative effect of up to about $-1.0\,W\,m^{-2}$. The climate forcing associated with anthropogenic aerosols could offset a substantial fraction of the warming associated with greenhouse gases. If the ratio of aerosol emissions to greenhouse gas emissions remains constant, however, these impacts should decrease in the future, because aerosols have a lifetime in the atmosphere of order five days so that their concentrations do not increase as rapidly as those of most greenhouse gases (many of whom have lifetimes that are much longer). In addition, it should be noted that if SO_2 emissions are controlled in the future, more rapid warming should occur.

EFFECTS OF AEROSOLS ON CLOUDS

While estimates of climate forcing by aerosols associated with the direct reflection and absorption of radiation are

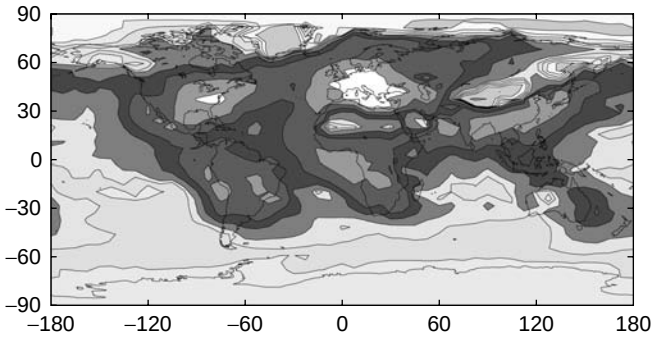


Figure 1 The sum of direct forcing by anthropogenic sulfate aerosols, biomass aerosols and organic matter and black carbon from fossil fuels. (Reprinted with permission from Penner *et al.*, 1998, © Springer-Verlag)

reasonably straightforward, estimates of the indirect effects of aerosols on clouds are much more difficult. This is mainly because the indirect effect of aerosols depends on knowing the natural as well as anthropogenic sources, the state of mixing of the different aerosol compounds, the fate of gas phase precursors which may either nucleate to form new particles that grow to a size range that would act as a CCN or condense on pre-existing aerosols, and the response of clouds (i.e., their droplet concentration and their vertical and horizontal extent) to changes in aerosols and droplet concentrations.

The state of mixing of aerosol has been shown to lead to uncertainties ranging from $-0.4\,\mathrm{W\,m^{-2}}$ to $-1.5\,\mathrm{W\,m^{-2}}$ in forcing by sulfate aerosols associated with changes in drop size distributions (see Chuang *et al.*, 1997). Changes in droplet concentrations due to carbonaceous aerosols from biomass burning and fossil fuel/industry lead to estimates of forcing that range from -1.2 to $-1.6\,\mathrm{W\,m^{-2}}$ depending on the magnitude of natural organics assumed in the calculation (Penner *et al.*, 1999). Changes in the magnitude of natural sources of dust and sea salt could change this estimate by a factor of 2. The estimates by Chuang and co-workers as well as those by Penner *et al.* (1999), assume that the total lifetime and extent of clouds remain constant. If this constraint is relaxed, then models have shown that the total forcing can more than double. Thus, the total indirect forcing by anthropogenic aerosols may range from about -0.3 to $-3.0\,\mathrm{W\,m^{-2}}$. It is not clear yet whether the upper range of these estimates for indirect forcing are far too large, but if they are correct, then our understanding of recent climate changes and the consistency of the historical record of temperature change with the record of radiation forcing would need to be altered.

As noted above, increases in high cirrus clouds may act to warm the climate. Aerosols from aircraft have been associated with observed increases in cirrus clouds that have occurred since the early 1970s (Fahey *et al.*, 1999). However, these increases may also be associated with surface level emissions, or they may be associated with other, natural cycles. Thus, there is not yet an estimate for the impact of aircraft exhaust on cirrus cloudiness and the associated radiative forcing.

FUTURE EFFECTS OF AEROSOLS

Scenarios for future emissions of SO_2 and other aerosols are highly uncertain. Some scenarios have SO_2 emissions increasing along with industrial emissions of CO_2. In these scenarios, the impact of aerosols on radiative forcing does not keep pace with the radiative forcing associated with CO_2 because the aerosols are removed from the atmosphere far more quickly than is CO_2. In other scenarios, SO_2 emissions are controlled (as they are now controlled in much of the developed world). These scenarios imply more rapid warming than scenarios which allow SO_2 emissions to grow as the overall industrial activity grows. However, additional effects must also be considered. For example, the aerosols from so-called natural sources may change. This is because the aerosols associated with sea salt, dust and DMS generation are dependent on wind speed and also on soil moisture in the case of dust. Aerosols from natural terpene emissions depend on temperature and light levels as well as changes in soil moisture. Thus, changes in environmental conditions associated with climate change itself can alter emissions of aerosols. Initial estimates of the possible changes in the emissions of DMS (*see* Dimethylsul de (DMS), Volume 1), terpenes, and dust aerosols by the year 2100 indicate that increases in DMS emissions may be relatively small, while both dust and terpene emissions may increase by about 10 and 20% on an annual average basis, respectively. Sea salt emissions may increase by 75% in the year 2100. Such changes could lead to an additional negative forcing of up to approximately $-1\,\mathrm{W\,m^{-2}}$ from direct forcing and an additional -1 to $-2\,\mathrm{W\,m^{-2}}$ from indirect forcing. Future climate models must consider both the projected changes in anthropogenic aerosols as well as the changes in natural aerosols and aerosol precursors associated with climate change.

See also: Aerosols, Stratosphere, Volume 1.

REFERENCES

Chuang, C C, Penner, J E, Taylor, K E, Grossman, A S, and Walton, J J (1997) An Assessment of the Radiative Effects of Anthropogenic Sulfate, *J. Geophys. Res.*, 102, 3761–3778.

Fahey, D W, Schumann, U, Ackerman, S, Artaxo, P, Boucher, O, Danilin, M Y, Karcher, B, Minnis, P, Nakajima, T, and Toon, O B (1999) Aviation-produced Aerosols and Cloudiness, in *Aviation and the Global Atmosphere*, eds J E Penner, D H Lister, D J Griggs, D J Dokken, and M McFarland, Cambridge University Press, Cambridge.

IPCC (1996) *Climate Change 1995: The Science of Climate Change*, eds J T Houghton, L G Meira Filho, B A Callander, N Harris, A Kattenberg, and K Maskell, Cambridge University Press, Cambridge.

IPCC (2001) *Climate Change 2001: The Scienti c Basis*, eds J T Houghton, Y Ding, D J Griggs, M Noguer, P J van der Linden, X Dai, K Maskell, and C A Johnson, Cambridge University Press, Cambridge.

Penner, J E, Chuang, C C, and Grant, K (1998) Climate Forcing by Carbonaceous and Sulfate Aerosols, *Clim. Dyn.*, 14, 839–851.

Penner, J E, Chuang, C C, and Grant, K (1999) Climate Change and Radiative Forcing by Anthropogenic Aerosols: a Review of Research During the Last Five Years, La Jolla International School of Science, The Institute for Advanced Physics Studies, La Jolla, CA, submitted for publication.

Tegen, I and Fung, I (1995) Contribution to the Atmospheric Mineral Aerosol Load from Land Surface Modification, *J. Geophys. Res.*, 100, 18 707–18 726.

Tegen, I, Lacis, A A, and Fung, I (1996) The Influence on Climate Forcing of Mineral Aerosols from Disturbed Soils, *Nature*, 380, 419–422.

Aerosols, Polar Stratospheric Cloud

Giovanni Pitari
Università L'Aquila, L'Aquila, Italy

Polar stratospheric cloud (PSC) aerosols are one of the most important types of particles present in the middle atmosphere. They are normally made of nitric acid hydrates or pure water ice and may only be formed at very cold temperatures (below about 200 K) in the wintertime polar vortices. These polar particles are regarded as one of the major factors responsible for the formation of the Antarctic ozone hole and for the presently detected negative trends of ozone in the Northern Hemisphere. Ef cient heterogeneous chemical reactions may in fact take place on the surface of these particles, producing large increases of reactive chlorine and bromine in the polar vortices with simultaneous presence of sunlight. Future increases of stratospheric water vapor and nitric acid, coupled to extensive stratospheric cooling caused by increasing burdens of greenhouse gases, are believed to be potentially capable of enhancing PSC formation (IPCC, 1999), thus delaying the future recovery of stratospheric ozone.

Sulfuric and Polar stratospheric cloud (PSC) particles are regarded as the most important aerosols in the stratosphere. They alter the ozone s chemical budget and may then affect the amount of ultraviolet radiation reaching the ground surface (*see* Stratosphere, Chemistry, Volume 1). Heterogeneous chemical reactions (see Table 1) may take place on the surface of these particles, and represent an important sink of reactive nitrogen at mid-high latitudes. In turn, a decrease of odd nitrogen makes chlorine and bromine catalytic cycles for ozone destruction faster than in a pure gas phase chemistry atmosphere. Reactions involving chlorine nitrate ($ClONO_2$) and hydrochloric acid (HCl) are very efficient on PSC particles and many scientific investigations have pointed out their role in the formation of the Antarctic springtime ozone hole. In the presence of sunlight, the temporary reservoir species formed by these reactions (Cl_2, HOCl, $ClNO_2$) are readily photodissociated, forming reactive chlorine. Aircraft data collected in the 1992 Arctic winter show dramatic losses of HCl and $ClONO_2$ with corresponding build-up of ClO when entering the polar vortex from the mid-latitudes (World Meteorological Organization (WMO), 1995).

PSC aerosols are complex particles made essentially by condensation of nitric acid and water vapor on preexisting nuclei (mostly sulfuric aerosols). However, while sulfuric acid condensation may occur everywhere in the middle atmosphere below about 35 km, nitric acid and water vapor condensation take place only at very cold temperatures in the wintertime polar vortices ($\approx$10–25 km altitude), forming particles of typical sizes of 1–10 μm radius. This implies that PSC particles may interact with the chemistry of the middle atmosphere, not only via heterogeneous chemistry, but also by irreversibly removing reactive nitrogen and water vapor via gravitational sedimentation, which may prolong ozone depletion (WMO, 1999). Denitrification and dehydration are the terms usually employed to refer to these kinds of irreversible loss of nitrogen and water. De-NO_x-ification, on the other hand, is the NO_x loss process associated with heterogeneous reactions that produce a net conversion of NO_x temporary reservoirs (N_2O_5, $ClONO_2$, $BrONO_2$) into nitric acid.

Chlorine activation, denitrification and de-NO_x-ification result in a perturbed stratospheric composition that is more favorable to ozone destruction (National Aeronautics and Space Administration, 1992). In fact, because of low levels

Table 1 Heterogeneous reactions on PSC and sulfuric aerosols relevant to atmospheric chemistry

$N_2O_5 + H_2O$	$\longrightarrow 2HNO_3$
$BrONO_2 + H_2O$	$\longrightarrow HOBr + HNO_3$
$ClONO_2 + H_2O$	$\longrightarrow HOCl + HNO_3$
$ClONO_2 + HCl$	$\longrightarrow Cl_2 + HNO_3$
$HOCl + HCl$	$\longrightarrow Cl_2 + H_2O$
$HOBr + HCl$	$\longrightarrow BrCl + H_2O$
$N_2O_5 + HCl$	$\longrightarrow ClNO_2 + HNO_3$

of NO_x and chlorine nitrate, some chlorine activation will tend to persist even after PSC particles disappear in the vortex. Horizontal mixing of air from mid-latitudes is the competitive process limiting the time extent of these low NO_x conditions. Pure H_2O ice particles are extensively observed only in the Antarctic polar vortex as winter progresses and upper air layers cool down below the frost point of water vapor in the polar stratosphere (≈187 K). These are normally coarse particles that are efficiently removed by gravitational sedimentation, with the net effect of dehydrating the south polar lower stratosphere (Kelly *et al.*, 1989). The observed dehydration and denitrification of the Antarctic winter stratosphere also have the effect of reducing the frequency of cloud formation at a specific temperature threshold.

Some properties of PSC particles have been defined by a decade of satellite observations (SAM II) from 1979 to 1989 (McCormick *et al.*, 1982; McCormick and Trepte, 1986; McCormick and Trepte, 1987). PSC frequency of observation peaks in August over the Antarctic region (60% PSC detection in the 18–20 km layer) and in the second half of January–first half of February over the Arctic region (10% PSC detection in the 18–22 km layer) (WMO, 1992). The observed optical depth at 1 μm wavelength is about 0.01 during the austral winter months over Antarctica (excluding the years perturbed by volcanic eruptions). Another class of upper atmospheric polar H_2O cloud particles is normally referred to as noctilucent clouds, but while PSC aerosols are found in the polar lower stratosphere, noctilucent cloud particles reside in the mesosphere at more than 80 km above the ground.

Future increases of stratospheric water vapor and nitric acid, coupled to extensive stratospheric cooling caused by increasing burdens of carbon dioxide and other greenhouse gases (methane and N_2O in particular), might enhance PSC formation over wider regions of the polar stratosphere and for longer periods of time. These atmospheric composition changes may then slow down the recovery rate of stratospheric ozone in the next few decades, as expected following the Montreal protocol compliance for chlorofluorocarbons and halons (WMO, 1999). Future trends of water vapor may also be a consequence of climate changes: in particular a warming/cooling of the tropical tropopause may be a key climate factor determining the distribution of stratospheric H_2O. In fact, the most important mechanisms for producing H_2O in the stratosphere are methane oxidation and upward transport in the lower stratospheric tropical pipe (i.e., the upper extension of the Hadley circulation). Future climate changes may affect the temperature of the tropical tropopause and then the saturation water vapor pressure in this critical region, from where the whole stratosphere is dynamically fed.

High-flying supersonic aircraft might be another potential cause of a future increase of frequency of PSC occurrence. Only a few studies, applying detailed PSC microphysical schemes, have been made to investigate the effects of stratospheric aircraft emissions on PSC properties (Peter *et al.*, 1991; Pitari *et al.*, 1993). A thorough review of the effects on PSC surface area densities, applying a selection of state-of-the-art atmospheric models with varying degrees of microphysical complexity, has been provided by IPCC (1999). It is shown by the International Panel on Climate Change (IPCC) that *in situ* aircraft emissions of NO_x and H_2O in the lower stratosphere may create more favorable conditions for nitric acid and water vapor condensation into PSC particles.

Common to most modeling studies (Pitari *et al.*, 1993, as an example) is the assumption of the existence of two types of solid phase PSC particles, forming at certain threshold temperatures (e.g., the nitric acid trihydrate (NAT) condensation temperature and the ice frost point). Subsequent research has revealed the existence of at least three PSC particle types, showing complicated transition processes between liquid and solid type particles, that depend on the temperature histories of individual particles (Larsen, 2000).

The most important physico-chemical processes involved in our current understanding of PSC formation and growth are the evolution of liquid PSC and the mechanisms of liquid-to-solid and solid-to-liquid transitions. The background stratospheric aerosol particles are mostly supercooled sulfuric acid/H_2O liquid droplets. Their composition is strongly dependent on temperature: as temperature falls, they absorb water vapor. Below 200 K the uptake of HNO_3 becomes important, and these droplets change in composition and grow in volume (Carslaw *et al.*, 1994). If local temperature falls below the ice frost point, these particles freeze, forming a mixture of ice and solid hydrates of the acids.

Several mechanisms have been suggested for the formation of solid phase particles at temperatures above the ice frost point. Koop *et al.* (1995) have argued that the $HNO_3/H_2SO_4/H_2O$ particles do not freeze above the ice freezing temperatures, if the HNO_3 and H_2SO_4 weight fractions are close to their equilibrium values. On the other hand, when the $HNO_3/H_2SO_4/H_2O$ droplets evolve away from the thermodynamic equilibrium (i.e., during fast cooling and heating events), the $HNO_3/H_2SO_4/H_2O$ particles could become similar to HNO_3/H_2O pure solution droplets: these can freeze a few degrees above the ice freezing temperatures (between 190 K and 193 K), forming solid NAT (Meilinger *et al.*, 1995). Cooling/heating events associated with orographically generated gravity waves are an important driving mechanism for the formation of these solid particles. Other theories are based on freezing nucleation mechanisms (Jensen *et al.*, 1991; Luo *et al.*, 1994).

REFERENCES

Carslaw, K S, Luo, B P, Clegg, S L, Peter, T, Brimblecombe, P, and Crutzen, P J (1994) Stratospheric Aerosol Growth and HNO_3 Gas Phase Depletion from Coupled HNO_3 and Water Uptake by Liquid Particles, *Geophys. Res. Lett.*, 21, 2479–2482.

IPCC (1999) *Special Report on Aviation and the Global Atmosphere*, eds J E Penner, D H Lister, D J Griggs, D J Dokken, and M McFarland, Cambridge University Press, Cambridge.

Jensen, E J, Toon, O B, and Hamill, P (1991) Homogeneous Freezing Nucleation of Stratospheric Solution Droplets, *Geophys. Res. Lett.*, 18, 1857–1860.

Kelly, K K *et al.* (1989) Dehydration in the Lower Antarctic Stratosphere During Late Winter and Early Spring, 1987, *J. Geophys. Res.*, 94, 11 317–11 357.

Koop, T, Biermann, U M, Raber, W, Luo, B P, Crutzen, P J, and Peter, T (1995) Do Stratospheric Aerosol Droplets Freeze above the Ice Frost Point? *Geophys. Res. Lett.*, 22, 917–920.

Larsen, N (2000) *Chemical Understanding of Ozone Loss - Heterogeneous Processes: Microphysical Understanding and Outstanding Issues*, Proceedings of the Fifth European Symposium on Polar Stratospheric Ozone Research, St Jean de Luz, France, September 1999, Air Pollution Research Report, European Commission.

Luo, B P, Peter, T, and Crutzen, P (1994) Freezing of Stratospheric Aerosol Droplets, *Geophys. Res. Lett.*, 21, 1447–1450.

McCormick, M P, Steele, H M, Hamill, P, Chou, W P, and Swissler, T J (1982) Polar Stratospheric Cloud Sightings by SAM II, *J. Atmos. Sci.*, 39, 1387–1397.

McCormick, M P and Trepte, C R (1986) SAM II Measurements of Antarctic PSCs and Aerosols, *Geophys. Res. Lett.*, 13, 1276–1279.

McCormick, M P and Trepte, C R (1987) Polar Stratospheric Optical Depth Observed Between 1978 and 1985, *J. Geophys. Res.*, 92, 4297–4306.

Meilinger, S K, Koop, T, Luo, B P, Huthwelker, T, Carslaw, K S, Krieger, U, Crutzen, P J, and Peter, T (1995) Size-dependent Stratospheric Droplet Composition in Lee Wave Temperature Fluctuations and their Potential Role in PSC Freezing, *Geophys. Res. Lett.*, 22, 3031–3034.

National Aeronautics and Space Administration (1992) The Atmospheric Effects of Stratospheric Aircraft: a First Program Report, eds M J Prather, H L Wesoky, R C Miake-Lye, A R Douglass, R P Turco, D J Wuebbles, M K W Ko, and A L Schmeltekopf, *NASA Ref. Publ.*, 1272, 64–91.

Peter, T, Bruhl, C, and Crutzen, P J (1991) Increase in the PSC-formation Probability Caused by High-flying Aircraft, *Geophys. Res. Lett.*, 18, 1465–1468.

Pitari, G, Rizi, V, Ricciardulli, L, and Visconti, G (1993) High-speed Civil Transport Impact: Role of Sulphate, Nitric Acid Trihydrate, and Ice Aerosols Studied with a Two-dimensional Model Including Aerosol Physics, *J. Geophys. Res.*, 98, 23 141–23 164.

WMO (1992) Scientific Assessment of Ozone Depletion: 1991, Global Ozone Research and Monitoring Project, *WMO Rep.*, 25.

WMO (1995) Scientific Assessment of Ozone Depletion: 1994, Global Ozone Research and Monitoring Project, *WMO Rep.*, 37.

WMO (1999) Scientific Assessment of Stratospheric Ozone: 1998, Global Ozone Research and Monitoring Project, *WMO Rep.*, 44.

Aerosols, Stratosphere

Giovanni Pitari

Università L'Aquila, L'Aquila, Italy

A variety of aerosol particles can be found in the middle atmosphere, with sulfate and polar stratospheric cloud (PSC) particles being the most important for the chemical budget of the stratosphere and for their potential impact on climate. Other aerosol types detectable in the stratosphere are meteoritic dust and carbonaceous particles. The latter can be injected in situ by aircraft engines and can also be transported into the lower stratosphere by deep convective storms. This vertical mixing mechanism moves upward black carbon soot and organic particles produced at the surface by fossil fuel and biomass burning, forest res and terpene oxidation.

The main process responsible for stratospheric aerosol production (other than transport from the troposphere or from space) is gas condensation on pre-existing nuclei or on molecular clusters formed through homogeneous nucleation mechanisms. Sulphuric acid condensation may occur everywhere in the middle atmosphere below about 35 km, producing the observed size distribution of sulfate aerosols. Nitric acid and water vapour condensation takes place only at very cold temperatures in the wintertime polar vortices, forming PSC particles. In situ injection from aircraft plumes is another important production mechanism for sulfate aerosols: future supersonic high- ying aircraft will greatly increase the particle number and surface area density (SAD) in the mid-stratosphere (IPCC, 1999).

Stratospheric sulfate aerosols (SSA) may be formed *in situ* following ultraviolet (UV) photolysis of carbonyl sulfide (OCS), forming sulfur dioxide (SO_2), and also through oxidation of SO_2 penetrating the tropical tropopause. SO_2 is oxidized in a three-body reaction with OH forming sulphuric acid (H_2SO_4) as the final product of a reaction chain. Sulphuric acid and water vapour condensation takes place on pre-existing nuclei: these are Aitken particles coming from the tropical tropopause (mainly carbonaceous) that are first coated by H_2O/H_2SO_4, and also ultrafine sulfate particles formed by homogeneous nucleation immediately below the tropical tropopause (Weisenstein *et al.*, 1997).

The particle size distribution at a given location is determined by simultaneous microphysical processes (heterogeneous and homogeneous nucleation, condensation growth, evaporation, coagulation) and by large-scale transport and gravitational sedimentation.

Background sulfate aerosols are those photochemically produced in steady-state conditions of stratospheric sulfur abundance. These background sulfate particles are often perturbed by new aerosol formation starting from SO_2 or H_2S injected *in situ* by explosive volcanic eruptions. In the case of large eruptions, the aerosol optical thickness may increase by orders of magnitude.

Removal of stratospheric aerosols is produced by large-scale downward transport at mid–high latitudes, which moves the particles down into the troposphere where they are washed out or deposited at the surface.

Gravitational sedimentation is another important removal mechanism, although the lower stratospheric settling deposition is faster than advective downward motion only for particles with radius larger than 1 μm. This process, however, is important in determining the aerosol vertical profile in the mid-stratosphere.

Sulfate and PSC aerosols (*see* Aerosols, Polar Stratospheric Cloud, Volume 1) are of primary importance for the chemical budget of the middle atmosphere: heterogeneous reactions taking place on the surface of these particles represent an important sink of reactive nitrogen at mid–high latitudes. In turn, a decrease of odd nitrogen makes chlorine and bromine catalytic cycles for ozone destruction faster with respect to a pure gas phase chemistry atmosphere. Chlorine activation and de-NO_x-ification result in a perturbed stratospheric composition that is more favourable to ozone destruction (NASA, 1992). Sporadic sulfur injection in the stratosphere produced by large volcanic eruptions may produce catastrophic ozone losses (Prather, 1992).

Stratospheric aerosols have been sampled using balloons and high-altitude aircraft (Hofmann and Rosen, 1981; Rosen *et al.*, 1975) and from ground-based lidar measurements (pioneering observations were those by Fiocco and Grams, 1964). The most extensive coverage of stratospheric aerosols has been made by the SAGE satellite: these satellite data have provided information on the three-dimensional global-scale distribution of several aerosol properties, as extinction, optical thickness, mass density and SAD, acidity, and size distribution (McCormick *et al.*, 1979; Kent and McCormick, 1984; Yue *et al.*, 1986; Yue *et al.*, 1994). The range of typical non-volcanic values is summarized in Table 1.

Table 1 Summary of background stratospheric aerosol properties

Extinction at 1.0 μm (10^{-5} km^{-1})	1–100
Mass density (μgm^{-3})	0.01–0.5
SAD (μm^2 cm^{-3})	0.1–2
H_2SO_4 weight fraction	0.4–0.9
Number density of optically active particles ($r > 0.15$ μm) (part mg^{-1} air^{-1})	1–10
Stratospheric optical depth at 1.0 μm	5×10^{-4}–2×10^{-3}
Log-normal mode radius (μm)	About 0.07
Log-normal dispersion (μm)	About 1.8
Mass burden (Tg-S)	About 0.14

Stratospheric sulfate aerosols (SSA) play a key role in the chemical budget of the stratosphere. Several chemical reactions may efficiently take place on the surface of these particles (JPL, 1997): these heterogeneous reactions include hydrolysis of N_2O_5, $BrONO_2$, $ClONO_2$ and others involving hydrochloric acid (reacting with $ClONO_2$, HOCl, HOBr; see Table 2) (*see* Stratosphere, Chemistry, Volume 1). Reaction probabilities are normally highly dependent on particle acidity, with the important exception of $N_2O_5 + H_2O$: in this case the sticking coefficient is rather insensitive to temperature and humidity conditions. Many laboratory studies have been made on the efficiency of heterogeneous reactions involving Cl/Br species, due to their potential importance in polar ozone depletion (Hanson and Ravishankara, 1991; Hanson and Ravishankara, 1995; Hanson *et al.*, 1994; Hanson *et al.*, 1996).

The net effect of these heterogeneous reactions on SSA surfaces is to lower the NO_x to NO_y ratio (Fahey *et al.*, 1993), where NO_x is the sum of reactive nitrogen and temporary reservoirs ($NO + NO_2 + NO_3 + HNO_4 + N_2O_5 + ClONO_2 + BrONO_2$), while NO_y is the sum of NO_x and nitric acid (HNO_3). Heterogeneous reactions on SSA are in fact the most important loss mechanisms for NO_x in the lower stratospheric mid–high latitudes, along with the three-body reaction of NO_2 with OH forming nitric acid. This NO_x removal ends up in a ClO, BrO increase due to the decreased amount of chlorine and bromine nitrates, thus making Cl/Br catalytic cycles for ozone destruction dominant at about 50 hPa (Weisenstein *et al.*, 1991).

The 1991/92 mid-latitude observations collected by Fahey *et al.* (1993) clearly show the effect of the additional Pinatubo aerosols on the stratospheric chemistry: NO_x/NO_y at about 20 km and 45 °N falls from about 0.12 with a SAD of 0.5 μm^2 cm^{-3} (typical background value) to 0.09 with

Table 2 Heterogeneous reactions on SSA relevant to atmospheric chemistry

$N_2O_5 + H_2O$	$\longrightarrow 2HNO_3$
$BrONO_2 + H_2O$	$\longrightarrow HOBr + HNO_3$
$ClONO_2 + H_2O$	$\longrightarrow HOCl + HNO_3$
$ClONO_2 + HCl$	$\longrightarrow Cl_2 + HNO_3$
$HOCl + HCl$	$\longrightarrow Cl_2 + H_2O$
$HOBr + HCl$	$\longrightarrow BrCl + H_2O$

SAD = 4 μm^2 cm^{-3} and to 0.07 with SAD = 20 μm^2 cm^{-3} (during March 1992, with substantial Mt. Pinatubo aerosol penetration into the Northern Hemisphere, *see* Volcanic Eruption, Mt. Pinatubo, Volume 1).

The corresponding reactive chlorine fraction increases in the total chlorine budget (Cl_y): ClO/Cl_y ranges between 0.02–0.03 with SAD < 4 μm^2 cm^{-3} and between 0.04–0.06 with SAD = 20 μm^2 cm^{-3}. On the other hand, model calculations assuming pure gas phase chemistry give a much higher NO_x/NO_y ratio (0.19). The role of aerosol variations in anthropogenic ozone depletion at northern mid-latitudes in the last two decades has been shown by Solomon *et al.* (1996). A time-dependent model exercise carried out for the ozone assessment of the World Meteorological Organization (WMO) has shown that the photochemical perturbation produced by the Pinatubo aerosols during 1992/93 can explain to a large extent the total ozone depletion observed by the total ozone mapping spectrometer (TOMS) (WMO, 1999). Additional radiation scattering and absorption by volcanic particles may also perturb the ozone distribution through photolysis rates and circulation changes in the stratosphere (Pitari and Rizi, 1993; Kinne *et al.*, 1992).

Future trends of anthropogenic SO_2 emissions may significantly perturb mass and surface area densities of SSA. This is because not only will the total emission change, according to IPCC (2000) future scenarios, but also because large regional changes will occur with decreasing emission in western industrialized countries and large increases over the Middle East, China, India, Africa, and Central and South America. Enhanced SO_2 release in the tropics may significantly perturb the net amount of sulfur dioxide reaching the tropical tropopause and then entering the stratosphere.

Assuming as a reference case the IPCC (2000) A2 scenario, SO_2 emissions should peak around the year 2030, with a 60% increase with respect to present time; an increase of lower stratospheric SSA SAD associated to this larger tropospheric influx could affect the rate of recovery of stratospheric ozone. A faster heterogeneous chemical production of reactive Cl/Br may also be expected as a consequence of future supersonic aircraft emissions of ultrafine sulfate particles (IPCC, 1999).

REFERENCES

Fahey, D W, Kawa, S R, Woodbridge, E L, Tin, P, Wilson, J C, Jonsson, H H, Dye, J E, Baumgardner, D, Borrmann, S, Toohey, D W, Avallone, L M, Proffitt, M H, Margitan, J, Loewenstein, M, Podolske, J R, Salawitch, R J, Wofsy, S C, Ko, M K W, Anderson, D E, Schoeberl, M R, and Chan, K R (1993) In Situ Measurements Constraining the Role of Sulfate Aerosols in Mid-latitude Ozone Depletion, *Nature*, 363, 509–514.

Fiocco, G and Grams, G (1964) Observation of Aerosol Layer of 20 km by Optical Radar, *J. Atmos. Sci.*, 21, 323–324.

Hanson, D R and Ravishankara, A R (1991) The Reaction Probabilities of $ClONO_2$ and N_2O_5 on 40 to 75% Sulfuric Acid Solutions, *J. Geophys. Res.*, 96, 17 307–17 314.

Hanson, D R, Ravishankara, A R, and Solomon, S (1994) Heterogeneous Reactions in Sulfuric Acid Aerosols: a Framework for Model Calculations, *J. Geophys. Res.*, 99, 3615–3629.

Hanson, D R and Ravishankara, A R (1995) Heterogeneous Chemistry of Bromine Species in Sulfuric Acid Under Stratospheric Conditions, *Geophys. Res. Lett.*, 22, 385–388.

Hanson, D R, Ravishankara, A R, and Lovejoy, E R (1996) Reactions of $BrONO_2$ with H_2O on Submicron Sulfuric Acid Aerosol and the Implications for the Lower Stratosphere, *J. Geophys. Res.*, 101, 9063–9069.

Hofmann, D J and Rosen, J M (1981) On the Background Stratospheric Aerosol Layer, *J. Atmos. Sci.*, 21, 323–324.

IPCC (1999) *Special Report on Aviation and the Global Atmosphere*, eds J E Penner, D Lister, D J Griggs, D J Dokken, and M McFarland, Cambridge University Press, Cambridge.

IPCC (2000) *Special Report on Emission Scenarios*, eds N Nakicenovic and R Swart, Cambridge University Press, Cambridge.

JPL (1997) Chemical Kinetics and Photochemical Data for Use in Stratospheric Modeling, JPL Publication 97-4, JPL, Pasadena, CA.

Kent, G S and McCormick, M P (1984) SAGE and SAM II Measurements of Global Stratospheric Aerosol Optical Depth and Mass Loading, *J. Geophys. Res.*, 89, 5303–5314.

Kinne, S, Toon, O B, and Prather, M J (1992) Buffering of Stratospheric Circulation by Changing Amounts of Tropical Ozone: a Pinatubo Case Study, *Geophys. Res. Lett.*, 19, 1927–1930.

McCormick, M P, Hamill, P, Pepin, T J, Chou, W P, Swissler, T J, and McMaster, L R (1979) Satellite Studies of the stratospheric aerosol, *Bull. Am. Meteorol. Soc.*, 60, 1038–1046.

NASA (1992) The Atmospheric Effects of Stratospheric Aircraft: a First Program Report, eds M J Prather, H L Wesoky, R C Miake-Leye, A R Douglass, R P Turco, D J Wuebbles, M K Ko, and A L Schmeltekopf, NASA Ref. Publ. 1272, 64–91.

Pitari, G and Rizi, V (1993) An Estimate of the Chemical and Radiative Perturbation of Stratospheric Ozone following the Eruption of Mt. Pinatubo, *J. Atmos. Sci.*, 50, 3260–3276.

Prather, M J (1992) Catastrophic Loss of Stratospheric Ozone in Dense Volcanic Clouds, *J. Geophys. Res.*, 97, 10 187–10 191.

Rosen, J M, Hofmann, D J, and Laby, J (1975) Stratospheric Aerosol Measurements, II, The Worldwide Distribution, *J. Atmos. Sci.*, 32, 1457–1462.

Solomon, S, Portmann, R W, Garcia, R R, Thomason, L W, Poole, L R, and McCormick, M P (1996) The Role of Aerosol Variations in Anthropogenic Ozone Depletion at Northern Midlatitudes, *J. Geophys. Res.*, 101, 6713–6727.

Weisenstein, D K, Ko, M K W, Rodriguez, J M, and Sze, N D (1991) Impact of Heterogeneous Chemistry on Model-calculated Ozone Change Due to High Speed Civil Transport Aircraft, *Geophys. Res. Lett.*, 18, 1991–1994.

Weisenstein, D K, Yue, G K, Ko, M K W, Sze, N D, Rodriguez, J M, and Scott, C J (1997) A Two-dimensional Model of Sulfur Species and Aerosols, *J. Geophys. Res.*, 102, 13 019–13 035.

WMO (1999) *Scienti c Assessment of Stratospheric Ozone: 1998, Global ozone research and monitoring project*, WMO rep # 44.

Yue, G K, McCormick, M P, and Chou, W P (1986) Retrieval of Composition and Size Distribution of Stratospheric Aerosols with the SAGE II Satellite Experiment, *J. Atmos. Oceanic Technol.*, 3, 372–380.

Yue, G K, Poole, L R, Wang, P H, and Chiou, E W (1994) Stratospheric Aerosol Acidity, Density, and Refractive Index Deduced from SAGE II and NMC Temperature Data, *J. Geophys. Res.*, 99, 3727–3738.

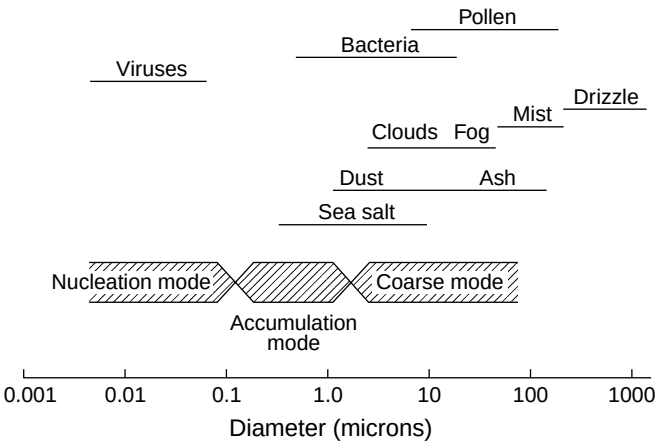


Figure 1 Sizes of some typical aerosols

Aerosols, Troposphere

Peter Brimblecombe
University of East Anglia, Norwich, UK

Aerosols are popularly seen as the products of spray cans. This misconception is understandable given the scienti c dif culties in de ning a term that was coined at the end of World War I to describe arsenical smokes. Aerosol has come to mean a suspension of solids or liquid droplets in air and can be regarded as an analogue of a hydrosol, the suspension of solid particles in water. The broadest de nition of aerosol embraces earlier terms such as: dusts, smokes, fogs, mists, fumes. The troposphere is lled with such aerosols, exhibiting a wide range of sizes due to their various sources and having a range of consequences. The study of aerosols has coped with treating both the individual particles and the bulk properties of the clouds they form.

It is size that most broadly characterizes aerosols and is critically important for their properties. There is an upper limit on the size that can remain suspended for any period of time. Occasionally very large particles are found in the atmosphere, perhaps 100 microns (μm) across. These are short lived, but include droplets, mists and volcanic particles. In general aerosol particles that are most often collected are about 20 microns or less in diameter. Figure 1 shows typical sizes of aerosols from a range of sources. The size of particles often reflects the processes responsible for their production. Those less than 0.1 microns are freshly formed by condensation or produced from reactions in the gas phase. This size range is known as the *nucleation mode*. Such small particles soon grow by coagulation or further condensation and lead to particles belonging to the *accumulation mode*. The largest particles (*coarse mode*) are usually formed by disaggregative processes, such as grinding or shattering after impacts at the surface of the Earth.

The very small size of aerosol particles has often made individual particle studies very difficult. Their properties may be studied optically, but bulk collection is perhaps the most common approach. Typically aerosols are collected onto filters of fibrous cellulose or quartz, or membranes of polycarbonate or Teflon. Membranes have the advantage of retaining particles on the surface, so they can later be examined by microscopy. The filters thus collected can later be weighed or analyzed. Collectors are often designed to admit particles in specific size ranges relevant to the planned study.

Size can be difficult to define because not all aerosols have regular shapes. This means that descriptions of aerosols have to define exactly what is meant by the size or diameter. Sometimes particles are assumed to be spheres with their volume treated as if it were spread within a sphere of a given diameter (*equivalent diameter*). When particles are observed under optical microscopy, the maximum edge to edge distance can be gauged (*Feret's diameter*). The rate at which fine particles fall in air (*settling velocity*) is dependent on their size. It provides another method of measuring particle size, so some studies express size as an *aerodynamic diameter*. This is defined as the diameter of a sphere of unit density (i.e., $1\,\mathrm{g\,cm^{-3}}$) that attains the same terminal velocity as the particle in question.

The aerosols found in the atmosphere have sizes that are often close to the wavelength of light. This enables them to interact with light and cause a wide range of optical phenomena. The reddish glow in the sky at sunrise and sunset is enhanced when large fires or volcanic eruptions have forced large quantities of particulate material into the atmosphere. This glow is produced as small particles scatter blue light and allow red to penetrate. It was first treated theoretically by Rayleigh late last century, but it has become apparent that the scattering process he described was restricted to

particles smaller than 30 nm (0.03 microns). In the present century a more complete description has been developed to describe what is called *Mie scattering*. It accounts for scattering taking place with particles close to the wavelength of light that could not be handled by the Rayleigh theory. These theories describe the optical properties of spherical particles, typically droplets, but are more difficult to apply to ice crystals, for example. More complex optical phenomena arise from particles of particular shape. Perhaps the best known of those originating from non-spherical particles, including phenomena such as the sun pillars (vertical shafts of light extending upward or downward from the sun).

SOURCES OF AEROSOLS

Although size reflects the process of aerosol production, composition is potentially a more informative indicator. *Primary particles* are those generated directly by their sources, as is seen clearly with wind blown dusts. The composition of dust particles is very similar to that of the soil from whence they came, so such particles include much silicon, calcium, aluminum, iron, etc. (*see* Dust, Volume 1). Although dust composition changes slightly over time, as particles age, much of the material is relatively insoluble and resistant to weathering.

The oceans are also a source of aerosols. Breaking waves produce sea spray, which can dry and yield salt particles in the atmosphere. However, salt particles from spray tend to be fairly large and short-lived. The minute bubbles that burst at the ocean surface, however, are more significant sources of sea salt particles. As this happens, the film from the cap of the bubble shatters, giving minute droplets, 6–30 microns in diameter that dry to form salt particles 1.5–8 microns across. Additionally, as the cavity left by the void that replaces the bubble collapses, a small jet pushes further droplets into the atmosphere; these are larger and produce salt particles of some 25 microns across. Although these droplets are mostly sea salt (sodium chloride with traces of other salts), there is the potential for other substances to be incorporated as the bubble forms. The surface of the sea is sometimes covered with a thin organic film that can interact with metals present in seawater and enhance their concentration in bubble caps. As bubbles burst, these metals are incorporated into the droplets released to the atmosphere, so we can find some elements enriched in the marine aerosol.

Volcanic particles contain rock fragments, but are often associated with sulfuric, hydrofluoric and hydrochloric acids from the hot vented gases. The high temperature production process gives the potential for the relocation of moderately volatile materials onto finer particles. These include metallic compounds or oxides, such as boron oxide, which can be found as boric acid in aerosols.

Biological activities are also an important source of particles, relevant to human health, as hay fever suffers well know. Bio-aerosols can be associated with the transmission of infectious diseases. The dispersal of pollen and spores are necessary for the propagation of plants. There are both viable and non-viable particles of primary origin. Non-viable particles include plant fragments, insect parts, feces, dander, etc. The seasonal distribution of biological particles is quite distinctive. In the spring, there is an important dominance of pollen grains that lie in the 10–30 micron range. In the autumn and winter, smaller biological particles dominate as bacterial aerosols arise from the decay processes.

Many of the particles discussed above are *primary particles* produced directly by processes such as combustion or weathering at the Earth s surface and carried aloft. Although this seems a very obvious source, it is not necessarily the most important. Much of the particulate material in the atmosphere, especially in the smaller size ranges, is produced by chemical reactions in the atmosphere. Such particles are called *secondary aerosols*. Typically atmospheric aerosols contain a significant proportion of sulfate (Figure 2). Much of this comes from the oxidation of atmospheric sulfur dioxide or sulfides that produce sulfuric acid. Sulfuric acid has an extremely strong affinity for water and, once formed, rapidly becomes incorporated in minute droplets. This acid can often enhance the uptake of alkaline ammonia from the atmosphere, such that fine particles are frequently found to be ammonium sulfate or bisulfate. Nitric acid is also produced in the atmosphere and can lead to the presence of an ammonium nitrate fraction within atmospheric aerosols. Thus, we have a situation where the soluble component of fine particles might be imagined as a $H^+/NH_4^+/SO_4^{2-}/NO_3^-$ system. Over the oceans, coarse

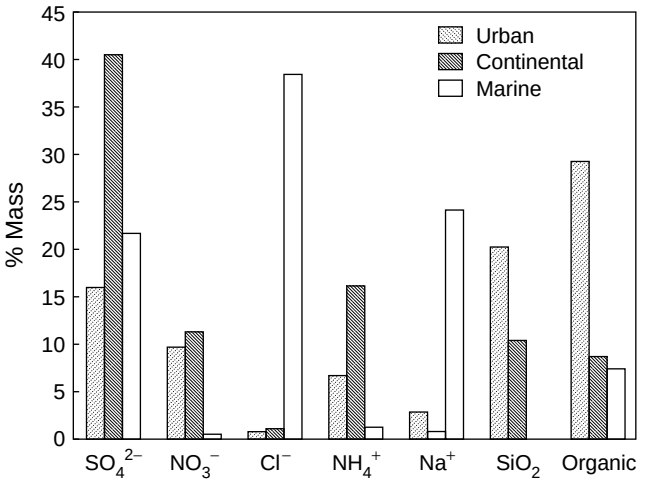


Figure 2 Percentage mass of various fractions in a typical urban, marine and continental aerosol (data from Colbeck, 1998)

sea salt particles of mostly Na^+/Cl^- are found together with this finer fraction. By contrast the finer continental aerosol, far from the oceans, is mostly produced from reactions in the atmosphere and will often be present with a coarser primary fraction composed of soil-derived dusts.

In addition to these inorganic materials, there is an organic component, that has not been as well studied. This typically includes acids or aldehydes, because they are so readily produced through the oxidation of volatile organic compounds (VOCs) in the atmosphere. Simple acids such as formic acid and acetic acid are abundant; they can only be removed (by conversion to droplets) if there is a significant amount of liquid water, as in clouds. Less volatile acids, and notably the dicarboxylic acid, oxalic acid, have a low volatility and high solubility and so become associated with aerosols in most conditions. Volatile hydrocarbons from plants include a wide range of terpenes. A common terpene, such as alpha-pinene, is released in large quantities by forests. It is readily oxidized in the atmosphere to a range of compounds, which include pinonic and pinic acids (derivatives of cyclobutane). These have been identified as components of forest aerosols.

Agriculture is one of the most natural human activities. However, it stirs up dust, and the pesticides that are in use also become part of the atmospheric aerosol loading. Fires that are so often a part of our agricultural endeavor are a further important source of atmospheric aerosols. Mention should also be made of widespread forest fires that are also a result of human activities, most notably in tropical South America (1970s and 1980s) and Southeast Asia (1990s). There has been much discussion that the impacts these fires, often set to clear land of vegetation, have on human health and regional air chemistry, including raising ozone concentrations. Biomass burning yields many millions of tonnes of soot every year. The soot has a graphitic structure, but there are some oxidized organic and hydrocarbon entities associated with it. Potassium and zinc are also likely to be found in the particles from forest fires.

Aerosols are usually present at high concentrations in cities. Combustion is one of the most important sources of particles, and soot or smoke has long been the most obvious urban aerosol. The larger smoke particles have become less common in the urban atmosphere as coal has been abandoned as the main domestic and industrial fuel. However, the rise of diesel powered vehicles represents a particularly important source of carbon particles. These particles are often extremely small and typically lie in the 10–80 nm range, but frequently cluster together to form larger conglomerates. Even so, diesel-derived particles are less than 2.5 microns in diameter. Furthermore, carbonaceous diesel particles are frequently associated with low volatility organic compounds such as the polycyclic aromatic hydrocarbons (PAHs). These compounds are of concern because they are often carcinogenic. Both motor vehicles and stationary combustion sources seem responsible for various PAHs, of which benzo(*a*)pyrene is widely studied, partly because it is a well known carcinogen. Recently it has become evident that PAHs can react in the atmosphere to form nitro- or chloro-PAHs. These compounds may well have enhanced carcinogenicity compared with their parent PAH. While the threat to human health is very real, there is evidence that the changes in combustion processes in urban areas over the last 50 years have brought about a decline in the concentration of major PAH fractions such as benzo(*a*)pyrene. Nevertheless, the poorly understood organic fraction remains a major component of respirable aerosol particles in urban areas.

Combustion is also responsible for the release of ashes. In the past, fly ash, especially from coal combustion, was rich in calcium and iron. Such alkaline aerosols may have neutralized some of the sulfuric acid that came from coal use and contributed to maintaining an acid balance when combustion products returned to the surface. Today, in many cities, ash particles are likely to derive from oil. Both vanadium and nickel are enriched in petroleum oils, so they are also present at high relative abundance in oil fly ash and their presence can be used as tracers of oil combustion.

Dust in cities is generated by localized activities. In arid regions, wind blown dusts from surrounding areas can be a further source that adds to anthropogenic particles in urban air.

The polluted atmosphere is very reactive and produces many secondary aerosols. The oxidation of sulfur dioxide and the nitrogen oxides to sulfuric acid and nitric acids provides important sources of these compounds in urban particles. Hydrogen chloride is released from industrial sources and incineration. These acids become associated as sulfates, nitrates and chlorides on the fine aerosols. As in remote continental air, they often become associated with ammonium salts.

Although much of the respirable particulate load in urban air is composed of organic matter, it is not clear what proportion of this is of primary origin. Oxidative reactions in the urban atmosphere produce relatively low volatility acids and aldehydes. The highly oxidized oxalic acid is often found associated with the fine particles of smog. This suggests that oxalic acid is an example of an organic aerosol component of secondary origin.

The particles in the atmosphere are removed in two ways. Over time, some slowly sediment and deposit on the ground, often filtered out by vegetation (*dry deposition*). A more important process involves their incorporation into cloud droplets and ultimate removal in rainwater (*wet deposition*). Particles typically remain in the lower troposphere (below 1.5 km) about one week before removal, although their residence times are dependent on weather conditions. Higher in the troposphere, residence times are as much as three weeks.

IMPACTS OF AEROSOLS

Particles present in air have a range of effects. In the past, the high concentrations of soot in cities were apparent from the black encrustations that disfigured buildings. Declining coal use within major cities gave some relief to this problem, but in Europe soiling once again appears to be on the increase. Now it is associated with diesel soot rather than solid fuel combustion.

Health impacts are one of the major problems likely to arise from the high concentrations of suspended particulate matter in cities. In recent years, regulatory attention has shifted to the finer fraction of the urban aerosol. Size differentiated measurements of respirable particles of 10 microns (PM_{10}) have been made for more than a decade. As more detailed studies have become available, there is evidence that the health impacts appear to be associated with yet finer fractions, so emerging regulations are likely to focus on $PM_{2.5}$ (particles of 2.5 microns diameter or less).

As we breathe in particles, some, especially the larger ones, can be trapped in the mucous that coats our airways. This is a useful protective mechanism as the mucous is driven upwards by the beating of fine hairs (*cilia*), so that the particles can ultimately be swallowed rather than reach the lung. Many finer particles are not effectively removed by the mucous and penetrate into terminal airways and alveoli. Here, macrophages engulf the particles and migrate to the ciliated parts of the respiratory system, effectively removing them from the lung. The activities of the macrophages, although important, release inflammatory compounds. The inflammatory effects may easily be transmitted to the blood. Other particles can cross the epithelium and become associated with inflammation of interstitial tissue. This pulmonary inflammation then leads to cardiovascular problems, so an enhanced death rate in particle laden air would seem an understandable outcome.

It has been difficult to evaluate the effect of urban aerosols on human health through controlled exposure of volunteers. This method is common when estimating the effects of pollutant gases, but more difficult for particles that are not so easily defined. Most evidence for the harmful effects of aerosols has come from epidemiological studies of exposed populations in cities, for example, by examining death rates and the frequency of hospital admissions. Studies in the US have indicated that a $10\,\mu g\,m^{-3}$ rise in PM_{10} concentration is likely to cause a 1% increase in daily mortality. The strongest association between exposure and death is related to the preceding 5 days exposure, which suggests a relatively rapid effect on mortality. Such fatalities can be viewed as deaths brought forward, because the affected individuals may already have been unwell, so that a high air pollution episode may cause them to die sooner. In addition to this rapid effect on mortality, lengthy exposure to high particle concentrations in the urban atmosphere may well contribute to a slower development of the debilitating diseases. Furthermore, lung cancer is a potential threat from long-term exposure to the PAHs present on urban particles.

Tropospheric aerosols have effects not only at the surface but they can change the radiative budget of the Earth. In addition to the potential influence of aerosols from human activities, it has often been argued, through the popular Gaia hypothesis, that natural marine sulfate aerosols have helped in maintaining relatively constant global temperatures. These sulfate aerosols are derived from the oxidation of dimethylsufide released to the atmosphere by phytoplankton (*see* Dimethylsul de (DMS), Volume 1). It has been argued that this suggests a biological control on global climate. Human activities have further enhanced the presence of sulfate aerosols, particularly in the Northern Hemisphere, by releasing large amounts of sulfur dioxide during fossil fuel combustion. Sulfate aerosols increase the number of cloud condensation nuclei and make clouds more reflective, with the potential for cooling the troposphere. Although at their most obvious in cities, tropospheric aerosols can thus have global impacts (*see* Aerosols, Effects on the Climate, Volume 1).

FURTHER READING

Colbeck, I (1998) *Physical and Chemical Properties of Aerosols*, Blackie Academic and Professional, London.

Harrison, R M and van Grieken, R E (1998) *Atmospheric Particles*, John Wiley & Sons, Chichester.

Agassiz, Louis

(1807- 1873)

Louis Agassiz was born in Switzerland in 1807. He became the leading proponent of the ice age theory and the existence of large continental ice sheets covering much of northern Europe and North America. While a professor at Neuchatel from 1835 to 1845, Agassiz studied the geological formations of Switzerland, in comparison to those of England and central Europe. He related the scratched and polished rocks, erratic boulders, and the geomorphology of the landscape to the presence of a large continental ice sheet. Although

he was not the first to develop the glacial theory (e.g., Jean-Pierre Perraudin, Ignace Venetz, Jean de Charpentier), his Discourse of Neuchatel, delivered in 1837 to the Swiss Society of Natural Sciences, brought the theory to prominence. He later extended this theory to North America. Early on, he won the acceptance of such notable scientists as the Reverend William Buckland, Charles Lyell and Charles Darwin, although they accepted a more general theory that included a combination of glaciers, icebergs, terrestrial melt water, and ocean currents to explain the features of glaciation. As a result, Agassiz s theory first became more obscure before becoming more widely accepted in the 1870s (American Council of Learned Societies, 1970; Hansen, 1970; Imbrie and Imbrie, 1986).

Agassiz obtained his PhD in 1829 at the Universities of Munich and Erlangen, studying the fishes of Brazil, and his Doctor of Medicine at Munich in 1830. In 1832 he married Cecile Braun and became a professor at the College of Neuchatel, which is where he published his ice age-related work. In 1846, he lectured at Lowell Institute in Boston. The following year his wife died (they had three children), and he started a professorship at the Lawrence Scientific School of Harvard University, where he taught until his death in 1873. He remarried in 1850 to Elizabeth Cabot Cary.

Agassiz considered himself more of a naturalist than a specialist in any one field, and he conducted exhaustive studies of the characteristics of marine biology, freshwater fishes, embryology, and fossil fishes. He also developed a love for the American landscape, and contributed to the study and education of natural history in the United States. Agassiz s Protestant background led to a belief that signs of the Creator exist in all flora and fauna. As such he became a major opponent to Darwin s theory of evolution. Whereas both Lyell and Darwin accepted Pleistocene glaciation as a mechanism for genetically linking species separated geographically, Agassiz believed that there was no relationship between different species as the Creator endowed each with its own qualities. He did not hesitate to use the public forum to try to gain acceptance for his views, although unsuccessfully. Agassiz s scientific reputation generally diminished towards the end of his career, but he did make some important contributions to the Museum of Comparative Zoology and the National Academy of Sciences (American Council of Learned Societies, 1970).

Photo: Courtesy of the Library of Congress.

See also: Earth System History, Volume 1.

REFERENCES

American Council of Learned Societies (1970) *Dictionary of Scienti c Biography*, ed C C Gillispie, Charles Scribner, New York, 72–74.

Hansen, B (1970) The Early History of Glacial Theory in British Geology, *J. Glaciol.*, 9(55), 135–141.

Imbrie, J and Imbrie, K P (1986) *Ice Ages: Solving the Mystery*, Harvard University Press, Cambridge, MA, 1–224.

BENJAMIN S FELZER *USA*

Air Pressure

Air pressure is the force exerted by the air; differences in pressure produce forces that set air into motion, thus driving winds and storms.

Pressure is the force exerted on one body by another. For fluids like air and water, pressure is measured in terms of the force exerted per unit area. In the atmosphere, this force results from the weight of the air above the measuring point. For example, the mass of the column of air above each square centimeter of the Earth s surface is about 1 kg.

Air pressure may be expressed in several ways. In English engineering units, air pressure is often measured in pounds per square inch; in this unit, the average sea-level pressure is $14.7\,\mathrm{lb\,in^{-2}}$. However, in meteorology it is more common to use metric units such as the millibar, defined as 1000 dynes per square centimeter or the more modern unit of the same magnitude, the hectopascal. Yet another common measure of pressure is the height of a barometer s mercury column that can be supported by the ambient air pressure. Thus, normal ground level pressure is sometimes expressed as $14.7\,\mathrm{lb\,in^{-2}}$, 1013 millibars, 1013 hectopascals, or 29.92 inches or 760 millimeters of mercury. In practice, pressures are often reported as sea level pressures, i.e., the pressure that would be observed at the bottom of a vertical shaft extending downwards to sea level.

Air pressure decreases rapidly with height; at an altitude of about 6 km, pressure is about half that at the surface. The pressure of air on our bodies is balanced by the pressure within, and thus does not affect us. However, differences in pressure do exert forces on the air and can set it into motion. Differential heating and cooling over the face of the globe produce regions of higher and lower pressure. As these force the air into motion, the rotation of the Earth causes the winds to adjust their direction (in the Northern Hemisphere) so that lower pressure is found to the left of the wind and higher pressure to the right. In the Southern Hemisphere, this pattern is reversed. Thus, the familiar centers of high and low pressure found on weather maps result from a complex set of processes involving heating and cooling, horizontal and vertical air currents, and condensation and precipitation processes on a rotating planet.

JOHN S PERRY *USA*

Air Quality, Arctic

see Arctic Air Quality (**Volume 1**)

Air Quality, Global

Douglas Whelpdale and Elizabeth Bush
Meteorological Service of Canada, Downsview, Canada

Excluding water vapor, 99.9% of the Earth's atmosphere is made up of nitrogen (N), oxygen, and the noble gases. Their concentrations have been relatively stable for the past billion years. The remaining 0.1% is composed of many different gases and particles that are present in only trace amounts. Air quality is a consequence of the relative concentrations of these trace species; poor air quality results when concentrations of trace species that harm either the Earth's environment or inhabitants are enhanced. The term global air quality (or synonymously, global air pollution, background air pollution, background air quality), although imprecisely de ned, is used here as a subjective or comparative descriptor of atmospheric quality resulting from the concentrations of the more important of these trace species at locations distant from their anthropogenic sources.

Trace gases (and small particles) in the atmosphere originate from both natural processes and human activities. Air pollution is sometimes thought of as material in the atmosphere, regardless of concentration, caused by human activity. This de nition is of limited use because atmospheric measurements are often unable to distinguish between anthropogenic and natural sources of the same species. Another de nition of air pollution relies on the derivation of threshold pollutant concentrations, above which adverse effects occur on selected receptors. This approach is falling out of favor - improved measurement techniques are lowering thresholds - and in any case, it is not particularly relevant to a discussion of global air quality. A third, and more practical, approach is to contrast air quality in areas known to be affected by human activity with air quality in locations distant from such pollution sources. One caution, however, is that the atmosphere distant from anthropogenic sources of pollution is not necessarily a spatially and chemically uniform mixture of pollutants.

The contribution made by a given trace species to the air quality at a given location depends on the species' atmospheric lifetime. The length of time that a substance remains after release into the atmosphere depends upon a number of factors that control its susceptibility to chemical destruction or physical removal: solubility, reactivity, density, shape, height of injection, location of release, etc. A substance with a lifetime of a few hours will not travel far beyond its source region. A substance with a lifetime of several weeks released in the windy mid-latitudes will be dispersed widely in the hemisphere. A substance with a lifetime of years will be transported globally and will exhibit a relatively uniform global distribution. Shorter-lived gases and particles exhibit heterogeneous distributions, with concentrations highest near sources and generally lowest in remote areas.

In this article, trace species that are commonly implicated in global air quality will be discussed. A number of potentially important species (the hydroxyl radical, DMS, trace metals, and many organic chemicals), have rather arbitrarily been excluded from this discussion.

MAJOR GLOBAL AIR MONITORING PROGRAMS

A number of global monitoring networks have been established to help characterize global air quality. Monitoring stations are located at remote sites around the globe, although Africa, South America and the southern oceans are still under-represented. The US National Oceanic and Atmospheric Administration/Climate Monitoring and Diagnostics Laboratory s (NOAA/CMDL) global cooperative air sampling network has approximately 46 sites, including four baseline observatories that have been operating since the late 1950s (Barrow, Alaska; Mauna Loa, Hawaii; American Samoa and South Pole) (Climate Monitoring and Diagnostics Laboratory, CMDL, 1998). The Advanced Global Atmospheric Gases Experiment (AGAGE) network has five remote observing stations established in 1978, and a measurement program focused on halocarbons and other species implicated in stratospheric ozone loss (Prinn *et al.*, 2000). The World Meteorological Organization s (WMO) Global Atmosphere Watch (GAW) program was established in 1989, integrating two previous programs, the Background Air Pollution Monitoring Network and the Global Ozone Observing System, established in the 1960s and late 1950s, respectively. The GAW network currently consists of about 20 global (remote) stations and about 300 regional stations. Some NOAA/CMDL and AGAGE stations are part of this network. Data from GAW stations are sent to six World Data Centers for archiving and analysis (Miller and Young, 1997) (*see* GAW (Global Atmosphere Watch), Volume 1).

In addition to these major systematic observing programs, many special short-term field campaigns have conducted extensive sampling programs. Supplemental observations are also provided by ship-, aircraft-, balloon-, and satellite-based measurement programs. Information on air pollutant concentrations prior to establishment of modern measurement programs is provided by proxy records, particularly

ice cores obtained from Greenland and Antarctica (Bradley, 1999).

CARBON DIOXIDE (CO_2)

Carbon dioxide is an important trace constituent because of the role it plays in radiative forcing in the climate system. Its concentration has been increasing dramatically since direct measurements began at the South Pole in 1957 and at Mauna Loa in 1958. Carbon dioxide concentrations are now routinely monitored at many locations globally. Data from before the late 1950s are available from ice core measurements, and extend back 400 000 years.

Carbon dioxide is a long-lived gas with an atmospheric lifetime of 50–200 years, and is therefore well-mixed globally. No single lifetime for CO_2 can be defined because of different rates of uptake and removal by different processes (Intergovernmental Panel on Climate Change, IPCC, 1996). Although CO_2 concentrations exhibit seasonal, latitudinal and hemispheric variation, annually averaged concentrations at individual remote stations are close to the global annual average. Concentrations at Mauna Loa have risen from approximately 315 parts per million by volume (ppmv) in 1958 to 368 ppmv in 1999. Over the years 1983–1997, the growth rate in CO_2 varied between 1–3 ppmv $year^{-1}$ (World Data Centre for Greenhouse Gases and Other Atmospheric Gases, WDCGG, 2000) (*see* Carbon Dioxide, Recent Atmospheric Trends, Volume 1). Pre-industrial (circa 1850) concentrations were approximately 280 ppmv, and ice core data reveal that concentrations were relatively stable around the 280 ppmv mark for the 1000 years prior to the industrial age (Intergovernmental Panel on Climate Change, IPCC, 1996) (*see* Intergovernmental Panel on Climate Change (IPCC): an Historical Review, Volume 4).

METHANE (CH_4)

Methane is the next most abundant atmospheric trace gas. It plays a key role in tropospheric chemistry, in the atmospheric greenhouse effect and in stratospheric ozone loss. Systematic direct measurement of CH_4 began in 1978 as part of the Atmospheric Lifetime Experiment (ALE), and has continued through the evolution of this program into AGAGE. The NOAA/CMDL program began making CH_4 measurements in the mid-1980s. Ice core data provide CH_4 concentrations as far back as 160 000 years ago (Chappellaz *et al.*, 1990).

Methane is a relatively long-lived gas with an atmospheric lifetime of approximately 12 years. This has been defined as an adjustment time that takes into account the indirect effect of CH_4 on its own lifetime (Intergovernmental Panel on Climate Change, IPCC, 1996). Methane exhibits a relatively even distribution globally in the upper atmosphere, albeit with latitudinal, hemispheric and seasonal variations. Near-surface concentrations are higher in the Northern Hemisphere as a result of proximity to source regions. The 1997 annual average atmospheric CH_4 concentration at the site with the longest running direct measurement program, Cape Grim, Tasmania, is 1690 parts per billion by volume (ppbv) (Carbon Dioxide Information Center, CDIAC, 2000). The 1998 global annual average atmospheric CH_4 concentration, calculated from sites contributing data to the World Data Centre for Greenhouse Gases, is 1749 ppbv. From the early 1980s to the late 1990s, CH_4 concentrations grew at a rate of 9 ppbv $year^{-1}$ (World Data Centre for Greenhouse Gases and Other Atmospheric Gases, WDCGG, 2000, 2). The atmospheric CH_4 concentration was relatively stable at approximately 700 ppbv from 1000–1800 AD, at which time concentrations began to increase (Etheridge *et al.*, 1998) (*see* Methane, Volume 1).

NITROUS OXIDE (N_2O)

Nitrous oxide is a trace gas that is implicated in both the greenhouse effect and in stratospheric ozone loss. The longest N_2O monitoring programs are those of the AGAGE network and the CMDL baseline observatories, which began in 1978 and 1977, respectively. Ice cores have also provided information on historic N_2O concentrations.

Nitrous oxide has an atmospheric lifetime of approximately 120 years (Intergovernmental Panel on Climate Change, IPCC, 1996, 15) and is a globally distributed pollutant. Global average N_2O concentrations have increased from approximately 275 ppbv in pre-industrial times (World Meteorological Organization, WMO, 1999, 1.17) to current (2000) concentrations of about 315 ppbv (Elkins, 2000). Annual average concentrations in the Northern Hemisphere are consistently about 1–2 ppbv higher than in the Southern Hemisphere. The trend in N_2O has been nearly linear over the past 13 years, averaging about 0.75 ppbv $year^{-1}$ (Elkins, 2000) (*see* Nitrous Oxide, Volume 1).

HALOCARBONS

Systematic measurements of halocarbons began in the late 1970s in the AGAGE and NOAA/CMDL programs, shortly after they were first detected in the troposphere. Measurements were launched in response to revelations that these compounds were implicated in stratospheric ozone depletion; it is now known that they are implicated in global climate change as well.

Halocarbons are carbon compounds containing a halogen, such as chlorine, fluorine and bromine. Included in this term are *chloro uorocarbons* (CFCs), hydrochlorofluorcarbons

(HCFCs), *hydro uorocarbons* (HFCs), methyl halides (e.g., methyl chloride, methyl bromide), chlorinated solvents (e.g., methyl chloroform, carbon tetrachloride) and halons (brominated carbons). Except for the HCFCs, methyl chloride, methyl bromide and a few others, these compounds are generally long-lived and eventually reach the stratosphere, where their destruction occurs. Most *halocarbons* are solely of anthropogenic origin, and thus did not occur in the atmosphere prior to the beginning of their manufacture.

Generally, concentrations of the major CFCs rose steeply from the time of their introduction into the environment until about 1990 at which time they stabilized, beginning a declining phase in the early 1990s that reflects reductions in emissions mandated through the international Montreal Protocol and its Amendments. The exception is CFC-12 whose concentration is still increasing, albeit at a reduced rate. Current concentrations of major halocarbons (e.g., CFC-11, 12 and methyl chloride) are approximately 0.5 ppbv. The concentrations of the other halocarbonss range from ~0.1 ppbv (e.g., carbon tetrachloride, methyl chloroform) to ~0.1 parts per trillion by volume (pptv) (bromocarbons) (World Meteorological Organization, WMO, 1999). The concentrations of CFC substitutes such as HCFCs and HFCs are rising rapidly but are still only present in the pptv range (Prinn *et al.*, 2000) (*see* Chloro uorocarbons (CFCs), Volume 1).

CARBON MONOXIDE (CO)

Carbon monoxide adversely affects human health and is also important in tropospheric chemistry; it is therefore of interest as a global pollutant. Carbon monoxide is monitored globally, the most extensive program being the NOAA/CMDL global sampling network (Novelli *et al.*, 1998). Ice core analyses have recently provided perspectives on historical CO concentrations over the last 2000 years (Haan and Raynaud, 1998).

Carbon monoxide has a relatively short atmospheric lifetime, ranging from weeks to months, although this can be extended to over a year in the winter polar atmosphere (Hollaway *et al.*, 2000). The concentration and distribution of CO in the atmosphere is variable, being dependent on the spatial pattern and variability of both sources and sinks, and on the effects of transport. As a consequence, CO exhibits a non-uniform distribution globally, with strong latitudinal and longitudinal gradients and strong seasonality (Novelli *et al.*, 1998). Based on data (1996) from clean, remote, continental and oceanic sites, the global average concentration is approximately 90 ppbv (World Data Centre for Greenhouse Gases and Other Atmospheric Gases, WDCGG, 2000). Annual average concentrations are about 120 ppb in the Northern Hemisphere and about 40 ppbv in the Southern Hemisphere (US Environmental Protection Agency, EPA, 2000). Ice core measurements confirm that there has been little change in CO concentrations in Antarctica over the last 2000 years (~50 ppbv) while CO measurements in Greenland ice cores suggest that there has been an increase in CO concentrations at high northern latitudes over the last century (Haan and Raynaud, 1998) (*see* Carbon Monoxide, Volume 1).

SULFUR DIOXIDE (SO_2) AND SULFATE (SO_4)

Trace amounts of several sulfur species exist as gases, particles and in precipitation in the atmosphere. The two compounds most relevant to air quality are SO_2 and SO_4; other important species include hydrogen sulfide, dimethyl sulfide, carbon disulfide and carbonyl sulfide. Both SO_2 and SO_4 are human health and environmental pollutants, causing cardio–respiratory problems and contributing to acid deposition. Sulfate particles are very small (~1 μm), hence they contribute also to visibility degradation, and interfere with the atmospheric radiation balance, both directly by reflecting incoming solar radiation and indirectly as a cloud nucleating agent.

The tropospheric lifetime of SO_2 is several hours to a few days, and that of SO_4 is several days. As a result, these pollutants are primarily regional rather than global pollutants. Sulfur concentrations are measured routinely in global networks such as the GAW, in regional networks in North America, Europe and Asia, and in many national networks. Typical SO_2 concentrations range from 20 pptv in the marine boundary layer to 200 pptv in clean continental North America to 1500 pptv in polluted continental air (Seinfeld and Pandis, 1998, 61). Non-sea-salt SO_4 concentrations in air range from less than 1 nmol m^{-3} (0.1 μg m^{-3}) in the remote marine environment to approximately 10–20 nmol m^{-3} (1–2 μg m^{-3}) in clear continental North America to >100 nmol m^{-3} (>10 μg m^{-3}) in and downwind of heavily industrialized regions (Whelpdale and Kaiser, 1996). While concentrations of SO_2 and SO_4 in remote regions have increased little since the onset of industrialization, concentrations in the most heavily industrialized regions have risen up to 50-fold.

NITROGEN OXIDES AND NITRATE

Similar to sulfur oxides, several nitrogen oxide compounds exist as gases, particles and in precipitation, and are implicated in human respiratory problems (both directly and through promoting the formation of ozone), acid deposition, and visibility degradation (*see* Nitrogen Cycle, Volume 2). The two main reactive gaseous species are nitric oxide (NO) and nitrogen dioxide (NO_2); they are often considered together as NO_x. Nitrate (NO_3) is found in both particles and precipitation. Among the other reactive

nitrogen species, nitric acid (HNO_3) and peroxyacetylnitrate (PAN) are also of interest from an air quality perspective. Sources of nitrogen oxides have high temporal and spatial variability, and this, combined with their relatively short atmospheric lifetimes of hours to a few days, results in concentrations which are also highly variable in space and time, with highest concentrations near source regions. Therefore, like the sulfur compounds, these compounds are primarily regional rather than global pollutants.

In a recent summary, Emmons *et al.* (1997) found that in clean regions median boundary layer NO concentrations are <10 pptv and NO_x <50 pptv. In highly industrialized North America and Europe NO concentrations can be >1 ppbv and NO_x >2 ppbv. Particulate nitrate concentrations range from $\sim 1\,\mathrm{nmol\,m^{-3}}$ ($0.06\,\mu\mathrm{g\,m^{-3}}$) in remote marine locations to $5\,\mathrm{nmol\,m^{-3}}$ ($0.3\,\mu\mathrm{g\,m^{-3}}$) in clean continental locations and typically up to $40\,\mathrm{nmol\,m^{-3}}$ ($2.5\,\mu\mathrm{g\,m^{-3}}$) in industrialized regions, but reaching as high as $100\,\mathrm{nmol\,m^{-3}}$ ($6\,\mu\mathrm{g\,m^{-3}}$) (Whelpdale and Kaiser, 1996). Concentrations of reactive N species have increased significantly since industrialization, as much as 50-fold in regions of heavy emissions. This trend is likely to continue as NO_x emissions continue to increase.

OZONE (O_3)

Ten percent of the Ozone in the Earth s atmosphere is found in the troposphere, the rest being stratospheric O_3. Ozone is not emitted directly, but formed in a series of complex reactions involving a number of precursor molecules that are of both natural and anthropogenic origin. In the troposphere, O_3 is considered a pollutant, both because of its adverse effects on human health and vegetation and because it is a greenhouse gas. The lifetime of O_3 is typically a few days to two months, and its global distribution is highly variable in time and space. The best O_3 data sets are from the last 20 years or so, during which time O_3 measurements were integrated into programs at sites in the GAW. The World Data Centre for Surface O_3 has only recently become operational (1998) but will soon provide an ongoing archive of surface O_3 data. National programs have also been established in several countries to monitor surface O_3.

Annual average hemispheric surface O_3 values for 1995 at remote sites were 41 ppbv in the Northern Hemisphere and 22 ppbv in the Southern Hemisphere (Oltmans *et al.*, 1998). Over the 20 years of record, most of the nine sites have experienced an increasing trend, although the magnitude varies considerably. At the South Pole, a notable decline in surface O_3 concentrations has occurred that is most likely related to the decline in stratospheric O_3 over the Antarctic (*see* Ozone Hole, Volume 1). European data for the late 1800s and early 1900s indicate that monthly average O_3 concentrations in the northern midlatitudes were $\sim$10 ppbv (Sandroni *et al.*, 1992). Modeled pre-industrial concentrations of surface O_3 for the Northern Hemisphere mid-latitudes are about twice this value (Lelieveld and Dentener, 2000). Therefore, there has been at least a doubling and possibly a several-fold increase in Northern Hemisphere mid-latitude O_3 concentrations over the last century. The Southern Hemisphere has also experienced an increase, but it is likely only on the order of 50% (Crutzen, 1995) (*see also* Stratosphere, Ozone Trends, Volume 1; Troposphere, Ozone Chemistry, Volume 1).

VOLATILE ORGANIC COMPOUNDS (VOCs)

The term VOCs is an umbrella term for those hydrocarbons (excluding CH_4) that are of low enough molecular weight to be volatile and therefore present in the atmosphere (i.e., non-CH_4 hydrocarbons) as well as all other reactive organics in the troposphere including alcohols, aldehydes and organic acids (Ehhalt, 1999). VOCs are primarily of interest for their role as precursors to tropospheric O_3. Most VOCs react locally, and therefore the concept of a global concentration for VOCs is not really applicable.

VOCs can be measured either individually or collectively. Individual VOCs (of which there are thousands) are typically present in the atmosphere in the pptv concentration range in remote and rural areas, and in the ppb range in urban locales (Finlayson-Pitts and Pitts, Jr, 2000, 591–593). For example, representative concentrations (pptv) of two major VOCs, benzene and toluene are, respectively, 0.008–0.2 and 0.01–0.25 in remote areas, 0.1–0.6 and 0.05–0.8 in rural environments, and 0.9–26 and 2–39 in urban areas (Finlayson-Pitts and Pitts, Jr, 2000, 591) (*see* Volatile Organic Compounds, Biogenic (VOCs), Volume 2).

AEROSOLS

The atmosphere contains not only trace gases but also suspensions of microscopic solid and liquid particles, termed aerosols (*see* Aerosols, Troposphere, Volume 1). Atmospheric aerosols are of both natural and anthropogenic origin, range in size from a few nm (10^{-9} m) to about a hundred μm (10^{-6} m), and can be of extremely complex chemical composition. Particles of natural origin, such as mineral and soil dust, pollen, spores, and sea-salt spray tend to be relatively larger (coarse-particle mode). Particles emitted from combustion sources or formed in the atmosphere through gas-particle conversion are smaller in size (fine-particle mode). The chemical composition of fine aerosols is complex and can include sulfates, nitrates, inorganic and organic carbons, and heavy metals. Atmospheric aerosols are of interest both because of their impact on

human health and their role in radiative forcing, where they are now believed to exert a cooling effect.

The lifetimes of particles in the atmosphere depend on their size, and on their chemical and other physical properties, and may range up to several days or a few weeks in the troposphere. The combination of diverse sources and complex chemical composition of particles ensures that the distribution of aerosol characteristics – physical and chemical – is far from uniform. Atmospheric aerosol measurements are too scarce to properly assess global levels. A long-term monitoring program has recently been initiated through the GAW that will supplement the many national and regional programs. Parameters measured may vary from network to network, and intercomparability is often lacking.

Nonetheless, some generalizations about aerosol mass concentrations in different locations can be made (Warneck, 1999). Typically, the concentration of fine aerosols in Polar Regions, especially the Antarctic, is very low ($\sim 1\,\mu g\,m^{-3}$) due to the relative absence of sources. Marine background concentrations are also generally low ($\sim 10\,\mu g\,m^{-3}$) although in windy conditions, and if sea-salt spray is included in the mass measurement, mass concentrations can become much higher. Continental concentrations generally reflect proximity to sources and vulnerability to exposure through prevailing winds. Rural levels are generally in the $30-50\,\mu g\,m^{-3}$ range. In urban centers, fine aerosol concentrations of $50-100\,\mu g\,m^{-3}$ are typical although concentrations $>100\,\mu g\,m^{-3}$ have been recorded (Finlayson-Pitts and Pitts, Jr, 2000, 618–619). Extremely high (several hundred $\mu g\,m^{-3}$) aerosol concentrations are found over and downwind of deserts and forest fires. Aerosol mass measurements are dependent on the particle sizes included in measurements. Information on trends in aerosol concentrations is very limited away from the immediate neighborhood of sources.

REFERENCES

Bradley, R S (1999) *Paleoclimatology: Reconstructing Climates of the Quarternary*, 2nd edition, International Geophysics Series, Academic Press, San Diego, CA, Vol. 64.

Chappellaz, J, Barnola, J M, Raynaud, D, Korotkevich, Y S, and Lorius, C (1990) Ice-core Record of Atmospheric Methane Over the Past 160 000 Years, *Nature*, 345, 127–131.

Carbon Dioxide Information Center (CDIAC) (2000) *CDIAC Online, Current Greenhouse Gas Concentrations*, October 2000 Update, (http://cdiac.esd.ornl.gov/pns/current_ghg.html).

Climate Monitoring and Diagnostics Laboratory (CMDL) (1998) *Summary Report No. 24*, 1996–1997, December 1998.

Crutzen, P J (1995) Overview of Tropospheric Chemistry: Developments During the Past Quarter Century and a Look Ahead, *Faraday Discussions*, 100, 1–21.

Ehhalt, D H (1999) Gas Phase Chemistry of the Troposphere, in *Topics in Physical Chemistry Volume 6: Global Aspects of Atmospheric Chemistry*, ed R Zellner, Springer, Berlin.

Elkins, J (2000) Trends of Atmospheric Nitrous Oxide, in *CMDL FY2000 3rd Quarter Milestones, online edition*, (http://www.cmdl.noaa.gov/milestones/2000/fy2000_3.html).

Emmons, L K, Carroll, M A, Hauglustaine, D A, Brasseur, G P, Atherton, C, Penner, J, Sillman, S, Levy, II, H, Rohrer, F, Wauben, W M F, Van Velthoven, P F J, Wang, Y, Jacob, D, Bakwin, P, Dickerson, R, Doddridge, B, Gerbig, C, Honrath, R, Hubler, G, Jaffe, D, Kondo, Y, Munger, J W, Torres, A, and Volz-Thomas, A (1997) Climatologies of NO_x and NO_y: a Comparison of Data and Models, *Atmos. Environ.*, 31(12), 1851–1904.

Etheridge, D M, Steele, L P, Francey, R J, and Langenfelds, R L (1998) Atmospheric Methane Between 1000 AD and Present: Evidence of Anthropogenic Emissions and Climatic Variability, *J. Geophys. Res.*, 103, D13, 15 979–15 993.

Finlayson-Pitts, B J and Pitts Jr, J N (2000) *Chemistry of the Upper and Lower Atmosphere: Theory, Experiments and Applications*, Academic Press, San Diego, CA.

Haan, D and Raynaud, D (1998) Ice Core Record of CO Variations During the Last Two Millennia: Atmospheric Implications and Chemical Interactions Within the Greenland Ice, *Tellus*, 50B, 253–262.

Hollaway, T, Levy, II, H, and Kasibhatla, P (2000) Global Distribution of Carbon Monoxide, *J. Geophys. Res.*, 105, D10, 12 123–12 147.

Intergovernmental Panel on Climate Change (IPCC) (1996) *Climate Change 1995: the Science of Climate Change*, Second Assessment Report, WG1, Cambridge University Press, Cambridge.

Lelieveld, J and Dentener, F J (2000) What Controls Tropospheric Ozone? *J. Geophys. Res.*, 105, D3, 3531–3551.

Miller, J M and Young, J (1997) The Global Atmosphere Watch – Contributing to our Understanding and Protecting our Changing Atmosphere, *WMO Bull.*, 46(2), 117–122.

Novelli, P C, Masarie, K A, and Lang, P M (1998) Distributions and Recent Changes of Carbon Monoxide in the Lower Troposphere, *J. Geophys. Res.*, 103, D15, 19 015–19 033.

Oltmans, S J, Lefohn, A S, Scheel, H E, Harris, J M, Levy, II, H, Galbally, I E, Brunke, E G, Meyer, C P, Lathrop, J A, Johnson, B J, Shadwick, D S, Cuevas, E, Schmidlin, F J, Tarasick, D W, Claude, H, Kerr, J B, Uchino, O, and Mohnen, V (1998) Trends of Ozone in the Troposphere, *Geophys. Res. Lett.*, 25(2), 139–142.

Prinn, R G, Weiss, R F, Fraser, P J, Simmonds, P G, Cunnold, D M, Alyea, F N, O Doherty, S, Salameh, P, Miller, B R, Huang, J, Wang, R H J, Hartley, D E, Harth, C, Steele, L P, Sturrock, G, Midgley, P M, and McCulloch, A (2000) A History of Chemically and Radiatively Important Gases in Air Deduced From ALE/GAGE/AGAGE, *J. Geophys. Res.*, 105, D14, 17 751–17 792.

Sandroni, S, Anfossi, D, and Viarengo, S (1992) Surface Ozone Levels at the End of the Nineteenth Century in South America, *J. Geophys. Res.*, 97, 2535–2539.

Seinfeld, J H and Pandis, S N (1998) *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*, John Wiley & Sons, New York.

US Environmental Protection Agency (EPA) (2000) *Air Quality Criteria for Carbon Monoxide*, EPA 600/p-99/001F, June 2000.

Warneck, P (1999) Fundamentals, in *Topics in Physical Chemistry Volume 6: Global Aspects of Atmospheric Chemistry*, ed R Zellner, Springer, Berlin.

Whelpdale, D M and Kaiser, M S (1996) *Global Acid Deposition Assessment*, WMO GAW Report No. 106.

World Data Centre for Greenhouse Gases and Other Atmospheric Gases (WDCGG) (2000) *WMO WDCGG Data Summary*, WDCGG No. 22, GAW Data, Volume IV – Greenhouse Gases and Other Atmospheric Gases, March 2000.

World Meteorological Organization (WMO) (1999) *Scienti c Assessment of Ozone Depletion: 1998*, World Meteorological Organization, WMO Global Ozone Research and Monitoring Project – Report No. 44, Geneva.

Albedo

Garth W Paltridge
University of Tasmania, Tasmania, Australia

The original Latin meaning of albedo (i.e., whiteness) has given way to the formal de nition found in most dictionaries and textbooks - namely: albedo is the ratio of the total radiant energy ux re ected or scattered in all directions by an object to the total radiant ux intercepted by the object. Even this formal de nition is strictly appropriate only in the context of certain special cases. These concern re ection of sunlight from entire and isolated bodies such as planets, or the re ection of sunlight (or indeed any radiation) from individual particles or molecules. In the latter case, the ratio of the uxes has the speci c name of single scattering albedo, and is a basic parameter used in the mathematical theories of the scatter of radiation in media such as the atmosphere.

Ground-level measurements of albedo are not made in a way that is consistent with the textbook definition given above. Using a horizontal surface as an example, its measured albedo is the ratio of the upward component of the flux density (flux per unit area) of radiation scattered or reflected from the surface to the downward component of the incoming flux density. The measurement is normally taken in the context of solar radiation reflected by the ground or by some other horizontal surface such as a deck of cloud. It is obtained by an albedometer, which measures separately and simultaneously both the up-welling and down-welling flux densities at some small distance above the surface in question. The sensors are flat-plate absorbers. Thus, the typical observation of albedo does not involve measurement of the radiation scattered in all directions by a particular small area of the surface. Rather, it involves measurement of the radiation scattered into a horizontal unit area above the surface from all parts of the surface within view of the downward-looking flat-plate sensor. The two situations are equivalent only if the surface is horizontally uniform and is of sufficient extent to fill the effective view of the sensor. The effective view is determined by the fact that the sensor response to radiation from an isotropically scattering surface is dominated by radiation in the mid-range of the angles between 0 and 90° from the vertical, and tails off to zero at 0 and 90°. This response pattern ensures, among other things, that the character of the surface immediately beneath the instrument has little impact on the actual measurement.

Albedo is not an intrinsic single-valued property of a surface, but is normally a function of the manner in which the surface is illuminated. In cloud-free skies where the direct beam of the Sun dominates the down-coming solar radiation, the albedos of virtually all ground surfaces increase with solar zenith angle. The increase is usually most significant at zenith angles greater than about 60° and is dependent on surface character. There can be exceptions, however, where the shadowing by surface roughness elements increases with solar zenith angle faster than does the albedo of those portions of the surface that are sunlit. Factors contributing to the angular dependence may include the broad forward-peak in the scatter pattern of individual particles making up the surface (clouds and snow are good examples); Fresnel reflection from horizontally-oriented platelets making up the surface (the leaves of certain types of vegetation are an example); and the fact that the depth of penetration of incoming radiation into a surface, and hence the depth of the actual reflection process, may be reduced as the angle of incidence decreases. The variation with zenith angle is important in the context of climate because of the many geographic and seasonal situations where the daily-average solar zenith angle is large enough for the variation to be highly significant.

Albedo is a factor determining the energy balance of surfaces, objects, or planets as a whole (*see* Energy Balance and Climate, Volume 1). It determines their net solar energy input and hence their temperature and overall thermodynamic behavior. Thus, the usual reference to albedo concerns the total spectrum of incoming solar radiation, although there can be specialist uses that refer to the spectral albedo of particular wavelength bands. In the context of photosynthesis, for instance, the interest may be in the albedo at visible wavelengths of plant leaves. The spectral albedo can vary enormously with wavelength (*see* PAR (Photosynthetically Active Radiation), Volume 2).

The climatologist is usually concerned with the average albedo over an area, which may be several hundred

kilometers on a side. Over land and sometimes over the sea, such areas are made up of a mosaic of individual patches, each of which has its own albedo and albedo behavior. Thus, much effort is currently devoted to the derivation of surface albedo from satellite measurements. In order to obtain data with sufficient resolution, and to satisfy other remote sensing applications, the satellite observations are, of necessity, narrow-beam measurements of the intensity of radiation reflected directly to the satellite from a particular small area on the ground. By using many satellite passes to take multiple observations at different viewing angles and at different solar zenith angles, it is possible to measure what is known as the *bidirectional re ection distribution function* – in other words the overall pattern of the intensity of the sunlight scatters in all directions from the surface in question – and to measure the variation of this function with solar zenith angle and with wavelength. Provided corrections are made for the scatter and absorption of the intervening atmosphere, the intensities can be integrated appropriately over angle and wavelength so as to calculate the upward flux density required to determine albedo. An albedo derived in this way is, in principle, close to the albedo of the formal textbook definition.

The Earth s temperature is highly dependent on its albedo because its thermal radiation to space must, on average, balance its absorption of solar energy. The lower the albedo, the more solar energy will be absorbed, the more thermal energy must be radiated, and the higher must be the temperature to produce that thermal radiation. The Earth s albedo is currently about 0.30, with most of the effect being caused by clouds. As a rough guide, and other things being equal, the Earth s surface temperature would rise about 2 K if the overall albedo of the planet were to be lowered by 0.01. This has an interesting consequence in the context of climate change and of the suggestion that more forests might be grown so as to limit global warming by increasing the uptake of atmospheric carbon dioxide (CO_2). Referring to Table 1, significant increases in forest cover would generally decrease the Earth s albedo. Such a darkening of the planet would tend to reduce the benefit of the increased uptake of CO_2.

Table 1 Approximate values of low solar zenith angle albedo

Surface	Albedo
Smooth water	0.02
Rough sea	0.05
Evergreen and tropical rainforest	0.1
Dry and eucalypt forest	0.15
Dry grassland and semi-desert	0.2
Desert sand to white sand	0.3–0.6
Old snow to new snow	0.4–0.8
Cirrus to middle and low-level cloud	0.3–0.6
Overall planet Earth	0.3

AMIP (Atmospheric Model Intercomparison Project)

The AMIP is an international project in which the performance of atmospheric general circulation models integrated under standardized conditions is evaluated in comparison with observations over a specified period of years. As originally formulated by the International Working Group on Numerical Experimentation of the World Climate Research Program, the AMIP protocol requires the use of the observed monthly mean sea-surface temperature and sea-ice distribution during the decade from 1979 to 1988 as lower boundary conditions for the atmospheric calculations, use of fixed values of the solar irradiance and atmospheric CO_2 concentration, and the calculation of a specific set of standard output variables. More than 30 modeling groups, representing virtually the entire international atmospheric modeling community, are participants. Results of AMIP analyses show the presence and extent of systematic errors in the simulation of many variables, and a marked scatter in the simulations of the mean seasonal cloudiness and precipitation. Comprehensive documentation and databanks for both model output and observations are maintained by the Program for Climate Model Diagnosis and Intercomparison at the Lawrence Livermore National Laboratory (Livermore, CA), while specific regional and phenomenological aspects of the models performance have been examined in a series of diagnostic subprojects conducted by the international modeling and diagnostic communities. AMIP has demonstrated the value of systematic diagnosis and validation in the process of model improvement, and a second phase (AMIP 2) is providing for expanded output and further diagnostic and experimental subprojects (Gates, 1992, 1999).

See also: Model Simulations of Present and Historical Climates, Volume 1.

REFERENCES

Gates, W L (1992) AMIP – The Atmospheric Model Intercomparison Project, *Bull. Am. Meteorol. Soc.*, 73, 1962–1970.

Gates, W L, Boyle, J S, Covey, C, Dease, C G *et al.* (1999) An Overview of the Results of the Atmospheric Model Intercomparison Project (AMIP 1), *Bull. Am. Meteorol. Soc.*, 80, 29–56.

W LAWRENCE GATES *USA*

Angular Momentum

see Atmospheric Angular Momentum and Earth Rotation (**Volume 1**)

Antarctica

Charles R Bentley
University of Wisconsin-Madison, Madison, WI, USA

Antarctica, the continent around the South Pole, comprises two geologically distinct regions, East Antarctica and West Antarctica, separated by the great Transantarctic Mountains but joined together by the all-encompassing ice sheet. The presence of the high ice sheet and the polar location make Antarctica a powerful heat sink that strongly affects the climate of the whole Earth. Furthermore, the annual sea ice cover around the continent, which seasonally reaches an area greater than that of the continent itself, modulates exchanges of heat, moisture, and gases between the atmosphere and ocean and, through salt rejection when it freezes, forces the formation of cold oceanic bottom waters that spread out under the world's oceans.

Both the ice sheet and the sea ice are potentially subject to change in a changing climate; the ice sheet, in fact, may be changing now in response to past climate change. The greatest threat to the inhabited world comes from the West Antarctic ice sheet (WAIS), which rests on a bed far below sea level and so may have the potential for rapid shrinkage. The Antarctic is so vast, remote, and dif cult to monitor, however, and the physical behavior of the ice sheet so complex, that there is as yet no de nitive demonstration (or disproof) of secular change, even though a pronounced climatic warming is ongoing in one northerly portion of the continent. Measurements from satellites early in the 21st century should settle the question of current growth or shrinkage, but prediction will remain problematic for many years.

GEOGRAPHY AND GEOLOGY

Antarctica is Earth s fifth largest (area *circa* 14 200 000 km^2) and most southerly continent. It comprises two major regions: West Antarctica, which includes the Antarctic Peninsula, and East Antarctica. They are so named because they lie mostly (although not entirely) in the Western and Eastern Hemispheres, respectively. The boundary between the two is the vast Transantarctic Mountains, which extend 3000 km between the Atlantic and Pacific shores of the continent.

Most of the continent is covered with ice. The Antarctic ice cover comprises three components: the thick, grounded, inland ice that rests on a (more or less) solid bed, the thinner, permanent floating ice shelves, and the ephemeral sea ice that surrounds the continent. (By definition, the inland ice and the ice shelves together constitute the ice sheet.) The inland ice ranges in thickness up to 5 km, with an average of about 2400 m. In central East Antarctica, the ice surface rises to a height of 4000 m. Because of the ice sheet, Antarctica is, on average, by far the highest of the continents.

Antarctica lies in a circumpolar position (the South Pole lies within the continent but not at its center) completely surrounded by the ocean. Its nearest neighbor continent is South America, 1000 km from the Antarctica Peninsula, across the Drake Passage. Antarctica and South America are indirectly connected by the long, curve of islands known as the Scotia Arc (Figure 1).

The continents of the Southern Hemisphere were not always so far apart, however; 200 million years ago they were all joined in the super-continent known as Gondwana. Then, one after another the other continents separated from Gondwana until, last of all, the Drake Passage opened up about 30 million years ago. That opening completed the isolation of Antarctica and permitted the establishment of the uninterrupted Circumpolar Current in the Southern Ocean. The combination of continuous oceanic and atmospheric circumpolar circulation was instrumental in allowing the continent to cool enough for the ice sheets to become established.

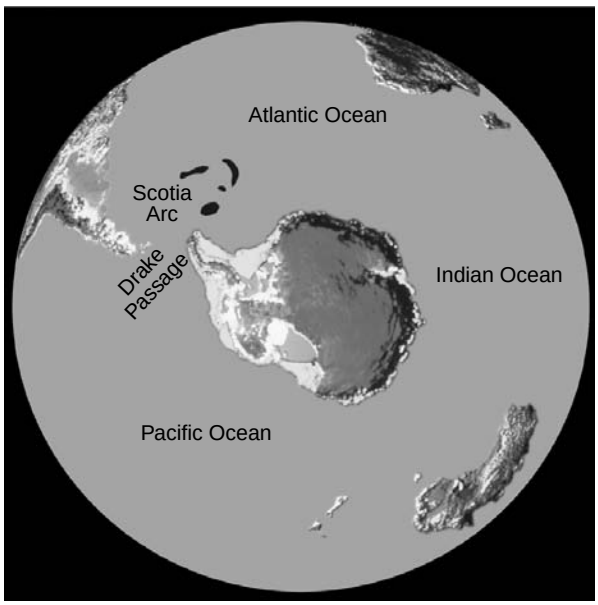


Figure 1 Antarctica in the Southern Ocean. Light-shaded area around the continent is the continental shelf. Other land masses, clockwise from the top, are Africa (with Madagascar), Australia (with Tasmania), New Zealand, and South America

East and West Antarctica are geologically very different. If all Antarctic ice were removed and the solid earth beneath were given time to rise in response to the removal of its weight (isostasy), East Antarctica would be one large continental land mass, roughly the size of Australia, with some inland seas. West Antarctica, on the other hand, would comprise several large mountainous islands separated from East Antarctica by open ocean, much of it over 1 km deep. Geographically, the whole region would resemble Australia and New Guinea, except that the intervening body of water would be much deeper in Antarctica (Figure 2).

East Antarctica is truly continental, with crustal thicknesses comparable to the other continents and with rocks of all ages (the most ancient are more than three billion years old). West Antarctica is geologically much younger with a separate history of continental drift and a central region of thin, stretched crust covered with old marine sediments.

South of the Antarctic Circle, where the vast majority of the continent lies, the Sun does not set in mid-summer and does not rise in mid-winter. At the South Pole itself, the year comprises one 6 month day and one 6 month night.

The South Pole, on the spin axis of the Earth and at high altitude on dry land (unlike the North Pole, in the middle of the Arctic Ocean), has characteristics valuable for certain types of observations, which are carried out at the United States Amundsen-Scott Station there. The axial location provides a unique simplification of some earthquake waves. The half-year-long winter night, together with the extreme clarity of the thin atmosphere, provide unparalleled opportunities for long-term, continuous astronomical observations. And instruments emplaced deep in the 2800 m thick ice sheet count flashes of light from the rare collisions between neutrinos (which have passed through the entire Earth) and water (ice) molecules.

WEATHER AND CLIMATE

Antarctica is the coldest place on the surface of the Earth. It is polar, so the Sun shines only obliquely on the surface; its surface is at mountain-top heights, in the colder part of the troposphere; and the bright snow-covered surface has a very high albedo), about 85% – so that little of the

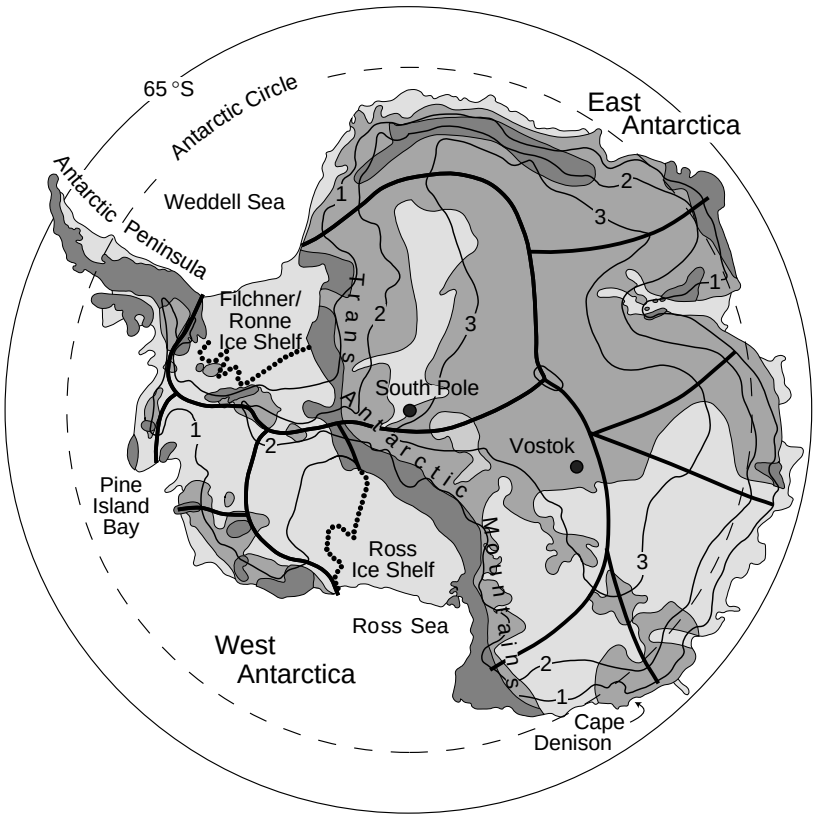


Figure 2 Physical characteristics of Antarctica. Shading indicates elevation of the continental surface. Darkest shade denotes mountains and other mostly exposed rock; lightest shade denotes bed elevations below sea level; intermediate shade denotes a subglacial bed above sea level. Thin, numbered lines give elevations of the continental surface in kilometers. Heavy dotted lines mark the grounding lines of the two biggest ice shelves. Heavy black lines demark ice drainage basins

incoming solar radiation is absorbed. The coldest recorded temperature is close to $-90\,^{\circ}$C ($-130\,^{\circ}$F) at the Russian Station Vostok in central East Antarctica. Especially during the long night, Antarctica is a powerful sink for heat – air that circulates southward in the upper atmosphere cools, sinks, and flows out along the surface, creating katabatic winds, which blow particularly strongly near the coast (at an Australian station on Cape Denison the wind averaged nearly $20\,\mathrm{m\,s^{-1}}$ (43.5 mph) for an entire year). The strong offshore winds in some places blow the sea ice away from the shore as fast as it can freeze, producing ice-free coastal polynyas, which have an important effect on both oceanic circulation and the regional climate (see later section on Sea Ice, Ocean, and Atmosphere).

The katabatic winds force a cyclonic circulation higher in the atmosphere over the continent. The resulting circumpolar vortex isolates the Antarctic atmosphere from incursions by the storms that proliferate in the oceans around the continent. It is that isolation, together with the occurrence of polar stratospheric clouds, that allows stratospheric ozone depletion (the *Antarctic ozone hole*), to occur annually.

Summers, with 24-hour daylight, are as much as $50\,^{\circ}$C warmer and less stormy than winters. Although the ice sheet has a refrigerating effect that makes it rare for the temperature to exceed freezing on its surface, more balmy temperatures abound in coastal areas where there is seasonally snow-free ground and the dark, low-albedo rocks absorb much of the incoming solar radiation. However, even in January, the warmest month, a mean temperature of $-10\,^{\circ}$C is found just a short distance inland from the edge of the ice sheet. That low temperature near the margin means there would be little surface melting even if the Antarctic climate were to warm by 4 or $5\,^{\circ}$C in an enhanced-greenhouse-effect world (see later section on The Ice Sheet).

In a CO_2-doubled world, the temperature in Antarctica may increase by several degrees; in some projections the temperature rise in Antarctica would be more than elsewhere, although in other projections it is less. Regardless of projections, a pronounced warming has been observed over the last several decades, but only along the Antarctic Peninsula. Elsewhere around the continent there appear to be warming trends in a few places and cooling trends in a few, but no significant average trend has been discerned.

An indirect way to look for temperature changes on a longer time scale than is covered by instrumental records is to examine the change in snow accumulation rates with time. Greater snowfall is naturally associated with warmer temperatures, as a warmer atmosphere can hold more water vapor. Ice cores from a majority of sites scattered about the ice sheet, both fairly near the coast and far inland, reveal an increasing snowfall over recent decades (up to a century or so), but the trend is not universal; a significant minority of sites shows the opposite trend or none at all. Furthermore, one or two sites show a reversal of trend, from increasing accumulation to decreasing accumulation, in just the last decade or two. On average for the continent, exclusive of the Antarctic Peninsula, there is no significant trend in snowfall either.

THE UPPER ATMOSPHERE

The near-surface magnetic field of the Earth can be approximated by that of an imaginary bar magnet at the Earth s center but tilted about 12° relative to the Earth s axis. The geomagnetic poles, the points on the surface directly above the ends of the imaginary bar magnet, are the points toward which the force lines of the Earth s external magnetic field converge. The South Geomagnetic Pole lies near Vostok Station.

At heights of many Earth radii, the magnetic lines of force are distorted by the stream of charged particles from the Sun, known as the solar wind, to form the magnetosphere, which deflects the solar wind away from the Earth s surface everywhere except at the geomagnetic poles, where those lines come to Earth. The charged particles flowing down through the polar cusps (the holes at the geomagnetic poles) in both polar regions collide with the atoms and molecules of the atmosphere at elevations of a few hundred kilometers to cause the auroras (*aurora australis* in the south). The auroras are merely the visually most spectacular phenomena associated with the polar cusps and the solar winds. Others include magnetic storms and worldwide radio blackouts.

These solar-terrestrial phenomena are not particularly subject to change because of human or natural activities on Earth, although they are strongly affected by variation in solar activity. The importance of Antarctica with regard to these polar phenomena is that it provides a stable platform from which they can be observed, in notable contrast to the Arctic Ocean.

THE ICE SHEET

The ice sheet is actually a giant glacier, and like most glaciers it forms by the continual accumulation of snow on its surface. As successive layers of snow build up, the layers beneath are gradually compressed into solid ice. The snow mass input is balanced by a glacial outflow, so the height of the ice sheet stays approximately constant through time. The ice is driven by gravity downhill, i.e., in the direction of the surface slope, from the highest points in the interior to the coast. There it mostly breaks off as icebergs, which are carried away by the ocean currents and eventually melt, thus returning the water to the ocean whence it came.

Only a small proportion of the mass loss from the ice sheet occurs by melting from the surface – summertime melt from the margins of the ice sheet occurs only in the northern Antarctic Peninsula and the northernmost fringes of East Antarctica.

The outflow from the inland ice is organized into a series of drainage basins, separated by ice divides, much like water flow on other continents. Most drainage basins concentrate the flow into either narrow outlet glaciers (particularly through the Transantarctic Mountains) or fast-moving ice streams, which are surrounded by slow-moving ice rather than rock walls (Figure 3). Ice in a few drainage systems flows in a diverging pattern that spreads out across a wide section of the coast. The speed of flow increases towards the coast, both because of the increasing area of snowfall upstream and because the ice generally thins outward. Where there is convergence into an outlet glacier or ice stream, the speed of flow is increased even more, attaining values that are more than a kilometer per year in a few places.

The ice sheet in West Antarctica flows mostly into the Ross Ice Shelf, at the head of the Ross Sea, the Filchner/Ronne Ice Shelf (really two separate but connected ice shelves) at the head of the Weddell Sea, and Pine Island Bay, which contains no ice shelf. The two great ice shelves are similar in size; each about the area of Spain. Ice shelves are also driven by gravity, spreading out over the ocean under their own weight. They are supplied with ice partly by the glaciers and ice streams that flow into them and partly by the snow that falls on their surfaces. Unlike the inland ice, they are subject to substantial mass exchange at their bases, melting in some places and freezing in others.

Despite the extreme cold on the high surface of the inland ice sheet, the heat from the interior of the Earth is sufficient to cause melting at the base of the ice across wide areas; the overlying ice acts as an effective protective blanket. The melting is only a few mm year^{-1} or less, but over the long history of the ice sheet it has been enough to create dozens of lakes beneath the ice, as the water flows into low spots in the bed. The largest and best studied (solely by remote sensing), Lake Vostok, whose southern end underlies Vostok Station, is the size of Lake Ontario and is apparently hundreds of meters deep. Since it is only the very old basal ice that melts, Lake Vostok and its companions contain an environment record, and possibly life, that has been isolated from the rest of the world for a million years or more. Researchers eagerly await the development of contamination-free methods of sampling the lakes (*see* Vostok, Subglacial Lake, Volume 1).

Of all the possible changes in Antarctica the one with the greatest potential for destructive consequences around the world is a diminishment in mass of the Antarctic inland ice sheet, with a corresponding, direct, and immediate rise in sea level. If all the inland ice were to melt (melting ice shelves, which are already afloat, does not change sea level), some of the resulting water would be used to fill in the depressions where the base of the ice is now below sea level, but most would go to raising sea level, by about 70 m. There is no chance of that happening within the time scale of human interest, but the loss of even a significant fraction of it could have serious consequences. Interest is concentrated particularly on the West Antartic ice sheet, because it rests on a bed far below sea level, which may make it particularly vulnerable to accelerated discharge into the ocean. If the entire WAIS were to disappear, sea level would rise by 5 or 6 m, causing substantial disruption to human life, to say nothing of coastal facilities and activities, around the world.

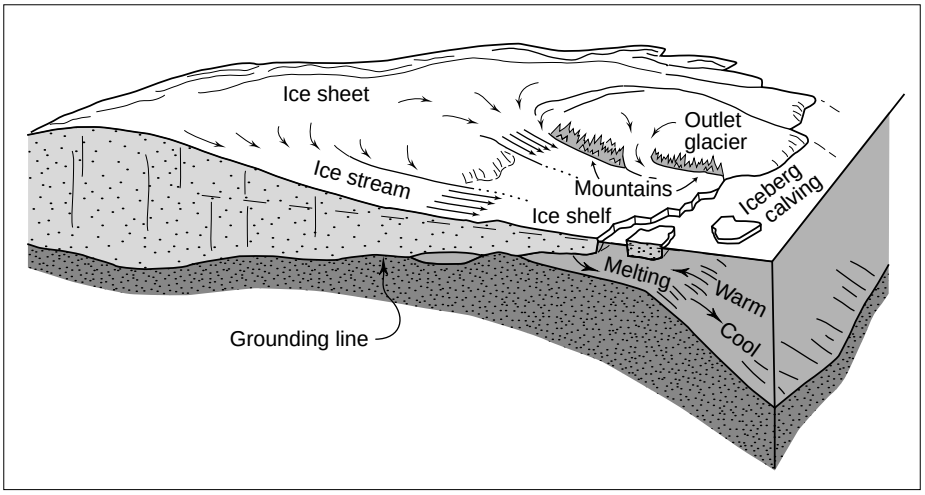


Figure 3 Diagrammatic sketch of an ice sheet containing ice streams, an outlet glacier, an ice shelf, and icebergs. Dark stippled, rock; light stippled, ice; solid gray, water. Arrows show directions of ice and water flow

At present, both the East and West Antartic ice sheets (WAISs) are nearly in mass balance, meaning that the mass inputs and outputs (both about 6 mm of sea level equivalent per year) are nearly the same. Unfortunately, measurements across the huge continent are not accurate enough overall to be more precise – the ice sheet could be producing as much as half of the 2 mm year^{-1} sea level rise or it could actually be drawing water out of the ocean (*see* Sea Level, Volume 1). Complementary satellite measurements of changes in surface height and changes in gravitational attraction (directly dependent on mass) in the early 21st century should decrease the uncertainty greatly.

Predicting the future of the WAIS, however, is more difficult. It seems clear that the WAIS substantially disappeared (and then reformed) sometime (presumably during an interglacial time) in the last 600 000 years, a time span that covers half a dozen ice age cycles. That suggests, but does not prove, that it could happen again, perhaps during the current interglacial epoch in which we now live. The crucial questions, though, from a societal standpoint, are how soon could it happen and how fast.

The only way that short-term climate change (including anthropogenic change) could effect shrinkage of the ice sheet is through oceanic warming. Because the ice shelves are in contact with the ocean, and the ocean can respond much faster to rising temperatures than the solid ice can, the undersides of the ice shelves are the most likely loci of climate-induced change – if warmer waters circulate under the ice shelves in the future, melting could be enhanced, and ice shelves could thin. One modeled result is that the grounding line, the boundary between inland grounded ice and the floating ice shelf, would retreat unstably into the interior of the ice sheet by a process of rapid acceleration of ice stream discharge. Other models, however, suggest that the ice streams would respond stably to restore a mass balance, so that nothing drastic would occur. The answer is simply not known on a theoretical or experimental basis because of too many uncertainties about the physics of ice stream flow. However, there may be a clue from Pine Island Bay, where the two fastest-flowing ice streams in Antarctica terminate. The absence of an ice shelf there might make those ice streams particularly susceptible to rapid acceleration. In fact, both of them do show shrinkage, according to recent experimental results reported (as of 2001).

Past climate may be more important to present WAIS behavior than present climate; because the large size of the ice sheet means that it takes thousands of years to respond fully to external changes. During the last Ice Age the WAIS was perhaps twice as large as now and it may not yet have achieved equilibrium. The glaciological and geological records of changes in the last few thousand years demonstrate no shrinkage of the ice sheet rapid enough to cause it to disappear in the next few hundred years, but indicate that there remains the possibility of a disappearance in the next few thousand years. In terms of sea level, the disappearance of the WAIS in 5000 years would be equivalent to an average rise in sea level of 1 mm year^{-1}, 50% of the current rate. That is enough to be worthy of serious study, but it is certainly not catastrophic.

PALEOCLIMATE

The continual accumulation of snow on the interior ice sheet without melting provides an environmental record of incomparable value. Air in the precipitated snow is preserved in bubbles in the ice, which thus contains a sampling of ancient atmospheres back to the age of the deepest, oldest ice, perhaps more than a million years old. Oxygen and hydrogen isotopes in the ice itself and the temperatures in the borehole provide additional information about past climate. A continuous, 400 000-year record has been collected by ice core drilling to a depth of 3623 m at Vostok Station.

Paleoclimatic records of the last 75 000 years or more also have been recovered from deep ice cores at other sites, in central East Antarctica, central West Antarctica, and East Antarctica near the Ross Sea. All (including Vostok) except the near-coastal site show paleo-temperatures that are different in detail from the record in central Greenland, while still showing the main broad changes. The record near the Ross Sea, however, more closely resembles that from Greenland than those from elsewhere in Antarctica. The implication is that some of the rapid changes generally associated with events in the North Atlantic Ocean (e.g., Heinrich events) have Antarctic counterparts, but that they may not have propagated far into the interior.

Ice cores have a high temporal resolution that provides a superb complement to marine sediment cores, which stretch farther back in time but with poorer resolution. Early in the 21st century additional long paleoclimatic records will be collected by ice coring in both East and West Antarctica, including at a West Antarctic site where the snow accumulation rate is high enough to anticipate dating of annual layers. These new cores will yield much new information about climatic teleconnections between the north and the south.

SEA ICE, OCEAN, AND ATMOSPHERE

In winter, the continent is surrounded by a belt of sea ice hundreds to two thousand kilometers wide, whose total area is more than that of the continent itself (*see* Sea Ice, Volume 1). A fourth of that area, however, is open water in leads and polynyas. By the end of the summer, 80–90% of the sea ice has melted away, which allows ship access

directly to most of the continent (Figure 4). It is then that supplies and personnel are brought to the Antarctic bases that dot the coast.

In contrast to the multi-year ice in the Arctic Ocean, Antarctic ice is relatively thin, mostly less than a meter. Nonetheless, it forms an effective barrier to heat and moisture exchange between the ocean and atmosphere and thus has a fundamental influence on the climate of the Southern Hemisphere. Sea ice freezes at its base, a process that leaves salt in the uppermost water, increasing its density and leading to the sinking of cold, relatively saline water that spreads northward to form the bottom waters of the world s oceans. Additionally, the sea ice zone is a critical habitat for marine algae, the primary producers of life in the ocean, and hence for higher level organisms as well.

Because Antarctic sea ice is so thin, it is particularly vulnerable to a diminishment in extent in a warmer climate; the consequent effect on physical and biological processes in the sea ice zone could be of major significance. Diminished oceanic coverage (including more extensive polynyas and leads within the sea ice zone) would increase the heat and water exchange with the atmosphere, which would tend to increase cloudiness, wintertime air temperatures, and moisture in the atmosphere. That could enhance the increase in snowfall over the continent that is expected from the direct effect of climatic warming. Diminished sea ice production near the fronts of the ice shelves could lead to decreased oceanic circulation beneath the shelves; this is a feedback that tends to counteract enhanced basal melting under the shelves from the influx of warmer ocean water. At the same time, the production of Antarctic Bottom Water would be affected (*see* Ocean Circulation, Volume 1).

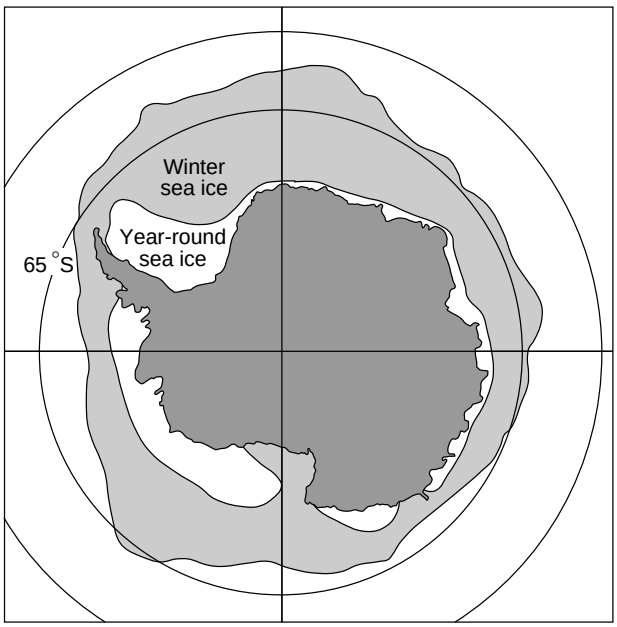


Figure 4 Average extent of sea ice around Antarctica in summer (white) and winter (white plus gray)

Sea ice is also an important regulator of gas exchange between the ocean and the atmosphere. The ocean can be either a source or a sink for carbon dioxide, depending on regional variations in physical and biological circumstances. Sea ice acts as a physical barrier to gas exchange; cold waters in the Southern Ocean tend to remove CO_2 from the atmosphere, so a decreasing sea ice extent might increase CO_2 absorption. Even more important may be the role of the sea ice in controlling oceanic primary production, which is a powerful factor in removing CO_2 from the atmosphere and, eventually, depositing it on the ocean floor (the biological pump). Changes in sea ice cover and changes in seasonal behavior of the marginal sea ice zone could have feedback effects on atmospheric CO_2, although even the sign of the change is difficult to predict.

More open water decreases the albedo of the surface, increasing the absorption of radiation. However, the sea ice zone tends to be cloudy and an increase in cloud cover will increase the albedo, thus providing a negative feedback. Generally speaking, clouds can act either to increase or decrease surface temperatures, depending on their type and height. Thus the net effect of changing cloudiness as a result of climatic change over Antarctica cannot yet be predicted, even though that effect may be important.

Satellite measurements of the extent of the Antarctic sea ice have been continuous since the early 1970s. Trends of increasing or decreasing ice cover over a decade or more have been observed in some sectors of the ocean around Antarctica and large interannual and regional variations are common, but no general secular trend has yet been detected.

INTERNATIONAL RESEARCH

Since the International Geophysical Year (IGY) in 1957–1958, there has been a continuous international research effort in Antarctica. More than twenty countries now have permanent, year-round stations in the Antarctic region. Research programs are coordinated by the Scientific Committee on Antarctic Research (SCAR), to which all nations with Antarctic research programs belong.

Anthropocene

The *Anthropocene* has been proposed by Paul J Crutzen and Eugene Stoermer (2000) as a new geological epoch

to mark the strong, in many cases dominating, impact of humans on the Earth s ecology and geology. The Anthropocene covers the most recent 200 years of the Holocene, the geological epoch of the past 10–12 thousand years since the last glacial (*see* Earth System History, Volume 1).

The proposal to introduce the Anthropocene is based on the following arguments. The expansion of humankind, both in numbers and *per capita* exploitation of Earth s resources, has been astounding. During the past three centuries, human population has increased 10-fold to 6000 million, accompanied, for instance by a growth in cattle population to 1400 million. Urbanization has increased 10-fold in the past century. Over a few centuries, humankind will virtually exhaust the fossil fuels that were generated over several hundred million years by the biosphere. The release of sulfur dioxide (SO_2) to the atmosphere by coal and oil burning, globally about 160 Tg $year^{-1}$, is at least two times larger than the sum of all natural emissions. About 30–50% of the land surface has been transformed by human action. More artificial nitrogen is now applied in the form of fertilizers in agriculture than is fixed naturally in all terrestrial ecosystems. The escape into the atmosphere of nitric oxide (NO) from fossil fuel and from biomass combustion in the tropics likewise is larger than the natural inputs, giving rise to photochemical ozone (smog) formation in extensive regions of the world. More than half of all accessible fresh water is used by humankind. Human activity has increased the species extinction rate one thousand to 10 000-fold in the tropical rain forests. The concentrations of several climatically important greenhouse gases have substantially increased in the atmosphere: carbon dioxide (CO_2) by more than 30% and methane (CH_4) by about 150% (*see* Carbon Dioxide, Recent Atmospheric Trends, Volume 1; Methane, Volume 1). Furthermore, humankind releases many toxic substances into the environment, and even some that are not directly toxic to humans cause significant impacts through indirect pathways. For example, the increasing concentrations of chlorofluorocarbon gases have caused the Antarctic ozone hole and would have destroyed much of the ozone layer if no international regulatory measures to end their production had been taken (*see* Ozone Hole, Volume 1). Coastal wetlands have also been affected by humans, resulting in the loss of 50% of the world s mangroves. Fishing removes more than 25% of the primary production of the oceans in the upwelling regions and 35% in the temperate continental shelf regions. Because of the anthropogenic emissions of CO_2, surface temperatures are projected to increase by at least several degrees, and sea level by a meter or more over the next few centuries, with some part of the changes lasting many thousands of years and perhaps may even alter the natural occurrence of glacial–interglacial cycling (*see* Projection of Future Changes in Climate, Volume 1; Sea Level, Volume 1).

To assign a more specific date to the onset of the Anthropocene is somewhat arbitrary, but we propose the latter part of the 18th century as when the global effects of human activities started to become noticeable. This date is supported by data retrieved from glacial ice cores that show the beginning of the human-induced increase in the atmospheric concentrations of several greenhouse gases, in particular CO_2 and CH_4. Such a starting date also coincides with James Watt s invention of the steam engine in 1784.

FURTHER READING

Berger, A and Loutre, M-F (1996) Modelling the Climate Response to Astronomical and CO_2 Forcings, *C.R. Acad. Sci. Paris*, 323(II A), 1–16.

Crutzen, P J and Graedel, T E (1986) *Sustainable Development of the Biosphere*, eds W C Clark and R E Munn, Chapter 9, Cambridge University Press, Cambridge.

Crutzen, P J and Stoermer, E (2000) The Anthropocene, Global Change International Geosphere Biosphere Programme, Newsletter 41, May, 2000, c/o The Royal Swedish Academy of Sciences, Stockholm, Sweden.

Pauly, D and Christensen, V (1995) Primary Production required to Sustain Global Fisheries, *Nature*, 374, 255–257.

Turner, II, B L, Clark, C, Kates, R W, Richards, J F, Mathews, J T, and Meyer, W B (1990) *The Earth as Transformed by Human Action*, Cambridge University Press, Cambridge.

Vitousek, P M, Mooney, H A, Lubchenco, J, and Melillo, J M (1997) Human Domination of Earth s Ecosystems, *Science*, 277, 494.

Watson, R T, Meira Filho, L G, Sanhueza, E, and Janetos, A (1990) *Climate Change. The IPCC Scienti c Assessment*, eds J T Houghton, G J Jenkins, and J J Ephraums, Chapter 1, Cambridge University Press, Cambridge.

Wilson, E O (1992) *The Diversity of Life*, Penguin Books, London.

PAUL CRUTZEN *Germany*

Anthropogenic

see Anthropogenic (**Volume 3**)

Anthropogenic Climate Change

see Depletion of Stratospheric Ozone (*Opening essay*, **Volume 1**); Projection of Future Changes in Climate (*Opening essay*, **Volume 1**)

Anthropogenic Composition Changes

see Air Quality, Global (**Volume 1**); Trends in Global Emmisions: Carbon, Sulfur, and Nitrogen (*Opening essay*, **Volume 3**)

Anticyclone

An *anticyclone* is a region of high atmospheric pressure with horizontal winds that spiral outward from its center, rotating clockwise in the Northern Hemisphere (counterclockwise in the Southern Hemisphere). Anticyclones typically have a horizontal extent of a few thousand kilometers. Anticyclones are associated with sinking air, which replaces the air at the surface that is spreading outward while rotating under the influence of the Coriolis effect due to the Earth s rotation. The sinking motion, referred to as subsidence, leads to warming that suppresses cloud development and traps air pollution, so anticyclones are generally associated with fair weather, and where there are sources of pollution, with smog episodes.

See also: Atmospheric Motions, Volume 1.

KEITH L SEITTER *USA*

Arctic Air Quality

Leonard A Barrie
Pacific Northwest National Laboratory, Richland, WA, USA

Arctic pollution was rst noticed and called Arctic haze by Canadian pilots in the late 1940s and by American of cers aboard weather reconnaissance ights in the Alaskan Arctic in the 1950s. As an environmental issue, Arctic haze lay dormant until the 1970s and 1980s. It was then investigated by atmospheric scientists and found to be widespread and linked with acid rain type air pollution originating mainly from Europe and Russia (Barrie, 1986). At that time, there was growing awareness worldwide that air pollution is not just con ned to small areas around urban or industrial sources but is transported long distances before being removed to the Earth's surface. This leads to regional problems such as acid rain and northern air pollution. Subsequent research revealed the presence of a host of other contaminants in Arctic air and snow. These include: greenhouse gases (GHG), nitrogen oxides, reactive hydrocarbons, persistent organic pollutants (POPs) such as herbicides and pesticides and other potentially toxic substances such as radionuclides, lead (Pb) and mercury (Hg) (AMAP, 1998; MacDonald et al., *2000).*

A frequent question is where does Arctic pollution come from. Of course air pollution originates from human activities somewhere on Earth, but generally not in the Arctic. These substances must be transported by winds to the Arctic. The potential upwind source region of a particular Arctic air pollutant depends on how long-lived it is in the environment. GHG, POPs and mercury are very persistent. They survive for years in the environment. This is long enough for winds to mix them throughout the Earth s atmosphere. Long-lived contaminants can therefore originate from sources anywhere on Earth or, at least those in the Northern Hemisphere (Wania and Mackay, 1996). Shorter-lived soluble or reactive gases (atmospheric lifetimes of less than three months) such as sulfur oxides, nitrogen oxides and non-methane hydrocarbons as well as particulate matter containing sulfates, heavy metals and radionuclides come from an upwind air shed that is mainly restricted to the mid to high-latitudes of the Northern Hemisphere (Figure 1). Arctic haze is formed from this shorter-lived pollution. The main potential source regions of shorter-lived Arctic haze in the Northern Hemisphere are Europe/Russia, eastern North America and Southeast Asia. However, in practice, seasonal variations in rain or snow as well as in wind trajectories conspire to deliver most of this pollution to the Arctic from Europe and Russia over frozen snow-covered land during winter and spring (see arrows in Figure 1). Most pollution from eastern North America and Southeast Asia blows eastward over the stormy Atlantic and Pacific Oceans where it is removed by rain and dissolution in ocean water. In the period December–April, atmospheric concentrations of sulfate haze aerosols in the Arctic are 10–20 times higher than in the rest of the year and they follow trends in emissions of sulfur dioxide (SO_2) from Europe and Russia (Sirois and Barrie, 1999).

Another frequent question is what impact does this pollution have on people and the environment. The primary effect of air pollution are contamination of food sources by airborne POPs, mercury and radionuclides deposited to Arctic ecosystems (AMAP, 1998). Arctic people are generally exposed to radionuclides through terrestrial food webs and to POPs and mercury through aquatic food chains. They are most exposed to polychlorinated biphenyls (PCBs), certain pesticides and mercury through long aquatic food webs that result in bioaccumulation in mammals, birds and

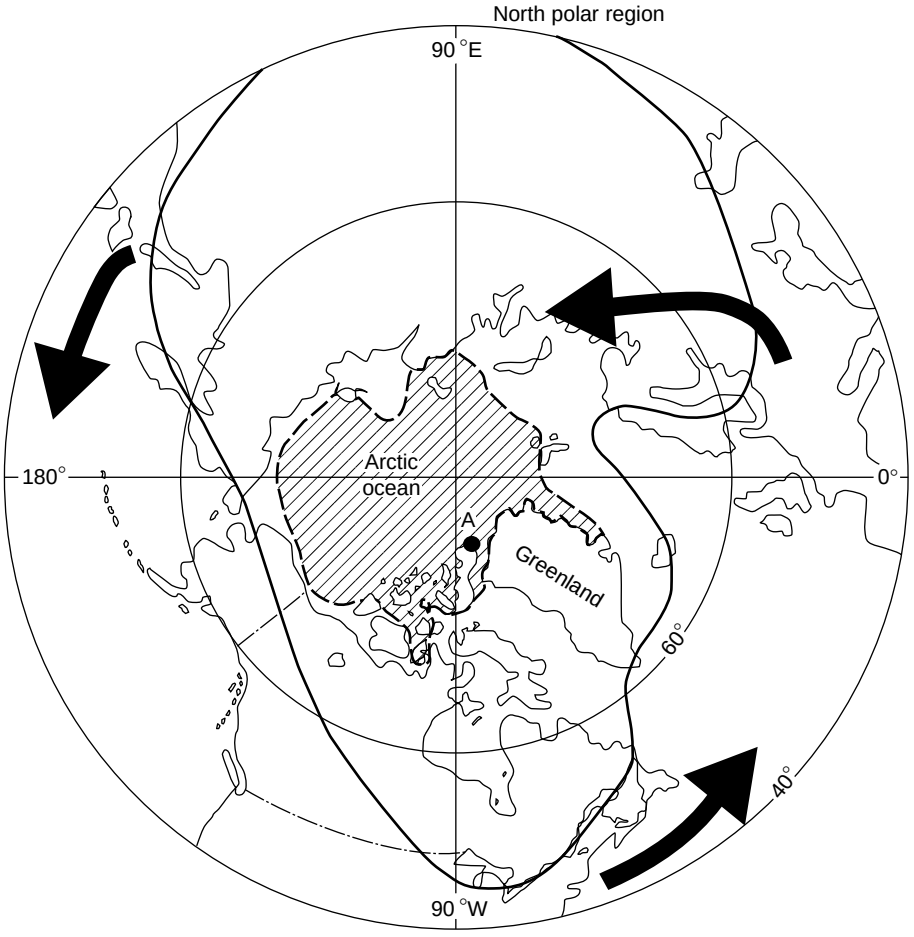


Figure 1 The Northern Hemispheric source region of shorter-lived Arctic haze type air pollution, the main pathways of air out of the three major continental source regions (arrows), the mean position of the Arctic front in winter (dark line) and the location of ice cover on the Arctic ocean at its minimum in summer (hatched area). European and Russia sources are channeled directly into the Arctic in winter while sources in eastern North America and Southeast Asia move eastward over stormy ocean (adapted from AMAP (1998) with permission)

fish. Exposure to radionuclides from nuclear bomb tests or nuclear power plant accidents is mainly through atmospheric transport and deposition to terrestrial ecosystems. Particular soil and vegetation characteristics concentrate some radionuclides in plants and animals (reindeer/caribou, game, mushrooms).

POPs consist of two classes of persistent synthetic organic compounds: those synthesized for industrial purposes and those for pesticidal and herbicidal applications. The ones that are seen in remote regions such as the Arctic have largely been banned from usage in many countries but their presence continues to be felt because of persistence and/or continued use in some countries (MacDonald *et al.*, 2000; Harner *et al.*, 2000). Compounds that have been detected in the Arctic include eight industrial POPs and 10 pesticidal/herbicidal POPs or their metabolites. They have been measured throughout the Arctic in the 1990s (Halsall *et al.*, 1998).

The most abundant POP in the Arctic Ocean is hexachlorocyclohexane (HCH). It is the only one that has been measured long enough to correlate levels in Arctic air with global usage in agriculture (Li *et al.*, 1998). It serves as an example of the complex nature of pathways of these substances and of the unique physical and chemical properties of the Arctic environment that lead to preferential cold-trapping of this substance in the polar region. Production of HCH started approximately in 1950 and peaked in 1980. Most use was in China, India and the former Soviet Union. This high level of use coupled with strong partitioning of HCHs into colder water resulted in a preferential out-gassing in warm climates, transport to the Arctic and cold-trapping by the Arctic Ocean. It is estimated that several thousand tonnes of HCHs are now contained within the surface layer of the Arctic Ocean (MacDonald *et al.*, 2000) providing a reservoir for sea-atmosphere exchange and a continuous source of contamination in the Arctic food

chain. By 1993, HCH emissions declined to approximately 10% of their 1980 levels as a result of bans or restrictions in China and other major use countries. Arctic atmospheric levels of HCH have paralleled usage trends decreasing to a point that this reversibly soluble gas is flowing back out of Arctic Ocean waters into the Arctic atmosphere.

The Arctic atmosphere near the ocean surface after polar sunrise from February–May has a unique chemistry that destroys atmospheric ozone and enhances mercury deposition. Discovered in the late 1980s (Barrie *et al.*, 1988), surface ozone depletion occurring in spring over vast areas of the Arctic Ocean and Hudson Bay in the lower atmospheric (0–2 km) is a phenomenon somewhat analogous to stratospheric ozone destruction but with a difference. It is driven by sea salt bromine (Br) and chlorine (Cl) generated from chemical reactions occurring in sunlight on surface snow and ice, whereas in the stratosphere, ozone depletion is driven by bromine and chlorine from solar degradation of gases such as chlorofluorocarbons (CFCs) and methyl bromide (CH_3Br) used in termite control . The most remarkable impact is on the biosphere through enhancement of mercury deposition by a factor of approximately six (Schroeder *et al.*, 1999). Elevated mercury levels in Arctic char and lake trout are in part due to this effect.

REFERENCES

AMAP (1998) *AMAP Assessment Report: Arctic Pollution Issues*, Arctic Monitoring and Assessment Program (AMAP), Oslo, Norway, 12, 859.

Barrie, L A (1986) Arctic Air Pollution: an Overview of Current Knowledge, *Atmos. Environ.*, 20, 643–663.

Barrie, L A, Bottenheim, J W, Schnell, R C, Crutzen, P J, and Rasmussen, R A (1988) Ozone Destruction and Photochemical Reactions at Polar in the Lower Arctic Atmosphere, *Nature*, 334, 138–141.

Harner, T, Jantunen, L M M, Bidleman, T F, Barrie, L A, Kylin, H, and Strachan, W M J (2000) Microbial Degradation is a Key Elimination Pathway of Hexachlorocyclohexanes from the Arctic Ocean, *Geophys. Res. Lett.*, 27, 1155–1158.

Halsall, C J, Bailey, R, Stern, G A, Barrie, L A, Fellin, P, Muir, D C G, Rovinsky, F Ya, Kononov, E Ya, and Pastukov, B (1998) Organohalogen Pesticides in the Arctic Atmosphere, *Environ. Pollut.*, 102, 51–62.

Li, Y-F, Bidleman, T F, Barrie, L A, and McConnell, L L (1998) Global Hexachlorocyclohexane Use Trends and their Impact on the Arctic Atmospheric Environment, *Geophys. Res. Lett.*, 25, 39–41.

MacDonald, R, Barrie, L A, Bidleman, T F, Diamond, M L, Gregor, D J, Semkin, D J, Strachan, R G, Li, W M J, Wania, Y F, Alaee, F *et al.* (2000) Contaminants in the Canadian Arctic: Five Years of Progress in Understanding Sources, Occurrence and Pathways, *Sci. Tot. Environ.*, 254(2 and 3), 93–234.

Schroeder, W H, Anlauf, K G, Barrie, L A, Lu, J Y, Steffen, A, Schneeberger, D R, and Berg, T (1998) Mercury Vapor Depletion in Arctic Air during Springtime, *Nature*, 394, 331.

Sirois, A and Barrie, L A (1999) Arctic Lower Tropospheric Aerosol Trends and Composition at Alert, Canada: 1980–1995, *J. Geophys. Res.*, 104D, 11 599–11 618.

Wania, F and Mackay, D (1996) Tracking the Distribution of Persistent Organic Pollutants, *Environ. Sci. Technol.*, 30, A390–A396.

Arctic Climate

Gunter Weller

University of Alaska Fairbanks, Fairbanks, AK, USA

The climate of the Arctic has produced one of the most inhospitable and extreme environments on Earth, characterized by limited sunlight for much of the year, extreme temperature excursions and a short growing season. The Arctic is often de ned by climatic parameters, e.g., as the area where the average temperature for the warmest month is below 10 °C (Köppen, 1936). Sea ice, snow cover, glaciers, tundra, permafrost, boreal forests and peatlands are expressions of this severe climate, as well as being sensitive indicators of climatic change. Their presence and extent are susceptible to subtle variations in sunlight, surface temperature, and heat transport through the atmosphere and ocean.

The Arctic also plays an important part in the complex interactions of the Earth System. The mean global circulation patterns of the atmosphere and ocean are controlled by equator- pole temperature differences. Polar feedback processes affect the global climate and global warming due to the greenhouse effect, amplifying this effect at high latitudes. Greenhouse warming in the Arctic is likely to have major impacts on the environment, including melting of sea ice and glaciers, and thawing of permafrost, with potentially profound societal impacts, not only regionally but also around the globe. The Arctic climate is thus governed by many complex interactions, which are part of the global climate system, and in turn climatic conditions shape much of the Arctic environment.

THE CURRENT ARCTIC CLIMATE

Temperature and Precipitation

The Arctic climate varies greatly by location and season. Sea ice, snow cover, glaciers, tundra, permafrost, boreal forests and peatlands (Figure 1) are expressions of this severe climate, as well as being sensitive indicators of climatic change. Mean annual surface temperatures range

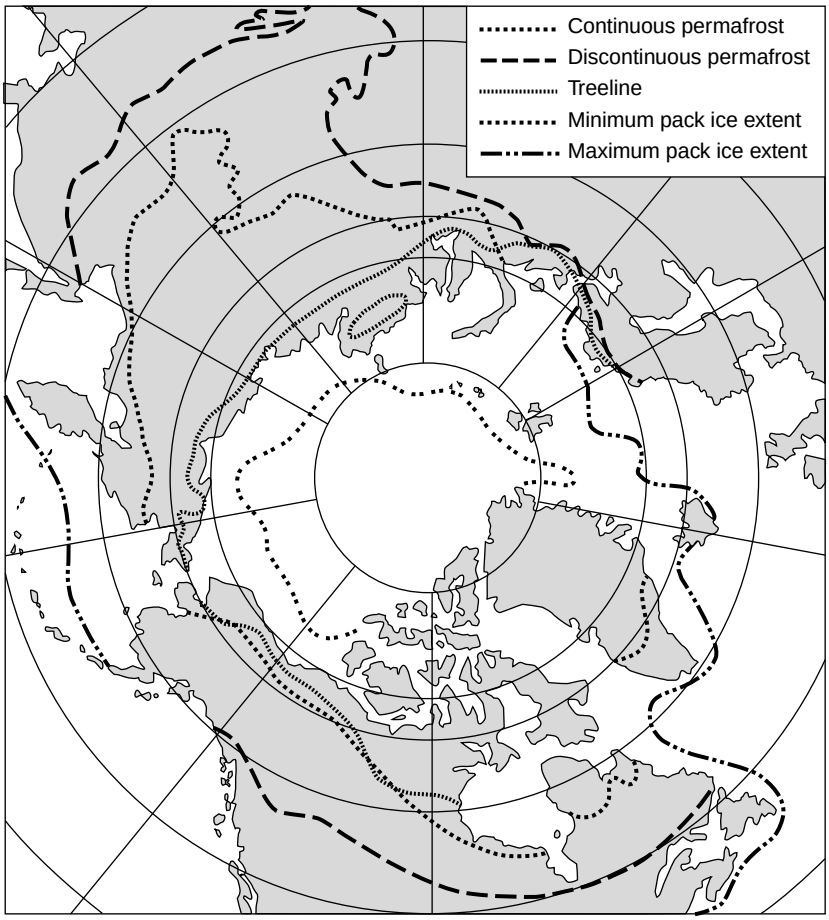


Figure 1 Climate-related features of the Arctic and sub-Arctic, including continuous and discontinuous permafrost boundaries and the tree-line on land, and maximum and minimum pack ice extent in the ocean

from 0 °C at Murmansk, Russia (69 °N) to −12.2 °C at Point Barrow, Alaska (71.3 °N), −16.2 °C at Resolute, Canada (74.7 °N), −18 °C over the central Arctic Ocean to −28.1 °C at the crest of the Greenland ice sheet (about 71 °N). Some of these differences are due to the poleward intrusion of warm ocean currents such as the Gulf Stream and the southward extension of cold air masses. Arctic tundra areas in North America have at least a 50% frequency of being in Arctic air in July and the median location of the polar front in July corresponds approximately with the northern limit of the boreal forest (Barry, 1967).

Precipitation in the Arctic is difficult to measure, because it is generally light and is associated with storms and also because for the greater part of the year it falls as dry snow, which is redistributed by winds according to exposure and local topography. With an annual precipitation of 200–300 mm and frequently less than 100 mm (e.g., 104 mm year^{-1} at Barrow, Alaska; 130 mm year^{-1} at Resolute, and 95 mm^{-1} measured in the central Arctic Ocean on Soviet ice islands (NP 7–8)), the Arctic is comparable to arid regions elsewhere.

Much of the Arctic consists of cold deserts largely devoid of vegetation. It should be noted, however, that climatically arid regions can have wet surfaces when permafrost prevents drainage of the seasonally thawed layer, even though the precipitation may be low.

Seasonal Variability

Winter in the Arctic tundra is 6–9 months long and is characterized by a relatively shallow (30–40 cm) snow cover, by darkness, and by January temperatures that average about −30 °C in North America, somewhat lower in Asia (Siberia) and much higher (−10 °C) in Northern Europe. Spring and autumn are short transition periods, respectively. Late winter and spring are generally characterized by clear skies and receipt of a high percentage of the possible solar radiation. Temperatures begin to rise, but lag 4–6 weeks behind the increase in solar radiation. Mean daily temperatures are above freezing during the short summer, but over the Arctic Ocean and the coastal areas cloudiness and fog prevent temperatures from rising much above 5 °C even in

July. Temperatures increase from the coast inland to the forest-tundra ecotone, which can have July mean temperatures of 10 °C and higher.

Atmospheric Circulation

Surface weather systems associated with a large-scale cold-cored westerly circumpolar circulation are distinctive of the polar atmosphere. Cyclones and anticyclones are embedded in, and are steered by, this flow, and changes in this circulation can radically affect the weather and climate of the planet. The synoptic (large-scale instantaneous state of the atmosphere) activity in the Arctic between 1952 and 1989 has been described by Serreze *et al.* (1993). They found that winter cyclone activity is most common near Iceland, between Svalbard and Scandinavia, the Norwegian and Kara Seas, Baffin Bay and the Eastern Canadian Arctic. Cyclone tracks in winter most frequently enter the Arctic from the Norwegian and Barents Seas. Winter anticyclones are most frequent and strongest over Siberia and Alaska/Yukon, with additional but weaker systems over the Central Arctic Ocean and Greenland. During summer, cyclones are common in the same regions as in winter.

Since 1952, cyclone numbers in winter, spring and summer have increased significantly, as have anticyclone numbers in spring, summer and fall (Serreze *et al.*, 1993). Proshutinsky and Johnson (1997) have shown that the atmospheric circulation over the Arctic Ocean alternates between typical anticyclonic and cyclonic circulation regimes approximately every 10–15 years. This effect has been named the Arctic Oscillation (*see* Arctic Oscillation, Volume 1) and has important implications for the Arctic s thermohaline (temperature and salinity) structure, salinity anomalies observed in the Greenland Sea, and the variability of sea ice in the Arctic Ocean.

ARCTIC MICROCLIMATES

Arctic Tundra Microclimates

Arctic tundra covers about 5.5% of the land surface of the Earth and constitutes a major surface feature of the Arctic. The word *tundra* is broadly used to refer to the landscapes that are found above the altitudinal or latitudinal treeline (*see* Tundra, Volume 2). The boundary between tundra and the northern edge of the boreal forests, the forest-tundra *ecotone* (*see* Ecotones, Volume 2), is not very well defined. The most common description of the tundra is treelessness, and trees may be defined in this context as woody plants with a single central stem at least 2 m tall. Of the world s major ecosystems, tundra has the lowest temperatures and the shortest growing season. The tundra plants are adapted to the cold winters and chilly summers of Arctic and alpine regions. For these species the environment is not severe either physically or biologically.

The environmental conditions within a few meters above and below the ground surface, where most biological activities take place, constitute the microclimate of a region. The microclimate is characterized by the radiation, temperature, and moisture regimes of these near surface layers. It is determined largely by the regional climate, as modified by local topography and by the vegetation cover. In the Arctic tundra environment with its low and exposed vegetation, large variations in microclimate may exist within a small area. Even a few centimeters difference in microtopography can produce marked differences in soil temperature, depth of snow, wind effects, snow drifting and resultant protection of plants.

Energy Balance and Surface Conditions

Quantitatively, the microclimate is described by the energy balance, which expresses how the available net radiative energy is converted into sensible and latent heat to warm the air and ground and to melt snow or evaporate water from the surface. The microclimate of the Arctic coastal tundra is characterized by an 8-month winter phase with net radiative losses, compensated primarily by sensible heat fluxes towards the surface, in the presence of steep temperature inversions (Figure 2). This is followed by a brief, but distinct, phase related to snow melting. Towards the end of this phase, the tundra is completely wet and evaporation rates reach 6 mm day^{-1}, with mean values of 4.5 mm day^{-1}. Through the melting phase, within a span of 2–3 weeks, the net radiation increases by a factor of 10 and the latent heat flux by a factor of 40. Evaporation continues throughout summer, but by the time freezing of the tundra occurs in early September, net radiation has been reduced appreciably and evaporation rates are down to about 1 mm day^{-1}. Soon thereafter, snow covers the tundra and the radiation balance becomes negative again.

PAST CLIMATES

Climate Variability

Historical climate records generally do not go back more than a few hundred years but past climates can be reconstructed from many different proxy indicators, including tree rings (1000 years), ice cores (100 000 years), lake sediments (1 million years), and marine sediments (10 million years), among others. From these indicators, the climate of the Arctic is known to have undergone large variations in the past. The Little Ice Age and Medieval Warm Period were the most recent examples of cooler and

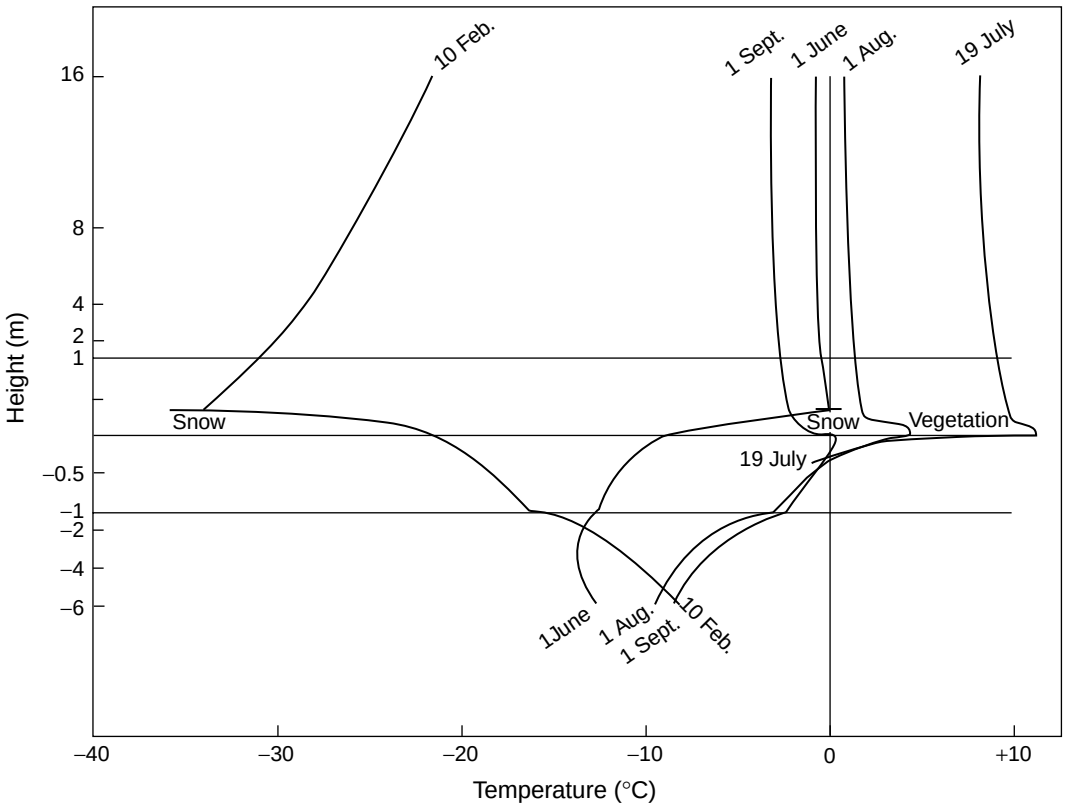


Figure 2 Typical temperature gradients in air, snow, vegetation and soil at Barrow, Alaska. The vertical height scale from +1 to −1 m is expanded. Note the effects of snow and vegetation (from Weller and Holmgren, 1974)

warmer climates, respectively. The North Atlantic region was unusually mild when the Norse first settled Greenland in the 10th century AD, then plunged into a 500-year cold spell known as the Little Ice Age, starting about 1300 AD (*see* Little Ice Age, Volume 1; Medieval Climatic Optimum, Volume 1). By 1500, archaeological and historical evidence shows that the Norse settlers had vanished. The changing climate was undoubtedly a factor and illustrates the potential impacts of climate change on humans and human activities.

The Ice Core Record

Ice ages have occurred at roughly 100 000 year intervals during the Pleistocene. Indications are that they are caused by changing amounts and distribution of sunlight on the planet due to long-term variations in the Earth s orbit and the inclination of its spin axis to the Sun (the so-called Milankovich cycles after the Yugoslav mathematician who computed them 75 years ago; *see* Climate Model Simulations of the Geological Past, Volume 1; Milankovitch, Milutin, Volume 1). During ice ages, large ice sheets covered parts of the Northern Hemisphere continents and the Bering Strait was above sea level. Recent results obtained from the deep ice cores recovered by European and American researchers from the Greenland ice sheet have produced remarkable new insights into short- and long-term climatic changes, with the abrupt changes revealed by the ice cores being a particular surprise.

The ice core record shows great detail of past climatic variability. At the end of the last ice age, about 13 000 years ago, the climate was beginning to warm during a period called the B lling–Aller d when it suddenly plunged back to ice age conditions. This 1300 year long cold period is named the Younger Dryas because the polar wildflower *Dryas octopetala* had a resurgence in Europe during this period. Temperatures in Greenland dropped by about 7 °C back to full ice age conditions. At the end of the Younger Dryas the climate returned to warmer and wetter interglacial conditions. Further back, during the last ice age, there are unexpected abrupt climate shifts. These oscillations, called *Dansgaard- Oescheger cycles*, (*see* Dansgaard- Oescheger Cycles, Volume 1) lasted from centuries to millennia, jumping abruptly from cold to warm climates before slowly reverting to cold conditions again. Many of these oscillations seem to be associated with the collapse of big ice sheets, which sent numerous icebergs into the North Atlantic, the so-called Heinrich events.

PRESENT AND FUTURE CLIMATE TRENDS

Observed Climate Trends in the Arctic

Chapman and Walsh (1993) examined the climate trends in the Arctic for the period 1961–1990, using the climate data set of the Climate Research Unit of the University of East Anglia. The results, including updated data for 1966–1995, indicate considerable warming over the landmasses of Eurasia and North America, particularly in winter and spring. Over the last three decades trends have been towards higher surface temperatures in these areas at rates of up to $1\,^{\circ}\mathrm{C}\,\mathrm{decade}^{-1}$ in the annual mean and up to $2\,^{\circ}\mathrm{C}\,\mathrm{decade}^{-1}$ in winter. On the other hand, there are also smaller areas of cooling of similar magnitude within the Arctic regions, particularly in the South Greenland and Davis Strait area. There is little change in summer and slight cooling in the fall for most areas.

Pronounced reductions in seasonal snow, glaciers, permafrost and sea ice, following a warmer climate in the Arctic, have also been observed. Some of the changes include:

- Sea ice extent in the Bering Sea has been reduced by about 5% over since the 1950s, with the steepest decrease occurring in the late 1970s. Sea ice extent has also decreased in the East Siberian Sea.
- Sea ice thickness, a sensitive indicator of climate change, seems to have decreased between 1976 and 1987, based on limited submarine sonar records.
- Glaciers have generally receded, with typical ice thickness decreases of 10 m since the 1950s, but some glaciers have thickened in their upper regions. A warming of $1\,^{\circ}\mathrm{C}$, if sustained, appears to reduce glacier lengths by about 15%.
- Borehole measurements in continuous permafrost have shown warming of up to $2–4\,^{\circ}\mathrm{C}$ over the last century.
- Discontinuous permafrost throughout Alaska has warmed, and some of it is currently thawing from the top and bottom.
- Cyclone and anticyclone frequency have increased over the Arctic between 1950–1990.
- Annual snowfall has increased in the same period over northern Canada (North of $55\,^{\circ}\mathrm{N}$) by about 20% and by about 11% over Alaska.

Modeling Future Climates

The future climate of the Arctic may be strongly affected by the increasing greenhouse effect. Numerous general circulation models (GCMs) have attempted to simulate the global climate for this effect. All of these models show a temperature amplification in the annual mean in the Arctic; the warming is even more pronounced in winter and spring. These results are the basis of the World Meteorological Organization (WMO)/United Nations Environmental Programme (UNEP) Intergovernmental Panel on Climate Change (IPCC) (IPCC, 1996) projections that the Arctic will warm more than the global mean, particularly in winter. The IPCC report also projects that Arctic landmasses will warm more rapidly than the ocean, although some models show the greatest temperature increases over the Arctic Ocean.

A comparison of GCM performances in the Arctic as part of the Atmospheric Model Intercomparison Project (AMIP) shows a wide divergence of results, however. The temperature simulations for the seasonal mean surface air temperatures in the Arctic Ocean of 19 GCMs were compared with observations. The majority of the models produced temperatures that were too cold. Simulation of other climate parameters also showed large differences. Sea level pressure was well simulated by several models, particularly the ones with higher resolution. Precipitation was too large, in some cases by more than a factor of two. There were also wide variations in cloud cover. The reliability of the simulated climate change scenarios is likely not to be high, but model performance is improving.

IMPACTS OF CLIMATE CHANGE

The IPCC Report

The IPCC report *Climate Change 1995* (IPCC, 1996) includes a chapter (Chapter 7 of Working Group II on the cryosphere) on climate change and its impacts on the polar regions for a simulated doubling of atmospheric CO_2. The report provides the following projections, including the degree of confidence in each case:

- Many components of the cryosphere are sensitive to changes in atmospheric temperature because of their thermal proximity to melting. The extent of glaciers has often been used as an indicator of past global temperatures (High Confidence).
- Projected warming of the climate will reduce the area and volume of the cryosphere. This reduction will have significant impacts on related ecosystems, associated people and their livelihoods (High Confidence).
- There will be striking changes in the landscapes of many high mountain ranges and of lands at northern high latitudes (High Confidence). These changes may be exacerbated where they are accompanied by growing numbers of people and increased economic activities (Medium Confidence).

The chapter goes on to state that the following changes and associated impacts on the polar regions are likely:

- Pronounced reductions in seasonal snow, permafrost, glacier and periglacial features with a corresponding shift in landscape processes (High Confidence).

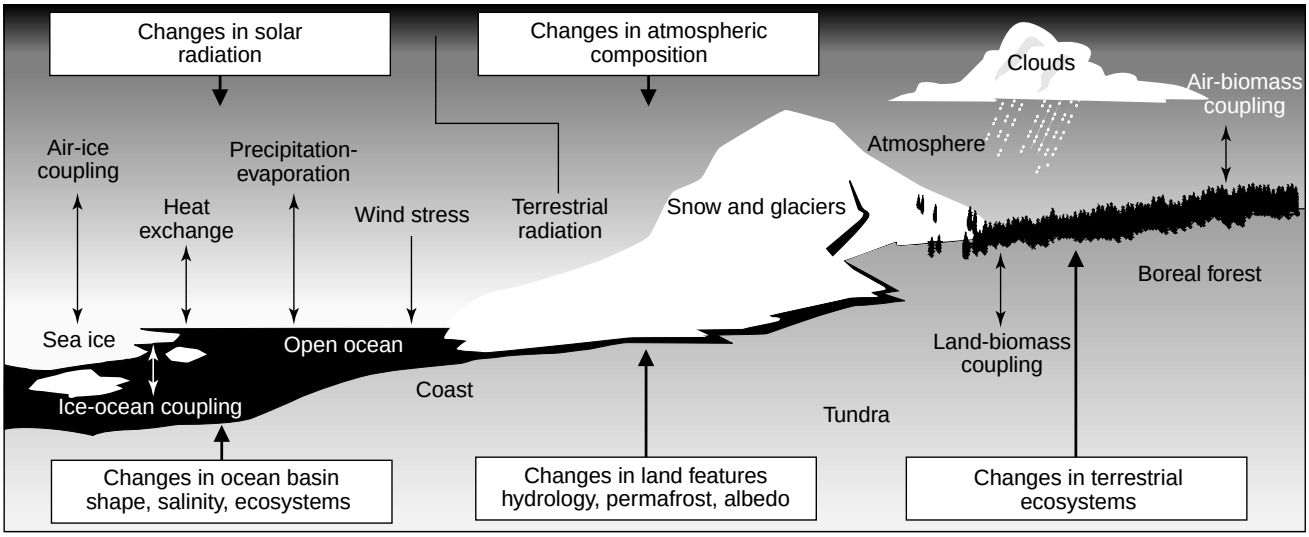


Figure 3 A diagrammatic transect across the Arctic landscape, showing the principal external forcing functions of change (top boxes), the Arctic internal interactions (middle arrows), and the effects on the Arctic (bottom boxes). (Illustration used by permission of the Arctic Research Consortium of the United States)

- Increases in the thickness of the active layer of permafrost and the disappearance of most of the ice rich discontinuous permafrost over a century-long time span (High Confidence).
- Disappearance of up to a quarter of the presently existing mountain glacier mass (Medium Confidence).
- Less ice on rivers and lakes. Freeze-up dates will be delayed, and break-up will begin earlier. The river ice season could be shortened by up to a month (Medium Confidence).
- A large change in the extent and thickness of sea ice, not only from warming but also from changes in circulation patterns of both atmosphere and oceans. There is likely to be substantially less sea ice in the polar oceans (Medium Confidence).

Figure 3 shows a schematic diagram of the causes, effects, interactions and feedbacks of climate change involving the various components of the Arctic environment.

Economic Impact of Climate Change

The economical impacts of a major climate change in the Arctic regions are still largely unknown but some impacts in recent decades have been summarized by Weller and Lange (1999). As the climate continues to get warmer, widespread loss of discontinuous permafrost will trigger erosion or subsidence of ice rich landscapes. Permafrost thawing will reduce slope stability and increase the incidence of natural hazards for people, structures and communication links. Buildings, roads and pipelines will be threatened. Thawing of permafrost could also lead to disruption of petroleum production and distribution systems in the tundra, unless mitigation techniques are adopted. Engineering and agricultural practices will need to adjust to changes in snow, ice and permafrost distributions. Arctic fisheries and forestry may also be affected.

Not all of the impacts are likely to be negative. For example, improved opportunities for water transport, tourism and trade are expected from a reduction in sea, river, and lake ice. Reduced sea ice may also aid new exploration and production of oil in the Arctic Basin. Agriculture and forestry may benefit from a warmer climate. These will have important implications for the people and economies of the Arctic.

See also: Arctic Air Quality, Volume 1; Arctic Ocean, Volume 1; Arctic Oscillation, Volume 1.

ACKNOWLEDGMENTS

This article is based in part on a paper by the author entitled *The Weather and Climate of the Arctic* (Weller, 1999), with permission from the publishers, Harwood Academic Press.

REFERENCES

Barry, R G (1967) Seasonal Location of the Arctic Front in North America, *Geogr. Bull.*, 9, 79–95.

Chapman, W L and Walsh, J E (1993) Recent Variations of Sea Ice and Air Temperatures in High Latitudes, *Bull. Am. Meteorol. Soc.*, 74(1), 33–47.

IPCC (1996) *Climate Change 1995. Impacts, Adaptations and Mitigation of Climate Change: Scienti c-technical Analysis Contributions of Working Group II to the Second Assessment*

Report of the Intergovernmental Panel on Climate Change, eds R T Watson, M C Zinyowera, R H Moss, and D J Dokken, WMO-UNEP, Geneva, Cambridge University Press, New York.

Koppen, W (1936) Das geographische System der Klimate, in *Handbuch der Klimatologie*, eds W Koppen and R Geiger, Berlin.

Proshutinsky, A and Johnson, M (1997) Two Circulation Regimes of the Wind-driven Arctic Ocean, *J. Geophys. Res.*, 102(12), 493–514.

Serreze, M C, Box, J E, Barry, R G, and Walsh, J E (1993) Characteristics of Arctic Synoptic Activity, 1952–1989, *Meteorol. Atmos. Phys.*, 51, 147–164.

Weller, G and Holmgren, B (1974) The Microclimates of the Arctic Tundra, *J. Appl. Meteorol.*, 13, 854–862.

Weller, G and Lange, M (1999) Impacts of Global Climate Change in the Arctic Regions, Workshop on the Impacts of Global Change, 25–26 April 1999, Troms , Norway, Published for International Arctic Science Committee by Center for Global Change and Arctic System Research, University of Alaska Fairbanks.

Weller, G (1999) The Weather and Climate of the Arctic, in *The Arctic - A Guide to Research in the Natural and Social Sciences*, eds M Nuttall and T V Callaghan, Harwood Academic Press, Reading, 143–160.

Arctic Climate System Study (ACSYS)

see ACSYS (Arctic Climate System Study) (**Volume 1**)

Arctic Ocean

Claire L Parkinson
NASA Goddard Space Flight Center, Greenbelt, MD, USA

The Arctic Ocean is the smallest of the Earth's four major oceans (the Paci c, Atlantic, Indian, and Arctic), covering 14 $\times$ 10^6 km^2 located entirely within the Arctic Circle (66° 33′N). It is a major player in the climate of the north polar region, has a variable sea ice cover, and is home to a multitude of plant and animal species. Its temperature, salinity, and ice cover have all undergone changes in the past several decades, although it is uncertain whether these predominantly re ect long-term trends, oscillations within the system, or natural variability.

The Arctic Ocean is surrounded largely by the land masses of Eurasia, Greenland, and North America (Figure 1). Its principal connection to the rest of the Earth s oceans lies between Greenland and Scandinavia, where it connects to the North Atlantic. Smaller connections include the narrow (85 km wide) Bering Strait, linking it to the North Pacific, and the passageways within the Canadian Archipelago and between the Archipelago and Greenland, leading to Baffin Bay and thence to Davis Strait and the North Atlantic.

The Arctic Ocean has an unusually broad and shallow continental shelf on the Eurasian side, extending more than 1000 km northward from Scandinavia and approximately 800 km northward from Siberia. The deeper portion of the Arctic is divided by the Lomonosov Ridge into two main basins, the Canadian Basin and the smaller and deeper Eurasian Basin, which has a maximum depth exceeding 5000 m.

Water flows into the Arctic principally from the Atlantic as a warm, salty undercurrent. There are also smaller oceanic inputs through the Bering Strait and cold, freshwater inputs from many rivers, most significantly the Lena, Yenisei, and Ob rivers in Russia and the Mackenzie River in Canada. A large part of the outflow from the Arctic is through the Fram Strait between Greenland and Svalbard. Surface currents in the Arctic tend to be clockwise in the Canadian Basin, with occasional reversals to this flow, and more linear along the Transpolar Drift Stream flowing from north of Russia, across the Eurasian Basin and the vicinity of the North Pole, and out through Fram Strait (*see* Ocean Circulation, Volume 1).

The water inflows and outflows play a major role in the temperature and salinity structure of the Arctic (*see* Salinity Patterns in the Ocean, Volume 1). The warm, salty Atlantic layer from the Atlantic inflows is most prominent closest to the entrance regions but is also apparent throughout the Arctic at depths exceeding 200 m. Overlying the Atlantic layer, the Arctic surface water, in contact with

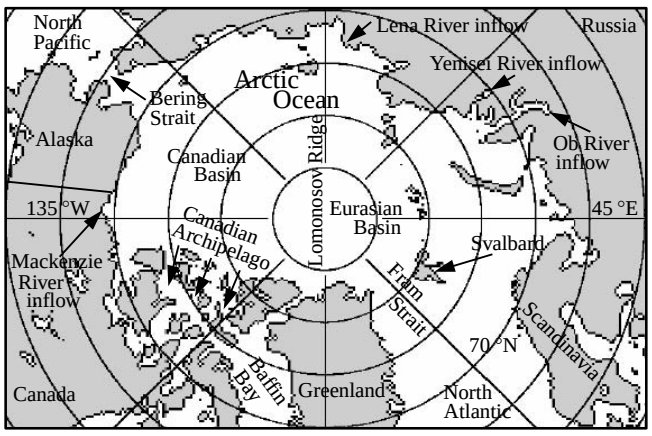


Figure 1 The Arctic Ocean and its surroundings

the cold Arctic atmosphere and subject to the freshwater inputs from the surrounding rivers, is colder and less saline. The upper 30–50 m of the surface water tends to be fairly well mixed vertically, with temperatures near the freezing point and salinities ranging from highs exceeding 34 parts per thousand (ppt) near the North Atlantic to lows below 29 ppt near river inflows. Surface salinities in the Bering Strait are approximately 31 ppt. Vertically, salinities tend to increase with depth from the bottom of the mixed layer down to the Atlantic layer, with this vertical variation forming a prominent halocline, especially in the Eurasian Basin. The resulting stable density stratification hinders the warm Atlantic layer waters from upwelling to the surface.

Importantly, the Arctic Ocean is largely capped by a thin, broken layer of sea ice, generally less than 6 m thick and covered by snow (*see* Sea Ice, Volume 1). The sea ice restricts exchanges of heat, mass, and momentum between the ocean and the overlying atmosphere and, due to its high reflectivity, also tremendously restricts the input of solar radiation to the ocean. Ice covers almost all of the Arctic Ocean in winter, to an ice concentration (percent areal coverage) of at least 90%, and most of the Arctic Ocean in summer. Sea ice is considerably less saline than the ocean water from which it forms and tends to decrease in salinity over time, as more of the salt content is washed downward through the ice during periods of summer melt.

The Arctic ice cover is in constant flux, being melted by solar radiation, augmented by additional freezing, and moved by winds, waves, and currents. As ice floes separate, openings appear, called *leads* when linear and *polynyas* when large and non-linear. In contrast, when the forces acting on the ice result in floes colliding forcefully together, the ice breaks and piles of ice rubble form. The above-water portions of these are called *ridges*, and the more massive underwater portions are called *keels*. A ridge/keel combination can have an ice thickness of 30 m or more, far exceeding the level-ice thickness.

Despite the cold temperatures, the Arctic is home to a host of plant and animal life, including algae colonizing the sea ice, sometimes at a concentration of millions in a single ice floe, and protozoans, crustaceans, and nematodes, most of which are smaller than 1 mm in length, also living in the ice and feasting upon the algae. Although biomass tends to be low under the permanent ice pack, high phytoplankton and zooplankton concentrations are frequently found in the ice-free waters. The resulting availability of food makes the ice-free waters popular for numerous species of birds and marine mammals. Amongst the larger animals, polar bears and Arctic foxes roam over the ice, and seals, walruses, and whales live in the ocean waters.

The Arctic has received considerable attention during the late 20th century and the start of the 21st century because of various changes reported to be occurring in it and the sense that these could be related to a possible global warming. Among the changes are the following:

1. A warming and spatial expansion of the Atlantic layer, at depths of 200–900 m, determined from ship-based conductivity–temperature–density (CTD) measurements in the 1990s versus data from 1950–1989 (Morison *et al.*, 2000; Serreze *et al.*, 2000).
2. A warming of the upper ocean in the Arctic s Beaufort Sea (north of Alaska) from 1975 to 1997, found from *in situ* measurements (McPhee *et al.*, 1998).
3. A considerable thinning, perhaps as high as 40%, of the Arctic sea ice cover in the second half of the 20th century, found from submarine data (Rothrock *et al.*, 1999).
4. A lesser and uneven retreat of the ice cover, averaging approximately 3% per decade between late 1978 and the end of 1996, found from satellite data (Bjørgo *et al.*, 1997; Parkinson *et al.*, 1999), and a related shortening of the length of the sea ice season throughout much of the region of the Arctic Ocean s seasonal sea ice cover, also found from satellite data (Parkinson, 2000).
5. An increase of 5.3 days per decade in the length of the melt season on the perennial ice cover, found from satellite data for 1979–1996 (Smith, 1998).
6. A decrease in the salinity of the upper 30 m of the central Beaufort Sea from 1975 to 1997, found from *in situ* measurements. This freshening of the water has been attributed largely to sea ice melt (McPhee *et al.*, 1998) and to increased runoff from the Mackenzie River (Macdonald *et al.*, 1999).
7. A mixed pattern of salinity increases and decreases through the expanse and depth of the rest of the Arctic Ocean (Morison *et al.*, 2000). This includes an increase in the salinity of the surface waters in the mid-Eurasian Basin during the 1990s, found from submarine data, and a thinning of the halocline separating the surface from the warm Atlantic layer waters (Steele and Boyd, 1998; Morison *et al.*, 2000).

In view of the highly coupled nature of the Arctic climate system, many of the changes occurring within it are likely connected. In particular, the freshening of the upper ocean in the Beaufort Sea is likely a response in part to the thinning of the ice, as ice melt adds freshwater to the upper ocean which consistently is much less saline than the ocean average. Similarly, the reduction in the sea ice cover and warming of the upper ocean are probably both connected to the Arctic surface air temperature increases reported, for instance, by Serreze *et al.* (2000).

The causes of the late-20th century changes in the Arctic system remain uncertain, although it is likely that several factors are involved. The increase in carbon dioxide and

other greenhouse gases in the global atmosphere, is believed to cause atmospheric warming, although some human influences, such as the increase in particulate matter in the atmosphere, tend to offset a portion of the warming (*see* Arctic Climate, Volume 1). Atmospheric warming contributes to oceanic warming, sea ice melt, and upper ocean freshening, all observed in recent decades in portions of the Arctic.

Other potential influences, however, are more oscillatory in nature, such as the impacts of the North Atlantic Oscillation (NAO) and the Arctic Oscillation (AO), two major decadal-scale oscillations in atmospheric pressure patterns, or the impacts of El Niño/La Niña cycles. The NAO and AO in particular have received attention because many of the patterns of change in the ocean and ice cover of the Arctic can be explained by changes in the NAO and AO in recent decades (e.g., Parkinson *et al.*, 1999; Morison *et al.*, 2000). It remains uncertain, however, whether the changes in the NAO and AO are exclusively natural fluctuations in the climate system or are related to long-term, perhaps anthropogenically induced climate change (*see* Arctic Oscillation, Volume 1; El Niño, Volume 1; North Atlantic Oscillation, Volume 1).

REFERENCES

Bjørgo, E, Johannessen, O M, and Miles, M W (1997) Analysis of Merged SMMR-SSMI Time Series of Arctic and Antarctic Sea Ice Parameters 1978–1995, *Geophys. Res. Lett.*, 24(4), 413–416.

Macdonald, R W, Carmack, E C, McLaughlin, F A, Falkner, K K, and Swift, J H (1999) Connections among Ice, Runoff and Atmospheric Forcing in the Beaufort Gyre, *Geophys. Res. Lett.*, 26(15), 2223–2226.

McPhee, M G, Stanton, T P, Morison, J H, and Martinson, D G (1998) Freshening of the Upper Ocean in the Arctic: Is Perennial Sea Ice Disappearing? *Geophys. Res. Lett.*, 25(10), 1729–1732.

Morison, J, Aagaard, K, and Steele, M (2000) Recent Environmental Changes in the Arctic: A Review, *Arctic*, 53(4), 359–371.

Parkinson, C L (2000) Variability of Arctic Sea Ice: The View from Space, an 18-Year Record, *Arctic*, 53(4), 341–358.

Parkinson, C L, Cavalieri, D J, Gloersen, P, Zwally, H J, and Comiso, J C (1999) Arctic Sea Ice Extents, Areas, and Trends, 1978–1996, *J. Geophys. Res.*, 104(C9), 20 837–20 856.

Rothrock, D A, Yu, Y, and Maykut, G A (1999) Thinning of the Arctic Sea-ice Cover, *Geophys. Res. Lett.*, 26(23), 3469–3472.

Serreze, M C, Walsh, J E, Chapin, III, F S, Osterkamp, T, Dyurgerov, M, Romanov, V, Oechel, W C, Morison, J, Zhang, T, and Barry, R G (2000) Observational Evidence of Recent Change in the Northern High-latitude Environment, *Clim. Change*, 46, 159–207.

Smith, D M (1998) Recent Increase in the Length of the Melt Season of Perennial Arctic Sea Ice, *Geophys. Res. Lett.*, 25(5), 655–658.

Steele, M and Boyd, T (1998) Retreat of the Cold Halocline Layer in the Arctic Ocean, *J. Geophys. Res.*, 103(C5), 10 419–10 435.

Arctic Oscillation

Drew Shindell

Columbia University, New York, NY, USA

The Arctic Oscillation (AO) is a hemispheric scale seesaw of heat and mass between the Arctic region and Northern Hemisphere middle latitudes. This pattern dominates the variability in the Northern Hemisphere winter, and has the second largest impact of all the Earth's climate cycles (after the El Niño/La Niña cycle). The strength of the AO pattern is closely associated with surface wind speeds and temperatures in the Northern Hemisphere. Variability induced by the AO is especially important in the wintertime, when the oceans are warmer than the land because they have retained more heat from the previous summer. Increased winds then carry relatively warm Atlantic oceanic air across northern Europe and Asia, causing a local rise in temperature. Similarly, onshore winds from the Paci c warm North America, while surface temperatures fall in the vicinity of the Bering Strait and Newfoundland due to the offshore ow of colder continental air. A bias in the AO has been observed since the early 1970s, and is associated with roughly half of the observed wintertime surface warming over Europe and Siberia, and at least part of the observed cooling in the vicinity of Newfoundland and the Labrador Sea.

The AO, sometimes called the *Northern Annular Mode*, is created by internal atmospheric interactions of eddy momentum and heat fluxes with the mean flow. These affect temperatures, pressures, and atmospheric motions, all of which exhibit AO-related variability. The most common definition of the AO is based on *Empirical Orthogonal Functions* (EOFs). These are mathematical representations of a data set by a number of fixed spatial patterns, each derived to explain the maximum amount of variance possible, while remaining independent of the other patterns. The AO pattern is the first EOF of sea-level pressure (Figure 1), and therefore is the pattern that explains the most variance. It accounts for roughly 1/4 of the Northern Hemisphere variability during winter, more than twice that of the next pattern. The variability associated with this pattern extends from the surface up into the stratosphere during winter, and the variability pattern can be equivalently defined as being composed of variability at all levels

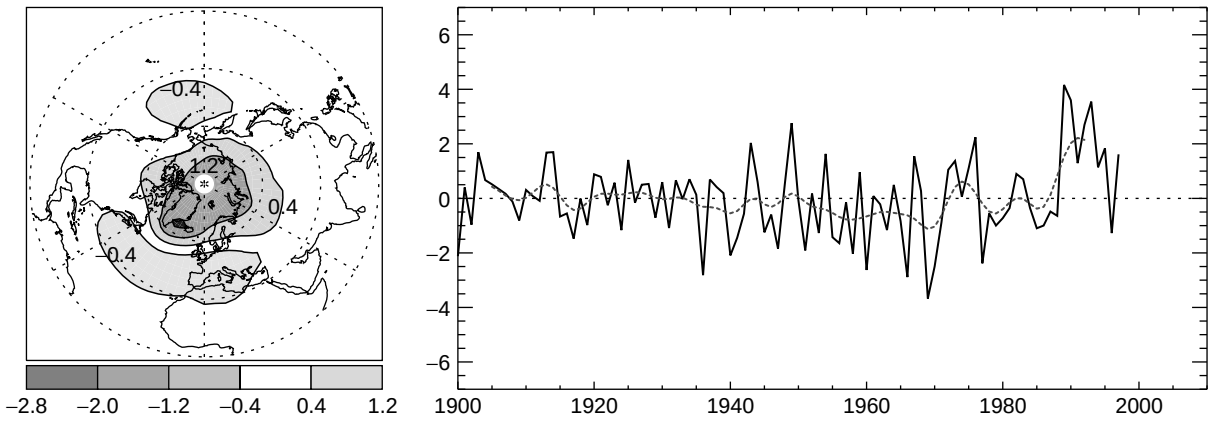


Figure 1 The AO and the AO index. The left side shows the leading variability pattern of Northern Hemisphere wintertime sea level pressure (arbitrary sign), while the right side shows the amplitude of this pattern over time. The index is scaled to units of pressure (mbar or hPa), while the spatial pattern has unit amplitude poleward of 60°N. The dashed line is a 10-year smoothed fit to the index. (Reproduced by permission of *Nature* (Shindell *et al.,* 1999) © (1999) Macmillian Magazines Limited)

from the surface through the lower stratosphere. A particular AO pattern has a typical persistence time of about 1–2 weeks.

The time series of the size of the first EOF is often called the *AO index* (Figure 1). The classic North Atlantic Oscillation contrast between subtropical and high-latitude sea-level pressure is the subset of the broader AO which appears in the Atlantic region (*see* North Atlantic Oscillation, Volume 1). The AO variability pattern is present throughout the year, but it is largest in winter, when its amplitude actually increases with altitude. The AO also shows significant wintertime correlations in the lower stratosphere with the Quasi-Biennial Oscillation (*see* Quasi- Biennial Oscillation (QBO), Volume 1). Similar variability is present in the Southern Hemisphere, showing almost identical patterns in sea-level pressure, zonal winds, and temperatures. It has been called the *Antarctic Oscillation* (AAO) or *Southern Annular Mode*.

During the high phase of the AO, sea-level pressure is lowered in the Arctic as the atmosphere at high latitudes cools, while at middle latitudes it warms. Consistent with these temperature changes, the speed of the westerlies blowing at Northern Hemisphere mid-latitudes increases. The prevailing westerly winds are caused by the natural difference in temperature between the colder polar region and the warmer lower latitudes, so that an increase in the temperature gradient with latitude strengthens the winds. The high phase of the AO therefore weakens the subtropical jet and strengthens the subpolar vortex jet. The laws of physics governing the atmosphere require consistency between changes in motion, pressure and temperature, so that the AO can be equally well described by the variability in westerly wind speed or in the temperature, mass, or sea-level pressure patterns. In addition to its effects on zonal winds, the AO also modulates the mean meridional circulation and hence the total column ozone, as well as tropopause height.

Observations show that the AO index remained relatively constant from around 1900 to 1970, perhaps even decreasing very slightly. Beginning in the 1970s, it began to increase. This apparent trend in the AO accounts for slightly more than half the total observed trend in sea-level pressure, and nearly the entire trend towards increasing subpolar westerly winds. Because the strength of the AO affects continental surface temperatures, this is associated with a significant portion of the observed wintertime warming. This warming has been 4–6 °C (~7–11 °F) over large areas of the Northern Hemisphere continents from the mid-1960s to the mid-1990s. During this period, the AO contribution to the Siberian warming has been +3–4 °C (~5–7 °F), while it has contributed roughly 2–4 °C (~4–7 °F) cooling to the Newfoundland/Labrador Sea area (this cooling may also be related to oceanic circulation changes). Over the same period, globally and annually averaged surface temperatures have increased by only about 0.5 °C (0.9 °F).

It is not yet certain whether the observed high values of the AO index will continue, or what is causing them. There is clear evidence from observations that the AO pattern at the ground is closely related to the wind pattern in the stratosphere, however (*see* Stratosphere, Temperature and Circulation, Volume 1). Observations also suggest that changes in the two coupled patterns typically originate in the stratosphere, and then propagate downwards to the Earth s surface. Changes in the stratosphere may therefore play a large role in the apparent AO trend. Possible causes for the initial changes in the stratosphere are anthropogenic greenhouse gas emissions, Arctic

ozone depletion, or solar variability. Climate model simulations show that it is possible for all of these to alter the latitudinal temperature gradient in the stratosphere. In the case of increasing greenhouse gases or ozone depletion, this leads to a strengthening of the westerly jet stream in the lower stratosphere, and hence of the Arctic polar vortex. The effect works its way down as the enhanced stratospheric winds alter the propagation of energy coming out of the troposphere, which then changes the wind speed at progressively lower altitudes. Observations and model simulations both show a distinct shift in energy flow over the past few decades, supporting the mechanism outlined above. Furthermore, a trend has been observed in the AO only during wintertime, when variability in the lower atmosphere is strongly coupled to the stratosphere by planetary waves, and is therefore the only season in which this mechanism is effective.

If in fact the increased amplitude of the AO index is caused by human activities, then it will likely continue to increase. Furthermore, a sizeable fraction of the large wintertime continental warming trends should therefore be attributed to human impacts on climate, and not to natural variability.

See also: Arctic Climate, Volume 1; Natural Climate Variability, Volume 1.

FURTHER READING

Baldwin, M P and Dunkerton, T J (1999) Propagation of the Arctic Oscillation from the Stratosphere to the Troposphere, *J. Geophys. Res.*, 104, 30 937–30 946.

Graf, H-F, Kirchner, I, and Perlwitz, J (1998) Changing Lower Stratospheric Circulation: The Role of Ozone and Greenhouse Gases, *J. Geophys. Res.*, 103, 11 251–11 261.

Hartmann, D L, Wallace, J M, Limpasuvan, V, Thompson, D W J, and Holton, J R (2000) Can Ozone Depletion and Global Warming Interact to Produce Rapid Climate Change? *Proc. Natl. Acad. Sci.*, 97, 1412–1417.

Kuroda, Y and Kodera, K (1999) Role of Planetary Waves in the Stratosphere–Troposphere Coupled Variability in the Northern Hemisphere Winter, *Geophys. Res. Lett.*, 26, 2375–2378.

Ohhashi, Y and Yamazaki, K (1999) Variability of the Eurasian Pattern and its Interpretation by Wave Activity Flux, *J. Meteorol. Soc. Japan*, 77, 495–511.

Perlwitz, J and Graf, H-F (1995) The Statistical Connection between Tropospheric and Stratospheric Circulation of the Northern Hemisphere in Winter, *J. Clim.*, 8, 2281–2295.

Shindell, D T, Miller, R L, Schmidt, G A, and Pandolfo, L (1999) Simulation of Recent Northern Winter Climate Trends by Greenhouse Gas Forcing, *Nature*, 399, 452–455.

Shindell, D T, Schmidt, G A, Miller, R L, and Rind, D (2001) Northern Hemisphere Winter Climate Response to Greenhouse Gas, Volcanic, Ozone and Solar Forcing, *J. Geophys. Res.*, 106, 7193–7210.

Thompson, D W J and Wallace, J M (1998) The Arctic Oscillation Signature in the Wintertime Geopotential Height and Temperature Fields, *Geophys. Res. Lett.*, 25, 1297–1300.

ARM (Atmospheric Radiation Measurement) Program

The ARM is a highly focused observational and analytical research effort that collects data for comparison with and improvement of the atmospheric component of global climate models (AGCMs). The primary focus for ARM has been on acquiring the data needed to improve understanding of the role of clouds and the treatment of cloud-radiation feedbacks (*see* Cloud- Radiation Interactions, Volume 1). The program is sponsored by the US Department of Energy as part of its contribution to the US Global Change Research Program, and is also closely coordinated with a number of international research programs.

To gather the needed data, ARM observatories have been established at a number of locations experiencing quite different types of cloud cover. The first ARM site was the Cloud and Radiation Test-bed (CART) facility set up in the southern Great Plains of the US; this site experiences a wide variety of continental cloud types and began collecting data in the spring of 1992. The second ARM site, which is located in the tropical western Pacific Ocean and which began operations in September 1996, has stations in Manus and Nauru. The third ARM site, located on the North Slope of Alaska, began operations in the summer of 1997. Data from each of the sites are assembled using a centralized data system and are made available in near real-time to participating scientists. Data for the general science community are made available through a data archive established at the Oak Ridge National Laboratory.

Using the data, the program s objectives are to quantify the radiation balance from the surface to the top of the atmosphere, to determine the atmospheric characteristics responsible for this balance, to improve the parameterization of the formation and evolution of clouds in climate models, to operate an experimental test-bed for testing process models being used in AGCMs, and to provide satellite ground-truth measurements. The research program gathers data both from networks of ground-based remote-sensing instruments and, during focused campaign studies, from manned and unmanned aircraft. Measurements include

vertical profiles of temperature, water vapor, trace gases, aerosols, and solar and infrared radiation. In addition, ARM gathers satellite data to supplement the ground-based and *in situ* data. The ARM data are intended to be used as a testbed for evaluating the process models and their treatments of the cloud-climate feedbacks in AGCMs and mesoscale models.

REFERENCE

http://www.archive.arm.gov/data/ordering.html.

MICHAEL C MACCRACKEN *USA*

Arrhenius, Svante

(1859- 1927)

Svante Arrhenius, Swedish physicist and chemist, received the Nobel Prize in Chemistry in 1903 for his theory on electrolytic dissociation. In environmental science he is best known for his pioneering work on the atmospheric greenhouse effect of carbon dioxide.

Arrhenius got his PhD from Uppsala University in 1884 based on a thesis in which he developed a theory for electrical conductivity of aqueous electrolytes. During the following years he spent long periods at various European universities. Among his achievements during these years was the formulation, inspired by J H van t Hoff in Amsterdam, of the well-known law for the temperature dependence of chemical reactions.

In 1891 Arrhenius was employed as a teacher at Stockholm s Hogskola (now Stockholm University). In 1895, he became a full professor, a position he held until his retirement in 1927. He was the rector of the Hogskola from 1896 to 1902. Soon after he joined the Hogskola in 1891, Arrhenius took the initiative to form the Stockholm Physics Society, which brought local scientists in physics, chemistry and geosciences together for informal and very fruitful discussions on topics such as climate variations and other large-scale interdisciplinary issues related to the oceans, the atmosphere, the solid Earth and space. One of the discussion topics was the cause of the Quaternary glaciation cycles. The geologist Arvid Hogbom had suggested that volcanic eruptions might have caused large variations in the atmospheric concentration of carbon dioxide and Arrhenius wanted to find out if such variations might explain the temperature variations during the glacial cycles. Arrhenius was not first to describe what is now referred to as the greenhouse effect (the credit for this should go to the French scientist Joseph Fourier who, in 1824, introduced the glass bowl analogy to the heating effect of infrared absorbing gases in the atmosphere), but Arrhenius was the first to make a quantitative estimate of the magnitude of the greenhouse effect. In a now classic paper published in 1896 (On the Influence of Carbonic Acid in the Air on the Temperature of the Ground, *The Philosophical Magazine*, 41, 237–276) Arrhenius estimated that a doubling of the carbon dioxide concentration would lead to a 5.7 °C increase in the mean surface temperature. In view of the very limited data Arrhenius had at his disposal, this estimate is surprisingly close to the present estimate of 1.5–4.5 °C. In his calculation, Arrhenius also took into account the positive feedback on the temperature of the increasing amount of water vapor in a warmer atmosphere.

Although Arrhenius first estimate of the greenhouse effect of carbon dioxide was made with the purpose of finding a possible cause of the glacial temperature cycles, he soon realized its relevance for the issue of a human-induced impact on climate. His first estimate of a man-made global temperature change, due to carbon dioxide emissions from fossil fuel combustion, was also published in 1896. He then estimated that the burning of coal alone would thus be capable of raising the temperature of the earth by somewhat more than one thousandth of a degree centigrade per annum. This is about 20 times smaller than current estimates, the reason for the difference being mainly Arrhenius failure to anticipate the enormous growth in fossil fuel use that has taken place since his days.

Arrhenius spent much time and effort educating the general public about the new domains conquered by science. His books *Textbook of Cosmic Physics* (1903) and *Worlds in the Making: The Evolution of the Universe* (1906), were translated into several languages.

Photo: © Royal Swedish Academy of Sciences.

FURTHER READING

Crawford, E (1996) Arrhenius: From the Ionic Theory to the Greenhouse Effect, *Philos. Mag.*, 41, 237–276.

Rodhe, H and Charlson, R (1998) *The Legacy of Svante Arrhenius: Understanding the Greenhouse Effect*, Royal Swedish Academy of Sciences and Stockholm University, Stockholm.

HENNING RODHE *Sweden*

Asteroids and Comets, Effects on Earth

Curt Covey
Lawrence Livermore National Laboratory, Livermore, CA, USA

Collisions of asteroids and comets with Earth are rare events, but over the long span of Earth's history such events have taken place many times. They will continue to occur unless future technology is used to prevent them. An asteroid or comet collision produces an explosion of enormous energy, equivalent to thousands of times the explosive energy of all the world's nuclear weapons at the height of the Cold War. Early in Earth's history these collisions occurred more frequently than they do today. They may have hindered the early evolution of life. On the other hand, comets may have provided Earth with much of the water and other substances necessary for life to exist.

A variety of disastrous effects would result from any such collision. Likely global-scale effects include stratospheric ozone depletion, acid rain, global wild res, and enough smoke and dust to block sunlight and cool the surface for months. After the smoke and dust cleared, substantial global warming would result if the collision had caused large amounts of water or carbon dioxide to enter the atmosphere. The most recent time that a large asteroid or comet collided with Earth was 65 million years ago, at about the same time that the dinosaurs and many other species became extinct. It seems probable that that collision caused or at least contributed to these extinctions. Several other mass extinction events in Earth's history may be connected with asteroid or comet impacts.

An asteroid or comet collision as severe as the one 65 million years ago is unlikely to happen again during the next several million years. Less energetic impacts occur more frequently. The probability is quite low that a collision severe enough to endanger the global environment will occur during the next few human generations. It is an open question whether such an unlikely but gravely consequential event warrants human concern at this time.

IMPACTS OVER EARTH HISTORY

The 1994 collision of Comet Shoemaker-Levy 9 with Jupiter served as a reminder that collisions of asteroids and comets with planets occur from time to time. The effects on Jupiter s clouds were visible even in small telescopes for months afterward. A quarter of a century of spacecraft exploration preceding this spectacular event revealed that impact craters are the dominant landforms on most of the solid bodies (planets, moons and asteroids themselves) in the solar system. Indeed, these bodies – as well as the cores of giant gaseous planets like Jupiter – are believed to have formed over 4.5 billion years ago by the accretion of numerous smaller rocky and icy planetesimals.

The crater-covered surface of the Earth s Moon provides the closest and most detailed record of asteroid and comet impacts over the past several billion years. Dating of lunar rocks returned to Earth by astronauts and automated spacecraft, together with mapping of craters by telescopes and lunar-orbiting satellites, gives a history of asteroid and comet impacts spanning the past 4 billion years. At the beginning of this period, impacts were several times more frequent than today, but they decreased rapidly to near present day rates by about 3 billion years ago. The rapid pace of cratering nearly 4 billion years ago is sometimes referred to as the late heavy bombardment. It may have been the final phase of the Moon s – and presumably Earth s – accretion from planetesimals.

Due to erosion and other geologic processes, craters on Earth are not nearly as well preserved as they are on the Moon. It is also necessary to distinguish impact craters from volcanic craters. Careful surveys have revealed over 150 impact craters on the continents with a few new discoveries made each year (Figure 1). Among the most evident is Meteor Crater in Arizona, 1.2 km in diameter, but others are considerably larger. Examples include the Manicouagan and Sudbury craters in Canada, of 100 and 140 km diameter, respectively, and the comparably sized Chicxulub crater discussed later. Over the past several hundred million years the rate at which sizeable craters have appeared on Earth, according to the geologic record, is roughly consistent with both the lunar cratering rate inferred for the same period and the number of asteroids and comets that astronomers observe today.

The role of such impacts during the earliest phases of Earth history can only be guessed. Because water and carbon-containing molecules are present in some asteroids and all comets, it is possible that these materials – which are essential for life as we know it – were supplied to Earth during the time of the late heavy bombardment. For example, comets are nearly half water (ice), and the number of comets impacting Earth since its formation could provide more than enough water to supply all of the present day oceans. On the other hand, comet impacts may have been sufficiently energetic to blast as much water as the comets contained back into space. Recent measurements of the composition of three comets suggest that this process (impact erosion) may have prevented comets from supplying a significant amount of water to Earth. The ratio of heavy to light isotopes in the hydrogen in these comets is about twice that of ocean water, implying a source other than (or in addition to) comets. Conventional wisdom holds that Earth s water – and other volatile materials

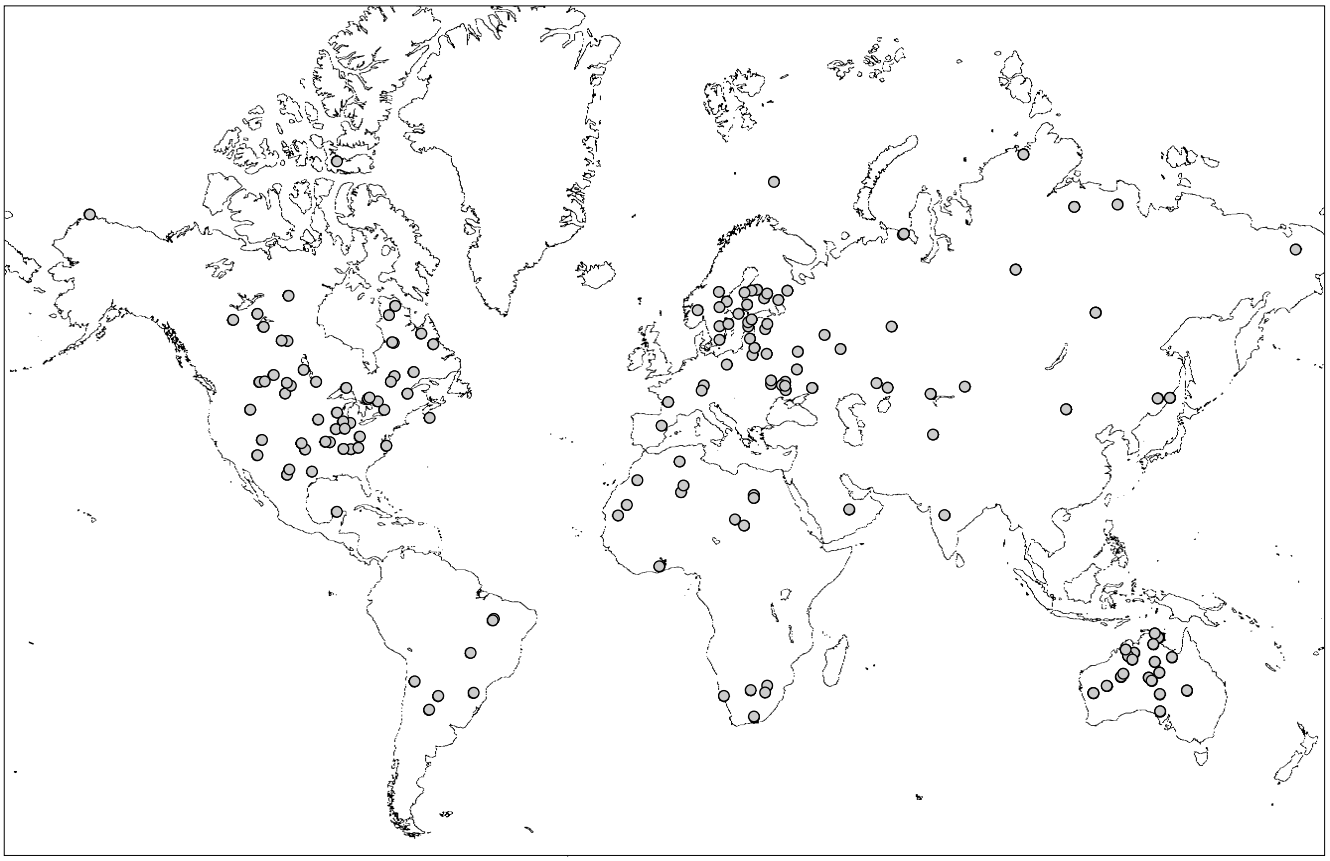


Figure 1 Locations of currently known terrestrial impact craters. Updated from Grieve and Shoemaker (1994). (Reprinted by permission of the University of Arizona Press)

such as the gases comprising the atmosphere – were chemically bound to the rocky material that originally formed the planet. According to this theory, the Earth s ocean and atmosphere formed gradually as the volatile materials outgassed from the interior of the planet in volcanic eruptions.

Rather than promoting life, the late heavy bombardment may have actually hindered its emergence. The lunar record implies that not only the frequency, but also the energy of impacts on the very early Earth was far greater than it has been during more recent times. Collisions prior to about 4 billion years ago could have repeatedly evaporated the oceans and destroyed all forms of life. It is possible that life originated several times on the early Earth but only survived to give rise to the known biosphere after the late heavy bombardment diminished. This scenario is perhaps consistent with the fossil record, which indicates that life has existed continuously on Earth for at least the past 3.5 billion years. During this time, individual impacts were not sufficiently energetic to extinguish all life, and the extinction events they caused may have opened new niches for other forms of life including humans.

ASTEROID AND COMET POPULATIONS TODAY

Asteroids are rocky bodies ranging in size from ~1000 km diameter downward. Although the vast majority of them reside between Mars and Jupiter, a substantial number are perturbed into orbits that approach or cross that of Earth. Over 250 such near-Earth asteroids greater than about 1 km in diameter have been found in systematic surveys by telescopes. It is estimated that twice that number exist but are undiscovered as yet. If the orbits of the known near-Earth asteroids are typical of the entire population, then about five such objects greater than 1 km in diameter would be expected to collide with Earth every million years. (A far greater number of very small asteroids in Earth-crossing orbits provide the source of ordinary meteorites: about 4000 larger than 1 kg in mass strike Earth each year.) From this cosmic perspective the big collisions are like unusual rolls of dice in a game. They are rare events, randomly spaced in time, but virtually guaranteed to occur repeatedly if the game goes on long enough.

Comets also collide with Earth. These bodies are distinguished from asteroids not only by their icy

composition – which generates a bright tail of gas and dust when comets are sufficiently near the Sun – but also by their highly eccentric and often retrograde orbits about the Sun. From these orbital characteristics, it appears that comets originate from among an enormous number of icy bodies residing at very great distances from the Sun (about 40 000 times the Earth–Sun distance on average). Random external perturbations such as a passing star occasionally send an icy body much closer to the Sun, where it becomes visible as a comet and often passes inside the orbit of Earth. Telescopic surveys and statistical theory, similar to the considerations discussed above for asteroids, imply that a solid comet nucleus greater than about 1 km in diameter should collide with Earth roughly every 2–10 million years.

Earth s collisions with comets are thus estimated to be less frequent than its collisions with asteroids. This situation may not always have been the case. Stars or interstellar clouds passing near the solar system could send a greatly enhanced number of comets into the planetary region. From the number and movement of stars and clouds in our part of the Galaxy, such events have been estimated to occur roughly every 10 000–400 000 years. Tentative evidence for these comet showers has been inferred from the geologic record.

Collision frequencies depend strongly on the size of the objects considered (Figure 2). Larger asteroids and comets are far less abundant than smaller ones, and thus they collide with Earth much less often. For example, the average rate of asteroid collision is believed to be roughly once every few hundred thousand years for kilometer-size objects but only once every 50–100 million years for objects 10 or more kilometers in diameter.

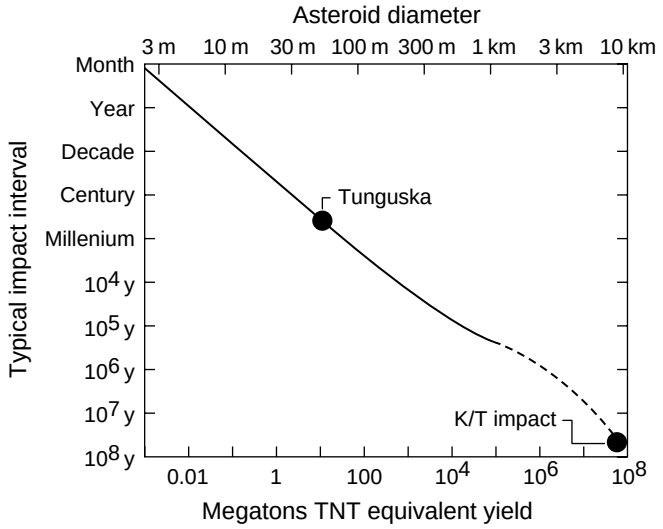


Figure 2 Typical interval between impacts as a function of asteroid diameter, or equivalently (for typical asteroid density and velocity) the explosive energy of impact. The solid curve is obtained from crater counts. The dashed curve extrapolates the data to the K–T impact. The other point locates the diameter of the Tunguska impacting body on the assumption that it was an asteroid rather than a comet (*see* **Tunguska Phenomenon**, Volume 1). From Morrison *et al.* (1994). (Reprinted by permission of the University of Arizona Press)

THE CRETACEOUS–TERTIARY IMPACT

The geologic record contains evidence of several impacts over the past several hundred million years. By far the most complete evidence to be discovered comes from 65 million years ago, the transition between the Cretaceous and Tertiary eras (or K–T boundary). The discussion below focuses on this one event, but similar effects could be expected from any impact of similar magnitude (*see* Extinctions and Biodiversity in the Fossil Record, Volume 2).

Primary evidence for an impact 65 million years ago comes from sediments deposited at that time. The K–T transition can be precisely identified in sedimentary rock layers by an abrupt disappearance of fossilized microscopic Cretaceous plants and animals: they are far less abundant immediately above the boundary than they are immediately below it. It has been estimated that half the genera of living organisms perished at the end of the Cretaceous – including dinosaurs, many land plants, and marine creatures of all sizes in addition to most of the microscopic plankton – setting the stage for the proliferation of other life forms including, ultimately, humanity.

K–T sediments contain greatly enhanced concentrations of iridium, an element that is relatively depleted in Earth s crust but more abundant in meteorites (which originate from asteroids). Enhanced iridium at the K–T boundary has now been detected at over 80 locations worldwide. Enhanced concentrations of other so-called platinum-group elements appear at the K–T boundary, in the same ratios to each other as in meteorites (which distinguishes the K–T sediments from volcanic deposits). Peculiar rounded grains (spherules) of glassy material – apparently arising at impact from the melting and mixing together of different types of rock – are also present at the K–T boundary. So are grains of quartz containing microscopic bands of deformation, seen elsewhere only in rocks near known impact craters or in nuclear explosion craters, and evidently produced by shock waves passing through the target rock.

After discovery of the evidence described above, a worldwide search to locate the site of the K–T impact culminated in discovery of a buried crater, at least 180 km in diameter, near Puerto Chicxulub (pronounced cheek-shoe-lube) on the Yucatan coast of Mexico. The Chicxulub crater is buried under sediments deposited during the slow sinking of the Yucatan area over the past several million years. The crater can be seen in maps depicting subtle irregularities of Earth s gravitational or magnetic fields. Drilling samples

from Chicxulub (originally obtained by geologists exploring for oil in the area) match the isotope composition of deposits at the K–T boundary in neighboring regions of the Caribbean and Gulf of Mexico. These deposits appear to record the passage of enormous ocean waves (tsunami) at the time of impact (Figure 3). The Chixulub samples themselves contain rocks that were melted at the precise time of the K–T event, according to inference from their radioactive isotope composition.

Taken as a whole the evidence suggests that the impacting body 65 million years ago was an asteroid, although it is possible it may have been a comet nucleus. The worldwide amount of excess iridium and other trace elements deposited at the K–T boundary, together with the concentrations of this element found in asteroids, implies that the impacting body was about 10 km in diameter. If it struck Earth at a velocity typical of asteroids, 15 km s^{-1}, the resulting energy released on impact would have been equivalent to the detonation of about 50 million megatons (or 50 trillion tons) of high explosive. If the impacting body was a comet, the impact energy would likely have been much larger – perhaps as large as a billion megatons – due mainly to the higher impact velocities typically arising from comet orbits. These enormous numbers contrast with values of the order of 10 megatons for the energies of both Meteor Crater s formation 50 000 years ago and the 1908 Tunguska explosion (*see* Tunguska Phenomenon, Volume 1).

Fifty million megatons is thousands of times the total explosive energy of all the world s nuclear arsenals at the height of the Cold War. A single explosion of this magnitude would create a crater of the order of 100 km in diameter (roughly consistent with the size of the Chixulub crater). Blast and earthquake damage would extend for thousands of kilometers and – if the impact was located in an ocean – tsunami would devastate the surrounding coastal margins. Fully global effects, however, would arise only indirectly as a consequence of the explosion s interaction with the atmosphere, as discussed in the following section.

Figure 3 The K–T boundary in north-eastern Mexico. This region was underwater in the Gulf of Mexico at the close of the Cretaceous. At and below the geologist's hand is impact debris, unusually thick because this site is near the impact point at Chixulub. The layers above record the passage of enormous ocean waves (tsunami) that broke up and mixed together the original ocean bottom deposits. The lowest sediments date from before the impact. From Smit (1994). (Reprinted by permission of the University of Arizona Press)

GLOBAL CONSEQUENCES

Even before an asteroid or comet strikes the Earth s surface, its passage though the atmosphere would produce shock waves in which air is heated and nitrogen and oxygen (the two most common atmospheric constituents) chemically combine to form nitrogen oxide. This compound destroys ozone in a highly efficient catalytic process. Together with other atmospheric changes, nitrogen oxide production could cause severe depletion of the stratospheric ozone layer, although the resulting increase in harmful solar ultraviolet radiation might well be intercepted by impact-generated pollutants before it reaches the surface. Acid rain would also be produced as the nitrogen oxide converts to nitric acid in cloud droplets (and possibly by analogous processes involving sulfate compounds). These changes in atmospheric chemistry are difficult to quantify but probably would produce a substantial stress on the biosphere.

The energy of an asteroid or comet impact would be sufficient to melt or vaporize all of the impacting body, together with a comparable amount of target rock, and to launch the resulting debris on ballistic trajectories into space. (For a body the size of the K–T impactor, with a size about twice the average depth of the oceans, this would occur whether the impact occurred on land or sea.) Friction with the atmosphere on reentry of this material could produce enough heat to ignite wildfires worldwide. This sequence of events appears to have occurred following the K–T impact. Soot is deposited at the K–T boundary in amounts equivalent to 1–2% of the aboveground carbon content of the Cretaceous biosphere. Since fires convert only a small fraction of carbon to soot, the implication is that much of the Cretaceous land biomass (considerably more than today s aboveground plant material) was consumed.

Apart from the catastrophic direct effects of global wildfires, the soot they produce would strongly absorb sunlight, preventing it from reaching the surface. In amounts found at the K–T boundary – over 10 mg for each cm^2 of Earth s surface area – an atmospheric layer of soot absorbs essentially all incident sunlight. Impact-produced dust would have a similar effect. In addition to melting or vaporizing a considerable amount of material, the energy of impact would pulverize a much larger quantity of target rock, at least 100 times the mass of the impacting body according to

theoretical calculations, and propel it into the atmosphere. Shocked quartz grains in K–T deposits provide observational evidence of this process. A fraction of this material would be microscopic in size. Upon reaching the stratosphere, it could be expected to remain there for a year or two, by analogy with the microscopic sulfate particles that major volcanic eruptions produce. Asteroid or comet impacts themselves might produce sulfate particles if sufficient amounts of sulfur were present in either the impacting body or the target rock. The latter case appears to have occurred at the K–T impact site.

All of the above effects would reverse the normal situation in which most incident sunlight penetrates the atmosphere and is absorbed at Earth s surface. Instead, following a major impact most solar energy would either be absorbed high in the atmosphere or reflected back to space. In either case the result would be surface cooling. Computer simulations of climate (similar to those used to predict human-induced global warming) imply that the interiors of continents would cool rapidly, approaching the freezing point of water within a week or two after the impact, even in summertime. Ocean and coastal areas would not cool significantly, however, because the large volume of water in the upper oceans stores enough heat to mitigate a loss of sunlight for several years (see Figure 4). As noted above, dust and other particles are expected to remain in the stratosphere no longer than a few years. The tremendous perturbations to surface and atmospheric temperature during this time would undoubtedly lead to major changes in weather patterns. One possibility implied by the computer simulations is global drought resulting from the stabilized atmosphere (higher temperatures above cooler temperatures) and the resulting suppression of cloud formation.

After the dust, smoke and sulfate particles cleared from the atmosphere, it is possible that the global cooling they caused could be replaced by global warming due to an enhanced greenhouse effect. This effect would occur if an asteroid or comet impact – or its primary effects – injected sufficient quantities of carbon dioxide or water vapor into the atmosphere. Wildfires of the extent inferred for the K–T event would produce several times the carbon dioxide present in today s atmosphere. Impact into carbonate sediments could add to the effect of fires. On the other hand, the Cretaceous atmosphere is thought to have already contained enhanced carbon dioxide levels compared to the present, and it is not clear if the additional enhancement resulted in noticeable long-term warming. An ocean impact would inject far more water into the stratosphere than is normally present there, although most of it would be removed at least as fast as smoke and dust particles.

In light of the overwhelming evidence that a substantial sized ($\sim$10 km diameter) asteroid or comet collided with Earth at the end of the Cretaceous, and the possible global

effects listed previously, it seems reasonable to conclude that the impact was responsible for the K–T mass extinction. This inference, however, is weakened by the scarcity of evidence geologists and paleontologists have for such a remote time. Nor can one exclude the possibility that separate but simultaneous effects were operating. Indeed, there is evidence that an upsurge in volcanic activity occurred at the end of the Cretaceous, possibly adding to the carbon dioxide greenhouse effect mentioned above. Perhaps the safest thing to say is that the major asteroid or comet impact that evidently occurred at the K–T transition would surely have pushed many plants and animals in the direction of extinction.

A natural follow-up question is whether other impacts caused other mass extinction events. The K–T boundary marks the most recent of five great mass extinction events since the detailed fossil record began 570 million years ago. Some evidence of an asteroid or comet impact has been found for other mass extinction events, for example shocked quartz grains at the boundary between the Triassic and Jurassic eras 205 million years ago. In addition, enrichment of platinum-group elements has been detected for the times of smaller extinction events 11, 34 and 91 million years ago. It is possible – although difficult to prove – that asteroid or comet impacts contributed to most of the extinction events that mark transitions between the major subdivisions of Earth s geologic history.

THE IMPACT HAZARD TODAY

As discussed previously, both the record of crater formation on Earth and the Moon, and the numbers of asteroids and comets found in near-Earth orbits, provide estimates of how often impacts of such bodies on Earth could be expected. The very low probability obtained for a sizeable impact is consistent with the apparent absence of any during recorded history. The largest observed event is the Tunguska explosion, probably caused by an asteroid of the order of 100 m in diameter. Impacts of this size are expected to occur once every few centuries. Evidently the damage they cause has gone largely unnoticed, an unsurprising situation considering the local nature of the destruction and the relatively low human population during most of recorded history. The Tunguska explosion, for example, flattened a remote Siberian forest but resulted in only two reported fatalities.

For asteroids and comets about half a kilometer in diameter, the typical impact interval is between 50 000 and 100 000 years. Objects this large would produce craters about 10 km in diameter and extensive regional damage rivaling or exceeding that of the greatest natural disasters in human experience. Blast, earthquake and fire damage would extend for 100 km or more. If the impact occurred in an ocean basin, the resulting tsunami would sweep over all adjacent coastlines. At somewhat larger size – about 1–2 km diameter – asteroids and comets would begin to produce global effects as discussed above for the K–T impact.

Should asteroid and comet impacts be a cause for concern? Although the likelihood of a sizeable collision in the near future is exceedingly small, the results if one occurred would be disastrous beyond human experience. How to deal with such low probability, high consequence events is not obvious. For example, one might assume that 1-km objects strike Earth roughly once every 300 000 years. One such impact could kill something like a billion people if worldwide disruption of agriculture resulted. Multiplying these two numbers together ($1\,000\,000\,000 \times 1/300\,000$) gives an average fatality rate of about 3000 people per year. What if anything this number implies is a matter of individual interpretation.

Some protective action could be undertaken with today s technology. A large fraction of near-Earth asteroids greater than 1 km in size have already been identified. Nearly all of these objects could be tracked with a modest extension of present day efforts. Because an asteroid destined to collide with Earth will typically make a large number of close approaches before impact, the observing system would likely provide at least several decades of warning. During that time the missile and nuclear weapon technologies of the Cold War might be adapted to deflect the asteroid. This leisurely schedule would not apply to long-period comets – which enter the planetary region with rarely more than 1 year of advance notice – or to smaller but still dangerous asteroids that would be detected only when they are extremely close to Earth. Examples of the latter class include three asteroids (all with diameters greater than 100 m) that in the past decade approached Earth to within twice the Earth–Moon distance. It is not obvious that ensuring protection against all such objects is achievable with present day technology.

See also: Earth System History, Volume 1.

ACKNOWLEDGMENTS

I am grateful to the editors of this encyclopedia and to Paul Weissman of the Jet Propulsion Laboratory for valuable comments on this article.

REFERENCES

Covey, C, Ghan, S J, Walton, J J, and Weissman, P R (1990) Global Environmental Effects of Impact-generated Aerosols: Results from a General Circulation Model, in *Global Catastrophes in Earth history: an Interdisciplinary Conference on Impacts, Volcanism, and Mass Mortality*, eds V L Sharpton and P D Ward, Geological Society of America, Boulder, CO, Special Paper 247, 263–270.

Grieve, R A F and Shoemaker, E M (1994) The Record of Past Impacts on Earth, in *Hazards Due to Comets and Asteroids*, ed T Gehrels, University of Arizona Press, Tucson, AZ, 417–462.
Morrison, D, Chapman, C R, and Slovoc, P (1994) The Impact Hazard, in *Hazards Due to Comets and Asteroids*, ed T Gehrels, University of Arizona Press, Tucson, AZ, 59–92.
Smit, J (1994) Extinctions at the Cretaceous Tertiary Boundary: the Link to the Chicxulub impact, in *Hazards Due to Comets and Asteroids*, ed T Gehrels, University of Arizona Press, Tucson, AZ, 859–878.

FURTHER READING

Alvarez, W (1997) *T. Rex and the Crater of Doom*, Princeton University Press, Princeton, NJ, 1–185.
Beatty, J K, Petersen, C C, and Chaikin, A, eds (1999) *The New Solar System*, 4th edition, Sky Publishing Corporation, 1–421.
Gehrels, T, ed (1994) *Hazards Due to Comets and Asteroids*, University of Arizona Press, Tuscon, AZ, 1–1300.
Thomas, P J, Chyba, C F, and McKay, C P, eds (1997) *Comets and the Origin and Evolution of Life*, Springer-Verlag, New York, 1–296.
Toon, O B, Zahnle, K, Morrison, D, Turco, R P, and Covey, C (1997) Environmental Perturbations Caused by the Impacts of Asteroids and Comets, *Rev. Geophys.*, 35, 41–78.

Atlantic Ocean

The *Atlantic Ocean* is the body of water bordered by the American continents, Greenland, Europe, Africa, and the Antarctic region. It is connected to the Indian Ocean south of Africa and to the Pacific Ocean south of South America by a continuous current (Antarctic Circumpolar Current) that flows eastward around Antarctica and to the Arctic Ocean through passages across the ridge stretching from Greenland to the UK (*see* Ocean Circulation, Volume 1). The average depth of the Atlantic is around 4000 m. Its maximum depth is 8648 m, in the Puerto Rico Trench. The Reykjanes Ridge and Mid-Atlantic Ridge are significant topographic features that separate the western and eastern portions of the Atlantic throughout its length. Three marginal seas impact the water properties of the Atlantic Ocean: the Mediterranean Sea through its high salinity, the Inter-American Seas (Caribbean region) as a conduit for part of the Gulf Stream, and the Arctic Ocean through its production of very dense water. Some areas that are part of the open North Atlantic Ocean have specific geographical names: the Labrador Sea between Labrador and Greenland, the Irminger Sea between Greenland and the Reykjanes Ridge, and the Sargasso Sea south and east of the Gulf Stream and west of the Mid-Atlantic Ridge.

The Atlantic is the most saline of the three major oceans, as a result of higher net evaporation rates over a large area (*see* Salinity Patterns in the Ocean, Volume 1). The northern North Atlantic is one of the two global sites for deep-water formation, the other being around Antarctica. Deep waters in the northern North Atlantic are approximately 20 years old, compared with 500 years for the deep North Pacific where there is no deep-water source. Through ocean–atmosphere coupling, the tropical Atlantic provides a major inter-annual climate signal that impacts the neighboring regions. Coupling with the atmosphere in the North Atlantic and Arctic provide decadal climate signals called the *North Atlantic* and the *Arctic Oscillations* (*see* Arctic Oscillation, Volume 1; North Atlantic Oscillation, Volume 1), which impacts weather and especially precipitation in Europe and eastern North America.

LYNNE D TALLEY *USA*

Atmospheric Angular Momentum and Earth Rotation

Richard D Rosen
Atmospheric and Environmental Research, Inc., Lexington, MA, USA

The study of angular momentum provides a natural framework for understanding the dynamics of rotating systems. Much of our knowledge about the general circulation of the atmosphere has been gained through studies of its angular momentum about the polar axis, an integrated measure over all parcels of air of the product of their mass, westerly component of motion, and distance from the axis. As a whole, the atmosphere is rotating faster than the Earth beneath, so its relative angular momentum is positive (on average, equal to about 1.4×10^{26} *kg m*^{2} *s*$^{-1}$*). The main sources and sinks of angular momentum at the atmosphere's lower boundary, as well as the role of large-scale waves in maintaining the mean circulation through the transport of angular momentum from the tropics to the midlatitudes, have been well documented (Peixoto and Oort, 1992). Atmospheric angular momentum should, therefore, serve as a fundamental measure of the atmosphere's dynamic response to global environmental change.*

The development of improved data sets over the past two decades has heightened awareness of the close relationship

between changes in the axial component of the global atmosphere s angular momentum (AAM) and that of the solid Earth, the latter being reflected in small, but measurable, changes in the planet s rate of rotation. In the absence of external torques on the coupled dynamical system comprised of the atmosphere and the solid Earth, the law of angular momentum conservation dictates that changes in the momentum about the polar axis of one component of the system must be simultaneously matched by equal, but opposite, changes in the momentum of the other. Given the size of the changes observed in the large-scale wind field responsible for changes in AAM, this relationship implies that compensating changes in Earth s rotation rate, reckoned in terms of changes in the length of the day (ΔLOD), will be no more than fractions of a millisecond (10^{-3} s). Remarkably, modern space geodetic techniques are capable of detecting such tiny fluctuations in the length of the day, and once the effects of known tidal forces are removed from time series of ΔLOD, this series correlates strongly with series of AAM over a wide range of time scales (Rosen, 1993). In conforming to the momentum conservation principle, this agreement attests to the quality of modern atmospheric and geodetic data sets, and to the secondary role played by other geophysical processes (such as changes in ocean currents or the redistribution of fresh water on land) in affecting ΔLOD.

Much of the variability in AAM and ΔLOD that occurs on interannual time scales is related to the El Niño/Southern Oscillation (ENSO) phenomenon (*see* El Niño/Southern Oscillation (ENSO), Volume 1). When water temperatures in the tropical Pacific Ocean are unusually high, as during the warm phase of ENSO, winds in the subtropical jet stream become anomalously strong, leading to larger values than normal of AAM and ΔLOD. This signal has been reproduced in simulations of the recent past by atmospheric general circulation models that are driven by historical sea surface temperature (SST) fields (Hide *et al.*, 1997). From an extended run of one such model attempting to simulate the climate of the 20th century, the angular momentum of the model atmosphere was analyzed for interannual and multidecadal signals (Rosen and Salstein, 2000). The interannual variability of the model s AAM compares well with that derived from observed AAM values available since the middle of the century. In both the model and observed series, there is a notable increase in year to year variability since the 1970s, coinciding with a marked change in atmospheric circulation regimes at that time. In the model, the recent high level of variability is comparable to that during the first two decades of the century, but this result is strongly dependent on the quality of the SST field used in the experiment, because of the strong linkage between AAM variability and ENSO. Observations do point to a tendency towards more ENSO warm events and a warmer tropical Pacific in general by the end of the 20th century, and as a consequence, a trend in AAM towards larger values is discernible. Based on the observations of AAM available since around 1950, this trend, in equivalent units of LOD, amounts to lengthening the day at a rate of around 0.5 millisecond (ms) per century.

How will AAM and ΔLOD respond to a possible warming of the globe in the 21st century? This question was first addressed by Rosen and Gutowski (1992) but more recently by Huang *et al.* (2001) using a coupled atmosphere–ocean global climate model in which the concentration of greenhouse gases is allowed to increase through the year 2100. A significant increase in tropical SSTs is found in this experiment, accompanied by a similar rise in AAM. The implied increase in the length of the day during the 21st century is, again, about 0.5 ms. This possibility of human intervention with Earth s rotation, via the atmosphere s response to an accumulation of greenhouse gases, is noteworthy.

See also: Tides, Oceanic, Volume 1.

REFERENCES

Hide, R, Dickey, J O, Marcus, S L, Rosen, R D, and Salstein, D A (1997) Atmospheric Angular Momentum Fluctuations During 1979–1988 Simulated by Global Circulation Models, *J. Geophys. Res.*, 102, 16 423–16 438.

Huang, H-P, Weickmann, K M, and Hsu, C J (2001) Trend in Atmospheric Angular Momentum in a Transient Climate Change Simulation with Greenhouse Gas and Aerosol Forcing, *J. Clim.*, 14, 1525–1534.

Peixoto, I P and Oort, A H (1992) *Physics of Climate*, American Institute of Physics, New York.

Rosen, R D (1993) The Axial Momentum Balance of Earth and its Fluid Envelope, *Surv. Geophys.*, 14, 1–29.

Rosen, R D and Gutowski, W J (1992) Response of Zonal Winds and Atmospheric Angular Momentum to a Doubling of CO_2, *J. Clim.*, 5, 1391–1404.

Rosen, R D and Salstein, D A (2000) Multidecadal Signals in the Interannual Variability of Atmospheric Angular Momentum, *Clim. Dyn.*, 16, 693–700.

Atmospheric Chemistry, Stratospheric

see Stratosphere, Chemistry (**Volume 1**)

Atmospheric Chemistry, Tropospheric

see Troposphere, Ozone Chemistry (**Volume 1**)

Atmospheric Composition, Past

James C G Walker
The University of Michigan, Ann Arbor, MI, USA

In recent decades, scientists have been able to demonstrate that the composition of Earth's atmosphere has changed from the natural composition prevailing over the previous 10 millennia. The records and analyses indicate that the atmosphere is an ephemeral reservoir, easily altered by the exchange of material with the much more massive ocean and solid Earth. Ice core records deciphered during recent decades have also demonstrated that the composition of Earth's atmosphere can change quickly and has changed signi cantly over the entire course of the Earth's history. Change results from both biological and geological processes, which continually convert atmospheric constituents into solid and liquid compounds or restoring them to gaseous form. At the same time, changes in atmospheric composition have affected the metabolic capabilities of living creatures as well as the climate. Through these connections, it is clear that the various components of the Earth System are closely linked and evolve together. Human beings are part of the Earth System, able to affect its evolution, but not able to escape its constraints.

Given the evidence of both natural and human-induced change, it now seems dif cult to understand how scientists could, at one time, have taken for granted an assumption of an unchanging atmosphere under which Earth and life evolved. Perhaps scientists, like many other people, were more comfortable in a world without change; perhaps an unchanging atmosphere seemed most consistent with the in uential doctrine of uniformity or perhaps an assumption of no change was the most reasonable one in the absence of unambiguous evidence concerning past composition.

*Early interest in atmospheric composition was stimulated largely by environmental issues. The scienti c study of air pollution in Southern California in the 1960s called attention to the chemical reactivity of many atmospheric constituents. Research on the depletion of stratospheric ozone in the 1970s and 1980s made it clear that seemingly small, local additions of gas to the atmosphere could have global consequences. There was also an unresolved theoretical question dating from the end of the 19th century about the fate of the carbon dioxide (CO_2) released to the atmosphere as a result of the burning of fossil fuels. This question was resolved, in part, by the work of C. D. Keeling, who in 1960 reported unambiguous measurements of an increasing concentration of CO_2 in the atmosphere (*see *Keeling, Charles David, Volume 1).*

RESERVOIRS, FLUXES, AND RESIDENCE TIMES

Stimulated by issues ranging from air quality to climate change, atmospheric scientists have learned to think about atmospheric composition in terms of the processes that add gases to or remove gases from the atmosphere and the rates of addition or removal compared with the amounts in the atmosphere. The atmosphere is thought of as a reservoir that contains trace amounts of a particular chemical compound, carbon dioxide (CO_2), for example. Various processes, called *sources*, add that compound to the atmospheric reservoir at rates specified as amount added over a given time. Other processes, called *sinks*, remove the compound from the atmospheric reservoir at rates specified as amount removed over a given time. The sources and sinks together are called fluxes, as they represent flows of matter into or out of the atmospheric reservoir. If the sum of all the sources of a particular compound is equal to the sum of all the sinks, the amount of that compound in the atmosphere is not changing with time. But if the sources exceed the sinks, the compound must be accumulating in the atmosphere; that is the amount of that compound in the atmosphere, its concentration, must be increasing.

Scientists get a sense of how rapidly the concentration of a particular compound can change by dividing the amount of the compound in the atmosphere by the flux in or out, the sources or the sinks. The ratio is called the residence time (*see* Residence Time (of an Atom, Molecule or Particle), Volume 1) of that compound in the atmosphere. The concentration of a given compound in the atmosphere would rise from zero to its actual value in one residence time if the sources were constant and there were no sinks. Alternatively, the atmospheric reservoir would empty in one residence time if the sinks were constant and there were no sources.

Careful evaluation of the sources and sinks and reservoir sizes of the various gases in the atmosphere has revealed that the concentration of all of the gases could have varied during the course of Earth history (*see* Earth System History, Volume 1). The most abundant gases, oxygen and nitrogen, could have varied only over periods of tens to hundreds of millions of years because their residence times are this long. The less abundant and chemically more reactive gases, like hydrogen and methane (CH_4), have residence times of years to decades and so can vary significantly over the course of a human generation if sources and sinks are not equal.

Thousands of Years

The first assessments of past atmospheric composition were theoretical attempts to deduce the properties of the atmosphere when the fluxes of particular gases were very different. Oxygen attracted the most interest. The oxygen in

the modern atmosphere is almost entirely a product of photosynthesis by plants and algae. So was there any oxygen in the atmosphere before photosynthesis originated, and if so, how much? This topic is still of great interest. Oxygen affects many of the chemical processes in the atmosphere and oceans, and it obviously affects life. But photosynthesis appears to have originated more than a billion years ago. This article deals first with the more recent past.

Remarkable data on past atmospheric composition have been derived from measurements of bubbles of air (sometimes called antique-air) trapped in polar ice. Snow accumulates each year on the ice caps of Greenland and Antarctica. This snow is gradually compacted to ice by the weight of subsequent snowfalls. In the process, air, more or less from the time of accumulation, is trapped in the ice. The age of the ice and of its entrapped air increases with depth. Annual accumulations are apparent in ice cores and can be counted. So the ice preserves samples of air of known age.

The analysis has now been carried back about 420 000 years, which is probably close to the limit of what is possible. Old ice flows gradually away from the interior of the ice cap toward lower elevations. Annual layers therefore become thinner and more distorted as the age and depth increase. There may be older ice at the bottom of the ice caps in Greenland and Antarctica, but its age is hard to determine.

For the last two million years, Earth has endured a recurrent sequence of ice ages separated by warm interglacial periods. We are currently enjoying an interglacial that began with a sudden warming and ice melting about 10 000 years ago. The previous interglacial (*see* Eemian, Volume 1) began about 140 000 years ago. The intervening ice age took the form of a gradual decrease in the average surface temperature of the globe, accompanied by a gradual advance of ice sheets over Europe and North America. Earlier ice ages followed the same pattern of gradual cooling followed by sudden warming repeated at approximately equal intervals of time. These periodic fluctuations in Earth s climate are caused by regular changes in Earth s orbit about the Sun that result from gravitational interactions between the Sun, Earth, and the other planets, but it is not clear just how the orbital changes cause major changes in climate. Nor is there agreement among scientists about why the ice ages began just two million years ago.

The ice core data on atmospheric composition reveal a clear record of change closely correlated with the waxing and waning of the last four ice ages. Although other atmospheric constituents, including dust, are preserved, of most interest are the measurements of CO_2 and CH_4. Both of these greenhouse gases increased suddenly in concentration during the sudden onset of an interglacial period and decreased gradually during the subsequent gradual cooling of the globe. The range of variation is between 200 and 280 parts per million by volume (ppmv) for CO_2 and between 0.35 and 0.65 ppmv for CH_4. Both gases contribute to the warming of Earth s surface by absorbing outgoing infrared radiation and returning heat energy to the ground, but their contributions to the greenhouse effect are not large enough to account for ice age changes in Earth s average temperature.

Much of the background CH_4 concentration in the atmosphere is a result of the decay of vegetation in poorly aerated swamps. It seems that swamps were more widespread during interglacial periods when there was more rain in the tropics. Glacial periods were characterized by dry conditions at low latitudes. CH_4 is destroyed in decades by chemical reactions in the atmosphere, so the concentration in the atmosphere also changes whenever the source strength changes. A change in global climate results in a significant change in atmospheric composition.

Explanations of the changing concentration of CO_2 remain controversial. The gas is exchanged between the atmosphere and other large reservoirs, the carbon dissolved in the ocean, and the carbon stored in living matter and decaying vegetation. It may be that ocean chemistry or circulation changes in such a way as to absorb more atmospheric CO_2 during ice ages and releases more during interglacial periods. Alternatively, the mass of carbon stored in forests, swamps, and soil may increase during ice ages and decrease during interglacial periods. It is quite likely that both processes contribute to the observed changes.

Millions of Years

As far as we know, there are no preserved samples of ancient air older than a few hundred thousand years. Inferences concerning atmospheric composition earlier in Earth s history depend on theoretical arguments and indirect indicators (*see* Carbon Dioxide Concentration and Climate Over Geological Times, Volume 1).

The theoretical method begins with an analysis of the processes controlling the present composition of the atmosphere. If we understand these processes well enough, we can imagine how they may have changed in the past and can estimate how the atmosphere would have responded to the changes we imagine. Among the changes that need to be considered are the evolution of life and the metabolic processes that consume and release atmospheric constituents. The production of oxygen and consumption of CO_2 by photosynthesis is a prime example, but organisms release a multitude of other gases to the atmosphere, including CH_4, hydrogen, and nitrous oxide. Climate change affects temperatures and rainfall, and thus the abundance of living creatures and the rates of chemical reactions between atmospheric gases and between gases and rocks or gases and seawater. Climate also controls the circulation of the ocean, and thus the exchange of gas between ocean and atmosphere.

A third potential cause of change in atmospheric composition is geological activity, including the release of gas by volcanoes, the lifting up and eroding away of mountains, and the drifting of continents across the surface of the globe. Chemical interactions between atmospheric gases and the rocks and minerals of the solid Earth are slow, but the Earth contains so much more material than the atmosphere that these interactions determine atmospheric properties over geological times lasting hundreds of millions of years. Mountain building and continental drift affect the atmosphere indirectly through their influence on global climate and ocean circulation and directly through their influence on the composition of the rocks exposed to air and the rates at which these rocks react with atmospheric constituents.

Theoretical analysis can provide an indication of past atmospheric composition, but observational confirmation is certainly needed. The fossil record of life is one important source of information. As long as there have been animals on Earth that breathe oxygen, there must have been enough oxygen in the air, or dissolved in water, to sustain them. On these grounds, we can be confident that the atmosphere has contained more than a few percent of oxygen for the last billion years or more. This is not a trivial observation. Earth, like the solar system and the universe, is reducing in overall composition – it contains more than enough iron and other elements that react with oxygen to consume all of the oxygen in the atmosphere. Reactions between air and rocks are fast enough to deplete atmospheric oxygen in less than 10 million years in the absence of a source. That oxygen has never been depleted in more than a billion years implies that there must be a steady, reliable source of oxygen. The fossil record tells us the source is green plants, which release oxygen during photosynthesis. They have caused the atmosphere to become rich in oxygen and they have maintained enough oxygen to sustain animal life.

On the other hand, plants require CO_2. Carbon dioxide dissolves in rainwater and seawater, where it reacts with the products of rock weathering to precipitate as solid carbonate minerals. These processes are fast enough to remove all of the CO_2 from the atmosphere in less than a million years, in the absence of a source. Yet for billions of years there has been enough CO_2 in the atmosphere to sustain plant life. The sustaining source is Earth s hot interior. The high temperatures there drive CO_2 out of its solid compounds and release it as a gas from volcanoes and hot springs. Over periods of millions to billions of years, the amount of CO_2 in the atmosphere is set by a balance between this volcanic source and the consumption of CO_2 in reactions with cold rocks at the surface of the Earth. On the shorter time scales of hundreds to thousands of years, atmospheric CO_2 concentrations depend on exchange with the large oceanic and terrestrial living reservoirs of carbon.

Some idea of the past CO_2 concentration can be gained from paleoclimate studies because CO_2 is a greenhouse gas, warming Earth s surface. It appears that Earth has been cooling for the last 60 million years, a cooling that culminated in the onset of the ice ages two million years ago. This cooling is attributed, in part, to a decreasing concentration of CO_2 in the atmosphere, caused by some combination of reduced volcanic activity and an increased consumption of CO_2 in weathering reactions with rocks. The rate of weathering may have been increased by mountain building, which exposes fresh reactive rock at the surface, or by the activities of life, which may have evolved more aggressive ways to attack and dissolve rocks to obtain essential mineral nutrients.

Additional support for the notion that CO_2 has decreased comes from the fossil plant record and from measurements of carbon isotopes. Plants have recently, in geological terms, developed a photosynthetic pathway that is adapted to low concentrations of CO_2. This development may have been a response to the decreasing CO_2 concentration. Furthermore, photosynthesis favors the abundant, light isotope of carbon, tending to leave the rare, heavy isotope in atmospheric CO_2 while producing organic matter enriched in the abundant, light isotope. This fractionation of the carbon isotopes is strongest when CO_2 is abundant. Careful measurements of the isotopic composition of ancient organic matter in sediments compared with carbonate minerals of the same age reveal larger fractionation in the past, indicative of higher concentrations of CO_2.

No single argument is persuasive. The reconstruction of past atmospheric composition calls for a careful reconciliation of suggestive observations and incomplete theoretical understanding. At present, it seems reasonable that the concentration of CO_2 60 million years ago may have been as much as several times its pre-industrial level. This is an active area of research.

Billions of Years

Uncertainties are even larger further back in time. The state of the Earth in those days is hard to imagine and to analyze theoretically, and the early rock record is sparse because most old rocks have been destroyed by more recent geological activity. Research on atmospheric composition in the first few billion years of Earth s history has focused mainly on CO_2 and on oxygen.

The CO_2 discussion centers again on its climatic effects. The problem is straightforward. The Sun is getting brighter as it ages and gradually consumes its nuclear fuel. The young Earth, warmed by the faint, young Sun, would have been completely frozen without a greenhouse effect very much larger than that of the present day. Ancient sedimentary rocks make it clear that liquid water was abundant at Earth s surface as long ago as 3.8 billion years. Various greenhouse gases have been evaluated as potential solutions to the cool sun problem. None is as plausible as CO_2, possibly augmented by haze.

This analysis suggests an atmosphere four billion years ago consisting largely of CO_2 with a surface pressure of only a few bars.

Perhaps this suggestion should not surprise us. The atmospheres of Mars and Venus consist largely of CO_2, and the surface pressure on Venus is 100 bars. The volcanic sources of CO_2 would have been large when the Earth was young with a hot interior. The consumption of CO_2 in reactions with rocks may have been slow when there was little land exposed to the atmosphere and no terrestrial life to promote weathering. It is believed that the CO_2 concentration decreased through time as landmasses increased in area, the interior of the Earth cooled, and the Sun grew brighter, warming Earth s surface. Weathering reactions are faster at higher temperatures.

Deductions concerning oxygen are also based on a combination of theory and observation. The oxygen-rich atmosphere we now enjoy, which is unique among known planets, is sustained by the photosynthetic activities of green plants. There can have been little oxygen in the atmosphere before green plants evolved. Sedimentary rocks provide the best evidence concerning the rise of atmospheric oxygen. The preponderance of this evidence indicates an oxygen-free environment for the weathering and deposition of sedimentary rocks before about 2.3 billion years ago. Oxygen accumulated in the atmosphere at this time, but the deep ocean remained anaerobic for hundreds of millions of years. During this period, sediments continued to form, containing abundant iron, which must have been dissolved in seawater free of oxygen. The difference between atmosphere and deep ocean arose because the exchange of material between these reservoirs is slow, the deep ocean receives reduced compounds from Earth s interior, and the oxygen source, which requires sunlight, is located at the surface. By about 1.5 billion years ago, it seems that the photosynthetic oxygen source was large enough to overwhelm the source of reduced compounds in most parts of the deep sea. There have been speculative discussions about subsequent changes in atmospheric oxygen, but the evidence is not convincing.

ORIGIN OF THE ATMOSPHERE

Earth s atmosphere consists of gases that react with the materials of the Earth to form solid compounds. It is deficient, relative to the Sun, in the inert gases that do not react. This observation has been understood for nearly half a century to imply that Earth s materials accreted originally as solid compounds, not as gases, and that high temperatures inside the Earth drove the gases of the atmosphere, along with ocean water, out of these compounds. It was, at one time, thought that Earth was initially cold, and that atmosphere and oceans accumulated gradually as the interior of the planet heated up. It now seems more likely that Earth had a hot origin followed by gradual cooling. This view implies that most of Earth s volatile constituents were driven out of the interior to form a dense original atmosphere while the planet was accreting. This atmosphere was probably composed mainly of steam. When the planet stopped releasing gravitational heat, the steam condensed to form oceans, and the remaining gases, mainly CO_2, began to react chemically with water and with rock to form dissolved and solid compounds. The components of Earth s atmosphere and ocean are gradually returning to the interior of a cooling planet.

See also: Carbon Dioxide Concentration and Climate Over Geological Times, Volume 1; Climate Model Simulations of the Geological Past, Volume 1; Earth System History, Volume 1.

Atmospheric Composition, Present

Peter Brimblecombe
University of East Anglia, Norwich, UK

Although the atmosphere can be seen as an inert matrix of nitrogen and oxygen, many of the gases found at lower concentration are highly reactive, with only short lifetimes, such that their concentrations are maintained through a balance between sources and removal processes. Sources may be geochemical activities that lead to volcanic emissions or wind-blown dust and sea-spray. Biological releases, particularly those from microorganisms, are important in producing a wide range of organic trace gases and human activities can now affect atmospheric composition on a global scale, re ected most importantly in the carbon dioxide driven greenhouse effect.

The composition of the atmosphere was of interest to scientists from the earliest times. The ancient Greeks regarded the air as an element, although educated farmers were already aware that it carried other substances that could damage or benefit crops. With the growth of modern science in the 17th century, there was further interest in the atmosphere. Robert Boyle, remembered for Boyle s Law, studied the composition of the air and was aware that there were minute trace components that contributed to its corrosiveness.

In the 18th century the Phlogiston Theory was developed to explain the properties of air, but the discovery of oxygen

by Scheele and Priestley overthrew this theory. Dalton s atomic theory gave a better understanding of air that was proving to be a mixture, not a compound, but a formula such as $N_{15}O_4$ seemed unlikely. Evidence against air being a compound became so strong that the notion had to be discarded.

We now know air to be composed of nitrogen and oxygen, which occupy almost constant proportions throughout the atmosphere and make up almost 99% of its dry volume (Table 1). Oddly enough, carbon dioxide, a much lesser component, was discovered much earlier than either oxygen or nitrogen, when the Scottish chemist Joseph Black investigated the properties of carbonate minerals in the 1750s.

Throughout the 19th century scientists became aware of many trace components in air: ozone, ammonia, hydrogen sulfide, sulfur dioxide, etc. However, the most surprising discovery came at the end of the century when Ramsay discovered argon in the atmosphere, which made it the third most abundant compound in dry air. Argon is formed from the radioactive decay of potassium-40 over very long time periods. This revelation was gradually followed by the discovery of a number of other inert gases.

So far, we have stressed that the air is dry. This is not true, because air also contains water vapor, but at highly variable concentrations. The amount of water is dependent on temperature and relative humidity. At 20 °C and 100% relative humidity, air is about 2.3% water vapor.

In recent decades, scientists have come to realize that the atmospheric composition has a more dynamic nature, with none of the constancy assumed when looking at the major gases, such as nitrogen, oxygen and argon. It is often convenient to think of the gases in the atmosphere as falling between two extreme types, stable and highly reactive gases.

Long-lived gases are at a relatively constant concentration throughout the atmosphere. Generally these are the gases that are unreactive – nitrogen, oxygen and the noble gases (helium, neon, argon, krypton, xenon).

Table 1 The concentration of some components of dry tropospheric air

	ppm
Nitrogen	781 000
Oxygen	209 000
Argon	9340
Carbon dioxide	360[a]
Neon	18
Helium	5.24
Methane	1.75[a]
Krypton	1.14
Hydrogen	0.5
Nitrous oxide	0.3[a]
CFC-11	0.0003[a]

[a] Changing due to human activities.

Many other gases are not so stable, but still long-lived enough to be well mixed in the atmosphere.

These include carbonyl sulfide, nitrous oxide, methane, and carbon monoxide. Although these gases have lifetimes in the order of 1 year, they are maintained at steady state concentrations by production at the Earth s surface. The sources are often geochemical or biological. Their long lifetimes mean that they may not necessarily degrade in the lower atmosphere, but cross the tropopause into the stratosphere, where they make contributions to stratospheric chemistry. Methane contributes to water vapor in the upper atmosphere, carbonyl sulfide to the sulfate aerosol layer and nitrous oxide is involved in stratospheric ozone depletion.

Gases that have lifetimes of just a few days tend to be present at highly variable concentrations in the atmosphere. We can understand this in terms of radon-222. This is a radioactive gas that emanates from land. It decays with a half-life of 3.8 days, so it is easy to imagine that its concentrations far out over the ocean will be low.

The oceans cover 70% of the Earth s surface and represent an enormous source of trace gases for the atmosphere. Seawater has high concentrations of chloride and sulfate, so it is hardly surprising to find that the oceans are large sources of sulfur and chlorine-containing trace gases. In particular we find hydrochloric acid, derived from sea salt above the oceans, and alkyl chlorides important. The long-lived carbonyl sulfide also comes from the ocean.

The oceans are far less rich in nitrogen, so we often look to the land as a predominant source of nitrogen compounds. Nitrogen oxides arise from soil respiration. Ammonia comes from the degradation of proteinaceous material or urea. Wetlands also prove an important source of methane. They are also sources of other trace gases that seem especially unusual: dimethyl selenides or diphosphane. This latter gas may ignite the methane from marshes and be responsible for will-o-the-wisps.

Human inputs also change the composition of the atmosphere. The combustion of fossil fuels has caused the concentration of carbon dioxide to rise by 70 ppm in the last century (*see* Carbon Dioxide, Recent Atmospheric Trends, Volume 1). This change in concentration is recognized as enhancing the greenhouse effect (*see* Greenhouse Effect, Volume 1). The concentrations of other greenhouse gases are also on the increase. Methane has increased substantially over the last few centuries, driven by increased agriculture: cattle and paddy fields are important sources, but mining is also significant (*see* Methane, Volume 1). Carbon monoxide has natural sources, but human activities have probably been responsible for the increase to a peak in the early 1990s, but levels have declined since then.

Nitrous oxide has also been on the increase (*see* Nitrous Oxide, Volume 1).

Anthropogenic gases, such as the chlorofluorocarbons (*see* Chloro uorocarbons (CFCs), Volume 1), cause so much concern because, although unreactive in the lower atmosphere, they degrade the ozone layer (*see* Stratosphere, Ozone Trends, Volume 1). These gases have shown sharp changes in concentration. First they rose, as their use increased, and subsequently declined as production was curtailed in response to various international protocols. They have often been replaced by hydrofluorocarbons and hydrochlorofluorocarbons that are now released in substantial amounts, but have smaller long-term impacts on global atmospheric chemistry.

FURTHER READING

Brimblecombe, P (1997) *Air Composition and Chemistry*, Cambridge University Press, Cambridge.

Atmospheric Composition, Recent Changes

see Air Quality, Global (**Volume 1**); Trends in Global Emmisions: Carbon, Sulfur, and Nitrogen (*Opening essay*, **Volume 3**)

Atmospheric Electricity

R Giles Harrison and Karen L Aplin
University of Reading, Berkshire, UK

By atmospheric electricity, we usually mean the composite effect of all electrically charged atmospheric matter, which includes molecular ions and charged aerosol particles, as well as the charge exchanged between the ice, liquid and vapor phases of water. Global climate change may increase planetary thunderstorm activity. Associated with this could be a change in lightning distribution that would affect, in turn, the xation of atmospheric nitrogen, carbon dioxide concentrations from lightning-initiated forest res and atmospheric ozone concentrations. Ions owing in the atmospheric electrical circuit may be implicated in condensation nucleus formation; thus the atmospheric electrical system may also have an indirect effect on cloud formation.

PLANETARY ELECTRIFICATION

Earth is not alone in having an electrified atmosphere. If we regard lightning as a key signature of atmospheric electrification, then at least Jupiter, and Saturn s largest moon – Titan – probably qualify. In order for a planet s atmosphere to become electrified, it seems it must contain polar molecules (such as ammonia or water) present in different phases, and strong convection must be taking place, usually driven by powerful differential heating. Lightning is not a recent phenomenon on Earth. When lightning strikes sand it fuses it into a glassy column, which, if preserved, becomes a fossil. Such fossilized lightning strikes, or *fulgurites*, have been found in sediments dated to be at least 250 million years old.

Two pioneering 18th century scientists are particularly associated with early work on atmospheric electricity and lightning. Benjamin Franklin (*see* Franklin, Benjamin, Volume 1) is well known for his kite-flying experiments in which he discovered the similarity between atmospheric charge and that produced in laboratories artificially, but John Canton was performing much more sensitive experiments in England at a similar time. In 1754, Canton observed that clouds that didn t produce lightning could also be electrified, and could carry positive or negative charge. He also identified hail as being important in the electrification of thunderclouds.

THE ATMOSPHERIC ELECTRICAL SYSTEM

The atmospheric electrical system arises because of the vast differences in scale between the different forms of charge-exchange occurring on the planet: charge is principally generated in vigorous thunderstorms, and dissipated elsewhere in regions where the weather is less disturbed. If we regard the thunderstorms, mostly occurring in the tropics, as the source of charge, rather like a battery, and the currents flowing in the undisturbed regions as the sink of the electrical energy, then it is possible to construct a simple electrical circuit (Figure 1). Although simplistic, this model is a useful description of the system. The current flowing in the circuit is not constant, although typically about 2000 Amperes, as it depends on the global activity of thunderstorms, which varies as a result of surface heating. An ampere (often abbreviated to amps or A) is the unit of electric current, which is the rate at which electric charge is passed in a circuit. Common domestic electrical currents vary from a few thousandths of an amp to operate a digital clock, to tens of amps to operate an electric cooker. Monitoring the total current flowing by measuring the voltage of the upper atmosphere, using balloon-carried probes and sounding rockets, shows that there is a daily cycle in the voltage associated with a similar cycle in global thunderstorms as the planet rotates.

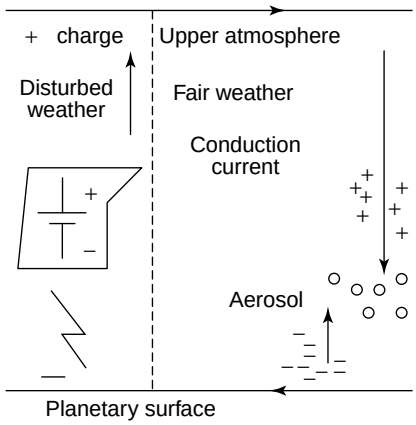


Figure 1 Simplified schematic description of the atmospheric electrical circuit. Charge separation in thunderstorms causes a current of about 2000 A to flow between conducting regions in the upper atmosphere and the surface, causing the upper atmosphere to charge positively to about 300 kV. The majority of the 2000 A current arises from charge exchange in lightning flashes: a smaller fraction occurs from non-lightning point discharges beneath clouds. It is not yet known how above-cloud discharges (e.g., sprites) contribute to the global circuit. In fair weather regions (which cover a much greater area of the planetary surface than thunderstorms), a compensating conduction current flows, consisting of ions formed in the atmosphere by cosmic ray-induced ionization. Aerosol in the lower atmosphere removes some of the ions and reduces the local electrical conductivity

This cycle was first identified from the research voyages of the vessel Carnegie in the 1920s. The Carnegie was a research ship entirely constructed of wood so that it could be used for geophysical (including terrestrial magnetic and electrical) studies. No metal was permitted on board. It was ultimately destroyed by fire. This voltage variation has since been shown to follow the variations in thunderstorms caused by the sequential heating of continental landmasses as the planet rotates. Thunderstorms generated on the African landmass dominate the global electrification in the late afternoon and early evening, Greenwich time.

The ionosphere and the Earth s surface also form a cavity within which radio waves can propagate, although the radio properties of the ionosphere vary with the time of day and the solar cycle. Lightning discharges generate radio waves at a variety of frequencies, some of which can be detected at great distances from the lightning discharge and can therefore be used in lightning location systems. Of particular interest is the *Schumann resonance*: this is an extra low-frequency (8 Hz) radio wave, which has a wavelength equal to the equatorial circumference. The amplitude of this radio signal can be used to monitor the amount of global lightning.

THUNDERSTORMS

Thunderstorms occur when large cumulonimbus clouds become electrically charged. The exact mechanism for this charging has been disputed for many years, but it is increasingly accepted that it is due to electrical interactions in the collisions between soft hail (known as *graupel*) and ice. The falling graupel carries negative charge downwards, and ice crystals rise in vigorous central updrafts to the upper parts of the cloud, where they form the characteristic anvil-shaped tops associated with such clouds. The base of the cloud often contains some positive charge (Figure 2). Charging continues for the life of the thundercloud, which is about an hour, and lightning flashes within and from the cloud result.

LIGHTNING

Lightning occurs when the electric fields generated by the cloud are so intense that air ceases to be an insulator (*see* Lightning, Volume 1; Lightning and Atmospheric Electricity, Volume 1). Breakdown of air occurs and a transient electrical discharge – lightning – results. A typical lightning strike may be several kilometers in length. Lightning exists in many forms. If it occurs within clouds it is known as *intracloud*, between clouds *intercloud*, below clouds *cloud to ground* (CG) and above clouds variously as *red sprites* or *blue jets*. Sprites and jets have only been studied in the last few years when satellite and space shuttle observations became available. Sprites are particularly intriguing forms of lightning because they occur on such massive scales – they are 5–30 km wide, and about 50 km high. Although not well understood, they are thought to be associated with positive lightning strokes to ground. Intracloud discharges are most common, but the greatest visual spectacle is CG, often seen as forked lightning, which accounts for about 20% of discharges. A CG strike leads to tens of Coulombs of charge, most commonly negative, flowing from the cloud. (The Coulomb is the unit of electric charge. One ampere is the current flowing when one coulomb passes each second.)

CG lightning begins as a weak initial discharge (the stepped leader), which, after reaching the surface, is followed by the luminous return stroke. A stepped leader propagates at about $100\,\mathrm{km\,s^{-1}}$, and within a narrow channel of about 5 m. When the stepped leader is near the surface, intense local electric fields are generated, and charges (streamers) begin moving upwards to the leader. If the upward streamer charge contacts the leader, the vigorous return stroke results, and currents of tens of thousands of amps flow within tens of microseconds. During the remainder of the discharge, the current falls to hundreds of amps for a longer period of several milliseconds. If there is no more charge in the cloud, the lightning discharge will then

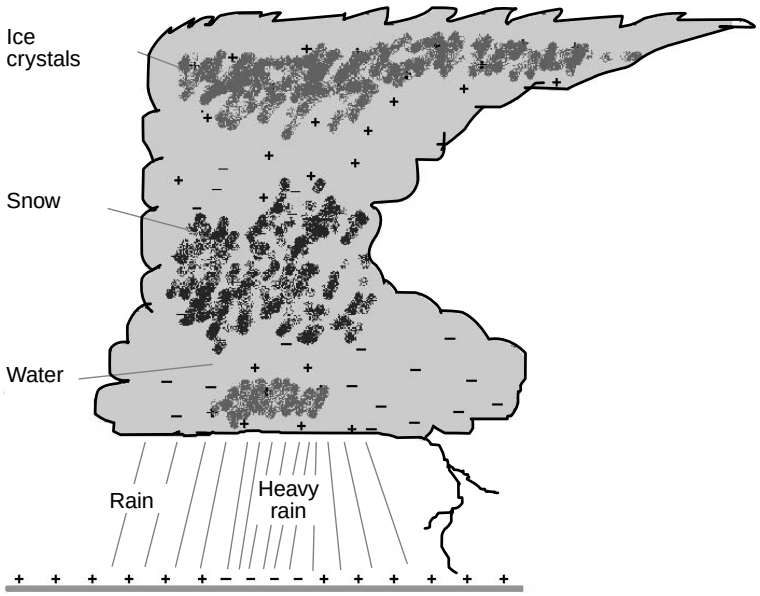


Figure 2 Thunderstorms are the most spectacular demonstration of the interactions between meteorology and electricity. Charging occurs because of collisions between ice crystals and soft hail (graupel) in the intense vertical updrafts within a deep convective cloud. The upper anvil of a thundercloud contains positively charged ice crystals, and the middle part of the cloud negative charge. There is a small positive charge in the base of the cloud, the temperature of which is usually above the freezing threshold

end, but it is possible for several strokes to flow along the same ionised channel. Crackles due to the electrical energy from thunderstorms, with fundamental frequencies of about 10 kHz, can be picked up by radio receivers and are known as *sferics*. Radio location of thunderstorms is used to assist forecasts of disturbed weather.

Associated with the return stroke channel are temperatures and pressures much higher than those of the surrounding air, and the channel expands, generating an acoustic shock wave. This shock is heard as thunder, some time after the lightning is seen because the speed of sound in air is less ($\approx 300\,\mathrm{m\,s^{-1}}$) than the speed of light. At significant distances from the storm, the sound is more reminiscent of a rumble rather than a bang, because of acoustic refraction along the sound path. Thunder lasts from about one-tenth to $1-2$ s.

Until recently, it was not possible to monitor lightning occurrences globally, but a new National Aeronautics and Space Administration (NASA) satellite instrument, the optical transient detector, can record lightning from space in daylight. These new observations have led to data such as the annual total of detected lightning flashes across the world, as shown in Figure 3. Because the satellite orbits the Earth frequently, daily variations in thunderstorm activity can also be extracted from its observations. This system has detected electrical activity patterns very similar to the ones observed during the Carnegie voyages. Global flash rates are now thought to be about 50 flashes $\mathrm{s^{-1}}$, although the precise figure currently remains uncertain. It is considered important to quantify this more thoroughly in the context of global change, because the flash rates may vary with local surface temperature, and therefore perturb the nitrogen cycle and ozone production.

FAIR WEATHER ELECTRIFICATION

Although thunderstorms are the most dramatic and apparent manifestation of atmospheric electricity, undisturbed (or fair weather) electrical regions cover most of the planet for most of the time. Figure 1 illustrates the atmospheric electrical system, in which the fair weather component was seen essentially to provide the conduction current. The conduction currents can only flow because there are charged particles (small ions) present that are sufficiently mobile to be accelerated by the small electric fields present in the atmosphere. Small ions are formed by bombardment of air molecules by natural radioactivity and cosmic rays, making the air slightly electrically conductive. The air also contains a myriad of small particles with a spectrum of sizes, from cloud droplets down to molecular clusters, known as aerosol. Small ions can become attached to aerosol particles, causing them to become electrically charged. Some aerosol particles can act as nuclei for cloud formation, and it is also apparent that clusters of small ions can facilitate the production of ultrafine aerosol particles. Fair weather electricity is, therefore, also increasingly becoming identified as relevant to the climate system.

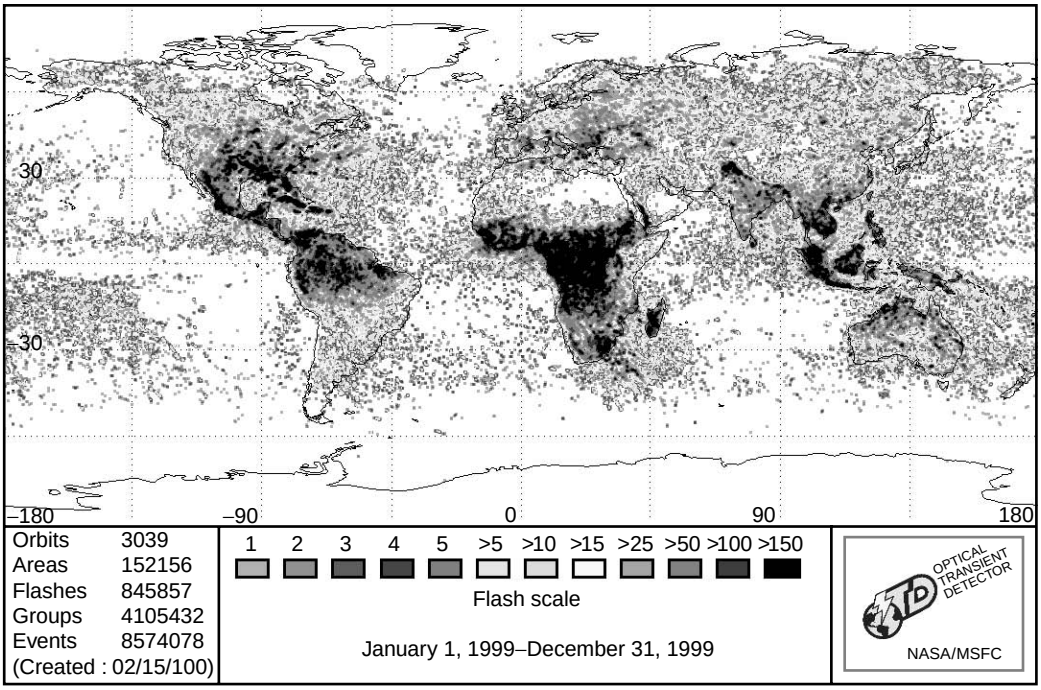


Figure 3 Annual global lightning flashes for 1999, as determined by the NASA satellite-borne optical transient detector. The lightning flash scale indicates the number of flashes per square kilometer per year. Darker regions indicate a greater number of flashes. (Data provided by the Global Hydrology Resource Center at the Global Hydrology and Climate Center, Huntsville, AL (see http://thunder.msfc.nasa.gov/data))

EFFECTS OF GLOBAL CLIMATE CHANGE

The effects of global climate change are uncertain, but one possible implication is a change in the planetary thunderstorm activity. This could lead to a modified lightning distribution, which may have significant feedbacks, because lightning affects the fixation of atmospheric nitrogen, forest fire initiation and ozone concentrations. If climate change does affect lightning, this may indirectly affect the other components of the global atmospheric electrical circuit, although the magnitude and sign of the effect are difficult to determine. For example, the ions comprising the fair weather current flow in the circuit may be relevant in condensation nuclei formation, which may influence cloud formation and properties. The effects of global climate change on such complex and sensitive processes are not yet known.

FURTHER READING

Chalmers, J A (1967) *Atmospheric Electricity*, 2nd edition, Pergamon Press, Oxford.

MacGorman, D R and Rust, W D (1998) *The Electrical Nature of Storms*, Oxford University Press, Oxford.

Rogers, R R and Yau, M K (1989) *A Short Course in Cloud Physics*, 3rd edition, Pergamon, Oxford.

Atmospheric Electricity, Relation to Lightning

see Lightning and Atmospheric Electricity (**Volume 1**)

Atmospheric Model Intercomparison Project (AMIP)

see AMIP (Atmospheric Model Intercomparison Project) (**Volume 1**)

Atmospheric Motions

Anders Persson

European Centre for Medium Range Weather Forecasts (ECMWF), Reading, UK

With the first pictures looking down on the Earth from a space platform in the 1960s, the beauty and complexity of

the atmospheric motions were immediately evident: the long streaks of jet stream clouds, the large spirals of developing mid-latitude storms, and the strings of pearls of towering thunderstorm clouds in the equatorial regions.

It was a beautiful view, but it was not entirely unexpected. By the end of the 19th century meteorologists had managed to deduce the essential parts of the global circulation, from soundings by balloons and cloud drift observations at higher levels. Steady improvements of observational technology, radar, remote sensing and, above all satellite observations have provided an ever-increasing amount of data on the state and motion of the atmosphere, not only in the lowest 10- 15 km (the troposphere), but also at higher levels (the stratosphere, mesosphere and ionosphere).

There were surprises to come. In the 1890s, when balloon soundings started to reach above 10 km, they recorded that the temperature, from having rapidly decreased, suddenly at some height became more or less constant. This marked the discovery of the tropopause, the cold dividing boundary between the troposphere and the stratosphere. During the First World War, when the thunder on the western front could be heard in London, meteorologists began to speculate that the only explanation must be a warm layer at a height of about 50 km; the existence of the stratopause, the warm boundary between the stratosphere and the mesosphere was later con rmed by more advanced measurements. The third great surprise was of course the jet streams: long, narrow bands of very strong winds of 30 m s^{-1} or more at high altitudes. In the 1960s observations from 30- 50 km height in the equatorial region revealed that the winds turned from westerly 20 m s^{-1} to easterly 50 m s^{-1} over a period of, oddly, 13 months. In recent decades the interactions between the oceans and the atmosphere, most notably the El Niño, have added to our understanding of the complexities of the atmosphere.

The main task for meteorology is to explain, or at least describe, the general circulation of the atmosphere and its variations in accordance with the basic principles of uid dynamics, radiation and thermodynamics. To what extent are the processes that give rise to these complicated circulation systems really understood? The answer depends to a great extent on what is meant by an explanation.

INTRODUCTION

Computer Models

In 1956, Norman A Phillips, an American meteorologist, achieved something remarkable: he fed a computer with the basic facts about the size of the Earth, its gravitational pull, its rotational velocity and the chemical composition of the gases that constitute its atmosphere. Programmed with simplified versions of the mathematical equations that describe the atmosphere s motions and physical processes, the computer was able to generate the observed three-dimensional general circulation of the Earth s atmosphere, including jet streams, travelling cyclones, fronts and high-pressure areas.

Doubling of computer power every one and a half years has since then provided an ever increasing ability to simulate and forecast most atmospheric features. Numerical simulations of the general circulation using supercomputers serve as an atmospheric laboratory, in which we can perform experiments, observe the effects and hopefully predict its future state. However, many meteorologists are not satisfied and think that it is not enough that the computers seem to understand. We want to know, for example, why the westerly winds so dominate the mid-latitudes, why the Sun s radiation reaches the Earth s surface without being hindered by clouds in the subtropical areas, why northwestern Europe s winters are so mild compared to those at the same latitude in northeastern North America and why the storms in the tropics are so different from the storms that rage over the Atlantic or the Pacific.

Watching Atmospheric Motion

Our understanding of the atmosphere is like that of the spectator at a football match. Without any insight into the rules of the game, it will only appear to be a lot of people randomly running around chasing a ball. Those who know the rules and the logic of the game might not be able to make predictions of what will happen in the next minute, but can make sense of what is unfolding in front of their eyes.

Our knowledge of the atmosphere is often summarized in a statistical way. But using statistics brings advantages and disadvantages. We gain in simplicity, but also lose because we have thrown something away. Whenever there are averages, there are fluctuations or deviations from these averages. In addition, a mean picture does not necessarily represent a typical picture. In environmental studies, we are for example interested to know the evolution and trajectory of individual air parcels: Where does the moist air go? From where comes the polluted water? What makes the air lose its ozone? One good illustration of how knowledge can enhance understanding is the movement of a water droplet in a simple water wave.

Two Ways to Look at Motion

Consider a train of waves on a water surface moving from left to right, as depicted in Figure 1(a). The fluid is rising as the ridge of the wave approaches and falling after it passes. The mean motion, as inferred by averaging

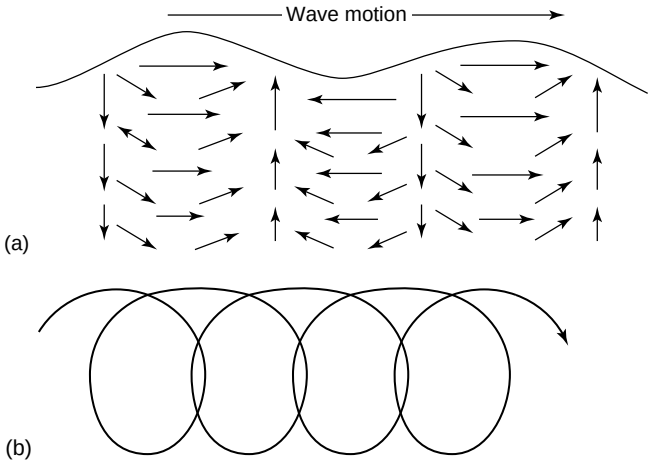


Figure 1 (a) A train of waves moving from left to right. For further discussion, see text. (b) The individual particles in the water are subjected to a net displacement to the right

the motions at individual points in space over a large number of waves, will everywhere be zero because all motions, vertical and horizontal in the long run cancel out (sometimes called *Eulerian averaging*). However, if we concentrate on individual parcels, we can see that they move in clockwise circuits. If we look carefully, we see that the extent of the motion decreases with depth, so we may realize that parcels move a little farther towards the right at the top of their path, than they move toward the left farther down. Hence, there is a net rectifying displacement toward the right, as pictured in Figure 1(b). To understand the motions of the air we must therefore also make use of averages that are geared to the movement of individual air parcels (sometimes called *Lagrangian averaging*).

Motion without Fricton

The motions of the atmosphere show patterns, regularities and mutual relationships, which make sense if certain basic mechanisms are understood. Many of these patterns are easily accommodated with everyday experience: the bubbles in a pan of boiling water resemble the production of convective clouds; the downhill flow of water resembles the katabatic winds that flow down mountain slopes; etc. What is difficult is the seemingly simple. Many processes, particularly in the free atmosphere, occur under almost frictionless conditions. Pure inertia is actually a very difficult process to comprehend with our senses because our everyday life very much involves friction. The fact that the motion takes place on a rotating planet also has some counterintuitive consequences.

These energy conversions occur in the interplay between gravity, the effect of the Earth s rotation and friction.

THE DRIVING AND RETARDING FORCES

The Sun provides the energy that drives the atmospheric circulation, but this energy is not evenly distributed over the globe. While gravity acts to equalize these differences by setting the air in motion, friction seeks to bring all motion ultimately to rest.

The Incoming Radiation from the Sun

Understanding the atmosphere involves the unraveling of a lot of paradoxes. It is, for example, not primarily the Sun that warms the atmosphere, but the Earth. Or rather the Sun warms the air indirectly by first warming the surface of the Earth, which then warms the air. The radiation from the Sun is short wave (visible) and the cloud-free atmosphere allows much of it to pass through without much absorption. Of all the solar energy that initially enters the Earth s atmosphere, about half reaches the surface to warm the land masses and oceans. About 30% of incoming solar radiation is reflected back to space, mostly by clouds, snow and ice; and about 20% is absorbed in the atmosphere (*see* Energy Balance and Climate, Volume 1).

When the land and oceans heat up, they emit long-wave (infrared) radiation, to which the atmosphere is not very transparent. The atmospheric absorption of long-wave radiation is due to the presence of radiatively active gases, the most important of which is water vapor. Together with carbon dioxide, ozone and the other trace gases (methane, nitrous oxides, chlorofluorocarbons, etc.), these gases account for the normal greenhouse effect, without which most parts of the Earth would be uninhabitably cold because much more radiation would leave the atmosphere and much less downward long-wave radiation would warm the surface (*see* Infrared Radiation, Volume 1). Although there is a debate about the magnitude of the *enhanced* greenhouse effect caused by human emissions, there is no such debate about the greenhouse effect as such (*see* Greenhouse Effect, Volume 1). What further complicates this already complex balance of in- and out-going radiation is that it depends on latitude, the seasons and the distribution of land and sea.

The Uneven Heating and Cooling

The farther away from the tropics, the lower the Sun rises over the horizon. This means that the Earth s surface is not as efficiently heated in the higher latitude regions as in lower latitudes. The incoming solar radiation depends also on the orientation of the hemispheres relative to the Sun; thus it changes with the seasons. The summer hemisphere

receives more insolation than the winter hemisphere; during the summer the polar regions receive as much solar radiation per day as the tropics!

The outgoing radiation is, however, less dependent on latitude. Because land and sea have different physical properties, they react differently to the influx of solar radiation. Land has a comparatively small heat capacity and its immobility means that heat can only spread through conduction and diffusion. Land surfaces can therefore go through temperature changes on the order of several degrees within hours down to depths of several centimeters. Water bodies have a much greater heat capacity and a lower albedo (reflectivity) than land; water s mobility also enables an efficient transfer of heat vertically as well as horizontally. For these reasons, oceans absorb more solar radiation and to greater depth than land. The vertical mixing in the ocean spreads the heat deeper, so that the sea surface temperature changes little even on seasonal time-scales (*see* Sea Surface Temperature, Volume 1). In the latitude band up to 40° north and south the continents are on average warmer than the oceans; poleward of 40° the oceans are warmer. These differences in heating and cooling create differences in the density and the mass distribution of air. It is these contrasts that gravity tries to equalize.

Gravity – The Great Equalizer

If the ultimate source of energy is the Sun, the ultimate mover is the Earth s gravitational pull. Gravitation is always trying to turn everything into perfect spheres, including the Earth and its atmosphere. Due to the centrifugal effect of the Earth s rotation, its shape is a sort of oblate ellipsoid, bulging outwards at the equator (this shape is referred to as the *geoid*). Any material surface that does not follow such a geoid, but stands out against it is subject to a gravitational pull.

Pressure falls off more slowly with altitude in warm air than it does in cold air. Therefore, the same atmospheric pressure is found at higher elevations in the tropics than in the polar region. If the pressure at a certain level is not horizontally uniform, gravity accelerates the air from high to low pressure to equalize the pressure distribution. We talk about this acceleration as due to the *pressure gradient force* (PGF), which is not any new kind of force, just the name we have given to the effects of gravity in this context.

Gravity also generates vertical motions by making colder and heavier air sink and warmer and lighter air rise. Rising air experiences decreasing pressure with height, and so expands and cools; sinking air encounters increasing pressure, and so is compressed and warms. Rising air therefore tends to have a cooling effect on the environment, sinking air to have a warming effect. One important exception is, as we will see, rising moist air where the water vapor condenses.

The Three Phases of Water

Heat melts ice into water, and vaporizes water into a gas. The energy used for these processes is stored in liquid water and water vapor, respectively. When water vapor condenses into liquid water, or water freezes into ice, this *latent heat* is released again. A rainfall of 8 mm, which is the same as 8 liters per square meter, releases the same amount of energy as the same area would have received from the Sun during one day. The same happens during snowfall when liquid water freezes to ice, although the amount of energy released is less. A large proportion of the energy received from the Sun is stored in this way and this plays an important role in the dynamics of the Earth s atmosphere because energy can be transported long distances before it is released as latent heat.

Ice, water and water vapor are important also for their reflective characteristics. The proportion of solar radiation that is reflected back to space (albedo) varies significantly with location because it depends on the proportion of ice-covered land and sea, and on the ice crystals in the upper parts of the clouds. Water vapor also absorbs and emits radiation differently from liquid water. The water drops and ice crystals in the clouds reflect the Sun s short wave radiation, which prevents it from reaching the Earth s surface. The expected cooling of the air beneath the cloud is to some extent compensated by warming from long-wave radiation that is trapped by the same cloud.

Finally, water has a high specific heat capacity; that is, water takes a lot of energy to warm, and it cools slowly. Energy stored in the oceans at tropical latitudes does not dissipate rapidly when transported to higher latitudes, serving as a regulator and conveyor of energy. Without these ocean currents, the temperature difference between poles and equator would be 90–100 K instead of the 40–45 K that is observed. The energy conversions related to water in all three forms, and their different radiative properties, contribute to making predictions of any change in the climate very difficult.

Friction

Friction plays a double role: on the one hand it is slowing down all motion; on the other hand, it is through friction (and gravity) that the atmosphere feels that it is on a rotating Earth. Large-scale friction like lee-effects of the large mountain ranges affects the atmosphere s tendency to form large-scale vortices.

THE RELATIONSHIP BETWEEN WIND AND PRESSURE

When the PGF starts to equalize by accelerating the air from high to low pressures, the rotation of the Earth tries

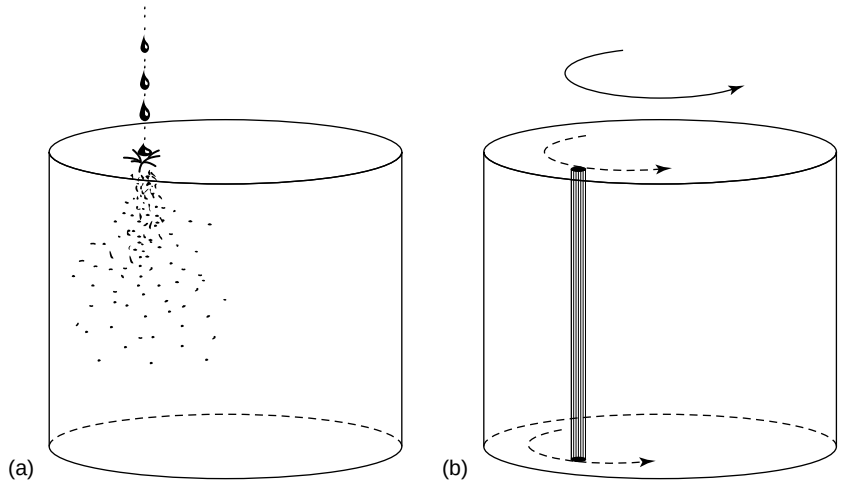


Figure 2 (a) Ink inserted into a non-rotating water tank would disperse immediately. (b) The same ink inserted into a rotating tank would, due to the Coriolis effect, form vertical columns. Such vertical columns are called ''Taylor columns'' after the British scientist who first conducted this experiment

to return the air from where it came. The struggle between these two contesting forces shapes the pressure and wind patterns.

The Effect of the Rotation of the Earth on an Air Parcel

Due to the Earth s rotation, wind and ocean currents are deflected to the right in the Northern Hemisphere and to the left in the Southern Hemisphere. This is due to the so-called *Coriolis force* (fV, where f is the Coriolis parameter and V the wind speed) (*see* Coriolis Effect, Volume 1); it has the effect of opposing any displacement by trying to restore the air to its initial position.

A striking demonstration of this can be made using a rotating tank of water. When a drop of ink is inserted in the water, it falls slowly downward. But instead of dispersing and coloring the water, it remains in a vertical column moving around with the tank as a rigid body. What happens is that when the ink particles start to spread out horizontally, they are immediately affected by the Coriolis force (fV), which forces them into circular motions. If the tank rotates with one revolution in two seconds ($\omega = 2\pi/2 = 3.14\,\mathrm{rad\,s^{-2}}$), and the ink spreads out with a velocity of $1\,\mathrm{cm\,s^{-1}}$, the fV would yield circles with radius of less than 2 mm (Figure 2).

An airflow in the atmosphere of $10\,\mathrm{m\,s^{-1}}$, only affected by the fV, would be confined in a circular trajectory with a radius of 100 km, about the size of London or Los Angeles. In the equatorial regions, where the horizontal component of the fV is relatively weak, velocities of the order of $30\,\mathrm{m\,s^{-1}}$ would be needed to move an air parcel from the equator to poleward of 20 °latitude. Were it not for the existence of strong PGFs, it would be very difficult for air parcels in the atmosphere to move any considerable distance.

The Relationship Between the Wind and Pressure

With two effects, the PGF and the fV, pulling in opposite directions, it is no surprise that they quite often balance each other,

$$f \cdot V = PGF$$

If this is the case, the air moves by itself with no acceleration (Figure 3). Winds blowing under these conditions are called *geostrophic* (from Greek *geos* meaning Earth and *strophein* meaning turning). For such conditions,

$$V_{\mathrm{geostrophic}} = \frac{PGF}{f}$$

The farther away from the equator, the stronger the PGF has to be to balance the fV. That is why pressure maps over the tropics and subtropics contain relatively

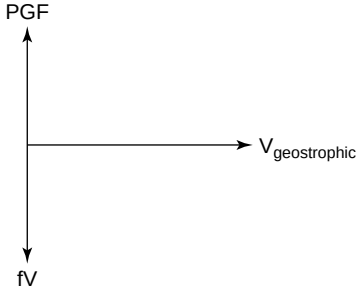


Figure 3 The wind is geostrophic if the PGF and the fV, balance each other

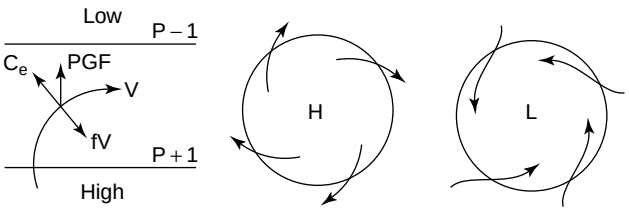


Figure 4 An air parcel accelerating into a pressure field will describe a curved motion and be affected also by a centrifugal force (C_e). A wind accelerated out from a high pressure area by a PGF will be deflected to the right by the fV and, after adjustment describe a clockwise circulation; a wind accelerated into a low pressure area by PGF will be deflected to the right by the fV and, after adjustment describe an anti-clockwise circulation

fewer isobars (lines of equal pressure) than at higher latitudes.

The Effect of Curved Flow

When the pressure field drives the air into a curved motion, it is affected also by a *centrifugal force* (C_e) that disrupts the bilateral balance between the fV and the PGF (Figure 4). Centrifugal forces are always directed outward from the center of rotation and, depending on how the trajectory is curved, the C_e will support either the fV or the PGF (Figure 5).

In a stationary high-pressure system (where the air moves in clockwise trajectories), the C_e supports the PGF; in a stationary low-pressure system (where the air follows anticlockwise trajectories), the C_e supports the fV. This is why high pressure systems, where the C_e supports the PGF, normally have weaker pressure gradients (fewer isobars); and low pressure systems, where the PGF must balance both the fV and the C_e, generally have stronger pressure gradients (more isobars).

Airflow in which the PGF, the fV and the C_e balance each other is called *gradient flow*. In small, intense vortices, with high velocities and small radii, such as in tropical cyclones, the C_e may dominate over the fV. These are called *cyclostrophic flows* (Figure 6).

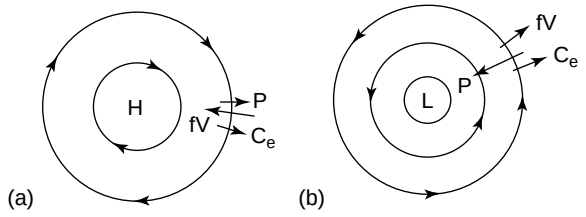


Figure 5 (a) The gradient wind approximation for cyclonic and anticyclonic flows. For curved air trajectories around a high pressure area (H), the C_e will support the PGF P; (b) for curved trajectories around a low pressure area (L), the C_e will support the fV

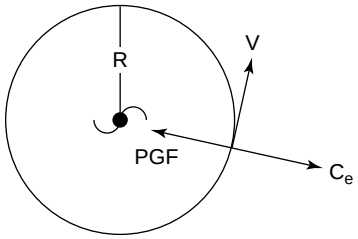


Figure 6 For very intense vorticies like mature tropical cyclones, the C_e dominates over the fV, which can be neglected, although it may have played a decisive role in the initial stages of the formation

The Effect of a Moving Flow Pattern

If the circulating system is not stationary, but moving, the relation between the C_e, the PGF and the fV is more complicated. The trajectories will alternately be more or less curved, compared to those of a stationary vortex. The reason why the winds on the southern side of a cyclone moving eastward close to geostrophic is because the air trajectories are rather straight (a large radius of curvature). On the other hand, on the northern side, the winds are much weaker, although the pressure gradients might be the same, because the air trajectories are more curved (Figure 7).

The Dynamics of Pressure Changes

Had the wind in the free atmosphere been in complete geostrophic, gradient or cyclostrophic wind balance, it would be frozen in a permanent state. What instead keeps the atmospheric circulation changing is that these balances are never fulfilled. It is the *ageostrophic* wind component, the part of the wind that is not in geostrophic balance, that drives the atmospheric patterns. If the wind at some instant happens to be in geostrophic balance, it sooner or later moves into a region where it is no longer in geostrophic balance (Figure 8). The PGF and the fV then provide a mechanism for restoring the balance by moving mass horizontally between lower and higher pressure.

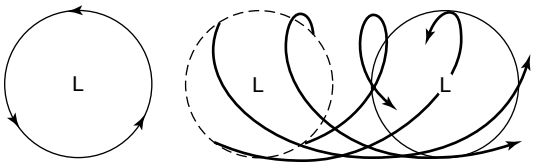


Figure 7 For an eastward moving circular vortex the trajectory of an air parcel on the southern side (where the air is moving in the same direction as the vortex) will have less curved trajectory than on the opposite side, where the air is moving in the opposite direction the trajectory will be more curved

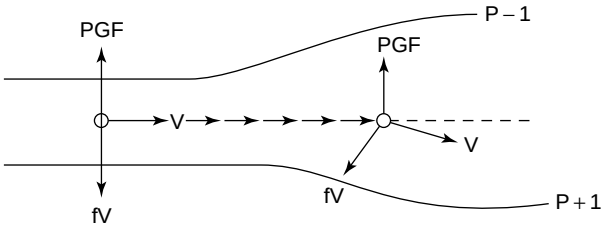


Figure 8 The wind is only temporarily in geostrophic balance; the motion of the air soon brings it into areas where this balance is no longer valid. In the schematic picture the wind is in geostrophic balance upstream where the pressure gradient is stronger than further downstream. When the wind enters this downstream region, it becomes supergeostrophic, and is deflected to the right by fV. The flow will thereby transport air towards higher pressure, which will strengthen the pressure gradient at the same time as the wind weakens it, until a new balance is reached

If the wind is sub-geostrophic, i.e., weaker than the geostrophic balance requires, the PGF will accelerate the air towards lower pressure. Mass will be transported from higher to lower pressure and thereby weaken the PGF. At the same time the wind will increase, and so will the fV, until a new temporary balance with the PGF is established (Figure 9a).

If the wind is supergeostrophic, i.e., stronger than the geostrophic balance requires, the fV will drive air towards higher pressure. Doing so, an air mass will be transported from low to high pressure, thereby increasing the PGF. At the same time, the wind will weakens and so will the fV, until a new temporary balance with the PGF is established (Figure 9b).

The Ice Skater Effect

An important mechanical effect in the atmosphere is the *ice skater* effect. By contracting or widening their arms, ice skaters increase or decrease their rate of rotation. Something similar happens for a body of gas (or fluid) which contracts or expands. But there are also vertical motions in the gas. Rising motion is associated with an inflow of air at lower layers and an outflow in upper layers. Sinking motion is associated with an inflow of air at upper layers and an outflow in lower layers. Due to the fV, the air flowing inwards forms an anticlockwise spiral; air flowing outwards forms a clockwise spiral. The pressure then adjusts to create a low pressure distribution for the inward spiral or a high pressure distribution for the outward spiral (Figure 10).

THE ATMOSPHERIC ENGINE

On average the equatorial region receives more heat from the Sun than is sent back to space, while the opposite is true at high latitudes. The accumulated result of the heating is an expansion of the air at low latitudes, setting the atmospheric engine in motion.

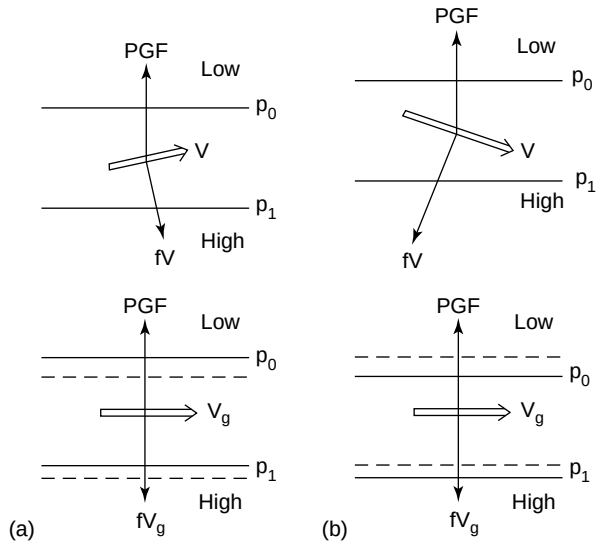


Figure 9 (a) When the PGF is stronger than the fV the wind *accelerates* towards lower pressure. At the same time the PGF weakens when mass is transported from high to low pressure. The wind moves parallel to the lines of equal pressure (isobars) when the two forces fV and PGF have reached a balance. (b) When the PGF is weaker than the fV the wind *decelerates* towards higher pressure. At the same time the PGF strengthens when mass is transported from low to high pressure. The wind moves parallel to the isobars when the two forces fV and PGF have reached a balance

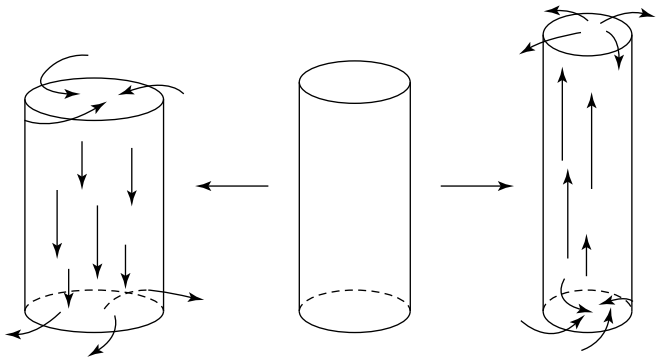


Figure 10 Schematic illustration of the ice skater effect. In the Northern Hemisphere, upward motion is associated with air spiraling anticlockwise inward at lower levels, clockwise outward at higher. For downward motion the opposite is the case

Hadley Circulation

Because pressure falls off more slowly with height in warm air than in cold air, the same pressure in the atmosphere is

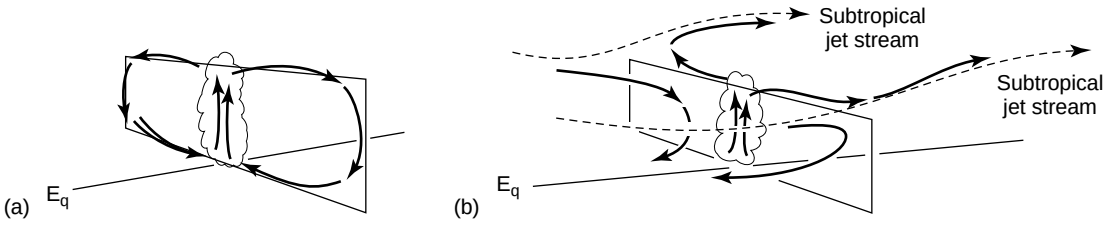


Figure 11 The Hadley cell is characterized by rising motion in the equatorial region, poleward flow at upper levels, sinking at higher latitudes and a return flow equatorward at low levels. (a) The traditional Eulerian view, which depicts the average north–south motion and vertical motion; (b) an attempt at a Lagrangian view, which aims at depicting the average motion of individual air parcels which necessarily involves a three dimensional image including and east–west motion

found at higher elevations in the tropics than elsewhere. From these upper levels, air is accelerated towards the poles. This is the driving mechanism behind the Hadley circulation, named after the 18th century British scientist who first suggested the Earth s rotation and differential heating were major components in the atmosphere s general circulation (Figure 11). The removal of air from the upper levels is felt farther down as a decrease in pressure and leads to a groove of low pressure at low levels along the Equator. From the surrounding areas of relatively higher surface pressure, air is accelerated into this groove, creating the *Trade Winds*. Due to the fV, the Trade Winds arrive at the equator from a northeasterly direction in the Northern Hemisphere and from a southeasterly direction in the Southern Hemisphere (*see* Hadley Circulation, Volume 1).

Intertropical Convergence Zone

A large part of the heating in the equatorial region does not originate directly from the Sun, but through a delayed effect. Through evaporation from rivers, seas and the vegetation the Sun s energy is stored in the atmosphere as water vapor. It is then transported by the Trade Winds and converges in a band along the equatorial zone called the Intertropical Convergence Zone (ITCZ). The ITCZ is a surprisingly narrow belt of vigorous convective activity and heavy precipitation, usually made up of a large number of distinct cloud clusters a few hundred kilometers wide separated by clear skies. Of particular importance are the so-called *hot towers*, large cumulonimbus clouds that concentrate the main heating.

Over the oceans the ITCZ is not located exactly at the equator. Sometimes it is split up into two parallel bands on either side of the equator. This is because of the relatively low sea surface temperature often present at the equator. How this colder water has been brought there is worth a short explanation because it involves the action of the fV. The steady Trade Winds from the east drive the surface water towards the west. Although the fV is weak close to the equator, it has an accumulated deflective effect on the water due to the steadiness of the driving wind, and so fV deflects the surface water away from both sides of the equator. This leads to an outflow of surface water, which causes an upwelling of deeper and colder water from deeper layers. This mechanism is particularly prominent over the eastern Pacific Ocean where an equatorial pool of cold water generally prevails. This is known as La Niña (Figure 12) (*see* Intertropical Convergence Zone (ITCZ), Volume 1).

Monsoons

The fact that land has a small heat capacity compared to ocean is profoundly significant for many atmospheric features. The absorption of solar radiation raises the surface temperature over the land much more rapidly than over the oceans and causes and creates local wind systems. In the smallest scale we find the sea breeze, in the largest scale the monsoon, a gigantic land and sea-breeze circulation caused by the seasonal warming and cooling.

The largest monsoon circulation is created by the vast Asian landmasses and the Himalayas. During the cold season, the cooling of the Asian continent north of the Himalayas causes cold air to sink. Sinking motion is linked with inflow at upper levels and outflow at lower levels, the ice skater effect. The fV drives these flows into outward spiraling anticyclonic and inward spiraling cyclonic patterns, respectively. Due to the pressure adjustment, a vigorous surface high-pressure area is created over Siberia, with a tendency to lower pressure aloft. The cold air from Siberia moves southward over China and Southeast Asia. This is the winter monsoon.

During the warm season the heating creates a vast upper air high-pressure system that is centered over the Indian peninsula and the Himalayan mountain plateau. This leads to an upper outflow; a surface low-pressure area is formed and moist air is drawn in from the surrounding areas. The weakness of the fV over the Indian Ocean allows air, even from south of the equator, to be drawn into the low-pressure region over the Indian Peninsula. The influx of moisture gives rise to the formation of strong

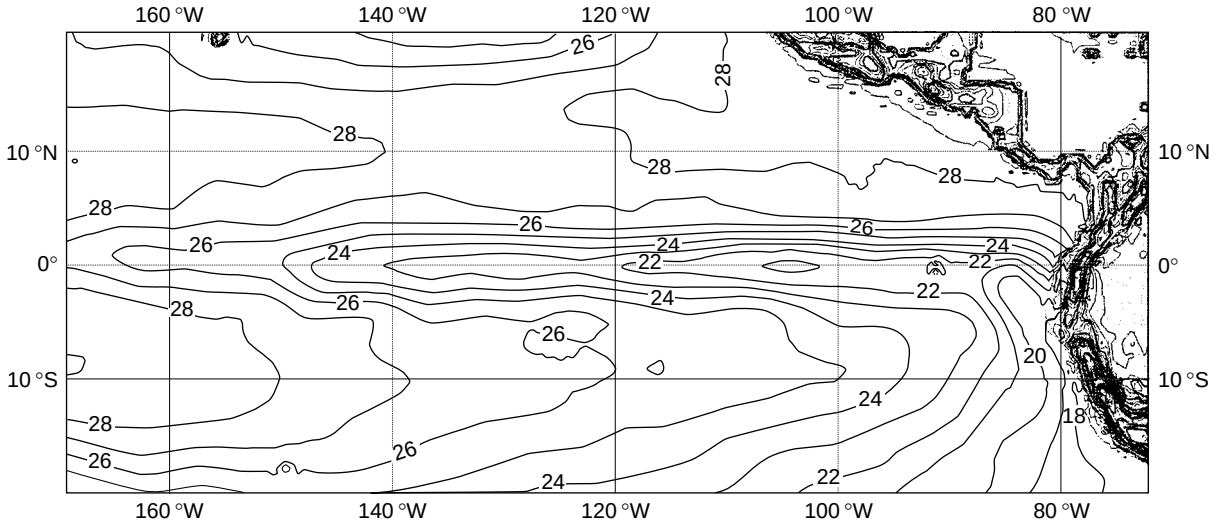


Figure 12 Mean sea surface temperature of the eastern part of the Pacific Ocean for the period 5 Sep to 5 Oct, 1998. The relatively colder water along the equator has been brought up from deeper layers through the action of the fV. Although the fV is weak close to the equator, the steady trade winds from the east account for a small accumulated effect, so that water on both sides of the equator is deflected away from it. This divergence causes an upwelling of deeper and colder water

cumulonimbus convection that leads to further heating of the air due to the release of latent heat.

The monsoon circulation, in particular the summer monsoon, also has an impact on the neighboring continents. The dry regions over northeast Africa are assumed to have their climate determined by the intensity of the Indian monsoon. One contributing factor to this westward directed influence might be the formation of a strong easterly jet stream formed over the northern part of the Indian Ocean on the southern flank of the upper air high pressure system over the Indian peninsula mentioned above (*see* Monsoons, Volume 1).

Easterly Waves and Tropical Cyclones

The Trade Winds blow steadily, interrupted only by minor westward moving easterly waves. When they interact with disturbances that have come down from higher latitudes, they develop into subtropical storms, which can become very intense due to the release of latent heat. These disturbances can sometimes move into higher latitudes and affect developments there.

During the warm season, when the subtropical high-pressure belts are displaced poleward, the easterly waves appear in regions with slightly stronger fV. This allows individual easterly waves to develop into small, intense tropical cyclones, hurricanes (or typhoons as they are called over the Pacific, Willy–willys in Australia) driven by the release of latent heat (*see* Hurricanes, Typhoons and other Tropical Storms - Descriptive Overview, Volume 1; Hurricanes, Typhoons and other Tropical Storms - Dynamics and Intensity, Volume 1). The fV, which has been instrumental during the formation of the initial vortex, soon becomes overpowered by the C_e, resulting in a cyclostrophic flow. The outward pushing C_e causes a sucking effect that draws down dry air from high levels down into the center of the storm. This is the cloudless eye of the storm.

Because tropical cyclones depend crucially on latent heat release from moist air, their development occurs almost exclusively during the peak of the warm season. The air and water temperature (which peak at different times) should not be below $+28\,^{\circ}$C. Most tropical cyclones die out over land due to friction and lack of supply of sufficiently warm and moist air from below. Because of the cold ocean water from the Antarctic, there are never any hurricanes in the South Atlantic and Southeast Pacific.

Walker Circulation

A third major tropical circulation system driven by heating from below is the so-called *Walker circulation*. Its rising branch is over the Indonesian archipelago and the western part of the equatorial Pacific Ocean, a region of very warm sea surface temperatures and high mountains with intense rainfall, causing extensive release of latent heat. The descending branch is over the central parts of the Pacific where the sea surface is relatively cooler (Figure 13).

Unlike the Hadley circulation and the Monsoon, the Walker circulation is hardly affected by the fV. The air that has been brought to high altitudes over Indonesia is blown straight eastward toward the eastern Pacific. The return flow over the equatorial Pacific is associated with a

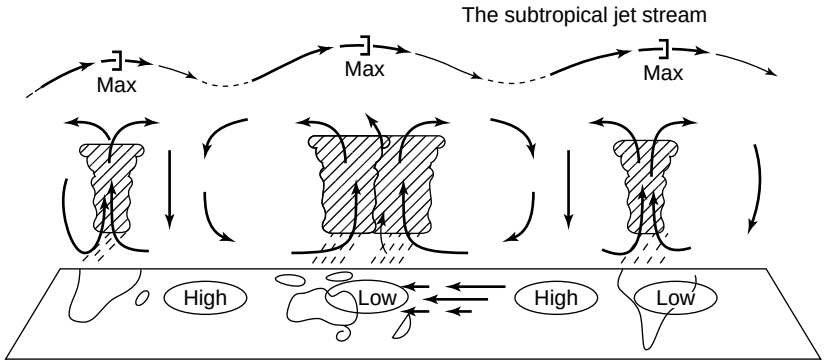


Figure 13 The Walker circulation has rising motion due to convection over Africa, the Indonesian archipelago and South America. The surface flow over the Pacific is easterly, which creates an upslope in the water north of Australia. Shown schematically above these cells are the approximate locations and wind velocity tendencies of the subtropical jet streams that surrounds the Walker circulation at about 30°N and 30°S. With the long distance between the center of the subtropical jet over Indonesia and the center over South America, the Coriolis fV might manage to deflect the jet stream from both hemispheres to cross the equator

weak pressure gradient, with high pressure over the Pacific and low pressure over Indonesia. For not quite understood reasons, this gradient reverses its direction with a period of 2–5 years. These semi-regular variations are known as the Southern Oscillation (*see* Southern Oscillation, Volume 1), which is associated with one of the most important semi-regular variations in the ocean, the El Niño (*see* Walker Circulation, Volume 1).

El Niño and La Niña

The steady easterly Trade Winds coming from the Pacific into the Indonesian region cause a pile-up of warm surface water, creating an upslope hill. From time to time the Trade Winds weaken, allowing the water to spread back towards the central Pacific. This is the El Niño, which is Spanish for the Christ child. This term was originally applied to a warming of the coastal waters of Peru and Ecuador that occurs around Christmas time. As the warmest water moves eastward, the tropical convective clouds and rain and heating maximum follows suit, moving towards the central Pacific. In this way, the whole Walker circulation is displaced eastward, which leads to profound consequences throughout the tropics, with changed patterns of rainfall leading to floods in some regions and drought in others (Figure 14).

This shift extends its influence over North and South America, and probably also to the Atlantic where the frequency of tropical storms (hurricanes) is affected. These changes constitute a striking example of how feedbacks between the atmospheric and ocean circulations act together to drive irregular or semi-regular variations, influencing climate and climate change (*see* El Niño and La Niña: Causes and Global Consequences, Volume 1).

Because the El Niño (EN) and Southern Oscillations (SO) are highly interrelated phenomena, they are often treated together under the name ENSO. It is not quite known what triggers the change. It could be a change in either the atmospheric or the ocean component of the system. Some observations point to an influence of the decaying monsoon over the Indian Ocean. Still, the ENSO events are becoming predictable, at least to some extent, by computer models. Also, to some extent, their anomalous weather consequences for other parts of the tropics are also predictable, including some aspects affecting North and South America and the Atlantic. An important question is whether the El Niño would have the same character in a warmer atmosphere resulting from enhanced greenhouse warming.

The Formation of the Subtropical Jet Stream

Why does the Hadley circulation extend only to 30° and not to the poles? Part of the answer is contained in an elementary calculation (often found in old meteorological textbooks). The student is invited to calculate the poleward acceleration of air due to the meridional pressure difference at upper levels. A thermal contrast of 40 K is assumed between pole and equator. It turns out that the corresponding acceleration is only $0.7\,\mathrm{mm\,s^{-2}}$. This does not sound like much, but in 24 hours this tiny acceleration would have increased wind speeds to velocities of $60\,\mathrm{m\,s^{-1}}$ and carried the air a distance of more than 2500 km, which is more than 20 degrees latitude.

The accelerated air is increasingly affected by the fV, which tries to make it return to the equator. A balance is eventually established when the air on the one hand is prevented from moving farther poleward by the fV, and on the other is prevented from returning equatorward by a strengthened PGF. With the typical heating and rotation of our particular planet, this balance between wind and

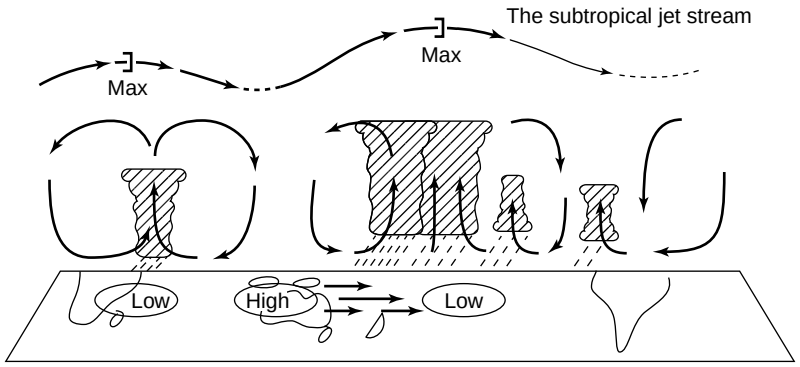


Figure 14 The El Niño occurs when the warm ocean water in the western Pacific spreads back towards the east. The main area of convection follows suit, which leaves Southeast Asia in the subsiding, dry region. With the eastward shift of the heating, the subtropical jet stream is also shifted eastward. This reduces the distance to the next center over South America and effectively prevents any deflections of the jet winds over the equator

Figure 15 A schematic illustration of the formation of the subtropical jet stream. Air accelerated by the upper-tropospheric PGF gains speed poleward, but is simultaneously affected by the fV, which, by deflecting it clockwise in the Northern Hemisphere (anti-clockwise in the Southern Hemisphere) tries to bring the air back towards the equator. A balance is reached around 30 ° latitude

pressure is established around 30 °latitude. Here we find at 7–13 km height a strong westerly flow with winds of 30–80 m s^{-1}. This is the subtropical jet stream, the largest, strongest and most persistent wind system in the Earth s atmosphere. It also sets the poleward limit of the Hadley circulation (Figure 15).

With a faster rotation of the planet, and/or weaker Hadley circulation, the subtropical jet would have been closer to the equator and weaker. If the Earth on the other hand rotated more slowly and/or the Hadley circulation had been stronger, we would have a much stronger jet stream at higher latitudes.

The Longitudinal Variations of the Subtropical Jet

The strongest heating of the atmosphere occurs over the Caribbean and the Amazon basin, over the Sahel and the African rainforests, and over the Indonesian archipelago and Australia. This is where the upper level high-pressure systems are strongest, and it is poleward from these regions where the subtropical jet has wind maxima. Over the oceans, where the heating from below is less intense, the meridional pressure differences are weaker and so are the PGFs.

Due to the irregular distribution of continents and oceans, these regions of maximum heating are not evenly distributed longitudinally. The maximum over East Asia and Australia around 120 °E is far upstream of the next downstream maximum over the Americas, which is located around 80 °W, almost half the circumference of the Earth. The Subtropical jet stream, which originates over the western Pacific, therefore has a long distance to the next heat source over Latin America. During the long passage over the Pacific, the fV has plenty of time to affect the jet stream, trying to drive it into a circular motion toward the equator. This is facilitated by the SST distribution, with the warmest water in the west Pacific. As the temperature of the ocean surface water decreases eastward, the upper level PGF therefore weakens and allows branches of the subtropical jet occasionally to cross the equator from either hemisphere (Figure 16a).

During an El Niño this does not happen. Then the relatively cold sea surface water in the central and eastern Pacific is replaced by warm water, the heating increases and so has the upper PGF, which prevents the fV from deflecting the winds over the equator. The extra heat source over the central and eastern part of the equatorial Pacific creates an upper-air high-pressure area that effectively prevents the fV from deflecting the wind across the equator (Figure 16b).

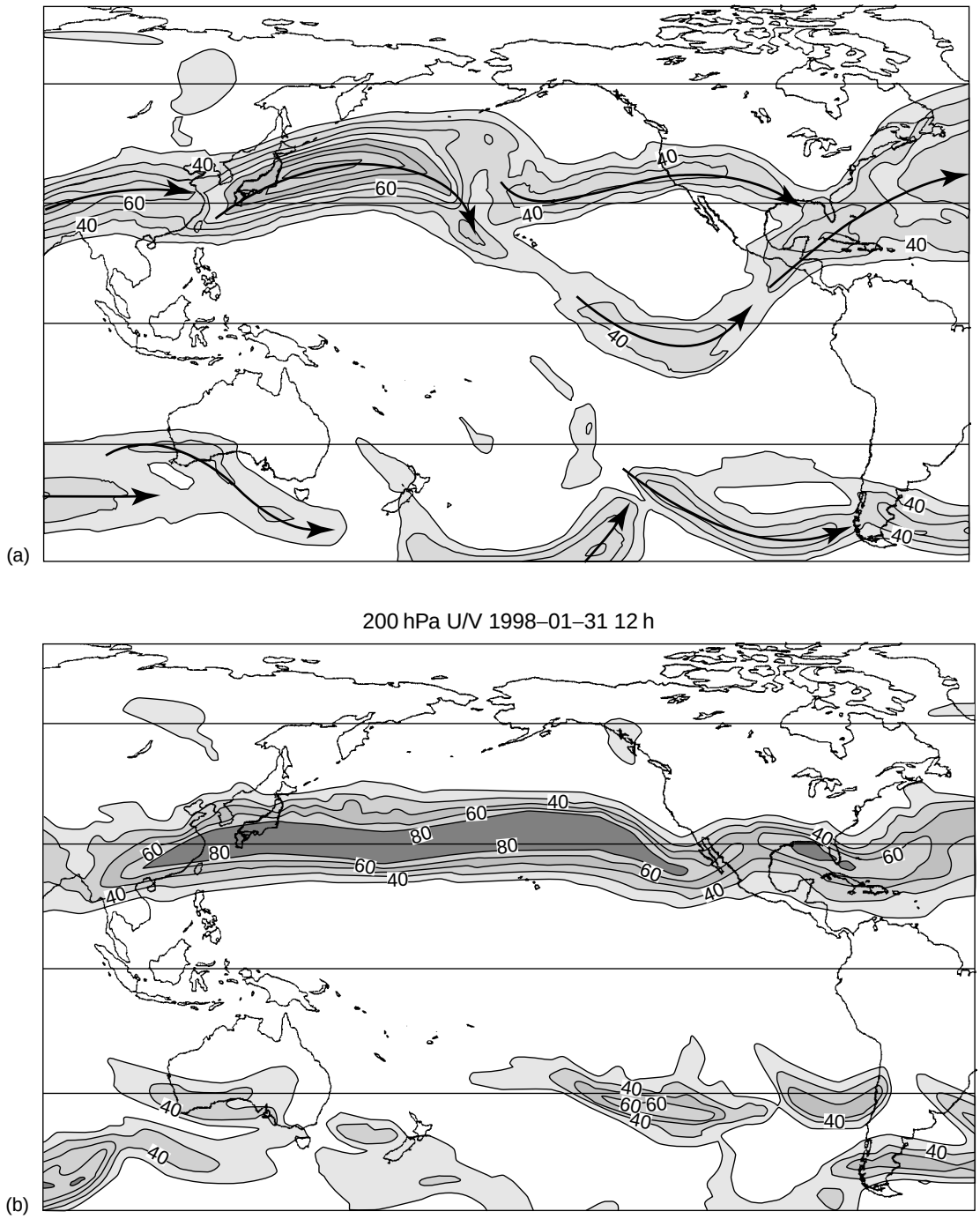


Figure 16 (a) The winds at 200 hPa (approx. 12 km) on 19 February, 1999. The isolines indicate the wind speed in intervals of 10 m/s. The subtropical jet stream over eastern Asia is almost permanently positioned over southern Japan. Halfway across the North Pacific, due to the action of the fV and the reduced heating from below, the wind can be deflected across the equator. Because the fV deflects winds to the left in the Southern Hemisphere, the jet wind is normally deflected back to the Northern Hemisphere. (b) The same as (a) for 31 January, 1998. This is a typical upper tropospheric flow pattern over the Pacific during an El Niño period with the two subtropical jet streams clearly separated in each hemisphere (*Source:* ECMWF analyses)

Seasonal Variations of the Subtropical Jet

The subtropical jet streams show large seasonal variations, particularly in the Northern Hemisphere. Indeed, there is nothing that shows the close connection between the subtropical jet stream and large-scale heat sources better than the extraordinary change in the subtropical jet that regularly occurs from winter to summer. When summer approaches, the heat source for the atmosphere is extended well into the middle latitudes. Due to strong insolation over the large subtropical land masses of the Northern Hemisphere, the large mass circulation of winter that transports heat poleward is greatly weakened and the westerly subtropical jet stream disappears as a circumpolar phenomenon (Figure 17a,b).

The Caging In of the Atmospheric Machine

The fact that air moving from the equator into the subtropical jet stream is, on the one hand, prevented from moving farther poleward by the fV, and on the other, prevented from returning equatorward by the PGF, causes a congestion of air in the upper troposphere. This is felt at lower levels as a small (1–2%) increase in the pressure. This creates the subtropical high-pressure belts that are found around 30° latitude in each hemisphere underneath

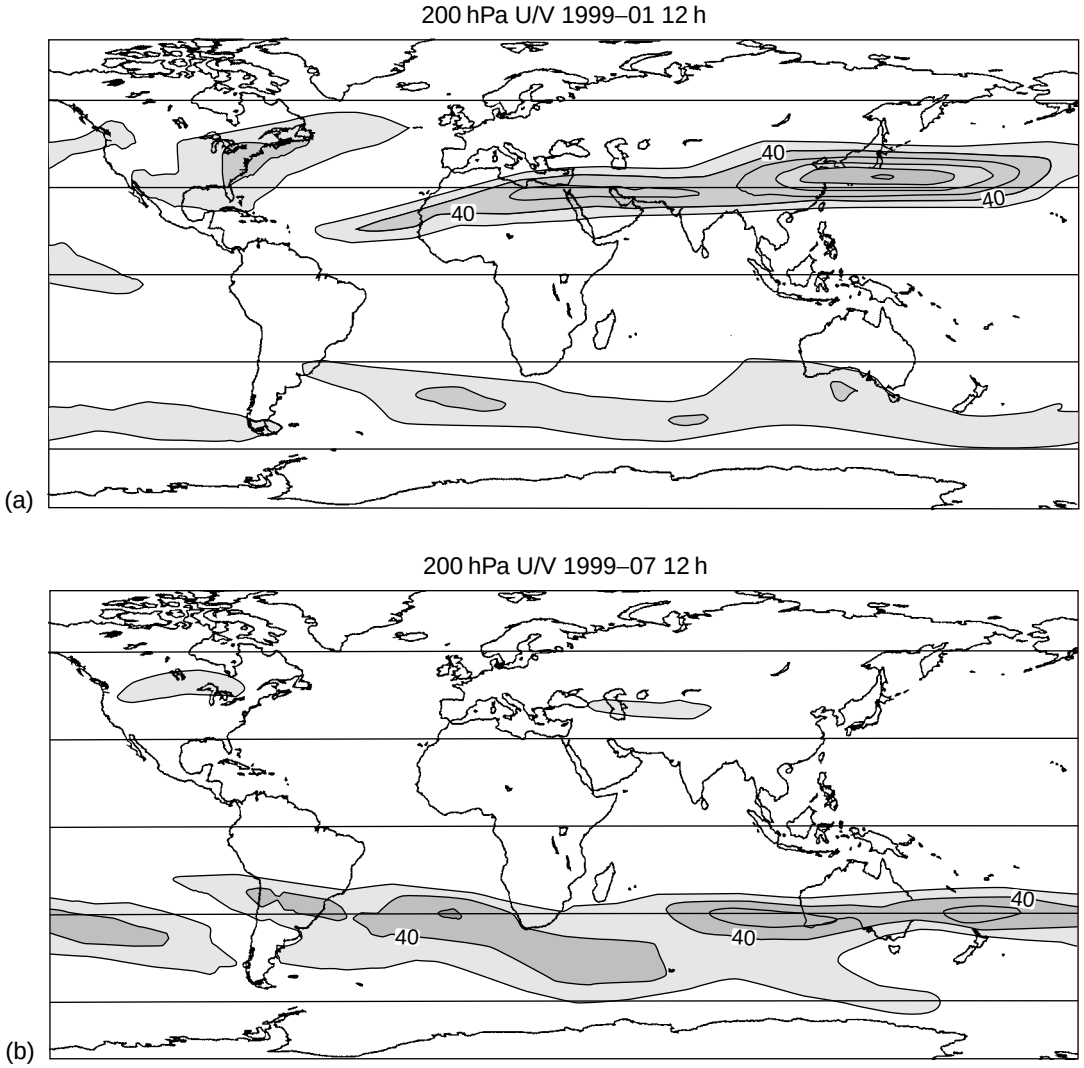


Figure 17 (a) The mean wind speed at 200 hPa (ca 12 km) during February 1999. The Northern Hemisphere subtropical jet stream is clearly seen around 30 °N, whereas its counterpart in the Southern Hemisphere is weak. The mid-latitude jet streams around latitude 50° do not come out strongly on the monthly averages due to their great geographical variability. (b) The same as (a) but for August 1999. The subtropical jet stream in the Northern Hemisphere has disappeared, but its counterpart in the Southern Hemisphere is clearly seen around 30°. The mid-latitude jet streams around 50 °S are still visible due to their strength and interaction with the subtropical jet stream

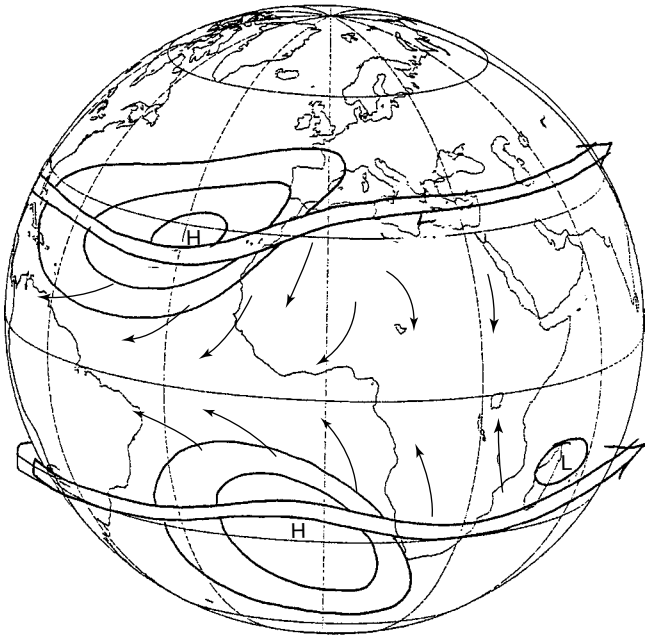


Figure 18 The subtropical high pressure belt is located beneath the subtropical jet stream. The high pressure is particularly strong over the oceans, which coincides with the left entrance of the jet stream where strong sinking motion occurs

the subtropical jets (Figure 18). The convergence of air at high levels also leads to sinking motions, which in turn leads to a drying of the air and dissipation of the clouds. The clear skies at subtropical latitudes lead to increased outward radiation into space and cooling, which is compensated during daylight hours by dry convection, bringing air heated by the underlying surface to higher levels in the atmosphere.

In the subtropical belt, individual high-pressure centers are formed over the subtropical oceans due to a combination of dynamic and thermal effects. These were called the *Doldrums* or the *Horse latitudes* by seamen due to their weak winds. The subtropical highs are closest to the equator during the respective winters, and displaced poleward during the respective summers.

The Polar Regions – Another Source of Energy

Not only do the tropics supply energy to the atmosphere, but so do the polar regions to a small degree. Around the Pole, air is constantly cooled, which makes it heavier and causes it to sink. As we saw with the ice skater effect, sinking motion is linked with inflow of air at upper levels and outflow at lower levels. The upper air becomes a gigantic inward spiral, the lower an equally large outward spiral. Due to the Coriolis effect and the adjustment between pressure and wind, a low-pressure region is found at upper levels and high pressure at lower levels. From this polar cap cell, air is accelerated away from the polar regions and deflected by the fV to create a regime of northeasterly winds down to 60–70 °N. During summertime, with its strong insolation, this process disappears.

WHERE THE ENERGY IS LOST

If it is in the tropics that the Sun s radiative energy is transformed into energy of different kinds, then it is in the mid-latitude westerlies (the roaring forties) that most of this energy, released in storm developments, finally returns to frictional heat and is radiated back into space.

The Mid-latitude Westerlies

The mid-latitude westerlies mirror the tropical easterlies. While the tropical easterlies blow in the opposite direction to the Earth s rotation, and thereby slow it down, the extratropical westerlies speed up the Earth s rotation. These two effects balance, on average, but due to the seasonal variations of the wind regimes, the Earth s rotation, and thereby the length of the day, the Earth s rotation undergoes a periodic variation of the order of milliseconds (*see* Atmospheric Angular Momentum and Earth Rotation, Volume 1).

But why do we have westerlies at all in the mid-latitudes? And why are they not, as in the Tropics, compensated at upper levels by winds in the opposite direction? The reason is that the winds in the mid-latitudes are created and maintained by other mechanisms than those decisive in the tropics. One is the effect of the Earth s rotation; while it can be neglected in the tropics, in the mid-latitudes it plays a very important role.

When the wind is accelerated poleward from the subtropical high-pressure belt, it is affected by an increasing fV. A strong adjustment between wind and pressure is continuously taking place. As a result isobars typically become oriented from WSW to ENE in the Northern Hemisphere, WNW to ESE in the Southern Hemisphere. Where the fV tries to drive the air equatorward, back towards higher pressure, the pressure gradient sharpens. Part of the kinetic energy of the winds is then converted back into potential energy, which delays its further transport up into the mid-latitudes. Pressure decreases with height more slowly in warm air than in cold, so the meridional temperature contrast gives rise to pressure differences at upper levels. Due to the strong fV, the wind and the pressure field rapidly adjust, resulting in wind from a westerly direction.

The strong fV is also the reason why, on average, the subtropical air masses are confined to latitudes equatorward of 45°. Also, at high latitudes the strong Coriolis deflection makes it difficult for the cold air masses from the high-pressure areas around the poles to extend very far. Arctic air masses are normally found poleward of 65 °latitude,

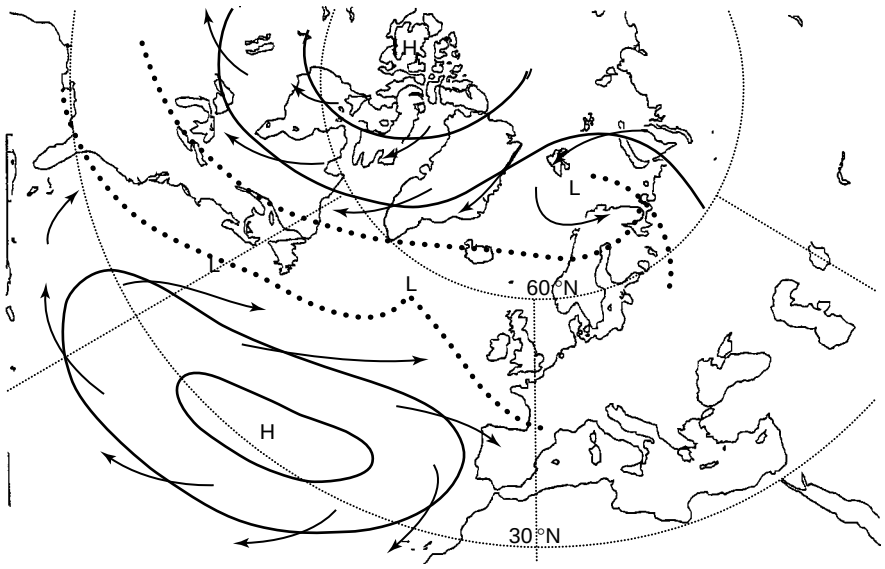


Figure 19 The acceleration of the winds out from the subtropical high pressure belt and the polar cap high pressure system. Winds moving poleward from the subtropical high pressure belt encounter an increasing fV, and a strong adjustment with the pressure field takes place. The fV acts to confine air masses so that cold air from the poles cannot easily penetrate equatorward of 60–65° and subtropical air masses not poleward of 40–50°. The air mass boundary around 45° is called the Polar Front, which in winter extends from the southwestern USA across the Atlantic to the British Isles, and from the northern Mediterranean eastward into Asia. The borderline around 60°, the Arctic Front, is most pronounced over Alaska, western Canada and northen Europe. Another branch of the Polar Front extends from the waters south of Japan across the Pacific Ocean to the US-Canadian border

and then only in wintertime. In summertime the insolation prevents cold air masses from forming. The intermediate latitude band between 45° and 65° is occupied by stagnant air, either from the tropics which has cooled, or from the Arctic regions, which has warmed (Figure 19).

Jet Streams and Fronts

In contrast to the tropics, where the main motion is due to vertical temperature contrasts that create convection that heats the atmosphere, the important driving forces in the mid-latitudes are horizontal temperature contrasts between air masses of different temperature and density. Through the action of the fV, these contrasts are concentrated in narrow zones, so called fronts or Polar Fronts.

In the broad river of air which constitutes the mid-latitude westerlies, long, narrow bands are imbedded between 8 and 13 km altitude where the winds reach 30–80 m s^{-1}, and are on rare occasions 100–150 m s^{-1}. These are the Polar Front jet streams. In contrast to the subtropical jet, which gains its strength from winds accelerated in the Hadley circulation, the Polar Front jet streams thrive on the thermal contrasts along the Polar Front (*see* Fronts, Volume 1; Jet Stream, Volume 1).

The fronts and their associated jet streams do not form a homogenous, uninterrupted band around the hemisphere, but rather separate bands stretching over 60–90° longitudinal sectors where the temperature contrasts and the horizontal pressure differences are more concentrated and the winds stronger than in the intermediate sectors.

As the air moves through the non-uniform pressure field, it is constantly accelerated at the upstream end of a major jet stream, decelerated at the downstream end. With the acceleration, kinetic energy is drawn from the potential energy associated with the pressure and thermal contrasts; with the deceleration, kinetic energy is converted back to potential energy associated with the pressure and thermal contrasts. When the potential energy decreases upstream, these contrasts (gradients) weaken; when potential energy is increased downstream, the contrasts are strengthened. This has the consequence that the whole pressure and temperature pattern will slowly be displaced downstream, while the wind rapidly moves through (Figure 20a–c).

Vertical Motions around Fronts and Jet Streams

Cold and warm air masses of different densities would under buoyancy forces (gravity) tend to arrange themselves horizontally, with cold dense air sinking under the warm light air. Doing so, the system s center of gravity is lowered. But this re-arrangement would involve horizontal motion that would immediately be affected by the fV, which would try to bring the air back. The frontal surfaces are left to slope when a balance is reached between buoyancy forces, which try to create a horizontal stratification, and

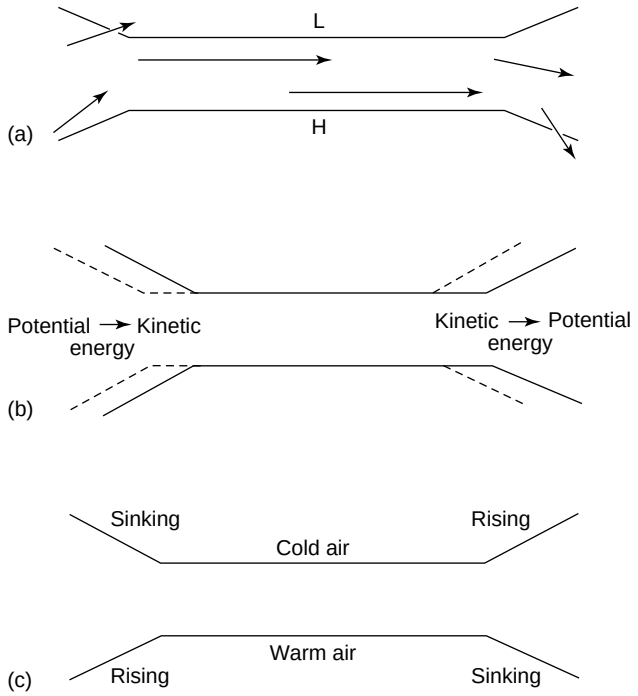


Figure 20 (a) The inflow and outflow in a jet stream. The acceleration of the air at the entrance implies an increase of kinetic energy at the expense of potential energy, the deceleration at the exit implies a decrease of kinetic energy and an increase of the potential energy. (b) Because the level of potential energy is associated with the strength of the pressure differences, these decrease at the entrance and increase at the exit, which leads to the whole pressure system being slowly moved in the direction of the flow. (c) A decrease in potential energy means that the system's center of gravity is lowered. This implies rising of warm, light air and sinking of cold heavy air. With warm air on the equatorward side and cold air on the poleward side, this means at the entrance of the jet rising motion to the right, sinking motion to the left. With a similar reasoning, it can be understood why, at the exit of the jet where potential energy is increasing, cold air to the left is rising and warm air to the right is sinking

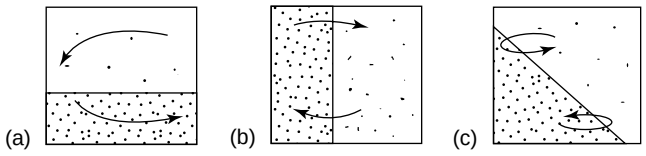


Figure 21 The motion at frontal boundaries is determined by two opposing mechanisms: (a) the thermal contrast tries to create a *horizontal* stratification; (b) the rotation of the Earth tries to create a *vertical* stratification. (c) Neither of the mechanisms is dominant, leaving the frontal surfaces to slope with an inclination of the order of 1 : 50 to 1 : 200

the rotation of the Earth, which tries to create a vertical stratification (Figure 21a–c).

The reversal of this process, against the buoyancy forces, and thus against gravity, involves lifting of cold air and sinking of warm air. Because this would raise the center of gravity and increase the potential energy, it can only occur through external mechanical forcing. It is the fV that provides this mechanical forcing when it, being stronger than the PGF, drives the air towards higher pressure, which slows down the wind. Note that it is the PGF, not the fV, that slows down the wind. This can be compared to a ball rolling down a hill. Gravity accelerates the ball when the kinetic energy, increases. Inertia might then drive the ball uphill, but the work, the conversion from kinetic to potential energy, is still made by gravity slowing down the speed of the ball.

The induced vertical motions affect the temperature, sharpening or even creating fronts. This conversion between potential and kinetic energy is reflected in the waves that develop and the accompanying changes of the frontal slopes: flattening when kinetic energy is created, steepening when the kinetic energy returns to potential (Figure 22). During these transitions the frontal surface oscillates like a membrane. Frequently, for reasons which are not well understood, the waves along the frontal boundary escalate into intense vortices.

The Development of Extra-tropical Storms

The subtropical highs are created and maintained by air at higher levels converging into the subtropical jet stream.

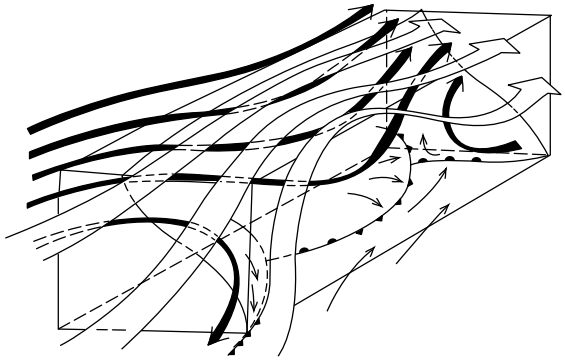


Figure 22 The circulation in extra-tropical weather systems involves a complex interplay of rising and descending motions. Air masses coming from higher and lower latitudes converge in the western part, the ''entrance,'' and diverge in the eastern part, the ''exit''. While staying separate, they undergo large vertical displacements. Whereas at the left entrance and right exit, where the air is sinking, there is a tendency for downward motion and dynamical stability, at the right entrance and left exit of a jet stream, where the air is rising, there are favorable areas for storm development. The combined sinking of cold air and ascent of warm air represents a conversion of potential to kinetic energy, which maintains the jet stream. Heat and moisture is transported downstream, maintaining and even sharpening and extending the frontal zone

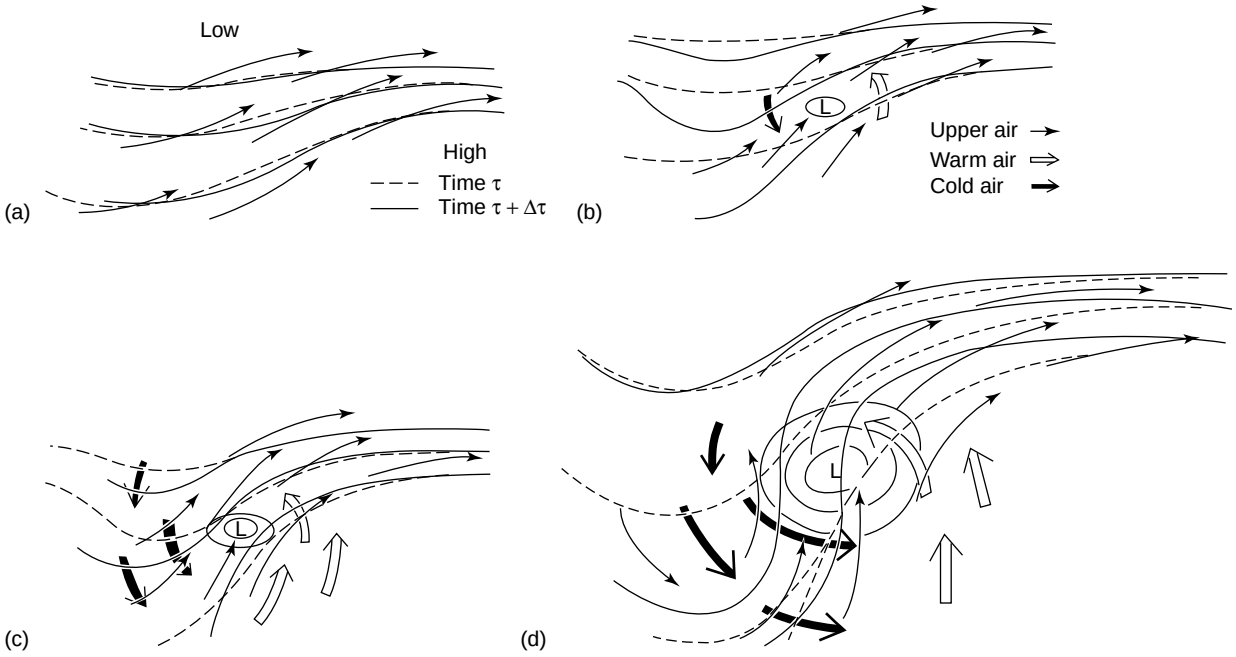


Figure 23 (a) The initial development of a deepening storm. The zone between warm air to the south and cold air to the north; strong pressure differences and a jet stream have developed at high levels. The wind is accelerated into the jet stream and the gradient is weakening [(- - - - -) lines for equal pressure at an early stage, (———) at a later stage]. (b) More air is exported out of the region than imported. This upper divergence causes the pressure to fall at lower levels. A new circulation is rapidly established due to the balance between the PGF and the fV. (c) The transport of warm air northwards makes the upper flow more anticyclonic and increases the acceleration of air out from the region; the transport of cold air southward makes the flow more cyclonic and weakens the acceleration of the wind. (d) The changes in the pressure field lead to an increase in the wind divergence. A chain reaction has come under way with the pressure at lower levels falling, increasing the thermal transports and the amplification of the flow pattern

The opposite process, divergence of the wind at higher levels, creates pressure drops at low levels that often initiate the development of mid-latitude cyclones. There are various mechanisms causing upper-air divergence, the most common of which is when the wind is accelerated into or out of a jet stream. The wind is subject to a strong acceleration, partly reflected in the curvature of the wind flow, from a cyclonically (anti-clockwise) curved flow to an anticyclonically (clockwise) one in the Northern Hemisphere (Figure 23a).

If the air leaving a region is not replaced by the same amount of air coming in, this is felt at lower levels as a drop in pressure. As the wind at lower levels adjusts to the pressure field, a cyclonic circulation is created (Figure 23b). This circulation affects the temperature distribution, which close to jet streams is normally characterized by strong horizontal temperature differences.

East of the low-pressure system the wind at lower levels transports warm air poleward. Pressure decreases more slowly upward in warm air and the influx of warm air increases the pressure at upper levels. This intensifies the anticyclonic circulation at upper levels and accelerates the wind even more. As a consequence, more air is exported out of the region (Figure 23c).

To the west of the low, the induced wind circulation transports cold air equatorwards. Because pressure decreases more rapidly upward in cold air, the influx of cold air decreases the pressure at higher levels and intensifies the cyclonic (anticlockwise) circulation. The increased curvature of the airflow further slows down the wind velocity and less air is imported into the region.

With more air exported, and less air imported, the pressure at lower levels will continue to fall, the circulation around the low-level system will intensify, and so will the transport of warm air northward and cold air southward (Figure 23d). A chain reaction has started that leads to a major storm development with strong winds circulating around the low pressure area, warm air rising above cold air at the front of the system, and cold air sinking below warm air at the rear of the system. It comes to an end when the circulation has spread from lower levels to higher levels and established the warm air in a position above the cold air.

Upper air divergence can also be caused or enhanced by latent heat release when moist air condenses. The heating of the air raises the pressure at higher levels and further enhances the upper outflow of air. This is the main driving mechanism in tropical and subtropical storms, but also plays an important role in extra-tropical developments. The latent

heat release at the right entrance of the jet maintains the thermal contrasts in the upstream part of the jet and keeps it in a westerly position where new cyclones can form. This can lead to a succession of frontal cyclones forming and, within a couple of days, to the creation of a cyclone family.

Downstream Development

The kinetic energy released during the cyclonic development takes two routes: one into the storm itself, where it is soon to a large extent lost in the frictional process at the Earth s surface; the other route into the mid- and upper-troposphere jet winds where it is rapidly transported downstream. Arriving in the next system downstream, these huge amounts of energy can have a profound effect, either directly or through conversion back to potential energy through the action of the fV (forcing the PGF to do negative work on the air). Although the density of the air in the upper troposphere is only 1/3 of the density at the ground, the wind velocities are about ten times stronger, which makes the kinetic energy 30 times larger. The kinetic energy of an upper level jet stream of $60\,m\,s^{-1}$ is the same as a hurricane force wind at the ground of $35\,m\,s^{-1}$.

When the downstream storm develops, which receives the energy in the same way, it transports part of its released kinetic energy into the next system downstream, and so on. As long as the dynamic conditions are favorable, the influence of one storm rapidly spreads downstream to the next in a hemispheric domino effect (Figure 24). The speed of this downstream development is roughly $30\,m\,s^{-1}$, which corresponds to $30^{\circ}\,day^{-1}$ at 45° latitude, but slightly less in summertime. This should be compared with the typical phase speed of $10^{\circ}\,day^{-1}$ of a normal frontal cyclone wave. In the Southern Hemisphere the typical velocity of the downstream influence is $40^{\circ}\,day^{-1}$ due to the predominant zonal flow.

The fact that the influence of a weather development spreads three times faster than the weather system has consequences for weather forecasting. A five day weather forecast for Europe is dependent on the initial atmospheric conditions over North America, a five day forecast for North America on the conditions over the Northwest Pacific, although the weather systems in those areas may not necessarily arrive in the target area.

Cut-off Lows and Blocking Highs

The downstream development not only deepens cyclones, it also amplifies high-pressure systems. A strong explosive cyclogenesis creates strong pressure differences high up in the troposphere that force the jet stream poleward and helps to amplify a downstream ridge. The poleward transport of warm air further increases the pressure and supports the creation of a high pressure system through a deep vertical layer, a so called *blocking high*. The upper current moving over the ridge, partly due to the anticyclonic curvature, accelerates and make the fV stronger than the PGF. The wind, thus driven towards higher pressure by the fV, creates a low-pressure system, a so-called cut-off low (Figure 25).

Flow Patterns Forced by the Earth's Surface

The uneven heating and cooling of the oceans and the continents also contribute towards shaping the large-scale

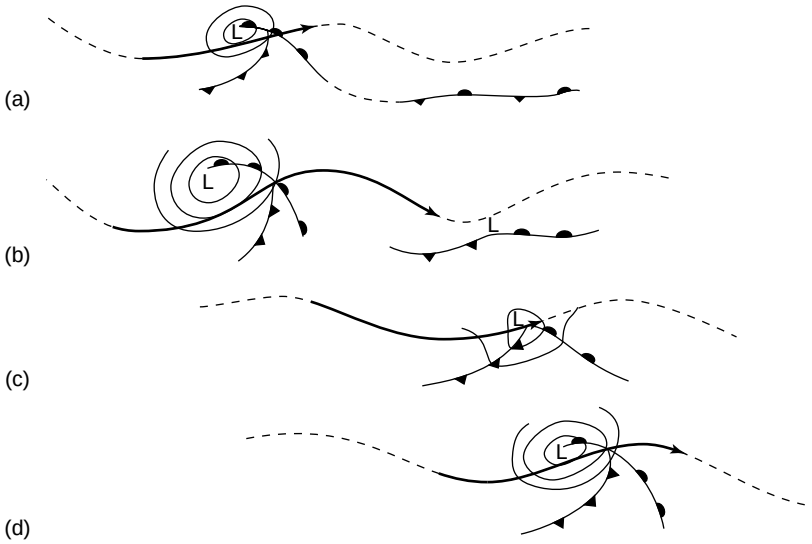


Figure 24 Down stream development. The kinetic energy released during the development of one storm (a) is transported downstream in the jet stream (b) and in a short period of time affects the next system (c),which then during its development releases additional kinetic energy into the upper tropospheric flow (d)

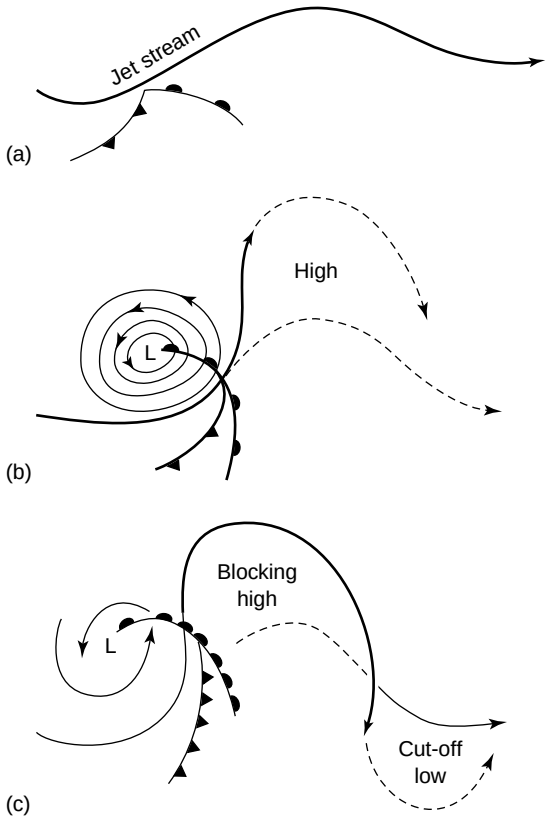


Figure 25 (a) Blocking highs tend to develop as a response to explosive cyclones. (b) When huge amounts of kinetic energy are ejected into the upper atmospheric flow, bringing into a new direction and often amplifying a downstream high pressure system. (c) When the strong winds in the free atmosphere reach the downstream side of these amplified high pressure systems, there is a tendency to form low pressure circulations, so called ''cut-off lows''

flow over the Earth. In wintertime there is a strong cooling over the continents due to radiative losses. The cooled air has a general tendency to sink. This sinking motion is associated with upper inflow and, in line with the ice skater effect, cyclonic circulation. At lower levels there is outflow and anticyclonic circulation.

When the cold air is transported out over surrounding warm ocean waters, which normally are to the east of the continents, there is widespread and intense rising motion due to heating from below. This creates favorable conditions for low pressure and low level cyclonic circulation. Cyclones develop easily and strongly by drawing from these thermal contrasts, which are found in the western Atlantic and the northwest Pacific. In summertime the heating of the continents contributes to rising motion with low level inflow and upper level outflow, with a tendency for anticyclonic circulation.

Another influence, which contributes to the shape of the large-scale flow, is from the major mountain ranges. The atmospheric flow tends to accumulate air on the windward side, creating or strengthening a PGF across the mountain. This accelerates the wind to an extent that the fV becomes stronger than the PGF on the lee-side of the mountain ridge. The wind therefore starts to deviate towards higher pressure and, similar to the cut-off process described above, creates a cyclonic circulation and a low-pressure system. The tendency to have cyclonic shaped flow in the lee of major mountain ranges facilitates the development of cyclones in this area, whereas there is a enhanced tendency for high pressure systems on the windward side.

The waves generated by a combination of land-sea contrasts and the orography, further enhanced by cyclone developments where the dynamic conditions are favorable, ultimately create geographically very large waves or weather zones, so called Planetary or Rossby waves (*see* Rossby Waves, Volume 1). They are responsible for periods of persistent dry or wet, cold or warm weather types. The so called *beta-effect*, the latitudinal variation of the fV, is particularly prominent in systems of large north–south extent and may make the waves become stationary and even move slowly westward, against the general current (*see* Coriolis Effect, Volume 1).

A SUMMARY OF THE ATMOSPHERIC CIRCULATION

A simple two-dimensional summary of the general circulation of the atmosphere must be accompanied by some words of caution when the picture is interpreted. As was discussed in the Introduction, average values do not necessarily represent what is typical and it is not normally possible to deduce the path of air particles from average wind fields.

Two Ways to Look at the Meridional Circulation

The commonly depicted, tri-cellular, cross-section of the atmosphere with rising motion in the tropics and the mid-latitudes, sinking motion in the polar regions and subtropics should not be interpreted as showing the approximate path of any individual air parcel. The cross-section is based on averages of winds in fixed locations (called Eulerian averages) (Figure 26a). A picture of the average routes of air parcels can only be accomplished by averaging in terms of trajectories (called Lagrangian averaging). If one wishes to determine the likely spread of a pollutant in the atmospheric circulation, the Lagrangian view would give a much better idea than the Eulerian (Figure 26b).

The Eulerian and Lagrangian methods do not necessarily have to show different pictures. Due to the steadiness of the Hadley cell, the two averages portray similar pictures of the rising branch within the summer hemisphere and large mass transports into the winter hemisphere. But at subtropical and middle latitudes, where the motion is more complicated (see

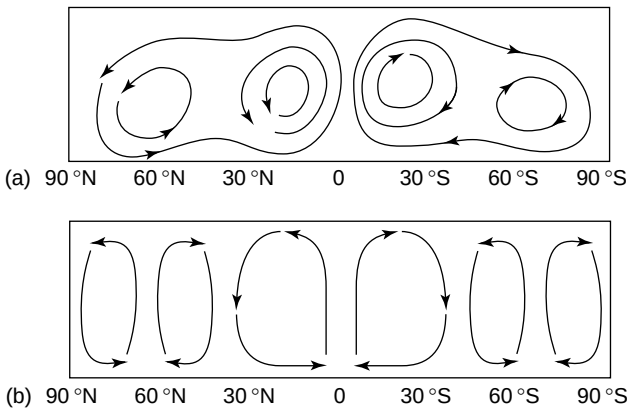


Figure 26 Cross sections of the atmospheric circulation using: (a) Lagrangian averages; (b) Eulerian averages. A common mistake is to visually interpret the Eurlerian average, which denotes streamlines, as if it were Lagrangian, which represents the tracks of individual air parcels (trajectories)

Figure 11) the two averages display very different images. Instead of three cells in the Eulerian picture, the Lagrangian view depicts one single cell extending from the tropics to polar latitudes.

Correlated Fields

The (Eulerian) average wind flow depicts rising motion at mid-latitudes between 50° and 70°, sinking at subtropical latitudes around 35°. Because the air is colder at higher latitudes than lower, such a circulation seems to imply rising of cold air and sinking of warm, which would imply an increase in potential energy. This interpretation has supported the notion that between the Polar Cap cell and the tropical Hadley cell, both sources of kinetic energy, there is a third cell, the *Ferrel cell*, that is supposed to convert kinetic energy back to potential. But this notion of a mid-latitude cell is due to a misinterpretation of statistics.

It is true that the 50° – 70 °latitude band has a tendency for upward motion, but these are mostly associated with rising tongues of warm air in connection with passing cyclones. The same applies for the subsidence around 35° subtropical latitude, which is associated with outbreaks of cold air that are generally associated with sinking motion (Figure 27). The relevant correlation, the one which can be given a physical interpretation, is not between mean values, but between instantaneous values of temperature and vertical velocity.

The Atmosphere's Energy Budget

The interpretation of the atmosphere s energy budget must be done with care. On a long-term average, the potential energy of the atmosphere is converted into kinetic, which, in turn, is lost in friction against the Earth s surface and internal friction (viscosity) (Figure 28a). On average, the contributions from the Sun to the potential energy makes up for the average loss in friction. However, this is the

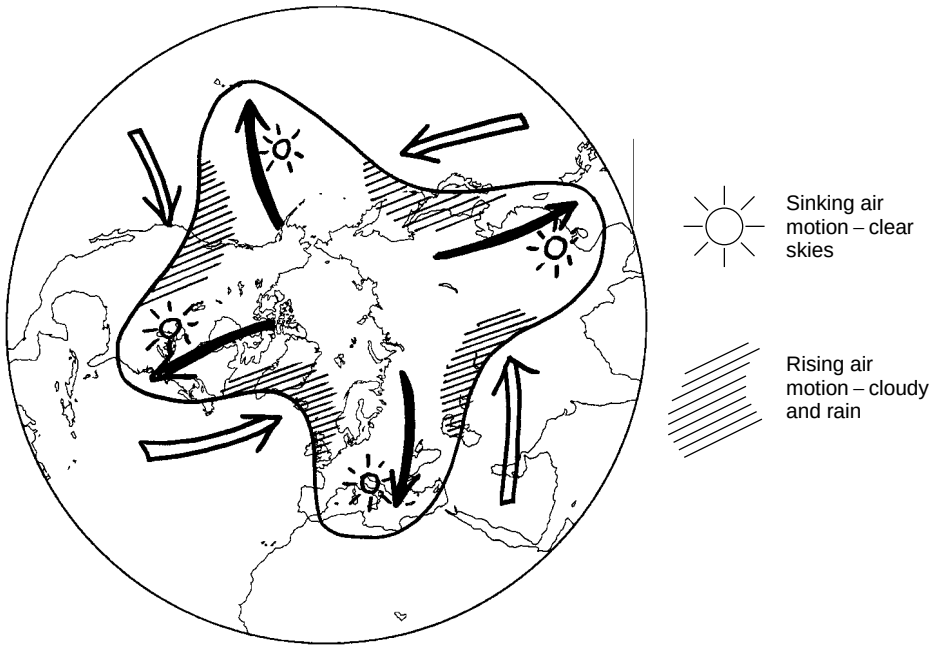


Figure 27 Advances of warm air poleward is often associated with rising motion (denoted by hatching symbolizing precipitation) whereas cold outbreak to low latitudes often are accompanied with sinking air and rather cloudless conditions (here symbolized with a sun)

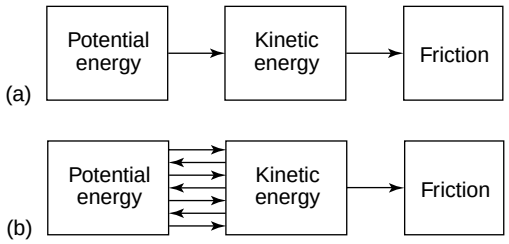


Figure 28 (a) The net energy budget shows that potential energy is lost to kinetic energy, which in its turn is lost to friction. (b) Only the kinetic energy to friction conversion is irreversible, the potential to kinetic is highly reversible

net effect, or budget, and does not necessarily tell what is going on at every moment at any place where there is a constant conversion between potential and kinetic energy (Figure 28b).

In the atmosphere, it is the subtropical high pressure belts that act as sources for kinetic energy; the low pressure systems in the mid-latitudes act as sinks for kinetic energy due to frictional losses. But before the kinetic energy is lost to friction, it has been recycled several times through reversible conversions between potential to kinetic energy in the mid-latitude cyclones, most clearly reflected in the downstream development process.

A schematic cross-section of the atmosphere is given in Figure 29. The Hadley cell circulation should be seen as a longitudinally elongated spiral into the diagram where air rising at one longitude descends at another far downstream. Due to the seasonal variations of the maximal heating from the Sun, the Hadley cell is slowly oscillating around the equator and also varies in intensity.

The fronts and jet streams in the westerlies are drawn in a way to suggest their general orientation from WSW to ENE. The vertical motion in the area might on average be upward, but, again, this is just the net result of very strong up- and downward motions.

The Stratosphere

The flow in the lower stratosphere, the part of the atmosphere at 10–25 km height, is generally characterized by a rising motion of cold air from the equatorial belt in the summer hemisphere to a sinking motion of warmer air in the mid-latitudes and around the poles. With this circulation, converting kinetic energy into potential energy, its energy cycle differs considerably from that of the atmosphere as a whole. It is rather like a heat engine in reverse, or a thermodynamic refrigerator. This can only come about through mechanical forcing, and it is the motions in the atmosphere below that provide this mechanism (*see* Stratosphere, Temperature and Circulation, Volume 1).

However, the stratosphere also has dynamics of its own where instabilities create strong vortices or blocking highs, so called *sudden warming*. The dynamics of the stratosphere is further complicated by a complex interaction between radiative, dynamical and photochemical processes in the ozone layer.

The stratosphere has been under intense study since it was discovered 100 years ago, still new discoveries

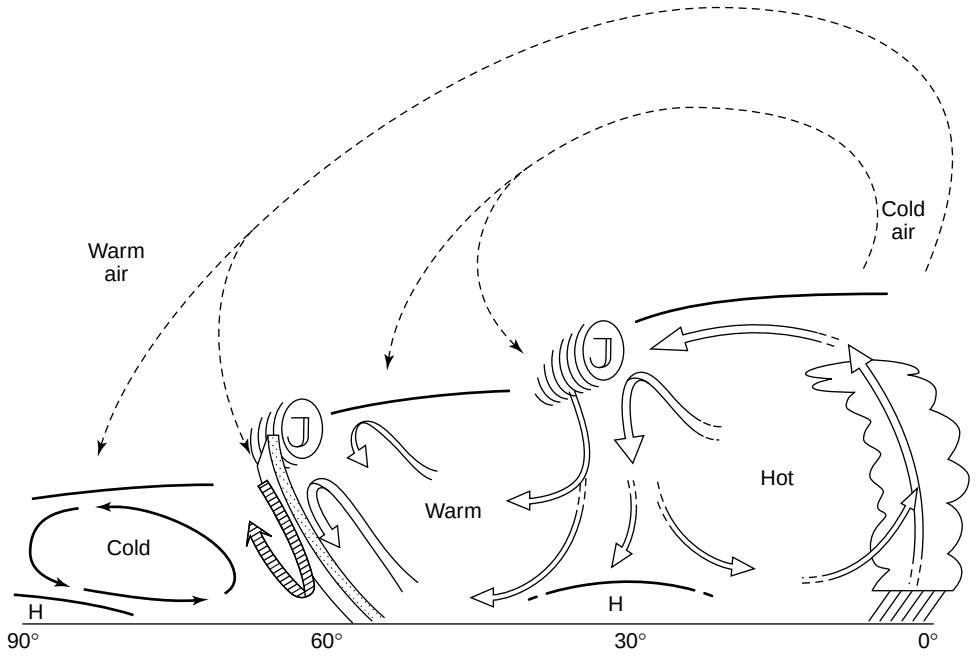


Figure 29 A schematic cross section of the atmosphere (troposphere and lower stratosphere)

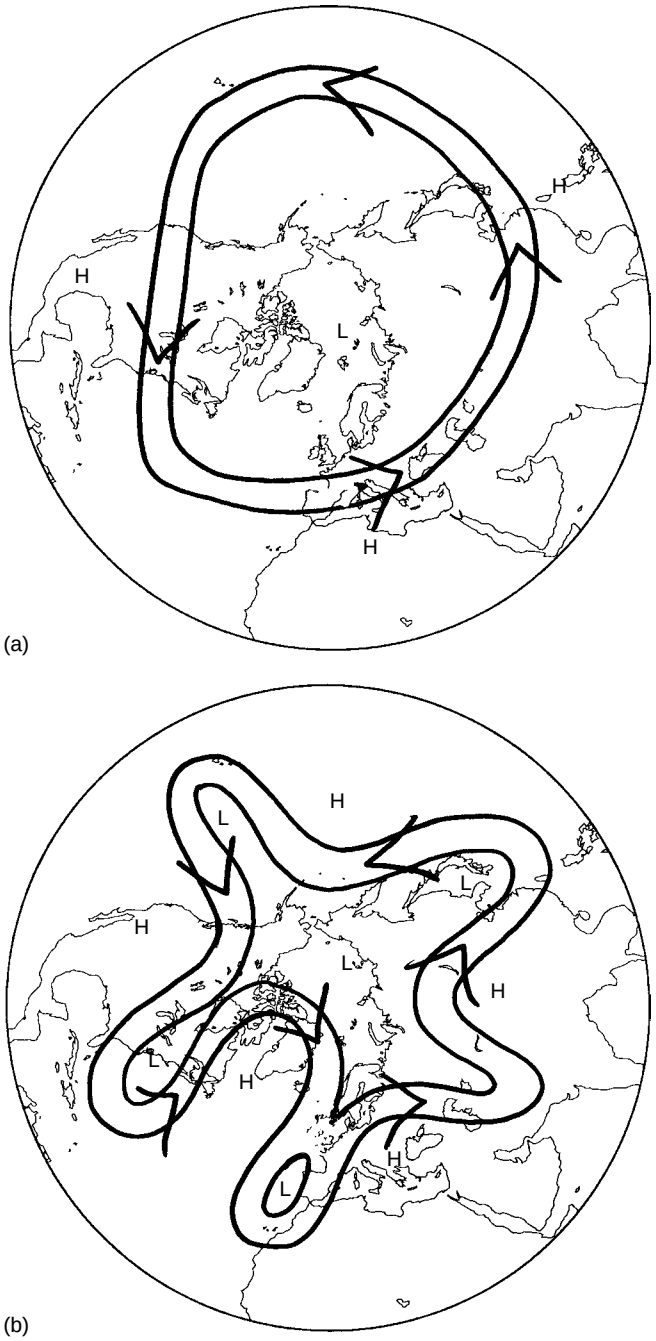


Figure 30 The atmospheric flow vacillates between two regimes, (a) one zonal (mostly east-west flow) and (b) one meridional (much north-south flow) with pronounced high pressure formations (''blocking'') and small, isolated low pressure systems (''cut-offs''), both created when there are strong imbalances between the pressure field and the winds

are made almost every decade. So for example, a 26-month oscillation in the upper winds has been found in the tropics. This Quasi-Biennial Oscillation (QBO) is still not fully understood, nor the general interaction between the stratosphere and the underlying atmosphere (*see* Quasi- Decadal Oscillation, Volume 1). There are still discoveries to be made and surely many surprises in store!

Can the Atmospheric Flow Change?

There has been much research into the dynamics of the large-scale flow to acquire a deeper understanding of past climate changes (the ice ages) and possible scenarios for the future. It is not yet clear why the atmospheric flow can be anomalous for relatively short periods like several months or years. Why were the mid-1990s characterized by dry conditions over Western Europe, why have the last five years been wetter than before?

One process where an understanding is growing concerns the mechanisms that induce wave patterns because the land-sea contrasts and the orography may not necessarily pull the atmosphere in the same direction. There is also a seasonal variation in the land-sea contrast, which is only slightly evident in the orographic forcing. Due to this misfit, the large scale flow is never quite in a stable state and tends to vacillate between two preferred states: one relatively straight zonal type with a well-developed and fairly broad westerly current in middle latitudes, and the other large meridional flow patterns which extend meridionally with strong cyclonic vortices at low latitudes and anticyclonic flow at high latitudes (Figure 30). Each pattern persists for several weeks, whereas the transition from one to the other is relatively brief.

There are reasons to assume that the typical flow regimes that we regard as normal might be just one of several possible quasi-stable states of the atmosphere. Any change of the overall temperature of the atmosphere will most likely change the flow pattern as well. This might have unexpected consequences; an over-all warming of the atmosphere might lead to some regions getting colder temperatures due to the change of flow regimes. This can only be fully understood if the motion of the general circulation also includes the motions of ocean currents. It might be that neither the atmosphere nor the oceans can undergo any radical changes on their own, they might only be able to change if they change together.

Atmospheric Processes and Interactions

see Earth System Processes (*Opening essay*, **Volume 1**)

Atmospheric Radiation Measurement Program (ARM)

see ARM (Atmospheric Radiation Measurement) Program (**Volume 1**)

Atmospheric Structure

Peter Brimblecombe
University of East Anglia, Norwich, UK

Atmospheric structure is usually represented in terms of temperature pro les which de ne the boundaries of the troposphere and stratosphere. It is also possible to describe the structure of the atmosphere in terms of composition. The best known example of this is the ionosphere where free electrons abound. However, high in the atmosphere gases separate gravitationally (heterosphere) and at its extreme edges (exosphere) molecules so rarely collide that they can undergo ballistic trajectories and escape from the atmosphere.

The atmosphere appears to be a transparent gas and initially devoid of obvious structure. However, more careful examination reveals that there are layers. Clouds, especially those with flat bases make this fairly obvious.

Although there is a gradual decrease in pressure with altitude, it is the thermal changes with altitude that are most often used to define structure. Through the lowest part of the atmosphere, the troposphere, temperature decreases with altitude at about $6.5\,^{\circ}\mathrm{C\,km^{-1}}$ (*see* Lapse Rate, Volume 1). Air closest to the ground is warmest and this is the region we are most familiar with. It is here that our weather takes place. Convection can be active in the troposphere and is dramatically seen in the formation of thunderclouds. At times, air close to the ground can become cold, particularly at night. Under these *inversion* conditions, the air is resistant to vertical mixing, pollutants can accumulate or fogs can form.

Higher up in the atmosphere the absorption of radiation by the ozone layer warms the atmosphere. Thus, temperature increases above the *troposphere* (Figure 1), in a layer called the *stratosphere*. The increase in average temperature with height means that the stratosphere is very stable with respect to vertical mixing, hence the tendency for well defined strata of air to form. Yet it is so cold that water is present in very small amounts and clouds are very rare and

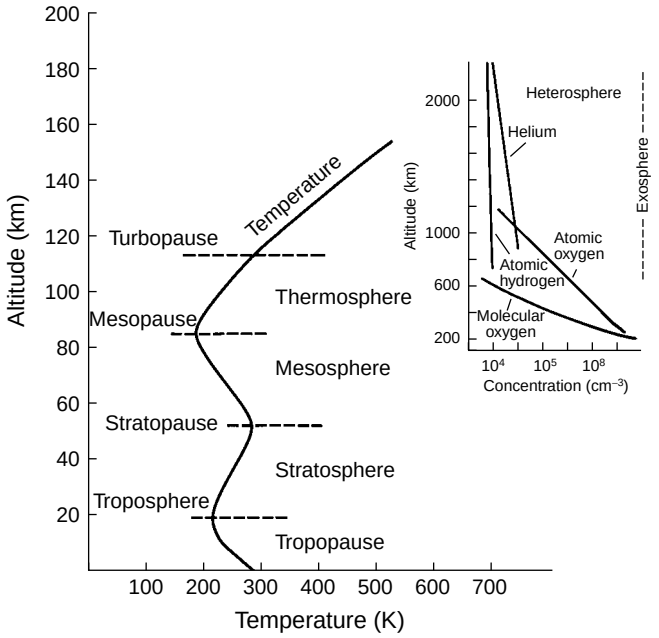


Figure 1 The structure of the atmosphere. Note how the boundaries at the top of each of these layers are called the tropopause, stratopause, mesopause. (Adapted from Andrews *et al.*, 1996)

consist of sulfuric and nitric acid containing droplets or icy crystals.

Above the stratosphere, the air cools again in the *mesosphere*. Here, noctilucent clouds can be observed in the summer at high latitudes. These are formed through water nucleation onto dust or ions, possibly from meteor trails, which penetrate the mesosphere. The water probably derives from the photooxidation of methane, which has led to speculation that increases in atmospheric methane has caused an increase in the frequency of noctilucent clouds.

Even higher, we encounter the *thermosphere*, where the atmosphere has a very low density and the molecules are warmed by direct impact of hard photons from the Sun. Its high temperature is sensitive to output of solar radiation, which varies as the Sun passes through its quiet and active phases every 11 years. Although the temperatures are very high, the gas is so tenuous that it carries little heat, so does not control the temperature of objects such as spacecraft operating at these altitudes.

There are other ways we can define atmospheric structure. The lower atmosphere is well mixed, where turbulent processes are stronger than the gravitational settling of the gases. The top of this well mixed part of the atmosphere, above which gases settle out dependent on their molecular weight, is called the *turbopause*. Below the turbopause the atmosphere is often termed the *homosphere* where the gases are well mixed. Above the turbopause it is called the

heterosphere where we find the light gases beginning to dominate at the top of the atmosphere (Figure 1).

The inset of Figure 1 shows the heterosphere and we can see that the gases have separated out due to gravitational settling, but many are not the familiar gases we know at ground level (*see* Atmospheric Composition, Present, Volume 1). Even oxygen and nitrogen are not in their familiar diatomic form. The photodissociative processes at high altitude break the molecules down into atoms, thus, at the very top of the atmosphere we find a dominance of hydrogen and helium.

These divisions of the atmosphere have only considered the neutral molecules. The upper atmosphere is affected by ions in a region known as the *ionosphere*. These ions are at relatively low concentration, but their effects are so pronounced that they have been much studied. In particular the way they affect radio waves has been of importance for much of the 20th century. The ionosphere is structured into a number of regions. Early studies used radio waves as probes, which detect changes in electron density in the upper atmosphere. This meant that early studies focussed on the electron distribution of the ionosphere, although scientists were well aware that the charge would be balanced by positive ions. The advent of rockets allowed the positively charged species to be identified.

In the upper atmosphere incoming meteors, although contributing only small amounts of material, have noticeable effects because the atmosphere is so tenuous. As the meteors ablate volatile materials such as water they leave a trail. In addition, less volatile materials such as sodium, magnesium and iron are ablated and form distinctive layers.

At the very outer extent of the atmosphere we encounter the *exosphere*. Here, the gas is so tenuous that molecules follow ballistic trajectories, kilometres in length before colliding with another molecule. Some of the molecules travel so fast that their ballistic trajectories at thermospheric temperatures allow them to escape from the atmosphere completely and become lost to space. In the case of the upper atmosphere of the Earth it is hydrogen and to a lesser extent helium that escape with residence times of about a thousand and many millions of years, respectively. The rapid escape of hydrogen is important as it allows atmospheres to become more oxidizing over time.

See also: Atmospheric Motions, Volume 1.

REFERENCE

Andrews, J E, Brimblecombe, P, Jickells, T D, and Liss, P S (1996) *An Introduction to Environmental Chemistry*, Blackwell Science, Oxford.

FURTHER READING

Brekke, A (1997) *Physics of the Upper Polar Atmosphere*, John Wiley & Sons, Chichester.

B

Balloon Measurements

see Radiosondes (**Volume 1**)

Bjerknes, Jacob

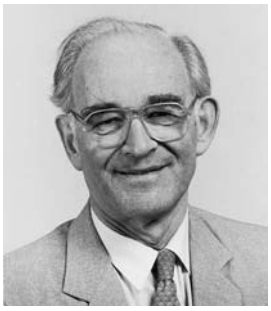
(1897- 1975)

Jacob Aall Bonnevie (Jack) Bjerknes, perhaps now best known for his work on ocean–atmosphere interaction, especially El Niño and the Southern Oscillation (ENSO), was one of the founders of modern theoretical and forecast meteorology. He was part of the Bergen School directed by his father, the theoretical physicist Vilhelm Bjerknes. This group was influential in changing meteorology from a largely observational science to one grounded in physical principles. Bjerknes 1918 paper, *On the Structure of Moving Cyclones*, written at age 20, introduced his frontal cyclone model. This theory, extended in a 1922 paper (with H Solberg), *The Life Cycle of Cyclones and the Polar Front Theory of Atmospheric Circulation*, formed the basis for weather forecasting as well as explicating the principal mechanism for meridional heat transport in the mid-latitude atmosphere. In the 1920s and 1930s he further developed both theory and forecasting practice, including important extensions to waves in the upper troposphere.

Bjerknes 1939 lecture tour in the US became a permanent move after the German invasion of Norway. During the war he organized a training school for air corps weather officers. In 1940, Bjerknes began an association with the University of Los Angeles (UCLA) that lasted until his death. In 1945, the Department of Meteorology was established at UCLA with Bjerknes as head.

In the late 1950s, Bjerknes changed his research direction to ocean–atmosphere interaction. Based largely on data from the 1957 International Geophysical Year, he realized that an El Niño event, then thought of solely as a warming along the South American coast, typically involved a rise in sea surface temperature (SST) in the vast equatorial region from the coast to the dateline. He also noted that it was accompanied by all the atmospheric changes connected to a low phase of Sir Gilbert Walker s Southern Oscillation. In his 1969 and 1972 papers, Bjerknes went beyond the empirical connection to suggest that ENSO was part of a single coupled phenomenon generated by a chain reaction between the ocean and atmosphere in the tropical Pacific. Bjerknes was also the first to propose that the changes in atmospheric heating associated with tropical Pacific SST anomalies cause changes in mid-latitude circulation patterns, generating the global response Walker had found to be associated with the Southern Oscillation. His hypothesis, augmented by the linear ocean dynamics shown to be needed by the observational work of Klaus Wyrtki, is the foundation of our present understanding of ENSO (*see* El Niño and La Niña: Causes and Global Consequences, Volume 1).

MARK CANE *USA*



Bolin, Bert

(1925-)

Bert Bolin, Swedish scientist and science organizer, received a PhD degree in Meteorology at Stockholm University in 1956 based on a thesis dealing with the interaction between wind and pressure fields and its application to numerical weather prediction. His subsequent research has been mainly focused on global biogeochemical cycles and on the global carbon cycle in particular. His 150 scientific papers and books also include work on the use of chemical

and radioactive tracers in the atmosphere and in the sea, the acidification problem, atmospheric greenhouse gases and aerosols, and their role in determining the climate and impacts of climate change.

As a student of meteorology in the late 1940s and early 1950s, Bolin was supervised by Professor C-G Rossby (*see* Rossby, Carl-Gustaf, Volume 1), in whom he found long-lasting inspiration. Bolin was appointed Professor of Meteorology at Stockholm University in 1961, a position he held until his retirement in 1990.

Bolin served as president of the Committee of Atmospheric Sciences (CAS) of the International Council of Scientific Unions (ICSU) from 1964 to 1967, and was involved in the proposal for the creation of the Global Atmospheric Research Programme (GARP). In 1967 he organized the scientific planning conference that led to the launching of GARP jointly by ICSU and the World Meteorological Organization (WMO). He served as the first president of the Joint Organizing Committee (JOC) of GARP from 1967 to 1971. As a member of JOC he organized the first international conference for the formulation of a World Climate Research Programme (WCRP), which was later launched in 1980.

In 1970 Bolin took the initiative leading to the preparation of Sweden s Case Study for the United Nations Conference on the Human Environment, held in Stockholm in 1972. This case study contained the first comprehensive description of the acid deposition problem that had then recently been discovered.

Bolin took part in the first assessment of the climate change issue in 1979, organized by the US National Academy of Sciences. He later led the first broad international assessment (1983–1986), which was asked for by the United Nations Environmental Programme (UNEP), WMO and ICSU. The final report adopted in 1985 served as one of the basic documents in the preparation of the UN Report *Our Common Future: The Brundtland Report.*

Bolin served as chairman (1985–1986) for the committee that proposed the initiation by ICSU of the International Geosphere-Biosphere Programme (IGBP), launched in 1986.

Despite the many previous important achievements as organizer of international scientific cooperation, Bolin s most fundamental contribution in this field is probably his role in leading the Intergovernmental Panel on Climate Change (IPCC) (*see* Intergovernmental Panel on Climate Change (IPCC): an Historical Review, Volume 4), which was set up in 1988 by UNEP and WMO. Under Bolin s leadership (1988–1997) IPCC produced two major assessments of the climate change issue (1990 and 1995). With his deep and broad knowledge of the subject, his moral integrity and his engagement, Bolin represented the community of climate scientists in an outstanding way in the transfer of information to the policy makers during the critical early period of international negotiations on a climate convention.

Bolin is a member of some ten science academies and has received numerous awards including the International Meteorological Organization (IMO) Prize from WMO in 1981, the Tyler Prize from the University of California, USA in 1988, and the Blue Planet Prize from the Asahi Glass Foundation, Japan in 1995.

FURTHER READING

Rodhe, H (1991) Bert Bolin and his Scientific Career, *Tellus*, 43(AB), 3–7.

HENNING RODHE *Sweden*

Broecker, Wallace S

(1931-)

Wallace Smith Broecker is a renowned chemical oceanographer and one of the foremost interpreters of how the Earth operates as an intercoupled biological, chemical, and physical system. He is particularly well known for utilizing isotopic tracers to document his conceptualization of the global ocean circulation as a giant conveyor belt that is driven in large part by conditions in the North Atlantic Ocean.

Broecker was born in Chicago, Illinois on November 29, 1931. After completing his first 3 years at Wheaton College in Wheaton, Illinois, Broecker took a summer job at the Lamont–Doherty Geological Observatory (now Lamont–Doherty Earth Observatory, LDEO) and he has stayed on throughout his entire career. He completed his senior year of undergraduate work at Columbia University and received his AB Degree in 1953. Broecker s research at LDEO, working with carbon-14, provided the basis for his dissertation research, and he received his PhD from Columbia in 1958. He was appointed assistant professor at Columbia University in 1961, and full professor in 1977. Since 1979, Dr Broecker has been the Newberry Professor of Geology.

Broecker began his research in the 1950s with the development of techniques for measuring the radiocarbon content of ocean waters, and using this information to determine the ages and accumulation rates of deep sea and lake sediments. Broecker also developed techniques for using uranium daughter products as time clocks and environmental tracers, thereby creating new tools for use in paleoclimatic studies. Using such techniques, Broecker and his colleagues dated marine shells, for example, found in sediment deposits on the sea bottom, contributing to reconstruction of glacial and interglacial periods and transitions.

Broecker pioneered the study of the circulation of chemical elements in the sea, which provided detailed information about the mixing of surface and deep waters of the ocean that takes place over periods of 1000–2000 years. Urged on by oceanographer Henry Stommel, Broecker was instrumental in the creation of the Geochemical Ocean Sections (GEOSECS) program. This program, carried on in the 1970s, gathered a wealth of isotopic information about the world s oceans, providing a much clear picture of how ocean waters circulated. Broecker s book *Tracers in the Sea*, published in 1982 and co-authored with T H Peng, has become an indispensable guide for geochemists. Using the accumulated base of isotopic information, Broecker constructed a theory of the global ocean circulation, often termed the conveyor belt (*see* Ocean Conveyor Belt, Volume 1). Not only do the isotopic tracers indicate the path taken, but also sediment records can be used to determine the speed of the conveyor belt and to identify periods when it has essentially stopped.

While researching changes in the Earth s climate that occurred in the past 200 000 years, Broecker discovered that major climate shifts may have occurred much more rapidly than previously had been believed possible. For example, Broecker and colleagues found that temperatures in Northern Europe suddenly plummeted about 11 000 years ago and remained low for about 1000 years before snapping back to the warm conditions that exist today. The actual transition from the prevailing warm conditions to the cold conditions of the Younger Dryas (*see* Younger Dryas, Volume 1) may have taken as little as 20 years. Broecker has suggested that this cold spell was caused by a temporary disturbance in the global circulation of the world s oceans (*see* Thermohaline Circulation, Volume 1). Because this global current is linked in complex ways to the atmosphere and the precipitation cycle, he has suggested that the carbon dioxide (CO_2) emissions from fossil fuel combustion could trigger a major disruption of this circulation, causing a significant disturbance of at least the climate of the North Atlantic region, particularly Europe, although he thinks such transitions may not be as likely during warm climatic conditions as during cold conditions.

Broecker s interest in the carbon cycle began early in his career, often working with his colleague Taro Takahashi. In 1965, Broecker served, along with Roger Revelle, C D Keeling (*see* Keeling, Charles David, Volume 1; Revelle, Roger Randall Dougan, Volume 1), Harmon Craig, and Joseph Smagorinsky, on a panel that advised the President s science advisory committee panel on environmental pollution on the emerging issue of the growing concentration of atmospheric CO_2, the first such recognition of this issue by the US government.

A key area of study has been the uptake of carbon by the oceans, which could be tracked using tracers of various types. For example, in the 1980s, Broecker and colleagues used information about the distributions of radioactive fallout from nuclear bomb tests conducted in the 1960s as chemical tracers to study the rate of uptake of fossil fuel CO_2 by the ocean. More recently, Broecker has organized the research program for Biosphere 2, a very large glass house in Arizona that originally was home to the Biospherians, but is now used for scientific research about how increased CO_2 concentrations may affect ecosystems (e.g., how higher CO_2 levels will alter the oceanic composition in ways that weaken coral structures).

In recognition of his wide-ranging studies, Broecker has received many honors. He was elected a member of the US National Academy of Sciences in 1979, and has been a member of the American Academy of Arts and Sciences since 1976. Broecker was the recipient of the Maurice Ewing Medal in 1979, the Arthur L Day Medal in 1984, the AG Huntsman Award for Excellence in the Marine Sciences in 1985, the Urey Medal, the Alexander Agassiz Medal, and the UM Goldschmidt Award in 1986, the Vetlesen Award in 1987, the Joseph Priestly Award and the Wollaston Medal in 1990, and the American Geophysical Union s Roger Revelle Medal in 1995. In 1996, he was named winner of the Blue Planet Prize for his achievements in global environmental research, an honor equivalent in this field to winning the Nobel Prize.

Photo: by Sally Savage.

ACKNOWLEDGMENTS

This biography draws from, among other sources, the biography prepared in recognition of the award of the Blue Planet Prize, and Broecker (2000).

REFERENCE

Broecker, W S (2000) Converging Paths Leading to the Role of the Oceans in Climate Change, *Annu. Rev. Energy Environ.*, 25, 1–19.

MICHAEL C MACCRACKEN *USA*

Budyko, Mikhail Ivanovich

(1920-)

Mikhail Ivanovich Budyko is an internationally recognized Russian (Soviet) environmental scientist whose work has focused on the study of natural and anthropogenic climate change, surface hydrology, and the Earth s radiative balance. He is a pioneer in the development of the theory of anthropogenically induced greenhouse gas climate change and a leader in the field of climatological studies of the thermal balance of the Earth s surface.

Budyko received his Master of Science degree in 1942 from the Physics Division of the Leningrad Polytechnic Institute. He was first employed as a researcher in the Voeikov Main Geophysical Observatory (MGO). While at the MGO, Budyko developed a procedure to calculate the components of the thermal balance of the Earth s surface, including a method that allowed calculation of the components of the heat balance from measurements of the lapse rate of atmospheric temperature and the humidity of the surface layer of the atmosphere. Using these techniques, Budyko compiled the first maps of the components of the annual thermal balance of the southern European part of the USSR, determined the latitudinal distribution of the thermal and moisture balance components of land and ocean surfaces for the Northern Hemisphere, and established the factors governing the characteristics of this distribution. In 1951, Budyko received the degree of Doctor of Science from the MGO, and in 1954 he was chosen to be its Director. In his follow-up research, Budyko developed a method for calculating the radiation balance of land from data on the water balance. For his work on the thermal balance, Budyko was awarded the Lenin National Prize in 1958.

As early as 1961, Budyko recognized the possible initiation of anthropogenic global warming. Primarily through analysis and interpretation of observational data, Budyko developed a quantitative relationship between surface temperature and incoming solar radiation, which he used to formulate one of the earliest one-dimensional global climate models, the *energy-balance model* (*see* Energy Balance Climate Models, Volume 1). Using it he deduced that the Earth s climate might be sensitive to small disturbances in the radiative balance. In 1964, Budyko became a corresponding member of the Academy of Sciences of the USSR.

Since 1970, Budyko and his colleagues have carried out theoretical and experimental investigations to explore the dependence of the photosynthetic productivity of agricultural crops on the principal meteorological factors. Also, the thermal conditions of people in different climates have been studied by calculating the thermal balance of a person s body. For this research Budyko received the Professor Lithke Gold Medal of the Russian Geographical Society in 1972.

Since 1975, Budyko has led research at the Department for Investigating Climatic Changes and the Hydrologic Cycle of the Atmosphere in the State Hydrological Institute (Leningrad). Since 1980, he has taken the lead in applying the paleoanalog approach to project future anthropogenic global warming. By doing this Budyko opened the possibility of estimating climate sensitivity via the reconstruction of greenhouse gas radiative forcing and retrieving past global temperature changes for the same periods. In 1987, Budyko was awarded the International Meteorological Organization (IMO) Prize by the World Meteorological Organization (WMO). In 1989, the Russian Academy of Sciences awarded him the A P Vinogradov Prize, and in 1991 he received a Diploma of the First Degree from the Russian Knowledge Society. In 1992, Budyko became an Academician of the Russian Academy of Sciences, and in 1995 he won its A Grigoryev Prize.

Since 1972, Budyko has played an important role in studies of the greenhouse effect under the auspices of Working Group VIII of the US-USSR Bilateral Agreement on the Protection of the Environment. In 1994, Budyko received the Professor R Horton Medal of the American Geophysical Union for his contribution to the study of the hydrologic cycle, and in 1998 he was awarded the prestigious Blue Planet Prize.

NATALIA G ANDRONOVA *USA*

C

Carbon Dioxide Concentration and Climate Over Geological Times

Elizabeth K Berner and Robert A Berner
Yale University, New Haven, CT, USA

The concentration of atmospheric carbon dioxide (CO_2) has been directly measured since 1958 and during this time has risen rapidly from 315 ppmv to 367 ppmv in 1998. The source of this rise is primarily anthropogenic fossil fuel emissions and to a lesser extent deforestation. Measurements of CO_2 concentrations in air bubbles trapped in glacial ice cores from Antarctica and Greenland have provided a means to extend the record of atmospheric CO_2 concentrations back in time. The concentration of CO_2 was about 280 ppmv in the pre-industrial era (before 1850) and varied only about plus or minus 10 ppmv during the preceeding 1000 years. Ice core data also provide a record of CO_2 and temperature changes for the last 400 000 years, including four glacial- interglacial cycles. CO_2 concentrations dropped 80- 100 ppmv during cold glacial periods to a glacial minimum of ~180 ppmv and rose to an interglacial maximum of 280- 300 ppmv accompanied by a glacial- interglacial surface temperature variation of about 12 °C at Vostok in Antarctica.

Atmospheric CO_2 concentrations from 500 000 to 550 million years ago varied considerably and generally were higher than any time during the past 500 000 years. Information about atmospheric CO_2 in this interval comes from a variety of indirect methods and the variations can be explained by models of the long-term carbon cycle. CO_2 is released by volcanoes and removed from the atmosphere by weathering of silicate rocks and the burial of organic matter. Warm climates are generally associated with high concentrations of CO_2 and cold climates with low CO_2 concentrations. Prior to 550 Myr during the Precambrian, CO_2 fell from very high concentrations shortly after the formation of the Earth to much lower concentrations about 0.5 Myr. This drop in CO_2 was accompanied by a rise in solar radiation and also a rise in atmospheric oxygen (O_2) resulting from the beginning of photosynthesis.

LAST 500 000 YEARS (HOLOCENE AND LATE PLEISTOCENE)

Atmospheric Carbon Dioxide (CO_2) Measurements Since 1958

Beginning in 1958, accurate direct measurements of atmospheric CO_2 have been made on top of Mauna Loa on the island of Hawaii by C D Keeling and associates (Keeling and Whorf, 1999). It has been observed that the CO_2 concentrations which began at 315 ppm in 1958 increased rapidly over time. By 1998, the CO_2 concentration measured was 367 ppmv (Figure 1). The growth rate of CO_2 in the atmosphere increased exponentially for much of the 20th century, except during World War II and since the late 1980s when the rate of increase was roughly linear rather than exponential. The source of this increase in atmospheric CO_2 is largely due to anthropogenic emissions from fossil fuel combustion and cement production, which amounted to about 6.5 gigatons of carbon per year (Gt C year^{-1}) in 1996 (Ledley *et al.*, 1999). The rise in atmospheric CO_2 tends to parallel fossil CO_2 emissions, providing support for isotopic evidence that associates the rise with anthropogenic emissions. Another smaller anthropogenic source is changes in forest storage of carbon due to deforestation which is thought to be partially balanced by forest regrowth (1.6 ± 1.0 Gt C year^{-1}). The CO_2 curve from Mauna Loa also shows a yearly cycle with a drop in CO_2 due to net photosynthesis and carbon storage during the Northern Hemisphere spring and summer and a rise in CO_2 due to net respiration or decay of plant material during the fall and winter.

Atmospheric CO_2 absorbs infrared radiation emitted by the Earth s surface and atmosphere. This tends to warm the atmosphere, which in turn reradiates energy back to the Earth s surface and lower atmosphere, keeping them warmer than they would be otherwise. This is commonly referred to as the *greenhouse effect*. The concern over

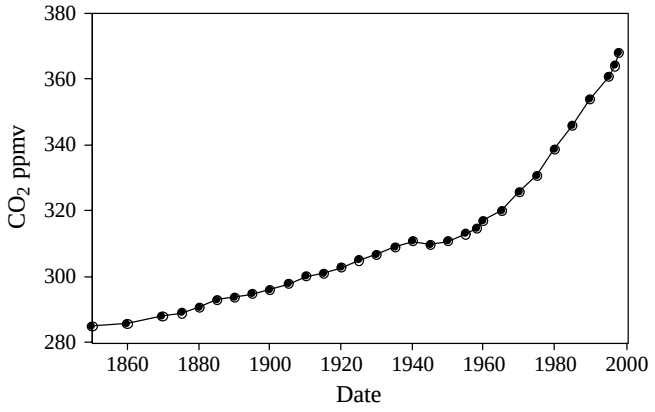


Figure 1 Atmospheric CO_2 concentrations in ppmv from 1850 to 1998. The data, beginning in 1958, are direct measurements made on Mauna Loa, Hawaii (Keeling and Whorf, 1999). The prior data comes from air bubbles in ice cores from Antarctica. (Reproduced using data from Hansen *et al.*, 1998)

possible global warming due to an enhanced greenhouse effect has led to attempts to determine how much change there had been in the CO_2 concentration since pre-industrial times (generally thought to be since about 1850) when anthropogenic changes became important.

Changes in Atmospheric CO_2 During the Industrial Era (1850 to Present) and in Pre-industrial CO_2 During the Last Millenium

Information about atmospheric CO_2 concentrations prior to 1958 comes from air bubbles trapped in glacial ice in Greenland and Antarctica. The ice must be unmelted, unfractured and not contaminated. Antarctic ice appears to yield more consistently reliable CO_2 measurements than Greenland ice. Recent ice core data are compared with direct atmospheric CO_2 measurements for verification. The ice data show a steep rise in CO_2 from about 280 ppm in 1800 to 315 ppm in 1958 when direct measurements began (Figure 2).

Pre-industrial CO_2 levels from ice cores over the last 1000 years vary about ± 10 ppmv from the average value of 280 ppmv (Barnola *et al.*, 1995). The largest change occurred from about 277 ppmv in 1200 AD to 285 ppmv around 1350 AD, followed by a slow decrease to 280 ppmv in 1800 during what was known as the Little Ice Age in medieval Europe (Barnola *et al.*, 1995). These natural changes are much smaller than the 75 ppmv increase during the industrial era (see Figure 2).

Another indication of the introduction of fossil fuel CO_2 to the atmosphere during the period 1800–1950 was the drop in the ratio of atmospheric ^{14}C to total carbon preserved in coral and tree rings in this interval. Fossil fuel contains no ^{14}C because it is old and the ^{14}C has been lost due to radioactive decay.

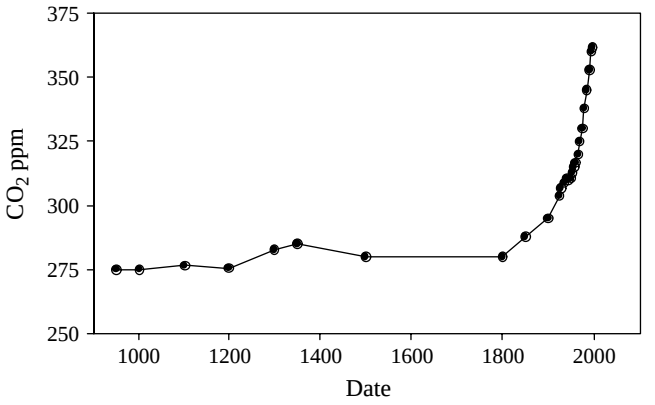


Figure 2 Atmospheric CO_2 from 950 until present based on ice core data until 1958. After 1958 the data is from direct measurements. (Reproduced using data from Hansen *et al.*, 1998; Barnola *et al.*, 1995)

Temperature Changes Since 1850 Compared to Changes Over the Past Millenium

Changes in the CO_2 concentration over the past millennium prior to the industrial period are very small so, not surprisingly, they do not correlate well with temperature changes (Barnola *et al.*, 1995), which are more controlled by changes in some of the various other factors that influence temperature. Surface temperatures would be expected to rise, however, in response to the considerable rise in atmospheric CO_2 over the last 150 years. Various authors have estimated an increased global mean annual temperature of 0.3–0.6 °C in this interval (Nicholls *et al.*, 1996). However, this temperature increase is not much greater than the natural variations in annual temperature due to various natural causes. Proxy temperature measurements such as retreat of valley glaciers (due to melting) also imply warmer temperatures. In addition, based on studies of proxy temperature indicators such as tree ring width and delta ^{18}O in ice cores, Mann *et al.* (1999) concluded that the late 20th century is anomalously warm in comparison with the rest of the millennium, and that the 1990s were the warmest decade and 1998 the warmest year of the past 1000 years. In addition, the 20th century warming overcame a millennium scale cooling trend, that may be due to long-term variations in solar radiation reaching the Earth due to changes in astronomical factors (Mann *et al.*, 1999).

Glacial–interglacial CO_2 and Temperature Changes from Ice Core Records for the Last 420 000 Years

When account is taken of time lags coupling temperature and the carbon cycle, changes in CO_2 and temperature correlate reasonably well over four glacial–interglacial cycles from Vostok, Antarctica, ice core records going back 420 000 years (Petit *et al.*, 1999). CO_2 concentrations

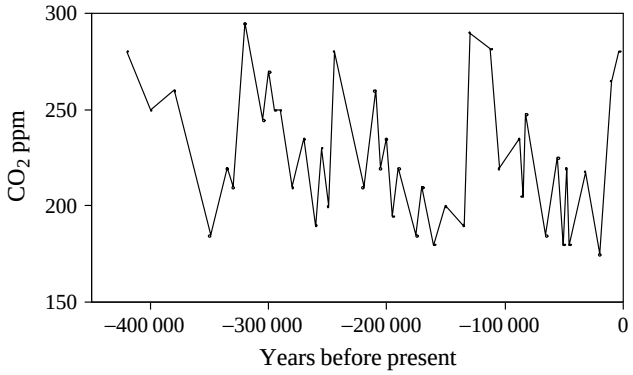


Figure 3 Atmospheric CO_2 for the past 420 000 years based on ice core data from Vostok, Antarctica. (Reproduced using data from Petit *et al.*, 1999)

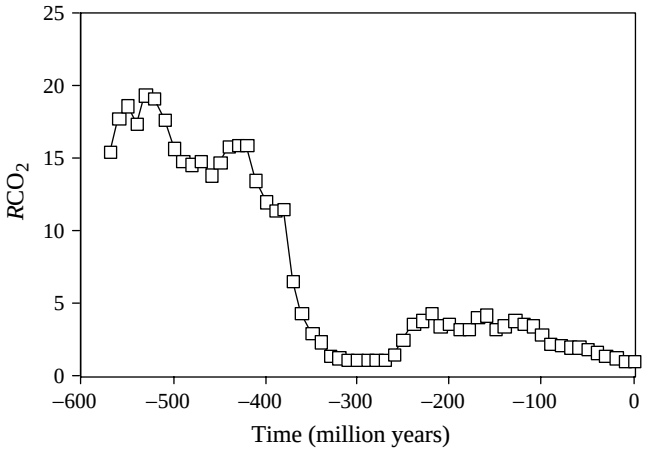


Figure 4 Atmospheric CO_2 over the past 550 million years. The parameter RCO_2 is the ratio of CO_2 at some past time to that for the pre-industrial present (300 ppmv). Results are based on the theoretical long-term carbon cycle model of Berner (1994) as amended by Berner (1997). The theoretical curve agrees, within modeling sensitivity error (±50%), with many different independent estimates of paleo-CO_2 (for a summary see Berner, 1997). Shorter-term CO_2 variations, such as what occurred around 65 million and 55 million years ago, cannot be discerned by the model because it calculates only longer-term average values

are generally lower by 80–100 ppmv during cold glacial periods than during warm interglacial periods (see Figure 3). The glacial–interglacial surface temperature variation (in Antarctica) as calculated from isotopic information in the Vostok ice record indicates a temperature range of 12 °C. CO_2 decreases more slowly to its minimum (~180 ppmv) during glacial times than it rises to its maximum (280–300 ppmv) during interglacial times. The drop in CO_2 occurs several thousand years after the temperature drop according to the ice core record. Methane (CH_4) concentrations also correlate with CO_2 and temperature in Antarctica. The maxiumum CO_2 concentrations recorded in the ice are 280–300 ppmv, considerably less than the present day CO_2 concentration of about 368 ppmv.

The correlation between temperature and CO_2 and methane concentrations suggests that these gases have contributed to glacial–interglacial changes by amplifying the effects of solar radiation changes resulting from variations in the Earth s orbital parameters (precession, obliquity, and eccentricity). The Earth s eccentricity has a strong ~100 000 year cycle that is particularly evident in the records of ice core temperature and gas concentration changes. The sequence suggested by Petit *et al.* (1999) during each glacial termination is that orbital forcing (due to changes in eccentricity) is followed first by amplification by greenhouse gases (CO_2 and methane) and then by deglaciation and ice-albedo feedback. (The melting of the glacial ice reduces the surface albedo, which increases absorption of solar radiation because ice is more reflective than bare earth; this warms the Earth).

CARBON DIOXIDE FROM 550 MILLION TO 500 000 YEARS AGO (PHANEROZOIC TIME)

The Phanerozoic Eon, or last half billion years of Earth history, is the period of time over which the vast majority of organisms we see today and in the fossil record arose and evolved. During this time, the level of atmospheric CO_2 varied considerably (see Figure 4) and most of the time it was higher than during the past half million years. The causes of this variation can be explained in terms of the long-term carbon cycle.

The Long-term Carbon Cycle

The term *carbon cycle*, as normally used, refers to the exchange of carbon between the atmosphere, oceans, biota (mainly terrestrial vegetation), and soils, and it is this cycle that has dominated the controls on atmospheric CO_2 over the past 500 000 years and which acts as a modern sink for anthropogenically produced CO_2. However, there is another cycle that operates more slowly over millions of years and during which it dominates CO_2 concentration. It is the long-term carbon cycle. According to this cycle carbon is exchanged between rocks and a surficial reservoir made up of the combination of the atmosphere, oceans, biota, and soils. Carbon is added to rocks by the burial of calcium and magnesium carbonates and organic matter in sediments. These forms of carbon represent the remains of organisms, both their hard parts (carbonates) and soft parts (organic matter). Some of this carbon is eventually returned to the Earth s surface by the thermal breakdown of carbonates and organic matter at great depth to form CO_2, which is then outgassed to the atmosphere and oceans by volcanoes, hot springs, and by simply seeping out of

the ground. The rest of the carbon is eventually uplifted onto the continents as old organic matter and carbonates in sedimentary rock, which are subjected to the release of carbon to surface waters and to the atmosphere by the process of rock weathering (*see* Carbon Cycle, Volume 2).

Factors Affecting CO_2 and the Long-term Carbon Cycle

A key process in the long-term carbon cycle is the chemical weathering of calcium and magnesium silicate minerals on the continents. This involves the conversion of atmospheric CO_2 to dissolved bicarbonate ions in ground water and river water. After rapid delivery to the sea by rivers, the bicarbonate undergoes reaction with dissolved calcium and magnesium to form calcium and magnesium carbonates. In this way, atmospheric CO_2 is removed from the atmosphere and buried in sedimentary rocks.

The combined processes of Ca–Mg silicate weathering and Ca–Mg carbonate burial have been greatly modified over the past 550 million years. This is partly because of the rise and evolution of land plants. Approximately 400 million years ago, trees began to grow on the land in upland areas and to send down extensive root systems that secreted organic acids for attacking Ca and Mg silicate minerals. This gave rise to enhanced weathering (dissolution) of these minerals and enhanced removal of CO_2 from the atmosphere, ultimately to form Ca–Mg carbonates in marine sediments. This increase in weathering rate is evident in Figure 4 by the large drop in CO_2 that occurred between 400 and 350 million years ago. Added to enhanced weathering by trees was the production of a new type of organic matter, in the form of wood, that is resistant to microbial decomposition and can be buried in sediments. This wood production led to the increased burial of organic matter in sediments, which resulted, after deep burial, in the formation of coal and coal-like organic matter. This accelerated burial of organic matter led to a further drop in atmospheric CO_2 concentration between 350 and 280 million years ago (Figure 4).

The very high values of atmospheric CO_2 found before 400 million years (early Paleozoic) were likely due mainly to the lack of large terrestrial plants, such as trees, to bring about enhanced weathering and CO_2 removal. The moderately high values during the period 250–60 million years ago (Mesozoic Era) were due to a greater rate of degassing of CO_2 to the atmosphere by volcanoes and hot springs, combined with decreased rates of CO_2 uptake by weathering due to a relative lack of high mountains. (High mountains accelerate the chemical weathering of Ca and Mg silicate minerals by causing greater rainfall and by the exposure of the minerals to acid-bearing waters by the erosional removal of protective clays.) The decrease in CO_2 from 100 million years ago to the beginning of the industrial era was due mainly to decreased degassing and the uplift of major mountain belts, such as the Himalayas.

Climate Over the Past 550 Million Years

Accompanying the changes in CO_2 over the past half billion years have been very large changes in climate (Crowley and North, 1991). The early Paleozoic (550–350 million years ago) was marked by warm seas and little evidence of polar glaciation except for a short-lived glaciation around 440 million years ago. The period from about 350 to 280 million years ago was marked by a major Southern Hemisphere glaciation where at times glaciers reached within 30° of the equator. During Mesozoic time (250–60 million years ago) climates were warm and marked by the absence of major polar glaciation and the presence of unusually high temperatures at high latitudes. This is shown by such things as the remains of forests in Antarctica and palm trees and alligators in northern Canada and Greenland, and this was the time of the dinosaurs. During the Cenozoic (past 60 million years) there was gradual cooling of the land and the sea, leading ultimately to the late Cenozoic (Pleistocene) glaciation discussed in a previous section.

These climate changes correlate with levels of atmospheric CO_2 shown in Figure 4, within the limits of a million year time scale resolution. Warm climates are associated with elevated CO_2 levels and cool climates with lowered CO_2 levels. This gives added evidence that the atmospheric greenhouse effect has been important in affecting climate during the geological past. The overall downward trend of the CO_2 curve in Figure 4 is due to an additional factor, the aging of the Sun. Solar output has increased by about 6% since 550 million years ago. This has meant an acceleration of the silicate weathering component of the long-term carbon cycle and, consequently, a gradual lowering of CO_2 concentration. The aging Sun effect helps to explain why the climates of the early Paleozoic and Mesozoic were equally warm, even though the CO_2 level was considerably higher during the early Paleozoic. A warmer Sun combined with a lower CO_2 level during the Mesozoic gives the same warming as a cooler Sun and higher CO_2 level during the early Paleozoic (Crowley and Baum, 1992).

CARBON DIOXIDE BEFORE 550 MILLION YEARS AGO (PRECAMBRIAN)

The Earth originated at about 4.6 Gyr (billion years ago) from the solar nebula of gas and solid material (planetesimals). The planetesimals accreted together to form the Earth, but the primary gaseous atmosphere was lost. The present atmosphere is secondary and was derived from volatiles trapped in the original planetesimals, and released

upon heating. More volatiles were derived from subsequent meteor impacts. The earliest Earth had a steam atmosphere plus CO_2, carbon monoxide (CO) and nitrogen (N_2). At 4.6 Gyr the concentrations of CO_2 are thought to have been more than 300 times the present atmospheric level and the surface temperature was probably about 80–90 °C (versus 15 °C today) (Kasting, 1993). By about 4.5 Gyr the steam rained out to form an ocean, leaving the atmosphere dominated by CO_2, N_2 and water (H_2O), and possibly CH_4 (Kasting, 1997).

The reason that such high early concentrations of CO_2 (or another greenhouse gas such as methane) are thought to have been present is to compensate for a 30% lower solar luminosity at 4.6 Gyr. Without higher concentrations of greenhouse gases in the atmosphere, this reduced solar heating would have caused the Earth s surface temperature to drop below the freezing point of water and should have resulted in near global glaciation. However, the geologic evidence shows that this did not happen and points to a warmer climate due to higher greenhouse gas concentrations. It is known that the first sedimentary rocks (including carbonates derived from CO_2) were laid down in water at 3.8 Gyr and life originated in the presence of liquid water by 3.5 Gyr. Sufficient warmth was needed to prevent liquid water from freezing.

Significant meteor bombardment of the Earth stopped by about 3.8 Gyr and life began about this time (Des Marais, 1994). The atmosphere at 3.8 Gyr was dominated by CO_2 and N_2 with lesser amounts of CH_4, CO, hydrogen (H_2) and reduced sulfur gases, and oxygen was essentially absent near the surface. This contrasts to our present atmosphere containing 21% oxygen (O_2) and 0.036% CO_2.

Atmospheric oxygen was primarily produced by photosynthesis by cyanobacteria and this required the development of life. Among the earliest life forms were probably mats of cyanobacteria (blue–green algae) which produce organic matter by photosynthesis from H_2O and CO_2 and give off excess O_2 to the atmosphere. If the organic matter produced is buried and not broken down, the O_2 remains in the atmosphere and CO_2 is essentially removed from the atmosphere. A fossil record of these algal mats, referred to as stromatolites, is found by 3.5 Gyr, indicating O_2 production by photosynthesis should have been occurring by then. However, atmospheric O_2 probably did not build up rapidly, because of reaction with reduced iron and sulfur.

The first known glaciation did not occur until 2.3–2.5 Gyr. Although reduced solar luminosity at that time was compensated by higher concentrations of atmospheric CO_2 than at present, CO_2 was thought to have dropped sufficiently that by 2.5 Gyr when glaciation occurred, the CO_2 level was only 10–100 times the present atmospheric value (Kasting, 1993, 1997; Rye *et al.*, 1995).

Evidence for low atmospheric O_2 until ~2.0 Gyr comes from several sources. Minerals, such as uraninite and pyrite, which would decompose under today s O_2 levels, were carried in stream gravels before 2.0 Gyr. Soils which formed during this time and are preserved, show evidence of iron transported in solution, which is only possible with low atmospheric O_2. There appears to have been little O_2 at depth in the oceans, allowing iron to occur there in solution. When the dissolved iron was carried into ocean surface water, it encountered dissolved O_2 and precipitated out as the so-called banded iron formations (a rich Precambrian iron ore). These deposits were only formed until about 1.8–1.9 Gyr, by which time atmospheric O_2 was too high for an anoxic deep ocean.

The major way that CO_2 was removed from the pre-Cambrian atmosphere was by weathering Ca and Mg silicate rocks, and the major source of atmospheric CO_2 was from volcanic emissions. Thus, the Archean concentration of atmospheric CO_2 depended upon the balance between these two processes. Because land area was small, CO_2 removal by weathering was probably reduced, which would have contributed to higher CO_2 and temperature (Des Marais, 1994). This was particularly true before large-scale production of buried organic matter by photosynthesis, which contributes to both CO_2 removal from the atmosphere and O_2 production.

After ~2.0 Gyr, atmospheric O_2 appears to have increased considerably and atmospheric CO_2 continued to drop. By 0.6 Gyr when another glaciation occurred at the end of the Precambrian, atmospheric CO_2 was ~1–10 times the present atmospheric level (Kasting, 1993).

See also: Earth System History, Volume 1.

REFERENCES

Barnola, J M, Anklin, M, Pocheron, J, Raynaud, D, Schwander, J, and Stauffer, B (1995) CO_2 Evolution During the Last Millenium as Recorded by Antarctic and Greenland Ice, *Tellus*, 47B, 264–272.

Berner, R A (1994) GEOCARB II: a Revised Model for Atmospheric CO_2 Over Phanerozoic Time, *Am. J. Sci.*, 294, 56–91.

Berner, R A (1997) The Rise of Plants and their Effect on Weathering and Atmospheric CO_2, *Science*, 276, 544–546.

Crowley, T J and Baum, S K (1992) Modeling Late Paleozoic Glaciation, *Geology*, 20, 507–510.

Crowley, T C and North, G R (1991) *Paleoclimatology*, Oxford University Press, Oxford.

Des Marais, D J (1994) The Archean Atmosphere: Its Composition and Fate, in *Archean Crustal Evolution*, ed K C Condie, Elsevier, Amsterdam, 505–523.

Hansen, J E, Sato, M, Lacis, A, Ruedy, R, Gegen, I, and Mattews, E T (1998) Climate Forcings in the Industrial Era, *Proc. Natl. Acad. Sci. USA*, 95, 12 753–12 758.

Kasting, J F (1993) Earth s Early Atmosphere, *Science*, 25, 920–926.

Kasting, J F (1997) Planetary Atmospheres – Warming Early Earth and Mars, *Science*, 276, 1213–1215.

Keeling, C D and Whorf, T P (1999) Atmospheric CO_2 Records from Sites in the SIO Air Sampling Network, in *Trends: a Compendium of Data on Global Change*, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, TN.

Ledley, T S, Sundquist, E, Schwartz, S E, Hall, D K, Fellows, J, and Killeen, T (1999) Climate Change and Greenhouse Gases, *EOS*, 80(39), 453–458.

Mann, M E, Bradley, R S, and Hughes, M K (1999) Northern Hemisphere Temperatures During the Past Millenium: Inferences, Uncertainties, and Limitations, *Geophys. Res. Lett.*, 26, 759–762.

Nicholls, N, Gruza, G V, Jouzel, J, Karl, T R, Ogallo, L A, and Parker, D E (1996) Observed Climate Variability and Change, in *Climate Change 1995: The Science of Climate Change*, eds J T Houghton, L G M Filho, B A Callendar, N Harris, A Kattenberg, and K Maskell, Cambridge University Press, Cambridge, 133–192.

Petit, J R, Jouzel, J, Raynaud, D, Barkov, N I, Barnola, J-M, Basil, I, Bender, M, Chappellaz, J, Davis, M, Delaygue, G, Delmotte, M, Kotlyakov, V M, Legrand, M, Lipenkov, V Y, Lorius, C, Pepin, L, Ritz, C, Saltman, E, and Stievenenard, M (1999) Climate and Atmospheric History of the Past 420 000 Years from the Vostok Ice Core, Antarctica, *Nature*, 399, 429–436.

Rye, R, Kuo, P H, and Holland, H D (1995) Atmospheric Carbon Dioxide Concentrations Before 2.2 Billion Years Ago, *Nature*, 378, 603–605.

FURTHER READING

Berner, R A (1999) A New Look at the Long-term Carbon Cycle, *GSA Today*, 9, 1–6.

Berner, E K and Berner, R A (1996) *The Global Environment: Water, Air and Geochemical Cycles*, Prentice Hall, Upper Saddle River, NJ, 1–376.

Schimel, D, Enting, I G, Heimann, M, Wigley, T M L, Raynaud, D, Alves, D, and Siegenthaler, U (1995) CO_2 and the Carbon Cycle, in *Climate Change 1994: Radiative Forcing of Climate Change and an Evaluation of the IPCC 1992 Emission Scenarios*, eds J T Houghton, L G Meira Fihlo, J Bruce, B A Hoesung Lee, B A Callender, E Haites, N Harris, and K Maskell, Cambridge University Press, Cambridge, 37–71.

Carbon Dioxide, Recent Atmospheric Trends

Kaz Higuchi[1] and Takakiyo Nakazawa[2]

[1]*Meteorological Service of Canada, Toronto, Canada*
[2]*Tohoku University, Tohoku, Japan*

Analyses of air bubbles retrieved from ice cores tell us that the atmospheric carbon dioxide (CO_2) concentration, before the Industrial Revolution began in Europe during the 1800s, was around 280 parts per million by volume (ppmv). Since that time, the increase in the CO_2 concentration has paralleled the amount of CO_2 emitted by fossil fuel combustion, with about 50% of the fossil fuel CO_2 remaining in the atmosphere, and the rest going into the land biosphere and the oceans. The present atmospheric CO_2 concentration is close to 370 ppmv, representing more than a 30% increase in the CO_2 concentration in the atmosphere since the mid 1800s. In addition to the global fossil fuel CO_2 emission of over 6 gigaton of carbon (Gt C) year^{-1} (1 Gt C = 10^{15} grams C), it is believed by many scientists that the deforestation activities in the tropics are emitting a net CO_2 ux of about 1.5 Gt C year^{-1} into the atmosphere at the present time. This means that the land biosphere and/or oceans are absorbing more CO_2 from the atmosphere than hitherto recognized. The carbon cycle scientists call this the missing carbon sink problem.

The future level of the atmospheric CO_2 is very dif cult, if not almost impossible, to predict accurately. It depends greatly on human behavior, because the main cause of the increase in the atmospheric CO_2 concentration over the last 150 years or so has been the anthropogenic combustion of fossil fuels to satisfy the almost exponential need for power to feed the lifestyles of the industrialized nations. The oceans and the land biosphere have been absorbing almost half of what humans have been emitting into the atmosphere. But there is some scienti c evidence now to indicate that, under a warming climate, these two major carbon reservoirs will take up less and less CO_2 from the atmosphere. For example, under global warming, the sea surface temperature will increase, causing the oceans to absorb less CO_2 from the atmosphere. Also, the soil temperature will increase, causing the land biosphere to emit more CO_2 into the atmosphere through accelerated decay and respiration. If the industrialized nations continue to do what they have been doing, then the atmospheric CO_2 concentration will certainly double. This problem will be compounded by the almost certain possibility that developing nations will increasingly burn fossil fuels to satisfy their needs for power. The biggest problem we have right now is to predict when the doubling, or tripling for that matter, of the atmospheric CO_2 concentration will occur.

CO_2 CONCENTRATION IN THE ATMOSPHERE

The first systematic and accurate monitoring of the CO_2 concentration in the atmosphere started during the International Geophysical Year (IGY), which took place from July 1, 1957 to December 31, 1958 (*see* IGY (International Geophysical Year), Volume 1). The IGY involved more than 5000 investigators from over 60 countries; these scientists measured and observed major phenomena of the Earth

simultaneously. One of the those who participated was a young scientist from the Scripps Institution of Oceanography, La Jolla, California. His name was Charles David Keeling (*see* Keeling, Charles David, Volume 1). His responsibility was to initiate and maintain measurements of the atmospheric CO_2 concentration at the South Pole and Mauna Loa, Hawaii sites. These two locations are remote and were chosen to represent background conditions of the CO_2 gas in the atmosphere. Sampling and analysis for atmospheric CO_2 was not an easy task at that time, being fraught with technical difficulties, some of which we still face today. (A brief description of measurement procedures for obtaining CO_2 concentration values is provided in Appendix A.) Keeling showed tenacity in his work and overcame many of these difficulties. Thanks to his efforts, we have a 45-year systematic and accurate record of atmospheric CO_2 concentration that can be used to examine long-term trends, as well as inter-annual variations in the seasonal cycle of atmospheric CO_2 concentration.

Figure 1 shows the time series of CO_2 from Mauna Loa (on a mountain above the trade wind inversion) and the South Pole, from 1958 to the present. One of the striking features is that the Mauna Loa station time series shows a very distinct sinusoidal oscillation (seasonal cycle) in the Northern Hemisphere, with annual maxima and minima in early spring and late summer, respectively. This variation is caused mostly by the metabolic cycle of the mid to high latitude land biosphere, with photosynthetic uptake of atmospheric CO_2 by the vegetation dominating the release of CO_2 by respiration and decay during the summer season. The CO_2 seasonal cycle observed at the South Pole shows very small amplitude compared to that seen at Mauna Loa, because there is very little land biosphere in the Southern Hemisphere. It is also opposite in phase from Mauna Loa because the seasons are out of phase between the Northern and Southern Hemispheres. There is also an exchange of air between the two hemispheres, but this atmospheric mixing takes about 6–12 months, adding to the phase difference.

If the seasonal cycle of the atmospheric CO_2 concentration is caused mainly by the seasonal cycle of the land biospheric activity, then it is logical to assume that any change in the biological activity of the land vegetation would be reflected in the inter-annual variation of the atmospheric CO_2 seasonal cycle. Many scientists have exploited this property of the global carbon cycle to find out if the land biosphere has become more active in recent years, thus absorbing more and more CO_2 from the atmosphere. One such study was conducted by Keeling and his colleagues (Keeling *et al.*, 1996). They showed that the seasonal amplitudes observed at Mauna Loa, Pt Barrow (Alaska) and Alert (northern tip of Ellesmere Island in the Canadian Arctic) have been increasing during recent times. The scientists interpreted the results to mean that the Northern Hemisphere land biosphere has become more active. They also found that the drawdown of the atmospheric CO_2 concentration (that portion of the seasonal cycle that begins with

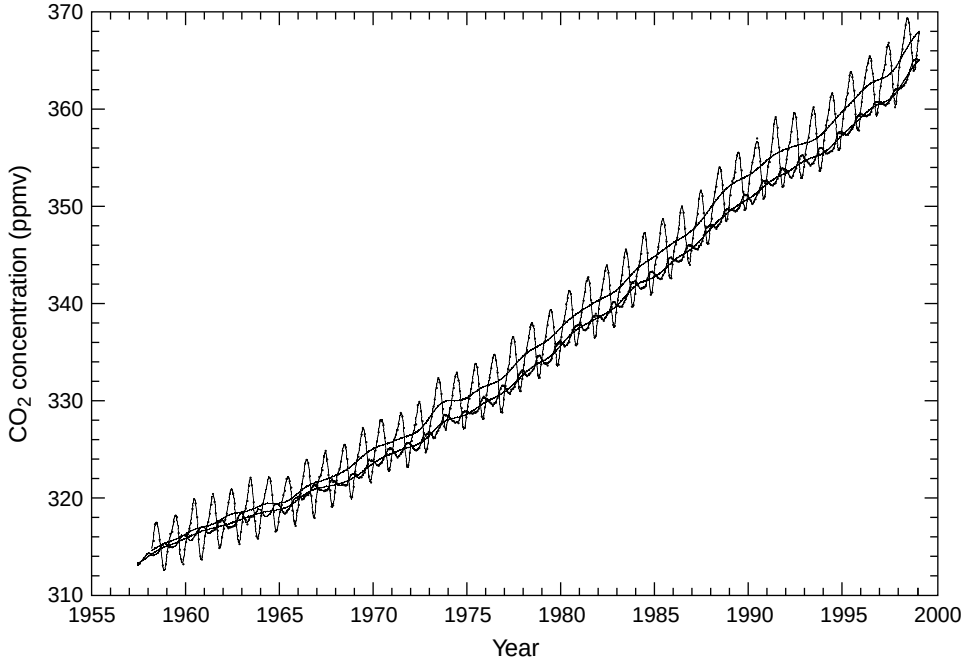


Figure 1 Time series of the atmospheric CO_2 concentration measured at Mauna Loa, Hawaii (larger seasonal cycles) and at the South Pole (smaller seasonal cycles), from 1958 to 1998. The secular increase observed at Mauna Loa is indicated by the upper solid line, while that at the South Pole is indicated by the lower solid line

a decrease in the CO_2 concentration each spring) begins earlier now. This result indicates that the photosynthetic process is starting earlier each year, causing the land biosphere to absorb an increasing amount of atmospheric CO_2 each year. The researchers have attributed the earlier starting of the photosynthesis to the increase in the global average surface temperature.

Another interesting feature of Figure 1 is that the Mauna Loa values show an annual concentration level that, year-to-year, is about 3–4 ppmv higher than the concentrations observed at the South Pole. This difference, or gradient, has been maintained from year to year over the last few decades. The secular upward trend we see at these stations (as well as at other sites added to the global network in the last 20 years, now totaling over 50 stations) in recent atmospheric CO_2 concentration is caused mainly by the accelerating increase in the emissions of CO_2 from burning of fossil fuels by the industrial nations located mainly in the Northern Hemisphere. The inter-hemispheric difference is therefore noticeably smaller at the beginning of the record. Figure 2 shows a worldwide network of atmospheric CO_2 sampling sites operated by the NOAA/CMDL; included in the network are those sites where samples are collected co-operatively with agencies of other countries.

At the beginning of the CO_2 measurement programs at Mauna Loa and the South Pole, the annually averaged atmospheric concentration was around 315 ppmv. Since then, the annually averaged concentration has risen to almost 370 ppmv in the Northern Hemisphere in the year 2000, an increase of 17% from the value observed in the late 1950s. Averaged over a number of years, it has been found that the atmosphere retains about 50% (or even a little more) of the CO_2 emitted into the atmosphere from fossil fuel combustion. Figure 3 shows fossil fuel CO_2 emission from 1850 to 1997. Using mathematical models of varying hierarchical complexity of the global biogeochemical cycle of carbon, scientists have determined that the remaining portion of the fossil fuel CO_2 in the atmosphere is absorbed mainly by the world s oceans. (It should be noted here that many scientists involved in carbon cycle research accept the fact that there is an additional anthropogenic net source of CO_2 resulting from deforestation in the tropics. An amount of atmospheric CO_2 almost equivalent to that emitted by the deforestation activities is conjectured by a number of scientists to be absorbed by the terrestrial biosphere in the Northern Hemisphere. Since the land biospheric uptake is so uncertain, it is a hotly debated issue in the scientific community, and is sometimes referred to as the *missing carbon sink* issue.) An example of how sensitive the atmospheric CO_2 concentration is to fossil fuel combustion emission can be seen in Figure 3. During the so-called oil crisis in 1973 and 1974, the consumption of fossil fuels (and oil in particular) by the industrial nations dramatically declined. This contributed to the dip seen in the secular CO_2 trend at Mauna Loa around the same time; the dip was also observed at other CO_2 monitoring stations located world-wide. The growth rate of the atmospheric CO_2 concentration also declined during the major energy crisis in the early 1980s (see Figure 3), but its impact on the actual CO_2 concentration appears to be less obvious in Figure 1 than the impact of the early-1970s oil crisis. This is because such natural phenomena as the El Niño and the Mount St Helen (State of Washington, USA) volcanic explosion in the early 1980s also influenced the atmospheric CO_2 concentration growth. It should be added that an El Niño did occur during the 1973–1974 period of the first oil crisis identified in Figure 3, but it was much weaker than the one that occurred during the 1982–1983 period.

ATMOSPHERIC CO_2 FROM ICE CORES

If the recent changes in the CO_2 concentration level observed in the atmosphere are caused mainly by the combustion of fossil fuels by the industrial nations for power, then it is only logical to ask what the atmospheric CO_2 level was before industrialization began. To find the answer to this question, we go to an icy world.

In places like Antarctica, Greenland and high-altitude mountains, annual-averaged air temperatures are below $0\,^{\circ}C$, the melting point of snow and ice. Even in summer, air temperatures at these locations do not usually rise above $0\,^{\circ}C$. Snow that falls does not melt and thus accumulates year after year. A fresh layer of snow has a fluffy texture because the space between flakes is filled with air at the time of the snowfall. As more and more layers of snow accumulate on top of this layer, it becomes more and more compacted under the overlying pressure, eventually turning into ice. Some of the air that filled the space between the snowflakes gradually becomes trapped in the ice as air bubbles. This has popularly been called *antique air*. During the transition of the layer from snow to ice, the air in the layer is still in contact with the atmosphere. It takes many decades before the air bubbles become completely sealed off from the atmosphere. This means that the air bubbles are younger than the ice layer in which they are embedded. The information on the relative gaseous composition of the atmosphere contained in the air bubbles reflects atmospheric composition at a time more recent than the age of the ice layer from which these bubbles are retrieved. This fact is taken into account when scientists assign an age to the CO_2 concentration value analyzed from air bubbles obtained from a certain layer in the ice core. If we can retrieve and analyze the mix of gases in air bubbles present throughout the depth of an ice core, say, from Antarctica (or Greenland or any other location where snow does not melt during the year), then we can find out what the atmospheric

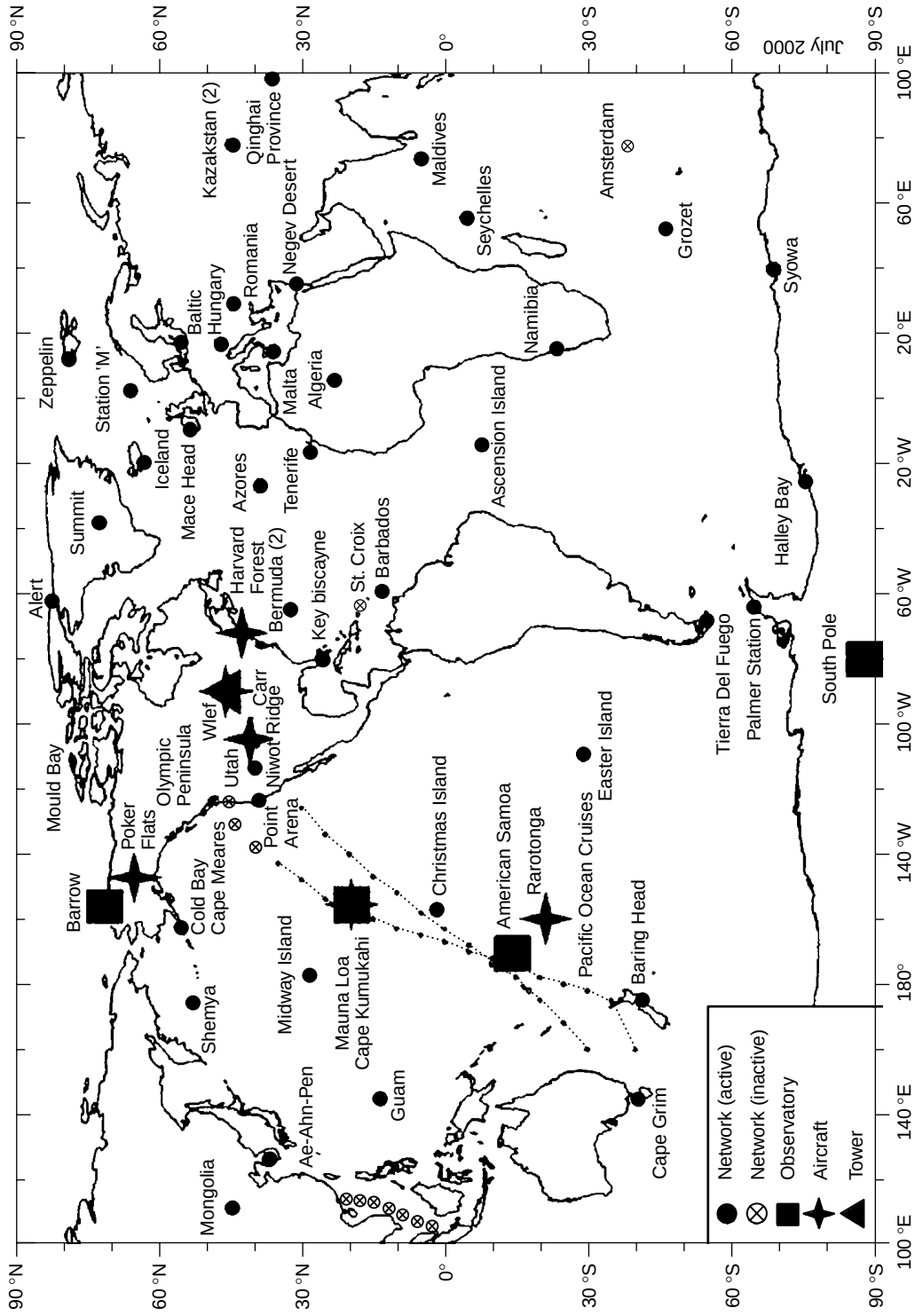


Figure 2 The world-wide network of CO_2 sampling stations operated by National Oceanic and Atmospheric Administration/Climate Monitoring and Diagnostics Laboratory (NOAA/CMDL). The network includes those stations where atmospheric samples are collected co-operatively with agencies of other countries. Inclusion of all stations operated by other countries would total more than 50 sites at the present time. (*Source:* http://www.cmdl.noaa.gov)

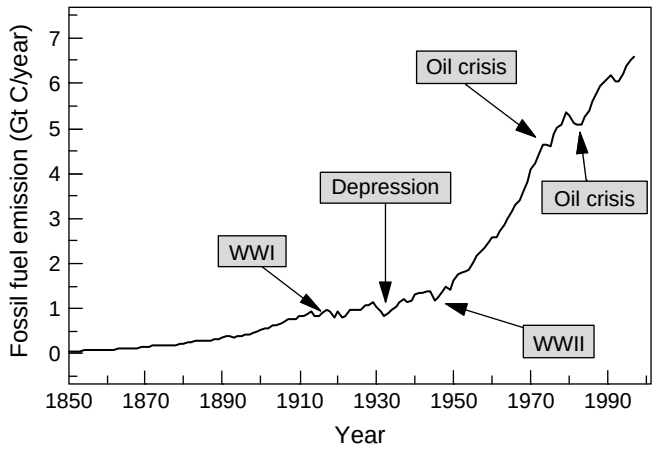


Figure 3 Fossil fuel CO_2 emissions, in Gt C/year (or billions of metric tons per year), from 1850 to 1997. Times of major economic perturbation are identified. Over the last 30 years, two major energy crises (one in the early 1970s, and another one in the early 1980s) have contributed to a slight, short-lived reduction of the total global fossil fuel CO_2 emission. (Data obtained from the Carbon Dioxide Information Analysis Center, Department of Energy, USA)

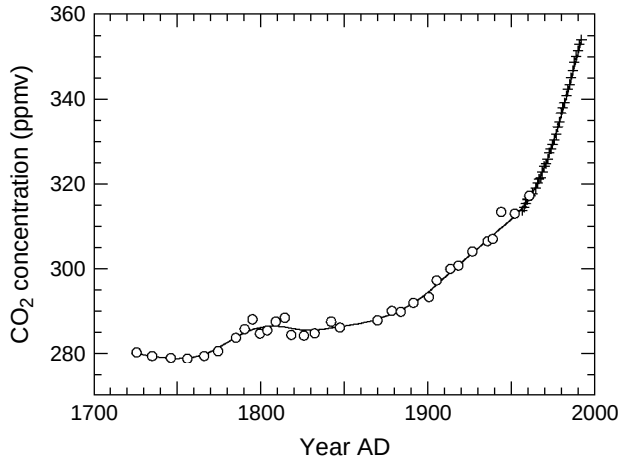


Figure 4 Atmospheric CO_2 concentration values analyzed from trapped air bubbles extracted from an ice core obtained at an Antarctic site known as H15. The CO_2 concentration in the atmosphere has risen from about 280 ppmv prior to the Industrial Revolution to about 355 ppmv around 1990. By the end of 20th century, the level had risen to almost 370 ppmv

CO_2 concentration levels were in the recent past. (A brief description of the retrieval and analysis procedures is given in Appendix B.)

Figure 4 shows the atmospheric CO_2 concentration values from the early 1700s to around 1960, as obtained from analysis of the composition of air trapped air bubbles in a high quality ice core recovered from an Antarctic site known as H15 (69°05′ S, 40°47′ E). The H15 ice core drilling site is located at 1057 m above sea level, and has a mean annual air temperature of −20.5 °C. Prior to the Industrial Revolution that started in Europe around the mid 1800s, the air bubbles from the H15 ice core clearly show CO_2 concentration levels around 280–285 ppmv. From around the mid 1800s, the atmospheric CO_2 concentration starts to rise exponentially, merging very nicely into the instrumental CO_2 record shown in Figure 1. So, since the beginning of the Industrial Revolution, the atmospheric CO_2 concentration level has increased by over 30%, not a trivial amount! This rise in the concentration is consistent with the exponential release of CO_2 into the atmosphere from the burning of fossil fuels by the industrialized nations. Analyses of ice cores from other sites in Antarctica and Greenland show a similar rise in the atmospheric CO_2 concentration.

INTER-ANNUAL VARIATION IN THE NET CO_2 FLUX FROM NATURAL SOURCES

The relationship between the rise of the atmospheric CO_2 concentration level and the increasing amount of CO_2 emitted by anthropogenic activities, particularly the fossil fuel combustion, is valid on a time scale of several years or more. On an inter-annual year-to-year basis, however, the relation is weaker because there are other natural processes that add CO_2 to, or remove CO_2 from the atmosphere.

A close examination of the CO_2 time series in Figure 1 shows that the rate of increase in the atmospheric CO_2 concentration is not exponentially smooth, and is not always correlated very well with the anthropogenic emission rate of CO_2 into the atmosphere from year to year. There are times, such as the oil crisis of 1973–1974, when the lower emission of fossil fuel CO_2 is reflected in a slowed rate of growth of CO_2 in the atmosphere. There are also times when changes in the atmospheric CO_2 concentration from one year to the next depend significantly on the processes governing the net flux of CO_2 between the atmosphere and the other two major carbon reservoirs in the oceans and the land biosphere. Thus, in order to explain the secular increase and the inter-annual variability of the atmospheric CO_2 concentration, as well as developing techniques to predict future concentration levels, we need to achieve a greater understanding of the relative contributions of the oceans and the land vegetation to the global carbon cycle.

In order to obtain quantitative estimates of the relative contributions that the oceans and the land biosphere make to the secular increase and the inter-annual variability in the atmospheric CO_2 concentration, scientists have exploited chemical and physical properties of carbon isotopes in various mathematical models of the global carbon cycle. One of the isotopic signatures of CO_2 is labelled $\delta^{13}C$. It

is a ratio of ^{13}C to ^{12}C ($^{13}C/^{12}C$), and is defined as:

$$\delta^{13}C = \left(\frac{\left(\frac{^{13}C}{^{12}C}\right)_{\text{sample}}}{\left(\frac{^{13}C}{^{12}C}\right)_{\text{standard}}} - 1 \right) \times 1000$$

At the present time, the atmospheric reservoir has a $\delta^{13}C$ value of $-8\,\text{mil}^{-1}$ (parts per thousand), while the oceanic and the land biospheric reservoirs have, on average, about $+1\,\text{mil}^{-1}$ and $-25\,\text{mil}^{-1}$, respectively. The ^{13}C isotopic signature of each carbon reservoir is caused by the isotopic fractionation effect. For example, when a plant absorbs CO_2 during its photosynthesis, it prefers CO_2 molecules containing ^{12}C (six protons and six neutrons in each carbon atom) rather than those containing ^{13}C (six protons and seven neutrons in each carbon atom) because they are lighter. This differentiation, or isotopic fractionation, process causes, during photosynthesis, lighter CO_2 molecules (those containing ^{12}C) to be absorbed into the plant preferentially, leaving heavier ^{13}C to be left in the atmosphere. That is why the land biosphere shows a deficit in ^{13}C ($-25\,\text{mil}^{-1}$), compared to that of the atmosphere ($-8\,\text{mil}^{-1}$). There is very little net isotopic fractionation in the CO_2 exchange across the air-sea interface.

Both ^{12}C and ^{13}C are called stable isotopes of CO_2. However, there is one more isotope of CO_2, labeled ^{14}C, which is radioactive. Typically, every millionth CO_2 contains ^{14}C. Radioactive CO_2 is created as a by-product of cosmic radiation bombardment with the atmosphere. Its half-life is around 5700 years, and after about 40 000 years the amount becomes so small that it is very difficult to measure. Fossil fuel CO_2 contains practically no radioactive ^{14}C. Therefore, it decreases the $^{14}C/^{12}C$ ratio ($\delta^{14}C$) in the atmosphere; this process is called the *Suess Effect* (*see* Suess Effect, Volume 2). Therefore, measurements of $\delta^{14}C$ in atmospheric CO_2 can be used to differentiate the contribution, to the atmospheric CO_2 variability, of the fossil fuel CO_2 from that of the land vegetation; these two carbon reservoirs have similar $\delta^{13}C$ signature but very different $\delta^{14}C$ signature. Explosions of nuclear bombs in the late 1950s and the early 1960s have made the task very difficult, however.

When CO_2 enters the atmosphere from either the oceans or the land biosphere, it carries with it the $\delta^{13}C$ value of its source. By doing so, it can change the atmospheric $\delta^{13}C$ value. Fossil fuels have a similar $\delta^{13}C$ signature to that of the land vegetation because fossil fuels are just fossilized vegetation. By measuring the inter-annual variations in the atmospheric CO_2 concentration and its corresponding $\delta^{13}C$ value, scientists can estimate inter-annual variations in the net flux of CO_2 from the oceans and the land biosphere. This mathematical technique is called the *inverse method*. In this method, the effect of the fossil fuel CO_2 emission on the atmospheric $\delta^{13}C$ value is eliminated quite early in the calculation. Scientists can do this because they know relatively well how much fossil fuel CO_2 is being emitted into the atmosphere every year.

Figure 5 shows a result of a study conducted by a group of scientists from Japan and Canada using the inverse method (Morimoto *et al.*, 2000). The result is compared to those obtained from studies by Roger Francey from Australia and his colleagues (Francey *et al.*, 1995) and by CD Keeling from the US and his colleagues (Keeling *et al.*, 1995). Flux anomaly values in each study are obtained by subtracting an average calculated over the period of data record. Figure 5 shows the relative change from year to year in the net CO_2 flux across the atmosphere–ocean and the atmosphere–land biosphere interfaces. There are differences among the results of these studies; however, within the context of trying to obtain some understanding of the secular and inter-annual changes in the atmospheric CO_2 record, we note that the variations in the net flux from either the oceans or the land biosphere can be almost as large as 50% of the fossil fuel CO_2 flux of about 5–6 Gt

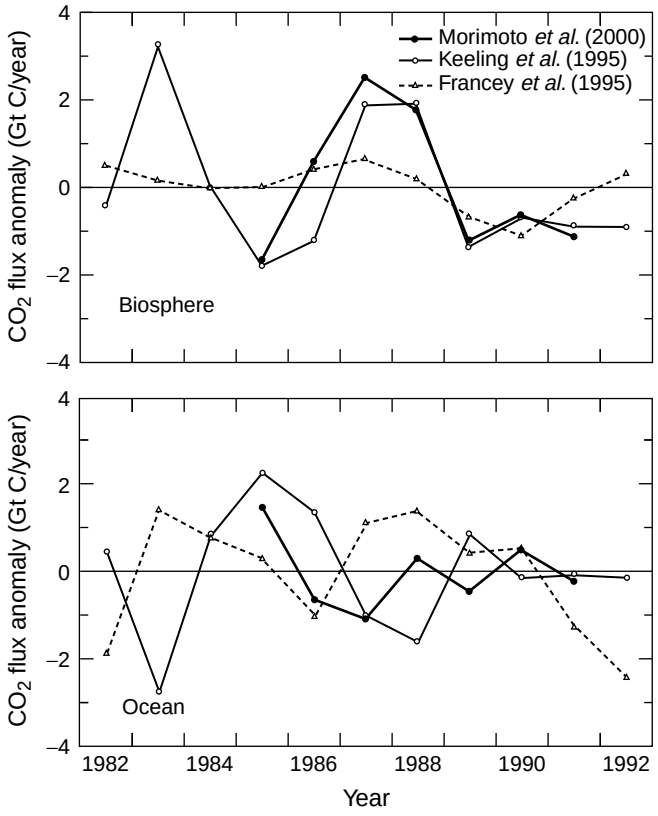


Figure 5 Calculated inter-annual variation in the net CO_2 flux anomaly across the atmosphere–ocean and the atmosphere–land biosphere interfaces. Note that the year to year variability can be as large as 50% of the annual fossil fuel CO_2 emission into the atmosphere. (*Source:* Morimoto *et al.*, 2000)

C per year So the future year-to-year trend in atmospheric CO_2 depends greatly on the behavior of the land biospheric and the oceanic carbon reservoirs, particularly in determining how much of the anthropogenic CO_2 emitted into the atmosphere will be absorbed by these two large carbon reservoirs.

See also: Carbon and Energy: Terrestrial Stores and Fluxes, Volume 2; Carbon Cycle, Volume 2; Boreal Forest Carbon Flux and it's Role in the Implementation of the Kyoto Protocol Under a Warming Climate, Volume 4.

APPENDIX A

One of the simplest and most economical ways of measuring atmospheric CO_2 concentrations is to take samples of air in small containers (flasks) on a regular basis, usually about once every week. These containers vary in size and composition, depending on the laboratory taking the samples. Nonetheless, they are usually made of Pyrex glass, aluminum or stainless steel. They usually vary in size from 2 to 5 liters. Once the atmospheric samples are taken, they are shipped back to a main laboratory for analysis of the CO_2 concentration.

At these laboratories, the air samples are injected into a CO_2 analyzer called a non-dispersive infrared (NDIR) analyzer. The NDIR system is basically composed of two cells, one of which is filled with a standard gas of known CO_2 concentration. (For the convenience of our discussion, we can call the cell containing the standard gas the standard gas cell.) Infrared (IR) light is emitted from one side of the standard gas cell and is registered as an electric current at the receptor on the other side. Because CO_2 is a good absorber of the IR radiation, the amount of IR light that reaches the receptor depends on the CO_2 concentration in the standard gas; the more CO_2, the less IR and thus less electric current.

The other cell in the NDIR system is the one in which we put an air sample whose CO_2 concentration we would like to determine. Into this sample gas cell we shine IR light, as we did in the standard gas. By comparing the electric currents from the standard gas cell and the sample gas cell, we can determine the CO_2 concentration of the air sample. The overall precision is about ± 0.05 ppmv.

Flask sampling is a good monitoring method for measuring the long-term secular trend in atmospheric CO_2 concentrations, as well as some basic aspects of the CO_2 seasonal cycle. However, some of the important scientific issues regarding temporal and spatial distributions of carbon sinks and sources require more continuous monitoring of the behavior of the atmospheric CO_2 level. To provide a temporal resolution of minutes to hours, the NDIR system is used directly at the monitoring station. This is typically done at such reference stations as Mauna Loa, South Pole, Pt Barrow, Alert and Cape Grim (Tasmania).

APPENDIX B

Ice coring procedures allow scientists to retrieve long cylindrical cores of ice from ice caps on Greenland and Antarctica, as well as from glaciers on top of tall mountains. Deeper layers of ice in the ice core represent older ice containing correspondingly older bubbles of air. Thus, analyses of air bubbles from successive layers of ice represent historical time series of the composition of trace gases in the atmosphere. The extraction and analysis procedures for CO_2 in the air bubbles differ from laboratory to laboratory, but they usually follow the procedure outlined below.

For the purpose of extracting air samples, ice samples of several 10 to a few 100 g each are typically used for determining the CO_2 concentration only. The ice sample is placed in a stainless steel chamber after removing contamination from the surface by using a bandsaw and a knife. The ice sample is then placed in an evacuated chamber at $-20\,^{\circ}$C for at least two hours, to sublimate the ice surface for further cleaning. After the cleaning process is finished, the ice sample is crushed into fine powder to release air from the bubbles. After passing through cold traps held at $-100\,^{\circ}$C to remove water vapor, the air released from the ice sample is collected into a sample tube cooled to $-269\,^{\circ}$C by liquid helium. Using a gas chromatograph, the CO_2 concentration of the air samples thus extracted is determined by comparison with standard gas mixtures with concentrations ranging from 200 to 300 ppmv. The overall precision of this kind of air extraction and analysis procedure is estimated to be better than ± 1 ppmv. A more technically-based description can be found in a paper published by a group of Japanese scientists (Nakazawa *et al.*, 1993).

REFERENCES

Francey, R J, Tans, P P, Allison, C E, Enting, I G, White, J W C, and Trolier, M (1995) Changes in Oceanic and Terrestrial Carbon Uptake Since 1982, *Nature*, 373, 326–330.

Keeling, C D, Whorf, T P, Wahlen, M, and van der Plicht, J (1995) Interannual Extremes in the Rate of Rise of Atmospheric Carbon Dioxide Since 1980, *Nature*, 375, 666–670.

Keeling, C D, Chin, J F S, and Whorf, T P (1996) Increased Activity of Northern Vegetation Inferred from Atmospheric CO_2 Measurements, *Nature*, 382, 146–149.

Morimoto, S, Nakazawa, T, Higuchi, K, and Aoki, S (2000) Latitudinal Distribution of Atmospheric CO_2 Sources and Sinks Inferred by δ^{13}C Measurements From 1985 to 1991, *J. Geophys. Res.*, 105, 24 315–24 326.

Nakazawa, T, Machida, T, Esumi, K, Tanaka, M, Fujii, Y, Aoki, S, and Watanabe, O (1993) Measurements of CO_2 and CH_4 Concentrations in Air in a Polar Ice Core, *J. Glaciol.*, 38, 209–215.

Carbon Monoxide

The gas *carbon monoxide* (CO) is produced by natural and human activities. Roughly half to two-thirds of atmospheric CO is emitted from fossil fuel combustion and industrial activities. The remaining CO comes from biomass burning, oceanic and land plant emissions, and the oxidation of methane (CH_4) and non-methane hydrocarbons. Roughly 1800–2800 Tg CO year^{-1} are released each year. The largest sink for CO is reaction with OH, accounting for 85% of the CO. An additional, smaller sink is uptake by soils.

Present day concentrations of CO range from 40–200 parts per billion volume (ppbv), with a global average of roughly 90 ppbv. CO results mainly from land-based sources. Thus, CO concentrations are higher in the Northern Hemisphere, which has both more land and human sources than the Southern Hemisphere. The concentrations in the Southern Hemisphere are lower (roughly 60 ppbv) and more uniformly mixed, both horizontally and vertically. Because CO is emitted almost exclusively from surface sources, concentrations also tend to be lower in the free troposphere than in the boundary layer.

The concentration of CO varies with season. Because the largest CO sink is reaction with OH, as the concentration of OH increases in the summer, the concentration of CO tends to decrease. The residence time for CO in the troposphere is roughly one month in the tropics, and four months in the mid-latitudes. Pre-industrial concentrations of CO (prior to 1850) appear to be have been ~55–60 ppbv in the Southern Hemisphere and ~90 ppbv in the Northern Hemisphere (based on ice core data).

See also: Troposphere, Ozone Chemistry, Volume 1.

CYNTHIA S ATHERTON *USA*

Carbon-14

see Radionuclides, Cosmogenic (**Volume 1**)

CEOS (Committee on Earth Observation Satellites)

CEOS is an international organization charged with coordinating civil space-borne missions designed to observe and study our planet. Created in 1984, it now comprises 41 space agencies and other national and international organizations and is recognized as the main international forum for the coordination of Earth observation satellite programs and for the interaction of these programs with users of satellite data and information worldwide. Individual participating agencies use their best efforts to implement CEOS recommendations. The main goal of CEOS is to ensure that critical scientific questions relating to Earth observation and global change are covered and that satellite missions do not unnecessarily overlap each other. CEOS s three primary objectives are: (1) to optimize benefits of space-borne Earth observations through cooperation of its participants in mission planning and in development of compatible data products, formats, services, applications and policies; (2) to serve as a focal point for international coordination of space-related Earth observation activities; and (3) to exchange policy and technical information to encourage complementarity and compatibility of observation and data exchange systems.

CEOS is working closely with the Integrated Global Observing Strategy (IGOS) partnership in the development and implementation of the IGOS partnership. IGOS intends to unite the major satellite and ground-based systems for global environmental observations and monitoring of the atmosphere, oceans, land and life, in a framework that delivers maximum benefit and effectiveness in their final use. IGOS focuses on the observing dimension of the process of providing environmental information for decision-making IGOS, while linking research and operational programs as well as data producers and users.

CEOS meets in plenary annually, with activities and coordination occurring throughout the year. The Chair organization of CEOS rotates at the annual plenary. A permanent secretariat provides most of the coordination between plenary sessions and is maintained by the European Space Agency, the National Aeronautics and Space Administration, jointly with the National Oceanographic and Atmospheric Administration of the United States, and the Science and Technology Agency jointly with the National Space Development Agency (NASDA) of Japan, and is headed by the CEOS chair organization. CEOS has two permanent working groups which meet regularly. The objectives of the Working Group on Calibration and Validation are to enhance coordination and complementarity, to promote international cooperation and to focus activities in the calibration and validation of Earth observations. The Working Group on Information Systems and Services seeks to facilitate data and information management and services for users and data providers in dealing with global, regional and local issues.

CEOS issues a number of information products and services which are available from the CEOS homepage (www.ceos.org.) or from the secretariat (points of contact

also available on the website). CEOS also creates *ad hoc* working groups when needed to address specific issues such as the development of Earth observation education and training or the role of remote sensing in natural disaster monitoring and management.

See also: EOS (Earth Observing System), Volume 1.

LESLIE CHARLES *USA*

CFCs (Chlorofluorocarbons)

see Chlorofluorocarbons (CFCs) (**Volume 1**)

Chandler Wobble

Richard S Gross
Jet Propulsion Laboratory, Pasadena, CA, USA

The Earth rotates about its axis once a day, but does not do so uniformly. Instead, the rate of rotation uctuates by up to a millisecond per day, and the Earth wobbles as it rotates. Much like the wobble of an unbalanced automobile tire, the Earth wobbles because the mass of the Earth is not balanced about its rotation axis.

The wobbling motion of the Earth was first detected by Seth Carlo Chandler, Jr, in 1891 and has been under observation ever since. In these observations, the wobble manifests itself as an oscillation of the rotation pole about the North Pole of the Earth. This oscillation can be characterized by its *period*, which is the time that it takes the rotation pole to complete one circuit about the North Pole, and by its *amplitude*, which is the size of the offset of the rotation pole from the North Pole. Analyses of these observations reveal that the Earth has, in fact, two dominant wobbling motions:

1. an annual wobble with a period of 12 months and an amplitude of about 3 m; and
2. the Chandler Wobble with a period of 14 months and an amplitude that varies between about 3–6 m.

The annual wobble is a forced motion of the Earth that is caused largely by the annual appearance of an atmospheric high pressure system over Siberia every winter. This system annually loads the Siberian crust, causing the Earth to wobble with an annual period. The Chandler Wobble on the other hand is not a forced motion of the Earth, but is rather a resonant motion that was first predicted by the Swiss mathematician Leonhard Euler in 1765. Euler studied the general translational and rotational motion of rigid bodies and, by applying his theory to the Earth, predicted that if the Earth s mass were not balanced about its rotation axis, then the Earth should wobble as it rotates. Assuming that the solid Earth is rigid, the period of this Eulerian wobble is about 10 months.

Despite numerous subsequent attempts by astronomers to detect this predicted wobbling motion of the Earth, it was not until 1891 that Chandler finally detected such a motion, but with a period of 14 months. A year later, Simon Newcomb explained the lengthening of the period from 10 to 14 months by the fact that the solid Earth is not perfectly rigid and will therefore deform as it wobbles, and by the fact that the oceans will respond to the wobbling of the underlying solid Earth. Further analysis of the observations by Chandler revealed the additional presence of a wobble with a period of 12 months. The 14-month wobble is now known as the Chandler Wobble in honor of the person who first detected it.

Frictional forces associated with the wobble-induced deformation of the solid Earth would cause the Chandler Wobble to freely decay with an exponential time constant of about 68 years if no mechanism or mechanisms were acting to excite it. Observations of the Chandler Wobble taken during the past century show that there are times when the amplitude of the Chandler Wobble has actually increased. Thus, some mechanism or mechanisms must clearly be acting to excite the Chandler Wobble. Because it is known that the annual wobble is largely caused by changes in atmospheric pressure over land, and because the period of the Chandler Wobble is close to that of the annual wobble, many studies have been conducted to determine if atmospheric pressure fluctuations over land are also energetic enough to excite the Chandler Wobble. However, these studies have generally concluded that atmospheric pressure changes over land have only about a quarter of the power needed to excite the Chandler Wobble. Only recently, more than a century after its discovery, has it been shown that the combination of atmospheric and ocean-bottom pressure changes are likely to have enough power to excite the Chandler Wobble.

The recent availability of numerical general circulation models of the oceans has allowed the impact of oceanic processes on the Earth s rotation to be studied. In particular, such models have been used to show that the change in the load on the oceanic crust due to changes in the weight of the overlying column of water associated with changes in the distribution of the oceanic mass is about twice as effective in exciting the Chandler Wobble as is the changing atmospheric pressure over land. In fact, the sum of the

changing atmospheric pressure over land and the changing ocean-bottom pressure can fully explain the excitation of the Chandler Wobble.

Atmospheric pressure changes are the primary cause of the annual wobble, and the sum of atmospheric and ocean-bottom pressure changes are the primary excitation source of the Chandler Wobble. Thus, pressure changes associated with weather systems cause the Earth to wobble as it rotates. In addition, the dominant cause of changes in the rotation rate of the Earth, or, equivalently, changes in the length of the day (*see* Atmospheric Angular Momentum and Earth Rotation, Volume 1), are changes in the strength and direction of atmospheric winds. In fact, changes in the strength and direction of the jet stream associated with the El Niño/Southern Oscillation phenomenon have been shown to cause changes as large as a half a millisecond in the length of the day. Thus, among its many other consequences, climate change also affects the rotation of the Earth.

ACKNOWLEDGMENTS

The work described in this paper was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration (NASA). Support for this work was provided by NASA s Office of Earth Science.

FURTHER READING

Lambeck, K (1988) *Geophysical Geodesy*, Oxford University Press, Oxford.

Chaos and Predictability

Cecil E Leith

Lawrence Livermore National Laboratory, Livermore, CA, USA

Perhaps the oldest and most ubiquitous of prediction problems is that of the weather. It is, of course, well recognized that there is a practical limit of a few days to the range of useful weather forecasts. Less well known is the fact that there is a theoretical limit to the useful range of such forecasts, even if they are based on a perfect forecasting model. This is owing to the inherent growth of the initial errors that arise from the necessarily imperfect knowledge of the initial state of the atmosphere.

Such sensitive dependence on initial conditions characterizes dynamical systems that are called chaotic. Thus, although the underlying dynamics are purely deterministic, the evolving behavior appears to be increasingly random and unpredictable and leads to chaos.

HISTORICAL BACKGROUND

During the 19th century, much philosophical discussion dealt with the question of whether the world behaved in a way that was completely deterministic, or did not, and thus in some way allowed room for individual free will to be exercised. The Scottish mathematician and physicist J Clerk Maxwell entered the debate with an essay on the possible role of the physical sciences in considering this question. In his text *Matter and Motion*, Maxwell (1876) summarized his views in a statement of General Maxim of Physical Sciences:

> There is a maxim which is often quoted, that The same causes will always produce the same effects. There is another maxim that must not be confounded with that quoted . . ., which asserts that like causes produce like effects.

Maxwell gives the rather dramatic example:

> . . . in which a small initial variation may produce a very great change in the final state of the system, as when the displacements of the points causes a railway train to run into another instead of keeping its proper course.

To which he adds in a footnote:

> . . . In so far as the weather may be due to an unlimited assemblage of local instabilities, it may not be amenable to a finite scheme of law at all.

It is this sensitive dependence on initial conditions that characterizes a dynamical system that is called *chaotic*. In this context, chaos becomes a technical term for the behavior induced by a chaotic dynamical system. It imposes a limit on the predictability of such a system.

PREDICTABILITY OF THE WEATHER

Following the proposal made in about 1911 by Richardson (1922), numerical weather prediction was attempted by John von Neumann and colleagues at the Institute for Advanced Study in Princeton in the early 1950s. The first such effort was able to track a November storm for a few days across the United States with moderate success. However, in the course of time the forecast bore no resemblance to the actual evolving weather. The question naturally arose as to the possibility that even a perfect prediction model would be limited in the range of its skill, as suggested by Maxwell, by the growth of the inevitable

errors in the specification of the initial atmospheric state, i.e., weather may be chaotic.

Lorenz (1963) suggested that there were three ways to quantitatively determine the rate of this error growth. The most direct was to rerun a numerical forecast model with a small difference in its initial state and to note the rate at which this difference grew. This, of course, revealed the predictability of the model atmosphere but perhaps not that of the real atmosphere. The second (the *analogue method*) was to search the weather records for pairs of times at which the atmosphere was in nearly the same state and then to record the rate at which the ensuing states diverged. Although this approach dealt with the predictability of the real atmosphere, the records were unfortunately not long enough to provide any close pairs, and an estimate had to be made based on extrapolation to close pairs of results obtained from pairs that were not very close. Finally, one could turn to a somewhat dubious theory of the turbulent nature of the large-scale motions in the atmosphere in order to estimate the expected error growth rate.

Robinson (1967) used this turbulence approach with an assumption that the classical energy-cascading inertial range for three-dimensional isotropic turbulence proposed by Kolmogorov (1941) was relevant to atmospheric motions, and made the rather pessimistic estimate that prediction could not be useful beyond a day or so. However, the planetary scales of motion in the atmosphere have horizontal length scales of thousands of kilometers. These are much larger than the thickness of the atmosphere, which thus behaves more in accordance with turbulence confined to two dimensions rather than with the three-dimensional isotropic turbulence theory of Kolmogorov. This leads to a replacement of the classical $(-5/3)$ power-law energy spectrum of Kolmogorov (1941) with the (-3) power-law spectrum considered by Batchelor (1969) and Kraichnan (1967) and Leith (1971). The two-dimensional turbulence approach gives a result for the error growth rate, which is more consistent with the other two approaches, namely, that the doubling time of the root-mean-square error is about two days and thus that useful predictions may be possible for a week or more.

The question of the predictability of the weather was considered of such importance that it led to the establishment of an international program called the Global Atmospheric Research Programme (GARP), described by Robinson (1967) (*see* GARP (Global Atmospheric Research Program), Volume 1). The success of this program and its successor, the World Climate Research Programme, helped to establish the basis for current operational weather and climate prediction.

DYNAMICAL PHASE SPACE

A numerical model used for weather prediction utilizes of the order of a million numbers to specify the hydrodynamic and thermodynamic state of the atmosphere at a particular time. The model provides a computational process that is used to advance the state to a new state for a slightly later time, and this process continues to be repeated in order to determine the evolving state of the atmosphere for a forecast period of several days. Such a set of a million numbers can be considered as coordinates of a point in a million-dimensional space whose motion defines a curve or orbit in that space. In general, for such a dynamical system, this is called a *phase orbit* in the *phase space* of the system. Lorenz (1963) examined the properties of such a phase orbit in a phase space in which he had drastically reduced the dimension from one million to only three. He was then looking at only the grossest scales of motion in his model atmosphere. However, his three modes were coupled through non-linear interaction terms in his equations of motion that simulate such interactions in the real atmosphere. It was then fairly straightforward for him to make a plot of the evolving orbit. He found that these orbits were attracted to, and finally confined to only a part of the three-dimensional phase space rather than wandering around throughout the whole space.

These final sets have become known as *strange attractors* owing to their unusual topological properties. They appear to be two-dimensional sheets, but they are twisted and stacked in a strange way. They also appear to have two separate components such that an orbit will remain in each component for a period of time with an occasional transition to the other. Lorenz (1963) called such a bimodal system *almost intransitive* to distinguish it from a classical transitive or ergodic system in which the phase orbit fills a region of the phase space more or less uniformly.

An interesting possibility is that there exists a strange attractor for the atmospheric dynamical system. Such an attractor would lower the dimension of the dynamical phase space and might, thereby, save variables and computing time in numerical prediction. However, the dynamical equations for evolution confined to the attractor would be so complicated compared to those currently in use that any potential computational gain most probably would be lost.

Searches for bimodal behavior in the real atmosphere have not been particularly successful, although there is some suggestion that the Himalayan Massif may act as a fluidic switch for the flow of the jet stream about it. There is clearer evidence of such bimodal behavior in the currents of the North Atlantic Ocean, and this, of course, can have an indirect effect on the atmosphere and its predictability.

PREDICTABILITY OF TURBULENT FLOWS

Turbulent flows in general have limited predictability and thus are chaotic. As mentioned previously, there is a difference between the predictability of turbulence in two and in three dimensions.

In the classical analysis (Kolmogorov, 1941) of three-dimensional turbulence, one considers a statistically stationary situation in which a stirring source feeds kinetic energy at large scales into a fluid at a fixed rate ε. This energy is then transferred at the same rate through an inertial range of smaller scales until it is finally dissipated, still at the same rate ε, at scales small enough for viscous dissipation to be effective. Dimensional analysis is then based on the assumption that the energy spectrum function $\mathrm{E}(k)$ of the inertial range depends only on the energy cascade rate $\varepsilon[L^2T^{-3}]$ and the wave number $k[L^{-1}]$, defined as $k = 2\pi/\delta$, where δ is the scale wavelength. The total kinetic energy per unit mass is $E[L^2T^{-2}] = \int \mathrm{E}(k)\,\mathrm{d}k$; thus for dimensional consistency one must have $\mathrm{E}(k)[L^3T^{-2}] = C_{\mathrm{K}}\varepsilon^{2/3}k^{-5/3}$ where C_{K} is the dimensionless Kolmogorov constant found experimentally to be about equal to 2.

For two-dimensional turbulence, as noted, for example, by Batchelor (1969), Kraichnan (1967) and Leith (1971), there exist additional inviscid integrals of the motion based on the detailed conservation of vorticity $\omega[T^{-1}]$, i.e., the curl of the velocity field, which is always normal to the plane of motion and can, therefore, be treated as a scalar rather than as a vector quantity. In particular, the enstrophy $\Omega[T^{-2}]$, defined as half the square of the vorticity, is a conserved quantity that cascades from large scales to small ones at a constant rate $\eta[T^{-3}]$. Now the dimensional scaling arguments lead to a different energy spectrum function $E(k)[L^3T^{-2}] = C_{2D}\eta^{2/3}k^{-3}$, where $C_{2\mathrm{D}}$ is another dimensionless constant also about equal to 2.

The predictability of turbulent flow depends on how rapidly the errors in very small scales, thus at large values of k, will induce, through non-linear interactions, errors at somewhat smaller values of k, and lead to a back-cascade of error to ever smaller k and larger scales. At a particular wave number k, the non-linear interaction rate is found by dimensional analysis to be proportional to $[k^3\mathrm{E}]^{1/2}$. For a $(-5/3)$ power-law spectrum, this rate becomes extremely rapid for large values of k, whereas for a (-3) power-law spectrum the rate is independent of k, and the back-cascade of error is therefore much slower.

Although the dynamical behavior of the atmosphere is far more complex than that of two-dimensional turbulence, the analogy is close enough so that a (-3) power-law spectrum is, in fact, observed for the larger scales of motion of global atmospheric flow. The atmosphere is therefore far more predictable than it might have been had Robinson s original estimate been valid.

PREDICTABILITY OF THE CLIMATE

One might, at first, think that since the weather is of limited predictability, the climate is also. However, the climate consists of the probabilistic properties of the weather. As Lorenz has put it climate is what you expect; weather is what you get. Most numerical climate models at present are, however, really weather models that are run for long times from arbitrary initial states with an average taken over time in order to estimate climate properties.

The interesting climate prediction questions have to do with determining the sensitivity of the climate to changing influences.

This is done by making a long baseline run with a numerical model and then repeating the run with some specified perturbation in the influence of interest, such as the concentration of carbon dioxide in the model atmosphere. Usually these artificial changes are made rather large in order to obtain a significant response, which is then scaled down to the size of interest with the assumption of a linear dependence of the response on the perturbation.

Here it is important to consider the sampling fluctuations associated with any estimate of climate properties based on a finite time sample of either a numerical model or of the real atmosphere. It is known, for example, that a collection of N independent samples drawn from an infinite population has a sample mean which, as an estimate of the population mean, has a sampling error proportional to $1/N^{1/2}$.

To apply this result to a statistically stationary time series requires knowledge of the number N of effectively independent samples in a time interval of length T. This depends on the time interval τ between effectively independent samples which is given by $\tau = \int R(|s|)\,\mathrm{d}s$, where $R(|s|)$ is the correlation between values of the series at times separated by a time interval s. For typical atmospheric variables, this correlation time τ is about one week, and thus, for example, the effective sample size for a season is $N \sim 13$. Clearly then, much of the variability of seasonal average quantities from one year to the next is simply the consequence of sampling fluctuations. This variability is not unlike the sort of variability found in hands dealt in the card game of contract bridge.

In more general human endeavors, we must often deal with the relative importance of skill versus luck. A cloud of uncertainty always obscures predictive skill. It is clearly important to predict not only a quantity of interest but also to predict the uncertainty in that prediction, that is, to track the associated chaos. Gauss, as a teenager, gave an early example of this when he provided astronomers with a useful prediction of when and in what region of the sky the asteroid Ceres would be found again following its initial discovery in 1801 and its later obscuration by sunlight. Although Gauss was, at the time, already well known in the pure mathematics community, this prediction made him so famous in the larger scientific community that his mathematical colleagues accused him of selling out to popular applied science.

A straightforward way of predicting uncertainty along with a forecast is to carry out, in fact, a collection or ensemble of forecasts starting from a cluster of initial states that reflect the initial uncertainties in observations and analysis (Leith, 1974). The mean and standard deviation of the ensemble provide estimates of the forecast and of its uncertainty. Such ensemble forecasts have become standard in operational weather prediction. Although an ensemble average forecast may technically minimize the mean square error, it suffers from producing a weather map that looks unreasonably smooth compared to those observed.

SUMMARY AND OUTLOOK

Chaotic dynamical systems are those for which small disturbances lead to large consequences. The prototypical example is that of the weather system for which errors in the determination of the initial state of the atmosphere have an rms doubling time of about two days. The theoretical limit of predictive skill with a perfect forecasting model is, thus, about 10 days. In fact, the practical range of forecast skill is only about five days owing to model imperfections. Of course, the major imperfections in early forecast models have already been recognized and removed. The remaining imperfections are, thus, individually small although large in number. It will therefore be a slow and tedious process to identify and remove them.

It seems likely that such a multitude of small imperfections plague attempts to build models to predict the evolution of other more complex chaotic dynamical systems, such as the economy, and that the weather system provides a relatively well understood example of a general problem.

See also: Natural Climate Variability, Volume 1; Chaos and Cycles, Volume 2; Non-linear Systems, Volume 2; Monitoring in Support of Policy: an Adaptive Ecosystem Approach, Volume 4.

REFERENCES

Batchelor, G K (1969) Computation of the Energy Spectrum in Homogeneous Two-dimensional Turbulence, *Phys. Fluids*, 12(Suppl II), 233–239.

Kolmogorov, A N (1941) The Local Structure of Turbulence in Incompressible Viscous Fluid for Very Large Reynolds Number, *Dokl. Akad. Nauk. SSSR*, 30, 301–305.

Kraichnan, R H (1967) Inertial Ranges in Two-dimensional Turbulence, *Phys. Fluids*, 10, 1417–1423.

Leith, C E (1971) Atmospheric Predictability and Two-dimensional Turbulence, *J. Atmos. Sci.*, 28, 145–161.

Leith, C E (1974) Theoretical Skill of Monte Carlo Forecasts, *Mon. Weather Rev.*, 102, 409–418.

Lorenz, E N (1963) Deterministic Nonperiodic Flow, *J. Atmos. Sci.*, 20, 130–141.

Maxwell, J C (1876) *Matter and Motion*, Society for Promoting Christian Knowledge, London, reprinted in 1952 by Dover, New York.

Richardson, L F (1922) *Weather Prediction by Numerical Process*, Cambridge University Press, London, reprinted in 1965 by Dover, New York.

Robinson, G D (1967) Some Current Projects for Global Meteorological Observation and Experiment, *Q. J. R. Meteorol. Soc.*, 93, 409–418.

Charney, Jule Gregory

(1917- 1981)

Jule Charney, one of the greatest meteorologists of the 20th century, made original and fundamental contributions to the fields of meteorology and oceanography.

Charney was born January 1st, 1917 in San Francisco, CA. He received his undergraduate education in mathematics and physics at the University of California, Los Angeles. While pursuing further graduate studies, he became interested in fluid dynamics and meteorology and took part, initially as a student and then teacher, in the university s emerging meteorology program.

As World War II ended, Charney, with little or no guidance or supervision from the faculty, began work on the mathematical theory of instability in atmospheric flow. The 1947 publication of the resulting PhD thesis established him as a world leader in dynamic meteorology. Displaying deep physical insight, extraordinary brilliance, and creative genius, he provided a theoretical explanation for the genesis, intensification and propagation of large-scale waves in the atmosphere – the theory of atmospheric baroclinic instability. He showed that the observed large-scale fluctuations in the lowest 10 km of the atmosphere are manifestations of hydrodynamic instability of the extra-tropical middle-latitude westerly winds, the speeds of which increase with height between the warmer tropical and colder polar regions of the rotating planet.

The next few years took Charney to Chicago, Bergen (Norway), and then Princeton, while his interests turned

to the forecasting problem. He applied a systematic scaling technique to develop a simplified set of equations of motion, referred to as the quasi-geostrophic equations, that can be used for numerical weather prediction. In collaboration with the noted mathematician, John von Neumann, and others, he pioneered the application of electronic computers to weather forecasting, leading to the first successful numerical predictions in 1952. Charney s deep physical and theoretical insight transformed weather prediction from guesswork to science.

In the following years, Charney made numerous other significant scientific contributions on diverse problems, including: the growth of hurricanes; the Inter-Tropical Convergence Zone (ITCZ); the assimilation of satellite data; the dynamics of the Gulf Stream; the equatorial undercurrent in the ocean; ocean boundary currents; geostrophic turbulence; the theory of desertification; the theory of blocking and multiple flow equilibria; and the predictability of monsoons. In 1956, he joined the faculty of the Massachusetts Institute of Technology (MIT), which remained his primary affiliation for the remainder of his career.

In 1957, Charney was appointed to the US National Academy of Sciences Commission on Meteorology, where he called attention to the potential of global satellites, new instrumentation for detailed observation, and electronic computers – factors that led him to promote formation of the National Center for Atmospheric Research and eventually the Global Atmospheric Research Programme. The Global Weather Experiment of 1979, the first major international global weather measurement field program, is considered his brainchild.

Elected to Membership in the National Academy of Sciences in 1964, Charney continued to participate in many of its studies. Notably, he led a brief study in the summer of 1979 that produced an influential estimate of the influence of the increasing atmospheric carbon dioxide concentration on climate. That estimate of the range in global climate sensitivity, namely a warming of 1.5−4.5 °C for a doubling of the carbon dioxide (CO_2) concentration, has been since then the range used in developing estimates for future climate change.

Charney was an active participant in the anti-war efforts of the US academic community during the Vietnam War. After the US invasion of Cambodia, he was very active in raising funds for the anti-war candidates for political offices in the 1970 elections.

Charney died June 16th, 1981 in Boston, MA. Throughout his academic career, Charney mentored a stream of talented students recruited from varied countries and disciplines. His theoretical insights, the predictive capabilities we now enjoy, and perhaps above all the productive scientists whom he nurtured, remain his enduring legacy.

Photo: reproduced by permission of NASA.

JAGADISH SHUKLA *USA*

Chlorofluorocarbons (CFCs)

Mack McFarland
DuPont Fluoroproducts, Wilmington, DE, USA

Chloro uorocarbons (CFCs) are compounds containing only chlorine, uorine and carbon. The normal state of these compounds at standard temperature and pressure ranges from gases to liquids to waxes to solids. CFCs with suf cient vapor pressure to escape to and remain in the atmosphere contribute to both ozone-depletion and global climate change. These compounds are not destroyed in the troposphere and have very low water solubility. They are destroyed primarily by photolysis in the stratosphere where the chlorine is released. The chlorine acts as a catalyst for ozone destruction. Because transport to the stratosphere is a slow process, atmospheric lifetimes of CFCs are 45 years or longer. CFCs have signi cant infrared absorption cross-sections and, thus, contribute to global climate change with global warming potentials (GWPs) of 4600 or larger. CFCs are controlled under the Montreal Protocol for the Protection of the Ozone Layer and, except for minor exempted uses, their production for use has been phased out in developed countries and is to be phased out in developing countries by 2010. The most commonly used CFCs were CCl_3F (CFC-11), CCl_2F_2 (CFC-12), $CClF_3$ (CFC-13), CCl_2FCClF_2 (CFC-113), $CClF_2CClF_2$ (CFC-114) and $CClF_2CF_3$ (CFC-115).

CFCs were developed in 1929 as efficient alternative refrigerants to replace toxic and flammable fluids, including ammonia and sulfur dioxide, in use at that time. Commercial production began in 1931. CFCs are nonflammable, low in toxicity, thermally and chemically stable and relatively low in cost. These properties led to development of a wide range of other applications. After about 1950 they began to be used as aerosol propellants, cleaning agents and blowing agents for plastic insulating and cushioning foams. For example, through the early 1990s, CFC-12 was the primary refrigerant in home refrigerators/freezers. The insulating foam in the walls of those appliances was produced with CFC-11. A combination of properties including low vapor phase thermal conductivity contributed to efficient insulation in the small space available. CFC-11 as well as other CFCs were also used in many other thermal insulating plastic foams for the construction industry. Up until about 1995 CFC-12 was used as the refrigerant in automobile air conditioners. The CFC properties listed above in addition to desirable solvency and surface tension properties contributed to the widespread use of CFC-113 as a cleaning agent in the electronics industry, as well as specialized metal cleaning

applications. Both CFC-12 and CFC-11, and other CFCs to a lesser degree, were used as propellants in a wide variety of consumer products including hair sprays and perfumes. CFCs were chosen for these applications, even though they were more expensive than hydrocarbon propellants, because they are non-flammable and because of their chemical stability, appropriate vapor pressures at room temperature and miscibility with the various products. Consumption of CFCs expanded rapidly during the 1960s and early 1970s reaching almost a billion kg year^{-1} in 1974 with almost 70% of that use as an aerosol propellant (Figure 1).

As a result of the publication of the ozone depletion hypothesis by Molina and Rowland (1974) use of the compounds as aerosol propellants declined rapidly, especially in North America and Scandinavia where bans on the use for this application were eventually enacted. However, use of the compounds in other applications continued to grow. Continuing advances in the understanding of the potential impact of CFCs and other chlorine and bromine containing compounds on the ozone layer led to negotiation and adoption of the Vienna Convention for the Protection of the Ozone Layer in 1985. This agreement called for international cooperation to protect the ozone layer and provided a framework for agreements to control use of human produced ozone depleting substances such as CFCs. Further advances in the understanding of the science of ozone depletion provided the basis for the Montreal Protocol on Substances that Deplete the Ozone Layer (The Protocol), an instrument under the Vienna Convention, in 1987 (*see* Ozone Layer: Vienna Convention and the Montreal Protocol, Volume 4). This agreement initially required a 50% reduction in consumption of CFCs in developed countries over a ten-year period beginning in 1989 with similar controls for developing countries delayed by ten years. The Protocol also contained provisions for periodic review and modification of controls based on advances in understanding of the scientific, technical and economic issues. By early 1988, focused and extensive scientific studies initiated after the discovery of the seasonal declines in stratospheric ozone over Antarctica, the Antarctic ozone hole, in 1985 had established a link between CFCs and other compounds and ozone depletion. This link and other continuing advances in scientific understanding as well as the development of technical options for eliminating the use of CFCs and other ozone-depleting substances led to a series of increasingly stringent controls under the Protocol. Every major ozone-depleting substance is now controlled under the Protocol. Except for exempted uses, production of CFCs for use in developed countries ceased at the end of 1995. Production of CFCs for use in developing countries will cease by 2010.

These regulations have led to the development and implementation of a wide range of alternative fluids and technologies to replace CFCs. Hydrofluorocarbons (HFCs) and hydrochlorofluorocarbons (HCFCs) replaced CFCs in most refrigerant applications with hydrocarbons and ammonia playing a relatively minor role. Hydrocarbons and alternative dispensing technologies such as pump sprays were the primary alternatives for CFCs used as aerosol propellants.

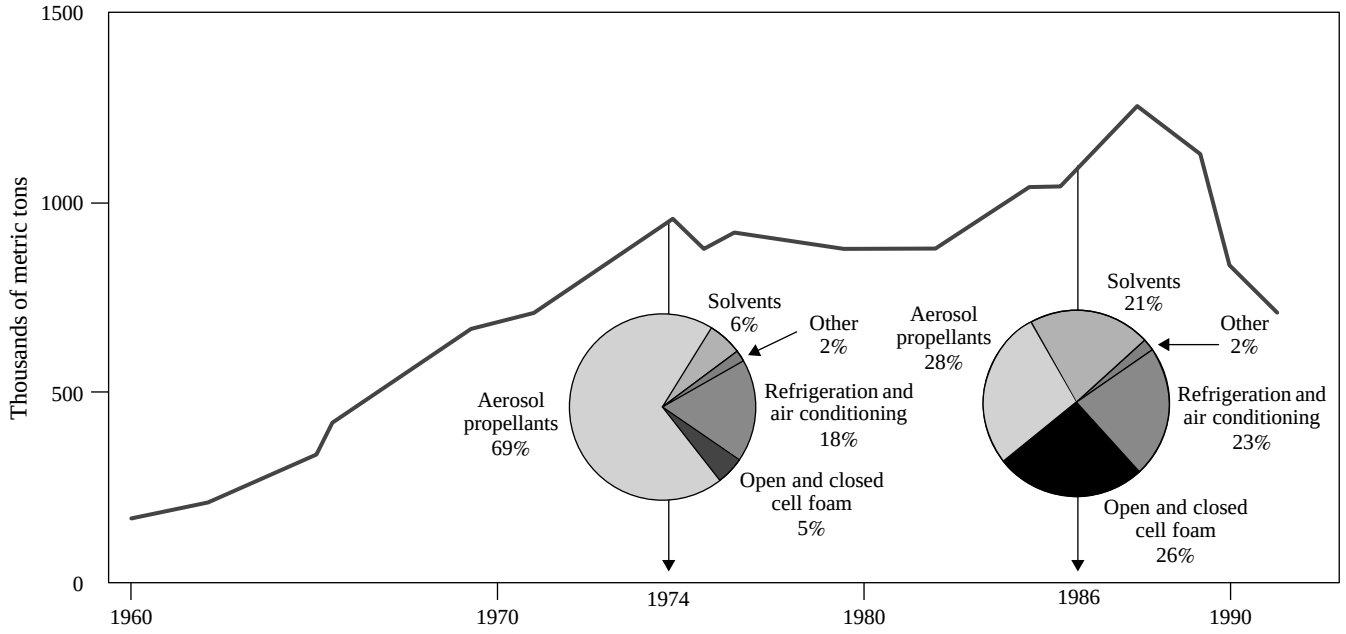


Figure 1 Estimated global CFC consumption. (Reprinted from McFarland (2000) with kind permission from Kluwer Academic Publishers)

HCFCs, water and carbon dioxide were the primary replacements for insulating foams. A wide range of technologies including water, hydrocarbons and changes in production techniques such that no cleaning was required displaced CFC use as cleaning agents.

Production of CFCs peaked at about 1.2 billion kg year^{-1} in 1988 before beginning to decline as a result of the internationally agreed phase-out. The cumulative atmospheric concentration of these compounds peaked at a level below 1 ppb around the year 2000.

See also: Stratosphere, Chemistry, Volume 1; Depletion of Stratospheric Ozone, Volume 1; Ozone Depletion Potential (ODP), Volume 1; Ozone Hole, Volume 1; Stratosphere, Ozone Trends, Volume 1.

REFERENCES

McFarland, M (2000) Applications and Emissions of Fluorocarbon Gases: Past, Present, and Prospects for the Future, in *Non-CO_2 Greenhouse Gases: Scienti c Understanding, Control and Implementation*, eds J van Ham, A P M Baede, L A Meyer, and R Ybema, Kluwer, Dordrecht, 65–82.

Molina, M J and Rowland, F S (1974) Stratospheric sink for chlorofluoromethanes: chlorine atom-catalyzed destruction of ozone, *Nature*, 249, 810.

Circulation, Atmospheric

see Atmospheric Motions (**Volume 1**)

Clathrates

see Methane Clathrates (**Volume 1**)

CLIC (Climate and Cryosphere)

CLIC is a coordinated study of the role of all components of the cryosphere in the global climate system. The principal scientific questions being considered include:

- change in global sea level;
- energy and water cycles in regions with sea ice;
- land-ice and frozen ground;
- the role of polar fresh water in the ocean thermohaline circulation;
- the interactions of snow cover, sea ice and seasonally frozen ground with other climate processes;
- albedo feedback.

A CLIC Science and Co-ordination Plan has now been prepared, with systematic studies in the following areas:

- cryosphere–atmosphere interactions on a global scale;
- cryosphere–ocean interactions on a global scale;
- atmosphere–sea-ice–ocean interactions;
- atmosphere–land-ice–sea-level;
- cryospheric indicators of climate change.

Under CLIC, the World Climate Research Programme Arctic Climate System Study (ACSYS) will also be completed (*see* ACSYS (Arctic Climate System Study), Volume 1). Since 1993, ACSYS has been assembling a series of comprehensive Arctic climate data sets, including Arctic Ocean hydrography, sea-ice characteristics, the Arctic atmosphere and hydrological cycle. Efforts are also being made to improve representation in models of the physical processes dominant in Arctic climate. Another primary objective is the estimation of the Arctic freshwater budget in order to assess the role of the Arctic in driving the ocean thermohaline circulation. An extensive field program is involved with a range of Arctic Ocean hydrographic and shelf surveys, and an extensive historical Arctic Ocean climate data base has been compiled.

Further information may be obtained from: International ACSYS/CLIC Project Office, The Polar Environmental Center, N-9296 Troms , Norway, Fax: +47-77-75-05-01, E-mail: tordis.villinger@npolar.nom http://www.npolar.no./acsys.

JOHN S PERRY *USA*

CLIMAP (Climate: Long-range Investigation, Mapping, and Prediction)

The CLIMAP project was an international research project funded by the United States National Science Foundation s International Decade of Ocean Exploration program in 1971. The original goal of the project was to reconstruct the history of the North Pacific and North Atlantic during the last 700 000 years (the Brunhes Normal magnetic epoch). This research resulted in developing the link

between changes in orbital insolation and the glacial cycles by examining the oxygen isotope ratios from ocean cores and proving that changes in oxygen isotope ratios mostly record ice volume rather than ocean temperature. In 1973, the objective expanded to map the surface of the Earth during the Last Glacial Maximum (LGM) (*see* Last Glacial Maximum, Volume 1), and to measure oscillations of the Pleistocene climate. One of the better-known products of CLIMAP was the mapping of LGM sea surface temperatures (SSTs) from planktonic assemblages in ocean cores. The CLIMAP (1981) results showed that SSTs were only a couple of degrees colder than present, which has recently been questioned by newer data showing even colder SSTs. In the final stages, the group also tried to map the Eemian (*see* Eemian, Volume 1) world. CLIMAP was originally composed of researchers from Lamont-Doherty Geological Observatory of Columbia University, Brown University, and Oregon State University, and later included the University of Maine and Princeton University. Eventually nearly 100 researchers from throughout the world were involved (Imbrie and Imbrie, 1986).

REFERENCES

CLIMAP (1981) *Seasonal Reconstructions of the Earth's Surface at the Last Glacial Maximum*, Geological Society of America Map Chart Series, MC-36.

Imbrie, J and Imbrie, K P (1986) *Ice Ages: Solving the Mystery*, Harvard University Press, Cambridge, MA, 1–224.

BENJAMIN S FELZER *USA*

Climate

In popular usage, *climate* is the term used to describe the average of the weather during a long period of time at a particular location or over a particular region. In traditional climatological studies for the period when instrumental records are available, averages are usually taken over a designated 30-year period, although sometimes the lack of observational data requires the use of other time intervals. For pre-instrumental periods, efforts are made to reconstruct at least some of the variables defining the climate by use of proxy variables that are responsive to climate, such as tree ring width, pollen distribution, snow amount or other variables, always attempting to consider the representativeness of the measure. In more technical usage, and particularly in connection with modeling, climate is the set of statistics characterizing the state of the atmospheric, oceanic, cryospheric and land surface components of the physical climate system over any period of time longer than about one month. This definition thus includes a description of the variability, the seasonal cycle, and other statistical measures of the system, but is often restricted to the mean values characterizing the atmosphere.

W LAWRENCE GATES *USA*

Climate Agenda

Michael Coughlan
World Meteorological Organization, Geneva, Switzerland

The Climate Agenda was jointly developed by a number of international organizations carrying out signi cant climate-related activities in response to the requirements stated in Agenda 21. It is a comprehensive integrating framework for cooperation in climate data collection and application, climate system research and studies of socio-economic impacts of climate variability and their effects on ecosystems.

The *Climate Agenda* is a comprehensive integrating framework for international cooperation on climate-related programs, especially those relating to data collection and application, climate system research and studies of socio-economic impacts of climate variability and their effects on ecosystems.

The Climate Agenda was jointly developed by a number of organizations carrying out significant climate-related activities in response to the requirements stated in Agenda 21 and the United Nations (UN) Framework Convention on Climate Change (*see* United Nations Framework Convention on Climate Change and Kyoto Protocol, Volume 4). The international organizations are: Food and Agriculture Organization, the International Council for Science, UN Environment Programme, UN Educational, Scientific and Cultural Organization and its Intergovernmental Oceanographic Commission, World Health Organization and World Meteorological Organization (WMO).

The Climate Agenda is designed to enable governments, and international governmental and non-governmental organizations to coordinate their contributions to national and international climate programs. This is to be done in a way that allows the participating bodies to benefit from complementary activities of the global effort. International climate-related activities under the Climate Agenda concentrate on the following areas, called *thrusts*:

- new frontiers for climate science and prediction;
- climate services for sustainable development;
- studies of climate impact assessments and response strategies to reduce vulnerability;
- dedicated observation of the climate system.

The principal international programs comprising the Climate Agenda include the World Climate Programme (WCP), the International Geosphere–Biosphere Programme, the International Human Dimensions Programme for Climate Change and the Global Climate Observing System. The periodic assessments carried out by the Intergovernmental Panel on Climate Change are also considered major contributions to the Climate Agenda. Further information can be obtained from the WCP Department of the WMO in Geneva.

Climate Analogues

Climate analogues are past climates that may be similar to possible future climates. Typically the term refers to past climates analogous to a 21st century climate influenced by global warming due to human-produced greenhouse gases such as carbon dioxide. The Russian climatologist M I Budyko argued that a study of climate analogues would lead to more reliable predictions of global warming than can be made with climate models. To develop the analogues, he drew upon paleoclimatic results from warm periods from 6000 to 100 million years ago to characterize the geographic patterns of warm climates. Among the periods he focused on were times characterized by high atmospheric concentrations of carbon dioxide (e.g., the mid-Cretaceous era) or when warming had been induced by changes in the orbital characteristics of the Earth about the Sun. More recently, it has been recognized that the rapid pace of human-induced climate change limits the applicability of analogs from past climates and thus makes a detailed agreement between past and future climatic conditions unlikely (Crowley, 1990). However, there may be some validity in deducing the long-term global-mean climate sensitivity to greenhouse gases and other forcing factors using analogs (Hansen *et al.*, 1993; Hoffert and Covey, 1992). When this is done, the results are in reasonable agreement with the range of estimates developed using climate models.

REFERENCES

Crowley, T J (1990) Are there any Satisfactory Geologic Analogs for a Future Greenhouse Warming? *J. Clim.*, 3, 1282–1292.

Hansen, J, Lacis, A, Ruedy, R, Sato, M, and Wilson, W (1993) How Sensitive is the World s Climate? *Natl. Geogr. Res. Explor.*, 9, 142–158.

Hoffert, M I and Covey, C (1992) Deriving Global Climate Sensitivity from Paleoclimate Reconstructions, *Nature*, 360, 573–576.

CURT COVEY *USA*

Climate, Arctic

see Arctic Climate (**Volume 1**)

Climate, Asteroid and Comet Effects

see Asteroids and Comets, Effects on Earth (**Volume 1**)

Climate Change

Traditionally, *climate change* is defined as a local, regional or large-scale change in the long-term average regime of temperature, precipitation, circulation and related weather variables characterizing the climate. For the period since 1860 with instrumental records, the definition is usually applied over a designated 30-year period. For the pre-instrumental record, changes in climate can usually only be resolved on multi-decadal to millennial time scales. More recently, and particularly in the context of modeling, climate change is the change in the statistics characterizing the physical state of the atmosphere, ocean, cryosphere and land surface over a specified region over a specified period of time. This definition thus includes changes of the mean values, variability and other statistical properties of the physical variables describing the climate, and one can speak, for example, of a change in the January climate at a specific site, a change in the decadal climate over Europe, or the change in the global climate over a century.

W LAWRENCE GATES *USA*

Climate Change, Abrupt

Frank Oldfield and Keith Alverson
PAGES (Past Global Changes) International Project Office, Bern, Switzerland

Some of the most exciting new insights into the nature of global change have arisen from the growing body of evidence for rapid changes in the climate system. Recent evidence from ice cores and elsewhere shows that global climate, together with the other aspects of the Earth System to which it is intimately linked, has undergone repeated oscillations in the geologically recent past. Moreover, the emerging evidence indicates that the most dramatic of these switches in climate occurred within a matter of decades and was at least hemispheric in extent, with major consequences for temperatures, hydrological regimes, plants and animals. The new evidence is also providing insights into the climatic backdrop for the development of human societies in the past. Abrupt change also raises major challenges for projecting future climate change. If the potentially disastrous impacts of climatic surprises like the ones that have already occurred are to be avoided in the future, research devoted to deepening our understanding of abrupt changes in the past is of paramount importance.

INTRODUCTION

Any definition of abrupt climate changes is necessarily somewhat subjective, because it depends in large measure on the sample interval used in a particular study and on the pattern of longer-term variation within which the sudden shift is embedded. Here, we avoid any attempt at a general definition but focus attention on four types of rapid transition found in the paleo-record from the geologically recent past:

- On the time-scale over which Milankovitch forcing is recognizable, the evidence from marine, polar ice core and terrestrial records spanning the last 800 kyr, all highlight the sudden and dramatic nature of glacial terminations, the shifts in global climate that occurred as the world passed from dominantly glacial to interglacial conditions.
- Within the glacial periods themselves, and especially the most recent, spanning the period from around 70–11.6 kyr ago, there are very large climate oscillations recorded in archives from the polar ice caps, high to middle latitude marine sediments, lake sediments and loess sections. These oscillations occur on millennial time-scales and are usually referred to as Dansgaard–Oescheger (D/O) cycles (*see* Dansgaard- Oescheger Cycles, Volume 1).
- During the first half of the Holocene period, from 11.6 kyr to around 6 kyr, evidence from lower latitudes especially points to rapid shifts in climate during the period when ice volume, sea level and vegetation were changing in the wake of the last glacial termination and the latitudinal and seasonal distribution of incoming solar radiations differed significantly from that experienced today (*see* Holocene, Volume 1).
- Over the last 6 kyr, the patterns of solar and volcanic forcing are thought not to have differed greatly from the conditions prevailing over the last few centuries. Evidence for rapid changes within this period is therefore of critical interest as context for projecting future climate change.

GLACIAL TERMINATIONS

Figure 1 shows the record of paleo-temperature and atmospheric trace gas concentrations over the last 450 thousand years as recorded in the Vostok ice core from Antarctica (Petit *et al.*, 1999). That glacial terminations are much more rapid than glacial onsets is immediately evident. Trace gas concentrations closely track, but do not lead, the evidence for warming. These data suggest that under natural conditions, greenhouse gases have played a crucial role at the end of each glacial cycle, probably through feedback effects rather than as initiators of change.

The glacial termination for which we have most information is the last one, beginning some 16 kyr ago and ending around 11.6 kyr BP with the beginning of the Holocene (Figure 2). Two features tend to qualify any simple view of the termination as a single, rapid, globally coherent, monotonic process. When the inferred temperature changes in Antarctica and Greenland are synchronized on the basis of their records of changing atmospheric methane concentrations, it is clear that during the course of the warming trend from Glacial to Holocene, the record from Antarctica leads that from Greenland. Moreover, during the period of warming, the sharp fluctuations in temperature that mark the Greenland record in particular, are largely in anti-phase with the changes in much of Antarctica. Recent high resolution U–Th dating reinforces the idea of an Antarctic lead, since it suggests that the penultimate glacial termination occurred 135 ± 2.5 kyr ago (Henderson and Slowey, 2000). This date precedes the Northern Hemisphere insolation peak by nearly 10 kyr; thus, although consistent with primary deglaciation forcing by insolation levels in the Southern Hemisphere, or the tropics, this new evidence runs counter to the view that Milankovich forcing at high northern latitudes was the initial trigger for deglaciation. Any concept of a rapid shift associated with glacial terminations thus has to be qualified by the realization that the rate

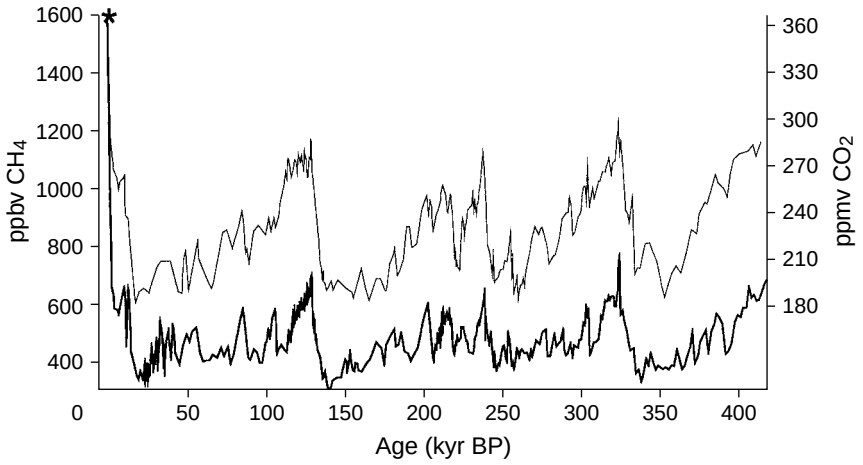


Figure 1 The sequence of changes in the concentration of atmospheric greenhouse gases methane (bold line) and carbon dioxide (narrow line) as recorded in trapped air bubbles in the Vostok ice core over the past four glacial cycles. In this figure, the ice core records have been extrapolated to the modern concentrations of both gases, 365 ppmv and 1600 ppbv for CO_2 and CH_4, respectively, as measured in the atmosphere in Antarctica (indicated by the asterisk). This extrapolation clearly shows that the modern levels are unprecedented over the entire record. (Reproduced with permission from Raynaud *et al.*, 2000)

Figure 2 The termination of the last glacial period was an abrupt climatic change. The associated Younger Dryas cold event some 11.6 kyr ago is, in this record from Central Greenland, manifested as a warming of around 10 °C, accompanied by a doubling in annual precipitation volume, occurring in about a decade. Shown here are accumulation (thick line) and oxygen isotope records (shown as inferred temperature (thin line)) from the GISP2 ice core. (Reproduced with permission from Alley *et al.*, 1993)

of change is not uniform over the whole globe and that the oscillations that are superimposed on the trend are not globally parallel in sign. The rapid oscillations that are superimposed on the warming trend and which tend to emphasize the dramatic nature of the termination are the last of the D/O cycles that are considered in the next section.

MILLENNIAL SCALE CHANGES: THE D/O CYCLES

One of the most unexpected outcomes of the detailed study of Greenland ice cores that began in the 1960s was the discovery that between 70 kyr and 11 kyr ago, all the climate indicators in the cores consistently indicated rapid,

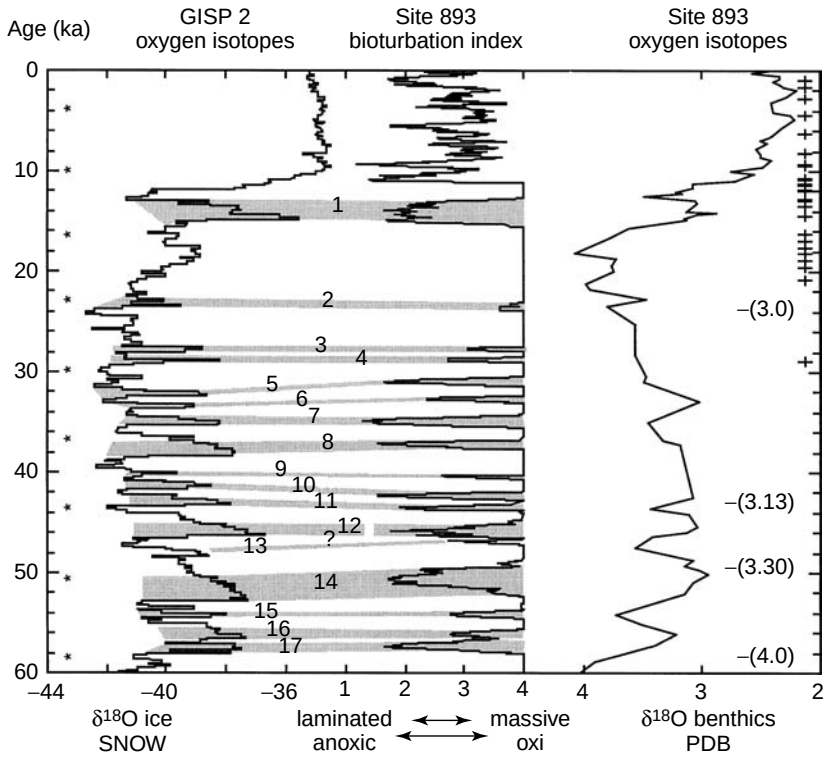


Figure 3 Millennial scale climate variability recorded in the oxygen isotope record from Greenland ice cores (GISP 2) and marine sediments in the Santa Barbara basin (Site 893). The bands linking the two records show the correlations proposed by the authors. The numbers represent the sequence of D/O oscillations identified in the Greenland record. The ice core oxygen isotope record is dominated by past changes in air temperature. The bioturbation index in the marine core reflects the oxygen status of the water column. The apparent coherence of the two widely separated records of environmental change points to climate linkages at least hemispheric in extent. (Reproduced with permission from Behl and Kennet, 1996)

high amplitude switches between full glacial and relatively mild interstadial conditions. Many of these switches appear to have been well over half the amplitude of the temperature change from the glacial maximum to peak Holocene warmth. Corroborating evidence for these oscillations has now come from both marine and continental sediment records (Figure 3). In the North Atlantic, each of the most severe of the cold shifts was associated with evidence for a massive southerly extension of the limit of iceberg extent, confirmed by the presence of ice-rafted detritus (IRD) in sediment cores. Moreover, the events themselves seem to be grouped in cycles (Bond Cycles) beginning with the most severe and gradually declining in amplitude until the next multiple cycle is initiated.

By combining the stratigraphic record of these rapid D/O cycles oscillations with attempts to model the mechanisms responsible, the dominant hypothesis to emerge sees the events as a product of the instability of the ice sheets that girdled the North Atlantic. Rapid discharge of ice into the ocean is thought to have reduced salinity to the point where North Atlantic Deep Water (NADW) formation was inhibited. This cut-off of NADW formation may then have so modified ocean circulation as to have influenced climate on a global scale. Such a hypothesis is consistent with the anti-phase relationship between Greenland and much of Antarctica often referred to as a *bipolar see-saw*.

It seems possible that the variations in amplitude of the cold swings reflect changes between iceberg surges reflecting instability of the Fennoscandian ice sheet, with its relatively more rapid response time, and the Laurentide ice sheet with its longer response time. Surges of the latter are thought to have been responsible for the most extreme events, those that gave rise to the main Heinrich IRD layers. The periodicity of the D/O events is not regular and no credible external forcing mechanism has been proposed. The bipolar anti-phase relationship demonstrated for the major D/O events suggests that they are largely the product of cryosphere–ocean dynamics.

The last of these events, often referred to as the *Younger Dryas*, is by far the best documented (*see* Younger Dryas, Volume 1). It marks the final stage of the last glacial termination, the transition to the Holocene already referred to above (Figure 2). Some sense of the suddenness of this transition in central Greenland may be obtained from

the record of ice accumulation and temperature (Alley *et al.*, 1993). Most of the transition, a near doubling of snow accumulation rate and a temperature shift of around 10 °C, was accomplished within a decade or less. By any standards, this is a truly dramatic, abrupt shift in climate, yet the generally accepted external forcing, an increase in solar irradiance, was relatively modest and certainly gradual. It appears that external forcing pushed the Earth system over some crucial threshold beyond which internal dynamics could generate extremely rapid responses through a range of, as yet poorly understood or quantified, feedback mechanisms.

Over much of the Northern Hemisphere, at least, the period of deglaciation between 16 and 11.6 kyrs ago is marked by a series of rapid shifts in climate, with minor, short-term oscillations superimposed on the major swings from full glacial to the Bolling–Allerod interstadial, the onset of the Younger Dryas and the final termination at the opening of the Holocene. These sharp, lower amplitude changes span decades to centuries and there is increasing evidence that they are spatially coherent and synchronous over large areas, including not only Central Greenland, Eastern Canada and Western Europe, but low latitude sites like the Cariaco Basin off the coast of Venezuela where laminated sediments provide an annually resolved paleoproductivity signal linked to climate changes.

THE FIRST HALF OF THE HOLOCENE

Although the Glacial/Holocene boundary is, over much of the world, a clearly recognizable and apparently synchronous stratigraphic feature, this does not imply that the Earth system as a whole experienced an instantaneous and complete shift. Several of the responses to the rapid changes taking place at the boundary took centuries, even millennia to complete. These include processes like the melting of the continental ice sheets, the recovery of global sea level to something approaching its present height, the recolonization of extensive areas by vegetation adapted to changed thermal and hydrological regimes, and the maturation of soils that goes hand in hand with increasingly stable vegetation cover where this had been absent during glacial times. Hence, it is important to remember that the period was one of transition and readjustment.

This said, it is equally important to dispel the view that the Holocene as a whole was a period of relatively constant climate. This proposition, arising from the stable isotope record in Central Greenland, is misleading. Not only is there now clear evidence of climate variability during the Holocene in Greenland itself, but evidence for Holocene climate variability at lower latitudes is also very strong.

Evidence is accumulating from several quite widely spaced sites for a sudden climate anomaly lasting for several hundred years centered around 8.2 kyr BP. This event shows up as a sharp decline in inferred temperature and in methane concentrations in the Greenland Icecore Project (GRIP)/Greenland Ice Sheet Project Two (GISP 2) ice core records from Central Greenland. An almost perfectly parallel change in $\delta^{18}O$ derived temperature can be seen in the evidence from Ammersee, in southwest Germany, based on stable isotope analyses of ostracod remains (Von Grafenstein *et al.*, 1998). Growing numbers of records appear to indicate that this oscillation had a widespread effect. In terms of timing and signature in the Greenland ice cores records, it appears to be in many ways comparable to the preceding, higher amplitude D/O oscillations and some link to ice melt and NADW formation may be involved, but there is also the possibility that the event, especially the methane anomaly, is linked to lower latitude changes in the extent of tropical and subtropical wetlands or the effects of rising sea level. Certainly, the whole of the early Holocene is marked by dramatic shifts in lake level and wetland extent in Africa and Central America.

The records of changing lake levels at a range of sites in central Africa constitute dramatic evidence of the major changes in hydrology that occurred during the early Holocene (Gasse, 2000) (Figure 4a). Equally powerful indications of variability come from reconstructions of the changing extent of wetlands and vegetative cover in the Sahara/Sahel region and from evidence for high altitude lake level variations in the Altiplano of the central Andes. Gasse (2000) summarizes evidence showing strong coherence between salinity changes in the North Atlantic and the lake level variations recorded across central Africa. The dynamics of these dramatic, low latitude hydrological changes are further complicated by the demonstration that the changes in surface hydrology and vegetation themselves have a crucial feedback to the regional climate. Without including this feedback alongside other forcing, including sea-surface temperatures, it is impossible for existing models to come even close to simulating the extent of wetlands and plant cover in the Sahara/Sahel region during much of the early–mid Holocene (DeMenocal *et al.*, 2000).

THE SECOND HALF OF THE HOLOCENE

Before sharpening the focus on the late Holocene, it is useful to consider briefly some of the evidence that has been interpreted as reflecting the persistence of modes of variability throughout the Holocene. For example, the lake level record from central Africa shows evidence for rapid changes after 6 kyr BP. The changes, though no less sudden than those in the early Holocene, become somewhat reduced in overall amplitude. In the North Atlantic there are

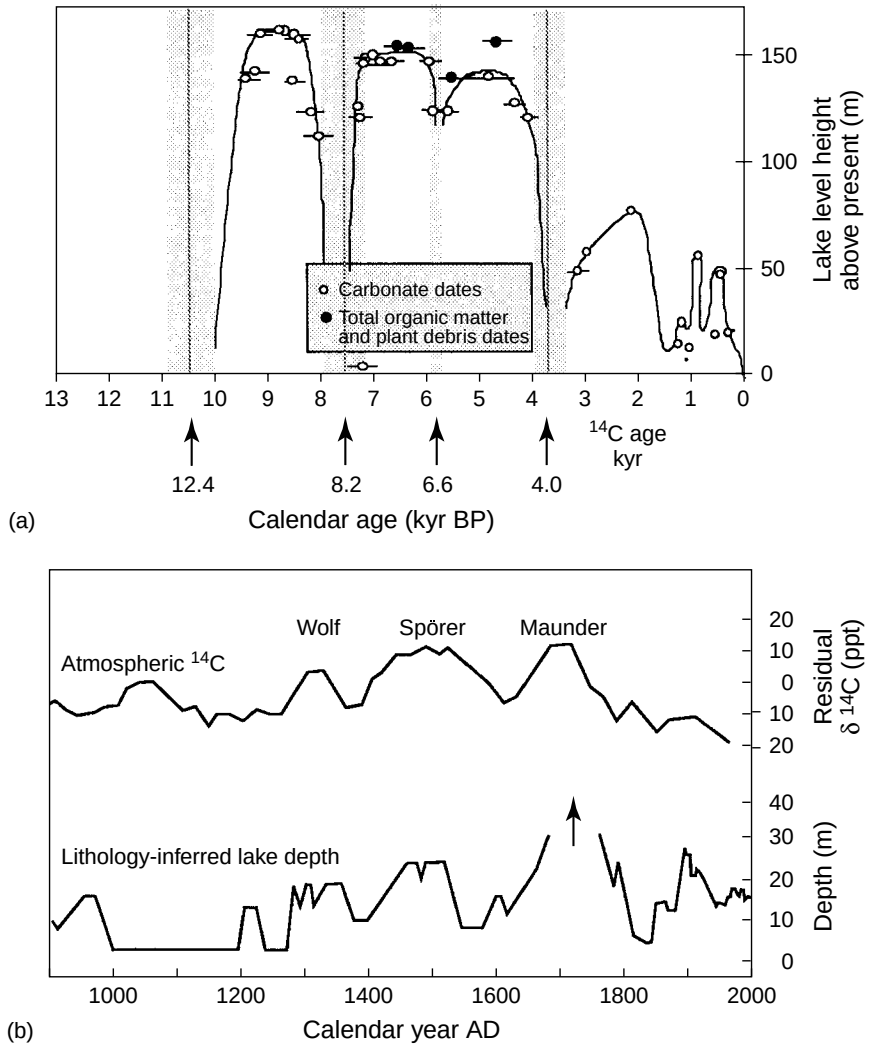


Figure 4 Holocene lake level records from much of Africa show dramatic oscillations well beyond the range of variability captured by the instrumental record. (a) the record from lake Abhe Ethiopia showing periods of extreme drought lasting for centuries with lake level changes of over 100 m occurring on these timescales. (b) A high resolution lake level record for the past millennium from Lake Naivasha, Kenya showing periods of multidecadal drought and lake level variations of over 10 m. Over this time period, lake level is well correlated with atmospheric ^{14}C, a possible proxy for solar activity. (Reproduced with permission from Gasse, 2000 and Verschuren *et al.*, 2000)

also signs that the mode of variability expressed in earlier times successively as D/O events, the Younger Dryas and the 8.2 kyr event persists throughout the Holocene with a somewhat irregular periodicity within the thousand to three thousand year range (Bond *et al.*, 1999). Similar evidence, also indicative of rapid climate shifts, can be cited from many parts of the world, but better chronological control is needed before a clearer picture emerges of the temporal and spatial coherence of these variations.

Turning now to the later part of the Holocene, additional records, many of which have annual resolution, become available. Moreover, the concept of rapid or sudden change becomes something that can be better quantified over large areas and more readily compared with the variations recorded during the recent period of instrumental observations which, for most of the world, seldom exceeds the last 150 years. Attention has focused on the so-called Little Ice Age (LIA) and Medieval Warm Period (MWP) because these are claimed to have had a strong influence on settlements and prosperity especially in Europe. Chronologically constrained evidence for climate changes broadly paralleling but rarely perfectly synchronous with the European MWP–LIA sequence has been identified in records from around the Northern Hemisphere and into low latitudes – for example in the record from the Quelcaya Ice Cap in Peru. Where it has proved possible to analyze high

frequency modes of variability such as El Niño–Southern Oscillation (ENSO) in the time domain, clear evidence is emerging of low frequency abrupt shifts in mode involving both amplitude and periodicity. Increasingly, as detailed chronologies become applicable to a wider range of sites and archives, key time intervals both within and beyond the last two millennia are showing evidence of rapid climatic change with signatures and impacts over a wide area (Mann *et al.*, 1999).

Out of this mix of indications, some generalizations about forcing are beginning to emerge. One of the influences frequently cited is solar variability. In some cases, links between climate and solar variability appear to be present on decadal to century timescales (Verschuren *et al.*, 2000) (Figure 4b) although the precise mechanisms involved are still a matter of controversy. This example is but one of many documenting climate shifts at the regional level and pointing to changes in drought frequency and amplitude well beyond those documented in the instrumental record. However, when archives as well dated as high latitude tree rings are compared across the whole of the Northern Hemisphere for their paleoclimatic response, non-synchroneities are clear. By contrast the brief and transient, but sometimes quite extreme, responses to volcanic eruptions are demonstrably synchronous over wide areas.

IMPLICATIONS FOR THE FUTURE

Growing attention has been paid to the possibility that anthropogenically driven climate change in the future may lead to surprises, major shifts well beyond the range of variability upon which planning and construction schemes are based and even outside the envelope of scenario projections generated by currently available climate models. The paleo-record does not preclude such possibilities, among which major changes to NADW formation have received most attention. Oceanic General Circulation Models raise the possibility that an inflow of sufficient fresh water from additional high latitude precipitation and runoff could have the effect of reducing NADW formation to the point of seriously altering the ocean thermohaline circulation. The climate changes linked to such an event would certainly constitute a surprise and, for many parts of the world, even a catastrophe.

The lessons from the paleorecord for the future are not limited to warnings about potential surprises on the scale of a cessation of NADW formation. Extending the record of climate variability back through time reveals changes, often sudden and sometimes persistent on decadal to century time-scales, which lie outside the range of instrumental records. The concepts of future sustainability, water supply and food security are limited and potentially dangerously short sighted if they fail to accommodate the evidence from the longer-term record.

See also: Earth System History, Volume 1.

REFERENCES

Alley, R B, Meese, D A, Shuman, C A, Gow, A J, Taylor, K C, Grootes, P M, White, J W C, Ram, M, Waddington, E D, Mayewski, P A, and Zielinski, G A (1993) Abrupt Increase in Snow Accumulation at the End of the Younger Dryas Event, *Nature*, 362, 527–529.

Behl, R J and Kennet, J P (1996) Brief Interstadial Events in the Santa Barbara Basin, NE Pacific, During the Past 60 kyr, *Nature*, 379, 243–246.

Bond, G C, Showers, W, Elliot, M, Lotti, R, Hajdas, I, Bonani, G, and Johnson, S (1999) The North Atlantic s 1–2 kyr Climate Rhythm: Relation to Heinrich Events, Dansgaard/Oescheger Cycles and the Little Ice Age, in *Mechanisms of Global Climate Change*, eds P U Clark, R S Webb, and L D Keigwin, American Geophysical Union, Washington, DC, 35–58.

DeMenocal, P, Ortiz, J, Guilderson, T, Adkins, J, Sarnthein, M, Baker, L, and Yarusinsky, M (2000) Abrupt Onset and Termination of the African Humid Period: Rapid Climate Responses to Gradual Insolation Forcing, in *Past Global Change and their Signi cance for the Future*, eds K Alverson, F Oldfield, and R S Bradley, Elsevier, Amsterdam, Vol. 9, 347–362.

Gasse, F (2000) Hydrological Changes in the African Tropics Since the Last Glacial Maximum, in *Past Global Change and their Signi cance for the Future*, eds K Alverson, F Oldfield, and R S Bradley, Elsevier, Amsterdam, Vol. 9, 189–211.

Grafenstein, U, Von Erlenkeuser, H, Muller, J, Jouzel, J, and Johnse, S (1998) The Cold Event 8200 Years Ago Documented in Oxygen Isotope Records of Precipitation in Europe and Greenland, *Clim. Dyn.*, 14, 73–81.

Henderson, G M and Slowey, N C (2000) Evidence from U–Th Dating Against Northern Hemisphere Forcing of the Penultimate Deglaciation, *Nature*, 404, 61–66.

Mann, M E, Bradley, R S, and Hughes, M K (1999) Northern Hemisphere Temperatures During the Past Millenium: Inferences, Uncertainties, and Limitations, *Geophys. Res. Lett.*, 26, 759–762.

Petit, J R, Jouzel, J, Raynaud, D, Barkov, N I, Barnola, J M, Basile, I, Benders, M, Chappellaz, J, Davis, M, Delaygue, G, Delmotte, M, Kotlyakov, V M, Legrand, M, Lipenkov, V Y, Lorius, C, Pepin, L, Ritz, C, Saltzman, E, and Stievenard, M (1999) Climate and Atmospheric History of the Past 420 000 Years from the Vostok Ice Core, Antarctica, *Nature*, 399, 429–436.

Raynaud, D, Barnola, J-M, Chappellaz, J, Blunier, T, Indermuhle, A, and Stauffer, B (2000) The Ice Record of Greenhouse Gases: a View in the Context of Future Changes, in *Past Global Change and their Signi cance for the Future*, eds K Alverson, F Oldfield, and R S Bradley, Elsevier, Amsterdam, Vol. 9, 9–18.

Verschuren, D, Laird, K R, and Cumming, B F (2000) Rainfall and Drought in Equatorial East Africa During the Past 1100 Years, *Nature*, 403, 410–414.

Climate Change, Anthropogenic

see Projection of Future Changes in Climate (*Opening essay*, **Volume 1**); Climate Change, Detection and Attribution (**Volume 1**)

Climate Change, Detection and Attribution

Myles R Allen
Rutherford Appleton Laboratory, Didcot, UK

Research in the detection and attribution of climate change addresses two basic questions. First, is the climate changing by more than we would expect from internally generated variability (the detection problem)? Second, can we assign this change to a speci c external cause (such as human activity) or combination of causes (the attribution problem)? The main challenge is that we can only observe climate change in the real world once: controlled, repeatable experimentation on the planet being infeasible (not to mention unethical). Hence detection and attribution involve detailed comparisons of multiple model simulations with the observed record, comparing modeled and observed patterns of change to build a statistical picture of what is going on. Formal statements of causal attribution, in any eld, will always be subject to some level of uncertainty: the goal is to be as quantitative as possible in establishing the level of uncertainty and reducing it as far as possible.

Although human activities have been changing atmospheric composition for two centuries, prior to the mid-1990s most scientists were reluctant to attribute observed large-scale changes in climate specifically to human influence. The Second Assessment Report (SAR) of the Intergovernmental Panel on Climate Change (IPCC), (Santer *et al.*, 1996a), broke significant ground in concluding that the balance of evidence suggests a discernible human influence on global climate (*see* Intergovernmental Panel on Climate Change (IPCC): an Historical Review, Volume 4). While a major step forward in building confidence in overall scientific understanding of the problem, this statement was still very cautious. In particular, the SAR did not attempt to quantify the magnitude of this human influence, and it also avoided attributing all or even part of the observed change to a specific forcing agent, such as anthropogenic greenhouse gases.

CORRELATIONS BETWEEN MODELED AND OBSERVED CHANGES

The reason for this caution was that the SAR statement was primarily based on the evidence provided by correlations between observed and model-simulated patterns of trends in near surface (Santer *et al.*, 1994) and zonal mean atmospheric temperatures (Santer *et al.*, 1996b; Tett *et al.*, 1996) over the past 30–50 years. These correlations were found to be significantly higher than would be expected by chance due to internal climate variability as simulated by control integrations of a range of climate models. Although attractively simple to implement and explain to non-specialists, the pattern correlation approach had two acknowledged shortcomings. First, correlations are insensitive to the relative magnitude of observed and model simulated changes, so they do not convey any information about the size of the observed change or whether the models are over- or underestimating it. Second, a positive correlation between an observed change and a single model simulated response pattern does not necessarily preclude some other external influence, which happens to generate a similar pattern of response, confounding the picture.

For example, the combination of an increase in solar activity causing warming of the troposphere and depletion of stratospheric ozone causing cooling of the stratosphere (Ramaswamy *et al.*, 1996) might, together, generate a pattern of atmospheric temperature change rather similar to that expected from the anthropogenic increase in greenhouse gas levels. Stratospheric ozone depletion is itself primarily anthropogenic, so this possibility does not bring into question the conclusion of human influence of some form. Nevertheless, because ozone depletion is expected to reverse over the coming decades while greenhouse gas increases are expected to dominate anthropogenic change over the 21st century, the evidence for anthropogenic greenhouse influence is of greater practical significance than the evidence for human influence *per se*.

In summary, the evidence available at the time of the SAR was sufficient to conclude that a significant climate change was occurring that could be explained by human influence. Nevertheless, both the size of the contributions from different anthropogenic agents, notably greenhouse gases, sulfate aerosols and stratospheric ozone depletion, and the extent of other natural influences, such as changes in solar or volcanic activity, remained uncertain. Thus, as was noted at the time, the IPCC statement in the SAR allowed for the possibility of a statistically significant anthropogenic climate change that was too small to have any practical importance.

NEW DEVELOPMENTS IN DETECTION AND ATTRIBUTION

Five years on, the situation has changed considerably, and we are in a position to make a more quantitative assessment of the extent to which recent observed climate changes are attributable to human influence, particularly the impact of anthropogenic greenhouse gas emissions.

The first new development is an apparently strengthening signal in the observational record: the 1990s was the warmest decade, and 1998 the warmest year, almost certainly over the past century and very possibly over the past millennium (Mann *et al.*, 1998). The availability of reconstructions of past temperatures using indirect proxy indicators extending the temperature record back beyond the 140 years of reasonably global direct measurements is an important new development, because it places recent changes in the context of a longer record. Individual years or decades should not, however, be over-interpreted: in particular, an unusually strong El Niño occurred in 1998. Confidence in detection and attribution of human influence on climate only emerges as model-based predictions are borne out over a much longer period.

A second important development is increased confidence in estimates of internal climate variability. These proxy climatic records also provide a check on the realism of model simulated internal variability. In general, the more up-to-date models generate internal variability similar in magnitude to that observed once changes attributable to natural and anthropogenic external factors are taken into account (Hegerl *et al.*, 1996; Allen and Tett, 1999; Crowley, 2000; Stott *et al.*, 2000b; Allen *et al.*, 2001). None of these models generates warming trends as large as that observed over the past century through internal variability alone, even when run for several thousand years.

A third important development since the SAR is the wider availability of comprehensive model simulations of the climate response to a range of different external forcing scenarios, including various combinations of anthropogenic forcing agents and natural external influences such as variations in solar and volcanic activity (e.g., Tett *et al.*, 1999). A crucial element of any quantitative assessment of the extent to which an observed change can be attributed to a particular external factor is an estimate of the expected change due to that factor, together with estimates of expected changes due to physically plausible alternatives. In simple terms, to assess whether a signal is there, we need to know both what we are looking for and what we need to discriminate against.

Finally, we are now using a wider range of quantitative tools for model–data comparison that take into account the relative magnitude of model simulated and observed changes, allowing us to assess whether models over- or underestimate the observed response. These tools are primarily based on the *ngerprinting* approach of Hasselmann (1979, 1993), also referred to as *optimal averaging* (Bell, 1986) or *optimal ltering* (North *et al.*, 1995). The first applications of the fingerprinting approach (Hegerl *et al.*, 1996) considered only a single pattern of anthropogenic influence, and were therefore subject to the problem of confounding signals mentioned above. More recent studies, from Hegerl *et al.* (1997) onwards, allow for multiple uncertain contributions to recent climate change, assuming that the observed change consists of the sum of a set of model simulated changes, including the response to greenhouse gases, sulfate aerosols, and solar and volcanic activity, for example. Allen and Tett (1999) discuss the physical interpretation of this multi-fingerprinting algorithm in some detail. In essence, the parameters being estimated are the factors by which it is necessary to scale model-simulated signals to reproduce observed climate changes. A scaling factor consistent with zero implies that this particular model-simulated signal is not detectable in the observed climate record using the data set under consideration, while a scaling factor consistent with unity implies the model simulated amplitude could be correct. In a very similar vein, Leroy (1998) reformulated the approach of North and Stevens (1998) in terms of Bayesian estimation theory, a more flexible framework that allows multiple lines of evidence to be drawn together in a formal manner. The algorithms of Allen and Tett (1999) and Leroy (1998), are identical in the absence of any prior expectation about (or equivalently, given infinite prior variance in) scaling factors to be applied to individual signals.

EVIDENCE FROM PATTERNS OF SURFACE TEMPERATURE TRENDS

Hegerl *et al.* (1997) concluded that both anthropogenic greenhouse gas and sulfate influences were required to account for the observed spatial pattern of boreal summertime warming trends over the past fifty years, while cautiously noting that some influence of solar variability might be required to account for the warming observed in the early decades of the 20th century. In a number of follow-up papers (Barnett *et al.*, 1999; Hegerl *et al.*, 2000), the authors have examined the sensitivity of this result to the model used to simulate the characteristics of anthropogenic signals and internal climate variability, finding that their original conclusions regarding sulfate influence on climate were particularly sensitive to the model used and assumptions made in the analysis.

EVIDENCE FROM THE TIME-HISTORY OF RECENT CHANGES

Tett *et al.* (1999) applied a similar approach to observed spatio-temporal patterns of surface temperature change

(Hegerl *et al.* (1996) and subsequent papers focused on spatial patterns of linear temperature trends) and also concluded that both greenhouse and sulfate influences are required to account for observed near-surface temperature changes in the latter half of the 20th century, with a possible role for solar variability in the warming that occurred over the period 1910–1940. With ensemble simulations of the response to natural external forcing agents at their disposal, they were able to draw stronger conclusions than previous authors regarding the difficulty of accounting for observed changes in exclusively natural terms.

Stott *et al.* (2001) show that the inclusion of seasonal information strengthens the evidence for solar influence on the early part of the 20th century. In contrast, Delworth and Knutson (2000), using a different model to estimate internal variability which displays greater variance on 50–80 year time scales, conclude that the warming that occurred early in the century can be accounted for through a combination of anthropogenic influence and internal variability. There is no inherent inconsistency between these two results: Stott *et al.* (2001) were using a more powerful analysis, in that they included both seasonal information and a model based estimate of the spatio-temporal pattern of response to solar forcing (not available to Delworth and Knutson (2000), because their model had not yet been run with estimates of changing solar forcing). It is not surprising that a more powerful analysis detects a weak signal when a less powerful one fails, but this sensitivity to the diagnostic and/or model used indicates that the evidence for a solar role in early 20th century warming remains weaker than the evidence for anthropogenic influence in more recent decades.

The earliest attribution studies, such as Hegerl *et al.* (1996) and Hegerl *et al.* (1997), used signals based either on the equilibrium response of a climate model to a particular forcing or on the transient response over some period in the future. In both cases, the forcing would be sufficiently strong that sampling uncertainty in the model-simulated response would be negligible: if the experiment were repeated, essentially the same model-simulated signal would be obtained. In contrast, more recent studies, beginning with Tett *et al.* (1996), have used signals based on model-simulated responses to forcing changes over the same period as the observations used in the comparison. This approach simplifies interpretation of results, but introduces the complication that climate forcing is not as strong over the 20th century, so model-simulated signals are themselves contaminated by internal variability. This contamination can be reduced through the use of ensemble simulations, as in Tett *et al.* (1999), or through the use of noise free-climate models, as in North and Stevens (1998). The ensemble approach is expensive, so ensembles are not available for all models, while any climate model that is free of internal variability is also likely to be lacking key non-linear feedback processes that may affect the model simulated signal.

Allen *et al.* (2000) and Stott *et al.* (2000a) describe and test a modification to standard fingerprinting that accounts explicitly for the presence of sampling noise in model simulated signals based on single member or small size ensembles. They follow the standard total least squares (TLS) estimation procedure documented in detail by van Huffel and Vanderwaal (1994). Stott *et al.* (2000a) demonstrate that this revised algorithm is particularly important if small (one or two member) ensembles are used in climate model simulations, although it can still make a significant difference even if larger ensembles are available, particularly on upper bounds of uncertainty ranges and when considering relatively weak model-simulated signals (such as the response to natural external influences). Allen *et al.* (2001) apply this procedure to a wide range of model simulations of recent near-surface temperature change, while Stott *et al.* (2000b) and Tett *et al.* (2001) investigate the most up-to-date and comprehensive set of model simulations available to date for detection and attribution studies.

CONCLUSIONS OF KEY RECENT STUDIES

1. When the spatio-temporal patterns of change are taken into account, recent changes cannot be accounted for as internal variability even if the amplitude of model-simulated variability is increased by a factor of two or more.
2. Natural external forcing (solar and volcanic activity) cannot account for the observed acceleration in warming between the 1960–1970s and 1980–1990s, even if amplified by unknown feedback mechanisms, because solar forcing leveled off over this period and volcanic activity was acting as a cooling influence.
3. Solar activity may have contributed significantly to the warming observed in the early 20th century, but uncertainties in the reconstruction of past forcings make this only a tentative conclusion.
4. There is evidence for anthropogenic sulfate and ozone influence on recent near-surface and stratospheric temperature changes, respectively, but the estimated size of this influence depends acutely on the models used to estimate both forcing and response.
5. By far the strongest and clearest signal in the recent climate record is the influence of the anthropogenic increase in greenhouse gas concentrations, with the warming trend attributable to greenhouse gases being at least half and possibly up to three times the total observed warming, depending on the diagnostic used and period considered.

Drawing this information together, the *IPCC Third Assessment Report* (Mitchell *et al.*, 2001) concludes there

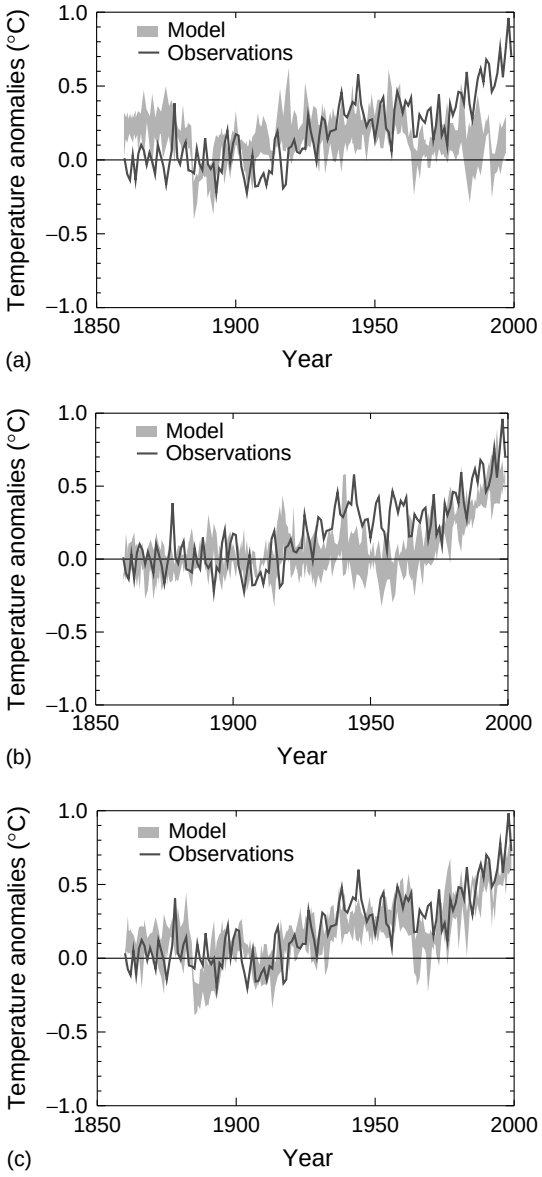


Figure 1 Global mean temperature from 1861 to the present as observed (solid lines) and as simulated by the HadCM3 climate model from the UK Met Office (gray bands, with the width of the bands representing internal model variability within four-member ensembles): (Stott *et al.*, 2000b). In panel (a), the model is forced with the main natural influences on climate, namely solar variability and volcanic activity: the model clearly fails to reproduce the rapid warming over the past three decades. In panel (b), the model is forced with anthropogenic influences alone, particularly greenhouse gases, sulfate aerosols, and stratospheric and tropospheric ozone: the model now fails to simulate the observed early-century warming. In panel (c), all forcings are imposed, and the agreement between model and data is remarkably good. (Figure as published in the Summary for Policymakers of the IPCC Third Assessment Report, prepared by Peter Stott of the UK Met Office and Paul van der Linden of the IPCC WGI Technical Support Unit, and used by permission)

is new and stronger evidence that most of the warming observed over the past 50 years is attributable to human activity, specifically the influence of anthropogenic greenhouse gases.

Figure 1, from Stott *et al.* (2000b), summarizes this evidence: panel (a) shows that the model-simulated response to natural forcings alone cannot explain the warming trend since 1980, panel (b) shows that anthropogenic influence alone can explain the most recent warming but leaves some discrepancies earlier in the century, while panel (c) shows that a combination of natural and anthropogenic influences account, to a remarkably high degree, for observed large-scale temperature changes over the full 20th century.

While much progress has been made in the detection and attribution of temperature changes on continental and global scales, attribution of regional changes and of changes in non-temperature variables remains much more speculative. Changes have been observed that are consistent with those we would expect to result from anthropogenic climate change, including in particular a 10–20 cm rise in sea level over the past century, a significant fraction of which is attributable to an increase in global ocean heat content. Other changes that may be partly due to human influence include a 10% decrease in snow cover since the 1960s; substantial declines in Arctic sea ice thickness and extent; widespread retreat of mountain glaciers; over land a decrease in the temperature difference between night time and day time, and between winter and summer; and, possibly, some increase in the frequency and intensity of heavy rain and snowfall events over Northern Hemisphere temperate regions.

The lack of long observational records or reliable model-based estimates of internal variability in these non-temperature indicators makes it impossible formally to quantify the probability that any of these changes, taken individually, is human-induced rather than natural. Taken together, however, they present a collective picture of a changing global climate increasingly under the influence of human activity.

ACKNOWLEDGMENTS

Myles Allen would like to acknowledge support from the UK Natural Environment Research Council through an Advanced Research Fellowship and the US Department of Energy and National Ocean and Atmosphere through the International Ad Hoc Detection Group.

See also: Model Simulations of Present and Historical Climates, Volume 1.

REFERENCES

Allen, M R and Tett, S F B (1999) Checking Internal Consistency in Optimal Fingerprinting, *Clim. Dyn.*, 15, 419–434.

Allen, M R, Stott, P A, Mitchell, J F B, Schnur, R, and Delworth, T (2000) Quantifying the Uncertainty in Forecasts of Anthropogenic Climate Change, *Nature*, 407, 617–620.

Allen, M R, Gillett, N P, Kettleborough, J A, Hegerl, G C, Schnur, R, Stott, P A, Boer, G, Covey, C, Delworth, T L, Jones, G S, Mitchell, J F B, and Barnett, T P (2001) Quantifying Anthropogenic Influence on Recent Near-surface Temperature Change, *Surv. Geophys.*, in press.

Barnett, T P, Hasselmann, K, Chelliah, M, Delworth, T, Hegerl, G, Jones, P, Rasmusson, E, Roeckner, E, Ropelewski, C, Santer, B, and Tett, S (1999) Detection and Attribution of Recent Climate Change: A Status Report, *Bull. Am. Meteorol. Soc.*, 80, 2631–2659.

Bell, T L (1986) Theory of Optimal Weighting to Detect Climate Change, *J. Atmos. Sci.*, 43, 1694–1710.

Crowley, T J (2000) Causes of Climate Change over the Past 1000 Years, *Science*, 289, 270–276.

Delworth, T L and Knutson, T R (2000) Simulation of Early 20th Century Global Warming, *Science*, 287, 2246–2250.

Hasselmann, K (1979) On the Signal-to-noise Problem in Atmospheric Response Studies, in *Meteorology of Tropical Oceans*, ed T Shawn, Royal Meteorological Society, London, 251–259.

Hasselmann, K (1993) Optimal Fingerprints for the Detection of Time Dependent Climate Change, *J. Clim.*, 6, 1957–1971.

Hegerl, G, Hasselmann, K, Cubasch, U, Mitchell, J F B, Roeckner, E, Voss, R, and Waszkewitz, J (1997) On Multi-fingerprint Detection and Attribution of Greenhouse Gas and Aerosol Forced Climate Change, *Clim. Dyn.*, 13, 613–634.

Hegerl, G C, von Storch, H, Hasselmann, K, Santer, B D, Cubasch, U, and Jones, P D (1996) Detecting Greenhouse Gas-induced Climate Change with an Optimal Fingerprint Method, *J. Clim.*, 9, 2281–2306.

Hegerl, G C, Stott, P A, Allen, M R, Mitchell, J F B, Tett, S F B, and Cubasch, U (2000) Optimal Detection and Attribution of Climate Change: Sensitivity of Results to Climate Model Differences, *Clim. Dyn.*, 16, 737–754.

Leroy, S (1998) Detecting Climate Signals, some Bayesian Aspects, *J. Clim.*, 11, 640–651.

Mann, M E, Bradley, R S, and Hughes, M K (1998) Global-scale Temperature Patterns and Climate Forcing over the Past Six Centuries, *Nature*, 392, 779–787.

Mitchell, J F B, Karoly, D J, Hegerl, G C, Zwiers, F W, Allen, M R, and Marengo, J (2001) Detection of Climate Change and Attribution of Causes, in *Climate Change 2001, The Science of Climate Change*, eds J T Houghton, Y Ding, D J Griggs, M Noguer, P J van der Linden, X Dai, K Maskell, and C A Johnson, Cambridge University Press, Cambridge.

North, G R, Kim, K Y, Shen, S S P, and Hardin, J W (1995) Detection of Forced Climate Signals, 1: Filter Theory, *J. Clim.*, 8, 401–408.

North, G R and Stevens, M J (1998) Detecting Climate Signals in the Surface Temperature Record, *J. Clim.*, 11, 563–577.

Ramaswamy, V, Schwarzkopf, M D, and Randel, W J (1996) Fingerprint of Ozone Depletion in Spatial and Temporal Patterns of Recent Lower Stratospheric Cooling, *Nature*, 382, 616–618.

Santer, B D, Bruggemann, W, Cubasch, U, Hasselmann, K, Hock, H, Maier-Reimer, E, and Mikolajewicz, U (1994) Signal-to-noise Analysis of Time-dependent Greenhouse Warming Experiments, Part 1: Pattern Analysis, *Clim. Dyn.*, 9, 267–285.

Santer, B D, Wigley, T M L, Barnett, T P, and Anyamba, E (1996a) Detection of Climate Change and Attribution of Causes, in *Climate Change 1995, The Science of Climate Change*, eds T J Houghton, L G Meira Filho, B A Callander, N Harris, A Kattenberg, and A Maskell, Cambridge University Press, 411–443.

Santer, B D, Taylor, K E, Wigley, T M L, Johns, T C, Jones, P D, Karoly, D J, Mitchell, J F B, Oort, A H, Penner, J E, Ramaswamy, V, Schwarzkopf, M D, Stouffer, R J, and Tett, S (1996b) A Search for Human Influences on the Thermal Structure of the Atmosphere, *Nature*, 382, 39–46.

Stott, P A, Allen, M R, and Jones, G S (2000a) *Estimating Signal Amplitudes in Optimal Fingerprinting II: Application to General Circulation Models*, Technical report 20, Hadley Centre for Climate Prediction and Research.

Stott, P A, Tett, S F B, Jones, G S, Allen, M R, Mitchell, J F B, and Jenkins, G J (2000b) External Control of Twentieth Century Temperature by Natural and Anthropogenic Forcings, *Science*, 290, 2133–2137.

Stott, P A, Tett, S F B, Jones, G S, Allen, M R, Ingram, W J, and Mitchell, J F B (2001) Attribution of Twentieth Century Climate Change to Natural and Anthropogenic Causes, *Clim. Dyn.*, 15, 1–22.

Tett, S F B, Mitchell, J F B, Parker, D E, and Allen, M R (1996) Human Influence on the Atmospheric Vertical Temperature Structure: Detection and Observations, *Science*, 247, 1170–1173.

Tett, S F B, Stott, P A, Allen, M R, Ingram, W J, and Mitchell, J F B (1999) Causes of Twentieth Century Temperature Change Near the Earth s Surface, *Nature*, 399, 569–572.

Tett, S F B, Jones, G S, Stott, P A, Hill, D C, Mitchell, J F B, Allen, M R, Ingram, W J, Johns, T C, Johnson, C E, Jones, A, Roberts, D L, Sexton, D M H, and Woodage, M J (2001) Estimation of Natural and Anthropogenic Contributions to Twentieth Century Temperature Change, *J. Geophy. Res.*, in press.

van Huffel, S and Vanderwaal, J (1994) *The Total Least Squares Problem: Computational Aspects and Analysis*, SIAM, Philadelphia, PA.

Climate Change, Framework Convention

see United Nations Framework Convention on Climate Change and Kyoto Protocol (**Volume 4**)

Climate Change, Natural Records of

see Natural Records of Climate Change (**Volume 1**)

Climate, Cryosphere Interaction (CLIC)

see CLIC (Climate and Cryosphere) (**Volume 1**)

Climate Extremes

see Climatic Extremes (**Volume 3**)

Climate, Energy Balance

see Energy Balance and Climate (**Volume 1**)

Climate Feedbacks

L D Danny Harvey
University of Toronto, Toronto, Canada

Climate can change for a variety of reasons, including changes in the composition of the Earth's atmosphere, changes in the energy output from the Sun, changes in the Earth's orbit around the Sun, or changes in the position and area of continents or in the nature of the land surface. These factors cause the climate to change by upsetting the balance between the absorption of solar energy by the sun and the emission of infrared radiation to space by the earth and atmosphere. The change in the net radiation at the tropopause in response to some perturbing factor is called the radiative forcing, and has units of $W\,m^{-2}$ *(watts per square meter), and is what drives changes in the surface climate. The radiative forcing is translated into a temperature change by a variety of feedbacks which either amplify or diminish the initial radiative forcing. A positive climate feedback adds to the initial radiative perturbation and so increases the eventual change in climate, while a negative feedback subtracts from the initial perturbation and so reduces the eventual change in climate.*

The temperatures of the atmosphere and surface tend to adjust themselves such that there is a balance between the absorption of energy from the sun and the emission of infrared radiation to space. If, for example, there were to be an excess of absorbed solar energy over emitted infrared radiation (as occurs with the addition of greenhouse gases to the atmosphere), temperatures would increase but, in so doing, the emission of infrared radiation to space would increase. This would reduce the initial imbalance, and eventually a new balance would be achieved, but at a new, higher temperature. The more rapidly that infrared emission to space increases with increasing temperature, the less that the temperature must increase in order to restore balance.

However, as temperatures increase, other quantities would also change, which would further alter the net emission of radiation to space. For example, an increase in the amount of water vapor as temperature increases would tend to counteract the effect of warmer temperatures in increasing the emission of infrared radiation to space, since water vapor absorbs infrared radiation. As a result, infrared emission would increase more slowly and a greater temperature increase would be required to restore radiative balance. A reduction in the amount of solar energy reflected to space would also tend to offset the effect of increasing infrared emission to space. The resulting rate at which the net emission of radiation to space increases is called the *radiative damping*, λ. The term damping is used because it is the increase in net emission to space as temperatures warm that limits or dampens the increase in temperature. A larger radiative damping results in a smaller temperature increase.

This discussion can be succinctly summarized by the equations

$$\Delta T = \frac{\Delta R}{\lambda} \tag{1}$$

and

$$\lambda = \frac{\mathrm{d}F}{\mathrm{d}T} - \frac{\mathrm{d}Q}{\mathrm{d}T} \tag{2}$$

where ΔT is the eventual change in global average temperature, ΔR is the global average radiative forcing, F is the global average emission of infrared radiation to space, Q is the global average absorption of solar radiation, and T is the global average temperature (strictly speaking, F should be evaluated at the tropopause and Q should be based on the absorption of solar radiation by the surface and troposphere only, since we are interested in surface temperature change).

In the absence of any feedback except the increase in temperature itself, the only way that the emission of radiation to space increases is through the direct dependence of emission on temperature through the Stefan–Boltzman Law. In this case, $\lambda = 3.76\,\mathrm{W\,m^{-2}\,K^{-1}}$. Since a doubling of atmospheric CO_2 (carbon dioxide) causes a global mean radiative surplus ΔR of $3.5–4.0\,\mathrm{W\,m^{-2}}$, the temperature increase required to restore radiative balance in this case would be $\Delta R/\lambda = 1.0\,\mathrm{K}$ (a warming of $1.0\,^{\circ}\mathrm{C}$). However, as noted above, the very change in temperature would cause other atmospheric and surface properties to change, which would lead to further alterations in the energy balance and require further temperature changes through a series of feedback processes.

MAJOR CLIMATE FEEDBACKS

Feedbacks translate a radiative forcing into a change in temperature by altering the radiative damping. Among the feedbacks that have to be considered are the following:

- *Water vapor amount*: in a warmer climate the atmospheric concentration of water vapor will increase. Because water vapor is a greenhouse gas, this is a positive feedback – that is, it amplifies the initial change.
- *Water vapor distribution*: if, as the total water vapor amount increases, the percentage increase is greater at higher altitudes than at lower altitudes, this will result in a greater climate warming than if the water vapor amount increases by the same percentage at all altitudes, as explained later. The case where the percentage increase in water vapor is greater at higher altitude is equivalent to increasing the water vapor amount by the same percentage at all altitudes, then shifting the distribution upward. This serves as a positive feedback. Conversely, a downward shift in the water vapor distribution as the climate warms would serve as a negative feedback.
- *Clouds*: changes in clouds can serve as a positive or negative feedback on climate. Clouds have a cooling effect by reflecting sunlight, but they also exert a warming effect by absorbing the infrared radiation emitted by the surface and re-emitting a smaller amount of radiation due to the fact that the cloud top is colder than the surface. The net feedback of changes in clouds depends on the details of when and how clouds change, so even the sign of the feedback is uncertain (*see* Clouds, Volume 1).
- *Atmospheric temperature structure*: changes in the rate of decrease of tropospheric temperature with height (or lapse rate) will alter the relationship between surface temperature and infrared emission to space, thereby exerting a feedback effect on climate.
- *Areal extent of ice and snow*: a reduction in the area of sea ice and seasonal snow cover on land as climate warms will reduce the surface reflectivity, thereby tending to produce greater warming (a positive feedback). However, concurrent changes in cloud cover that could be induced by the change in ice or snow cover could significantly alter the net feedback, as explained later.
- *Vegetation*: changes in the distribution of different biomes or in the nature of vegetation within a given biome can also lead to changes in the surface reflectivity, thereby exerting a feedback effect on climatic change.
- *The carbon cycle*: the effect of climate on the terrestrial biosphere and the oceans is likely to alter the sources and sinks of CO_2 and CH_4 (methane) leading to changes in their atmospheric concentrations and hence causing a radiative feedback.
- *The sulfur cycle*: changes in climate or associated changes in the vertical mixing of nutrients in the ocean are likely to lead to changes in the production of dimethyl sulfide (DMS) by bacteria in the ocean, which in turn could lead to changes in the optical properties of clouds (*see* Dimethylsul de (DMS), Volume 1).
- *Atmospheric chemistry*: changes in climate will also lead to changes in the chemistry of the atmosphere, in turn indirectly affecting the concentration of a number of greenhouse gases and thereby altering the emission of infrared radiation to space. In particular, in a warmer climate there will be a greater concentration of water vapor and hence of the hydroxyl radical (which is formed by dissociation of water vapor by ultraviolet radiation). The hydroxyl radical in turn determines the life-span and hence atmospheric concentration of methane, and influences the concentration of tropospheric ozone.

Of these feedbacks, those involving water vapor, the lapse rate, atmospheric chemistry, and clouds respond essentially instantaneously to climatic change, while those involving sea ice and snow respond within a few years. These feedbacks are therefore referred to as fast feedbacks. Some vegetation and carbon cycle processes are relevant on a time-scale of decades. Other feedbacks, such as dissolution of carbonate sediments in the ocean and enhanced chemical weathering on land (both of which tend to reduce the atmospheric CO_2 concentration), require hundreds to thousands of years to unfold. Feedbacks involving continental ice sheets occur over periods of thousands of years.

The term *climate sensitivity* refers to the ratio of the long term (steady-state) increase in the global and annual mean surface air temperature to the global and annual mean radiative forcing (*see* Climate Sensitivity, Volume 1). It is standard practice to include only the feedbacks due to water vapor, seasonal ice and snow, and clouds – all of which are fast feedback processes – in the calculation of climate sensitivity using computer climate models. Feedbacks involving DMS and clouds have been excluded because they are too uncertain, while atmospheric chemistry feedbacks have not yet been included because coupled climate–chemistry models have only recently been developed (atmospheric chemistry feedbacks are not expected to be large). In the long-term, a warming of the climate might induce increases in the atmospheric concentration of CO_2, but these and other slow feedback processes are not included in the conventional definition of climate sensitivity. To calculate these feedbacks requires the use of coupled climate–carbon cycle models, with a representation of the geographical distribution of different ecosystems and of biological and chemical processes affecting the cycling of carbon in the oceans. Unlike the fast feedback processes, which are driven by overall climatic change independently of how fast the change occurs, many potential

climate–carbon cycle feedbacks depend to some extent on the rate of change as well as on the absolute change, and thus depend on the particular future human emissions of greenhouse gases and aerosols. Thus, there are distinct advantages in restricting the term climate sensitivity to the fast feedback processes while recognizing the potential for further, longer term feedback effects.

EVALUATION OF THE FAST FEEDBACK PROCESSES USING CLIMATE MODELS

The only models with the potential to reliably simulate the individual feedback processes are three-dimensional (3D) atmospheric general circulation models (AGCMs) with high spatial resolution and a diurnal cycle. However, the complexity of the processes involved in these feedbacks and the need for rather coarse resolution in AGCMs means that results obtained from AGCMs are still rather uncertain.

For example, the amount and distribution of water vapor depends on evaporation of surface water, horizontal and vertical transport of moisture by large-scale winds and small-scale eddies, upward movement by dry convection and within convective clouds, detrainment of moisture from cloud tops, dissipation of non-precipitating clouds, partial re-evaporation of precipitation, and the downward movement of dry air between rising convective columns at low latitudes. At present, only simplified representations of these processes are included in AGCMs, due mainly to the relatively coarse resolution of these models (from many tens to several hundred kilometers) compared with the scales of the most important processes (micrometers to kilometers).

Another example of difficult but important processes to simulate involves clouds. The initiation of clouds in nature depends on large scale motions, the coupling of these motions to the surface layer of the atmosphere (which feeds moisture into the overlying air), and the spatial distribution and intensity of surface–air fluxes of heat and moisture. Once clouds have been initiated in a model, a number of cloud properties and processes need to be simulated, including:

- the liquid and ice water content of clouds and their horizontal variability within model grid cells;
- the mean water droplet or ice crystal sizes and the distribution of sizes around the mean size;
- the optical properties of clouds for a given mean particle size and distribution – something that is complicated by the potential presence of impurities (such as soot) and is particularly difficult to predict for ice crystals because of their irregular shape;
- the fall velocity and hence life-span of water droplets and ice crystals;
- the cloud geometry (thickness, patchiness);
- the detrainment of cloud moisture and mixing with surrounding air, which is especially important to the formation of cirrus anvils from cumulus clouds;
- the formation of precipitation and the partial re-evaporation of falling precipitation, and;
- convective-scale downdrafts and their interaction with the surface layer of the atmosphere.

Most of these processes and properties have been incorporated in 3D climate models, but they are treated in a highly simplified and in some cases a rather *ad hoc* manner because most of them occur at scales well below the resolution of global models. This introduces the potential for substantial error in the simulated changes in cloud properties, locations, and amounts.

Clear-sky Feedbacks

The combination of changes in the amount of water vapor, in the vertical distribution of water vapor, and in the lapse rate in cloud-free regions is referred to as the clear-sky feedback. Analyses with a variety of climate models indicate that the clear-sky feedback is positive and sufficient to increase the climate sensitivity by 50–100% (Cess *et al.*, 1990; Harvey, 2000a, b). That is, the expected warming for a doubling of the atmospheric CO_2 concentration ranges from about 1.5 to 2.0 °C, rather than 1.0 °C (as in the case with no feedbacks other than a spatially uniform increase in temperature). The water vapor feedback is positive in these models due to the simple fact that the amount of water vapor in the atmospheric increases as the climate warms. The lapse rate feedback is negative at low latitudes and positive at high latitudes.

To understand the lapse rate feedbacks, suppose that there is an initial heating perturbation (such as a CO_2 increase) that begins to warm the climate. At low latitudes the lapse rate decreases as the climate warms, meaning that temperature begins to fall less rapidly with increasing height as the climate warms. This results in a larger increase in temperature in the upper troposphere than if the lapse rate were constant. This in turn causes a faster increase in the emission of infrared emission to space as surface temperature increases, which is the same as saying that the radiative damping is larger. As a result, the surface temperature does not need to warm up as much in order to restore radiative balance, which means that the climate sensitivity is smaller. At high latitudes, the opposite changes occur – the lapse rate increases as the climate warms (temperature falls more rapidly with increasing height), so there is less warming of the upper troposphere and less increase in the emission of infrared radiation to space, so a larger surface warming can occur.

The reason the lapse rate tends to decrease at low latitudes as the climate warms is that, in a warmer climate, there is more upward transport of water vapor by

convective clouds and a greater release of latent heat in the upper troposphere through condensation. In current models, some of the extra upward moisture transport leads to greater moistening of the upper transport, rather than immediate condensation and rainout. Thus, greater moistening of the upper troposphere (a stronger positive water vapor feedback) is correlated with greater warming of the troposphere through condensational heating (a stronger negative lapse rate feedback), which reduces the variation in the overall feedback between model versions in which these processes are parameterized in different ways (Zhang *et al.*, 1994).

The increase in the amount of water vapor in the upper troposphere is particularly important to the overall water vapor feedback and hence to climate sensitivity. This is so for two reasons: (1) the effect on an atmospheric layer s emissivity (its ability to absorb and re-emit infrared radiation) for a given increase in water vapor amount is greater the smaller the initial amount of water vapor (layers in the upper troposphere have less water vapor than layers of equal mass in the lower troposphere). (2) The effect of a given increase in emissivity on the emission of infrared radiation to space is larger the colder the layer in which the emissivity increases (the upper troposphere is colder than the lower troposphere). In most AGCMs, the amount of water vapor in the upper troposphere increases as the climate warms, but by less than what would be expected if the relative humidity (RH) were constant: in the Hadley Centre AGCM, RH decreases by 2–3% of the saturation value between 550 and 250 mb (Mitchell and Ingram, 1992); in the (National Center for Atmospheric Research (NCAR), Boulder, CO) AGCM, RH decreases throughout the tropical troposphere by up to 20% of the initial RH, but increases in the upper troposphere outside the tropics (Zhang *et al.*, 1994; Figure 16). Qualitatively similar behavior occurs in the (Geophysical Fluid Dynamics Laboratory (GFDL), Princeton, NJ) AGCM in response to a quadrupling of atmospheric CO_2, at least over continents (Wetherald and Manabe, 1995; Figure 6). Finally, Colman *et al.* (1997) report decreases in upper tropospheric RH in the (Bureau of Meteorology Research Centre (BMRC), Melbourne, Australia) AGCM in regions where strong increases in convection occur. However, in the (Goddard Institute for Space Studies (GISS), New York, USA) AGCM, RH is constant near the surface but increases in the global mean by 5% of the saturation value at the 200 mb level (Del Genio *et al.*, 1991). The reason that the GISS model gives an increase in upper tropospheric RH as climate warms, while the other models give a decrease, appears to be because greater sublimation of ice crystals that are detrained from cumulus anvils occurs in the GISS model. As noted above, more water is pumped into the upper tropical troposphere by convection in a warmer climate. In all models, a new steady state balance will be achieved in which essentially all of the extra upward moisture transfer will eventually fall out as precipitation. In the GISS model, the new steady state requires a greater increase in upper tropospheric moisture than in other models. In the other models, the required increase in upper tropospheric moisture is smaller because more of the anvil water immediately falls as precipitation.

The fraction of upward moisture transfer by convective clouds that immediately falls as precipitation is referred to as the precipitation efficiency (the rest detrains from convective columns and moistens the upper troposphere). This is an important consideration at low latitudes especially. In most AGCMs, the precipitation efficiency is not explicitly considered but would need to be diagnosed; in many cases it seems to be constant. This in turn explains why, in current models, an increase in moisture pumping leads to an increase in both condensational heating and moistening of the upper troposphere, causing variations in the strength of the positive water vapor feedback and the negative lapse rate feedback to partly cancel. However, if, as upward moisture pumping increases, a smaller fraction goes into moistening the upper troposphere and more into immediate precipitation, then the net effect will be a less positive change in upper tropospheric humidity and a more negative lapse rate feedback. That is, the difference in the strength of the two feedbacks would reinforce each other, and reduce climate sensitivity, rather than partly cancel each other. Calculations by Sun and Lindzen (1993; Figure 20) indicate that the time required for raindrops to form decreases as temperature increases, which implies that the precipitation efficiency should increase. This in turn might mean that the water vapor feedback at low latitudes (where convective processes are important) is somewhat too strong and the lapse rate feedback somewhat too weak.

Outside the tropics, the increase in water vapor as the climate warms is more directly related to the increase in the surface vapor pressure and in the atmosphere s ability to hold water, which in turn are directly related to the increase in temperature through the dependence of saturation vapor pressure on temperature as measured in the laboratory.

In summary, the water vapor feedback in climate models is strongly positive (sufficient to increase climate sensitivity by 50–100%), but in low latitudes is partly linked to the lapse rate feedback. The effect on climate sensitivity of changes in upper tropospheric water vapor and temperature, when both increase as overall water vapor pumping increases, is minimal because such changes have opposing effects. This diminishes the importance of changes in upper tropospheric water vapor due to changes in moisture pumping. However, if the precipitation efficiency increases as moisture pumping increases, there will be less cancellation of the negative lapse rate feedback by the positive water vapor feedback. Thus, the clear sky radiative feedbacks in AGCMs depend critically on the way in which cloud processes are treated, and are thus coupled to the cloud processes and associated radiative feedbacks.

Table 1 Qualitative effect of various changes in cloud characteristics that could occur as the climate changes[a]

Cloud property change	Effect on climate	Example reference
Decrease in amount of low cloud	Warming	Le Treut and Li (1991)
Increase in amount of high cloud	Warming[b]	Wetherald and Manabe (1988)
Increase in cloud height	Warming	Mitchell and Ingram (1992)
Increase in water content of stratus clouds	Cooling	Somerville and Remer (1984)
Increase in water content of cirrus clouds	Warming or cooling	Lohmann and Roeckner (1995)
Increase in ratio of water droplets to ice crystals	Cooling through enhanced reflectivity	Senior and Mitchell (1993)
	Warming through increased particle fall velocities	Senior and Mitchell (1993)

[a] Also given are references where the indicated effect is quantitatively assessed or included in climate simulations. The actual changes in cloud property that occur as the climate warms may be different from indicated here, and sometimes differ from model to model (from Harvey, 2000a).
[b] This assumes that high clouds have a net heating effect. However, Del Genio *et al.* (1996) calculate a net cooling effect for high tropical cirrus clouds.

Role of Cloud Feedbacks

Another important source of uncertainty concerning climate sensitivity is the role of cloud feedbacks. Table 1 lists the feedback effect of various cloud changes, if they were to occur in association with a warming of the climate. Clouds reflect solar radiation, which has a cooling effect. This effect depends on the difference between the cloud and surface albedo (reflectivity), and on the amount of incident solar radiation. The smaller the underlying albedo and the greater the incident solar radiation, the greater the increase in the reflection of solar radiation to space due to clouds. The cooling effect thus depends on where and when the clouds occur. Clouds also absorb infrared radiation emitted from the surface and re-emit their own radiation, but the amount re-emitted is smaller than the amount absorbed because the tops of clouds are colder than the underlying surface. The higher the cloud, the colder the cloud top and the greater the reduction in net infrared emission to space. High clouds thus tend to have a net warming effect because the reduction in infrared emission to space is greater than the extra reflection of solar radiation. Low clouds, in contrast, have a net cooling effect. Thus, an increase in the amount of high cloud as the climate warms usually serves as a positive feedback, while an increase in the amount of low cloud serves as a negative feedback. Similarly, an increase in the height of existing clouds, because this makes them colder and reduces the emission of radiation to space, has a warming effect on surface climate. Stratus clouds have an emissivity of 1.0 and so absorb all of the radiation emitted from the surface. An increase in the water content of these clouds increases the albedo but has no effect on infrared emission to space, so there is a net cooling effect. Cirrus clouds, on the other hand, have an emissivity less than 1.0. An increase in water content increases the albedo and the emissivity, so both the cooling and heating effects of the clouds increases. Thus, an increase in the water content of cirrus clouds can have a net warming or a net cooling effect. Ice crystals tend to be much larger than water droplets and so are less reflective, but they also have greater fall velocities. Thus, a transition from ice crystal to water droplet clouds makes clouds more reflective (a cooling effect) but presumably decreases the life-span of clouds which, in most cases, will be a warming effect.

The directions of the cloud changes listed Table 1 are given for illustrative purposes only. In some cases the direction of change that would occur in reality is unknown, as illustrated in the following examples: First, Senior and Mitchell (1993; Figure 18) show that the direction and magnitude of the changes in cloud fraction depend on the way in which cloud amounts are parameterized. Second, we are reasonably confident that convective clouds will rise to greater heights in a warmer climate, but the extent to which this occurs depends on the choice of convective parameterization scheme (Cunnington and Mitchell, 1990). Third, Sinha and Shine (1994) show that implementation of a feedback between temperature and ice water content in cirrus clouds changes the sign of the feedback due to increasing cloud height from positive to negative. For fixed cloud height and for clouds cold enough to remain as ice crystal clouds, warmer temperatures lead to a greater ice water content, which serves as a positive feedback. However, if cloud height increases, this feedback is greatly weakened because there is a smaller increase in cloud ice content due to the fact that the cloud hardly warms at all. Fourth, it was widely believed that the optical thickness of liquid water clouds would increase as climate warms, due to an increase in cloud droplet size (Somerville and Remer, 1984). Now it appears that concurrent changes in cloud thickness – which depend on changes in atmospheric vertical stability in regions of subsiding motion – can override the relationship between temperature and optical depth that one would otherwise expect (Del Genio, 1996). Thus, the direction of the feedback between climate and cloud optical thickness is also uncertain.

Table 2 lists the changes in cloud amount and height that occurred in response to an increase in CO_2 in a number of simulations with AGCMs or coupled atmosphere–ocean

Table 2 Changes in cloud amount and cloud height in response to a warming of the climate, as simulated by various AGCMs (from Harvey, 2000a)[a]

Model	References	Change in cloud amount			Mean cloud height	Type of experiment
		Total or low	Mid	High		
ARPEGE	Timbal *et al.* (1997)	Total and mid-level clouds decrease (with magnitude dependent on forcing, generally around 1–2%)		Positive except at ITCZ	Increased	Forced with $2 \times CO_2$ SST anomalies from ECHAM1 and UKMO models
BMRC	Colman and McAvaney (1995)	Low = +0.23%	−3.7%	−2.1%	High clouds occur higher	$2 \times CO_2$ equilibrium
BMRC	Colman and McAvaney (1997); Colman *et al.* (1997)	Total clouds = −0.45% for 0–2 K Δ SST		−1.2% for 0–2 K Δ SST Mainly in tropics	Increased by about 30 m	Perpetual July, ± 2 K SST
CCC	Boer *et al.* (1992)	Total = −2.2% (−1.9% winter, −2.4% summer)		Positive except between ±30°		$2 \times CO_2$ equilibrium
GFDL	Wetherald and Manabe (1988)	Cloud cover decreases in the upper troposphere at mid and low latitudes, except for an increase near the tropopause, especially at high latitudes. Low cloud cover increases at high latitudes				$2 \times CO_2$ equilibrium
GISS	Rind *et al.* (1995); Cess *et al.* (1996)	−2.3% global average; +0.1% polar		−0.3% global average; +0.4% polar		$2 \times CO_2$ equilibrium
LMD	Le Treut and Li (1991)	Cloud cover decreases almost everywhere at all heights, except that it increases near the tropopause at most latitudes				Perpetual July, ± 2 K SST
NCAR	Zhang *et al.* (1994)	−0.5%	−3.5%	−3.5%		Perpetual July, ± 2 K SST
NCAR	Washington and Meehl (1989, 1996)	Increase at high latitudes		Increase		Transient $2 \times CO_2$
UKMO	Murphy and Mitchell (1995)	NH = −0.2% SH = +1.1 Increase in Arctic	NH decrease	NH = +0.3% SH = −0.7%		Transient $2 \times CO_2$, values given after 75 years

[a] ARPEGE, Action De Recherche, Petite Echell Grande Echelle, Meteo France, Toulouse, France; CCC, Canadian Climate Centre, Victoria, British Columbia, Canada; LMD, Laboratoire De Meteorologie Dynamique, Paris, France; UKMO, United Kingdom Meteorological Office, Bracknell, UK (now the Hadley Centre for Climate Prediction and Research). See text for definitions of other acronyms.

(GCMs) atmosphere–ocean general circulation model (AOGCMs). There is a general tendency for cloudiness to decrease everywhere, except for low clouds at high latitudes and for clouds near the tropopause at most latitudes. There are, however, many exceptions.

In order to accurately simulate cloud feedbacks, at least four conditions need to be satisfied: (i) the relevant feedback processes must be incorporated in the cloud parameterization; (ii) these processes have to be incorporated in a way that mimics nature; (iii) changes in the

large-scale factors that influence clouds must also be correctly simulated, and this depends on the fidelity of all the other processes in the climate model; and (iv) the present-day distribution and characteristics of clouds must be accurately simulated. Because conditions (i) and (ii) must both be satisfied, the incorrect incorporation of a feedback process could result in worse results than if the feedback were neglected altogether. An example might be the incorporation of a strong temperature– optical depth feedback, when the true feedback is much weaker. Thus, more complex cloud parameterization schemes are not necessarily better than simpler schemes, unless the individual processes added to the complex scheme are separately validated against observations.

An analysis of cloud feedbacks published by Cess *et al.* (1996) indicates that the effect of cloud feedbacks in recent models ranges from a 27% reduction in climate sensitivity to a doubling of climate sensitivity, compared with the sensitivity expected based on the clear-sky radiative damping alone. In this analysis, a globally uniform increase in sea surface temperature (SST) was imposed and the changes in radiative fluxes with fixed and variable clouds were determined. From this, the radiative damping λ, and the contribution of clouds to λ, can be computed. If the calculated feedbacks are applicable to a CO_2 doubling perturbation, then the expected global average climatic response ranges from 1.3 to 3.3 °C.

However, the magnitude and even the sign of the net cloud feedback depends in part on the spatial *pattern* of temperature change. In the afore-mentioned analysis of cloud feedbacks published by Cess *et al.* (1996), a globally uniform increase in SST was imposed. In reality, the surface warming will vary with latitude and with longitude, and both variations can cause a markedly different cloud feedback than for the cause of globally uniform warming. With regard to east–west variations, the critical region appears to be the tropical Pacific Ocean. Today, the eastern Pacific is relatively cold due to the upwelling of deep water, while the western Pacific is relatively warm. This temperature contrast drives an east–west circulation known as the Walker circulation, with rising motion over the western Pacific and sinking motion over the eastern Pacific. In the GFDL AOGCM, greater warming occurs in the eastern than in the western tropical Pacific Ocean when CO_2 is increased, thereby reducing the temperature contrast and weakening the Walker circulation (Knutson *et al.*, 1997). In experiments in which a globally uniform change in SST is imposed, Del Genio *et al.* (1996) found that the global mean cloud feedback was slightly negative. However, when greater warming in the eastern than in the western Pacific Ocean was allowed, the cloud feedback was positive. This is because the weakening of the Walker circulation in the latter case resulted in less anvil cloud in the western Pacific, which otherwise has a cooling effect. The warming effect of less anvil cloud in the western Pacific Ocean was only partly offset by increased anvil cloud in the eastern Pacific Ocean. Another instance where the spatial patterns of temperature change could determine the sign of the cloud feedback is in the relative difference between Northern Hemisphere (NH) and Southern Hemisphere (SH) warming (Murphy and Mitchell, 1995). An additional factor complicating the picture is that both the east–west difference in warming in the tropical Pacific Ocean, and the inter-hemispheric difference in warming between the two hemispheres, could be quite different from the equilibrium patterns during the transition to a new climate (Harvey, 2000a). Thus, the sign and strength of the cloud feedback in some regions could very well change over time.

FEEDBACKS INVOLVING LAND SNOW COVER

A reduction in the extent of snow cover on land as climate warms is widely expected to serve as a positive feedback, contributing to a greater warming than in the global average in high latitude regions. This is because snow tends to have a higher albedo (reflectivity) than snow-free land, so a reduction in snow cover leads to an increase in the amount of solar radiation that is absorbed at the surface. This will lead to further warming and thus acts as a positive feedback. However, analyses by Cess *et al.* (1991) and Randall *et al.* (1994) indicate that a number of competing changes come into play when snow cover changes, and these changes can substantially weaken the positive snow cover feedback that would otherwise occur or even change the snow cover feedback into a weak negative feedback.

First, changes in snow cover can induce changes in the amount, location, or optical properties of clouds. In particular, cloudiness tends to increase over snow-free regions, so the reduction in albedo at the top of the atmosphere (the planetary albedo) is much smaller than the reduction in surface albedo. It is even possible for the planetary albedo to increase if the clouds are more reflective than the former snow cover. Second, once the snow melts away, the surface temperature is free to rise above 0 °C, so the emission of infrared radiation from the surface and thence to space can increase. This serves as a negative feedback. Third, the atmospheric lapse rate tends to increase after snow retreats (due to greater surface heating), and this serves as a positive feedback. Finally, the drying of the surface after the snow all melts reduces the water vapor content of the overlying atmosphere, which serves as a negative feedback. Overall, the shortwave feedback tends to be positive, while the longwave feedback tends to be negative. Either feedback can be stronger, so the net feedback due to retreat of snow cover can be positive or negative.

Cess *et al.* (1991) analyzed the snow cover feedback in 17 different AGCMs, and they found that the effect of

snow cover feedback ranged from a reduction in climate sensitivity by 10% to almost doubling the climate sensitivity, based on the calculated radiative damping for experiments in which the models continuously simulate the month of April (a perpetual April experiment). That is, the global mean temperature response can as much as double, with even larger changes (by up to a factor of 3–4 or more) in regions where snow retreats. However, snow cover feedbacks are expected to be strongest during transitional seasons, so the mean annual climate feedback in these models is certainly smaller than indicated by the perpetual April experiments.

Feedbacks Involving Sea Ice

Changes in the areal extent of sea ice can also exert a strong feedback effect on climate by changing the radiative balance. As well, changes in the area and thickness of sea ice alter the fluxes of latent and sensible heat between the ocean and atmosphere, which has a strong effect on the seasonality of temperature change at high latitudes. However, except through possible interactions with cloud amount, these changes will not significantly alter the planetary radiative balance and so will not noticeably affect the climate sensitivity.

The simulation of sea ice in global-scale climate models is a particularly challenging problem (Randall *et al.*, 1998). Sea ice involves vertical growth (at the base of the ice), lateral growth (into cracks or leads created by divergent ice motions), surface and basal melting, and lateral melting (from leads in summer). The surface albedo of sea ice depends on its thickness and the possible presence of snow cover or surface melt ponds; the way in which the surface albedo is computed generally has a dramatic effect on the simulation of sea ice for the present climate (Shine and Henderson-Sellers, 1985), and on the sensitivity of sea ice to changes in climate (Meehl and Washington, 1990). Ice motions, caused by surface winds and ocean currents, have a significant effect on the distribution of ice thickness within a given region, on the occurrence of leads and larger areas of open water, and on the spatial extent of sea ice. There are a number of important interactions between the thermodynamics of sea ice (processes related to freezing and melting) and the dynamics of sea ice (processes related to ice motions). However, only six of the 16 coupled climate models discussed in Gates *et al.* (1996) included sea ice dynamics; in the other 10, only sea ice thermodynamics is explicitly considered, with dynamical effects included only through the specification of a minimum lead fraction.

Pollard and Thompson (1994) found that, in simulations with ice dynamics, the ice was less compact for the control simulation than for the case without sea ice dynamics. Thus, the total area of sea ice was smaller, and because of this, the temperature–albedo feedback was weaker. This reduced the global mean equilibrium warming for a doubling of CO_2 from 2.27 to 2.06 °C, with much larger effects at high SH latitudes in winter and spring. Watterson *et al.* (1997) also found that the introduction of sea ice dynamics reduces the global mean climate sensitivity, in their case from 4.8 to 4.3 °C using the Commonwealth Scientific and Industrial Research Organization (CSIRO) AGCM. The reduced sensitivity with sea ice dynamics is consistent with the more general result that the climate model sensitivity tends to be smaller, the smaller the initial amount of sea ice, all else being equal. The weaker high-latitude feedback in the presence of ice dynamics implies a smaller polar amplification of the response, with important implications for the change in soil moisture in mid-latitudes (Harvey, 2000a; Chapter 10.2).

The simulated high-latitude climatic change and hence the global mean climate sensitivity are also quite sensitive to the initial ice thickness. The initial sea ice thickness in turn is very sensitive to the heat flux from the underlying water to the base of the ice, which is related to the poleward oceanic heat flux simulated for the present climate. Rind *et al.* (1995) found that, when the GISS model was adjusted so that the sea ice thickness (which were initially too great) matched observational estimates, the global mean warming for a CO_2 doubling increased from 4.17 to 4.78 °C. This was due to the fact that, when initial ice thicknesses in the model match those observed for the present climate, it is easier for the ice to retreat when the CO_2 concentration is increased.

INTERACTIVE EFFECT OF FAST FEEDBACK PROCESSES IN AGCMS COUPLED TO MIXED LAYER-ONLY OCEAN MODELS

The effect of any given feedback on the climate sensitivity to a radiative perturbation depends on what other feedbacks are already present (Harvey, 2000a; Chapter 3). This is because there is a synergistic or multiplicative effect between the different feedbacks. Thus, if there is a positive cloud feedback, this will increase the warming at high latitudes, thereby increasing the effect of the positive snow and ice feedback, which in turn will amplify the effect of the positive cloud feedback. These interactions can be simulated with climate models.

As previously noted, the strength of the cloud feedback depends in part on the spatial patterns of climatic change. These in turn depend on regional variations in the heat uptake or release by the oceans (during the transition from one climate to another) and in changes in ocean currents. Ideally, coupled AOGCMs should be used to compute the patterns of climatic change and from these, the feedback strengths and climate sensitivity. However, these models are not practical for evaluating the steady state (long-term) response to an increase of CO_2. This is because integrations on the order of 1000 years are required to reach close

to steady state, and this is prohibitively expensive. Since the feedback processes that determine climate sensitivity operate in the atmosphere or at the surface, an alternative is to couple the AGCM to an ocean model consisting of the mixed layer only. This permits steady state changes to be attained after only a few decades simulation. The disadvantage is that horizontal heat transport in the oceans, which depends on the 3D overturning circulation, cannot be simulated. Instead, the oceanic heat transports required to give the present day variation of temperature with latitude are prescribed. These are then held constant as the climate changes. Spatial variations in the warming can develop, in response to regional variations in the feedback strengths, but the spatial patterns in the new steady state climate would undoubtedly have been different in an AOGCM simulation due to changes in the horizontal transport of heat by the oceans. Not only would this affect regional patterns of climatic change, it would also likely affect cloud feedbacks and hence the global mean climate responsiveness to a doubling of atmospheric CO_2.

Table 3 lists the steady state, global mean change in surface air temperature obtained with coupled AGCM-mixed layer models for a doubling of atmospheric CO_2. In most cases, the global mean warming for a CO_2 doubling is 2.0–4.0 °C. This is consistent with the discussion of feedbacks presented above.

Table 3 Global mean change in surface air temperature (ΔT, K) for a doubling of atmospheric CO_2 as obtained by various AGCMs coupled to a slab ocean[a,b]

Model	ΔT	Reference
BMRC	2.1	Colman and McAvaney (1995)
Genesis[c]	2.1	Pollard and Thompson (1994)
	2.3	Pollard and Thompson (1994)
UKMO	2.5	Johns *et al.* (1997)
	2.8	Murphy and Mitchell (1995)
GLA	2.6[d]	Sellers *et al.* (1996)
	2.8[e]	Sellers *et al.* (1996)
CCC	3.5	Boer *et al.* (1992)
GFDL	4.0	Manabe *et al.* (1992)
NCAR	4.0	Washington and Meehl (1989)
	4.6	Washington and Meehl (1993)
CSIRO	4.3	Watterson *et al.* (1997)
GISS	4.2[f]	Rind *et al.* (1995)
	4.8[g]	Rind *et al.* (1995)

[a] Acronyms not defined below are defined in Table 2 (from Harvey, 2000a).
[b] CSIRO, Commonwealth Scientific and Industrial Research Organization, Aspendale, Australia; GLA, Goddard Laboratory for Atmospheres, Greenbelt, MD, USA.
[c] The Genesis model is an outgrowth of the NCAR AGCM.
[d] Standard model version.
[e] Version with reduced stomatal conductance in response to a higher atmospheric CO_2 concentration.
[f] Standard model version, having sea ice that is too thick for the present climate.
[g] Model version in which sea ice thickness for the present climate matches observational estimates.

DIRECT OBSERVATIONS OF THE WATER VAPOR FEEDBACK

The preceding discussion has highlighted the complexity of the feedback processes that determine climate sensitivity in nature, and of the difficulties and uncertainties associated with trying to simulate these processes – and their interactive effects – in computer climate models. The increase in the amount of water vapor in the atmosphere as climate warms is projected to be the single largest feedback, and to be strongly positive. Cloud feedbacks could also be important, but the net feedback could be either positive or negative. However, the importance of cloud (and surface albedo) feedbacks depend on the overall strength of the water vapor feedback, due to the synergistic effects mentioned above. The primary factors driving the overall increase in atmospheric water vapor are the global mean warming and its latitudinal distribution. The primary factors governing the vertical distribution of changes in water vapor at low latitudes are likely to be how the intensity and height of convection changes as the climate warms, and how the partitioning of convective moisture pumping between upper tropospheric heating and humidification changes as convection changes. At middle and high latitudes, the changes in the vertical moisture transport by large-scale eddies (i.e., storms) are likely to be the dominant control on the vertical distribution of changes in water vapor.

The amount of water vapor and its vertical distribution can be directly observed at present. This provides the possibility of directly inferring the strength of the feedback between temperature and water vapor from observations. There are three different kinds of observations that shed light on the water vapor feedback: the relationship between surface temperature and the total amount of water in the atmospheric column (the precipitable water); the relationship between surface temperature and the vertical distribution of water vapor; and the relationship between specific processes (such as deep convection) that influence water vapor and the greenhouse trapping (GHT) of infrared radiation.

Variations in the Total Amount of Water Vapor

There is no doubt that the total amount of water vapor in the atmosphere will increase as the climate warms. This follows from very fundamental principles involving the fact that the driving force for evaporation from the ocean – the difference between surface and atmospheric vapor pressures – tends to increase as the surface temperature increases, combined with the fact that the atmosphere s ability to hold water increases as the air warms. Observations made during recent decades indicate that the atmospheric water vapor content has indeed increased as the climate warmed (Harvey, 2000a; Chapter 5). The relationship between surface temperature and precipitable water in

the tropics is weak at the monthly and annual time scales, where concurrent changes in atmospheric dynamics complicate the picture, but is stronger at the decadal time scale (Gaffen *et al.*, 1992).

Interannual Variations in the Amount and Vertical Distribution of Water Vapor

Available balloon-based observations are not reliable for determining long-term trends in the amount of water vapor in the upper troposphere, while satellite observations have not been available for long enough to determine long-term trends. However, the available datasets are adequate for assessing the spatial distribution of inter-annual changes in water vapor and their correlation with inter-annual changes in surface temperature. Much of the inter-annual variability is related to the El Niño oscillation, which involves large-scale oscillations in surface temperatures in the tropical Pacific Ocean and a shift in the main region of convection between Indonesia and the eastern Pacific Ocean. Since the changes in surface temperature are associated with major shifts in the tropical circulation that are unique to the El Niño oscillation, the local correlation between temperature and humidity will not be applicable to long-term, large-scale climatic change unless long-term climatic change happens to involve some circulation changes.

Sun and Oort (1995) have attempted to circumvent this problem by examining the correlation between atmospheric water vapor and surface temperature averaged over the entire tropics (30 °S–30 °N). Using balloon-based data for the period May 1963 to December 1989, they found that the amount of water vapor in the tropical atmosphere increased with increasing temperature as follows:

- by 70–75% of the increase that would occur with fixed RH in the 1000–900 mb layer;
- by only 15% of the increase expected with fixed RH at 700 mb, and;
- by 70% of the increase expected with fixed RH at 300 mb.

In the planetary boundary layer (1000–900 mb), the change in water vapor is most directly tied to the thermodynamically-driven increase in the rate of evaporation. There is also a sizeable relative increase in the upper troposphere, related no doubt to an increase in the rate of detrainment of moisture from the tops of cumulus cloud columns. However, in the middle troposphere there is almost no increase in the amount of water vapor, which can be explained by an increase in the rate of subsidence of relatively dry air from the upper troposphere (even though this air becomes moister, it is still much drier than the air below).

Since AGCMs simulate close to constant RH at all heights as the climate warms, the results of Sun and Oort (1995) – assuming that they are applicable to long-term climatic change – imply that the water vapor feedback is too strong in AGCMs. This is turn could be due to the convective coupling between the moist surface layer and the middle troposphere being too strong, so that water vapor increases enough at all heights to keep RH constant as temperatures increase (Sun and Held, 1996). Sun and Oort (1995) compared the effect of their observed water vapor variation on the outgoing emission of infrared radiation with that assuming constant RH. Their results imply that the assumption of constant RH increases $|\mathrm{d}F/\mathrm{d}T|$ by about $0.8\,\mathrm{W\,m^{-2}\,K^{-1}}$ compared with the observed water vapor variations. Since this error is applicable only to cloud-free regions in the tropics, the error in the global mean $\mathrm{d}F/\mathrm{d}T$ is about $0.2\,\mathrm{W\,m^{-2}\,K^{-1}}$. This in turn implies that the total water vapor feedback strength as computed by AGCMs is about 10% too strong.

Direct Observations of Processes Governing Upper Tropospheric Water Vapor

An alternative to examining the correlation between changes in surface temperature and atmospheric water vapor is to examine the relationship between specific processes and atmospheric water vapor. In particular, one would like to know how the upper tropospheric humidity, averaged over the entire tropics, changes through time as the frequency of deep convection, averaged over the entire tropics, changes. Unfortunately, currently available data are not reliable enough to answer this question clearly.

An alternative (and more physically meaningful) approach is to directly examine the relationship between clear-sky GHT and SST, where GHT is simply the difference between the surface emission (σT^4) and the outgoing emission for clear skies at the top of the atmosphere (which can be easily measured by satellite). For the period April 1985 to December 1987, Soden (1997) found that the tropical-mean GHT increased in association with an increase in tropical-mean SST, as conditions changed from La Niña to El Niño. This is illustrated in Figure 1. GHT increased in regions where deep convection increased, and decreased elsewhere (due to the drying effect on the middle troposphere of an increase in subsidence). There were marked intra-annual variations in GHT that are not reflected in changes in the tropical-mean SST. This must be related to changes in the spatial patterns of SST, which are known to be as important to the occurrence of deep convection as absolute temperatures (Zhang *et al.*, 1994). Since it is not clear how the spatial patterns of SST will change as the climate warms or whether the patterns of temperature change will resemble El Niño patterns, the correlation seen over the course of a La Niña–El Niño transition might not apply to a warmer climate.

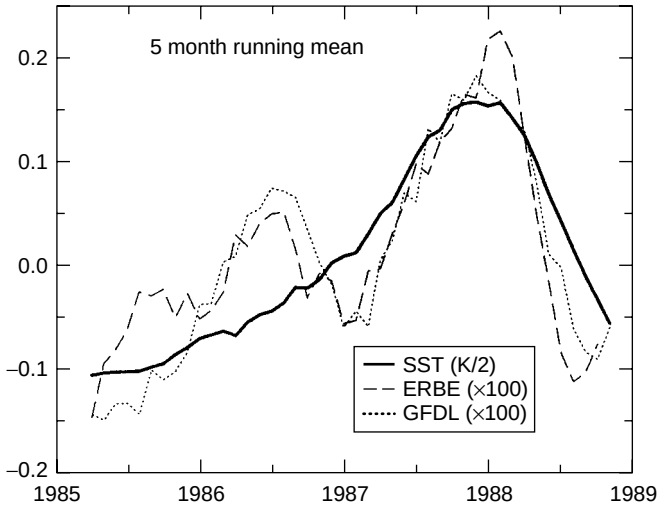


Figure 1 Variation (times 100) from 1985 to 1989 in the GHT ($W M^{-2}$) as measured by the Earth Radiation Budget Experiment (ERBE) satellite observations and as computed by the GFDL AGCM when driven by observed SST variations. Results are averaged over the entire region from 30°S to 30°N. Also given are the average SST variations divided by two. (Reproduced from Soden, 1997)

However, the GFDL AGCM has successfully simulated the overall variation in tropical-mean GHT from 1985 to 1989, in tropical-mean intra-annual fluctuations, and the local patterns of increasing and decreasing GHT when given the observed changes in SST. The tropical-mean variations are shown alongside the observed variations in Figure 1. This model, like all AGCMs, produces a positive long-term water vapor feedback. Thus, although observations of intra-annual, longer term, and spatial patterns of GHT do not directly tell us anything about the multi-decadal water vapor feedback, they can be used to test the fidelity of convective process embedded in AGCMs, the same processes that give rise to the long-term feedback in AGCMs. This test indicates that, given the correct patterns of SST change, current AGCMs should do a reasonable job in simulating the water vapor feedback in tropical regions.

Outside the tropics, Del Genio *et al.* (1994) find that, in AGCMs, synoptic-scale storms and the east–west averaged motions are the dominant factors influencing moisture in the upper troposphere. Although their analysis is based on model simulations rather than on nature (for which adequate observations are lacking), the processes in question are fully resolved by AGCMs. This is in contrast to convective processes, which must be parameterized because they occur at scales smaller than the model grid resolution. Thus, we can have more confidence in these results than in model results concerning the role of convection. The analysis of Del Genio *et al.* (1994) indicates that the moistening effect of storm systems and average motions would both increase as the climate warms.

Thus, the evidence based on observations in the tropics, where convection is an important process, and diagnoses of model behavior outside the tropics (where we have more confidence in the models) supports the conventional wisdom that the water vapor feedback is positive. Observations indicate that total precipitable water increases as the climate warms, which will exert a positive feedback. Observations of inter-annual variability indicate that the amount of water vapor also increases in the upper troposphere, although the increase in the middle troposphere is probably less than simulated by current AGCMs. Observations also show that the clear-sky GHT, averaged over the tropics, increases as tropical average SST increases during the course of the La Niña–El Niño cycle. Although this particular change might not be a valid analogue for long-term climatic changes, the GFDL AGCM does an excellent job in simulating this correlation. This serves as a validation of key water vapor processes in this AGCM, an AGCM that, like other models, simulates a significant positive water vapor feedback for long-term climatic change.

SLOW FEEDBACK PROCESSES

The above discussion has focused on the fast feedback processes that determine climate sensitivity, as traditionally defined. These feedback processes are relevant if one were to imagine a prescribed and fixed increase in greenhouse gas concentrations (other than water vapor), and if one wanted to know the eventual climatic response to the prescribed concentration changes. In reality, the change in climate would induce further changes in greenhouse gas concentrations, on a variety of time scales ranging from a decade to several centuries. The major potential climate–greenhouse gas concentration feedbacks involving CO_2 and CH_4 – the two most important gases directly affected by human activities – are outlined below. A more detailed discussion and extensive references can be found in Harvey (2000a; Chapter 8).

A warming of the climate would alter the uptake by or release of CO_2 from both the terrestrial biosphere and the oceans. First, a warmer climate will lead to changes in the rate of photosynthesis (generally an increase). Second, a warmer climate will generally lead to increased rates of respiration by plants and of soil organic matter by micro-organisms. Climate-induced changes in the rate of respiration of soil organic matter depend on the concurrent changes in soil moisture. If soils are saturated, a partial drying would facilitate greater decomposition (respiration) of soil organic matter, while if soil moisture is already limited, further drying could reduce rates of decomposition. Another potential terrestrial biosphere feedback could arise if climatic zones shift substantially faster than the rate at which forests can migrate

(as is likely under business-as-usual scenarios). As the ocean surface warms, the solubility of CO_2 in the surface layer will decrease, thereby reducing the ocean s ability to absorb anthropogenic CO_2. A change in the ocean circulation would alter the upwelling of nutrients to the surface layer, thereby altering the intensity of marine photosynthesis and the downward transfer of carbon.

A warming of the climate would alter the production of CH_4 from wetlands – the largest natural source – in two ways. First, it would change the productivity of wetland ecosystems and hence of the supply of organic matter, a portion of which is anaerobically decomposed and converted to CH_4. Second, a change in temperature would alter the rate of decomposition of the available organic carbon. A warmer climate would also increase the rate of removal of CH_4 from the atmosphere since, under a warmer climate, more hydroxyl radicals would be created through the splitting of water molecules by ultraviolet radiation, due to the greater water vapor content of a warmer atmosphere (reaction with hydroxyl is the main removal process for atmospheric CH_4). There is also the possibility that a warmer climate could provoke the decomposition and release of methane clathrates – ice–methane mixtures found in continental slope sediments worldwide and in permafrost regions on land. This latter possibility has been investigated by Harvey and Huang (1995) and found not to be a significant except for scenarios leading to very large climate change even in the absence of methane clathrate destabilization. The methane released in this way would be converted to CO_2 on a time scale of 10 years, which would then be gradually removed from the atmosphere in the same way as direct anthropogenic emissions (*see* Methane Clathrates, Volume 1).

See also: Earth System Processes, Volume 1; Feedbacks, Chemistry - Climate Interactions, Volume 1.

REFERENCES

Boer, G J, McFarlane, N A, and Lazare, M (1992) Greenhouse Gas-induced Climate Change Simulated with the CCC Second-generation General Circulation Model, *J. Clim.*, 5, 1045–1077.

Cess, R D, Potter, G L, and Blanchett, J P (1990) Intercomparison and Interpretation of Climate Feedback Processes in Nineteen Atmospheric General Circulation Models, *J. Geophys. Res.*, 95, 16 601–16 615.

Cess, R D, Potter, G L, and Zhang, M H (1991) Interpretation of Snow-climate Feedback as Produced by 17 General Circulation Models, *Science*, 253, 888–892.

Cess, R D, Zhang, M H, and Ingram, W J (1996) Cloud Feedback in Atmospheric General Circulation Models: an Update, *J. Geophys. Res.*, 101, 12 791–12 794.

Colman, R A and McAvaney, B J (1995) The Sensitivity of the Climate Response of an Atmospheric General Circulation Model to Changes in Convective Parameterization and Horizontal Resolution, *J. Geophys. Res.*, 100, 3155–3172.

Colman, R A and McAvaney, B J (1997) A Study of General Circulation Model Climate Feedbacks Determined from Perturbed Sea Surface Temperature Experiments, *J. Geophys. Res.*, 102, 19 383–19 402.

Colman, R A, Power, S B, and McAvaney, B J (1997) Non-linear Climate Feedback Analysis in an Atmospheric General Circulation Model, *Clim. Dyn.*, 13, 717–731.

Cunnington, W M and Mitchell, J F B (1990) On the Dependence of Climate Sensitivity on Convective Parameterization, *Clim. Dyn.*, 4, 85–93.

Del Genio, A D (1996) GCM Implications for Mechanisms Determining Cloud and Water Vapor Feedbacks, in *Climate Sensitivity to Radiative Perturbations: Physical Mechanisms and their Validation*, ed H Le Treut, NATO ASI Series, Springer-Verlag, Berlin, 107–125, Vol. I 34.

Del Genio, A D, Kovari, W, and Yao, M-S (1994) Climatic Implications of the Seasonal Variation of Upper Tropospheric Water Vapor, *Geophys. Res. Lett.*, 21, 2701–2704.

Del Genio, A D, Lacis, A A, and Ruedy, R A (1991) Simulations of the Effect of a Warmer Climate on Atmospheric Humidity, *Nature*, 351, 382–385.

Del Genio, A D, Yao, M-S, Kovari, W, and Lo, K K-W (1996) A Prognostic Cloud Water Parameterization for Global Climate Models, *J. Clim.*, 9, 270–304.

Gaffen, D J, Elliott, W P, and Robock, A (1992) Relationships Between Tropospheric Water Vapor and Surface Temperature as Observed by Radiosondes, *Geophys. Res. Lett.*, 19, 1839–1842.

Gates, W L, Henderson-Sellers, A, Boer, G, Folland, C K, Kitoh, A, McAvaney, B J, Semazzi, F, Smith, N, Weaver, A J, and Zeng, Q-C (1996) Climate Models – Evaluation, in *Climate Change 1995: The Science of Climate Change*, eds J T Houghton, L G F Filho, B A Callander, N Harris, A Kattenberg, and K Maskell, Cambridge University Press, Cambridge, 229–284.

Harvey, L D D (2000a) *Global Warming: The Hard Science*, Prentice Hall, Harlow, UK, 336.

Harvey, L D D (2000b) An Assessment of the Potential Impact of a Downward Shift of Tropospheric Water Vapor on Climate Sensitivity, *Clim. Dyn.*, 16, 491–500.

Harvey, L D D and Huang, Z (1995) An Evaluation of the Potential Impact of Methane-clathrate Destabilization on Future Global Warming, *J. Geophys. Res.*, 100, 2905–2926.

Johns, T C, Carnell, R E, Crossley, J F, Gregory, J M, Mitchell, J F B, Senior, C A, Tett, S F B, and Wood, R A (1997) The Second Hadley Centre Coupled Ocean–Atmosphere GCM: Model Description, Spinup and Validation, *Clim. Dyn.*, 13, 103–134.

Knutson, T R, Manabe, S, and Gu, D (1997) Simulated ENSO in a Global Coupled Ocean–Atmosphere Model: Multidecadal Amplitude Modulation and CO_2 Sensitivity, *J. Clim.*, 10, 138–161.

Le Treut, H and Li, Z-X (1991) Sensitivity of an Atmospheric General Circulation Model to Prescribed SST Changes: Feedback Effects Associated with the Simulation of Cloud Optical Properties, *Clim. Dyn.*, 5, 175–187.

Lohmann, U and Roeckner, E (1995) Influence of Cirrus Cloud Radiative Forcing on Climate and Climate Sensitivity in a General Circulation Model, *J. Geophys. Res.*, 100, 16 305–16 323.

Manabe, S, Spelman, M J, and Stouffer, R J (1992) Transient Responses of a Coupled Ocean–Atmosphere Model to Gradual Changes of Atmospheric CO_2. Part II: Seasonal Response, *J. Clim.*, 5, 105–126.

Meehl, G A and Washington, W M (1990) CO_2 Climate Sensitivity and Snow-sea-ice Albedo Parameterization in an Atmospheric GCM Coupled to a Mixed-layer Ocean Model, *Clim. Change*, 16, 283–306.

Mitchell, J F B and Ingram, W J (1992) Carbon Dioxide and Climate: Mechanisms of Changes in Cloud, *J. Clim.*, 5, 5–21.

Murphy, J M and Mitchell, J F B (1995) Transient Response of the Hadley Centre Coupled Ocean–Atmosphere Model to Increasing Carbon Dioxide. Part II: Spatial and Temporal Structure of Response, *J. Clim.*, 8, 57–80.

Pollard, D and Thompson, S L (1994) Sea-ice Dynamics and CO_2 Sensitivity in a Global Climate Model, *Atmosphere-Ocean*, 32, 449–467.

Randall, D A, Cess, R D, and Blanchet, J P (1994) Analysis of Snow Feedbacks in 14 General Circulation Models, *J. Geophys. Res.*, 99, 20 757–20 771.

Randall, D A, Curry, J, and Battisti, D (1998) Status of and Outlook for Large-scale Modelling of Atmosphere-ice-ocean Interactions in the Arctic, *Bull. Am. Meteorol. Soc.*, 79, 197–219.

Rind, D, Healy, R, Parkinson, C, and Martinson, D (1995) The Role of Sea Ice in $2 \times CO_2$ Climate Model Sensitivity. Part I: The Total Influence of Sea Ice Thickness and Extent, *J. Clim.*, 8, 449–463.

Sellers, P J, Randall, D A, Collatz, G J, Berry, J, Field, C, Dazlich, D A, Zhang, C, and Bounoua, L (1996) A Revised Land Surface Parameterization (SiB2) for Atmospheric GCMs. Part I: Model Formulation, *J. Clim.*, 9, 676–705.

Senior, C A and Mitchell, J F B (1993) Carbon Dioxide and Climate: The Impact of Cloud Parameterization, *J. Clim.*, 6, 393–418.

Shine, K P and Henderson-Sellers, A (1985) The Sensitivity of a Thermodynamic Sea Ice Model to Changes in Surface Albedo Parameterization, *J. Geophys. Res.*, 90, 2243–2250.

Sinha, A and Shine, K P (1994) A One-dimensional Study of Possible Cirrus Cloud Feedbacks, *J. Clim.*, 7, 158–173.

Soden, B J (1997) Variations in the Tropical Greenhouse Effect During El Niño, *J. Clim.*, 10, 1050–1055.

Somerville, R and Remer, L (1984) Cloud Optical Thickness Feedbacks in the C_2 Climate Problem, *J. Geophys. Res.*, 89, 9668–9672.

Sun, D-Z and Held, I M (1996) A Comparison of Modeled and Observed Relationships Between Interannual Variations of Water Vapor and Temperature, *J. Clim.*, 9, 665–675.

Sun, D-Z and Lindzen, R S (1993) Distribution of Tropical Tropospheric Water Vapor, *J. Atmos. Sci.*, 50, 1643–1660.

Sun, D Z and Oort, A H (1995) Humidity-temperature Relationships in the Tropical Troposphere, *J. Clim.*, 8, 1974–1987.

Timbal, B, Mahfouf, J-F, Royer, J-F, Cubasch, U, and Murphy, J M (1997) Comparison Between Doubled CO_2 Time-slice and Coupled Experiments, *J. Clim.*, 10, 1463–1469.

Washington, W M and Meehl, G A (1989) Climate Sensitivity Due to Increased CO_2: Experiments with a Coupled Atmosphere and Ocean General Circulation Model, *Clim. Dyn.*, 4, 1–38.

Washington, W M and Meehl, G A (1993) Greenhouse Sensitivity Experiments with Penetrative Cumulus Convection and Tropical Cirrus Albedo Feedbacks, *Clim. Dyn.*, 8, 211–223.

Washington, W M and Meehl, G A (1996) High-latitude Climate Change in a Global Coupled Ocean-atmosphere-sea ice Model with Increased CO_2, *J. Geophys. Res.*, 101, 12 795–12 801.

Watterson, I G (1997) The Diurnal Cycle of Surface Air Temperature in Simulated Present and Doubled CO_2 Climates, *Clim. Dyn.*, 13, 533–545.

Watterson, I G, O Farrell, S P, and Dix, M R (1997) Energy and Water Transport in Climates Simulated by a General Circulation Model that Includes Dynamic Sea Ice, *J. Geophys. Res.*, 102, 11 027–11 037.

Wetherald, R T and Manabe, S (1988) Cloud Feedback Processes in a General Circulation Model, *J. Atmos. Sci.*, 45, 1397–1415.

Wetherald, R T and Manabe, S (1995) The Mechanisms of Summer Dryness Induced by Greenhouse Warming, *J. Clim.*, 8, 3096–3108.

Zhang, M H, Hack, J J, Kiehl, J T, and Cess, R D (1994) Diagnostic Study of Climate Feedback Processes in Atmospheric General Circulation Models, *J. Geophys. Res.*, 99, 5525–5537.

Climate Feedbacks, Chemical Interactions

see Feedbacks, Chemistry – Climate Interactions (**Volume 1**); Earth System Processes (*Opening essay*, **Volume 1**)

Climate History

see Earth System History (*Opening essay*, **Volume 1**)

Climate, Human Effects

see Model Simulations of Present and Historical Climates (*Opening essay*, **Volume 1**); Projection of Future Changes in Climate (*Opening essay*, **Volume 1**)

Climate, Land Cover Influences

see Land Cover and Climate (**Volume 1**)

Climate Model Simulations of the Geological Past

André Berger

Institut d'Astronomie et de Géophysique G. Lemaître, Louvain-la-Neuve, Belgium

The accumulating amount of information on the Earth's climatic history can be greatly augmented by carrying out climate model simulations of periods in the geological past. Such model simulations have been extensively used to better understand the in uence on the climate of tectonic uplift, movement of the continents, and changes in atmospheric composition, land-surface properties, the intensity of solar radiation, and the seasonal and latitudinal distribution of solar radiation caused by cyclic variations in the shape and timing of the orbit of the Earth around the Sun. By accounting for changes in these factors, major changes in the Earth's climate history can be simulated with some success. During the early life of the planet, high levels of atmospheric carbon dioxide are believed to have kept the Earth's temperatures reasonable even with lower luminosity of the Sun. Over the last one billion years, changes in the locations of continents caused by plate tectonics have modulated the global climate and were likely the origin of the late Ordovician and Permo-Carboniferous Ice Ages. The very warm climate of the Cretaceous (~100 Myr ago) and the Miocene (~17 Myr ago) were likely due to very high concentrations of greenhouse gases and large poleward heat transport by the oceans. The onset of the Quaternary Ice Age about three million years ago is most probably due to the uplift of the Tibetan Plateau, a decrease of the carbon dioxide level, and some important changes in continental con guration. The glacial- interglacial cycles during this Ice Age result from cyclic changes in the Earth's orbit around the Sun; and the more rapid changes of the Late Glacial period are likely to be associated with complex interactions between the ice sheets, the ocean circulation and the hydrological cycle.

INTRODUCTION

The Earth s climate has always been changing and will undoubtedly continue to change. The Earth s history indicates that quite large variations in climate can occur (*see* Earth System History, Volume 1). For example, when the dinosaurs dominated the planet during the Mesozoic (250–65 Myr BP), the climate was very warm with tropical-like conditions in high latitudes. There is extensive evidence from the Mid-Cretaceous (120–90 Myr BP) indicating that the Earth was ice-free with a significant poleward displacement of many types of warm climate flora and fauna.

However, at the time of the last glacial maximum (LGM) 20 kyr ago, the climate was very different, with surface temperature likely being about 5 °C lower than at present. At this time, sea level was about 110 m below the present-day level, the North American and Eurasian ice sheets were much larger (reaching about 45×10^6 km^3 of ice), and the carbon dioxide concentration in the atmosphere was only about 200 parts per million by volume (ppmv) (against 367 in 1999).

In the future, climate will surely continue to change, driven by natural causes such as fluctuations in the Earth s orbit (*see* Orbital Variations, Volume 1) and in solar activity (Rind *et al.*, 1999; Bertrand *et al.*, 1999). But future climate will also be influenced by human activities (Houghton *et al.*, 1996), which even might interfere with the natural evolution of the climate (Loutre and Berger, 2000).

Observations indicate that climate has fluctuated at many different time scales, ranging from seasonal to the tens of millions of years associated with the tectonic evolution of the Earth. New observational techniques, accurate dating methods and comprehensive program of climate reconstruction are leading to an improved understanding of the past evolution of the atmosphere and the oceans, the shift of continents, the waxing and waning of the ice sheets, and the growth and retreat of forests and deserts.

The study of the past climate of the Earth provides a useful time perspective for viewing the present climate and projecting future climate change. The first goal of paleoclimate research is to describe variations in climate that occur over intervals well before the availability of instrumental records. The ultimate objective is to identify the physical causes of past climate variations and to understand the complex behaviour of the whole climate system. This is a key problem because past climatic conditions are not thought any more to be good analogs of future climate changes, future climate forcing and boundary conditions going likely to be largely different from those of the past. Finally, the need to verify models of climate change has stimulated study of the Earth s history in order to reap the potential benefits of testing models of differing complexity at many different time scales and conditions.

MODELING PAST CLIMATES WITH GENERAL CIRCULATION MODELS

General Circulation Models

Paleoclimates are being studied using a wide range of types of models (*see* Models of the Earth System, Volume 1). Climate models are frequently divided into three very broad categories:

1. General circulation models (GCMs);
2. Statistical–dynamical models that include a two-dimensional (2D) representation of the Earth (either latitude–vertical or latitude–longitude) and;
3. Thermodynamic and energy balance models.

Three-dimensional (3D) GCMs include, in the most explicit form, processes that depend on details of the atmospheric flow. They are primarily generally used for simulating geographic features of paleoclimates. However, because of the large computational costs that are involved, coupled ocean–atmosphere GCMs are generally used for time spans not exceeding a few thousands of years (Von Storch *et al.*, 1997). Most of the time these models are only able to provide snapshot views of the climate in equilibrium with specific boundary conditions. These boundary conditions include the seasonal and latitudinal changes of solar radiation at the top of the atmosphere (resulting from variations in the Earth s orbit about the Sun and in its axis of rotation) as well as variations in sea-surface temperature (SST), in major ice sheets, in atmospheric composition and opacity, in land-surface properties, and in location of continents for simulations distant enough in the past.

Most of these modeling experiments have focused on extreme climates of the Quaternary and Cretaceous periods.

Cretaceous (*see* Cretaceous, Volume 1)

Early modeling studies of the ice-free climates of the mid-Cretaceous focused on understanding the effects of the significant changes in continental positions that occurred during this time (the break up of the supercontinent of Pangaea). To explain the global average temperature during the mid-Cretaceous, including estimated Cretaceous global temperatures 6–14 °C above the present temperature and very high temperatures in polar regions, it is however necessary to increase the poleward heat transport by the ocean and to invoke higher atmospheric CO_2 levels (up to 6–8 times the present level, Barron *et al.*, 1993) which might have resulted from increased sea floor spreading rates and volcanism. Sellwood *et al.* (1994) have however, challenged the evidence for such Cretaceous warmth, which further questions the role of atmospheric carbon dioxide in determining Cretaceous climate.

Pliocene

Although cooler than the Cretaceous, the Pliocene (about 5–2 Myr BP) may be the best example of a climate that was significantly warmer than now over large areas. In addition to a CO_2 concentration that was about twice its present-day value, changes in oceanic heat transport seem to have been needed to explain the magnitude of high latitude warming in the winter, especially for the last 40 Myr. Recent evidence suggests indeed that warmer temperatures are not necessarily always linked to higher CO_2 levels and vice versa. Pagani *et al.* (1999), for example, presented evidence, later contested, for surprisingly low CO_2 levels of about 180–290 ppmv through the warm Miocene (17–12 Myr BP), invoking an increase of other greenhouse gas concentrations (like water vapor and methane) or changes in continental geography in order to explain the climate of those times.

Entering the Quaternary Ice Age (*see* Quaternary, Volume 1)

There is strong evidence that the Quaternary Ice Age began to develop in the Northern Hemisphere around 2.7 Myr BP. Simulations in the Eurasia sector (Kutzbach *et al.*, 1993) show that surface uplift of the Tibetan Plateau produced marked increases in the intensity of the summer and winter monsoons over Asia and a general cooling in northern polar latitudes. Thus, the Tibetan uplift could therefore have contributed significantly to the late Cenozoic climate change. However, some other mechanisms are needed to explain the full range of the observed trend toward the Quaternary Ice Age. These might include a decrease of atmospheric carbon dioxide, which could result from increased global chemical weathering and erosion, and a change of continental configurations with openings and closings of oceanic gateways and consequent impacts on ocean heat transport.

Last Interglacial (*see* Eemian, Volume 1)

During the last interglacial, 125 kyr ago (called the *Eemian interglacial*), the atmospheric CO_2 concentration was above pre-industrial levels, and sea level was somewhat higher than now. The Greenland ice sheet was perhaps smaller and orbital parameters favoured greatly enhanced Northern Hemisphere seasonality. In agreement with geological reconstructions, modeling experiments suggest warmer conditions, especially in high latitudes. Reduced sea-ice extent, enhanced northern tropical monsoons, and poleward displacement of the tundra and ta ga biomes (Harrison *et al.*, 1995) are other indications of the sign of the change. This warm interglacial ended around 115 kyr ago with a major transition to cooler conditions. Although orbital parameters induced a strong cooling at these times, it does not seem to be sufficient to initiate glaciation when other components of the climate system are kept as present. Actually, it is the southward migration of the boreal forest/tundra limit which helps to create favourable conditions for continental ice-sheet growth (de Noblet *et al.*, 1996). These results support the fundamental role that biogeophysical feedbacks play in initiating glaciation: if the climates were to cool initially and the boreal forests were to retreat, then feedbacks, associated in the model with albedo differences between forest

and bare ground-tundra during the season of snow cover, might cause additional cooling.

Abrupt Climatic Changes of the Late Glacial

Deep-sea cores from the Atlantic (Heinrich, 1988) and ice cores from the Greenland ice cap (Dansgaard *et al.*, 1993) suggest that many major abrupt century-scale warm–cold oscillations have occurred through the Late Glacial over the last 60 kyr (*see* Climate Change, Abrupt, Volume 1). The so-called *Dansgaard- Oeschger events*, clearly recognized in both types of cores, represent a climate system oscillating between significantly different states (*see* Dansgaard- Oescheger Cycles, Volume 1). These events, apparently related to reorganization of the ocean–atmosphere system, also seem to group into larger oscillations, each terminating in a massive discharge of icebergs into the North Atlantic. These discharge events, known as *Heinrich events* (*see* Heinrich (H-) Events, Volume 1), appear to be related to instabilities in the glacial ice sheets, particularly the North American Laurentide ice sheet, and result in abrupt changes in oceanic circulation and North Atlantic climate.

Modeling experiments seeking to understand these abrupt climatic changes have so far been carried out using low-order dynamical systems models. These studies have focused on simulating the Dansgaard–Oeschger oscillation (Sakai and Peltier, 1999) and Heinrich events (Verbitsky and Saltzman, 1994) and their relation to the thermohaline circulation (Stocker, 1999). These model experiments confirm that abrupt climate changes appear to be caused by changes in the freshwater balance of the ocean s sea surface, but suggest also that the response of the ocean circulation depends on the amplitude and temporal evolution of the perturbation. The coupled ocean–atmosphere model developed at the National Oceanic Atmospheric Administration s Geophysical Fluid Dynamics Laboratory develops a reversal in the normal thermohaline circulation (i.e., with no deep water formation in the North Atlantic Ocean) resulting from a massive discharge of freshwater into the high North Atlantic latitudes. The simulation shows that the ocean circulation transforms back to the original direct thermohaline circulation as soon as the fresh water discharge into the North Atlantic Ocean is terminated (Manabe and Stouffer, 1999). Moreover, this inactive mode of the reverse thermohaline circulation is not a stable equilibrium for a large coefficient of vertical subgrid scale eddie mixing.

The *Younger Dryas* event is another example of a rapid climatic change that occurred during the Late Glacial period (*see* Younger Dryas, Volume 1). The Younger Dryas terminated the relative warmth of the Bolling/Allerod period in the middle of the last deglaciation. It lasted about 1300 years before there was a rapid change from glacial to interglacial conditions at about 11.5 kyr BP. The Younger Dryas event could have been triggered by rapid insertions of glacial ice or melt water into the North Atlantic, which might have induced changes in the thermohaline circulation. However, based upon paleoceanographic evidence, Manabe and Stouffer (1999) suggested that the stable state of the reverse thermohaline circulation mentioned above, did not prevail during the cold periods of the Younger Dryas. Instead it is likely that the thermohaline weakened temporarily, but reintensified before it fully reached the state of the reverse phase.

The Last Glacier Maximum (LGM)

Several GCMs have been used to simulate the climate of the LGM (*see* Last Glacial Maximum, Volume 1). These simulations have helped to clarify the relative roles of changes in continental ice sheets, CO_2 concentration, sea ice, ocean temperature, and land albedo in producing major changes in atmospheric circulation, surface temperature, and precipitation patterns. An early model simulation by Broccoli and Manabe (1987) demonstrated the sensitivity of glacial-age climate simulations to the lowered level of the glacial-age atmospheric CO_2 concentration. The movement of wind blown dust and the cycling of water isotopes have also been simulated by models with tracer capabilities (Jouzel *et al.*, 2000).

Changes in ocean circulation associated with the LGM have been simulated using dynamical ocean models forced by surface wind, temperature and salinity information from atmospheric models. Using this procedure, Lautenschlager *et al.* (1992) found a much reduced thermohaline circulation in the Atlantic and an increased one in the Pacific during this extreme phase of the Late Glacial.

Mid-Holocene (*see* Holocene: Climate Changes and Society, Volume 3)

Experiments with climate models show that a significant fraction of the widespread warmth and enhanced northern summer monsoon circulation of the mid-Holocene can be simulated correctly by including the seasonal and latitudinal changes of solar radiation (Kutzbach, 1999). The northward shifts of the main regions of monsoon precipitation over Africa and India are also well reproduced in the 18 simulations conducted for the Paleoclimate Modeling Intercomparison Project (PMIP) (Joussaume *et al.*, 1999). However, all the models underestimate the magnitude of the monsoon increase over northern Africa.

This increase in precipitation in a warm climate changed the vegetation cover significantly. Paleobotanic evidence indicates that during the early to middle Holocene, boreal forests extended north of the modern treeline. Foley *et al.* (1994) demonstrated that this northward expansion of forest itself, triggered by changes in the Earth s orbit, likely

amplified the initial warming; however, this vegetation-climate interaction needed to be tested using coupled atmosphere-vegetation models (Claussen, 1994). These model simulations also confirmed an earlier suggestion that the vegetation snow-albedo feedback is positive. The importance of the synergism between terrestrial and marine feedbacks has already been demonstrated by Ganopolski *et al.* (1998a) using a lower order coupled atmosphere–ocean-vegetation model.

The same biospheric feedback also operates in subtropical latitudes. Climate reconstructions indicate that North Africa was much greener in the mid-Holocene than today (Jolly *et al.*, 1998). This was more or less well simulated by Claussen and Gayler (1997), who found a strong feedback between vegetation and precipitation due to an interaction between high albedo of Saharan sand deserts and atmospheric circulation. In Ganopolski *et al.* (1998a), the biospheric feedback dominates and its synergism with an increase in monsoon precipitation owing to increased SST results in only a little further amplification (see also Berger, 1999). It also appears that multiple equilibria exist in the atmosphere-vegetation system, mainly in the subtropical areas of North Africa. The existence of these would be important as they could explain the abrupt transitions in vegetation structure, such as occurred at the end of the mid-Holocene wet phase in North Africa around 5 kyr BP (Claussen *et al.*, 1999).

TRANSIENT SIMULATIONS

Due to the length of the needed simulations, the time-dependent behavior of the fully coupled climate system over geological time scales has only been able to be simulated with simplified models. These models incorporate the slow-response climate processes involving changes in ice volume, bedrock depression, deep-ocean temperature, and the atmospheric concentrations of greenhouse gases (Saltzman, 1990).

Most of the simulations with such models have examined Quaternary climatic variations. Some studies have focused on understanding the forced response of the coupled climate system to changes of astronomical parameters, while other have demonstrated that the coupled systems themselves exhibit free oscillatory behavior. Many long, time-dependent integrations have been performed with such models of intermediate complexity (Berger, 1995; Berger and Loutre, 1998; Ganopolski *et al.*, 1998b; Peltier and Marshall, 1995).

In particular, the Louvain-la-Neuve (LLN) 2D model is an altitude–latitude model that links the Northern Hemisphere atmosphere, ocean mixed layer, sea-ice, ice sheets and continents. In each latitudinal belt, the surface is divided into at most seven oceanic or continental surface types, each of which interacts separately with the subsurface and the atmosphere. The oceanic surface can be ice-free or sea-ice covered, while the continental surfaces can be covered by snow and includes Northern Hemisphere ice sheets. This model is actually able to simulate long-term variations of the Northern Hemisphere ice volume using either a constant CO_2 value or the CO_2 record reconstructed from the Vostok ice core (Jouzel *et al.*, 1993). The model also simulates successfully the initiation of glaciation at around 2.75 Myr BP, the late Pliocene–early Pleistocene 41 kyr cycle, the emergence of the 100 kyr cycle around 900 kyr BP, the glacial–interglacial cycles of the middle and late Pleistocene and the importance of CO_2 in shaping the climate of isotopic stages 11 and 1 (our present Holocene interglacial).

This model confirms the hypothesis of Hays *et al.* (1976) that the orbital forcing acts as a pacemaker for the glacial–interglacial cycles. However, changes in the CO_2 concentration are also shown to play an important role in the amplification of the climatic response to orbital forcing. The other important processes that need to be accounted for in the model simulations are albedo-temperature feedback (in particular related to the snow covered ta ga versus tundra in high polar latitudes), water vapor–temperature feedback, altitude and continental effects on snowfall over the ice sheets, the snow aging process, and isostatic rebound.

Due to its ability to simulate the past climate, the LLN 2D model has also been used to project the climate of the next 130 kyr making different assumptions about the chemical composition of the atmosphere (Loutre and Berger, 2000). Most of the natural scenarios indicate that the present interglacial is likely to last several tens of thousands of years, which is significantly longer than previous interglacials. In addition, according to the modeling experiments and as a result of human activities, the Greenland ice sheet might start to melt significantly within a few thousands of years, its return to a natural behavior not being expected before 35 000 years from now.

CONCLUSIONS

In summary, many of these modeling experiments indicate an acceptable agreement between their results and geological reconstructions. This validation of models has increased confidence in scientific understanding of the behavior of the climate system in the model simulations of the changes of the climate under the forcing of human activities.

REFERENCES

Barron, E J, Fawcett, P J, Pollard, D, and Thompson, S (1993) Model Simulations of Cretaceous Climates: The Role of Geography and Carbon Dioxide, *Philos. Trans. R. Soc. London B*, 341, 307–316.

Berger, A (1995) Modelling the Response of the Climate System to Astronomical Forcing, in *Future Climates of the World, a Modelling Perspective*, Vol. 16, ed A Henderson-Sellers, World Survey of Climatology, ed H E Landsberg, Elsevier, Amsterdam, 21–69.

Berger, A and Loutre, M F (1998) Climate Model in an Astronomical Perspective, in *Do We Understand Global Climate Change?* NTVA Report 2, Norwegian Academy of Technological Sciences, Trondheim, 149–184.

Berger, A (1999) The Role of CO_2, Sea-level and Vegetation During the Milankovitch Forced Glacial–interglacial Cycles, Workshop on Geosphere-Biosphere Interactions and Climate, ed L Bengtsson, Pontifical Academy of Sciences, Vatican City.

Berger, A (2000) Orbital Variations, in *Encyclopedia of Global Environmental Change*, Vol. 1, eds T Munn, M MacCracken, and J S Perry, John Wiley & Sons, New York.

Bertrand, C, van Ypersele, J P, and Berger, A (1999) Volcanic and Solar Impacts on Climate Since 1700, *Clim. Dyn.*, 15(5), 355–367.

Broccoli, A J and Manabe, S (1987) The Influence of Continental Ice, Atmospheric CO_2, and Land Albedo on the Climate of the Last Glacial Maximum, *Clim. Dyn.*, 1(2), 87–100.

Claussen, M (1994) On Coupling Global Biome Models with Climate Models, *Clim. Res.*, 4, 203–221.

Claussen, M and Gayler, V (1997) The Greening of Sahara During the Mid-Holocene: Results of an Interactive Atmosphere-biome Model, *Global Ecol. Biogeogr. Lett.*, 6, 369–377.

Claussen, M, Kubatzki, C, Brovkin, V, Ganopolski, A, Hoelzmann, P, and Pachur, H J (1999) Simulation of an Abrupt Change in Saharan Vegetation in the Mid-Holocene, *Geophys. Res. Lett.*, 26(14), 2037–2040.

Dansgaard, W, Johnsen, S J, Clausen, H B, Dahl-Jensen, D, Gundestrup, N S, Hammer, C U, Hvldborg, C S, Steffensen, J P, Svelnbjornsdottir, A E, Jouzel, J, and Bend, G (1993) Evidence for General Instability of Past Climate from a 250 kyr Ice-core Record, *Nature*, 364, 218–220.

De Noblet, N I, Prentice, I C, Joussaume, S, Texier, D, Botta, A, and Haxeltine, A (1996) Possible Role of Atmosphere–Biosphere Interactions in Triggering the Last Glaciation, *Geophys. Res. Lett.*, 23(22), 3191–3194.

Foley, J A, Kutzbach, J E, Coe, M T, and Levis, S (1994) Feedbacks Between Climate and Boreal Forests during the Holocene Epoch, *Nature*, 371(6492), 52–54.

Ganopolski, A, Kubatzki, C, Claussen, M, Brovkin, V, and Petoukhov, V (1998a) The Influence of Vegetation–Atmosphere–Ocean Interaction on Climate During the Mid-Holocene, *Science*, 280(5371), 1916–1919.

Ganopolski, A, Rahmstorf, S, Petoukhov, V, and Claussen, M (1998b) Simulation of Modern and Glacial Climates with a Coupled Global Model of Intermediate Complexity, *Nature*, 391(6665), 351–356.

Harrison, S P, Kutzbach, J E, Prentice, I C, Behling, P J, and Sykes, M T (1995) The Response of Northern Hemisphere Extratropical Climate and Vegetation to Orbitally Induced Changes in Insolation During the Last Interglaciation, *Quatern. Res.*, 43, 174–184.

Hays, J D, Imbrie, J, and Shackleton, N J (1976) Variations in the Earth s Orbit: Pacemaker of the Ice Ages, *Science*, 194, 1121–1132.

Heinrich, H (1988) Origin and Consequences of Cyclic Ice Rafting in the Northeast Atlantic Ocean During the Past 130 000 Years, *Quatern. Res.*, 29, 142–152.

Houghton, J T, Meira Filho, L G, Callander, B A, Harris, N, Kattenberg, A, and Maskell, K (1996) *Climate Change 1995: The Science of Climate Change*, Intergovernmental Panel on Climate Change, Cambridge University Press, 1–572.

Jolly, D, Harrison, S P, Damnati, B, and Bonnefille, R (1998) Simulated Climate and Biomes of Africa During the Late Quaternary: Comparison with Pollen and Lake Status Data, *Quatern. Sci. Rev.*, 17, 629–657.

Joussaume, S, Taylor, K E, Braconnot, P, Mitchell, J F B, Kutzbach, J E, Harrison, S P, Prentice, I C, Broccoli, A J, Abe-Ouchi, A, Bartlein, P J, Bonfils, C, Dong, B, Guiot, J, Herterich, K, Hewitt, C D, Jolly, D, Kim, J W, Kislov, A, Kitoh, A, Loutre, M F, Masson, V, McAvaney, B, McFarlane, N, de Noblet, N, Peltier, W R, Peterschmitt, J Y, Pollard, D, Rind, D, Royer, J F, Schlesinger, M E, Syktus, J, Thompson, S, Valdes, P, Vettoretti, G, Webb, R S, and Wyputta, U (1999) Monsoon Changes for 6000 Years Ago: Results of 18 Simulations from the Paleoclimate Modelling Intercomparison Project (PMIP), *Geophys. Res. Lett.*, 26(7), 859–862.

Jouzel, J, Barkov, N I, Barnola, J M, Bender, M, Chapellaz, J, Genthon, C, Kotlyakov, V M, Lipenkov, V, Lorius, C, Petit, J R, Raynaud, D, Raisbeck, G, Ritz, C, Sowers, T, Stievenard, M, Yiou, F, and Yiou, P (1993) Vostok Ice Cores: Extending the Climatic Records over the Penultimate Glacial Period, *Nature*, 364(6436), 407–412.

Jouzel, J, Hoffmann, G, Koster, R D, and Masson, V (2000) Water Isotopes in Precipitation: Data/model Comparison for Present-day and Past Climates, *Quatern. Sci. Rev.*, 19(1–5), 363–380.

Kutzbach, J E, Prell, W L, and Ruddiman, W F (1993) Sensitivity of Eurasian Climate to Surface Uplift of the Tibetan Plateau, *J. Geol.*, 101, 177–190.

Kutzbach, J E (1999) Simulations of the Climate of the Holocene. Perspectives Gained with Models of Different Complexity, Workshop on Geosphere–Biosphere Interactions and Climate, ed L Bengtsson, Pontifical Academy of Sciences, Vatican City.

Lautenschlager, M, Mikolajewicz, U, Maier-Reimer, E, and Heinze, C (1992) Application of Ocean Models for the Interpretation of Atmospheric General Circulation Model Experiments on the Climate of the Last Glacial Maximum, *Paleoceanography*, 7(6), 769–782.

Loutre, M F and Berger, A (2000) Future Climatic Changes: Are we Entering an Exceptionally Long Interglacial? *Clim. Change*, 46(1–2), 61–90.

Manabe, S and Stouffer, R J (1999) Are Two Modes of Thermohaline Circulation Stable? *Tellus*, 51A(3), 400–411.

Pagani, M, Arthur, M A, and Freeman, K H (1999) Miocene Evolution of Atmospheric Carbon Dioxide, *Paleoceanography*, 14(3), 273–292.

Peltier, W R and Marshall, S (1995) Coupled Energy-balance/ice-sheet Model Simulations of the Glacial Cycle: A Possible Connection Between Terminations and Terrigenous Dust, *J. Geophys. Res.*, 100(D7), 14 269–14 289.

Rind, D, Lean, J, and Healy, R (1999) Simulated Time-dependent Climate Response to Solar Radiative Forcing Since 1600, *J. Geophys. Res.*, 104(D2), 1973–1990.

Sakai, K and Peltier, W R (1999) A Dynamical Systems Model of the Dansgaard Oeschger Oscillation and the Origin of the Bond Cycle, *J. Clim.*, 12(8) Part 1, 2238–2255.

Saltzman, B (1990) Three Basic Problems of Paleoclimatic Modeling: A Personal Perspective and Review, *Clim. Dyn.*, 5(2), 67–78.

Sellwood, B W, Price, G D, and Valdes, P J (1994) Cooler Estimates of Cretaceous Temperatures, *Nature*, 370(6489), 453–455.

Stocker, T F (1999) Past and Future Reorganizations in the Climate System, *Quatern. Sci. Rev.*, 19(105), 301–320.

Verbitsky, M and Saltzman, B (1994) Heinrich-type Glacial Surges in a Low-order Dynamical Climate Model, *Clim. Dyn.*, 10(172), 39–47.

Von Storch, J S, Kharin, V, Cubasch, U, Hegerl, G C, Schriever, D, von Storch, H, and Zorita, E (1997) A Description of 1260-year Control Integration with the Coupled ECHAM1/LSG General Circulation Model, *J. Clim.*, 10(7), 1525–1543.

Climate Model Simulations, Future

see Projection of Future Changes in Climate (*Opening essay*, **Volume 1**)

Climate Model Simulations, Geological

see Climate Model Simulations of the Geological Past (**Volume 1**)

Climate Model Simulations, Present and Historical

see Model Simulations of Present and Historical Climates (*Opening essay*, **Volume 1**)

Climate Models, Earth System

see Models of the Earth System (*Opening essay*, **Volume 1**)

Climate Models, Energy Balance

see Energy Balance Climate Models (**Volume 1**)

Climate Models, General Circulation

see General Circulation Models (GCMs) (**Volume 1**)

Climate, Mountain

see Mountain Climates (**Volume 1**)

Climate, Orbital Influences

see Climate Model Simulations of the Geological Past (**Volume 1**)

Climate Predictability

see Chaos and Predictability (**Volume 1**)

Climate Proxies

see Natural Records of Climate Change (**Volume 1**)

Climate, Radiative Forcing

see Radiative Forcing (**Volume 1**)

Climate Sensitivity

Michael E Schlesinger and Natalia Andronova
University of Illinois at Urbana-Champaign, Urbana, IL, USA

The importance of human-induced climate change depends critically on the temperature sensitivity of the climate system, measured by the change in global-average near-surface temperature resulting from a doubling of the preindustrial carbon dioxide (CO_2) concentration, denoted by ΔT_{2x}. If ΔT_{2x} is small, then the problem of human-induced climate change may not be acute. If ΔT_{2x} is large, then human-induced climate change may be one of the most severe problems of the 21st century.

Estimates of the value of ΔT_{2x} *have been based on mathematical climate models, instrumental measurements of temperature since about the mid 19th century, and surrogate indicators of temperature prior to the instrumental record. Based predominantly on climate model simulations, the Intergovernmental Panel on Climate Change (IPCC) has stated that* $1.5\,^{\circ}C \leq \Delta T_{2x} \leq 4.5\,^{\circ}C$. *However, a recent study based on the instrumental temperature record finds that there is a significant (>50%) probability that* ΔT_{2x} *lies outside this range.*

Progress in reducing the uncertainty in the value of ΔT_{2x} *will require reducing the uncertainty in the anthropogenic factors such as non-sulfate aerosols, and natural factors such as changes in the outputs of the Sun and volcanoes, which may have contributed to the observed temperature changes. It will also require a longer record of instrumentally observed near-surface temperatures to enhance the signal of forced climate change relative to natural climate variations. Thus, it is quite likely that the formulation and negotiation of policies to abate human-induced climate change will, for the foreseeable future, continue to be made against a backdrop of deep uncertainty.*

INTRODUCTION

Imagine that, at some time t_o, the amount of energy given off by the Sun were to instantaneously increase, say by 1%. Such an increase in solar radiation would imbalance the energy budget of the Earth s climate system by causing more solar (shortwave) radiation to enter the system across the top of the atmosphere than the amount of terrestrial (longwave) radiation then being emitted by the system to space. This net energy gain will cause the system to warm. As it does, the longwave radiation emitted to space by the system will increase and the system will thereby eventually achieve a new balance or equilibrium wherein there is no net energy gain or loss by the system. What will be the change in global-average temperature of the air near the surface of the Earth, ΔT_e, when equilibrium is restored? The answer to this question is defined as the climate sensitivity, expressed either as the temperature change, ΔT_e, in degrees Celsius ($^{\circ}$C) for a prescribed radiative forcing F in $W\,m^{-2}$ at time t_o, such as a 1% increase in solar radiation, or as their ratio:

$$G_f = \frac{\Delta T_e}{F} \tag{1}$$

the change in near-surface temperature per unit of radiative forcing, in units of $^{\circ}C/W\,m^{-2}$. For convenience, F is usually taken not at the top of the atmosphere, but instead at the tropopause – the boundary between the lowest layer of the atmosphere, the troposphere, and the overlying atmosphere. In this case, F is called the adjusted radiative forcing and includes not only the original instantaneous radiative forcing at the tropopause, but also the additional radiative forcing there due to the resulting change in temperatures above the tropopause, with all quantities in the troposphere being held fixed (Andronova *et al.*, 1999).

The change in equilibrium near-surface temperature due to F can be expressed by Equation (1) as:

$$\Delta T_e = G_f F \tag{2}$$

Because G_f is positive, based on the first law of thermodynamics, positive F results in near-surface warming, $\Delta T_e > 0$, as in the example above, and negative F results in near-surface cooling, $\Delta T_e < 0$. The larger the value of G_f, the larger the equilibrium near-surface warming or cooling for a given positive or negative F.

The actual warming or cooling at any time t after t_o, but before the Earth s climate system restores its equilibrium, is the realized temperature change, $\Delta T(t)$, which is smaller in magnitude than the equilibrium temperature change. This occurs because of the thermal inertia of the climate system, principally as a result of the transfer of heat downward into the Earth s oceans. While the initial response of the near-surface temperature to the radiative forcing is rapid, full restoration of equilibrium can take thousands of years to achieve as heat is slowly transferred into the deep ocean from above. Only after this long time does $\Delta T(t) = \Delta T_e$ for a time-independent F.

Imagine now that the amount of energy given off by the Sun were to change not instantaneously, but as a function of time t. Then both F and ΔT_e will change with time, the latter following Equation (2), $\Delta T_e(t) = G_f F(t)$, and:

$$\Delta T(t) = G_f F(\tau) = \Delta T_e(\tau) \quad t_o < \tau \leq t \tag{3}$$

Consequently, the realized global-average temperature change at time t equals the equilibrium global-average temperature change at some earlier time τ and the climate system is continually striving to achieve equilibrium.

Equation (3) shows that the magnitude of the realized global-average temperature change at any time depends on the size of the climate sensitivity, G_f. Accordingly, the larger the value of G_f, the larger the impacts of any radiative forcing will be. For example, through the burning of fossil fuels – coal, gas and oil – which contain carbon, humanity has been increasing the concentration of CO_2 in the Earth s atmosphere since the beginning of the industrial revolution in the mid 18th century. Everything else being equal, this increased atmospheric CO_2 would reduce the amount of radiation emitted by the Earth to space, thereby leading to a time-dependent, positive radiative forcing. In response, according to Equations (2) and (3), both the equilibrium and realized near-surface air temperatures of the Earth would increase.

Several computer models have been developed to assess, in an integrated manner, the economic effects of such

a human-induced warming (e.g., Nordhaus, 1994). These integrated assessment models have taken the cost to the world economy of global warming to be:

$$C(t) = \alpha \left[\frac{\Delta T(t)}{3} \right]^n \tag{4}$$

where $C(t)$ is the costs of climate-change impacts in terms of their percent reduction in gross world productivity (GWP), $n = 2$ or 3, and the value of α – the economic damage for a 3 °C global-average warming – ranges from 0.5–20% of GWP. Although such damage functions can be criticized for their amalgamation into a single metric of the regional and cultural dependencies of climate-change impacts, they do serve to illustrate that such impacts may depend nonlinearly on climate sensitivity, for as given by Equations (3) and (4),

$$C(t) = \alpha G_f^n \left[\frac{F(\tau)}{3} \right]^n, \quad \tau_0 \leq t \tag{5}$$

Accordingly, uncertainty in the value of climate sensitivity G_f translates into an amplified uncertainty in the costs of climate-change impacts.

In view of this result, what is the accepted range in the value of the climate sensitivity G_f, or equivalently,

$$\Delta T_{2x} = G_f F_{2x}, \quad F_{2x} = 3.7\,\mathrm{W\,m^{-2}} \tag{6}$$

the change in equilibrium global-average near-surface air temperature in response to the radiative forcing equivalent to that due to a doubling of the pre-industrial CO_2 concentration, F_{2x}? The range given by the IPCC is $1.5\,^\circ\mathrm{C} \leq \Delta T_{2x} \leq 4.5\,^\circ\mathrm{C}$ (Houghton *et al.*, 2001), which translates into $0.4 \leq G_f \leq 1.2\,^\circ\mathrm{C/W\,m^{-2}}$. By Equation (5), this three-fold uncertainty in climate sensitivity translates into a nine-fold uncertainty in climate costs for $n = 2$ and a 27-fold uncertainty for $n = 3$. Such a large uncertainty in climate costs can confound attempts to develop and agree to policies that abate human-induced climate changes.

Why is there such a large uncertainty in the value of the climate sensitivity? To answer this question it is necessary to understand both the concept of feedbacks in the climate system and the computer models of climate that underpin the IPCC assessment of climate sensitivity.

CLIMATE SYSTEM FEEDBACKS

In the examples above, for the 1% increase in energy emitted by the Sun and the doubling of the pre-industrial CO_2 concentration, the climate system responds to the respective radiative forcing F by changing its near-surface air temperature T and other climate-system properties such as atmospheric water vapor and clouds, which we shall denote by X, to restore the net radiation at the top of the Earth s atmosphere N back to zero. For a small change Δ, this equilibration can be expressed by the linear approximation:

$$F + \frac{\partial N}{\partial T}\Delta T_\mathrm{e} + \left(\frac{\partial N}{\partial X}\frac{dX}{dT} \right) \Delta T_\mathrm{e} = 0 \tag{7}$$

where $(\partial N/\partial T)$ is the change in the net radiation due to the change in near-surface air temperature T alone, and $(\partial N/\partial X)(dX/dT)\Delta T_\mathrm{e}$ is the change in N due to the change in everything else in the climate system. The latter includes, for example, the vertical profile of temperature, the amount of water vapor in the atmosphere, and the amount and radiative properties of clouds. Solving for ΔT_e in Equation (7) yields Equation (2) with the sensitivity (gain) of the climate system with feedback given by:

$$G_f = \frac{G_o}{1 - f} \tag{8}$$

where

$$G_o = -\left(\frac{\partial N}{\partial T} \right)^{-1} = \frac{T}{(1 - \alpha_\mathrm{p})S_o} = 0.3\,^\circ\mathrm{C}\,(\mathrm{W\,m^{-2}})^{-1} \tag{9}$$

is the sensitivity (gain) of the climate system without feedback, with $S_o (= 1370\,\mathrm{W\,m^{-2}})$ the amount of energy from the Sun incident on the top of the Earth s atmosphere, $\alpha_p (= 0.30)$ the reflectivity (albedo) of the Earth, and $T (= 288\,\mathrm{K})$ the global-average near-surface air temperature; and:

$$f = G_o \frac{\partial N}{\partial X}\frac{dX}{dT} \tag{10}$$

is the feedback (Schlesinger, 1985, 1988).

If N were independent of all climatic quantities X, hence $\partial N/\partial X = 0$, or if X were independent of the near-surface air temperature T, hence $dX/dT = 0$, then $f = 0$ by Equation (10). In this zero-feedback case, $G_f = G_o$ by Equations (8) and (9) and $(\Delta T_{2x})_o = 1.1\,^\circ\mathrm{C}$ by Equation (6).

If instead, N depends on one or more climatic quantities, hence $\partial N/\partial X \neq 0$, and they in turn depend on T, hence $dX/dT \neq 0$, then $f \neq 0$ by Equation (10). In this nonzero-feedback case, by Equation (8), G_f will either be larger or smaller than G_o, depending on whether the feedback f is positive (amplifying) or negative (damping).

For example, suppose X represents the amount of water vapor in the atmosphere and that this quantity increases with increasing near-surface air temperature, that is, $dX/dT > 0$. An increase in atmospheric water vapor, which is a greenhouse gas like CO_2, will decrease the amount of terrestrial radiation emitted to space by the climate system; hence the net incoming radiation at the top of the atmosphere will increase, i.e., $\partial N/\partial X > 0$. By Equations (10), (2) and (6), $f > 0$, $G_f > G_o$ and $\Delta T_{2x} > (\Delta T_{2x})_o$. The response of the

climate system with positive feedback is larger than the response without feedback.

In contrast, suppose X represents the amount of cloud cover and that cloud cover increases with increasing T, i.e., $dX/dT > 0$. Increasing cloud cover will allow less solar radiation to enter the climate system and less terrestrial radiation to escape to space. For low and middle altitude clouds, the solar effect dominates, hence $\partial N/\partial X < 0$. Now by Equations (10), (2) and (6), $f < 0$, $G_f < G_o$ and $\Delta T_{2x} < (\Delta T_{2x})_o$. The response of the climate system with negative feedback is smaller than the response without feedback.

As previously stated, the range for ΔT_{2x} given by the IPCC is $1.5\,^{\circ}\mathrm{C} \leq \Delta T_{2x} \leq 4.5\,^{\circ}\mathrm{C}$. To what values of feedback do these limits correspond? To answer this question we can rearrange Equation (8) to be:

$$f = 1 - \frac{G_o}{G_f} = 1 - \frac{(\Delta T_{2x})_o}{(\Delta T_{2x})} \tag{11}$$

the latter by Equation (6). Thus, $\Delta T_{2x} = 1.5\,^{\circ}\mathrm{C}$ and $\Delta T_{2x} = 4.5\,^{\circ}\mathrm{C}$ correspond respectively to $f = 0.27$ and $f = 0.76$. Why is this large range in positive feedback obtained? As shown by Equation (10), f depends on three quantities: G_o, the value of which is known; $\partial N/\partial X$, which can be calculated using a radiative-transfer model; and dX/dT, which can be determined only by a mathematical climate model. As described below, it is the uncertainty in dX/dT that results in the threefold uncertainty in f, and thus in G_f and ΔT_{2x}.

MATHEMATICAL CLIMATE MODELS

The earliest estimate of ΔT_{2x} using a simple mathematical climate model for the temperatures of the Earth s surface and atmosphere was made by Arrhenius in 1896 who found $\Delta T_{2x} = 5.4\,^{\circ}\mathrm{C}$ (Arrhenius, 1896; Houghton *et al.*, 2001).

The IPCC s estimated range of $1.5\,^{\circ}\mathrm{C} \leq \Delta T_{2x} \leq 4.5\,^{\circ}\mathrm{C}$ is based on simulations of climate change by the most comprehensive type of model in the hierarchy of mathematical climate models, namely, the general circulation model (GCM) of the atmosphere and ocean. GCMs are the only type of mathematical climate model that determines the geographical and vertical distributions of an ensemble of climatic quantities, including temperature, wind, water vapor, clouds and precipitation in the atmosphere; soil moisture, soil temperature and evaporation on the land; and temperature, currents, salinity and sea ice in the ocean.

The mathematical equations in GCMs that express the fundamental governing laws of nature, such as conservation of energy, conservation of mass, and Newton s second law of motion, are so complex, however, that they can be solved only at discrete geographical and vertical locations, and only over discrete time intervals. Thus, a typical GCM divides the atmosphere into thousands of three-dimensional volumes, each having dimensions of about 250 km (155 mi) in both the north–south and east–west directions, and 1 km (0.6 mi) in the vertical direction. Reducing these dimensions is severely constrained by the speed of available supercomputers. For example, increasing the horizontal resolution of a GCM 10-fold from 250 km (155 mi) to 25 km (16 mi) would increase the required computer time 1000-fold, from about two weeks to compute a greenhouse-gas induced change in equilibrium climate to more than 30 years.

The constraint on horizontal resolution imposed by available supercomputers also has profound significance for the physical processes that can be included in a GCM. In particular, the Earth s climate system encompasses physical processes ranging in horizontal size from 40 055 km (24 889 mi) for disturbances that span the entire circumference of the Earth, to about one-millionth of a meter (0.00004 in) for the condensation of water vapor onto aerosol particles and dust to form clouds. GCMs using supercomputers explicitly include only physical processes having horizontal sizes of about 250 km (155 mi) and larger. Worse yet, the physical processes smaller than 250 km (155 mi) that are not resolved cannot be ignored because their effects significantly determine climate and climate change. Thus, climate modelers face the dilemma that their models cannot resolve all the important physical processes, but they cannot ignore their effects. This is the major difficulty in modeling the Earth s climate.

The approach taken to overcome the computational limitation on a GCM s spatial resolution is to determine the effects of the unresolved physical processes on the scales resolved by a GCM by using information available only on the resolved scales. This approach is called parameterization (*see* Parameterization, Volume 1). The principal differences among GCMs lie in their different parameterizations of the unresolved subgrid-scale physical processes, particularly those for clouds and precipitation.

The parameterizations used in GCMs have a significant influence on the climate changes simulated by these models, i.e., the different parameterizations used for the unresolved physical processes are the principal cause for the differences in f, G_f and ΔT_{2x} simulated by GCMs. Put another way, the parameterizations can be tuned to yield a wide range of ΔT_{2x} values. This dependency can be eliminated by either increasing the spatial resolution of a GCM such that it resolves all the important physical processes, thereby eliminating the need for parameterizations, or by providing a target value of ΔT_{2x} for a GCM s parameterizations to replicate. While the first option may not be impossible in principle, it is not now possible in practice because it requires a hypercomputer that is many powers of 10 faster than today s supercomputers. The second option requires estimating ΔT_{2x} empirically from observed climate changes.

EMPIRICAL ESTIMATES OF CLIMATE SENSITIVITY

The value of climate sensitivity has been estimated empirically using both the historical record of temperatures measured by thermometers since about the middle of the 19th century, and non-thermometer, surrogate (proxy) indicators of temperature for ancient (paleo)climates. The instrumental record consists of near-surface air temperature measurements obtained at land stations and sea-surface temperature measurements obtained by ships. The pre-instrumental proxy record consists of quantities that are sensitive to temperature, such as the thickness and isotopic composition of the annual growth rings of trees and the annual layers of glacial ice; the relative abundance and isotopic composition of planktonic (living near the sea surface) and benthic (living near the sea floor) foraminifera (shell-covered species) that, after dying, fall to the sea floor where they are covered up by the sedimentary material raining down from above; and the relative abundance of pollen in the annual growth layers of sediment at the bottom of lakes. These proxy data are converted into temperatures using statistical relations that have been developed for the present climate.

The temperature so reconstructed for any paleoclimate gives its temperature difference from the present, ΔT_e. This is used in Equation (2), together with an estimate of the radiative forcing, F, to estimate the climate sensitivity, G_f. This approach was pioneered by Budyko and colleagues using estimates of ΔT_e and F for six paleoclimates – (1) Late Cretaceous, 101–67 million years ago (Mya); (2) Paleocene, 67–58 Mya; (3) Eocene, 58–37 Mya; (4) Oligocene, 37–25 Mya; (5) Miocene, 25–9 Mya; and (6) Pliocene, 9–2 Mya – with the result $2.8\,^{\circ}\mathrm{C} \leq \Delta T_{2x} \leq 3.5\,^{\circ}\mathrm{C}$ (Budyko, 1977); also, see Chapter 6 of (MacCracken *et al.*, 1990). A more recent estimate based on reconstructed temperatures for the Cretaceous and the Last Glacial Maximum, 21 000 years ago, yielded a similar result, $1.4\,^{\circ}\mathrm{C} \leq \Delta T_{2x} \leq 3.2\,^{\circ}\mathrm{C}$ (Hoffert and Covey, 1992).

There are at least three factors that lead to uncertainty in the estimates of ΔT_{2x} by the paleocalibration method. First, the proxy data for temperature are not global in extent, hence their global average is uncertain, and their conversion to temperature is also uncertain. Second, estimation of the radiative forcing for paleoclimates relative to the present climate is difficult and thus uncertain. Third, the sensitivity of paleoclimate temperature changes from the present climate, G_f and ΔT_{2x}, may be different from the sensitivity of future human-induced temperature changes from the present. This is possible because in Equation (8) for G_f, G_o in Equation (9) depends on the state of the climate (T, α_p, S_o) that is being perturbed, and f depends on which feedback processes are active, both of which factors can vary with paleoclimate. Accordingly, it is desirable to estimate G_f and ΔT_{2x} using the instrumental temperature record.

Two approaches have been used to estimate the climate sensitivity from the observed departure of near-surface temperature from a 30-year average temperature, $\delta T_O(t)$, shown by the dashed curve in Figure 1. One method uses a statistical model such as:

$$\delta T_M(t) = \beta_0 + \beta_1 \delta T_M(t-1) + \beta_2 F(t) + \varepsilon(t) \quad (12)$$

where $\delta T_M(t)$ and $F(t)$ are the model s temperature departure and radiative forcing in year t, and $\varepsilon(t)$ is an error term of mean zero. The values of the coefficients β_0, β_1 and β_2, are estimated statistically by fitting $\delta T_M(t)$ to $\delta T_O(t)$ such that the sum over all years of $[\delta T_M(t) - \delta T_O(t)]^2$ is minimized. The climate sensitivity is then given by $G_f = \{[(1-\beta_1)/\beta_2] - k\}^{-1}$ where k characterises the heat transfer from the upper ocean to the deeper ocean and ΔT_{2x} by Equation (6). This approach was pioneered by Miles and Gildersleeves who obtained $\Delta T_{2x} = 1.4 - 2.2\,^{\circ}\mathrm{C}$ (Miles and Gildersleeves, 1977). A major problem with the statistical model arises because we do not have observations

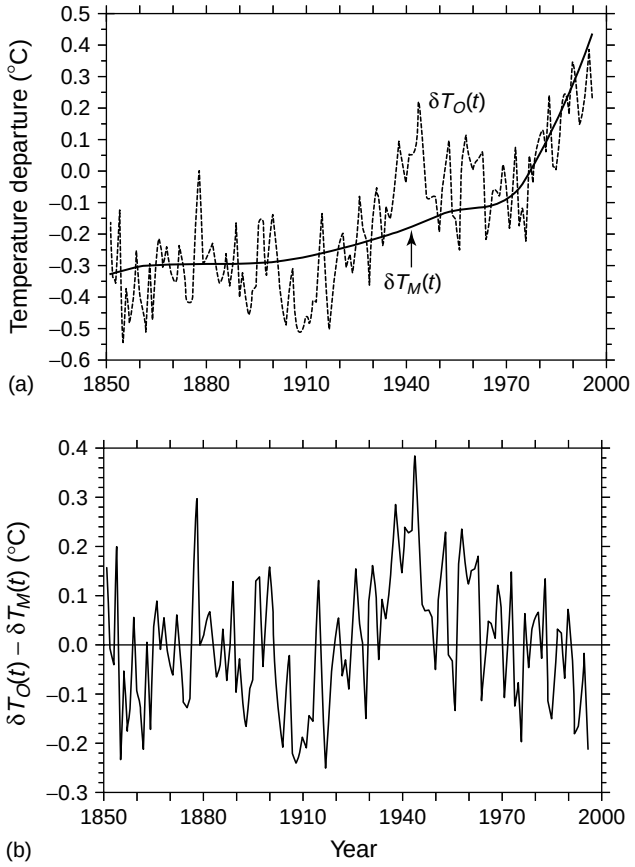


Figure 1 (a) The observed and model-simulated departures of near-surface temperature from the 1961–1990 average temperature, and (b) their difference for the GT radiative-forcing model with greenhouse gases (GHGs) and tropospheric ozone

of the temperature change of the deep ocean. Accordingly, an assumption must be made regarding k, and this assumption strongly influences the estimated value of the climate sensitivity. A way to overcome this problem is to use a simple, physically based mathematical model of the atmosphere/ocean system.

In one such mathematical model (Schlesinger *et al.*, 1997), the temperatures of the atmosphere and ocean are simulated, the latter as a function of depth from the surface to the ocean floor. In the model, the ocean is subdivided vertically into 40 layers, with the uppermost being the mixed layer and the deeper layers each being 100 m thick. Also, the ocean is subdivided horizontally into a polar region where bottom water is formed, and a nonpolar region where there is upwelling. In the nonpolar region, heat is transported upwards toward the surface by the water upwelling there and downwards by physical processes whose effects are treated as an equivalent diffusion. Heat is also removed from the mixed layer in the nonpolar region by a transport to the polar region and downwelling toward the bottom, this heat being ultimately transported upward from the ocean floor in the nonpolar region. The atmosphere in each hemisphere is subdivided into the atmosphere over the ocean and the atmosphere over land, with heat exchange between them. The governing (differential) equations in this model are stepped forward in time from some initial time, circa 1850, in steps of a year or less, to obtain $\delta T_M(t)$, an example of which is shown by the solid curve in Figure 1. The climate sensitivity is determined such that the sum of $[\delta T_M(t) - \delta T_O(t)]^2$ over all years of observations is minimized.

The value of ΔT_{2x} estimated by the simple atmosphere/ocean model is presented in Figure 2 for four different radiative forcing models. The GT model includes the radiative forcing by the GHGs CO_2, methane, nitrous oxide and the chlorofluorocarbons CFC-11 and CFC-12, as well as tropospheric ozone. For this radiative forcing model, $\Delta T_{2x} = 1.1\,^{\circ}\mathrm{C}$, which is the same as $(\Delta T_{2x})_o$, the climate sensitivity with zero feedback.

The GTA model includes the radiative forcing by sulfate (SO_4) aerosol, which is created in the atmosphere from SO_2 gas generated by the burning of coal and oil that contains sulfur. In contrast to the positive radiative forcing by the GHGs, the SO_4 aerosol produces negative radiative forcing by two effects, the enhanced scattering to space of solar radiation incident on the aerosol in the cloud-free atmosphere, and the formation (nucleation) on the aerosol of cloud droplets that are smaller than they would be in the absence of the aerosol, thereby preferentially scattering solar radiation backwards to space and enhancing the longevity of the clouds. For the GTA radiative forcing model, $\Delta T_{2x} = 5.0\,^{\circ}\mathrm{C}$ ($f = 0.78$), which exceeds the upper bound of the IPCC estimate, $\Delta T_{2x} = 4.5\,^{\circ}\mathrm{C}$. This large value of ΔT_{2x} is required

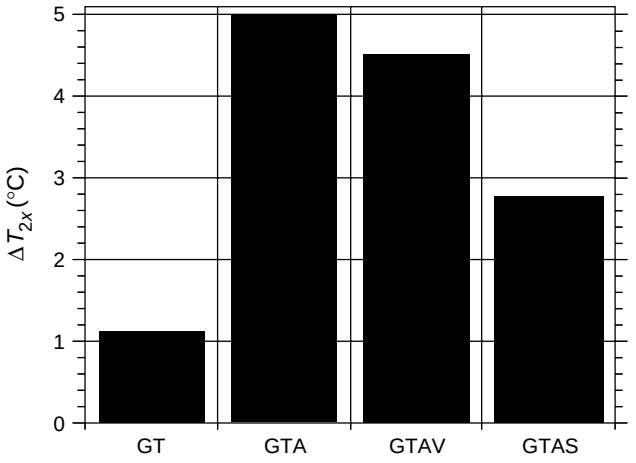


Figure 2 The climate sensitivity for four radiative-forcing models: (1) GT, GHGs and tropospheric ozone; (2) GTA, as in GT, but with sulfate aerosols; (3) GTAV, as in GTA, but with volcanoes; and (4) GTAS, as in GTA, but with solar irradiance variations

for the simulated temperature departure to reproduce the observed temperature departure for the smaller net positive radiative forcing that results from the partial cancellation of the positive GHG forcing by the negative aerosol forcing.

When the negative radiative forcing by volcanoes is included, as in the GTAV radiative forcing model, the estimated value of ΔT_{2x} is reduced slightly to $4.5\,^{\circ}\mathrm{C}$ ($f = 0.76$). This occurs because there was more negative radiative forcing by volcanoes in the 19th century than in the 20th century, and this reduced negative forcing with increasing time is equivalent to a positive forcing during the period of the observed warming, hence the value of ΔT_{2x} needed for the simulated temperature departure to agree with the observed is thereby diminished.

Lastly, the GTAS radiative forcing model shows the effect of conjectured variations in the energy emitted by the Sun. Solar radiative forcing larger than the 0.1% variation of the solar irradiance observed by satellites since 1978 over two 11-year sunspot cycles has been hypothesized to have occurred before 1978, based on the observed variations of other characteristics of the Sun and Sun-like stars. Comparison of the climate sensitivity for the GTAS and GTA radiative forcing models shows that inclusion of the conjectured solar radiative forcing reduces ΔT_{2x} by 46% to $2.7\,^{\circ}\mathrm{C}$ ($f = 0.41$). Clearly, the uncertainty in whether or not the Sun s irradiance has changed by more than the 0.1% observed by satellites is a major uncertainty in the determination of ΔT_{2x}.

Yet another uncertainty in estimating ΔT_{2x} is displayed in Figure 1(b) which shows the difference between $\delta T_O(t)$

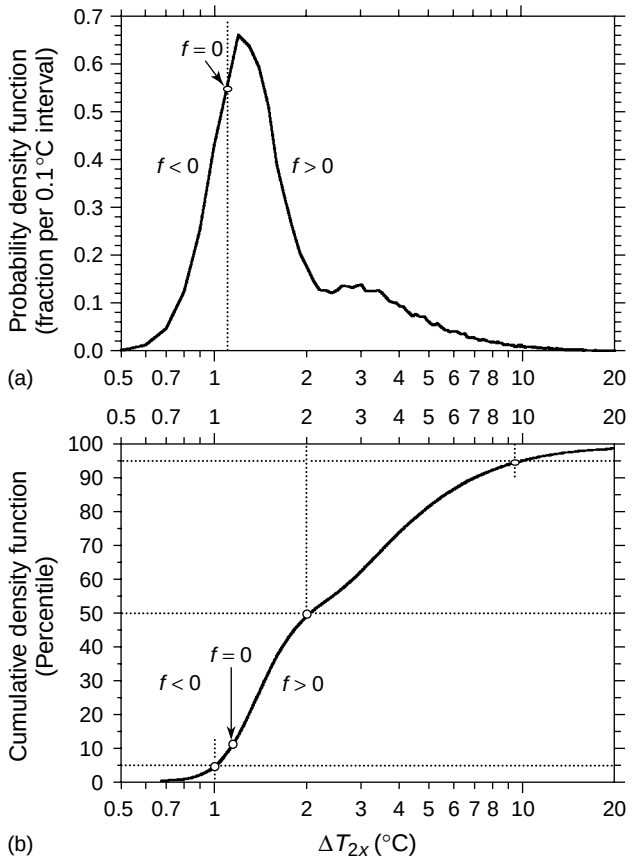


Figure 3 Differential (a) and cumulative (b) density functions for the climate sensitivity

and $\delta T_M(t)$ in Figure 1(a). This temperature difference represents the part of the observed temperature change that is not explained by the radiative forcings described above, hence it represents the natural variability of the climate system, of which we have only the single realization shown in Figure 1(b). If other realizations were available, our estimates of ΔT_{2x} based on them plus the forced temperature changes would likely yield different results. In a recent study we have attempted to quantify this uncertainty by statistically generating 80 000 realizations of the natural variability based on sampling $\delta T_O(t) - \delta T_M(t)$, adding to each the model-simulated forced temperature departures, and then estimating ΔT_{2x} for each of the resulting 80 000 surrogate observed temperature records (Andronova and Schlesinger, 2001). The result is presented in Figure 3 in terms of both the differential and cumulative probability density functions for ΔT_{2x}. The latter shows that there is a 33% probability that $\Delta T_{2x} \leq 1.5\,^{\circ}\mathrm{C}$ and a 79% probability that $\Delta T_{2x} \leq 4.5\,^{\circ}\mathrm{C}$. Accordingly, the combined uncertainties in the radiative forcing model and the natural variability of the climate system yield a 54% $[= 33 + (100 - 79)]$ probability that ΔT_{2x} lies outside the IPCC range of $1.5\,^{\circ}\mathrm{C} \leq \Delta T_{2x} \leq 4.5\,^{\circ}\mathrm{C}$.

THE WAY FORWARD

To reduce the uncertainty in the estimate of ΔT_{2x} requires learning whether or not the energy emitted by the Sun before 1978 varied by more than the 0.1% variation that has been observed by satellites since 1978. It also requires a longer record of instrumentally observed near-surface temperatures to enhance the signal of forced climate change relative to natural climate variations. It is quite likely that a significant reduction in the uncertainty in ΔT_{2x} will take a large part of the 21st century. Thus formulation and negotiation of policies to abate human-induced climate change will continue for the foreseeable future to be made against a backdrop of deep uncertainty. This argues for adaptive climate-change policies, which make midcourse corrections based on what is observed and learned in the future (Lempert *et al.*, 1996). Such adaptive policies can be constructed that are robust against our present uncertainty about the magnitude of the climate-change problem (Lempert and Schlesinger, 2000).

REFERENCES

Andronova, N G, Rozanov, E V, Yang, F, Schlesinger, M E, and Stenchikov, G L (1999) Radiative Forcing by Volcanic Aerosols from 1850 Through 1994, *J. Geophys. Res.*, 104(D14), 16 807–16 826.

Andronova, N G and Schlesinger, M E (2001) Objective Estimation of the Probability Density Function for Climate Sensitivity, *J. Geophys. Res.*, (in press).

Arrhenius, S (1896) On the Influence of Carbonic Acid in the Air Upon the Temperature of the Ground, *Lond., Edin., Dub. Phil. Mag. J. Sci.*, 41(251), 237–277.

Budyko, M I (1977) *Present-day Climate Change*, Gidrometeoizdat, Leningrad, 1–44.

Hoffert, M I and Covey, C (1992) Deriving Global Climate Sensitivity from Paleoclimate Reconstructions, *Nature*, 360, 573–576.

Houghton, J T, Ding, Y, Griggs, D J, Noguer, M, van der Linden, P J, and Xiaosu, D, eds (2001) *Climate Change 2001: The Scienti c Basis*, Cambridge University Press, Cambridge.

Lempert, R J and Schlesinger, M E (2000) Robust Strategies for Abating Climate Change, *Clim. Change*, 45(3–4), 387–401.

Lempert, R J, Schlesinger, M E, and Bankes, S C (1996) When We Don t Know the Costs or the Benefits: Adaptive Strategies for Abating Climate Change, *Clim. Change*, 33, 235–274.

MacCracken, M C, Budyko, M I, Hecht, A D, and Izrael, Y A (1990) *Prospects for Future Climate: a Special US/USSR Report on Climate and Climate Change*, Lewis Publishers, Chelsea, MI, 1–270.

Miles, M K and Gildersleeves, P B (1977) A Statistical Study of the Likely Causative Factors in the Climatic Fluctuations of The Last 100 Years, *Meteor. Mag.*, 106, 314–322.

Nordhaus, W D (1994) *Managing the Global Commons: the Economics of Climate Change*, The MIT Press, Cambridge, 1–213.

Schlesinger, M E (1985) Feedback Analysis of Results from Energy Balance and Radiative–Convective Models, in *The Potential Climatic Effects of Increasing Carbon Dioxide*, eds M C MacCracken and F M Luther, US Department of Energy, 280–319.

Schlesinger, M E (1988) Quantitative Analysis of Feedbacks in Climate Model Simulations of CO_2-induced Warming, in *Physically Based Modelling and Simulation of Climate and Climatic Change*, ed M E Schlesinger, NATO Advanced Study Institute Series, Kluwer, Dordrecht, 653–736.

Schlesinger, M E, Andronova, N G, Entwistle, B, Ghanem, A, Ramankutty, N, Wang, W, and Yang, F (1997) Modeling and Simulation of Climate and Climate Change, in *Past and Present Variability of the Solar-Terrestrial System: Measurement, Data Analysis and Theoretical Models. Proceedings of the International School of Physics "Enrico Fermi" CXXXIII*, eds G Cini Castagnoli and A Provenzale, IOS Press, Amsterdam, 389–429.

Climate Simulations, Future

see Model Simulations of Present and Historical Climates (*Opening essay*, **Volume 1**)

Climate Simulations, Geological Past

see Climate Model Simulations of the Geological Past (**Volume 1**)

Climate Simulations, Historical

see Model Simulations of Present and Historical Climates (*Opening essay*, **Volume 1**)

Climate Simulations, Present

see Model Simulations of Present and Historical Climates (*Opening essay*, **Volume 1**)

Climate, Solar Irradiance Effects

see Solar Irradiance and Climate (**Volume 1**)

Climate Variability and Predictability (CLIVAR)

see CLIVAR (CLImate VARiability and Predictability) (**Volume 1**)

Climate Variability, Natural

see Natural Climate Variability (**Volume 1**)

Climate, Volcanic Effects

see Volcanic Eruptions (**Volume 1**)

Climatology

F Kenneth Hare
University of Toronto, Toronto, Canada

Climatology is the scienti c eld that focuses on observation and documentation of the climate (generally of the Earth's climate, but it can refer to study of planetary climates as well). Climate, in turn, is the term used to describe the way the atmosphere has been and can be over time scales from days to millions of years, and space scales from kilometers to spanning the Earth. From its beginnings in the 19th century, dependent largely on volunteer observers and on the simplest archiving facilities, climatology has become a richly equipped science, able to draw on a sophisticated observational system using orbiting and geostationary satellites; a world-wide network of free-air soundings by balloons, rockets, and airplanes; state-of-the-art data reduction and storage systems; and, above all, comprehensive climate models that can be used to make predictions of climate change. During the last few decades of the 20th century, study of the climate has become very encompassing, so that it now refers to study of the interconnected system involving the atmosphere, the oceans, the land surface and its vegetation, and the snow and ice surfaces. Over recent decades, studies of the oceans and atmosphere have been brought closer together, and there are emerging collaborative efforts that include consideration of the biosphere. Building on the base of information provided by climatology, full understanding and prediction of how the climate will evolve has become a central objective - perhaps the

ultimate objective - of those studying the Earth and its environment (Manabe, 1998).

The Earth s climate system is composed of the atmosphere, the oceans, the land surface and its vegetation, and the snow and ice surfaces (*see* The Earth System, Volume 1). All climatic activity, however, traces back to the interaction of solar radiation – insolation – with the Earth in ways that create temperature gradients that, in turn, drive the circulation of the lower atmosphere and ocean, and thereby determine the world s climatic zones. A major advance in scientific understanding during the 20th century was to recognize that, in nature, nothing exists in isolation and that all aspects of the climate system are connected. Thus, those seeking to understand the climate must be involved in observing, documenting, and analyzing a wide range of geophysical and biological parameters extending from the Sun to the bottom of the oceans.

Study of the climate system, although it was implicit in the work of such pioneers as Augustin de Candolle and Thomas Blodget in the 19th century, did not really win full recognition until the 1970s, and then in the context of pending climatic change and its consequences (e.g., Gates and Mintz, 1975). The United Nations (UN) Conference on the Human Environment at Stockholm in 1972 was of particular importance in advancing study of the climate system, because it focused major attention on three connected environmental issues: atmospheric pollution, climatic stability and ecosystem health. To advance study of these topics, two parallel initiatives were taken by committees of the International Council of Scientific Unions (now ICSU, the International Council for Science). These initiatives have contributed greatly to advancing understanding of the Earth s climate system, including its physical, chemical, and biological environment. The first was the development of the World Climate Programme by ICSU, the World Meteorological Organization, the UN Environment Programme, and other UN agencies, with a World Climate Research Programme (WCRP) at its core. The second was the development of the International Geosphere–Biosphere Programme (IGBP), which brought Earth scientists and biologists together. Both initiatives, which are still nominally independent (although strong couplings are developing), have dealt squarely with the issue of climatic change: with climate s past, largely in the IGBP area, and its present and future state, largely in WCRP s. The net result has been the emergence of the climatic system as the common object of study.

In the study of the climate system, climatologists are pursuing three key approaches. The first is analysis of climatic archives. The instrumental network began in the 19th century (with a few pioneering efforts reaching back to the 17th century). By the late 19th century, pioneers such as Laplace, Ferrel, Maxwell and von Helmholtz had shown what needed to be observed to understand the Earth s fluid domains – water, air and ice. Observation systems were initiated, and archives of such variables as temperature, air pressure, humidity and precipitation were generated. Not only are individual records analyzed, but connections, correlations, teleconnections, and apparent contradictions are being carefully investigated. The next phase, only just beginning, will be to identify the regularities that have still to be isolated (since most lie unsuspected) in the archives.

The second approach to understanding the Earth s climate system is the validation of the various theoretically based models of the system (*see* Models of the Earth System, Volume 1), of which the most encompassing are the general circulation models of the atmosphere and ocean. These models are large and comprehensive, and give solutions that create model climates that need to be compared with one another, and with the observed climate of today. Validating that the models do a reasonable job of reproducing present and past climates is a key step in building confidence in their use for predicting potential changes in future climatic conditions (*see* General Circulation Models (GCMs), Volume 1).

This is a formidable challenge; the need is to predict the present day mean fields of temperature, precipitation and winds, and also the set of space and time anomalies that are internal to the present-day climate, but which may be greatly changed in the future. These anomalies include such perturbations as the El Niño–La Niña events of the Pacific, which appear to affect much of the globe (McPhaden, 1999); (*see* El Niño/Southern Oscillation (ENSO), Volume 1) and the Arctic oscillation (with its relative the North Atlantic Oscillation) (*see* Arctic Oscillation, Volume 1; North Atlantic Oscillation, Volume 1). In the same category, but lying more in the oceanographic domain, are predictions of changes in ocean currents and in the thermohaline circulation within the oceans, whose bottom water global coldness (below 4 $^{\circ}$C) is dependent on high-latitude processes.

A major international effort has been made in the past 10 years to carry this validation task forward, thanks to the work of the WCRP s Atmospheric Model Intercomparison Project begun in 1989 (*see* AMIP (Atmospheric Model Intercomparison Project), Volume 1). Directed by W L Gates, and based at the US Lawrence Livermore National Laboratory, the very large effort has confirmed (with reservations) the professional confidence in general circulation modeling (Gates *et al.*, 1999). Thirty-one such models have been examined (virtually the entire international effort). No single model has proved superior across all the range of outputs. Modeled seasonal distributions of pressure, temperature and circulations are reasonably close to observed values (apart from the El Niño and the Southern Oscillation (ENSO) anomalies over the Pacific), but precipitation is less so; and cloudiness (for which the observational

basis is poor) is not yet well predicted. Numerical modeling of this sort is, by common agreement, the best form of experiment possible within the observational sciences, where controlled isolation of the object, in this case the Earth s climate system, is unattainable. Years ago, John von Neumann identified the prediction of weather and climate as among the most difficult problems, and chose to put such numerical predictions high on the list of tasks for the digital computers then being launched. Von Neumann s prediction has proved true.

Third in the list of approaches has been focusing on obtaining a close understanding of processes taking place at the Earth s surface. There has been a long-standing interest among geographers, ecologists, and others interested in the climate near the ground, defined since Rudolph Geiger s times as microclimatology, because it deals with local sites. More recently, the students of surface climates have called themselves physical climatologists (especially among geographers), including the specialized group of urban climatologists (e.g., Bailey *et al.*, 1997). In parallel with such efforts, and often independently, there has been long-standing work on the atmospheric and oceanic boundary layers, including the study of turbulence, and of the role of vegetation in modulating exchanges of heat and moisture with the atmosphere (Monteith, 1976). This is territory that has been explored by several different sub-disciplines that have tended to disregard one another, until challenged by the global climatic change problem. Among climatologists, it was Mikhail Budyko (*see* Budyko, Mikhail Ivanovich, Volume 1) and his Russian colleagues who first overcame this arm s-length attitude (Budyko, 1959).

Here again there has been a concerted international effort to reach agreement on the role of the land surface in weather and climate, including the provocative question: does the land surface matter? As wind passes over plant-covered terrain, it encounters roughness elements (trees, shrubs, grass, etc.) that are crudely represented in terms of their roughness length in boundary layer theory. The boundary layer then experiences fluxes of momentum, water (evapotranspiration), sensible heat, carbon dioxide and many other trace gases and particles. Organic processes within the vascular plant cover, also usually parameterized, modulate the fluxes. The immense complexity of the land surface means that all winds, as they pass along their trajectories, must be considered to be interacting with the surface, although the full range of interactions is rarely capable of being specified. Yet this interface is crucial to determining the general circulation of the global atmosphere, for which it acts as a brake, a heating element, a source of trace gases, and a reservoir for major exchanges of water and carbon dioxide. Pitman *et al.* (1999) have reported the results of an IGBP program called Biospheric Aspects of the Hydrological Cycle (*see* IGBP Core Projects, Volume 2). The results are tentative, yet very impressive. The writers conclude that accurate representation of land-use changes will be central to the ability to simulate the climate of the 21st century.

Two other themes for climatological study are also emerging. One is the reconciliation of traditional climatology with hydrology and glaciology. These fields grew up rather separately in most countries, in different government agencies, and often without reconciling, for example, conflicting fields of precipitation and run-off (Hare and Hay, 1970; Hare, 1980). For the reason that the transfer and storage of water substance is at the very center of operation of the climatic system, however, a major effort is underway to unify the treatment of water within an enlarged climatology (or if another term is preferred, simply global system science). The Global Energy and Water Cycle Experiment (GEWEX) of the WCRP has been in progress for some time, and has already produced some landmark syntheses (see, for example, Stewart *et al.*, 1998) (*see* GEWEX (Global Energy and Water Cycle Experiment), Volume 1).

The second theme has emerged as a result of the astonishing findings of paleoclimatology, which have pushed the observational record back very far indeed, although not, as yet, in as great detail as the recent instrumental records. In broadest outline, the world contains numerous sources of laminated material – glacial ice, peat and lake-bottom organic sediments, ocean floor sediments, tree-rings and cave calcite deposits – in which the laminations represent annual, seasonal or longer-term repositories for proxy indicators of climate. These can be dated by a variety of isotope indicators, that can, in addition, offer evidence of the temperature of the past. The written record has also been used to great effect, and the world s volcanoes have helped by spreading ash layers that serve as markers in many deposits. Remarkably detailed reconstructions of climate have thus been created for the Holocene and late Pleistocene epochs, and some indications of past climates are available back into the Proterozoic, that is to say, more than 550 Myr. The relevant sciences are geology, geophysics, geochemistry, paleoecology and many others (not least isotope chemistry). A happy consequence of the global change movement has been to bring these sciences into direct contact with the climatological community.

Based on the results of these various approaches, the findings of climatologists have become of great interest to the world community. Although climate is not alone in causing global change, concerns, even fear, about the potential consequences of global warming, especially after the mid-1970s, have been a primary factor in promoting the global change movement. Through the 1960s and early 1970s, fear arose from early model predictions that rising carbon dioxide concentrations would lead to rising surface temperatures (Callendar, 1938; Plass, 1956; Manabe and Wetherald, 1975). In those decades, global mean temperatures appeared to be stable, or even falling, and there were warnings of

possible reversion to glacial conditions, so it was difficult for society to accept the notion of potential warming. Since 1975, however, increasingly accurate estimates of global surface temperatures have revealed a strong and continued warming even more marked than that between 1920 and 1945. Today, most climate scientists are convinced by the observational evidence that there has been a global average warming of about 0.7 °C since 1900 (*see* The Global Temperature Record, Volume 1), although they are not yet sure how much of the warming is due to human influences. This was the view cautiously expressed by the Intergovernmental Panel on Climate Change (IPCC, 1996), and it has underlain the successive rounds of discussions between governments that are trying to control greenhouse gas emissions. Improving confidence in our understanding of the Earth s climate and helping society to consider how to deal with, and respond to the potential for change, are key activities of climatologists.

REFERENCES

Bailey, W G, Oke, T R, and Rouse, W R, eds (1997) *The Surface Climates of Canada*, McGill-Queens University Press, Montreal.

Budyko, M (1959) *The Heat Balance of the Earth's Surface*, translator N S Stepanova, US Department of Commerce, Washington, DC. (Original 1956 Teplovoi balanszemnoi poverkhnosti, Leningrad, Gidrometeorologischeskoe izdatel stvo.)

Callendar, G S (1938) The Artificial Production of Carbon Dioxide and its Influence on Temperature, *Q. J. R. Meteorol. Soc.*, 64, 223–240.

Gates, W L, Boyle, J S, Covey, C, Davies, C G, Doutriaux, C M, Drach, R S, Florino, M, Glecker, P J, Hnilo, J J, Marlais, S M, Phillips, T J, Potter, G L, Santer, B D, Sperber, K R, Taylor, K E, and Williams, D N (1999) An Overview of the Results of the Atmospheric Model Intercomparison Project (AMIPI), *Bull. Am. Meteorol. Soc.*, 80, 29–50.

Gates, W L and Mintz, Y (1975) *Understanding Climatic Change*, National Academy of Sciences, Washington, DC.

Hare, F K (1980) Long-term Annual Surface Heat and Water Balances Over Canada and the United States South of 60 °N: Reconciliation of Precipitation, Run-off and Temperature Fields, *Atmosphere-Ocean*, 18, 127–153.

Hare, F K and Hay, J E (1970) Anomalies in the Large-scale Annual Water Balance Over Northern North America, *Can. Geogr.*, 15, 79–94.

IPCC (1996) *IPCC Second Assessment Climate Change 1995*, in four volumes, WMO and UNEP, Geneva.

Manabe, S (1998) Study of Global Warming by GFDL Climate Models, *Ambio*, 27, 182–186.

Manabe, S and Wetherald, R T (1975) The Effect of Doubling the CO_2 Concentration on the Climate of a General Circulation Model, *J. Atmos. Sci.*, 32, 3–15.

McPhaden, M J (1999) Genesis and Evolution of the 1997–1998 El Niño, *Science*, 283, 950–954.

Monteith, J K, ed (1976) *Vegetation and the Atmosphere*, 2 Volumes, Academic Press, New York.

Pitman, A, Pielke, Sr, R, Avissar, R, Clausen, M, Gash, J, and Holman, H (1999) The Role of the Land Surface in Weather and Climate: Does the Land Surface Matter? *Global Change Newslett., ICSU*, 39, 4–11.

Plass, G N (1956) The Carbon Dioxide Theory of Climate Change, *Tellus*, 8, 140–154.

Stewart, R E, Leighton, H G, Marsh, P, Moore, G W K, Ritchie, H, Rouse, W R, Soulls, E D, Strong, G S, Crawford, R W, and Kochtubajda, B (1998) The MacKenzie GEWEX Study: the Water and Energy Cycles of a Major North American River Basin, *Bull. Am. Meteorol. Soc.*, 79, 2665–2683.

FURTHER READING

Porter, S C and Haldorsen, S (1999) Environmental Change in the Glacial Ages, *Sci. Int., ICSU, Newslett.*, 70, 10–12.

Schellinhuber, H J (1999) Earth System Analysis and the Second Copernica Revolution, *Nature*, 402(Supplement), C19–24.

CLIVAR (CLImate VARiability and Predictability)

CLIVAR is an international scientific program under the aegis of the World Climate Research Programme (WCRP) to extend the investigation of climate variability and predictability to large geographic regions and longer time-scales and the response of the climate system to anthropogenic forcing.

The specific objectives of CLIVAR are:

- to describe and understand the physical processes responsible for climate variability and predictability on seasonal, interannual, decadal, and centennial time-scales;
- to extend the record of climate variability through paleoclimatic and instrumental data;
- to extend the range and accuracy of seasonal to interannual climate prediction through the development of global coupled predictive models;
- to understand and predict the response of the climate system to increases of radiatively active gases and aerosols, and to compare these predictions to the observed climate record in order to detect the anthropogenic modification of the natural climate signal.

The CLIVAR program is organized into three component programs:

CLIVAR-GOALS examines the variability and predictability of the Global Ocean Atmosphere Land System on seasonal-to-interannual time scales, building on the

successes of the Tropical Oceans Global Atmosphere (TOGA) program, by:

- developing observational capabilities;
- further developing models and predictive skill for SST and other climate variables on seasonal to interannual time scales around the entire global tropics;
- building understanding and predictive capabilities of the interaction of monsoons with the Indian Ocean, El Niño–Southern Oscillation and land surface processes;
- understanding climate variability and predictability arising from the interaction between the tropics and extratropics;
- exploring the predictability of extratropical seasonal to interannual climate variability induced by the interaction of the atmosphere with oceans, land surface processes, and sea-ice processes and developing means to exploit any such predictability.

See GOALS (Global Ocean- Atmosphere- Land System), Volume 1.

CLIVAR-DecCen determines the mechanisms of variability and predictability of climate fluctuations on Decadal-to-Centennial time-scales with a special emphasis on the role of the oceans in the global coupled climate system by:

- describing and understanding the patterns of global decadal-to-centennial climate variability in the instrumental, paleoclimatic, and model records;
- extending the records of climatic variability by concerted efforts of data recovery, reanalysis of existing atmospheric, oceanic and paleoclimatic data, finding new paleoclimatic indices, and instituting new oceanographic monitoring sites;
- developing and implementing appropriate observing, modeling, computing, and data collection and dissemination systems needed to describe, understand, and predict global decadal variability;
- identifying and studying the oceanic regions and processes, such as water mass transformation regions, strong boundary currents and return path choke points, through which the ocean and atmosphere interact to produce decadal-to-centennial climate variability.

CLIVAR-ACC studies the response of the climate system to anthropogenic climate change by:

- developing understanding, modeling and predictive capabilities of the response of the climate system to the anthropogenic increases in radiatively active gases and changes in aerosols;
- identifying the patterns of anthropogenic modification to the mean state and to the variability of the climate system;
- using the understanding of natural climate variability derived from the other two CLIVAR component programs as a baseline for detecting the trends and signatures associated with increases in greenhouse gases and the effects of other anthropogenic changes.

A scientific steering group has oversight of the development of the CLIVAR implementation. Topical workshops and symposia with broad international participation are organized to gather input for these plans and to refine them. CLIVAR implementation is closely coordinated with and builds on national programs and other international climate-related activities. Joint projects are developed, and maximum use is made of existing infrastructure. Two CLIVAR Numerical Experimentation Groups coordinate international modeling studies.

For further information, contact the WCRP, Case postale 2300, 1211 Geneva 2, Switzerland.

See also: Natural Climate Variability, Volume 1.

JOHN S PERRY *USA*

Cloud–Radiation Interactions

Richard C J Somerville
University of California, San Diego, CA, USA

Only a few years ago, clouds were regarded as picturesque but minor components of the hydrological cycle. Today, they are believed to be the single most important source of uncertainty in climate model forecasts of future climate change, because their interaction with solar and terrestrial radiation is such an important aspect of the climate system. A current dilemma of climate modeling is that model results are strongly sensitive to the treatment of several poorly understood physical processes, especially cloud- radiation interactions. Thus, different models with alternative plausible parameterizations often give widely varying results. Yet, we typically have had little basis for estimating which parameterization is more realistic, although most of the global differences in results between leading climate models, as measured by their sensitivity to greenhouse gases, can be traced to different model treatments of cloud- radiation interactions.

Cloud- radiation interactions are regarded today as one of the most critical areas in global change research. In particular, when climate models are intercompared, cloud-radiation parameterizations are responsible for most of the global-mean differences in sensitivity to greenhouse gas increases. This fact has become well established through parameterization transplant experiments.

In such computational experiments, transplanting the cloud- radiation algorithm from one model to another typically causes the recipient model to closely replicate the climate sensitivity of the donor model. The uncertainty in model responses is directly due to a lack of fundamental understanding of the physical processes involved. A major research effort is underway worldwide in response to this challenge. Furthermore, closely related research areas, such as the role of atmospheric aerosols in climate, are also beginning to receive the attention they deserve.

On even the simplest theoretical grounds, it is not surprising that climate is extremely sensitive to cloud amount. Similar arguments can be made to show that climate also ought to depend strongly on other cloud properties, such as cloud height and cloud liquid water or ice content. In fact, an easily stated but still unsolved major problem is to understand why the global cloud cover is now about 60% and why the planetary albedo is now about 30%. Were these quantities the same during the ice ages? What mechanisms maintain the system at the present values of these key parameters, and how stable are these mechanisms to perturbations, such as those due to changing greenhouse gas concentrations? We simply do not know. Until we nd out, the con dence limits or error bars on the results of climate model greenhouse simulations will be much too large.

For many years, virtually all global climate model or general circulation model (GCM) treatments of clouds were based on simple algorithms relating cloud amount to relative humidity. Such parameterizations usually produced positive global average cloud–radiation feedbacks in numerical experiments simulating greenhouse-induced climate change. For example, in a typical integration performed with a GCM developed in the 1970s, a climate warming due to increased atmospheric carbon dioxide (CO_2) concentrations would lead to increased average cloud heights and/or decreased average cloud amounts. It is easy to understand qualitatively why such feedbacks were positive. First, higher clouds are colder and so less effective infrared emitters, and they generally have lower albedos than lower clouds, so the cloud-height feedback was positive (i.e., the change in clouds produced by the warming tended to amplify the warming). Second, average model clouds, like average real clouds, contribute more strongly to the planetary albedo than to the planetary greenhouse effect (in technical terms, the shortwave cloud forcing is larger than the longwave cloud forcing by about $20\,W\,m^{-2}$). Hence, a reduction in cloud amount reduces the shortwave effect more than the longwave effect of clouds. Thus, the cloud amount feedback is also positive.

Climate models are now more numerous and more complicated, however, and model responses to increased greenhouse gas concentrations are more varied. GCMs today attempt to take into account a broader range of physical processes involved in cloud–radiation feedbacks. The climate modeling community now realizes clearly that cloud feedback processes are not limited to macrophysical cloud properties, such as cloud amount and cloud altitude. In recent years, many GCMs have begun to include cloud parameterizations, which include explicit treatments of cloud physics.

Clouds have a powerful effect on the radiation budget of the Earth. The reasons are obvious: clouds contribute to both the greenhouse effect, which warms the planet, and to the Earth s reflectivity, or albedo, a competing cooling effect. Studies in which alternative cloud parameterizations can be tested against observations show great promise for improving our understanding of clouds and their influence on climate. Ultimately, this improved understanding will find its way into the representations of clouds used in GCMs.

Improving the realism of this aspect of models is a key to improving the model simulations of climate change. We know that the leading models of today differ by a factor of three among themselves, when they are compared in terms of their simulation of the global average surface temperature increase, due to a prescribed climate forcing, such as doubling the atmospheric concentration of CO_2. Most of this factor of three is due to the different ways in which the models represent clouds and cloud–radiation interactions. We know, for example, that in many climate models, cloud amounts in a warmer climate tend to be somewhat less than in the present-day climate, and cloud altitudes tend to be somewhat greater. These changes lead to positive feedbacks, increasing the apparent sensitivity of the model climate to the imposed forcing which led to the original warming, such as an increase in CO_2. However, the nature and strength and especially the regional aspects of these feedbacks differ greatly from model to model. For this reason, many scientists regard cloud–radiation processes as the most critical area of climate modeling research, deserving highest priority for climate research resources.

Only about 70% of the sunlight intercepted by the Earth is available to drive the climate system. The other 30% (the planetary albedo) is simply reflected to space, mainly by clouds, which cover some 60% of the surface of the planet (*see* Energy Balance and Climate, Volume 1). On average, clouds reduce the global average absorbed solar radiation by about $50\,W\,m^2$. Clouds also help to trap terrestrial radiation, contributing about $30\,W\,m^2$ to the greenhouse effect. The net cloud radiative forcing is the difference, approximately $20\,W\,m^2$. Thus, the albedo effect dominates, and the net effect of clouds at present is to cool the Earth.

Simple blackbody radiative equilibrium calculations suggest that changes in cloud amount by only 1–2% might double or halve the model sensitivity to CO_2. One simple way to appreciate the climatic significance of clouds is

to compare the cloud radiative forcing magnitudes given above with the direct radiative effect of doubling the concentration of atmospheric CO_2, which is only about $4\,W\,m^{-2}$. Thus, if a climate change caused by increased CO_2 were to result in even a small change in the cloud amount, the cloud feedback effect might well be important. Furthermore, even if global average cloud amount did not change appreciably in response to a changed climate, the spatial and temporal distribution of clouds might well be altered, as might other critical quantities, such as cloud altitude and cloud radiative properties. All of these changes in clouds could lead to significant feedback effects.

Until cloud processes are much better understood, and until this understanding is incorporated in our models, the model results will always be subject to major uncertainties. However, the plain fact is that we lack a basic understanding as to why the global cloud amount is about 60%, why the planetary albedo is about 30%, how these and other fundamental quantities may have changed as climate changed over geological time, and how they may change in the future.

Much research is now underway exploring the role that microphysical properties of clouds might play in affecting climate change. These processes might well lead to strong feedbacks. For example, as the climate warms because of an increase in the greenhouse effect, the entire hydrological cycle may accelerate. More water may evaporate from the oceans, and the atmospheric concentration of water vapor may increase. Because of the greater availability of water vapor, some clouds in the warmer climate may have more liquid water or ice than their counterparts in today s climate. In general, a higher liquid water or ice content is thought to lead to a higher albedo, hence a negative feedback. However, for thin clouds, particularly cirrus, the cloud greenhouse effect may also increase. In addition, it is not at all clear that cloud water content will change in any systematic way as climate alters. It may also be too simplistic to look on temperature as a dominant controlling factor for cloud microphysics.

Furthermore, the way in which cloud water or ice content depends on temperature, even in the present climate, is not well understood. Simple theory and aircraft data and some modeling studies support the idea of higher cloud water contents in warmer clouds. Some recent interpretations of satellite data, however, suggest that even the sign of the temperature dependence may be in doubt. It seems unlikely that any simple universal relationship is valid. Additionally, the radiative properties of clouds also depend on factors such as the size distribution of cloud droplets, the shape of ice particles, and other factors. Despite much observational and theoretical work in recent years to explore these issues, we are still far from a comprehensive physical understanding of them.

Nevertheless, several leading atmospheric GCM groups in different countries have now incorporated this class of cloud feedback mechanisms in one way or another. In a typical approach, cloud liquid water or ice content is included in the model as an additional prognostic variable, just like temperature, wind velocity and water vapor. The physical processes that act as sources and sinks for cloud water or ice, such as evaporation, condensation and precipitation, are simulated parametrically. In other words, the effects of these processes on the cloud water and ice budget are represented by simple formulae relating these processes to the large-scale variables that the model predicts explicitly.

The results of climate change simulations with these models confirm the strong sensitivity of climate to cloud microphysics. In one striking set of numerical experiments, a British GCM group (Senior and Mitchell, 1993) produced global average surface temperature changes (due to doubled CO_2) ranging from $1.9-5.4\,^{\circ}C$, simply by altering the way in which these cloud–climate feedback mechanisms were treated in the model. They tested four different parameterizations, successively incorporating relative humidity cloud, prognostic cloud water, phase changes from water to ice, and interactive radiation dependent on cloud microphysics. Their cloud water algorithm was that of Smith (1990). It is somewhat unsettling that the results of a complex model can be so drastically altered by what amounts to changing a few lines of code, essentially replicating the factor of three difference in global sensitivity between GCMs that has been revealed by extensive model intercomparisons (Cess *et al.*, 1989). Clearly, further research is urgently required to understand this class of physical processes better and to incorporate this understanding in models.

One particularly promising avenue of research is to combine process modeling with intensive field observations and with research using GCMs. For too long, research in this field has been characterized by too many plausible cloud–radiation parameterizations and too little effort to test them empirically. Now that appropriate observations and novel modeling tools are at last becoming available, we may anticipate rapid progress in this critical area of climate research.

Recent research has led to a greatly increased understanding of the uncertainties in today s climate models. In attempting to predict the climate of the 21st century, we must confront not only computer limitations on the affordable resolution of global models, but also a lack of physical realism in attempting to model key processes. Until we are able to incorporate adequate treatments of critical elements of the entire biogeophysical climate system, our models will remain subject to these uncertainties, and our scenarios of future climate change, both anthropogenic and natural, will not fully meet the requirements of either policymakers or the public. The areas of most-needed model improvements are thought to include air–sea exchanges,

land surface processes, ice and snow physics, hydrologic cycle elements, and especially the role of aerosols and cloud–radiation interactions.

A serious dilemma of climate modeling today is that model results are extremely sensitive to parameterizations of several poorly understood physical processes. As a result, models with different plausible parameterizations give very different results. Unfortunately, we have no firm basis for knowing which parameterization is more nearly correct. Perhaps the most dramatic example of this dilemma is the mystery of cloud–radiation interactions.

Unfortunately, it is not known which of these parameterizations is the most realistic, or even if any of them captures the essential feedback processes of actual clouds. Furthermore, other GCM groups have obtained different results by trying other ways of incorporating cloud microphysical processes and their radiative interactions (e.g., Le Treut and Li, 1988, 1991; Roeckner *et al.*, 1987), in contrast to the approach that Senior and Mitchell (1993) followed. There is thus a clear need to intercompare these approaches with one another, and, even more importantly, to validate them against observations, so as to evaluate the strengths and weaknesses of each.

It is noteworthy that the sensitivity of model-simulated climates to changes in atmospheric CO_2 concentration has undergone major fluctuations in recent years. The equilibrium global average surface temperature change in response to a CO_2 doubling, based on GCM results from models developed in the mid-1970s, was typically between 2–3 °C. By the middle to late 1980s, the range of typical GCM sensitivities was between 4–5 °C. Nearly all of the increase in sensitivity could be traced to cloud–radiation interactions. More recently, several GCMs incorporating more complex cloud algorithms, including some feedbacks arising from cloud microphysical processes, have shown reduced sensitivity to changing greenhouse gas concentrations (Senior and Mitchell, 1993).

In the earlier models, clouds were treated in a very simplistic way, and their ability to undergo changes, and thus to influence climate variability, was limited. In some GCMs, in fact, clouds and their radiative properties were prescribed once and for all and then held constant, so that no feedbacks were possible. Later models, by contrast, featured clouds that could and did change their absolute amount and their height distribution in response to changes in atmospheric water vapor content (e.g., Slingo, 1987). As the simulated clouds changed, so did their ability to contribute to both planetary reflectivity, or albedo, and to the greenhouse effect.

One type of problem is to characterize clouds, once they are formed in GCMs, i.e., to determine their radiative properties. Many unanswered questions are tied to this type of problem. For example, does carrying cloud liquid water as a prognostic variable offer real advantages in terms of being able to specify cloud radiative properties realistically? Or is it feasible to specify these properties directly from the other large-scale GCM fields? Some recent work (e.g., Tselioudis *et al.*, 1992) suggests that it may be difficult or impossible to infer cloud radiative properties as simple functions of temperature and other variables carried explicitly by GCMs, but much research remains to be done on cloud characterization.

Another class of problems involves cloud formation. Here the goal is to develop parametric treatments that enable GCMs to simulate when and where clouds occur. In current GCMs, typical algorithms relate cloud amount to GCM variables such as relative humidity. Then partial cloud cover is handled by weighting clear-sky and overcast radiative calculations by the predicted cloud fraction. One promising alternative approach is to specify clouds stochastically, i.e., to develop parameterizations that yield probability distributions for variables such as the size and spacing of clouds (e.g., Malvagi *et al.*, 1993).

New theoretical tools have been developed to aid in validating parameterizations against observational data. One such tool is the single-column model (SCM) (Somerville, 2000). An SCM is a computationally efficient and economical one-dimensional (vertical) model, resembling a single column from a GCM grid (e.g., Iacobellis and Somerville, 1991a,b). The model contains a full set of modern GCM parameterizations of subgrid physical processes. To force and constrain the model, the advective terms in the budget equations are specified observationally (Randall *et al.*, 1996).

The trend of increased reliance on observational field programs, which can provide both satellite and *in-situ* data, together with the development of SCMs and other means of using these data, is a powerful combination (e.g., Iacobellis and Somerville, 2000). This trend holds great promise for improving our understanding of cloud–radiation processes and for the future improvement of the treatment of these processes in climate models (e.g., Lee *et al.*, 1997).

See also: Climate Feedbacks, Volume 1; Earth System Processes, Volume 1.

REFERENCES

Cess, R D and 19 others (1989) Interpretation of Cloud Climate Feedback as Produced by 14 Atmospheric General Circulation Models, *Science*, 245, 513–516.

Iacobellis, S F and Somerville, R C J (1991a) Diagnostic Modeling of the Indian Monsoon Onset, I: Model Description and Validation, *J. Atmos. Sci.*, 48, 1948–1959.

Iacobellis, S F and Somerville, R C J (1991b) Diagnostic Modeling of the Indian Monsoon Onset, II: Budget and Sensitivity Studies, *J. Atmos. Sci.*, 48, 1960–1971.

Iacobellis, S F and Somerville, R C J (2000) Implications of Microphysics for Cloud–Radiation Parameterizations: Lessons from TOGA-COARE, *J. Atmos. Sci.*, 57, 161–183.

Lee, W H, Iacobellis, S F, and Somerville, R C J (1997) Cloud–Radiation Forcings and Feedbacks: General Circulation Model Tests and Observational Validation, *J. Clim.*, 10, 2479–2496.

Le Treut, H and Li, Z X (1988) Using Meteosat Data to Validate a Prognostic Cloud Generation Scheme, *Atmos. Res.*, 21, 273–292.

Le Treut, H and Li, Z X (1991) Sensitivity of an Atmospheric General Circulation Model to Prescribed SST Changes: Feedback Effects Associated with the Simulation of Cloud Optical Properties, *Clim. Dyn.*, 5, 175–187.

Malvagi, F, Byrne, N, Pomraning, G, and Somerville, R C J (1993) Stochastic Radiative Transfer in a Partially Cloudy Atmosphere, *J. Atmos. Sci.*, 50, 2146–2158.

Randall, D A, Xu, K M, Somerville, R C J, and Iacobellis, S (1996) Single-column Models and Cloud Ensemble Models as Links Between Observations and Climate Models, *J. Clim.*, 9, 1683–1697.

Roeckner, E, Schlese, U, Biercamp, J, and Loewe, P (1987) Cloud Optical Depth Feedbacks and Climate Modeling, *Nature*, 329, 138–140.

Senior, C A and Mitchell, J F B (1993) Carbon Dioxide and Climate: The Impact of Cloud Parameterization, *J. Clim.*, 6, 393–418.

Slingo, J (1987) The Development and Verification of a Cloud Prediction Scheme for the ECMWF Model, *Q. J. R. Meteorol. Soc.*, 113, 899–927.

Smith, R N B (1990) A Scheme for Predicting Layer Clouds and their Water Content in a General Circulation Model, *Q. J. R. Meteorol. Soc.*, 116, 435–460.

Somerville, R C J (2000) Using Single-column Models to Improve Cloud–Radiation Parameterizations, in *General Circulation Model Development: Past, Present and Future*, ed D A Randall, Academic Press, New York, 641–657.

Tselioudis, G, Rossow, W B, and Rind, D (1992) Global Patterns of Cloud Optical Thickness Variation with Temperature, *J. Clim.*, 5, 1484–1495.

Clouds

Marcia B Baker
University of Washington, Seattle, WA, USA

Clouds cover two-thirds of the global surface. They re ect incoming solar radiation and inhibit outgoing longwave radiation at the top of the atmosphere, thus providing strong controls on the global and regional energy balances. The evolution of clouds is strongly in uenced by: (i), the moisture content, roughness, and radiative characteristics of the Earth's surface; (ii), the thermal and humidity structure of the atmosphere; and (iii), atmospheric particles and gases. Each of these factors may change as the Earth's climate changes, and each factor can, in turn, be modi ed by clouds. Thus, clouds and their environments are parts of one coupled system, and the responses of this system to global change must be considered in a uni ed way.

Traditionally, clouds were studied only by observation from the Earth's surface. Instrumented aircraft and balloons launched from the surface and aircraft became useful in directed eld experiments in the 1940s. Radar is now used routinely to monitor the evolution of cloud and precipitation particles, both from aircraft and surface stations. In the last decades, satellites carrying sophisticated sensors have become useful indicators of cloud distributions and temperatures, ice and water distributions, and particles and gases in clear air. These data are used in numerical cloud models that provide a great deal of information on cloud processes and serve as the basis for routine weather prediction. The cloud models consist of dynamic, thermodynamic, and particle evolution equations that are solved on grids, the spacing of which range from several meters to hundreds of kilometers. Cloud-related processes are responsible for large uncertainties in current global climate models, and different representations of these processes lead to quite different predictions of climate evolution.

*Clouds determine and are determined by their exchanges of heat, moisture, and chemical species with their environments. These exchanges can be perturbed through external factors, including land-use modi cation, and perturbations in circulation patterns and surface uxes that accompany changing climate. Due to the fact that few systematic records of cloud parameters over periods longer than about 40 years exist, little is known about the historical evolution of clouds and cloud impacts during the 19th and early 20th centuries. We are now much better equipped to study these processes, but we are at this point unsure what sign and magnitude and what geographical distribution of cloud-related effects to project for the future. We do know that clouds provide some of the dominant feedbacks in determining climate change (*see *Climate Feedbacks, Volume 1).*

CLOUDS AND PRECIPITATION

Figure 1 shows the most important cloud-related processes. Clouds form when moist air rises as a result of mechanical lifting, convergence, or positive buoyancy. The rising air expands and cools, causing the relative humidity to increase. This stage of cloud formation is driven by surface heat and moisture sources, as well as rising motion in the lower atmosphere. At a high enough relative humidity, vapor condenses to form liquid drops (at temperatures $T \geq 0\,^{\circ}\mathrm{C}$) and small ice particles (at lower temperatures). Drops form by condensation on small (submicron) solid and partially liquid particles, termed atmospheric aerosol particles, that are ubiquitous in the troposphere.

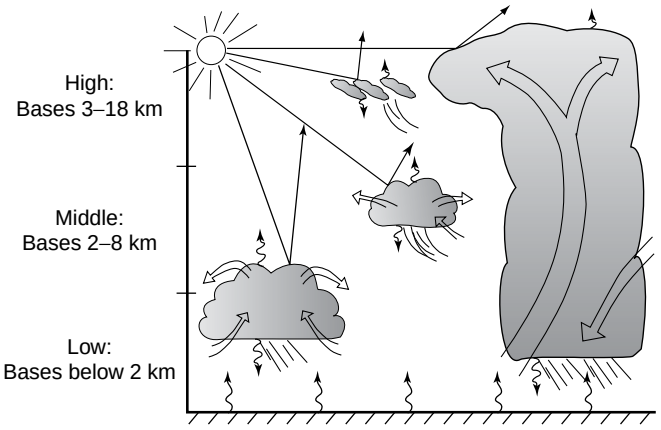


Figure 1 A schematic view of clouds, and of their interactions and environments. Thin lines indicate precipitation and hollow arrows indicate airflow, which brings environmental gases and particles into cloud. Wiggly arrows indicate infrared radiation, which clouds receive from Earth. They reemit this radiation back to Earth and to space. The lengths of the wiggly arrows indicate the relative magnitudes of the radiative fluxes. The thick lines indicate solar radiation, which is reflected by clouds. The lengths of these lines emanating from the cloud top suggest the relative magnitudes of solar reflectivity of different cloud types

The subset of aerosol particles that are large enough and soluble enough to produce cloud drops under normal atmospheric conditions are called *cloud condensation nuclei* or CCN. These are generally larger than $\sim 0.1\,\mu m$ in size, and consist of sulfates, salts, organic materials, black carbon, and minerals, in varying proportions, depending on their origins. In remote regions, typical droplet concentrations are around $10-100\,cm^{-3}$, whereas in polluted regions, the droplet concentrations can be on the order of $1000\,cm^{-3}$.

The $0\,^{\circ}C$ isotherm is normally at an altitude of $2-4\,km$ over much of the Earth s surface. The cloud must rise higher than this level in order to become partially or totally glaciated. Ice formation appears to be facilitated by certain aerosol particles termed ice nuclei (IN). Very little is known about the so-called IN; there are apparently very few of them relative to the number of CCN present in the atmosphere; at $-10\,^{\circ}C$, for example, there are only a few ice particles per liter. IN are thought to be largely composed of minerals (dust), but have been associated with oceanic sources in the Arctic and with biological debris. Anthropogenic processes appear to increase the concentration of IN under some circumstances and decrease it in others. It is possible that adsorption of atmospheric constituents, such as sulfates and aliphatic alcohols, can either *activate* or *deactivate* certain classes of atmospheric aerosols, rendering them more or less efficient as IN.

Both anthropogenic and *natural* modification of the chemical composition and/or concentrations of CCN and IN may modify the development and characteristics of clouds by modifying the evolution of the droplets and ice particles within them.

Precipitation Formation

When some of the droplets in a rising cloudy air parcel become large enough to fall through the updraft, they can collide with smaller, rising droplets below them. This process becomes effective when the droplets reach around $20\,\mu m$ in radius. For a given total amount of water, the droplets are larger and more likely to reach this size if the droplet concentration is relatively low. Coalescence of the colliding drops results in even larger drops that fall as *warm rain*. As a general rule of thumb, something like one million cloud drops must coalesce to form a single raindrop. The rain falls through the cloud and into the subcloud air, where it either evaporates or continues to the surface.

The effect of a few ice particles on precipitation in a supercooled cloud can be large. The equilibrium vapor pressure over ice is lower than that over liquid water at the same temperature, so that when they coexist, vapor flows from liquid to ice; the ice particles grow rapidly and deplete the liquid water. Collisions among these particles lead to precipitation in the form of rain (if the ice particles melt on their way to the surface), snow (aggregates of vapor-grown ice particles), sleet (freezing rain), or hail (rimed ice). Much of the precipitation in middle and high latitudes begins in the ice phase, although the hydrometeors can melt on the way down and fall to the surface as rain. Cloud lifetime is, in general, shortened by the development of ice, which removes water substance rapidly.

It is evident from the above that precipitation increases as a result of increasing uplift (through local surface heating, or convergence in the lower atmosphere, or orographically forced rise over mountains), increasing moisture content in the air (and thus, proximity to surface water, as well as increasing surface and atmospheric temperatures), relatively low concentrations of CCN (so the liquid water is divided among a few relatively large drops), and relatively high concentrations of IN (so that ice particles can form at midtropospheric temperatures and grow, sedimenting out of cloud base). Modification of the thermodynamic, chemical, or dynamic properties of the air in the cloud can have dramatic influences on precipitation formation. In turn, the formation of clouds and precipitation modifies the circulations, atmospheric temperatures, relative humidities, and the radiative fluxes below and above the clouds.

CLOUDS AND RADIATIVE FLUXES

Clouds and Shortwave Radiation

The fraction of downwelling solar radiation that is reflected back to space by clouds is called the *cloud albedo* (A_{cl}). The

A_{cl} generally ranges from about 0.4 to 0.8 for low stratiform clouds, increasing with cloud thickness, cloud liquid water content, and the number concentration of cloud drops. The albedo is usually assumed to be higher for water clouds than for ice clouds, all other parameters remaining constant, but if the ice particles are very small, the opposite is true. Thus, the cooling impact of clouds, which stems from the reflection of incoming solar radiation, increases with all the factors that increase cloudiness and with drop concentrations, and can decrease or increase with increasing ice fraction (for constant total condensate). Increased concentrations of anthropogenic aerosol particles and precursor gases can increase drop concentrations in warm clouds, and possibly, the concentrations of ice particles in colder clouds. This effect would increase the cloud albedo; however, in reality it is very difficult to quantify the anthropogenic effects on cloud albedo from large-scale observations because of the complexity of the clear air to cloudy air transition. However, in plumes and downwind of large population centers, the effect is clearly seen in satellite and aircraft measurements. There is some indication that the mean droplet size in low clouds in the Northern Hemisphere is around 1 μm smaller than that in the Southern Hemisphere, and models suggest that this difference would give rise to a net cooling of the Earth–atmosphere system on the order of $1\,W\,m^{-2}$ (about 25% of the magnitude of the warming due to the increase in CO_2 since the industrial revolution).

Clouds and Longwave Radiation

While the reflection of solar radiation by clouds cools the surface under the clouds, absorption and the subsequent reemission of terrestrial radiation by clouds warms the underlying surface, and lessens the flux of outgoing radiation at the top of the atmosphere. The impacts that clouds have on downwelling longwave radiation at the surface are felt by anyone who goes camping on cool nights; the approach of a cloud signals a rise in surface temperature.

The longwave radiative properties of a cloud can be expressed in terms of its emissivity, a number between 0 and 1 that depends on cloud properties and the wavelength of the radiation. The higher the emissivity, the more efficiently the cloud can absorb and emit radiation. The emissivities of clouds depend largely on the cloud temperature and total column water substance, and for thin clouds, the emissivity decreases as the mean dropsize increases, all other factors remaining equal.

Thick liquid water clouds tend to have emissivities near 1 at all wavelengths in the infrared; therefore, small perturbations in their properties do not change their longwave radiative impacts. On the other hand, the emissivities of thinner clouds (such as those that form in very cold regions and at higher levels) can be much less than 1. High CCN concentrations in these clouds could result in smaller drops for the same amount of condensed water; this might lead to increased cloud emissivities, thus increasing the greenhouse effect.

Clouds reflect downwelling solar radiation and emit infrared radiation to space, as well as back to the surface; therefore, they alter the fluxes of radiation at the top of the atmosphere, and they also modulate the fluxes of shortwave and longwave radiation at the Earth s surface. These cloud–radiation interactions determine the present climate and are subject to change as the climate changes (*see* Cloud- Radiation Interactions, Volume 1).

PRECIPITATION CHEMISTRY: ACID RAIN

In remote locations, the chemical composition of cloud water is dominated by the chemical content of the CCN and the absorption of atmospheric gases, such as CO_2 and NH_3. The predicted average pH of background cloud water is thus not 7, but approximately 4.5–6 i.e., it is slightly acidic. Atmospheric gases, such as SO_2, HNO_3, NO_x, and NH_3, and gaseous organics are all taken up to some extent by liquid water, which is the most efficient solvent known. In general, these decrease the pH of the cloud water further. The chemical composition of the rainwater produced exclusively by warm rain processes is usually dominated by the same ions as those in cloud water. The acidity of rain falling in and downwind of urban and industrial areas is often as low as 3–4. A general downward trend in the pH of rainwater in the north-east US, Eastern Canada, and Scandinavia appears to have ceased in the 1980s, possibly due to the control of emitted particles and SO_2 (which reacts with water to produce condensed sulfate). Contamination of rainwater by industrial emissions and by biomass burning remains a problem in many areas of the world.

The ice lattice, unlike liquid water, does not readily accept most non-water molecules. Therefore, the composition of precipitation formed in the ice phase is largely dominated by the *impurities* present in drops that freeze, and it has been shown in both urban and remote areas that the composition of snow that falls to the ground depends on the extent of riming. The composition of precipitation on the ground at any location depends on the trajectories of the air masses prior to precipitation; significant contamination from sulfates and trace metals has been found in newly fallen snow in the remote Alps and even in the high Caucasus Mountains.

CLOUDS IN URBAN AND AGRICULTURAL REGIONS

Anthropogenic effects on cloud and cloud processes have been observed for centuries over urban regions and those

intensively used for agriculture and grazing. In the last century, these effects have increased markedly. While the strictly urban effects are of limited importance globally, due to their small coverage of the planet, they give an indication of the kinds of impacts on cloudiness to be expected as the fraction of the Earth s surface modified by human activities continues to grow.

There are many (often conflicting) reports, based on surface measurements, of large changes in cloud-related quantities near urban areas. For example, low cloud cover, F_{cloud}, increased by 16.5% over Moscow in the years 1958–1993, together with surface temperature (increase by 1.3 °C), and a decrease of 6–7% in downwelling shortwave radiation. Recent satellite observations also suggest an urban effect on low cloud cover. The effect is due, in part, to perturbations in water or heat fluxes from the surface and partly to perturbations in the nature and concentrations of the aerosol particles in the air. There have been several studies that suggest increasing convective rainfall, thunderstorm activity, and hail production over and just downwind of highly urbanized areas during the 20th century. On the other hand, there are also a number of reports indicating decreased ice formation efficiency in polluted air masses. These changes in cloud properties are likely to be due to the anthropogenic modification of one or more of the parts of the cloud–environment system defined in the introduction.

Surface Moisture, Temperature, Albedo, and Roughness

Estimates of the impact of overgrazing alone suggest an increase of roughly 0.1 in the mean surface albedo due to overgrazing over the past 6000 years. This increase would have lowered the mean surface temperature by around 0.8 °C; larger effects have been reported in local and regional areas due to changes in land use. The reduction in surface moisture in these areas means that the surface heats more than in the damp surrounding countryside, a situation that enhances convergence and uplift over the dry area and subsidence outside it.

The results from mesoscale numerical models generally indicate that the accurate inclusion of heterogeneous soil moisture is of first-order importance in simulating cloud structures and movement over land surfaces. Agriculture, particularly when accompanied by surface irrigation and/or grazing, and deforestation modify surface heat and moisture fluxes, surface albedo, surface roughness, boundary-layer temperature structure, and the air chemistry. Gradients in vegetation type are thought to produce gradients in surface moisture and heat fluxes that modulate the mesoscale circulations associated with summer thunderstorms over the central US.

Stability of the Lower Atmosphere

While the air over cities tends to be drier, and therefore, the cloud bases tend to be higher than those over more moist surfaces, the convergence and enhanced updrafts over the city compensate for this effect, and cloudiness and rain often increase over and downwind of urban areas.

Chemical Composition of the Air

Emissions of pollutant gases and liquids modify the growth of droplets from the haze to the cloud droplet stage. Adsorption of such materials can change the droplet/vapor surface energy and droplet composition (i.e., equilibrium vapor pressure). These effects may increase droplet concentrations under some circumstances and decrease them in others. The rate of ice formation may also be sensitive to trace amounts of organic solutes or thin films on aerosol particles that modify the ice nucleating efficiencies. One possible mechanism for urban enhancement of storm activity might be the enhancement of ice formation by increased production of large CCN, and therefore, large drops, in urban boundary-layer clouds, or by enhanced IN activity. (Ice formation accelerates upward motion and precipitation evolution in general.) On the other hand, if CCN concentrations rise, the cloud droplets might be smaller, and the frequency and intensity of precipitation might decrease. Thus, it is not possible to make general predictions about the impacts of future anthropogenic changes in the atmospheric chemistry on cloudiness.

CLOUDS IN POLAR REGIONS

Clouds in polar regions are white like the underlying surface, and they tend to be associated with atmospheric temperature inversions so that their longwave emissivity (and not just their shortwave reflectivity) tends to be similar to that of the underlying surface. These properties make them very difficult to identify from satellite measurements. Nevertheless, the apparently small cloud-induced changes in both the reflected solar radiation and the emitted terrestrial radiation can have enormous impacts in polar regions. The Arctic is particularly susceptible to any perturbation of surface temperature, because the ensuing modifications in ice cover can then drive changes in cloud cover. Due to the complexities of these multiple interactions, the largest disagreements in models of global circulation are found in their predictions of cloud properties in the polar regions.

Surface Properties

Arctic clouds increase the downwelling longwave radiation and decrease the downwelling shortwave radiation at the surface, while they have opposite effects on the fluxes of

radiation leaving the system at the top of the atmosphere. Model results indicate that the sea-ice cover increases with increasing low cloud cover, and it increases with increasing cloudiness in summer, presumably because the clouds increase the system albedo. The predicted sea-ice cover also increases with high liquid to ice ratio in the clouds because, in general, liquid clouds have higher albedos than ice clouds. (On the other hand, as we have pointed out, if the ice particles are numerous and small, they may increase cloud albedo.) Also, the sea-ice cover decreases with increasing coverage by high clouds, the main radiative impact of which is to inhibit outgoing longwave radiation. Modifications in precipitation and in the cloud radiative fluxes can modify the temperature structure and stability of the ice, which, in principle, can modify the thermohaline circulation in the oceans by perturbing the freshwater flow to the sea. Moreover, the surface albedo decreases when sea ice melts, causing more heat to be absorbed by the surface. This amplifies the temperature increase. Model simulations of climate change due to increased greenhouse gas emissions indicate that, because of the response of the sea ice, the zonal mean temperature rise in the Arctic will be 2–3 times larger than the global temperature rise.

Structure of the Lower Atmosphere

The ice thickness and surface ice coverage determine the fluxes of heat and moisture from the surface into the lower atmosphere. These, together with fluxes from lower latitudes, determine the atmospheric temperature and humidity profiles to which the cloud cover responds.

In the winter, clouds are sparse in the Arctic, but the low-level atmosphere tends to hold *clear air ice precipitation*. This consists of ice crystals that apparently form at fairly low relative humidities, often in the plumes of moisture emanating from leads or cracks in the ice that expose liquid water to the air. The ice particles sediment to the surface, dehydrating the lower atmosphere. Although the clear air precipitation appears visually thin, it can contribute up to $80\,W\,m^{-2}$ to the downwelling longwave flux at the surface, which is equivalent to a heating rate of a few degrees per day.

Chemical Composition of the Air

Arctic clouds persist in summer and tend to be mainly liquid [for temperatures over $\sim 255\,K$ ($-18\,^{\circ}C$)]. It is thought that the CCN on which spring and summer Arctic cloud drops form are derived from sulfate aerosols. Some of the particles are produced photochemically from dimethyl sulfide emitted at the ocean surface. However, chemical analysis, showing that these particles contain black carbon and other elements typical of pollution sources, suggests that most are advected from midlatitudes in spring and summer. The advected aerosol particles are apparently inefficient at promoting ice formation, perhaps because sulfate particles coagulate with and deactivate dust or other potential ice nucleants; perhaps some other deactivation mechanism is present in the advected air masses. During winter, circulation patterns do not move air into the Arctic from lower latitudes as efficiently as in summer; the clouds tend to be completely glaciated. When a few ice particles form, they grow rapidly. Growth is faster at lower temperatures, causing the particles to fall from the clouds, taking most of the condensed phase water with them. Thus, cloud lifetime (and thus, cloud-related fluxes of heat and moisture, and therefore, the response of the underlying surface) depends sensitively on the ice/water ratio, which in turn depends on the particles and gases in the environment.

The results of these observational and modeling studies suggest that, if the mean surface temperature increases in the Arctic, the cloud cover and cloud lifetime might increase.

FURTHER READING

Abakumova, G, Feigelson, E M, Russak, V, and Stadnik, W (1976) Evaluation of Long-term Changes in Radiation, Cloudiness, and Surface Temperature on Territory of the Former Soviet Union, *J. Clim.*, 9, 1319.

Baker, M B (1997) Cloud Microphysics and Climate, *Science*, 276, 1072–1078.

Cotton, W R and Pielke, R A (1995) *Human Impacts on Weather and Climate*, Cambridge University Press, Cambridge, MA.

Houze, R A (1993) *Cloud Dynamics*, Academic Press, San Diego, CA.

Kolb, C E, Worsnop, D R, Zahniser, M S, Davidovits, P, Keyser, C F, Leu, M T, Molina, M J, Hanson, D R, Ravishankara, A R, Williams, L R, and Tolbert, M A (1995) Laboratory Studies of Atmospheric Heterogeneous Chemistry, in *Current Problems and Progress in Atmospheric Chemistry*, ed J R Barker, World Scientific Publishing, Singapore, 771–875, Vol. 3.

Randall, D A, Curry, J A, Battisti, D, Flato, G, Grumbine, R, Hakkinen, S, Martinson, D, Preller, R, Walsh, J E, and Weatherly, J (1998) Status of and Outlook for Large-scale Modeling of Atmosphere–Ice–Ocean Interactions in the Arctic, *Bull. Am. Met. Soc.*, 79, 197–219.

Rodhe, H and Herrera, R (1988) *Acidi cation in Tropical Countries*, John Wiley & Sons, Chichester, (SCOPE 36), 3–39.

Comets and Asteroids, Effects on Climate

see Asteroids and Comets, Effects on Earth (**Volume 1**); Tunguska Phenomenon (**Volume 1**)

Committee on Earth Observation Satellites (CEOS)

see CEOS (Committee on Earth Observation Satellites) (**Volume 1**)

Composition, Atmospheric

see Carbon Dioxide, Recent Atmospheric Trends (**Volume 1**); Air Quality, Global (**Volume 1**); Atmospheric Composition, Past (**Volume 1**); Atmospheric Composition, Present (**Volume 1**)

Continental Drift

Naomi Oreskes
University of California, San Diego, CA, USA

Continental drift is the motion of the Earth's continents over geological time. Since the 17th century, cartographers have noticed the jigsaw puzzle fit of the continental edges; in the 19th century, paleontologists discovered that some fossil plants and animals were extraordinarily similar across the globe. Some rock formations in distant continents were also surprisingly similar. To account for these similarities, Austrian geologist Eduard Suess proposed the theory of Gondwanaland: a giant supercontinent that had once covered the entire Earth surface before breaking apart to form continents and ocean basins. In the early 20th century, German meteorologist Alfred Wegener suggested an alternative explanation: continents drift. The paleontological patterns could be explained if the continents migrated, sometimes joining together, sometimes breaking apart.

Continental drift was not accepted when first proposed, but in the 1960s it became a cornerstone of the theory of global plate tectonics. Continental drift is now explained as a consequence of moving plates. The continents are embedded within the lithospheric plates that comprise the upper 80–100 km of the Earth, and are carried along with the plates as they migrate at average rates of 3–10 cm year^{-1}. As a result, the global configuration of continents and oceans is constantly changing. For several hundred million years during the late Paleozoic and the Mesozoic eras, the continents were united into a supercontinent called Pangea. The breakup of Pangea produced the configuration of the continents we have today.

Continental drift is a fundamental natural cause of global change. The major geological processes (earthquakes, volcanoes, mountain building) are caused by plate motions and interactions. These processes control the large-scale physiography of the globe, and with it, the distribution of habitat. In addition, many facts of evolution are explained as effects of continental drift. With the breakup of Pangea, previously unified populations began to diverge and speciate in response to their new environmental conditions. Other types, such as South American emus and African ostriches, retained obvious similarities even while physically segregated. A major part of paleoclimate change is also explained by continental drift, as the direct result of continents moving passively through climate zones, and the indirect impact of the reconfiguration of continents and oceans on oceanic circulation and climate patterns.

The development of the theory of continental drift and plate tectonics illustrates the tortuous process by which new scientific knowledge is established. The idea of globally moving continents was adamantly rejected when first widely debated in the 1920s, then established as scientific fact only 40 years later. Studying this history illuminates some of the reasons why scientific communities resist new ideas.

HISTORICAL BACKGROUND: THE ORIGIN OF MOUNTAINS AND CONTRACTION THEORY

One of the central scientific questions of 19th century geology was the origin of mountains. How were they formed? What process squeezed and folded rocks like putty? What made the Earth's surface move? Most theories invoked terrestrial contraction as a causal force. It was widely believed that the Earth had formed as a hot, incandescent body, and had been steadily cooling since the beginning of geological time. Because most materials contract as they cool, it seemed logical to assume that the Earth had been contracting as it cooled.

In Europe, Austrian geologist Edward Suess (1831–1914) popularized the image of the Earth as a drying apple: as the Earth contracted, its surface wrinkled to accommodate the diminished surface area. Suess assumed that the Earth's initial crust was continuous, but broke apart as the Earth's interior shrunk; the collapsed portions formed the ocean basins, the remaining elevated portions formed the continents. With further cooling, the continents became unstable and collapsed to form the next generation of ocean floor; what had formerly been ocean now became dry land. The interchangeability of continents and oceans explained the presence of marine deposits on land (which had long before puzzled Leonardo Da Vinci), and the extensive interleaving of marine and terrestrial materials in the stratigraphic record. Suess's theory also explained the widely known similarities of fossil assemblages in parts of India, Africa, and South America by attributing

them to an early geological period when these continents were still contiguous. He called this ancient supercontinent Gondwanaland.

In North America, a different version of contraction theory was developed by James Dwight Dana (1813–1895). Dana suggested that the Earth s continents had formed first, when minerals with relatively low-fusion temperatures such as quartz and feldspar had solidified. Then the globe continued to cool and contract, until the high temperature minerals such as olivine and pyroxene finally solidified: on the moon, to form the lunar craters, on Earth, to form the ocean basins. As contraction continued after the Earth was solid, it induced surface deformation. The greatest pressure was experienced at the boundaries between the oceanic and continental blocks, explaining the concentration of mountains along continental margins. Because continents and oceans were viewed as globally permanent features, Dana s account came to be known as permanence theory.

In North America, permanence was linked to the theory of geosynclines (subsiding sedimentary basins along continental margins) developed by Dana and James Hall (1811–1889), state paleontologist of New York and the first president of the Geological Society of America (1889). Hall noted that the Appalachian Mountains mostly consisted of folded sequences of shallow water sedimentary rocks, thousands of meters thick. How did thick sequences of shallow water deposits form? How were they folded and uplifted into mountains? Hall suggested that materials eroded off the continents and accumulated in the adjacent marginal basins, causing the basin to subside. Subsidence allowed more sediments to accumulate, causing more subsidence, until finally the weight of the pile caused the sediments to be heated, lithified, and then uplifted into mountains. Dana modified Hall s view by arguing that thick sedimentary piles were not the cause of subsidence but the result of it. Either way, the theory provided a concise explanation of how thick sequences of shallow water rocks could form, but was vague on the question of how they were transformed into mountain belts.

CONTINENTAL DRIFT AS AN ALTERNATIVE TO CONTRACTION THEORY

In the early 20th century the contraction theory was refuted by three independent lines of evidence. First, field mapping in the Swiss Alps and the North American Appalachians demonstrated hundreds of kilometers of shortening of strata. This would require impossibly huge amounts of terrestrial contraction to explain. Second, geodesists studying the problem of surface gravitational effects showed that the surface mass associated with mountains was counterbalanced by subsurface mass deficit. Mountains were held aloft not by their internal strength, but by floating: a concept called isostasy. Contra Suess, continents and oceans were not interchangeable. Third, physicists discovered radiogenic heat, which refuted the basis of contraction theory. With contraction no longer axiomatic, Earth scientists were motivated to search for other driving forces of deformation. Many did; Alfred Wegener (1880–1930) is the most significant, for his theory was the most widely discussed (*see* Wegener, Alfred, Volume 1).

Primarily known as a meteorologist and author of a pioneering textbook on the thermodynamics of the atmosphere, Wegener realized that paleoclimate change could be explained if continents had migrated across climate zones, and the reconfiguration of continents and oceans altered the Earth s climate patterns. However, continental drift was more than just a theory of paleoclimate change. It was an attempt to unify disparate elements of Earth science: on the one hand, paleontological evidence that the continents had once been connected; on the other, geodetic evidence that they could not be connected in the way European contractionists had supposed. Wegener s answer was to reconnect the continents by moving them laterally.

Wegener s theory was widely discussed in the 1920s and early 1930s. It was also hotly rejected, particularly by geologists in the United States who labeled it bad science. The standard explanation for the rejection of continental drift is the lack of a causal mechanism, but this explanation is false. There was a spirited and rigorous international debate over the possible mechanisms of continental migration. Much of it centered on the implications of isostasy: if continents floated in a denser substrate, then this substrate had to be plastic or fluid and continents could at least in principle move through it. The Fennoscandian rebound (the progressive uplift of central Scandinavia since the melting of Pleistocene glacial ice) provided empirical evidence that they did, at least in the vertical direction and at least in the Pleistocene period. However, in Scandinavia the cause of motion was generally agreed upon, first the weight of the glacial ice, then the pressure release upon its removal. What force would cause horizontal movement? Would the substrate respond comparably to horizontal as to vertical movement? Debate over the mechanisms of drift concentrated on the long-term behavior of the substrate and the forces that could cause continents to move laterally.

In the United States, the question was addressed by Harvard geology Professor Reginald A Daly (1871–1957), that country s strongest defender of continental drift. Daly argued that the key to tectonic problems was to be found in the Earth s layered structure. Advances in seismology demonstrated that the Earth contained three major layers: crust, substrate (or mantle) and core. The substrate, he suggested, might be glassy, and therefore could flow in response to long-term stress just as old plates of glass gradually thicken at their downward edges and glassy lavas flow downhill. Continents might do the same. Building on the geosyncline concept of Dana and Hall, Daly suggested

that sedimentation along the continental margins resulted in subtle elevation differences, which in turn produced isostatic instabilities. Eventually, the continent could rupture, sliding down over the glassy substrate under the force of gravity. The sliding fragment would then override the other half (an early suggestion of subduction) and, over time, the accumulation of small increments of sliding would result in global continental drift.

Daly admonished his American colleagues to take up the question of drift, but few did. Reaction in Europe was more favorable. Irish geologist John Joly (1857–1933) linked the question to discoveries in radioactivity. Trained as a physicist, Joly had demonstrated that pleochroic haloes in mica were caused by radiation damage from tiny inclusions of U and Th-bearing minerals, such as apatite. Radioactive elements were therefore ubiquitous in rocks, suggesting that radiogenic heat was also ubiquitous. If it were, then it could be a force for geological change. Joly proposed that as radiogenic heat built up, the substrate would begin to melt. During these periods of melting, the continents would move under the influence of small forces (such as small gravitational effects) that would otherwise be ineffectual. These were the periods of global orogeny (i.e., periods of mountain formation), such as the Appalachian/Caledonian that crossed Europe and North America in the Paleozoic and the Alpine/Dinaric era that crossed Europe in the Cenozoic age.

Joly s theory responded to a geophysical complaint against a plastic substrate: that the propagation of seismic waves indicated a fully solid and rigid Earth. He pointed out that although the Earth was solid now, it might not always have been. More widely credited was the suggestion of British geologist Arthur Holmes (1890–1965) that the substrate was partially molten or glassy. Underscoring arguments made by Wegener, Holmes emphasized that the substrate did not need to be liquid, only plastic, and that it might be rigid under high strain rates during seismic events yet still be ductile under the low strain rates that prevailed under most geological conditions. If it were plastic in response to long-term stress, then continents could move within it. Holmes s driving force was convection currents in the mantle. He argued that radiogenic heat would generate convection currents: the mid-ocean ridges were the sites of upwelling convection currents, where continents had split, and the ocean deeps (geosynclines) were the sites of downwelling currents, where continents were deformed as the substrate descended. Between the ridges and the trenches, continents were dragged along in conveyor-like fashion.

THE REJECTION OF CONTINENTAL DRIFT

Arthur Holmes s papers were widely read and cited; many geologists thought he had found the cause of continental drift. However, opposition was none the less strong, particularly in the United States, where reaction to Wegener s theory was vitriolic. More was at stake than a matter of scientific fact.

Three factors contributed to the American animosity to continental drift. One, Americans were widely committed to the method of multiple working hypotheses, and Wegener s work was interpreted as violating it. For Americans, the right scientific method was empirical, inductive, and required weighing observational evidence in light of alternative explanatory possibilities. Good theory was also modest, holding close to the objects of study. Most closely associated with the University of Chicago geologist T C Chamberlin (1843–1928), who named it, the method of multiple working hypotheses reflected American ideals expressed since the 18th century linking good science to good government: Good science was anti-authoritarian, like democracy. Good science was pluralistic, like a free society. If good science provided an exemplar for good government, then bad science threatened it. To American eyes Wegener s work was bad science: It put the theory first and then sought evidence for it. It settled too quickly on a single interpretive framework. It was too large, too unifying, too ambitious. In short, it was seen as autocratic. Features that were later viewed as virtues of plate tectonics were attacked as flaws of continental drift.

Second, continental drift was incompatible with the version of isostasy to which Americans subscribed. In the late 19th century, two accounts of isostatic compensation had been proposed: John Henry Pratt (1809–1871) attributed it to density variations, George Biddell Airy (1801–1892) attributed it to differences in crustal thickness. Until the early 20th century, there had been no empirical confirmation of the concept beyond the original evidence that had inspired it, nor any means to differentiate the two explanations. Then American geodesists John Hayford (1868–1925) and William Bowie (1872–1940) used Pratt s model to demonstrate that isostatic compensation was a general feature of the crust. By making the assumption of a uniform depth of compensation, they were able to predict the surface effects of isostasy to a high degree of precision throughout the United States. At first, their work was hailed as proof of isostasy in general, but in time it was viewed as confirmation of the Pratt model in particular. However, if continental drift were true, then the large compressive forces involved would squeeze the crust to generate thickness differences. Continental drift seemed to refute Pratt isostasy, which had worked for Americans so well. Rather than reject Pratt isostasy, they rejected continental drift.

Third, Americans rejected continental drift because of the legacy of uniformitarianism. By the early 20th century, the methodological principle of using the present to interpret the past was deeply entrenched in the practice of

historical geology. Many believed this was the only way to interpret the past: that uniformitarianism made geology a science because, for without it what proof was there that God hadn t made the Earth in seven days, fossils and all? Historical geologists routinely used faunal assemblages to make inferences about climate zones, but according to drift theory, continents in tropical latitudes did not necessarily have tropical faunas, because the reconfiguration of continents and oceans might change matters altogether. Wegener s theory raised the spectre that the present was not the key to the past: that it was just a moment in Earth history, no more or less characteristic than any other. This was not an idea American scientists were willing to accept.

In North America, the debate over continental drift was quelled by an alternative account of the faunal evidence. In 1933, geologists Charles Schuchert (1858–1942) and Bailey Willis (1857–1949) proposed that the continents had been intermittently connected by isthmian links, as the isthmus of Panama presently connects North and South America and the Bering Land Bridge recently connected North America to Asia. The isthmuses had been raised up by orogenic forces, then subsided under the influence of isostasy. This explanation was patently *ad hoc*: there was no evidence of isthmian links other than the paleontological data that they were designed to explain (away). Nevertheless, the idea was widely accepted, and a major line of evidence of continental drift undercut. In 1937, South African geologist Alexander du Toit (1878–1948) published *Our Wandering Continents*, a comprehensive synthesis of the geological evidence of continental drift, but it had little impact in North America. The matter rested there for two decades, until the debate was reopened on the basis of entirely new evidence.

FROM CONTINENTAL DRIFT TO PLATE TECTONICS

In the 1950s, continental drift was revived by British geophysicists studying rock magnetism as a means to understand the Earth s magnetic field, one group at Imperial College led by P M S Blackett (1897–1974), and one at Cambridge (later at Newcastle) led by S K Runcorn (1922–1995). Both groups found evidence that rocks had moved relative to the Earth s magnetic poles, so either the continents or the poles had moved. Initially geophysicists were more receptive to the idea of polar wandering, but by the late 1950s comparative evidence from India and Australia pointed in the direction of moving continents. Inspired by these results, American geologist Harry Hess (1906–1969) revived the idea earlier proposed by Arthur Holmes: that convection currents drove continental motions.

Hess suggested that mantle convection drives the crust apart at mid-ocean ridges and downward at ocean trenches, forcing the continental migrations in their wake. He interpreted the oceanic crust as a hydration rind on a serpentinized mantle; Robert Dietz (1914–1995) modified this to generate an oceanic crust by submarine basalt eruptions, and gave it the name it holds today: sea floor spreading. Dietz s interpretation was later confirmed by direct examination of the sea floor. Meanwhile, geophysicists had demonstrated that the Earth s magnetic field has repeatedly and frequently reversed its polarity. Magnetic reversals plus sea floor spreading added up to a testable hypothesis, proposed independently by Canadian Lawrence Morley and British geophysicists Frederick Vine and Drummond Matthews. If the sea floor spreads while the Earth s magnetic field reverses, then the basalts forming the ocean floor will record these events in the form of a series of parallel stripes of normal and reversely magnetized rocks. Since World War II, the United States Office of Naval Research had been supporting sea floor studies for military purposes, and large volumes of magnetic data had been collected. American and British scientists examined the data, and by 1966 the Vine and Matthews hypothesis had been confirmed. In 1967–1968, the evidence of drifting continents and the spreading sea floor was unified into a global framework. Working independently, Daniel P McKenzie and Robert L Parker at the Scripps Institution of Oceanography, and Jason Morgan at Princeton University, showed that existing data could be used to analyze crustal motions as rigid body rotations on a sphere. The result became known as *plate tectonics* (*see* Plate Tectonics, Volume 1); by the early 1970s it was the unifying theory of the Earth sciences.

Continental drift is now subsumed into global plate tectonics, but the problems it was designed to explain (global physiography, disjunctively distributed fauna, and paleoclimate change) are still explained by the same basic idea: Continents drift. So faunal assemblages are divided or united, climate patterns are altered, and the Earth s physiography is transformed.

LESSONS FROM CONTINENTAL DRIFT: THE TORTUOUS DEVELOPMENT OF SCIENTIFIC KNOWLEDGE

Most people believe that when the weight of evidence becomes sufficiently great, scientists will accept the reality of new phenomena and the truth of new theories. The case of continental drift suggests that reality is more complex. Scientific theories are not judged only in the light of evidence, but also in the light of methodological standards and epistemic preferences. Because these standards and preferences are forged prior to the onset of any theoretical debate, the legacies of past intellectual debates may weigh heavily in the outcomes of new ones.

The problem with continental drift was not a lack of evidence, nor a lack of causal explanation. The evidence presented by Wegener and du Toit has been largely confirmed by global plate tectonics, and Holmes s causal account (mantle convection) is now generally accepted as the cause of plate tectonics. The problem with continental drift was a conflict with prior intellectual commitments. Between the 1920s and the 1960s, these earlier commitments (to Pratt isostasy, to the method of multiple working hypothesis, to uniformitarianism) were loosened, modified, or abandoned altogether. With this, the debate could be reopened on the basis of new evidence, and a previously discarded idea resurrected.

See also: Earth System History, Volume 1.

FURTHER READING

Frankel, H (1985) The Continental Drift Debate, in *Resolution of Scienti c Controversies: Theoretical Perspectives on Closure*, eds A Caplan and A T Englehardt, Cambridge University Press, Cambridge, 312–373.

Glen, W (1982) *The Road to Jaramillo: Critical Years of the Revolution in Earth Science*, Stanford University Press, Stanford, CA.

Greene, M T (1992) *Geology in the 19th Century: Changing Views of a Changing World*, Cornell University Press, Ithaca, NY.

Le Grand, H E (1988) *Drifting Continents and Shifting Theories*, Cambridge University Press, Cambridge.

Marvin, U B (1973) *Continental Drift: The Evolution of a Concept*, Smithsonian Institution Press: Washington, DC.

Oreskes, N (1999) *The Rejection of Continental Drift: Theory and Method in American Earth Science*, Oxford University Press, New York.

Oreskes, N and Le Grand, E (2001) *Plate Tectonics: Stories of Discovery*, Columbia University Press, New York.

Stewart, J A (1990) *Drifting Continents and Colliding Paradigms: Perspectives on the Geoscience Revolution*, Indiana University Press, Bloomington, IN.

Strahler, A N (1998) *Plate Tectonics*, Geobooks Publishing, Cambridge, MA.

Wood, R M (1985) *The Dark Side of the Earth*, Allen and Unwin, London.

Convection

Roger A Pielke, Sr

Colorado State University, Fort Collins, CO, USA

The term convection refers to vertical mixing through the exchange of mass that is driven by buoyancy forces. This buoyancy occurs when the density of a uid parcel is different from the surrounding uid. In the oceans, for example, cold and saline water is denser than warm and fresh water. As a result cold water above warm water will vertically mix. Similarly, salty water over less saline water will vertically mix. In the atmosphere, a warm air parcel will rise and a cold air parcel will sink. In an unsaturated atmosphere, this convection is called dry convection*; while in an atmosphere that is partially or completely saturated it is referred to as* moist convection*. A third type, called* forced convection*, occurs when air blows over a rough surface or up a slope.*

Dry convection occurs in a dry atmosphere when the vertical decrease of temperature with altitude exceeds in magnitude the dry adiabatic lapse rate (1 °C per 100 m), the rate at which a rising parcel of dry air cools due to expansion (*see* Lapse Rate, Volume 1). When the ambient lapse rate is greater than this, a rising air parcel is warmer, and thus less dense than its surrounding air, and buoyancy forces accelerate its upward motion. In a saturated atmosphere, *moist convection* happens when the release of latent heat in a rising air parcel, combined with the effect of expansion, is sufficient to make it warmer than the surrounding air at the same altitude. This occurs when the temperature change with altitude outside the parcel (referred to as the environmental *lapse rate*) decreases more rapidly than the cloud cools as it ascends. The cloud cools less rapidly than a dry air parcel, since condensation of water vapor into liquid water, the freezing of liquid water, and the deposition of water vapor into ice introduce the heat of condensation, freezing and deposition, respectively.

This lapse rate, called the moist adiabatic lapse rate, is always smaller than the dry adiabatic lapse rate. At very cold temperatures, where the actual amount of water vapor in the air must be small, the moist and dry lapse rates become nearly identical. Moist convection is also often referred to as *cumulus convection*. The specific lapse rate of a real cloud as it ascends depends on its initial temperature, the actual amounts of water vapor in the air, and the amount of lateral mixing with the surrounding cooler and drier air. This lateral mixing is called *entrainment*.

Downward convection in the atmosphere can be initiated when rainwater evaporates. If the cooling by evaporation produces air parcels that are cooler than the surrounding air, vertical mixing will occur. When rainwater remains within the parcel, this downward convection is referred to as a *wet microburst*. The term *downburst* has also been used. If the rainwater is completely evaporated, yet the parcel remains colder than the environmental air, the term *dry microburst* is used. These microbursts are a hazard to aircraft landing and takeoffs, since the wind shear created as the convection reaches the surface can be quite large.

Clouds formed through convective processes are referred to as cumuliform clouds (*see* Clouds, Volume 1). Shallow cumulus clouds, as often occur on otherwise sunny days, are called *cumulus humulis*. They are also known as fair weather cumulus. Towering cumulus clouds that extend upwards to several kilometers or more are called *cumulus congestus*. Cumulus clouds, which extend to the tropopause, and have an anvil shape at their top, are referred to as *cumulonimbus*. The anvil forms as the cloudy air becomes colder or is of the same temperature as the surrounding clear air, as it impinges into the lower levels of the stratosphere. Cumulonimbus clouds are also called thunderstorms, since these clouds are electrified enough to produce lightning. Downward bubbles of cloud that often form below the anvils of thunderstorms are called *mammatus* clouds. These clouds are actually elevated wet downbursts.

Convection occurs in the atmosphere, lakes and the oceans. Changes in the atmospheric temperature lapse rate associated with environmental change would directly affect the intensity of both dry and moist convection. Alterations in the amount of fresh water inflow from rivers would change the intensity of convection in the oceans. The Younger Dryas period (which occurred early in the Holocene), for example, was a period of brief return to glacial climate conditions, as massive inflows of fresh water from melting glaciers caused suppression of vertical mixing in the North Atlantic, since the fresh water had a lower density than the underlying ocean water (*see* Younger Dryas, Volume 1). This reduced vertical mixing caused much colder temperature in the region, which subsequently cooled the entire Northern Hemisphere.

Atmospheric convection is strongly controlled by the energy and water balance at the Earth s surface. As a result, alterations in land surface conditions; such as due to landscape changes (deforestation, overgrazing, etc.) will directly influence dry and moist convection. Examples of studies that have explored these effects include Pielke *et al.* (1998); Pitman *et al.* (1999); Xue (1996); and Polcher and Laval (1994).

See also: Climate Feedbacks, Volume 1; Cloud- Radiation Interactions, Volume 1; Earth System Processes, Volume 1.

REFERENCES

Pielke, R A, Avissar, R, Raupach, M, Dolman, H, Zeng, X, and Denning, S (1998) Interactions Between the Atmosphere and Terrestrial Ecosystems: Influence on Weather and Climate, *Global Change Biol.*, 4, 461–475.

Pitman, A, Pielke, R A, Avissar, A, Claussen, M, Gash, J, and Dolman, A (1999) The Role of the Land Surface in Weather and Climate: Does the Land Surface Matter? IGBP Global Change Newsletter No. 39, September.

Polcher, J and Laval, L (1994) The Impact of African and Amazonian Deforestation on Tropical Climate, *J. Hydrol.*, 155, 389–405.

Xue, Y (1996) The Impact of Desertification in the Mongolian and the Inner Mongolian Grassland on the Regional Climate, *J. Clim.*, 9, 2173–2189.

Conveyor Belt

see Ocean Circulation (**Volume 1**); Ocean Conveyor Belt (**Volume 1**)

Coriolis Effect

Anders Persson

European Center for Medium-range Weather Forecasts (ECMWF), Reading, UK

As a consequence of the Earth's rotation, wind and ocean currents are de ected to the right in the Northern Hemisphere and to the left in the Southern Hemisphere. This de ection is attributable to a Coriolis force, a so-called ctitious force, which, like the equally ctitious centrifugal force, is essentially a consequence of inertia. The Coriolis de ection is always at right angles with the motion, which means that it cannot change the velocity and kinetic energy of a body (do work). Still the Coriolis force has a decisive in uence on many atmospheric and oceanic ow patterns of almost all scales, from the diurnal turning of the sea breeze circulation and the drifting of icebergs, to large-scale circulation systems like the Indian monsoon and the Gulf Stream.

THE HISTORY OF THE CORIOLIS FORCE

That the motion of the Earth has a profound effect on the motion of the atmosphere has been recognized since the 17th century, most famously by George Hadley (1685–1768). A general mathematical expression for the apparent motion within a rotating system was derived around 1800 by Pierre Simon Laplace (1749–1827), but he did not understand the implications for the atmospheric circulation. In the 1830s, the French mathematician and physicist Gaspard Gustav Coriolis (1792–1843) wanted to calculate the centrifugal force on a body that was not fixed to the rotating system, but moving within it. The

total centrifugal force C, or *compound* centrifugal force as Coriolis called it, could be calculated by adding to the normal centrifugal force ($C_0 = m\omega^2 R$), an extra force, the Coriolis force $m^2\omega V_r$ (where ω is the angular velocity of the rotation, R the distance to the axis of rotation, V_r the relative motion with in the rotating system and m the mass). This means that physically the Coriolis force is just an extra component of the centrifugal force for a body moving at the same time it is taking part in a rotation.

The relevance of the Coriolis effect to Earth sciences was not realized by Coriolis and his contemporaries. This recognition only came in the wake of the famous pendulum experiment by Jean Bernard Léon Foucault (1819–1868). Inspired by this experiment and Laplace s work, the American William Ferrel (1817–1891) concluded in 1856 that the direction of the wind is parallel to the isobars, and that its speed is dependent on the latitude and the horizontal pressure gradient. At the same time, but independently of Ferrel, the Dutch meteorologist C H D Buys-Ballot (1817–1890) published his empirical rule according to which low pressure is to the left if the wind is blowing into one s back. Neither Foucault, nor Ferrel and Buys-Ballot had read Coriolis 1835 paper. The first time his work was brought into Earth Sciences was in 1859 at the French Academy in a comprehensive discussion of the deflective force in relation to the problems associated with water currents in channels or rivers. But it took many years before Coriolis name started to appear in the meteorological literature. When meteorologists in the second half of the 19th century discussed the effects of the rotation of the Earth, it was the findings of Foucault, Ferrel and Buys-Ballot that were referenced. The name Coriolis force was not used until the early 1920s, and then was simply attached to a mechanical effect that had been explored by others.

BASIC MECHANISM

A quantitative measure of the horizontal deflection is fV_r where V_r is the velocity of the body and $f = 2\Omega \sin\varphi$, the so called Coriolis parameter, where Ω is the angular velocity of the Earth s rotation and φ the latitude. The horizontal deflection is strongest at the poles, zero at the equator. At 43° latitude, where $f = 10^{-4}\,s^{-1}$, a body moving with a speed of $10\,ms^{-1}$ will in less than a minute have deviated 1 m from its original course. Because the Coriolis force is always perpendicular to the motion, it tries to bring back the moving body from where it came by forcing it into a circular orbit (Figure 1). The role of the Coriolis force on the global system is presented in Atmospheric Motions, Volume 1; Ocean Circulation, Volume 1.

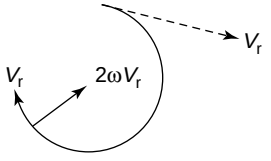


Figure 1 A body (of unit mass) moving with a velocity V_r in a system rotating with an angular velocity ω under no other force than the Coriolis force $2\omega V_r$ will follow a circular path with a radius of curvature $V_r/2\omega$

THE CORIOLIS FORCE ON A TURNTABLE

To acquire a physical understanding of the Coriolis force, consider a body m in a curved motion, like a vehicle taking a sharp bend. In the driver s frame of reference there will be (per unit mass) a centrifugal force C (Figure 2a) (Equation 1).

$$C = \frac{V^2}{R} \tag{1}$$

where V and R are the velocity and radius of curvature of the trajectory, both viewed from the fixed frame of reference.

A special, but common, case is a body, fixed to a rotating system, moving with a constant angular velocity ω under a central force in a circle with a radius R_0. In this case the total centrifugal force agrees with what is normally referred to as the centrifugal force C_0 (Figure 2b) (Equation 2).

$$C_0 = m\omega^2 R_0 \tag{2}$$

The next step is to consider the total centrifugal force on a body that is not fixed to the rotating system, but moving within it with a relative velocity V_r. The trajectory seen from the fixed frame of reference will then not necessarily be a circle (where Equation (2) applies), but a curved path (where the more general Equation (1) applies).

If the body moves radially *inwards* the trajectory, seen from the fixed frame of reference, it will follow an inward spiral with $R < R_0$, with R and C pointing in a different direction than R_0 and C_0 (Figure 2c). If the body moves radially *outward*, the trajectory in the fixed frame of reference will be an outward spiral with $R > R_0$ with R and C_0 pointing in a different direction to R_0 and C_0 (Figure 2d). The difference between the centrifugal force C for a moving body and C_0 for a fixed body, will in both cases be in the same direction relative to V_r (to the right for anticlockwise rotation) (Equation 3)

$$C - C_0 = 2\omega V_r \tag{3}$$

This is also true if the body moves tangentially. If the motion is in the direction of the rotation, the centrifugal force C will increase; if the body moves against the direction of the rotation, the centrifugal force C will decrease (Figure 2e). The difference $C - C_0$ will again for both

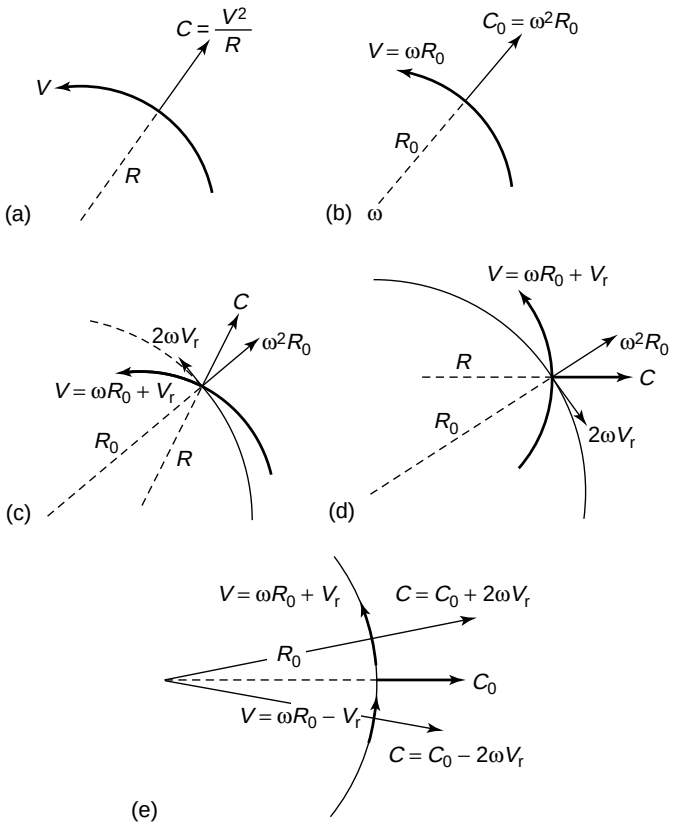


Figure 2 (a) The total centrifugal force is always directed perpendicularly to the trajectory of the body moving in a curved path in a fixed frame of reference; (b) for a body stationary in a rotating system, the trajectory in a fixed frame of reference is a circle and the total centrifugal force is directed radially out from the center of rotation; (c) for a body moving inward with velocity V_r, the trajectory is an inward spiral and the total centrifugal force is pointing outward but not exactly radially; (d) for a body moving outward with velocity V_r, the trajectory is an outward spiral and the total centrifugal force is pointing outward but not exactly radially; (e) for a body moving tangentially, the total centrifugal force is pointing radially outwards, but is weakened or strengthened, respectively, depending on whether the motion is against or with the rotation

cases be in the same direction relative to V_r (to the right for anticlockwise rotation). On a turntable any body will be accelerated away from the center of rotation more or less independent of whether the body is moving on it or not. This does not happen to air, water or any other mass element on the Earth, because the Earth s gravitation acts as a balancing force.

THE CORIOLIS FORCE ON A ROTATING PLANET

During the formation of the Earth, the centrifugal force due to the Earth s rotation shaped it into an oblate ellipsoid. A balance was reached when the combination of the centrifugal force C and the gravitational force g^* became perpendicular to the Earth s surface. The resultant of C and g^* is the weight of a body, g, called the effective gravity (Figure 3a). We can also consider the vertical and horizontal components of C and g^*. The vertical components C_V and g_V^* make up the effective gravity g, whereas the horizontal components C_H and g_H^* balance each other (Figure 3b).

However, this balance $C_H + g_H^* = 0$ with g perpendicular to the Earth s surface is only valid for stationary bodies; a non-stationary object will, due to its motion, change the total centrifugal force C in the way discussed above: east–west motions (relative to the rotating system) will change it through a change of the absolute velocity, a south–north motion will change it because of a change of the radius of curvature (both in magnitude and direction). Thus, the resultant of g^* and C will no longer be perpendicular to the Earth s surface but have components in the vertical and horizontal directions.

The modification of effective gravity g in magnitude and direction of a body due to its motion relative to the Earth s surface is the physical mechanism behind the Coriolis effect on a rotating planet. For motion westward, it is g_H^*, the horizontal component of gravitation, which has become stronger than the horizontal component of the centrifugal force C_H, that will physically displace the body towards the poles, to the right of the relative motion V_r. A body

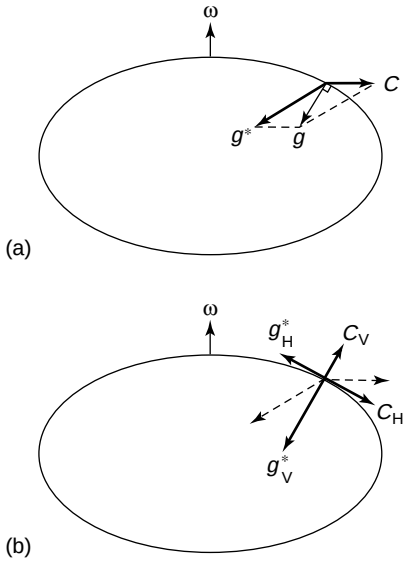


Figure 3 (a) The combined effect of the centrifugal force C and gravitational force g^* affecting a stationary body, results in the force of gravity g which is perpendicular to the earth's surface. (b) The earth's centrifugal force C and gravitational force g^* decomposed into horizontal and vertical components. C_V and g_V^* constitute effective gravity. The horizontal components C_H and g_H^* balance each other provided the body is stationary

moving westward also becomes heavier, moving eastward lighter, due to the vertical Coriolis deflection. This so-called *Eötrös effect*, named after the Hungarian physicist Lorand Eotros (1848–1919), has long been known to geodesists measuring the Earth s gravitational field.

The Coriolis force in vector form is $-2\omega \times V_r$ where the cross multiplication reflects the fact that the Coriolis force is strongest for motions perpendicular to the axis of rotation and is zero for motions parallel to it (Figure 4).

The angular velocity ω for the Earth s angular velocity should not relate to the solar day (24 hours) but the sidereal day (23 hours and 56 minutes) because this represents the rotation versus the fixed stars. By using the sidereal day instead of the solar day, we take effectively into account the Coriolis effect due to the orbiting of the Earth around the Sun!

Due to the latitudinal variation of the Coriolis force, a body moving under inertia will therefore not follow a perfectly circular but a slowly westward translating spiral (Figure 5). This is because the radius of curvature is shorter at higher latitudes than at lower ones. This so-called *beta-effect* plays an important role in the dynamics of the atmosphere and oceans.

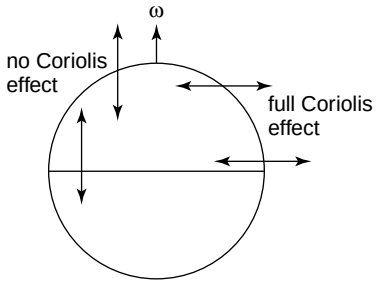


Figure 4 The Coriolis force affects all motions with maximal strength when they are perpendicular to the Earth's axis. When the motion is parallel to the axis there is no Coriolis deflection

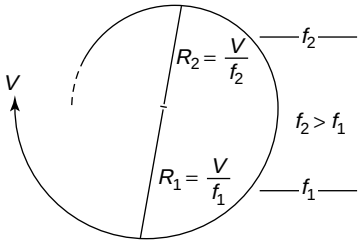


Figure 5 The variation of the Coriolis force makes an inertial motion slowly spiral westward (the beta-effect). A body moves with a velocity V and is deflected at a northerly latitude where the Coriolis parameter is f_2 with a radius of curvature R. Arriving at a southerly latitude with Coriolis parameter $f_1 < f_2$ the radius of curvature. $R_1 > R_2$ and the body will not exactly follow a circular path

Cretaceous

The *Cretaceous* is the last of three geologic periods in the Mesozoic Era (225–65 Ma), lasting 71 million years from 136 to 65 Ma, and is the final period of the existence of the dinosaurs. The Cretaceous Period follows the Jurassic Period and is followed by the Tertiary Period in the Cenozoic Era, which follows the Mesozoic (*see* Earth System History, Volume 1). The origin of the name, originally coined by William D Conybeare (1787–1857) and William Phillips (1775–1828), is the clay (*creta* in Latin) found in the White Cliffs of Dover and throughout Northern Europe (Prothero, 1990). The beginning of the Cretaceous period is marked by the separation of North America from Eurasia (Laurasia), and South America from Africa (Gondwanaland), while the uplift of the Rocky Mountains began towards the end of the Cretaceous. Sea level was, on average, over 300 m higher than present, mainly due to global tectonic processes (Southam and Hay, 1981). The resulting paleogeography led to a near-equatorial current in the so-called Tethys Sea, covering large parts of western Europe and North Africa, as well as a large, shallow interior sea in western North America, extending from the Gulf of Mexico throughout the western US and Canada (Crowley and North, 1991). The end of the Cretaceous, known as the K–T (Cretaceous [C stands for Carboniferous]–Tertiary) boundary, occurred catastrophically with an asteroid impact that is thought to have wiped out the dinosaurs and 75% of all living species. Geological evidence shows that the mid-Cretaceous was warmer than present in high latitudes and possibly ice-free. Modeling studies have explored various reasons for this enhanced high latitude warmth, including land–sea configuration, ocean heat transport, land–surface processes such as vegetation, and CO_2 levels. In all of the simulations, higher CO_2 levels are needed to produce the required warmth. There is some evidence that higher CO_2 levels in the mid-Cretaceous may be the result of greater volcanism due to enhanced sea-floor spreading and reduced continental area, which leads to less weathering and therefore less uptake of CO_2 (Crowley and North, 1991).

REFERENCES

Crowley, T J and North, G R (1991) *Paleoclimatology*, Oxford University Press, New York, 339 [Crowley and North summarize the results of many modeling studies].

Prothero, D R (1990) *Interpreting the Stratigraphic Record*, W H Freeman, New York, 1–410.

Southam, J R and Hay, W W (1981) Global Sedimentary Mass Balance and Sea Level Changes, in *The Sea*, Vol. 7, ed C Emiliani, Wiley Interscience, New York, 1617–1684.

BENJAMIN S FELZER *USA*

Crutzen, Paul J

(1933-)

Paul J Crutzen was born in Amsterdam, The Netherlands, in December 1933. He first studied civil engineering and worked at the Bridge Construction Bureau in his native city between 1954 and 1958. He then moved to Sweden where he received a degree in Meteorology (1968) and a few years later his doctoral degree (1973). His thesis work was performed in part at Oxford University in the UK. In 1974, Paul Crutzen moved to the United States and became a research scientist at the National Center for Atmospheric Research (NCAR) in Boulder, CO, and at the Aeronomy Laboratory of the National Oceanic and Atmospheric Administration, also in Boulder. In 1977, he was appointed as the Director of the Air Quality Division of NCAR. In 1980, he returned to Europe and became a member of the Max Planck Society for the Advancement of Science, and the Director of the Atmospheric Chemistry Division at the Max Planck Institute for Chemistry in Mainz, Germany. Crutzen held several part-time teaching positions including a professorship at the universities of Chicago (1987–1991), Stockholm (1991–1992), California at San Diego (since 1992), Mainz (since 1993), and Utrecht (since 1997).

Paul Crutzen has made a large number of innovative contributions in many areas of atmospheric chemistry. In the 1970s, he studied the photochemical mechanisms affecting ozone in the stratosphere and established that the most efficient ozone destruction mechanism is provided by a catalytic cycle involving the presence of nitrogen oxides (*see* Stratosphere, Chemistry, Volume 1). This discovery led H Johnston at the University of California to suggest that the emission of nitrogen oxides by a projected fleet of high-altitude aircraft would destroy large amounts of stratospheric ozone.

Crutzen also showed that the presence of ozone in the troposphere was not due only to the intrusion of ozone-rich stratospheric air masses or the formation of ozone in urban areas, but that on the global scale, the photooxidation of methane and carbon monoxide catalyzed by the presence of nitrogen oxides could produce substantial quantities of background tropospheric ozone. This finding suggested that photochemical pollution was not limited to industrial areas, but was potentially global in nature. Crutzen suggested that an important source of ozone precursors (including carbon monoxide and nitric oxide) was provided by biomass burning in the tropics during the dry season.

Together with John Birks at the University of Colorado, Crutzen pointed out that a large number of intense fires could be triggered by the explosion of nuclear bombs, should a nuclear war occur. Such perturbation would not only alter dramatically the chemical composition of the atmosphere, but it would considerably modify the climate and produce for the entire population of the Earth what Crutzen and Birks named a *nuclear winter* (*see* Nuclear Winter, Volume 3). In the 1980s, Crutzen contributed to the understanding of the chemical mechanisms involved in the formation of the Antartic ozone hole (*see* Ozone Hole, Volume 1).

Paul Crutzen has received honorary doctoral degrees from several universities: York (Canada), Louvain (Belgium), East Anglia (UK), San José (Costa Rica), Liège (Belgium), Tel Aviv (Israel), Oregon State (US), Burgundy (France), and Athens (Greece). He is a member of the Royal Swedish Academies of Sciences and of Engineering, a foreign Associate of the US National Academy of Sciences, a Foreign member of the Italian Academy of Sciences, a corresponding member of the Royal Netherlands Academy of Science, and a founding member of the Academia Europaea. Crutzen is the recipient of numerous prestigious awards including the Tyler Prize for the Environment (1989), the Volvo Environmental Prize (1991), the German Environmental Prize of the Federal Foundation for the Environment (1994), and the United Nations Environment Program (UNEP) Global Ozone Award (1995). In 1995, together with Drs M Molina (*see* Molina, Mario J, Volume 1) and F S Rowland (*see* Rowland, F Sherwood, Volume 1), Crutzen was awarded the Nobel Prize in Chemistry for their contributions to understanding the stratospheric ozone layer and the cause of ozone depletion.

GUY BRASSEUR *Germany*

Cryosphere

The term *cryosphere* traces its origins to the Greek word *kryos* for frost or icy cold. It is used collectively to designate regions of the Earth s surface where water is in a solid form. The components of the cryosphere include: sea ice, lake ice, river ice, snow cover, glaciers, ice caps and ice sheets, and seasonally frozen or perennially frozen (permafrost) ground. Seasonally, snow and ice cover between 8 and 15% of the Earth s surface. Snow cover has the largest areal extent of any component of the cryosphere, with a mean maximum areal extent of

approximately 47 million km^2 in January–February, mostly located in the Northern Hemisphere. Sea ice covers much of the polar oceans and some adjacent seas. In the Arctic Ocean there is perennial ice cover (*see* Arctic Climate, Volume 1), whereas around Antarctica the ice is mostly seasonal. Sea ice has highly variable thickness, as new ice forms in openings and level ice is compressed into pressure ridges (*see* Sea Ice, Volume 1). Freshwater ice forms on rivers and lakes in response to seasonal cooling. Ice on rivers may form ice jams that block the flow and cause flooding prior to spring breakup. Ground seepages and springs in northern land areas commonly freeze to form surface icings. Permafrost (perennially frozen ground) may occur where mean annual air temperatures (MAAT) are less than -1 or $-2\,^{\circ}$C and is generally continuous where MAAT are less than $-7\,^{\circ}$C. It underlies 24% of exposed Northern Hemisphere land areas. Thicknesses exceed 600 m along the Arctic coast of northeastern Siberia and Alaska, but toward the margins, permafrost becomes thinner and horizontally discontinuous. Land ice occurs in the two large ice sheets of Greenland and Antarctica, smaller ice caps and mountain glaciers (*see* Greenland, Volume 1; Antarctica, Volume 1). The ice sheets hold approximately 77% of global freshwater, corresponding to 80 m of world sea-level equivalent. Of this, Antarctica accounts for 90%, Greenland nearly 10%, other ice bodies and glaciers less than 0.5%. In Antarctica, large ice streams extend out into the ocean, forming floating ice shelves that are the primary source of Southern Ocean icebergs. The cryosphere is an integral part of the global climate system with important linkages and feedback generated through its influence on surface energy and moisture fluxes, clouds, precipitation, hydrology, and atmospheric and oceanic circulation. Deep permafrost also contains methane gas hydrates trapped in the ice lattice (clathrates) which could be a significant long-term source of additional greenhouse gas burden for the atmosphere.

ROGER G BARRY *USA*

D

Dansgaard–Oescheger Cycles

Dansgaard-Oescheger (*D-O*) *cycles* are rapid warm–cold cycles that occurred on millennial time scales during the last glaciation, which began in earnest around 75 000 years ago, and reached its maximum extent about 21 000 years ago. The D-O cycles are sawtooth-shaped cooling cycles comprised of long cold periods (*stadials*) followed by abrupt warming (*interstadials*) of up to 5 °C. Most, but not all of the cool events were terminated by what is called a Heinrich event (Bond *et al.*, 1993) (*see* Heinrich (H-) Events, Volume 1).

See also: Earth System History, Volume 1.

REFERENCE

Bond, G, Broecker, W, Johnsen, S, McManus, J, Labeyrie, L, Jouzel, J, and Bonani, G (1993) Correlations Between Climate Records from North Atlantic Sediments and Greenland Ice, *Nature*, 365, 143–147.

BENJAMIN S FELZER *USA*

Desertification

see Desertification (**Volume 3**); Desertification, Definition of (**Volume 3**)

Deserts

Martin Williams
Adelaide University, Adelaide, Australia

The word desert comes from the Latin verb deserere, meaning to forsake or abandon, as in the expression deserted. The Arabic word Sahara also implies a vast and empty wilderness, bereft of life. A subtle and hidden implication of the term desert is that these now dry areas have been deserted, but were once able to support more abundant life. There is growing concern today that human actions are contributing to the spread of desert-like conditions in previously fertile and well-vegetated land, a process termed desertization or deserti cation. The causes of deserti cation are complex and include droughts, human mismanagement, and the aftermath of war. Two thousand years ago (ka), the great Roman historian, Tacitus, wrote caustically of the scorched earth policies favored by some of the Roman emperors and their generals: "Ubi solitudinem faciunt pacem appellant (They create a desert, and call it peace)." Once again, the use of the Latin word solitudinem *(solitude) in this context connotes emptiness and an absence of life.*

WHAT IS A DESERT?

A desert is a region where the rainfall is too little and too erratic and the evaporation too high to allow many plants and animals to survive except in a few favored localities. Over very long intervals of time the plants and animals living in desert areas have become well adapted through their physiology and behavior to using scarce water efficiently. The sparse human populations in deserts have also evolved long-term behavioral adaptations to the harsh extremes of desert climate, particularly through a nomadic life style designed to make optimum use of sporadic rains and ephemeral grazing. However, low precipitation is a necessary but not a sufficient cause of aridity. In certain cold areas of the world the rates of evaporation may be low enough to compensate for the low rates of precipitation, allowing a relatively dense plant cover and sometimes even a wetland flora to exist in spite of low overall amounts of annual precipitation. In such instances, the effective precipitation is said to be high.

Some definitions of a desert are essentially economic, with a desert being defined as a region where viable agriculture is not possible without irrigation. If we accept this working definition, then roughly 36% of the land area of the globe is either arid or semi-arid. Despite their aridity, these

lands provide a home for about 15% of the world s present human population. It is also worth noting that, although deserts share a number of common attributes or diagnostic characteristics, each desert is unique, and reflects the subtle interplay between local biophysical influences, including rock type, tectonic history, climate, and biota. This account will deal with desert conditions in all continents except Antarctica.

WORLD DISTRIBUTION OF DESERTS

The distribution of the world s major deserts (Figure 1) is closely linked to latitude and to distance from the sea. The great deserts of the Sahara and Arabia and the Rajasthan Desert of India with its counterpart the Cholistan Desert of Pakistan lie astride or close to the Tropic of Cancer. The deserts of Australia, the Kalahari and the Atacama are likewise traversed by the Tropic of Capricorn. The deserts of Central Asia, including the Taklimakan and Gobi Deserts of China and Mongolia, are situated in the interior of mid-latitude continental regions. A number of deserts are also located in the rain-shadow of high mountain ranges such as the Andes, the Rockies, the Himalayas and the Altai, Tien Shan and Kunlun Ranges in Central Asia. Why is this so?

CAUSES OF PRESENT-DAY ARIDITY

There are four main reasons why deserts have such little rain (Abrahams and Parsons, 1994). The two most important factors are location in latitudes dominated by dry subsiding air and location inland far from sources of moist maritime air. The remaining two factors are location in the rain-shadow of high mountain ranges and location on a coast flanked by cold ocean currents or cold upwelling ocean water. The first and fourth factors are a direct product of the global atmospheric circulation system. Atmospheric circulation is determined by solar radiation modulated by latitude, by the distribution of land and sea, and by the topography of the land.

Solar radiation is greatest at the equator because the sun is almost directly overhead there for much of the year. Away from the equator, progressively more incoming radiation is reflected or absorbed by the earth s atmosphere since the sun s rays travel ever more obliquely through the atmosphere as a result of the curvature of the earth. Because of the tilt of the earth s axis, the sun is directly over the Tropic of Cancer/Capricorn once a year, at the summer solstice. If the axial tilt were greater, the sun would appear to travel farther from the equator during summer, and the converse would apply if the tilt were less.

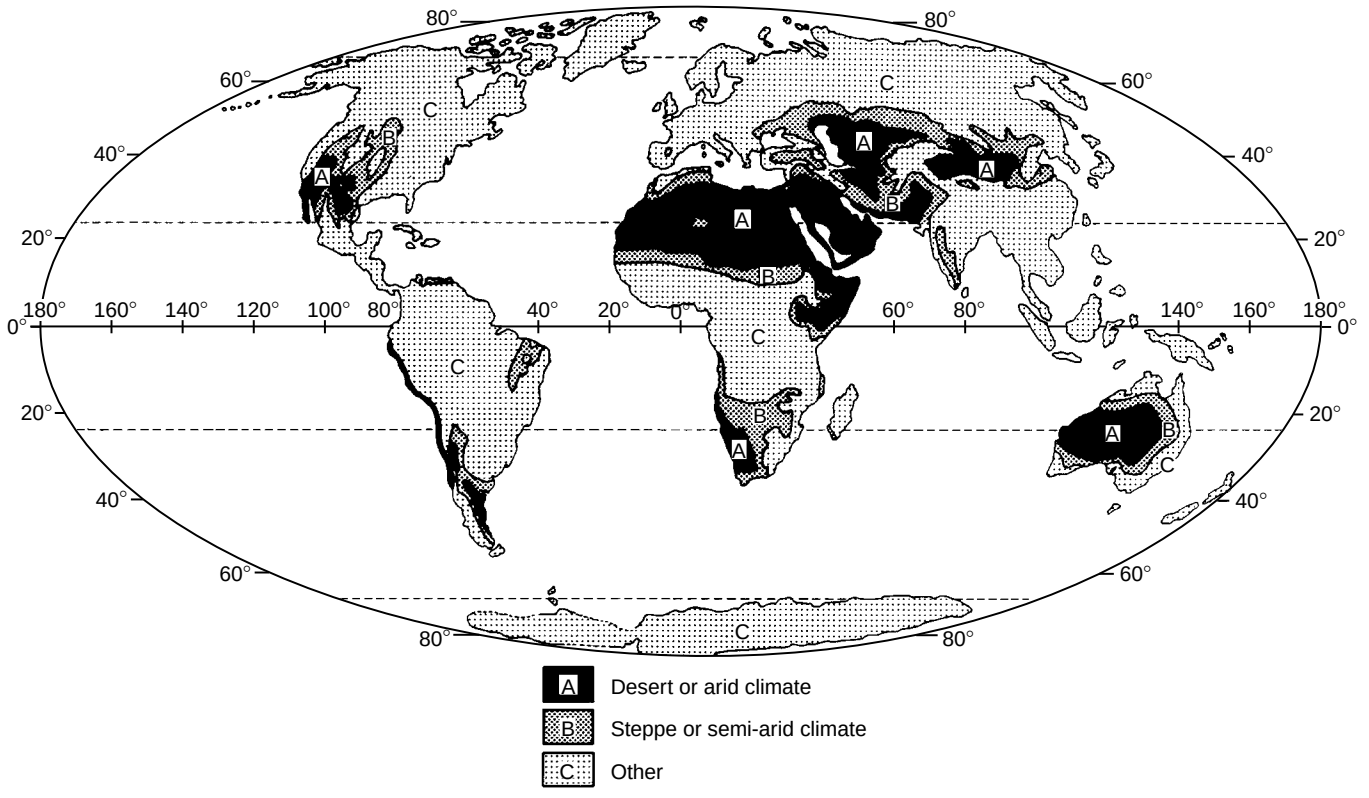


Figure 1 Present-day distribution of tropical and temperate deserts and semi-deserts. (Reproduced from Williams *et al.*, 1998)

Hot tropical deserts like the Sahara and Arabia are in latitudes where the air aloft is dry and subsiding, and the atmospheric pressure is high for much of the year. The surface winds in deserts are therefore generally directed outwards, towards areas of lower atmospheric pressure, so that there is minimal inflow of moisture from surface winds. As the air over the deserts subsides it is compressed and becomes warmer, so that its capacity to absorb additional water vapor is increased. The result is that the relative humidity of desert air is usually very low. The dew point (100% relative humidity) is only reached when the night temperatures fall sufficiently for desert dew to precipitate on chilled rock surfaces. This ephemeral dew allows some desert antelopes to survive despite the lack of surface water. Shortly before sunrise small herds of gazelle may sometimes be seen licking the dew from the surface of small piles of desert rocks in North Africa and Arabia, a behavioral strategy that enables them to survive in otherwise waterless conditions.

The tropical anticyclonic deserts are a direct result of the atmospheric circulation cells (often termed Hadley cells) located between the equator and the Tropics of Cancer and Capricorn, so that their location is determined by latitude rather than by the regional distribution of land and sea. The two polar deserts also come under the influence of semi-permanent anticyclones and of cold dry subsiding air. Because the distribution of high pressure cells (anticyclones) is closely related to latitude, the oceans in both polar and strictly tropical latitudes receive very little precipitation, and are the arid marine counterparts of the continental deserts.

The second major cause of aridity is a geographical location sufficiently far inland to be away from the influence of moist maritime air masses. Rainfall decreases rapidly away from the coast in all parts of the world except those close to the equator. Distance inland is sometimes described as *continentality*, and naturally applies to all big deserts, including the great tropical deserts of Arabia, Australia and the Sahara. In the case of these hot tropical deserts, the effects of continentality accentuate those of latitude. Other examples of continental interior deserts are the great mid-latitude deserts of central Asia, Mongolia and Western China. Bitterly cold in winter, with temperatures falling as low as minus 40 °C, summer temperatures can rise to 50 °C, which is nearly as hot as the southern Libyan Desert in July and August.

Two other factors may often enhance the aridity resulting from latitude and continentality, or may be the direct and dominant cause of reduced precipitation. These two factors are the proximity of cold oceanic water immediately offshore, and the rain-shadow effect generated by high mountains. They may operate individually or together.

The presence close offshore of cold upwelling water or a cold ocean current is an effective cause of coastal aridity in tropical and even in equatorial latitudes such as the arid Horn of Africa. The cold Peru current flows north parallel to the coast of the Atacama Desert in northern Chile and the coastal desert of Peru, and the cold Benguela current flows north parallel to the Namib Desert in southern Africa. The cold West Australian current likewise flows north parallel to the arid west coast of Australia, but the situation here is more complex, with the warm Leeuwin current flowing somewhat erratically from the Indonesian Warm Pool to the north, to counteract the desiccating effect of the cold West Australian current. In fact, the western borders of all the great tropical or trade wind deserts in both hemispheres are washed by cool ocean currents associated with the oceanic circulation cells or gyres which flow clockwise in the Northern and anticlockwise in the Southern Hemisphere. If cool moist maritime air blows onshore, it often meets a land surface that is warmer than the adjacent ocean surface, at least in summer and during the day. The cool maritime air mass becomes warmer on contact with the warm surface of the land. The relative humidity of this air mass is therefore decreased, and its ability to absorb additional moisture from surface evaporation is increased. The air therefore has a desiccating effect upon the land. This situation is only reversed if the land temperatures become significantly cooler than those of the adjacent ocean or if the sea surface temperatures become periodically warmer, as happens off the coast of Peru during El Niño years. Otherwise, the major source of moisture in these often quite narrow coastal deserts are the coastal fogs that blow inland in winter when the land has cooled down relative to the sea surface temperatures. Coastal fogs are quite common in deserts where mountain ranges like the Andes or the Rockies or uplands of more moderate elevation like the Red Sea Hills run parallel to and close to the shore. For example, Erkowit in the Red Sea Hills of the eastern Sudan is a mist oasis and supports a spectacular flora of tall *Euphorbia candelabra* trees in the dry valleys between its rocky granite hills.

The fourth and final general cause of aridity is the rain-shadow effect, which is a global phenomenon linked to topography and is not restricted to deserts. Wherever ranges of hills or mountains lie close to the coast, forming a physical barrier to onshore winds, the incoming moist maritime air will be forced upwards. Moist air cools adiabatically as it rises, attains vapor saturation, and sheds its condensed water vapor as rain or snow. The air then passes over the coastal ranges and flows downhill, becoming warmer and drier. The country inland of the coastal ranges is said to be in the *rain-shadow* of the ranges, a somewhat poetic term for lack of rain. The inland-facing slopes of high mountains are almost invariably drier than the coastal foothills, hence the great aridity of the Tibetan Plateau in contrast to the extreme wetness of the Assam foothills of the Himalayas. The wind-swept upland plains of the Bolivian Altiplano

and of Patagonia lie in the rain-shadow of the Andes. The rain-shadow deserts of New Mexico and Arizona are situated downwind of the Rockies. Other examples are the very gently undulating semi-arid western plains of Queensland and New South Wales inland of the Eastern Highlands of Australia, and the exceptionally hot, dry and rugged Afar Desert bounded by the Ethiopian highlands to the south and west. Both the Afar Depression and the Dead Sea Rift are flanked by very high mountainous escarpments and occupy low-lying fault-troughs or rifts that in places descend as much as 150 m below sea level. If the region inland of the coastal mountains contains high mountains, these will be a focus for some additional orographic or relief rain, but if the area is lacking in relief, there will be no opportunity for any such precipitation. The high gravel plains of the southern Libyan desert (serir) and the sandstone plateaus (hamada) of the central Sahara, the Gobi plains of northern China and Mongolia and the gibber plains of central Australia are all good examples. On these stony desert surfaces, rainfall and runoff are at a minimum and plants and animals are exceedingly rare, even by desert standards.

EVIDENCE OF FORMERLY WETTER CLIMATES IN NOW ARID AREAS

After considering why deserts are arid it is pertinent to ask whether they have always been so. The answer is unequivocally no. Scattered across the Sahara Desert are the silicified trunks of tall trees that once grew in abundance in the forests that covered the Sahara over 100 million years ago. The prehistoric hunters who roamed the Sahara at intervals during the last million years made good use of this silicified fossil wood to fashion the stone tools they used to hunt the great herds of savanna animals that also inhabited this now empty region. Later still, Neolithic pastoralists grazed their brindled herds of cattle, sheep and goats at numerous localities throughout the Sahara (Frostick and Reid, 1987). They left behind them an enduring legacy of superbly executed color paintings and engravings on the smooth rock faces of the Tassili sandstone plateau in Algeria, the A r mountains in Niger and the sandstone plateaus and granite massifs of the Libyan and Egyptian deserts. Another indication of formerly wet climates in the drier parts of Africa, Asia and Australia is the ubiquitous presence of deeply weathered and chemically altered bedrock. To achieve such a degree of intense leaching and new mineral formation requires considerable rainfall, a relatively dense vegetation cover, and very low rates of physical denudation – conditions more akin to the wet tropics than the arid tropics.

Perhaps the most striking evidence of previously wetter conditions are the remains of the once integrated river systems that used to flow through every major modern desert, providing the sandy alluvium that was later fashioned by wind into the imposing sand seas and associated dunes popularly considered synonymous with deserts. Indeed, one of the most characteristic features of all deserts is their lack of a perennial and integrated system of drainage (Cooke *et al.*, 1993). Desert streams are ephemeral. They flow episodically, for variable distances, depending upon the intensity and duration of sporadic rainstorms in their upper catchments. Even great rivers like the Nile, the Tigris and the Euphrates, which flow through deserts, originate in well watered uplands far beyond those deserts. All rivers that flow through deserts constantly lose water by evaporation and by seepage to the local aquifers. Most desert rivers never reach the coast and flow into closed depressions like the Tarim basin in China or the Lake Eyre basin in Australia. Such rivers are termed endoreic, in contrast to exoreic rivers like the Nile.

The fossil river valleys of the Sahara (Figure 2), the Gobi and Western Australia have long provoked the curiosity of geologists. Today they are broad linear depressions filled with Cenozoic alluvium that is often cemented with iron, silica or calcium carbonate. Some of these former valleys now form low erosional remnants or even extensive sheets of resistant ferricrete, silcrete or calcrete. In Mauritania, Namibia and Western Australia these valley-fill calcretes may also contain variable amounts of secondary uranium minerals precipitated out of slowly moving groundwater originating from the Pre-Cambrian host rocks, which form the valley interfluves.

EVIDENCE OF PREVIOUSLY GREATER ARIDITY AND DESERT EXPANSION

Just as now-arid areas retain evidence of once wetter climates, so is the converse equally true. Along the now-vegetated and stable margins of all the great deserts there is abundant evidence of former aridity including former desert dunes, salt lake and evaporite deposits, and now vegetated mantles of desert dust. Some care is needed in using such evidence to reconstruct past aridity, particularly in the case of desert dunes.

Desert dunes currently occupy about one-fifth of the Sahara and nearly two-fifths of the Australian arid zone. In North Africa, the 150 mm isohyet is a good indicator of the boundary between active and vegetated dunes (Figure 3), although precipitation is not the only factor responsible for dune mobility. Sand supply, wind velocity, surface roughness and evaporation rates also have an important influence on the movement of sand grains, as the presence of active coastal dunes in relatively wet areas with strong winds and abundant sand attests. Another exception to the general rule that dunes cease to be mobile once the rainfall exceeds about 150 mm are the source-bordering dunes that form downwind of sandy channels in semi-arid areas. There are three main prerequisites for the formation of

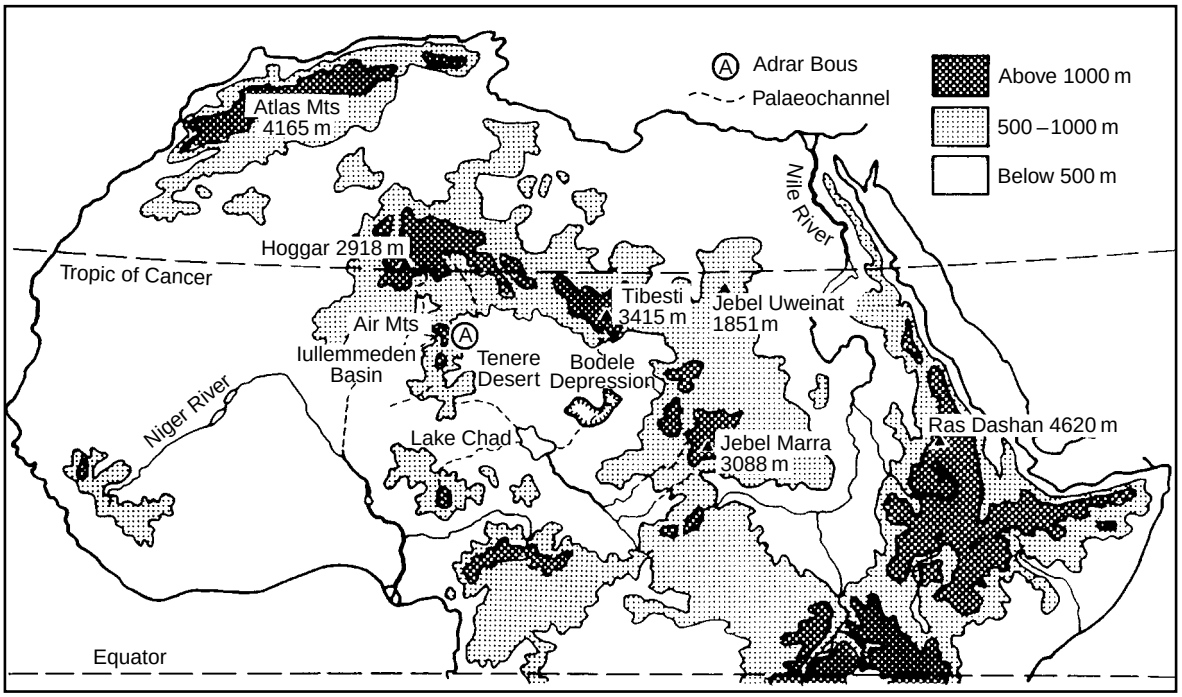


Figure 2 Distribution of major uplands in the Sahara

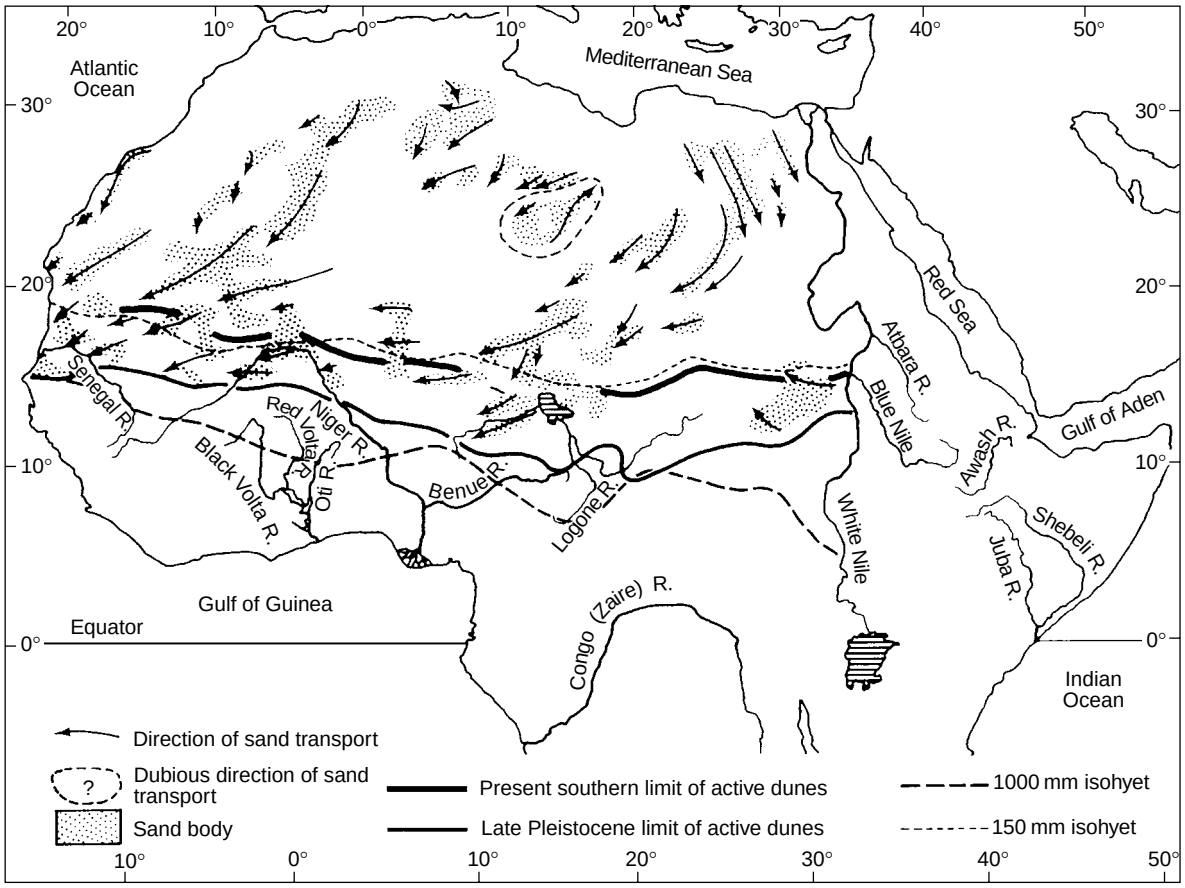


Figure 3 Distribution of active and stable sand dunes in the Sahara and its southern margins. (Reproduced from Williams *et al.*, 1998)

source-bordering dunes. First is a regular, usually seasonal replenishment of river channel sands or sandy beaches by long-shore drift in deep lakes. Second is a strong seasonal unidirectional wind and third a lack of riparian or lake-margin vegetation. The first prerequisite, a regular renewal of the sand supply from seasonally active rivers, precludes a fully arid climate.

Bagnold s classic observations in the Libyan desert and his detailed experimental work demonstrated that the volume of desert sand transported by wind increased exponentially with wind velocity above a certain threshold value (Pye and Tsoar, 1990; Cooke *et al.*, 1993; Lancaster, 1995). Where sand supply and wind velocity are not limiting factors, dune mobilization will increase as vegetation cover decreases. Rainfall and evapotranspiration are the primary controls over plant cover in arid areas. There is therefore a close relationship between the amount of rainfall and the average outer limit of active dunes in such deserts as the Great Indian Desert or the Sahara, where the southern limit of presently mobile dunes coincides remarkably closely with the 150 mm isohyet (Figure 3). The belt of fixed and vegetated dunes along what are now the semi-arid margins of these two deserts has been mapped in detail from air photographs and satellite imagery. Assuming that the relationship between rainfall and dune mobility held in the recent past, then the presence of these fixed dunes indicates that the effective range of the Sahara once extended 400–600 km farther south and that of the Rajasthan Desert some 350 km to the southeast.

ANTIQUITY OF DESERT LANDFORMS

Desert landscapes are made up of very young depositional landforms and very old erosional landforms (Frostick and Reid, 1987; Abrahams and Parsons, 1994; Thomas, 1997).

Figure 4 Ténéré desert dunes from summit of Air Mountains, Republic of Niger. (Photo: J Desmond Clark, January 1970)

Figure 5 Jebel Kassala, a desert inselberg or monolith, NE Sudan. (Photo: Martin Williams, December 1971)

The young landforms include dunes (Figure 4), alluvial fans, salt lakes and alluvial channels. The old landforms include mountains, hills and plateaus. It would be misleading to assume that the landform assemblage that is found in deserts had developed entirely under arid conditions. Most of the major erosional landforms formed under previously wetter climates and have been preserved from further erosion by the onset of aridity. Many desert landforms are exceedingly old. The vast desert plains of the central Sahara and Western Australia have been exposed to subaerial denudation for well over 500 million years. Desert monoliths such as Ayers Rock (Uluru) in Central Australia, or the granite inselbergs of the Sahara (Figures 5 and 6), far from being diagnostic of aridity, owe their present morphology to prolonged and repeated phases of weathering and erosion under a succession of former climates, few of which were particularly arid.

The abrupt juxtaposition of very ancient erosional landforms and very young depositional landforms gives desert landscapes their peculiar and somewhat paradoxical character, together with the absence of vegetation and the sharp breaks of slope. These young landforms and sediments, whether aeolian or fluviatile or lacustrine, contain the best record of past environmental changes, most notably the rapid climatic fluctuations of the last few million years.

LINKS BETWEEN CENOZOIC TECTONISM, GLOBAL COOLING AND DESICCATION

The onset of late Cenozoic aridity some 10–15 million years ago and the resultant slow emergence of the deserts were a result of global tectonic events that led to changes in global atmospheric circulation linked to changes in the global distribution of land and sea. For example, the origin of the Sahara as a desert was associated with several independent tectonic events. Slow northward movement of

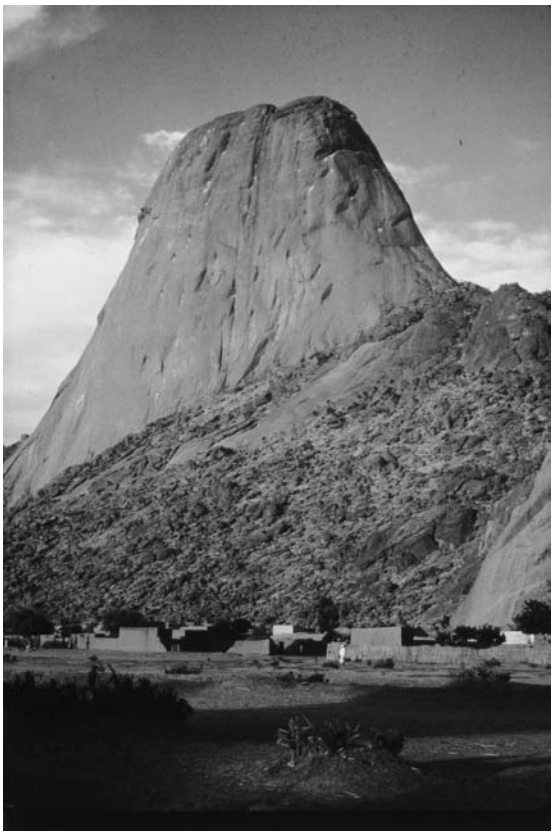


Figure 6 Jebel Kassala, a desert inselberg or monolith, NE Sudan. (Photo: Martin Williams, December 1971)

the African plate during the last 100 million years (late Mesozoic and Cenozoic) resulted in the migration of North Africa from wet equatorial into dry tropical latitudes. A slight clockwise rotation of Africa began over 60 million years ago (60 Ma) and continued through the Miocene and Pliocene, bringing Africa into contact with Europe. This displacement was accompanied by crustal deformation and rapid uplift in the Atlas region, and by volcanic eruptions and gentle updoming in Jebel Marra (3042 m), Tibesti (3415 m), the Hoggar (2918 m) and the A r Mountains (Figure 2). Owing to their altitude, the high mountains of the central and southern Sahara have always been wetter than the surrounding desert plains and so may have served as refugia for plants, animals and humans throughout the Quaternary.

Two additional factors were responsible for the late Cenozoic desiccation of North Africa. One was the gradual expansion of continental ice in high latitudes associated with the cooling of the Southern Ocean. The separation of Australia from Antarctica some 45 Ma ago culminated in the establishment of a large ice cap on Antarctica by 15 Ma ago. Closure of the Panama Isthmus and diversion of warm water into the North Atlantic allied to high northern latitude cooling provided the impetus for a sudden increase in the volume of Northern Hemisphere ice caps towards 2.5 Ma ago. One effect of the progressive build-up of high latitude ice sheets was to steepen the temperature and pressure gradients between the equator and the poles, resulting in increased trade wind velocities. Faster trade winds were better able to mobilize the alluvial sands of an increasingly dry Sahara and to fashion them into desert dunes. For example, the first appearance of wind-blown quartz sands in the Chad basin is towards the end of the Tertiary, when they occur interstratified among late Pliocene to early Pleistocene fluviatile and lacustrine sediments. The associated lacustrine diatom flora indicate temperatures cooler than those now prevalent in this region. The combined evidence suggests that the late Pliocene was both cooler and drier along the tropical borders of the Sahara. The diatom and pollen evidence from a large late Pliocene lake in the south-eastern uplands of Ethiopia is also consistent with the inference that intertropical cooling and desiccation may have been closely bound up with the expansion of Northern Hemisphere ice caps towards 2.5 Ma ago.

A further influence contributing to the drying out of the Sahara was late Cenozoic uplift of the Tibetan plateau and the ensuing creation of the easterly jet stream which brought dry subsiding air to the incipient deserts of Pakistan, Arabia, Somalia, Ethiopia and the Sahara. Isotopic analysis of fossil soils and fossil herbivore teeth collected from the Potwar Plateau of Pakistan indicates a major change in flora and fauna between 7.3 and 7.0 Ma ago. Until about 7.3 Ma ago, forest and woodland dominated the landscape. After 7.0 Ma ago, there was a rapid expansion of tropical grassland at the expense of forest. This change in vegetation may indicate the inception (or strengthening) of the Indian summer monsoon 7 Ma ago.

Other factors have probably contributed to intertropical cooling and desiccation during the past 30 Ma. One was the progressive shrinkage of the Paratethys Sea. This warm shallow sea once stretched across Eurasia but shrank gradually during the Oligocene and Miocene. As the once extensive sea shrank, the rainfall that was previously well distributed throughout the year became progressively more seasonal.

A further agent of late Cenozoic cooling was the decrease in atmospheric carbon dioxide (CO_2) associated with increased erosion, weathering and associated consumption of CO_2 caused by the late Cenozoic uplift of the Himalayas, the Rockies, the Andes, the Ethiopian uplands and perhaps also the Transantarctic Mountains. The global increase in plants using C_4 photosynthesis (*see* C_3 and C_4 Photosynthesis, Volume 2) and the reduction in C_3 plants between about 8 and 6 Ma ago is certainly consistent with a decrease in the concentration of atmospheric CO_2. The threshold for C_3 photosynthesis is higher at warmer latitudes, and so it is not surprising that the initial change from C_3 to C_4 plants was in the lowland tropics first. Climatic cooling was

probably also triggered by the eruption of the voluminous Ethiopian flood basalts over a period of no more than a million years towards 30 Ma ago.

Changes in the Cenozoic flora and fauna of the Sahara show a similar trend to that inferred for the Himalayan foothills of Pakistan. During the Palaeocene and Eocene much of the southern Sahara was covered in equatorial rainforest, and there was widespread deep weathering at this time. During the Oligocene and Miocene, much of what is now the Sahara was covered in woodland and savanna woodland, but by Pliocene times many elements of the present Saharan flora were already present. Pollen preserved in scattered localities in northern Africa shows that replacement of tropical woodland by plants adapted to aridity was already under way during the late Miocene and early Pliocene, a conclusion consistent with the pollen evidence preserved in deep sea cores off the north-west coast of Africa.

From about late Pliocene times onwards, the great tropical inland lakes of the Sahara, Ethiopia and Arabia began to dry out. The formerly abundant tropical flora and fauna of the well-watered Saharan uplands became progressively impoverished as entire taxa became extinct, and a once integrated and efficient network of major rivers became increasingly obliterated by wind-blown sands. In the Chad basin there is good evidence of wind-blown desert dune sands deposited between alluvial and lacustrine sediments that were laid down well over two million years ago. Farther north, in the Tibesti and Hoggar mountains of the central Sahara, the evidence from fossil pollen grains shows that some of the plants growing in this region were already adapted to aridity at about the time that the desert sands made their first appearance in the Chad basin. In Central China the first appearance of wind-blown desert dust has been reliably dated to *circa* 2.4 Ma ago. In Central Australia the inception of aridity seems to have started much later, with major inland lakes persisting until at least a million years ago before giving way to desert playa lakes and dunes fashioned from wind-blown quartz and gypsum sand-sized particles. Farther afield, in the present Congo/Za re basin of Central Africa, there are much older deposits of red desert sands that pre-date the oldest sands of the present Kalahari desert. Allowing for local differences in timing linked to regional climatic and tectonic factors, a similar sequence of events is also true of the deserts of China, India, Australia and southern Africa.

Three conclusions may be drawn from this brief survey. First, the aridity of the Sahara, northern China and Central Australia pre-dates any human presence in those regions and so humans cannot be considered the cause of these deserts, as some have claimed. Second, the inception of aridity was not everywhere synchronous, suggesting that local influences need to be considered. Third, in some presently humid areas there is evidence of former aridity just as some previously wetter areas are now arid.

DESERTS AS GEOMORPHOLOGICAL, ARCHAEOLOGICAL AND PALAEOCLIMATIC MUSEUMS

Deserts contain superb evidence of past climatic events. The very aridity to which they owe their existence also promotes the preservation of landforms, soils and sediments formed under quite different climatic conditions. Many desert landforms reflect the influence of weathering and erosional processes that are seldom active today. Every desert contains dry or saline lakes that once belonged to major exoreic river systems. The salt lakes of Western Australia are all that remain of the great rivers that flowed when Australia was joined to Antarctica over 45 Ma ago. Likewise, the fossil fauna of great deserts like the Sahara and the Gobi evoke distant echoes of a time long gone when dinosaurs roamed these once luxuriantly vegetated lands. In more recent times, the Sahara was host to great herds of elephants and giraffes, and a widespread aquatic fauna of turtles, hippos, crocodiles and Nile perch. The contrast with the present waterless wilderness could hardly be more stark. Some elements of this savanna fauna survive rather precariously. In sheltered valleys in the rugged A r massif in Niger, near the southern margin of the Sahara, there are remnant populations of baboons (*Papio anubis*) and patas monkeys (*Cercopithecus patas*), which must have reached these desert mountains during times of wetter climate when the West African savanna woodland was more extensive than it is today. They probably made use of riparian forest corridors that grew along once permanent rivers flowing south from the mountains. The Awash River that flows from the Ethiopian highlands down into the Afar Desert is a possible modern analogue. The dwarf crocodiles that lived in some of the permanent waterholes in the Tibesti Mountains of the south-central Sahara in the 1950s but have since been hunted to extinction were part of this palaeoclimatic legacy.

INTERACTIONS BETWEEN PREHISTORIC PEOPLES AND THE DESERTS

Scattered throughout the Sahara are abundant remains of the stone tools left behind by the Early, Middle and Late Stone Age peoples who once roamed the Sahara during wetter climatic intervals. The Late Stone Age hunters preyed upon the savanna herbivores and were also gifted artists, leaving behind them an enduring legacy of rock engravings and rock paintings depicting the animals they knew so well.

With the onset of plant and animal domestication by small groups of Neolithic herders and farmers some

10 000 years ago, the focus of the paintings changed. Cattle camps showing herds of brindled cattle guarded by men with bows and arrows and dogs were painted on suitable smooth rock faces in mountainous areas throughout the Sahara. Some of these paintings show women in their finery, riding oxen just as they do today among the Baggara cattle-owning Arabs of the western Sudan during the summer migrations north into the desert. Others show papyrus or reed canoes similar in design to those still made and used on Lake Tana near the Ethiopian headwaters of the Blue Nile. The plains adjacent to high mountains were preferred occupation sites for these Neolithic pastoralists. Small lakes and permanent springs were a guarantee of survival in years when the summer rains failed. The cattle herders ranged as far as Jebel Uweinat in south-east Libya, the Tassili sandstone plateau in southern Algeria and the A r massif in Niger (Figure 2). The nature of the interactions between prehistoric peoples and their environment remains the subject of enduring archaeological enquiry. Given the former presence of domesticated cattle, sheep and goats in areas no longer able to sustain them, it is tempting to speculate that they themselves may have accelerated their exodus from the desert. An obvious question to ask is to what extent did Neolithic overgrazing by large herds of hard-hoofed cattle, sheep and goats accelerate soil erosion by wind and water, and initiate humanly induced processes of desertification, especially in the drier second half of the Holocene?

GLACIAL ARIDITY OR GLACIAL PLUVIAL?

Nineteenth century geologists working in the semi-arid inter montane basins in the US identified and mapped the shorelines of a series of former very large lakes. The close association between glacial moraines and high-level shorelines or strandlines showed that the lakes were high during glacial episodes. The notion, that glacial climates were wetter than today, was quickly adopted in Europe. During the first half of the 20th century former high lake levels discovered by European and American geologists in the deserts of Asia and Africa were regarded as pluvial lakes formed during glacial times. The notion of glacial pluvial climates became solidly entrenched and the so-called pluvial chronology was even used to provide a relative chronology for prehistoric sites in East Africa, with each alleged pluvial being equated with one of the glaciations identified in the European Alps. The far-traveled coarse alluvial deposits of great rivers like the Nile were also interpreted as having originated during glacial pluvial climates.

Although there were dissenting voices, it was not until the high strandlines of the East African Rift lakes and the abundant remnants of former lakes scattered across the Sahara were directly dated in the late 1960s, that the glacial pluvial concept was finally abandoned for the arid and semi-arid tropics. The African high lake levels were found to be 9000 rather than 18 000 radiocarbon years old, that is, of early Holocene rather than of Last Glacial Maximum age. The alternative hypothesis, glacial aridity, was rapidly adopted and by 1975 there was growing acceptance that late Pleistocene intertropical aridity had been synchronous in both hemispheres.

It is now widely accepted that during times of maximum glaciation the tropical deserts were even drier than they are today and during the interglacial phases they were somewhat wetter. For example, during the last glacial maximum, dunes were active well beyond their present limits, and considerable volumes of desert dust were deposited downwind of the desert margins in Central Asia and China, northern India, northern Nigeria and south-eastern Australia (Pye and Tsoar, 1990; Williams *et al.*, 1998). These dust mantles are now vegetated and relatively stable. Maximum concentrations of desert dust in deep-sea cores from the equatorial Atlantic, coincide with glacial maxima during the last half a million years, and probably for far longer. Such dust is easily recognized by its high degree of sorting.

A similar pattern of glacial aridity is evident also in the Gulf of Aden and the Red Sea. The isotopic composition of planktonic foraminifera from deep-sea cores in this region shows that during the last 250 ka, at least, glacial maxima were times of extreme aridity, with increased sea-surface salinity reflecting even higher rates of evaporation than prevail there today. We can therefore conclude that in many of the world s hot deserts, the dominant climate during the last glacial maximum (18 ± 3 ka) was drier, windier and colder than today, although the summers may still have been very hot. In addition, the desert lakes that had been full and fresh until about 18–20 ka dried out or became hyper-saline, previously perennial desert rivers became seasonal, and seasonal rivers became intermittent or ephemeral streams. With glacially lowered sea levels, land areas were greater and so the aridity associated with enhanced continentality was accentuated. Stronger trade winds associated with steeper pressure gradients between equator and pole caused increased upwelling of cold water close offshore, further accentuating the aridity of coastal deserts.

The early Holocene climates of the tropical deserts were wetter than today, with highest lake levels towards 9 ka. Similar climatic conditions occurred in the last interglacial towards 125 ka. The desert environments no doubt oscillated between these two extremes, with the interglacials slightly warmer and very much wetter than today, and the glacial maxima colder and mostly drier. However, not all arid phases coincide with glacial maxima, nor do all humid phases coincide with interglacial times. For instance, Lake Chad in the southern Sahara and Lake Abhe in the Afar desert of Ethiopia were both very high for at least

10 000 years before 18 ka, when they fell rapidly. They were then intermittently dry (Lake Chad) or dry (Lake Abhe) until 12 ka, rising rapidly thereafter to reach peak levels at 9 ka. Since about 4.5 ka both lakes have remained low apart from occasional brief transgressions. The $>30-18$ ka phase of high lake levels could be regarded as a humid glacial phase, and the $18-12$ ka regression as an arid glacial phase. Similarly, the early Holocene transgression represents a humid interglacial phase; and the late Holocene interval of low lake levels a dry interglacial phase. This simplified four-fold subdivision ignores local hydrological and geomorphic controls over rainfall, runoff, evaporation, seepage losses and groundwater inflow, but is probably closer to reality than the simple dichotomy between arid glacial and humid interglacial climates.

AQUIFER RECHARGE AND FOSSIL GROUNDWATER

The contrast between the terminal Pleistocene aridity of the Sahara, the Gobi and the Namib Deserts and their well-vegetated and well-watered early Holocene condition had an enormous impact on the Late Stone Age and early Neolithic peoples who benefited from these changes. As the great continental ice sheets melted, temperatures and sea levels rose around the world. Evaporation from the intertropical oceans increased once sea surface temperatures became warmer. The summer monsoons of tropical Australia, India and Africa brought reliable rainfall to the seasonally wet margins of the tropical deserts. Aquifers were replenished and groundwater levels rose, sometimes feeding springs and small lakes. Formerly dry lakes refilled. The mobile late Pleistocene dunes became vegetated and stable. Savanna woodland and grassland re-occupied what are today the semi-arid regions of the world. Rivers flowed once more in many parts of the Sahara and Arabia. Countless small freshwater lakes supported communities of Upper Palaeolithic or Late Stone Age hunter-fisher-gatherer communities who left behind their middens and hearths as enduring testimony of these brief episodes when the deserts were green once more.

Some groundwater recharge also took place during previous interglacial periods. However, since each interglacial episode was of relatively short duration, or roughly one-tenth of each glacial-interglacial cycle, the duration of these recharge intervals was quite short. The consequence of all this is that the groundwater in major aquifers like the Great Artesian Basin in Central Australia or the Nubian Sandstone aquifers in North Africa may be up to a million years old. Current use of the fossil groundwater in Australia, Arabia and in some of the North American deserts is well in excess of current recharge rates. In essence, we are mining a non-renewable resource.

AMBIGUOUS QUALITY OF THE EVIDENCE FOR CLIMATIC CHANGE

Early studies of desert regions tended to focus upon specific desert landforms such as dunes, alluvial fans, river terraces, playa lakes, and deflation hollows. In the last 30 years, particularly since the widespread use of radiocarbon dating, palaeoclimatic research in deserts has focused upon using alluvial and lacustrine deposits and their associated plant and animal fossils to reconstruct the history of desert lakes and rivers (Cooke *et al.*, 1993; Abrahams and Parsons, 1994; Thomas, 1997). One of the problems inherent in using high lake levels as evidence of formerly wetter climates lies in the complex hydrology of many desert lakes. Some are fed primarily from groundwater, and may respond slowly to local changes in climate. Others may be fed solely from surface runoff. If the rivers that flow into these lakes originate in some distant well-watered uplands, the lake levels will fluctuate in response to distant changes in rainfall and may again not accurately reflect local conditions. Where the lakes are full and overflowing, and merely enlarged portions of a through-flowing river system, they will also tend to be highly insensitive to local climatic fluctuations. Finally, is a lake high because of high rates of precipitation over the lake basin or because of much lower rates of evaporation related to colder or cloudier conditions? Interpreting river terraces is equally fraught with ambiguity. Does widespread sedimentation reflect a river no longer able to transport its load because of aridity in the headwaters and reduced discharge? Or does it reflect an increase in the supply of sediment from increased erosion in the headwaters, perhaps related to glaciation? Or might it represent a change from regular perennial flow to a more seasonal flow regime? To use a river terrace to infer a particular climate and then use the inferred climate to interpret other river terraces is to indulge in circular argument. None of these questions is easy to answer. Each requires accurate dating and careful scrutiny of many independent lines of evidence for its proper resolution. An excellent example of this approach is the loess record from China.

THE LOESS RECORD IN CHINA

The loess sequence in China illustrates how accurate dating and careful evaluation of different lines of evidence are essential in reconstructing environmental change in deserts. Fine-grained wind-blown dust accumulated at intervals throughout the Quaternary on the downwind margins of deserts in Africa, Australia, South America and Asia. In terms of their sorting and mineral composition, they are virtually identical to the central European and North American loess mantles, which accumulated downwind of the fluvioglacial outwash plains, so that the term *loess* is used

here for any fine-grained aeolian deposit irrespective of its original provenance. The aeolian dust deposits in the Loess Plateau of Central China are the thickest and most extensive loess deposits in the world (Liu, 1991; Pye, 1987). They cover an area of 275 000 km^2 and attain thicknesses commonly in excess of 100 m and more than 300 m near the city of Lanzhou. Detailed studies over the past three decades have made the Chinese loess sequence with its alternation of unweathered loess and fossil soils one of the most informative sequences covering the last 2.5 Ma that exists on earth (Liu, 1991).

Detailed sampling of stratigraphic sections located on a set of west-east and north-south transects has revealed a sequence of 37 loess-soil couplets spanning the past 2.5 Ma. Each couplet represents a cold and dry phase of rapid dust accumulation and an ensuing wet and warm phase of weathering and soil formation. A soil is defined as weathered loess if it shows at least as much pedological organization as the widespread early Holocene soil at the top of the loess sequence. Interpretation of the loess-soil couplets is based on high-resolution sampling and detailed analyses of grain size, magnetic susceptibility, organic carbon (C), sediment micromorphology and mineralogy, calcium carbonate content and mollusc species. The soils are thought to indicate a weaker winter monsoon and a stronger summer monsoon. Conversely, the coarser-grained unweathered loess with generally much weaker magnetic susceptibility values is regarded as evidence of a stronger and more extensive winter monsoon and a weaker summer monsoon.

Chronological control is based on the palaeomagnetic timescale, cross-correlation with the marine oxygen isotope record and a combination of radiocarbon and thermoluminescence dates for the more recent part of the sequence. Comparison of the Chinese loess record with evidence from deep sea cores and the Greenland and Antarctic ice cores strongly confirms the climatic interpretation of the loess-soil couplets, with glacial maxima synchronous with times of maximum dust deposition and interglacials with times of maximum weathering and soil development.

INFLUENCE OF DESERT DUST ON LOCAL AND REGIONAL CLIMATE

In general, glacial maxima were drier than today and interglacial maxima were as wet or wetter. The deserts of North Africa, Arabia, Central Asia, China, Patagonia and Australia all display evidence of more vigorous aeolian dust flux during glacial maxima. Nor should it be forgotten that aeolian dust could have an influence upon local and regional climates. In tropical West Africa from 15 ka to about 7 ka the rivers were mainly depositing clays and after 7 ka they mostly carried sands. French research has attributed this abrupt hydrological change to a change in the size of raindrops. Abundant atmospheric dust would provide the nuclei for many small raindrops to form, resulting in gentle, non-erosive rains. Conversely, a reduction in atmospheric dust load would result in large, highly erosive drops. The momentum of a falling raindrop is the product of its mass and velocity, so that this interpretation is physically plausible (see also Tucker *et al.*, 1991).

Other workers have noted that the very high inputs of aeolian dust to Central Antarctica and Greenland during the Last Glacial Maximum, which are clearly evident in the Antarctic and Greenland ice cores, are consistent with shorter dust washout times and a weaker global hydrological cycle. It is also possible that a high concentration of atmospheric dust may in itself have contributed to the lowering of sea surface temperatures evident in the tropical western Pacific, especially in the warm shallow seas or Warm Pool immediately to the north of Australia. Given the growing recognition of the interactions between present-day desertification processes, dust generation, and the impact of dust particles in scattering incoming solar radiation, it seems highly plausible that wind-blown dust would be both a cause and an effect of Quaternary climatic fluctuations.

MODERN MYTHS RELATING TO DESERT ENCROACHMENT AND DESERTIFICATION

The great African droughts of the 1930s and the 1970s triggered considerable speculation as to the respective role of climatic fluctuations and human influences on land degradation along the margins of the Sahara. Many observers blamed the destruction of the vegetation and the mobilization of hitherto fixed and stable dunes entirely on human mismanagement in the form of overgrazing and clearing of forest and woodland for arable land. Some even argued that the Sahara Desert was itself at least in part a product of adverse human impact. Phrases like desert encroachment or the advancing desert were bandied about with more enthusiasm than accuracy. Such sensationalism took no account of the long geological history of the deserts, nor of the inconvenient fact that the deserts pre-date human origins. While the deserts may have helped to shape the prehistoric cultures of the occasional desert dwellers, there is scant evidence that the reverse applies, at least until more recent times.

The prolonged drought that afflicted the Sahel region of Africa as well as Ethiopia from 1968 onwards led to renewed debate about the role of humans in accentuating normal climatic variability and helping to prolong severe drought. The argument may be summarized thus. Destruction of the vegetation through overgrazing increases the surface albedo (*see* Albedo, Volume 1). If albedo increases, the surface will become cooler. A cooler surface will mean less convection and hence less instability in the lower atmosphere. The result will be a reduction in convectional rain. Less rain will mean less plant growth, and so a vicious spiral of degradation or desertification will ensue along the

desert margins. Although plausible, the albedo hypothesis fails to account for the globally synchronous pattern of major floods and droughts in both hemispheres. Nor does it account for even more severe historic droughts in Africa well before population increase and overgrazing were evident. The current consensus is that sea surface temperature anomalies control the global incidence of floods and droughts. The El Niño/Southern Oscillation (ENSO) events are perhaps the best known of these phenomena.

ENVIRONMENTAL CHANGE IN DESERTS AS AN INTEGRAL ASPECT OF GLOBAL CHANGE

In a number of areas bordering deserts, including Southern and Eastern Africa, North-east Brazil, New Mexico, Eastern and Northern Australia, Central India, and North-eastern China, the incidence of wet and dry years is strongly influenced by the incidence of ENSO events. Many of the rivers in these areas are sensitive to small changes in rainfall, and ENSO events will amplify their already highly variable flow regimes. The sensitive hydrological response of desert rivers to global sea surface temperature anomalies is an integral part of global environmental change, and is likely to remain so in the future.

This immediately raises the question of how the various deserts are likely to respond to global warming over the next 50–100 years. Williams and Balling (1996) have examined this issue in some detail in their book on *Interactions of Deserti cation and Climate*. They evaluated a number of numerical climate models and concluded that desert regions were likely to warm substantially over the next century, particularly in the higher latitudes. Increases in evaporation associated with temperature increases were likely to cause a reduction in soil water content, which could in turn lead to a reduction in vegetation cover, leaving the desert margins even more vulnerable to accelerated erosion by wind and water. The precipitation predictions for individual arid and semi-arid areas are even less certain than the temperature predictions, but overall the models suggest greater aridity and an increase in extreme climatic events for the temperate and tropical deserts and their borders. It is worth noting that the broad trends of temperature and aridity in most but not all desert areas over the past century are generally consistent with the predictions of the models.

See also: Deserti cation, Volume 3; Groundwater Withdrawal and the Development of the Great Man-made River Project, Libya, Volume 3; Deserti cation Convention, Volume 4.

REFERENCES

Abrahams, A D and Parsons, A J, eds (1994) *Geomorphology of Desert Environments*, Chapman and Hall, London.

Cooke, R, Warren, A, and Goudie, A (1993) *Desert Geomorphology*, UCL Press, London.

Frostick, L E and Reid, I, eds (1987) *Desert Sediments: Ancient and Modern*, Blackwell, Oxford.

Lancaster, N (1995) *Geomorphology of Desert Dunes*, Routledge, London.

Liu, T S, ed (1991) *Loess, Environment and Global Change*, Science Press, Beijing.

Pye, K (1987) *Aeolian Dust and Dust Deposits*, Academic Press, London.

Pye, K and Tsoar, H (1990) *Aeolian Sand and Sand Dunes*, Unwin Hyman, London.

Thomas, D S G, ed (1997) *Arid Zone Geomorphology: Process, form and Change in Drylands*, 2nd edition, John Wiley & Sons, New York.

Tucker, C J, Dregne, H E, and Newcomb, W W (1991) Expansion and Contraction of the Sahara Desert from 1980 to 1990, *Science*, 253, 299–301.

Williams, M A J and Balling, Jr, R C (1996) *Interactions of Deserti cation and Climate*, WMO and UNEP with Arnold, London.

Williams, M, Dunkerley, D, De Deckker, P, Kershaw, P, and Chappell, J (1998) *Quaternary Environments*, 2nd edition, Arnold, London.

FURTHER READING

Goudie, A, Livingstone, I, and Stokes, S, eds (1999) *Aeolian Environments, Sediments and Landforms*, John Wiley & Sons, Chichester.

Detection and Attribution of Climate Change

see Climate Change, Detection and Attribution (**Volume 1**)

Dimethylsulfide (DMS)

Meinrat O Andreae
Max Planck Institute for Chemistry, Mainz, Germany

The presence of dimethylsul de (DMS, $(CH_3)_2S$) in the oceans and atmosphere was discovered in the early 1970s by James Lovelock (Lovelock et al.*, 1972). Subsequently, DMS was identi ed as the most abundant biogenic sulfur compound in the oceans and as the dominant natural source of*

sulfur to the atmosphere (Andreae, 1990). DMS is produced from dissolved sulfate by the marine plankton community, and to a lesser extent also by terrestrial plants (Bates et al.*, 1992). Based on these observations, Charlson, Lovelock, Andreae and Warren (CLAW) (Charlson* et al.*, 1987) proposed a biospheric climate feedback, in which DMS is released by marine phytoplankton, enters the troposphere and is oxidized to sulfate particles, which then act as cloud condensation nuclei (CCN) for marine clouds. Changes in CCN concentration affect the cloud droplet number concentration, which in uences cloud albedo and consequently climate. Initial estimates have suggested that the effect of a 33% increase in CCN production from DMS would be of the same order of magnitude, but opposite direction, as that associated with present-day greenhouse gas forcing. Large-scale climate change, in turn, affects the phytoplankton in the oceans and thereby closes the feedback loop. In spite of considerable research effort, fundamental gaps remain in understanding the key issues in this biosphere-climate interaction. Critical information is missing regarding the processes which regulate the concentration of DMS in seawater, the rate of transfer across the air/sea interface, the mechanism and rate of CCN production from DMS oxidation, and the effect of climate on DMS production in the sea.*

The concentration of DMS in the surface ocean is regulated by a complex set of interactions, shown schematically in Figure 1 (Bates *et al.*, 1994; Kiene *et al.*, 1996). DMS is produced from a metabolic precursor, dimethylsulfoniopropionate (DMSP), the intracellular concentration of which varies between different phytoplankton species over a range of five orders of magnitude. DMSP is released into the water column predominantly as a result of senescence or grazing by viral, bacterial, and zooplankton. The subsequent breakdown of DMSP to DMS, which occurs with turnover times on the order of hours to days, is microbially mediated and has a highly variable DMS yield (10–70%). Once produced in the marine mixed layer, DMS is lost again by a number of mechanisms including bacterial and photochemical decomposition, emission to the atmosphere, and downward mixing, with a total turnover time of one to a few days. The removal rates by these sink mechanisms are highly variable as a function of ecological and meteorological conditions, but are of comparable overall importance for the removal of dissolved DMS. Because of this complexity, attempts to predict the concentration of DMS in surface waters by a process model have been successful only on a regional scale. As a result of this limited understanding at the process level, current estimates of the global DMS concentration and emission field are based on an heuristic extrapolation scheme (Kettle *et al.*, 1999). This precludes prediction of the distribution of DMS in the oceans and its emission to the atmosphere under scenarios of future global change.

The ocean-to-atmosphere flux of DMS is estimated from the air/sea concentration gradient and an empirically determined exchange coefficient. The different formulations for the wind speed dependence of this coefficient, based on wind tunnel, radiocarbon, and tracer measurements, range in their predictions by about a factor of two. At present, there is no objective way to select one formulation over the

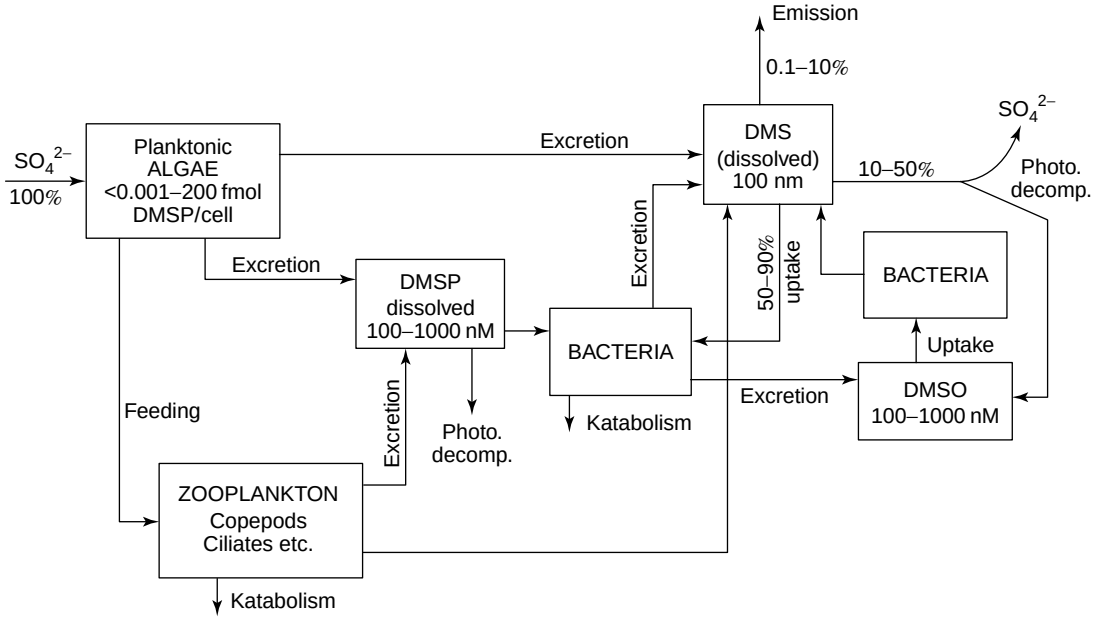


Figure 1 Model of DMS cycle in surface ocean

others, and it appears that parameters other than windspeed, e.g., the coverage of the water surface with organic films, must also be considered. Largely as a consequence of this uncertainty, current estimates of global annual sea/air flux of DMS range from 16 to 40 Tg S year^{-1}, with frustratingly little improvement in the width of this range over the last decade.

The main oxidant in the atmospheric photo-oxidation of DMS is the OH radical, initiating a reaction sequence leading to the production of SO_2 and lesser amounts of methanesulfonic acid (MSA), dimethylsulfoxide and dimethylsulfone (Davis *et al.*, 1999). Under some circumstances, the NO_3 radical can also play an important role, especially in winter. This could explain the observed seasonal cycle in the MSA/non-seasalt-sulfate ratio, which shows a pronounced minimum in winter. Recent modeling studies suggest that reactions with halogen radicals in the marine boundary layer (MBL), such as BrO, Cl, and Br, may also contribute to the oxidation of DMS. Such additional sinks and the observed large diel (24-hour) amplitude of DMS would imply sea surface emission fluxes near the higher end of the range of estimates.

Present evidence from field measurements, laboratory studies, and model calculations suggests that SO_2 is the dominant oxidation product of DMS under most environmental conditions and that this SO_2 is the main source of non-seasalt sulfate aerosol in the unpolluted MBL. In order to be relevant to climate forcing under the CLAW hypothesis, however, DMS oxidation must increase the number concentration of CCN, i.e., particles that can nucleate droplets in clouds. This can either be achieved by the creation of new particles from the gas phase, or by growth of existing smaller particles beyond the minimum diameter necessary for cloud droplet nucleation (ca. 0.05 μm diameter, assuming the particles are completely soluble).

The formation of new particles from gaseous H_2SO_4 molecules is a highly non-linear process, which is strongly influenced by the levels of H_2O and NH_3 present. Since deposition of H_2SO_4 molecules on existing aerosol surfaces competes with nucleation, it has been argued that nucleation of sulfate aerosol can only take place in the MBL if the normally present aerosol population has been depleted (Clarke *et al.*, 1998a). Recent measurements of newly formed particle concentrations over the remote Southern Ocean provide strong evidence that, at least in this region, most of the formation of new particles does indeed take place in the free troposphere around the detraining regions of clouds, at altitudes from 2 to 6 km (Clarke *et al.*, 1998b). These new particles are initially too small to act as CCN, but can be entrained from the free troposphere into the MBL and subsequently grow either by gas-to-particle deposition or by cloud processes. By this mechanism, DMS oxidation can lead to the formation of new CCN both by providing the starting material for the nucleation of new particles in the free troposphere, and subsequently by supporting the growth of smaller particles into CCN.

Given the strong predicted climate response to even relatively modest changes in the DMS-derived CCN number concentration (Charlson *et al.*, 1987), it would be important to know how DMS emission and CCN production might change over the next 100 years (*see* Aerosols, Effects on the Climate, Volume 1). If we assume that DMS levels in the ocean remain constant, the climate change predicted over the next century would only result in a slight increase ($\ll$10%) in the marine DMS emission. It is quite likely, however, that changing ecological conditions in the oceans will influence the concentration of DMS. Unfortunately, in the absence of a process-based model of the regulation of DMS concentrations in the surface ocean, we are presently unable to predict by how much, and even in which direction, this change would be. Since the rate of sulfate aerosol production depends strongly on the amount of OH radical in the atmosphere, changes in UV radiation and the oxidizing properties of the atmosphere may also influence the climate forcing due to DMS in ways that are difficult to predict at the present state of knowledge.

REFERENCES

Andreae, M O (1990) Ocean–Atmosphere Interactions in the Global Biogeochemical Sulfur Cycle, *Mar. Chem.*, 30, 1–29.

Bates, T S, Kiene, R P, Wolfe, G V, Matrai, P A, Chavez, F P, Buck, K R, Blomquist, B W, and Cuhel, R L (1994) The Cycling of Sulfur in Surface Seawater of the Northeast Pacific, *J. Geophys. Res.*, 99, 7835–7843.

Bates, T S, Lamb, B K, Guenther, A, Dignon, J, and Stoiber, R E (1992) Sulfur Emissions to the Atmosphere from Natural Sources, *J. Atmos. Chem.*, 14, 315–337.

Charlson, R J, Lovelock, J E, Andreae, M O, and Warren, S G (1987) Oceanic Phytoplankton, Atmospheric Sulphur, Cloud Albedo, and Climate, *Nature*, 326, 655–661.

Clarke, A D, Davis, D, Kapustin, V N, Eisele, F, Chen, G, Paluch, I, Lenschow, D, Bandy, A R, Thornton, D, Moore, K, Mauldin, L, Tanner, D, Litchy, M, Carroll, M A, Collins, J, and Albercook, C (1998a) Particle Nucleation in the Tropical Boundary Layer and its Coupling to Marine Sulfur Sources, *Science*, 282, 89–92.

Clarke, A D, Varner, J L, Eisele, F, Mauldin, R L, Tanner, D, and Litchy, M (1998b) Particle Production in the Remote Marine Atmosphere: Cloud Outflow and Subsidence During ACE 1, *J. Geophys. Res.*, 103, 16 397–16 409.

Davis, D, Chen, G, Bandy, A, Thornton, D, Eisele, F, Mauldin, L, Tanner, D, Lenschow, D, Fuelberg, H, Huebert, B, Heath, J, Clarke, A, and Blake, D (1999) Dimethyl Sulfide Oxidation in the Equatorial Pacific: Comparison of Model Simulations with Field Observations for DMS, SO_2, H_2SO_4(g), MSA(g), MS, and NSS, *J. Geophys. Res.*, 104, 5765–5784.

Kettle, A J, Andreae, M O, Amouroux, D, Andreae, T W, Bates, T S, Berresheim, H, Bingemer, H, Boniforti, R, Curran, M A J, DiTullio, G R, Helas, G, Jones, G B, Keller, M D,

Kiene, R P, Leck, C, Levasseur, M, Maspero, M, Matrai, P, McTaggart, A R, Mihalopoulos, N, Nguyen, B C, Novo, A, Putaud, J P, Rapsomanikis, S, Roberts, G, Schebeske, G, Sharma, S, Simó, R, Staubes, R, Turner, S, and Uher, G (1999) A Global Database of Sea Surface Dimethylsulfide (DMS) Measurements and a Simple Model to Predict Sea Surface DMS as a Function of Latitude, Longitude and Month, *Global Biogeochem. Cycles*, 13, 399–444.

Kiene, R P, Visscher, P T, Keller, M D, and Kirst, G O, eds (1996) *Biological and Environmental Chemistry of DMSP and Related Sulfonium Compounds*, Plenum Press, New York, 1–430.

Lovelock, J E, Maggs, R J, and Rasmussen, R A (1972) Atmospheric Dimethyl Sulphide and the Natural Sulphur Cycle, *Nature*, 237, 452–453.

Downscaling

Brad Bass
University of Toronto, Toronto, Canada

Climate change is a global phenomenon, yet many of the impacts will differ substantially between regions and between speci c locations within regions. Due to the computer time required to simulate the climate, it is necessary to ignore variations at higher resolutions. Downscaling offers a means to derive higher-resolution climate scenarios in order to assess the impacts of climate change at a regional or local scale. The major approach to downscaling is stochastic or empirical linking of the large-scale atmospheric circulation with site-speci c observations in a manner similar to the perfect prognosis (PP) method used in weather forecasting. An alternative dynamic approach uses the outputs from a global climate model as boundary conditions for regional climate models (RCM) or higher-resolution mesoscale atmospheric models, which simulate the physics and dynamics of the atmosphere over a smaller area. Even though both approaches have their drawbacks, downscaling has been recognized as having a valuable role to play in providing input for assessing the risks of climate variability and change.

Many of the impacts of climate change will be experienced as changes in a broad range of biophysical or socio-economic systems such as agriculture, hydrology, forestry and health. Biophysical and social impacts often occur on the order of tens of kilometers in response to weather events that range in size from a few to hundreds of kilometers and last a few hours to a few days. Simulations of future climates, such as produced by global climate models (components of which are general circulation models (GCMs)) (*see* General Circulation Models (GCMs), Volume 1), are not able to represent these smaller scale atmospheric events. Downscaling offers one approach to providing data for climate impact assessment that are more representative of the scales at which the impacts will be observed (Giorgi and Mearns, 1991). Downscaling approaches provide higher resolution data through quantifying the links between synoptic scale patterns and site-specific observations and using those links to translate the large, atmospheric patterns from the GCM to site-specific climate variables.

The major approach to downscaling is stochastic or empirical (Bárdossy and Plate, 1992; Cubasch *et al.*, 1996; Wilby *et al.*, 1998). In this approach, the large-scale atmospheric circulation is empirically linked with site-specific observations in a manner similar to the PP method used in weather forecasting. PP assumes that numerical simulations are able to provide an accurate representation and forecast of the large-scale atmospheric circulation. The variables that characterize this larger-scale circulation, typically free atmospheric variables, for instance geopotential heights and/or surface variables such as sea surface temperature, are related in an empirical manner to site-specific observations, such that site-specific variables can be derived from these more reliably simulated variables.

Downscaling applications can proceed in the same manner, either for monthly or daily data. Although linear methods are commonly used, good results have been achieved with non-linear methods for variables that are considered to be difficult such as wind (Faucher *et al.*, 1999). In some applications, the variables that characterize the large-scale circulation are used to classify the atmospheric circulation into regular patterns. Each pattern is associated with a specific probability distribution of temperature, precipitation or other climatic variables. The classification and association can occur at daily or monthly time steps.

One proposed extension of downscaling has been to the integrated assessment of multiple air issues, specifically climate change, ultraviolet radiation, smog, acid rain, suspended particulates and the transport of hazardous pollutants (Bass and Brook, 1997). It is very difficult to assess the integrated risk of multiple issues because the risks are not additive as the occurrence of one may affect the occurrence of another. However, these atmospheric issues are sensitive to a similar suite of climatic variables, and if these can be downscaled, it may be possible to develop a set of integrated scenarios for these issues under current climate conditions and under climate change.

These stochastic approaches have been acknowledged as being quite valuable for producing high-resolution data sets under current atmospheric conditions. When used to

produce climate change scenarios, they have been criticized with respect to the basic underlying assumption – the stochastic parameters are invariable under a climate change scenario. In those applications that rely on classified circulation patterns, it is further assumed that the patterns do not change, but their frequency of occurrence may change. Thus, climate change becomes defined as the change in the frequency of atmospheric circulation patterns, but this assumption has also been criticized (Wilby *et al.*, 1998). An alternative is to embed a higher resolution atmospheric model into the GCM or to use a dynamic approach to downscaling by coupling the GCM to a higher resolution model (Bass and Brook, 1997).

The dynamic approach uses the outputs from a GCM as boundary conditions for regional climate models (RCM) or higher-resolution mesoscale atmospheric models at scales ranging from 5 to 50 km (Mearns *et al.*, 1999). These models simulate the physics and dynamics of the atmosphere over a small area. Statistical models, called weather generators, can be used with stochastic downscaling to create long multiple-year scenarios. Dynamic approaches are often used over shorter time periods due to the demanding computational requirements, but are especially useful in applications where stochastic approaches are unreliable due to an inability to classify an atmospheric pattern, sparseness of data or complex terrain. In addition, there are several variables for which initial values must be provided in order to run an RCM and a mesoscale model (Bass and Brook, 1997).

Stochastic approaches are also useful in conducting risk assessments due to the large amounts of data and the multiple realizations that can be generated through the use of Monte Carlo techniques. However, any other errors or uncertainties that are inherent in the GCM will be translated into the new higher-resolution climate data sets. In fact, with respect to the accuracy of the GCM in simulating the future, all downscaling approaches will most likely increase the uncertainty of the scenario.

An international research program, the Weather Generator Project, was created to evaluate different approaches to downscaling conducted under the auspices of the International Geosphere-Biosphere Programme – Biospheric Aspects of the Hydrological Cycle (BAHC). One of the conclusions that emerged from the Weather Generator Project is that an exclusive focus on high-resolution climate change scenarios would only allow for a limited assessment of the risks to important biophysical systems. A more comprehensive approach has been proposed in BAHC based on a vulnerability assessment (Box 1). Methods such as downscaling, and in particular stochastic downscaling, have been recognized as having a valuable role in vulnerability assessment, particularly in providing input for assessing the risks of climate variability and change (Hutjes *et al.*, 1998).

See also: Climate Feedbacks, Volume 1.

Box 1 Vulnerability assessment

Vulnerability assessment refers to the evaluation of the sensitivity of a particular ecosystem, resource or activity to a broad range of environmental and socioeconomic stresses. Depending on the particular case, the assessment could be conducted through determining critical thresholds of different stresses and risks associated with those thresholds. Stochastic downscaling is a suitable method for evaluating the risks associated with climatic thresholds. For a definition of vulnerability, *see* **Precautionary Principle**, Volume 5.

REFERENCES

Bárdossy, A and Plate, E J (1992) Space-time Model for Daily Rainfall Using Atmospheric Circulation Patterns, *Water Resour. Rev.*, 28, 1247–1259.

Bass, B and Brook, J R (1997) Downscaling Procedures as a Tool for Integration of Multiple Air Issues, *Environ. Assess. Monit.*, 46, 151–174.

Cusbasch, U, von Storch, H, Waszkewitz, J, and Zorita, E (1996) Estimates of Climate Changes in Southern Europe Using Different Downscaling Techniques, *Clim. Res.*, 7, 129–149.

Faucher, M, Burrows, W R, and Pandolfo, L (1999) Empirical-statistical Reconstruction of Marine Winds Along the Western Coast of Canada, *Clim. Res.*, 11, 173–190.

Giorgi, F and Mearns, L O (1991) Approaches to the Simulation of Regional Climate Change: A Review, *Rev. Geophys.*, 29, 493–501.

Hutjes, R W A, Kabat, P, Running, S W *et al.* (1998) Biospheric Aspects of the Hydrological Cycle, *J. Hydrol.*, 212- 213, 1–21.

Mearns, L O, Bogardi, I, Giorgi, F, Matyasovszky, I, and Palecki, P (1999) Comparison of Climate Change Scenarios Generated from Regional Climate Model Experiments and Statistical Downscaling, *J. Geophys. Res.*, 104(D6), 6603–6622.

Wilby, R L, Wigley, T M L, Conway, D, Jones, P D, Hewitson, B C, Main, J, and Wilks, D S (1998) Statistical Downscaling of General Circulation Model Output: A Comparison of Methods, *Water Resour. Res.*, 34, 2995–3008.

Dust

Irina N Sokolik
University of Colorado, Boulder, CO, USA

Mineral aerosols, or dust, are airborne mineral particles that range in size from about 0.01- 100 μm *and are a dominant component of the total atmospheric aerosol burden. Most dust particles are lofted into the atmosphere by aeolian (wind) erosion of arid and semi-arid lands, which cover*

approximately 33% of the global land area. However, any type of land surface can be a potential source of dust particles. Current estimates of global annual mean dust burden are in a range from 1000 to 5000 Tg year^{-1}. With an average lifetime of up to several weeks, mineral particles can be transported long distances downwind from the source. Each year, large plumes of dust are transported from sources in North Africa across the Atlantic Ocean, reaching the Caribbean and the southeastern United States, as well as across the Mediterranean into Europe. Large quantities of dust, originating in Central Asia, are also carried out over the North Paci c to the west coast of the United States.

Dust particles are not only of natural origin, but they are also produced as a result of human activities. Various human activities, e.g., road construction, can extend the geographical area of dust sources and increase dust loading into the atmosphere. This portion of dust is called anthropogenic dust, and it is of special interest in climate change studies. Recent estimates show that the anthropogenic fraction of dust could be as much as 30- 50% of total dust production, but this remains uncertain (Tegen et al., *1996). The dependence of dust emissions on climatic parameters, such as wind speed and rainfall, strongly suggests that the atmospheric dust load could be signi - cantly affected by any climatic change that may result from human activities. This fraction of the mineral dust particles is considered as anthropogenic dust too, but it is largely unquanti ed.*

Once lifted into the atmosphere, both anthropogenic and natural components of mineral aerosols play an important role in air quality, atmospheric chemistry, ecology, remote sensing, Earth's radiation budget, and, hence, regional and global change (Duce, 1995).

DUST IMPACTS

The direct radiative effects of mineral particles are due to their ability to scatter and absorb ultra-violet (UV), visible, and infrared (IR) radiation (Sokolik and Toon, 1996). To quantify the effects of aerosols on the Earth s radiative balance, the concept of direct radiative forcing is often employed. Dust can cause either positive or negative radiative forcing, leading to heating or cooling of the climate system. The sign of the direct radiative forcing is determined by the optical properties of the dust as well as atmospheric conditions and surface reflectance. For instance, dust plumes over dark surfaces such as the ocean result in negative forcing because reflection reduces the amount of solar radiation absorbed, while their radiative forcing is positive over bright surfaces such as snow and ice because aerosols darken the surface–troposphere system and increase the amount of solar radiation absorbed.

Moreover, the presence of dust alters the surface radiation budget, thereby affecting surface temperature and various surface–air exchange processes. In addition, radiative heating or cooling occurring within the dust layer itself affects the temperature profile and, hence, atmospheric dynamics. Overall, the radiative impact of mineral aerosols is important relative to that of other types of aerosols, such as sulfates and smoke particles, due to their widespread distribution and large optical depth.

The direct radiative impact of dust is augmented by indirect radiative effects. Aerosol-induced variations of clouds, followed by the alteration of the radiative budget, are referred to as indirect radiative forcing of aerosols on climate. Some mineral particles are able to take up water and can act as cloud condensation nuclei, altering cloud properties and amount. The insoluble fraction of dust may serve as ice forming nuclei.

In addition, dust can alter atmospheric chemical processes and air composition by a variety of means: altering UV radiative fluxes, and, hence, the photochemistry of the atmosphere; influencing the partitioning of semi-volatile trace species; providing additional reaction pathways for reactions involving H_xO_v, nitric oxides (N_xO_v), sulfur dioxide (SO_2), ozone (O_3) and semi-volatile organics (VOCs) (Denterner *et al.*, 1996). In turn, chemical processes occurring on the surfaces of dust particles may result in the formation of multicomponent aerosols that can have drastically different properties and, hence, radiative impact, from those at the dust source.

Dust also has a significant impact on the chemical and biological processes in the oceans. For instance, aeolian mineral dust deposited in the oceans provides soluble iron, which is essential for the growth of plankton (Duce, 1995). As a result, iron may affect the biogeochemical cycle of sulfur-stimulating planktons to grow and produce more dimethyl sulfide, DMS. The transport of micronutrients (such as phosphate and potassium) associated with dust particles also can be important for terrestrial systems.

The degradation of visibility and health problems caused by dust make it an important air quality issue (Wilson and Spengler, 1996). During a heavy dust storm, the visibility might be as low as a few meters. Some dust particles are toxic or can act as the transport mechanism for toxic species that can be inhaled and pose a health threat.

The fundamental problem in estimating the diverse impacts of dust comes from its complex nature. Mineral aerosol is a collective term referring to widely varying mixtures that may include many constituents such as quartz, various clays, calcite, gypsum, hematite and others. These minerals have very different physical and chemical properties (e.g., size distribution, particle shape, density, spectral refractive index, solubility, chemical reactivity, etc.).

In addition, dust sources and sinks are not uniformly distributed, and the lifetime of mineral dust in the atmosphere is relatively short, ranging up to a few weeks. Consequently the dust burden has a complex spatial and temporal pattern. Moreover, the physical and chemical characteristics of dust, originally determined by a given source and production mechanism, may vary during transport. Individual particle analysis of samples collected at various locations reveals that dust particles often occur as irregular aggregates of minerals and/or other aerosol species.

Therefore, the quantification of dust impacts must include various processes governing the changes of the physical and chemical properties of dust along with dust emission and atmospheric transport.

DUST EMISSION

Mineral dust is injected into the atmosphere mainly during sporadic events of high intensity. Strong winds capable of entraining dust particles occur under different meteorological conditions and at a variety of spatio–temporal scales (Pye, 1996). It has been demonstrated that the amount of uplifted dust is proportional to:

$$(u - u^*)u^2 \tag{1}$$

where u is the surface wind speed, and u^* is the threshold friction velocity required to initiate particle motion (Gillette, 1978). However, the threshold friction velocity depends on several factors and can vary from a few tens to a few hundreds of cm s^{-1}. The main factors influencing dust production are wind stress, surface roughness, the presence of non-erodible elements (such as rocks), soil mineralogy, size of erodible particles, vegetation, and soil moisture. Both natural and anthropogenic changes to any of these factors can drastically alter dust production.

In recent decades, the dramatic increase in the human population has had a crucial impact on Earth s drylands, causing their degradation, and desertification has been a result. At present, desertification threatens about one-third of the world s land surface, spreading over every continent with an estimated rate of about 6×10^4 km^2 $year^{-1}$ (*see* Deserti cation, Volume 3). Over-cultivation of poor soils, inappropriate irrigation practices, human-induced wind erosion, severe trampling and heavy grazing, deforestation, urbanization, building and road construction; mining and industry, off-road vehicle use, and tourism are among the major factors that contribute to anthropogenic dust loading. For example, the presence of Saharan dust is thought to be an indicator of desertification of the Sahel region in Africa. In the United States, the highest dust production occurs in regions where the soil has been disrupted by cultivation and other intensive land uses. Although there are many examples of humans altering the environment, thereby causing an additional dust burden, it is a challenging problem to separately quantify the dust emission rates from natural and anthropogenic (disturbed) sources with the pertinent temporal and spatial resolution.

EVOLUTION OF DUST PROPERTIES DURING ATMOSPHERIC TRANSPORT

Once mobilized, the suspended dust particles can be vertically mixed in the atmosphere and transported downwind from the source. Measurements made at the surface often show a bimodal dust particle size distribution, with the larger particle mode composed primarily of quartz and the small particle mode consisting of smaller quartz grains and clay minerals. Because of size-dependent removal processes, the relative importance of these smaller particles increases with increasing distance from the source. Some particles are deposited back to the surface due to aerodynamic resistance (referred to as dry deposition) (*see* Deposition, Dry, Volume 2). Larger particles fall out rapidly near the source region due to gravitational settling. Rain can also remove some or all dust particles from the atmosphere (referred as wet removal). Depending on specific meteorological conditions, dust can be transported downwind for thousands of kilometers, e.g., across the Atlantic Ocean from West Africa to the Caribbean. Long-lived dust particles that are transported long distances are typically characterized by one size mode with median mass diameter of about 1–3 μm. Particle concentration generally decreases rapidly with distance from the source. At the source, particle concentrations can be as high as a few thousands particles per cubic centimeter, but the concentration normally drops to a few tens of particles per cubic centimeter over distances of several thousand kilometers.

Along with concentration and size distribution, composition of dust particles may evolve during transport. Observations indicate that dust particles are often internally mixed with other aerosol species. Dust particles coated by sulfate, nitrate, sea-salt or carbonaceous matter have been measured by both ground-based and aircraft measurements. This can occur as a result of coagulation of dust with other aerosols, uptake of gases on dust particle surfaces followed by heterogeneous chemistry, and/or cloud processing of dust particles. Whereas the mechanism for coagulation is well understood, it is not clear how heterogeneous chemistry and cloud processing are accomplished.

The evolution of dust particles also depends on the prevailing physical and meteorological characteristic of a particular dust source, as well as transport route. For instance, some plumes of Asian dust can travel through large urban areas that are heavily polluted by anthropogenic sulfates and soot aerosols. By the time Asian dust reaches the marine environment, it can have drastically different properties from those at the source.

Many studies demonstrate that dust may be transported in one or several layers located at different altitudes in the atmosphere. Normally, long-range transport of dust by strong winds occurs above the boundary layer but below 10 km. Due to stratification of the atmosphere, dust layers can have drastically different properties depending on the history of the air mass found at a given altitude. If there were to be significant changes in atmospheric circulation due to natural climate variability or climate change, the atmospheric transport and, hence, the evolution of dust plumes and their spatial distribution could be very different.

MEASUREMENTS AND MODELING OF DUST PROPERTIES

Various experimental and modeling studies have been conducted to characterize the properties of dust and evaluate its impacts. However, most observations that exist are from short-duration ground-based studies at a limited number of locations. As a result, the diverse impacts of dust cannot be fully quantified from existing measurements.

Recently considerable progress has been made in employing satellite observations to characterize dust properties and transport on a global scale. For instance, since 1978, the Advanced Very High-Resolution Radiometer (AVHRR) on board National Oceanic Atmospheric Administration (NOAA) satellites has been used to derive long-term climatologies of aerosol amount over the oceans. Data from the National Aeronautics and Space Administration (NASA) Total Ozone Mapping Spectrometer (TOMS), operating at UV wavelengths, have been used to monitor dust production and transport over land as well as over the ocean since 1979. Several new space sensors that have been explicitly designed for aerosol study are becoming available (King *et al.*, 1999).

Satellite observations confirm that mineral dust is a major aerosol component over many marine and continental areas and reveal a complex structure of the spatial and temporal distribution of dust plumes. Satellite imagery, providing global-scale aerosol distributions, can be used to guide the development of models of dust sources, transport, and deposition.

Process-oriented models (such as climate models, global circulation models (GCMs), and atmospheric chemical transport models) have become the most comprehensive tools available for predicting the diverse impacts of dust. A common approach to simulate dust fields employing an atmospheric transport model includes several steps:

1. characterize the dust source;
2. use a dust production scheme to simulate dust emission; and
3. use transport and dust removal schemes to predict the size-resolved concentration fields or total mass concentration fields as a function of time and space depending on the resolution of a given model.

Because of the complexities involved and the lack of understanding of the interactions between dust and the chemistry and physics of the atmosphere, model predictions typically rely on oversimplified parameterizations of dust processes. Nevertheless, the estimated global-mean radiative forcing by anthropogenic dust is in the same range as that due to anthropogenic sulfate aerosols and smoke from biomass burning. Models predict that the presence of dust will affect surface temperature, rainfall, and tropospheric winds.

Modeling studies further demonstrate that the diverse effects of dust may induce various feedbacks on the climate system. Several competitive feedbacks have been suggested; for example, reduction of the surface winds is likely a negative feedback, whereas a decrease in rainfall would likely lead to a positive feedback. However, an understanding of the overall response of the climate system to the diverse impacts of mineral aerosols is still incomplete.

See also: Aerosols, Effects on the Climate, Volume 1; Deserts, Volume 1; Land Cover and Climate, Volume 1.

REFERENCES

Denterner, F, Carmichael, G, Zhang, Y, Crutsen, P, and Lelifeld, J (1996) The Role of Mineral Aerosols as a Reactive Surface in the Global Troposphere, *J. Geophys. Res.*, 101, 22 869–22 890.

Duce, R A (1995) Sources, Distributions, and Fluxes of Mineral Aerosols and their Relationship to Climate, in *Aerosol Forcing of Climate*, eds R J Charlson and J Heintzenberg, John Wiley & Sons, New York, 43–72.

Gillette, D A (1978) A Wind Tunnel Simulation of the Erosion of Soil: Effects of Soil Texture, Sandblasting, Wind Speed, and Soil Consolidation on Dust Production, *Atmos. Environ.*, 12, 1735–1743.

King, M D, Kaufman, Y J, Tanré, D, and Nakajima, T (1999) Remote Sensing of Tropospheric Aerosols from Space: Past, Present and Future, *Bull. Meteorl. Soc.*, 80, 2229–2259.

Pye, K (1996) *Aeolian Dust and Dust Deposits*, Academic Press, Cambridge, 1–334.

Sokolik, I N and Toon, O B (1996) Direct Radiative Forcing by Anthropogenic Mineral Aerosols, *Nature*, 381, 681–683.

Tegen, I, Lacis, A A, and Fung, I (1996) The Influence on Climate Forcing of Mineral Aerosols from Disturbed Soils, *Nature*, 380, 419–422.

Wilson, R and Spengler, J, eds (1996) Particles in our Air: Concentrations and Health Effects, Harvard University Press, Cambridge, MA, 1–254.

FURTHER READING

Prospero, J M (1996) Saharan Dust Transport over the North Atlantic Ocean and Mediterranean: an Overview, in *The Impact*

of Desert Dust Across the Mediterranean, eds S Guerzoni and R Chester, Kluwer Academic, Norwell, MA, 133–151.

DVI (Dust Veil Index)

see DVI (Dust Veil Index) (**Volume 3**)

Dynamics, Atmospheric

see Atmospheric Motions (**Volume 1**)

Dynamics, Oceanic

see Ocean Circulation (**Volume 1**)

E

Earth, Internal Structure
see Earth System History (*Opening essay*, **Volume 1**)

Earth Observing System (EOS)
see EOS (Earth Observing System) (**Volume 1**)

Earth Observing Systems
see Earth Observing Systems (*Opening essay*, **Volume 1**)

Earth Rotation
see Atmospheric Angular Momentum and Earth Rotation (**Volume 1**)

Earth System
see The Earth System (*Introductory essay*, **Volume 1**)

Earth System History
see Earth System History (*Opening essay*, **Volume 1**)

Earth System Models
see Models of the Earth System (*Opening essay*, **Volume 1**)

Earth System, Processes and Interactions
see Earth System Processes (*Opening essay*, **Volume 1**)

EBCMs (Energy Balance Climate Models)
see Energy Balance Climate Models (**Volume 1**)

Eemian

The *Eemian* refers to the last interglacial period, which began about 127 000 years ago at the end of the previous glaciation. The onset of the last glacial cycle followed at 115 000 years ago. The Eemian is also known as the Riss-Wurm interglacial (Berggren and Couvering, 1977) and corresponds to oxygen isotope stage 3 from ocean sediments. The transition from full interglacial to full glacial conditions has generally been relatively slow (taking tens of thousands of years), while the transition from glacial-to-interglacial conditions is more rapid (taking thousands of years). This asymmetric timing creates a sawtooth-shaped pattern, with the entire process occurring in cycles that have averaged about 100 000 years. The glacial-interglacial cycles are thought to be related to changes in the Earth s orbit (*see* Orbital Variations, Volume 1) that regulate the amount of solar radiation reaching different latitudes of the Earth during different seasons (there is little overall change in the annual, global integral of incoming radiation). The most recent maximum in the Northern Hemisphere, summertime solar insolation, occurred about 10 000 years ago (the start of the Holocene). During the Eemian interglacial, the insolation maximum during July at high northern latitudes was even greater than 10 000 years ago, suggesting

that the Eemian could have been warmer than the beginning of the Holocene. Paleogeologic proxy data do indeed indicate that land temperatures in some locations were a couple degrees warmer than present, even though sea surface temperatures seem to have been similar to present. Sea level, on average, is estimated to have been 6 m higher than at present, possibly due to melting of some or all of the West Antarctic or Greenland ice sheets. Crowley and North (1991) summarize the results from many studies.

See also: Earth System History, Volume 1.

REFERENCES

Berggren, W A and Van Couvering, J A (1978) The Quaternary, in *Treatise on Invertebrate Paleontology, Part A., Introduction: Fossilization (Taphonomy), Biogeography and Biostratigraphy*, eds R A Robinson and C Teichert, Geological Society of America, Boulder, CO, A505–A543.

Crowley, T J and North, G R (1991) *Paleoclimatology*, Oxford University Press, New York, 1–339.

BENJAMIN S FELZER *USA*

El Chichón Volcanic Eruption

see Volcanic Eruption, El Chichón (**Volume 1**)

El Niño

El Niño is a warming of the tropical Pacific that occurs roughly every 3–7 years and typically lasts for 12–18 months. It is dynamically linked to the Southern Oscillation, a seesaw in surface atmospheric pressure between the Australian–East Asian region and the eastern tropical Pacific. During El Niño, the trade winds weaken along the equator as atmospheric pressure rises in the western Pacific and falls in the eastern Pacific. Weakened trade winds allow warm surface water, normally confined to the western Pacific, to migrate eastward. Wind-driven upwelling, a process which brings cold water to the surface along the equator and along the west coasts of North and South America, is also greatly reduced, causing sea surface temperatures to rise. Upwelled waters are rich in nutrients that support biological productivity, so that reduced upwelling adversely affects marine ecosystems and economically valuable fish stocks.

In the atmosphere, deep cumulus clouds and heavy rains typically occur in the western Pacific over the warmest surface waters. As these waters migrate eastward during El Niño, so does the heavy rainfall. Condensation of water vapor into liquid water releases heat into the middle and upper troposphere. This heat provides a source of energy to drive global wind fields that extend El Niño s influence to remote parts of the planet. Altered circulation patterns produce droughts, floods, unusual storminess, heat waves, and other weather extremes that have serious social, economic, and public health consequences. However, El Niño can also have a positive influence on weather, such as bringing milder winters to North America and suppressing Atlantic hurricane formation.

El Niño in Spanish means the child, with specific reference to the Christ child. It was the name given by fishermen to a warm southward flowing coastal ocean current that occurred around Christmas time off the west coast of Peru and Ecuador. These coastal warmings are linked to the much more extensive ocean warmings that occur across the width of the tropical Pacific Basin. Thus the term El Niño has now become synonymous with these extensive ocean warmings because of their impacts on global climate. El Niño is sometimes also referred to as the warm phase of the El Niño/Southern Oscillation (ENSO) cycle (*see* El Niño/Southern Oscillation (ENSO), Volume 1).

FURTHER READING

Diaz, H F and Markgraf, V (1999) *El Niño and the Southern Oscillation*, Cambridge University Press, Cambridge.

MICHAEL J MCPHADEN *USA*

El Niño and La Niña: Causes and Global Consequences

Michael J McPhaden

National Oceanic and Atmospheric Administration, Seattle, WA, USA

El Niño is a warming of the tropical Paci c that occurs roughly every three to seven years and lasts for 12- 18 months. It is dynamically linked to the Southern Oscillation, a see-saw in surface atmospheric pressure between the Australian- East Asian region and the eastern tropical Paci c. During El Niño, the trade winds weaken along the equator as atmospheric pressure rises in the western Paci c and falls in the eastern Paci c. Weakened trade winds allow warm surface water, normally con ned to the western Paci c, to migrate eastward. Wind-driven

upwelling, a process that brings cold water to the surface along the equator and along the west coasts of North and South America, is also greatly reduced, causing sea surface temperatures to rise. Upwelled waters are rich in nutrients that support biological productivity, so that reduced upwelling adversely affects marine ecosystems and economically valuable fish stocks.

In the atmosphere, deep cumulus clouds and heavy rains typically occur in the western Pacific over the warmest surface waters. As these waters migrate eastward during El Niño, so does the heavy rainfall. Condensation of water vapor into liquid water releases heat into the middle and upper troposphere. This heat provides a source of energy to drive global wind fields that extend El Niño's influence to remote parts of the planet. Altered circulation patterns produce droughts, floods, unusual storminess, heat waves, and other weather extremes that have serious social, economic, and public health consequences. However, El Niño can also have a positive influence on weather, such as bringing milder winters to North America and suppressing Atlantic hurricane formation.

La Niña is a climatological phenomenon akin to El Niño, but with opposite tendencies in the tropical Pacific Ocean and atmosphere. La Niña is characterized by stronger than normal trade winds and colder than normal tropical Pacific sea surface temperatures. It is also characterized by unusually high surface atmospheric pressure in the eastern tropical Pacific and low surface pressure in the western tropical Pacific in association with the Southern Oscillation. La Niña's effects on global weather are roughly opposite to those of El Niño. As a result, El Niño, La Niña and Southern Oscillation are often referred to collectively as El Niño/Southern Oscillation (ENSO), a cycle that oscillates year-to-year between warm, cold and neutral states in the tropical Pacific.

After the disastrous 1982-1983 El Niño, which was neither predicted nor even detected until nearly at its peak, a 10-year international scientific research program was undertaken between 1985-1994 to improve the understanding, detection, and prediction of ENSO-related variability (see TOGA (Tropical Ocean Global Atmosphere), Volume 1). New networks of moored and drifting buoys, island and coastal tide gauges, and ship-based measurements were initiated. Satellites provided global observations of the ocean and the atmosphere. Computer models were developed to forecast El Niño and La Niña events with lead times of up to one year. Scientific progress in these areas was highlighted during the 1997-1998 El Niño, which by some measures was the strongest on record. This El Niño was monitored day-by-day at a level of detail never before possible. In addition, early warning impending impacts, made possible by new data and forecasting techniques led to planning efforts in some parts of the world that helped to reduce the toll the El Niño, might otherwise have taken.

The strength of the 1997-1998 El Niño, the tendency for more frequent El Niño events and less frequent La Niña events in the past 25 years, and the prolonged 1991-1995 El Niño have raised questions about the possible influence of global warming on the ENSO cycle. Some recent computer model simulations suggest that the ENSO cycle may be more energetic in a warmer world. However, no firm conclusions can be drawn at present because models used to simulate the interaction between global warming and ENSO are limited in their ability to accurately represent key physical processes. Also, available data are insufficient to unambiguously establish that recent changes in the ENSO cycle are outside the range of natural variability.

HISTORICAL BACKGROUND

El Niño in Spanish means *the child*, with specific reference to the Christ child. It was the name given by fishermen to a warm southward flowing coastal ocean current that occurred every year around Christmas time off the west coast of Peru and Ecuador. The term was later restricted to unusually strong warmings every few years that disrupted local fisheries, led to massive bird die offs, and brought torrential rains to the region. These coastal warmings and episodes of heavy rainfall have been linked by researchers over the past 40 years to much more extensive ocean warmings that occur across the width of the tropical Pacific basin (Figure 1). The term El Niño has now become synonymous with these extensive ocean warmings because of their impacts on global climate (Philander, 1990).

El Niño is coupled to an atmospheric phenomenon known as the Southern Oscillation, first defined by Sir Gilbert Walker in the early 20th century. Walker was an Englishman who was appointed director general of observatories in India in 1904. He arrived in the wake of a devastating drought and famine that claimed more than a million lives in 1899–1900. Most of the rain in India falls during the summer monsoon. When the monsoons fail, the results can be disastrous for Indian agriculture. Walker believed that summer monsoon rainfalls might be predictable, so he set out to find possible precursors, gathering weather information from all over the world. In the 1920s and 1930s, he published a series of papers describing a year-to-year see-saw in surface atmospheric pressure between the Eastern and Western Hemispheres (Figure 2). He found that when pressure was unusually low over the Indian Ocean region and the western tropical Pacific, it was high east of the international date line in the southeastern tropical Pacific. This situation corresponded to periods of good monsoon rains in India. When this pressure pattern was reversed, with low pressures east of the date line and high pressures west of the date line, rainfall was deficient not only in India,

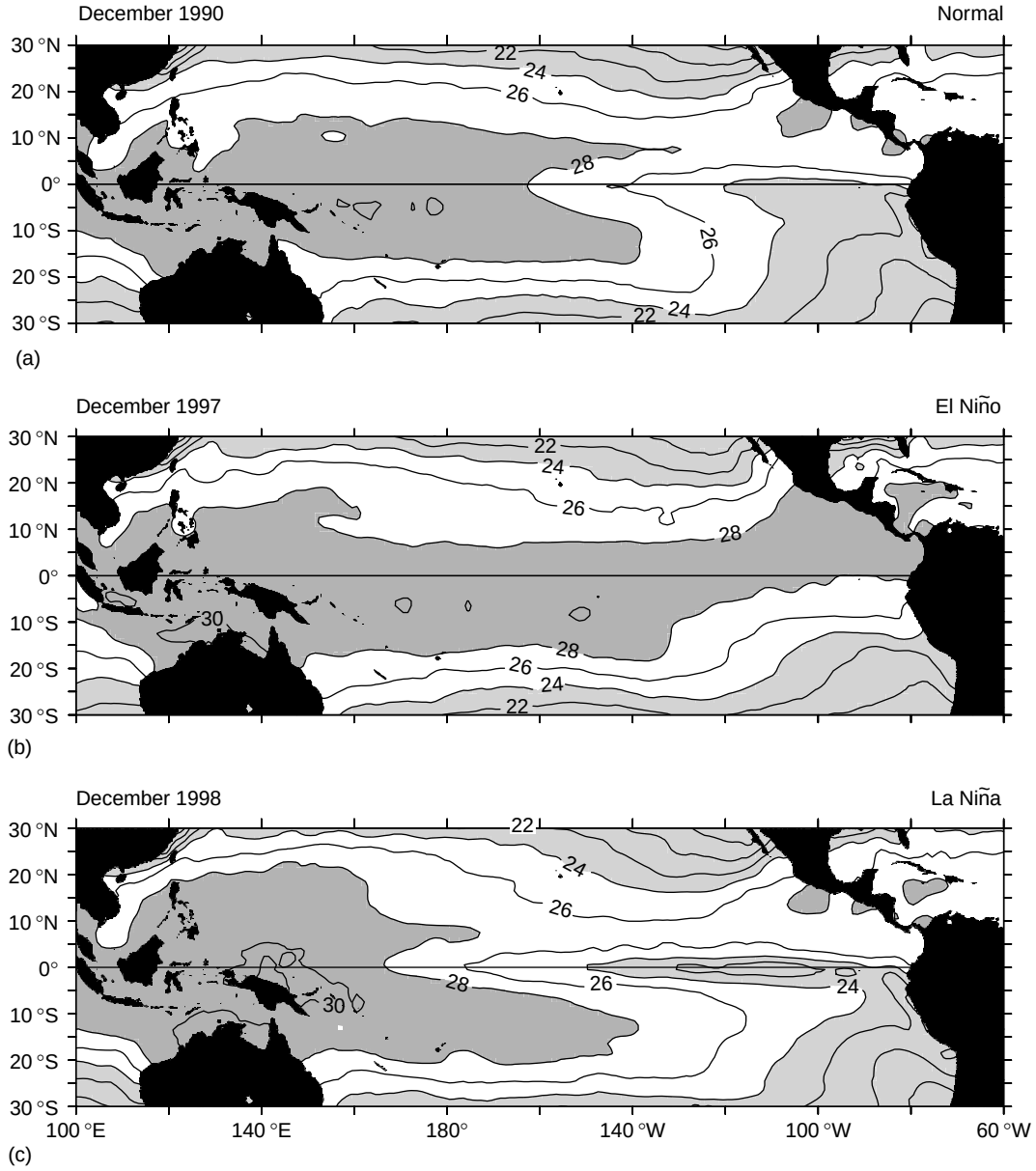


Figure 1 Monthly averaged sea surface temperature (in °C) illustrating normal conditions in December 1990 (a), El Niño conditions in December 1997 (b), and La Niña conditions in December 1998 (c)

but also in neighboring countries such as Australia and Indonesia.

Unfortunately, for Walker, variations in the Southern Oscillation lagged rather than led those associated with Indian summer monsoon rainfall, so that his index was not useful for monsoon predictions. Curiously, he also found that the low index phase of the Southern Oscillation was correlated with warm winters over Canada and dry conditions in parts of South Africa. His work was met with a great deal of skepticism though because he could not provide a physical explanation for the observed connections between the Southern Oscillation and climatic variations in widely separated parts of the globe. This skepticism was evident in his obituary, published in the *Quarterly Journal of the Royal Meteorological Society* (1959, p. 185):

> Walker s hope was presumably not only to unearth relations useful for forecasting but to discover sufficient and sufficiently important relations to provide a productive starting point for a theory of world weather. It hardly seems to be working out like that.

Ironically, Walker died in 1958 during the 1957–1958 International Geophysical Year (IGY), the first truly international scientific study of the global oceans, atmosphere, and

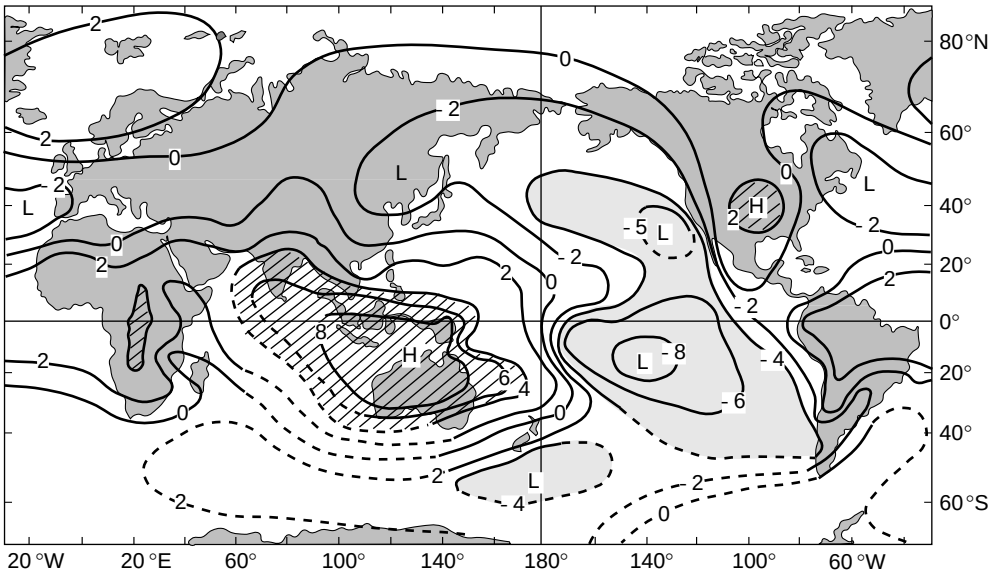


Figure 2 Spatial pattern of annual mean sea level pressure anomalies associated with the Southern Oscillation. Solid cross hatching indicates regions where sea level pressure varies in phase with Darwin, Northern Australia, and light shading indicates regions where sea level pressure varies out of phase with Darwin. Units are in correlation coefficient ($\times$10), for which large absolute values indicate a more consistent relationship with Darwin (Trenberth and Shea, 1987)

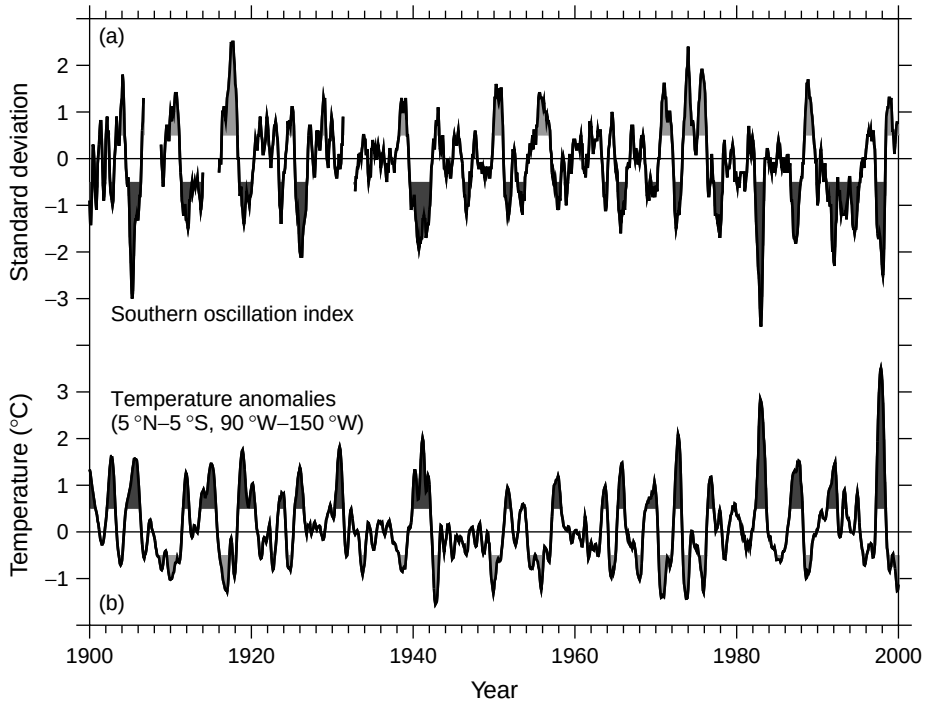


Figure 3 (a) SOI, which is a normalized difference between Tahiti, French Polynesia minus Darwin, Australia surface air pressure, after typical seasonal variations have been subtracted out. Periods of SOI greater in magnitude than 0.5 are shaded to emphasize the relationship with El Niño and La Niña episodes. Low SOI is associated with warm sea temperatures (El Niño), and high SOI with cold sea temperatures (La Niña). (b) Sea surface temperature anomalies (that is, deviations from normal) for the region 5 °N–5 °S, 90 °W–150 °W from 1900–1999. Positive anomalies greater than about 0.5 °C indicate El Niño events. Negative anomalies less than about −0•5 °C indicate La Niña events

solid earth (*see* IGY (International Geophysical Year), Volume 1). Intensive measurements were being undertaken in a coordinated effort to better understand the processes involved in shaping the global environment. As fate would have it, 1957–1958 coincided with a strong El Niño event. Measurements from the various observing systems deployed during the IGY stimulated Jacob Bjerknes to examine how interactions between the ocean and the atmosphere in the tropical Pacific could affect weather patterns (*see* Bjerknes, Jacob, Volume 1). Bjerknes was a Norwegian meteorologist who emigrated to the US just before World War II. He had established his reputation in the early part of the century by making fundamental contributions to the understanding of fronts and weather systems at middle and high latitudes. Using IGY data in the 1960s, Bjerknes would be the first to identify the relationship between El Niño and the Southern Oscillation, to propose physical mechanisms by which they might interact, and to describe some of the effects of El Niño on North American weather.

The close connection between El Niño and the Southern Oscillation over the past 100 years can be seen in the strong inverse relationship between two commonly used indicators (Figure 3). The Southern Oscillation index (SOI) is based on departures from normal of surface atmospheric pressure at Tahiti, French Polynesia, and Darwin, Northern Australia. The difference of Tahiti minus Darwin pressure provides a measure of the pressure force that drives the trade winds across the Pacific basin. When this index is negative (low pressure at Tahiti relative to Darwin), the trade winds are weak. When the index is positive, the trade winds are strong.

A commonly used index for El Niño is the areal averaged sea surface anomaly (i.e., deviation from normal) in the region 5 °N to 5 °S, 90 °W to 150 °W. Large positive values of this index define warm El Niño conditions, while large negative values indicate cold La Niña conditions. It is evident that the warm surface temperatures are strongly linked to weak trade wind (negative values of the SOI) and vice versa. These indices show that the 1982–1983 El Niño and the 1997–1998 El Niño were the strongest of the 20th century.

El Niño events exhibit some common characteristics (such as unusually warm sea surface temperatures and weak trade winds in the tropical Pacific) that allow them to be classified as a distinct phenomenon (see Box 1). However, they often differ among themselves in duration, intensity and in the details of their development. No two El Niños are exactly alike, and the same is true of La Niña events. Thus, the ENSO cycle displays a degree of irregularity for which there is yet no simple explanation.

Box 1 El Niño and La Niña, Past and Present

El Niño and La Niña events generally last 12–18 months, and tend to be most fully developed during the Northern Hemisphere winter season. Thus, they often span two calendar years. Commonly used indices like those in Figure 3 show that El Niño events over the past 50 years occurred in 1951, 1953, 1957–1958, 1963, 1965–1966, 1969, 1972–1973, 1976–1977, 1982–1983, 1986–1987, 1991–1995 and 1997–1998. La Niña events of the past 50 years occurred in 1949–1950, 1955–1956, 1964, 1970–1971, 1973–1975, 1988–1989, 1995–1996 and 1998–2000. Weak cold events also occurred in 1967–1968 and 1984–1985, but these were not accompanied by positive values of the SOI.

Instrumental records available for identifying El Niño and La Niña years prior to the mid-19th century are very limited. However, it has still been possible to identify El Niño events in the distant past using anecdotal historical information from the reports of early explorers and settlers in those areas bordering the Pacific where in modern times El Niño consistently affects weather and climate. Quinn *et al.* (1987) produced a chronology of El Niño events going back to 1525 using historical reports of conditions from the coastal region and adjacent waters of northwestern South America. Their work suggests, e.g., that Francisco Pizarro's conquest of the Incas in 1531–1532 coincided with an El Niño event. Heavy rains and swollen rivers, which typically occur only during El Niño years in Peru, delayed Pizarro's advance through the countryside. On the other hand, the same rains produced abundant vegetation, providing plentiful fodder for his horses, which were one of the chief tactical advantages (along with swords) that his small contingent of soldiers had over the natives.

Quinn (1992) extended this El Niño chronology back to 622 AD using maximum yearly Nile River flow data at Cairo, and related historical information on African droughts, floods, plagues and famines. Careful records of Nile stream flow were kept because it formed the basis of Egyptian agriculture for thousands of years. The stream flow records reveal year-to-year changes that can be related to variations in summer monsoon rains over the highlands of Ethiopia. These rains, and the stream flows they feed, are typically reduced in El Niño years.

It is possible to extend the record of ENSO time scale variations even further back in time by developing climate reconstructions using proxy data from tree rings, laminated lake sediments, glacier ice cores and tropical Pacific corals. These proxies preserve year-by-year records of environmental parameters related to physical climate, such as temperature, precipitation, or rainwater runoff. The relevant climate information is encoded in the chemical and isotopic composition of coral skeletons, the water content and chemistry of ice cores, the thickness and grain sizes of sediment layers, and the thickness and density of tree rings. Properly calibrated against modern instrumental records, proxy data allow investigations on ENSO-related climate variability to extend many thousands of years into the past.

EL NIÑO PHYSICS AND THE ENSO CYCLE

To understand El Niño, we must first understand what is considered normal. Sunlight reaching the Earth s surface is more intense in the tropics than at higher latitudes, so that the warmest ocean temperatures are found near the equator. Air masses over warm tropical waters extract heat and moisture from the ocean, expand, become less dense than surrounding air, and ascend to higher altitudes. The rising air cools and condenses, producing towering cumulus clouds and heavy precipitation through a process referred to as deep convection. At upper levels of the troposphere (the lowest 10–12 km of the atmosphere affected by weather) these air masses flow poleward, then sink in regions of high surface pressure over the subtropical oceans of the Northern and Southern Hemispheres. The rising tropical air masses in turn are fed by air drawn in towards the equator near the surface from the subtropical high-pressure zones. This circulation cell on the meridional plane is called the Hadley circulation in honor of George Hadley, an 18th century British meteorologist who was the first to provide a theoretical explanation for it (*see* Hadley Circulation, Volume 1).

The equatorward flowing surface air masses are deflected westward because of the rotation of the Earth about its axis (the Coriolis effect). The result is easterly tradewind systems in the Northern and Southern Hemispheres. The northeast and southeast trade winds meet in the Intertropical Convergence Zone, situated on average about five to ten degrees of latitude north of the equator in the Pacific. This is a region of deep convection, cumulus cloud formation, and heavy rainfall that comprises the ascending branch of the Hadley circulation. A second zone of converging surface winds, the South Pacific Convergence Zone, is on average located several degrees south of the equator in the western tropical Pacific.

Along the equator, the trade winds normally drive surface flow westward in the south equatorial current (Figure 4). This current piles up the warm surface water in the western Pacific, and drains it from the eastern Pacific. The thermocline, which is the sharp vertical temperature gradient separating the warm surface layer from the cold deep ocean, is pushed down to a depth of 150 m in the west, but shoals to 50 m depth in the east. Sea level tends to mirror thermocline depth since water expands when heated. Thus, while the thermocline tilts downward towards the west along the equator, sea level rises to the west where it stands about 60 cm higher than in the eastern Pacific.

The relative shallowness of the thermocline in the eastern Pacific facilitates the upward transport of cold interior water by the trade winds, and a cold tongue develops in sea surface temperature from the coast of South America to near the international date line. The east–west surface temperature contrast reinforces the easterly trade winds, since low atmospheric surface pressure is associated with

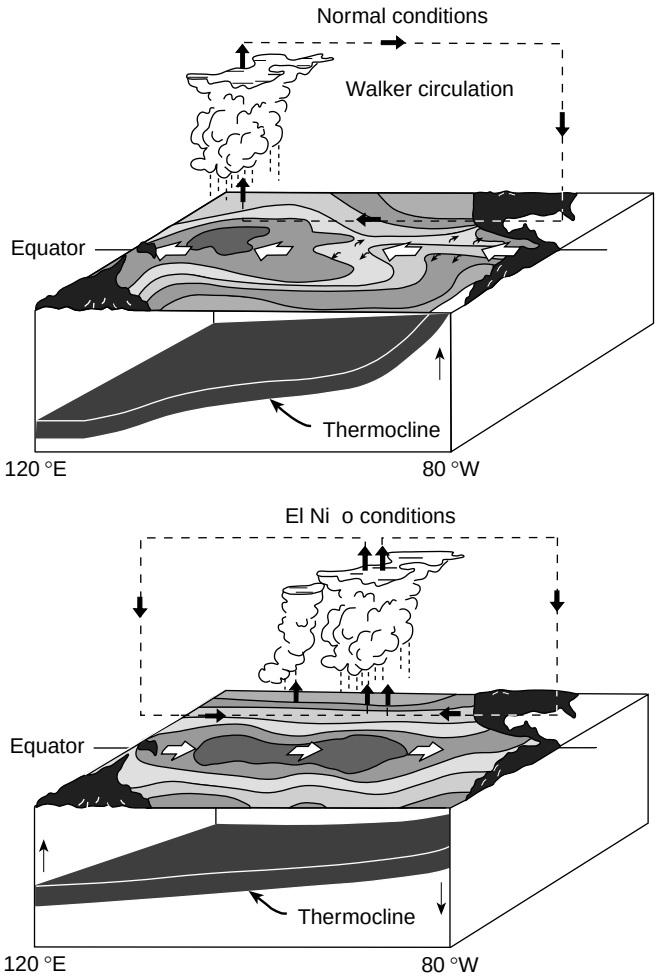


Figure 4 Schematic of normal and El Niño conditions in the equatorial Pacific

warm water in the west and high surface pressure with cooler water in the east. Also, as the trade winds flow from east to west, they pick up heat and moisture from the ocean. The warm, humid air mass becomes less dense and rises over the western Pacific warm pool where deep convection leads to towering cumulus clouds and heavy precipitation. Ascending air masses in this region of deep convection return eastward in the upper levels of the troposphere, then sink over the cooler water of the Eastern Pacific. Bjerknes labeled this atmospheric circulation pattern on the equatorial plane the Walker circulation in honor of Sir Gilbert Walker (*see* Walker Circulation, Volume 1).

Upwelling is a key oceanic process that regulates sea surface temperatures along the equator and along much of the west coast of the Americas. Wind-driven surface currents flow to the right of the winds in the Northern Hemisphere and to the left of the winds in the Southern Hemisphere because of the Coriolis force. Easterly trade wind-driven currents therefore flow poleward in opposite directions near the equator in the Northern and Southern

Hemispheres. This divergent surface flow is fed from below by upwelled thermocline water to create the equatorial cold tongue (Figure 1). A similar process operates off the west coasts of the Americas. Equatorward winds on the eastern flanks of subtropical high-pressure systems drive surface waters offshore. These offshore flows are fed from below by cooler thermocline waters, which then lower coastal sea surface temperatures.

During El Niño, the trade winds weaken in the western and central equatorial Pacific as atmospheric pressure rises in the west and falls in the east (Figure 4). Weakened trade winds generate waves in the ocean interior, which radiate along the equator both to the east and to the west away from the region of wind forcing (see Box 2). Over the course of a few months, these waves elevate the thermocline in the west and push it down in the east. In the cold tongue region, depression of the thermocline leads to surface warming as equatorial upwelling can no longer efficiently tap into the cold water reservoir below. Westward flow in the south equatorial current also weakens and reverses when the trade winds relax, allowing the western Pacific warm pool to migrate eastward. Sea level drops in the west and rises in the east in response to the redistribution of warm surface waters.

As sea surface temperatures warm east of the date line, the pattern of deep convection and precipitation also shifts eastward. This reinforces the reduction in trade wind intensity, because westerly winds flow into the convective center from the west. Pressure continues to fall in the east and rise in the west as the region of convection and rainfall migrates eastward. The system becomes locked in a positive feedback loop, with warming surface temperatures leading to weakened trade winds and vice versa. This positive feedback is eventually broken when the oceanic waves that elevated the thermocline in the western Pacific at the onset of the event bounce off the land masses bordering the western Pacific and reflect back toward the east. As the reflected waves propagate eastward along the equator, they elevate the thermocline and initiate surface cooling. Cooler surface waters in the eastern Pacific increase the east–west atmospheric pressure gradient and the strength of the trade winds, which in turn reinforces surface cooling through intensified upwelling. These interactions between the ocean and the atmosphere terminate El Niño typically 12–18 months after its onset. The ocean returns to near normal conditions, or it may overshoot into a cold La Niña state.

While El Niño involves an intimate coupling of the atmosphere and the ocean, the two fluids respond to each other on very different time scales. The atmosphere is a thousand times less dense than the ocean, and therefore responds almost immediately to forcing from the ocean surface. The ocean is massive by comparison, so its dynamical response to atmospheric forcing is relatively sluggish. For example, thermocline depths may take many months to adjust to changing winds in the tropics, whereas atmospheric convection can develop within hours over warm surface water. It is the slow evolution of the upper ocean thermal field that provides the memory for the climate system in the tropical Pacific, and it is this thermal inertia that provides the basis for predictability of ENSO time scale variations.

Box 2 Equatorial Waves

One early theory for El Niño postulated that the winds blowing equatorward off the west coast of South America weakened during an El Niño event, so that coastal upwelling would be reduced and sea surface temperatures would rise. However, Klaus Wyrtki, an oceanographer at the University of Hawaii, demonstrated in the mid-1970s that winds along the coast of South America actually strengthened during El Niño (Wyrtki, 1975). He found instead that weakening of the tradewinds thousands of kilometers to the west in the central Pacific was related to the development of El Niño along the coast of South America some months later. Based on these results, Wyrtki suggested that large scale equatorial ocean waves were the mechanism by which wind variations in the central Pacific could lead to the onset of El Niño in the eastern Pacific.

Two classes of oceanic waves are important for understanding the cycle of El Niño and La Niña variations. One class is referred to as Kelvin waves, named after Lord Kelvin (William Thompson), a 19th century British physicist who was the first to theoretically predict such waves in rotating fluids. The other class is referred to as Rossby waves, named after Carl-Gustaf Rossby, a Swedish-born meteorologist who in the 1930s first discovered this kind of wave in the atmosphere (*see* **Rossby Waves**, Volume 1; **Rossby, Carl-Gustaf**, Volume 1). Both types of wave are generated in the equatorial ocean by large-scale variations in surface winds.

Kelvin waves propagate eastward along the equator and Rossby waves propagate westward. They are evident below the surface as undulations of the thermocline, causing it to rise and fall by tens of meters as the waves pass by. Equatorial waves also affect sea level and the intensity and direction of ocean currents. The Earth's rotational forces trap these waves within several hundred kilometers of the equator in the open ocean, so they transfer energy very efficiently over many thousands of kilometers in the east–west direction.

Kelvin waves take about two months to cross the Pacific basin and Rossby waves take about six months to cross. When they reach the landmasses at the eastern and western boundaries of the ocean, they reflect back into the interior and, in the case of Kelvin waves, leak energy to higher latitudes along the west coasts of the Americas. The life cycle of these waves, evolving over many months and seasons in response to changing winds, are a critical aspect of ocean dynamics controlling the evolution of El Niña and La Niña events.

EL NIÑO WEATHER IMPACTS

One impact of the eastward shift in rainfall along the equator during El Niño is that drought develops in Australia, Indonesia and neighboring countries. On the other hand, the island states of the central Pacific and the west coast of South America are inundated with heavy rains. Heavy rainfall bands normally situated north and south of the equator in the Intertropical Convergence Zone and the South Pacific Convergence Zone also shift equatorward as surface waters warm. These latitudinal shifts contribute to unusually heavy rains near the equator in the central and eastern Pacific, and to drought conditions at higher latitudes in regions such as New Caledonia and Fiji to the south, and Hawaii to the north.

Heat released into the troposphere from deep tropical convection is one of the principal driving forces for the global atmospheric circulation. Changes in the location of tropical heat sources during El Niño therefore lead to widespread changes in wind and weather patterns outside the tropical Pacific. These remote effects of El Niño are referred to as teleconnections. In the extratropics, teleconnections are usually most pronounced during the winter season, though they are detectable in other seasons as well.

Anomalous tropospheric heating in the central tropical Pacific during El Niño generates quasi-stationary atmospheric wave trains that radiate poleward and eastward. In the Northern Hemisphere, these waves set up the Pacific North American Teleconnection Pattern, or (PNA) pattern, which is a series of high- and low-pressure centers extending from the central North Pacific to North America (Figure 5). During El Niño, the Aleutian low pressure center over the North Pacific deepens, high pressure develops over western North America, and low pressure prevails over the southeastern US. These pressure changes steer warmer air masses from southern latitudes into the Pacific Northwest and southern Canada, so that much of Canada and the northwestern US tend to experience mild winters. Low pressure in the southeastern US associated with the PNA pattern brings wetter, cooler conditions to the

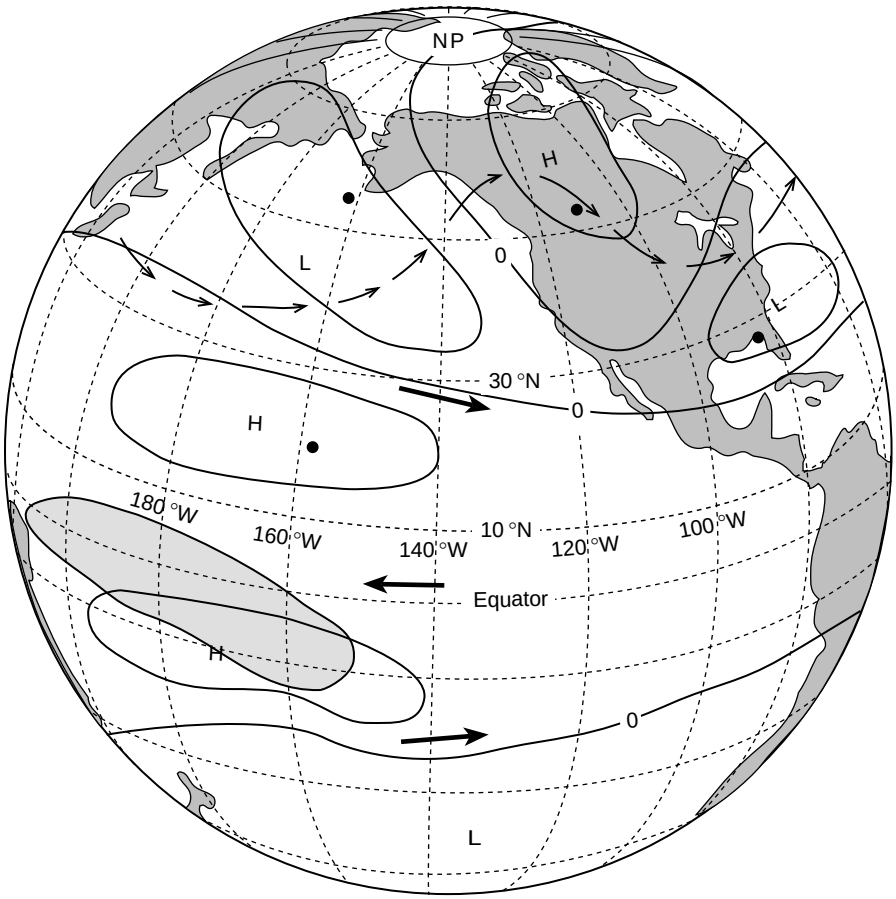


Figure 5 Schematic diagram of the PNA pattern of middle tropospheric pressure anomalies during Northern Hemisphere winter coinciding with an El Niño event. Dark arrows indicate strengthening of the subtropical westerly jet streams and, along the equator, weakening of the upper branch of the Walker circulation. Lighter arrows depict mid-tropospheric flows as distorted by the anomalous high and low pressure patterns. Rainfall and cloudiness are enhanced over the shaded region near the equator (Horel and Wallace, 1981)

states bordering the Gulf of Mexico. Similar wave trains are excited in the Southern Hemisphere, but they are weaker and more variable than those in the north.

Teleconnections also affect the subtropical jet streams which are swift air flows girdling the Earth at altitudes centered between about 10 000 m and 12 000 m (*see* Teleconnection, Volume 1). The eastward migration of deep convection during El Niño causes the core of these jet streams to intensify and shift southeastward in the central and eastern Pacific (Figure 5). The jet streams are a major influence on the location of storm tracks, so that southern California and northern Chile typically experience stormier and wetter weather in their respective winter seasons during El Niño.

Teleconnections affect surface conditions over land masses and the ocean. At extratropical latitudes of the North and South Pacific, for example, sea surface temperatures cool in response to intensified surface wintertime westerlies. Over land, changes in temperature and precipitation can affect soil moisture and evapotranspiration. These changes in surface conditions can feed back to the atmosphere, affecting the overall climatic response to El Niño forcing.

A global composite of typical rainfall and air temperature anomalies associated with El Niño is presented in Figure 6. In addition to effects already mentioned, El Niño years generally bring drought in northeastern Brazil, southern Africa, and the western Pacific, and wetter conditions to southern Brazil, Uruguay, Peru, and equatorial East Africa (Ropelewski and Halpert, 1987). Indian monsoon summer rains tend to be weaker during El Niño events, consistent with the strong relation between the Southern Oscillation and Indian monsoon rainfall noted by Walker.

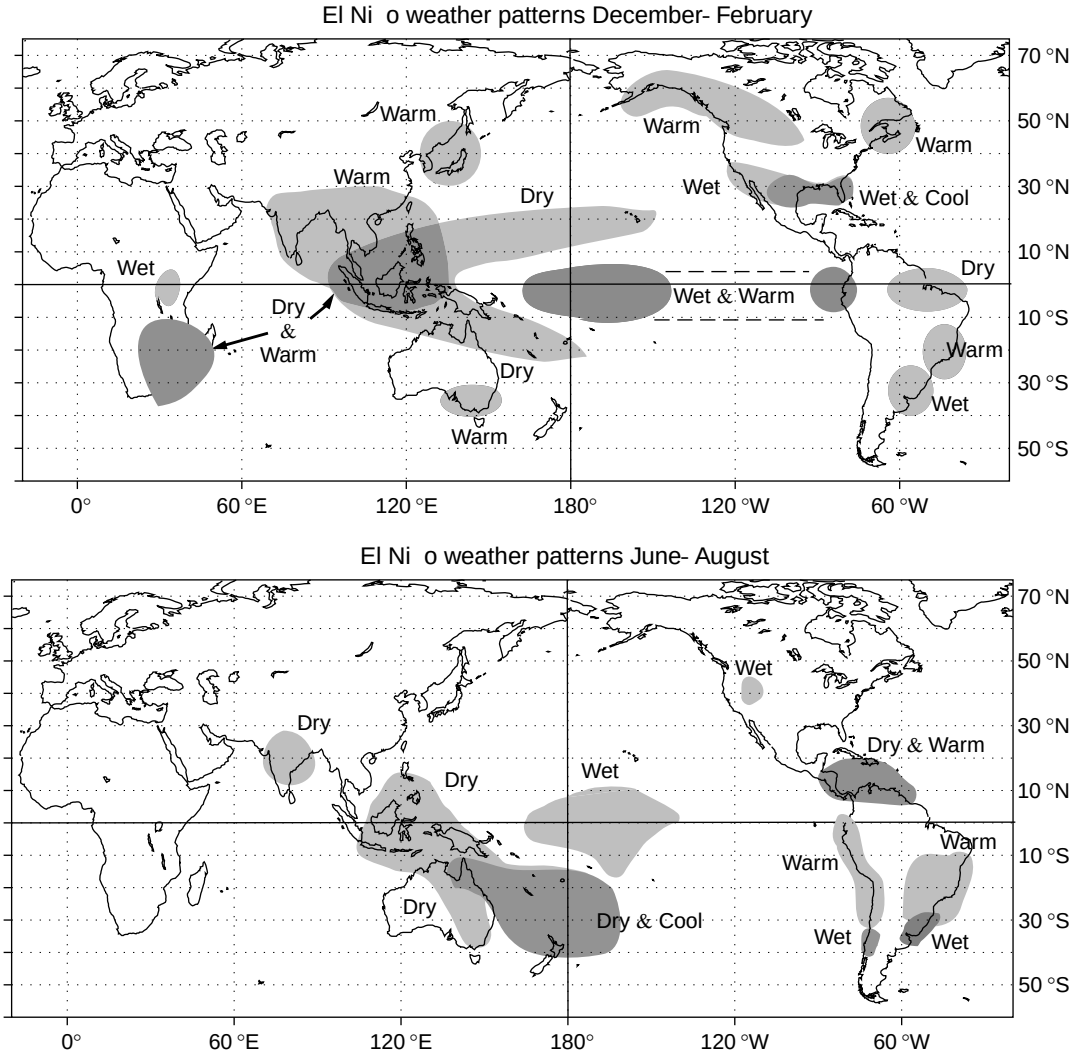


Figure 6 Schematic diagram showing temperature and precipitation anomalies associated with El Niño. To a first approximation, the impacts of La Niña are similar, but with opposite sign (Ropelewski and Halpert, 1987 and Halpert and Ropelewski, 1992)

El Niño also affects tropical storm frequency, intensity, and spatial distribution (Gray, 1984). In the Atlantic, 10 named tropical storms, of which six reach hurricane intensity (wind speeds greater than 33 ms^{-1}), form during a typical June to November hurricane season. However, during El Niño, an intensified subtropical jet stream shears off the top of fledgling storms before they fully develop. Thus, fewer tropical storms and hurricanes form, and those that do form tend to be weaker and shorter lived. The summer of 1997 was a particularly striking example of reduced Atlantic hurricane activity during El Niño. Only seven named storms formed, three of which reached hurricane intensity.

Intense tropical storms in other regions (referred to as hurricanes in the northeastern Pacific, typhoons in the northwestern Pacific, and cyclones in the Indian ocean and Southern Hemisphere) are affected more in terms of their spatial distribution and intensity than in terms of total numbers (*see* Hurricanes, Typhoons and other Tropical Storms - Descriptive Overview, Volume 1; Hurricanes, Typhoons and other Tropical Storms - Dynamics and Intensity, Volume 1). Heat from the underlying ocean surface is a major source of energy for intense tropical storms, which are almost always generated over waters warmer than 26 °C. Therefore, extensive warming of the eastern Pacific during El Niño allows hurricanes to travel farther westward and northward and, in some cases, attain greater strength. Similarly, the eastward extension of warm tropical surface waters along the equator during El Niño expands the area over which typhoons and cyclones may spawn. Changes that can occur in tropical storm activity in the eastern Pacific during El Niño are exemplified by the 1997 season when hurricane Linda grew to record strength and hurricane Nora tracked far to the north, bringing heavy rains to normally arid northern Mexico and the southwestern US. In the South Pacific during the 1982–1883 El Niño, ocean surface warming allowed six cyclones to strike French Polynesia, a region not usually prone to experiencing intense tropical storms (Canby, 1984).

El Niño affects the coastal zones of the Americas all the way to Alaska in the Northern Hemisphere and at least to central Chile in the Southern Hemisphere. When eastward propagating oceanic waves generated by weakening of the trade winds at the onset of El Niño reach the coast of South America, part of the wave energy is reflected back into the open ocean, but part also radiates northward and southward along the coasts. The coastal waves push the thermocline down and reduce the upwelling of cool water as they progress poleward. Within one to two months after the first equatorial waves reach South America, sea surface temperatures rise all along the coast to the north and south. North of central California, southerly surface winds related to the PNA pattern also contribute to coastal warming (Ramp *et al.*, 1997). These winds both bring warm subtropical air northward, and also drive on-shore ocean currents that converge at the coast to push the thermocline down.

The unusually thick layer of warm surface water along the west coast causes near shore sea levels to rise by 15–30 cm during El Niño. In addition, offshore winter storms associated with intensified jet streams in the subtropics can generate unusually high ocean swells and storm surges in coastal zones. Storm-generated surface waves and currents, and high tides, can add to already elevated sea levels high tides to cause severe coastal erosion. In communities with major shoreline development such as in southern California, the combined assault from high seas and heavy rains can lead to significant loss of property and life.

The impacts of El Niño on weather are most consistent from event to event in the tropical Pacific and bordering areas. Impacts are prominent, but less consistent, at higher latitudes and in other ocean basins where regional influences can become important. Thus, although El Niño increases the probability of a particular kind of weather pattern occurring in particular regions of the globe, actual impacts may vary from those expected for any given event (Trenberth *et al.*, 1998).

ECOSYSTEMS IMPACTS

Narrow coastal and equatorial upwelling zones account for much of the biological productivity in the world s oceans. One-celled plants known as phytoplankton are at the base of the aquatic food chain in these regions, and all higher organisms depend directly or indirectly on them as a food source. Upwelling ecosystems are rich in species diversity and abundance, including many types of phytoplankton, zooplankton, shellfish, crustaceans, fish, marine mammals, and sea birds. Many of the fish and shellfish species such as tuna, anchoveta, sardines, squid, and shrimp are valuable commercially.

Primary productivity is the rate at which carbon is taken up by phytoplankton via photosynthesis. Nutrient limitation controls primary productivity in tropical and subtropical regions where sunlight is plentiful. The principal source for nutrients, away from river outflows and areas of significant coastal runoff, is the thermocline. Small organisms that have died near the surface decay and remineralize as they slowly sink out of the euphotic zone (the well lit upper layer of the ocean). Fecal pellets from larger organisms likewise sink and decompose at depth. These processes create a pool of nutrients and remineralized carbon in the thermocline which, when upwelled into the euphotic zone, are available for uptake by phytoplankton. Lateral flows in the thermocline may bring nutrients and carbon from great distances before they are upwelled, concentrating primary production in narrow equatorial and coastal zones.

During El Niño, ocean dynamical processes depress the thermocline in the eastern and central equatorial Pacific, and along the coasts of North and South America. The supply of nutrients to the euphotic zone drops or may be cut off entirely. Primary production decreases, with effects that ripple through the entire food chain (Barber and Chavez, 1984). Zooplankton that feed on phytoplankton decrease in abundance. Fish, sea birds and marine mammals die off or migrate to more productive regions in search of food. Undernourished sea birds and marine mammals may experience reproductive failures or abandon young when food becomes scarce. In extreme cases, decimated populations may require one or more years to fully rebound.

El Niño can also cause bleaching of tropical corals when water temperatures become too warm (Strong *et al.*, 2000) (*see* Coral Reefs: an Ecosystem Subject to Multiple Environmental Threats, Volume 2 and Coral Bleaching (1997- 1998), Volume 2). Corals thrive in a narrow range of temperatures between about 18 and 28 °C. At temperatures above the upper threshold, they expel symbiotic algae that reside in their polyps. The algae add color to healthy corals, and their metabolic by-products are essential for coral survival. Coral can recover from bleaching unless water temperatures remain too high for too long. The concern over lethal bleaching is that it threatens the vitality of coral reef ecosystems, which support local fisheries and provide tourist income. Massive and widespread coral bleaching occurred during 1998 in the Galapagos Islands, off the coast of Panama, in the Great Barrier Reef of Australia and elsewhere in the tropics in response to the exceptionally strong 1997–1998 El Niño. Decadal warming trends in tropical ocean temperatures contributed to this bleaching, by elevating background temperatures on which El Niño warming was superimposed. These decadal trends, possibly related to global warming, have raised concerns about prolonged deterioration in the health of coral reefs.

El Niño can also dramatically affect fisheries, a particularly striking example of which was the collapse of the Peruvian anchoveta fishery following the 1972–1973 El Niño (Glantz, 2001). In the late 1960s and early 1970s, Peru was the most productive fishing nation in the world. The annual catch of anchoveta was over 10 million metric tons, one fifth of the global total. Most of the catch was exported for use as a feed supplement in poultry farms, with the exports accounting for nearly one third of Peru s foreign exchange earnings. Concerns were voiced at the time that overfishing might be depleting the stocks from sustainable levels. Those concerns proved to be well founded. Intense fishing pressure and extraordinarily high mortality rates during the 1972–1973 El Niño caused the fishery to crash. Effects were long lasting. For at least 10 years following the 1972–1973 El Niño, anchoveta landings ranged between only 1/10–1/3 the catch of 1970.

It was the collapse of the Peruvian anchoveta fishery that focused international attention on the socio-economic consequences of El Niño. A reduction in the supply of fishmeal caused many farmers in the US to switch from planting wheat to planting soybeans. Soy is an alternative source of poultry feed supplement and more lucrative than wheat as a cash crop. However, wheat harvests had declined in many nations, particularly the Soviet Union, because of widespread droughts in 1972. Staple food shortages developed in some regions and prices steeply increased for basic commodities in others. These events highlighted the role of climate in affecting global economies, and specifically focused attention on the need for a better understanding of the El Niño phenomenon.

DEVELOPMENT OF ENSO OBSERVING AND FORECASTING SYSTEMS

Research efforts in both oceanography and meteorology in the decade following the 1972–1973 El Niño gathered momentum, motivated by heightened awareness of El Niño s global socio-economic consequences, and the prospect that it might be predictable. It was recognized that advance warning of an impending El Niño would potentially be extremely valuable for either disaster preparedness or for exploiting opportunities created by altered environmental conditions. Thus, in the early 1980s, the international research community began mobilizing resources for a major 10-year study of El Niño. This study, called the Tropical Ocean-Global Atmosphere study, or (TOGA), took place from 1985–1994.

The 1982–1983 El Niño, the strongest in over 100 years, occurred as plans for TOGA were being formulated. This El Niño left a global swath of devastation in its wake. It had not been predicted because El Niño forecasting capabilities had not yet been developed. Even more stunning to the research community, however, was that it was not even detected until nearly at its peak. At the time, most oceanographic data were available for analysis only months, or in some cases years, after they had been collected. A handful of scattered but more readily available ocean and weather data from islands and volunteer observing ships suggested the development of unusually warm conditions in mid-1982, but these reports were discounted as erroneous. One reason was that the 1982–1983 El Niño developed differently than expected from conventional wisdom. There had been no prior strengthening of the trade winds, and no warming off the west coast of South America in early 1982, both considered at the time to be necessary precursors of an impending El Niño. To complicate matters, eruptions of the Mexican volcano El Chichón in March–April 1982 injected a massive cloud of aerosols (dust, soot, and other fine particles) into the lower stratosphere, where prevailing winds spread it around the globe at low latitudes within

weeks. This cloud of volcanic aerosols produced undetected cold biases of several degrees Celsius in satellite measurements of sea surface temperature. It was only after definitive reports were received from a research cruise in the eastern equatorial Pacific that the scientific community realized to what extent existing data sources and simplistic notions of El Niño evolution had misled them.

The lessons from this experience were clear: there was urgent need for improved monitoring, detection, and understanding of El Niño, as well as an urgent need to develop reliable El Niño forecast models. Observational requirements were met in part by utilizing data from a constellation of satellites viewing the Earth from space. However, satellites require Earth-based measurements for calibration and verification. Also, satellites do not see below the ocean surface, where critical oceanic processes operate to produce El Niño events. An ocean observing system specifically designed for El Niño was therefore developed under the auspices of the TOGA program (McPhaden *et al.*, 1998) (*see* TOGA (Tropical Ocean Global Atmosphere), Volume 1). It included an extensive network of moored buoys, drifting buoys, volunteer observing ships, and tide gauges (Figure 7). One of the most important attributes of this system was that most of the data were transmitted via satellite relay to shore stations within hours of collection. The observing system was implemented through an international collaborative effort spanning the full 10 years of TOGA, and was completed only in the last month of the program (December 1994). Continued after TOGA, it is now called the ENSO Observing System. The 1997–1998 El Niño was the first for which the ENSO Observing System was in place from start to finish, so that this event was not only among the strongest on record, but also the best ever documented (McPhaden, 1999).

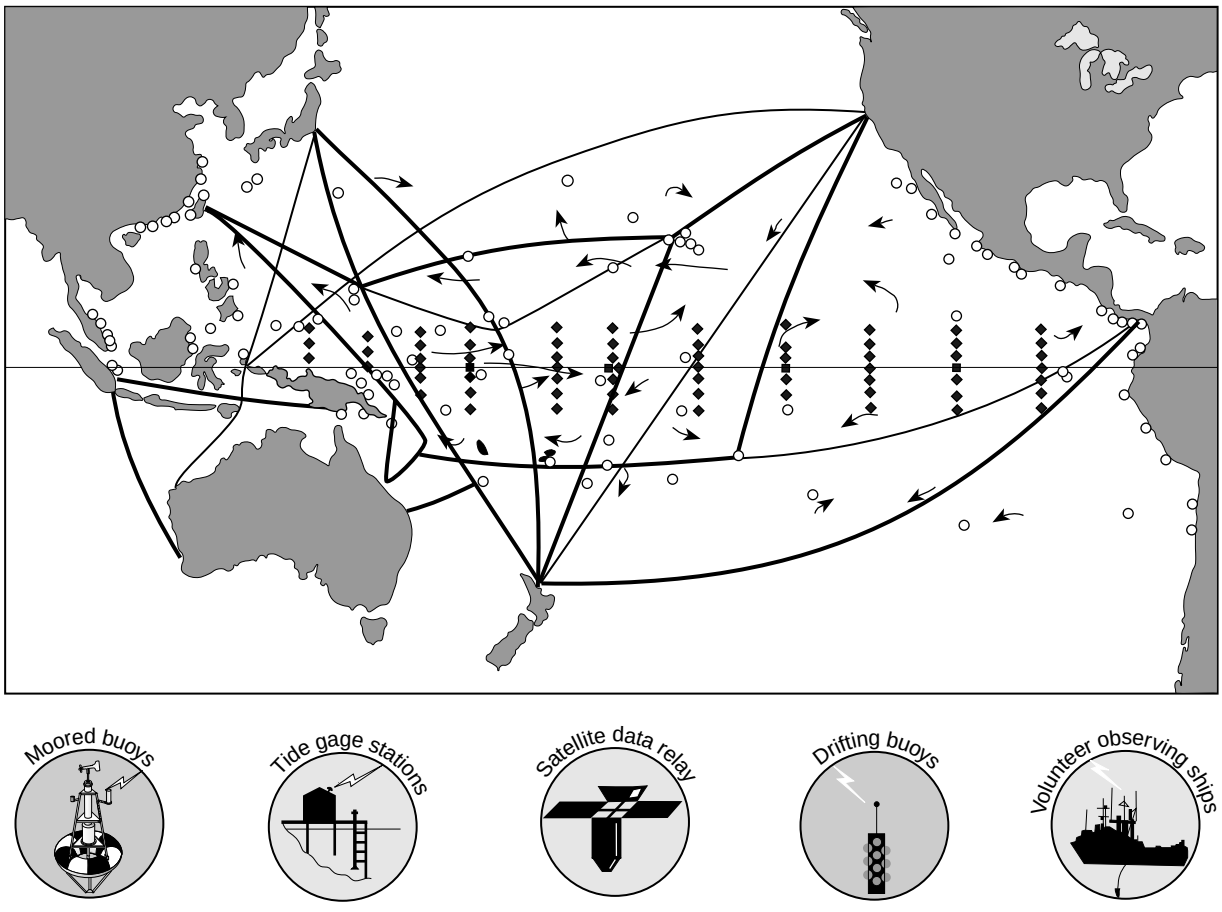


Figure 7 The ENSO Observing System. The four major elements of this observing system are; (1) a volunteer observing ship program for surface marine meteorological observations and ocean temperature profiles (shown by schematic ship tracks); (2) an island and coastal tide gauge network for sea level measurements (circles); (3) a drifting buoy network for sea surface temperatures and ocean currents (shown schematically by curved arrows); (4) a moored buoy array for wind, ocean temperature, and ocean current measurements (squares and diamonds). Thick ship tracks indicate repeat transects 11 or more times per year; thin ship track indicate repeat transects six to ten times per year. One schematic drifting buoy arrow represents 10 actual randomly distributed drifters (McPhaden *et al.*, 1998)

Progress in the development of the ENSO Observing System has been paralleled in the development of ENSO forecast models (National Research Council, 1996). These models range from statistical approaches based on the average behavior of previous ENSO variations, to complex dynamical models that explicitly represent the physical processes at work in the coupled ocean-atmosphere system. The 1986–1987 El Niño was the first to be successfully predicted with these techniques, which have since undergone continual refinement and improvement. As one measure of success, many ENSO forecast models predicted that 1997 would be unusually warm in the tropical Pacific at least one to three seasons in advance (Barnston *et al.*, 1999). Long-range weather forecasting schemes that included information about tropical Pacific conditions during the 1997–1998 El Niño likewise had success in predicting surface air temperature and precipitation patterns in widely disparate parts of the globe months in advance. For instance, the forecast for wintertime precipitation and temperature issued by the National Centers for Environmental Prediction in Washington, DC, in the fall of 1997 were among the most accurate ever for the continental US.

Successful seasonal forecasts, unprecedented high definition ocean measurements from the ENSO Observing System, and record warmth in the tropical Pacific all combined to capture the attention of the public in 1997–1998. Media coverage was so intense that El Niño became a household word all over the world. As a result of the heightened public awareness about El Niño and its consequences, many individuals, municipalities, businesses, and in some cases national governments mobilized resources in an effort to prepare for El Niño s onslaught. It is likely that without the advance warning made possible by new measurement and forecasting capabilities, the toll in terms of lives, property damage, and economic losses due to the 1997–1998 El Niño would have been much higher.

SOCIO-ECONOMIC IMPACTS OF THE 1997–1998 EL NIÑO

The 1997–1998 El Niño affected the lives of tens of millions of people around the globe. Significant disruptions were experienced in agriculture, forestry, fisheries, transportation, communications, power generation, public health and other climate-sensitive areas of human endeavor. Weather-related disasters and disease outbreaks during this El Niño claimed over 22 000 lives worldwide and caused US$36 billion dollars in economic losses (Sponberg, 1999).

The 1997–1998 El Niño brought torrential rainfalls and flooding to parts of California, the southeastern US, equatorial East Africa and Chile. It was also responsible for severe droughts in Papua New Guinea, Indonesia, Central America, and northeastern Brazil (Supplee, 1999; WMO, 1999). Many regions experienced extensive crop failures and livestock losses due to these droughts and floods. Life-threatening food and drinking water shortages developed in New Guinea, prompting international relief efforts. In some drought-stricken regions, forest fires raged for months, devastating the countryside. The drought and fires in Indonesia produced deadly smog that covered an area one half the size of the continental US, causing widespread respiratory ailments, and contributing to airline and shipping disasters. Unusually low water levels in the lakes that feed the Panama Canal forced officials to restrict traffic through the waterway for the first time in 15 years.

Flood-contaminated water supplies in some regions contributed to outbreaks of cholera and dysentery. Stagnant pools of floodwater also provided ideal breeding grounds for mosquitos and other insects that spread infectious diseases like malaria and dengue fever. Public health emergencies developed in Southeast Asia, South America, and Africa as these diseases afflicted thousands of people.

Heavy rains and flooding from El Niño damaged or destroyed roads, bridges, buildings and phone and power lines in many parts of the globe. Thousands of individuals and families lost homes and personal property due to storms, floods, mud slides and land slides. Hydroelectric energy generation was curtailed in regions of severely reduced rainfall and stream flow. Many business lost income as a result of weather-related closures, commodity shortages or shifts in market demand for consumer goods. Tourism was affected because of extreme weather in areas such as southern California and Florida. The catch of Peruvian anchoveta fell sharply as in previous El Niño years, and the fishery was briefly closed by government decree as a precautionary measure to limit further depletion of the remaining stock. Economic hardship followed severe declines in catch experienced in three of the four major fisheries (tuna, squid, sardine) off the west coast of Baja California.

Despite losses of life, property, and income, 1997–1998 El Niño produced some benefits as well (Changnon, 1999). Losses from Atlantic hurricanes were greatly reduced, as only one of three hurricanes that formed made landfall. Parts of the US Midwest and the Great Lakes region experienced the mildest winter in over 100 years, as temperatures warmed to record highs between November 1997 and February 1998. The milder than normal winter reduced heating bills, weather-related travel delays for the transportation industry, and deaths from exposure. It has been estimated that for the US as a whole, the 1997–1998 El Niño produced a net economic gain of about $16 billion dollars and resulted in 650 fewer deaths than would have otherwise occurred.

Similarly, not all fisheries were hit hard by the 1997–1998 El Niño. Though severe declines in most

commercial fisheries were experienced off the west coast of Baja California, these declines were partially offset by an eight-fold increase in the catch of more lucrative shrimp. Likewise, shrimp catch increased off the coast of Ecuador, with export revenues rising by 40% in 1997 (WMO, 1999). Northward migration of warm water fish species such as marlin, tuna and mahi-mahi was also a boom to the sport-fishing industry off the west coast of the US.

The 1997–1998 El Niño highlighted societal sensitivities and vulnerabilities to year-to-year climate fluctuations on a global scale. It also demonstrated that losses and gains are inherently unevenly distributed across economic sectors and national boundaries. Spurred in part by the dramatic consequences of this El Niño, new areas of interdisciplinary research are emerging to examine the use of El Niño-related climate information in economic planning and policy decisions. These efforts involve both social and physical scientists, with a focus on not only how forecasts are generated, but also how they are disseminated, interpreted, and acted upon.

LA NIÑA

La Niña is the cold phase of the ENSO cycle (Philander, 1990). It is characterized by stronger than normal trade winds, colder tropical Pacific sea surface temperatures, and positive values of the SOI (Figure 3). It represents a situation in which oceanic and atmospheric processes which give rise to normal conditions illustrated in Figure 4 are exaggerated, with deep convection and heavy rainfall along the equator shifted even further to the west and confined to a narrower range of longitudes. The term La Niña (*the girl*) was coined in the mid-1980s by scientists investigating the year-to-year oscillations between warm and cold conditions in the tropical Pacific. La Niña has also been referred to as anti-El Niño, ENSO cold event, and El Viejo (the old man). Like El Niño, La Niña typically lasts 12–18 months.

La Niña s effects on global weather variability are roughly (though not exactly) opposite to those of El Niño (Halpert and Ropelewski, 1992). The atmospheric response to weak-to-moderate cold and warm tropical Pacific sea surface temperature anomalies tends to be similar in magnitude, but opposite in sign. However, the atmospheric response to strong La Niña events tends to be weaker than the response to strong El Niño events. This difference arises because tropical rainfall (which translates into atmospheric heating and associated teleconnection patterns) has a lower limit of zero, regardless of how cold the tropical Pacific may get. On the other hand, the upper limit for rainfall and hence atmospheric heating during strong El Niño events is not as rigidly constrained.

Examples of La Niña weather impacts include an increased probability of unusually rainy conditions in southern Africa and northeastern Brazil, and in the monsoon regions of India, Indonesia and Northern Australia. La Niña is often linked to drier than normal conditions over equatorial East Africa, southern Brazil and Uruguay. The subtropical jet streams over the Pacific weaken, shift poleward, and become more variable in strength during La Niña winters. Blocking highs tend to develop in the eastern North Pacific, accompanied by more frequent cold polar air outbreaks over western North America. These changes in atmospheric pressure patterns and circulation result in winters that tend to be colder and stormier than normal over Alaska, western and central Canada, and the northern US. Conversely, the northward shift in the subtropical jet stream leads to warmer and drier than normal conditions in the southeastern US, and drier than normal conditions in the southern plains of the US.

La Niña, like El Niño, affects tropical storm frequency, intensity, and geographical distribution through changes in sea surface temperature and atmospheric circulation. Pacific hurricanes, typhoons and cyclones are more restricted in their geographical extent because of colder underlying sea surface temperatures. On the other hand, Atlantic hurricanes tend to increase in number as the subtropical jet stream shifts northward, reducing the shear between upper and lower level tropospheric winds where hurricanes form. The 1995 Atlantic hurricane season is a good example of what can happen during La Niña. That year witnessed a bumper crop of 19 named tropical Atlantic storms, including 11 hurricanes, almost double the usual number.

La Niña s socio-economic impacts can be as impressive as those resulting from El Niño. La Niña-related drought and heat waves in the US Midwest during the summer of 1988 constituted one of the worst natural disasters in US history, causing US$40 billion in agricultural and other losses, and claiming approximately 7500 lives from heat stress. Also, with more and stronger Atlantic hurricanes during La Niña, their destructive potential and probability of making landfall in populated regions increases. For the US, it is three times more likely that an intense hurricane will make landfall during a La Niña than during an El Niño. Hurricane Mitch, one of the strongest Atlantic hurricanes on record and the deadliest in 200 years, was spawned during the 1998 La Niña. Mitch devastated Central America, claiming 10 000 lives, leaving over three million homeless, and causing US$6 billion in damage.

La Niña can also be beneficial to some regions of the globe. For instance, higher monsoon rainfall totals over the Indian subcontinent, the western Pacific, and northeastern Brazil can support greater agricultural production and economic growth. Enhanced winter snowpack

in the mountains of the Pacific Northwest provide for extra hydroelectric power production, ample summer water supplies, and improved freshwater habitat for salmon. In terms of marine ecosystems, primary productivity, driven by more intense equatorial and coastal upwelling, is generally enhanced. Higher-level marine organisms tend to thrive under these conditions, and severe disruptions in fisheries, like those during El Niño, are not so evident.

DECADAL VARIATIONS AND GLOBAL WARMING

The frequency and amplitude of El Niño and La Niña events are modulated on decadal and longer time scales. The 1930s to the early 1950s, for example, was a period of relatively few El Niños, while the period since 1976 has witnessed more frequent, stronger, and in the case of 1991–1995, longer lasting El Niños. In addition, the magnitude of the strongest El Niños over the past 50 years seems to be increasing while at the same time there has been a trend toward fewer La Niña events (Figure 3).

Variations in the ENSO cycle may simply be a result of random or chaotic fluctuations in the Earth s highly complex climate system. However, there may be other factors at work as well. Temperatures have been elevated in the tropical Pacific since the mid-1970s in association with a recently discovered, naturally occurring phenomenon known as the Pacific Decadal Oscillation, or (PDO) (Mantua *et al.*, 1997) (*see* Paci c- Decadal Oscillation, Volume 1). This oscillation alternates between phases of warm and cold tropical Pacific sea temperatures lasting 20–25 years. The physical mechanisms that account for the PDO are at present poorly understood. However, the PDO affects the background oceanographic and atmospheric conditions on which ENSO events develop. Warm phase PDO and El Niño sea surface temperature anomalies are roughly additive. Under these conditions, the global atmospheric teleconnection patterns set up by elevated tropical Pacific sea surface temperatures can be amplified.

In addition to naturally occurring phenomena such as the PDO, global warming may also affect the ENSO cycle (Trenberth and Hoar, 1996). The warmest years on record were 1998 and 1997, in that order. Occurrence of 1997–1998 El Niño contributed in part to these extremes, because global temperatures rise a few tenths of a degree Celsius following the peak El Niño warming in the tropical Pacific. However, aside from the record warmth in 1997–1998, there has been an underlying trend toward increased globally averaged air temperatures of about 0.6 °C over the past century. Moreover, tree ring and ice core data indicate that the 20th century has been the warmest century and the 1990s has been the warmest decade over the past millennium. The consensus of the scientific community at present is that the recent increase in global temperatures is human-induced through fossil fuel combustion and deforestation (Houghton *et al.*, 1996).

It is reasonable to assume that anthropogenic greenhouse gas warming will affect ENSO because both phenomena are involved in regulating the Earth s heat balance. In addition, an altered ENSO cycle may leave its imprint on the global carbon balance (see Box 3) and on patterns of regional climate change. If global warming were to lead to intensified, more frequent, or prolonged El Niños, those regions affected by El Niño-related drought or flood in the present climate might well be more prone to such natural disasters in the future. How global warming may affect ENSO and other natural modes of climate variability is not known at present though. Computer models used for developing scenarios of how the Earth s climate will change in response to elevated levels of carbon dioxide (CO_2) and other greenhouse gases in the atmosphere differ on how the tropical Pacific Ocean will be affected (Meehl *et al.*, 2000). Many coupled ocean-atmosphere models predict a permanent El Niño-like warming of the tropical Pacific due to increased forcing from greenhouse gases. Some models also exhibit more energetic and higher frequency year-to-year ENSO-like fluctuations superimposed on this permanent warming. However, realistic greenhouse gas forcing in other models can lead to quite different scenarios. The differences among model simulations arise in part because results are very sensitive to the representation of poorly understood but critical processes such as the formation of clouds and their radiative feedbacks on the global energy balance, and the transport of heat by ocean currents and turbulent mixing. Thus, uncertainties in the simulated effects of global warming on ENSO and ENSO-like variability are large, and further model improvements are required to gain confidence in the details of future climate projections based on such simulations.

Instrumental records have also been studied extensively for clues about the relation between ENSO and global warming. Unfortunately, these records are limited to the past 100–150 years, and are too short to describe the full range of natural climatic variability on inter-annual and decadal time scales. Thus, from instrumental data alone it is not possible to unambiguously answer the question of whether the unusual character of ENSO during the past 25 years is within the range expected for natural variations, or whether it is anthropogenically forced. On the other hand, paleoclimatic reconstruction based on proxy data allow for investigation of ENSO variability from periods predating the instrumental record. For example, year-to-year fluctuations in the Earth s climate are evident in New England lake sediments dating from near the end of the last ice age (15 000 years before present) when the Earth was cooler (Rittenour *et al.*, 2000), and

Box 3 El Niño and the Global Carbon Cycle (*see* **Carbon Cycle**, Volume 2)

Each year, seven billion tons of carbon in the form of CO_2 are released into the atmosphere from anthropogenic sources. About half this amount remains in the atmosphere, while the other half is taken up approximately equally by the ocean and the terrestrial biosphere. Ocean uptake occurs at high latitudes where the deep waters of the world's oceans form. Uptake also occurs in the subtropics where surface waters are carried down into the thermocline.

The tropical oceans are a major source of CO_2 to the atmosphere, with the Pacific Ocean dominant because of its great size. Equatorial upwelling outgases about one billion tons of carbon in the form of CO_2 per year into the atmosphere, as water rich in inorganic carbon is brought up from the thermocline to the surface. The amount of carbon released to the atmosphere in the equatorial Pacific would be even higher were it not for photosynthesis, which converts one billion tons of upwelled inorganic carbon each year into living organisms.

Year-to-year variability in global atmospheric carbon concentrations is dominated by the ENSO cycle (Rayner *et al.*, 1999). During El Niño, equatorial upwelling is suppressed in the eastern and central Pacific, significantly reducing the supply of CO_2 to the atmosphere oceanic (Feely *et al.*, 2000). As a result, the global rise in CO_2 noticeably slows down during the early stages of an El Niño event. However, during the later stages of El Niño, global CO_2 concentrations rise sharply, possibly reflecting the delayed response of the terrestrial biosphere to El Niño-induced changes in weather patterns. Widespread droughts and elevated temperatures in the tropics during El Niño contribute to an increase in the number of forest fires, modifying the balance between respiration and photosynthetic uptake of CO_2 in land plants. These processes could result in an anomalous increase in the supply of CO_2 to the atmosphere sufficient to override the reduction in CO_2 from decreased equatorial upwelling.

In addition to year-to-year variations, lower frequency changes in the ENSO cycle can affect the global atmospheric carbon balance. In particular, the 1990s was an unusual decade in that it witnessed both a prolonged El Niño in 1991–1995 and an extraordinarily strong El Niño in 1997–1998. These El Niño events resulted in an overall decrease of oceanic input to the atmosphere of about two to three billion tons of carbon during the 1990s relative to the 1980s (Feely *et al.*, 2000).

in fossil Pacific coral records from the last interglacial period (124 000 years before present) when the Earth was slightly warmer (Hughen *et al.*, 1999). These and other proxy records exhibit decadal modulations in the amplitude of ENSO time scale variations similar to those observed today. Moreover, compared to some coral climate reconstructions, the increased frequency and strength of El Niño events since 1975 could be considered unusual relative to earlier epochs.

However, using paleocliamtic data as a benchmark for evaluating global warming effects in the modern ENSO record is complicated by the fact that background climatic conditions on which ENSO developed during earlier periods of the earth s history have frequently been very different from those which exist today (Tudhope *et al.*, 2001). A further complication is that there are relatively few reliable proxy records capable of defining year-to-year variations in the distant past, and those records are sparsely distributed in both space and time. Finally, the information in proxy records is subject to contamination by non-climatic chemical, biological and physical changes related to micro-environmental factors. Therefore, conclusions based on comparisons with proxy data about the effects of global warming on the ENSO cycle are subject to considerable uncertainty at this time.

CONCLUSION

The dramatic 1997–1998 El Niño highlighted progress made in ENSO research and forecasting since the pioneering work of Bjerknes nearly four decades ago. New observational techniques now allow us to take the pulse of the tropical Pacific day-by-day, and computer forecast models can predict the broad outlines of ENSO-related climatic perturbations up to one year in advance. The accuracy of these models is sufficient in some cases to allow for practical applications of forecast information for planning and mitigation purposes.

These scientific advances, though impressive, are just the beginning. Theories and observations have illuminated many basic features of the ENSO cycle, but fundamental questions such as why the trade winds initially weaken at the onset of El Niño still appear to have no simple answer. How ENSO interacts with the PDO and how it may be affected by global warming are not at present well understood. Though satellites can monitor the globe from space, comprehensive ocean-based observational networks as required for climate research and prediction are limited primarily to the tropical Pacific. This limitation hinders attempts to examine interactions between the ocean and the atmosphere in regions outside the tropical Pacific where important physical processes may affect, or be affected by, ENSO variations.

ENSO forecast systems are in their infancy, in some ways analogous to the early stages of numerical weather forecasting 40 years ago. As might be expected, despite the many seasonal forecasting successes during 1997–1998, there were notable failures. None of the forecast models predicted the ultimate intensity of the El Niño before its onset, and one previously successful model failed completely.

Similarly, predictions of severe drought in Australia and Zimbabwe, and reduced summer monsoon rainfall over India, proved to be erroneous. The reasons for these failures have yet to be fully determined.

The oscillation between warm, neutral, and cold states in the tropical Pacific associated with the ENSO cycle involves massive redistributions of upper ocean heat content. For instance, the accumulation of excess heat in the eastern Pacific due to the eastward movement of warm water and to the depression of the thermocline during a strong El Niño is approximately equivalent to the output of a million 1000 megawatt power plants operating continuously for a year. The magnitude of these natural variations clearly indicates that society cannot hope to consciously control or modify the ENSO cycle. Rather, we must learn to better predict it, and to adapt to its consequences.

The challenge for physical scientists then is to improve ENSO forecast models, to improve our understanding of underlying physical processes at work in the climate system, and to improve the observational data base in support of these objectives. Capitalizing on advances in the physical sciences for practical purposes is a challenge for social scientists, economists, politicians, business leaders, and the citizenry of those countries affected by ENSO variations. The promise of the future is that continued research on ENSO and related problems will be rewarded with scientific breakthroughs that translate into a broad range of applications for the benefit of society.

REFERENCES

Barber, R T and Chavez, F P (1983) Biological Consequences of El Niño, *Science*, 222, 1203–1210.

Barnston, A G, Glantz, M H, and He, Y (1999) Predictive Skill of Statistical and Dynamical Models in SST Forecasts During the 1997–1998 El Niño Episode and the 1998 La Niña Event, *Bull. Am. Meteorl. Soc.*, 80, 217–243.

Canby, T Y (1984) El Niño s Ill Wind, *Natl. Geogr.*, 165(2), 144–183.

Changnon, S A (1999) Impacts of the 1997–1998 El Niño-Generated Weather in the US, *Bull. Am. Meteorl. Soc.*, 80, 1819–1827.

Feely, R A, Boutin, J, Cosca, C E, Dandonneau, Y, Etcheto, J, Inoue, H Y, Ishii, M, LeQuere, C, Mackey, D, McPhaden, M J, Metzl, N, Poisson, A, and Wanninkhof, R (2000) Seasonal and Inter-Annual Variability in CO_2 in the Equatorial Pacific, *Deep-Sea Res.*, in press.

Glantz, M H (2001) *Currents of Change: El Niño's Impact on Climate and Society*, Cambridge University Press, Cambridge, 1–252.

Gray, W M (1984) Atlantic Hurricane Frequency: El Niño and 30 mb Quasi-biennial Oscillation Influences, *Mon. Weather Rev.*, 112, 1649–1668.

Halpert, M S and Ropelewski, C F (1992) Surface Temperature Patterns Associated with the Southern Oscillation, *J. Clim.*, 5, 577–593.

Horel, J D and Wallace, J M (1981) Planetary Scale Atmospheric Phenomena Associated with the Southern Oscillation, *Mon. Weather Rev.*, 109, 813–829.

Houghton, J T, Meira Fihlo, L G, Callender, B A, Harris, N, Kattenberg, A, and Maskell, K, eds (1996) *Clim. Change (1995) The Science of Climate Change*, Cambridge University Press, Cambridge, 549.

Hughen, K A, Schrag, D P, and Jacobsen, S B (1999) El Niño During the Last Interglacial Period Recorded by a Fossil Coral from Indonesia, *Geophys. Res. Lett.*, 26, 3129–3132.

McPhaden, M J (1999) Genesis and Evolution of the 1997–1998 El Niño, *Science*, 283, 950–954.

McPhaden, M J, Busalacchi, A J, Cheney, R, Donguy, J R, Gage, K S, Halpern, D, Ji, M, Julian, P, Meyers, G, Mitchum, G T, Niiler, P P, Picaut, J, Reynolds, R W, Smith, N, and Takeuchi, K (1998) The Tropical Ocean-Global Atmosphere (TOGA) Observing System: a Decade of Progress, *J. Geophys. Res.*, 103, 14 169–14 240.

Mantua, N J, Hare, S R, Zhang, Y, Wallace, J M, and Francis, R C (1997) A Pacific Inter-decadal Climate Oscillation with Impacts on Salmon Production, *Bull. Am. Meteorl. Soc.*, 78, 1069–1079.

Meehl, G A, Zwiers, F, Evans, J, Knutson, T, Mearns, L, and Whetton, P (2000) Trends in Extreme Weather and Climate Events: Issues Related to Modelling Extremes in Projections of Future Climate Change, *Bull. Am. Meteorl. Soc.*, 81, 427–442.

National Research Council (1996) *Learning to Predict Climate Variations Associated with El Niño and the Southern Oscillation. Accomplishments and Legacies of the TOGA Program*, National Academy Press, Washington, DC, 171.

Philander, S G H (1990) *El Niño, La Niña, and the Southern Oscillation*, Academic Press, San Diego, CA, 293.

Quinn, W H (1992) A Study of Southern Oscillation-related Climatic Activity for AD 622–1900 Incorporating Nile River Flood Data, in *El Niño - Historical and Paleoclimatic Aspects of the Southern Oscillation*, eds H Diaz and V Markgraf, Cambridge University Press, Cambridge, 119–149.

Quinn, W H, Neal, V T, and Antunez de Mayolo, S E (1987) El Niño Occurrences over the Past Four and a Half Centuries, *J. Geophys. Res.*, 92, 14 449–14 461.

Ramp, S R, McClean, J L, Collins, C A, Semtner, A J, and Hayes, K A S (1997) Observations and Modeling of the 1991–1992 El Niño Signal off Central California, *J. Geophys. Res.*, 102, 5553–5582.

Rayner, P J, Enting, I G, Francey, R J, and Langenfelds, R (1999) Reconstructing the Recent Carbon Cycle from Atmospheric CO_2, $\bullet\ ^{13}C$ and O_2/N_2 observations, Tellus Publications, Boston, MA, 51B, 213–232.

Rittenour, T M, Brigham-Grette, J, and Mann, M E (2000) El Niño-like Climate Teleconnections in New England During the Late Pleistocene, *Science*, 288, 1039–1042.

Ropelewski, C F and Halpert, M (1987) Global and Regional Scale Precipitation Patterns Associated with the El Niño/Southern Oscillation, *Mon. Weather Rev.*, 115, 1606–1626.

Sponberg, K (1999) *Compendium of Climatological Impacts*, National Oceanic and Atmospheric Administration, Washington, DC, 62.

Strong, A E, Kearns, E J, and Govig, K K (2000) Sea Surface Temperature Signals from Satellites – an update, *Geophys. Res. Lett.*, in press.
Supplee, C (1999) El Niño/La Niña, *Nat. Geo.*, 195(3), 72–95.
Trenberth, K E and Shea, D J (1987) On the Evolution of the Southern Oscillation, *Mon. Weather Rev.*, 115, 3078–3096.
Trenberth, K E and Hoar, T J (1996) The 1990–1995 El Niño-Southern Oscillation Event: Longest on Record, *Geophys. Res. Lett.*, 23, 57–60.
Trenberth, K E, Branstator, G W, Karoly, D, Kumar, A, Lau, N C, and Ropelewski, C (1988) Progress During TOGA in Understanding and Modeling Global Teleconnections Associated with Tropical Sea Surface Temperatures, *J. Geophys. Res.*, 10, 14 291–14 324.
Tudhope, A W, Chilcott, C P, McColloch, M T, Cook, E R, Chappell, J, Ellam, R M, Lea, D W, Lough, J M, and Shimmeild, G B (2001) Variability in the El Ninño-Southern Oscillation through a glacial-interglacial cycle, *Science*, 291, 1511–1517.
WMO (1999) *The 1997- 1998 El Niño Event: a Scienti c and Technical Retrospective*, WMO No. 905, World Meteorological Organization, Geneva, Switzerland, 93.
Wyrtki, K (1975) El Niño – The dynamic response of the equatorial Pacific Ocean to Atmospheric Forcing. *J. Phys. Oceanogr.*, 5, 572–584.

El Niño/Southern Oscillation (ENSO)

Collectively, El Niño, La Niña, and the *Southern Oscillation* are referred to as El Niño/Southern Oscillation, or ENSO for short. These phenomena are intimately related and arise from coupled interactions between the atmosphere and the ocean in the tropical Pacific. El Niño is sometimes referred to as the warm phase of ENSO and La Niña the cold phase of ENSO. The ENSO cycle is irregular, having a period of roughly 3–5 years, with individual El Niño and La Niña events lasting typically 12–18 months (*see* El Niño and La Niña: Causes and Global Consequences, Volume 1). Shifts in tropical rainfall patterns associated with ENSO result in anomalous tropospheric heating that alters the global atmospheric circulation, producing droughts, floods, unusual storminess, heat waves and other weather extremes in many regions of the globe. The social, economic, and public health consequences of ENSO-related climate variability have motivated coordinated international efforts since the mid-1980s to better understand, monitor, and predict the evolution of the ENSO cycle.

See also: El Niño, Volume 1.

MICHAEL J MCPHADEN *USA*

El Viejo

El Viejo is synonymous with the coupled ocean–atmosphere phenomenon more commonly known as La Niña. Literally, El Viejo is Spanish for the *old man* (*see* El Niño and La Niña: Causes and Global Consequences, Volume 1; La Niña, Volume 1).

MICHAEL J MCPHADEN *USA*

Electricity, Atmospheric

see Atmospheric Electricity (**Volume 1**)

Emissions, Aircraft

see Aircraft Emissions (**Volume 3**)

Emissions, Carbon

see Trends in Global Emmisions: Carbon, Sulfur, and Nitrogen (*Opening essay*, **Volume 3**)

Emissions, Nitrogen

see Trends in Global Emmisions: Carbon, Sulfur, and Nitrogen (*Opening essay*, **Volume 3**)

Emissions, Sulfur

see Trends in Global Emmisions: Carbon, Sulfur, and Nitrogen (*Opening essay*, **Volume 3**)

Emissions, Trends in Carbon, Sulfur and Nitrogen

see Trends in Global Emmisions: Carbon, Sulfur, and Nitrogen (*Opening essay*, **Volume 3**)

Energy Balance and Climate

Gerald R North
Texas A and M University, College Station, TX, USA

The global average temperature of the Earth changes little from year to year. Until the 20th century, during which the global average temperature warmed by more than 0.60 °C, paleoclimatological evidence from the last 10 000 years suggests that even over hundreds of years the planet's average temperature has not changed by more than about 0.25 °C. One reason for this stability is the powerful constraint imposed by the conservation of energy. The global and annual average temperature is determined by a balance of the rate of solar radiated energy being absorbed by the planet and the rate of infrared radiant energy being emitted from the top of the atmosphere to space. The energy balance of these streams of radiation into and out of the system in part determine the present climate and its sensitivity to system perturbations, such as changes in the solar brightness.

BALANCING THE STREAMS OF SOLAR AND INFRARED RADIATION

The Earth s energy balance is determined by considering the incoming energy from the Sun and the outgoing energy from the Earth. If the planet is in thermal equilibrium with these flows of energy (which typically occurs when the climate is stable), the planet s annual average rates of absorption and emission will be the same.

The amount of solar radiation reaching the Earth at its average distance from the Sun (averaged through the year) is approximately $1365\,\mathrm{W\,m^{-2}\,s^{-1}}$. Satellite measurements indicate that about 70% of this radiation is absorbed by the surface-atmosphere system, the rest being reflected directly back to space, giving the Earth system an albedo of about 30%. Bright areas such as deserts, white cloud tops and ice fields are highly reflective, whereas darker areas such as forests are absorbing. In addition, some solar radiation is absorbed as it passes through the atmosphere itself by such constituents as water vapor and (black) soot particles. Most of the solar radiation that is absorbed, however, is absorbed at the ground or ocean surface. When the radiation is absorbed at the surface, it warms the immediate surroundings, including the surface material (e.g., soils, ocean waters), the overlying air, etc. This local warming can set off many different responses, including raising the local temperature, causing evaporation of wet surfaces, initiating convection, and causing the emission of thermal infrared radiation.

All material bodies radiate and absorb electromagnetic energy due to collisions and other accelerations experienced in the randomized translational, vibrational and rotational motions of the molecules comprising the body. The radiation emitted from a material surface surrounding such a mass is called *thermal radiation*. An important feature of thermal radiation is its distribution of power as a function of wavelength. The amount of thermal radiation that is released depends on the nature of the material surface (or composition of the emitting gas). An *ideal black body* (sometimes called a perfect black body) is a material substance whose surface behaves in such a way that any electromagnetic radiation impinging on it is totally absorbed. The Stefan–Boltzmann Law of thermal radiation states that the rate of radiation energy per unit surface area from a black body is proportional to the fourth power of the absolute temperature (in Kelvins, or K; 0 K is −273•2 °C). As a result, warm bodies emit far more radiation than cool ones.

The power distribution or spectrum for black bodies was first explained by the German physicist Max Planck at the end of the 19th century; Planck s discovery marked the birth of the quantum theory. According to this distribution, radiation from hot bodies tends to be concentrated in shorter wavelengths and that from cooler bodies in longer wavelengths. An example of this is the radiation from a body as hot as the Sun. The effective radiating temperature of the Sun is about 5000 K, although at longer wavelengths the effective radiating temperature is closer to 6000 K. These temperatures lead to radiation across the visible part of the spectrum. Overall, the solar radiation distribution is somewhat bell shaped and peaked at a wavelength of about 0.5 μm. This kind of electromagnetic radiation or light is visible to the eye and its perceived color is white (with a slight emphasis toward yellow in that it is peaked near the center of the visible spectrum). Somewhat cooler bodies such as glowing embers and burning flames (with temperatures of 1500–2000 K) have their peaks in the red part of the visible spectrum, whereas warmer bodies such as very bright stars have their peaks in the blue part of the visible spectrum. For a much cooler body such as the Earth, which has a surface temperature of about 288 K (about 15 °C), the peak of the radiation distribution is about 10 μm, which is in the wavelength region called the *infrared*. The infrared portion of the electromagnetic spectrum is not visible to the naked eye, but can easily be detected by such common devices as infrared cameras, etc. Figure 1 shows two distributions, one from a 5000 K black surface the other from a 300 K black surface.

Kirchhoff s Law states that a black body is both a maximal absorber and maximal emitter of electromagnetic radiation for a given temperature; that is, good emitters are good absorbers and good absorbers are good emitters. However, not all bodies are perfect black bodies. A non-black body will absorb and emit only a fraction of impinging radiation. The fraction of the black body radiation that is emitted is termed the surface s *emissivity*. Because of a

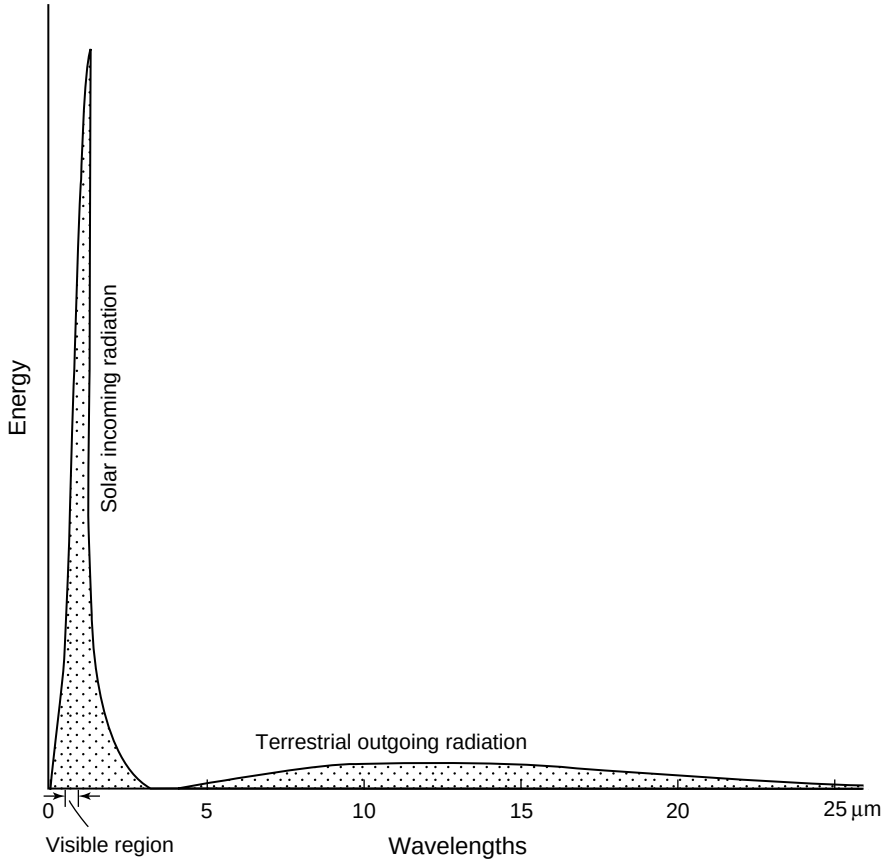


Figure 1 Two distributions, one from a 5000 K black surface the other from a 300 K black surface

material s characteristics, these relationships can be a function of the wavelength of the radiation. That is, because of the properties of a material, a body or layer of gas might have a low absorptivity in the visible part of the spectrum and a high value in the infrared or *vice versa*. For the atmosphere–surface system, there is a complex mixture of situations.

THERMAL RADIATION FROM THE EARTH

The infrared radiation originating in the surface–atmosphere system and eventually released to space has a number of sources and typically goes through a number of absorption and re-emission cycles (*see* Infrared Radiation, Volume 1). Consider first the emission from the ground. By cooling of the surface, this emission causes the transfer of radiation upwards in the atmosphere. Some of the upwelling infrared radiation is absorbed by various greenhouse gases in the atmosphere (a greenhouse gas is one which is largely non-absorbing of radiation with wavelengths in the visible, but generally strongly absorbing of radiation with wavelengths in the infrared). The most important of these greenhouse gases is water vapor, which is confined mainly to the lowest few kilometers of the atmosphere because at higher levels the air is so cold that condensation of water vapor leads to its precipitation out of the atmosphere as rain or snow. Another important absorber of infrared radiation is carbon dioxide, which is present uniformly throughout the atmosphere in concentrations of about 370 parts per million by volume. Other trace gases, including ozone, methane, nitrous oxide, and the halocarbons are also gases that absorb and emit infrared radiation. In addition, clouds are very important, both because they scatter solar radiation and because they act like black bodies in the infrared portion of the electromagnetic spectrum, both absorbing and emitting infrared radiation.

It is important at this point to note that the temperature of the atmosphere decreases with altitude in the troposphere at a rate of about $6\,\mathrm{K\,km^{-1}}$ (this rate is called the *atmospheric lapse rate*; *see* Lapse Rate, Volume 1). Because of the atmospheric absorption and reemission of infrared radiation by greenhouse gases and clouds, the emission to space from the surface–atmosphere system appears, because of its variation in wavelength, to have the appearance of being emitted by a body at a temperature that is cooler than the surface. Averaged over wavelength, this effective radiating altitude is roughly 5 km. Thus, while most of the emission

begins at the surface, some of it is absorbed and re-emitted, often several times, as it makes its way out of the system. As a result, the effective emission temperature to space is that of the air about 5 km above the surface where it is typically about 30 K cooler than at the surface. Hence, the effective emission temperature is considerably less than the surface temperature.

ESTIMATION OF PLANETARY TEMPERATURE

Based on this reasoning, a very simple model of the planetary temperature can be constructed. The rate at which solar radiation absorbed by the surface–atmosphere system is 1365 $W m^{-2} s^{-1}$ times 0.70 (this latter factor, the fraction absorbed is called the planetary co-albedo) gives a net absorption of solar radiation of 956 $W m^{-2} s^{-1}$. The cross-section of the planet as seen by the Sun is a disk, but the total energy absorbed is spread over the entire surface of the sphere. Thus, the actual amount of energy absorbed per square meter averaged over the Earth is one quarter of the absorbed amount of a surface directly facing the Sun; that is, average absorption is about 238 $W m^{-2} s^{-1}$. The Stefan–Boltzmann Law can then be used to determine the temperature of a black body that would emit this much energy per square meter per second. Insertion of this radiation rate into the Stefan–Boltzmann formula leads to a radiating temperature of about 255 K. Using the effective altitude of emission of 5 km, we can then apply the standard lapse rate of the atmosphere to obtain a surface temperature of about 285 K (about 13 °C), which is in very good agreement with the Earth s observed global average surface temperature (which is close to 15 °C). Readers should appreciate that this is not a rigorous theory of the planetary average temperature, but rather a check on the consistency of such assumptions as the 5 km effective emission altitude, as well as the other assumptions such as uniform lapse rate.

LATITUDINAL DISTRIBUTION OF THE ENERGY BUDGET

Figure 2 shows a schematic graph of absorbed incoming and emitted outgoing radiation energy ($W m^{-2}$). The information is averaged around the Earth s latitude circles (zonal averages). The curves indicate that the outgoing and incoming radiation rates do not balance latitude by latitude. The outgoing radiation rate at the poles exceeds the incoming rate of radiation absorbed while the reverse is true in the tropical latitudes. For this to be sustained over time, there must be a net transport of heat from the equatorial latitudes towards the poles by the ocean–atmosphere system.

One can proceed to make a model of the zonal average temperatures. Consider, for example, finite width strips around the planet parallel to the latitude circles in the Northern Hemisphere. The energy budget for the strip consists of four terms: (1) the absorbed solar radiation in the strip; (2) the outgoing radiation to space leaving the area in the strip; (3) the incoming stream of heat entering the strip from the south; and (4) the outgoing stream of heat leaving the strip to the north. Terms (1) and (2) depend upon the absorption and emission characteristics of the surface–atmosphere system in the strip, which involve the amount and types of cloud, snow cover, lapse rate, etc. Terms (3) and (4) depend upon heat conduction via the motions of the atmosphere and the ocean, which on the average carry warm parcels of air and water poleward and cool parcels equatorward. To actually compute these transport terms the detailed motions of the atmosphere and ocean have to be included. This can be done in a number of ways. For example, the transport terms can be estimated by constructing a mathematical system based on highly simplified assumptions such as making the average north–south heat transport by the atmosphere–ocean system proportional to the north–south temperature gradient (i.e., proportional to the difference between average temperatures in neighboring strips divided by their latitudinal width). This means the flow northwards across the southern edge of the strip is proportional to the local slope of the (zonal average) surface temperature, and likewise for the flow of heat out of the strip to the north. Such a simplifying assumption allows the entire problem to be cast in terms of the surface temperature as a function of latitude. The proportionality factor to convert temperature gradient into heat flow can, for example, be an empirically or phenomenologically determined macro-scale thermal conductivity that is derived based on fitting available observations of temperature and energy flows. Using such a model, one can then solve the model for the zonally averaged surface temperature, and for changes that might occur given a range of variations in factors affecting the energy flows. Models of this type are called *energy balance climate models* (EBCMs) and have been used extensively over the last 25 years as guides in the problem of climate change (*see* Energy Balance Climate Models, Volume 1).

However, EBCMs are not dynamical models and they cannot handle such subtleties as changes in atmospheric circulation, cloud cover and cloud type. Theoretically, these are important omissions because clouds, water vapor concentrations and other factors change as climate changes. The only way to include such changes is through invocation of additional conservation laws beyond just the conservation of energy. A minimal set of principles to include involves:

1. the conservation of momentum (Newton s Second Law for the atmosphere–ocean fluid parcels; i.e., each parcel is accelerated by the vector sum of forces acting on it);

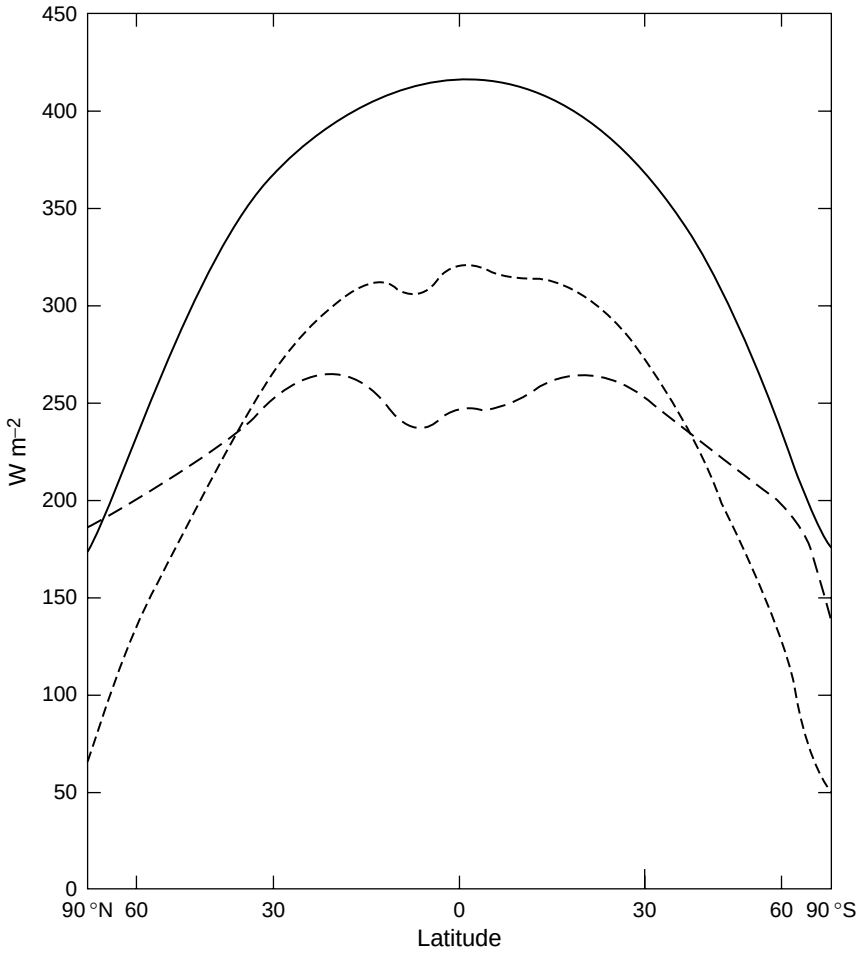


Figure 2 Zonal and annual mean radiation budget for the atmosphere. The solid curve is the net amount of radiation reaching the top of the atmosphere. The short-dashed line indicates the amount of radiation that is absorbed by the land, atmosphere, and ocean. The long-dashed line is the net longwave emission back to space. (These data are from the Earth Radiation Budget Experiment for the period February 1985 through January 1986)

2. compliance with the First and Second Laws of Thermodynamics (the sum of all forms of energy is conserved for a closed system and the entropy of a system and its surroundings always increases or stays the same in a transition);
3. the conservation of water substance in its several forms; vapor, liquid and solid surface water and cloud particles;
4. conservation of the mass of air and other materials, which acts as a constraining equation.

A model that includes these features is called a *general circulation model* (GCM). The GCM might include the motions and thermodynamics of the ocean as well as the atmosphere and ice components.

In order to implement a GCM, one needs to partition the atmosphere (and/or ocean) into small boxes fixed in space, each being a small accounting center wherein the budgets of all the relevant quantities (atmospheric or oceanic pressure, temperature, ocean salinity, wind or current speed and direction, etc.) are tracked. In modern simulations with GCMs, millions of quantities are computed for each time step in the calculation. Obviously, the simulation of a climate with given external conditions (solar constant, greenhouse gas composition, etc.) requires very fast computers with large storage capacity. The detail of such a simulation is generally limited by the availability of computing resources.

Climate simulations with GCMs require two steps. First, one must establish a control climate; i.e., a simulation of the climate under given fixed conditions. As the simulation proceeds, one finds that there is considerable natural variability, as evidenced by the quasi-random fluctuations that are common in nonlinear dynamical systems. Because of this natural variability, simulations (called control runs) to establish equilibrium conditions must typically span hundreds to thousands of years to establish baseline statistics for the control climate. The next step is to perturb the climate with a scenario of future change or, in a pure

sensitivity experiment, to change one control parameter and wait for the new equilibrium climate to establish itself. In these cases, it is often necessary to perform either a very long control run at the new parameter values or, in the case of a time dependent scenario, to perform many realizations of the experiment in order to get a statistically rigorous estimate of the rate of change. The realizations (often called an ensemble) can be averaged to obtain a probable scenario of the response of the model to the external perturbation. If the control runs are sufficiently long and/or the number of realizations are sufficiently many, the change of climate under various perturbations can be estimated.

CLIMATE SENSITIVITY

To get a sense of the sensitivity of the climate to changes in various parameters, it is useful to consider certain hypothetical standard experiments one might perform on the Earth system if one could isolate it in a laboratory (*see* Climate Sensitivity, Volume 1). One such hypothetical experiment is to consider the effect on global average temperature of an increase in the solar constant of 1%. In the case of a black body with no atmosphere, the change can be calculated from simple energy balance equations, and the result if a warming of about 0.6 K. As the actual conditions of the surface–atmosphere system are added to the calculation of sensitivity, complexities arise deriving from the subtle mechanisms operating in the atmospheric column. For example, the presence of greenhouse gases, such as naturally occurring water vapor in the column of air, lead to an increase in the estimated sensitivity. The way this increase manifests itself depends on many details of atmospheric physics. For example, because the emission to space effectively comes from an altitude of about 5 km, the behavior of the lapse rate and other mechanisms controlling the surface temperature are critical. While the change at 5 km might be about 0.6 K, the behavior at the surface might, for example, be very different. The connection between the temperature at 5 km and the surface is mediated by a number of processes, including downwelling as well as upwelling infrared radiation from one layer to another; warming and cooling at the surface; convection of sensible heat by vertical motions of buoyant air (which tends to cool the surface and warm the air above); evaporation of water from the surface and recondensation above in clouds (which also cools the surface and warms the air above); advection (horizontal displacement of air) of air from other locations that changes the vertical temperature structure of the atmosphere; and change in turn of the lapse rate. Other subtle changes such as changes of cloud cover aloft or of the condition of ice and snow at the surface can also be factors in the change of surface temperature for a 1% change of solar constant. This multitude of factors that can intervene to alter the sensitivity from that of a black body are called *feedback mechanisms* (*see* Feedbacks, Chemistry - Climate Interactions, Volume 1).

Because more water vapor will be evaporated, it is generally thought that the concentration of water vapor in the lowest few kilometers will increase if the surface temperature is increased, with the relative humidity remaining roughly constant. This is generally observed to be the case in seasonal changes of the climate and is verified by most GCM simulations that include water vapor as an interactive constituent. With more water vapor in the atmosphere, the absorption of outgoing infrared radiation is increased, leading to increased downward radiation to the surface, causing a warming. Simulations with GCMs as well as simpler models suggest that this positive (because it is amplifying) water vapor feedback mechanism, may be responsible for doubling the sensitivity of global average surface temperature, to such external influences as changes in the solar constant or doubling the concentration of the carbon dioxide concentration.

Another rather simple feedback mechanism is that of snow and ice. One part of this feedback involves the strong effect of snow and ice cover on planetary albedo. This feedback can be illustrated by considering a simple model of the planet with ice caps at the poles. If the surface temperature drops because of, for example, the presence of a dust veil overhead, the area of the ice caps would grow, leading to an increased albedo. Through this mechanism, the initial cooling of the surface leads to a more reflective planet and then further cooling. This particular positive feedback mechanism is somewhat less strong than that due to water vapor primarily because the present ice caps are so small (for colder conditions when the ice caps are more extensive, this feedback can become relatively larger). Nevertheless, it illustrates how rather simple physical processes at the global level can enhance the climate sensitivity.

There are many other feedback mechanisms, the most troublesome of which is the cloudiness (*see* Cloud - Radiation Interactions, Volume 1). Clouds cover on the average about 50% of the planet in the present climate. Clouds affect the incoming solar radiation by being highly reflective; hence, more clouds mean a cooler climate. But clouds also act as an absorber and re-emitter of infrared radiation, leading to a warming influence at the surface. The two effects nearly cancel over much of the Earth, the altitude of the cloud top often determining the local winter. All this means that if the cloudiness properties (percent coverage, height of tops, size of droplets, etc.) change in the course of a forced climate change, the feedbacks could be very subtle indeed. The challenge of determining the net effects of these competing influences is that the mechanisms which might cause these properties of cloud to change are governed by

processes that operate on scales that are finer than can be resolved by the grid structure used in GCMs. This leads to some extremely difficult problems for climate researchers, and even the sign of cloud feedback is not certain.

Based on a wide range of GCM experiments and from analogues derived from past climatic behavior, the sensitivity of the climate to a doubling of carbon dioxide (roughly equivalent to a 2% increase in the solar constant) has been estimated since the late 1970s to be in the range of 1.5–4.5 K, indicating an uncertainty of about 50% (i.e., $3 \pm 1{\bullet}5$ K). Because of the difficulties in incorporating the feedback mechanisms, refining this estimate will be the subject of continued research for many years to come.

See also: Chaos and Predictability, Volume 1.

SUMMARY

Considering the Earth s energy flows and energy balance is a very powerful way of understanding the climate. Averaged through the year, the surface–atmosphere system adjusts in ways that tend toward a balance of the energy absorbed from solar radiation with the energy emitted to space in the infrared part of the spectrum. In coming to this balance, the atmosphere–ocean system transports heat from tropical latitudes towards the poles. Some of the first models of the climate system were based on this kind of constraint, and reasonably simulated the surface temperature distribution even with only very simplified schemes for horizontal heat transport. The principle of energy balance also provides insights about the sensitivity of the climate to natural or human induced perturbations such as changes in the solar constant or the doubling of carbon dioxide concentration. The sensitivity is controlled by a multitude of subtle feedback mechanisms that can amplify or diminish the response to an external stimulus to climate change. The difficulty with using solely an energy balance approach to such analyses is in the transformation of the estimate of the change at 5 km of altitude where the radiation to space is effectively emitted to an estimate of the change at the surface where people live. Even with its limitations, however, consideration of the energy balance is a key constraint on the system that must be satisfied in all techniques used to estimate changes in the global climate.

FURTHER READING

Gilgen, H and Ohmura, A (1999) The Global Energy Balance Archive, *Bull. Am. Meteorol. Soc.*, 80, 831–850.

Kiehl, J T and Trenberth, K E (1997) Earth s Annual Global Mean Energy Budget, *Bull. Am. Meteorol. Soc.*, 78, 197–208.

Energy Balance Climate Models

L D Danny Harvey
University of Toronto, Toronto, Canada

In order to simulate climate effectively with a numerical model, it is necessary to take into account all of the processes and interactions that in uence the temperature, winds, precipitation, soil moisture, and other variables. These processes must be represented by mathematical equations, but the complexity of the equations and the number of calculations required are such that the number of calculations far exceeds what can be done without a computer. The mathematical formulation of the climate processes is therefore implemented in a computer program, which is referred to as a climate model. There are many types of climate models, varying in the number of dimensions, the number of components (i.e., atmosphere, oceans, biosphere, glaciers) that are included, the variables that are explicitly represented, and the mathematical relationships used to represent climate-related processes. Energy balance climate models (EBCMs) are one of the simpler, but very useful, types of climate models.

EBCMs WITHIN THE CLIMATE MODELING HIERARCHY

In a climate model, physical quantities that vary continuously in three dimensions are represented by a finite number of points laid out on a grid. The most complex models use a three-dimensional (3D) grid to represent the global environment (*see* General Circulation Models (GCMs), Volume 1). In other models, averaging has been performed over one, two, or all three dimensions. The important processes for climate range in size from continental-scale wind patterns, to microphysical processes in clouds at a scale of millimeters or microns, to molecular-scale scattering, absorption, and emission of radiation. The most detailed global models have a horizontal resolution of about 200 km (i.e., calculations are performed on a 200×200 km grid covering the globe). Because not all important processes can be explicitly represented on these scales, *all* climate models, require that special steps be taken to simulate subgrid-scale processes that cannot be directly represented by the model. To accomplish this, subgrid processes have to be represented by simplified, parametric relationships called *parameterizations* (*see* Parameterization, Volume 1). Parameterizations use observed correlations between the variables that are included in the model and the unresolved processes to represent the effect of the

unresolved processes on the resolved variables. The fewer the dimensions retained in a model, the fewer the number of processes that can be explicitly represented in the model, and the greater the reliance on parameterization. In order to understand how EBCMs fit into the climate modeling hierarchy and the contributions they can make, it is necessary to identify the major categories of climate processes and then indicate how these processes are represented.

The major climate processes involve (i) dynamics, (ii) radiation, and (iii) thermodynamics (*see* Earth System Processes, Volume 1). Dynamics refers to forces and motions, and is involved in the computation of atmospheric winds and ocean currents. Two kinds of radiation are relevant to climate: solar radiation (wavelengths of 0.1–4.0 μm) and infrared radiation (wavelengths of 4.0–50.0 μm). Solar radiation is subject to scattering and absorption by the atmosphere, and absorption or reflection by the surface. The Earth s surface emits infrared radiation, some of which is transmitted to space and some of which is absorbed by the atmosphere. The atmosphere also emits infrared radiation, both to space and back to the surface (*see* Infrared Radiation, Volume 1). Thermodynamics refers to heat and changes in temperature.

Climate models can be distinguished in terms of complexity and in terms of comprehensiveness (*see* Models of the Earth System, Volume 1). Complexity refers to the number of processes included in a model and the detail with which they are included. The most complex climate models are coupled, 3D atmosphere–ocean models which are referred to as atmospheric general circulation models (AGCMs) and oceanic general circulation models (OGCMs). These models are capable of representing dynamic processes, so they explicitly compute winds and ocean currents. Because the atmosphere is divided into 10–20 layers in these models, they can perform relatively detailed calculations for the vertical transfer of solar and infrared radiation. Comprehensiveness refers to the number of components of the climate system that are included in a model. The longer the time scale under consideration, the more components that are relevant (at short time scales, some components can be regarded as unchanging). The most comprehensive climate models include the atmosphere, oceans, terrestrial biosphere, seasonal snow and sea ice, glaciers, and the Earth s mantle, at least as far as its response to loading by ice sheets or sea water is concerned. An important principle in modeling the climate system is that the components that are coupled together should be of comparable complexity and resolution, because, otherwise, the errors arising from the most simplified component would likely govern the overall response of the system. It is thus viewed as relatively ineffective to couple a highly detailed representation of one component to a highly simplified representation of another component.

STRUCTURE OF EBCMs

In EBCMs, dynamic processes are not directly represented, and averaging is performed over one or usually two spatial dimensions. In the original EBCMs (Budyko, 1969; Sellers, 1969), the only dimension that was included was latitude. Each point in the model grid represented an east–west and vertical average at that point, without distinguishing between the atmosphere and surface. The Earth, in effect, was represented by a series of east–west strips.

In this type of EBCM, which is still widely used, it is assumed that the transport of heat between latitude strips can be represented by a diffusion process – i.e., the amount of heat transferred is assumed to depend on a diffusion coefficient (designed to represent both the mean transport and eddy-mixing) times the rate of variation of temperature with latitude. This is an example of a parameterization. These models often also represent the net emission of radiation to space as a linear function of the combined surface–atmosphere temperature – another parameterization. The absorption of solar radiation by the latitude strip is represented in terms of the incoming solar radiation, the surface reflectivity or *albedo*, and the transmissivity and reflectivity of the atmosphere. The surface albedo is made a function of a fixed snow-free albedo, with an adjustment for the presence of snow or ice. The latter in turn depends on temperature, which provides the basis for a third parameterization – the relationship between temperature and surface albedo. The temperature of each latitude strip is governed by the balance between absorbed solar radiation, emitted infrared radiation, and the difference between heat diffusing in or out across the two boundaries of the given strip. In mathematical terms, the governing equation is

$$R\frac{\mathrm{d}T}{\mathrm{d}t} = (\text{absorbed solar}) - (\text{emitted IR}) - \text{div}(\text{horizontal heat flux}) \quad \bullet 1 \bullet$$

where $\mathrm{d}T/\mathrm{d}t$ is the rate of change of temperature with time (t), R is the thermal inertia of the latitude strip, and the last term is the divergence of the north–south heat transport. The thermal inertia accounts for the fact that the temperature of the atmosphere–land–ocean system cannot change instantly when heat is added or removed. The divergence of the heat flux is the heat flowing out of the strip across one boundary, minus the heat inflow across the other boundary, divided by the area of the strip. If there is a net inflow of total energy, the strip warms up, while the opposite occurs if there is a net outflow.

The basic one-dimensional (1D) EBCM can be extended to include (a) land–sea resolution, and (b) surface–air resolution (e.g., Harvey, 1988). When land–sea resolution

is introduced, a simple parameterization of east–west heat exchange (e.g., based on the east–west temperature difference) must be added. When surface–air resolution is introduced in the vertical, separate energy balance equations must be solved for the atmosphere and surface, and terms representing vertical heat exchanges due to sensible and latent heat must be added. Heat transport by the atmosphere and the oceans must then be separately parameterized, usually using an approximation based on diffusion. Rather than directly parameterizing the surface albedo, the presence or absence of snow and sea ice can be explicitly computed. This can be done by, for example, (a) specifying the present day pattern of precipitation but using the model-computed air temperature to determine if the precipitation falls as snow or rain, or (b) using the model-computed surface temperature to determine if and how much of the snow should melt on a given time step. Freezing and melting rates of sea ice can be computed based on the energy balance for the ocean surface layer. Figure 1 illustrates an EBCM with surface–air and land–sea resolution and with explicit sea ice and snow processes.

Two-dimensional (2D) EBCMs have also been developed. In one case, the two dimensions are latitude and height (e.g., Peng *et al.*, 1982), and a height-dependent diffusion coefficient for north–south heat exchange is used. In the other case, the atmosphere and surface are collapsed into a single layer, but latitude and longitude are resolved (e.g., North *et al.*, 1983). Even though atmospheric winds and heat exchange are stronger in the east–west than in the north–south direction, an isotropic (same in all directions) heat diffusion coefficient has been used in these models.

In the vast majority of EBCMs, the only variable that is explicitly computed is temperature, although in some cases involving latitudinal or latitude–height EBCMs, precipitation rates have also been computed (e.g., Gallimore, 1983; Fanning and Weaver, 1996). Representing precipitation requires separately parameterizing the north–south transport of sensible heat and of water vapor. In the simplest case, the precipitation rate in a given latitude zone is given by the convergence of the north–south moisture flux plus the rate of evaporation within the zone. That is, all the water that enters the atmospheric box in a zone is assumed to either fall out as precipitation or to be transferred to the next zone, and atmospheric water vapor content is not explicitly computed.

Not all EBCMs parameterize horizontal heat and moisture transports as a diffusion process. In a few cases (e.g., Sellers, 1973), the effects of advective transports at low latitudes, associated with the Hadley cells, have been separately parameterized. In other models, transport patterns drawn from observations or from higher dimensional models are used.

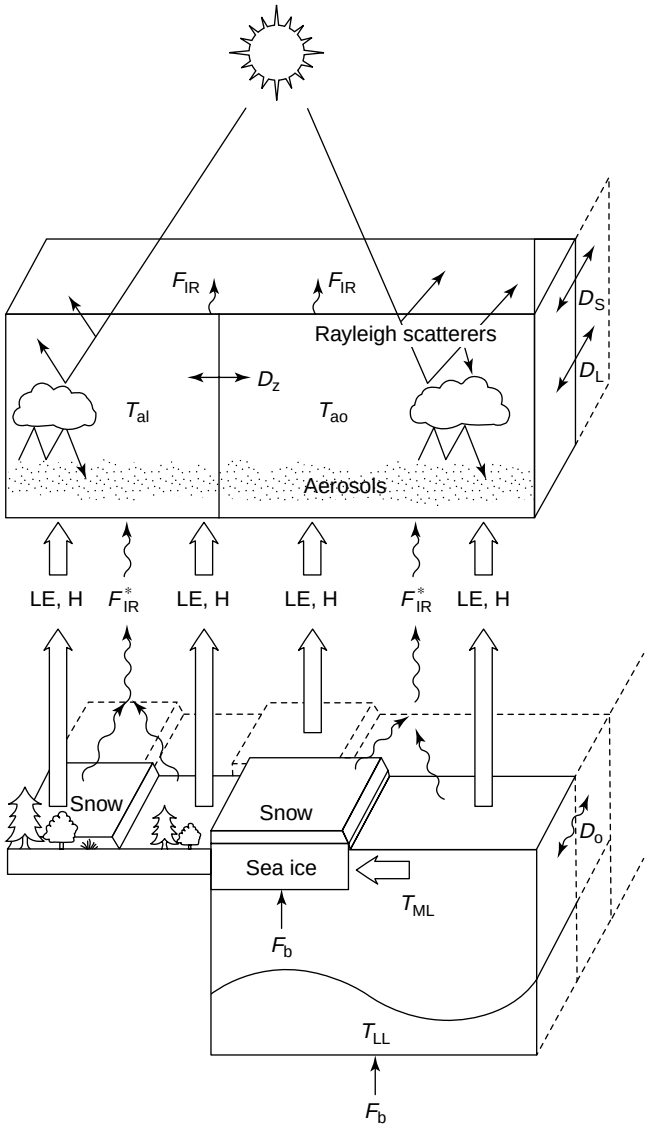


Figure 1 Illustration of a single latitude zone in the EBCM of Harvey (1988). In each latitude zone, a distinction is made between land and ocean and the surface and atmosphere. Partial snow cover on land and on sea ice in the ocean box can also occur. Vertical heat transfer occurs as sensible (H) and latent heat (LE), infrared radiation F_{IR}, and multiple scattering of solar radiation, while horizontal heat transfer is represented by diffusion (D)

The direct parameterization of the infrared emission to space as a function of temperature typically accounts for all the variables that change with temperature and that affect infrared emission. Foremost among these is the atmospheric water vapor content, which is not explicitly modeled in most EBCMs. There are several models in which the atmospheric water vapor content has been explicitly modeled (e.g., Gallimore, 1983; Fanning and Weaver, 1996); this requires a budget equation for atmospheric moisture content and internally calculated precipitation rates. Precipitation

in this case no longer exactly balances the convergence of the horizontal moisture flux plus evaporation. Rather, the small imbalance between these terms leads to changes in the atmospheric water vapor content. The explicit representation of atmospheric water vapor content is useful only if combined with a scheme for computing infrared emission to space that uses this information, rather than being based on temperature alone.

EBCMs with latitudinal, land–sea, and or surface–air resolution (as depicted in Figure 1) can be readily coupled to a 2D (latitude–depth) ocean model (e.g., Stocker *et al.*, 1992; Harvey, 1992). Such models do very well in simulating the large-scale, east–west averaged overturning circulation in the oceans (the so-called thermohaline overturning). Such a formulation allows a more realistic representation of horizontal heat transport by the oceans. In such models, there are two components to the oceanic heat transport: the transport by horizontal gyres and eddies within the surface layer (which is represented by a diffusive process); and the heat transport associated with the explicitly modeled overturning. The thermohaline overturning is driven by horizontal variations in the density of ocean water, which in turn depend on horizontal variations in both temperature and salinity (*see* Thermohaline Circulation, Volume 1). To simulate the distribution of oceanic salinity requires including the effects of precipitation, evaporation, runoff from continents, and the freezing and melting of sea ice, all of which can be computed within the EBCM structure shown in Figure 1.

Finally, EBCMs can be averaged in both horizontal dimensions, and coupled to a vertically resolved (but also globally averaged) ocean model (e.g., Hoffert *et al.*, 1980). In effect, the global atmosphere and ocean surface layer are represented by single boxes, which are underlain by the deep ocean. The two vertical mixing processes in this model are (i) the large-scale thermohaline overturning, and (ii) mean transport and eddy mixing, represented by diffusion. Such a model is referred to as an upwelling-diffusion model. In nature, deep water forms in the polar regions, sinks to the ocean bottom, spreads laterally, and upwells. This is represented in the upwelling-diffusion model by specifying a rate and temperature at which water leaves the mixed layer box, and injecting this water into the lowest model layer. This model structure is shown in Figure 2.

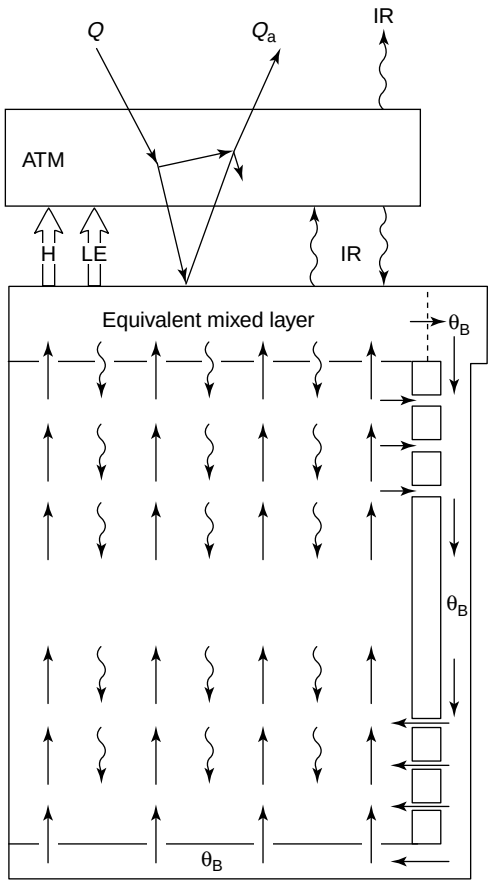


Figure 2 Illustration of a 1D upwelling diffusion model. The model consists of boxes representing the global atmosphere and ocean mixed layer, underlain by a deep ocean. Bottom water forms at temperature θ_B, sinks to the bottom of the ocean, and upwells with an upwelling velocity w. The temperature profile in the deep ocean is governed by the balance between the upwelling of cold water and downward diffusion of heat from the mixed layer. Latent (LE) and sensible (H) heat, and infrared radiation (IR) are transferred between the atmosphere and the surface. Q represents solar radiation

GENERAL APPLICATIONS OF EBCMs

Most researchers would probably agree that the primary purposes for constructing climate models are (i) to understand the factors responsible for the present climate, (ii) to understand the causes of past climatic changes, and (iii) to project future climatic changes for given scenarios due to human activities (emissions of greenhouse gases, aerosols, and their precursors; changes in the land surface). One of the most crucial factors in explaining past climatic change or projecting future climatic change is the *climate sensitivity* – the ratio of global mean temperature response to global mean radiative forcing (*see* Climate Sensitivity, Volume 1). In 3D AGCMs, the climate sensitivity is computed based on the detailed interaction of as many processes as can be incorporated. In EBCMs, only the most basic processes (infrared emission to space; reflection, scattering, and absorption of solar radiation; horizontal heat transport) are included, and these are included in a highly simplified manner. Other variables (such as cloudiness and cloud properties) are implicitly held constant as the climate changes. Although more of the relevant

physics is directly computed in AGCMs, the resulting climate sensitivity is still highly dependent on the ways in which subgrid scale processes (largely involving clouds) are parameterized. The primary advantage of EBCMs is that they can be easily used to study the interactions between those processes that are included, interactions that also occur in AGCMs but are difficult to decipher in AGCMs. Because EBCMs represent only the most critical processes, they are relatively easy to understand and inexpensive to run, so that multiple diagnostic tests can be executed.

Another primary advantage of EBCMs is that they can be readily coupled to other components of the climate system that one would wish to represent in a simplified manner (such as the deep oceans, continental-scale ice sheets, and the Earth's mantle), and the coupled system can be used for very long (up to millions of years) simulations (*see* Climate Model Simulations of the Geological Past, Volume 1). The subcomponents of such models can be constrained to replicate the overall behavior of the more complex models of each of the subcomponents in the coupled system. Because the coupled system is comprehensive but simple, a wide range of interactions can be investigated and important qualitative (and sometimes quantitative) insights can be gained.

An EBCM can also serve as the core of a climate modeling system that includes only those components and processes relevant to the build up of atmospheric carbon dioxide, climate change, and sea level rise during the next century or two. The coupled model can then be used to investigate the consequences of a wide range of different scenarios concerning emissions of greenhouse gases and aerosol precursors, along with different assumptions concerning climate sensitivity, the response of the terrestrial biosphere and oceanic circulation to changes in climate, and other system parameters. The Intergovernmental Panel on Climate Change, for example, has used EBCMs in this manner to compare the climatic response to different emissions scenarios (Cubasch and Meehl, 2001).

SPECIFIC EXAMPLES OF THE USE OF EBCMs

As examples of the types of uses of EBCMs, we provide a brief listing of some of the ways in which EBCMs, for example, has been used to improve our understanding of climate and climatic change. The 1D, latitudinally resolved EBCM, for example, has been used in the following ways, alone or in conjunction with other climate system model components:

- To determine the conditions under which the simulated climate is stable and under which more than one climate state is possible (Cahalan and North, 1979; North, 1984; Matteucci, 1993).
- To investigate the effect of different formulations of horizontal heat transport on climate sensitivity through the interaction between heat transport and high latitude albedo feedback (Held and Suarez, 1974; Lindzen and Farrel, 1977; Held *et al.*, 1981).
- To determine the potential effect on climate sensitivity of changes in surface albedo resulting from climate-induced changes in vegetation (Cess, 1978; Otterman *et al.*, 1984).
- To investigate the basic mechanisms by which changes in the seasonal and latitudinal distribution of incoming solar energy due to changes in the Earth's orbit lead to ice age-scale changes in climate (Schneider and Thompson, 1979; Birchfield and Weertman, 1983; Watts and Hayder, 1984; Lin *et al.*, 1992).
- To simulate glacial–interglacial climatic changes with coupled EBCM-ice sheet-mantle models (Pollard *et al.*, 1980; Deblonde and Peltier, 1991).
- To assess alternative hypotheses concerning the cause of the Younger Dryas cold episode (around 10 000–10 800 years ago) (Harvey, 1989).
- To assess the role of changes in ocean circulation on glacial climates (Bintanja and Oerlemans, 1996).

The 2D (latitude–longitude) EBCM has been used for:

- Elucidation of the relationship between orbital-induced changes in solar radiation and the driving force (land–ocean temperature contrast) for the monsoons (Short and Mengel, 1986).

The 1D (globally averaged) upwelling-diffusion model has been used in a variety of ways, including:

- Investigation of the role of the oceans in delaying the climatic response to increasing greenhouse gas concentrations and the role of ocean mixing-climate feedback in modifying the transient response (e.g., Hoffert *et al.*, 1980; Harvey and Schneider, 1985; Morantine and Watts, 1990).
- Exploring the importance of natural variability in observed global mean temperature variations during the past century (Wigley and Raper, 1990; Schlesinger and Ramankutty, 1994).
- Setting constraints on the magnitude of the global mean aerosol cooling effect (Wigley, 1989).
- Assessing the relative roles of greenhouse gases, aerosols, and solar variability in explaining global mean temperature variations during the past century (Kelly and Wigley, 1992; Schlesinger and Ramankutty, 1992).

Thus, EBCMs, although only one among several types of model being used (*see* Models of the Earth System,

Volume 1), has proven extremely useful for improving understanding of many types of problems.

See also: Energy Balance and Climate, Volume 1.

REFERENCES

Bintanja, R and Oerlemans, J (1996) The Effect of Reduced Ocean Overturning on the Climate of the Last Glacial Maximum, *Clim. Dyn.*, 12, 523–533.

Birchfield, G E and Weertman, J (1983) Topography, Albedo-temperature Feedback, and Climate Sensitivity, *Science*, 219, 284–285.

Budyko, M I (1969) The Effect of Solar Radiation Variations on the Climate of the Earth, *Tellus*, 21, 611–619.

Cahalan, R F and North, G G (1979) A Stability Theorem for Energy-balance Climate Models, *J. Atmos. Sci.*, 36, 1178–1188.

Cess, R D (1978) Biosphere-albedo Feedback and Climate Modeling, *J. Atmos. Sci.*, 35, 1765–1769.

Cubasch, U, Meehl, G A, Boer, G J, Stouffer, R J, Dix, M, Noda, A, Senior, C A, Raper, S, and Yap, K S (2001) Projections of Future Climate Change, in *Climate Change 2001: The Scienti c Basis*, eds J T Houghton, Y Ding, D J Griggs, M Noguer, P J van der Linden, X Dai, K Maskell, and C A Johnson, Cambridge University Press, Cambridge, 525–582.

Deblonde, G and Peltier, W R (1991) Simulations of Continental Ice Sheet Growth over the Last Glacial–Interglacial Cycle: Experiments with a One-level Seasonal Energy Balance Model Including Realistic Geography, *J. Geophys. Res.*, 96, 9189–9215.

Fanning, A F and Weaver, A J (1996) An Atmospheric Energy-moisture Balance Model: Climatology, Interpentadal Climate Change, and Coupling to an Ocean General Circulation Model, *J. Geophys. Res.*, 101, 15 111, 15 128.

Gallimore, R G (1983) A Water Vapor–Energy Balance Model Designed for Sensitivity Testing of Climatic Feedback Processes, *J. Atmos. Sci.*, 40, 39–52.

Harvey, L D D (1988) A Semi-analytic Energy Balance Climate Model with Explicit Sea Ice and Snow Physics, *J. Clim.*, 1, 1065–1085.

Harvey, L D D (1989) Modelling the Younger Dryas, *Quat. Sci. Rev.*, 8, 137–149.

Harvey, L D D (1992) A Two-dimensional Ocean Model for Long-term Climatic Simulations: Stability and Coupling to Atmospheric and Sea Ice Models, *J. Geophys. Res.*, 97, 9435–9453.

Harvey, L D D and Schneider, S H (1985) Transient Climatic Response to External Forcing on 10^0–10^4 year Time Scales, 1, Experiments with Globally Averaged Coupled Atmosphere and Ocean Energy Balance Models, *J. Geophys. Res.*, 90, 2191–2205.

Held, I M and Suarez, M J (1974) Simple Albedo Feedback Models of the Icecaps, *Tellus*, 26, 613–630.

Held, I M, Linder, D I, and Suarez, M J (1981) Albedo Feedback, the Meridional Structure of the Effective Heat Diffusivity, and Climate Sensitivity: Results From Dynamic and Diffusive Models, *J. Atmos. Sci.*, 38, 1911–1927.

Hoffert, M I, Callegari, A J, and Hseih, C-T (1980) The Role of Deep Sea Heat Storage in the Secular Response to Climatic Forcing, *J. Geophys. Res.*, 85, 6667–6679.

Kelly, P M and Wigley, T M L (1992) Solar Cycle Length, Greenhouse Forcing and Global Climate, *Nature*, 360, 328–330.

Lin, R-Q, Huang, R-X, and Apel, J R (1992) A Study of the Astronomical Theory of Ice Ages in a Two-dimensional Nonlinear Climate Model, *J. Geophys. Res.*, 97, 10 029–10 036.

Lindzen, R S and Farrell, B (1977) Some Realistic Modifications of Simple Climate Models, *J. Atmos. Sci.*, 34, 1487–1501.

Matteucci, G (1993) Multiple Equilibria in a Zonal Energy Balance Climate Model: The Thin Ice Cap Instability, *J. Geophys. Res.*, 98, 18 515–18 526.

Morantine, M and Watts, R G (1990) Upwelling Diffusion Climate Models: Analytical Solutions for Radiative and Upwelling Forcing, *J. Geophys. Res.*, 95, 7563–7571.

North, G R (1984) The Small Ice Cap Instability in Diffusive Climate Models, *J. Atmos. Sci.*, 41, 3390–3395.

North, G R, Mengel, J G, and Short, D A (1983) *J. Geophys. Res.*, 88, 6576–6586.

Otterman, J, Chou, M-D, and Arking, A (1984) Effects of Nontropical Forest Cover on Climate, *J. Clim. Appl. Meteorol.*, 23, 762–767.

Peng, L, Chou, M-D, and Arking, A (1982) Climate Studies with a Multi-layer Energy Balance Model. Part I: Model Description and Sensitivity to the Solar Constant, *J. Atmos. Sci.*, 39, 2639–2656.

Pollard, D, Ingersoll, A P, and Lockwood, J G (1980) Response of a Zonal Climate-ice Sheet Model to the Orbital Perturbations During the Quaternary Ice Ages, *Tellus*, 32, 301–319.

Schlesinger, M E and Ramankutty, N (1992) Implications for Global Warming of Intercycle Solar Irradiance Variations, *Nature*, 360, 330–333.

Schlesinger, M E and Ramankutty, N (1994) An Oscillation in the Global Climate System of Period 65–70 years, *Nature*, 367, 723–726.

Schneider, S H and Thompson, S L (1979) Ice Ages and Orbital Variations: Some Simple Theory and Modeling, *Quat. Res.*, 12, 188–203.

Sellers, W D (1969) A Global Climatic Model Based on the Energy Balance of the Earth-atmosphere System, *J. Appl. Meteorol.*, 8, 392–400.

Sellers, W D (1973) A New Global Model, *J. Appl. Meteorol.*, 12, 241–254.

Short, D A and Mengel, J G (1986) Tropical Climatic Phase Lags and Earth's Precession Cycle, *Nature*, 323, 48–50.

Stocker, T F, Wright, D G, and Mysak, L A (1992) A Zonally Averaged, Coupled Ocean-atmosphere Model for Paleoclimate Studies, *J. Clim.*, 5, 773–797.

Watts, R G and Hayder, E (1984) A Two-dimensional, Seasonal, Energy Balance Climate Model with Continents and Ice Sheets: Testing the Milankovitch Theory, *Tellus*, 36A, 120–131.

Wigley, T M L (1989) Possible Climate Change Due to SO_2-derived Cloud Condensation Nuclei, *Nature*, 339, 365–367.

Wigley, T M L and Raper, S C B (1990) Natural Variability of the Climate System and Detection of the Greenhouse Effect, *Nature*, 344, 324–327.

ENSO (El Niño/Southern Oscillation)

see El Niño/Southern Oscillation (ENSO) (**Volume 1**)

EOS (Earth Observing System)

EOS is a global observation, data acquisition and analysis program undertaken by the US National Aeronautics and Space Administration (NASA). EOS is meant to provide, for a minimum of 15 years, sustained measurements of the key physical, chemical and biological parameters that define the state of the global Earth environment. The EOS program is a pioneering effort of NASA and the centerpiece of the US Global Change Research Program, a major contribution to the international effort to understand and predict the evolution of climate and Earth environment as a single interactive system.

EOS is unique in that the program combines a major technology push to develop more sensitive, more accurate, and better calibrated instruments for observing the Earth from space, and advanced data processing systems to match, with unprecedented involvement of a broad community of Earth system scientists in the design of improved instruments, data analysis methods and algorithms. EOS is also unique in its interdisciplinary breadth, which encompasses atmospheric physics and chemistry, oceanic circulation and biological productivity, and land vegetation and hydrology.

The first series of EOS missions includes three large polar-orbiting, observatory-type satellites, each carrying several sensors for simultaneous measurement of multiple Earth system parameters, and several smaller missions dedicated to a single measurement. The observatories will be launched during the period 1999 to 2003: Terra on a morning equator-crossing orbit for ocean and land observation, Aqua on an afternoon orbit for atmospheric climate observation, and Aura for measurement of atmospheric composition and chemical reactions. The next series of EOS missions is being planned to facilitate the transfer of advances in science, instrument technology and algorithms achieved in the first series to the next generation of operational observing systems, such as the US National Polar-orbiting Environmental Satellite System, that can continue the same measurements in the long term.

See also: Earth Observing Systems, Volume 1.

PIERRE MOREL *USA*

Equilibrium Response

The term *equilibrium response* is used to indicate the steady-state response (e.g., including magnitude and pattern) that a system eventually exhibits when exposed to a sustained change (or perturbation) in forcing or boundary conditions. In studies of the global environment, systems can range from limited domains (e.g., a chemical reaction chamber and the response to adding a new compound) to quite large domains (e.g., the full Earth System and the response to changing the atmospheric composition). Perturbations can range from changes in one parameter (e.g., how a change in the carbon dioxide concentration affects plant growth) to changes in a set of conditions (e.g., how a change in climate affects the distribution of ecosystems). In all cases, the equilibrium response represents the change that would occur as a result of altered forcing or boundary conditions that are assumed to exist indefinitely (e.g., the temperature change in a cold room that ultimately would result from turning on an electric heater). The time for a new equilibrium to be established can be quite long, because the response of the system is normally much slower than the time to impose the perturbing influence.

Investigations of the equilibrium response are made using a wide variety of techniques, broadly described in Models of the Earth System, Volume 1. Small-scale systems can be studied in the laboratory by actually imposing the perturbation; large-scale systems are most often studied using numerical models, although historical and geological records can also be used to study the response to natural perturbations that occurred in the past. An important part of such studies is to investigate not only the immediate, direct response, but also the effects of various positive and negative feedbacks (*see* Climate Feedbacks, Volume 1). Such feedbacks can often take extensive time to develop and create a new equilibrium. For example, it is likely to take thousands of years for the deep ocean to warm and for ice sheets to adjust in response to the atmospheric warming that will be caused by the ongoing increase in greenhouse gas concentrations. Therefore, great care must be taken to allow time in the experiments for all feedbacks to interact and adjust.

For comparison across perturbations of various magnitudes and types, the equilibrium response is often normalized (e.g., by dividing a measure of the response by the forcing) to determine the sensitivity of a system to a unit change in the perturbing force. For example, wide ranges of approaches are used to investigate the response of the climate to various perturbing factors (e.g., changes in solar radiation, volcanic aerosol loading, atmospheric composition). To estimate the climate sensitivity (*see* Climate Sensitivity, Volume 1), these different forcings are normalized by dividing the global average temperature response

by the global average of the change in radiative forcing across the tropopause (*see* Radiative Forcing, Volume 1). The units of climate sensitivity are then $^{\circ}C\ (W\,m^{-2})^{-1}$; i.e., the temperature change for a perturbing influence of $1\,W\,m^{-2}$.

While useful for many purposes, studies of the equilibrium response are highly idealized experiments. For example, perturbations in the real world do not persist indefinitely and multiple factors are changing at the same time. Because of the many interacting processes, calculations for many types of systems would take decades to centuries or longer for the response to come to a new steady state, by which time other factors and forcings may have changed. Developing estimates of the equilibrium response is thus only a first step in understanding how and why systems from as small as a plant to as large as the global environment are changing and may change in the future. The more challenging studies involve examining the time-varying response to individual perturbations and then to the full coupled set of interacting forces and perturbations because it is the time-dependent changes that will actually be experienced (*see* Transient Response, Volume 1).

MICHAEL C MACCRACKEN *USA*

Extremes, Climate

see Climatic Extremes (**Volume 3**)

F

Feedbacks, Chemistry – Climate Interactions

Joyce E Penner
University of Michigan, Ann Arbor, MI, USA

Several important greenhouse gases and aerosols have atmospheric abundances that generally depend on atmospheric chemistry. These gases include methane (CH_4), nitrous oxide (N_2O), hydro uorocarbons, per uorocarbons, sulfur hexa uoride - as well as the chloro uorocarbons, the hydrochloro uorocarbons, and the halons. In addition, the abundance of tropospheric ozone (O_3) and stratospheric water vapor (H_2O) is also determined by chemistry. Tropospheric H_2O generally depends on climatic conditions and the hydrological cycle (including changes in cloud microphysics), and changes in tropospheric H_2O will greatly impact chemistry. A number of gases, termed indirect greenhouse gases, do not directly affect climate, but they control the abundances of direct greenhouse gases through atmospheric chemistry. These gases include the reactive gases carbon monoxide (CO), volatile organic compounds (VOCs), and nitrogen oxides ($NO_x = NO + NO_2$).

Projections of future atmospheric composition depend not only on changes in the projections of population, social structures, economics, and technology and their attendant changes in emissions, but also on the possible effects of climate change on the so-called natural emissions of gases and aerosols. These will change as a result of the biosphere s response to global change and will impact future atmospheric composition. The anticipated changes in climate (e.g., temperature, precipitation) and in chemistry will alter ecosystems and, thus, the natural, background emissions of trace gases. The biospheric response, however, is difficult to project because we have only a poor understanding of how biological communities may change as a result of climate change. For example, it is still not known whether the marine phytoplankton that produce dimethylsulfide (DMS) would increase or decrease as a result of increases in global temperatures (*see also* Dimethylsul de (DMS), Volume 1). However, several biogenic emissions depend directly on physical factors whose changes might be reasonably described if it is assumed that the biosphere itself remains unchanged. This article summarizes the factors that should be considered in determining how changes in climate might affect future atmospheric emissions and composition. A robust calculation of future climate change should, of course, include these changes in composition in order to self consistently determine the future climate.

In addition to climate-driven changes to the terrestrial and marine biosphere, there are also complex interactions that result from increases in nitrogen (N) and other nutrient deposition. Furthermore, elevated surface ozone (O_3) abundances may have an opposite influence on plant carbon dioxide (CO_2) uptake to that of nutrient deposition. Also, increased nitrogen availability from atmospheric deposition and direct fertilization accelerates the emission of nitrogen-containing trace gases (nitric oxide (NO), nitrous oxide (N_2O) and ammonia (NH_3) as well as methane (CH_4)). These complex interactions represent chemistry–biosphere feedbacks that may alter the direct warming influence of greenhouse gases.

The physical factors that determine emissions of important gases and aerosols include changes in wind speed, light levels, ocean circulation, precipitation, lightning frequency and temperature. The compositional changes associated with these changes in physical factors have not been systematically studied. This article provides only a rough outline of possible changes. Table 1 lists the long-lived, radiatively active trace species and provides an estimate of total present-day emissions and the percent of total emissions that are associated with natural processes.

WATER VAPOR (H_2O)

According to the Clausius–Clapeyron relation, as temperature increases, the atmosphere will be able to hold more water vapor. As the evaporation and precipitation cycle moves toward a new equilibrium, the total amount of water

Table 1 Total emissions of trace species important for determining climate, and the percent of emissions associated with natural emissions that may change as a result of climate change

Species	Total emissions	Percentage of total associated with natural emissions
CH_4	550 Tg CH_4 $year^{-1}$	30–45
N_2O	16.4 Tg N $year^{-1}$	60
VOCs	590 Tg C $year^{-1}$	70
NO_x	52 Tg N $year^{-1}$	20
CO	2800 Tg CO $year^{-1}$	40
Sulfate aerosol precursors	88 Tg S $year^{-1}$	30
Organic aerosols and precursors	160 Tg $year^{-1}$	50
Black carbon aerosols	13 Tg C $year^{-1}$	3
Dust	2200 Tg $year^{-1}$	50
Sea salt	3300 Tg $year^{-1}$	100

vapor present within the atmosphere will increase. Changes to cloud microphysics driven by changes in aerosols may also lead to changes in water vapor, although this process has not been systematically studied. Through its reaction with the excited oxygen atom, $O(^1D)$, H_2O is the main source of atmospheric hydroxyl radical, OH (*see* OH- Radical: is the Cleansing Capacity of the Atmosphere Changing?, Volume 2):

$$H_2O + O(^1D) \longrightarrow 2OH \quad (1)$$

Because OH is the main chemical sink for a large number of gaseous compounds, increases in atmospheric H_2O will cause a multitude of chemical changes. The atmospheric lifetime of CH_4 will decrease, as will the lifetimes of the VOCs and carbon monoxide (CO). In addition, atmospheric sulfur dioxide (SO_2) and DMS will be more rapidly converted to sulfate, while organic aerosol precursors will be more rapidly oxidized.

The most important feedbacks caused by increases in H_2O are the climate feedbacks. Because H_2O is a greenhouse gas, increased amounts in the troposphere will cause further increases to tropospheric temperature. Increases in stratospheric H_2O will cause increases in the radical HO_x pool, as well as acting to locally cool the stratosphere with attendant changes in chemistry through changes to stratospheric reaction rate coefficients.

STRATOSPHERIC WATER VAPOR (H_2O)

Water vapor in the lower stratosphere is a very effective greenhouse gas. Natural levels of stratospheric H_2O are controlled primarily by the temperature of the tropical tropopause, which is expected to vary with changes in climate. However, the oxidation of CH_4 is a second important source of mid-stratospheric H_2O and is the reason why there is an increase in stratospheric H_2O abundance from about 3 parts per million by volume (ppmv) at the tropopause to about 6 ppmv in the upper stratosphere. H_2O is also injected directly into the lower atmosphere by aircraft, and future high-flying aircraft may cause even larger increases in stratospheric H_2O than has the present subsonic fleet (*see* Aircraft Emissions, Volume 3). Increases in stratospheric H_2O have been observed, but the vertical profile and amplitude of these changes do not correspond quantitatively with those expected from the recognized anthropogenic sources (CH_4 oxidation and subsonic aircraft).

TROPOSPHERIC HYDROYL (OH)

Tropospheric OH abundances depend on abundances of nitrogen oxides (NO_x), CH_4, CO, VOC, O_3, as well as H_2O. In addition, the amount of solar ultraviolet (UV) (>300 nm) that reaches the troposphere determines the rate of photolysis of O_3 and the production rate of $O(^1D)$. Stratospheric circulation and the distribution of stratospheric O_3 control the penetration of solar UV into the lower atmosphere. As a consequence, OH varies widely with geographical location, time of day, and season. Likewise, the local loss rate of CH_4 also varies. Due to its dependence on CH_4 and other pollutants, tropospheric OH is very likely to have changed since the pre-industrial era and is expected to change again as a result of future emissions (*see* OH- Radical: is the Cleansing Capacity of the Atmosphere Changing?, Volume 2).

METHANE (CH_4)

The abundance of atmospheric CH_4 depends on the balance between emissions and its reaction with OH:

$$CH_4 + OH \longrightarrow CH_3 + H_2O \quad (2)$$

This removal reaction occurs mainly in the troposphere. Climate changes will, as noted above, tend to increase H_2O, which will tend to increase OH. However, as noted above, the concentrations of OH depend as well on a number of other complex, coupled interactions.

Substantial pre-industrial abundances of CH_4 are found in the bubbles of ancient air trapped in ice cores. CH_4 is the natural by-product of anaerobic fermentation, and its large natural emission rates have varied as the past climate has changed. Thus, in addition to the chemical interactions indicated above, a change in climate may well bring about changes in the natural sources of CH_4. Methane emissions from wetlands are sensitive to temperature, as are the rates of carbon and nitrogen cycling. Because there is an optimum temperature for production of CH_4, changing emissions may increase or decrease depending on the response of the microbes that produce CH_4 to the changing conditions. In addition, if wetland extent either increases or decreases in response to changes in the hydrological cycle, the natural sources of CH_4 from these regions will also be affected (*see* Methane, Volume 1).

NITROUS OXIDE (N_2O)

The abundance of N_2O is determined by its removal in the stratosphere through reaction with $O(^1D)$ and photolysis at short wavelengths (<240 nm):

$$N_2O + h\nu \longrightarrow N_2 + O \quad (3)$$

$$N_2O + O(^1D) \longrightarrow 2NO \quad (4)$$

$$N_2O + O(^1D) \longrightarrow N_2 + O_2 \quad (5)$$

As a result of the time required to transport N_2O to the region of chemical loss, its atmospheric lifetime is long, i.e., about 120 years. For this reason, its concentration will respond only slowly to changes in emissions (*see* Nitrous Oxide, Volume 1).

Climate change, including the radiative cooling of the stratosphere caused by increased levels of CO_2, is expected to alter stratospheric circulation and stratospheric O_3 concentrations. As a result, the global mean loss rates of the long-lived gases that are removed in the stratosphere, of which N_2O is arguably the most important for climate change, will change. In these ways, stratospheric circulation, and the distribution of O_3, control the transport of the long-lived greenhouse gases to regions of photochemical loss. At the same time, N_2O supplies ozone-depleting radicals (e.g., NO) to the stratosphere, thereby providing a feedback between the gas and its loss rate.

Soil N_2O emissions are sensitive to temperature and soil moisture as well as to the rates of carbon and nitrogen cycling. The abundance of N_2O deduced from ice cores is consistent with there being large, natural emissions, the rates of which have varied significantly as climate has changed. The atmospheric abundance of N_2O over the next 100 years will likely be driven by changes in both anthropogenic and natural sources, as well as the chemical interactions noted above. This is another example of a biogeochemical coupling with climate change.

OZONE (O_3)

Tropospheric O_3 is both generated and destroyed by photochemistry within the atmosphere (*see* Troposphere, Ozone Chemistry, Volume 1). Its *in situ* sources, such as the sequence shown below, have grown with the increasing industrial emissions of the ozone precursors: CH_4, NO_x, CO, and VOC:

$$CO + OH \longrightarrow H + CO_2 \quad (6)$$

$$H + O_2 + M \longrightarrow HO_2 + M \quad (7)$$

$$HO_2 + NO \longrightarrow OH + NO_2 \quad (8)$$

$$NO_2 + h\nu \longrightarrow NO + O \quad (9)$$

$$O + O_2 + M \longrightarrow O_3 + M \quad (10)$$

$$\text{Net: } CO + 2O_2 \longrightarrow O_3 + CO_2 \quad (11)$$

where M denotes an inert atom or molecule that absorbs excess energy but does not react.

In addition, there is substantial transport of ozone from the stratosphere to the troposphere.

The chemistry of tropospheric O_3 is closely tied to that of OH, and its abundance also varies with changing precursor emissions. The chemistry of the troposphere is also directly influenced by the stratospheric burden of O_3, climatic changes in temperature (T) and humidity (H_2O), as well as by interactions between tropospheric aerosols and trace gases. Such couplings provide a feedback between the climate change induced by increasing greenhouse gases and the concentration of these gases. Another feedback, internal to chemistry, is the impact of changes in CH_4 on OH and hence on O_3. These feedbacks are expected to be important for both tropospheric O_3 and OH.

Natural emissions of the pollutants NO_x, CO, and VOC play an important role in production of tropospheric O_3 and the abundance of OH. These emissions are also subject to changes induced by changes to the physical climate, the biosphere, and chemical systems. Terrestrial and aquatic ecosystems, in turn, respond to near-surface pollution (O_3, NO_2, acidic gases and aerosols) and to inadvertent fertilization through deposition of reactive nitrogen (often emitted from the biosphere as NO or NH_3). This response can take the form of die-back, reduced growth, or changed species composition that may alter trace-gas surface exchange and ecosystem health and function. As with other biosphere feedbacks, the magnitude of the coupling of this feedback system between build-up of greenhouse gases, human-induced climate change, ecosystem responses, and trace-gas exchange at the surface is difficult to assess.

OTHER CHEMISTRY PROCESSES AND COUPLING WITH CLIMATE

Stratospheric O_3 chemistry is determined by reactions on the surfaces of aerosols. Similar reactions have been identified in the troposphere. In particular, the conversion of NO_x to nitrates on aerosol surfaces is now routinely included in chemistry models. A wide variety of other processes may also be important including: removal of HO_x, altering the UV flux and hence photodissociation rates, and catalyzing more exotic reactions leading to the release of NO_x or halogen radicals. These processes are highly sensitive to the properties of the aerosol and the local chemical environment; their importance on a global scale is not yet established. However, if climate change leads to a change in relative humidity, aerosol surface area may change. This would lead to an overall change in the rates of such reactions over and above any change associated with direct changes in emissions of aerosols or aerosol precursors.

Stratospheric O_3 depletion has led to increases in tropospheric UV levels that, as noted above, tend to increase OH abundances. The ozone depletion itself, however, may have reduced the influx of O_3 from the stratosphere, thereby counteracting the tendency to increase OH. Since future changes in stratospheric O_3 may be driven by increases in CO_2 (which reduces stratospheric temperature) as well as possible increases in CH_4 and N_2O, adding radical compounds that may deplete ozone, such processes may play a role in future climate.

IMPACTS OF PHYSICAL CLIMATE CHANGE ON ATMOSPHERIC CHEMISTRY

The most direct changes to atmospheric chemistry that may be anticipated as a result of climate change are associated with the anticipated rise in temperature and water vapor. The first factor will change chemical reaction rates while the second will lead to enhanced photochemical destruction of tropospheric O_3. Recent studies estimate that tropospheric O_3 might decrease by about 5–10% in response to a warmer, more humid climate projected for the middle of the 21st century, as compared to an atmosphere with current temperature and H_2O.

Thunderstorms, and their associated lightning, are a component of the physical climate system that provide a direct source of a key chemical species, NO_x. The magnitude and distribution of this lightning NO_x source provide a background upon which anthropogenic perturbations are determined. The uncertainty in the source strength of lightning NO_x, together with uncertainty in the prediction of the frequency of thunderstorms, precludes an easy determination of how important this feedback might be (*see also* Lightning, Volume 1).

Convection is another climate process whose magnitude may well change as climate changes. Because convection is an important process for delivery of a variety of surface-level pollutants to the upper troposphere, a change in the effectiveness of this process could affect the chemistry of this greenhouse-gas important region. Convection is a parameterized process in climate models, and its treatment tends to be quite different from one model to the next. Hence, evaluation of the significance of any change in this process for chemistry-climate coupling awaits determination.

AEROSOLS

Aerosols have both direct sources that depend on the physical climate system as well as photochemical sources that depend on OH and gas-phase aerosol precursors (*see* Aerosols, Troposphere, Volume 1). Changes in aerosol loading and type are evident in ice core records, so consideration must be given to the possible changes in natural sources in association with future changes in climate. For example, levels of insoluble species, such as dust, recorded in Greenland ice cores were more than 10 times higher during the last glacial maximum compared to present levels. Additionally, sea salt concentrations were much higher during the last two glacial periods than now. It has been suggested that these changes resulted from either higher wind speeds during the glacial maximum or more efficient meridional (north–south) transport in the Northern Hemisphere. In Antarctica, similar increases in dust loading during the last glacial maximum have been interpreted as being associated with an expansion of arid areas and exposure of the continental shelves. Another aerosol species that has been found to have been enhanced in Antarctica was sulfate. Because a similar enhancement of methane sulfonic acid (which is associated with marine biological sources of sulfate) was also recorded, it is thought that the Antarctic increase was due to enhanced biological activity during the glacial maximum. No evidence of an increase in biological activity is evident in the ice cores from Greenland, however.

The increases in sea-salt and sulfate aerosols would have led to additional cooling during ice ages, but the increased dust, because of its ability to absorb solar and thermal radiation, could actually have led to regional warming events. This absorption might help to explain the abrupt warming events that took place as the glacial era was terminating.

In addition to possible changes in sea salt, sulfates and dust associated with changes in wind speeds, precipitation, and biological activity, future climate change may lead to changes in emissions of terpenes. Terpenes are emitted from vegetation and their source rates depend on temperature, precipitation and light levels. They undergo photochemical

oxidation (initiated by reaction with OH, O_3, and NO_3) and produce condensable compounds.

One recent study estimated an increase of about 20% in the emissions of terpenes associated with climate changes projected for 2100, while dust emissions would increase by 10%. A major feedback is associated with sea salt emissions, because the production rate of sea salt is a very strong function of wind speeds. The study estimated that these emissions might increase by 75%. Such a change would result in a significant negative feedback that would tend to moderate climate change; this factor has not yet been included in climate models.

CONCLUSION

A variety of feedbacks to atmospheric chemistry will arise as a result of climate change. Most of these have not been sufficiently studied to allow estimates of their importance at the present time. However, some feedbacks, such as the estimated change in sea salt emissions, are clearly important on the time scale of the next 100 years and should be further studied. Other feedbacks, such as possible changes to the stratosphere and troposphere that result from changes in the exchange of air between these reservoirs, or changes to the troposphere that result from changes in convective processes, will be difficult to assess until much higher resolution climate models become available. Changes in the biosphere itself will lead to changes in emissions rates that may be substantial, given current estimates of the role of the biosphere and other natural emissions in current budgets of the trace species. However, a much better understanding of the processes controlling species abundance and emissions are needed to determine the magnitude of the feedbacks that result from climate changes to the biosphere.

See also: Climate Feedbacks, Volume 1.

FURTHER READING

Atherton, C S, Penner, J E, Price, C, and Walton, J J (1995) Climate Change and its Effect on Tropospheric Ozone, in *Atmospheric Ozone as a Climate Gas General Circulation Model Simulations*, eds W C Wang and I S A Isaksen, Springer-Verlag, 65–85.

Ehhalt, D, Prather, M, Dentener, F, Derwent, R, Dlugokencky, E, Holland, E, Isaksen, I, Katima, J, Kirchhoff, V, Matson, P, Midgley, P, and Wang, M (2001) *Chapter 4: Atmospheric Chemistry and Greenhouse Gases*, Intergovernmental Panel on Climate Change, Report to IPCC from the Scientific Assessment Working Group (WGI), Cambridge University Press, Cambridge, in press.

Oltmans, S J and Hofmann, D J (1995) Increase in Lower Stratospheric Water Vapor at Midlatitude Northern Hemisphere, *Nature*, 374, 146–149.

Penner, J E, Andreae, M, Annegarn, H, Barrie, L, Feichter, J, Hegg, D, Jayaraman, A, Leaitch, R, Murphy, D, Nganga, J, and Pitari, G (2001) *Chapter 5: Aerosols, their Direct and Indirect Effects*, Intergovernmental Panel on Climate Change, Report to IPCC from the Scientific Assessment Working Group (WGI), Cambridge University Press, Cambridge, in press.

Feedbacks, Climate

see Climate Feedbacks (**Volume 1**)

Feedbacks, Cloud–Radiation

see Cloud–Radiation Interactions (**Volume 1**)

Fingerprinting

Fingerprinting is a term used in the Earth System sciences to describe an approach for distinguishing among the different factors that are likely to be contributing to changing the climate. Although problems can sometimes be introduced because of the limited length of records (especially for the period prior to human influence on the climate), relatively straightforward statistical techniques can be used to detect if any of a number of climate variables is changing (e.g., the temperature record can be analyzed to determine if the world is warming, *see* The Global Temperature Record, Volume 1). However, because the climate system varies naturally (and is thus said to be noisy) and because the potential changes to the climate are varied and the factors contributing to them may generate signals that are comparable in size, fingerprinting techniques have been developed to use in attributing particular amounts of the change to each of the many possible causal factors. This two-step process of detection and attribution is analogous to a doctor first determining that a patient has a temperature, and then using diagnostic techniques to determine the cause of the illness – both steps are essential, and each has its special set of approaches best suited to the goal.

Fingerprinting techniques, sometimes called *optimal weighting*, *optimal filtering*, or *pattern analysis*, were first conceived about 20 years ago when the initial warming due to human-induced activities was suspected (Hasselmann, 1979; MacCracken and Moses, 1983; Bell, 1986; Santer *et al.*, 1994; North *et al.*, 1995). Considering each of the possible factors that might contribute to climate change, it was realized that the particular radiative forcing (*see*

Box 1 Factors that might contribute to climate change

Agent forcing climate change	Effect on surface and tropospheric temperature	Effect on stratospheric temperature
Increase (decrease) in solar radiation	Warms (cools)	Warms (cools)
Increase (decrease) in the incidence of major volcanic eruptions	Cools (warms)	Warms (cools)
Increase in greenhouse gas concentrations	Warms	Cools the stratosphere as a whole
Increase (decrease) in sulfate aerosol loading	Cools (warms), mostly in Northern Hemisphere	Little effect
Depletion of stratospheric ozone	Slight cooling	Cooling, mostly of lower stratosphere

Radiative Forcing, Volume 1) of each potential causal factor would induce a different pattern of change (see Box 1). With each causal factor having its particular character, the challenge is to figure out which factors are contributing to the observed, time-varying patterns of change, requiring in each case that the response of each factor be proportional to its forcing at a given time (i.e., without very good reason, a small change in one factor cannot be assumed to lead to a large response while a large change in another factor leads to only a small response).

The application of the fingerprinting approach to the latitude-vertical pattern of temperature change (Santer *et al.*, 1996a,b) led the Intergovernmental Panel on Climate Change (IPCC) to its pioneering and widely quoted conclusion that the balance of evidence suggests that there is a discernible human influence on global climate (IPCC, 1996). Basically, with the stratosphere cooling strongly and the troposphere warming, particularly in the Southern Hemisphere, it became quite apparent that the temperature changes over the latter part of the 20th century were due primarily to increases in greenhouse gas concentrations (cooling the stratosphere and warming the lower atmosphere), increased sulfate aerosol loading (cooling the Northern Hemisphere more than the Southern Hemisphere) and stratospheric ozone depletion (causing the sharp cooling of the lower stratosphere). Studies since the IPCC s Second Assessment Report have carried the analysis back through the 20th century, finding that an increase in solar radiation may have contributed to some of the warming in the first half of that century, but that the warming in the second half of the 20th century is primarily due to human activities (IPCC, 2001); for a discussion of these developments (*see* Climate Change, Detection and Attribution, Volume 1).

REFERENCES

Bell, T L (1986) Theory of Optimal Weighting to Detect Climate Change, *J. Atmos. Sci.*, 43, 1694–1710.

Hasselmann, K (1979) On the Signal-to-Noise Problem in Atmospheric Response Studies, in *Meteorology of Tropical Oceans*, ed T Shawn, Royal Meteorological Society, London, 251–259.

IPCC (Intergovernmental Panel on Climate Change) (2001) *Climate Change 2000: The Scienti c Basis*, Working Group I Report, Cambridge University Press, Cambridge, 1–881.

MacCracken, M C and Moses, H (1983) The First Detection of Carbon Dioxide Effects: Workshop Summary, *Bull. Am. Meteorol. Soc.*, 63, 1164–1168.

North, G R, Kim, K Y, Shen, S S P, and Hardin, J W (1995) Detection of Forced Climate Signals: 1. Filter Theory, *J. Clim.*, 8, 401–408.

Santer, B D, Bruggemann, W, Cubasch, U, Hasselmann, K, Hock, H, Maier-Reimer, E, and Mikolajewicz, U (1994) Signal-to-noise Analysis of Time-dependent Greenhouse Warming Experiments. Part 1: Pattern Analysis, *Clim. Dyn.*, 9, 267–285.

Santer, B D, Wigley, T M L, Barnett, T P, and Anyamba, E (1996a) Detection of Climate Change and Attribution of Causes, in *Climate Change 1995, The Science of Climate Change*, Chapter 8, Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, 441–443.

Santer, B D, Taylor, K E, Wigley, T M L, Johns, T C, Jones, P D, Karoly, D J, Mitchell, J F B, Oort, A H, Penner, J E, Ramaswamy, V, Schwarzkopf, M D, Stouffer, R J, and Tett, S (1996b) A Search for Human Influences on the Thermal Structure of the Atmosphere, *Nature*, 382, 39–46.

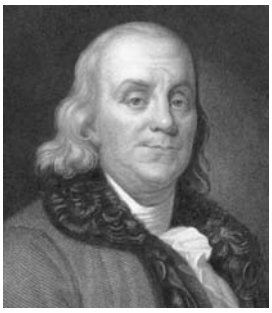
MICHAEL C MACCRACKEN *USA*

Forecast

see Prediction in the Earth Sciences (**Volume 1**)

Framework Convention on Climate Change

see United Nations Framework Convention on Climate Change and Kyoto Protocol (**Volume 4**)



Franklin, Benjamin

(1706- 1790)

Benjamin Franklin was a colonial and US printer, publisher, prolific author, scientist, inventor, philanthropist, and statesman. He brought attention to a central concern of modern environmentalists with a major essay, *Observations Concerning the Increase of Mankind*, published in 1755. Noting that the population of British descent in North America was then one million, but that only 80 000 had immigrated, Franklin calculated that the population would double every 20–25 years. It did for the next century, until immigration surged. In this essay, Franklin also explicitly anticipated the limitations stressed by Thomas Malthus in his famous tract of 1798.

Franklin did much else pertinent to environmental science. He gave a classic explanation of a seemingly paradoxical observation: why storms coming from the northeast actually move in the opposite direction (e.g., hitting Philadelphia hours before Boston). He noted that on lighting a fire in a chimney, the heated air rises, so nearby cooler air flows in to replace it, and other air to replace that. The air near the distant door is last to move, even though the flow is toward the chimney. Accordingly, he suggested that air heated over Florida and the Gulf of Mexico is replaced by cooler air flowing in from the north, generating storms that begin in the south and move to the northeast, directed by the coastline and the Appalachian mountain range.

Other meteorological imaginations and conjectures offered by Franklin include his analysis of the formation of hailstones and his proposal that major climatic changes may result from the vast quantity of smoke, long continuing to issue from volcanic eruptions. He also devoted much effort to measuring ocean temperature profiles from which he mapped the Gulf Stream.

In a cautionary letter written when he was 80 years old, Franklin compiled observations and episodes he had noted over six decades which convinced him of the danger of lead poisoning. For instance, he attributed a sudden attack of dry bellyach that had beset a family to drinking rainwater collected from a leaded roof. He noted that trees planted around the house years before had grown tall enough to shed leaves on the roof, thereby creating acid that corroded the lead and furnished the water ... with baneful particles . He lamented how long a useful truth may be known before it is generally received and practiced on.

Born in Boston, Franklin was the 15th child among 17 (perhaps presaging his concern about population). He attended school only two years, but was an avid, life-long student of literature, philosophy and science. Resentful of ill treatment as a printer s apprentice, at 17 he ran off to Philadelphia. His industry and skill enabled him to acquire a newspaper, *The Pennsylvania Gazette*, which became by the 1730s the most widely read paper in the colonies. His *Poor Richard's Almanack* (1732–1759) also was extremely popular and fostered his reputation for wit and wisdom. He promoted many civic projects, including paving and lighting streets, a circulating library, a fire company, a police force, a hospital, and an academy that later became the University of Pennsylvania. In 1748 Franklin retired from business and thereafter devoted himself chiefly to scientific research and public service. As a diplomat, he lived for 25 years in England and France.

Franklin was greatly esteemed as a scientist as well as a sage. In the early 18th century, electricity was a greater mystery than gravity had been a century earlier. Contemporary scientists hailed Franklin as the Newton of electricity. His book, *Experiments and Observations on Electricity, made at Philadelphia in America*, consisting of letters he had sent to a colleague in England, was a sensation in Europe: it went through some 10 editions (1751–1774), including translations into French and German. Like all of his writings, the book was accessible to a wide public. Franklin s invention of the lightning rod in 1750 had great impact, cultural as well as practical. Before him, lightning was considered a supernatural phenomenon. It was customary to ring church bells, to break the thunder, and store munitions in churches. The presumption of divine protection had long persisted, although many bell ringers were killed and churches damaged by lightning. In 1767 a stroke detonated tons of powder in a Venice church, killing 3000 people. Despite ample demonstrations of the efficacy of his lightning rods, Franklin had to weather thunderous attacks from clergy for his audacity in intruding on a prerogative of the Almighty. He remained unruffled, writing in 1753: Surely the thunder of heaven is no more supernatural than the rain, hail or sunshine of heaven, against the inconveniences of which we guard by roofs & shades without scruple.

Photo: Copyright, The Early American Digital Library.

See also: Lightning and Atmospheric Electricity, Volume 1; Volcanic Eruptions, Volume 1.

DUDLEY HERSCHBACH *USA*

Fronts

Fronts are the boundaries between differing air masses. Typically, the term is used for the long (order of hundreds of kilometers) and narrow (order of tens of kilometers) zone of transition between large air masses of different air densities; that is, warmer and/or moister in one air mass and cooler and/or drier in the other. Fronts are further identified by their motion: cold fronts move such that colder (denser) air replaces warmer (less dense) air; warm fronts move such that warmer air overtakes cooler air; and stationary fronts, as the name implies, show little movement. The frontal zone separating the air masses typically slopes with height over the cold (dense) air with slopes on the order of 1 km in the vertical for every 100 km in the horizontal. The circulation associated with fronts is such that, regardless of the frontal movement, the air on the warmer, moister side of the front tends to ascend the sloping frontal surface, leading to the formation of one or more bands of clouds and precipitation parallel to the front. Cold fronts, as they undercut warm and moist air ahead of them, may trigger convection leading to narrower, more intense bands of precipitation. Fronts are typically associated with mid-latitude cyclones and are often easily identifiable on satellite imagery. The term *front* is also used for the leading edge of smaller air masses, such as the gust front at the leading edge of the cool thunderstorm outflow or the sea-breeze front at the leading edge of the cool sea breeze.

See also: Atmospheric Motions, Volume 1.

KEITH L SEITTER *USA*

G

GAIM (Global Analysis, Interpretation and Modeling) Program

The Global Analysis, Interpretation, and Modeling Program (GAIM) is one of the framework activities of the International Geosphere Biosphere Programme (IGBP). Its goal is to advance the study of the coupled dynamics of the Earth System using both data and models as tools. GAIM emphasizes activities designed to expand upon the development, testing, and analysis of integrative data sets and models of those aspects of the Earth System where IGBP has the scientific lead, and it commits to collaborate on aspects of the Earth System where the World Climate Research Programme (WCRP) has the lead.

In order to achieve its goal, the GAIM Task Force has established specific objectives dealing with development, evaluation, and application of comprehensive prognostic models of the global biogeochemical subsystem; experiments with existing models; clarifying key scientific issues facing the development of global biogeochemical models and the coupling of these models to general circulation models; contributing to international assessment exercises, particularly the Intergovernmental Panel on Climate Change (IPCC) process; advising the IGBP Scientific Committee, and maintaining scientific liaison with the WCRP Steering Group on Global Climate Modeling.

GAIM coordinates a diverse set of research projects dealing with modeling of the carbon cycle, trace gas processes, and climate–vegetation interactions on a number of time scales and periods ranging from beyond 20 000 years ago through the near- and long-term future. Numerous publications in international journals have resulted from GAIM coordinated research, and a number of international workshops have been convened. For further information, contact GAIM Task Force Office, Institute for the Study of Earth, Oceans, and Space, Morse Hall, University of New Hampshire, Durham, NH 03824-3525, Tel.: +1-603-862-0322/Fax: +1-603-862-1915.

JOHN S PERRY *USA*

GARP (Global Atmospheric Research Program)

GARP was a massive international research effort to advance understanding of weather and climate, with a view to improving weather forecasting and projection of both natural and human-induced climate changes.

The program was motivated by the development of meteorological satellites in the early 1960s, the development of computer-based numerical weather prediction, and research indicating possibilities for markedly improved forecasts. In 1961, US President John F Kennedy called upon the United Nations (UN) to mount international collaborative efforts in weather prediction. The UN requested the World Meteorological Organization (WMO) and the International Council for Science (ICSU) to plan and organize such a program, which was developed through a series of international scientific conferences beginning in 1967.

The focus of GARP was on improving understanding of the processes determining short-term changes in the atmosphere, which we call the weather; and the factors determining its long-term statistical characteristics, which we call the climate. The central activity was a global observing experiment, designed and implemented over a decade of international work. By augmenting operational observing systems, enhancing and coordinating satellites, and launching special *in situ* observing systems, the atmosphere and oceans of the planet would be observed in unprecedented detail for a full year. The resulting data would permit development of greatly improved models of weather and climate. In parallel with the global experiment, a number of regional observing and research programs were designed.

The Global Weather Experiment, as it came to be called, was very successfully carried out in 1979 with contributions by virtually every nation on Earth. The unique high-quality research data sets obtained have been employed ever since in the development of ever-improving numerical forecast models, and have led directly to the extended-range predictions that are provided today by the world s weather services. These same improvements in numerical models

have advanced our understanding of natural climate variability and human influences on climate.

JOHN S PERRY *USA*

GAW (Global Atmosphere Watch)

Many urgent environmental problems are connected with human-induced changes in the composition of the atmosphere. Long-term data on the chemical composition and physical parameters of the atmosphere are therefore needed. The GAW of the World Meteorological Organization (WMO) was established in 1989 to coordinate international collection of such data.

The GAW provides information on the chemical composition and related physical characteristics of the atmosphere and their trends. The data are particularly essential to understanding the relationship between changing atmospheric composition and changes of global and regional climate, long-range atmospheric transport and deposition of potentially harmful substances, natural cycling of chemical elements in the global atmosphere–ocean–biosphere system, and anthropogenic impacts on these systems.

The GAW monitoring network, consisting of global and regional monitoring stations, relies largely on activities of WMO member countries. Central facilities handle standardization, calibration, data processing, central archiving, publication, training, and analysis. Global stations carry out a complete range of measurements including ozone, other greenhouse gases (carbon dioxide, chlorofluorocarbons, methane, nitrous oxide, water vapor), solar radiation, ultraviolet (UV)-B, atmospheric turbidity, total aerosol load, reactive gas species (SO_2, NO_x, carbon monoxide), chemical composition of rain and atmospheric particles, radionuclides and meteorological parameters. Regional stations assess regional environmental problems such as acid rain, increased surface ozone, deterioration of ecosystems, air pollution in rural areas, airborne pollution of the seas, etc.

Six World Data centers have been established: ozone and UV-B (Toronto, Canada), greenhouse gases (Tokyo, Japan), precipitation chemistry (Albany, USA), solar radiation (St Petersburg, Russia) surface ozone (Oslo, Norway) and aerosols (Ispra, Italy). Quality Assurance Science Activity Centres to oversee the data quality have been established in Germany, USA and Japan.

Results from the GAW serve as the scientific basis for a number of international conventions and declarations aimed at protecting the atmosphere and the environment as a whole.

See also: Air Quality, Global, Volume 1; Stratosphere, Ozone Trends, Volume 1.

JOHN S PERRY *USA*

GCDIS (Global Change Data and Information System)

The Global Change Data and Information System (GCDIS) was originally conceived in the late 1980s as a mechanism for making available data and information produced by the US Global Change Research Program. This concept came from a recognition, by key senior managers in the agencies conducting global change research, that data and information had considerable value beyond their initial use, and that a coordinated effort across the federal agencies would be required. As GCDIS evolved, it was recognized that the interconnected nature of global change research required more than the observations made by the program, extending to virtually all of the environmental data and information and much of the socioeconomic data held by the US Federal Government. It also needed to extend to include process science, in addition to measurements and model outputs. Moreover, not only scientists, but legislators, educators, and the general public have a need in gaining access to these data and information. Thus, the GCDIS became the largest effort in the world to provide environmental data and information to anyone who needs it.

GCDIS, as it currently is structured, uses a World Wide Web based gateway (http://www.gcdis.usgcrp.gov) to direct users to appropriate federal data centers where data and information reside. Data sets are described in searchable catalogs, such as the National Aeronautics and Space Administration (NASA) Global Change Master Directory, which refer users to data and information in over 70 participating data centers (and centers of data) throughout the US Federal Government. These catalogs are maintained by the individual agencies responsible for their data holdings and are supported by linked web sites in a distributed system environment. Users who need additional help may submit electronic questions to Dr Global Change and receive answers directing them to data and information available on the web relating to their global change question. Additional user services are provided by the Global Change Research Information Office and by each participating agency data center. Participating centers adhere to the principle of full and open non-discriminatory data sharing, except for those

data which are restricted by copyright, trade secret, or by law. GCDIS also reaches out to the international community through a series of linked catalogs to include holdings worldwide.

TOM MACE *USA*

GCMs (General Circulation Models)

see General Circulation Models (GCMs) (**Volume 1**)

GCOS (Global Climate Observing System)

GCOS was established in 1992 to ensure that the observations and information needed to address climate-related issues were obtained and made available to all potential users. It is co-sponsored by the World Meteorological Organization (WMO), the Intergovernmental Oceanographic Commission of the United Nations Educational, Scientific and Cultural Organization (UNESCO), the United Nations Environment Programme (UNEP) and the International Council for Science (ICSU).

GCOS is intended to be a long-term, user-driven operational system capable of providing the comprehensive observations required for monitoring the climate system, for detecting and attributing climate change, for assessing the impacts of climate variability and change, and for supporting research toward improved understanding, modeling and prediction of the climate system. It addresses the total climate system including physical, chemical and biological properties, and atmospheric, oceanic, hydrologic, cryospheric and terrestrial processes.

GCOS does not itself directly make observations nor generate data products. It stimulates, encourages, coordinates, and otherwise facilitates the taking of the needed observations by national or international organizations in support of their own requirements as well as common goals. It provides an operational framework for integrating, and enhancing, as needed, observational systems of participating countries and organizations into a comprehensive system focussed on the requirements for climate issues. GCOS builds upon, and works in partnership with other existing and developing observing systems such as the Global Ocean Observing System, the Global Terrestrial Observing System, and the Global Observing System and Global Atmosphere Watch of the WMO.

GCOS is directed by a Steering Committee that provides guidance, coordination and oversight to the program. Four panels, reporting to the Steering Committee, have been established to prepare specific program elements (atmosphere, ocean, land and space-based observations) and to make recommendations for implementation.

HANS TEUNISSEN *Switzerland*

General Circulation Models (GCMs)

Michael C MacCracken
Lawrence Livermore National Laboratory, Livermore, CA, USA

General Circulation Models (GCMs) are numerical models developed to simulate the 3-dimensional, time dependent dynamics and thermodynamics of the global atmosphere or ocean. Atmospheric GCMs normally treat, for example, the movement of the air horizontally and vertically, temperature and relative humidity, surface pressure and surface- atmosphere interactions, precipitation and storms, and much more. In simulating the geographical distribution of the climate, GCMs provide quantitative frameworks for assembling objective, quantitative representations of the full range of physical, chemical, and biological processes that determine the behavior of the atmosphere and oceans.

The first attempt to construct a comprehensive numerical model of the atmosphere was made by Lewis Fry Richardson (*see* Richardson, Lewis Fry, Volume 1) during the latter part of World War I. He constructed a series of boxes covering much of Europe and the adjacent Atlantic Ocean that represented the state of the atmosphere over this region. Then, starting from a baseline set of observations, he used the equations of motion, which had been developed by Norwegian scientist Vilhelm Bjerknes to explain the movement of weather systems, to calculate the evolution of the weather over the next eight hours. Although his hand calculations took over 100 times longer to complete than for the weather to actually evolve, and although the calculations gave results different than for the real atmosphere,

Richardson was not discouraged. The conclusion of his book (Richardson, 1922) envisioned the creation of a forecast factory staffed by a very large number of teams distributed around the various balconies of an opera house. He proposed that each team would receive information about observations, make calculations with a fast mechanical calculator, and flash their results to neighboring teams needing such information and to a central conductor who would keep the teams synchronous in time and would assemble the results for distribution to the public. This is much the way, in practice, that weather forecasts are made today, except they are done using modern parallel processing computers rather than human calculators.

Studies following Richardson s pioneering effort found a number of shortcomings that had to be overcome to actually construct a functioning model of the general circulation of first the atmosphere, then the oceans, and finally the coupled atmosphere–ocean system. A global grid (actually, at first only hemispheric) had to be used because what happens at remote locations affects the positions and movement of the major circulation features (e.g., the El Niño in the Pacific Ocean has effects far from where it warms ocean temperatures). A more comprehensive and timely set of observations was needed, emphasizing the need for a globally connected network of observation stations spread across the world. The set of observations had to be carefully processed to overcome the problems of unrepresentative data and the need for the data to be dynamically and thermodynamically self-consistent when making the calculations. More advanced numerical techniques were needed to solve the sets of equations that represented the dynamics of the atmospheric motion. These techniques had to be designed to filter out sound waves that traveled so fast that model time steps would have had to be very short. The techniques also had to be designed to conserve key quantities such as atmospheric mass and energy. The set of equations had to include treatment of atmospheric thermodynamics (which controls the air temperature) and of atmospheric water vapor (which determines where clouds form and how latent heat is added to or removed from the system). For longer calculations, changes in surface conditions had to be included in that these can cause changes in where energy is added to and removed from the atmosphere. Finally, faster computers were needed if the rate of calculations for the numerical atmosphere was going to keep up with and then exceed the evolution of the real atmosphere (a necessity if the calculations were going to produce useful predictions).

Following the development of electronic computers during World War II, the Princeton mathematician John von Neumann assembled a team to attempt to actually simulate the evolution of the atmosphere for a few days. With the impressive, but still quite limited, computers available at the time, von Neumann s team constructed a model using a somewhat simplified set of equations that treated mainly atmospheric dynamics. They completed their first trial forecast in 1950, treating a region covering North America using a time step of 3 h and a grid of 270 points about 700 km apart (Charney *et al.*, 1950). While achieving only limited success, this first effort led, as computers advanced, to several additional efforts around the world to start making real time forecasts, led in Sweden by Carl Gustav Rossby and in the US by Jule Charney (*see* Charney, Jule Gregory, Volume 1; Rossby, Carl-Gustaf, Volume 1).

As the speed of computers increased, the next step was to turn weather modules into climate models. The first to do this was Norman Phillips, who constructed a two-layer quasi-geostrophic atmospheric model (i.e., it was based on conservation of vorticity) and carried out a 1 month simulation at the Institute for Advanced Studies in Princeton (Phillips, 1956). This initial effort to simulate the atmosphere s general circulation was extended to a more complete representation of the equations of motion of the atmosphere in the early 1960s. Joseph Smagorinsky led a team within the General Circulation Research Section of the US Weather Bureau in Washington, DC, to construct a hemispheric GCM (Smagorinsky, 1963). Beginning shortly thereafter, Cecil Leith led a team at what is now the Lawrence Livermore National Laboratory to construct a full global GCM (Leith, 1965).

These early models represented the vertical structure of the atmosphere in several layers and had a horizontal resolution of about 5–10 degrees in latitude and longitude (roughly 500–1000 km at the equator, with finer resolution near the poles). The interchange of mass, momentum, and energy among grid cells was governed by use of a set of fundamental conservation equations (*see description in* Projection of Future Changes in Climate, Volume 1) and representations of particular processes (*see* Earth System Processes, Volume 1). Descriptions of the structure and characteristics of GCMs can be found in the References and Further Reading sections.

Because of limits in computer resources and data availability, the initial calculations with this type of model typically would simulate a perpetual January condition by assuming fixed sea surface and land temperatures and then letting the model simulation run until the influence of initial atmospheric conditions disappeared. As the global observation network grew and satellites started to provide data over the oceans, these atmospheric models developed further over the next decade and were soon serving as the basis for numerical weather prediction at several centers around the world.

As a complement to the atmospheric models, Kirk Bryan and Syukuro Manabe, working with Smagorinsky, led the effort to develop the first global ocean model. A key limitation in developing and testing ocean models was the lack of knowledge about the oceans, there being reasonable

measurements for only the upper ocean. Whereas the atmospheric models could readily be tested against observations, the ocean model simulations generated currents and ocean flows that had never been seen. Such observations helped to prompt field programs that would go to search for evidence of deep ocean behavior (e.g., the GEOSECs program begun in the 1970s started a systematic effort to sample the world ocean from the surface to the bottom).

The first simulations of extended periods of atmospheric behavior and of the seasonal evolution of the global climate took place in the 1970s, carried out initially by Syukuro Manabe (*see* Manabe, Syukuro, Volume 1) and colleagues working at the Geophysical Fluid Dynamics Laboratory under Joseph Smagorinsky, which had been relocated to Princeton, NJ. The initial simulations typically had a schematic distribution of the continents, cloud fractions prescribed at observed values, and ocean temperatures specified based on observations. Improvements came rapidly, however, and soon the continents were properly shaped, clouds varied based on relative humidity and atmospheric stability, and oceans were represented as a well mixed layer of water about 100 meters deep.

By the mid 1970s, even though the treatment of model physics remained relatively limited, the model simulations were deemed sufficiently interactive and representative to start using the models as experimental tools. Using the numerical models in the same way using an experiment a physical model could be conducted by running the numerical models with multiple sets of conditions. For example, a perturbation study could be run by using one set of conditions (normally representing the present) as a *control case* (or control simulation). Then rerunning the simulation after arbitrarily changing one or more parameters (such as atmospheric composition), and using this result as a *perturbed case* (or perturbed simulation). One of the first such experiments was to calculate the effects on the climate of a hypothetical doubling of the atmospheric carbon dioxide concentration (Manabe and Wetherald, 1975). These early simulations suggested that a warming of about 2.5–3 °C would result, which was not much different than the initial estimates from mathematical analysis made by Arrhenius in 1897 (*see* Arrhenius, Svante, Volume 1). Additional studies using a model developed at the Goddard Institute of Space Studies by James Hansen and colleagues provided the basis for an initial estimate of the climate sensitivity by the US National Research Council (1979) (*see* Climate Sensitivity, Volume 1).

During the 1980s and 1990s, the development of GCMs was relatively rapid, with models being designed to include more detailed representations of the various physical processes, with more accurate and numerically efficient numerical techniques introduced, with finer spatial resolution in the vertical and horizontal, and with fully interactive coupling of the atmospheric and oceanic components of the climate system. In addition, with the widespread availability of supercomputers, GCMs have been developed and are being used at many centers around the world for a wide range of simulations and investigations. Gates *et al.* (1999) and Lambert and Boer (2001) report on efforts to evaluate model performance of atmospheric and coupled atmosphere–ocean models (*see* Model Simulations of Present and Historical Climates, Volume 1). The Intergovernmental Panel on Climate Change (IPCC, 2001) reports on the use of models to study the potential impacts of the increasing concentration of greenhouse gases on the atmosphere (*see* Projection of Future Changes in Climate, Volume 1).

REFERENCES

Charney, J G, Fj rtoft, R, and von Neumann, J (1950) Integration of the Barotropic Vorticity Equation, *Tellus*, 2, 237–254.

Gates, W L, Boyle, J S, Covey, C, Dease, C G, Doutriax, C M, Drach, R S, Fiorino, M, Gleckler, P J, Hnilo, J J, Marlais, S M, Phillips, T J, Potter, G L, Santer, B D, Sperber, K R, Taylor, E, and Williams, D N (1999) An Overview of the Results of the Atmospheric Model Intercomparison Project (AMIP), *Bull. Am. Meteorol. Soc.*, 80, 29–56.

IPCC (Intergovernmental Panel on Climate Change) (2001) *Climate Change 2000: The Scienti c Basis, Working Group I Report*, Cambridge University Press, Cambridge, 1–881.

Lambert, S J and Boer, G J (2001) CMIP1 Evaluation and Intercomparison of Coupled Climate Models, *Clim. Dyn.*, 17, 83–106.

Leith, C E (1965) Numerical Simulation of the Earth s Atmosphere, *Methods in Computational Physics, Applications in Hydrodynamics*, Academic Press, New York, 1–28, Vol. 4.

Manabe, S and Wetherald, R T (1975) The Effects of Doubling the CO_2 Concentration on the Climate of a General Circulation Model, *J. Atmos. Sci.*, 32, 3–15.

Phillips, N A (1956) The General Circulation of the Atmosphere: A Numerical Experiment, *Q. J. R. Meteorol. Soc.*, 82, 123–164.

Richardson, L F (1922) *Weather Prediction by Numerical Process*, Cambridge University Press, Cambridge, (reprinted by Dover Publications, NY in 1965).

Smagormsky, J (1963) General Circulation Experiment with the Primitive Equations. 1. The Basic Experiment, *Mon. Weather Rev.*, 91, 98–164.

FURTHER READING

Mote, P and O Neill, A (2000) *Numerical Modeling of the Global Atmosphere in the Climate System, NATO Advanced Science Institute*, Kluwer, Dordrecht.

National Research Council (1979) *Carbon Dioxide and Climate: a Scienti c Assessment*, National Academy of Sciences, Washington, DC.

Nihoul, J (1985) *Coupled Ocean- Atmosphere Models*, Elsevier, Amsterdam.

Ojima, D (1992) *Modeling the Earth System*, University Corporation for Atmospheric Research/Office for Interdisciplinary Earth Studies, Boulder, CO.

Schlesinger, M (1988) *Physically Based Modeling and Simulation of Climate and Climate Change*, NATO Advanced Science Institute, Kluwer, Dordrecht.

Schlesinger, M (1991) *Greenhouse Gas Induced Climatic Change: A Critical Appraisal of Simulations and Observations*, Elsevier, Amsterdam.

Washington, W and Parkinson, C (1986) *An Introduction to 3-Dimensional Climate Modeling*, University Science Books, Oxford University Press, Oxford.

Geoengineering

see Geoengineering: a Way to Stabilize Global Climate? (**Volume 3**)

Geological Cycling

Robert Raiswell
Leeds University, Leeds, UK

Understanding the role of time lies at the heart of many geological controversies. In the Bible, the Book of Genesis asserts that the Earth was formed in stages over six days. This view has an identi able starting point, from which time is directly coupled to progress (and by implication global change). There is, however, a less well-known alternative in the Book of Ecclesiastes where Koheleth the Sage states that all streams run into the sea, yet the sea never over ows: back to the place from which the streams ran they return to run again. In this view, the world is maintained in an unchanging state by the operation of processes that are counter-balanced to produce no overall change. Here, time and progress are decoupled. There is no identi able starting point and no overall progress, only a cyclical passage through continuously repeating states. This latter view is the basis of geological cycling.

The debate between a linear (catastrophic) and a cyclic (uniformitarian) origin for the Earth occupied much of the 18th and 19th centuries (Gregor, 1988), and was particularly sharply focused in the dispute over the origin of the oceans. It was comparatively easy to comprehend that river ow could only be maintained by the continued precipitation of water derived from evaporating the oceans because the processes involved are readily observable and occur on a timescale that is short relative to the human lifespan. Thus, the idea of a water cycle was established. However, it was much more dif cult to comprehend the fate of the salts carried by river water into the sea. Many geologists simply regarded the ocean as an accumulation of salts, and becoming progressively saltier with time. Refuting this view proved dif cult and it gave way only gradually in the face of growing evidence that marine sediments hundreds of millions years old were essentially similar to those of the present day. This was impossible to reconcile with an ocean that must have changed in composition as it became saltier.

The uniformitarian nature of geological processes is conceptualized in the rock cycle (Figure 1), even the earliest representations of which show nearly all of the familiar geological processes (the only significant absences relate to plate tectonic processes). Magma is erupted and crystallizes into *igneous rocks*, releasing volatiles and water. These *igneous rocks* are then weathered chemically (by reaction with air and water) and physically (by abrasion through the action of wind, tides and running water). Soluble components are dissolved from the rocks and carried by rivers into the oceans. The solid residues may remain behind for considerable periods of time, where they constitute a soil, but ultimately these too are transported via rivers into the oceans. Here, additional solids are contributed by biogenic materials derived from organisms, and the mixture of weathering products and biogenic material settles to the bottom of the sea to form sediment. Burial occurs with the continued addition of sediment and is accompanied by chemical alteration, which together convert sediments into *sedimentary rocks*. Continued burial is accompanied by heating (the Earth has an average geothermal gradient of $28\,^{\circ}\mathrm{C\,km^{-1}}$), and new rock minerals are formed which are stable at these higher temperatures and pressures, converting *sedimentary rocks* into *metamorphic rocks*. Still deeper burial may ultimately result in melting, which reproduces new *igneous rocks*; the type of material at the starting point of the cycle.

However, this rock cycle is not a simple circle, from which it differs in several important ways. Firstly, the cycle can be short-circuited. For example, sediments need not necessarily become buried and transformed into metamorphic rocks, but may instead be uplifted to the surface of the Earth where they are eroded and weathered into new sediments. Secondly, the cycle is not completely closed. Melting existing rocks may produce magma, but there is also a small amount of new magma introduced to the cycle from the interior of the Earth. Equally, there is a small amount of meteoritic material that enters the cycle from outer space. Strictly speaking, these additions constitute a change such that passage around the cycle may not allow a return to precisely the same starting point. Progression

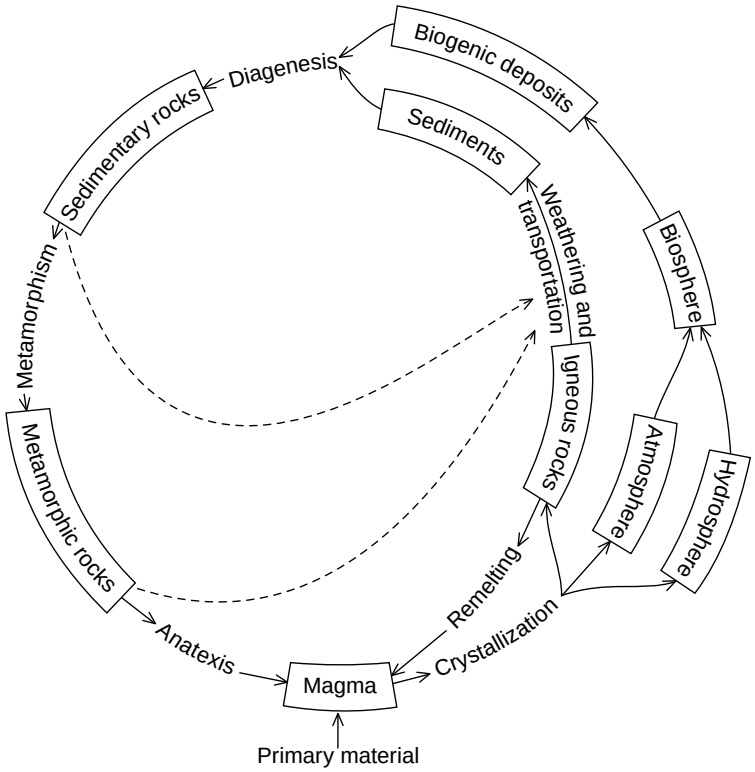


Figure 1 A conceptualization of the geological or rock cycle

in this manner can be envisaged by representing the rock cycle in three dimensions as a coil made up of loops of exactly the same diameter. Viewed end on, effectively in two dimensions, progression down the length of the coil would not be apparent and only a circular movement would be seen. Nevertheless, progress around the cycle does not result in a return to precisely the original state, and thus there is a linear progression (or global change component) to cyclicity, which could be considered analogous to aging.

The rock cycle in Figure 1 is constructed by drawing boxes (or reservoirs) that are connected by arrows (or fluxes). A flux occurs where a geological process (e.g., river transport) transfers material from one reservoir to another, and is expressed in units of mass per unit time. A reservoir represents a mass of material that possesses characteristic physical, chemical or biological properties, and is normally considered to be well mixed and thus homogeneous. This is a reasonable approximation for some reservoirs (e.g., the atmosphere) but is clearly a very poor approximation for the sedimentary rock reservoir (see below). Heterogeneity is an obvious feature of the sedimentary rock reservoir, where the common rocks (limestones, shales, sandstones) are all very different in composition while additionally numerous, less common, rocks exist with still other compositions.

A cycle approach, although an over-simplification, does possess the advantage of giving a good over-view of the relationships between different reservoirs and their connecting fluxes. This can sometimes be useful in identifying those parts of the cycle that are poorly understood. However there are disadvantages in that this approach provides no information on the processes that occur within reservoirs, or on the exact nature of the fluxes between them. Furthermore, on a large scale, it is clearly necessary to use average values, which may have to account for considerable spatial variability (as in the sedimentary rock reservoir). Such averages are difficult to obtain accurately, and may, therefore, be a poorly recognized source of uncertainty.

Nevertheless, cycles do possess the overwhelming advantage of encouraging a quantitative view of geological processes. Figure 2 shows a simplified rock cycle (Gregor, 1988) with the reservoirs and fluxes quantified, and which provides a greatly improved insight into the timescales over which the cycle operates.

(i) A *residence time* (R) can be calculated as:

$$R = \frac{\text{Mass in reservoir}(M)}{\text{Flux or fluxes}(F)} \qquad (1)$$

The residence time represents the time it would take to empty the reservoir (if all the input fluxes ceased) or the time it would take to double in mass (if all the output fluxes ceased). The residence time of

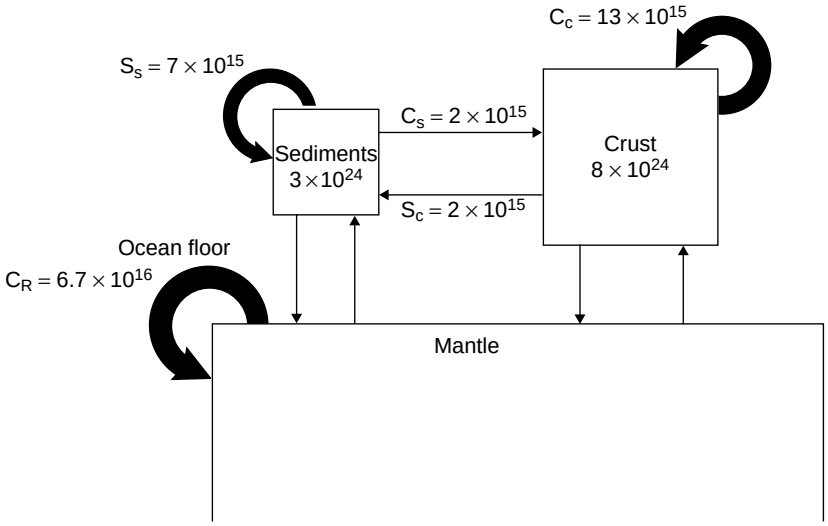


Figure 2 The rock cycle, modified from Gregor (1988). S, sediments; C, crystalline continental crust; S_s, erosion of sediments; C_s, metamorphism; S_c, erosion of crystalline rocks; C_c, recycling of crystalline rocks in tectogenesis. The ocean floor loop (C_r) represents the production and subduction of basaltic rocks (data from Veizer and Jansen, 1985). Reservoirs in grams, fluxes in grams year^{-1}

the sediment reservoir in Figure 2 is about 300×10^6 years, which represents the average time spent by sediments and sedimentary rocks in their reservoir, calculated with respect to the sum of the known input or output fluxes.

(ii) Many geological cycles are considered to represent a steady state, in which the fluxes in and out of a particular reservoir are exactly balanced. Any reservoir that is not so balanced must change in mass with time. Thus, the crust reservoir in Figure 2 is balanced because the flux of material brought in by sedimentation matches that removed to form metamorphic rocks. Geological cycles are usually assumed to be in a steady state because the rock record has remained uniform for a comparatively long time (probably $1-2 \times 10^9$ years) compared to the residence times of typical rock reservoirs (e.g., the sediment reservoir). This indicates that sufficient time has elapsed for non-steady state conditions to produce observable changes. The existence of a steady state allows a balance sheet to be drawn up comparing the sum of all the input fluxes to the output fluxes. The absence of a balance may indicate the existence of an unknown flux, whose operation is required to produce a steady state. Such budgets are thus extremely valuable, although it must be remembered that there can be considerable error in estimating an unknown flux, which is the difference between two comparatively large figures that represent the sums of the input and output fluxes.

(iii) These cycles operate by first order kinetics, i.e., the flux out of a reservoir (F) is assumed to be proportional to the reservoir mass (M). Thus,

$$F = kM \tag{2}$$

where k is a first-order rate constant. Writing $1/k = M/F$ shows from Equation (1) that residence time is the reciprocal of the first order rate constant. Now the rate of change of any reservoir ($\mathrm{d}M/\mathrm{d}T$) is given by:

$$\mathrm{d}M/\mathrm{d}T = S - F$$

where S is the sum of the input fluxes. At a steady state $\mathrm{d}M/\mathrm{d}T = 0$ and so $S = F$, as defined above. However, for a reservoir which is not in a steady state:

$$\mathrm{d}M/\mathrm{d}T = S - F = S - kM$$

This differential equation can be solved with the initial conditions that $M = M_0$ at $T = 0$, $M = M_1$ at $T = t$, which on integrating gives:

$$M_t = M_1 - (M_1 - M_0)\mathrm{e}^{-kt} \tag{3}$$

The term e^{-kt} becomes very small ($\sim$0.05) when $t = 3/k$ or $3R$ and Equation (3) then shows that M_t closely approaches the new value M_1 after the passage of three residence times. This represents the time taken by a reservoir to respond to change by reaching a new steady state.

The rock cycles in Figures 1 and 2 require that matter flows between the reservoirs; thus energy must also be transferred. Many different forms of energy are used in the rock cycle, either within, or between, reservoirs:

solar radiation, kinetic and potential energy, and chemical energy. Solar radiation is the only external source, and without this input, the second law of thermodynamics predicts that a closed system (like an Earth without an external input) would ultimately run down and become disordered.

Operation of the rock cycle is only sustained by solar radiation. The solar radiation absorbed by the crust of the Earth is unevenly distributed across the Earth s surface such that the equatorial regions gain more energy than the poles. This imbalance of energy represents an unstable situation, which winds and currents attempt to eliminate by moving heat from hot to cold regions. Overall, the energy absorbed by the crust is converted to kinetic energy, which appears as winds, currents, glaciers and rivers. These geological processes in turn cause the erosion of rocks.

The utilization of energy in, and the more dynamic aspects of, the rock cycle are more evident in Figure 3, which depicts the land–air–water system of our planet as a giant factory (Garrels and Mackenzie, 1971; Siever, 1974). This representation bears little resemblance to Figure 1, but instead likens the chemical and physical processes occurring on the Earth to the successive steps in an industrial process. There is, however, one major difference between the land–air–water factory and an industrial plant; the former neither requires any raw materials nor creates any finished product. However, the analogy is valid in that the different parts of the land–air–water factory do receive, process and transfer materials through successive operations, just as in an industrial plant.

There are two separate but linked cycles in Figure 3. One involves gas–liquid interactions in the atmosphere, and the other solid–liquid interactions between the crust and the hydrosphere. The gas–liquid system utilizes solar radiation to produce organic matter and oxygen via photosynthesis, with the reverse of this reaction returning carbon dioxide (CO_2) to the atmosphere by respiration and decay:

$$6CO_2 + 6H_2O(\text{water}) + \text{nutrients} \longrightarrow 6O_2(\text{oxygen}) + C_6H_{12}O_6(\text{carbohydrate})$$

Sunlight also drives the distillation plant to evaporate water from the hydrosphere, which is condensed as rain and enters the liquid extractor. Once in the liquid extractor, the second solid–liquid cycle operates on the crushed rock materials delivered by the pulverizer, which crushes rocks through erosion by wind, water and ice. The soluble rock components are leached out from the crushed rocks by water that contains dissolved CO_2 and water. Rivers quickly return the leachates to the oceans and, more slowly, also the residual solids.

The solids delivered to the oceanic reactor slowly deposit out to the bottom of the vessel, and are in turn covered by the following layers of sedimenting material. This vessel also acts as an evaporator, because salts can be formed by the evaporation of seawater in certain parts of the marine environment. Continued burial consolidates the accumulating sediments, first into sedimentary rocks and later into metamorphic rocks (see previous). These changes are produced by a radioactively heated pressure furnace, which can reach high enough temperatures to melt rocks and form magma.

A further benefit of Figure 3 is that it also gives some indication of the spatial variability of rock reservoirs. For example, the oceanic reactor vessel contains four different types of sediment that are spatially separated from each other. The surface of the Earth actually shows considerably more spatial variability in sediment composition than this, but Figure 3 does at least illustrate two important points. Firstly, different types of sediments can be produced in different parts of the ocean as a result of physical separation (by grain size principally) and chemical reaction (often precipitation by organism activity but also hydrothermal addition, evaporation and reaction with seawater). Secondly, it is difficult to reverse this spatial variability. There are no widespread geological processes that are particularly effective at mixing solids, although the rock flour produced by glaciers does provide a reasonable estimate of the average composition of the bedrocks from within the drainage basin. In the absence of effective mixing processes it is hardly surprising that sediments show considerable variation in age and composition across the surface of the Earth.

These observations on sediment composition and mixing can be developed in a general way to apply to all reservoirs in the rock cycle, whether gas, liquid or solid. Quite simply, a reservoir is only homogeneous when the rates at which fluxes add or remove material are slower than the rates of mixing. Because different chemical components can be added and removed at different rates, it follows that a reservoir can be homogeneous with respect to some components and heterogeneous with respect to others. The ocean provides some clear examples of this. Life processes are sufficiently rapid to deplete nutrients (nitrate, phosphate) from the surface layers of the ocean but, beyond the depth of light penetration, photosynthesis becomes impossible and nutrient concentrations are significantly higher at depth. This nutrient/depth gradient persists throughout many areas of the ocean, because the surface and deep layers of the ocean (which lie beyond the stirring effects of wind and wave action) are mixed together only slowly on a timescale of tens of thousands of years. By contrast, nutrients are being consumed and released on an annual basis by plankton. Nutrient mixing is very clearly much slower than rates of removal, and heterogeneity is established. However, the main components of seawater (e.g., sodium, chloride) are removed only slowly into sediments (on a timescale of hundreds of millions of years), and the mixing is here effective at maintaining similar proportions of sodium and chloride at surface and at depth.

The issues of mixing and homogeneity are also of importance in environmental problems, where anthropogenic activity has locally enhanced pollutant levels. In some cases, dispersion of a pollutant throughout a reservoir might greatly mitigate any potential hazard; in other cases dispersion may be undesirable because a more sensitive reservoir may be affected. The ideas developed on the geological cycle can also be applied to environmental pollutants, whether these are natural components whose concentrations have been increased by anthropogenic activity (e.g., CO_2) or are new compounds (e.g., dichlorodiphenyltrichloroethane, polychlorinated biphenyls). The cycle approach can be used to estimate the extent to which a reservoir has been perturbed, the time it will take to reach a new steady state, and the impact of the redistribution of the pollutant to other reservoirs. These questions, which lie at the very heart of global change, require the integration of cyclic processes with linear change (whether by natural aging or by anthropogenic activity).

See also: Carbon Dioxide Concentration and Climate Over Geological Times, Volume 1; Depletion of Stratospheric Ozone, Volume 1.

REFERENCES

Garrels, R M and Mackenzie, F T (1971) *Evolution of Sedimentary Rocks*, Norton, New York.

Gregor, C B (1988) Prologue: Cyclic Processes in Geology, a Historical Sketch, in *Chemical Cycles in the Evolution of the Earth*, eds C B Gregor, R M Garrels, F T Mackenzie, and J B Maynard, John Wiley & Sons, New York, 5–16.

Siever, R (1974) The Steady State of the Earth s Crust, Atmosphere and Oceans, *Sci. Am.*, 230, 72–79.

Veizer, J and Jansen, S L (1985) Basement and Sedimentary Recycling 2. Time Dimension to Global Tectonics, *J. Geol.*, 93, 625–643.

Geomorphology

The origin of the term *geomorphology* probably lies with two officers of the US Geological Survey in the 1880s, J W Powell and W J McGee. In 1891, McGee wrote The phenomena of degradation form the subject of geomorphology, the novel branch of geology (McGee, 1891). He plainly regarded geomorphology as a practice that would enable geological history to be reconstructed, not by fossils and sediments, as was the norm, but by reconstructing past erosional history. Many scientists had studied the development of erosional landforms before the term was thus defined, and since that time its meaning has become broader. It is true that until the early 1960s the reconstruction of landscape history by studying the nature of surfaces planed off by erosive processes was a dominant approach. Since that time there has been a tendency for geomorphologists to become more deeply involved with understanding not just erosional history (denudation chronology), but also the processes of erosion. In addition, geomorphologists are concerned with weathering, transport and depositional processes, with measuring the rates at which processes operate, and with quantitative analysis of the forms of the ground surface (morphometry) and of the materials of which they are composed. There are many sub-branches of geomorphology. *Climatic geomorphology* attempts to interpret landforms in terms of the climatic conditions under which they developed while *applied geomorphology* seeks to use geomorphology to deal with real world issues of resource location and hazard reduction and increasingly, *anthropogeomorphology*, where the role of humans in modifying the land surface is considered. Geomorphology has much to contribute to the study of the impacts of future global change, because global warming is likely to have a profound effect on ice cap and glacial melting, permafrost degradation, dune reactivation, sea level change, and soil erosion.

REFERENCE

McGee, W J (1891) *The Pleistocene History of Northeastern Iowa*, US Geological Survey, 11th Annual Report, 189–577.

ANDREW GOUDIE *UK*

Geothermal Heat

The interior of the Earth comprises a large reservoir of heat. This thermal energy is transported to the surface by conductive and advective mechanisms where it is transferred to the oceans or atmosphere and eventually radiated to space. The sources of the internal heat have two time scales: *primordial heat* associated with the very early history of the Earth, and *radiogenic heat* delivered slowly over the lifetime of the Earth. The primordial energy is derived from three sources. First, from gravitational energy liberated (converted to heat) as the Earth was accreted from in-falling rock and dust debris distributed through the early solar system. Subsequently, heat was derived from major impacts such as that which ejected the moon from the Earth and finally gravitational energy liberated by the segregation and sinking of the Earth s metallic core from its silicate mantle and crust, and heat from the radioactive decay of the short-lived isotopes aluminium 26 (^{26}Al), chlorine 36 (^{36}Cl) and iron 56 (^{56}Fe). All of this energy was delivered to or generated within the Earth in the first few hundred million years after the planet s birth. Subsequently, heat continued to be generated within the Earth by radioactive decay of long-lived isotopes: uranium 238 (^{238}U), uranium 235 (^{235}U), thorium 232 (^{232}Th) and potassium 40 (^{40}K), with half-lives of 4.5, 0.7, 14.1 and 1.25 billion years, respectively. Such half-lives are on the order of the 4.6 billion-year age of the Earth, and thus the heat liberated by these isotopes has been added to the internal reservoir slowly over the lifetime of the Earth. At the Earth s surface the internal heat is manifest through phenomena such as geysers, hot springs, steam vents, fumaroles, and volcanic eruptions. At shallow depths below the surface the heat is apparent in the elevated temperatures encountered in mines, and in the steady increase of temperature with depth as found in deep boreholes.

See also: Ground Temperature, Volume 1.

HENRY N POLLACK *USA*

GEWEX (Global Energy and Water Cycle Experiment)

GEWEX was initiated by the World Climate Research Program to observe, understand and model the hydrological cycle and energy fluxes in the atmosphere, at the land surface and in the upper oceans. GEWEX is an integrated program of research, observations, and science activities involving hundreds of scientists throughout the world (Figure 1).

The goal of GEWEX is to reproduce and predict, by means of suitable models, the variations of the global hydrological regime, its impact on atmospheric and surface dynamics, and variations in regional hydrological processes and water resources and their response to changes in the

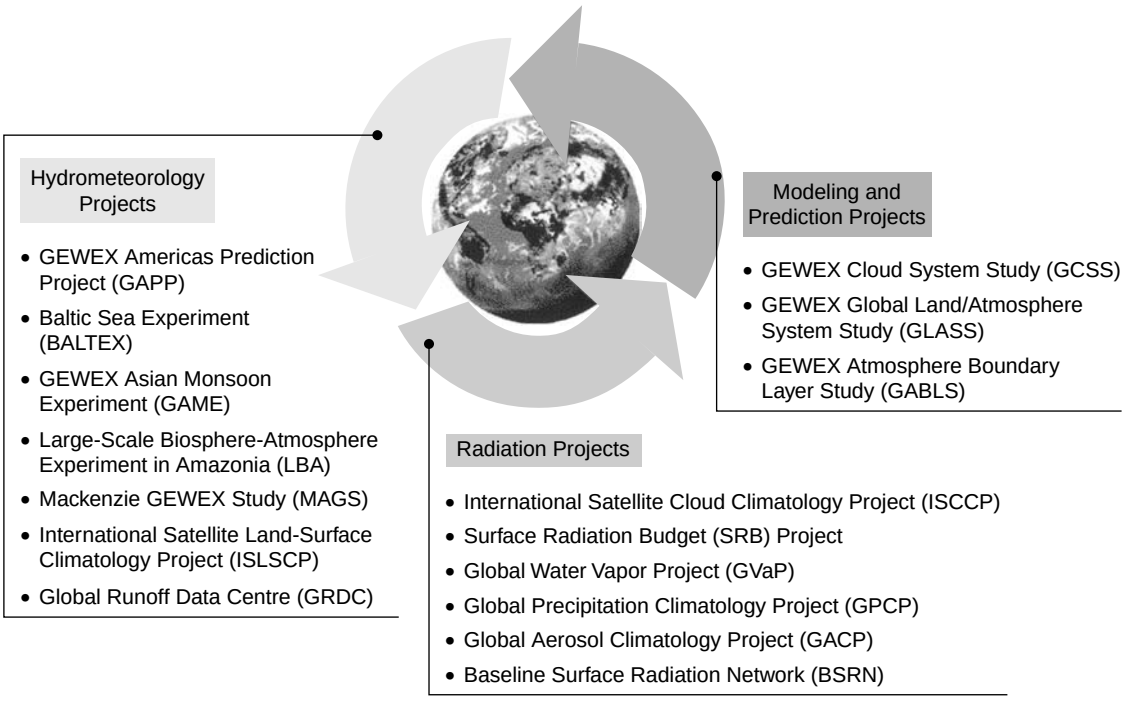


Figure 1 GEWEX projects

environment. GEWEX will provide an order of magnitude improvement in the ability to model global precipitation and evaporation, as well as an accurate assessment of the sensitivity of atmospheric radiation and clouds to climate change.

The objectives of GEWEX are to:

- determine the hydrological cycle and energy fluxes by means of global measurements of atmospheric and surface properties;
- model the global hydrological cycle and its impact on the atmosphere, oceans and land surfaces;
- develop the ability to predict the variations of global and regional hydrological processes and water resources, and their response to environmental change; and
- advance the development of observing techniques, data management, and assimilation systems for operational application to long-range weather forecasts, hydrology, and climate predictions.

GEWEX Research Foci:

- *Radiation*: determine atmospheric and surface radiation fluxes and heating with the precision needed to predict transient climate variations and decadal-to-centennial climate trends;
- *Hydrometeorology*: demonstrate skill in predicting changes in water resources and soil moisture on time scales up to seasonal and annual as an integral part of the climate system;
- *Modeling and Prediction*: develop accurate global model formulation of the energy and water budget and demonstrate predictability of their variability and response to climate forcing.

GEWEX projects are organized by their relationship to the GEWEX research focus areas.

The GEWEX Program Strategy is to build on existing programs and data; produce global data sets of clouds, radiation, water vapor, precipitation, and land surface characteristics; conduct modeling programs to model all aspects of the hydrologic and energy cycles with evolving fully coupled atmosphere–land–ocean components; and conduct pilot studies with international participation encompassing the full range of experimental scales (regional, continental, and global scales) (Figure 2).

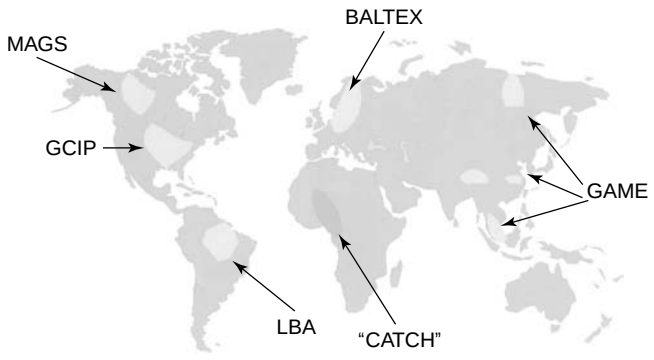


Figure 2 Regional data

A central activity of GEWEX addresses the influence of continental hydroclimatic processes on the predictability of global atmospheric circulation patterns and changes in water resources, with a particular focus on heat source and sink regions, using data sets collected from the GEWEX Continental Scale Experiments and related components of GEWEX, as well as other global hydrology initiatives (website: www.gewex.com⟨http://www.gewex.com/⟩).

PAUL TRY *USA*

Glaciers

Mark F Meier
University of Colorado at Boulder, Boulder, CO, USA

People have long been fascinated with glaciers. This is because of their obvious changes from year to year in response to a fluctuating climate, as well as their striking appearance and the apparent enigma of a brittle, crystalline substance visibly flowing down valleys. Now glaciers are subject to renewed scientific inquiry due to their usefulness as indicators of environmental change through the study of their internal layering and entrapped samples of ancient atmospheres. In addition, their role in causing rising sea levels and other effects on the environment is a cause for concern.

We discuss here all perennial ice masses lying on the Earth s land surface, except for the ice sheets of Greenland and Antarctica (*see* Antarctica, Volume 1; Greenland Ice Sheet, Volume 1). Thus we include mountain glaciers, ice fields, ice caps, and all other masses of glacier ice of less than continental dimension. Several hundred thousand of these glaciers exist on Earth. The combined area of this ice is about 680 000 km^2 with an estimated volume of 180 000 km^3. Were this ice to melt, sea level would rise about 0.5 m. Glaciers occur at almost all latitudes, including on or near the equator in South America, Africa, and Southeast Asia, to around the margins of the Greenland and Antarctic ice sheets (Table 1).

CLIMATE VARIATIONS AND GLACIER FLUCTUATIONS

Most glaciers advance and retreat in response to changes in climate; few are large enough to influence climate directly. The glacier–climate interaction can be pictured as a chain of processes, progressing from climate, through the glacier mass balance, to the resulting change in a glacier. The atmospheric environment (weather, climate) leads to processes of heat and mass exchange at a glacier surface. Mass gains (accumulation) are primarily due to snowfall, and mass losses (ablation) are usually due to melting followed by runoff. The difference between these gains and losses is called the net mass balance. Accumulation, ablation, and balance can be measured at a point or integrated over a total glacier area. Not all meltwater runs off and causes loss of mass to the glacier; some may refreeze in cold snow (internal accumulation). Glaciers reaching the ocean may also lose mass due to the discharge of icebergs.

Glaciers flow from an accumulation area where accumulation usually exceeds ablation and where the mass balance is defined as positive, to an ablation area where ablation usually exceeds accumulation and the mass balance is negative. The boundary between these two areas is the equilibrium line, and its position is a sensitive measure of a glacier s state of health (growth or shrinkage) at any given time. Glacier flow carries the excess mass of the accumulation area to where it is lost in the ablation area.

If the climate changes so that the rates of accumulation or ablation change, this changes the amount of mass added or lost from the glacier and thus its thickness. The rate of glacier flow depends critically on thickness, so the flow changes as the thickness changes. This produces change in the glacier s length (advance or retreat of the terminus), width, thickness, and volume. This change can be observed directly, or inferred from geological evidence of past positions (moraines, trimlines, etc.). Feedbacks occur in this process; for instance, the changed area affects the mass balance because it is now integrated over a different area. These fluctuations tend to bring the glacier to a new equilibrium state (Paterson, 1994).

The discussion above refers to normal glaciers. Two unusual classes of glaciers have inherent instability mechanisms that cause them to fluctuate independently of climate change. Surging glaciers oscillate between short intervals of rapid flow during which they can move up to several hundred meters per day (typical flow speeds of normal glaciers are of the order of 0.1–1.0 m day^{-1}) and periods of slow recovery from an extended position. These oscillations can be repeated at regular intervals from decades to centuries. Tidewater glaciers, which discharge icebergs into the sea and rest on bedrock (not floating), tend to advance slowly when the terminus is in relatively shallow water. But if the terminus backs off into deep water (which might be caused by a minor retreat off a moraine bank), the glacier may retreat irreversibly and rapidly until the terminus is again in shallow water. Such a situation can produce an oscillation of slow advance and rapid retreat; the period of this oscillation for tidewater glaciers in Alaska is of the order of 1000 years. Neither of these oscillations is directly

Table 1 Total areas of glacier ice in major regions, and the dimensions of some well known glaciers[a]

Region	Glacier type	Area (km^2)	Length (km)
Arctic Islands		245 000	
Mer de Glace Agassiz (Ellesmere, Canada)	Icecap	20 400	
Au. Broggerbreen (Svalbard, Norway)	Valley glacier	6.1	6
Greenland[b]		70 000	
Hans Tausen (Pearyland)	Ice cap	95	
Asia		119 000	
Guliya (Kunlun, China)	Ice cap	376	
Fedchenko (Pamir, Kirghizstan)	Valley glacier	824	77
Ts. Tuyuksu (Tien Shan, Kazahkistan)	Valley glacier	2.7	3.1
Europe		17 800	
Vatnajökull (Iceland)	Ice cap	8400	
Storglaciären (Kebnekaisa, Sweden)	Valley glacier	3.1	3.7
Hintereisferner (Alps, Austria)	Valley glacier	8.9	7.7
Djankuat (Caucasus, Russia)	Valley glacier	3.1	4.2
Africa		5	
Lewis (Mt. Kenya, Kenya)	Valley glacier	0.24	0.9
Irian Jaya		3	
North America		124 000	
Bering (Chugach/St. Elias, US/Canada)	Piedmont glacier	5200	200
Peyto (Rockies, Canada)	Valley glacier	3	5.3
Blue (Olympics, US)	Valley glacier	5.5	4.3
South America		25 900	
So. Patagonia I.F. (Andes, Chile/Argentina)	Ice field	14 000	
New Zealand		1160	
Tasman (So. Alps, New Zealand)	Valley glacier	98	28
Subantarctic Islands		7000	
Antarctica[b]		>70 000	
Regional Total		>679 000	

[a] Data from several published sources. All areas and lengths are in km^2 and km, respectively.
[b] Not including the ice sheet.

related to climate fluctuations, but they could be affected by climate changes.

GLACIER MASS BALANCES

One approach to modeling glacier variations is to use knowledge of climate history, working forward through the mass balance and dynamic response steps. A second approach is to work backward from the size variations through the dynamic response physics to infer the changes in the mass balance. However, it is impossible then to infer the causal climatic changes without additional information, because the net mass balance is the difference between two quantities (input minus output). Thus, understanding the mass balance occupies a central position in investigations of climate/glacier interactions.

Although one cannot infer climatic variables directly from glacier net mass balances, it is often possible to make estimates of snow accumulation and melting based on seasonal data. During winter, accumulation normally predominates over ablation, so a change in balance from the beginning to the end of the winter accumulation season (winter balance) provides an approximate measure of winter precipitation. The summer balance of the rest of the year is related largely to summer melting, which may be estimated using summer temperature. This approximation does not hold true for glaciers in monsoonal or tropical climates, but is useful in most glacier covered areas.

The change in net mass balance caused by a 1 $^\circ$C change in air temperature (mass balance sensitivity) is often used in glacier–climate modeling (e.g., Wigley and Raper, 1995). Oerlemans and Fortuin (1992) show that this sensitivity is also a function of precipitation. This concept has been extended to sensitivities of winter and summer balances to temperature. Numerical models of the changes in glacier mass balances due to variations in annual or winter precipitation and annual or summer temperature have been used extensively, and some of these incorporate glacier dynamics

to simulate glacier advances or retreats. However, these models must be calibrated against a period of observational mass balance data in order to extend the results back in time or into the future.

The World Glacier Monitoring Service in Zurich is charged with archiving and distributing mass balance observations collected around the world (website: http://www.geo.unizh.ch/wgms/). Cogley *et al.* (1998) and Dyurgerov and Meier (1997) have used these data together with additional observations and careful data quality analyses to produce enhanced data sets on the time series of mass balances. A major problem is the fact that before the late 1950s less than 10 glaciers were regularly measured, and since 1960 only 30–85 glaciers have been measured for mass balance. The paucity of the data is exacerbated because most of the measured glaciers are concentrated in Europe, parts of central Asia, and western North America, leaving large gaps in coverage in the Southern Hemisphere and in the polar regions.

These compilations show generally negative mass balances since the 1960s, with increasingly negative values since 1990, and much year-to-year variability (Figure 1). These trends show a crude relation to trends in mean annual Northern Hemisphere air temperature.

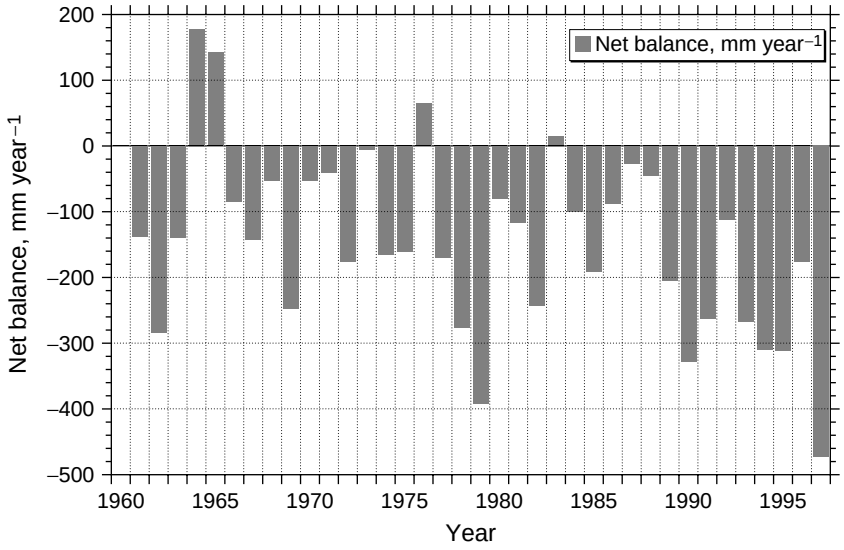


Figure 1 Annual average values of glacier net balances, in mm year^{-1} of water equivalent. The observed values have been weighted by the areas of major glacier covered regions and combined to form a global average. (Data compiled by Dyurgerov, 1999)

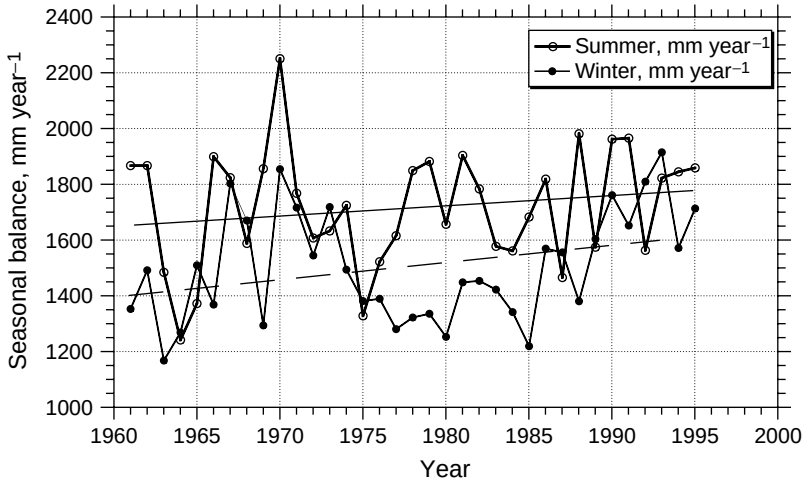


Figure 2 Winter (dots, thin line) and summer (circles, heavy line) balances of the observed glaciers of the world. The observed values have been weighted by the areas of major glacier covered regions to form a global average. Long straight lines indicate trends in the data. (Data compiled by Dyurgerov, 1999)

The trend in seasonal mass balances shows the relation between glacier changes and climate more directly (Figure 2). The trends in both winter accumulation and summer ablation since 1960 have been intensifying, but the annual fluctuations in these balances are often of different sign. During this period the altitude of the equilibrium lines has been rising, leading to diminishing accumulation area and increasing ablation area. Glacier mass balances also show teleconnections related to large-scale climatic patterns (Hodge *et al.*, 1998; McCabe *et al.*, 2000).

ICE CORES

An enormous amount of climatic information is deposited within the layered snowpack at the surface of a glacier. However, most glaciers now experience some summer (or all-year) melting, and this tends to wash out or otherwise degrade much of the interesting climatic information. Very high altitude glaciers (above 5000 m in altitude) at temperate or tropical latitudes, along with polar ice caps and ice sheets, may experience relatively little melting and thus preserve useful climatic information. These records may provide a rich history of paleoclimatic and paleoenvironmental information: accumulation, fallout, chemistry, and stable isotopes (Thompson *et al.*, 1993). These can be used to provide information on both spatial and temporal changes. Such cores have been obtained from high mountains at many latitudes, even close to the equator. The tropical glacier records, especially, show striking evidence of recent warming that is also, unfortunately, degrading these unique environmental and climatic records (Thompson *et al.*, 1993).

THE GLACIER DYNAMICS PROBLEM

Changes in mass balance cause changes in length, width, thickness, and volume of a glacier, but the pattern and rate of the geometric changes involves a dynamic (ice flow) response. Glaciers move by internal deformation caused simply by gravity acting on the weight of the ice, and, in the case of glaciers, which are at the melting point at their base, by sliding over the bed. These processes are very non-linear (Paterson, 1994). They are also sensitive to temperature: temperate glaciers (those at the melting point of ice) flow and react fairly rapidly, whereas cold or polar glaciers (those with internal temperatures below freezing) react much more slowly. Thus glaciers with similar sizes and shapes may react differently, depending on location, to a climate change.

Changes in the glacier s dimensions are not instantaneous so even with a step change in climate from one constant regime to another, the glacier will be out of equilibrium for some time. This response time (defined as the time it takes to reach 63% of the final value for a step change in climate) is typically of the order of 10 to perhaps 1000 years. Jóhannesson *et al.* (1989) have shown that the response time can be estimated as mean glacier thickness divided by mass balance at the terminus. Recent work has shown that response time depends on other factors also, and it is possible for large glaciers to have somewhat shorter response times than small glaciers, contrary to intuition.

In order to model the time series of changes in volume of a glacier, a simple decay (exponential) function may be used with response time controlling the rate of change (e.g., Wigley and Raper, 1995). The mass balance over time can be estimated from observed or projected temperature values using an observed mass balance sensitivity. Compilations of global or regional totals are difficult because each glacier has individual values of each of these parameters, and most glaciers have not been measured. The theory of stochastic scaling (Bahr, 1997) offers an opportunity to circumvent at least some aspects of this problem. It is relatively easy to determine glacier area, from a limited number of glacier inventories or by using aerial photography or satellite imagery. Scaling may, for instance, use the equations of glacier continuum mechanics to relate glacier size to other parameters that are difficult to measure, such as glacier volume and response time. These results can then be used to obtain distributions of the parameter values, which are needed for global or regional synthesis. This technique holds promise.

CHANGES IN GLACIER SIZE

Measurements of fluctuations of glacier length have been collected since at least the 14th century, and in 1894 the International Commission on Glaciers (now the International Commission on Snow and Ice) was formed to encourage and coordinate such efforts. These observations have been extended back in time by other indicators such as old paintings, moraine dating, lichen ages, tree rings, and radiocarbon dating. Although marked differences exist from region to region, these long-term records generally show minimum lengths of glaciers during the Medieval Warm Period, advances in many areas during the Little Ice Age from the 15th through the 19th centuries, and retreats since then (*see* Little Ice Age, Volume 1; Medieval Climatic Optimum, Volume 1). An example is South Cascade Glacier in Washington State (Figures 3 and 4).

An estimate of the time of an early advance of this glacier (5000 years ago) was obtained from radiocarbon dating of tree stumps that had been overridden by the advancing glacier and perfectly preserved until exposed at the terminus by glacier recession in 1958. Other information on terminus positions was obtained from tree ring counts or actual measurements. These data show that the fluctuations

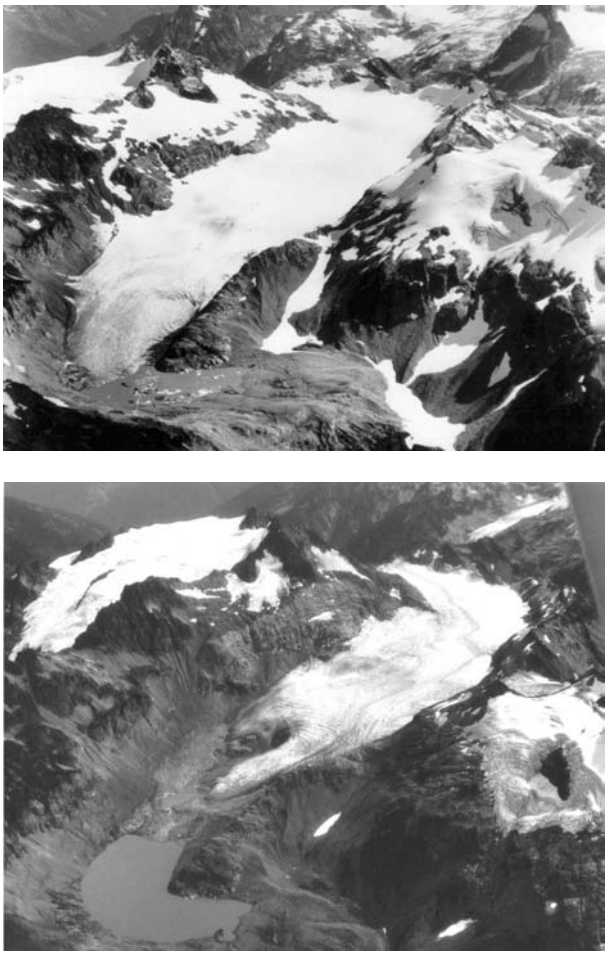


Figure 3 Photographs of South Cascade Glacier, in the North Cascades of Washington State, in 1958 (a) and in 1993 (b). Note the striking recession. (a) photograph by Austin Post, University of Washington; (b) photograph by Robert Krimmel, US Geological Survey

in length varied only between 3.6 and 4.6 km throughout the Medieval Warm Period, the Little Ice Age, and other climatic oscillations over the past 5000 years, whereas in the 20th century the glacier diminished to 2.9 km in length. This recent recession exceeds the 5000 year range of normal climate variability.

Fluctuations in glacier volume or mass balance can be measured. Many glaciers have been mapped repeatedly and some have also had measurements of thickness. For those without thickness data, it is often possible to estimate volume using scaling. A major loss of glacier ice from the late 19th to the late 20th century has occurred (Table 2). As much as half of the ice has disappeared in this century in many temperate and tropical latitudes. The recent loss of ice in Southeast Asia (Irian Jaya) and Africa has been striking, but these tiny glaciers are in a marginal climatic environment for the preservation of ice. Glaciers in the high Arctic have been shrinking very little. These high latitude

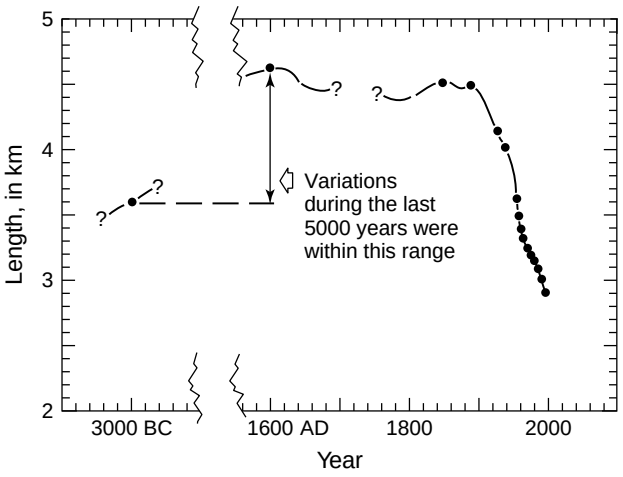


Figure 4 The history of advance and retreat (changes in length) of South Cascade Glacier. The position of the glacier between 3500 BC and 1600 AD is unknown but it had to have been between the limiting positions of these two dates because no evidence of advances beyond the 1600 AD position exists, and the dated tree stump would have rotted had it been exposed before 1958

glaciers are well below freezing in temperature throughout most of their mass, thus global warming may have produced more surface melting, but refreezing of this water below the surface has only warmed the glaciers, producing very little loss of mass.

ENVIRONMENTAL EFFECTS OF GLACIER CHANGES

Runoff, Outburst Floods, and Riparian Ecosystems

Glaciers have a self-regulating effect on streamflow: at a time of climatic warming, increased ablation increases runoff, while runoff may decrease in non-glacier streams. The additional increment of water from glacier wastage is thus especially valuable in semiarid agricultural regions such as along the Andes and in central Asia. But as the glaciers lose mass and area, this extra increment of water disappears, leaving, in a global warming scenario, a significant loss of water supply. Such an effect is expected in 10–100 years.

Glacier outburst floods, the abrupt release of water stored under a glacier in volcanic areas or due to the damming of a tributary river, can have a devastating effect on transportation corridors, other works of humankind, and riparian ecosystems (Bjornssonn *et al.*, 1998). Glacier thinning and retreat may cause a temporary increase in outburst floods but is also likely to prevent others from forming. The net effect is not known but will be very specific to the local environment.

Glacier fed runoff usually has a strong seasonal component (maximum flow in midsummer), and an equally strong

Table 2 Long-term changes in area and volume of glaciers in certain glacier covered regions[a]

Area	Time period	Loss of area (%)	Loss of volume (%)
Changes in mid-latitude glaciers			
Switzerland	1850–1973	27	31
Alps total	1850–1994	35	50
New Zealand	1890–1996	26	
Caucasus	1894–1970	29	50
TienShan	1955–1995	15	22
Changes in tropical glaciers			
Irian Jaya	1936–1990	77	
Andes[b]	1920–1970	18	
Africa	ca. 1910–1990	84	
(Lewis Gl., Kenya)	1899–1990	62	92
Changes in arctic glaciers			
Franz Josef Land	1953–1993	1.5	2.0
Novaya Zemlya	1880–1980	4	4.5
Severnaya Zemlya	1880–1980	0	1.5

[a] Data from various published sources. Some of the volume changes have been derived from areas using scaling.
[b] Huascarán-Chopicalqui Massif only.

diurnal component (maximum flow in late afternoon). Glaciers also provide finely ground rock particles (flour) to the stream. As a result, glacier rivers are always out of equilibrium and full of sediment due both to the rock flour and the constant adjustment of channel geometries. This produces distinctive types of river flood plains, geomorphic forms, and ecosystems. Removal of the glacier will change these characteristics in ways that are not easy to predict.

Icebergs

Iceberg discharge, called calving, is important in some areas such as Alaska, the Arctic, the Antarctic Peninsula, and high latitudes in South America. The discharge of ice from small glaciers may have major local effects on transportation and marine ecosystems in fjords. Calving may increase due to global change because of the warming of high latitude glaciers and the resulting increases in flow rate, but the effect has not yet been quantified.

Sea Level Rise

Many studies have been made to estimate the contribution of glacier runoff to present day sea level changes, and to project the future contribution to sea level rise for various temperature and precipitation scenarios. The 1995 Intergovernmental Panel on Climate Change (IPCC) (Houghton *et al.*, 1996) estimated that this contribution was between 2 and 5 cm over the last 100 years, out of a total measured sea level rise of 10–25 cm. More recent studies generally fall within this range, but some suggest more and some much less than the IPCC values. The IPCC used simple models to project this component of sea level rise into the future using agreed on scenarios of the growth of greenhouse gases and sulfate aerosols, and thus future regional climates. The central estimates project a sea level rise due to glacier wastage by the year 2100 of about 16 cm, but with a wide spread of possible values. Thus, about a third of the small glacier ice on Earth is likely to disappear during the 21st century (Church, 2001).

Existing estimates range widely due to differing methods of analysis and techniques for averaging or combining data from known glacier observations. Long-term records of mass balance exist for only a tiny fraction (<0.01%) of the world s glaciers, and rather drastic assumptions are needed to develop regional or global averages. Climatological data are more complete, but are seriously inaccurate in high mountain areas, and the problem of calibrating models using mass balance sensitivity still exists because of the paucity of data. The problem is compounded by insufficient knowledge of glacier response times, a very difficult parameter to measure. Modern mathematical techniques, such as scaling, may prove to be useful in developing distribution functions of glacier parameters, which then can be used to make analyses of present change and projections into the next century with much greater confidence. This is a fertile field for research.

REFERENCES

Bahr, D B (1997) Global Distributions of Glacier Properties: A Stochastic Scaling Paradigm, *Water Resour. Res.*, 33, 1669–1679.

Bjornssonn, H (1998) Hydrological Characteristics of the Drainage System Beneath a Surging Glacier, *Nature*, 395, 771–774.

Church, J A, Gregory, J M, Huybrechts, P, Kuhn, M, Lambeck, K, Nhuan, M T, Qin, D, and Woodworth, P L (2001) Changes in Sea Level, in *Climate Change 2001: The Scienti c Basis*, eds J T Houghton, Y Ding, D J Griggs, M Noguer,

P J van der Linden, X Dai, K Maskell, and C A Johnson, Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, 639–693.
Cogley, J G and Adams, W P (1998) Mass Balance of Glaciers other than the Ice Sheets, *J. Glaciol.*, 44, 315–325.
Dyurgerov, M B and Meier, M F (1997) Mass Balance of Mountain and Subpolar Glaciers: A New Global Assessment for 1961–1990, *Arct. Alp. Res.*, 29, 379–391.
Hodge, S M, Trabant, D C, Krimmel, R M, Heinrichs, T A, March, R S, and Josberger, E G (1998) Climate Variations and Changes in Mass of Three Glaciers in Western North America, *J. Clim.*, 11, 2161–2179.
Houghton, J T, Filho, L G M, Callander, B A, Harris, N, Kattenberg, A, and Maskell, K (1996) *Climate Change 1995, The Science of Climate Change*, Cambridge University Press, Cambridge.
Johannesson, T, Raymond, C E, and Waddington, E D (1989) Time-scale for Adjustments of Glaciers to Changes in Mass Balance, *J. Glaciol.*, 35, 355–369.
McCabe, G J, Fountain, A G, and Dyurgerov, M B (2000) Variability in Winter Mass Balance of Northern Hemisphere Glaciers and Relations With Atmospheric Circulation, *Arct. Ant. Alp. Res.*, 32, 64–72.
Oerlemanns, J and Fortuin, J P F (1992) Sensitivity of Glaciers and Small Ice Caps to Global Warming, *Science*, 258, 115–117.
Paterson, W S B (1994) *The Physics of Glaciers*, 3rd edition, Pergamon Press, Oxford.
Thompson, L G, Mosley-Thompson, E, Davis, M, Lin, P N, Yao, T, Dyurgerov, M, and Dai, J (1993) Recent Warming: Ice Core Evidence from Tropical Ice Cores with Emphasis on Central Asia, *Global Planet. Change*, 7, 145–156.
Wigley, T M L and Raper, S C B (1995) An Heuristic Model for Sea Level Rise due to the Melting of Small Glaciers, *Geophys. Res. Lett.*, 22, 2749–2752.

Global Analysis, Interpretation and Modeling (GAIM) Program

see GAIM (Global Analysis, Interpretation and Modeling) Program (**Volume 1**)

Global Atmosphere Watch (GAW)

see GAW (Global Atmosphere Watch) (**Volume 1**)

Global Atmospheric Electric Circuit

see Lightning and Atmospheric Electricity (**Volume 1**)

Global Atmospheric Research Program (GARP)

see GARP (Global Atmospheric Research Program) (**Volume 1**)

Global Change and Data Information System (GCDIS)

see GCDIS (Global Change Data and Information System) (**Volume 1**)

Global Climate Observing System (GCOS)

see GCOS (Global Climate Observing System) (**Volume 1**)

Global Energy and Water Cycle Experiment (GEWEX)

see GEWEX (Global Energy and Water Cycle Experiment) (**Volume 1**)

Global Ocean–Atmosphere–Land System (GOALS)

see GOALS (Global Ocean–Atmosphere–Land System) (**Volume 1**)

Global Plate Tectonics

Global plate tectonics provides the theoretical framework of modern Earth sciences. The theory holds that the Earth s surface layer, the *lithosphere*, is broken into 12 major rigid pieces or plates that move relative to each other. Plate motions are possible because the lithosphere rests upon a partially molten layer, the *asthenosphere* (*see* Plate Tectonics, Volume 1).

The major tectonic features of the globe are caused by plate interactions along three types of boundaries: convergent, divergent, and transform. At *convergent boundaries*, plates collide to generate mountain belts and volcanoes. At *divergent boundaries*, plates separate to form the mid-ocean

ridges and continental rifts. At *transform boundaries*, plates slide past one another to form transcurrent faults, such as the San Andreas, which marks the boundary between the North American and Pacific plates. Most earthquakes and volcanoes occur along plate boundaries, and the world s major physiographic features are records of past or present interactions.

Global plate tectonics is crucial to our understanding of the Earth s dynamic nature. Because the plates are in motion, the configuration of continents and oceans is constantly shifting. This explains continental drift, with its impact on evolution and speciation. Plate tectonics also plays a role in climate change. Throughout geological history, plate motions have directly affected regional climates, as continents moved passively though climate zones, and as the uplift of mountains transformed local climate patterns. Globally, the reconfiguration of continents and oceans alters oceanic circulation, which in turn affects global climate patterns. Furthermore, changes in weathering rates and volcanic activity caused by plate interactions can affect the long-term concentration of atmospheric carbon dioxide (CO_2), which may account for a significant fraction of the long-term cooling from the Cretaceous to the Pleistocene. Tectonics is thus an essential part of understanding both physical and biological global change.

See also: Continental Drift, Volume 1.

NAOMI ORESKES *USA*

Global Temperature Record

see The Global Temperature Record (*Opening essay,* **Volume 1**)

Global Warming Potential (GWP)

Global warming potentials (GWPs) provide a means of quantifying the relative potential integrated forcing on climate from emissions of various greenhouse gases (*see* Radiative Forcing, Volume 1). In the international assessments by the Intergovernmental Panel on Climate Change, GWPs have been defined as the time-integrated radiative forcing from the instantaneous release of a unit mass of a gas, expressed relative to that of the same mass of the reference gas, generally taken as carbon dioxide, the gas of most current concern to climate change. Thus, the concept of GWPs is an index to estimate the relative impact of emission of a fixed amount of one greenhouse gas compared to another, for the globally averaged radiative forcing over a specified time scale. GWPs provide a better measure of the relative greenhouse impacts than radiative forcing alone, as they help differentiate between gases that would reside in the atmosphere for vastly different amounts of time, from days to, in some case, many centuries. The GWP concept is based on the science of greenhouse gas effects, but does not include climatic or biospheric feedbacks nor consider resulting impacts on the environment. GWPs are normally only applied to gases that are well mixed in the atmosphere.

Table 1 gives the GWPs for many compounds of interest, either because of their direct potential importance to climate or because they are being considered as replacements for chlorofluorocarbons (CFCs) such as the hydrochlorofluorocarbons (HCFCs) and hydrofluorocarbons (HFCs). The GWPs are shown for integrated time horizons of 20, 100 and 500 years. No single choice matches all of the applications of this concept. Policymakers often use the 100-year time horizon. In addition to the direct radiative effects of these gases on climate, many of these gases also have indirect effects on climate due to their chemical effects on other radiatively important gases like ozone. For example, the GWP values for methane in the table include its indirect effects on tropospheric and stratospheric ozone and on stratospheric water vapor.

See also: Depletion of Stratospheric Ozone, Volume 1.

REFERENCE

World Meteorological Organization (1999) *Scienti c Assessment of Ozone Depletion: 1998*, Report No. 44, Geneva.

DONALD J WUEBBLES *USA*

GOALS (Global Ocean–Atmosphere–Land System)

GOALS is an interdisciplinary research component of the World Climate Research Programme (WCRP) CLIVAR (Climate VARiability and Predictability) aimed at improving predictions of global climate variations on seasonal to interannual time scales (*see* CLIVAR (CLImate VARiability and Predictability), Volume 1). It is based on a broad range of empirical and modeling evidence that suggests

Table 1 Direct GWPs relative to carbon dioxide[a]

Gas	Formula	GWPs (time horizon in years)		
		20	100	500
Carbon dioxide	CO_2	1	1	1
Methane	CH_4	64	24	7.5
Nitrous oxide	N_2O	330	360	190
CFCs				
CFC-11	CCl_3F	6300	4600	1600
CFC-12	CCl_2F_2	10 200	10 600	5200
CFC-13	$CClF_3$	10 000	14 000	16 300
CFC-113	CCl_2FCClF_2	6100	6000	2700
CFC-114	$CClF_2CClF_2$	7500	9800	8700
CFC-115	CF_3CClF_2	7100	10 300	14 300
HCFCs				
HCFC-21	$CHCl_2F$	700	210	65
HCFC 22	$CHClF_2$	5200	1900	590
HCFC-123	CF_3CHCl_2	390	120	36
HCFC-124	CF_3CHClF	2000	620	190
HCFC-141b	CH_3CCl_2F	2100	700	220
HCFC-142b	CH_3CClF_2	5200	2300	720
HCFC-225ca	$CF_3CF_2CHCl_2$	590	180	55
HCFC-225cb	$CClF_2CF_2CHClF$	2000	620	190
HFCs				
HFC-23	CHF_3	11 700	14 800	11 900
HFC-32	CH_2F_2	2900	880	270
HFC-41	CH_3F	460	140	43
HFC-125	CHF_2CF_3	6100	3800	1200
HFC-134	CHF_2CHF_2	3400	1200	370
HFC-134a	CH_2FCF_3	4100	1600	500
HFC-143	CHF_2CH_2F	1200	370	120
HFC-143a	CF_3CH_3	6800	5400	2000
HFC-152	CH_2FCH_2F	140	43	13
HFC-152a	CH_3CHF_2	630	190	58
HFC-161	CH_3CH_2F	33	10	3
HFC-227ea	CF_3CHFCF_3	5800	3800	1300
HFC-236cb	$CH_2FCF_2CF_3$	3500	1400	430
HFC-236ea	CHF_2CHFCF_3	3100	1000	310
HFC-236fa	$CF_3CH_2CF_3$	7500	9400	7300
HFC-245ca	$CH_2FCF_2CHF_2$	2300	720	220
HFC-365mfc	$CF_3CH_2CF_2CH_3$	2600	910	280

[a] *Source:* World Meteorological Organization (1999).

there may be many predictable components of fluctuations on these time scales. It will extend our understanding, analytical, modeling and predictive capabilities of natural climate variability on seasonal to interannual time scales, building on the knowledge and experience gained in the Tropical Oceans Global Atmosphere (TOGA) program (*see* TOGA (Tropical Ocean Global Atmosphere), Volume 1). Because GOALS has an emphasis on prediction, it will also take an interest in slow components of the climate system that give it a memory. Therefore, how the atmosphere interacts with the oceans, the land surface, and sea and land ice fields are key aspects of GOALS. The potential is to greatly improve seasonal predictions of climate variations and bring that knowledge to applications that would greatly benefit society.

CLIVAR/GOALS will examine the variability and predictability of the climate system on seasonal to interannual time scales by:

1. developing observational capabilities to describe seasonal and interannual climate variability;
2. further developing models and predictive skills for sea surface temperature and other climate variables on seasonal to interannual time scales around the entire global tropics;
3. building understanding and predictive capabilities of the interaction of monsoons with the oceans, El Niño/Southern Oscillation (ENSO) and land surface processes;
4. extending predictions from the tropics to higher latitudes, by improving understanding of CLIVAR arising

from the teleconnection linkages from lower latitudes, and interactions between the tropics and extratropics;

5. exploring the predictability and building predictive models of extratropical seasonal to interannual climate variability induced by the intrinsic low-frequency components from high latitudes such as extratropical sea surface temperatures (SSTs), soil moisture and other land surface processes, sea ice, snow cover, along with other forcings such as from volcanic eruptions. Key aspects will include improved representation in models of critical feedbacks, oceanic water mass formation, the hydrological cycle and sea-air coupling.

Initial foci of GOALS include projects dealing with global aspects of ENSO, the Variability of the American Monsoon System, the Asian–Australian Monsoon system, and African Climate Variability. More information on GOALS may be obtained from the worldwide web at http://www.clivar.org.

KEVIN E TRENBERTH *USA*

Greenhouse Effect

The term *greenhouse effect* refers to the reduction in outgoing infrared radiation to space due to the presence of the atmosphere. It can be and has been directly measured, based on the surface emission (which depends only on temperature) and the emission at the top of the atmosphere as observed by satellites. The greenhouse effect arises because the atmosphere is colder than the surface of the earth, so that the amount of radiation that the atmosphere emits to space is less than what it absorbs from the surface, giving a net reduction in emission to space. The largest contributor to the natural greenhouse effect is water vapor, followed by carbon dioxide and ozone. The natural greenhouse effect causes the global average surface temperature to be about 33 °C greater than it would be otherwise.

The physical mechanisms by which the presence of the atmosphere warms the climate and the primary mechanism that causes a glass greenhouse to be warm are different. A greenhouse heats up by day as the air within it is heated by the Sun. Outside the greenhouse, near-surface air that is heated through absorption of solar radiation by the ground surface is free to rise and be replaced with colder air from above. This cannot happen in a greenhouse because the heated air is physically prevented from being replaced with colder air. The so-called atmospheric greenhouse effect involves the net trapping of infrared radiation, which occurs independently of the movement of individual air parcels or the lack thereof. The term *greenhouse effect* is therefore a misnomer; it would be more appropriate to call it the atmospheric effect. There is, nevertheless, a weak similarity between the effect and what happens in a greenhouse. The glass enclosure will absorb some (close to 100%) of the radiation emitted from within the greenhouse, but will re-emit a smaller amount of radiation to the sky if the outer surface of the glass is colder than the interior of the greenhouse, in the same way that the emission of radiation to space is reduced if the atmosphere is colder than the surface of the earth.

Humans have already started to strengthen the natural greenhouse effect, primarily through the buildup of atmospheric carbon dioxide and, to a lesser extent, methane. As the climate warms in response to the buildup of these and other greenhouse gases emitted by humans, the amount of water vapor in the atmosphere will increase, thereby strengthening the greenhouse effect further (*see* Climate Feedbacks, Volume 1; Energy Balance and Climate, Volume 1; Infrared Radiation, Volume 1).

L D DANNY HARVEY *Canada*

Greenhouse Gases, Trends

see Air Quality, Global (**Volume 1**); Trends in Global Emmisions: Carbon, Sulfur, and Nitrogen (*Opening essay*, **Volume 3**)

Greenland

Anna Mia Stampe
Danish Space Research Institute, Copenhagen, Denmark

Greenland, the largest island in the world, is the site of much research contributing to our understanding of the climate and climate changes. Results from Greenland may be seen as a few pieces in that enormous bio-geo-chemical puzzle of the global environment. Climate research involves phenomena over a tremendous range, from oscillations in local weather on decadal scales to the global-scale glacial cycles a hundred thousand years in length.

This article provides:

- *a short description of the geography of Greenland;*
- *a short description of Greenland's contemporary climate;*

- *a discussion of why Greenland is so important in climate research, and finally;*
- *a brief overview of essential results from various scientific fields carried out in Greenland.*

FACTS ABOUT GREENLAND

Geography

Greenland is part of the Kingdom of Denmark, but is not a member of the European Economic Union and it has had home rule since 1979.

Greenland is the largest island in the world. It belongs to the North American continent and is situated within the area 59.5 to 83.4° North and −11.2 to −73.1° East. The ice-free area covers roughly 410 000 km^2, which counts for only 15% of Greenland's total area of 2 166 086 km^2. Figure 1 shows the island with the most important geographical locations indicated. The ice-sheet, or Inland Ice extends 2500 km north–south and up to 1000 km east–west (*see* Greenland Ice Sheet, Volume 1). At the center, the ice is more than 3000 m thick and represents 10% of the world's total fresh water reserves. The highest point is 3700 m high at Gunnbjørns Field on the east coast between Tasiilaq and Illoqqortoormiut. Today the northeastern part – approximately one-third of the whole country – is protected as the world's largest national park (Greenland Statistical Yearbook, 1998).

Greenland is the site of the oldest sedimentary rocks ever dated: 3700 million years. In comparison the Earth is about 4600 million years old (Minik, 1999). The ice-free coastlands are characterized by mountainous terrain. The highest peaks are situated on the east coast where peaks reach almost 4000 m above sea level. Greenland's topology was shaped by the glaciers of previous ice ages, but in several locations traces of geological processes have been found, such as earthquakes, formation of mountain ranges, volcanic activity, etc.

Contemporary Climate

The climate of Greenland is arctic and primarily controlled by the relatively warm surface of the North Atlantic Ocean and cold surface of the ice cap, which rises to altitudes of more than 3000 m. However, the presence of sea ice, cooling of the North American continent in the winter and the steep coasts of Greenland have a significant impact on the local climate, especially to the south. The annual mean temperature is below zero and summer temperatures rarely exceed 10 °C. Only inside deep fjords in the south are temperatures above this level found in summer. The lowest temperature ever recorded in the country was −70 °C at the northernmost tip of the ice cap. The coldest month is February and the warmest month August (in the coastal area: July). To the north the pattern is of continental or high arctic type (cool summers and very cold winters). The range between the coldest and warmest months is as much as 30 °C. To the south the pattern is oceanic (cool summers and relatively mild winters) with temperature ranges less than 10 °C.

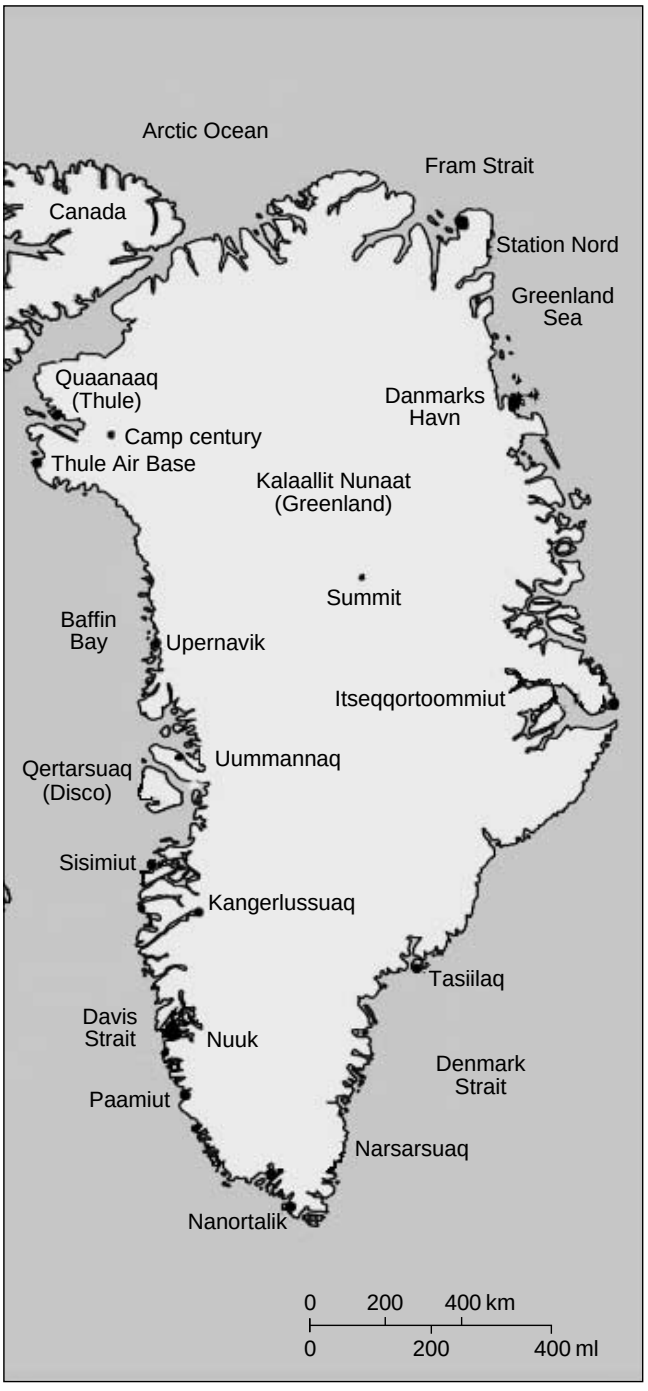


Figure 1 A simple map of Greenland showing the most important geographical locations and deep drilling sites

Precipitation varies considerably, from 800–1400 mm at the coastal stations in South Greenland to only 200 mm in the northernmost areas, which must be considered an arctic desert (Greenland Statistical Yearbook, 1998).

Most of Greenland is surrounded by sea ice, except for the southwestern shores, which are normally free of sea ice throughout the year. Several meters thick, sea ice of Arctic Ocean origin drifts southwards in the East Greenland Current and deteriorates near Cape Farewell. Locally formed sea ice (thickness: about 1 m) covers most of the Davis Strait and Baffin Bay from mid-winter until late summer. Permafrost, i.e., layers of permanently frozen soil, prevails over all of northern, and partially southern, Greenland.

In winter an area of high pressure normally exists over the northernmost part of Greenland (and northern Canada and the Arctic Ocean). Cooling of the atmosphere, primarily due to the negative radiation balance over the ice cap or sea ice covered waters far from open water, contributes to formation of this high-pressure region. The high pressure is accompanied by northerly winds (a winter monsoon) prevailing in the west and northeast Greenland waters. At the same time a low-pressure area, extending from Newfoundland via Iceland to the Norwegian Sea, together with an area of similar low pressure along the southwest coast, reflects the main zone of cyclonic activity. It appears that Greenland generally receives its stormy weather from the southwest. About 60–70% of all cyclones approaching South Greenland arrive from directions between west and south-southeast.

In summer, the mean gradient of sea level pressure around Greenland is weak and no prevailing wind direction is discernible. Cyclonic activity may take place anywhere in the Greenland area. During the year the highest pressure normally occurs around April. The lowest pressure occurs in September/October over Canada and in December/January over Greenland.

Due to the large temperature gradients and instability of the atmosphere many Atlantic cyclones develop and pass near the southern tip of Greenland, frequently causing very strong winds off south and southeast Greenland, often intensified by the orography in the fjords or at the coast.

WHY IS GREENLAND INTERESTING FOR CLIMATE RESEARCH?

Greenland is of tremendous interest to climatologists and other geophysicists for the following reasons:

- Greenland s Inland Ice represents a store of water and energy making it an important parameter in climate modeling.
- Almost all living species (flora or fauna, including humans) in an arctic region like Greenland, are on the edge of existence and to a very high degree depend on the mercy of nature.
- Climate research has shown that regions at high latitudes in general tend to be much more sensitive to climate change than are low-latitude regions. This means that if any global changes in climate occur they are more likely to be discovered in the Arctic, where impacts most probably will be more pronounced.
- Greenland is among the cleanest environments in the world, meaning that it is possible to get samples that are not polluted by contemporary waste products.
- Finally, we find in the ice caps a neatly stacked archive of information on the climate. The ice itself contains physical and chemical indications of the past. The world s two biggest ice caps, the Antarctic and the Greenlandic, provide us with time series of climate data reaching several hundred thousands years back in time. These Greenlandic data are important climatic indicators for the North Atlantic region, and in many cases also for global climate conditions as the ocean currents link the Northern and Southern Hemisphere.

When it is said that Greenland (and the Arctic in general) is a clean environment, it should in the same breath be mentioned that it is at the same time extremely sensitive to anthropogenic pollutants. Pollution in the Arctic affects nature as well as the people, of whom the majority depend on wildlife harvesting, which is the basis of their society, culture, and economy.

Sadly and ironically, Arctic people are scarcely themselves responsible for any of the major pollutants threatening the environment. On behalf of eight Arctic countries (Canada, Denmark/Greenland, Finland, Iceland, Norway, Sweden, Russia, and US) comprehensive investigations of the extent and levels of anthropogenic pollutants have been carried out by the Arctic Monitoring and Assessment Programme (AMAP) (1997). AMAP reports show that, in particular, among the harmful substances are a number of persistent organic pollutants (POPs), heavy metal contaminants from industrial sources in Europe and North America, sulfur and nitrogen compounds from sources associated with industries, energy production and transport, and finally radioactive contamination from nuclear plants. The contaminants are spread though the atmosphere and the oceans from which various processes remove them and make them available to plants and animals. This often leads to an accumulation within the food chain, leaving the higher mammals such as seals, polar bears and humans as the final destinations for these substances.

The above mentioned pollution sources are all remote and controlled by foreign political decisions. Thus, keeping the Arctic clean depends to a great extent on the common sense and goodwill of other nations.

TIME SERIES OF ENVIRONMENTAL DATA

The most important things to be aware of when working with climate data, typically a time series of various parameters, are the time scale and the limits of interpretation. Often we see time series of historical data, apparently showing a significant trend. However, if the data were viewed on a longer time scale, they would not have exhibited the same dramatic effect. The problem is that most often we only have data from recent years, a decade or a century, making it impossible to see the longer variations. Continuous and reliable measurements of atmospheric or hydrospheric parameters have only been carried out since the middle of the 19th century. When we are talking about climate variations, we should always bear in mind that the small portion of data we have does not represent the general system. Neither will a model, solely based on these data, be able to describe climatic regimes different from the present one. Therefore, it is tremendously valuable, when sources such as ice and sea-sediment cores can provide us with data series reaching much further back.

Sea-sediment cores are excellent long-scale cross-references to the glacial cores. Because the annual deposition of organic matter on the seabed is much lower than the annual snow deposition on the ice caps, the yearly resolution is much lower than for the ice cores. On the other hand, from sea-sediment cores we can obtain much older information.

CLIMATE RESEARCH IN GREENLAND

Greenland is the setting for a great many research projects, including biological, chemical, geophysical and meteorological ones. All of them contribute to a better understanding of climate and the Earth as a complex system. The following merit special mention:

- Extensive mapping of the distribution of flora and fauna. The Zackenberg Station on the west coast is the base for this work, which contributes to the general picture of Greenland s present bioclimate (Rasch, 1999).
- Meteorological measurements, which (as part of the global meteorological network) contribute to climate research as well as forecasting.
- Ozone measurements are also closely linked to the Global Ozone Monitoring Network. This is an important task because the thinning of the stratospheric ozone layer is worst over the Arctic and Antarctic regions (Goutail *et al.*, 1999).
- Magnetic field measurements: little is known of the extent to which our climate, especially cloud formation, is affected by cosmic particles. Measuring the Earth s magnetic field, which is highly influenced by particles produced by magnetic eruptions on the Sun, may provide further insight.
- Oceanography: the sea and the currents surrounding Greenland are very important parts of the global circulation.
- Sea ice studies: the extent of sea ice and the associated drift patterns are closely watched, partly for safety reasons because sea ice can be extremely dangerous to navigation, and partly because sea ice plays an important role in the complex climate system.
- Glaciology: in Greenland this primarily involves ice core drilling and analysis, mass balance studies and glacial morphology. All three fields contribute to understanding of the past climate.

A few words about the Inland Ice would be appropriate here, because it is the most characteristic feature of Greenlandic geography. The Inland Ice is the world s second largest ice cap. It covers approximately 1.7 million km^2 and 65% of this is situated above 2000 m altitude.

To the early Inuit people, the interior of the Inland Ice was a mythological, dangerous area only occupied by spirits and wickedness. This barren ice desert was no place for humans to go. To the Europeans, who later came to the island, the Inland Ice was primarily an obstacle blocking easy travel from one coast to another. Eventually explorers saw it as a challenge; arm wrestling with Nature. Who would be the first to win? Although many attempts were made, it was the then only 27 year old Norwegian, Fridjof Nansen and his party who first crossed the Greenland Inland Ice on their first polar expedition in 1888. Subsequent explorations included those by the American, Robert Peary and the Dane, Ludwig Mylius-Erichsen. Many different scientific (especially meteorological and geodetic) expeditions followed during the 20th century. Among other scientists, the German, Alfred Wegener, and later the Frenchman, Poul-Emile Victor, completed several important projects. Willi Dansgaard from Denmark initiated the era of climate research through ice core drilling in the late 1960s.

The ice cap is created when moist air from the sea is forced by the wind over the mountains in the coastal areas (orographic lifting). This causes the air to lose its humidity in the form of snow, which does not melt because of the northern location and high altitude. The annual accumulation is, on average, 35 cm. The snow is slowly compressed due to subsequent depositing of snowfall; within approximately 100 years the ice is transformed into glacial ice. Glacial ice is characterized by tiny bubbles of air, which were originally between the loose snowflakes, trapped within the ice matrix. These bubbles provide well preserved samples of atmospheric composition from the time when they were sealed off (in Greenland typically at a depth of 70 m). By extracting this air from ice cores from Greenland, and in addition Antarctica, we have a record of carbon dioxide (CO_2) reaching several hundred thousands years back (compared with approximately 45 years of operational atmospheric measurements). The annual layers of ice are

perfectly preserved down through the ice cap. However, due to gravity, they are stretched more and more in lower layers near the bedrock, because the ice, being a plastic material, slowly flows towards the coastal areas, where, in the form of huge glaciers, the ice either melts or breaks off as icebergs. This is illustrated in Figure 2. A significant part of the ice is also lost through sublimation from the surface. All in all, this leaves the ice cap in dynamic equilibrium, as long as the system is not disturbed by exterior forcing such as changes in temperature (for example due to changing ocean currents), precipitation, or solar irradiation.

The idea behind ice core drillings is that atmospheric impurities remain on the surface once they are deposited there. The impurities may be deposited either though simple fallout or as captured by precipitation. The substances include windblown continental dust and biological material, volcanic compounds, sea salt, cosmic particles and isotopes produced by cosmic radiation. The temperature on top of the Inland Ice is too low for rainfall or melting, processes that could remove the substances. The low temperature also prevents post-depositional chemical reactions. Nothing is added and nothing is removed; therefore, the impurities remain undisturbed in the ice as indicators of chemical and physical conditions of the past climate. The composition of the annual layer is only altered through decay of radioactive impurities and through diffusion processes, the latter being a significant process during the firnification (the transformation by compression of snow to ice). Chronological layering is the reason why, by drilling ice cores from the ice cap, we can obtain continuous time series of physical and chemical parameters of our past climate.

The first ice core drilled to bedrock was in 1966 at Camp Century (77.2 °N, 61.1 °W) 225 km east of Thule. This was a scientific breakthrough and an impressive engineering achievement. Other drilling projects followed and from

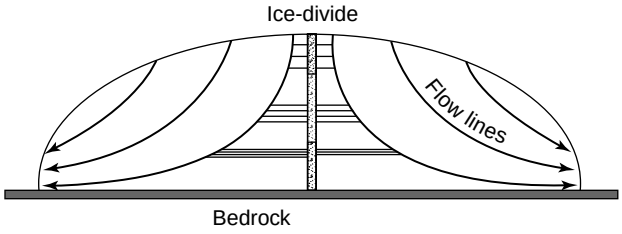


Figure 2 A schematic view of the cross-section of the ice sheet resting upon a horizontal subsurface. Ice particles deposited upon the snow surface will follow lines that travel closer to the base the farther inland the site of deposition. An ice mass formed around the divide (the location where the ice flow has only a vertical component) will be plastically deformed (thinned) with depth as suggested by the thin horizontal lines. The black arrows are flow-lines. They show the assumed profile of the horizontal velocity component

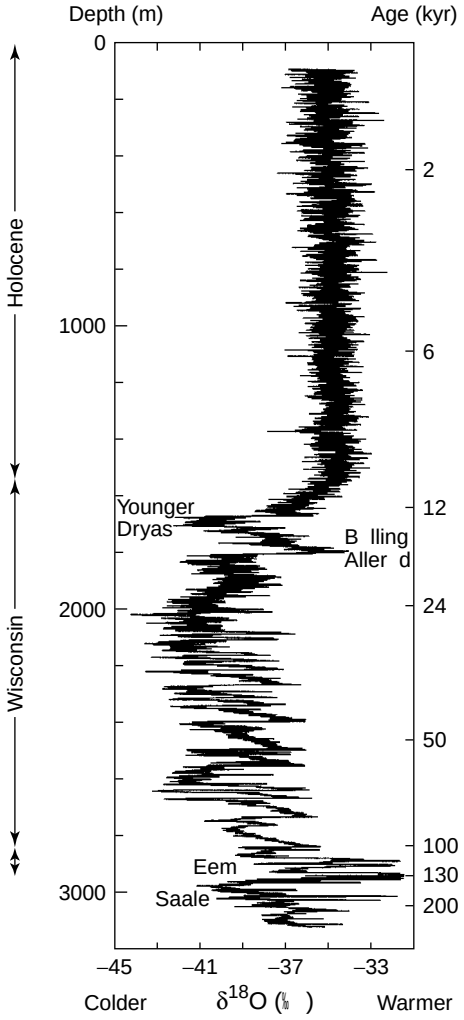


Figure 3 Stable isotope profile of the more than 3000 m long ice-core from the GRIP project drilled at Summit. This curve is produced on the basis of 55 cm mean values. The delta value is a proxy for temperature, the lower the delta value, the colder the air. The *y*-axis to the right indicates the corresponding age. (Reproduced by permission of The Glaciology Group, Department of Geophysics, University of Copenhagen)

the more than 3000 m long ice cores from the Greenland Ice Core Project (GRIP) (72.6 °N, 37.6 °W) and the Greenland Ice Core Project 2 (GISP2) (72.6 °N, 38.5 °W), both close to Summit, we now have a very accurate and detailed temperature profile reaching back about 250 000 years, i.e., back through the last glaciation (the North American Wisconsin Stage corresponding to the European Wurm Glacial Stage), through the previous warm period (the Eemian interglacial period in Europe and Sangamon in North America) and well into the previous ice age (the Saale period in Europe and the Illinoian Glacial Stage of North America) (Dansgaard *et al.*, 1993; Johnsen *et al.*, 1992). Camp Century and Summit are both marked on Figure 1. The

temperatures obtained through ice cores are only proxy values, calculated on the basis of the relation between the stable oxygen isotope ratio (which can be measured very accurately using a mass spectrometer) and the atmospheric temperature at the time of the precipitation. The stable oxygen isotope ratio is measured relative to a sample of standard mean ocean water and given in units per mille. This value is called δ^{18}O. The lower the δ^{18}O value, the colder was the atmosphere. Typical values for Greenland are between -25 per mille (contemporary) and -35 per mille (glacial values). The empirical relationship between the δ^{18}O value and temperature for Greenland is: $\delta^{18}\mathrm{O} = 0.34T - 13.7$. The δ^{18}O–temperature curve obtained from the GRIP core is shown in Figure 3.

Among the major results from the ice cores are the discovery of rapid oscillations of the climate during the Eem period. From marine oxygen isotope records it was known that climatic conditions driving the Eem were at least as warm as today, probably a bit warmer, and apparently also as stable. But the two Greenlandic cores indicate that the Eem was far from being a climatically stable period such as our present Holocene. On the contrary, the rapid regional-to-global climate oscillations that are found during most of the last 110 000 year glacial period, also persisted through the previous warm Eem period. Actually, unstable conditions seem to be more the norm than the exception. This indicates that our recent Holocene period is unique regarding climate. It is not pure guesswork to suggest that this is one of the major reasons for cultural developments in recent millennia in Asia and Europe. Other things being equal, mankind has more energy to develop agriculture and social culture when he doesn t need to fight the elements each day just to survive.

Close to bedrock (where the Eemian ice is to be found), the δ^{18}O profile of the two deep cores, GRIP and GISP2, disagree. To resolve the ambiguities, another big deep-drilling project, NorthGRIP, was initiated in north-central Greenland in 1996. Figure 4 shows the camp as it looked during the first drilling season in 1996. The drilling is expected to terminate during the summer 2001 season. Because the Eemian ice at NorthGRIP is situated higher above bedrock, the layers are less influenced by the bedrock topography; thus, this new core should be able to provide us with the most detailed and reliable data yet of the Eemian period.

From these ice cores we now know that Greenland is roughly 26 °C warmer today than during the recent ice age. This has been confirmed also by direct measurements of the preserved ice age temperature in the ice sheet (Dahl-Jensen *et al.*, 1998). When Eric the Red arrived at Greenland s west coast in 982 AD, the land probably lived up to its name. The climate was milder than today, providing relatively hospitable conditions for settlements to be established and grow. However, during the Little Ice Age (a somewhat ambiguous definition of several colder periods from approximately 1200 AD, to 1850 AD) the Norse cultures in Greenland died out (Dansgaard *et al.*, 1975) (*see* Little Ice Age, Volume 1). The coincidence of events is of course no proof, but it is believable that there is a connection, especially when one degree Celsius means dramatic changes in the ecosystems that are so vital to the population.

The ice cores also show that many of these oscillations, some of which, such as the Younger Dryas and Bølling/Allerød, have been known from sea-sediment cores, were even more rapid than hitherto anticipated. Studies

Figure 4 The NorthGRIP camp as it looked in 1996 where the NGRIP deep drilling project started. (Reproduced by permission of The Glaciology Group, Department of Geophysics, University of Copenhagen)

have shown that the climate can change abruptly within decades, and it has been suggested that the trigger of these fast climatic shifts could be changes in the global ocean circulation, a very important process, which through the Gulf Stream is closely coupled to the Greenland region.

The global circulation system transports enormous water masses which carry energy (or heat) that is critical to the Earth s climatic condition. An important part of this circulation system is the so-called Nordic Sea heat pump (Broecker, 1997; in popular form by Calvin, 1998). Hypersaline water (the result of heavy evaporation due to cold, dry winds from Canada) of the Nordic Seas (the Greenland Sea and the Norwegian Sea) pours like a sub-surface waterfall south into the lower levels of the North Atlantic Ocean with a ferocity of more than thirty times the Amazon River s combined output.

Cold water also sinks in the Labrador Sea between Canada and the southern tip of Greenland. The strength of the process is related to the North Atlantic Oscillation (NAO). The NAO is a major disturbance of the atmospheric circulation and climate of the North Atlantic–European region that is linked to a waxing and waning of the dominant middle-latitude westerly wind flow during winter. Salt sinking on such a grand scale in the Nordic Seas causes warm water to flow much farther north than it might otherwise do. The result is that northern Europe experiences a milder climate than it s latitudinal position otherwise would suggest.

ACKNOWLEDGMENTS

Thanks to Keld Q Hansen, Danish Meteorological Institute Ice and Remote Sensing Division for meteorological assistance and to J P Steffensen, Department of Geophysics at the University of Copenhagen for providing ice core data.

REFERENCES

AMAP (1997) Arctic Pollution Issues: A State of the Arctic Environment Report. A condensed version of the Scientific/Technical AMAP Assessment, 1–188.

Broecker, W S (1997) Thermohaline Circulation, the Achilles Heel of Our Climate System: Will Man-Made CO_2 Upset the Current Balance? *Science*, 278(5343), 1582–1588.

Calvin, W H (1998) The Great Climate Flip-flop, *Atlantic Monthly*, 281(1), 47–64.

Dahl-Jensen, D, Mosegaard, K, Gundestrup, N, Clow, G D, Johnsen, S J, Hansen, A W, and Balling, N (1998) Past Temperatures Directly from the Greenland Ice Sheet, *Science*, 282, 268–271.

Dansgaard, W, Johnsen, S J, Reeh, N, Gundestrup, N, Clausen, H B, and Hammer, C U (1975) Climatic Changes, Norsemen and Modern Man, *Nature*, 255(5503), 24–28.

Dansgaard, W, Johnsen, S J, Reeh, N, Gundestrup, N, Clausen, H B, and Hammer, C U (1993) Evidence for General Instability of the Past Climate from a 250-kyr Ice Core Record, *Nature*, 364(6434), 218–220.

Greenland Statistical Yearbook (1998) Naatsorsueqqissaartarfik/Statistics Greenland, Nuuk, Greenland, Kalaallit Nunanni.

Goutail, F, Pommereau, J P, Phillips, C, Deniel, C, Sarkissian, A, Lefévre, F, Kyro, E, Rammukainen, M, Eriksen, P, Andersen, S B, Kaastad-Hoiskar, B A, Braathen, G, Dorokhov, V, and Khatatov, V U (1999) Depletion of Column Ozone in the Arctic during the Winters of 1993–1994 and 1994–1995, *J. Atmos. Chem.*, 32, 1–34.

Johnsen, S J, Clausen, H B, Dansgaard, W, Fuhrer, K, Gundestrup, N, Hammer, C U, Iversen, P, Steffensen, J P, Jouzel, J, and Stauffer, B (1992) Irregular Glacial Interstadials Recorded in a new Greenland Ice Core, *Nature*, 359(6393), 311–313.

Minik, T R (1999) C13-Depleted Carbon Microparticles in >3700 Ma Sea-Floor Sedimentary Rocks from West Greenland, *Science*, 283, 674–676.

Rasch, M (1999) *Zackenberg Ecological Research Operations*, 4th Annual Report 1998, Danish Polar Center, Ministry of Research and Information Technology.

Greenland Ice Sheet

Waleed Abdalati
NASA Goddard Space Flight Center, Greenbelt, MD, USA

The Greenland ice sheet contains roughly 6% of the world's fresh water and 8% of all terrestrial ice, and with a volume of nearly 3 million km^3, it contains enough ice to raise sea level approximately 7 m. Through its exchange of energy, moisture, and mass with the atmosphere and surrounding seas, the ice sheet is an integral part of the regional climate system. Warmer than its austral counterpart, Antarctica, more than half of the Greenland ice sheet experiences some surface melt during the summer. This melt, with its associated reduction in albedo, creates a positive feedback by increasing the amount of absorbed solar energy at the surface, which in turn enhances the melt process. Consequently, this relatively temperate ice sheet is believed to be particularly sensitive to changes in climate. What is not known, however, is how cloud, precipitation, and other feedbacks will affect the overall ice sheet balance as climate changes. A widely accepted notion is that in a warming climate, at least initially, the lower and more temperate edges of the ice sheet will lose mass due to melting, while the upper portions are likely to thicken with increased accumulation. The result is expected to be a net loss of mass, contributing to an increase in sea level.

At a time when record high temperatures are being reported globally, and sea level is rising as much as 2 mm year^{-1}, by some estimates, the current state of balance of the Greenland ice sheet has important implications, yet until very recently, scientists have been unable to say with any certainty whether the ice sheet is growing or shrinking. Assessing and understanding the balance of the ice sheet have been approached by examining individual components of the large-scale mass balance equation (accumulation, melt-runoff, and discharge) and by direct measurement of the combined effects of these components.

Satellite radar altimetry has been used to measure ice sheet surface elevations for over 20 years in southern Greenland and roughly nine years for the ice sheet as a whole. By comparing observed elevations of locations at different points in time, the change in elevation, and by inference mass balance, can be determined. However, because of the physics of the radar interaction with the snow surface, the best results are generally limited to the flat high portions of the ice sheet. As a result, the most dynamic regions are either omitted from such studies, or their findings are widely questioned.

Recently, however, advances in global positioning system (GPS) techniques and airborne laser-ranging have enabled the measurement of ice sheet surface elevation from aircraft to better than 10 cm accuracy over most locations on the ice sheet. Comprehensive repeat elevation surveys with a five-year time separation between re-surveys were completed in the spring of 1999 (Figure 1). By comparing results from the two sets of measurements, the elevation change and subsequent mass balance were estimated at negative 51 km^3 year^{-1} ice (51 billion tons of water). Essentially all of this change was occurring at the lower elevations (below 2000 m), especially in southeast Greenland where thinning on one glacier exceeded 10 m year^{-1}. The upper areas, above 2000 m, which comprise approximately 60% of the ice sheet were essentially in balance over the 5-year period, but with some regional variability associated with topography and climate patterns. High elevation balance was also observed on the 20-year time scale by comparing ice flux through the 2000 m contour line with the accumulation within the region. These flux estimates were based on measured surface velocities, and accumulation rates were derived from an extensive network of ice cores. These higher elevation estimates are in broad agreement with satellite-derived estimates for the same regions.

The spatial variability that is observed at high elevations is consistent with changes in accumulation and precipitation rates over the observation period, suggesting that accumulation is the dominant mechanism controlling high elevation balance. Such a relationship is reasonable since above 2000 m, melt is very limited, except in the south, and flow rates are relatively small and stable. Precipitation estimates for 1985–1999, derived from sophisticated atmospheric models of the whole ice sheet, indicate that areas north of 70 °N have been experiencing a positive trend of just a few mm year^{-1} of water, whereas regions south of 70 °N show a reduction in precipitation on the order of 1–2 cm year^{-1}, increasing to more than 3 cm year^{-1} closer to the southeast and southwest coasts, where precipitation rates are high.

The regions below 2000 m experience seasonal melt, and as such are most sensitive to small temperature perturbations. Satellite measurements of emitted microwave energy have been successfully used to assess the spatial extent of melt, a proxy for surface ablation, by exploiting a distinct melt signature in the microwave emission of snow. Analysis of the spatial extent of summer melt shows a slightly increasing melt trend over the last 21 years on the order of 1% per year. This correlates with summer coastal temperature increases over the same period of nearly one-quarter of a degree. Similar trends are also reflected in the albedo values, which have generally decreased since 1981. This melt trend was more than four times as strong during the 1980s and early 1990s, although the aerosol loading of the stratosphere created by the Pinatubo eruptions in 1991 contributed to cooler temperatures globally and the lowest melt year (1992) in the 21 year record. Regression analysis of the melt/temperature relationship suggests a 1 °C temperature rise will increase the spatial melt extent by nearly 50%.

Coastal weather station temperature observations have been useful in helping scientists understand recent conditions of the ice sheet. However, of more importance are the climate conditions directly over the ice sheet, for which there is a paucity of data. With a growing network of 21 automatic weather stations initiated in 1995, the ice sheet climate record has been and will continue to be dramatically improved. Ice sheet temperature maps developed with data from 18 of these stations show significant increases in temperature on the ice sheet when compared to similar maps for the 1951–1960 time period. At elevations above 2000 m, warming of approximately 2 °C has been observed, reducing to 1 °C farther coastward in the 1000–2000 m elevation area. Such a change is important because warmer snow generally has a lower albedo and also requires less energy to melt.

When viewed as a whole, there are no significant recent accumulation trends on the ice sheet, but temperatures have risen and melt rates have increased. These factors contribute to the ice sheet balance, which at present is negative. Discharge rates are just now beginning to be reliably calculated for a number of outlet glaciers, so assessments of how these vary in time have yet to be made. With the complex nature of the climate system, the non-linear geological controls of the ice sheet, and the uncertainty in the feedback

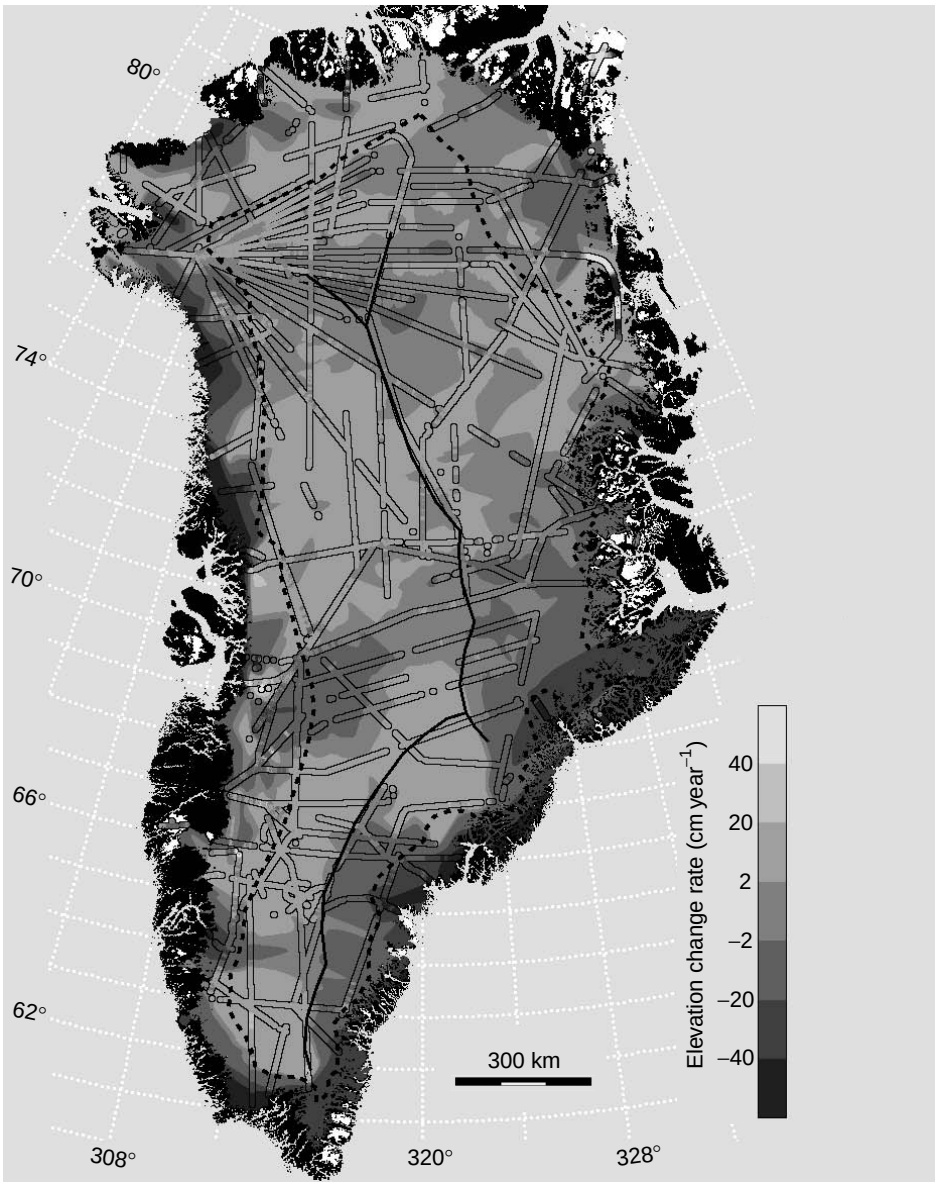


Figure 1 Observed 5-year elevation change of the Greenland ice sheet derived from airborne laser elevation surveys completed in June, 1999. The black line-pairs within the ice margin are the flight lines along which the surveys were made. The dashed line is the 2000 m contour and the solid line is the ice sheet central ridge. The region between ±2 cm represents an area of no change within the measurement uncertainty

mechanisms, it is difficult to assess how the current trends relate to the backdrop of the long-term conditions. Extrapolating these relatively brief observations to the longer time scales is not possible. However, with advances in technology and modeling, we are able to make a comprehensive assessment of the present state in the ice sheet, the factors that contribute to its balance, and how they are changing. Understanding these trends and the responsible mechanisms is an important step toward predicting their future behavior, and significant progress is being made toward that end (*see* Greenland, Volume 1).

FURTHER READING

Dahl-Jensen, D (2000) The Greenland Ice Sheet Reacts, *Science*, 289, 404–405.

Krabill, W, Abdalati, W, Frederick, E, Manizade, S, Martin, C, Swift, R, Sonntag, J, Thomas, R, Wright, W, and Yungel, J (2000) Greenland Ice Sheet: High-Elevation Balance and Peripheral Thinning, *Science*, 289(5478), 428–429.

McConnell, J R, Arthern, R J, Mosley-Thompson, E, Davis, C H, Bales, R C, Thomas, R, Burkhart, J F, and Kyne, J D (2000) Changes in Greenland Ice Sheet Elevation Attributed Primarily to Snow Accumulation Variability, *Nature*, 406, 877–879.

Ogilvie, A E and Jonsson, T, eds (2000) The Iceberg in the Mist: Northern Research in pursuit of a little age , *Clim. Change*, 48, 1–272 (Special Issue).

PARCA investigators (2001) Special Section on NASA s Program for Arctic Regional Climate Assessment (PARCA), *J. Geophys. Res.: Atmos.*, in press.

Weidick, A (1995) *Satellite Image Atlas of the Glaciers of the World, Greenland*, United States Geological Survey Professional Paper 1386-C, eds R S Williams and J G Ferrigno, US Government Printing Office, Washington, DC.

Ground Temperature

Henry N Pollack and Shaopeng Huang
University of Michigan, Ann Arbor, MI, USA

Temperatures beneath the surface of the Earth have long been measured and used in various aspects of the geological and agricultural sciences, and now are being exploited in environmental science as indicators of climate change.

The temperature of the ground surface on which we live is the outcome of coupled physical and biological processes in both the atmosphere and the ground (Geiger, 1965; Sellers, 1992) that are quite complex in detail. The most signi cant factor is the balance of incoming radiation, principally solar, and outgoing terrestrial radiation. Other important factors include latent and sensible heat uxes, ground or vegetative surface roughness and albedo, soil chemistry, hydraulic conductivity, thermal diffusivity, and moisture content, wind, chemical uxes to and from the atmosphere and more. Moreover, the relative signi cance of the factors varies with time through the diurnal and seasonal oscillations, as well as through longer term changes associated with changing land use, vegetation and climate. Because the energy ux from the Sun is some 4000 times the energy ux from the Earth's interior, the surface temperature of the planet is dominated by external, rather than internal, forcings.

Given the largely external control of the surface temperature, temperatures in the Earth s near subsurface (the outer few hundred meters) are governed principally by two processes. The first is the outward flow of heat from the deeper interior, comprising heat residual from the accretion of the Earth, and radiogenic heat produced over the lifetime of the Earth (*see* Geothermal Heat, Volume 1). The second is the spatial and temporal variations of temperature at the ground surface as it interacts with the atmosphere. The deeper heat flow reflects long-term deep-seated geological processes, and has a characteristic time scale of change on the order of millions of years. The surface temperature variability derives from topography, vegetation patterns, and climatological processes that have diurnal, seasonal, decadal and millennial characteristic time scales, much shorter than the geological processes that determine the deeper heat flow. The climatological perturbations can be thought of as transient effects superimposed on the quasi-steady-state temperature regime of the deeper heat flow.

Temperatures beneath the surface are determined in a number of ways, but the most common method is simply to lower a thermometer to a given depth in a borehole and observe the temperature at that point. Measurements can be repeated at successively greater depths until a profile of temperature versus depth is obtained (Figure 1). Another measurement scheme utilizes a number of temperature sensors fixed at various positions along a cable, which is then introduced into a borehole quasi-permanently for long-term monitoring of temperatures at the selected depths. For measurements at or very near the surface, thermometers can be buried directly in the soil. Temperatures have also been obtained at various working levels of underground mines, where thermometers are inserted into horizontal borings in the walls of the shafts to reach virgin rock temperatures undisturbed by the opening of the mineshafts and tunnels. In all these settings, various types of thermometers have been used, liquid-in-glass in some earlier surveys and electrical resistance thermometers for most modern logs.

In the absence of any climatic or other surficial perturbations, the subsurface temperature distribution reflects only the deep heat flow, and in a homogeneous medium is characterized by a linear increase of temperature with depth. The rate of increase, known as the geothermal gradient, is directly proportional to the amount of the heat flow from below, and inversely proportional to the thermal conductivity of the rock through which it is being transported. The geothermal gradient averages around $25\,^{\circ}\mathrm{C\,km^{-1}}$, but varies regionally in the range of $10-50\,^{\circ}\mathrm{C\,km^{-1}}$ in most geological settings on the continents. The regional variation arises both from regional differences in the heat flow from the interior and in the rock type at the surface. The geothermal gradient is apparent in the deeper parts of the borehole temperature profiles shown in Figure 1.

Perhaps the earliest interest in subsurface temperatures arose in the operation of underground mines where, because of the geothermal gradient, elevated temperatures at the working levels necessitated ventilation and cooling for work at deeper horizons. In the petroleum industry subsurface temperatures also play a significant role in the kinetics of the organic chemical reactions that produce petroleum and natural gas. Coal grade is also dependent in part on the temperature to which the coal has been subjected in the subsurface.

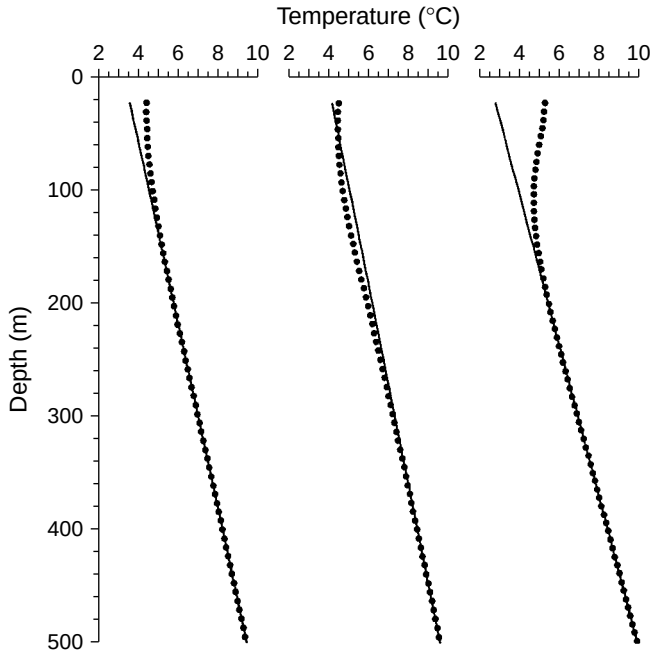


Figure 1 Temperature measurements (shown by dots) at various depths below the ground surface in three boreholes in eastern Canada. The curvature in the upper parts of the profiles is a response to temperature changes at the surface. The linear increase of temperature with depth in the deeper sections of the holes is the undisturbed geothermal gradient. The line is the extrapolation of the geothermal gradient up toward the surface; its intercept at the surface indicates what the mean surface temperature at each site was prior to the temperature changes at the surface

Beginning in the 1960s geologists and geophysicists recognized that present-day heat flow and subsurface temperatures were important parameters in understanding the composition, structure and dynamics of the Earth. Accordingly, a major global effort was made to measure these quantities in as many places as possible. Three decades later, subsurface temperatures had been determined in some 10 000 boreholes distributed unevenly over all the continents (Pollack *et al.*, 1993).

For the purpose of determining the terrestrial heat flow, the measurements of subsurface temperatures are typically made in boreholes a few hundred meters deep. The geothermal gradient, if not apparent in the upper few tens of meters because of near surface perturbations, usually can be recognized at depths below 100 m or so. The near surface observations, noise in terms of determining the geothermal gradient, were recognized as containing useful information about processes and conditions at the surface or in the shallow subsurface environment.

Climate change involving variations in the surface air temperature was identified as one possible cause of the subsurface temperature perturbations. Some of the early interest in climate change was in how it might affect the estimate of the geothermal gradient and thus the determination of heat flow from below. Accordingly, various climate change scenarios, such as the generally increasing temperatures as Earth emerged from the Last Glacial Maximum about 20 000 years ago, were constructed from other geological evidence and climate proxies. The temperature history at the surface reconstructed from such sources was then used as a time-varying boundary condition for a forward model of the perturbed subsurface temperatures and temperature gradients, which in turn could be used to make a climatic correction to the observed heat flow (Birch, 1948). These calculations revealed that climatic events occurring in the early Holocene (i.e., about 10 000 years ago) perturbed temperatures to depths of 1–2 km, and had a small but measurable effect on the estimate of the terrestrial heat flow (Beck, 1977).

There also was some early interest in using the observations of temperatures in deep boreholes and mines as constraints for the reconstruction of Holocene climate history (Lane, 1923; Cermak, 1971). More recent climate changes, such as the warming that has been observed instrumentally since the mid-to late 19th century (Jones *et al.*, 1999), the Little Ice Age of the 17th and 18th centuries, and the Medieval Climate Optimum occurring in the 13th and 14th centuries AD, would leave signatures confined to the upper few hundred meters, a depth range sampled by many of the boreholes used to determine the terrestrial heat flow (*see* Little Ice Age, Volume 1; Medieval Climatic Optimum, Volume 1). Of particular interest has been the warming of the late 19th and 20th centuries (Lachenbruch and Marshall, 1986; Pollack *et al.*, 1998) and the importance of determining the relative strengths of natural and anthropogenic forcings over the same time interval. The latter forcing is thought to arise from the well-documented increases in the concentrations of atmospheric greenhouse gases in the industrial era of human history.

Temperature changes at the Earth s surface propagate into the subsurface by heat conduction through the soil and rock. The process is entirely analogous to the warming of a cold ceramic cup following the pouring of hot tea into it. The interior surface of the cup experiences an increase of temperature, which then propagates through the wall of the cup and can be sensed a short time later on the exterior surface. The pace at which the signal propagates is governed principally by the thermal diffusivity of the conducting medium, a physical property that is directly proportional to the thermal conductivity of the material. The characteristic values of these properties in most Earth materials is such that all temperature changes at the surface during the last millennium would be confined (at present levels of detection ability) to the upper 1 km of the Earth. This slow diffusion places Earth materials in the category

of thermal insulators, as compared with metals, which transport heat more easily.

All fluctuations of temperature at the Earth s surface can be spectrally decomposed into thermal waves of different periods, with variable amplitudes and phase relations. The theory of heat conduction shows that thermal waves diminish in amplitude exponentially with depth, but that the attenuation is frequency dependent. Longer period waves attenuate more slowly with depth than do shorter period waves. Thus the Earth acts as a low-pass frequency filter, allowing a long period wave to propagate to a greater depth than a short period wave before being attenuated below the level of detection. The diurnal and seasonal variations of surface temperature, although relatively large in amplitude compared to long climatological trends, cannot be detected at depths greater than a few meters and few tens of meters, respectively. Thus, at increasing depths below the surface, only progressively longer term temperature trends are imprinted.

Virtually all of the common interpretation schemes that reconstruct ground surface temperature history from subsurface temperature profiles are cast in terms of the one-dimensional theory of heat conduction. In such a representation, the link between a present-day temperature-versus-depth profile and the past surface temperature history that produced that present-day profile is principally through the thermal diffusivity of the medium, the material property that governs space-time relations in a conductive system.

As noted earlier, many thousands of boreholes around the world have been subjected to temperature logging in the course of measuring the terrestrial heat-flow. Thus, there exists an abundance of observations. However, because of the many investigators and different measurement practices and techniques, the data are of a highly heterogeneous character. Temperature logs usually comprise measurements of temperature at discrete depths in the borehole, commonly at intervals of 5, 10 or 20 m. The boreholes penetrate to depths ranging from a few tens of meters to more than 1 km. The thermometers used have different sensitivities, generally in the range of 0.1–0.01 °C. Some borehole logs are accompanied by lithologic information and site characteristics, but most are not. Even with such heterogeneity, however, quality data are sufficiently abundant and the analysis tools sufficiently flexible to allow credible climate reconstructions from these data at many sites around the world.

Caution must be exercised in selecting data for analysis and in interpreting the reconstructions, because in addition to climatological perturbations of the subsurface temperature regime, there are many other non-climatological perturbations to contend with. Because the theoretical framework of all techniques of climate reconstruction is the one-dimensional theory of heat conduction in a laterally homogeneous medium (i.e., only vertical conductive heat transfer), any subsurface temperature variations that arise from conditions that depart from that theoretical model have the potential to be incorrectly interpreted as a climate change signature. Thus, surface topography and vegetative patterns, lateral heterogeneity in the subsurface arising from geological structures juxtaposing different rock types, and advective (non-conductive) heat transfer by groundwater movement all can alter the subsurface temperature regime with perturbations of the same order of magnitude as a climate change signature.

Many of these non-climatological disturbances can be quantitatively modeled and their significance estimated if sufficient information about the setting of a borehole is available. Alternatively, one can average the individual results from a number of boreholes spread across a region. Because it is unlikely that each borehole would have the same surficial topography and vegetative cover, or subsurface geological structure and hydrologic regime as the other boreholes, a common signal seen in the ensemble of reconstructions can more safely be attributed to climate change (Pollack *et al.*, 1996).

Analyses of several hundred temperature profiles (Pollack *et al.*, 1998; Huang *et al.*, 2000) from boreholes on six continents, taken as a global ensemble, indicate a temperature increase over the past five centuries of about 1 °C, half of which has occurred in the 20th century alone (see Figure 2). This estimate of 20th century warming is very similar in trend to the instrumental record of surface warming determined from meteorological stations (Jones *et al.*, 1999). Individually, and not surprisingly, the boreholes show a fair amount of variability, with some recording cooling and others indicating warming. But more than 75% of the borehole sites show a warming of some magnitude over

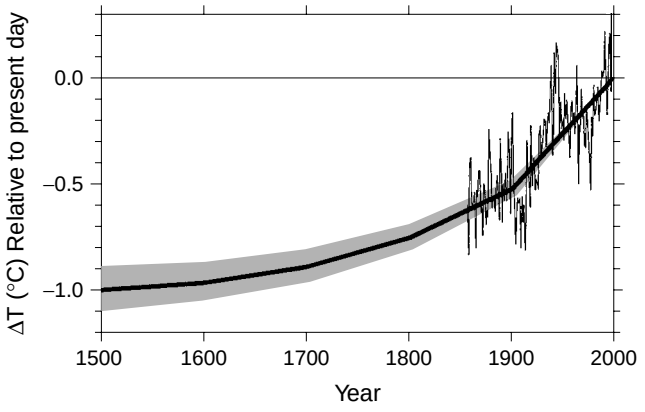


Figure 2 A reconstruction of the global mean ground surface temperature history as inferred from more than 600 borehole temperature profiles from six continents, displayed relative to the present-day temperature. The shading represents the standard error of the estimate. Also shown for comparison is a five-year running mean of the globally averaged instrumental record of surface air temperature since 1860 (modified from Huang *et al.,* 2000)

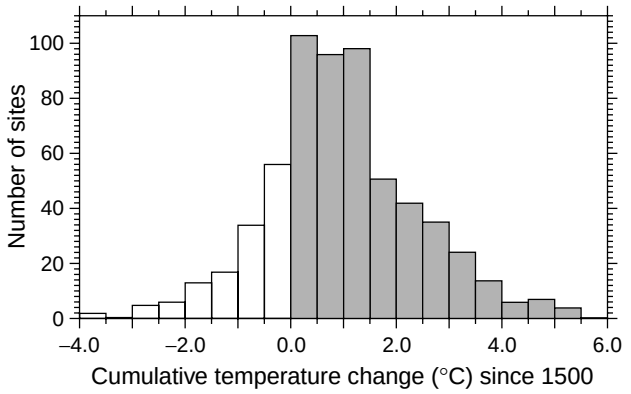


Figure 3 Distribution of temperature change since 1500 at the boreholes used in the reconstruction shown in Figure 2. Columns with shading indicate a net warming over the past five centuries, and columns without shading indicate a net cooling. More than 75% of the sites have experienced warming in this time interval

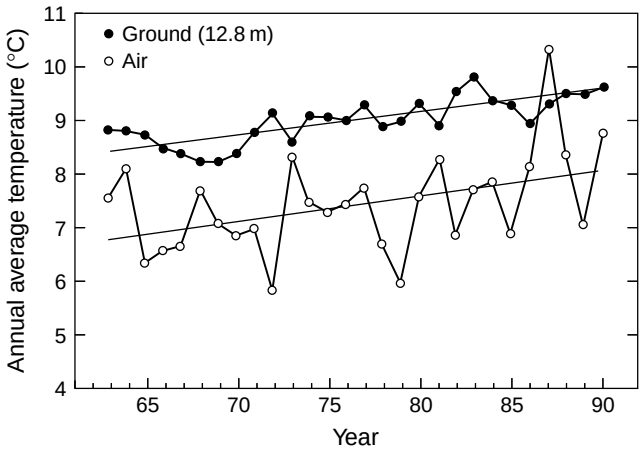


Figure 4 Annual mean soil and air temperatures at a site in north-central USA in the interval 1963–1990. Soil temperature is at a depth of 12.8 m. The lines are temperature trends over the 28-year time interval. (Modified from Baker and Ruschy, 1993)

the past five centuries (see Figure 3), clearly confirming the overall warming of the ground surface in that time interval.

Another large body of shallow subsurface temperature data comes from measurements of soil temperature at meteorological and agricultural research and experimental stations, and in boreholes drilled into permanently frozen ground (see, e.g., Baker and Ruschy, 1993; Osterkamp and Romanovsky, 1994). These measurements typically are made at depths of a few centimeters to a few tens of meters, with permanent installations of the thermometers to enable monitoring temporal changes. They frequently are accompanied by other instruments that determine air temperature and other meteorological variables at or near the surface, and soil moisture and various chemical parameters in the near subsurface. Diurnal temperature changes are observed in the upper two meters of the soil, and annual (seasonal) changes in the upper twenty meters. Annual time series for soil temperatures are rarely a century long, and usually much shorter.

An example of a soil temperature time series is displayed in Figure 4 (modified from Baker and Ruschy, 1993). At the 12.8 m depth of observation, inter-annual variations are still apparent, as well as an upward trend of temperature of $0.04\,^{\circ}\mathrm{C\,year^{-1}}$ over the 28 year time interval of the measurements. Although these shallow measurements can provide information about the surface temperature history only for the time interval represented by the time series, they are useful complements to the longer but less well resolved histories determined from the deeper borehole temperature profiles.

Although the ground temperature and the air temperature near the surface generally show similar trends of temperature change (Figure 4), the air and ground temperatures are rarely the same. Differences of several degrees are not uncommon, with the ground temperature usually warmer than the air temperature. These differences sometimes arise from simple and obvious causes, such as the effect of snow cover on the ground during winter months, which insulates the ground from the much colder diurnal and seasonal air temperature variations. The latent heat effects during the freezing and thawing of soil moisture also hold the soil temperature near $0\,^{\circ}\mathrm{C}$, independent of what the air temperature may be doing. If there are long-term changes in soil moisture that parallel long-term changes in ground surface temperature, then the trends of air and ground temperature may slowly diverge, because the latent heat effects in the soil play a greater (or lesser) role (Majorowicz and Skinner, 1997).

See also: The Global Temperature Record, Volume 1.

REFERENCES

Baker, D G and Ruschy, D L (1993) The Recent Warming in Eastern Minnesota Shown by Ground Temperatures, *Geophys. Res. Lett.*, 20, 371–374.

Beck, A E (1977) Climatically Perturbed Temperature Gradients and Their Effect on Regional and Continental Heat-flow Means, *Tectonophysics*, 41, 17–39.

Birch, F (1948) The Effects of Pleistocene Climatic Variations Upon Geothermal Gradients, *Am. J. Sci.*, 246, 729–760.

Cermak, V (1971) Underground Temperature and Inferred Climatic Temperature of the Past Millennium, *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, 10, 1–19.

Geiger, R (1965) *The Climate Near the Ground*, revised edition, English translation, Harvard University Press, Cambridge, 611.

Huang, S, Pollack, H N, and Shen, P Y (2000) Temperature Trends over the Past Five Centuries Reconstructed from Borehole Temperatures, *Nature*, 403, 756–758.

Jones, P D, New, M, Parker, D E, Martin, S, and Rigor, I G (1999) Surface Air Temperature and its Changes Over the Past 150 Years, *Rev. Geophys.*, 37, 173–199.

Lachenbruch, A H and Marshall, B V (1986) Changing Climate: Geothermal Evidence from Permafrost in the Alaskan Arctic, *Science*, 234, 689–696.

Lane, E C (1923) Geotherms of the Lake Superior Copper County, *Bull. Geol. Soc. Am.*, 34, 703–720.

Majorowicz, J A and Skinner, W R (1997) Anomalous Ground Warming Versus Surface Air Warming in the Canadian Prairie Provinces, *Clim. Change*, 35, 485–500.

Osterkamp, T E and Romanovsky, V E (1994) Characteristics of Changing Permafrost Temperatures in the Alaskan Arctic, USA, *Arctic Alpine Res.*, 28, 267–273.

Pollack, H N, Hurter, S J, and Johnson, J R (1993) Heat Loss from the Earth s Interior: Analysis of the Global Data Set, *Rev. Geophys.*, 31(3), 267–280.

Pollack, H N, Shen, P Y, and Huang, S (1996) Inference of Ground Surface Temperature History from Subsurface Temperature Data: Interpreting Ensembles of Borehole Logs, *Pure Appl. Geophys.*, 147, 537–550.

Pollack, H N, Huang, S, and Shen, P Y (1998) Climate Change Record in Subsurface Temperatures: a Global Perspective, *Science*, 282, 279–281.

Sellers, P J (1992) Biophysical Models of Land Surface Processes, in *Climate System Modeling*, ed K E Trenberth, Cambridge University Press, Cambridge, 451–490.

GWP (Global Warming Potential)

see Global Warming Potential (GWP) (**Volume 1**)

H

Hadley Circulation

The *Hadley circulation* is one of the simplest and best-known conceptual pictures (models) of the prevailing circulation of the Earth s atmosphere (*see* Atmospheric Motions, Volume 1). George Hadley originally described the conceptual model in 1735, based on a combination of sparse, mostly near-surface, atmospheric observations and 18th century physical understanding. Despite these limitations, the conceptual model is still valid, with some modifications.

In the modern picture, the stronger solar heating at low latitudes relative to high latitudes leads to rising air at or near the equator, and sinking air near the poles. To close the circulation, air flows towards the poles in the upper levels of the atmosphere, and towards the equator near the surface. There are then two equator-to-pole circulation patterns, or Hadley cells, one in each hemisphere. The Hadley cells convert potential plus thermal energy to kinetic energy. As a by-product of this conversion, they transport heat from low to high latitudes, cooling the lower latitudes and warming the higher latitudes.

The Hadley circulation also exerts a profound influence on the prevailing atmospheric winds, producing the eastward jet streams in the mid-latitude upper troposphere, and the low-latitude surface easterlies and mid-latitude surface westerlies. Because the Earth is spinning eastward on its axis, and the distance of the atmosphere from the axis of rotation decreases from equator to pole, the poleward moving air in the upper troposphere acquires eastward velocity relative to the ground, due to the tendency to conserve angular momentum (much like a spinning skater pulling in his or her arms). Similarly, air in the equatorward moving lower branches of the Hadley cells eventually acquires westward velocity, but with weaker magnitude than the upper level winds due to surface friction. As the field of meteorology developed after Hadley, an important modification to Hadley s original picture was to take into account the passage of air through storm systems at middle and high latitudes, which reduces the eastward velocities in the upper levels, and causes apparently poleward prevailing midlatitude surface winds.

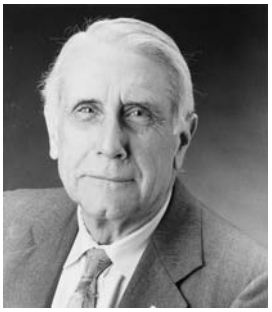
EDWIN K SCHNEIDER *USA*

Halocarbons

Halocarbons are a category of chemicals containing carbon and at least one of the halogens; fluorine, chlorine, bromine and iodine. Compounds in this category are of both natural and anthropogenic origin. Applications for the human-produced halocarbons include pharmaceuticals, crop protection, refrigerants, plastics as well as a wide range of other industrial and consumer products. Halocarbons involved in global environmental change possess three common characteristics: the compounds are persistent in the environment (generally decades to centuries), mobile in the environment, and either unique to the environment or they are emitted in sufficient quantities to provide a significant enhancement to a natural background concentration. Several classes of gaseous halocarbons that contain chlorine or bromine, including chlorofluorocarbons (CFCs) (*see* Chloro uorocarbons (CFCs), Volume 1), hydrochlorofluorocarbons (HCFCs), halons and some chlorocarbons, are involved in the stratospheric ozone-depletion issue. Production and use of these compounds are controlled under an international agreement, the Montreal Protocol on Substances that Deplete the Ozone Layer (*see* Ozone Layer: Vienna Convention and the Montreal Protocol, Volume 4). Those classes of compounds plus other gaseous halocarbons including hydrofluorocarbons (HFCs) and perfluorocarbons (PFCs) are involved in the global climate change issue. HFCs and PFCs are among the compounds listed for control under the Kyoto Protocol to the United Nations Framework Convention on Climate Change. Several of the higher molecular weight chlorocarbons are implicated in a third global environmental change issue, that of persistent organic pollutants (POPs). In addition to the three

characteristics listed above, these POP compounds bioaccumulate through the food web.

MACK MCFARLAND *USA*



Hare, Kenneth

(1919-)

F Kenneth Hare, geographer, climatologist and environmental scientist, was born in Wylye, England, in 1919. He graduated from King s College, University of London, and during World War II served as an operational meteorologist and climatologist with the United Kingdom Air Ministry.

Kenneth Hare came to Canada in late 1945 to teach at McGill University where he became Department of Geography chairman and later Dean of Arts and Science (1962–1964). For his research in arctic climatology and biogeography he was awarded a PhD (University of Montreal, 1950). At McGill he became a leading authority on heat and water balances, on climate change, and on environmental and land-use matters while serving as a senior officer of Canadian and American meteorological societies, the Arctic Institute, the Canadian Association of Geographers, and the American and Royal Geographical Societies. He was named a member of the National Research Council of Canada in 1962–1963.

In 1964, Dr Hare became Master of Birkbeck College (1964–1966) and professor at King s College, University of London, where he chaired/participated in many official enquiries and became president of the Royal Meteorological Society. He was a founding member of the UK Natural Environment Research Council in 1965–1968. He returned to Canada to become president of the University of British Columbia and was elected a fellow of the Royal Society of Canada in 1968. Dr Hare returned to teaching and research at the University of Toronto and was later seconded to Environment Canada as director-general of research coordination. In 1974–1979 he was the director of the Institute for Environmental Studies at the University of Toronto.

As program organizer, Dr Hare played a leading role in the World Meteorological Organization s 1979 World Climate Conference and was the first chairman of the Advisory Group on Greenhouse Gases. In 1988, he was honoured with international meteorology s premier award, the International Meteorological Organization (IMO) Medal. In Canada, Dr Hare served as Provost of Trinity College (1979–1986), Chancellor of Trent University (1988–1995) and chairman of the University of Toronto s Advisory Board of the Institute of International Programs (1990–1994). He has chaired/participated in the Royal Society of Canada s programs and studies on acid precipitation, nuclear winter, Pb in the environment, asbestos and global change. Also, Dr Hare was chairman of the Canadian Climate Planning Board (1979–1990) and of the Technical Advisory Panel on Nuclear Safety of Ontario Hydro (1991–1994). He has been much involved in nuclear waste management in Canada and elsewhere.

Dr Hare s published books include *The Restless Atmosphere* (1953), *On University Freedom* (1967), *Climate Canada* (co-author, 1974 and 1979), and editor of *The Experiment of Life* (1982). He has also published about 250 articles on geography, climatology and the environment in professional and popular scientific journals and yearbooks. Dr Hare has received honorary doctorates from eleven Canadian and Australian Universities and is a fellow of the American Meteorological and Geographical Societies and was president of Sigma Xi in 1986–1987. He was awarded the Patterson Medal in 1973, the Massey Medal in 1978, the Patron s Medal of the Royal Geographical Society in 1977 and the Cullum Medal of the American Geographical Society in 1987.

Made an Officer of the Order of Canada in 1979, Dr Hare was elevated to Companion, Canada s highest honour, in 1987. He is now a University Professor Emeritus at the University of Toronto.

Photo: reproduced by permission of Oakville Beaver.

MORLEY THOMAS *Canada*

HCFCs

see Hydrofluorocarbons (**Volume 1**)

Heat, Latent

see Latent Heat (**Volume 1**)

Heat, Sensible

see Sensible Heat (**Volume 1**)

Heinrich (H-) Events

John T Andrews
University of Colorado, Boulder, CO, USA

Heinrich (H)-events were first specifically described in 1988 and named by Broecker and others in 1992. They were defined as sharp peaks in the ratio of {% lithics/(% lithics + % foraminifera)} in sand-sized late Quaternary sediments from Mt Dreizack, a seamount west of Portugal. Because most deep-sea sediments consist mainly of foraminifera (sand-size carbonate tests of marine animals) and clay-size material, the dramatic increase in the numbers of lithics (i.e., grains of quartz, feldspar, and other minerals), required a significant increase in the flux of sediments from land to the oceans. The only reasonable source for this sized material was glacial sediment entrained in icebergs from the large ice sheets, which existed on the landmasses around the North Atlantic, and the massive North American Ice Sheet in particular. Research indicated that the major source of sediment for deep-sea H-events is the Hudson Strait region of Canada, a major trough which controlled the major drainage from the ice sheet toward the North Atlantic.

Evidence from cores in the North Atlantic indicates that the ocean circulation of the past 100 000 years (100 kyr) has undergone dramatic changes (Bond *et al.*, 1992). During some intervals, which are called Heinrich (H)-events, the Gulf Stream, that provides the heat to warm northwest European winters virtually disappeared. These changes appear to be associated with the introduction of freshwater into the North Atlantic basin. An understanding of the mechanism, which led to these abrupt changes, may provide some guidance as to what the future may hold under various global change scenarios.

H-events can be characterized as short-lived phenomena, lasting several hundred years and separated by about 7 kyr on average. These events are associated with plumes of glacially derived sediments being transported to the south and east across the North Atlantic (Figure 1). Studies have used a variety of methods to date the individual H-events and their duration (Bond *et al.*, 1992; Hillaire–Marcel *et al.*, 1994). The most recent 4–5 events can be dated directly by radiocarbon methods, whereas events older than 40 kyr are dated based on assumptions about sediment rate. The youngest widespread event, H-0 is of Younger Dryas age and thus had a duration of 1 kyr and occurred between 10 and 11 kyr. The best estimates for the dates of other events are: H-1 = ±14.5 kyr, H-2 = ±20.5 kyr, H-3 = ±27 kyr, and H-4 = ±36 kyr. All the events are estimated to have a duration of a few hundred years to 1 kyr.

During the 1960s and into the 1980s the prevailing paradigm for Quaternary ice sheet growth and decay was based largely on the notion that these changes were forced by solar insolation in the northern summer months (the Milankovitch theory, Imbrie *et al.*, 1992). Thus, ice sheets developed over a 100 kyr interval with stadial and

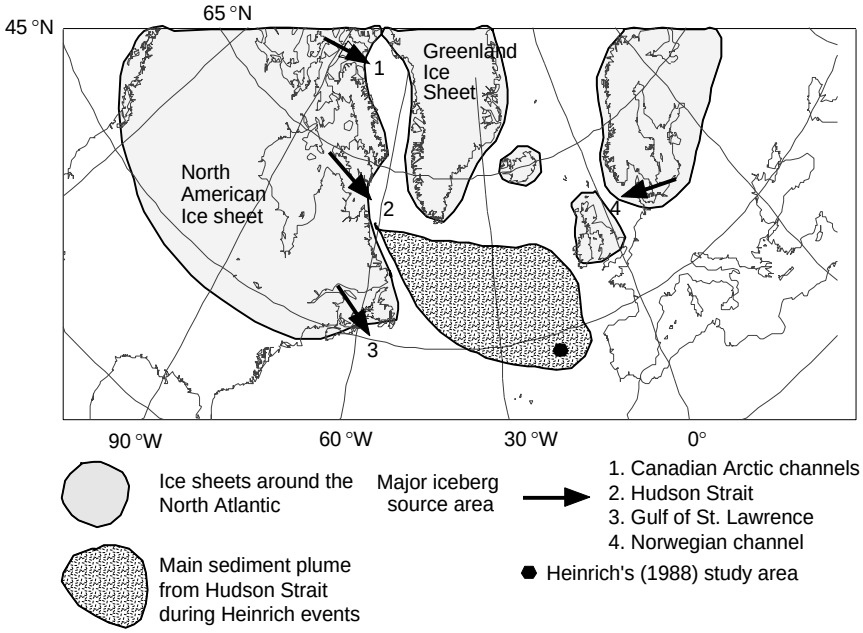


Figure 1 Map showing the North Atlantic and the main area of deposition of IRD during H-events primarily derived from the export of glacial-eroded sediment through the Hudson strait

interstadial events associated with the 21 and 41 kyr periodicities. At the same time, however, glaciologists were advancing the notion that marine-based ice sheets, i.e., ice sheets grounded below sea level, were inherently unstable. Such a designation applies to the present West Antarctic Ice Sheet, and applied to the former North American and Fennoscandian ice sheets. In the late 1970s and the 1980s researchers working on marine sediments off NE Canada had noted distinct changes in sediment color and mineral composition, especially the occurrence of intervals dominated by detrital carbonate. However, a detailed radiocarbon chronology was not available at this time, thus they were unable to assign an age or rate of sediment accumulation to these intervals (Andrews, 1998). However, in the 1980s, the advent of accelerator mass spectrometer (AMS) radiocarbon (^{14}C) dating of small (2–6 mg) samples of foraminifera revolutionized our concepts of ice sheet–ocean interactions.

H-events were first described in detail in a paper by Heinrich (1988) and their paleoclimatic significance developed by Broecker *et al.* (1992). Heinrich examined the sand fraction $>150\,\mu$m and counted the number of foraminifera, and the number of lithic (mineral grains) at closely spaced intervals in cores from Mt Dreizack, a seamount west of Portugal. He noted variations in the number of lithic grains which he ascribed, following Ruddiman (1977), to deposition from far-travelled icebergs (so-called iceberg-rafted debris, IRD). Heinrich expressed his data as a ratio of {% lithics/(% lithics + % foraminifera)}. His work showed a series of 6 discrete intervals (H-events) of high IRD and/or low foraminifera numbers within the last glacial cycle (marine isotope stage (MIS) stages 2–4). However, it was the work of Bond and co-authors (Andrews and Tedesco, 1992; Bond *et al.*, 1992) that demonstrated:

1. the widespread distribution of these events in the North Atlantic;
2. the widespread occurrence within the H-sediments of detrital carbonate;
3. the probable source for the IRD being the area of Hudson Strait (Dowdeswell *et al.*, 1995; Hemming *et al.*, 1998), which served as the major drainage conduit for the Laurentide ice sheet (LIS).

Dansgaard/Oeschger (D/O) events are those with H-intervals occurring during intervals of extreme cold (as seen in the Greenland ice sheet isotope records), but which then rapidly give way to warm conditions. Several authors, e.g., Lowell *et al.* (1995) have suggested correlations between H-events and other proxies for global change in areas removed from the North Atlantic, whereas others have noted that IRD events in Nordic Seas sediments do not have the same pattern as those off northeast Canada and in the IRD belt (Figure 1).

MECHANISMS

The discussion on the mechanism(s) for H-events lies at the heart of their importance to the global environment (Alley and MacAyeal, 1994; Clarke *et al.*, 1999). Two hypotheses have been proposed: in the first case, generally favoured by glaciologists, it is argued that H-events represent mechanical instabilities in the LIS, possibly associated with changes in the basal temperature regime; the second hypothesis argues that the global nature of short-lived events occurring at the same time as H-events indicates that they are caused by global atmospheric forcing. However, a rapid change in the atmospheric temperature cannot be transmitted quickly to the bed of an ice sheet because of the thermal properties of ice. Andrews (1998), however, presented a qualitative argument for the links between the atmosphere, ice sheet bed, and changes in relative sea-level.

OUTSTANDING PROBLEMS

Three outstanding issues related to the study of H-events remain. These are:

1. the specific glaciological mechanism(s) which led to the massive flux of icebergs into the North Atlantic;
2. the distribution of sediment within the ice sheets which allowed such long transport routes of sediment entrained in the icebergs and;
3. the chicken and egg question – are H-events the cause of, or the product of, global environmental change?

See also: Climate Change, Abrupt, Volume 1; Earth System History, Volume 1.

REFERENCES

Alley, R B and MacAyeal, D R (1994) Ice-rafted Debris Associated with Binge–Purge Oscillations of the Laurentide Ice Sheet, *Paleoceanography*, 9, 503–511.

Andrews, J T (1998) Abrupt Changes (Heinrich Events) in Late Quaternary North Atlantic Marine Environments: A History and Review of Data and Concepts, *J. Q. Sci.*, 13, 3–16.

Andrews, J T and Tedesco, K (1992) Detrital Carbonate-rich Sediments, North Western Labrador Sea: Implications for Ice-sheet Dynamics and Iceberg Rafting (Heinrich) Events in the North Atlantic, *Geology*, 20, 1087–1090.

Bond, G, Heinrich, H, Broecker, W S, Labeyrie, L, McManus, J, Andrews, J T, Huon, S, Jantschik, R, Clasen, S, Simet, C, Tedesco, K, Klas, M, Bonani, G, and Ivy, S (1992) Evidence for Massive Discharges of Icebergs into the Glacial Northern Atlantic, *Nature*, 360, 245–249.

Broecker, W S, Bond, G, McManus, J, Klas, M, and Clark, E (1992) Origin of the Northern Atlantic s Heinrich Events, *Clim. Dyn.*, 6, 265–273.

Clarke, G K C, Marshall, S J, Hillaire–Marcel, C, Bilodaeu, G, and Veiga-Pires, C (1999) A Glaciological Perspective on Heinrich Events, in *Mechanisms of Global Climate Change*

at Millennial Time Scales, eds P U Clark, R S Webb, and L D Keigwin, American Geophysical Union, Washington, DC, 243–262.

Dowdeswell, J A, Maslin, M A, Andrews, J T, and McCave, I N (1995) Iceberg Production, Debris Rafting, and the Extent and Thickness of Heinrich Layers (H-1, H-2) in North Atlantic Sediments, *Geology*, 23, 301–304.

Heinrich, H (1988) Origin and Consequences of Cyclic Ice Rafting in the Northeast Atlantic Ocean During the Past 130 000 years, *Q. Res.*, 29, 143–152.

Hemming, S R, Broecker, W S, Sharp, W D, Bond, G C, Gwiazda, R H, McManus, J F, Klas, M, and Hajdas, I (1998) Provenance of Heinrich Layers in Core V28-82, North Eastern Atlantic: 40Ar/39Ar Ages of Ice-rafted Hornblende, Pb Isotopes in Feldspar Grains, and Nd-Sr-Pb Isotopes in the Fine Sediment Fraction, *Earth Planet. Sci. Lett.*, 164, 317–333.

Hillaire–Marcel, C, de Vernal, A, Bilodeau, G, and Wu, G (1994) Isotope Stratigraphy, Sedimentation Rates, Deep Circulation, and Carbonate Events in the Labrador Sea During the Last ~200 kyr, *Can. J. Earth Sci.*, 31, 63–89.

Imbrie, J, Boyle, E A, Clemens, S C, Duffy, A, Howard, W R, Kukla, G, Kutzbach, J, Martinson, D G, McIntyre, A, Mix, A C, Molfino, B, Morley, J J, Peterson, L C, Pisias, N G, Prell, W L, Raymo, M E, Shackleton, N J, and Toggweiler, J R (1992) On the Structure and Origin of Major Glaciation Cycles, 1, Linear Responses to Milankovitch Forcing, *Paleoceanography*, 7, 701–738.

Lowell, T V, Heusser, C J, Andersen, B G, Moreno, P I, Hauser, A, Heusser, L E, Schluchter, C, Marchant, D R, and Denton, G H (1995) Interhemispheric Correlation of Late Pleistocene Glacial Events, *Science*, 269, 1541–1549.

Ruddiman, W F (1977) Late Quaternary Deposition of Ice-rafted Sand in the Sub-polar North Atlantic (40–60 N), *Geol. Soc. Am. Bull.*, 88, 1813–1827.

HFCs (Hydrofluorocarbons)

see Hydrofluorocarbons (**Volume 1**)

History, Earth System

see Earth System History (*Opening essay*, **Volume 1**)

Holocene

The *Holocene* epoch is the current interglacial, or post glacial epoch following the Pleistocene. Interglacials are defined by the widespread presence of broad-leafed trees, such as oaks, in Central Europe, whereas glacial periods are defined by the replacement of deciduous forests by grasslands (Imbrie and Imbrie, 1986). The Holocene began about 10 000 years ago, when there were still remnant Laurentide and Fennoscandian ice sheets. However, greater solar insolation in Northern Hemisphere summers than the present led to warmer conditions than present (or pre-industrial) times in many places not under the direct influence of the ice sheets. The more positive orbital insolation forcing 10 000 years ago was the result of differences in the Earth s precession, which determines the date of perihelion (Earth s closest approach to the sun), and obliquity, or tilt (*see* Orbital Variations, Volume 1). Today, perihelion occurs in Northern Hemisphere winter, whereas at the start of the Holocene, perihelion occurred during Northern Hemisphere summer, resulting in warmer summers and colder winters in the Northern Hemisphere. The Earth s tilt was nearly 1° greater 10 000 years ago than the present, resulting in more extreme summers and winters in both hemispheres (Kutzbach *et al.*, 1998). These differences in orbital insolation also resulted in a stronger Southwest Asian monsoon, resulting in higher lake levels throughout Asia and Africa during the early Holocene (Prell and Kutzbach, 1992). The remaining ice sheets were mostly melted by 6000 years ago. However, partly due to changing orbital insolation, the late Holocene during the last 4500 years was cooler and drier than the early Holocene. This most recent time frame is often referred to as the Neoglacial (Crowley and North, 1991).

See also: Deserts, Volume 1.

REFERENCES

Crowley, T J and North, G R (1991) *Paleoclimatology*, Oxford University Press, New York, 1–339.

Imbrie, J and Imbrie, K P (1986) *Ice Ages: Solving the Mystery*, Harvard University Press, Cambridge, MA, 1–224.

Kutzbach, J, Gallimore, R, and Laarif, F (1998) Climate and Biome Simulations for the past 21 000 years, *Quatern. Sci. Rev.*, 17(6/7), 473–506.

Prell, W L and Kutzbach, J E (1992) Sensitivity of the Indian Monsoon to Forcing Parameters and Implications for its Evolution, *Nature*, 360, 647–652.

BENJAMIN S FELZER *USA*

Holocene, Climate Change and Society

see Holocene: Climate Changes and Society (**Volume 3**)

Houghton, John Theodore

(1931-)

Sir John Theodore Houghton was born in 1931, and became a leading UK satellite and climate scientist and author. For more than 40 years, Sir John Houghton has exerted a strong and beneficial influence on the development of the science of meteorology and its applications to practical problems throughout the world.

Sir John is well known internationally for his research in remote sensing of the atmosphere from space. In cooperation with Professor Desmond Smith, he developed the Selective Chopper Radiometer for the Nimbus 4 and 5 satellites in the early 1970s. Other instruments developed by the group at Oxford University led by Sir John, provided, for the first time, global information of the structure of the stratosphere and mesosphere and played a major part in the implementation of a detailed study of the radiation, dynamics and chemistry of the whole atmosphere. In cooperation with Professor Fred Taylor, then at the Jet Propulsion Laboratory, Sir John developed the Pioneer Venus Orbiter in 1978.

As chief executive (formerly director-general) of the UK Meteorological Office he was responsible for major innovations in the meteorological service and was influential in developing the United Kingdom s strong international reputation in many areas of climate research. Sir John was the driving force in establishing the Hadley Centre for Climate Prediction and Research which, since its inception, has sought to collaborate in, and contribute to, the growing international research effort on climate change.

As permanent representative of the United Kingdom with the World Meteorological Oganization (WMO) for 8 years (and third vice president 1987–1991), his combination of thorough technical knowledge and his appreciation of practical problems enabled him to exert a valuable influence on the development of WMO policy.

Due to his gift for achieving consensus at meetings, he is held in high regard as a chairman. Between 1981 and 1984 Sir John was chairman of the Joint Scientific Committee for the World Climate Research Programme and from 1992 to 1996 was chairman of the Joint Scientific Committee for the Global Climate Observing System. He was invited in 1993 to chair the Intergovernmental Meeting on the World Climate Programme: The Climate Agenda.

In 1988, he was appointed chairman of the Scientific Assessment Working Group of the Intergovernmental Panel on Climate Change (and is currently co-chairman). This working group, IPCC, Working Group I, has produced five major reports under his leadership (see Intergovernmental Panel on Climate Change (IPCC): an Historical Review, Volume 4; IPCC (Intergovernmental Panel on Climate Change), Volume 1). To achieve this has required great sensitivity to the political aspects of climate change, combined with firmness to ensure that the reports remained faithful to the underlying science.

Sir John has continued to represent meteorology in the higher echelons of British science. From 1994 to 2000 he was one of the five members of the British Government Panel on Sustainable Development. He became a member of the Royal Commission on Environmental Pollution in 1991 and from 1992 to 1998 was chairman of this distinguished body.

Sir John is a committed Christian and has sought to build intellectual bridges between science and religion. He has authored two books on this subject, in addition to his well known books on the physics of the atmosphere and climate change.

Sir John was elected a fellow of the Royal Society in 1972. In recognition of his outstanding services to science and meteorology, Sir John received the Queen s Honour of Commander, Order of the British Empire in 1983 and, in 1991, the Royal Accolade of Knight Bachelor.

DAVID J GRIGGS *UK*

Human-induced Changes in Atmospheric Composition

see Anthropogenic (**Volume 3**)

Humidity

Humidity refers to the water vapor content of the air, and may be expressed in several ways.

The *absolute humidity* is the mass of water per unit volume of air. For example, extremely humid air at normal room temperature contains about 20 g of water vapor in each cubic meter of air. The most common and most useful measure of humidity, however, is the relative humidity, because this governs processes of evaporation, condensation, and precipitation. *Relative humidity* is defined as the ratio between the actual moisture content of the air and the saturation moisture content. The latter is simply the

maximum amount of moisture that can be put into the air by evaporation. In other words, when the air is saturated, surfaces of water gain as many molecules of water from the air as they lose, and no net evaporation takes place.

Low values of relative humidity indicate dry air, and encourage high rates of evaporation. When the relative humidity increases to reach or to briefly exceed saturation levels (as occurs, for example, when air cools by radiation at night under clear skies or in rising air currents), water may condense on tiny particles in the atmosphere to form fog, clouds, or precipitation, or to form dew or frost on cold surfaces.

Another measure of humidity often encountered is the *dew point*. This is the temperature at which moisture will condense on a cold surface. The closer this temperature is to the ambient temperature, the higher the humidity.

The amount of water vapor in the atmosphere generally decreases with altitude, and varies markedly over the globe. However, relative humidity can sometimes increase with altitude, mainly as a result of cooling of air through ascent or loss of heat through radiation. These processes govern the formation of fog, clouds, rain, and snow that in turn determine the habitability of our climate and our planet. In summary, water is both the most variable and the most vital trace component of our atmosphere.

JOHN S PERRY *USA*

Hurricanes, Typhoons and other Tropical Storms – Descriptive Overview

Kendal McGuffie
University of Technology, Sydney, Australia

Tropical cyclones (TCs), known regionally as typhoons, hurricanes and cyclones, are devastating weather events to which the human population is increasingly vulnerable. The physical character, development and effects of these storms are the same in the North Atlantic (hurricanes), in the Southwest Paci c (TCs), North and South Indian Ocean (cyclones) and in the Northwest Paci c (typhoons). The terms hurricane and typhoon can therefore be considered as regionally speci c names for severe tropical cyclonic storms. These storms, no matter what oceanic region they occur in, can be referred to as TCs without loss of correctness. Tropical cyclone is a generic term for a non-frontal synoptic scale low pressure system that originates and develops over tropical or subtropical waters. When the sustained surface wind speed in such a storm is less than 17 m s^{-1}, the term tropical depression is used. Once the winds reach 17 m s^{-1} and until the sustained wind reaches 33 m s^{-1}, the term tropical storm is used. When the sustained wind exceeds 33 m s^{-1}, the storms are called hurricanes in the North Atlantic, Northeast Paci c (east of the international dateline) and in the Southeast Paci c (east of 160°E).

The term *typhoon* applies in the Northwest Pacific and storms in the Southwest Pacific, Southeast Indian ocean are called *severe tropical cyclones*. The term severe cyclonic storm is applied in the North Indian Ocean, and in the southwestern Indian Ocean the term *tropical cyclone* is used. It is a popular myth that TCs are referred to as *willy-willys* in Australia, but this is not the case (the term applies to dust devils). The only term in common use in Australia is tropical cyclone. The spectrum of storms is named differently in all six tropical cyclone basins, and in the North Indian Ocean the names for different storm intensities vary by country.

Despite this diversity in naming conventions, the same processes occur in all these storms and, with the exception of the fact that cyclonic rotation in the Northern Hemisphere is anti-clockwise and is clockwise in the Southern Hemisphere, the following discussion applies to all such storms. The concept of sustained surface wind varies somewhat between agencies. The World Meteorological Organization (WMO) suggests use of a 10-minute average and this is utilized by most countries. However, the Joint Typhoon Warning center in Guam utilizes one minute winds, which results in some discrepancies when comparing statistics between regions (usually referred to as basins).

ORIGIN OF NAME

The origin of the regional names is not a modern imposition. The word *hurricane* derives originally from the Mayan god, Hurakan, who blew his breath across the chaotic water and created the dry land. The word *typhoon* is believed to have come from the Mandarin word t ai fung, meaning great wind. The term *kamikaze* also derives from a 14th century typhoon, which devastated a Mongol fleet that was about to invade Japan. The kamikaze, or divine wind, was therefore a source of great fortune for Japan. The only term in common use in Australia is tropical cyclone. The term super-typhoon is used in the Northwest Pacific to refer to typhoons that have a maximum sustained wind speed of at least 65 m s^{-1}. This is the equivalent of a category four or five hurricane in the North Atlantic or a category five cyclone in the Australian region (Table 1). The term

Table 1 The Saffir/Simpson scale used to categorize North Atlantic hurricanes compared with the Australian scale developed by the Australian Bureau of Meteorology showing approximate correspondence of the categories

Classification and Saffir/Simpson Scale	Effects	Australian category definitions
Tropical depression winds: $<17\,m\,s^{-1}$	Negligible house damage. Damage to some crops, trees and caravans	
Tropical storm winds: $17–32\,m\,s^{-1}$		Category 1: winds $17–25\,m\,s^{-1}$
	Minor house damage, significant damage to signs, trees caravans. Heavy crop damage. Risk of power failure, small craft may break moorings	Category 2: winds $26–34\,m\,s^{-1}$
Category 1: winds $33–42\,m\,s^{-1}$	No real damage to building structures. Damage primarily to unanchored mobile homes, shrubbery, and trees. Some small craft in exposed moorings may be torn free	Category 3: winds $35–45\,m\,s^{-1}$
Category 2: winds $43–49\,m\,s^{-1}$	Some roofing material, door, and window damage. Considerable damage to foliage and some trees blown down. mobile homes, etc. Flooding damages piers and small craft in unprotected moorings may break their moorings	Category 4: winds $46–62.5\,m\,s^{-1}$
Category 3: winds $50–58\,m\,s^{-1}$	Some structural damage to small residences and utility buildings, with a minor amount of curtainwall failures. Foliage removed from trees and large trees blown down. Mobile homes are destroyed. Flooding near the coast destroys smaller structures with larger structures damaged by floating debris. Terrain may be flooded well inland	
Category 4: winds $59–69\,m\,s^{-1}$	More extensive curtain wall failures with some complete roof structure failure on small residences. Complete destruction of mobile homes. Major erosion of beach areas. Terrain may be flooded well inland	Category 5: winds $63+\,m\,s^{-1}$
Category 5: winds $70+\,m\,s^{-1}$	Complete roof failure on many residences and industrial buildings. Some complete building failures with small utility buildings blown over or away. Flooding causes major damage to lower floors of all structures near the shoreline. Massive evacuation of residential areas may be required	

cyclone was itself coined in 1855 by Henry Piddington, a 19th century British meteorologist, from the Greek word describing the coil of a snake, recalling the characteristic pattern of flow towards the center of the storm. Cyclones also occur outside the tropics but these storms are different in character from TCs. TCs are driven by energy from the warm tropical ocean. The energy source for extra-TCs or mid-latitude cyclones comes from the strong latitudinal temperature gradient that exists in these regions.

NAMING OF INDIVIDUAL STORMS

The naming of storms seems to have arisen as a matter of convenience for weather forecasters. The attachment of a single label to a storm means that the storm can be unambiguously tracked throughout its life cycle. The first use of a proper name for a TCs is attributed to Clement Wragge, an Australian meteorologist. An informal naming system operated during World War II in the Pacific and from 1952, storms in the North Atlantic were given names from the phonetic alphabet (alpha, baker, charlie ... etc.). Women s names replaced these in 1953. The current convention is for a pre-arranged list of male and female names, which are created for each basin. Some basins (such as the North Atlantic) start at A each year, whereas others cycle through the list continuously. Where a storm causes significant damage, deaths or distress, the name is retired. Examples of retired names are Camille (1969), Andrew (1992) and Tracy (1974). North Indian ocean cyclones are not named. Despite occasional criticism of the practice, it is well established and understood by both meteorologists and the general public. Table 2 shows the lists of proposed cyclone names for the North Atlantic for 2003 and 2004.

Table 2 Lists of proposed North Atlantic storm names for 2003 and 2004. Naming of storms reduces confusion when multiple storms occur at the same time. Storm names are usually retired when a storm causes significant damage or loss of life (e.g., Andrew, Camille, Tracy)

2003	2004
Ana	Alex
Bill	Bonnie
Claudette	Charley
Danny	Danielle
Erika	Earl
Fabian	Frances
Grace	Gaston
Henri	Hermine
Isabel	Ivan
Juan	Jeanne
Kate	Karl
Larry	Lisa
Mindy	Matthew
Nicholas	Nicole
Odette	Otto
Peter	Paula
Rose	Richard
Sam	Shary
Teresa	Tomas
Victor	Virginie
Wanda	Walter

CONDITIONS FOR TROPICAL CYCLONE DEVELOPMENT

There is a popular conception that warm ocean temperatures are the prime genesis factor for TCs. While these are important, they are only one of a number of conditions that are necessary for the development of a tropical cyclone. The first three parameters can be said to define the dynamic potential for storm occurrence and the last three define the thermodynamic potential. These necessary conditions include:

- a pre-existing weather disturbance, which provides a means for extracting energy from the ocean;
- warm tropical oceans with temperatures of at least 26.5 °C that provide a source of energy for the storm;
- a minimum distance of at least 500 km from the equator to ensure that a rotating storm can develop;
- suitable temperature and moisture distribution within the troposphere; and
- relatively light winds aloft to ensure that the vertical structure of the storm is maintained.

Even the achievement of all these conditions is not sufficient for the development of a tropical cyclone. We know this because these conditions can be satisfied without any cyclone forming. The exact processes that result in the genesis of a tropical cyclone are still a topic of active research.

The areas of the globe in which TCs occur is indicated in Figure 1. Some areas of the tropical ocean virtually never experience storms intense enough to be described as TCs. The distribution also makes it clear that cyclonic storms never cross the equator. TCs are a seasonal occurrence with most basins having a maximum frequency during late summer to early autumn. Frequency of occurrence of TCs by basin is presented in Table 3.

STRUCTURE OF TYPHOONS

The most distinctive feature of a typhoon or tropical cyclone is the eye (visible in Figure 2). The eye is an approximately circular area of light winds found at the center of a severe tropical cyclone. Complete calm exists at the center of rotation of the storm, but significant winds can be found away from the center of the eye. In some circumstances the eye can be cloud free (in which case it is very easy to locate on satellite imagery) and stars or blue sky may be visible to an observer at the surface. In other cases, the eye may be covered by high cirrus clouds. The eye is surrounded by a ring

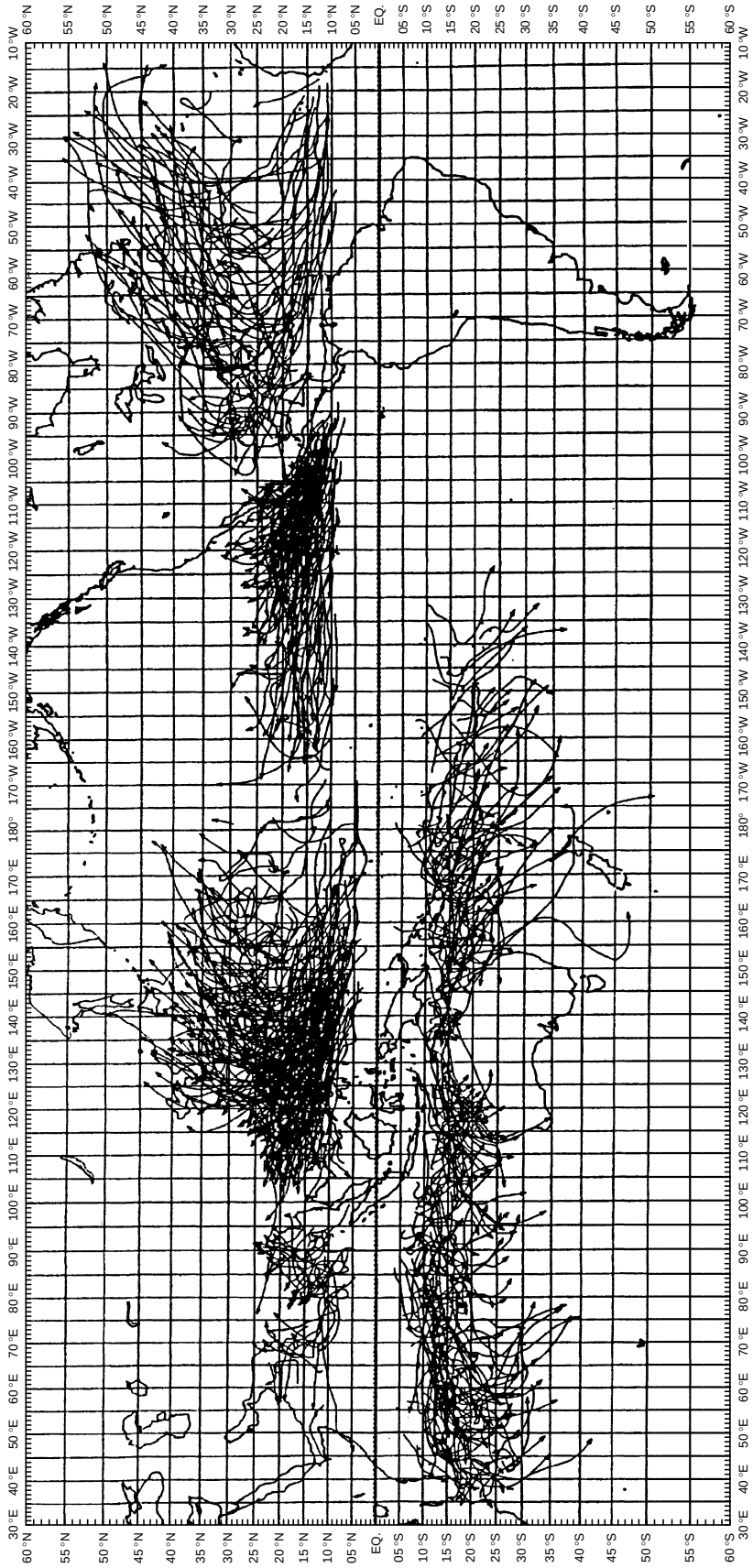


Figure 1 Global map showing regions of tropical cyclone occurrence (World Meteorological Organization, 1993). The map shows tracks of storms with maximum winds $>17.5\,m\,s^{-1}$ for the period 1979–1988

Table 3 Averaged annual total numbers of TC (wind at least 17 ms^{-1}) and intense TC (wind at least 33 ms^{-1}) and their standard deviations for all tropical cyclone basins (number per year)[a]

		North Atlantic Basin	East North Pacific Basin	North Indian Basin	South-west Indian Basin	South-west Pacific Basin	North-west Pacific Basin	Totals
Mean	TC	9.3	17.8	5.2	10.6	16.4	26.8	86.1
	Intense TC	5.0	10.3	2.0	4.8	7.5	16.4	45.9
Standard deviation	TC	2.6	4.7	2.2	3.2	4.6	3.9	7.9
	Intense TC	1.7	3.5	1.9	2.6	2.6	3.4	7.0

[a] Data are retrieved from US National Climate Data Center data for 1970–1995 (Henderson-Sellers *et al.*, 1998)

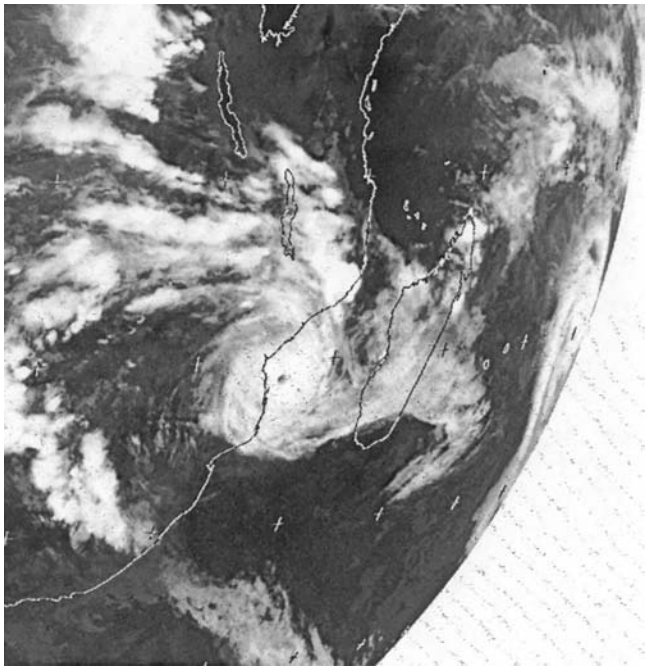


Figure 2 Meteosat image of tropical cyclone Eline as it approaches the coast of Mozambique on 21st February, 2000. Earlier rains had already swollen rivers and displaced 200 000 people. The central clear eye is visible in this image. (Image courtesy of University of Nottingham)

of deep convection known as the eye wall, the region of most destructive winds. Beyond this central core, the storm tends to be organized into spiral rainbands, which are coherent bands of convective storms that extend to the periphery of the storm and well away from the region of maximum winds. These rainbands are important in the maintenance of the central core of the storm as they consist of convective cells that move (spiral) in towards the eye. Typhoon eyes range in size and structure and are not necessarily exactly circular. They are typically 30–60 km across, though they can be as small as 8 km or as large as 200 km. If the cyclone is particularly intense, it may exhibit a double eye wall. The level of observational detail that is required to monitor the changes in the nature of the eye wall is considerable, and, because these storms originate and develop over the ocean, detailed observations are not readily available for all cyclones. As a consequence, the behavior of the eye wall, which is generally the most destructive region of the storm, is still a topic of active research.

DEVELOPMENT AND MOVEMENT OF TROPICAL STORMS

Typically, a tropical cyclone will develop from a tropical depression and progress through more intense stages, although the rate of development can vary quite considerably. In the North Atlantic, TCs typically begin their life in atmospheric wave disturbances (about 60% of all hurricanes and 85% of intense or severe hurricanes). In other regions, cyclones are spawned in monsoon troughs and associated with mesoscale convective clusters that provide the pre-existing weather disturbance for the cyclone to form. The exact course of an individual storm is difficult to predict, as tropical storms move and develop in response to a complex array of forcings. The storm movement follows the overall environmental flow in the troposphere, but interactions with mid-latitude weather systems such as cold fronts and upper level jet streams act to deflect this otherwise smooth motion in a relatively unpredictable manner. The storms rely on a steady supply of energy supplied by the evaporation from the warm tropical ocean surface. As the storms move over cooler water or over land, they rapidly decrease in intensity. As the storms move polewards they carry significant amounts of moisture to higher latitudes, which often results in heavy rain and flooding in these extratropical locations.

MEASURING AND OBSERVING TROPICAL STORM CHARACTERISTICS

Each year, approximately 80–90 TCs occur around the globe. The level of variability in this annual total is around 10%. For individual basins, the variation can be much larger and no obvious correlation exists between individual

basins. Since the advent of effective satellite observation of typhoons, cyclones and hurricanes, our ability to forecast the behavior of storms has increased. As a consequence, the number of deaths that have occurred due to TCs has decreased dramatically, particularly on the high seas. Since 1944, regular aerial reconnaissance of hurricanes has been undertaken in the North Atlantic; aircraft flights into these storms do not occur in any other basins on a regular basis. Such aircraft reconnaissance is extremely expensive (and potentially dangerous), but provides valuable information about the nature and position of the storm that can be fed into numerical models to predict the future track and development. Conventional satellite imagery provides useful information on position and intensity. The Dvorak technique is an empirical technique that uses pattern matching to estimate the intensity of a storm from the patterns of clouds visible on satellite images. Observations using both passive microwave sensors and windfields derived from satellite mounted radar scatterometers provide additional information on the nature and position of the storm. The benefits of enhanced communications such as the World Wide Web now mean that advanced satellite products are available in real time to all forecasting centers with minimal additional investment.

IMPLICATIONS OF CLIMATE CHANGE

The instrumental record of relevance to interannual variability and trends in tropical cyclone, hurricane and typhoon activity is rather short and extensive analysis is therefore difficult. The quality of databases on tropical cyclone activity is highly variable. Even within a single basin, changes in observational networks, techniques and definitions may produce biases and errors in the datasets that have implications for studies of natural variability and trends in cyclone activity.

The balance of evidence collected by the Intergovernmental Panel on Climate Change (IPCC) suggests that the increases in greenhouse gas concentrations in the atmosphere are causing detectable climate change. A number of factors will offer improved understanding of the impacts of global change on tropical cyclone activity.

- Increased realism in three dimensional coupled climate models.
- Improved observations of air–sea interaction and other aspects of storm genesis and evolution.
- Further studies of past changes through palaeo-evidence.

Human vulnerability to typhoons, hurricanes and TCs is increasing as the population of coastal regions increases. At the same time, the costs of tropical cyclone impacts are increasing as infrastructure and insurance costs increase. Concerns about the impacts of global climate change therefore relate to many aspects of the storms: the frequency of occurrence; intensity (mean and maximum); structure and rainfall patterns. We still do not know how to predict the maximum intensity of today s storms, nor do we have a complete predictive model of storm genesis.

Although datasets are limited, some analysis of trends and variability has been undertaken at a regional scale. The number of typhoons in the Northwest Pacific showed a steady increase from 1980–1996; however, this period followed a decrease of almost identical magnitude in the previous 20 years. Difficulties in obtaining consistent wind measurements from the historical database have limited comparative studies of intensity. In the North Atlantic, there is substantial year to year variability, but no significant trend.

SUMMARY

Observational networks and technology have improved substantially over the last 40–50 years with a significant reduction in the loss of life resulting from improved warning to shipping and coastal communities. Changes in global climate in response to increased greenhouse gases have the potential to affect the frequency of occurrence, intensity and distribution of these storms. Because a number of factors influence the genesis of TCs, the exact influence of global change remains unclear. Although sea surface temperatures are predicted to rise, other changes in the structure of the atmosphere are likely to have a mitigating effect on storm formation and intensity and the 26.5 °C threshold is thought likely to rise under greenhouse conditions. Thermodynamic modeling studies of the effects of doubling the carbon dioxide concentration have indicated that there is likely to be a moderate increase in maximum potential intensity of storms; evidence of increased area of occurrence derived from regional scale models, however, is tentative and preliminary. There is no certainty that a storm will reach its thermodynamic maximum potential intensity and there are many uncertainties in other aspects of storm structure and behavior, particularly in the way the storm exchanges energy with the ocean.

REFERENCES

Henderson-Sellers, A, Zhang, H, Berz, G, Emanuel, K, Gray, W, Landsea, C, Holland, V, Lighthill, J, Sheih, S-L, Webster, P J, and McGuffie, K (1998) Tropical Cyclones and Climate Change: a Post-IPCC Assessment, *Bull. Am. Meteorol. Soc.*, 79, 19–38.

World Meteorological Organization (1993) *Global Guide to Tropical Cyclone Forecasting*, Tropical Cyclone Program Report No. TCP-31, WMO/TD No. 560, World Meteorological Organization, Geneva.

FURTHER READING

Australian Bureau of Meteorology: http://www.bom.gov.au.

Bunbury, B (1994) *Cyclone Tracy: Picking Up the Pieces*, Freemantle Arts Center Press, South Fremantle, 148.

Chris Landsea s Hurricane FAQ: http://www.aoml.noaa.gov/hrd/tcfaq/tcfaqHED.html.
Diaz, H F and Pulwarty, R S (1997) *Hurricanes: Climate and Socioeconomic Impacts*, Springer-Verlag, Berlin, 292.
Hong Kong Observatory: http://www.info.gov.hk/hko/index.htm.
Joint Typhoon Warning Center: http://www.npmoc.navy.mil/.
Simpson, R H and Riehl, H (1981) *The Hurricane and its Impact*. Louisiana State University Press, Baton Rouge, LA.
US National Hurricane Center: http://www.nhc.noaa.gov.

Hurricanes, Typhoons and other Tropical Storms – Dynamics and Intensity

Jenni L Evans
The Pennsylvania State University, University Park, PA, USA

Tropical cyclones are the most destructive natural phenomenon on Earth. Winds, rain, river ooding, mud slides, coastal inundation and erosion contribute to loss of life and property due to these storms. In the aftermath of a storm moving inland from the sea (making landfall), disease outbreaks are a constant danger due to contamination of the local fresh water supply (from brackish water contamination) and the prevalence of standing water (creating new breeding areas for disease vectors such as mosquitoes).

An overview of current understanding of tropical cyclone formation (cyclogenesis), seasonal and interannual variability, intensity change, and storm motion is presented. This provides a background from which to infer potential changes in tropical cyclone characteristics in a warmer climate. Such changes in tropical cyclones in a warmer climate are strongly tied to changes in the tropospheric wind, temperature and moisture structure of the global tropics. The more sensational claims of increased tropical cyclone occurrence are not supported and, while the strongest storms may increase in intensity by 5- 10%, evidence is lacking for the effect on average storms. Further, a 5- 10% difference is within the range of interannual variability in the present climate and so will not be readily detected initially. While global storm statistics may be largely unchanged, recent evidence of a tendency for a more El Niño/Southern Oscillation-like structure in warmer climate regimes suggests regional changes in tropical cyclone frequency and possibly in preferred storm tracks.

INTRODUCTION

Tropical cyclones are clearly the most destructive natural phenomenon on Earth, visiting death and destruction across the global tropics each year. As a tropical cyclone passes by, the winds and rain become stronger and stronger until they die off rapidly as the eye of the storm passes over. This brings a time of deceptive calm (often accompanied by blue skies). Then, as quickly as they ceased, the winds and rain return and the storm may rage for hours more. To anyone who has experienced or witnessed the aftermath of one of these storms, they stand out as terrifying examples of the power of weather. Satellite images (e.g., Figure 1) illustrate many aspects of the structure of a tropical cyclone: the central, calm and (usually) cloud free eye region surrounded by an eyewall of deep thunderstorms and bands of cloudiness that spiral outward from the eyewall (although these are sometimes obscured by high clouds). These tropical low pressure systems are typically 12–15 km tall and the strongest winds may extend from a few kilometers to hundreds of kilometers from the center.

While we have only named individual storms since the 1960s, this storm type has different names originating in the folklore of the various regions in which it occurs: *hurricane* in the Atlantic and *typhoon* in the western Pacific. These terms are used to designate the strongest of these storms in each region. A tropical cyclone begins life as a depression. Each depression is tracked, but it is not a named tropical storm until its intensity (area averaged sustained surface winds over its center or minimum pressure) exceeds $17\,\mathrm{m\,s^{-1}}$ (with stronger gusts). Once the sustained winds are above $34\,\mathrm{m\,s^{-1}}$, a tropical storm becomes a tropical cyclone, and at $67\,\mathrm{m\,s^{-1}}$, it becomes a Category one hurricane or typhoon. There are five hurricane (typhoon) categories, with storms in Category three, four or five being known as severe tropical cyclones (intense hurricane in the Atlantic or supertyphoon in the Western North Pacific). Threshold values for these categories vary slightly between the different basins, as does the time over which the sustained winds are recorded (one minute in the Atlantic and eastern Pacific, 10 minutes elsewhere) (*see* Hurricanes, Typhoons and other Tropical Storms - Descriptive Overview, Volume 1).

TROPICAL CYCLONE CLIMATOLOGY

Annual global tropical cyclone activity is remarkably stable. Approximately eighty tropical storms are observed around the globe each year and of these, two thirds reach hurricane (typhoon) intensity (Figure 2). The average annual variation of global occurrence is only $\pm 7\%$, although regional variations are much wider and are largely uncorrelated from one region to another.

Tropical storms exhibit a definite seasonal variation. The season extends from June to November in the Northern

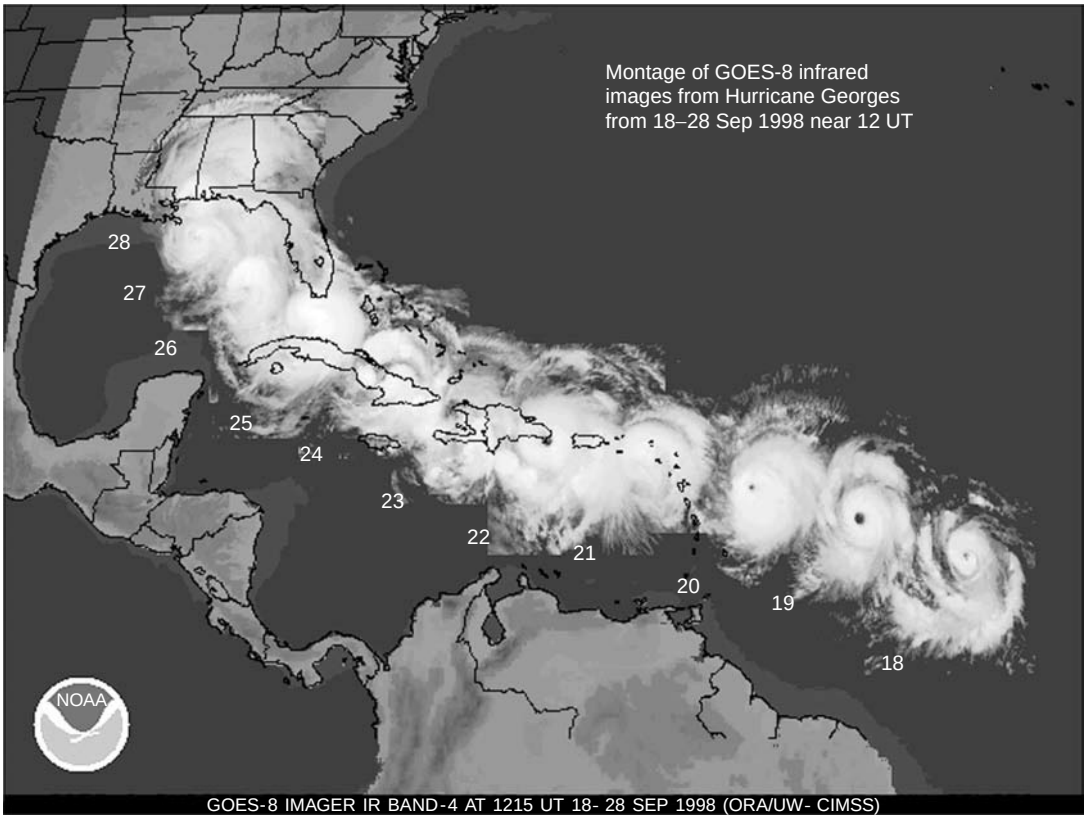


Figure 1 Montage of infrared (IR) satellite images of Hurricane Georges (1998). The data are from the GOES-8 satellite and cover the period 18–28 September 1998. This image was provided by Gary Wade at NOAA/NESDIS/ORA. Between 18 and 28 September 1998, Georges progressed from a minimal hurricane (984 hPa on 18 September) to a peak intensity of 937 hPa (Cat-4 Hurricane; 20 September) and eventually decayed after making landfall about 1130 UTC on 28 September 1998 near Biloxi, MI. After reaching its peak intensity on the 20th, Georges crossed a series of Caribbean islands on the 21st: Antigua (0430 UTC), St Kitts (0800 UTC), Puerto Rico (2200 UTC), and the Dominican Republic early on the following morning (0430 UTC). Hurricane Georges then proceeded towards Cuba, making landfall on the 23 September at 2130 UTC, before crossing Key West as a minimal hurricane (25 September at 1530 UTC). The variations in the storm intensity are generally evident in the variations of the cloud signature

Hemisphere (NH) and December to May in the Southern Hemisphere (SH). The western North Pacific basin is the only region in which tropical cyclones have been observed in all months of the year. Peak tropical cyclone activity occurs in late summer/early autumn in each hemisphere (August–October in the NH and January–March in the SH). This seasonal bias coincides with the maximum in regional sea surface temperatures (SSTs) (exceeding the critical lower limit of 26.5 °C) and favorable large-scale winds.

CONDITIONS OBSERVED TO BE NECESSARY FOR TROPICAL CYCLONE FORMATION

Formation of tropical cyclones (tropical cyclogenesis) occurs over oceans in preferred regions: the poleward side of the monsoon trough in the Pacific and Indian Ocean basins and easterly waves in the North Atlantic and eastern North Pacific basins (Gray, 1968) (Figure 3). Tropical cyclones are not observed in the South Atlantic or eastern South Pacific basins.

A number of temperature and moisture (thermodynamic) and windfield (dynamic) conditions have been observed to be necessary for tropical cyclone formation, however no single condition has been determined to be sufficient. That is, all of the factors described below could be present and yet a tropical cyclone may not form, but a tropical cyclone will not form if any of these factors is missing. This lack of predictability of tropical cyclone formation makes it difficult to study because, by the time we re sure that a storm is forming, it already has!

Thermodynamic (temperature or moisture) factors favorable for tropical cyclogenesis include warm SST over the oceanic mixed layer, high humidity values well above the surface (5–7 km) and the potential for tall thunderstorm growth (Gray, 1968). In the present day climate, warm oceans correspond to SST $\geq$26.5 °C; areas of deep upward motion (also required for thunderstorms to grow) in the

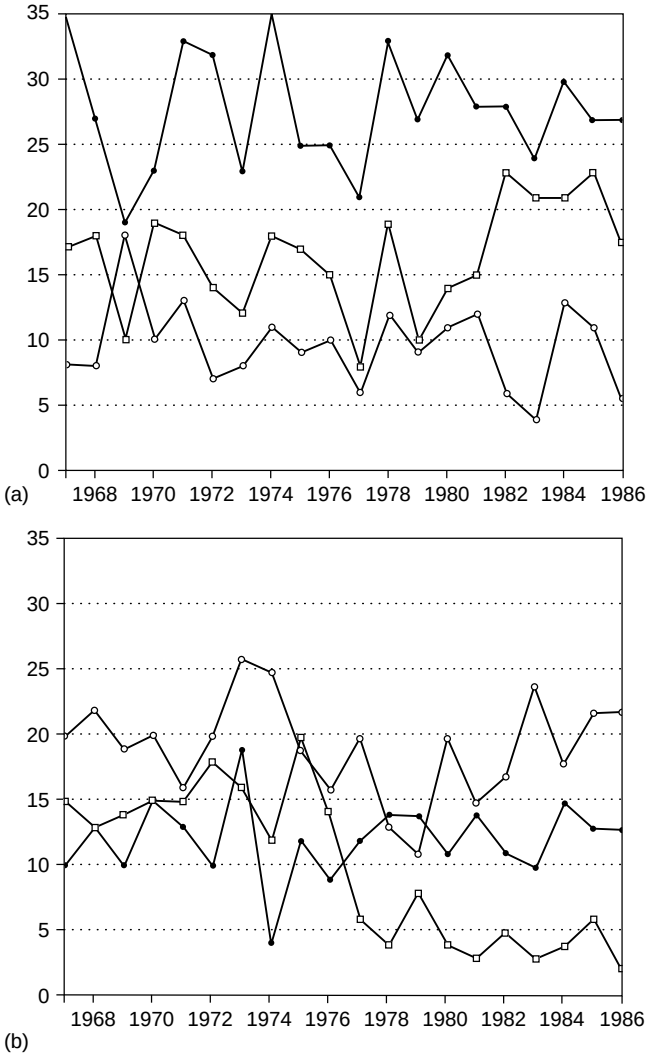


Figure 2 Annual tropical cyclone activity in each basin from 1967 to 1986: (a) North Atlantic (●), western North Pacific (○) and eastern North Pacific (□); and (b) north (□) and south (●) Indian Oceans and South Pacific–Australian region (○)

tropics are always located over waters with SST $\geq 26.5\,^{\circ}C$, so it is possible that this SST threshold condition is tied to the convection. This potential link becomes important when projecting tropical cyclone activity into the future as the climate regime changes.

McBride (1981) and others observe that these thermodynamic conditions are common in the tropical atmosphere, especially during the tropical cyclone season. McBride compared developing systems with non-developing ones that had been deemed likely to become tropical cyclones. The thermal characteristics of the developing and non-developing systems were essentially the same, leading McBride (1981) to conclude that these thermodynamic conditions were seasonal indicators of the likelihood of tropical cyclogenesis rather than indicators that an individual event was about to occur. This result is also important when considering tropical cyclogenesis in differing climate regimes.

Because the thermodynamic requirements for tropical cyclogenesis have been shown to be present throughout the cyclone season, the day to day variability of cyclone activity must be related to the dynamics of the genesis process (Gray, 1979).

The dynamic (wind related) factors associated with tropical cyclogenesis have been identified as locally enhanced, near-surface rotation (absolute vorticity) and small changes in the horizontal wind with height (vertical wind shear) over the center of the growing storm (Gray, 1968, 1979; and numerous others). Included in the absolute vorticity requirement is the condition that the system should not be too close to the equator (even though this is the region of warmest SST) because Earth rotation contributes to its development. In the Atlantic, easterly waves provide a dynamical seed for tropical cyclogenesis. In all other basins, the most dynamically favorable area for tropical cyclogenesis is in the region just poleward of the equatorial monsoon trough. (Gray, 1968). The monsoon trough is a region of both small vertical wind shear and enhanced near surface absolute vorticity; to both its north and south, the vertical and horizontal wind shear increases.

Watterson *et al.* (1996) used the presence of these thermodynamic and dynamic factors to diagnose tropical cyclogenesis from 10 years of gridded observations. Broad basins of tropical cyclogenesis were correctly identified, but the combination of these conditions did not isolate tropical cyclogenesis locations and neither did they replicate interannual variability. This study indicated that, if tropical cyclogenesis is deterministic, some undetected factor must also be present. Studies of the 30–60 day variation of tropical convection within a season (known as the Madden Julian Oscillation (*see* Madden- Julian Oscillation, Volume 1); Madden and Julian, 1972) indicate that inclusion of this ocean-scale modulation of the atmospheric winds and convection contributes to the likelihood of tropical cyclogenesis (Hendon and Liebmann, 1996).

In summary, the large-scale factors necessary for tropical cyclogenesis are present for the majority of the cyclone season. While they must be present for tropical cyclogenesis to occur, their presence does not guarantee that a storm will form.

INTERANNUAL FLUCTUATIONS IN TROPICAL CYCLONE NUMBERS

Interannual variability in any one of the cyclone regions is large, in spite of the overall global stability of cyclone frequency (Figure 2). Variations of two large-scale atmospheric and oceanic oscillations, the El Niño/Southern

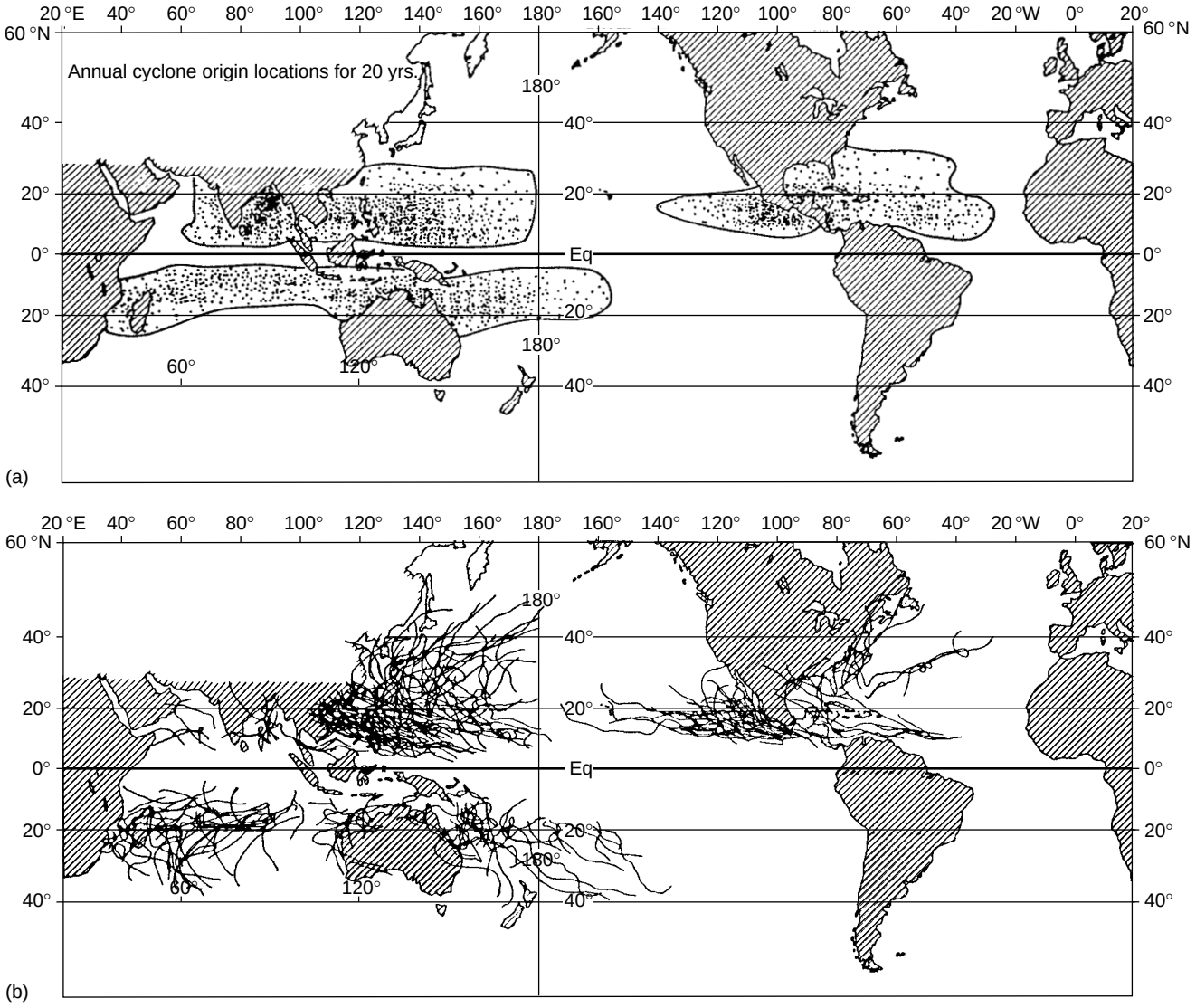


Figure 3 Global tropical cyclones: (a) genesis locations for a 20-year period and (b) tracks for a 3-year period. (Reproduced from Gray, 1979)

Oscillation (ENSO) and the Quasi-biennial Oscillation (QBO) (Gray, 1984; Shapiro, 1989), explain much of this variability (*see* El Niño and La Niña: Causes and Global Consequences, Volume 1).

Strong and significant links between tropical cyclone activity and the phase of ENSO have been detected in a number of ocean basins. In the western North Pacific and Australian regions, tropical cyclone activity is suppressed in stronger ENSO years and increased in the opposite extreme of anti-ENSO (Chan, 1985; Nicholls, 1989). Evans and Allan (1992) compiled the relative distributions of tropical cyclone occurrence in the Australian region for five ENSO and five anti-ENSO storm seasons. On the western Australian coast, tropical cyclogenesis extends further into the Indian Ocean in ENSO years compared to other seasons. This spread of tropical cyclogenesis away from the coast in an ENSO year is also evident in eastern Australia. This shift of storm activity results in decreased danger of tropical cyclone damage along the eastern Australian coast, while islands further out in the Pacific face an increased risk of cyclone strike. Similarly, ENSO enhances tropical cyclone activity in the central and eastern North Pacific (Chan, 1985).

ENSO effects on tropical cyclones in the eastern Pacific and Atlantic basins are out of phase: ENSO brings decreasing storm numbers in the Atlantic and increases in the eastern Pacific (Gray, 1984).

ENSO events occur at spacings of 2–10 years. The cause (or causes) of ENSO are still being investigated, but bursts of westerly (eastward) winds in the equatorial western

Pacific region just prior to the onset of an El Niño cycle have been well documented. Tropical cyclone activity close to the equator has been observed in advance of ENSO onset, suggesting that a tropical cyclones may enhance the equatorial westerly winds, providing one trigger for ENSO initiation. Other investigators argue that ENSO is purely ocean driven. Hence, while we have some understanding of the role of ENSO in moderating tropical cyclone variability, the question of whether tropical cyclones modulate ENSO remains open.

TROPICAL CYCLONE INTENSITY CHANGE

Tropical cyclone intensity is defined in terms of either the minimum central pressure or the maximum sustained surface wind of the storm. An intensity increase (increasing rotating wind and decreasing central pressure) is referred to as intensification or deepening; an intensity decrease (decreasing wind and increasing pressure) is referred to as weakening. Although heavy rainfall causes much damage associated with tropical cyclones, rainfall is not used when evaluating storm intensity.

Theories for Tropical Cyclone Intensification

Theories of intensification address limits on tropical cyclone intensity in the absence of adverse impacts of other weather systems, which are supposed to arrest intensification or even to cause the storm to decay. These theories predict the potential intensity (PI) of a tropical cyclone, which is the theoretical limit of tropical cyclone intensity if the environmental conditions are ideal. It is not clear that other weather systems provide only weakening effects on a tropical cyclone, but no existing theory incorporates external weather systems in a positive (intensifying) role.

Theories proposed to explain tropical storm intensification have key common elements. Heating due to a moist energy source is released into the atmosphere via deep thunderstorms. Because the storm traps the heat locally, it warms the air above the storm center and forces some air laterally out of the top of the column, dropping the surface pressure and increasing the surface winds. The stronger surface winds resulting from this intensification bring more moist energy into the storm and the cycle of intensification continues (Emanuel, 1999; Holland, 1997). The candidates for the moist energy source include atmospheric moisture convergence or ocean surface fluxes; air forced downward (subsiding) in the tropical cyclone eye is relatively dry, but also contributes to the atmospheric warming and the ultimate central pressure drop.

Observed Limits on Tropical Cyclone Intensification

A number of observationally based studies (e.g., Miller, 1958; Merrill, 1988; Evans, 1993; Emanuel, 1999) have examined the relationship between tropical cyclone intensity evolution and its environment. Miller (1958) and Merrill (1988) related the minimum probable central pressure for a hurricane to its underlying SST and its synoptic environment at 200 hPa. Miller (1958) concluded that maximum deepening is achieved with maximum outflow and minimum inflow at 200 hPa over water warmer than 26 °C. Merrill (1988) concurred, noting that while SST provides an upper bound, adverse environmental effects such as strong vertical wind shear across the storm center and may prevent a storm attaining its PI. Thus, the wind field surrounding the tropical cyclone has either a neutral or negative (weakening) effect on tropical cyclone intensification. Emanuel (1999) considered both SST and outflow (top of storm) temperatures to be critical in limiting a storm PI.

Evans (1993) examined the relationship between tropical cyclone intensity and the underlying SST for storms in all ocean basins. Scatter plots of storm intensity versus SST in the western North Pacific and North Atlantic basins for the 20 year period 1967–1986 are reproduced in Figure 4. Clearly, the highest intensities are achieved over tropical waters (SST $\geq$26.5 °C), although tropical storm intensities are also achieved over cooler waters (usually due to weakening storms moving out of the tropics). An alternative perspective can be gained by stratifying each observation by its SST, then compiling frequency distributions of observed storm intensity within these SST bins. When the entire lifecycle of each storm is considered, this analysis reveals that storms are likely to be the strongest over cooler, tropical water (e.g., stronger over 27 °C compared to 30 °C; Figure 5a). This is evidence of the evolving storm life cycle, because the mature storms move over cooler waters as they move further out of the tropics. If only the peak intensity of each storm is included, no discernible pattern relating SST and intensity emerges (Figure 5b). Clearly, SST does not override environmental interactions in determining the peak intensity achieved by an individual tropical cyclone.

Observed and theoretical peak intensities for the majority of storms occurring in the North Atlantic and western North Pacific basins over more than 40 years are analyzed by Emanuel (2000). He finds the cumulative distribution function for normalized peak winds of these storms to be linear, implying that there is a nearly equal likelihood that any given storm will achieve any intensity up to the potential intensity (PI). Further, the normalized lifecycles of Pacific and Atlantic storms (for seven days centered on the time of peak observed storm intensity) are remarkably similar.

In summary, observations demonstrate that a theoretical PI is rarely, if ever, reached by real storms (Figure 4). PI provides a useful indicator of the maximum devastation possible in a given climate regime, but does not provide guidance on the distribution of actual storm intensities likely in a given region.

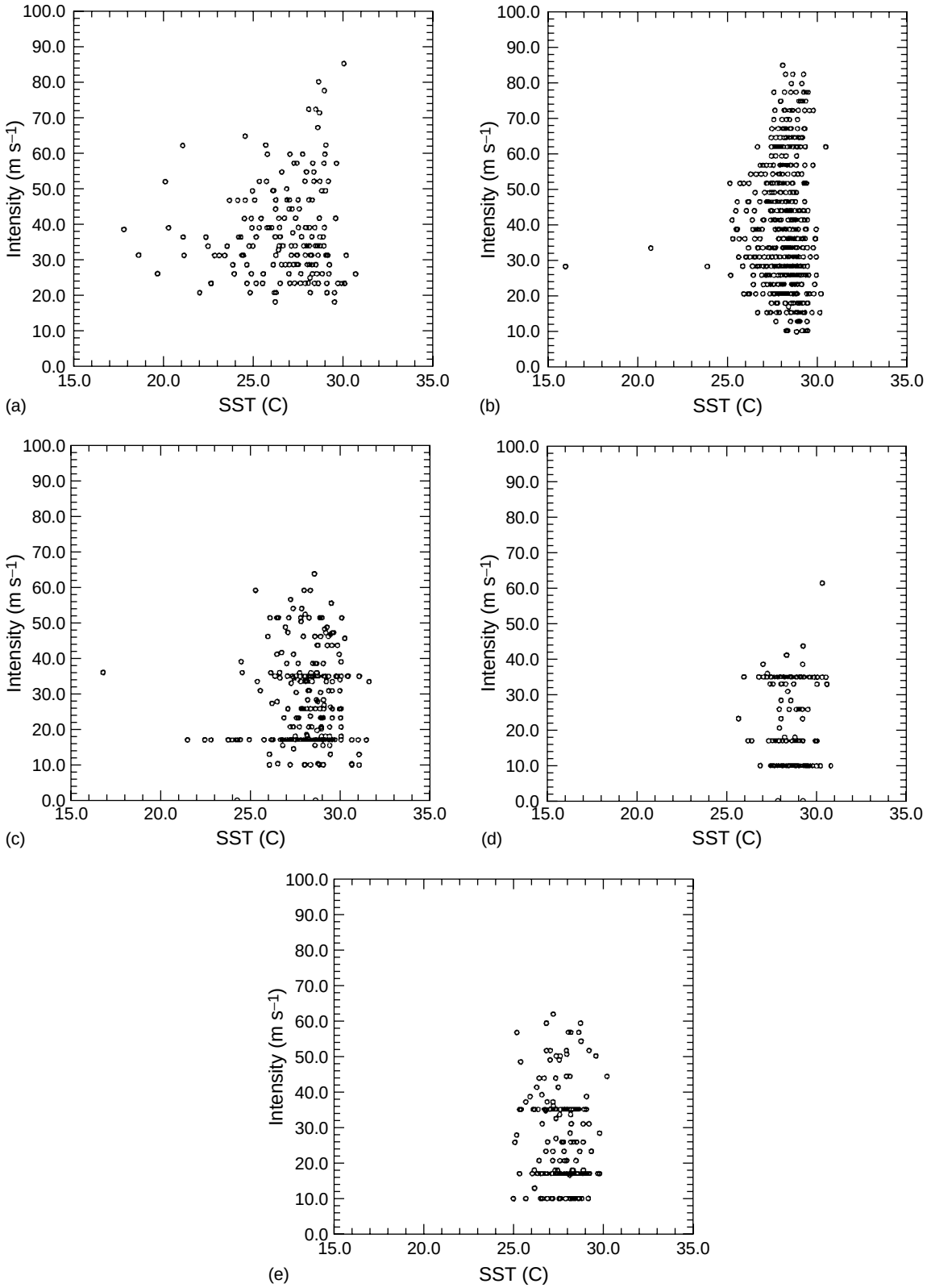


Figure 4 Empirically-derived relationship between SST and the observed maximum intensity over the lifetime of each tropical storm in (a) the North Atlantic, (b) the western North Pacific, (c) the South Pacific–Australian Region; (d) the north and (e) the south Indian Oceans. (Reproduced from Evans, 1993, Figure 4)

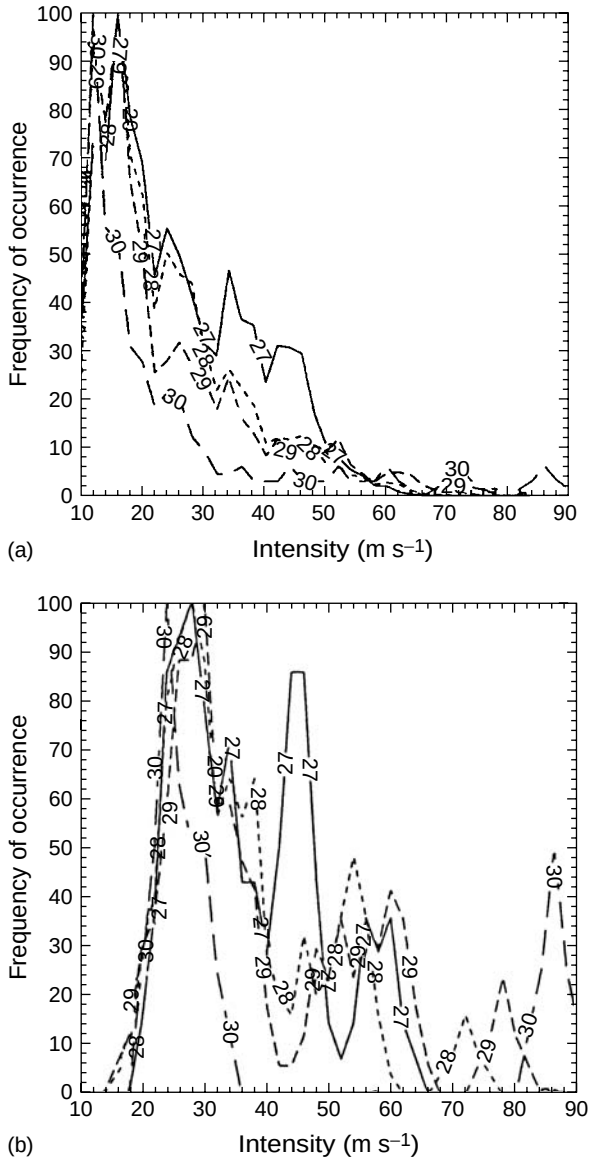


Figure 5 Frequency distribution of storm intensity as a function of SST (in 0.5 °C bins) for (a) all observed times for all storms and (b) the peak intensity for each storm. Atlantic tropical storm data for the period 1967–1986 are used here, but the results hold in other basins

TROPICAL CYCLONE MOTION

Early investigators believed that the mean horizontal wind averaged through the depth of the system provided reasonable estimates of the motion of a tropical cyclone (Helmholtz, 1867; Bjerknes, 1921). For this reason, it is referred to as the steering flow. However, there is a consistent discrepancy between the movement of a tropical cyclone and the steering flow due to spatial variation in the steering flow (Evans *et al.*, 1991) and to interactions of the storm with the Earth s rotation (Rossby, 1948). Other processes that potentially contribute to tropical cyclone motion are not yet well understood.

POSSIBLE CHANGES IN TROPICAL CYCLONES IN A WARMER CLIMATE

General circulation models (GCM) are the most sophisticated and flexible tool available for sensitivity studies of the climate system. The Atmospheric Model Intercomparison Project (AMIP), the Coupled Model Intercomparison Project (CMIP) and related endeavors have highlighted the strengths and weaknesses of GCM in studies of the Earth s climate. While concerns about model details and uncertainties on the effects of clouds remain, the major features of the present climate are well represented (Meehl *et al.*, 2000a; Easterling *et al.*, 2000).

There are many commonalties between simulations of climates with elevated levels of greenhouse gases by different GCM. Common elements include globally averaged warmer surface temperature, cooler temperatures high above the Earth s surface, and some evidence of more intense rain events (heavy rainfalls) and fewer light showers (Meehl *et al.*, 2000b). Much temporal and regional variability is simulated, with the largest surface warming and upper atmosphere cooling occurring inland and closer to the poles, generally milder winters and warmer nighttime temperatures. Coupled models indicate the possibility of more frequent ENSO-like patterns in both the atmosphere and ocean (Meehl and Washington, 1996).

Changes in the mean climate may be very different from changes in variability (Meehl *et al.*, 2000a). For example, the simulated warmer nighttime temperatures indicate less temperature change between night and day (less variability), but result in an overall warming.

The impacts of these simulated changes in the Earth s climate on tropical cyclones must be deduced based on an understanding of these systems in the present climate (Evans, 1993). Thus, we contemplate the most likely effects of this new, warmer climate regime on tropical cyclones as we now know them.

Mean Genesis Locations and Interannual Variability

Warming of SSTs and increasing moisture in the middle troposphere will extend the area of the tropics for which the thermal conditions for tropical cyclone genesis in the current climate are satisfied. However, the area of the tropical atmosphere that supports deep convection is unlikely to change (Royer *et al.*, 1996; Dutton *et al.*, 2001).

New results from coupled climate models indicate the possibility of a warmer climate that is more ENSO-like. In this scenario, the average tropical atmosphere and ocean circulations look more like the climate extreme that we refer

to as ENSO. Thus, we might expect the mean location of the monsoon trough to move and the distribution of tropical storm numbers among different ocean basins to change. If ENSO-like patterns prevail under these warmer climate conditions, storm numbers in the Atlantic and western Pacific Oceans may decrease and increased storm numbers in the eastern and central Pacific are possible.

Climate models are just beginning to approach the grid spacings necessary to start to resolve the large-scale features of individual tropical cyclones. In one of the first of these studies, Bengtsson *et al.* (1994) diagnose a global decrease in tropical cyclone numbers, by counting individual systems. However, concerns about cloud and ocean flux treatments in this and other GCM render this result preliminary.

Cyclone Intensity

The possibility of increasingly intense tropical cyclones in a warmer environment has been raised by a number of authors. This prediction is largely based on the potential for more intense storms over increasingly warmer SST (Lighthill *et al.*, 1994). Indeed, the modeling studies of Knutson *et al.* (1998) indicate that the theoretical maximum PI of a tropical storm in a warmer climate could increase by 5–10%. Because only a few intensity sensitivity studies are available, however, these data are preliminary.

Empirical studies of tropical cyclones demonstrate that the majority of storms never achieve their PI (e.g., Emanuel, 2000) and because a 5–10% intensity difference is well within the present range of intensity variability, such an increase would be difficult to detect for some time. Thus, while the potential likely exists for the strongest storms to become stronger in a warmer climate, the change in the peak intensity indicated (in the few studies available) is within the present range of interannual variability and so not possible to observe.

Tropical Cyclone Motion

Changes in the large-scale wind pattern could affect the preferred cyclone paths currently observed. These changes would impact rainfall, storm surge and wind damage statistics at storm affected locations. Should a warmer climate regime have more ENSO-like wind structure, the overall storm track pattern would be modified in response to altered steering flow; however, individual tracks will continue to show substantial variation from the average basic state (as they do in the present climate).

REFERENCES

Bengtsson, L, Botzet, M, and Esch, M (1994) Will Greenhouse Gas-induced Warming over the Next 50 Years Lead to a Higher Frequency and Greater Intensity of Tropical Cyclones? *Max-Planck-Institut fur Meteorologie Report No. 139*, Hamburg, August 1994.

Bjerknes, V (1921) On the Dynamics of the Circular Vortex with Applications to the Atmosphere and Atmospheric Vortex and Wave Motions, *Geof. Publ.*, 2, 1–88.

Chan, JC-L (1985) Tropical Cyclone Activity in the Northwest Pacific in Relation to the El Niño–Southern Oscillation Phenomenon, *Mon. Weather Rev.*, 113, 599–606.

Dutton, J F, Poulsen, C J, and Evans, J L (2000) The Effect of Global Climate Change on the Regions of Tropical Convection in CSM1, *Geophys. Res. Lett.*, 27, 3049–3052.

Easterling, D R, Evans, J L, Ya Groisman, P, Karl, T R, Kunkel, K E, and Ambenje, P (2000) Observed Variability and Trends in Extreme Climate Events: A Brief Review, *Bull. Am. Meteorl. Soc.*, 81, 417–425.

Emanuel, K A (1999) Thermodynamic Control of Hurricane Intensity, *Nature*, 401, 665–669.

Emanuel, K A (2000) A Statistical Analysis of Tropical Cyclone Intensity, *Mon. Weather Rev.*, 128, 1139–1152.

Evans, J L (1993) Sensitivity of Tropical Cyclone Intensity to Sea Surface Temperature, *J. Clim.*, 6, 1133–1140.

Evans, J L and Allan, R J (1992) El Niño/Southern Oscillation Modification to the Structure of the Monsoon and Tropical Cyclone Activity in the Australasian Region, *Int. J. Climatol.*, 12, 611–623.

Evans, J L, Holland, G J, and Elsberry, R L (1991) Interactions between a Barotropic Vortex and an Idealized Subtropical Ridge. Part I: Vortex Motion, *J. Atmos. Sci.*, 48, 301–314.

Gray, W M (1968) Global View of the Origin of Tropical Disturbances and Storms, *Mon. Weather Rev.*, 96, 669–700.

Gray, W M (1979) Hurricanes: Their Formation, Structure and Likely Role in the Tropical Circulation, in *Meteorology Over the Tropical Oceans*, ed D B Shaw, Royal Meteorological Society, 155–218.

Gray, W M (1984) Atlantic Seasonal Hurricane Frequency, Part I: El Niño and the 30 mb Quasi-biennial Oscillation Influences, *Mon. Weather Rev.*, 112, 1649–1668.

Helmholtz, H (1867) On Integrals of the Hydrodynamical Equations which Express Vortex Motion, *Philos. Mag.*, 33, 485–512.

Hendon, H H and Liebmann, B (1994) Organization of Convection Within the Madden Julian Oscillation, *J. Geophys. Res.*, 99, 8073–8083.

Holland, G J (1997) The Maximum Potential Intensity of Tropical Cyclones, *J. Atmos. Sci.*, 54, 2519–2541.

Holland, G J and Merrill, R T (1984) On the Dynamics of Tropical Cyclone Structural Changes, *Quart. J. R. Meteorl. Soc.*, 110, 723–745.

Knutson, T, Tuleya, R E, and Kurihara, Y (1998) Simulated Increase of Hurricane Intensities in a CO_2 Warmed Climate, *Science*, 279, 1018–1020.

Lighthill, J, Holland, G, Gray, W, Landsea, C, Craig, G, Evans, J, Kurihara, Y, and Guard, C (1994) Global Climate Change and Tropical Cyclones, *Bull. Am. Meteorl. Soc.*, 75, 2147–2157.

Madden, R A and Julian, P R (1972) Description of Global-scale Circulation Cells in the Tropics with 40–50 Day Period, *J. Atmos. Sci.*, 29, 1109–1123.

McBride, J L (1981) Observational Analyses of Tropical Cyclone Formation: III, Budget Analysis, *J. Atmos. Sci.*, 38, 1117–1131.

Meehl, G A, Zwiers, F, Evans, J L, Knutson, T R, Mearns, L O, and Whetton, P (2000a) An Introduction to Trends in Extreme Weather and Climate Events: Observations, Socioeconomic Impacts, Terrestrial Ecological Impacts, and Model Projections, *Bull. Am. Meteorl. Soc.*, 81, 413–416.

Meehl, G A, Zwiers, F, Evans, J L, Knutson, T R, Mearns, L O, and Whetton, P (2000b) Trends in Extreme Weather and Climate Events: Issues Related to Modeling Extremes in Projections of Future Climate Change, *Bull. Am. Meteorl. Soc.*, 81, 427–436.

Meehl, G A and Washington, W M (1996) El Niño-like Climate Change in a Model with Increased Atmospheric CO_2 Concentrations, *Nature*, 382, 56–60.

Merrill, R T (1988) Environmental Influences on Hurricane Intensification, *J. Atmos. Sci.*, 45, 1678–1687.

Miller, B I (1958) On the Maximum Intensity of Hurricanes, *J. Meteorl.*, 15, 184–195.

Nicholls, N (1984) The Southern Oscillation, Sea Surface Temperature and Interannual Fluctuations in Australian Tropical Cyclone Activity, *J. Climatol.*, 4, 661–670.

Rossby, C-G (1948) On Displacements and Intensity Changes of Atmospheric Vortices, *J. Mar. Res.*, 7, 175–187.

Royer, J-F, Chauvin, F, Timbal, B, Araspin, P, and Grimal, D (1998) A GCM Study of the Impact of Greenhouse Gas Increase on the Frequency of Occurrence of Tropical Cyclones, *Clim. Change*, 38, 307–343.

Shapiro, L J (1989) The Relationship of the Quasi-biennial Oscillation to Atlantic Storm Activity, *Mon. Weather Rev.*, 117, 1545–1552.

Watterson, I G, Evans, J L, and Ryan, B F (1995) Seasonal and Interannual Variability of Tropical Cyclogenesis: Diagnostics from Large-scale Fields, *J. Clim.*, 8, 3052–3066.

FURTHER READING

Houghton, J T, Meira Filho, G, Callandar, B A, Harris, N, Kattenberg, A, and Maskell, K (1996) *Climate Change 1995: The Science of Climate Change*, Cambridge University Press, Cambridge.

Houghton, J T, Ding, Y, Griggs, D J, Noguer, M, van der Linden, P J, Dai, X, Maskell, K, and Johnson, C A (2001) *Climate Change 2001: The Scienti c Basis*, Intergovernmental Panel on Climate Change Working Group I, Cambridge University Press, Cambridge, 1–881.

Hydrofluorocarbons

Hydro uorocarbons (HFCs) are compounds containing hydrogen, fluorine and carbon. The normal state of these compounds at standard temperature and pressure ranges from gases to liquids and waxes to solids. The gaseous HFCs can contribute to global climate change if they are emitted to the atmosphere. There are no known significant natural sources of HFCs. HFCs are removed from the atmosphere primarily through reaction with hydroxyl (OH) and their lifetimes range from about 1.5 to 250 years. These atmospheric lifetimes, combined with relatively strong infrared band strengths, lead to global warming potentials (GWPs) in the range 190–15 000 (*see* Global Warming Potential (GWP), Volume 1). For the most part, HFCs have been developed to replace ozone-depleting substances, primarily chlorofluorocarbons and hydrochlorofluorocarbons, that are being phased out under the Montreal Protocol on Substances that Deplete the Ozone Layer (*see* Depletion of Stratospheric Ozone, Volume 1; Ozone Layer: Vienna Convention and the Montreal Protocol, Volume 4). The commercially produced compounds are either non-flammable or low in flammability, low in toxicity, chemically and thermally stable and are compatible with a wide range of construction materials. The primary current and projected applications of HFCs are refrigeration, air conditioning, heat pumps and plastic foams for thermal insulation. In addition, small quantities of HFCs are used in specialty cleaning applications, for dispersing lubricants on computer discs, as a propellant in metered dose inhalers (e.g., asthma sprays), and as a non-flammable propellant in safety applications. The most abundant HFC in the atmosphere in 1995 was HFC-23 with a concentration of about 10 ppt. HFC-23 is produced as an unintended byproduct during the manufacture of HCFC-22 and much of what was produced has been emitted to the atmosphere.

MACK MCFARLAND *USA*

Hydrogen Peroxide Trends in Greenland Glaciers

Detlev Möller
Brandenburg Technical University, Cottbus, Germany

Hydrogen peroxide (H_2O_2), which is an important atmospheric oxidant, has been found at increasing concentrations in Greenland ice cores over the past 200 years. Most of the increase has occurred over the past 20 years (Sigg and Neftel, 1992; Anklin and Bales, 1997). Many trace species have been measured in Antarctic and Arctic ice cores that are unique in documenting the Earth's our past global

environment. Thus, it is possible to get not only information on the paleoclimate several thousands of years back, e.g., carbon dioxide (CO_2) concentrations over the last ice ages, and the occurrence of volcanic eruptions, but also on the increase of some compounds (e.g., nitrate, sulfate) with industrialization about 150 years ago. So-called primary compounds (e.g., heavy metals) may give an ice core concentration proportional to the emission strength. Secondary products (e.g., nitrate and sulfate) depend on the oxidation capacity of the atmosphere, because they are formed along different oxidation pathways from the primary ones.

Being a natural component of the atmosphere (having been found in rainwater more than 100 years ago), H_2O_2 is one of the species which is, to a large extent, responsible for oxidation of sulfur dioxide (SO_2) in the aqueous phase (cloud and rainwater) in the troposphere. H_2O_2 is produced in radical reactions that follow ozone photolysis. There is evidence for direct emission from biomass burning and non-radical and non-photolytic chemical formation. However, all atmospheric H_2O_2 sources are linked with human activities. Peroxides (like H_2O_2) have been found to be harmful. Radicals, subsequently produced after absorption, might impact enzymatic and genetic cell functions, called oxidative stress. Organisms have learned to build up antioxidant systems for repairing structural damages, which, however, will not completely compensate for this stress.

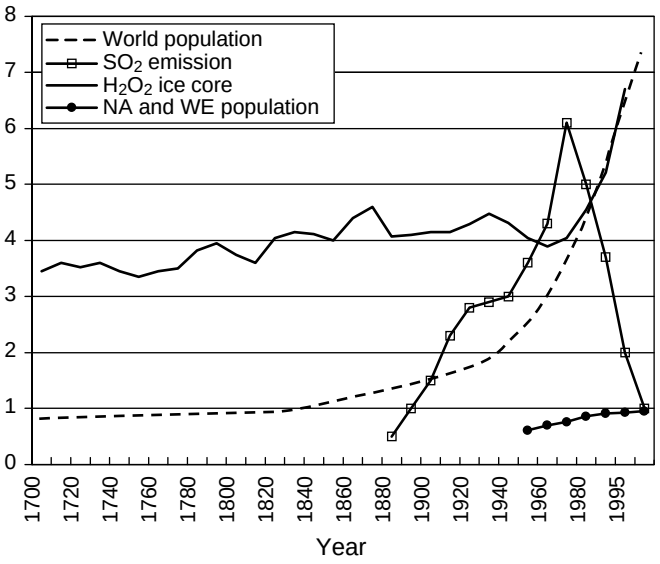


Figure 1 Comparison of 10 year averaged H_2O_2 concentrations in Greenland ice core in μM, world population and sum of population from Northern America and Western Europe (NA and WE) in 10^9 people and sum of Northern American and Western European SO_2 emission in 10^7 tons S year^{-1}; typical error bars ±10% for ice core concentration, ±20% for retrospective man-made SO_2 emission, ±50% for the 1995 figure and ±100% for the year 2000 figure (for references see Möller, 1999)

To begin to understand how the ice core concentration changes shown in Figure 1 may have arisen, we first need to consider the formation pathways of H_2O_2. In the gas phase the predominant mechanism is via $HO_2 + HO_2 \rightarrow H_2O_2$, with the reaction rate strongly depending on the water vapor pressure. The HO_2 radical required for this reaction is mainly formed through ozone, (O_3) photolysis via OH, which is in turn converted by carbon monoxide, CO and volatile organic compounds (VOC) to HO_2. Under NO rich conditions (polluted), HO_2 will be transformed back to OH, decreasing the H_2O_2 yield. Under NO poor conditions (clean background air), however, HO_2 reacts with O_3 resulting in a diurnal anticorrelation between O_3 and H_2O_2. Thus, only under moderately polluted conditions is HO_2 recombination a source of H_2O_2. H_2O_2 is a rather stable compound and does not undergo fast photochemical and gas phase reactions. The only sinks are dry deposition and scavenging by clouds and precipitation. Thus, transport of H_2O_2 over long distances is possible during essentially cloud-free conditions. Increasing emissions of CO and VOC are generally associated with the world s increasing population due to biomass burning and industrial activities. Consequently, one has to expect a positive correlation between the H_2O_2 signal and the world population (Figure 1), which, however, can be increased or weakened by other processes (*see* Troposphere, Ozone Chemistry, Volume 1).

H_2O_2 reacts in the liquid phase with dissolved sulphur dioxide, SO_2, forming sulfate. Despite the low volume fraction of cloud water in air (typically 0.3×10^{-6}) and the limited occurrence of clouds in time and space, it is now widely accepted that on average more than 80% of atmospheric non-seasalt sulfate is produced via the cloud phase. Thus, atmospheric H_2O_2 could be controlled to a large extent by SO_2 via liquid phase chemical processes.

To complicate matters, H_2O_2 in cloud drops may arise not only from scavenging of gaseous H_2O_2, but also from *in situ* liquid phase chemical reactions (Gunz and Hoffmann, 1990). One route involves the uptake of HO_2 and subsequent reactions, but we think (Moller and Mauersberger, 1992) that this source is negligible ($\sim 10^{-7}$ parts per billion (ppb) s^{-1}) compared to the gas phase production rate. Another aqueous phase formation mechanism for H_2O_2 may involve electron transfer to dissolved O_2 (Faust and Allen, 1992) where chromophores (e.g., aromatic carbonyles) in the droplets may be responsible for producing electrons after absorption of sunlight. The chromophores themselves could be products of biomass burning. Hence, besides sunlight, no gas phase oxidant precursor is needed in contrast to all other known H_2O_2 formation processes. Using the aqueous chemical rates, and considering cloud abundance within the planetary boundary layer (a factor of ≤ 0.1 has to be taken into account) we conclude that the mean liquid phase production of H_2O_2 may be of the same

order as the gas phase production ($\sim 10^{-5}$ ppb s^{-1}). This rate is related to summer and midday conditions. This leads us to conclude that H_2O_2 has a considerably lower annual mean production than given above, and large seasonal amplitude.

Recently, it has been proposed that biomass burning could be a direct source of H_2O_2. Another non-radical and light-independent H_2O_2 source is the gas-phase reaction between O_3 and olefins (ozonolysis). Under high humidity conditions, H_2O_2 may be formed directly; the mechanism for indirect formation via intermediate HO_2 radicals has been known for a long time (see references in Moller, 1989).

Besides the sulfur oxidation pathway, H_2O_2 is also consumed in its liquid phase via the Fenton reaction ($H_2O_2 + Fe^{2+} \rightarrow OH + OH^- + Fe^{3+}$) which, however, is relatively slow (Mn and Cu undergo similar Fenton-like reactions). Numerical modeling shows (Moller and Mauersberger, 1992) that cloud water accumulates H_2O_2 at all for concentrations of SO_2 smaller than 1 ppb.

Considering the production and loss terms for H_2O_2, the following relationships may be stated:

- Precursor emissions (CO, VOC) are proportional to human population when no alternative technologies and control techniques have been applied.
- Atmospheric H_2O_2 increases with increases in precursors (via O_3 formation).
- Atmospheric H_2O_2 increases with biomass burning (including wood burning).
- Atmospheric H_2O_2 is also produced photochemically within the gas and liquid phase as well at similar rates.
- Atmospheric SO_2 consumes H_2O_2 via the liquid phase.
- Seasonal amplitudes of atmospheric H_2O_2 (summer maximum) and emission of SO_2 (winter maximum) are anticorrelated; however, the amplitude for SO_2 is much smaller than that for H_2O_2.
- H_2O_2 ice core records represent the atmospheric burden of H_2O_2.

Considering all of these factors, we can now devise a scheme to explain the H_2O_2 trends in Greenland ice cores:

1. no significant trend in H_2O_2 in the period 1300–1750 (mean of $3.6 \pm 0.3\,\mu M$);
2. slow increase (0.2% $year^{-1}$) of H_2O_2 proportional to human population between 1750 and 1850 (i.e., wood burning as the main human energy source);
3. steady (mean of $4.0 \pm 0.6\,\mu M$), or decreasing H_2O_2 between 1850 and 1975 due to masking of the production term by the consumption of H_2O_2 due to highly dissolved SO_2 from fossil fuel combustion;
4. a marked increase in H_2O_2 by about a factor of 2 (5% $year^{-1}$) after 1975 due to widespread use of flue gas desulphurization and consequently decreasing SO_2 emissions as shown in Figure 1.

From the above arguments we calculate an annual flux of H_2O_2 in the Northern Hemisphere of about 0.9 Teq $year^{-1}$ (annual production rate half the summer time value given above, and a mean mixing depth of 1500 m). The corresponding SO_2 fluxes from North America and Western Europe (the principal sources of SO_2 to Greenland, see Figure 1) have dropped from 1.9 Teq $year^{-1}$ in 1970 to an estimated 0.3 Teq $year^{-1}$ in 2000. Clearly then, there has been a change from an oxidant- to a sulfur-limiting process. The recent ice core record of Anklin and Bales (1997) seems to confirm this. It shows an increase in the seasonal amplitude of H_2O_2 between 1988 and 1995 by a factor of about 5, due principally to higher summertime values. This agrees with our belief that H_2O_2 formation is dominated by the summer photochemical process.

An atmospheric increase of hydrogen peroxide in the Northern Hemisphere, which has to be expected as a consequence of the relationships presented here (and already forecast as a consequence of SO_2 emission abatement by Moller and Mauersberger, 1992), has not been checked with experimental data due to the lack of long-term gas phase H_2O_2 measurements. Such monitoring extending over several years is from Harwell (Dollard and Davies, 1993), showing a significant increase in gaseous H_2O_2 between 1987 and 1991 of $+0.028$ ppb $year^{-1}$ ($r^2 = 0.83$).

The hypothetical strong increase in H_2O_2 in the Northern Hemisphere must be seen in the context of a very slow population increase since 1950 (Figure 1) and may therefore only be a result of the strong SO_2 emission decrease. Considering the Southern Hemisphere, where most of the increase in the world population is occurring (Figure 1) and due to SO_2 emission abatement strategies probably applying in the future, a much stronger effect on atmospheric H_2O_2 could be possible. The consequences of increasing H_2O_2 are oxidative stress to vegetation (Moller, 1989) and possibly also to humans and animals, and may in future exceed, by several times, those arisen by O_3.

REFERENCES

Anklin, M and Bales, R C (1997) Recent Increase in H_2O_2 Concentration at Summit, Greenland, *J. Geophys. Res.*, 102, 19 099–19 104.

Dollard, G J and Davies, T J (1993) Measurements of Rural Photochemical Oxidants, in *The Chemistry and Deposition of Nitrogen Species in the Troposphere*, ed A T Cox, Royal Society Chemistry, Cambridge, 46–77.

Faust, B C and Allen, J M (1992) Aqueous-phase Photochemical Sources of Peroxy Radicals and Singlet Molecular Oxygen in Clouds and Fog, *J. Geophys. Res.*, 97, 12 913–12 926.

Gunz, D W and Hoffmann, M R (1990) Atmospheric Chemistry of Peroxides: A Review, *Atmos. Environ.*, 24A, 1601–1633.

Moller, D (1989) The Role of H_2O_2 in New-type Forest Decline, *Atmos. Environ.*, 23, 1625–1627.

Moller, D (1999) Explanation for the Recent Dramatic Increase of H_2O_2 Concentrations Found in Greenland Ice Cores, *Atmos. Environ.*, 33, 2435–2437.

Moller, D and Mauersberger, G (1992) Cloud Chemistry Effects on Tropospheric Photooxidants in Polluted Atmosphere, *J. Atmos. Chem.*, 14, 153–165.

Sigg, A and Neftel, A (1992) Evidence for a 50% Increase in H_2O_2 over the past 200 years from the Greenland Ice Core, *Nature*, 351, 557–559.

Hydrologic Cycle

Thomas Pagano and Soroosh Sorooshian
University of Arizona, Tucson, AZ, USA

The importance of water on Earth cannot be underestimated. Water is transported endlessly throughout the various components of the Earth's climate system, affecting every component along the way. Clouds and water vapor in the atmosphere influence the energy balance of the Earth, and snow and ice-covered surfaces reflect a significant amount of the Sun's radiation back to space. Water at and under the land's surface, in the form of stream flow or groundwater, plays an important role in the maintenance of living organisms and human societies. However, while water is a benefit, if it arrives at the wrong time, in the wrong quantity, or is of poor quality, it can be a severe hazard. To mitigate these hardships, humans have significantly altered the hydrologic cycle with construction of dams, cultivation of farmland, urbanization, draining of swamplands, etc. The local consequences of these changes can be dramatic, leading to environmental changes and, in some cases, local degradation. The global impacts are difficult to ascertain, however. With the increasing demand for freshwater resources and increased societal vulnerability to climate extremes, the effects on humans by water-related global environmental change remain an interesting but as of yet unresolved question.

INTRODUCTION

The hydrologic cycle is the perpetual movement of water throughout the various components of the Earth's climate system. Water is stored in the oceans, in the atmosphere, as well as on and under the land surface. The transport of water between these reservoirs in various phases plays a central role in the Earth's climate. Water evaporates from the oceans and the land surface into the atmosphere, where it is advected across the face of the Earth in the form of water vapor. Eventually, this water vapor condenses within clouds and precipitates in the forms of rain, snow, sleet, or hail back to the Earth's surface. This precipitation can fall on open bodies of water, be intercepted and transpired by vegetation, and become surface runoff and/or recharge groundwater. Water that infiltrates into the ground surface can percolate into deeper zones to become a part of groundwater storage to eventually reappear as streamflow or become mixed with saline groundwater in coastal zones. In this final step, water re-enters the ocean from which it will eventually evaporate again, completing the hydrologic cycle. The hydrologic cycle qualitatively, quantitatively, and conceptually is depicted in Figures 1–3.

The important reservoirs within the hydrologic cycle include:

Ocean

This vast body of salt water covers 70% of the Earth's surface; it stores and circulates enormous amounts of water and energy. In addition, patterns of ocean surface temperatures can exert a strong influence on circulation patterns in the atmosphere. Frequently, the ocean is divided into two parts, an upper and lower zone. The upper zone is considerably warmer and less saline than the lower zone, and the two are separated by a relatively sharp thermocline. The depth from the ocean surface to the thermocline can be as much as 400 m, but is generally less than 150 m.

Atmosphere

Water can be stored in the atmosphere as liquid in clouds or as water vapor. Water vapor content of the atmosphere is described by its humidity. Specific humidity is a measure of the water content per unit of dry atmosphere (typical values are $1-20\,g\,kg^{-1}$); relative humidity is the amount of water vapor present relative to the amount of water vapor that would saturate the air at a particular temperature. The presence of water in the atmosphere alters the radiation budget of the atmosphere, directly through latent heat and indirectly as both a reflector and absorber of radiation. Water in the atmosphere is the most significant contributor to the natural greenhouse effect.

Cryosphere

The largest stores of fresh water on the Earth are contained in glaciers and icecaps, primarily at high latitudes. The cryosphere has a significant impact on the climate of the Earth because snow and ice-covered surfaces have a very high albedo (comparable to that of clouds). The large volume of runoff from northern high-latitude rivers also influences the Arctic and Atlantic Ocean circulation, which impacts the climate in those regions. Despite the importance

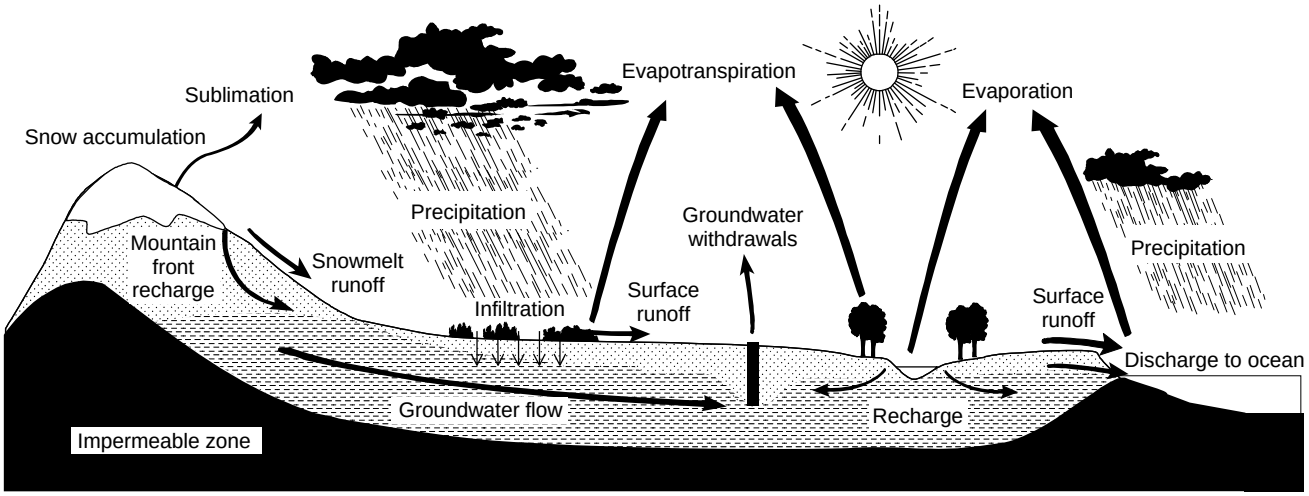


Figure 1 A schematic diagram of various fluxes within the hydrologic cycle. (Figure prepared by B Imam)

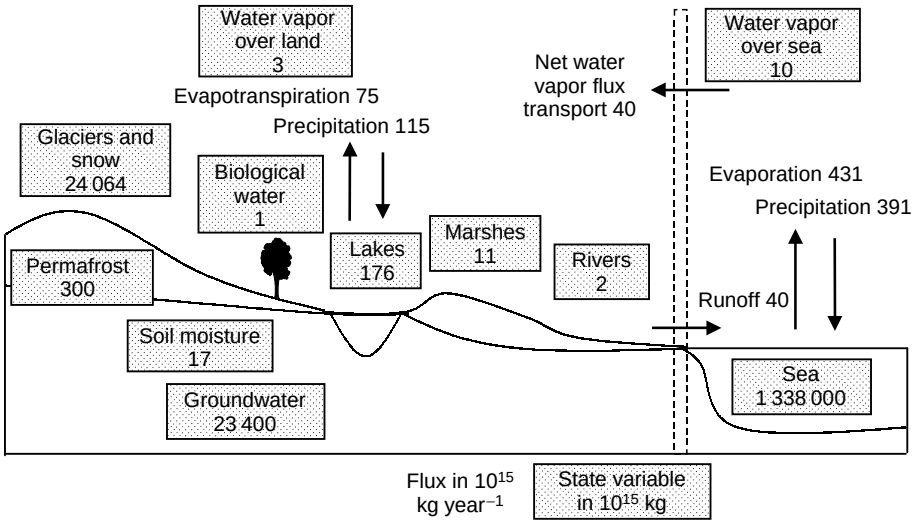


Figure 2 A diagram of the various fluxes and reservoirs within the hydrologic cycle with their yearly average magnitudes (after Oki, 1999) (reprinted with permission of Cambridge University Press). The magnitudes given are approximate, and differ from other authors. For example, see Chahine (1992); Figure 1 for comparison

of the cryosphere to the hydrologic cycle, relatively little is understood about this part of the climate system, partially because of the lack of adequate data in these often remote and difficult to access areas.

Groundwater

Water beneath the land surface can be classified in a variety of ways. Water closest to the surface (within a few meters) is considered soil moisture, and this water influences the evapotranspiration rate of water from the surface. Soil moisture that is frozen year-round is called permafrost. Deeper below the surface is the aquifer, where the water concentration in the rock and soil is sufficient for withdrawal by pumping. Groundwater for human activities is contained primarily in the aquifer. In this saturated zone, all available spaces within the rock and soil are filled with water. Between the aquifer and soil moisture lays an unsaturated intermediate (vadose) zone that has a lesser influence on the atmosphere than soil moisture. Despite the great societal importance of groundwater supplies, quality spatially distributed subsurface data are elusive.

Land Surface

Water on land can be contained in lakes and marshes as well as rivers and within living organisms (biological water). The volume of water stored on land is relatively small, but the flux of water throughout these systems is relatively

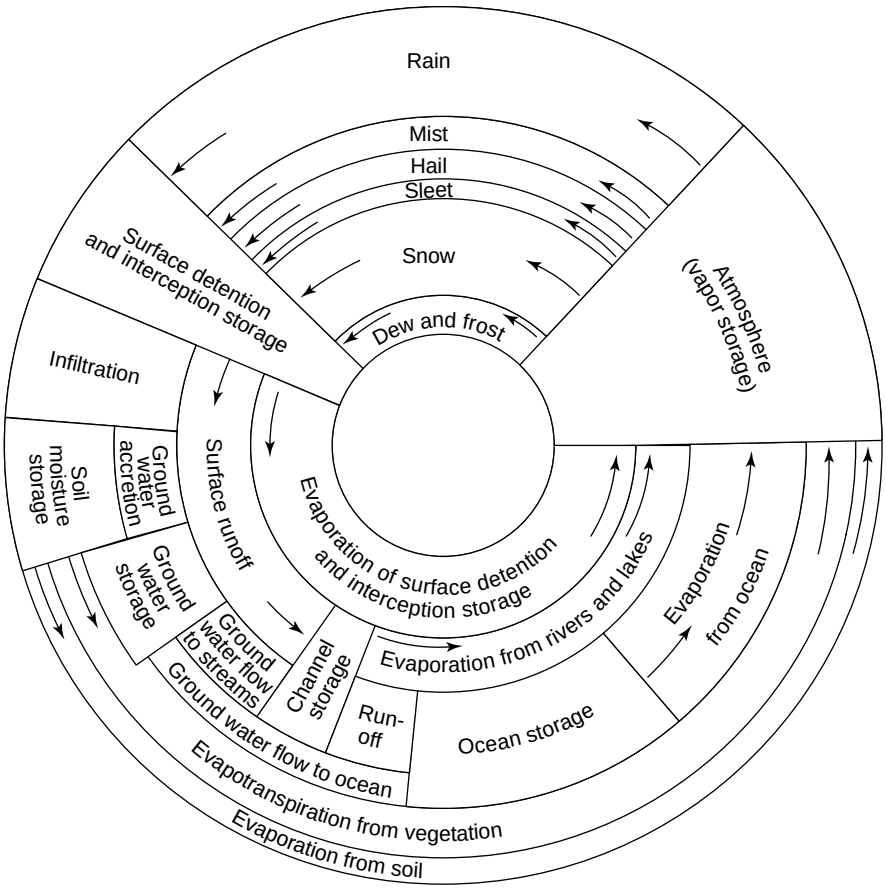


Figure 3 A conceptual diagram of the hydrologic cycle (after Wisler and Brater, 1959). (Reproduced by permission of John Wiley & Sons)

high. The relevance of this water to human activities is paramount.

In the broadest sense, the *major uxes* between reservoirs are:

Precipitation

Precipitation is the fall of solid or liquid water over land and oceans, and is the major driver of the hydrologic cycle over land. Hydrologists have traditionally recognized precipitation as the start of the hydrologic cycle because all other hydrologic phenomena (e.g., evaporation, runoff, recharge) result from it. The importance of precipitation to the hydrologic cycle cannot be overstated.

Evapotranspiration

Evaporation is the return of water from bare soil or open bodies of water (mainly the ocean surface) to the atmosphere. Transpiration is the transfer of water to the atmosphere through the stomata of vegetation. Collectively, they are considered evapotranspiration.

Runoff

Runoff is the transport of liquid water across the surface of the Earth. Excess water in saturated soils flows into rivers to the ocean, to terminal lakes or swamps. Groundwater can interact with streamflow in rivers if the water table is near the surface.

Water Vapor Transport

Atmospheric water vapor transport is the redistribution of atmospheric water vapor. Globally, there is a net transfer from over ocean to over land. This process is known as advection, and this flux is the major source of water vapor for precipitation over land, aside from recycling.

DESCRIPTIONS OF THE HYDROLOGIC CYCLE

Mathematical Models

Mathematically, the movement of water throughout the hydrologic cycle can be described using the hydrologic continuity equation:

$$I - O = \frac{\Delta S}{\Delta t} \quad (1)$$

where input (I) and output (O) depend on the reservoir in question (e.g., evapotranspiration is an input to the atmosphere, whereas precipitation is an output). The change in storage (S) in time describes the removal from or addition to present supplies to make up for the imbalance between input and output (in the case of the atmosphere, change in storage would signify a change in specific humidity).

In contrast to the atmosphere, the water balance of a surface portion of a river basin is considerably more complex. Water is input into this system through precipitation, surface runoff, and groundwater inflow from other parts of the basin. Water is lost through surface runoff, groundwater outflow, and evapotranspiration. The change in storage is reflected in changes in soil moisture content. A graphical depiction of a water balance for a portion of the atmosphere and land surface is shown in Figure 4.

On a global basis, the Earth is effectively a closed system, and the amount of water present remains relatively constant (i.e., $\Delta S/\Delta t \approx 0$). However, input and output rates of the hydrologic cycle vary regionally and on a wide range of time scales. Describing, quantifying, and predicting these variations are, in essence, major tasks in contemporary hydrology.

To describe and predict variations within the hydrologic cycle, considerable effort has been invested in developing computer-based numerical models of hydrology and climate. Every component of the climate system has its own models (from groundwater to oceanography and the atmosphere to the land surface) and, within disciplines, there are too many different models to be completely described here. For example, surface hydrologic models range from simple statistics, to regression models (such as the antecedent precipitation index and the soil conservation service curve number model) to more complex conceptual rainfall-runoff models (like the Sacramento model) and finally to the physically based distributed models such as HEC-1 and KINEROS. The spatial and temporal scales of their applications vary from model to model, but ranges from tens to thousands of square kilometers and from minutes and hours to days and years, respectively (Sorooshian *et al.*, 1996). The most complex models available are general circulation models (GCMs; *see* General Circulation Models (GCMs), Volume 1), which have full global representations of the ocean, atmosphere, cryosphere, and land surface. Most of the early work on GCMs related to refining the treatment of the ocean–atmosphere interface. Recently, increasing emphasis has been put upon the land surface–atmosphere interface, improving such models as the Biosphere Atmosphere Transfer Scheme and the Simple Biosphere model (*see* Land Surface, Volume 1). Lau *et al.* (1995) compared the ability of 29 GCMs in simulating various aspects of regional hydrologic processes and found them insufficient for use in climate studies related to continental scale water balance. Regardless, this is an area of very active research and as computing power rapidly increases in the near future, one can expect these models to improve.

Data

To date, there are several definitive works providing quantitative descriptions of the global hydrologic cycle

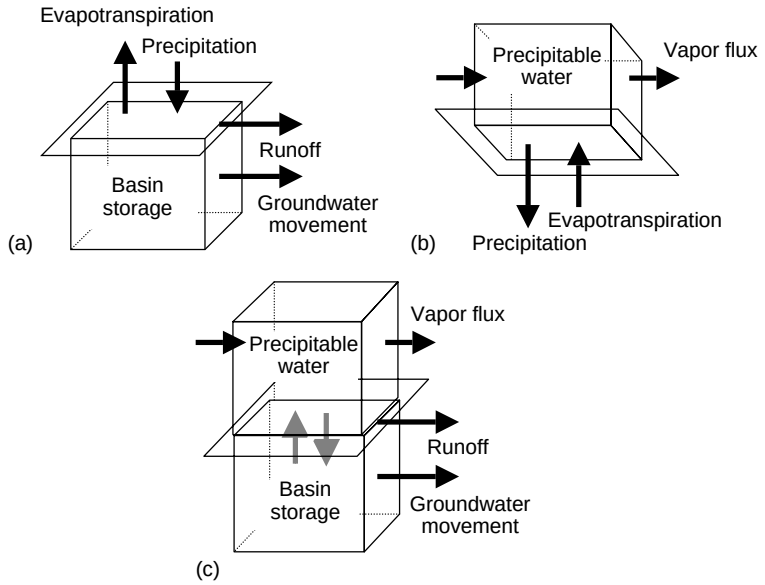


Figure 4 Mathematical schematic of a water balance for: (a) the land surface; (b) the atmosphere; and (c) the combined atmosphere and surface (from Oki, 1999). (Reprinted with permission of Cambridge University Press)

(for example, Korzun, 1978; Piexoto and Oort, 1992; Oki, 1999). The most comprehensive review of freshwater resources (supply and use) can be found in Shiklomanov (1999) http://espejo.unesco.org.uy/ and Shiklomanov (1997) http://pangea.upc.es/orgs/unesco/webpc/world_water_resources.html. Earlier works quantifying the complete water cycle have attempted analysis using sparse measurements; the creation of re-analyzed data sets by the European Center for Medium-Range Weather Forecasts and the National Center for Atmospheric Research (NCAR)/National Center for Environmental Prediction (NCEP) represents a significant advance to these kinds of studies. These data sets blend measurements from rawinsondes, satellite temperatures and moisture, cloud track winds, surface observations by ships, ocean buoys, land stations, aircraft reports, and GCM analysis of atmospheric dynamics. At present, reanalyzed data sets represent the best available global atmospheric water balance measurements.

Quality measurements of individual components of the hydrologic cycle also exist (*see* Earth Observing Systems, Volume 1). In particular, global records of precipitation and streamflow are maintained by the Global Precipitation Climatology Center (GPCC; http://www.dwd.de/research/gpcc/) and the Global Runoff Data Center (GRDC http://www.bafg.de/grdc.htm), respectively. Within the US, precipitation records are available from the National Climatic Data Center (NCDC; http://www.ncdc.noaa.gov/), and runoff is available from the US Geologic Survey (USGS; http://www.water.usgs.gov). Of all surface hydrologic variables, precipitation and runoff are the best measured, whereas data quality is considerably lower for other variables, such as snow, soil moisture, and evapotranspiration (Sorooshian *et al.*, 1996). One of the greatest barriers to adequately measuring the hydrologic cycle is the lack of spatially distributed data. Radar and satellite measurements (e.g., the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) system, Sorooshian *et al.*, 2000) hold great promise in overcoming this barrier, but many of their benefits have yet to be fully realized.

OBSERVED CLIMATOLOGY

The quantities of water contained within the different components of the Earth s system are listed in Table 1.

Table 1 Distribution of water on Earth[a]

Form of water	Area covered (1000 km^2)	Volume (1000 km^3)	Share of world reserves of total water (%)	Share of world reserves of fresh water (%)
World ocean	361 300	1 338 000	96.5	–
Total groundwater[b] (saturated and vadose)	134 800	23 400	1.7	–
Predominantly fresh groundwater	134 800	10 530	0.76	30.1
Soil moisture	82 000[c]	16.5	0.001	0.05
Glaciers and permanent snow cover	16 000	24 000	1.74	68.7
Antarctica	14 000	22 000	1.56	61.7
Greenland	1800	2300	0.17	6.68
Arctic islands	230	83.5	0.006	0.24
Mountainous areas	220	40.6	0.003	0.12
Ground ice in zones of permafrost strata	21 000	300	0.022	0.86
Water reserves in:				
Lakes	2000	180	0.013	–
Fresh water	1240	91	0.007	0.26
Salt water	820	85	0.006	–
Marsh water	2700	11.5	0.0008	0.03
Water in rivers	148 800[d]	2.12	0.0002	0.006
Biological water	510 000	1.12	0.0001	0.003
Atmospheric water	510 000	12.9	0.001	0.04
Total water reserves	**510 000**	**1 390 000**	**100**	**–**
Fresh water	148 800	35 000	2.35	100

[a] The individual number and column totals may not exactly agree due to rounding. Table modified from Korzun (1978) (reproduced by permission of the United Nations Educational, Scientific and Cultural Organization).
[b] Not including groundwater reserves in Antarctica, broadly estimated at 2 million km^3.
[c] Soil moisture is the water contained within the top 2 m of soil. This does not include ice and snow covered land areas, permafrost regions and arid and semi-arid areas (which collectively cover approximately 65.5 million km^2).
[d] The area of rivers here is the entire land surface. The true area of rivers, in the sense of channels with flowing surface water, is extremely small (<1 million km^2).

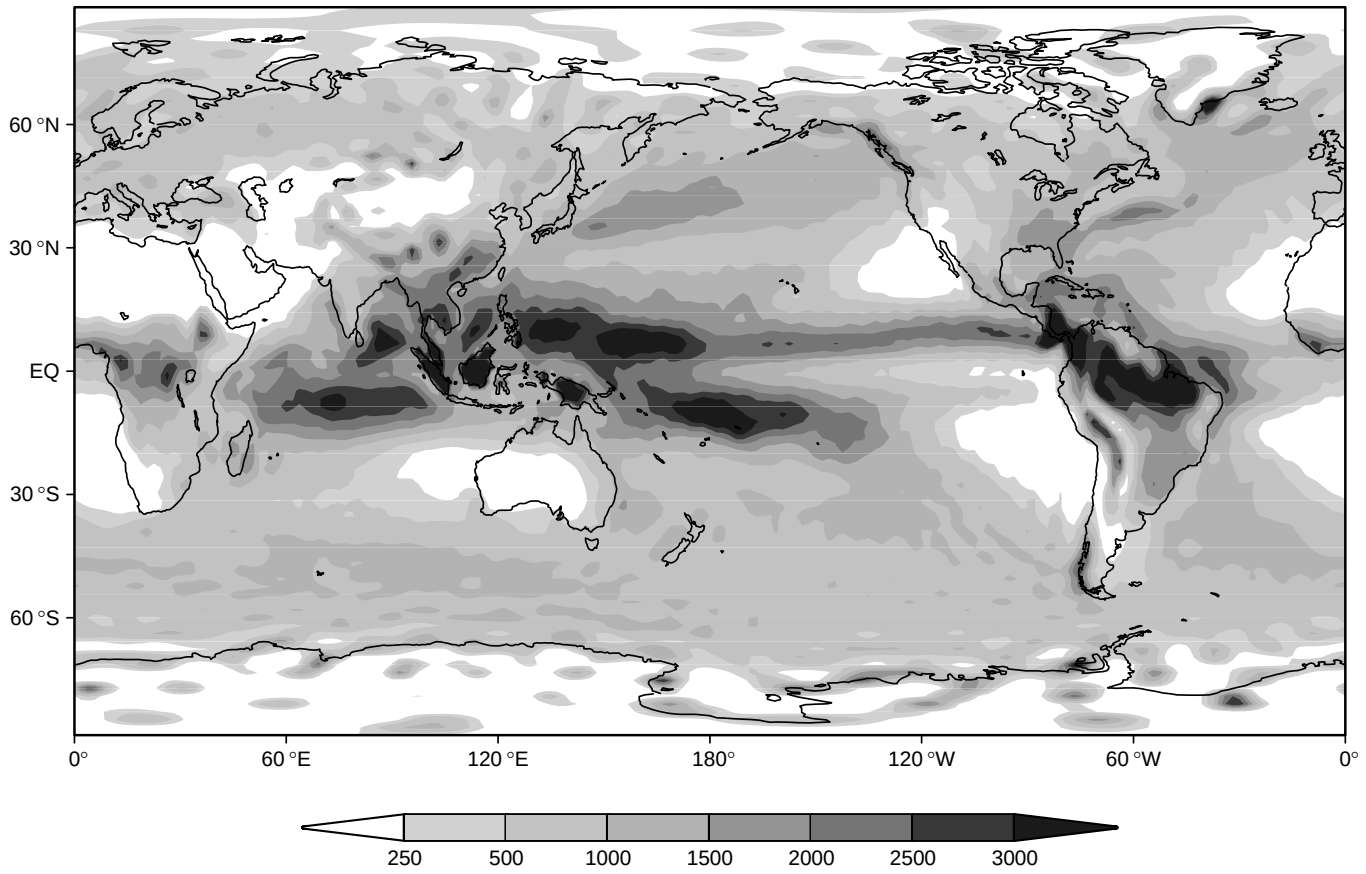


Figure 5 Yearly average precipitation from the NCAR/NCEP re-analysis of data from 1979–1995. Data are shown in mm year^{-1}. (Data provided by S Williams of UCAR)

The vast majority of the Earth s total water is contained in the oceans (96.5%, 1.34 billion km^3). The atmosphere contains a relatively small amount of the Earth s total water (0.001%, 12 900 km^3). Of the water not in the ocean (3.5%, 52 million km^3) and if the amount of water contained in polar ice caps and saline waters is excluded, only 15 million km^3 (31% of fresh, 0.8% of total) remain. If the water on the Earth fitted proportionately within a 200-liter (45 imperial gallon) barrel, the water contained in the atmosphere would fill just under half a liter, and freshwater lakes would fill a tablespoon. The water of all the Earth s rivers, from the Nile to the Amazon and the Yangtze to the Mississippi, would fit in the eye of a needle.

The average rate of exchange between the various components can be derived from Figure 2. The residence time (the average length of time that an element of water would spend in a component before moving on) is the total storage of the component divided by the rate of flux out of the component. For example, the residence time of water in the atmosphere is less than 10 days, whereas the ocean has a residence time of over 3200 years (longer for deep parts of the ocean). Depending on the aquifer, residence times for deep groundwater can be very long ($>$10 000 years). On human time scales, many groundwater supplies can be effectively considered non-renewable. Measurements of residence times are important in detecting, for example, an intensification of the hydrologic cycle associated with climate change. This intensification would appear as a decrease in the residence time of water in the atmosphere.

There is considerable spatial variation within the hydrologic cycle. Precipitation is generally higher in the tropics and along mid-latitude storm tracks (Figure 5). Precipitation patterns also vary by season, following the migration of the intertropical convergence zone, storm tracks, and monsoon systems. Global average precipitation is approximately 970–1000 mm year^{-1}, where average precipitation over the ocean is roughly 1100 mm year^{-1} and over land is 700–750 mm year^{-1} (Table 2). The average annual local precipitation rate can be as high as 13 300 mm year^{-1} in Lloro, Columbia and as low as 0.76 mm year^{-1} in Arica, Chile (US Army Corp of Engineers, 1997), although there are dry valleys in the interior of Antarctica that are believed to have not experienced rainfall in the last 2 million years. Globally, precipitation and evaporation are approximately in balance, although the ocean is considered a source for the

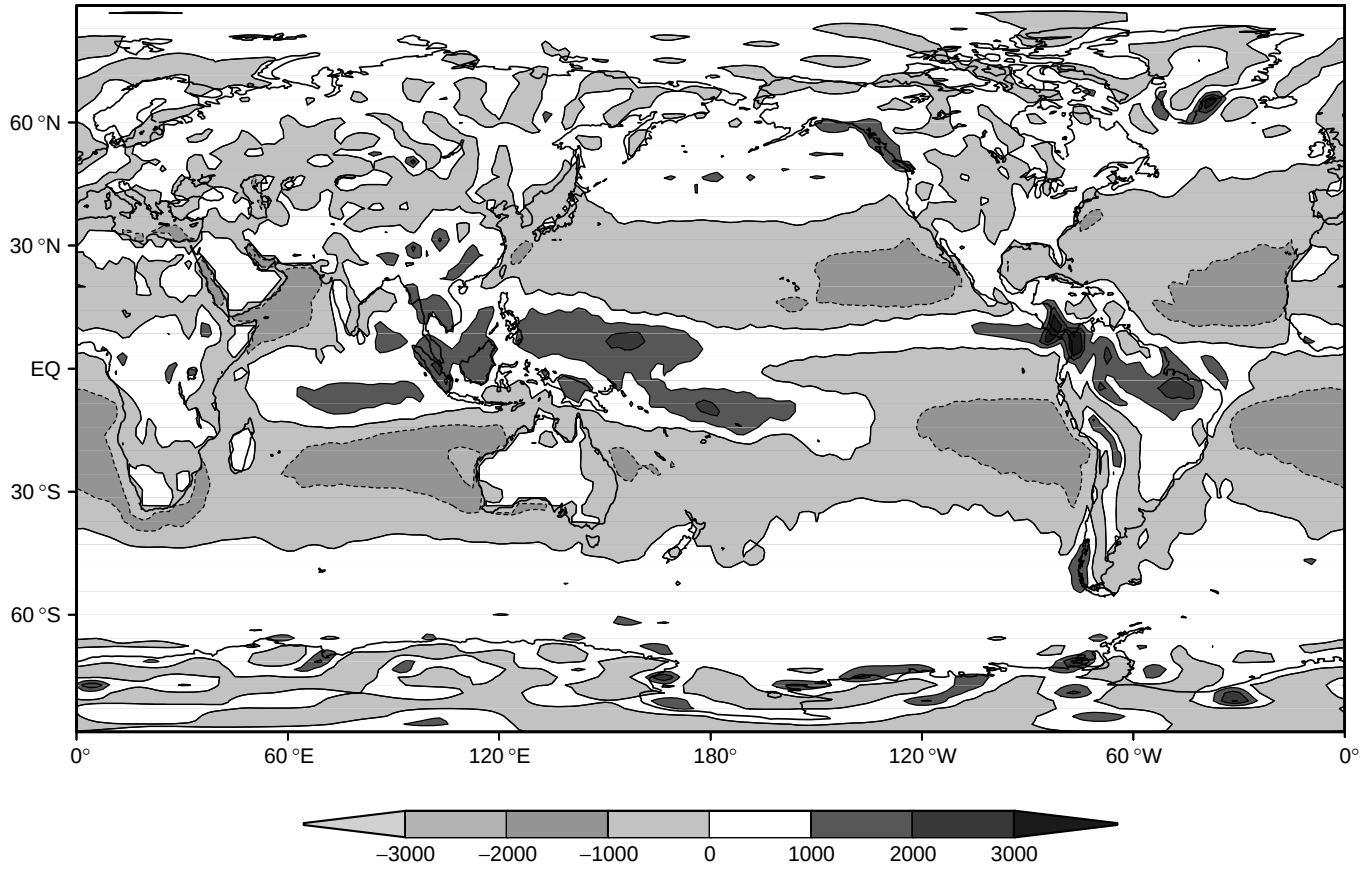


Figure 6 Yearly average precipitation-evaporation from NCAR/NCEP re-analysis data in millimeters year^{-1}. See also Figure 5

atmosphere with average evaporation-precipitation being 110 to 130 mm year^{-1}, and the land is considered a sink at -260 to -310 mm year^{-1} (Figure 6).

The numbers presented here are not exact and may disagree with those of other authors, emphasizing the level of uncertainty present when computing a water balance. The disagreement between authors can become large at finer spatial scales. For example, comparisons of two authors estimates of precipitation, evaporation, and runoff by continent and ocean are presented in Table 2.

VARIATIONS IN THE HYDROLOGIC CYCLE

Although general descriptions of the hydrologic cycle (such as those presented at the beginning of this article) make it seem relatively simple, once one starts to consider the smaller-scale aspects of the hydrologic cycle, it is, in reality, quite complex. Variations within the hydrologic cycle span a wide range of spatial and temporal scales. There are distinct seasonal variations to precipitation and evaporation, and fluxes within the hydrologic cycle vary with latitude and by continent, as stated earlier. However, precipitation also varies with altitude and orientation to local mountains, creating an enormous diversity of microclimates across the globe. Likewise, local weather conditions can change in a matter of minutes, leading to punctual events such as flash floods and microbursts, or they can evolve very slowly, such as long-term drought. Several studies have detailed the static aspects of the global hydrologic cycle, and the interesting new questions in hydrology and meteorology concern the hydrologic cycle as a dynamic system operating and interacting at a variety of scales. What causes drought? What makes one winter wetter or drier than any other? What are the implications of climate change for extreme events? All of these are questions facing researchers today.

One of the areas of increasingly more active research concerns variations within the Earth s climate on interannual, interdecadal, and longer time scales. Prior to the 1970s, solar variations were the most common natural explanation for year to year variations in climate and the hydrologic cycle. Streamflow of certain rivers could be correlated with various sunspot cycles, such as the 11-year cycle. Coincidentally, the earliest attempts at climate forecasting at the turn of the century were born out of attempts to link sunspot cycles with the periodicities found in climate time series, such as Indian monsoon rainfall (the variability of which is now known to be influenced by El Niño (*see* El Niño, Volume 1). However, the predictive skill of sunspot–climate relationships is low, and the relationships are unstable in space and time (Korzun, 1978; Allan *et al.*, 1996).

In recent decades, scientists have developed an appreciation for the relationship between the ocean and the atmosphere, in particular how seasonal patterns of local climate can be affected by ocean temperatures in remote areas (also known as teleconnections following Bjerknes, 1969). The El Niño/Southern Oscillation (ENSO, *see* El Niño/Southern Oscillation (ENSO), Volume 1) cycle is the most widely studied of this type of phenomenon. Recent large events have occurred in 1982–1983 and 1997–1998. The ENSO cycle is a coupled variation of the ocean and atmosphere in the tropical Pacific that impacts precipitation in many locations across the globe, from Peru to Africa, Australia and the US, among others (Ropelewski and Halpert, 1987). During warm ENSO events, ocean temperatures become warmer than usual in a region about the size of the US, extending from the coast of Peru to across the international dateline. This warming causes shifts in the patterns of convection in the tropics and, in turn, impacts the global atmospheric circulation. This favors seasonal precipitation anomalies in specific regions, such as floods and droughts. During cold ENSO events (La Niña), the same region of the ocean becomes colder than usual, and the global impacts are generally but not exactly opposite to those of warm ENSO events. These events occur irregularly every two to eight years and can last from one to several years. The breadth and magnitude of variations in the Earth s climate due to ENSO are second only to the changing of the seasons (Allan *et al.*, 1996). Predictions of ENSO have contributed successfully to seasonal forecasts of precipitation, which, in turn, can be very useful for water management (for example, Changnon and Vonnhame, 1986; Stern and Easterling, 1999; Changnon and Bell, 2000).

Variations and teleconnections on longer time scales also have major impacts on the hydrologic cycle. Two phenomena that have received considerable attention and research are the North Atlantic Oscillation (NAO, *see* North Atlantic Oscillation, Volume 1) and the Pacific Decadal Oscillation (PDO, *see* Paci c- Decadal Oscillation, Volume 1). The PDO concerns ocean temperature variations in the northern Pacific, whereas NAO concerns the northern Atlantic Ocean and atmosphere. The PDO operates on a time scale of 20–30 years, with observed shifts in the 1890s, 1920s, 1940s, 1970s and possibly the mid-1990s. The various phases of the PDO have been associated with wet and dry periods in North America and can serve to enhance or cancel the effects of ENSO, depending on their states. The NAO exerts a considerable influence on the hydroclimatology of Europe, Northern Africa, and the Middle East, among other regions. There is some evidence supporting the idea that variations in the North Pacific and North Atlantic are coupled and should be thought of as two manifestations of a single underlying phenomenon. Developing predictors of both of these oscillations remains an area for active research.

On centennial scales and longer, the hydrologic cycle has changed in both subtle and dramatic ways. Changes in the orbital parameters of the Earth with respect to the Sun have

been responsible for glacial periods in the Earth s history. During glacial periods, the extent of glaciers increases to much more than that of today. For example, during the most recent glacial maximum 20 000 years before the present interglacial, Chicago lay beneath more than 1 km of ice. Changes in the Earth s temperature and the volume of water trapped in glaciers caused the sea level to be about 100 m lower than today, drying up the English Channel. Clearly, large-scale weather patterns were altered, changing patterns of precipitation and the partitioning of precipitation into rain and snow. Likewise, continental streamflow would have increased as the glaciers melted. Changes in streamflow can impact ocean circulation. As the glaciers retreated from the most recent glacial maximum, and their waters released into the North Atlantic, the density of ocean surface water was reduced, and the formation of deep ocean water in this region was halted. As a result, Europe s temperature and hydrology was much different than today. On climate time scales, this switching was extraordinarily rapid, on the order of decades (*see* Thermohaline Circulation, Volume 1; Younger Dryas, Volume 1).

Aside from ocean and atmospheric variability, the land surface itself can impact precipitation in local and remote areas. Variations in soil moisture influence precipitation due to the amount of moisture that is recycled over land. Chahine (1992) estimated that a full 65% of land-falling precipitation comes from evaporation over the land, most of it advected in from other locations (see Trenberth, 1999a for further discussion on evaporation and moisture recycling). The amount of moisture available, as well as its spatial distribution, is important, even at fine scales; sharp soil–moisture gradients have been known to influence the development of tornadoes, for example. Discontinuities in soil moisture, such as those found at the interface between irrigated agriculture and native vegetation, tend to enhance shallow convective precipitation.

Vegetation also regulates the availability of soil moisture to the atmosphere through the opening and closing of plant stomata. Changes in surface vegetation can induce changes in local meteorology and climate. In particular, decreases in precipitation are believed to be associated with changes in land cover in south–east Asia and the Amazon (and in Africa; *see* Land Cover and Climate, Volume 1). Charney (1975) found that the absence of significant moisture sources might help to maintain deserts. In other words, the lack of local moisture for recycling purposes makes a desert a stable system and that, once an area has been made into a desert, it is difficult to change without external forcing (*see* Deserts, Volume 1).

HUMANS AND THE HYDROLOGIC CYCLE

One component of the hydrologic cycle that is frequently not directly included in its descriptions is human activity. The hydrologic cycle would continue, irrespective of human activities, but humans do have a significant impact on the terrestrial component of the hydrologic cycle. Likewise, changes in the hydrologic cycle can dramatically impact human activities, for better or for worse. Under growing population pressures, decreasing availability of freshwater per person, and potential global climate change, how this feedback will develop in coming years remains an interesting yet unresolved question.

Are humans affected by the hydrologic cycle? Certainly. Water is essential to life on Earth. The availability of water has shaped where civilizations have developed and thrived, just as lack of water has caused great hardships. Water is both a necessity and a resource for financial gains. Yet, while it is a benefit, it is also a hazard. On average, over $8 billion damages per year have resulted from flooding and hurricanes in the US alone (Kunkel *et al.*, 1999). Indeed, this does not include pandemic health hazards created by poor water quality.

Will humans affect the hydrologic cycle in the future? Undoubtedly. There is considerable evidence that humans have affected the hydrologic cycle in the past (Figure 7). Large dams, reservoirs, and extensive canal systems are perhaps the most visible testimony to this. In several basins across the globe, surface water resources have been so intensively developed that major rivers periodically cease flowing to the ocean (such as the Colorado River in the US and the Yellow River in China). Inter- and intra-basin transfers have significantly interfered with the natural distribution of water.

The most pervasive change to the hydrologic cycle due to human activities is associated with land-use change. These changes are very important to consider as, according to van Dam (1999, xiii), the effects of climate variability and change on the hydrological cycle will be coincident with those of changes in land use, which could be of the same order of magnitude. The various types of land-use changes range from deforestation to agriculture, urbanization to draining swamplands. The impact of these land uses on streamflow is presented in Table 3. The impacts arise from changes in surface albedo, surface roughness, surface permeability (the ability of water to pass through a surface, such as concrete), and the ability of the surface to intercept and evaporate moisture. These impacts are inherently scale-dependent, and most local land-use change will not have a major impact on the continental and global hydrologic cycle. However, the extent of land-use change in total is considerable; Flohn (1973) suggested that, over the last 8000 years, approximately 11% of the land surface has been converted to arable land, and 31% of forests have been modified from their original condition. Additionally, certain regions are poised to have a disproportionately strong impact on global circulation. For example, there is debate as to whether Amazon deforestation will

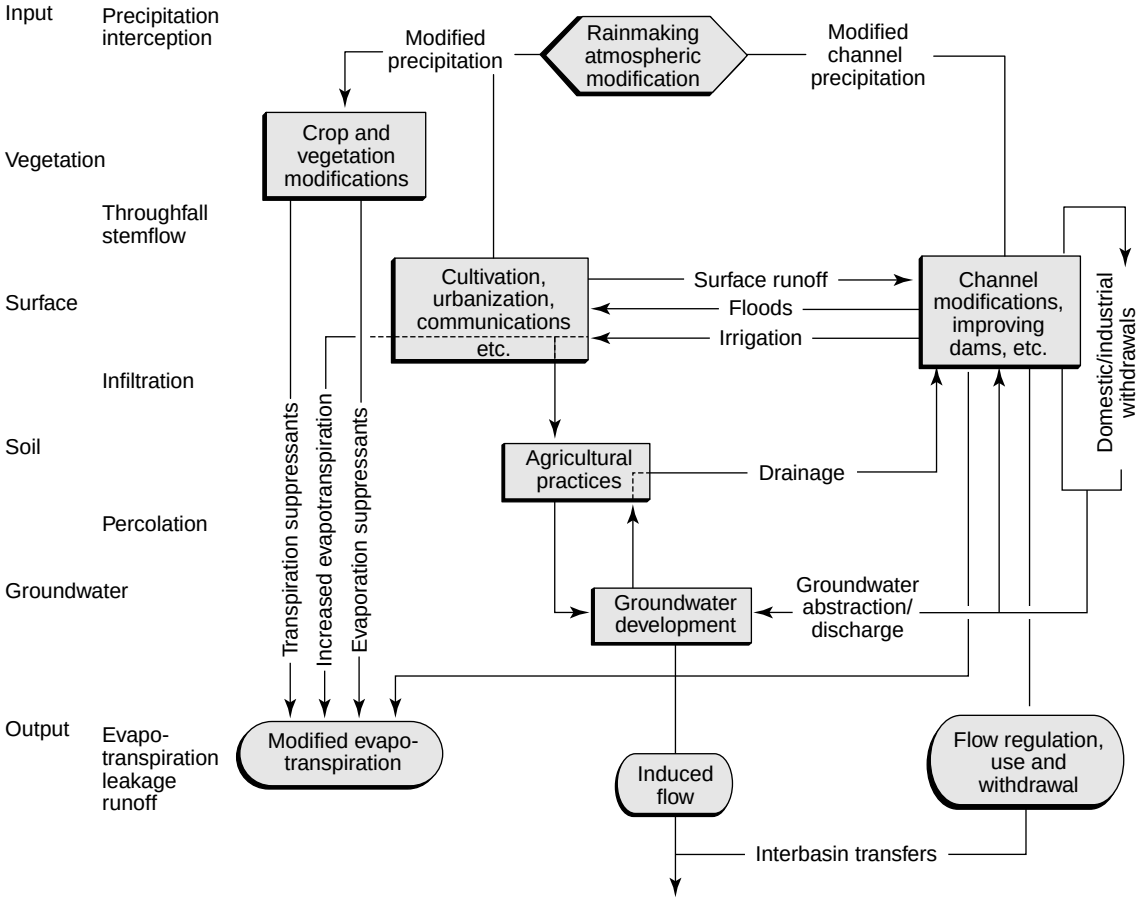


Figure 7 Systems diagram of the impacts of human activities on streamflow (from Ward, 1990). (Reproduced with the permission of McGraw-Hill Publishing Company)

have an impact on tropical and extratropical climate out of the region. Although this is an area of active research, the expected long-range impacts from Amazon deforestation outside the region remain unclear (Gash and Nobre, 1996).

While global fluxes or distributions of water may not be influenced by water quality, humans significantly impact water quality at every step of the hydrologic cycle. Since the 1970s, the primary atmospheric water quality concern has been acid rain. Acid rain damages trees, particularly at high elevations, and contributes to the acidification of lakes and streams. Regions already affected include North America and Northern Europe. Contamination of water at and below the land surface poses a significant threat to potable water supplies (see Fetter, 1998 and Bedient *et al.*, 1999 for further reading on groundwater contamination). Water quality can be affected by human activities in a multitude of ways, including effluent, leeching from landfills, industrial and mining activities, and agricultural fertilizer and pesticide runoff. In particular, the long-term isolation of hazardous radiological byproducts from water supplies poses a special challenge.

GLOBAL CLIMATE CHANGE AND THE HYDROLOGIC CYCLE

As mentioned previously, the magnitude of the impact of human activities on the hydrologic cycle is highly scale-dependent, and few activities manifest themselves on a global scale. However, the extra release of carbon dioxide (CO_2) into the atmosphere and the resulting possible changes in the hydrologic cycle are the subject of a considerable amount of research effort and political attention. A summary of these changes is provided in Table 4. While there is some uncertainty and debate concerning global warming, the theory as to why the hydrologic cycle may be affected is relatively straightforward.

Greenhouse gases in the atmosphere reflect and re-radiate outbound long-wave radiation back to the Earth s surface, increasing the surface temperature. CO_2 is one such gas, while water vapor, methane, and nitrous oxide are among some of the others. The relationship between the concentration of CO_2 in the atmosphere and the temperature of the Earth has been thoroughly documented on geologic

Table 4 Best estimates of climate change projections over the next 50–100 years

Indicators	Annual average change	Distribution of changes		Interannual variability	Significant transients	Confidence of projection	
		Regional average	Change in seasonality			Global average	Regional average
Temperature	+1 to +3.5 °C	−3 to +10 °C	Yes	Down?	Yes	High	Medium
Sea level	+15 to +95 cm	–	No	?	Unlikely	High	Medium
Precipitation	+7 to +15%	−20 to +20%	Yes	Up	Yes	High	Low
Direct solar radiation	−10 to +10%	−30 to +30%	Yes	?	Possible	Low	Low
Evapotranspiration	+5 to +10%	−10 to +10%	Yes	?	Possible	High	Low
Soil moisture	?	−50 to +50%	Yes	?	Yes	?	Medium
Runoff	Increase	−50 to +50%	Yes	?	Yes	Medium	Low
Severe storms	?	?	?	?	Yes	?	?

Source: Schneider *et al.* (1992) (reproduced by permission of John Wiley & Sons) and S Schneider (personal communication, July 27, 2000).

time scales. For instance, during the last glacial period when the atmospheric CO_2 concentration was closer to 200 parts per million by volume (ppmv), the Earth was 6–8 °C cooler than during pre-industrial conditions of 280 ppmv. The current atmospheric concentration of CO_2 is about 370 ppmv, and predictions of future concentrations range from 450–1000 ppmv in the year 2100. The anticipated increase in global surface temperatures over this period is from 1.4 to 5.8 °C (IPCC, 2001).

If the amount of long-wave radiation reflected back to the Earth s surface is increased, the potential evaporation from the ocean and the land surface may also increase. Likewise, because warmer air can potentially hold more water (the amount of water that it would take to saturate a parcel of air increases exponentially with temperature), humidity may increase. If the increase in water vapor storage in the atmosphere does not balance the increased evaporation, precipitation rates should change, on the average (*see* Water Vapor: Distribution and Trends, Volume 1). Due to complex feedbacks within the Earth s climate system, the anticipated impacts of global warming on the hydrologic cycle become considerably more difficult to predict beyond these first few conceptual steps. For example, a more vigorous atmospheric water cycle may involve increased low clouds that might reduce the amount of incoming radiation to the Earth s surface, leading to negative feedback and cooling. However, an increase in high clouds, which are efficient at reflecting long-wave radiation back to the Earth s surface while letting short-wave radiation pass through, would lead to positive feedback of further warming.

If understanding the impact of climate change on all aspects of the hydrologic cycle is difficult, translating these impacts into their effects on human activities is an even greater challenge. Humans are most sensitive to hydrologic variations, extremes, and certain sequences of events. Such questions as how climate change will affect the occurrence of drought in a particular region remain open to debate (although considerable effort has been devoted towards trying to answer these questions, see National Assessment Synthesis Team, 2000). Some scientists believe that the clearest changes associated with global warming will be in the occurrences of extreme events. Assuming that a climatological variable (e.g., precipitation, temperature) has a normal distribution, changes in the mean will bring proportionally greater shifts in intensity and frequency at the tails of the distribution (Trenberth, 1999b; see also Easterling, 2000a for an excellent discussion of the observed and anticipated changes in extreme events induced by climate change).

The atmosphere is not the only component of the hydrologic cycle that may be affected by global climate change; the ocean and cryosphere may also be impacted. Simulations of climate change have predicted that the greatest warming will occur at high latitudes. Locally, this poses a threat to regions with permafrost, where melting can cause land-surface subsidence and damage to structures. However, if temperatures are warm enough, glacial melt and thermal expansion of the oceans also may cause sea level changes over the next 100 years. This is of concern because a 1% decrease in glacial water content translates to approximately a 30 cm increase in sea level. To help put this in perspective, a 2-m rise in sea level could entirely inundate the Republic of the Maldives. Changes in surface albedo associated in the change from ice to bare ground or open ocean is a positive feedback, associated with increased absorption, higher temperatures, and possibly increased melting. The predictions of sea level increase range from 13–94 cm by the year 2100 (IPCC, 1995). Such a change could have dire consequences for coastal ecosystems.

The response of the hydrologic cycle to climate forcing can also be nonlinear. Large temperature rises at high

latitudes, increases in precipitation, and glacial melt may alter the vertical density profile of the North Atlantic Ocean, causing a slowing or collapse of the deep-water formation, greatly impacting the climate of Europe, as was seen during the younger Dryas period. Model simulations of doubling CO_2 concentrations predict a decrease in deep-water formation, with eventual recovery after concentration stabilization, whereas quadrupling concentrations leads to thermohaline collapse and very slow recovery. These simulations are highly model-dependent, but they do emphasize the presence of thresholds and triggers within the climate system (*see* Climate Change, Abrupt, Volume 1).

The expected impacts of climate change on surface runoff and groundwater recharge remain unknown, in part due to the fine spatial scale being considered, and unknown changes in seasonality and timing of precipitation. However, the possible ramifications of these changes should not be underestimated. In regions where water resources are fully committed or over-committed, declines in supply can lead to regional disputes and international conflict. A comprehensive database of articles related to water resources and climate change in the US has been compiled and is available at http://www.pacinst.org/CCBib.html.

20TH CENTURY OBSERVED CHANGES IN THE HYDROLOGIC CYCLE

One of the most difficult challenges in detecting changes in the hydrologic cycle is the lack of globally complete, high-quality long-term measurements. In particular, existing observations of soil moisture and evaporation are unsuitable for studies of long-term changes. Likewise, it is difficult to estimate the underlying climatological frequency of rare events (e.g., floods) from existing data sets, much less detect changes in the frequency or magnitude of these events. Attributing the observed changes to human activities or natural variability represents a special challenge in and of itself. Nonetheless, there are some detectable trends in the modern instrumental precipitation data, primarily towards drier conditions in the tropics and wetter conditions in the extra-tropics. Fewer trends have been detected in streamflow. A summary of recent hydrologic trends is presented in Table 5.

A thorough review of the observed global changes in precipitation is provided by IPCC (1995). A review of changes in the US was done by the National Assessment Synthesis Team (2000). Annual precipitation in the US has increased by 5%, primarily in autumn and in the eastern two–thirds of the country. These increases have been associated with modest increases in heavy rainfall events (i.e., $>5\,\mathrm{cm\,day^{-1}}$, Karl *et al.*, 1995). Streamflow trends are consistent with these increases; geographically widespread regions of the US have experienced increases in the annual median and minimum daily flows. This suggests that, overall, there is more water, in general, in US rivers today and that droughts have not been as extreme as they were in the 1950s–1970s (Lins and Slack, 1999). Easterling *et al.* (2000b) found that precipitation extremes are on the rise in many locations across the globe, but reminded us of the danger of drawing firm conclusions based on analysis of relatively short data records. Likewise, recent changes in the frequency of warm ENSO events and the state of the PDO (since the 1970s) have made sole attribution of the observed climate changes to anthropogenic activities difficult.

Table 5 Observed trends in the hydrologic cycle

Variable	Observed trend	Confidence
Ocean		
High clouds	Increase 1951–1981	Low
Mid-level clouds	Increase in Northern Hemisphere, mid-latitude 1951–1981	Low
Convective clouds	Increase 1951–1981	Low
Fair weather cumulus clouds	Decrease 1951–1981	Low
Water vapor	Increase 1973–1988	Low
Evaporation in tropics	Increase 1949–1989	Medium
Land		
Mid- to high-latitude clouds	Increasing 1900–1980s	Medium
Mid- to high-latitude precipitation	Increasing since 1900	Medium
Northern Hemisphere subtropical precipitation	10% decrease since 1970	Medium
Evaporation in US and FSU[a]	Decreasing since 1950	Low
Soil moisture in FSU[a]	Increasing 1970s–1990s	Low
Runoff	Pattern consistent with precipitation changes	Medium

Source: IPCC, 1995. (Reproduced by permission of the IPCC).
[a] Former Soviet Union.

TOWARDS IMPROVED UNDERSTANDING OF THE HYDROLOGIC CYCLE

Although the hydrologic cycle has been studied for well over a century, with considerable work being done in recent decades, many unresolved questions and expanding

frontiers in water cycle research remain. In particular, understanding the impacts of human activities on the hydrologic cycle is an area of active research. Considerable effort is also being devoted to understanding the linkages between the ocean and atmosphere in the tropics (such as ENSO), as well as at mid-latitudes (through PDO, NAO, and through changes in ocean thermohaline circulation). Increased attention has been devoted to land surface and hydrologic representation in large-scale computer models, the intent of which is to improve predictions of variations in the hydrologic cycle on interannual and longer time scales. Medium-range weather predictions can also improve through having a better understanding and model representation of soil–moisture processes (Chahine, 1992). Finally, groundwater and surface interactions, ranging from the amount of water recharged into an aquifer from snowmelt to the hydrology of natural springs, are some of the least understood components of basin-scale water balances.

Improved understanding of the hydrologic cycle is not limited to the understanding contained within the scientific research community. It is also that of the operational water management community, the general public, and other water resources stakeholders. Concepts such as climate stationarity (the belief that the underlying statistics of the climate of a particular region are constant in time) are key assumptions for most water management planning and design, although the research community has long recognized the flaws in this assumption. For example, for structural design purposes, relatively brief historical records are used to estimate the magnitude of the flood that would happen once in a 100 years. If the historical period being considered contains unusually wet or dry spells, this approach will not give a representative estimate of what may happen in the future. Likewise, most regions lack legal recognition of the connection between the groundwater and surface-water components of the hydrologic cycle. The impacts of excessive groundwater withdrawals on streamflow are fairly well understood by the scientific research community, although few, if any, regions have laws that reflect this understanding.

There are several major research programs designed to develop a more sophisticated understanding of the hydrologic cycle. One such program, the Global Energy and Water cycle EXperiment (GEWEX, *see* GEWEX (Global Energy and Water Cycle Experiment), Volume 1), was initiated in 1988 by the World Climate Research Programme (WCRP, 1990; Chahine, 1992). Part of GEWEX is designed to observe and model the hydrologic cycle, with the ultimate goal of predicting global and regional climate change. The program includes large-scale field activities and intensive measurements, as well as modeling and research. GEWEX has contributed to the development of improved numerical models and the creation of state of the art climate data sets, and it is also helping to achieve its goals of improving resource management through scientific outreach to the engineering and other user communities.

See also: Hydrology, Volume 2; Water Use: Future Trends, and Environmental and Social Impacts, Volume 3; Circulating Freshwater: Crucial Link between Climate, Land, Ecosystems, and Humanity, Volume 5.

ACKNOWLEDGMENTS

We gratefully acknowledge assistance of S Williams and B Imam in data and graphical assistance and C Theis and C Sprout in proofreading and editing the manuscript. Financial support was provided by NASA/EOS Grant #OSSA-A/88, by CLIMAS (Southwest Climate Assessment Program) under NOAA Office of Global Programs Grant #NA86GP0061 and by SAHRA (Sustainability of semi-Arid Hydrology and Riparian Areas) under the STC program of the National Science Foundation, agreement No. EAR-9876800.

REFERENCES

Allan, R, Lindesay, J, and Parker, D (1996) *El Niño Southern Oscillation and Climatic Variability*, CSIRO, Collingwood, VIC, Australia.

Bedient, P B, Rifai, H S, and Newell, C J (1999) *Ground Water Contamination: Transport and Remediation*, Prentice-Hall, Englewood Cliffs, NJ.

Bjerknes, J (1969) Atmospheric Teleconnections from the Equatorial Pacific, *Mon. Weather Rev.*, 97, 163–172.

Calder, I R (1993) Hydrologic Effects of Land Use Change, in *Handbook of Hydrology*, ed D R Maidment, McGraw-Hill, New York, 13.1–13.50.

Chahine, M T (1992) The Hydrological Cycle and Its Influence on Climate, *Nature*, 359, 373–380.

Changnon, S A and Bell, G D, eds (2000) *El Niño, 1997- 1998: the Climate Event of the Century*, Oxford University Press, New York.

Changnon, S A and Vonnhame, D R (1986) Use of Climate Predictions to Decide a Water Management Problem, *Water Resour. Bull.*, 22, 649–652.

Charney, J (1975) Dynamics of the Deserts and Drought in the Sahel, *Q. J. R. Meteorol. Soc.*, 101, 193–202.

Easterling, D R, Meehl, G A, Parmesan, C, Changnon, S A, Karl, T R, and Mearns, L O (2000a) Climate Extremes: Observations, Modeling, and Impacts, *Science*, 289, 2068–2074.

Easterling, D R, Evans, J L, Groisman, P Y, Karl, T R, Kunkel, K E, and Ambenje, P (2000b) Observed Variability and Trends in Extreme Climate Events: a Brief Review, *Bull. Am. Meteorol. Soc.*, 81, 417–425.

Fetter, C W (1998) *Contaminant Hydrogeology*, Prentice-Hall, Upper Saddle River, NJ.

Flohn, H (1973) Globale Energiebilanz und Klimaschwankungen, in *Bonner Meteorologische Abhandlungen, Westdeutscher Verlog*, 75–117.

Gash, J H C and Nobre, C A, eds (1996) *Amazonian Deforestation and Climate*, John Wiley & Sons Ltd, New York.

IPCC (1995) *Climate Change 1995: The Science of Climate Change*, Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge.

IPCC (2001) *Climate Change 2001: The Scienti c Basis*, eds J T Houghton *et al.*, Cambridge University Press, Cambridge, 1–881.

Karl, T R, Knight, R W, and Plummer, N (1995) Trends in High-frequency Climate Variability in the Twentieth Century, *Nature*, 377, 217–220.

Korzun, V I (1978) *World Water Balance and Water Resources of the Earth*, UNESCO.

Kunkel, K E, Pielke, R A J, and Changnon, S A (1999) Temporal Fluctuations in Weather and Climate Extremes that Cause Economic and Human Health Impacts: A Review, *Bull. Am. Meteorol. Soc.*, 80, 1077–1098.

Lau, W K M, Sud, Y C, and Kim, J H (1995) *Intercomparison of Hydrologic Processes in Global Climate Models*, Technical Memorandum 104617, NASA, Washington, DC.

Lins, H F and Slack, J R (1999) Streamflow Trends in the United States, *Geophys. Res. Lett.*, 26, 227–230.

National Assessment Synthesis Team (2000) *Climate Change Impacts on the United States: The Potential Consequences of Climate Variability and Change*, US Global Change Research Program, Cambridge University Press, New York.

Oki, T (1999) The Global Water Cycle, in *Global Energy and Water Cycles*, eds K A Browning and R J Gurney, Cambridge University Press, Cambridge, 10–29.

Piexoto, J P and Oort, A H (1992) *Physics of Climate*, American Institute of Physics, New York.

Ropelewski, C F and Halpert, M S (1987) Global and Regional Scale Precipitation Patterns Associated with the El Niño/Southern Oscillation, *Mon. Weather Rev.*, 115, 1606–1626.

Schneider, S H, Mearns, L O, and Gleick, P H (1992) Climate-Change Scenarios for Impact Assessment, in *Global Warming and Biological Diversity*, eds R Peters and T Lovejoy, Yale University Press, New Haven, CT, 38–55.

Sellers, W D (1965) *Physical Climatology*, University of Chicago, Chicago, IL.

Shiklomanov, I (1999) *World Water Resources and their Use*, St Petersburg, Russian Federation, State Hydrological Institute/UNESCO [CD-ROM].

Shiklomanov, I A (1997) *Comprehensive Assessment of the Freshwater Resources of the World: Assessment of Water Resources and Water Availability in the World*, Report 556.18 SHI, World Meteorological Organization, Geneva, Switzerland.

Sorooshian, S, Gupta, H V, and Rodda, J C, eds (1996) *Land Surface Processes in Hydrology*, Springer-Verlag, New York.

Sorooshian, S, Hsu, K, Gao, X, Gupta, H V, Imam, B, and Braithwaite, D (2000) Evaluation of PERSIANN System Satellite-Based Estimates of Tropical Rainfall, *Bull. Am. Meteorol. Soc.*, 81(9), 2035–2046.

Stern, P C and Easterling, W E, eds (1999) *Making Climate Forecasts Matter*, National Academy Press, Washington, DC.

Trenberth, K E (1999a) Atmospheric Moisture Recycling: Role of Advection and Local Evaporation, *J. Clim.*, 12, 1368–1381.

Trenberth, K E (1999b) Conceptual Framework for Changes of Extremes of the Hydrological Cycle with Climate Change, *Clim. Change*, 42, 327–339.

Urbonas, B R and Roesner, L A (1993) Hydrologic Design for Urban Drainage and Flood Control, in *Handbook of Hydrology*, ed D R Maidment, McGraw-Hill, New York, 28.1–28.52.

US Army Corps of Engineers (1997) *Weather and Climate Extremes*, TEC-0099, National Technical Information Service, Springfield, VA.

van Dam, J C, ed (1999) *Impacts of Climate Change and Climate Variability on Hydrological Regimes*, Cambridge University Press, Cambridge.

Ward, R C (1990) *Principles of Hydrology*, McGraw-Hill, London.

WCRP (1990) *Scienti c Plan for the Global Energy and Water Cycle Experiment*, WCRP-40, World Climate Research Programme, Geneva.

Wisler, C O and Brater, E F (1959) *Hydrology*, John Wiley & Sons, New York.

Hydrology

The science of water, the study of the distribution, conservation and use of the water of the Earth: these are examples of some of the dictionary definitions of *hydrology*. They are usually too broad, however. As an earth science, hydrology invariably deals with fresh and brackish water. Salt water belongs to oceanography, while the study of water in the atmosphere is part of meteorology. Indeed some would argue that hydrology starts where meteorology stops, but there is an area of overlap termed *hydrometeorology*. There are similar overlaps with a number of other sciences, e.g., that with geology is *hydrogeology*. Hydrology developed largely as an applied science, but in recent years it has metamorphosed into a distinct discipline.

The purview of hydrology is best described by two widely accepted definitions (UNESCO/WMO, 1992).

1. Science that deals with the waters above and below the land surfaces of the Earth, their occurrence, circulation and distribution, both in time and space, their biological, chemical and physical properties, their reaction with the environment, including their relation to living things.
2. Science that deals with the processes governing the depletion and replenishment of the water resources of the land areas of the Earth, and treats the various phases of the hydrological cycle.

Hydrology can also be seen through the eyes of those bodies with programs in the sciences. For example, the International Association of Hydrological Sciences (*see* IAHS (International Association of Hydrological Sciences), Volume 1) deals with: surface water, groundwater, erosion, snow and ice, water quality, water resources systems and atmosphere–soil–vegetation relations and the use

of remote sensing and tracers. The Association focuses on the study of the hydrological cycle on the Earth and the waters of the continent; including their physical chemical and biological processes, their relation to climate and to other physical and geographical factors as well as the inter-relations between them. Assessment of water resources, hydrological forecasting and flood and drought prediction, for design purposes are germane to its purpose.

See also: Hydrologic Cycle, Volume 1.

REFERENCE

UNESCO/WMO (1992) *International Glossary of Hydrology*, 2nd edition, Paris, Geneva.

JOHN RODDA *UK*

I

IAHS (International Association of Hydrological Sciences)

The IAHS is the oldest and foremost international non-governmental organization (NGO) in the fields of hydrology and water resources. Incorporating the International Commission on Glaciers, which had been set up in 1894, it was established in 1922 with the aim of bringing together hydrologists from all countries to promote the hydrological sciences.

IAHS promotes the study of all aspects of hydrology; fosters discussion, comparison and publication of research results; and initiates and coordinates research that requires international cooperation. To fulfill these aims, the association organizes assemblies, symposia and workshops in various parts of the world, publishes the proceedings of these meetings as well as a scientific journal, contributes to a wide range of international initiatives and generally fosters activities and collaboration in hydrology and water resources.

IAHS is one of the seven associations that form the International Union of Geodesy and Geophysics (IUGG), which was established in 1919 to promote physical, chemical and mathematical studies of the Earth and its spatial environment. IUGG is itself one of the 20 scientific unions that are grouped within the International Council for Science (ICSU), an international NGO, which was established in 1931 to promote international scientific activity for the benefit of mankind.

IAHS maintains close collaboration with the United Nations Educational, Scientific and Cultural Organization (UNESCO) and the World Meteorological Organization (WMO) through their respective hydrology and water resources programs, the International Hydrological Programme (IHP) of UNESCO, and the Hydrology and Water Resources Programme (HWRP) of WMO.

For further information, contact Dr Pierre Hubert, Secretary General, IAHS, Ecole des Mines de Paris, 35 rue Saint Honoré, 77305 Fontainebleau, France.

JOHN S PERRY *USA*

IAMAS (International Association of Meteorology and Atmospheric Sciences)

The IAMAS is one of seven associations making up the International Union of Geodesy and Geophysics (IUGG), which is in turn a component of the International Council for Science (ICSU). These bodies provide an organizational structure for the scientists of the world to come together to present, discuss, and promote the newest achievements in their fields of research. The objectives of IAMAS are to promote the study of all problems relating to the sciences of the atmosphere; initiate, facilitate, and coordinate research that requires international cooperation; and to stimulate discussion and provide for publication of the results of research in this field. The IAMAS is organized into ten commissions, each of which focuses on a particular aspect of this research field. The international commissions presently active are the International Commission of Atmospheric Chemistry and Global Pollution (ICACGP), International Commission for Atmospheric Electricity (ICAE), International Commission of Climate (ICCl), International Commission for Clouds and Precipitation (ICCP), International Commission for Dynamic Meteorology (ICDM), International Commission of Meteorology for the Middle Atmosphere (ICMMA), International Commission for Planetary Atmospheres and their Evolution (ICPAE), International Commission for Polar Meteorology (ICPM), International Ozone Commission (IOC), and the International Radiation Commission (IRC).

A major role of IAMAS and the commissions is to organize international conferences and symposia. In addition, IAMAS has recently undertaken a major activity in cooperation with the World Meteorological Organization to strengthen linkages with universities around the world as a means of improving the ability of scientists in this field to respond to society s needs for information. Further information is available at http://iamas.org/.

MICHAEL C MACCRACKEN *USA*

IAPSO (International Association for the Physical Sciences of the Oceans)

The IAPSO is one of seven associations making up the International Union of Geodesy and Geophysics (IUGG), which is in turn a component of the International Council for Science (ICSU). These bodies provide an organizational structure for the scientists of the world to come together to present, discuss and promote the newest achievements in their fields of research. IAPSO became a separate organization in 1929 and has held periodic assemblies every 2–4 years since that time (except during World War II). The primary goal of IAPSO is to promote study of scientific problems relating to the oceans and the interactions taking place at the sea floor, coastal, and atmospheric boundaries. To address this goal, IAPSO organizes international meetings and fora, establishes commissions to encourage and co–ordinate research activities, provides basic services significant to the conduct of physical oceanography and publishes proceedings and reference materials. IAPSO commissions include the Commission on Mean Sea Level and Tides, the Commission on Sea Ice, the Commission on Cooperation with Developing Countries, and the Tsunami Commission, which is jointly sponsored by the IUGG s International Association of Seismology and Physics of the Earth s Interior. The IAPSO also is a sponsor of the Permanent Service for Mean Sea Level and the IAPSO Standard Seawater Service. IAPSO also maintains formal liaison with ICSU s Scientific Committee on Oceanic Research (SCOR) and United Nations Educational, Scientific and Cultural Organization s (UNESCO) Intergovernmental Oceanographic Commission (IOC). Further information is available at http://www.olympus.net/iapso/.

MICHAEL C MACCRACKEN *USA*

Icebergs

Icebergs are large pieces of floating glacier ice (not sea ice) that protrude more than 5 m above the water and are more than about 10 m across. Smaller pieces of floating glacier ice are called *growlers* if they are almost awash, protruding no more than about a meter, or *bergy bits*, if they are intermediate in size. Eighty to ninety percent of the mass of a freely floating iceberg lies below the water surface. Often, due to faster melting of the ice above water, an iceberg has an underwater projection, dangerous to ships, called a *ram*.

Icebergs form by calving from the termini of glaciers where they meet a lake or the ocean. Antarctic ice shelves calve huge tabular (flat-surfaced) icebergs; the largest ever reported was twice the area of Connecticut. Icebergs in Antarctica mostly remain in the waters near the continent and are seldom found in shipping lanes. This is not so in the Northern Hemisphere. Icebergs from Greenland are commonly found in the North Atlantic and a rapid retreat of the terminus of Columbia Glacier in Alaska led to a large influx of icebergs into Prince William Sound, which is used by supertankers carrying oil from the Alaska pipeline.

Icebergs are responsible for almost all of the ice loss from Antarctica ($2000\,km^3\,year^{-1}$) and much of that from Greenland ($300\,km^3\,year^{-1}$). Exact discharge rates are not known, partly because the calving of large icebergs is infrequent. An ice front may creep forward for decades or centuries and then re-establish its earlier configuration in a single calving event. Conversely, the calving of a state-sized iceberg does not imply climatic change. Glaciologists do not have a satisfactory predictive model of calving and consequently cannot forecast the effect of climatic change on iceberg production. However, it is noteworthy that several small ice shelves in the northern Antarctic Peninsula have broken up into icebergs in the last decade, during a pronounced warming in that region.

See also: Antarctica, Volume 1; Glaciers, Volume 1.

CHARLES BENTLEY *USA*

IGAC (International Global Atmospheric Chemistry)

The IGAC Project was initiated in 1988 by the Commission on Atmospheric Chemistry and Global Pollution of the International Association of Meteorology and Atmospheric Sciences (IAMAS) in response to growing concern about observed changes in atmospheric composition and their potential impacts on humankind. IGAC became a Core Project of the International Geosphere–Biosphere Programme in 1990 (*see* IGBP Core Projects, Volume 2).

IGAC s goals are to:

- develop a fundamental understanding of the processes that determine atmospheric composition;

- understand the interactions between atmospheric chemical composition and physical, biospheric, and climatic processes;
- predict the impact of natural and anthropogenic forcings on the chemical composition of the atmosphere.

The atmosphere is chemically complex and dynamic, interacting internally between troposphere and stratosphere, and externally with oceans, land, and biota. Atmospheric composition has changed over geologic time, and ice core records indicate that it has changed markedly so over at least the last 400 000 years. Human activities have raised concentrations of radiatively active greenhouse gases and aerosols, and diminished the ozone layer, potentially impacting the biosphere.

After atmospheric carbon dioxide (CO_2), which undergoes exchanges with the land and ocean but is not chemically altered in the atmosphere, the most important long-lived greenhouse gas is methane. Methane has major natural and anthropogenic sources and is destroyed mostly by oxidation in the atmosphere. Nitrous oxide (N_2O), with a similar range of sources, and the purely anthropogenic chlorofluorocarbons are also greenhouse gases, although their climate forcing is offset partially by the changes in concentration of stratospheric ozone they destroy.

Ozone is a key atmospheric chemical and protective ultraviolet shield. Its chemistry is influenced by many other trace species. It is also toxic to humans and vegetation, and is an important greenhouse gas.

Gaseous sulfur compounds, both natural and anthropogenic, are oxidized to form sulfate aerosols which have an important effect on albedo, partly counteracting climate forcing by the greenhouse gases.

Atmospheric chemistry is closely linked to industrial activity, climate, and land-use through many interrelated processes. IGAC is dedicated to understanding this complex system through a combination of observations, theory, and laboratory and modeling studies.

ALEX PSZENNY *USA*

IGBP (International Geosphere–Biosphere Programme)

see IGBP (International Geosphere–Biosphere Programme) (**Volume 2**); IGBP Core Projects (*Opening essay*, **Volume 2**)

IGY (International Geophysical Year)

The IGY (1958–1959) was a highly successful international program conducted with the goal of observing geophysical phenomena and securing data from all parts of the world, and conducting this effort on a coordinated basis by fields, and in space and time, so that results could be collated in a meaningful manner.

In 1952, the International Council of Scientific Unions (later renamed the International Council for Science (ICSU)) proposed the IGY, modeled on the International Polar Years of 1882–1883 and 1932–1933. The aim was to allow scientists from around the world to take part in a series of coordinated observations of various geophysical phenomena employing rapidly advancing observational techniques, notably the then-new resources of space technology.

The IGY was overseen by an ICSU Special Committee for the IGY, which acted as the program s governing body. Supporting activities were organized in many countries. For example, a US National Committee, under the aegis of the US National Academy of Sciences, organized US participation in the IGY. Specialized groups developed programs in many areas of geophysics: aurora and airglow, cosmic rays, geomagnetism, glaciology, gravity, ionospheric physics, longitude and latitude determination, meteorology, oceanography, rocketry, seismology, and solar activity. A major goal, achieved by the Soviet Union followed shortly by the US, was to launch the first artificial Earth satellites. In successfully launching science into space, the IGY may have scored its greatest breakthrough.

IGY activities literally spanned the globe from the North to the South Poles, and ultimately 67 countries participated. Although much work was carried out in the arctic and equatorial regions, special attention was given to the Antarctic, where research contributed to improved meteorological prediction, advances in the theoretical analysis of glaciers, and better understanding of seismological phenomena in the Southern Hemisphere.

JOHN S PERRY *USA*

IHP (International Hydrological Program)

Water – the source of life and human civilization – is surely one of the major issues of the 21st century. During

the coming years, problems of water quality and availability will affect the lives of virtually everyone on the planet. Regions of the world facing water shortages are growing in area and number while the demand for water is estimated to have risen six or seven times from 1900 to 1995 – more than double the rate of population growth. Issues concerning the world s fresh water highlight the dilemma facing humankind. Can competition between the environment and development be transformed into a partnership between the two, so that the goal of sustainable development is met?

The IHP is an intergovernmental applied research and education activity under United Nations Educational, Scientific and Cultural Organization (UNESCO) auspices and is designed to help nations to improve their knowledge of the water cycle and thereby increase their capacity to better manage and develop their water resources. It aims at the improvement of the scientific and technological basis for the development of methods for the rational management of water resources.

The program started as the International Hydrological Decade (IHD, 1965–1974) and was followed by the IHP as a long-term program executed in phases of a six-year duration. The general objective of the IHD, and later of the IHP, is to improve the scientific and technological basis for the development of methods and the human resource base for the rational management of water resources, including the protection of the environment. Over time, the program has evolved from being research oriented to placing increased emphasis on practical aspects of hydrology and water resources. The program s major current theme is *Hydrology and Water Resources Development in a Vulnerable Environment*. Eight subsidiary themes have been identified that cut across different hydrological scales and different climatic regions, but have integrated water management in a vulnerable environment as a common issue:

- global hydrological and biochemical processes;
- ecohydrological processes in the surficial zone;
- groundwater resources at risk;
- strategies for water resources management in emergency and conflicting situations;
- integrated water resources management in arid and semi-arid zones;
- humid tropics hydrology and water management;
- integrated urban water management;
- transfer of knowledge, information and technology.

The IHP carries out a wide variety of activities to address these issues. For example, IHP sponsors postgraduate hydrology courses in the field of hydrology and water resources management with emphasis on a interdisciplinary approach. It is also engaged in preparing computer-based learning materials for *distance learning* (*see* Distance Learning and Environment, Volume 4) and, in cooperation with other United Nations agencies, a broad range of publications.

At the global level the intergovernmental council for IHP plans, coordinates and monitors all IHP activities, which are executed by the secretariat either directly or with the help of committees, working groups or rapporteurs. At the national level, the IHP is coordinated and executed by national committees. The UNESCO regional offices are responsible for regional implementation.

For further information, contact Dr A Szollosi-Nagy, UNESCO, 7 Place de Fontenoy, 75352 Paris Cedex-07, France.

JOHN S PERRY *USA*

Imbrie, John

(1925-)

John Imbrie, US paleoclimatologist and paleoceanographer masterminded the analysis of paleoceanographic data to demonstrate that the Earth s orbital variations in seasonal insolation paced the multiple glacial/interglacial cycles of the current ice age. Having begun his career as a Paleozoic brachiopod paleontologist with geology degrees from Princeton (BA in 1948) and Yale (PhD in 1951), Imbrie achieved considerable recognition for his contributions to paleoecology and for introducing multivariate statistical methods to the study of sediments. He did this research while teaching at Columbia University from 1952–1967, where he became chair of the Geology Department in 1966. When he moved to Brown University in 1967, he switched to studying microfossils in the Quaternary sediments of marine cores. He saw a major paleoclimatic opportunity to study the past distribution of microscopic marine organisms within the vast core collection at the Lamont–Doherty Geological Observatory at Columbia University. Working with Nilva Kipp as his assistant and with the numerous collaborators within *Climate: Mapping and Prediction* (CLIMAP) (*see also* CLIMAP (Climate: Long-range Investigation, Mapping, and Prediction), Volume 1), Imbrie helped found and direct this National

Science Foundation-sponsored multi-institutional project that extended from 1971 to 1981 within the International Decade of Ocean Exploration, Imbrie was co-author on three papers (Imbrie and Kipp, 1971; Hays *et al.*, 1976; and CLIMAP, 1976) that gained him international recognition and election to the National Academy of Sciences in 1978. In addition to providing convincing evidence for the Milankovitch hypothesis of orbitally forced glacial/interglacial cycles (*see* Milankovitch, Milutin, Volume 1; Orbital Variations, Volume 1), these papers also showed how to estimate sea-surface temperatures from microfossils in marine cores, and provided the boundary conditions for running a global climate model to simulate the climates of the last glacial maximum, 21 000 years ago.

From 1973 to 1975, Imbrie served on a National Research Council panel that developed a 10-year research program for advancing understanding of the climate system. With George Denton and Wallace Broecker, Imbrie wrote an appendix to the main panel report that helped shape paleoclimatic research for the next decade, during which paleoclimatology became central to an understanding of climate dynamics in atmospheric and ocean sciences. Imbrie and Imbrie (1979) provide a highly readable summary of the historical development of the astronomical theory of ice ages and of Imbrie s contributions to it. This book won the Phi Beta Kappa Science Award for science writing. From 1981 to his retirement in 1990, Imbrie helped direct SPECMAP, another multi-institutional project concerned with spectral analysis of 100 000-year or longer paleoclimatic time series. This project resulted in multi-authored papers (Imbrie *et al.*, 1992, 1993) summarizing the structure and origin on the major glaciation cycles. He was a MacArthur Fellow from 1981–1985 and has received numerous honors and medals.

REFERENCES

CLIMAP (1976) The Surface of the Ice-Age Earth, *Science*, 191, 1131–1137.

Hays, J D, Imbrie, J, and Shackleton, N J (1976) Variations in the Earth s Orbit: Pacemaker of the Ice Ages, *Science*, 194, 1121–1132.

Imbrie, J and Imbrie, K P (1979) *Ice Ages: Solving the Mystery*, Enslow Press, Short Hills, NJ.

Imbrie, J and Kipp, N (1971) A New Micropaleontological Method for Quantitative Paleoclimatology: Application to a Late Pleistocene Caribbean Core, in *The Late Cenozoic Glacial Ages*, ed K Turekian, Yale University Press, New Haven, CT, 71–181.

Imbrie, J, Boyle, E A, Clemens, S C, Duffy, A, Howard, W R, Kukla, G, Kutzbach, J E, Martinson, D G, McIntyre, A, Mix, A C, Molfino, B, Morley, J J, Peterson, L C, Pisias, N G, Prell, W L, Raymo, M E, Shackleton, N J, and Toggweiler, J R (1992) On the Structure and Origin of Major Glaciation Cycles, Part 1: Linear Responses to Milankovitch Forcing, *Paleoceanography*, 7, 701–738.

Imbrie, J, Boyle, E A, Clemens, S C, Duffy, A, Howard, W R, Kukla, G, Kutzbach, J E, Martinson, D G, McIntyre, A, Mix, A C, Molfino, B, Morley, J J, Peterson, L C, Pisias, N G, Prell, W L, Raymo, M E, Shackleton, N J, and Toggweiler, J R (1993) On the Structure and Origin of Major Glaciation Cycles, Part 2: the 100 000 Year Cycle, *Paleoceanography*, 8, 699–735.

THOMPSON WEBB, III *USA*

Infrared Radiation

V Ramaswamy

Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA

Infrared radiation is one of the two critical components in the heat balance of the planet taken as a whole, being distinct from the other component (viz., solar radiation). It is comprised of radiative energy emitted by the Earth's surface and the atmosphere. In contrast to solar radiation, which is a source of heat from the Sun into the climate system, the infrared radiation represents a loss of heat from Earth. The absorption and emission of infrared radiation by the Earth's surface and atmosphere constitute the basis for the greenhouse effect, and play a critical role in governing the climate of the planet.

CONCEPTS

Infrared radiation (also known as longwave, terrestrial or thermal infrared radiation) consists of electromagnetic waves propagating at the velocity of light ($3 \cdot 10^8\ m\ s^{-1}$), with wavelengths greater than about 3 μm (micrometers or microns). Defined in this manner, the outgoing radiation from Earth also includes the traditional microwave spectrum. Instead of wavelength, the infrared radiation can be described in terms of frequency space, where a unit commonly used is the wavenumber, which is inversely proportional to wavelength and is expressed in inverse centimeter units. In these units, the outgoing infrared radiation consists of wavenumbers less than about 3000 cm^{-1}. From quantum mechanics, the propagation of electromagnetic radiation can also be thought of as occurring in the form of discrete particles called photons. The spectrum of Earth s longwave radiation is almost completely separated from that of solar radiation, with most of the energy contained

in the 5–100 μm range (or 2000–100 cm^{-1}). Longwave radiation is absorbed and emitted by the planetary surface and by atmospheric constituents including gases, aerosols, and clouds. On an annual basis, the longwave irradiance or flux (energy per unit area per unit time) emitted by the Earth s surface–atmosphere system, very nearly balances the net input of solar radiation, such that the climate system is in radiative equilibrium in the absence of externally applied forcing (e.g., due to natural or human induced factors).

The fundamental laws governing longwave radiation are formulated with respect to a black body, i.e., a conceptually idealized object that is a perfect absorber of radiation of all wavelengths incident on it and that also emits the maximum energy possible at a given temperature. Planck s law states that the intensity (energy per unit time per unit area per unit solid angle per unit wavelength) of radiation emitted by a black body at any wavelength is uniquely determined by its temperature; the emitted radiation is isotropic, i.e., the intensity is independent of direction. The total irradiance emitted by a black body integrated over all wavelengths, given by the Stefan–Boltzmann law, is proportional to the fourth power of the absolute temperature. According to Wien s law, the wavelength of maximum radiation by a black body is inversely proportional to its absolute temperature.

In practical terms, the emission and absorption of radiation by the surface and atmospheric constituents (e.g., molecules and particles) are described with reference to a black body. A measure of how strongly a body radiates at any wavelength is given by the emissivity (a fraction between zero and one), which is the ratio of the actual emission to a black body emission at the same wavelength and temperature. Likewise, the absorptivity of a material is defined as the ratio of the radiation absorbed by that material to that absorbed by a black body. According to Kirchhoff s law, in thermodynamic equilibrium (valid below an altitude of about 70 km), any selective absorber of radiation at a given wavelength is also a selective and an equally effective emitter of radiation at the same wavelength. At any wavelength, the intensity of the radiation emitted by a non-black body is the product of its emissivity and the black body emission at that temperature.

ATMOSPHERIC GASES

The absorption and emission of longwave radiation by atmospheric gases occurs at specific wavelengths depending on their atomic and molecular structure. The atmospheric gaseous constituents can be viewed as entities possessing energy states associated with the vibrations of the atoms about their mean positions in a molecule and the rotation of the molecules about their center of mass. Further, according to quantum mechanics, electrons can occupy only certain orbital configurations within atoms, and only certain vibrational and rotational states are permitted for a molecule. Specific combinations of the electron orbits and the vibrational and rotational states determine the energy level of a molecule. A molecule may make a transition to a higher energy level (an excited state) by absorbing photons, or through collisions with other molecules. Likewise, it may proceed from a higher to a lower energy level by emitting photons either spontaneously or via collisions. The interaction of a molecule with photons is quantized, in that only certain discrete changes in energy levels are permitted, and these remain the same for both absorption and emission. Both the wavelengths (or frequencies) at which such transitions are possible and the energies associated with them are characteristic for each molecule, giving rise to line like features of distinct strength (absorption lines) in the observed spectra of the molecule. The individual spectral lines of the molecules are not sharp, but rather have finite widths in wavelength space, according to the Heisenberg uncertainty principle. Within the atmosphere, the width of the spectral lines is further broadened due to the random thermal motions of the molecules (Doppler broadening) and frequent collisions between molecules (collisional broadening). Below about 30 km altitude, the widths are determined largely by collisional broadening. For any specific molecule, a large series of closely spaced spectral lines in a narrow interval of the spectrum is referred to as its absorption band. Isotopes of molecules can also possess significant absorption bands; these bands are displaced in wavelength with respect to that of the parent molecule depending on the mass of the isotope.

If radiation at a particular wavelength incident on a gas cannot excite its atoms and molecules, energy is neither absorbed nor emitted at that wavelength. The major constituents of the atmosphere by volume viz., nitrogen (N_2) and oxygen (O_2), have no absorption bands in the longwave spectrum. Thus, only gases present in trace amounts in the atmosphere are infrared active and responsible for the longwave radiative process. Molecules that have significant absorption bands in the 5–100 μm wavelength (2000–100 cm^{-1}) region include water vapor (H_2O), carbon dioxide (CO_2), ozone (O_3), methane (CH_4), nitrous oxide (N_2O), and the halocarbons (such as chlorofluorocarbons, CFCs); these constitute the so-called greenhouse gases. Water vapor is the strongest longwave gaseous absorber, absorbing virtually throughout the longwave spectrum. It has strong absorption bands extending from 17 μm (588 cm^{-1}) toward longer wavelengths, and also in the 5–8 μm (2000–1250 cm^{-1}) region. In other spectral regions, water vapor absorption, though not as strong, is not negligible.

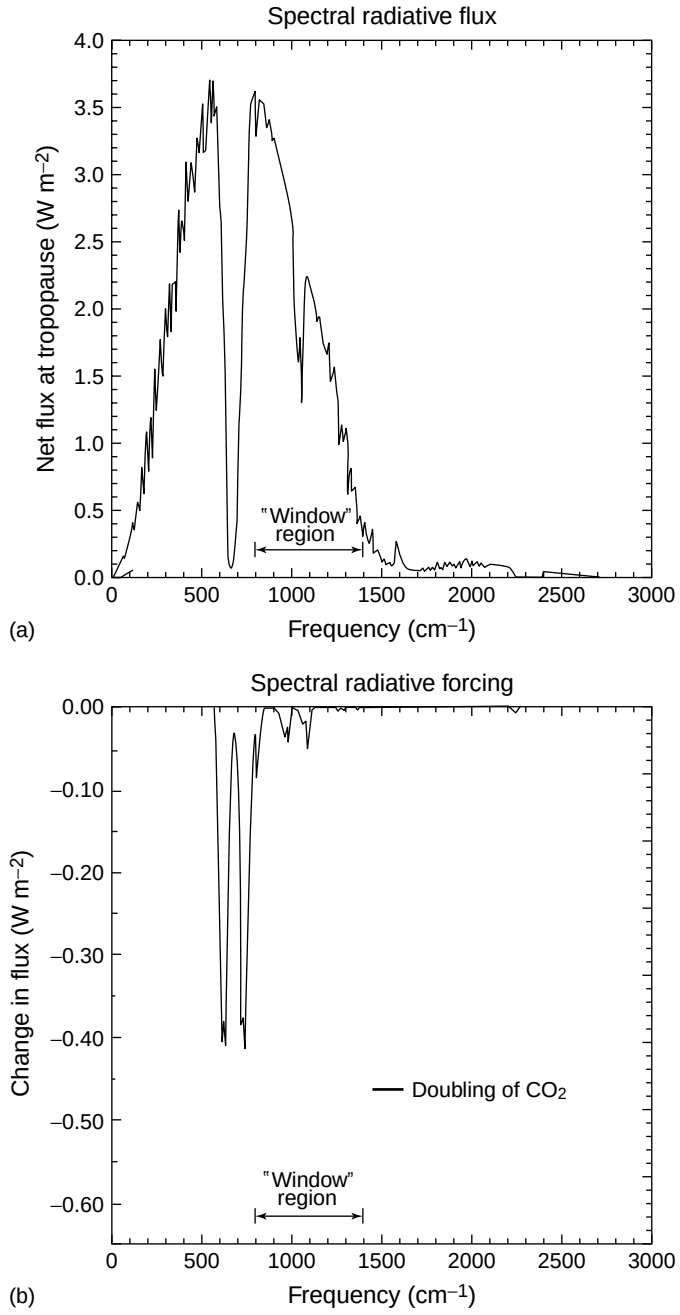


Figure 1 (a) Infrared radiation spectra of the net outgoing radiative flux at the tropopause (i.e., surface-troposphere system) in clear sky, typical mid-latitude summer conditions containing water, carbon dioxide and ozone. The concentration of carbon dioxide (300 ppmv) corresponds approximately to the value at the beginning of the 20th century, while the other gases are held at 1980 climatological values in this calculation. The so-called *window region* is labeled. The spectral absorption by the three gases becomes evident by examining the reduction seen in the net outgoing flux at the respective band locations of the concerned gases (see text). (b) Change in the spectral flux due to an instantaneous doubling of the carbon dioxide concentration. This is referred to as a radiative forcing of the surface-troposphere system. Note that the center of the carbon dioxide band (667 cm^{-1}) does not yield as substantial a reduction in the flux as do the wings of the band. This is because the center of the band is strongly absorbing and is already quite saturated for the base concentration assumed, so that additional absorption in this region due to an increase in carbon dioxide leads to only a small decrease in the spectral flux (or, equivalently, a small enhancement of the spectral greenhouse effect). However, away from the band center and in the wings, where the absorption strength of the molecule is less relative to the center, there is no saturation. Here, there results a substantially larger decrease of the spectral outgoing infrared flux (or a larger enhancement of the spectral greenhouse effect) due to the increase in carbon dioxide

The strongest carbon dioxide absorption bands are centered in the 15 μm (667 cm^{-1}) region (see Figure 1), with a minor band at about 10 μm (1000 cm^{-1}). Among the gaseous absorbers, water vapor and carbon dioxide together dominate the longwave radiative process in the troposphere. Ozone has an important band centered at 9.6 μm (1042 cm^{-1}), which is especially important for the longwave radiative process in the stratosphere. Among the species present in lesser concentrations, methane has a band centered at 7.8 μm (1282 cm^{-1}), nitrous oxide has bands centered at 4.5 (2222 cm^{-1}), 7.8 (1282 cm^{-1}), and 17 μm (588 cm^{-1}), and the halocarbon molecules (e.g., CFCs) absorb in the 7–14 μm (1429–714 cm^{-1}) region. The 8–12 μm (1250–833 cm^{-1}) spectral region in general exhibits the weakest gaseous absorption. At any wavelength, the absorptivity of a gas depends on its concentration and on the strength of its absorption band. The intensity of the emitted radiation depends on the absorptivity and temperature of the gas.

ATMOSPHERIC PARTICLES

Particles in the atmosphere also absorb and emit longwave radiation. Particles active in this process include aerosols (such as sulfate, sea salt, dust, and soot), water drops and ice crystals. The absorption and emission of longwave radiation by these particles depends on their size, shape, and composition. Aerosols, with the exception of dust and soot, consist essentially of spherical particles; water clouds also consist of spherical drops, but ice clouds contain nonspherical crystals. The longwave absorption property of the particles varies with wavelength, although this is less pronounced than with the gaseous absorbers. Within liquids and solids, the absorption and emission takes place throughout a continuous spectrum of wavelengths, in contrast to molecular absorption and emission. The region of most significance for longwave radiative interactions with particles is the 8–12 μm (1250–833 cm^{-1}) portion of the spectrum, because this is where gaseous absorption is weakest, while aerosols and clouds have moderate to strong absorption features here. As in the case of gases, the absorptivity of particles, at any wavelength, is governed by the absorption strength and concentration of the species. Clouds composed of water drops in the lower portions of the atmosphere tend to act as black bodies; that is, they are almost completely absorbing at all wavelengths in the longwave spectrum. In general, clouds composed of ice crystals and aerosol layers tend not to behave as black bodies and have absorptivities less than unity. Again as for gases, the intensity of radiation emitted by a particle at any wavelength depends on absorptivity and temperature. For most practical applications, the various types of surfaces on Earth can be considered to have uniform properties and to absorb and emit longwave radiation like black bodies. This assumption holds true for water, snow, and vegetated surfaces; the exceptions are desert surfaces, whose longwave properties depart from that of a black body.

RADIATION PROCESS

In addition to the selective characteristics of spectral absorption and emission by atmospheric species and the surface, the amount of longwave radiation absorbed, transmitted, and emitted in the atmosphere is governed by the vertical variation of the atmospheric mass, the species concentrations, and the temperature. If the total absorptivity at any wavelength due to all the species is less than unity, a portion of the radiation at that wavelength is transmitted. If there is no absorption at that wavelength, the radiation passes unimpeded. Thus, radiation emitted under cloudless skies from a black body surface is strongly absorbed in the major bands of the various gases, notably water vapor and carbon dioxide; a certain amount is emitted by these gases in the same bands. The net effect at the top of the atmosphere is a decrease of the radiation in these bands compared with that emitted by the surface. In contrast, a large part of the 7–13 μm (approximately 1400–800 cm^{-1}) radiation emitted from the surface escapes to tropopause and then to the top of the atmosphere, to space. This spectral region is commonly referred to as a window region (see Figure 1). The portion of the longwave radiation that does not escape to space is trapped by the infrared active gases, leading to a warming of the surface and the troposphere (*see* Greenhouse Effect, Volume 1).

At different altitudes, the absorption and emission processes, when integrated over all wavelengths, result in either a net loss of longwave radiative energy (radiative cooling) or a net gain (radiative warming). This is in contrast to solar radiation, which always acts to warm the atmosphere. If more radiation is emitted than is absorbed in any region, radiative cooling occurs; the converse leads to radiative warming. In the clear sky troposphere (below about 12 km), longwave cooling is essentially due to water vapor and carbon dioxide. The cooling decreases with height, becoming small near the tropopause.

In cloudy atmospheres, layers containing thick clouds tend to behave very nearly as black bodies, especially in the case of water clouds. For such layers, the general tendency is for a radiative heating near the base of the cloud (due essentially to absorption of radiation from below) and a cooling at the top (due essentially to emission of radiation). However, the magnitudes and thus the net radiative gain or loss by the cloud as a whole depends on the vertical location of the cloud, the cloud top and cloud base. Layers containing clouds in the lower troposphere experience strong radiative cooling. The thicker the cloud is, the more it acts like a black body and the stronger the cooling. In layers with

cirrus clouds, however, in the upper troposphere (around 12–17 km), there tends to be a net radiative warming. In the lower stratosphere (around 20 km), ozone absorption causes warming; in the rest of the stratosphere (above about 25 km), there is cooling owing to water vapor, carbon dioxide, and ozone. On an annual basis, the global mean stratospheric longwave cooling very nearly balances the solar heating owing to absorption of solar radiation by ozone, resulting in a state of radiative equilibrium for that region. This is in contrast to the troposphere, where longwave cooling and solar heating do not balance each other; the residual energy is manifest in non-radiative processes such as dynamical motions and latent and sensible heat fluxes. Above about 60 km (the mesosphere), the longwave cooling is due to carbon dioxide.

The Earth s surface is also affected by atmospheric longwave radiation. Most of the radiation received at the surface is due to emission by tropospheric water vapor and carbon dioxide. In the 8–12 μm (1250–833 cm^{-1}) region, there is a strong contribution resulting from clouds, particularly from low lying clouds that radiate virtually as black bodies.

PHYSICAL EFFECTS

Perturbations in amounts of longwave radiation occur in response to changes in the concentrations of the various absorbers, gases, aerosols, and clouds. As a result of human activities, the concentrations of certain greenhouse gases (carbon dioxide, methane, nitrous oxide, and halogenated species including CFCs) have been increasing over the past century or so. These have increased the trapping of longwave energy in the atmosphere, giving rise to concerns about possible global climatic warming. If increases in water vapor accompany the increases in trace gases, this would exacerbate surface warming. Changes in the concentrations of the longwave absorbers, including aerosols and clouds, also perturb the radiative cooling of the atmosphere. The destruction of ozone in the stratosphere by the CFCs (the ozone hole phenomenon), increases in greenhouse gases, and growing concentrations of stratospheric aerosols following volcanic eruptions perturb the longwave radiative cooling of the stratosphere, causing changes in temperatures there as well.

Over the past approximately four decades, different kinds of instruments that measure longwave radiation have been devised to make inferences about the climate system. These fall into two broad categories: some measure radiant energy at different wavelengths; others measure the total longwave irradiance. Over the past two decades, satellites have permitted a reliable global monitoring of the longwave irradiance emitted by the Earth. Observations confirm that, in clear skies, water vapor is the dominant infrared absorber (i.e., major component in the present day greenhouse effect). Observations also indicate that longwave trapping is most effective in the moist tropics and least effective in the drier subtropical regions. Clouds are even more effective in trapping longwave radiation, especially the deep convective clouds in tropical regions whose tops are located at altitudes in the upper troposphere. The present global effect of clouds in reducing the longwave irradiance at the top of the atmosphere is equivalent to about seven times more than the effect that would result from an instantaneous doubling of carbon dioxide, and about 14 times more than the effect of increases in greenhouse gases over the past century. Considering the secular increases in greenhouse gases, the accompanying feedbacks due to increases in water vapor and changes in clouds (amounts, locations and properties) constitute critical factors in quantifying global climate change.

The effect of clouds on longwave radiation is the opposite of their effect on solar radiation. As a consequence of the absorption and the emission processes, the longwave radiation spectrum at the different altitudes and at the top of the atmosphere is a function of the vertical distribution of temperature and the absorptivity of the various species. For this reason, remote sensing by satellites at different wavelengths in the longwave (including the microwave) spectrum has been found to be useful in extracting information about the atmosphere. This principle has been exploited to infer temperature and species concentrations, especially above the tropopause. It is also being used to deduce physical properties of clouds.

See also: Energy Balance and Climate, Volume 1.

FURTHER READING

Peixoto, J P and Oort, A H (1992) Radiation Balance, in *Physics of Climate*, American Institute of Physics, New York, 91–130.

Ramanathan, V, Cess, R D, Harrison, E F, Minnis, P, Barkstrom, B R, Ahmad, E, and Hartmann, D (1989) Cloud Radiative Forcing and Climate: Insights from the Earth Radiation Budget Experiment, *Science*, 243, 57–63.

Ramaswamy, V (1996) Longwave Radiation, in *Encyclopedia of Climate and Weather*, ed S Schneider, Oxford University Press, Oxford, Vol. II.

Ramaswamy, V, Schwarzkopf, M D, and Trueman, D (1990) Line-by-line Characterization of the Radiative Effects and the Greenhouse Warming Potential due to Various Halogenated Compounds, *Proceedings of Seventh Conference on Atmospheric Radiation*, American Meteorological Society, 438–441.

Shine, K, Fouquart, Y, Ramaswamy, V, Solomon, S, and Srinivasan, J (1995) Radiative Forcing, in *Climate Change (1994): Radiative Forcing of Climate Change and an Evaluation of the IPCC IS92 Emission Scenarios*, Intergovernmental Panel on Climate Change, eds J T Houghton, L G Meira Filho,

B A Callander, N Harris, A Kattenberg, and K Maskell, Cambridge University Press, 163–203.

Interdecadal Oscillation
see Quasi–Decadal Oscillation (**Volume 1**)

Intergovernmental Panel on Climate Change (IPCC)
see IPCC (Intergovernmental Panel on Climate Change) (**Volume 1**); Intergovernmental Panel on Climate Change (IPCC): an Historical Review (**Volume 4**)

International Association for the Physical Sciences of the Oceans (IAPSO)
see IAPSO (International Association for the Physical Sciences of the Oceans) (**Volume 1**)

International Association of Hydrological Sciences (IAHS)
see IAHS (International Association of Hydrological Sciences) (**Volume 1**)

International Association of Meteorology and Atmospheric Sciences (IAMAS)
see IAMAS (International Association of Meteorology and Atmospheric Sciences) (**Volume 1**)

International Geophysical Year (IGY)
see IGY (International Geophysical Year) (**Volume 1**)

International Geosphere–Biosphere Programme (IGBP)
see IGBP (International Geosphere–Biosphere Programme) (**Volume 2**)

International Global Atmospheric Chemistry (IGAC)
see IGAC (International Global Atmospheric Chemistry) (**Volume 1**)

International Hydrological Program (IHP)
see IHP (International Hydrological Program) (**Volume 1**)

International Oceanographic Commission
see IOC (Intergovernmental Oceanographic Commission) (**Volume 4**)

International Organizations in the Earth Sciences
see International Organizations in the Earth Sciences (*Opening essay*, **Volume 1**)

International Panel on Climate Change
see IPCC (Intergovernmental Panel on Climate Change) (**Volume 1**); Intergovernmental Panel on Climate Change (IPCC): an Historical Review (**Volume 4**)

International Research Institute for Climate Prediction (IRI)
see IRI (International Research Institute for Climate Prediction) (**Volume 1**)

International Satellite Cloud Climatology Project (ISCCP)

see ISCCP (International Satellite Cloud Climatology Project) (**Volume 1**)

International Satellite Land Surface Climatology Project (ISLSCP)

see ISLSCP (International Satellite Land Surface Climatology Project) (**Volume 1**)

International Union of Geodesy and Geophysics (IUGG)

see IUGG (International Union of Geodesy and Geophysics) (**Volume 1**)

Intertropical Convergence Zone (ITCZ)

Rainfall over many parts of the tropical oceans occurs primarily in a long, zonally oriented band that is more or less parallel to and slightly off the equator. This band is referred to as the *Intertropical Convergence Zone* (ITCZ). As well as high rainfall, the ITCZ is a band of relatively low pressure and high convective cloud amount; it marks the meeting of the Northern and Southern Hemisphere trade winds (the Doldrums). The ITCZ occurs over areas of high sea surface temperature in the Atlantic and Indian Oceans. The ITCZ tends to follow the annual latitudinal cycling of the Sun, migrating seasonally towards the warmer hemisphere. However, in the central and eastern Pacific the ITCZ tends to remain in the Northern Hemisphere. Air that converges in the lower levels of the atmosphere due to the meeting of the trade winds rises through the clouds in the ITCZ and emerges in the upper troposphere. After it emerges, the air participates in the Hadley Circulation (*see* Hadley Circulation, Volume 1), which carries it into the subtropics, where the air descends, eventually becoming entrained again into the source region for the trade winds.

The ITCZ is one of three characteristic near-equatorial large-scale precipitation features that is evident in the seasonal or annual mean motions of the atmosphere. The zonal character of the ITCZ is broken over Africa and South America, as well as over the Indonesian maritime continent. Over these continental areas, the precipitation is spread over a larger band of latitudes than in the ITCZ. In contrast to the ITCZ, the continental-type precipitation features, which are associated with the rising branches of the Walker Circulation, straddle the equator, at least in the annual mean. During an El Niño, the Indonesian continental-type precipitation region moves eastwards and replaces the ITCZ in the central Pacific (*see* El Niño and La Niña: Causes and Global Consequences, Volume 1). Also, there is another persistent band of precipitation in the Southern Hemisphere tropical Pacific known as the South Pacific Convergence Zone that slants away from the equator from west to east.

See also: Atmospheric Motions, Volume 1.

EDWIN K SCHNEIDER *USA*

Ionosphere

Ionosphere is the upper region of the atmosphere that contains a significant concentration of ions and free electrons. The ions are a result of the absorption of solar radiation and cosmic rays by atoms and molecules and by ionization due to collision among the particles. The ionosphere is usually identified as the region above 60 km above the Earth s surface, and is often divided into three layers. The *D region*, which exists from about 60 km to about 120 km in altitude, has a significant diurnal variation with high concentrations of ions during daylight hours, when solar radiation creates ions at a rate higher than natural ion removal, and much lower concentrations at night, when electrons naturally recombine with positive ions. The higher regions, the *E region* that occupies the layer from about 120 km to about 180 km, and the *F region* identified as the region above about 180 km, do not show a strong diurnal variation as both the rates of creation and removal of ions are much slower due to the lower densities there. The ionosphere is important for radio communication because longer wavelength (amplitude modulation (AM) radio frequencies) radio waves are absorbed and reflected by the D and E regions of the ionosphere, establishing limits to the effective transmissions of these wavelengths (though with a diurnal variation to these limits as the weakened D region does not impede transmission at night). Shortwave radio wavelengths can penetrate the lower D and E regions but are reflected by the F region, allowing transmission to longer distances through reflection off the higher layer. Much shorter wavelengths

(frequency modulation (FM) radio and television, radar, and visible light) are not reflected by the ionospheric layers.

KEITH L SEITTER *USA*

IPCC (International Panel on Climate Change)

The IPCC was established jointly by the World Meteorological Organization and the United Nations Environment Programme in 1988. The IPCC operates independently, but in support of the United Nations Framework Convention on Climate Change (UNFCCC). The IPCC s mandate is: (i) to assess available information on the science, the impacts and the socio-economic aspects and consequences of climate change and on the policy options to address climate change; and (ii) to provide, on request, scientific/technical advice to the Conference of the Parties to the UNFCCC and its bodies. While the IPCC assessments strive to be policy relevant and analyze policy implications, they do not recommend international or national policies. The First Assessment Report completed in 1990 in three volumes, including suggested possible elements for a framework convention, played a key role in the negotiations leading to the establishment of the UNFCCC. The Second Assessment Report, completed in 1995 in four volumes, played a similar role in support of the negotiation of the Kyoto Protocol. The IPCC s Third Assessment Report is to be published in four volumes during 2001. In addition, a number of special reports and papers have been prepared dealing with such issues as regional climate change, climate change and aircraft, technology transfer, and carbon sequestration.

Authors for the chapters of an IPCC report are selected from among nominations by governments of scientists and other experts. Author teams normally include at least one expert from the developing world. The authors draft chapters that undergo: (i) review by experts in the subject throughout the world, and (ii) technical review by governments. The Summaries for Policymakers of the reports are approved line-by-line; the underpinning scientific reports are accepted or adopted in whole at the plenaries. Differing but technically well-founded views are stated as such.

The IPCC also develops methodologies such as Guidelines for National Greenhouse Gas Inventories, Technical Guidelines for Assessing Climate Change Impacts and Adaptations.

See also: Intergovernmental Panel on Climate Change (IPCC): an Historical Review, Volume 4.

N SUNDARARAMAN *Switzerland*

IRI (International Research Institute for Climate Prediction)

Scientists have long recognized a connection between the ocean and the atmosphere. This interaction influences rainfall and temperature patterns around the world. Droughts and floods associated with climate variability frequently cause economic and social damage in many regions. Advances in Earth science research have provided breakthroughs in forecasting related impacts, such as those associated with El Niño.

The IRI was first proposed at the United Nations Conference on Environmental and Development (UNCED) in 1992. IRI was envisaged as a multinational, end-to-end, climate research and applications system for producing and distributing experimental seasonal-to-interannual climate forecasts, complementing existing global ocean and atmospheric observing systems and process studies.

IRI was formally launched during the International Forum on Forecasting El Niño in 1995, and later established in 1996 through an agreement between the US National Oceanic and Atmospheric Administration (NOAA) and Columbia University s Lamont-Doherty Earth Observatory located in Palisades, NY. The Taiwan Central Weather Bureau joined as a major supporter of IRI in 2000. IRI is currently organized to integrate five major functions: modeling and prediction, experimental forecasting, applications, climate monitoring and dissemination, and training.

IRI links the physical, environmental, social and economic sciences with the goals of understanding impacts and vulnerabilities, and developing key applications and capacity building projects to demonstrate the value of climate forecasts. IRI acts to help decision-makers to manage risks in climate sensitive sectors such as agriculture, water management, fisheries, disaster relief, human health and energy. Working with international and regional partners, IRI brings together scientists with potential users of climate information helping countries in Latin America, the Caribbean, southern Africa, Southeast Asia, South Asia and North America to prepare for or to mitigate the severe

weather-related impacts of climate fluctuations. In this way, IRI catalyzes regional responses to global problems.

Further information on IRI may be obtained by contacting: The International Research Institute for Climate Prediction, 61 Rt. 9W, PO Box 1000, Palisades, NY 10964, USA. Tel. 845-680-4468; Fax: 845-680-4866; http://iri.ldeo.columbia.edu/

See also: El Niño and La Niña: Causes and Global Consequences, Volume 1.

J MICHAEL HALL *USA*

ISCCP (International Satellite Cloud Climatology Project)

ISCCP was organized in 1982 as the first project of the World Climate Research Programme to obtain the global, multi-year satellite observations of cloud properties, needed to determine how clouds alter the radiation balance of Earth, and what role they play in the global hydrological cycle (*see* Cloud- Radiation Interactions, Volume 1). Cloud–radiation and cloud–precipitation interactions produce a complex set of feedbacks on climate change that are not yet well understood, limiting the certainty that can be placed on projections of future climate. Since July 1983, institutions in five countries have collected infrared and visible radiances measured by the international constellation of weather satellites. The radiances are sampled in space and time, calibrated, geo-located and placed into a common format; the first ever global radiance dataset with 30 km, 3 h time resolution was released in 1984. The radiances are then analyzed to characterize cloud property variations over diurnal to decadal time scales and mesoscale (at 50–100 km) to global scales; the first global cloud data products were released in 1988. Currently, these global datasets cover a time period from 1983 to 1999; processing is expected to continue until at least 2005. Together with cloud datasets from surface observations and weather balloon measurements and additional special satellite measurements, the ISCCP cloud datasets describe the variations of cloud cover, top and base heights, optical and layer thicknesses, particle sizes and column water amounts. All of this information is now being used to quantify the effects of clouds on Earth s planetary and surface radiation balances, to determine how atmospheric motions produce clouds, and to determine the relationships between cloud properties and precipitation. More details about ISCCP can found in a series of articles in the *Bulletin of the American Meteorological Society* (Schiffer and Rossow, 1983; Schiffer and Rossow, 1985; Rossow and Schiffer, 1991; Rossow and Schiffer, 1999). Project status, as well as complete documentation of all datasets, can be found on the ISCCP Web site at http://isccp.giss.nasa.gov.

REFERENCES

Rossow, W B and Schiffer, R A (1991) ISCCP Cloud Data Products, *Bull. Am. Meteorol. Soc.*, 72, 2–20.

Rossow, W B and Schiffer, R A (1999) Advances in Understanding Clouds From ISCCP, *Bull. Am. Meteorol. Soc.*, 80, 2261–2287.

Schiffer, R A and Rossow, W B (1983) The International Satellite Cloud Climatology Project (ISCCP): the First Project of the World Climate Research Programme, *Bull. Am. Meteorol. Soc.*, 64, 779–784.

Schiffer, R A and Rossow, W B (1985) ISCCP Global Radiance Data Set: a New Resource for Climate Research, *Bull. Am. Meteorol. Soc.*, 66, 1498–1505.

WILLIAM B ROSSOW *USA*

ISLSCP (International Satellite Land Surface Climatology Project)

The ISLSCP operates under the auspices of the World Climate Research Programme (WCRP). Established in 1983, originally under the United Nation s Environment Programme, the goal of ISLSCP is to promote the use of satellite data for improving land-surface data used in climate studies (*see* Land Surface, Volume 1). The project has been instrumental in developing global land-surface data employed in climate models and in the design and implementation of important field experiments.

The First ISLSCP Field Experiment (FIFE) was conducted between 1987 and 1989 in the Konza Prairie in Kansas, USA. The aim of FIFE was to develop satellite-derived data on land-surface conditions relevant to biomass, vegetation cover type and temperature, and vegetation and soil processes such as transpiration and photosynthesis. Between 1994 and 1996 ISLSCP continued and extended its field activities to the Boreal ecosystems–atmosphere study (BOREAS) in Canada (*see* Boreal Ecosystem- Atmosphere Study (BOREAS), Volume 2).

The stated objectives of the ISLSCP are:

- to demonstrate the types of surface and near-surface satellite measurements that are relevant to climate and global change studies;
- to develop and improve algorithms for the interpretation of satellite measurements of land-surface features;
- to develop methods to validate area-averaged quantities derived from satellite measurements for climate simulation models;
- to prepare the groundwork for future operational production of land-surface data sets, which can be directly applied to climate problems.

ISLSCP was integrated into the WCRP as part of the Global Energy and Water Cycle Experiment in 1992 (*see* GEWEX (Global Energy and Water Cycle Experiment), Volume 1). Since then the project s emphasis on global satellite-derived information pertaining to soil and vegetation has become an important input into numerical climate models. An on-going goal of ISLSCP is to identify influences that are responsible for changes in the interactions between the atmosphere and land-surfaces in terms of both physical and biological quantities and relate these to the exchange of energy and water between the land surface and atmosphere, weather and climate.

ANN HENDERSON-SELLERS *Australia*

Isostasy

The word *isostasy* (Gr. *isostasios*, in equipoise) is widely employed in the subjects of geophysics and geology to denote the state of gravitational equilibrium into which the Earth would evolve (the state of isostatic equilibrium) should the gravitational forces acting on the system become unbalanced for any reason. The concept was originally introduced to understand the subsurface distributions of mass that would have to exist beneath continental mountain ranges in order that the anomalously light material of which they are constructed be in gravitational equilibrium with their surroundings (see e.g., Holmes, 1965). However, it is also employed to understand the nature of the deformation of the solid Earth that is produced where large ice sheets develop on the continents as a consequence of climate change. In this situation the surface of the solid Earth comes to be viscoelastically depressed by the weight of a load of large spatial scale to a degree such that the Archimedes force of buoyancy due to the surface displacement balances the weight of the load. One then speaks of the system as being in a state of glacial isostatic equilibrium.

REFERENCE

Holmes, A (1965) *Principles of Physical Geology*, Nelson, London.

W RICHARD PELTIER *Canada*

ITCZ (Intertropical Convergence Zone)

see Intertropical Convergence Zone (ITCZ) (**Volume 1**)

IUGG (International Union of Geodesy and Geophysics)

IUGG is an international scientific non-governmental organization, established in 1919.

This organization is one of the 25 scientific unions presently grouped within the International Council for Science (ICSU – formerly International Council of Scientific Unions).

IUGG is dedicated to the scientific study of the Earth, including applications to societal needs, such as mineral resources, mitigation of natural hazards and environmental preservation. Its objectives are the promotion and coordination of physical, chemical and mathematical studies of the Earth and its environment in space. They include the shape of the Earth, its gravitational and magnetic fields, the dynamics of the Earth as a whole and of its component parts, the Earth s internal structure, composition tectonics, the generation of magmas, volcanism and rock formation, the hydrological cycle including snow and ice, all aspects of the oceans, the atmosphere, ionosphere, magnetosphere and solar terrestrial relations, and analogous problems associated with the moon and other planets.

The IUGG is comprised of seven associations, each responsible for a specific range of topics or themes within

the overall scope of the union s activities and each with a sub-structure. These seven international associations are:

International Association of Geodesy (IAG);
International Association of Seismology and Physics of the Earth s Interior (IASPEI);
International Association of Volcanology and Chemistry of the Earth s Interior (IAVCEI);
International Association of Geomagnetism and Aeronomy (IAGA);
International Association of Meteorology and Atmospheric Sciences (IAMAS);
International Association of Hydrological Sciences (IAHS);
International Association for the Physical Sciences of the Ocean (ASPSO);

Besides, inter-association committees and commissions have been established, e.g.,

The Committee on Mathematical Geophysics (CMG);
The Committee on the Study of the Earth s Deep Interior (SEDI);

The union also co-sponsors the Federation of Astronomical and Geophysical data analysis Services (FAGS) and is a partner with other Unions of ICSU in inter-union bodies such as the International Lithosphere Program.

Members to the union are countries; there are 75 such members as of January 1st, 2000. They participate in the union through IUGG committees set up by the national academy or another body that adheres to the union.

IUGG holds General Assemblies at 4-year intervals and each of its associations also organizes a scientific assembly in its own area of research between general assemblies of the union.

GEORGES BALMINO *France*

J

Jet Stream

Reid Bryson
University of Wisconsin, Madison, WI, USA

The jet stream or band of maximum winds in the upper troposphere was first recognized as an important global phenomenon in 1944. Normally it consists of two segments in the Northern Hemisphere, one from western North America to Northern Europe, and the other from western North Africa across Asia and the Pacific. Such segments tend to spiral toward the pole in both hemispheres, and change latitude with the seasons. The main jet streams are in the westerlies, but there are also some easterly jet streams, mostly in the tropics.

DEFINITION

The standard glossary definition of the jet stream is:

> Relatively narrow river of very strong horizontal winds (usually 50 knots or greater) embedded in the planetary winds aloft. These jets are typically located in the upper troposphere above regions of strong horizontal temperature contrasts (fronts) ••• . Several major jet streams include the polar jet and the subtropical jet.
>
> (Geer *et al.*, 1996)

The definition omits the fact that these elongated west–east bands are nearly always composed of high-speed westerlies, but does include the relationship between the horizontal temperature contrast and the winds aloft over that zone of contrast. That is more obvious in climatic data than in daily weather.

ORIGIN AND DESCRIPTION

The basic equations of motion require that, once above the friction layer, the change of wind velocity with height must be proportional to the horizontal temperature gradient (contrast). Thus, the higher one goes in this region of contrast, the stronger will be the winds. Indeed, the upper westerlies of both Northern and Southern Hemispheres exist because the poles are colder than the equator. Above about 10 km elevation the south–north temperature gradient tends to reverse, so that the winds decrease upward above that level. Consequently, the level of the jet stream core of maximum winds is generally at 10–13 km above sea level.

Thus, one expects strong winds above the regions of the earth where there is a strong temperature gradient. In the European sector, this means a jet stream in the Mediterranean region where Europe to the north is much colder than Africa to the south, and in the Scandinavian–Baltic area where the air flowing into Europe from the Atlantic is warmer than the cold Arctic air flowing from the ice covered Arctic Ocean. The former position has far more temperature contrast than the latter. Similarly, in the North American sector, the land–Gulf of Mexico and Arctic Ocean–land contrasts give two regions of high probability of strong westerlies aloft. Again, the lower latitude jet stream is the stronger, climatically. The locations of these core areas change with the seasons.

Over the Atlantic and Pacific Oceans there is essentially only one such region of strong temperature gradient each. These lie roughly over the Gulf Stream and Kuroshio oceanic warm currents, respectively. Consequently there is not a continuous or multiple or single jet stream running around the Northern Hemisphere (Figure 1). The numbers along the heavy line give the mean wind speed in meters per second at maxima and significant points.

In January, and the winter season in general, a westerly jet is found over the southern United States and off Southeast Asia, with maxima near the southeast coasts. In crossing the oceans, both shift northward, the Asian jet less than the North American jet, and in the mean they weaken near the eastern edge of the oceans. The main Eurafrican jet starts near the Atlantic over North Africa, weakens in the vicinity of the Himalayas and India, and then strengthens as it becomes the southeastern Asian jet. Thus, the jet stream is not one continuous band of high winds about the Northern Hemisphere.

In July, the jet axis comes closer to being a single river of air but the winds are weaker and more indefinite in position near the west coasts of Eurasia and North America.

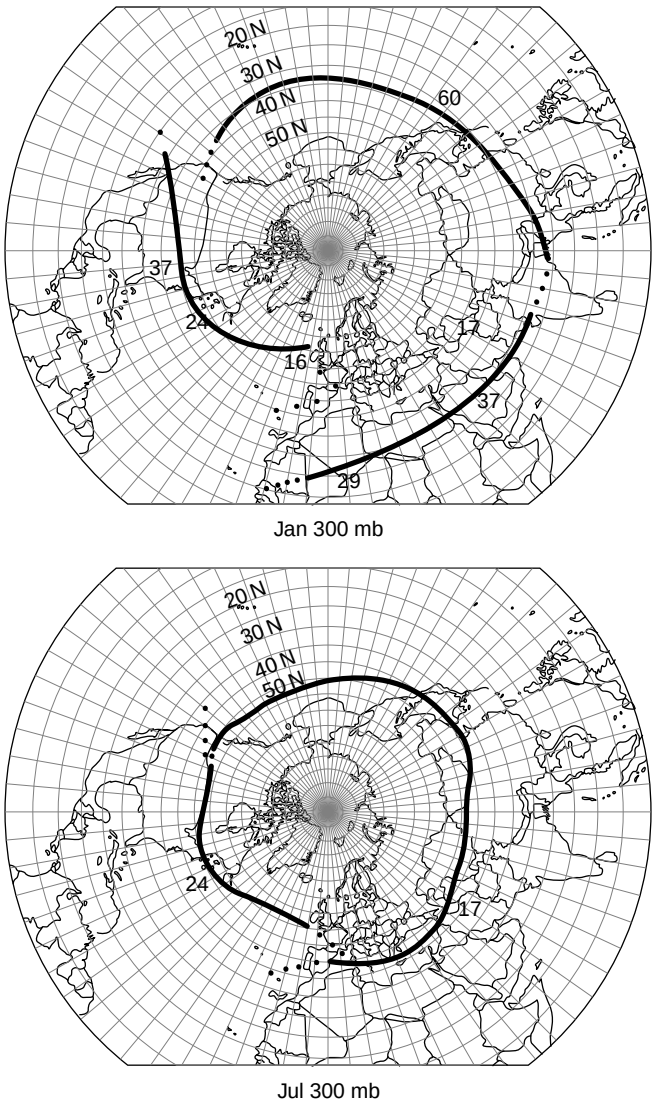


Figure 1 The heavy lines show the position of the axis of the jet stream at 300 mb. The values are mean wind speed in meters per second. By far the strongest westerlies along the jet axis in January are off Japan

The Southern Hemisphere has corresponding jet streams, but in that largely water hemisphere the pattern is simpler. There are also segments of weaker, easterly jets at times and in places where the equatorial region is cooler than the tropical region somewhat to the poleward, e.g., subsaharan Africa.

HISTORY

The first observations of large-scale features of the atmosphere with winds in excess of what might occasionally be observed at the ground came with the development of high-flying aircraft. In the mid-1930s there were some encounters with winds on the order of 80 m s^{-1} or less. With the advent of the Second World War, aircraft in Europe occasionally met even stronger winds, but there was little notice taken in the meteorological community at large, possibly due to restrictions in access to information.

During the same time period the Japanese became aware of high winds aloft. With the advent of high altitude bombing missions over Japanese targets in the fall of 1944, a different wind problem arose than in Europe, and produced widespread interest and discussion.

As shown in Figure 1, there are two main locations for branches of the jet stream in Europe: one in the Mediterranean–North African latitudes and one in the Scandinavian area, both with monthly average wind speeds ranging up to 40 m s^{-1} in the south and less than half that in the north. Most aircraft operations were between the two branches. On the other hand the monthly mean speeds in eastern North America and in the Pacific near Japan are 50–100% greater, and the aircraft operations in the Pacific crossed the jet stream position nearly every day in the cooler months. Winds of nearly 120 m s^{-1} were observed. This was sufficiently close to the speed of the aircraft to pose serious military problems. One might say that this was when the jet stream as a large-scale, significant feature of the atmospheric circulation was discovered (Bryson, 1994).

Immediately after World War II a great deal of research effort was devoted to the study of this newly recognized large-scale feature of the atmospheric circulation (Reiter, 1967).

SIGNIFICANCE

The jet stream, lying above the greatest temperature gradients, i.e., the fronts, is therefore the approximate path of frontal cyclones. However, the fast-moving winds of the jet stream must follow the laws of air dynamics and thus do not follow the fronts in detail. The daily adjustment of the one to the other is a large part of mid-latitude meteorology. Climatically speaking, the concordance of mean frontal position and mean jet stream position is much better, as indicated previously. Changes of mid-latitude climate and changes in the mean position of the jet stream are therefore closely related. For example, periods when the hemisphere is warmer are times when the jet stream and associated fronts are generally farther poleward than in cooler times.

See also: Atmospheric Motions, Volume 1.

REFERENCES

Bryson, R A (1994) The Discovery of the Jet stream, *Wisconsin Acad. Rev.*, 40(3), 15–17.

Geer, I W, Ginger, K M, Moran, J M, Hopkins, E J, Weinbeck, R S, and Smith, D R, eds (1996) *Glossary of Weather and Climate, Am. Meteorol. Soc.*, Boston, 1–272.

Reiter, E R (1967) *Jet Streams*, Doubleday, New York, 1–189.

JGOFS (Joint Global Ocean Flux Study)

The role of the ocean in controlling climate change through the storage and transport of heat (and liquid water) was recognized early on by the World Climate Research Programme. The joint Scientific Committee on Oceanic Research (SCOR)/International Oceanographic Commission (IOC) Committee on Climatic Change and the Ocean (CCCO) proposed a global survey of the oceanic carbon dioxide budget field. The World Ocean Circulation Experiment (WOCE) agreed to make berths available on ships taking part in the WOCE Hydrographic Programme for the necessary measurements. The formal organization of JGOFS was carried out by SCOR. JGOFS is now a core project of the International Geosphere–Biosphere Programme, although responsible directly to SCOR.

The oceans contain some 50 times as much carbon dioxide as the atmosphere, and small changes in the ocean carbon cycle can therefore have large atmospheric consequences. Many scientists worldwide are addressing aspects of the ocean carbon cycle, but to determine overall net fluxes, and the processes controlling them, is beyond the capability of any one nation. JGOFS therefore plans and executes research that requires international cooperation. Twenty-one countries contributed to JGOFS planning or field work.

The scientific goals of JGOFS are:

- to determine and understand on a global scale the processes controlling the time-varying fluxes of carbon and associated biogenic elements in the ocean, and to evaluate the related exchanges with the atmosphere, sea floor and continental boundaries;
- to develop a capacity to predict on a global scale the response to anthropogenic perturbations, in particular those related to climate change.

Operationally, the program s activities seek to assess more accurately and improve understanding of the processes controlling regional to global and seasonal to interannual fluxes of carbon between the atmosphere, surface ocean and ocean interior, and their sensitivity to climate changes.

Further information may be obtained from: JGOFS International Project Office, Centre for Studies of Environment and Resources, University of Bergen, Bergen High-Technology Centre, 5020 Bergen, Norway, Tel. (• 47) 555 84246, Fax: (• 47) 555 89687, E-mail: JGOFS@smr.uib.no.

JOHN S PERRY *USA*

Junge Layer

The *Junge layer* is the name given to the background stratospheric sulfate aerosol (SSA) layer that was first directly observed by Professor Christian Junge in 1961 (Junge *et al.*, 1961). This layer is located at an altitude of about 20 km. Junge and his colleagues found that this aerosol layer was composed mostly of sulfuric acid (H_2SO_4) particles. They found it particularly interesting that the layer seemed to be present even long after volcanically injected sulfur dioxide (SO_2) should have oxidized to form H_2SO_4 and been removed from the stratosphere over periods of a couple of years (*see* Volcanic Eruptions, Volume 1).

Observations of the Junge layer during volcanically quiescent periods suggest that the sulfur contributing to the SSA layer may be coming from oxidation of the upward mixing of SO_2 emitted in the lower troposphere, from the upward transport of carbonyl sulfide (COS) that is emitted naturally from the ocean and anthropogenically by industry, or from the upward transport of SO_2 emitted by jet aircraft. However, the recent upward trends in the SSA, which are not readily explained by these factors, suggest that perhaps volcanically injected sulfur may be remaining in the stratosphere longer than had previously been thought (Godin and Poole, 1998).

Professor Junge (1912–1996), who was a leading German scientist, is also well known for many contributions to atmospheric research. Among other insights, he is credited with being the first to describe the continuous size distribution of atmospheric aerosols (*see* Aerosols, Stratosphere, Volume 1; Aerosols, Troposphere, Volume 1). In addition, Junge s book in 1963 on *Atmospheric Chemistry and Radioactivity* established the subject as an important new research field. Up until that time, there were few atmospheric chemists; and national meteorological services generally considered the field to be outside their terms of reference.

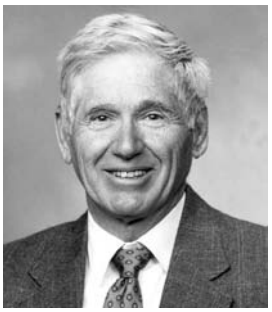
REFERENCES

Godin, S and Poole, L R (1998) Global Distributions and Changes in Stratospheric Particles, in *Scienti c Assessment of Ozone Depletion: 1998*, World Meteorological Organization, Report No. 44, Geneva, Switzerland.

Junge, C E, Changnon, C W, and Manson, J E (1961) Stratospheric Aerosols, *J. Meteorol.*, 18, 81–108.

MICHAEL C MACCRACKEN *USA*

K



Keeling, Charles David

(1928-)

The epochal discovery that the concentration of carbon dioxide (CO_2) in the Earth s atmosphere is rising was made by C D Keeling, based on a series of extremely accurate measurements beginning in 1958 and continuing to the present time. Keeling, born in 1928, graduated from the University of Illinois and earned a PhD in chemistry from Northwestern University. He began a postdoctoral fellowship at the California Institute of Technology in 1953, where he developed an interest in measuring atmospheric CO_2 that would prove to dominate his professional life. While a postdoctoral fellow, Keeling designed and built a manometer to measure CO_2 accurately, and found that a concentration of about 310 parts per million by volume (ppmv) was apparently characteristic of the free atmosphere over large regions of the Northern Hemisphere. This surprising result, which seemed clear to Keeling by 1956, would soon upset the prevailing opinion at the time that CO_2 concentrations varied widely over a range of about 150–450 ppmv.

Keeling received important early encouragement from Harry Wexler, in charge of research at the US Weather Bureau in Washington, DC, and from Roger Revelle, director of the Scripps Institution of Oceanography in La Jolla, CA (*see* Revelle, Roger Randall Dougan, Volume 1). Both Wexler and Revelle were instrumental in planning the International Geophysical Year (IGY), an 18-month global program that took place in 1957–1958. At Revelle s invitation, Keeling moved to Scripps in 1956 and, except for brief visits elsewhere, has spent his entire career there. As part of the IGY, Keeling began measurements in Antarctica and at the Mauna Loa Observatory in Hawaii. The Mauna Loa record of CO_2 measurements began in 1958 and quickly yielded important new results. By 1960, in a paper published in *Tellus*, Keeling could report a distinct seasonal cycle of CO_2 concentration in the Northern Hemisphere. Based on carbon isotopic ratio data from his early measurements as a postdoctoral fellow, Keeling proposed in the 1960 *Tellus* paper that the activity of plants growing on land caused the seasonal cycle. CO_2 concentrations were largest in northern spring, when most plants begin to grow, drawing down the atmospheric CO_2 levels. Finally, in the same paper, Keeling noted the possibility of a worldwide increase from year to year.

The Mauna Loa record of CO_2 concentration continues to the present day, and the graph of rising CO_2 concentration as a function of time over more than four decades is now known as the Keeling curve (see Figure 1). This record, now supplemented by measurements at other sites, demonstrates conclusively that CO_2 concentrations are rising and that the primary cause is the burning of fossil fuels: coal, oil and natural gas (*see* Carbon Dioxide, Recent Atmospheric Trends, Volume 1; Carbon Cycle, Volume 2). All discussions of the possibility of global climate change due to human activities begin with this solid empirical evidence, which no reputable scientist doubts. Charles David Keeling thus deserves credit for alerting humankind to the fact that it is changing the chemical composition of the global atmosphere. In addition, the CO_2 concentration observations have yielded a rich harvest of insights into the carbon cycle. It is important to realize that these measurements, which are now sustained by international collaboration, were originally made simply because of the dedication, perseverance and skill of one scientist. In the research community, Keeling s stubborn insistence on highly accurate and completely trustworthy CO_2 concentration measurements is legendary. In fact, Keeling has continually struggled with government agencies for financial support to maintain these measurements according to the high scientific standards that he considers necessary. He has documented these efforts in detail in an autobiographical article (Keeling, 1998). A popular account of Keeling s work in the context of global change science, rich in anecdotal material, has been provided by Weiner (1990). Keeling won the Blue Planet Prize in 1993 for his work.

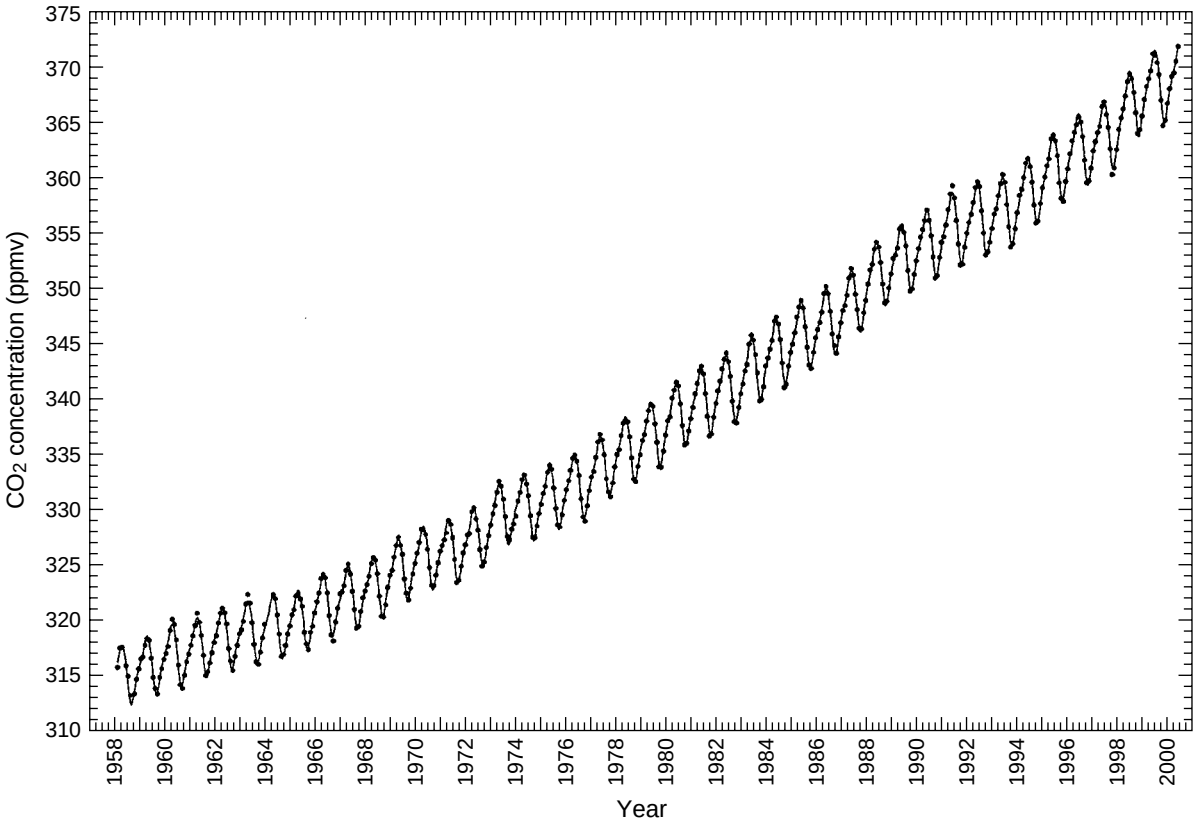


Figure 1 The Keeling curve

REFERENCES

Keeling, C D (1998) Rewards and Penalties of Monitoring the Earth, *Annu. Rev. Energy Environ.*, 23, 25–82.

Weiner, J (1990) *The Next One Hundred Years*, Bantam, New York.

RICHARD C J SOMERVILLE *USA*

Kondratyev, Kirill Yakovlevich

(1920-)

Professor Kirill Kondratyev, a Russian atmospheric scientist of considerable renown, has spent most of his life in Leningrad/St. Petersburg. He entered University there in 1938 but his studies were interrupted by World War II. He joined the army and was wounded during the blockade of Leningrad. After recovering in a hospital in western Siberia, he trained with a parachute division and then returned to the front line west of Moscow. Wounded again in 1944, Kondratyev was demobilized.

When the war ended, Kondratyev resumed his academic studies at the University of Leningrad, graduating in 1946. He immediately became Assistant Professor, then Professor, then Head of the Department of Atmospheric Physics, and finally Rector (1964–1970). At the same time, he was Chief of the Department of Radiation Studies at the Main Geophysical Observatory in Leningrad. Kondratyev became world famous during this period for his many publications in the field of atmospheric radiation. One of his major works was his monograph *Radiative Heat Exchange in the Atmosphere* (Kondratyev, 1965). He won the prestigious World Meteorological Organization/International Meteorological Organization (WMO/IMO) Prize in 1967.

In the 1970s, Kondratyev broadened his scientific interests, becoming a leading figure in remote sensing of the atmosphere from satellites, and today he is still Editor-in-Chief of the journal, *Studying the Earth from Space*. For years, he has collaborated with the leading specialists in

space research (including the astronauts themselves) in his country. One of his important findings during this period was that stratospheric aerosols are significant absorbers in both the short- and the long-wave regions of the spectrum; thus aerosols may cause either cooling (by reflection) or warming (by absorption).

In 1982, Kondratyev joined the Institute for Lake Research, and then in 1992, the Research Centre for Ecological Safety. The changes in affiliation reflect his broadening interest in larger-scale environmental issues:

- climate change, both natural and anthropogenic: in the 1980s Kondratyev became convinced that greenhouse-gas warming was almost inevitable;
- geosphere-biosphere interactions, especially those triggered by massive human disruptions of the Earth System;
- sustainable development: by the early 1990s, Kondratyev was thinking about the urgent need for social scientists to become involved with physical scientists in studies of the global environmental problems.

In addition, Kondratyev has contributed to the solution of the problem of the optimization of global-change observing systems.

Professor Kondratyev has published many research papers and monographs. In 1982, he was elected a full member of the Russian Academy of Sciences, and he has received many international honors and awards. These include the Symons Gold Medal of the Royal Meteorological Society (UK), as well as the IMO Prize. He is *Honoris Causa* of the Universities of Lille, Athens and Budapest. He is co-chairman of the Board of Nansen, a newly created Russian-Norwegian environmental and remote sensing center.

REFERENCE

Kondratyev, K (1965) *Radiative Heat Exchange in the Atmosphere*, Pergamon Press, New York.

FURTHER READING

Taba, H (1998) The Bulletin Interviews Professor K Y Kondratyev, *WMO Bull.*, 47, 2–10.

R E MUNN *Canada*

Krakatau Volcanic Eruption

see Volcanic Eruption, Krakatau (**Volume 1**)

L

La Niña

La Niña is a climatological phenomenon akin to El Niño, but with opposite tendencies in the tropical Pacific Ocean and atmosphere. La Niña is characterized by stronger than normal trade winds and colder than normal tropical Pacific sea surface temperatures. It is also characterized by, unusually high surface atmospheric pressure in the eastern tropical Pacific and low surface pressure in the western tropical Pacific in association with the Southern Oscillation. Like El Niño, La Niña typically lasts 12–18 months. La Niña s effects on global weather variability are roughly (though not exactly) opposite to those of El Niño.

El Niño, La Niña, and the Southern Oscillation are often referred to collectively as ENSO (El Niño/southern oscillation), a cycle that oscillates from year to year between warm, cold, and neutral states in the tropical Pacific. The term La Niña (Spanish for the girl) was coined in the mid-1980s by scientists investigating the oscillations between warm and cold conditions in the tropical Pacific. La Niña has also been referred to as anti-El Niño, the cold phase of ENSO, and El Viejo (the old man).

See also: El Niño and La Niña: Causes and Global Consequences, Volume 1; El Niño/Southern Oscillation (ENSO), Volume 1.

MICHAEL J MCPHADEN *USA*

Lamb, Hubert H

(1913- 1997)

Hubert Lamb was a pioneering climatologist and founding Director of the Climatic Research Unit at the University of East Anglia. For much of his career he found himself fighting the received wisdom that climatology was purely a statistical exercise and an academic backwater. Lamb did more than any other climatologist to convince the skeptics that climate change was relevant to the modern world although, paradoxically, he remained skeptical of the enhanced greenhouse gas effect as a forcing factor.

Professor Lamb was born on September 22, 1913 in Bedford, England to the expectations of a family with a distinguished academic pedigree. His grandfather, Horace Lamb, was an eminent mathematician, known to generations of meteorologists through his textbook on hydrodynamics. His father was a professor of engineering. As a child, Lamb was a frequent visitor to tea with Lewis Fry Richardson. Richardson s Quaker philosophy was to be a strong influence throughout the rest of Lamb s life although they never talked about weather or climate; this was deeply ironic since Richardson is well known as the pioneer of numerical weather forecasting (*see* Richardson, Lewis Fry, Volume 1).

Lamb started a degree in natural sciences at the University of Cambridge, but eventually switched over to geography. This training inspired his openness to interdisciplinary studies and fostered his interest in climate–human interactions. In 1936 he took a post at the UK Meteorological Office, but was posted to the Irish Meteorological Service during World War II. Producing weather forecasts for the new transatlantic passenger flights was a challenge because British observations were not available to the neutral Irish. Lamb gleaned most of his information from arriving aircrews. A perfect safety record was early, and convincing, evidence of the scientific insight and intuition that was to characterize much of his professional life. This was also evidenced by his prediction, when a meteorologist on a whaling ship in the Southern Ocean, of the existence of a major topographic barrier in the interior of Antarctica on the basis of synoptic analysis.

On his return to a land posting, Lamb devised and refined a daily weather classification for the British Isles that may prove to be one of his most lasting memorials to climatology. Starting in 1861, and still continuing today, it represents the longest daily record of airflow records for any part of the world. One of the most striking consequences

of the Lamb weather type catalogue was the way he used it to demonstrate the nature of circulation changes, especially in westerly flow, around the British Isles. This observation provided the stimulus for a significantly growing interest in climate change. Lamb s work on weather types is highly cited, and this approach has been shown to have a wide utility for examining the links between circulation and surface temperature, rainfall, pollution, and even ecology. Because of their links with surface variables, Lamb weather types have also been used to assess global climate model results and to generate scenarios of the future climate.

Another significant contribution that Lamb has made to the study of climate was the reconstruction of monthly mean sea level pressure maps over the North Atlantic and Europe using old meteorological archives extending back to the 1750s. The work for which he is most likely to be remembered by those outside climatology was his reconstruction of the characteristics of the Medieval Warm Period (*see* Medieval Climatic Optimum, Volume 1) and of the Little Ice Age period (*see* Little Ice Age, Volume 1), which he suggested extended over parts of the Northern Hemisphere from about the middle of the 16th century into the 17th century with also some later, but shorter, periods. In these studies, Lamb exploited a myriad of sources with great imagination, and well and truly established the importance of climate in the history of our forebearers.

Lamb also started to make connections between sea surface temperature and the atmospheric circulation. This work deepened his conviction of the importance of climate change on time scales of significance to modern mankind. Lamb was also fascinated by the links between climate and human history, and started to work with geologists, botanists and historians on this issue. He became as well known to the practitioners of other disciplines as he did to climatologists.

In 1970 he produced one of his landmark publications; a chronology of volcanic eruptions and a dust veil index (*see* DVI (Dust Veil Index), Volume 3) following each eruption back to the 1500s. It also contained a treatise of the scientific basis of the volcanic forcing mechanism, and an analysis of the observational evidence linking climate change and volcanic events. This work was a classic example of Lamb s interdisciplinary skill, and this was his major strength: combining knowledge from different fields and setting the agenda for the next generation of researchers. This demeanor was crucial for the eventual success of the Climatic Research Unit, which he founded in 1972.

Lamb died on June 28, 1997. His research had been recognized formally through awards such as the Murchison Award of the Royal Geographical Society, the Symons Gold Medal of the Royal Meteorological Society, and the Gold Medal of the Swedish Geographical Society. His legacy is the scientific method of the interdisciplinary study that climatology has become.

See also: Volcanic Eruptions, Volume 1.

TREVOR D DAVIES *UK*

Land Cover and Climate

Roger Pielke, Sr
Colorado State University, Fort Collins, CO, USA

Land cover includes soil type, vegetation type and fractional coverage, snow, lakes, marshes and ponds. Land cover is a dynamic component of the Earth's climate system. It changes on all time and space scales, and is altered by human activity. It is conventional to separate land cover interactions into short (seconds to days), medium (weeks to seasons and years), and long (decade to century and longer) time scales. With respect to vegetation interactions, short, medium, and long-term interactions within the climate system are referred to as biophysical, biogeochemical, and biogeographical processes, respectively. Natural changes have occurred over all of these time scales. For example, during the Pleistocene geological time, continental-scale glaciers covered large portions of the higher latitude landmasses of the Northern Hemisphere. In this article, the focus will be on land cover change and climate that has occurred, and is expected to continue to occur, on decade to century time scales.

The conversion of a landscape by human activity or natural processes is an obvious environmental change. The biodiversity of an area, for example, will be altered when a landscape is changed (Stohlgren, 1999). Natural landscape change includes lightning-started res, certain insects and disease infestations, vegetation succession during recovery from disturbance, and weather extremes such as drought and oods in watersheds not affected by human activities. Human-caused landscape changes include, among other activities, clear-cutting of forests, agriculture, urbanization, draining of wetlands, human-ignited res, and the introduction of exotic plants and animals.

These changes have obvious local effects, but regional and global in uences have also been reported. Moreover, it is actually arti cial to separate natural from human landscape disturbance if adjacent and/or distant human alteration of the environment in uences what occurs over the natural landscape. For example, air pollution from an urban area that is near a national park or wilderness area may alter vegetation growth due to wet and dry disposition of pollutants.

LOCAL SCALE LANDSCAPE CHANGES

Figure 1 provides a schematic illustration of how the surface energy budget and overlying atmosphere are altered as a result of the conversion of a forested area to agriculture. This figure, based on real observational data, indicates that there are several significant alterations of the surface heat energy budget, including different amounts of absorbed solar insolation, different partitioning between sensible and latent heat (evaporative and transpired) turbulent fluxes, and soil heat flux. The water budget, including runoff and infiltration, is also changed. Landscape alterations also change biodiversity, with new plants (crops) being introduced to replace the natural vegetation. Exotic invasive plants, insects and other non-indigenous animals (e.g., goats, etc.) may be introduced, and even invade adjacent forested areas.

Other obvious anthropogenic landscape changes include urbanization, irrigated crops, afforestation, and domestic grazed areas. Each of these local changes can affect the local and nearby climate and other aspects of the environment. The effect of long-term land use change on the climate usually needs to be investigated with a model. Only rarely are long-term observed weather data available whose collection began when the landscape was natural. Figure 2 illustrates the effect in a model simulation of a landscape change from a natural short grass prairie to the current landscape of mixed agriculture and short grass in the central Great Plains of the US. With the current landscape (top), deep cumulonimbus convection (which was observed) results. When the natural landscape was assumed as the bottom boundary condition, only shallow cumulus clouds developed. The reason for the difference is

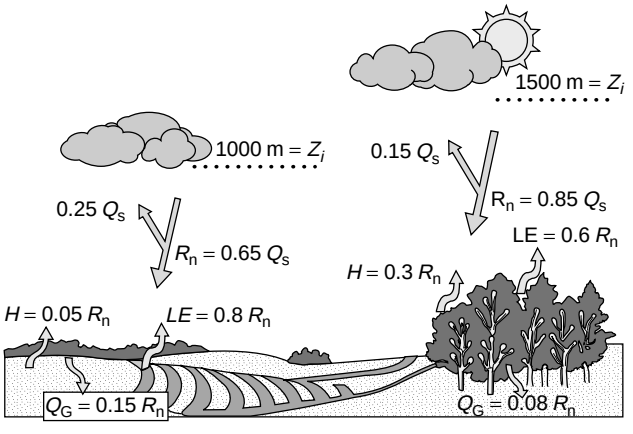


Figure 1 Schematic diagram of the differences in surface heat energy budget and the planetary boundary layer over a forest and over a cropland. Horizontal fluxes are left off the Figure. The symbols refer to: H, sensible turbulent heat flux; Q_s, solar insolation; Z_i, top of the atmospheric boundary layer; LE, latent turbulent heat flux; Q_G, ground heat flux; R_n, net radiative fluxes. (Adapted from P Kabat, personal communication, 1999)

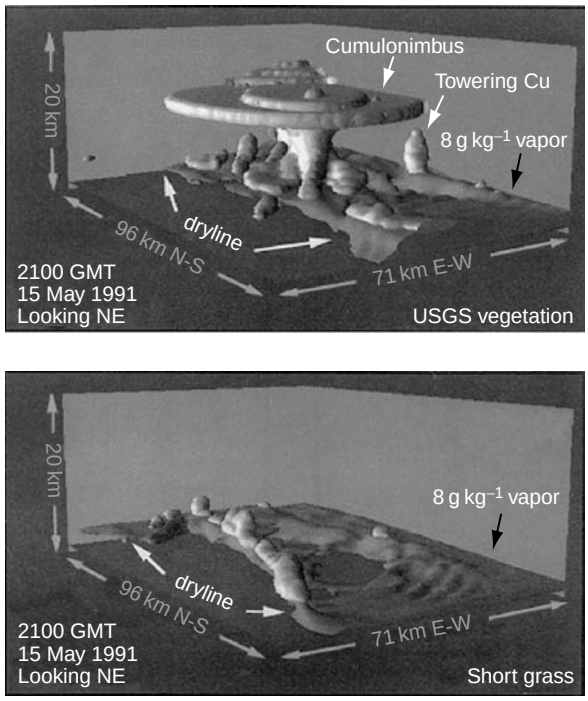


Figure 2 Mesoscale atmospheric model output of cloud and water vapor mixing ratio fields for a region in the Texas and Oklahoma panhandle for 21 GMT on May 15, 1991 with the current landscape (a) and the natural landscape (b). The initial atmospheric structure and lateral boundary conditions are identical for both landscapes. The clouds are depicted by white surfaces with the sun illuminating the clouds from the west. The water vapor mixing ratio of 8 g kg^{-1} in the planetary boundary layer is depicted by the grey surface. Visible ground indicates that the water vapor mixing ratio is less than 8 g kg^{-1} through the entire boundary layer. The vertical axis is height, and the backgrounds are the north and east sides of the domain (United States Geological Service (USGS)). (Reproduced from Pielke *et al.*, 1997)

that surface moisture fluxes were increased by the agricultural landscape because the area of transpiring leaf surface became larger. The reduced water vapor in the atmosphere with the natural landscape resulted in a smaller amount of energy for thunderstorm development.

REGIONAL SCALE LANDSCAPE CHANGES

Landscape changes occur over regional areas. Most of the eastern two-thirds of the US has been changed, for example, as vast areas have been converted to agriculture and deforested. Figure 3 documents estimates of regional changes of landscape for several specific regions of the US. Several regions (such as the Phillipines) have had a rather consistent decrease in forest cover during the 20th century, while New England has become reforested after deforestation prior to the mid-19th century. The effect of weather on these human-caused changes in landscape

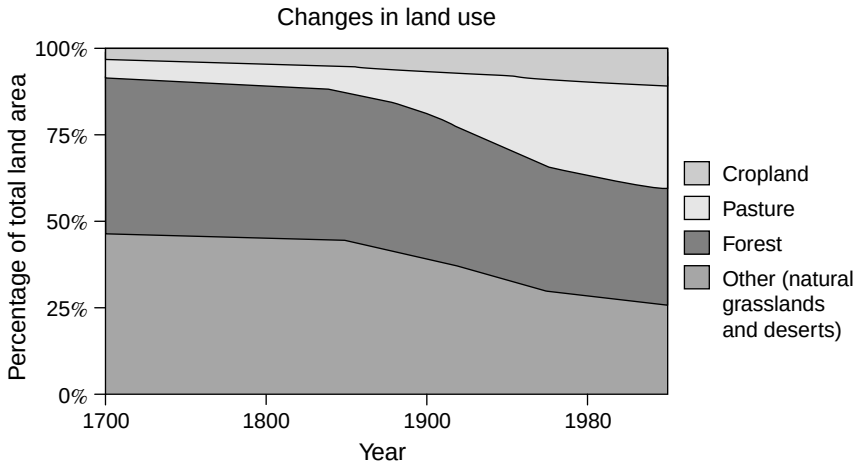


Figure 4 Estimated global land use changes from 1700 to 1995. The land use from top to bottom in the figure are cropland, pasture, forest and others, respectively. (Reproduced from Klein, 2001)

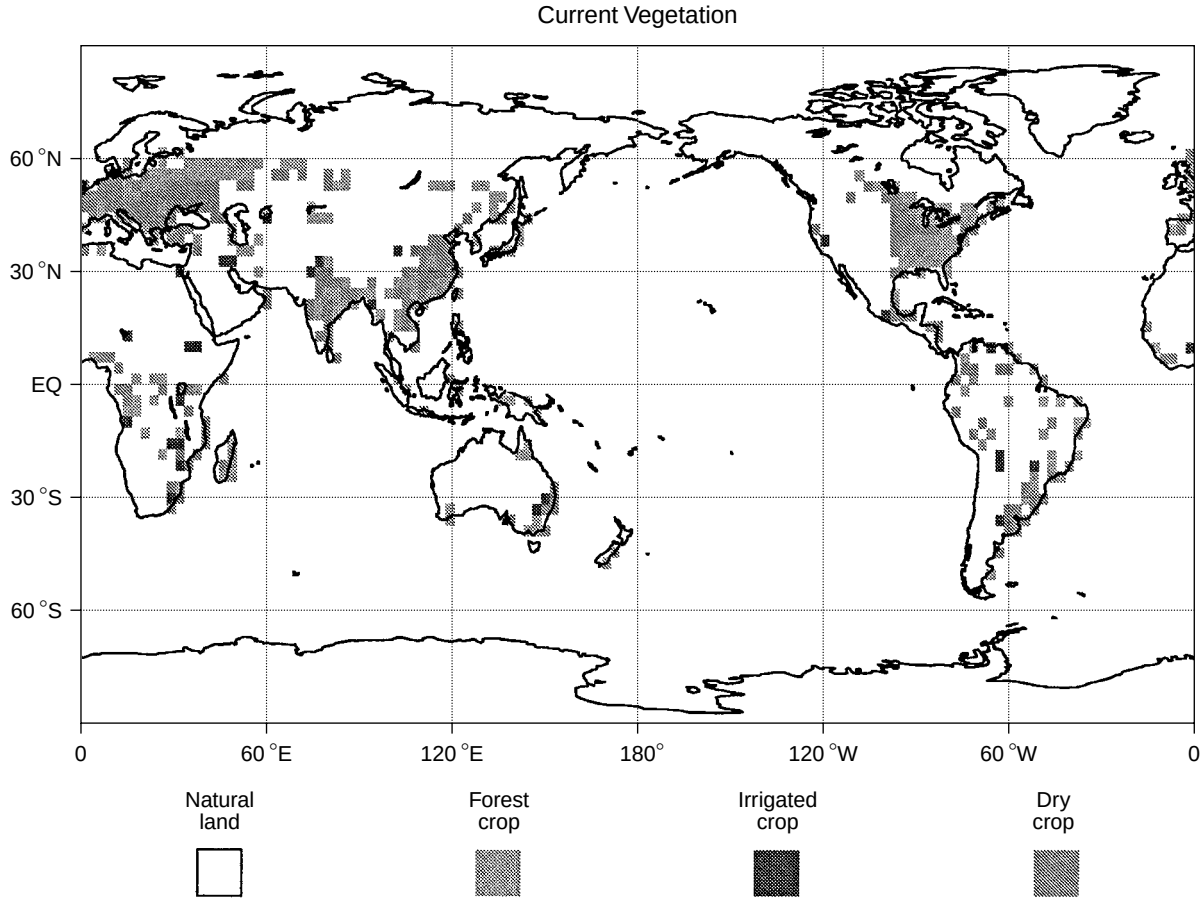


Figure 5 Regions in the NCAR Community Climate Model – 3 GCM, where land was changed from the natural landscape to its human disturbed form. (Reproduced from Chase *et al.*, 2000)

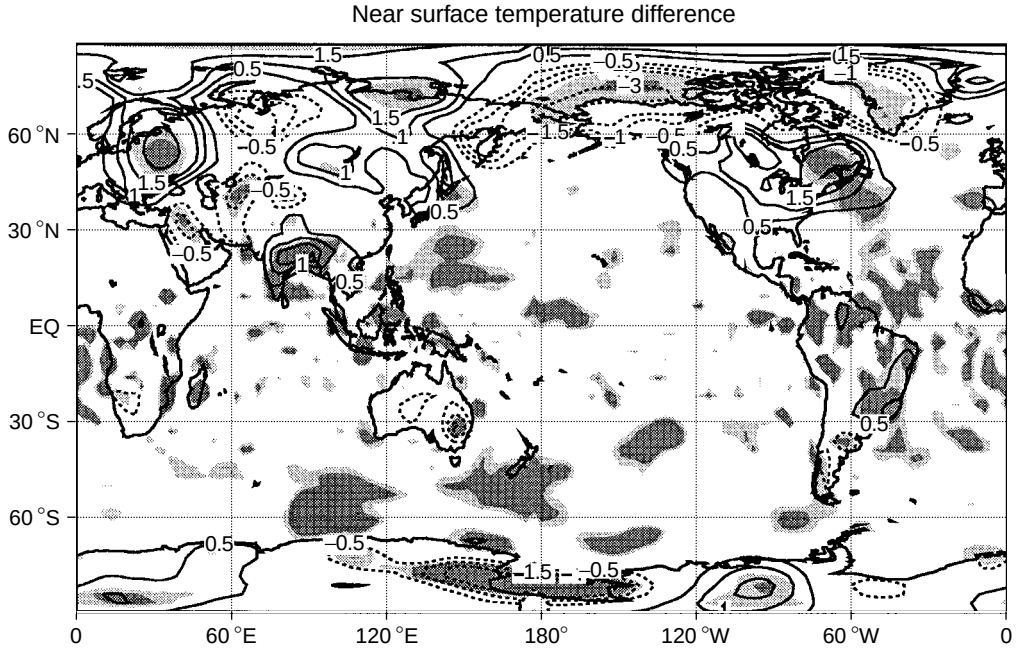


Figure 6 Difference in January global temperature, averaged over 10 years of simulation, where the current landscape instead of the Earth's natural landscape is used in the regions shown in Figure 5. Locations which are statistically different are shown by gray shading. That statistically significant changes appear distant from regions where surface cover changes is a result of changes in the global atmospheric circulation (*see* **Pacific–North American (PNA) Teleconnection**, Volume 1). (Reproduced from Chase *et al.*, 2000)

can be illustrated based on results from a recent study. Assuming identical large-scale weather conditions based on observations, an atmospheric model was integrated for a two-month summer period using observed landscapes of South Florida for 1900, 1973, and 1993 (Pielke *et al.*, 1999). The 1900 landscape was essentially the natural landscape of the region (Costanza, 1975). Compared with the model runs when the 1900 landscape was used, there was a 9% decrease in rainfall averaged over South Florida using the 1973 landscape, and an 11% decrease when the 1993 landscape was used. Corresponding to the decrease in rainfall, the South Florida region warmed an average of about 0.5 °C. The limited available precipitation and temperature records are consistent with this trend.

GLOBAL SCALE LANDSCAPE CHANGES

The cumulative change of landscape by people is estimated to affect directly 40% or more of the Earth's land surface. Figure 4 illustrates the change of different land cover since about 1700, as estimated by the Land Use Change Committee of the International Geosphere–Biosphere Programme (IGBP). A concern, which these data illustrate, is that most land use change occurred in the 1900s, and the change continues to accelerate. Recent general circulation model (GCM) sensitivity simulations suggest that these landscape changes have global effects. Because landscape changes are long-term, cover a large area, and directly affect thunderstorm activity, the influences on weather can extend over large distances. Such a response is similar to El Nino and La Nina effects on weather, except that the land use change persists for a very long time (unlike the tropical Pacific sea surface temperature anomaly, which flip-flops between warm and cold sea surface temperatures).

As an example of the global effect of land use change, Figure 5 indicates regions where land cover is assumed to have been modified by human activities in a GCM study of the effects of changes in the natural landscape. Even over a short time period of 10 years, the two different landscapes produce different large-scale weather patterns. Figure 6 shows how much these landscape changes have altered January temperatures in the model. Regional patterns of warming and cooling occur, as the polar jet stream is changed in its long-term average position at this time of year, as a result of the land use change.

CONCLUSIONS

A major conclusion is that human and natural landscape changes have had effects on the climate on regional and global scales. The human-caused land cover changes in the last 100 years, and even more seriously in recent years,

appear likely to have had an increasingly significant effect on climate and other aspects of the environment, including biodiversity, water quality and quantity.

REFERENCES

Chase, T N, Pielke, Sr, R A, Kittel, T G F, Nemani, R, and Running, S W (2000) Simulated Impacts of Historical Land Cover Changes on Global Climate in Northern Winter, *Clim. Dyn.*, 16, 93–105.

Costanza, R (1975) *The Spatial Distribution of Land Use Subsystems, Incoming Energy and Energy Use in South Florida from 1900- 1973*, MS Thesis, Department of Architecture, University of Florida, FL.

Dobson, A P, Bradshaw, A D, and Baker, A J M (1997) Hopes for the Future: Restoration Ecology and Conservation Biology, *Science*, 277, July 25, 515–522.

Klein, G K (2001) Estimating Global Land Use Change over the Past 300 Years: The Hyde 2.0 Database, Global Biogeochem. Cycles, in press.

Pielke, R A, Lee, T J, Copeland, J H, Eastman, J L, Ziegler, C L, and Finely, C A (1997) Use of USGS-provided Data to Improve Weather and Climate Simulations, *Ecol. Appl.*, 7, 3–21.

Pielke, Sr, R A, Walko, R L, Steyaert, L T, Vidale, P L, Liston, G E, Lyons, W A, and Chase, T N (1999) The Influence of Anthropogenic Land Use Changes on Weather in South Florida, *Mon. Weather Rev.*, 127, 1663–1673.

Stohlgren, T J (1999) The Rocky Mountains, in *Status and Trends of the Nation's Biological Resources*, eds M J Mac, P A Opler, C E Puckett Haecker, and P D Doran, Biological Resources Division, US Geological Survey, Reston, VA, 473–504, Vol. 2.

FURTHER READING

Claussen, M (2001) Biogeophysical Feedbacks and the Dynamics of Climate, in *Global Biogeochemical Cycles in the Climate System*, eds E D Schulze, S P Harrison, M Heimann, E A Holland, J Lloyd, I C Prentice, and D Schimel, Academic Press, San Diego, CA, in press.

O Brien, K L (2000) Upscaling Tropical Deforestation: Implications for Climate Change, *Clim. Change*, 44, 311–329.

Pitman, A, Pielke, Sr, R, Avissar, R, Claussen, M, Gash, J, and Dolman, H (1999) The Role of the Land Surface in Weather and Climate: Does the Land Surface Matter, *IGBP Newslett.*, 39, September, 4–9.

Land–Ocean Interactions in the Coastal Zone (LOICZ)

see LOICZ (Land–Ocean Interactions in the Coastal Zone) (**Volume 2**)

Land Surface

Ann Henderson-Sellers

Lucas Heights Science and Technology Centre, New South Wales, Australia

The land surface is at the front line of concerns about global environmental change. People live on the land surface: it is where we grow our food, build our homes and collect our drinking water. The land surface is vital for our survival. However, exploitation of large areas of the continents for current societal needs is causing massive disturbances that are contributing to and are impacted by global changes.

Human-induced land-use change (the activity) and consequent land-cover change (the result) are inescapable features of global environmental change. Over half the world's continental surface has already been discernibly altered by people: forests replaced by fields, valleys flooded for reservoirs, and roads and cities sprawled across coastal regions and many valleys. People's activities cause, or contribute to not only desertification, deforestation, salinization and soil erosion, but also afforestation, irrigation and landscape development.

INTRODUCTION

Ten percent of the ice-free land area of the world (about 13 500 million ha) is under permanent cultivation, 25% is pasture, 30% is forest and woodland, and the remaining 35% comprises tundra, desert, degraded land, built-up areas and parks. *Per capita* contributions now amount to eroding, lifting and removing about 10 tonnes of minerals every year. The global total of about 50 Pg exceeds by at least a factor of three the sediment load transported annually to the sea by rivers, which also includes a significant input resulting from human activities.

The major types of human-induced land surface change are deforestation (and reforestation), desertification (and irrigation of vegetation), agricultural expansion (and land abandonment) and soil erosion (and fertilization). These changes are widespread and prompt additional impacts such as soil degradation and salinization; increases in aerosol concentrations; changes to greenhouse gas concentrations and nutrient loading of waterways and coastal waters. Dam-building, reclamation and erosion of coastal zones, irrigation of arid land, urbanization, and industrialization are often very important locally, but seem unlikely to cause anything other than local environmental changes.

Richards (1986, 1990) studied the global environmental aspects of land-cover changes over the past three centuries and calculated that woodlands and forests have decreased in area by 1200 million ha and that croplands have increased

by 1200 million ha while grasslands and pasture have declined in area by about 500 million ha. Over the much shorter period of only 15 years from 1973 to 1988, the FAO (1990) estimated that croplands have increased by 57 million ha, pastures decreased by 11 million ha, and woodlands and forests decreased by 141 million ha. The rate of global urbanization is very difficult to estimate but seems likely to be at least 2 million ha year^{-1}. These estimates seem to indicate that land surface changes due to human activities are accelerating. Land surface changes in the future will be driven primarily by population change but will also depend and also on affluence/poverty, technological development, economics, and dietary preferences. In the 21st century, the greatest pressure for changes to the land surface seems likely to be in the tropics.

LAND SURFACE–CLIMATE INTERACTIONS

The relationships between the land surface and the global environment are very complex. The potential contributions to global environmental change due to land surface changes include: greenhouse gas emissions; aerosol (particulates and droplet) loading; surface reflectivity (albedo) and emissivity disturbances; and impacts on the surface roughness and surface hydrology. Land clearance and land-use changes are now recognized as having the potential to affect local, regional and global characteristics (Figure 1). However, the mechanisms by which large-scale changes occur and the types of feedback likely are, as yet, not understood. For example, Zhang *et al.* (1996) have identified a means by which tropical deforestation can produce global-scale impacts in addition to the known contribution to the greenhouse gas burden and disturbance to regional hydrology.

Henderson-Sellers (1995) illustrates the potential importance of land surface changes with a series of simple calculations. Assuming that one-quarter to one-third of the contribution to the projected doubling of the atmospheric CO_2 will be due to land-cover change (e.g., Houghton *et al.*, 1990, 1996) implies that its contribution to radiative forcing will be about $+1\,W\,m^{-2}$. The possible effect of aerosols from both human-induced and natural biomass

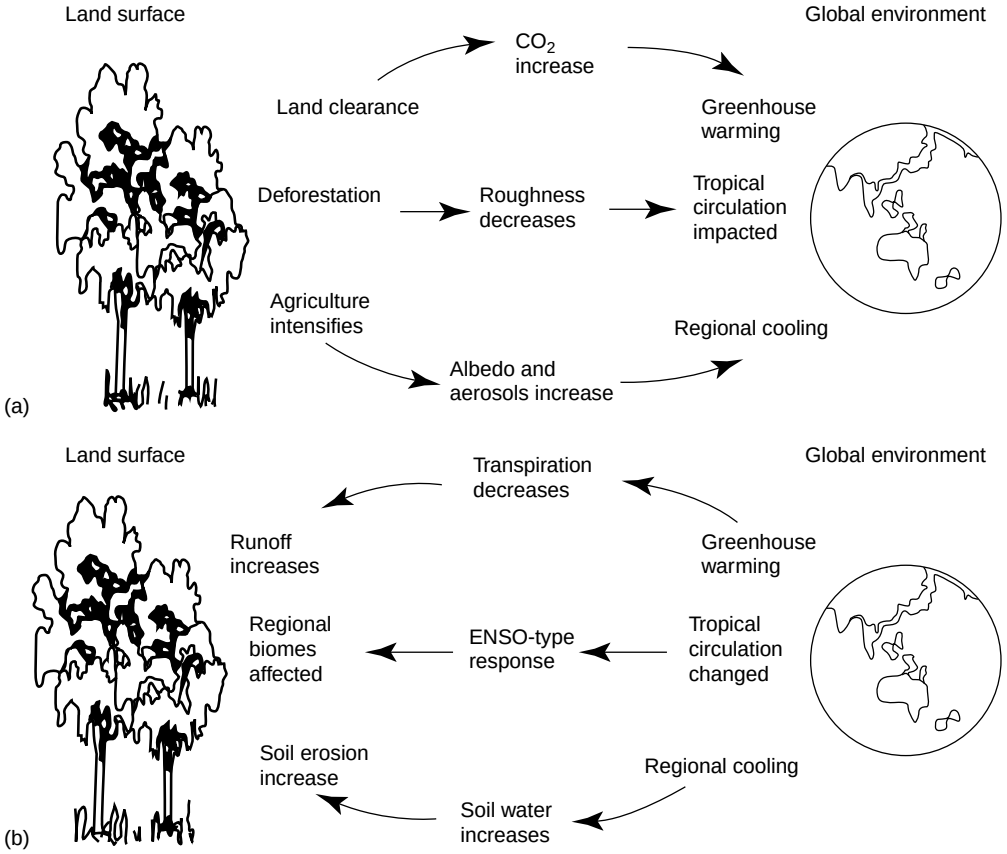


Figure 1 Schematic of some of the ways in which land surface changes and global environmental changes may interact: (a) land surface changes prompt disturbances that contribute to global changes; and (b) global environmental changes affect the land surface impacting vegetation, soils and hydrology. The examples (picked from a wide range of possible interactions) are illustrative also of the likelihood of synergies (e.g., between greenhouse warming and ENSO-like disturbances) and of feedback (e.g., intensification of agriculture associated with increased soil erosion)

burning is of a similar magnitude but in the opposite sense: a cooling of $-1\,W\,m^{-2}$ (e.g., Andreae, 1995). A plausible increase in continental albedo of +5% (from a forest albedo of 15% to a grassland albedo of 20%) would give a decrease in absorbed solar radiation of around $-3\,W\,m^{-2}$ (Henderson-Sellers and Gornitz, 1984). These very simplistic calculations suggest that the global environmental impact of land surface changes through surface albedo increases, combined with possible aerosol cooling, could very roughly counterbalance the currently projected radiative forcing from a CO_2 doubling (Figure 1). While the actual effects may be different, the need to include land surface changes as part of global change calculations is clear (*see* Land Cover and Climate, Volume 1).

In that over 70% of the energy absorbed into the climate system is absorbed at the surface and one-third of the surface is land, experiments with climate models have identified sensitivities of the global environment to conditions at the land surface, the hydrology and to the vegetation cover. For example, Shukla and Mintz (1982) show that moisture evaporated from the surface has a considerable impact on rainfall. In this early global climate model experiment, evaporation from the land surface was forced to be equal to either the potential evaporation (assuming land was fully wet) or set equal to zero (assuming land was totally dry) at all land-surface points. The calculated values of July precipitation were sharply reduced in the latter case and there was also a shift in the position of the remaining maxima of rainfall over the continental areas. This experiment, although testing a limit, indicated the importance of treating evapotranspiration from the land surface to accurately model the global hydrological regime and prompted an intensified focus on land surface modeling (*see* Earth System Processes, Volume 1).

REPRESENTATIONS OF LAND SURFACES

The treatment of the land surface in models has changed markedly over the history of climate modeling (e.g., McGuffie and Henderson-Sellers, 1997). The Budyko or *bucket* model dates back to 1969 (Manabe, 1969). This bucket (Figure 2a) assumed a maximum soil water depth, termed the *eld capacity*. The bucket fills when precipitation exceeds evaporation and, when it is full, excess water is assumed to run off. The bucket model has been demonstrated to be inadequate for heating the diurnal cycle, particularly when details of changes of the atmosphere and surface are needed. As a result, treatments of energy and moisture exchanges at the land surface have been made more detailed and, hopefully, their results have become more realistic (Figure 2b, cf. Figure 3).

It has also been found necessary to treat changes in land cover. Land surface change typically modifies the surface albedo and can also involve an alteration in the roughness of the surface: for example, removing trees decreases the roughness (Figure 1). In simulating the effects of changes in land cover, the canopy area and the resistance of the vegetation to transpiration may also be altered, as may the soil state and its characteristics, including drainage. Surface albedo and roughness have been found to be the two factors likely to be the most important factors in affecting the regional climate.

EFFECTS OF LAND SURFACE CHANGES

Two types of human-generated land surface changes, deforestation and desertification, have been simulated in a number of global climate models (GCMs). Among the relevant properties of tropical forests are that they have a very low surface albedo throughout the year because their leaf area and stem area are larger than those of any other vegetation and the trees are tall. Replacing the tropical forest with grasslands would lead to three primary changes at the land surface; similar changes would also occur in the event of desertification.

1. The surface albedo would be increased, which reduces absorption of solar radiation and causes a reduction in net radiation available at the land surface.
2. Reductions in the leaf area and stem area would lead to a decrease in the water holding capacity of the vegetation and thus reduce the re-evaporation of the intercepted precipitation and the flux of water from the soil moisture reservoir to the atmosphere through transpiration.
3. The grassland replacing the tropical forest (or, for desertification, the bare soil replacing the scrub) would be shorter and smoother than the rainforest so that the surface roughness is reduced and the surface frictional forcing is weakened.

When deforestation is simulated, changes in both albedo and roughness length occur, whilst for desertification only the surface albedo is usually modified because any roughness length change is assumed to be small (although the ratio of old to new roughness lengths is roughly commensurate in the two cases). Henderson-Sellers (1995) summarizes the mechanisms implicated in global change.

Tropical deforestation and desertification cause surface albedo increases of between about 0.01 and 0.10. Assuming annual incoming solar radiation of $300\,W\,m^{-2}$ (typical for the tropics), the absorbed solar radiation in these locations could be reduced by up to $30\,W\,m^{-2}$ (e.g., Mylne and Rowntree, 1992). If net surface longwave radiation remains unchanged, less energy is therefore available for latent or sensible heat flux from the surface and for transfer into the ground, causing the temperature to decrease. Secondary effects include decreased evaporation and decreased precipitation. The decrease in net radiative heating of the

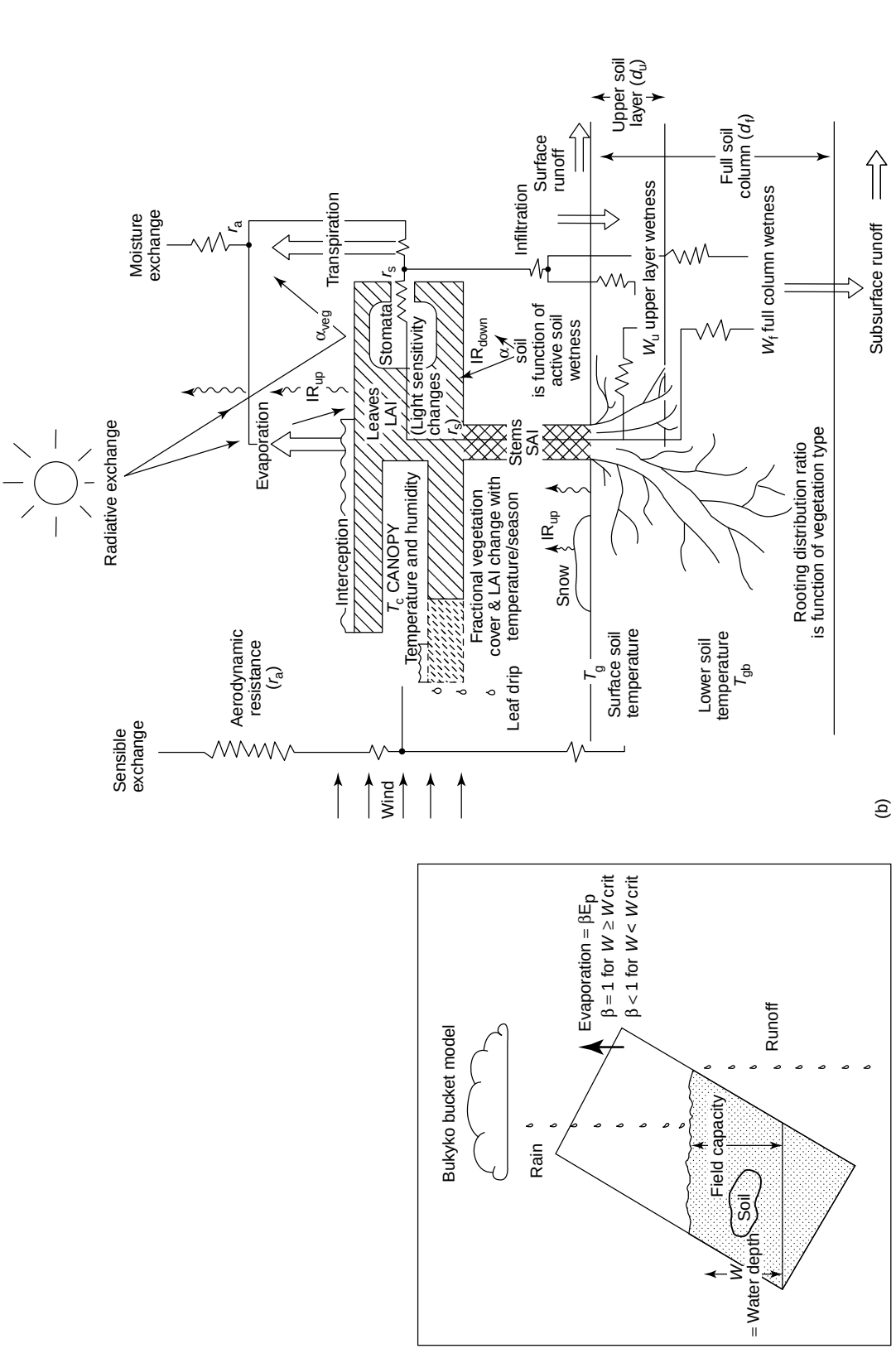


Figure 2 (a) An illustration of the simple bucket land surface scheme (from *A Climate Modelling Primer,* McGuffie and Henderson-Sellers, 1997. © John Wiley & Sons Limited. Reproduced with permission). (b) An illustration of the processes included in more complex land surface schemes. Such a change includes representations of the radiative, latent and sensible heat fluxes occurring at the surface, and simulates the movement of soil moisture below the ground and through the plants (from *A Climate Modelling Primer,* McGuffie and Henderson-Sellers, 1997. © John Wiley & Sons Limited. Reproduced with permission)

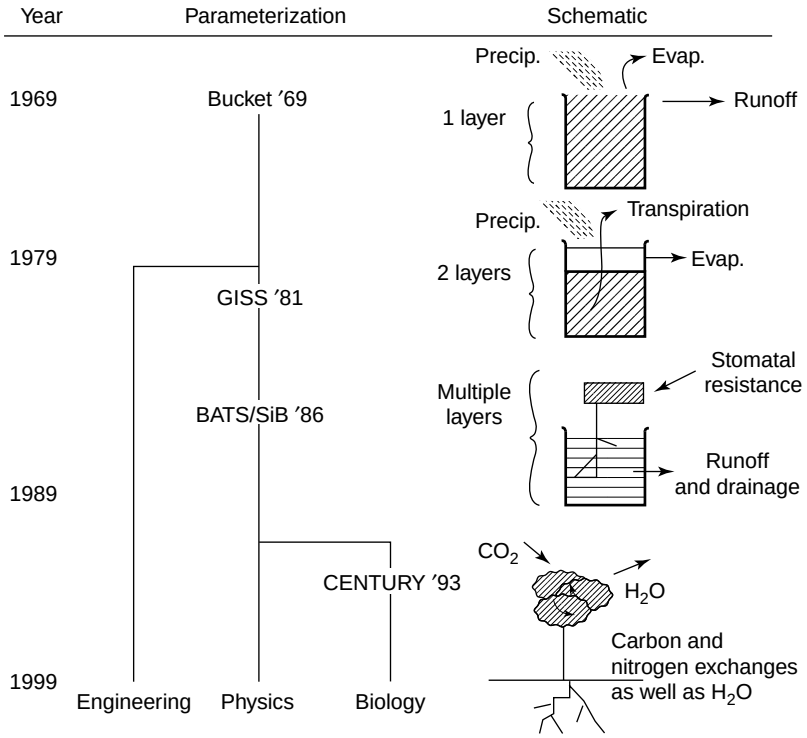


Figure 3 Schematic showing the time-line of development of soil moisture schemes used in models of global environmental change since the development of models to simulate the global climate. These range from the engineering solution of the Manabe bucket (Manabe, 1969), through the soil physics parameterizations of the Goddard Institute for Space Studies (GISS) GCM (Hansen *et al.*, 1981) and the biosphere–atmosphere transfer scheme (BATS) and simple biosphere (SiB) schemes (Dickinson *et al.*, 1986; Sellers *et al.*, 1986) to carbon sequestration models such as CENTURY (Parton *et al.*, 1993)

atmosphere and the surface leads (e.g., Charney, 1975) to a reduction in the upward movement of the air, reducing moisture convergence and rainfall. These changes in turn prompt changes in the surface water budget and hence in soil moisture. If the decrease in precipitation exceeds that in evaporation, drying out of the soil occurs, prompting further decreases in evaporation. As is evident in desert areas, this can lead to an increase in sensible heat flux and an increase in surface temperature even though the albedo increases. On the other hand, the combined impact of multiple processes may be sufficient to cause a larger decrease in evaporation than in precipitation, tending to increase soil moisture. Such a reduction in precipitation would tend to further increase albedo as vegetation and soils dry.

Deforestation causes a decrease in surface roughness from a forest roughness length of around 2 m to a grassland value of about 0.05 m (somewhat less for desertification). Decreasing the roughness reduces the turbulent fluxes of sensible and latent heat and of momentum between the land surface and the atmosphere. This reduction causes an increase in the surface temperature that prompts a negative feedback because the warming will tend to increase sensible and latent heat fluxes. The effect of an increase in surface temperature and thus increased longwave radiation from the surface is to decrease the net radiation, prompting decreased ascent and inhibiting precipitation. In addition to the reduction in evaporation due to the reduced roughness length, there is a further reduction caused by the reduction in the transpiration from the smaller amount of biomass (i.e., less vegetation and the much shorter roots of grasses than those of trees). This combined reduction in evaporative flux acts to further inhibit precipitation.

The decrease in surface roughness also decreases the turbulent momentum flux, which decreases the surface drag and increases the surface wind. Decreased drag results in decreased cross-isobaric flow and hence decreased moisture convergence. This further contributes to a reduction in the tendency for the air to rise and hence tends to decrease precipitation. A decrease in precipitation will cause a further decrease in soil moisture, although this will be somewhat counter-balanced by the decreased evaporation. Not only is the net outcome of these two effects uncertain but there are also possibilities for positive feedback on evaporation caused by the changes in soil moisture.

As can be surmised from this analysis, it is very difficult to anticipate the direction and magnitude of the regional environmental changes following from desertification and deforestation. The imposition of an albedo increase might

cool the surface and thus prompt atmospheric descent. The imposition of a reduced roughness length is likely to reduce evaporation, and hence warm the surface, but also to prompt descent or at least decreased ascent in the air.

The overall result of increasing albedo (whether from deforestation, salinization, or other factors) will depend crucially on the relative strengths of the processes, the strengths of the identified feedback, the time periods over which each operates, and other impacts, such as changes in cloudiness (e.g., reduced cloud cover is likely to allow increased absorption of solar radiation and thus increase the net radiation balance, thereby acting as a negative feedback on the imposed albedo increase itself).

Zhang *et al.* (1996) examined the surface and atmospheric energy budgets derived from a GCM in order to improve understanding of the simulated climatic changes after tropical deforestation. Generally, the canceling effects of increased surface albedo and decreased evapotranspiration explain the small changes in surface temperature over deforested regions. Their analysis also suggests that:

1. Cloud radiative forcing plays an important role in shortwave radiation processes, while the influence of increased surface albedo is mitigated dramatically by the decrease of cloud amount.
2. The effect of the change in net longwave radiation is larger than the effect of the shortwave radiation change, a result not emphasized in previous simulations.
3. Changes in cloud radiative forcing increase the daily variability of the surface temperature even though the daily mean surface temperature may only show small changes.
4. In spite of the increase of surface albedo over deforested regions, the net radiative energy heating the atmosphere is actually increased by the changes in cloud radiative forcing and longwave radiation.
5. The role of changes in latent heat flux is the most important factor in the net reduction of the atmospheric energy budget.

Based on this evaluation, Zhang *et al.* (1996) asserted that changes in latent heat flux, themselves the result of the reduction of surface roughness length and the changes in surface radiative energy, are the primary reasons for the calculated reduction of precipitation over deforested regions from which rainforest has been removed.

Given the number of interacting mechanisms and various types of situations, however, the net effect of feedbacks and differing magnitudes and time-scales of responses remains uncertain. In addition, it is not yet known how to relate these local impacts to regional changes such as atmospheric moisture convergence, which is mainly controlled by larger-scale circulation characteristics (Rowntree, 1991). The possibility of regional-scale atmospheric responses to land-cover change has been noted, for example, by, Shukla *et al.* (1990) with reference to tropical deforestation and by Charney *et al.* (1977) with reference to desertification. Major atmospheric features being investigated in current simulations include connections to the Hadley and Walker circulations and the tropical jets, because these dynamical processes can cause the regional influences to have effects outside of where they are initiated. Because the tropics are the primary energy source (because of high solar irradiation) and the moisture source (because of high sea surface temperatures) for the global climate, changes in the land surface in the tropics can influence global climatic conditions.

Given the many types of possible interactions, it is important to treat the state of the continental surface, particularly the biomes it will support, as coupled and integral parts of the global environment (e.g., Kump and Lovelock, 1995). The rates and types of land-use changes, both anthropogenic and in response to climate, are thus likely to modify the climate as well as the sources and sinks of greenhouse gases, aerosols and moisture. The resulting changes will also alter the locations in which land cover is sustainable and offer insights into the consequences of long-term changes in land use.

At present, the computational links between biology and climate in global environmental models are in a developmental stage. While global maps agree on the locations of the major deserts, ice caps and forests, many differences are evident when particular regions are examined in detail. The challenge being undertaken is to reconcile these results so that the land-surface in all its manifestations from the locus of human life to a driver of global climate is fully considered.

See also: Energy Balance Climate Models, Volume 1; Land Cover and Climate, Volume 1; Modeling Regional Climate Change, Volume 1; Biological and Ecological Dimensions of Global Environmental Change, Volume 2; Biome- BGC Ecosystem Model, Volume 2; Large Scale Biosphere- Atmosphere Experiment in Amazonia (LBA), Volume 2; Natural Systems: Impacts of Climate Change, Volume 2; Vegetation Ecosystem Model and Analysis Project (VEMAP), Volume 2; Deforestation, Tropical: Global Impacts, Volume 3; Global land cover and land use trends and changes, Volume 3.

REFERENCES

Andreae, M O (1995) Climatic Effects of Changing Atmospheric Aerosol Levels, in *Future Climates of the World*, ed A Henderson-Sellers, Elsevier, Amsterdam, 347–398.

Charney, J G (1975) Dynamics of Deserts and Drought in the Sahel, *Q. J. R. Meteorol. Soc.*, 101, 193–202.

Charney, J G, Quirk, W J, Chow, J H, and Kornfield, J (1977) A Comparative Study of the Effects of Albedo Change on Drought in Semi-arid Regions, *J. Atmos. Sci.*, 34, 1366–1385.

Dickinson, R E, Henderson-Sellers, A, Kennedy, P J, and Wilson, M F (1986) *Biosphere- atmosphere Transfer Scheme (BATS) for the NCAR Community Climate Model*, National Center for Atmospheric Research, Boulder, BO, Tech Note/ TN−275+STR.

FAO (1990) *FAO Production Yearbook*, FAO, Rome, Vol. 43.

Hansen, J, Johnson, D, Lacis, A, Lebedeff, S, Lee, P, Rind, D, and Russell, G (1981) Climate Impact of Increasing Atmospheric Carbon Dioxide, *Science*, 213, 957−966.

Henderson-Sellers, A (1995) Human Effects on Climate Through the Large-scale Impacts of Land-use Change, in *Future Climates of the World*, ed A Henderson-Sellers, Elsevier, Amsterdam, 433−475.

Henderson-Sellers, A and Gornitz, V (1984) Possible Climatic Impacts of Land Cover Transformations, with Particular Emphasis on Tropical Deforestation, *Clim. Change*, 6, 231−258.

Houghton, J T, Jenkins, G J, and Ephraums, J J, eds (1990) *Climate Change. The IPCC Scienti c Assessment*, Cambridge University Press, Cambridge.

Houghton, J T, Meira Filho, L G, Callander, B A, Harris, N, Kattenberg, A, and Maskell, K, eds (1996) *Climate Change 1995. The Science of Climate Change*, Cambridge University Press, Cambridge.

Kump, L R and Lovelock, J E (1995) The Geophysiology of Climate, in *Future Climates of the World*, ed A Henderson-Sellers, Elsevier, Amsterdam, 537−553.

Manabe, S (1969) Climate and the Ocean Circulation: 1, The Atmospheric Circulation and the Hydrology of the Earth s Surface, *Mon. Weather Rev.*, 97, 739−805.

McGuffie, K and Henderson-Sellers, A (1997) *A Climate Modelling Primer (and CD)*, 2nd edition, John Wiley & Sons, Chichester.

Mylne, M F and Rowntree, P R (1992) Modelling the Effects of Albedo Change Associated with Tropical Deforestation, *Clim. Change*, 21, 317−343.

Parton, W J, Scurlock, J M O, Ojima, D S, Gilmanov, T G, Scholes, R J, Schimel, D S, Kirchner, T, Menaut, J-C, Seastedt, T, Garcia Moya, E, Kamnalrut, A, and Kinyamario, J I (1993) Observations and Modeling of Biomass and Soil Organic Matter Dynamics for the Grassland Biome Worldwide, *Global Biogeochem. Cycles*, 7, 785−809.

Richards, J F (1986) World Environmental History and Economic Development, in *Sustainable Development of the Biosphere*, eds W C Clark and R E Munn, Cambridge University Press, Cambridge, 53−71.

Richards, J F (1990) Land Transformation, in *The Earth as Transformed by Human Action: Global and Regional Changes in the Biosphere over the Past 300 Years*, eds B L Turner, II, W C Clark, R W Kates, J F Richards, J T Mathews, and W B Meyer, Cambridge University Press, Cambridge, 163−178.

Rowntree, P R (1991) Atmospheric Parameterization Schemes for Evaporation over Land: Basic Concepts and Climate Modelling Aspects, in *Land Surface Evaporation*, eds T J Schmugge and J-C André, Springer Verlag, New York, 5−29.

Sellers, P J, Mintz, Y, Sud, Y C, and Dalcher, A (1986) A Simple Biosphere Model (SiB) for Use within General Circulation Models, *J. Atmos. Sci.*, 43, 505−531.

Shukla, J and Mintz, Y (1982) Influence of land-surface evapotranspiration on the Earth s climate, *Science*, 215, 1498−1501.

Shukla, J, Nobre, C, and Sellers, P J (1990) Amazon Deforestation and Climate Change, *Science*, 247, 1322−1325.

Zhang, H, McGuffie, K, and Henderson-Sellers, A (1996) Impacts of Tropical Deforestation. Part II: The Role of Large-scale Dynamics, *J. Clim.*, 9, 2498−2521.

Lapse Rate

The *lapse rate* is the rate of decrease of temperature per unit change of height.

The temperature of the atmosphere normally decreases with increasing altitude in the lower part of the atmosphere, the *troposphere*. The rate of decrease is termed the *lapse rate*, and is on a global average $6.5\,^{\circ}\mathrm{C}\ 1000\,\mathrm{m}^{-1}$ (or $3.5\,^{\circ}\mathrm{F}\ 1000\,\mathrm{ft}^{-1}$). This decrease continues upward to about 10−15 km reaching the tropopause, which is the boundary with the stratosphere. Within the stratosphere, the temperature increases or varies little with height, and the lapse rate is thus close to zero.

The Earth s surface and lower atmosphere are heated by solar radiation, while the upper atmosphere is cooled by radiation to space. It is thus not surprising that temperature declines with height. A more important factor, however, is the change in temperature produced by upward and downward motions of air. As an air parcel rises, the pressure of the surrounding air decreases; the air parcel therefore expands and its temperature drops. The reverse occurs as air descends. This rate of change is about $10\,^{\circ}\mathrm{C}\ 1000\,\mathrm{m}^{-1}$ ($5.5\,^{\circ}\mathrm{F}\ 1000\,\mathrm{ft}^{-1}$).

Rising air often contains water vapor that condenses as the air cools, forming clouds and rain and releasing the heat originally absorbed in evaporation (latent heat). Because of this heat release, the rising air parcel may become warmer, and therefore lighter, than its surrounding air. The resulting buoyancy accelerates its ascent, producing even more cooling and condensation. These processes of radiation, cooling and heating by rising and descending air, and condensation in rising air columns determine the actual ambient lapse rate.

The lapse rate in turn influences the rates of vertical overturning of air that disperses pollutants and drives storm systems. When the lapse rate is small, or even negative (an *inversion*, with temperature increasing with height), upward and downward movements of air are inhibited. For example, an inversion layer formed by strong cooling near the surface can trap pollutants over a city until solar radiation warms the layer enough to start convective mixing. As another example, a Los Angeles-type inversion based at 500 m or so (a so-called *subsidence inversion*) also traps pollutants in the air below. On a global scale, the degree of stability

or instability of the atmosphere as a whole is an important factor in determining the number and intensity of storm systems.

JOHN S PERRY *USA*

Last Glacial Maximum

The *Last Glacial Maximum* (LGM) was the time of maximum extent of the continental ice sheets during the most recent glacial period. The LGM peaked about 21 000 years ago (*see* Earth System History, Volume 1). The most recent glacial cycle began about 115 000 years ago, with major development of ice sheets starting at 75 000 years ago. Continental ice sheets, which could have been as much as 4000 m thick, covered large portions of North America and Eurasia. Major ice sheets included the Laurentide ice sheet in eastern North America, the Cordilleran ice sheet in western North America, the Fennoscandian ice sheet in northwestern Europe, the Barents and Kara Sea ice sheets in eastern Eurasian Russia, and the West Antarctic ice sheet. The LGM is defined by oxygen isotope stage 2 from ocean sediments. While the overall timing of the LGM has been based on oxygen isotope ratios of ocean sediments, it corresponds to land-based periods such as the Wisconsin (maximum extent in eastern North America), the Weichselian (maximum extent in western Europe), and the Wurm (maximum extent in the Alps) (Crowley and North, 1991). In addition to the presence of the ice sheets, CO_2 levels at the LGM were about 200 ppmv (compared to 265–275 ppmv from 10 000 years ago to the start of the industrial period in the early 19th century), and solar insolation was close to its present seasonal and latitudinal pattern (Kutzbach *et al.*, 1998). Sea level, was, on average, about 120 m lower than present (Fairbanks, 1989). During the LGM, temperatures are estimated to have been from 10 to 21 °C colder in Greenland, Antarctica, and tropical ice caps (Dahl-Jensen *et al.*, 1998; Severinghaus and Brook, 1999; Jouzel, 1987; Thompson *et al.*, 1995). Lowland tropical temperatures and sea surface temperatures, on the other hand, are estimated to have been about 5 °C colder (Stute *et al.*, 1995) or perhaps lesser amounts (CLIMAP, 1981). The world was more arid than present, although circulation changes resulted in certain areas being wetter.

REFERENCES

Crowley, T J and North, G R (1991) *Paleoclimatology*, Oxford University Press, New York, 339.

CLIMAP (1981) *Seasonal Reconstructions of the Earth's Surface at the Last Glacial Maximum*, Geological Society of America Map Chart Series MC-36.

Dahl-Jensen, D, Mosegaard, K, Gundestrup, N, Clow, G D, Johnsen, S J, Hansen, A W, and Balling, N (1998) Past Temperatures directly from the Greenland Ice Sheet, *Science*, 282, 268–271.

Fairbanks, R G (1989) A 17 000-year Glacio-eustatic Sea Level Record: Influence of Glacial Melting Rates on the Younger Dryas Event and Deep-ocean Circulation, *Nature*, 342, 617–637.

Jouzel, J, Lorius, C, Petit, J R, Genthon, C, Barkov, N I, Kotlyakov, V M, and Petrov, V M (1987) Vostok Ice Core: a Continuous Isotope Temperature Record over the Last Climatic Cycle (160 000 years), *Nature*, 329, 403–408.

Kutzbach, J, Gallimore, R, and Laarif, F (1998) Climate and Biome Simulations for the past 21 000 years, *Quatern. Sci. Rev.*, 17(6/7), 473–506.

Severinghaus, J P and Brook, E J (1999) Abrupt Climate Change at the end of the Last Glacial Period Inferred from Trapped Air in Polar Ice, *Science*, 286, 930–934.

Stute, M, Forster, M, Frischkorn, H, Serejo, A, Clark, J F, Schlosser, P, Broecker, W S, and Bonani, G (1995) Cooling of Tropical Brazil (5 °C) During the Last Glacial Maximum, *Science*, 269, 379–383.

Thompson, L G, Mosley-Thompson, E, Davis, M E, Lin, P N, Henderson, K A, Cole-Dai, J, Bolzan, J F, and Liu, K B (1995) Late Glacial Stage and Holocene Tropical Ice Core Records from Huascaran, Peru, *Science*, 269, 47–50.

BENJAMIN S FELZER *USA*

Latent Heat

Latent heat is the internal potential energy of a substance by virtue of its molecular configuration and structuring described by the phase or state of the substance, i.e., solid, liquid, or vapor. This potential energy is distinct from the internal mean molecular kinetic energy measured by temperature. A phase change results in a transformation between latent heat and sensible heat. Depending on the direction of transformation, the result is an internally produced internal source or sink of sensible heat for the substance. This source or sink affects the internal mean molecular kinetic energy (temperature) of the substance in the same way as sensible heat transfer between the substance and surrounding material. Latent heat energy is thus latent or hidden from the internal mean molecular kinetic energy until phase change occurs, when it becomes manifest as a source or sink of sensible heat for the substance.

In environmental applications, latent heat usually refers to the latent heat associated with water. Water exists readily in all three phases (ice, liquid and vapor) and commonly undergoes phase changes which are important for

influencing temperature in the atmosphere and at the Earth s surface. For the case where the air temperature is less than that of the human body, a human hand held up in the wind will feel cooler if it is wet than if it is dry. For the dry hand situation, the hand feels cool because of a sensible heat transfer from the hand to the air. For the wet hand situation, the water on the hand will evaporate (a phase change from liquid to vapor), which converts sensible heat to latent heat, representing a sensible heat loss from the water which lowers its temperature. This cooling increases the sensible heat transfer from the human hand, which then feels cooler.

The evaporation of water at the Earth s surface and condensation higher up in the atmosphere results in a net transfer (flux) of latent heat from the land and ocean surface to the atmosphere and thereby is a net sensible heat source for the atmosphere. This transfer between the Earth s surface (ocean and land) and atmosphere is an important factor in the global energy balance and temperature distribution. The upward latent heat energy transfer is three times as large as the sensible heat transfer and accounts for 15% of the total upward energy transfer on a global and annual average basis. (Note: radiation accounts for 80% of the transfer of energy from the surface to the atmosphere.)

DAVID HOUGHTON *USA*

Lifetime (of a Gas)

Lifetime of a gas in the atmosphere is a characteristic of its rate of removal from the atmosphere. The *e-folding lifetime* is generally used, although *half-life* or *half-lifetime* is sometimes given. The *e*-folding lifetime is the time required for the atmospheric concentration of a gas to decrease to $1/e$, or 36.79% of its original concentration in that $e = 2{\bullet}7183$. The half-lifetime is the time required for the concentration of the gas to decrease to half of its original value.

The total lifetime of the gas can be represented in terms of the partial lifetimes due to different loss processes, such as those due to chemical or photochemical reactions, loss to the ocean or soil, or washout. The total lifetime, τ, is related to the sum of the partial lifetimes, τ_i, where i represents the individual loss processes through the identity

$$\frac{1}{\hat{\tau}} = \sum \frac{1}{\tau_{\mathrm{i}}} = \frac{1}{\tau_1} + \frac{1}{\tau_2} + \frac{1}{\tau_3} + \cdots \qquad {\bullet}1{\bullet}$$

In atmospheric chemistry, atmospheric lifetime of a gas is often used to denote the turnover time. Turnover time (T) is the ratio of the mass (M) of the atmospheric content of the gas to the rate of removal (S), such that $T = M{\bullet}S$. For most gases, the turnover time is equivalent to the adjustment time or response time (T_{a}), the timescale characterizing the decay of an instantaneous pulse emission of the gas into the atmosphere. For example, for chlorofluorocarbon (CFC)-11 ($CFCl_3$), $\tau = T = T_{\mathrm{a}} = 50$ years due to its removal via photochemical processes.

Where the removal frequency is not constant or where several reservoirs exchange with the atmosphere, the equality between T and T_{a} no longer holds. An important example of this is carbon dioxide (CO_2). Because of the rapid exchange of CO_2 between the atmosphere and the oceans and the terrestrial biosphere, the residence time of a particular CO_2 molecule is only about four years. However, a high fraction of the CO_2 removed from the atmosphere each year is returned to the atmosphere from these reservoirs within a few years. In this case, the adjustment time is more representative of the actual removal of atmospheric CO_2 (the net lifetime of all additional CO_2 molecules in the atmosphere). Calculating this net response time gives a result of roughly 100 years, which is primarily dependent on the rate of transport of CO_2 from the upper levels of the ocean to deeper layers. For CO_2, the adjustment time varies with time, being shorter just after the addition of a pulse as the CO_2 pulse redistributes in part to the land and upper ocean, and then longer once the land and upper ocean have reached a balance with the atmosphere. Methane (CH_4) is another gas where T does not equal T_{a} as a result of a nonlinear relationship between CH_4 and its loss through reaction with OH.

DONALD J WUEBBLES *USA*

Lightning

A natural electrical discharge in the atmosphere of several kilometers in length is called a *lightning ash*. These flashes are usually associated with cumulonimbus clouds (thunderclouds), but also occur in nimbostratus clouds, snowstorms, dust storms, and sometimes in the erupting gas of a volcano. The flash originates from a region of net charge, either positive or negative, when the electric field exceeds the dielectric strength (insulator strength) of air. Within a thunderstorm the flash can occur within a cloud, between clouds, or between a cloud and air, and these are all referred to as *cloud ashes*. If the flash occurs between the cloud and the ground, the flash is called a ground flash. Cloud flashes are the most common, followed by *ground ashes*.

Only cloud-to-ground lightning flashes are composed of one or more lightning strokes, which the eye can perceive as

a flicker. Although these flashes are less frequent than cloud flashes, they represent about 20% of lightning flashes and cause virtually all the damage associated with lightning. The cloud-to-ground lightning flash begins with electrical breakdown in the cloud from the charged region. A faint luminous process in regular distinct steps, of approximately 50 m length, at time intervals of 50 μs, descends in a downward branching pattern toward the ground. Carrying currents on the order of hundreds of amperes, this stepped leader or initial stroke propagates at a velocity of about $1•5 \times 10^5\,\mathrm{m\,s^{-1}}$, or one 2/1000th the speed of light. Within 50 m or so of the ground, the stepped leader is met by an upward discharge moving at one-third the speed of light which produces the luminous return stroke with peak currents averaging 30 000 amperes and temperatures of 30 000 K, or five times hotter than the surface of the Sun. Several subsequent strokes usually occur and have lower peak currents. The subsequent strokes complete the lightning flash, all occurring within about 200 ms.

RICHARD E ORVILLE *USA*

Lightning and Atmospheric Electricity

Colin Price
Tel Aviv University, Ramat Aviv, Israel

Although invisible to the eye, we are continuously surrounded by natural electric elds and electric currents owing in the atmosphere. This electricity in the atmosphere is generated by global thunderstorm activity, which results in 50- 100 lightning ashes every second somewhere on the globe. These electric elds and currents would disappear within a few minutes if all thunderstorm activity were to cease on our planet. The global electric parameters can easily be monitored in fair-weather regions of the globe, and in recent years researchers have shown that these global electric parameters may be useful tools for studying global climate change.

In 1752, Lemonnier discovered that the air above the Earth had a persistent electric field, which in fair-weather conditions is directed downwards, and has an average magnitude of 130 Volts meter^{-1} ($\mathrm{V\,m^{-1}}$). Knowing the dielectric coefficient of air, it was found that the charge on the Earth, as a result of this field, was negative, and has a magnitude of approximately half a million Coulombs. Since air is an excellent insulator, it was believed for more than 100 years that this was a permanent static feature of the Earth System. In 1887, Linss discovered that the atmosphere had a slight conductivity due to ions in the atmosphere. This resulted in the measurements of fair-weather atmospheric currents of approximately $2\,\mathrm{pA\,m^{-2}}$ flowing from the atmosphere to the Earth s surface. Over the entire Earth s area, this results in a current of 1000 A, which should neutralize the Earth s charge in less than 10 min. However, the charge remains.

In the 1920s, it was suggested that in order to maintain the Earth s negative charge there needs to be a generator. Global thunderstorm activity was proposed as this generator, and it was shown soon after that the diurnal variations of the surface electric field, known as the Carnegie curve, are well correlated with the diurnal variations of global thunderstorm activity.

With the discovery of the ionosphere in the 1920s, a global picture was proposed, called the *global atmospheric electric circuit*. The overall picture is that of a spherical capacitor formed by a conducting Earth, a conducting upper atmosphere, and a leaky dielectric in between. Currents in the atmosphere are generated in regions of thunderstorms by lightning and point discharge currents (St Elmo s fire). These currents flow upward toward the ionosphere. At altitudes of 60–100 km the atmosphere becomes highly conductive, and these currents spread laterally around the globe, returning to the Earth s surface in the fair-weather regions of the world, and then flowing through the conductive Earth back toward the thunderstorm regions, closing the global circuit. The resultant ionospheric potential between the ionosphere and the Earth s surface is approximately 250 kV.

The variability of the global circuit can theoretically be monitored from only a single site on the Earth s surface. Indeed, simultaneous measurements of the global circuit on opposite sides of the globe show remarkable agreement. This implies that global thunderstorm activity that varies on hourly, daily, seasonal and interannual time scales can also be successfully monitored via the global circuit from only a few stations. Furthermore, because lightning and thunderstorm activity are closely related to the global climate system, changes in the global climate may be reflected in changes in the global electric circuit.

Recently it has been shown that the variability of the global electric circuit, measured from a single location, is well correlated with tropical and global surface temperatures, as well as other important climate parameters such as upper tropospheric water vapor. Moreover, modeling studies using global climatic models indicate that global lightning frequencies are very sensitive to changes in global surface temperatures.

See also: Atmospheric Electricity, Volume 1.

FURTHER READING

Rycroft, M J, Israelsson, S, and Price, C (2000) The Global Atmospheric Electric Circuit, Solar Activity and Climate Change, *J. Atmos. Solar-Terrestrial Phys.*, 62, 1563–1576.

plants); in others, energy for organisms (heterotrophs) is dominated by plant material originating in the catchment. A third source of energy comes from chemical reactions, and some organisms (chemoautotrophs) are able to facilitate these to derive their energy.

PIERRE HORWITZ *Australia*

Limnology

The word *limnology* stems from *limnos*, Greek for lake, originally implying a freshwater, still and permanent water body; this stereotypic notion has been expanded in recent usage to include:

- flowing and non-flowing waters (ponds, lakes, creeks, rivers, swamps, etc.);
- constructed wetlands (drains, channels, dams, reservoirs, etc.);
- subterranean water (namely aquifers, cave water, water-saturated substrates).

Limnology includes those fields dealing with bodies of freshwater, including the nature of a basin, the substrate and its form (geology and geomorphology), the way in which water moves over the surface or underground (hydrology), movement within a water body (hydrodynamics), or interaction between the water and the atmosphere (meteorology). Other central fields include the chemical constituents and processes in water (water chemistry), along with its contamination and use (environmental sciences). Studies of the microbes, plants and animals which live in, or rely on, water bodies (biology), their productivity and utilization (for example, aquaculture) and the way they interact with each other and their non-living surroundings, indicate that limnology can be regarded as a significant branch of ecology.

Important differences in the subject matter of the science arise from:

- Degrees of permanence of the water (from permanently inundated, through seasonally and predictably filled, to intermittently and unpredictably filled hollows).
- Degrees of salinity encountered in inland waters, which range from fresh at one end of a continuum of ionic concentrations, to brackish, saline and hypersaline waters; anions and cations originate from sea spray, groundwater and/or soil and rocks. Salt lakes can be internally drained evaporative basins which accumulate salts, or those that occur in catchments influenced by the discharge of saline groundwater.
- Energy source: some waters have energy pathways dominated by photosynthetic organisms (photoautotrophs) living in the water itself (like algae and water

Lithosphere

The rigid outermost part of the solid Earth is called the *lithosphere*. Its outer boundary forms the surface of the land and the floor of the ocean. The lithosphere interacts with the overlying convecting atmosphere and hydrosphere and with the convecting mantle below. Organisms live on and interact with the surface of the lithosphere and bacteria live inside it to depths of perhaps 1 km. Time scales of interaction at the Earth s surface range from fractions of a second, in lightning strikes, through tens of thousands of years, in soil formation, to hundreds of millions of years in basins of sedimentary rock deposition. The lithosphere is broken into about a dozen large plates that together completely cover the Earth s surface. The plate boundaries form a continuous network in which most of the Earth s volcanic activity, earthquakes and mountain-building are concentrated. Lithospheric plates move horizontally with respect to each other in response to ridge-push and slab-pull forces applied at their margins. Although the plates rotate rigidly across the Earth s surface at velocities in the range of $2-10\,\mathrm{cm\,year^{-1}}$ they respond by downward flexure, on wavelengths of 100–1000 km, to loads (including ice-sheets and newly deposited sedimentary rocks) applied from above. Areas centered on the Baltic Sea and on Hudson s Bay are presently flexing upward at a few $\mathrm{mm\,year^{-1}}$ in response to the removal, within the past 20 000 years, of ice sheet loads.

New lithosphere is continuously being generated at oceanic spreading centers and in complementary destruction, oceanic lithosphere up to 170 million years old is being taken into the Earth s interior at subduction zones. This cyclical process, called *plate tectonics*, which serves to remove most of the Earth s internally generated heat, has been going on for about 4 million years. Production of the continents, which are less susceptible to recycling than is the ocean floor, has been a substantial by-product of the plate tectonic process over that time. About 25% of the heat generated in the Earth s interior is conducted to the surface through the lithosphere at a gradient of approximately $20\,^{\circ}\mathrm{C\,km^{-1}}$. That gradient can be strongly perturbed by the passage of hot or cold fluids and by temperature changes

at the Earth s surface in response to global climate change. The mechanical lithosphere is divided into chemically distinct upper and lower parts. The upper part, called the *crust*, is made of rocks containing 50% silica under the ocean floor and of rocks containing 65% silica under the continents. The lower part of the lithosphere, the *lithospheric mantle*, is everywhere made of rock containing 42% silica. Solid Earth scientists take great interest in the behavior of the lithosphere in plate boundary zones where tectonic activity is concentrated. At spreading center boundaries in oceans, the thickness of the lithosphere approaches zero, although it rapidly thickens as newly formed ocean floor cools and ages. In mountain belts at active convergent boundaries the crust thickens and becomes hot. Deep under the mountains the temperature in degrees Kelvin exceeds 65% of the melting points of the constituent rocks, at which temperature the lithosphere loses its rigidity.

See also: Earth System History, Volume 1.

KEVIN BURKE *USA*

Little Ice Age

Michael E Mann
University of Virginia, Charlottesville, VA, USA

*The term Little Ice Age was originally coined by F Matthes in 1939 to describe the most recent 4000 year climatic interval (the Late Holocene) associated with a particularly dramatic series of mountain glacier advances and retreats, analogous to, though considerably more moderate than, the Pleistocene glacial uctuations. This relatively prolonged period has now become known as the Neoglacial period. The term Little Ice Age is, instead, reserved for the most extensive recent period of mountain glacier expansion and is conventionally de ned as the 16th- mid 19th century period during which European climate was most strongly impacted. This period begins with a trend towards enhanced glacial conditions in Europe following the warmer conditions of the so-called medieval warm period or medieval climatic optimum of Europe (*see *Medieval Climatic Optimum, Volume 1), and terminates with the dramatic retreat of these glaciers during the 20th century. While there is evidence that many other regions outside Europe exhibited periods of cooler conditions, expanded glaciation, and signi cantly altered climate conditions, the timing and nature of these variations are highly variable from region to region, and the notion of the Little Ice Age as a globally synchronous cold period has all but been dismissed (Bradley and Jones, 1993; Mann* et al., *1999). If de ned as a large-scale event, the Little Ice Age must instead be considered a time of modest cooling of the Northern Hemisphere, with temperatures dropping by about 0.6 °C during the 15th- 19th centuries (Bradley and Jones, 1993; Jones* et al., *1998; Mann* et al., *1998, 1999).*

Documentary accounts of dramatic mountain glacier retreats and advances during past centuries, widespread historical documentation of weather conditions (e.g., Pfister, 1995, 1998) and even a handful of several centuries-long thermometer measurements (e.g., Bradley and Jones, 1993) provide incontrovertible evidence of the occurrence of the Little Ice Age in Europe and other regions neighboring the North Atlantic during the 16th–19th centuries. This climatic era has, in fact, been pictorially captured in paintings detailing the greatly expanded range of various mountain glaciers in the French and Swiss Alps during past centuries. The juxtaposition of such early artists renderings of such glaciers against their modern photographic counterparts, provides a particularly graphic illustration of the dramatic climatic changes associated with the Little Ice Age in Europe (Figure 1).

These dramatic glacial advances often had important practical consequences for nearby human populations. In the Chamonix valley near Mont Blanc, France, numerous farms and villages were lost to the advancing front of a nearby mountain glacier. The damage was so threatening that the villagers summoned the Bishop of Geneva to perform an exorcism of the dark forces presumed responsible (this procedure, as for most human attempts at weather modification, does not appear to have been successful). Such societal threats were common during the late 17th and early 18th centuries, as many glaciers expanded well beyond their previous historical limits. Colder conditions combined with altered patterns of atmospheric circulation, appear to be tied to the prevalent crop failures in the more northern areas of Europe of the time. There are widespread reports of famine, disease, and increased child mortality in Europe during the 17th–19th century that are probably related, at least in part, to colder temperatures and altered weather conditions. Certainly, not all consequences of the associated climate changes were deleterious for European society. In London, the freezing of the Thames River, commonplace during the era, was celebrated with a winter carnival. The colder climate, furthermore, appears to have served as inspiration for writers of the time. The greater frequency of cold, icy winters sentimentally framed author Charles Dickens notion of the old-fashioned white Christmas. The unusually cold summer of 1816 (the year without a summer – see discussion herein) forced Mary Shelley to spend her summer vacation at Lake Geneva indoors, where she and her husband entertained each other with horror

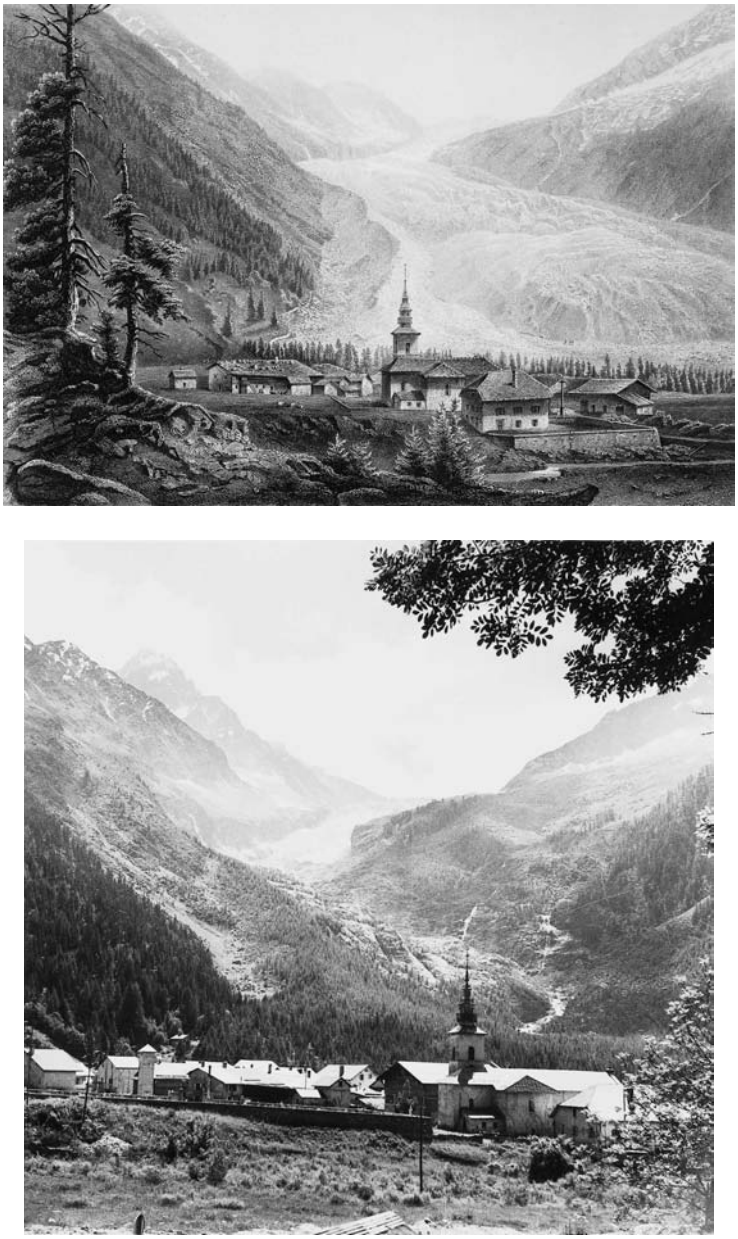


Figure 1 A portrait of the Argentiere glacier in the French Alps from an etching made between 1850 and 1860 just prior to its dramatic withdrawal, and a modern photograph of the glacier from a similar vantage point taken in 1966. (Reproduced by permission of Doubleday and Company, Inc., from Le Roy Ladurie, E, 1971 (Plates XXI and XXII))

stories, one of which resulted in her writing the novel *Frankenstein* (see Le Roy Ladurie, 1971).

The Little Ice Age may have been more significant in terms of increased variability of the climate, rather than changes in the average climate itself. The most dramatic climate extremes were less associated with prolonged multiyear periods of cold than with year to year temperature changes, or even particularly prominent individual cold spells, and these events were often quite specific to particular seasons. In Switzerland, for example, the first particularly cold winters appear to have begun in the 1550s, with cold springs beginning around 1568: the year 1573 had the first unusually cold summer (Pfister, 1995). The increased variability of the climate may have led to alternations between unusually cold winters and relatively warm summers. A severe winter preceded the hot summer that precipitated the Great Fire of London in 1666. A harsh winter followed by a warm summer may have added to the discontent of peasants who stormed the Bastille in Paris during the summer of 1789.

The cooling of the Little Ice Age has frequently been blamed for the demise of Norse settlements in Greenland

that had been established during the early centuries of the second millennium. This premise appears, however, to have only limited validity. Expanded sea ice extent in the North Atlantic certainly created problems for fishermen in Iceland and Scandinavia, and for the Norse settlements in Iceland and Greenland. Increased winter sea-ice cover closed off previously accessible trade routes between Scandinavia and Greenland during the late 14th century, cutting off trade with mainland Europe, upon which the Norse settlements relied. The collapse of Norse colonies in Greenland, however, represented a complex reaction to changing climate conditions (including seasonal precipitation/snowfall as well as temperature patterns), and their interaction with societal dynamics, and cannot be understood simply in terms of a lowering of temperatures in the region. In fact, as discussed below, temperature variations in western Greenland in past centuries bear only a limited resemblance to those in Europe, which conventionally have best defined the Little Ice Age.

Outside of the North Atlantic region, the large-scale signature of the Little Ice Age becomes even less clear. Unlike the true ice ages of the Pleistocene, which were marked by a clear global expression associated with dramatic growth of all the major continental ice sheets and a substantial lowering of global temperatures (probably $2-3\,^{\circ}C$ below current levels), the available evidence does not support the existence of a continuous period of cooler global temperatures synchronous with cold conditions in Europe. What evidence is available suggests, instead, generally colder conditions anywhere from the 13th through 19th century, quite variable in timing from region to region, and in most cases punctuated with intermittent periods of warmth (see Bradley and Jones, 1993; Pfister, 1995). Direct indications of climate variability (e.g., long thermometer measurements or reliable historical documentary records) are rarely available outside Europe and neighboring regions. There are some long instrumental climate records in North America dating back to the mid 18th century and historical information exists in the form of documentary evidence from parts of Canada and the northeastern US from early European settlers, and anecdotal reports from Native Americans in western North America. Such evidence gives a mixed picture of the Little Ice Age in North America. For example, the 17th century, the coldest century in Europe, does not appear to have been unusually cold in North America. By contrast, during the 19th century, as Europe was recovering from Little Ice Age conditions, North America was experiencing some of its coldest temperatures. There are reports, for example, that New York harbor froze over during this period of time. There is also some long-term human documentary evidence of climate changes in Russia and China that provide yet a different picture of temperature variations in past centuries from that of Europe (see Bradley and Jones, 1993 and references therein).

There is geological information from the position of moraines or till left behind by receding glaciers that provide a more global, albeit indirect, picture of the advances (and, less precisely, the retreats) of mountain glaciers. Such evidence suggests, for example, increased glaciation in certain regions of the world outside Europe prior to the 20th century, including Alaska in North America, and New Zealand and Patagonia in the Southern Hemisphere (see Grove, 1988). However, the precise timing of glacial advances in these regions (and even between the western and eastern Alps) differs considerably from region to region, suggesting the possibility that they represent roughly coincident, but independent regional climate changes, rather than globally synchronous increased glaciation. Owing to the complex balance between local changes in melting and ice accumulation, and the effects of topography, all of which influence mountain glacier extent, it is difficult to ascertain the true nature of climate change simply from evidence of retreat of mountain glaciers alone. For example, both increased winter precipitation (through greater accumulation) and lower summer temperatures (through decreased melting or ablation) can lead to increases in glacial mass. Furthermore, the inertia of large glaciers dictates that they respond relatively slowly, and with delays of decades to centuries, in response to any contemporaneous climate changes.

Other information is thus necessary to assess the global-scale climate variations of past centuries. A variety of other types of information is fortunately available to help provide a truly global-scale picture of climate change during past centuries (Bradley and Jones, 1993). These types include a few long-term historical documentary records (particularly outside Europe), supplemented by proxy climate records such as growth and density measurements from tree rings, laminated sediment cores, annually resolved ice cores, isotopic indicators from corals, and long-term ground temperature trends from borehole data. While these indirect measurements of climate change vary considerably in their reliability as indicators of long-term temperatures (varying in the degree of influence by non-climatic effects, the seasonal nature of the climate information recorded, and the extent to which the records have been verified by comparison with independent data), they are, nonetheless, essential for documenting global-scale patterns of temperature change in past centuries.

Comparing temperature estimates over the Northern Hemisphere and globe during past centuries from these different sources provides considerable insight into the regional variability and extent of the Little Ice Age (Figure 2). Slightly preceding the increased growth of mountain glaciers across Europe during the 17th–19th centuries is evidence of a depression of temperatures in the region (Central England – panel e) relative to modern levels by approximately $0.4\,^{\circ}C$ during the period 1500–1800 and by more than $0.6\,^{\circ}C$ during the 17th century period

of peak cold. The conclusion that the region exhibited its coldest conditions during the 17th century, and began warming significantly during the 19th century, is independently confirmed from tree ring reconstructions of European summer temperatures (see Jones *et al.*, 1998). A Little Ice Age is not as plainly evident in temperature estimates for Western Greenland (panel d), reinforcing the notion that the collapse of Norse civilization in Greenland was not simply a response to cooling temperatures. Temperature trends in Scandinavia (panel f) show some similarities with both those of Greenland and central England, further emphasizing the importance of regional variation in temperature trends over the past few centuries even in the North Atlantic and neighboring regions. In fact, the Little Ice Age appears to have been most clearly expressed in the North Atlantic region in terms of altered patterns of polar atmospheric circulation. Such altered patterns of circulation likely impacted seasonal snowfall patterns in a way that the Norse could not easily adapt to, and may explain differences between temperature variations in Europe and those in Iceland and Greenland. For example, the winter of 1833/1834, an unusually warm winter in central Europe,

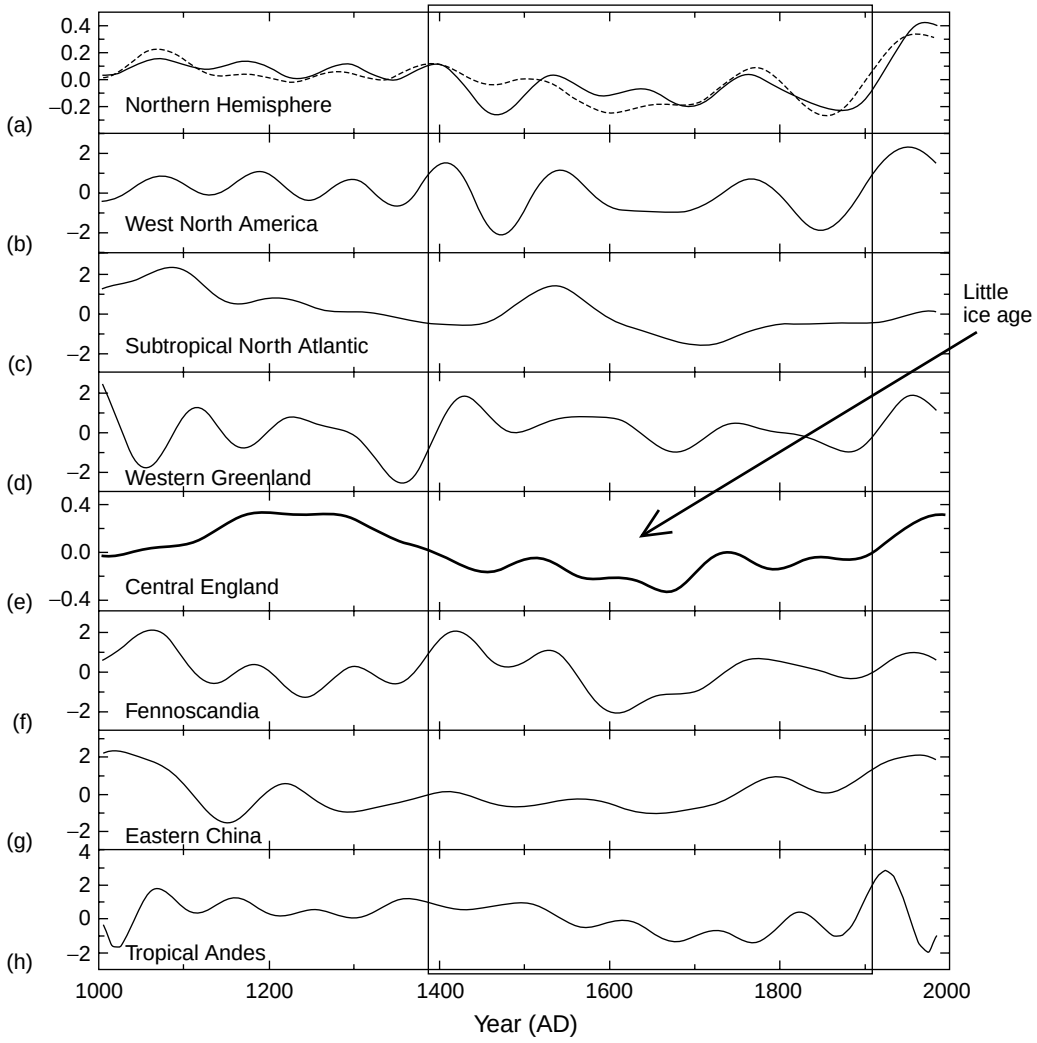


Figure 2 Estimated relative temperature variations during the past millennium for different regions. The records have been smoothed to emphasize century and longer-term variations. (a) Northern Hemisphere mean temperatures as estimated for the annual mean over the entire hemisphere (solid – Mann *et al.*, 1998, 1999) and over the extratropical region during the warm season (dashed – Jones *et al.*, 1998) based on global databases of proxy climate indicators, (b) tree ring data from western North America, (c) sediment record from the Sargasso Sea of the tropical North Atlantic, (d) ice cores from western Greenland, (e) a combination of thermometer, historical and proxy data records from central England, (f) tree ring data from Fennoscandia, (g) phenological evidence from eastern China and (h) ice core data from the tropical Andes of South America. Temperature scale is in °C for (a) and (e), and indicates relative temperature variations otherwise. Panel (e), which best defines the European Little Ice Age, is highlighted, with the large rectangle (extending from about AD 1400 to 1900) indicating more broadly the consensus among regions of the coldest period

appears to have been associated with dramatic changes in storm tracks over Europe, consistent with colder than normal conditions evident in Iceland that same year. While the 17th century appears to represent the timing of peak cooling in Europe, the 19th century was more clearly the period of peak cold in North America (panel b). In both the subtropical North Atlantic (panel c) and the tropical Andes of South America (panel h), peak cooling is evident during the 17th and 18th centuries. Even farther a field in eastern China (panel g), there is less evidence of any distinct cold period during the latter centuries of the millennium, with temperatures rather relatively uniformly depressed from about AD 1100–1800.

Some of the regional variability evident during the Little Ice Age can be understood in terms of changes in atmospheric circulation patterns. Such patterns, particularly the North Atlantic Oscillation (*see* North Atlantic Oscillation, Volume 1) – the dominant mode of atmospheric circulation variation in the North Atlantic and neighboring regions – have a particularly strong influence on winter temperatures in Europe. While the coldest year overall in Europe, 1838, was indeed one of the coldest over much of the Northern Hemisphere (the late 1830s were generally quite cold, perhaps due to the effects of the large volcanic eruption in Coseguina, Nicaragua during 1835 – see discussion herein), conditions were, nonetheless, relatively mild over significant portions of Greenland and Alaska. In fact, unusually cold, dry winters in central Europe (e.g., 1–2 °C below normal during the late 17th century) appear to have been associated with the flow of continental air from the northeast towards western Russia and Europe, conditions which, along with warmer than normal temperatures in Greenland and other regions, are consistent with the positive phase of the North Atlantic Oscillation. Warmer than normal winters in Europe, and cooler than normal temperatures in Greenland, tend to arise during the opposite phase.

While the peak maximum cooling occurred at quite different times throughout the Northern Hemisphere, an overall pattern does emerge when one composites regional variations into an estimate of large-scale mean temperature changes (see Mann *et al.*, 1998, 1999; Jones *et al.*, 1998). For the Northern Hemisphere on the whole (both annual mean and extratropical summer – see Figure 2, panel a), the period 1400–1900 appears to have been moderately cooler (approximately 0.3 °C) than the earlier period AD 1000–1400 and about 0.8 °C colder than the late 20th century. Following the cold late 15th century, the 17th and 19th centuries appear as the coldest centuries within this period (although, as discussed above, the spatial pattern of this cooling is quite distinct for the two periods). If one wishes to define the Little Ice Age proper as a time of large-scale cooling, it must be defined as a period of only moderately cooler Northern Hemisphere temperatures from about 1400–1900 (see Figure 2), preceding the rapid warming of the 20th century. For the Southern Hemisphere, evidence for a comparable Little Ice Age is far more diffuse (e.g., Jones *et al.*, 1998).

The existence of the Little Ice Age (whether defined by the particularly cold conditions in Europe during the 16th–18th centuries, or the more modest large-scale cooling of the 15th–19th centuries) invites questions as to what factors may have led to such a cooling.

This unusual period in climate history occurred before the likely influence of human activity (e.g,. the burning of fossil fuels associated with the industrial revolution). Though some of the long-term cooling of the climate prior to the 20th century might have been associated with astronomical factors, such factors cannot explain the pronounced and relatively short-duration cooling observed in many regions.

The explanation for the Little Ice Age must thus lie in other natural causes, whether associated with external forces, or internal noise in the climate system. The injection of sunlight-reflecting sulfate aerosols by explosive volcanic eruptions, for example, may be responsible for some of the cooling of the early and mid 19th century, in particular (see Lean *et al.*, 1995; Mann *et al.*, 1998). A prominent example is the 1815 eruption of Tambora in Indonesia that is typically blamed for the year without a summer. While parts of eastern North America and Europe experienced notable cooling, the observation that other regions, including the western US and the Middle East, appear, in fact, to have been warmer than usual is consistent with a hypothesized relationship between volcanic forcing of climate and the response of the North Atlantic Oscillation. The longer-term variations, and in particular cooler temperatures during the 17th century and warmer temperatures during the 18th century were likely to have been related to a concomitant increase in solar output by the Sun by approximately 0.25% following the Maunder Minimum of the 17th century (Lean *et al.*, 1995; Mann *et al.*, 1998) (*see* Maunder Minimum, Volume 1). Finally, changes in the ocean circulation (e.g., the Gulf Stream) of the North Atlantic, and associated impacts on North Atlantic storm tracks, may have emphasized temperature changes in Europe. The relative influences of these various external and internal factors on climate change during past centuries are an area of active climate research.

See also: Ground Temperature, Volume 1.

REFERENCES

Bradley, R S and Jones, P D (1993) Little Ice Age Summer Temperature Variations: their Nature and Relevance to Recent Global Warming Trends, *Holocene*, 3, 367–376.

Grove, J M (1988) *The Little Ice Age*, Methuen, London.

Jones, P D, Briffa, K R, Barnett, T P, and Tett, S F B (1998) High-resolution Palaeoclimatic Records for the Last Millennium: Interpretation, Integration and Comparison with General

Circulation Model Control Run Temperatures, *Holocene*, 8, 477–483.
Lean, J, Beer, J, and Bradley, R S (1995) Reconstruction of Solar Irradiance Since 1610: Implications for Climatic Change, *Geophys. Res. Lett.*, 22, 3195–3198.
Le Roy Ladurie, E (1971) *Times of Feast, Times of Famine, a History of Climate Since the Year 1000*, Doubleday, New York.
Mann, M E, Bradley, R S, and Hughes, M K (1998) Global-scale Temperature Patterns and Climate Forcing Over the Past Six Centuries, *Nature*, 392, 779–787.
Mann, M E, Bradley, R S, and Hughes, M K (1999) Northern Hemisphere Temperatures during the Past Millennium: Inferences, Uncertainties, and Limitations, *Geophys. Res. Lett.*, 26, 759–762.
Matthes, F (1939) Report of Committee on Glaciers, *Trans. Am. Geophys. Union*, 20, 518–535.
Pfister, C (1995) Monthly Temperature and Precipitation in Central Europe 1525–1979: Quantifying Documentary Evidence on Weather and its Effects, in *Climate Since A.D. 1500*, revised edition, eds R S Bradley and P D Jones, Routledge, London, 118–142.
Pfister, C (1998) Winter Air Temperature Variations in Western Europe during the Early and High Middle Ages (AD 750–1300), *Holocene*, 5, 535–552.

FURTHER READING

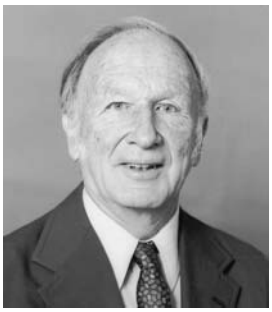
Ogilvie, A E J and Jonsson, T (2000) *The iceberg in the Mist: Northern Research in Pursuit of a "Little Ice Age"*, Proc. Symp., Climatic Change, 48, 1–263.

LOICZ (Land–Ocean Interactions in the Coastal Zone)

see LOICZ (Land–Ocean Interactions in the Coastal Zone) (**Volume 2**)

Lorenz, Edward N

(1917-)



Even if it is not common that a discovery in one field of science turns out to be of fundamental importance in several other fields, it does occur. And certainly it did occur when Edward Lorenz published his findings on the limitations of the predictability of the atmosphere, which in turn led to his pioneering work on deterministic chaos.

Before elaborating on the importance of these and other scientific contributions by Professor Lorenz we should identify some of the milestones in his life beginning with his birth May 23rd, 1917 in West Hartford, CT, USA, where he also grew up. His father was a mechanical engineer, and his mother did social work. Already as a child he demonstrated an unusual skill in solving mathematical puzzles, and as a young boy his favorite occupations were playing chess and collecting stamps.

After finishing school in 1934 he went to Dartmouth College, where he majored in mathematics, and in 1938 he entered the Mathematical Department of Harvard University. Here he was fortunate in having an eminent mathematician as advisor, Professor George Birkhoff, and he received his Master s degree in 1940.

Then came Pearl Harbor and the entrance of the USA into World War II, causing a radical turn of the Lorenz career. He applied for a meteorological course at Massachusetts Institute of Technology (MIT) and served in the US Army Air Corps 1942–1946 as weather forecaster in different places in the tropics, including Saipan and Okinawa.

After the war, Lorenz decided to continue his studies in meteorology. He returned to MIT, where he obtained his doctorate in 1948 and was offered a job by Professor Henry Houghton as a research scientist at its Department of Meteorology. Given his interest in mathematics and the experience he had achieved as a forecaster during the war, it was natural that he became actively engaged in research within the field of dynamic meteorology.

Thus, he became actively involved (together with Professor Victor Starr) in a project on the general circulation of the atmosphere. This research led to his introduction of the concept of available potential energy and clarification of how the work performed by the atmospheric engine is used to maintain the kinetic energy of the circulation against a continuous drain of energy by frictional dissipation (Lorenz, 1955).

In 1962, he was appointed Professor of Meteorology at MIT and at about that time he became deeply engaged in the problem of the predictability of weather. In a paper published in 1963, he was able to prove that there is a maximum time beyond which it is impossible to compute a useful forecast. After that time, any small error in the initial state used as input to the computations (for example caused by the flaps of a butterfly s wings) would have been amplified to such an extent that the value of the forecast would have become totally degraded (Lorenz, 1963).

It was in connection with this research that Lorenz s epoch-making work on deterministic chaos began. His papers focused on the behavior of physical systems that

seem to proceed according to chance even though their behavior is determined by precise physical laws. That is, the variations of such systems are not random, but only appear to be random. Although his thinking was stimulated by his familiarity with atmospheric processes, his ideas have since permeated virtually every branch of science (Lorenz, 1993).

Professor Lorenz was elected as a Member of the US National Academy of Sciences in 1975. He has also been elected as foreign member by academies of sciences in many other countries, and has received numerous honorary degrees and awards including the Kyoto Prize, the Roger Revelle Medal, the Holger and Anna-Greta Crafoord Prize, and most recently the IMO Prize of the WMO (AMS, 2001).

See also: Chaos and Predictability, Volume 1.

REFERENCES

AMS (2001) AMS Hosts Special Ceremony for Presentation of IMO Prize to Edward Norton Lorenz, *Bull. Am. Meteor. Soc.*, 82, 319–320.

Lorenz, E N (1955) Available Potential Energy and the Maintenance of the General Circulation, *Tellus*, 7, 157–167.

Lorenz, E N (1963) Deterministic Nonperiodic Flow, *J. Atmos. Sci.*, 20, 130–141.

Lorenz, E N (1993) *The Essence of Chaos*, University of Washington Press, Seattle, WA.

BO R DOOS *Austria*

Lovelock, James

(1919-)

James Lovelock is an inventor and independent scientist. In the 1950s Lovelock carried out pioneering work on the detection of trace gases. This work culminated in the invention of the electron capture detector, a device capable of detecting electron absorbing compounds in the gas phase at concentrations as low as 10^{-14} (1 part in 100 trillion). Such instrumental capability revolutionized environmental chemistry. For example, halogenated compounds are efficient at electron capture and some are important anthropogenic environmental contaminants, both as persistent toxins and in relation to the catalytic destruction of stratospheric ozone (*see* Depletion of Stratospheric Ozone, Volume 1). Other compounds that absorb electrons are products of microbial processes. The ability of Lovelock s detector to measure trace amounts of such compounds as methyl iodide and dimethyl sulfide opened up new vistas in knowledge of global biogeochemical cycles.

Lovelock s other great contribution to environmental science arose directly from his studies of trace gases in the atmosphere. In 1972 he put forward the revolutionary idea that the composition of Earth s atmosphere is a construct of the biosphere acting as a quasi-organismic system. He gave the system the name Gaia (*see* Gaia Hypothesis, Volume 5). Lovelock published three books on this stimulating, controversial way of looking at the interaction of biosphere, atmosphere, hydrosphere and lithosphere (*Gaia: A New Look at Life on Earth*, *The Ages of Gaia: A Biography of our Living Earth*, and *Healing Gaia: Practical Medicine for the Planet*).

Since the 1970s Lovelock has worked outside the institutional confines of academia, government and industry and this independence was integral to his personal, radical approach to doing science. He has been a Fellow of the Royal Society since 1974, President of the Marine Biological Association from 1986–1990, and in 1997 he was a recipient of the Blue Planet Prize.

Photo: Reproduced by permission of Bruno Comby Institute, www.comby.org. James Lovelock s website: www/ecolo.org/lovelock/; EFN s website: www.ecolo.org.

REFERENCE

Lovelock, J (2000) Homage to Gaia: The Life of an Independent Scientist, Oxford University Press, 396.

G R WILLIAMS *Canada*

M

Madden–Julian Oscillation

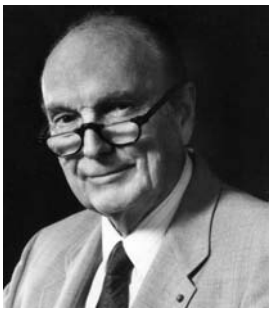
The *Madden-Julian Oscillation* (MJO) is the major mode of variability in the tropics on time scales of one to a few months. For this reason, it is often referred to as the 30–60 or 40–50 day oscillation. The MJO is particularly evident in the variation of tropical easterlies (the west to east winds) over the tropical Pacific and Indian Oceans. It is also evident in the variations of the sea surface temperature, cloudiness, and rainfall. The MJO is named after Roland Madden and Paul Julian, both at the National Center for Atmospheric Research in Boulder, CO, who in the early 1970s identified the oscillation when analyzing the fluctuating character of the winds over the tropical Pacific Ocean.

In that the major seasonal to interannual mode of climate variability in the same regions is the El Niño/Southern Oscillation, there is great interest in understanding the various interactions between these two oscillations. For example, there are some suggestions that the intensity of El Niño events might be influenced by the phase of the MJO when the El Niño begins (*see* El Niño and La Niña: Causes and Global Consequences, Volume 1).

MICHAEL C MACCRACKEN *USA*

Malone, Thomas F

(1917-)



Meteorologist and seasoned initiator and organizer of cooperative scientific activities, Thomas Malone has greatly influenced the linking of research and public policy relating to environmental understanding at the global, national, and local levels. Born in South Dakota, USA in 1917, and completing his ScD at Massachusetts Institute of Technology (MIT) in 1946, Malone served successively on the Faculty of MIT, 1941–1955; Research Director for the Travelers Insurance Company, 1965–1970; Dean at the University of Connecticut, 1970–1973; Director of the Holcomb Research Institute, Butler University, 1973–1983; Scholar in Residence, St. Joseph College, 1983–1991; and Director, Sigma Xi Center, Research Triangle, NC, 1992–1998.

At the international level, Malone was a prime mover in the creation of the International Geosphere-Biosphere Programme, as outlined in the book of which he was co-editor (Malone and Roederer, 1985). He was an organizer and first Secretary-General of the International Council for Science (ICSU) Scientific Committee on Problems of the Environment (SCOPE) (*see* SCOPE (Scienti c Committee on Problems of the Environment), Volume 4). In that capacity, he helped initiate studies on the design of global environmental monitoring systems and on the environmental effects of large dams. Later, he was a principal consultant to the World Conference on the Human Environment at Stockholm, and then an active participant at the Earth Summit at Rio in 1992. He was an influential member of the international committee of scientists that assessed the possible environmental effects of a nuclear war in 1985 and of the United Nations (UN) Task Force that outlined the basis for UN consideration of this issue. He also was a key figure in the early establishment of the Global Atmospheric Research Programme, having negotiated the cooperative agreement between ICSU and the World Meteorological Organization leading to worldwide experiments in 1978 that evolved into the World Climate Research Programme (*see* WCRP (World Climate Research Programme), Volume 1).

At the national level, Malone was an active member of the National Academy of Sciences, and steered its foreign activities during 1978–1982. He has chaired an array of National Research Council units that examined programs in atmospheric sciences, geophysics, environmental studies, and international scientific relations, including

global change (1962–1991). This same skill in dealing with a diversity of both scientific and organizational criteria has enabled him to play a role in organizing the National Center for Atmospheric Research at Boulder, CO. He also has devoted time to the application of scientific judgment to problems of motor vehicle safety and assessment of acid precipitation.

Although Malone has had a profound influence on national and international research and policy, he has provided continuing support for the constructive roles of local governments and of individual scientists in dealing with environmental problems. Thus, in the period 1988–1998 he served first as executive scientist for the Connecticut Academy of Science and Engineering on problems of sustainability, and then as director of the national Sigma Xi Center addressing current scientific issues.

To a large degree, the current organization for understanding and dealing with environmental issues reflects the vision and persistence that Malone has brought to finding constructive, solidly based lines of action (Malone, 1992).

REFERENCES

Malone, T F and Roederer, J G, eds (1985) *Global Change*, Cambridge University Press, Cambridge, MA.

Malone, T F (1992) Global Change–The Human and International Dimensions of Science: View of the Possible, *Interdisciplinary Sci. Rev.*, 17, 137–142.

GILBERT WHITE *USA*

Manabe, Syukuro

(1931-)

Dr Syukuro Manabe has been a world renowned pioneer and leader in mathematical modeling of human-caused climate change since the 1960s. He was a senior scientist at the Geophysical Fluid Dynamics Laboratory (GFDL), National Oceanic and Atmospheric Administration (NOAA) (1958–1997, now retired) and is currently a senior scientist at the Earth Frontier Research System, Japan (1997–present).

Manabe was born in Shingu-Mura, Uma-Gun, Japan on September 21, 1931. His father and other relatives were strongly centered within the medical profession. Manabe often humorously commented that, in his family s view, he had abandoned the highest calling, medical practice, for the lesser pursuit of climate science. The very presence of his biography in this Encyclopedia suggests that his medical relatives have undoubtedly reconsidered their original assessment by now.

Manabe received his PhD in 1958 from Tokyo University in Geophysics. Shortly thereafter, he joined Joseph Smagorinsky in the General Circulation Research Section of the US Weather Bureau, in Washington, DC, which soon was renamed the Geophysical Fluid Dynamics Laboratory (GFDL). GFDL is now located at Princeton University and is part of the US National Oceanic and Atmospheric Administration (NOAA).

Over his more than 40-year career, Manabe has been the world s leader in developing the field of climate modeling, its use in understanding the workings of the climate system, and in defining the quantification of human-caused climate change (greenhouse warming) (*see* General Circulation Models (GCMs), Volume 1). Recently, each of these research challenges has become a major international priority. Remarkably, the essence of the field today remains very similar to Manabe s pioneering vision.

Over his first decade of research (basically, the 1960s), he developed the world s first successful multilevel, 3D atmospheric circulation model in collaboration with Joseph Smagorinsky, and applied it to a pioneering determination of the role of the hydrological cycle in the climate system. He led the effort to produce the first model that included the stratosphere and the associated transport and chemistry of ozone. He and Kirk Bryan developed the world s first coupled atmosphere–ocean model. Such coupled models have now become indispensable for studying natural and human-caused climate changes. Most importantly, during that decade, he and Richard Wetherald were the first to develop a self-consistent simple 1D (altitude only) model of the global heat balance and apply it to assess the sensitivity of climate to changes in the chief radiatively important gases, carbon dioxide (CO_2), water vapor (H_2O), and ozone (O_3). In addition, Manabe and Wetherald presented the first assessment of the positive feedback effect of water vapor on CO_2 induced global warming. This work also highlighted the strong cooling of the stratosphere in response to increased atmospheric CO_2 concentrations. The greenhouse theory of climate change took its modern shape from these results, which were later supported by comprehensive three-dimensional models. Radiative-convective models similar to this one are still important analytical tools for interpreting the results from 3D models. It is not surprising that this work is still frequently quoted.

In his second decade of research during the 1970s, Manabe continued his systematic effort to model and understand the Earth s climate system. During the first half of this period, he published the first realistic simulations of the global hydrological cycle and the circulation of the tropics using the pioneering 3D atmospheric model he developed during the preceding decade. He continued his collaborative efforts with Kirk Bryan and Michael Cox to develop a model of the coupled atmosphere–ocean climate system, producing ever more realistic model simulations. It was during this second decade that he began his now world-renowned effort to understand the 3D, global sensitivity of the climate system to various perturbations such as added CO_2. This led to pioneering and still central studies on the effects of increasing CO_2 and changes in the solar output. It was this effort, more than any other, that led to today s enlightened understanding and appreciation of the greenhouse climate warming problem. Others have followed with major advances, but Manabe s research achievements still provide the foundation for the ongoing public debate on human caused climate change.

In many respects, the third decade of Manabe s climate research was even more noteworthy. His prodigious energy, diagnostic genius, and unique creativity pushed climate dynamics into new frontiers during the 1980s. He, Theodore Terpstra, and Douglas Hahn demonstrated the role of mountains in shaping the circulation of the atmosphere, most notably, that of the Indian summer monsoon. With Ronald Stouffer and Richard Wetherald, he continued his attack on the CO_2 climate warming problem by exploring regional and seasonal effects, as well as the first studies of the influence of changing clouds on climate sensitivity. His pioneering work that first projected a substantial drying of summer soil moisture in semi-arid continental regions in a greenhouse-warmed Earth is now the focus of major scientific, economic, and political attention. Also, in collaboration with Kirk Bryan and Anthony Broccoli, he attempted to simulate the warm CO_2-rich climate of the Cretaceous period and the cold climate of the last glacial maximum. These studies provided many new perspectives on the lessons of paleo-climate for today s climate change problems.

Manabe s fourth research decade in the 1990s, showed no loss of his research creativity and energy. Working with his colleagues, Thomas Delworth, Thomas Knutson, and Ronald Stouffer, Manabe and his team produced a number of insights on the roles of the ocean in climate change, including the possibility of multiple equilibrium states. During this period, he also pioneered the investigation of a possible slowdown or collapse of the overturning circulation of the North Atlantic caused by increasing greenhouse gases. The details of this phenomenon are still under intensive investigation today by a number of researchers. This period was especially marked by Manabe s pioneering attempt to simulate and understand decadal-scale natural climate variability, today a major international research priority. Referring to their successful simulation of low frequency variability of global mean surface temperature, Ronald Stouffer and Manabe concluded that the 20th century surface warming 0.6 °C cannot result solely from natural climate variations. They then concluded that the observed warming should, to a significant degree, be caused by the thermal forcing due to increased greenhouse gases in the atmosphere.

Manabe and his colleagues then took on the problem of time-dependent responses of the climate system to slowly increasing greenhouse gases. This work eventually produced a heightened awareness of the daunting character of the greenhouse warming problem by investigating the climate s response along its entire natural time scale of many hundreds of years. Even today, his original choice of investigating the problem out to doubled and quadrupled atmospheric CO_2 levels over the pre-industrial amounts still brackets the policy debates. Doubled CO_2 is essentially the commitment that society has made, and quadrupled CO_2 is where we are currently heading without focused mitigation of CO_2 emissions. More recently, in cooperation with the Intergovernmental Panel on Climate Change (IPCC) 1995 assessment process, scientists from the UK Hadley Centre and Manabe s team independently showed that the added greenhouse gases, offset somewhat by human-emitted sulfate aerosol particles, match the global-mean surface-air temperature warming over the past century rather well. Together with the study of natural climate variability mentioned above, these groups results played a key role in the formulation of the IPCC s now famous statement, The balance of evidence suggests a discernible human influence on global climate .

Now in his fifth decade of research, Manabe is at the Earth Frontier Research System in Japan. He is currently heavily involved in assisting the daunting Japanese effort to produce a very high resolution, comprehensive climate model using the world s most powerful supercomputer. The outcome of this effort is still unknown. However, if it succeeds, Manabe s uniquely perceptive insights and helpful advice will most likely have played a very substantial guiding role.

Manabe has achieved numerous national and international recognitions for his climate modeling research. These include: the Japan Meteorological Society s Fujiwara Award; the American Meteorological Society s Meisinger Award, Second Half Century Award, as well as the Carl-Gustaf Rossby Research Medal, Fellow, and Honorary Member; the American Geophysical Union s Roger Revelle Medal and Fellow; European Geophysical Society s Milutin Milankovich Medal; Member of the National Academy of Sciences; Foreign Member of Academia Europa; Foreign Fellow of the Royal Society of Canada; US Department

of Commerce Gold Medal; Presidential Meritorious Rank Award; Blue Planet Prize; Asahi Prize; and the Volvo Environmental Prize.

The world of government, commerce, and environment has now increasingly realized how vitally important it is to understand climate, predict its variations, and project its long-term changes. Without Syukuro Manabe s pioneering accomplishments and intellectual leadership, today s global awareness of climate issues might still be but a dim hope in the minds of a small number of scientists and visionaries.

See also: Projection of Future Changes in Climate, Volume 1.

JERRY D MAHLMAN *USA*

Maunder Minimum

Sunspots are relatively dark locations on the Sun s surface. The earliest observations of sunspots appear to have been made in China and Korea around 800 BC, and scattered records of sunspots exist over the centuries. Regular telescopic observations of the Sun began in the early 17th century, including records of sunspots. It was noted very early that the number of sunspots varied over about an 11-year cycle (*see* Sunspots, Volume 1). However, beginning in about 1645, the number of sunspots that were observed precipitously declined, and from 1645 to 1715 sunspots were rare events. This period is now known as the *Maunder Minimum*, after the solar astronomer E W Maunder (1890, 1894), who investigated the dearth of sunspots following earlier work by Sporer (1887). The reality and significance of the phenomenon were clearly established by exhaustive research by J Eddy (1976).

Reconstructions of solar activity based on various proxy data sources suggested that *solar irradiance* was significantly reduced during this period (*see* Solar Variability, Long-term, Volume 1). However, only in the last two decades have very accurate measurements been possible using satellites that can view the Sun from above the atmosphere. These measurements indicate that variations in solar radiation over the solar cycle vary by a few tenths of a percent. Roughly simultaneous to the Maunder Minimum, generally cooler climatic conditions prevailed in Europe. The exceptionally cold winters associated with these conditions led to suggestions that the world was experiencing a Little Ice Age, although newer evidence suggests that the severe cold was not a global phenomenon (*see* Little Ice Age, Volume 1). How the cooling in Europe, the winter climate of which is determined mainly by oceanic conditions in the Atlantic Ocean, may be tied to variations in solar output remains to be fully elaborated.

REFERENCES

Eddy, J A (1976) The Maunder Minimum, *Science*, 192, 1189.
Maunder, E W (1890) *MNRAS*, 50, 251.
Maunder, E W (1894) A Prolonged Sunspot Minimum, *Knowledge*, 17, 173.
Sporer, F W G (1887) *Vierteljahrsschr. Astron. Ges. Leipzig*, 22, 323.

JOHN S PERRY *USA*

Medieval Climatic Optimum

Michael E Mann
University of Virginia, Charlottesville, VA, USA

*The Medieval Climatic Optimum (also known as the Little Climatic Optimum, Medieval Warm Period, or Medieval Warm Epoch) refers to a period of climatic history during which temperatures in Europe and neighboring regions of the North Atlantic are believed to have been comparable to, or to have even exceeded, those of the late 20th century. This period is conventionally believed to have occurred from approximately 900- 1300 AD, terminating with the more moderate conditions of the 15th century, and the Little Ice Age (*see *Little Ice Age, Volume 1) which impacted Europe during the 16th- mid 19th centuries. The Medieval Climatic Optimum appears to have been in large part a feature of the North Atlantic and neighboring regions (Wigley* et al., *1981). Indeed, when Lamb (1965) coined the term Medieval Warm Epoch, it was based on evidence largely from Europe and parts of North America. Regional temperature patterns elsewhere over the globe show equivocal evidence of anomalous warmth (see Wigley* et al., *1981; Hughes and Diaz, 1994) and, as Lamb (1965) noted, episodes of both cooler as well as warmer conditions are likely to have punctuated this period.*

It is evident that Europe experienced, on the whole, relatively mild climate conditions during the earliest centuries of the second millennium (i.e., the early Medieval period). Agriculture was possible at higher latitudes (and higher elevations in the mountains) than is currently possible in many regions, and there are numerous anecdotal reports of especially bountiful harvests (e.g., documented yields of grain) throughout Europe during this interval of time.

Grapes were grown in England several hundred kilometers north of their current limits of growth, and subtropical flora such as fig trees and olive trees grew in regions of Europe (northern Italy and parts of Germany) well north of their current range. Geological evidence indicates that mountain glaciers throughout Europe retreated substantially at this time, relative to the glacial advances of later centuries (Grove and Switsur, 1994). A host of historical documentary proxy information such as records of frost dates, freezing of water bodies, duration of snowcover, and phenological evidence (e.g., the dates of flowering of plants) indicates that severe winters were less frequent and less extreme at times during the period from about 900–1300 AD in central Europe. Lamb (1965) (*see* Lamb, Hubert H, Volume 1) concluded that winters in Europe were less severe, and summers far drier, during the interval from AD 1080–1200. Farther south in the subtropical North Atlantic, there is also evidence for warmer sea surface temperatures during Medieval times (Keigwin, 1996).

Some of the most dramatic evidence for Medieval warmth has been argued to come from Iceland and Greenland (see Ogilvie, 1991). In Greenland, the Norse settlers, arriving around AD 1000, maintained a settlement, raising dairy cattle and sheep. Greenland existed, in effect, as a thriving European colony for several centuries. While a deteriorating climate and the onset of the Little Ice Age are broadly blamed for the demise of these settlements around AD 1400, the best evidence suggests that it was a combination of societal factors and trade relationships with mainland Europe. These in turn were probably influenced by a variety of seasonal climatic changes that were occurring throughout the North Atlantic region, rather than any simple local cooling trend (see McGovern, 1981, and also *see* Little Ice Age, Volume 1).

Although Lamb (1965) did not argue for a globally-synchronous warm period, his characterization has often been taken out of context, and used to argue for global-scale warmth during the early centuries of the millennium comparable to or greater than that of the latter 20th century. The best available evidence does not support such a notion. Outside of Europe and other regions neighboring the North Atlantic, the evidence for a Medieval Warm Period is indeterminate, at best (see Hughes and Diaz, 1994). Even those regions which appear to have experienced greater warmth exhibited it at quite different times. Indirect estimates of temperatures over the globe (based on proxy climate indicators such as tree rings, ice cores, and ocean sediments, and in certain regions, human documentary and phenological evidence – *see* Little Ice Age, Volume 1, Figure 2) provide an estimate of the considerable regional variations in timing of cold and warm periods around the globe during the Medieval period. Estimates of long-term changes in Central England temperatures (the basis, in large part, for the original definition of the Medieval Warm Period) suggest warmth during the period from about AD 1150–1350 (though the reliability of these estimates has been called into question – see Hughes and Diaz, 1994). In contrast, estimates of temperatures in western Greenland from ice cores (relevant to the earlier discussion of the Norse colonization of Greenland) suggest anomalous warmth locally only around AD 1000 (and to a lesser extent, around AD 1400), and in fact, quite cold temperatures during the latter part of the 11th century. The seasonality of this warmth (e.g., winter or summer) indicated by such proxy information is, however, not clear. Estimates of both sea surface temperatures in the subtropical North Atlantic from sediment cores and tree rings from Scandinavia and Eastern China imply unusually warm conditions only during the 11th and early 12th centuries. There is no evidence of unusual warmth in either tree-ring estimates of western North American temperatures or ice-core based estimates of temperatures in the tropical Andes of South America.

Figure 1 compares estimated temperature variations for the Northern Hemisphere as a whole (based on combined temperature information over the globe from indirect sources) with estimated temperature trends in Central England alone. Northern Hemisphere annual mean temperatures (Mann *et al.*, 1999) and extratropical summer temperatures (Jones *et al.*, 1998), suggest only slightly warmer temperatures (a couple of tenths of a °C) during the

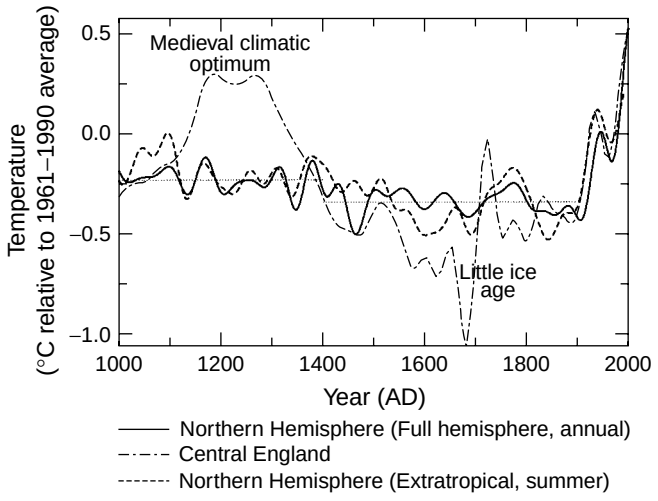


Figure 1 Estimated temperature variations during the past millennium for (1) the entire Northern Hemisphere estimated for the annual mean over the entire hemisphere (solid – Mann *et al.*, 1999) and over the extratropical region during the warm season (dashed – Jones *et al.*, 1998) based on global databases of proxy climate indicators, and (2) central England, based on a combination of thermometer, historical and proxy data records from central England (Lamb, 1965). Horizontal dashed lines indicate the moderately different Northern Hemisphere annual mean temperatures during the periods AD 1000–1400, and AD 1400–1900

period AD 1000–1400 relative to the later, colder period AD 1400–1900 (the latter associated with the Little Ice Age). Moreover, unlike European temperatures that indeed indicate a distinct warm phase earlier in the millennium, the large-scale trend represents a relatively monotonic long-term cooling. The less variable long-term fluctuations in temperature for the entire Northern Hemisphere result from the fact that the timings of cold and warm periods, so highly variable from region to region, tend to cancel in a hemispheric average. If one were to define hemispheric cold and warm periods during the past millennium by modern standards, only the 20th century could be termed a warm period; the period AD 1000–1400 would be termed a moderately warm period, and the period 1400–1900 a moderately cold period. Evidence for the Southern Hemisphere is far sketchier, and it is difficult as yet to reach any confident conclusions, although estimates of Southern Hemisphere temperatures (Jones *et al.*, 1998), uncertain as they are owing to the small amount of available information, show no evidence of a Medieval Climatic Optimum.

Thus, current evidence does not support the notion of a Medieval Climatic Optimum as an interval of hemispheric or global warmth comparable to the latter 20th century. Astronomical climate forcing may have contributed to a long-term cooling trend throughout the second millennium that terminated in the 20th century. Increased northward heat transport by an accelerated Atlantic thermohaline ocean circulation during Medieval times may have warmed the North Atlantic and neighboring regions, causing the warmest temperatures to be evident in Europe and lands neighboring the North Atlantic (albeit at notably varying times within the broader period of AD 900–1300).

A variety of factors thus may have contributed to both the moderate warmth of the Northern Hemisphere and the more sizeable and distinct North Atlantic/European warming during the early centuries of the second millennium.

REFERENCES

Grove, J M and Switsur, R (1994) Glacial Geological Evidence for the Medieval Warm Period, *Clim. Change*, 26, 143–169.

Hughes, M K and Diaz, H F (1994) Was there a Medieval Warm Period and if so, Where and When? *Clim. Change*, 26, 109–142.

Jones, P D, Briffa, K R, Barnett, T P, and Tett, S F B (1998) High-resolution Palaeoclimatic Records for the last Millennium: Interpretation, Integration and Comparison with General Circulation Model Control Run Temperatures, *Holocene*, 8, 477–483.

Keigwin, L (1996) The Little Ice Age and Medieval Warm Period in the Sargasso Sea, *Science*, 274, 1504–1508.

Lamb, H H (1965) The Early Medieval Warm Epoch and its Sequel, *Palaeogeogr., Palaeoclimatol., Palaeoecol.*, 1, 13–37.

Mann, M E, Bradley, R S, and Hughes, M K (1999) Northern Hemisphere Temperatures During the Past Millennium: Inferences, Uncertainties, and Limitations, *Geophys. Res. Lett.*, 26, 759–762.

McGovern, T H (1981) Economics of Extinction in Norse Greenland, in *Climate and History*, eds T M L Wigley, M J Ingram, and G Farmer, Cambridge University Press, Cambridge, 404–443.

Ogilvie, A E J (1991) Climatic Changes in Iceland, AD 865 to 1598, *Acta Archeol.*, 61, 233–251.

Wigley, T M L, Ingram, M J, and Farmer, G (1981) Past Climates and their Impact on Man: A Review, in *Climate and History*, eds T M L Wigley, M J Ingram, and G Farmer, Cambridge University Press, New York, 3–50.

Mesosphere

The *mesosphere* is the layer of the Earth s atmosphere from about 50 km to about 85 km that lies above the stratosphere and in which the temperature decreases with height. A peak in temperatures occurs at the boundary between the stratosphere and mesosphere, called the *stratopause*, the result of heating due to absorption of ultraviolet solar radiation by ozone. The temperature decreases with height throughout the mesosphere up to the *mesopause*, above which (in what is known as the *thermosphere*) the temperature begins increasing with height due to absorption of solar radiation by oxygen molecules. It should be noted that the *ionosphere*, which overlaps both the mesosphere and thermosphere, is defined in terms of the ion concentrations present (also created by the absorption of solar radiation), while the stratosphere, mesosphere, and thermosphere are defined by the vertical temperature structure of those layers.

See also: Ionosphere, Volume 1; Stratosphere, Volume 1.

KEITH L SEITTER *USA*

Mesozoic

The *Mesozoic* era was the middle era during the last 570 million years (Ma). It followed the Paleozoic, with the transition at 220 Ma before present (BP), and came before the Cenozoic, which started 65 Ma BP. The Mesozoic lasted 155 Ma (Matthews, 1984). The Mesozoic era contained three periods, the Triassic, Jurassic, and Cretaceous (*see* Cretaceous, Volume 1) in order from earliest to latest. The

Mesozoic is best known as the age of the dinosaurs. The Permo–Triassic and Cretaceous–Tertiary boundaries represent two of the most severe extinction events in Earth s history, each resulting in a dramatic change in the Earth s biota. The term Mesozoic was first coined by John Phillips in 1840, based on the fossils found in specific stratigraphic formations in Wales (Prothero, 1990). The paleogeography of the Mesozoic began as a single pole-to-pole supercontinent known as *Pangaea*. During the course of the next 155 Ma, North America separated from Africa and South America, Europe separated from Africa, Africa separated from India and Antarctica, and later from South America, and India and Australia separated from Antarctica, forming the equatorial Tethys Sea and later the Atlantic Ocean (Scotese and Golonka, 1993). The mid-late Cretaceous represents a time of high sea levels (300–500 m greater than present) and high atmospheric carbon dioxide (CO_2) concentrations due to sea floor spreading (resulting from increased volcanism) and was most likely a warm, ice-free period. The entire Mesozoic probably represents a relatively warm period in Earth s history (Crowley, 1991). Petroleum reserves in Saudi Arabia were formed during the Jurassic to mid-Cretaceous, while the uplift of the Rocky Mountains began during the Cretaceous (Matthews, 1984).

See also: Earth System History, Volume 1.

REFERENCES

Crowley, T J and North, G R (1991) *Paleoclimatology*, Oxford University Press, New York, 339.

Matthews, R K (1984) *Dynamic Stratigraphy: an Introduction to Sedimentation and Stratigraphy*, Prentice Hall, Englewood Cliffs, NJ, 489.

Prothero, D R (1990) *Interpreting the Stratigraphic Record*, W H Freeman, New York, 410.

Scotese, C R and Golonka, J (1993) *Paleogeographic Atlas*, Mobil Exploration and Production Services.

BENJAMIN S FELZER *USA*

Meteorology

Meteorology is the study of the atmosphere. The term has traditionally been used to refer to all aspects of the study of the atmosphere and atmospheric phenomena, but in recent decades has been more often applied to theory and applications associated with understanding and predicting the weather. Current usage treats meteorology as a subset of the broader atmospheric sciences.

KEITH L SEITTER *USA*

Methane

Michael C MacCracken

Lawrence Livermore National Laboratory, Livermore, CA, USA

Methane (CH_4) is the most abundant organic compound in the atmosphere. Although chemically very simple, methane is an unusual gas in that it is: (a) a source of energy when combusted; (b) able to absorb infrared energy and so is a contributor to the Earth's greenhouse effect; and (c) chemically decomposed in the atmosphere along chemical pathways that affect atmospheric chemistry and the cleansing capacity of the atmosphere. Methane is produced as a consequence of biological processes, being created when biological materials decompose in the absence (or near absence) of oxygen. Because of this, methane is most often generated underground or underwater or in the guts of animals. The methane that comes out of deep wells in the Earth was created when very old biological matter decomposed far underground. This methane is often associated with oil and coal-containing formations that derive from biological matter buried tens to hundreds of millions of years ago. Because methane is valuable as a source of energy when burned (known familiarly as natural gas), every effort is made to make sure that this methane is piped without leakage to combustion devices; this is largely successful and improvements are being made. Methane also comes from decomposition that takes place in the organic mass of accumulated materials in wetlands, below rice paddies, land lls, and in the guts of ruminant animals and termites.

Because methane has not been fully oxidized, its chemical bonds contain energy that can be released through chemical reactions in both the stratosphere and troposphere. In the stratosphere, methane affects the ozone concentration and is also the primary reactant in chlorine deactivation (Cl^- reacting to form hydrochloric acid (HCl)). After finally breaking down chemically, water is a major by-product, and the increasing methane concentration is leading to an increase in the stratospheric water vapor concentration. In the troposphere, the primary chemical destruction mechanism is initiated via the reaction with the hydroxyl radical (OH), which is a reactive molecule that helps break down a wide range of pollutant compounds that reach the atmosphere (*see* OH- Radical: is the Cleansing Capacity of the Atmosphere Changing?, Volume 2). As a result, methane affects the concentrations of OH, tropospheric ozone (O_3), and formaldehyde (COH_2). The present atmospheric lifetime of methane as a result of this and other minor reactions is estimated to

be about 8.4 years, which is the shortest lifetime of the various greenhouse gases whose concentration is being significantly affected by human activities. Increases in the methane concentration tend to reduce the OH concentration, thereby lengthening the lifetime of methane in the atmosphere.

Ice core records indicate that the concentration of methane has varied over the past 420 000 years, being about 400 parts per billion by volume (ppbv) during glacial peaks when wetland extent was likely at a minimum and about 700 ppbv during the peaks of interglacials when conditions were warmer and wetter. For the 10 000 years prior to the Industrial Revolution, the concentration was relatively steady at about 750 ppbv, indicating that natural emissions must have been roughly 200–250 teragrams of methane per year.

Since about 1750, the methane concentration has risen by about 150% and is now about 1750 ppbv. This amount is consistent with a human-induced augmentation of emissions to a level that is about two to three times the natural level. Because methane is a result of such basic human activities as growing rice and cattle, projections are that the emissions are likely to grow with world population unless control measures are taken. Observations do indicate that the increase in concentration is continuing, but there are year-to-year variations that are not fully understood and the rate of rise has been dropping somewhat, although it is not yet well established if this is due to reductions in human-induced emissions, reductions in natural emissions due to climate fluctuations, or changes in atmospheric chemistry.

On a per mass basis, at present methane and carbon dioxide (CO_2) concentrations (and assuming a 100-year time frame), methane is over 20 times as effective at trapping infrared radiation compared to carbon dioxide (*see* Global Warming Potential (GWP), Volume 1). For this reason, although the methane concentration is only about 1/200th the carbon dioxide concentration, the increase in the methane concentration has augmented radiative forcing by about one-third the increment caused by the increase in the carbon dioxide concentration. In addition, chemical model studies indicate that the OH concentration drops about 0.3% for each 1% increase in the methane concentration. Because of its relatively large effect, because it has a relatively short atmospheric lifetime, and because capture of methane is relatively inexpensive and when captured methane can be used as a fuel, significant attention is being devoted to controlling the emissions of methane that result from human activities. Control measures include capture at landfills and sewage treatment plants and altering agricultural activities, including the growing of rice and the feed provided to cattle.

An important potential climatic feedback involving methane is that global warming may lead to melting of the permafrost in high latitudes, and that this warming would accelerate the release of methane that is trapped in the ice below the land surface and in the sediments of the continental shelf. Release of this trapped methane (*see* Methane Clathrates, Volume 1) and the more rapid decomposition of biological materials would enhance the release of methane to the atmosphere, accelerating global warming.

Methane Clathrates

Nick Sundt
Office of the US Global Change Research Program, Washington DC, USA

Under high pressure, low temperatures and relatively quiescent conditions, very large amounts of gaseous methane (CH_4) have been trapped in frozen water within ocean sediments and on land in association with permafrost. The methane molecules are enclosed within tiny cages that are formed as water crystallizes around the gas molecules, thereby forming clathrates. The term clathrate refers to the enclosure of one kind of molecule within a crystal lattice formed by another compound.

In addition to the methane within clathrates themselves, additional methane may be trapped as clathrate layers, which block routes through which methane otherwise would freely escape. Various land and ocean drilling programs have found methane clathrates in permafrost and concentrated in oceanic sediments along the margins of continents.

Clathrates form when methane is introduced to a favorable combination of low temperatures, high pressure and water. Among major sources of methane are bacteria breaking down organic matter and geologic faults that allow older methane to rise from deeper below the Earth s surface. Over millennia, methane sources have fed the process in which clathrates trap large volumes of the gas that otherwise might have been vented to the atmosphere.

Estimates of the amount of methane locked up by clathrate formations vary widely across studies, although scientists agree that vast quantities are present. According to the US Department of Energy (Tomer, 2000), worldwide oceanic hydrate resources are estimated to range from 850 to ~1 400 000 trillion cubic meters (30 000–49 000 000 trillion cubic feet (Tcf)) and worldwide continental resources are estimated to range from 140

to 340 000 trillion cubic meters (5000–12 000 000 Tcf). By comparison, conventional worldwide gas resources are estimated to be about 370 trillion cubic meters (13 000 Tcf). Indeed, some estimates of methane clathrate resources exceed the combined global reserves of conventional gas, coal and petroleum.

Within the context of climate change, the clathrates are important for two reasons. First, methane clathrates hold very large amounts of energy. However, because the resource often is diffuse and difficult to access, it is not known what fraction might be economically and safely exploited within the foreseeable future. Indicative of the potential importance of methane clathrates, the US Department of Energy has calculated that even if only 1% of the methane hydrate resource (in the US, mainly in Alaska and on adjacent continental shelves) could be made technically and economically recoverable, the US could more than double its domestic natural gas resource base.

Combustion of such large volumes of methane would add carbon dioxide emissions to the atmosphere. However, if methane were burned instead of coal, the total emissions of carbon dioxide would be relatively lower because methane is characterized by lower carbon emissions per unit of energy produced. In addition to the carbon dioxide emissions, it would be very likely that unburned methane would be emitted directly to the atmosphere at various stages in the fuel cycle. This would be particularly important because methane is about 10–20 times more potent as a greenhouse gas than carbon dioxide, although its lifetime in the atmosphere is much shorter (*see* Global Warming Potential (GWP), Volume 1).

The second reason that methane clathrates are potentially important contributors to climate change is that, under the right environmental conditions (lower pressures or higher temperatures, for example), clathrates can relatively quickly release their captive methane to the atmosphere. If such disturbances were to occur on a large enough scale, sufficient methane could be released to the atmosphere to noticeably affect climate. Indeed, recent research suggests that enormous amounts of methane have been periodically vented from clathrates in the distant past. Norris and Rohl (1999) present evidence suggesting that, over a period of a few thousand years or less, a methane release of an estimated 1200–2000 billion tonnes of carbon (Gt C) occurred 55 million years (Myr). For comparison, the pre-industrial atmospheric burden of methane was about 1.5 Gt C and the current burden is about 3.5 Gt C, with current annual emissions of about 0.3 Gt C year^{-1} each occurring due to both human and natural activities (IPCC, 2001). Significant global ecological impacts followed the release 55 Myr ago and included widespread extinctions of species living on or near the seabed.

Norris and Rohl found evidence indicating that the initial release of methane was triggered by slope failure, i.e., submarine landslides. They further concluded that the initial pulse of methane helped warm the atmosphere, which then warmed the oceans, leading to further slope failures and melting of hydrates. They also suggest that subsequent releases might have continued this positive feedback process for about 30 000 years.

Other scientists have identified other possible occurrences of massive releases of methane clathrates, including some at about 120 Myr and about 90 Myr ago (reported in Kerr, 2000). The events were followed by dramatic and widespread ecological changes. For example, Kennett *et al.* (2000) report evidence from the Santa Barbara Basin off the coast of California of at least four episodes of brief, but large releases of methane from continental margin sediments, each of which was associated with millennial scale bottom water temperature changes. Their findings support mounting geologic evidence for past pervasive, massive methane releases from the marine sediment reservoir.

Findings that there have been massive methane releases in the past raise concerns that the human induced buildup of greenhouse gases in the atmosphere could trigger similar methane releases in the future. Of concern too is the possibility that large-scale commercial exploitation of methane hydrates could disturb the reservoirs enough to inadvertently prompt large and rapid methane releases.

The economic potential of the resource, however, has prompted significant interest by governments and industry in tapping methane clathrates for energy. In the US, the Methane Hydrate Research and Development Act was signed into law in May 2000, authorizing a five year, $47.5 million effort directed by the Department of Energy. The Japanese research and development effort is led by the Ministry of International Trade and Industry, and a drilling ship has been used to obtain samples and learn more about the resource.

For the future, balancing the interest in exploiting the energy resource with the potential for its effects on climate (which could be moderating, if methane reduces coal use, or amplifying, if substantial methane escapes to the atmosphere) will be a key challenge.

See also: Methane, Volume 1.

REFERENCES

IPCC (Intergovernmental Panel on Climate Change) (2001) *Climate Change 2000: The Scienti c Basis, Working Group I Report*, Cambridge University Press, Cambridge, 1–881.

Kennett, J P, Cannariato, K G, Hendy, I L, and Behl, R J (2000) Carbon Isotopic Evidence for Methane Hydrate Instability during Quaternary Interstadials, *Science*, 288, 128–133.

Kerr, R A (2000) Quakes Large and Small, Burps Big and Old, *Science*, 287, 576–577: Discussion of B N Opdyke, E Erba and R L Larson, *Hot LIPs, Methane, and the Carbon Record of the Apticore*, paper presented to Fall 1999 meeting of the

American Geophysical Union (13–17 December 1999, San Francisco, CA).

Norris, R and Rohl, U (1999) Carbon Cycling and Chronology of Climate Warming during the Palaeocene/Eocene Transition, *Nature*, 401, 775–778.

Tomer, B (2000) DOE National Hydrates Program Overview, *Gulf of Mexico Hydrates R & D Workshop Proceedings*, US Department of Energy, National Energy Technology Laboratory.

FURTHER READING

Blunier, T (2000) Frozen Methane Escapes from the Sea Floor, *Science*, 288, 68–69.

Hesselbo, S P, Grocke, D R, Jenkyns, H C, Bjerrum, C J, Farrimond, P, Morgans-Bell, H S, and Green, O R (2000) Massive Dissociation of Gas Hydrate during a Jurassic Oceanic Anoxic Event, *Nature*, 406, 392.

Methane, Industrial Sources

see Methane: Industrial Sources (**Volume 3**)

Methyl Bromide

James H Butler

National Oceanic and Atmospheric Administration, Boulder, CO, USA

Unlike most of the halogenated gases involved in stratospheric ozone depletion, atmospheric methyl bromide has both natural and anthropogenic sources. Also, its sinks include more than just atmospheric reactions - it is removed by the ocean, by soils, and possibly by plants. These differences from chloro uorocarbons (CFCs), chlorocarbons, and halons, along with the fact that the atmospheric budget of this gas is unresolved, have made it more dif cult to estimate accurately the contribution of anthropogenic methyl bromide to stratospheric ozone depletion. Considerable progress in understanding the behavior of this ozone-depleting gas has been made over the past decade, but signi cant uncertainties remain.

Methyl bromide (CH_3Br) is an atmospheric trace gas of both natural and anthropogenic origin. It is present in the atmosphere at a mole fraction, or volume-mixing ratio, of around 10 parts per trillion (ppt $= 10^{-12}$ moles of specific gas per mole of air). Its known sources include oceanic emission, biomass burning, agricultural application as a biocide, leaded gasoline combustion, and disinfestation of buildings and structures. Until the 1990s, little attention had been paid to this gas in the atmosphere, in part, because of its low mixing ratio and short atmospheric lifetime.

Methyl bromide is of concern because it is the primary carrier of bromine to the stratosphere and because reaction-rate enhancements due to bromine make it 50–60 times more effective than chlorine on a per atom basis in removing ozone from the stratosphere (Kurylo *et al.*, 1999). Thus, 10 ppt of bromine would be the equivalent of 500–600 ppt of chlorine. (The 1999 amount of organic chlorine in the atmosphere is about 3500 ppt, of which only 600–700 ppt occurs naturally.) Normally, a short atmospheric lifetime, such as that for CH_3Br, reduces the significance of the compound in depleting stratospheric ozone by maintaining a low mixing ratio in the atmosphere. This is reflected in the calculation of its ozone depletion potential (ODP) (*see* Ozone Depletion Potential (ODP), Volume 1), which has become an essential consideration in determining whether an anthropogenic, ozone-depleting compound should be scheduled for phase out. However, because CH_3Br contains bromine, its ODP remains high, even with its relatively short lifetime.

In the early 1990s atmospheric methyl bromide was thought to emanate naturally from a large oceanic source and to be destroyed exclusively by reactions in the atmosphere, predominantly with tropospheric OH (Albritton and Watson, 1992) (*see* OH- Radical: is the Cleansing Capacity of the Atmosphere Changing?, Volume 2). Anthropogenic emissions, mainly from disinfestation of soils, commodities, and structures, were considered responsible for 3 ppt of CH_3Br in the atmosphere. Biomass burning and emissions from burning leaded gasoline were thought to be possible contributors, but were not quantified at that time. Recognizing that there was a paucity of information on this important atmospheric gas, scientists began working to understand its cycling and atmospheric budget more completely. The results were surprising in a number of areas.

The first of these surprises was that the ocean was not the large source it was thought to be, but rather, a small net sink for atmospheric CH_3Br (e.g., Lobert *et al.*, 1995). This net sink, however, results from rapid aquatic production and degradation working in opposition in the surface ocean, leaving it largely undersaturated. In some areas, where production exceeds degradation, the ocean is supersaturated in methyl bromide, but most of the time, in most of the surface ocean, methyl bromide is undersaturated. Due to the fact that the degradation rate of CH_3Br is so high virtually everywhere in the surface waters, it had to be included as a significant component of the atmospheric lifetime computation. Subsequent calculations of

Table 1 The influence of newly identified plant sources on the atmospheric budget of CH_3Br[a]

Source type	Emissions (Gg year^{-1})	Sink type	Uptake (Gg year^{-1})
Oceans	56 (5–130)[b]	Oceans	77 (37–133)[b]
Disinfestation – soils	26.5 (16–48)	OH and hν	86 (65–107)
Disinfestation – durables[c]	6.6 (4.8–8.4)	Soils[d]	46.8 (32–154)
Disinfestation – perishables[c]	5.7 (5.4–6.0)	Plants[d,e]	
Disinfestation – structures	2 (2–2)		
Gasoline	5 (0–10)		
Biomass burning	20 (10–40)		
Wetlands[d]	4.6 (?)		
Salt marshes[d]	14 (7–29)		
Plants – rapeseed[d]	6.6 (4.8–8.4)		
Rice fields[d]	1.5 (0.5–2.5)		
Fungus[d]	1.7 (0.5–5.2)		
Total	150[f] (56–290)		210 (134–394)[f]

[a] Summary taken from Yvon-Lewis (1999), which lists specific references. All values are from Chapter 2 of the *1999 WMO Scientific Assessment of Ozone Depletion* (Kurylo *et al.*, 1999) and references therein except where noted. The numbers in parentheses represent the range of uncertainty for the best estimate shown.
[b] The ranges in the oceanic source and sink terms must keep the accepted range in the net flux of -3 to -32 Gg year^{-1}.
[c] The use of methyl bromide as a biocide includes treating agricultural products for export, including items such as timber (durables) and fresh fruits and vegetables (perishables).
[d] Items are from work that was published after the WMO report was finalized.
[e] Global estimate for green plants not yet available.
[f] Totals are rounded to the nearest integer.

the atmospheric lifetime of CH_3Br yielded a rate that was almost equal to the loss due to reaction with OH in the troposphere (Yvon-Lewis and Butler, 1997). This alone lowered the atmospheric lifetime of CH_3Br from 2.0 to 1.0 years.

At about the same time, studies of the terrestrial environment revealed additional sinks and sources of atmospheric CH_3Br (e.g., Shorter *et al.*, 1995). The discovery that CH_3Br was degraded rapidly in a variety of soils, mainly by prokaryotic bacteria, lowered the atmospheric lifetime even further. Studies of isolated plant leaves and stems from over 100 species of plants demonstrated that the biosphere was also involved in the degradation of methyl bromide (Jeffers and Wolfe, 1997). Whether this loss to plants turns out to be a significant sink or not depends upon further research. At this time, it appears to be small on a global basis. The few field studies of CH_3Br fluxes between plants and plant ecosystems and the atmosphere, however, reveal net emissions from the plants rather than net losses (Table 1). These are each small, but significant, in the global atmospheric budget of this gas.

Our current understanding of atmospheric CH_3Br is that of a gas with numerous, diverse sources and significant sinks on land, in the ocean, and in the atmosphere. Its lifetime, including atmospheric, oceanic, and soil sinks, is now computed at 0.7 years, but its calculated atmospheric budget is largely out of balance, with sinks outweighing sources by $\sim$40%. New findings continue to reveal previously unidentified sources, which seem gradually to begin closing the gap between calculated sources and sinks (Table 1). Anthropogenic emissions of CH_3Br are scheduled for phase-out by the year 2005 in developed countries and by 2015 in developing countries. However, the extent to which this will actually reduce the atmospheric burden of methyl bromide depends in part upon how the atmospheric budget is ultimately resolved. Finally, a question that will become more pressing with global change is: how will the fluxes of methyl bromide between the Earth s surface and atmosphere change in the future? A change in the sea-surface or soil temperature will certainly affect methyl bromide fluxes, as will changes in precipitation or land-use patterns. It is possible that such alterations of natural fluxes could outweigh today s direct anthropogenic emissions, but we cannot know until we understand the natural cycles of this important, ozone-depleting gas more fully.

See also: Stratosphere, Chemistry, Volume 1; Depletion of Stratospheric Ozone, Volume 1.

REFERENCES

Albritton, D L and Watson, R T (1992) Methyl Bromide: Its Atmospheric Science, Technology, and Economics – Montreal Protocol Assessment Supplement, United Nations Environment Programme, Nairobi, Kenya.

Jeffers, P M and Wolfe, N L (1997) Green Plants: A Terrestrial Sink for Atmospheric CH_3Br, *Geophys. Res. Lett.*, 25(1), 43–46.

Kurylo, M J, Rodriguez, J M, Andreae, M O, Atlas, E L, Blake, D R, Butler, J H, Lal, S, Lary, D J, Midgley, P M, Montzka,

S A, Novelli, P C, Reeves, C E, Simmonds, P G, Steele, L P, Sturges, W T, Weiss, R F, and Yokouchi, Y (1999) Short-lived Ozone-related Compounds, in *Scienti c Assessment of Ozone Depletion: 1998*, ed C A Ennis, World Meteorological Organization, Geneva, Switzerland.

Lobert, J M, Butler, J H, Montzka, S A, Geller, L S, Myers, R C, and Elkins, J W (1995) A Net Sink for Atmospheric CH_3Br in the East Pacific Ocean, *Science*, 267, 1002–1005.

Shorter, J H, Kolb, C E, Crill, P M, Kerwin, R A, Talbot, R W, Hines, M E, and Harriss, R C (1995) Rapid Degradation of Atmospheric Methyl Bromide in Soils, *Nature*, 377(6551), 717–719.

Yvon-Lewis, S A (1999) Methyl Bromide in the Atmosphere and Ocean, *IG Activities Newsletter*, December 1999.

Yvon-Lewis, S A and Butler, J H (1997) The Potential Effect of Oceanic Biological Degradation on the Lifetime of Atmospheric CH_3Br, *Geophys. Res. Lett.*, 24(10), 1227–1230.

Microwave Sounding Unit (MSU)

see MSU (Microwave Sounding Unit) (**Volume 1**)

Milankovitch, Milutin

(1879- 1958)

Milutin Milankovitch was a Serbian mathematician who specialized in astronomy and geophysics. He was born on 28th May 1879 in Dalj, Slavonia (part of Austro–Hungary until 1919; then part of Jugoslavia and now in Croatia). He died in Belgrade on 12th December 1958. He graduated as a doctor of technical sciences on 17th December 1904 from the Technical High School of Vienna. On the 1st October 1909, he was elected professor at the University of Belgrade where he lectured on Rational Mechanics, Theoretical Physics and Celestial Mechanics, until his retirement 46 years later. He was a member of the Serbian Academy of Sciences, Jugoslav Academy of Sciences and Arts, German Academy of Naturalists Leopoldine Halle and Italian Institute of Paleontology.

Milutin Milankovitch was a contemporary of Alfred Wegener (1880–1930) with whom he became acquainted through Vladimir Koppen (1846–1940), Wegener s father in law (*see* Wegener, Alfred, Volume 1). It is roughly between 1915 and 1940 that Milankovitch put the astronomical theory of the Pleistocene ice ages on a firm mathematical basis. His first book, written in French, dates from 1920, but his massive special publication of the Royal Serbian Academy of Sciences on *Kanon der Erdbestrahlung* was published in German in 1941 and was translated into English in 1969.

Milankovitch s main contribution was to explore the solar irradiance at different latitudes and seasons in great detail, to compute its long-term variations from the orbital parameters and to relate them in turn to the climate (*see* Orbital Variations, Volume 1). His theoretical investigation provided the basis for the core of his argument that under those astronomical conditions in which the heat budget around the summer solstice falls below average, so will summer melt, with uncompensated glacial advance being the result. The essential product of the Milankovitch theory is therefore his curve demonstrating how the intensity of summer sunlight varied over the past 600 000 years. This curve was used to identify the European ice ages reconstructed by Albrecht Penck and Eduard Bruckner, from which he concluded that these geological data constituted a verification of his theory.

Up to the 1960s, the Milankovitch theory was disputed as a result of discussions based on fragmentary geological records and because the climate was considered too resilient to react to such small changes as observed in the caloric summer insolation. Milankovitch was little disturbed by these different opinions, believing firmly that his theory was correct. In the late 1960s a systematic approach and use of modern techniques led to major discoveries, which progressively supported his essential concept, namely that orbital variations exert a significant influence on climate (Imbrie and Imbrie, 1979; Berger *et al.*, 1984; Berger, 1995).

Milutin Milankovitch wrote some 70 books and papers, including some on popular astronomy. He was involved in the reform of the Julian calendar, became even a writer and wrote his memoirs: not because I thought I was such an important person, but because I have lived in an historically interesting and turbulent period, and I described these events as a trustworthy witness. My work spanning some 30 years, has been closely connected with the work of other scientists who have used my results in their respective fields. The mutual collaboration has been documented with more than 600 letters and 100 publications. Therefore, these memoirs are, for a good part, the history of a branch of the sciences called Astronomical Theory of Climatic Changes . These memoirs were the basis of the unique

biography written in English on Milankovitch by his son (Milankovitch, 1995).

See also: Climate Model Simulations of the Geological Past, Volume 1.

REFERENCES

Berger, A, Imbrie, J, Hays, J, Kukla, G, and Saltzmann, B, eds (1984) *Milankovitch and Climate. Understanding the Response to Orbital Forcing*, D Reidel, Dordrecht, 1–895.

Berger, A (1995) Modeling the Response of the Climate System to Astronomical Forcing, in *Future Climates of the World, a Modelling Perspective*, Vol. 16, ed A Henderson-Sellers, World Survey of Climatology, ed H E Landsberg, Elsevier, Amsterdam, 21–69.

Imbrie, J and Imbrie, K P (1979) *Ice Ages, Solving the Mystery*, Enslow, New Jersey.

Milankovitch, M M (1920) *Théorie Mathématique des Phénomènes Thermiques Produits par la Radiation Solaire*, Académie Yougoslave des Sciences et des Arts de Zagreb, Gauthier-Villars.

Milankovitch, M (1941) *Kanon der Erdbastrahlung und seine Anwendung auf das Eiszeitenproblem* (Canon of Insolation and the Ice Age Problem, English Translation by Israel Program for Scientific Translation and published for the US Department of Commerce and the National Science Foundation, Washington DC), Royal Serbian Sciences, Special Publication 132, Section of Mathematical and Natural Sciences, Belgrade, Yugoslavia.

Milankovitch, M (1995) *Milutin Milankovic, from his Autobiography with Comments by his Son, Vasko, and a Preface by André Berger*, European Geophysical Society, Katlenburg, Lindau, Germany, 1–181.

FURTHER READING

Henderson-Sellers, A, ed (1995) *Future Climates of the World: A Modelling Perspective. World Survey of Climatology*, ed H E Landsberg, Elsevier, Amsterdam, 1–608, Vol. 16.

ANDRÉ BERGER *Belgium*

Model Simulations, Future

see Projection of Future Changes in Climate (*Opening essay*, **Volume 1**)

Model Simulations, Geological Past

see Climate Model Simulations of the Geological Past (**Volume 1**)

Model Simulations, Present and Historical Climates

see Model Simulations of Present and Historical Climates (*Opening essay*, **Volume 1**)

Modeling Regional Climate Change

Filippo Giorgi

Abdus Salam International Center for Theoretical Physics, Trieste, Italy

The determination of climatic changes at the regional scale is one of the central issues within the global change debate. Regional-scale information is needed to assess the impacts of climatic changes on the environment and society, and to determine suitable policies to react to such impacts. Despite the importance of this issue, the current uncertainties in the projection of regional climatic changes for the next several decades are still very high. This is due to the complexity of the processes that determine regional climate change and the need to develop more comprehensive modeling tools and research strategies to address this problem. This article discusses the basic processes that regulate regional climates and some of the modeling approaches developed to study regional climate change. The term climate change here refers to the climatic effects due to anthropogenic forcings for the coming decades (e.g., greenhouse gas and aerosol emissions). However, the modeling techniques discussed can be extended to any kind of regional climate study.

THE PROCESSES OF REGIONAL CLIMATE CHANGE

It is useful to provide an operational definition of regional scale, as different definitions are often implied in different research contexts. Here the term regional scale covers the relatively broad range of 10^3–10^7 km^2. The upper end of the range (10^6–10^7 km^2) is also commonly referred to as sub-continental scale, and phenomena occurring at broader scales are dominated by general circulation processes. The mid and lower end of the range (10^3–10^6 km^2) covers scales that are commonly referred to as mesoscale to synoptic scale, and are characteristic of the weather phenomena that affect our daily life. Scales finer than 10^3 km^2 are here referred to as local scale, while scales

coarser than 10^7 km^2 are referred to as large (or global) scale.

Following these definitions, the climate of a given region is determined by processes and interactions that occur at the global, regional and local scale. Large-scale processes and forcings regulate the general circulation of the atmosphere, which determines the sequence of weather events and regimes that characterize the climate of a region. Embedded within the large-scale circulation systems, regional and local forcings, along with mesoscale circulations, modulate the spatial and temporal structure of the regional climate signal. Examples of large-scale forcings that affect the global circulation are the solar radiation flux, the concentration of long-lived greenhouse gases, large scale snow and sea ice distribution, ocean–continent contrasts and continental scale topographical features. Examples of regional and local scale forcings are those due to complex topography, land use characteristics, inland bodies of water, complex coastlines, short-lived radioactive gases and aerosols, and regional snow and sea ice distributions.

In addition, because of the non-linear nature of atmospheric dynamics, processes occurring over given regions can influence the climate of distant regions through teleconnection patterns. This occurs, for example, in the El Niño Southern Oscillation (ENSO) and the North Atlantic Oscillation (NAO) phenomena, whereby regional anomalies in sea surface temperature and sea level pressure patterns can affect weather regimes over areas far from the source of the anomaly.

Finally, it is important to recognize that climate, both at the global and regional scale, is determined by atmospheric processes as well as complex interactions between different components of the climate system, i.e., the atmosphere, the hydrosphere, the cryosphere, the biosphere and the chemosphere. Processes within and among climate system components take place at a wide range of temporal scales and can be highly non-linear in nature.

The difficulty of modeling regional climate change is thus evident. The effects of climate system interactions, forcings and circulations at the global, regional and local scale need to be properly described, along with the teleconnection effects of regional forcing anomalies. This presents a formidable cross-disciplinary and multi-scaled scientific task that cannot be fully achieved by any individual modeling tool presently available, but requires the combined use of different modeling techniques.

MODELING REGIONAL CLIMATE CHANGE

Coupled Atmosphere–Ocean General Circulation Models (AOGCMs) are the most powerful tools available today to simulate long-term climate change. AOGCMs are three-dimensional numerical representations of the governing equations that describe the dynamics and physics of the coupled global atmosphere–ocean–sea ice system. State-of-the-art AOGCMs also include descriptions of land biosphere and atmospheric chemistry processes. For the simulation of climatic changes over centennial time scales, the model equations are numerically integrated over equivalent three-dimensional grids in which the distance between adjacent grid points determines the resolution at which processes can be explicitly represented. Processes occurring at sub-grid scales need to be expressed, or parameterized, in terms of quantities resolved by the model. The higher the model resolution, the more grid points are used in the numerical solution of the model equations, and the more computer power is necessary to integrate the model. Crudely speaking, an increase in horizontal resolution of a factor of two requires an eight-fold increase in computer power because, with increasing resolution, the integration time step must be reduced to preserve computational stability and accuracy. Therefore, computer resources are one of the factors that influence the resolution used in AOGCMs.

Despite the continuous development of increasingly powerful computing systems, the horizontal resolution of the atmospheric component of present-day AOGCMs is still relatively coarse, in the order of 300–500 km. This resolution is primarily due, in an indirect sense, to the long temporal scales involved in simulating the long-term changes in the ocean circulation, because of the very extensive computer resources that are then needed when integrating a coupled AOGCM for century time scales or longer (as is typically done in climate change experiments). Other factors limit the resolution of AOGCMs, such as the inclusion in the models of increasingly comprehensive descriptions of physical, chemical and biological processes and the effect that increasing resolution has on the AOGCM performance. While some aspects of model performance improve with increasing resolution, others deteriorate.

As a result, it can be expected that the standard atmospheric horizontal resolution of long-term AOGCM simulations may remain of the order of a few hundred km down to approximately 100 km for years to come. At these resolutions, it has been shown that AOGCMs can reproduce the fundamental characteristics of the general circulation of the atmosphere, and of large scale circulation features that affect regional climates, such as planetary wave patterns, storm tracks, jet streams and the Hadley circulation (e.g., IPCC, 1996, Chapter 5; *see* Model Simulations of Present and Historical Climates, Volume 1). AOGCMs have also shown encouraging performance in simulating important aspects of observed climate variability, such as related to ENSO or NAO (e.g., IPCC, 1996, Chapter 5).

On the other hand, resolution limitations prevent AOGCMs from capturing the fine scale signal that characterizes regional and local scale climate over many

regions of the world. As mentioned, this fine scale structure is caused by the effects of local and regional scale forcings which occur at sub-AOGCM grid scales and thus cannot be captured by AOGCMs. Lack of representation of these forcings can induce large errors in regional climate simulations by AOGCMs. An example can illustrate this point.

The Great Basin of the US, and especially its southern portion, is the driest region of the continental US and one of the driest in the globe. The Basin is enclosed by the Sierra Nevada range to the west and the Rocky Mountains to the east (Figure 1a). The dryness of the basin is caused by two primary factors. One is of large scale nature, i.e., the fact that the region lies to the south of the main Pacific storm track that impinges upon the northwestern US. The other is regional in nature, and is due to the precipitation shadowing effect of the Sierra Nevada mountains. As westerly moving systems cross the Sierra Nevada, air moving upslope cools down and condenses, generating clouds and precipitation in the upstream side of the range. Conversely, in the lee of the mountain range, downward moving air warms up, causing cloud evaporation and inhibiting precipitation over the southern Great Basin region (Figure 1b). As the maps indicate, the fine scale precipitation patterns observed over the western US are indeed caused by the complex topographic forcing over the region (Figure 1b).

One of the consequences of the coarse resolution of AOGCMs is that the surface topography seen by the model is highly smoothed. Over the continental US this results in merging the Rockies and the Sierras into one single mountain system centered just east of the Great Basin (Figure 2a). As a result, the southern Great Basin in a coarse AOGCM instead of lying in the lee of the Sierras and being shielded from precipitation, lies in the upslope of the model Rockies, thereby undergoing an upslope topographic enhancement of precipitation. Therefore, for the problem of simulating the climate of the Great Basin, while AOGCMs can capture the large scale aspects of it, e.g., the position

(a)

Figure 1 (a) Topography of the western United States; (b) average observed January precipitation over the western United States. Units are inches. (Reproduced from Giorgi *et al.*, 1992)

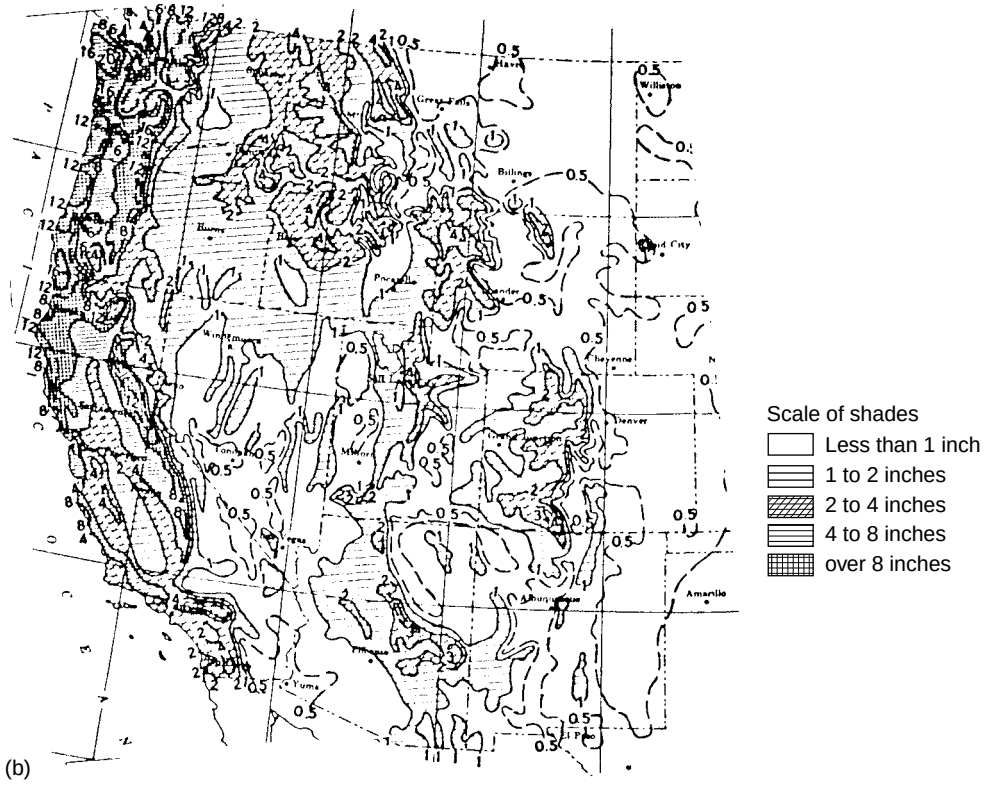


Figure 1 (*Continued*)

of the Pacific storm track, they entirely miss the aspects related to forcing by regional topographical features. With a topographic representation such as that of Figure 2(a), an AOGCM cannot capture the spatial detail of precipitation shown in Figure 1(b).

Various techniques have been developed over the last decade to enhance the regional information of AOGCMs, with the aim of describing at increased spatial resolution the effects of regional and local forcings. These are commonly referred to as regionalization techniques, and can be divided into three categories:

1. Use of high resolution and variable resolution atmospheric general circulation models (GCMs) in time slice experiments. With this technique, the atmospheric component of an AOGCM is run for pre-selected time periods, or slices (typically 10–30 years in length), with sea surface temperature forcing taken from the corresponding AOGCM simulation. Because only the atmospheric component is integrated for a limited period of time, a higher resolution can be reached (of the order of 100 km to date). In variable resolution models, the horizontal resolution is gradually increased over the region of interest (up to approximately 50 km to date).
2. Regional climate modeling. With this technique, a high resolution limited area model is used to simulate the climate of a region using large scale forcing meteorological fields from GCM simulations.
3. Statistical downscaling methods. Using this approach, statistical relationships are developed between large scale predictors and local/regional predictands using either observed or modeled data. These relationships are then applied to the output of AOGCM simulations to estimate changes in local and regional climate.

These techniques have different advantages and limitations that are discussed in a number of articles and references cited therein (Giorgi and Mearns 1991, 1999; Cubasch *et al.*, 1995; Deque *et al.*, 1995; IPCC, 1996, Chapter 6; Hewitson and Crane, 1996). It is beyond the purpose of this article to discuss all these techniques. Rather, for illustrative purposes we focus on one of them, i.e., the use of regional climate models, with the premise that many of the considerations presented also apply to the other regionalization techniques.

The Great Basin example given above has some historical significance. It is in order to simulate possible climatic changes over the Great Basin of the US that in the late 1980s a research group at the National Center for Atmospheric Research, in Boulder, Colorado, pioneered the use of limited area atmospheric models for climate applications (Dickinson *et al.*, 1989; Giorgi, 1990). When used

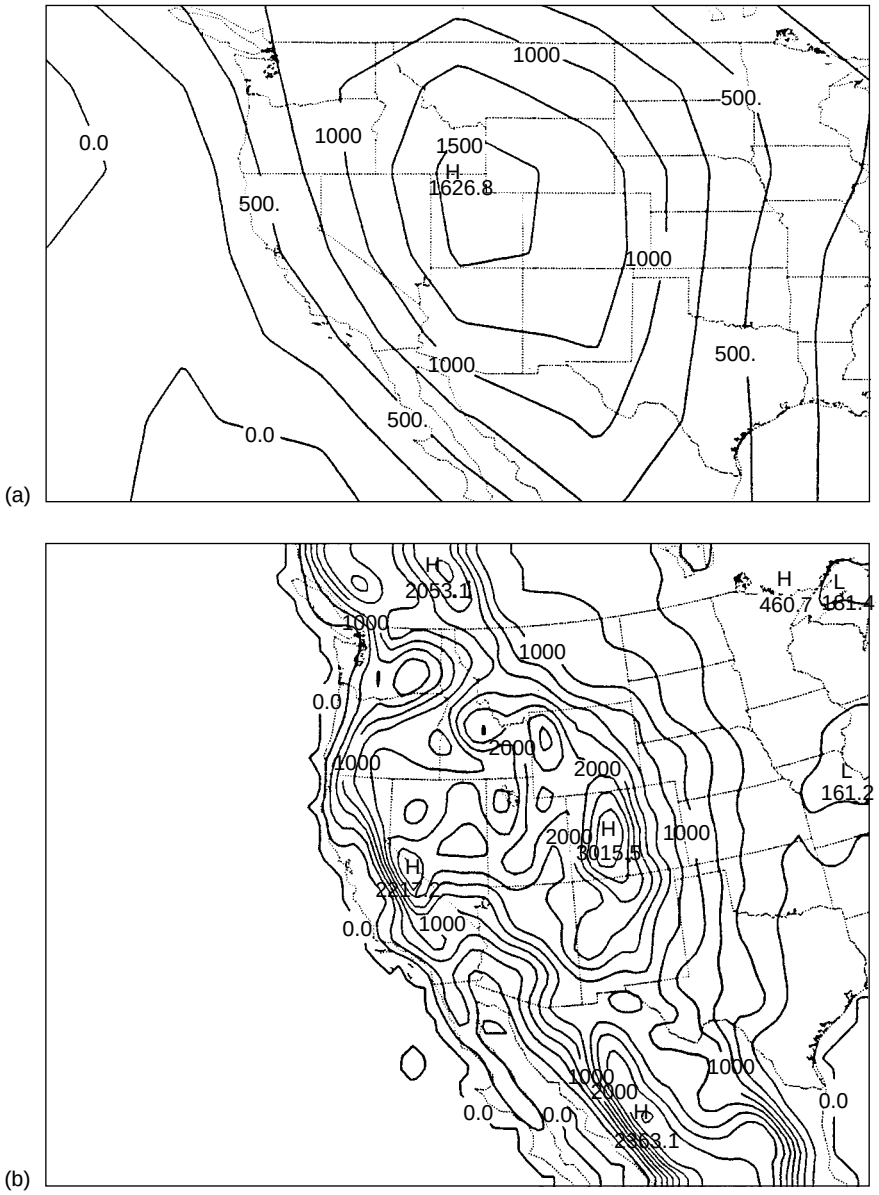


Figure 2 (a) Representation of the topography in western United States in a coarse resolution GCM (approximately 500 km grid size); (b) Representation of the topography in western United States in a RCM at 60 km grid size. Units are meters. (Reproduced from Giorgi *et al.,* 1992)

in climate studies, these models are usually referred to as regional climate models (RCMs). Essentially, RCMs are composed of the same set of equations as GCMs and describe the same physical processes, but they only cover limited area horizontal domains typically of 10^6 to $5 \times 10^7\,\mathrm{km}^2$ in size. As a consequence, for a given amount of computing power, an RCM can be run over a selected region with a much higher horizontal resolution than a GCM and can thus provide a more spatially detailed representation of the physical processes that affect regional climate. For the Great Basin example, Figure 2(b) shows the topographic representation of an RCM over the western US at a 60 km grid size. It is evident that the basic regional topographical features of the area are much better represented than in a coarse resolution GCM (Figure 2a).

RCMs require the provision of time-dependent meteorological conditions (wind components, temperature, water vapor and atmospheric pressure) at the lateral boundaries of the integration domain. In addition, they require the specification of sea surface temperature and initial meteorological and land surface conditions (e.g., soil moisture).

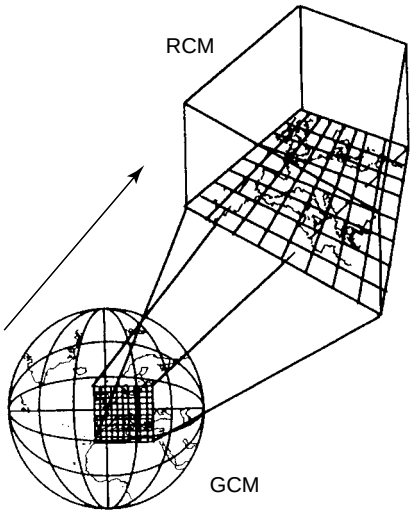


Figure 3 Conceptual representation of the one-way GCM–RCM nesting technique. See text for details. (Reproduced from Beniston, 1998)

These driving fields can, for example, be provided by observations or, for climate change studies, by the output of AOGCM (or more generally GCM) simulations. This technique is referred to as one-way nesting of an RCM within a GCM (Figure 3, Beniston, 1998). The term one-way derives from the fact that, while information from the driving GCM is supplied to the nested RCM, the feedback from the RCM fields to the GCM is not represented. The underlying strategy of this approach is that the GCM can provide the response of the general circulation to large scale forcings along with the teleconnection response to broad regional forcing anomalies. Having as input this information, the nested RCM can describe the effects of local and regional forcings on regional climate in a physically based way.

Limited area models have been used for many years in numerical weather prediction, and in the study of atmospheric phenomena such as extratropical cyclones, mesoscale convective complexes, tropical storms, sea breeze circulations. As a result, several advanced and widely tested limited area modeling systems are available as the end product of a multi-decadal experience in mesoscale modeling. These models can provide robust frameworks for the development of RCMs.

The relative simplicity of the one-way nesting methodology constitutes an important aspect of nested modeling. The method most commonly used to provide lateral boundary conditions to RCMs (the relaxation method) consists of smoothly blending the RCM model solution towards the forcing fields in a lateral buffer area of the domain. With several variants available, this method has proven to be effective in providing good consistency between forcing and simulated large scale fields, and in achieving little generation of boundary numerical noise.

Because RCMs only cover limited areas, they can reach relatively high horizontal resolutions. To date, the majority of RCM experiments utilize grid sizes of the order of 50 km. However, some experiments have been conducted at grid sizes of 10–20 km and even higher. When running an RCM at very high resolution it may be necessary to first run the RCM at an intermediate resolution and then use fields from this experiment to drive the very high resolution experiments. This multiple nesting prevents an excessively abrupt resolution transition from the GCM to the RCM.

It is important to understand the fundamental limitations of the nested RCM technique (Giorgi and Mearns, 1991, 1999). One is the lack of two-way feedbacks from regional scale circulations produced by the RCM to the large scale circulations. This limitation can be more or less important depending on the region under consideration. Another limitation is that the nested RCM cannot correct large errors present in the forcing GCM fields. Most RCM work has shown that the basic features of large scale circulations in an RCM are essentially dominated by the GCM forcing that, originating from the lateral boundaries, pervades the interior of the RCM domain reaching a dynamical equilibrium with the internal model solution. The role of the RCM is essentially to modulate the GCM signal by adding fine scale information generated by the internal high resolution model physical and dynamical forcings.

As an example, if the GCM has large errors in simulating the storm tracks that affect a region, the nested RCM will not be able to compensate for that error. It is thus of foremost importance that the driving GCM fields be of good quality, and indications are that the general circulation features simulated by the latest generation of AOGCMs are considerably improved compared to earlier ones. Various model intercomparison projects, the largest one being the Project to Intercompare Regional Climate Simulations (Takle *et al.*, 1999), have been recently organized to elucidate strengths and weaknesses of RCMs.

EXAMPLES OF REGIONAL CLIMATE CHANGE SIMULATIONS

A number of steps need to be taken before completing a nested RCM simulation of regional climate change. The first is to analyze the RCM behavior over the region of interest using observations as driving fields to simulate actual periods. The model results can then be compared with climate observations for the period of simulation. Fields from global analyses of observations produced, for example, by the European Centre for Medium-range Weather Forecasts or by the National Center for

Environmental Prediction can be used for this purpose. Experiments of this type are referred to as perfect boundary condition experiments, because the quality of the driving fields, although still not error-free because of deficiencies in the observing networks and analysis models, is the best available.

These experiments serve different purposes. One is to evaluate and optimize the model performance. In fact, systematic model errors revealed by perfect boundary condition experiments can be attributed primarily to deficiencies in aspects of the internal model physics and dynamics, so that these aspects can be suitably improved. Studies have shown that, when driven by good quality boundary conditions, RCMs can reproduce the observed structure of surface regional climate variables such as temperature and precipitation in terms of their average climatology, intra- and inter-seasonal characteristics, and inter-annual variability (Giorgi and Mearns, 1999).

Another primary purpose of these experiments is to select the most appropriate model domain and horizontal resolution. The choice of domain and resolution is an important and not trivial issue, as it depends on both physical and computational considerations. The model domain should be large enough to allow the RCM to develop its internal mesoscale circulations and to include processes and forcings that are important for the climate of the region. Similarly, the resolution should be sufficiently fine to appropriately describe such forcings and processes. On

Figure 4 Cold season precipitation over Great Britain. (a) Present day climate simulation with a driving GCM; (b) present day climate simulation with a nested RCM; (c) observations from the Climatic Research Unit (CRU) of East Anglia University. Units are mm day^{-1}. (Reproduced from Jones *et al.*, 1995)

the other hand, the computational requirements to run an RCM rapidly increase with domain size and model resolution, so that ultimately a compromise needs to be reached between physical and computational constraints. Once the model is fully tested and configured using perfect boundary condition experiments, the RCM can be run with driving fields from GCM simulations of present day climate and can be validated against climatological observations. It is important to identify systematic biases in the simulated present day climate by the nested GCM–RCM system because these are fundamental for a correct interpretation of the climate change simulations. GCM-driven simulations of present day climate are also important because they help to establish the added value of using RCM nesting. In GCM-driven experiments, the simulation period should be of sufficient length to produce robust climate statistics.

In general, the RCM performance in reproducing present-day climate in GCM-driven experiments tends to deteriorate compared to that of experiments driven by

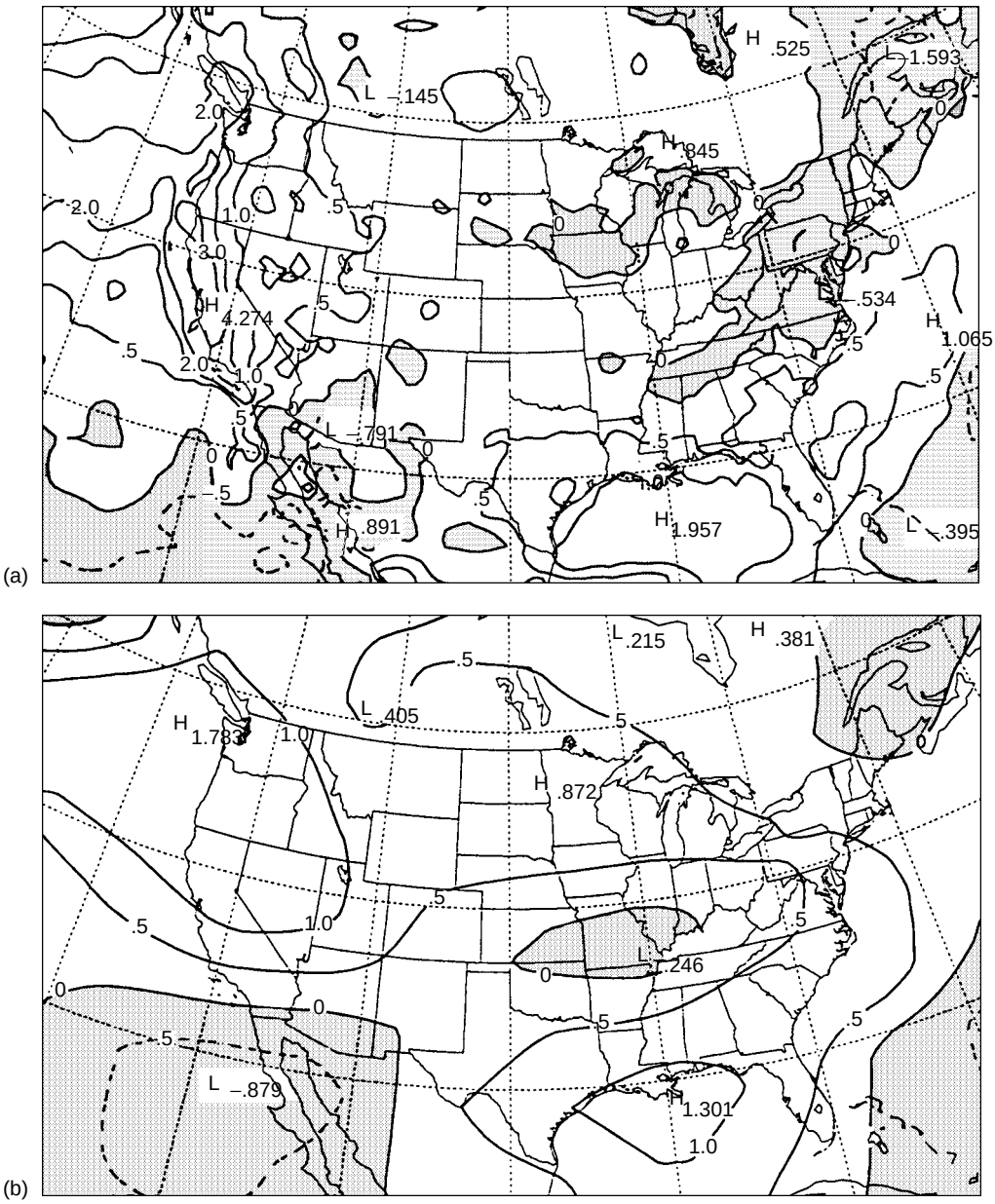


Figure 5 Cold season precipitation difference between climate conditions at doubling of greenhouse gas concentration and present day climate conditions over the continental United States as simulated with a nested RCM (a) and the driving GCM (b). Units are mm day^{-1}, and the contour intervals are at −3, −2, −1, −0.5, 0, 0.5, 1, 2, 3, mm day^{-1}. Shaded areas denote negative values. (Reproduced from Giorgi *et al.*, 1994)

observations, because errors in the large scale GCM forcing fields are transmitted to the RCM through the lateral boundary conditions. Compared to the driving GCMs, RCMs have been shown to enhance the simulation of spatial patterns of surface climate as forced by topography and other sub-GCM scale processes (IPCC, 1996, Chapter 6; Giorgi and Mearns, 1999). The increased resolution of RCMs also allows the simulation of a broader and more realistic spectrum of weather events, in particular as concerning higher order climate statistics such as daily precipitation intensity distributions and extreme events.

After systematic biases in the nested GCM–RCM modeling system are identified, we are ready to perform simulations for climate change conditions. In some circumstances it may be possible, and thus advisable, to further test the model performance by carrying out simulations for known climate conditions different from present, such as palaeoclimates. This can be an important test of the model, because future climate simulations are not directly verifiable.

To date, a wide range of regional climate simulations have been carried out (see for example Giorgi and Mearns, 1991, 1999; IPCC, 1996, Chapter 6; McGregor, 1997). Early experiments, in which the technique was still in its development stage, were typically of a few years in length, i.e., relatively short compared to the interdecadal time scales characteristic of the climate change issue. Nevertheless, these simulations allowed the development of the technique, and an understanding of first order effects. It is important to realize that a fundamental methodological step was taken when extending limited area model simulations beyond the several days length of weather prediction application. Several recent RCM simulations have reached lengths of 20–30 years, and a full 140-year RCM simulation of transient climate change has been recently completed.

A few examples are useful to illustrate the functioning of the nested modeling technique. Figure 4 (Jones *et al.*, 1995) shows a simulation of present day wintertime precipitation over Great Britain by a nested RCM and the driving GCM. Results are compared with high resolution observations. A substantial enhancement in the simulation of spatial precipitation detail by the RCM is evident. Figure 5 (Giorgi *et al.*, 1994) shows a simulation of wintertime precipitation change over the continental US in conditions of CO_2 doubling by a nested RCM and the corresponding driving GCM. It can be seen that the broad scale patterns of precipitation change in the two models are similar, for example, the location of broad regions of positive and negative change. This is because these broad scale signals are due to general circulation processes (e.g., shifts in the storm tracks) and are essentially regulated by the driving GCM. The fine scale structure of the signal, however, can be quite different in the two models. For the western US, which is characterized by complex topographical features, much of the precipitation increase simulated by the RCM is concentrated over the western coastal regions, while the GCM-simulated increase spreads much farther inland. This is because of the precipitation shadowing effect of the western US coastal ranges, which is better represented in the high resolution RCM than the low resolution GCM.

FUTURE PERSPECTIVES IN RCM RESEARCH

The use of RCMs (as well as other regionalization techniques) has filled an important gap in the spatial-temporal scale of climate modeling, i.e., that of fine spatial scales in conjunction with long temporal scales. This has opened the opportunity to a wide range of innovative applications, of which the simulation of regional climate change is only one.

Since the late 1980s, many RCM systems have been developed, employing a wide variety of model dynamical and physical frameworks, numerical techniques and methods of nesting. RCM applications have essentially covered all regions of the globe, and studies have ranged from paleoclimate and present day climate processes to multi-decadal future climate simulations. The use of RCMs for seasonal prediction has recently received considerable attention and will likely represent an important area of application of regional models. Multiple nesting has already allowed RCM resolutions of a few 10s of km, and it can be expected that, as computing power continues to increase, regional multi-decadal simulations at resolutions of less than 10 km will become feasible in the near future, i.e., that regional climate simulations reaching the local scale will be possible. Work on two-way coupling between GCMs and RCMs is also underway to examine the importance of representing regional-to-global feedbacks, although these feedbacks can be alternatively represented through the use of variable resolution AGCMs.

One of the research directions that offers the most promising and exciting perspectives of innovative research is the coupling of RCMs with other components of the climate system to generate regional climate system models (Figure 6, Giorgi, 1995), and research efforts in this direction have been initiated recently. In particular, regional coupled models can be used to study climate system interactions and feedbacks with the advantage that the coupling is accomplished at scales that are most consistent with relevant processes and available models (Giorgi, 1995). Finally, regional modeling can also be very useful for global modeling in different ways: by providing information on interactions at scales that global models will presumably reach in the future, and by helping

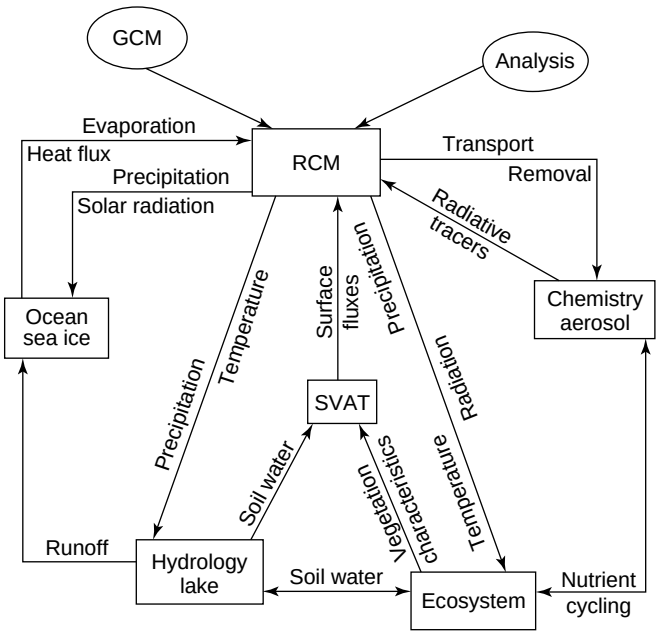


Figure 6 Conceptual design of a regional climate system model. A SVAT, or Surface–Vegetation Atmosphere Transfer scheme, is a model of land surface processes that can act as interface between different components of the climate system. (Reproduced from Giorgi, 1995)

in developing improved representations of dynamical and physical processes.

See also: Projection of Future Changes in Climate, Volume 1.

REFERENCES

Beniston, M (1998) *From Turbulence to Climate: Numerical Investigations of the Atmosphere with a Hierarchy of Models*, Springer-Verlag, Berlin, 1–328.

Cubasch, U, Waszkewitz, J, Hegerl, G, and Perlwitz, J (1995) Regional Climate Changes as Simulated in Time-Slice Experiments, *Clim. Change*, 31, 273–304.

Deque, M and Piedelievre, J Ph (1995) High Resolution Climate Simulation over Europe, *Clim. Dyn.*, 11, 321–340.

Dickinson, R E, Errico, R M, Giorgi, F, and Bates, G T (1989) A Regional Climate Model for the Western United States, *Clim. Change*, 15, 383–422.

Giorgi, F (1990) On the Simulation of Regional Climate using a Limited Area Model Nested in a General Circulation Model, *J. Clim.*, 3, 941–963.

Giorgi, F (1995) Perspectives for Regional Earth System Modeling, *Global Planet. Change*, 10, 23–42.

Giorgi, F and Mearns, L O (1991) Approaches to Regional Climate Change Simulation: A review, *Rev. Geophys.*, 29, 191–216.

Giorgi, F and Mearns, L O (1999) Regional Climate Modeling Revisited. An Introduction to the Special Issue, *J. Geophys. Res.*, 104, 6335–6352.

Giorgi, F, Bates, G T, and Nieman, S (1992) Simulation of the Arid Climate of the Southern Great Basin using a Regional Climate Model, *Bull. Am. Meteorol. Soc.*, 73, 1807–1822.

Giorgi, F, Shields Brodeur, C, and Bates, G T (1994) Regional Climate Change Scenarios Over the United States Produced with a Nested Regional Climate Model, *J. Clim.*, 7, 375–399.

Hewitson, B C and Crane, R G (1996) Climate Downscaling: Techniques and Application, *Clim. Res.*, 7, 85–95.

IPCC (1996) *Climate Change 1995, The Science of Climate Change, Second Assessment Report of the Intergovernmental Panel on Climate Change*, eds J T Houghton, L G Meira Filho, B A Callander, N Harris, A Kattenberg, and K Maskell, Cambridge University Press, Cambridge, 1–572.

Jones, R G, Murphy, J M, and Noguer, M (1995) Simulations of Climate Change over Europe using a Nested Regional Climate Model. I: Assessment of Control Climate, including Sensitivity to Location of Lateral Boundaries, *Q. J. R. Meteorol. Soc.*, 121, 1413–1449.

McGregor, J L (1997) Regional Climate Modeling, *Meteorol. Atmos. Phys.*, 63, 105–117.

Takle, E S, Gutowski, W J, Arritt, R W, Pan, Z, Anderson, C J, da Silva, R R, Caya, D, Chen, S-C, Giorgi, F, Christensen, J H, Hong, S-Y, Juang, H-M H, Katzfey, J, Lapenta, W M, Laprise, R, Liston, G E, Lopez, P, McGregor, J, Pielke, R A, and Roads, J O (1999) Project to Intercompare Regional Climate Simulations (PIRCS): Description and Initial Results, *J. Geophys. Res.*, 104, 19 443–19 461.

Models, Earth System

see Models of the Earth System (*Opening essay*, **Volume 1**)

Models, Energy Balance

see Energy Balance Climate Models (**Volume 1**)

Models, General Circulation

see General Circulation Models (GCMs) (**Volume 1**)

Models, Radiative–Convective

see Models of the Earth System (*Opening essay*, **Volume 1**)

Molina, Mario J

(1943-)

Mario Molina is an atmospheric chemist and professor at the Massachusetts Institute of Technology (MIT) in Cambridge, MA. In 1995, he shared the Nobel Prize in Chemistry with Sherwood Rowland (*see* Rowland, F Sherwood, Volume 1) and Paul Crutzen (*see* Crutzen, Paul J, Volume 1) for their contributions to understanding the effects of chlorofluorocarbons (CFCs) (*see* Chloro uorocarbons (CFCs), Volume 1) on the chemistry of the stratosphere and the factors contributing to stratospheric ozone depletion (*see* Depletion of Stratospheric Ozone, Volume 1).

Professor Molina was born in Mexico City on March 19, 1943 to Roberto Molina Pasquel and Leonor Henriquez de Molina. His father was a lawyer who also taught at the National University of Mexico (Universidad Nacional Autonoma De Mexico, UNAM) and later served as Mexican Ambassador to Ethiopia, Australia, and the Philippines. Molina attended elementary school and high school in Mexico City, where he was fascinated by science, particularly chemistry. In 1960, following graduation from high school, he enrolled in the chemical engineering program at UNAM, with plans to become a physical chemist. After graduating in 1965, Molina enrolled at the University of Freiburg in Germany and received a postgraduate degree in 1967, having done research in the kinetics of polymerizations. He then returned to Mexico as an Assistant Professor at the UNAM and set up its first graduate program in chemical engineering. In 1968, he enrolled in the graduate program at the University of California (UC) at Berkeley to pursue studies in physical chemistry, using chemical lasers to investigate the distribution of internal energy in the products of chemical and photochemical reactions. Molina received his PhD in physical chemistry in 1972. While at Berkeley he also met Luisa Tan, who was a fellow graduate student and who later became his wife and scientific collaborator.

In the fall of 1973, Molina made a move that would set the stage for the development of his career in atmospheric chemistry. Leaving the UC Berkeley, he became a postdoctoral fellow in the group of Professor Sherwood Rowland at the UC Irvine. Offered a number of research options, Molina chose the task of finding out the environmental fate of certain very inert industrial chemicals – the CFCs – that had been accumulating in the atmosphere. CFCs had been introduced as refrigerants early in the 20th century, being safer and more efficient than earlier refrigerants, namely ammonia and sulfur dioxide. At that time, CFCs were thought to have no significant effects on the environment. Within three months of arriving at UC Irvine, however, Rowland and Molina had developed the CFC–ozone depletion theory, which explains how chlorine (Cl) atoms produced by the decomposition of the CFCs catalytically destroy ozone. The importance of this destruction mechanism became clear when they compared the emissions of CFCs to the amounts of nitrogen oxides that, based on studies a few years earlier by Paul Crutzen of the natural factors influencing stratospheric chemistry, were controlling ozone levels. With the possibility that the continued release of CFCs into the atmosphere would cause a significant depletion of the Earth s stratospheric ozone layer, which protects life at the surface from the highly energetic ultraviolet radiation from the Sun, Rowland and Molina contacted Professor Harold Johnston at UC Berkeley, who was a pioneer in studying the impact of the release of nitrogen oxides from the proposed supersonic transport aircraft on the stratospheric ozone layer. Johnston informed them that atmospheric chemists Ralph Cicerone and Richard Stolarski had arrived at similar conclusions concerning the catalytic properties of Cl atoms in the stratosphere, in connection with the release of hydrogen chloride, either from volcanic eruptions or from the ammonium perchlorate fuel planned for the space shuttle. The results of these studies were published in *Nature* in 1974 (Molina and Rowland, 1974), setting in motion a chain of events that led first to the curtailment of the use of CFCs in aerosol spray cans and as solvents, and later to radical reductions in global production of CFCs under the Montreal Protocol on the Protection of the Ozone Layer (*see* Ozone Layer: Vienna Convention and the Montreal Protocol, Volume 4). Even though the potential influence was realized in the 1970s, studies indicate it will take until the middle of the 21st century for CFC-induced Cl levels in the stratosphere to drop to levels that do not create ozone holes in the polar regions of the stratospheric ozone layer.

In 1975, Molina was appointed to the faculty of the UC Irvine, where he studied the spectroscopic properties of several unstable and difficult to handle compounds of atmospheric importance, including hypochlorous acid, chlorine nitrite, chlorine nitrate, peroxynitric acid, etc. In 1982, he joined the Molecular Physics and Chemistry Section at the Jet Propulsion Laboratory in Pasadena, CA, where he led a group conducting measurements and developing techniques for the study of newly emerging problems, particularly relating to the peculiar chemistry that is promoted by polar stratospheric clouds, some of which consist of ice crystals. They were able to show that Cl-activation reactions take place very efficiently in the presence of ice under polar stratospheric conditions. With his wife, they carried

out experiments with chlorine peroxide, a new compound that turned out to be important in providing the explanation for the rapid loss of ozone in the polar stratosphere. In 1989, Molina accepted a joint appointment in the departments of Earth, Atmospheric and Planetary Sciences and of Chemistry at MIT, where he has continued his research on global atmospheric chemistry issues. He was named MIT Institute Professor in 1997. Molina is a member of the US National Academy of Sciences, the Institute of Medicine, and the Pontifical Academy of Sciences. He has served on the US President s Committee of Advisors in Science and Technology, the Secretary of Energy Advisory Board, the National Research Council Board on Environmental Studies and Toxicology, and on the boards of US-Mexico Foundation of Science and other non-profit environmental organizations. In winning the Nobel Prize in 1995, Molina became the first Mexican–American to win a Nobel Prize and indeed the first Mexican-born Scientist to become a Nobel Prize Winner.

See also: Stratosphere, Ozone Trends, Volume 1.

REFERENCE

Molina, M J and Rowland, F S (1974) Stratospheric Sink for Chlorofluorocarbons. Chlorine Atom-Catalyzed Destruction of Ozone, *Nature*, 249, 810–812.

MICHAEL C MACCRACKEN *USA*

Monitoring Systems, Global Geophysical

Christopher N K Mooers
University of Miami, Miami, FL, USA

To help understand and assess the state of the Earth system, monitoring systems are used to observe (or sample) geophysical variability in a sustained, standardized fashion, as opposed to conducting one-time exploratory surveys or short-term observational experiments. Ideally, monitoring systems would sample a comprehensive set of geophysical variables to a uniformly high standard on a global basis; the observed data would be freely available to all concerned parties; and the data, or derived, high-order information products, would be well utilized by local, regional, and global decision-makers. This ideal remains to be ful lled in the perhaps not too distant future due to the onrush of technology, scienti c understanding, and societal need.

The Earth system is variable on space and time scales that depend upon the dynamics of the intrinsic and extrinsic processes producing natural variability. Convolved with natural variability is that due to activities of human society, usually termed anthropogenic variability. For Earth system management, it is important to be able to deconvolve the patterns of natural and anthropogenic variability in order to contemplate effective mitigating action of human-induced change. However, natural and anthropogenic variability may or may not be readily distinguishable depending upon whether their characteristic space and time scales and space- time patterns are similar or dissimilar, as dictated by the attributes of their forcing. Therefore, a close understanding of natural and anthropogenic processes is also often required to distinguish anthropogenic from natural variability.

Below, several general principles and issues of geophysical monitoring are discussed, some aspects of discipline-based geophysical monitoring are outlined, and a future outlook is conveyed for geophysical monitoring in support of the new global environmental ethic needed for sustainable Earth system management.

GENERAL CONCEPTS

Historically, geophysical monitoring systems have evolved from ancient astronomical observatories, record keeping by religious missionaries and imperial forces, and exploratory and expeditionary activities. Modern autonomous instrumentation facilitates systematic observations by large numbers of entities and at diverse locations and, with a modicum of effort, in a coordinated fashion. Understandably, monitoring systems, driven by particular societal needs or scientific opportunities, have developed on a disciplinary basis, and often on a national basis. However, this disciplinary and national partitioning of monitoring efforts may lead to lost opportunities and other inefficiencies as well as undermining a unified, comprehensive approach. Because economic or military advantage can be derived from monitoring certain geophysical variables (e.g., atmospheric winds and oceanic currents), the data accrue value, which leads to their treatment as proprietary, thereby undermining the principle of free exchange of geophysical data.

For enlightened stewardship of the Earth s environment, natural processes must be understood. This requires the use of the modern, combined approach of observations, experiments, theory, and numerical simulations (modeling). The most reliable and quantitative observations are obtained from scientific instrumentation. However, instrumental records, as opposed to anecdotal records or proxy (paleo) records, have begun to become available over only the past century or two. Hence, it is important to conduct process studies to understand the relationship between the

proxy (usually geochemical, biological (e.g., microfossil), or biogeochemical in nature) and instrumental records. With that understanding, the proxy records may be more fully interpreted (*see* Natural Records of Climate Change, Volume 1), and the instrumental records may be extrapolated backwards in time.

Because geophysical variables vary with spatial position, observations are generally required at numerous locations to form the basis for producing maps in the horizontal plane and transects in vertical planes. Because they also vary with time, time series of observations are required, too. Ideally, the data product consists of time sequences of maps and transects and plots in a coordinate plane defined by time and one spatial dimension (such diagrams are commonly called Hovmuller diagrams). To isolate the essential physics, it is frequently effective to examine geophysical variables in Fourier space for one or more of the four coordinates (three spatial dimensions plus time).

Exceptions to the previous remarks occur. For example, some monitoring strategies are focused on space–time events that may be restricted in space or time and have characteristic signatures (i.e., patterns). Such events can be targeted with a specific, tailored monitoring strategy (which can be termed a strategic sentinel), analogous in a very general fashion to the canary in the coal mine paradigm.

Enlightened stewardship of the environment also requires assessments of trends in geophysical variables to identify statistically significant departures from normal conditions. For this purpose, systematic observations need to be made routinely to specified accuracy and precision. Observations need to be sampled frequently enough in time and space to avoid aliasing from high temporal or spatial frequencies and over long enough durations and distances to avoid biasing from low temporal or spatial frequencies, respectively. One approach is to install temporally sampling sensors at various fixed sites and then connect these sensors into a network or observing system (e.g., *see* Air Quality, Global, Volume 1). Another approach is to use a platform; such as, an artificial satellite to convey remote sensing systems; the satellite can either be an Earth orbiting satellite which periodically repeats its ground track or a geostationary satellite which continuously views a fixed sub-domain of the Earth (*see* EOS (Earth Observing System), Volume 1).

Such routine, sustained, professional observational campaigns define geophysical monitoring, which is only interrupted due to a technical or some other kind of a disturbance, upgrading of the instrumentation, or implementation of an observing system redesign. These routine monitoring programs are usually carried out by governmental agencies. The governmental monitoring services generally provide documentation of sensor calibrations and network nodes (e.g., date of installation, geographical coordinates, and altitude relative to mean sea level); installation, operation, and maintenance of the sensor network; and data acquisition, transmission, quality control, archival, and distribution. Error estimates for the observations (e.g., those due to sensor noise) and the overall observing system network (e.g., those due to small scale (unresolved) geophysical noise) are byproducts of the calibration and monitoring activities that are essential for quantitative analysis of the results. With the advent of modern technology, network sensors and nodes have become electronic and automated, and data are typically telemetered from remote nodes to central sites in real time via landlines, radio links, communication satellites, or equivalent means.

Scientific researchers have several roles to play regarding monitoring systems. For example, they are often responsible for new sensor development. Similarly, they are often responsible for the design and advocacy of sensor networks, and the conduct of sampling experiments to determine the network s design parameters. Less commonly, they may create or assist in operating a monitoring system. Very commonly, researchers rank, along with environmental decision-makers, as prime users of the data sets derived from monitoring systems.

An interesting irony is that when a new observational capability bursts upon the scene, the data acquisition activities are seen as a source of exciting science. Once the sensing technology becomes proven, routine, and institutionalized, the data acquisition activities are viewed as routine monitoring, or even mindless monitoring. However, once the observing network has operated long enough to document at least a large fraction of one cycle of energetic, longer period variability, the data become highly valuable again for scientific research, and the monitoring system achieves heroic status. The only way to avoid the mindless monitoring appellation, is to keep research scientists engaged with the monitoring process, which requires support for critical investigations conducted with the monitoring data. Such scientific investigations also provide invaluable feedback on the performance of the monitoring system.

Another irony is that geophysical data are useful to the largest number of users, which includes managers and operators as well as researchers, when they are fresh; i.e., available in real time. Subsequently, their utility languishes until the data record becomes long enough to provide a useful historical and statistical perspective and then recognition and appreciation of the value of the historical data record greatly increases.

It is generally understood that it is impractical, or even impossible, to observe every geophysical variable of interest with sufficient accuracy and space–time resolution, and that dynamical models must be used for full field estimates. However, the models are very dependent upon the adequacy of observations for initial and boundary conditions,

forcing functions, and adjustment of internal values through data assimilation. In other words, the available observations are used to constrain the models; conversely, the models are used to interpolate and extrapolate dynamically the inevitably sparse observations.

These comments have only addressed a few aspects of the synergistic interchange between observations and models. For example, pure model simulations should be validated with observations to determine whether the model s verisimilitude is sufficient before attempting data assimilation. As another example, by sub-sampling a model simulation according to the template of a candidate observing system and carrying out various analysis protocols, models can be used to help assess the efficacy of the design of a candidate observing system network on an *a priori* basis.

Overall, the modern concept of geophysical monitoring systems includes the incorporation of models and observations to provide optimal field estimates that, in turn, can be analyzed and synthesized to produce higher order information products (replete with error estimates) than those based on observed or modeled variables alone. Additionally, a modern monitoring system includes a vital subsystem for information management that treats such functionalities as data acquisition, quality control, archival, retrieval, display, transmission, and distribution.

Given a geophysical monitoring system and a data archival strategy, a natural adjunct is the formation of a climatology for broad use that characterizes a variety of properties of the evolving data set in statistical and geophysical terms. Because numerous assessments for societal purposes and scientific analyses are likely to stem from a climatology, it is essential to document its attributes and understand its quality.

Having reviewed some of the general principles and issues associated with geophysical observing systems, attributes of the monitoring systems of several geophysical disciplines will next be considered. Of interest is their adequacy for fulfilling the needs of society in addressing stewardship issues associated with a sustainable Earth system, and in protecting its inhabitants from natural hazards by issuance of early warnings and otherwise mitigating their impacts.

STATUS OF GLOBAL GEOPHYSICAL MONITORING SYSTEMS

Global geophysical monitoring systems tend to be organized by scientific disciplines. Increasingly, however, scientific and logistical advantages in coordinating such efforts are being recognized. Hence, where possible, these synergistic arrangements or opportunities will be noted. The three fluid geophysical disciplines, namely meteorology, hydrology, and oceanography, will be addressed first because they have much in common. Then, the monitoring systems of the four solid earth geophysics disciplines; viz., seismology, geodesy, geomagnetism, and volcanology are addressed.

Meteorology

Probably the most advanced geophysical discipline in modern monitoring strategies is meteorology due to the everyday interest in forecasting the weather, and now short-term climate, by observing and modeling the atmosphere. Certainly, it is the most advanced in pioneering the assimilation of observations from a very diverse observing system into numerical models to make optimal estimates of atmospheric variables and for the initialization of those models to facilitate their use in the forecast mode. This task, and the infrastructure necessary to accomplish it, is typically termed operational meteorology. Beginning over a century ago, military, agricultural, and aviation applications provided the motive force for the development of operational meteorology, which now serves innumerable users, including the general public. A combination of observing systems is used, including ground stations to observe winds, wet and dry temperature, pressure, visibility, etc. radiosonde stations to observe vertical profiles of temperature and humidity; remote sensing radars for vertical profiles of winds; upper tropospheric winds and temperatures from commercial aircraft; and satellite imagers for surface temperature and clouds and sounders for vertical profiles of temperature. Most nations have national weather services that coordinate their operational meteorology activities through the World Meteorological Organization (WMO). WMO is a United Nations (UN) affiliated agency, which oversees, for example, the global telecommunication system (GTS) for aggregating and disseminating real time data reports and numerical model products as part of the World Weather Watch (WWW). *En passant*, the research meteorology (and hydrology and oceanography) community is also a beneficiary of the WWW data sets and relies on them to a significant degree.

Hydrology

Hydrological observing systems serve a variety of needs ranging from water resource management to study of climate issues ranging from droughts and floods to hazardous events such as flash floods and heavy and harmful precipitation. Such observing systems include rain gauges, stream gauges (for river flows), snow pack thickness, plus standard surface atmospheric variables. With so much in common with operational meteorology, operational hydrology is often conducted by the national weather services for real time hydrology, and is facilitated by WMO. Additional gauging and modeling are done by the engineering

and geological components of hydrology dealing with water resource management on a seasonal and longer-term basis. With the growing concern for climate and global change, the scientific issue of understanding the global water cycle has emerged as a priority, necessitating the use of new technologies such as satellite remote sensors for producing global, synoptic estimates of precipitation and ice cover.

Oceanography

Ocean observing systems serve the needs of mariners who navigate the ocean; the industrial entities involved in offshore operations (oil and gas extraction, waste disposal, fishing, aquaculture, and so forth); the environmental managers responsible for sustainable marine ecosystems and public health; plus those who are engaged in climate analysis and prediction. The systematic and cooperative acquisition of oceanic data by the maritime industry and navies of the world began in the 1850s, leading to bottom topographic charts of the seafloor and climatological atlases of ocean surface currents and atmospheric surface winds, for example. Despite this early start, operational oceanography has been slow to develop except in the context of several leading navies in recent times. However, today there exists an embryonic ocean observing system which is in the process of being upgraded through the Global Ocean Observing System (GOOS), a program of the Intergovernmental Oceanographic Commission (IOC), an agency of United Nations Educational, Scientific and Cultural Organization (UNESCO) (*see* Global Ocean Observing System (GOOS), Volume 2).

The principal elements of the *de facto* ocean observing system are the coastal sea level network, consisting of hundreds of telemetering coastal tide gauges that are non-uniformly distributed and biased toward the Northern Hemisphere; extensive, high quality marine lower atmospheric and upper oceanic data from dozens of merchant vessels in the Ships of Opportunity and Volunteer Observing Ship (VOS) programs; satellite-tracked, surface layer drifters for surface currents, temperature, and atmospheric pressure; and satellite infrared imagery from which sea surface temperature is determined. Rapidly emerging components for the *de facto* ocean observing system include the autonomous profiling (for temperature and salinity), telemetering Palace Floats, with 3000 to be deployed as the global ocean; Array for Real-time Geostrophic Oceanography (ARGO), which will also provide current estimates at their rest depth (circa 1 km) as well as at the sea surface; satellite radar altimeters, which determine the sea surface height with high precision and characterize jets and eddies; and satellite radar scatterometers that measure surface winds with high accuracy.

The international Global Ocean Data Assimilation Experiment (GODAE) program is under development; it aims to make ocean state estimates with global ocean models that assimilate Palace Float temperature and salinity profiles, altimetric sea surface heights, and so forth. The results will allow improvement of the design of GOOS and help facilitate the growth of operational oceanography. To obtain absolute values for sea surface heights measured from satellite radar altimeters, it is essential to have an accurate geoid; conversely, sea surface height variability due to ocean dynamics constitutes a noise source for geodetic measurements. Similarly, for coastal tide gauges to provide absolute sea levels for use in sea level rise studies, they must be connected to a geodetic grid to take into account vertical displacements due to long period tectonic motions and glacial rebounds. Thus, there are strong needs for coordination between oceanography and geodesy.

With the ocean system strongly forced by atmospheric and hydrologic processes on time scales of the order of a year or less, and with oceanic processes forcing the atmospheric and the hydrologic systems on longer time scales, there is an increasingly strong basis for cooperation between the fluid geophysics disciplines. Because of the societal concern for possible changes in the contemporary climate regime, and because of the inadequacies of the *ad hoc*, *de facto* monitoring systems currently used for climate change assessments, the international meteorological, oceanographic, and hydrological communities, under the leadership of WMO, are developing the Global Climate Observing System (GCOS), which is an intersection between WWW and GOOS, plus some enhancements (*see* GCOS (Global Climate Observing System), Volume 1).

Seismology

Driven by societal concern for earthquake prediction and nuclear weapons testing detection, as well as basic research, numerous arrays of seismometers have been established, some of which are of a regional nature, while others are of a global nature. Seismic wave propagation theory is sufficiently advanced that it can be used together with the seismic array data to estimate synergistically geophysical and geochemical information about the structure of the Earth s crust and mantle using inverse methods. Much of the seismological observing system is organized on a national basis, and some on an international basis. Frequently, arrays are operated by consortia of research institutions and/or agencies. Because two-thirds of the Earth s surface is covered by the ocean, it is essential to extend the seismological arrays from the continents to the deep sea for global coverage. For logistical and technological reasons, as well as some scientific reasons, this need is leading to a partnership between marine seismologists and those oceanographers concerned with developing and operating long-term seafloor observatories, as in the International Ocean Network (ION) program.

Geodesy

Geodesy is concerned with the accurate and precise shape of the Earth s surface (i.e., the geoid) and its temporal variation due to tectonic activity, which has innumerable societal applications ranging from accurate topographic mapping to property rights to land-use planning to flood control strategies, to name several. Hence, mapping the Earth s gravitational field and determining precisely the rate of rotation of the Earth (equivalently, the length of the day) are central tasks for geodesy (*see* Atmospheric Angular Momentum and Earth Rotation, Volume 1). The monitoring methods of geodesy include the use of networks of gravimeters, both in the stationary and portable modes. They also include the use of precise navigation systems; e.g., as the satellite based global positioning system (GPS) that utilizes radio signals whose propagation is perturbed by atmospheric effects, necessitating collaborations between geodesists and meteorologists. Conversely, the refracted radio signals can be used, through inverse methods, to estimate the thermal structure of the atmosphere. The methods of geodesy can be combined to provide precise positioning, in three space dimensions and time, of moving as well as stationary objects. Marine geodesists have been extending precise positioning to the seafloor and the ocean interior in recent years, making feasible programs such as ION, as mentioned above.

Geomagnetism

The geomagnetic field is highly variable in time as well as space due to solar–terrestrial interactions stemming from solar flares and involving ionospheric processes that affect radio propagation and electrical grids on the Earth s surface. Thus, a network of geomagnetic observatories, equipped with magnetometers, was initiated over a century ago to monitor geomagnetic storms in the ionosphere. With the use of upper atmospheric models, forecasting of ionospheric conditions is emerging as an operational capability, for which coordination with operational meteorology is essential. Magnetometer networks are being established on the seafloor, as well.

Volcanology

Volcanism is monitored by methods ranging from seismometers to synthetic aperture radar (SAR) imagery and GPS receivers to detect vertical movements of the surface for advanced warning of society as well as for basic research.

SUMMARY

Over the past two decades, with the rising concern for global sustainability issues, the maturing of enabling technologies; e.g., satellite remote sensing and numerical modeling, and the increasing scientific attention devoted to global processes, the new multidisciplinary field of Earth system science has emerged. Earth system science encompasses global geophysics, biogeochemistry, and marine and terrestrial ecology.

Earth system science aims to understand how the physical, chemical, and biological systems of the Earth cofunction (*see* The Earth System, Volume 1). Thus, global geophysics, geochemistry, and geo-ecology are beginning to share common interests and strategies for monitoring (both observing and modeling), including an abiding need for observed data. It is often possible for these disciplines to co-locate observing networks for mutual advantage, both logistically and scientifically, and to share platforms (e.g., satellites and seafloor observatories), information management systems, and other resources. In this spirit, the Global Terrestrial Observing System (GTOS) is being developed under the leadership of the United Nations Environmental Programme (UNEP), which adds terrestrial and marine ecology to the disciplinary mix (*see* Global Terrestrial Observing System (GTOS), Volume 2). At another level of integration, GOOS, GCOS, and GTOS are being linked through the Integrated Global Observing Strategy (IGOS).

As an entity in the International Council for Science (ICSU), the International Union of Geodesy and Geophysics (IUGG), through its seven constituent disciplinary associations, fulfills several roles in global geophysical monitoring (*see* IUGG (International Union of Geodesy and Geophysics), Volume 1). For example, it has promoted and fostered several disciplinary observing system networks through various permanent services. It was a leading coordinator for the International Geophysical Year (IGY) in 1957, that provided a one-time, integrated global geophysical survey (*see* IGY (International Geophysical Year), Volume 1). Above all, it periodically assembles geophysicists from the less developed and more developed countries to discuss scientific results based, in part, on data from the observing systems and to plan extensions of these systems. Such communications and collaborations play a crucial role in technology transfer from the geophysicists of the more developed countries to those from the less developed countries. In turn, the collegial relations so engendered help form the basis for access to the less developed nations for installation of monitoring sites to extend the global networks and fill data gaps. In 1999, IUGG adopted a resolution advocating the advance of Integrated Global Earth Monitoring Systems (IGEMS) by supporting the various observing system networks in existence and planned and promoting the free exchange of geophysical data. It can be anticipated that this new resolution will lead, in due course, to the IUGG playing a planetary stewardship role by assembling and promulgating global assessments on the state of the Earth system.

The future of global geophysical monitoring systems is strongly linked to the rising moral imperative (well represented in the deliberations of the United Nations Commission on the Environment and Development (UNCED), popularly known as the Earth Summit, Rio de Janeiro, Brazil, 1992) of understanding the global environment at a level firmly grounded in observations, as well as theory and modeling, such that increasingly reliable assessments and projections, with defensible error estimates, can be made about the state of the Earth system in support of Earth management protocols designed to achieve sustainability. If done well, the world community will then be able to move beyond the present confusion and uncertainty over the prospects for climate and global change and make more rational and timely decisions dealing with anthropogenic influences on the Earth system.

Monsoons

The term *monsoon*, believed to be of Arabic origin, refers to large-scale seasonally varying wind flows associated with major continental land masses and neighboring oceans. In summer, air over the land is heated, expands (so becomes less dense), rises, cools and spreads outward at high altitude. Near the surface, moist air flows in from the ocean to replace it, bringing clouds and rain as rising and cooling take place. In winter, the atmosphere over land areas cools, becomes more dense, subsides, warms as it is compressed, and dry currents flow from land to sea. The phenomenon is in essence a sub-continental scale version of the sea breeze/land breeze regime observed at coastlines around the world.

The largest and best known monsoon system is that centered on the Indian subcontinent, although monsoonal climates are also found in tropical regions of Australia, Africa, Central America and the southwestern US. In March and April the Indian sub-continent begins to heat strongly, with some of the highest surface temperatures of the year occurring by May. A large difference between land surface temperature and sea surface temperature results. As a consequence, the wintertime flows of dry air from the land to the ocean are replaced by moist winds coming in off the ocean due to the low pressure created over the land by the warm and rising air currents. A large low pressure system dominates Southwest Asia, while the Himalayas and Hindu Kush mountains trap warm air within the Indian ocean basin, preventing it from continuing north without rising and precipitating out its moisture. Intense winds blow from the southwest around the low pressure area and over the Indian Ocean, coinciding with the seasonal northward shift in the intertropical convergence zone (ITCZ). As the winds cross the Indian Ocean they pick up moisture that is released in tremendous convective storms (usually beginning in early June) as they pass over the sub-continent. Most of South Asia receives nearly all of its rainfall from the summer monsoon, with the months of May, June, and July being the wettest of the year. Variability in monsoon rainfall can make the difference between starvation and prosperity for millions. The Indian Meteorological Department issues regular forecasts of monsoon onset and precipitation based primarily on statistical methods calibrated by the long history of past events.

In the winter, a reverse process occurs. The land surface cools faster than the sea surface as a result of water s capacity to retain heat. The winds turn and carry continental air out over the ocean, the ITCZ retreats to the south, subsiding air dominates the Indian subcontinent, and dry conditions prevail.

This simple picture is but the most prominent feature of a complex circulation system that dominates a major portion of the planet. Other branches of the Asian monsoon control the seasonal precipitation of Southeast Asia and Southern China. Moreover, the strength and timing of the Asian monsoon is linked with the Pacific-wide Walker circulation associated with the El Niño/Southern Oscillation phenomenon (*see* El Niño and La Niña: Causes and Global Consequences, Volume 1). The latter connection offers some promise of improving predictions of the Asian monsoon.

Similar, though less intense, monsoonal circulations are found in West Africa, Australia, the southwestern US, and northern Mexico. In the context of a changing global environment, it is natural to expect changes in the massive monsoonal circulations. Indeed, Chinese investigators (Fu and Yuan, 2001) have identified connections between human-induced land-use changes over China, consequent changes in surface heating and evapotranspiration, and monsoon intensity. Future changes in global circulation associated with global warming may be expected to influence the timing and intensity of monsoons as well. Because the behavior of the Asian Monsoon is a major factor in life-support systems for millions of peoples in Asia (*see*, for example Fisheries: Pollution and Habitat Degradation in Tropical Asian Rivers, Volume 3; Indus Basin: a Case Study in Water Management, Volume 3), research into the effects of global environmental change on the behavior of the Asian monsoon is of highest research priority.

See also: Atmospheric Motions, Volume 1.

REFERENCE

Fu, C B and Yuan, H L (2001) A Virtual Numerical Experiment to Understand the Impacts of Recovering Natural Vegetation on the Summer Climate and Environmental Conditions in East Asia, *Chin. Sci. Bull.*, 46, 691–695.

FURTHER READING

Webster, P J (1987) The Elementary Monsoon, in *Monsoons*, eds J S Fein and P L Stevens, John Wiley & Sons, New York.

JAGADISH SHUKLA AND JOHN S PERRY *USA*

Montreal Protocol

see Ozone Layer: Vienna Convention and the Montreal Protocol (**Volume 4**)

Mountain Climates

Roger G Barry
University of Colorado, Boulder, CO, USA

Mountain climates are highly variable spatially as a result of the combined effects of latitude, altitude, distance from the ocean, and local relief.

Mountain ranges occupy about 20% of the total land surface of the Earth. A mountain can be defined as terrain having a sufficient altitudinal range to cause vertical differentiation of climatic elements and vegetation cover. The common elements of mountain climates include:

1. reduced air density, pressure and oxygen availability with consequences for physiology and bioclimatology; about 50% of the population will experience some symptoms of hypoxia above 3 km altitude;
2. reduced vapor pressure;
3. vertical differences in mean temperature over 1 km in elevation equivalent to the gradient over about 1000 km distance poleward;
4. corresponding vertical/latitudinal differences in the duration of snow cover which increases nearly linearly with height;
5. solar radiation (including UV–B) increases of 5–15% km^{-1}, due to decreased attenuation by moisture and aerosols;
6. strong local climatic contrasts, termed topoclimates, as a result of slope orientation, topographic shading and air drainage.

The height, length and width of a mountain barrier determine its effects on airflow, clouds, precipitation and frontal systems over a range of scales. A high, broad barrier may block the airflow causing it to flow through gaps or around the mountains, thereby often creating local winds. The temperature conditions on an isolated mountain are closer to those in the free air than over an elevated plateau or mountain range where the terrain modifies the lower atmosphere through energy transfers and orographicaly-induced cloud cover.

Mountain climates are highly variable as a result of the basic controls of latitude, proximity to an ocean in the upstream direction of prevailing winds and weather systems, and relief. The imprint of latitude on mountain climates is evidenced by the diurnal and seasonal rhythms of solar radiation and temperature. High equatorial mountains experience summer every day and winter every night (Hedberg, 1964), whereas high latitude mountains have a polar night and summer midnight sun. An upwind ocean is crucial as a moisture source. In North America, the western Cordilleras have large annual precipitation totals, up to 80% of it falling as snow in the winter half year. Precipitation amounts decrease progressively inland, with exaggerated dry zones (rain shadows) in the interior valleys and basins, although western slopes of the successive ranges have augmented totals. In South America, the western slopes of the Andes in Peru and northern Chile are semi-arid because the winds are mostly from the east and the main moisture source is the Amazon Basin.

Wind speeds increase steadily with altitude in the mid-latitude westerlies but in the tropical easterlies there is little altitudinal change and winds may even decrease upward. Orography gives rise to local winds of dynamic origin (foehn, barrier winds) and thermodynamic origin (anabatic/katabatic up/down slope winds; mountain/valley winds, respectively, blowing from those directions). Altitudinal gradients of precipitation vary widely. In mid-latitudes, annual precipitation totals tend to increase with altitude on windward slopes up to about 3 km as the increase in wind velocity offsets the decrease in moisture content. Because evaporation also tends to decrease with altitude in extratropical latitudes, the runoff increases, with altitude, at a faster rate than the precipitation increases. In equatorial regions, the largest rainfall totals are usually on the lowest slopes and amounts decrease above. In the easterly trade winds, the maximum occurs on windward slopes between about 700 and 1200 m altitude, near the mean cloud base level. Some of the heaviest annual precipitation totals are recorded in mountain areas – 11 m $year^{-1}$ in the South Island of New Zealand Alps around 2000 m altitude, 11–12 m $year^{-1}$ around Cherrapunji (1300 m) in Assam from the southwest monsoon, and 12 m $year^{-1}$ at 1570 m elevation on Mount Waileale, Kauai, where the 4200 m summit receives less than 50 cm.

Mountain observatories in Europe especially provide century-long records of climatic data. Stations in the

Alps and Pyrenees indicate a mean annual temperature increase of about 1 °C per 100 years which at three stations exceeds that observed at lowland stations in western Europe. However, in central Europe the period 1930–1950 was warmest and, in eastern Europe and the Caucasus, there has been little twentieth century warming. For 30–70 °N most stations at elevations above 1500–2000 m show a diurnal asymmetry in warming trends with minimum temperatures for 1951–1989 increasing more than maximum temperatures, although in the Alps both maximum and minimum temperatures show an increase.

There has been widespread warming on tropical mountains during 1975–1990 as evidenced by a rise in the freezing level at radiosonde stations; this is particularly pronounced between 15 °N and 15 °S. This appears to be responsible for a dramatic shrinkage of tropical glaciers in South America and Irian Jaya although decreased precipitation is suggested for glaciers on East African peaks.

REFERENCE

Hedberg, O (1964) Features of Afro-Alpine Plant Ecology, *Acta Phytogeogr. Suec.*, 1–144.

FURTHER READING

Barry, R G (1992) *Mountain Weather and Climate*, 2nd edition, Routledge, London, 1–402.

Diaz, H E, Beniston, M, and Bradley, R S, eds (1997) *Climatic Change at High Elevation Site*, Kluwer Academic Publishers, Dordrecht, 1–530.

Ives, J D and Barry, R G, eds (1974) *Arctic and Alpine Environments*, Methuen, London, 1–999.

Price, M F and Barry, R G (1997) Climate Change, in *Mountains of the World. A Global Priority*, eds B Messerli and J D Ives, Parthenon Publishing, New York, 409–445.

Whitehouse, C D (2000) *Mountain Meteorology: Fundamentals and Applications*, Oxford University Press, Oxford, 368.

MSU (Microwave Sounding Unit)

The MSU was part of the atmospheric sounding instrument suite on nine National Oceanic and Atmospheric Administration (NOAA) polar orbiting spacecraft beginning with TIROS-N in December 1978 and ending with NOAA-14, launched in late 1994.

The MSU measures the intensity of microwave radiation emitted from atmospheric oxygen at four frequencies near the 60 GHz absorption band; 50.30, 53.74, 54.96 and 57.95 GHz. The intensity of the radiation is proportional to the temperature of oxygen and thus to that of the atmosphere. The emissions arise from broad vertical layers with approximate levels of maximum emissions being the surface, 6, 15 and 20 km for the four frequencies, respectively.

The MSU is a cross-track scanning instrument, with 11 Earth-views per 25.6 s scan. After each set of 11 Earth-views, the rotating antenna points to cold space and then to a hot target on board for calibration points. The temperatures of cold space (2.7 K) and the hot target (by embedded thermometers) are known and from this information, the representative temperature of the Earth-views may be calculated. Adjustments, however, are required to account for the effects of changing altitude, longitudinal drifting of the satellite orbit and instrument gain fluctuations. Once adjusted, the time series reveal exceptionally close agreement with independent measurements produced from weather balloons.

Several data products have been generated, including a low-mid tropospheric (surface to 8 km) and lower stratospheric (17–23 km) temperature series. For the period 1979–2000, the global mean low-mid troposphere indicated little trend either up or down while for the same period, the conventional surface temperatures indicated a significant increase (*see* Tropospheric Temperature, Volume 1). The lower stratospheric temperatures reveal significant cooling. Theoretical calculations have shown that these three distinct layers will respond differently to climate forcing and thus reveal varying trends.

JOHN R CHRISTY *USA*

Munk, Walter

(1917-)

Walter Munk, one of the great founding fathers of modern oceanography, was born in Austria in 1917. At age 14 he was sent to New York to be trained in the financial tradition of his maternal family. After 3 years in finances, he enrolled in the California Institute of Technology (Caltech) and graduated in applied physics in 1939.

Subsequently, he obtained a masters degree in geophysics from Caltech and was admitted to the PhD program at Scripps Institution of Oceanography, where, for some time, he was the only student under Harald Sverdrup, Scripps director. During World War II, he served for 18 months in the Field Artillery and the Ski Troops. Afterwards, he joined Sverdrup, Roger Revelle and Richard Fleming, the leading oceanographers of the time, at the US Navy Radio and Sound Laboratory. He finished his PhD dissertation in 1947 and joined Scripps Institution as assistant professor of geophysics. He has remained at Scripps ever since, becoming associate professor in 1949 and full professor in 1954. He served as director of the La Jolla Laboratory of the Institute of Geophysics and Planetary Physics in 1959–1982; retired as full professor in 1988 but was recalled to active duty in the same year.

Walter Munk is the author of 250 publications in international journals and volumes. He has contributed seminal papers in all his fields of interest, most often being the initiator of the field itself. In the period 1942–1950, his research focussed on wind waves, where he pioneered the first wind wave prediction model that was of crucial importance for the US Navy in maritime operations. He successively worked on the wind-driven ocean circulation, producing a series of fundamental papers that first explained how the surface currents of the ocean are driven by the wind stress field. In 1950–1960 his interest concentrated on the Earth s wobble and spin and, with Gordon MacDonald, he wrote an encompassing geophysical discussion of the rotation of the Earth which is a classic in the literature. Deep-sea drilling ventures, sun glitter and radar clutter, and the study of southern swell, surfbeats, edge waves, tsunamis and tides kept him busy through the 1960s. In the early 1970s his curiosity was attracted by internal waves, leading to the Garrett-Munk spectrum for internal waves, the explanation of which is still a challenge for oceanographers. In the last 15 years, Munk s major interest has been focussed on the exploration of the ocean interior through sound waves, inventing, with Carl Wunsch, the field of Ocean Acoustic Tomography, that has culminated with the Acoustic Thermometer of Ocean Climate (ATOC) Experiment (*see* Acoustic Thermometry, Volume 1).

Walter Munk is a member of all the major scientific societies including the US National Academy of Sciences, the Russian Academy of Sciences, and the Royal Society of London. He has received honorary degrees from the University of Bergen, Crete, Cambridge (UK) and more than 20 other national and international medals and awards, including the National Medal of Science in 1983 from President Reagan. In 1999 he was awarded the prestigious Kyoto Prize in Basic Sciences.

PAOLA MALANOTTE-RIZZOLI *USA*

Munn, Robert Edward (Ted)

(1919-)

Robert Edward Munn was born July 26th, 1919 in Winnipeg, Manitoba, Canada. Armed with a BA in Mathematics from McMaster University in 1941, Munn joined the Meteorological Division of the Canadian Department of Transport. After 7 months of courses at the University of Toronto learning how to forecast the weather, he was posted to Gander, Newfoundland in 1943 where he was part of a team that forecast for transatlantic flights and North Atlantic operational patrols.

In 1948 Munn moved to the Public Weather Office in Halifax where he published his first paper, *A Survey of the Persistence of Precipitation at Halifax*. In 1956 he became an air pollution research meteorologist in Windsor, Ontario, seconded to the Canada–US International Joint Commission. While there he earned a PhD from the University of Michigan in 1961 on intermediate-range Lagrangian diffusion modelling, and published his first book, *Descriptive Micrometeorology*.

In 1959 Munn became head of the new Micrometeorology Section in Transport Canada s Meteorological Branch in Toronto. He published extensively throughout the 1960s and 1970s on air pollution, turbulence and micro- and meso-scale meteorology, lake-breeze studies and urban heat-island effects. In 1971 Munn founded *The International Journal of Boundary-layer Meteorology* and served as editor in chief for more than 25 years. Over the years he managed to combine journal editorship with his many other duties by setting aside the early morning hours from 5 to 7.30 for editing and proof reading. He was editor in chief for Scientific Committee on Problems of the Environment (SCOPE) John Wiley Environmental Monographs from 1979–1998 and is currently a member of five journal editorial boards (*see* SCOPE (Scienti c Committee on Problems of the Environment), Volume 4).

Munn spent 6 months as a visiting professor at the University of Stockholm in 1971. He remained with the Atmospheric Environment Service (AES) (now known as the Meteorological Service of Canada (MSC)) until 1977 when he retired at age 58, having given 35 years to public service, and began a new career, as an Associate at the Institute for Environmental Studies at the University of Toronto. He became increasingly interested in environmental policy, risk assessment and sustainable development during the period

from 1977 to 1985. He made good use of his experience in evaluating and synthesising scientific research results and served as an interpreter of these results for the broader, non-scientific audience of policy makers. Munn was a visiting scientist at Chelsea College, University of London in 1979.

Between 1985 and 1988 he served as Leader of the Environment Program and deputy director at the International Institute for Applied Systems Analysis (IIASA) in Laxenburg, Austria. On his return to the Institute for Environmental Studies in 1989 Munn focused on long-term global change and global environmental issues, contributing frequently to interdisciplinary syntheses. Munn was one of the principal editors of the United Nations Environment Programme (UNEP) 1992 book: *The World Environment 1972- 1992*. He was editor for a 1996 monograph on Policy Making in an Era of Global Environmental Change for the Royal Dutch Academies of Engineering and International Council for Science (ICSU), and advisor to the United Nations Conference on Environment and Development (UNCED) and ICSU on an important project to define environmental issues in the next millennium.

In addition to his extensive editorial work, Munn has published or edited 20 books and monographs, authored more than 200 papers and received numerous international and national fellowships, awards, medals and prizes, including: Fellow of the Royal Society of Canada; Fellow of the American Meteorological Society (AMS); Fellow of the American Association for the Advancement of Science; and Fellow of the Canadian Meteorological and Oceanographic Society; the Frank A. Chambers Award of the Air Pollution Control Association; the Paterson Medal (highest Canadian award for meteorology); 1998 AMS Award for Outstanding Achievement in Biometeorology; and most recently 1999 AMS Award: The Walter Orr Roberts Lectureship.

His latest challenge is as editor in chief of this John Wiley *Encyclopedia of Global Environmental Change*. Despite his heavy workload he is always relaxed, never seems to be rushed, spends lots of time at galleries and concerts and, somehow, gets everything done.

DOUGLAS WHELPDALE *Canada*

N

Natural Climate Variability

Michael Ghil
University of California, Los Angeles, CA, USA

Climate varies on all time scales, from one year to the next, as well as from one decade, century or millennium to the next. The complex nature of this variability is a major obstacle to the reliable identi cation of global changes brought about - in the past, present or future - by the presence and activities of humanity on this planet.

This article discusses observed climate variability on various time scales, paleoclimatic, interdecadal to centennial, seasonal to interannual, and intraseasonal. Key ideas to explain the observed variability are presented, and some of the models used to explore these ideas are mentioned. Finally, we discuss brie y the implications of these ideas for the detection and prediction of climate change.

Climate variations happen on all time scales, as well as on all spatial scales, from the regional to the global. As an example, the oceanic El Niño (*see* El Niño and La Niña: Causes and Global Consequences, Volume 1) phenomenon is most pronounced in the tropical Pacific off the coast of Peru, but the associated Southern Oscillation in the atmosphere has far-reaching, nearly global implications. The combined El Niño/Southern Oscillation (ENSO) (*see* El Niño/Southern Oscillation (ENSO), Volume 1) variability has a seasonal component, hence the name El Niño, due to its preference for appearing first at Christmas time. It also has a quasi-biennial component, with a characteristic recurrence time of 2–2.5 years, and a low-frequency one, with a recurrence of 4–5 years. The interaction of these three distinct modes of variability renders the evolution of sea-surface temperatures across the tropical Pacific fairly irregular. An additional cause of the irregularity observed in these temperatures lies in the frequent weather perturbations that affect the ocean s surface.

Natural climate variability includes in general, as it does in the ENSO example above, three types of phenomena.

1. Variations that are directly driven by a purely periodic external force, like the diurnal or the seasonal cycle of insolation, are the easiest to understand and predict. The diurnal variations are due to the rotation of Earth on its axis, the seasonal ones to its revolution around the Sun. They bring about the temperature and precipitation variations between day and night and between summer and winter, respectively. On much longer time scales, multiply periodic (often called quasi-periodic) variations in Earth s orbit around the Sun affect the intensity of the solar radiation that reaches the Earth s surface.
2. Variations due to the non-linear interplay of feedbacks within the climate system are harder to understand and to predict than the previous ones. For instance, a temperature drop within the system will increase the amount of snow and ice, and thus lead to further cooling; this is the so-called ice-albedo feedback, explored in some detail by energy-balance models (*see* Energy Balance Climate Models, Volume 1). On the other hand, the increase of trace-gas concentrations in the atmosphere, such as that of carbon dioxide (CO_2), will increase surface temperatures through the greenhouse effect (*see* Projection of Future Changes in Climate, Volume 1). This temperature increase will release even greater amounts of trace gases from the upper ocean or, on the contrary, trap them in terrestrial vegetation. Each climate feedback can enhance or countervail the effect of another feedback (*see* Climate Feedbacks, Volume 1). The distinct feedback mechanisms identified so far in the climate system are numerous and complex. Their number and complexity contribute significantly to the difficulty of reliably detecting human-induced climate change (*see* Climate Change, Detection and Attribution, Volume 1).
3. Variations associated with random fluctuations in physical or chemical factors are hardest to predict in detail for any length of time. These factors can be external to the climate system itself, such as aerosol loading due to volcanic eruptions (*see* Volcanic Eruptions, Volume 1). They can also be internal to the system, such as weather fluctuations. The latter are known to be unpredictable on the longer time scales of seasons

to millennia. Still their averaged effect over these time scales may result in heat-transport variations between the equator and the poles.

The periodically driven variations – in the absence of any other source of variability – are themselves purely periodic and thus highly predictable. A single sudden change imposed on the system, whether in the solar irradiance received at the top of the atmosphere, in the atmosphere s reflectivity or opacity, or in other forcings or parameters of the system, has a relatively simple effect, conceptually speaking, on such a periodic variation.

A sudden impulse, as described in the previous paragraph, can change, for example, either the mean about which hemispheric temperature oscillates or the amplitude of the oscillation or both. Such a change would thus be easily detectable, speaking again in relative terms, when compared with one that a sudden impulse would cause in the other two types of variations, those resulting from the interplay of non-linear feedbacks (Lorenz, 1963) or those produced by stochastic forcing (Hasselmann, 1976). The characteristics of the latter two types of variations may be affected by an imposed change in diverse and complicated ways, ways that may be hard to distinguish from those of a spontaneous change within the system (*see* Chaos and Predictability, Volume 1).

An artist s rendering of climate variability on all time scales is provided in Figure 1(a). It is meant to summarize our knowledge of the relative power S, i.e., the amount of variability in a given frequency band, between f and $(f + \delta f)$; here frequency f is the inverse of period T, $f = 1 \bullet T$, and δf indicates a small increment. This power spectrum is not computed directly by spectral analysis from a time series of a given climatic quantity, such as (local or global) temperature; indeed, there is no single time series that is 10^7 years long and has a sampling interval of hours, as the figure would suggest.

Instead, Figure 1(a) represents a composite of information obtained by analyzing the spectral content of many different climatic series. The figure reflects the three types of variability mentioned earlier: sharp lines that correspond to periodically forced variations, at one day and one year; broader peaks that arise from internal modes of variability; and a continuous portion of the spectrum that reflects stochastically forced variations, as well as deterministic chaos. The latter represents the irregular variations that result from the deterministic interplay of non-linear feedbacks.

Between the two sharp lines at one day and one year lies the synoptic variability of mid-latitude weather systems, concentrated at 3–7 days, as well as intraseasonal variability, i.e., variability that occurs on the time scale of 1–3 months. The latter is also called atmospheric low-frequency variability. This name refers to the fact that the variability in question has lower frequency, or longer period, than the so-called baroclinic instability of large-scale atmospheric flow that gives rise to the development of weather systems. The periods associated with intraseasonal variability exceed even the duration of these weather systems complete life cycle, from their birth in the storm tracks off the east coasts of the major landmasses to their decay further to the east, across an entire ocean basin. Intraseasonal variability comprises phenomena such as the 40–50-day Madden-Julian oscillation of winds and cloudiness in the tropics, as well as the alternation between episodes of zonal and blocked flow in mid-latitudes. Both of these phenomena involve exchanges of angular momentum

between the atmosphere, the oceans and the land: as the winds speed up, the Earth rotation slows down, and vice-versa (Ghil and Childress, 1987; Ghil and Robertson, 2000).

Immediately to the left of the seasonal cycle in Figure 1(a) lies interannual, i.e. year to year, variability. This variability is dominated by ENSO-related phenomena and involves the interaction of the seasonal cycle with internal modes of variability of the ocean-atmosphere system in the tropical Pacific (Philander, 1990; Neelin *et al.*, 1998). Additional forms of interannual and interdecadal variability in higher latitudes are associated with the North Atlantic Oscillation, the Pacific Decadal Oscillation, and the Arctic Oscillation (*see* Arctic Oscillation, Volume 1; North Atlantic Oscillation, Volume 1; Pacific Decadal Oscillation, Volume 1).

The emphasis of climate research in the second half of the 20th century has shifted farther and farther to the left in the composite spectrum of Figure 1(a). It has proceeded from the study of weather systems in the 1950s and 1960s to that of intraseasonal variability in the 1970s and 1980s and on to interannual variability in the 1980s and 1990s. The greatest excitement – among scientists as well as the public – is currently being generated by interdecadal variability, namely climate variability on the time scale of a few decades, i.e., the time scale of an individual human's life cycle.

This progression of interest towards longer and longer time scales is due to two complementary causes. First, the non-linear feedbacks between different components of the climate system that affect the longer-term interactions pose a greater challenge to the scientific community. Second, the greater scientific challenge goes hand in hand with humanity's increasing desire to predict and eventually control its environment farther and farther into the future (National Research Council, 1995).

Clearly, controlling our environment is predicated upon the ability to understand and predict it first. Thus, weather prediction can be performed with the same accuracy now for 3–5 days as it was for 12–24 hours in the 1960s, due to the very substantial progress in understanding and modeling weather phenomena over the last 40 years. We have learned to carry out predictions of certain climate variables over certain areas, with some useful skill, for up to six months in advance. This skill is a result of considerable advances in understanding ENSO and modeling coupled ocean-atmosphere interactions over the last 20 years. It is reasonable, therefore, to attack now the even more challenging problems of understanding and predicting climate variability for the next decade and century.

Prediction is greatly facilitated by regularity. Thus, the prediction of sea-surface temperatures on seasonal to interannual time scales is helped by the near-repetition of warmer temperatures in the eastern tropical Pacific every boreal winter, as well as roughly every second year, and every fourth or fifth year. The positive interference of the quasi-biennial mode (2–2.5 years) with the low-frequency mode (4–5 years) yields major warm and cold events, but the episodes of mutual reinforcement between modes still occur rather irregularly. This irregularity precludes, at present, greater certainty in predictions for 6 months or longer (Neelin *et al.*, 1998).

The power spectrum shown in Figure 1(b) is based on the longest instrumentally measured record of any climatic variable, the 335 year long record of Central England temperatures. It represents an up-to-date blow-up of the interannual to interdecadal portion of Figure 1(a). The broad peaks are due to the climate system's internal processes: each spectral component can be associated, at least tentatively, with a mode of interannual or interdecadal variability. Thus the rightmost peak, with its period of 5.2 years, can be confidently attributed to the remote effect of ENSO's low-frequency mode. The 7.7-year peak is currently believed to capture a North Atlantic mode of variability that arises from the Gulf Stream's interannual cycle of meandering and intensification (Moron *et al.*, 1998; Speich *et al.*, 1998). The two interdecadal peaks, near 14 and 25 years, are also present in global records, instrumental as well as paleoclimatic. They seem to be associated with oscillatory modes in the global oceans' thermohaline circulation (*see* Ocean Circulation, Volume 1) and its coupling to the atmosphere above (Ghil and Robertson, 2000).

Finally, the leftmost part of Figure 1(a) represents paleoclimatic variability. The information summarized here comes exclusively from proxy indicators of climate. These indicators include coral records and tree rings for the historic past, and marine-sediment and ice-core records for the last two million years of Earth history, known as the Quaternary. During the Quaternary era, large ice sheets were present on the planet, especially in its Northern Hemisphere. This era is therefore termed an ice age, when compared to such ice-free epochs of the past as the Cretaceous.

Ice ages occupy on the whole no more than about one tenth of documented Earth history. Their main interest lies in the fact that natural climatic variability is higher during an Ice Age than during more benign geological times. The Quaternary represents an alternation of warmer and colder episodes, called glaciation cycles. This cyclicity is manifest in the broad peaks present in Figure 1(a) between roughly 1 kyr and 1 Myr. Of these, the three peaks near 20 kyr, 40 kyr and 400 kyr reflect quasi-periodic variations in Earth's orbit. These orbital variations (*see* Orbital Variations, Volume 1) are due to the perturbing forces exerted by Jupiter and the other planets on Earth's otherwise purely periodic, Keplerian motion around the Sun. They represent, respectively, variations in precession, obliquity and eccentricity, three parameters that are used to describe Earth's secularly varying orbit.

The large 100-kyr peak continues to puzzle paleoclimatologists. It has been attributed variously to an internal mode

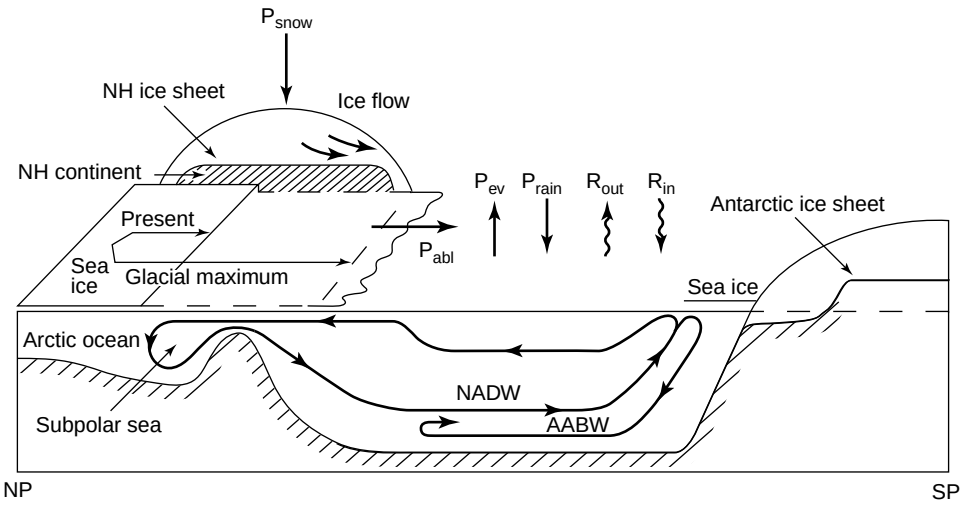


Figure 2 Schematic diagram of the Atlantic Ocean's thermohaline circulation and of the radiative, hydrologic and cryospheric processes likely to affect it (after Ghil *et al.*, 1987)

of the climate system, to the system s resonant response to small eccentricity variations or to a combination tone between orbital frequencies (Ghil and Childress, 1987). In the latter theory, the dominant 100-kyr peak corresponds to the non-linearly resonant amplification of a difference tone between the two distinct precessional lines at 19 kyr and 23 kyr. Certain climate simulations from the geological past attribute a critical role to natural changes in CO_2 levels, both in the initiation of the Quaternary as a whole and in generating the 100-kyr peak during the Late Pleistocene, i.e., the last one million years (*see* Climate Model Simulations of the Geological Past, Volume 1).

The peak at 6–7 kyr was predicted theoretically in the early 1980s as an internal oscillatory mode due to the interplay between the positive ice-albedo feedback and the negative precipitation-temperature feedback. In this theory, when coupling an energy-balance model with an ice-sheet model, global temperature drops as ice mass increases. The temperature drop diminishes the evaporation in low latitudes and hence the ice accumulation in high latitudes; this in turn decreases the ice extent and allows temperatures to rise again. Heinrich events (*see* Heinrich (H-) Events, Volume 1) were discovered in the late 1980s and found in the 1990s to have a near-periodicity of 6–7 kyr, as predicted by this theory (Ghil, 1994).

The peak at 2–2.5 kyr is associated with the names of W Dansgaard and H Oeschger and was discovered in Greenland ice cores (*see* Dansgaard- Oescheger Cycles, Volume 1). It seems to arise from the alternative intensification and weakening of the Atlantic s thermohaline circulation, as represented in Figure 2. Both the region of formation of North Atlantic Deep Water (NADW) and its flux (i.e., its mass transport per unit time) vary in a fairly irregular manner, but with a millennial-scale mean

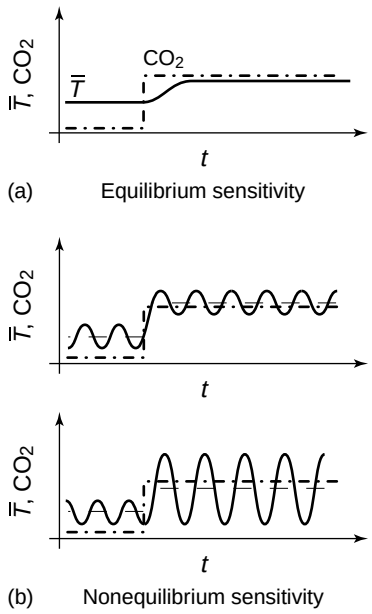


Figure 3 Climate sensitivity for (a) an equilibrium model and (b) a nonequilibrium, oscillatory model: as a forcing parameter (atmospheric CO_2 concentration, dash-dotted line) changes suddenly, global temperature (heavy solid) undergoes a transition. In (a) only the mean temperature (still heavy solid) changes; in (b) the oscillation's amplitude can also decrease (upper panel), increase (lower panel) or stay the same, while the mean (light dashed) adjusts as it does in panel (a)

periodicity. They are associated with variations in sea-ice cover, as well as in temperature and precipitation over the adjacent land areas.

What are the implications of the natural variability described so far on our understanding of global climate

Box 1 Computations of the climate's equilibrium sensitivity are often based on the linear, scalar ordinary differential equation

$$c\dot{\theta} = \lambda\theta + Q \qquad \bullet 1 \bullet$$

where θ is global temperature and

$$\dot{\theta} \equiv \frac{d\theta}{dt}$$

is its rate of change with time t.

Here

$$\lambda = \sum \lambda_i \qquad \bullet 2 \bullet$$

is the sum of the (linear) feedbacks λ_i, positive or negative, and

$$Q = \sum Q_j \qquad \bullet 3 \bullet$$

is the sum of the (positive or negative) forcings Q_j, also called source (or sink) terms (see Ghil and Robertson, 2000, and references therein). Typically, Q_j might include absorption by greenhouse gases (positive forcing) and reflection by aerosols (negative forcing).

If climate change during the industrial era were truly governed by simple linear mechanisms, according to Equations (1–3) above, it would be fairly smooth. Provided that only gradual increases in aerosols and greenhouse gases were at work, it would look schematically like the light solid line in Figure 4. In fact, the observed temperature variations in the Northern Hemisphere looked, again schematically, like the heavy solid line in the figure. The difference between the two curves, light and heavy, is due to the presence of natural variability, as well as to that of variable natural forcings, like volcanic eruptions and solar-irradiance variability.

This difference implies that the actual climate system's evolution is more complex than that dictated by a linear response to simple forcings. To simulate and predict this complex evolution requires the use of a set of non-linear (ordinary or partial) differential equations, depending on time t and on a set of parameters μ:

$$\dot{X} = F \bullet X \bullet t \bullet \mu \bullet \qquad \bullet 4 \bullet$$

Here X is a vector of state variables that describes the atmosphere, ocean and other climate subsystems. The components of the right-hand side vector F give the rates of change of the components of X. The vector F depends on the state X itself, time t (in the form of time-dependent forcing), and the parameter set μ. This set includes but is not limited to those present in the former scalar equation; typical parameters are those that determine the strength of various (non-linear) feedbacks and the net insolation at the top of the atmosphere.

Gradual changes in μ can lead to sudden changes in system behavior, from steady state to purely and multiply periodic, and on to chaotic and fully turbulent; these changes are called bifurcations. As a result, the system's predictability can change dramatically.

change? If no variability whatsoever were present, i.e., if the climate system were in equilibrium, external changes would result in a simple shift of this equilibrium (see Figure 3a). The ratio of the amount of this shift, in global mean temperature say, to the equivalent change in irradiance at the top of the atmosphere say, is usually defined as climate sensitivity (*see* Climate Sensitivity, Volume 1). The actual forcing change under consideration might be in carbon dioxide concentration or in atmospheric opacity due to aerosol loading or in the mean distance between Earth and Sun (see Box 1 for details).

Figure 3(b) shows how the response of the system would be modified if the system were undergoing periodic oscillations, either as the result of purely periodic forcing or due to an internal mode that is purely oscillatory. In this case, a shift in the mean would be accompanied by a change in the oscillations amplitude, increasing or decreasing it. In fact, for a non-linear system, a change in an external parameter can lead from the system s being in equilibrium to its undergoing self-sustained oscillations, i.e., as a result of global change the system might be destabilized and become oscillatory.

The climate system s behavior, however, is much more complicated than being in equilibrium or in a state of purely periodic oscillations. Thus, the effects of natural or anthropogenic changes in the system s forcing or parameters, including but not restricted to net insolation, cannot be measured by a single quantity like climate

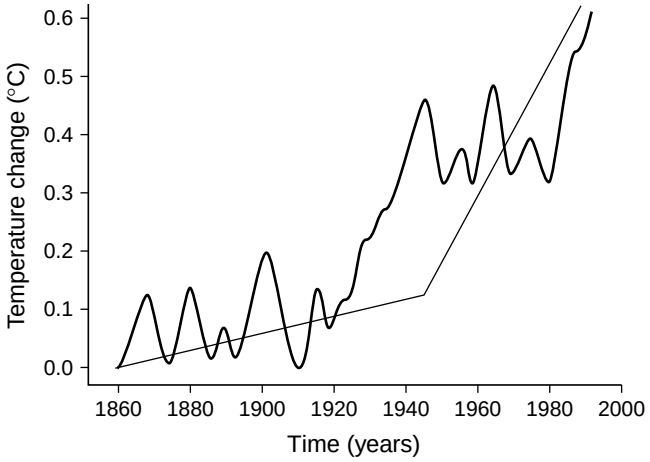


Figure 4 The role of natural variability in climate-change detection and attribution: equilibrium response of global temperatures to changes in aerosols and trace gases (light solid line) vs. observed temperature variations (heavy solid)

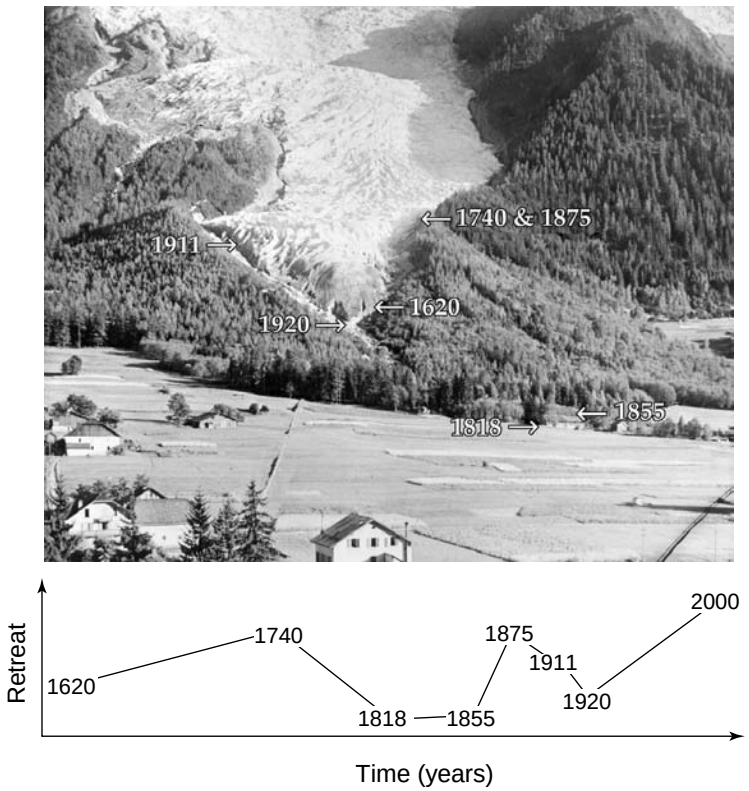


Figure 5 Retreats and advances of an alpine glacier over the last 300 years (upper panel courtesy of Musée Alpin–Les Amis du Vieux Chamonix)

sensitivity. Resonances may lead to the amplification of certain oscillatory modes and the entire behavior of the system may change, becoming more or less predictable. The situation we are confronted with is illustrated schematically in Figure 4. The figure highlights the discrepancy between a simple response to human-induced forcings and the temperature variability recorded during the last century-and-a-half (see Box 1 for explanations).

A simple example of the difficulties in assessing human-induced climate change in the past, and hence predicting it for the future, is given in Figure 5. Valley glaciers in France s Chamonix Valley, and elsewhere in the Alps, have been retreating recently. This retreat, however, is far from unprecedented, as the figure shows. Is the present retreat reversible – as it was in the latter part of the 18th century or the early part of the 20th – or is it not?

REFERENCES

Ghil, M (1994) Cryothermodynamics: The Chaotic Dynamics of Paleoclimate, *Physica D*, 77, 130–159.

Ghil, M and Childress, S (1987) *Topics in Geophysical Fluid Dynamics: Atmospheric Dynamics, Dynamo Theory and Climate Dynamics*, Springer-Verlag, New York.

Ghil, M, Mullhaupt, A, and Pestiaux, P (1987) Deep Water Formation and Quaternary Glaciations, *Clim. Dyn.*, 2, 1–10.

Ghil, M and Robertson, A W (2000) Solving Problems with GCMs: General Circulation Models and their Role in the Climate Modeling Hierarchy, in *General Circulation Model Development: Past, Present and Future*, ed D Randall, Academic Press, San Diego, CA.

Hasselmann, K (1976) Stochastic Climate Models, Part I: Theory, *Tellus*, 28, 473–485.

Lorenz, E N (1963) Deterministic Nonperiodic Flow, *J. Atmos. Sci.*, 20, 130–141.

Moron, V, Vautard, R, and Ghil, M (1998) Trends, Interdecadal and Interannual Oscillations in Global Sea-surface Temperatures, *Clim. Dyn.*, 14, 545–569.

National Research Council (1995) *Natural Climate Variability on Decade-to-century Time Scales*, eds D G Martinson, K Bryan, M Ghil, M M Hall, T R Karl, E S Sarachik, S Sorooshian, and L D Talley, National Academy Press, Washington, DC.

Neelin, J D, Battisti, D S, Hirst, A C, Jin, F-F, Wakata, Y, Yamagata, T, and Zebiak, S E (1998) ENSO Theory. *J. Geophys. Res.*, 103, 14 261–14 290.

Philander, S G H (1990) *El Niño, La Niña, and the Southern Oscillation*, Academic Press, San Diego, CA.

Plaut, G, Ghil, M, and Vautard, R (1995) Interannual and Interdecadal Variability in 335 years of Central England Temperatures, *Science*, 268, 710–713.

Natural Hazards

see Natural Hazards (**Volume 3**)

Natural Records of Climate Change

John S Perry[1] and Thompson Webb, III[2]
[1]*Alexandria, VA, USA*
[2]*Brown University, Providence, RI, USA*

Climate scientists study records of past climates to better understand natural climate variations and the signi cance of human-induced changes to global and regional climate. Of great importance, data on past climatic variations provide information about a rich array of natural experiments from which insights can be gained into the mechanisms that determine and can change the climate. Data on past climates can also be used to validate explanatory and predictive models of the climate system.

Instrumental records provide information about the climate for only the past century and a half, and for much of this time, station data provide uniform and dense coverage over only a few subcontinental areas. Because of these limitations, the earlier climatic (i.e., paleoclimatic) record must be pieced together by other means. Fortunately, many organisms and natural processes are affected by the climate and have left records that can be deciphered and calibrated to provide proxy records of climate. Historical records also contain many direct and indirect observations about climate. By taking advantage of these non-instrumental sources of climate information, research in many areas of the Earth and life sciences has led to an increasingly comprehensive and reliable base of information describing the history of Earth's climate.

INTRODUCTION

If the Earth had from its birth been covered with a dense network of modern automatic weather stations (perhaps operated by the Martian Weather Service), then the study of past climates would be a simple exercise in arithmetic and bookkeeping. However, such is far from being the case. Even today, with satellites peering at the Earth from all angles, its surface is only incompletely and imperfectly observed. Frustratingly large blank areas persist in global maps of weather observations. As we look back in time, these blanks rapidly become larger, and the coverage of reliable quantitative instrumental data rapidly shrinks to a sparse network of cities and sailing routes. If we seek to look back some hundreds of years, we soon precede the invention of thermometers and barometers and find only a handful of careful observational records from China and the Greco-Roman Empire.

The reconstruction of past climates therefore cannot rely on official observers who used thermometers and carefully filled out observation forms. Other sources of information have to be found and utilized. Through clever analysis, natural records that have long been accumulating can be used to reconstruct past conditions. Rocks have eroded under rain, wind, and ice sheets and glaciers. Dunes, wadis (dry streambeds), and lake deposits grew or waned as deserts advanced and retreated. Glaciers and rivers flowed or shrank. Plants and animals and their associated communities flourished, died, evolved, or migrated as the climate became more or less favorable for their growth. Even atmospheric composition has changed; for example, oxygen was produced as life first evolved (which in turn allowed the UV-absorbing stratospheric ozone layer to form) and then and later the amount of carbon dioxide and other greenhouse gases changed in concentration.

Many of the processes affected by and affecting past climates have left their traces in the world of today, much as even the most careful criminals leave clues at the scene of their crimes. Paleoclimatologists act much like detectives and forensic pathologists in seeking to reconstruct descriptions of past climates. Fortunately, many of the indirect clues left in the Earth s vast storehouse can be related quantitatively to climate variables. Such clues serve as substitutes for the unavailable instrumental records and have therefore been termed *climate proxies*. For example, the width and density of the annual rings in ancient wood yields information on past growth rates, which depended in turn on the temperature and precipitation when the ring formed hundreds and even thousands of years ago. Reliable quantitative data on the geographic distribution of plant, animal, and human populations of the past can also provide valuable insights into past climatic conditions. For example, during the Last Glacial Maximum 21 000 years ago, spruce trees like those growing in central Canada today grew in Kentucky and southern Ohio.

Paleoclimatic data can also aid in the interpretation of recent variations and changes in climate. For example, knowledge of the climate patterns that accompanied past warm periods can suggest how particular climatic features may change as temperatures rise, and may offer useful insights about the specific climate patterns that could accompany global warming (e.g., US/USSR, 1990). The responses of past animal and plant populations to times of rapid climate change may also indicate how rapidly they can adapt to change. Past climate states can also serve

as natural experiments for testing the reliability of the numerical models of the Earth system that are used to predict the future course of climate. Such natural experiments can be selected from times with higher and lower concentrations of greenhouse gases, altered values of seasonal solar radiation, lower and higher sea levels, and more and less abundant ice extent. The ability of models to explain the climates of the past also tells us something about their reliability as guides for projecting climates of the future (Joussame and Taylor, 2000).

The list of data sources employed as paleoclimate observations is long and growing (Ruddiman, 2000). Bradley (1999) broadly categorizes the potential sources of information as: (1) historical documents; (2) geomorphological features such as dunes, glacial moraines, and river terraces; and (3) cores from ice sheets, glaciers, marine sediments, lake sediments, peat deposits, soils, rocks, trees, cave formations, and corals as providing the natural archives in which data on past climates are preserved. Some of these archives, such as rocks, span billions of years with relatively coarse resolution whereas others, such as tree rings, corals, and annually laminated sediments, provide annual records over hundreds to thousands of years. Other articles in this volume touch upon a number of these sources of information; here, we include an illustrative list and provide some information about their interpretation and availability.

HISTORICAL INFORMATION

Historical documents (i.e., written records) contain a wealth of information about past climates and some of the factors (such as volcanic eruptions) that have caused them to change (Bradley and Jones, 1995). Observations of weather and climatic conditions can be found, among other sources, in ships logs, farmers diaries, travelers accounts, newspaper articles, court and church records, and other written material (Lamb, 1982). For example, the dates of ice breakup in harbors, lakes, and rivers have often been recorded for centuries, as have the dates of first and last frosts, flowering, and harvest (*see* Phenology, Volume 2). Tax records indicate when highland farms were flourishing, abandoned, or even covered by glaciers as some Norwegian farm fields were at times during the Little Ice Age extending roughly from 1450–1850 AD (Grove, 1988). Shipping records indicate both weather conditions and the market conditions for weather dependent crops. Reports on the brilliance of past sunsets, especially when available from many locations, can tell of the timing and approximate magnitude of major volcanic eruptions. Even paintings contain clues about the prevailing climate of their period, such as the Breughel paintings of ice covered canals in the Netherlands. When properly evaluated, historical data can yield both qualitative and quantitative information about past climate.

TREE RINGS

Tree rings provide an annual record of local climate during the life of the tree (Fritts, 1976). Many trees are hundreds of years old, and a few live thousands of years. Moreover, much ancient wood is preserved in arid areas, in the beams of buildings and ruins, or petrified, for example, in wetlands. The rings of a tree record the influence of the environment on its growth over the tree s lifetime. In a tree, the cells that will become wood or bark, the cambium, grow in a light layer during late spring/early summer, changing to a dark layer in later summer/early fall. A light and dark ring pair thus represents one year in temperate latitudes. These annual rings can be counted to provide an exact chronology, and because there is more growth under good conditions, the growth patterns can be studied to determine the conditions a tree lived through, such as forest fires, drought, insect attack, or floods. The rings of a tree may be studied in either a cross section of the trunk or a cut of a fallen tree, or in a core taken from a living tree. By matching tree ring sequences from a series of overlapping wood samples, continuous chronologies of climate can be developed extending many centuries to thousands of years into the past. For example, rings from Douglas fir and bristlecone pine yield a continuous record of over 8000 years in the western United States. Quantitative relationships are derived between recent tree ring variations and observed climate; these relationships are then employed to estimate past climatic conditions from ancient trees. In addition to their scientific value, such data have important practical applications because they enable climatologists and hydrologists to extend the instrumental record of precipitation and river flow to produce more representative and reliable statistics for design of dams and irrigation projects.

CORALS

Corals build their hard skeletons from calcium carbonate, a mineral extracted from seawater. Because of variations in growth rates related to temperature and cloud cover conditions, the skeleton formed in the winter has a different density than that formed in the summer. Thus, corals exhibit seasonal growth bands much like those observed in trees. The carbonate contains oxygen and the isotopes of oxygen, as well as trace metals, may be used to determine the temperature of the water in which the coral grew (Dunbar *et al.*, 1996). These temperature recordings can then be used to reconstruct climatic conditions during that period of time that the coral lived. Many corals grow only in shallow waters, and raised terraces of these coral deposits provide definitive indications of past sea levels in coastal regions, such as Barbados and New Guinea, although care must be taken to account for tectonic uplift (*see* Isostasy, Volume 1). The decay of natural radioactive

elements incorporated into the coral permits the age of coral to be determined. Because of the amount of information that can be derived from contemporary and fossil corals, much research attention is being devoted to such studies, and a network of coral cores from all around the world is being established. Data derived from corals has been particularly valuable in studying the variability of the El Niño/Southern Oscillation (ENSO) over time (Dunbar *et al.*, 1996).

FOSSIL POLLEN

Wind pollinated plants produce millions of durable pollen grains to allow for the inefficiency of the fertilization process, and many of these grains end up in lakes and bogs where they are well preserved in the sediments or peats. Palynologists collect cores of these sediments. The cores can be radiocarbon dated and treated with strong acids and bases to dissolve away the sediments, leaving a dated residue rich in pollen. Samples of the residue are then spread on microscope slides. The distinctive shape of each microscopic grain allows each to be counted and identified to the level of genus or family (and sometimes even the species). After 300 or more grains are counted in each sample, the pollen data are then used to estimate the percentage of each taxon present in the sample. From the types of pollen found in each sample, inferences can then be made about the vegetation that produced the pollen and the climatic conditions that allowed the vegetation to grow.

Time series of pollen data covering the last 10 000 or more years are available from all continents excepting Antarctica. The maps of past vegetation and climates constructed from these data show the changing patterns of vegetation since the Last Glacial Maximum for Europe and North America. These maps have been used to check the past climates simulated by climate models for each 3000–5000 year interval over the past at 21 000 years ago. For example, Huntley and Webb (1988) provide descriptions of the interpretive methods and vegetation history for Australia, New Zealand, Eurasia, and North America. Wright *et al.* (1993) used pollen data to reconstruct the climate changes of the past 21 000 years (i.e., since the Last Glacial Maximum) and also to check the results of climate model simulations of these climates.

ICE SHEETS AND GLACIERS

Moraines mark the extent of past ice sheets and glaciers, and, when accurately dated, can be used to construct the size and equatorward or downward extent of these flowing bodies of ice. The growth of the Laurentide and Scandinavian ice sheets to heights of 1–3 kilometers during the Last Glacial Maximum ultimately lowered sea levels by over 100 m. Although the successive advances and retreats of continental ice sheets can wipe away evidence of earlier glaciations, carved landscapes and the rebound rates of formerly glaciated lands can give indications of the mass, flow rates, and duration of such ice masses (Peltier, 1994).

ICE CORES

In mountain glaciers and polar ice caps, snowfall has been accumulating for centuries and millennia. The annually varying deposits of snow gradually become squeezed into thinner and thinner layers of ice as snow continues to accumulate. Layers of ash from dated volcanoes, lead deposited during Greek and Roman times, correlations with climatic variations recorded in other media, and other techniques (in addition to simply counting) can be used to confirm the ages of the various layers.

Drilling as much as 4 km downward through these ice masses, glaciologists have extracted ice cores that can be analyzed for clues about climatic conditions as far back as 420 000 years ago. These ice cores contain dust and tiny bubbles of air entrained in the snowfall from which the ice originated. The isotopic composition of the water making up the ice and the chemical composition of the entrained air can be used to reconstruct the past climatic conditions of the area and the past composition of the atmosphere. The ratio of oxygen isotopes in the ice indicates the temperature at which the snow was formed. The layers of ice and snow that make up glaciers and ice sheets thus form a natural archive of local, regional, and global environmental information (Bradley, 1999; Ruddiman, 2000).

Extensive drilling programs have been conducted in Greenland, Antarctica, and a number of mountain glaciers. Analyses of the ice and the material trapped in it indicate both natural and human made environmental variations over the time period during which the ice has accumulated. Ice core records now extend over time scales up to hundreds of thousands of years. Cores recording conditions going back thousands of years show how the concentration of carbon dioxide has increased from pre-industrial levels of 280 ppmv to current levels of over 360 ppmv, thus revealing the effects of human activities and the burning of fossil fuels. The core extracted at the Vostok station in Antarctica extends back 420 000 years, and reveals swings in temperature and atmospheric carbon dioxide concentration during the last four glacial/interglacial cycles. Such records thus provide powerful evidence for a linkage between carbon dioxide and climate (see Figure 1).

LAKE SEDIMENTS

Large amounts of sediment accumulate in lake basins of the world each year. Most of the sediments come from algae and other plankton that grow in the lake waters each year. Soil, leaves, seeds, and pollen grains also wash into

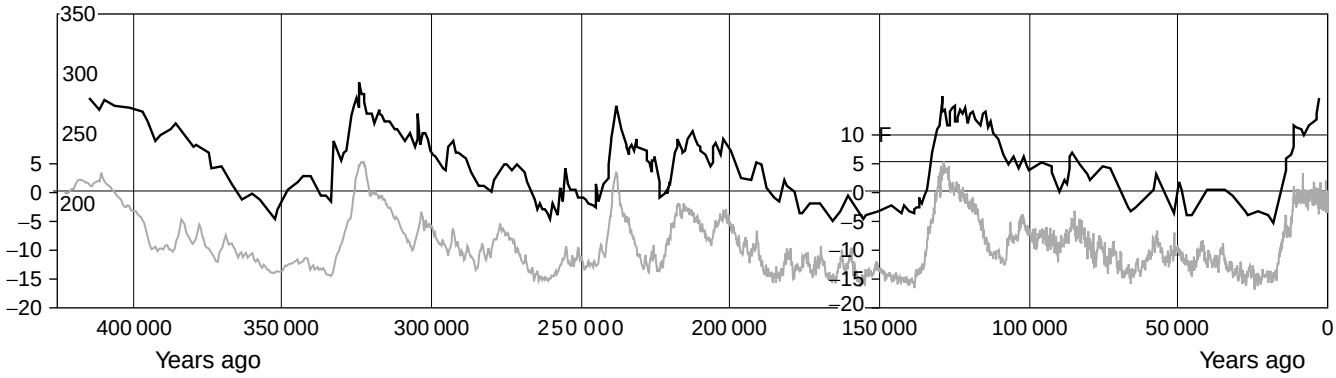


Figure 1 Changes in the global average concentration of carbon dioxide and the local surface air temperature have been reconstructed for the past 420 000 years using information derived from an ice core drilled at the Vostok station in Antarctica (Petit *et al.*, 1999). The local temperature record is derived from measurements of oxygen-18 isotope concentrations in the water frozen as snow. The record shows a series of long-term variations in the lower tropospheric temperature (above the inversion layer) that is similar to changes in solar radiation caused by changes in the Earth's orbit around the Sun. For most of past 420 000 years, temperatures in Antarctica (and by implication the globe) have been lower than recent values. Independent geological evidence indicates that glacial ice amounts peaked on Northern Hemisphere continents during these cold periods, most recently about 20 000 years ago. The very brief warm periods coincide with interglacial periods over the world's continents, with the Eemian interglacial of about 120 000 years ago being the last warm period until the present interglacial started about 10 000 years ago. In the absence of human influences on the climate, models of the advance and retreat of glaciers that include representations of changes in the Earth's orbit, natural variations in atmospheric composition, effects of climate change on land cover, sinking and rising of land areas due to the presence or absence of glaciers, and other factors suggest that the Earth would not return to glacial conditions for many thousands of years (Berger *et al.*, 1999). These studies also suggest that global-scale glaciation would be unlikely if the carbon dioxide concentration is above about 400 ppmv

rivers and lakes. These sediments accumulate millimeter by millimeter, forming a record of the history of environmental conditions in the watershed that feeds the lake. The high organic content of most lake sediments allows for radiocarbon dating back to 40 000 years or more. During the current interglacial of the last 11 500 years, mid-latitude lakes have accumulated sediments at average rates of 70 cm per 1000 years, which is about an order of magnitude faster than sediments are accumulating in the oceans.

In some lakes, the contents and color of sedimentary layers can vary with the seasons and striations are evident. These striations, called varves, are found in both contemporary lakes and in ancient lake sediments that have been converted into sedimentary rock. They allow for studies with annual resolution.

Lakes can often be viewed as giant rain gauges whose water levels record the changing moisture conditions of the watershed and region. Former shorelines can appear like bathtub rings around lakes that are now at lower levels than they were in the past. The Great Salt Lake in Utah is such a lake today. The changing shorelines indicate that the lake was many meters higher when Lake Bonneville existed during the last glacial period. Evidence of changes in the salinity of lakes is another clue to changes in moisture balance. Changes in the composition of diatoms and other plankton can also indicate moisture and water level changes. Continental maps of past lake level changes provide a strong test of climate simulations of past climates (Wright *et al.*, 1993).

OCEAN SEDIMENTS

Billions of tons of sediment accumulate on the seabeds of the world each year. Part of this rain of sediment consists of the remains of tiny zooplankton and algae that inhabit the near surface layers, and the excrement of other organisms that eat them. Just as on land, the particular mixture of plants and animals is dependent on the temperature and other variables, and some species even provide additional clues by, for example, changing which way they coil depending on the temperature. The remains of other bottom living (or benthic) organisms that derive sustenance from the downward flow of detritus can provide clues to deep ocean conditions, which in turn provides clues about the surface climate.

These sediment layers accumulate very slowly, sometimes only a few centimeters per thousand years, but they provide a record of Earth s climate history going back many millions of years. Because the accumulation rate is typically very slow, time markers are based on radiometric dating, on changes in isotope ratios (which can indicate, for example, the amount of glacial ice), and on the alignment of ferromagnetic particles that record reversals in the Earth s magnetic field (the last of which occurred about 700 000 years ago).

Ocean sediment records are particularly useful at deriving long histories of climatic conditions (*see* Earth System History, Volume 1). For example, the international Climate: Long-range Investigation, Mapping and Prediction (CLIMAP) project employed deep sea sediment cores to reconstruct the global pattern of sea surface temperature changes at the peak of the last glaciation, now dated at 21 000 years ago. Mathematical techniques were applied to derive quantitative estimates of surface temperatures from the relative proportions of different species of foraminifera, diatoms, and radiolaria in samples taken from the cores. Spectral analysis of the changes in marine plankton during the last 400 000 years was key to showing that variations in the Earth s orbit pace the glacial interglacial cycles (*see* Imbrie, John, Volume 1).

OTHER DATA SOURCES

This selection of data sources represents only a sampling of the numerous techniques employed to reconstruct the Earth s climatic history. For example, the sedimentary rocks that constitute much of the Earth s crust reveal much information about the chemistry and climate of the periods in which they were deposited. The rise and fall of global sea level as the great ice sheets waxed and waned is recorded in coastal terraces; for example, on Barbados, coral reefs grew during high sea stands and have left terraces as the ocean level fell. The fortuitous slow uplift of the island itself left these terraces as a sort of strip chart recording of glacial/interglacial history. Indeed, virtually any biological organism that has left discernible traces may be investigated as a proxy indicator of climate. For example, European caves inhabited by owls for thousands of years preserve on their floors layers of the tiny bones of the small rodents on which the owls subsisted. Changes in the relative numbers of various species have been correlated with changes in climatic conditions.

CONCLUSION

The billions of years of the Earth s history before the invention of thermometers left no giant filing cabinets of observations. However, the Earth and its inhabitants did leave for us a wealth of records of other kinds. These natural records of past climates vary greatly in their accuracy and resolution, and require great diligence and skill to acquire and interpret. However, they provide powerful searchlights into the obscurity of the past. They permit us to build remarkably detailed pictures of climates very different from our own; pictures that can help us to understand the processes determining the climate of today and to improve our capability to predict the climate of the future (see Ruddiman, 2000; Bradley, 1999; Crowley and North, 1991).

REFERENCES

Berger, A, Loutre, M F, and Melice, J L (1999) The 100 kyr Period in the Astronomical Forcing, Scientific Report 1999/6, Institut D Astronomie et de Geophysique G Lemaitre, Université Catholique de Louvain, Belgium.

Bradley, R S (1999) *Paleoclimatology: Reconstructing Climates of the Quaternary*, Academic Press, San Diego, CA.

Bradley, R S and Jones, P D (1995) *Climate Since AD 1500*, Routledge Press, London.

Crowley, T J and North, G R (1991) *Paleoclimatology*, Oxford University Press, New York.

Dunbar, R B, Linsley, B K, and Wellington, G M (1996) Eastern Pacific Coral Monitor El Niño/Southern Oscillation, Precipitation and Sea Surface Temperature Variability over the Past 3 Centuries, in *Climate Variations and Forcing Mechanisms of the Last 2000 Years*, eds P D Jones, R S Bradley, and J Jouzel, Springer-Verlag, Berlin, 373–405.

Fritts, H C (1976) *Tree Rings and Climate*, Academic Press, London.

Grove, J M (1988) *The Little Ice Age*, Methuen, London.

Joussaume, S and Taylor, K E (2000) The Paleoclimate Modeling Intercomparison Project, in *PMIP, 2000: Paleoclimate Modeling Intercomparison Project (PMIP)*, Proceedings of the Third PMIP Workshop, Canada, 4–8 October 1999, ed P Braconnot, WCRP Report WCRP-111, WMO/TD-No 1007, 9–24 (see also http://www-pcmdillnlgov/pmip/).

Lamb, H H (1982) *Climate History and the Modern World*, Metheun, London.

Peltier, W R (1994) Ice Age Paleotopography, *Science*, 265, 195–201.

Petit, J R, Jouzel, J, Raynaud, D, Barkov, N I, Barnola, J-M, Basile, I, Bender, M, Chappellaz, J, Davis, M, Delaygue, G, Delmotte, M, Kotlyakov, V M, Legrand, M, Lipenkov, V Y, Lorius, C, Pepin, L, Ritz, C, Saltzman, E, and Stievenard, M (1999) Climate and Atmospheric History of the Past 420 000 Years from the Vostok Ice Core, Antarctica, *Nature*, 399, 429–436.

Ruddiman, W F (2000) *Earth's Climate Past and Present*, W H Freeman, New York.

US/USSR (1990) *Prospects for Future Climate: A Special US/USSR Report on Climate and Climate Change*, eds M C MacCracken, M I Budyko, A D Hecht, and Y A Izrael, Lewis Publishers, Chelsea, MI.

Wright, Jr, H E, Kutzbach, J E, Webb, III, T, Ruddiman, W F, Street-Perrott, F A, and Bartlein, P J (1993) *Global Climates Since the Last Glacial Maximum*, University of Minnesota Press, MN.

Nitrous Oxide

Nitrous oxide (N_2O) is a radiatively and chemically active gas that is present in the atmosphere in trace amounts.

Nitrous oxide is a natural component of the atmosphere and is a primary source of nitrogen oxides in the stratosphere, where they have an important effect on the concentration of stratospheric ozone. Nitrous oxide is emitted to the atmosphere by a variety of biological sources in the water and soils and also is created as a by-product of atmospheric chemical processes involving the breakdown of ammonia (NH_4) (itself a biological product). Estimates of emissions from these natural sources are approximately 10 teragrams of nitrogen (TgN) per year, with an uncertainty of about 50%. Ice core records indicate that the pre-industrial concentration of nitrous oxide in the atmosphere was approximately 285 parts per billion by volume (ppbv), implying a total atmospheric burden of about 1380 TgN. Dividing the burden by the natural flux gives an atmospheric lifetime of about 140 years, which is in reasonable accord with estimates of the rate of chemical destruction, which give a lifetime in the range of about 120–150 years.

In addition to the natural sources of nitrous oxide, there are sources from human activities. The primary sources are estimated to be agricultural soils (largely as a result of fertilizer application), cattle and feedlots, industrial sources, and biomass burning. Total human-induced emissions are estimated to be about $7\,\mathrm{TgN\,yr^{-1}}$ from natural sources. Consistent with the augmentation of natural emissions, observations indicate that the atmospheric concentration of nitrous oxide has increased to about 315 ppbv. The concentration is slightly higher in the Northern Hemisphere than in the Southern Hemisphere, which is consistent with the human-induced sources being larger in the Northern Hemisphere. As fertilizer is applied more widely, and as climate changes soil conditions, the emissions from human activities are likely to change, and the newest emission scenarios generally project nearly constant or rising emissions during the 21st century. Because of the long lifetime, the nitrous oxide concentration is therefore projected to continue to rise due to human activities.

The increase in the nitrous oxide concentration of about 30 ppbv to date has created a radiative forcing of about $0.15\,\mathrm{W\,m^{-2}}$, which is about equivalent to the contribution from the most important of the chlorofluorocarbons and about 10% of the effect of the rising concentration of carbon dioxide (CO_2). Because nitrous oxide is so closely associated with the production of food and because it has a relatively small influence compared to carbon dioxide, the focus of efforts to limit greenhouse warming has been less intense on nitrous oxide, although it is included in the basket of gases created by the Kyoto Protocol. Nonetheless, there have been efforts to reduce industrial emissions of nitrous oxide as a contribution to the overall effort to limit global warming.

MICHAEL C MACCRACKEN *USA*

North Atlantic Oscillation

Christof Appenzeller
Swiss Federal Office of Meteorology and Climatology, Zurich, Switzerland

The North Atlantic Oscillation (NAO) is a large-scale uctuation in atmospheric mass and heat that strongly affects the climate of the North Atlantic sector and its surrounding continents. Its impact is strongest in the winter season. During a positive NAO phase the mean westerly winds are increased and the European winters are milder than normal whereas in the western Atlantic cold conditions prevail. In a negative phase the opposite occurs. A remarkable feature of the NAO is its long-term variability, with a tendency towards a more positive phase over the last 30 years. This upward trend in NAO accounts for a good part of the observed surface warming over Eurasia and also contributes to the Northern Hemisphere warming (Hurrell, 1996).

The NAO is defined in terms of an index (Hurrell, 1995) that represents a normalized pressure difference between the high latitude low pressure region centered near Iceland and the high pressure region in the subtropical Atlantic (Figure 1). A large pressure gradient between these two centers indicates stronger than average westerlies across the North Atlantic. This leads to increased flow of relatively warm and moist oceanic air towards the European continent, which in turn leads to warmer winter temperatures in Eurasia and colder conditions over the western Atlantic. Also wetter conditions occur over Iceland and Scandinavia and drier ones over Southern Europe. During negative phases of the NAO, the mild westerlies are reduced and Europe experiences a cold, more continental winter climate. The cold European winters of the 1960s as well as the relatively cold winter of 1995–1996 were associated with negative phases of the NAO, while the warm winters of the late 1980s accompanied positive phases.

NAO-like climate fluctuations can be observed in a large number of key environmental parameters and many of these show distinct changes in recent decades. Examples are changes in storm track intensity and location, in the temperature of the polar stratosphere and the strength of the Earth s ozone shield, in the intensity of ocean convection in the Labrador and Greenland–Iceland sea and its implication for the Gulf stream system, in sea ice cover over the northern North Atlantic and Arctic ocean, in retreat and advances of Scandinavian and Alpine Glaciers, in marine and terrestrial ecosystems such as the distribution of fish and the length of the growing season over Europe. See for example Uppenbrink (1999), and Marshall *et al.* (2000) for further information.

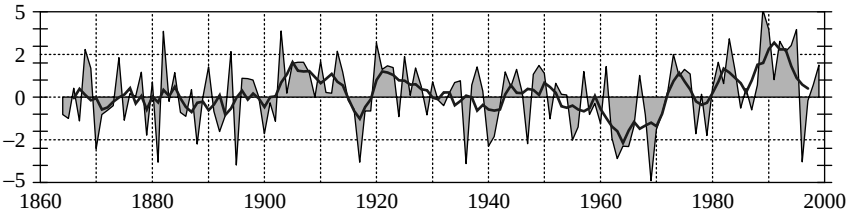


Figure 1 Winter NAO index based on the observed pressure difference between Iceland and Portugal. Data are a December to March average and each station is normalized relative to the period 1864 to 1983. The thick line is a 5 year running average. (Data provided by the Climate Analysis Section, NCAR, Boulder, CO, see also Hurrell (1995))

The NAO index shows both inter-annual and decadal fluctuations. A distinct trend towards more positive NAO values occurred in the decades from 1970 to 1990, with the highest value on the instrumental record occurring in winter 1988–1989. Such decadal climate variability can substantially mask or enhance a human-induced warming trend and can make the detection of the global warming more difficult. Reconstructed NAO indices using paleoclimatic indicators suggest that prolonged positive and negative NAO phases have occurred in the past (Appenzeller *et al.*, 1998). In addition the NAO index took a deep dive in the winter 1995–1996 although it has since returned to more average values. Hence, the fluctuations over the last decades could simply be due to natural variability. On the other hand, numerical as well as observational studies have shown that external climate forcing such as human-made greenhouse gases can affect natural climate variability in a non-linear way, e.g., Palmer (1999). If this is the case the positive NAO trend might be a consequence of global change.

Although the NAO seesaw of cold–warm winters between Iceland and Northern Europe has long been known, a comprehensive knowledge of the mechanisms responsible for this climate oscillation is still lacking. Some modeling and observational evidence supports the idea that the NAO can occur as a purely internal atmospheric process. While such variations could explain the observed coherent spatial structure, it is unlikely that they are a source of energy for the observed variability on multi-annual or longer time scales. Low frequency variability could also be occurring due to coupling of the atmosphere with the underlying ocean. There are clear indications that the temperatures at the surface as well as in the uppermost ocean layer are varying with the NAO. Coupled interaction with the wind-driven gyre circulation or the large-scale thermohaline circulation can also result in NAO-like variability. Hence, the sensitivity of the middle-latitude atmosphere to changes in surface boundary conditions, such as the sea surface temperature in the North Atlantic, the extent of sea-ice or some land property seems to have a controlling effect.

On the other hand, it has been suggested that the NAO should be considered as a sub-part of a more hemispheric pattern referred to as Northern Hemisphere annular mode (NAM) or Arctic oscillation (AO) (*see* Arctic Oscillation, Volume 1). These ideas point in particular to the role of the stratosphere within the climate system. If this is the case chemical and dynamical processes that affect the stratospheric circulation such as volcanic eruptions, ozone depletion or greenhouse gases might also be affecting the NAO long-term variability.

See also: El Niño and La Niña: Causes and Global Consequences, Volume 1; Natural Climate Variability, Volume 1.

REFERENCES

Appenzeller, C, Stocker, T F, and Anklin, M (1998) North Atlantic Oscillation Dynamics Recorded in Greenland Ice Cores, *Science*, 282, 446–449.

Hurrell, J W (1995) Decadal Trends in the North Atlantic Oscillation Regional Temperatures and Precipitation, *Science*, 269, 676–679.

Hurrell, J W (1996) Influence of Variations in Extratropical Wintertime Teleconnections on Northern Hemisphere Temperature, *J. Geophys. Res.*, 23, 665–668.

Marshall, J, Kushnir, Y, Battisti, D, Chang, P, Hurrell, J, McCartney, M, and Visbeck, M (2000) Atlantic Climate Variability, *Geophys. Rev.*, in press, copies available from *http://geoid.mit.edu/ACCP/avehtml.html*.

Palmer, T N (1999) A Nonlinear Dynamical Perspective on Climate Prediction, *J. Clim.*, 12, 575–591.

Uppenbrink, J (1999) The North Atlantic Oscillation, *Science*, 283, 948–949.

O

Observing Systems

see Earth Observing Systems (*Opening essay*, **Volume 1**); Monitoring Systems, Global Geophysical (**Volume 1**)

Observing Techniques, Ocean

see Ocean Observing Techniques (**Volume 1**)

Ocean Circulation

Lynne D Talley

Scripps Institution of Oceanography, University of California, La Jolla, CA, USA

Ocean circulation is one of the major elements of climate, moving heat, freshwater, nutrients and dissolved gases. Ocean circulation at the largest spatial scales and long time scales contains many elements that are in common in the different oceans. Understanding the processes that drive these different elements has progressed a long way and is essential for verifying climate models. The ocean circulation is forced (1) by the winds, through a thin frictional layer called the Ekman layer and through the response of the interior ocean to mass convergences and divergences in this layer, and (2) by heating and cooling that change the temperature and processes that change the amount of fresh water and hence the salinity. The first type of circulation is wind driven and the second is the thermohaline circulation.

The prevailing, large-scale winds of the Earth are trades (easterlies) in the tropics, westerlies at mid-latitudes. The wind-driven ocean circulation that results contains subtropical gyres around ocean high pressure regions and subpolar gyres around low pressure regions. These gyres are asymmetric, with strong western boundary currents (velocities of about $1\,m\,s^{-1}$*) that extend to great depth and relatively weak currents elsewhere. With increasing depth, each subtropical gyre shrinks poleward and westward towards its western boundary current. Shallow eastern boundary currents are created by local upwelling along eastern boundaries, driven by equatorward winds. In the tropics, the trade winds drive westward surface flow and create a strong and very thin eastward-flowing undercurrent on the equator. At the latitude of Drake Passage, between South America and Antarctica, the ocean is open all the way around the Earth. The westerly winds here create the strong, deep-reaching eastward flowing Antarctic Circumpolar Current.*

Except in the strong western boundary currents and the Antarctic Circumpolar Current, the deep flow is driven by thermohaline forcing, in which dense waters are made in isolated locations at high latitudes and spread through the oceans via relatively strong deep western boundary currents, which differ in mechanism and often direction from the wind-driven western boundary currents. Abyssal flow away from the deep western boundary currents is dominated by topography, and is often found to move parallel to the topography, with low pressure in the center.

Ocean currents transport heat from the tropics to the poles. Most of the heat is lost at mid-latitudes, where vigorous western boundary currents bring warm water to the latitude of cold, dry air outbreaks from the continents. This part of the heat transport is associated with the shallowest, wind-driven part of the subtropical gyre circulations (Talley, 1999), where the thermohaline circulation, forced by heating and cooling, is overshadowed by the wind-driven circulation. The deep thermohaline circulation is asymmetric, with deep water formation only in the northern North Atlantic and its adjacent seas and along the margins of Antarctica. No deep water is formed at the sea surface in the North Pacific or northern Indian Ocean. The global thermohaline circulation thus consists of two intersecting cells, one with sinking in the North Atlantic and the other with sinking in the Antarctic.

INTRODUCTION

The circulation of the ocean has been of practical interest to fishermen, traders and navies for as long as humans have gone to sea. As knowledge about ocean currents and capabilities to observe it below the surface grew, curiosity about currents below the sea surface resulted in increasingly detailed descriptions of the ocean as a three dimensional fluid. These observations have led to an improved understanding of the processes that govern the circulation. In this day of concern about the Earth s climate and our burgeoning ability to understand and model it with ever better computer resolution and power, knowledge of ocean circulation is as essential as knowledge of atmospheric circulation and the biosphere. In this article, the basic circulation patterns as we understand them now are described briefly including some historical perspective on this knowledge, followed by a section describing the physical processes that we understand to govern the circulation. In the last section, ideas about the role of ocean circulation in climate are presented.

Fluid flows, including those of the ocean circulation, are continuous and turbulent, and contain a wide range of behaviors from very small waves to the global circulation that is described here. Thus approach them in a simplified way, focusing on the time and space scales for considering the potential for and character of global environmental change. What is described here is often called the *general circulation*, meaning that it is the circulation at the largest space scales and longest time scales. Most elements of the general circulation are nearly unchanging in time within a given major climate regime; while given currents may be somewhat stronger or weaker in different seasons or years, or found in a slightly different location, the currents are always present in some form. Thus, we can hope to understand the underlying reasons for their existence in the present interglacial state and therefore be able to surmise the circulation patterns in other major climate regimes. Based on our understanding of the oceans, we have learned that, even with the massive climate shifts associated with glacial/interglacial switches, the basic elements of the ocean circulation would not have changed, with western boundary currents, wind driven gyres, eastern boundary currents, tropical circulation, and thermohaline circulation driven by dense water formation. However, strengths, depths, and horizontal positions of ocean currents would have shifted with large-scale shifts of wind patterns and changes in the relative strength of thermohaline forcing in each ocean basin. The circulation with quite different ocean geometry resulting from shifting continents can also be surmised given understanding of the processes that govern the present circulation.

The space scales of the ocean circulation range from about 100 km, which is the width of a strong current like the Gulf Stream off the east coast of North America, to several thousand kilometers, which is the width of the gyres that extend across each ocean basin, to global scale. Speeds of horizontal currents range from less than a centimeter per second in some areas of the deep ocean and far from the ocean sides to about $100\,\mathrm{cm\,s^{-1}}$ ($1\,\mathrm{m\,s^{-1}}$) in the strongest currents such as the Gulf Stream. Because the ocean depth averages around 5 km, and is at most about 10 km, the vertical circulation is limited and is in fact so much slower than the horizontal circulation that it is almost impossible to measure. Nevertheless, this very weak vertical circulation can be important for connecting the layers of the ocean, particularly for the circulation forced by water density changes.

Ocean currents are geostrophic (see section on processes below), like the large-scale atmospheric patterns that are associated with high and low pressure centers (*see* Atmospheric Motions, Volume 1). The force due to the difference in pressure pushes water from high to low pressure. However, because the Earth rotates, the water turns to the right in the Northern Hemisphere, and therefore circulates clockwise around high pressure centers and counterclockwise around lows; this tendency is due to the Coriolis acceleration (*see* Coriolis Effect, Volume 1). In the Southern Hemisphere, the water turns to the left and hence flows counterclockwise around lows, etc. High pressure at the sea surface in the ocean is due to water being piled up there, so the sea surface is slightly higher, up to one meter, than in the low pressure regions. Most surface height differences that drive the ocean circulation are much smaller, on the order of 5–10 cm. Spatial atmospheric pressure variations are too small to drive ocean circulation.

The ocean s circulation is forced almost exclusively at the sea surface by the winds and by changes in water density resulting from heating, cooling, evaporation and precipitation. These forcings create the high and low pressures through flows that are not geostrophic. We refer to the first type of forcing as *wind forcing*. The second is often called *buoyancy forcing*, because making water denser makes it less buoyant, or *thermohaline forcing*, because it results from changes in temperature and/or salinity (*see* Thermohaline Circulation, Volume 1). The ocean circulation can be reasonably well separated into portions that result from wind and from thermohaline forcing. Wind driven circulation is mainly confined to ocean basins while some parts of the thermohaline circulation extend from one end of the ocean to the other. Wind forcing dominates the surface circulation and creates the strongest currents, carrying the largest volumes of water, but thermohaline circulation dominates in the deep ocean. The wind driven and thermohaline circulations are of course connected with each other; a given water parcel will be subject to both forcings. The elements of the circulations are described in the next two sections. Our understanding of how the circulation is forced is described in more depth in the section on physical processes.

SURFACE CIRCULATION

The circulation at the ocean s surface is mainly produced by the winds. Mapping attempts have been made throughout the history of navigation, with a major growth in description and insight in the 17th century in the wake of global navigation. An excellent review, replete with reproductions of the significant charts of surface circulation, can be found in Peterson *et al.* (1996). The first to make the connection between winds and ocean currents was the German physician Varen (1622–1650), followed by the first global mapping and description of mid-latitude subtropical gyres in each ocean by the Dutch scholar Vos (1618–1689). Production of various schemes of surface circulation blossomed in the 19th century, with various degrees of realism, following invention of the chronometer and ability to determine longitude. Sobel (1995) provides a popular account of the British government s quest to solve the problem of measuring longitude, and the nearly unrewarded success of British clockmaker John Harrison in his lifelong work on the marine chronometer, a precise clock that worked even on moving and rolling ships.

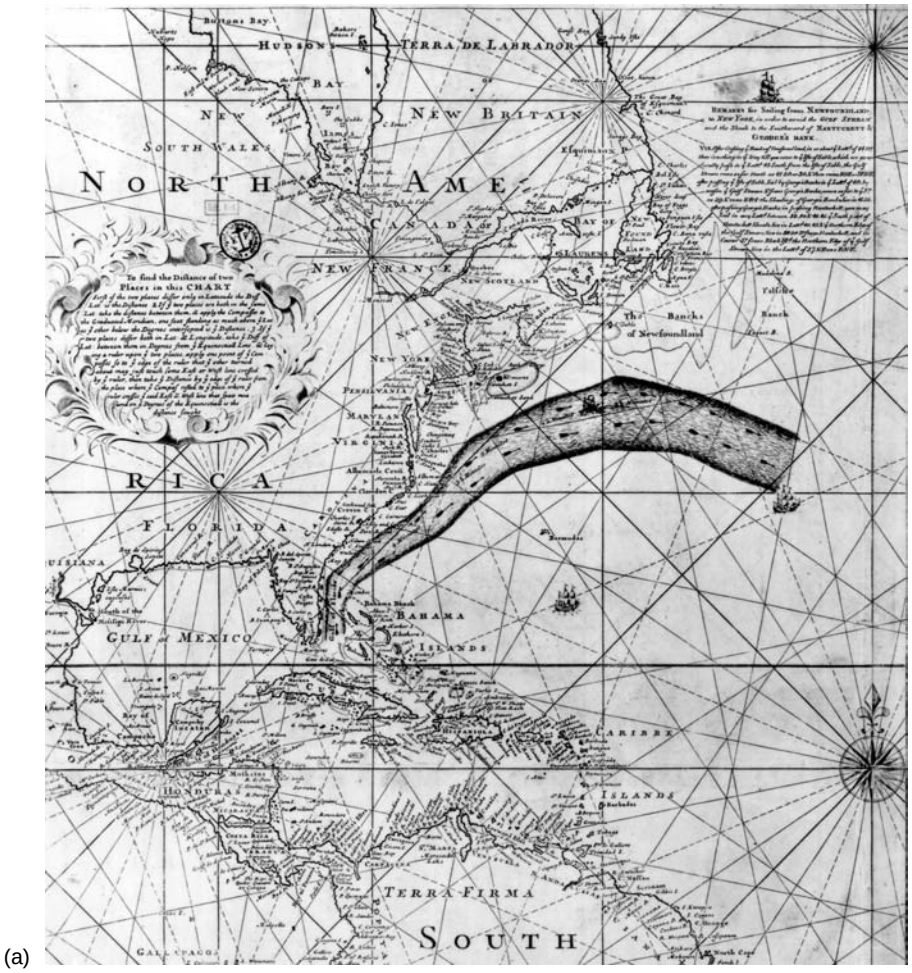
The strongest currents that are part of the general circulation are about $1\,\mathrm{m\,s^{-1}}$, or about 2 knots ($1\,\mathrm{knot} = 1$ nautical mile $\mathrm{h^{-1}}$). Tidal currents can exceed this, but are not found at this strength in the open ocean (*see* Tides, Oceanic, Volume 1). Currents of this strength can impact the progress of ships. An early map of the Gulf Stream (Figure 1a) was produced by Benjamin Franklin for the mail service from Britain to America, based on observations by his cousin Captain Timothy Folger, who was a whaling captain based in Nantucket. The original map was unearthed in an archive in Paris (Richardson, 1980) and is remarkably accurate, unlike a miscopied version that had been widely circulated and which placed the Gulf Stream in the wrong location (*see* Franklin, Benjamin, Volume 1).

The surface circulation for the globe has been summarized in schematic form by Schmitz (1996), and is reproduced here with some changes (Figure 2). Each of the largest ocean basins (North and South Atlantic, North and South Pacific, Indian) has a subtropical gyre, in which the currents circulate clockwise in the Northern Hemisphere and counterclockwise in the Southern Hemisphere. These gyres extend all the way from the western to the eastern boundary. The subtropical gyres are driven by westerly winds at mid-latitudes and easterly trade winds equatorward of about 30 °latitude. The communication of this forcing to the general circulation is roundabout and is described in the section on physical processes below.

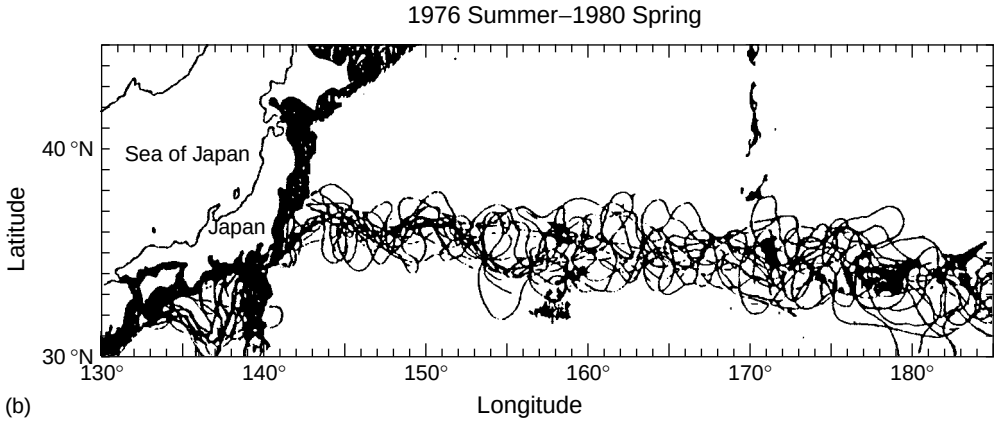
The subtropical gyres are asymmetric; the currents at the western boundaries are much stronger in all oceans than anywhere else in the gyres. These western boundary currents are the fastest currents in the ocean circulation. (Some tidal and strait flows are faster.) The western boundary currents include the Gulf Stream (North Atlantic), the Kuroshio (North Pacific), the Brazil Current (South Atlantic), the East Australia Current (South Pacific) and the Agulhas Current (Indian). All of these subtropical western boundary currents carry warm water away from the tropics towards the cooler regions at mid-latitudes; each loses a tremendous amount of heat during the process. In fact, the ocean s total heat loss is dominated by the losses in these western boundary currents. (The ocean s total heat gain is dominated by input from the Sun in the tropics.) As can be seen from the Franklin Gulf Stream chart (Figure 1a) and the modern satellite image of the Gulf Stream (Figure 1c), the Gulf Stream separates from the coast at North Carolina and moves out into the open ocean, remaining a strong current for a long distance out to sea. This behavior is characteristic of all of the other western boundary currents as well, in that all of them flow strongly along the boundary, then separate and flow offshore as strong and narrow currents for about 1000 km.

Western boundary currents are approximately 100 km in width, even after they separate from the coast and flow out to sea. After the currents leave the coast, they retain their width, but they also meander widely before they finally lose their energy. The meanders fill a wide swath, which for the Gulf Stream is almost exactly the envelope shown in the Franklin/Folger chart (Figure 1a). The meandering paths of the Kuroshio are shown in (Figure 1b). The meanders are about the same length and fill about the same envelope most of the time but are very dynamic, with a time-scale for change of two to four weeks. The meanders pinch off regularly both north and south of the current axis to form rings, which are 100–200 km across (e.g., Gulf Stream rings seen in Figure 1c). The rings that form on the north side of the current contain water from the south, warm side of the current and so are called *warm core rings*, while rings on the south side contain cold water and are called *cold core rings*. The rings generally migrate westward and sometimes rejoin the current.

All of the western boundary currents extend to the ocean bottom when they are along the boundary. The western boundary is not a vertical wall but is rather the continental slope between the land and the abyssal ocean. Before the separation point, the boundary current rides along the slope, not in the deepest water. After separation, the boundary currents flow out into the deeper part of the ocean and many of them then extend down to the ocean bottom at 4000–6000 m due to their tremendous energy. The highest speeds are always at the ocean surface, with decay to very low speeds at depth, reflecting the source of the western boundary currents as part of the wind driven circulation.



(a)



(b)

Figure 1 (a) Map of the Gulf Stream, produced by Benjamin Franklin based on information from Captain Timothy Folger, and rediscovered in a Paris archive by Richardson (1980). (b) Meanders of the Kuroshio after it separates from the Japanese coast (Mizuno and White, 1983). (c) Shown as a bright red band, the Gulf Stream is about 27 °C (~80 F) in this sea surface temperature image of the western North Atlantic during the first week of June 1984. There is a large temperature difference between the Gulf Stream and the surrounding waters and so the current and it meanders and rings are visible in sea surface temperature. This image is based on data from NOAA-7 Advanced Very High Resolution Radiometer (AVHRR) infrared observations. Warmer hues denote warmer temperatures. (Courtesy of O. Brown, R. Evans and M. Carle, University of Miami, Rosenstiel School of Marine and Atmospheric Science, Miami, Florida.)

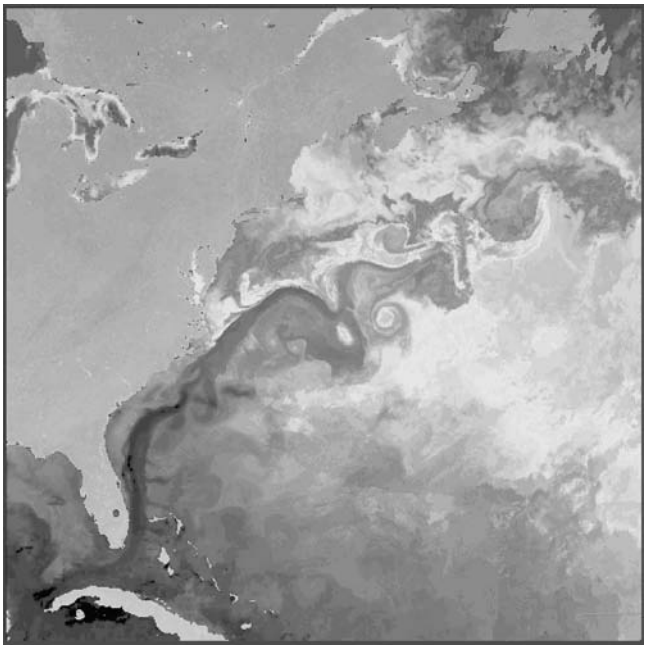
(c) 

Figure 1 (*Continued*)

While every ocean has a subtropical gyre with a western boundary current, each western boundary current has its own peculiarities, arising from the shape or existence of the western boundary and its intersection with the part of the wind pattern that dictates the separation point (see the section on physical processes below). The Brazil Current of the South Atlantic is the most canonical, following a single boundary until the separation point. The North Pacific s Kuroshio is nearly as simple, although it is complicated along the western boundary because of the numerous island chains and marginal seas. In the North Atlantic, there are actually two subtropical western boundary currents: the Gulf Stream, which we have discussed, and the North Atlantic Current, which is in a sense a continuation of a portion of the Gulf Stream, with the coast of Newfoundland as its western boundary because the winds dictate that the subtropical gyre should extend as far northward as Newfoundland. In the South Pacific, the East Australia Current flows southward to the latitude of New Zealand and then shoots straight eastward across the Tasman Sea to the northern tip of New Zealand, where it then recommences its southward flow for a short distance as a western boundary current before separation. In the Indian Ocean, the Agulhas flows southward to the Cape of Good Hope and would keep going if Africa extended farther south because the wind pattern does not dictate a separation until farther south. So the Agulhas essentially runs past the western boundary and makes a tremendous loop south of Africa back towards the east into the Indian Ocean. This is called the Agulhas retroflection. The retroflection is highly unstable and large rings of Indian Ocean water are produced at the place where it curves back. These rings move westward into the South Atlantic and are a major source of Indian Ocean water for the Atlantic Ocean.

The strength of the western boundary currents is related to the strength of the wind forcing that drives the subtropical gyres. A useful measure of the strength of a current is the total amount of water that it transports, which means looking for the maximum depth and horizontal extent of the current and computing the net flux of water. The transport is the sum (integral) of all of the velocities through a chosen area. The usual unit for ocean transport is the Sverdrup, where 1 Sverdrup $= 1 \times 10^6\,\mathrm{m^3\,s^{-1}}$, which is velocity times area. Subtropical western boundary currents typically carry 60–100 Sverdrups. All of this water must be returned equatorward in the interior of the ocean, away from the boundary, and then rejoin the western boundary current.

Most of the area of the subtropical gyres is occupied by the broad, slow eastward flow on the poleward side of the gyre, broad, slow equatorward flow all the way across the ocean, and broad, slow westward flow on the equatorward side of the gyre. The eastward flows are sometimes referred to as the west wind drift; in the Northern Hemisphere they are also referred to as the North Atlantic Current and the North Pacific Current. The westward flows in both Northern Hemisphere oceans are called the North Equatorial Current, not because they are along the equator but because they are on the tropical side of the gyres. The westward flows in the Southern Hemisphere oceans are called the South Equatorial Current in each ocean. The eastward and westward slow flows are divided at the sea surface by a narrow front called the Subtropical Front, which is at around 30 °latitude.

An important feature of the upper ocean subtropical circulation, called *subduction*, is associated with the density structure of the ocean, in which density increases with depth, and surface density increases with distance from the equator, that is, surface water is warmest in the tropics and colder at higher latitudes. When the warm surface water flows poleward in the western boundary current, it is subjected to intense cooling. The water that emerges from the separated western boundary current is then cooler than the water to the south. This emerging water circulates back towards the south, but when it does so, it encounters warmer surface water. The warming rate at the surface is not large enough to change the temperature and density of this equatorward flow, and so it submerges beneath the lighter surface layers. This process occurs for all waters that enter the subtropical gyres and that must make their way back towards the tropics. The submerging process was named subduction by Luyten *et al.* (1983) who were the first to fully describe the process and provide a theory for it.

Each subtropical gyre has an eastern boundary current as well as a western boundary current. All eastern boundary currents are significantly weaker and much shallower than western boundary currents. In snapshot observations, eastern boundary currents often resemble a string of highly time dependent eddies, with weak flow moving around and between the eddies. Eastern boundary currents are important for the ocean s biological productivity because they are driven by winds that cause upwelling of nutrient rich waters from about 100 m depth, below the sunlit euphotic zone where most ocean life occurs. The upwelled water is also cool, and is responsible for foggy coastal conditions in the eastern boundary regions. The simple theory for eastern boundary currents is included in the section on physical processes below. The eastern boundary currents are: Canary Current (North Atlantic), California Current (North Pacific), Benguela Current (South Atlantic), Peru Current (South Pacific) and the Leeuwin Current (Indian). Alone among the eastern boundary currents, the Leeuwin Current flows poleward, against the flow of the subtropical gyre.

A second major feature of the surface circulation that occurs in most ocean basins is a subpolar gyre. These are counterclockwise circulations found in the North Atlantic and North Pacific, and clockwise circulations in the Weddell

Sea (Atlantic sector of the Antarctic) and Ross Sea (Pacific sector of the Antarctic). The subpolar gyres are driven by westerly winds that are strongest at the subtropical/subpolar gyre and weaker at higher latitudes. The Northern Hemisphere subpolar and subtropical gyres are separated from each other by a narrow front, called the Subarctic or Subpolar Front. The Antarctic region is open to ocean circulation all the way around Antarctica, at the latitude of the maximum westerly winds. A major current, the Antarctic Circumpolar Current, flows eastward around Antarctica at this latitude. Its maximum speeds are about half those of the Gulf Stream. The subpolar circulations in the Southern Hemisphere are south of the Antarctic Circumpolar Current, and are separated from the subtropical circulations by the two main fronts of the Antarctic Circumpolar Current, called the Subantarctic and Polar Fronts (*see* Southern Ocean, Volume 1).

The subpolar gyres, like the subtropical gyres, are asymmetric, with strongest circulation at the western boundary. Subpolar circulation extends more vigorously throughout the gyre to the ocean bottom than does subtropical circulation, and so there tend to be focused currents along all boundaries in subpolar gyres, but the western boundary currents are the strongest. The western boundary current in the Weddell Sea is the simplest, flowing northward along the Antarctic Peninsula. The Ross Sea gyre is less confined by a western boundary and part of the flow continues westward west of the Ross Sea along the coast of Antarctica. The North Atlantic s subpolar gyre has two western boundaries, one along Greenland and the other along Labrador. There are western boundary currents in each of these areas, called the East Greenland Current and the Labrador Current. The North Pacific s subpolar gyre has a leaky boundary due to the many island chains. The main western boundary current is the Oyashio, which flows southward along the southern part of the Kuril Islands and along the coast of Hokkaido before it separates and flows eastward. The western boundary current in the Bering Sea and along the Northern Kuril Islands is referred to as the East Kamchatka Current because it mainly flows along the coast of Kamchatka. There is also a western boundary current along the coast of Sakhalin within the Okhotsk Sea, called the East Sakhalin Current. And finally, the shape of the Gulf of Alaska (the area of the eastern subpolar gyre surrounded by Alaska and Canada) allows a weak western boundary current called the Alaskan Current along the Alaska Peninsula and the eastern Aleutian Islands.

The third major surface circulation feature is the complex circulation in the tropics, which is described separately in the section on tropical circulation, including its vertical structure, which is also complex. At the sea surface, the Pacific and Atlantic tropical circulations include a North Equatorial Current (which was already mentioned as being part of the subtropical gyre circulation and which is not close to being on the equator, but rather is a westward flow north of 10 °N), the North Equatorial Countercurrent (eastward flow at about 5 °N), and the South Equatorial Current (westward flow at the surface on the equator and in the Southern Hemisphere).

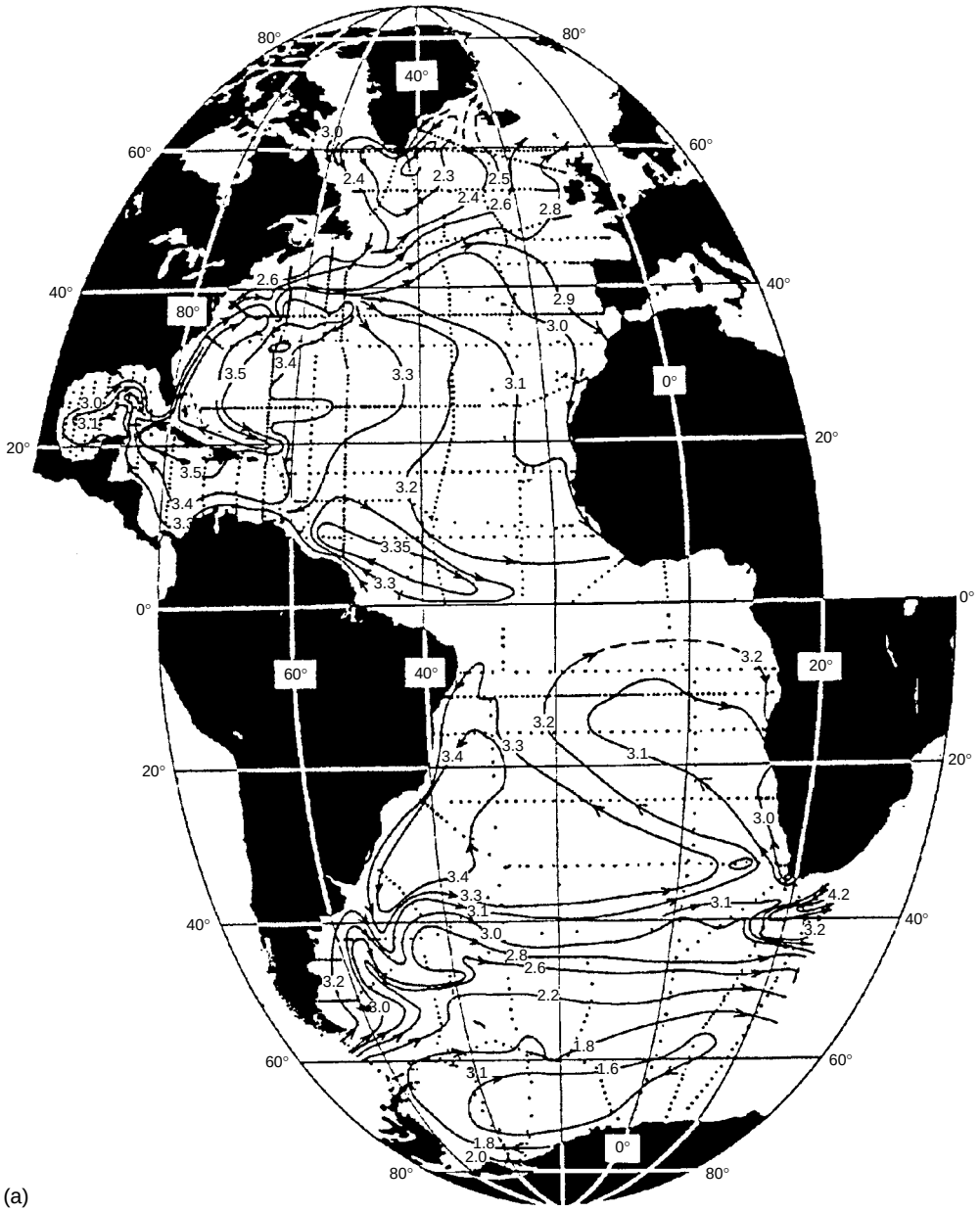
The surface circulation in the tropical Indian Ocean is the most complex of all of the general circulation patterns because it responds to the strong seasonal monsoonal winds, which reverse direction. West of India, in the Arabian Sea, the circulation is clockwise (around a high pressure center) during the southwest monsoon in late summer, i.e., like a normal Northern Hemisphere subtropical gyre. The northward flowing western boundary current is called the Somali Current. During the Northeast Monsoon, which occurs in late winter, the wind blows from the land to the sea and the Arabian Sea circulation is counterclockwise around a low. The Somali Current then reverses direction and flows southward.

Finally for the surface circulation, the connections between basins and gyres are shown in Figure 2. The most important connecting currents for the global ocean circulation are the westward flow from the Pacific to the Indian Ocean through the Indonesian passages, the northward flow from the Pacific to the Arctic through Bering Strait, and the connection of Agulhas Current waters into the South Atlantic s circulation. These flows are not necessarily part of the wind driven circulation, but are part of the shallow portion of the thermohaline circulation, described below.

THE MID-DEPTH CIRCULATION

The surface circulations described in the previous section do not extend unchanged to depth. One major change with increasing depth is that current speeds of the general circulation generally decrease considerably. This is most true for currents that are driven by the winds. There is another part of the circulation that is driven by high latitude cooling and resulting deep convection. These currents, although weak in comparison with the surface flows of western boundary currents such as the Gulf Stream, may be strongest close to the ocean bottom. A second major change is that the gyre circulations shift and shrink with depth. A third major change is that topography becomes more important in steering the flows (most apparently for the abyssal circulation described in the section on deep circulation). And fourthly, the equatorial circulation has a large amount of vertical structure (see section on tropical currents below), differing in this way from circulation outside the tropics.

The mid-depth circulation of the Atlantic and Pacific Oceans, from Reid (1994, 1997) is depicted in Figures 3(b) and 4(b). For comparison, Reid s maps for the surface

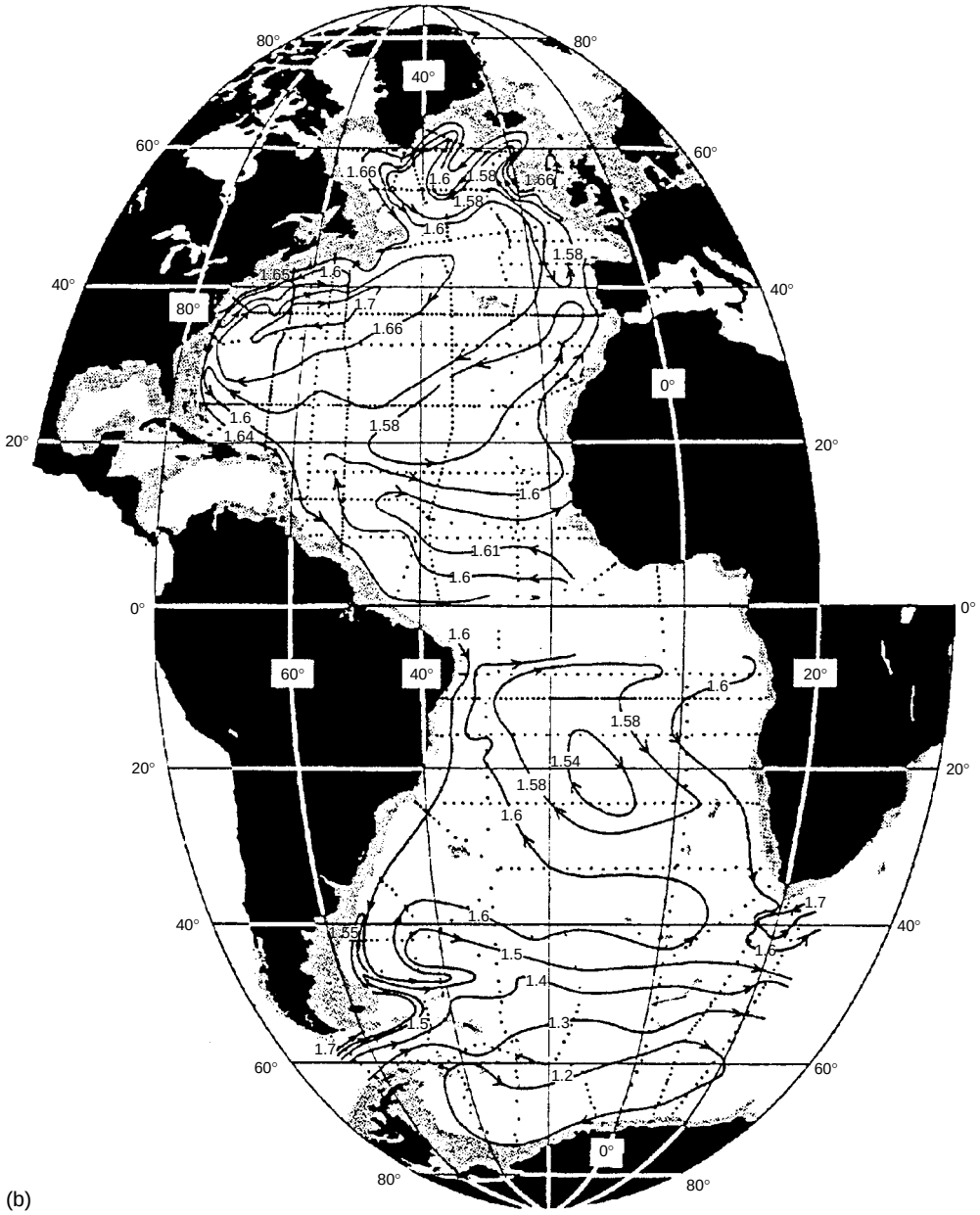


(a)

Figure 3 Depictions of the circulation of the Atlantic Ocean at the (a) sea surface, (b) at 1500 m depth, and (c) at 4000 m depth (all from Reid, 1994). The contoured quantity in each map is called the streamfunction. Where the contoured surface is high is the high pressure region described in the text. The general circulation follows the contours, that is, with flow around the highs and lows. The strength of the currents is proportional to the distance between the curves; the closer they are the stronger the current

circulation are included because the circulations in Figure 2 were schematic. As of now there is no product like those of Figures 3 and 4 for the Indian Ocean, but many of the comments made here about the subsurface circulation apply also to the Indian Ocean.

The subtropical gyres of all of the four basins (North and South Atlantic, North and South Pacific) in Figures 3 and 4 can be seen to have contracted significantly towards the western boundary and also poleward. This contraction begins just below the sea surface, as can be seen in Reid (1994, 1997), where by 200 m depth, the gyres are already considerably smaller than the surface gyres. What is constant with increasing depth in these gyres is the position of the western boundary current and that the gyres remain asymmetric, with the western boundary current still the most vigorous part of the flow.



(b)

Figure 3 (*Continued*)

The subpolar gyres of the North Pacific and North Atlantic also change with depth. According to Reid (1997) the North Pacific s subpolar gyre becomes more strongly confined to the north, with the subtropical gyre shifting farther north with increasing depth. The North Atlantic s subpolar gyre is not very different at 1500 m than at the surface, although the flow is more strongly constrained by topography because the Mid-Atlantic Ridge encroaches on the northern side of the gyre. The Weddell Sea gyre at 1500 m is about half the strength of the surface Weddell gyre, but with no change in location or horizontal extent. The clockwise Ross Sea gyre is not strongly defined at any level in Reid (1997) in that a western boundary current is not indicated, possibly due to lack of observations. However, the clockwise flow is apparent and is little changed in location at 1500 m compared with the surface.

The Antarctic Circumpolar Current is at about the same location at 1500 m as at the sea surface.

Outside the surface intensified gyres, weaker gyre, which, have no parallels at the sea surface appear at 1500 m in Reid s (1994, 1997) analyses. Most of these gyres have not been studied carefully or mapped outside this particular set of studies. Direct observations of flow in most mid-latitude regions, using either current meters or drifting floats, show

(c)

Figure 3 (*Continued*)

that currents are dominated by time dependent eddies rather than by the very slow broad flows depicted in Figures 3 and 4. The primary exception to this eddy dominance is in the tropical region where the flows are nearly due east-west and somewhat faster than in the mid-latitudes (Davis, 1998).

Because the mid-depth circulation outside the tropics is so weak, the most useful measures of its impact and direction may be the total amounts of water moving from one latitude to another, as summarized in the section on global thermohaline circulation.

THE DEEP CIRCULATION

Near the ocean bottom, the topography of the ocean bottom significantly affects the circulation, because the mid-ocean ridges and island chains have larger and larger

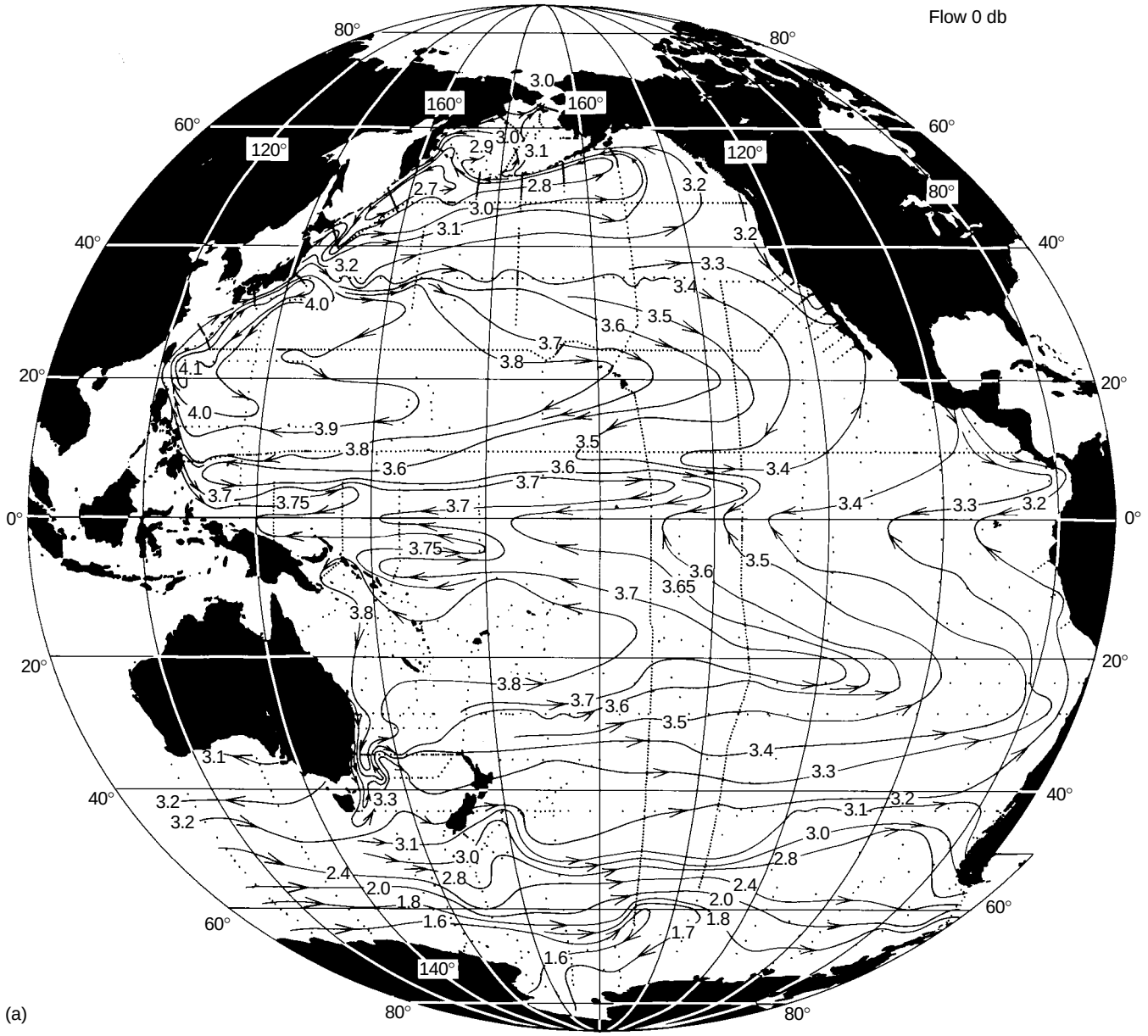


Figure 4 Depictions of the circulation of the Pacific Ocean at the (a) sea surface, (b) at 1500 m depth, and (c) at 4000 m depth, all from Reid (1997). The contoured quantity in each map is the streamfunction. Where the surface is high is the high pressure region described in the text. The general circulation follows the contours, that is, with flow around the highs and lows. The strength of the currents is proportional to the distance between the curves; the closer they are the stronger the current

cross-sections. As seen in Figures 3(c) and 4(c) from Reid (1994, 1997) the subtropical gyres shrink all the way to the western boundary and separation region, where the most vigorous surface currents are found and which therefore penetrate to great depth. The subpolar circulations reach to the ocean bottom, but must wind their way around the various topographic obstacles, particularly in the North Atlantic. The Weddell Sea gyre shrinks to the westernmost part of its basin.

The most significantly different aspects of the abyssal ocean s circulation are concentrated western boundary currents that have nothing to do with the surface intensified gyre s western boundary currents. The clearest of these deep western boundary currents in Figures 3(c) and 4(c) is a southward flow along the western boundary of the Atlantic, extending from the northern subpolar region to South America to about 40 °S. This deep flow was first detected by Wust (1935), although an explanation for

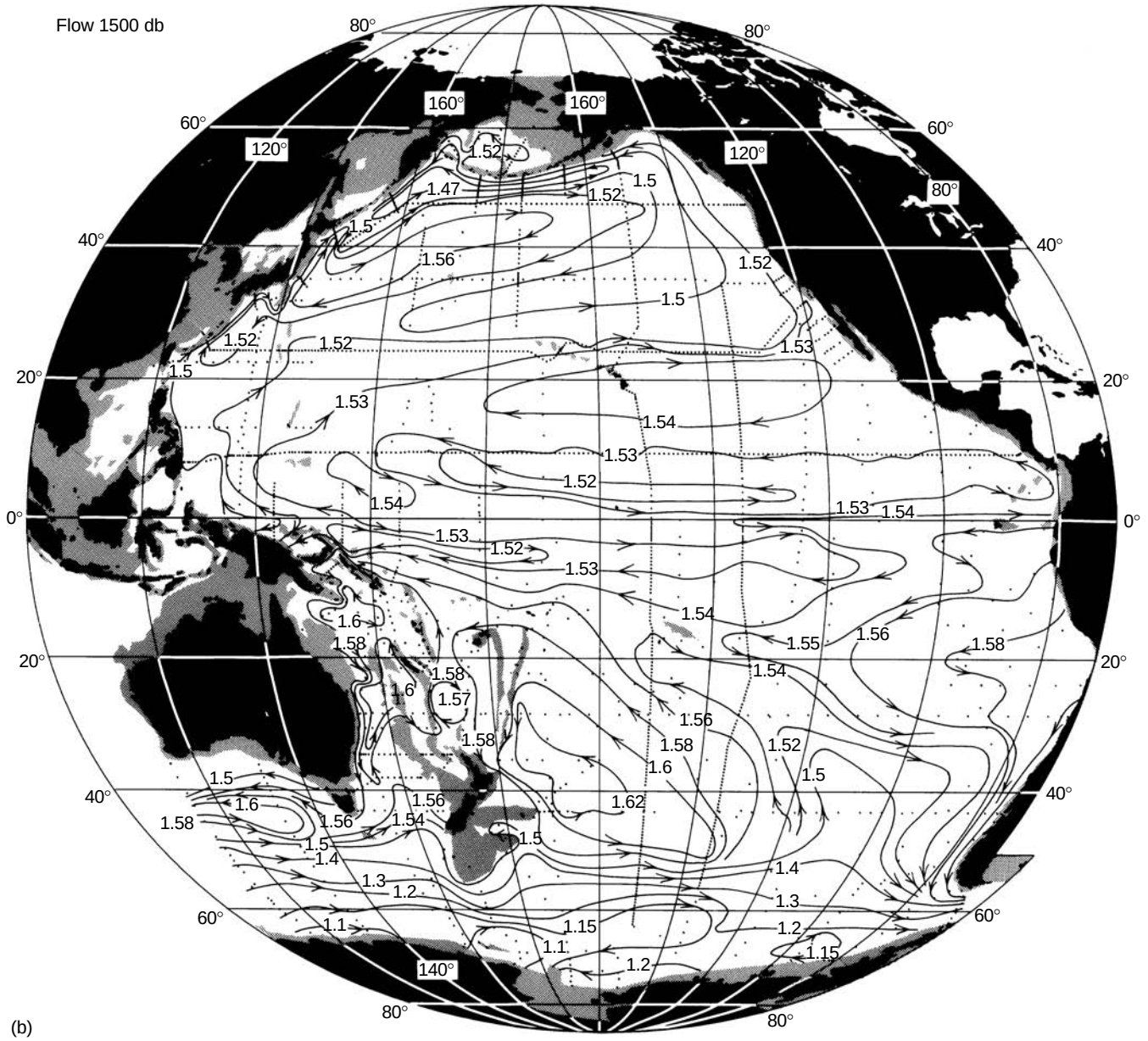


Figure 4 (*Continued*)

its existence did not come along until the early 1960s (Stommel and Arons, 1960). The deep western boundary currents are the most obvious and measurable part of the deep thermohaline circulation, in which deep waters are formed at high latitudes and spread via the boundary currents and very slow interior flow to fill the whole of the deep oceans below about 1500 m depth.

A deep western boundary current flows northward in the Pacific Ocean (Figure 4c), bringing deep water from the Antarctic (of both North Atlantic and Antarctic origin) to the deep South Pacific, across the equator and into the deep North Pacific.

In mid-latitude regions of the Atlantic and South Pacific, the deep flow that is not associated with the deep extension of the subtropical gyres or with the deep western boundary currents tends to be centered around low pressures (counterclockwise in the Northern Hemisphere and clockwise in the Southern Hemisphere), as shown by Reid in the maps reproduced in Figures 3(c) and 4(c).

THE TROPICAL CIRCULATION

The surface circulation in the equatorial Pacific and Atlantic is westward, and is directly connected to the westward

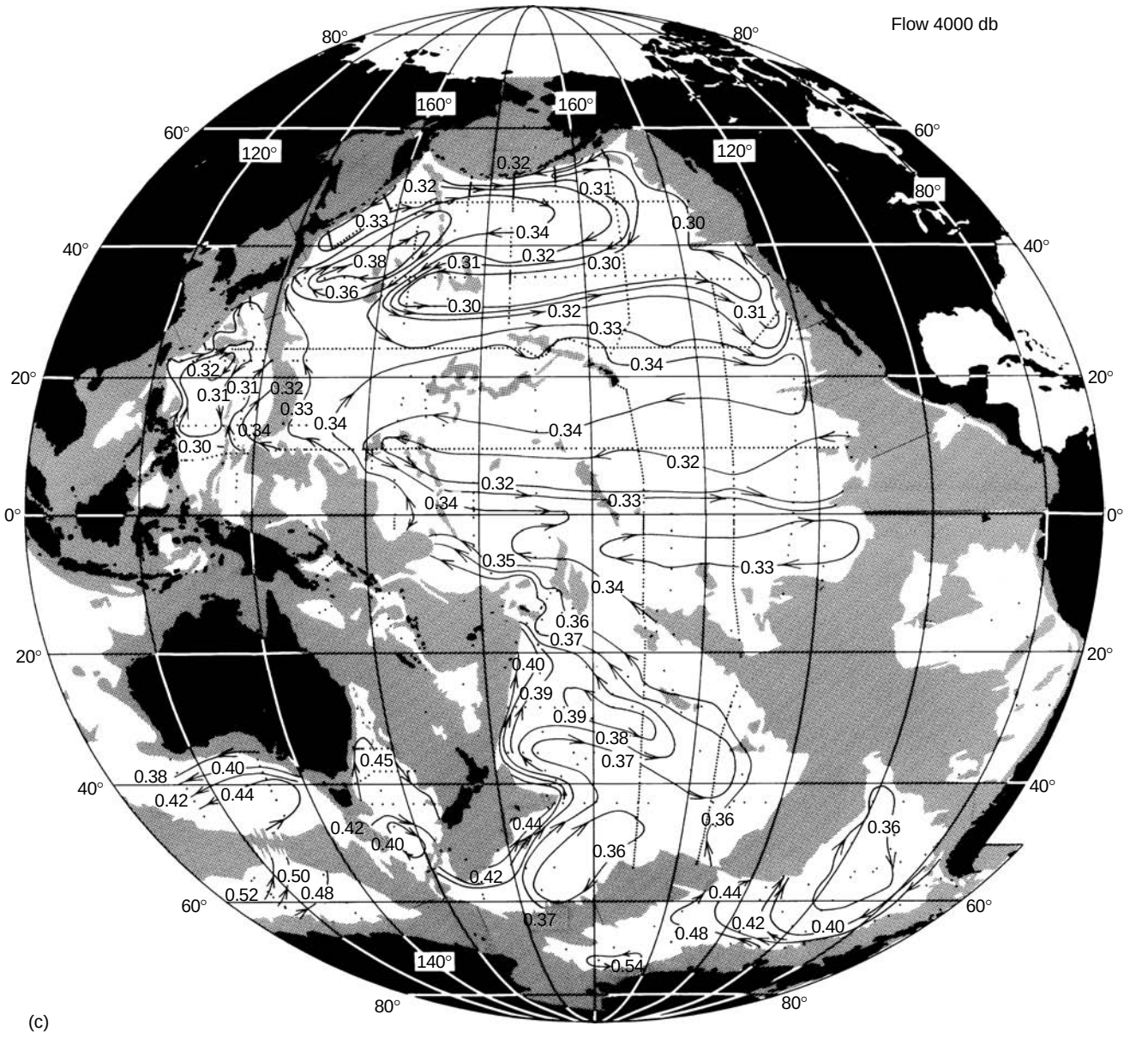


Figure 4 (*Continued*)

flow of the Southern Hemisphere subtropical gyres. The westward surface flow is separated from the Northern Hemisphere subtropical gyres by a narrow and strong eastward flow at about 5 °N, called the North Equatorial Countercurrent. This current probably reaches very deep into the ocean. It is the southern side of a very long and narrow counterclockwise circulation. The Southern Hemisphere does not usually have an equatorial countercurrent because the trade wind pattern is different from the Northern Hemisphere. The difference lies in the average position of the Intertropical Convergence Zone north of the equator, which creates the North Equatorial Countercurrent (*see* Intertropical Convergence Zone (ITCZ), Volume 1).

North of the North Equatorial Countercurrent lies the North Equatorial Current, which is also the westward flow of the subtropical gyre and which was mentioned in the section on surface circulation. The North Equatorial Current flows westward to the western boundary and splits into the northward western boundary current for the subtropical gyres (Gulf Stream and Kuroshio) and into a southward flow for the long, narrow clockwise tropical gyre (Antilles Current in the North Atlantic and Mindanao Current in the North Pacific).

The westward equatorial surface flow is driven by the trade winds (easterlies). When the trade winds weaken or even reverse, the flow of water westward at the equator

weakens or reverses and upwelling weakens or stops. In the Pacific, this is the case during an El Niño (*see* El Niño and La Niña: Causes and Global Consequences, Volume 1). In the Indian Ocean, this change in the trade winds happens twice a year during the transitions between monsoons; when the winds are from the west along the equator, the equatorial surface flow becomes strong and eastward and is referred to as the Wyrtki Jet.

Immediately below the surface currents, the flow directly on the equator is eastward and very strong (greater than $1\,m\,s^{-1}$, that is, even stronger than the Gulf Stream). The strong flow is centered around 150 m depth, and is shallower to the east. It is no more than about 100 m thick and is confined to within 1 °latitude from the equator. This continuous flat ribbon of flow (100 m thick by 200 km wide) is called the Equatorial Undercurrent.

Below the Equatorial Undercurrent, the equatorial currents have complicated structure in the vertical. Immediately below the undercurrent, the flow is westward and is called the Equatorial Intermediate Current. This current extends down to about 900 m depth. Between 900 and 1800 m, the flow reverses direction every 150 m. These are referred to as the stacked jets. Finally at the bottom on the equator is a layer several thousand meters thick that flows eastward (Firing, 1987).

The subsurface flow just north and south of the equator is eastward, down to at least 400 m (Figure 5). These flows are called the North and South Subsurface Countercurrents, or the Tsuchiya Jets. They sometimes appear to be slightly deeper poleward extensions of the Equatorial Undercurrent.

THE GLOBAL THERMOHALINE CIRCULATION

The circulations at individual levels shown in the previous three sections provide a view of the complex horizontal flows that transport waters around the basins. It is difficult from these views to discern the throughputs of water that connect the basins and oceans. However when the ocean s properties are viewed along cross-sections from top to bottom and, say, from the northernmost to the southernmost part of each hemisphere, it is clear that there must be connections between the ocean basins. For instance, north–south cross-sections of salinity through the middle of each ocean in the article on salinity patterns (*see* Salinity Patterns in the Ocean, Volume 1) show that waters can be detected for a long distance from their high or low salinity sources. These properties are carried by the circulation. The circulation that carries water properties from one wind driven regime to another is the thermohaline circulation.

The thermohaline circulation is driven by cooling, evaporation and salinization through sea ice formation, which increases density, and by heating and precipitation, which decrease density. The thermohaline circulation is much weaker than the wind driven circulation and so it is difficult to see its effect on currents in the upper ocean where the wind driven circulation is vigorous. In the deep ocean, the permanent circulation, particularly when quantified by transports of currents carrying water from one region to another, is dominated by thermohaline forcing.

Cooling is probably the most significant of the various processes in driving a discernible, easily quantified circulation. This is because deep convection created by cooling is localized to well-defined regions of the oceans, and produces waters that are injected below the depth of the vigorous part of the wind driven circulation. References to global thermohaline circulation usually mean this mechanism of driving circulation. While cooling at mid-latitudes in the separation regions of the wind driven western boundary currents is enormous, the associated convection is not deep. The cooled waters are not dense enough to penetrate below the thermocline. It is nearly impossible to quantify how much circulation is associated with mid-latitude cooling because the circulation above the thermocline is strong and associated mainly with wind driving. Mid-latitude cooling is, however, an important part of the global heat budget (poleward heat transport), resulting in a shallow overturn, with somewhat cooler waters returning towards the equator, to be heated and returned poleward in the western boundary current.

Evaporation is large in the subtropical gyres, under the atmospheric high pressure zones (*see* Salinity Patterns in the Ocean, Volume 1). As with mid-latitude cooling, the effect of the evaporation can be seen almost exclusively in changes of properties and in calculations of net freshwater transport into the evaporation regions, but not in the effect of evaporation on the circulation itself. On the other hand, brine rejection associated with sea ice formation also increases salinity and is a significant, measurable thermohaline process in producing denser water because it is localized, like high latitude deep convection driven by cooling (*see* Sea Ice, Volume 1).

Globally the heating of the Earth balances cooling. Most heating is in the tropics. This means that deep waters cooled at high latitudes must eventually be heated and returned to the surface. This process is probably distributed over most of the oceans, with deep waters gradually warming by downward mixing of heat until the water parcels rejoin the surface circulation and move to the cooling regions. The upwelling rates are very low and regions of larger and smaller upwelling have yet to be well determined.

Precipitation s role in the thermohaline circulation is mostly to inhibit convection. A large-scale example is the subpolar North Pacific, where the surface layer is relatively fresh due to local precipitation, large amounts of runoff from the coast of North America and sea ice melt in the Bering and Okhotsk Seas. The result is a surface layer that cannot convect through the resulting halocline even when

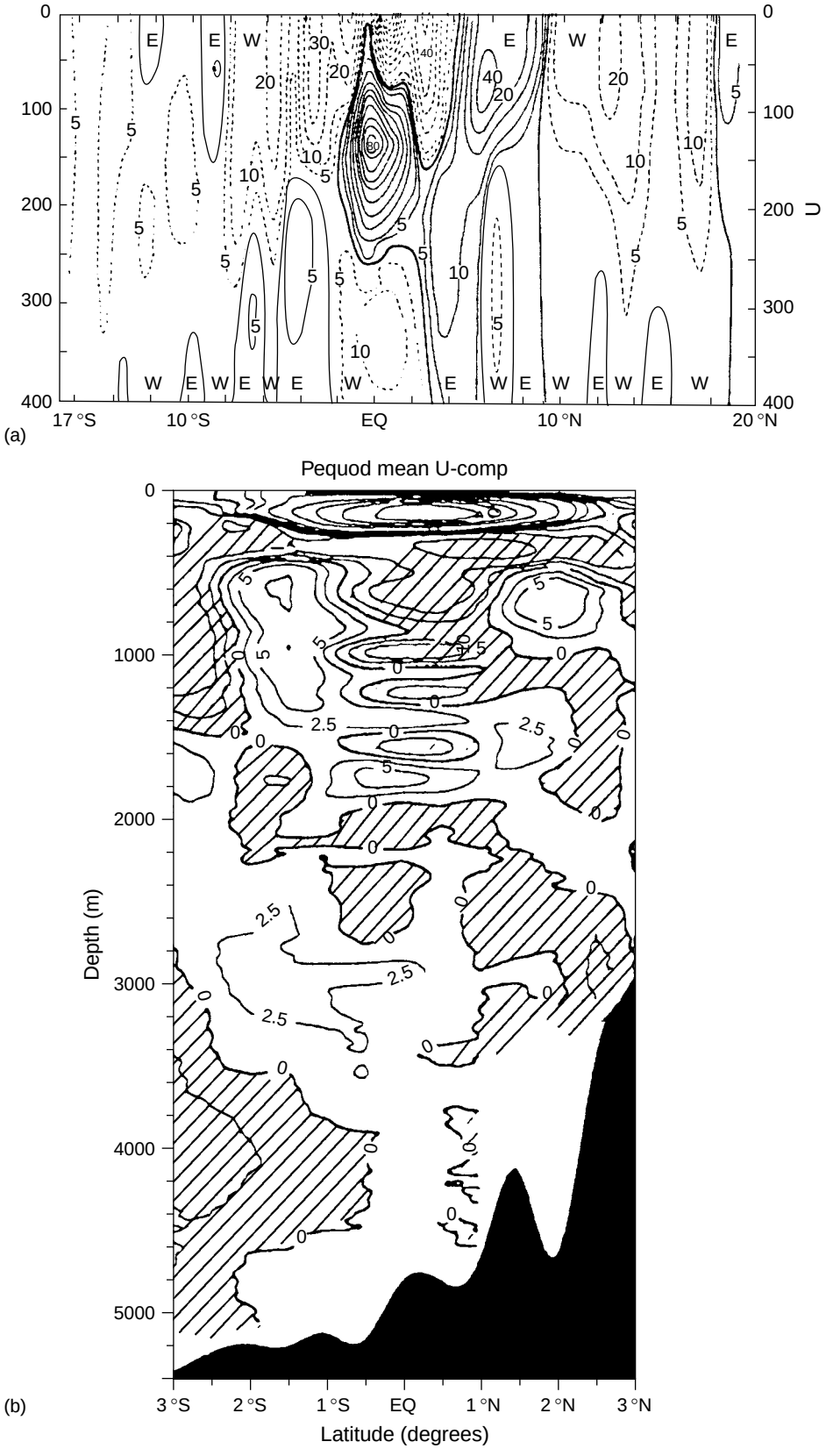


Figure 5 East–west circulation within several degrees of the equator in the central Pacific, based on Wyrtki and Kilonsky (1984) for the upper 1000 m and Firing (1987) for the full water column

the surface temperature is reduced to the freezing point. There are also well-documented events of large (hundreds of square kilometers) patches of low salinity at the sea surface in the subpolar North Atlantic that can inhibit the usual convection as they move slowly counterclockwise with the wind driven circulation (Dickson *et al.*, 1988).

Because the thermohaline circulation is weaker than the wind driven circulation and because it inherently involves overturning, a common way to quantify it is to compute how much water in different layers is transported into and out of given areas. A common choice is to calculate how much water flows northward/southward across an east–west section that goes completely across an ocean basin. For instance, transports are often calculated at 24 °N in the Atlantic and Pacific, where there are data sets that go from coast to coast. To be most meaningful, oceanographers divide the water column up into layers associated with surface waters, thermocline water, and various intermediate and deep layers. The resulting transports can show water moving northward across that latitude in say the uppermost, warm layer, and returning southward in a deep, but not bottom layer.

Transport calculations were introduced above for the wind driven circulation. It was seen above that the wind driven western boundary currents carry about 50–100 Sverdrups. The global thermohaline circulations described here carry about 15–20 Sverdrups, which is considerably less than the wind driven circulation. However, the thermohaline circulation is important because it connects all regions of the global ocean.

The recognizable elements of the thermohaline circulation are the deep western boundary currents, which were mentioned in section 4 in reference to Figures 3(c) and 4(c). The pattern of flow that fills in the interior of the basins from the deep western boundary currents is strongly affected by topography.

There are two global thermohaline circulations involving deep convection at high latitudes: the North Atlantic Deep Water (NADW) and the Antarctic Bottom Water (AABW) cells (Figure 6). These two global overturning circulations are interconnected. Smaller thermohaline cells are more connected with the wind driven circulation and quite complex (see Schmitz, 1995) and will not be described here. NADW and AABW are water masses, that is, particular types of waters that can be recognized by their special properties, such as a salinity maximum (for NADW) or extreme cold temperature (AABW). Because thermohaline circulation involves transformation of surface waters into something else, it is often and perhaps most easily described in terms of the plethora of recognized water masses. Besides the NADW and AABW, we will introduce several other water masses that are associated with the overturning circulations of these two.

The NADW global overturning has sometimes been referred to as the global conveyor belt (Broecker, 1991), referring to its extent from the northern North Atlantic to the Pacific Ocean via the Antarctic, its upwelling, and its return of warm water through the Indonesian archipelago, across the Indian Ocean and back up into the Atlantic (*see* Ocean Conveyor Belt, Volume 1). The actual NADW cell has considerably more detail and many additional elements (Worthington, 1976; Schmitz and McCartney, 1993; Reid and Lynn, 1971; review in Talley, 1996). The NADW cell is depicted reasonably well by Schmitz (1995, 1996) (Figure 6).

The deep water in the NADW cell is formed in the northern North Atlantic. There are three northern North Atlantic sources. First is localized deep convection in the Nordic Seas, which lie north of the sills connecting Greenland, Iceland and Scotland. The renewed Nordic Seas water is dense and plunges over the sills into the subpolar North Atlantic; the most important pathway is through Denmark Strait between Greenland and Iceland. This dense Nordic Sea overflow water (NSOW) reaches the bottom of the northern North Atlantic and flows westward and southward, following the continental slope. The second source is localized intermediate depth convection in the Labrador Sea, between Labrador and Greenland. This area is part of the subpolar North Atlantic. The convected LSW fills the mid-depth of the Labrador Sea and is moved eastward out of the Labrador Sea to fill the subpolar North Atlantic and southward along the continental slope, just above the Nordic Seas water. Together the LSW and Nordic Seas overflow water form the deep western boundary current in the North Atlantic. At mid-latitudes these waters are joined by those from the third source, which is the Mediterranean Sea. Within the Mediterranean, deep convection driven by heat loss and evaporation creates a dense, very saline water that flows back into the North Atlantic, plunging down over the sill at the Strait of Gibraltar and filling intermediate depths at mid-latitudes. This MOW is recognized by its high salinity (*see* Salinity Patterns in the Ocean, Volume 1). The MOW joins the deep western boundary current in the tropics and together the renewed North Atlantic waters flow southward into the South Atlantic as the NADW. By the time they reach the southern South Atlantic, the individual sources of the NADW have mixed together and cannot be recognized; the NADW becomes a generally saline and well-oxygenated water mass.

NADW in the South Atlantic joins the Antarctic Circumpolar Current and flows eastward towards the Indian Ocean. As the NADW is carried eastward in the Antarctic Circumpolar Current, its high salinity gradually becomes fresher through mixing with other waters, but it remains recognizable as a salinity maximum all the way through the Indian and Pacific Oceans. Some of the high salinity water flows northward into the Indian and Pacific Oceans,

and is one of the two major components of the deep water for these oceans. In the Indian Ocean, the northward flow of deep water occurs in three regions, separated by mid-ocean ridges. In the southwestern Indian Ocean, a branch of NADW flows northward into the region east of the Agulhas. In the Pacific Ocean, the main northward flow is just east of New Zealand, in the deep western boundary current described in the section on tropical circulation. This deep flow fills the South Pacific and moves northward through the Samoan Passage into the tropics and then into the North Atlantic.

Where does NADW upwell to eventually complete the circuit back to the northern North Atlantic? This is a very difficult question to answer. Within the North Atlantic, by computing net transports across sections from coast to coast, one can show that the northward warm flow is in the Gulf Stream and is mostly in and above the thermocline. The calculations become more difficult the more removed they are from the North Atlantic. What is clear is that there is no wholesale upwelling in the Pacific Ocean that feeds the warm westward flow through the Indonesian archipelago, as depicted in Broecker (1991) (*see* figure in Thermohaline Circulation, Volume 1). In the Antarctic Circumpolar Current, NADW is joined by bottom waters formed in the Antarctic. These flow together into the deep Indian and Pacific Oceans and both upwell into older deep water layers (between 1500 and 3000 m depth) in those oceans. The IDW and PDW re-enter the Antarctic Circumpolar Current and dilute the NADW there. South of the Antarctic Circumpolar Current, these mingled deep waters upwell to near the surface, only to have some portion convected or salinified through brine rejection to become

deep and bottom waters in the Antarctic (the second major global overturning cell).

Some of the upwelled deep waters might also take part in formation of the thick mixed layer waters just north of the Antarctic Circumpolar Current. These thick mixed layers vary in properties around Antarctica, becoming colder and fresher eastward from South America until reaching their most extreme values just west of Chile (McCartney, 1977). The thick mixed layers in general are called Subantarctic Mode Water (SAMW). The coldest, freshest SAMW is modified by intermediate depth convection just west of Chile and becomes a fresh intermediate water called Antarctic Intermediate Water (AAIW). AAIW fills all of the Southern Hemisphere oceans north of the Antarctic Circumpolar Current, by spreading northward by the broad subtropical gyre circulations. SAMW, which is warmer than AAIW, also spreads northward from the Antarctic Circumpolar Current, via the subtropical gyre circulations.

SAMW and AAIW have three sources: water that flows southward in the Southern Hemisphere s subtropical gyre western boundary currents, water from south of the Antarctic Circumpolar Current that is pushed northward across the current, and upwelling from the underlying deep waters, most likely as the SAMW and AAIW circulate around the subtropical gyres.

Schmitz (1995) shows an upwelling of deep water to intermediate water within the subtropical gyres and then a conversion of intermediate waters to warm upper waters, primarily in the eastern South Pacific and eastern South Atlantic. Many of the conjectures about upwelling are, however, based on mass balances that have significant errors. What Schmitz s (1995) schematics and budgets do clearly show is that the global thermohaline system is fairly complicated.

The second major global overturning circulation is that of AABW (or Lower Circumpolar Deep Water, CDW), with formation of this water in localized regions in the Antarctic (Orsi *et al.*, 1999). Two primary sources of AABW have been found in the Weddell Sea and in the Ross Sea. Both of these areas have clockwise subpolar-type gyres, as described in the first section. Formation is through brine rejection, which occurs during formation of sea ice (*see* Salinity Patterns in the Ocean, Volume 1; Sea Ice, Volume 1). A third distributed source has been found along the Antarctic coast south of Australia (Rintoul, 1998), also presumed to be brine rejection in coastal areas. AABW spreads northward to the Antarctic Circumpolar Current through the gyre circulations south of the current. From the Antarctic Circumpolar Current, it enters the deep Atlantic Ocean as a northward flowing deep western boundary current, beneath the southward flowing deep western boundary current that carries NADW. AABW is the bottom water of the South Atlantic and also of the North Atlantic, extending as far north as Bermuda. By the time it reaches Bermuda, it all upwells into the NADW layer above it, and the bottom waters north of this region are NADW.

In the Indian and Pacific Oceans, AABW flows northward along with the high salinity NADW core that is also within the Antarctic Circumpolar Current. Oceanographers usually do not refer to these cores as AABW or NADW, because they are significantly diluted by the old deep waters from the Indian and Pacific Oceans, which themselves are formed from upwelled AABW and NADW. Throughout its path, AABW upwells into the deep water layer above it, and so it does not fill the northern reaches of either the Atlantic or Pacific Oceans.

What is the warm water return path to the AABW source? As mentioned above, the water that is found at the surface in the Antarctic regions where AABW is formed, is upwelled deep water from the Antarctic Circumpolar Current.

The magnitude of the currents associated with the NADW and AABW overturning circulations is about $5\,\mathrm{cm\,s^{-1}}$. The total transport of NADW out of the North Atlantic is 15–20 Sverdrups. The transport of AABW into the Antarctic Circumpolar Current is not as well measured, but is between 10–30 Sverdrups.

PHYSICAL PROCESSES THAT CREATE THE GENERAL CIRCULATION

Up to this point, the circulation of ocean waters has been described without much mention of the physical processes that oceanographers consider to be central. These are described here. The preceding sections could have been couched in these physical process terms, much as is the description of global circulation and water masses in a recent text by Tomczak and Godfrey (1994) and in a review of what is known about Pacific circulation with emphasis on comparison with theory (Talley, 1995). The central concepts for ocean circulation are: geostrophy, Ekman transport, Sverdrup transport, western boundary current theory, subduction, abyssal circulation theory, and equatorial circulation theory. These are not a complete set of the important concepts, but serve as a framework for much of what was described in the previous sections. Reasonable descriptions of these processes without extensive mathematical formulations can be found in texts such as Tomczak and Godfrey (1994).

A basic tenet for general circulation, which has very large spatial scales except in boundary currents, and long temporal scales, is that friction (viscosity) is very weak. Molecular-scale diffusion is indeed extraordinarily weak compared with the diffusion and friction that would be needed to have any impact on the general circulation. The ocean does have some frictional responses, including the Ekman transport and western boundary current formations described below. Because we observe these behaviors, which occur only in boundary layers at the sides of the

ocean where flow changes rapidly with distance or depth, and because we also observe that most of the ocean acts as if there were no friction, we conclude that something like friction must be acting near boundaries. Based on where the pseudo-frictional behavior occurs, we empirically conclude that small-scale turbulence, and possibly eddies up to a scale of a few kilometers, create something like friction and diffusion for the general circulation.

The most central concept for the general circulation is *geostrophy*, which is a balance between the force resulting from differences in pressure and the Coriolis force resulting from the Earth s rotation (*see* Coriolis Effect, Volume 1). Geostrophic flows in the Northern Hemisphere are always exactly to the right of the pressure difference force, while in the Southern Hemisphere they are always exactly to the left. All general circulation flows, except in the very surface layer (top 50 m) and right on the equator, are geostrophic. Tests of geostrophy using direct current measurements have borne this out. The pressure differences, especially for the upper ocean circulation, arise from differences in sea surface height, which allows more water mass to appear in one place than another. The difference in sea surface height across the strongest currents such as the Gulf Stream are no more than about 1 m and height difference for weaker currents are much smaller. This cannot be measured directly with today s technology. Therefore, we have no direct knowledge of the absolute pressure difference between two points. What we can construct though from the ocean s density distribution is how the pressure differences change with depth, which means that the velocity changes with depth. For instance, for a wind driven circulation such as the Gulf Stream, the velocity is highest at the sea surface, which means that the largest pressure difference is at the surface. Velocity decreases with depth, which has implications for the density distribution with depth, resulting in density surfaces that tilt across the current (see the textbook descriptions). We then must somehow measure or guess the true pressure difference at one reference depth. This is done in many ways, all of which require intensive work.

The second central concept is of *Ekman transport*, which is the flow driven directly by the winds in the surface layer. The winds push on the water through friction, and the frictional flow in the very top layer (say 1–2 m) is conveyed through friction to the next layer and so forth. This downward conveyance of the frictional driving dies out with depth, and vanishes around 50 m. Because of the Earth s rotation, each layer is driven slightly to the right of the one above in the Northern Hemisphere (left in the Southern Hemisphere). If the total flow in this 50 m thick layer, which was first described by Ekman (1905), is added up, the net flow is exactly to the right of the wind (Northern Hemisphere). This net flow is the Ekman transport and is how the wind directly forces the ocean.

Ekman transport has immediate consequences for the ocean when winds blow along a coast. If, as along California, they blow from the north in the Northern Hemisphere, then the Ekman transport is offshore (i.e., to the west), which results in upwelling at the coast to fill in waters pushed westward by the Ekman transport. The upwelled water comes from about 100–200 m depth and brings with it significant nutrients and colder water, as described in the first section. The offshore transport also causes the sea surface to be slightly lower at the coast than offshore, which results in a geostrophic flow along the west coast of North America, creating the California Current, which is the eastern boundary current described in the first section.

In the middle of the oceans, in places where the winds vary with position, the Ekman transport also varies with position. This results in upwelling if the Ekman transport has to be supplied (Ekman divergence), or downwelling if it has to be removed (Ekman convergence). In subtropical gyres, which are driven by westerly (west to east) winds on the higher latitude side and trade winds (east to west) on the lower latitude side, Ekman transport is towards the equator on the higher latitude side and towards the pole on the lower latitude side. This results in downwelling in the gyre. Subpolar gyres on the other hand are characterized by upwelling, which explains their much higher surface nutrient content and hence much higher productivity than subtropical gyres. The biology in surface waters uses up whatever food is available in a short time, and the dead parts and fecal pellets that fall below the surface layer provide nutrients back into the water through bacterial decay.

The surface circulation described in the first section was the purely geostrophic circulation, based on pressure differences and Coriolis acceleration. Ekman transport is not geostrophic, and is superimposed on top of the geostrophic flow. Thus, the surface circulation, as described here, is really more like the true circulation about 50 m below the sea surface, assuming only small changes in the pressure difference force between the surface and 50 m.

The third concept for general circulation is *Sverdrup transport*. This is the reaction of the full water column to Ekman convergence or divergence. When there is Ekman convergence, and hence downwelling, it is as if the whole water column is being squashed. Because the water column is effectively rotating because the Earth is rotating, squashing will cause its rotation rate to flow (much like a spinning ice skater who extends her arms and slows). The rotation rate change is accomplished by the movement of the water column to a different latitude, with slower spinning being towards the equator and faster spinning towards the poles. This response to the winds/Ekman transport was theorized by Sverdrup (1947). Sverdrup transport is geostrophic, i.e., the forcing that causes the northward or southward movement is very small, so that what we actually observe is that

the pressure field is set up to produce geostrophic flow in the correct direction.

Because the average wind patterns for the globe are east–west, the mid-ocean Sverdrup transport tends to be all northward or almost all southward at a given latitude. Therefore Sverdrup transport can work only if there are north–south boundaries along which the flow can return in a narrow boundary current that has different physics from the Sverdrup transport. Support for the Sverdrup transport idea lies in the existence of the subtropical and subpolar gyres, in which flow is equatorward over most of the ocean in Ekman convergent regions, and poleward in Ekman divergent regions. This requirement for boundaries explains why the ocean does not exhibit a clear Sverdrup transport response to the winds in the Antarctic where there is no land (at the latitude of Drake Passage), and why the atmosphere which obeys the same physical laws as the ocean does not exhibit Sverdrup transport.

Narrow western boundary currents are required to return the Sverdrup transport back to its original latitude, to close the mass balance (because there are no holes in the ocean). These boundary currents are always on the western side of the ocean. This is because they are removing energy (or rotation changes) put in by the large-scale winds through the Sverdrup transport. The physics of the boundary currents that dictates that the currents must be on the western side is friction (or viscosity), which can act if there are large changes in flow in a short space scale. In a subtropical gyre of the Northern Hemisphere for instance, the winds put in negative rotation (reducing the rotation), with the response that the water moves equatorward. The water must move back poleward to complete the gyre. This is only possible if there is a source of positive rotation. Friction acting on boundary current at the western wall can input positive rotation because the current will have no velocity at the wall, and maximum velocity at some distance offshore (about 50–100 km). This difference in velocity (zero at the wall and maximum offshore) will thus create positive rotation (think of a paddle wheel sitting in the flow next to the wall) and can balance the Sverdrup transport forcing. The theory of frictional western boundary currents was introduced by Stommel (1948) and Munk (1950) (*see* Munk, Walter, Volume 1). A frictional boundary current on the eastern wall however would have negative rotation with zero velocity at the wall and maximum poleward velocity offshore and so cannot balance the Sverdrup transport forcing. Like Sverdrup transport and unlike Ekman transport, the flow in western boundary currents is almost geostrophic i.e., the frictional part, while important in determining the direction and shape of the flow, is still much weaker than geostrophy, and so it is possible to use estimates of the horizontal changes in pressure to compute the speed of the currents.

Eastern boundary currents are of a completely different nature from western boundary currents, and were explained above when Ekman transport was introduced. They are the geostrophic flow that goes with the coastal upwelling created by offshore Ekman transport (equatorward if there is upwelling and poleward if there is downwelling). Upwelling is very shallow because the ocean is strongly stratified and it is not possible to bring very dense water all the way up to the surface, and so eastern boundary currents are also very shallow.

Returning to the interior circulation of the ocean gyres, we consider again the topic of subduction. As already explained, subduction is a phenomenon in the subtropical gyres in which the cooler surface waters from the higher latitudes in the gyre must be moving equatorward because of the Sverdrup transport. In the steady state ocean circulation, the surface waters are warmer and lighter towards the equator. Because there is no whole scale heating at the sea surface that would change the cooler water into exactly the right warmer waters, the cooler waters subduct beneath the warmer ones. Subduction is the main mechanism for setting the vertical temperature (and salinity) structure in the subtropical gyres, down to the depth of the coldest/densest water that outcrops at the highest latitude of the subtropical gyres. This latitude is around $40-55^{\circ}$ depending on the wind patterns in each of the ocean basins. The subducted waters typically reach no deeper than about 500–1000 m in the centers of the subtropical gyres.

Part of the deep ocean circulation (i.e., the thermohaline part) is driven by sinking of cold, dense water at high latitudes and upwelling of these waters elsewhere. Abyssal circulation theory, introduced by Stommel and Arons (1960), explains the current structure that results from isolated sinks of water, the spread of these waters to fill the oceans, and a general upwelling to return the cold water back to the upper ocean. The latter process must be accompanied by downward diffusion of heat, of course, but is not included explicitly in the Stommel/Arons theory. The crux of their theory is that most of the ocean is responding to the general upwelling, because the sinking is so localized (for the ocean now this would be in the Nordic Seas and the Antarctic). If the upwelling is uniform, i.e., the same size everywhere in the ocean, and if the ocean had a flat bottom, then the upwelling acts on the deep ocean like Ekman divergence, acting to create Sverdrup transport in the wind driven ocean circulation. That is, the response is for the water columns to move poleward (slowly) because their rotation rate is being increased by the upwelling.

There must of course be a connection between the sources of dense water at high latitudes and all of this upwelling and poleward flow. Stommel and Arons showed that the connection is through deep western boundary currents, called deep because this mechanism has to do with circulation in the deep layer fed by the high-latitude

sinking. Because the actual deep ocean is far from flat bottomed, and because the upwelling itself very likely is not uniform (although we do not yet know what its distribution is), the slow poleward flow of the Stommel/Arons theory is not really observed. However, the deep western boundary currents that connect all of the sinking and upwelling together are clearly observed and are the primary evidence that something like their mechanism is at work. Again, as for Sverdrup transport and wind driven western and eastern boundary currents, the abyssal circulation driven through this sinking and upwelling is basically geostrophic.

The equatorial circulation is a special topic because geostrophy breaks down at the equator. The breakdown is apparent because the Coriolis force has one sign in the Northern Hemisphere (flow is to the right of the pressure difference force) and the opposite sign in the Southern Hemisphere (flow to the left of the pressure difference). Right on the equator, the Coriolis force vanishes. Flow directly on the equator responds directly to the pressure difference force and flows from high pressure to low pressure (instead of being geostrophic). The surface frictional flow on the equator is pushed in the direction of the winds (rather than at right angles as in Ekman transport). In the Pacific and Atlantic, the equatorial winds are almost always easterlies (trades), and blow the surface water to the west. In the Indian Ocean, the winds reverse seasonally and sometimes blow the surface water to the west and sometimes to the east. Because all of these oceans have side boundaries at the equator, the wind blown water piles up at the boundary. This creates a sea surface height difference between the western and eastern parts of the oceans. In the Pacific and Atlantic, the water is piled up at the western boundary. This creates a pressure difference force that drives water back towards the eastern boundary. The flow created by this pressure difference force is the equatorial undercurrent.

A small distance away from the equator (within 30 km), the Earth s rotation begins to assert itself on the equatorial circulation. Ekman transport (at right angles to the wind forcing) begins to assert itself and flow also begins to respond geostrophically to pressure differences. The Ekman response to the prevailing trade winds in the Pacific and Atlantic is transport away from the equator both north and south of the equator. This creates upwelling in the equatorial region, which is important for setting sea surface temperature there and hence has a direct impact on climate (see section below).

There are many additional aspects to ocean circulation theory at the equator, including reasons for the existence of such very complicated vertical structure, but there is not room here to go into them because explanations require introducing the very large-scale waves that adjust the ocean circulation to changes in forcing.

THE ROLE OF OCEAN CIRCULATION IN CLIMATE

The ocean affects the atmosphere and hence the Earth s climate by directly heating and cooling the atmosphere. The ocean also affects climate through its effect on the Earth s albedo, which is how strongly the Earth reflects solar radiation back to space, and by its role in greenhouse gas cycles. Climate change also has an expression in the ocean through changes in temperature and salinity distributions, largely due to circulation changes, which in turn affect biological distributions in the ocean.

The circulation, temperature and salinity changes that are associated with present day climate fluctuations, particularly those centered at mid to high latitudes such as the North Atlantic Oscillation and the Arctic Oscillation, are linked to only small perturbations to the general circulation (*see* Arctic Oscillation, Volume 1; North Atlantic Oscillation, Volume 1). The tropical climate fluctuations such as El Niño are associated with relatively larger changes, but even these are only large at the sea surface, with only small changes to the deeper tropical circulation. During the much longer glacial cycles, the circulation can be significantly perturbed, but the basic physical elements described in the section above remain intact even if altered.

Ocean circulation thus affects climate through its influence on sea surface temperature, ice formation, and the upwelling/downwelling that affect greenhouse gas sequestration. The ocean s surface temperature distribution is a major forcing factor for the atmosphere, particularly in the tropics where atmospheric convection to great heights occurs. The ocean has a central role in the tropically centered climate cycle of El Niño (*see* El Niño and La Niña: Causes and Global Consequences, Volume 1). Sea surface temperature in the equatorial oceans is affected by the depth of the thermocline there, by the extent to which there is upwelling of colder waters from below the thermocline, and by east–west circulation that moves warm and cold waters around. The normal trade wind forced Pacific Ocean circulation, with surface waters pushed to the west and upwelling driven mainly by Ekman transport away from the equator, results in warm temperatures in the west and cold in the east. When the trade wind strength is reduced, and winds in the western Pacific even reverse, upwelling is reduced and warm water surges to the central Pacific. This reduces the temperature difference between the western and eastern equatorial regions, which in turn changes the strength and shifts the location of the atmosphere s tropical convective cell (the Walker circulation). This shift in the atmosphere further reduces the trade winds hence the cycle has positive feedback. Similar feedbacks between the ocean and atmosphere are being explored for the tropical Atlantic and Indian Oceans.

Within the lowest layer of the atmosphere the ocean has a major impact on local climate, as clearly seen in coastal

regions. Water has a much larger heat capacity than air, which means that water temperature changes in response to gain or loss of a given amount of heat are much smaller than air temperature changes. Ocean temperatures vary within a much smaller range than air temperatures. Coastal climates have much smaller extremes than inland climates because of this relative oceanic stability.

At mid and high latitudes, the direct feedback of the ocean s surface temperature on climate is less marked because deep atmospheric convection is a tropical phenomenon. The general surface temperature difference from equator to pole, much of which is due to the ocean, drives the atmosphere s Hadley circulation with rising air in the tropics and sinking air at mid-latitudes. The large heat loss from the ocean (heat gain by the atmosphere) within the subtropical western boundary current separation regions affects the major storm tracks. Several climate modeling studies have suggested the importance of mid-latitude ocean temperature distributions in the mid-latitude decadal climate oscillations (e.g., Latif and Barnett, 1994). The lure of using ocean circulation to explain these variations is in the match of the time scale of the oscillations to the time scale of the upper ocean circulation. This takes a decade or two depending on ocean width, with the Pacific width and time scales about twice those of the Atlantic. Mid-latitude temperature distributions are affected by the strength of the western boundary currents, by sequestration of temperature anomalies through subduction in the subtropical gyres with reemergence of the anomalies either in the tropics or at the western boundary, and by westward, slow propagation of large-scale planetary waves that carry information about the ocean s thermocline depth and hence upper ocean temperature distribution.

The ocean s various thermohaline circulations help to redistribute the Earth s heat, with the oceans carrying about one-third to one-half of the Earth s heat from the tropics to the poles and the atmosphere carrying the remainder. Where circulation is dominated by wind driving, such as in the North Pacific, almost all of the heat is carried in the shallow wind driven circulation, and is associated with subduction. Where circulation also has a major thermohaline component, and if the thermohaline component is associated with temperature changes of more than about 5 °C, both the wind driven and thermohaline circulations are important for heat transport. Such is the case in the North Atlantic, where there is both major heat loss in the Gulf Stream separation region, associated with the wind driven subtropical gyre, and heat loss in the eastern subpolar gyre and in the Nordic Seas, associated with the deep thermohaline overturning. Warmth is carried northward by the ocean circulation, and the cooled waters return southward in both the subducted upper layers of the subtropical gyre and in the intermediate and deep waters formed farther north.

Changing sea ice distributions (*see* Sea Ice, Volume 1) may have a stronger effect on climate than mid-latitude sea surface temperature changes, through both the much larger temperature contrast between the ice and water than between different waters and by the role of ice in the Earth s albedo (reflectivity). Sea ice distributions are affected to some extent by ocean circulation. A marked example is in the Nordic Seas between Greenland and Norway and north of Iceland, where warm northward flowing North Atlantic waters keep the eastern part of the Nordic Seas ice-free all year round. An increase in the flow would decrease the sea ice area and move the ice edge. During glacial periods, sea ice extended much farther equatorward and the atmospheric circulation would have had significantly different strength as a result.

Within the ocean, climate change, including interannual and decadal variations as well as much longer time scales, has a demonstrable effect on local biological productivity and species distribution. In the North Atlantic for instance, observed changes in temperature in the subpolar gyre over the past 100 years have had major consequences for fisheries, particularly spawning grounds which depend on temperature (Dickson *et al.*, 1988). In the North Pacific, El Niño changes the temperature distribution of the tropical and eastern boundary regions, including flow of warmer waters poleward along the eastern boundaries, which brings tropical species to higher latitudes, and which reduces the efficiency of local upwelling and hence reduces nutrient supply to the local fisheries.

At the western boundary of the North Pacific, the position of the subpolar western boundary current separation (that is, maximum southward intrusion of Oyashio waters) is carefully monitored by Japanese fisheries agencies because of the profound impact of presence or absence of the nutrient rich Oyashio waters. Within the subpolar North Pacific, the position of the subarctic front also has a large impact on fisheries, because of the large-scale upwelling of the subpolar gyre north of the front. The latitude at which the subpolar and subtropical gyre circulations split in the eastern North Pacific impacts the paths followed by returning spawning salmon. (North of Vancouver Island the fish belong to Canadian fisherman and south of Vancouver Island they belong to US fisherman.) Many other examples of how ocean and circulation changes impact biota abound, including in paleooceanographic studies, in which changes in marine species are used to reconstruct past ocean circulations.

The role of the ocean circulation in climate was the main subject of the international World Ocean Circulation Experiment (WOCE), whose field study phase was 1991–1998, and which has sponsored major advances in general circulation modeling and ocean data assimilation (*see* WOCE (World Ocean Circulation Experiment), Volume 1). Because it was recognized that the general circulation, as it exists today, is central to climate, much of WOCE dwelt on describing and modeling the present ocean circulation. WOCE results have been reported in several thousand scientific publications, and

have been summarized in a recent book (Siedler *et al.*, 2001). The WOCE data sets and model advances will be used for decades to come to continue to advance understanding of the general circulation and to provide necessary checks for climate models, which couple all of the relevant systems; atmosphere, ocean, land, and biology.

See also: Atmospheric Motions, Volume 1; Ocean Observing Techniques, Volume 1; Sea Surface Temperature, Volume 1; Sverdrup, Harald Ulrik, Volume 1.

REFERENCES

Broecker, W (1991) The Great Ocean Conveyor, *Oceanography*, 4, 79–89.

Davis, R E (1998) Preliminary Results from Directly Measuring Mid-depth Circulation in the Tropical and South Pacific, *J. Geophys. Res.*, 103, 24 619–24 639.

Dickson, R R, Meincke, J, Malmberg, S-A, and Lee, A J (1988) The Great Salinity Anomaly: In the Northern North Atlantic 1968–1982, *Prog. Oceanogr.*, 20, 103–151.

Ekman, F W (1905) On the Influence of the Earth s Rotation on Ocean Currents, *Arkiv. Matem, Astr. Fys. (Stockholm)*, 2(11), 53.

Firing, E (1987) Deep Zonal Currents in the Central Equatorial Pacific, *J. Mar. Res.*, 45, 791–812.

Latif, M and Barnett, T P (1994) Causes of Decadal Climate Variability over the North Pacific and North America, *Science*, 266, 634–637.

Luyten, J R, Pedlosky, J, and Stommel, H (1983) The Ventilated Thermocline, *J. Phys. Oceanogr.*, 13, 292–309.

McCartney, M S (1977) Subantarctic Mode Water, *Deep-Sea Res.*, 24, Suppl., 103–119.

Mizuno, K and White, W B (1983) Annual and Interannual Variability in the Kuroshio Current System, *J. Phys. Oceanogr.*, 13, 1847–1867.

Munk, W H (1950) On the Wind Driven Ocean Circulation, *J. Meteor.*, 7, 79–93.

Orsi, A H, Johnson, G C, and Bullister, J L (1999) Circulation, Mixing and Production of Antarctic Bottom Water, *Prog. Oceanogr.*, 43, 55–109.

Peterson, R G, Stramma, L, and Kortum, G (1996) Early Concepts and Charts of Ocean Circulation, *Prog. Oceanogr.*, 37, 1–115.

Reid, J L (1994) On the Total Geostrophic Circulation of the North Atlantic Ocean: Flow Patterns, Tracers, and Transports, *Prog. Oceanogr.*, 33, 1–92.

Reid, J L (1997) On the Total Geostrophic Circulation of the Pacific Ocean: Flow Patterns, Tracers, and Transports, *Prog. Oceanogr.*, 39, 263–352.

Reid, J L and Lynn, R J (1971) On the Influence of the Norwegian-Greenland and Weddell Seas upon the Bottom Waters of the Indian and Pacific Oceans, *Deep-Sea Res.*, 18, 1063–1088.

Richardson, P (1980) Benjamin Franklin and Timothy Folger s First Printed Chart of the Gulf Stream, *Science*, 207, 643–645.

Rintoul, S R (1998) On the Origin and Influence of Adelie Land Bottom Water, Ocean, Ice and Atmospheric Interactions at the Antarctic Continental Margin, *Ant. Res. Ser.*, 75, 151–171.

Schmitz, W J (1995) On the Interbasin-scale Thermohaline Circulation, *Rev. Geophys.*, 33, 151–174.

Schmitz, W J (1996) On the World Ocean Circulation: Volume II, The Pacific and Indian Oceans/A Global Update, Woods Hole Oceanographic Institution Technical Report WHOI–96–08.

Schmitz, W J and McCartney, M S (1993) On the North Atlantic Circulation, *Rev. Geophys.*, 31, 29–49.

Siedler, G, Church, J, and Gould, J (2001) *Ocean Circulation and Climate: Observing and Modelling the Global Ocean*, Academic Press, New York.

Sobel, D (1995) *Longitude: The True Story of a Lone Genius who Solved the Greatest Scienti c Problem of his Time*, Walker, New York.

Stommel, H (1948) The Westward Intensification of Wind Driven Ocean Currents, *Trans. Am. Geophys. Union*, 29, 202–206.

Stommel, H M and Arons, A B (1960) On the Abyssal Circulation of the World Ocean, I, Stationary Planetary Flow Patterns on a Sphere, *Deep-Sea Res.*, 6, 140–154.

Sverdrup, H U (1947) Wind Driven Currents in a Baroclinic Ocean; With Application to the Equatorial Currents of the Eastern Pacific, *Proc. Natl. Acad. Sci. USA*, 33, 318–326.

Sverdrup, H U, Johnson, M W, and Fleming, R H (1942) *The Oceans: Their Physics, Chemistry and General Biology*, Prentice Hall, Englewood Cliffs, NJ.

Talley, L D (1995) Some Advances in Understanding of the General Circulation of the Pacific Ocean, with Emphasis on Recent US Contributions, *Rev. Geophys. Suppl.*, 1335–1352.

Talley, L D (1996) North Atlantic Circulation and Variability, Reviewed for the CNLS Conference, *Physica D*, 98, 625–646.

Talley, L D (1999) Some Aspects of Ocean Heat Transport by the Shallow, Intermediate and Deep Overturning Circulations, in *Mechanisms of Global Climate Change at Millennial Time Scales*, eds P U Clark, R S Webb, and L D Keigwin, *Geophys. Mono. Ser.*, 112, American Geophysical Union, 1–22.

Tomczak, M and Godfrey, J S (1994) *Regional Oceanography: an Introduction*, Pergamon, Oxford.

Worthington, L V (1976) *On the North Atlantic Circulation*, Johns Hopkins Oceanographic Studies, 6.

Wust, G (1935) Schichtung und Zirkulation des Atlantisches Ozeans. Die Stratosphare, *Wissenschaftliche Ergebnisse der Deutschen Atlantischen Expedition auf dem Forschungs-und Vermessungsschiff, Meteor 1925- 1927*, 6, 109–288.

Wyrtki, K and Kilonsky, B (1984) Mean Water and Current Structure during the Hawaii-to-Tahiti Shuttle Experiment, *J. Phys. Oceanogr.*, 14, 242–253.

Ocean Conveyor Belt

The global-scale thermohaline circulation of the world ocean has been termed the *ocean conveyor belt*, because it forms a continuous loop carrying water and its properties around the world (Gordon, 1986; Broecker, 1991), see Figure 1. A description may conveniently begin in the

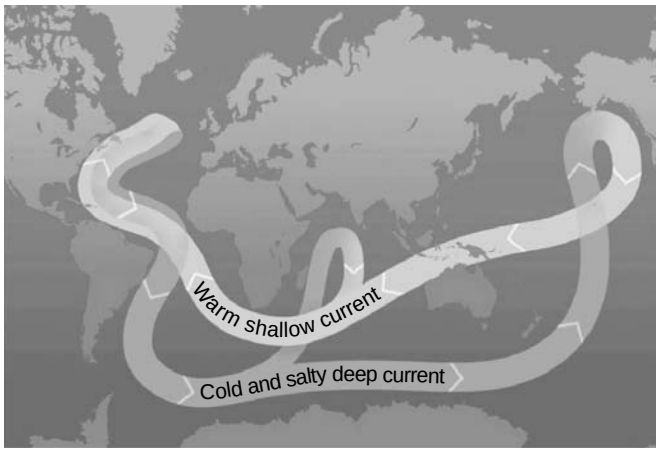


Figure 1 The ocean plays a major role in the distribution of the planet's heat through deep sea circulation. This simplified illustration shows this *conveyor belt* circulation, which is driven by differences in heat and salinity. Records of past climate suggest that there is some chance that this circulation could be altered by the changes projected in many climate models, with impacts to climate throughout lands bordering the North Atlantic. (Redrawn from Broecker, 1991)

North Atlantic, where the well-known Gulf Stream carries northward relatively warm and salty surface waters that are formed, in large part, in the Mediterranean and the subtropical Atlantic Ocean. In the sea between Norway and Greenland, Arctic winds cause evaporation and chilling. The heat and moisture transferred to the air warm Europe and provide its precipitation, particularly during the winter. The waters are also made saltier as a result of brine rejection during ice formation (*see* Ocean Circulation, Volume 1; Salinity Patterns in the Ocean, Volume 1).

The resulting cold, salty, and therefore dense water sinks to become an immense southward-flowing stream of North Atlantic deep water. Because of the Earth s rotation, this deep-ocean current hugs the western edge of the Atlantic basin, running southward under the northward moving Gulf Stream that occupies the upper ocean layer. The cold, deep water continues to flow southward into the South Atlantic, where it joins with even colder, dense, oxygen-rich water sinking from the seas bordering Antarctica to form a pool of circumpolar deep water (*see* Southern Ocean, Volume 1).

Two branches of the global circulation emerge from this Antarctic abyss. One branch flows northward along the eastern coast of Africa to mix with the waters of the Indian Ocean. The other continues around the eastern coast of Australia and then northward, eventually mixing upward with surface waters in the tropical and North Pacific. The surface waters in the Pacific Ocean then move equatorward, then westward, driven by the trade winds through the Indonesian Archipelago, across the South Indian Ocean, around the Cape of Good Hope, and ultimately back into the equatorial Atlantic, where the waters are taken up by the Gulf Stream and begin another round the world, thousand-year excursion.

Because of the heat provided to Europe in the North Atlantic region and the nutrients carried upward as deep waters rise in the tropical and North Pacific, variations in the strength and path of this conveyor system could have major implications for regional and global climates. During glacial periods, as ice covered the North Atlantic, the Gulf Stream shifted location and moved across the Atlantic toward France, thereby contributing to the chilling and glaciation of northern Europe. There is also emerging evidence that a sudden slowdown in the thermohaline circulation contributed to the return of glacial conditions to Europe during the Younger Dryas about 11 000 years ago (*see* Younger Dryas, Volume 1). It has been suggested that the slowdown occurred because of a sudden release of glacial meltwaters just before the Younger Dryas, and that these waters freshened the North Atlantic and thereby stopped the downward push by dense, salty waters that drives the global thermohaline circulation.

That the thermohaline circulation has changed in the past has raised questions about how it might change in the future (e.g., Calvin, 1998). In contrast to the freshening mechanism that caused the Younger Dryas, it is suggested that future freshening could result from the sharp increase in high-latitude precipitation (and ensuing increased river runoff) that are projected as the world warms, and from an increase in the rate of glacial calving from Greenland. Many of the coupled ocean–atmosphere model simulations of potential climate change during the 21st century project a slowing down of the thermohaline circulation (Cubasch and Meehl, 2001), with some indication that the likelihood and magnitude of the reduction in strength increase as the rate of change in carbon dioxide concentration increases. Such a slowdown would reduce the amount of warming over Europe and the North Atlantic. While no models currently show a complete collapse of the thermohaline circulation during the 21st century, there is a possibility that such a flip-flop could occur at later times if the emissions of greenhouse gases continue to grow at high rates.

REFERENCES

Broecker, W S (1991) The Great Ocean Conveyor, *Oceanography*, 4, 79–89.

Calvin, W H (1998) The Great Climate Flip-Flop, *Atl. Month.*, 281(1), 47–64.

Cubasch, U and Meehl, G A (2001) Projections of Future Climate Change, in *Climate Change 2000: The Scienti c Basis*, Intergovernmental Panel on Climate Change, World Meteorological Organization, Cambridge University Press, Cambridge, 525–582.

Gordon, A L (1986) Interocean Exchange of Thermocline Water, *J. Geophys. Res.*, 91, 5037–5046.

JOHN S PERRY *USA*

Ocean Drilling Program

The *Ocean Drilling Program* (ODP) is an international partnership of scientists and research institutions organized to explore the evolution and structure of Earth. ODP provides researchers around the world with access to a vast repository of geological and environmental information recorded far below the ocean surface in seafloor sediments and rocks. By studying ODP data we gain a better understanding of Earth s past, present, and future.

The drill ship, *JOIDES Resolution*, is the centerpiece of the ODP. With this ship, ODP can drill cores (long cylinders of sediment and rock) in water depths up to 8.2 km. Built in 1978 in Halifax, Nova Scotia, the drill ship was originally a conventional oil drilling ship. She was refitted in 1984 and is now equipped with some of the world s finest shipboard laboratories. The ship s complement for each cruise is a mixture of 30 scientists from around the world, 20 engineers and technicians, and a crew (including drilling personnel) of 52.

Each year JOIDES Resolution departs on six scientific expeditions, each approximately 2 months in length. Every expedition has specific scientific goals chosen through a careful review process. The research ship has drilled in the Atlantic, Pacific, Indian, and Arctic Oceans. Since January 1985, ODP has recovered more than 160 000 m of cores. Upon completion of an expedition, the cores are transported to one of four repositories for curation, storage, and future research. ODP scientists are able to use these repositories much as the general public uses a library. For further information, see http://www.oceandrilling.org.

JOHN S PERRY *USA*

Ocean Observing Techniques

Robert A Weller

Woods Hole Oceanographic Institution, Woods Hole, MA, USA

Ocean observations are essential to understanding the ocean's role in climate, particularly how it transports and stores heat, freshwater, and carbon dioxide (CO_2). The principal observations being made are of temperature, salinity, and ocean currents. Vertical pro les and time series are collected using pro lers and moorings so that the transport and storage of heat and freshwater by the ocean can be determined. Additional observations are made of chemical tracers, nutrients, and other quantities to elucidate biological, chemical, and geological as well as physical pathways and processes in the ocean. To date, the ocean has been sparsely sampled, and an international effort is underway to institute comprehensive global ocean observations.

Several characteristics of the ocean guide our choice of what to measure and of the methods chosen when making observations in support of understanding and predicting change in the ocean and how that change impacts the global climate system. First, the ocean is a large reservoir, covering about 70% of the Earth s surface, with great capacity to store and transport heat and climatically important compounds such as carbon dioxide (CO_2). Second, the entire oceanic volume is not, however, in immediate contact with the atmosphere. Generally, a shallow, warm and buoyant layer several tens to one hundred meters thick is found at the surface. This plays an important role in short-term atmosphere–ocean interaction while also insulating the interior of the ocean. Third, the ocean has a complex, three-dimensional circulation that impacts global climate over a wide range of time scales (*see* Ocean Circulation, Volume 1). Horizontal flows in the surface layer can be strong, up to several meters per second, and can move warm water from the tropics to higher latitudes. Surface water can also be exchanged with water in the interior, sinking rapidly in some locations when strongly cooled by the atmosphere, but more typically moving much more slowly along three-dimensional pathways that take several years to exchange surface water with intermediate depths and up to hundreds of years to reach the deepest basins. Finally, the ocean is a challenging environment for observations. Seawater is corrosive. Biofouling and vandalism damage instrumentation. Much of the ocean is remote and seldom visited, and the radio telemetry that we rely on in space exploration cannot be used. Because radio waves do not penetrate, oceanographic remote sensing is largely limited to the use of sound and instruments placed deep in the ocean have had to record their data for later recovery.

The upper ocean has been the focus for most observations, in part motivated by the potential for improvements in weather prediction as well as climate models and also because of ease of access. Observations of the intermediate and deep regions are not routine and have been made mainly during oceanographic research programs. The majority of ocean observations to date have been of the physical variables, temperature and salinity, and of the ocean currents.

Ocean temperature is measured primarily to determine how much heat the ocean stores in what locations and also to combine the temperature with salinity to determine the

density of seawater. Temperatures range as high as about 33 °C at the sea surface under strong sunlight and in low wind and can be as cold as 1–2 °C in the deep basins. The sea surface temperature (SST) provides a measure of the ocean s thermal forcing of the atmosphere and a key surface boundary condition for numerical models of the atmosphere. Observations of the temperature and thickness of the surface mixed layer provide a measure of the heat stored in the ocean that can readily be released to the atmosphere. SST is measured globally to an accuracy of about 0.5 °C using satellite observations of infrared radiation from the sea surface calibrated by and merged with direct measurements by drifting buoys, moored buoys, and merchant ships. These *in situ* measurements have been found to be essential to the interpretation of the satellite data.

Vertical profiles of temperature in the upper ocean are collected from some of the same merchant ships (volunteer observing ships, VOS) that volunteer to drop expendable temperature profilers (known as expendable bathythermographs, XBTs), typically to 750 m depth. Research ships also drop XBTs, which are accurate to about 0.1 °C. To obtain more accurate measurements, oceanographers lower the more precise, reusable profilers, conductivity, temperature, and depth (known as CTDs) down as far as the ocean bottom. Temperature accuracy is 0.005 °C, and the conductivity measurement allows salinity to be determined. Time series of temperature at one point can be collected from moorings, of either a surface buoy or a subsurface buoy. These moored instruments record their data internally, with sampling as frequently as every minute. Such moorings are recovered by a research ship months to years after deployment in order to collect the data and recover the instruments for reuse. Recently efforts have been made to also get some of the data from moorings back more rapidly, and instruments are being developed to transmit temperature data up the mooring line, to be relayed via satellite from the surface buoy or from a buoyant data capsule that floats to the surface.

Salinity and temperature are needed to compute the density of seawater. The temperature and salinity of seawater are also useful as tracers to track the movement of the water. Salinity is more difficult to measure, and the most precise measurements made with CTDs achieve an accuracy of about 0.005 practical salinity units, psu (*see* Salinity Patterns in the Ocean, Volume 1) when the CTD is calibrated with water samples drawn from the ocean when the profile is collected. Many of the same techniques used for measuring temperature can also be used. Expendable profilers (XCTDs) are used less frequently than XBTs due to their greater cost. CTDs deployed from research ships have collected much of the historical salinity data. In recent years, new salinity sensors have led to underway surface salinity observations from merchant ships and the deployment of salinity sensors on moorings. These new, more stable salinity sensors will also be used on a new generation of instruments known as Lagrangian profilers. These profilers are autonomous; they will be released around the world and allowed to drift. They have an internal pump and piston that allows them to change their own density and dive from the surface to about 1500–2000 m depth. They remain there for 1–2 weeks, then become more buoyant and rise to the surface, where temperature and salinity data collected on the way down and the way up are relayed back via satellite and where the float s position is determined by the global positioning system (GPS).

Measurements of ocean currents are needed to understand how and where the ocean transports heat, salinity and water mass (referred to as freshwater). Ocean current measurements also help us understand the dynamics of the ocean and provide benchmarks for developing ocean models. Ocean currents can be measured by tracking drifters. This is commonly done in the surface layer, where a small float little influenced by the surface wind has a large drogue at 15 m depth and where the drifters positions can be tracked by satellite. It is also done by using sound to track floats ballasted to remain at a fixed depth. Ocean currents in the interior of the ocean are often measured by current meters attached to moorings as shown in Figure 1. Acoustic techniques have been developed that use the Doppler shift of sound reflected from moving particles to measure ocean currents. Acoustic Doppler current profilers (ADCPs) mounted to look down from ships and on moorings can obtain velocity profiles over depth ranges of 100–1000 m. In the interior of the ocean, the geostrophic flow (where the pressure gradients are balanced by the Coriolis force) can be calculated from temperature and salinity profiles made using CTDs. A new class of instruments, called moored profilers, combines the measurement capabilities of a CTD with a current meter. These instruments measure salinity, temperature, and velocity as they move up and down along mooring cables.

Additional observations useful for climate studies are of tracers and of quantities involved in biological, chemical, and geological pathways and processes. Helium/tritium and other isotopic ratios are measured, and when there is a well-defined source, provide tracers to better understand the slow, three-dimensional circulation of the ocean. Nutrients (phosphate, silicate) and dissolved oxygen are measured using water samples collected at various depths during research cruises, adding to the evidence used to determine these slow pathways. Efforts are increasing to more routinely measure dissolved CO_2 and particulate fluxes of carbon in the ocean.

The major challenges for understanding the ocean s role in climate are first to adequately sample the global ocean

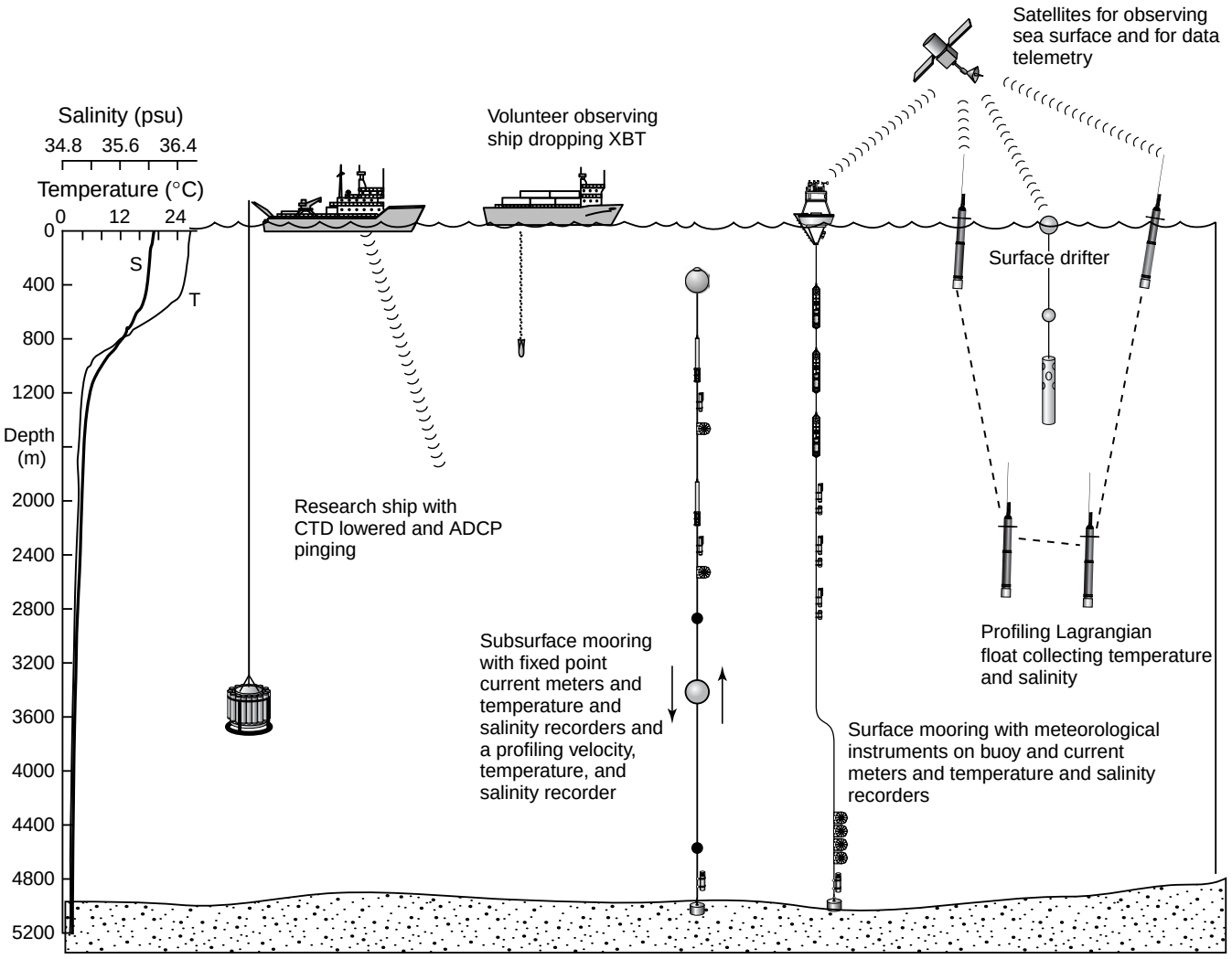


Figure 1 Typical mid-North Atlantic temperature and salinity profiles to the left and various oceanographic sampling techniques to the right. Research ships lower CTDs to obtain accurate temperature and salinity profiles. Research ships also have hull-mounted acoustic Doppler current profilers (ADCPs) to measure ocean currents in the upper ocean. Voluntary observer ships (VOSs) from the merchant fleet drop XBTs to obtain temperature profiles. Subsurface and surface moorings anchored to the bottom can be equipped with current meters to measure ocean currents and salinity and temperature records. Drifting buoys that float on the surface can be fitted with drogues to measure shallow currents and SST. New floats that can dive to 1500–2000 m are being used to collect temperature and salinity profiles that are relayed back via satellite

and second to understand the processes that govern change in the ocean and its role in the climate system. Comprehensive global observations of the ocean are not yet routine. Only a handful of oceanographic time series span more than a decade. As a result, we have a very sparse data base for detecting and attributing changes in ocean temperature and salinity that might result from combinations of natural and anthropogenic variability in the hydrologic cycle, the surface heating of the ocean or changing patterns of surface winds. The need to develop ongoing, operational ocean measurements is one of the motivations for the Global Ocean Observing System (GOOS) now being designed. Present plans call for the deployment of moorings at fixed points and 3000 Lagrangian profilers to obtain new observations to complement the present satellite and VOS measurements. Measurements of acoustic travel time between the fixed moorings may be used to provide integral measures of change in the temperature of the ocean. Efforts will increase to add additional biological, chemical, and geological sensors. New research programs, such as the Climate Variability and Predictability program (CLIVAR) sponsored by the World Climate Research Programme, are being developed to make the more detailed measurements needed to resolve and understand the processes.

See also: CLIVAR (CLImate VARiability and Predictability), Volume 1; EOS (Earth Observing System), Volume 1; Global Ocean Observing System (GOOS), Volume 2.

Ocean, Salinity Patterns

see Salinity Patterns in the Ocean (**Volume 1**)

Oceanography

Broadly defined, *oceanography* comprises the scientific study of all aspects of the ocean, which covers about 70% of the planet s surface. Thus, oceanography is inescapably a huge field of science. Some of the areas of study involved illustrate its diversity: ocean bottom geographic features (depths, volcanoes, trenches, oceanic plateaus, fracture zones, ridges and basins), plate tectonics, chemical composition of ocean water (salt content, carbon dioxide (CO_2) levels, density, etc.), interaction between the ocean and the atmosphere (Coriolis effect, weather systems, hurricanes, monsoons, cyclones, El Niño, etc.), ocean climates, ocean currents, waves, tides, chemical cycling in the ocean, plankton and sediments, and all the diverse varieties of marine life. The interactions between global environmental change and the oceans vividly demonstrate this diversity. For example, understanding the global carbon cycle and its modifications by human activity requires understanding of ocean chemistry (shaped by ocean basin geology, river runoff, etc.), biological processes, turbulent exchange between ocean and atmosphere, and ocean circulation on time scales from days to millennia.

Oceanography has been defined by one oceanographer (Chamberlin, 2001) as an interdisciplinary science aimed at understanding the relationships between physical, geological, chemical and biological processes in the sea. Oceanography therefore involves:

- mathematics and physics – to understand ocean circulation and climate;
- chemistry – to understand the carbon cycle and cycling of other elements;
- biology – to understand marine life and natural resources; and
- geology – to understand records of past climate and the development of the oceans.

Thus, a number of subdisciplines may be identified:

- chemical oceanography: study of the chemical composition and quality of water, and the chemical processes taking place in the ocean;
- geological oceanography: study of the geological and geophysical, and surface and subsurface characteristics of ocean basins and coastal margins;
- marine biology: study of life processes in marine environments; and
- physical oceanography: study of the physical and dynamical properties and processes of the ocean.

Because of the vastness of the ocean, and the difficulty of observing its depths, oceanography is uncommonly dependent on large and expensive observing platforms and systems, and on advanced data processing and computing facilities. Fleets of specialized oceanographic vessels are maintained by many nations to sound the ocean depths, assay its chemistry, and inventory its life. Buoys, both surface and subsurface, free-floating and moored, are increasingly being employed in ocean-wide networks. Satellites provide detailed measurements of sea-surface temperature, wind stress and waves, and indicators of biological activity. Because of the dynamic characteristics of ocean circulation, modeling its circulation requires immense computational power, while huge data archives are needed to manage the vast amounts of data currently being acquired. Thus, very much like space exploration, oceanography demands collaborative work by multi-disciplinary teams employing large and complex observational and research facilities.

REFERENCE

Chamberlin, W (2001) www.oceansonline.com.

JOHN S PERRY *USA*

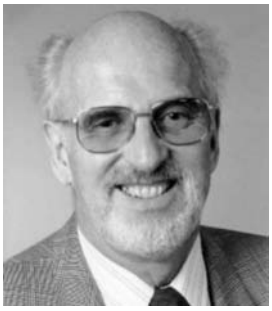
Oceans

see Arctic Ocean (**Volume 1**); Atlantic Ocean (**Volume 1**); Ocean Circulation (**Volume 1**); Southern Ocean (**Volume 1**)

ODP (Ozone Depletion Potential)

see Ozone Depletion Potential (ODP) (**Volume 1**)

Oeschger, Hans A



(1927- 1998)

Hans A Oeschger was a Swiss physicist and paleoclimatologist. His career included: PhD University of Bern, 1955; Venia Docendi, University of Bern, 1962; vice-president of the Swiss National Academy of Sciences, 1975–1981; and Dean of the Faculty of Science, University of Bern, 1978–1979. Professor Oeschger s honors include the Harold C Urey Medal from the European Association of Geochemistry, the Seligman Crystal from the International Glaciological Society, the Tyler Prize for Environmental Achievement, and the Roger Revelle Medal from the American Geophysical Union.

Hans Oeschger trained as an experimental physicist at Eidgenoessische Technische Hochschule (ETH) Zurich. He was greatly influenced by his mentor, Fritz Houtermans, at the University of Bern to follow a career in isotope geochemistry. Houtermans inspired him to reach out beyond the traditional boundaries of physics and, intrigued by Libby s work with the isotopes of carbon and hydrogen, he and Oeschger established a radiocarbon laboratory, the Laboratory for Low Level Counting and Nuclear Geophysics. During his 40-year tenure as a researcher and professor, and together with the splendid team of colleagues that he assembled at the Physics Department at the University of Bern, Professor Oeschger pioneered the application of the techniques of modern physics to the investigation of the Earth System. He and his colleagues developed many innovative techniques for measuring radiocarbon on very small samples of carbon, oxygen isotopes, and radiocarbon dating of ice.

On a visit to La Jolla in 1958, Oeschger interacted with scientists just developing the field of environmental geochemistry. David Keeling had just begun carbon dioxide (CO_2) measurements on Mauna Loa and Hans Suess and Oeschger made the first radiocarbon measurements on bicarbonate samples in water from the deep Pacific Ocean. Oeschger realized the great potential of natural radioactive and stable isotope studies for many scientific fields. With Johannes Geiss he analyzed the radioactivity in meteorites and later also in lunar samples. With Hugo Loosli, he succeeded in measuring the radioactive isotopes ^{37}Ar, ^{39}Ar, and ^{81}Kr in atmospheric samples. The measurements of ^{37}Ar provided information on underground nuclear weapon testing. ^{39}Ar was employed as a tracer for deep ocean circulation and ^{39}Ar and ^{81}Kr analyses helped to determine the age of underground water. With Uli Siegenthaler, he modeled bomb-^{14}C and fossil fuel CO_2 uptake by the ocean using a box diffusion model; and with W Berner and Bernhard Stauffer he made the first measurements of the CO_2 content of glacial age air trapped in ice cores.

In the late 1970s, Oeschger became interested in the issue of the increasing concentration of atmospheric CO_2 and its consequences for global environmental change. He realized the importance of knowing the preindustrial concentration prior to the direct atmospheric measurements that had only recently begun. At a meeting in Austria, Oeschger and Chet Langway found a common interest in the potential of natural ice of known age as an archive for atmospheric gasses and for information on other Earth System phenomena. Together with Willi Dansgaard, a collaboration evolved to sample the world s great ice sheets. They successfully drilled to bedrock at the Dye3 Station in south Greenland and recovered ice samples dating back into the Eemian (*see* Eemian, Volume 1) of the last interglacial. This was the fore-runner of the more extensive deep drilling projects, Greenland Ice Core Project (GRIP) and Greenland Ice Sheet Project 2 (GISP2), at the summit of the Greenland ice sheet in the early 1990s and the recovery of the suite of deep cores from the Russian Vostok Station in Antarctica under the leadership of Claude Lorius.

The drilling techniques and analytical methods that were devised by Oeschger and his colleagues enabled them to establish a history of variation on greenhouse gas concentration, providing an unprecedented reconstruction of climate change over the past 150 000 years, and demonstrating the reality of the abruptness of some past changes in climate. These results stimulated further studies on the mechanisms of climate change, the relationship of climate to Earth s carbon system and inter-hemispheric climate coupling during glacial–interglacial cycles.

Because Oeschger understood that the Earth must be seen as a complex system controlled by its biogeochemical systems, he was instrumental in establishing the International Geosphere–Biosphere Programme (IGBP) to assess the consequences of the human impact on the Earth System. In emphasizing the importance of paleoinformation, he led the way in initiation of the Past Global Changes (PAGES) project, which was charged with producing a quantitative understanding of the Earth s past environmental history and defining the envelope of natural environmental variability.

Oeschger transformed paleoclimatology into a quantitative science that today has a central and essential position within the global change sciences. His career exemplifies innovative research, but he was also a strong advocate for science on the service of society; he considered our planetary environment as his personal responsibility.

HERMAN ZIMMERMAN *USA*

OH–Radical and Atmospheric Cleansing Capacity

see OH–Radical: is the Cleansing Capacity of the Atmosphere Changing? (**Volume 2**)

Orbital Variations

André Berger

Institut d'Astronomie et de Géophysique G. Lemaître, Louvain-la-Neuve, Belgium

The amount of incoming solar radiation received at any point on the Earth has an annual periodic variation caused by the Earth's elliptic translation motion around the Sun. In addition, the seasonal and latitudinal distributions of this solar radiation have a long-period variation caused by the so-called secular variations in the astronomical elements. These are: the eccentricity, e, *a measure of the shape of the Earth's orbit around the Sun; the obliquity,* ε, *the tilt of the equator on the Earth's orbit; and the climatic precession,* $e \sin \tilde{\omega}$, *a measure of the distance from Earth to Sun at the summer solstice (*$\tilde{\omega}$ *being the longitude of the perihelion measured from the moving equinox). As determined from celestial mechanics, the secular variations of these elements of the Earth's orbit and rotation are due to the gravitational perturbations which the Sun, the other planets, and the Moon exert on the Earth. These astronomical elements form the basis of any astronomical theory of paleoclimates for explaining the quaternary glacial- interglacial cycles (*see *Quaternary, Volume 1). Actually, the long-term variations of the incoming solar radiation are pacing climatic changes at time scales of tens to hundreds of thousands of years. The exceptional strength of the 100-kyr cycle requires some non-linear ampli cation through feedback mechanisms related to the albedo, water vapor and isostatic rebound, in particular.*

Astronomical computations allow reconstruction of past (and prediction of future) variations with a good accuracy. The short-term variations of the astro-climate elements (e, ε, $e \sin \tilde{\omega}$) have been computed for the past few millennia (Loutre *et al.*, 1992) using one of the classical secular planetary theories built for the ephemerides by Bretagnon (Paris) and used in astronomical observations (another well-known expression was calculated by Simon Newcomb in 1898 for the eccentricity as a function of time). Spectral analyses of these astronomical parameters for the past 5000 years show periodicities ranging from 2.67 to 250 years. Particularly significant for the eccentricity and climatic precession are the periods of 2.67, 3.98, 5.93 and 11.9 years. Obliquity shows, in addition, periods of 29 and 18.6 years, the last one arising from the nutations related to the Moon and well known in the theory of the Earth s tides. The amplitude of the induced variations in the incoming solar radiation are, however, quite small, reaching a maximum of $0.2\,\mathrm{W\,m^{-2}}$ at these frequencies.

For the long-term variations over the last few million years, the equations for the motion of the planets around the Sun, for the precession, and for the obliquity are integrated, and their numerical solutions are expressed in trigonometrical form as quasi-periodic functions of time:

$$e = e^* + \sum E_i \cos\bullet\lambda_i t + \varphi_i\bullet \qquad \bullet 1\bullet$$

$$\varepsilon = \varepsilon^* + \sum A_i \cos\bullet\gamma_i t + \zeta_i\bullet \qquad \bullet 2\bullet$$

$$e \sin \tilde{\omega} = \sum P_i \sin\bullet\alpha_i t + \eta_i\bullet \qquad \bullet 3\bullet$$

where the amplitudes P_i, A_i, E_i, the frequencies α_i, γ_i, λ_i and phases η_i, ζ_i, φ_i are given in Berger (1978).

In the term $e \sin \tilde{\omega}$, the amplitude of $sin\tilde{\omega}$ is modulated by *eccentricity*, and the envelope of $e \sin \tilde{\omega}$ is given exactly by e because the frequencies of e originate strictly from combinations of the frequencies of $\tilde{\omega}$. Figure 1 shows the long-term variations of these three astronomical parameters over the past 200 000 years and the next 130 000 years.

Over the past 3×10^6 years, the eccentricity of the orbit varies between near circularity ($e = 0$) and slight ellipticity ($e = 0\bullet07$) at a period with a mean of about 100 000 years. The most important terms in the series expansion occur, however, at 413 000, 95 000, 124 000, 99 000, and 131 000 years (in decreasing order of amplitude). The tilt of the Earth s axis varies between about 22 and 25° at a period of nearly 41 000 years. Although this period corresponds in Equation (2) to the amplitude that is by far the largest, there are two other important terms, one with a period of 54 000 and one with 29 000 years.

As far as *precession* is concerned, two motions must be considered. The first is the axial precession, in which the torque of the Sun, the Moon, and the planets on the Earth s equatorial bulge causes the axis of rotation to wobble like that of a spinning top. The net effect is that the North Pole describes a clockwise circle in space (provided the *nutations* – the terms with small amplitude and much smaller periods – are neglected), with a period of about 25 800 years corresponding to the period of the vernal equinox against a fixed reference point. The second is related to the fact that the elliptical figure of the Earth s orbit is itself rotating counter-clockwise in the same plane, leading to an absolute motion of the perihelion with a period, measured relative to the fixed stars, of about 100 000 years (the same as that of the eccentricity).

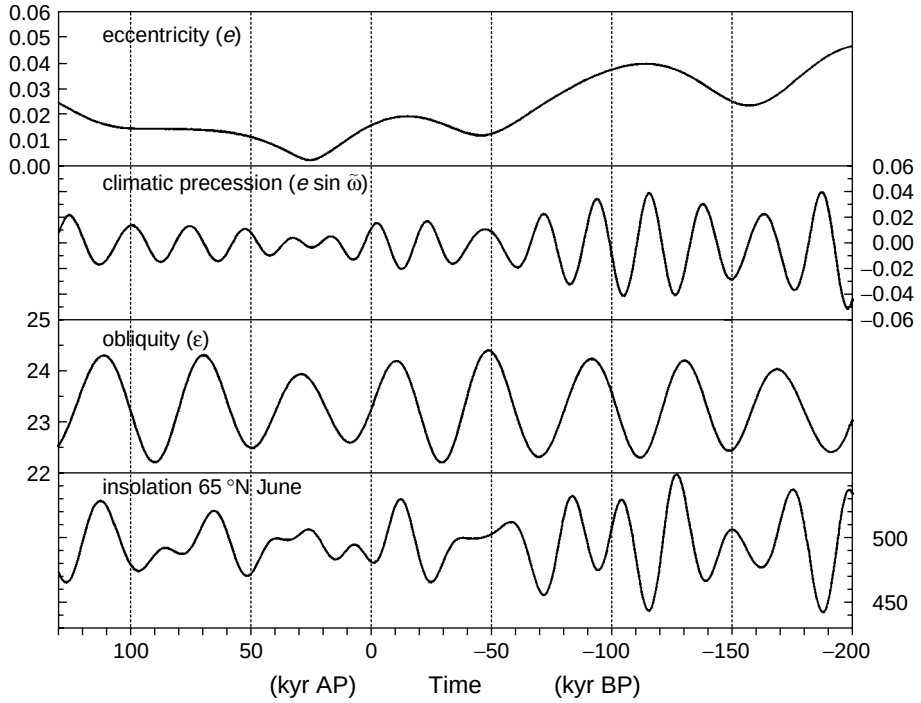


Figure 1 Long-term variations of eccentricity, climatic precession, obliquity and June insolation at 65 °N from 200 000 years ago to 130 000 years in the future (Berger, 1978). This latitude is taken as a typical example. According to Milankovitch (*see* **Milankovitch, Milutin**, Volume 1), it is indeed a key latitude where most of the ice sheets develop in the Northern Hemisphere. Time is going from the right (the past 200 kyr) to the left (the next 130 kyr) according to the geological tradition. The very particular future behavior of the orbital parameters and insolation can lead to an exceptionally long interglacial lasting up to 50 kyr AP, even if humans were not having an influence. In a scenario where the CO_2 concentration reaches 750 ppmv within the next two centuries as a result of human activities, the Greenland ice sheet might melt after 6 kyr, the human influence on the global climate disappearing only after 50 kyr AP (Loutre and Berger, 2000)

The two effects together result in what is known as the climatic precession of the equinoxes, a motion mathematically described by $\tilde{\omega}$, and in which the equinoxes and solstices shift slowly around the Earth s orbit relative to the perihelion with a mean period of 21 000 years (this period actually results from the existence of two close periods of 23 000 and 19 000 years).

Therefore, while today the winter solstice occurs near perihelion, some 11 000 years ago it occurred near aphelion. Moreover, because the lengths of the seasons vary in time according to Kepler s second law, the solstices and equinoxes have occurred at different calendar dates during the geological past. Presently in the Northern Hemisphere, the longest seasons are spring (92 days, 19 h) and summer (93 days, 15 h); fall (89 days, 20 h) and winter (89 days) are definitively the shortest. In about 1250 AD spring and summer had exactly the same length as fall and winter because the winter solstice was occurring at the perihelion. About 4500 years from now, the Northern Hemisphere spring and winter will have the same shorter length and, consequently, summer and fall will be equally long.

The combined influence of changes in e, ε, and $e \sin \tilde{\omega}$ produces a complex pattern of insolation variations. A detailed analysis of the daily solar irradiance (Berger *et al.*, 1993) shows that it is principally affected by variations in *precession*, although the *obliquity* plays a more important role as latitudes increase, mainly in the winter hemisphere. At the equinoxes, insolation for each latitude is only a function of precession. At the solstices, both precession and obliquity are present, although precession dominates for most latitudes.

Changes in incoming solar radiation caused by changes in tilt are the same in both hemispheres during the same local season: an increase of ε leads to an increase in insolation in the summer hemisphere and to a decrease in the winter hemisphere. Because the strength of the effect is small in the tropics and maximal at the pole, an increase in obliquity tends to amplify the seasonal cycle in the high latitudes of both hemispheres simultaneously.

The precession effect can cause warm winters and cool summers in one hemisphere while doing the opposite in the other. For example, portions of the present winter in the Northern Hemisphere receive as much as 10%

more insolation than they did 11 000 years ago, when the perihelion occurred in the Northern Hemisphere summer.

The spectral power of the long-term variations of different insolation parameters depends on the type of parameters considered. For example, the total energy received during any astronomical season for a given latitude is only a function of obliquity (as a consequence of the second law of Kepler). This implies that the total amounts of energy received at a given latitude in the Northern and Southern Hemispheres during the same season, are exactly out of phase, but in phase if the local season is considered. If the mean irradiance is considered, precession again starts to play an important role depending upon the latitude (this is because the length of the astronomical seasons is only a function of precession).

Many other insolation parameters can be elaborated (like the latitudinal gradient and the seasonal contrasts), all displaying a spectrum with different precession and obliquity powers. In particular, the total energy received by the Earth over a full year varies according to the eccentricity only, being maximum for an elongated orbit and minimum for a circular one. In the intertropical regions, because the Sun passes overhead twice a year, precession can lead there to a cycle about 10 kyr long.

But the theoretical frequencies in Equations (1)–(3), are not stable in time (Berger *et al.*, 1998). For eccentricity, the 400-kyr period is weak before 1 Myr BP, it strengthens starting around 900 kyr BP and is particularly strong over the next 400 kyr, when the components in the 100-kyr band will be exceptionally weak. This 400-kyr cycle plays a very important role because the amplitude of precession is modulated by eccentricity. Minima corresponding to this 400-kyr cycle being very low ($e \sim 0$), insolation remains about constant during these times, allowing the other factors to become more important. For *obliquity*, the main period is quite stable, but the spectra of the amplitude and frequency modulations display significant power at 191 and 97 kyr. (Obliquity is the angle between the equatorial plane and the plane of the orbit of the Earth around the Sun (the ecliptic). It is also the tilt of the Earth axis of rotation at an angle, presently $23^{\circ}27'$, away from a vertical drawn to the plane of the orbit.) However, these modulations explain very little (less than 2%) of the variance of insolation (at 65°N in June, for example) over the last million years. On the other hand, the ice ages themselves might have had an influence on the astronomical parameters (Peltier and Jiang, 1994). The waxing and waning of the ice sheets might indeed have altered the Earth s dynamical flattening and so the axial tilt. However, this ice sheet effect is partly cancelled by viscous compensation in the mantle, so that the net effect remains unknown.

For pre-Quaternary times in the Cenozoic, the chaotic nature of the orbital motion of the planets in the solar system limits the possibility of obtaining an accurate solution over more than 35–50 Myr (Laskar, 1999). For earlier times, sensitivity analyses of the astronomical frequencies has suggested (Berger *et al.*, 1992) that the shortening of the Earth–Moon distance and of the length of the day would induce a shortening of the fundamental periods for obliquity and climatic precession. Over the last half-billion years, these periods would accordingly change, respectively, from 54 to 35, 41 to 29, 23 to 19 and 19 to 16 thousand of years. Over such remote times and earlier, the obliquity of the Earth s orbit might have been chaotic with very large variations.

The orbital hypothesis was first quantitatively formulated by the Serbian astronomer Milutin Milankovitch in the 1920s and 1930s (*see* Milankovitch, Milutin, Volume 1) to explain the Quaternary s long-term climatic variations. He argued that insolation changes in the high northern latitudes during the summer season were critical to the formation of continental ice sheets. During periods when insolation in the summer was reduced, the snow of the previous winter would tend to be preserved – a tendency that would be enhanced by the high albedo of the snow and ice fields. Eventually, the effect of this positive feedback would lead to the formation of ice sheets.

A simple linear version of the Milankovitch model would therefore predict that the total ice volume and climate over the Earth would vary with the same regular pattern as the insolation. This means that the proxy record of climate variations would reflect those frequencies of the astronomical parameters that are responsible for changing the seasonal and latitudinal distribution of incoming solar radiation. Investigations during the past 20 years have indeed demonstrated that the 19 000, 23 000, and 41 000 year periodicities actually occur in long records of the Quaternary climate (Hays *et al.*, 1976), that the climatic variations observed in these frequency bands are linearly related to the orbital forcing functions, and that there is a fairly consistent phase relationship among insolation, sea–surface temperature, and ice volume (Imbrie *et al.*, 1992). The geological observation of the bipartition of the precessional peak that was confirmed to be real in astronomical computations (Berger, 1978), was one of the first and most impressive of all tests for the Milankovitch theory. The same investigations, however, identified the largest climatic cycle as 100 000 years, at least over the last 1 million years. Because the eccentricity signal is very weak in the insolation, it cannot be related to orbital forcing by any simple linear mechanism. Actually, the variance components centered near this 100 000-year cycle seem to be in phase with the eccentricity cycle, but its exceptional strength requires some non-linear amplification by the glacial ice sheets themselves, involving mechanisms such as ice albedo feedback and isostatic rebound of the lithosphere, by the carbon dioxide atmospheric concentration, or by the oceanic circulation. It might also be that other astronomical parameters

must be looked for, like the inclination of the Earth s orbit on a reference plane or the axial tilt although careful analysis shows that that is much less probable (Ridgwell *et al.*, 1999).

For the last few millions of years, for which a reasonably reliable astronomical solution exists, the orbital parameters provide a clock with which to date old sediments with a precision several times greater than permitted by physical techniques. Instead, for the last 2–3 tens of millions of years, the climatic cycles are used to create a time scale by assuming that the observed quasi-cyclicity is a response to astronomical forcing (Shackleton *et al.*, 1999). Finally, careful analysis of much older proxy records (even hundreds of millions of years ago) reflects also the astronomically induced climatic variations.

See also: Climate Model Simulations of the Geological Past, Volume 1; Earth System History, Volume 1.

REFERENCES

Berger, A (1978) Long-term Variations of Daily Insolation and Quaternary Climatic Changes, *J. Atmos. Sci.*, 35(12), 2362–2367.

Berger, A, Loutre, M F, and Laskar, J (1992) Stability of the Astronomical Frequencies over the Earth s History for Paleoclimate Studies, *Science*, 255, 560–566.

Berger, A, Loutre, M F, and Tricot, C (1993) Insolation and Earth s Orbital Periods, *J. Geophys. Res.*, 98, 10 341–10 362.

Berger, A, Loutre, M F, and Mélice, J L (1998) Instability of the Astronomical Periods from 1.5 Myr BP to 0.5 Myr AP, *Paleoclimates Data and Modelling*, 2(4), 239–280.

Hays, J D, Imbrie, J, and Shackleton, N J (1976) Variations in the Earth s Orbit: Pacemaker of the Ice Ages, *Science*, 194, 1121–1132.

Imbrie, J, Boyle, E A, Clemens, S C, Duffy, A, Howard, W R, Kukla, G, Kutzbach, J, Martinson, D G, McIntyre, A, Mix, A C, Molfino, B, Morley, J J, Peterson, L C, Pisias, N G, Prell, W L, Raymo, M E, Shackleton, M J, and Toggweiler, J R (1992) On the Structure and Origin of Major Glaciation Cycles. 1. Linear Responses to Milankovitch forcing, *Paleoceanography*, 7(6), 701–738.

Laskar, J (1999) The Limits of Earth Orbital Calculations for Geological Time-scale Use, Astronomical (Milankovitch) Calibration of the Geological Time Scale, eds N J Shackleton, I N McCave, and G P Weedon, *Philos. Trans. R. Soc. London, Ser. A*, 357, 1735–1759.

Loutre M F, Berger A, Bretagnon P, and Blanc, P L (1992) Astronomical Frequencies for Climate Research at the Decadal to Century Time Scale, *Climate Dyn.*, 7, 181–194.

Loutre, M F and Berger, A (2000) Future Climatic Changes: Are we Entering an Exceptionally Long Interglacial? *Climatic Change*, 46, 61–90.

Peltier, W R and Jiang, X (1994) The Precession Constant of the Earth: Variations Through the Ice Age, *Geophys. Res. Lett.*, 21, 2299–2302.

Ridgwell, A J, Watson, A J, and Raymo, M E (1999) Is the Spectral Signature of the 100 kyr Glacial Cycle Consistent with a Milankovitch Origin? *Paleoceanography*, 14(4), 437–440.

Shackleton, N J, McCave, I N, and Weedon, G P, eds (1999) Astronomical (Milankovitch) Calibration of the Geological Time Scale, *Philos. Trans. R. Soc. London, Ser. A*, 357(1757), 1731–2007.

Oscillation, Arctic

see Arctic Oscillation (**Volume 1**)

Oscillation, El Niño/Southern

see El Niño/Southern Oscillation (ENSO) (**Volume 1**)

Oscillation, Madden-Julian

see Madden–Julian Oscillation (**Volume 1**)

Oscillation, North Atlantic

see North Atlantic Oscillation (**Volume 1**)

Oscillation, Pacific–Decadal

see Pacific–Decadal Oscillation (**Volume 1**)

Oscillation, Quasi–Biennial

see Quasi–Biennial Oscillation (QBO) (**Volume 1**)

Oscillation, Quasi–Decadal

see Quasi–Decadal Oscillation (**Volume 1**)

Oscillation, Southern

see Southern Oscillation (**Volume 1**)

Ozone Chemistry, Stratosphere

see Stratosphere, Chemistry (**Volume 1**)

Ozone Chemistry, Troposphere

see Troposphere, Ozone Chemistry (**Volume 1**)

Ozone Depletion Potential (ODP)

Ozone depletion potentials (ODPs) provide a relative measure of the cumulative impact on stratospheric ozone from trace gas emissions, and are used by policymakers to examine the relative effects on ozone from chlorofluorocarbons (CFCs), halons, and other halocarbons. The ODP concept is also used to evaluate the potential effects of possible replacement compounds. Emissions of gases containing chlorine, bromine, or iodine can particularly affect stratospheric ozone (but not the emissions of water-soluble gases like hydrochloric acid, which would likely be washed out of the atmosphere and not reach the stratosphere). Chlorine, bromine, and iodine react extremely efficiently with ozone in catalytic processes, and these atoms can destroy thousands of ozone molecules if they reach the stratosphere. The ODP concept arose as a means of determining the relative ability of a chemical to destroy ozone. If a compound does not contain chlorine, bromine or iodine, then it is less likely to affect the stratosphere. Exceptions could occur if the compound can achieve such significant use that it would become a source of stratospheric nitrogen oxides, hydrogen oxides, or sulfuric particles.

The ODP of a gas is defined as the integrated change in total ozone per unit mass emission of the gas, relative to the change in total ozone per unit mass emission of CFC-11, one of the gases of most concern to ozone change. As a relative measure, ODPs are subject to fewer uncertainties than estimates of the absolute percentage of ozone depletion caused by different gases. As a result, ODPs are an integral part of national and international considerations on ozone-protection policy, including the Montreal Protocol and its amendments, and the United States Clean Air Act. Table 1 shows ODPs for several of the gases of most concern to ozone depletion. Replacement compounds like hydrochlorofluorocarbons (HCFCs) and hydrofluorocarbons (not shown because their ODPs are close to zero) have much smaller ODPs than the CFCs and halons.

Table 1 Steady-state ODPs[a]

Gas	Formula	ODP
CFCs		
CFC-11	CCl_3F	1.0
CFC-12	CCl_2F_2	0.82
CFC-113	CCl_2FCClF_2	0.90
CFC-114	$CClF_2CClF_2$	0.85
CFC-115	CF_3CClF_2	0.40
Bromocarbons		
Methyl bromide	CH_3Br	0.37
Halon-1301	CF_3Br	12.0
Halon-1211	CF_2ClBr	5.1
HCFCs		
HCFC 22	$CHClF_2$	0.034
HCFC-123	CF_3CHCl_2	0.012
HCFC-124	CF_3CHClF	0.026
HCFC-141b	CH_3CCl_2F	0.086
HCFC-142b	CH_3CClF_2	0.043
HCFC-225ca	$CF_3CF_2CHCl_2$	0.017
HCFC-225cb	$CClF_2CF_2CHClF$	0.017
Others		
Carbon tetrachloride	CCl_4	1.20
Methyl chloroform	CH_3CCl_3	0.11
Methyl chloride	CH_3Cl	0.02

[a] *Source:* World Meteorological Organization (1999).

See also: Depletion of Stratospheric Ozone, Volume 1; Stratosphere, Chemistry, Volume 1.

REFERENCE

World Meteorological Organization (1999) *Scienti c Assessment of Ozone Depletion: 1998*, Report No. 44, Geneva, 1999.

DONALD J WUEBBLES *USA*

Ozone Hole

The term *ozone hole* arose in the media following the discovery in 1985 of the large stratospheric ozone depletion that has been occurring during the springtime (September–November) over much of the Antarctic continent since the late 1970s. Joseph Farman and colleagues first documented this rapid springtime decrease in Antarctic ozone over their British Antarctic Survey (BAS) station at Halley Bay, Antarctica. These analyses were confirmed by satellite data and it was soon found that decreases in the total ozone column were greater than 50% compared with historical values observed by both ground-based and

satellite techniques. Measurements have found 60–90% decreases in monthly average ozone amounts between 12 and 20 km, for the springtime period, relative to pre-ozone-hole levels, with essentially all the ozone at these altitudes destroyed at times. Since 1992, the Antarctic ozone holes have been the biggest (in areal extent) and the deepest (in terms of minimum amounts of ozone overhead), with ozone being locally depleted by more than 99% between about 14 and 19 km in October (*see* Stratosphere, Ozone Trends, Volume 1). There is no indication yet that the magnitude of the Antarctic ozone hole has peaked or begun to decline.

The cause of the ozone hole is as follows:

- Air over the Antarctic becomes extremely cold during the winter as a result of the lack of sunlight over the polar region and because of greatly reduced mixing of the lower stratospheric air over this region with air outside this region.
- During the winter, a circumpolar vortex, also called the *polar winter vortex*, forms which isolates the air in the polar region from that outside of the region as a result of a stratospheric jet of wind circulating between approximately 50 °S and 65 °S.
- The extremely cold temperatures inside the vortex lead to the formation of clouds in the lower stratosphere (from roughly 12 to 22 km), called *polar stratospheric clouds*.
- Reactions occur on the cloud particles that convert less reactive forms of chlorine and bromine gases to much more reactive ones that then can react to destroy ozone.
- When daylight starts to occur over Antarctica in the early spring, this chlorine and bromine is available to react with and destroy ozone. The ozone destruction continues until the polar vortex breaks up, usually in November.

In recent years, it has been discovered that these processes also occur in the Northern Hemisphere, although the weaker strength of the arctic polar vortex generally has resulted in much smaller ozone depletions. However, total ozone column reductions of up to 25% and local destruction of ozone of 70% were discovered during the 1999–2000 winter and early spring.

See also: Depletion of Stratospheric Ozone, Volume 1.

DONALD J WUEBBLES *USA*

Ozone, Montreal Protocol

see Ozone Layer: Vienna Convention and the Montreal Protocol (**Volume 4**)

Ozone, Stratospheric Depletion

see Depletion of Stratospheric Ozone (*Opening essay*, **Volume 1**); Stratosphere, Chemistry (**Volume 1**); Stratosphere, Ozone Trends (**Volume 1**)

Ozone, Tropospheric

see Troposphere, Ozone Chemistry (**Volume 1**)

P

Pacific–Decadal Oscillation (PDO)

Nathan J Mantua
University of Washington, Seattle, WA, USA

The Paci c (inter) Decadal Oscillation (PDO) is an El Niño-like oscillatory pattern of climate variability centered over the Paci c Ocean and North America. Over North America, PDO variations are most energetic in the boreal winter and spring. The PDO has considerable in uence on climate-sensitive natural resources in the Paci c and over North America, including the water supplies and snow pack in some selected regions in North America, and major marine ecosystems from coastal California north to the Gulf of Alaska and the Bering Sea. In fact, it was sheries scientist Steven Hare that coined the term Paci c Decadal Oscillation (PDO) in 1996 while researching connections between Alaska salmon production cycles and Paci c climate (Hare, 1996).

Three main characteristics distinguish PDO from El Niño/Southern Oscillation (ENSO): rst, 20th century PDO events persisted for 20- 30 years, while typical ENSO events persisted for 6 to 18 months; second, the climatic ngerprints of the PDO are most visible in the North Paci c/North American sector, while secondary signatures exist in the tropics, while the opposite is true for ENSO; and third, the mechanisms that cause PDO are not currently known, while causes for ENSO are relatively well-understood (Mantua et al., *1997; Zhang* et al., *1997).*

From a societal impacts perspective, recognition of PDO is important because it shows that normal climate conditions can vary over time periods comparable to the length of a human's lifetime.

Typical surface climate anomaly patterns for warm and cool phases of the PDO are shown in Figure 1. In warm PDO phases (Figure 1a), sea surface temperatures (SSTs) tend to be anomalously cool in the central North Pacific coincident with anomalously warm SSTs along the west coast of the Americas. For October–March, warm PDO sea level pressure (SLP) anomalies vary in a wave-like pattern. Low pressures over the North Pacific cause enhanced counterclockwise winds over the North Pacific, and high SLP over western North America and the subtropical Pacific cause enhanced clockwise winds in those regions. PDO circulation anomalies extend through the depth of the troposphere, and are well expressed as persistence in the Pacific North America (PNA) teleconnection pattern described by Wallace and Gutzler (1981) (*see* Paci c- North American (PNA) Teleconnection, Volume 1). Cool PDO climate anomalies are simply opposites of those for warm PDO phases (Figure 1b).

Tracking PDO variations is typically done with indices constructed from observed Pacific SST and SLP patterns. A representative pair of PDO indices are shown in Figure 2

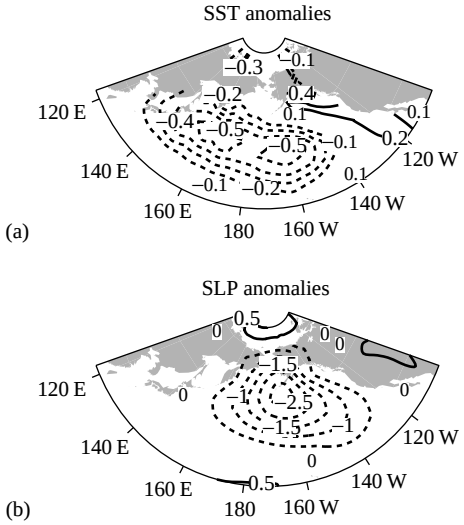


Figure 1 Warm phase PDO surface climate anomalies: (a) sea surface temperature anomaly pattern, dashed contours depict cooler than average temperatures, while solid contours reflect warmer than average temperatures, contour interval is 0.1 °C; (b) atmospheric SLP anomaly pattern, dashed contours depict lower than average SLPs, while solid contours reflect higher than average pressures, contour interval is 0.5 h Pa

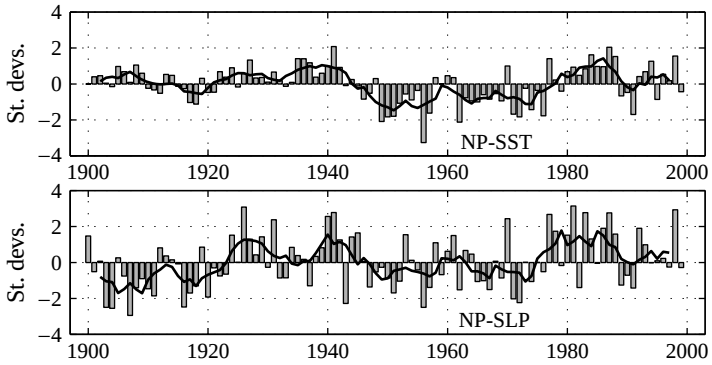


Figure 2 October–March averaged PDO indices based upon projections of observed North Pacific SST and SLP patterns onto those shown in Figure 1. Positive values indicate warm phases of PDO, while negative values indicate cool phases of PDO. Index values are normalized for the entire period of record shown. Solid curve shows a 5-year running average for each time series. October–March averages are used because year to year PDO fluctuations are most energetic during these months

(see Trenberth and Hurrell, 1994; Mantua *et al.*, 1997). When SSTs are anomalously cool in the interior North Pacific and warm along the Pacific Coast, and when SLPs are below average over the North Pacific, the two indices have positive values, indicating warm phase PDO conditions. A notable feature of the indices is their tendency for year to year persistence, with positive/warm or negative/cool index values tending to prevail for 20–30 year periods. However, within the 20–30 year regimes there are several short-lived sign reversals in the indices, including three year reversals from 1959–1961 and again from 1989–1991. Note that these (and other) indices depict two full PDO cycles in the 20th century: cool PDO regimes prevailed from 1890–1924 and again from 1947–1976, while warm PDO regimes dominated from 1925–1946 and from 1977 to (at least) the mid-1990s (Mantua *et al.*, 1997; Minobe, 1997). Minobe (1999) has shown that 20th century PDO fluctuations were most energetic in two general periodicities, one from 15–25 years, and the other from 50–70 years.

The North American climate anomalies associated with PDO are broadly similar to those connected with El Niño and La Niña, though generally not as extreme (Latif and Barnett, 1996) (*see* El Niño and La Niña: Causes and Global Consequences, Volume 1). The warm phase of PDO is correlated with anomalously warm and dry winter/spring conditions in the northern half of North America and anomalously wet conditions over the southern US and Northern Mexico (El Niño-like). The cool phase of PDO is correlated with the opposite (La Niña-like) climate patterns over North America. PDO variability is strongly expressed in regional snow pack and stream flow anomalies, especially in western North America (see Cayan, 1996; and Mantua *et al.*, 1997), and appears to influence summer rainfall and drought in the US (Nigam *et al.*, 1999). A summary of major climate anomalies associated with PDO is given in Table 1.

Combining PDO and ENSO information may enhance the skill of empirical North American climate forecasts. Empirical studies suggest that ENSO influences on North American climate are strongly dependent on the phase of the PDO, such that the canonical El Niño and La Niña patterns of North American climate anomalies are most likely to occur during years in which ENSO and PDO

Table 1 Summary of Pacific and North American climate anomalies associated with extreme phases of the PDO

Climate anomalies	Warm phase PDO	Cool phase PDO
Ocean surface temperatures in the northeastern and tropical Pacific	Above average	Below average
October–March northwestern North American air temperatures	Above average	Below average
October–March Southeastern US air temperatures	Below average	Above average
October–March southern US/Northern Mexico precipitation	Above average	Below average
October–March Northwestern North America and Great Lakes precipitation	Below average	Above average
Northwestern North American spring time snow pack and water year (October–September) stream flow	Below average	Above average

extremes are in phase (i.e., with warm PDO coincident with El Niño, and cool PDO coincident with La Niña, but not with other combinations) (Gershunov and Barnett, 1998).

Important changes in Northeast Pacific marine ecosystems have been correlated with the PDO (Francis *et al.*, 1998). For example, warm PDO phases have favored high salmon production in Alaska and low salmon production off the west coast of California, Oregon, and Washington states. Conversely, cool PDO eras have favored low salmon production in Alaska and relatively high salmon production for California, Oregon, and Washington (Hare, 1996; Hare *et al.*, 1999).

Causes for the PDO are not currently known. Some climate simulation models produce PDO-like oscillations (e.g., Latif and Barnett, 1994), although often for different reasons (NRC, 1998). The mechanisms giving rise to PDO will determine whether skillful PDO climate predictions for up to one or more decades into the future are possible. Even in the absence of a theoretical understanding, PDO climate information improves season-to-season and year-to-year climate forecasts for North America because of its strong tendency for multi-season and multi-year persistence.

Much controversy exists over how PDO works, and how it might best be monitored, modeled and predicted. The stakes in PDO science are high, as an improved PDO understanding offers sharper views of future climate than those now provided by ENSO science alone.

See also: El Niño and La Niña: Causes and Global Consequences, Volume 1; Natural Climate Variability, Volume 1.

REFERENCES

Cayan, D R (1996) Interannual Climate Variability and Snowpack in the Western United States, *J. Clim.*, 9, 928–948.

Francis, R C, Hare, S R, Hollowed, A B, and Wooster, W S (1998) Effects of Interdecadal Climate Variability on the Oceanic Ecosystems of the Northeast Pacific, *Fish. Oceanogr.*, 7, 1–21.

Gershunov, A and Barnett, T P (1998) Interdecadal Modulation of ENSO Teleconnections, *Bull. Am. Meteorol. Soc.*, 79, 2715–2726.

Hare, S R (1996) Low Frequency Climate Variability and Salmon Production, PhD dissertation, School of Fisheries, University of Washington, Seattle, WA.

Hare, S R, Mantua, N J, and Francis, R C (1999) Inverse Production Regimes: Alaska and West Coast Pacific Salmon, *Fisheries*, 24, 6–14.

Latif, M and Barnett, T P (1994) Causes of Decadal Climate Variability Over the North Pacific and North America, *Science*, 266, 634–637.

Latif, M and Barnett, T P (1996) Decadal Climate Variability Over the North Pacific and North America: Dynamics and Predictability, *J. Clim.*, 9, 2407–2423.

Mantua, N J, Hare, S R, Zhang, Y, Wallace, J M, and Francis, R C (1997) A Pacific Interdecadal Climate Oscillation with Impacts on Salmon, *Bull. Am. Meteorol. Soc.*, 78, 1069–1079.

Minobe, S (1997) A 50–70 Year Climatic Oscillation Over the North Pacific and North America, *Geophys. Res. Lett.*, 24, 683–686.

Minobe, S (1999) Resonance in Bidecadal and Pentadecadal Climate Oscillations Over the North Pacific: Role in Climatic Regime Shifts, *Geophys. Res. Lett.*, 26, 855–858.

NRC (National Research Council) (1998) *Decade-to-Century Scale Climate Variability and Change: a Science Strategy*, National Academy Press, Washington, DC (http://www.nap.edu).

Nigam, S, Barlow, M, and Berbery, E H (1999) Analysis Links Pacific Decadal Variability to Drought and Streamflow in the United States, *EOS, Trans. Am. Geophys. Union*, 80(51).

Trenberth, K E and Hurrell, J W (1994) Decadal Atmosphere–Ocean Variations in the Pacific, *Clim. Dyn.*, 9, 303.

Wallace, J M and Gutzler, D S (1981) Teleconnections in the Geopotential Height Field During the Northern Hemisphere Winter, *Mon. Weather Rev.*, 109, 784–812.

Zhang, Y, Wallace, J M, and Battisti, D S (1997) ENSO-like Interdecadal Variability: 1900–1993, *J. Clim.*, 10, 1004–1020.

Pacific–North American (PNA) Teleconnection

Yochanan Kushnir

Columbia University, Palisades, NY, USA

*The PNA pattern is the dominant atmospheric teleconnection pattern in the western half of the Northern Hemisphere. The PNA in uence is largest during the boreal winter (December, January, and February) when it extends from the eastern seaboard of Asia, across the North Paci c and North America to the western North Atlantic (see Figure 1). It is weaker during fall and spring and least in uential in the boreal summer months. The PNA displays a wide range of time scales, from month-to-month variability to multidecadal undulations that bear relevance to the broader issue of climate change. It is most frequently associated with the El Niño/Southern Oscillation (ENSO) phenomenon and is viewed as the latter's extratropical arm (*see *El Niño/Southern Oscillation (ENSO), Volume 1).*

The PNA is generally depicted in terms of its corresponding mid-tropospheric, monthly, geopotential height variations. These form a chain of four large-scale, oval-shaped elements. The most prominent of the four is centered

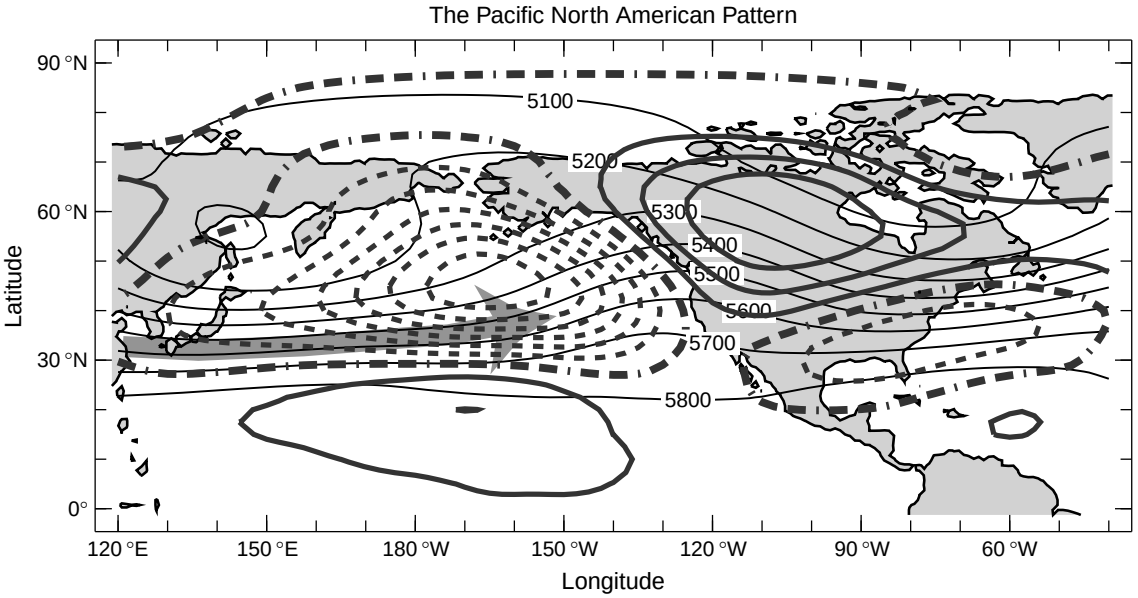


Figure 1 A map of the western sector of the Northern Hemisphere, showing the averaged difference between winters (December–February) in which the PNA was in its positive phase and winters with a negative phase of the phenomenon. The PNA pattern is depicted by thick contours, drawn every 20 m with positive contours solid, negative contours dashed, and the zero line dot dashed. The composite is based on the PNA index as defined in Wallace and Gutzler (1981). (Data available from http://tao.atmos.washington.edu/data_sets/pna). The index values for the winters of 1959–1997 were sorted and the average of the 10 most negative index years was subtracted from that of the 10 most positive. The PNA composite is superimposed on the normal, wintertime, 500 hPa geopotential height field (thin black contours every 100 m). The thick gray arrow over the Pacific Ocean depicts the mean position of the jet stream

in the vicinity of the Aleutian Islands and influences the entire northern North Pacific Basin. To the south lies a much weaker, yet significant, tropical center. Geopotential height fluctuations in these two centers are anti-correlated (varying out of phase with one another) and form a seesaw straddling the mean position of the Pacific subtropical jet stream. The third PNA element overlays western Canada and the northwestern US and also varies out of phase with the main North Pacific center. The fourth element lies over the southeastern half of North America and varies in-phase with the Aleutian center and out of phase with the American Northwest one. The significance of these locations and their respective phases are in their relation to the normal atmospheric circulation (see Figure 1). They represent a variation in the waviness of the atmospheric flow in the western half-hemisphere and thus the changes in the north–south migration of the large-scale Pacific and North American air masses and their associated weather.

References to an atmospheric connection between the Pacific and North America appear in the literature for more than half a century. Wallace and Gutzler (1981) gave a detailed historical account of these early studies and introduced the term PNA in their seminal work on Northern Hemisphere teleconnections. The PNA has been associated with significant variations of surface air temperature and precipitation over North America. The spatial pattern of these effects reflects changes in air-mass movement and in the paths of wintertime storms. When the PNA is in its positive phase (with pressure lower than normal over the North Pacific and higher than normal over Canada) winter temperatures are higher than normal in Alaska, western Canada, and the northwestern US. Positive PNA winters are also associated with cooler conditions in eastern Mexico and the southeastern US. Alaskan winters tend to be drier than normal when the PNA is positive, while northern California and Florida are wetter than normal.

The pioneering work of Bjerknes (1969) on the nature of El Niño and its observed global links was one of the first to divulge the origins and physical mechanisms of the PNA. In the ensuing years this work was extended and the PNA links to El Niño explained as the dynamical atmospheric response to equatorial Pacific sea surface temperature variations and the associated shifts in the massive tropical convection centers (see for example, Rasmusson and Wallace, 1983). However, significant PNA variability exists even in the absence of ENSO, as is evident from observations and climate model experiments that do not include sea surface temperature variability. Non-ENSO variability in variations in the PNA pattern is most probably related to non-linear processes in the atmosphere in which various scales of motion interact with one another (particularly weather disturbances with the more permanent part of the circulation). Thus, while advances in ENSO prediction provide hope for predicting the PNA and its climatic

impacts, such forecasts will have to remain probabilistic in nature, reflecting the random, unpredictable element of midlatitude variability.

Changes in surface winds and air mass movement over the North Pacific Ocean associated with PNA variability have a profound impact on the temperature and circulation of the upper ocean layer and consequently on the distribution and abundance of fish species and other marine life (Mantua *et al.*, 1997). These oceanic effects have been studied extensively in recent years in connection with the multi-decadal component of PNA variability (Trenberth and Hurrell, 1990; Zhang *et al.*, 1997). The existence of such slow components in climate variability is important not only because of their direct impacts but also because of their significance to climate change detection and attribution (IPCC, 1990, 1996).

See also: Natural Climate Variability, Volume 1.

REFERENCES

Bjerknes, J (1966) A Possible Response of the Atmospheric Hadley Circulation to Equatorial Anomalies of Ocean Temperature, *Tellus*, XVIII, 821–828.

IPCC (1990) *Climate Change: The IPCC Scienti c Assessment*, eds J T Houghton, G J Jenkins, and J J Ephraums, Cambridge University Press, Cambridge.

IPCC (1996) *Climate Change 1995: The Science of Climate Change*, eds J T Houghton, L G Meiro Filho, B A Callendar, N Harris, A Kattenberg, and A Maskell, Cambridge University Press, Cambridge, 1–572.

Mantua, N J, Hare, S R, Zhang, Y, Wallace, J M, and Francis, R C (1997) A Pacific Interdecadal Climate Oscillation with Impacts on Salmon Production, *Bull. Am. Meteorol. Soc.*, 78, 1069–1079.

Rasmusson, E M and Wallace, J M (1983) Meteorological Aspects of the El Niño/Southern Oscillation, *Science*, 222, 1195–1202.

Trenberth, K E and Hurrell, J W (1994) Decadal Atmosphere–Ocean Variations in the Pacific, *Clim. Dyn.*, 9, 303–319.

Wallace, J M and Gutzler, D S (1981) Teleconnections in the Geopotential Height Field during the Northern Hemisphere Winter, *Mon. Weather Rev.*, 109, 784–812.

Zhang, Y, Wallace, J M, and Battisti, D S (1997) ENSO-like Decade-to-Century Scale Variability: 1900–93, *J. Clim.*, 10, 1004–1020.

PAGES (Past Global Changes)

PAGES is the International Geosphere–Biosphere Programme Project charged with providing a quantitative understanding of the way the Earth System has changed during the geologically recent past, in response to both natural processes and human activities. Coordination of the PAGES project, which evolved during the 1990s, is undertaken by an international office located in Bern, Switzerland and jointly funded by the US and Swiss National Science Foundations.

One of the central tasks of PAGES, through the organization of coordinated national and international scientific efforts, is to obtain and interpret high quality paleoclimatic records and thereby provide the data essential for the assessment and validation of predictive climate models. PAGES provides an essential framework for the integration and intercomparison of ice, ocean and terrestrial paleo-records and encourages the creation of consistent analytical and data base methodologies in paleoscience.

The responsibility of PAGES goes beyond paleo-climate to include, for the geologically recent past, reconstruction of all the major physical, chemical and biological processes that, through their interactions, regulate the total Earth System. Increasingly, as we approach the present day, this implies a major concern with the role of human activities as an integral component. A more complete understanding of the Earth system as it has changed through time, at both global and regional level, is crucial to providing a robust scientific basis for sustaining human societies in the future. PAGES, therefore, works to identify and elucidate those aspects of past environmental change that are of the greatest significance for human societies. It seeks to promote and coordinate the broad and diverse area of research that extends knowledge of our changing environment back beyond the short, recent period for which we have direct observations or instrumental records.

Further information on the organization of the project, its personnel, priorities and products, can be obtained by consulting its website: http://www.pages.unibe.ch/.

FRANK OLDFIELD *Switzerland*

Paleoclimate

see Earth System History (*Opening essay*, **Volume 1**)

Paleoclimatology

Paleoclimatology is the study of past climates that predate measurement by standard meteorological instruments. The

realm of inquiry stretches from the last century back to the origin of the Earth, 4.5 billion years ago, and starts with the changing composition of the atmosphere as life evolved (*see* Carbon Dioxide Concentration and Climate Over Geological Times, Volume 1). When did photosynthesis begin and when did oxygen concentrations become high enough that an ozone layer could form and absorb ultraviolet radiation? Atmosphere, biosphere, and lithosphere interactions were central to these developments. Paleoclimatologists also use data, models, and theoretical reasoning to study what past climates were like, when and how they changed, and what caused the changes (*see* Earth System History, Volume 1). Hypotheses about the causes of past changes in climate have abounded. However, recent advances in dating, data, data analysis, and modeling have led to information about past climates that is accurate and precise enough that hypotheses can be thoroughly tested, rejected or refined, working toward a well-substantiated and self-consistent experimental and theoretical explanation. Imbrie and Imbrie (1986) provide record of this evolution in paleoclimatic research by tracing early 19th century ideas about ice ages through to the funding of major international, interdisciplinary programs that characterized a significant share of late 20th century research into past climates.

Central to the advances that have been made were the improvement in dating methods and the proliferation of data sets that contain climatic information. Data in terms of geochemical records and numbers and types of fossils come from ice cores, marine, lake, and cave sediments, soils, rocks, and trees (Bradley, 1999) (*see* Natural Records of Climate Change, Volume 1). Tax records, and geomorphological patterns (e.g., glacial moraines and ancient shorelines) on the Earth s surface also provide information. The data are then interpreted in terms of past changes in ice volume, vegetation, plankton abundance, animal distributions, carbon dioxide concentrations, wetland extent, permafrost, and sand-dune formation – all of which are indicative of past changes in climate (Crowley and North, 1996). Through synthesis and modeling, paleoclimatologists have constructed time series and maps of past climates. In an iterative process, these data are being used to test simulations of past climates by climate models, while the model results are being used to help explain the patterns in the data (*see* Climate Model Simulations of the Geological Past, Volume 1).

REFERENCES

Bradley, R S (1999) *Paleoclimatology: Reconstructing Climates of the Quaternary*, 2nd edition, Academic Press, San Diego, CA.

Crowley, T J and North, G R (1996) *Paleoclimatology*, Oxford University Press, New York.

Imbrie, J and Imbrie, K P (1986) *Ice Ages: Solving the Mystery*, Harvard University Press, Cambridge, MA.

THOMPSON WEBB, III *USA*

Paleozoic

The *Paleozoic* era lasted 350 million years (Ma), beginning 570 Ma before present (BP) at the end of the Precambrian and ending with the start of the Mesozoic at 220 Ma BP (Matthews, 1984). The Paleozoic era contains six periods. In order, from earliest to latest, they are: the Cambrian, Ordovician, Silurian, Devonian, Carboniferous (subdivided into the Mississippian and Pennsylvanian), and Permian, in order from earliest to latest. The lower (older) boundary with the Precambrian marks the beginning of hard-bodied organisms that yield the first fossil remains, while the upper (younger) boundary with the Triassic marks one of the major extinction events in Earth s history. The term Paleozoic was coined by Adam Sedgewick based on sedimentary units in Wales, but modified by John Phillips based on fossil assemblages (Prothero, 1990). The paleogeography during the Paleozoic went through major transitions, from circum-equatorial continents, including the supercontinent Gondwanaland (Africa, South America, India, Antarctica, and Australia), to the pole-to-pole supercontinent Pangaea. The transition from Gondwanaland to Pangea required Gondwanaland to pass over the South Pole, and North America, Europe, and Siberia to shift to the Northern Hemisphere (Matthews, 1984). The polar position of Gondwanaland resulted in two major glaciations during the Paleozoic. The first glaciation was the Ordovician–Silurian in Northern Africa and the second the Permo–Carboniferous glaciation in South Africa, Australia, South America, India, and Antarctica. Prior to the first glaciation, high sea levels are indicative of high atmospheric carbon dioxide (CO_2) concentrations (related to tectonic activity) and a period of relative warmth. During the late Paleozoic, the existence of Pangea led to an arid, continental climate (Crowley, 1991). Coal formation in the US and Europe peaked during the late Paleozoic (Crowley, 1991) and the uplift of the Appalachians also began during the Ordovician.

See also: Earth System History, Volume 1.

REFERENCES

Crowley, T J and North, G R (1991) *Paleoclimatology*, Oxford University Press, New York, 339.

Matthews, R K (1984) *Dynamic Stratigraphy: an Introduction to Sedimentation and Stratigraphy*, Prentice Hall, Englewood Cliffs, NJ, 489.

Prothero, D R (1990) *Interpreting the Stratigraphic Record*, W H Freeman, New York, 410.

BENJAMIN S FELZER *USA*

Parameterization

The term *parameterization* refers to a formula that is used to represent the effect of processes that are too small to be explicitly included in a model because they occur at scales smaller than the model grid-point spacing. Parameterizations are necessary because, on the one hand, physical quantities in nature vary continuously in three dimensions, while on the other hand, computers are capable of storing the values of model variables at only a finite number of grid points. The spacing between the points of the model grid is the *spatial resolution*. The typical resolution that can be used in a climate model is hundreds of kilometres in the horizontal. Many important elements of the climate system (e.g., clouds, turbulent eddies in the atmosphere and ocean, irregularities in the land surface) have scales much smaller than this. An example of a parameterization would be a formula used to compute the fraction of a grid box in an atmospheric model that is filled with cloud based on the average relative humidity of the box and the direction and magnitude of vertical air motion (high relative humidity or strong upward motion should lead to greater cloudiness). Both of these parameters would be computed by the atmospheric model on the grid-point array, and are in effect being used to diagnose what should be happening between the grid points. Parameterizations are usually developed based on an intuitive, physical understanding of the key driving variables combined with empirical (observed) relationships. However, because the empirical relationships or physical principles that underlie parameterizations are not rigorously derived from fundamental principles, the observed correlation might not be applicable if the underlying conditions change i.e., as the climate changes. For this reason, the use of parameterizations introduces uncertainties in the climatic changes predicted by climate models. All climate models, no matter how complex and how fine the resolution, rely on parameterizations, because there will always be some processes that cannot be explicitly represented (*see* Climate Feedbacks, Volume 1).

See also: Downscaling, Volume 1; Energy Balance Climate Models, Volume 1.

L D DANNY HARVEY *Canada*

Past Global Changes (PAGES)

see PAGES (Past Global Changes) (**Volume 1**)

Perfluorocarbons (PFCs)

Per uorocarbons (PFCs) are compounds containing only fluorine and carbon. The normal state of these compounds at standard temperature and pressure ranges from gases to liquids and waxes to solids. Very few of these compounds occur naturally in the environment. Only the gaseous PFCs are involved in the global environmental change issue: global climate change. These gases are very resistant to decomposition, with photolysis and ion reactions in the mesosphere being the only significant loss processes. Since transport to these altitudes is slow, the atmospheric lifetimes of PFCs are thousands of years. Due to a combination of the long lifetime and strong infrared absorption, atmospheric emissions of gaseous PFCs make disproportionately large contributions to global climate change with global warming potentials (GWPs) over 5700 (*see* Global Warming Potential (GWP), Volume 1). Intentional production of gaseous PFCs is generally for use in very specialized applications such as the manufacture of semiconductors. The semiconductor industry uses CF_4, C_2F_6 and C_3F_8 in both the plasma etching of thin films and in plasma cleaning chemical vapor deposition tool chambers because of their unique characteristics. Perfluoromethane (CF_4) is the most abundant PFC in the atmosphere with a 1995 concentration of about 80 ppt. Current anthropogenic emissions of this gas are more than 1000 times larger than natural emissions. However, because of the extremely long lifetime of CF_4, over 50 000 years, about half of the current atmospheric burden is of natural origin. The largest anthropogenic source of CF_4 is a result of unintended production during the manufacture of aluminum. PFCs are among the compounds listed for control under the Kyoto Protocol to the United Nations Framework Convention on Climate Change (*see* Ozone Layer: Vienna Convention and the Montreal Protocol, Volume 4).

MACK MCFARLAND *USA*

Permafrost

Gunter Weller
University of Alaska Fairbanks, Fairbanks, AK, USA

Permafrost, or perennially frozen ground, is de ned as ground that is continually at a temperature below 0 °C for a

period of two years or more. It is therefore directly related to climate, and it can be highly sensitive to climate change. Permafrost is characteristic of the polar regions, and underlies approximately 25% of the land areas of the Northern Hemisphere; it is widespread in Alaska, Canada, Russia, Mongolia, China, as well as Antarctica. It also exists offshore along the arctic coasts as a remnant from the last ice age, when the global sea level was lower and the offshore regions were exposed to lower air temperatures. In addition, it occurs in some high mountains in lower latitudes and also on Mars and possibly other planets.

OCCURRENCE AND DISTRIBUTION

Figure 1 shows the distribution of permafrost in the Northern Hemisphere. Brown *et al.* (1997) have produced a map with greater detail. The distribution of permafrost is not controlled by climatic factors alone but also by geologic, hydrologic, topographic and botanical features. Where thick, insulating vegetation or peat layers protect the underlying soil from the Sun s direct rays, or on north-facing slopes, the upper permafrost boundary is close to the surface. The annually thawed top layer of the ground is called the active layer. It is a few decimeters thick in the high Arctic, but can be over 1 m thick in the sub-Arctic. Since the ground is frozen, allowing no groundwater discharge, the overlying tundra or boreal forest is often waterlogged, even in areas with low precipitation.

PERMAFROST TYPES AND FEATURES

Permafrost is classified as discontinuous, continuous, sporadic and mountain, and sub-sea permafrost, as shown in Figure 1, depending on its regional distribution and thickness. Discontinuous permafrost is patchy and relatively thin and extends south to about 50° latitude in China and in Canada s Hudson Bay area; often there are islands of thawed ground surrounded by permafrost. Continuous permafrost is extensive in aerial coverage, has no permafrost-free patches, except under deep lakes and river channels, and is thick. Known permafrost thicknesses range from about 1500 m in Siberia to less than 1 m at the southern boundary of the discontinuous or sporadic permafrost zone. The largest extent of mountain permafrost is on the Tibetan Plateau but it is also widespread in other alpine areas. Sub-sea permafrost is widespread on the Siberian and North American continental shelves in shallow water.

Permafrost terrain can often be recognized by geometrically arranged regular features on the surface, which are produced by freezing and thawing processes. This is called patterned ground and resembles large polygons, from a meter to many meters across, separated from one another by shallow troughs. Another feature is the *thermokarst*, which is an irregular depression in the ground where ice masses in permafrost terrain have thawed due to a warmer climate. *Pingos* are surface blisters or ice mounds, reaching tens of meters in height and upwards of 150 m across, which result as ice forms and expands during the re-freezing of thawed sediments. They are often located in old lakebeds and may take hundreds of years to form.

THE ROLE OF PERMAFROST IN HUMAN ACTIVITIES

One of the major obstacles to human settlement in and development of the high-latitude regions is ice-rich permafrost. The regions underlain by permafrost, with considerable amounts of ground ice, pose unusual and serious engineering problems related to the design, construction and maintenance of roads, pipelines, airfields, houses and other structures. The construction of the 1300-km long Trans-Alaska hot-oil pipeline, for example, required elevating the pipelines above ground in permafrost terrain and keeping

the support structures for the pipe artificially frozen. These problems require special engineering techniques, procedures, and materials to minimize disruption of the natural environment, but also to provide the most economical and sound methods of development.

Large amounts of ground ice are often present in the form of ice wedges and massive layers of ice lenses (Figure 2). Ice wedges form as contractions in the soil, which, on freezing and cooling, produce cracks that eventually fill with water and freeze. This process is repeated over long periods of time with subsequent expansion of the ice wedges and displacement of the surrounding frozen sediments and soils. Their presence is of paramount importance because thawing of this ice can cause subsidence, loss of soil strength, and other disastrous effects on man-made structures. Some of the biggest permafrost problems have been observed in Russia where apartment blocks, built from 1950 to the present, have begun to collapse because the climatic warming experienced since then has weakened the bearing capacity of the permafrost on which the buildings stand. Russian engineers expect most of these structures, of which there are hundreds in cities like Yakutsk, Vorkuta and Tiksi, to fail in the next 30 years if present climate trends continue.

EFFECTS OF CLIMATE CHANGE ON PERMAFROST

Permafrost is very sensitive to climate change and its profile of temperature with depth can indicate the past climatic history of the site. For example, borehole temperature measurements in continuous permafrost in Alaska have shown temperature increases of $2-4^{\circ}C$ over the past century (Lachenbruch and Marshall, 1986). Since the ground temperatures in the discontinuous permafrost zone are close to the melting point of ice, this permafrost is particularly susceptible to changes, and a small temperature rise could thaw all of it over times ranging from decades to centuries.

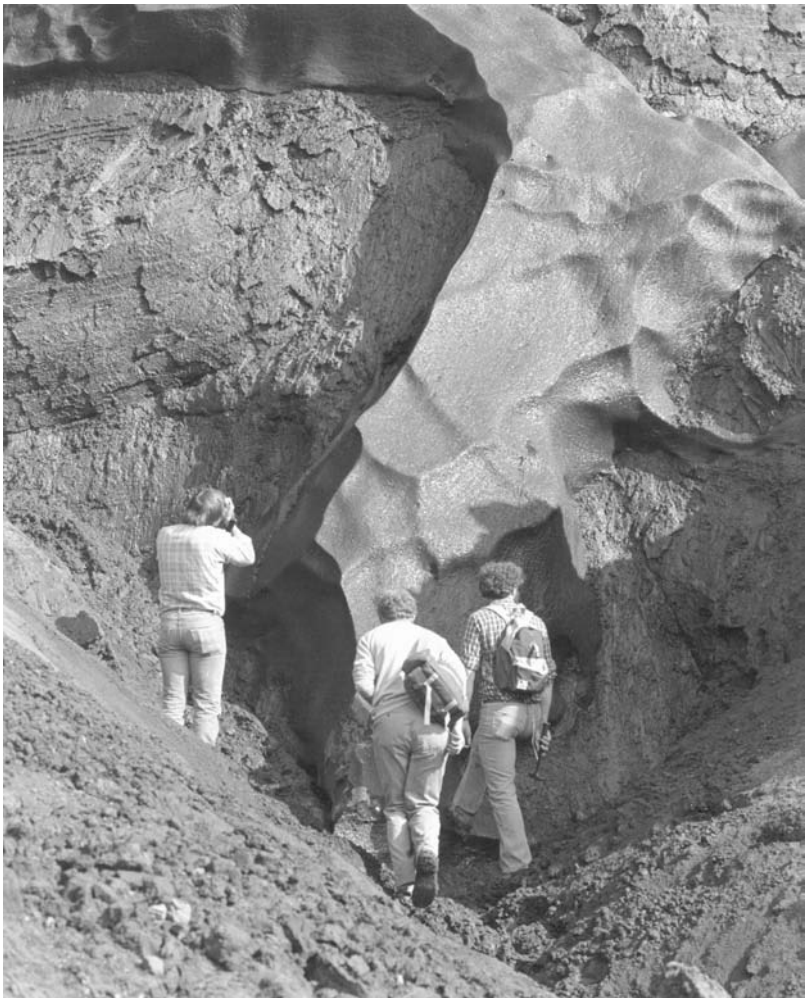


Figure 2 Massive ice wedges in permafrost terrain near Fairbanks, Alaska, exposed by gold mining operations. (Geophysical Institute photo, University of Alaska Fairbanks)

Table 1 Projected impacts due to a 3 °C warming of permafrost (modified from Weller and Lange, 1999)

	Continuous permafrost terrain			Discontinuous permafrost terrain		
Feature/parameter	Low	Moderate	Severe	Low	Moderate	Severe
Thaw lakes	X					X
Coastal processes		X				X
Eolian activity	X			?←	X	
Vegetation		X→	X	X→	X	
Active layer thickness	X					X
Permafrost thawing	X					X
Thaw settlement	X					X
Slope instability		X				X
Erosion		X				X
Solifluction		X				X
Engineering impacts		X				X

Temperature increases in the Arctic in recent decades have already reduced the volume of permafrost by thawing. In some areas, this has led to major erosion problems and landslides, for example in the Mackenzie River. Permafrost thawing has also caused major landscape changes from forest to bogs, grasslands and wetland ecosystems in Alaska. Table 1 summarizes some of the expected changes for a 3 °C temperature rise. *Soli uction* in permafrost terrain is the process of slow downward flow of water-saturated soil on slopes, which is promoted by the permafrost trapping snow and ice melt within the surface layer making it more fluid.

See also: Arctic Climate, Volume 1; Glaciers, Volume 1.

REFERENCES

Brown, J, Ferrians, Jr, O J, Heginbottom, J A, and Melnikov, E S (1997) Circum-Arctic Map of Permafrost and Ground-Ice Conditions, US Geological Survey Circum-Pacific Map CP-45, 1 : 10 000 000, Reston, VA.

Heginbottom, J A, Brown, J, Melnikov, E S, and Ferrians, Jr, O J (1993) Circumarctic Map of Permafrost and Ground Ice Conditions, in *Proceedings: Sixth International Conference on Permafrost*, Vol. 2, South China University of Technology Press, Wushan, Guangzhou, China, 1132–1136.

Jorgenson, M T, Racine, C H, Walters, J C, and Osterkamp, T E (2001) Permafrost degradation and ecological changes associated with a warming climate in central Alaska, *Climatic Change*, 48, 551–579.

Lachenbruch, A H and Marshall, B V (1986) Changing Climate: Geothermal Evidence from Permafrost in the Alaskan Arctic, *Science*, 234, 689–696.

Nelson, F E, Anisimov, O A, and Shiklomanov, N I (2001) Subsidence risk from thawing permafrost, *Science*, 410, 889–890.

Sturm, M, Racine, C, and Tape, K (2001) Increasing shrub abundance in the Arctic, *Nature*, 411, 546–547.

Weller, G and Lange, M, eds (1999) *Impacts of Global Climate Change in the Arctic Regions*, Report from a Workshop on the Impacts of Global Change, 25–26 April 1999, Tromsø, Norway, Published for International Arctic Science Committee by the Center for Global Change and Arctic System Research at the University of Alaska Fairbanks, 1–59.

PFCs (Perfluorocarbons)

see Perfluorocarbons (PFCs) (**Volume 1**)

Photochemical Reactions

Douglas Kinnison

National Center for Atmospheric Research, Boulder, CO, USA

Photochemical reactions are reactions that are initiated by the absorption of a photon and that lead to the formation of new chemical species. These reactions are of particular importance because they are the link between the energy provided by solar radiation and the chemical composition of the atmosphere. Interestingly, for example, photochemical reactions driven by solar radiation are essential to the formation of the stratospheric ozone layer, which is essential to life on Earth because it screens out most solar UV-radiation.

Photochemical reactions occur in several different ways. In a photodissociation reaction, a photon splits apart a molecule:

$$AB + h\nu \longrightarrow AB^* \longrightarrow A + B \qquad \bullet 1 \bullet$$

The asterisk (*) represents an intermediate excited state of the *AB* molecule. There are also other photochemical processes, in addition to photodissociation, that can be described as photochemical reactions or photochemically driven, e.g.,

excitation:

$$A + h\nu \longrightarrow A^* \qquad \bullet 2 \bullet$$

insertion:

$$A^* + B \longrightarrow AB \qquad \bullet 3 \bullet$$

abstraction:

$$A^* + B \longrightarrow C + D \qquad \bullet 4 \bullet$$

dissociative excitation:

$$A^* + CD \longrightarrow A + C^* + D \qquad \bullet 5 \bullet$$

In the above examples, *A* through *D* species could be an atom or molecule.

The most important type of photochemical reaction in atmospheric chemistry is photodissociation (*see* Stratosphere, Chemistry, Volume 1; Troposphere, Ozone Chemistry, Volume 1). For example, the nitrogen dioxide (NO_2) molecule absorbs solar radiation in the visible and near ultraviolet region:

$$NO_2 + h\nu \longrightarrow NO_2^* \qquad \bullet 6 \bullet$$

If the wavelength of the photon is less (shorter) than approximately 420 nm, the photon has sufficient energy to exceed the bond strength of the NO_2 molecule and stimulate creation of nitric oxide (NO) and atomic oxygen (O) in its ground electronic state (^{3}P):

$$NO_2 + h\nu \longrightarrow NO_2^* \longrightarrow NO + O \bullet {}^3P \bullet \qquad \bullet 7 \bullet$$

This reaction, followed by the reaction of O(^{3}P) with molecular oxygen (O_2) is the source of *in situ* produced ozone (O_3) in the troposphere. Another example of a photochemical reaction that is important in atmospheric chemistry is:

$$O_3 + h\nu \bullet \lambda \bullet \ 310\,\text{nm} \bullet \longrightarrow O \bullet {}^1D \bullet + O_2 \qquad \bullet 8 \bullet$$

The products of the O_3 photolysis induced by radiation with wavelengths shorter than 310 nm are electronically excited oxygen atom (O(^{1}D)) and molecular oxygen (O_2). O(^{1}D) is a very important species and reacts with numerous chemical constituents in the atmosphere, e.g., O(^{1}D) plus H_2O is the primary process for forming hydroxyl radical (OH) in the troposphere; $O \bullet {}^1D \bullet + N_2O$ is the primary processes for forming nitrogen oxides in the stratosphere.

The additional energy of the absorbed photon above the bond dissociation energy is distributed as electronic, vibrational, rotational, and translational energy of the product molecules (and/or atoms).

In order to assess the atmospheric importance of any photochemical reaction, one must derive the photochemical rate (*see* Depletion of Stratospheric Ozone, Volume 1; Stratosphere, Chemistry, Volume 1; Troposphere, Ozone Chemistry, Volume 1). In the NO_2 example given above, the photochemical rate or change in the concentration of NO_2 with time is (e.g., in units of molecules $\text{cm}^{-3}\,\text{s}^{-1}$):

$$\frac{-\text{d}[NO_2]}{\text{dt}} = j[NO_2] \qquad \bullet 9 \bullet$$

The concentration of NO_2 is designated with square brackets (in units of molecules cm^{-3}). The rate constant is designated as j and has units of s^{-1} and is commonly termed the *photolysis rate coef cient* or *photolysis frequency*. This rate constant is the integral over all wavelengths for the following terms:

$$j = \int_\lambda F \bullet \lambda \bullet \sigma \bullet \lambda \bullet \phi \bullet \lambda \bullet \text{d}\lambda \qquad \bullet 10 \bullet$$

where F is the flux of photons (quanta $\text{cm}^{-2}\,\text{s}^{-1}\,\text{nm}^{-1}$) within the wavelength region, $\text{d}\lambda$ (nm); σ is the absorption cross-section of the molecule (cm^2 molecule^{-1}) within the same wavelength region; and ϕ is the primary quantum yield of the photochemical process (molecules quanta^{-1}) also within the same wavelength region. The primary quantum yield is defined as:

$$\phi_i = \frac{\text{number of excited molecules following the pathway for process } i}{\text{total number of photons absorbed}} \qquad \bullet 11 \bullet$$

For example, the nitrate radical (NO_3) absorbs light in the visible region of the solar spectrum:

$$NO_3 + h\nu \bullet 600 \bullet \ \lambda \bullet \ 700\,\text{nm} \bullet \longrightarrow NO_3^* \qquad \bullet 12 \bullet$$

This electronically excited NO_3 molecule then follows one of two photochemical pathways, with primary quantum yields given as ϕ_1 and ϕ_2:

$$NO_3^* \longrightarrow NO_2 + O \qquad \phi_1 \qquad \bullet 13 \bullet$$

$$NO_3^* \longrightarrow NO + O_2 \qquad \phi_2 \qquad \bullet 14 \bullet$$

By definition, the sum of all primary quantum yields is unity for all photochemical and photophysical processes. An example of a photophysical processes is fluorescence:

$$NO_3^* \longrightarrow NO_3 + h\nu \qquad \phi_3 \qquad \bullet 15 \bullet$$

For additional information on both photochemical and photophysical processes, see Atkins (1990) and Finlayson-Pitts and Pitts (2000).

REFERENCES

Atkins, P W (1990) *Physical Chemistry*, 4th edition, W H Freeman, New York.

Finlayson-Pitts, B J and Pitts, Jr, J N (2000) *Chemistry of the Upper and Lower Atmosphere, Theory, Experiments, and Applications*, Academic Press, San Diego, CA.

Planetary Boundary Layer

Roger A Pielke, Sr

Colorado State University, Fort Collins, CO, USA

The planetary boundary layer (PBL) is the layer of the atmosphere immediately above the Earth's surface through which turbulent eddies of air, generated at the surface, extend. The atmosphere above the PBL is called the free atmosphere. These eddies vertically mix heat, water vapor, and atmospheric trace gases such as carbon dioxide, methane and radon. These turbulent air motions can be caused by surface heating, which produces convective turbulent fluxes, and/or by the shearing stress of the wind as it blows over the surface. The shearing stress produces what are called mechanical turbulent fluxes.

The depth of the PBL, in a cloud free atmosphere, can reach 5 km or more over regions of the Earth with strong surface heating and a temperature change with altitude that does not significantly suppress vertical mixing (Figure 1). This temperature change (called the lapse rate) has a value close to the adiabatic lapse rate of $-1\,^{\circ}\mathrm{C}$ per 100 m. At some height within the troposphere, there will be a layer of air that has a lapse rate that decreases less rapidly with altitude, and could even increase with altitude for a short distance. The height at which this first occurs is referred to as the base of the temperature inversion. This height corresponds to the top of the PBL.

When deep cumuliform clouds (cumulonimbus clouds) are present, the depth of the PBL can extend to the tropopause. In the tropical regions, this corresponds to altitudes of 17 km and higher. Cumulonimbus clouds, therefore, are a mechanism by which surface conditions can be rapidly transported through the entire depth of the troposphere. The PBL is generally subdivided into the convective boundary layer, the neutral boundary layer, the stable boundary layer, and the disturbed boundary layer. The *convective boundary* occurs when surface heating dominates the vertical turbulent mixing. It most often develops during daylight over land, where the ground surface is strongly heated by the Sun and over oceans when cold air blows over warmer water.

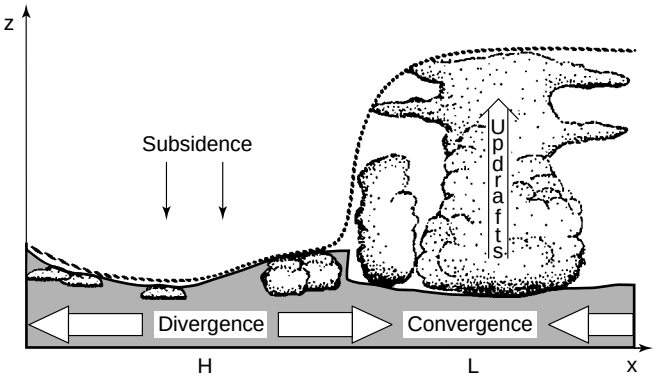


Figure 1 Schematic of synoptic-scale variation of boundary layer depth between centers of surface high (H) and low (L) pressure. The dotted line shows the maximum height reached by surface modified air during a one-hour period. The solid line encloses the shaded region, which is most studied by boundary-layer meterologists. (Reproduced from Stull (1989) by permission of Kluwer Academic Publishers)

Among the major characteristics of the surface that influence the convective PBL are its roughness, albedo, and conductance of water vapor. The partitioning of the convective heat flux into its sensible and latent turbulent heat components directly affects the boundary layer, because only sensible heating directly forces the vertical turbulent mixing. Latent convective fluxes result from the evaporation of water vapor from the soils or water surfaces, and from the transpiration of water vapor through the stomata of leaves. This latent flux is passively transported by the turbulence generated by the sensible heating of the surface until condensation and/or deposition of some of the water vapor occur. The latency of the heat is realized when this condensation and/or deposition occurs. This heat release can then directly affect vertical turbulent mixing. We see this effect visually in the bubbly, cauliflower form of cumulus clouds.

The *neutral boundary layer* occurs when there is no surface heating, but the winds are strong enough to continuously generate turbulence through shearing stresses at the surface. There is a shearing stress at the surface, because the wind speed is zero at that level, yet non-zero just above the surface. Under this condition, the wind change with altitude in the lowest 10% of the PBL is accurately represented as a linear function of the natural logarithm of the height above the surface. This wind profile is called the logarithmic wind profile.

The *stable boundary layer* develops when the vertical turbulent mixing is caused by the shear-generated turbulent fluxes. The turbulent heat flux, however, will actually be towards the surface when this mixing occurs, because the

surface is cooler than the overlying air. A stable boundary layer will tend to suppress shear generated turbulence.

The stable boundary layer commonly occurs on light-wind, clear nights, so that it is also referred to as the *nocturnal boundary layer*. Cooling by radiation is typically an important aspect of a stable boundary layer. This cooling can actually be strong enough to completely suppress vertical mixing. Such a boundary layer is then referred to as a *radiation boundary layer*. Vertical turbulence can also occur intermittently in a stable boundary layer, because the shearing stress periodically becomes large enough to generate turbulence, despite the stabilizing influence of the vertical temperature profile. A radiation boundary layer is also called a *very stable boundary layer*. The boundary layer itself can be decomposed into two layers. These layers are most clearly defined for the convective boundary layer and the neutral boundary layer. In the lower 10% of the boundary layer, wind direction changes little with altitude and can be generally ignored. This layer is called the *surface layer*. Above this layer to the top of the planetary boundary is called the *transition layer*. As you ascend within the transition layer, the atmospheric structure is approaching that of the free atmosphere. The transition layer is also referred to as the *Ekman layer*. The Ekman layer is a mathematical derivation used to explain the approach of the wind to its free atmospheric value, because the frictional effect of the surface becomes smaller with altitude (see Figure 2).

When adjacent regions of the Earth s surface have different characteristics, PBLs of different depths can occur. When the wind blows from one surface type to another, a transition boundary layer will occur that has properties of the new surface in the lowest levels, but properties of the older surface above. The region below the interface between these two regions of the boundary layer is called the *internal boundary layer*.

The physics governing the boundary layer for slowly varying weather conditions over relatively homogeneous, flat surfaces on Earth for clear or shallow cloud conditions is reasonably well understood. The modeling of weather and climate almost always uses parameterizations of the PBL that were developed from observations made over such simple landscapes. Our understanding of the boundary layer in more complex landscapes, in disturbed weather, and for rapidly changing weather conditions, however, remains poor.

In the context of environmental change, any anthropogenic or natural effect that alters the Earth s surface will immediately and directly alter the structure and dynamics of the PBL. Anthropogenic effects include landscape changes, such as the conversion of a tropical forest to grassland. Such changes in the boundary layer, through its effect on cumulonimbus convection, can influence the global atmosphere circulation (e.g., Chase *et al.*, 2000). Natural effects include snow accumulation and melt, as discussed for example in Liston (1999).

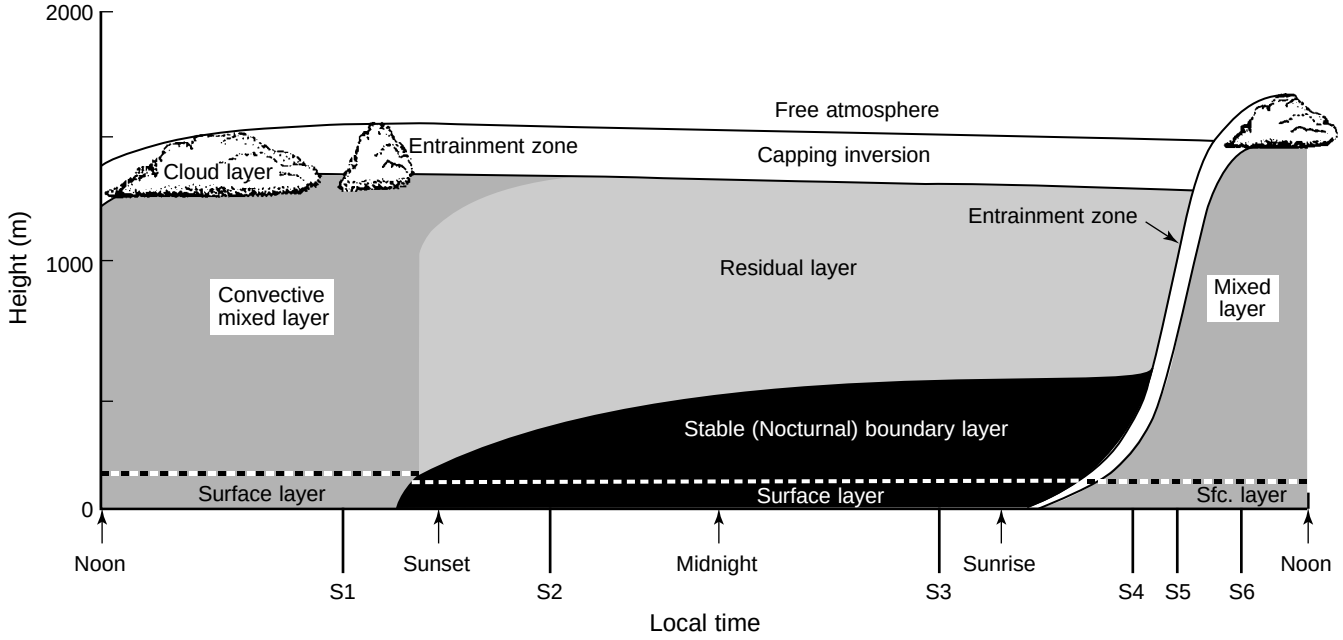


Figure 2 The boundary layer in high pressure regions over land consists of three major parts: a very turbulent mixed layer; a less turbulent residual layer containing former mixed layer air; and a nocturnal stable boundary layer of sporadic turbulence. The mixed layer can be subdivided into a cloud and a subcloud layer. Time markers are indicated by S1–S6. (Reproduced from Stull (1989) by permission of Kluwer Academic Publishers)

REFERENCES

Chase, T N, Pielke, R A, Kittel, T G F, Nemani, R R, and Running, S W (2000) Simulated Impacts of Historical Land Cover Changes on Global Climate in Northern Winter, *Clim. Dyn.*, 16, 93–105.

Liston, G E (1999) Interrelationships among Snow Distribution, Snowmelt, and Snow Cover Depletion: Implications for Atmospheric, Hydrologic, and Ecologic Modeling, *J. Appl. Meteorol.*, 38(10), 1474–1488.

Stull, R B (1989) An Introduction to Boundary Layer Meteorology, Kluwer, Dordrecht, 1–666.

Plate Tectonics

W Brian Simison
University of California, Berkeley, CA, USA

The evolutionary history of our planet's varied environments is intricately tied to the processes and products of plate tectonics. Over great spans of time, the movement and deformation of the Earth's crust has brought different environments together to form new environments and separated other environments to evolve in isolation. Environments have been lifted from low to high elevations during the formation of mountains. Volcanic activity has wiped out vast environments and simultaneously laid down rich soils for the formation of new ones. The moving continents have in uenced global weather patterns, which in turn have affected the global distribution of environments. All of these scenarios are products of plate tectonics and part of the evolutionary history and global environmental change of our planet.

Plate tectonics is the primary process controlling the geographic distribution of continents and oceans and the geological dynamics of the ever-changing Earth. It is also the mechanism responsible for many geological features and events, including earthquakes, volcanoes, geysers, mountains, hot vents, island arcs, and the distribution of mineral resources.

Our understanding of plate tectonics is still developing; prior to 1960, very little was known about the processes controlling the planet s surface. The first to suggest that the surface of the planet was not static was Alfred Wegener (1880–1930) (*see* Wegener, Alfred, Volume 1). While examining maps and globes, Wegener noticed that the continents seemed to fit together like a puzzle. In 1912, he proposed that the continents were once united into a single protocontinent, which he called *Pangaea* (meaning all lands), and that over time they drifted apart.

There were many well-known geological and paleontological clues supporting his hypothesis. The best-known geological evidence was the observation that ancient gouges left by glaciers, known as *striae*, were very similar along the matching shorelines of Africa and South America. Wegener s hypothesis could explain this observation by placing the two continents together during an ancient ice age, thus permitting glaciers to cross the now separated margins of the continents and leave similar striae on their coasts. Well-documented paleontological evidence indicated the presence of the same fossilized plants and animals on different continents. Several hypotheses had been proposed to explain this observation; among the most popular of that time was the idea that land bridges spanned the vast oceans, thereby permitting plants and animals to cross the oceans. Wegener s hypothesis of continental spreading, however, provided an alternative explanation for the occurrence of similar fossils separated by great seas.

Wegener also argued that his hypothesis could account for the formation of the planet s mountain ranges and the differences in their apparent ages. The theory of mountain formation, or *orogenesis*, being discussed during his time was the Contraction theory, which suggested that the planet was once a molten ball and in the process of cooling, the surface cracked and folded upon itself to create mountains. The big problem with this idea was that all mountain ranges should be approximately the same age, and this was known not to be true. Wegener s explanation suggested that as the continents moved, the leading edge would encounter resistance, and thus compress and fold upwards, forming mountains near the leading edges of the drifting continents. The Sierra Nevada Mountains on the Pacific coast of North America and the Andes on the coast of South America were cited as examples. Wegener also correctly suggested that India drifted northward into the Asian continent, thus forming the Himalayas.

While his ideas elegantly explained many geological phenomena, Wegener did not have a viable mechanism that could move continents across the surface of the planet. This kept the scientific community from accepting his ideas. In 1929, just as Wegener s theory was being dismissed, Arthur Holmes elaborated on one of Wegener s many hypotheses; the idea that the mantle undergoes thermal convection. This idea is based on the property of heating fluids from below: the density of a heated fluid will decrease and therefore rise to the surface where it cools and subsequently sinks. This repeated rising and sinking results in a circulating current that could provide the necessary force to move continents. Arthur Holmes suggested that this thermal convection was like a conveyor belt and that the upwelling pressure could break apart a continent and force the fragmented continent to ride the convection currents in opposite directions.

Unfortunately, there was very little evidence that the surface of the planet was riding on a swirling sea of fluid

rock. However, scientists began documenting features of the ocean floor and discovered mid-oceanic ridges; parallel to the ridges were symmetrical bands of alternating geomagnetic polarity (Figure 1). Scientists have since learned that the Earth periodically undergoes reversals of its magnetic fields, where the geomagnetism that pulls the compass needle towards the North (magnetic) Pole suddenly switches to the South (magnetic) Pole. Geomagnetic direction is permanently recorded in rock as it cools from a molten to solid phase. Scientists studying the sea floor noticed that the geomagnetic bands along the mid-oceanic ridges formed parallel patterns of varying widths and that these patterns formed mirror images on either side of the ridges. This and other clues led R Deitz (1961) and Harry Hess (1962) to publish similar hypotheses based on mantle convection currents, now known as sea floor spreading. Their idea was the same as that proposed by Holmes over 30 years earlier, but now there was much more evidence to further develop and support the idea.

Our current understanding of plate tectonics involves four main features:

1. the Earth s surface is covered by a series of crustal plates;
2. the plates are continually moving, with their ocean edges continually spreading from the mid-oceanic ridges, their landward edges sinking at the plate margins, and the descending edges being reabsorbed in the mantle;
3. crustal plates ride convection currents in various directions;
4. the source of heat driving the convection currents is radioactive decay deep in the Earth s mantle.

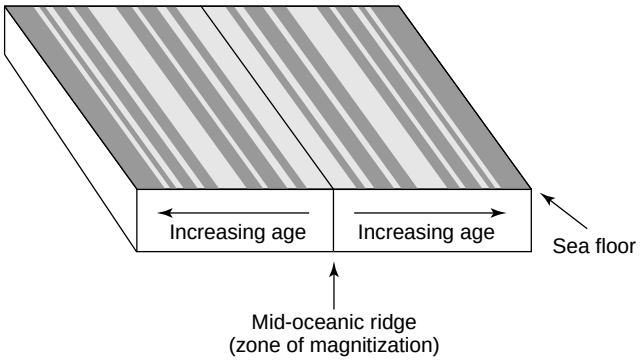


Figure 1 This figure demonstrates the effects of periodic magnetic polar reversals along mid-oceanic ridges. The dark stripes represent the ocean floor generated during *reversed* polar orientation and the lighter stripes represent the polar orientation we have today. Notice that the patterns on either side of the line representing the mid-oceanic ridge are mirror images of one another. Molten rock is extruded at the mid-oceanic ridge and hardens with the current geomagnetic direction. The stripes represent older and older rock as they move away from the mid-oceanic ridge. Geologists have determined that rocks found in different parts of the planet with similar ages have the same magnetic characteristics

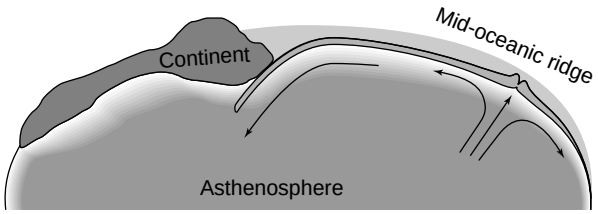


Figure 2 A cross-section of the Earth revealing a continental plate, an oceanic plate, an ocean and the asthenosphere. The arrows indicate the convection patterns of the fluid asthenosphere underneath the solid plates or lithosphere. These circulating convection currents are responsible for the movement of the plates

Underneath the Earth s crust (the *lithosphere*: a solid array of plates (*see* Lithosphere, Volume 1)) is a malleable layer of heated rock known as the asthenosphere, which is heated by the radioactive decay of elements such as uranium, thorium, and potassium. Because the radioactive source of heat is deep within the mantle, the fluid asthenosphere circulates as convection currents underneath the solid lithosphere. The asthenosphere is the source of lava released by volcanoes, the source of heat that drives hot springs and geysers, and the source of raw material that creates the mid-oceanic ridges and forms new ocean floor. Magma continuously wells upwards at the mid-oceanic ridges (Figure 2, arrows), producing currents of magma flowing in opposite directions that generate the forces pulling the sea floor apart at the mid-oceanic ridges. As the ocean floor is spread apart, cracks appear in the middle of the ridge, allowing molten magma to surface through the cracks to form the newest ocean floor. As the ocean floor moves away from the mid-oceanic ridge, it encounters continental plates that drive it back into the asthenosphere, where the plate edge is melted and returns to a heated state.

The large-scale effects of plate tectonics are important determinants of the global environment, with mountains affecting global weather, ocean basins affecting sea levels, shape of continents determining geography, and continental dynamics affecting the evolution of biological diversity.

See also: Earth System History, Volume 1; Global Plate Tectonics, Volume 1.

FURTHER READING

Hamblin, W K (1975) *The Earth's Dynamic Systems: a Textbook in Physical Geology*, Burgess, MN, 1–578.

Kearey, P and Vine, F J (1996) *Global Tectonics*, Blackwell Science, London.

Montgomery, C W (1987) *Physical Geology*, W C Brown, Chicago, IL.

Uyeda, S (1978) *The New View of the Earth*, W H Freeman, San Francisco, CA, 1–217.

Pleistocene

Pleistocene refers to the ice age that was the geologic epoch prior to the present Holocene within the Quaternary Period, which is part of the Cenozoic Era. The Pleistocene began 1.5–1.6 Ma and ended with the start of the Holocene, 10 000 years ago (Berggren and Couvering, 1978).

See also: Earth System History, Volume 1; Holocene, Volume 1; Quaternary, Volume 1.

REFERENCE

Berggren, W A and Van Couvering, J A (1978) The Quaternary, in *Treatise on Invertebrate Paleontology, Part A., Introduction: Fossilization (Taphonomy), Biogeography and Biostratigraphy*, eds R A Robinson and C Teichert, Geological Society of America, Boulder, CO.

BENJAMIN S FELZER *USA*

Polynyas

Polynyas are large, persistent areas of open ocean surrounded by sea ice or sea ice and land. They are larger, less elongated, and less ephemeral than the channels of open water known as *leads*, which form where the drift of the sea ice locally diverges. They tend to recur seasonally in the same place – sometimes every year, sometimes very irregularly.

Coastal polynyas commonly form in regions where wind blowing from land (often ice covered) forces newly forming sea ice away from the shore. In the polar winter, the heat loss from the open water to the cold atmosphere may be a hundred times greater than that through the surrounding sea ice. The latent heat released by the freezing of the new ice helps replenish that lost heat, allowing the polynya to remain unfrozen. At the same time, salt rejected in the freezing increases the salinity of the water and hence its density. The denser water sinks; consequently, coastal polynyas, particularly around Antarctica, contribute importantly to the generation of the bottom waters of the world s oceans.

A second mechanism of polynya formation is the upwelling of relatively warm, deeper water. Such upwelling may be caused by the interaction of ocean currents with features of the submarine topography.

Polynyas are important to climate because, together with leads, they produce the bulk of the heat and water vapor exchange between the ocean and the atmosphere in regions of sea ice cover. They are also the loci of strong biological activity, like the marginal sea ice zone around the outer edges of the ice pack.

A particularly large (several hundred square kilometers), mysterious polynya formed in the Weddell Sea near Antarctica for several austral winters during the 1970s. Its cause was probably a combination of atmospheric and oceanic circulation patterns, but marine expeditions to study it were largely unsuccessful because the polynya never reappeared after 1978. A more reliable polynya known as North Water forms annually at the northern end of Baffin Bay, west of Greenland. It too results from a combination of the two mechanisms. North Water exerts a pronounced effect on the regional climate.

Polynyas, because they are delicately balanced between competing oceanic and climatic processes, are likely to respond sensitively to environmental change. Thus they may act both to amplify changes and to provide early evidence, through their own alterations, of those changes.

See also: Sea Ice, Volume 1; Southern Ocean, Volume 1.

CHARLES R BENTLEY *USA*

Precipitation

see Hydrologic Cycle (**Volume 1**); Rain (**Volume 1**); Snow (**Volume 1**)

Predictability, Climate

see Chaos and Predictability (**Volume 1**)

Prediction in the Earth Sciences

Michael C MacCracken

Lawrence Livermore National Laboratory, Livermore, CA, USA

The great Danish physicist Neils Bohr once commented that "Prediction is very dif cult, especially about the future" (also sometimes attributed to Mark Twain and, later, to Yogi Berra). Although there is some public understanding that the existing pattern of the weather allows for prediction of how

the weather will evolve over the coming few days, there is little understanding of the basis for or accuracy of longer-term prognostications. Indeed, the notion that scientists can make statements about what is likely to happen 100 or more years into the future can seem overly presumptuous if recognition is not given to the careful ways that scientists use words to convey their views about the possibility/probability of future happenings.

In public discussion about weather and climate, the words prediction, forecast, projection, and scenario are often used almost interchangeably, as if their meanings are synonymous. However, the scientific community researching global change generally uses these words quite carefully, with each word having a distinctive meaning and significance. These differences and meanings are important to understand because they convey an important set of assumptions underpinning their usage.

Although there is not universal agreement, distinctions of the following types are often made:

- A *prediction* is usually a probabilistic statement that something will happen in the future based on conditions that are known today and assumptions about the physical processes that will determine these changes. A prediction generally assumes that future changes in factors other than those being predicted will not have a significant influence on what is to happen. In this sense, a prediction is most influenced by the initial conditions, that is, on the current conditions that are known through observations. Thus, a weather prediction indicating a major snowstorm will develop over the next few days is based on the state of the atmosphere today (and its conditions in the recent past) and not on unpredictable changes of other potentially influential factors that serve as boundary conditions, such as how ocean temperatures or human activities may change over the next few days. A prediction is made probabilistic by accounting for various types of uncertainties, for example, in the accuracy of observations, in the chaotic state of the atmosphere, etc. (*see* Chaos and Predictability, Volume 1). For decision-makers, what is important is that a prediction is a statement about an event that is likely to occur no matter what they do (i.e., policymakers cannot change tomorrow s weather).
- A *forecast* is generally more specific and more routinely prepared than a prediction, often being made by a particular person (or group) or with a particular technique or representation of current conditions. Whereas a prediction might suggest that there is a 70% chance of rain in the area, an example of a forecast would be a statement by a particular weather forecaster (or forecasting group) that it is expected to rain in the mid-afternoon tomorrow in a particular community. Basically, a general prediction is being made more specific using best judgment of some type. For a decision-maker, the credibility of the forecast depends critically on the credibility of the forecaster (and the forecasting technique). For many forecasters, this can be evaluated because there is a history of their records of success and failure. For example, weather forecasts are made every day, so statistics on their performance can be developed and evaluated. The recent development of ensemble forecasts (i.e., drawing upon, in a probabilistic way, an assembly of a set of forecasts that are each based on a separate technique or set of initial conditions) is a means of transforming a forecast into a prediction.
- A *projection* is usually a probabilistic statement that it is possible that something will happen in the future if certain conditions develop. In contrast to a prediction, a projection specifically allows for significant changes in the set of boundary conditions that might influence the prediction. As a result, what emerges are conclusions of the type if this happens, then this is what is expected. The simplest type of projection is to extrapolate into the future assuming all of the boundary conditions remain the same or the same trends prevail. For projections extending well out into the future, however, this is often a poor assumption, so scenarios (or story-lines) are developed of what could happen given various assumptions and judgments (see below). For example, the Intergovernmental Panel on Climate Change (IPCC, 2001) recently projected a range of possible temperature increases for the 21st century that calculations indicated would result in the event that the world followed a number of plausible story-lines concerning population and economic growth, energy technologies and emissions, and demographics and international relationships – while also assuming no agreements to limit emissions due to concerns about climate change. By considering how the resulting changes in atmospheric composition would affect the climate using seven different climate models, each with its own particular climate sensitivity (the temperature change that would result from a CO_2 doubling; *see* Climate Sensitivity, Volume 1), the projections of climate change accounted, to a reasonable extent, for a wide range of possibilities of both societal development and climate behavior. This is clearly a projection of what could happen if certain assumed conditions prevailed in the future; it is neither a prediction nor a forecast of what will or is likely to happen, especially given knowledge of what these projections portend. For a decision-maker, a projection

is thus an indication of a possibility, and normally of one that could be influenced by the actions of the decision-maker (*see* Projection of Future Changes in Climate, Volume 1).

- A *scenario* is a set of conditions describing how the future may evolve based on a set of assumptions. While the climate system depends on many rigorous physical laws, there are no such laws governing societal behavior and occurrences such as the invention of new technologies, the choices that lead to war and peace, the prevalence and intensity of diseases and cures, and the growth of the world population. Most important, society also evolves as a result of thinking about the future, perhaps the most unique aspect of human and societal capabilities. To provide a basis for taking advantage of human insight and in recognition that many relevant parameters can vary within such wide ranges that specific predictions are not possible, scenarios are used to envision possible alternative futures, hypothesizing that various factors could turn out in particular ways without predicting that this will occur. Because we cannot know what will happen, scenarios are constructed that describe the possible evolution of relevant parameters in a manner that is detailed, internally consistent, and scientifically sound. Usually, an envelope of possible futures (or scenarios) is designed that allows exploration or illustration of important issues that can enable scientists and decision-makers to think logically about the future. Another important reason for considering such a set of alternative futures is so that policymakers can conduct analyses and design management strategies that do not, for example, foreclose options or lock society into a particular, possibly adverse outcome. For example, one particular scenario might assume world population will reach a particular level, renewable energy sources will become less expensive and more available at a certain rate, the economy will grow at a certain rate, the world s nations will cooperate in sharing technologies, etc. Another scenario might be constructed with a different combination of expectations. For a planner, a set of scenarios is about possibilities, and the frequent shortcoming is in not being imaginative enough. It is for this reason that many such efforts are either too limited, or have resulted in such large ranges of possibilities that the resulting projections at first seem inconclusive. However, it is just this type of outcome that can help decision makers consider the wide range of possible futures (*see* Futures Research, Volume 5; Scenarios, Volume 5).

While scientists generally tend to use these words quite carefully, this is not always the case when the media report on science. For example, in the early 1990s, a survey of scientists was widely reported to indicate that many scientists had low confidence in predictions for the 21st century. This was taken by some to mean that scientists did not believe in the greenhouse effect. However, the result of the survey is not at all surprising given how prediction is defined and generally understood by scientists – of course, scientists cannot offer a definitive indication about what will happen in the 21st century because it depends on energy technologies and political decisions as well as how the climate will respond. However, without contradicting the survey results, the Intergovernmental Panel on Climate Change can report that there is an international consensus among most members of the scientific community on its projections of a temperature rise during the 21st century if emissions follow the projected emissions scenarios.

While these differences in terminology may seem subtle, they are essential for understanding the confidence to have in conclusions and to consider how they can be responsibly used. Outside the scientific community – for example in military and financial planning – preparing scenarios and then using them to explore their impacts on related parameters by constructing projections is a widely used approach. In such studies, it is understood that these are projections of what could happen if a particular situation developed, not predictions of what will happen. Although everyone would like to have reliable (and verifiable) predictions, the complexities of society and climate are such that we are forced to rely on projections if we want to use our capability for understanding the Earth system to look forward into the future in a useful way; otherwise we are limited to advancing forward blindly because we can only rely with certainty on mindless extrapolations of changes that have been observed in the past.

REFERENCE

IPCC (Intergovernmental Panel on Climate Change) (2001) *Climate Change 2000: The Scienti c Basis*, Working Group I Report, Cambridge University Press, Cambridge, 1–881.

Processes, Atmospheric

see Earth System Processes (*Opening essay,* **Volume 1**)

Projection

see Prediction in the Earth Sciences (**Volume 1**)

Q

Quasi–Biennial Oscillation (QBO)

Kevin Hamilton
University of Hawaii, Honolulu, HI, USA

Observations and theory of the Quasi-Biennial Oscillation (QBO) in the stratospheric circulation are reviewed. The QBO is a cyclic variation with slightly irregular periods averaging about 28 months. It appears most clearly in observations of the prevailing winds in the tropical stratosphere, but has signi cant effects on stratospheric temperature and circulation in the extratropics as well. The QBO also has a major impact on the interannual variation in stratospheric ozone, which must be considered when evaluating possible anthropogenic effects in the observed ozone record.

The prevailing east–west winds near the equator at heights between about 18 and 40 km are observed to undergo a remarkable oscillation from strong easterlies (up to about 30 m s^{-1} at some levels) to strong westerlies (about 20 m s^{-1}) with a mean period of about 28 months. This so-called QBO is by far the dominant component of interannual variability in the tropical stratospheric circulation and has a significant influence on variability of temperature and wind in the extratropical stratosphere as well. The variations in atmospheric circulation associated with the QBO significantly influence the ozone layer. The QBO variations in total column ozone can be as much as 6% of the mean value. Thus, QBO effects need to be accounted for when evaluating observed ozone trends.

The first information on the winds at high altitudes near the equator was obtained in 1883 when the dust cloud produced by the explosive eruption of Mount Krakatoa (in modern-day Indonesia) was observed to travel westward around the equator at a fairly constant speed of about 33 m s^{-1}. It was reasonably inferred that the winds near the equator at the heights estimated for the dust cloud (about 20–40 km) were steady easterlies of about 30 m s^{-1}. In the first half of the 20th century a few scattered balloon observations of stratospheric winds were taken at different times and locations near the equator. These sometimes showed the presence of easterlies but also sometimes westerlies. These puzzling results were reconciled at the time by assuming that the Krakatoa easterlies were a broad current in which thin ribbons of meandering westerlies were imbedded. In the 1950s, the first regular daily balloon observations of stratospheric winds and temperatures were begun at several stations near the equator. By 1960 enough data existed for R Reed (University of Washington) and, independently, R Veryard and R Ebdon (United Kingdom Meteorological Office) to determine that the monthly-mean winds in the tropical stratosphere in fact had very little geographical variation around the equator, but had a strong temporal variation, with prevailing easterlies at any height replaced by westerlies roughly every other year. Initially it was thought that the oscillation might have a period of exactly 2 years, but subsequent observations have shown that the QBO is somewhat irregular from cycle to cycle, but with an average period of about 28 months.

Figure 1 shows a long time series of monthly-mean values of the east–west (zonal) component of the wind at 50 hPa (approximately 21 km) computed from regular daily balloon observations at Singapore (1.3 °N). The presence of the QBO is obvious, and this oscillation utterly dominates any other aspect of variability in the record (such as the annual cycle, high frequency month-to-month variability or very long term trends). While the oscillation has been fairly stable through the almost 50 years for which there are detailed observations, Figure 1 also shows some modest irregularity from cycle to cycle. In particular, the period of the oscillation appears somewhat variable, with some anomalously long cycles (e.g., 1966–1969) and short cycles (e.g., 1971–1972). The variations in monthly-mean zonal wind observed at other stations close to the equator are very similar. Thus, the QBO can be thought of as an oscillation in the zonally averaged flow (i.e., the circulation averaged around latitude circles).

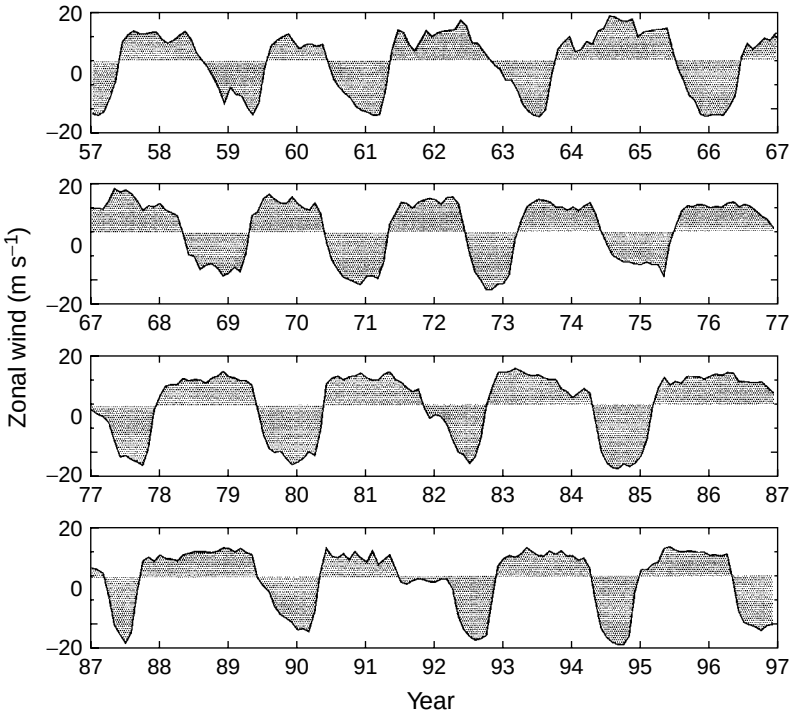


Figure 1 Monthly-mean values of the east–west wind for 40 years (January 1957 through December 1996) measured at 50 hPa at Singapore (1.3 °N, 103.9 °E). Positive values denote westerly winds

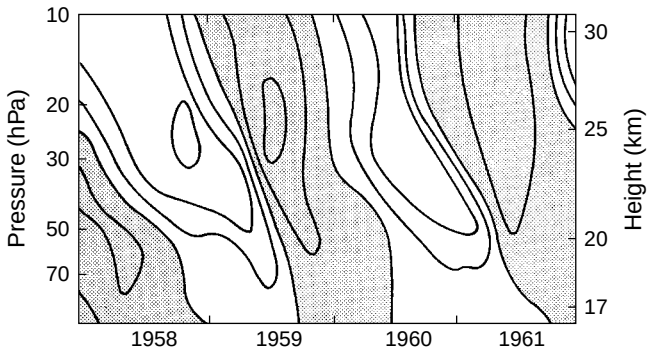


Figure 2 Height–time section of the monthly-mean east–west wind measured at Canton Island (2.8 °S, 171.7 °W) for a period of four years. The contour interval is 10 m s^{-1} and shading denotes westerly wind. Results shown for the height range 17–30 km

The QBO in zonal wind also has an interesting height structure. Figure 2 shows contours of the height–time evolution of the monthly-mean zonal wind over 4 years as measured at Canton Island (2.8 °S). At any level the oscillation between easterlies and westerlies is quite apparent, but the changes do not occur simultaneously at all heights. In fact there is a clear downward progression of the wind reversals at a rate of between 1 and 2 km month^{-1}. This is a consistent feature in each cycle of the QBO that has been observed. The oscillation is strongest near about 30 hPa (• 24 km), and has a large amplitude (generally greater than 20 m s^{-1} peak to peak) over a broad height region (about 20–40 km). Below 20 km the amplitude drops off rapidly and the QBO is not seen clearly in the troposphere. Similarly, above 40 km the QBO amplitude drops strongly, although the QBO is now known to have subtle influences at higher altitudes, particularly in terms of modulating higher-frequency components of the circulation.

Associated with the QBO in zonal wind are oscillations in zonal-average temperature and in the circulation in the meridional plane (i.e., the north–south and vertical wind components). The variations in temperature have been measured by daily balloons and from remote sensing instruments on satellites. The temperature QBO is as large as 8 °C, peak to peak. The QBO in the vertical wind is too small to be directly observed, but it has effects on stratospheric composition. In fact a quite strong QBO in tropical stratospheric ozone concentration is observed. The QBO in total integrated column ozone amounts has a peak to peak amplitude of about 6% of the mean column at the equator. The maximum in the ozone column at low latitudes is coincident with the maximum westerly wind at 50 hPa, clearly showing the strong connection between dynamical and ozone QBOs. The QBO is a very important component of natural variability and must be taken into account when evaluating possible

long-term trends in observed ozone, as well as temperature and wind.

The observed amplitude of the QBO in zonal wind drops off roughly exponentially with latitude in both hemispheres and has only half its equatorial amplitude at about 12° latitude. However, the QBO does appear to have at least some influence on the extratropical stratosphere, although these effects can only be seen as statistical effects in long observational records. In particular, the Northern Hemisphere polar stratosphere tends to be warmer (and the circumpolar vortex weaker) on average in winters when the equatorial QBO in the lower stratosphere is in its easterly phase. It is thought that the correlation with the equatorial QBO can explain about 25% of the interannual variance observed in the high latitude stratosphere. The effects of the stratospheric QBO on aspects of tropospheric circulation have also been investigated. While the possible tropospheric connections remain somewhat controversial, the phase of the QBO (which itself is reasonably predictable for months or even years) has been used as an input into long-term weather forecasts.

In the now standard theory for the QBO, developed initially in 1968 by R Lindzen (University of Chicago) and J Holton (University of Washington), higher frequency (periods of hours to days) wave motions excited by moist convection in the tropical troposphere act as catalysts transferring zonally averaged zonal momentum between different layers of the atmosphere. When a broad spectrum of waves is allowed to interact in this manner with the mean flow, the result can be a long-period oscillation of the zonally averaged flow with the downward progression characteristic of the observed QBO. Random variations in the strength of the convectively forced waves may then be invoked to explain the cycle to cycle variability seen in the QBO record. This model of the QBO has been used as a basis for speculation concerning long-term anthropogenic modification of the QBO. The effects of increasing greenhouse gas concentrations could include modifications to the convective forcing of high-frequency waves in the tropics and to the propagation and dissipation characteristics of the waves in the stratosphere. All these changes should have effects on the properties of the QBO. Continued monitoring of the QBO for possible long-term trends may provide evidence for anthropogenic effects on the atmosphere.

See also: Natural Climate Variability, Volume 1; Stratosphere, Temperature and Circulation, Volume 1.

REFERENCES

Andrews, D, Holton, J R, and Leovy, C B (1987) *Middle Atmosphere Dynamics*, Academic Press, New York.

Hamilton, K (1998) Dynamics of the Tropical Middle Atmosphere: a Tutorial Review, *Atmosphere-Ocean*, 36, 319–354.

Quasi–Decadal Oscillation

Lon L Hood
University of Arizona, Tucson, AZ, USA

A 10- 12 year oscillation of the heights of constant-pressure surfaces in the Northern Hemisphere lower stratosphere was rst reported by Karin Labitzke of the Free University of Berlin (Germany) and Harry van Loon of the National Center for Atmospheric Research (Boulder, CO, USA) in the late 1980s. In addition to this meteorological quasi-decadal oscillation, a similar decadal oscillation of total column ozone exists with a global mean amplitude of 1.5- 2.0%. Like the meteorological oscillation, the quasi-decadal oscillation of total ozone is approximately in phase with the solar cycle (ozone maxima tend to coincide with solar activity maxima) and maximum amplitudes of the ozone oscillation occur near 30° latitude in both hemispheres. Current general circulation models that account for known direct effects of solar ultraviolet variability (changes in ozone production, radiative heating, upper stratospheric winds) as well as indirect dynamical interactions are able to simulate qualitatively several characteristics of the observed quasi-decadal oscillation. However, signi cant quantitative differences remain.

A current issue in global environmental research is the extent to which solar variability can influence climate change on time scales ranging from decades to centuries. In particular, it is important to determine the relative contributions of increasing solar energy output and increasing emissions of human-induced increases in greenhouse gas concentrations in producing the observed surface temperature rise of the past century. Although numerical atmospheric circulation models represent a useful means of estimating the climatic effects of solar forcing, it is essential that these models be tested and validated observationally. One means of testing Sun-climate models may be provided by the *Quasi-decadal Oscillation*, an observed temporal cycle of stratospheric and tropospheric meteorological parameters that correlates approximately with the Sun s 11-year activity cycle.

A 10–12 year oscillation of the heights of constant-pressure surfaces in the Northern Hemisphere lower stratosphere was first reported by Karin Labitzke of the Free University of Berlin (Germany) and Harry van Loon of the National Center for Atmospheric Research (Boulder, CO) in the late 1980s (*see* Stratosphere, Temperature and Circulation, Volume 1). As shown in Figure 1, correlations of 30 mb heights with the solar 10.7 cm radio flux (a ground-based proxy for solar ultraviolet variability)

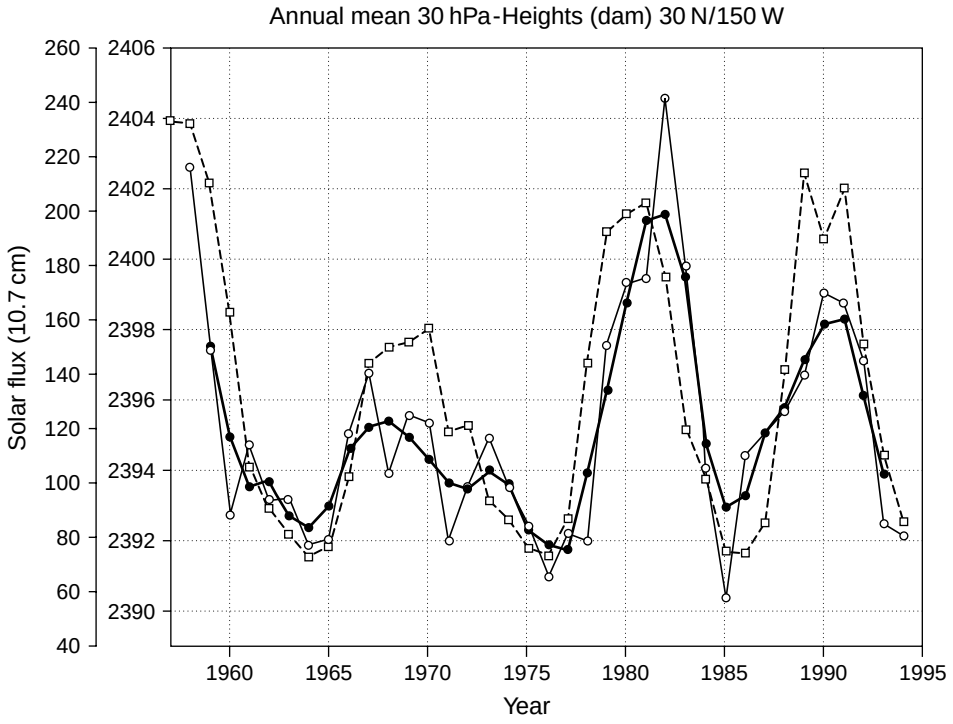


Figure 1 Time series of annual mean 10.7 cm solar flux (open squares), annual mean 30 mb height at 30 °N 150 °W (open circles), and a 3-year running mean of the 30 mb heights for each year (heavy solid line) (K Labitzke, private communication). The heights are in decameters (dam). (Reproduced by permission of *Tellus*)

can be quite high near 30 °N latitude. A similar correlation is present in the Southern Hemisphere with maximum amplitude near 30 °S latitude in winter. Because the heights of constant-pressure surfaces are determined in part by the temperature at lower levels, some evidence also exists for an 11-year solar cycle signal in temperature records for the lower stratosphere and upper troposphere of the Northern Hemisphere.

In addition to the meteorological quasi-decadal oscillation, a decadal oscillation of total column ozone exists with a global mean amplitude of 1.5–2.0%. Because most of the ozone column is in the lower stratosphere and the ozone lifetime there is much longer than dynamical time scales, total ozone is a sensitive measure of circulation changes in this part of the atmosphere. Like the meteorological oscillation, the quasi-decadal oscillation of total ozone is approximately in phase with the solar cycle (ozone maxima tend to coincide with solar activity maxima) and maximum amplitudes of the ozone oscillation occur near 30 °latitude in both hemispheres.

Solar-related energy inputs that could be involved in producing the quasi-decadal oscillation include total irradiance changes (which penetrate directly to the troposphere), changes in solar ultraviolet flux (which affect the production rate of ozone and are largely absorbed in the stratosphere), and changes in energetic particle fluxes and shorter wavelength radiation (which primarily affect the upper atmosphere above the stratopause). Because total irradiance changes are relatively small (about 0.1% from solar minimum to maximum) and the observed oscillation has a maximum amplitude in the lower stratosphere, changes in solar ultraviolet flux, which reach amplitudes of 6–7% near 200 nm wavelength, are considered to be the most likely forcing mechanism. However, indirect effects of solar variability on the upper atmosphere can not yet be eliminated as a significant contributor.

Current general circulation models that account for known direct effects of solar ultraviolet variability (changes in ozone production, radiative heating, upper stratospheric winds) as well as indirect dynamical interactions are able to simulate qualitatively several characteristics of the observed quasi-decadal oscillation. However, significant quantitative differences remain. For example, the observed ozone decadal oscillation differs markedly in latitude and altitude dependence from the simulations of the present climate by existing models. This suggests that an important element of the solar forcing process is not included in the models.

Recent work indicates that a full understanding of the quasi-decadal oscillation, and, by implication, a full understanding of how solar variability affects climate, requires a more detailed investigation of how solar variability may affect another atmospheric oscillation, the quasi-biennial oscillation (QBO) of zonal wind in the equatorial lower

stratosphere (*see* Quasi- Biennial Oscillation (QBO), Volume 1). Specifically, statistical studies suggest that this roughly 2-year oscillation, which is driven internally by upward propagating equatorial wave disturbances originating in the troposphere, may be modulated slightly by the solar activity cycle. Because the equatorial QBO strongly influences stratospheric dynamics on a global scale, any solar-induced perturbation of the QBO would imply dynamical effects in the lower stratosphere and upper troposphere that are not currently accounted for in climate model simulations of solar interactions. Previous work has documented that the QBO influences the occurrence of major mid-winter warmings in the polar zones of the Northern Hemisphere such that they occur primarily during the easterly phase of the QBO. However, as first noted by Labitzke during the 1980s, major mid-winter warmings also occur during the westerly QBO phase under solar maximum conditions. This results in an effective solar cycle variation of January and February mean lower stratospheric temperatures on a decadal time scale when the data are separated according to the phase of the QBO. Recent studies suggest that the westerly phase of the QBO wind oscillation itself in the equatorial lowermost stratosphere (near 50 mb) tends to be shorter in duration near solar maxima than near solar minima. The mechanisms by which solar variability is apparently able to modulate the equatorial QBO have not been fully established. Until these issues are resolved, the ability of current general circulation models to accurately simulate the solar-induced component of climate change will remain limited.

FURTHER READING

Hood, L L (1997) The Solar Cycle Variation of Total Ozone: Dynamical Forcing in the Lower Stratosphere, *J. Geophys. Res.*, 102, 1355–1370.

Intergovernmental Panel on Climate Change (1996) Climate Change 1995, in *The Science of Climate Change*, eds J T Houghton, L G Meira Filho, B A Callander, N Harris, A Kattenberg, and K Maskell, Contribution of Working Group 1 to the Second Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge.

Labitzke, K (1987) Sunspots, the QBO, and the Stratospheric Temperatures in the North Polar Region, *Geophys. Res. Lett.*, 14, 535–537.

Salby, M L and Callaghan, P (2000) Connection between the Solar Cycle and the QBO: The Missing Link, *J. Clim.*, 12, 328–338.

Shindell, D, Rind, D, Balachandran, N, Lean, J, and Lonergan, J (1999) Solar Cycle Variability, Ozone, and Climate, *Science*, 284, 305–308.

van Loon, H and Shea, D J (1999) A Probable Signal of the 11-year Solar Cycle in the Troposphere of the Northern Hemisphere, *Geophys. Res. Lett.*, 26, 2893–2896.

Quaternary

The *Quaternary Period* is the geologic period following the Pliocene, beginning 1.5–1.6 Ma. The Quaternary consists of the Pleistocene and Holocene epochs (*see* Holocene, Volume 1; Pleistocene, Volume 1). The Pleistocene was a period encompassing many glacial-interglacial cycles; the Holocene is the current interglacial period. The Quaternary Period has been marked by repeated cycles of Northern Hemisphere glaciations, which began prior to the start of the Quaternary around 2.5 Ma. The glaciations became more intense starting 700 000 years ago, changing from 41 000 year cycles to 100 000 year cycles (Webb and Bartlein, 1992). The glacial-interglacial cycles appear to be driven by changes in the latitudinal-seasonal pattern of insolation reaching the top of the Earth s atmosphere (caused by cyclic changes in the Earth s orbit around the Sun) and by other internal feedback mechanisms of the climate system (*see* Orbital Variations, Volume 1). Other factors that distinguish the Quaternary are repeated latitudinal climate displacements and changes in oceanic circulation and sea level. Originally the Quaternary was thought to be correlated with the first hominids, but their appearance is now thought to extend back into the Pliocene. The name *Quaternary*, originally used by Marcel De Serres, was coined by Paul Desnoyers (1829) based on nearly horizontal, relatively unconsolidated younger strata in lowland regions, although the term was also earlier used by Lyell to refer to beds containing fossils of relatively young age (Berggren and Couvering, 1977). The term *alluvium*, referring to recent loose and unconsolidated sediments (Werner) and *diluvium*, referring to sediments resulting from the noachian flood (Buckland), are also associated with Quaternary sediments (Prothero, 1990). More robust definitions of the start of the Quaternary (Pliocene-Pleistocene boundary) are based on the Calabrian Formation in Italy, as well as on certain index fossils (Berggren and Couvering, 1977).

See also: Earth System History, Volume 1.

REFERENCES

Berggren, W A and Van Couvering, J A (1978) The Quaternary, in *Treatise on Invertebrate Paleontology, Part A., Introduction: Fossilization (Taphonomy), Biogeography and Biostratigraphy*, eds R A Robinson and C Teichert, Geological Society of America, Boulder, CO, A505–A543.

Prothero, D R (1990) *Interpreting the Stratigraphic Record*, W H Freeman, New York, 1–410.

Webb, III, T and Bartlein, P J (1992) Global Changes During the Last 3 million Years: Climatic Controls and Biotic Responses, *Annu. Rev. Ecol. Syst.*, 23, 141–173.

BENJAMIN S FELZER *USA*

Radiative Forcing

L D Danny Harvey

University of Toronto, Toronto, Canada

Surface and tropospheric temperatures are tightly coupled by non-radiative heat exchanges (namely, by sensible and latent heat fluxes). However, the coupling between the stratosphere and troposphere is rather weak. Thus, if the solar luminosity were to change, or the concentration of greenhouse gases or aerosols were to change, the surface and tropospheric temperatures would respond together, while the surface–troposphere system and stratosphere would respond more or less independently of each other. The net energy gain or loss from the surface–troposphere system is given by the net radiation at the tropopause. For this reason, when climatic change is driven by changes in net radiation, the changes in the surface temperature are driven by changes in the net radiation at the tropopause, not at the surface or at the top of the atmosphere. This change in net radiation is called the radiative forcing, because it is what forces or drives the change in surface climate. The radiative forcing involves the change in the net downward flux of solar energy at the tropopause, the change in the upward emission of infrared radiation at the tropopause, and the change in the downward emission of infrared radiation from the stratosphere.

The change in net radiation at the tropopause before any temperatures are allowed to change is called the *instantaneous radiative forcing*. As stratospheric temperatures change in response to the perturbation in the stratospheric radiative energy balance, the downward emission of infrared radiative energy will change, which will alter the subsequent change in the tropospheric and surface temperatures. Of course, any change in surface–troposphere temperatures will also alter the radiative balance of the surface–troposphere system (by changing the upward emission of infrared radiation), but such changes are part of the feedback which determine the ultimate response of the surface and troposphere. Because the stratosphere responds quickly (within months) and independently of the surface–troposphere system, the effect of any changes in stratospheric temperatures should be included when computing the radiative forcing. This gives the *adjusted radiative forcing*.

In the case of a doubling of the atmospheric concentration of CO_2, the global mean instantaneous forcing, as computed by a variety of researchers, is about 4.0–4.5 W m^{-2} (Cess *et al.*, 1993). Of this, about 1.5 W m^{-2} – or 40% – is due to extra downward emission from the stratosphere. This extra downward radiation causes the stratosphere to cool at the same time as the surface and troposphere warm. The cooling of the stratosphere in turn reduces the downward emission of radiation from the stratosphere by about 0.5 W m^{-2} (for a doubling of CO_2), thereby offsetting part of the initial increase in downward emission caused by the

increase in CO_2. Hence, the global mean adjusted radiative forcing for a CO_2 doubling is $3.5–4.0\,W\,m^{-2}$. On the other hand, the increase in downward radiation at the surface is only about $1.4\,W\,m^{-2}$ in the global mean, but the surface temperature change does not depend on the direct surface forcing. These flux changes, and the associated changes at the top of the atmosphere, are shown in Figure 1. Note that, before adjustment of stratospheric temperatures, the reduction in upward radiation at the top of the atmosphere is considerably less than the forcing at the tropopause, whereas after adjustment, the two are the same.

Table 1 lists the instantaneous and adjusted direct radiative forcings for changes in the concentration of selected greenhouse gases. The instantaneous and adjusted radiative forcings are almost identical in all cases except for CO_2, where the adjusted forcing is about 7% smaller than the instantaneous forcing, and for depletion of stratospheric O_3, where the two forcings differ in sign.

The size of the difference between the instantaneous and adjusted radiative forcings is a direct reflection of the extent of adjustment in stratospheric temperature that occurs. The main energy balance in the stratosphere (in the global mean) is between absorption of ultraviolet radiation by O_3 and emission of infrared radiation by CO_2. Thus, an increase in CO_2 concentration cools the stratosphere, causing the adjusted forcing to be smaller than the instantaneous forcing. Because O_3 is the prime gas responsible for heating the stratosphere, a decrease in stratospheric O_3 causes a large cooling of the stratosphere and a reduction in downward radiation that more than offsets the extra penetration of solar radiation. This in turn changes the sign of the surface–troposphere forcing from positive to negative. A more extensive analysis, explaining the reasons for the differences between CO_2 and other well-mixed gases, is presented in Chapter 10.1 of Harvey (2000).

Table 1 Instantaneous and adjusted direct radiative forcings for changes in the concentration of selected greenhouse gases and for a change in solar luminosity, as summarized by Harvey (2000)[a]

Forcing mechanism	Instantaneous forcing ($W\,m^{-2}$)	Adjusted forcing ($W\,m^{-2}$)
CO_2: 300 → 600 ppmv	4.63	4.35
CH_4: 0•28 → 0•56 ppmv	0.53	0.52
N_2O: 0•16 → 0•32 ppmv	0.96	0.93
CFC-11: 0 → 1 ppbv	0.21	0.22
CFC-12: 0 → 1 ppbv	0.26	0.28
SF_6: 0 → 1 ppbv	0.56	0.62
Stratospheric O_3: 50% reduction	0.26	−1.23
Solar luminosity: 2% increase	4.48	4.58

[a] The forcings for greenhouse gases have an uncertainty of ±10–15%.

To summarize, the surface and troposphere respond as a tightly coupled system to any of the factors that alter the Earth s radiative energy balance. To a good approximation, this coupled response is governed by the change in net radiation at the tropopause, after accounting for any adjustments in stratospheric temperatures that might occur. The adjustment in stratospheric temperature is particularly important in computing the radiative forcing due to changes in stratospheric ozone and to a lesser extent for changes in CO_2.

See also: Infrared Radiation, Volume 1; Projection of Future Changes in Climate, Volume 1.

REFERENCES

Cess, R D, Zhang, M-H, Potter, G L *et al.* (1993) Uncertainties in Carbon Dioxide Radiative Forcing in Atmospheric General Circulation Models, *Science*, 262, 1252–1255.

Harvey, L D D (2000) *Global Warming: The Hard Science*, Prentice Hall, Harlow, UK.

Radiative–Convective Models

see Models of the Earth System (*Opening essary,* **Volume 1**)

Radionuclides, Cosmogenic

Robert Finkel
Lawrence Livermore National Laboratory, Livermore, CA, USA

Cosmogenic nuclides *are radioactive and stable nuclei found in the Earth, moon and meteorites. They are produced by energetic particles from outside the solar system called* cosmic rays*. Amounts and ratios of radionuclides can be used to derive a great deal of information about Earth system history and solar activity.*

The most widely studied cosmogenic nuclides (with their half-lives) are: ^{3}He (stable), ^{7}Be ($t_{1/2} = 53$ days), ^{10}Be ($t_{1/2} = 1{\cdot}5 \times 10^{6}$ years), ^{14}C ($t_{1/2} = 5730$ years), ^{21}Ne (stable), ^{26}Al ($t_{1/2} = 7{\cdot}0 \times 10^{5}$ years), ^{36}Cl ($t_{1/2} = 3{\cdot}0 \times 10^{5}$ years). Because they are produced only in the atmosphere or outer layers of solid bodies in the solar system, cosmogenic nuclides are especially useful for studying surface processes. Thus far they have been used to study the timing of glaciations, to assess past solar energy output, to trace the mixing of the atmosphere, to determine how fast mountains erode, to estimate earthquake recurrence intervals, to investigate the rate of climate change, and to probe many other environmental processes. Their ability to help determine both the rates at which Earth system processes occur and the dates of events in the Earth s history makes cosmogenic nuclides extremely valuable tools in Earth science research.

The employment of cosmogenic nuclides in terrestrial environmental studies was at first limited by their extreme rarity in nature. However, developments in noble gas mass spectrometry and accelerator mass spectrometry during the last two decades improved measurement sensitivity to the point that as few as a million atoms of a cosmogenic nuclide could be accurately determined. With this remarkable increase in sensitivity, it became possible to use cosmogenic nuclides to address a number of open problems in the Earth and environmental sciences. A brief and selected survey of both the technique and the application of cosmogenic nuclides in the Earth sciences follows.

^{14}C is the most widely used cosmogenic nuclide. It is produced primarily via the exothermic ^{14}N(n,p)^{14}C reaction between cosmic ray–derived neutrons and atmospheric nitrogen. ^{14}C produced in the atmosphere rapidly equilibrates with volatile carbon species so that the ^{14}C/C ratio in the atmosphere is nearly homogeneous. Once ^{14}C containing carbon is incorporated into plant material, no further chemical exchange occurs and thereafter the ^{14}C content decreases slowly with time due to radioactive decay with a half-life of 5730 years. The age of organic material can therefore be determined by comparing the ^{14}C content of plant carbon with the ^{14}C content of atmospheric carbon. In practice, corrections must be made for variations in the atmospheric content of radiocarbon due to small changes in the cosmic ray flux and due to exchange between the atmosphere and other carbon reservoirs. These corrections can be made by measuring the ^{14}C content in samples whose age is otherwise known, e.g., in tree rings whose age has been determined by direct counting. First used to date cultural artifacts from the last 50 kyr, ^{14}C has now found wide application to dating such natural events as the timing of climate change recorded in marine and terrestrial sediments and the change in ocean circulation over time (*see* Younger Dryas, Volume 1).

^{10}Be and ^{36}Cl are produced in the atmosphere through cosmic ray interaction with atmospheric oxygen and argon, respectively. When deposited in glaciers by snowfall, these nuclides record the cosmic ray flux at the time the snow fell. Ice cores from polar regions contain a record of the cosmogenic nuclide content of the atmosphere for hundreds of thousands of years in the past. The effectiveness with which the solar magnetic field shields the Earth from cosmic rays waxes and wanes with overall changes in solar activity. This fact has allowed the ice core record to be used to derive past solar energy output in the period long before instrumental measurements began to be made. For example, the historic cold period in the 17th and 18th centuries, known as the *Little Ice Age*, was a time with very low sunspot activity and with elevated ^{10}Be and ^{36}Cl in ice cores. Similar periods of elevated cosmogenic nuclide content occur throughout the ice core record and are being used to investigate the possibility that variations in solar output may have had an effect on climate.

The atmosphere also contains ^{7}Be, a short-lived isotope of beryllium, that has been combined with ^{10}Be to probe atmospheric mixing. Because ^{10}Be and ^{7}Be are produced with a fixed ratio in the atmosphere and because ^{7}Be has a half-life which is of the same order as the atmospheric mixing time, the ratio of these nuclides has been used to date air masses and to trace the exchange of air between the troposphere, below, and the stratosphere, above.

Cosmogenic nuclides are also produced in rocks at the Earth s surface, even though the flux of cosmic rays decreases a thousand-fold by the time they reach the bottom of the atmosphere. By measuring the concentrations of the radioactive nuclides ^{10}Be–^{26}Al–^{36}Cl and the stable nuclides ^{3}He and ^{21}Ne, geologists have been able to determine the length of time a rock has spent at the Earth s surface. This dating technique – cosmogenic nuclide surface exposure dating – has been applied to a number of problems in geochronology which were previously intractable: direct dating of glacial moraines, fluvial terraces, debris flows, and lava flows. These ages have been important in determining the timing of advances

and retreats of mountain glaciers and in understanding how these regional events relate to global climate change recorded in marine sediments and ice cores.

The field of earthquake studies has also benefited from the application of cosmogenic nuclide surface exposure dating. There is much interest in determining the recurrence interval between earthquakes, a parameter important in assessing the seismic risk to a particular area. ^{14}C in organic material collected from displaced formations in trenches can often be used to determine the timing of fault motion. When organic material is lacking, as is often the case in arid regions, or when fault motion occurred before the $\sim$40 kyr time limit accessible with ^{14}C, surface exposure dating of features which have been offset by earthquake motion such as alluvial fans or moraines has been used to determine slip rates. Several such studies have focused on the uplift of the Tibetan plateau. This method is now being extended to western North America and other regions in the world.

Even subtle questions such as why some hills are covered with soil while others consist of exposed bedrock can be approached using cosmogenic nuclides. The surface of the Earth is constantly changing due to competition between landform displacement and erosion. The rates of these processes have been very difficult to determine. Because the concentration of cosmogenic nuclides in a rock depends on erosion as well as on the exposure time, cosmogenic nuclide concentrations in surface rocks have provided clues to help understand why some landscapes erode quickly and others are more stable, to determine what controls the conversion of bedrock to soil, and to study the rate at which cobbles migrate along stream catchments.

See also: Earth System History, Volume 1; Natural Records of Climate Change, Volume 1; Solar Variability, Long-term, Volume 1.

Radiosondes

Walter F Dabberdt

National Center for Atmospheric Research, Boulder, CO, USA

The radiosonde is an expendable balloon-borne electronics and sensor package that routinely measures the variation with altitude of temperature, humidity and pressure from the ground surface to heights up to about 30 km (a pressure altitude of about 11 hPa). When the device also measures winds, it is more correctly referred to as a rawinsonde, although the term radiosonde is commonly applied to both. In some cases, the balloon alone is tracked by either optical or radar techniques in order to measure only winds. This type of balloon is known as a pilot balloon or simply a pibal. The height pro le of these important meteorological parameters constitutes an upper-air sounding. In 1999, there were 100 operational radiosonde stations in the US that made a daily average of 182 soundings. In the continental US, an average distance of 315 km separates radiosonde stations. In 1999, there were 992 radiosonde stations worldwide that made an average of 1209 soundings each day, while an additional 65 pibal stations made 576 wind soundings daily. The number of soundings are down considerably from their peak values in 1988 of 1660 and 964, respectively. Since 1957, all stations make their soundings at the same times, 00Z and 12Z (Greenwich Mean Time or universal standard time), although many stations outside the US have reduced soundings to once per day due to budgetary constraints. Countries launching operational radiosondes are members of the World Meteorological Organization's World Weather Watch program; as such, they freely share their sounding data with each other. Shortly after an operational upper-air sounding is completed, a standard data message is prepared and made available to all nations using the Global Telecommunications System. The message is transmitted in a universal format that reports conditions at various standard or so-called mandatory (pressure) levels, as well as at signi cant levels, which represent levels where prescribed changes in meteorological conditions occur. There are two primary purposes of upper-air soundings: to analyze and describe current weather patterns; and to provide inputs to short- and medium-range computer-based weather forecast models. Other uses include climate studies, air pollution investigations, aviation operations, and national security applications. The radiosonde continues to be the backbone of an eclectic suite of measurement technologies (both in situ and remote measurements, ground-based and satellites) that are used to provide data for input to numerical weather forecast models.

The radiosonde is carried aloft by a balloon as part of a flight train. The balloon itself may be made of either natural rubber (latex) or synthetic rubber (neoprene). The mass of the flight train, the desired ascent rate, the type of gas used, and the maximum height of the sounding determine the size of the balloon that is used. Operational radiosonde systems typically use balloons that may weigh anywhere from 300 to 1200 g and are filled to ensure an ascent rate of 300 m min^{-1}. Hydrogen is the gas most commonly used to inflate the balloon and provide its lifting capacity, although helium and natural gas are sometimes used for special applications. The flight train consists of four components: a parachute that is used to safely bring the radiosonde back to the surface after the balloon bursts; 20–36 m of nylon line (to isolate the radiosonde s sensors from water vapor contamination by the balloon surface);

a de-reeler to let out the nylon line; and the radiosonde itself. A few countries, such as the US and Switzerland seek to recover and then re-use their radiosondes; in the US, it is estimated that about 18% are reused after extensive refurbishment.

The radiosonde comprises three major sections: a suite of sophisticated meteorological sensors; signal processing electronics; and a radio transmitter to relay the measurements back to a receiver at the radiosonde launch station (Figures 1 and 2). The frequency with which meteorological measurements are made and reported varies according to the type of radiosonde, with a full set of data measured every 1–6 s. The meteorological community has been assigned two radio frequency bands for use in transmitting meteorological data: 400–406 and 1675–1700 Mhz.

Thermodynamic sensor types vary widely among radiosondes currently in use throughout the world. Temperature sensors are of three designs: thermistors, resistance wires, and bimetallic elements. The two common humidity elements are carbon hygristors and thin-film capacitance sensors. Pressure measurements are typically made with either an aneroid cell or a piezoresistance element. There are about a dozen different radiosonde designs presently in use throughout the world. The global radiosonde network has evolved over the last half-century, incorporating technological advances in radiosonde design to meet the changing needs of the world s weather forecasting centers (Table 1). However, the changes in radiosonde design and technology have created special challenges and limitations to climatologists seeking to piece together a consistent and homogeneous multi-decadal global database to analyze and understand climate change. As a result, climate researchers have had to account for biases in the historical records due to changes in instrumentation and observation methods, many of which have poor or no documentation. In the US alone, these changes have been varied and significant. Four distinctly different humidity sensors have been in use since 1943. Temperature measurements have undergone major changes, including sensor type, size, and coating, exposure to the air stream, and corrections to account for radiation biases. At present, the US National Weather Service uses radiosondes from two different manufacturers, each having its own distinct set of pressure, temperature and humidity sensors.

Figure 1 First radiosonde prototype built in 1931 by Professor Vilho Väisälä in Ilmala, Nr Helsinki, Finland. The sonde weighed 420 g and measured 30 cm in length and 7.5 cm in diameter. It included only one sensor, a bimetallic thermometer. The signal from the sonde's radio transmitter was received up to a height of about 7 km

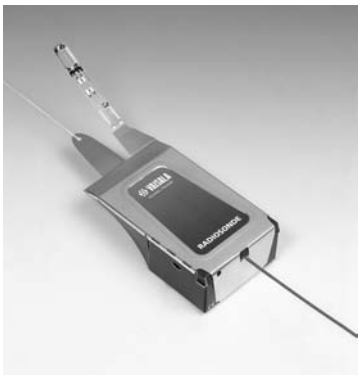


Figure 2 The Vaisala RS90 radiosonde was introduced in 1995. It weighs 290 g and measures $15 \times 9 \times 5$ cm. The arm (top right) contains the temperature sensor and dual heated thin-film humidity sensors. The pressure transducer and the navigation-signal receiver are mounted inside the housing with the battery, electronics and transmitter

Wind measurement methods vary widely as well. Winds can be measured directly, or they may be determined from the drift of the balloon. One class of wind measurement techniques tracks the balloon using any one of three methods: (1), an optical system that uses a theodolite to visually track the balloon s azimuth and elevation; (2), a radio theodolite that tracks a radio signal sent from a transmitter on the radiosonde, again to obtain azimuth and elevation information; and (3), a radar system that tracks a radar retroreflector suspended from the balloon to obtain slant range, azimuth and elevation. The first two methods require knowledge of the balloon height, which is determined from the pressure measurement. Theodolite measurements suffer from the requirement for clear conditions between the ground station and the balloon. All three of these tracking methods determine the wind from the change in the observed location of the radiosonde. Tracking methods lose accuracy as the distance to the balloon increases and its elevation angle decreases.

The second class of wind measurement techniques uses various navigation systems to determine the change in position of the radiosonde or measure the drift velocity

Table 1 Historical Milestones Leading to Development of the Modern Meteorological Radiosonde

Year	Milestone
1643	Evangelista Torricelli invents the barometer in Florence, Italy
1648	French mathematician Blaise Pascal observes the decrease of atmospheric pressure with altitude
1749	Alexander Wilson, Glasgow, Scotland, uses kites to study the variation of temperature with altitude
1783	The French Montgolfier brothers, Jacques Etienne and Joseph, invent the balloon
1783	Jacques Alexandre César Charles, Paris, France, uses a manned balloon to make first measurements of variations of pressure and temperature with altitude
1784	Englishman John Jeffries, London, and Frenchman Jean-Pierre Blanchard begin the systematic study of the atmosphere using manned balloons
1804	French physicists Louis Gay-Lussac and Jean Baptiste Biot ascend to 7 km in a balloon and discover that water vapor decreases with altitude
1822	Englishmen Sir Edward Parry and Rev. George Fisher use kites with recording thermometers to study the Arctic atmosphere
1892	Frenchmen H. Hermite and G. Besancon launch the first free-flying weather balloon with a mechanical recording system (the *meteorograph*)
1847	William Radcliff Birt is first to measure winds aloft (and temperature) with a kite flown from Kew Observatory, London
1893	Lawrence Hargrave, Sydney, Australia, invents the box kite; by the end of the decade, many major observatories are using box kites routinely to measure the atmosphere, including: Blue Hill Observatory (n. Boston), Central Physical Observatory (Moscow), Trappes (n. Paris), Kew Observatory (n. London), Lindenberg (Germany), and Ilmala (Finland)
~1900	British scientist W.H. Dines invents the mechanical meteorograph design that is widely used until 1939
1901	Richard Assmann, Germany, is first to use *extensible* rubber balloons for free-flying soundings with meteorographs
1917	Germans F. Herath and M. Robizsch use the *telemeteorograph* to transmit meteorological data from a kite using the steel kite cable as the signal cable
1920	US Weather Bureau and Army Air Corps establish a program of daily upper-air soundings using airplanes at 20 locations nationwide
1921	US Weather Bureau establishes a kite network for routine upper-air observations; remains in operation until 1933
1928	Professor P. Moltchanoff, Director of the Russian Aerological Observatory in Slutzk, publishes a paper that outlines the design of a meteorological sonde that uses a radio transmitter to send data back to recording stations
1931	Moltchanoff (Russia), Paul Duckert (Germany) and Vilho Väisälä (Finland) independently devise and fly the first radiosonde using radio oscillators that are controlled by output of temperature, humidity and pressure sensors
1974	The National Center for Atmospheric Research (Boulder, CO, USA) develops a special radiosonde called the dropsonde that is launched from research aircraft and measures winds, pressure, temperature and humidity while descending on a parachute
1976	The Vaisala Oy company (Helsinki, Finland) introduces the first computer-controlled upper-air sounding systems
1982	The US National Oceanic and Atmospheric Administration begins routine use of dropsondes for hurricane research; 1 year later, the US Air Force initiates its hurricane reconnaissance program
1995	First commercial radiosonde systems using the Global Positioning System (GPS)[a] to measure winds are introduced by the Atmospheric Instrumentation Research Company (Boulder, CO, USA) and the Vaisala Oy Company (Helsinki, Finland)

[a] GPS is a constellation of 24 orbiting satellites that provide precise measurements of time that are used to locate the position and determine the movement of surface-based and airborne objects to very high levels of accuracy.

directly. Two systems currently in use employ the LORAN-C navigation system and various VLF systems, such as the Russian ALPHA system and the US Navy s VLF system. A transponder on the balloon receives the navigation (time-of-arrival) signals and rebroadcasts them back to the ground station with the meteorological measurements. Both systems require pressure to determine the height of the balloon. These systems are being phased out as the world community ceases to use these navigation signals.

A new navigation-based windfinding technique is now coming into widespread usage. A GPS receiver inside the radiosonde very accurately measures the horizontal and vertical Doppler velocity of the radiosonde with respect to the individual GPS satellites that can be observed at any given time (typically, 4–8 satellites). Other types of GPS receivers also observe the latitude, longitude and altitude of the radiosonde. In both cases, the GPS receiver directly measures the drift velocity of the balloon, and hence, the wind. Two major advantages of the GPS-based

techniques are the high accuracy and precision of the wind measurements, together with the worldwide coverage of GPS. A current impediment is the relatively high cost of GPS technology for expendable radiosonde applications.

The airborne counterpart to the conventional radiosonde (sometimes also called an upsonde) is the dropsonde, which is ejected from research aircraft and floats downward on a special balloon-like parachute. Dropsondes are used primarily in weather research studies and for hurricane reconnaissance. Current state-of-the-art dropsonde sensors include capacitance fine-wire sensors to measure temperature, capacitance silicon pressure sensors, and GPS receivers to measure winds. Humidity is measured with a pair of thin-film capacitance sensors that are alternately heated to avoid condensation on descent from colder to warmer air. All measurements are made twice every second while the dropsonde falls at a rate of $10-25\,\mathrm{m\,s^{-1}}$. Experience has shown that dropsonde hurricane reconnaissance observations have appreciably improved the accuracy of forecasts of hurricane landfall.

See also: Tropospheric Temperature, Volume 1.

Rain

Rain is liquid water precipitation made up of drops larger than 0.5 mm diameter. Precipitation with drop sizes less than 0.5 mm is referred to as drizzle. Precipitation forms when one or more processes lead to the aggregation of a large number of cloud droplets (on the order of one million) into a single drop or ice crystal. Clouds form when air is lifted, typically the result of the lifting by fronts and in cyclones, or by buoyant thermals initiating convection, resulting in expansional cooling that leads to condensation of water vapor into cloud droplets. Condensation occurs on aerosols called *cloud condensation* nuclei (microscopic particles, usually dust or salt particles from sea spray). It is very common for a cloud to be made up of liquid water droplets, even if all or a portion of the cloud is at a height where the temperature is well below freezing, because nuclei that support direct sublimation of water vapor to ice are very much rarer in natural air than condensation nuclei, and generally do not support sublimation unless the temperature is lower than $-10\,^{\circ}\mathrm{C}$. Droplets not in contact with a surface can remain liquid for temperatures down to $-40\,^{\circ}\mathrm{C}$. The typically rare ice crystal in a cloud can grow quickly at the expense of the water droplets, however, because the saturation vapor pressure over ice is lower than that over water, resulting in the liquid water droplets evaporating and supplying their water to the ice crystals. This process, called the Bergeron–Findeisen process, can create ice crystals large enough to fall as precipitation. As the ice crystal falls, cloud droplets it collides with are frozen on contact, increasing the mass of the crystal, causing it to fall faster, sweep out more cloud droplets, and grow even more quickly. Almost all mid-latitude rain forms initially through ice crystal processes in cloud regions where the temperature is below freezing, and melts as it falls into warmer air on its way toward the surface. Rain can also form when there are no ice crystals present through the collision and coalescence process. Here, cloud droplets that are slightly larger than their neighbors fall relative to them, striking some and coalescing into a larger droplet that falls faster, sweeps out more droplets, and grows even more quickly. This process is important for the formation of rain in the tropics, where clouds at temperatures above freezing throughout can produce heavy rainfall, but it also clearly contributes to rainfall in clouds whose upper portions are cold enough to initiate precipitation through ice processes as those crystals fall and melt in the lower portions of the cloud.

See also: Clouds, Volume 1; Hydrologic Cycle, Volume 1.

KEITH L SEITTER *USA*

Remote Sensing, Terrestrial Systems

see Remote Sensing, Terrestrial Systems (**Volume 2**)

Residence Time (of an Atom, Molecule or Particle)

Residence time is the average time spent in a reservoir by an individual atom, molecule or particle. Thus, the residence time is the average length of time that a material will remain in the reservoir before it is removed or destroyed. Alternatively, the residence time is indicative of the age of the molecule or material when it leaves the reservoir.

Residence time is often used interchangeably with lifetime, particularly for gases or aerosols in the atmosphere (*see* Lifetime (of a Gas), Volume 1). Like lifetime, residence time is generally given in terms of the time to reach $1/e$ of its original value if all sources of the material into the reservoir were turned off. For chemically reactive gases or particles, lifetime tends to be used to represent the rate

of chemical reactivity. However, lifetime is also used in the more general sense. Residence time is associated with all processes that can remove the material from the reservoir including transport from one reservoir to another or chemical reactions.

If we consider a box model representation of a reservoir, then material in the reservoir can be removed either by flowing out of the reservoir or by processes destroying the material within the reservoir, such as by chemical reactions changing the nature of the material. At steady state, the total loss of the material in the reservoir through transport out of the reservoir or by destruction within the reservoir will be equal to the total sources through transport into the reservoir or production within the reservoir. The residence time is then equivalent to the total amount of the material in the reservoir divided by the total losses or total sources.

The case of carbon dioxide, CO_2 presents an interesting illustration of the difference between lifetime and residence time. For CO_2, the residence time of a particular molecule in the atmosphere is only about 4 years, because of CO_2 exchanges with the land surface and upper ocean. The lifetime of a pulse injection of CO_2, however, is roughly 100 years because the atmospheric, terrestrial, and upper ocean carbon reservoirs are nearly in balance. For CO_2 containing carbon-14 from atmospheric nuclear testing, however, the lifetime for the pulse injection in the 1960s was initially only about 4 years because the terrestrial and oceanic reservoirs were empty and so did not return carbon-14 to the atmosphere. Understanding these differences is critical to avoiding confusion about the long time-scale of the use of increasing amounts of fossil fuels.

DONALD J WUEBBLES *USA*

Revelle, Roger Randall Dougan

(1909- 1991)

Roger Revelle was one of the 20th century s most eminent, multi-faceted, productive, and influential environmental scientists. His interests, research, and achievements spanned the physical, biological, and social sciences, engineering, and the humanities. He contributed immensely to the development of modern oceanography, but also brought fresh insights to issues of population, world poverty, and hunger. At the same time, he was an inspiring and effective leader of scientific enterprises, an insightful educator, intellectual architect of a great university, and powerful supporter of international programs. He was one of the first leading figures to place the issue of global climate change on the front burner of world concern, and may be viewed as the intellectual godfather of today s international programs dealing with global environmental change.

Roger Randall Dougan Revelle was born in Seattle, Washington State, USA, on March 7, 1909 into a family of lawyers and schoolteachers. He grew up in Pasadena, California and, showing formidable intellectual powers at an early age, he was admitted to Pomona College at the age of 16. His interests first turned to geology, and after receiving his BA degree in 1929, he undertook graduate studies at the University of California, Berkeley, in 1930, where he began to study marine sedimentation as a graduate research assistant at the Scripps Institution of Oceanography (SIO) in La Jolla, CA. In 1931, he married Ellen Clark, a student at nearby Scripps College and descendant of the founder of SIO.

At Scripps, he pursued his doctoral studies on the composition and physical properties of deep-sea sediments, participating in numerous oceanographic cruises. His research focused on the nature of marine carbonates, and the buffering capacity of seawater to control carbonate solubility, research that built a strong foundation for his later work on the carbon dioxide (CO_2) issue.

In 1936, Revelle was awarded his doctorate from the University of California, and he then undertook postdoctoral studies at the Geophysical Institute in Bergen, Norway. It was during this period that he began his active involvement in international scientific affairs. Returning to California, he taught marine geology and physical oceanography at Scripps and the University of California at Los Angeles.

During World War II, Revelle served in the US Navy, conducting research on radar, sonar, and application of oceanographic research to wartime needs. After the war, he organized the scientific program connected with early atomic bomb tests, an activity that left him with a lifelong concern for the marine environment and for the need to avert nuclear war. Transferred to the Office of Naval Research, he served as head of its Geophysics Branch and left the Navy with the rank of Commander in 1948.

Returning to SIO, he led a survey in the Gulf of California to address fisheries problems. Revelle developed this into a continuing interdisciplinary program that brought together fisheries biologists and marine scientists, demonstrating the holistic approach that was the hallmark of his work. In 1951, he became director of SIO, and initiated a rapid expansion of the institution in terms of facilities, staff, and sphere of interest. Scripps s field observations

during this period played a crucial role in developing a more detailed scientific understanding of the Pacific Ocean.

Revelle s own research addressed the continuing addition of CO_2 to the atmosphere by burning of fossil fuels, and its implication for global climate. In a seminal 1957 paper, Revelle and Hans Suess termed this a great geophysical experiment. He commissioned Charles David Keeling to initiate a program of precision atmospheric CO_2 measurements on Mauna Loa, Hawaii and at the South Pole that starkly revealed the extent of human influences on the global atmosphere (*see* Keeling, Charles David, Volume 1). He later chaired a National Research Council study of the impact of energy production on climate, and contributed greatly to a subsequent broader study. At the same time, Revelle vigorously expanded SIO s work in biochemistry and microbiology, and chaired several national committees in this area, including, in 1965, chairing the subpanel that prepared the US Government s first report on CO_2-induced climate change for the President s Scientific Advisory Committee Panel on Environmental Pollution. Internationally, he played an important role in the creation of the International Council for Science (ICSU) Scientific Committee on Problems of the Environment, and the International Geosphere–Biosphere Programme (*see* IGBP (International Geosphere- Biosphere Programme), Volume 2; SCOPE (Scienti c Committee on Problems of the Environment), Volume 4). He remained deeply involved in a wide variety of United Nations (UN) and ICSU activities for nearly four decades, including the Scientific Committee on Ocean Research, serving as its first president (*see* SCOR (Scienti c Committee on Oceanic Research), Volume 1).

Expansion of the University of California system led to development of that institution s San Diego campus adjacent to SIO in La Jolla. Revelle played a central role in the intellectual development of the new university, and the recruiting of its faculty. He then moved on to Washington DC as science advisor in the Department of the Interior, back to California as University Dean of Research, and to Harvard University as Director of the Center for Population Studies. These positions led him into a diverse range of issues: pesticide damage, irrigation, global food supply, and population – issues still high on the contemporary agenda. A particular concern was the plight of the developing countries, and he was a frequent and valued visitor to India. He also chaired the National Academy of Sciences Board on Science and Technology for International Development, and advised many US and UN agencies.

In 1976, Revelle returned to San Diego to become professor of science and public policy. He kept up an active program of interdisciplinary seminars and teaching until his death in 1991. In his honor, SIO named a multidisciplinary research ship the *RV Roger Revelle*. The Revelle College of the University of California, San Diego, also memorializes his name.

Roger Revelle earned many academic and professional honors. He was elected to the US National Academy of Sciences in 1958, and received its Agassiz Medal in 1963. He was a member and fellow of numerous other learned societies. His prizes included the Order of the Sitara-Imtiaz from Pakistan, the William Bowie Medal of the American Geophysical Union, the Tyler Medal, and the National Medal of Science.

JOHN S PERRY *USA*

Richardson, Lewis Fry

(1891- 1953)

Lewis Fry Richardson is today best known for his pioneering attempt at weather prediction by numerical methods. However, during his long and productive career he made seminal contributions to a remarkable range of scientific and social problems, and stands as one of the most influential scientific figures of the 20th century.

Born of a Quaker family in Northumberland, UK, Richardson was educated at the University of Newcastle and then Cambridge University. Faithful to his pacifist values, he performed non-combatant service during World War I. While at the front, he began to address the problem of weather prediction by numerical solution of the fundamental physical equations, and actually carried out a numerical prediction of pressure change. Inadequacies in both data and numerical technique produced unrealistic results, but the giant step from witchcraft to quantitative science had been taken. Subsequently, in his famous book *Weather Prediction by Numerical Process*, he imagined how real-time predictions might be produced by a vast auditorium full of clerks orchestrated by a central director – a vision uncannily resembling the organization of parallel-processing computers of today!

Richardson s fundamental research in atmospheric fluid dynamics centered on the cascade of energy from larger scales of motion down to smaller eddies to be ultimately dissipated through viscosity, a process that he immortalized in a celebrated poem:

> Big whorls have little whorls that feed on their velocity, and little whorls have smaller whorls and so on to viscosity – in the molecular sense•••.

As a consequence, Richardson s name is attached to the non-dimensional number associated with fluid convective stability. At the same time, he saw how these eddy processes governed turbulent diffusion and limited the range of predictability. In the course of this work, he developed concepts of scaling and fractality that have had broad application to scientific and practical problems in many fields.

From 1920 onwards, he worked in academia in England and Scotland, and was elected a Fellow of the Royal Society in 1926. His interests turned increasingly to questions of war and peace, where he applied mathematical modeling techniques and statistical analysis. These studies and his database on deadly quarrels have influenced many peace researchers and continue to be a rich source of ideas. This work led to books such as *Arms and Insecurity*, and *Statistics of Deadly Quarrels*, and continues at the Richardson Institute for Peace Studies (University of Lancaster, UK).

See also: General Circulation Models (GCMs), Volume 1; Models of the Earth System, Volume 1.

JOHN S PERRY *USA*

Roberts, Walter Orr

(1915- 1990)

In the more than four decades of his career in scientific research and institution building, Walter Orr Roberts became internationally recognized as one who broadened the vision of meteorologists and astronomers and who deepened the understanding of science by industrialists, political leaders, and other policy makers.

Born in 1915, Roberts grew up in Massachusetts. He received his PhD in astronomy from Harvard University in 1943. His early work focused on observations of the Sun, especially solar activity and the chromosphere. This work, conducted from a mountain observing station, which he developed, led him to establish a new organization, the High Altitude Observatory (HAO). At HAO he assembled a staff of solar physicists, scientists with interests in solar–terrestrial interactions, planetary scientists, and those interested in meteorology.

This diversity of activity was unusual among astronomical observatories, but it proved appropriate for exploiting the rapid expansion of observational opportunities for measurements of the Sun, the space between the Sun and the Earth, and the Earth s atmosphere.

In 1960, Roberts became director of the newly established National Center for Atmospheric Research (NCAR) in Boulder, CO. He brought to this institution the same belief that a scientifically diverse staff was more in keeping with modern needs than a narrowly specialized one. The early recruitments to the NCAR staff included physicists, engineers, applied mathematicians, and chemists as well as meteorologists.

Roberts also pursued his interest in possible Sun–weather connections. His careful approach to this topic restored some respectability to a subject damaged by a plethora of published correlations of sunspot numbers with records ranging from typhoon frequencies to wine vintage quality. He also became prominent in the activities leading to the Global Atmospheric Research Program (GARP) (*see* GARP (Global Atmospheric Research Program), Volume 1), and he was instrumental in the design of an oversight mechanism for the program involving cooperation between the World Meteorological Organization and the non-governmental International Council of Scientific Unions. The GARP field programs – in the Caribbean, the tropical Atlantic, monsoon regions, and global – produced observations and studies that built a foundation for subsequent rapid progress in the quality of weather forecasts and in the understanding of climate and the possibility of a human-induced climate change.

His scientific work expanded to include the climate–agriculture connection, the world s food system and, indeed the world s future. As an organizer of conferences, a popular speaker, and an advisor to political and industrial leaders Roberts was a force in bringing a possible climate change to the attention of thoughtful policy makers years before the idea achieved the currency and validity that it has today. Roberts died in Boulder in 1990, having spent much of his later years participating in the communications revolution made possible by personal computers. He had earlier been a pioneer in reaching out to scientists in the Soviet Union during the uneasiest times of the Cold War, and he organized an ongoing computer conference between US and Soviet scientists discussing how a severe climate change might affect relations between these two giants.

JOHN FIROR *USA*

Rossby, Carl-Gustaf

(1898- 1957)

Carl-Gustaf Rossby was a Swedish–American meteorologist, preeminent as a theoretician, and also one of the first scientists to realize the importance of environmental influences on the global climate. His achievements were equally significant in the realms of organization and in theoretical meteorology. In both, his characteristically broad view and bold approach made spectacular achievements possible.

After a BSc at the University of Stockholm in 1919, Rossby worked for 7 years as a weather forecaster in Norway and Sweden. In 1926 he moved to the United States where he created the first airways weather service. As a professor at the Massachusetts Institute of Technology from 1928 to 1939, he established the first advanced educational program in meteorology. In 1940 he moved to the University of Chicago where he organized university training programs for thousands of American meteorologists during World War II. After the war he led an intense effort to study the general circulation, in particular the jet streams. He also reorganized and revitalized the American Meteorological Society, and established *Journal of Meteorology* (nowadays *Journal of Atmospheric Sciences*) as a leading professional journal.

During the last 10 years of his life Rossby spent an increasing amount of time in Sweden, where he established the International Institute of Meteorology at University of Stockholm, founded the geophysics journal *Tellus* and took part in the development of the first computer-produced weather forecasts. He also started an ambitious research program in atmospheric chemistry.

Rossby was associated with the Woods Hole Oceanographic Institution from its founding. In addition to membership in the National Academy of Sciences in the United States and being an honorary member of the Royal Meteorological Society, he was a foreign member of the Norwegian Academy of Science and the Royal Swedish Academy of Science.

In the 1930s, the severe drought in the United States led to government support of Rossby s studies in atmospheric motion and the general circulation of the atmosphere. During attempts to devise a 5-day forecasts system he became interested in the interactions between fluctuations in the strengths of the westerlies and the movements and developments of the large-scale anticyclones and cyclones. He saw a striking similarity between the long-period variations of the world weather patterns and the week-to-week cycle of alternating zonal and meridional flows (*see* Rossby Waves, Volume 1). He believed that climate variability on almost all time scales, from inter-annual to centennial and beyond, could be understood in terms of changes of states of the atmospheric flows.

During his last years he placed increasing emphasis on environmental aspects of the atmosphere and oceans including their chemistry. In a speech at the annual meeting of the American Meteorological Society just before he died, he outlined his vision of the future of meteorology. With computers capable of assuming much of the routine work of the forecaster, the meteorologist should be freed to direct his efforts toward more constructive applications of meteorological knowledge. This would permit meteorology to expand and broaden so that the profession would attract and utilize many highly qualified people from mathematics, physics, geophysics, chemistry, etc.

See also: Atmospheric Motions, Volume 1.

REFERENCES

Palmer, T (1998) Nonlinear Dynamics and Climate Change: Rossby s Legacy, *Bull. Am. Met. Soc.*, 79, 1411–1423.

Phillips, N A (1998) Carl Gustav Rossby: His Times, Personality, and Action, *Bull. Am. Met. Soc.*, 79, 1097–1112.

Rossby, C G (1959) Current Problems in Meteorology, in *The Atmosphere and the Sea, in Motion, Scienti c Contributions to the Rossby Memorial Volume*, ed B Bolin, The Rockefeller Institute Press, New York, 9–50.

ANDERS PERSSON *UK*

Rossby Waves

The atmospheric circulation in the extratropical troposphere is directed mainly from west to east. However, there are significant departures from this zonal flow, in the form of wave-like disturbances giving the wind northward or southward components at different longitudes. These disturbances have wavelengths of thousands of kilometers in the longitudinal direction and, relative to the Earth s surface, can be either stationary or traveling. Such disturbances affect the tracks of the wandering highs and lows that account for much of the changing surface weather patterns.

The physical basis for these wave-like features was first proposed by Rossby (1939) (*see* Rossby, Carl-Gustaf, Volume 1); as a result they are now universally referred to as *Rossby waves*. The underlying basis derives from

Kelvin s circulation theorem, related to the principle of conservation of angular momentum. Because the Earth is spherical, its local vertical component of angular velocity, with respect to the fixed stars, is a maximum at the pole, and zero on the equator. Because of this, fluid parcels displaced north or south from some reference line of latitude can lose or gain vorticity (a measure of the local angular velocity of the air) as determined in a frame co-moving with the Earth. Rossby was able to show that these induced changes in fluid vorticity would maintain a wave-like disturbance in velocity and pressure, with a westward phase speed relative to the eastward-directed flow.

The stationary Rossby waves mentioned above are forced primarily by near-surface flow over the major orographic features of the Earth s surface, and by thermal forcing from ocean–continent surface temperature contrasts. The traveling Rossby waves, on the other hand, are free-wave solutions, not tied to any specific surface forcing. Consistent with theory, lines of constant phase are always observed to retrogress to the west relative to the eastward-directed (basic-state) flow.

Modifications of Rossby s theory have been proposed. In one of the most important of these, Charney and DeVore (1979) have shown that nonlinear interactions between the east-west flow, and orographically-forced stationary Rossby waves can give rise to multiple equilibrium solutions. In one such solution, the Rossby waves have weak amplitude and the eastward flow is strong. In another solution, the Rossby waves have strong amplitude and the eastward flow is weak. These two equilibrium states correspond to two dominant types of circulation regime that occur in the extratropical atmosphere, the so-called zonal and blocked regimes, respectively. In some sense, these regimes can be considered important building blocks of the climate system, and may be important elements for understanding the impact of anthropogenic forcing on climate (Palmer, 1999).

Rossby-wave dynamics are also relevant in the tropics and in the oceans. For example, the El Niño phenomenon cannot be understood without considering the effect of westward propagating Rossby waves in the near-equatorial Pacific Ocean. In particular, according to the so-called delayed oscillator theory (e.g., Battisti and Hirst, 1989), the time it takes for equatorial oceanic Rossby and Kelvin waves to propagate across the Pacific basin, affects the lifetime of an El Niño event.

See also: Atmospheric Motions, Volume 1; Rossby, Carl-Gustaf, Volume 1.

REFERENCES

Battisti, D S and Hirst, A C (1989) Interannual Variability in a Tropical Atmosphere–Ocean Model. Influence of the Basic State, Ocean Geometry and Non-linearity, *J. Atmos. Sci.*, 46, 1687–1712.

Charney, J G and DeVore, J G (1979) Multiple Flow Equilibria in the Atmosphere and Blocking, *J. Atmos. Sci.*, 36, 1205–1216.

Palmer, T N (1999) A Non-linear Dynamical Perspective on Climate Prediction, *J. Clim.*, 12, 575–591.

Rossby, C-G (1939) Relation Between Variations in the Intensity of the Zonal Circulation of the Atmosphere and the Displacements of the Semi-permanent Centers of Action, *Tellus*, 2, 275–301.

TIM PALMER *UK*

Rowland, F Sherwood

(1927-)

F Sherwood Rowland founded the Department of Chemistry of the University of California, Irvine where he serves as Bren Professor of Chemistry and of Earth System Science. In 1995, he was Nobel Laureate (with Molina (*see* Molina, Mario J, Volume 1) and Crutzen (*see* Crutzen, Paul J, Volume 1)) and has been Foreign Secretary of the US National Academy of Sciences (NAS) since 1998.

After completing his PhD in physical chemistry at the University of Chicago in 1952, Rowland s early career focused on radiochemistry and chemical kinetics. He performed and interpreted laboratory experiments designed to study reaction rates and mechanisms. He continued this kind of research well into the 1980s and established himself as a world leader through careful, accurate and well-selected experiments.

In atmospheric chemistry, Rowland has been a pioneer. He and his postdoctoral fellow, Mario Molina, discovered the role of chlorofluoromethanes and other chlorofluorocarbons (CFCs) as chlorine-atom donors in the stratosphere. They also showed in 1974 that chlorine atoms could catalyze the destruction of large amounts of ozone. This work was highly original and also exceptionally complete and careful. In a 1975 paper, they investigated many possible mechanisms that might destroy CFCs (besides ultraviolet light in the upper atmosphere), investigations that were repeated a number of times later by others, always confirming the original results (*see* Depletion of Stratospheric Ozone, Volume 1).

An early 1980s discovery (with his former student Don Blake) that atmospheric methane amounts were increasing

worldwide confirmed (less conclusive) reports by two other groups. The clear and careful experimental methods of Blake and Rowland established this fact, and many scientists worldwide are working to continue to define the upward trend and to explain its causes (propositions include emissions from rice paddies, landfills, coal mines, natural gas usage, biomass burning associated with the clearing of tropical forests, and also diminished intensity of atmospheric photochemistry that consumes methane). Atmospheric methane s alteration of the radiative energy balance of the Earth s surface is second only to carbon dioxide (CO_2), and the need to consider methane s changes has greatly complicated the study of climate change, but has also sharpened the science.

With his student, Neil Harris, he examined several long time series of ozone data from around the world and noted that decreasing amounts were being observed in middle to high latitudes in late winter and spring seasons, suggesting a chemical loss process in wintertime. Neither the empirical fact from the data or the thought of such a loss had been seen or conceived previously. Combined with a suggestion from Rowland (in Feldafing, Germany in 1984) that hydrochloric acid (HCl) and $ClNO_3$ (chlorine nitrate) might be shorter lived in the lower stratosphere than gas-phase processes alone would dictate, this work seeded the idea that heterogeneous chemical processes might cause the Antarctic ozone hole by liberating ozone-active chlorine.

The impact of Rowland s research on the fields of atmospheric chemistry and related laboratory physical chemistry is enormous. His research altered the worldwide CFC industry and led to a greatly increased awareness of environmental chemistry and human impact on the global environment by industry, government and the general public, and enabled the Montreal Protocol, a 1987 international treaty (updated in Copenhagen and London in 1990 and 1992) that launched international phase-outs of fully halogenated CFCs and related compounds.

As a communicator of science and a scientist-citizen in matters of public policy, Rowland has been powerful. He has spoken clearly and forcefully, always true to the science and its limits.

For his research and its impact on environmental preservation, Rowland has received many awards, including the Tyler Prize for the Environment, the Dana Award, and the Japan Prize, membership in the NAS and many honorary degrees. The Dana Medal recognizes that his research and his public speaking increased worldwide awareness of skin cancer and the role of excessive exposure to solar ultraviolet light. An accomplished tennis player, Rowland also played intercollegiate basketball and baseball while at Chicago, and semi-professionally in baseball. In 2000, he received the Foundation of Economic Teaching (FTE) Academic All American Hall of Fame Award for combining scientific excellence with high achievement in intercollegiate athletics.

RALPH J CICERONE *USA*

Runoff

When precipitation or meltwater from snow and ice reaches the ground surface, it is partitioned between water which moves across the surface, the *overland ow*, and that entering the soil, the *in ltration*. Initially most water may enter the soil, but the proportion will decrease with time, as infiltration capacity is approached and exceeded. Then water starts to pond on the surface and when this ponded water spills, it becomes overland flow, or quick flow, which collects in rills, gulleys and small channels. These channels may also receive water emerging from the ground; altogether water from these different sources flows into the larger channels, streams and rivers. These watercourses contain the runoff from the basin. This is the traditional Hortonian view of runoff (Horton, 1933) with its division between surface flow and subsurface flow. The subsurface flow is usually separated between the *inter ow* (that part of the infiltration, which has traveled laterally through the upper soil horizons and the base flow) and the base flow, or *groundwater ow* (that part of the runoff contributed by drainage from the aquifers in the basin). This is the flow that sustains the runoff in dry weather. Flow velocities are considered to decrease from overland flow, to interflow, to base flow, giving rise to the possibility of differences in the age of the water that may be in one increment of runoff.

There are a number of factors that govern the runoff process such as: intensity and duration of the precipitation, soil type, geology, land use, the characteristics of the surface and the stream channel network. That all parts of a basin do not react to precipitation in the same manner results from the mix of these factors and has led to the concept of the contributing area (Betson, 1964) and the further developments which have been incorporated into contemporary runoff models.

REFERENCES

Betson, R P (1964) What is Watershed Runoff? *J. Geophys. Res.*, 69, 1452–1541.

Horton, R E (1933) The Role of Infiltration in the Hydrologic Cycle, *EOS Trans. Am. Geophys. Union*, 14, 446–460.

JOHN RODDA *UK*

S

Salinity

Salinity is a general concept related to the amount of dissolved material in sea water, measured historically as grams of salt per kilogram of sea water. Seawater density depends on temperature, pressure and the amount of dissolved matter, which primarily consists of various salts. Because the salts come from continental weathering and deep-sea vents, the total amount of salt in the whole of the world s oceans varies on a time scale (order 100 000 years) that is much longer than the 1000-year time scale of ocean circulation. Therefore, the relative proportion of the constituents of sea salt is nearly constant throughout the oceans. However, the total amount of water in the ocean is not constant; it is affected by evaporation, precipitation, and runoff. The total amount of dissolved matter per kilogram of seawater therefore varies.

Historically, salinity was measured by evaporating the water and weighing the residual matter. This method was replaced by a titration method to determine the amounts of chlorine, bromine and iodine, which were then related to salinity through an empirical formula. Since the International Geophysical Year (1957), the standard method for measuring salinity uses seawater conductivity, which is related through an empirical formula to salinity. A seawater standard is required to provide high accuracy. The conductivity-based method is far more precise (0.00015) and accurate (0.001) than the previous methods. In 1978, the United Nations Educational, Scientific and Cultural Organisation (UNESCO) Joint Panel on Oceanographic Tables and Standards produced a standard algorithm for salinity based on conductivity, this was the practical salinity scale 78 (PSS-78), reported in a 1981 publication. Until that time, salinities were commonly reported in units of (parts per thousand). The UNESCO panel recommended that, henceforth, no units should be used for salinity, which is the convention followed here. Within the oceanographic community, the unit practical salinity scale (psu) came into common usage with PSS-78; psu and remain in common use. The range of salinity in the open ocean, away from continental margins, is about 31–40, with the mid-latitude oceans in the range 34–36. Lowest values occur at high latitudes where ice melt lowers the salinity, and highest values are found in strongly-evaporating regions such as the Red and Mediterranean Seas.

See also: Ocean Circulation, Volume 1; Salinity Patterns in the Ocean, Volume 1.

LYNNE D TALLEY *USA*

Salinity Patterns in the Ocean

Lynne D Talley
Scripps Institution of Oceanography, University of California, La Jolla, CA, USA

Ocean salinity varies geographically and with time. Fresh water input occurs at the sea surface due to precipitation and river in ow, reducing salinity. Salinity is increased by evaporation and also as a by-product of sea ice formation. Because freshening and salini cation occur in different places, salinity at a particular location re ects the upstream source of the water there. In subtropical latitudes, high surface evaporation creates high salinity near the sea surface. In subpolar latitudes, high precipitation creates low salinity near the sea surface. As these waters ow into the ocean interior, they create layers of high and low salinity.

At mid-depth (i.e., around 1000 to 2000 m deep), out ows from the highly evaporative Mediterranean and Red Seas create a vertical salinity maximum in the North Atlantic and Indian Oceans, respectively. Also at mid-depth in the subtropical and tropical regions, the relatively fresh, but dense, surface water from higher latitudes ows in and creates a vertical salinity minimum, most prevalent in the Southern Hemisphere and North Paci c. The North Atlantic is the most saline ocean and the North Paci c the freshest.

Salinity affects sea water density and therefore can be a controlling factor for the depth of the ocean surface mixed

layer. This depth also depends on wind speed and cooling at the sea surface. Feedbacks can occur as precipitation and evaporation are affected by the atmosphere, which in turn is affected by sea surface temperature. The atmosphere in turn is affected by sea surface temperature, which is affected by mixed layer depth.

The distribution of dissolved salts (*see* Salinity, Volume 1) in the oceans and adjacent seas varies in space and time. Salinity is changed near the sea surface by precipitation and evaporation of fresh water and by salty water produced sea ice forms and excludes salt. Geographical variations in inputs create regional differences in salinity at the sea surface. As seawater circulates down into the ocean away from the sea surface, it carries these differences along, creating large-scale salinity patterns throughout the ocean. Changes in time in the inputs at the sea surface also affect the salinity distribution. By mapping the salinity distribution and tracking changes in salinity over time, much insight is gained into both the basic ocean circulation and the processes that change the ocean circulation and temperature distribution.

Salinity is a seawater property related to the amount of matter, mainly consisting of salts, dissolved in the water (*see* Salinity, Volume 1). The original definition of salinity was in terms of grams of dissolved salts per kilogram of seawater. Salinity is now defined in terms of seawater conductivity, and is currently calibrated using a standard solution prepared at a laboratory in the UK. The quantity of dissolved salts, hence the salinity, affects seawater density, which also depends on temperature and pressure. Quantifying these dissolved constituents as the single parameter, salinity, is possible because the source of solids is weathering, which occurs on geological time scales. Ocean circulation and mixing, on the other hand, distribute the solids throughout the World Ocean on a time scale of no more than thousands of years, and so the various salt constituents are essentially well mixed. Therefore, salinity differences are created only by dilution or concentration as fresh water is added or removed, or as salty water is rejected from sea ice as it freezes. Other dissolved non-salt constituents such as carbon, calcium and silica, which are affected by biological cycles, have enough spatial variability to affect seawater density (e.g., Millero *et al.*, 1978); this small variation is not usually included in salinity and density studies.

In many ocean regions, particularly in the tropics, the subtropics and the deep ocean, temperature variations are more important than salinity in changing density. However, at higher latitudes, the presence of freshened surface waters resulting from high precipitation and/or ice melt has a strong influence on whether the water column can be cooled enough to mix vertically with other layers. Because winter surface cooling at high-latitudes determines water properties for the global ocean at depths below about 1500 m, salinity can control the occurrence of convection and its penetration depth. With a surface layer that is relatively fresh compared with the underlying water that came from another ocean region, surface waters can cool to freezing without overturning. If the high-latitude surface waters are not much freshened, as is the usual case in the northern North Atlantic, then overturn to great depth at the end of winter can occur with even modest surface cooling, leaving the surface waters well above the freezing point. Because ocean surface temperatures are part of the forcing for the atmosphere, changes in high-latitude salinity distribution can be involved in climate change.

In addition to its effect on seawater density, salinity is useful in itself as a tracer of ocean circulation and mixing. Salinity is established in the ocean s surface layer by local processes, and therefore salinity variations within the ocean depend on the surface origin of a parcel of water. Mixing between waters of different salinities occurs within the ocean, and so the original salinity changes gradually. However, mixing rates are low enough that salinity (and other chemical tracers) can provide a great deal of information on the pathways of flow.

SEA SURFACE FORCING OF SALINITY: PRECIPITATION, EVAPORATION, RUNOFF AND ICE PROCESSES

The total amount of salt in the ocean is constant on time scales of millions of years, and hence is assumed constant for studies at almost all time scales. However, the total amount of water in the oceans varies daily due to exchange at the sea surface. Rainfall, runoff from land, and ice melt all reduce salinity by increasing the amount of fresh water. Evaporation increases salinity by decreasing the amount of fresh water.

Ice formation also locally increases salinity through a process called brine rejection. As ice forms from seawater, the dissolved salts are rejected from most of the ice and collect in pockets in the ice. This briny water drips through the base of the forming ice. More brine is released during slow ice formation than if the ice forms quickly, so the strength of brine rejection depends on the local air temperature. The brine mixes with the seawater beneath the ice, raising the seawater s salinity. The net salinity increase depends on the thickness of the mixing layer below the ice. In shallow locations, such as on shallow continental shelves, salinity might increase more than over a deep bottom where mixing occurs to greater depth. Because the water below sea ice is usually very cold, the addition of salty brine may increase salinity without increasing temperature. This results in a density increase in the seawater. This denser water then sinks until it reaches a depth where surrounding waters have the same density. Because these deeper waters may have come, at least in part, from somewhere else (e.g., in the Arctic from

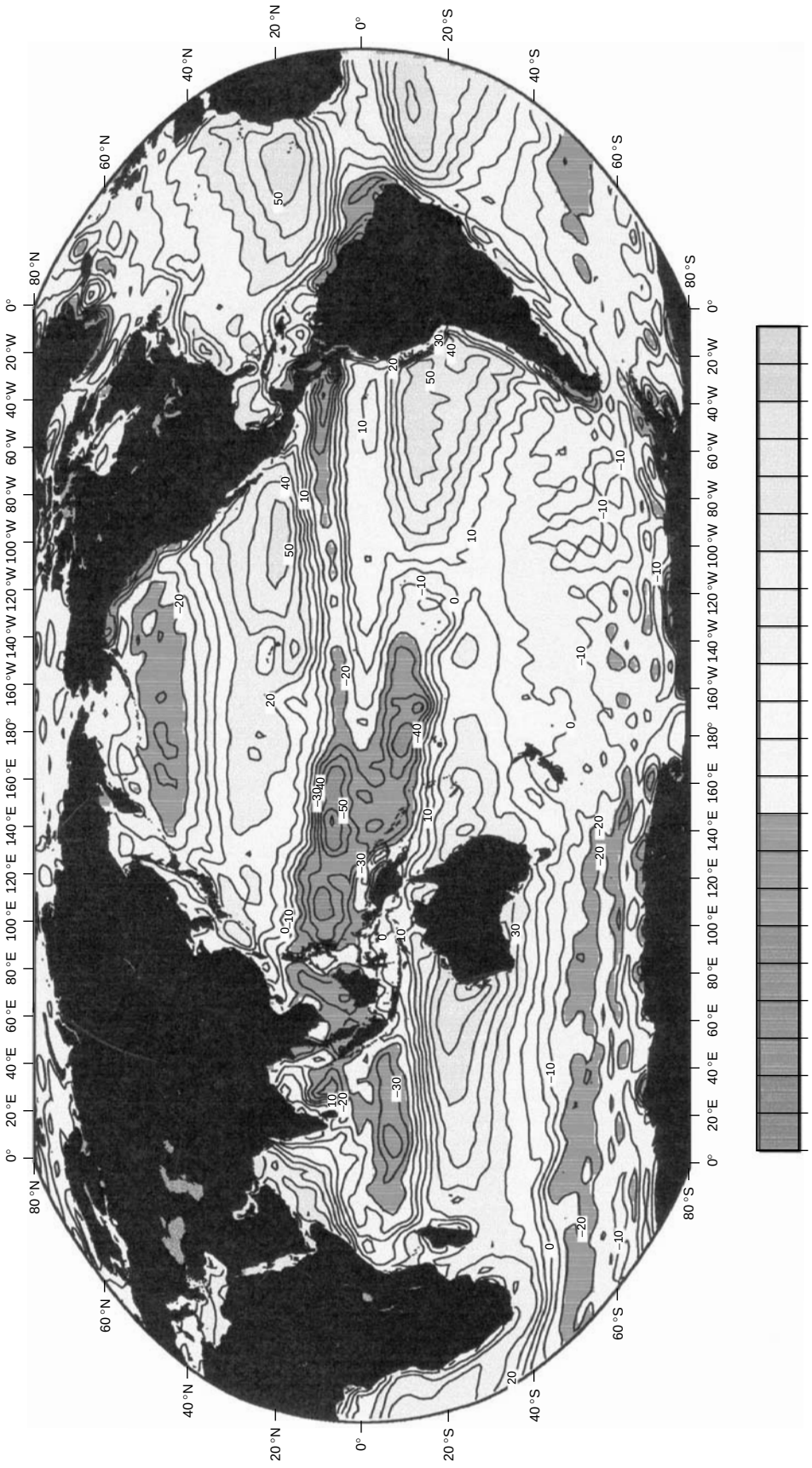


Figure 1 Evaporation minus precipitation (cm year^{-1}), annual average. The major rivers of the world are also indicated. (Reproduced using data from the National Centers for Environmental Prediction Reanalysis Project, Kalnay *et al.*, 1996)

the North Atlantic Ocean), these waters will usually not be as cold as the new ice-rejected brine. Although warmer, their density will be equal to the ice-rejected brine because they are even saltier. That is because the ice-rejected brine was formed from near-freezing surface waters that initially had much lower salinity (e.g., because they had a higher proportion of meltwater and river runoff), and therefore lower density, than the deeper waters. Although it seems paradoxical, the ice formation and brine rejection processes are thus able to create the coldest and freshest waters on a surface of constant density within the ocean.

Because sea ice has lower salinity than the original seawater from which it formed, melting of sea ice creates a fresh surface layer of low-density waters that float over the more saline waters below.

Ocean salinity patterns at the largest scales are affected primarily by the annual average of evaporation and precipitation, with additional major effects from ice melt and freezing at high-latitudes. Evaporation and precipitation, usually reported in cm year^{-1}, are primarily calculated from observations collected on a routine basis by most ships. The difference between evaporation (E) and precipitation (P) has been calculated from the complete historical data set compiled from these ship observations (Figure 1). The range of E$-$P is about -10 to 10 cm year^{-1}. There is more evaporation than precipitation in the subtropical regions between about 15° and 40 °S and N latitude, under the atmosphere s large-scale high pressure centers, with descending, dry air. At higher latitudes, there is more precipitation than evaporation, under the atmospheric low-pressure centers. In the tropics, especially at about 10 °N over the Pacific and Atlantic, surface air rises in the intertropical convergence zone (ITCZ), forming enormous cumulus clouds that produce large amounts of rain.

Runoff from land affects the coastal salinity distribution. On a global scale, runoff is most important at the major river outlets, indicated in Figure 1. The net fresh water input is not shown in Figure 1, but the effect of these major rivers is apparent in the surface salinity distribution considered next.

SURFACE SALINITY DISTRIBUTION

Salinity at the sea surface (Figure 2) resembles the evaporation$-$precipitation pattern, with some differences due to runoff, ocean currents and ice melt. The total salinity range of the open ocean is about 30$-$40 parts per thousand or (*see* Salinity, Volume 1). The international convention is that salinity has no units; its value is approximately grams salt per kilogram of sea water. Many authors continue to use the practical salinity unit (psu) rather than leaving the quantity unitless. While salinity in coastal estuaries can be much lower, it does not affect large oceanic-scale patterns. Lowest salinities occur in the Arctic and Antarctic where there is both net precipitation and seasonal ice melt. Highest salinities occur in the Red Sea and Persian Gulf, both located in the northwestern Indian Ocean, where net evaporation is high. High salinity is also found in the Mediterranean Sea. In the open ocean, high salinity occurs in the subtropical areas of net evaporation seen in Figure 1. A band of low salinity underlies the ITCZ at 10 °N.

The effect of continental runoff is apparent in lowered surface salinity near the mouths of major rivers such as the Amazon and the Congo and the numerous large rivers that empty into the Bay of Bengal, east of India, including the Ganges and Brahmaputra. Runoff from many rivers around the Gulf of Alaska in the northeastern Pacific and around the Arctic Ocean is important in the lowered salinities of high-latitude ocean regions.

VERTICAL STRUCTURE OF THE SALINITY DISTRIBUTION

Waters at the sea surface flow down into the interior. Salinities from the sea surface (Figure 2) carried with this flow into the ocean interior are helpful in identifying sources and pathways of the circulation. The two primary mechanisms for this downward flow are: (1) flow following constant density surfaces that slope gently into the ocean interior, and (2) vertical convection and brine rejection at higher latitudes.

Flow down along constant density surfaces is primarily through a process called *subduction*, occurring in subtropical regions between about 15° and 40 °S and N latitude. In this process, waters in the surface mixed layer circulate equatorward, encountering warmer, less dense waters to the south, which they are carried beneath. Surface waters are most saline in the middle of these subtropical regions. Therefore, all equatorward subtropical regions are characterized by a subsurface vertical maximum in salinity caused by subduction of more saline waters from the center of the gyre, as described later (*see* Ocean Circulation, Volume 1).

Convection occurs in many places, but reaches most deeply into the ocean at high latitudes. Convection is not so much a downdraft of water from the surface as it is a deep mixing, which brings surface water to great depth, from whence the water spreads laterally along constant density surfaces. Major deep convection areas can be identified from subsurface properties, such as the salinity distributions discussed below. In the North Atlantic and Arctic, these sites are in the Labrador Sea between Labrador and Greenland, in the Greenland Sea east of Greenland and north of Iceland, and in the Mediterranean Sea. In the North Pacific, convection in the Okhotsk and Japan Seas is significant in modifying water properties at mid-depth. In the Southern Hemisphere, convection occurs all around Antarctica at about 40° to 45 °S, which is just north of the major current

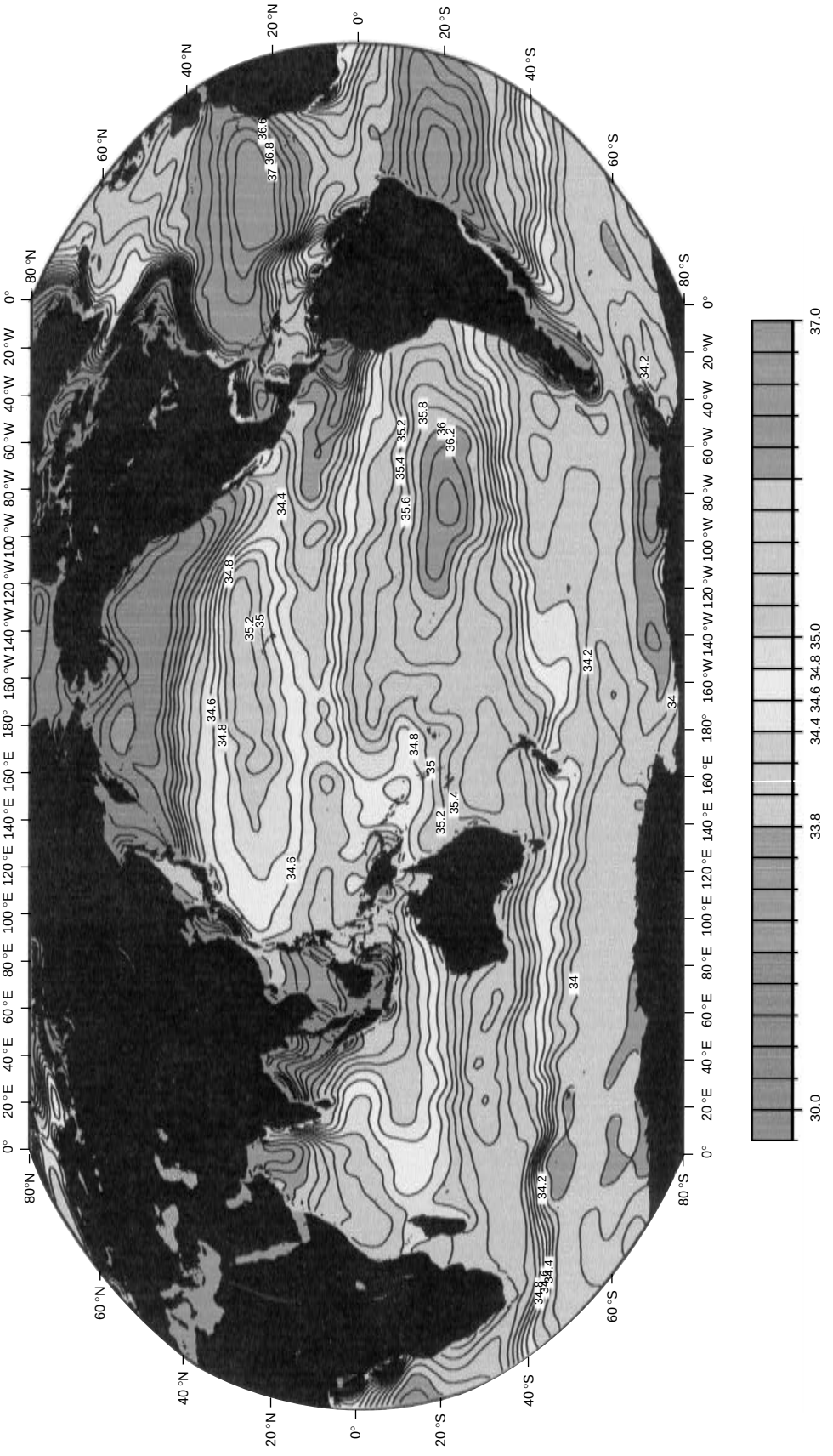


Figure 2 Surface salinity (g salt per kg of seawater), using annually averaged ship observations. (Reproduced using data from Levitus *et al.*, 1994)

called the Antarctic circumpolar current. Brine rejection is important for subsurface water properties primarily in the Antarctic (Weddell and Ross Seas and other locations around the continent), the Okhotsk Sea, and the interior of the Arctic Ocean.

The following definitions are useful for discussing the salinity distribution in the vertical: (1) *halocline* – where the salinity value changes rapidly in the vertical in comparison with changes above and below the halocline; (2) *halostad* – where salinity changes very slowly in the vertical; (3) *intermediate water* – usually referring to a minimum or maximum in salinity in the vertical which is found over a large horizontal area at mid-depth in the ocean. Equivalent terms for vertical variations in temperature and density are *thermocline*, *thermostat*, *pycnocline* and *pycnostad*.

SUBTROPICAL SALINITY STRUCTURE

In the subtropical regions of every ocean, salinity (Figures 3 and 4a) is high close to the sea surface due to strong evaporation. At the location of the highest surface salinity (Figure 2), the highest salinity in vertical profiles is at the sea surface. This high salinity water is carried downward and equatorward through subduction. Therefore, equatorward of the center of the highest surface salinity, there is a vertical maximum in salinity. This salinity maximum is at about 50 m depth, which is very shallow compared with the ocean depth. These subsurface salinity maxima are sometimes called *Subtropical Underwater* (Worthington, 1976).

At all locations in the subtropical gyres, and below the near-surface salinity maximum, salinity decreases smoothly with depth to a salinity minimum. Temperature also decreases smoothly and rapidly (Figure 4b). This layer of changing properties is called the *thermocline* (region of high temperature change in the vertical), and is also often called *central water*. Central water is the subducted water of the subtropical gyre; the colder waters come from farther north within the subtropical circulation, with little input from the subpolar region. It is clear from Figure 4(b) that the North Atlantic central waters are the most saline and the North Pacific the freshest.

Below the central waters, at depths between about 500 and 1500 m, each subtropical gyre (except in the North Atlantic) has a vertical minimum in salinity. The salinity minimum in the North Pacific is called *North Paci c intermediate water* (Reid, 1965). It arises from cold, low salinity surface waters in the subpolar region of the North Pacific, specifically in the Okhotsk Sea and near the Kuril Islands (Talley, 1991, 1993). These waters flow southward along the western boundary in the Oyashio current and enter the subtropical circulation east of Japan. The existence of the low salinity layer throughout the subtropical circulation reflects the presence of this subpolar water underlying the waters from subtropical sources.

The salinity minimum in the Southern Hemisphere is called *Antarctic intermediate water* (Reid, 1965). It comes from low salinity surface waters found just west of Chile and just east of Argentina (McCartney, 1977). Antarctic intermediate water in the Pacific Ocean originates primarily from the Chilean source. The Atlantic and Indian Ocean Antarctic intermediate waters originate from the Argentinean source. The Argentinean waters arise from Chilean surface waters, which flow eastward through the northern side of Drake Passage between South America and Antarctica. The Chilean waters are modified somewhat during their passage into the Atlantic. Therefore, the properties of the Atlantic and Indian Antarctic intermediate waters differ slightly from those of the Pacific Antarctic intermediate water.

The salinity structure in the subtropical North Atlantic is somewhat different. While there is a salinity minimum at the base of the central water, this results primarily from the northernmost subducted waters of the subtropical region, and is not the major fresh intermediate water of the North Atlantic, which dominates the subpolar region (described later). The main fresh intermediate water is displaced in the subtropical gyre by the high salinity outflow at mid-depth from the Mediterranean Sea. This outflow creates an intermediate depth salinity maximum in the vertical.

In the Indian Ocean, a dense, high salinity source in the Red Sea creates a high-salinity intermediate water in the north, similar to the Mediterranean water of the Atlantic. Because there are no northern high-latitude sources of water in the Indian Ocean, there is no fresh intermediate layer arising from the north. The fresh water at the surface in the Bay of Bengal arising from major river runoff is much too warm to sink to mid-depth. Thus, the mid-depth northern Indian Ocean is dominated by the highly saline outflow from the Red Sea.

Beneath the low salinity intermediate waters of the North Pacific and Southern Hemisphere, salinity increases towards the bottom. The higher salinity of the deep waters of the Southern Hemisphere and North Pacific results from input of saline North Atlantic waters. Beneath the saline intermediate water of the North Atlantic and tropical Indian Oceans, salinity decreases to the bottom. In the South Atlantic Ocean, low-salinity intermediate water (Antarctic intermediate water) lies above higher salinity deep waters from the North Atlantic. Beneath the high-salinity North Atlantic deep water, salinity decreases to the bottom. The lower salinity of the bottom waters in the Atlantic and Indian Oceans results from deepwater sources in the Antarctic that are of lower salinity.

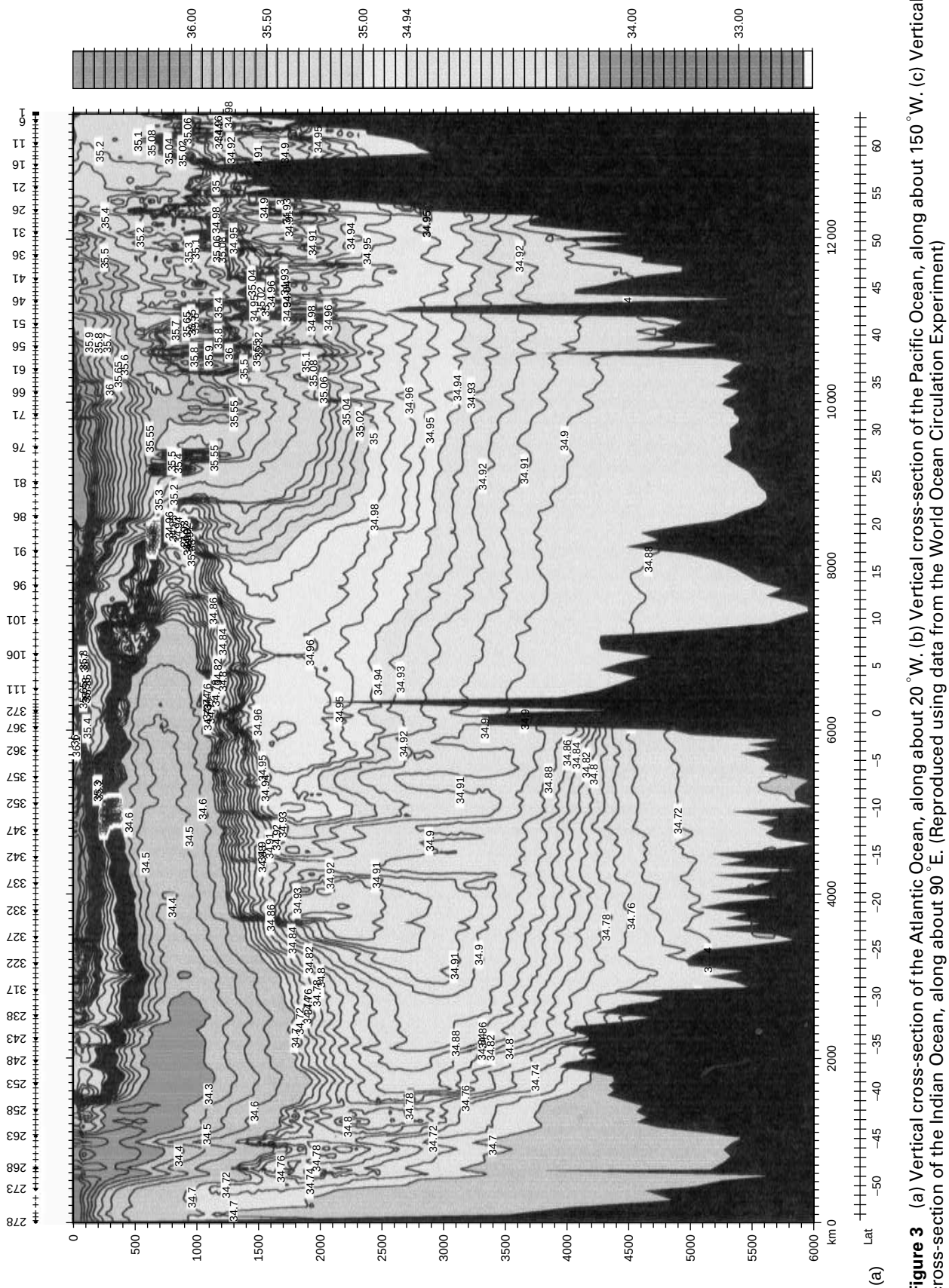


Figure 3 (a) Vertical cross-section of the Atlantic Ocean, along about 20 °W. (b) Vertical cross-section of the Pacific Ocean, along about 150 °W. (c) Vertical cross-section of the Indian Ocean, along about 90 °E. (Reproduced using data from the World Ocean Circulation Experiment)

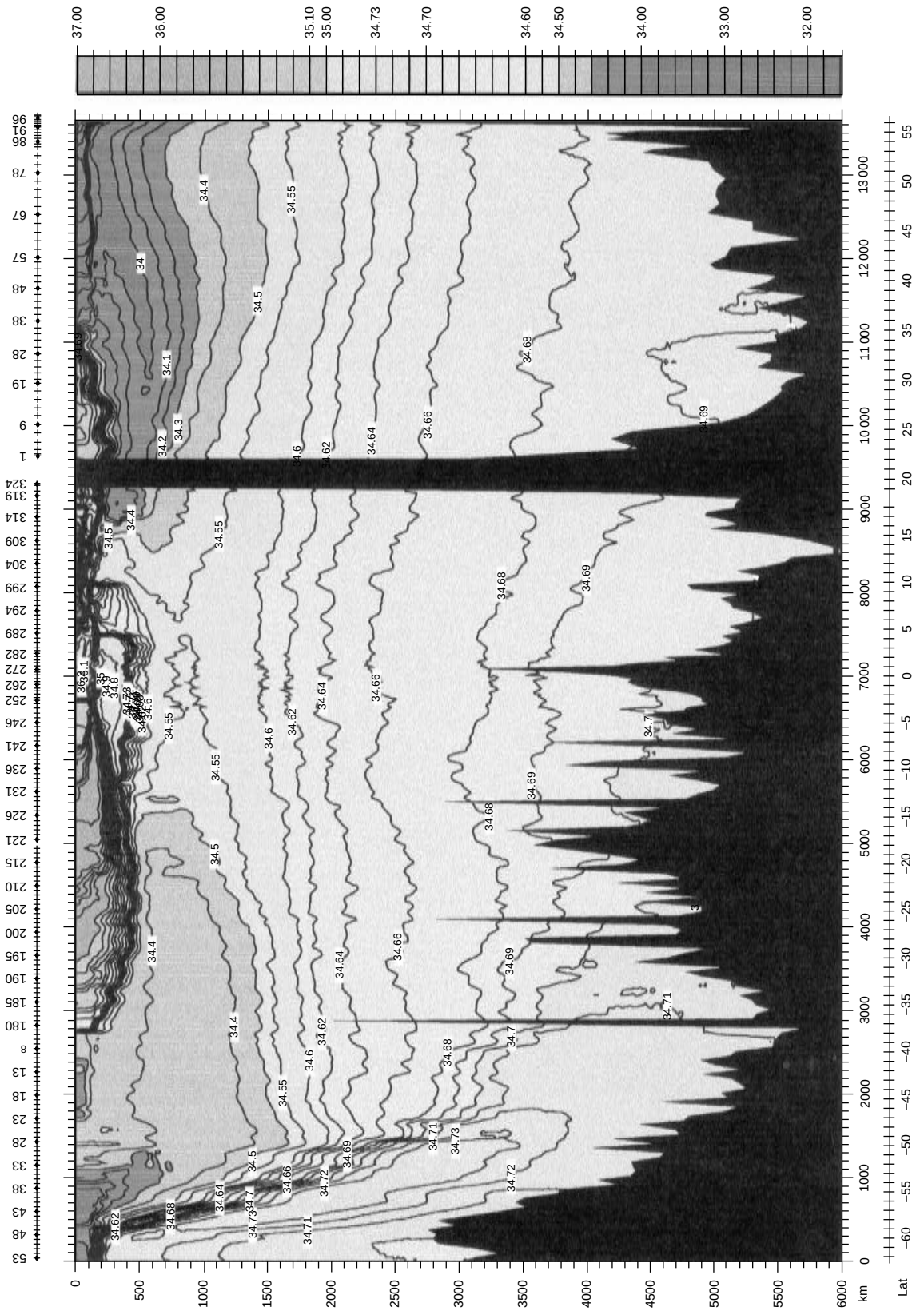


Figure 3 (*Continued*)

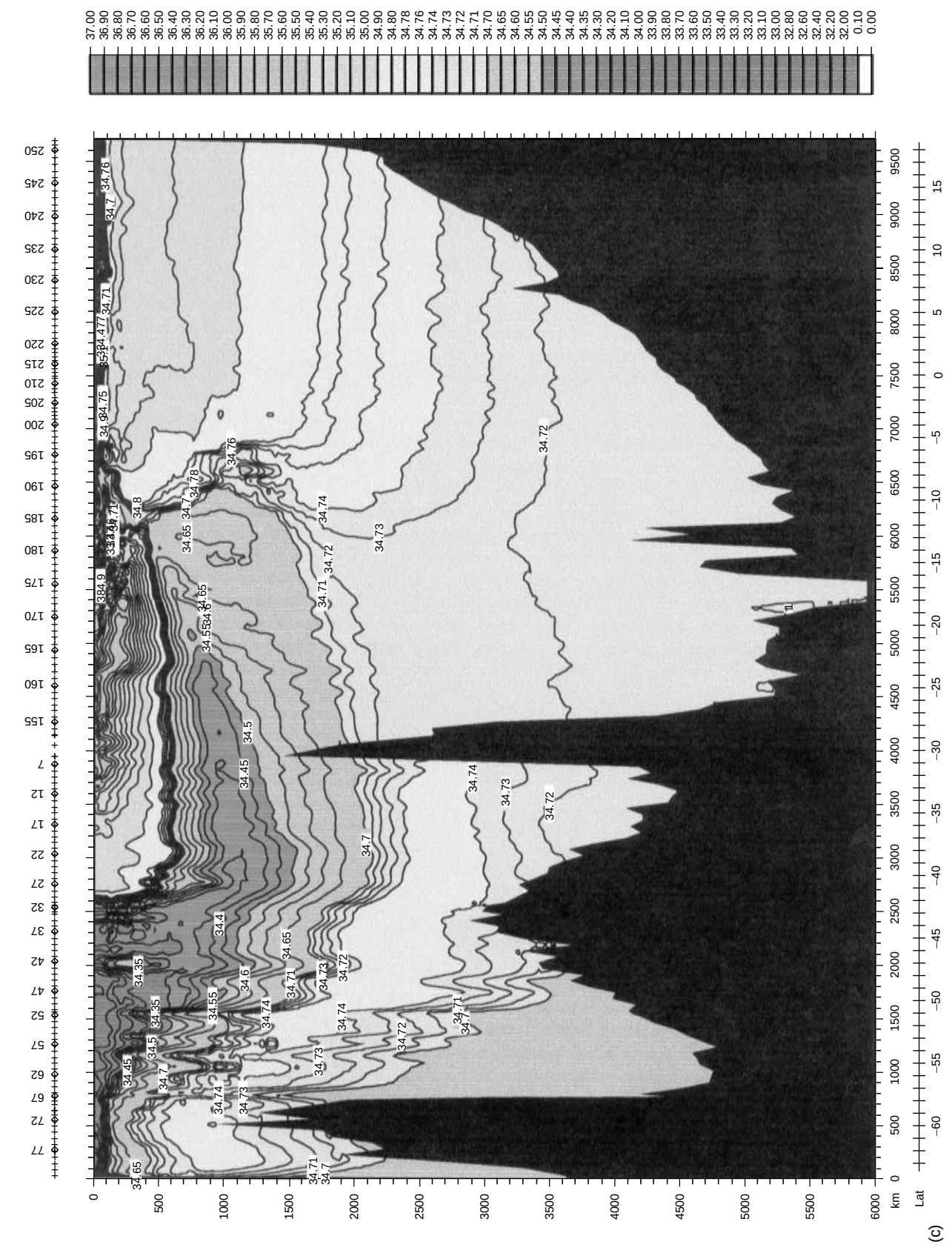
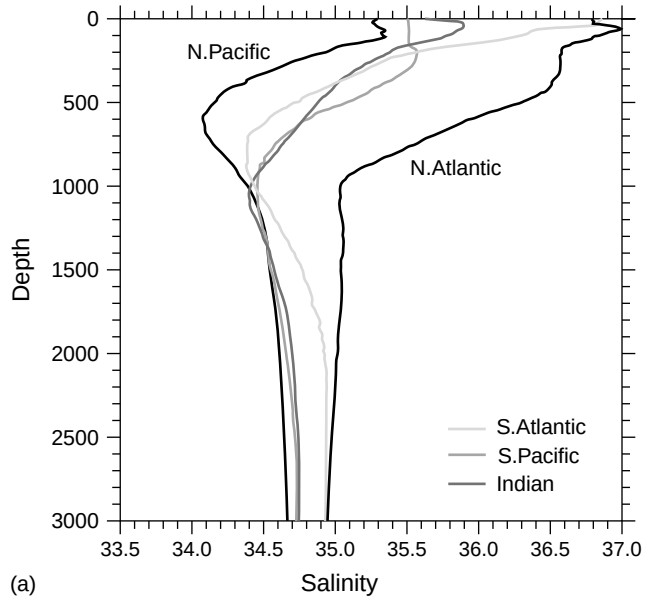


Figure 3 (*Continued*)



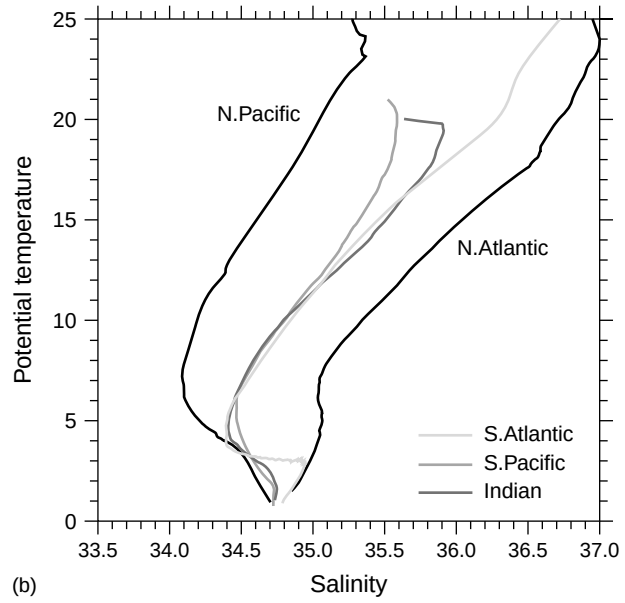


Figure 4 (a) Salinity distribution with depth from each subtropical gyre. All profiles are chosen from the centers of the subtropical circulations, at about 25 °N (Northern Hemisphere) and 30 °S (Southern Hemisphere), and from the center of the basins rather than near the boundaries. (b) Potential temperature as a function of salinity for each subtropical ocean basin

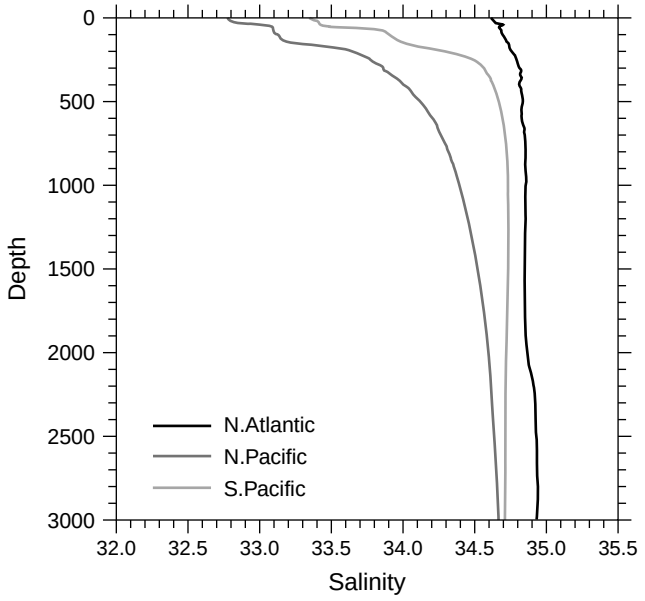


Figure 5 Salinity distribution with depth from the subpolar regions. The Northern Hemisphere profiles are from about 50 °N. The Southern Hemisphere profile is from the Pacific Ocean at about 65 °S

SUBPOLAR SALINITY STRUCTURE

In high latitude regions dominated by precipitation and ice melt, salinity is lowest at the sea surface and increases downward through a halocline (Figure 5, North and South Pacific profiles). In the North Pacific and Southern Hemisphere, and also in most of the Arctic Ocean, the halocline is very sharp, with a large change in salinity over several tens of meters. Because of the salinity structure, the surface layer can be colder than the water below the halocline. This pre-conditions the region for sea ice formation. The cooled surface waters can circulate away from their winter outcrop locations and appear as a subsurface temperature minimum between warmer water below from inflow from other regions and warmer water above from local seasonal warming.

The North Atlantic s subpolar gyre is affected by inflow of a thick layer of saline, warm surface water from the subtropical gyre. Therefore, the halocline typical of the subpolar Pacific and subpolar Southern Hemisphere is absent in most of the subpolar North Atlantic (Figure 5, North Atlantic profile). The saline flow into the central and eastern subpolar gyre is affected by precipitation, and the surface waters become gradually fresher as they circulate counterclockwise. When the surface waters enter the Labrador Sea, they are much fresher than when they first enter the subpolar region. Winter convection to about 1500 m depth in the Labrador Sea then mixes the relatively fresh water to intermediate depth. This spreads away from the Labrador Sea and forms a salinity minimum throughout the subpolar gyre, beneath the inflowing warm, saline waters. This differs from subpolar regions in the other oceans where the surface waters are strongly dominated by local fresh water sources. Therefore, most of the North Atlantic s subpolar gyre has a vertical salinity structure similar to a subtropical structure, with saline waters at the surface and an intermediate depth salinity minimum.

TROPICAL SALINITY STRUCTURE

The upper ocean salinity distribution depends on local precipitation patterns. Thus, a band of low salinity is found under the high precipitation of the ITCZ at 5–10 °N in the Pacific. This band is separated from the saline waters below it by a halocline. Excess precipitation is also found in the western Pacific at the equator, creating a halocline there as well. The eastern equatorial Pacific on the other hand is dominated by evaporation and so there is no halocline there. The Bay of Bengal in the Indian Ocean is the site of major runoff from Asia; this relatively fresh, warm water creates a shallow halocline throughout the northeast Indian Ocean. The intermediate and deepwater salinity distributions in each ocean within 15° of the equator are similar to those of the adjacent subtropical regions.

GLOBAL SCALE SALINITY PATTERNS

As indicated above, there are significant differences in salinity between oceans. Considering the total ocean basin, the North Atlantic is the most saline and the North Pacific the freshest of the major oceans. Because of significant river inflows, the surface waters of the Arctic are the freshest. The three Southern Hemisphere oceans are connected around Antarctica and hence their salinities are similar, midway between the North Atlantic and North Pacific. The main reason for the differences between the North Pacific and North Atlantic salinities is the difference in net evaporation per unit area. There is relatively more evaporation in the North Atlantic s subtropical regions than in the North Pacific s subtropics. The evaporative regions occur where there are easterly trade winds. The North Atlantic s trade winds originate over the deserts of Africa and the Middle-East. The North Pacific s trade winds originate over the much smaller continental areas of Central America and Mexico. Therefore, the North Atlantic s trade winds are drier than the North Pacific s and cause more evaporation per unit area. High salinity in the North Atlantic is present at intermediate depth because of the Mediterranean Sea. The North Atlantic exchanges water with the Mediterranean Sea through the Strait of Gibraltar. Evaporation and cooling in the Mediterranean cause the outflowing water to be much denser than the inflow. This dense outflow sinks to intermediate depth and creates the mid-depth high salinity that marks North Atlantic waters. This input of high salinity at mid-depth in the North Atlantic is one reason for the overall higher salinity of the North Atlantic compared with the North Pacific. Similarly, in the Indian Ocean, high evaporation in the Arabian region creates very saline and dense water in the Red Sea, which enters the Indian Ocean and raises the salinity at mid-depths compared with the inflowing fresher deep waters from the Antarctic.

ROLE OF SALINITY IN CLIMATE CHANGE

Salinity is a major factor in climate change because it can affect the depth of mixing at the sea surface. If for instance there is a shallow, fresh layer atop a denser, more saline layer, even large surface cooling can act only on the surface layer, creating a large temperature drop without deep convection. In the absence of a halocline, heat can be removed from a much thicker layer, which then results in a smaller temperature change. The processes that set the salinity of the waters beneath the surface layer, and hence the strength of the halocline, are complex. This salinity control of surface-layer processes in the present climate has been studied, e.g., in the tropical Pacific, as part of the El Niño and La Niña cycle. In the normal state in the northern Atlantic, up to and including the region between Greenland and Norway, surface salinities are not much lower than those of the underlying waters, permitting overturn to more than several hundred meters depth. This overturn is part of the global thermohaline circulation in which deep waters are produced from surface waters at high-latitudes, mainly through heat loss to the atmosphere. (The presence of a permanent and strong halocline in the northern North Pacific and in the Arctic Ocean prevents deep overturning of waters there.) During observed decadal climate changes in the North Atlantic, a halocline sometimes forms as a result of excessive ice melt and runoff and is associated with shutdown of deep wintertime convection. This then affects the sea surface temperature as described previously, and can then affect the atmosphere. From ocean computer modeling experiments it is hypothesized that similar controls by salinity on deep overturn have occurred on much longer climate timescales.

See also: Broecker, Wallace S, Volume 1; Ocean Circulation, Volume 1.

REFERENCES

Kalnay, E, Kanamitsu, M, Kistler, R, Collins, W, Deaven, D, Gandin, L, Iredell, M, Saha, S, White, G, Woollen, J, Zhu, Y, Chelliah, M, Ebisuzaki, W, Higgins, W, Janowiak, J, Mo, K C, Ropelewski, C, Wang, J, Leetmaa, A, Reynolds, R, Jenne, R, and Joseph, D (1996) The NCEP/NCAR 40-Year Reanalysis Project, *Bull. Am. Meterol. Soc.*, 77, 437–471.

Levitus, S, Burgett, R, and Boyer, T (1994) *World Ocean Atlas 1994*, Vol. 3, Salinity, NOAA Atlas NESDIS 3, 99, US Government Printing Office, Washington, DC.

McCartney, M S (1977) Subantarctic Mode Water, in *A Voyage of Discovery*, ed M Angel, Pergamon Press, New York, 103–120.

Millero, F, Forsht, J D, Means, D, Gieskes, J, and Kenyon, K (1978) The Density of North Pacific Ocean Waters, *J. Geophys. Res.*, 83, 2359–2364.

Reid, J L (1965) *Intermediate Waters of the Paci c Ocean*, Number 2, Johns Hopkins Oceanographic Studies, Baltimore, MD.

Talley, L D (1991) An Okhotsk Sea Water Anomaly: Implications for Subthermocline Ventilation in the North Pacific, *Deep-Sea Res.*, 38, S171–S190.
Talley, L D (1993) Distribution and Formation of North Pacific Intermediate Water, *J. Phys. Oceanogr.*, 23, 517–537.
Worthington, L V (1976) On the North Atlantic Circulation, Number 6, Johns Hopkins Oceanographic Studies, Baltimore, MD.

Satellite Systems

see Earth Observing Systems (*Opening essay*, **Volume 1**)

Satellites, Atmospheric Measurements

see Earth Observing Systems (*Opening essay*, **Volume 1**)

Satellites, Land Measurements

see Earth Observing Systems (*Opening essay*, **Volume 1**); Remote Sensing (**Volume 2**)

Satellites, Ocean Measurements

see Earth Observing Systems (*Opening essay*, **Volume 1**)

Scenario

see Prediction in the Earth Sciences (**Volume 1**); Business-as-usual Scenarios (**Volume 5**)

SCOR (Scientific Committee on Oceanic Research)

SCOR was created by the International Council for Science (ICSU) in 1957 as the first of its interdisciplinary bodies, and is charged with the promotion of international activities in oceanography. SCOR s scientific activities fall into two categories: small, short-lived working groups formed to address narrowly focused scientific topics, and longer-term, large-scale international research programs in oceanography designed to address issues of the role of the ocean in global climate change. SCOR also serves as an official scientific advisory body to the Intergovernmental Oceanographic Commission (IOC) of the United Nations Educational, Scientific and Cultural Organization (UNESCO). Lastly, SCOR administers a program of travel awards to marine scientists from developing countries. National committees for SCOR exist in many countries.

SCOR cooperates with ICSU and IOC in the development of projects such as the Global Ocean Observing System (*see* Global Ocean Observing System (GOOS), Volume 2) and the Joint Ocean Flux Study (*see* JGOFS (Joint Global Ocean Flux Study), Volume 1). The latter program has completed its worldwide sequence of field data-collection cruises, and is entering an intensive period of data assembly, analysis, modeling, interpretation and synthesis. Its objective is to develop a global picture of the oceanic carbon cycle.

Another major SCOR initiative is the Global Ocean Ecosystem Dynamics effort, which is one of the core projects of the IGBP.

For further information contact: Elizabeth Gross, Executive Director of SCOR, Department of Earth and Planetary Sciences, The Johns Hopkins University, Baltimore, MD 21218, USA. Tel.: +1-410-516-4070; Fax: +1-410-516-4019; Email: scor@jhu.edu.

JOHN S PERRY *USA*

Sea Ice

Douglas G Martinson
Columbia University, Palisades, NY, USA

Sea ice is ice that forms by the freezing of ocean seawater. It is distinctly different from icebergs (of Titanic fame), which are pieces of glacial ice formed by the compaction of snow that have broken off from continental ice sheets (on Greenland and Antarctica) or from alpine glaciers.

Sea ice is important to global environmental change because it resides at the interface between the ocean and the atmosphere. Global climate patterns in space, and variations in time, re ect complex interactions between the atmosphere, ocean and land surfaces (including biosphere).

When and where sea ice is present, the nature of the planet's surface is dramatically altered (from a dark liquid ocean surface to a white solid ice surface), as are the interactions between the atmosphere and ocean. These surface alterations are the primary means by which sea ice influences the environment, both in the immediate region where it is present, as well as on global scales.

Sea ice formation is dependent upon distinct atmospheric and oceanic conditions. The conditions are met predominantly in the polar seas (Figure 1), specifically in the Arctic Ocean and neighboring seas, and in the seas surrounding the Antarctic continent (Figure 2). However, despite what might appear to be some rather obvious similarities between the Arctic and Antarctic polar seas including, the presence of a sea ice cover included, there are some rather dramatic differences. Most notable is the fact that the land-locked Arctic Ocean and neighboring polar seas are blanketed by a sea ice cover that averages 3–4 m thick and covers nearly 15×10^6 km^2 (an area comparable in size to the US). About half of this ice cover is perennial (8×10^6 km^2), i.e., year round. Conversely, the Antarctic seas, unbounded to the north, are blanketed by a rather thin seasonal sea ice cover, typically less than 1 m thick and covering approximately 20×10^6 km^2. Like ice on a pond, most of the Antarctic sea ice melts away in the spring, is absent in summer (except for 4×10^6 km^2 of thicker perennial ice) and reforms the following fall and winter. The dissimilarities between Arctic and Antarctic sea ice reflect a wealth of differences in the manner by which the ocean, ice, and atmosphere interact in the two regions.

Regardless of differences, one might wonder why an expanse of relatively thin floating ice in the most remote reaches of the Earth would have any influence on climate outside of the polar regions themselves. From the broadest perspective, this influence is not unexpected when considering that climate is the consequence of the natural redistribution of excess heat in the tropics to the heat-starved polar regions. This meridional heat imbalance reflects the fact that, in the tropics, the Sun (the planet s primary source of surface heat) beats down almost directly overhead throughout the year, whereas in the polar regions, the Sun does not even rise above the horizon for much of the year. When it finally does, in the polar summer near the top of the spherical Earth, the Sun is never directly overhead. Consequently, the incoming sunlight strikes the surface at an oblique angle, reducing its heating effectiveness. The oblique angle also causes much of the solar radiation to be reflected back into space, further reducing its heating potential. The vast fields of highly reflective snow and sea ice (and glacial ice) greatly enhance this reflective loss.

This differential heating leads to a strong temperature contrast between the equator and poles. Heat always flows from warm to cold regions, and this flow, complicated by a rotating Earth and complex distribution of land and sea, drives all of the atmosphere and ocean circulation systems that form the basis of our climate. In the crudest sense, the stronger this temperature contrast, the more vigorous the atmospheric circulation and climate system. Changes in the temperature gradient from the equator to the pole may thus be expected to change the basic state or vigor of the climate system.

Such changes can be induced by changes in the tropics, the polar regions, or anywhere in between. Because of sea ice, the polar regions are particularly susceptible to change. This susceptibility stems from a variety of ocean–atmosphere feedbacks, unique to sea ice, and

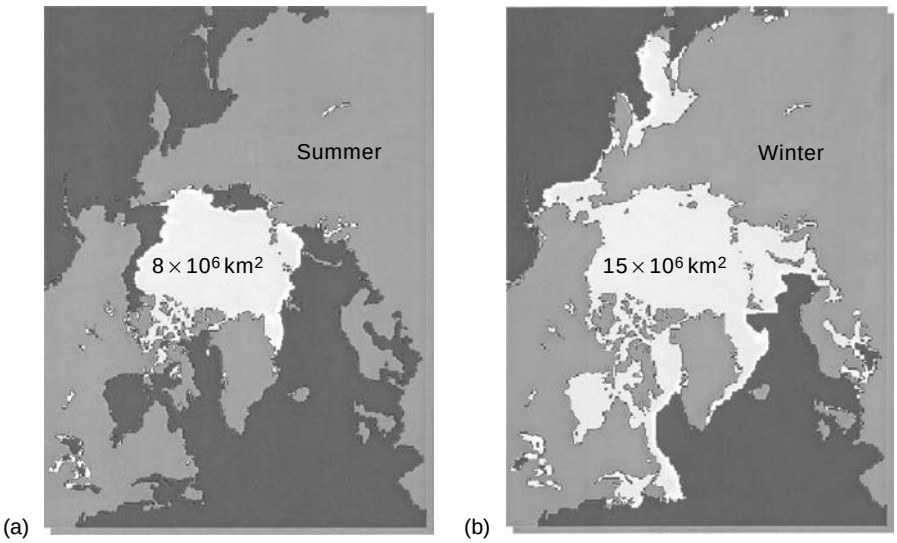


Figure 1 Summer (a) and winter (b) sea ice distribution in the Arctic polar region

Figure 2 Photograph of newly formed sea ice in the Antarctic. When ice crystals first float to the surface of the ocean, the wave motion causes them to coalesce into the characteristic shape seen here. This new ice, approximately 30 cm thick, is referred to as pancake ice (though it often looks more like lily pads, especially given the turned edges resulting from the collision of pancakes amongst themselves in the wave field). For scale, the fur seal resting on one of the larger pancakes is approximately 2 m in length. (Photo by Richard A. Iannuzzi)

capable of amplifying small sea ice changes into relatively large impacts. Of these, two dominate: (1) the ice–albedo feedback; and (2) the ice–ocean feedback. The former influences the effectiveness of solar heating, and is primarily active in summer when sunlight is present. In the polar regions, as previously noted, much of the cooling is due to the reflectance of the solar radiation from the sea ice or its snow cover. The degree of reflectance is indicated by the albedo (*see* Albedo, Volume 1). The ocean has an albedo of ~0.2 (it absorbs most of the incoming radiation), whereas fresh snow has an albedo of ~0.8 (it is a strong reflector). Thus, if the ocean is covered with ice overlain with snow, the effectiveness of surface radiation absorption (surface warming) is dramatically reduced (by 75%). This leads to a colder atmosphere (which draws heat from the warmth of the surface). Such conditions favor the formation of more ice, reflecting back more radiation, chilling the atmosphere further, causing more ice growth, etc. If the ice cover disappears, the opposite occurs, resulting in a rapid increase in local warming, increased ice loss, more warming, etc. This enhanced warming with the loss of a sea ice cover, or cooling with the addition of a sea ice cover describes the ice–albedo feedback mechanism (*see* Climate Feedbacks, Volume 1).

This self-amplifying cascade of effects, the ice–albedo feedback, is most active and effective in summer when the Sun is above the horizon, and in the Arctic because of the perennial sea ice. If the Arctic were to begin losing its perennial ice cover, exposing the ocean, with its significantly lower albedo, to summer sunlight, one might expect a self-amplifying warming to begin, driving more summer ice loss and further warming, all the while altering the equator–pole temperature gradient, and presumably climate. This feedback is inefficient in the Antarctic because most of the sea ice fields are absent in summer (and because the land-based Antarctic icecap is extremely cold and unlikely to significantly melt). However, unlike the Arctic Ocean, the Antarctic sea ice fields extend equatorward of the polar circle. Therefore, they are exposed to some (though not much) sunlight in winter, and even more in the fall and spring, thus allowing some ice-albedo influence at those times.

The ice–ocean feedback encompasses an ensemble of interactions between the ocean, ice and atmosphere. Of these, the primary one involves the insulation effect of sea ice. This is a positive feedback mechanism that complements the ice–albedo feedback by operating primarily in winter. Specifically, sea ice, floating on the ocean water, serves to isolate the frigidly cold winter atmosphere from the relatively warm ocean. That is, the water cannot get colder than the freezing point ($-2\,^{\circ}$C for seawater) or it will freeze, whereas the winter atmosphere over ice, in the absence of warming sunlight, cools to approximately -15 to $-40\,^{\circ}$C. The layer of ice lying on top of the water insulates the water just as double-glazed glass insulates a home – if the window is opened (the ice removed), the warm ocean immediately loses heat to the atmosphere and, because the ocean water contains considerably more heat than the overlying atmosphere, the latter will warm to the water temperature, e.g., from – 40 to $-2\,^{\circ}$C, about a $40\,^{\circ}$C warming! Thus, as sea ice forms, it insulates the water. This leads to a colder atmosphere favoring more ice growth (the thicker the ice and overlying snow cover, the better the insulation), and thus colder atmospheric temperatures, etc. In the warmer summer, the air and water temperatures both hover near the freezing point, so this insulating effect and positive feedback occurs primarily in winter, when the atmosphere is at its coldest and the ocean is its only source of heat.

Even small cracks in the ice (known as leads), like a window slightly open, allow a considerable amount of heat to escape. In fact, heat is lost through leads 100 times more effectively than it is lost via conduction through the thick Arctic ice cover, and 20 times more effectively than it is through the relatively thin Antarctic ice cover. In both cases, this means that even a small increase in the number and size of leads can result in a tremendous amount of atmospheric warming. More leads also decrease the area covered by the highly reflective ice, which reduces the albedo as well, further warming the system.

Leads reflect a balance between the rate at which winds and tides open them, by moving ice apart, relative to how fast new ice growth in the leads closes them. Because of

their importance, we typically classify a sea ice cover not only as perennial or seasonal, according to its persistence in time, but also by the fraction of leads within an ice cover. A 95% ice concentration, or ice cover, indicates that 95% of the area is covered by ice, while the remaining 5% represents leads, exposing the ocean directly to the atmosphere. It is difficult to sustain a 100% ice cover over broad areas because irregular winds tend to blow the ice apart in some regions, while forcing it to collide in other locations. Where it collides, it often crumbles into broken and overridden blocks, forming rough and irregular ridges. This is *ice rafting*.

In summary, both the albedo and insulation of the ice cover work to reduce the polar atmospheric temperature and thus exacerbate the equator–pole temperature contrast, and increase the vigor (and with that, the characteristics) of our climate. In addition, a multitude of secondary feedbacks, unique to the sea ice region, also operate in response to sea ice changes. For example, as sea ice forms, and leads close, the atmosphere is impeded in its ability to draw moisture from the ocean for local cloud formation. Clouds serve as warm insulating blankets by trapping heat, but they also serve as umbrellas that cool the surface by blocking sunlight. Whether a change in cloud cover leads to a net warming or cooling in polar regions is a matter of debate (there is some indication that the net influence of clouds is different in polar regions relative to their net influence in tropical regions; recent evidence suggests clouds lead to a net warming in the Arctic – that is, the blanket warming dominates over the umbrella cooling).

Other aspects of the ice–ocean feedback can influence climate by modifying the ocean s ability to transport heat. The slow but unremitting ocean circulation is part of the Earth s arsenal for moving heat from the equator to the poles, though the ocean often accomplishes this through convoluted and circuitous paths. In the Southern Hemisphere, the circulation of the deep waters accomplishes this ocean transport of heat to the polar region. In the Antarctic polar seas, ocean circulation, driven by the winds, forces the deep water to the surface where the excess heat (relative to the freezing point) is vented to the atmosphere, accomplishing the delivery of heat to this cold region. How effective this heat transfer is depends on ocean–ice interactions in both the Arctic and Antarctic polar seas.

For example, sea ice formation works like a distillery, separating the freshwater from the salt in seawater, resulting in nearly freshwater ice (though a small amount of salt is still trapped between the crystals of frozen water) and saltier seawater. The density of seawater is controlled by its saltiness in regions where the water is very cold (in warmer regions, temperature controls the density). In polar regions, the water column can be divided predominantly into two layers, a relatively cold but fresh buoyant surface layer (typically the uppermost 100–200 m) and warmer but saltier deep water (several kilometers thick). As the seawater freezes, the salt is released into the ocean, which makes the surface water saltier and thus denser. This denser water begins to sink because it is heavier than the water immediately beneath it (which is not as dense). The warmer, underlying deep water replaces the sinking water as it mixes upward to replace the sinking surface water. This delivers heat to the surface that reduces ice growth or melts existing ice (a primary reason why the Antarctic sea ice is so thin). This process is not active in the Arctic because the surface ocean layer is so fresh; even with salinization it does not grow dense enough to sink and bring warmer deep water upward, which is why Arctic ice is so thick. If enough ice forms in the Antarctic, the surface layer may become so dense that the entire upper layer mixes downward into the deep layer, destroying the ocean stratification and delivering considerable heat to the ocean surface. In fact, enough heat can be delivered in this manner to eliminate all of the overlying sea ice (forming what is called a *polynya* – an area of ice-free water surrounded by ice). This sinking water will also replenish the deep-water circulation because the sinking of dense surface waters drives that circulation.

Conversely, as sea ice is blown by the winds (generally at 2–4% of the wind speed), it drifts into regions where it eventually melts, freshening the surface ocean layer. This can influence the global deep-water circulation, which begins in the northernmost North Atlantic, where warm and salty subtropical waters, transported northward by the extension of the Gulf Stream are cooled (warming the neighboring atmosphere and helping to keep NW Europe warm). As it cools, the salty subtropical water grows denser, eventually sinking (while still warmer than the freezing point) north of Iceland and southwest of Greenland. This initiates southward flow to the Antarctic and eventually the rest of the world s oceans. If an unusual amount of sea ice from the Arctic drifts into these sinking regions and melts, the buoyant melt water prevents sinking, which inhibits, or shuts down the deep-water circulation (and its ability to deliver heat to the Antarctic). It also has the potential for altering the Atlantic surface circulation, because warm subtropical water no longer needs to flow northward to replace the water from that region that is sinking and flowing southward. Such influences have been hypothesized as being responsible for abruptly altering climate, particularly in Europe, because that region depends on the warm extension of the Gulf Stream to keep the winters moderate. However, counter arguments state that an unusual sea ice incursion is more likely a consequence of an already changing climate, not a cause. Comparable arguments suggest such sea ice incursions and their impact on the deep-water circulation can initiate an avalanche

of global feedbacks that ultimately trigger the great ice ages.

How sensitive global climate is to changes in the sea ice distribution and its growth/melt processes is still undetermined. This reflects the complexity of the processes involved and the myriad of feedbacks operating elsewhere in the climate system that may ultimately diminish or enhance the sea ice driven impacts. At present, our best estimates come from model studies. However, because of the complexities and still imperfect understanding, many of the sea ice processes and feedbacks are represented in the climate models in quite simplified ways, if at all. Consequently, a recent comprehensive international study found that the most important reason for differences among the results of global climate models was related to differences and inadequacies in their treatment of the polar regions.

However, some model studies explicitly targeting the impact of sea ice on climate have shown that sea ice fields play a major role in determining the extent of change in global climate. For example, an investigation using NASA s Goddard Institute of Space Studies climate model to simulate the climate under conditions of doubled atmospheric CO_2, found that 37% of the global warming simulated by that model was a consequence, directly and indirectly, of the modeled changes in the sea ice fields. While this estimate cannot be viewed as precise, it serves to underscore just how important the treatment of sea ice is in models, and it suggests how significant the various sea ice related feedbacks can be.

In the final decades of the 20th century, perennial sea ice appeared to have gotten thinner, and to have been slowly diminishing in areal extent, retreating in the Arctic at a rate of approximately 4% decade^{-1}. If sustained, the perennial ice cover would be eliminated in a few hundred years. However, theoretical and modeling studies suggest that, unless some other compensating changes are introduced elsewhere in the climate system, the various feedbacks will work to accelerate the rate of retreat. One recent study suggested that the perennial ice cover could be eliminated within 50 years (though it will likely always be cold enough to grow some winter sea ice). At present, it is not clear why the ice is retreating, but one detailed model study suggests that the decrease is consistent with the accelerated global warming experienced over the latter quarter of the 20th century. One model suggests that such a sustained retreat is highly unlikely to arise as a result of natural climate variability and that the melting is more consistent with anthropogenic warming. Clearly these ideas need further testing and analysis as understanding improves and improvements are made to models to better represent sea ice-related processes (progress is slow given the difficulty of studying these processes in the remote and hostile polar environment).

Finally, in addition to having the potential of playing a major role in determining climate, sea ice has another interesting characteristic that makes it valuable for climate studies. Sea ice has been identified as a sensitive indicator of climate change because global temperature changes are amplified in the polar regions: warming in the temperate zones is accompanied by even stronger warming in the polar regions. Consequently, as we seek to identify clear signals of anthropogenic warming, it has been noted that such a signal may first appear above the background noise (natural variability) in an amplified response of the polar regions. Furthermore, in the Arctic, where the ocean offers little resistance to ice growth, changes in atmospheric temperature may be reflected as changes in sea ice extent. Sea ice has the distinct advantage of being a highly visible medium, easily monitored from space via satellites. Thus, changes in sea ice may be easily observed and serve as an early warning indicator of global warming. However, while this basic concept appears to be valid, whether such warming is due to anthropogenic change or natural variability requires additional investigation, such as the modeling studies mentioned above. Besides the obvious physical influences and implications, the sea ice fields play a major role in the seasonal evolution of the unique and limited polar flora and fauna. The sea ice is thought to help establish surface ocean conditions that allow the large and impressive phytoplankton blooms to occur in spring and summer. In the Antarctic polar oceans, these blooms are exceptionally large, playing a role in the global uptake of atmospheric CO_2 and providing food for larger animal life, such as krill, penguins and whales. In the Arctic, the sea ice provides the means by which polar bears access the highest northern latitudes where they feed and live during the summer months. Without the sea ice, the bears would be limited to the continental regions surrounding the Arctic. Thus, changes in sea ice extent, and its seasonal cycle, may have profound, although unknown implications to the polar biology.

See also: Arctic Climate, Volume 1; Arctic Ocean, Volume 1; Cryosphere, Volume 1; Climate Feedbacks, Volume 1; Polynyas, Volume 1; Southern Ocean, Volume 1.

FURTHER READING

Johannessen, O M, Muench, R D, and Overland, J E, eds (1994) *The Polar Oceans and Their Role in Shaping the Global Environment, Geophysical Monograph*, American Geophysical Union, Washington, DC, 1–525.

Untersteiner, N (1986) *The Geophysics of Sea Ice*, NATO ASI Series, Physics, Vol. 146, Plenum Press, New York, 1–1196.

Wadhams, P (2000) *Ice in the Ocean*, Gordon and Breach, London, 1–351.

http://www.socc.uwaterloo.ca. General site on the State of the Canadian Cryosphere; and excellent general location regarding sea ice and other aspects of the Earth s cold regions.

http://nsidc.org/NASA/SOTC/index.html. General site of the National Snow and Ice Data Center, reporting all relevant information on the State of the Cryosphere – excellent site with considerable introductory material and information for the general public.

Sea Level

Michael C MacCracken
Lawrence Livermore National Laboratory, Livermore, CA, USA

Oceans currently contain about 97% of the Earth's water and cover about 70% of the Earth's surface. At the edges of oceans, shorelines have been created. The geological record provides clear evidence that there have been very large changes in both the level of the oceans and in the locations of shorelines over the history of the Earth. While the natural processes governing sea level continue to play out, human activities have introduced some new ways in which sea level, and thereby coastlines and the coastal environment, can be changed. For example, projections are that human activities are likely to cause sea level rise during the 21st century to be several times larger than the rise of about 10- 20 cm that occurred over the 20th century.

OBSERVATIONS OF SEA LEVEL

Viewed from an airplane, the different nature of the land and oceans is quite apparent. The land is very rough and rigid, with mountains reaching up to almost 9 km. From a distance, however, the ocean seems quite smooth. For several reasons, this smoothness is somewhat deceptive. First, soundings of the ocean bottom indicate that ocean depth varies about as much as the land surface, and it is just the effect of gravity on the water that is creating the apparently smooth surface. Second, while the water surface can be smooth locally, the various currents, ocean ridges, temperature variations, and gravitational effects cause the ocean height, relative to the geoid, to vary by a few tens of meters (with extremes of as much as 100 m).

Viewed from up close, the oceans are also seen to be always moving. Walking near a shoreline, the most obvious feature of the ocean is perhaps how much the height of the water is going up and down. The level rises and falls with each wave and how it strikes the shoreline, the level varies through a semidiurnal and daily cycle of tides and these tides vary in location along coastlines and estuaries. The heights of the tides also vary through the month, roughly in phase with lunar influences (*see* Tides, Oceanic, Volume 1). Ocean currents and variations in the heights of waves and tides also depend on the weather and on the nature of the location (e.g., the depth of the water and the shape of the coastline). In addition, changes in winds and sea level pressure can push the ocean down or allow it to surge upward relative to the level established by the tides (e.g., the winds and low pressure during a hurricane or typhoon can cause water levels to rise up to several meters, depending on the storm and location).

In many regions, sea level varies with what is happening in the global ocean, including whether there is an El Niño or La Niña or other such global climatic fluctuation. Historic records also document that the level of the ocean is changing relative to the level of the land. For example, islands present in historic times in the Chesapeake Bay in the eastern US are now covered with water, and the state of Louisiana is losing about 50 $km^2\,year^{-1}$ to inundation. In addition, anthropological records suggest that the Americas were populated by peoples migrating across what are now the Bering Strait and Bering Sea at a time when the oceans had receded to expose lands currently up to 100 m below sea level. Geological records provide even more dramatic evidence of there being changes in ocean locations and boundaries, with what are clearly ocean bottom sediments now located well inland and lofted to become parts of mountains. Quite clearly, the location and the level of the sea has varied quite extensively due both to climate variations and to land movements, and there is no basis for thinking that long-term changes due to these natural processes will not continue into the future.

For the purposes of considering the potential importance to society of sea level change, what is important, however, is to determine how the average sea level may change over the time scales of societal interest. Tide predictions and weather and storm forecasts focus on how sea level will fluctuate over hours to days. Satellites now provide relatively accurate estimates of the variations in mid-ocean sea level due to changes in ocean currents and temperature, allowing forecasting, for example of El Niño and La Niña events (*see* El Niño and La Niña: Causes and Global Consequences, Volume 1).

In the context of long-term environmental change, the time periods of interest are decades to centuries, even though the effects of sea level rise are likely to be most noticed during times of extreme weather (e.g., hurricanes and typhoons). In addition, to provide context for understanding whether such changes are significant and how they may affect society, gaining an understanding of changes experienced over historical times are of interest.

Until recently, virtually all high-precision sea-level measurements have been made at coastlines. Coastal measurements typically involve creating some vertical scale that is solidly anchored (e.g., initially markings on a rock or on a piling). Boxes or baffles are also constructed to reduce the variations in level caused by the waves. A time-varying history of changes is recorded and then a time-averaged value is generated. As shown in Figure 1, however, even averaging over a year does not create a smooth record. That there is general agreement in interannual departures between locations as far apart as San Diego and San Francisco, for example, is an indication that large-scale oceanic conditions are playing a role in determining these fluctuations. This suggests the need to average over much longer times (e.g., a century) to get a trend. This has been done for a number of stations around the world. What becomes evident is that the long-term changes vary with where the measurement is taken, suggesting that long-term regional-scale processes are also important to consider.

Until very recently, measurements of sea level along coastlines have been relative sea level measurements because they have not been able to determine the separate contributions due to changes in the absolute level of the ocean and changes in the level of the land. This becomes an important issue because there are a number of factors that can cause the level of the land to change. These include compaction of newly deposited materials (which is particularly important along river deltas), subsidence caused by the extraction of water or other substances from the ground, vertical shifts caused by earthquakes, and the long-term upward or downward isostatic adjustment of the land surface (*see* Isostasy, Volume 1) occurring in response to the melting of the very heavy glacial ice that covered many continental areas from roughly 20 000 to 10 000 years ago. Such changes in land elevation can introduce an important bias into observations made at particular locations (e.g., Scandinavia, northeastern North America). To derive a meaningful estimate of the absolute change in sea level, this bias needs to be accounted for, either by removing the bias (if it can be determined, e.g., by using the new global positioning satellites) or by considering only results from stations where the land is not moving vertically. When the upward and downward movement of observation stations is accounted for, the results indicate that, over the 20th century, the average sea level rose by about 10–20 cm (Church *et al.*, 2001), with perhaps a tendency toward the upper limit (Douglas, 1997).

In that coastal observations at a global set of stations go back only of order 100–150 years, historic records can be used to estimate sea level at earlier times. For example, some Roman baths were apparently built at sea level over 2000 years ago, and until recently have remained near sea level. Heights of coral and of beaches also suggest that global average sea level has been relatively constant for several thousand years, with estimates being that sea level has varied by less than about 0.3–0.5 m over the past 6000 years. These records also indicate that the long-term average rates of change of sea level have been only about 1–2 cm per century, or about one-tenth of the rate of change during the 20th century (Church *et al.*, 2001).

Quite certainly, observations indicate that the historical development of coastal cities occurred during times of relatively stable sea level, but that coastal dwellers are now facing a rising sea level.

Further back than 6000 years, large ice sheets covered much of North America and Eurasia, with maximum extent peaking during the Last Glacial Maximum about 20 000 years ago (*see* Last Glacial Maximum, Volume 1). At that time, sea level was roughly 125 m below current levels. This is equivalent to about 3% of the oceans waters being accumulated on land as glacial ice. Although there are some uncertainties due to uncertainties about the amount of subsurface waters, this about doubled the amount of water not in the oceans. At the last interglacial (the Eemian), which peaked about 125 000 years ago and was probably a bit warmer than at present, there are some indications that sea level may have been a few meters higher than at present. This suggests that warming could indeed induce some melting of either the Greenland and/or West Antarctic ice sheets.

FACTORS CAUSING SEA LEVEL CHANGE

Global sea level is affected by many factors, some obvious and some quite subtle. In analogy with a bathtub, the level of the water is affected by the shape of the tub, the density of the water in the tub (less dense water takes up more space, so levels are higher), and by the amount of water put into the tub.

Over very long time scales, the shape of the ocean basins is changing (*see* Continental Drift, Volume 1). While continents are drifting relative to each other at rates of centimeters per year, the rate of continental drift is unlikely to be changing and this factor is not contributing significantly to changes in sea level, especially the increase rate of rise during the 20th century. In addition, the relative stability of sea level over the past few thousand years suggests that these contributions to sea level change must be small. Sediments are also carried into the oceans in rivers, thereby displacing some of the water, but this process is likely to be quite small. For the distant future, the melting of methane clathrates in the ocean sediments may become a contributing factor (*see* Methane Clathrates, Volume 1). Because these substances expand when frozen (just as ice takes up more volume than water), melting would create a tendency for ocean levels to drop slightly. However, such a process would likely take many hundreds to thousands of years to occur to cause a significant change.

In addition to changes in sea level that might result from changes in the shape of the ocean basins, sea level is also dependent on the density of the water. As for other fluids, the warmer it is, the less dense it is; as a result, the warmer it is, the more volume a fixed mass of water needs to occupy. Thus, sea level depends on the temperature of the oceans, and changes in sea level can result from changes in the temperature of the oceans. Although data sets of the temperature of the oceans at depth are relatively limited, recent compilations of the observations indicate that the upper 1–2 km of the oceans are warming, and that this is contributing to sea level rise (Levitus *et al.*, 2000). Combined with model simulations, the observations suggest that this thermal expansion effect has caused a rise in sea level of roughly 3–7 cm during the 20th century.

Finally, there are quite a number of factors that can affect the amount of water that is in the oceans. Among the major factors are the amount of ice and snow on land, the amount of water in underground aquifers, the amount of water stored in permafrost, and the amount of water stored on the land in reservoirs. In considering the transformation of snow and ice to ocean waters, it is important to recognize that a significant time lag is created by the time it takes to accumulate the energy to carry out the melting. Thus, while warming of the atmosphere can warm the ice, melting takes much longer because as much energy is needed to melt the ice as it takes to warm the ice by 80 °C. This large energy requirement has the effect of essentially anchoring the temperature at the melting point for a considerable time as warming is occurring.

Based on studies reported in Church *et al.* (2001) that account for the appropriate delay times, Table 1 provides a summary of the various factors that may change the amount of water in the oceans. The table also indicates the best estimates for how changes in these terms may have affected sea level during the 20th century.

- Mountain glaciers and icecaps (excluding Antarctica and Greenland) store an amount of water equivalent to about 50 cm of sea level (*see* Global Plate Tectonics, Volume 1). Over the past 100 years, the retreat of many glaciers has been evident in mountain ranges around the world from the tropics to upper mid-latitudes. Although some mountain glaciers in high latitudes are advancing, the overall effect is estimated to have caused a rise in sea level of 2–4 cm.
- Greenland is an ice sheet that is quite old, with some drill holes finding ice as old as 200 000 years. Other geological evidence suggests the ice sheet is likely to be much older. The Greenland ice sheet holds enough water to raise sea level by about 7 m were it all to melt, a long-term possibility suggested by the near absence of large ice sheets at the same latitude on adjacent continents. There are relatively limited observations of changes in the Greenland ice sheet until quite recently, when a network of surface stations was established. Surface and satellite observations indicate that some regions are currently melting while other regions are accumulating snow (*see* Greenland Ice Sheet, Volume 1). For the 20th century, the limited

Table 1 Estimates of contributions to changes in sea level for the period 1910–1990, based on observations and on modeling studies (Church *et al.*, 2001)

Factors contributing to sea level rise	Lower bound estimate (cm)	Central value estimate (cm)	Upper bound estimate (cm)
Thermal expansion of seawater due to oceanic warming	3	5	7
Greater retreat than growth of mountain glaciers and icecaps	2	3	4
Greater melting than snow accumulation on Greenland	0	0.5	1
Accumulation of snow on Antarctica	−2	−1	0
Long-term adjustment of Greenland and Antarctica in response to end of last glacial period	0	2.5	5
Melting of some permafrost	0	0.25	0.5
Changes in the amount of water stored in reservoirs and aquifers and changes in land cover	−11	−3.5	4
Rise in sea level due to flow of sediments into the ocean via rivers	0	0.25	0.5
Sum of estimates of individual contributions to sea level rise (1910–1990)	−8	7	22
Observed rise in sea level during the 20th century (Church *et al.*, 2001)	10	15	20
Observed rise in sea level during the 20th century (Douglas, 1997)	17	18	19

observations combined with modeling studies suggest that a small melting of the Greenland ice sheet has occurred, causing a sea level rise of up to 1 cm.

- Antarctica is a very large ice sheet, holding enough water to raise sea level by roughly 60–70 m were it all to melt (the range resulting from how the long-term adjustment of the land is accounted for after melting occurs). The Antarctic ice sheet is believed to be many millions of years old and it would take a very long time to melt, especially because of its location near the South Pole where the sun angle is quite low. Although observations are quite limited, the best observational and modeling results suggest that the current climatic conditions are causing snow to accumulate over much of Antarctica, and that this accounted for the removal of about 0–2 cm of sea level equivalent from the oceans during the 20th century.
- In addition to the estimates of actual accumulation and melting over Greenland and Antarctica due to recent climatic change, geological data suggest that changes in these ice sheets are still occurring as a continuing adjustment to the warming and sea level rise that took place as the climate shifted from glacial to the interglacial conditions. The best estimates, which are constrained by the rates of crustal rebound, are that these changes could have contributed 0–5 cm to sea level rise during the 20th century, although this number is quite uncertain.
- Permafrost is observed to be melting across locations in North America and Eurasia. This has caused the release of water to rivers that in turn flow into the oceans. This is estimated to have caused up to 0.5 cm of sea level rise.
- Changes in the storage of water on land, in aquifers and in reservoirs, as well as changes in land cover, could have caused quite large changes in sea level, although estimates are still quite uncertain. The ways in which this could occur include the pumping out of groundwater, the drawdown of lakes, impoundment in and downward infiltration of water from reservoirs, infiltration through irrigation, runoff from urbanization, deforestation, and changes in evapotranspiration. Estimates in Church *et al.* (2001) suggest the net effect on sea level change could have reduced sea level by as much as 11 cm or to raised it by as much as 4 cm during the 20th century.

As summarized in Table 1, these terms would allow for a sea level change of from −8 to 22 cm. This range brackets the independently determined observations of sea level rise, which suggest a rise of 10–20 cm during the 20th century (Church *et al.*, 2001), perhaps more likely to be in the upper half of this range according to Douglas (1997). That the agreement is not better is somewhat discouraging, in that each centimeter of sea level rise is equivalent to about 4000 km^3 of water or ice, and to be missing several times that much ice or water seems quite an oversight. However, just to make clear how hard even such a large amount is to detect, when spread across Antarctica, for example, each centimeter of sea level amounts to only about 30 cm of ice, and so, over a century, even seemingly large amounts of water can become hidden from direct measurement, even though other constraints (as for example on changes in the rotation rate of the Earth) may point indirectly to what is happening.

PROJECTED CHANGES IN SEA LEVEL IN THE FUTURE

The projected rate of sea level rise for the 21st century is likely to be significantly larger than for the 20th century. Warming of the global ocean waters will become a much more important contributor to sea level rise during the 21st century. With projections that the global average temperature is likely to rise a few degrees rather than about 0.5 °C, thermal expansion is projected to contribute to an increase in sea level of roughly 10–40 cm over the 21st century, with comparable amounts during ensuing centuries even if the rate of rise in global temperatures is slowed because the deep ocean will be continuing to warm to catch up with surface warming. At the same time, there are likely to be contributions to sea level due to the warming of the land surface. This is likely to lead to the lowering of some lakes, increased pumping of some aquifers, and changes in land surface vegetation. While these factors are likely to contribute to sea level change, estimates of their potential influence remain quite uncertain.

The increased rate of warming is also projected to substantially increase the rate of melting of mountain glaciers and icecaps. Estimates are that this could lead to a loss of up to 30–50% of the mass of mountain glaciers, which would contribute to a potential rise in sea level of roughly 15–25 cm. In addition, the melting of permafrost will likely accelerate, contributing an additional small amount to sea level rise. Other changes due to factors involving storage or depletion of water in reservoirs and aquifers and due to changes in vegetation are quite uncertain.

For Greenland and Antarctica, the situations are quite complex. At least for some conditions, observations and model calculations suggest that warming of the adjacent oceans and the very cold air over the ice sheets is very likely to lead to increased snowfall, tending to pull sea level down. At the same time, the ice at the edges and at lower elevations will be subjected to warmer conditions. For Greenland, there are already indications that deterioration of the ice sheet is occurring in some regions (*see* Greenland Ice Sheet, Volume 1). For Antarctica, there is melting in the area of the Antarctic Peninsula due to regional warming, but, although some ice streams are moving in unusual ways (*see* Antarctica, Volume 1), there are no indications of warming-related losses for the East or West Antarctic ice sheets. However, there is expected to be a continuing adjustment of these ice sheets to the long-term rise in sea level and warming that occurred coming out of the Last Glacial Maximum. The net effect on sea level due to changes in these two ice sheets together is estimated here to range very roughly from −15 cm to +15 cm during the 21st century, depending on the climate change scenario, with Greenland adding to a tendency for sea level rise and Antarctica to a tendency for a sea level drop.

In their analysis of all of these factors, the Intergovernmental Panel on Climate Change (2001) used a number of model-based analyses to arrive at a best estimate that sea level would rise by from 9 to 88 cm during the period 1990–2100, with a mid-range estimate of about 48 cm. Such a rate of rise would be roughly 2–5 times higher than during the 20th century, and a rate not experienced during historical times. For those living along coastlines, converting these estimates of change in global sea level to particular locations requires adding (or subtracting) the local rate of subsidence (or uplift) of the land surface. In particular locations, these vertical adjustments can amount to as much as several tens of centimeters per century.

What might be most important, however, is that during the 21st century, warming could pass a threshold that would start a multi-century long deterioration of the Greenland and Antarctic ice sheets. For example, Church *et al.* (2001) cite studies indicating that a 3–6 °C warming of Greenland, if sustained, could lead to a few meter rise in sea level over the next 1000 years, and some projections of warming are indeed this large. Similarly, such a rise in Antarctica, if sustained, could lead to a similar size rise in sea level over many centuries. Thus, while not at all yet certain, the warming projected for the 21st century could initiate events that would lead to significant inundation of coastal regions over the very long term.

SUMMARY

After apparently showing relatively little change over the past several thousand years, sea level started to rise during the late 19th century and this upward trend continued through the 20th century. Projections indicate that the rate of rise is likely to further increase during the 21st century, posing serious threats to such low-lying areas as river deltas, barrier islands, coastal wetlands, and, in many regions, coastal cities and infrastructure.

REFERENCES

Church, J A, Gregory, J M, Hubbrechts, P, Kuhn, M, Lambeck, K, Nhuan, M T, Qin, D, and Woodworth, P L (2001) Changes in Sea Level, in *Climate Change 2001: the Scienti c Basis*, eds J T Houghton, Y Ding, D J Griggs, M Noguer, P J van der Linden, X Dai, K Maskell, and C A Johnson, Intergovernmental Panel on Climate Change Working Group I, Cambridge University Press, Cambridge.

Douglas, B C (1997) Global Sea Level Rise: a Redetermination, *Surv. Geophys.*, 18, 279–292.

Douglas, B C (2001) An Introduction to Sea Level, in *Sea Level Rise: History and Consequences*, eds B C Douglas,

M S Kearney, and S P Leatherman, Academic Press, San Diego, CA, 1–11.

Douglas, B C, Kearney, M S, and Leatherman, S P (2001) *Sea Level Rise: History and Consequences*, Academic Press, San Diego, CA, 1–231.

Intergovernmental Panel on Climate Change (2001) *Climate Change 2001: the Scienti c Basis*, eds J T Houghton, Y Ding, D J Griggs, M Noguer, P J van der Linden, X Dai, K Maskell, and C A Johnson, Cambridge University Press, Cambridge.

Levitus, S, Antonov, J I, Boyer, T P, and Stephens, C (2000) Warming of the World Ocean, *Science*, 287, 2225–2229.

Sea Surface Temperature

Yochanan Kushnir
Columbia University, Pailsades, NY, USA

The oceans affect climate primarily through their surface temperature. Determined through a complex interaction between the atmosphere and oceans, sea surface temperature (SST) is arguably the most important indicator of the oceans' climate. It is important not only because the oceans cover about 70% of the Earth's surface, but also because it is a rough indicator of the heat contained in the upper ocean. The difference between SST and the overlying surface air temperature is one of the factors determining the heat exchange between oceans and atmosphere in the form of evaporation, sensible heat ux, and infrared radiation. Month-to-month and year-to-year changes in SST in uence the state of the atmosphere, particularly in the tropics, where they can modulate the strength and location of deep convection, the engine that drives the global atmospheric circulation. Information on the climatological (mean) SST distribution and its temporal uctuations is essential for simulating the current climate and predicting its future evolution.

Because the ocean is continually mixed by the action of winds and gravity, a homogeneous layer of some thickness forms at its surface. This mixed-layer exhibits a vertically uniform temperature (and salinity) distribution represented by the SST (and surface salinity). The depth of the oceanic mixed layer generally ranges between 20 and 200 m (and in some locations is even deeper), depending on the season and the geographical location (Figure 1). At the ocean surface, the specific heat of water is about four times greater than that of air and its density a thousand times larger; thus the mixed layer capacity to store heat far exceeds that of the entire atmosphere (see e.g., Pexióto and Oort, 1992). This large heat capacity explains the damping effect of the upper ocean on the thermal variability of the marine atmosphere, as well as the moderate nature of coastal climates compared to those of inland regions. The heat contained in the mixed layer is available for transport by the upper ocean circulation and can affect regions far away from where the thermal properties of the water were initially shaped.

The horizontal and seasonal distribution of SST is set by a balance between net incoming solar radiation (what is left after absorption in the atmosphere and reflection from clouds, aerosols, and the ocean surface) and the net loss of heat by infrared radiation (the difference between the surface emission and the heat radiated downward from the atmosphere). Solar radiation is absorbed in the ocean s upper few tens of meters, while outgoing infrared radiation cools just the surface. Evaporation (latent heat flux) and thermal conduction (sensible heat flux) also cool the surface by moving moisture and heat from the ocean to the atmosphere (only in a very few locations, primarily during summer, does the atmosphere lose heat to the ocean). Under the action of gravity (water parcels that cool at the surface, become denser than the water below and sink down) and the mechanical mixing induced by the surface wind stress, a nearly uniform temperature distribution is established within the mixed layer (Figure 1). At its interface with the deep ocean, changes in mixed-layer depth cause the entrainment of cold water from below, imposing additional constraints on SST. Where wind-induced mixing and surface heat loss are large, as is the case over the mid-latitude and subpolar oceans during winter, the mixed layer is deep and relatively cold. In the subtropics and tropics, where winds are usually weak and solar heating is large, gravitational and mechanical mixing in the upper ocean are weak and the mixed layer is generally shallow and warm. Weak winds and strong solar heating also cause the seasonal thinning of the mixed layer during summer compared to winter (Figure 1).

In addition to these air–sea processes, oceanic heat advection in the horizontal and vertical also affects SST. Advection involves the movement of heat by the flow of water masses from warm to cold regions and *vice versa*, damping or enhancing the influence of surface fluxes, depending on the geographical location and time of year. The interactions governing SST are further complicated by the fact that surface fluxes depend not only on atmospheric condition but also on ocean heat advection, which can also occur in response to forcing by local and distant surface winds.

The mean seasonal variation of SST is shown in Figure 2, which depicts the average January and July conditions over the global ocean. Year round, SST is warmest in the tropics and coldest in the polar latitudes, a reaction to the latitudinal variation of available solar radiation between the equator

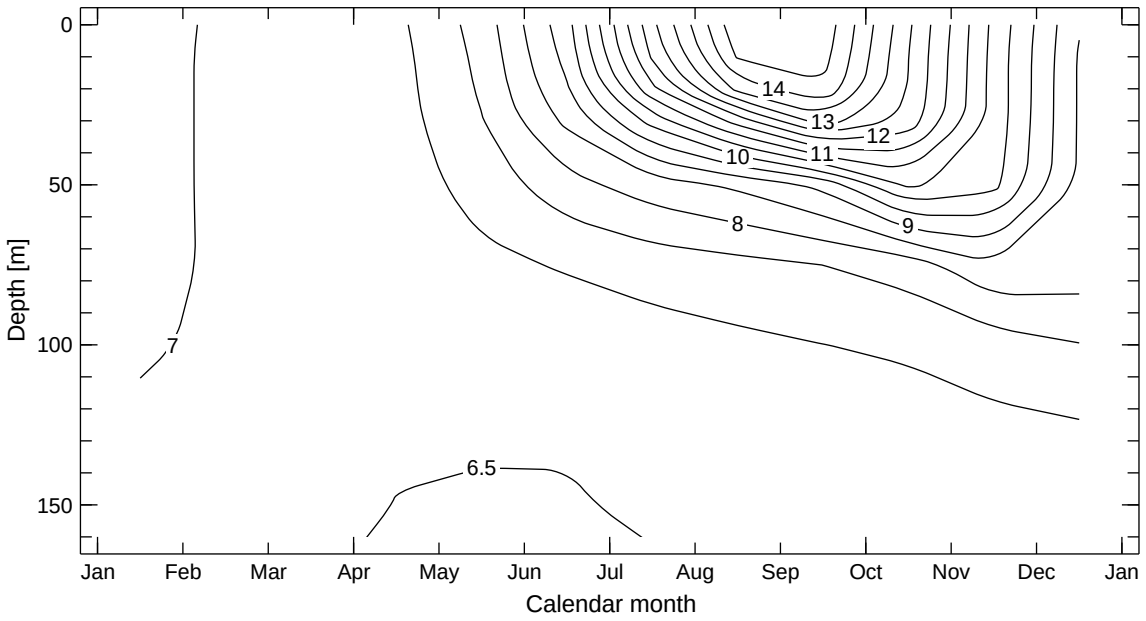


Figure 1 The annual march of upper ocean temperature in the North Pacific (175 °E–150 °W, 40 °N–50 °N) as a function of depth. Contour interval is 0.5 °C. The abscissa is the calendar month and the ordinate is depth in meters. The figure depicts the mixed layer of the ocean. Temperature is uniform from the surface to more than 150 m in March and April, but only to 20–50 m between July and November. (Data are from Levitus and Boyer, 1994)

and the poles. The large influx of solar heating over the tropical and subtropical oceans is offset mainly by the cooling effects of evaporation and ocean heat transport. On the eastern side of the tropical Pacific and Atlantic Oceans, the upward movement of deep cold-water masses in response to the pumping action of the surface winds (called *upwelling*) also causes surface cooling.

An east–west contrast in SST is also found in subtropical latitudes (i.e., between 15 and 30° north and south of the equator) where it is believed to be shaped by the oceans subtropical gyre circulation (the latter is driven by the action of the surface atmospheric winds over the entire basin). The ocean circulation transports water masses from the tropics poleward on the western side of the ocean basins and from the subpolar regions equatorward on their eastern side, thus warming the western side and cooling the eastern side. The Kuroshio (North Pacific), Gulf Stream (North Atlantic), and Brazil Current (South Atlantic) move warm tropical water poleward to warm western boundary SST. The California (North Pacific), Peru (South Atlantic), Canary (North Atlantic) and Benguela (South Atlantic) currents move subpolar water equatorward to cool eastern boundary SST. Low-level atmospheric clouds (stratocumulus) that prevail over the eastern subtropical oceans (partly because SST is colder there) enhance the cooling effects of the eastern boundary currents by sheltering the ocean surface from the full effect of solar radiation. In the eastern coastal waters, upwelling forced by off-shore winds also cools the surface. In the polar latitudes of the Northern Hemisphere oceans, the western basins are colder than the eastern ones. Here the Oyeshio (Pacific) and Labrador (Atlantic) currents bring cold, polar water south, to meet with the warm western boundary currents and create strong horizontal temperature gradients. On the eastern side of the high latitude Northern Hemisphere oceans, the eastern extensions of the western boundary currents (the North Pacific and North Atlantic Currents, respectively) are responsible for relatively warm SSTs. A different, more zonal SST distribution exists in the Southern Ocean. Unperturbed by continental boundaries but constrained by the steep, subsurface topography, the Southern Ocean responds to the strong prevailing surface winds (westerlies) to produce a strong eastward flowing Circumpolar Current that is delineated by a strong SST gradient along the latitude of 40 °S (the Roaring Forties).

The seasonal contrast in surface temperature is depicted in Figure 3, which shows the difference between the January and July averages. In the tropics the differences are small and limited to the far-eastern Pacific and Atlantic Oceans. These eastern ocean differences result from the larger rates of upwelling during boreal summer compared to winter (a direct result of surface wind circulation changes there). Outside of the tropics, local summer SST values are higher than in winter. In the mid-ocean basins, along the latitudes of about 40° north and south of the equator, the difference between January and July reaches up to 6 °C. Due to the large ocean storage capacity and horizontal heat transport, this SST range is much smaller than the corresponding range in land-surface air temperature at the same latitude.

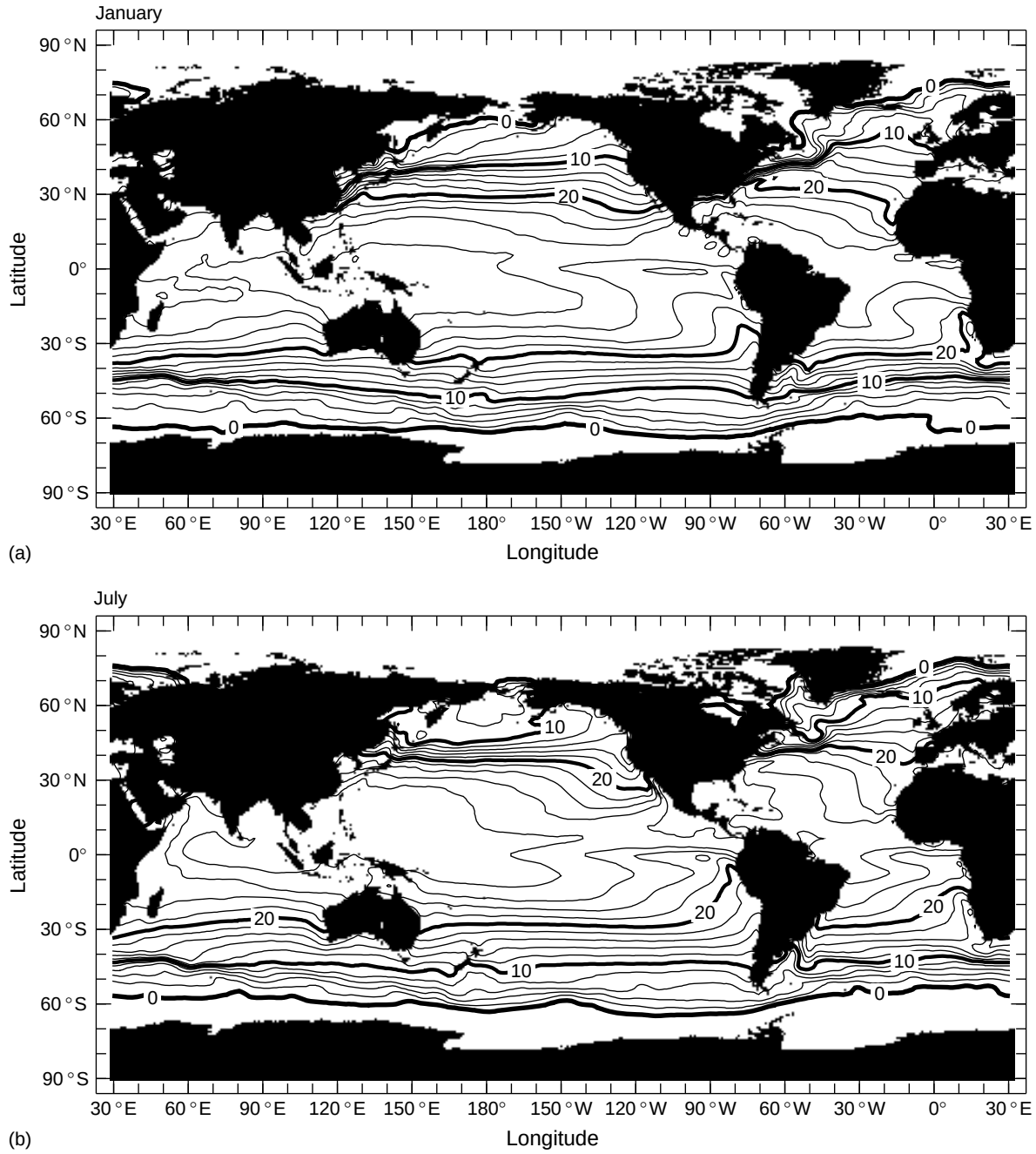


Figure 2 Mean SST for January (a) and July (b). Contour interval is 2 °C. The zero contour is heavy and so are the contours for 10 and 20 °C. (Data are from National Oceanic and Atmospheric Administration (NOAA)/National Centers for Environmental Predictions (NCEP) as described in Reynolds and Smith, 1994)

In the waters along the eastern seaboards of Asia and North America, the seasonal differences reach maximum values. This is due to a combination of low wintertime solar radiation compared to summer and the vigorous wintertime off-shore movement of cold, dry air from the continental interior, which causes ocean-to-atmosphere heat fluxes, by evaporation and conduction, to be very large. Additional discussion of the SST climatology and the associated mechanisms can be found in Peixóto and Oort (1992), and Hartmann (1994).

The departure of SST from its mean state is considered an important measure of the oceans contribution to climate variability (*see* Natural Climate Variability, Volume 1). These departures can result from the passive, local response of the ocean to atmospheric forcing or from an active involvement of ocean heat transport. The

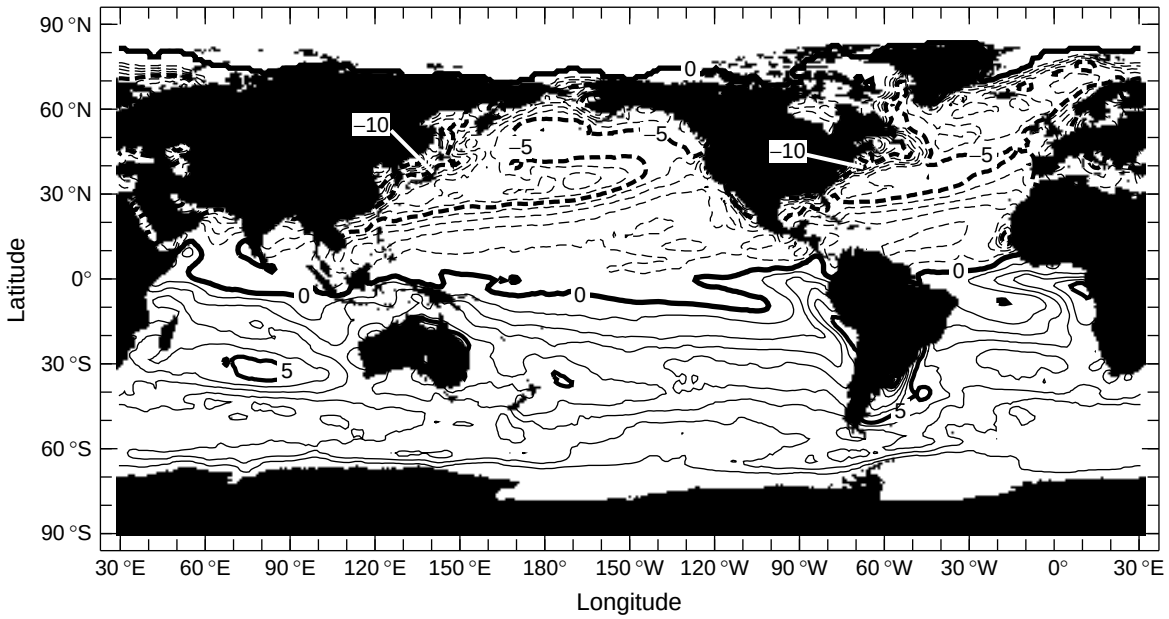


Figure 3 The difference between mean January and July SST of Figure 2. Contours are every 1 °C. Negative contours are dashed. Contours at 0 °C and at 5° intervals thereof are heavy

local interaction is most clearly observed in the extratropics during winter. To a large extent, SST variability in that region results from the slow adjustment of the oceanic mixed layer temperature to random, day-to-day fluctuations of wind, temperature and humidity in the overlying atmosphere, the latter arising from changes in wind speed and direction. The mixed layer s slow adjustment is due largely to its large heat capacity and its feedback on the atmospheric surface air temperature and absolute humidity. The time-averaged effect of fast, chaotic atmospheric fluctuations on the extratropical ocean mixed layer produces anomalies that vary much more slowly than the atmospheric forcing (see excellent review by Frankignoul, 1985 for a mathematical discussion of these processes).

The atmosphere also displays low-frequency variability in the form of large-scale, month-to-month and year-to-year variations in its circulation over the ocean basins (*see* North Atlantic Oscillation, Volume 1; Paci c- North American (PNA) Teleconnection, Volume 1). These low-frequency fluctuations organize the more random variability of weather and leave a basin scale imprint on the SST field (see Figure 4 and Frankignoul, 1985). Recent experiments with climate models show that the response of the extratropical ocean mixed layer to anomalous atmospheric forcing has a positive influence on the latter. Apparently the reaction of the ocean surface temperature to changes in surface windspeed, temperature, and humidity, reduces the thermal drag on the atmosphere. This effect enhances the amplitude and persistence of anomalies in the atmospheric circulation over the ocean (see e.g., Lau, 1997).

The effect of ocean circulation on SST and the maintenance of its anomalies is evident everywhere because the mean circulation can transport SST anomalies and change local conditions elsewhere (*see* Ocean Circulation, Volume 1). Fast responding currents forced by surface wind stress anomalies (often referred to as *Ekman currents*) can instantaneously enhance or dampen the effect of atmospheric flux forcing, depending on their direction with respect to the mean regional SST gradient. The role of the ocean circulation in climate is of particular significance in regions where it responds strongly to changes in surface winds and occurs with a long delay. Such delayed response can affect SST long after the action of the atmosphere has passed and open the way to semi-regular and possibly predictable ocean–atmosphere oscillations. Such oceanic response is involved in the El Niño–Southern Oscillation phenomenon (*see* El Niño/Southern Oscillation (ENSO), Volume 1), where the interaction between the immediate and delayed interaction between the atmosphere and ocean enables the prediction of the phenomenon. Scientists have speculated that a delayed circulation response, such as that acting in the equatorial Pacific Ocean, can also be found in the extratropical oceans where patterns of decadal climate variability are apparent (Latif and Barnett, 1994).

The debate on global climate change drew attention to a possible widespread rise in SSTs, particularly in the tropics. Such a change could be of large consequence to climate worldwide because of the implied potential for increased evaporation. An increase in evaporation must lead to an increase in globally averaged precipitation. Higher SSTs would also lead to an increase in atmospheric water vapor

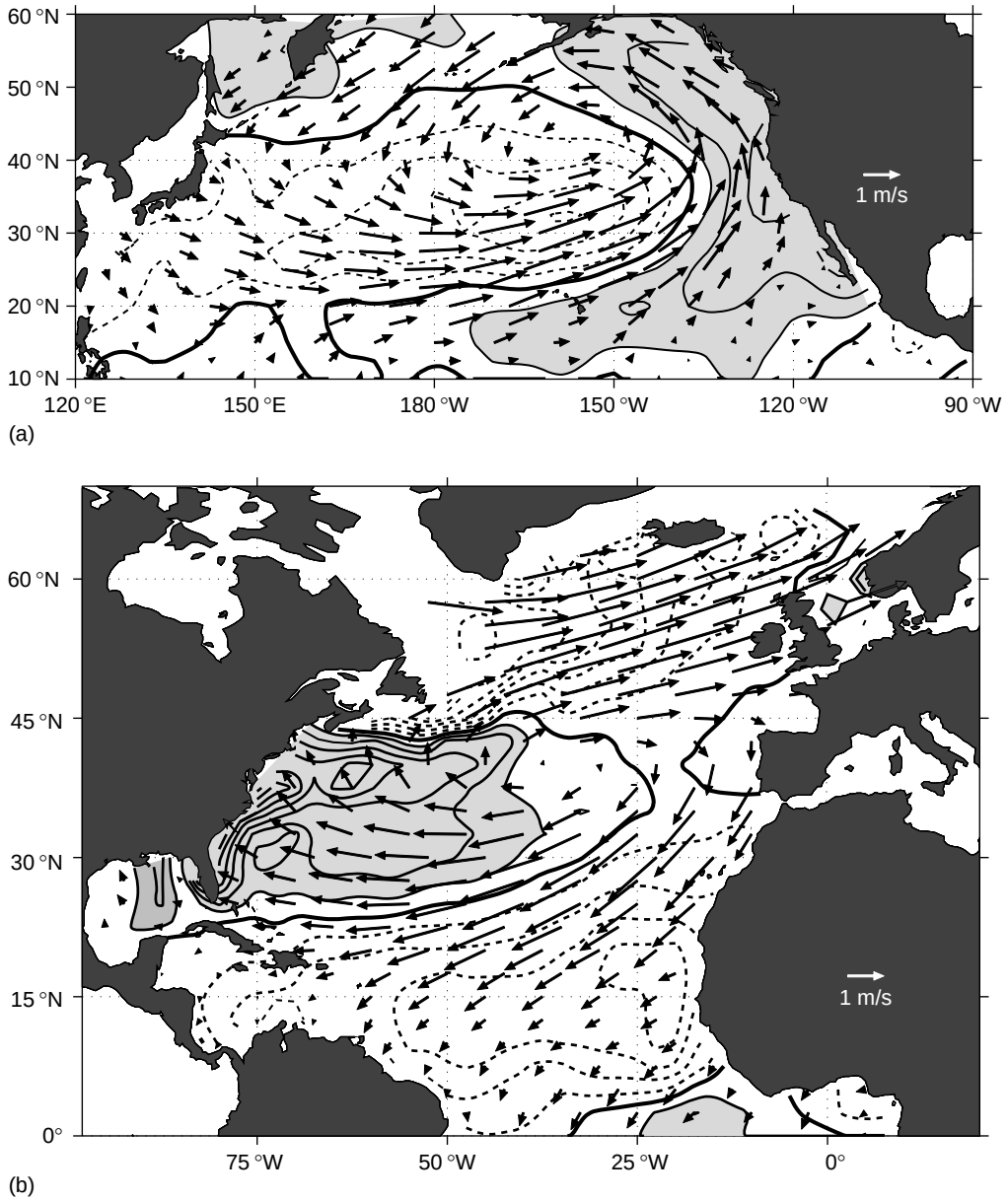


Figure 4 Surface atmospheric wind anomalies (arrows) and SST anomalies (contours) associated with typical wintertime (January–March) climate fluctuations in the North Pacific (a) and North Atlantic (b) Ocean Basins. Contours are every 0.1 °C with heavy zero contours and dashed negative ones. The agreement between the surface wind features (intensity and direction) and the regions of large SST anomalies reflects the role of atmospheric forcing in the generation and maintenance of the latter. (The data used for this figure are from the NCEP/National Center for Atmospheric Research (NCAR) Climate Data Assimilation System 1 (CDAS–1) dataset described in Kalnay *et al.*, 1996)

concentration and this enhances the ability of the atmosphere to absorb and emit infrared radiation (the so-called greenhouse effect). Other influences of higher SSTs could include an increase in the strength of tropical convection including the occurrence and strength of tropical storms (see Henderson-Sellers *et al.*, 1998). The outcome of such tropical changes could be dynamical changes in such extratropical circulation features as the location of jet streams and the seasonal low and high-pressure centers. Warming of the tropical oceans can also lead to changes in the strength and frequency of El Niño (Timmermann *et al.*, 1999) and its global teleconnections.

Because the surface of the Earth is covered mostly by ocean, estimation of the global surface temperature change with time, a primary indicator of global warming, requires knowledge of the concomitant change in SST. Routine SST measurements date back to the mid-19th century when the world merchant fleet initiated weather observations (as

a result of an international agreement). In the 1970s the world climate community began assembling comprehensive global archives of ship weather data in order to reconstruct a century long, monthly history of SST over the World Ocean. Several national meteorological centers were involved in this initiative. The compilation of marine data became available for research and applications in the late 1980s (e.g., Woodruff *et al.*, 1987; Bottomley *et al.*, 1990) and hinted that the ocean was warming. These data influenced many of the recent advancements in understanding and modeling of the climate system and its atmospheric and oceanic components.

The compilation of SST data involved formidable efforts to overcome the errors and biases in the raw data. The errors were due to multiple sources. One such source was the uneven distribution of ship observations in time and space (see e.g., Trenberth *et al.*, 1992; Parker *et al.*, 1995). Such non-uniformity affects not only the calculation of regional and global averages, but also increases the uncertainty in the local, monthly mean SST estimates. Today, satellite infrared radiation measurements are processed to derive more evenly spaced weekly estimates of SST that can be blended with ship data to overcome the latter s uneven coverage (Reynolds and Smith, 1994). The impact of uneven sampling of historical SST can be alleviated by sophisticated space-filling procedures (Kaplan *et al.*, 1998). Large biases in SST also resulted from changes in observational practices, such as the time of day when SST was measured and the type of bucket used to draw the seawater samples. Early buckets were porous and cooled off slightly because of evaporation lowering the deck temperature measurement. Later, better buckets were designed to overcome this bias. A large discontinuity in SST data was found around 1941 and attributed to a sudden shift from bucket measurements to temperature readings at the ship engine cooling water intake. This practice and probably adopted in response to wartime restrictions but was widely embraced in the post war era. Bias correction work has been essential for obtaining a self-consistent picture of worldwide SST evolution during the industrial era, one that is consistent with changes in other variables such as the change in air temperature over ocean and land (Parker *et al.*, 1995). It is important to note that SST bias correction and space-filling methodologies continue to evolve and the details of these corrections do have a limited effect on the final estimates of global SST change.

Since 1979, several estimates of the global (and hemispheric) changes in SST since the mid-19th century have been published, together with the concomitant land surface air temperature data (e.g., see summary in IPCC, 1996). The picture emerging from these global SST estimates is that there has been an overall rise in ocean surface temperature that follows most of the details found in the land record. Comparison of the two records shows a slower rise in SST compared to land surface air temperature during the years 1920–1940 as well as in the most recent two decades. These differences are likely the result of real physical processes, such as the horizontally non-uniform change in the atmospheric circulation associated with the global temperature rise (Wallace *et al.*, 1995). There also remains a large difference between the land and ocean records during the last two decades of the 19th Century that seems to be due to data problems.

The importance of the oceans in climate and climate variability is strongly connected to the ability of the oceans to affect SST and the sensitivity of the atmosphere to the resulting SST anomalies. Extensive modeling and observational studies show that in the tropics, particularly in the tropical Pacific, the ocean can be as important in generating SST anomalies as the atmosphere. Tropical SST anomalies also exert a strong impact on the local atmosphere by affecting tropical convection. This in turn produces significant changes in remote locations, including regions outside the tropics. In the extratropical ocean, SST anomalies are produced primarily by concurrent changes in atmospheric conditions. The effect of extratropical SST anomalies on the atmosphere is rather subtle, probably mainly local, and not yet fully resolved or understood.

See also: Earth System Processes, Volume 1; Salinity Patterns in the Ocean, Volume 1.

REFERENCES

Bottomley, M, Folland, C K, Hsiung, J, Newell, R E, and Parker, D E (1990) Global Ocean Surface Temperature Atlas. Meteorological Office Bracknell UK and Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, HMSO, London, 20, 313 plates.

Frankignoul, C (1985) Sea Surface Temperature Anomalies, Planetary Waves and Air–sea Feedback in Midlatitudes, *Rev. Geophys.*, 23, 357–390.

Hartmann, D L (1994) *Global Physical Climatology. International Geophysical Series*, Vol. 56, eds R Dmowska and J R Holton, Academic Press, San Diego, CA, 411.

Henderson-Sellers, A, Zhang, H, Berz, G, Emanuel, K, Gray, W, Landse, C, Holland, G, Lighthill, J, Shieh, S L, Webster, P, and McGuffie, K (1998) Tropical Cyclones and Global Climate Change: a Post-IPCC Assessment, *Bull. Am. Meteorl. Soc.*, 79, 19–38.

IPCC (1996) *Climate Change 1995: The Science of Climate Change*, Cambridge University Press, Cambridge, 572.

Kalnay, E *et al.* (1996) The NCEP/NCAR 40-year Reanalysis Project, *Bull. Am. Meteorl. Soc.*, 77, 437–471.

Kaplan, A, Cane, M A, Kushnir, Y, Clement, A C, Blumenthal, M B, and Rajagopalan, B (1998) Analyses of Global Sea Surface Temperature 1856–1991, *J. Geophys. Res.*, 103, 18 567–18 589.

Latif, M and Barnett, T P (1994) Causes of Decadal Climate Variability over the North Pacific and North America, *Science*, 266, 634–637.

Lau, N C (1997) Interactions between Global SST Anomalies and the Midlatitude Atmospheric Circulation, *Bull. Am. Meteorl. Soc.*, 78, 21–33.

Levitus, S and Boyer, T (1994) *World Ocean Atlas 1994*, Vol. 4, Temperature. NOAA Atlas NESDIS 4, US Department of Commerce, Washington, DC.

Parker, D E, Folland, C K, and Jackson, M (1995) Marine Surface Temperature: Observed Variations and Data Requirements, *Clim. Change*, 31, 559–600.

Peixóto, J P and Oort, A H (1992) *Physics of Climate*, American Institute of Physics, New York, 520.

Reynolds, R W and Smith, T M (1994) Improved Global Sea Surface Temperature Analyses, *J. Clim.*, 7, 929–948.

Timmermann, A, Oberhuber, J, Bacher, A, Esch, M, Latif, M, and Roeckner, M (1999) Increased El Niño Frequency in a Climate Model Forced by Future Greenhouse Warming, *Nature*, 398, 694–697.

Trenberth, K E, Christy, J R, and Hurrell, J W (1992) Monitoring Global Monthly Surface Temperature, *J. Clim.*, 5, 1424–1440.

Wallace, J M, Zhang, Y, and Renwick, J (1995) Dynamic Contribution to Hemispheric Mean Temperature Trends, Science, 270, 780–783.

Woodruff, S D, Slutz, R J, Jenne, R L, and Steurer, P M (1987) A Comprehensive Ocean-atmosphere Data Set, *Bull. Am. Meteorl. Soc.*, 68, 1239–1250.

molecular kinetic energy. Temperature change due to sensible heat transfer for this case is directly proportional to the change in enthalpy.

The temperature change of a body due to sensible heat transfer depends not only on the amount of sensible heat transferred and the associated change in volume of the body but also on the mass of the body, and chemical characteristics of the body described by its specific heat. Energy transfer by electromagnetic radiation can also change the temperature of a body. This is an important factor for environmental situations. Sometimes the colloquial terminology *radiative heat* is used when radiation is involved in the increase of internal energy (temperature) of a body. An example of this occurs when sitting near a hot stove.

Sensible heat transfer between the Earth s surface (ocean and land) and the atmosphere is an important factor in the global energy balance and temperature distribution. The upward sensible heat energy transfer is one-third as large as the latent heat transfer and accounts for 5% of the total upward energy transfer on a global and annual average basis. [Note: radiation accounts for 80% of the transfer of energy from the surface to the atmosphere.]

DAVID HOUGHTON *USA*

Sensible Heat

Sensible heat (sometimes just called heat) is the energy transferred between two material bodies by intermolecular interactions. This transfer occurs when the internal mean molecular kinetic energy (measured by temperature) is different between the two bodies. The transfer goes from the warmer to the cooler body. Sensible heat can be sensed or explicitly recognized, hence the term *sensible*. An example would be the feeling of warmth or coolness to a person when touching a substance that has a temperature higher or lower, respectively, than that of the human body. A warm feeling results when sensible heat is being transferred from the substance to the human body. In the case of sensible heat transferred from the human body, a cold feeling results.

In many environmental situations, warming and cooling associated with sensible heat transfer occur under constant-pressure conditions. In this case, the energy transferred accounts both for changes in the internal mean molecular kinetic energy (temperature) of the body and for the energy of work associated with a change in the volume of the body. This situation can be conveniently handled by the thermodynamic variable *enthalpy*, defined by adding the product of pressure and volume to the internal mean

Snow

David A Robinson
Rutgers University, Piscataway, NJ, USA

The impact of snow on humans and the environment is considerable. Snow covers approximately 30% of the Earth's land surface on a seasonal basis, with additional coverage at high elevations, over polar ice sheets and sea ice. Snow lying on the ground, or on ice in uences hydrologic, biologic, chemical, and geologic processes. Snow exerts an impact on activities as diverse as engineering, agriculture, travel, recreation, commerce and safety. In turn, the presence and state of snow are in uenced by weather, climate, topography, proximity to water bodies and humankind.

INTRODUCTION

The low heat conductivity, high thermal emissivity, low vapor pressure and high reflectance of snow differ greatly from snow-free land. The accurate forecasting of local daily temperatures, regional climatic anomalies and the location and strength of cyclonic systems relies, in part, on knowledge of the distribution and state of regional

snow cover (Walsh and Ross, 1988; Groisman *et al.*, 1994; Clark *et al.*, 1999). Model simulations of climate change show that spatial decreases in snow extent amplify global warming (Meehl and Washington, 1990).

Recent years have seen the availability of more accurate and complete information on the spatial extent and physical state of snow. This is leading to a better understanding of the variability of snowfall and snow cover on annual to decadal scales, of cryosphere–climate interactions, and of the role snow may play in regional and global climate change.

OBSERVING SNOW

Station observations are the primary means of monitoring snowfall, despite the spatial limitations of the global observing network. Data-scarce areas include high latitudes and mountainous regions. Where station density is high, these point measurements provide data of sufficient accuracy for baseline climatologic applications. There is no hemispheric snowfall product derived from station data for either daily or monthly totals.

Ground-based data on snow cover depth are also relatively sparse outside of the lower elevations of the middle latitudes. Included, are station observations made at a single point, and along less abundant, several kilometer long snow courses. Data on the amount of water in a snow pack (water equivalent) are gathered at a very few of the point sites, along snow courses, and by snow pillows that measure snow mass in some mountainous regions.

Visible and passive microwave sensors, on geostationary and polar orbiting satellites, record information used to monitor snow cover on regional to hemispheric scales. Snow cover extent is best identified on visible imagery by recognizing characteristic textured surface features and brightness. Shortcomings include the inability to detect snow cover when solar illumination is low or when skies are cloudy, and the lack of all but the most general information, on pack depth.

The recognition of snow using microwave techniques results from differences in the emissivity of snow-covered and snow-free surfaces across several different frequency ranges. Information on water equivalent can be obtained, although not of the accuracy necessary for climatologic studies. Clouds and low solar illumination are not problems when using microwave data to chart snow cover; however there are difficulties in identifying shallow or wet snow.

SNOW VARIABILITY

The considerable interannual variability of snow makes long-term data sets a necessity when investigating means, trends, low frequency events or interactions of snow with other climatic elements. The standard 30-year climatic normal used for variables such as temperature and precipitation may not be of a sufficient length for realistic cryospheric means. Unfortunately, consistent decades-long data are scarce. Satellite-derived snow maps have been available since the late 1960s. Station data exist for approximately the past century, though of decreasing abundance early on. In all cases, limitations in the collection, quality control, archiving and synthesis of snow data must be considered before applying the data in long-term climatologic investigations.

For the past several decades, weekly maps of Northern Hemisphere snow extent have been produced by US National Oceanic and Atmospheric Administration (NOAA) analysts (Robinson *et al.*, 1993). This constitutes the longest satellite-derived data set of any environmental variable. While produced in a relatively consistent manner since late 1966, original observer inexperience and lower image resolution necessitated a reanalysis of the interval prior to 1972. The earlier period is currently being reanalyzed, but until it is completed, only the time series from 1972 to the present is available for analysis. Microwave records go back to the late 1970s, but studies of time series over this period are just beginning.

According to the NOAA maps, the mean annual Northern Hemisphere snow cover extent is 25.3 million km^2, with 14.7 million km^2 over Eurasia and 10.6 million km^2 over North America (including Greenland). Snow cover was more extensive in the first half of the satellite record than in the past decade (Figure 1). Between 1972 and 1985,

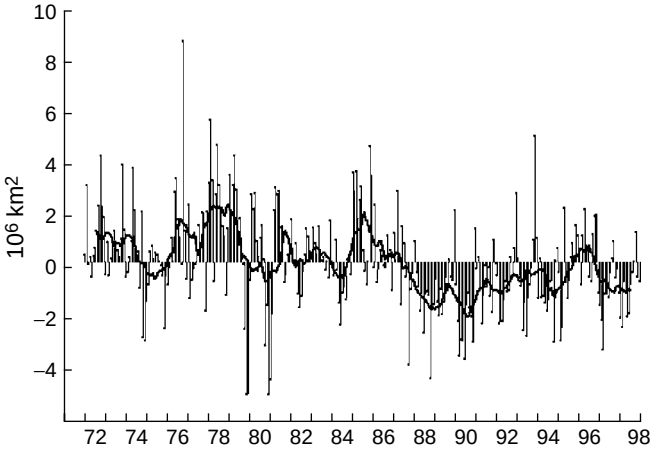


Figure 1 Anomalies of monthly snow cover extent over Northern Hemisphere lands (including Greenland) between January 1972 and December 1998. Also shown are 12-month running anomalies of hemispheric snow extent, plotted on the seventh month of a given interval. Anomalies are calculated from NOAA snow maps. Mean hemispheric snow extent is 25.3 million km^2 for the full period of record. Updates to the time series may be viewed at http://climate.rutgers.edu/snowcover

annual means of snow extent fluctuated around a mean of 25.9 million km^2. An abrupt transition occurred in 1986 and 1987, and since then the mean annual extent has been 24.2 million km^2. Monthly anomalies from the long-term mean are most often less than 3 million km^2, however on occasion they are 4 and 5 million km^2, with October 1976 having a positive anomaly of over 8 million km^2.

Recent decreases in snow extent are large during the spring and summer, while winter and fall extents show no statistically significant change. The tendency toward less late-season cover in recent years begins in February. During seven of the first 15 years of record, February snow extent exceeded the January value. This has occurred only once in the past 12 years.

Station records for portions of the Northern Hemisphere continents suggest that spring and early summer snow extents in the past decade may be at their lowest values of the century. Otherwise, seasonal extents in the latter portion of this century exceed those of earlier years. Annual and decadal fluctuations are embedded in this upward trend, and at least over portions of central North America, the variability of snow cover duration has also increased throughout the century (Frei *et al.*, 1999). The increase in snow cover duration in this region is accompanied by statistically significant increases in seasonal snowfall. In recent decades, snowfall has also been heavier to the lee of the North American Great Lakes than earlier in the century (Leathers and Ellis, 1996). These findings are in line with observations from Canada and the former Soviet Union, with all areas appearing to be part of a trend towards increased precipitation over the middle latitudes lands in the Northern Hemisphere (Brown, 2000).

CONCLUSIONS

An increasing body of evidence suggests that snow plays a critical role in the climate system and will continue to be an important climate variable to monitor when assessing climate change. However, there remain many unanswered questions regarding the variability of snow and the long-term representativeness of recent empirical studies of cryosphere–climate interactions, such as links between snow extent and monsoon strength. Additional information is needed over space and time, including better information over sea ice and ice sheets in both hemispheres. Historic station data sets of snowfall and snow cover remain to be collected and analyzed. Recent efforts to integrate visible and microwave satellite and station data must continue for past intervals and future years. Studies need to examine variations in snow water equivalent, and work begun recently to improve model simulations of snow must continue.

Should increasing levels of greenhouse gases result in warmer conditions over the coming decades, it can be expected that the coverage of snow across continental landmasses will decrease, which in turn should enhance the warming. Snow cover is likely to become established later in the fall, and melt sooner in the spring. Areas where snow cover is ephemeral throughout the winter season will see the ground covered less often. Likewise, it is possible that warming will be accompanied by an increased flux of moisture into the high latitudes, resulting in a deeper winter snow pack in some regions, though probably in a shortened snow season.

See also: Climate Feedbacks, Volume 1; Permafrost, Volume 1; Sea Ice, Volume 1.

REFERENCES

Brown, R (2000) Northern Hemisphere Snow Cover Variability and Change, 1915–1997, *J. Clim.*, 13, 2339–2355.

Clark, M P, Serreze, M C, and Robinson, D A (1999) Atmospheric Controls on Eurasian Snow Extent, *Int. J. Climatol.*, 19, 27–40.

Frei, A, Robinson, D A, and Hughes, M G (1999) North American Snow Extent: 1900–1994, *Int. J. Climatol.*, 19, 1517–1534.

Groisman, P Y, Karl, T R, and Knight, R W (1994) Observed Impact of Snow Cover on the Heat Balance and the Rise of Continental Spring Temperatures, *Science*, 263, 198–200.

Leathers, D J and Ellis, A W (1996) Synoptic Mechanisms Associated with Snowfall Increases to the Lee of Lakes Erie and Ontario, *Int. J. Climatol.*, 16, 1117–1135.

Meehl, G A and Washington, W M (1990) CO_2 Climate Sensitivity and Snow-sea Ice-albedo Parameterization in an Atmospheric GCM Coupled to a Mixed-layer Ocean Model, *Clim. Change*, 16, 283–306.

Robinson, D A, Dewey, K F, and Heim, Jr, R R (1993) Global Snow Cover Monitoring: an Update, *Bull. Am. Meteorol. Soc.*, 74, 1689–1696.

Walsh, J E and Ross, B (1988) Sensitivity of 30-day Dynamical Forecasts to Snow-cover, *J. Clim.*, 1, 739–754.

Soil Moisture

Soil moisture is a measure of the amount of water contained in a soil column (a vertical section of arbitrary areal dimension from the soil surface to the underlying bedrock). Soil moisture is described in terms of the volume of water contained compared with the entire volume of soil. It is also described with reference to the soil s water holding capacity as a ratio of the current volume of water to the maximum possible content if fully saturated. The soil column is fully saturated when water drains readily into or away from the underlying rock. In the absence of replenishing precipitation, this downward drainage to the rock will cease, leaving the voids (capillaries) in the soil still full of water.

In this state, the soil moisture is said to be equal to the *field capacity*: the water content of soil after gravitational free drainage ceases.

Evaporative removal of soil water in the absence of precipitation reduces the soil moisture to below field capacity, causing a soil moisture deficit. Even after direct evaporation from the soil ceases, plants can continue to transpire, further depleting the soil moisture. As the soil water deficit increases, plant transpiration also ceases and the soil moisture is said to have reached the *wilting point*. Although evaporative demand cannot further reduce soil moisture, the soil water content is rarely zero (i.e., oven drying could remove additional water). The environmental limits of soil moisture (or synonymously soil water) are therefore field capacity and wilting point.

Models employed for global change simulations attempt to capture soil moisture characteristics and variability within the constraints of data availability and computational power. The first global soil moisture scheme, developed by Manabe (1969), was a bucket with a maximum soil water capacity equal to an approximated world mean field capacity of 15 cm. This scheme did not, as originally formulated, include representation of the transfer of water from the soil to the atmosphere by plants (i.e., transpiration). More recent schemes attempt to parameterize soil evaporation and plant transpiration by employing multiple soil layers. For example, a two-layer soil scheme transfers water directly from soil to air from an upper thin soil layer by evaporation and also allows plants to access soil water in a deeper soil layer through transpiration. Current formulations include many other embellishments, which have tended to increase the attributes and complexity of soil moisture parameterization schemes through time.

REFERENCE

Manabe, S (1969) Climate and the Ocean Circulation: 1. The Atmospheric Circulation and the Hydrology of the Earth s Surface, *Mon. Weather Rev.*, 97, 739–805.

ANN HENDERSON-SELLERS *Australia*

Solar Irradiance and Climate

Judith Lean
Naval Research Laboratory, Washington, DC, USA

Earth's primary energy source is the Sun. Solar radiation heats the surface of the Earth, which emits infrared radiation that is absorbed by gases in the atmosphere. The greenhouse that results from downward reradiation of this energy further increases the surface temperature so that together the Sun and atmospheric gases permit habitability. Although relatively steady and reliable, the Sun's radiation is not, in fact, constant, and its fluctuations must be considered as possible causes of climate change that may exacerbate or mitigate anthropogenic influences.

SOLAR RADIATION

The Sun radiates energy with a spectrum, shown in Figure 1. The spectrum is similar in shape to that of a black body near 5770 K, the approximate temperature of the Sun s visible surface. Solar irradiance is the radiant power (energy per unit time) incident on unit area of the Earth s surface when it is separated from the Sun by one astronomical unit ($\mathrm{AU} = 1.49 \times 10^{13}$ cm). The total irradiance of the contemporary Sun is, on average, $1366\,\mathrm{W\,m^{-2}}$ with a measurement uncertainty of $\pm 2\,\mathrm{W\,m^{-2}}$.

Solar radiation ranges in wavelength across the entire electromagnetic spectrum, from very short wavelength X-rays, to ultraviolet, visible, infrared and very long wavelength radio waves. Solar spectral irradiance, which is the power per unit area attributable to radiation within a specified wavelength interval, varies with wavelength over five orders of magnitude, with peak values in the visible spectrum. The Sun s irradiance at wavelengths from 0.16 to 5 micron emerges from the vicinity of the Sun s surface in a layer of the solar atmosphere a few hundred kilometers thick called the photosphere.

Most of the Sun s irradiance penetrates the Earth s atmosphere to below 15 km (i.e., into the troposphere). As Figure 1 shows, the spectrum reaching the surface (0 km) comprises radiation mainly in the wavelength region from 0.3–1 micron (μm). Atmospheric gases including water and carbon dioxide control the actual penetration of infrared radiation through the troposphere to the surface. Above the troposphere, other atmospheric gases, primarily ozone, oxygen and nitrogen, absorb the ultraviolet, extreme ultraviolet and X-ray radiations that would have detrimental biological effects were this energy able to penetrate to the surface. This shorter wavelength, higher energy solar radiation provides the primary energy input to the Earth s atmosphere, where it initiates and controls the ozone layer in the stratosphere and the neutral upper atmosphere and its embedded ionosphere at higher altitudes.

SOLAR IRRADIANCE VARIABILITY

Although the Sun s total irradiance has in the past been called the solar constant, in reality the levels of total and spectral irradiance fluctuate continuously. Prominent in the space-based record of total solar irradiance, shown

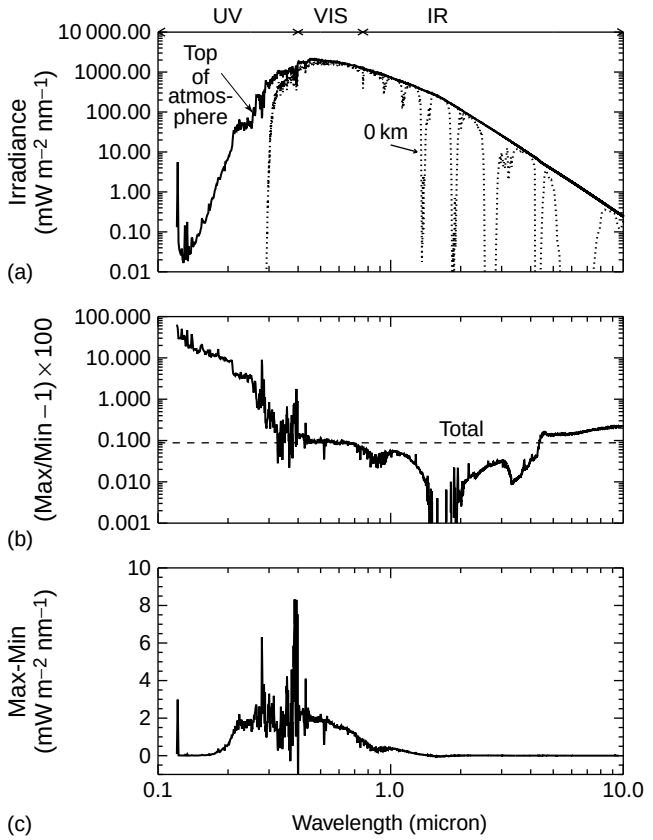


Figure 1 Shown in (a) is the spectrum of the Sun's radiation incident at the top of the Earth's atmosphere (solid line), and at the surface (0 km, dashed line). The broad spectral bands identified along the top of this figure are designated the ultraviolet (UV), visible (VIS) and infrared (IR). Shown in (b) are estimates of the percentage changes in solar spectral irradiance during a recent 11-year activity cycle. At wavelengths longer than 400 nm these estimates are speculative because very few observations exist. The dashed line indicates the percentage variation of total solar irradiance. In (c) the spectrum of irradiance changes during the solar cycle is shown in energy units (maximum minus minimum irradiances)

in Figure 2, is an 11-year cycle upon which are superimposed shorter-term variations (shown in detail in the upper left panel) associated with the Sun s rotation on its axis (approximately once every 27 days). The irradiance cycle is but one manifestation of an 11-year cycle of magnetic activity thought to be driven by a dynamo residing near the base of the Sun s convection zone. During recent maxima of this cycle (near 1980 and 1990), mean levels of total irradiance were observed to be higher by almost 0.1% (1.3 W m^{-2}) relative to levels during cycle minima (near 1986 and 1996), with dips of a few tenths percent often occurring during solar rotation. Wavelength-dependent spectral irradiance changes accompany these total irradiance fluctuations as Figure 1 shows for a recent 11-year cycle (middle and lower panels). Solar cycle increases are of order a few tenths percent at visible and near infrared wavelengths, increasing to 8% in the ultraviolet spectrum near 200 nm and factors of two and more at shorter extreme ultraviolet and longer radio wavelengths. Note that the actual radiant energy falling on the Earth varies by ±3% annually because of regular changes in the distance of the Earth from the Sun. This wavelength-independent fluctuation is a geometrical effect of the Earth s orbit around the Sun. Solar irradiance variability refers to changes intrinsic to the Sun, specified at a fixed Sun–Earth distance of 1 AU.

The primary cause of solar irradiance variability is changes in the activity of the Sun. Solar activity refers to the amount of magnetic flux present in the Sun s atmosphere. Some magnetic fields coalesce to form features that are small and dark, called sunspots, in which the upward flow of energy from the convection zone is inhibited. Sunspots are therefore cooler (by about 1000) and darker (by about 30%), than the surrounding photosphere. Magnetic fields also exist in much less compact features called faculae that are bright rather than dark. Faculae occur both widely dispersed over the solar disk and clumped in contiguous active regions. Sunspots and faculae, evident in the images in Figure 2, are both significant sources of solar irradiance variability. Records of the number of dark sunspots on the solar disk exist since about 1600 and provide the longest direct observation of solar activity. Cycles near 11 years and 27 days are prominent in the sunspot record. As Figure 2 clearly demonstrates, the 11-year cycle in total solar irradiance tracks the annual mean sunspot numbers during the solar activity cycle.

When solar activity is high (such as in 1980 and 1990, and in the left image in Figure 2), many sunspots and faculae are present in the Sun s atmosphere. These magnetic features respectively reduce and enhance solar radiation locally, by different amounts depending on wavelength. Although the local radiation enhancement in faculae (a few percent) is much less than the reduction in sunspots (∼30%), their greater dispersion over the solar surface makes them as important as sunspots in modifying the radiative output of the Sun. Solar cycle irradiance variability reflects the net effect of the these two competing influences, whose numbers and areas evolve continually. Total solar irradiance (Figure 2) is enhanced during solar maximum epochs because the increase in facular brightening during the 11-year activity cycle exceeds (by about a factor of two) the increase in sunspot darkening. Thus, total solar irradiance and sunspot numbers, which provide a generic indicator of solar activity, vary approximately in phase during the 11-year cycle. Solar rotation further modulates irradiance by presenting different populations of these magnetic features to the Earth s view, thereby producing a strong 27-day cycle, evident in Figure 2. On

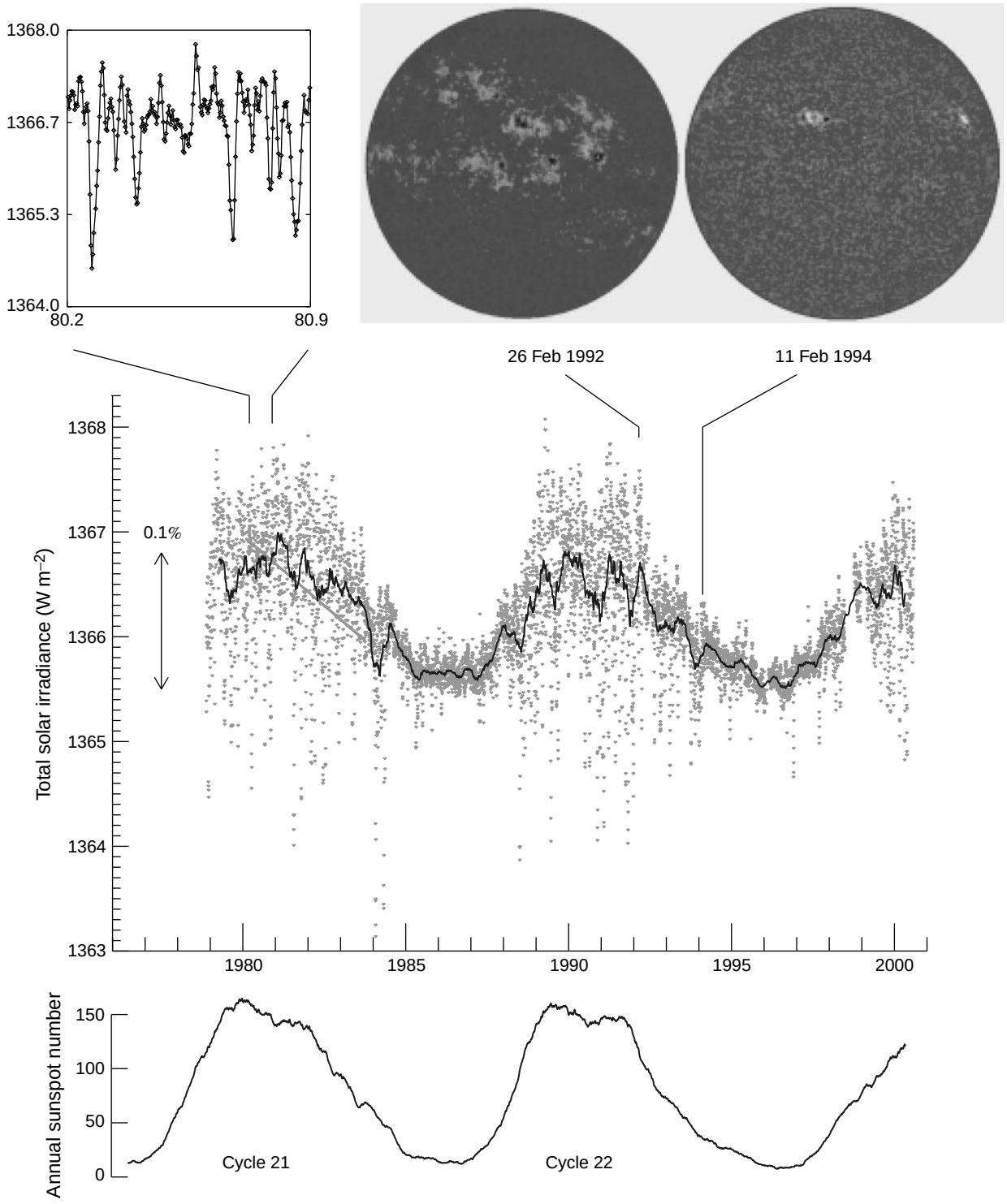


Figure 2 Daily mean values of the Sun's total irradiance since 1979 are shown in the middle panel. Fröhlich and Lean (1998) constructed this composite irradiance record from observations made by four different solar radiometers at various times throughout the epoch. A pronounced 11-year cycle with an approximate peak to peak magnitude of 0.1% is evident. The 11-year total irradiance cycle tracks the general level of solar activity, indicated by the sunspot numbers, in the lower panel. Superimposed on the 11-year cycle are larger weekly to monthly fluctuations of a few tenths percent, shown in detail in the upper left panel, that are associated with the Sun's rotation on its axis, approximately every 27 days. Solar activity and solar rotation irradiance modulation occurs in response to changes in the distribution of bright and dark magnetic features present in the solar atmosphere, as shown in the two solar images in the upper right of the figure. The images are made by the Big Bear Solar Observatory (located at Big Bear Lake, CA) in the Ca K line at 393 nm

these shorter time scales total solar irradiance and sunspot numbers often fluctuate out of phase because sunspot darkening frequently exceeds facular brightening, especially near maximum activity.

Quantitative knowledge – of the locations, brightnesses and areas – of all magnetic features in the solar atmosphere permits the construction of models of solar irradiance variability. Ground-based visible light solar images supply information about areas and locations of sunspots, from which sunspot darkening functions are calculated. Solar images made in other spectral regions, such as the images made at the Ca K Fraunhofer line (393 nm) in Figure 2, record the locations and areas of enhanced facular regions, seen in their chromospheric counterparts (called plage) and in the surrounding bright network. Solar variability models that use wavelength dependent combinations of sunspot and faculae parameterizations can account for much of the variability observed in the total irradiance and the ultraviolet spectrum (there are insufficient observations of the visible and infrared spectral regions to test the models). A model with significant contributions from both sunspot darkening and facular brightening accounts for more than 80% of the variance in the daily mean total irradiance data over nearly two decades, shown in Figure 2.

LONG-TERM IRRADIANCE TRENDS

Direct observations of solar irradiance exist only for the last 20 years, which is a very short period on climatological and solar time scales. However, selected solar, stellar, geomagnetic and geological records contain information about long-term solar activity fluctuations. From these proxy records can be inferred irradiance trends that presumably accompany fluctuating solar activity, not just during 11-year cycles but also on longer time scales as well. Solar activity proxies unanimously suggest that solar activity during the era of space-based solar irradiance observations is higher than in past centuries and that the possibility exists for past and future solar irradiance variability to exceed the range observed thus far.

Sunspots, which provide the only direct observations of historical solar activity, appear frequently during recent 11-year activity cycles, but in the past have disappeared from the Sun s disk entirely for years at a time. Such episodes of solar quiescence, typified by the Dalton, Maunder and Sporer minima from 1790–1820, 1645–1715 and 1425–1575, respectively, are apparent at semi-regular intervals in the Sun s multi-centennial past, and expected in its future. They are evident also in indirect cosmogenic proxies of historical solar activity recorded by $\Delta^{14}C$ in tree-rings and $\Delta^{10}Be$ in ice-cores, whose variations reflect changes in galactic cosmic ray fluxes due to solar activity modulation of the magnetic coupling of the Sun and the Earth (*see* Radionuclides, Cosmogenic, Volume 1). Cosmogenic isotopes exhibit long-term changes that exceed the amplitudes of their 11-year cycles, decreasing from relatively high levels in the recent past, concurrent with increasing sunspot number cycle amplitudes and geomagnetic indices since 1880. Geomagnetic activity recorded by magnetometers at the Earth s surface varies because solar activity influences the solar wind, which modulates the interaction of interplanetary and terrestrial magnetic fields. In suggesting that the Sun s present activity is relatively high compared with the historical range, the evidence from sunspots, cosmogenic isotopes and geomagnetic indices is furthermore consistent with the range of flux levels present in a collection of Sun-like stars.

Solar irradiance is expected to track solar activity on longer time scales, as it does during the 11-year cycle, because solar activity alters the magnetic features that are primary causes of irradiance variability. Solar irradiance has been reconstructed during the past from associations of sunspots and faculae with historical solar activity proxies, using assumptions about additional long-term trends consistent with the range of variability in Sun-like stars and in the cosmogenic isotope and geomagnetic records. The irradiance reconstructions shown in Figure 3 suggest long-term total irradiance changes of order 0.2–0.4%, and associated spectral irradiance changes that are also larger (by about a factor of two) than their solar cycle amplitudes. However, the possibility remains that long-term irradiance variability is confined to 11-year cycles, a scenario that Figure 3 also illustrates in the upper panel. That solar activity during the past 8000 years has ranged from low levels similar to the Maunder Minimum to high levels on a par with the present suggests that even if long-term irradiance variations exist, future levels may be comparable to or less than present values. However, projections of future solar activity are exceedingly difficult for even one 11-year solar cycle, and are essentially impossible for the long-term.

EMPIRICAL SUN–CLIMATE RELATIONSHIPS

An array of empirical relationships implicates the role of the Sun in climate change on Earth. Warmer (colder) climate epochs in the past millennium generally correspond with higher (lower) levels of solar activity, while climate parameters of many types often exhibit cycles that also occur in solar activity proxy records (near 11, 22, 88, 210 and 23 000–25 000 years). Solar variability offers a plausible explanation for significant pre-industrial climate change in the recent Holocene, and possibly also in present day processes.

Of cooler terrestrial epochs in past millennia, the Little Ice Age (1450–1850) is the most recent (*see* Little Ice Age, Volume 1). Surface temperatures then were from

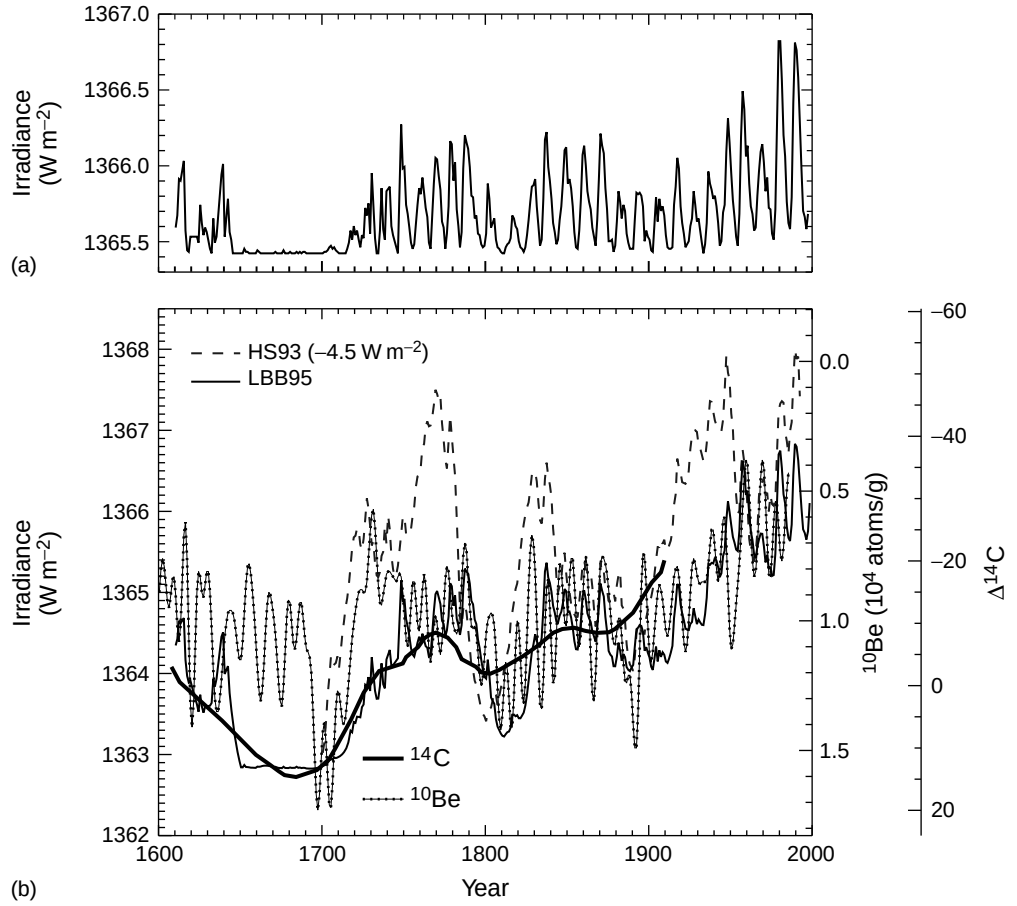


Figure 3 Reconstructions of the Sun's total irradiance, assuming in (a) that the 11-year cycle alone is the only source of long term variability, and in (b) that the 11-year cycle combines with a longer term component. In the LBB95 reconstruction of Lean *et al.* (1995), this long-term component is based on the average amplitude of each sunspot cycle, whereas in the HS93 reconstruction by Hoyt and Schatten (1993), the adopted long-term component is related to sunspot cycle length, rather than amplitude. Also shown for comparison are changes in cosmogenic isotopes (^{14}C and ^{10}Be) that are terrestrial proxies of solar activity

0.6 °C to 1 °C colder than at present (depending on geographical location) and solar activity was lower than at present because of the occurrence of the Maunder and Sporer minima. For almost 200 years, from 1600–1790, surface temperatures tracked solar activity, as the lower panel of Figure 4 shows. The apparent relation of the Little Ice Age and the Maunder Minimum is the most recent example of a general coincidence of increased $\Delta^{14}C$ production (i.e., lower solar activity) and decreased $\delta^{18}O$ (colder temperatures) which may reflect centennial climate change in response to net forcing from combinations of solar periods.

Many climate records (such as temperature, rainfall and drought) exhibit quasi cycles on decadal time scales. Large stochastic climate variability often obscures these cycles, which are not truly deterministic, nor global, nor always present. Although their physical origins remain ambiguous, the commonality of quasi-decadal climate cycles with solar activity (for example near 11, 22 and 88 years) suggests that external solar forcing plays a role. As well, some climate cycles correspond to combinations or harmonics of solar frequencies (7.8, 18, 30, 45 years), as expected for non-linear systems. Comparison of recent annual global surface temperature anomalies and annual total solar irradiance in the upper panel of Figure 4 illustrates the significant decadal variability present in both, compounded in the surface temperature record with shorter term volcanic cooling. The correspondence of higher solar activity with generally enhanced terrestrial warmth also occurs in both the tropospheric and stratospheric regions of the Earth s atmosphere during recent solar cycles. However, decadal Sun–climate correlations appear to be considerably more complicated than simple linear associations, and may involve phase locking by solar forcing of internal atmosphere–ocean variability modes that have distinct regional patterns whose temporal evolution may at times obscure the global signal.

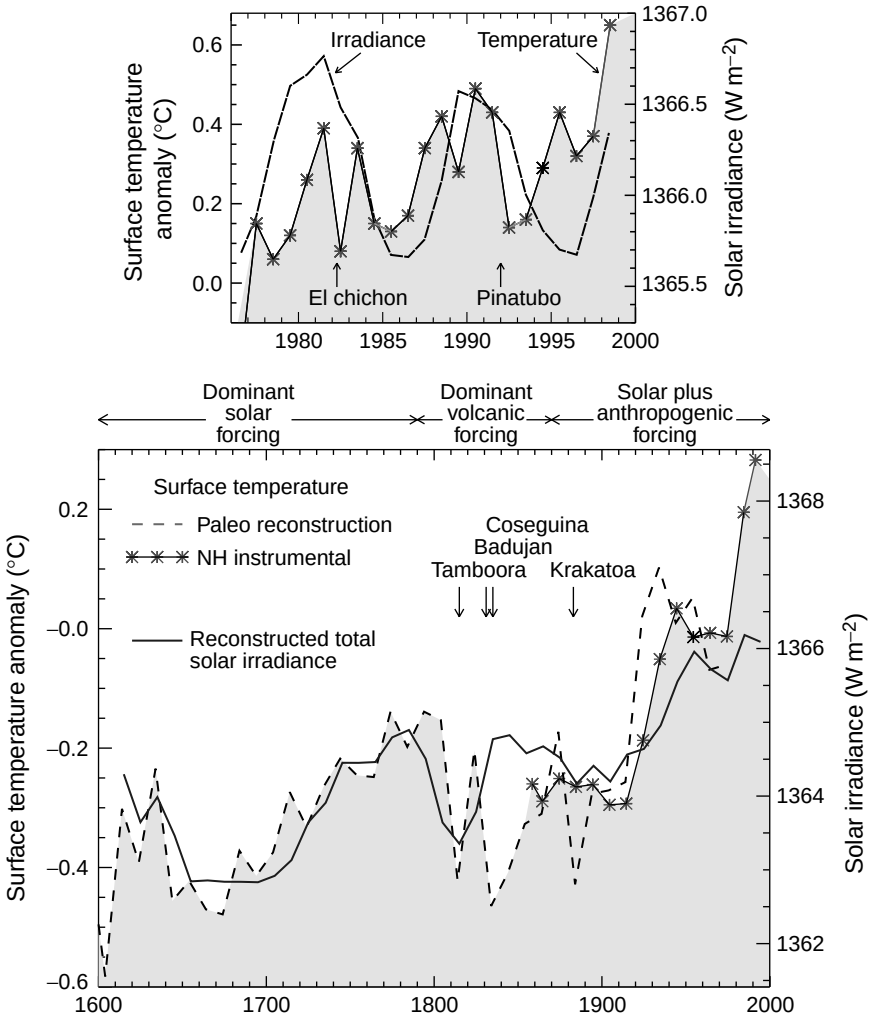


Figure 4 Comparisons are shown in the upper panel of total solar irradiance and annual mean global surface temperatures since 1976 and in the lower panel of a reconstruction of historical total solar irradiance (LBB95 from Figure 3) and surface temperatures since 1600. The symbols show actual surface temperature measurements whereas the solid dashed line is a paleoreconstruction based primarily on tree rings, by Bradley and Jones (1993). Significant volcanic events (Tamboora in 1815, Badujan in 1831, Coseguina in 1835, Krakatoa in 1883, El Chichon in 1982 and Pinatubo in 1991) are clearly evident as surface temperature depressions of a few tenths degree. Nineteenth century volcanism may have prolonged the Little Ice Age beyond that arising from low solar activity, alone

CLIMATE RESPONSE TO SOLAR FORCING

Quantifying the physical processes that facilitate solar-induced climate change is essential for validating and interpreting empirical Sun–climate relations. The balance between incoming solar radiation and outgoing terrestrial radiation determines the equilibrium surface temperature of the Earth. Climate responds to a perturbation in either of these quantities (whether by natural solar variability or anthropogenic changes in greenhouse gas concentrations) by adjusting to a new equilibrium surface temperature. The magnitude of surface temperature adjustment depends on climate sensitivity, which results from a number of feedback processes involving water vapor, sea ice/snow cover, clouds and ocean transport. Climate sensitivity is generally considered to be in the range of $0.2–1\,^{\circ}\mathrm{C}\,(\mathrm{W\ m^{-2}})^{-1}$ and likely depends on the geographical, altitudinal and historical character of the forcing.

Climate models with sensitivities that are generally consistent with paleoclimate evidence simulate global surface temperature changes of about $0.5\,^{\circ}\mathrm{C}$ since 1650, in response to plausible variations in solar irradiance (such as those shown in the lower panel of Figure 3). In the simulations, regional changes can exceed the global mean as a result of dynamical motions driven by differential heating of the land and oceans. Of the $0.5\,^{\circ}\mathrm{C}$ global surface temperature increase, 35% is thought to result from direct surface heating and 65% from feedbacks involving water

vapor (35%) which increases, cloud cover (20%) which also increases, and sea ice/snow cover (10%) which decreases. Figure 5 compares simulated solar-induced surface temperature fluctuations with actual changes determined from instrumental measurements and proxy reconstructions based on tree-rings, ice-cores and corals. Simulated changes of order 0.3 °C during the pre-industrial epoch agree well with reconstructed surface temperatures. However, the simulated 0.25 °C increase from 1900 to 1990 is notably less than the observed 0.6 °C increase. These scenarios are consistent with the empirical results in Figure 4 and suggest that solar irradiance variability contributes to climate change but is not the primary causes of the 20th century warming (*see* Climate Change, Detection and Attribution, Volume 1).

Establishing the contributions to 20th century climate change of solar variability, anthropogenic influences (by greenhouse gases, sulfate aerosols, ozone depletion and albedo changes) and other natural processes (volcanoes and unforced internal oscillations such as the El Niño Southern Oscillation, the Arctic Oscillation and the Quasi Biennial Oscillation) is a challenging task. With the exception of greenhouse gas concentrations, the forcings themselves are uncertain, as is climate sensitivity to each forcing. Also uncertain is the geographical and latitudinal extent of actual climate change. Furthermore, attribution scenarios must account simultaneously for pre-industrial surface temperature changes of about 0.3 °C and post industrial changes in excess of 0.6 °C, by different combinations of forcings. As Figures 4 and 5 illustrate, solar forcing alone permits a satisfactory explanation for the 0.3 °C warming from 1650 to 1790, when changes in greenhouse gas concentrations and aerosols, shown in the upper panel of Figure 5, were small. However, with realistic climate sensitivity, changes in solar forcing and greenhouse gas concentrations together contributed a factor of two or more warming since 1850 than is observed. An attribution scenario mutually consistent in both periods includes cooling by volcanic and sulfate aerosols in addition to warming by solar variability and greenhouse gases. During the past two decades, identification of the Sun s role in long-term climate change is more secure because direct observations of solar irradiance and surface temperature both exist (as Figure 4 shows). That solar radiation levels were approximately equal during two successive solar cycle minima (in 1986 and 1996) suggests negligible long-term solar forcing of climate during a period when measured surface temperatures nevertheless increased 0.2 °C.

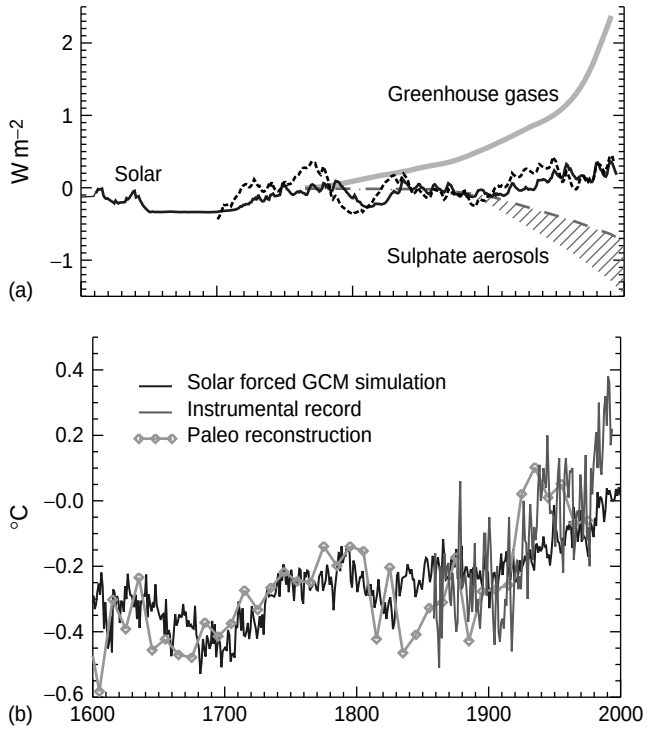


Figure 5 Climate radiative forcing by solar irradiance variability (from Figure 3) is compared in (a) with anthropogenic forcings by changes in concentrations of greenhouse gases (thick solid line) and sulfate aerosols (thick dashed line), including indirect effects (hatching). In (b) are compared annual simulations of surface temperature response to solar forcing by Rind *et al.* (1999) using the Goddard Institute for Space Studies moderate resolution general circulation model, with the annual instrumental record and with the paleoreconstruction by Bradley and Jones (1993) from Figure 4

An upper limit to the surface temperature changes to be expected from solar irradiance variability may be 0.5 °C. Both climate model simulations and empirical Sun–climate relations suggest that the Sun s variability since the Maunder Minimum has contributed about this level of surface warming. It is consistent, as well, with solar-induced climate change inferred from the correlation of $\Delta^{14}C$ (solar activity proxy) and $\delta^{18}O$ (surface temperature proxy) during the past millennium. Solar activity is presently at high levels relative to the historical record of the past 8000 years. This suggests that, were solar activity to undergo another Maunder Minimum type event over the next 200 years, surface cooling would be 0.5 °C or less, as a result; and future long-term solar forcing would either be small, or negative relative to the climate forcing by greenhouse gases, whose concentrations are projected to increase.

Of course, even when long-term solar forcing is minimal, as is apparently the case during the past few decades, decadal climate variability may accrue from prominent 11-year irradiance cycles in ways implied by empirical associations, but not presently understood or accounted for in climate models. Current climate models fail to simulate significant decadal variability in response to 11-year cyclic solar forcing in part because the feedbacks fail to reach their full equilibrium responses, which mutes surface temperature changes. An additional reason may be that the general circulation models do not adequately simulate

natural modes of atmosphere–ocean variability (such as the El Niño Southern Oscillation, the North Atlantic Oscillation and the Pacific Decadal Oscillation) which themselves exhibit decadal variability patterns that may become phase locked with the solar cycle.

Indirect solar-driven processes are expected to influence climate, in addition to direct forcing by solar irradiance variations in the visible and near-infrared spectrum. Most important of these indirect processes are variations in solar ultraviolet irradiance, which varies by an order of magnitude more than total irradiance. The ultraviolet irradiance variations alter ozone concentrations and middle atmosphere temperature, thereby facilitating radiative and dynamical coupling of the middle atmosphere with the biosphere. Changes in solar energetic particles and in galactic cosmic rays, whose penetration to the lower atmosphere depends on solar activity, are also postulated to affect climate indirectly through modification of cloud formation and other processes. Ongoing research is presently exploring such Sun–climate pathways. Improved understanding of all processes by which solar radiation interacts with the Earth is essential for better specifying the impact of solar variability on climate change. Ultimately, a more refined determination of the Sun s role in climate change requires long duration, high precision measurements from space of solar radiation over the entire spectrum, indefinitely.

See also: Solar Variability, Long-term, Volume 1.

REFERENCES

Bradley, R S and Jones, P D (1993) Little Ice Age Summer Temperature Variations: Their Nature and Relevance to Recent Global Warming Trends, *Holocene*, 3, 367–376.

Frohlich, C and Lean, J (1998) The Sun s Total Irradiance: Cycles, Trends and Climate Change Uncertainties Since 1976, *Geophys. Res. Lett.*, 25, 4377–4380.

Hoyt, D V and Schatten, K H (1993) A Discussion of Plausible Solar Irradiance Variations, 1700–1992, *J. Geophys. Res.*, 98, 18 895–18 906.

Lean, J, Beer, J, and Bradley, R (1995) Reconstruction of Solar Irradiance Since 1610: Implications for Climate Change, *Geophys. Res. Lett.*, 22, 3195–3198.

Rind, D, Lean, J, and Healy, R (1999) Simulated Time-Dependent Climate Response to Solar Radiative Forcing Since 1600, *J. Geophys. Res.*, 104, 1973–1990.

FURTHER READING

Cubasch, U, Voss, R, Hegerl, G C, Waszkewitz, J, and Crowley, T J (1997) Simulation of the Influence of Solar Radiation Variations on the Global Climate with an Ocean–Atmosphere General Circulation Model, *Clim. Dyn.*, 13, 757–767.

Foukal, P V (1990) *Solar Astrophysics*, John Wiley & Sons, New York.

Foukal, P and Lean, J (1990) An Empirical Model of Total Solar Irradiance Variations between 1874–1988, *Science*, 247, 556–558.

Lean, J (1997) The Sun s Variable Radiation and its Relevance for Earth, *Ann. Rev. Astron. Astrophys.*, 35, 33–67.

Lean, J and Rind, D (1998) Climate Forcing by Changing Solar Radiation, *J. Clim.*, 11, 3069–3094.

Solar Variability, Long-term

Juerg Beer

Swiss Federal Institute of Environmental Science and Technology (EAWAG), Duebendorf, Switzerland

The Sun is the engine that drives the climate. It emits a total power of 3.9×10^{26} W (luminosity). At the top of the Earth's atmosphere the arriving power still amounts to 1.7×10^{17} W (irradiance or solar constant). From the environmental point of view the most important questions are: how variable is the solar irradiance and how does solar variability affect the Earth System. Today it is well known that the Sun is indeed a variable star as are many other solar type stars. It shows variability on time scales ranging from minutes to billions of years.

The power is generated in the core of the Sun by fusion of hydrogen to helium. It is then transported by radiation and for the last about 30% of its way to the surface by convection. The standard solar model reveals that, after the formation of the solar system 4.5 billion years ago, the luminosity was only 75% of its present value. Since then it has been steadily growing and will continue to do so for another 4- 5 billion years.

Variability on time scales shorter than 100 000 years is produced within the convection zone by the complex interplay between circulating plasma, magnetic elds and the differential rotation of the Sun (Lang, 1995).

DIRECT PARAMETERS OF SOLAR VARIABILITY

We can distinguish between direct and indirect data of solar variability. Direct data are continuously or regularly recorded from ground based as well as from space borne instruments. In general their quality is high, but the time interval they cover is limited to a few decades or centuries.

The main direct parameters of solar variability are listed in Table 1. All of them have a temporal resolution of at least 1 year.

Table 1 Direct parameters of long-term solar variability

Solar variability	Records since
Sunspots	~1600
Aurorae	~1700
aa-index	1843
Neutrons	1937
Solar irradiance	1978

During the last several decades the number of direct observational parameters has increased significantly (http://www.spaceweather.com/, http://umbra.nascom.nasa.gov:80/sdac.html, http://www.ngdc.noaa.gov/stp/stp.html, http://www.sunspotcycle.com/).

Historically the sunspot number is the most famous measure of solar variability. However, only after the discovery of the 11-year cycle by Schwabe in 1843 was a scientific observational program set up leading to a continuous record from about 1600 AD to the present (Figure 1). Sunspots occur on sites where magnetic flux tubes cross the solar surface. Because the strong magnetic fields hinder the upward movement of hot plasma, the local temperature is lower by some 1000 K compared to the surroundings (Hoyt and Schatten, 1997).

Aurorae are produced by solar plasma (solar wind) streaming out of coronal holes. The interaction of this solar wind with the upper part of the Earth s atmosphere leads to ionization processes, which can be seen as spectacular colored phenomena at high latitudes in both hemispheres.

The solar wind not only causes ionization in the atmosphere, it also disturbs the geomagnetic field. The aa-index is a measure for this disturbance. The earliest measurements go back to the time when the magnetometer was invented (Figure 1) (Mayaud, 1973).

The neutron flux is the result of the interaction of cosmic radiation (mainly protons and helium nuclei) with the atmosphere. The cosmic ray flux is modulated by the solar wind and the geomagnetic field. During periods of a very active Sun, the strong solar wind shields the cosmic ray flux and therefore reduces the neutron production. The first instruments monitoring the cosmic ray flux became operational in 1937.

Finally, after a long period of unsuccessful attempts to detect changes in solar irradiance (solar constant) from the Earth s surface, satellite based radiometers launched in 1978 showed a positive correlation between solar irradiance and solar activity (Lean, 1997). Although the changes observed during an 11-year Schwabe cycle are small (0.15%), they raised the fundamental question whether larger changes occur on longer time scales that could affect significantly the global climate.

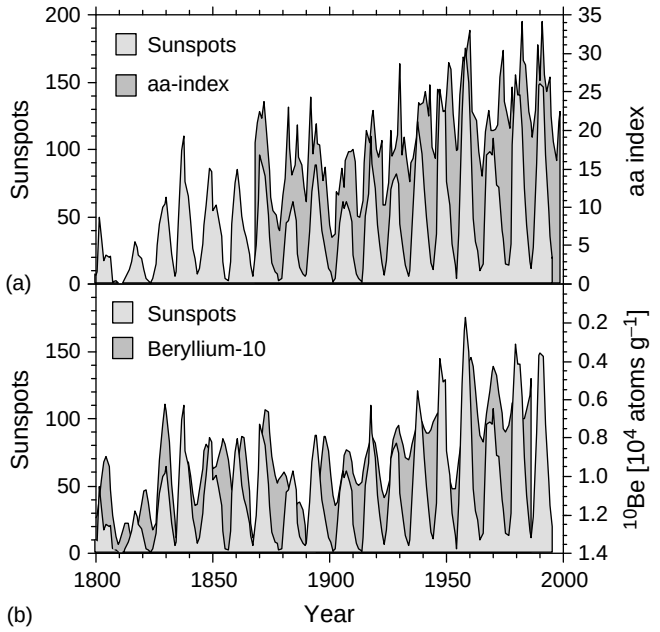


Figure 1 Comparison of two direct parameters of solar variability: (a) the aa-index with the sunspot number, and (b) the indirect parameter ^{10}Be with the sunspot number. Note that the ^{10}Be data (number of atoms per gram of ice) are plotted on an inverse scale

INDIRECT PARAMETERS OF SOLAR VARIABILITY

All parameters discussed so far are observational and therefore restricted in time. Observations in the pretelescopic era before 1600 AD are sparse and often not very reliable. Indirect parameters, however, have the potential to extend the variability record of the Sun back much further in time. The most promising tools for this purpose are cosmogenic radionuclides such as ^{10}Be and ^{14}C (*see* Radionuclides, Cosmogenic, Volume 1). They are produced in the atmosphere by the interaction of cosmic ray particles (mainly neutrons) with nitrogen and oxygen. As a result of the shielding effect of the solar wind on the cosmic rays, the production rate of these nuclides depends on the solar activity (Masarik and Beer, 1999).

In order to reconstruct the past, solar variability changes in the production rate of radionuclides must be recorded in an appropriate archive. Nature provides such archives in the form of ice sheets for ^{10}Be and tree rings for ^{14}C. Ice sheets contain the precipitation of more than 100 000 years. After its production in the atmosphere ^{10}Be becomes attached to aerosols and is removed within 1–2 years, mainly by precipitation. ^{14}C on the other hand forms CO_2 and exchanges then occur between the atmosphere, biosphere, and ocean. Tree rings therefore reflect the $^{14}C/^{12}C$ ratio at the time when they are formed. As a consequence of the ^{14}C exchange between the reservoirs, its system behavior is

different from that of ^{10}Be and has to be taken into account using a carbon cycle model.

Since radionuclide records in natural archives reflect the superposition of solar and geomagnetic effects on the production rate and changes within the system, the interpretation is not always straightforward. System effects can be separated from the production signal by comparing ^{10}Be with ^{14}C variations making use of the fact that variations in the two nuclides are very similar regarding production, but completely different regarding system effects.

LONG-TERM SOLAR VARIABILITY DERIVED FROM COSMOGENIC RADIONUCLIDES

Cosmogenic radionuclides provide long-term information on cyclic and non-cyclic solar variability. The well known 11-year Schwabe cycle was found for the first time in the ^{10}Be record of an ice core from Greenland (Beer *et al.*, 1990) (Figure 1) and extended beyond the sunspot record back to 1423. The 88-year Gleissberg and the 207-year De Vries cycle are also present in ^{10}Be ice core records (Beer *et al.*, 1994). The De Vries cycle is the most prominent solar periodicity in the radiocarbon tree ring record (Stuiver and Braziunas, 1993). Shorter cycles are dampened by the large sizes of the carbon reservoirs and are therefore very difficult to detect. Longer cycles (e.g., around 2000 years) do exist, but cannot yet be attributed unambiguously to solar variability.

Especially interesting features are periods of low solar activity such as the Maunder minimum (*see* Maunder Minimum, Volume 1), which is characterized by the almost complete absence of sunspots. The detrended Δ^{14}C record (Stuiver *et al.*, 1998) shows that several of these minimum periods occurred (Figure 2) during the last 7000 years. The fact that these periods often coincide with a change from warmer to colder climates indicates that solar forcing plays an important role in climate change (Beer *et al.*, 2000). However, the underlying mechanisms are still subject to investigation.

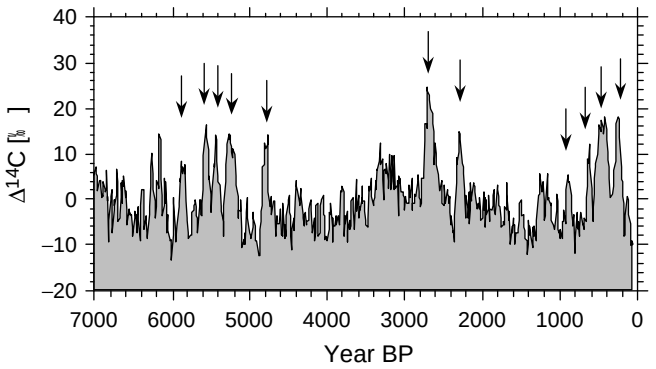


Figure 2 Δ^{14}C (deviation of the atmospheric $^{14}C/^{12}C$ ratio from a standard in permil) for the last 7000 years (BP, before 1950) after removing the long-term trend. The main periods of low solar activity are marked by arrows. The last arrow on the right hand side corresponds to the Maunder minimum (1645–1715 AD)

See also: Earth System History, Volume 1; Solar Irradiance and Climate, Volume 1; Sunspots, Volume 1.

REFERENCES

Beer, J, Blinov, A, Bonani, G, Finkel, R C, Hofmann, H J, Lehmann, B, Oeschger, H, Sigg, A, Schwander, J, Staffelbach, T, Stauffer, B, Suter, M, and Wolfli, W (1990) Use of ^{10}Be in Polar Ice to Trace the 11-year Cycle of Solar Activity, *Nature*, 347, 164–166.

Beer, J, Joos, C F, Lukasczyk, C, Mende, W, Siegenthaler, U, Stellmacher, R, and Suter, M (1994) ^{10}Be as an Indicator of Solar Variability and Climate, in *The Solar Engine and its In uence on Terrestrial Atmosphere and Climate*, ed E Nesme-Ribes, Springer-Verlag, Berlin, 221–233, Vol. I.25.

Beer, J, Mende, W, and Stellmacher, R (2000) The Role of the Sun in Climate Forcing, *Quat. Sci. Rev.*, 19, 403–415.

Hoyt, D V and Schatten, K H (1997) *The Role of the Sun in Climate Change*, Oxford University Press, New York.

Lang, K R (1995) *Sun, Earth, and Sky*, Springer, Berlin Heidelberg.

Lean, J (1997) The Sun s Variable Radiation and its Relevance for Earth, *Annu. Rev. Astron. Astrophys.*, 35, 33–67.

Masarik, J and Beer, J (1999) Simulation of Particle Fluxes and Cosmogenic Nuclide Production in the Earth s Atmosphere, *J. Geophys. Res.*, 104, 12 099–13 012.

Mayaud, P N (1973) A Hundred Year Series of Geomagnetic Data 1868–1967, *Bull. Int. Assoc. Geomagn. Aeron.*, 33, 1–251.

Stuiver, M and Braziunas, T F (1993) Sun, Ocean, Climate and Atmospheric $^{14}CO_2$: an Evaluation of Causal and Spectral Relationships, *Holocene*, 3, 289–305.

Stuiver, M, Reimer, P J, Bard, E, Beck, J W, Burr, G S, Hughen, K A, Kromer, B, McCormac, G, Van der Plicht, J, and Spurk, M (1998) INTCAL98 Radiocarbon Age Calibration, 24 000–0 cal BP, *Radiocarbon*, 40, 1041–1083.

Southern Ocean

John A Church

Antarctic CRC and CSIRO Marine Research, Tasmania, Australia

The Southern Ocean plays a critical role in decadal and centennial climate variability and in determining the timing and regional impact of anthropogenic climate change. Its circumpolar extent allows inter-basin exchange and thus adds a global dimension to the ocean's vertical overturning

circulation that dominates ocean heat transport and storage. Over 50% of the global ocean's water mass properties are determined through interaction with the atmosphere and cryosphere in the Southern Ocean region. These waters carry their heat, freshwater, dissolved oxygen, nutrient and carbon dioxide content to lower latitudes and into the ocean interior.

Projections of 21st century climate change indicate that surface warming will be a minimum in the Southern Ocean. The Southern Ocean absorbs additional heat from the enhanced greenhouse effect that is exported to lower latitudes, resulting in sea-level rise through ocean thermal expansion. Formation and subduction of Antarctic water masses result in the Southern Ocean being the largest long-term sink for anthropogenic carbon dioxide. These water masses are transported into the ocean interior, thus preventing further exchange of carbon dioxide with the atmosphere for decades or longer.

Observations suggest that changes in the properties of water masses formed north of the Antarctic circumpolar current (ACC) and subducted into the subtropical oceans may have occurred in recent decades. These changes include a warming, a freshening, a decrease in dissolved oxygen and an increase in dissolved inorganic carbon concentrations. Some climate models also project a decrease in the meridional overturning of the Southern Ocean. However, the relevant processes are not well represented in the current generation of models and con rmation of this decrease awaits further observations and analysis.

The Southern Ocean, with an area of $7.6 \times 10^{13}\,\mathrm{m}^2$ south of $40\,^{\circ}\mathrm{S}$, is the only circumpolar ocean and forms a zonal connection between the southern parts of the Pacific, Atlantic and Indian Oceans. The major topographic features and horizontal transports are shown in Figure 1 and a schematic of the vertical overturning circulation is shown in Figure 2. The eastward flowing *Antarctic Circumpolar Current* (ACC) is constrained by choke points south of America, Australia/New Zealand and Africa. The ACC extends over the full depth of the Southern Ocean and has an estimated transport of about $134 \times 10^6\,\mathrm{m}^3\,\mathrm{s}^{-1}$ through Drake Passage and about $145 \times 10^6\,\mathrm{m}^3\,\mathrm{s}^{-1}$ south of Australia (Rintoul *et al.*, 2001). The surface layer south of the ACC is freshened by an excess of precipitation over evaporation and the melting of sea ice. This surface water is advected northward across the ACC in the surface Ekman transport (Figure 2).

On the northern side of the ACC, heat loss during winter results in surface mixed layers several hundred meters deep and the formation of sub-Antarctic mode water (SAMW). From there, SAMW s distinctive temperature/salinity characteristics, low vertical density gradient and high oxygen values can be tracked to lower latitudes as it is advected into the thermocline of the subtropical gyres of the Southern Hemisphere oceans. The coldest and densest form of SAMW, Antarctic intermediate water (AAIW), is formed in the South–East Pacific.

Several studies have shown significant changes over recent decades of SAMW and AAIW properties in the South Pacific and southern Indian Oceans (Dickson *et al.*, 2001). The changes are consistent with a warming and a freshening in the formation regions. Wong *et al.* (1999) argue that the freshening of AAIW, together with an increased salinity of water masses formed at lower latitudes, is consistent with increased precipitation at high latitudes and an intensification of the hydrological cycle as projected in climate-change simulations. Banks *et al.* (2000) see similar changes in global climate model results and argue that this climate change signal is most easily detected in the southern Indian Ocean. These same water mass formation and circulation processes are responsible for a significant uptake of anthropogenic carbon (C) and its storage in the ocean (Wallace, 2001). From analysis of model results, Cai and Whetton (2000) argue that the advection of warmed SAMW into the Pacific equatorial band is likely to result in greater warming in the eastern rather than the western Pacific Ocean during the 21st century (i.e., more El Niño-like conditions). The subduction of heat into the ocean with warmed SAMW is an important component of ocean thermal expansion and steric sea-level rise (Church *et al.*, 2001).

Relatively warm ($2\,^{\circ}\mathrm{C}$) salty water upwells south of the ACC and is transported onto the Antarctic shelf, where it loses heat to the atmosphere and ice shelves. Its density is also increased through the addition of salt rejected during sea-ice formation. The resultant dense Antarctic bottom water (AABW) cascades down the slope and flows northward into the ocean interior. The main formation site for AABW is the Weddell Sea, with lesser amounts being formed around the Antarctic coastline, particularly the Adelie Coast and the Ross Sea (Rintoul *et al.*, 2001). A number of studies (e.g., Comiso and Gordon, 1998; Jacobs and Giulivi, 1998) have documented changes in AABW and high salinity shelf water properties in the Weddell and Ross Seas, but the reasons for these changes are unclear. Farther north in the South Atlantic, observed changes in deep-water properties are attributed to variability in open-ocean convection in the Weddell Sea (Coles *et al.*, 1996; Zenk and Hogg, 1996).

Anthropogenic climate change has the potential to result in a reduction in the area of the Southern Ocean covered by sea ice from its present values of about $3 \times 10^6\,\mathrm{km}^2$ in summer and about $19 \times 10^6\,\mathrm{km}^2$ in late winter. Such a change would reduce the albedo of the Southern Ocean (a positive feedback), freshen the surface ocean and allow greater air sea exchange of heat, freshwater and carbon dioxide (CO_2). In a controversial result, de la Mare (1997)

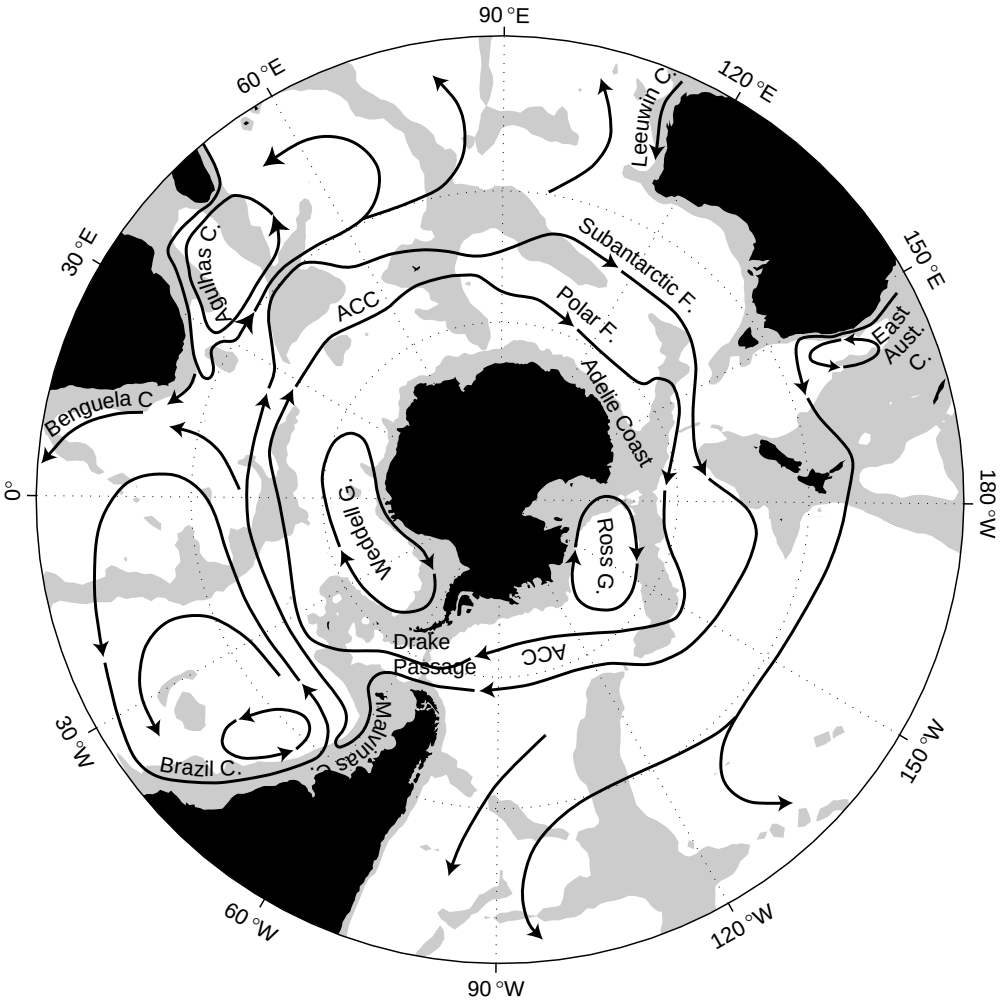


Figure 1 Schematic map of the major current systems. Depths shallower than 2500 m are shaded. The two major cores of the ACC are shown, the Subantarctic front (SAF) and the polar front (PF) (Abbreviations used are F for front, C for current and G for gyre). (Reproduced by permission from Academic Press in Rintoul *et al.*, 2001)

argues, from records of the location of whaling vessels close to the ice edge, that there was a significant retreat in the edge of the sea-ice zone during the central decades of the 20th century. While there has been a 20% decline in sea ice extent in the Amundsen and Bellingshausen Seas during the two decades from 1973 (Jacobs and Comiso, 1997), there appears to have been a slight increase in the area of the Southern Ocean covered by sea ice since modern satellite records commenced in the early 1970s (Cavalieri *et al.*, 1997).

Some recent coarse-resolution climate models project an increased stability of the Southern Ocean as a result of surface warming and freshening. This leads to a reduction in the rate of formation of AABW and a decrease of downwelling adjacent to the Antarctic coast by 20% by 2000, with a virtual shutting down of ventilation of waters below 2000 m after a further 300 years (Hirst, 1998). The slowdown in the overturning circulation, together with surface warming, may reduce the oceanic uptake of carbon dioxide (Matear and Hirst, 1999) and lead to a reduction in ocean oxygen levels, similar to observed reductions south of Australia (Matear *et al.*, 2000). Further changes to ocean uptake of carbon dioxide as a result of biological response to the changed physical conditions are uncertain but potentially important (Sarmiento and Le Quéré, 1996). The observational estimate of the rate of formation of AABW comes from ocean chlorofluorocarbon (CFC) inventories (Rintoul *et al.*, 2001). Broecker *et al.* (1998, 1999) use CFC, nutrient and $\triangle^{14}$C distributions to infer that there has been a recent reduction in the formation of Antarctic deep water. However, an analysis of these changes in the formation of deep water masses are confounded by insufficient observations, differing definitions of deep water masses and a lack of confidence that global models adequately represent the small-scale processes responsible for the formation of AABW and other deep water masses.

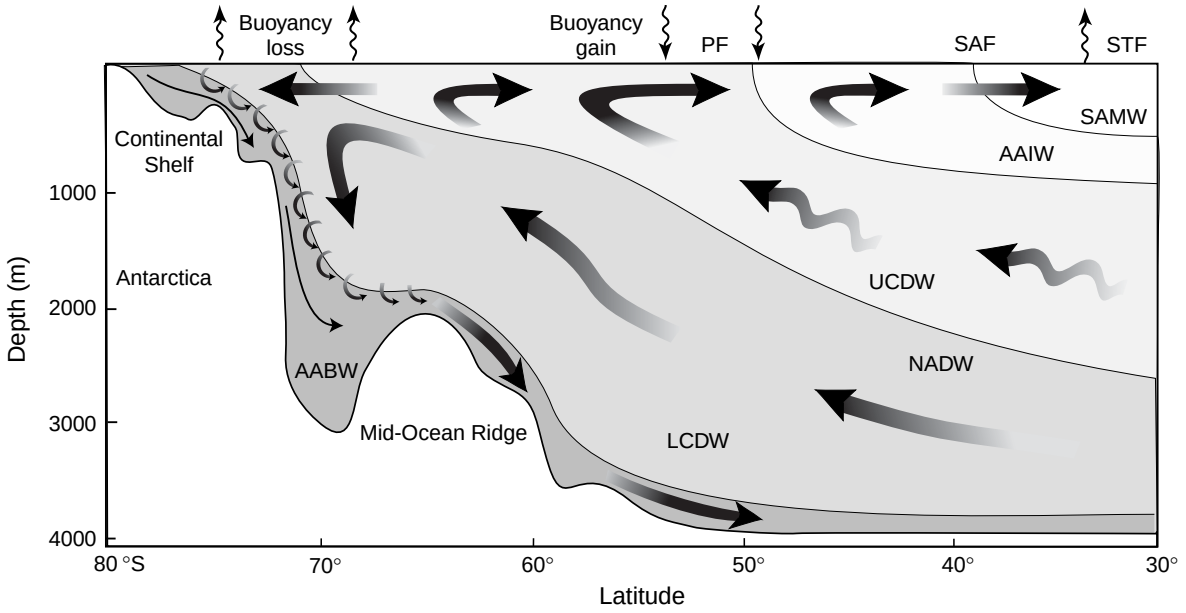


Figure 2 A schematic view of the meridional overturning circulation in the Southern Ocean. Abbreviations used are SAMW, AAIW, UCDW (upper circumpolar deep water), NADW (North Atlantic deep water), LCDW (lower circumpolar deep water), AABW, STF (subtropical front), SAF, PF. (Reproduced by permission from American Meteorological Society in Speer *et al.*, 2000)

Changes in the Southern Ocean have global as well as local significance. Increased confidence in potential changes requires ongoing observations, numerical modeling and analysis.

See also: Antarctica, Volume 1; Ocean Circulation, Volume 1; Salinity Patterns in the Ocean, Volume 1; Sea Ice, Volume 1.

REFERENCES

Banks, H T, Wood, R A, Gregory, J M, Johns, T C, and Jones, G S (2000) Are Observed Decadal Changes in Intermediate Water Masses a Signature of Anthropogenic Climate Change?, *Geophys. Res. Lett.*, 27, 2961–2964.

Broecker, W S, Peacock, S L, Walker, S, Weiss, R, Fahrbach, E, Schroeder, M, Mikolajewicz, U, Heinze, C, Key, R, Peng, T-H, and Rubin, S (1998) How Much Deep Water is Formed in the Southern Ocean?, *J. Geophys. Res.*, 103, 15 833–15 844.

Broecker, W S, Sutherland, S, and Peng, T-H (1999) A Possible 20th Century Slowdown of Southern Ocean Deep Water Formation, *Science*, 286, 1132–1135.

Cai, W and Whetton, P H (2000) Evidence for a Time-Varying Pattern of Greenhouse Warming in the Pacific Ocean, *Geophys. Res. Lett.*, 27, 2577–2580.

Cavalieri, D J, Gloersen, P, Parkinson, C L, Comiso, J C, and Zwally, H J (1997) Observed Hemispheric Asymmetry in Global Sea Ice Changes, *Science*, 278, 1104–1107.

Church, J A, Gregory, J M, Huybrechts, P, Kuhn, M, Lambeck, K, Nhuan, M T, Qin, D, and Woodworth, P L (2001) Changes in Sea Level, in *Climate Change 2001: The Scienti c Basis. Contribution of Working Group 1 to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, eds J T Houghton, Y Ding, D J Griggs, M Noguer, P van der Linden, X Dai, K Maskell, and C I Johnson, Cambridge University Press, Cambridge, in press.

Coles, V J, McCartney, M S, Olson, D B, and Smethie, W M (1996) Changes in Antarctic Bottom Water properties in the Western South Atlantic in the 1980s, *J. Geophys. Res.*, 101, 8957–8970.

Comiso, J C and Gordon, A L (1998) Interannual Variability in Summer Sea Ice Minimum, Coastal Polynyas and Bottom Water Formation in the Weddell Sea. In Antarctic Sea Ice: Physical Processes, Interactions and Variability, ed M O Jeffries, *Antarct. Res. Ser.*, 74, 293–315.

de la Mare, W (1997) Abrupt mid-20th Century Decline in Antarctic Sea-Ice Extent From Whaling Records, *Nature*, 389, 57–60.

Dickson, B, Hurrell, J, Bindoff, N, Wong, A, Arbic, B, Owens, B, Imawaki, S, and Yashayaev, I (2001) The World during WOCE, in *Ocean Circulation and Climate*, eds G Siedler, J Church, and W J Gould, Academic Press, London.

Hirst, A C (1998) The Southern Ocean Response to Global Warming in the CSIRO Coupled Ocean-Atmosphere Model, *Environ. Model. Software*, 14(Special Issue), 227–241.

Jacobs, S S and Comiso, J C (1997) Climate Variability in the Amundsen and Bellingshausen Seas, *J. Clim.*, 10, 697–709.

Jacobs, S S and Giulivi, C F (1998) Interannual Ocean and Sea Ice Variability in the Ross Sea. In Ocean, Ice and Atmosphere: Interactions at the Antarctic Continental Margin, eds S S Jacobs and R F Weiss, *Antarct. Res. Ser.*, 75, 135–150.

Matear, R J and Hirst, A C (1999) Climate Change Feedback on the Future Oceanic CO_2 Uptake, *Tellus*, 51B, 722–733.

Matear, R J, Hirst, A C, and McNeil, B I (2000) Changes in Dissolved Oxygen in the Southern Ocean with Climate Change, *Geochem., Geophys., GeoSyst.*, 1.

Rintoul, S R, Hughes, C, and Olbers, D (2001) The Antarctic Circumpolar Current System, in *Ocean Circulation and Climate*, eds G Siedler, J Church, and W J Gould, Academic Press, London.

Sarmiento, J L and Le Quéré, C (1996) Oceanic Carbon Dioxide Uptake in a Model of Cenury-Scale Global Warming, *Science*, 274, 1346–1350.

Speer, K G, Rintoul, S R, and Sloyan, B M (2000) The Diabatic Deacon Cell, *J. Phys. Oceanogr.*, 30, 3212–3222.

Wallace, D W R (2001) Storage and Transport of Excess CO_2 in the Oceans: the JGOFS/WOCE Global CO_2 Survey, in *Ocean Circulation and Climate*, eds G Siedler, J Church, and W J Gould, Academic Press, London.

Wong, A P S, Bindoff, N L, and Church, J A (1999) Large Scale Freshening of Intermediate Waters in the Pacific and Indian Oceans, *Nature*, 400, 440–443.

Zenk, W and Hogg, N (1996) Warming Trend in Antarctic Bottom Water flowing into the Brazil Basin, *Deep-Sea Res.*, 43, 1461–1473.

Southern Oscillation

The *Southern Oscillation* is a year-to-year seesaw in low latitude surface atmospheric pressure between the eastern and western hemispheres. This pressure oscillation is intimately related to the climatic conditions known as El Niño and La Niña. The British meteorologist Sir Gilbert Walker first defined the southern oscillation in the early 20th century.

One phase of the Southern Oscillation occurs when surface pressure is unusually high over the tropical Indian Ocean region and the western tropical Pacific, and unusually low east of the international date line in the southeastern tropical Pacific. During these periods, the pressure force across the Pacific basin that drives the trade winds is reduced. The trade winds therefore weaken, resulting in the abnormally high sea surface temperatures that characterize El Niño. Conversely, when the surface pressure is unusually high east of the date line and low west of the date line, the trade winds are unusually strong and cold sea surface temperatures develop in association with La Niña.

A commonly used indicator for the Southern Oscillation, the Southern Oscillation Index, is based on departures of surface atmospheric pressure from normal at Tahiti, French Polynesia and Darwin, Northern Australia. When this index is negative (low pressure at Tahiti relative to Darwin), the trade winds are weak (El Niño conditions). When the index is positive, the trade winds are strong (La Niña conditions).

Collectively, El Niño, La Niña, and the Southern Oscillation are referred to as El Niño/Southern Oscillation (ENSO). These phenomena are intimately related and arise from coupled interactions between the atmosphere and the ocean in the tropical Pacific. Shifts in tropical rainfall patterns associated with ENSO result in anomalous tropospheric heating that alters the global atmospheric circulation, producing droughts, floods, unusual storminess, heat waves and other weather extremes in many regions of the globe. The social, economic, and public health consequences of ENSO-related climate variability have motivated international efforts since the mid-1980s to better understand, monitor, and predict the evolution of the ENSO cycle (*see* El Niño/Southern Oscillation (ENSO), Volume 1).

MICHAEL J MCPHADEN *USA*

SPARC (Stratospheric Processes and their Role in Climate)

SPARC is a research project established in 1992 by the Joint Scientific Committee of the World Climate Research Programme (WCRP) to investigate the role played by the stratosphere in the climate system. While intensive research was carried out on stratospheric ozone depletion, there were important areas that were not receiving sufficient attention, notably those concerning the role of the stratosphere in changing the climate and the feedback of climate change on the stratosphere. The mandate of SPARC was to stimulate research in those areas that require its attention.

SPARC deals with the role of stratospheric dynamical, chemical and radiative processes in the global climate of the troposphere–stratosphere system. The key scientific issues were identified by the SPARC Scientific Steering Group in 1992 as follows: the influence of the stratosphere on climate, the physics and chemistry associated with stratospheric ozone decrease, stratospheric variability and its monitoring, and changes in ultraviolet radiation.

Since 1992, SPARC s activities have been structured around the three following issues: stratospheric processes and their relation to climate, stratospheric indicators of climate change, and modeling stratospheric effects on climate. Results from SPARC s activities have been used in *Scienti c Assessment of Ozone Depletion, 1998*, trends in temperature and changes in the ozone vertical profile were assessed. The results of the water vapor assessment have been made available for the Intergovernmental Panel on Climate Change s Third Assessment Report (2001).

SPARC interacts with the other WCRP projects concerned with different aspects of the climate system, acting as the expert group on the atmosphere in the vicinity of the tropopause and above. Some of the research interests of SPARC are closely related to those of the International Global Atmospheric Chemistry project of the International Geosphere–Biosphere Programme, although with a different focus.

Contact: sparc.office@aerov.jussieu.fr.

MARIE-LISE CHANIN *France*

Storm Surge

Orrin H Pilkey[1] and David M Bush[2]
[1]*Duke University, Durham, NC, USA*
[2]*State University of West Georgia, GA, USA*

One of the deadliest aspects of hurricanes and other coastal storms is the storm surge - the local and temporary rise in water level above the normal astronomical tide. Historically, 90% of hurricane-related deaths have been attributed to the storm surge, such as the 1900 hurricane that killed at least 6000 in Galveston, TX. In the US today, storm surge is no longer such a threat because of watches and warnings, evacuation, and sheltering. Worldwide, however, storm surge is still a devastating killer. The storm surge associated with great hurricanes can range from 4 to 7 m or greater.

The US National Hurricane Center (NHC) defines storm surge as an abnormal rise in sea level accompanying a hurricane or other intense storm, and whose height is the difference between the observed level of the sea surface and the level that would have occurred in the absence of the cyclone. Storm surge is usually estimated by subtracting the normal or astronomic high tide from the observed storm tide. The NHC defines storm tide as the actual level of sea water resulting from the astronomic tide combined with the storm surge. Storm tide is poor terminology because of the confusion with astronomical tides. Another common term, and perhaps less ambiguous, is simply storm high water level.

Storm surge impacts include flooding, floating structures off of their foundations, and floating debris inland. The initial flow over and around obstructions (e.g., pilings) may cause scouring and sediment transport. The rising water also elevates waves and increases their landward incursion resulting in a wider zone of potential destructive impact. Waves combined with storm surge act to wash beach and dune sand landward forming washover deposits. Associated currents can transport sediment offshore. Saltwater flooding and spray kills or damages inland plants.

All coastal storms can cause storm surge. Hurricanes usually get most of the attention because of the potential for very high storm surge owing to their fierce winds and rapid forward movement, but they typically have only a localized impact (*see* Hurricanes, Typhoons and other Tropical Storms - Descriptive Overview, Volume 1; Hurricanes, Typhoons and other Tropical Storms - Dynamics and Intensity, Volume 1). Extratropical cyclones (winter storms, northeasters, southwesters) usually have markedly lower storm surge than hurricanes but they may affect much longer stretches of coastline.

CONTROLS ON STORM SURGE

The major components of storm surge can be considered as: (1) those relating to the storm physical characteristics; (2) those relating to the movement of the storm; and (3) those relating to the characteristics of the shoreline of impact. In addition, timing of hurricane landfall with astronomical tides can raise or lower the storm high-water level.

Storm physical characteristics include wind speed (faster means higher potential storm surge), wave height (nearshore, larger breaking waves mean more water transported shoreward thus higher storm surge), and storm size (radius of maximum winds for hurricanes, overall size for extratropical storms – the larger the storm the longer stretch of shoreline will be impacted by surge). Also important is the barometric effect, i.e., the air pressure of the storm system relative to surrounding normal atmospheric pressure. In general, a 1-mb drop in air pressure translates into about a 1-cm rise in sea level. Overall, the barometric effect is not the major contributor to storm surge. A Category five hurricane (the strongest on the Saffir–Simpson Hurricane Scale) can have a central pressure of less than 920 mb, compared to about 1013 mb for normal atmospheric pressure, that means that even for the strongest storms only about 1 m of the total storm surge can be due to atmospheric pressure differences. Also of major importance is the rotational effect of the water column created by the inward-spiraling winds. As the rotating water column beneath the hurricane moves into shallower water, the water is forced to rise upward to conserve vorticity – to keep the same volume of water rotating. That is, the water column now limited by depth must rise, further increasing the storm surge. Hurricanes are very fast moving compared to most winter storms so the duration of hurricane winds is not a major controlling factor in hurricane storm surge. Winter storms, however, may have surge levels that continue to build somewhat as the storm persists for several days.

Movement of the storm for extratropical storms means how long the storm persists to impact the coast; for hurricanes, the more important factors are how fast the storm center is approaching the coast (30 mph (48 km h^{-1}) or faster increases potential storm surge by at least one category on the Saffir–Simpson Scale) and what the angle of approach to the shoreline is (perpendicular approach means higher potential storm surge).

The shape of the shoreline of impact can be considered in profile and in plan view. In profile, storm surge is augmented by shelf width just as tidal amplitude is. On narrow shelves there is little shallow water to be impacted by the storm frictional processes. Narrow shelves such as those off the Caribbean Islands, Cape Hatteras, NC, or Miami Beach, FL, have inherently lower maximum potential surge elevations. On wide shelves, a broad expanse of shallow water can be piled up as a storm surge. Storm surges are potentially higher along the Gulf Coast and southern Atlantic Coast of the US between Cape Hatteras, NC and Cape Canaveral, FL. In plan view, a concave shoreline shape tends to funnel surge water. The 1993 storm of the century northeaster blew straight onshore in the concave Apalachee Bay area of Florida, creating storm surges of nearly 30 ft (9 m). Straight and especially convex shorelines have the opposite effect.

Under the right conditions, large estuaries can have profound storm surges. Hurricane Emily s (1993) winds blew offshore over Pamlico Sound, NC, causing maximum storm surge on the back side of Hatteras Island. Surge elevations were amplified by the concavity of the sound side shoreline of the Cape Hatteras cuspate foreland.

STORM CURRENTS

Storm surges are also responsible for bottom currents, which are set up because of the local rise in sea level creating a downwelling pressure gradient current. These types of currents are not very strong, but they can be responsible for moving great quantities of sediment away from the beach and into deeper water. Once sediment is resuspended by storm waves, only a slight current is needed to transport sediment offshore and possibly permanently out of the beach/dune system.

STORM-SURGE EBB

Storm-surge ebb is an important process where the storm surge water flows back to sea, either by the force of gravity alone or driven by offshore-blowing winds, generating an erosive ebb current. This type of current is different from downwelling pressure gradient currents because it occurs while the storm is moving out of the area or diminishing. Storm-surge ebb can cause an existing inlet to change shape, breach a spit or island creating a new inlet, scour shallow cross-island channels, transport storm debris (including houses) offshore, and cause permanent removal of sand from the beach/dune system to the deeper offshore. After Hurricane Hugo, the shoreface in front of Myrtle Beach, SC was covered with deep scour tracks, perpendicular to the shoreline, formed by the storm-surge ebb.

DEFINING THE STORM SURGE RISK IN US

A computer simulation model developed for the US National Weather Service called SLOSH (sea, lake and overland surges from hurricanes) is used in the US to predict the storm surge. The SLOSH model is essential for developing hurricane evacuation plans in exposed coastal areas. Such indundation information is available from state or local emergency management offices, the US Federal Emergency Management Agency (FEMA), the US National Weather Service, as well as local governments. Another measure of storm surge is relative to the 100-year flood. FEMA publishes maps called flood insurance rate maps (FIRMs) showing various flood zones for the entire US, in order to identify flood hazard areas (riverine as well as coastal). The highest-hazard coastal areas, according to the FIRMs, are A-zones (100-year flood zone; a statistical 1% chance of flood reaching or exceeding a predetermined area in any given year), and V-zones (portions of the 100-year flood zones subject to waves 3 ft (0.91 m) or greater). For further information on SLOSH modeling and hurricane processes, visit the US National Oceanographic and Atmospheric Administration (NOAA) web site: http://www.noaa.gov. For further information on FIRMs and related hazard management issues, visit the US FEMA web site: http://www.fema.gov.

See also: Tides, Oceanic, Volume 1; Tsunamis, Causes and Consequences, Volume 1.

Stratosphere

The *stratosphere* is the layer of the Earth s atmosphere from about 11 km to about 50 km in which the temperature is isothermal or increases with height. The temperature structure of the stratosphere is a result of atmospheric ozone, which absorbs ultraviolet solar radiation and warms the air. The stratospheric ozone concentrations are at a maximum near 25 km, but the temperature increases with height to about 50 km because the ozone concentrations above the ozone maximum are sufficient to effectively intercept ultraviolet solar radiation, warming these levels, and allowing less ultraviolet radiation to penetrate to the

lower levels, where it could be absorbed by the higher concentrations of ozone. The lower air densities at higher altitudes also contribute to the absorbed radiation yielding higher temperatures. The peak in temperature defines the top of the stratosphere, known as the *stratopause*, above which the temperature decreases with height in the mesosphere.

See also: Mesosphere, Volume 1; Tropopause, Volume 1; Troposphere, Volume 1.

KEITH L SEITTER *USA*

Stratosphere, Chemistry

Douglas Kinnison
National Center for Atmospheric Research, Boulder, CO, USA

In this overview of stratospheric chemistry, the important chemical processes are outlined from a historical point of view, starting with the discovery of molecular oxygen and continuing with a detailed focus on the chemical processes that affect both the distribution and trend of stratospheric ozone. The increasing realization that the stratosphere is an important and interesting component of the Earth system has gone hand in hand with: our capability to monitor the stratosphere (stimulated in part by the International Geophysical Year (IGY) eld programs in the 1950s); our need to know more about the stratosphere for operational aviation reasons, e.g., the high- ying supersonic transports (SSTs) that began to emerge in the late 1960s; and our growing realization that the stratosphere was a major player in a number of global environmental issues, including the stratospheric ozone hole over Antarctica and greenhouse gas-induced climate warming.

*Air generally enters the stratosphere in the equatorial region, carrying upward chemical constituents from the troposphere to the lower stratosphere. In the troposphere vertical mixing is on the order of hours to days, while in the stratosphere, vertical mixing tends to require months to years. Chemical constituents that have a relatively long chemical lifetime in the troposphere will be transported into the stratosphere. These species (e.g., methane (CH_4), nitrous oxide (N_2O), and chloro uorocarbons (CFCs)) are typically designated as long-lived and the photochemical destruction of these species within the stratosphere produces short-lived radicals that are important in stratospheric ozone (O_3) balance (*see *Lifetime (of a Gas), Volume 1; Residence Time (of an Atom, Molecule or Particle), Volume 1; Stratosphere, Temperature and Circulation, Volume 1).*

*Stratospheric ozone absorbs both solar and infrared radiation, which impacts the thermal structure of the atmosphere. These radiative processes help to determine the circulation of the global atmosphere. O_3 is also an important greenhouse gas. The radiative balance in the mid-to-upper troposphere and lowermost stratosphere is quite sensitive to the O_3 concentration, so changes in O_3 concentration in these regions are very important. In addition, because O_3 absorbs potentially harmful solar ultraviolet radiation, any depletion of stratospheric O_3 increases the amount of solar ultraviolet radiation reaching the Earth's lower atmosphere and the surface. Increased ultraviolet radiation is known to affect many chemical and biological processes and to increase the rate of skin cancers in humans and animals (*see *Depletion of Stratospheric Ozone, Volume 1; Ultraviolet Radiation, Volume 1).*

EARLY HISTORY OF GAS-PHASE STRATOSPHERIC CHEMISTRY

The stratosphere is composed primarily of molecular nitrogen (N_2) and molecular oxygen (O_2), 78.084% and 20.946% respectively. In the stratosphere, N_2 is unreactive; however, O_2 can be appreciably photolyzed, producing atomic oxygen (O) down to levels as low as 20 km. Both Priestly and Schelle independently discovered O_2 in 1774. As discussed below, photolysis of O_2 is responsible for production of stratospheric ozone (O_3). The development of the field of stratospheric chemistry is closely tied to understanding of the chemical processes that affect stratospheric O_3 abundance.

Knowledge of the physical properties of O_3 was greatly enhanced during the 19th century. O_3 was discovered and named in 1839 by Schonbein (from the Greek word *ozein*, to smell). In 1864, Soret presumed that O_3 had a structure made up of three-oxygen atoms (O_3). Cornu (1878) observed that there was a sudden break off of the solar spectrum at wavelengths shorter than 300 nm. Hartley (1880–1881) discovered a strong absorption band between 200 nm and 320 nm. Experimental and observational work led Hartley to conclude that the absorption feature noted by Cornu was due to O_3 ultraviolet absorption in the middle atmosphere. Chappuis (1880) discovered eleven absorption bands of O_3 contained within the visible spectrum between 500 nm and 700 nm. Finally, Huggins (1890) observed ultraviolet absorption bands between 320 nm and 360 nm in the spectrum of the star Sirius and concluded that this absorption was also due to O_3. The absorption by O_3 of solar radiation in the visible has a significant influence on the dynamics of the middle atmosphere. In addition, the absorption by O_3 in the ultraviolet region is necessary for

life to exist on Earth (*see* Atmospheric Composition, Past, Volume 1; Earth System History, Volume 1; Solar Irradiance and Climate, Volume 1; Stratosphere, Temperature and Circulation, Volume 1; Ultraviolet Radiation, Volume 1).

The chemical processes that influence stratospheric O_3 abundance have been examined and updated throughout the 20th century, with most of the significant advances in theoretical understanding coming in the last three decades. In 1913, the first quantitative measurement of O_3 in the atmosphere was made by Fabry and Buisson. Dobson (1926) developed an ultraviolet spectrometer that measured the total column of O_3 routinely. Total column O_3 is the integrated O_3 measured in a vertical column from the surface to the top of the atmosphere and is often expressed in Dobson units (DU), where 1 DU is equivalent to a column of O_3 of 10^{-5} meters in height at standard temperature and pressure (STP) (1 atm pressure and 273.15 K). Because approximately 90% of atmospheric O_3 is located within the stratosphere (the other 10% is located within the troposphere), perturbations to stratospheric O_3 will significantly impact the total column O_3 abundance. In 1929, Dobson established a surface network of stations at various latitudes. The present day O_3 monitoring network still includes many of these same stations (*see* Stratosphere, Ozone Trends, Volume 1).

In 1930 Chapman proposed that photochemical processes produced O_3, specifically by the photolysis of O_2 (Equation 1), producing two oxygen atoms (*see* Photochemical Reactions, Volume 1).

$$O_2 + h\nu \longrightarrow O + O \quad (1)$$

$$\underline{O + O_2 + M \longrightarrow O_3 + M} \text{ (twice)} \quad (2)$$

$$\text{NET: } 3O_2 + h\nu \longrightarrow 2O_3$$

Subsequent photolysis of O_3 (Equation 3) was not considered a net loss, because the oxygen atom produced would rapidly react in a three-body reaction to produce O_3 again (Equation 2). The third body reactant (i.e., M) is typically either N_2 or O_2. In many reactions, this chemical constituent is necessary to remove excess energy for the stabilization of the O_3 bond. Because Equations (2) and (3) do not result in a loss of O_3, this is typically called a null cycle.

$$O_3 + h\nu \longrightarrow O + O_2 \quad (3)$$

$$\underline{O + O_2 + M \longrightarrow O_3 + M} \quad (2)$$

NET: Null Cycle

Chapman proposed two reactions that would reduce stratospheric O_3 abundances. The first reaction is between an oxygen atom and an O_3 molecule (Equation 4). This is the primary stratospheric loss reaction for O_3 in the odd-oxygen (i.e., $O + O_3$) family of chemical constituents. We use the term odd-oxygen because any oxygen atom produced reacts rapidly with an oxygen molecule (Equation 2) to form O_3, therefore, a loss of oxygen atom is essentially a loss of O_3 in the stratosphere. A schematic of the Chapman mechanism is shown in Figure 1. In the scientific literature, the expressions, odd-oxygen loss and O_3 loss are used interchangeably. In this article we will use O_3 loss .

$$O + O_3 \longrightarrow 2O_2 \quad (4)$$

$$O + O + M \longrightarrow O_2 + M \quad (5)$$

Laboratory measurements indicated that Chapman s reaction (Equation 5) was too slow to impact O_3 abundance in the stratosphere, although this reaction becomes more important above the stratopause, within the mesosphere.

In 1936, Wulf and Deming were the first to calculate the O_3 profile using a photochemical model. They incorporated the Chapman mechanism into their photochemical model and used a solar intensity that corresponded to a black body temperature of 6000 K, typical of solar output. Using measured photochemical rate constants for that period, they were able to calculate a vertical O_3 profile. Three years earlier, Gotz (1934) developed the Umkehr method to measure the O_3 vertical profile. This method used spectra taken from scattered radiation from the overhead sky through sunrise and sunset. The results of the Umkehr method showed O_3 to have a vertical profile that peaked approximately between 22–25 km in the Northern Hemisphere, mid-latitudes. Considering the uncertainties in the photochemical rate constants and solar irradiance for that period, the calculated versus observed O_3 vertical profiles were in good agreement. However, in the 1950s, U–2 rockets were used for the first time to directly measure the solar intensities throughout visible and ultraviolet wavelengths. It was found that the solar intensity could not be represented by a blackbody temperature profile of 6000 K. Correcting with this discrepancy and making additional changes in

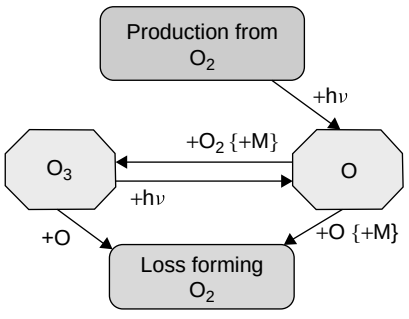


Figure 1 Schematic diagram of the principal reactions making up the Chapman mechanism. The rounded rectangular boxes represent production and loss processes of odd-oxygen species. The octagonal boxes represent the O_3 and O odd-oxygen radicals. The arrows show the possible reaction pathways

photochemical rate constants, model-derived vertical profiles of O_3 were found to be more than the observed values.

During the 1960s, scientific research was focused on finding additional chemical processes that would bring theory into alignment with measurements. McGrath and Norrish (1960) experimentally showed that the hydroxyl radical could be produced through reactions of water (H_2O) vapor with electronically excited singlet atomic oxygen or $O(^1D)$. The initiation step for this process occurs when O_3 is photolyzed at wavelengths less than 310 nm, forming $O(^1D)$ and O_2 (Equation 6). The subsequent $O(^1D)$ reaction with H_2O produces two hydroxyl (OH) radicals (Equation 7).

$$O_3 + h\nu(\lambda < 310\,\text{nm}) \longrightarrow O(^1D) + O_2 \quad (6)$$

$$O(^1D) + H_2O \longrightarrow 2OH \quad (7)$$

The OH radical is then available to react with O_3, producing the hydroperoxyl (HO_2) radical and O_2 (Equation 8). Because the concentration of OH is several orders of magnitude less than O_3, Equation (8) will have negligible effects on O_3 abundance. However, the HO_2 radical produced by Equation (8) can react with another O_3 to reproduce the OH radical. This chain process will continue thousands of times before it is interrupted, and it is therefore termed a catalytic cycle. The net effect of Equations (8) and (9) is the loss of two O_3 molecules, forming three O_2 molecules. This catalytic cycle is important in the lower stratosphere.

$$OH + O_3 \longrightarrow HO_2 + O_2 \quad (8)$$

$$\underline{HO_2 + O_3 \longrightarrow OH + 2O_2} \quad (9)$$

$$\text{NET: } 2O_3 \longrightarrow 3O_2$$

Figure 2 shows the schematic of the hydrogen oxide (HO_x) family. As in Figure 1, the octagons represent species that interconvert rapidly with each other (i.e., H, OH, and HO_2). Species highlighted within rectangles are considered temporary reservoir species. For the simple schematic shown in Figure 2, H_2O_2 would be a temporary reservoir species. A reservoir species by definition is a species that removes the active radicals from the O_3 destruction cycle(s).

In 1966, Hunt modeled the atmosphere using photochemical reactions that corresponded to the Chapman mechanism plus the water reactions. He found that, by adjusting the values for the rate constants of reactions (8) and (9), a calculated vertical O_3 profile could be derived that matched the observed profile. However, when more accurate values for the rate constants in the above mechanism became available, the theoretical value of both the O_3 profile and total column O_3 still exceeded observational data.

In 1968, Murcray and colleagues measured the concentration of nitric acid (HNO_3) in the stratosphere to be in the parts per billion by volume (ppbv) range. Expanding

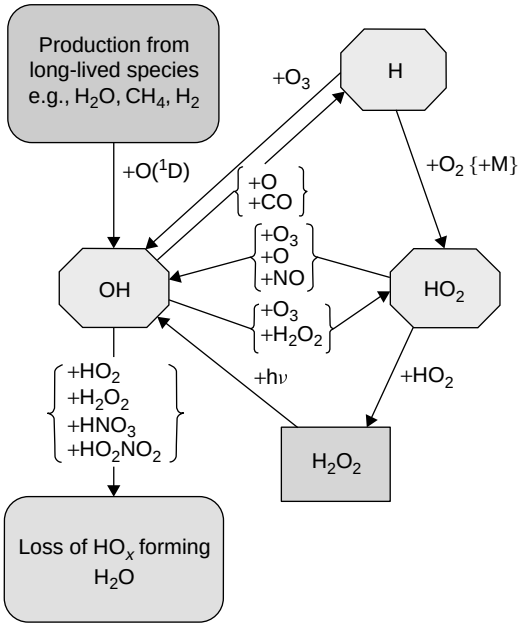


Figure 2 Principal reactions for the HO_x chemical family. The rounded rectangular boxes represent production and loss processes of odd-hydrogen species. The octagonal boxes represent HO_x radicals that have a short atmospheric lifetime (i.e., >1 day for H, OH, and HO_2). The species highlighted within the rectangular box is considered a temporary reservoir species (i.e., H_2O_2)

on this, Crutzen (1970) reasoned that the observed HNO_3 concentration implied enough nitric oxide (NO) and nitrogen dioxide (NO_2) to support a catalytic cycle that would reduce O_3 in a manner similar to the hydrogen oxide catalytic cycle (*see* Crutzen, Paul J, Volume 1).

$$NO + O_3 \longrightarrow NO_2 + O_2 \quad (10)$$

$$\underline{NO_2 + O \longrightarrow NO + O_2} \quad (11)$$

$$\text{NET: } O + O_3 \longrightarrow 2O_2$$

In 1971, Crutzen and Nicolet independently proposed the primary source of stratospheric nitrogen oxides (i.e., $NO_x = NO + NO_2$) to be reaction of nitrous oxide (N_2O) with $O(^1D)$:

$$N_2O + O(^1D) \longrightarrow 2NO \quad (12)$$

See Figure 3 for a schematic of the NO_x family of reactions. N_2O is primarily produced and subsequently released from the surface by bacterial nitrification and denitrification in soils and aquatic systems. N_2O is not appreciably destroyed in the troposphere, but is removed in the stratosphere by both photolysis and by the above-mentioned reaction with $O(^1D)$. The NO_x production in the stratosphere is a prime example of how, through modification to the biospheric nitrogen cycle, life on Earth controls natural levels of stratospheric O_3.

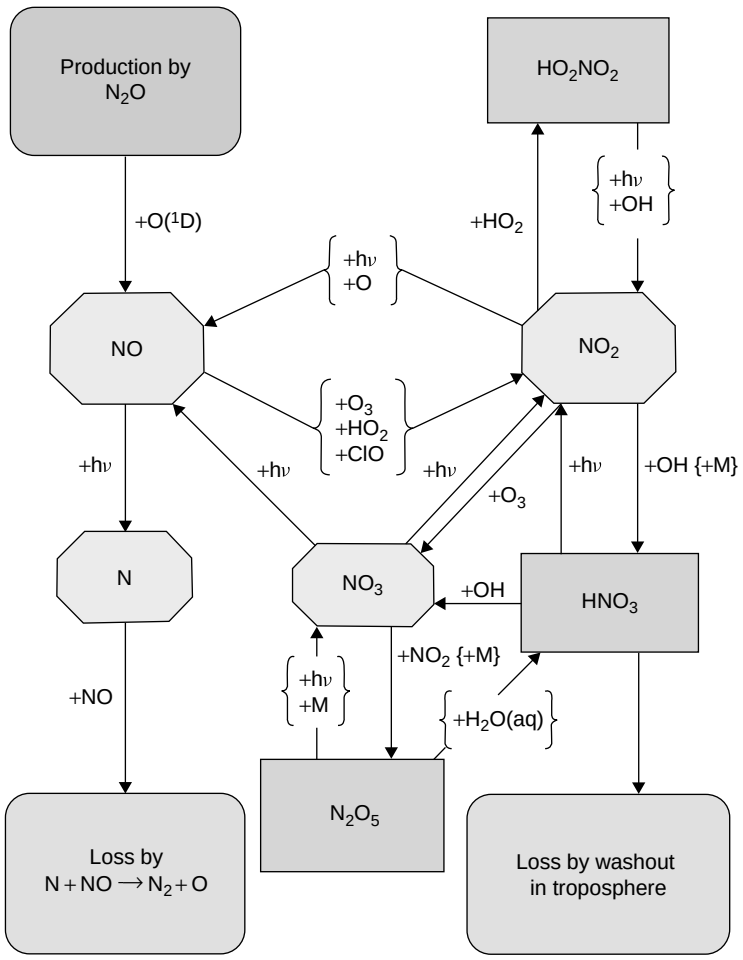


Figure 3 Principal reactions for the NO_x chemical family. The rounded rectangular boxes represent production and loss processes of odd-nitrogen species. Species contained within octagonal boxes represent NO_x radicals with short atmospheric lifetimes (i.e., N, NO, NO_2, and NO_3). Species highlighted within rectangular boxes are considered temporary reservoir species (i.e., N_2O_5, HO_2NO_2, and HNO_3)

An additional source of NO_x in the stratosphere is believed to be from galactic cosmic ray (GCR) production of NO. These processes occur through secondary electrons ejected by heavy cosmic particles. This production process is not insignificant, especially in the polar regions, where N_2O oxidation is slow. NO_x could also be produced in the lower stratosphere by a fleet of SSTs. Johnston (1971) clarified the role that a proposed fleet of SSTs would have on stratospheric NO_x and subsequent changes in O_3. NO_x is formed from the breakdown of molecular nitrogen from the high combustion temperatures contained within the jet engines. NO_x increases in the lower stratosphere from a fleet of proposed SSTs could increase ambient NO_x by approximately 50%. To investigate this potential environmental issue, the US Department of Transportation created the Climatic Impact Assessment Program (CIAP). This program was responsible for not only increasing understanding of SST impacts on stratospheric O_3, but also, potentially more important, for sponsored research that greatly increased understanding of physical, dynamical, and chemical processes related to stratospheric O_3.

NO_x is eventually removed in the upper stratosphere and lower mesosphere by reaction of atomic nitrogen (N) with NO, forming chemically inert molecular nitrogen (N_2).

$$NO + N \longrightarrow N_2 + O \tag{13}$$

NO_x can also be removed by the formation of HNO_3 (Equation 14). Because HNO_3 does not readily regenerate NO_x, it is considered a temporary reservoir species for NO_x. In addition, HNO_3 is water-soluble and any HNO_3 transported into the troposphere can be precipitated out of the atmosphere (*see* Nitrogen Cycle, Volume 2). Therefore, Equation (14) effectively short-circuits the NO_x catalytic cycle (Equations 10 and 11).

$$NO_2 + OH + M \longrightarrow HNO_3 + M \tag{14}$$

Shortly after the importance of NO_x was proposed by Crutzen and Johnston, an additional catalytic cycle based on chlorine oxides (ClO_x) (i.e., $ClO_x = Cl + ClO$) was also proposed. Stolarski and Cicerone (1974) realized that chlorine, if in high enough concentrations, could be important in controlling stratospheric O_3 abundance. They proposed a catalytic cycle similar to the NO_x cycle proposed by Crutzen.

$$Cl + O_3 \longrightarrow ClO + O_2 \quad (15)$$

$$\underline{ClO + O \longrightarrow Cl + O_2} \quad (16)$$

$$\text{NET: } O + O_3 \longrightarrow 2O_2$$

The source of ClO_x radicals in the stratosphere was identified by Molina and Rowland (1974) (*see* Molina, Mario J, Volume 1; Rowland, F Sherwood, Volume 1). See Figure 4 for a schematic of the ClO_x family of reactions. They suggested that CFCs were insoluble and relatively unreactive in the lower atmosphere and could be transported into the stratosphere (*see* Chloro uorocarbons (CFCs), Volume 1; Halocarbons, Volume 1). It is known that CFCs and other long-lived molecules that are heavier than air can rise up into the upper atmosphere based on prevailing winds and atmospheric mixing processes. This large-scale upward motion in the tropics and downward motion toward the poles is known as the Hadley circulation (*see* Hadley Circulation, Volume 1). This air motion that mixes molecules to greater altitudes is faster than gravitational settling. Once in the stratosphere, CFCs are photodissociated, releasing ClO_x radicals. This source of ClO_x was of particular importance because CFCs are primarily of anthropogenic origin (human-related). This pointed to the need to investigate the potential impacts that anthropogenic emissions of CFCs may have on future stratospheric O_3 abundance (*see* Depletion of Stratospheric Ozone, Volume 1; Stratosphere, Ozone Trends, Volume 1).

ClO_x is removed from the atmosphere primarily by the formation of hydrochloric acid (HCl). HCl, like HNO_3, is water-soluble and will be removed in the troposphere.

$$Cl + CH_4 \longrightarrow HCl + CH_3 \quad (17)$$

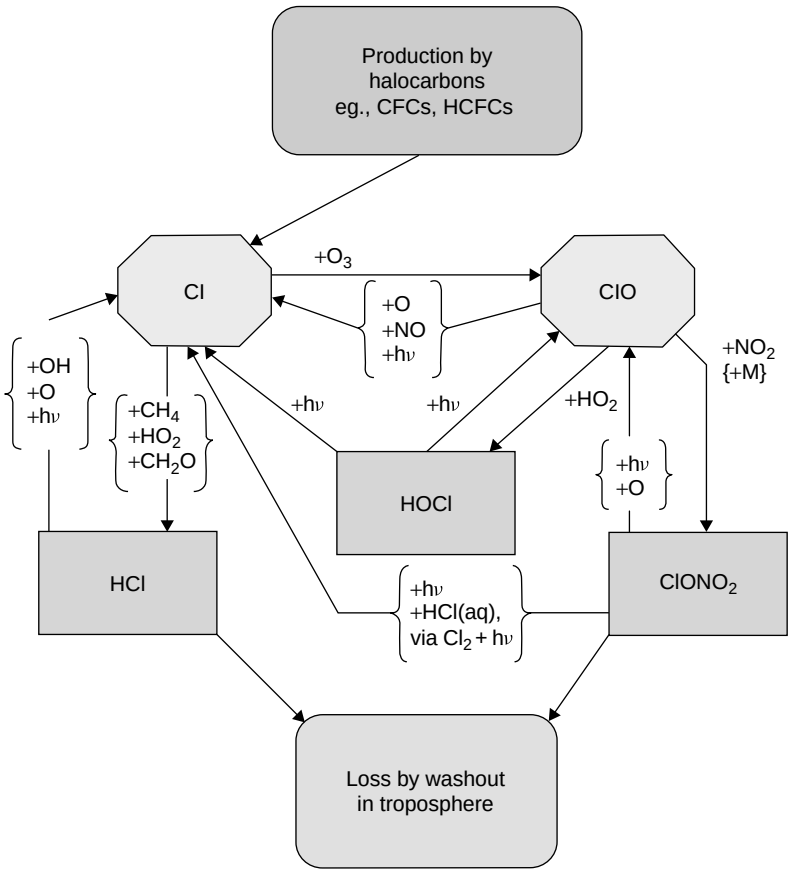


Figure 4 Principal reactions for the ClO_x chemical family. The rounded rectangular boxes represent production and loss processes of odd-chlorine species. Species contained within octagonal boxes represent ClO_x radicals with short atmospheric lifetimes (i.e., Cl and ClO). Species highlighted within rectangular boxes are considered temporary reservoir species (i.e., HOCl, $ClONO_2$, and HCl)

In addition, Wofsy *et al.* (1975) proposed that bromine oxide (BrO_x) (i.e., $BrO_x = Br + BrO$) radicals, like ClO_x, would be an effective catalytic O_3 reducing agent in the stratosphere.

$$Br + O_3 \longrightarrow BrO + O_2 \quad (18)$$

$$\underline{BrO + O \longrightarrow Br + O_2} \quad (19)$$

$$\text{NET: } O + O_3 \longrightarrow 2O_2$$

The source of BrO_x radicals is from halons and bromocarbons (*see* Halocarbons, Volume 1). The present day concentration of BrO_x (<20 pptv) is considerably less than ClO_x (~3.5 ppbv), however, per molecule, the reactivity of BrO_x is considerably greater, being up to 45 times more effective than ClO_x in global O_3 destruction. The analogous BrO_x reaction to the ClO_x (Equation 17) does not occur; however, other reactions do occur that convert active BrO_x to HBr, $BrONO_2$, and HOBr. These species are all water-soluble and will be removed in the troposphere.

Figure 5 shows a comparison of a model-derived O_3 distribution using only the Chapman mechanism relative to a mechanism that also includes all the important gas-phase species contained within the HO_x, NO_x, ClO_x, and BrO_x chemical families. Including all the above-mentioned gas-phase processes (reference atmosphere in Figure 5) yields a model-calculated vertical O_3 profile significantly less than the Chapman only mechanism. The reference atmosphere in Figure 5 is consistent with the mid-latitude observed distribution of O_3 typical of 1960 conditions (i.e., prior to significant loading of CFCs).

The work of Crutzen (1970) and Rowland and Molina (1974) was cited in award of the 1995 Nobel Prize in Chemistry. Additional reading on the history and theory of gas-phase chemical processes that affect stratospheric O_3 abundance can be found in Wayne (1985); Brasseur and Solomon (1986), Johnston (1992); Brasseur *et al.* (1999); and Finlayson-Pitts and Pitts (2000).

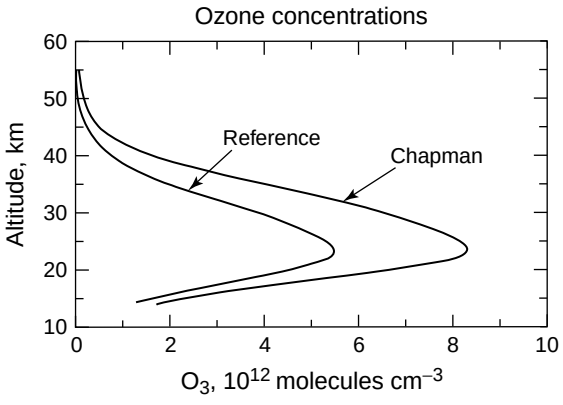


Figure 5 Model-calculated ozone vertical profile for both a Chapman mechanism and a reference atmosphere including a complete O_x, NO_x, HO_x, and ClO_x mechanism. The reference atmosphere is representative mid-latitude 1960 conditions. (Reproduced with permission from Kinnison *et al.*, 1988, copyright by the American Geophysical Union)

AEROSOL INFLUENCES ON STRATOSPHERIC CHEMISTRY

The basic understanding of gas-phase (also defined as homogeneous) chemical processes was essentially complete by the beginning of the 1980s. There certainly were needs for refinements in laboratory kinetics, in the development and evaluation of multi-dimensional models, and comprehensive multi-platform measurements (e.g., *in situ* aircraft, and remote satellite measurements) of the above mentioned chemical species. However, this confidence in our basic understanding of stratospheric chemistry was short-lived and changed dramatically after the release of a 1985 paper by Farman and colleagues in the journal *Nature*.

Farman *et al.* (1985) reported a dramatic decrease in the total column O_3 abundance in the austral spring over Antarctica (*see* Depletion of Stratospheric Ozone, Volume 1; Stratosphere, Ozone Trends, Volume 1). This decrease in total column O_3 was unprecedented and pointed to the need for the atmospheric science community to understand the physical, chemical, or dynamical process (or combination thereof) that was responsible for this unprecedented O_3 loss.

It quickly became apparent that known gas-phase chemical processes could not explain this dramatic change in O_3, which had been occurring since the mid-to-late 1970s. Farman *et al.* (1985) had suggested that anthropogenic changes in chlorine might be responsible for the observed O_3 change. However, the reactions that control O_3 loss in the ClO_x catalytic cycle (Equations 15 and 16) were not important at high latitudes in the 15–25 km altitude range where this O_3 depletion was believed to occur. These reactions were more important at 35–45 km. Gas-phase chemical processes might be responsible for total column O_3 depletion of 5%, but certainly not the observed 50% depletion.

Solomon *et al.* (1986) suggested heterogeneous chemical processes were the explanation; specifically, that reactions of gas-phase constituents on stratospheric particles (or aerosols) were necessary to explain this rapid O_3 loss. It was well known that ice particles or mother-of-pearl clouds were common during the cold (<190 K) Antarctic winter/spring. Solomon *et al.* (1986) specifically proposed that the reaction of HCl and $ClONO_2$ on polar stratospheric clouds (PSCs) would convert the temporary chlorine reservoir species (i.e., HCl and $ClONO_2$) into active ClO_x radicals, capable of destroying O_3 (*see* Aerosols, Polar Stratospheric Cloud, Volume 1; Aerosols, Stratosphere,

Volume 1; Depletion of Stratospheric Ozone, Volume 1):

$$HCl(ads) + ClONO_2(g) \longrightarrow Cl_2(g) + HNO_3(ads) \quad (20)$$

In Equation (20), HCl is adsorbed (ads) on the surface of the particle and $ClONO_2$ is in the gas-phase. When a collision occurs between $ClONO_2$ and the HCl adsorbed on the surface of the particle, molecular chlorine (Cl_2) is formed and released into the gas-phase. HNO_3 remains adsorbed on the surface of the particle. This process occurs throughout the winter period under polar night conditions. When sunlight returns in the spring, Cl_2 is photolyzed and the active chlorine is released to destroy O_3. However, as mentioned above, under these conditions, the ClO_x catalytic cycle (Equations 15 and 16) is not efficient, therefore a new catalytic cycle was proposed by Molina and Molina (1987).

$$Cl_2 + h\nu \longrightarrow 2Cl \quad (21)$$

$$Cl + O_3 \longrightarrow ClO + O_2 \text{ (twice)} \quad (22)$$

$$ClO + ClO + M \longrightarrow Cl_2O_2 + M \quad (23)$$

$$Cl_2O_2 + h\nu \longrightarrow Cl + ClOO \quad (24)$$

$$\underline{ClOO + M \longrightarrow Cl + O_2 + M} \quad (25)$$

$$\text{NET: } 2O_3 \longrightarrow 3O_2$$

This set of reactions (Equations 22–25) regenerates atomic chlorine and therefore is catalytic. A similar catalytic cycle has been proposed with BrO_x species. The heterogeneous ClO_x activation process is represented in Figure 4 (see HCl (aq), via $Cl_2 + h\nu$).

One very important consequence of heterogeneous reactions like Equation (20) is that the HNO_3 remaining on the particle will eventually settle with the particle and the stratosphere will become denitrified. Denitrification (i.e., loss of $NO_y = NO_x + HNO_3$) will tend to promote O_3 depletion by keeping the NO_x abundance low. If NO_x is low, the ClO_x and BrO_x radicals responsible for O_3 depletion cannot readily reform the temporary reservoir species $ClONO_2$ and $BrONO_2$, therefore extending the O_3 loss from catalytic cycles like Equation (22) through Equation (25).

$$ClO + NO_2 + M \longrightarrow ClONO_2 + M \quad (26)$$

$$BrO + NO_2 + M \longrightarrow BrONO_2 + M \quad (27)$$

There are additional heterogeneous reaction pathways similar to Equation (20) that deplete reservoir species (e.g., $ClONO_2$, $BrONO_2$, HOBr, HOCl, and HCl) and activate both ClO_x and BrO_x radicals. These reactions can occur not only on PSCs composed of water-ice aerosols, but also on PSCs that are composed of solid nitric acid and liquid sulfuric acid aerosols (*see* Aerosols, Polar Stratospheric Cloud, Volume 1; Depletion of Stratospheric Ozone, Volume 1; Aerosols, Stratosphere, Volume 1).

Shortly after the discovery of the Antarctic O_3 hole, Hofmann and Solomon (1989) proposed that heterogeneous chemical processes occur not only at high-latitudes during polar winter/spring, but also at mid-latitudes in the lower stratosphere throughout the year. These heterogeneous processes typically occur on liquid sulfate aerosols (*see* Aerosols, Stratosphere, Volume 1).

Sulfate aerosols are formed in the stratosphere primarily from oxidation of carbonyl sulfide (COS). At temperatures above approximately 200 K, sulfate aerosols are composed of a binary mixture of sulfuric acid (H_2SO_4) and H_2O. As the temperature decreases, uptake of HNO_3 occurs, forming a ternary solution of H_2SO_4, HNO_3, and H_2O. One of the most important heterogeneous reactions on the surface of liquid sulfate aerosols (at mid-latitudes) is the reaction of dinitrogen pentoxide (N_2O_5) with H_2O. As shown in Figure 3, N_2O_5 is a temporary reservoir species for NO_x and is formed from the three-body reaction of NO_2 and nitrogen trioxide (NO_3).

$$NO_2 + NO_3 + M \longrightarrow N_2O_5 + M \quad (28)$$

$$N_2O_5 + H_2O \text{ (aq)} \longrightarrow 2HNO_3 \quad (29)$$

Equation (29) will convert N_2O_5, a temporary NO_x reservoir, into HNO_3, effectively reducing the importance of the NO_x catalytic cycle on O_3 loss (Equations 10 and 11). Since the HO_x, NO_x, ClO_x, and BrO_x O_3 loss processes are tightly coupled, reducing the importance of the NO_x catalytic cycle increases the importance of the other O_3 depleting catalytic cycles. In fact, increased mid-latitude O_3 depletion is exacerbated following large volcanic eruptions (e.g., the recent eruptions of El Chichon and Mt. Pinatubo) that introduce sulfur dioxide (SO_2) gas into the stratosphere. The SO_2 is oxidized to form liquid sulfate aerosol, increasing the surface area density and therefore, increasing the rate of O_3 depletion by heterogeneous reactions (*see* Depletion of Stratospheric Ozone, Volume 1; Volcanic Eruptions, Volume 1; Volcanic Eruption, El Chichón, Volume 1; Volcanic Eruption, Mt. Pinatubo, Volume 1).

Additional reading on the history and theory of heterogeneous chemical processes that affect stratospheric O_3 abundance can be found in World Meteorological Organization (WMO) (1999); Brasseur *et al.* (1999); Solomon (1999); and Finlayson-Pitts and Pitts (2000).

PROSPECTS FOR THE 21st CENTURY

Future stratospheric research will be conducted within the context of the Earth system and its response to natural and anthropogenic-induced changes. The physical, dynamical, and chemical couplings between the different regions of the atmosphere (i.e., troposphere, stratosphere, mesosphere, and thermosphere), the ocean, and the biosphere need to be better understood in order to assess the impact

of natural and anthropogenic influences on stratospheric chemistry and climate in the 21st century (*see* Earth System Processes, Volume 1).

New observational platforms (e.g., satellite-based instruments) will be developed and launched to study global change. These satellite-based measurements will be compared with ground-based and *in situ* measurements from balloons and aircraft. These data will be used to evaluate Earth system models (ESMs) that attempt to represent the above-mentioned components of the Earth s system. In turn, these ESMs will be used to assess the impacts of future anthropogenic emissions on atmospheric composition of not only stratospheric O_3, but also the climate impacts that changes in stratospheric O_3 may induce.

One of the most important environmental issues to address in the 21st century will be the future change in the stratospheric O_3 layer. Will the magnitude of the Antarctic O_3 hole depletion increase? Will the temporal and spatial area of the observed O_3 depletion increase? Will a similar decrease in O_3 occur in the Arctic? When (if ever) will stratospheric O_3 return to natural abundances? How will the increase in greenhouse gas abundances influence stratospheric chemistry and climate? These questions are currently at the forefront of scientific research related to the stratosphere (*see* Arctic Climate, Volume 1; Depletion of Stratospheric Ozone, Volume 1).

Since the days of Priestly, Schelle, and Schonbein, there have been great advances in our understanding of stratospheric chemistry and the role that the stratosphere plays in global chemistry and climate. As in the past, there most likely will be many surprises in store for researchers in the 21st century.

REFERENCES

Brasseur, G P and Solomon, S (1986) *Aeronomy of the Middle Atmosphere*, 2nd edition, Atmospheric Sciences Library, D Reidel, The Netherlands.

Brasseur, G P, Orlando, J J, and Tyndall, J S (1999) *Atmospheric Chemistry and Global Change*, Oxford University Press, Oxford.

Crutzen, P J (1970) The Influence of Nitrogen Oxides on the Atmospheric Ozone Content, *Q. J. R. Meteorol. Soc.*, 96, 320–325.

Farman, J C, Gardiner, B G, and Shanklin, J D (1985) Large Losses of Total Ozone in Antarctica Reveal Seasonal ClO_x/NO_x Interaction, *Nature*, 315, 207–210.

Finlayson-Pitts, B J and Pitts, Jr, J N (2000) *Chemistry of the Upper and Lower Atmosphere, Theory, Experiments, and Applications*, Academic Press, San Diego, CA.

Hofmann, D J and Solomon, S (1989) Ozone Destruction Through Heterogeneous Chemistry Following the Eruption of El Chichon, *J. Geophys. Res.*, 94, 5029–5041.

Johnston, H S (1971) Reduction of Stratospheric Ozone by Nitrogen Oxide Catalysts from Supersonic Transport Exhaust, *Science*, 173, 517–522.

Johnston, H S (1992) Atmospheric Ozone, *Annu. Rev. Phys. Chem.*, 43, 1–32.

Molina, M J and Rowland, F S (1974) Stratospheric Sink for Chlorofluoromethanes: Chlorine Atom-catalysed Destruction of Ozone, *Nature*, 249, 810–812.

Solomon, S, Garcia, R R, Rowland, F S, and Wuebbles, D J (1986) On the Depletion of Antarctic Ozone, *Nature*, 321, 755–758.

Solomon, S (1999) Stratospheric Ozone Depletion: A Review of Concepts and History, *Rev. Geophys.*, 37, 275–316.

Stolarski, R S and Cicerone, R J (1974) Stratospheric Chlorine: A Possible Sink for Ozone, *Can. J. Chem.*, 52, 1610–1615.

Wayne, R P (1985) *Chemistry of Atmospheres, an Introduction to the Chemistry of the Atmospheres of Earth, the Planets, and their Satellites*, Oxford Science Publication, Clarendon Press, Oxford.

World Meteorological Organization (WMO) (1999) *Scienti c Assessment of Ozone Depletion: 1998*, Global Ozone Research Monitoring Project – Report No. 44.

Stratosphere, Ozone Trends

Rumen D Bojkov

World Meteorological Organization, Geneva, Switzerland

Results from more than 40 years of continuous observations, laboratory experiments, and modeling studies have led to a scienti c consensus regarding the fragile state of the stratospheric ozone layer. Starting in the 1970s, a stratospheric ozone decrease has been observed, and it is continuing at variable rates. In general, the decrease is twice as strong in the winter and spring (>5% per decade decrease) as during the summer and autumn. The ozone decrease is most extensive over the Antarctic, especially during September to November (>20% per decade decrease). Over the Northern Hemisphere mid- and polar latitudes, the ozone decline is also quite pronounced, especially during the cold winter- spring seasons of the 1990s. Only the equatorial belt (20°S- 20°N) is experiencing no signi cant changes. Although tropospheric ozone has increased, especially over the Northern Hemisphere mid-latitudes, this increase cannot compensate for the large decline in stratospheric ozone and the resulting increase in ultraviolet (UV) radiation at the surface.

Theoretical, experimental, and observational studies all con rm that the multi-decade stratospheric ozone depletion is related to increased human-induced emissions of chlorine- and bromine-containing compounds such as chloro uorocarbons (CFCs) and halocarbons. Low ozone values observed over short time periods are also in uenced

by atmospheric circulation processes. In addition to their effect on UV-radiation, the ozone changes contribute to radiative forcing of the climate comparable to other minor greenhouse gases. With the very cold lower stratospheric temperatures that are characteristic of springtime in both the Arctic and Antarctic, the ampli ed losses of ozone in polar and mid-latitudes are likely to persist until the overall stratospheric chlorine loading decreases to below ~3 parts per billion by volume (ppbv). A recovery of the global ozone layer to its pre 1970s conditions is not expected to occur until after ~2050, when the chlorine loading is expected to fall below 2 ppbv. This will only occur if all countries strictly adhere to the Montreal Protocol requirements for phasing out ozone-depleting compounds.

INTRODUCTION

A German physicist (C F Schonbein) discovered ozone when observing electrical discharges at the University of Basel in 1839, but it was not until after 1850 that it was determined to be a natural atmospheric constituent. The word *ozone* comes from the Greek word meaning smell, a reference to ozone s distinctively pungent odor when present in large concentrations. Surface ozone measurements were made at hundreds of places during the second half of 19th century. However, total column ozone measurements were initiated only at the end of 1920s using a few Dobson spectrophotometers. Only starting in the 1960s has their number exceeded 100.

Ozone (O_3) is a form of the element oxygen (O) that has three atoms in each molecule instead of the two atoms making up normal oxygen molecules (O_2). In the atmosphere, most ozone is formed in the tropical middle stratosphere by the action of solar radiation on oxygen molecules in a process called photolysis. In this process, O_2 molecules are broken down to yield atomic oxygen, which in turn combines with molecular oxygen to produce ozone. Ozone is then transported at stratospheric levels from the tropics to higher latitudes and then downward to the lower stratosphere. In the lower stratosphere, the ozone lifetime is ~1 year. Ozone is destroyed naturally through a series of catalytic cycles involving reactive nitrogen (NO_y), chlorine, bromine and hydrogen species (HO_x) (*see* Stratosphere, Chemistry, Volume 1; Stratosphere, Temperature and Circulation, Volume 1).

The stratosphere (extending ~10–50 km above the Earth s surface) contains ~90% of all the ozone in the atmosphere. The ozone molecules are concentrated mainly between altitudes of 15–35 km. Looking up through the atmosphere, the ozone column has its maximum partial pressure in the lower stratosphere at a level of 19–23 km above the surface. The balance between photolytic production, transport, and chemical destruction determines the abundance of ozone at any particular stratospheric location. This balance is also strongly seasonally dependent, having a maximum in the middle and polar latitudes in the winter-spring and a minimum in the autumn.

Ninety-nine per cent of the air we breathe is nitrogen (78%) and oxygen (21%). This ratio has not changed for millions of years. On average, trace components such as water vapor, carbon dioxide, methane, nitrous oxide, ozone and inert gases (e.g., argon, helium, neon) make up less than 1% of the volume of air. Although concentrations vary widely throughout the atmosphere, on average, only three of every 10 million molecules of air are ozone molecules.

Total ozone is a measure of the amount of ozone contained in a vertical column of air reaching from the surface up through the whole atmosphere. Total ozone is expressed in units of pressure (milli-atmosphere centimeters, or matm-cm), being the depth the ozone layer would be at standard pressure and temperature. If all the ozone in the atmosphere were moved toward the Earth s surface to calculate the total ozone, it would, averaged around the world, assume a thickness of about 3 mm (300 matm-cm). Except when the ozone hole is present (see herein), the lowest monthly mean values of total ozone typically occur over the equatorial belt (~220 matm-cm) and the amount increases toward the poles. The maximum monthly mean value that is observed reaches ~480 matm-cm in the Northern Hemisphere. These variations in total ozone at any particular place are mainly determined by the interplay between large-scale atmospheric dynamics and atmospheric chemistry.

Although only present at exceedingly low concentrations, ozone molecules play a vital role in the life of our planet. Of most importance, the stratospheric ozone layer normally absorbs virtually all of the highly energetic solar UV radiation (i.e., radiation having wavelengths below about 320 nm that can cause sunburn and skin cancer), thereby providing an important shield that protects humans and all other animals and plants. Ozone also largely determines the thermal structure of the stratosphere (10–50 km) where temperature increases with height, and, by absorbing infrared radiation, ozone serves to enhance the Earth s greenhouse effect.

Since the early 1970s, ozone, as an atmospheric trace gas, has gone from being of interest only to a small group of scientists to an issue of global prominence. This leap occurred because it was determined that the normal concentration of atmospheric ozone is under attack from chemicals released by human activities (*see* Depletion of Stratospheric Ozone, Volume 1). The Nobel Prize in Chemistry for 1995 was given to three ozone researchers for their contributions in working out the chemical destruction mechanisms (*see* Crutzen, Paul J, Volume 1; Molina, Mario J, Volume 1; Rowland, F Sherwood, Volume 1).

The ozone decline that these scientists projected was detected from data collected by the World Meteorological Organization s (WMO) Global Ozone Observing System,

which has operated more than 150 stations around the world since the mid-1950s. In the last 20 years, ozone depletion has also been detected from a few specialized satellites. In the mid-1980s, evidence of ozone destruction also became evident as a result of identification of the dramatic ozone decline in the Antarctic spring, popularly labeled the ozone hole.

That these downward trends in total ozone were due to human factors has also been established. Because stratospheric ozone is primarily created by UV radiation, the Sun s output affects the rate at which it is produced. Indeed, the Sun s energy release in the UV part of the spectrum varies significantly over the well-known 11-year sunspot cycle (*see* Solar Variability, Long-term, Volume 1; Sunspots, Volume 1). Observations over several solar cycles since the 1950s show that total global ozone levels decrease by 1–2% from the maximum to the minimum of a typical solar cycle. The long-term ozone decline observed during the last two decades, however, has been much greater than this, making clear that this decline cannot be due to changes in solar activity.

Based on more than 25 years of research by scientists from around the world, including laboratory studies, field measurements and theoretical investigations, a strong link has been clearly established between human-made compounds and these ozone losses. Even as the Sun s energy produces new ozone in the upper stratosphere, natural compounds containing oxygen, nitrogen, hydrogen and chlorine or bromine continuously destroy ozone molecules. Such chemicals were all present in the stratosphere in small amounts long before humans began altering the composition of the atmosphere. Nitrogen compounds come from soils and the oceans, hydrogen comes mainly from atmospheric water vapor and some chlorine comes from the oceans in the form of methyl chloride and methyl bromide. However, human activities have significantly upset the delicate balance of production and destruction. By releasing relatively large amounts of chlorine- and bromine-containing chemicals (e.g., *see* Chloro uorocarbons (CFCs), Volume 1; Halocarbons, Volume 1) into the atmosphere, societal activities have enhanced the destruction of ozone, leading to lower ozone concentrations in the stratosphere.

It was on the basis of this information, summarized and evaluated in a series of WMO Ozone Assessment Reports (see references), that countries agreed to the first environmental Convention for the Protection of the Ozone Layer (Vienna, 1985) and its Montreal Protocol (1987) (*see* Ozone Layer: Vienna Convention and the Montreal Protocol, Volume 4). Their implementation represents an environmental success story. The countries took action to halt and then reverse an emerging problem. Results of these treaties provide a model for further international action against global threats to the environment. The action to defend the ozone layer will rank as one of the great international achievements of the 20th century. There are still some uncertainties regarding the future of the ozone layer. For example, its recovery in the next 50–60 years depends on the ability of nations to comply with international agreements and curb emissions of substances containing chlorine and bromine that deplete stratospheric ozone.

The opposite process is occurring in the troposphere, which is the lower part of the atmosphere extending up to 10–14 km. Here, mainly as a result of combustion processes, the local concentrations of ozone in the Northern Hemisphere mid-latitudes have nearly tripled in the last 100 years (*see* Troposphere, Ozone Chemistry, Volume 1). This tropospheric ozone increase, however, cannot compensate for the effects of the decline in stratospheric ozone on UV radiation, although the changes could influence the radiative balance of the surface-atmosphere system.

OBSERVED CHANGES IN THE GLOBAL OZONE

The ozone concentration varies over many time and space scales. A major component of the global ozone variation is its annual cycle. Total ozone in the polar and mid-latitudes reaches its annual maximum in the winter–spring and minimum in the autumn months. In the pre-1976 period, the annual peak to trough differences were ∼180 and ∼145 matm-cm in the polar and mid-latitude regions, respectively. Since about 1985, these peak to trough differences have decreased by 15–20%, due mainly to the decline of total ozone during the period of the ozone annual maximum.

In addition to the annual cycle, the natural variability of atmospheric ozone is evident over a broad spectrum of time scales, from day to day and month to month, to interannual and interdecadal. Atmospheric circulation, chemistry and radiative processes all play important roles in this ozone variability. For example, the amplitude of synoptic-scale (i.e., weather-driven) variations can be ∼30% of the monthly mean over polar and mid-latitudes. This compares to the peak to peak amplitude of the annual cycle at particular latitudes ranging from less than 6% in the equatorial and tropical belts to ∼30% at sub-polar latitudes.

At longer time scales, a number of factors contribute to the variability. For example, the quasi-biennial oscillation (QBO) (*see* Quasi- Biennial Oscillation (QBO), Volume 1) contributes to interannual variability. The amplitude of the QBO-induced variation ranges from 3 to 7–8%, depending on latitude. Variations in solar activity over the 11-year sunspot cycle are observed to range between 1 and 2%. Large volcanic eruptions reaching the stratosphere also can cause temporary reductions in the average ozone loading of from 2% at low latitudes up to 4–5% over high latitudes in winter.

In addition to the natural variations, the comprehensive international ozone assessment reports sponsored by the WMO, which are based on extensive sets of papers analyzing ground station observations and satellite radiances, all note that total ozone has been decreasing significantly in mid- and polar latitude regions over the last 20–25 years. Only in the equatorial belt (20 °S to 20 °N) are the trends not statistically significant.

Based on ground-based and satellite observations, Figure 1 shows the estimated annual cycle of average monthly means for two different periods. The years 1964 to 1976 are taken as the base period. During these years, the concentrations of anthropogenic compounds that led to the destruction of ozone were quite low and no systematic ozone trends were noticed. The later period of 1986–1999, however, was the time of maximum emission of CFCs and halocarbons. The annual cycle for the base period has two maxima, corresponding to the time of the Northern Hemisphere spring ozone increase in March–April and to the maximum reached in September–October in most of the Southern Hemisphere. The overall average total ozone for the globe was ∼307 matm-cm, with more in the Northern Hemisphere than in the Southern Hemisphere by ∼4%.

Data for the last 14 years show very significant ozone destruction. The depletion is concentrated during the austral spring and during the winter–spring period in northern latitudes. These changes have substantially altered the amount of ozone and the shape of the annual cycle. In the 1986–1999 period, the average global ozone amount declined to ∼296 matm-cm. The secondary maximum in September–November has disappeared due to the drastic ozone decline over Antarctica during these months.

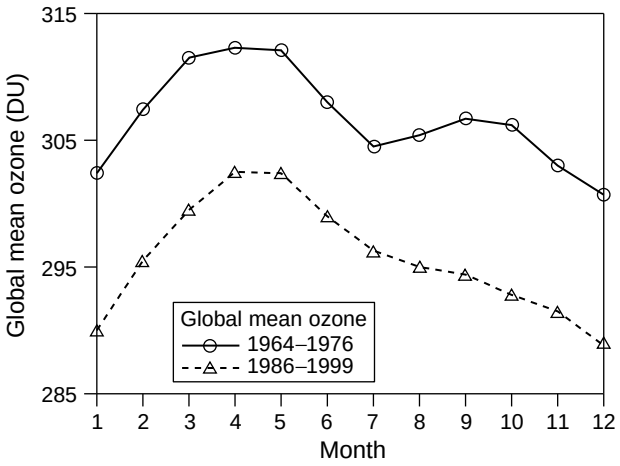


Figure 1 Monthly global average ozone values (in Dobson units, DU) for two periods show substantial overall decline during the latter years. (Updated from Bojkov, WMO, 1999)

Although there has been relatively little change over equatorial regions, the overall ozone decline for the rest of the globe between these two periods was greater than 3.5%.

Figure 2 shows total ozone mapping spectrometer (TOMS) instrument satellite data indicating the change in total ozone for 60 °N–60 °S, adjusted for seasonal, QBO and solar influences. The linear trend prior to May 1991 is −2.3% per decade. The strong, nearly linear, decrease in total ozone throughout the early 1990s is followed by a sudden decrease in 1992–1993, possibly related to effects of the Mt Pinatubo eruption (*see* Volcanic Eruption, Mt.

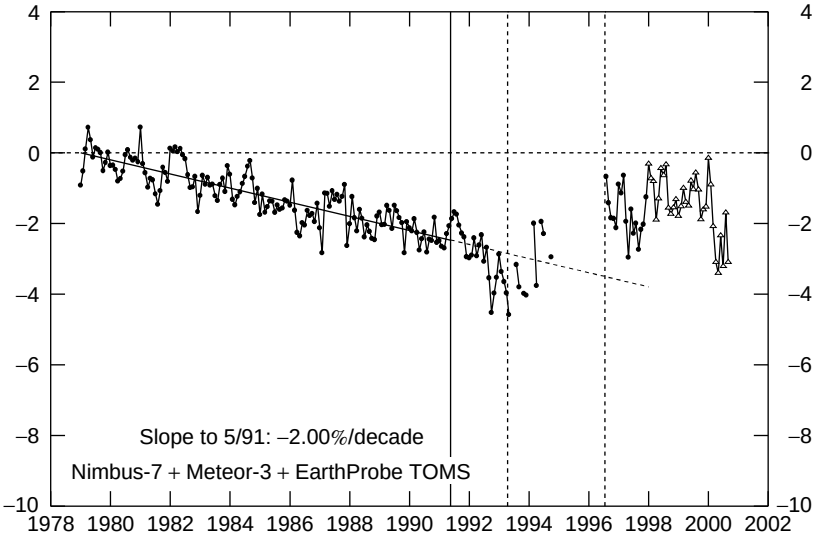


Figure 2 Deviations in total ozone, area weighted over 60 °S–60 °N adjusted for seasonal, QBO and solar effects. Straight line is the least squares fit. (*Source:* WMO Assessment-98, updated to 30 September 2000 by L Bishop)

Pinatubo, Volume 1). Thereafter, the trend is no longer a linear extension from the previous period, but is leveling off. However, this apparent leveling off may be an artifact because reliable ground station data indicate that the preliminary data from the TOMS instrument on the Earth Probe satellite that are available since September 1996 are showing ~1–1.5% higher values than the ground station data. Some unusually warm Arctic winters also contributed to the appearance of higher ozone in 1998 and 1999. In addition, it is also possible that the long-term influence of changes in solar radiation is not represented well enough in the statistical model, and so this may account for some of the leveling off. It should be noted that this reduction of the downward trend should not be interpreted as a recovery of the ozone layer.

In both hemispheres, the negative ozone trends are stronger during the winter–spring seasons than during the summer–autumn seasons. The calculated annual average trend for the period 1979–2000 (in % decade^{-1}) for the latitude band 35–65 °N is roughly constant at ~−2.4% decade^{-1}. In the southern latitudes, there is a significant difference between the trends in the bands 35–50 °S and 50–65 °S, with values of ~−1.7 and −4.1% decade^{-1}, respectively. If one considers also the stronger ozone decline in the polar regions, the ozone amount outside the tropics is estimated to be 6–8% less than it was about 25 years ago.

Observations, experiments and model simulations confirm that chlorine- and bromine-activated heterogeneous reactions in the stratosphere are responsible for the main part of the observed ozone destruction. However, parts of the low ozone and its fluctuations are related to variations in the atmospheric circulation wave activity. During the past two decades, ozone has been varying in unison with the fluctuations in the lower stratospheric temperatures. Much of the interannual and seasonal variations in both ozone and temperature are related to the specifics of the atmospheric circulation, and were present before the anthropogenic influence started to dominate the ozone decline. In particular, changes in the upwelling planetary wave activity from the troposphere that enters the stratosphere have the effect of modulating the diabatic circulation and thus the ozone transport. These effects are considered to play an important role in the persistence of the Arctic vortex, in determining springtime stratospheric temperatures in polar regions, and in interannual ozone variations.

The chemical processes leading to ozone depletion can be summarized in terms of a relatively few chemical reactions (*see* Depletion of Stratospheric Ozone, Volume 1). The depletion of ozone in the polar lower stratosphere is due to a sequence of chemical reactions that starts with the conversion of halogen-containing (chlorine and bromine) gases with long lifetimes into less stable reservoir forms. The reservoirs of chlorine are hydrochloric acid (HCl) and chlorine nitrate ($ClONO_2$). Dinitrogen pentoxide (N_2O_5) and nitric acid (HNO_3) are reservoirs of nitrogen oxides. The Cl-containing reservoirs are then broken down by reactions involving polar stratospheric clouds (PSCs) to release high concentrations of reactive chlorine and chlorine monoxide radicals (Cl and ClO). These in turn contribute to catalytic cycles that can destroy ozone very rapidly in the presence of sunlight.

Laboratory experiments suggest that the most important reactions are:

$$HCl + ClONO_2 \longrightarrow HNO_3 + Cl_2 \quad (1)$$

$$ClONO_2 + H_2O \longrightarrow HNO_3 + HOCl \quad (2)$$

$$HCl + HOCl \longrightarrow H_2O + Cl_2 \quad (3)$$

$$N_2O_5 + H_2O \longrightarrow 2HNO_3 \quad (4)$$

Particles that support heterogeneous reactions exist throughout the lower stratosphere (e.g., sulfate aerosol particles, sulfur dioxide). The reactions involving chlorine species are temperature dependent and become much more efficient on liquid PSCs in the cold (<−78 °C) polar lower stratosphere. Furthermore, at mid-latitudes aerosol reactions can also perturb the chemistry, especially when large amounts of aerosol are inserted by a large volcanic eruption.

Following the heterogeneous production of active chlorine, sunlight can drive the following catalytic cycles in the polar region:

(I) $ClO + ClO + M \longrightarrow Cl_2O_2 + M$

$Cl_2O_2 + h\nu \longrightarrow Cl + ClO_2$

$ClO_2 + M \longrightarrow Cl + O_2 + M$

$2 \times (Cl + O_3 \longrightarrow ClO + O_2)$

net: $2O_3 \longrightarrow 3O_2$

(II) $ClO + BrO \longrightarrow Br + Cl + O_2$

$Cl + O_3 \longrightarrow ClO + O_2$

$Br + O_3 \longrightarrow BrO + O_2$

net: $2O_3 \longrightarrow 3O_2$

The dimer of the ClO radical (Cl_2O_2) involved in cycle I is thermally unstable, and the cycle is most effective at low temperatures. It is thought to be responsible for most of the ozone loss in Antarctica, with a large contribution from cycle II involving reactive bromine. In the warmer Arctic, a larger proportion of the loss may be driven by cycle II. In the mid-latitude lower stratosphere, the coupled ClO + BrO (Cycle II) and $BrO + HO_2$ cycles are responsible for increased halogen-induced chemical depletion.

The roles of these reactions are coupled to the climatic conditions in each hemisphere. The next two sections cover the effects in the Northern and the Southern Hemisphere. Although much attention is devoted to the ozone hole in the Southern Hemisphere, the changes in the Northern Hemisphere may well emerge as most important for society in the 21st century, so that is covered here first (see Box 1).

OZONE DECLINE IN THE NORTHERN HEMISPHERE

In the mid-latitudes in general, the natural day to day variations of ozone amounts are seldom greater than 10% of the mean. The changes vary from day to day and are regionally distributed as a function of meteorological conditions. The variations increase toward the polar latitudes. Therefore, one indication of ozone decline would be if the surface area covered with deviations of greater than -10% from the pre-1976 mean (i.e., areas experiencing $>10\%$ depletion) is increasing.

Figure 3 shows the fraction of the entire surface poleward from $35\,^{\circ}$N with contours greater than $+10\%$ and -10% ozone deviations occurring during the winter–spring period. It is clear that the area of the negative deviations has grown from $\sim$15% in the beginning of 1980s to $>$45% in the last 11 years, an increase of a factor of three. The highest fractions of $>70\%$ correspond to major ozone declines in 1993 and 1995. Over the same period, the fraction of the area covered by positive deviations of greater than $+10\%$ has declined from $\sim$15% to $\sim$5%.

Figure 4 shows the total ozone data from a dozen European and North American stations having long-term observations, smoothed to generate a 12-month running mean. The underlying decline from the 1970s to the early 1990s is modulated by large, sawtooth-like variations over time scales of every two to two and a half years that are related to the QBO (*see* Quasi- Biennial Oscillation (QBO), Volume 1). The QBO oscillations are best expressed in the periodic easterly to westerly changes of stratospheric winds in the equatorial belt. The QBO modulates the diabatic circulation, which is accompanied by a similar modulation

Box 1 State of the Ozone Layer – 2000

- The total abundance of ozone-depleting compounds in the troposphere peaked in about 1994 at about 3.7 parts per billion by volume and is now starting to decline slowly, indicating that the measures formulated by the Montreal Protocol are having an effect. Total bromine, however, is still increasing. The combined abundance of chlorine and bromine in the stratosphere peaked around the year 2000 and is expected to remain high for the next decade or two.
- The observed total column ozone losses for Northern Hemisphere mid-latitudes from the mid-1970s to 1998–1999 are close to 7% in winter–spring and $\sim$3% in summer–autumn. Year-round losses in the Southern Hemisphere mid-latitudes are less than 5%. The strongest rate of decline in stratospheric ozone at mid-latitudes has been during 1985–1995. Some of the higher ozone values observed in the northern latitudes in 1998 and 1999 winter–spring periods were related to an intensified transport of sub-tropical air that had a low ozone concentration. However, in the spring of 2000, the ozone chemical destruction in the European sector of Arctic and adjacent middle latitudes reached nearly 60% at 18 km altitude. There are no statistically significant trends in total ozone in the equatorial region ($20\,^{\circ}$S to $20\,^{\circ}$N).
- In the Arctic, late winter-spring ozone values were unusually low in seven out of the last 11 years (since 1990). The seven years were unusually cold and this led to protracted stratospheric winters with minimum Arctic vortex temperatures near the threshold for large chlorine activation. Ozone has declined during some spring months by 25–30% below the pre-1976 average. For short-periods, declines of 200 matm-cm (a decrease of 45%) were registered. Continued presence of an elevated stratospheric halogen abundance over the next decade or two in conditions of low stratospheric temperatures would imply the Arctic's continuing vulnerability to very large ozone losses.
- Above Northern Hemisphere mid-latitudes, the downward trend is largest near 40 km and 15 km ($>$7% per decade) and is smallest at 30 km (2% per decade). The bulk of the column ozone decline is between the tropopause and $\sim$24 km.
- The springtime Antarctic ozone hole continues unabated. Although the extent of ozone depletion was strongest in 1998, the depletion has, during the last 3–4 years, remained generally similar to depletions in the early 1990s, with monthly total ozone in September and October continuing to be 40–55% below the pre-ozone hole values. The austral spring of 2000 did show an early (in late August–early September) record ozone decline. However, due to an early polar vortex breakdown (in the beginning of November), the seasonal deficiency was not the strongest.
- Field studies show that anthropogenic emissions of ozone precursors (nitrogen oxides, carbon monoxide and hydrocarbons) lead to large-scale production of ozone in the troposphere. Through long-range transport, these emissions influence the ozone concentration in large regions of the troposphere in both hemispheres. Trends in tropospheric ozone since 1970 in the Northern Hemisphere show large regional differences, with increases over Europe and Japan, decreases over Canada, and only small changes over the US. In the Southern Hemisphere, small increases have now been observed in surface ozone.

Box 1 (*continued*)

- The inverse dependence between the decline of overhead amount of ozone by 1% causing about similar increase of surface UV radiation, has been further quantified, but this ratio is strongly dependent on the presence of clouds, particles and surface reflectivity.
- Stratospheric ozone losses have caused a cooling of the global lower stratosphere and contributed to a negative radiative forcing to the climate system. Much of the observed downward trend in lower stratospheric temperatures (more than 0.6 °C per decade from 1979–1999) is attributed to the ozone loss in the lower stratosphere. The stratospheric ozone losses since 1980 may have offset about 30% of the positive forcing caused by increases in the well-mixed greenhouse gases (i.e., carbon dioxide, methane, nitrous oxide and the halocarbons) over the same time period.
- The increase of ozone in the troposphere since preindustrial times is estimated to have augmented the average radiative forcing by $+0.35\,W\,m^{-2}$, which is thought to have contributed 10–20% of the warming attributed to the increase in radiative forcing of long-lived greenhouse gases during the same period.
- The most vulnerable period for ozone depletion will be the next decade or two. Evidence for the recovery of the ozone layer lies still further off. Because of the slow rate at which natural processes remove these compounds from the stratosphere, the drop in total chlorine and bromine abundances in the stratosphere in the 21st century will be much slower than the rate of increase observed in past decades. Thus, because of natural ozone variability and changing atmospheric conditions, unambiguous detection of the recovery of the ozone layer may not be possible for perhaps another 20 years.
- The issues of ozone depletion and climate change are interconnected. Changes in ozone affect the Earth's climate, and changes in climate and meteorological conditions affect the ozone layer. This is the case because the ozone depletion and climate change phenomena share a number of common physical and chemical processes. Stratospheric cooling caused by increased concentrations of greenhouse gases could prolong the period for ozone recovery well beyond the middle of this century.

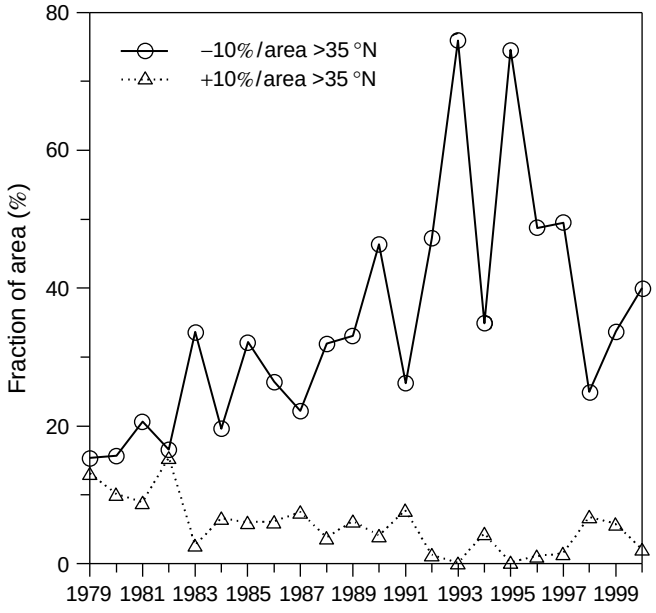


Figure 3 The fraction of the −10 and +10% contours surface area from the entire hemisphere poleward of 35 °N, during 1 January–15 April of each year since 1979 (updated from Bojkov, WMO, 1999)

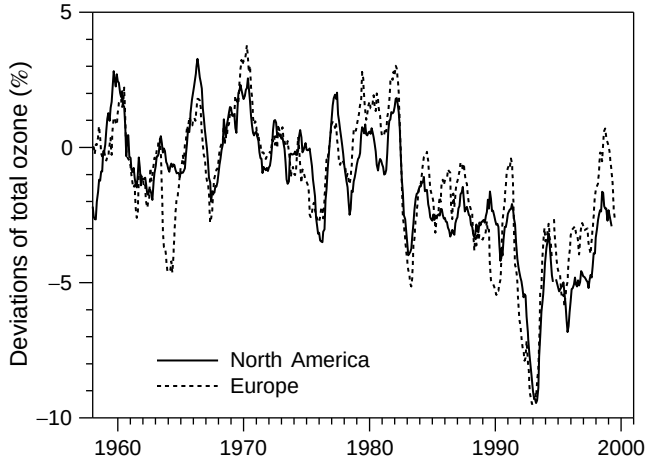


Figure 4 The deviations (in % from the pre-1976 average) of the total ozone measured at the European and North American stations (40–64 °N) with long-term observations, smoothed by a 12-month running mean. A substantial decline of the ozone column starting in the 1970s and major fluctuations during the years are obvious

of the ozone transport poleward. The QBO-induced changes in the planetary waves cause the appearance of a more disturbed and warmer polar vortex that exhibits higher ozone during the easterly phase of the equatorial winds than during the time of westerly phase. These ozone variations in middle and polar latitudes can be 3–7%. Detailed analyses of the long-term ozone records also show a solar signal of up to 2% per solar cycle. Slightly higher ozone values have been detected at the time of the solar spot maximums (about 1958, 1969, 1980, 1991 and 2000).

The behavior of the ozone deviations since the deep decline in the early 1990s has been somewhat different. Although there continues to be winter–spring ozone

destruction, the 12-month smoothed mean ozone values have been variable around a fairly constant level. There is even an indication of an increase in 1998 and 1999 followed by a substantial decrease in early 2000 (not shown on the plot).

In addition to the significant ozone decline in the last few decades, of particular interest is the variability of the ozone amount from year to year, and appearances of periods with major ozone deficiencies (sometimes >25%) that last for number of weeks over the Northern Hemisphere mid- and polar latitudes. Figure 5 shows the evolution of the total ozone area-averaged departures from the pre-1976 averages occurring during December–January–February–March (winter–spring season) over North America, Europe and Siberia during 1975–2000. Negative deviations exceeding two standard deviations for the seasonal mean have occurred in seven of the last 11 years. The major deviations in 1983 and 1992–1993 are after major volcanic eruptions (El-Chichón in 1982 and Mt Pinatubo in 1991), each of which introduced aerosols into the stratosphere that facilitated chemical destruction. However, the atmospheric circulation during the westerly phase of QBO also contributed to the low ozone levels in these winter–spring periods. QBO was also a contributing factor to the low ozone values in 1995.

Differences in the strength of the ozone departures also exist between the three continental-scale regions of the Northern Hemisphere. These relate mainly to the position and state of the Arctic stratospheric vortex. Its position is influenced by the planetary wave structures, which also modulate the circulation and resulting upwelling of air from the troposphere. In most of the springs with major negative deviations, the Arctic vortex was very strong and the lower stratospheric temperatures fell below −76 °C. Such low temperatures generate PSCs and facilitate ozone destructive reactions. From late February through early March 1996, there is evidence from UARS-HALOE satellite measurements, for example, that cold chemically preconditioned air, known to have excess ClO content, moved over the sunlit latitudes where some of the highest ozone destruction (>100 matm-cm) has been detected.

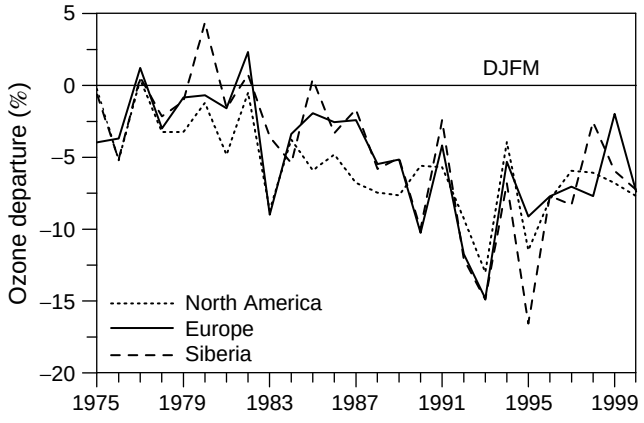


Figure 5 Total ozone departures (% from the pre-1976 average) from Dec–Jan–Feb–Mar over North America, Europe and Siberia in the 45–65 °N belt during 1975–2000. Two standard deviations of the seasonal mean is ∼6%. (Updated from 4–20 of the WMO Ozone Assessment-98)

During most of the winter–springs of the 1990s, the vortex was tilted from the central Arctic toward Siberia, frequently reaching upper middle latitudes where the solar exposure was significant enough to trigger ozone destruction even over adjacent areas. This is the main reason that the four-month averaged deviations over Siberia exceed those over the other regions. For a large number of days in 1995 and 1997, the negative ozone deviations there reached 35–40%. The regional deviations during 1990, 1995 and 1997, which exceeded 10% for the entire 4-month season, as shown in Figure 5, suggest that the presence of volcanic aerosols (as in 1992 and 1993) is not the only reason for the appearance of record low ozone values. Recent studies have demonstrated that the appearance of low ozone values, particularly over the Northern Hemisphere mid-latitudes, is related to the variability of wave activity affecting the stratosphere, changes in diabatic circulation, and transport of ozone-poor air from the sub-tropical latitudes.

The spatial distribution of the mean February–March 1990–2000 total ozone departures (in % from the pre-1976 averages) over the Northern polar and middle latitudes shows that the strongest deviations (i.e., those exceeding 15–20% for the last 12-year period) were observed over the Arctic. However, the strongest deviations were tilted (displaced) toward Siberia and deviations exceeding 10% extended toward northern and central Europe. This shape is mostly due to the location of the polar vortex.

In the Arctic vortex and adjacent areas, ozone loss reached about 100 matm-cm below the pre-1976 averages of 420–450 matm-cm. During the 1990s, extremes lasting a week or so exceeding 200 matm-cm also occurred. International campaigns conducted in the 1990s (AASE, EASOE, SESAME, THESEO, etc.) have estimated cumulative chemical ozone destruction of between 100 and 140 matm-cm during each Arctic winter–spring since 1989, except in 1998 and 1999. In the vortex in the lower stratosphere, ozone mixing ratios as low as ∼1 ppmv were observed in 1995, 1996, 1997 and 2000. In the same seasons, the concentrations of ozone depleting substances were similar to those in the Antarctic spring and, consequently, daily ozone-loss rates of ∼40 ppbv have been observed for short periods of time. The ozone loss at 18 km even reached 60% during the spring of 2000.

During the last 10–15 years, record low ozone values of less than 240 matm-cm were observed for a few days over very large regions in the October–February period. Such major negative ozone deviations are related not only to the

increased chlorine loading in the stratosphere, but also to the unusually low winter–spring temperatures facilitating ozone destruction. The atmospheric circulation causing the injection of subtropical low-ozone content tropospheric air into the lower stratosphere of the middle and polar latitudes also played a role in causing the extremely low ozone values during these short periods. Appearance of such episodes over the European–North Atlantic region has shown some increase occurred during the 1990s. Their occurrence could contribute noticeably to the observed ozone deficiency in mid-latitude ozone levels. However, to clearly distinguish between the relative role of the chemical loss and the atmospheric transport processes has proven to be extremely difficult.

To characterize the ozone changes quantitatively and to be able to objectively compare year to year and region to region variations, it is useful to introduce the concept of ozone mass deficiency (O_3MD). In order to calculate the O_3MD, the deviations from the pre-1976 zonal averages are first determined daily from homogenized satellite data using a $3^\circ \times 5^\circ$ grid. If the deviations are found to be greater than -10% (assumed scale of natural fluctuations $\sim$2 standard deviations of the typical latitudinal mean) and knowing its value in matm-cm, it is a simple matter to determine the respective ozone mass (in megatons, Mt). The daily O_3MD are subsequently integrated over time and/or space to arrive at an estimate of the overall mass deficiency for a given period and/or region.

As already indicated herein, there is a remarkable correlation between the changes in lower stratospheric temperature and in total ozone. Figure 6 shows the O_3MD amounts integrated for the 59 days of February–March period for the northern polar region (60–90 °N) together with the deviations from the pre-1978 average of the 50 hPa temperatures for the last two decades. The integrated O_3MD for February–March is capturing more than 70% of the entire winter–spring ozone deficiency. As can be estimated from Figure 6, the integrated deficiency over the polar region has increased from an average of less than 500 Mt for the early and mid-1980s to greater than 2000 Mt in 1996 and 1997. The increasing strength of the negative deviations of the stratospheric temperatures during the 1990s is also obvious. The interruption of the temperature decline and appearance of a warming in 1998 and 1999 is the main cause of the observed reduction of the strength of the O_3MD. In February–March of 1999, the ozone deficiency was, quite surprisingly, back to its normal pre-1976 level. However, with circulation changes causing a cooler stratosphere and a stronger polar vortex in the spring of 2000, the O_3MD substantially increased to $\sim$1500 Mt.

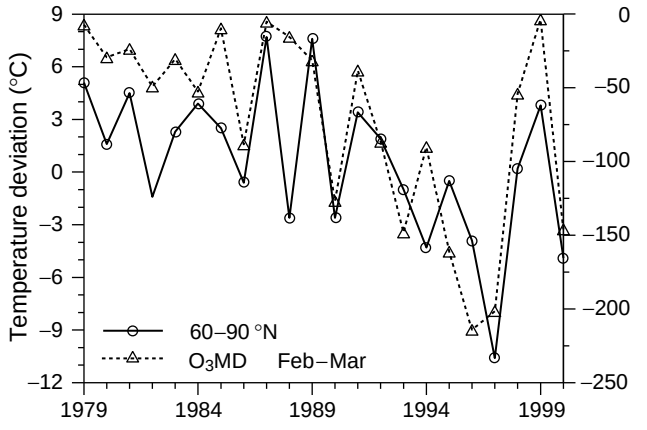


Figure 6 The Feb–Mar 50 hPa temperature deviations from the pre-1978 averages in the polar region (60–90 °N) – continuous line, and integrated for the same months and region O_3MD [Mt] (dashes with triangles) for the period 1979–2000

The O_3MD poleward of 35 °N, integrated for the first 105 days of each year, has increased from $\sim$2800 Mt in the early 1980s to an average of $\sim$7400 Mt in the last 11 years. It should be recalled that the O_3MD in the northern spring times of 1993 and 1995 exceeded 12 000 Mt. This quantity is comparable with the average O_3MD for the last 10 austral springs. Figure 7 shows the distribution of the O_3MD of February–March in three main zones: within the vortex, in the southern part of the middle latitudes (35–45 °N), and in their northern part (45–65 °N). It is clear that the northern belt is contributing in absolute value the largest portion of the O_3MD poleward of 35 °N. The strength of the deficiency in this belt has increased similarly to the intensification seen in the vortex, whereas over the more southern part of the middle latitudes the increase of the O_3MD is marginal. This part of the middle latitudes is also more frequently influenced by tropospheric transport from the sub-tropical belt.

Table 1 indicates the relative (normalized by area) contribution of the different latitudinal belts during the past 11 years. The area poleward of 35 °N is taken as unity, and then the O_3MD values over the three belts, and separately for within the vortex, are considered. The most interesting result is that the contribution to the entire O_3MD poleward of 35 °N, on a per unit basis of the vortex area, is more than four times greater than the contribution of the southern middle latitudes and twice as much as that of the band from 45–65 °N.

With this overview of the total ozone changes in the northern latitudes, it is interesting to examine where the observed ozone has declined in the vertical plane. Figure 8 shows the vertical variation of ozone trends (% per year) for 40–52 °N over the period 1979–1996 calculated from three different measuring systems. A variety of data sets are shown: ground-based data are Dobson/Umkehr measurements (triangles and vertical bars are averages for below 20 km and above 37 km); SAGE I/II satellite sunrise and sunset observations (diamonds); and BUV/2 satellite observations (circles), which are known with less confidence

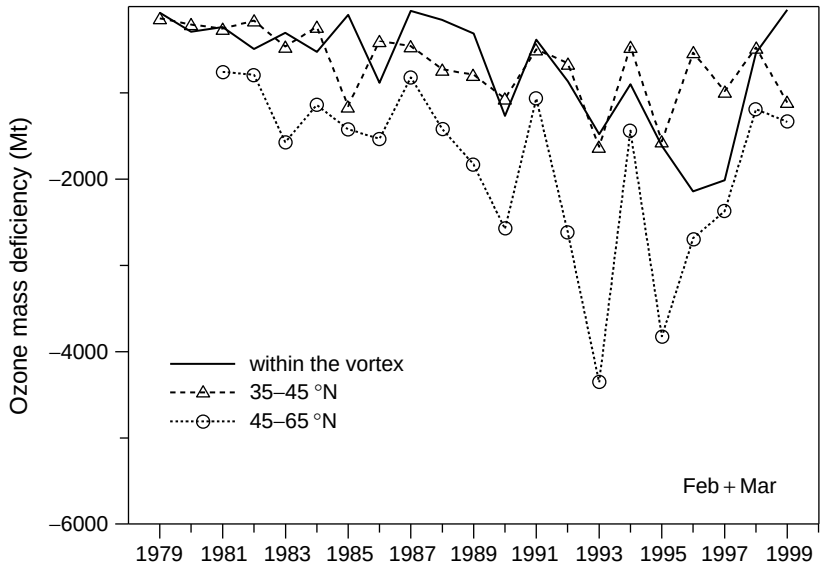


Figure 7 The O_3MD from the pre-1976 averages integrated for 1 Feb–31 Mar within the Arctic vortex and within the −10% deviation contours in the 35–40 °N and 45–65 °N belts during 1979–1999

Table 1 The top row gives the area of selected latitudinal belts as a fraction of the entire area poleward from 35 °N[a]

	Latitudinal domain				
Parameter	North of 35 °N	35–45 °N	45–65 °N	65–90 °N	Northern Hemisphere polar vortex
Area in %	100	33	45	22	variable
O_3MD for 1990–2000 (Mt)	4700	860	2310	1530	1160
O_3MD north of 35 °N (%)	100	18	49	33	25
Relative contribution to change in O_3MD north of 35 °N	1.0	0.5	1.1	1.5	>2.0

[a] The second and third rows give the O_3MD [in Mt] integrated for 105 days (1 January–15 April) within the −10% deviation contours in the specified belts, averaged for 1990–2000, and as % contribution to the entire O_3MD north of 35 °N. The bottom row gives the contribution of a 1% area of each belt (and of the vortex) to the overall O_3MD >35 °N, thereby indicating the relative contribution to the change by latitudinal band.

due to potential drift problems with this satellite. There is good agreement between the independent Umkehr and SAGE I/II measurements. All three systems show a general decline, with statistically significant peak values of −5 to −8% decade^{-1} at 40–45 km and the smallest decline at about 30 km. Balloon ozone-soundings also indicate that there is second maximum of ozone depletion of about −7% decade^{-1} in the lower stratosphere at about 15 km.

STRATOSPHERIC OZONE CHANGES OVER ANTARCTICA

Since the early 1980s, the total ozone over the southern polar region has exhibited significant declines. Each year, starting in the sunlit periphery of the Antarctic in August, but mainly from September through November, there has been a sharp ozone reduction, reaching ~40–50% in the 1990s as compared to its pre-1976 values. A number of field measurement campaigns and analysis programs have confirmed that chlorine-activated heterogeneous reactions in the stratosphere are responsible for the observed ozone destruction. The stratospheric temperature and enriched aerosol content (especially soon after volcanic eruptions reaching the stratosphere, like Mt Pinatubo in 1991) have been identified as key controls over the processes at the current burden of chlorine loading. They determine the extent of the depletion and, together with the dynamics, the overall surface and length of time the ozone hole persists.

Ozone loss is especially strong over the Antarctic because the winter circumpolar vortex prevents extensive air exchange with mid-latitudes. This produces very low stratospheric temperatures (below −80 °C) lasting for months,

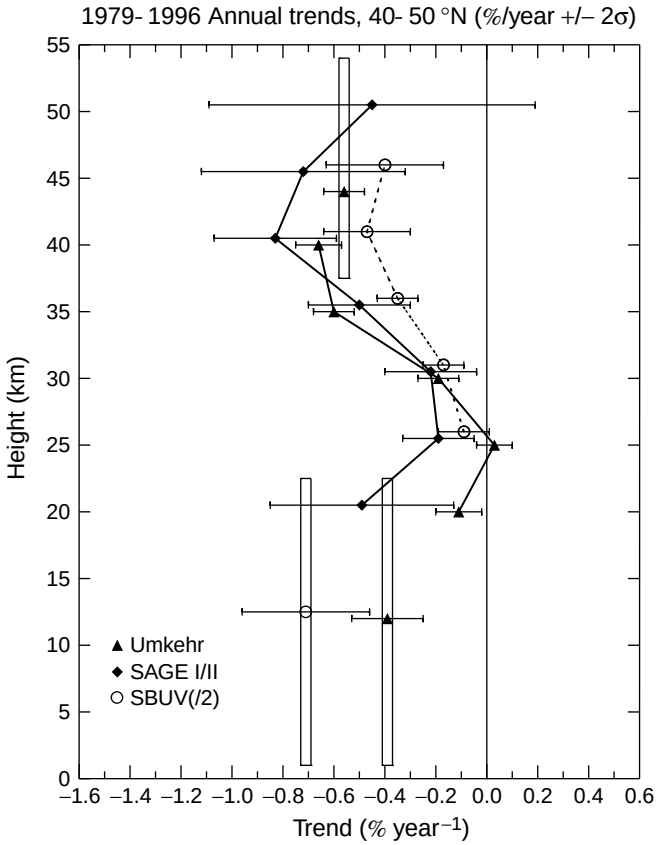


Figure 8 Ozone trends (% year^{-1}) at different altitudes deduced from three independent systems for the 40–52 °N belt over the 1979–1996 period. Horizontal bars show two standard deviations. (*Source:* Figures 4–32 from WMO Ozone Assessment-98)

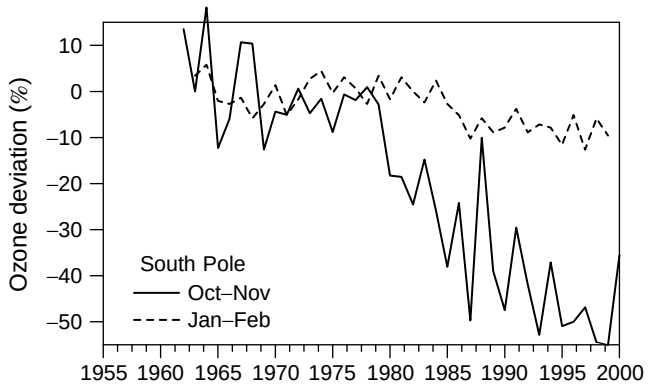


Figure 9 Ozone deviations from pre-1976 averages for Oct–Nov (solid) and Jan–Feb (dashed) over the South Pole (NOAA station) show rapid decline since the early 1980s, mainly during the late spring months

which favor the generation of PSCs of ice particles. Normally, chlorine and bromine are locked into stable reservoir compounds (such as $ClONO_2$, $BrONO_2$ and HCl). The ice particles attract water vapor and absorb nitrogen compounds, then fall to lower levels of the atmosphere, thereby dehydrating and denitrifying the air in the stratosphere. With the return of the sunlight in the early spring, these reservoir compounds are converted to active chlorine and bromine species on the surfaces of the PSC ice particles and/or on stratospheric sulfate aerosol. The released substances can split ozone molecules with amazing efficiency.

The drastic ozone decline over Antarctica was very strong during the 1990s and continues unabated. Figure 9 shows the deviations from the pre-1976 averages of the ozone values during two late spring and two summer months at the National Oceanic and Atmospheric Administration Amundsen-Scott station at the South Pole. In addition to typical interannual variability related to the specifics of the atmospheric circulation, the data show a drastic decline in October–November ozone levels, dropping 45–55% during most of the 1990s. During the summer months, the decline is about 9%. Over Antarctica the negative total ozone trends since 1979, in percent per decade, were for December–March, ~5%, for September–November ~22%, and ~9% for the annual average. By comparison, the negative trend of the annual averages in the 35–50 °S belt is only ~2% and in the 50–65 °S belt is ~4.3% per decade.

Observational and modeling studies indicate that the total ozone loss can be primarily ascribed to the depletion in the lower stratosphere. Syowa ozonesonde data from 69 °S provide information stretching back to 1968 and show that ozone depletion of ~90% occurred in the 12–20 km layer for a few consecutive weeks in September–October during the 1990s. Similar changes are observed over the Neumayer (70 °S) and South Pole stations. Springtime ozone depletion has been especially severe since 1992, reaching near complete ozone loss in the ~14–20 km layer from the end of September through mid-October. Such ozone losses in the 12–20 km layer are in agreement with modeling studies for the observed increase in stratospheric halogen concentrations from ~2.4–3.4 ppbv.

The decline of lower stratospheric ozone usually begins in early August on the sunlit edges of the vortex. The highest partial pressure, ~17 mPa before August, drops to lower than 1 mPa in the second half of September and early October. Temporal cross sections from the Neumayer and Syowa ozone soundings show a weakening of the polar vortex in November. Then, as a result of transport from the middle latitudes, ozone concentrations begin to increase, first above 25 km and then slowly propagate downward.

The ozone loss is the main reason for the temperature decline in the Antarctic lower stratosphere. In the 70–90 °S belt during September–October–November–December of the last 10 years, the temperature has dropped by about 2, 10, 12 and 5 °C, respectively, during these months. Lowering of the temperatures prolongs the persistence of the

strong stratospheric vortex. It is observed that the springtime stratospheric warming and the related vortex breakdown have been delayed by about a month during the last 10–15 years. Further stratospheric temperature declines, due to increasing emissions of greenhouse gases in the atmosphere, would facilitate even further ozone depletion, especially in the Arctic region, as shown in recent modeling studies.

Related to the ozone decline, a major change appears in the annual course of the total ozone over Antarctica. Comparing the average September–October–November values in 1985–1999 with the pre-1976 period, the total ozone was severely reduced: at Halley (76 °S) by 37–49%, at Syowa (69 °S) by 30–38% and at Faraday-Vernadsky (65 °S) by 25–31%. In December, the decline is 17–24%. As a result, the natural annual ozone maximum has been reduced by close to 30% and its occurrence shifted from November to December. The severe Antarctic decline has also affected the global ozone content by practically eliminating its secondary maximum, previously found in September–October as shown in Figure 1.

Ozone hole values (<220 matm-cm), representing a reduction of more than 30% from the pre-1976 averages, have covered an area not greater than 3–4 million km^2 at the time of ozone hole maximum during the early 1980s. This sunlit area regularly exceeded 10–13 million km^2 in the second half of the 1980s. In 1998 for the first time, the ozone hole covered an area greater than 25 million km^2 and persisted for more than 20 consecutive days. In September 2000, the ozone hole spread over such a large area for seven days. A weakening of the polar vortex in the beginning of November 2000 prevented repetition of characteristics similar to the previous two record setting seasons. In the springs of 1998 and 1999, ozone hole values prevailed for about 100 days over an area greater than 10 million km^2. This area is as big as Europe from the Gulf of Biscay to the Ural Mountains. The number of days with areas of 15 and 20 million km^2 also increased, which is indicative of strengthening of the ozone depletion in the 1990s (see Table 2). In the process of polar vortex rotation and expansion toward the lower latitudes, the very low ozone hole values were measured for a few days over the southern tip of South America (~56 °S) during each of the last 10 years.

The basic characteristics of the ozone hole each year since 1979 are shown in Figure 10. The maximum extent of the ozone hole area, which for the first time exceeded 10 million km^2 sunlit area in the mid-1980s, exceeded 22 million km^2 each year since 1990. The lowest ozone values within the ozone hole have usually occurred at the end of September and/or early October. The daily-minimum total ozone values have fallen from ~200 matm-cm in the early 1980s to ~150 matm-cm in the mid-1980s and down to values a few units less than 100 matm-cm in most of the 1990s. The date of disappearance of the ozone hole values

Table 2 Number of days when the ozone hole area (total ozone <220 matm-cm) covered a sunlit area surface greater than 10, 15, 20 and 25 [km^2 × 10^6] in the last eight years

	Number of days when ozone hole sunlit surface area (in millions of km^2) was larger than indicated			
Year	>10	>15	>20	>25
1993	79	52	39	1
1994	67	56	33	1
1995	75	68	36	–
1996	88	70	35	–
1997	73	56	25	–
1998	101	70	48	20
1999	98	79	50	–
2000	75	54	41	7

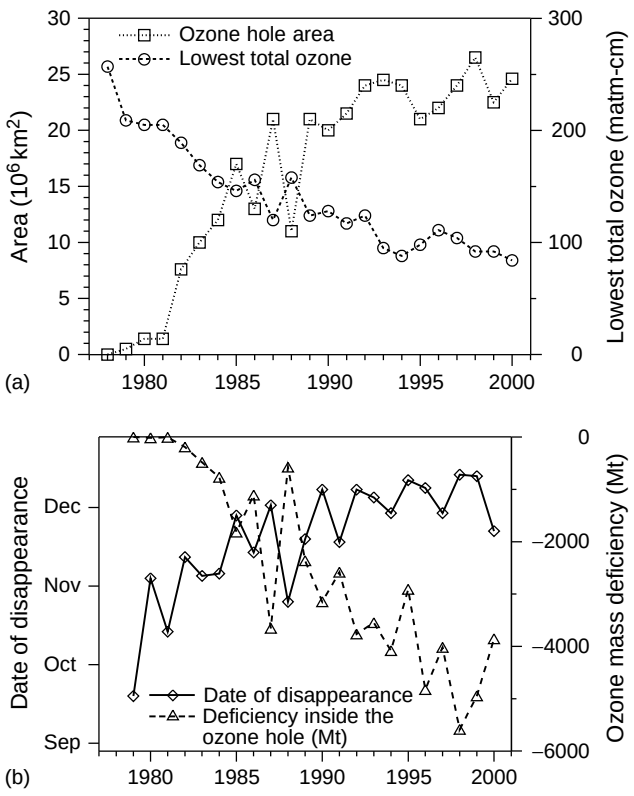


Figure 10 Evolution of basic characteristics of the ozone hole since 1979. (a) (□) maximum sunlit day-area covered by <220 matm-cm in [millions of km^2]; (○) lowest total ozone measured each year. (b) O_3MD [in Mt] as difference from pre-1976 averages integrated for 1 Sep–15 Dec (△); and date of disappearance of values less than 220 matm-cm (◇) [updated from Bojkov, WMO, 1999]

(<220 matm-cm) has been delayed from mid-October and early November in the first half of the 1980s, toward early December for most of the last 10 years. The date of disappearance is very closely related to the breakdown of the polar stratospheric vortex each spring.

Another important characteristic of ozone decline plotted on Figure 10 is O_3MD. This characteristic over different latitudinal belts or for different seasons is directly comparable on a molecule per molecule basis, independent of the different background levels. The O_3MD is calculated as deviations from pre-1976 zonal averages on a daily base using satellite observations homogenized over $3^\circ \times 5^\circ$ grid. If the deviations were found to be greater than -10% ($\sim$2 standard deviations for most locations), and knowing its value in matm-cm, it is a simple matter to determine the respective ozone mass. The O_3MD over the sunlit area poleward of 60 °S, integrated for the 1 September–15 December period of each year, shows a continuous increase within the ozone hole area since early 1980. The average O_3MD for the last six austral springs (September–October–November for 1995–2000) was $\sim$6400 Mt. In the first six years (1979–1984) this amount was only $\sim$2100 Mt.

Table 3 provides numerical expressions for the changes in area with deviations stronger than -10% (from the pre-1976 averages) in September–October–November and the corresponding integrated O_3MD for three time-periods. The average O_3MD for the last three springs (1998–2000) was close to six times stronger than for the years of initial ozone hole appearance (1979–1981). This measure is now 15% stronger than the previously record-setting 1992–1994 season! The increase of the O_3MD between the three periods is strongest in November followed by September, indicating expansion of the time with ozone destructive processes. Looking at the summary line of Table 3, it is clear that between the two end periods, the area with anomalies greater than -10% increased by almost a factor of 2.5. The O_3MD increased five times to 1992–1994 and nearly six times to 1998–2000, demonstrating an intensification of the ozone depletion per unit area in the later periods.

Table 3 Area (10^6 km^2) with deviations greater than -10% poleward of 60 °S (left entry), and O_3MD (in Mt) over these sunlit areas integrated for each of September, October, November (right entry, in italics) for three different time-periods

	Area and O_3MD for three different periods		
Month	1979–1981	1992–1994	1998–2000
September	330/*360*	670/*1770*	800/*2300*
October	370/*470*	740/*2010*	800/*2320*
November	230/*270*	700/*1720*	680/*1820*
Average	930/*1100*	2110/*5500*	2280/*6440*

Figure 11 compares the values of O_3MD over the southern polar region with values over the southern middle latitudes. In the 61-day period (1 September through 31 October) the integrated values of O_3MD over the middle latitudes show great variability but have grown by very little since the mid-1980s. Contrary to this stability, in the polar region, as already discussed, the O_3MD has continued to increase even during the late 1990s.

It is also important to provide perspective on events in the Northern Hemisphere. Over the mid-latitudes (35–55°) the O_3MD in the two hemispheres was about equal in the early 1980s, but in 1993–1998 the deficiencies over the Northern Hemisphere mid-latitudes have been significantly greater ($\sim$30%). This fact is also supported by the smaller negative trends calculated for the Southern Hemisphere mid-latitudes mentioned earlier. The stronger decline over the Northern Hemisphere mid-latitudes could be related to the sequence of colder springs for most of the 1990s and more extensive meridional exchange of chemically disturbed air from the Arctic vortex to the sunlit middle latitudes. The latter is in contrast to a predominantly zonal circulation over the Southern Hemisphere mid-latitudes and the nearly completely isolated Antarctic vortex, but this is not completely understood and requires further study.

The O_3MD ratio is generally opposite over the two polar regions (>60°). At the beginning of the rapid ozone decline in the early 1980s, Antarctic O_3MD was double that of the Northern Hemisphere polar region. Even for the last six years, it remains more than one-third greater than for the Arctic. If the O_3MD is taken relative to the actual surface area of the -10% deviation contours during the respective spring period in both hemispheres, the total ozone over the last six years is $\sim$120 matm-cm less over the

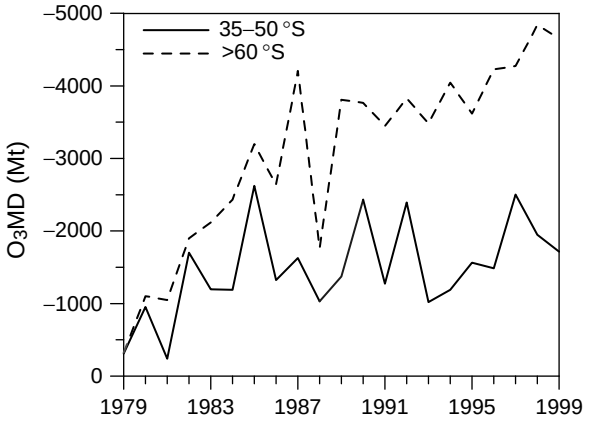


Figure 11 Ozone-mass deficiency (O_3MD) from pre-1976 averages, inside the -10% deviations contours, integrated for 61 days (1 Sep–31 Oct) in Mt for Southern Hemisphere polar region (dashes) and middle latitudes (continuous line) in the period 1979–1999. the data show a major O_3MD increase only over the polar region [updated from Bojkov, WMO, 1999]

Antarctica and ~80 matm-cm less over the Arctic. It should be remembered that the pre-1976 winter–spring averages over Antarctica were ~300–320 matm-cm, and over the Northern Hemisphere polar region, ~420–450 matm-cm. Therefore, the actual total column ozone thickness in the northern polar region is still, on average, well above the conventional ozone hole values of <220 matm-cm.

RADIATIVE EFFECTS OF OZONE CHANGES

Both ozone and halocarbons are greenhouse gases. Much like carbon dioxide (CO_2), they intercept and re-radiate the Earth s outgoing infrared radiation, thereby helping to warm the lower atmosphere. However, neither ozone nor its changes is uniformly distributed. As a result, ozone s radiative forcing is more complicated than those of the other principal greenhouse gases that have a long life span, and are therefore well mixed throughout the global atmosphere. To characterize a potential for climatic effects, radiative forcing (in units of $W\,m^{-2}$) is frequently used. It estimates the net solar and thermal infrared radiance at the tropopause level. These estimates are much less dependent on uncertainties concerning the combined interactive role of clouds, aerosols or oceans than for predicting surface climate change. A positive value signifies increased energy retention, producing a warming influence; a negative value implies greater energy loss, producing a cooling influence. In general, ozone increases in the troposphere, especially near the tropopause, induce warming, whereas ozone decreases in the stratosphere induce cooling. The observed temperature decline (0.6–0.8 °C per decade) in the 12–20 km layer over the middle latitudes during the past 25 years seems to be consistent with expected radiative impacts of a stratospheric ozone decrease. In the Antarctic in October–November, the decline is much stronger (10–12 °C).

An accurate assessment of the radiative effect of ozone changes is limited by the lack of detailed information on the variation in vertical distribution of ozone with latitude and longitude. However, recent calculations support earlier conclusions that stratospheric ozone depletion in recent decades has resulted in a negative radiative forcing of about $-0.2\,W\,m^{-2}$ (causing a cooling effect on the climate). Thus, stratospheric ozone change has offset, by about 30%, the positive greenhouse forcing due to increases in all other gases in the same period.

The Northern Hemisphere mid-latitude increase of tropospheric ozone since pre-industrial times is estimated to be $0.35\,W\,m^{-2}$, which has enhanced the total greenhouse forcing by as much as 15–20%. The net global mean radiative forcing arising from ozone changes is likely to have been positive since pre-industrial conditions, contributing about 15% of all the greenhouse gas-caused increase of radiative forcing in that period. Such changes could have an impact on the radiative balance of the Earth-atmosphere system and the thermal structure of the atmosphere; this could cause as yet unpredictable changes to atmospheric circulation patterns.

FUTURE OF THE OZONE LAYER

The results of nearly three decades of research focused on the ozone issue have provided a progressively better understanding of the interaction of human activities and the chemistry and physics of the global atmosphere. New policy-relevant insights of the roles of trace atmospheric constituents have been conveyed to decision-makers in governments and industry through the WMO/United Nations Environment Programme coordinated ozone assessment process (see references).

Large reductions in the production and atmospheric release of ozone-depleting substances (O_3-DS) have been achieved by international regulations. Without the Montreal Protocol and assuming a (conservative) 3% annual growth rate in production, O_3-DS would have led to an equivalent chlorine loading of around 17 ppbv in 2050 and massive ozone depletion (50–70% of pre-1976 levels). Furthermore, the impacts would have continued to grow in the years beyond 2050. This is schematically shown in Figure 12. The original Montreal Protocol (1987) would have reduced

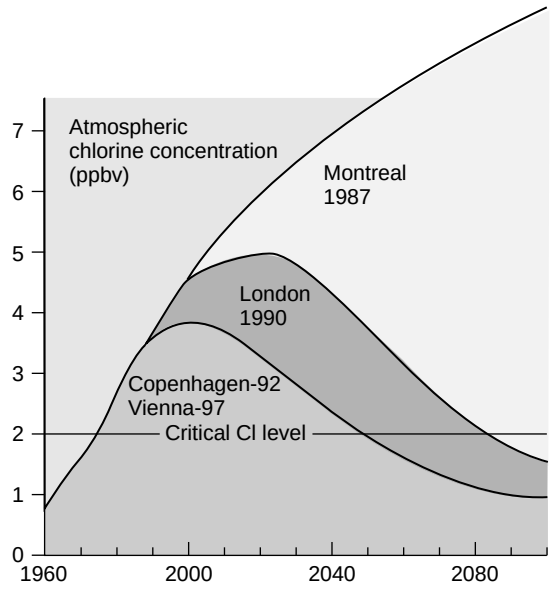


Figure 12 Schematic presentation of measured chlorine-equivalent concentrations in the stratosphere and projections according to various Montreal Protocol requirements for reductions and/or phase out of CFCs, halocarbons, and other O_3-DS. The ozone hole appeared when chlorine equivalent concentrations reached ~2 ppbv, and severe ozone destruction became predominant during polar springtimes when levels approached ~3 ppbv. (*Source:* Bojkov, WMO, 1995)

the O_3-DS concentration to ~9 ppbv. The London amendments (1990) are expected to lead to a reduction to about 4.6 ppbv, and the amendments and adjustments of Copenhagen (1992), Vienna (1995) and Montreal (1997) have further reduced the projected level to about 2.0 ppbv around the year 2050 (approximating the 1980 abundance, when the ozone hole first appeared).

Following the ozone losses of the 1980s and 1990s, recovery, in the sense of there being a negligible impact of chlorine compounds, is unlikely to occur before ~2050, when chlorine levels are expected to have been reduced considerably (<2 ppbv). At that time, the Antarctic ozone hole is projected to become as small as in the early 1980s or even disappear entirely. By that time, however, the recovery will be to a different atmosphere, in the sense that increases in greenhouse gases will have cooled the stratosphere, resulting in a reduction in homogeneous ozone loss, while increases in N_2O may have increased tropospheric NO_x to counteract this effect. Changes in H_2O will likely have complicated effects due to both its radiative and chemical impacts. Thus, due to changes in temperature, transport, water vapor and nitrogen species, the ozone vertical distribution is likely to be different in the year 2050 from that during the period before substantial ozone depletion was observed. Based on past emissions of O_3-DS and a projection of the maximum allowances of emissions under the Montreal Protocol, maximum ozone depletion is estimated by 2- and 3-dimensional modeling studies to occur within the next two decades, which will be the most vulnerable period for ozone depletion.

However, identification of, and the evidence for, the recovery of the ozone layer lie still further off. Global ozone will start to recover slowly within coming decades towards its pre-1980 level once the stratospheric abundances of ozone-depleting (halogen) gases decrease below ~3 ppbv. Because of the slow rate at which natural processes remove these compounds from the stratosphere, the reduction of total chlorine and bromine abundances in the stratosphere will be much slower than the rate of increase observed in past decades.

The main drivers of ozone change, however, could be natural and anthropogenic processes unrelated to chlorine and bromine compounds, but to which the ozone layer is sensitive because of the elevated abundances of O_3-DS. For example, the increases in N_2O and aerosol amounts will delay recovery of ozone while increases in CH_4 will have the opposite effect. Increases in NO_x emissions by supersonic aircraft in the mid-stratosphere will lead to ozone losses by gas phase destruction. Increases in aerosols provide sites for heterogeneous chemistry leading to ozone loss in the presence of halogen compounds. In contrast, increases in CH_4 lead to increased production of HCl and a reduction in the amount of active chlorine-catalyzed ozone loss. Furthermore, extreme perturbations (e.g., natural events, such as volcanic aerosols reaching the stratosphere, and/or extreme cold temperatures) could increase the loss from ozone-depleting chemicals. Arctic ozone depletion could be increased by cold polar stratospheric temperatures ($<-76\,^{\circ}$C), which, in turn, could be linked to increase of greenhouse gases or to natural temperature fluctuations.

The effect of increasing fleets of sub- and super-sonic aircraft with flight altitudes of 9–13 and 17–20 km, respectively, is sometimes listed as a factor in increasing of the tropospheric ozone and of delaying of the ozone recovery respectively (IPCC, 1999). The principal emissions of these aircraft are CO_2, CH_4, NO_x, SO_2, H_2O and particles. Their effects are highly dependent on the altitude of injection due to the changing chemical and dynamical regimes in the lower stratosphere. The available modeling results summarized below have considerable uncertainty.

Subsonic emissions result in a significant increase in the concentrations of NO_x (factor of about two increase at 12 km) and a decrease in the lifetime of CH_4 as a result of increases in tropospheric OH concentration. In all models, ozone is projected to increase, but the impact is restricted to the most frequent flight corridors over northern mid- and high-latitudes. For the latitude band 30–60 °N, the estimated maximum average tropospheric ozone increase at cruise altitudes (above background with no aircraft) is 6% for the 1992 level of emissions and 13% for 2050 emissions. The response is nearly linear, with changes in NO_x emission. The effects on the column of ozone are an increase by 0.4% in 1992, and by 1.2% in 2050.

The assessment of the impact of a projected fleet of up to 1000 super-sonic aircraft for the year 2015 atmosphere is a slightly negative (−0.4%) column ozone change, reflecting a negative change of 40–60 ppbv at 30 km and a positive change of 15 ppbv at 20 km.

There will also be effects from increasing water vapor and sulfur emissions in the stratosphere. The response to sulfur emissions depends on how much of the fuel sulfur is converted to sulfate particles in the aircraft exhaust plume, with larger ozone loss from larger conversion to aerosols. The overall impact on ozone of this probably unrealistically large projected fleet of supersonic aircraft is small compared to perturbations due to halogens.

Because there is less variability in the ozone loss, the detection of the onset of ozone recovery from halogen-induced depletion should be possible earlier in the Antarctic than in the Arctic or globally. In particular, reduction of the rates of ozone destruction in the 12–20 km layer during the Antarctic spring would be an early indicator of the beginning of the ozone recovery. However, this would not be expected before the chlorine

loading declines to less than 3 ppbv, which requires a decade or two according to current Montreal Protocol requirements.

Despite the uncertainties regarding the modeling simulations, some delay in Arctic ozone recovery is likely because of the impact of climate change on lowering the stratospheric temperature. Climate-induced changes are projected to lead to increased production of PSCs in the lower stratosphere and to enhanced heterogeneous reactions. Results from 3-dimensional models project the same or greater ozone depletion in the Arctic in 2015 as was observed around 1995, despite the assumed reduction in chlorine loading of between 3 and 16%.

Because the issues of ozone depletion and climate change are interconnected, so too are the Montreal and Kyoto Protocols. That the ozone depletion and climate change phenomena share a number of common physical and chemical processes means that changes in ozone concentrations will affect the Earth s climate and changes in climate and meteorological conditions will affect the ozone layer. Decisions taken – or not taken – under one Protocol, therefore, impact the aims of the other. For example, decisions made under the Kyoto Protocol with respect to CH_4, N_2O and CO_2 will affect the rate of recovery of ozone, while decisions regarding controlling hydrofluorocarbons (used as substitutes for some CFCs) may affect decisions regarding the ability to phase out O_3-DS.

Although the Montreal Protocol is reducing O_3-DS, springtime depletion of ozone in the Antarctic continues unabated at the same levels as observed in the early 1990s and large depletions of ozone have been observed since 1990 in the Arctic in most years characterized by unusually cold and protracted winters. The total stratospheric loading of O_3-DS did reach its maximum at about year 2000; however, it will stay at rather high levels (>3 ppbv) for many years. Therefore, the ozone layer is at present, and will, during the coming decade, be in its most vulnerable state.

REFERENCES

IPCC (Intergovernmental Panel on Climate Change) (1999) *Aviation and the Global Atmosphere*, eds J E Penner, D H Lister, D J Griggs, D J Dokken, and M McFarland, Geneva, 1–373.

WMO (1988) *Report of the International Ozone Trends Panel*, Global Ozone Research and Monitoring Project, Report 18, Geneva.

WMO (1994) *Scienti c Assessment of Ozone Depletion*, Global Ozone Research and Monitoring Project, Report 37, Geneva.

WMO (1998a) *IO$_3$C-SPARC Assessment of Trends in the Vertical Distribution of Ozone*, Global Ozone Research and Monitoring Project, Report 43, Geneva.

WMO (1998b) *Scienti c Assessment of Ozone Depletion*, Global Ozone Research and Monitoring Project, Report 44, Geneva.

Stratosphere, Temperature and Circulation

Karin Labitzke

Meteorologisches Institut der FU-Berlin, Germany

In the Arctic summer, the temperatures are higher than in winter at both the tropopause and the stratopause, but this is not so above 58 km, where the mesosphere is extremely cold in summer and where Noctilucent clouds can form. The Arctic troposphere is cold in winter, and the stratosphere too, but the tropopause is not well marked since the temperature often continues to fall into the lower stratosphere, though at a slower rate than in the troposphere. On the lee side of mountains, especially in Scandinavia, Mother-of-Pearl clouds may form where the temperature is below $-80\,^{\circ}C$; nowadays such clouds are called Polar Stratospheric clouds.

MEAN CONDITIONS IN THE STRATOSPHERE

Global Temperatures

Figure 1 shows the average vertical profile of temperature in three instances; the temperature can be appreciably higher or lower on individual days. The dashed curve is for the equator, where the tropopause is cold and lies at high elevation. The two other curves depict conditions in the Arctic, the solid line in summer and the dotted line in winter. The distribution of ozone with height in February at about $50\,^{\circ}N$ is shown in the inset at the lower right-hand side of the figure; the units are in nanobar (nb), a measure of the partial pressure of ozone (Labitzke and van Loon, 1999).

The vertical temperature profile is shown in another manner in Figure 2, which is a latitude versus height section where the temperature averaged around latitude circles (the zonal mean) in January is shown from pole to pole (Barnett and Corney, 1985). Certain associations are more easily seen in such a presentation: for instance, the tropopause lies at different elevations in different latitudes; in the tropics and subtropics it is found between 15 and 18 km, whereas it is as low as 7 to 9 km in middle and high latitudes.

The stratopause lies, on average, at a height of almost 50 km, at a pressure of about 1 hPa. Only above the winter pole (Arctic, left in Figure 2) is the stratopause on average markedly higher. The temperature in the stratopause region ranges from $-18\,^{\circ}C$ above the winter pole to $+12\,^{\circ}C$ above the summer pole (Antarctica, right in Figure 2). These are all average conditions. The deviations from the mean can be appreciable for individual situations; for instance, during a major warming in the northern stratosphere in winter, when

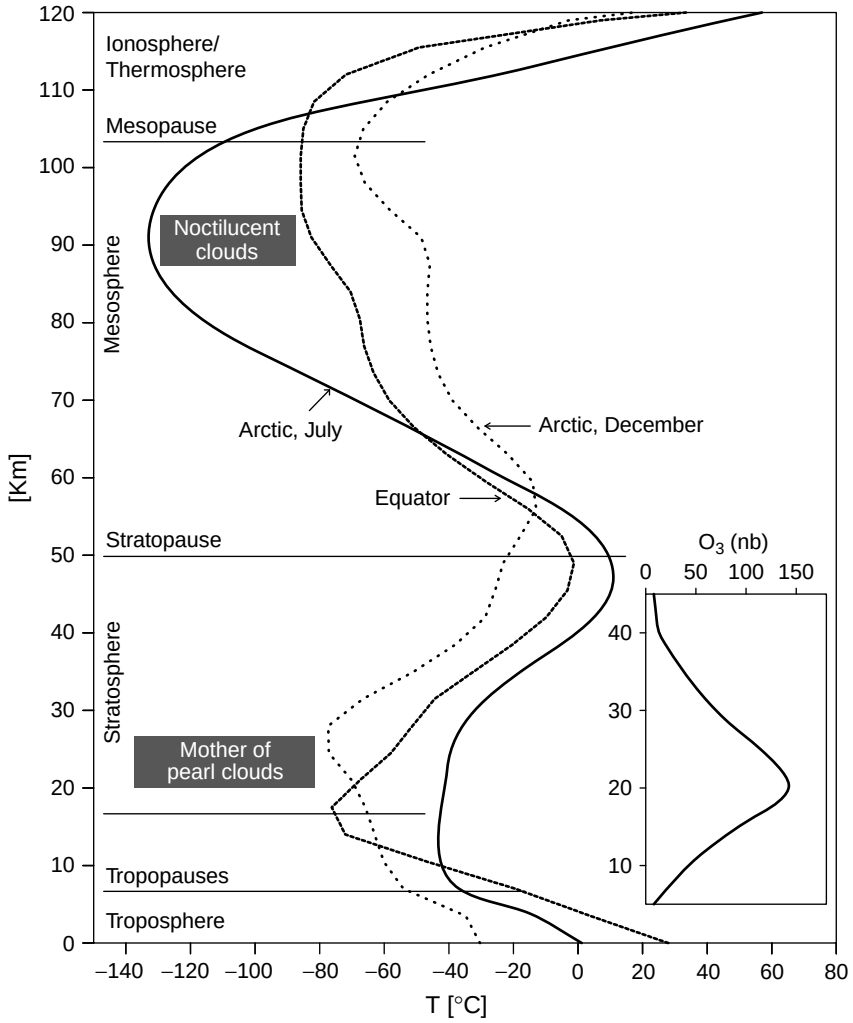


Figure 1 The layers of the atmosphere (Labitzke and van Loon, 1999)

the stratopause lowers and temperatures as high as $+49\,^{\circ}C$ have been measured at 40 km, by meteorological rockets and lidar (Scherhag, 1952; Labitzke, 1972; von Zahn *et al.*, 1998). More details on this curious phenomenon are contained in the last subsection in this article.

There are other important facts about the winter pole that must be mentioned: the temperature at high latitudes does not rise immediately above the tropopause, as it does elsewhere; in these latitudes it falls further with height to a level near 25 hPa (about 25 km) but more slowly than in the troposphere (compare Figure 1). The temperatures in the winter stratosphere over Antarctica are substantially lower than those above the Arctic in winter, (see Figure 3).

Global Meridional Circulation

Differences in temperature are always associated with pressure differences, which the atmosphere tries to even out by moving air from high to low pressure. Due to the deflection of air motion caused by the rotation of the Earth (the so-called Coriolis force, *see* Coriolis Effect, Volume 1), the air does not move directly from high to low pressure around the center of low pressure. The attempt of the atmosphere to eliminate the temperature and height differences is thus responsible for the strong west and east winds which encircle the globe.

If one averages over a longer time interval, it turns out that there is also a north–south gradient in the troposphere. This gradient is responsible for the south–north circulation cells in the vertical plane that meteorologists call meridional cells. Hadley (1735) described the general features of these cells by means of numerous observations from sailing ships: warm moist air rises in the tropics, expands and cools as the pressure decreases upward. The moisture condenses to form high reaching clouds that sometimes build upwards to the tropopause in what is known as the Intertropical Convergence Zone (ITCZ), where the trade

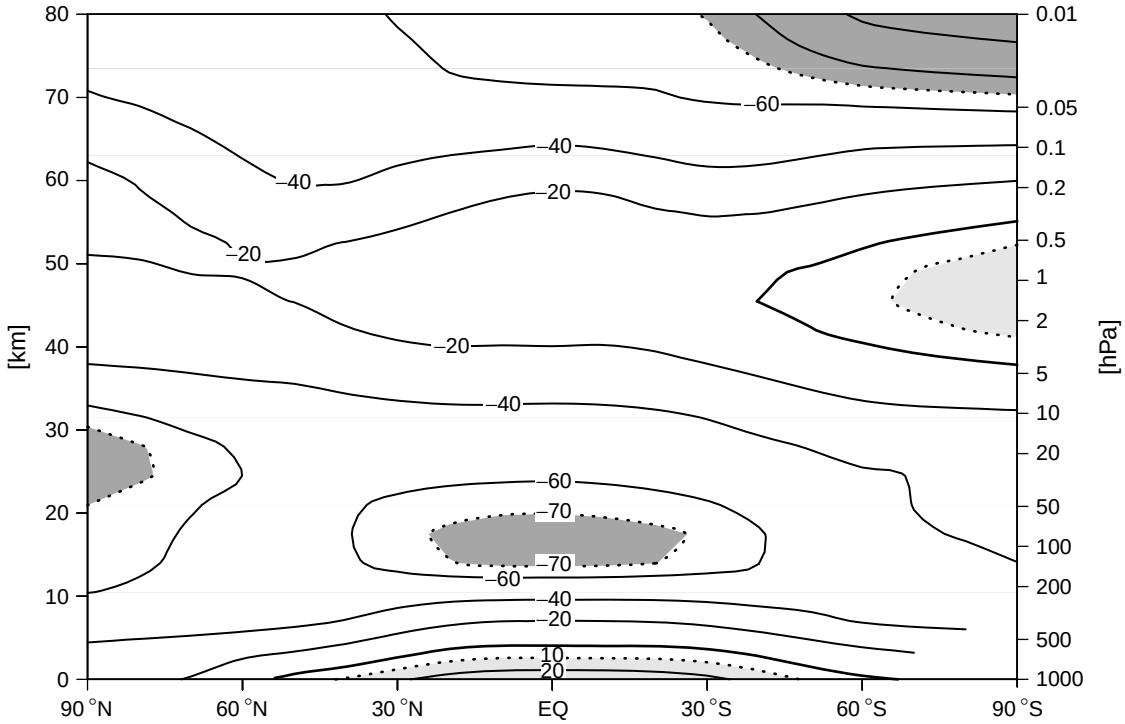


Figure 2 The mean distribution of zonally averaged temperature (°C) with height in January – winter in the Northern and summer in the Southern Hemisphere. The ordinate is in km on the left and in pressure (hPa) on the right (Barnett and Corney, 1985)

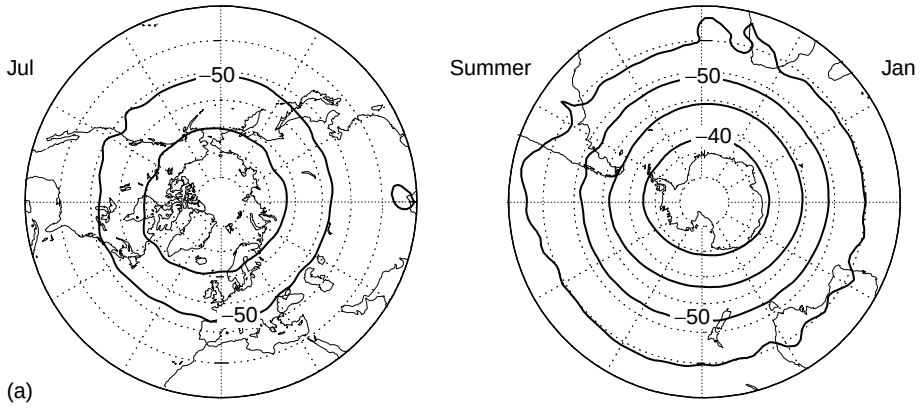


Figure 3 Monthly mean maps of the 30-hPa temperatures (°C); (a) summer, (b) winter, (c) spring over both hemispheres. Interval between isotherms is 5 °C. Data: NCEP/NCAR reanalyses (Kalnay *et al.*, 1996), 1968–1997; (d) monthly mean maps of total ozone (Dobson Units, interval 30 units) March, Northern Hemisphere (left), and September, Southern Hemisphere (right); mean total ozone based on Total Ozone Mapping Spectrometer data from National Aeronautics and Space Administration, period: 1979–1992. Polar stereographic projection, outer limit is 20 °latitude. (Adapted from Labitzke and van Loon, 1999)

winds of the two hemispheres clash. The ITCZ moves with the Sun and brings the rainy season to the various tropical regions. In the upper troposphere the air moves towards the poles in the meridional cell but it sinks in the subtropical high pressure areas, is compressed, and dries out; this is where deserts such as the Sahara and the Kalahari form. North of the subtropical highs in the troposphere are the familiar, mainly eastward-moving high and low pressure weather systems (*see* Atmospheric Motions, Volume 1).

Dobson (1930) realized through his early observations of ozone that there had to be a meridional circulation in the stratosphere as well, in order to bring the ozone created

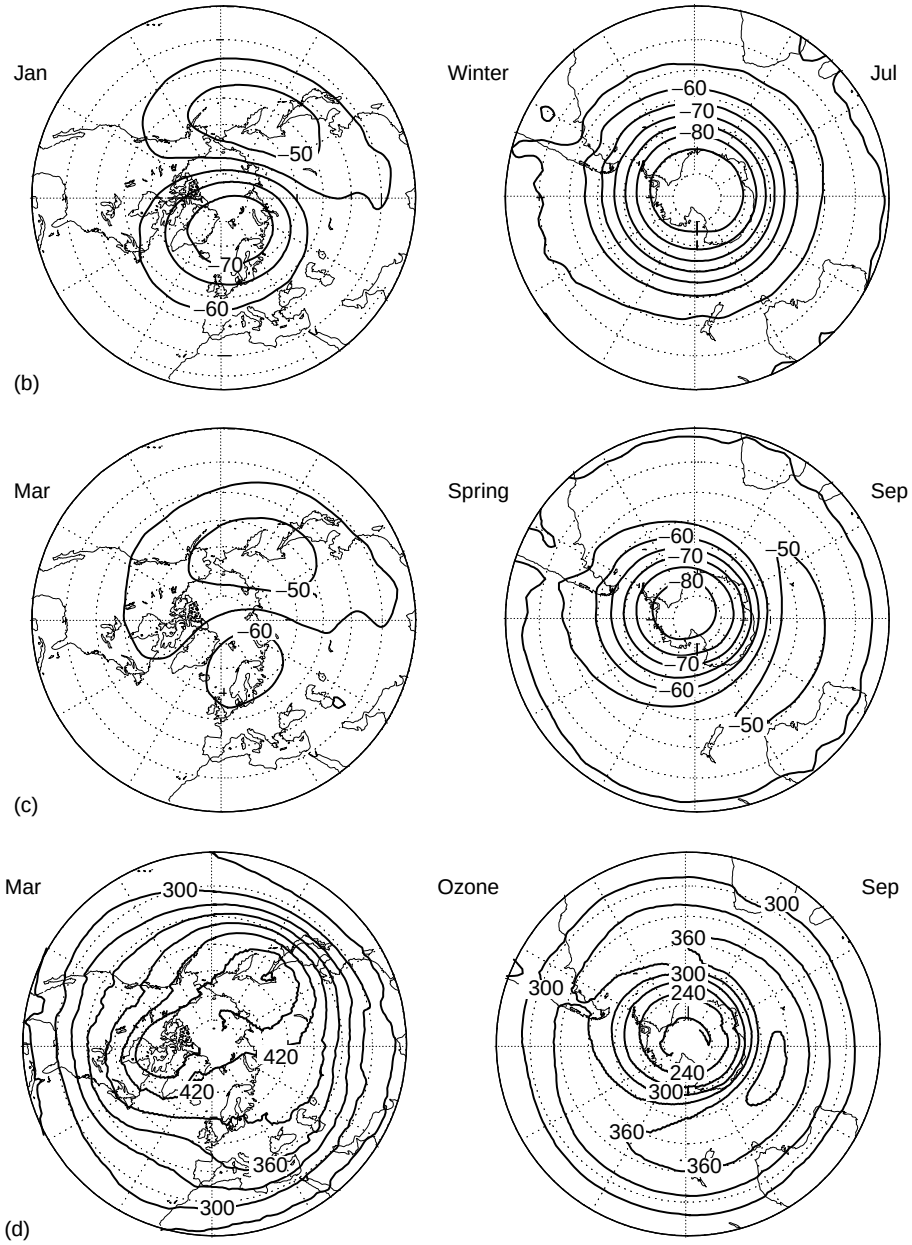


Figure 3 (*Continued*)

in the tropical stratosphere to higher latitudes. Today we know that the meridional circulation in the stratosphere (the Brewer-Dobson Circulation) is responsible for the transport of long-lived trace elements, for instance the chlorofluorocarbons (CFCs). These trace gases manage to enter the tropical stratosphere and from there to reach the polar regions in winter. It is in this way that trace gases produced in the Northern Hemisphere, like the CFCs, reach Antarctica where they destroy the ozone in spring (*see* Depletion of Stratospheric Ozone, Volume 1; Ozone Hole, Volume 1.).

The air that rises in the tropics is entrained into a large, poleward directed circulation through middle to high latitudes, where it sinks in winter; and the compression causes the sinking air to warm. This explains why the arctic mesosphere is warmer in winter than in summer (Figure 1).

The meridional cells are mainly driven by gravity waves of different scales that, under certain conditions, propagate from the troposphere to the stratosphere and mesosphere, deposit their momentum there and influence the zonal and meridional circulations.

Monthly Mean Temperature Maps

Extreme Seasons

The maps in Figure 3 show the temperature distribution at the 30-hPa level averaged over 30 years. The maps are shown for summer, winter and spring, for both hemispheres, and the 30-hPa level was chosen because this is the highest level for which reliable data are available for both hemispheres for a time span of about 30 years. Also, this is a level in the middle stratosphere where the most important features of the stratospheric temperatures are obvious and where the hemispheres can be suitably compared.

The temperatures during summer are shown in Figure 3(a). The high temperatures measured during this time of year are due to absorption, by ozone, of ultraviolet radiation coming from the Sun, and as there is little influence on the stratosphere from below, the distribution of temperature is very symmetric, parallel to the latitude circles. The temperatures are very similar over both hemispheres, but the air over Antarctica is somewhat warmer. This is due to the fact that the Earth is in perihelion (closest to the Sun) during the southern summer and thus receives more insolation; the south polar stratosphere, therefore, can absorb more short-wave radiation than the Arctic during the northern summer when the Earth is in aphelion (farthest from the Sun).

The transition from summer to fall is similar over both hemispheres (not shown). With the approach of winter, the picture changes (Figure 3b): although the sinking of air over the polar regions continues through the winter, in the stratosphere this warming is more than offset by the radiational cooling in the polar night. Over both polar regions, we find temperature minima during this part of the year that are surrounded by belts of higher temperatures. However, there are important differences between the polar regions: (1) the Antarctic polar region is much colder, and at the South Pole the mean temperature for July (midwinter) is close to $-90\,^{\circ}$C, which is $15\,^{\circ}$C lower than the mean over the Arctic; (2) the pronounced asymmetry of the northern stratosphere is not present in the southern winter stratosphere.

The reason is that the tropospheric winter jet stream in the western North Pacific is stronger than that in the western South Pacific and that, therefore, the sinking of the air in the stratosphere on the poleward side of the jet is more intense in the North Pacific. This leads to the warming over the Aleutian–Kamchatka region (Figure 3b, left) and to the build up of the so-called Aleutian high, which is the mark of large planetary waves that transport heat into the polar region, and help to ameliorate the cooling caused by the outgoing radiation. The Antarctic stratosphere does not have the same possibilities to retard or offset the cooling, and it therefore cools continuously during the polar night.

Spring

The transition from the winter (cyclonic) to the summer (anticyclonic) circulation in the stratosphere (final warming) takes place about two months earlier in the Arctic than in the Antarctic. The late interseasonal transition in the Antarctic – after the Sun has returned to the polar region – produces the meteorological conditions that are necessary for the formation of the ozone hole over the Antarctic in spring (*see* Ozone Hole, Volume 1). A comparison of the temperature (Figure 3c) and ozone (Figure 3d) distribution on the two hemispheres in spring (March and September) shows the marked differences between the polar regions: in September the temperature is still extremely low over the Antarctic when the Sun returns and enables the ozone-destroying chemistry to take place, and we find very low values of total ozone over Antarctica. In contrast, the mean temperature above the Arctic in March is high and planetary waves, which are themselves responsible for the higher temperatures, have also transported much ozone into the polar region.

There are, of course, also very late transitions (final warmings) in the Arctic, and in such a year, ozone is destroyed over the Arctic as well, but the total ozone values remain significantly higher than observed over Antarctica.

Interannual Variability

The standard deviations (scatter about the mean) of the monthly mean 30-hPa temperatures averaged along latitude circles in Figure 4 serve as measure of the interannual variability. This variability is small in the tropics and on the summer hemisphere, but increases with latitude and season. It is largest in the Arctic in winter, and then much larger than in the Antarctic winter. The larger arctic variability is related to larger wave activity there, and to the frequent occurrence of warmings during the northern winters, (see Figure 5).

During winter, the wave activity is large in southern middle latitudes as far south as $60\,^{\circ}$S, and the influence of the waves becomes noticeable in the Antarctic polar region after August. When the final warmings take place in the Antarctic spring (see previous), the interannual variability increases there, and it is then larger than in the corresponding months on the Northern Hemisphere.

Variability During the Arctic Winters

The monthly mean 30-hPa temperature at the North Pole in January–February ranges from below $-80\,^{\circ}$C to $-54\,^{\circ}$C, so we are dealing with a frequent change between warm and cold winters (Figure 5). The variability is the largest observed globally at this pressure level, (compare Figure 4). This variability is due to the occurrence or non-occurrence of stratospheric warmings. The World Meteorological Organization recognized the importance of these stratospheric

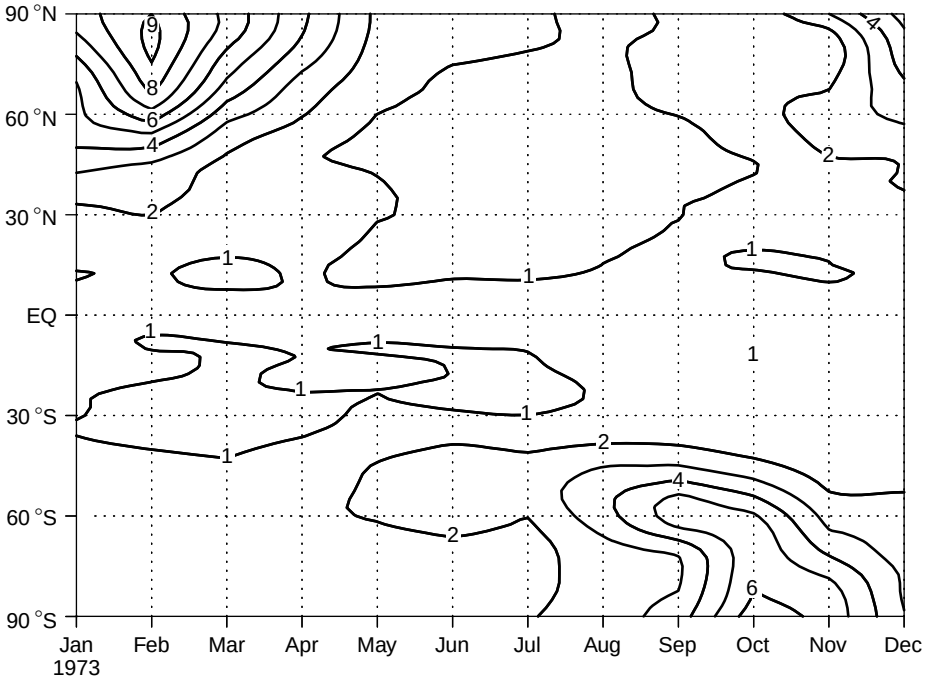


Figure 4 The global distribution of standard deviations (°C) of the 30-hPa monthly mean temperatures for the period 1973–1995, based on NCEP/NCAR reanalyses (Labitzke and van Loon, 1999, Figure 2.11)

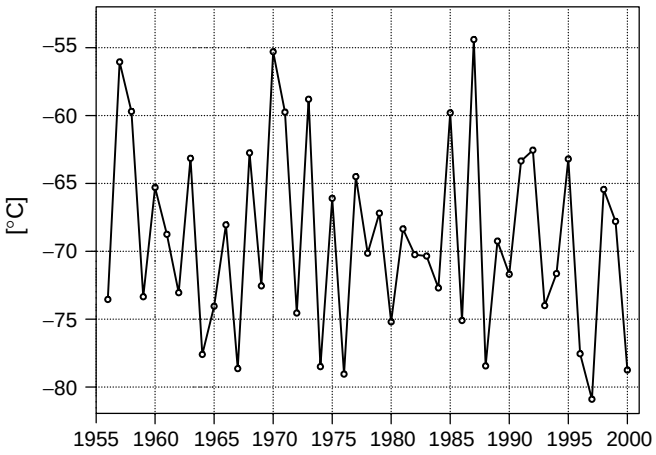


Figure 5 Time series of the 30-hPa temperature (°C) at the North Pole in January–February, 1956–2000; data: Meteorological Institute, Free University Berlin (Labitzke and van Loon, 1999, Figure 3.1, updated)

phenomena, and in 1964 asked a few meteorological institutes to issue sudden warming advisories regularly during winter, the so-called stratalerts. This alerting system is still ongoing. Four different types of warmings are considered:

- *Major midwinter warmings* occur mostly in January–February. In addition to warming of the north polar region and reversal of the temperature gradient, they are also associated with a breakdown of the polar westerly vortex, which is replaced by a high. That is, the definition of a major midwinter warming requires not only the warming but also a total change of circulation. The definition of a breakdown of the polar vortex is that the usual westerlies in the Arctic at 10 hPa are replaced by easterlies, so that the center of the vortex moves south of 60–65 °N. The vortex is either displaced entirely or split into two. This type of warming has not been observed in the Antarctic.
- *Minor warmings* can indeed be intense and sometimes also reverse the temperature gradient, but they do not result in a reversal of the circulation at the 10-hPa level. They are found in the Antarctic as well.
- *Canadian warmings* often happen in early winter. They take place when the Aleutian stratospheric high intensifies and moves poleward. The Canadian warmings can reverse the temperature gradient and sometimes briefly change the wind direction over the polar cap, but they do not, however, lead to a breakdown of the cyclonic polar vortex.
- *Final warmings* mark the transition from the cold cyclonic vortex in winter to the warm high centered on the pole in summer. Their intensity varies by a considerable amount, and they can be divided into major and minor final warmings. The time when the final warmings take place – when the westerlies of winter are replaced by the easterlies of summer – also varies a good deal, so they are further

divided into early and late final warmings. Naturally, there are also final warmings on the Southern Hemisphere.

What causes this high variability in the arctic stratosphere, or what makes a winter extremely warm or cold? We can still supply only a preliminary answer to this question, because one can only guess at the interaction between the various forcings from outside the stratosphere, and at the role played by the random variability inherent in the internal dynamics of the troposphere and stratosphere.

Meteorologists agree that planetary-scale waves, under given conditions, propagate from the troposphere into the stratosphere, and also that the explanation of why these waves are enhanced in the stratosphere must be sought in the troposphere. It is also evident, at least for the Northern Hemisphere, that much of the difference among stratospheric circulations in different winters is associated with the known oscillations in the atmosphere such as:

- The Southern Oscillation, an irregular oscillation in the tropical-subtropical sea level pressure and in the sea surface temperature of the tropical Pacific Ocean. Its period lies between two and eight years, on average four years (*see* Southern Oscillation, Volume 1).
- The Quasi-Biennial Oscillation, an oscillation in the equatorial stratospheric winds with a period between 19 and 36 months, the mean period being about 28 months (*see* Quasi- Biennial Oscillation (QBO), Volume 1).
- The North Atlantic Oscillation, a seesaw in atmospheric pressure between the subtropical high, centered in the mean near the Azores, and the Icelandic low whose center is in the Greenland–Iceland region. It has a periodicity on the order of 27 to 28 months (Hurrel and van Loon, 1997) (*see* North Atlantic Oscillation, Volume 1).

But there are other forcings from outside the stratosphere such as:

- the 11-year sunspot cycle, (*see* Solar Irradiance and Climate, Volume 1); and
- such volcanic eruptions that are violent enough for their emission to enter the stratosphere.

Even with the knowledge of these influences, it is not yet possible to forecast if a winter is going to be colder or warmer than normal in the north polar regions (i.e., if the polar vortex is going to be strong or weak), because we cannot determine how large a part the individual forcing will play – nor can we foretell the future state of the forcing, or if a volcano will erupt. Quite apart from that, the internal dynamics of the atmosphere can randomly, without any outside forcing, move the stratosphere from one state to another. The models still do not provide satisfactory longer-range forecasts of these complex associations.

But it has become more important than ever to understand the cause(s) of the large variability, because one might, for instance, then be able to explain why, so far, there has been no ozone deficit in the arctic stratosphere comparable to that above Antarctica in spring. Understanding why a winter turns out to be warm or cold in the stratosphere would also help us to understand the interaction between the troposphere and stratosphere (Labitzke and van Loon, 1999).

See also: Stratosphere, Ozone Trends, Volume 1.

REFERENCES

Barnett, J J and Corney, M (1985) Middle Atmosphere Reference Model Derived From Satellite Data, *MAP Handbook*, 16, 47–85.

Dobson, G M B (1930) The Ozone in the Earth s Upper Atmosphere, *Beit. Phys. Atmos.*, 16, 76–85.

Hadley, G (1735) Concerning the Cause of the General Trade Winds, *Philos. Trans. R. Soc. London*, 39, 58, reprinted in *Smithsonian Inst. Misc. Collections*, 51, 5–7 (1910).

Hurrel, J W and van Loon, H (1997) Decadal Variations in Climate Associated with the North Atlantic Oscillation, *Clim. Change*, 36, 301–326.

Kalnay, E M, Kanamitsu, R, Kistler, W, Collins, D, Deaven, L, Gandin, M, Iredell, S, Saha, G, White, J, Woollen, Y, Zhu, A, Leetmaa, R, Reynolds M, Chelliah, W, Ebisuzaki, W, Higgins, J, Janowiak, K, Mo, C, Ropelewski, C, Wang, J, Jenne, R, and Joseph, D (1996) The NCEP/NCAR 40-year Reanalysis Project, *Bull. Am. Meterol. Soc.*, 77, 437–471.

Labitzke, K (1972) Temperature Changes in the Mesosphere and Stratosphere Connected with Circulation Changes in Winter, *J. Atmos. Sci.*, 29, 756–766.

Labitzke, K and van Loon, H (1999) *The Stratosphere: Phenomena, History, and Relevance*, 1–190, Springer, Berlin.

Scherhag, R (1952) Die Explosionsartigen Stratospharenerwarmungen des Spatwinters 1951/52, *Ber. Deutsch. Wetterdienstes US-Zone*, 6(38), 51–63.

von Zahn, U, Fiedler, J, Naujokat, B, Langematz, U, and Kruger, K (1998) A Note on Record-high Temperatures at the Northern Polar Stratopause in Winter 1997/98, *Geophys. Res. Lett.*, 25, 4169–4172.

Stratospheric Ozone, Depletion

see Depletion of Stratospheric Ozone (*Opening essay*, **Volume 1**)

Stratospheric Processes and their Role in Climate (SPARC)

see SPARC (Stratospheric Processes and their Role in Climate) (**Volume 1**)

Sub-grid Processes

In nature, physical quantities such as temperature, pressure, atmospheric humidity, and the concentration of gases or nutrients in the ocean, vary continuously in three dimensions. Nature can thus be thought of as consisting of an infinite number of infinitely closely spaced points. However, in a computer model, only a finite number of points is possible, because the computer memory capacity is finite, and only a finite number of calculations can be performed. In three-dimensional models, the points are arranged in a three-dimensional grid. The spacing between the grid points is called the *spatial resolution*. Processes that occur at spatial scales smaller than the spatial resolution, and which therefore cannot be explicitly represented in models, are referred to as *sub-grid scale processes*. In recent three-dimensional, global-scale atmospheric models, the typical spatial resolution is about 200 km in the horizontal, with the atmosphere divided into 10–20 layers of uneven thickness in the vertical. Examples of sub-grid scale processes in these models include: updrafts and downdrafts associated with convective clouds (100 m spatial scale), convective clouds themselves (1 km spatial scale) and convective cloud systems (10 km spatial scale), microphysical processes within clouds of all types (at a scale of microns), turbulence in and above the surface layer of the atmosphere (meters to 1 km in scale). The exchange of heat, moisture, and momentum between a land surface and the atmosphere involves surface irregularities that occur below the model resolution, so the processes governing these exchanges occur at the sub-grid scale. In ocean models, important sub-grid scale processes include diffusive mixing by a variety of physical processes, and convective overturning. Finally, climate models are marched through time by computing rates of change at a given time and extrapolating these to the next point in time. The length of extrapolation is called the *time step*, and can be thought of as the grid resolution in the time dimension. Many important processes (particularly involving clouds) are sub-grid in the time dimension, as well as in the spatial dimensions.

See also: Downscaling, Volume 1; Parameterization, Volume 1.

L D DANNY HARVEY *Canada*

Sunspots

Judith Lean
Naval Research Laboratory, Washington, DC, USA

Sunspots are small dark features on the Sun's disk. Observed with telescopes since the 17th century, they were recognized in the 1850s to exhibit an 11-year cycle. The number of sunspots present on the solar disk viewed from Earth is the longest, directly observed record of solar variability, and the only such record prior to the 20th century. Space-based observations of the Sun made in the past few decades con rm the signi cance of the sunspot record as a generic indicator of solar activity that relates to a wide variety of solar and terrestrial phenomena. However, in most cases these relations are neither simple nor linear.

Sunspots are compact clusters of magnetic fields in the Sun s atmosphere. Seen as small dark features in the image in Figure 1, sunspots are regions where magnetic fields erupt into the solar atmosphere from below the visible solar surface, after being twisted and stretched by the Sun s rotation on its axis (once approximately every 27 days). Sunspots appear darker than the surrounding solar atmosphere because they are cooler by about

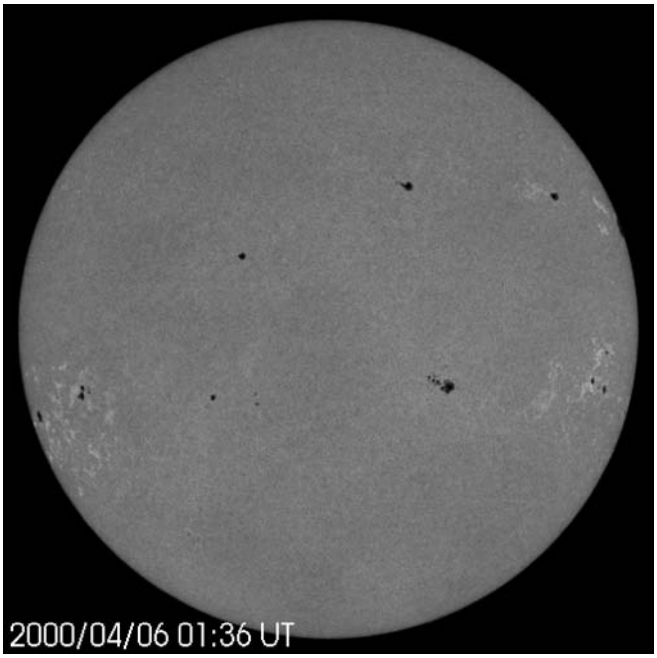


Figure 1 Shown is an image of the Sun recorded in visible radiation by the Michelson Doppler Imager on the Solar and Heliospheric Observatory on the 6th April 2000, during a period of high solar activity in cycle 23, with many sunspots present. Bright faculae are evident surrounding sunspots near the east and west limbs of the disk

1000 K. They are cooler because their magnetic fields form compact structures, as Figure 2 shows, that inhibit the upward flow of energy from the solar interior. Sunspots appear typically at mid heliocentric latitudes during the beginning of an 11-year cycle, and move to lower latitudes as the cycle progresses. Because they rotate with the Sun s atmosphere, sunspots appear to move in a band across the face of the Sun viewed from the Earth. They grow to a typical size of 50 million km^2, evolving and eventually decaying over time scales of weeks. Bright regions called *faculae*, in which magnetic fields are more dispersed and energy flow enhanced, usually accompany sunspots. Faculae are also evident in Figure 1 near the limb of the Sun s disk where their brightness relative to the background radiation is higher than at disk center.

When regular telescopic observations of sunspots commenced in the 17th century, solar activity was at lower levels and sunspots less frequent than today, as the record of sunspot numbers in the top panel of Figure 3 shows. Sunspots occur often, and with larger areas when solar activity is high. Solar activity refers in general terms to the amount of magnetic flux present in the Sun s atmosphere, not just in sunspots but also in other features, such as faculae. The nominal 11-year cycle of magnetic activity is asymmetric, having a rise time of <5 years and a decay time of >6 years. The middle and lower panels of Figure 3 show that the amplitude of this cycle varies by two orders of magnitude while the phase ranges from 8 to 15 years. Solar cycles with lower amplitudes tend to last longer, but the amplitude (Figure 3c) and length (Figure 3d) of solar cycles are only loosely related (correlation coefficient of 0.34). This suggests the behavior of a non-linear oscillator but the true origin of solar activity and the sunspot cycle is still uncertain. Also evident in the sunspot numbers are longer periods, near 90 years, called the *Gleisberg* cycle, and near 200 years.

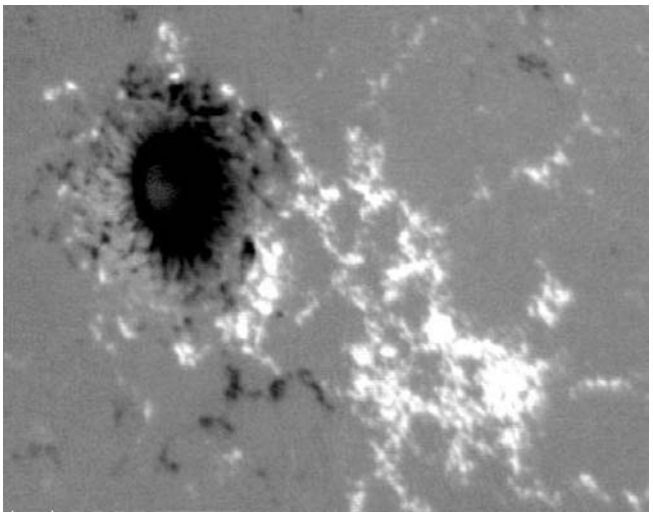


Figure 2 Shown is an image of a sunspot and surrounding magnetic features in NOAA region 7912 on the 25th October 1995 recorded by the Big Bear Solar Observatory. Bright and dark regions signify magnetic fields of different polarity

Even though the Sun s 11-year cycle of magnetic activity is often termed the sunspot cycle, sunspots are only one manifestation of numerous solar processes that vary in response to solar activity. When sunspot numbers are high, such as near maxima of 11-year cycles (1980, 1990 and 2000), bright features (faculae, plages and hot coronal loops) are also larger and more frequent. The overall amount of radiation from the Sun (irradiance) increases, eruptive events such as flares and coronal mass ejections occur more often, and the Sun emits into space a greater flux of energetic particles. Enhanced plasma flow from the solar corona modifies the solar wind and interplanetary magnetic field, which increases the complexity and strength of the magnetic coupling between the Sun and the Earth. This inhibits the penetration of galactic cosmic rays into the Earth s atmosphere, and the levels of cosmogenic isotopes such as ^{14}C and ^{10}Be (in tree rings and ice cores, respectively) are reduced correspondingly (*see* Radionuclides, Cosmogenic, Volume 1). Many solar and terrestrial phenomena have temporal variations that correlate with sunspot numbers and with each other as a result of their common connection, albeit by different physical processes, to solar magnetism.

Because of its duration and apparent homogeneity, the sunspot record is often used in the search for Sun–climate relations as a proxy for the geophysically relevant solar energy outputs, even though the actual physical relationships are complex. Sunspots are direct sources of irradiance variability because their presence in the Sun s atmosphere depletes radiation locally and reduces irradiance at Earth. But the bright faculae that accompany sunspots enhance irradiance so that the observed irradiance variability actually reflects the competing effects of two sources, each of which correlates with the sunspot number. Reconstructions of historical solar irradiance adopt either the cycle amplitude (Figure 3c) or cycle length (Figure 3d) to represent an assumed long-term component that is present in terrestrial geomagnetic and cosmogenic proxies of solar activity in addition to the 11-year cycle. Although times of many sunspots at the Sun s surface signify the presence of hot coronal loops in the outer solar atmosphere and enhanced plasma and particle emissions that also couple with the Earth, the physical associations of sunspots with geomagnetic indices and cosmogenic isotopes is complicated, and remains to be quantitatively specified. The record of sunspots is therefore a generic indicator of solar variability during the past four centuries. An important, unresolved issue for climate change is the question of long-term variability of the Sun s irradiance: Are sunspots

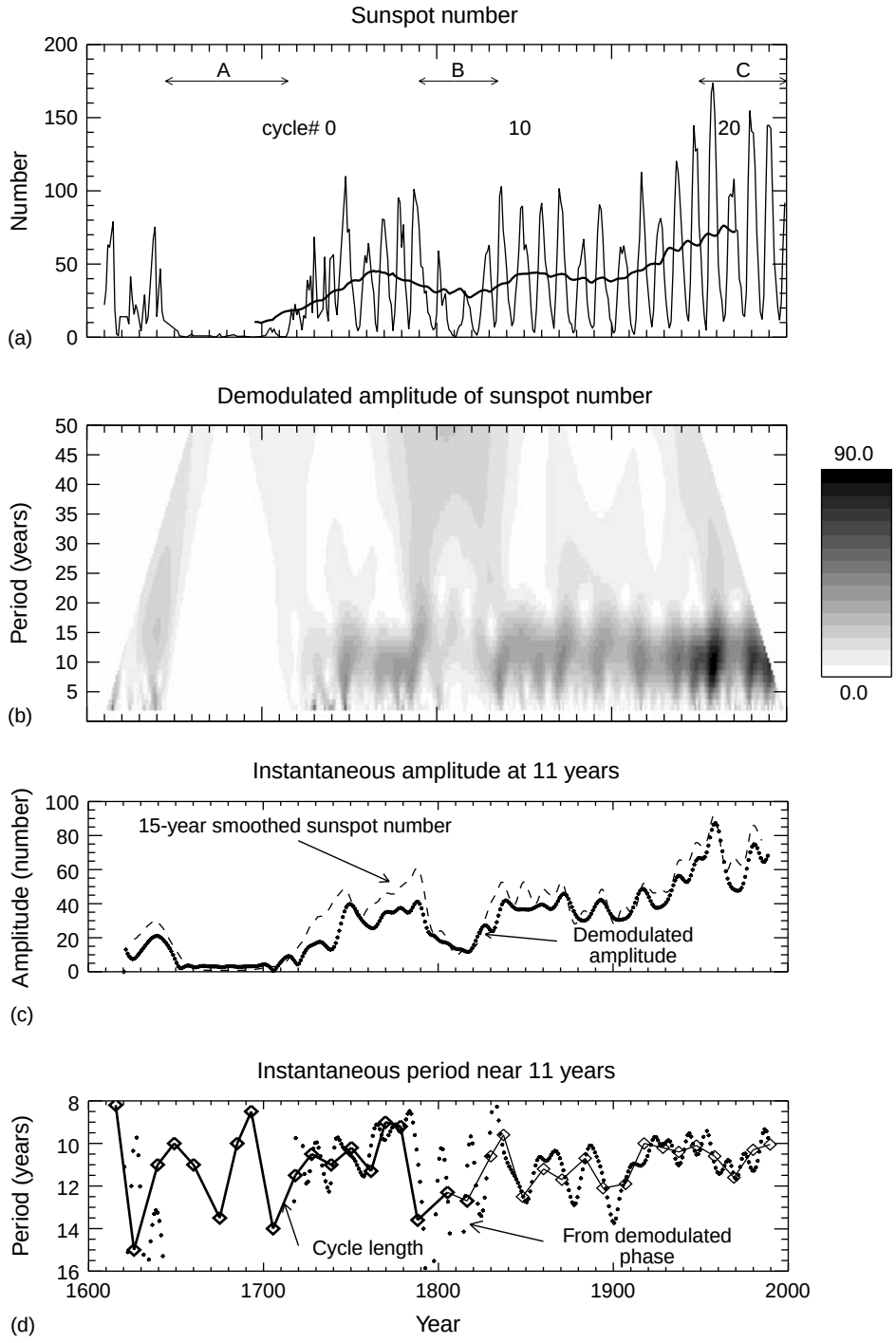


Figure 3 Shown in (a) is the 400 year record of sunspot numbers. The numbering of 11-year cycles commenced in 1750, with cycle 0. The letters A, B and C designate specific epochs called, respectively, the Maunder Minimum (*see* **Maunder Minimum**, Volume 1), Dalton Minimum and Modern Maximum. Space-based observations of the Sun exist only during the Modern Maximum. Shown in (b) are the amplitudes of periods less than 50 years present in the sunspot number record. Most prominent is the nominal 11-year cycle. Variations in its amplitude and length are shown in (c) and (d)

directly indicative of irradiance changes or are longer-term components, such as those evident in terrestrial proxies, also present?

See also: Little Ice Age, Volume 1; Maunder Minimum, Volume 1; Solar Irradiance and Climate, Volume 1; Solar Variability, Long-term, Volume 1.

FURTHER READING

Bray, R J and Loughhead, R E (1964) *Sunspots*, Dover Publications, New York.

Hoyt, D V and Schatten, K H (1994) The One Hundredth Year of Rudolf Wolf s Death: Do We Have the Correct Reconstruction of Solar Activity? *Geophys. Res. Lett.*, 21, 2067–2070.

Suomi, Verner Edward

(1915- 1995)

Verner Edward Suomi was a US scientist and educator. Universally recognized as the father of satellite meteorology, Suomi s legacy to humanity is a dramatically improved understanding of the weather, the atmosphere, and the Earth obtained from sensors in space.

His early instruments flew around the Earth, he designed space probes to study Venus and Jupiter, and he invented the camera that produced the first satellite images of the Earth and its weather now seen daily throughout the world. He was eulogized as a giant of modern science. His inventions were simple and elegant, and their consequences are ubiquitous. Suomi opened the way to a global view from space that reshaped the Earth sciences in the latter decades of the 20th century.

Born in Eveleth, MN, Suomi received a BSc degree in 1938 from the Winona State Teachers College and taught high school until 1941, when he began to study and teach meteorology in a wartime program. He joined the faculty of the University of Wisconsin-Madison in 1948 and received his doctorate from the University of Chicago in 1953. He remained on the University of Wisconsin faculty for his entire career and founded the Space Science and Engineering Center there, directing it until his retirement in 1988.

Suomi s career was shaped by his early attempts to measure the fluxes of radiation involved in the Earth s energy balance. For his PhD thesis, he developed a radiometer to measure radiative fluxes in a cornfield and was led to think about the energy balance of the planet. He created an ingeniously simple flat plate radiometer that was first flown in space in 1959. Combining his observations from space with similar observations from weather balloons, he established the critical role of clouds in the planetary energy balance and stimulated the integration of satellite observations in meteorological practice and research.

His most renowned invention was the 1963 spin-scan camera that typified Suomi s genius for penetrating simplicity. The then new communications satellites, in orbits that remain stationary over a point on the equator, were spin-stabilized. Suomi developed an electronic telescope to scan adjacent sections of the Earth from the rapidly spinning satellite. A picture of the whole disc was obtained by lining up the images of the sections. Later versions of the instrument included color images and then the capability to measure vertical profiles of temperature and water vapor in the atmosphere below.

Recognizing that advancing computer capabilities provided the opportunity to obtain even greater value from his satellite observations, Suomi led efforts to develop interactive data analysis systems that could provide replays of weather events and combine data fields for detailed analysis. These systems proved valuable as the Global Atmospheric Research Programme (GARP), conceived by Suomi and other distinguished scientists, led to worldwide observation initiatives. At the time of his death, he was exploring the uses of the new global positioning system as the basis for new satellite observations of the atmosphere.

Suomi was a tireless advocate of the benefits of science and served on a wide variety of scientific and policy bodies. He was a member of the US National Academy of Engineering and similar academies and honorary bodies of the US and other nations. Awards, honors, and prizes from around the world recognized his creativity and contributions.

Verner Suomi was an unusual individual in many ways. He was patient with other people, but always impatient for progress. His unbounded energy and enthusiasm were contagious. But most of all, he combined a deep physical intuition and understanding with the insight and ingenuity that led to powerful instruments. He described his work as dabbling around in the wonders of creation, and through his bold and sweeping grasp of life and science, he helped the people of the world touch the wonders that he loved.

Photo: by UW–Madison News Service. Reproduced by permission from the University of Wisconsin Board of Regents.

See also: Earth Observing Systems, Volume 1.

JOHN A DUTTON *USA*

Surface Observing Systems

see Earth Observing Systems (*Opening essay*, **Volume 1**); The Global Temperature Record (*Opening essay*, **Volume 1**); Air Quality, Global (**Volume 1**); Monitoring Systems, Global Geophysical (**Volume 1**)

Sverdrup, Harald Ulrik

(1888- 1957)

Harald Ulrik Sverdrup, a Norwegian by birth, was the pre-eminent oceanographer of the first half of the 20th century. Sverdrup s two grandfathers were churchmen. His paternal grandfather had five sons, all of whom became churchmen, and three daughters who married churchmen. But Sverdrup did not follow the family tradition. After a brief stint in astronomy, Sverdrup accepted one of the renowned Carnegie assistantships to Professor V Bjerknes, under whose guidance he wrote the doctoral dissertation *Der Nordatlantische Passat* in Leipzig during World War I (on a starvation diet, as he later reported). Here, Sverdrup s style of mathematics as a tool for quantitative description rather than for deductive reasoning is representative of all his later work. This was also the style of his famous colleague and friend C-G Rossby.

At the age of 29, Sverdrup took charge of the scientific work (with additional duties as navigator and cook) aboard the *Maud* on Amundsen s North Polar expedition. The goal to duplicate Nansen s drift of the *Fram* across the North Pole was never realized. The expedition was planned to last for 3 years, but winter after winter the *Maud* was frozen fast, and she did not return for 7 years. Sverdrup made good use of the time, taking measurements of the physical properties of sea water and sea ice, of marine geologic features, aurora borealis, gravity, magnetism, atmospheric electricity, and even of the animal and bird life on the drift ice. At one time, he left the ship to spend 8 months as the lone white man with the nomadic Chukchi tribe of north-eastern Siberia, an experience he later described in a delightful book, *Reindeer Hair in your Soup* (now published in English). The most significant of the *Maud* papers was *Dynamics of Tides on the North Siberian Shelf*, which was completed on board ship (oceanographers will appreciate this feat).

Upon his return, Sverdrup succeeded V Bjerknes in the Chair of Meteorology in Bergen and became Research Professor at the Michelsen Institute. In 1936 he accepted the Directorship of the Scripps Institution of Oceanography in La Jolla, CA, for a 3-year period. The war intervened, and Sverdrup did not permanently return to Norway for 12 years. During his tenure, Scripps grew from a land-locked marine station to a general oceanographic institute with an active seagoing program. He personally made significant scientific contributions to the war effort.

In the postwar years, Sverdrup devoted much effort to writing (with Johnson and Fleming) *The Oceans, their Physics, Chemistry and Marine Biology*. This is the last comprehensive work covering all aspects of oceanography in a single volume. The book reflects Sverdrup s broad exposure during his 7 years in the Arctic.

Sverdrup s final accomplishment while in America is now broadly known as Sverdrup dynamics. This mathematical representation of the ocean circulation relates northward mass transport $\rho \int v\,\mathrm{d}z$ to $\beta\ \mathrm{curl}\,\tau$, where τ is the wind-stress on the sea surface, and $\beta = \mathrm{d}f/\mathrm{d}y$ is the northward derivative of the Coriolis parameter f. This simple torque relation avoids the equatorial singularity of the traditional force equation, a singularity that had plagued oceanographers for generations. As is typical of his work, the origin of Sverdrup dynamics came from analysis of data and not from deductive analysis.

In 1948, at the age of 65, Sverdrup returned to Norway. He had greatly admired Nansen s decision to totally drop all scientific activity at this time of life. He did likewise. Sverdrup took charge of a Norwegian project for improving fishing in India. He also chaired a committee for re-organizing the Norwegian educational system. In August 1957, Sverdrup died from heart failure in Oslo, 2 days after the death of his good friend Carl-Gustav Rossby in Stockholm, also from heart failure. The death of the two giants drew to a close the era of explorer–scientist from the countries of the large Coriolis parameter.

See also: Ocean Circulation, Volume 1.

WALTER MUNK *USA*

T

Tambora Volcanic Eruption
see Volcanic Eruption, Tambora (**Volume 1**)

Teleconnection
see Pacific–North American (PNA) Teleconnection (**Volume 1**)

Temperature

Temperature is a measure of the magnitude of mean molecular kinetic energy in a physical substance. This energy, called *internal energy*, is defined by the motions of the individual molecules relative to the motion of the center of mass of the physical body. The temperature difference between adjacent physical bodies determines the direction of transfer of this internal energy between the bodies by intermolecular interactions (sensible heat transfer). The transfer is from the body with the higher temperature to that with the lower temperature. No such transfer occurs if the temperatures are the same. Temperature may be measured by a number of physical properties of material substances. Examples include the pressure of a gas in a fixed volume container, the volume of a gas or liquid under constant pressure conditions (such as the standard mercury thermometer), the resistance of an electrical conductor, and the electromotive force (voltage) produced by temperature differences in an electric circuit with two different conductors. The magnitude of temperature is represented by several linear scales, the most common of which are the Kelvin (K), Celsius (C), and Fahrenheit (F) scales. The Kelvin scale is an absolute scale in that magnitudes are proportional to internal energy so that a temperature of 0 K represents a reference condition of no internal energy, i.e., the coldest temperature possible. In these scales and at standard atmospheric sea-level pressure, the freezing point for pure water is 273.15 K, 0 °C and 32 °F, respectively, and the boiling point for pure water is 373.15 K, 100 °C, and 212 °F, respectively.

DAVID HOUGHTON *USA*

Temperature, Atmosphere
see The Global Temperature Record (*Opening essay*, **Volume 1**)

Temperature, Ground
see Ground Temperature (**Volume 1**)

Temperature, Sea Surface
see The Global Temperature Record (*Opening essay*, **Volume 1**)

Temperature, Stratosphere
see Stratosphere, Temperature and Circulation (**Volume 1**)

Temperature, Surface
see The Global Temperature Record (*Opening essay*, **Volume 1**)

Temperature, Troposphere

see Tropospheric Temperature (**Volume 1**)

Thermohaline Circulation

Thermohaline circulation is the part of the ocean circulation that is driven by differences in heating/cooling and evaporation/precipitation/salinity forcing, as opposed to the part of the circulation driven by winds or tides (*see* Ocean Circulation, Volume 1; Salinity Patterns in the Ocean, Volume 1). The upper ocean thermohaline circulation is difficult to detect because of the strength of the wind-driven circulation there. However, thermohaline forcing (i.e., density differences caused by salinity and temperature differences) dominates the deep ocean circulation, below about 2000 m, which is the depth of the large-scale, wind-driven gyres. In contrast to the wind-driven circulation, the thermohaline circulation is of global scale. The abyssal (deep ocean) thermohaline circulation is driven by deep convection caused by large air–sea heat exchange in winter in isolated high-latitude locations. Such convection is very localized, in patches of about 1 km horizontal extent. Convection in these regions reaches 1500–2000 m depth but rarely to the ocean bottom. The abyssal circulation is also driven by rejection of saline brine during the formation of sea ice in narrowly defined regions – the briny waters increase the density of the underlying seawater. Deep water formation sites include the Greenland, Weddell and Ross Seas whose convection fills the abyssal oceans below about 2000 m and the Labrador and Okhotsk Seas, which fill intermediate depths below about 1000 m. Upwelling of these dense waters is considered to occur throughout the world s oceans and is a major cause for the large-scale, but weak, abyssal currents. The convective source regions and ocean basins are connected through deep western boundary currents (DWBCs). DWBC velocities can reach $0.1\,m\,s^{-1}$, while interior ocean velocities due to the thermohaline circulation are usually less than $0.01\,m\,s^{-1}$. In contrast, speeds of the wind-driven currents of the upper 2000 m of the ocean are about 10 times faster, reaching $1\,m\,s^{-1}$ in boundary currents and about $0.1\,m\,s^{-1}$ in mid-ocean.

See also: Ocean Conveyor Belt, Volume 1.

LYNNE D TALLEY *USA*

Tidal Waves

see Tsunamis, Causes and Consequences (**Volume 1**)

Tides, Atmospheric

Atmospheric tides are regular but generally small oscillations of the pressure at the Earth s surface, and also in the wind, caused principally by the Sun but to a smaller extent by the moon. The solar-generated tides follow the Sun through all meridians with nearly constant amplitude in a given latitude and nearly constant time delay. Hence, at a given place on the Earth s surface, the period of oscillation is 24 h or a submultiple thereof. Daily changes of the wind in the stratosphere and above are observed with amplitude of many kilometers per hour, but, near the surface, tides can be easily detected in barometric records only in the tropics, where they have a dominant period of 12 h. Here, typical amplitudes of order 1 millibar (mb) (100 Pascals (Pa)) are observed with maximum pressure at about 10:00 and 22:00 h and minima at about 4:00 and 16:00 h, local solar time. The amplitude falls off with increasing latitude, and in any case it is swamped by random noise due to weather at mid to high latitudes. The atmospheric tide is basically caused by the daily cycle of thermal expansion in the upper atmosphere. The reason for its twice-daily manifestation at the surface remained a mystery until about 1967 when it was shown theoretically to be due to the near-inability of the atmosphere to transmit waves of long period from the stratosphere to the surface. However, a small oscillation of period 24 h at the surface does exist in addition to the predominantly 12-h variation.

The smaller atmospheric tidal oscillation due to the moon has a dominant period of 12.5 h and amplitudes of order 0.1 mb (10 Pa). This lunar tide is driven similarly to the oceanic tide, by variations in the minute tilt of gravity from the vertical caused by the moon s orbit, but it is also affected by direct interaction with the oceanic tide itself.

DAVID CARTWRIGHT *UK*

Tides, Oceanic

David Cartwright
Petersfield, Hampshire, UK

Any regular oscillation of large horizontal scale in the sea that is related to the motion of the Moon and the Sun is known as a tide. Oscillations in the horizontal are strictly called tidal streams or tidal currents, and the closely related vertical oscillations of the surface are called tidal elevations, but the word tide is commonly used to denote both types of motion.

The lunar (and solar) components of the oceanic tides are caused by the out of balance stresses at the Earth s surface between the gravitational pull of the Moon (Sun) and the centrifugal acceleration of the orbit. These stresses tilt the directions of the vertical, towards the Moon (Sun) on the near hemisphere and away from the Moon (Sun) in the opposite hemisphere, as illustrated in Figure 1. Tilting the vertical causes currents, which in turn are modified by the Coriolis effect of the Earth s rotation (*see* Coriolis Effect, Volume 1), by the shape of the continents, and by variations in the ocean s bottom topography. The currents cause tidal elevation to change in order to preserve continuity of mass. As the Earth turns relative to the pattern of stresses, tides result.

The principal spectral component of the tide at most places has a period of half a lunar day, that is 12 h 25 min; this component is known as M_2. Next in importance is the principal solar component with period exactly 12 h; this is known as S_2. These are examples of the semidiurnal tides. The ratio of amplitudes S_2/M_2 varies, but a typical value is 0.4. M_2 and S_2 reinforce each other about every 15 days to produce the *spring tides* of larger vertical range. These higher tides come a day or two after full and new Moons, the delay being due to friction. *Neap tides* are the correspondingly weak tidal ranges occurring when M_2 and S_2 are in anti-phase at the quarter phases of the Moon. Other factors that intermittently produce larger tides are the phases of the Moon and the passage of the equinoxes in March and September.

Diurnal components of the tide, with periods close to a lunar day and a solar day (known as O_1 and K_1), also exist. These components are almost negligible in some seas and oceans, but predominate in others. The diurnal tides are largest when the Moon and Sun have a high declination (i.e., angular distance from the equator), usually near the solstices in December and June. Spectral analysis of tides also reveals many other periodic components related to the motion of the Moon and Sun, generally called harmonic tidal constituents. Predictive tide tables are computed by summing many such harmonic constituents.

Contrary to a popular misconception, tidal elevation is not in general greatest when the Moon has greatest elevation in the sky, but may lag by any fraction of a period, according to oceanic position. The ocean basins resonate with natural periods comparable with the tidal periods. As a result, the amplitude also varies with position from zero at certain nodal positions (known as *amphidromes*) to as much as 5 m in some coastal seas. Accurate global maps showing contours of phase lag and amplitude are now available from high-precision satellite altimetry for all the main harmonic constituents. A global map for the principal lunar tide M_2 is shown in Figure 2.

In shallow seas exposed to the ocean, the tidal tilt of the vertical is less important than the slopes resulting from wave transmission of the local oceanic tides across the shelf boundary. Here the current speed is greatly magnified by transmission from deep to shallow water, and the rate of

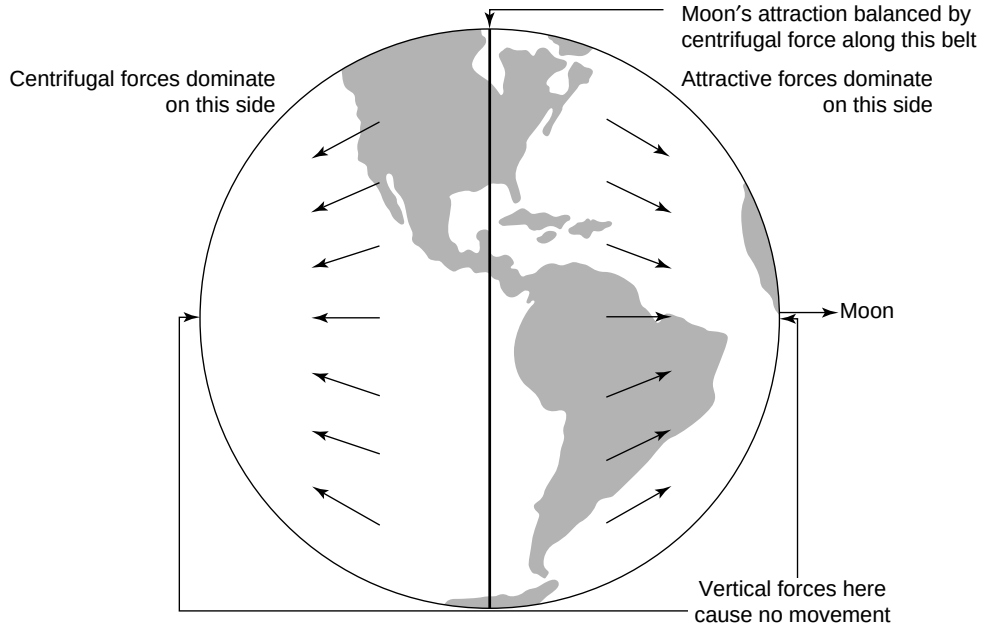


Figure 1 The pattern of the horizontal stresses that generate the tides, with the Moon in the direction shown. The effective tilting of the vertical is greatest, of order 10^{-7}, at 45° and 135° from nadir, and is zero at nadir and at 90°. When the orbit coincides with the equatorial plane, as shown, there are two equal tides per day. In general the Moon may be up to 28.5° above or below the equator, creating an asymmetry that then results in diurnal tides also being generated

energy transmission (being proportional to the square root of the depth) is reduced, with consequent magnification of the vertical amplitude. Small-scale coastal irregularities also produce local alterations. The net result is that maps of shallow coastal seas tend to have larger amplitudes and much closer spacing of contours, typically one-tenth of that of oceanic tidal maps. A little of the detailed structure may be seen in the more geographically constrained areas shown in Figure 2, particularly in the seas surrounding Britain and in the shelf seas east of Argentina and east of Asia.

Another characteristic of tides in shallow seas is that the wave form is distorted from that of a quasi-sinusoid, most commonly rising more rapidly and falling more slowly. An extreme example of this phenomenon is a *tidal bore*, familiar in some river estuaries. These distortions are due to the rise of vertical amplitude in relation to local depth, and to friction on the sea bed being stronger at low than at high tide. The distortion manifests itself spectrally in compound harmonics of the main tidal frequencies, known as *overtides*. In some coastal regions, particularly in parts of the English Channel, the overtides actually produce double high tides or double low tides, or a *stand* in level lasting a few hours.

The largest tidal ranges in the world, 12.9 m during spring tides, occur at the head of the Bay of Fundy, Nova Scotia. The large range is due to the shallow depth and dimensions of this bay together with the outlying continental shelf; in combination, these factors cause the bay to resonate to wave motions with period about 12.5 h. Similar local resonances with somewhat less tidal range are known in other shelf-sea locations.

The net effect of friction in the world s shallow seas and certain internal processes in the deep ocean is to apply a steady torque opposing the rotation of the Earth (*see* Atmospheric Angular Momentum and Earth Rotation, Volume 1). The torque lengthens the day by about 2.3 ms and causes the Moon to recede from the Earth by 3.8 m every century. These forces represent a dissipation of energy at the rate of 3.5 million MW. By contrast, the largest power station built to harness tidal power, which is located on the estuary of the river Rance in Brittany, generates less than 250 MW. A proliferation of such *clean* tidal power stations would therefore hardly harness one thousandth of the power dissipated by natural processes and would have a negligible effect in lengthening the day.

The oceanic tides are a fairly robust physical system. Little change in current strength, vertical amplitude or timing is likely to result from the expected trends in global warming or various large-scale engineering projects (*see* Tidal Power Development, Volume 3). For example, studies of secular changes at tide-gage stations that have existed for over a century in exposed locations have

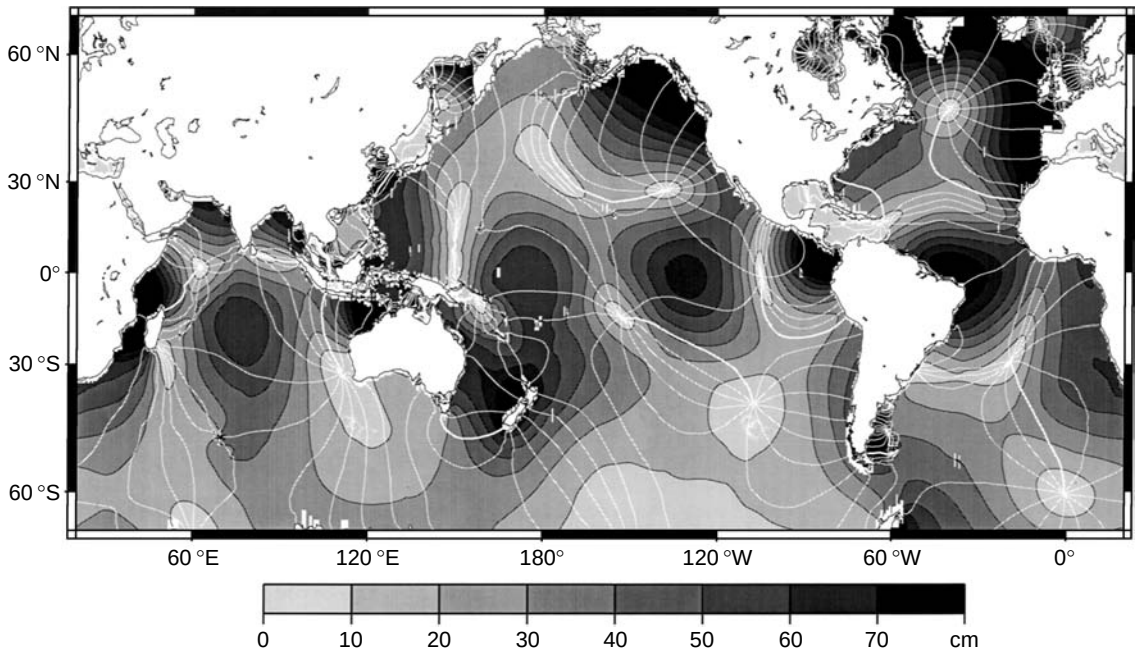


Figure 2 The principal lunar tide M_2 in the deep ocean shows a complex geographic pattern. White lines indicate the phase lag after lunar transit at Greenwich in 12 steps of 30°, each step being equivalent to 1.035 h. Phase lag is zero along the thicker white lines. Amplitude in centimeters according to shading, with black indicating the largest amplitude, pale-gray the least. The amplitude is zero at the nodal points (amphidromes) about which the phase lines radiate. (Figure by courtesy of Richard D. Ray, NASA)

shown changes of only a few centimeters and a few minutes in timing. The largest changes have been observed where a shallow river estuary such as the Thames has been intensively dredged, resulting in the tide arriving at London about 1 h earlier than in medieval times, but with little change in amplitude. A rise in mean sea level by 1 m or so is negligible compared with oceanic depths in its effect on global tidal dynamics. Of far greater consequence is the expected rise in mean sea level itself and its potential for flooding low-lying coastal terrain (*see* Intertidal Zones, Volume 2; Sea Level, Volume 1).

TOGA (Tropical Ocean Global Atmosphere)

The TOGA Program was a 10 year (1985–1995) international program of the World Climate Research Programme with formal objectives:

1. to gain a description of the tropical oceans and the global atmosphere as a time-dependent system, in order to determine the extent to which this system is predictable on time scales of months to years, and to understand the mechanisms and processes underlying that predictability;
2. to study the feasibility of modeling the coupled ocean–atmosphere system for the purpose of predicting its variations on time scales of months to years;
3. to provide scientific background for designing an observing and data transmission system for operational prediction if this capability is demonstrated by the coupled ocean–atmosphere system.

Because of resource limitations, the TOGA Program was conducted primarily in and over the Pacific Ocean and was dedicated to the elucidation of the El Niño/Southern Oscillation phenomenon. The TOGA Program met its third objective by not only designing an observing system, but by deploying a unique system of 70 moorings (the Tropical Atmosphere Ocean, TAO array) deployed along the entire tropical Pacific within 8° of the equator and reporting its observations daily and making the observations available to all (http://www.pmel.noaa.gov/toga-tao/realtime.html).

The first two objectives were aided by this system: predictability was demonstrated by using the data from this system and actually making skillful predictions of sea surface temperature in the tropical Pacific, showing the effects of these predictions on aspects of mid-latitude climate, and showing that the predictions of both tropical and mid-latitude climatic affects had societal value.

Even though TOGA ended in 1995, it has left a number of legacies: the TAO array; the initiation of climate prediction activities on seasonal to interannual times scales in many institutions throughout the world; and the creation of an institute designed to make and apply the predictions to societal benefit, especially in the developing world, the International Research Institute for Climate Prediction (http://iri.ldeo.columbia.edu/). A set of review papers covering the results of the entire international TOGA program was published as a special volume (see Anderson *et al.*, 1998). The US TOGA Program as seen in the context of the International TOGA Program was published in the National Research Council (1996).

FURTHER READING

Anderson, D L T, Sarachik, E S, and Webster, P J, eds (1998) *J. Geophys. Res.*, 103(C7), 14, 167–14, 512.

National Research Council (1996) *Learning to Predict El Niño: Accomplishments and Legacies of the TOGA Program*, National Academy Press, Washington, DC, 1–171 (available online at http://www.nap.edu).

EDWARD SARACHIK *USA*

Tornadoes

Howard B Bluestein
University of Oklahoma, Norman, OK, USA

Tornadoes are rapidly rotating columns of air that are in contact with the ground. They are observed underneath convective storms or underneath convective clouds that are quickly developing into storms but have not yet produced precipitation at the ground. Some tropical cyclones spawn tornadoes when they make landfall. Tornadoes are often, but not always, visible as a condensation funnel underneath cloud base. Some tornadoes are visible only as a rotating column of debris, while others may be hidden behind a curtain of precipitation. A tornado over water is called a waterspout.

Wind speeds at the ground in tornadoes range from around $18\,\mathrm{m\,s^{-1}}$ (40 mph) to as high as $140\,\mathrm{m\,s^{-1}}$ (312 mph) in the most violent case. The intensity of tornadoes is often rated on the Fujita (F) scale, which is based

on the nature of the damage inflicted and ranges from 0 to 5. The correspondence between F-scale rating and wind is as follows: F0 ($18-32\,m\,s^{-1}$), F1 ($33-49\,m\,s^{-1}$), F2 ($50-69\,m\,s^{-1}$), F3 ($70-92\,m\,s^{-1}$), F4 ($93-116\,m\,s^{-1}$), F5 ($117-142\,m\,s^{-1}$). The last can remove well-built houses from their foundations. The highest theoretically possible wind speed depends on both the vertical profile of temperature and moisture (which determine the potential buoyancy) and the wind field; thus, the so-called thermodynamic speed limit can be exceeded, owing to rapid accelerations of air parcels near the ground. The intensity of tornadoes is inversely related to their frequency of occurrence, since it is relatively unlikely that all the conditions necessary for the most intense ones occur.

In the Northern Hemisphere (in the Southern Hemisphere) the winds in most tornadoes rotate in a counterclockwise (clockwise) manner; in some, however, the air rotates in the opposite direction. Ranging in diameter from around 100 m to 1.5 km, some tornadoes are so shallow that they barely extend up to cloud base, while others can extend for 10 km or more, throughout the depth of their parent storm. They can last from a few seconds up to one hour or longer. Most are on the order of hundreds of meters across and last for around tens of minutes. Some convective storms produce a series of tornadoes, which altogether can produce a damage track a 100 km or longer.

Tornadoes move along with their parent storms, but may also rotate around a larger-scale vortex embedded within the storms, or may be displaced laterally by evaporatively cooled gusts of relatively dense air near the ground. Some tornadoes contain smaller-scale vortices, which may be very intense, rotating around a common axis; these multiple satellite vortices are likely responsible for the sometimes-observed spotty nature of tornado damage. Results from laboratory model experiments and computer simulations suggest that whether a tornado has multiple vortices or not depends upon the ratio of the amount of circulation about the tornado vortex to the amount of convergence into the vortex, the *swirl ratio*.

Studies of laboratory models and computer simulations indicate that in single-cell tornadoes, air converges into the vortex near the ground and rises at its center; compensating sinking air is found outside the tornado. In two-cell tornadoes, air that converges into the vortex near the ground rises before it reaches the center; sinking motion is found at the center. Two-celled tornadoes tend to occur when the swirl ratio is relatively high. There is both theoretical and observational evidence that the horizontal wind field in a tornado behaves like a rotating solid body (which has constant vorticity) surrounded by a fluid having no vorticity.

Observed over much of the world, both at low elevations and over mountainous terrain at high elevations, tornadoes are most frequent over a portion of the central US sometimes referred to as Tornado Alley, where a north–south oriented mountain range to the west and a warm body of water to the south, often create favorable conditions for storms that can spawn tornadoes. Although tornadoes can occur at any time of the day or night, or in any season, they are found most frequently in the late afternoon and early evening during the spring.

Most significant tornadoes are spawned in convective storms called *supercells*. Unlike ordinary thunderstorm cells, each of which lasts for less than one hour, supercells last for much longer periods of time, owing to a cyclonically-rotating column of air, the mesocyclone, which is several kilometers wide, and contains air ascending as rapidly as $50\,m\,s^{-1}$ (110 mph) or more. Supercells tend to move to the right (left) of the tracks of ordinary thunderstorms in the Northern (Southern) Hemisphere and often produce large hail.

Supercells can form when there is relatively strong vertical wind shear, i.e., when the wind speed and/or direction at low levels is much different from the wind speed and/or direction at mid- and upper-levels in the troposphere. Since underneath the jet stream there is often strong vertical wind shear, supercells tend to be found during those seasons when the jet stream is near. Since a buoyant updraft is also required, supercells are found when there is relatively warm, moist air near the ground and much cooler air aloft.

Tornadoes in supercells tend to be found near the edge of rain and/or hail, where air is descending, but attached to a lowered cloud base called a wall cloud, in which air is ascending. The tornado is often adjacent to the region where the largest hail reaches the ground. Tornadoes are produced in fewer than half and possibly in fewer than a quarter of all supercells for reasons which are not completely understood. It is also not known what role the mesocyclone has in producing tornadoes. Some tornadoes in supercells form first aloft and build downward; others appear in an air column simultaneously. The detection of rotation in a mesocyclone or developing tornado by a Doppler radar is often used as a basis for the issuance of a tornado warning, often as much as 20–30 minutes in advance. Although there has been much speculation that tornadoes may be related to electrical activity in thunderstorms, it appears that electrical effects are not responsible for their formation (Figure 1).

Other tornadoes can also occur in ordinary convective storms. Unlike supercell tornadoes, which are associated with mesocyclones, tornadoes in ordinary convective storms acquire their rotation from the boundaries along which their parent storms form.

Most tornado damage occurs as a result of wind stress, which varies as the square of the wind speed, while other damage occurs when objects become airborne debris and

Figure 1 Photo of a Tornado taken on 5th of June, 1999, southwest of Bassett, NE. (Copyright Howard B Bluestein)

then make impact. Although the pressure inside a tornado can be as much as 10% less than that of its environment, most damage to buildings is not caused by rapid pressure changes. The safest place to be when a tornado strikes is underground. If it is not possible to get to an underground shelter, one should go to an interior closet or bathroom and protect one s head from flying debris.

It is very difficult to determine whether or not the frequency of occurrence of tornadoes or their geographic locations has changed on decadal time scales, even in the US where they are most frequently observed. Reports of tornadoes are heavily weighted by population density, and have been affected by the recent public awareness of weaker tornadoes, which were not reported as often in earlier years. Estimates of tornado intensity using the F scale are not very reliable when the nature of structures damaged by the tornado is not well known; furthermore, many tornadoes strike unpopulated areas where little if any damage can be found.

Estimating future trends in tornado frequency and intensity requires knowledge of why some convective storms produce tornadoes, while others do not. One could estimate the likelihood that the frequency of occurrence of supercells would change from a predicted change in the vertical wind-shear profile and potential buoyancy. The former depends on the nature of the general wind circulation and the latter depends on the vertical distribution of temperature and moisture; both the former and latter might be sensitive to global climate change.

FURTHER READING

Bluestein, H B (1999) *Tornado Alley: Monster Storms of the Great Plains*, Oxford University Press, New York.

Doswell, III, C A and Burgess, D W (1988) Some Issues of United States Tornado Climatology, *Mon. Weather Rev.*, 116, 495–501.

Fujita, T T (1973) Tornadoes Around the World, *Weatherwise*, 26, 56–62, 78–83.

Fujita, T T (1987) *US Tornadoes: Part 1, 70-year statistics*, University of Chicago, Chicago, IL.

Kelly, D L, Schaefer, J T, McNulty, R P, Doswell, III, C A, and Abbey, Jr, R F (1978) An Augmented Tornado Climatology, *Mon. Weather Rev.*, 106, 1172–1183.

Schaefer, J T, Livingston, R L, Ostby, F P, and Leftwich, P W (1993) The Stability of Climatological Tornado Data, in *The Tornado: its Structure, Dynamics, Prediction, and Hazards*, eds C Church, D Burgess, C Doswell, and R Davies-Jones, AGU Geophysical Monograph 79, 459–466.

Trade Winds

The surface winds over most of the tropics are from the east and directed towards the equator. These are the (easterly) *trade winds*. The trade winds are much steadier in both speed and direction than the surface winds at higher latitudes. The mechanism maintaining the trades is understood as a consequence of the existence of the Hadley Circulation (*see* Hadley Circulation, Volume 1), which operates as a heat engine to transfer heat away from the warmer low latitudes to the cooler high latitudes. Because of momentum conservation, the rotation of the Earth causes the trade winds to blow from east to west.

See also: Atmospheric Motions, Volume 1.

EDWIN K SCHNEIDER *USA*

Transient Response

In climate change research, the term *transient response* is used to indicate the time-dependent character of how a system (typically a numerical model) responds to a perturbation. Because it is sometimes difficult to specify the time-dependent changes in forcings or other boundary conditions, initial studies often are designed to determine the response to a sustained perturbation of some arbitrary size (e.g., doubling the atmospheric carbon dioxide (CO_2) concentration). Such a response indicates the equilibrium response of the system (*see* Equilibrium Response, Volume 1). However, because the forcing factors are likely to be changing at the same time other factors are changing (e.g., the CO_2 concentration is slowly increasing at

the same time solar radiation is varying) and because time lags are varied and complex, conditions are rarely the same long enough for a system to reach new steady state conditions. Thus, the system is always trying to catch up to the changes in forcings that are taking place, and even if the forcing stabilizes at some new value, there will be continuing change as the system tries to equilibrate to the new conditions.

For example, to determine how the climate is likely to respond to changes in atmospheric composition, a rough sense is provided by determining, using a numerical model, the sensitivity of the climate to a steady state doubling of the CO_2 concentration (*see* Climate Sensitivity, Volume 1). However, although some have used such results to try to interpolate how the climate might respond to a gradually increasing CO_2 concentration, this introduces quite a number of problems because feedback processes involving such processes as ice and snow cover and their albedos exhibit nonlinearities and thresholds and because the time for new equilibrium conditions to develop can be very long. For this reason, it has proven essential to attempt to represent the time-dependent nature of the forcings and to simulate the time-dependent (or transient) response.

In simulating the global climate, model simulations are being carried out that involve, for example, specifying the time-varying composition of the atmosphere and then calculating the time-dependent response of the climate system. Early model experiments of this type were based on simplified approximations to changes in atmospheric composition (e.g., the CO_2 concentration increasing at a compound rate of 1% per year). More recently, model experiments have involved simulation of the response to reconstructions of the historic rates of change in atmospheric concentration and scenarios of how the composition may change in the future.

Although such attempts to simulate the observed history of the climate would seem to be just the type of information needed, there are several challenges. A key one is in initializing such simulations. In that the climate has always been changing, and has been changing particularly rapidly over the past century due to human activities, determining how best to start the experiment can have important influences on the results. For example, if the problem is started based on observations of current conditions, a way has to be found to initialize the changing state of all of the various systems (including particularly the ocean and land vegetation). If the problem is started back in time a couple of hundred years, so prior to human influences, there are far fewer observations of key variables with which to verify and initialize the model experiment even though the approximation of near steady state may be more appropriate, allowing the model itself to develop the initial conditions. In addition, starting back in time and running up to the present requires information on the time histories of the various factors that influenced the climate over this period. And further, because the climate is a chaotic system (*see* Chaos and Predictability, Volume 1), an ensemble of simulations needs to be conducted to ensure a statistically satisfactory representation of how the model climate is responding.

Although sometimes difficult to conduct and plagued by limitations in observations to use to drive and test the models, such transient experiments are vital to developing an understanding of how rapidly or slowly the interacting set of climate processes and feedbacks will respond. Although complicated by the natural variability of the system, studies are now underway to compare the results of model simulations of the time-dependent changes in the global climate over the past 100 years with observations of these changes (*see* Climate Change, Detection and Attribution, Volume 1). It is these studies that led the Intergovernmental Panel on Climate Change (IPCC) to conclude that the balance of evidence suggests that there is a discernible human influence on global climate (IPCC, 1996). More recently, studies are focusing on the temporal and spatial patterns of the transient response of the climate to human-induced activities.

REFERENCE

IPCC (1996) *Climate Change 1995: the Science of Climate Change*, eds J T Houghton, L G Meira Filho, B A Callander, N Harris, A Kattenberg, and K Maskell, Cambridge University Press, Cambridge, 1–572.

MICHAEL C MACCRACKEN *USA*

Tree Rings

see Natural Records of Climate Change (**Volume 1**)

TRMM (Tropical Rainfall Measuring Mission)

TRMM is the first space mission dedicated to measuring tropical rainfall through microwave and infrared sensors; a key instrument is the first weather radar in space.

Measuring tropical rainfall and its variability is crucial to understanding and predicting global climate change. By use

of a low-altitude orbit of 350 km (217 miles), TRMM will increase knowledge of how rainfall releases heat energy to drive the atmospheric circulation. TRMM s orbit ranges between 35 °N and 35 °S so that it flies over each surface position at a different local time, permitting determination of how rain varies over the 24-h day.

TRMM is a joint project between the United States and Japan. The National Space and Development Agency of Japan provided the precipitation radar and the H-II rocket launch on November 27th, 1997 for a 3-year mission. TRMM completed 3 years of a very successful mission in November 2000, with enough fuel remaining to continue taking data for 2–4 additional years. NASA s Goddard Space Flight Center provided the observatory, three instruments, the Science Data Processing System, and operation of the TRMM satellite via the Tracking and Data Relay Satellite System. TRMM s primary instruments are a multi-channel passive microwave instrument, the rain radar, and a visible and an infrared radiometer. In addition, TRMM flies a Lightning Imaging Sensor (from Marshall Space Flight Center) and a Clouds and Earth Radiant Energy Sensor (from Langley Research Center).

TRMM runs a continuous validation program using data from four primary sites with calibrated surface radars, rain gauges, and data networks. Four intensive field programs with instrumented aircraft underflights took place over both land and ocean in 1998/1999. Some of the TRMM products involve the use of data from other civilian satellites, such as those in geosynchronous orbit, and from military satellites that fly multi-channel passive microwave instruments in polar orbits.

More information on TRMM and access to its data may be obtained from http://trmm.gsfc.nasa.gov.

See also: Earth Observing Systems, Volume 1; Hydrologic Cycle, Volume 1.

JOANNE SIMPSON *USA*

Tropical Ocean Global Atmosphere (TOGA)

see TOGA (Tropical Ocean Global Atmosphere) (**Volume 1**)

Tropical Rainfall Measuring Mission (TRMM)

see TRMM (Tropical Rainfall Measuring Mission) (**Volume 1**)

Tropical Storms

see Hurricanes, Typhoons and other Tropical Storms – Descriptive Overview (**Volume 1**); Hurricanes, Typhoons and other Tropical Storms – Dynamics and Intensity (**Volume 1**)

Tropopause

The *tropopause* is the boundary separating the troposphere (below) from the stratosphere (above) at a height of around 11 km in midlatitudes. The tropopause is highest in equatorial regions (order 17 km) and lowest over the poles (order 8 km). It is typically defined as the height at which the temperature structure changes from decreasing with height below the tropopause, to being nearly isothermal with height above it. As the bottom boundary of the stable isothermal layer making up the lower regions of the stratosphere, the tropopause represents a barrier to vertical motion that impedes exchange of air between the troposphere and stratosphere. The barrier is not impenetrable, however, and some mixing of stratospheric and tropospheric air occurs when vigorous updrafts penetrate the tropopause or when stratospheric air is drawn downward near upper level fronts.

See also: Stratosphere, Volume 1; Troposphere, Volume 1.

KEITH L SEITTER *USA*

Troposphere

The *troposphere* is the layer of the Earth s atmosphere from the surface to about 11 km in mid-latitudes, in which the temperature normally decreases with height. On clear calm nights, however, the air near the ground cools, causing a rise in temperature in the lowest 100 m or so (called a *radiation inversion*). It is in this layer that most of the phenomena described as weather occur. The troposphere s temperature structure is a result of the vertical mixing that is prevalent in this layer. Convection, both on local and global scales, leads to the troposphere often being vertically well mixed, and the expansional cooling associated with upward motion as well as the compressional warming associated with downward motion leads to an average decrease in temperature with height of about 6.5 °C km^{-1}. The troposphere is capped by

a stable isothermal layer, the bottom of which is referred to as the *tropopause*. This stable layer impedes further vertical motion so that circulations producing clouds and precipitation are confined to the troposphere.

See also: Convection, Volume 1; Tropopause, Volume 1.

KEITH L SEITTER *USA*

Troposphere, Ozone Chemistry

Cynthia S Atherton
Lawrence Livermore National Laboratory, Livermore, CA, USA

Tropospheric ozone refers to ozone (O_3, a tri-atomic form of oxygen) present in the lowest layer of the atmosphere. The troposphere extends from the surface up to 10- 16 km. Because tropospheric ozone can come close to the surface, it damages materials, plants, and is a respiratory irritant. Ozone can also act as a greenhouse gas. Tropospheric ozone arises from two main sources - in situ *chemical production and transport from the stratosphere. The relative importance of these two processes is still debated.*

Few long-term records of ozone concentrations exist. Tropospheric ozone levels have increased substantially since pre-industrial times (~1850). Current models predict that tropospheric ozone may increase signi cantly between now and 2050. These increases affect not only the chemical cycle of the atmosphere, but also the amount of radiation that reaches the surface of the Earth.

Ozone (O_3) is a key atmospheric gas. Ozone in the stratosphere is beneficial because it absorbs incoming ultra-violet (UV) radiation from the Sun. Stratospheric ozone depletion, e.g., the Antarctic ozone hole, therefore, is potentially harmful because it increases human and environmental exposure to UV radiation. In the troposphere, however, a high ozone episode damages materials, vegetation, and is a respiratory irritant. Tropospheric ozone is both chemically produced *in situ* and transported down from the stratosphere, where concentrations are much higher. Within the troposphere, ozone forms when carbon monoxide (CO), methane (CH_4), and non-methane hydrocarbons (NMHCs) react in the presence of nitrogen oxides ($NO_x = NO + NO_2$) and sunlight (Crutzen, 1973). Both natural (biogenic) and human-induced (anthropogenic) processes emit these precursors. Sources include fossil fuel combustion, biomass burning, lightning and emissions from plants and soils. Transport from the stratosphere (Danielsen, 1968) occurs in extratropical latitudes, mainly due to tropopause folding events and the decay of cut-off low-pressure weather systems.

MEASUREMENTS SINCE PRE-INDUSTRIAL TIMES

Ozone concentrations appear to have increased since pre-industrial times (~1850). During the late 1800s, ozone was measured using the Schonbein method on several continents. Unfortunately, the method was susceptible to interferences, and ozone was measured for long times at only a few locations. Data, however, have been reanalyzed for Paris, Moncalieri (Italy), South America, Pic Du Midi (southwestern France), Zagreb (Yugoslavia), and other locations (see, e.g., Volz and Kley, 1988; Anfossi *et al.*, 1991; Sandroni *et al.*, 1992; Marenco *et al.*, 1994). While the results are somewhat qualitative, they suggest that in the late 19th century, annual average surface ozone concentrations may have been as low as roughly 10 parts per billion volume (ppbv). Current measurements in similar locations, as shown in Figure 1, indicate surface ozone concentrations may have increased by a factor of 2–5, corresponding to a rate of 1.6% year^{-1} for the past 100 years (Marenco *et al.*, 1994; Staehelin *et al.*, 1994).

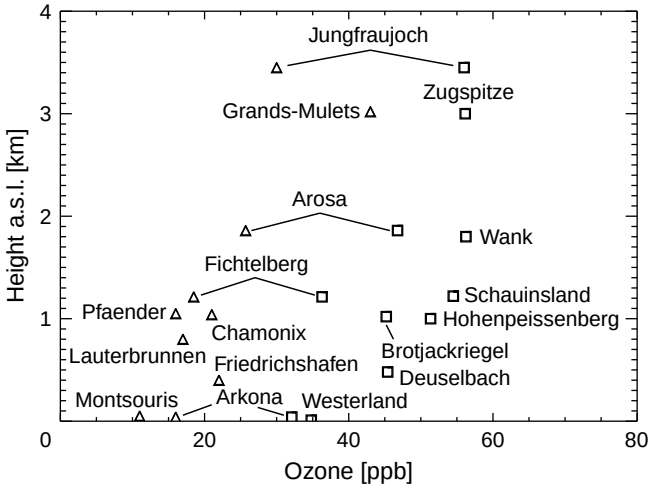


Figure 1 Measurements of surface ozone concentrations from different locations in Europe performed before the end of the 1950s (△) and in recent years (1990–1991; □) during August and September, as a function of altitude. (Reprinted from *Atmospheric Environment*, Staehelin *et al.*, 1994, with permission from Elsevier Science)

CURRENT CLIMATOLOGIES AND RECENT TRENDS (20TH CENTURY)

Climatologies of tropospheric ozone have been published using surface, ozonesonde, and satellite measurements (Fortuin and Kelder, 1998; Logan, 1999). Concentrations vary with season, location, local meteorology, and other factors (WMO, 1998). Multi-decadal records of tropospheric ozone exist for only a handful of sites that are not influenced by local sources. At a small number of locations, vertical distributions are measured by ozonesondes, which are lightweight electrochemical cells carried on balloons to roughly 30 km. These measurements are most uncertain in the troposphere. Their geographic distribution is limited, and instrument calibrations, sampling frequency, changes in methods, etc., affect results.

Tropospheric column ozone (the amount of ozone in a column from the Earth s surface to the tropopause) has recently been measured directly by satellites (Munro *et al.*, 1998). It is also derived as the difference (residual) between satellite measurements of total column and stratospheric column ozone. This residual indicates high ozone levels in the Northern Hemisphere in the summer and plumes in the South Atlantic Ocean that arise from biomass burning (Fishman *et al.*, 1991).

Tropospheric ozone trend studies of the past 30–35 years use data from a small number of ozonesondes and surface sites. In specific regions, tropospheric ozone may increase if precursor emissions increase, or decrease if stratospheric ozone decreases. At high latitudes (e.g., over northern Canada, Alaska, the South Pole), near-surface and free tropospheric ozone appear to be decreasing, perhaps due to decreasing stratospheric ozone concentrations (Oltmans *et al.*, 1998; Tarasick *et al.*, 1995). The largest decrease is over the South Pole, with a trend of $-0{\bullet}70 \pm 0{\bullet}20\%\,\text{year}^{-1}$. Ozonesondes at five Canadians sites showed negative trends for 1981–1993 of -1.0 to -2.2%, depending on height in the troposphere.

In the northern mid-latitudes, two eastern United States (US) sites show only slight increases in tropospheric ozone since 1970 (Oltmans *et al.*, 1998). Over much of the US, surface ozone did not show an appreciable trend during 1980–1995, and showed decreasing trends primarily near New York City, Los Angeles, and Chicago (Logan, 1994; Fiore *et al.*, 1998).

In Arosa, a high elevation station in Switzerland, it appears that surface ozone concentrations have increased by a factor of 2.2 since the 1950s (Staehelin *et al.*, 1994). Although European ozone concentrations increased rapidly through the mid-1980s, there appears to have been no change or little decrease since then (Logan, 1994; Oltmans *et al.*, 1998). Increases in ozone over the mid-latitudes of Japan are largest in the lower troposphere; however, they have slowed in the last decade (Oltmans *et al.*, 1998).

Less data are available for the tropics and subtropics. The tropics generally have not experienced significant changes in tropospheric ozone (Oltmans *et al.*, 1998; WMO, 1998). Sites located at 34°S and 41°S may show very slight increases in ozone (Oltmans *et al.*, 1998).

MODELING RESULTS – PAST ATMOSPHERES

Global, three-dimensional atmospheric models have simulated both pre-industrial and present day atmospheres. They show ozone increases of 10–40 ppbv (a factor of 2–4 higher) since pre-industrial times. The largest increases tend to be in the Northern Hemisphere during summer. Although the largest percentage ozone increase occurs near the Earth s surface, the largest change in total ozone mass occurs in the upper troposphere. The tropospheric ozone burden (the total mass of ozone in the troposphere) appears to have increased by 40–70% since pre-industrial times.

Increases in tropospheric ozone lead to increases in the absorption of both longwave and shortwave radiation in the troposphere. The change in the net radiative flux at the tropopause (the top of the troposphere) is called *radiative forcing* (*see* Radiative Forcing, Volume 1). Changes in ozone near the tropopause are much more efficient at increasing radiative forcing than changes near the surface, by as much as a factor of 10 (Lacis *et al.*, 1990). Increases in tropospheric ozone since pre-industrial times may have increased global annual radiative forcing by $0.2–0.5\,\text{W}\,\text{m}^{-2}$. Regional and seasonal variations may range from 0.1 to $1.0\,\text{W}\,\text{m}^{-2}$. For comparison, the increase in carbon dioxide (CO_2) since pre-industrial times is predicted to have increased radiative forcing by $1.6\,\text{W}\,\text{m}^{-2}$ (IPCC, 1995).

MODELING PREDICTIONS – OZONE IN FUTURE SCENARIOS

Ozone is expected to increase as emissions of biogenic and anthropogenic O_3 precursors increase. For example, as temperature increases, emissions of isoprene (a hydrocarbon emitted by plants) soil NO_x, and lightning NO_x are expected to increase. Increased lightning NO_x may increase ozone, but soil NO_x and isoprene have not yet been studied globally.

Several models have predicted changes in ozone and radiative forcing due to increases in anthropogenic emissions between 1990 and 2050. Brasseur *et al.* (1998) project the largest relative change in zonal average O_3 concentrations will be ~30% (4 ppbv) near the surface in the tropics. The largest absolute change in zonal average O_3 concentrations is projected to be 35 ppbv (~25%) near 300 mb in the northern mid- and high-latitudes. Ozone in the boundary layer is predicted to increase by 10–20 ppbv (~50%)

over Asia, Central America, South Africa and Brazil. This would lead to a global average increase in radiative forcing due to O_3 of 0.41 W m^{-2} (Brasseur *et al.*, 1998), compared to projections in similar studies of 0.28 W m^{-2} (van Dorland *et al.*, 1997) and 0.21 W m^{-2} (Stevenson *et al.*, 1998). The models predict a higher increase in radiative forcing in the Northern Hemisphere compared to the Southern Hemisphere.

See also: Carbon Monoxide, Volume 1; Hydrogen Peroxide Trends in Greenland Glaciers, Volume 1; Methane, Volume 1; Stratosphere, Chemistry, Volume 1.

REFERENCES

Anfossi, D, Sandroni, S, and Viarengo, S (1991) Tropospheric Ozone in the Nineteenth Century: The Moncalieri Series, *J. Geophys. Res.*, 96, 17 349–17 352.

Brasseur, G P, Kiehl, J T, Muller, J F, Schneider, T, Granier, C, Tie, X X, and Hauglustaine, D (1998) Past and Future Changes in Global Tropospheric Ozone: Impact on Radiative Forcing, *Geophys. Res. Lett.*, 25, 3807–3810.

Crutzen, P (1973) A Discussion of the Chemistry of some Minor Constituents in the Stratosphere and Troposphere, *Pure Appl. Geophys.*, 106- 108, 1385–1399.

Danielsen, E F (1968) Stratospheric–Tropospheric Exchange based on Radioactivity, Ozone and Potential Vorticity, *J. Atmos. Sci.*, 25, 502–518.

Fiore, A M, Jacob, D J, Logan, J A, and Yin, J H (1998) Long-term Trends in Ground Level Ozone over the Contiguous United States, 1980–1995, *J. Geophys. Res.*, 103, 1471–1480.

Fishman J, Fakhruzzaman, K, Cros, B, and Nganga, D (1991) Identification of Widespread Pollution in the Southern Hemisphere Deduced from Satellite Analyses, *Science*, 252, 1693–1696.

Fortuin, J P F and Kelder, H (1998) An Ozone Climatology based on Ozonesonde and Satellite Measurements, *J. Geophys. Res.*, 103(31), 709–31 734.

IPCC (1995) *Climate Change*, eds J T Houghton, L G Meiro Filho, B A Callendar, N Harris, A Kattenburg, and K Maskell, Cambridge University Press, New York.

Lacis, A A, Wuebbles, D J, and Logan, J A (1990) Radiative Forcing of Climate by Changes in the Vertical Distribution of Ozone, *J. Geophys. Res.*, 95, 9971–9981.

Logan, J A (1999) An Analysis of Ozonesonde Data for the Troposphere: Recommendations for Testing 3D Models and Development of a Gridded Climatology for Tropospheric Ozone, *J. Geophys. Res.*, 104(16), 115–16, 149.

Logan, J A (1994) Trends in the Vertical Distribution of Ozone: an Analysis of Ozonesonde Data, *J. Geophys. Res.*, 99, 25 553–25 585.

Marenco, A, Gouget, H, Nédélec, P, Pagés, J P, and Karcher, F (1994) Evidence of a Long-term Increase in Tropospheric Ozone from Pic du Midi Data Series: Consequences: Positive Radiative Forcing, *J. Geophys. Res.*, 99, 16 617–16 632.

Munro, R, Siddans, R, Reburn, W J, and Kerridge, B J (1998) Direct Measurement of Tropospheric Ozone Distributions from Space, *Nature*, 392, 168–171.

Oltmans, S J, Lefohn, A S, Scheel, H E, Harris, J M, Levy, II, H, Galbally, I E, Brunke, E G, Meyer, C P, Lathrop, J A, Johnson, B J, Shadwick, D S, Cuevas, E, Schmidlin, F J, Tarasick, D W, Claude, H, Kerr, J B, Uchino, O, and Mohnen, V (1998) Trends of Ozone in the Troposphere, *Geophys. Res. Lett.*, 25, 139–142.

Sandroni, S, Anfossi, D, and Viarengo, S (1992) Surface Ozone Levels at the End of the Nineteenth Century in South America, *J. Geophys. Res.*, 97, 2535–2539.

Staehelin, J, Thudium, J, Buehler, R, Volz-Thomas, A, and Graber, W (1994) Trends in Surface Ozone Concentrations at Arosa (Switzerland), *Atmos. Environ.*, 28, 75–87.

Stevenson, D S, Johnson, C E, Collins, W J, Derwent, R G, Shine, K P, and Edwards, J M (1998) Evolution of Tropospheric Ozone Radiative Forcing, *Geophys. Res. Lett.*, 25, 3819–3822.

Tarasick, D W, Wardle, D I, Kerr, J B, Bellefleur, J J, and Davies, J (1995) Tropospheric Ozone Trends over Canada: 1980–1993, *Geophys. Res. Lett.*, 22, 409–412.

van Dorland, R, Dentener, F J, and Lelieveld, J (1997) Radiative Forcing due to Tropospheric Ozone and Sulfate Aerosols, *J. Geophys. Res.*, 102, 28 079–28 100.

Volz, A and Kley, D (1988) Evaluation of the Montsouris Series of Ozone Measurements made in the Nineteenth Century, *Nature*, 332, 240–242.

WMO (1998) Global Ozone Research and Monitoring Project – Report No. 44, in *Scienti c Assessment of Ozone Depletion: 1998, Tropospheric Ozone and Related Processes*.

Tropospheric Temperature

John R Christy

University of Alabama in Huntsville, Huntsville, AL, USA

Tropospheric temperature measurements, which are extensive enough to begin to address issues of global variability and change, have been reported since 1958 from balloons, and 1979 from satellites. Temperature variability is greater in the deep troposphere than observed at the surface, as can been seen in responses to transient climate events such as El Niño/Southern Oscillation (ENSO) events and volcanoes. Analyses of temperature trends in the troposphere from both balloons and satellites show a warming of approximately 0.10 °C decade^{-1} since 1958 and about 0.04 °C decade^{-1} since 1979. The rate since 1979 is about one-third the rate of temperature increase observed at the surface. In particular, the tropical latitudes show trend-differences of about 0.15 °C decade^{-1}, with the surface warming faster. Whether the period since 1979 is truly representative of the longer-term climate responses of the surface and atmosphere to in uences such as increasing greenhouse gas and aerosol

concentrations is made more dif cult to assess because of the relatively short length of time examined here.

INTRODUCTION

Research concerning climate variation and change is generally focused on the surface climate because this is where we as humans live and produce what is necessary for our survival. However, surface climate is dependent upon the behavior of the atmosphere as a whole. Above the Earth s surface is a deep layer of fairly well-mixed air called the troposphere (from Greek $\tau\rho o\pi\eta$, or turning) in which the familiar processes of weather occur. This layer stretches from the surface to altitudes of 16–17 km in the tropics, but only to about 8 km near the poles. In terms of mass, the troposphere represents over 80% of the global atmosphere. Above the troposphere lies the thin and stable stratosphere.

With the advent of aviation, and particularly the needs of aviators in World War II, knowledge of the behavior of the troposphere became vital. To measure such quantities as temperature and winds aloft, balloons were released from the surface carrying instrument packages which would transmit measurements by radio to a ground station as the balloon ascended (*see* Radiosondes, Volume 1). This radio-based system provided information necessary to reconstruct an atmospheric sounding (i.e., vertical profile), and so the balloon/instrument package was and is known as a radiosonde. Radiosonde releases were concentrated along airways in the 1940s and 1950s and so could not give a picture of, for example, the global variation of tropospheric temperature to answer scientific questions. The emphasis on global measurements during the International Geophysical Year of 1957–1958 brought several new stations into the mix so that attempts could begin to assess global temperature variations and trends, though vast oceanic regions were (and are) still unobserved.

In the 1970s, satellites were first used to create vertical profiles of atmospheric temperature as well. However, the satellite technique of temperature reconstruction, in most cases, is not a straightforward method. Unlike radiosondes, which report temperature at many discrete levels, satellite radiometers receive emissions from deep layers of the atmosphere, often several kilometers thick. Most weather forecasting and aviation problems require data at discrete levels, so estimating the true vertical profile of temperature at many levels using the broad satellite-measured layer values was and is a difficult challenge. However, it is relatively easy to produce from radiosonde data what would be sensed by the satellites. It is this type of comparison that will be examined below.

Most of the activity we call weather, such as clouds and storms, occurs within the troposphere, so that surface conditions are significantly determined by what happens in this deep layer. The connection between the surface and the troposphere, though, varies considerably around the globe. Over mid-latitude continents, for example, solar heating of the surface generates rising convective currents of air that produce substantial mixing throughout the troposphere. In these regions, the surface and tropospheric temperature tend to vary in unison, though the effect dominates during the daytime.

Over the polar ice caps, there are low-level temperature inversions due to radiation effects. In the tropics and subtropics, it is the anticyclonic subsidence inversion that is at play, which is associated with the trade-wind inversion. In such regions, the surface temperature is relatively cool and there is little mixing of surface air with the deep layer above so that a clear connection between the surface and tropospheric temperatures is not evident. Thus, the temperatures of the surface and troposphere represent the temperatures of different physical layers.

The combustion of carbon-based fuels for energy and transportation increased rapidly in the 20th century, resulting in increased atmospheric concentrations of carbon dioxide (CO_2), a greenhouse gas. In general, results from global climate models indicate that the surface and troposphere should warm as a consequence of increased CO_2 (and other greenhouse gases such as methane.) However, other emissions due to human activity, such as sulfur compounds or carbon soot, create aerosols which are thought to inhibit or enhance the flux of radiative energy to the surface, and so cause relative regional cooling or warming. When the estimated effects of all of these emissions are incorporated into climate models, the troposphere, particularly in the tropics, is predicted to warm at a more rapid rate than the average surface temperature. At the same time, other types of natural variations, such as EI Niño events and volcanic eruptions, and of human-induced changes, such as ozone depletion and land cover change, can also cause different temperature responses in the troposphere and at the surface. Thus, precise knowledge of the global surface and tropospheric temperature is needed to check against such modeled simulations and help provide clarity on the potential evolution of greenhouse-induced global climate changes.

The following sections describe the key methods used to determine tropospheric temperatures, as well as the major uncertainties associated with each. As will be shown, the remarkable correspondence between the temperatures measured by balloons and satellites independently give confidence that the long term variations they measure are substantially correct. One consistent result, that since 1979 the troposphere has warmed little while the surface has warmed considerably, is a curious finding which presents a challenge for climate researchers to explain at the dawn of the 21st century. Beginning with balloon data in 1958, the overall 40+ year tropospheric trend is very close to that of

the surface, indicating the period of the last two decades of the 20th century may be unusual.

DATA

There are three sources of tropospheric temperature data examined here. Radiosonde data are thought of as *in situ* data because the instrument carried aloft comes into direct contact with the air being measured. Satellites are remote measurements as they sense the intensity of radiation emitted by the atmosphere to space. And finally, reanalysis may be described as an intelligent combination of observations, both *in situ* and remote, with the added influence of mathematical relationships that approximate the physical laws of nature.

Radiosondes

Radiosondes (i.e., weather balloons) provide the longest record of upper air data and thus have been studied for clues as to the long-term variations of atmospheric temperature. However, changes in radiosonde instrumentation and data processing over the years have resulted in discontinuities in these temperature records (Gaffen, 1994; Parker *et al.*, 1997; Gaffen *et al.*, 2000a). Efforts to generate adjustments to account for these effects are a continuing area of research. One method uses models of the thermodynamics of radiosonde instruments to quantify the effect of varying radiation corrections applied in the processing of the raw data (Luers and Eskridge, 1998), but has yet to be applied on a daily basis to a global data set. A second method relies on statistical tests to identify change-points – sudden temperature shifts not likely to be natural (e.g., instrumental changes.) The global trend of a radiosonde dataset depends on how these inconsistencies are defined, which stations are utilized and the method of areal averaging (Santer *et al.*, 1999; Gaffen *et al.*, 2000a).

Since 1979, the temperatures from the microwave sounding unit (MSU) data have been available as a potential reference. In a third method, Parker *et al.* (1997) used the stratospheric MSU temperatures to adjust monthly radiosonde reports for stations in Australia and New Zealand at times when changes were reported in instrumentation or operational procedures. Some individual stratospheric corrections were as much as 3 K. The main disadvantage of Parker *et al.* s technique is that the raw MSU record has time-varying biases that must first be estimated and eliminated (Christy *et al.*, 2000).

For the analysis of trends, long station data records with minimal missing data are most desirable. The records for the 180 stations in the combined Global Climate Observing System Baseline Global Upper-Air Network, and the data sets that have been used in the earliest Intergovernmental Panel on Climate Change (IPCC, 1990, 1992) compiled by Angell (1988) do not generally meet this standard. In the CLIMAT dataset, a subset of about 400 radiosonde stations that report monthly statistics for tropospheric levels (850–100 hPa), only 74 stations reported enough data to produce monthly means for 85% of the overall period. Extending into the lower stratosphere (up to 30 hPa), only 22 stations meet this requirement. These deficiencies require using either relatively small networks of stations with adequate data or larger networks with lower quality, which may add uncertainty to resulting trend estimates.

The construction methodology of the various global balloon data sets is substantively different. Sterin (1997) from the Russian Research Institute for Hydrometeorological Information (RIHMI) use data from over 800 stations reported over the global telecommunication system and compiled into a special dataset known as the comprehensive aerological reference data set. The data were objectively interpolated to unobserved regions to produce global coverage. Parker *et al.* (1997) from the Hadley Center, United Kingdom Meteorological Office (HadRT) introduced a series of gridded datasets based on CLIMAT reports (HadRT2.0, ~400 radiosonde sites) containing only grids with *in situ*, unadjusted observations. Building upon HadRT2.0, derivative HadRT datasets progressively include more complete maps and adjusted values by infilling with statistical techniques and/or applying adjustments to the original data. Because the tropospheric temperatures from these datasets are relatively low in altitude, data availability is generally high and discontinuities are small (Parker *et al.*, 1997). Angell (1988) reports average volume-based temperature measurements of 63 stations with long data records.

Satellites

Unlike the discrete levels at which radiosondes measure temperatures, satellite instruments used to date observe the intensity of radiation from deep atmospheric layers. The MSU instrument detects emissions from atmospheric oxygen near the 60 GHz absorption line, the intensity of which is proportional to temperature. These satellites orbit the planet 14 times per day and measure emissions for over 30 000 systematically spaced locations every 24 h. About 98% of the Earth is observed over any 4-day period. Two temperature products made widely available are the low-mid troposphere (T2LT, surface to about 8 km) and the lower stratosphere (15–23 km, Christy *et al.*, 1998, 2000).

The MSU data have undergone three revisions since they first appeared in 1990, due to discoveries as the data were examined through time (Spencer and Christy, 1990). Mo (1995) reported that for one of the longest lived satellites (National Oceanic and Atmospheric Administration [NOAA] – 12, 1991–1998) the laboratory-tested,

non-linear calibration coefficients required corrections. In addition, Wentz and Schabel (1998) discovered that the loss of altitude during solar activity (when increased drag of the atmosphere causes orbit decay) induces gradual, spurious cooling in T2LT. Christy *et al.* (2000) found that instrument responses often differ between the laboratory assessments and on-orbit performance, and require additional corrections. Also, further adjustments to the MSU temperatures were made by explicitly recalculating temperature biases due to diurnal effects caused by east–west drift of the orbital path of the spacecraft.

Reanalyses

The methodology of reanalysis is to use historical data, along with a global computer model s data-ingestion scheme, to create a set of atmospheric data maps of weather parameters at many levels (Kalnay *et al.*, 1996). Within the ingestion scheme, potentially errant data are amended or excluded on the basis of comparisons with neighbors and/or expected conditions. However, small, time-dependent biases, of magnitudes potentially important for determining long-term warming or cooling, are virtually impossible to detect in a typical depiction of weather fields. Furthermore, in areas with few *in situ* data, the reanalyses may be affected by inadequate model physics or biased satellite data (different from products described above). Values of tropospheric temperature from the National Centers for Environmental Predication (NCEP) reanalysis are included for comparison purposes here, though it should be noted that discontinuities might exist in these data (Santer *et al.*, 1999).

RESULTS

Figure 1 shows tropospheric temperature variations since 1958 using representative datasets from those described above. Actual trend values ($^{\circ}$C decade^{-1}) since 1958 and 1979 are shown in Table 1. Consistent in the time series is the relative shift to warmer temperatures in the troposphere between 1975 and 1978 followed by large variations due to ENSO events, particularly in 1998, and volcanic events (Agung, 1963, El Chichón, 1982 and Mt Pinatubo, 1991). Christy and McNider (1994) and Michaels and Knappenberger (2000) show that over 60% of the variance in the global time series is due to these two recurring natural fluctuations with the atmospheric response lagging the event (i.e., ENSO or volcano) by a few months. In general, a sea surface temperature fluctuation during ENSO is communicated to the atmosphere through sensible and latent heat wherein a response factor of about 1.3 is measured (i.e., a rise in tropical sea water temperatures of 0.10 $^{\circ}$C translates to a rise in tropospheric temperatures of 0.13 $^{\circ}$C). Thus, the variance of the tropospheric temperature time series is greater than that of surface temperatures.

For the longer period (1958–2000), the tropospheric datasets in Table 1 indicate positive trends near $+0.10\,^{\circ}$C decade^{-1}. Since 1979, however, the tropospheric trends are less, around $+0.04\,^{\circ}$C decade^{-1}. This latter period represents an unusual result because the surface temperature trend during this period is calculated to be $+0.13$ to $+0.18\,^{\circ}$C decade^{-1} from various estimates, and is highlighted as evidence of human induced climate change (e.g., Santer *et al.*, 2000). (While all of the datasets described above are being used to detect small changes from year to

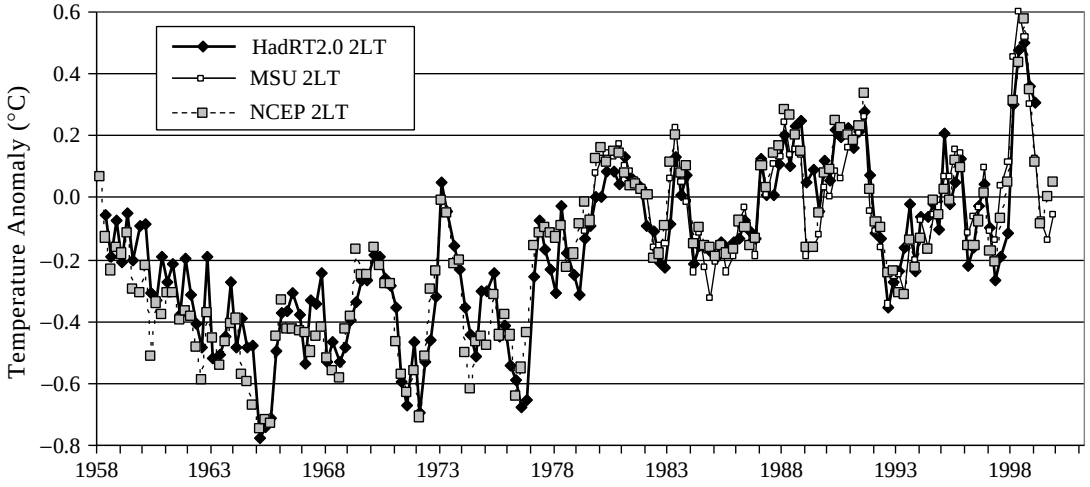


Figure 1 Seasonal global tropospheric temperature anomalies 1958–1999. HadRT2.0 is a compilation of radiosonde DATA, British Crown Copyright 2000 and was provided by Dr S J Brown, Hadley Centre, Meteorological Office, UK (Parker *et al.*, 1997). MSU 2LT represents satellite tropospheric temperatures as described in Christy *et al.* (2000). NCEP is the National Centers for Environmental Prediction (US) reanalysis (Kalnay *et al.*, 1996) converted to satellite temperatures by Dr M Stendel, Danish Meteorological Service

Table 1 Trends of annual anomalies of the following surface and tropospheric datasets computed by restricted maximum likelihood[a]

	Global		Tropics	
	1958–2000	1979–2000	1958–2000	1979–2000
Surface	+0.10	+0.16	+0.08	+0.10
Troposphere				
MSU	–	+0.04	–	−0.06
HadRT2.0	+0.11	+0.03	+0.13	−0.08
NCEP	+0.13	+0.01	+0.08	−0.07
RIHMI	+0.04	+0.00	+0.07	−0.06
Angell	+0.08	−0.02	+0.08	−0.05

[a] The surface data are comprised of land–air temperatures from the Climate Research Unit, University of East Anglia and the ocean near-surface temperatures from the Hadley Center, UK Met Office (Jones *et al.*, 1999; Folland *et al.*, 1995). MSU represents the MSU lower tropospheric values (Christy *et al.*, 2000), HadRT2.0: unadjusted radiosonde-simulated MSU temperatures (Parker *et al.*, 1997), NCEP, National (US) Centers for Environmental Prediction reanalysis-simulated MSU temperatures (Stendel *et al.*, 2000), RIHMI, 850–300 hPa radiosonde temperatures (Sterin, 1997) and Angell: 850–300 hPa thickness temperature (Angell, 1988).

year and decade to decade, it must be noted that none of the systems was specifically designed for this purpose.)

The major data uncertainties have been noted for the balloon, satellite and reanalysis temperatures. Because the balloon and satellite systems independently measure the temperature of the troposphere and have such remarkable agreement, it is highly likely that the actual slow variation (trend) of the troposphere since 1979 is fairly well characterized (Table 1 and NRC, 2000). The result implies that the bulk of the atmosphere s mass has shown little warming since 1979, and that the difference is significant (Christy *et al.*, 2001).

To further investigate this unusual phenomenon in which the surface temperatures apparently have warmed more rapidly than the tropospheric temperatures, Gaffen *et al.* (2000b) examined trends in tropical freezing elevation and lapse rate using radiosondes. They found that the mean elevation of freezing temperatures increased about 150 m between 1960 and 1980, indicating tropospheric warming. Since 1980, however, there has been a slight decrease (though not significant) in freezing height, consistent with a negative tropospheric temperature trend for this more recent period.

The lapse rate is simply the change in temperature versus altitude, which in the tropics (where the main difference in surface and tropospheric trends is found) is a slope of approximately $-6.5\,^{\circ}\mathrm{C\,km^{-1}}$. If the surface and satellite data shown here are both generally correct, then the lapse rate should reveal a tendency toward increasingly cooler values during the more recent period, as the upper air temperatures warmed less than lower level temperatures. One distinct advantage of utilizing this quantity is the elimination of radiosonde biases that come about when stations change methods or instrumentation because, here, the quantity measured is the relative difference in temperature at two levels, not the absolute temperature itself. Their results, also confirmed by Brown *et al.* (2000), indicate that since 1979 there has been a decrease of the tropical lapse rate (i.e., becoming even more negative). This provides further evidence that surface and tropospheric temperatures may respond with slightly different magnitudes of change to the sum of forces acting upon them.

Though there is less confidence in the balloon data prior to the mid-1970s, both Gaffen *et al.* and Brown *et al.* note that tropical lapse rates became more positive during the period 1958–1979, suggesting that in the two most recent decades, a reversal of the trend found in the earlier period has occurred. Upon examination of the entire time series since 1958, one notes that the surface and tropospheric trends are not significantly different.

Error estimates for the trends in temperature since 1979 are generally between ± 0.05 and $\pm 0.10\,^{\circ}\mathrm{C\,decade^{-1}}$ (Christy *et al.*, 2000). This is an estimate of measurement error of the systems themselves. In other words, this error-estimate answers the question – to what precision do we know the trend of the tropospheric temperature since 1958 or 1979?

There is another equally important question regarding the estimation of trends. To what extent is the specific period since 1979 representative of global temperature variations in general? This question leads to an examination of the statistical properties of the time series (Santer *et al.*, 1999). It is possible, for example, that since 1979, the Earth has experienced an especially anomalous period characterized by two large volcanic eruptions (El Chichón and Mt Pinatubo) as well as several substantial ENSO events. These types of events have considerable impact on the global temperature, which in turn has considerable impact on the calculation of the trend, given the relatively short climate period being studied (Christy *et al.*, 2000). In addition, the

surface and troposphere tend to respond in slightly different ways to these short-term climate fluctuations (i.e., the factor of 1.3, for example). As such, the period since 1979 might not be representative of a longer period. It is, therefore, also possible that this particular period is relatively unrepresentative of the manner in which the surface and troposphere might respond to slow increases in greenhouse gases.

See also: Climate Change, Detection and Attribution, Volume 1; Earth Observing Systems, Volume 1; The Global Temperature Record, Volume 1.

REFERENCES

Angell, J K (1988) Variations and Trends in Tropospheric and Stratospheric Global Temperatures, 1958–1987, *J. Clim.*, 1, 1296–1313.

Brown, S J, Parker, D E, Folland, C K, and Macadam, I (2000) Decadal Variability in the Lower-tropospheric Lapse Rate, *Geophys. Res. Lett.*, 27, 997–1000.

Christy, J R and McNider, R T (1994) Satellite Greenhouse Signal, *Nature*, 367, 325.

Christy, J R, Spencer, R W, and Lobl, E (1998) Analysis of the Merging Procedure for the MSU Daily Temperature Time Series, *J. Clim.*, 5, 2016–2041.

Christy, J R, Spencer, R W, and Braswell, W D (2000) MSU Tropospheric Temperatures, 2000: Dataset Construction and Radiosonde Comparisons, *J. Atmos. Oc. Tech.*, 17, 1153–1170.

Christy, J R, Parker, D E, Brown, S J, Macadam, I, Stendel, M, and Norris, W B (2001) Differential Trends in Tropical Sea Surface and Atmospheric Temperatures Since 1979, *Geophys. Res. Lett.*, 28, 183–186.

Folland, C K and Parker, D E (1995) Correction of Instrumental Biases in Historical Sea Surface Temperatures, *Quart. J. R. Meteorl. Soc.*, 121, 319–367.

Gaffen, D J (1994) Temporal Inhomogeneities in Radiosonde Temperature Records, *J. Geophys. Res.*, 99, 3667–3676.

Gaffen, D J, Sargent, M A, Habermann, R E, and Lazante, J R (2000a) Sensitivity of Tropospheric and Stratospheric Temperature Trends to Radiosonde Data Quality, *J. Clim.*, 13, 1776–1796.

Gaffen, D J, Santer, B D, Boyle, J S, Christy, J R, Graham, N E, and Ross, R J (2000b) Multi-decadal Changes in the Vertical Temperature Structure of the Tropical Troposphere, *Science*, 287, 1242–1245.

IPCC (1990) *Climate Change, The IPCC Scienti c Assessment*, eds J T Houghton, G J Jenkins, and J J Ephraums, Cambridge University Press, Cambridge, 1–365.

IPCC (1992) *Climate Change 1992: The Supplementary Report to the IPCC Scienti c Assessment*, eds J T Houghton, B A Callander, and S K Varney, Cambridge University Press, Cambridge, 1–198.

IPCC (1996) *Climate Change 1995: The Science of Climate Change*, eds J T Houghton, F G Meira Filho, B A Callander, N Harris, A Kattenberg, and K Maskell, Cambridge University Press, Cambridge, 1–570.

Jones, P D, Osborne, T J, and Briffa, K R (1997) Estimating Sampling Errors in Large-Scale Temperature Averages, *J. Clim.*, 10, 2548–2568.

Jones, P D, New, M, Parker, D E, Martin, S, and Rigor, I G (1999) Surface Air Temperature and its Changes Over the Past 150 Years, *Rev. Geophys.*, 37, 173–199.

Kalnay, E, Kanamitsu, M, Kistler, R, Collins, W, Deaven, D, Gandin, L, Iredell, M, Saha, S, White, G, Woollen, J, Zhu, Y, Leetma, A, and Reynolds, R (1996) The NCEP/NCAR 40-year Reanalysis Project, *Bull. Am. Meterol. Soc.*, 77, 437–471.

Luers, J K and Eskridge, R E (1998) Use of Radiosonde Data in Climate Studies, *J. Clim.*, 11, 1002–1019.

Michaels, P J and Knappenberger, P C (2000) Natural Signals in the MSU Lower Tropospheric Temperature, *Geophys. Res. Lett.*, 27, 2905–2908.

Mo, T (1995) A Study of the Microwave Sounding Unit on the NOAA-12 Satellite. IEEE Trans, *Geosci. Remote Sens.*, 33, 1141–1152.

NRC (National Research Council) (2000) *Reconciling Observations of Global Temperature Change*, National Academy Press, Washington, DC, 1–85.

Parker, D E, Gordon, M, Cullum, D P N, Sexton, D M H, Folland, C K, and Rayner, N (1997) A New Global Gridded Radiosonde Temperature Data Base and Recent Temperature Trends, *Geophys. Res. Lett.*, 24, 1499–1502.

Santer, B D, Hnilo, J J, Wigley, T M L, Boyle, J S, Doutriaux, C, Fiorino, M, Parker, D E, and Taylor, K E (1999) Uncertainties in Observational Estimates of Temperature Change in the Free Atmosphere, *J. Geophys. Res.*, 104, 6305–6333.

Santer, B D, Wigley, T M L, Gaffen, D J, Bengtsson, L, Doutriaux, C, Boyle, J S, Esch, M, Hnilo, J J, Jones, P D, Meehl, G A, Roeckner, E, Taylor, K E, and Wehner, M F (2000) Interpreting Differential Temperature Trends at the Surface and in the Lower Troposphere, *Science*, 287, 2000.

Spencer, R W and Christy, J R (1990) Precise Monitoring of Global Temperature Trends from Satellites, *Science*, 247, 1558–1562.

Stendel, M, Christy, J R, and Bengtsson, L (2000) Assessing Levels of Uncertainty in Recent Temperature Time Series, *Clim. Dyn.*, 16, 587–601.

Sterin, A M (1999) An Analysis of Linear Trends in the Free Atmosphere Temperature Series for 1958–1997, *Meteorol. Gidrol.*, 5, 52–68.

Wentz, F J and Schabel, M (1998) Effects of Satellite Orbital Decay on MSU Lower Tropospheric Temperature Trends, *Nature*, 394, 661–664.

Tsunamis, Causes and Consequences

Charles L Mader

Mader Consulting Co, Honolulu, HI, USA

Tsunamis occur at irregular intervals on the order of years. A tsunami is an event that results in the sea level rising

by tens of meters with widespread destruction to shoreline habitations and facilities. Few natural phenomena have popular conceptions that are more incorrect than those associated with tsunamis, including their common misnomer - tidal waves. Only recently have scienti c studies revealed that tsunami waves are primarily manifestations of the Earth's crustal readjustments caused by internal radioactive heating. Historically, an average of 10 large and 100 minor tsunamis per century occur in the Paci c. Thirteen large events have occurred during the last 250 years in the Atlantic and about 50 in the Indian Ocean and eastern Mediterranean Sea.

Tsunamis involve major, but temporary, changes in sea level that are caused most often by natural geological processes. For example, tsunamis have resulted from volcanic eruptions such as that of Krakatoa in 1883. That explosion was equivalent to 500 megatons of trinitrotoluene (TNT). It produced waves that drowned an estimated 36 000 people in the East Indies.

Earthquakes can also cause tsunamis. The tsunami waves caused by earthquakes have killed more people than the land movement itself. Over 55 000 lives have been claimed by tsunamis since 1900.

The ocean oors are composed of slowly moving crustal plates that plunge under the continental plates, creating deep ocean trenches and mountain ranges around the periphery of the ocean basins. The elastic strain at the plate edges is discontinuously relieved in the form of earthquakes. Vertical strains occur where the plates meet head on (Peru, Chile and the Aleutian Islands Arc) resulting in uplift or depression of large regions of the sea oor that results in a corresponding sea displacement, which causes the tsunami wave.

For example, the earthquake of March 28, 1964 in Alaska involved a dislocation of 2 m over 250 000 km^2 (three times the size of Florida). The resulting mound of water fanned out over the Paci c, creating its greatest damage at Crescent City, CA.

In the deep ocean a tsunami wave has a height of about 1 m and a wavelength of 100 to 500 km. Its period varies from a few minutes to 30 minutes. Tsunamis move at the speed of a jet airplane and, upon arriving at a seashore, the wavelength and speed decrease while the height increases greatly.

The severity of local wave effects at any place remote from the source of a large tsunami depends in a complicated way upon the source orientation, the ocean oor topography, and the local topography. One region of a shore may have a large amount of destruction while none occurs a few kilometers away; the reverse could occur for a tsunami from a different source. This directional dependence makes the issuing of accurate tsunami warnings and the evaluation of the tsunami hazard for any particular shoreline location very dif cult.

Advances in numerical modeling of ocean wave generation, propagation and interaction with real shoreline topography permit evaluation of the tsunami hazard resulting from historical tsunamis and from possible future tsunamis. Numerical modeling of tsunamis is being used by the West Coast- Alaskan Tsunami Warning Center to provide emergency managers along the Paci c coasts of Alaska, British Columbia, Washington, Oregon and California a reasonable range of expected wave heights during a tsunami warning for tsunamis generated in the northwest Paci c Ocean. This major improvement in tsunami warning is described in Whitmore and Sokolowski (1996).

EARTHQUAKE GENERATED TSUNAMIS

The Hilo, Hawaii Example

Hilo, Hawaii has received some of the most extensive damage from tsunami waves of any city in the US.

The tsunami of April 1, 1946 was caused by an earthquake of 7.5 magnitude off the Aleutian Islands. The source was located about 100 km SW of Scotch Cap, Unimak Island where the tsunami destroyed a lighthouse and radio towers located more than 30 m above sea level. The tsunami arrived at Hilo about five hours later with a large withdrawal of the ocean followed by several waves that flooded large regions of downtown Hilo. Losses included 173 lives with much property damage.

The tsunami of March 28, 1964 was caused by an earthquake of 8.4 magnitude in Alaska near Prince William Sound. Upon arrival at Hilo and other Hawaiian sites, the resulting tsunami was considerably smaller than the tsunami of April 1, 1946, even though the earthquake source was much larger.

The SWAN code for modeling shallow water waves has been developed over the last 30 years to numerically model tsunami waves (Mader, 1988). Flooding is described using the Mader flooding technique (Mader, 1990). The numerical model accurately simulated the actual flooding levels for the April 1, 1946 and March 28, 1964 tsunamis. The 7.5 magnitude 1946 tsunami flooding of Hilo was much greater than the 8.5 magnitude 1964 tsunami because of the directionality of the tsunami from its source.

The largest tsunami wave ever observed at Hilo, Hawaii occurred during the tsunami of May 23, 1960. The tsunami was caused by an earthquake of 8.5 magnitude occurring near Peru, Chile. The source of the tsunami was a sea floor displacement over an area 150 km wide and 800 km long, aligned along the N–S Chile coastline. The tsunami wave arrived 14 hours later at Hilo, Hawaii with the third wave being the largest and most destructive tsunami wave in Hilo s history. The waves arrived from the southeast at the Hawaiian islands. The calculated tsunami wave profile was

a 0.25 m, 3000 second half wave followed by a 0.5 m high, 2500 second wave and then by a 0.5 m 1500 third wave.

The tsunami wave interacted with the Hawaiian Islands and refracted such that the wave arrived from the east on the Hilo side of the Island of Hawaii. The calculated tsunami wave arriving outside Hilo Bay had an initial amplitude of 0.5 m and a period of 3000 second for 1500 s, followed by a 1.5 m, 1500 second wave and then by a 2.0 m 1500 second wave.

The calculated flooding levels agreed with those that occurred on May 23, 1960. The calculated third wave was the largest and steepest, in agreement with the observations. The technical details of the Hilo tsunami modeling are described in Mader and Curtis (1991).

The amount of information generated by such a series of calculations is very large. Following the complicated wave generation, propagation, interaction and flooding with single snapshot graphs of wave contours or picture plots is difficult. The wave patterns and their histories can be better understood by the generation of computer movies. A collection of over 50 tsunami computer movies is available from the author or The Tsunami Society, PO Box 37970, Honolulu, HI 96817 for tsunamis generated by earthquakes, landslides, and asteroids, showing their propagation and inundation of coastlines around the world.

Landslide Tsunamis

On July 7, 1998, a magnitude 7.1 earthquake occurred off the northern coast of Papua, New Guinea. The earthquake was observed at the Pacific Tsunami Warning Center, but because the earthquake was below the 7.2 threshold for significant earthquake-generated tsunamis, no warning was issued. Yet within fifteen minutes of the main shock, waves up to 15 m high caught the coastal inhabitants unawares. The tsunami killed more than 2200 villagers, including more than 230 children.

Apparently while the earthquake was too small to generate a significant tsunami wave, it did trigger a submarine landslide offshore that caused a large amplitude, short period tsunami wave over a relatively short (30 km) stretch of the coastline.

Tsunamis have also been generated by underwater landslides without an earthquake trigger. Because the landslides often occur when there is an extremely low tide, it appears that the pore pressure inside the slide mass can serve as a landslide trigger. So the very dangerous situation exists that tsunamis may be generated without any warning by an earthquake. The landslides often occur in regions with large river delta deposits. Such was the case in the Papua, New Guinea tsunami caused by an earthquake-triggered landslide and in the 1994 Skagway tsunami triggered only by a very low tide. Landslide generated tsunamis without an earthquake trigger such as the 1994 Skagway tsunami can occur along any coastal region without any warning to the residents of the coast.

The Skagway Tsunami

On November 3, 1994, at about 19:10 h, a tsunami wave with a period of about 3 minutes and maximum height of 8–9 m occurred in the Taiya Inlet at Skagway, Alaska. The event occurred at the time of a low tide of 1.2 m below lower low-level. The tidal range at Skagway is about 8 m. The wave was observed traveling along the Pacific and Arctic Railway and Navigation Company (PARN) dock from the South or deep end of Taiya Inlet.

The sea floor elevations before and after the event indicate that a considerable area of the sea floor was lower for about 1500 m down the inlet and then much of the sea floor was higher farther down the inlet as described by Campbell and Nottingham (1999). The Skagway River had been at its highest level in history the previous month and had deposited large amounts of sediment at the head of the landslide. The sea floor elevation changes had a volume of 15–18 million m^3.

The National Oceanic and Atmospheric Administration (NOAA) tide gauge was situated midway along the ore dock on the west side of the harbor. The tsunami recorded by this gauge on November 3, 1994, had a period of approximately three minutes, a maximum recorded amplitude of about 1 m and lasted for about 30 minutes. The Skagway tide gauge was a damped nitrogen bubbler analog gauge that gives a non-linear response at short periods. Thus, the recorded wave heights are considerably smaller than those of the actual tsunami. To determine the tsunami wave heights, Nottingham (1997) determined the water tower-generated tide gauge trace that would duplicate the tsunami tide gauge record. The tsunami wave had an initial amplitude of 3.6 m and period of 60 s followed by 3 waves with amplitudes of 1.8–2.7 m and period of 180 s. The Skagway landslide generated tsunami is the best documented and most extensively studied landslide tsunami.

The generation and propagation of the tsunami wave of November 3, 1994, in the Taiya Inlet was modeled using the SWAN code by Mader (1997) and by Kowalik (1997) using his shallow water code. The landslide model studied was that described by Campbell (1999) from the change in the sea floor topography that occurred before and after the tsunami. The Campbell model consists of three slide regions separated by ridges.

The material from the three slides moved down into the deep part of the inlet occupying an area with a volume of 16 million m^3.

The tide gauge duration and profile was best reproduced by the two-minute duration slide (about 28 m s^{-1}). The observed and calculated 30-minute duration of the tide gauge record is a result of the resonance for three-minute

waves in the harbor where the tide gauge is located. The waves decayed outside the harbor after 10 minutes. The Skagway landslide-generated tsunami is now a classic example of a tsunami generated without an earthquake trigger. Such tsunamis present the greatest hazard to the largest number of coastal dwellers because they strike with no warning.

Mega-tsunamis

Mega-tsunamis are defined as waves of over 100 m height in the deep ocean. The primary sources of mega-tsunamis are asteroid impacts with the ocean, volcanic explosions, and landslides. A mega-tsunami did occur in 1958 in Lituya Bay, Alaska.

Lituya Bay – a 520 m High Tsunami

Lituya Bay, Alaska is on the northeast shore of the Gulf of Alaska. It is an ice-scoured tidal inlet with a maximum depth of 220 m and a narrow entrance with a depth of only 10 m. It is a T-shaped bay, 11 km long and up to 3 km wide. The two arms at the head of the bay, Gilbert and Crillon Inlets, are part of a trench along the Fairweather Fault. On July 8th, 1958, a 7.5 magnitude earthquake occurred along the Fairweather fault with an epicenter near Lituya Bay.

A mega-tsunami wave was generated that washed out trees to a maximum altitude of 520 m at the entrance of Gilbert Inlet. A photograph of the region is shown in Figure 1. Much of the rest of the shoreline of the Bay was denuded by the tsunami from 30–200 m altitude.

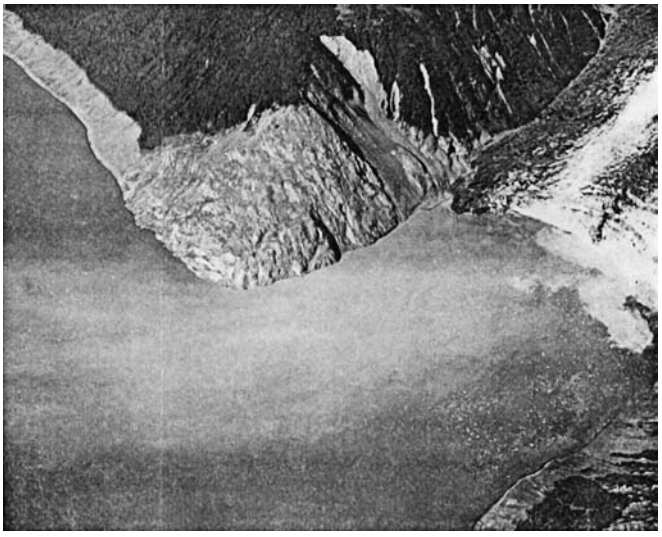


Figure 1 Photograph of where a mega-tsunami reached 520 m on the spur ridge going into Gilbert Inlet of Lituya Bay, Alaska on July 9, 1958. The camera is looking northwest at the head of the bay

During the last 150 years, five giant waves have occurred in Lituya. The previous event occurred on October 27, 1936, which washed out trees to a maximum altitude of 150 m and was not associated with an earthquake.

Miller (1960) recorded all that was known in 1960 about the giant waves in Lituya Bay.

The July 9th, 1958, earthquake occurred at about 22:15 h, which was still daylight at Lituya Bay. The weather was clear and the tide was ebbing at about plus 1.5 m. Bill and Vivian Swanson were on their boat anchored in Anchorage Cove near the western side of the entrance of Lituya Bay. Their astounding observations are recorded in Miller (1960) and were as follows:

> With the first jolt, I tumbled out of the bunk and looked toward the head of the bay where all the noise was coming from. The mountains were shaking something awful, with slide of rock and snow, but what I noticed mostly was the glacier, the north glacier, the one they call Lituya Glacier.
>
> I know you can t ordinarily see that glacier from where I was anchored. People shake their heads when I tell them I saw it that night. I can t help it if they don t believe me. I know the glacier is hidden by the point when you re in Anchorage Cove, but I know what I saw that night, too.
>
> The glacier had risen in the air and moved forward so it was in sight. It must have risen several hundred feet. I don t mean it was just hanging in the air. It seems to be solid, but it was jumping and shaking like crazy. Big chunks of ice were falling off the face of it and down into the water. That was six miles away and they still looked like big chunks. They came off the glacier like a big load of rocks spilling out of a dump truck. That went on for a little while – it s hard to tell just how long – and then suddenly the glacier dropped back out of sight and there was a big wall of water going over the point. The wave started for us right after that and I was too busy to tell what else was happening up there.

A 15 m high wave rushed out of the head of the bay toward Swanson s anchored boat. The boat shot upward on the crest of the wave and over the tops of standing spruce trees on the entrance spit of Lituya Bay. Swanson looked down on the trees growing on the spit and said he was more than 25 m above their tops. The wave crest broke just outside the spit and the boat hit bottom and foundered some distance from the shore. Swanson saw water pouring over the spit, carrying logs and other debris. The Swansons escaped in their skiff to be picked up by another fishing boat two hours later.

The front of Lituya Glacier, on July 10th, was a nearly straight, vertical wall almost normal to the trend of the valley. Comparisons with photographs of the glacier, taken July 7th, indicate that 400 m of ice had been sheared off the glacier front.

The partially subglacial lake west of the sharp bend in Lituya glacier was 30 m lower than it was two days before the earthquake.

After the earthquake there was a fresh scar on the northeast wall of Gilbert Inlet, marking the recent position of a large mass of rock that had plunged down the steep

slope into the water. The next day after the earthquake and tsunami, loose rock debris on the fresh scar was still moving at some places, and small masses of rock still were falling from the rock cliffs near the head of the scar. The volume of the rockfall was approximately 30 million m^3. The Swanson observations and numerical modeling studies by Mader (1999) indicate that a lot more than a simple landslide occurred. The Swanson observations suggest a water wave lifted the front of the glacier up and moved it out from its initial position and generated the 520 m high wave run-up.

George Pararas-Caryannis (1999) suggested that the wave was formed by a landslide impact similar to an asteroid impact, making a cavity in the inlet ocean to the depth of the inlet floor (120 m) near the landslide. The water in the inlet with the width of the landslide and between the landslide and the 520 m high run-up is sufficient to cover the run-up region to 100 m height. This high water layer is sufficient to form a wave that will reproduce the observed flooding of the bay beyond the inlet. Fritz *et al.* (2001) have shown with a physical model how the Lituya Bay landslide impact resulted in the observed movement of the water and glacier.

The Tlingit natives believe that a powerful spirit, Kah Lituya, lives in deep ocean caverns near the entrance to the bay. He resents being disturbed and rises up to shake the mountains and wash away intruders with giant waves.

Asteroid Tsunamis

Tsunami waves generated by earthquakes have typically had a maximum deep ocean amplitude of 10 m and periods of 500–2000 s. Tsunami waves with amplitudes of approximately 100 m are referred to as mega-tsunamis and may be generated by landslides, volcanic explosions, or by an asteroid impact with the ocean (*see* Asteroids and Comets, Effects on Earth, Volume 1). Landslide and small asteroid (less than 1 km in diameter) mega-tsunamis have short periods and thus the tsunami wave amplitude is not maintained as the wave travels from the source, as described in Crawford and Mader (1998). Large asteroid-generated tsunami waves have long periods and waves with large amplitudes that can propagate across an ocean basin.

The July 1994 Shoemaker-Levy-9 asteroid impact with Jupiter is typical of the type of event being investigated. To generate a 100 m high, 2000 second period tsunami that can traverse an ocean basin, an approximately 150 km wide cavity in the ocean to the ocean floor is required. Possible Jupiter asteroid parameters which could generate such a cavity are diameters of 5 km, densities of 5 g cm^{-1}, and velocities of 20 km s^{-1}. The proposed Jupiter asteroid characteristics vary at least by a factor of two in diameter and density, with velocities between 10 and 30 km s^{-1}. To create a mega-tsunami, the location of the asteroid impact would need to be in the deep ocean (at least 3000 m deep). A large, high-density asteroid could be located almost anywhere in a deep ocean basin; while a smaller, lower density asteroid would need to be located nearer the region of interest.

A tsunami wave with a 100 m height and period of 2000 s generated west of the Hawaiian island chain and its interaction with the islands has been modeled by Mader (1996). The asteroid tsunami wave would result in inundation of much of the Hawaiian Islands to between 150 and 300 m. Simulations have been done for all the ocean basins and detailed inundation studies have been performed for all the US, European, and Japanese coastal areas using the technique described by Mader (1998). These results indicate that all coastal regions below 300 m in elevation are at risk from tsunamis generated by Shoemaker-Levy-9 type asteroids.

Asteroid impacts are one of the few natural disasters that can be averted by technical means, as described by Solem (1999). The most cost-effective and only currently available means of deflection is a nuclear explosive. Asteroids more than 3 km in diameter can be most efficiently deflected with a surface burst; asteroids as small as half a kilometer can be effectively deflected with a stand-off device; and smaller asteroids are best pulverized.

The risk of being killed as a result of an asteroid impact is twice the risk of being killed in an airplane crash, as discussed by Solem (1999). Steele (1995) states that the comet impacts on Jupiter similar to the one that occurred in July 1994 were observed in 1883, 1928, 1939 and 1948, suggesting an occurrence rate of about five collision events per century. Jupiter is likely to be struck about 1000 times as often as the Earth which implies that the Earth would be hit, on the average, every 20 000 years by an asteroid that could generate a tsunami that was a hazard through the entire Atlantic or Pacific basin. However when large asteroids next impact the planet, tsunamis will probably be the least of the problems for life on our Earth.

REFERENCES

Campbell, B A and Nottingham, D (1999) Anatomy of a Landslide Created Tsunami at Skagway, Alaska November 3, 1994, *Sci. Tsunami Hazards*, 17, 19–48.

Crawford, D A and Mader, C L (1998) Modeling Asteroid Impact and Tsunami, *Sci. Tsunami Hazards*, 16, 21–30.

Fritz, H M, Hager, W H, and Minor, H E (2001) Lituya Bay Case: Rockslide Impact and Wave Run-Up, *Sci. Tsunami Hazards*, 19, 3–22.

Kowalik, Z (1997) Landslide-Generated Tsunami in Skagway, Alaska, *Sci. Tsunami Hazards*, 15, 89–106.

Mader, C L (1988) *Numerical Modeling of Water Waves*, University of California Press, Berkeley, CA.

Mader, C L (1990) Numerical Tsunami Flooding Study – I, *Sci. Tsunami Hazards*, 8, 79–96.

Mader, C L and Curtis, G (1991) Modeling of Hilo, Hawaii Tsunami Inundation, *Sci. Tsunami Hazards*, 9, 85–94.
Mader, C L (1996) Asteroid Tsunami Inundation of Hawaii, *Sci. Tsunami Hazards*, 14, 85–88.
Mader, C L (1997) Modeling the 1994 Skagway Tsunami, *Sci. Tsunami Hazards*, 15, 3–22.
Mader, C L (1998) Asteroid Tsunami Inundation of Japan, *Sci. Tsunami Hazards*, 16, 11–16.
Mader, C L (1999) Modeling the 1958 Lituya Bay Mega-tsunamis, *Sci. Tsunami Hazards*, 17, 57–67.
Miller, D (1960) *Giant Waves in Lituya Bay*, Alaska Geological Survey Professional Paper 354-C, US Government Printing Office, Washington, DC.
Nottingham, D (1997) The 1994 Skagway Tsunami Tsunami Tide Gauge Record, *Sci. Tsunami Hazards*, 15, 81–88.
Pararas-Carayannis, G (1999) Analysis of Mechanism of Tsunami Generation in Lituya Bay, *Sci. Tsunami Hazards*, 17, 193–206.
Solem, J C (1999) Comet and Asteroid Hazards: Threat and Mitigation, *Sci. Tsunami Hazards*, 17, 141–154.
Steele, D (1995) *Rogue Asteroids and Doomsday Comets*, John Wiley & Sons, London.
Whitmore, P M and Sokolowski, T J (1996) Predicting Tsunami Amplitudes along the North American Coast from Tsunamis Generated in the Northwest Pacific Ocean during Tsunami Warnings, *Sci. Tsunami Hazards*, 14, 147–166.

FURTHER READING

Dudley, W C and Lee, M (1998) *Tsunami*, University of Hawaii Press, Honolulu, HI.
Myles, D (1985) *The Great Waves*, McGraw-Hill, New York.

Tunguska Phenomenon

Sergey Yazev
Astronomical Observatory of Irkutsk State University, Irkutsk, Russia

The ight of a giant bolide observed early in the morning of 30 June 1908 ended in a powerful explosion in the basin of the Podkamennaya Tunguska river (East Siberia, Russia). The explosion yield is estimated to have been between 10- 15 and 40 megatons TNT. The event was followed by a number of consequences: seismoshock, global airwave, local jump of geomagnetic eld strength, highly scattered light in the atmosphere, extensive elds of luminescent clouds, anomalously bright airglow events, etc. Investigations of the Tunguska Event made by Russian scientists are brie y described. There exist over a hundred hypotheses of differing scienti c credibility to explain the Tunguska catastrophe. The basic theory accepted by most researchers postulates explosive evaporation, at 5- 6 km altitude, of the ice nucleus of a small comet about 60 m in diameter and with a mass from 100 000 to a million tons.

Early in the morning of 30 June 1908, a giant bolide was sighted from an area of 10^5–10^6 km^2 in Siberia (Russia s eastern part), along a path as long as 800 km. Numerous eye-witnesses described it as an incandescent sphere, a barrel-shaped body, a luminous sheaf, a log, or a blackness with flame-colored tails. Its brightness was estimated in terms of from like the Sun to not bright, could be looked at. The fireball in its passage left behind it a luminous smoky trail and made a variety of sounds (a hissing whistle, peals of thunder, and a roar resembling gun volleys). According to the data available before 1964, the travel direction was almost exactly from south to north. A more recent analysis showed that in the closing stage of the flight, the trajectory projection was directed from east to west (azimuth 90–95°). The inclination of the atmospheric trajectory is estimated, by most researchers, at 10–20°, although angles of 30–45° too have been reported.

The flight ended in an explosion in the basin of the Podkamennaya Tunguska river (indigenous name Katonga). The explosion yield is estimated to have been 10–15 (commonly accepted estimate) to 40 megatons TNT. A column of material shot up to a height of 20 km. The sound from the explosion could be heard for 1200 km. The shock wave felled the forest in the area of over 2000 km^2. Within up to 15 km of the explosion epicenter, surface layers of trees were scorched, and a devastating forest fire broke out. Seismographs recorded a shock in Irkutsk (800 km from the epicenter), Tashkent, Tbilisi, and Jena. The airwave passed twice round the globe (recorded in, for example, Copenhagen, Zagreb, Washington, Potsdam, London, Jakarta, etc.).

The Irkutsk magnetic observatory recorded a jump of geomagnetic field strength. The night of July 1st was unusually light at more than 150 locations in Siberia, Central Asia, the European part of Russia, and Western Europe. A highly scattered light in the atmosphere and extensive fields of luminescent clouds were observed. During the summer of 1908 (from June 21), anomalously bright airglow events were witnessed.

The first expedition (1927) to the site of meteoritic fall was organized by Leonid Kulik (who died in German captivity in 1942). He discovered a giant area of fallen trees, and a dead standing forest devoid of twigs and branches, with a large marsh in the middle. A search for the meteoritic remains made by Kulik s subsequent five expeditions was unsuccessful. Eye witness inquiries and analyses of the expeditions results were published in the 1940s in Russia by Eugene Krinov.

Later expeditions (1958–1962) were led by K Florensky. No meteorite craters were discovered, and those existing in the area were found to be of thermokarst origin. Metallic microinclusions in the soil were similar to those in the area of Moscow, as well as in Antarctica and were of extraterrestrial origin (cosmic dust), but were not associated with the Tunguska event. It was concluded that the explosion occurred at a height of 10–12 km. During a subsequent period of many years, a comprehensive volunteer expedition under the direction of N Vasiliev worked in the area. They used novel techniques for identifying extraterrestrial dust, and theoretical analyses of the explosion and ballistic wave and of the physics of the Tunguska explosion. An increased content of a radioactive isotope, cesium-137, was detected in layers formed after 1908. It was found that the soils within 30 km around the explosion center underwent a reversal of magnetization. During the summer of 1999, a Russian–Italian expedition also studied the area using remote sensing aircraft-based methods. Bottom set beds of Chako Lake in the vicinity of the epicenter were studied with a catamaran-based sonar. In spite of a plethora of false reports (in 1994, for example), no debris of the hypothetical meteorite has ever been found.

At the same time in Antarctica, for example, deposits from 1908 were found to contain great amounts of iridium of cosmic origin, supposedly scattered in the atmosphere during the Tunguska event.

There exist over a hundred hypotheses for explaining the Tunguska catastrophe. Among the most exotic conjectures are, the explosion of an extraterrestrial spacecraft, the collision of the Earth with a micro-black hole or a meteorite made up of antimatter, and detonation of a giant swarm of gnats. Some recognition was accorded the theory of destruction of a solar plasmoid, as well as the theory of interaction of tectonic (the explosion occurred near an ancient paleovolcanic chimney) and atmospheric electromagnetic processes. The basic theory accepted by most researchers, however, postulates explosive evaporation, at 5–6 km altitude, of the ice nucleus of a small comet (about 60 m in diameter and with a mass from 100 000 to a million tons, the former estimate being preferable), probably with inclusions typical of cometary nuclei. This explanation would account for the absence of any persisting fragments of the Tunguska body and the anomalous phenomena in the atmosphere induced by the injection of cometary dust and aerosols. Explaining the changes in the atmospheric flight of the Tunguska bolide is troublesome, and the problem is dramatized by a drastic diversity of eyewitness accounts (mostly from inexpert people questioned long after the event). A central hypothesis is that the body underwent fragmentation during its atmospheric flight.

See also: Asteroids and Comets, Effects on Earth, Volume 1.

Typhoons

see Hurricanes, Typhoons and other Tropical Storms – Descriptive Overview (**Volume 1**); Hurricanes, Typhoons and other Tropical Storms – Dynamics and Intensity (**Volume 1**)

U

UARS (Upper Atmosphere Research Satellite)

The UARS was launched by the National Aeronautics and Space Administration (NASA) on the Space Shuttle Discovery on September 12th, 1991 and is still providing useful data as of early 2000. The UARS was deployed from Discovery at an orbital altitude of 585 km and an inclination of 57°.

Ten instruments are contained on the UARS, which is about 10 m long and weighs 6804 kg. The mission goals of UARS were to study the physical processes acting within and upon the stratosphere, mesosphere, and lower thermosphere. Four instruments (Cryogen Limb Array Etalon Spectrometer (CLAES), Improved Stratospheric and Mesospheric Sounder (ISAMS), Microwave Limb Sounder (MLS) and Halogen Occultation Experiment (HALOE)) provide temperature profiles and concentrations of ozone, methane, water vapor and other key trace species such as chlorofluorocarbons, hydrogen chloride and chlorine monoxide.

Two instruments (High Resolution Doppler Imager (HRDI) and the Wind Imaging Interferometer (WINDII)) map the upper atmosphere wind fields that shape the global distributions of chemical species. Four instruments (Active Cavity Radiometer Irradiance Monitor (ACRIM), Particle Environment Monitor (PEM), Solar/Stellar Irradiance Comparison Experiment (SOLSTICE) and Solar Ultraviolet Spectral Irradiance Monitor (SUSIM)) monitor energy inputs from the Sun, which strongly influences both chemistry and dynamics. Since its launch, UARS data sets have contributed to many important scientific findings. For example, UARS measurements were used to confirm the role of chlorofluorocarbons in ozone depletion, clarify the chemical processes that cause the Antarctic ozone hole, demonstrate the dominance of human-made chlorofluorocarbons in the chlorine and fluorine amounts in the stratosphere, derive the movement and mixing of air in the stratosphere and mesosphere both globally and seasonally, and quantify the solar ultraviolet spectral irradiances over both maximum and minimum levels of solar activity. Eight of the original 10 instruments continue to take measurements (as of March 2000).

See also: Earth Observing Systems, Volume 1.

FURTHER READING

UARS Project Science Office homepage http://umpgal.gsfc.nasa.gov/.

MICHAEL C MACCRACKEN *USA*

Ultraviolet Radiation

E C Weatherhead and Amy Stevermer
University of Colorado, Boulder, CO, USA

Ultraviolet (UV) radiation is only a small portion of the radiation we receive from the Sun, but it has a large impact on biological activity. UV radiation has become a topic of increasing concern because of depletion of stratospheric ozone, observed since the 1970s.

Since the discovery of the Antarctic ozone hole and depletion of ozone at mid-latitudes, a number of scienti c efforts have been launched to monitor UV radiation. The interest in UV radiation arises for a number of reasons, including the following:

1. *detecting changes in the amount of radiation reaching the Earth;*
2. *increasing public awareness of potentially harmful exposures to UV radiation through daily reports carried on television weather channels and in newspapers;*
3. *assessing UV's biological effects;*
4. *establishing a basic UV radiation climatology.*

PHYSICAL DEFINITION

UV radiation is defined as all radiation between 100 and 400 nm. Although there are other sources of UV radiation,

such as welding arcs and some lamps, most UV radiation reaches the biosphere from the Sun. The UV portion is less than 10% of the total energy output from the Sun. Most of the UV radiation is then absorbed or scattered back to space by the Earth s atmosphere, resulting in very little UV radiation reaching the Earth s surface.

All radiation from the Sun travels in the form of electromagnetic waves. Wavelengths just shorter than those of light at the violet end of the visible spectrum are referred to as UV radiation.

UV radiation can be divided into three basic categories:

- UV-A (320–400 nm), containing most of the energy in the UV spectrum and most likely to reach the biosphere; weakly affected by ozone;
- UV-B (280–320 nm), some of which reaches the biosphere; strongly affected by ozone;
- UV-C (100–280 nm), little of which reaches the biosphere; primarily scattered and absorbed by atmospheric oxygen, nitrogen and ozone.

Some scientists consider 315 nm to be a more appropriate dividing point between UV-A and UV-B radiation categories; an absolute division between UV-A and UV-B has not yet been established.

UV LEVELS

The amount of UV radiation reaching the surface is measured at over 100 sites around the world, primarily in North America and Europe. Surface UV can be measured by ground-based instruments, such as a spectrophotometer, or estimated by satellite instruments, such as the Total Ozone Mapping Spectrometer (TOMS). Ground-based instruments are best at measuring the UV reaching the surface at a particular location, but require accurate calibrations and standardization to ensure accurate observations. Satellite instruments provide global coverage but the observations must be corrected for clouds and other atmospheric variables.

The ongoing measurements from both the surface and satellites show how a number of factors, including ozone, clouds, and ground cover, can affect UV levels. Generally, as stratospheric ozone amounts decrease, more UV radiation is able to penetrate to the surface. The relationship between ozone and surface level UV radiation is shown in Figure 1 for several locations around the world. In addition to ozone concentration, other important factors governing UV levels are sun angle and cloud amount. Differences in sun angle can strongly affect the amount of UV radiation received directly. Frederick and Snell (1990) found that clouds can reduce the amount of UV reaching the ground by as much as 90% on a given day.

UV levels can also be affected by surface reflectivity, altitude, and pollutant concentrations, including both aerosols and trace gases. Studies of the effects of these variables are important for several reasons. Understanding the processes affecting UV radiation allows estimation of UV levels at times and in places where measurements are not available. The improved understanding also helps scientists estimate how a changing environment will affect UV levels and vice versa.

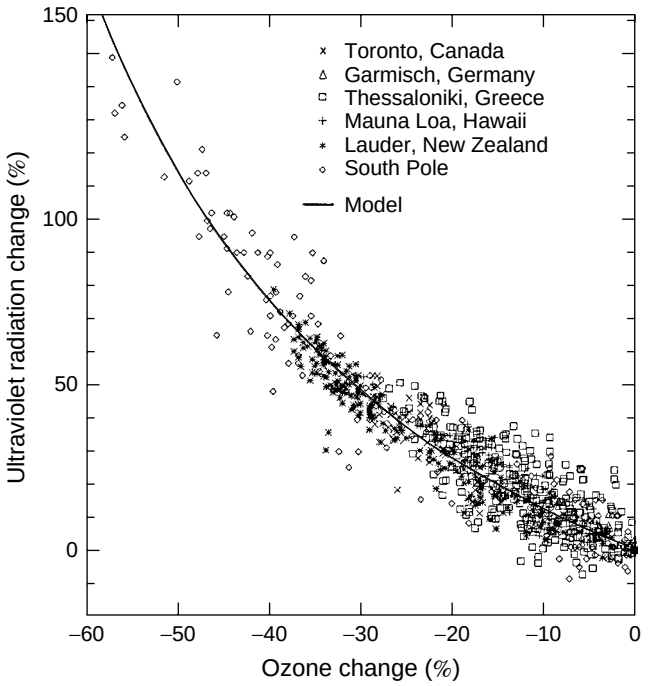


Figure 1 Increase in biologically effective UV radiation as a function of the decrease in ozone amount in the overlying atmosphere. (From the World Meteorological Organization Scientific Assessment of Ozone Depletion, 1998)

UV EFFECTS

UV radiation has a variety of effects on plants, animals, and materials in the biosphere, most of them detrimental. UV radiation is known to affect materials, particularly plastics and polymers with easily broken bonds. UV radiation provides the energy to drive both stratospheric and tropospheric chemistry (*see* Stratosphere, Chemistry, Volume 1; Troposphere, Ozone Chemistry, Volume 1). In the troposphere, UV radiation contributes to ozone formation. Unlike stratospheric ozone, ozone in the troposphere is a pollutant that can damage plants, human health, and natural materials.

The impacts of UV radiation on biological systems have been extensively studied. These effects range from human health impacts to impacts on single cell organisms. These effects tend to arise because each UV photon has more energy than most other photons encountered in nature. As shown in Figure 2, the biological response to UV radiation

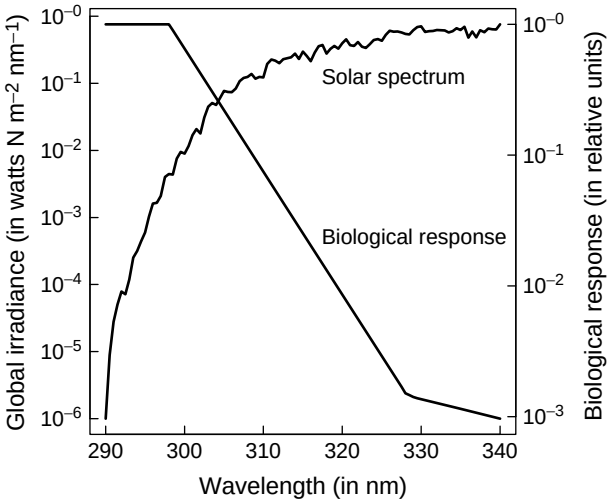


Figure 2 Biological response to UV radiation as a function of wavelength. Shorter wavelength UV radiation is the most important in causing biological damage. (Reproduced from Weatherhead and Webb, 1997)

is most significant at shorter wavelengths. Additionally, UV radiation, particularly UV-B radiation, is able to efficiently break DNA bonds. While UV radiation, particularly UV-A, can assist in repairing DNA damage, the primary effect of UV radiation appears to be damaging to both DNA and individual cells.

The effects of UV radiation on individual species and ecosystems have been studied both in the laboratory and in field experiments. UV radiation is known to have primarily deleterious effects on individual species, including phytoplankton, zooplankton, amphibians, terrestrial plants and fish. However, the effects of UV radiation on ecosystems can be unexpected or confounding. Studies by Bjorn *et al.* (see AMAP, 1998) have shown that the growth rates of some shrubs can be reduced or increased by additional UV radiation, depending on available water supplies. Other studies have shown that UV radiation can change ecosystems by favoring UV-resistant species. The effects can be strong and measurable, but are not always intuitive or obvious.

The most well-known effects of UV radiation on people include sunburn and snow blindness (photo-keratitis). UV radiation has also been linked to skin cancer, immune suppression, and cataract formation as well as to a number of other dermatological and ocular problems. These effects have been observed in controlled laboratory experiments and derived from epidemiological studies. These studies help to explain, for instance, the differences in skin cancer incidence by latitude and altitude. Skin cancer risks are often a function of lifestyle, increasing for those who pursue leisure activities outdoors. Persons who allow themselves considerable exposure to the Sun are at a higher risk, regardless of skin type. Intermittent but extreme bouts of sun exposure, such as those an indoor worker might receive when relaxing outdoors on days off work, have been linked to melanoma, a rare but life-threatening form of skin cancer.

RECENT CHANGES

Decreases in stratospheric ozone over the past three decades have been well documented and their causes are now fairly well understood (*see* Stratosphere, Ozone Trends, Volume 1). Figure 3 shows how the observed ozone decreases depend on latitude, with depletion being greatest in the mid-latitudes and at the poles. While we know from daily data that decreases in ozone amount lead to increases in UV radiation, long-term records of UV radiation are more difficult to interpret. Scotto *et al.* (1988) found that UV data from 1974 to 1985 indicated a decrease in UV radiation when an increase was expected. Weatherhead *et al.* (1997) later re-examined the long-term records and showed that the trends were inconclusive. Blumthaler and Ambach (1990) found an increase of almost 1% per year in spring alpine UV data. Using three years of data, Kerr and McElroy (1993) found an increase in UV radiation that agreed with the decrease in ozone observed during the same time period.

A question remains about UV levels in centuries past. No consistent UV monitoring existed before the 20th century, so there is no basis for a definitive answer. However, current understanding of the effects of UV allows some insights when examining biological records. While current estimates are very rough, it is likely that UV radiation was not much different during the past millennium than it was prior to the 1970s, when anthropogenic ozone depletion began.

Future UV levels are likely to be dictated by both anthropogenic factors and natural variations. Human activities are likely to affect the concentrations of ozone depleting

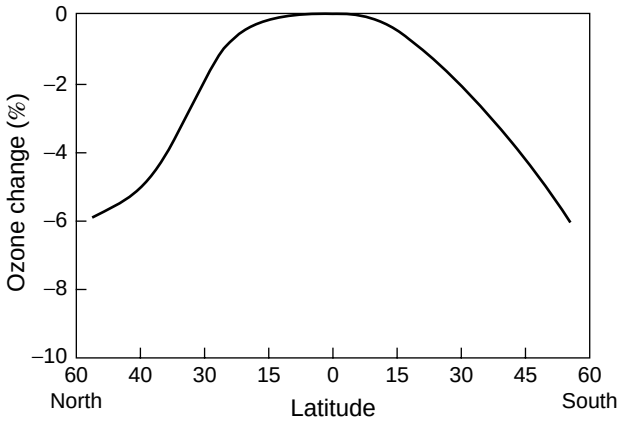


Figure 3 Schematic diagram of the latitudinal distribution of ozone depletion: 1979–1997. (Reproduced from the World Meteorological Organization Scientific Assessment of Ozone Depletion, 1998)

substances and to contribute to climate change in ways that are likely to lead to long-term changes in UV levels. The slow recovery predicted for the ozone layer and the uncertainty associated with enforcing international agreements to control emissions of ozone depleting substances are likely to mean that UV levels will be elevated above natural levels for many decades to come.

Whether elevated UV levels are a critical environmental threat or only one of the many changes affecting the environment is an important question. The answer depends on the total impact of UV to the environment, but our current understanding of the full impact of elevated UV levels does not allow a clear answer to this question.

REFERENCES

AMAP (Arctic Monitoring and Assessment Programme) (1998) *AMAP Assessment Report: Arctic Pollution Issues*, AMAP, Oslo, Norway.

Blumthaler, M and Ambach, W (1990) Indication of Increasing Ultraviolet-B Flux in Alpine Regions, *Science*, 248, 206–208.

Frederick, E and Snell, H E (1990) Tropospheric Influence on UV Radiation: the Role of Clouds, *J. Clim. Meteorol.*, 3, 373–381.

Kerr, J and McElroy, C T (1993) Evidence for Large Upward Trends of Ultraviolet-B Radiation Linked to Ozone Depletion, *Science*, 262, 1032–1034.

Scotto, J, Cotton, G, Urbach, F, Berger, D, and Fears, T (1988) Biologically Effective Ultraviolet Radiation: Surface Measurements in the United States, 1974–1985, *Science*, 239(4841), 762–763.

Weatherhead, E C, Tiao, G C, Reinsel, G C, Frederick, J E, DeLuisi, J J, Choi, D S, and Tam, W K (1997) Analysis of Long-term Behavior of Ultraviolet Radiation Measured by Robertson–Berger Meters at 14 sites in the United States, *J. Geophys. Res.*, 102(D7), 8737–8754.

Weatherhead, E C and Webb, A R (1997) International Response to the Challenge of Measuring Solar Ultraviolet Radiation, *Radiat. Prot. Dosimet.*, 72(3–4), 223–229.

Upwelling–Diffusion Model

see Energy Balance Climate Models (**Volume 1**)

Urban Heat Island

see The Global Temperature Record (*Opening essay*, **Volume 1**); Urban Heat Island (**Volume 3**)

V

VEI (Volcanic Explosivity Index)

see VEI (Volcanic Explosivity Index) (**Volume 3**)

Vertical Soundings

see Radiosondes (**Volume 1**)

Volcanic Eruption, El Chichón

On March 29th, April 3rd and April 4th, 1982, the El Chichón volcano in Southern Mexico erupted, injecting about 7 million metric tonnes of sulfur dioxide (SO_2) and 20 million metric tonnes total of particulate material into the stratosphere. The cloud of volcanic gases and particles traveled westward, circling the Earth in three weeks. The April 4th eruption was the largest of the three, and its cloud went right over Hawaii a couple days after the eruption, allowing detailed observations from instruments at the Mauna Loa Observatory. The cloud quickly developed into a band extending from the Equator to 30 °N for more than six months, and then gradually spread even more widely. This was one of the largest volcanic eruptions of the 20th century, only exceeded in stratospheric input by the 1991 Mt. Pinatubo eruption. The eruption took place without warning, and killed many people in towns near the volcano. El Chichón had never erupted before in historic times.

The eruption took place just as the largest El Niño of the century so far was beginning. (In fact the volcanic cloud in the stratosphere fooled the satellite sensors which monitor ocean temperatures into thinking ocean temperatures were normal, whereas they had warmed substantially. Thus, scientists were not aware of the El Niño until months after it had started (*see* El Niño, Volume 1). A dense network of buoys in the tropical Pacific now prevents that from happening.) Because of the simultaneous eruption and appearance of warm ocean temperatures, several scientists suggested that the El Chichón eruption caused the El Niño. But climate model simulations and detailed studies of past eruptions and El Niños have shown that there is no plausible mechanism connecting these two events, and in fact their timing was a coincidence.

Because of the simultaneous eruption and El Niño, the climatic system felt the impacts of both, and it was difficult to separate their effects on temperature. Normally a large eruption like this would cool the global climate, especially in the summer, but during the first year after the El Chichón eruption, no large cooling was observed, as the El Niño produced large compensating warming. Studies of these effects using climate models have improved understanding of the climate system and increased confidence in projections of global warming from anthropogenic greenhouse gases, which are made with the same models.

The climatic effects also included winter warming of Northern Hemisphere continents in 1982–1983, with the temperature over North America, Europe, and Siberia much higher than normal. During the same winter, it was colder than normal over Alaska, Greenland, the Middle East, and China. The volcanic aerosols in the stratosphere produced heating, which changed the atmospheric wind patterns, to one we now call the positive phase of the Arctic Oscillation (*see* Arctic Oscillation, Volume 1).

See also: Volcanic Eruptions, Volume 1.

ALAN ROBOCK *USA*

Volcanic Eruption, Krakatau

Krakatau (Krakatoa) is one of a linear series of Javan–Sumatran volcanoes that mark a destructive margin, where the Indo–Australian plate is subducted under the Asian plate. Krakatau was once one large island, which was destroyed in eruptions (possibly in the 6th century,

and possibly of global climatic significance), to leave three fragments (Panjang, Sertung, and Rakata Islands) of the original volcanic walls, in a broken ring. Renewed activity rebuilt one island back into the centre of the caldera. Following eruptions in 1680, it was dormant until May, 1883. The three peaks of this island, Perboewatan, Danan and Rakata, each activated as the cycle of building and destruction reached its peak. The sequence ended catastrophically on August 27th 1883, with an output of 18 km^3 of ejecta. The surrounding region, for a radius of 80 km, experienced 57 h of darkness. Some 36 000 people in the coastal lowlands of the Sunda Straits were killed by tsunami, giant waves up to 40 m in height. These were generated by the emptying of the magma chamber, which led to the outer walls failing, and two-thirds of the main island being displaced in a repeat of the pre-historic caldera collapse. The marine pressure waves were detected as far away as the English Channel, and the air waves from the eruption made four journeys around the globe to Krakatau s antipodes in Colombia. The violence ended abruptly leaving a greatly reshaped archipelago. In places, extensive new land surfaces had been created, and on the persisting areas of land, an average of 60–80 m of pyroclastic ashes had been emplaced. These completely barren landscapes were soon colonized by plants and animals, and succession led to forest closure by 1930. The (temporary) climatic impacts were global in extent as volcanic ash and gases were injected 40 km into the stratosphere, also producing unusual visual effects (e.g., broad opalescent corona (Bishop s ring) around the Sun and spectacular sunsets around the Northern Hemisphere). Comparative studies show that eruptions of the magnitude of the 1883 Krakatau eruption can decrease global temperatures by about 0.2–0.4° on an annual average for 1–2 years following the eruption. After 1883, Krakatau remained dormant until 1927, when renewed activity in the center of the caldera led to the formation (in 1930) of a new island, Anak Krakatau, now approximately 2 km in diameter and over 280 m in height.

See also: Volcanic Eruptions, Volume 1; Volcanoes and the Environment, Volume 3.

ROBERT WHITTAKER *UK*

Volcanic Eruption, Mt. Pinatubo

On June 15, 1991, the Mt. Pinatubo volcano, on Luzon Island in the Philippines, erupted, injecting about 20 million metric tonnes of SO_2 into the stratosphere. The cloud of volcanic gases and particles traveled westward, circling the Earth in 3 weeks. Although it started at 15 °N latitude, by the time it had circled the Earth once, the cloud straddled the Equator, extending from about 15 °S to 15 °N. This was the largest volcanic eruption of the 20th century.

The eruption was fairly well forecast by a team of Filipino and American geologists. Many lives were saved by these warnings, but a typhoon hit the area during and soon after the eruption. The thick layers of volcanic dust on the roofs of some of the shelters were soaked by the intense rain and collapsed, killing hundreds. The US Clark Air Force base at Angeles City, only 10 km from the eruption, was destroyed and never reopened.

The climatic effects included global cooling of about 0.3 °C for about 2 years. The cooling was particularly strong over Northern Hemisphere continents in the summer of 1992, with cooling more than 3 °C in the northern US and southern Canada. The cooling was temporary and 1998 was the warmest year in the past 1000 years. This cooling was predicted by climate models, which are also used for predicting global warming from anthropogenic greenhouse gases, thus giving strength to the global warming predictions.

Due to the heating of the lower tropical stratosphere by the volcanic aerosols, the enhanced temperature gradient produced a changed atmospheric wind pattern in the troposphere which had large impacts on winter temperatures. In the winter of 1991–1992, the temperature over North America, Europe, and Siberia was much higher than normal, and over Alaska, Greenland, the Middle East, and China it was colder than normal. In fact, it was so cold that winter that it snowed in Jerusalem, a very unusual occurrence.

See also: Volcanic Eruptions, Volume 1.

ALAN ROBOCK *USA*

Volcanic Eruption, Tambora

Alan Robock
Rutgers University, New Brunswick, NJ, USA

The Tambora volcano (8.25 °S, 118.00 °E) on the island of Sumbawa in Indonesia, erupted on April 10th, 1815, sending a massive cloud of aerosols into the stratosphere. This was the largest eruption of the past 500 years.

In the summer of 1816, Lord Byron and his friends Mary and Percy Bysshe Shelley spent their vacation on the shores of Lake Geneva in Switzerland. Rather than the normally delightful weather they expected, " it proved a wet, ungenial

summer, and incessant rain often con ned us for days to the house." (Shelley, 1818, p. xxii). In fact it was so gloomy that they held a writing contest to see who could come up with the best ghost story, the most appropriate expression of their common dismal feelings. Mary won, and the famous novel Frankenstein was born. The book both begins and ends with the monster trudging across a bleak ice-covered wasteland.

Byron was also inspired that summer and wrote *Darkness*, the beginning of which is reproduced below:

I had a dream, which was not all a dream.
The bright sun was extinguish d, and the stars
Did wander darkling in the eternal space,
Rayless, and pathless, and the icy earth
Swung blind and blackening in the moonless air;
Morn came and went – and came, and brought no day,
And men forgot their passions in the dread
Of this their desolation; and all hearts
Were chill d into a selfish prayer for light:
And they did live by watchfires – and the thrones,
The palaces of crowned kings – the huts,
The habitations of all things which dwell,
Were burnt for beacons; cities were consumed,
And men were gather d round their blazing homes
To look once more into each other s face; •••

The summer in New England that year also had extreme weather. There was snow every month of the year, and the poor harvests started a mass migration from the US East Coast across the Appalachian Mountains to the Midwest. For these reasons, 1816 has come to be known as the year without a summer. Although the cool summer of 1816 was not universally felt around the globe, large regions, including North America, Europe, Argentina, India, and China experienced record anomalies. Even portions of South Africa experienced cool and wet conditions in the growing season, but Southern Hemisphere climate would not be expected to change much in response to Tambora, due to the overwhelming oceanic influence.

The extensive records of the Hudson s Bay Company prove extremely useful in this regard for determining Canada s climatic record of this period. The effects of the 1816 summer on agricultural productivity of New England did not have much to do with the average temperature, which was low but not abnormally so, but with a series of killing frosts that reduced the growing season. This, along with a severe drought, reduced agricultural output to record low levels. In Europe, by contrast, there were record low temperatures accompanied by above average rainfall and cloudiness, the combination of which slowed the growth of crops and produced fungus and molds. Thus, food production was negatively affected in both regions, but by different mechanisms, pointing out the complex relationship between climate and human impacts.

The Tambora eruption killed more than 90 000 people in the immediate vicinity, 35% of the people on Sumbawa (48 000 in two years; 10 000 immediately and 38 000 from hunger and disease) and 44 000 on Lombok 160 km to the east. A more indirect effect of the Tambora eruption has also been suggested. Because of the bad weather, a typhus epidemic broke out in Ireland, spread to England and Scotland, and killed 65 000 people. Bad harvests in India may have contributed to a cholera epidemic, which spread to Asia and Europe producing the great cholera pandemic of the century.

In addition to historical records, proxy records such as those from tree rings allow us to explore the hemispheric temperature record. While global cooling followed the eruption, climate had already begun to cool before 1815. Ice core records from Greenland and Antarctica preserve sulfuric acid layers from volcanic eruptions, and while all records at both poles show a Tambora layer, they also all show a layer from an eruption almost as large in 1808 or 1809, still unidentified. Thus, the climate was probably already cooling from a previous large eruption and several smaller ones also observed between the two large ones. If Tambora had erupted in a warmer climate, the effects would have been less harsh. While this is true in an absolute sense, in a relative sense they would have been larger and more easily attributable to the volcano.

See also: Volcanic Eruptions, Volume 1.

REFERENCE

Shelley, M (1818) *Frankenstein*, Bantam Classic edition, Bantam Books, New York, 1–209.

Volcanic Eruptions

Alan Robock
Rutgers University, New Brunswick, NJ, USA

Volcanic eruptions can inject into the stratosphere tens of teragrams of chemically and microphysically active gases and solid aerosol particles that can affect the Earth's radiative balance and climate and disturb the stratospheric chemical equilibrium. The volcanic aerosol cloud forms in several weeks by the conversion of sulfur dioxide (SO_2) to sulfate aerosol and its subsequent microphysical transformations. The resulting cloud of sulfate aerosol particles, with a residence time of a couple of years, has important impacts on both shortwave and longwave radiation. The

resulting disturbance to the Earth's radiation balance affects surface temperatures through direct radiative effects as well as through indirect effects on the atmospheric circulation. These aerosol particles also serve as surfaces for chemical reactions that can deplete stratospheric ozone.

*Volcanism has long been implicated as a possible cause of weather and climate variations. Even 2000 years ago, Plutarch and others pointed out that the eruption of Mt. Etna in 44 BC dimmed the Sun and suggested that this caused crops to shrivel and produced famine in Rome and Egypt. Because of their large effects in the English-speaking world, the largest eruptions of the past 250 years have each drawn attention to the atmospheric and potential climatic effects. The 1783 eruption in Iceland produced large effects all that summer in Europe. Benjamin Franklin, the US ambassador to France published the first paper on the subject in more than 1800 years, suggesting that the Lakagigar eruption in Iceland in 1783 might have been responsible for the abnormally cold summer of 1783 in Europe and the cold winter of 1783- 1784 (*see *Franklin, Benjamin, Volume 1). If an eruption identical to the 1783 Laki eruption took place today, air traffic in the Northern Hemisphere would be disrupted for six months or more. The 1815 Tambora eruption produced the year without a summer in 1816 (*see *Volcanic Eruption, Tambora, Volume 1) and inspired Mary Shelly to write Frankenstein that summer while vacationing on the shores of Lake Geneva, Switzerland. She and her husband, Percy Bysshe Shelley and their friend Lord Byron had a contest to see who could write the scariest ghost story, inspired by the cold, gray weather that kept them housebound. Mary Shelley won.*

On average, major volcanic eruptions have occured about every 20 years over the last few centuries (see Table 1). The most extensive study of the impacts of a single volcanic eruption was carried out by the British Royal Society examining the 1883 Krakatau eruption (*see* Volcanic Eruption, Krakatau, Volume 1). The Society produced a beautiful volume including watercolors of the volcanic sunsets near London. This eruption was probably the loudest explosion of historic times, and the book includes color figures of the resulting pressure wave s four circuits of the globe as measured by microbarographs.

The 1963 Agung eruption produced the largest stratospheric dust veil in the Northern Hemisphere (NH) in more than 50 years, and inspired many modern scientific studies. While the Mt. St. Helens eruption of 1980 was very explosive, it was a lateral blast and did not inject much sulfur into the stratosphere, where aerosol particles can remain several years and have a world-wide effect on global average temperature. Because no materials reached the stratosphere, this eruption had very small global effects and its tropospheric effects lasted only a few days. However, it occurred in the US, and so received much attention. In contrast, the subsequent 1982 El Chichón and 1991 Mt. Pinatubo eruptions produced very large stratospheric aerosol clouds and large climatic effects (*see* Volcanic Eruption, El Chichón, Volume 1; Volcanic Eruption, Mt. Pinatubo, Volume 1). Despite their different climatic influences, quantification of the size of these eruptions has proven difficult because different measures reveal different information. For example, one could examine the total mass ejected, the explosiveness, or the sulfur input to the stratosphere.

Table 1 Major volcanic eruptions of the past 250 years

Volcano	Year of eruption
Grimsvotn (Laki or Lakagigar), Iceland	1783
Tambora, Sumbawa, Indonesia	1815
Cosiguina, Nicaragua	1835
Askja, Iceland	1875
Krakatau, Indonesia	1883
Okataina (Tarawera), North Island, New Zealand	1886
Santa Maria, Guatemala	1902
Ksudach, Kamchatka, Russia	1907
Novarupta (Katmai), Alaska, US	1912
Agung, Bali, Indonesia	1963
Mt. St. Helens, Washington, US	1980
El Chichón, Chiapas, Mexico	1982
Mt. Pinatubo, Luzon, Philippines	1991

Volcanic eruptions inject several different types of particles and gases into the atmosphere (Figure 1) (*see* Aerosols, Stratosphere, Volume 1). In the past, it was only possible to estimate these volatile inputs based on measurements from active, but not explosive, eruptions and from remote sensing of the resulting aerosol clouds using lidar, radiometers, and satellites. The serendipitous discovery of the ability to monitor SO_2 using the Total Ozone Mapping Spectrometer satellite instrument, however, has given us a new tool to directly measure stratospheric injection of gases from eruptions.

The major component of volcanic eruptions is magmatic material, which emerges as solid, lithic material or solidifies into large particles that are referred to as *ash* or *tephra*. These particles fall out of the atmosphere very rapidly, on time scales of minutes to a few days, and thus have no climatic impacts. However, this temporary large atmospheric loading can reduce the amplitude of the diurnal cycle of surface air temperature in the region of the tropospheric cloud, and this material is very dangerous for aircraft. But these effects disappear as soon as the particles settle to the ground. When an eruption column still laden with these hot particles descends down the slopes of a volcano, the resulting pyroclastic flow can be deadly to those unlucky enough to be at the base of the volcano. The destruction of Pompeii and Herculaneum after the 79 AD Vesuvius eruption is the most famous example.

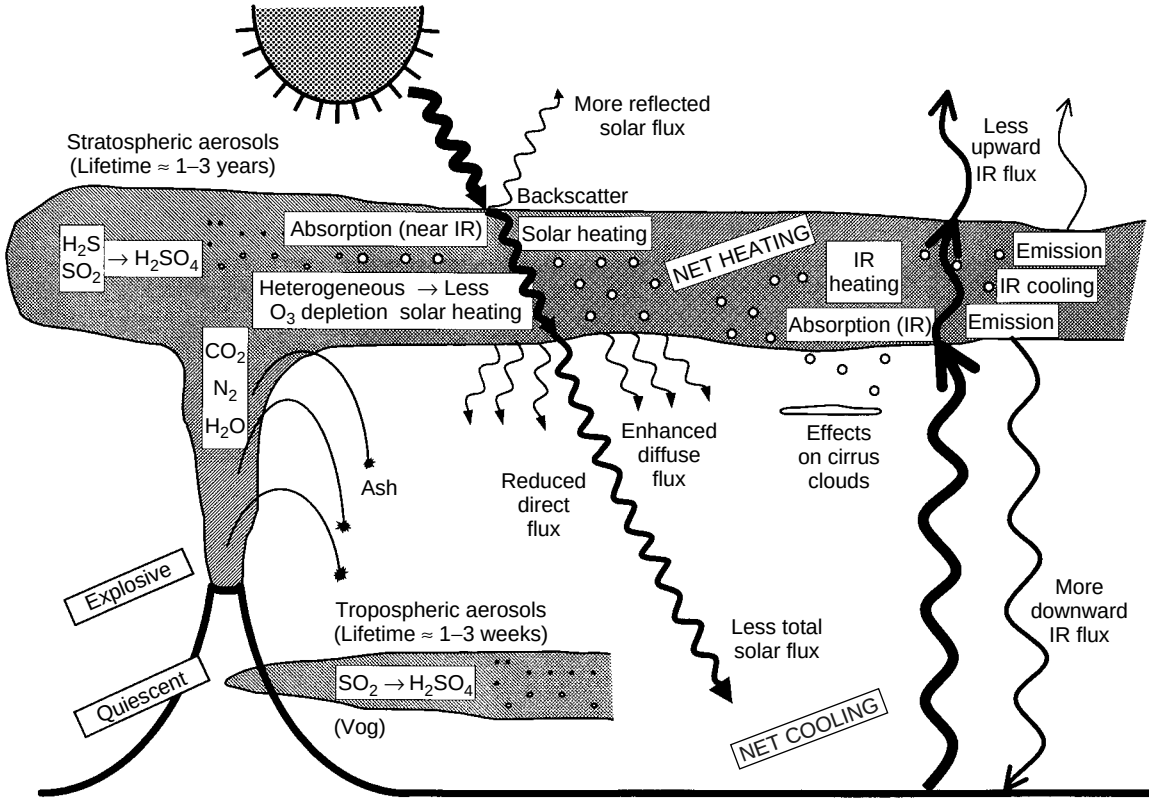


Figure 1 Schematic diagram of volcanic inputs to the atmosphere and their effects. (Adapted from Plate 1 of Robock, 2000, © Copyright, American Geophysical Union)

Volcanic eruptions typically also emit gases, with water (H_2O), nitrogen (N_2) and carbon dioxide (CO_2) being the most abundant. Over the lifetime of the Earth, these gases have been the main source of the planet s atmosphere and ocean, once the primitive atmosphere of hydrogen and helium was lost to space. The water has condensed into the oceans, the CO_2 has been changed by plants into oxygen (O_2), with some of the C turned into fossil fuels. Of course, we eat the plants and animals that eat the plants, we drink the water, and we breathe the oxygen, so each of us is made of volcanic emissions. The atmosphere is now mainly composed of N_2 (78%) and O_2 (21%), both of which had sources in volcanic emissions.

Of the less abundant gases, both H_2O and CO_2 are important greenhouse gases; however their atmospheric concentrations are so large (e.g., CO_2 is at only 370 ppm and growing) that individual eruptions have a negligible effect on their concentrations and do not directly impact the greenhouse effect. Rather, the most important climatic effect of explosive volcanic eruptions is through their emission of sulfur species to the stratosphere, mainly in the form of SO_2, but sometimes as hydrogen sulfide (H_2S). These sulfur species react with H_2O to form sulfuric acid (H_2SO_4) on a time scale of weeks, and the resulting H_2SO_4 aerosols produce the dominant radiative effect from volcanic eruptions. For example, the 1982 El Chichón eruption injected 7 megatonnes (Mt) of SO_2 into the atmosphere, and the 1991 Pinatubo eruption injected 20 Mt.

Once injected into the stratosphere, the large aerosol particles and small ones being formed by the sulfur gases are rapidly transported around the globe by stratospheric winds. Observations after the 1883 Krakatau eruption showed that the aerosol cloud circled the globe in two weeks. Both the 1982 El Chichón cloud and the 1991 Pinatubo cloud circled the globe in three weeks. Although El Chichón (17 °N) and Mt. Pinatubo (15 °N) are separated by only 2° of latitude, their clouds, after only one circuit of the globe, ended up separated by 15° of latitude, with the Mt. Pinatubo cloud straddling the equator and the El Chichón cloud extending approximately from the equator to 30 °N. Subsequent dispersion of a stratospheric volcanic cloud depends heavily on the particular distribution of winds at the time of eruption. For those trying to reconstruct the effects of older eruptions, this factor adds a further complication, as the latitude of the volcano is not sufficient information.

Quiescent, continuous volcanic emissions also add sulfates to the troposphere, but their lifetimes in that part of the atmosphere are much shorter. Global sulfur emission by volcanoes to the troposphere is about 14% of the total natural and anthropogenic emission, thereby leading to a

cooling influence at the surface. Only if there is a long-term trend in these emissions, however, will they be important contributors to climate change; nevertheless, they must be considered when evaluating the effects of anthropogenic sulfate emissions.

Figure 1 indicates the major radiative processes resulting from the stratospheric aerosol cloud from a major volcanic eruption. The most obvious and well-known effect is on solar radiation. Because the sulfate aerosol particles are about the same size as the wavelength of visible light, with a typical effective radius of 0.5 μm, and they have a single-scatter albedo of about 1, they strongly interact with solar radiation by scattering visible light. Some of the light is backscattered, reflecting sunlight back to space, increasing the net planetary albedo and reducing the amount of solar energy that reaches the Earth s surface. This is the dominant radiative effect at the surface and results in a net cooling there. Much of the solar radiation is forward-scattered, however, resulting in enhanced downward diffuse radiation that somewhat compensates for the large reduction in the direct solar beam.

The longest record of the effects of volcanic eruptions on atmospheric transmission of radiation is the apparent transmission record from the Mauna Loa Observatory in Hawaii (see Figure 2). The effects of the 1963 Agung, 1982 El Chichón, and 1991 Mt. Pinatubo eruptions can be clearly seen. Although the Mt. Pinatubo eruption produced the largest stratospheric input of the three, the center of the El Chichón cloud went directly over Hawaii, while only the edge of the Mt. Pinatubo cloud was observed. The Agung cloud was mostly in the Southern Hemisphere, so only the edge was seen in Hawaii.

The reflection of the setting Sun from the bottom of stratospheric volcanic aerosol layers produces the characteristic red sunsets that are used as one means of detecting past eruptions. The famous 1893 Edvard Munch painting, *The Scream*, shows a red volcanic sunset over the Oslo harbor produced by the 1892 Awu (Indonesia) eruption.

Volcanic eruptions can affect the climate system over many time scales (Table 2). The largest known eruption of the past 100 000 years was the great Toba (on Sumatra, Indonesia) eruption about 71 000 years ago, which occurred intriguingly close to the beginning of a major glaciation (although a cause and effect relationship has yet to be established). Volcanic aerosols can be important causes of temperature changes for several years following large eruptions, and, even on a 100-year time scale, they can be important when their cumulative effects are taken into account. This is a very significant consideration when analyzing the extent of global warming, because the impacts of anthropogenic greenhouse gases and aerosols on climate must be evaluated against a background of continued natural forcing of the climate system from volcanic eruptions, solar variations, and internal random variations from land–atmosphere and ocean–atmosphere interactions. New analyses of the volcanic loading of the atmosphere, based on ice core records, have helped

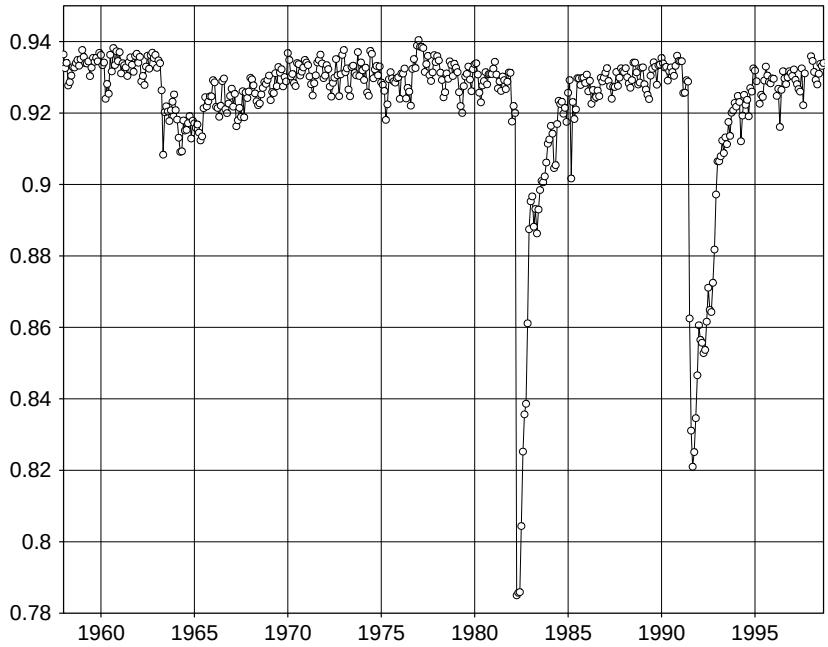


Figure 2 Broadband spectrally-integrated atmospheric transmission factor, measured at the Mauna Loa Observatory, Hawaii (19 °N). Effects of the 1963 Agung, 1982 El Chichón, and 1991 Mt. Pinatubo eruptions can clearly be seen. Years on abscissa indicate January of that year. (Data courtesy of E Dutton, NOAA CMDL, Boulder)

Table 2 Effects of large explosive volcanoes on weather and climate

Effect/mechanism	Begins	Duration
Reduction of diurnal cycle Blockage of shortwave and emission of longwave radiation	Immediately	1–4 days
Reduced tropical precipitation Blockage of shortwave radiation, reduced evaporation	1–3 months	3–6 months
Summer cooling of NH tropics and subtropics Blockage of shortwave radiation	1–3 months	1–2 years
Reduced Sahel precipitation Blockage of shortwave radiation, reduced land temperature, reduced evaporation	1–3 months	1–2 years
Stratospheric warming Stratospheric absorption of shortwave and longwave radiation	1–3 months	1–2 years
Winter warming of NH continents Stratospheric absorption of shortwave and longwave radiation, dynamics	6–18 months	1 or 2 winters
Global cooling Blockage of shortwave radiation	Immediately	1–3 years
Global cooling from multiple eruptions Blockage of shortwave radiation	Immediately	Up to decades
Ozone depletion, enhanced ultraviolet (UV) radiation Dilution, heterogeneous chemistry on aerosols	1 day	1–2 years

identify the effects of volcanic eruptions in the climate record.

Individual large eruptions certainly produce global or hemispheric cooling for two or three years, and this signal can now be more clearly in climate records. It has also been found that the winter following a large tropical eruption is warmer over the NH continents. This counter-intuitive effect is due to a nonlinear response through atmospheric dynamics. It has also recently been realized that volcanic aerosols provide a surface for heterogeneous chemical reactions that destroy ozone. Indeed observations following the Mt. Pinatubo eruption have documented midlatitude ozone depletion caused by volcanic effects.

The winter warming pattern is illustrated in Figure 3, which shows the global lower tropospheric temperature anomaly pattern for the NH winter of 1991–1992, following the 1991 Mt. Pinatubo eruption (*see* Tropospheric Temperature, Volume 1). This pattern is closely correlated with the surface air temperature pattern where the data overlap, but the satellite data allow global coverage. The temperatures over North America, Europe, and Siberia were much higher than normal, and those over Alaska, Greenland, the Middle East, and China were lower than normal. In fact, it was so cold that winter that it snowed in Jerusalem, a very unusual occurrence. As another example of the extreme conditions, Coral at the bottom of the Red Sea died that winter because the water at the surface cooled and convectively mixed the entire depth of the water. The enhanced supply of nutrients produced anomalously large algal and phytoplankton blooms, which smothered the coral. This coral death had only happened before in winters following large volcanic eruptions.

At the tropopause, the boundary between the troposphere and stratosphere, the strongest winds are found in the midlatitudes in the winter and are called the jet stream or polar vortex. The strength of the jet stream depends on the temperature difference (gradient) between the tropics and the polar region, which is largest in the winter when the polar regions cool. For a tropical eruption, the stratospheric heating from volcanic aerosols is larger in the tropics than in the high latitudes, producing an enhanced pole-to-equator temperature gradient, and in the NH winter, a stronger polar vortex and winter warming of NH continents. The stronger jet stream produces a characteristic wind pattern in the troposphere, which warms some regions and cools other ones. This indirect advective effect on temperature is stronger than the radiative cooling effect that dominates at lower latitudes and in the summer.

Because volcanic aerosols normally remain in the stratosphere no more than two or three years, with the possible exception of extremely large eruptions such as that of Toba approximately 71 000 years ago, the radiative effect of volcanoes is interannual rather than interdecadal in scale. A series of volcanic eruptions can, however, significantly raise the mean optical depth of the atmosphere over a longer period and thereby give rise to a decadal-scale cooling. If a period of active volcanism ends for a significant period, such as for the 50-year period from 1912 to 1963, the adjustment of the climate system to an absence of volcanic forcing can (and apparently did) contribute to warming over this period. Furthermore, it is possible that feedbacks involving ice and ocean, which act on longer time scales, can transform the short-term volcanic forcing into a longer-term effect. As a result, the possible role of volcanoes in

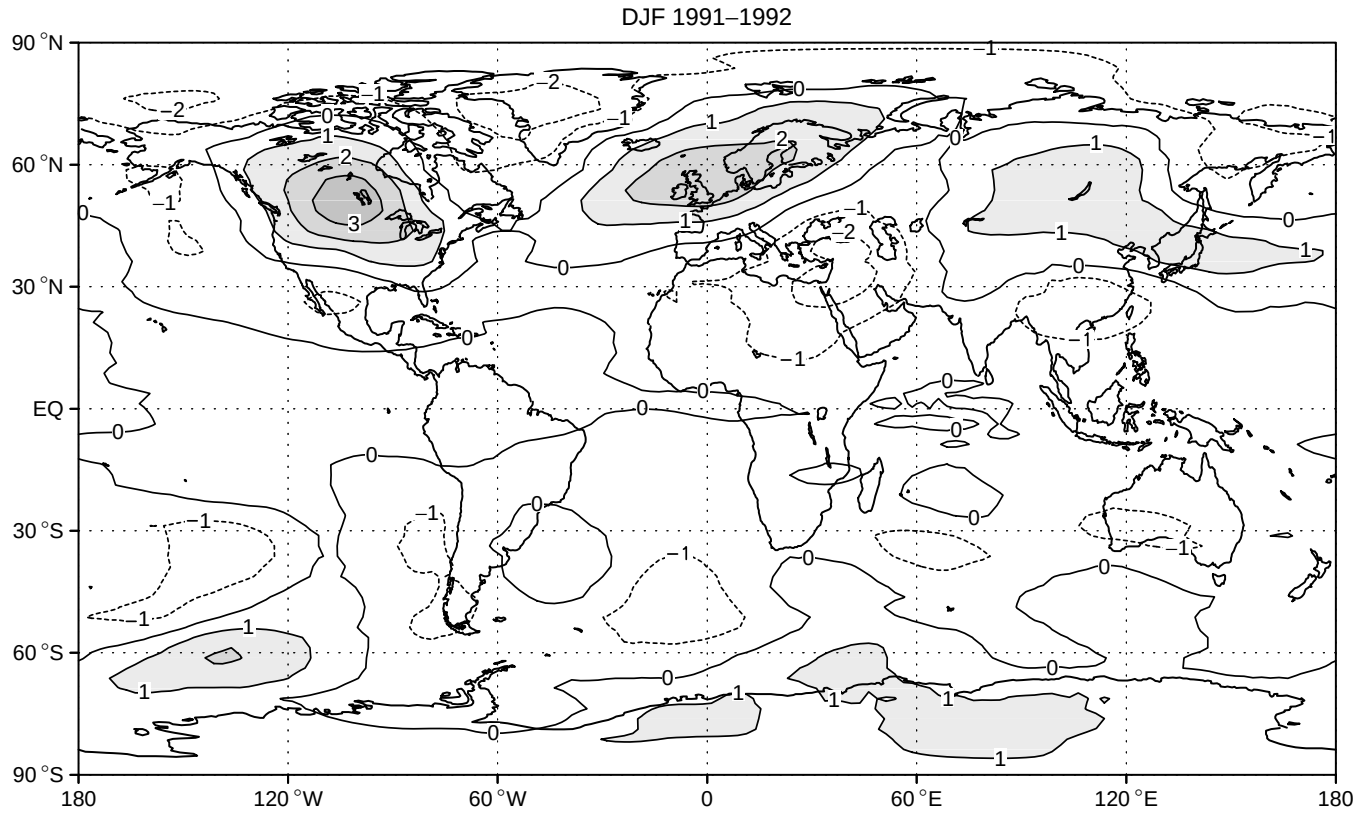


Figure 3 Winter (December January February, DJF) lower tropospheric temperature anomalies (K) (with the non-volcanic period of 1984–1990 used to calculate the mean) for the 1991–1992 NH winter (DJF) following the 1991 Mt. Pinatubo eruption. This pattern is typical of that following all large tropical eruptions, with warming over North America, Europe, and Siberia, and cooling over Alaska, Greenland, the Middle East, and China. (Data from Microwave Sounding Unit Channel 2LT, courtesy of J Christy, University of Alabama, Huntsville. Adapted from Plate 8, Robock, 2000, © Copyright American Geophysical Union.)

decadal-scale climate change remains unclear (*see* Natural Climate Variability, Volume 1). In particular, the current century is the warmest of the past five, with the early part of this period earning the moniker of the Little Ice Age due to its coldness (*see* Little Ice Age, Volume 1). Volcanic eruptions and solar variations are both likely to have been important contributors to the cooling of the Little Ice Age.

Volcanic aerosols have the potential to change not only the radiative flux in the stratosphere, but also its chemistry. The most important chemical changes in the stratosphere are related to ozone (O_3) which has significant effects on ultraviolet (UV) and longwave radiative fluxes. The reactions that produce and destroy O_3 depend on the UV flux, the temperature, and the presence of surfaces for heterogeneous reactions, all of which are changed by volcanic aerosols. The heterogeneous chemistry responsible for the ozone hole over Antarctica in October each year occurs on polar stratospheric clouds of water or nitric acid, which only occur in the extremely cold isolated spring vortex in the Southern Hemisphere (*see* Ozone Hole, Volume 1; Depletion of Stratospheric Ozone, Volume 1). These reactions make anthropogenic chlorine available for chemical destruction of O_3. Sulfate aerosols produced by volcanic eruptions can also provide these surfaces at lower latitudes and at all times of the year. In fact, after the 1991 Mt. Pinatubo eruption, a column O_3 reduction of about 5% was observed in midlatitudes, ranging from 2% in the tropics to 7% in the midlatitudes. Ozone depletion in the aerosol cloud is much larger and reaches 20%. The chemical ozone destruction is less effective in the tropics, but lifting of low ozone concentration layers with the aerosol cloud causes a rapid decrease in ozone mixing ratio in the low latitudes. Decrease of the ozone concentration reduces UV absorption in the stratosphere, which modifies the aerosol heating effect. The net effect of volcanic aerosols on the surface UV flux is to increase it because the aerosols backscatter less UV than the subsequent O_3 depletion allows through (*see* Ultraviolet Radiation, Volume 1).

The significant volcanic effect on O_3 chemistry is a new phenomenon, depending on anthropogenic chlorine in the stratosphere. While we have no observations, the 1963 Agung eruption probably did not deplete O_3, as

there was little anthropogenic chlorine in the stratosphere. Due to the Montreal Protocol and subsequent international agreements, chlorine concentration has peaked in the stratosphere and is now decreasing (*see* Stratosphere, Ozone Trends, Volume 1). Therefore, for the next few decades, large volcanic eruptions will have effects similar to Mt. Pinatubo, but after that, these O_3 effects will be significantly decreased and volcanic eruptions will have a stronger effect on atmospheric circulation without the negative feedback produced by O_3 depletion.

While the large 1982–1983 El Niño occurred just after the 1982 El Chichón eruption in Mexico, there is no evidence of a cause and effect relationship for this or any other eruptions. There is also no evidence that volcanic eruptions produce El Niño events. Because they can occur simultaneously, the climatic effects of El Niño and volcanic eruptions must be carefully separated to understand the climatic response to each.

Because volcanic eruptions and their subsequent climatic response represent a large perturbation to the climate system over a relatively short period, observations and the simulated model responses can serve as important analogs for understanding the climatic response to other perturbations. While the climatic response to explosive volcanic eruptions is a useful analog for some other climatic forcings, there are also limitations. The theory of nuclear winter, which focuses on the climatic effects of a massive injection of soot aerosols into the atmosphere from fires following a global nuclear holocaust, includes upward injection of the aerosols to the stratosphere, rapid global dispersal of stratospheric aerosols, heating of the stratosphere, and cooling at the surface under this cloud (*see* Nuclear Winter, Volume 3). Because testing this theory in the real world would involve global destruction, volcanic eruptions serve as analogues supporting various aspects of the theory.

While we have no idea when the next large tropical eruption will occur, our current understanding of the climatic impacts of volcanic eruptions suggest a likelihood of global cooling lasting about two years, and winter warming of the NH continents for one or two years.

See also: Stratosphere, Temperature and Circulation, Volume 1; DVI (Dust Veil Index), Volume 3; Lahars, Volume 3; VEI (Volcanic Explosivity Index), Volume 3; Volcanic Eruptions: Mt Merapi, Indonesia, Volume 3; Volcanoes and Cities, Volume 3; Volcanoes and the Environment, Volume 3.

FURTHER READING

Franklin, B (1784) Meteorological Imaginations and Conjectures, *Manchester Lit. Philos. Soc. Mem. Proc.*, 2, 122. [Reprinted in *Weatherwise*, 35, 262, 1982].

Genin, A, Lazar, B, and Brenner, S (1995) Vertical Mixing and Coral Death in the Red Sea Following the Eruption of Mount Pinatubo, *Nature*, 377, 507–510.

GRL Special Issue (1983) Studies of the 1982 El Chichón Eruption, *Geophys. Res. Lett.*, 10, 989–1060.

GRL Special Issue (1992) Studies of the 1991 Mt. Pinatubo Eruption, *Geophys. Res. Lett.*, 19, 149–218.

Robock, A (2000) Volcanic Eruptions and Climate, *Rev. Geophys.*, 38, 191–219.

Symons, G J, ed (1888) *The Eruption of Krakatoa, and Subsequent Phenomena*, Trubner, London.

Volcanic Explosivity Index (VEI)

see VEI (Volcanic Explosivity Index) (**Volume 3**)

Volcanoes and Cities

see Volcanoes and Cities (**Volume 3**)

Volcanoes and the Environment

see Volcanoes and the Environment (**Volume 3**)

Vostok Ice Core Record

see Climate Change, Abrupt (**Volume 1**)

Vostok, Subglacial Lake

J Cynan Ellis-Evans
British Antarctic Survey, Cambridge, UK

A vast lake, thought to be millions of years old, has been discovered beneath the Antarctic ice sheet. The deep sediments of Lake Vostok may contain a unique high-resolution record of climate change and continental-scale glaciation in Antarctica over the past 15- 30• million years. This extreme sub-glacial environment is also likely to contain novel microbiota, long isolated from the Earth's atmosphere and entirely dependent on geochemical energy. Lake Vostok

could feasibly provide a modern analogue for the Snowball Earth event some believe occurred during the neo-Proterozoic period (560- 700 million years ago) and which may have led to the subsequent enormous evolutionary radiation seen in the fossil record.

The presence of a water layer, due to geothermal heat production, beneath thick Antarctic ice sheets was predicted by Russian scientists almost 50 years ago and proven by airborne radio-echo studies in the vicinity of the Russian Vostok Station during the 1970s. By 1993, first evidence was emerging for the scale of water accumulation beneath the ice at Vostok and, in 1996 (Kapitsa *et al.*, 1996), it was confirmed that an enormous freshwater lake (over 280 km long and up to 1 km deep) was present beneath 4 km of glacial ice. Where the topographical ice sheet passes over the lake, surface features entirely flatten out and satellite radar altimetry (ERS-1 and RADARSAT) was used to map, at high resolution, the outline of the lake and its drainage basin (Figure 1). The lake has proved to be one of the largest freshwater bodies on the planet and represents a new extreme environment for humankind to explore. Subsequent coarse radio-echo surveys within a survey area that covers roughly half the Antarctic ice sheet have indicated in excess of 60 additional subglacial water bodies within this area. These additional lakes are significantly smaller than Lake Vostok (though many are around 20 km in length) and the majority are associated with ice sheet divides (start points for ice flow, equivalent to catchment boundaries) near Dome C and Ridge B in East Antarctica where the ice sheet is thickest (Figure 2).

Antarctica was separating from the other southern continents by 34 million years ago, and the resulting thermal isolation as the circum-polar Southern Ocean formed promoted development of the first ice sheets (Figure 3). Subglacial lakes formed in large rock depressions after the ice sheet had reached a thickness of at least 2 km. It is thought that the smaller subglacial lakes are perhaps a few million years old, but recent evidence increasingly points to a much older origin, and even to pre-glacial links, for Lake Vostok. With the Gamburtsev Mountains being the main point of glacial origin at the center of East Antarctica, the surrounding highland areas would have become ice covered at the outset of continent-wide glaciation and likely retained a thick ice sheet throughout the various glacial/interglacial events of the past 15–20 million years. A subglacial lake would have formed early and been maintained throughout this period in such a setting. The depth of sediment (up to 500 m) seen in Lake Vostok is also far greater than might be predicted from a few million years accumulation of glacial scour, which supports an earlier development.

Recent geophysical data have confirmed that the lake is associated with a major tectonic boundary, possibly still active, in the Vostok highlands, between the subglacial Gamburtsev Mountains to the west and the Aurora lowlands to the east (R Bell, personal communication). The current

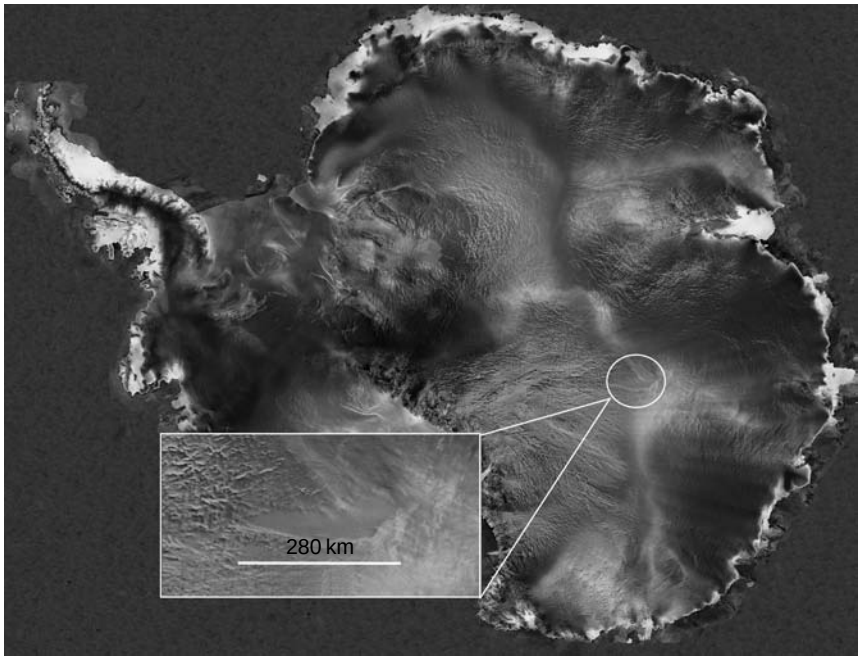


Figure 1 RADARSAT imagery, using radar altimetry data, clearly demonstrates the presence of Lake Vostok – seen as a very large area with an unusually flat surface on the East Antarctic ice sheet where the ice flows over the subglacial lake. (Image credits: NASA – Goddard Space Flight Center Scientific Visualization Studio)

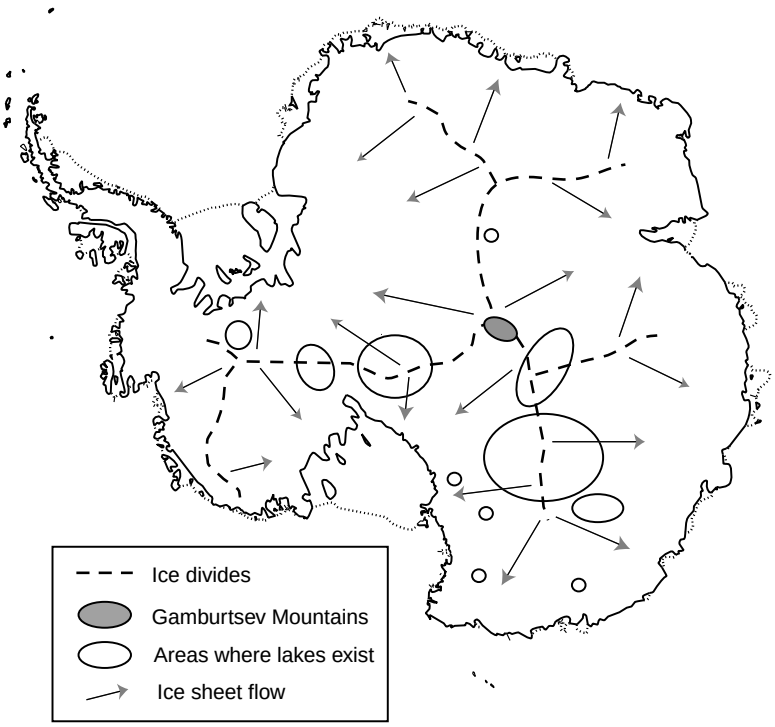


Figure 2 Dotted lines show the ice divides separating Antarctic ice catchments. Most of the subglacial lakes are located in close proximity to these ice divides where the greatest ice thickness occurs. (Modified from Barrett, 1999)

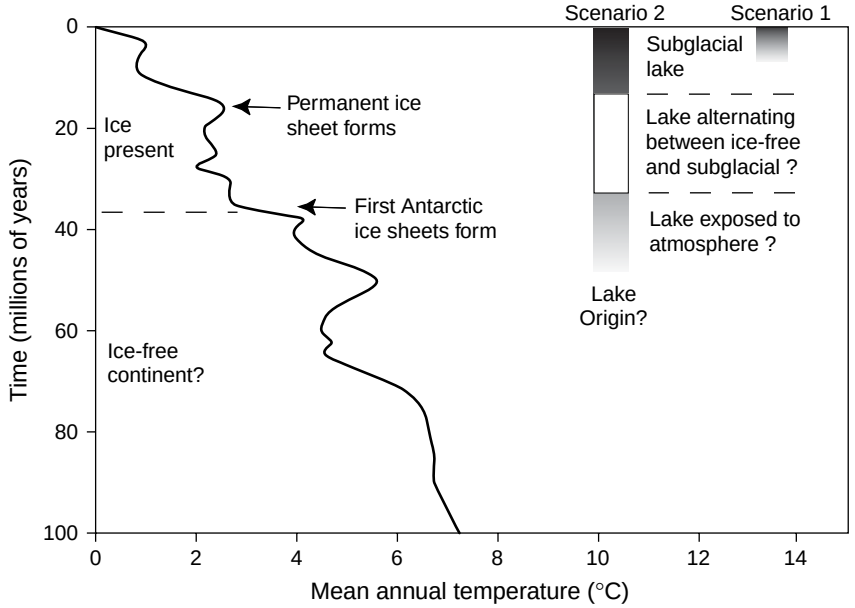


Figure 3 Time-temperature diagram illustrating the changes in mean annual air temperature in Antarctica over the past 100 million years and possible scenarios for the development of subglacial lakes over this time scale. (Modified from Barrett, 1999)

view is that Lake Vostok uniquely occupies an ancient rift valley basin. This raises the tantalizing possibility that a large lake may have occupied this basin before glaciation. A pre-glaciation Lake Vostok would likely have existed in a cold, temperate climate, similar to those now seen in southern Chile and Argentina, with diverse aquatic flora and fauna and a relatively barren catchment due to its alpine location. However, if the rift valley is comparatively

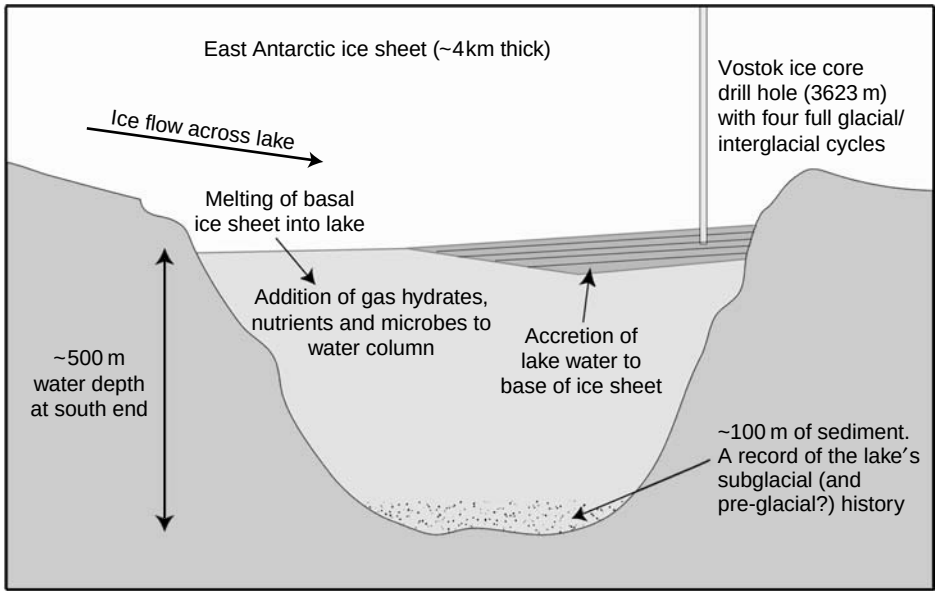


Figure 4 Vertical section through Lake Vostok (not to scale) illustrating salient features including the presence of areas of ice sheet melting, ice accretion and sediment accumulation

old, a lake could possibly have existed at that location for a substantial period extending into times of much warmer climate. Basal sediments in the current Lake Vostok would hopefully include sediments from any pre-glaciation lake. Lake Vostok would therefore appear to be unique, even amongst sub-glacial lakes, potentially containing a sediment record of the geological and climatic history of Antarctica extending throughout the period of glaciation and possibly further back into the pre-glacial Antarctic history. With this possibility, it is not surprising that the scientific community has become enormously interested in studying this particular lake.

The significance of Lake Vostok lies not only in its potential long sediment record but in the existing climate record serendipitously provided by the Vostok ice core which has drilled to within 150 m of the lake (Figure 4) and provided the longest ice climate record (four full glacial/interglacial cycles or 420 000 years) yet obtained. This exquisitely detailed record (Petit *et al.*, 1999) offers a means of cross-checking the responsiveness of a previously unexamined form of lake sediment record, though ice-covered surface lakes from polar regions are already known to record climate changes with great sensitivity.

The enormous catchment area associated with this lake constitutes a giant trap for terrestrial and extra-terrestrial material, which deposits on the ice sheet surface, migrates through the ice sheet and contributes to the sediments at the bottom of the lake. Examination of dust particle records in the ice sheet has demonstrated a previously unrecognized 100 000 year record that linked to climate changes over the past half million years. The lake sediments may not only contain this same record but also feasibly extend it through the millions of years of glaciation.

The ice record has also shown that viable microorganisms exist within the ice profile and could potentially seed the lake. The bottom of the Vostok ice core penetrated into ice that had been formed by freezing of lake water (accretion) to the base of the ice sheet. Within this accretion ice, recent studies have revealed viable microorganisms that in all probability came from the lake. Molecular fingerprinting (Priscu *et al.*, 1999) shows these bacteria to be identical to existing microorganisms, suggesting that they originated from aero biota that were carried through the ice sheet, deposited in the lake and subsequently trapped in accreted ice. However, the lake sediments may contain organisms derived from scoured bedrock and possibly from a pre-glaciation lake. Such organisms would potentially have had a longer period in which to adapt and evolve so could possibly prove to differ markedly from modern microbes. Information on such organisms could provide insights to the evolution of life on Earth and particularly as to how evolution might occur in a cold climate, whereas previously studies of the evolution of life have tended to be based on hot climate scenarios.

A recently presented and still controversial theory of evolution of the early Earth (the neo-Proterozoic period, 750–530 Ma) suggests that the Earth at that time was not simply cooling down from a hot start point, but instead went through cycles of hot climate (greenhouse) and cold climate (freezer) over periods of more than 10 million years. The *snowball Earth* theory is based on observations of extensive bands of carbonate cap deposits in ancient

rocks as the world s CO_2 became tied up during global glaciation and was subsequently released during greenhouse conditions (Hoffmann *et al.*, 1998). The switching between hot and cold climate has been proposed as a trigger for the enormous evolutionary radiations in biodiversity seen in the fossil record at the end of the Proterozoic. An old Lake Vostok would offer the only analogue possibility for a snowball Earth likely to be found on the planet, and would provide the opportunity to study microbial life at the limits of existence in a cold dark world.

To date very little is known about Lake Vostok and even less about other subglacial lakes, but the information already obtained hints at a potentially important role for Lake Vostok in deciphering the climate changes and evolutionary patterns of microbial life on Earth over the past tens of millions of years.

It would have been comparatively simple to have continued the Vostok ice core drilling project the last 150 m and penetrated to the lake. But the drilling operation uses a kerosene/antifreeze mixture as the drilling fluid, which would have risked contaminating this unique pristine environment, so drilling was stopped at a safe distance above Lake Vostok. Plans to penetrate to the lake in the latter half of this decade are now based on drilling new bore-holes with contamination-free technology currently under development. Initial entry will probably involve a non-return trip by robotic probes, which will set up monitoring observatories in the lake to characterize the environment s physical and chemical characteristics. A second drilling system will be deployed to obtain water and sediment samples for life detection studies and the final challenge will be to drill through the sediment profile and return a full sediment record to the surface without significantly disturbing the lake environment.

See also: Antarctica, Volume 1.

REFERENCES

Barrett, P (1999) Antarctic Climate History Over the Last 100 Million Years, in *Geological Records of Global and Planetary Changes*, eds P Barrett and G Orombelli, Terra Antarctica Reports, 3, 53–72.

Hoffman, P F, Kaufmann, A J, Halvorsen, G P, and Schrag, D P (1998) A Neoproterozoic Snowball Earth, *Science*, 281, 1342–1346.

Jean-Baptiste, P, Petit, J-R, Lipenkov, V Ya, Reynaud, D, and Barkov, N I (2001) Constraints on hydrothermal processes and water exchange in Lake Vostock from helium isotopes, *Nature*, 411, 460–462.

Kapitsa, A P, Ridley, J K, Robin, G, de Robin, Q, Siegert, M J, and Zotikov, I A (1996) A Large Deep Freshwater Lake Beneath the Ice of Central East Antarctica, *Nature*, 381, 684–686.

Petit, J R, Jouzel, J, Raynaud, D, Barkov, N I, Barnola, J-M, Basile, I, Benders, M, Chappellaz, J, Davis, M, Delaygue, G, Delmotte, M, Kotlyakov, V M, Legrand, M, Lipenkov, V Y, Lorius, C, Pepin, L, Ritz, C, Saltzmann, E, and Stievenard, M (1999) Climate and Atmospheric History of the Past 420 000 Years from the Vostok Ice Core, Antarctica, *Nature*, 399, 429–436.

Priscu, J C, Adams, E E, Lyons, W B, Voytek, M A, Mogk, D W, Brown, R L, McKay, C P, Takacs, C D, Welch, K A, Wolf, C F, Hirstein, J D, and Avci, R (1999) Geomicrobiology of Subglacial Ice Above Vostok Station, *Science*, 286, 2141–2144.

W

Walker Circulation

The atmospheric circulation near the equator varies strongly in longitude over scales of thousands of kilometers. A useful but perhaps oversimplified conceptual model of this variation is one of overturning cells in the longitude/height plane, along with associated patterns of surface temperature, sea level pressure and precipitation. The name given to this conceptual model is the *Walker Circulation* in honor of Sir Gilbert Walker, a Director-General of British observatories in India in the early part of the 20th century, who conducted much research directed at understanding and prediction of the Indian Monsoon. He also identified a number of relationships between seasonal climate variations in Asia and the Pacific region, notably the quasi-periodic variation known as the Southern Oscillation, which is associated with the El Niño.

The current widely used conceptual model and the term Walker Circulation both arise from the work of Jacob Bjerknes in the 1960s (*see* Bjerknes, Jacob, Volume 1). The Walker Circulation is comprised of rising motion in regions of high surface temperature (which in turn leads to low sea level pressure and high precipitation coming from deep convective clouds) and sinking motion in regions of low surface temperature (which in turn leads to high sea level pressure and low precipitation). The rising and sinking regions are connected by east/west motions in the upper and lower troposphere, forming the Walker Cells. The east/west motions near the surface are directed from regions of high to low surface pressure, and motions in the upper troposphere at the same longitude are in opposite directions to the surface flows, thereby closing the cells.

The Walker Circulation varies seasonally, but in the mean it has three important rising regions, located in the western Pacific near Indonesia, in equatorial Africa, and over equatorial South America. Following the Pacific cell as an example, the low-level, easterly trade winds form the base of the Walker Circulation, bringing warm moist air from over the Pacific Ocean towards the Indonesian region. Here, moving over normally very warm seas, intense convection takes place, and moist air rises to high levels of the atmosphere. The air then travels eastward before sinking over the eastern Pacific Ocean. The rising air is associated with a region of low air pressure, towering cumulonimbus clouds and rain. High pressure and dry conditions accompany the sinking air. Because this circulation takes place near the equator, the Coriolis force is absent or weak, and a direct zonal (i.e., east–west) circulation spanning the ocean is possible (*see* Atmospheric Motions, Volume 1).

Shifts in the location of the Indonesian rising region and the associated large scale changes in the Walker Circulation are a key feature in interannual climate variability associated with El Niño/Southern Oscillation, (*see* El Niño and La Niña: Causes and Global Consequences, Volume 1). The strength of the Walker Circulation over the Pacific is conventionally measured by a simple Southern Oscillation Index (SOI), which is essentially the difference in atmospheric pressure between Tahiti and Darwin. During El Niño years, the SOI becomes negative, the Walker Circulation is weakened or even reversed, with suppressed convection in the west and enhanced precipitation in the east over the warm waters characteristic of the El Niño. Positive values of the SOI are associated with stronger Pacific trade winds, warm sea surface temperatures in the western Pacific, and enhanced precipitation in Indonesia and eastern and northern Australia.

EDWIN K SCHNEIDER *USA*

Water Cycle

see Hydrologic Cycle (**Volume 1**)

Water Vapor: Distribution and Trends

Dian J Seidel
NOAA Air Resources Laboratory, Silver Spring, MD, USA

Water in the gaseous phase, water vapor, is the most signi cant atmospheric trace constituent vis-à-vis climate, weather, hydrology, and atmospheric chemistry. Water vapor is the most abundant atmospheric greenhouse gas, and without it the planet's surface temperature would be well below freezing. Phase changes involving water vapor - the condensation and evaporation processes - involve exchanges of latent heat energy that affect the vertical stability of the atmosphere, the structure and evolution of storm systems, and the energy balance of the global climate system (Chahine, 1992). Water vapor is also the prime source of the atmospheric hydroxyl radical, an oxidizing agent that cleanses the atmosphere of many air pollutants.

The total amount of water vapor in the atmosphere is about 13 • 10^{15} kg. For reference, oceans hold about 100 000 times as much water, and the total amount of water on land (as ice, snow, surface and ground water, and in living organisms) is about 4000 times as much as in the atmosphere. Nevertheless, changes in the amount and distribution of atmospheric water vapor, which can happen relatively quickly, both cause and provide evidence of other global environmental changes. For example, temperature and water vapor changes often go hand in hand, so that increases in water vapor can be indirect evidence of warming. As changes in climate are accompanied by changes in atmospheric circulation and the transport of water vapor, the patterns of precipitation will also change. The extent of such changes could be similar in magnitude to the shifts associated with the El Niño- Southern Oscillation phenomenon, which leads to droughts and ooding in various locations around the world.

The weight of the atmosphere s water vapor contributes to only about one-quarter of 1% of the total sea level pressure of all gases. If all the water vapor in the air were to condense and fall as rain, it would cover the Earth with a layer about 25 mm (or 1 inch) deep on average. This is called precipitable water. Since water vapor is not evenly distributed over the globe, there would be about 50 mm near the equator and less than one-tenth as much near the poles, as shown in Figure 1 (Randel *et al.*, 1996). The general decrease of precipitable water from equator to poles is a reflection of the global distribution of temperature, because warm air is capable of holding more moisture than cold air. There are exceptions in the major desert regions, where the air is very dry despite its high temperature. The most humid region is in the western equatorial Pacific, above the so-called oceanic warm pool, where the highest sea surface temperatures are found.

The average precipitation over the globe is about 1 m annually, so there must be swift transfer of atmospheric water vapor to the surface via precipitation and back again to the atmosphere via evaporation; the average water molecule spends about 9 days in the air before returning to the surface. This rapid turnover, combined with the variation of temperature with height and geography, causes water vapor to be distributed unevenly in the atmosphere, both horizontally and vertically. The map in Figure 1 is a 10-year average depiction that is smoothly varying. On any given day, water vapor exhibits much more complex patterns associated with storm systems and wind patterns.

The uneven distribution of water vapor is even more pronounced in the vertical dimension. Water vapor concentrations (expressed as the ratio of the volume occupied by water vapor to the total volume of air, most of which is nitrogen and oxygen) decrease rapidly with height, varying over four orders of magnitude, from a few percent near the surface to a few parts per million in the lower stratosphere. Nearly half the total water in the air resides below an altitude of about 1.5 km. Less than 5% is in the upper troposphere (above 5 km), and less than 1% is in the stratosphere, nominally above 12 km. This wide range of concentrations presents challenges in designing instruments for atmospheric water vapor measurements. Therefore, a plethora of technologies exist, from low-cost, expendable balloon-borne sensors designed for daily weather observations, to research-quality aircraft- and satellite-borne instruments.

Variations in atmospheric water vapor occur on timescales from milliseconds to decades. The fastest fluctuations are due to turbulence, and near the surface they

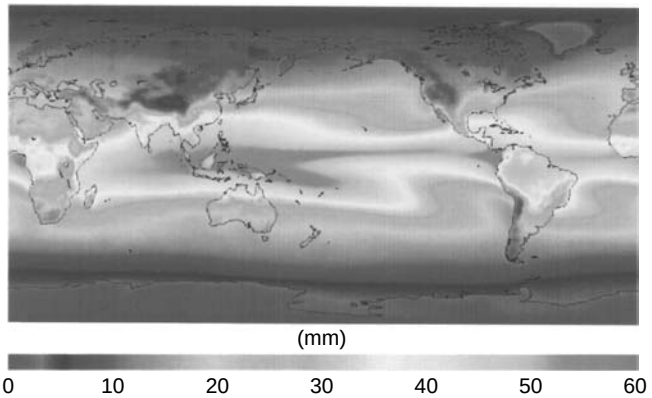


Figure 1 The global distribution of total atmospheric water vapor (precipitable water) above the Earth's surface. This depiction includes data from both satellite and weather balloon observations and represents an average for the period 1988–1997

are associated with changes in the vertical component of the wind and in evaporation. At higher altitudes, air currents generated by atmospheric instability redistribute water vapor and influence cloud formation over periods of minutes or hours. Land and sea breezes cause water vapor concentrations in coastal regions to vary considerably over the course of a day. The distribution of water vapor changes with seasonal changes in temperature and atmospheric circulation patterns. As with temperature, seasonal changes in water vapor appear to be stronger in the Northern Hemisphere than in the Southern Hemisphere. Since the Northern Hemisphere has more land, with lower heat capacity than the ocean, it shows faster temperature and water vapor changes.

On longer time scales, water vapor changes are thought to contribute to an important positive feedback mechanism for climate change, as follows. Warming of the surface, particularly the sea surface, leads to enhanced evaporation. Since warmer air requires more moisture to reach saturation, atmospheric temperature increases allow for increases in atmospheric water vapor. Due to the fact that water vapor is a greenhouse gas, enhanced water vapor in the lower troposphere results in further warming, allowing a higher water vapor concentration, thereby creating a positive feedback. This mechanism is well represented in global climate models, and observations of year to year variations in temperature and water vapor show high correlations (Wentz and Schabel, 2000). However, it is less clear whether this mechanism operates in the upper troposphere, where, despite its relatively low concentrations, water vapor has a more powerful greenhouse effect (Harries, 1996). The variability of upper-tropospheric water vapor with surface temperature is not well documented, mainly because of the lack of long-term measurements in this region (Rind, 1998). Mechanisms have been proposed by which upper-tropospheric water vapor changes, particularly in the tropics, may lead to negative feedbacks on surface temperature (Lindzen, 1990), but overall water vapor changes are thought to contribute to a positive feedback to global warming (Stocker, 2001).

With global warming, an increase in the rate of evaporation must be matched by an increase in the rate of precipitation, because the atmosphere holds so little water. This acceleration of the global hydrological cycle seems likely to lead to an increase in the frequency of extreme precipitation events, and an increase in the proportion of precipitation falling in extreme events (Trenberth, 1998, 1999), which has been observed in some regions (Folland and Karl, 2001). Such changes would have important consequences for agriculture, flood control, and other aspects of human society. Water vapor feedback may also be amplifying atmospheric temperature changes, suspected to be the primary driver of the observed rapid retreat, and in some cases the disappearance, of ice caps and glaciers at high elevations in the tropics and subtropics (Thompson, 2000). Evidence for long-term changes in atmospheric water vapor is limited by the availability and quality of measurements. However, data for the last three or four decades of the twentieth century indicate an increase in water vapor in the lower troposphere over parts of the Northern Hemisphere (Ross and Elliott, 2000).

Measurements are so limited in the stratosphere that detecting changes in stratospheric water vapor is very difficult. Although the spatial representativeness of the observations remains a question, observations made over Boulder, Colorado, in the 1980s and 1990s show an upward trend in water vapor concentration (Oltmans *et al.*, 1995). Unlike the troposphere, the stratosphere is not well mixed, and stratospheric water vapor is not controlled by the same processes operating in the troposphere. The observed stratospheric water vapor increases are larger than can be attributed to increases in methane, which is a chemical source of stratospheric water vapor, and do not appear compatible with the current understanding of changes in the temperature of the tropical tropopause, where tropospheric water vapor enters the stratosphere. Circulation changes are hypothesized to be a contributing factor. Changes in stratospheric water vapor, and related changes in stratospheric clouds, could alter the chemistry of the stratospheric ozone layer and contribute to a cooling of the stratosphere.

Thus, water vapor, a natural and familiar atmospheric constituent, plays many important roles in the global environment. In the greenhouse effect, the global water cycle, and global air chemistry, water vapor both brings about, and is a manifestation of, global environmental change.

See also: Hydrologic Cycle, Volume 1; The Global Temperature Record, Volume 1.

ACKNOWLEDGMENTS

This article is based, in part, on material in the American Geophysical Union Special Report (1995) *Water Vapor in the Climate System*, AGU, 2000 Florida Avenue, NW, Washington, DC, 20 009, 15.

REFERENCES

Chahine, M T (1992) The Hydrological Cycle and its Influence on Climate, *Nature*, 359, 373–380.

Harries, J E (1996) The Greenhouse Earth: a View from Space, *Q. J. R. Meteorol. Soc.*, 122, 799–818.

Folland, C K and Karl, T R (2001) Observed Climate Variability and Change, *IPCC WGI Third Assessment Report*, Chapter 2, in press.

Lindzen, R S (1990) Some Coolness Concerning Global Warming, *Bull. Am. Meteorol. Soc.*, 71, 288–299.

Oltmans, S J, Vomel, H, Hofmann, D J, Rosenlof, K H, and Kley, D (2000) The Increase in Stratospheric Water Vapor from Balloonborne, Frostpoint Hygrometer Measurements at

Washington, DC, and Boulder, Colorado, *Geophys. Res. Lett.*, 27, 3453–3457.
Randel, D L, Vonder Haar, T H, Ringerud, M A, Stephens, G L, Greenwald, and Combs, C L (1996) A New Global Water Vapor Dataset, *Bull. Am. Meteorol. Soc.*, 77, 1233–1246.
Rind, D (1998) Just Add Water Vapor, *Science*, 281, 1152–1153.
Ross, R J and Elliott, W P (2000) Northern Hemisphere Tropospheric Water Vapor Trends, *J. Clim.*, in press.
Stocker, T F (2001) Physical Climate Processes and Feedbacks, *IPCC WGI Third Assessment Report*, Chapter 7, in press.
Trenberth, K E (1998) Atmospheric Moisture Residence Times and Cycling: Implications for Rainfall Rates with Climate Change, *Clim. Change*, 39, 667–694.
Trenberth, K E (1999) Conceptual Framework for Changes of Extremes of the Hydrological Cycle with Climate Change, *Clim. Change*, 42, 327–339.
Thompson, L G (2000) Ice Core Evidence for Climate Change in the Tropics: Implications for our Future, *Q. Sci. Rev.*, 19, 19–34.
Wentz, F J and Schabel, M (2000) Precise Climate Monitoring Using Complementary Satellite Data Sets, *Nature*, 403, 414–416.

WCC (World Climate Conferences – 1979 and 1990)

Two international conferences, convened a decade apart by the United Nations World Meteorological Organization (WMO) in cooperation with a number of other international bodies, played a major role in rallying world support for international efforts in climate research and services.

The First World Climate Conference (FWCC), held in Geneva in 1979, was motivated by a number of growing concerns. Drought in Africa and fisheries collapse in South America had seriously disrupted international trade in food, causing significant price increases. At the same time, international conferences had called attention to the threat of human-induced climate change. With the availability of improved observing, data processing, and dissemination systems, the National Meteorological Services saw opportunities for providing greatly enhanced support. Finally, the Global Atmospheric Research Programme was drawing to a close, and ambitious plans were being developed to continue its research efforts relating to the climate system.

This convergence of motivations prompted the organization of the FWCC under the leadership of Dr Robert M White (US) as Conference Chairman and Sir David Arthur Davies, Secretary-General of the WMO. The week-long conference brought together hundreds of scientists and senior representatives from most of the world s nations and international bodies. The conference s recommendations formed the basis for establishment of the World Climate Programme (WCP), and firmly placed climate issues on the international agenda.

The Second World Climate Conference (SWCC), held in 1990, again in Geneva, assembled an expanded group of international representatives, including many ministers and several heads of state. It reviewed the very considerable progress of the WCP, strongly reaffirmed concerns for human influences on global climate, and made a number of important recommendations for the development of the WCP, notably in strengthening systems of observing and monitoring. In accord with the recommendations of the SWCC, the WCP now includes the four sub-programs dealing respectively with data and monitoring, applications and services, research, and impact assessment and response strategies.

JOHN S PERRY *USA*

WCP (World Climate Programme)

The WCP is a cooperative international scientific program whose goals are to improve understanding of the climate system and to apply that understanding for the benefit of societies coping with climate variability and change. The WCP was established following the staging of the First World Climate Conference in Geneva, Switzerland in February 1979. The major sponsoring organizations of the WCP at the international level are the World Meteorological Organization (WMO), the United Nations Environment Programme (UNEP), the Intergovernmental Oceanographic Commission of United Nations Educational, Scientific and Cultural Organization (UNESCO) and the International Council for Science (ICSU). The four components of WCP and their respective primary objectives are:

- World Climate Data and Monitoring Programme (WCDMP): to facilitate the effective collection and management of climate data and the monitoring of the global climate system, including the detection and assessment of climate variability and changes.
- World Climate Applications and Services Programme (WCASP): to foster the effective application of climate knowledge and information for the benefit of society and the provision of climate services, including the prediction of significant climate variations, both natural and as a result of human activity.

- World Climate Impact Assessment and Response Strategies Programme (WCIRP): to assess the impacts of climate variability and changes that could markedly affect economic or social activities. To advise governments thereon, and contribute to the development of a range of socioeconomic response strategies that could be used by governments and the community.
- World Climate Research Programme (WCRP): to improve the understanding of climate processes for determining the predictability of climate, including its variability and change, identifying the extent of human influence on climate and developing the capability for climate prediction (*see* GCOS (Global Climate Observing System), Volume 1).

The WCP and the Global Climate Observing System (GCOS) are the major international programs contributing to the implementation of Agenda 21.

MICHAEL COUGHLAN *Switzerland*

WCRP (World Climate Research Programme)

The WCRP was established in 1980, under the joint sponsorship of the International Council for Science (ICSU) and the World Meteorological Organization (WMO), and has also been sponsored by the Intergovernmental Oceanographic Commission (IOC) of the United Nations Educational, Scientific and Cultural Organisation (UNESCO) since 1993. The objectives of the program are to develop the fundamental scientific understanding of the physical climate system and climate processes needed to determine to what extent climate can be predicted and the extent of the influence of human activities on climate. The program encompasses studies of the global atmosphere, oceans, sea and land ice, and the land surface, which together constitute the Earth s physical climate system. WCRP studies are specifically directed to provide scientifically founded quantitative answers to the questions being raised on climate and the range of natural climate variability, as well as to establish the basis for predictions of global and regional climatic variations and of changes in the frequency and severity of extreme events.

WCRP activities thus forcefully address outstanding issues of scientific uncertainty in the Earth s climate system, including transport and storage of heat by the ocean, the global energy and hydrological cycle, the formation of clouds and their effects on radiative transfer, and the role of the cryosphere in climate. These activities match the scientific priorities identified by the Intergovernmental Panel on Climate Change (IPCC), and provide the basis for responding to issues raised in the United Nations Framework Convention on Climate Change (UNFCCC) (*see* Intergovernmental Panel on Climate Change (IPCC): an Historical Review, Volume 4). WCRP also lays the scientific foundation for meeting the research challenges posed in Agenda 21 (*see* Agenda 21, Volume 4). Together with the International Geosphere–Biosphere Programme (IGBP) and the International Human Dimensions of Global Environmental Change Programme (IHDP), WCRP provides the international framework for scientific cooperation in the study of global change (*see* IGBP (International Geosphere- Biosphere Programme), Volume 2; IHDP (International Human Dimensions Programme on Global Change), Volume 4).

In addition, WCRP organizes and/or sponsors *ad hoc* meetings to consider scientific and planning issues, and participates as appropriate in the planning of the ICSU/WMO/IOC/UNEP Global Observing System (GCOS). In contributing to the planning and oversight of GCOS and the Global Ocean Observing System (GOOS), WCRP cosponsors the Ocean Observation Panel for Climate (OOPC), which oversees the implementation of an ocean observation system for GCOS (or, equivalently, the climate component for GOOS). With GCOS, WCRP also oversees the Atmospheric Observation Panel for Climate (AOPC) and collaborates with IGBP and IHDP in the development of the System for Analysis, Research and Training (START).

Scientific guidance for the program is provided by a Joint Scientific Committee (JSC), which consists of 18 scientists selected by mutual agreement between the three sponsoring organizations. The scientists are selected so as to represent climate-related disciplines in atmospheric, oceanic, hydrological and polar sciences. Implementation of the program takes place through the Joint Planning Staff (JPS) in Geneva and the International Project Offices (IPOs).

WCRP has formulated a broad-based inter-disciplinary science strategy offering the widest possible scope for investigation of all physical aspects of climate and climate change. This strategy is reflected in the ongoing WCRP core projects. All projects are led by scientific steering or working groups, normally meeting once a year.

The Arctic Climate System Study (ACSYS), concentrating on the understanding of Arctic Ocean variability and change including sea ice processes, is being expanded into the Climate and Cryosphere Initiative (CLIC), which will investigate the role of the entire cryosphere for global climate and as an early indicator of change.

Climate Variability and Predictability (CLIVAR) is the main focus in WCRP for studies of climate variability, extending effective predictions of climate variation and refining the estimates of anthropogenic climate change.

CLIVAR is particularly attempting to exploit the memory in the slowly changing oceans and to increase understanding of the coupled behavior of the rapidly changing atmosphere and slowly varying land surface, oceans and ice masses as they respond to natural processes, human influences and changes in the Earth s chemistry and biota. CLIVAR will advance the findings of the successfully completed Tropical Ocean and Global Atmosphere (TOGA) project, and will expand on work now underway in WCRP s World Ocean Circulation Experiment (WOCE).

The Global Energy and Water Cycle Experiment (GEWEX) is the scientific focus within the WCRP for studies of atmospheric and thermodynamic processes that determine the global hydrological cycle and water budget and their adjustment to global changes such as the increase in greenhouse gases. One of the main thrusts of GEWEX is the implementation of a series of atmospheric/hydrological regional process studies, including the GEWEX Continental-scale International Project (GCIP) that embraces the whole Mississippi River basin, the GEWEX Asian Monsoon Experiment (GAME), and the Baltic Sea Experiment (BALTEX). Observational projects undertaken in order to satisfy specific and evolving scientific needs include the International Satellite Cloud Climatology Project (ISCCP), the International Satellite Land-Surface Climatology Project (ISLSCP), the Global Water Vapor Project (GVaP), and the Global Precipitation Climatology Project (GPCP). Progress has also been made in the GEWEX Cloud System Study (GCSS), which aims to develop improved parameterizations and models of cloud systems used in climate and numerical weather prediction studies.

Another WCRP research activity playing an important role in better understanding the climate system is the Stratospheric Processes and their Role in Climate (SPARC) study, which investigates the interaction of dynamical, radiative and chemical processes. Activities organized by SPARC include the construction of a stratospheric reference climatology and the improvement of understanding of trends in temperature, ozone and water vapor in the stratosphere. Gravity wave processes, their role in stratospheric dynamics and how they may be parameterized in models are also being investigated.

The WOCE is a fundamental element of the WCRP scientific strategy to understand and predict changes in the world ocean circulation, volume and heat storage that would result from changes in atmospheric climate and net radiation. WOCE is conducted by means of a combination of *in situ* oceanographic measurements, observations from space, and global ocean modeling.

The development of global climate models is an important unifying component of WCRP, building on scientific and technical advances in the more discipline oriented activities. These models are the fundamental tools for understanding and predicting natural climate variations and providing reliable estimates of anthropogenic climate change. Models also provide the essential means of exploiting and synthesizing in a synergistic manner all relevant atmospheric, oceanographic, cryospheric and land surface data collected in WCRP and other programs. The Working Group on Numerical Experimentation (WGNE), jointly sponsored by the JSC of WCRP and the WMO Commission for Atmospheric Sciences (CAS), leads the work on the development of atmospheric circulation models for both climate studies and numerical weather prediction. Each project has a numerical experimentation group, that supports the respective WCRP components.

In addition, the Working Group on Coupled Modelling (WGCM), jointly sponsored by JSC and CLIVAR, leads the development of coupled ocean/atmosphere/land models that are used for climate studies on longer time-scales. WGCM is also WCRP s link to the Earth system modeling in IGBP s Global Analysis Integration and Modelling (GAIM) and to the IPCC. Activities in this area concentrate on the identification of errors in model climate simulations and exploring the means for their reduction by organizing coordinated model experiments under standard conditions. Under the auspices of WCRP, the Atmospheric Model Intercomparison Project (AMIP) has been set up in which the 10-year period 1979–1988 has been simulated by 30 different atmospheric models under specified conditions. The comparison of the results with observations has shown the capability of many models to adequately represent mean seasonal states and large-scale interannual variability for several basic climate parameters. A similar Coupled Model Intercomparison Project (CMIP) is in progress.

Further information on many of these research programs are covered in separate articles in this volume. In addition, information may be obtained from the WCRP Joint Planning Staff, c/o World Meteorological Organization, 7 bis, Avenue de la Paix, Case Postale 2300, 1211 Geneva 2, Switzerland, email: dwcrp@gateway.wmo.ch.

JOHN S PERRY *USA*

Weather

Weather is the state of the atmosphere as perceived by a person: hot or cold, humid or dry, rain or snow, windy or calm, clear or cloudy, etc. Each of these weather elements has quantitative measures, e.g., temperature, humidity, precipitation amount, wind speed and direction, and cloud cover. *Climate* was once defined as the average of the various weather parameters. Climate was viewed as a static

measure of long-term conditions. In the past decade, however, the extension of the forecast period of long-range weather forecasting and the improved understanding of seasonal to interannual climate fluctuations has blurred the distinction between weather and climate. Climate may be viewed as what one would expect, whereas weather is what happens.

Weather prediction is made possible by observing the various atmospheric parameters, analyzing the resulting data by assimilating them into a comprehensive picture of what is occurring at a specific time, and using known relationships or models to project the changes in the initial parameters. These resulting values then must be related to the conditions that will be felt by human senses, i.e., weather. The weather observations may be direct, *in situ* measures of atmospheric parameters, such as in the case of thermometers and anemometers, or remotely sensed, and indirect measures, such as those taken by weather radars or environmental satellites. The challenge to the weather forecaster is to have sufficient information of the required accuracy to be able to specify the initial state of the atmosphere. The data assimilation process is crucial in this regard; it must blend varied types of data with varied degrees of accuracy and varied frequencies of observation. The challenge of the weather forecasting models is to be able to adequately represent the resolution of the weather events, as well as to represent the various physical, chemical, and dynamic processes that influence the future evolution of the state of the atmospheric system.

ELBERT W (JOE) FRIDAY, JR *USA*

Weathering

Weathering is the major force operating on the Earth s surface in opposition to mountain building. Weathering processes are the main agents of decay and disintegration of rocks at the surface of the Earth. Weathering processes include the effects of gravity but do not include significant sediment transport (such as occurs in rivers and streams, glaciers and aeolian environments). Weathering is caused by both physical and chemical processes and is commonly biologically mediated. The physical processes of weathering cause rocks to disintegrate via mechanical mechanisms. Mechanical disintegration can be caused by the freeze/thaw of water in cracks, roots expanding into cracks during plant growth, varying expansion rates of the component minerals during diurnal heating and cooling cycles (insolation), and salt precipitation and hydration in cracks. In contrast chemical processes cause rocks to decay by dissolution and re-precipitation processes. Chemical weathering is controlled by the composition of the parent material, climate, presence of water, rate of transport of solutes away from the site of weathering, redox potential (i.e., whether the pore waters are either relatively reducing or oxidizing; *see* Redox Potential, Volume 2), acidity of the ground water and kinetics of the reactions between the different minerals present and the ground water. Chemical weathering is particularly intense in soils where the combination of elevated (relative to the atmosphere) carbon dioxide (CO_2) concentrations resulting from microbial decay processes and acid ground waters cause minerals, which were stable in igneous settings, to be unstable in the weathering environment. Individual minerals are variously susceptible to weathering, with the common igneous minerals forming a weathering series (in which minerals such as olivine are much more susceptible to degradation than feldspar and quartz). During weathering some minerals (e.g., calcite) dissolve completely (congruent reactions), which liberates their component elements to the pore waters, while others partially dissolve (e.g., forsterite), which liberates material to the pore waters and leaves solid residues (incongruent reactions). In soils, the dissolved components commonly reprecipitate and form new mineral assemblages including clays, carbonate minerals, soil sesquioxides [aluminum (Al), iron (Fe), and manganese (Mn) oxides]. The chemical weathering of silicates and subsequent precipitation of carbonate minerals remove significant carbon from the atmosphere and is a major sink of carbon in the global carbon cycle. Weathering exerts a long-term (10^5–10^6 years) control on global warming.

See also: Geological Cycling, Volume 1; Soil Mineralization, Volume 2.

JOE MACQUAKER *UK*

Wegener, Alfred

(1880- 1930)

Wegener was born in Berlin on 1 November, 1880. From 1899 he studied astronomy and meteorology at Heidelberg, Innsbruck and Berlin. Whilst at university he also became very interested in geology. He received his doctoral degree

on 24 November 1904 for his thesis *Die Alfonsischen Tafeln für den Gebrauch des modernen Rechners* supervised by Wilhelm Foerster and Julius Bauschinger. From 1905/1906 he worked at the Aeronautical Observatory in Lindenberg (near Berlin) and participated in the Danish expedition for exploration of the northeastern coast of Greenland. In 1917 he became a professor of meteorology and practical astronomy at the University of Marburg and in 1924 he was appointed professor of meteorology and geophysics at the University of Graz.

Wegener s main work dealt with the physics of the upper atmosphere, especially the aurora, the so-called geo-corona, the nature of noctilucent clouds and the link between cyclones and Helmholtz s air billows. Another interest was in atmospheric optics in which he studied various aspects of halos. The main results of this work were published in his book on the physics of the atmosphere (1911).

In the 1920s, Wegener worked on the results of his exploration of Greenland, and he made plans to mount a new expedition to examine the inner ice shelf of Greenland. In 1929 he departed on a preparatory expedition to Greenland but died there from an illness on 1 April 1930.

The other major aspect of Wegener s work was his contribution to the theory of continental drift (*see* Continental Drift, Volume 1). The possibility of such drift had been proposed by H Wettstein as early as 1888. Wegener, in his well-known textbook of 1915, gave reconstructions of the distribution of the continents at various earlier geological epochs such as the Carboniferous (beginning about 362 million years ago) and Eocene (a geological epoch, subdivision of the stratigraphic column), supported by geological and other evidence. He envisaged continental drift as a flight from the Poles (Polflucht) due to gravitational attraction exerted by the Earth s equatorial bulge. This force, although real, is much too feeble and his theory was rejected by many scientists of the time for lack of a viable geophysical mechanism. Wegener s concept of continental drift has, however, been accepted since 1950 as a result of conclusive evidence from the new field of palaeomagnetism, and the hypothesis of convection in the Earth s mantle, which was made possible by solid-state creep. (It is of interest that Wegener, in 1929, discussed the role of convective streaming in the magmatical core as a possible source of drift). In recent years all this has been integrated neatly into a comprehensive theory of plate tectonics (*see* Plate Tectonics, Volume 1) and seafloor spreading, which explains much of the dynamic behaviour of the Earth s surface features.

REFERENCES

Ertel, H (1935) Die Berechnung der Polfluchtkraft, Gerlands Beitr, *Geophysik*, 43, 327–330.

Schroder, W (1999) Wegener s Work Included Studies of Noctilucent Clouds, Aurora, *EOS*, 80, 357.

Wegener, A (1915) *Die Entstehung der Kontinente und Ozeane*, Vieweg, Braunschweig.

Wegener, A (1911) *Thermodynamik der Atmosphäre*, Bahrt, Leipzig.

Wettstein, H (1880) *Die Strömungen des Festen, Flüssigen und Gasförmigen und ihre Bedeutung für Geologie, Astronomie, Klimatologie und Meteorologie*, Wurster & Cie, Zurich.

WILFRIED SCHRODER *Germany*

WGNE (Working Group on Numerical Experimentation)

The WGNE is an international group of experts concerned with the mathematical and physical aspects of atmospheric modeling for both weather prediction and climate simulation. The WGNE is jointly sponsored by the Joint Scientific Committee for the World Climate Research Programme (*see* WCRP (World Climate Research Programme), Volume 1) and the Committee on Atmospheric Sciences of the World Meteorological Organization. Originally established in the 1970s to support the Global Atmospheric Research Program (*see* GARP (Global Atmospheric Research Program), Volume 1), the WGNE has examined a wide variety of questions related to the formulation and solution of numerical atmospheric models, as well as questions related to the performance of the atmospheric components of climate models. The WGNE publishes an annual survey of international activities in atmospheric and oceanic modeling, and is the international host of the Atmospheric Model Intercomparison Project (AMIP) (*see* AMIP (Atmospheric Model Intercomparison Project), Volume 1).

W LAWRENCE GATES *USA*

Wind Chill

Wind chill is the effective temperature experienced by a person under a given combination of air temperature and wind speed. Wind chill is different than temperature in that it attempts to consider the additional cooling effect of wind

on the human body. Moving air can carry away heat from a person at a rate that increases with the wind speed, thus causing the person to perceive a lower temperature than actually exists. A person exposed to cold temperatures will be in more danger from hypothermia or frostbite as the wind speed increases.

A chart of wind chill factors may be found at *http://www.nws.noaa.gov/er/iln/tables.htm#wind chill*

The present wind chill values were determined in 1939 by experiments conducted in Antarctica. Containers of water were placed in sub-freezing temperatures and varying wind speeds. The time required for the water to freeze was recorded as a function of temperature and wind speed. The results were used to determine the empirical relationship that defines wind chill.

In the last few years, the use of the wind chill as a measure of the impact of the weather conditions on human health has come under criticism. Some scientists consider the index to be too simplistic because actual cooling is a function of other meteorological factors, such as humidity and solar radiation.

ELBERT W (JOE) FRIDAY, JR *USA*

WMO (World Meteorological Organization)

The WMO is a specialized agency of the United Nations (UN) system, whose purposes are to facilitate international cooperation in the establishment of networks of stations for making meteorological, hydrological and other observations; and to promote the rapid exchange of meteorological information, the standardization of meteorological observations and the uniform publication of observations and statistics. It also furthers the application of meteorology to aviation, shipping, water problems, agriculture and other human activities, promotes operational hydrology and encourages research and training in meteorology.

The WMO was created in 1947 by the World Meteorological Convention adopted at the Twelfth Conference of Directors of the International Meteorological Organization (IMO), which had carried out similar functions before World War II. It commenced operations in 1951 in Geneva with the late Sir David Arthur Davies (UK) as its first secretary general. There are at present 185 Members, each of which maintains its own Meteorological and Hydrological Service.

The WMO s supreme body, the World Meteorological Congress, meets every four years to determine policies, approve the program and budget, and adopt regulations. The executive council meets at least every year to provide timely review and guidance. Members of the WMO are grouped in six regional associations (Africa, Asia, South America, North and Central America, South-West Pacific and Europe), which meet regularly. Eight technical commissions are responsible for: aeronautical meteorology; agricultural meteorology; atmospheric sciences; basic systems; climatology; hydrology; instruments and methods of observation; and marine meteorology.

The WMO secretariat, numbering just under 250 international civil servants, is headed by the Secretary General, currently Professor G O P Obasi (Nigeria). Housed in an innovative modern building near other international bodies in Geneva, it serves as the administrative, documentation and information center of the organization.

The WMO has long played a central role in international environmental research and protection efforts, actively collaborating with other intergovernmental and non-governmental organizations. An early initiative in which WMO played an important role was the International Geophysical Year (1957–1958), which laid the groundwork for current WMO programs in ozone and air pollution monitoring. The WMO also contributed significantly to preparations for the 1972 UN Conference on the Human Environment in Stockholm, Sweden, which led to the establishment of the United Nations Environment Programme (UNEP). In 1975, WMO issued the first scientific statement on the modifications of the stratospheric ozone layer, which led to the formulation in 1977 of the first International Plan of Action for the Protection of the Ozone Layer. Subsequent assessments by WMO provided the basis for the 1985 Vienna Convention on the Protection of the Ozone Layer, the 1987 Montreal Protocol and subsequent amendments, bringing hope for the recovery of the ozone layer by the middle of the 21st century. Through the ongoing Global Atmosphere Watch, WMO is providing a vital input to the United Nations Framework Convention on Climate Change (UNFCCC).

One of the most massive international scientific undertakings of the 20th century was the Global Atmospheric Research Programme (GARP), a worldwide collaborative effort led by WMO, the International Council of Scientific Unions (now the International Council for Science), and the Intergovernmental Oceanographic Commission (*see* GARP (Global Atmospheric Research Program), Volume 1). Responding to a mandate from the UN, GARP sought to improve weather prediction capabilities and advance understanding of the physical basis of climate. A series of field observing experiments laid the groundwork for a Global Weather Experiment in

1979, in which the global atmosphere was, for the first time, observed in full detail for a meaningful period of time. The resulting data sets provided the basis for research that has greatly increased the accuracy and time range of weather prediction. WMO collaboration with other international organizations such as UNEP, the Food and Agriculture Organization, etc., has continued strongly with the establishment of the World Climate Programme (WCP) in 1979. Today, this program provides much of the scientific and administrative infrastructure supporting the UNFCCC, through which the nations are seeking to deal with the threat of human-induced climate change.

The WMO currently carries out a diverse suite of major programs of vital importance to science and society:

- The World Weather Watch (WWW) coordinates the gathering and dissemination of worldwide weather information through surface and space-based observing and communications systems. The global system comprises about 10 000 surface stations, hundreds of radiosonde stations, about 7000 ships, about 300 buoys, and a suite of four polar-orbiting and five geostationary satellites, all operated by national or regional organizations under complex international agreements on standards, codes, measurements and communications. The WWW also includes special activities on methods of observation and tropical cyclones.
- The WCP seeks to use existing climate information to improve economic and social planning; to improve the understanding of climate processes through research, so as to determine the predictability of climate and the extent of human influence on it; and to detect and warn governments of impending climate variations or changes, either natural or human-induced, which may significantly affect critical human activities. Established in 1979, the WCP comprises: the World Climate Data and Monitoring Programme; the World Climate Applications and Services Programme; the World Climate Impact Assessment and Response Strategies Programme; and the World Climate Research Programme. The WCP supports the Global Climate Observing System, encompassing all components of the climate system, atmosphere, biosphere, cryosphere and oceans.
- The Atmospheric Research and Environment Programme coordinates and fosters research on the structure and composition of the atmosphere, on the physics and chemistry of clouds and weather modification research, on tropical meteorology research and on weather forecasting. The Programme comprises the Global Atmosphere Watch – which incorporates long-standing WMO work on ozone and air pollution monitoring, the Weather Prediction Research Programme, the Tropical Meteorology Research Programme and the Physics and Chemistry of Clouds and Weather Modification Research Programme.
- The Applications of Meteorology Programme aims to help members in the application of meteorology and climatology to social and economic development, protection of life and property, and the well-being of humankind. It comprises the Agricultural Meteorology Programme, the Aeronautical Meteorology Programme, the Marine Meteorology and Associated Oceanographic Activities Programme and the Public Weather Services Programme.
- The Hydrology and Water Resources Programme concentrates on promoting worldwide cooperation in the evaluation of water resources and the development of hydrological networks and services, including data collection and processing, hydrological forecasting and warnings and the supply of meteorological and hydrological data for design purposes. It comprises the Operational Hydrology Programme–Basic Systems; the Operational Hydrology Programme–Applications and Environment; and the Programme on Water-related Issues.
- The Education and Training Programme encourages the exchange of scientific knowledge through special courses, seminars and training materials through advanced courses, surveys of personnel training requirements, development of appropriate training programs, establishment and improvement of regional training centers, and organization of training courses, seminars and conferences.
- The Technical Cooperation Programme seeks to bridge the gap between developed and developing countries by the systematic transfer of meteorological and hydrological knowledge and information, and by assisting members, especially developing countries, in obtaining the technical expertise and equipment for the development of their national meteorological and hydrological services. It works with major international partners such as the UNEP, the Global Environment Facility and regional development banks.
- The Regional Programme supports the regional associations in implementation of WMO programs and other activities through a network of regional offices.

Operational WMO activities are primarily funded by national members, which, for example, operate observing, communication and data-processing systems within the WMO framework. Members also contribute dues currently totaling about 255 million Swiss francs per year to support the organization s central activities. Additional funding for specific projects is provided by other national and international organizations.

Further information may be obtained from: WMO, 7 bis Avenue de la Paix, CP2300-1211 Geneva 2, Switzerland, http://www.wmo.ch.

JOHN S PERRY *USA*

WOCE (World Ocean Circulation Experiment)

The WOCE is a component of the World Climate Research Programme and is the most ambitious oceanographic experiment undertaken to date. In addition to global observations furnished by satellites, conventional *in situ* physical and chemical observations have been made by nearly 30 nations in four of the world s oceans. At the same time, global numerical ocean models are being developed to assimilate these measurements.

WOCE has two goals:

1. To develop models useful for predicting climate change and to collect the data necessary to test them. This will require determining and understanding on a global basis a number of aspects of the World Ocean Circulation (WOC) and their relation to climate:
    - the large-scale fluxes of heat and fresh water, their divergences over five years, and their annual and interannual variability;
    - the dynamical balance of the WOC and its response to surface fluxes;
    - components of ocean variability on scales of months to years and thousands of kilometers, and the statistics on smaller scales;
    - the rates and nature of the formation, ventilation and circulation of water masses that influence the climate system on time scales from 10–100 years.
2. To determine the representativeness of the specific WOCE data sets for the long-term behavior of the ocean, and to find methods for determining long-term changes in the ocean circulation. Specific objectives are:
    - to determine the representativeness of specific WOCE data sets;
    - to identify those oceanographic parameters, indices and fields that are essential for continuing measurements in a climate observing system on decadal time scales;
    - to develop cost-effective techniques suitable for deploying in a climate observing system.

WOCE used resources from nearly 30 countries to make unprecedented *in situ* and satellite observations of the global ocean between 1990 and 1998 and to observe poorly-understood, but important, physical processes. In parallel, great advances have been made during WOCE in the ability of ocean models to reproduce the known characteristics of the ocean and its circulation. Analysis, interpretation, modeling and synthesis of WOCE data will continue to 2002 but a prerequisite for its success will be the continuing assembly and quality control of WOCE data. The diverse WOCE data sets will serve as a unique resource for climate researchers and marine scientists for decades to come.

See also: Ocean Circulation, Volume 1.

FURTHER READING

WOCE International Project Office, *Ocean Circulation and Climate*, Brochure, Southampton Oceanography Center, South-ampton.

PIERS CHAPMAN AND JOHN S PERRY *USA*

Working Group on Numerical Experimentation (WGNE)

see WGNE (Working Group on Numerical Experimentation) (**Volume 1**)

World Climate Conferences (WCC)

see WCC (World Climate Conferences – 1979 and 1990) (**Volume 1**)

World Climate Programme (WCP)

see WCP (World Climate Programme) (**Volume 1**)

World Climate Research Programme (WCRP)

see WCRP (World Climate Research Programme) (**Volume 1**)

Y

Younger Dryas

Dorothy Peteet
Lamont Doherty Earth Observatory, Palisades, NY, USA and NASA/GISS, New York, NY, USA

With the dramatic rise in anthropogenic greenhouse gases over the past 100 years and evidence of recent global climate warming, questions have risen about the potential for triggering climate changes that might be very rapid. One of the best-studied examples of past climate change that was extremely rapid is the Younger Dryas (YD) interval. This pronounced cold event occurred about 13 000 calendar years ago and lasted over 1000 years. The YD was one of several climatic reversals in the warming trend that occurred as the world emerged from the last ice age. This period has generated considerable interest because evidence suggests that the close of this interval took place in a particularly short time interval in Greenland, with most of the change occurring in one to three years. This termination of the event indicates that it is possible for the climate to ip from a cold to a warm state very rapidly. Various physical and biological indicators reveal how rapidly plants and animal responded to this change.

The YD cold period was first discovered about 1900 as a result of an examination of plant macrofossils in a stratigraphic section in Denmark. The tiny leaves of *Dryas octopetala* (Rosaceae family) were found in layers above remains of shrubs and trees that indicated the climate was warming after the ice age. This return to Dryas, which is an indicator of climate conditions conducive to an arctic-alpine tundra environment, indicated that the vegetation was responding to a dramatic cooling. By the 1940s as the science of *palynology* (the study of pollen and spores) was established, investigations at many sites all over Europe showed signs of this pollen change from shrub/tree to tundra and then back again. Evidence for glacial advances during the cold snap was mapped in Scandinavia and even in Scotland, where glaciers actually reappeared after having melted away. By the early 1980s, scientists recognized the YD as a significant event that took place 11 000 to 10 000 carbon-14 (^{14}C) years ago throughout Europe. When looking for the event elsewhere in the world, this ^{14}C time interval was used as a radiocarbon *chronozone* (time slice).

Complications with these correlations have occurred with the recent discovery of differences between the radiocarbon time scale and the calendar time scale (a result of variations in the formation rate of ^{14}C because of variations in solar radiation). Several radiocarbon plateaus exist in the interval, and thus the YD interval actually spans more than 1400 calendar years. Thus, in calendar years, it is considered to have occurred between about 13 000 and 11 600 years ago. The estimates of the magnitude of the cooling range from • 10 °C in northern Europe to • 2 °C in southern Europe. Recent palaeoclimatic research based on the Greenland ice cores has also shown this as a major climatic event, resulting in a cooling estimated to be between • 7 and • 15 °C.

Evidence for the YD in the record of North Atlantic sea surface temperatures was found during the 1980s, as well as confirmation of its signature in eastern North America. Accelerator mass spectrometry (AMS) has been used to date lithologic, pollen, and macrofossil records from Atlantic Canada, southern New England, and New Jersey. The cooling in the eastern US is estimated to have ranged from • 3 to • 6 °C based on pollen/macrofossil and oxygen isotopic evidence from lake diatoms. Evidence indicates that the change from cold to warm conditions at the end of the YD in eastern North America occurred in less than 50 years.

The changes in the ocean and atmosphere have been linked to changes in the terrestrial system through modeling of this period and through data-model comparisons. Using general circulation models (GCMs), several research groups in the US and Europe have attempted to simulate the rapid cooling by experimenting with different climatic forcings. Colder ocean temperatures in the North Atlantic have been found to lead to colder air temperatures downstream in Europe and Asia, but not elsewhere around the globe. The subsequent discovery of the YD cooling through pollen and glacial records on the western coast of North America, and a coincident change in ocean circulation or

upwelling indicated in sedimentary records from the North Pacific Ocean (i.e., Santa Barbara Basin, Sea of Japan, Gulf of California), suggest that the North Pacific Ocean must be connected as a link in the causal sequence. GCM experiments forced by a slightly cooled North Pacific Ocean result in a significant cooling of the entire Northern Hemisphere because the North Pacific Ocean is such a large source of water vapor to the atmosphere (water vapor being the most important greenhouse gas). However, whether the Pacific change is caused by a change in the North Atlantic thermohaline circulation change (possibly as a consequence of the changes induced in the atmospheric circulation) or whether it is somehow generated independently, but at the same time, is still not resolved. Recent indications of changes in upwelling in the Cariaco Basin, Gulf of Mexico imply that the tropics also were involved.

The high-resolution Greenland ice cores indicate that the YD is just one of the dozens of pronounced and rapid climate changes that have occurred in the last 50 000 years. These abrupt events are indicated by such observations as an increase in the deposition of dust, a decline in annual snowfall, and a decrease in methane concentration. While some Antarctic ice cores show an asynchronous response between the Northern and Southern Hemispheres, with a lead of about 1000 years in the Southern Hemisphere, new results from the Taylor Dome in East Antarctica exhibit an apparently synchronous response.

Although significant progress has been made in recent decades in improving understanding of the YD event, the greater, spatial and temporal detail now being sought in terrestrial, sedimentary, and cryospheric domains promises significant advances in the decades ahead. These advances should improve understanding about how rapidly the climate can change and whether human activities (or natural events) might trigger such changes in the future. Higher resolution cores throughout the cryosphere, terrestrial, and marine realms will be the key to gaining new perspectives on the YD event.

See also: Climate Change, Abrupt, Volume 1; Earth System History, Volume 1.

Encyclopedia of
Global Environmental Change

egec

The volumes in this Encyclopedia are:

Volume One
The Earth system: physical and chemical dimensions of global environmental change

Volume Two
The Earth system: biological and ecological dimensions of global environmental change

Volume Three
Causes and consequences of global environmental change

Volume Four
Responding to global environmental change

Volume Five
Social and economic dimensions of global environmental change

Encyclopedia of
Global Environmental Change

Editor-in-Chief

Ted Munn

Institute for Environmental Sciences, University of Toronto, Canada

2

egec

The Earth system: biological and ecological dimensions of global environmental change

Volume Editors

Harold A Mooney

Stanford University, Stanford, USA

and

Josep G Canadell

GCTE/IGBP, CSIRO Sustainable Ecosystems, Australia

JOHN WILEY & SONS, LTD

Copyright © 2002 John Wiley & Sons, Ltd, Chichester
Baffins Lane, Chichester
West Sussex PO19 1UD, UK

National: 01243 779777
International: (• 44) 1243 779777
e-mail (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on http://www.wiley.co.uk
or http://www.wiley.com

Copyright Acknowledgments
A number of articles in the Encyclopedia of Global Environmental Change have been written by government employees in the United Kingdom, Canada and the United States of America. Please contact the publisher for information on the copyright status of such works, if required. In general, Crown copyright material has been reproduced with the permission of the Controller of Her Majesty's Stationery Office. Works written by US government employees and classified as US Government Works are in the public domain in the United States of America.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except under the terms of the Copyright Designs and Patents Act 1988 or under the terms of a license issued by the Copyright Licensing Agency, 90 Tottenham Court Road, London, W1P 0LP, UK, without the prior permission in writing of the publisher.

Other Wiley Editorial Offices

John Wiley & Sons Inc., 605 Third Avenue,
New York, NY 10158-0012, USA

Wiley-VCH Verlag GmbH, Pappelallee 3,
D-69469 Weinheim, Germany

Jacaranda Wiley Ltd, 33 Park Road, Milton,
Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01,
Jin Xing Distripark, Singapore 129809

John Wiley & Sons (Canada) Ltd, 22 Worcester Road,
Rexdale, Ontario M9W 1L1, Canada

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library.

ISBN: 0 471 97796 9

Typeset in 10pt Times by Laser Words Private Limited, Chennai, India
Printed and bound in Great Britain by Antony Rowe, Chippenham, Wiltshire

Editorial Board

Editor-in-Chief

Ted Munn
Institute for Environmental Studies, University of Toronto, Canada

Volume Editors

Volume One ***The Earth system: physical and chemical dimensions of global environmental change***
Dr Michael C MacCracken, *Lawrence Livermore National Laboratory, California, USA* and
Dr John S Perry, *formerly, National Research Council USA*

Volume Two ***The Earth system: biological and ecological dimensions of global environmental change***
Professor Harold A Mooney, *Stanford University, Stanford, USA* and
Dr Josep G Canadell, *GCTE/IGBP, CSIRO Sustainable Ecosystems, Canberra, Australia*

Volume Three ***Causes and consequences of global environmental change***
Professor Ian Douglas, *University of Manchester, Manchester, UK*

Volume Four ***Responding to global environmental change***
Dr Mostafa K Tolba, *International Center for Environment and Development, Cairo, Egypt*

Volume Five ***Social and economic dimensions of global environmental change***
Mr Peter Timmerman, *IFIAS, Toronto, Canada*

International Advisory Board

Dr Joe T Baker
Commissioner for the Environment, ACT, Australia

Professor Francesco di Castri
CNRS, France

Professor Paul Crutzen
Max Planck Institute for Chemistry, Germany

Professor Eckart Ehlers
IHDP/IGU Germany

Professor José Goldemberg
Universida de São Paulo, Brazil

Dr Robert Goodland
World Bank, USA

Professor Hartmut Grassl
WCRP, Switzerland

Professor Ronald Hill
University of Hong Kong, Hong Kong

Dr Yu A Izrael
Institute of Global Climate & Technology, Russia

Professor Roger E Kasperson
Clark University, USA

Professor Peter Liss
University of East Anglia, UK

Professor Jane Lubchenco
Oregon State University, USA

Mr Jeffrey A McNeely
IUCN, Switzerland

Professor Thomas R Odhiambo
Hon. President, African Academy of Sciences and
Managing Trustee, RANDFORUM, Kenya

Sir Ghillean T Prance
University of Reading, UK

Professor Steve Rayner
Columbia University, USA

Contents

Preface to the Encyclopedia of Global Environmental Change xi

Preface to Volume Two xvii

Biological and Ecological Dimensions of Global Environmental Change 1
Biological Invasions 11
Functional Biodiversity 20
IGBP Core Projects 37
Impacts of Global Environmental Change on Animals 56
Natural Systems: Impacts of Climate Change 67
Plant Dispersal and Migration 81
Plants – from Cells to Ecosystems: Impacts of Global Environmental Change 94
Sedimentary Records of Long-term Ecological Change 112
Terrestrial and Freshwater Ecosystems: Impacts of Global Change 122

Allometric 135
Angiosperm/Gymnosperm 135
Animal Physiology and Global Environmental Change 136
Anthropogenic Impacts on Atmospheric Oxygen 140
Autotroph/Heterotroph 143

Benthic/Pelagic 145
Biocenosis 145
Biocomplexity 145
Biodiversity in Freshwaters 146
Biodiversity in Soils and Sediments: Potential Effects of Global Change 152
Biodiversity: The UNEP Definition 159
Biogeochemical Cycle 159
Bioindicators 160
Biome 166
Biome Models 166
Biome–BGC Ecosystem Model 171
Biosphere 174
Biosphere Enhancement Ratio (BER) 175
Biosphere Reserves 175
Boreal Ecosystem–Atmosphere Study (BOREAS) 177
Boreal Forest 179
Buffering Capacity 184

C_3 and C_4 Photosynthesis 186
Carbon and Energy: Terrestrial Stores and Fluxes 190
Carbon Cycle 198
Carnegie–Ames–Stanford Approach (CASA) 202
CENTURY Ecosystem Model 206
Chaos and Cycles 209
Climax Vegetation 215
CO_2 Enrichment: Effects on Ecosystems 215
Community Richness 224
Conservation Biology 228
Controlled Environment Facilities in Global Change Research 228
Coral Bleaching (1997–1998) 231
Coral Reefs: an Ecosystem Subject to Multiple Environmental Threats 232
Crop Models 241

Daisyworld 247
Darwin, Charles Robert 247
Data Banks: Archiving Ecological Data and Information 248
De Wit, C T 259
Deposition, Dry 260
Deposition, Wet 261
Disturbance 261
DIVERSITAS 268

Ecosystem Services and Costing 272
Ecosystem Stability 281
Ecosystem Structure and Function 282
Ecotones 283
Ehrlich, Paul 288
Elton, C S 289
Environment 290
European Network for Research in Global Change (ENRICH) 290
Eutrophication 292
Evolutionary Processes in Ecosystems 292
Extinctions and Biodiversity in the Fossil Record 297
Extinctions (Contemporary and Future) 301

Fixed Nitrogen 308
Food Webs 308
Forest Gap Models 316
Forest Stand 323
Forest: the FAO Definition 324

Fragmentation and Corridors 324
Freshwater Fisheries 327

Gaia 332
Generic Decomposition and Yield (G'DAY): A Model of Global Change Impacts on Plant and Soil 332
Global Ocean Observing System (GOOS) 336
Global Terrestrial Observing System (GTOS) 337

Habitat 339
Halophyte/Halophobe 339
Holdridge Life Zone Classification 339
Holling, C S 340
Hutchinson, G Evelyn 341
Hydrology 343

IGBP (International Geosphere–Biosphere Programme) 350
IGBP-DIS (International Geosphere–Biosphere Program Data and Information System) 351
IGBP Terrestrial Transects 351
Infectious Diseases 357
Inter American Institute for Global Change (IAIGC) 363
International Biosphere Program (IBP) – a Foundation for Global Studies 364
Intertidal Zones 365
Iron Cycle 369

Keystone Species 375
Lakes and Rivers 375
Landscape Ecology 383
Large Scale Biosphere–Atmosphere Experiment in Amazonia (LBA) 383
Life Cycles 385
LOICZ (Land–Ocean Interactions in the Coastal Zone) 389
Long-term Ecological Research (LTER) System 390

Mangrove Ecosystems 395
Margalef, Ramon 401
May, Robert M 401
Mercury in the Environment 402
Metadata 409
Metapopulations 411
Microbial Diversity 421
Monsi, Masami 426
Mooney, Harold A 426

Natural Capital 428
Niche 429
Nitrogen Cycle 429
Nitrogen Deposition on Forests 435
Nitrogen Fixers 442
Nitrogen Mineralization 446
Non-equilibrium Ecology 446
Non-linear Systems 450

Odum, Eugene P 456
OH–Radical: is the Cleansing Capacity of the Atmosphere Changing? 457
Oligotrophic/Heterotrophic/Eutrophic 458

PAR (Photosynthetically Active Radiation) 460
Passerine 460
PET (Potential Evapotranspiration) 461
Phenology 461
Phenotype/Genotype 465
Phosphorus Cycle 466
Photosynthesis 470
Phytophagy 471
Plant Competition in an Elevated CO_2 World 471
Plant Functional Types 481
Plant Growth at Elevated CO_2 489
Plant Ontogeny 496
Plants – Carbon Balance and Growth 497
Population Sizes, Changes 504
Productivity 515
Productivity of Terrestrial Ecosystems 516

Raven, Peter 522
Redfield Ratio 522
Redox Potential 523
Remote Sensing 524
Remote Sensing, Terrestrial Systems 528
Resilience 530
Respiration 531
Restoration, Ecosystem 532
Rhizosphere 539
r–K Strategies 540
Rodin, Leonid E 540
Rosswall, Per Thomas 541

Salinization 543
Soil Mineralization 543
Stable Isotopes 544
Succession, Definition of 550
Succession 551
Suess Effect 557
Symbiosis 557

Tamm, Carl Olof 559
Taxon 560
Taxonomy/Systematics 560
Temperate Coniferous Forests 560
Temperate Deciduous Forests 565

Temperate Grasslands 569
Temperature Increase: Effects on Terrestrial Ecosystems 575
Thornthwaite, Charles Warren 581
Tropical Forests 582
Tropical Savannas 586
Tucker, Jim 592
Tundra 593

Vegetation Ecosystem Model and Analysis Project (VEMAP) 603
Vernadsky, Vladimir 604
Volatile Organic Compounds, Biogenic (VOCs) 605
von Humboldt, Alexander 609

Weed 611
Whittaker, Robert H 611
Wilson, Edward Osborne 612
Woodwell, George Masters 613

Alphabetical List of Articles 615

Preface to the Encyclopedia of Global Environmental Change

THE ENVIRONMENT

The word *environment*, whose dictionary meaning is simply that which surrounds, has in the last few decades become a buzzword , encompassing an exceedingly diverse array of elements and social issues. Taking the original meaning as a point of departure, it is clear that we humans depend totally on the environment provided by planet Earth for the food we eat, the water we drink, and the air we breathe. Thus changes in this environment must be of vital concern. Will Earth continue to sustain humans in a way that also encourages the ourishing of the other living things with whom we share the planet? This question has loomed ever larger as it has become more evident that human activities have been inducing major changes in all of the compartments of the global environment. We have converted forests and savannas to farms and cities; we have exhumed ancient treasures of fuels and minerals; we have used the rivers and winds as convenient sewers; and we have released entirely new chemical compounds and organisms into the environment. In the 1960s, the scienti c community began to use the word *environment* in this new non-specialist sense. Soon too, Departments of the Environment were created by many governments, and new scienti c journals began publication while others were re-named. For example, the *International Journal of Air Pollution* became *Atmospheric Environment*! In the ensuing decades, the world community has come to see the environment in many different ways, as a life-support system, as a fragile sphere hanging in space, as a problem, a threat and a home.

GLOBAL ENVIRONMENTAL CHANGE

A broader and deeper understanding of the global aspects of environmental concerns emerged in the 1970s and 1980s, and a new phrase *global environmental change* acquired popular currency. Paleoresearch had revealed that environmental change was far from new, and by no means the sole result of the actions of heedless humans. Since the planet s formation, virtually every element of its environment has been undergoing massive changes on all space and time scales. Oxygen waxed and carbon dioxide waned in the atmosphere. Continents moved about the planet s surface like scum on a soup kettle. Great ice sheets grew and shrank. Above all, the force of life (the biosphere) emerged as a dominant driver of planetary change.

However, another vital insight began to emerge about 1980: the inescapably interlinked nature of these many environmental changes. On the very longest time scales, continental drift moved lands into different climates but also changed the climate of the globe itself. Photosynthetic life changed the atmosphere but also made possible more advanced life forms that could take advantage of the new environment. On shorter time scales, atmosphere and ocean often interact to produce the massive changes in the Southern Paci c that we term El Niño and La Niña, whose consequences extend across the planet, and profoundly affect even our socioeconomic systems. Indeed, we have come to realize that human-induced perturbations in the environment are becoming increasingly large, and are potentially coming to dominate the natural workings of the complex and interdependent global system that sustains life on Earth. Humans and their global environment are no longer independent; they are ever-increasingly becoming interdependent components of a single global system.

Thus, the term *global environmental change* has come to encompass a full range of globally signi cant issues relating to both natural and human-induced changes in Earth s environment, as well as their socioeconomic drivers. This implicitly includes concerns for the capacity of the Earth to sustain life that have motivated the development of studies of global change and sustainable development in the last few decades. Analyses of global environmental change therefore demand input from the social sciences as well as the natural sciences (and indeed also from the engineering and health sciences) necessitating an inescapably interdisciplinary approach.

Scientists from many disciplines have been attracted to this growing eld of global environmental change. This is particularly noticeable in the biological sciences through the encouragement of IGBP (International Geosphere Biosphere Programme), which invites ecologists to expand their eld of vision from the plot and landscape scales to the regional, continental and global ones, and to interact with scientists from other disciplines in exploring environmental change at these larger scales. Indeed this trend has encouraged publishers and scienti c societies to introduce new journal titles, and these publications from all accounts appear to be ourishing.

At the same time, human social, economic, and cultural systems are rapidly changing under the in uence of growing globalization. In the economic sphere, for example, today s discourse centers on multinational corporations, the Global

Economy, and the Globalization of Trade. Thus, the agendas of most of the physical, life, and social sciences increasingly focus on global-scale changes.

THE EVER-CHANGING ENVIRONMENT

Since the launching of Earth-viewing satellites, we have been able to see a constantly updated moving picture of our planet. Great weather systems sweep around the globe, waves of green and brown ebb and ow over the continents with the seasons, as do white waves of ice and snow. Nevertheless, a remarkable feature of the Earth system has been its relative stability. Since the dawn of life, the planet's environment has remained within a range of conditions that has supported life. Moreover, with some notable and perhaps cautionary exceptions, changes from year to year, decade to decade, millennium to millennium, have been modest, particularly during the last 10 000 years. Many features of human society exhibit similar behavior. Civilizations and cultures evolve slowly for the most part in response to environmental, economic, and social driving forces over their long lives. Great cities like Rome endure through the ages. A citizen of ancient Babylon probably would have little difficulty understanding the politics of modern Toronto.

But today we appear to be entering an era of change unprecedented since Babylon was a cluster of mud huts. There are many reasons to believe that changes greater than humankind has experienced in its history are in progress and are likely to accelerate.

Virtually every measure of human society – numbers of people and automobiles, airplanes, energy consumption, generation of waste – is increasing exponentially. While the values of a few indicators are leveling off, the magnitudes of annual increases of others remain immense.

Major changes are becoming evident in many critical elements of the environment – increasing carbon dioxide concentrations, stratospheric ozone depletion (not to mention the stratospheric Antarctic ozone hole and the possibility of a similar Arctic event), rising sea level, declining productivity of soils, widespread collapses of fisheries, dramatic declines in biodiversity, destruction of tropical forests, etc. Within living memory, the countryside around large cities has been swallowed up by suburban developments and highways. Humans are stepping on the gas pedal of the planet's environment, and perhaps recklessly breaking the survivable speed limit of global change.

Paleoclimatological studies have shown that the stable environment that we take for granted has not always prevailed. Indeed, the relatively brief period in which human civilization has developed is somewhat unusual in its equable stability. Neolithic hunters in Europe some 13 000 years ago saw their climate shift from temperate to glacial in a single short lifetime. The evolution of human society is punctuated by wars, pestilence, and technological revolutions. Major – perhaps even catastrophic – change could occur in the future, because we know it has occurred in the past.

Rapidly advancing understanding of both natural and human systems – and above all the ability to translate that understanding into quantitative models – has enabled us to explore the future of our global society and its global climate with unprecedented realism. Although prediction of the single future path that we will follow is inherently unpredictable, it is possible to map a broad range of future environmental trajectories that we might take, each completely consistent with our understanding of how the system works. Such scenario-building exercises amply confirm our concerns that the changes of the 21st Century could be far greater than experienced in the last several millennia. Business-as-usual for human society appears to imply business-as-highly-unusual for the global environment.

THE INTERLOCKING BIOGEOPHYSICAL – SOCIOECONOMIC SYSTEMS

Recognition came in the 1970s that many of the environmental issues are inter-connected through the biogeochemical cycling of trace substances, especially carbon, sulfur, nitrogen and phosphorus. In fact, in a prescient statement on the main environmental research priority of the 1980s, Mostafa Tolba (then Executive Director of the United Nations Environment Programme) and Gilbert White (then President of the Scientific Committee on Problems of the Environment) drew the attention of both the scientific and the science-policy communities to the need to understand the major global biogeochemical cycles in order to maintain the global life support systems in a healthy state (Tolba and White, 1979). Quoting from that statement,

> We draw attention to the fundamental scientific importance of understanding the biogeochemical cycles which link and unify the major chemical and biological processes of the Earth's surface and atmosphere. The results will have practical significance for all of us who inhabit an Earth with limited resources and who, by our actions, increasingly affect the quality of the human environment.

So it is that many of the global environmental issues – acid rain, stratospheric ozone depletion, climate change, nitrogen over-fertilization – are inter-related through the global biogeochemical cycles.

One of the interesting results of the study of the Earth's history has been the discovery of the global teleconnections of the Earth system, with some major climatic shifts occurring simultaneously in the two hemispheres. In an analogous way, there has been recognition for a long time that human social, economic, and cultural systems are globally inter-related. That these are connected in turn to the

Earth system was implicitly recognized as early as the 1972 Stockholm Conference on the Human Environment.

To manage the human responses to this enormously varied but at least moderately coupled world system in an era of increasing global change through the diverse array of local, national and international organizations is indeed a daunting challenge. An essential first step is to describe past and present states, then to explain the various phenomena observed, and finally to develop predictive models (or at least a range of scenarios) describing the future behavior of this total Socioeconomic–Cultural–Environmental system. In this process, environmental scientists must learn how to assimilate better the new information constantly being received, although uncertainties, often large, will continue to exist. Within this mix, rational and effective policies must be developed that will balance the risks and costs of global environmental change in an adaptive way. Only then will we be able to even formulate, much less start to implement, rational and effective policies that cope effectively with global change. The prospect of change tempts us to think in terms of winners and losers. However, such analyses often do not play out in simple ways. For example, a modest increase in rainfall would cause farmers to rejoice while vacationers would despair. However, greater farm production can lead to lower commodity prices, thereby reducing farm incomes while making vacation food purchases less expensive. Human society is pretty well adapted to the present environment, so change is necessarily a challenge.

In the longer-term, much depends on the ability of societies to respond, to adapt. Societies will differ in the resources – natural, human, and technological – that are available to them. They will also differ in terms of the values and priorities they attach to physical, social, and environmental goals, and in the social and political mechanisms that they employ to seek these goals. These differences in the human world of generations yet unborn may be as great – and as significant – as the changes in the global environment. A major challenge for our times is to develop frameworks for understanding complex interdisciplinary issues of this complexity.

No discussion of change and the future can be complete without consideration of risk. Projections of the future, however imaginative or soundly based, necessarily center on plausible, surprise-free scenarios – population will increase, economies will advance, climate will change – all typically slowly and smoothly. However, such projections made at the beginning of the 20th Century would have missed two World Wars, the automobile, aviation, space travel, television, and McDonald's in Beijing. By definition, genuine surprises cannot be predicted. However, an understanding of the impacts of past surprises may help us to make our society and our world somewhat more robust in the face of the unknown surprises that await us.

THE HUMAN DIMENSIONS OF GLOBAL ENVIRONMENTAL CHANGE

The human dimension of global change has been defined as the ways in which individuals and societies contribute to global environmental change, are in uenced by it, and adapt to it. Many empirical studies have been undertaken that describe human activities in physical terms, based on various kinds of indicators. Rates of deforestation, urbanization, and changing levels of emissions of greenhouse gases are only some measurable contributors to environmental change. But the dynamics of human activities and global change are much more complex, and they re ect the complexities of the human–nature relationship.

While the sheer burden of human activity on the planet is important, we know that the major forces at work are human beings operating together through political systems, corporations, interest groups, and beliefs that sway whole peoples. These raise questions about the nature of choice, about the hopes, dreams, and frustrations that impel people forward, or block them. At the moment, for instance, the dream of the good life – which has been a critical element of every religious tradition in the world – is increasingly being defined in terms of material possessions and powerful images that are shown on global media. Can this particular version of human happiness be sustained on a limited planet? Some people say that it can; others say that it cannot.

The central questions about the role of human beings in global environmental change revolve around social, cultural, economic, ethical, and even religious issues. These are becoming more and more pressing, and more and more foundational as human beings deliberately or inadvertently modify more and more of the planet. It is also obvious that in this generation the modification of organisms and ecosystems may well extend to the modification of human beings themselves. Among the fundamental questions are: What motivates us towards saving or harming the environment? How do we see ourselves with regard to nature? What is our responsibility to this and future generations? Who do we think we are, and who would we like to become? Appropriately enough, Volume Five of this Encyclopedia wrestles, in many different voices, with these ultimate questions that remain intimately linked with the sweep of physical, chemical, biological, geographical, and institutional changes documented and discussed throughout the earlier four volumes.

The Brundtland philosophy urges us not to reduce options for future generations. Implementation of this idea is, however, difficult. For example, it often seems more compelling to alleviate current poverty than to protect the environment and renewable resources for future generations. Many scientists agree that new approaches are needed to meld the social with the natural sciences in the policy arena. Some of the new methodologies that go beyond the physical sciences

are described in Volumes Four and Five of this Encyclopedia, e.g., post-normal science, integrated environmental assessment, and the precautionary principle.

In a commentary in which he described the 21st Century as the century of the environment, Edward O. Wilson began by referring to the growing *consilience* (the interlocking of causal explanations across disciplines) so that the interfaces between disciplines become as important as the disciplines themselves. He then stated that this interlocking amongst the natural sciences will in the 21st Century also touch the borders of the social sciences and humanities. In the environmental context, environmental scientists in diverse specialities, including human ecology, are more precisely defining the arena in which that species arose, and those parts that must be sustained for human survival (Wilson, 1998).

This is a major challenge for environmental scientists. Already through DNA techniques, for example, it is possible to trace back connections amongst prehistoric peoples through the last Ice Age to modern times.

THIS *ENCYCLOPEDIA OF GLOBAL ENVIRONMENTAL CHANGE*

More than two million papers are published every year in science and medicine, a twentyfold increase since 1940 that taxes the resources of concerned citizens, scientists, university departments, research institutes and libraries. At the same time, many policy-motivated organizations find it difficult to draw together the necessary expertise for resolving the newly emerging environmental issues. The scientific literature relating to the environment is burgeoning. However, research syntheses are in general still scattered across a broad spectrum of journals and books, and information on global environmental change is not readily available in an inter-related way. In particular, it is quite uncommon for contributions from the natural sciences to appear in the same journal or workshop proceedings as contributions from the social sciences and the humanities.

This *Encyclopedia of Global Environmental Change* is a comprehensive and integrated reference in the broad area of global environmental change that will be conveniently accessible and productively usable by students, managers, administrators, legislators, and concerned citizens. The Encyclopedia consists of five volumes of interrelated material:

Volume 1 The Earth system: physical and chemical dimensions of global environmental change (Michael C MacCracken and John S Perry, editors)

Volume 2 The Earth system: biological and ecological dimensions of global environmental change (Harold A Mooney and Josep G Canadell, editors)

Volume 3 Causes and consequences of global environmental change (Ian Douglas, editor)

Volume 4 Responding to global environmental change (Mostafa K Tolba, editor)

Volume 5 Social and economic dimensions of global environmental change (Peter Timmerman, editor)

The first four volumes cover the broad issues of the science and politics of global environmental change. Volume Five adds an often understated but extremely important aspect, linking as it does, global environmental change to the socioeconomic, cultural and ethical dimensions of human societies. Here will be found a rich panoply of writings by people who are not natural scientists but who have thought deeply about the environment. It places global environmental change in a refreshing historical, sociological and cultural context. In many contributions, the time horizons of most interest are the last hundred and the next hundred years. However, some contributions dip backward millions of years. Throughout, the emphasis is upon the dynamics of the various processes discussed – how and why did they change? A second recurrent theme is the interconnection of processes and changes – What produces change? What is impacted by change? Finally, we attempt to deal even-handedly with natural and human-induced change, and with impacts on both the natural world and human society. From the numerous diverse articles in the Encyclopedia, we believe that the user can obtain a coherent picture of this complex and dynamic system of which we all are a part.

To assist in promoting this coherence, each volume begins with a group of extended essays on major topics that embrace the field covered in that volume. These are intended to provide an introduction to the topic, a convenient road map through cross-references for exploring the Encyclopedia. Then there follows, in alphabetical order, shorter articles on a variety of scientific topics, descriptions of scientific programs, definitions, acronyms, and biographies of leading contributors to the field – from Charles Darwin, through the Russian ecologist Vernadski, to the three most recent environmental Nobel Laureates – Crutzen, Rowland and Molina. Indeed, these definitions, biographies and acronym definitions are, we believe, a uniquely valuable feature of the Encyclopedia. In the case of acronyms of international organizations and programs, it is no exaggeration to state that a young environmental scientist requires not only a good understanding of science, but also a good knowledge of acronyms, if they are to follow the discussions taking place at many international meetings! Also included in the alphabetic listings are abundant cross-references to related topics in the same or other volumes.

The substantive scientific articles that comprise the meat of the Encyclopedia are original contributions by active scientists from around the world, and thus

represent authoritative and up-to-date summaries of the state of current knowledge, direct from the producers of this knowledge. A number of these articles break new ground in synthesizing and summarizing our understanding from novel viewpoints and in unconventional ways. Thus, readers will find a wide variety of styles and approaches within the articles, re ecting in a unique way the rich diversity of today's world science. The articles have also had the benefit of careful reviews, particularly by our Editors.

The scientific essays and some of the program descriptions begin with a few italicized paragraphs written for non-specialists. These are not intended to be abstracts of the paper to follow, but rather are aimed at providing the reader with an introduction into why the topic is important and where it fits into the broader aspects of global change – a kind of encouragement to read on. Reading on should not be too difficult a task, since most of the scientific essays are written at the level of journals such as *Scienti c American* and *AMBIO* that are intended for the non-specialist. The Encyclopedia will be a valuable source of information for everyone with a general interest or a need-to-know in the various environmental fields (the natural sciences, socio-economics, engineering, the health sciences, and policy analyses) particularly as they relate to global-scale environmental change, its drivers, and its consequences. It is also expected that among the audiences for each volume will be practitioners and researchers in the fields covered by the other four volumes. We believe that this rather unusual – indeed unique – Encyclopedia will be used in a variety of ways.

1. Some people will employ it simply as a convenient source of information on specific topics – What has been happening recently to the sea ice in the Arctic Ocean? Who was Roger Revelle? What is soil mineralization? What is ICSU or IGBP? What is the Kyoto Protocol? What is deep ecology?
2. Serious students in the environmental sciences may begin by reading one or all of the introductory essays, which present highly compressed crash courses in a range of central topics.
3. We believe that the substantive specialized articles that constitute the bodies of the volumes are productively and enjoyably readable in their own right.
4. Finally, a case has been made by one of the volume editors (Peter Timmerman) for encyclopedia browsing as an enjoyable and productive pastime (*see* Encyclopedias: Compendia of Global Knowledge, Volume 5).

We believe that these five volumes are in the tradition of the human aspiration towards the compilation of global knowledge that sparked the preparation of the first encyclopedias in the 18th Century.

Our Editors have had the benefit of a marvellous postgraduate education in the course of the Encyclopedia's evolution! We are immensely grateful to the many, many authors in our virtual university who have contributed their wisdom to this project.

ACKNOWLEDGMENT

Many helpful comments and draft paragraphs were received from the Volume Editors during the preparation of this Preface.

REFERENCES

Tolba, M K and White, Gilbert F (1979) Global Life Support Systems, SCOPE Newsletter, No. 7, October. 2 pp.

Wilson, Edward O (1998) Integrated science and the coming century of the environment, SCIENCE 279, 2048–2049.

Ted Munn
Editor-in-Chief

Preface to Volume Two

The rapid evolution of the science of global change over the last 15 years has provided major insights on the functioning of the biosphere and its interaction with the physical environment. As the realization grew that major changes of key environmental drivers were taking place, such as increased global temperature and atmospheric CO_2, Earth system science was challenged to better understand the function of the Earth system as a whole, with the biosphere both as a major driver and recipient of those environmental changes.

This volume of the *Encyclopedia of Global Environmental Change* deals with the fundamental science that underpins the understanding of quantities and processes that control the basic biogeochemical cycles such as carbon, nitrogen, and phosphorus, and the associated changes in ecosystem physiology and structure under current and predicted human-driven global environmental change. These changes also include the dramatic loss of species, and the world's reorganization of biota towards higher homogenization. Here we cover major themes relevant to various spatial scales of the biosphere including species, populations, communities, ecosystems, biomes, and the Earth system as a whole. In addition, this volume provides an important account of new research tools that are allowing the science of global ecology to quickly develop. Finally, a number of international programs have been fundamental in the development of thinking in this new field of science; and a description of the most important efforts is included.

The new understanding provided in this volume forms the basic building blocks which will allow us to (i) integrate the biospheric processes and quantities with the Earth's physical system into one single Earth system, (ii) detect changes in the world's ecosystems' function and structure, both as already visible impacts of global change and also as indicators of change, and (iii) develop the capability to predict impacts on the biosphere brought about by global environmental changes over the next few decades to century. This information, in turn, will form the basis to develop the technical capabilities and policy tools to mitigate and adapt to undesirable changes.

Volumes One and Two deal with the fundamental science that allows the development of capabilities to assess global change and its impacts on ecosystem goods and services critical to human well being (the latter largely dealt with in Volume Three). Both the realization that global change is happening now and its already noticeable impacts on terrestrial and aquatic ecosystems have prompted the action of new policies and institutions dealt with in Volumes Four and Five.

Harold A Mooney
Josep G Canadell
Editors of Volume Two

Biological and Ecological Dimensions of Global Environmental Change

JOSEP G CANADELL[1] AND H A MOONEY[2]

[1]CSIRO Sustainable Ecosystems, Canberra, Australia
[2]Stanford University, Stanford, CA, USA

A HUMAN DOMINATED WORLD

Humans have populated the Earth for thousands of years but not until the last few centuries have they played a major role in influencing the structure and metabolism of the global biosphere. Human-induced changes have today reached such an extent that, in some cases, they are as important or even more important than those caused by the geological and climatic forces that shaped the Earth's environment over the previous thousands of years.

There is no doubt that the growth of human populations and associated resource consumption have been the major drivers behind the large biospheric transformations that have occurred, particularly in the last few decades. Both the total and the *per capita* exploitation of the Earth's resources have been nearly exponential for some time now. The human population has increased tenfold to 6 billion over the last three centuries along with, for example, the growth in the number of cattle (to 1.4 billion) as well as in the number of automobiles. Figure 1 provides a vivid illustration of the extent of large-scale human interference with the natural environment.

Many changes are currently taking place (and at an accelerating pace) in the biosphere. See, for example, Vitousek *et al.* (1997) and Figure 1. These changes include:

1. The historic transformation, to various degrees of intensity, of 30–50% of the land surface, particularly the 10% converted from natural vegetation into cropland. Twenty percent of existing forest and woodlands has disappeared in the last three centuries (Figure 2).
2. Over the last four glacial periods, the atmospheric concentration of carbon dioxide oscillated between 180 and 280 ppm and not until recently have human activities (mainly industrialization and deforestation), driven carbon dioxide concentrations above this narrow window. The biogeochemical cycle of carbon is consequently perturbed on a grand scale, which must affect world vegetation patterns and climate likewise.
3. While carbon is the largest contributor to plant living matter, with about 45% of dry weight, no single life process would be possible without the action of highly specialized proteins (enzymes) for which nitrogen is the key element. Nitrogen is now being added to the biosphere by chemical synthesis (fertilizer production) and expansion of nitrogen-fixing crops at a similar rate to that of nitrogen produced naturally, thus doubling the planet's nitrogen production. The single largest change has resulted from the increased use of synthetic inorganic fertilizers in agriculture; half of the amount that has ever been used on Earth has been applied during the last 15 years. In some areas (with poorly buffered lakes and soils), the net result is acidification. In other areas, the net result is nitrogen over-fertilization, with the risk that the assemblage of species in whole biomes may change radically.
4. Humans use about 50% of the runoff that is fresh and reasonably accessible, irrigated agriculture being the most important single user. In the Northern Hemisphere, 77% of the water flowing into the oceans comes from rivers that have been regulated by dams or manipulated in other ways. This has had large impacts on the transport of sediment to the oceans and in the salinity of coastal waters.
5. Extinction rates of species due to human activities, and particularly due to changes in land use and cover, are calculated to be between 100 and 1000 times greater than in pre-human days. For instance, up to 25% of bird species have been driven to extinction. The current loss of species is comparable to those occurring during the five great extinction events of the geological past. *See*, for example, **Extinctions and Biodiversity in the Fossil Record**, Volume 2; **Extinctions (Contemporary and Future)**, Volume 2.
6. Invasive species are reshaping the biotic composition of the world's ecosystems. In just six nations studied, there have been over 100 000 introduced species (plants, animals and microbes) about 2000–50 000

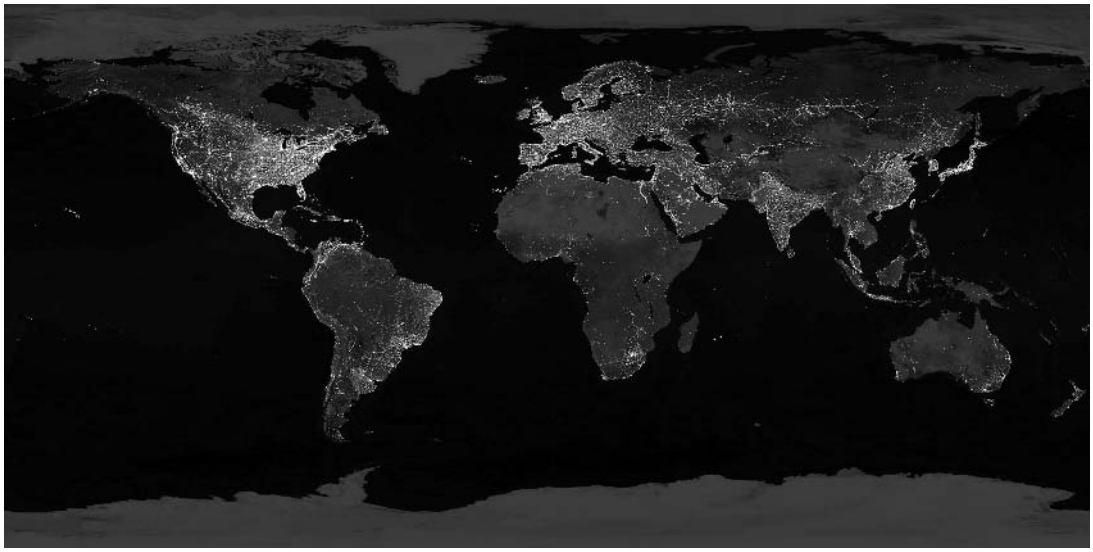


Figure 1 Earth city lights. (Image by Craig Mayhew and Robert Simmon, NASA GSFC, based on data courtesy of Christopher Elvidge, NOAA National Geophysical Data Center)

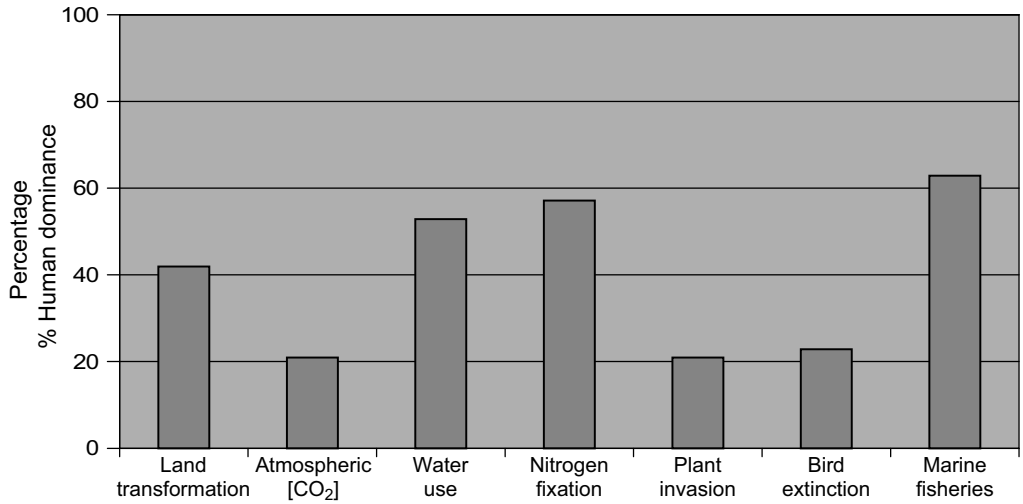


Figure 2 Human alteration of various biospheric compartments. (after Vitousek *et al.*, 1997)

species per country (*see also* **Biological Invasions**, Volume 2). Biological invasions are now considered the second largest threat to biodiversity losses in terrestrial ecosystems after habitat destruction (see for example Figure 3).

7. Fisheries around the world are known to be in trouble with about 22% of recognized fisheries being overexploited or already depleted (see Vitousek *et al.*, 1997). Global fish catches have leveled in recent years.

These examples illustrate the extent of human interventions in many of the Earth's key compartments and processes. Recognition of this situation has led Nobel Laureate Paul Crutzen (2000) to propose that mankind has entered a new biogeophysical era that he has called the *Anthropocene* (*see* **Anthropocene**, Volume 1). This idea is intended to sharpen the distinction between the current human-dominated world (starting around the 1800s) and the previous Holocene era that spanned the post-glacial period of the past 10 000–12 000 years when the Earth's climate and metabolism were exclusively controlled by geological forces. In this respect, *anthropocene* refers not only to the large scale of human interventions but also to the fact that these various drivers affect all terrestrial and aquatic ecosystems on Earth.

Finally, we should draw a distinction between biospheric changes due mainly to continental- and global-scale modifications of the chemical and radiative properties of the environment (climate greenhouse warming, stratospheric ozone depletion and acid deposition) with the direct changes

Figure 3 Invasion rate in San Francisco Bay (data from Cohen and Carlton, 1998). Image courtesy of MITI, ERSDAC, JAROS and the US/Japan ASTER Science Team

in the characteristics of the physical environment (forest clearing, urbanization, irrigation, etc.). The emphasis in this volume is on the former but in the real world, several stresses operate simultaneously on the biosphere. This kind of behavior will be considered in volume three, and the consequences for policy formulation will be discussed in volume four. The Intergovernmental Panel on Climate Change (IPCC) process (*see* **Intergovernmental Panel on Climate Change (IPCC): an Historical Review**, Volume 4) is a shining example of how to move from science to policy, certainly to be emulated with respect other biospheric changes in coming decades. To cite three examples,

1. How will wheat yields be affected over the next several decades, as a result of increasing frequencies and intensities of both ground-level ozone episodes, and of increasing heat waves and droughts? (not to mention globalization of markets, bioengineering, and changes in food preferences).
2. How will salmon and cod stocks respond to changes in technology, sea temperatures and ocean currents, and national and international regulations?
3. How will ecosystems respond to rising carbon dioxide concentrations and climate warming? (This question is addressed in several papers in this volume.)

THE CHANGING METABOLISM AND STRUCTURE OF THE BIOSPHERE

Over the last two centuries, the extraordinary changes that have taken place in the land surface and in the atmosphere

over the land have had large impacts on biospheric structure and metabolism. From an Earth metabolic perspective, the large conversion of land since the early 19th century has reduced the terrestrial biosphere's capacity to produce organic matter (amount of carbon stored in plants per year) by 5% and has released 182–199 Pg of carbon (1 Petagram is 1 billion tonnes) to the atmosphere. Changes in land use and cover have been responsible for 33% of the increased concentrations of carbon dioxide observed in the atmosphere.

Earth observing systems can already detect the changing metabolism and structure of the biosphere. For instance, analyses of monthly air samples around the world for carbon dioxide and other chemical species have shown that the difference between the higher concentrations in the winter, when carbon dioxide release from respiration is the prevailing process, and the lower concentrations in the summer, when carbon dioxide uptake is the prevailing process, has increased between 20 and 40% over the last two decades. This is the result of changes in the strength of key metabolic processes such as photosynthesis (carbon dioxide uptake) and plant and microbial respiration (carbon dioxide emissions), and the length and intensity at which these processes operate in different parts of the planet. Satellite-sensed canopy development (greening) at the beginning of the growing season shows an increased length of this biologically active season in the Northern Hemisphere, specifically an advance of about one week in the onset of the spring drawdown of carbon dioxide. Using similar techniques, it has been shown that terrestrial net primary productivity of some regions in the world has increased over the same period.

The primary driver of these changes is thought to be increased air temperature although higher photosynthesis rates due to carbon dioxide fertilization and nitrogen deposition may have also contributed to some degree.

Changes in the metabolism and structure of the biosphere may well be the best proof that global change is taking place. A large body of evidence on the existence of these changes has been compiled over the past few years, including changes in key physiological processes as well as species redistribution in land and aquatic systems.

As indicated above, changes in plant and animal phenology have also been linked to changes in temperature around the world but most especially at high latitude sites where temperature changes have been the largest. Earlier timing of reproduction has been found for many bird species in a number of countries, as well as advances in the arrival time of migratory birds. In the last few decades, an increase in forest production in several countries in Europe has been related to changes in climate. The lengthening of the growing season is the most common response found, particularly in cold ecosystems of high latitudes.

Changes in phenology and physiology ultimately result in changes in plant and animal abundance and distribution. These changes have been well documented in coastal communities of macro-invertebrate species and with plankton and fish species in the open ocean (increasing abundance of southern species and a decrease of northern species in the Northern Hemisphere). Poleward shifts and movement towards higher elevations of plant and animals have been associated with warming trends regionally with well document shifts in the ranges of many species, including American, British, and European butterflies and insects. Species shifts also include viruses and parasites harmful to human health.

Coastal zones are also undergoing multiple changes, some linked to increased frequency of the El Niño Southern Oscillation (ENSO), with effects on species distribution and stability of coral reefs around the world. Equally important, leveling of fishery harvests and loss of biodiversity is spreading in coastal zones due to dead zones caused by nitrogen pollution. Accelerated nitrogen cycling on land produces NPP (net primary production) decline and forest dieback under chronic heavy nitrogen loads.

Finally, land transformation affects the water cycle and land surface properties. Both have major feedbacks to local and regional weather and climate, which in turn influence the type of vegetation structure that can exist. To put this effect in perspective, catchment studies indicate that water yield decreases 25–40 mm for each 10% increase in tree cover while increases in transpiration of water taken up by deep roots may account for as much as 50% of the observed responses.

Many contributions in volumes 2–5 elaborate on these themes, but of special relevance are the following in Volume 2:

- A series of essays on many of the world's biomes – e.g., the boreal forest, temperate deciduous forests, temperate grass lands, etc. . . .
- **Natural Systems: Impacts of Climate Change**;
- **Plants – from Cells to Ecosystems: Impacts of Global Environmental Change**;
- **Biological Invasions**;
- **C_3 and C_4 Photosynthesis**;
- **Carbon and Energy: Terrestrial Stores and Fluxes**;
- **Carbon Cycle**;
- **CO_2 Enrichment: Effects on Ecosystems**;
- **Nitrogen Cycle**;
- **Nitrogen Deposition on Forests**;
- **Phenology**;
- **Plant Growth at Elevated CO_2**;
- **Plants – Carbon Balance and Growth**;
- **Remote Sensing, Terrestrial Systems**;
- **Temperature Increase: Effects on Terrestrial Ecosystems**.

THE EMERGING SCIENCE OF GLOBAL ECOLOGY

Realization of the existence of global environmental problems such as increasing atmospheric carbon dioxide, the stratospheric ozone hole, and other changes more continental in nature, such as nitrogen deposition and acidification of land and aquatic systems, has extended the ecological sciences towards the new field of *global ecology*. In this connection, the global environmental issues mentioned above have been found to be related through the global biogeochemical cycles. Thus understanding the interactions amongst these cycles is fundamental to human attempts to preserve global life-support systems. There is therefore an obvious practical need to develop a new type of global science.

The new global ecology saw its beginnings during the International Biosphere Programme (IBP) during the late 1960s with a focus on primary productivity of the earth's ecosystems (*see* **International Biosphere Program (IBP) – a Foundation for Global Studies**, Volume 2). It was not until the onset of the International Geosphere–Biosphere Programme (IGBP), however, that the ecological sciences together with other disciplines began to develop a more integrated bio-geophysical Earth system science. This new program also focused on the nature of the human perturbation and how it is changing the Earth's functioning. The IGBP has operated in more than 60 countries and has led the thinking on global change research for more than 10 years. It is now entering phase II with a further emphasis on integration and interdisciplinary research (*see* **IGBP Core Projects**, Volume 2).

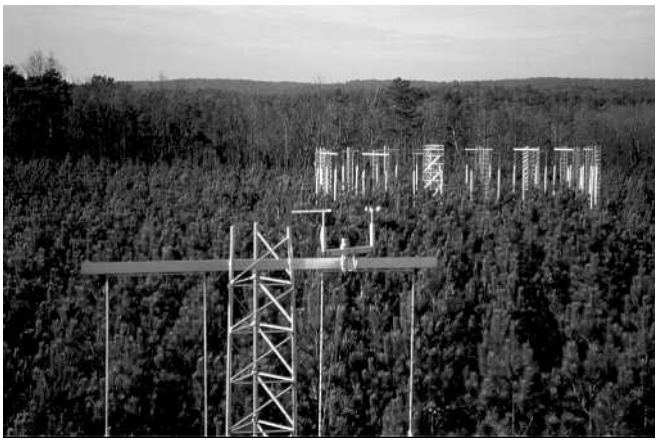
Some of the developments in global change research have been accompanied by technological innovation that has allowed measurements to be made over large regions of the globe, over the entire Earth for that matter. One such technological tool has been the satellite with the capacity to measure variables related to biospheric metabolism such as NPP and evapotranspiration (*see* **Global Ocean Observing System (GOOS)**, Volume 2; **Global Terrestrial Observing System (GTOS)**, Volume 2). The wide coverage and the high frequency of measurements has made it possible to track the phenology of plant cover (*see* **Phenology**, Volume 2) as well as historical changes in land surfaces.

New technology (e.g., eddy covariance, and free air carbon dioxide enrichment technology (FACE)) has also enabled us to manipulate and measure ecosystem quantities and processes at larger scales (dozens of m^2 to several hectares, and small watersheds) than previously in order to study the effects of present and future environmental conditions (Figure 4) (*see* **Long-term Ecological Research (LTER) System**, Volume 2). New knowledge resulting from observations and ecosystem experiments using these new technologies has been fundamental to increasing our understanding of ecosystem controls and the processes involved. This type of information is also the basis for the development and validation of biospheric models that simulate key functional and structural parameters of the Earth's ecosystems. *See*, for example, **Biome–BGC Ecosystem Model**, Volume 2.

The new Earth system science, fundamental to global change research, includes a large body of disciplinary work but the greatest challenge is in integrating research from multiple disciplines and developing appropriate linkages and constraints using both bottom-up and top-down approaches.

A good example of how such an integrated approach has been successful is the research done to explain the changes in interannual variability of atmospheric carbon dioxide concentrations. First, there is the global network of surface-based carbon dioxide monitoring stations, which provides the data on global trends in carbon dioxide concentrations. Then there is an integrated suite of top-down satellite data, plot-level bottom-up flux measurements (eddy covariance), and ecosystem experiments, which can explain why interannual variability occurs (Figure 5). The satellite-based normalized difference vegetation index (NDVI), for instance, might indicate earlier leaf-out in temperate and boreal forests. Ground-based measurements of leaf area document the phenomenon locally, and temperature measurements yield its probable cause. Eddy flux can then be used to measure the consequences of earlier leaf-out on the annual carbon balance for the ecosystem. Finally, ecosystem manipulative experiments (e.g., elevated carbon dioxide and temperature experiments) can provide the mechanistic information necessary to forecast changes under simulated global environmental change. Information on other intermediate-scale processes such as disturbances, succession and dispersion are also needed to understand the full range of controls on functioning and structure at multiple scales of space and time (*see also* **CO_2 Enrichment: Effects on Ecosystems**, Volume 2).

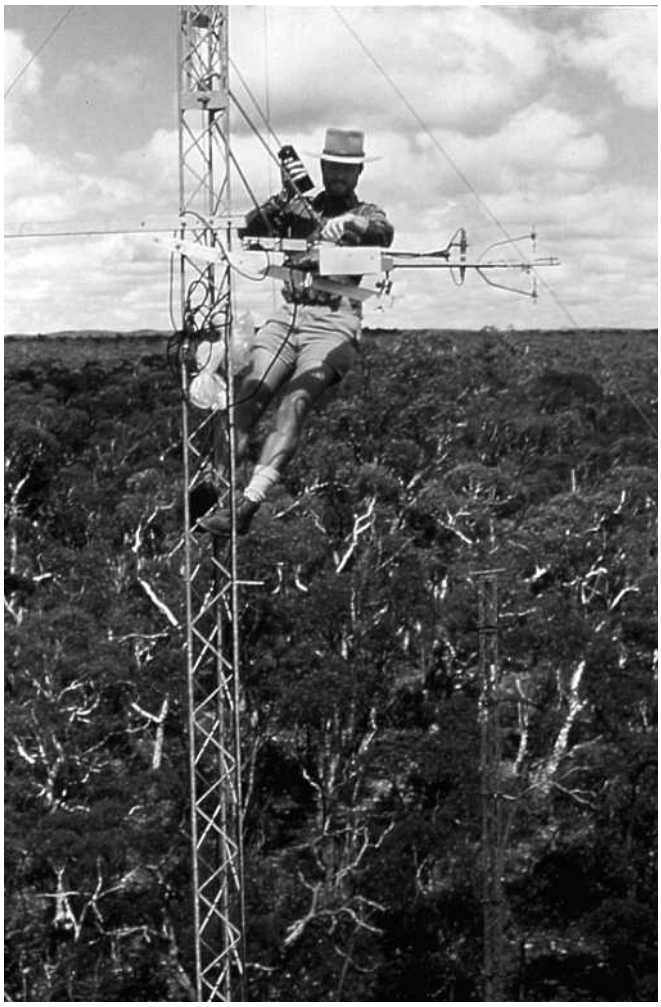
Global change research is also moving quickly towards better understanding of the nonlinear nature of the biosphere's behavior. Systems at any level of organization often display surprising and unexpected changes in response to continuous or sudden changes in the environment. Questions such as: Will NPP of terrestrial ecosystems change proportionally to changes in atmospheric carbon dioxide and temperature or are there thresholds, switch points or flip flops? Will the gradual deforestation of the tropics incrementally reduce biodiversity or are there fragmentation thresholds beyond which species extinction takes place more rapidly? The implications of nonlinear responses to global change are critical for the design of new research (*see* **Chaos and Cycles**, Volume 2; **Non-equilibrium Ecology**, Volume 2) and the development of suitable policies to deal with unexpected system behavior (*see* **Monitoring in**



(a)

(d)

(b)



(c)

(e)

Figure 4 Large-scale ecosystem manipulation experiments and observations: (a) FACE facility at Duke University, Durham, North Carolina, USA (photo: Pep Canadell); (b) ecosystem warming experiment using a passive nighttime warming roof in Denmark (CLIMOOR) (photo: Claus Beier); (c) Throughfall Displacement Experiment (TDE) in the Walker Branch Watershed temperate forest study area, Oak Ridge, Tennessee, USA (photo: Pep Canadell); (d) Elevated CO_2 open top chambers in Jasper Ridge, Stanford University, California, USA (photo: Pep Canadell); (e) measuring net ecosystem exchange of carbon dioxide using a eddy covariance system, Australia (photo: Greg Heath, CSIRO)

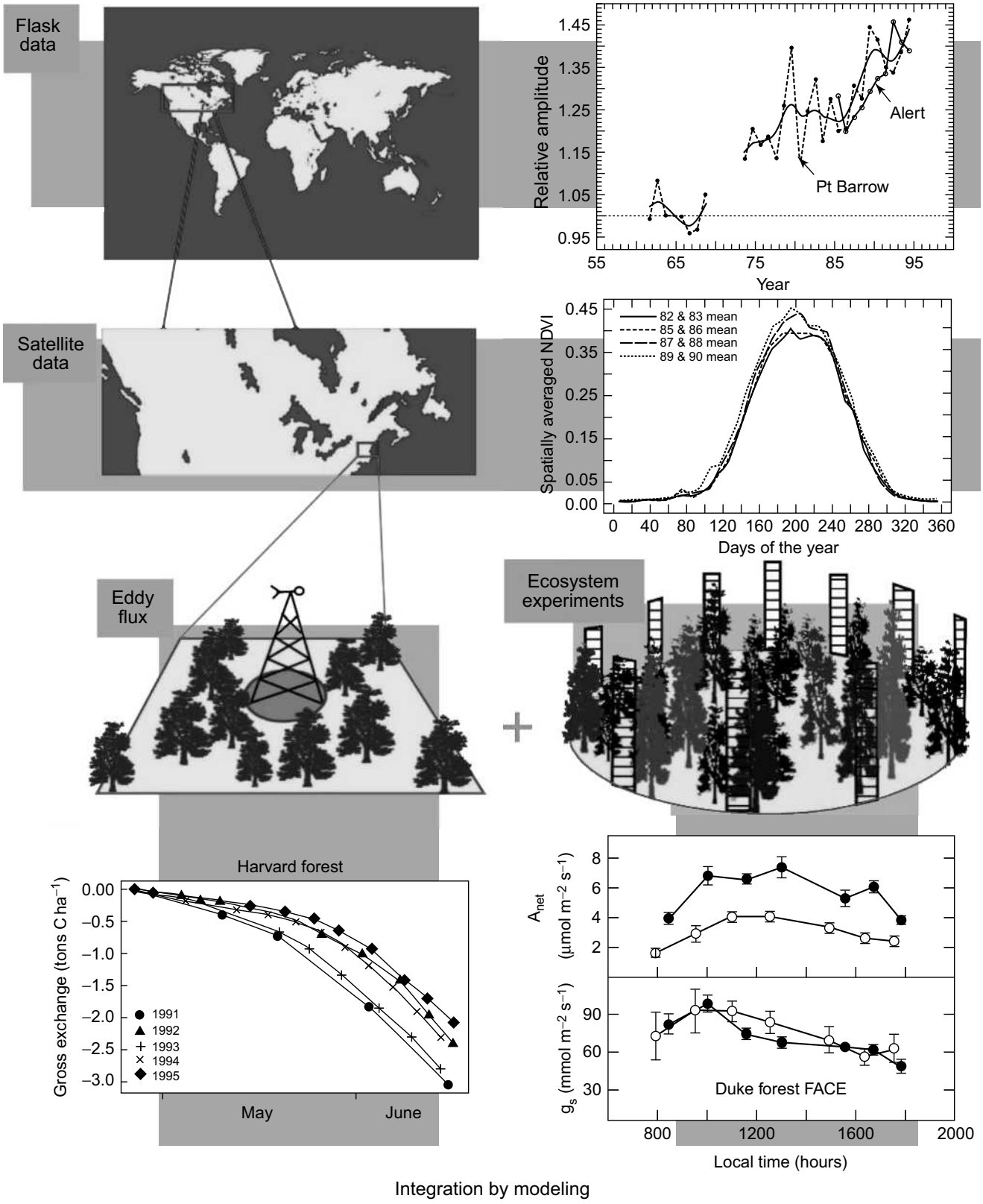


Figure 5 Measuring the carbon metabolism of the biosphere with a multi-technique approach (Canadell *et al.*, 2000)

Support of Policy: an Adaptive Ecosystem Approach, Volume 4).

Global change research also encompasses an understanding of the richness of the world's biota and its role in ecosystem functioning and stability. The diverse assemblages of organisms that dwell in the various components of terrestrial and aquatic systems perform critical functions such as decomposition of plant and animal detritus, nitrogen fixation, and a myriad of other essential functions. However, a large fraction of living species still remains unknown, particularly in oceans and soils (e.g., only 0.025–5% of nematode species have been described). This lack of information on the elemental constituents of an ecosystem poses large impediments to further understanding the role of biota on the functioning of the world's ecosystems. As loss of species continues, one could ask how many species are needed to maintain appropriate functioning and long-term stability of an ecosystem. In other words, from a functional point of view, how many species can we afford to lose? This challenge becomes even more complex when adding the rapid redistribution of the world's biota due to the increasing levels of material flows among many parts of the world, that is, the expansion of biological invasions (*see* **Biological Invasions**, Volume 2). Thus, a main challenge is the need for the development of new science and testable hypotheses linking biodiversity and ecosystem functioning (*see also* **Biodiversity in Soils and Sediments: Potential Effects of Global Change**, Volume 2; **Functional Biodiversity**, Volume 2).

CONCLUDING REMARKS

This volume contains a series of articles on the basic science that underpins current understanding of the functioning and structure of the biosphere and how it interacts with the physical environment. This knowledge allows us to evaluate the extent that the effects of global environmental change will have on the world's ecosystems and provides information on the development of concepts, theory, and tools that are used for simulating future trajectories under various scenarios of global change.

When individual species are grouped together into populations and communities, new processes shape their composition and structure such as herbivory, pollination, competition, succession, nitrogen fixation, and a critical understanding of life cycles as nonlinear and often chaotic systems.

The next level of biological organization and complexity concerns ecosystems with assemblages of multi-species, above and below ground interactions, and spatially explicit resource distributions. Contributions in this volume that provide understanding of ecosystem structure and functioning include: mineral cycling, disturbances, resilience, ecosystem invasions, and carbon and energy storage and fluxes. The volume also includes the latest understanding on current and future effects on the biosphere of increasing atmospheric carbon dioxide, nitrogen deposition, and rising air and soil temperatures.

Landscapes are the most common scale at which humans perceive and manage ecosystems. They comprise a mosaic of patches reflecting physical constraints, natural disturbance history, and land use patterns. This heterogeneity alone necessitates the use of approaches that account for the diversity of land surface properties and resulting fluxes, in addition to basic processes involved in landscape dynamics such as disturbances and dispersal, and dynamics at the interfaces (ecotones) in fragmented landscapes.

Landscape richness and distribution of species also have implications for the structure and functioning of ecosystems. Information on species richness and functional attributes is provided for some of the less studied components of the biosphere such as soils and freshwater sediments. Species assemblages sometimes change rapidly due to species extinctions and invasions of exotic species. Several contributions in this volume provide critical information on past, present, and expected future species extinctions with an analysis of what makes species most vulnerable. For individual plant and animal species, information on physiology, growth, and reproduction is also provided as key controls of structure and functioning.

The volume also reviews the major characteristics of a number of representative biomes across the world's climatic belts and their sensitivity to global environmental change (e.g., tundra is very sensitive to temperature changes). Special attention is given to some of the most vulnerable systems within biomes such as intertidal communities and coral reefs.

The highest level of complexity is the Earth system itself. This is the level at which some of the most fundamental life engines operate, particularly the various global biogeochemical cycles; carbon, nitrogen, iron. It is also the level at which major biosphere–atmosphere feedbacks take place with large impacts on regional and global climate.

The development of new tools has been critical to the advancement on global change research. The need to scale up to large regions and to the globe, and to project future trajectories of ecosystem functioning and structure has been a challenge. This volume describes some of these advances ranging from remote sensing techniques, ecosystem flux measurements (e.g., eddy covariance), manipulation experiments (e.g., FACE), the use of stable isotopes, ecotrons and other experimental facilities, etc. The information generated from the use of these techniques directly feeds into the development and testing of ecosystem and biospheric models also included in this volume.

The new research issues dealing with global environmental change have fostered the development of a new

type of science, a new type of scientist, and a new way to structure and coordinate research. The volume describes some of the international programs that have been critical to the development of the science of global change such as IGBP (*see* **IGBP Core Projects**, Volume 2), and how other international and regional programs and networks have played critical role in shaping the research agenda over the last decade. Such as the European Network for Research in Global Change, the Inter-America Institute for Global Change Research, the Asian Pacific Network for Global Change Research, the International Long Term Ecological Research network, and the International Global Observing Systems. Examples of regional research projects are also described such as the Boreal Ecosystems Atmosphere Study (BOREAS) and the Large-scale Biosphere–Atmosphere experiment in Amazonia (LBA) for their importance at pioneering integration within a large region (*see* **Boreal Ecosystem–Atmosphere Study (BOREAS)**, Volume 2; **European Network for Research in Global Change (ENRICH)**, Volume 2; **Inter American Institute for Global Change (IAIGC)**, Volume 2; **Large Scale Biosphere–Atmosphere Experiment in Amazonia (LBA)**, Volume 2; **Long-term Ecological Research (LTER) System**, Volume 2).

REFERENCES

Canadell, J G, Mooney, H A, Baldocchi, D D, Berry, J A, Ehleringer, J R, Field, C B, Gower, S T, Hollinger, D Y, Hunt, J E, Jackson, R B, Running, S W, Shaver, G R, Steffen, W, Trumbore, S E, Valentini, R, and Bond, B Y (2000) Carbon Metabolism of the Terrestrial Biosphere: a Multi-technique Approach for Improved Understanding, *Ecosystems*, **3**, 115–130.

Cohen, A N and Carlton, J T (1998) Accelerating Invasion Rate in a Highly Invaded Estuary, *Science*, **279**, 555–558.

Crutzen, P (2000) The Anthropocene, *IGBP Newsletter*, **41**, 17–18.

Vitousek, P M, Mooney, H A, Lubchenco, J, and Melillo, J M (1997) Human Domination of Earth's Ecosystems, *Science*, **277**, 494–499.

FURTHER READING

Canadell, J and Noble, I (2000) Changing Metabolism of Terrestrial Ecosystems under Global Change, *Ecol. Appl.*, **10**, 1551–1632.

Houghton, R A (1998) Historic Role of Forests in the Global Carbon Cycle, in *Carbon Dioxide Mitigation in Forestry and Wood Industry*, eds G H Kohlmaier, M Weber, and R A Houghton, Springer-Verlag, Berlin, 1–24.

Hughes, L (2000) Biological Consequences of Global Warming: is the Signal Already Apparent? *TREE*, **15**, 56–61.

Jackson, R B, Schenk, H J, Jobbagy, E G, Canadell, J, Colello, G D, Dickinson, R E, Dunne, T, Field, C B, Friedlingstein, P, Heimann, M, Hibbard, K, Kicklighter, D W, Kleidon, A, Neilson, R P, Parton, W J, Sala, O E, and Sykes, M T (2000) Belowground Consequences of Vegetation Change and its Treatment in Models, *Ecol. Appl.*, **10**, 470–483.

Mooney, H A (1998) The Globalization of Ecological Thought, in *Excellence in Ecology no. 5*, ed O Kinne, Ecology Institute, Oldendorf/Luhe, Germany.

Walker, B H, Steffen, W L, Canadell, J, and Ingram, J S I (1999) *The Terrestrial Biosphere and Global Change, Implications for Natural and Managed Ecosystems*, Cambridge University Press, London.

Biological Invasions

W M LONSDALE

CSIRO Entomology, Canberra, Australia

The world's natural and agricultural ecosystems are being invaded by a growing number of destructive species. Such invasive species are important agents of global change (Vitousek et al.*, 1996; Williamson, 1996; Simberloff, 1997a), and ecologists now believe that biological invasions are second only to habitat destruction as a threat to biodiversity (Walker and Steffen, 1997). Invasive weeds reduce agricultural productivity and out-compete the native ora, feral animals consume and degrade the resources that support our native animals and livestock, and insect pests and diseases threaten our natural vegetation and crops. Increasing numbers of exotic marine pests are invading our sheries and coastal environment. There are major impacts on animal health from exotic diseases, and on human health from allergies due to weeds. One of the main drivers of change through invasion has been trade. World trade grew 16-fold between 1950 and 1998 (The Economist, 1998), and although increasing globalization of trade represents great economic opportunities, there are increased risks to biodiversity from increasing movement of pests, weeds and diseases.*

DEFINITIONS

For Williamson (1996, 58–59), an *invader* or *invasive species* is any non-indigenous species – thus, it is a purely geographical term. Here, however, the terms *invader* and *invasive* are reserved for non-indigenous species that are in the process of increasing in range (Pysek, 1995). A further useful term is *cryptogenic*, defined as species that are neither clearly indigenous nor non-indigenous, and which may in some communities far outnumber the non-indigenous species (Carlton, 1996).

From a management point of view, it is useful to classify non-indigenous species by their population status. In ascending order of population size and stability (see Table 1), exotic species may be *adventive*, *naturalized*, or *invasive*. Each of these status groups has its counterpart for indigenous species.

A further classification, applying to both indigenous and non-indigenous species, is by function (*colonizers*, *dominants*) or by impact (*pests* or *weeds*). It is likely that most invasive species as defined above (i.e., non-indigenous species undergoing a range expansion) will have an economic or ecological impact. Indigenous species can cause damage, but a disproportionate amount of damage is caused by non-indigenous species, and they are major agents of global change. For example, in Australia, many of the pests and most of the weeds are non-indigenous.

Some other terms used in invasion studies include:

- *Disturbance*: removal of competing vegetation (Hobbs, 1991). Often regarded as the precursor to invasion.
- *Native species resistance to invasion*: competitive ability of native species with respect to invaders.
- *Resistance to disturbance*: ability of native ecosystems or species to recover from disturbance – often regarded as the inverse of invasibility.
- *Invasibility*: susceptibility of sites to invasion (e.g., Lonsdale, 1999).
- *Ecosystem resistance to invasion*: intrinsic resistance of native ecosystem to invasion through community structure (Williamson, 1996, 193–196). Probably the inverse of invasibility.
- *Invasion potential*: intrinsic ability of species to invade (di Castri, 1989).
- *Propagule pressure*: number of propagules arriving at a site (Williamson, 1996, 45).

THE INVASION PROCESS

Modes of Introduction

Many invasive species arrive as intentional introductions that then go awry. The proportion of harmful species that were introduced intentionally varies, however, between taxa. For weeds, 49% of Australian noxious weeds (Panetta, 1993) and at least 40% of invasive plants in North

Table 1 Population status of indigenous and non-indigenous species, and the words used to describe them

Population status	Indigenous	Non-indigenous (exotic, alien)
Uncommon, survival uncertain	Endangered	Adventive, casual, escaping
Stable	Stable	Naturalized
Range increasing	Expanding	Invasive
Capable of establishing in sparsely occupied habitats	Colonizers	
Most abundant species in community	Dominants	
Causing damage	Pests and weeds	

Eastern US (Mack and Lonsdale, 2001) were introduced intentionally. Similarly, most harmful fishes and terrestrial vertebrates in the US were intentional introductions (OTA, 1993, Figures 2–5). Invasive plants have been brought in as ornamentals, as pasture species, for erosion control, for afforestation, and as curiosities by botanic gardens (Mack and Lonsdale, in press). Mammals and fish are introduced for hunting and fishing. Although many are intentionally introduced, accidental introduction of fish, terrestrial vertebrates and plants also occurs.

By contrast, plant pathogens, mollusks and insect pests were mostly accidentally introduced (OTA, 1993, Figures 2–5). Pathogens and insect pests can arrive borne on air currents, or as unwanted passengers on imported plant material. Forest pests arrive on unprocessed timber imports. Mollusks frequently arrive in ballast water in ships, to be evacuated at the port of entry.

Rates of Incursion

Data on rates at which organisms become naturalized are quite scanty. For a range of organisms (Figure 1), the rate of establishment of species for the US, Australia and New Zealand varies between 16 species per annum (p.a.) (plant pathogens in Australia) to 0.5 species p.a. (fish in the US).

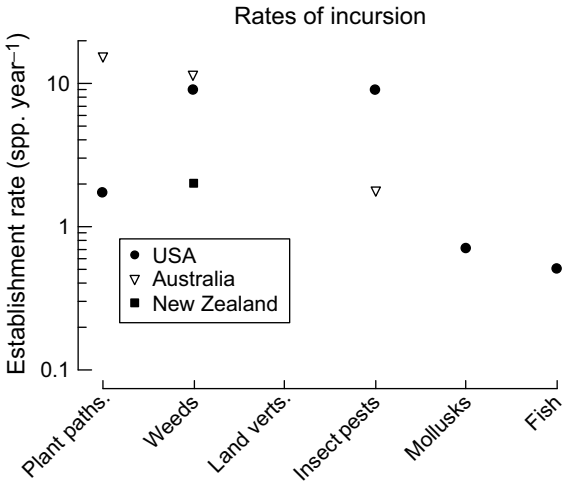


Figure 1 Rates of incursion of various exotic organisms (Plant paths are plant pathogens, Land verts are land vertebrates) into the USA, Australia and New Zealand (recalculated from data in OTA, 1993, Nairn *et al.*, 1996 and Buddenhagen *et al.*, 1998 respectively)

Lag Phases

Classically, invading populations have been characterized by a sigmoidal curve growth curve. This involves a delay, known as the lag phase, in the onset of the invasion process, followed by a phase of exponential increase until the system is saturated with the invader and the population size stabilizes. Cousens and Mortimer (1995, 25–27) have shown that the early stages of an exponential increase would, in many cases, be indistinguishable from a lag phase. Nevertheless, there are undoubtedly real cases of the phenomenon, e.g., the Brandenburg exotic ora (Kowarik, 1995), where the average time lag between introduction and invasion was 147 years. How general lag phases are for different organisms, and why they occur, remains the subject of conjecture. It seems, by contrast, that marine organisms have little lag phase, and that once introduced, rapidly move to pest status, with little wastage of species (J Carlton, personal communication, see section on rates of conversion to pest status herein).

Rates of Spread

Invaders exhibit a huge range of spread rates, over six orders of magnitude (Mack and Lonsdale, 2001). Most of this variation is found amongst invertebrates, which range in velocity of spread from 260 km year^{-1} (the Africanized honey bee in America; measured from Figure 6.2 of Williamson, 1996) down to 3 m year^{-1} for an introduced earthworm in Ireland (*Aporrectodea rosea*; Wellings, 1994). The highest spread rates are exhibited by diseases (Mack and Lonsdale, 2001) with in uenza the fastest at 1200 km year^{-1} (Williamson, 1996, 155). Amongst plants, tree spread rates are very much lower than those of herbs. This might be attributable to the greater maturation time of trees, though this perhaps cannot entirely account for

the two orders of magnitude of difference seen. Marine organisms are among the higher spread rates (Mack and Lonsdale, 2001), though not as high as might be expected given the velocity of surface currents (Grosholz, 1996). A considerable amount of effort has gone into the prediction of spread rates in invading organisms (see Shigesada and Kawasaki, 1997 for a review). However, analyses of the spread of invading organisms frequently use simple reaction–diffusion models that ignore the typically leptokurtic dispersal of propagules and tend to grossly underestimate the potential rate of spread (Kot *et al.*, 1996). Kot *et al.* (1996) showed that the rate of spread is extremely sensitive to the shape of the dispersal curve – for example, fat-tailed distribution curves can generate accelerating rates of spread. Similarly, Higgins and Richardson (1999) showed that a small proportion (0.001) of seeds moving long distances (1–10 km) can lead to an order of magnitude increase in predicted spread rate.

Rates of Conversion to Pest Status

Rates of conversion to pests can be termed *base rates* (Smith *et al.*, 1999) or *prevalences* (from epidemiology; see Lonsdale and Smith, 2001). Williamson (1996) and Williamson and Fitter (1996) have proposed a broad generalization, the *tens rule*, suggesting that roughly 10% of imported species become adventive, 10% of these become naturalized, and 10% of these become pests. Thus, about 0.1% of imported species become pests. The rule is useful as a reminder of the rarity of successful invaders, but in reality, insights will probably emerge from considering these rates as variable, and exploring the causes of variation.

The data analyzed so far (Williamson and Fitter, 1996) may in fact re ect the two aspects of the transitions. On the one hand, there are the transition rates between the pools of adventive, naturalized, and pest species and on the other, the ratio of the sizes of the pools. These are obviously two very different things. There is no reason why the transition rates should center on 10%, and indeed one would expect them to vary greatly, especially with the degree of assistance given to the arriving species (Simberloff, 1989; Williamson and Fitter, 1996). Data to test this critically are scarce because we can rarely follow a cohort of species from start to finish through the invasion process. However, for the biocontrol insects analyzed by Williamson and Fitter (1996), the rates are in fact quite high (averaging about 35% but varying between 20–60%).

On the other hand, the ratios of the sizes of the pools may possibly re ect the typical hierarchy of species abundance seen in communities (i.e., there are large numbers of less abundant species and a few very abundant ones). It is possible that these ratios may perhaps have predictable tendencies for different communities (cf. Sugihara, 1980), but again the data are lacking to test this idea. However, the ratio of the number escaping to the number imported is unlikely to exhibit any natural regularity at all, since it would be very dependent on the number imported, which could have almost any value.

INVASIBILITY

The idea that some communities are more invasible – i.e., more intrinsically susceptible to invasion – than others, has exercised the minds of invasion ecologists since Elton's classic book (Elton, 1958). Others have emphasized propagule pressure as the overriding factor in determining the degree to which systems are invaded, rather than the properties of ecosystems themselves (e.g., Newsome and Noble, 1986).

Analyses of global patterns of plant invasions (Lonsdale, 1999) showed that the Old World is less invaded than the New. Contrary to expectation, communities richer in natives had more, not fewer, exotics, while the temperate zone generally was more invaded than the tropics. While islands are more invaded, this is apparently not because of low native species diversity as the islands were no less rich in natives than mainland sites of similar size. Amongst continents, North America ranked as most invaded and Africa least, while temperate forest was the most invaded biome and savanna the least. However, there was considerable within-group variation in the mean degree of invasion. Incomplete as the data are for plants (little having been published on South American and Asian ecosystems), the data for other taxa are still less adequate.

However, it is difficult to conclude anything about relative invasibilities from this. This is because the number of exotics in a region must be the result of a ux between immigration and extinction. Comparisons between regions of invasibility, are, in fact, impossible to make, unless one can control for all the other variables that in uence exotic richness, including the rates of immigration of species, and the characteristics of the invading species themselves (Simberloff, 1989; Williamson, 1996; Lonsdale, 1999). This suggests that a true understanding of invasibility is likely to come only from experimentation, or from comparing rates at which groups of organisms become pests in different environments.

EXPLANATIONS FOR THE SUCCESS OF INVADERS

The asymmetry of transfer of weed species between the Mediterranean basin and the rest of the world is an interesting case study to explore the reasons for the success of invaders relative to native species. The five regions of the world sharing a mediterranean climate (Mediterranean basin, South Africa, southern Australia, California and Chile) have exchanged many weed species. There

was a primary invasion of mostly annual weeds from the Mediterranean basin to the other four regions, and a subsequent secondary invasion of woody species between the four regions and back to the Mediterranean basin (Fox, 1990). Overall, there has been a considerable asymmetry in this ux, with the species balance being outward from the Mediterranean (or at least from Europe) to the other four regions (di Castri, 1989; Fox, 1990; Groves, 1991). Even now, the degree to which the Mediterranean has been invaded is much less than that of the other regions. This has been attributed to a variety of causes including:

1. that the European species are inherently superior competitors to those elsewhere (di Castri, 1989);
2. that the European species had left behind their pests and diseases in the home range, and thus became superior competitors in moving overseas;
3. that the non-European ecosystems are more invasible (see section on Invasibility) or less resistant to disturbance than the European (di Castri, 1989);
4. that the package of grazing animals, plants, and diseases brought by immigrants from Europe proved, in combination, an unstoppable ecological juggernaut (Crosby, 1986);
5. that the propagule pressure was overwhelmingly outward from the Mediterranean;
6. that the invaders were ooding into vacant niches in the new range, especially the spring growing Mediterranean annuals moving into Australia and South Africa, which had largely summer-growing annuals (Fox, 1990);
7. that the Mediterranean species were in reality an assemblage of all-conquering cryptogenic (for meaning of this word, see line 6 of section on DEFINITIONS) species that had previously invaded the Mediterranean and were now moving on to overwhelm new territories.

These arguments have all been variously applied to other situations to explain the superior performance of invaders. It is nearly impossible to test the relative merits of these hypotheses critically. The fact that invasions are now occurring back towards the Mediterranean basin (Fox, 1990) demonstrates that the other regions' plants are not necessarily all ecological wimps. Moreover, there seems little doubt that propagule pressure was largely outward with the vast post-Columbian exodus of peoples and agricultural technologies, and could explain much of the observed variation between regions.

INVASIONAL MELTDOWN

Does the rate of acquisition of new species accelerate as the ecosystem becomes more invaded (Simberloff and von Holle, 1999)? The answer to this is likely to be scale dependent (at the local scale, saturation will occur, leading to deceleration, while at the geographical scale, no feedback will occur). It is not yet clear how many harmful invasive species can be tolerated in a region before the system collapses – if indeed this will happen – in a process of invasional meltdown (Simberloff and von Holle, 1999).

INTERACTIONS WITH OTHER AGENTS OF GLOBAL CHANGE

Climate change, nitrogen deposition, and elevated carbon dioxide concentration are all likely to favor range expansion of some invasive species while causing contraction of others, and it is not clear whether this will, in the end, increase or reduce the global pool of invasive species. However, the most important driver of global change – habitat destruction and land-use change – will probably increase the susceptibility of ecosystems to invasion. Thus, the interaction between invasive species and habitat destruction is likely to be positive.

WHAT ARE THE CONSEQUENCES OF BIOTIC INVASIONS?

The impact of an invader can be measured at five levels: (1) the phenotypic characteristics of individuals; (2) the genetics of populations; (3) population abundance and dynamics; (4) the structure and function of communities; and (5) abiotic ecosystem processes (Parker *et al.*, 1999). The least studied impacts are those of genetic changes and long-term evolutionary effects (Parker *et al.*, 1999), but most effects of invaders are not well documented, perhaps because of the inherent difficulties and expense of undertaking environmental assessment. These have been summarized by Underwood (1994). The two paramount problems in designing sampling programs to measure ecological impacts (including those of invasions) are that there is a large temporal variance in many populations, and that many populations show a marked lack of concordance in their temporal trajectories from one place to another (Underwood, 1994). It is clear that any invasion impact study will require substantial resources if it is to be authoritative.

Components of Impact

Overall impact I can be defined as the product of the range size of the species (R), its average abundance per unit area across that range (A), and the impact per individual (E), such that:

$$I = R \times A \times E \quad (1)$$

Clearly, all three of these entities are likely to vary over orders of magnitude between species, and have potentially very large associated errors of estimation (see also Parker *et al.*, 1999). *R* is probably the easiest component of overall impact to measure, and has potentially the smallest estimation error. *A* is more problematic, having a fairly wide range and a large estimation error. Perhaps the most important variable to focus on, however, in terms of the interspecific range in values, and the potential for enormous errors in estimation, is *E*. Although we lack data on *E*, we can draw on food web studies for insights. Studies where large numbers for food web links have been assessed (e.g., Wooton, 1997) suggest that interspecific interaction strengths in communities may vary over six or seven orders of magnitude. The variation for invasive species may be rather smaller (though we, as yet, have no data to test this) but will still be substantial: compare the likely *per capita* impact (not the impact across the region) of a herbaceous weed like purple loosestrife with that of an invasive tree like *Melaleuca*. So, if we are ever to be able to compare overall impact across a range of species, it seems likely that we must understand how *E* varies between species, as well as its presumably complex relationship with *R* and *A*.

Thresholds

Are there threshold densities below which an invasive organism is effectively harmless? Panetta and James (1999) conclude that, while thresholds may exist for invaders of natural ecosystems, experience from weeds in cropping systems suggests that even where damage thresholds do exist, they are of limited value for management purposes. Furthermore our ability to quantify thresholds in natural ecosystems is limited. Panetta and James argue instead that control measures for invaders of natural ecosystems are best determined by their cost and efficacy, rather than by waiting for some damage threshold to be measured (Panetta and James, 1999).

Environmental Consequences

Invasions that have the worst impacts are those that alter ecosystem function, alter the disturbance regime, or that facilitate other *invaders* (for a table of case studies of invasion impact, see http://www.ento.csiro.au and search on invaders).

Alteration of Ecosystem Function

The invasion of Hawaii Volcanoes National Park by *Myrica faya*, a small tree native to the Canary Islands, is transforming an entire ecosystem by fixing nitrogen in nitrogen-poor volcanic soils at a rate 90 fold greater than that added by native plants and over four times the rate at which precipitation adds nitrogen. Nitrogen fixed by *M. faya* is available to other non-indigenous plants, which favor fertile soils. Thus, *M. faya* is altering the course of succession in these communities (Vitousek *et al.*, 1987).

Ecosystem effects usually have economic consequences. The invasion of alien plants, especially acacias, pines and eucalypts, into South African native ecosystems, has caused loss of water (estimated at 3×10^9 m^3 year^{-1}; van Wilgen and van Wyk, 1999) from catchments through excessive transpiration rates, and the consequent loss of ow in rivers throughout the country, with many rivers ceasing to ow at all, or owing only infrequently. This has caused a loss of agricultural production as well as risking the loss by extinction of species from the unique Cape Flora.

Alteration of Disturbance Regime

It is argued that invaders that alter the disturbance regime are the most serious (Mack and D'Antonio, 1998), because they alter the rules of the game for all species in the ecosystem. Many invasive species cause change in ecosystems by fostering more frequent or intense fires, or both, to which key native species are not adapted. The paperbark tree has this effect in Florida (Schmitz *et al.*, 1997), as do numerous grasses worldwide (D'Antonio and Vitousek, 1992). In northern Australia, the native savanna vegetation has an annual grass understory that is being replaced by African perennial grasses of much higher biomass. These cure later in the dry season than the annual grasses they replace, thus promoting more intense fires. Although the native vegetation has evolved to cope with frequent fires, it is not certain that the increase in intensity is sustainable.

Facilitation

Given an invasion with *facilitation* (where one invader facilitates the invasion of another; see Simberloff and von Holle, 1999), the rate at which a species can swamp and transform an ecosystem can be dramatic. *Mimosa pigra* (mimosa), a tropical American shrub that invaded northern Australian wetlands in the late 1970s, expanded from a few hectares (ha) to an infestation that occupied about 30 000 ha of the Adelaide River oodplain between 1978 and 1984 (Lonsdale, 1993), and subsequently about 80 000 ha across the Northern Territory's wetlands. The expansion was all the faster because of facilitation by the Asian Water Buffalo, which had undergone a population explosion since 1970, and was overgrazing the native vegetation, reducing competition for mimosa seedlings and accelerating establishment. When the buffalo were largely removed, the rate of creation of new mimosa outbreaks slowed markedly (Lonsdale, 1993).

Economic Consequences

It has been estimated that invasive species in the USA cause economic damage of $138 billion p.a. (Pimentel *et al.*,

2000), roughly 2% of the gross domestic product. Some examples of economic costs from around the world include:

- Weeds cost Australian agriculture $3 billion p.a. in 1995 (Anon, 1999).
- Feral pigs cause $800 million in damages in the US annually (Pimentel *et al.*, 2000).
- $1.3 million was spent to control feral cats on Prince Edward Island, South Africa (Huntley, 1996).
- Estimated US losses to Russian wheat aphid from 1987–1989 exceeded $600 million (OTA, 1993).
- The estimated damage to livestock, wildlife and public health in Texas caused by the fire ant *Solenopsis invicta* is $300 million p.a. (Allen *et al.*, 1995).
- Bovine tuberculosis causes losses in the order of $63 million p.a. to Latin American countries (Cosivi *et al.*, 1998).

HOW ARE INVASIONS CONTROLLED?

Quarantine Methods

The ability of a nation to restrict the movement of invaders across its borders is governed by international treaties, of which a key agreement is the agreement on the Application of Sanitary and Phytosanitary Measures (SPS agreement; Anon, 1994). This agreement allows World Trade Organisation (WTO) members to take measures to restrict movement of species that may pose a threat to human, animal or plant life (Anon, 1994). The International Plant Protection Convention (IPPC) of 1951 deals with plant quarantine, and the IPPC secretariat coordinates phytosanitary standards (Anon, 1994). The SPS agreement requires WTO members to base any SPS measures on internationally agreed guidelines. Within these guidelines, countries like Australia have operated quarantine controls that have taken an innocent-until-proved-guilty approach, e.g., allowing new plant species in unless they are known weeds. This approach has been subjected to con icting pressures in recent years, with a movement to liberate trade and remove non-tariff trade barriers seeking to ease quarantine controls, and an opposing body of opinion that argues that the precautionary principle should apply, and that a guilty-until-proven-innocent approach should be used in quarantine protocols. The Australian Quarantine Inspection Service in 1997 adopted a risk assessment system for screening new plant imports based on their biological attributes and the consequent risk of invasiveness that they pose (see http://www.agric.wa.gov.au/progserv/plants/weeds/Weedsci.htm).

However, prediction of successful invaders from their biological attributes either has low predictive power (e.g., Perrins *et al.*, 1992) or, if high, has more limited applicability (e.g., to specific regions or taxa; Rejmanek and Richardson, 1996). In general, there is a tension here between quarantine risk assessment procedures that must be scientific, transparent, and of general application, supported by recent research that suggests that this may be achievable (e.g., Rejmanek and Richardson, 1996), and those who argue that the prediction of invasiveness will always be extremely difficult (Simberloff, 1989; Williamson, 1996).

Clearly, much research on accuracy of prediction remains to be done. In applying a risk assessment screening procedure as government policy, however, it is not just the accuracy that has to be considered. The rates at which species become naturalized, and the rate at which naturalized species become pests, are quite low. This low base rate for pests means that predictive power of any risk assessment has to be very high in order to reliably identify pests (Smith *et al.*, 1999). For example, a screening system of high accuracy (85%) will reduce the rate of ingress of pests from 1% of introductions to about 0.2%, but in doing so, it will exclude large numbers of non-pest species – 95% of the rejected species would not have been pests. This suggests that a quarantine system will risk causing intersectoral con ict between environmentalists desiring that the precautionary principle apply, and industries such as horticulture desiring free movement of species. It remains to be seen whether the effects of this degree of restriction on trade will be sustainable in the long run. It is also not yet known whether genetically modified organisms will be more or less predictable than unmodified organisms.

Pesticides

Pesticide trade and usage globally is growing – pesticide exports grew 28 fold between 1961 and 1997 (Food and Agriculture Organization, 1999). This is in itself an agent of global change – pesticides persisting in ground water, for example, had polluted 10% of wells in the US by 1991 (Environmental Protection Agency, 1991). Pesticides remain useful tools in the pest control armory, but research is often necessary to adapt them for use in natural ecosystems, since they are as a rule devised and tested for use in agro-ecosystems. Resistance of insect pests to insecticides emerged in the 1940s, while weeds started to show resistance to herbicides somewhat later, in the 1970s. The occurrence of pesticide resistance has increased the urgency to seek ways of minimizing chemical pesticide use. These include biological control, microbial pesticides, pheromones to alter pest insect behavior, novel technologies such as sterility vaccines for feral pests, and the general approach known as integrated pest or weed management (*see* Integrated Pest Management (IPM) in an Era of Global Environmental Change, Volume 4). On the other hand, enzyme systems that have evolved to break down pesticides are being extracted from pest organisms and adapted for use in bioremediation.

Biocontrol

Biocontrol is a form of planned invasion. It is an extraordinarily powerful tool when effective, particularly in its classical form. Recent successes have included the control of the oating fern Salvinia in southeast Asia (Room, 1990) and of water hyacinth in Lake Victoria. It has been used in the control of weeds, insect pests and vertebrates, and has a good reputation in Australia where strong legislation ensures a transparent release procedure for new agents in which a high degree of host specificity must be demonstrated (see Cullen and Delfosse, 1985). In the US, by contrast, there is no Federal statute that requires that biological control agents be reviewed before release, and the credibility of biological control in the US is under attack from environmentalists. This is despite the excellent quality of much of the science (e.g., McEvoy *et al.*, 1991), and the fact that many of the practitioners regard themselves as environmentalists reducing the use of pesticides. Well-publicized examples of off-target damage (e.g., Louda *et al.*, 1997) have added to the pressure. Testing of post-release host expansion, to verify that the host specificity claimed in pre-release studies has been maintained in the field, has been little studied in the US, and indeed anywhere in the world. This should be redressed as an aid to future risk assessments.

Eradication

Eradication of a non-indigenous species is sometimes feasible, particularly if it is detected early and resources can be applied quickly. However, many regulatory agencies tend to view a non-indigenous species as not worth the bother and expense of attempting to control, until it becomes so widespread that eradication is probably not an option (Simberloff, 1997). There are examples of successful eradication programs, however – one such was the recent program against the introduced black-striped mussel in Darwin harbor, in northern Australia. The infestation was small, and was discovered early, was contained within a marina, and control measures had strong government support because of the risk to fisheries posed by the pest. All these factors favored success. Eradication using chemicals succeeded, at a cost of A$1.5 million (R Thresher, personal communication).

Strategic Approach to the Control of Invaders

It is being increasingly recognized that a strategic approach is needed for the control of invaders (e.g., Luken and Thieret, 1997). Much pest control thinking, by contrast, has been short-term and limited to the immediately visible problems. The underlying philosophy of a pest management strategy should be to establish why pests are ourishing in a region, and address the underlying causes rather than just killing pests *per se*. System management, rather than species management, is the focus. For example, in a region where weed species are ourishing in multi-species stands, the introduction of control measures against one species only may be a waste of resources, unless there is a strategy in place for dealing with the remaining species. It is an approach that is particularly appropriate for conservation areas, but is seldom undertaken. Related concepts (generally used in agricultural systems) are integrated pest management (IPM) or integrated weed management, which involve combining control measures, crucially, with a knowledge of the pest or weed's ecology to reduce pesticide use and maximize impact of control. It is claimed that IPM measures have reduced pesticide use in Indonesia by 60% and increased rice yields by 40%, for example.

The strategic approach to the control of invaders has begun to be used at the national level. Australia, for example, has recently adopted a national weeds strategy aimed at reducing the impact of weeds (Anon, 1999). Similarly, in a project of extraordinary scale, South Africa has decided to clear the exotic woody weeds from its river catchments in a 20-year program (van Wilgen and van Wyk, 1999). The strategy involves manual clearing of thickets to allow native vegetation to re-establish, treatment of cut-stumps with mycoherbicides, and the use of biological control to prevent reinvasion. Although this will cost R900 million, this is still far cheaper than the alternative, a massive dam-building program (R12 000 million; van Wilgen and van Wyk, 1999).

CONCLUSION

Biological invasions are major agents of global change, possibly second only to habitat destruction, through their impact on managed and unmanaged ecosystems. This impact is likely to intensify as conservation areas become fragmented, and as disturbance increases. Trade, agriculture and forestry all play a part in the intentional or unwitting movement of species globally, and regulating this movement is problematic. Control of invasives is often subject to market failure as it can be expensive and others than those carrying out the control measures often experience the benefits. This implies that governments must increasingly shoulder the burden for the assault on invasive species.

See also: Plant Dispersal and Migration, Volume 2.

REFERENCES

Allen, C R, Lutz, R S, and Demarais, S (1995) Red Imported Fire Ant Impacts on Northern Bobwhite Populations, *Ecol. Appl.*, 5, 632–638.

Anon (1994) *Agreement on the Application of Sanitary and Phytosanitary Measures, Annex 1*, World Trade Organisation, Geneva.

Anon (1999) *The National Weeds Strategy: a Strategic Approach to Weed Problems of National Signi cance*, revised edition, Agriculture, Forestry and Fisheries Australia, Canberra, ACT, Australia, Available on the web at http://www.affa.gov.au/docs/nrm/nws.html.

Buddenhagen, C E, Timmins, S M, Owen, S J, Champion, P D, Nelson, W, and Reid, V A (1998) An Overview of Weed Impacts and Trends, in *Department of Conservation Strategic Plan for Managing Weeds*, ed J Owen, New Zealand Department of Conservation, Auckland, New Zealand, 11–21.

Carlton, J T (1996) Biological Invasions and Cryptogenic Species, *Ecology*, 77, 1653–1655.

di Castri, F (1989) History of Biological Invasions With Special Emphasis on the Old World, in *Biological Invasions: a Global Perspective*, eds J A Drake, H A Mooney, F di Castri, R H Groves, F J Kruger, M Rejmanek, and M Williamson, John Wiley & Sons, Chichester, 1–30.

Cosivi, O, Grange, J M, Daborn, C J, Raviglione, M C, Fujikura, T, Cousins, D, Robinson, R A, Huchzermeyer, H F A K, de Kantor, I, and Meslin, F X (1998) Zoonotic Tuberculosis due to Mycobacterium bovis in Developing Countries, *Emerg. Infect. Dis.*, 4, 1–16.

Cousens, R and Mortimer, M (1995) *Dynamics of Weed Populations*, Cambridge University Press, Cambridge.

Crosby, A W (1986) *Ecological Imperialism: the Biological Expansion of Europe, 900- 1900*, Cambridge University Press, Cambridge.

Cullen, J M and Delfosse, E S (1990) Progress and Prospects in Biological Control of Weeds, in *Proc. 9th Australian Weeds Conference*, ed J W Heap, Crop Science Society of South Australia, Adelaide, South Australia, 452–476.

D'Antonio, C M and Vitousek, P M (1992) Biological Invasions by Exotic Grasses, the Grass/Fire Cycle, and Global Change, *Annu. Rev. Ecol. Syst.*, 23, 63–87.

Elton, C (1958) *The Ecology of Invasions by Plants and Animals*, Methuen, London.

EPA (1991) *Phase 1 Report National Ground Water Survey*, Springfield, VI, US EPA, PB 91–125 715.

FAO (1999) *Pesticides Trade Survey*, http://apps.fao.org/, accessed 1999.

Fox, M D (1990) Mediterranean Weeds: Exchanges of Invasive Plants Between the Five Mediterranean Regions of the World, in *Biological Invasions in Europe and the Mediterranean Basin*, eds F di Castri, A J Hansen, and M Debussche, Kluwer Academic Publishers, Dordrecht, 179–200.

Grosholz, E D (1996) Contrasting Rates of Spread for Introduced Species in Terrestrial and Marine Systems, *Ecology*, 77, 1680–1685.

Groves, R H (1991) The Biogeography of Mediterranean Plant Invasions, in *Biogeography of Mediterranean Invasions*, eds R H Groves and F di Castri, Cambridge University Press, Cambridge, 427–438.

Higgins, S I and Richardson, D M (1999) Predicting Plant Migration Rates in a Changing World: the Role of Long-Distance Dispersal, *Am. Nat.*, 153, 464–475.

Hobbs, R (1991) Disturbance a Precursor to Weed Invasion in Native Vegetation, *Plant Protection Quart.*, 6, 99–104.

Huntley, B J (1996) South African Experience Regarding Alien Species: Impacts and Controls, in *Proc. Norway/UN Conference Alien Species*, eds O T Sandlund, P J Schei, and A Viken, Ministry of the Environment, Trondheim, Norway, 182–188.

Kot, M, Lewis, M A, and van den Driessche, P (1996) Dispersal Data and the Spread of Invading Organisms, *Ecology*, 77, 2027–2042.

Kowarik, I (1995) Time Lags in Biological Invasions with Regard to the Success and Failure of Alien Species, in *Plant Invasions - General Aspects and Special Problems*, eds P Pysek, K Prach, M Rejmanek, and M Wade, SPB Academic Publishing, Amsterdam, 15–38.

Lonsdale, W M (1993) Rates of Spread of an Invading Species: Mimosa pigra in Northern Australia, *J. Ecol.*, 81, 513–521.

Lonsdale, W M (1999) Concepts and Synthesis: Global Patterns of Plant Invasions, and the Concept of Invisibility, *Ecology*, 80, 1522–1536.

Lonsdale, W M and Smith, C S (2001) Evaluating Pest Screening Procedures – Insights from Epidemiology and Ecology, in *Proc. First Int. Workshop Weed Risk Assessment*, eds R H Groves, J Virtue, and D Panetta, CSIRO Publishing, Melbourne, Australia.

Louda, S M, Kendall, D, Connor, J, and Simberloff, D (1997) Ecological Effects of an Insect Introduced for the Biological Control of Weeds, *Science*, 277, 1088–1090.

Luken, J O and Thieret, J W (1997) *Assessment and Management of Plant Invasions*, Springer-Verlag, New York.

Mack, C and D'Antonio, C M (1998) Impacts of Biological Invasions on Disturbance Regimes, *Trends Ecol. Evol.*, 13, 195.

Mack, R N and Lonsdale, W M (2001) Humans as Global Plant Dispersers: Getting More Than we Bargained for, *Bioscience*, 51, 95–102.

McEvoy, P B, Cox, C S, and Coombs, E (1991) Successful Biological Control of Ragwort Senecio jacobaea by Introduced Insects in Oregon, *Ecol. Appl.*, 1, 430–442.

Nairn, M E, Allen, P G, Inglis, A R, and Tanner, C (1996) *Australian Quarantine - a Shared Responsibility*, Department of Primary Industries and Energy, Canberra, Australia.

Newsome, A and Noble, I R (1986) Ecological and Physiological Characters of Invading Species, in *Ecology of Biological Invasions*, eds R H Groves and J J Burdon, Cambridge University Press, Cambridge, 1–20.

OTA (1993) *Harmful Non-indigenous Species in the United States*, OTA-F-565, US Government Printing Office, Washington, DC.

Panetta, F D (1993) A System of Assessing Proposed Plant Introductions for Weed Potential, *Plant Protection Quart.*, 8, 10–14.

Panetta, F D and James, R F (1999) Weed Control Thresholds: a Useful Concept in Natural Ecosystems? *Plant Protection Quart.*, 14, 68–76.

Parker, I M, Simberloff, D, Lonsdale, W M, Goodell, K, Wonham, M, Kareiva, P M, Williamson, M H, Von Holle, B, Moyle, P B, Byers, J E, and Goldwasser, L (1999) Impact: Toward a Framework for Understanding the Ecological Effects of Invaders, *Biol. Invasions*, 1, 3–19.

Perrins, J, Williamson, M, and Fitter, A (1992) Do Annual Weeds have Predictable Characters? *Acta Oecol.*, 13, 517–533.

Pimentel, D, Lach, L, Zuniga, R, and Morrison, D (2000) Environmental and Economic Costs of Non-indigenous Species in the United States, *Bioscience*, 50, 53–65.

Pysek, P (1995) On the Terminology Used in Plant Invasion Studies, in *Plant Invasions - General Aspects and Special Problems*, eds P Pysek, K Prach, and M W M Rejmanek, SPB Academic Publishing, Amsterdam, The Netherlands, 71–81.

Rejmanek, M and Richardson, D M (1996) What Attributes Make Some Plant Species More Invasive? *Ecology*, 77, 1655–1660.

Room, P M (1990) Ecology of a Simple Plant–herbivore System: Biological Control of Salvinia, *Trends Ecol. Evol.*, 5, 74–79.

Schmitz, D C, Simberloff, D, Hofstetter, R H, Haller, W, and Sutton, D (1997) The Ecological Impact of Non-indigenous Plants, in *Strangers in Paradise: Impact and Management of Non-indigenous Species in Florida*, eds D Simberloff, D C Schmitz, and T C Brown, Island Press, Washington, DC, 39–61.

Shigesada, N and Kawasaki, K (1997) *Biological Invasions: Theory and Practice*, Oxford University Press, Oxford.

Simberloff, D (1989) Which Insect Introductions Succeed and Which Fail? in *Biological Invasions: a Global Perspective*, eds J A Drake, H A Mooney, F di Castri, R H Groves, F J Kruger, M Rejmanek, and M Williamson, John Wiley & Sons, Chichester, 61–75.

Simberloff, D (1997a) The Biology of Invasions, in *Strangers in Paradise: Impact and Management of Non-indigenous Species in Florida*, eds D Simberloff, D C Schmitz, and T C Brown, Island Press, Washington, DC, 3–17.

Simberloff, D (1997b) Eradication, in *Strangers in Paradise: Impact and Management of Non-indigenous Species in Florida*, eds D Simberloff, D C Schmitz, and T C Brown, Island Press, Washington, DC, 221–228.

Simberloff, D and von Holle, B (1999) Positive Interactions of Non-indigenous Species: Invasional Meltdown? *Biol. Invasions*, 1, 21–32.

Smith, C S, Lonsdale, W M, and Fortune, J (1999) When to Ignore Advice: Invasion Predictions and Decision Theory, *Biol. Invasions*, 1, 89–96.

Sugihara, G (1980) Minimal Community Structure: an Explanation of Species Abundance Patterns, *Am. Nat.*, 116, 770–787.

The Economist (1998) Time For Another Round, *The Economist; World Trade Survey*, 3 October 1998.

Underwood, A J (1994) On Beyond BACI: Sampling Designs That Might Reliably Detect Environmental Disturbances, *Ecol. Appl.*, 4, 3–15.

Vitousek, P M, Walker, L R, Whiteaker, L D, Mueller-Dombois, D, and Matson, P A (1987) Biological Invasion by Myrica faya Alters Ecosystem Development in Hawaii, *Science*, 238, 802–804.

Vitousek, P M, D'Antonio, C M, Loope, L L, and Westbrooks, R (1996) Biological Invasions as Global Environmental Change, *Am. Sci.*, 84, 468–478.

Walker, B H and Steffen, W (1997) An Overview of the Implications of Global Change for Natural and Managed Terrestrial Ecosystems, *Conserv. Ecol. Online*, 1(2), www.consecol.org/vol1/iss2/art2.

Wellings, P W (1994) How Variable are Rates of Colonization? *Eur. J. Entomol.*, 91, 121–125.

van Wilgen, B W and van Wyk, E (1999) Invading Alien Plants in South Africa: Impacts and Solutions, in *Proc. VI International Rangeland Congress, Vol 2*, eds D Eldridge and D Freudenberger, Aitkenvale, Queensland, Australia, 566–571.

Williamson, M (1996) *Biological Invasions*, Chapman and Hall, London.

Williamson, M and Fitter, A (1996) The Varying Success of Invaders, *Ecology*, 77, 1661–1665.

Wootton, J T (1997) Estimates and Tests of Per Capita Interaction Strength: Diet, Abundance, and Impact of Intertidally Foraging Birds, *Ecol. Monog.*, 67, 45–64.

Functional Biodiversity

SHAHID NAEEM

University of Washington, Seattle, WA, USA

One of the many values of biodiversity, such as untapped genetic resources and aesthetic values, is biodiversity's role in regulating environmental conditions such as atmospheric composition, water quality, and soil fertility. The biota move enormous quantities of elements, compounds, aerosols, and particles through the earth's atmosphere, hydrosphere and lithosphere every year and in so doing, modify the chemical and physical conditions of Earth's surface, creating an environment that sustains life. All species contribute to this environmental conditioning, but each in different ways. Some species are critical to the functioning of an ecosystem while others play minor or intermittent roles. Current limited evidence suggests that intermediate levels of biodiversity yield more ef cient ecosystems than depauperate systems while even higher levels of biodiversity provide even greater versatility and predictability in ecosystem response to environmental change. Our current knowledge of the functioning of biodiversity suggests that increasing loss of biodiversity may qualitatively and quantitatively change the way ecosystems contribute to environmental processes. Elucidating the functional roles of species and their contributions to ecosystem processes will be fairly important to better understand the environmental consequences of biodiversity loss.

BIODIVERSITY, ECOSYSTEMS, AND BIOGEOCHEMISTRY

Gigatons of elements such as carbon (C), hydrogen (H), nitrogen (N), oxygen (O), phosphorus (P) and sulfur (S) (Schlesinger, 1997) and teragrams of aerosols and particles are annually cycled among the atmo-, hydro-, and lithospheres by an extraordinarily diverse assemblage of species (Figure 1). Some cycling occurs passively, by geochemical and geophysical processes, but a significant portion, known as *biogeochemical cycling* (Schlesinger, 1997), is done by earth's biota. The earth's ecosystems and the biota within them modify the planet's surface conditions through a rich variety of biogeochemical activities, whether it is the production of stratospheric ozone, the cycling of greenhouse gases, nutrient cycling, or the production of aerosols that affect climate. These activities are sometimes referred to as *ecosystem services* (Daily, 1997) because they directly impact human economies and human welfare (Costanza *et al.*, 1997).

The function of the earth's extraordinary diversity in these biogeochemical processes is an important but difficult question. Ecologists began addressing the ecosystem consequences of biodiversity decline in the 1990s, motivated in part by dramatic declines in global biodiversity (Schulze and Mooney, 1993; Heywood, 1995; Mooney *et al.*, 1996). Three functional roles of biodiversity have been identified. First, high levels of biodiversity mean that biogeochemical processes occur over a wide range of habitats. Second, higher numbers of species potentially provide greater short-term efficiency of biogeochemical processes if species use their resources in complementary ways. Third, higher biodiversity may ensure greater reliability or predictability of ecosystem response to disturbance. Each of these functions is explained in greater detail later in this article.

Scientists have identified over 50 hypothetical ways in which biodiversity loss may impact ecosystem functioning (Schlapfer and Schmid, 1999), but most of these revolve around three basic ideas. First, beyond some minimum number of species necessary to carry out basic ecosystem processes, most species are equivalent (redundant) and their loss is of little significance. Alternatively, reductions in biodiversity may have significant but unpredictable (idiosyncratic) effects on ecosystem functioning. The third, and perhaps most widely cited idea, is that loss of a few species (like loss of a few rivets in a large machine) may at first have no apparent affect, but if loss continues unabated, the system eventually fails to function, much the way a machine would fail if it lost too many rivets (Ehrlich and Ehrlich, 1981).

The functioning of biodiversity therefore concerns whether depauperate ecosystems are functionally different from

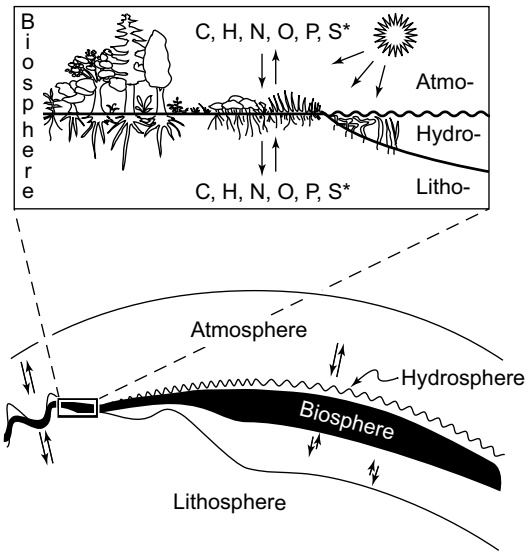


Figure 1 The role of the biota in biogeochemical processes. The bottom portion of the figure represents a cross section of the earth in which the atmo-, hydro-, and lithosphere are shown in relation to the biosphere. Note that this illustration is not to scale, (were the figure to scale the biosphere would be barely visible as a very thin line). The top part shows a section of the biosphere (shown as a rectangle on the lower figure) is expanded showing how the collective metabolic activities of the earth's biota move elements among the various spheres. This activity is referred to as biogeochemistry. The prefixes, atmo-, hydro-, and litho- refer to the major spheres illustrated in the bottom figure. Important organisms, such as microbes and animals, are not shown. Many other elements, including Fe, Mg, Ca, K, and more, are cycled but not illustrated

species-rich ecosystems. Large-scale human transformations of habitats into managed ecosystems have invariably led to the substitution of species-rich ecosystems with human-dominated, depauperate systems. For example, vast areas (about 99%) of North American prairie grasslands, once containing hundreds of plant species are now farms, pastures, urban areas, or other human-dominated ecosystems which contain relatively few species. Like prairies, crop farms and pastures are also temperate, herbaceous ecosystems, so these ecosystems may be similar in many respects to the prairies they replaced. Thus, the central issue is not so much the loss of species *per se* as it is the replacement of once species-rich ecosystems with modern species-poor systems.

To understand the possible consequences of such reductions in biodiversity over large stretches of habitat requires an understanding of the functioning of biodiversity. This article summarizes how scientists address this issue and reviews the functional classification of species and its use in understanding the functioning of biodiversity.

OBSERVATIONAL AND EXPERIMENTAL METHODS

Basic Approaches

Understanding the functioning of biodiversity involves the same approaches used to understand the functioning of any system's structure and parts. When faced with the task of understanding the functioning of a novel system, we begin by observing its activity and exploring its response to various manipulations. We then methodically disassemble the system and examine its parts and the relationship among its parts. If successful in our research, we can construct a blueprint for the structure of the system, explain its functioning and the role of its parts. We could also predict which parts are critical to functioning and how the system might respond to perturbations.

Ecologists approach ecosystems in much the same way. They begin by observing ecosystem activity, such as rates of energy ow and material cycling. They methodically explore the system to uncover the many species residing within the system and learn how these species interact with one another. They conduct experiments on the functional role of species by deleting certain species or groups of species and monitoring system response. Finally, they use this information from observations and experiments to predict how the ecosystem might respond to perturbation such as loss or addition of species or changes in environmental conditions. All these activities lead to the construction of a blueprint of the pattern of interactions among species and the ow of nutrients and energy through them.

Although the study of ecosystems and human engineered systems may seem similar, they are quite different in a number of respects. The first important difference is that parts and the relationship among parts in engineered systems generally re ect purposeful design. We therefore expect all parts to serve some function and our task is to identify that function. Purposeful design enormously simplifies the process of determining the functioning of a system and the functional role of its parts. Ecosystems, however, have no purposeful design. Species within ecosystems and the interactions among them are the result of evolution, biogeography, chance events, and other processes, none of which re ects deliberate design. Therefore, we should not expect that all species present serve roles in overall ecosystem functioning.

A second important difference is that engineered systems often exhibit some degree of economy or parsimony in design. Parts are seldom duplicated if they are costly, complexity is avoided, and efficiency of energy use is often maximized. Again, this human engineered economy of design simplifies analyzing functioning in machines. In contrast, there is no reason for, nor is there currently evidence that supports similar economy in ecosystem design.

Finally, the structure of an engineered system is generally static, whereas ecosystem structure is dynamic. Immigration, extinction, colonization, and evolutionary change are all important, dynamic properties of communities within ecosystems.

We should also note the way in which the words *function* and *functioning* are used in ecology. *Function* has two meanings, one referring to purpose and the other to activity. The term *functioning of biodiversity* may be misinterpreted as the purpose of biodiversity. Ecologically, however, the *functioning of biodiversity* refers simply to the role diversity plays in environmental processes, such as how biodiversity contributes to biomass production, nutrient cycling, soil acidity, soil water content, water quality, trace gas emissions, carbon dioxide (CO_2) ux, and many other processes. Similarly, the design of a natural ecosystem does not imply purposeful design, only the existence of structure and pattern. Managed ecosystems, however, can re ect considerable purpose in their design. In a farm, for example, the species composition, density, spatial arrangements, energy, water, and nutrient cycling may all be part of a manager's strategy to maximize production and yield.

Empirical Approaches

The functioning of biodiversity is determined by a variety of empirical methods. Observation of ecosystem behavior and dynamics, on the surface of the earth or from space through remote sensing, have provided ecologists with considerable knowledge about typical or average ecosystem behavior and activity for a large number of systems over many regions. These observations form the basis for expected or predicted ecosystem biogeochemical activity given geography, climate, and habitat.

Biodiversity, however, covaries with many environmental variables; the direct role of species in these observed ecosystem processes being best studied with experiments. For example, microbial biomass, pH, nitrogen content of litter, and other ecosystem properties show a variety of relationships with plant diversity in Swedish boreal forest islands (Wardle *et al.*, 1997). Plant diversity in these islands, however, is also associated with island size and fire frequency. Thus, in this system, like most natural systems, it is not possible without direct experimentation to determine whether biodiversity and ecosystem processes are associated because of confounding factors such as covarying island size and fire frequency. The primary value of such descriptive, observational studies is to establish the basic correlational pattern between biodiversity and ecosystem functioning as we move from habitat to habitat or ecosystem to ecosystem. Experimental approaches (discussed later) can then be used to determine how diversity affects ecosystem functioning within a particular habitat or ecosystem.

EXPERIMENTAL TESTS OF THE FUNCTIONING OF BIODIVERSITY

Experimental examinations of the functioning of biodiversity have focused on manipulating diversity and monitoring biogeochemical responses. This approach avoids the problems created by covarying or confounding factors typical of strictly observational studies. Any measure of the collective metabolic activities of species within a community represents a suitable measure of response, and may include production, nutrient retention, water retention, rates of decomposition, ux rates of oxygen (O_2), CO_2, or nutrients, or other biogeochemical processes.

Currently, the most common approach has been to create replicate ecological communities, varying the number of species among replicates and monitoring ecosystem response variables. Emphasis has been placed on plants because they acquire the energy and supply the organic material needed by all the organisms in an ecosystem (discussed later). A frequently used method is to clear the ground of vegetation, divide the cleared ground into equal-sized plots, then add plant species randomly to create monocultures, two-species plots, four-species plots, and so on. Once plants are established, production, decomposition, and other ecosystem measures are made on every plot. Averages for each level of species richness are calculated and compared.

Experiments come in a variety of shapes and sizes and have been conducted in growth chambers, greenhouses, in bottles and Petri dishes in the laboratory, in experimental plots in a single field, and in experimental plots established in replicate fields in different countries. Replicates have been as few as 14 growth chambers to as many as 460 microcosms. The majority of these experimental studies have shown that ecosystem processes respond in a variety of ways to experimental manipulations of species composition among replicate, model ecosystems (i.e., plots, pots, microcosms). A representative sample of results from some of these experiments is shown in Figure 2.

Most of these experiments have shown that ecosystem processes, especially the production of plant biomass, increases rapidly as one increases plant species richness from one species (monocultures) to multiple species (polycultures). In many cases, the increase in biomass diminishes as the number of plant species gets much higher than 5 or 10, suggesting ecological equivalency (redundancy) among species. This pattern can be seen quite clearly in Figure 2(a,b) which shows results from plant experiments using prairie grassland and English meadow species, respectively.

When one looks below ground, other interesting patterns are likely to emerge. For example, the biomass of microbes in the soil may show an increase with plant diversity (Figure 2c) and the number of mycorrhizal fungal species may affect plant production (Figure 2d). There are intimate

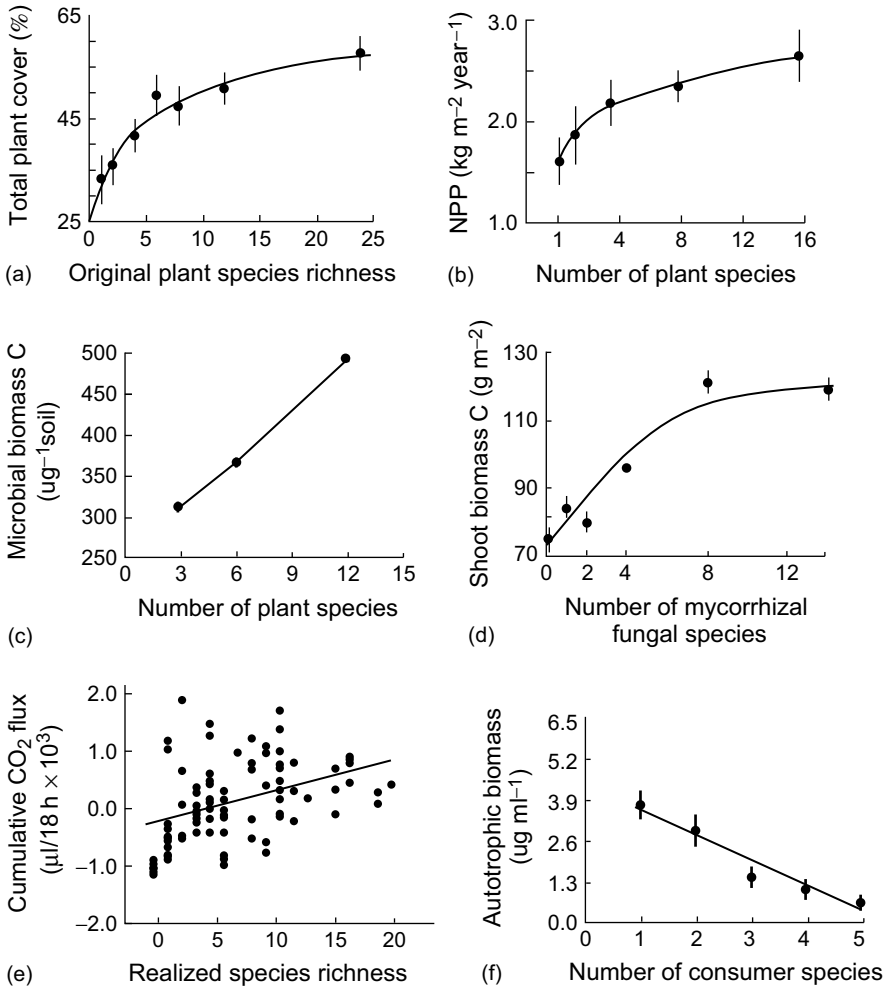


Figure 2 Experimental evidence that the diversity of an ecosystem's biota affects ecosystem processes. (a) The relationship between plant species richness of experimental grassland plots in Minnesota to percent of the ground covered with vegetation, a proxy measure for plant production (Tilman *et al.*, 1996). (b) The relationship between plant species richness of experimental old-field ecosystems and net primary production (NPP) (Chapin, III *et al.*, 1998). The ecosystems were modeled after English old-fields and maintained in controlled environmental chambers. (c) The relationship between plant species richness in Mediterranean, herbaceous ecosystems and microbial biomass, a proxy measure for microbial production. The experimental systems were intact sections of natural communities removed from the field and established in growth chambers within greenhouses in France (Chapin, III *et al.*, 1998). (d) The relationship between the number of mycorrhizal fungal species and shoot biomass, a proxy measure of plant production. The experimental systems were potted plants grown in greenhouses, modeled after North American and European plant communities (Van der Heijden *et al.*, 1998). (e) The relationship between CO_2 flux and microbial species richness. The experimental systems were modeled after microbial communities of protists and bacteria found in a local pond in New Jersey and maintained in bottles (McGrady-Steed *et al.*, 1997). Realized species richness is the actual number of species observed in the microcosm at the time of measurements (the initial diversity was higher, but local extinction occurred). (f) The relationship between heterotrophic protists, either bacterivores, algavores, or omnivores, and autotrophic (green alga) biomass, a proxy measure of production (Naeem and Li, 1997). The experimental systems were modeled after freshwater microbial communities and maintained in Petri dishes

connections between microbes in the soil and plants that suggest changes in diversity of one or the other will affect ecosystem functioning.

Microcosm experiments have shown that results from plant experiments may underestimate the complexity of ecosystem response to biodiversity loss. Microcosm experiments have the advantage that they can manipulate not just plant-like species (algae), but other trophic groups as well, such as herbivores (protists that eat algae), predators (protists that eat other protists), and decomposers (bacteria). Such experiments better capture the true complexity of natural ecosystems when compared

to plant-only experiments. Their smallness and artificiality, however, limit their applicability to predicting the response of much larger natural systems. Thus, microcosm experiments serve primarily as tests of basic ideas about the relationship between biodiversity and ecosystem functioning. For example, CO_2 flux is much higher with greater numbers of species in a microcosm containing many kinds of microbial species (Figure 2e) and algal biomass declines with greater numbers of protistan species as diversity increases (Figure 2f). These experiments point to the importance of distinguishing between plant, animal, and decomposer diversity when examining the relationship between biodiversity and ecosystem functioning, a fact that is often missed in plant-only experiments.

Why Does Biodiversity Affect Ecosystem Functioning?

These recent experiments have generally confirmed a variety of functional relationships between biodiversity and ecosystem processes, but causes for these relationships remain elusive. Three mechanisms have been proposed to explain such associations. One mechanism, commonly referred to as the *sampling effect*, is a probability-based explanation for the patterns observed. If a regional species pool contains a large number of species, some of which have strong impacts on ecosystem processes, then there is a high probability that replicate or local ecosystems which draw their biota from this pool (which sample the pool) will contain species with strong effects. If the species richness of the regional pool declines, then the probability that local systems get recolonized with strong-impact species declines, affecting ecosystem functioning.

Second, *species complementarity*, or species differing in the way they use resources, can create a positive relationship between diversity and functioning by leading to more thorough use of local resources in species-rich communities. For example, greater numbers of plant species, each with different rooting depths, may increase the thoroughness with which water and nutrients are used in a given volume of soil compared to monocultures that only root to one depth. Such principles of complementarity are well known and have been used in agroecological studies of intercropping. Temporal complementarity, in which certain species function only during certain environmental conditions, is known for microbes that appear at different times in degradative processes. Likewise, temporal complementarity may be found in stream invertebrates that form a shredder chain in which one group of species shreds incoming leaf litter into smaller pieces that are in turn consumed by other species. Similarly, temporal complementarity between early- and late-season owering plants can sustain needed pollinators throughout the year.

A third mechanism is that increasing the numbers of species increases the number of mutual, facilitative, or positive indirect effects among species in an ecosystem. This mechanism is poorly studied, so it remains unclear if such positive interactions are responsible for observed, positive effects between biodiversity and ecosystem functioning.

In most ecosystems, both the presence of species that have strong impacts on ecosystem processes and complementarity, may be important. In some cases, species may have such strong impacts on ecosystem functioning that complementarity may be less important than the presence or absence of strong-impact species.

Limitations in Experimental Approaches

Experimental approaches must contend with several problems all related to the practical difficulties associated with biodiversity research. First, the number of replicates required for statistical rigor is often prohibitive. Even small numbers of species provide for hundreds to thousands of possible combinations. Researchers invariably choose manageable numbers of species rather than representative numbers of species. Inferences from such experiments are therefore limited to cases in nature where severe reductions in biodiversity have already occurred, such as where natural habitats have been transformed to managed ecosystems. Second, experiments are small and short compared to the natural world. Third, additional factors, such as soil fertility, CO_2 concentrations, temperature, or other factors associated with global environmental change, are seldom manipulated in these experiments even though these factors interact with biodiversity and one another. As research programs grow in size and scope, many of these limitations may be minimized or overcome.

THE FUNDAMENTAL STRUCTURE (BLUEPRINT) OF ECOSYSTEMS

As in the exploration of any system, constructing a blueprint of system structure is a necessary step for understanding the functional significance of its parts. Evaluating the functioning of biodiversity, where we consider species to be the relevant parts or components of an ecosystem, involves the same exercise. Experimental and observational evidence already demonstrates that biodiversity is associated with ecosystem functioning. We know that the parts (species) are necessary. We do not know, however, if all the parts are necessary or if some parts are more important than others are. We address these issues by constructing the fundamental structure or blueprint of an ecosystem from which we can evaluate the relative importance of species found within it.

Evaluating the relative importance of species in ecosystem functioning is similar to evaluating the importance of

parts in an engineered system. In an engineered system, an important part is considered as one whose removal results in serious alteration of system performance or functioning. In ecological research, parts (or species) that are unique and represent critical steps in a series of linked, interdependent processes, are indispensable. Unnecessary parts (or species) are those whose removal or loss do not seriously compromise system functioning. Unlike parts in engineered systems, determining which species are critical or which species are expendable is quite difficult in ecological systems.

The relative importance of species in ecological processes is a well-developed research topic in ecology. Some species have long been recognized to be keystone species because their presence suppresses competitive dominants and permits greater diversity to persist within an ecosystem. Others consider the relative impact of the removal of a species to be a good measure of the relative importance of species to system functioning (Power *et al.*, 1996). By this approach, the species whose removal results in the largest change in the ecosystem is considered to have the highest *community importance value* (CIV) while all other species are given lower CIVs. For example, if the removal of one species leads to a dramatic reduction in production while the removal of another does not, we consider the first species to have a high CIV while the other species has a low CIV.

Other approaches consider those species most important whose activities modify the physical or chemical nature of the ecosystem (ecosystem engineering) such that the system supports many other species (Jones *et al.*, 1994). Beavers are considered a classic example of ecosystem engineers, building ponds within which many other species live. Clearly, the beaver is the most important species in the community as its activity, in large part, creates the ecosystem within which others live.

One final approach to assessing the relative importance of species is based on the impact a species has on the population dynamics of others. Some species are strong interactors, like top predators whose change in abundance has strong effects on the dynamics of other populations, while others are weak interactors, such as a rare, inferior competitor whose change in abundance has little effect on other populations (McCann *et al.*, 1998).

This variety of methods for evaluating the relative importance of species has a common theme. All these classification schemes, keystone versus non-keystone species, high versus low CIVs, ecosystem engineers versus non-ecosystem engineers, and strong versus weak interactors, imply that some species are indispensable in maintaining an ecosystem while others are potentially dispensable.

Identifying the Functional Roles of Species

Because all species contribute to ecosystem processes, these processes represent clear measures of the functioning of biodiversity. Organisms grow by the assimilation of elements and compounds. Either they assimilate materials from inorganic sources and convert them to organic materials or they assimilate materials from organic sources produced by other living organisms. Species acquire the energy they need for metabolic processes either from external sources such as light or energy stored in chemical bonds in organic or inorganic material. Regardless of how organisms acquire material and energy, or what their sources are, the end result is that all organisms of all species within an ecosystem chemically transform and physically move elements such as C, H, N, O, P, S, potassium (K), silicon (Si), calcium (Ca), magnesium (Mg), iron (Fe), and many others, into and out of the atmosphere, hydrosphere (e.g., lakes, streams, oceans), lithosphere (e.g., rocks or the mineral components of soil), and biosphere (Figure 1). The functional roles of species are therefore closely associated with their effects on these biogeochemical or ecosystem processes.

The functional roles of species are not necessarily related to taxonomic identity, thus taxonomy is of less interest in assessing the functioning of biodiversity. Taxonomic classification schemes are based on similarities of genetic information among groups of organisms reflecting shared evolutionary histories. Classification by genetic similarity, however, may show no correlation with classification by functional similarity. Two species in the same genus may not be similar in their contributions to biogeochemical

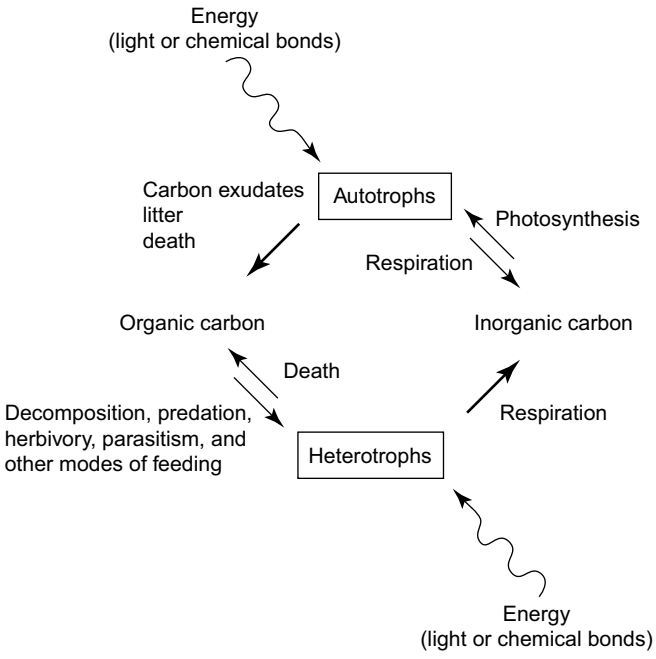


Figure 3 Fundamental trophic groups. All species can be divided into four fundamental groups distinguished by their source of carbon and their source of energy

cycles. Conversely, two species that are evolutionarily unrelated may be functionally quite similar. For example, the largest taxonomic division is that of the biological domains known as Eukarya, Bacteria, and Archaea, yet the ability to fix atmospheric nitrogen (convert N_2 to NO_2 or NO_3) is found in both Bacteria and Archaea, but not in Eukarya. In contrast, many species in the bacterial genus *Pseudomonas* do not fix nitrogen while some do (e.g., *Pseudomonas diazotrophicus*).

Source of Carbon versus Source of Energy as a Major Functional Distinction

To live and grow requires the acquisition and assimilation of carbon in one of two ways. First, one can acquire carbon by taking the carbon off of CO_2 and adding it to one's existing stock of organic carbon compounds. Alternatively, one can acquire carbon by consuming extrinsic sources of organic compounds. Acquiring carbon from CO_2 or other inorganic forms defines *autotrophy*. Acquiring carbon from organic or biological sources defines *heterotrophy*. These functional traits are illustrated in Figure 3.

This division is sometimes confused with the secondary division that concerns energy acquisition. If a species' source of energy for the acquisition of carbon from CO_2 is light, the species is a *photoautotroph*. If its source of energy is from chemical bonds it is a *chemoautotroph*. If the species is a heterotroph, but acquires its energy from light it is a *photoheterotroph*. If the species is a heterotroph but acquires its energy from chemical bonds, then it is a *chemoheterotroph*.

With respect to prokaryotes (Bacteria and Archaea) there are at least seven distinct kinds of autotrophic bacteria (Brock and Schlegel, 1989). Classification of the Eukarya, however, is less complex. The Eukarya consist of the protists (generally single celled organisms), fungi, metaphyta (multicellular organisms generally called plants) and metazoa (multicellular organisms generally called animals). The fungi and animals are heterotrophic because they obtain their carbon and energy from organic compounds. The protists can be photoautotrophs or heterotrophs or sometimes myxotrophic when their carbon may come from both CO_2 and organic material. Plants are generally photoautotrophic and obtain their carbon from CO_2 via photosynthesis. A more detailed functional classification of the earth's biota is provided at the end of this article.

Linkages Among Functional Groups

The construction of a general blueprint of ecosystem structure requires determining how many functional groups are necessary for an ecosystem to function and how many species are available to carry out each function. The majority of ecosystems have the essential features discussed in the simple terrestrial ecosystem portrayed in Figure 4. This figure can serve as a general blueprint of an ecosystem. Whether it is a grassland, savanna, forest, pond, lake, or marine ecosystem, they all contain living biomass, dead organic matter, and inorganic matter. We can therefore simplify Figure 4 by dividing ecosystem materials into two organic pools (living biomass and dead organic matter) and two inorganic pools (carbon and mineral or nutrient pools such as nitrogen, phosphorus, or sulfur). This simplification yields Figure 5, a generic model for an ecosystem more universal in its features than that shown in Figure 4. For simplicity, we have left out uptake facilitators (e.g., mycorrhizal fungi in terrestrial ecosystems and bacteria in aquatic systems) though they are commonly found in ecosystems.

Figure 5 shows six functional groups with each group occupied by a single species. In Figure 5, there are six serially dependent functional groups (autotrophs, decomposers, and two groups each of consumers of autotrophs and consumers of decomposers) each with a single species. *Serially dependent* means that one group cannot persist without the other because each is dependent on the functioning of the other. Autotrophs and decomposers, for example, cannot persist without each other because they are dependent upon one another for nutrient cycling. Similarly, consumers cannot persist without their prey. Each of these serially dependent groups is most likely occupied by many species. To produce a more useful blueprint, we can modify the model system in Figure 5 by adding a modicum of biodiversity and more clearly illustrating the serial dependency among the functional groups (Figure 6). Such a block diagram (Figure 6) provides a blueprint for an ecosystem in which each of six serially dependent functional groups is populated by six different species. In this hypothetical system, species are considered fully substitutable, though in real systems many species, even though occupying the same functional group, may not be so equivalent.

It is important to note that natural systems are likely extraordinarily complex compared to that shown in Figure 6. Even in the overly simplistic portrayal of an ecosystem in Figure 6, which contains only 36 species, there is already considerable complexity in the way species are linked to one another by nutrient and energy ows. In natural systems, however, there may be tens to hundreds of autotrophic species, hundreds to thousands of animal species, and thousands to tens of thousands of microbial species or strains. Nevertheless, the overall structure may still be understood in terms of functional groups and species within functional groups.

The generic blueprint of an ecosystem (Figure 6) reveals two major properties of ecosystems. First, the more serially dependent functional groups in an ecosystem process, the more likely functioning will not be achieved. Clearly the more groups of species dependent upon one another, the

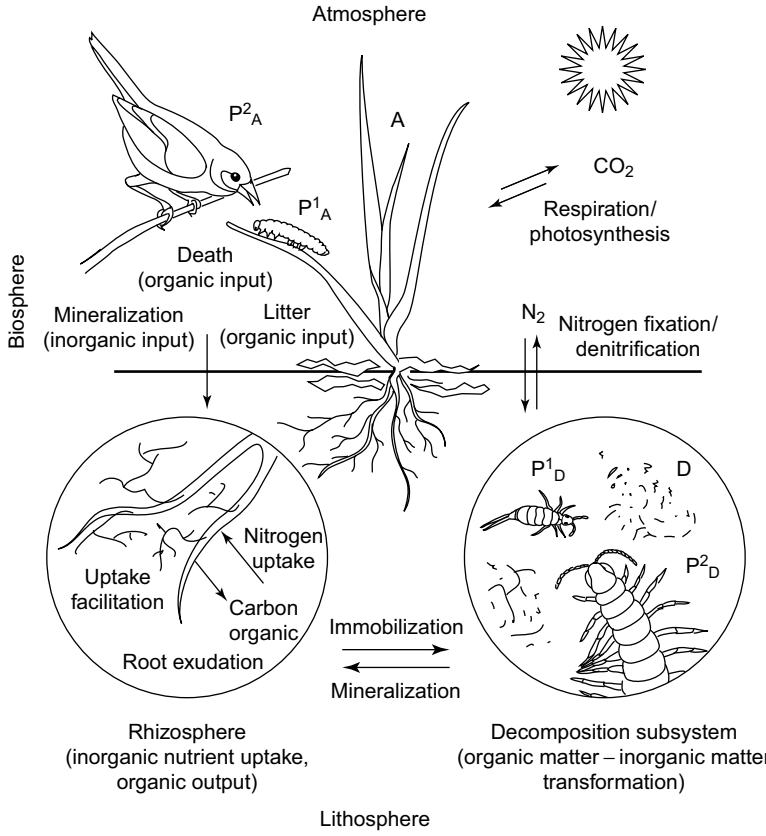


Figure 4 Common structure of an ecosystem as reflected in a model terrestrial ecosystem. This figure illustrates the core features of a simple terrestrial ecosystem. For simplicity, many processes and other functional groups of organisms are not illustrated. Circles represent magnified regions. The rhizosphere is illustrated by roots (larger, branched structure) and mycorrhizal fungi (thin lines entering roots). Root carbohydrate exudation and mycorrhizal facilitation of nutrient uptake are illustrated as arrows. The decomposition subsystem shows bacteria, fungi, a collembolan insect feeding on microbes and a small predatory centipede. Above-ground animal ecosystem processes are illustrated as mineralization, such as excretion of inorganic compounds, the entry into dead organic matter by death, or litter inputs from plants. A, photoautotrophs (e.g., plants). D, decomposers (e.g., bacteria and fungi). P^1_D, primary consumers of decomposers (e.g., bacterivores or fungivores). P^2_D, secondary consumers, decomposer-derived. P^1_A, primary consumers of autotrophs. P^2_A, secondary consumers, autotroph-derived. The atmosphere, lithosphere, and biosphere are illustrated to facilitate comparison with Figure 1. This figure should also be compared with Figures 5 and 6

more places functioning could change because one group is lost or adversely affected by environmental change. The second attribute apparent in the generic blueprint of an ecosystem (Figure 6) is that the more equivalent or redundant species one finds within a functional group, the greater the reliability of system functioning over long periods of time. More redundant species, like more redundant parts, means greater system reliability because loss of one or a few species is compensated for by the presence of other, substitutable species.

Difficulties in Determining Functional Groups and Species Redundancy

A strict definition of functional groups, one in which functional groups are serially dependent, is often not practical for ecosystems. Ultimately, the only truly serially dependent functional groups in continuously functioning ecosystems are likely to be decomposers dependent on organic carbon from autotrophs and autotrophs dependent on decomposers for material recycling. Most other species occupy two or more functional groups or their functional identity changes when conditions change. A rare herbivore, for example, may become a dominant regulator of plant biomass if the top predator vanishes. A weak interactor during one year may be a strong interactor when community composition or environmental conditions change.

FUNCTIONAL GROUPS AND FUNCTIONAL CLASSIFICATION OF SPECIES

As shown above, identifying functional groups and classifying species into functional groups provides an important

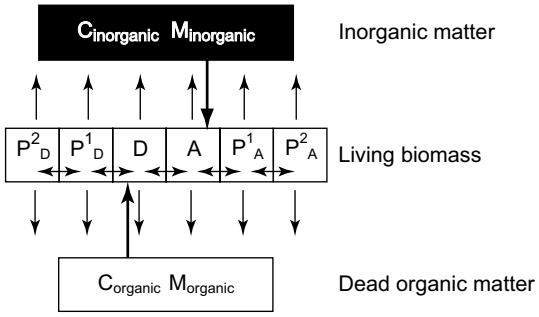


Figure 5 Common structure of ecosystems. This schematic portrays the common structure of an ecosystem by compartmentalizing material into common pools (organic and inorganic) and trophically-defined functional groups. Biological matter consists of living biomass and dead organic matter. Living biomass consists of autotrophs (A) and heterotrophs (see Figure 3), the latter of which consist of consumers (P^1_D, P^2_D, P^1_A or P^2_A) that are either derived from the consumption of decomposers (D) or the consumption of autotrophs (see Figure 4). Materials are divided into nutrients and mineral (M) or carbon (C), either bound in organic or inorganic compounds. Arrows indicate the flow of materials among the different compartments. Note that most ecosystems contain species that may exhibit complex feeding modes, such as omnivores that might feed on both autotrophs and decomposers or myxotrophs that are photoautotrophic but also consume other organisms. This figure should be compared with Figure 6

means for assessing the functioning of biodiversity. As discussed earlier, the major functional divisions of species are based on the different ways in which species acquire carbon and energy. Secondary and finer divisions are based on additional differences among species concerning their roles in other ecosystem processes. Following are two examples of functional group classification and then a full, but brief classification scheme is provided for the earth's biota.

Functional Groups of Plants

Many schemes exist for classifying plants into functional groups. Here, we consider major divisions based on how different groups of plants contribute to carbon cycling. We divide plants first into aquatic forms that generally acquire carbon in the form of HCO_3^- (carbonic acid or CO_2 dissolved in H_2O) versus terrestrial forms that fix atmospheric CO_2. We further divide plants into C_3, C_4 and Crassulacean acid metabolism (CAM) (*see* C_3 and C_4 Photosynthesis, Volume 2). Each of these groups of distinctly different plant physiologies (often associated with anatomical and taxonomic differences) leads to fixation or use of carbon dioxide in distinctly different ways (Pearcy *et al.*, 1989). C_4 plants, for example, are photosynthetically more efficient than C_3 plants whereas CAM plants fix CO_2 at night. Plants can be further divided into woody and herbaceous types. Woody plants produce compounds such as lignin that do not decompose readily and so remove carbon from the atmosphere for longer periods than herbaceous plants whose tissues are more readily decomposed.

We further divide plants into those that are associated with nitrogen-fixing symbionts and those which are not so associated. The rationale for this finer division is based on the observation that nitrogen-fixing plants, such as legumes, can have different C : N ratios, thereby producing litter that may decompose faster and tissues that are potentially more susceptible to herbivory. Nitrogen-fixing plants also have higher energy demands because they must use some of their photosynthate to support their energetically demanding nitrogen-fixing symbionts.

In functional groups, as in taxonomic groups, distinctions are not always clear and some groups have more representatives than others. For example, at least 22 species in eight different genera have been identified as intermediate between C_3 and C_4 and only one tree and a few bushes have been identified as woody C_4 plants.

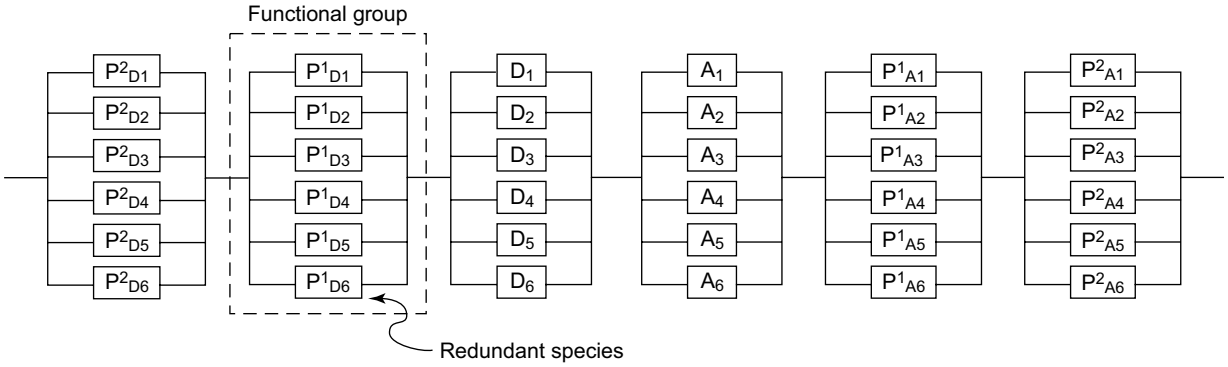


Figure 6 A block diagram or blueprint of an ecosystem ignoring material pools. Species are divided into six serially dependent functional groups within which redundant species are represented. As constructed in this ecosystem, normal functioning would only be obtained if at least one species in each functional group was present. Greater numbers of species insure that the system functions normally. See caption of Figure 5 for explanation of terms in boxes

Functional Groups of Nitrogen Cycling Organisms

Nitrogen cycling represents a good example of the lack of correspondence between taxonomic and functional classification schemes. Atmospheric nitrogen (N_2) is fixed only by prokaryotes (Bacteria and Archaea). Similarly, nitrification (the conversion of ammonia or NH_3 to NO_3^-) which can be taken up by plants and denitrification (the conversion of NO_3^- to atmospheric N_2), are both accomplished by prokaryotes. Thus, with the exception of lightning, which can convert N_2 to NO_3^-, the bulk of nitrogen cycling is accomplished by prokaryotes. Among these few functional groups (e.g., nitrogen-fixers, denitrifyers, and nitrifyers), however, many taxonomically distant species are involved. Over 20 families containing over 120 species, some more distantly related from one another than plants are from animals (i.e., Bacteria versus Archaea), exhibit N_2 fixing capabilities.

If we consider eukaryotic, symbiotic hosts of nitrogen-fixing prokaryotes, there is again an enormous diversity even though we lump them into a few functional groups. Among protists, some diatoms show symbiotic associations with cyanobacteria. Among plants, the cycads, liverworts, ferns, and *Gunnera* (an angiosperm) show symbiotic associations with cyanobacteria. Some fungi show associations with cyanobacteria yielding nitrogen-fixing lichens. Legumes show symbiotic associations with bacteria in the genus *Rhizobia*. A diverse array of terrestrial plants (greater than 200 species) show symbiotic associations with the bacteria in the genus *Frankia*. A few animals, including a ship worm (mollusk), have also been shown to exhibit symbiotic associations with nitrogen-fixing bacteria. Thus, although we can reduce the diversity of species associated with nitrogen-cycling to a small number of functional groups, taxonomic diversity is not only much greater, but shows little correlation with functional groups. Indeed, in the case of nitrogen-fixation, it is often necessary to classify symbiotic pairs of species (the nitrogen-fixing prokaryote and its host) as a single functional species.

A Brief Functional Classification of the Earth's Biota

The following is a brief functional classification of the earth's biota derived from the observations of how species affect ows of energy and materials as described in the previous sections. The classification is described in Table 1 and illustrated in Figure 7. Figure 7 also presents a contemporary taxonomic classification of biodiversity so that it can be compared with a functional classification.

Three things emerge from this simple functional classification. First, it is evident that many of the biogeochemical processes used in the classification are conducted predominantly by prokaryotes (Bacteria and Archaea). Second, it is also clear that there is no necessary correlation between functional groups and taxonomic groups (compare the top and bottom portions of Figure 7). Third, every group may be divided further. For example, within the chemoheterotrophs, sarcodinid protists may be divided into those producing calcium carbonate tests that serve as skeletons (foraminiferans) versus those that produce siliceous skeletons (radiolarians). Each of these subgroups significantly affects different elemental cycles (i.e., Ca, C, or Si). Other animal groups show such differences in use of elements. For example, arthropods primarily use chitin in their exoskeletons (an organic, carbon-based compound) sometimes fortified by calcium compounds, shelled mollusks use protein and calcium carbonate in their shells, vertebrates use protein and calcium salts in their bones, sponges use protein and spicules of either silica (e.g., certain *Demospongiae* and *Sclerospongiae*) or calcium carbonate (e.g., calcareous sponges), and corals use calcium carbonate. These organisms classified here simply as chemoheterotrophs would need further classification if calcium or silica were important in the functions being examined.

CONCLUSION

There is little doubt that the Earth's biota function critically in conditioning global environmental conditions. Historically, the appearance of oxygenic photosynthetic microorganisms 2.7 billion years ago led to the transformation of the earth's atmosphere from an anoxic environment to an oxygen-rich one. Later in the Permo-Carboniferous period (300 million years ago) the terrestrial invasion of photosynthetic organisms may have contributed to the draw-down of atmospheric CO_2 that led to global cooling and glaciation in the Southern Hemisphere. Currently, biogeochemistry provides abundant evidence for the role of the Earth's biota in governing cycles that affect greenhouse gases, soil fertility, and water quality. Without the biota, which means without biogeochemical processes, geochemical and physical processes would eventually transform the earth to an inhospitable planet (Lovelock, 1979). In fact, the foundation of the Gaia hypothesis (Lovelock, 1979) is based on the notion that this feedback between biogeochemistry and the biota creates a homeostatic regulation of earth's environment, an interesting hypothesis that is under investigation by many researchers (*see* Gaia, Volume 2; Gaia Hypothesis, Volume 5).

The role of the earth's extraordinary diversity, ranging from 10–100 million species, is only recently unfolding. Three important functions of this diversity have been identified. First, whereas the biota function as a whole in the transformation and movement of material among the earth's major spheres (hydro-, litho-, atmo-), it is the breadth of diversity within the biota that permits these cycles to occur over a physically and chemically heterogeneous planetary surface. The Earth's many biomes are defined

Table 1 Major functional divisions of the biota. See Figure 7 for illustration

Primary functional group	Secondary functional group	Finer divisions					Comments (examples)
Autotrophs: species that acquire carbon from inorganic sources	Photoautotrophs: acquisition of carbon from inorganic sources using light energy	Anoxygenic – photosynthesis					Species that do not produce oxygen as a by-product of (purple and green sulfur bacteria)
		oxygenic					species that produce oxygen as a by-product of photosynthesis
			Photosynthetic prokaryotes				(blue green algae)
			Photosynthetic protists	Without cell walls			(Example: cryptomonads, euglenids)
				With cell wall	Cellulose or organic material in cell wall or outer covering		(Some dinoflagellates)
					Siliceous cell walls		(Silicoflagellates)
					Calcified cell walls		(Some green and red algae)
			Green plants				
				Aquatic plants			Aquatic macrophytes
				Terrestrial plants	Grasses		Generally recover from above-ground grazing more readily, many species use silica to strengthen tissues
						C_3	Less efficient use of CO_2
						C_4	More efficient use of CO_2

Forbs, ferns and non-vascular plants		These organisms use less carbon, but require locally sufficient nitrogen
	C_3 forbs not associated with nitrogen-fixing symbionts	Less efficient use of CO_2
	Nitrogen fixers associated with nitrogen-fixing symbionts	These organisms use some of their acquired carbon to support symbionts, but have nitrogen supplied for them by their symbionts
	C_4	More efficient use of CO_2
Woody plants		Carbon used in structural compounds such as lignin that are more resistant to decomposition
	C_3 without nitrogen-fixing symbionts	Less efficient use of CO_2
	Associated with nitrogen-fixing symbionts	These organisms use some of their acquired carbon to support symbionts, but have nitrogen supplied for them by their symbionts
	C_4	Rare
CAM		Sequester CO_2 at night

(*continued overleaf*)

Table 1 (*continued*)

Primary functional group	Secondary functional group	Finer divisions		Comments (examples)
	Chemoautotrophs			Acquisition of carbon from organic material using inorganic chemical bonds (bacteria using energy stored in hydrogen, carbon monoxide, iron, sulfur, and nitrogen chemical bonds)
Heterotrophs	Photoheterotrophs: acquisition of carbon from organic sources using energy from light energy			(Examples: halobacteria and heliobacteria)
	Chemoheterotrophs: acquisition of carbon from organic sources using energy from organic compounds	Decomposers	Organic–inorganic transformers	Movement of nutrients to and from dead organic matter to (mineralization) and from (immobilization) inorganic pools (examples: many bacteria, fungi, including gut endosymbionts of herbivorous animals such as termites and ungulates)

	Uptake facilitators			Microbes that locally concentrate or mobilize nutrients which are taken up or supplied directly to autotrophs (examples: mycorrhizal fungi that facilitate uptake in terrestrial plants and bacteria that locally concentrate nutrients near phyto-plankton)
Consumers: direct consumption of living biomass	Decomposer-derived (examples: bacterivorous nematodes, nemato-phagous mites, predatory soil micro arthropods)	Primary: consume decomposer bacteria or fungi	Bacterivores	Consume bacteria
			Fungivores	Consume fungi
			Detritivores	Consume particulate matter containing microbes
		Secondary		Consume primary consumers in decomposer food chains
		Tertiary (and higher)		Consume secondary consumers in decomposer food chains

(*continued overleaf*)

Table 1 (*continued*)

Primary functional group	Secondary functional group	Finer divisions				Comments (examples)
				Omnivores		Consumes organisms from two or more trophic groups
			Autotroph-derived	Primary: consume autotrophic species such as plants or algae	Herbivores	Consume photosynthetic tissue
					Frugivores	Consume reproductive tissue
					Pollinators	Consume nectar and/or pollen
				Secondary		Consume autotrophic-derived primary consumers
				Tertiary (and higher)		Consume autotrophic-derived secondary or higher consumers
				Omnivores		
Myxotrophs						Acquires energy from inorganic and organic sources (example: photosynthetic protists that also consume bacteria such as euglenoid protists)

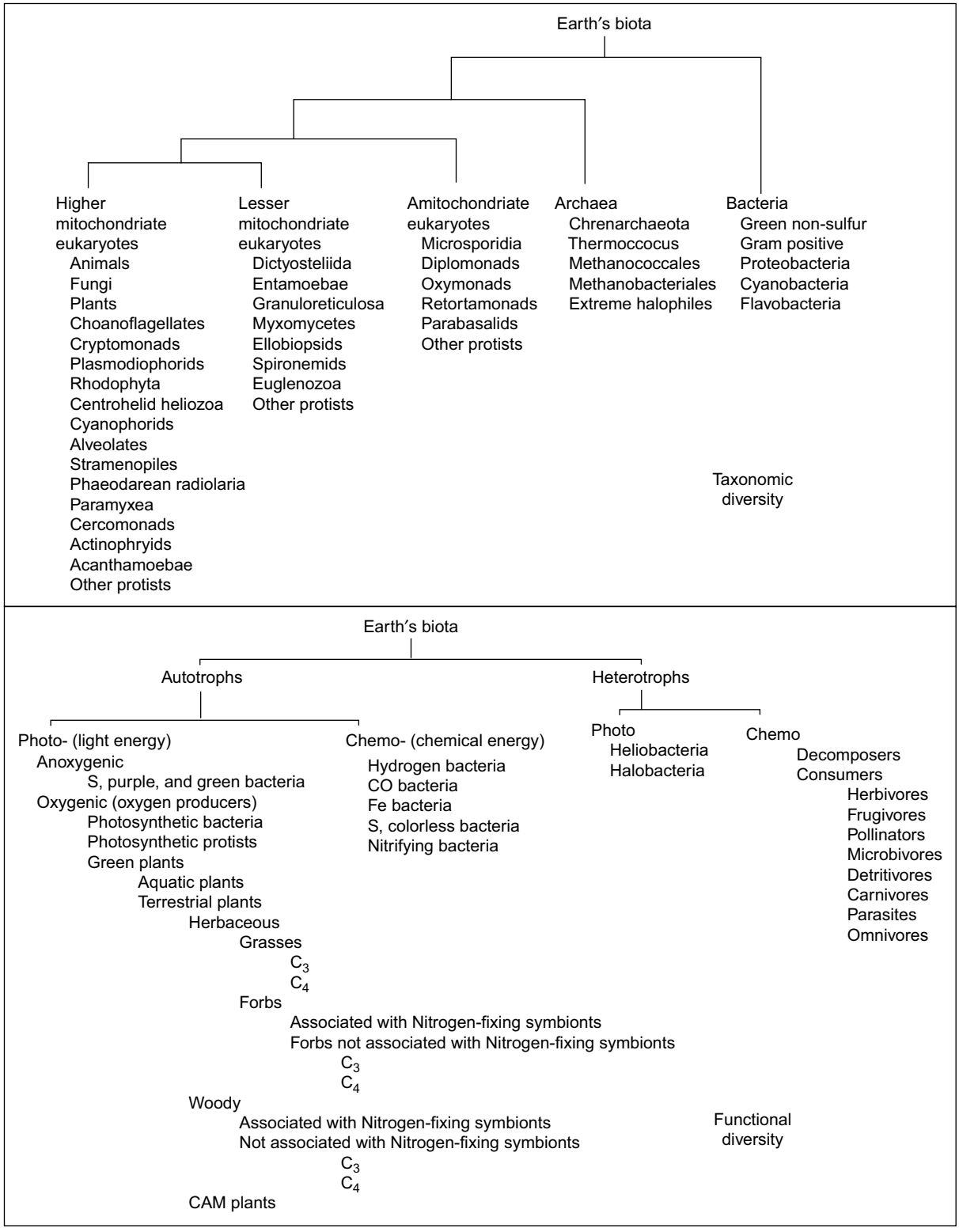


Figure 7 Taxonomic versus functional classifications of the biota. The top dichotomous tree represents species grouped by genetic or taxonomic similarity whereas the bottom represents species classified by their contributions to biogeochemical cycles (focusing primarily on carbon and nitrogen cycles). See text and Table 1 for further explanation

by climatically distinct regions, each containing numerous uniquely adapted species. Deserts, for example, contain drought-resistant plants, whereas tundras contain cold hardy plants and salt marshes contain halophytic or salt-loving plants. Such a diversity of differently adapted species permits carbon cycling and other biogeochemical processes to occur over much of the earth s surface.

Second, the ef ciency, in the sense of maximizing use of local resources, occurs when many complementary species are present, each using slightly different versions of a resource (e.g., different kinds of organic substrates) or each exploiting resources in slightly different ways (e.g., plants rooting at different depths). The rich variety of organic carbon sources, for example, that make up dead organic matter, is used the most when a wide diversity of species representing a wide diversity of enzymes for breaking down different kinds of organic compounds, is present in the decomposer functional group. This is the foundation of *bio-remediation*, which seeks to maximize the necessary microbial diversity to break down the complex milieu of organic compounds that comprise oil and other pollutants. *See* **Bioremediation**, Volume 3.

Third, having several species per functional group, or having greater redundancy of species, provides greater reliability of ecosystem functioning over a wide range of varying environmental conditions. For example, drought-tolerant plants may dominate communities during drought, thereby ensuring continued carbon cycling and production during drought years. By a functional classi cation, many species may appear redundant, but this redundancy is one of the most important values of diversity.

Interestingly, we tend to be disinclined, for whatever reasons, to consider familiar species as redundant even if they are functionally equivalent and taxonomically similar. Many, for example, would nd it worrisome to group zebra, wildebeest, giraffes, reed bucks, and elephants into a functional group called herbivores even though they are all members of a single, narrowly de ned taxonomic class and narrowly de ned functional group. On the other hand, species of microbes, which are less familiar to us, are frequently grouped into a microbial black box in ecological studies. Yet microbes in the microbial black box typically represent hundreds to thousands of species belonging to several phyla and representing all three biological domains (Eukarya, Bacteria, and Archaea). Awareness of the functional role of biodiversity draws attention to the importance of preserving both taxonomic and functional aspects of biodiversity.

The functioning of biodiversity is an inseparable issue from other topics in global environmental change. Certainly global environmental change affects biodiversity, but the converse, that changes in biodiversity may affect global environmental processes, is a rapidly growing research frontier.

REFERENCES

Brock, T D and Schlegel, H G (1989) Introduction, in *Autotrophic Bacteria*, eds H G Schlegel and B Bowein, Science Tech. Publishers, Madison, WI, 1 15.

Chapin, III, F S, Sala, O E, Burke, I C, Phillip Grime, J, Hooper, D U, Lauenroth, W K, Lombard, A, Mooney, H A, Mosier, A R, Naeem, S, Pacala, S W, Roy, J, Steffen, W L, and Tilman, D (1998) Ecosystem Consequences of Changing Biodiversity, *BioScience*, **48**, 45 52.

Costanza, R, D Arge, R, de Groot, R, Farber, S, Grasso, M, Hannon, B, Limburg, K, Naeem, S, O Neill, R V, Paruelo, J, Raskin, R G, Sutton, P, and van den Belt, M (1997) The Value of the World s Ecosystem Services and Natural Capital, *Nature*, **387**, 253 260.

Daily, G C, ed (1997) *Nature s Services*, Island Press, Washington, DC.

Ehrlich, P R and Ehrlich, A H (1981) *Extinction: The Causes and Consequences of the Disappearance of Species*, Random House, New York.

Heywood, V H, ed (1995) *Global Biodiversity Assessment*, UNEP and Cambridge University Press, Cambridge.

Jones, C G, Lawton, J H, and Shachak, M (1994) Organisms as Ecosystem Engineers, *Oikos*, **69**, 373 386.

Lovelock, J (1979) *Gaia*, Oxford University Press, Oxford.

McCann, K, Hastings, A, and Huxel, G R (1998) Weak Trophic Interactions and the Balance of Nature, *Nature*, **395**, 794 797.

McGrady-Steed, J, Harris, P M, and Morin, P J (1997) Biodiversity Regulates Ecosystem Predictability, *Nature*, **390**, 162 165.

Mooney, H A, Cushman, J H, Medina, E, Sala, O E, and Schulze, E D, eds (1996) *Functional Roles of Biodiversity: a Global Perspective*, John Wiley & Sons, Chichester.

Naeem, S and Li, S (1997) Biodiversity Enhances Ecosystem Reliability, *Nature*, **390**, 507 509.

Pearcy, R W, Ehleringer, J, Mooney, H A, and Rundel, P W, eds (1989) *Plant Physiological Ecology: Field Methods and Instrumentation*, Chapman and Hall, London.

Power, M E, Tilman, D, Estes, J A, Menge, B A, Bond, W J, Mills, L S, Daily, G, Castilla, J C, Lubchenco, J, and Paine, R T (1996) Challenges in the Quest for Keystones, *BioScience*, **46**, 609 620.

Schlapfer, F and Schmid, B (1999) Ecosystem Effects of Biodiversity: a Classi cation of Hypotheses and Exploration of Empirical Results, *Ecol. Appl.*, **9**, 893 912.

Schlesinger, W H (1997) *Biogeochemistry*, 2nd edition, Academic Press, San Diego, CA.

Schulze, E D and Mooney, H A, eds (1993) *Biodiversity and Ecosystem Function*, Springer Verlag, New York.

Tilman, D, Wedin, D, and Knops, J (1996) Productivity and Sustainability In uenced by Biodiversity in Grassland Ecosystems, *Nature*, **379**, 718 720.

Van der Heijden, M G A, Klironomas, J N, Ursic, M, Moutoglis, P, Streitwolf-Engel, R, Boller, T, Wiemken, A, and Sanders, I R (1998) Mycorrhizal Fungal Diversity Determines Plant Biodiversity, Ecosystem Variability and Productivity, *Nature*, **396**, 69 72.

Wardle, D A, Zackrisson, O, Hornberg, G, and Gallet, C (1997) The In uence of Island Area on Ecosystem Properties, *Science*, **277**, 1296 1299.

IGBP Core Projects

WILL STEFFEN

IGBP Secretariat Royal Swedish Academy of Sciences, Stockholm, Sweden

The overall goal of the International Geosphere–Biosphere Programme (IGBP) is to describe and understand the interactive physical, chemical and biological processes that regulate the Earth System, the unique environment it provides for life, the changes that are occurring and how they are influenced by human actions. To achieve this goal, the IGBP has organized its research effort into a number of core projects and framework activities. The core projects, largely oriented around broad disciplinary groupings, carry out the bulk of the scientific research within IGBP. The framework activities are designed to support the scientific endeavor carried out in the core projects by providing Global Analysis, Interpretation and Modeling (GAIM) data and information as required and the regional perspective on global change science.

Beyond the scientific research that is carried out within the core projects themselves, IGBP is a coherent scientific effort in its own right. To achieve its ultimate goal of understanding the changing Earth System as a whole, especially human perturbations to this system, requires that the work of the core projects be integrated and synthesized at a higher level. This task is tackled jointly by a suite of inter-core project activities, by the GAIM framework activity, and by the Scientific Committee of the IGBP and its executive body, the IGBP Secretariat.

THE CORE PROJECTS: OBJECTIVES AND STRUCTURE

The eight core projects of the IGBP (Figure 1) represent various ways of viewing the Earth System. Five of them – Global Change and Terrestrial Ecosystems (GCTE), Land–Ocean Interactions in the Coastal Zone (LOICZ), International Global Atmospheric Chemistry Project (IGAC), Global Ocean Ecosystem Dynamics (GLOBEC) and Joint Global Ocean Flux Study (JGOFS) – represent the land, coastal, atmospheric and oceanic compartments of the Earth System, respectively. Land-Use/Cover Change (LUCC) studies a global change driver, Biospheric Aspects of the Hydrological Cycle (BAHC) analyzes an Earth System cycle and Past Global Changes (PAGES) takes a time slice perspective on the changing Earth System as a whole.

The IGBP core projects also vary in their historical development. Two of them – IGAC and JGOFS – were initiated independently of IGBP, in fact before IGBP was conceived, and were later adopted by IGBP. Three of them – BAHC, GCTE, and PAGES – joined IGAC and JGOFS as the original core projects, all of which launched their implementation phase around 1990 or earlier. LOICZ, LUCC and GLOBEC were phased in throughout the 1990s, as their respective scientific communities developed coherent international frameworks, and resources were found to begin and coordinate the implementation phase. The scientific objectives of the core projects have evolved throughout the development of IGBP. The work of the five original core projects is strongly oriented towards basic Earth System science, with only GCTE featuring a strong effort on the impacts and consequences of changes in the Earth System. The three newer core projects – LOICZ, LUCC, and GLOBEC – all have a balance between basic Earth System science and impacts or consequences. In all core projects, there is a balance among the physical, chemical and biological sciences, but IGBP's strength lies in its ability to fully integrate the biological aspect into Earth System science.

Several core projects have co-sponsors. The Scientific Committee on Oceanic Research (SCOR) jointly sponsors both JGOFS and GLOBEC; the latter is also sponsored by the Intergovernmental Oceanographic Commission (IOC). IGAC is co-sponsored by the Commission on Atmospheric Chemistry and Global Pollution (CACGP) of the International Association of Meteorology and Atmospheric Science (IAMAS). LUCC is co-sponsored by the International Human Dimensions Programme on Global Environmental Change (IHDP).

Biospheric Aspects of the Hydrological Cycle (BAHC)

Understanding the biospheric controls on the hydrological cycle, and how changes to that cycle affect water quantity

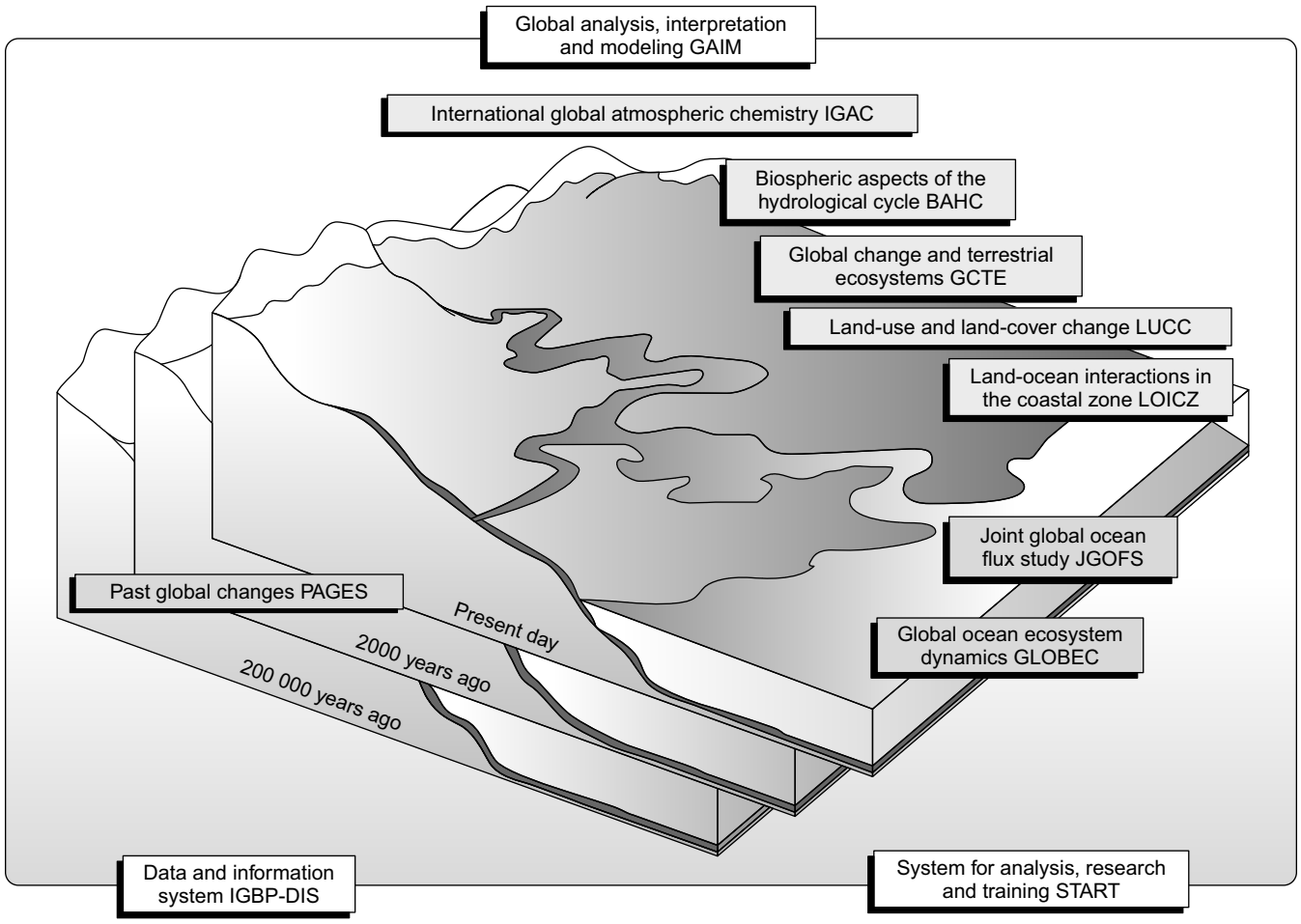


Figure 1 The IGBP core projects and framework activities

and quality, is crucial for sustainable development. BAHC's objectives are:

- to determine the biospheric controls of the hydrologic cycle through field experiments, for the purpose of developing models of energy and water fluxes in the soil–vegetation–atmosphere system at temporal and spatial scales ranging from vegetation patches to general circulation model (GCM) grid cells;
- to develop appropriate databases that can be used to describe the interactions between the biosphere and the physical Earth System, and to test or validate model simulations of such interactions.

BAHC is organized into 10 major areas, or key themes:

1. FLUXNET (a global network of stations for the routine measurement of water, energy and CO_2 fluxes – see the section on FLUXNET);
2. below-ground hydrological processes;
3. land-use change – climate/hydrology interactions at regional scales;
4. parameterization of land/atmosphere interactions;
5. global terrestrial–climate interactions (see the section on BAHC Earth System Modeling);
6. water transport/riverine systems;
7. contributions to developments of global databases;
8. mountain ecohydrology;
9. integrated terrestrial experiments;
10. impacts–feedbacks–vulnerability–water resources.

Global Change and Terrestrial Ecosystems (GCTE)

The health and productivity of the terrestrial biosphere is essential for human well-being; in addition, the terrestrial biosphere plays an important role in the dynamics of the Earth system. GCTE's objectives are:

- to determine the effects of changes in climate, atmospheric composition and land use on terrestrial ecosystems, including agriculture, forestry, soils, and ecological complexity;
- to determine the feedback of these changes to the atmosphere and the physical climate system.

GCTE's research effort is organized into four foci:

1. ecosystem physiology and global change;
2. ecosystem structure;
3. global change impact on agriculture, forestry and soils;
4. global change and ecological complexity.

In addition, GCTE has established a global change impacts center in Bogor, Indonesia, as the first in a set of such centers to help apply the results of its research to policy and resource management issues in the developing world.

Global Ocean Ecosystem Dynamics (GLOBEC)

The most extensive ecosystem on Earth is the marine, and the ocean is inextricably involved in the physical, chemical and biological processes that regulate the total Earth System. GLOBEC's objective is:

- to advance our understanding of the structure and functioning of the global ocean ecosystem, its major subsystems, and its response to physical forcing so that a capability can be developed to forecast the responses of the marine ecosystem to global change.

GLOBEC is organized into four foci:

1. retrospective analyses in the context of large-scale climatic changes;
2. process studies;
3. predictive and modeling capabilities;
4. feedbacks from changes in marine ecosystem structure.

International Global Atmospheric Chemistry Project (IGAC)

Human activities are now changing atmospheric composition rapidly compared to the past, and are significantly perturbing the natural processes operating in the atmosphere. IGAC's objectives are:

- to develop a fundamental understanding of the processes that determine atmospheric composition;
- to understand the interactions between atmospheric chemical composition and biospheric and climatic processes;
- to predict the impact of natural and anthropogenic forcings on the chemical composition of the atmosphere.

The implementation of IGAC is built around three foci:

1. biosphere–atmosphere interactions
2. oxidants and photochemistry
3. atmospheric aerosols

and two framework functions:

1. fundamental and cross-cutting activities;
2. capacity building.

Joint Global Ocean Flux Study (JGOFS)

Oceans are the sleeping giants in the uptake of increasing CO_2 in the atmosphere. The objectives of JGOFS are:

- to determine and understand on a global scale the processes controlling the time-varying fluxes of carbon and associated biogenic elements in the ocean, and to evaluate the related exchange with the atmosphere, sea floor and continental boundaries;
- to develop a capacity to predict on a global scale the response of oceanic biogeochemical processes to anthropogenic perturbations, in particular those related to climate change.

The implementation of JGOFS is built around large-scale surveys, process studies, time series, sedimentary record studies, and continental margin boundary fluxes, with data base development and management and global synthesis playing central roles. JGOFS has adopted a strongly regional approach to implementation, with work in the North Atlantic, equatorial regimes, the Southern Ocean, monsoon regimes, eastern and western boundary regimes and central gyres.

Land-Ocean Interactions in the Coastal Zone (LOICZ)

The coastal zone is a relatively small but highly dynamic and sensitive area that is coming under increasing pressure from population growth, urbanization and industrialization. The objectives of LOICZ are:

- to determine at global and regional scales:
    1. the fluxes of material between land, sea and atmosphere through the coastal zone;
    2. the capacity of coastal systems to transform and store particulate and dissolved matter;
    3. the effect of changes in external forcing conditions on the structure and functioning of coastal ecosystems.
- to determine how changes in land use, climate, sea level and human activities alter the fluxes and retention of particulate matter in the coastal zone, and affect coastal morphodynamics;
- to determine how changes in coastal systems, including responses to varying terrestrial and oceanic inputs of organic matter and nutrients, will affect the

global carbon cycle and trace gas composition of the atmosphere;
- to assess how responses of coastal systems to global change will affect habitation and usage by humans of coastal environments, and to develop further the scientific and socio-economic bases for the integrated management of the coastal environment;

LOICZ is structured around four foci:

1. the effects of changes in external forcing or boundary conditions on coastal fluxes;
2. coastal biogeomorphology and global change;
3. carbon fluxes and trace gas emissions;
4. economic and social impacts of global change in coastal systems.

Land-Use/Cover Change (LUCC)

For the Earth's land surface, the most important component of global change for the next several decades at least is land-use and land-cover change, more important than either changes in atmospheric composition or in climate. LUCC's objective is:

- to improve understanding and projections of the dynamics of LUCC as inputs to and consequences of global environmental change and as elements of sustainable development.

LUCC is organized into three foci:

1. land-use dynamics – comparative case study analysis;
2. land-cover dynamics – direct observation and diagnostic models;
3. regional and global models – framework for integrative assessments.

Past Global Changes (PAGES)

Paleo studies are powerful tools for understanding the natural variability of the Earth system, and for discerning human impacts on the system in the recent past. The objectives of PAGES are:

- to determine how global climate and the environmental systems linked to it changed in the past, the factors responsible for those changes, and how knowledge of past changes can be used to improve insight into likely future climate and environmental change;
- to determine the extent to which human activities have modified climate and the global environment, and to disentangle anthropogenically-induced change from natural responses to external forcing mechanisms and internal system dynamics;
- to determine the limits of natural greenhouse gas variation in the atmosphere, the natural feedback to the global climate system, and the sequence in which changes in greenhouse gas concentrations, surface climate and ecological systems occur;
- to determine the important forcing factors that produce climate change on societal time scales, the causes of abrupt climatic and environmental events and of the transitions between quasi-stable climatic states that occur on decade to century time scales.

The PAGES implementation is built around five foci:

1. palaeoclimates of the Northern and Southern Hemispheres;
2. palaeoclimate and environmental variability in polar regions;
3. human interactions in past environmental changes;
4. climate system sensitivity and modeling;
5. cross-project analytical and interpretive activities.

THE FRAMEWORK ACTIVITIES: OBJECTIVES AND STRUCTURE

The IGBP framework activities were all initiated at the beginning of IGBP in the late 1980s. More recently, Global Change System for Analysis, Research and Training (START) has become a jointly shared activity with the World Climate Research Programme (WCRP) and the IHDP.

Data and Information System (DIS)

Global change research requires global data, and an international research effort requires an integrated and efficient information system. The objective of DIS is:

- to improve the supply, management and use of the data and information that are needed to attain IGBP's scientific goals.

DIS is organized into three foci:

1. data set development;
2. data management and dissemination;
3. data coordination in an international context.

The data needs of IGBP, and the most appropriate structure required to meet them, were reviewed in 1998/1999 and a restructured DIS became fully operational in 2000. The main change in DIS was the new emphasis on service rather than system, and a more user-driven approach in response to the data needs of the core projects.

Global Analysis, Interpretation and Modeling (GAIM)

The work of IGBP is not complete without understanding the interactions between the core projects – the flows and feedbacks, the thresholds and non-linearities which are characteristic of complex systems. The goal of GAIM is:

- to advance the study of the coupled dynamics of the Earth System using as tools both data and models.

GAIM's implementation is based on a task force approach, engaging scientists who are involved in IGBP core projects and focusing on topics which involve more than one core project. GAIM activities to date have included intercomparison projects on ocean carbon models, atmospheric transport models, and models simulating net primary production of terrestrial vegetation; regional interaction of climate and ecosystems; climate–biosphere model studies of 6 k and 18 k years before present; a global carbon cycle model linkage project; and the characterization of global wetland distribution and functioning.

Global Change System for Analysis, Research and Training (START)

Increasing emphasis is being placed on the regional perspective of global change, dealing with the heterogeneity of the planet in terms of both Earth System processes and global change impacts. START employs a system of regional networks to meet its objectives:

- to conduct research on regional aspects of global change;
- to assess the impacts of regional findings;
- to provide regionally important integrated and evaluated information to policymakers and governments.

In its implementation START places emphasis on the development of regional networks in developing countries. It seeks to mobilize the resources required to augment existing scientific capabilities and infrastructures in developing countries.

IMPLEMENTATION STRATEGIES: THE VALUE-ADDING OF IGBP SCIENCE

Understanding the changing Earth System is beyond the capacity of the scientific communities of individual countries; it requires a coordinated international effort of unprecedented interdisciplinarity and cooperation. The implementation of IGBP's research program is based on voluntary contributions from a large number of national and regional global change research projects and the promotion and initiation of new work where critical gaps exist. The core projects knit these contributing projects together into a coherent effort, using a variety of value-adding activities. The basic principles on which the IGBP core projects undertake these activities are:

1. They must address questions which can only or best be answered through collaborative, interdisciplinary international research;
2. They must ensure that the scientific output of the activities is greater than what could be achieved by the individual contributing projects working on their own – the whole is more than the sum of its parts.

The most important of the value-adding activities are described below, with examples given for each.

Agreed International Framework

The first task of each of the core projects is to develop an agreed international framework under which the collaborative work will be undertaken. This in itself is a complex task, involving extensive consultation with the science and policy communities in many regions of the world. This process enables a wide range of ideas and hypotheses to be incorporated into the science plan, and entrains a large number of scientists into the implementation phase. The agreed framework ensures that a strong focus is maintained during the implementation of the research and facilitates the synthesis of results at regular intervals and the reformulation of the scientific agenda based on work completed under the agreed framework.

The LUCC Science Plan

The science plan developed by the land-use and land-cover change research community demonstrates not only the development of a widely accepted framework for international research, but also how that process of developing such a framework can itself promote interdisciplinary research. As noted earlier, the LUCC science plan is organized into three broad thematic areas, land-use dynamics – comparative case study analysis, land-cover dynamics – direct observation and diagnostic models; and regional and global models – framework for integrative assessments. Traditionally the theme of land-use dynamics has been the territory of the social sciences while the second theme, with its strong emphasis on remote sensing as a research tool, has been dominated by the natural sciences. However, in carefully developing a single, coherent science plan for the LUCC core project, the process of integrating the social and natural sciences had begun. For example, Figure 2 shows the numerous social and biophysical driving forces which affect land-use and land-cover change at a variety of scales. Such a conceptual framework in itself drives collaboration. Even more striking is the interaction between the natural and social sciences in the

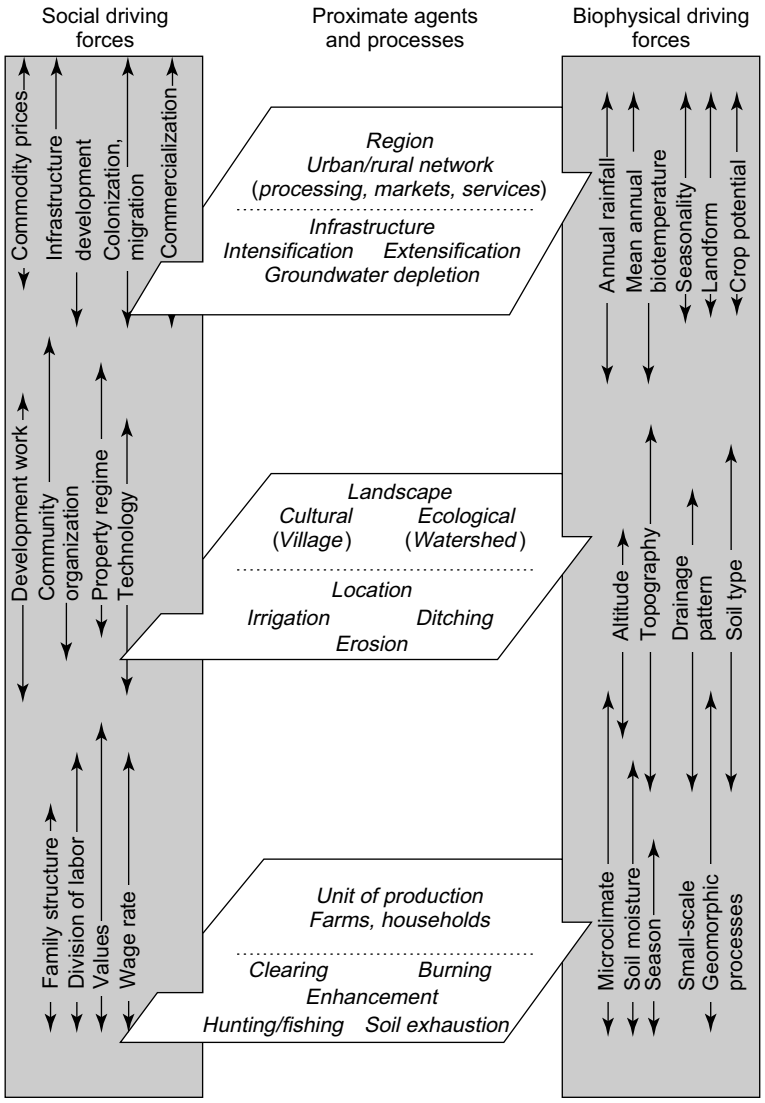


Figure 2 Multiscale driving forces in land-use and land-cover change

implementation of the LUCC Science Plan. An excellent example is the pioneering work on merging traditional social science research with remote sensing imagery, that is, LUCC's first two research themes. One of the concepts around which this interdisciplinary work is built is socializing the pixel, in which the biophysical data of a pixel (grid cell of a remotely sensed image) are interpreted in ways that yield social meaning or insights, or are used as a basis for social science modeling. A companion concept is pixelizing the social, in which social, political and economic data, which are usually aggregated by political unit (e.g., state or country) are disaggregated and made spatially explicit (Geoghegan *et al.*, 1998).

The Miombo Network Plan

At a regional scale, science plans are evolving to address regional issues specifically, but they are usually built around an integration of components of several IGBP (and WCRP and IHDP) core projects. The development of the Miombo Network, a study of global change in the seasonal woodlands of central and eastern Africa, is a good example of this approach. This framework adopts a modular approach in which work is organized around seven core studies: land-use dynamics, land-cover dynamics, ecosystem dynamics and disturbance, nutrient limitation in Miombo, biogeography and distribution of Miombo, regional primary production and carbon balance of Miombo, and sustainable natural resource management. The strengths of the Miombo Network plan include its broad community consensus and hence wide acceptance in the region, its emphasis on sustainable development–global change linkages, its explicit provision for activities in integration and synthesis and its incorporation of much existing appropriate research in

the region, thereby identifying where scarce new scientific resources are best placed.

Ef cient Use of Resources

A very practical benefit of IGBP research is the use of the agreed frameworks to make the most cost-effective use of scarce resources in achieving scientific results. This value-adding characteristic is most evident when large and expensive facilities, such as oceanographic research vessels, aircraft, or large *in situ* installations (e.g., free-air carbon dioxide enrichment (FACE) experiments) are required to carry out the work.

Aerosol Characterization Experiment

A good example of the coordination of scarce scientific resources is the series of Aerosol Characterization Experiment (ACE) campaigns carried out under the auspices of IGAC. For instance, the ACE-2 campaign was a major field effort over the North Atlantic in 1997 to study the radiative effects and controlling processes of anthropogenic aerosols from Europe and desert dust from Africa as they are transported over the North Atlantic Ocean (Raes and Bates, 1998). The campaign involved 60 coordinated aircraft missions with six aircraft, one ship, and ground stations on Tenerife, Madeira and mainland Portugal and over 250 scientists from the United States, five European countries and the European Union. ACE-2 was only possible through international collaboration, and was well beyond the scope of any of the individual countries or agencies.

Preliminary results from ACE-2 include:

- A wide variety of aerosol types was observed in the region studied, including background marine aerosols, European pollution-derived aerosols and African mineral dust. The last was encountered less frequently than expected.
- Aerosols originating over Europe and Africa and transported more than 3000 km from the source still impact on the chemistry and radiative properties of the atmosphere. For example, observations from a mountain ridge on Tenerife showed that the pH of cloud water was typically 5.5 during clean periods, but dropped sharply to 3.5 during polluted periods, with a substantial contribution of nitric acid to the acidity.
- Pollution from Europe and dust from Africa were transported in distinct layers, the optical depth of which varied over a wide range.

JGOFS pCO_2 Measurements

Another example of more efficient use of scientific resources is the suite of oceanic cruises undertaken by JGOFS, in collaboration with the WCRP's World Ocean Circulation Experiment (WOCE), to measure the exchange of CO_2 between the oceans and the atmosphere. If JGOFS and WOCE, with their detailed plans for international collaboration, did not exist, there likely would be a strong bias towards sampling the North Atlantic Ocean, around which much of the world's oceanographic research resources are concentrated. Instead, about 90% of the planned cruises over the past decade have actually been implemented, with a representative coverage of all of the world's oceans in all seasons of the year. The result is a vastly improved understanding of ocean–atmosphere carbon exchange – the breathing of the oceans.

At a broad scale the work has shown that the oceanic sink for atmospheric CO_2 is not quite as strong as estimated earlier by global models. The estimate has been reduced from about 2.0 Gt C annum^{-1} (e.g., IPCC, 1994) to about 1.6 Gt C annum^{-1} (Takahashi *et al.*, 1997). But the work also shows some fascinating details of oceanic exchange. For example, there is a large out-gassing of CO_2 in the equatorial Pacific Ocean, caused by strong upwelling in this region coupled with a lack of biological activity. The JGOFS work also shows that there is considerable variability in the strength of this source, with the almost complete disappearance of the source during El Nino years. In contrast to the equatorial Pacific, the North Atlantic Ocean is the most intense marine carbon sink on Earth, due to both physical and biological factors. As a result of the this research, the pattern, both in space and time, of oceanic exchange of CO_2 with the atmosphere is now much better known, yielding valuable insights into the nature of the processes responsible for this exchange and into the reliability of the oceanic sink into the future.

Network Approach

The backbone of IGBP's value-adding activities is the establishment and operation of international research networks (Ingram *et al.*, 1998). All core projects use this approach to varying degrees. Networks generate research which could not be done by the constituent groups acting independently (e.g., experiments where many individual sites or studies are required, yielding results and insights greater than the sum of the parts) sharing data, models and facilities as well as sharing costs. Experience in GCTE has shown that networks need a number of requirements for successful operation:

- The goals of the network must be clearly stated and scientifically exciting.
- The individuals in the network must themselves benefit in addition to the network as a whole.
- Success is maximized when there is a sense of ownership of the network by the members, which is best achieved by jointly designing network

Table 1 Key stages in the development of the GCTE Soil Erosion Network (from Ingram *et al.*, 1998)

Date	Event	Outcomes/product	Ongoing activities commenced
January 1991	GCTE Open Meeting (Brighton, UK); GCTE general planning	Draft GCTE Operational Plan	
September 1991	Focus 3 Open Workshop (Lunteren, NL): Focus 3 general planning/scoping exercise	Draft Focus 3 Operational Plan	
March 1993	Internal planning meeting (Oxford UK)	Task 3.3.2 (Soil Degradation) Inaugural Workshop designed	Inaugural Workshop planning
March 1994	Inaugural Workshop (Paris, France)	Task's major objectives agreed; major erosion models and monitoring projects identi ed; metadatabase structures developed; Task Working Group convened	Metadatabase structure developed for models, manipulative experiments and monitoring programs; invitation to join network widely distributed; follow-up workshop preparation
	Publication of Paris Workshop Report (GCTE Working Document II, Ingram, 1994)		
February 1995	Network Planning Workshop (Corvallis, US). Task Working Group meeting	Model sensitivity analysis planned for water erosion at plot scale. Four Core Research Projects (CRPs) planned; criteria for network membership established; inaugural membership agreed; Special Issue of *J. Soil & Water Conservation* designed	Preparation of inaugural members' metadata for submission to GCTE Scienti c Steering Committee for rati cation as GCTE Core Research; plot scale model sensitivity analysis workshop planning; JSWC Special Issue editing
April 1995	GCTE Scienti c Steering Committee meeting (Bariloche, Argentina)	Soil Erosion Network (GCTE-SEN) formally recognized as GCTE 'Core Research'	
September 1995	Network Workshop: 'Global Change: Modeling Soil Erosion by Water' (with NATO; Oxford, UK). Task Working Group meeting	Results of model sensitivity analysis reported and outlines of NATO ASI book chapters produced. Model sensitivity analysis planned for water erosion at catchment scale	Editing NATO ASI book chapters; catchment scale model sensitivity analysis workshop preparation
September 1995	Publication of revised Focus 3 Operational Plan including revisions to Task		
January 1996	Network Workshop:		
	'SALT' Experimentation Planning (Senegal, Burkina, Faso & Niger). Task Working Group meeting	Field collaboration planned along SALT transect (see Chapter 4)	
April 1996	Publication of Soils Activity (Activity 3.3) Implementation Plan (GCTE Report 12)		Editing metadata. Distribution of A3.3 Implementation Plan to current and prospective

Table 1 (*continued*)

Date	Event	Outcomes/product	Ongoing activities commenced
			members of GCTE-SEN; selected projects invited to contribute to GCTE-SEN; detailed planning for Utrecht workshop
October 1996	Publication of ve papers as Special Issue of *Journal of Soil and Water Conservation*, 51		Collaborative work on water erosion at eld-scale started in Senegal
December 1996	Publication of metadata for models, experiments and modeling programs (GCTE Report 6)		
Early 1997	Publication of Oxford NATO workshop book in NATO-ASI Series		
April 1997	Network Workshop: 'Modeling water erosional at catchment scale' (Utrecht, NL)	Series of papers for publication in book or journal special issue anticipated	Editing Utrecht papers. Detailed planning of GCTE presentations at Wind Erosion Congress, GCTE/LUCC Open Science Conference, Activity 3.3 Session at ISSS Congress, and Focus 3 Science Conference
November 1997	Network Workshop: 'Erosion Thresholds' (Almeria, Spain)	Erosion thresholds Core Research Project (CPR2) launched. Detailed plans for GCTE Erosion Network input to GCTE/LUCC Open Science Conference and GCTE Special Session at ISSS Congress; and proposals for input to Focus 3 Conference	
March 1998	GCTE/LUCC Open Science Conference (Barcelona, Spain)		Focus 3 session to include reporting of Soils Activity results
August 1998	ISSS Congress: GCTE Activity 3.3 Session (Montpellier, France)	Series of papers as Journal Special Issue	Editing ISSS Special Session papers
September 1999	Focus 3 Science Conference (Reading, UK)	To include a full review and synthesis of GCTE soil erosion research	Editing Focus 3 Conference papers

activities and by maximizing participation in synthesis workshops and publications.

- Both benefits and responsibilities of network membership must be clearly stated.

Table 1, based on the work of the GCTE Soil Erosion Network, is an excellent example of the activities and products of a successful network (Ingram *et al.*, 1998).

The examples below show the types of networks being used, some of the activities which networks use to meet their objectives, and the enhanced scientific understanding achievable through the network approach.

FLUXNET

Determining the pattern, strength and variability of the terrestrial sources and sinks of CO_2 is much more difficult

than for the ocean, due to the natural heterogeneity of the land surface as well as to human disturbances of the structure and functioning of terrestrial ecosystems. An IGBP initiative, led by the BAHC core project and involving GCTE and IGAC as well, aims to measure the exchange of water, energy and carbon between the land surface and the atmosphere at representative sites around the world, that is, to monitor the breathing of the planet's terrestrial biosphere. The project, called FLUXNET, is built around major independent flux-measuring networks being developed in Europe and North America, together with stations being established in other parts of the world. Figure 3 shows the current international array of flux-measuring stations (Baldocchi *et al.*, 1996).

FLUXNET demonstrates the utility of the network approach as a value-adding tool within IGBP. By forming an international network rather than operating as independent national or regional networks, several important advantages have been realized. First, a common measurement methodology, that of the EUROFLUX network (see http://www.unitus.it/eflux/euro.html for more information on EUROFLUX) has been adopted worldwide, thus greatly enhancing the intercomparability of network results. Second, a single international network will lead to a much more effective placement of future measuring sites, especially in the developing regions of the world. Choice of sites can more readily be based on scientific criteria, rather than on bilateral aid arrangements or more narrow national or regional considerations. Third, establishment and maintenance of an active international network promotes analysis and interpretation based on the results of the entire network, rather than on results of national or regional components of it.

The IGBP framework has added value to the FLUXNET enterprise in other ways as well. For example, collaboration with GCTE is leading to the choice of more representative sites in terms of ecosystem types, successional stages of vegetation, and degrees of human disturbance. In addition, the addition of ecological expertise to the network enhances the analysis and interpretation of results, especially with respect to the time dimension. Fluxes based on only a few years of measurements cannot usually be extrapolated to decade, or longer, fluxes in a straightforward fashion. Collaboration with IGAC will facilitate the use of the results in atmospherically based inverse calculations of carbon sources and sinks, which currently suffer from a lack of surface measurements to constrain the models.

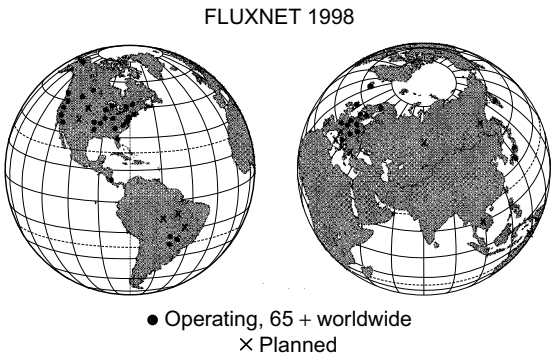


Figure 3 Location of FLUXNET sites (Baldocchi *et al.*, 1996). ●, Operating FLUXNET stations in 1998; ×, planned stations

LOICZ Biogeochemical Model Project

Another good example of the power of the networking approach, especially in terms of the adoption of common methodologies, is the LOICZ biogeochemical model project (Gordon *et al.*, 1995). This initiative demonstrates how a network based on a large number of individual investigators working at the site-specific level can, by the use of agreed techniques, yield results that can be readily scaled to yield regional and global understanding of the dynamics of biogeochemically important elements in the coastal zone. The common methodology was developed through a consensus process of all of the participating scientists in a series of international workshops. Although a suite of modeling techniques could have been adopted, the group opted initially for a simple approach based on stoichiometrically linked carbon, nitrogen and phosphorus budgets requiring relatively little data. This allows many more scientists and sites around the world to be included in the project. These simple but robust budget models of carbon, nitrogen and phosphorus fluxes can be used as the basis for developing more complex dynamic simulation models that can be used to explore the impacts of global change on important biogeochemical fluxes in the coastal zone.

GCTE Elevated CO_2 Network

Of the various IGBP core projects, GCTE has made the most extensive use of the network approach (Ingram *et al.*, 1998). One of the earliest networks to be formed in GCTE is that of researchers studying the effects of elevated CO_2 on terrestrial ecosystems, a network which has produced an impressive list of results in just seven years of operation. Established in 1991, the focus of the network is on ecosystem-level experiments and the analysis and interpretation of their results. Prior to the 1990s, there had been a large body of research on the effects of elevated atmospheric CO_2 on individual plants and small stands of plants, almost always in laboratories, greenhouses or other enclosures, but very little work on whole ecosystems *in situ*. The GCTE network aims to redress that balance.

Undertaking ecosystem-level elevated CO_2 experiments is challenging technically, and one of the first tasks of the GCTE network has been the evaluation and promotion of methodologies for undertaking such work,

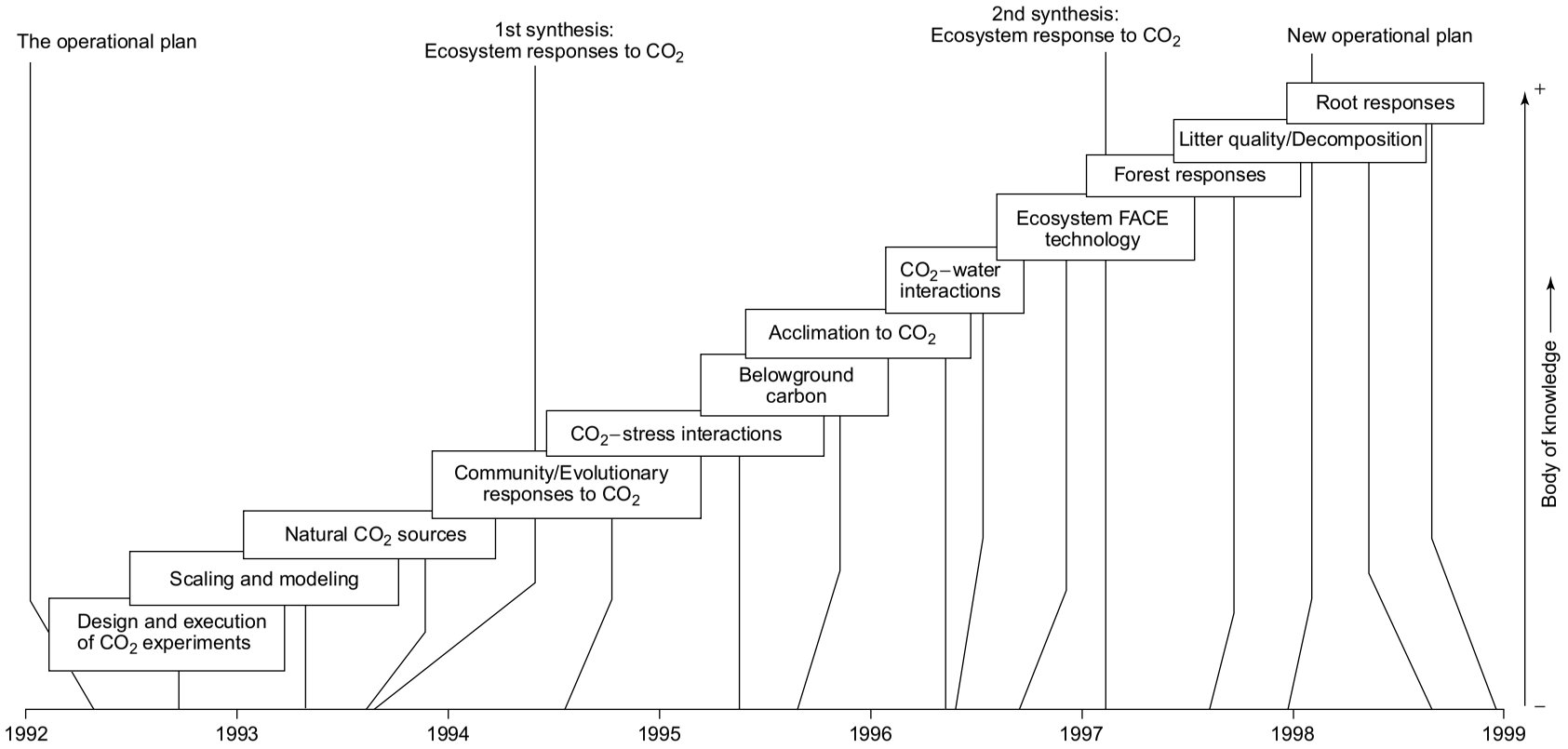


Figure 4 GCTE sponsored or co-sponsored activities in elevated CO_2 research, 1992- 1999. The increase in knowledge over time is enhanced by network activities; the network as a whole is greater than the sum of its constituent research projects. (Figure courtesy of Josep Canadell)

especially the FACE (Free Air Carbon Dioxide Enrichment) approach. This method does not use an enclosure of any kind and thus allows the ecosystem to interact with the ambient atmosphere without obstruction. FACE technology is complex and expensive, so the GCTE network has worked with principal investigators and agencies to move towards a more representative system of FACE sites around the world.

The elevated CO_2 network has operated in a rather less tightly organized way compared to FLUXNET in that no standardized methodologies have been prescribed. Nevertheless, as the network has developed, individual research groups have voluntarily moved towards more readily comparable experimental protocols and practices. The voluntary nature of network activities has worked remarkably well in terms of producing integrated scientific results. Network members have taken turns in hosting workshops on various focused aspects of elevated CO_2 effects, leading to a significant number of books and special journal editions based on network activities. The overall result is that the community's knowledge base has improved rapidly during the 1990s, much more rapidly than if the individual projects in the network had carried out their research independently (see Figure 4).

The most important findings include (Mooney *et al.*, 1998):

- Most whole ecosystems show a higher peak season net carbon uptake under elevated CO_2 compared to current ambient conditions. Above-ground productivity of grasslands increased by an average of about 15%, although there was wide variation among individual grassland communities, with some responses negative (Figure 5).

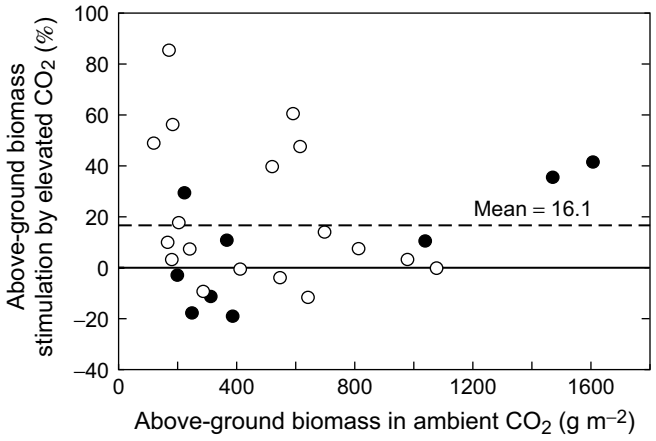


Figure 5 Stimulation of above-ground biomass of herbaceous systems by elevated CO_2 with and without addition of nitrogen (Reproduced by permission of Cambridge University Press in Mooney *et al.*, 1998). ●, Nitrogen addition; ○, no nitrogen addition

- Litter from plants grown under elevated CO_2 does not necessarily decompose more slowly. The ratio of carbon to nitrogen in the litter generally is not higher under elevated CO_2, as earlier predicted. This result has important implications for the terrestrial carbon cycle in general, as an important process in the cycle is the decomposition of soil organic carbon, derived from plant litter.
- Herbaceous plants show a reduced loss of water through transpiration under elevated CO_2 compared to current ambient conditions. This results in reduced loss of soil moisture, which in turn increases water availability to the vegetation. This secondary effect of elevated CO_2 is the dominant driver for increased carbon uptake in water-limited grassland systems.

Core Project Integration and Synthesis Activities

As the core projects progress further into their implementation phase, they are beginning to undertake within-project integration and synthesis activities. These represent significant leaps in understanding, and the examples described below have all had significant impacts on their own research areas.

PAGES Synthesis of Temperature Trends and Forcing Functions

One of the most intriguing questions in global change science is whether human perturbation of the climate is discernable. A synthesis of paleo-records provides some of the strongest evidence yet that humans are having a direct and measurable impact on climate (Mann *et al.*, 1998). The work is centered on integration of multi-proxy records – from tree rings, coral, ice core/melt and instrumental records – of Northern Hemisphere temperature from 1400 to the present, as shown in Figure 6. One of the most striking features of this record is the sharp increase in temperature in the 1990s. The three hottest years since 1400 occurred in this decade; their mean annual temperatures are greater than those of any other year since 1400 at a 99.7% level of certainty. Even more striking is the correlation of this temperature record with a corresponding record of three climatic forcing functions – atmospheric CO_2 concentration, solar irradiance and DVI, the latter an indication of aerosol inputs to the atmosphere from explosive volcanism (*see* **DVI (Dust Veil Index)**, Volume 3). Using a 200-year moving window to correlate the response of the climate to each of the three forcing variables, Mann *et al.* (1998) show that until the second half of the 20th century, solar irradiance was the primary driver of climate variability, with the dips in solar irradiance in the late 1600s and early 1800s and the rise through the first half of the 20th century clearly evident

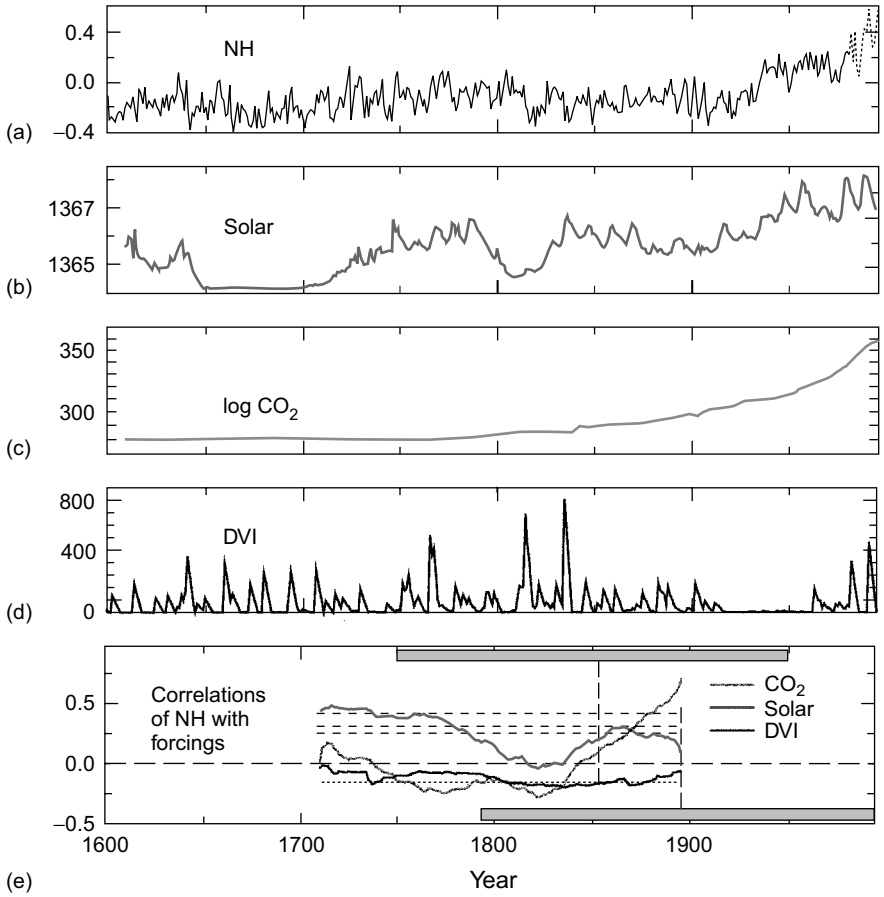


Figure 6 Relationships of Northern Hemisphere (NH) mean temperature with three candidate forcings between 1610 and 1995. (a) Reconstructed NH temperature series from 1610- 1980, updated with instrumental data from 1981- 1995. (b) Reconstructed solar irradiance. (c) Greenhouse gases represented by atmospheric CO_2 measurements. (d) Weighted volcanic dust veil index (DVI). (e) Evolving multivariate correlation of NH series with the three forcings solar, log CO_2, DVI. (Reproduced by permission of *Nature* in Mann *et al.*, 1998)

in the temperature record. However, the sharp increase in temperature in the last two decades does not coincide with a period of increase in solar irradiance, but rather correlates strongly with the anthropogenically driven rise in atmospheric CO_2 concentration.

The GCTE Terrestrial Carbon Cycle Synthesis

As part of its project-wide synthesis effort, GCTE analyzed the key terms of the terrestrial carbon budget, based on work carried out within GCTE networks, and projected likely trends in the terms (Walker *et al.*, 1998b). Three major areas of research were considered:

- *Land-use/cover change*: with the human population expected to increase by a billion every 12 years or so for the next several decades, most projections of land-use and land-cover change (e.g., Alcamo, 1994; Alcamo *et al.*, 1996) show increasing conversion of natural ecosystems to agriculture. In most cases conversion of ecosystems to agriculture or intensification of agriculture on land currently used for production leads to increased emissions of carbon from the vegetation removed in the conversion process and subsequently from the soil (e.g., Donigan *et al.*, 1995). Thus, the overall rate of carbon emissions from LUCC is likely to be maintained at current levels or increase for the foreseeable future.
- *Ecosystem physiology*: the GCTE synthesis analyzed the following factors:

  1. CO_2 fertilization, which is not likely to be as strong as earlier predicted;
  2. nitrogen deposition, which is likely to continue to act as a fertilizer for northern hemisphere forests, enhancing carbon uptake;
  3. nutrient limitations, which could be relieved as temperature increases and mobilizes nutrients through decomposition;
  4. physiological saturation, which will lead to reduced carbon uptake as global change proceeds

through a diminishing response of photosynthesis to increasing atmospheric CO_2 concentration and an increased release of carbon through decomposition as temperature increases.

- *Ecosystem structural change*: under global change, biomes will likely change through enhanced mortality of their current vegetation followed by establishment of a new vegetation assemblage. Mortality results in a fast release of carbon to the atmosphere while uptake of carbon by the new vegetation assemblage is a slower process. This mismatch in rates may lead to a significant transient pulse of carbon to the atmosphere (e.g., Smith and Shugart, 1993).

Based on likely trends in these areas, GCTE argued that the present rate of carbon uptake by terrestrial ecosystems, on a global scale, would be difficult to maintain. It projected that the terrestrial biosphere as a whole, including land used for production systems, would decrease in its ability to absorb carbon and may even become a net source of carbon sometime in the 21st century (Walker *et al.*, 1998b). This projection generated considerable controversy in the global change research community, but the results of more recent modeling studies are consistent with the GCTE projections (Cao and Woodward, 1998).

GAIM Model Intercomparisons

GAIM aims to integrate new understanding on biogeochemical cycling across the core projects. As a necessary precursor to that challenge, GAIM has first focused on compartments of the Earth System, undertaking model intercomparisons that are designed to test and improve our understanding of how important processes respond to anthropogenic perturbations.

Model intercomparisons are one of the most important of IGBP's value-adding activities. They accelerate model development by contrasting the varying degrees of success of a suite of models, which in turn leads to the identification of the causes of strengths and weaknesses of individual models. This process drives a more rational model adjustment and reformulation. In addition, model intercomparisons provide improved estimates of the uncertainty of model projections.

Below three GAIM-led model intercomparisons are described.

- *Ocean Carbon–Cycle Model Intercomparison*: the oceans are potentially the largest carbon sink on Earth, and understanding the partitioning of carbon between the oceans and the atmosphere is crucial to projecting future radiative forcing of climate. GAIM's ocean model intercomparison aims to improve 3D marine carbon cycle models for use in predicting future ocean–atmosphere partitioning of CO_2, and provide estimates of important but poorly known quantities relevant to the marine carbon cycle, such as new production, aphotic zone remineralization, and anthropogenic CO_2 fluxes and inventories. The first phase of the intercomparison forced the models with the history of pCO_2 change from 1765 (pre-industrial) to the present, based on ice core records and direct measurements. The results of the four models which participated in the project showed a good agreement on the overall oceanic uptake of carbon (a range of $1.5–2.2\,Pg\,C\,year^{-1}$), but they differed by more than 100% in their simulation of uptake of anthropogenic CO_2 by the Southern Ocean. This is a crucial shortcoming as the Southern Ocean is the largest oceanic sink of carbon and a good understanding of the processes and patterns of carbon uptake in this vast region is essential for projecting future atmospheric concentrations of CO_2. Comparison of the model output with measurements of pCO_2 (see the section on JGOFS pCO_2 Measurements) will also be important in improving model performance.
- *Atmospheric Transport Model Intercomparison*: the ultimate aim of this model intercomparison is to diagnose and quantify the uncertainty in inversion calculations of the global carbon budget due to errors in simulated transport. The first phase of the project, involving 12 atmospheric transport models, simulated the response of atmospheric CO_2 to anthropogenic emissions and to seasonal exchange with the biosphere. The results showed surprising variation in the interhemispheric concentration difference (a measure of transport of CO_2 from the Northern to the Southern Hemisphere) and in the simulation of the annual mean response to the biosphere exchange. Phase 2 of the project was based on intercomparison of the distribution of sulfur hexafloride (SF_6), which has slowly varying sources, no sinks (unlike CO_2), and is fairly well observed around the world. Evaluation of the outcomes of both phases suggests that the variation of the North–South gradient at the surface (where CO_2 concentration is measured) is not controlled primarily by interhemispheric mixing but rather by vertical mixing, which occurs at unresolved spatial scales and must be parameterized in the models. The third phase will intercompare inversion calculations of the atmospheric CO_2 budget, which estimate carbon fluxes from the observed distribution of measured atmospheric CO_2 concentrations. In the post-Kyoto era with its emphasis on terrestrial carbon sinks, such inverse calculations can lead to controversial estimates of regional sink strengths (e.g., Fan *et al.*, 1998). The GAIM-led intercomparison of inverse calculations hopes to produce a reliable estimate of the uncertainty in these calculations, especially the degree

of uncertainty due to the modeling of atmospheric transport.

- *Global Net Primary Productivity*: although net primary productivity (NPP) cannot be easily equated to long-term carbon storage in terrestrial ecosystems, it is a critical process in evaluating the role of the terrestrial biosphere in Earth System functioning. The GAIM NPP intercomparison, carried out in conjunction with GCTE and DIS, is the largest and most ambitious model intercomparison carried out in IGBP. It involved 17 modeling groups, two intensive workshops (in 1994 and 1995) and extensive analysis and interpretation following the formal intercomparisons. The biospheric productivity models participating in the project are complex. They relate geographically specific and comprehensive estimates of solar radiation, temperature and precipitation to the basic biospheric processes of photosynthesis, growth and maintenance respiration, water and nitrogen fluxes, allocation of assimilates within ecosystems, and the production and decomposition of litter. In addition to the driving variables of climate and radiation, soil characteristics and vegetation structure also have strong influences on the simulation of NPP, and these must either be specified or, in the case of vegetation structure for some models, simulated. The models simulated a global NPP of the terrestrial biosphere from 39.9 to 80.5 Pg C year^{-1} with an average of 54.9 Pg C year^{-1}. Despite the very significant differences in model philosophies and formulations, there was surprisingly good agreement on this value, leading to speculation that perhaps explicit or unconscious calibration to an assumed global NPP has occurred. It was significant that four of the six models which included nutrient constraints on NPP predicted lower global NPP values than the other models, and four of the six models which used a vapor pressure deficit calculation to determine NPP projected a higher global NPP. An important next step is to compare the model output with measured NPP, and to facilitate this, GAIM has begun the global primary production data initiative to collect observed estimates of NPP and to develop improved methodologies for acquiring and distributing these data.

IGBP-wide Integration and Synthesis Activities

As IGBP matures and more pieces of global change research yield new understanding in specific areas, there is more incentive to put the pieces of the puzzle together in new ways. Thus, integration and synthesis activities among sets of core projects, and across the program as a whole, are playing an increasingly important role in IGBP science.

BAHC Earth System Modeling

A first attempt to integrate the major components of the Earth System within a modeling framework is the development of intermediate complexity models to simulate Earth System functioning, undertaken by BAHC in collaboration with GAIM (Ganopolski *et al.*, 1998). The initial model, developed at the Potsdam Institute of Climate Impact Research, consists of coarse resolution but fully coupled atmospheric, oceanic and terrestrial vegetation components. The model can be run holding each of the components static in turn to determine the impacts of each component individually, and of two and three components synergistically. The results for present-day climate forcing and for the orbital forcing of 6000 years ago are striking. They show that the earlier notion that the atmosphere and oceans together largely determine climate and that the terrestrial biosphere is mainly a recipient of, rather than a player in, climate change is most likely wrong. In both time frames, the climate simulated when all three components are fully dynamic and interactive is much closer to the observed climate than when any component is held constant. The terrestrial biosphere is thus an important player in global climate, synergistically with oceans and climate. In particular, the results show the importance of a warming-driven northward expansion of boreal forest in the northern high northern latitudes leading to further warming via vegetation–snow–albedo feedback, and strong positive feedback between vegetation and precipitation in the Sahara region. In a broader sense, the simulations suggest that the Earth system may exist in semi-stable states, with sharp transitions between the states, a phenomenon that also features in ice core records (e.g., Alley *et al.*, 1993).

IGBP Terrestrial Carbon Working Group

Comprised of representatives from BAHC, GAIM, GCTE, IGAC and LUCC, the IGBP Terrestrial Carbon Working Group is an example of IGBP's ability to assemble multi-disciplinary teams on short notice to tackle pressing problems integrating science and policy. The working group was formed in March 1998 at a joint GCTE/LUCC Open Science Conference, it met in Stockholm in late April 1998, and produced in May 1998 an authoritative commentary on the implications of the science of the terrestrial carbon cycle for the Kyoto Protocol (IGBP Terrestrial Carbon Working Group, 1998). The paper was produced on a rapid timeframe as input to an important meeting of the Subsidiary Body on Scientific and Technical Advice to the Framework Convention on Climate Change.

The objective of the working group was to analyze our current understanding of the terrestrial carbon cycle, and, based on that understanding, provide an informed commentary on the implications of the Kyoto Protocol. The

Table 2 IGBP/DIS Land Cover Classi cation

1. Evergreen needleleaf forests: lands dominated by trees with a percent canopy cover >60% and height exceeding 2 m. Almost all trees remain green all year. Canopy is never without green foliage
2. Evergreen broadleaf forests: lands dominated by trees with a percent canopy cover >60% and height exceeding 2 m. Almost all trees remain green all year. Canopy is never without green foliage
3. Deciduous needleleaf forests: lands dominated by trees with a percent canopy cover >60% and height exceeding 2 m. Consists of seasonal needleleaf tree communities with an annual cycle of leaf-on and leaf-off periods
4. Deciduous broadleaf forests: lands dominated by trees with a percent canopy cover >60% and height exceeding 2 m. Consists of seasonal broadleaf tree communities with an annual cycle of leaf-on and leaf-off periods
5. Mixed forests: lands dominated by trees with a percent canopy cover >60% and height exceeding 2 m. Consists of tree communities with interspersed mixtures or mosaics of the other four forest cover types. None of the forest types exceeds 60% of landscape
6. Closed shrublands: lands with woody vegetation less than 2 m tall and with shrub canopy cover >60%. The shrub foliage can be either evergreen or deciduous
7. Open shrublands: lands with woody vegetation less than 2 m tall and with shrub canopy cover between 10 and 60%. The shrub foliage can be either evergreen or deciduous
8. Woody Savannahs: lands with herbaceous and other understory systems, and with forest canopy cover between 30 and 60%. The forest cover height exceeds 2 m
9. Savannahs: lands with herbaceous and other understory systems, and with forest canopy cover between 10 and 30%. The forest cover height exceeds 2 m
10. Grasslands: lands with herbaceous types of cover. Tree and shrub cover is less than 10%
11. Permanent wetlands: lands with a permanent mixture of water and herbaceous or woody vegetation that cover extensive areas. The vegetation can be present in either salt, brackish, or fresh water
12. Croplands: lands covered with temporary crops followed by harvest and a bare soil period (e.g., single and multiple cropping systems). Note that perennial woody crops will be classi ed as the appropriate forest or shrub land cover type
13. Urban and built-up: land covered by buildings and other structures. Note that this class will not be mapped from the AVHRR imagery but will be developed from the populated-places layer that is part of the Digital Chart of the World (Danko, 1992)
14. Cropland/natural vegetation mosaics: lands with a mosaic of croplands, forests, shrublands, and grasslands in which no one component comprises more than 60% of the landscape
15. Snow and ice: lands under snow and/or ice cover throughout the year
16. Barren: lands exposed to soil, sand, rocks, or snow and never having more than 10% vegetated cover during any time of the year
17. Water bodies: oceans, seas, lakes, reservoirs, and rivers. Can be either fresh or salt water bodies

working group pulled together IGBP and related research on the terrestrial carbon budget, the processes that control the terms in the budget, methodologies for measuring changes in terms of the budget and their implications for carbon accounting techniques, and the long-term dynamics of the terrestrial carbon cycle.

The formation of the working group and the publication of the critique of the Kyoto Protocol was the first stage in the formation of an IGBP-wide working group on the carbon cycle, adding the coastal, atmospheric and marine components to the existing terrestrial group. The objectives of the combined working group are to produce an up-to-date understanding of the carbon cycle as a whole, emphasizing and synthesizing the work within IGBP core projects on the carbon cycle, and to develop an agreed framework for a coordinated international, long-term effort on the dynamics of the global carbon cycle.

DIS Land Cover Classi cation

An integrating activity of a different type is the production by DIS of a land cover classification scheme for wide use in IGBP's terrestrially-oriented core projects. The lack of a common classification scheme for vegetation (e.g., biome classification) or land-cover often vitiates global change research on the terrestrial surface. Global vegetation models often define their own schemes, which are different from those adopted for remotely sensed databases, which in turn are invariably different from those of *in situ* databases. Intercomparison of models or comparison of models with databases is virtually impossible. DIS is attempting to rectify this situation within the IGBP family by the development of a single classification scheme, based on broad consultation with IGBP core projects (see Table 2). As part of the GCTE synthesis, this new classification scheme was used to compare global vegetation

and land cover models with each other and with databases (Walker *et al.*, 1998b).

TRENDS IN IGBP SCIENCE

As the more mature IGBP core projects (BAHC, GCTE, IGAC, JGOFS, PAGES) reach a decade of research, some underlying trends in global change science are becoming apparent:

- The original IGBP core projects were largely organized around compartments of the Earth system – the atmosphere (IGAC), the oceans (JGOFS), the terrestrial land surface (GCTE) and the coastal zone (LOICZ). As the research has progressed, however, there is growing interest in the interfaces between these compartments and in processes that move materials and energy between the compartments. Examples include the exchange of trace gases between the lower atmosphere and upper oceans, a corresponding exchange between the terrestrial biosphere and the troposphere, and the lateral transport of materials from inland terrestrial ecosystems through the coastal zone to the coastal seas.
- The original emphasis within IGBP was almost exclusively on fundamental Earth System science; only GCTE had a significant component (Focus 3: Global Change Impact on Agriculture, Forestry and Soils) on the biophysical consequences of changes in the Earth System to sectors of importance for human well-being. As IGBP has evolved, there is now a stronger emphasis on impacts and consequences of global change, building on a continuing solid core of basic Earth system science. All of the newer core projects – LOICZ, LUCC, and GLOBEC – have major components on impacts, and two of the older core projects – BAHC and IGAC – are placing much more emphasis on impacts of changes to their components of the Earth System (on water resources and on the human and ecosystem health effects of global atmospheric pollution, respectively) as they restructure.
- WCRP was founded nearly a decade before IGBP, with an obvious focus on the climatic part of the Earth System. As the IGBP core projects have matured and the relationship of climate and Earth System science has been more clearly recognized, collaborative research between IGBP and WCRP core projects has intensified. Examples include the Global Energy and Water Cycle Experiment (GEWEX), the BAHC work on the hydrological cycle, the joint WOCE/JGOFS oceanic cruises, the rapidly expanding interaction between the SPARC (Stratospheric Processes and their Role in Climate) and IGAC, and the development of joint work on climate variability involving the CLIVAR (Climate Variability Research Project) and PAGES. See Box 1.
- At the inception of IGBP, the human dimensions were recognized as a critically important aspect of global change but the international framework for a social science global change research program was not yet in place. By the late 1990s, however, the IHDP had developed rapidly into a focused, well-defined international research effort. Interfaces between IGBP and IHDP have subsequently developed strongly.
- The regional perspective on global change has gained prominence in the last several years as the policy and resource management implications of global change research have become increasingly appreciated. Impacts and adaptation are largely considered at regional and national, not global, scales. See Box 2.

THE FUTURE OF THE IGBP CORE PROJECTS

Over the 1998–2001 period, IGBP undertook a major integration and synthesis of its first decade of research. Each of the mature core projects – BAHC, GCTE, IGAC, JGOFS and PAGES – completed a synthesis of its area of global change science, with GCTE's effort being an update of an earlier synthesis (Walker *et al.*, 1998a). In addition to the core project efforts, there was an IGBP-wide synthesis, on the theme of the Earth System as an integrated whole. In each of the themes, the synthesis addressed both the fundamental Earth System science and the impacts and consequences for human societies.

As an integral part of the synthesis effort, the structure of IGBP is evolving to address the needs of the first

Box 1 Collaborative programs between IGBP and IHDP

Examples include joint work between IHDP's Global Environmental Change and Human Security (GECHS) program and IGBP on water resources (with BAHC) and on food security (with GCTE, LOICZ and GLOBEC), as well as interaction between the Institutional Dimensions of Global Environmental Change (IDGEC) core project of IHDP and LUCC and GCTE. LUCC itself is jointly sponsored by IGBP and IHDP.

Box 2 Examples of integrated regional programs within IGBP

Many core projects are now involved in integrated regional projects under the auspices of START, and several of the core projects - LOICZ, LUCC and GLOBEC - have a strong regional orientation to their implementation strategies. The IGBP Terrestrial Transects program is a good example of a regional approach to global change research, employing many of the value-adding activities described above (Koch *et al.*, 1995; Steffen *et al.*, 1998).

decade of the 21st century. The restructure will evolve naturally as the synthesis proceeds, based on the new scientific understanding and the new questions and gaps generated by the synthesis, taking into account the trends described above. Thus in the coming decade, the core projects within IGBP will almost surely be considerably different from the configuration of the 1990s, reflecting the changing scientific demands of a dynamic research program.

FURTHER INFORMATION

More information on IGBP research can be obtained from the IGBP Secretariat or from the international offices of the core projects:

IGBP Secretariat, Royal Swedish Academy of Sciences, Box 50005, SE-10405 Stockholm, Sweden, Email: sec@igbp.kva.se

Homepage: http://www.igbp.kva.se/

BAHC International Project Office, Potsdam Institute for Climate Impact Research, PO Box 601 203, D-14412 Potsdam, Germany, Email: bahc@pik-potsdam.de

Homepage: http://www.pik-potsdam.de/~bahc

GAIM Task Force Office, Institute for the Study of Earth Oceans and Space (EOS), University of New Hampshire, Morse Hall, 39 College Road, Durham, NH 03824-3525, USA, Email: gaim@unh.edu

Homepage: http://gaim.unh.edu

GCTE International Project Office, CSIRO Sustainable Ecosystems, PO Box 284, Canberra, ACT 2600, Australia, Email: pep.canadell@dwe.csiro.au

Homepage: http://GCTE.org

GLOBEC International Project Office, Plymouth Marine Laboratory, Prospect Place, Plymouth PL1 3DH, UK, Email: m.barange@pml.ac.uk

Homepage: http://www.npm.ac.uk/globec/

IGAC International Project Office, Massachusetts Institute of Technology, Building 24-409, Cambridge, MA 02139, USA, Email: pszenny@mit.edu

Homepage: http://web.mit.edu/igac/www/

JGOFS International Project Office, Center for Studies of Environment and Resources, University of Bergen, High Technology Centre, N-5020 Bergen, Norway, Email: jgofs@uib.no

Homepage: http://ads.smr.uib.no/jgofs/jgofs.htm

LOICZ International Project Office, Netherlands Institute for Sea Research, PO Box 59, NL-1790 AB Ben Burg-Texel, The Netherlands, Email: loicz@nioz.nl

Homepage: http://www.nioz.nl/loicz

LUCC International Project Office, University of Louvain, Department of Geography, 3, Place Louis Pasteur, B-1348 Louvain-la-neuve, Belgium, Email: lucc@geog.ucl.ac.be

Homepage: http://www.uni-bonn.de/ihdp

PAGES International Project Office, Bärenplatz 2, CH-3011 Bern, Switzerland, Email: pages@pages.unibe.ch

Homepage: http://www.pages.unibe.ch/

International START Secretariat, Suite 200, 2000 Florida Avenue, NW, Washington, DC 20009, USA, Email: start@agu.org

Homepage: http://www.start.org

REFERENCES

Alcamo, J, ed (1994) *IMAGE 2.0: Integrated Modeling of Global Change*, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Alcamo, J, Kreileman, E, and Leemans, R, eds (1996) Integrated Scenarios of Global Change, *Global Environ. Change* (Special Issue), **6**, 255–394.

Alley, R B, Meese, D A, Shuman, C A, Gow, A J, Taylor, K C, Grootes, P M, White, J W C, Ram, M, Waddington, E D, Mayewski, P A, and Zielinski, G A (1993) Abrupt Increase in Greenland Snow Accumulation at the End of the Younger Dryas Event, *Nature*, **362**, 527–529.

Baldocchi, D, Valentin, R, Running, S, Oechel, W, and Dahlman, R (1996) Strategies for Measuring and Modeling Carbon Dioxide and Water Vapor Fluxes Over Terrestrial Ecosystems, *Global Change Biol.*, **2**, 159–168.

Cao, M and Woodward, F I (1998) Dynamic Responses of Terrestrial Ecosystem Carbon Cycling to Global Climate Change, *Nature*, **393**, 249–252.

Donigan, Jr, A S, Patwardhan, A S, Jackson, IV, R B, Barnwell, Jr, T O, Weinrich, K B, and Rowell, A L (1995) Modeling the Impacts of Agricultural Management Practices on Soil Carbon in the Central US, in *Soil Management and Greenhouse Effect*, eds R Lal, J Kimble, E Levine, and B A Stewart, CRC Press, Boca Raton, FL.

Fan, S, Gloor, M, Mahlman, J, Pacala, S, Sarmiento, J, Takahashi, T, and Tans, P (1998) A Large Terrestrial Carbon Sink in North America Implied by Atmospheric and Oceanic Carbon Dioxide Data and Models, *Science*, **282**, 442–446.

Ganopolski, A, Kubatski, C, Claussen, M, Brovkin, V, and Petoukhov, V (1998) The Influence of Vegetation–Atmosphere–Ocean Interaction on Climate During the Mid-holocene, *Science*, **280**, 1916–1919.

Geoghegan, J, Pritchard, Jr, L, Ogneva-Himmelberger, Y, Chowdhury, R R, Sanderson, S, and Turner, II, B L (1998) "Socializing the Pixel" and "Pixelizing the Social" in Land-use and Land-cover change, in *People and Pixels*, eds D Liverman, E F Moran, R R Rindfuss, and P C Stern, National Academy Press, Washington, DC, 51–69.

Gordon, Jr, D C, Boudreau, P R, Mann, K H, Ong, J-E, Silvert, W L, Smith, S V, Wattayakorn, G, Wulff, F, and Yanagi, T (1995) LOICZ Biogeochemical Modeling Guidelines, LOICZ Reports & Studies No. 5, LOICZ Core Project Office, Texel, The Netherlands.

IGBP Terrestrial Carbon Working Group (1998) The Terrestrial Carbon Cycle: Implications for the Kyoto Protocol, *Science*, **280**, 1393–1394.

Ingram, J S I, Canadell, J, Elliot, T, Hunt, L A, Linder, S, Murdiyarso, D, Stafford Smith, M, and Valentin, C (1998)

Networks and Consortia, in *Global Change and the Terrestrial Biosphere: Implications for Natural and Managed Ecosystems. A Synthesis of GCTE and Related Research*, eds B Walker, W Steffen, J Canadell, and J Ingram, IGBP Book Series, No. 4, Cambridge University Press, Cambridge, 45–66.

IPCC (1994) *Climate Change 1994: Radiative Forcing of Climate Change and an Evaluation of the IPCC IS92 Emission Scenarios*, eds J T Houghton, L G Meria Filho, J Bruce, L Hoesung, B A Callander, E Haites, N Harris, and K Maskell, Cambridge University Press, Cambridge.

Koch, G W, Vitousek, P M, Steffen, W L, and Walker, B H (1995) Terrestrial Transects for Global Change Research, *Vegetatio*, **121**, 53–65.

Mann, M E, Bradley, R S, and Hughes, M K (1998) Global-scale Temperature Patterns and Climate Forcing Over the Past Six Centuries, *Nature*, **392**, 779–787.

Mooney, H A, Canadell, J, Chapin, III, F S, Ehleringer, J R, Körner, C, McMurtrie, R E, Parton, W J, Pitelka, L F, and Schulze, E-D (1998) Ecosystem Physiology Responses to Global Change, in *Global Change and the Terrestrial Biosphere: Implications for Natural and Managed Ecosystems. A Synthesis of GCTE and Related Research*, eds B Walker, W Steffen, J Canadell, and J Ingram, IGBP Book Series, No. 4, Cambridge University Press, Cambridge, 141–189.

Raes, F and Bates, T S (1998) ACE-2 Field Phase Completed. IGAC NewsLetter, IGAC Core Project Office, Cambridge, MA, 18–19, Vol. 11.

Smith, T M and Shugart, H H (1993) The Transient Response of Terrestrial Carbon Storage to a Perturbed Climate, *Nature*, **361** 523–526.

Steffen, W L, Valentin, C, Scholes, R J, Zhang, X-S, and Menaut, J-C (1998) The IGBP Terrestrial Transects, in *Global Change and the Terrestrial Biosphere: Implications for Natural and Managed Ecosystems. A Synthesis of GCTE and Related Research*, eds B Walker, W Steffen, J Canadell, and J Ingram, IGBP Book Series, No. 4, Cambridge University Press, Cambridge, 66–87.

Takahashi, T, Feely, R A, Weiss, R F, Wanninkhof, R H, Chipman, D W, Sutherland, S C, and Takahashi, T T (1997) Global Air–Sea Flux of CO_2: an Estimate based on Measurements of Sea–Air pCO_2 Difference, *Proc. Natl. Acad. Sci.*, **94**, 8292–8299.

Walker, B H, Steffen, W L, Canadell, J, and Ingram, J S I, eds (1998a) *Global Change and the Terrestrial Biosphere: Implications for Natural and Managed Ecosystems. A Synthesis of GCTE and Related Research*, IGBP Book Series No. 4, Cambridge University Press, Cambridge, 1–432.

Walker, B H, Steffen, W L, and Langridge, J (1998b) Interactive and Integrated Effects of Global Change on Terrestrial Ecosystems, in *Global Change and the Terrestrial Biosphere: Implications for Natural and Managed Ecosystems. A Synthesis of GCTE and Related Research*, eds B Walker, W Steffen, J Canadell, and J Ingram, IGBP Book Series, No. 4, Cambridge University Press, Cambridge, 329–375.

FURTHER READING

Danko, D M (1992) The Digital Chart of the World, *Geoinfosystems*, **2**, 29–36.

Impacts of Global Environmental Change on Animals

JOEL G KINGSOLVER

University of North Carolina, NC, USA

Global change represents a diverse assortment of human-induced environmental changes that collectively impact animals around the globe. These diverse environmental changes will necessarily have diverse biological impacts on the millions of species of animals alive today. Recent studies of fossil vertebrates suggest that human colonization and population expansion has been causing extinctions over the past 10 000 years and more: even small human populations with stone-age technology can be suf cient to elevate extinction rates. These effects have been greatest for endemic and ecologically specialized animals; typically habitat destruction and introduced species rather than direct human exploitation have been the primary causes of extinction. Anticipated climate change during the 21st century will alter seasonal, diurnal and latitudinal patterns of variation in temperature and precipitation. Changes in seasonal components of climate will likely be most important in determining the extent of range shifts in animals, particularly by altering the relative seasonal timing of animals and their resources or enemies. Recent evidence from butter ies, birds and other systems suggests that decadal-scale climate changes during the past century have already caused range shifts or local extinctions in some cases. Climate change will likely negatively impact animal species that are localized or habitat specialists, but will cause expanded ranges towards the poles in many generalist species, including agricultural pests. Useful predictions of climate change impacts must incorporate simultaneous effects of changes in carbon dioxide (CO_2) and seasonal temperature and precipitation. The effects of changes in climate on vector-borne diseases may depend critically on how climate alters the seasonal patterns of interactions among parasites, vectors and hosts.

INTRODUCTION

Global change is a misnomer, implying some singular alteration of the earth's physical and biological characteristics. Instead, global change is an assortment of quite diverse changes – climate change, environmental toxins, habitat loss and alteration, invading pests and pathogens, harvesting of natural resources; the common attribute of these changes is that they are the direct or indirect products of human activities, and collectively their impacts are worldwide. Clearly these different environmental changes operate at quite different characteristic scales of time and space. For example, an application of insecticide may act over days to weeks at a spatial scale of hectares; loss of habitat for a rare or endangered species or community may occur over years at local to regional scales; atmospheric carbon dioxide (CO_2) increases globally over a time scale of decades. These diverse environmental changes will necessarily have diverse biological impacts.

Similarly, the term animals refers to a natural grouping (the Metazoa) of enormous diversity – perhaps 10–100 million distinct species, each representing a unique evolutionary lineage with a distinctive set of biological and ecological characteristics. Although collectively animals inhabit the entire globe, the distribution of particular species is typically local to regional: there are few truly global species, *Homo sapiens* being one notable exception. When we consider some particular aspect of environmental change, e.g., habitat fragmentation, the responses of different animal species to such change may depend on their particular biological, ecological and distributional characteristics. What is the appropriate level of detail needed to make useful predictions about animal responses?

In this essay we will take an ecological and population perspective on the responses of animals to global changes: on how global change alters the distribution and abundance of populations and species. Environmental changes will affect the development, physiology, behavior, growth or reproduction of individual animals, but our primary concern is how these effects will impact the persistence and distribution of a population or species into the future. In this sense we will consider the biology of individuals primarily

as potential mechanisms of population-level responses. An ecological perspective also emphasizes the relationship of animal populations to their resources and to other animal species: that is, the ecological community in which the population is embedded. How important are these species interactions for ecological responses to global changes?

In this essay we will not attempt to systematically cover all of the diverse aspects of the global environment that are changing as a result of human activities. Rather, we will explore two quite different case studies that provide distinct but complementary messages about the past, present and future responses of animals to global changes. First, we will examine the early history of one globally invading species – *Homo sapiens* – and its consequences for extinctions of birds, mammals and other vertebrates during the past 100 000 years. Here we will focus on recent fossil evidence to provide some historical perspective on the roots of human-induced (anthropogenic) environmental change. Second, we will examine in some detail how weather and climate impact the distribution and abundance of animals. Here we will focus on ecological field studies of animal populations during the past century, and combine these with scenarios of global climate change to explore our capacity to predict the ecological consequences of ongoing and future climate change. Throughout the discussion we will emphasize three general themes:

1. Generalist and widely dispersive species will fare better than more specialized species in the face of environmental changes.
2. Indirect effects of environmental changes as mediated through species interactions are frequently as or more important than direct effects on individual survival or reproduction.
3. The details of the time and spatial scale of environmental change and the biology of particular species may sometimes matter critically to understanding population responses of animals.

ANIMAL EXTINCTION AND HUMAN COLONIZATION

There is abundant historical evidence that colonization and settlement by Europeans during the past 500 years has greatly increased rates of extinction of animals, and rates of extinction have accelerated as human population size and technology have exploded during the past century (Wilson, 1992). But of course most regions of the globe had been colonized by human populations prior to the past half-millenium. Only recently have we begun to understand how these earlier events in human evolutionary history may have impacted animal extinction rates. What population sizes and level of technological sophistication in human populations are required to cause substantial extinctions of animals, and what sorts of species are most likely to go extinct?

Molecular and fossil evidence suggests that modern humans (*Homo sapiens*) first emerged 100 000–200 000 years ago in Africa, then migrated out to other areas of the world. It is clear that other forms of *Homo* occurred in Europe, the Near East, Asia as well as Africa by 500 000–1 million years ago; and modern humans inhabited each of these continents before 60 000 years ago. Humans and their hominid relatives have been a part of the biological landscape of these regions for a long time.

Human colonization has been more recent for the remainder of the globe, including New Guinea, Australia and New Zealand, the Pacific Islands, and North and South America. Human settlement of these differing areas occurred at different times, thus providing a series of natural experiments to explore the rates of extinction following the appearance of humans. We shall first consider recent evidence about extinctions of birds on Pacific Islands, and then examine extinctions of vertebrates more generally on the continents of Australia and North America.

Bird Extinctions on Pacific Islands

There are more than 800 major islands in the Pacific Ocean; nearly all of these were initially colonized by humans during the period from 35 000 to 1000 years ago. Many of these islands are small, and some island groups (e.g., Hawaii, Easter Island) are quite isolated. As a result, the land birds on many islands are highly endemic: many species occur on only a single island or island group. There have been repeated evolutionary radiations of finches (including the Hawaiian honeycreepers and the Galapagos finches), ibises, parrots, pigeons and doves, and ightless species of ducks, geese and rails. Many of the larger islands possess lava tubes or limestone caves as well as archaeological sites, that in the past two decades have provided a new picture of past and present bird diversity on Pacific Islands (Steadman, 1995).

Islands in the Kingdom of Tonga in western Polynesia were settled by humans before 3000 BP (before present). Caves on Eua and other islands within Tonga have yielded substantial bird fossils both before and after human settlement. During the 10 000 years prior to human settlement (a period of substantial climate and sea-level changes), there were virtually no changes or extinctions in bird species on Tonga. However, the majority of endemic birds went extinct in the period following human settlement: e.g., of 27 land bird species known from prehuman times, only six survived into the past two centuries (the historical period) (Steadman, 1995). The losses were particularly heavy among ground-dwelling and forest-dwelling species, probably re ecting predation by humans and by mammals introduced by humans, especially rats, dogs and pigs. In fact, archaeological sites suggest that native birds, especially pigeons,

doves and rails, were a major part of the diet for early human settlers on these islands. Similar patterns of extinction are seen on the Cook Islands and other east Polynesian islands; in addition, many species remaining today have much more restricted geographic distributions than prior to human settlement (Steadman, 1995).

The isolation and ecological diversity of the Hawaiian Islands generated many unique and endemic groups of plants and animals, including birds, which evolved quite independently from related groups in the Polynesian heartland. The prehistorical fossil record for land birds on Hawaii, which was first settled about 1500–2000 BP, is extensive and has been well studied (Olson and James, 1982). About 60 species of endemic land birds went extinct in the period following human settlement (1500–200 BP); another 20–25 species have gone extinct during the past two centuries. These losses were particularly great for ightless ducks, geese and rails and for the honeycreepers and other finches. The extinctions likely have a variety of causes, including predation or disturbance by humans and other introduced mammals (rats and pigs), deforestation and agriculture, and invasion by new diseases (including malaria) and generalized birds (including crows).

One thousand years ago, New Zealand possessed a diverse endemic bird fauna including kiwis, many waterfowl and 11 species of large, ightless moas. The earliest human settlements on New Zealand by Polynesian colonists date to the 13th century AD (Holdaway and Jacomb, 2000); European settlement began in the 19th century AD. At least 44 endemic land bird species have become extinct since human settlement, including 27 of the 50 land birds on the South Island. A recent analysis of archaeological sites and demographic modeling of the moas of New Zealand suggest that these large ightless birds went extinct only 100–200 years after Polynesian settlement, as a result of habitat destruction and human predation on moa eggs and adults (Holdaway and Jacomb, 2000). Some of the 14th century archaeological sites reveal large quantities of moa bones, representing several hundred tons of moa meat, during the first several decades after settlement. Demographic models suggest that moas were probably particularly susceptible to extinction, because they were long-lived, slow to reach reproductive maturity, and had long reproductive rates (Holdaway and Jacomb, 2000).

This pattern of human settlement followed by elevated extinction rates of endemic birds is largely from remote Pacific Islands (Steadman, 1995). However, there is also evidence that similar effects may occur in larger and less isolated islands. For example, New Ireland in Papua New Guinea is a large, heavily forested island off the Huon Peninsula of New Guinea. It has been occupied continuously by humans for the past 35 000 years. Recent archaeological studies suggest that 20–30% of the more than 100 endemic bird species of New Ireland went extinct since human settlement (Steadman *et al.*, 1999). Most of the extinctions occurred between 10 000 and 18 000 BP, though some extinctions were more recent (1000–5000 BP). Although much of the extinction occurred well after first human settlement, it is likely that humans played a major direct or indirect role in the extinctions, probably as a result of habitat disturbance and predation by humans, pigs, dogs, rats and other introduced mammals (Steadman *et al.*, 1999).

The clear picture that emerges from these studies is that the colonization of islands by prehistoric humans repeatedly accelerated extinction rates of endemic birds, even when human population sizes were quite small. In some cases direct human predation was a major factor, but in most cases indirect effects, primarily habitat destruction and the accidental or intentional introductions of rats, pigs, chickens, dogs and other animals, were probably the dominant causes. Extinction rates were highest, up to 50–90% of land bird species, in smaller and more remote island groups; however, there is some evidence that accelerated extinctions also occurred on larger and less remote islands.

Islands are special places, both ecologically and evolutionarily. Because of their small size and isolation, islands tend to have relatively fewer species and more endemic species than continental regions; they sometimes lack large predators and other key players that may have disproportionate effects on animal communities. In this sense, the ecological disruption and species extinctions on small, remote islands caused by small, prehistoric human settlements may not be so surprising; and perhaps these lessons about human colonization and extinction simply do not apply to large, continental regions. How could small bands of humans armed with stone tools cause widespread extinctions on a continent? To address this, we shall consider extinctions of large vertebrates on Australia and North America.

Late Pleistocene Extinctions on Continents

It has been known for more than a century that Australia suffered major losses of its terrestrial vertebrates during the Pleistocene. Larger species were particularly affected. Recent analyses suggest that more than 85% of all Australian land genera with a body mass exceeding 44 kg became extinct during the late Pleistocene, including all 19 marsupial species greater than 100 kg (Miller *et al.*, 1999). These large vertebrates are sometimes called the Australian megafauna. In total about 60 vertebrate species became extinct, including large varanid (monitor) lizards, horned tortoises, and an 80–100 kg ightless bird, *Genyornis newtonii*; two somewhat smaller ightless birds, the emu and the cassowary, survived.

Humans first settled parts of Australia by about 55 000 BP. However, the timing of the Pleistocene extinctions of

Australian megafauna was unresolved until the development of new methods for dating these fossils. One recent study used analyses of fossil eggshell fragments, which are easily identified to species, abundant, and readily dated, to explore changes in abundance of *Genyornis* and emus over time from three different sites in Australia (Miller *et al.*, 1999). Analyses of more than 1200 shell fragments show that *Genyornis* was abundant 60 000 BP, but declined and disappeared suddenly at ~50 000 BP at all three sites. By contrast, emus were present continuously throughout the period in which *Genyornis* became extinct. Because *Genyornis* fossils are often associated with some of the youngest deposits containing remains of other large vertebrates, Miller *et al.* (1999) argue that the extinction of the Australian megafauna also occurred ~50 000 BP.

The rapid extinction of *Genyornis* within a few thousand years following human settlement clearly supports the idea that humans directly or indirectly contributed to the extinction of *Genyornis* in particular, and perhaps the Australian megafauna in general. However, only a single archaeological site provides clear evidence for direct predation by humans on *Genyornis*; this is also the case for other Australian megafauna. The main alternative hypothesis proposed for these extinctions has been climate change. However, 50 000 years ago Australia was experiencing only mild climatic cooling. By contrast, the last glacial maximum occurred about 20 000 BP, during which temperatures in Australia were up to 9 °C cooler than present, and much of the Australian continent was very arid. If climate change led to the extinction of the large, ightless *Genyornis* in Australia 50 000 BP, then one must explain why the large ightless moas of nearby New Zealand were unaffected, and thrived until just 800 years ago.

Of course, Australia was not the only continent to suffer megafaunal extinctions during the Pleistocene: major mass extinctions of large mammals occurred in both North America and northern Eurasia at the end of the Pleistocene between 12 500 and 10 000 BP. The extinctions in North America have been particularly well studied (Pielou, 1991).

The Pleistocene climate was dominated by the wax and wane of glaciation, and for much of the past 100 000 years, most of northern North America was covered by two major ice sheets, the Laurentide in the east and the smaller Cordilleran in the west. South of the ice sheets, a diverse and abundant endemic fauna of large mammals ourished, including mammoths and mastodons, sabertooth cats and American lions, five species of endemic horses, and giant ground sloths and beavers. There was also a large ice-free region known as Beringia, in the area that is now Alaska and the Yukon Territory, to the north and east of the Cordilleran Ice Sheet; during this period this region was connected to eastern Siberia via the Bering Land Bridge.

The last glacial maximum occurred about 20 000 years ago. These ice sheets largely disappeared during the subsequent 10 000–12 000 years following the glacial maximum; the retreat of the ice sheets to the north was particularly rapid between 14–12 000 BP. The melting of the ice was associated with increasing average temperature and increasing atmospheric CO_2 during this time. However, there was substantial periodic variation in temperature during this overall warming, that generated alternating warming and cooling cycles with a period of about 2500 years.

The diverse megafauna south of the ice sheets suffered major extinctions between 12 000 and 10 000 BP. More than 70% of all mammal species larger than 44 kg – a total of 35 species – went extinct during this period. The dramatic changes in both climate and vegetation in North America between 20 000 and 10 000 BP may have caused extinctions of large mammals at the end of the Pleistocene. However, global climatic warming had another, more indirect effect: the melting of the ice created an ice-free corridor between the Cordilleran and Laurentide sheets about 14–12 000 BP, allowing migration of animals and plants between Beringia to the north and the areas south of the ice sheets. Among the early migrants were humans, who migrated from Beringia into North America by 11 500 BP. Did these early human settlers cause the mass extinctions of large mammals in North America 12 000–10 000 BP?

One proposal, sometimes called the overkill hypothesis, suggests that direct predation by human hunters on large herbivores caused the extinctions. Hunting of large game was an important aspect of the lifestyle and diet of these early American colonists. One intriguing pattern in the North American megafaunal extinctions is that most of the species that went extinct, including mammoths, mastodons, shrub oxen and woodland muskoxen, evolved *in situ* south of the ice sheets: these species did not share a recent history of exposure to humans (Martin and Klein, 1984). By contrast, most of the species that survived the extinctions, including elk, moose, deer, caribou and muskox, were species that had immigrated from eastern Asia during the late Pleistocene, and hence shared their evolutionary history with humans. One interpretation of this pattern is that species that had not been previously exposed to humans were more vulnerable to human hunting and were more readily driven to extinction.

A major challenge for the overkill hypothesis is the lack of late Pleistocene archaeological sites in North America that possess remains of extinct mammals. In addition, most estimates indicate that human population sizes south of the ice sheets during this time were quite small. How could a few thousand human hunters with stone tools exterminate millions of large mammals so rapidly, and without leaving extensive evidence of the slaughter? It is also possible that the effects of humans were more indirect, e.g., as a result of introducing new diseases or parasites. Although the precise

causes of these extinctions remain unclear, there is now strong evidence that the colonization of both Australia and North America by human settlers was quickly followed by widespread extinctions of large vertebrate species.

PAST AND FUTURE CLIMATE CHANGE: HOW WILL ANIMALS RESPOND?

While global climate change has waxed and waned over the past 100 000 years, the human population expansion and colonization has continued unabated. In the past several centuries, the biological consequences of human population growth have been amplified by technological changes that have greatly accelerated rates of habitat and natural resource exploitation by humans. One of the many consequences of this exploitation has been increases in atmospheric CO_2 and other greenhouse gases that are expected to increase global mean temperatures during the coming decades. Evidence that such anthropogenic climatic warming has already begun is increasing: more importantly, nearly all present data and models concur that global mean warming of $1-4\,^{\circ}C$ will occur during the 21st century. What will be the biological consequences of global warming, given the backdrop of other anthropogenic environmental changes?

Studies of past climate changes offer several important lessons for predicting animal responses. First, climatic warming and cooling have occurred repeatedly in most geographic regions during the past one million years, at a variety of time scales. Recent studies have demonstrated that regional climate may change very rapidly on the scale of decades to centuries. For example, oxygen isotope records from lake sediments and ice cores indicate a nearly $7\,^{\circ}C$ increase in mean temperature over a 50 year period about 10 700 years ago (Houghton, 1997). Thus, most current animal species have experienced rapid climatic changes during their recent evolutionary histories. Second, palaeontological studies suggest that species respond individualistically, not as communities, in response to climate change. For example, mammals in North America and beetles in Britain show complex patterns in changes in geographic distribution following the climatic warming that started about 18 000 years ago, rather than straightforward northwards shifts in all species. As a result, predicting responses to climate change must begin by considering the characteristics of species.

Another important lesson is that climate change is much more than a simple change in global mean temperature. The predictions from global circulation models (GCMs) are in general consensus about several key ways in which greenhouse gases will alter climate (Houghton, 1997). First, warming will be greater at higher northern latitudes than at lower latitudes. Second, warming will be greater in winter than in summer, especially at higher latitudes. Third, diurnal temperature variation will be reduced, primarily because of increased nighttime temperatures in most seasons and geographical regions. Fourth, the mean and variability of global precipitation will increase, with greater winter precipitation at higher latitudes.

The message from these consensus predictions is that climate change will have greater effects on low temperatures than high temperatures, and will generally reduce diurnal, seasonal and latitudinal variation in temperature. A main theme of the rest of this essay is that patterns of variation in temperature, rather than mean temperature *per se*, are key to understanding the effects of climate on animal distributions.

Climate and the Distribution and Abundance of Animal Species

The traditional approach to understanding plant and animal distribution involves mapping the geographic distribution of a species (or community) onto the distribution of environmental factors. These environmental factors may be either physical or biological. By identifying critical values of environmental factors that coincide with particular geographic boundaries, one can develop hypotheses about biological or ecological characteristics of the species in relation to environment that may limit the species at the boundary. Correlations between species distributions and climatic factors have been widely identified for animals. For example, Root (1988) noted that the northern limits of the winter distributions of many North American birds coincide with particular isotherms of average minimum winter temperatures. Root used these associations to propose that the energy balance between food intake and metabolic output at cold temperatures was a major determinant for northern winter range boundaries in such species. A variety of statistical approaches have been used to quantify the associations between species distribution and climatic variables for many animals (Dennis, 1993; Rogers and Randolph, 1993); and most quantitative predictions about the effects of climate change on species distributions are based primarily on such statistical associations.

Although this is an appropriate starting point for predictions, there are several important problems with this traditional approach (Crozier, 2001; Parmesan *et al.*, 2001). First, as noted above there are many components of climate at a location involving mean, extreme and patterns of variability in temperature, precipitation and radiation at diurnal, seasonal and annual time scales; in the absence of biological information about the species, it is not clear which (if any) of these many components should be considered. The problem is compounded by the fact that many aspects of climate are strongly correlated spatially. Because geographic distribution of a species is by definition limited spatially, spurious correlations between range boundaries and climatic factors are likely to occur in the absence of

any causal relationship. Second, the approach implicitly assumes the species boundaries are in fact limited directly by (and in equilibrium with) climatic factors, rather than by (for example) food resources, habitat requirements, or other biological factors. Put another way, it assumes that each species occupies its entire fundamental niche (with respect to climate), so that changes in climate should directly translate into changes in geographic distribution. As a result, predictions based on the traditional approach are likely to be in error if (as detailed above) different aspects of climate change in different ways, or if species and their limiting resources or habitats respond to climate in different ways.

How do we identify those aspects of climate that determine the geographic boundaries of a species? I suggest that in many animals, specific seasonal components of climate acting at key stages in the life cycle are frequently primary determinants of species boundaries (*see* Phenology, Volume 2). Let us explore this suggestion using several examples from studies with butter ies in Great Britain. Butter ies may be of particular interest for studies of climate change, because they are important ecological indicators of diverse habitats, they are quite sensitive and responsive to weather variation, and in many areas their distributions, abundance and ecology are well known and routinely monitored (Pollard and Yates, 1993). In the first half of the 20th century, the White Admiral Butter y, *Ladoga camilla*, underwent a substantial northern and western extension of its range in southern Britain (Dennis, 1993). The expansion into new areas occurred primarily during the 1930s, a decade which experienced unusually warmer weather during the months of May–July. Detailed population studies near the range edge suggested two main causes of this effect of weather on butter y abundance and range extension. First, warmer conditions in June accelerated rates of larval and pupal development, reducing predation and mortality during this period. Second, warm, sunny weather in July increased the time available for ight and oviposition, increasing the eggs laid by females. Importantly, other components of weather appeared largely uncorrelated with population abundance and the range extension of White Admirals.

Subsequent analyses suggest seasonal weather effects may be important for much of the British butter y fauna (Dennis, 1993). For example, mean daily temperatures during June and July were the best predictors of butter y abundance in more than 80% of species surveyed. Demographic studies that indicate the direct effects of temperature and solar radiation on oviposition rates are probably most important in determining population abundance. Similar effects of diurnal mean temperatures and sunshine on ight activity, oviposition and population abundance have also been documented for North American butter ies (Kingsolver, 1989).

Minimum winter temperatures may also be important in determining range boundaries in insects, especially in species that lack adaptations to extreme cold. For example, the Sachem Skipper is a generalist, southern species that lacks freeze tolerance (Crozier, 2000). It has shifted its northern range boundary northwards by ~500 km into the Pacific Northwest US during the past 50 years. During this time the mean minimum January temperature has increased by ~2.5 °C in the Pacific Northwest (Crozier, 2000). Recent laboratory and field experiments suggest that over-wintering mortality is a major factor limiting Sachem populations near the current range boundary (Crozier, 2000). Because minimum winter temperatures are expected to increase more rapidly than other aspects of climate, this may be a particularly important type of animal response to climate change in cold-sensitive species.

Another important way that climate affects distribution and abundance is through effects on the timing of key life history events in relation to environmental factors and limiting resources (Harrington *et al.*, 1999). For example, for herbivorous insects the phenological timing between insect life cycle and plant resources is critical. In temperate regions, many insects over-winter in a diapause stage, and emergence from diapause is cued by photoperiodic and temperature conditions. For herbivores, the timing between emergence from diapause and the growth and development of host-plant resources is a major determinant of population success. Demographic studies with temperate butter ies show that caterpillars are impacted by seasonal weather in three important ways: by weather effects on the early growth of host-plant resources, by the direct effects of temperature on caterpillar growth, and by the timing of host-plant senescence. Because weather variation can have differing effects on plant growth and senescence than on insect growth and development, this can lead to phenological mismatches between plant and herbivore and cause high larval mortality and even local population extinction (Harrington *et al.*, 1999; Singer and Thomas, 1996). Similarly, in winter moths, an important forest pest, temperature has differing effects on the timing of larval emergence and budburst in Sitka spruce, a major host-plant, that may lead to phenological mismatches in response to climate change. However, the effects of climate on the timing of plants and herbivores may vary markedly among systems. For example, orange-tip butter ies in Britain diapause as pupae, emerging in spring to lay their eggs on the ower heads of their main host-plant, garlic mustard. Long-term survey and climatic data from Britain show that spring temperature variation has very similar effects on the timing of butter- y emergence and host-plant owering (Harrington *et al.*, 1999).

Analogous effects of climate on the phenology of breeding and reproduction in birds in relation to their food resources are also likely. For example, studies of great

tits (*Parus major*) in Western Europe indicate that nestling survival depends strongly on the abundance of winter moth larvae during the nesting period (van Noordwijk *et al.*, 1995). However, at high temperatures larval development rates are accelerated such that caterpillar abundance and the nesting period do not coincide, leading to higher chick mortality. Similar effects are likely to be particularly important for migratory birds that breed at higher latitudes, where breeding seasons may be very short, and where climatic warming is predicted to be greatest.

Shifts in relative phenology can arise when the climate has differing effects on the life cycles of animals and their food resources. More generally, analogous effects can occur whenever animals and their natural enemies respond differently to climate. For example, field studies and models of aphid pests and their ladybird beetle predator indicate that the outcome of predator–prey interactions can depend on temperature (Kingsolver, 1989). At low temperatures, predation rate is low relative to aphid development and reproduction, leading to rapid increases, in aphid abundance. However, as temperature increases, predation rate increases more rapidly than aphid development, leading to a decline in aphid abundance. These kinds of effects may occur widely with sedentary prey and actively foraging predators, when temperature has differing effects on (prey) growth and development than on (predator) foraging and prey capture (Harrington *et al.*, 1999; Kingsolver, 1989). Similarly, laboratory microcosm experiments with *Drosophila* illustrate how temperature can also affect the outcome of interspecific competition in ways that cannot be predicted solely from the responses of individual species to temperature (Davis *et al.*, 1998).

Our discussion so far has focused on effects of seasonal temperature and precipitation on interactions of animals and their food resources. The direct effects of CO_2 on plants can also alter their interactions with insect herbivores (Lincoln *et al.*, 1993). Experimental increases in atmospheric CO_2 concentration increase photosynthetic and growth rates of plants; in many plants increased CO_2 also decreases the nitrogen concentration in leaves and other tissues by up to 20–40%. Because nitrogen is frequently the main limiting nutrient for growth in insect herbivores, this has major consequences for insect feeding and growth. Watt *et al.* (1995) recently reviewed studies with 15 insect species feeding on plants grown in elevated CO_2. These studies showed that elevated CO_2 dramatically increased consumption rates of insects, often by 40% or more. Elevated CO_2 tended to decrease developmental rates and larval and pupal weights. These effects can be largely attributed to effects of CO_2 on leaf nitrogen: those plant species for which leaf nitrogen declined most strongly in elevated CO_2 had the greatest impacts on insect growth and development. Significantly, four of the seven studies to date showed that elevated CO_2 significantly decreased insect abundance, often by 30–50%.

These studies suggest that elevated CO_2 may decrease growth and abundance of herbivores. There are several important caveats, however. First, plants may acclimatize to elevated CO_2, altering the effects of CO_2 on leaf nitrogen (Lincoln *et al.*, 1993; Watt *et al.*, 1995); the insect feeding studies to date have used unacclimatized plants. Second, most studies have used leaf-feeding insects on annual plants; impacts for trees, and for sap-sucking and gall insects remain largely unexplored. More generally, the combined effects of elevated CO_2, which tends to decrease insect development rates and abundance, and increased temperature, which tends to increase insect development rates and abundance, have not been examined.

Taken collectively, these studies illustrate how specific seasonal components of climate during key life stages can dominate the responses of many species to climate change. These effects on a given animal species may occur directly, through effects of temperature or precipitation on growth, development, survival and reproduction; or indirectly by altering the timing and interactions of populations with their food resources or natural enemies. Indeed, shifts in relative phenology of animals and their food resources or enemies may be one of the most common effects of climate change for animals in seasonal environments, especially at higher latitudes (Harrington *et al.*, 1999). As a result, predicting responses to climate change will require both seasonal and geographic details of climate, and biological information about how climate determines distribution and abundance. Approaches that combine quantitative associations between current geographic distributions and climatic variables with demographic and phenological models of key, climate-dependent life stages will be most valuable in making quantitative predictions about responses of animal populations to climate change.

Recent Distributional Shifts and Climate Change

Geographic ranges and local abundance of animals are dynamic and may change for a variety of reasons. In addition, climate may change at a variety of temporal and spatial scales. Attempts to show that the geographic distribution or abundance of a given species has changed in response to climate change must do so against the backdrop of other natural and anthropogenic changes that are occurring. Nevertheless, there are clear indications that climatic warming events during the past century have caused species shifts in some cases. Once again, the evidence is most abundant for butter ies and birds in North America and Europe, for which we have substantial long-term geographic databases at the species level.

Parmesan (1996) provided a detailed analysis of range shift in Edith's Checkerspot butter y, which occurs in discrete local populations in western North America from

Mexico to southern Canada. During the past century, the probability of successful establishment of new populations has been highest in the northern part of the range, whereas the probability of population extinction has been highest in the southern part of the range. As a consequence, both the northern and southern boundaries of this species have shifted to the north: the mean location of populations has shifted nearly 100 km northward. A similar shift to higher altitudes has also occurred. These distributional changes coincide with a mean annual temperature increase of about 0.7 °C for this region, which is consistent with the magnitude of the range shift. Increased frequency of drought events and shifts in relative phenologies of host-plant senescence and larval development may account for the extinctions of southern populations (Parmesan, 1996; Singer and Thomas, 1996).

A more recent analysis of range shifts of non-migratory European butter ies also suggests the effect of climate change (Parmesan *et al.*, 1999). Of 35 species whose northern and southern boundaries were considered, 63% shifted northwards; 29% were stable at both northern and southern boundaries, and only 6% shifted southwards. The northern shift is largely due to extensions of the northern boundary: only 20% of species both extended the northern boundary and retracted the southern boundary. As a result, 46% of the species considered expanded the size of their geographic ranges. Again mean annual temperature in the region has increased about 0.8 °C during the past century, which is consistent with the magnitude of range shifts in species that extended northwards. This study provides perhaps the best evidence to date that sets of animal species are shifting their geographic distributions in response to recent climate change; however the mechanisms underlying these shifts and the relevant components of climate involved are unknown.

It is important to recognize that these studies do not indicate that all temperate butter ies are shifting distributions in response to climate. For example, to focus specifically on the possible effects of climate change, the above analyses (Parmesan *et al.*, 1999) excluded the majority of European butter ies: in particular, species were excluded from consideration if they were clearly limited by narrow habitat requirements, habitat loss, low tolerance for anthropogenic habitat modification, or host-plant distribution. Detailed analyses of butter ies in Britain and Holland show that half of all species have declined in abundance and/or geographic range since the 1890s, which would not be predicted from increased mean annual temperature during this period (Dennis, 1993). One general pattern in the data from Britain is that the decline in rare or localized species is largely associated with habitat loss and degradation, whereas many common species have expanded in abundance and northern distribution (Crozier, 2001; Dennis, 1993). Indeed, some butter y species have expanded northward in Britain during the past 50 years despite a reduction in preferred habitats. The overall pattern is that rare species became rarer or more localized, whereas common species became more common and widespread.

Studies with temperate birds provide some limited evidence of range shifts in response to recent climate change. A recent analysis of distributional changes of breeding birds in Britain from 1970 to 1990 indicates that many species had shifted their northern range boundaries by an average of 19 km; southern range boundaries showed no significant shift (Thomas and Lennon, 1999). The short time period and modest extent of any range shifts limit what one can conclude from these analyses. Root (1988) and Root and Schneider (1995) examined winter distributions of North American birds during the past half century, and showed northern range expansions in many common species, particularly in the northeastern and western US. These expansions are consistent with relaxed energetic constraints associated with increases in mean January temperatures of 1–2 °C in the past half century. However, in many cases non-climatic factors are certainly involved, including habitat modification and supplemental winter feeding, making it difficult to ascertain the role of climate in these changes.

Nearly all of the research discussed thus far involves animals in temperate or high latitude regions, but several recent studies suggest effects of recent climate change in tropical mountain systems. In the Monteverde Cloud Forest Reserve in Costa Rica, 20 of 50 species of frogs and toads, including the locally endemic golden toad, abruptly declined and disappeared in 1987, following a year of unusually warm dry weather associated with the El Nino-southern oscillation (ENSO) (Pounds *et al.*, 1999). Climatic analyses suggest that increases in sea surface temperatures in the equatorial Pacific are associated with increased air temperature and decreased dry-season precipitation since 1976. During the past 15 years, multiple population crashes in frogs, toads and anole lizards, as well as changes in bird communities, have coincided with unusually warm and dry conditions during the dry season. Pounds *et al.* (1999) propose that atmospheric warming has raised the average altitude of the cloud bank for the cloud forest, reducing the frequency of mist in the dry season, and causing population declines in amphibians and some lizards. Whether these effects are due to direct effects of desiccation on mortality, or due to indirect effects from stress and susceptibility to disease, is unknown. These widespread declines in amphibians in a relatively undisturbed highland tropical forest in association with climate change are sobering, particularly in light of recent amphibian declines around the globe.

Changes in sea surface temperature associated with ENSO have been implicated in biological change in a number of marine animal systems, including reef corals, marine fisheries and pelagic birds (Lubchenco *et al.*, 1993). However, it remains unclear whether short time scale oscillations

such as ENSO provide a useful analog for predicting responses to sustained, directional climate change. Evidence from longer-term studies is now beginning to accumulate. Northward range shifts of pelagic fishes in the north Pacific over the past 30–50 years have been documented, coincident with increased mean sea surface temperatures (McGowan *et al.*, 1998). A study of intertidal invertebrates and zooplankton in the English Channel indicated that the relative abundances of warm- and cold-water species have changed with alternating periods of warming and cooling during the past 75 years (Southward *et al.*, 1995). Similarly, detailed analyses of intertidal invertebrates in California showed that the abundance of southern species increased and of northern species decreased between 1933 and 1996, associated with a nearly 2 °C increase in mean ocean shoreline temperatures during summer over the interval (Sagarin *et al.*, 1999).

This brief survey suggests that we now have widespread evidence of important changes in distribution and abundance of animals from a variety of systems that are coincident with recent climatic changes during the past 20–100 years. These studies clearly demonstrate that anticipated anthropogenic climate change in the coming century will likely have major effects on many animal species. It is important to note, however, that many of these studies have not clearly distinguished climatic effects from other anthropogenic and natural components of environmental change, and that many animals do not appear to respond substantially to decadal-scale climate changes.

Climate Change, Pests and Disease

One general pattern that emerges from the examples above is that many animal species have expanded their ranges towards higher latitudes during recent periods of climatic warming. Significantly, the majority of such cases involve relatively abundant, geographically widespread species that are not narrow habitat specialists. One important class of abundant, widespread organisms is agricultural, forest and medical pests. Weather and climatic variation is known to be important in the incidence of outbreaks of many pests and disease organisms; as a result there has been widespread interest in modeling the consequences of global climate change for pests and disease (Martens, 1998; Parry, 1990). Initial analyses suggest that under most climate change scenarios, many agricultural and disease organisms will spread into regions that are presently too cool to sustain them. For example, modeling of the European corn borer (*Ostrinia nubilalis*), a major pest of maize in Europe, predicts that climate warming could lead to northward range extensions of 200–1000 km during the next 50 years (Porter, 1995). Similarly, climate warming simulations predict that malaria, a major tropic disease transmitted by anopheline mosquitoes, could extend northwards into central or northern Europe and much of North America during the next 50 years (Martens, 1998). These and similar predictions have raised concerns about the expansion of agricultural pests and vector-borne diseases as a result of climate change.

An important caveat is that most of these predictions are based on simulated changes in temperature (and sometimes CO_2) and not other components of climate; in particular, seasonal precipitation patterns are often omitted from both parameterization and simulation of the models. Because temperature and environmental moisture may interact to affect animal distributions, this has major implications for predictions about climate change. For example, Rogers and Randolph (2000) recently estimated how the present-day geographic distribution of malaria is related to the mean, maximum and minimum values of temperature, precipitation and saturation pressure, then used the resulting statistical model to predict the future malaria distribution in response to climate change. In contrast to previous models based solely on temperature, Rogers and Randolph (2000) predict that climate change will produce only modest changes in the distribution and human exposure to malaria during the next 50 years. Their results also indicate that the covariation between temperature and moisture variables was critical in determining the distributional limits predicted by the model. A similar situation may occur for the distribution of tsetse ies, the vector for trypanosomiasis (sleeping sickness) (Rogers and Randolph 1993).

Recent detailed analyses of climate and tick-borne encephalitis (TBE) are enlightening (Randolph and Rogers, 2000). TBE is a major vector-borne disease in Europe and Eurasia which is transmitted by several *Ixodes* tick species. TBE is maintained by natural tick-rodent cycles; humans may become infected if bitten by an infected tick. The incidence of infection has increased markedly in many areas in Europe during the last decade. Studies in Sweden indicate a northwards expansion in the geographic range of the *Ixodes* vector during the past 20 years, that is correlated with increased minimum winter temperatures. Farther south in Sweden, increases in tick abundance and the incidence of TBE were also correlated with milder winters. These and similar observations have been used to suggest that future climate change may increase the range and prevalence of TBE (Martens, 1998).

However, recent analyses indicate that seasonal patterns of climate are key to the dynamics of TBE (Randolph *et al.*, 1999). TBE infection cycles are maintained by transmission between infected nymphal ticks and uninfected larval ticks that are feeding on the same rodent host; a sufficient transmission rate to maintain infection requires specific seasonal patterns of tick population dynamics that are controlled by climate. In particular, it requires relatively warm summer temperatures (to allow rapid tick development), rapid cooling in autumn (to induce synchronous larval diapause and spring emergence) and adequate soil moisture

levels (Randolph *et al.*, 1999). Randolph and Rogers (2000) estimated how the recent (1960–1990) geographic distribution of TBE is related to the mean, maximum and minimum values of temperature, precipitation and saturation pressure, then used the resulting statistical model to predict the future distribution of TBE in response to climate change. Their simulations predict that the distribution of TBE will contract substantially, remaining only in higher-latitude and higher-altitude regions in Europe, as a result of higher summer temperatures and decreased moisture. They suggest that this predicted contraction will occur because the seasonal pattern of tick population dynamics required for successful transmission is disrupted by climate change.

These results reinforce our suggestion that changes in the seasonal patterns of climate may be critical to understanding animal responses to climate change. This may be particularly true in host–vector–parasite and similar systems of strongly interacting species, each of which may respond to climate differently. These results also highlight the importance of incorporating both temperature and moisture variables into models and simulations to predict responses to climate change. An important limitation, however, is that predictions about future patterns of precipitation and moisture from GCM models are less reliable at the regional level than are those for temperature. Moreover, likely future patterns of spatial and temporal covariation between temperature and environmental moisture, which may be key to animal distributions in some cases, are even more uncertain.

PREDICTING ANIMAL RESPONSES TO GLOBAL CHANGES

In this essay we have explored two complementary aspects of global changes: the history of human colonization and animal extinction, seen against a backdrop of climatic change during the past 100 000 years; and the effects of recent climate change on distribution and abundance of animals, in the context of other aspects of environmental change. There are several general lessons that we can draw from these considerations, as we try to anticipate the responses of animals to global changes in the near future. Human population expansion has been causing the extinctions of animals throughout much of the history of *Homo sapiens*. What is remarkable is that quite small human populations can be sufficient to cause substantial extinction, especially for endemic, local or ecologically specialized species. The effects are of course strongest on small or isolated islands, but can be seen as well in larger areas as well. Recent human population expansions have amplified these effects a million-fold and more; and as endemic species become increasingly restricted to small, isolated islands, their extinction rates will continue to accelerate. In addition, it is usually the secondary effects of humans, most notably habitat destruction and invading species, which are the primary causes of extinction, rather than the direct exploitation of animal resources.

While climate change has been part of the evolutionary histories of most current species, anthropogenic climate change will rapidly create a new constellation of seasonal and regional climatic conditions during the 21st century. The responses of animals to these changes will be as diverse as the animals themselves, but a few general patterns are likely. First, geographically localized and ecologically specialized animal populations are less likely to be able to track rapid climate change, and are more likely to go extinct; this will be a particular problem for populations already largely restricted to (geographically fixed) nature reserves and other habitat islands. Second, many species may shift range boundaries to higher latitudes, increasing the size of their geographical ranges. This effect will be most common in species that are not too strongly tied to a single habitat or food resource. In most cases, particular seasonal components of climate will be most important in determining the extent of range shifts. Third, because there will be major changes in seasonal patterns of climate in different regions, climate change may alter the relative seasonal timing of key life history events of an animal and its resources or enemies. Such relative shifts in phenology may have positive, negative or no effects on a particular animal species depending on the system in question, but are most likely to be important in strongly interacting ecological systems including host–vector–parasite and specialized insect-plant interactions. These indirect effects of climate change are likely to be as or more important than direct effects on survival and reproduction in understanding animal responses. Fourth, many animal distributions are directly or indirectly determined by specific combinations of seasonal temperature and environmental moisture; predictions based on temperature alone should be viewed with caution. This may be particularly true when considering tropical or subtropical species potentially shifting into more temperate regions. Fifth, range expansions by many temperate agricultural pests into higher latitudes seem likely. However, because of the strong interactions among parasites, vectors and hosts, it is less clear whether range expansions in response to climate change will be typical for vector-borne human diseases.

See also: Animal Physiology and Global Environmental Change, Volume 2; Extinctions and Biodiversity in the Fossil Record, Volume 2; Extinctions (Contemporary and Future), Volume 2.

REFERENCES

Crozier, L (2000) Climate Change and its Effect on the Range Shift of *Atalopedes Campestris*, the Sachem Skipper butter y: Experimental Evidence for a Climatic Range Limiting factor, *Ecol. Soc. Am. Abst.*, 80.

Crozier, L (2001) Climate Change and its Effect on Species Range Boundaries: a Case Study of the Sachem Skipper Butter y, *Atalopedes Campestris*, in *A Change in the Weather: How Will Plants and Animals Respond to Climate Change*, ed S H Schneider, Island Press, Washington, DC.

Davis, A J, Jenkinson, L S, Lawton, J H, Shorrocks, B, and Wood, S (1998) Making Mistakes When Predicting Shifts in Species Range in Response to Global Warming, *Nature*, 391, 783–786.

Dennis, R L H (1993) *Butter ies and Climate Change*, Manchester University Press, Manchester.

Harrington, R, Woiwood, I, and Sparks, T (1999) Climate Change and Trophic Interactions, *Trends Ecol. Evol.*, 14, 146–150.

Holdaway, R N and Jacomb, J (2000) Rapid Extinction of the Moas: Model, Test and Implications, *Science*, 287, 2250–2254.

Houghton, J (1997) *Global Warming: The Complete Brie ng*, Cambridge University Press, Cambridge.

Kingsolver, J G (1989) Weather and the Population Dynamics of Insects: Integrating Physiological and Population Ecology, *Physiol. Zool.*, 62, 314–334.

Lincoln, D E, Fajer, E D, and Johnson, R H (1993) Plant-insect Herbivore Interactions in Elevated CO_2 Environments, *Trends Ecol. Evol.*, 8, 64–68.

Lubchenco, J, Navarrete, S, Tissot, B N, and Castilla, J C (1993) Possible Ecological Responses to Global Climate Change: Nearshore Benthic Biota of Northeastern Pacific Coastal Ecosystems, in *Earth System Responses to Global Change: Contrasts Between North and South America*, ed H A Mooney, Academic Press, San Diego, CA, 147–165.

Martens, W J M (1998) *Health and Climate Change: Modeling the Impacts of Global Warming and Ozone Depletion*, Earthscan, London.

Martin, P S and Klein, R G (1984) *Quaternary Extinctions: a Prehistoric Revolution*, University of Arizona Press, Tucson, AZ.

McGowan, J A, Cayan, D F, and Dorman, I M (1998) Climate-ocean Variability and Ecosystem Response in the Northeast Pacific, *Science*, 281, 210–217.

Miller, G H, Magee, J W, Johnson, B J, Fogel, M L, Spooner, N A, McCulloch, M T, and Ayliffe, L K (1999) Pleistocene Extinction of *Genyornis Newtoni*: Human Impact on Australian Megafauna, *Science*, 283, 205–208.

Olson, S L and James, H F (1982) Fossil Birds from the Hawaiian islands: Evidence for Wholesale Extinction by Man Before Western Contact, *Science*, 217, 633–635.

Parmesan, C (1996) Climate and Species' Range, *Nature*, 382, 765–766.

Parmesan, C, Gaines, S, Gonzalez, L, Kaufman, D M, Kingsolver, J G, Peterson, A T, and Sagarin, R (2001) Empirical Perspectives on Species Borders: From Traditional Biogeography to Global Change, *Ecology*, in press.

Parmesan, C, Ryrholm, N, Stefanescu, C, Hill, J K, Thomas, C D, Descimon, H, Huntley, B, Kaila, L, Kullberg, J, Tammaru, T, Tennent, W J, Thomas, J A, and Warren, M (1999) Poleward Shifts in Geographical Ranges of Butter y Species Associated with Regional Warming, *Nature*, 399, 579–583.

Parry, M L (1990) *Climate Change and World Agriculture*, Earthscan, London.

Pielou, E C (1991) *After the Ice Age: the Return of Life to Glaciated North America*, University of Chicago Press, Chicago, IL.

Pollard, E and Yates, T (1993) *Monitoring Butter ies for Ecology and Conservation*, Chapman and Hall, London.

Porter, J (1995) The Effects of Climate Change on the Agricultural Environment for Crop Insect Pests With Particular Reference to the European Corn Borer and Grain Maize, in *Insects in a Changing Environment*, eds R Harrington and N E Stork, Academic Press, London, 94–123.

Pounds, J A, Fogden, M P, and Campbell, J H (1999) Biological Response to Climate Change on a Tropical Mountain, *Nature*, 398, 611–615.

Randolph, S E, Miklisova, D, Lysy, J, Rogers, D J, and Labuda, M (1999) Incidence from Coincidence: Patterns of Tick Infestations on Rodents Facilitate Transmission of Tick-borne Encephalitis Virus, *Parasitology*, 118, 177–186.

Randolph, S E and Rogers, D J (2000) Fragile Transmission Cycles of Tick-borne Encephalitis Virus May be Disrupted by Predicted Climate Change, *Proc. R. Soc. Lond., Ser. B: Biol. Sci.*, 267, 1741–1744.

Rogers, D J and Randolph, S E (1993) Distribution of Tsetse and Ticks in Africa – Past, Present and Future, *Parasitol. Today*, 9, 266–271.

Rogers, D J and Randolph, S E (2000) The Global Spread of Malaria in a Future, Warmer World, *Science*, 289, 1763–1766.

Root, T L (1988) Environmental Factors Associated With Avian Distributional Boundaries, *J. Biogeog.*, 15, 489–505.

Root, T L and Schneider, S H (1995) Ecology and Climate: Research Strategies and Implications, *Science*, 269, 334–341.

Sagarin, R, Barry, J P, Gilman, S E, and Baxter, C H (1999) Climate-related Change in an Intertidal Community over Short and Long Time Scales, *Ecol. Monogr.*, 69, 465–490.

Singer, M C and Thomas, C D (1996) Evolutionary Responses of a Butter y Metapopulation to Human and Climate-caused Environmental Variation, *Am. Nat.*, 148, S9–S39.

Southward, A J, Hawkins, S J, and Burrows, M T (1995) Seventy Years' Observations of Changes in Distribution and Abundance of Zooplankton and Intertidal Organisms in the Western English Channel in Relation to Rising Sea Temperature, *J. Therm. Biol.*, 20, 127–155.

Steadman, D W (1995) Prehistoric Extinctions of Pacific Island Birds: Biodiversity Meets Zooarchaeology, *Science*, 267, 1123–1130.

Steadman, D W, White, J P, and Allen, J (1999) Prehistoric Birds from New Ireland, Papua, New Guinea: Extinctions on a Large Melanesian Island, *Proc. Natl. Acad. Sci.*, 96, 2563–2568.

Thomas, C D and Lennon, J J (1999) Birds Extend their Ranges Northwards, *Nature*, 399, 213.

van Noordwijk, A J, McCleery, R H, and Perrins, C M (1995) Selection for the Timing of Great Tit Breeding in Relation to Caterpillar Growth and Temperature, *J. Anim. Ecol.*, 64, 451–458.

Watt, A D, Whittaker, J B, Docherty, M, Brooks, G, Lindsay, E, and Salt, D T (1995) The Impact of Elevated atmospheric CO_2 on Insect Herbivores, in *Insects in a Changing Environment*, eds R Harrington and N E Stork, Academic Press, London, 198–217.

Wilson, E O (1992) *The Diversity of Life*, Norton, New York.

Natural Systems: Impacts of Climate Change

S MARK HOWDEN AND GUY B BARNETT

CSIRO Sustainable Ecosystems, Canberra, Australia

*Climate is a key determinant of the location, structure and function of natural ecosystems. We see this in everyday observations of the landscapes in which we live; for example, progressive changes in vegetation up a mountainside or the relationship of forest types to rainfall. There have been many attempts to translate these types of observations into schemes that relate climate to vegetation at the global scale. One example is the Holdridge diagram, which relates the global distribution of natural vegetation in terms of temperature, precipitation and evaporation (*see *Holdridge Life Zone Classification, Volume 2). Similar attempts have been made with fauna, although mobility can make such analyses more complex. There are also many examples of how climatic change on geological timescales has impacted on ora and fauna populations and on survival and distribution of whole ecosystems. These observations lead us to predict that the climate changes forecast over the coming centuries will have profound impacts on natural ecosystems and their component species. This essay will review brie y these predicted impacts for natural systems, which here include managed, semi-natural systems such as grazed rangelands.*

UNCERTAINTIES

Climate change impacts on natural ecosystems will occur in conjunction with a host of other human-induced (anthropogenic) pressures. Land clearing for agriculture and urban development, intensification of use of other ecosystems, introduction of exotic plants, animals and diseases, nitrogen deposition, acid rain and atmospheric carbon dioxide (CO_2) increase are just a few of these stresses. It is the interaction between climate change and these other factors that will determine the nature and distribution of ecosystems in the future. These interactions are generally poorly understood but are thought to be important in that the effects of one source of change can be reinforced by the other stresses. For example, computer simulations of broadscale land clearing of tropical forests suggest that regional droughts will increase, leading to regional changes in ecosystem composition and function, threatening species and increasing susceptibility to climate change and other disturbances.

Some of these impacts have positive aspects. For example, increasing levels of atmospheric CO_2 can increase plant growth, particularly under dry conditions and ameliorate the deleterious effects of tropospheric ozone. Plant growth (net primary productivity) is positively related to the number of plant species per unit area and so at first glance this would appear positive for biodiversity. However, there is accumulating evidence that species differ in their response to increasing CO_2 levels, suggesting that changed competitive relationships may shift relative distribution and abundance of species. This could lead to changes in community composition that are difficult to predict, given the complexity of the interactions. Similar arguments for the likelihood of future alteration in community composition apply to many of the other factors listed above, including climate change. In addition to changes in community composition, there may be impacts on ecosystem function where the changes affect rooting depth, photosynthetic pathway, nutrient cycling or the structure of the community. Changes in ecosystem composition and function could impact on the free services we get from ecosystems such as fresh water, pollination and maintenance of soil structure and fertility. In agricultural systems, the impacts of increasing CO_2 are sufficiently large to justify their inclusion in any global change impacts assessments. However, for natural systems, the effects are less well known and often lead to uncertain assessments of the impact on individual species and their interactions, although the general impacts on growth can be estimated.

There are several other uncertainties in assessing potential impacts of climate change on natural ecosystems. Chief amongst these are:

- future emission trends (including global economic development pathways, population growth and resource-use equity, degree of technological change and other emission-reduction options and policies);

- the degree to which these emissions will affect global climate (including the cooling effects of sulfate aerosols, impacts on oceanic circulations, cloud and particulate dynamics, feedbacks between land cover changes from carbon sink activities and climate);
- the translation of global climate changes to regional climate effects (including impacts of topography, land cover, land-ocean interactions).

Additional important uncertainties relate to possible asymmetric warming (where night-time temperatures increase more than those in the daytime), changes in the variability of climate (e.g., increased frequency of El Nino events) and changes in climate extremes (oods, droughts, storms, heatwaves, cyclones). Changes in extremes can be particularly important as they may produce conditions that exceed critical environmental thresholds and trigger change in ecosystems by increasing mortality or enhancing reproductive success of species.

NATURAL SYSTEMS

Climate change impacts on natural systems are discussed below using a simple bio-regional ecosystem classification. The ecosystems used here are:

- grasslands and savannas;
- deserts;
- tropical forest;
- temperate forest;
- boreal forest;
- polar regions;
- alpine environs;
- oceans;
- coastal zones and small islands;
- freshwater and wetland systems.

The terrestrial bioregions represent latitudinal, rainfall and altitudinal gradients, as these factors strongly in uence biodiversity (as measured by species richness), productivity and ecosystem function, particularly water availability and water balance. However, natural systems within each of these regions interact strongly with other factors such as soils and so there can be large variation within regions.

Grasslands and Savannas

Grasslands and savannas cover approximately 15% of the world's land surface (*see* Temperate Grasslands, Volume 2; Tropical Savannas, Volume 2). This area is declining due to landuse change to cropping activities and in some locations substantial urbanization. Up to 71% of the grassland and savanna ecosystems are degraded to some extent as a result of overgrazing, salinization, alkalinization, acidification and other processes. The pervasiveness of human in uence means that in many cases grasslands and savannas can best be thought of as semi-natural ecosystems. Extensive clearing of the woody vegetation across landscapes and installation of artificial watering points has already had very significant negative impacts on the biodiversity of grasslands and savannas in some regions. High levels of sustained total grazing pressure by domestic livestock in conjunction with pest animals have further impacted on the native species through competition for food and occupation of habitat. The introduction of exotic predators and weeds in many locations has further reduced biodiversity. In many cases, the above changes have significantly altered the functioning of grasslands and savannas with decreased infiltration, greater run-off and erosion, more rapid cycling of nutrients and lowered carbon levels. In more humid climates, intensive pasture development has transformed these systems so that they have a marginal biodiversity value. The alteration in grassland function may produce secondary effects in adjacent regions through increased nutrient ows into groundwater and waterways and increased drainage components that can raise dryland salinity risk. In some locations, these systems are being impacted by increasing rates of nitrogen deposition (which can improve production in nitrogen-poor sites but also increase soil acidity) and increased exposure to ultraviolet-B (UV-B) radiation and tropospheric ozone. The impacts of climate change have to be viewed in the context of these existing pressures, which are expected to continue and grow rather than decline.

Many grasslands and savannas have already been exposed to climate changes over recent decades. Increasing temperatures, particularly at night, increases in rainfall intensity and changes in rainfall amounts (often upwards but sometimes downwards) have been observed. The impacts of such changes are hard to separate from the background noise caused by the high climate variability often found in these systems and the relatively indirect measures of production that arise from them. Increases in climatic variability are also suggested as a potential impact of climate change, making it even more difficult to separate a meaningful signal from the historical record. These trends in climate are expected to continue, and the location of many grasslands away from continental margins means that expected warming trends may be higher (1.4–6 °C by 2080 with significant variation by region and by scenario) than in areas adjacent to the coast.

Increased temperature will tend to reduce productivity in many locations. This occurs due to increased evaporation rates and lower water use efficiency, which can arise from lower relative humidity. For some species, temperature optima for photosynthesis may be exceeded. In contrast, in some locations where low temperatures limit growth in autumn or spring, but where soil moisture is available, there may be an increase in the length of the growing season. This can have significant positive impacts on

grazing animals. However, in already warm environments, increases in temperature will increase heat stress frequency in animals, reducing physiological activity, foraging levels, growth and reproduction rates but increasing water demand, thus limiting the extent of their range from watering points. If the water supplies themselves become more scarce or intermittent, there could be significant impacts on species distributions. Increases in temperature may also result in transient release of nutrients from soil pools as they adjust to higher temperatures and this may be important in nutrient-limited environments.

The water-limited nature of many grasslands and savannas means that they are sensitive to changes in effective rainfall. However, there is substantial regional variation in suggested rainfall trends. Many grasslands are assessed as likely to undergo small but significant (i.e., 0–10%) decreases in rainfall by year 2080. However, there are some regions such as in Central America and south–west Australia that, under high emissions scenarios, may have larger decreases (−25%) whilst others such as western China that could get much wetter (+25%). Whilst the implications of such changes on natural grasslands will be significant, improved regional scenarios are required to decrease existing uncertainties.

Atmospheric CO_2 levels are expected to rise rapidly over the next 100 years. This change will have profound impacts on these systems, particularly where they are water limited and warm. Grasses increase their water and nitrogen use efficiencies under higher CO_2 levels while cool season (C_3) grasses also increase their photosynthetic rate. Grasslands and savannas have been exposed to increases in atmospheric CO_2 over the past 100 years and this may have already enhanced production by up to 8%. Growth responses to doubling of current CO_2 levels may range from negligible in wet, cool grasslands to 45% in dry warm savannas. The level of this growth enhancement is also dependent on the nutrient status of the soil, with lower nutrients limiting the response. Increased CO_2 is likely to reduce the nitrogen content but increase the carbohydrate content of leaves. In the majority of savanna environments that are nutrient-poor, this will reduce growth and reproduction of grazing animals, affecting the whole food chain. Increases in temperature will also reduce the digestibility of savanna grasses. In both of these cases methane emissions from the animals will tend to increase, providing a positive feedback to climate change. In other more nutrient-rich environments, the increase in soluble carbohydrates may increase herbivore productivity.

The combined effects of increases in temperature and CO_2 may change both the grass composition and the structure of grasslands and savannas. Grasses are split into two groups based on their photosynthetic pathway; C_3 species, which tend to grow in cooler, moister climates, and C_4 species, which are found in warmer, drier climates (*see* C_3 and C_4 Photosynthesis, Volume 2). Both groups increase their water and nitrogen use efficiency under higher CO_2 levels, but only C_3 species increase their photosynthetic rates. The distribution of these two groups under current CO_2 levels is correlated with climatic parameters such as growing season temperature, which affects the relative productivity of the group. Thus, warming by itself would result in poleward migration of the component species, reducing the area occupied by C_3 species. However, increasing CO_2 levels will tend to counteract this movement because of the additional photosynthetic response of C_3 species to CO_2. The net effect is uncertain and appears to vary depending on the seasonality of rainfall against temperature and the nitrogen response of the species. In contrast, legume species that fix atmospheric nitrogen are considered likely to benefit from increasing CO_2 levels. If there are significantly different species responses, then there is likely to be a breakdown of the existing species assemblages and generation of new ones. This is likely to have biodiversity implications for associated vertebrate and invertebrate species.

Global change may also have implications for the structure of grasslands and savannas through changes in the dynamics of woody versus grass species. Vegetation structure is important as it affects biodiversity, animal productivity (as trees strongly depress grass productivity in savannas), carbon storage, nutrient cycling, hydrology and the interaction of the biosphere with the atmosphere. Woody species have the C_3 photosynthetic pathway. It has been suggested that this is a way to enhance their competitiveness over the C_4 grasses that currently dominate the understory in savannas. This would occur as increased CO_2 levels may give woody plants greater growth response, increased seedling survival due to improved soil water availability from lower transpiration rates and deeper rooting if climate becomes drier. However, where burning is practiced, CO_2-induced increases in grass fuel load may kill establishing cohorts of woody plants leading to long-term declines in the woody ecosystem component. Thus, the structure of these systems under global change is probably dependent on the disturbance or management regimes. Establishment of woody species is often critically dependent on a favorable sequence of seasons. Climate changes could either increase or decrease the frequency of such conditions, adding to the uncertainty of the outcomes.

Regional changes in rainfall are more uncertain than changes in CO_2 or temperature. In dry grasslands, plant productivity is often linearly related to rainfall; thus a 10% decline in rainfall translates to about a 10% decline in production. However, increases in CO_2 may be able to offset such small reductions in rainfall. Thus, in locations where climate changes are relatively favorable (i.e., a small decrease to an increase in rainfall), increased plant production may increase grass cover, thus providing lower

risks of soil erosion even whilst slightly increasing animal productivity and numbers. In wetter environments, other constraints limit growth and so there may not be the same sensitivity to rainfall. Some grasslands in Australia, Africa and the Americas are strongly in uenced by the El Nino/Southern Oscillation (ENSO) system which results in high rainfall variability. There have been projections that the frequency of El Nino events may increase with global warming. This could result in significant increases in the variability of climate, producing large impacts on the biodiversity, productivity and functioning of affected ecosystems. For example, migratory mammals and insects have movement patterns that are strongly linked to variability in seasonal climates which if altered may disrupt these movements. Such pressures for the mammals are likely to be compounded by increasing fragmentation of landscape, barriers such as fences and control of watering points.

In many of the world's grasslands and savannas, if there is an increase in biological productivity arising from global change, then economic and cultural pressures as well as natural increase in wild populations, will tend to increase herbivore populations thus maintaining pasture utilization and providing little benefit to the natural systems and their function. There may also be increased pressure to crop the land, and in some regions such as North America, large areas of grasslands may become suitable for cropping. In contrast, in areas where production may decline due to climate change, those cultural, economic and population pressures will tend to maintain previous grazing populations and so pasture utilization will tend to increase as has happened in the grasslands of temperate Asia. This will lead to increased pasture and soil degradation, increased invasion of unwanted grass and woody species and long-term decline in the productive capacity of these systems.

Desert

Deserts are characterized by low rainfall (<250 mm annual average), which is highly variable. They cover about 30% of the earth's land surface (*see* Deserts, Volume 1; Deserti cation, Volume 3). Most deserts are located in the mid latitudes between $20-40^{\circ}$ although at other latitudes deserts exist in rain shadows adjacent to mountains whilst others occur simply due to great distances from oceans. High levels of solar radiation and low relative humidity result in large daily temperature uctuations and high evaporation. Rainfall, when it occurs, is often intense and localized and is channeled into small parts of the landscape. This results in substantial variation in plant production over time and space because productivity is often linearly related to water use. This variation in productivity strongly in uences the distribution of fauna. Thus, water availability is often the dominant factor controlling ecosystem processes. Hence climate changes that in uence the availability or effectiveness of rainfall are likely to be the most critical in in uencing desert ecosystems. Changes in the effectiveness of rainfall may arise through changes in the amount, season and intensity of rainfall, temperature changes and the level of atmospheric CO_2.

Recent climate change scenarios suggest significant decreases in rainfall in many of the world's deserts over the next 100 years although there remains considerable uncertainty over such regional scenarios. One of the reasons for the projected decrease appears to be intensification and poleward movement of the mid-latitude high-pressure belts due to an intensification of the Hadley circulation whilst atmospheric concentrations of greenhouse gases are increasing. Reductions in rainfall will tend to result in approximately proportional reductions in plant productivity and this will tend to reduce the abundance of other organisms. In a few areas, such as the deserts of the western USA, there may be an increase in winter rainfall and a decrease in summer rainfall. Rainfall changes like this may not significantly alter total system productivity but may result in changes in species composition towards those plants that favor growth in cooler seasons. These plants often have the C_3 photosynthetic pathway rather than the C_4 pathway. Consequently changes in rainfall seasonality may result in a shift in the C_3/C_4 balance with consequent impacts on ecosystem functions due to changes in community structure, nutrient cycling and water use.

Increases in the intensity of rainfall may result in higher proportions of run-off, thus enhancing growth in those limited areas of the landscape that accumulate water. This will tend to focus biological activities on these small areas leading to possible increased threats to biodiversity from predators and from competition from domesticated livestock, as well as providing environments where pest animals and plants can establish themselves. In other environments, increased intensity of rainfall may enhance the establishment and survival of ephemeral species that grow after rainfall events and survive as seed until the next favorable growth opportunity. These changes in community composition have been observed over past decades in response to the increased intensity of rainfall in a range of environments. Increases in ephemeral species may affect the perennial species through competition for resources or by increasing fire frequency or grazing intensity. Many insects, such as locusts, have similar life cycles to ephemeral plants and they may be in uenced in a similar way by changes in the variability of climate.

Recent climate change scenarios suggest temperature increases in the range of $1-5\,^{\circ}$C for desert regions for the year 2080. Evaporation rates increase markedly with temperature and so a $2\,^{\circ}$C rise increases potential evapotranspiration by $0.2-2\,\mathrm{mm\,day^{-1}}$. By itself, this will result

in less efficient use of water by plants and more rapid depletion of water following rain, although in environments where cold weather limits growth in spring or autumn, warmer temperatures may extend the growing season provided water is available. For plants, the negative effects of such increases in temperatures on water use may be partly offset by the positive impacts of increasing levels of atmospheric CO_2, which increases plant water use efficiency and also increases photosynthetic rates in C_3 plants. By itself, doubling of CO_2 over current levels may increase plant growth by up to 50% in desert environments. However, these increases are likely to moderate the effect of temperature increases in combination with rainfall reductions so the net effect of global change on productivity may, on balance, be slightly negative, particularly in nutrient-poor environments where the CO_2 response will be limited. Furthermore, whilst the CO_2 response will tend to increase less rapidly with increased CO_2 levels, that due to climate change will not, thus suggesting possible early increases in productivity but increasingly negative responses over time. In deserts in which organisms rely on fog and dew deposition, temperature increases are likely to have particularly negative effects as fog frequency is likely to decline due to less frequent occurrences of the low temperatures which produce condensation. There are already strong reductions in fog frequencies in some environments, which are having demonstrated impacts on biodiversity.

Climate changes will not occur in isolation from existing pressures on biodiversity and natural systems in deserts. Unlike many other regions, pressures for landuse change are not strong; however, the existing land management has resulted in the majority of the world's deserts being classified as degraded to some extent. Important processes include overgrazing by domestic livestock, particularly near natural or manmade watering points, introduced feral predators and herbivores and weeds. The history of use of desert regions suggests that if there are any increases in productivity from climate change, corresponding increases in herbivore populations will occur, negating the benefits to the natural systems. In contrast, recent research suggests that declines in desert productivity will tend to result in transient increases in forage utilization until declining productivity or degradation causes adjustment in herbivore numbers or behavior. Thus, climate change may interact with other pressures resulting in even more rapid degradation and increased impacts on natural systems.

Many species of water birds, fish and invertebrates are highly attuned to ephemeral water storages in deserts. Moderate increases in the variability of the climate of desert environments are by themselves unlikely to affect these species negatively. Increases in the frequency of filling of such waterbodies could arise in some regions from a possible increase in ENSO events or from more poleward movements of tropical cyclones. However, if these changes in frequency are sufficiently great, then existing demands to use this water for irrigated agriculture may intensify, which may have serious, negative impacts on the natural ecosystems if development proceeds (*see* C_3 and C_4 Photosynthesis, Volume 2).

Tropical Forests

Tropical forests cover approximately 12% of the earth's land surface and 40% of the world's forested area, but contain almost half of the world's plant and animal species (*see* Tropical Forests, Volume 2). They are made up of a variety of different forest types, but the main ones are rainforest, drought–deciduous forest, and dry forest. They are found between 25 °N and 25 °S, distributed over five continents. Most of these tropical forests are located in the developing countries of the low latitudes, such as central Africa and much of southern America and southern Asia. These locations typically have high population densities and are undergoing rapid economic development, and thus predicting the impact of climate change requires consideration in the context of both current and future human activity. For instance, many tropical forests are currently being rapidly cleared, and if this continues, human-induced landuse change will have a much more significant impact on tropical forests than that of climate change. However, these impacts may interact to increase the hazard of losses in biodiversity.

Climate change projections for the year 2080 indicate that seasonal temperatures in tropical forests could increase by 1–4 °C, depending on scenario and location. This level of warming is unlikely to increase photosynthesis for any except tropical montane forests where low temperatures can limit growth. However, these higher temperatures will increase plant respiration, tending to lead to lower growth and increase rates of breakdown of litter, reducing ecosystem carbon levels. For cloud forests, which rely on droplet deposition for much of the water supply, increases in temperatures will tend to reduce the frequency of deposition events leading to substantial changes in the whole ecosystem. This already may be happening in forests such as the Monteverde forest in Costa Rica with impacts on birds, reptiles and amphibians.

Increases in temperature will increase evapotranspiration, leading to reduced soil water availability. The degree of this effect will be modified by any changes in rainfall patterns. Projected change in rainfall is highly regionally variable with expected increases (10–15%) in much of southeast Asia (but changes in seasonality), equatorial Africa and the Parana–Rio Grande region of South America but decreases for the Amazon and Central America. If these changes are coupled with possible changes in the frequency of ENSO events, the result could be prolonged dry seasons and drought in some regions, with more rains and ooding in others.

The impact of this potential change in soil water availability is likely to be most significant in those regions that contain boundaries of different biomes. In these areas, the impact could involve a shift in the current distribution and composition of tropical forest types. For instance, some research suggests that if precipitation in the seasonally dry deciduous forests decreases, then these forests will become drier, promoting the accumulation of organic matter and leading to a possible increase in the frequency of intense fires. Under this scenario, the seasonally dry deciduous forests would tend to be replaced by fire-tolerant savanna vegetation and those plant species that are sensitive to fire would be lost. The converse scenario might involve the expansion of those species that prefer moist and wet climates into regions from which they have been excluded in the past through a lack of soil water availability or low temperature. In either scenario, there would be subsequent impacts on vertebrates and invertebrates, many of which have life-cycles tightly linked to specific tropical forest plant species.

Tropical forests typically grow on infertile soils and, therefore, require efficient nutrient capture and cycling to ensure an adequate supply of plant-available nutrients. Phosphorus is the most limiting of these nutrients and it has been suggested that elevated CO_2 levels (which can enhance photosynthesis and water use efficiency), could increase both root growth and mycorrhizal development, thereby enhancing the ability of tropical forest vegetation to access soil phosphorus and to fix atmospheric nitrogen. In undisturbed tropical forests, it is likely that these effects may lead to a potential gain in productivity and net carbon storage, particularly where nutrient limitations are minimal. However, the degree to which these effects will occur in the wet tropical regions (particularly when subject to disturbance) is open to considerable debate. Nevertheless, in the drier tropical regions, elevated CO_2 levels are likely to increase water-use efficiency by plants, and hence potential vegetation productivity.

Given the tightly closed cycling of nutrients in tropical forests, it is clear that any disturbance could have a significant impact on nutrient status. Examples of how climate change may in uence disturbance include increased rainfall intensity leading to soil erosion, increased decomposition rates, altered leaching regimes, increased storms which cause wind-throw and increased drought frequency or intensity which enhances mortality rates from water stress or from drought-related fires. This, and the importance of the regeneration phase in determining forest composition, suggests that changes in disturbance regimes could in uence forest productivity and distribution.

Whether tropical forests as a whole will increase or decrease in area under climate change is unclear, although it is possible that the total area may not change, but rather the relative proportions and composition of each forest type.

Temperate Forests

Temperate forests are extremely diverse, sometimes being a transition zone between boreal and tropical forest and in other places being adjacent to semi-arid ecosystems (*see* Temperate Coniferous Forests, Volume 2; Temperate Deciduous Forests, Volume 2). They occur across wide spatial ($25-50\,^{\circ}$N and S) and climate ranges. The most extensive forest types include *Fagaceae*-dominated northern deciduous, temperate coniferous, and *Eucalyptus* and *Nothofagus*-dominated evergreen forests in the Southern Hemisphere. Temperate forests cover approximately 7% of the Earth's land surface and are primarily located in developed countries. Many have a significant history of human impact. They are now commonly fragmented or discontinuous vegetation blocks contained within a matrix of agricultural and urban landuse. However, many of these forests are presently undergoing significant regrowth following the cessation of clearing and wood removal, which is also being supplemented by reforestation. As a result, the area and biomass of temperate forest is currently increasing, and for this reason, temperate forests may act as a carbon sink over the forthcoming decades.

Climate change scenarios for the year 2080 are for increases in summer and winter temperatures of $1-5.5\,^{\circ}$C and generally small increases ($0-+10\%$) in rainfall, although some regions such as southern Canada and China are projected to have bigger increases, and some such as Spain, the southern United States and Australia may have decreases. The changes are highly variable between regions and scenarios. Due to increased evapotranspiration, many soils are expected to be drier in late summer even though annual rainfall may increase slightly. The frequency of summer droughts could increase, particularly for those regions with a continental rather than oceanic climate and where rainfall decreases. The result could include reductions in forest density, leaf area, growth and standing biomass, and ultimately the loss of some marginal temperate forests following repeated drought. For other forests, the end result could be earlier and warmer growing seasons (sometimes longer and sometimes shorter), with fewer cold temperature extremes and frosts in winter. But perhaps of more concern than the magnitude and direction of these changes is the speed at which this change is likely to occur. Many temperate forest species grow in a limited climatic range. For example, in Australia 25% of *Eucalyptus* species have mean annual temperature ranges spanning less than $1\,^{\circ}$C whilst 53% have a range less than $3\,^{\circ}$C and 23% have a range spanning less than 20% variation in rainfall. This limited climatic range suggests that many forest species may be vulnerable to climate change. The high rates of migration needed to stay in current climate envelopes seems unlikely, given habitat fragmentation and degradation due to clearing, introduced pests and diseases and changed fire regimes. The

net result is that significant changes in species composition may occur.

Other associated impacts of climate change might include an increase in the incidence of fire in response to drought, outbreaks of pests and pathogens that are advantaged by warmer climates, and limitations on reproduction in certain temperate forest plants. For the latter, this could involve inadequate chilling requirements that are needed to break dormancy in seeds, or a reduction in frost hardiness that leaves plants more susceptible to damage during extreme events.

The combination of climate change and elevated CO_2 are predicted to affect the net primary productivity of temperate forests, but the overall direction and magnitude of this affect are not always clear. For example, enhanced growth of young trees exposed to elevated CO_2 has been observed to occur in experiments, but the effects are less marked in longer-term experiments largely, it appears, due to other limitations to growth (e.g., low nutrient availability). Similarly, changes in net primary productivity remain sensitive to rainfall but the interaction with increased temperature and CO_2 levels are not fully understood.

Temperate forests are currently considered to be a carbon sink, but some temperate forests could potentially become a source of carbon if climate change results in forest decline or human activity causes forest degradation such as soil acidification. As most temperate forests are located in developed countries, there is at least some capacity and knowledge available to minimize climate change impacts. Such strategies could include the adoption of integrated fire, pest and disease management technologies, as well as the supplementation of existing forest through reforestation activities.

Boreal Forests

Boreal forests are dominated by evergreen, coniferous species (*see* Boreal Forest, Volume 2). They are found in the northern parts of North America, Europe and Russia and cover approximately 9% of the earth's land surface. They are constrained in the north by cold temperatures that limit tree height and reproduction, and in the south, by competition from broad-leafed and deciduous forests and savanna-woodland species, as well as water availability. Boreal forests are already experiencing a series of pressures. These include acid rain and soil acidification, increased tropospheric ozone exposure, increased UV-B exposure, increases in logging demands (e.g., 38–96% increase in logging from Asian forests projected over the next decades) and landuse change pressures from agricultural expansion. There have also been pressures on ora and fauna from overuse, increased wildfire frequency, habitat change and increased biotic invasions. Boreal forest growth rates have increased over recent decades due to warmer seasons, increased nitrogen deposition, increased CO_2 levels and changed management but there have also been increases in insect and disease attacks and damaging fires.

Climate projections are for significant warming over the boreal forest regions during the next 100 years, resulting in northwards expansion of boreal forest in excess of two–thirds of the area currently occupied by tundra ecosystems. These changes are due to increased reproduction and productivity in the warmer climate and longer growing seasons. This increased growth may be assisted by increased levels of atmospheric CO_2, in combination with enhanced nutrient availability from atmospheric deposition in some areas as well as transient nutrient release from the soil. However, on the southern margins there may be encroachment of broad-leaf, deciduous species and replacement by grasslands and shrublands due to warmer temperatures and reductions in summer rainfall. For North America this may result in about a 15% increase in total area of boreal forest but in Russia a 50% reduction in area has been suggested. Even in North America where net changes are positive, some 14–19% of boreal forests could undergo drought-induced biomass decrease. Palaeoecological evidence suggests that these forests may not be able to shift across the landscape as fast as climate may change. This may increase the likelihood of disruption, dieback and species extinction due to disequilibrium with climatic factors. If dieback in the south occurs faster than expansion in the north, then there may be a temporary decline in boreal forest area or a shift to drier types of boreal forest under warmer scenarios. Consequently, responses may vary with time as well as place.

Fire frequency and damage in boreal forests has been increasing over the past decades. This appears to be partly due to climatic effects but also from management that has tended to allow buildup of biomass in some regions. This buildup results in increased transpiration, increasing drought stress frequency and susceptibility to fire. Climate change is likely to increase the current fire risk in some regions (e.g., Russia and Canada) by prolonging the length of the season when fires can occur and increasing the frequency of climate sequences that generate larger, more intense and potentially destructive fires. This could result in replacement of existing stands by new species and in long-term negative impacts on soil properties, such as infiltration rates. There could also be significant, negative impacts on wildlife and biodiversity. Changes in fire regime may be moderated if rainfall increases as is expected under some scenarios in some regions.

Cold temperatures protect boreal forest trees from many insects and fungi that are prevalent farther south. Climate warming over past decades has increased the frequency and severity of insect attacks in boreal forest (e.g., budworms and bark beetles) by enabling more generations per year, increasing the physiological activity levels

and increasing over-wintering survival. These trends are expected to continue with climate change although the complex nature of interactions prevents firm prediction of the outcomes. These effects are more strongly felt on drought-stressed trees and increased drought frequency associated with warm temperatures; nitrogen overfertilization may also increase susceptibility to pest and disease attack.

Polar

The polar regions dealt with here are those land areas of the Arctic and Antarctic bounded by the Arctic Circle in the Northern Hemisphere and 58 °S latitude in the Southern Hemisphere, respectively. They are characterized by low temperatures, strongly seasonal radiation and the current dominance by ice and snow (*see* Tundra, Volume 2). There have been strong warming trends in both the Arctic and Antarctic during the past 50 years. These increases have been up to 5 °C, although some regions show neutral to cooling trends. In the Arctic, this has already enhanced the temperature environment for agriculture and forests with earlier springs and longer growing seasons, changing the reproductive cycles of some animals and encouraging the poleward movement of forests. In the Antarctic, this appears to have markedly increased growth and population sizes of both of the two vascular plant species that exist there. There have also been increases in rainfall in both these regions. In the Arctic, there is long-term decline in sea ice extent and the sea ice has thinned. In the Antarctic, there was a significant retreat of sea ice between the mid-1950s and early 1970s but subsequent changes have been regionally variable. However, there has been a marked collapse and retreat of ice shelves on the Antarctic Peninsula as a result of regional warming. Surface waters in both regions have freshened and warmed. Ground temperatures have increased, permafrost area has declined and there has been an increase in the number of melt days per summer. Glaciers, ice caps and snow-covered areas have generally diminished although in some areas there is expansion in the upper reaches due to increased precipitation. The above trends are expected to continue with global warming, with temperatures in the Polar Regions increasing more than other regions particularly in winter. Some projections indicate winter temperature increases in the Arctic of greater than 7 °C by 2080 under high greenhouse gas emissions scenarios. The Arctic region is likely to be impacted by climate change more rapidly and significantly than any other region on Earth, while the Antarctic may be less affected at least over the next 100 years.

Biological production in Polar Regions is temperature dependent, and therefore global warming is likely to generally increase productivity and abundance throughout the food chain, and earlier and longer growing seasons are likely to benefit many species. However, there are likely to be changes in species compositions due to the different competitive abilities. For example, the lichens and mosses of the tundra regions may be replaced by more productive, low-growing vascular plants. Exotic and other species such as insects, which cannot tolerate the current cold conditions, may start to colonize the Polar Regions whilst other species that already exist there may alter their relative abundances. These changes are expected to result in a general tendency for poleward migrations of species in the Arctic. For example, there may be gradual northward expansion of boreal forest into the tundra regions, affecting up to two–thirds of the tundra. Although more precipitation is generally expected, increases in evapotranspiration may lead to droughts late in summer that could place additional pressures on existing species that are not drought-adapted. In the Antarctic, warming may further diminish ice shelves, exposing more ground, resulting in significant changes in the terrestrial ecosystems, whilst in the Arctic, sea ice is expected to decline markedly. These changes may impact on terrestrial food chains through changes to predator and prey species, particularly where there is a substantial dependence on sea-ice for shelter, reproduction or feeding. Salinity and temperature changes will also have a dramatic effect on species composition, both in and out of the circumpolar oceans. Increased CO_2 levels will have only a limited affect in either of these regions due to the low temperature and nutrient availability. However, the increased frequency and intensity of stratospheric ozone depletion in both the Arctic and Antarctic is likely to increase UV-B radiation exposure and this may have negative impacts on many plants and animals. Limited evidence suggests that this may already be occurring in both plant and animal species.

The hydrology of polar regions is highly sensitive to temperature changes with increases in the proportion of rainfall compared with snowfall, increased and earlier melting events and increased infiltration in previously permafrost areas resulting in lower seasonal variations in runoff. Increase in winter flow under the ice may be beneficial to invertebrate and fish populations whilst reductions in ice-jam induced flooding may have serious negative implications on Arctic riparian ecosystems. Changes in total runoff are likely to be regionally variable, depending on the balance of increased precipitation and increased evapotranspiration. The thawing of large areas of permafrost soil, particularly in ice-rich areas, will change regional drainage, subsidence and hence vegetation (e.g., changes from forests to bogs) and landuse by increases in bog areas, and increase landscape erosion by slumping and landslides. If there are widespread increases in stream flow as a result of increases in rainfall and the above hydrological changes, there may be a freshening of the waters of the Arctic Ocean and the North Atlantic, reducing formation of deep ocean water and thus weakening the thermohaline circulation. If this circulation

undergoes rapid reorganization, then abrupt climate change may occur. Many of the above changes will have implications for atmospheric concentrations of greenhouse gases. For example, reduction of permafrost may release methane and weakening of ocean circulations will reduce oceanic carbon sinks.

Alpine

Alpine areas can be found scattered throughout the mountainous regions of the world. They are defined by the climatic limit of tree vegetation, referred to as the treeline, and the zone of permanent snow and ice. The altitude at which the treeline occurs varies around the world depending on latitude and proximity to the ocean. For instance, the treeline in the oceanic Kosciuszko Mountain Range of Australia is located at about 1800 m above sea level, while the treeline in the continental Rocky Mountains of the USA is around 3000 m above sea level.

Documented reductions in both alpine glaciers and permafrost ice provide some of the clearest scientific evidence for responses of natural ecosystems to recent climatic warming. Direct measurement shows a significant warming of the high latitude, mostly continental mountains of the European Alps and Asia, whilst upward movement of plant species has been widely reported. In the tropical alpine zone, evidence for responses to recent climate changes is less pronounced to date.

Climate change projections are highly variable and uncertain for the alpine zone due to the inherent complexity of mountain systems and the lack of appropriate spatial and topographic resolution in most climate models. However, it is generally considered that warming will reduce the period of time that snow cover persists, hence in uencing both the timing and amplitude of runoff during snowmelt. This shift towards earlier spring runoff may result in a prolonged summer, increasing the dryness of alpine vegetation and soil and consequently increasing the risk of fire. The seasonality and amount of precipitation may change, as might the proportion of rain to snow, but the magnitude and direction of this change will vary from region to region. These changes could be quite significant. For example, scenarios for the European Western Alps under doubling of CO_2 concentrations (about year 2080–2120) suggest an increase in winter precipitation of up to 15% but a decrease in summer precipitation and a warming of up to 4 °C. There could also be changes in the frequency of extreme climatic events (e.g., storms), but again, it is unclear how these changes will be manifested for each particular region.

Alpine zones are extremely vulnerable to the effects of climate change because they are generally characterized by specialized ora and fauna that have evolved to contend with harsh climatic conditions, such as intense cold, high winds and heavy snowfall. In the case of plants, these morphological and physiological adjustments often include the development of stunted growth forms and small leaves. With the onset of climate change, the unique environment to which these species are often restricted is likely to suffer from a reduction in extent and increased fragmentation. This will place added pressure and stress on a fragile alpine ora and fauna, much of which is already highly endemic and rare, due to a low ability to disperse or migrate. Alpine areas and other montane systems can be thought of as mountain islands. As temperature increases, species need to migrate higher and higher to stay in the same climatic zone resulting in progressive isolation of populations and reductions of suitable areas (as mountain tops are smaller than their bases). The end result could be the extinction of particularly sensitive alpine taxa, and a significant reduction in total alpine biodiversity. This process is already being observed widely with tree establishment at elevations above current treelines. In the European Alps, plant species richness has increased at 70% of the sites studied presumably from upwards colonization; some plant populations have migrated upwards at a rate of $1-4\,\mathrm{m\,decade^{-1}}$. However, even this rate may not be fast enough to maintain some species in their preferred environment as there is a general rule that a 1 °C rise in temperatures equates to about a 170 m increase in altitude. Extinctions have occurred in the past where plant species that were already restricted to high elevations have been exposed to environmental changes including warming.

For many alpine zones, predicting how the vegetation will be affected by climate change is complicated due to uncertainties about species-specific responses to increased CO_2 as well as regional uncertainties about the level of change in other important environmental variables. For instance, temperature, solar radiation, precipitation, wind speed and soil moisture will also in uence species responses, but the degree to which these factors change will vary considerably across regions. The overall effect may be an increase in competition between plant species as the ecological potential of sites change. For example, those plant species migrating upward in response to warming are likely to compete with other plant species that may be responding differently or more slowly.

In addition to the potential effect on biodiversity, climate change in the alpine zone could also affect the rate of land degradation. Alpine soils are highly susceptible to erosion, and it is possible that an increase in precipitation could also increase the incidence of debris ow, land slide and slope failure in many mountainous regions. This would obviously have many implications, but among the more serious for natural systems would be a potential deterioration of water quality in alpine lakes and streams, as well as a reduction in the overall productive potential of alpine soils.

While climate change is likely to have a significant impact on alpine biodiversity, it should be noted that many

of the alpine zones with greatest biodiversity, such as the Himalayas, are located in areas where human population pressures are also pronounced. Thus, the potential impact of climate change on the alpine zone must be considered within the context of other anthropogenic in uences, as well as the inherent fragility of alpine ecosystems.

See also: Mountain Climates, Volume 1.

Oceans

Oceans cover 71% of the earth. They are strongly in uenced by climate at small scales and over the short-term, through effects on waves and production and mixing of different water layers, and at larger scales through impacts on oceanic currents and heat storage. These factors in turn affect climate through redistribution of energy into the atmosphere and around the globe. For example, the ability of the oceans to store heat has mitigated the rate of recent climatic change. The ocean is also a major source and sink of greenhouse gases and aerosols containing some 60 times more carbon than the atmosphere. The strong linkages between the oceans and climate suggest that climate changes and oceanic changes will occur in tandem.

There are already many pressures on the world's oceans as a result of human activity. These include overfishing, population pressures in coastal zones, habitat destruction, increased pollution from atmospheric, terrestrial and oceanic sources and increased UV radiation. Climate change will add to these existing pressures. Climate change will affect the oceans through changes in sea surface temperature, sea level, ice cover, ocean circulation, and wave characteristics.

Ocean circulation is particularly important. Ocean currents transfer energy from the equatorial regions towards the poles and also vertically from the ocean surface into deeper waters; in this way these transfers also facilitate the movement of nutrients critical for many marine biological systems globally. For example, the highly productive areas of the Sargasso Sea, the eastern Pacific and the Southern Ocean are dependent on the nutrient transfers associated with these currents, and about 45% of the net primary production of the oceans occurs in the nutrient-rich upwelling zones of these currents. Changes in ocean currents have occurred over both historical and palaeological time frames, resulting in large impacts on marine ecosystems. One of the best-known examples of this is the impact of temporary cessation of the nutrient-rich cold-water upwelling off the South American coast during ENSO events, which results in large reductions in fish stocks. There are many other similar examples and these relationships are sufficiently strong, so that for many species, water temperature and current strength can be used to forecast abundance and distribution. A range of possible changes in oceanic currents could occur with climate change and all of these will have impacts on marine ecosystems. In the extreme, there is some concern about the stability of the largest current – the global conveyor belt or the thermohaline circulation – which sinks cold water in the North Atlantic which is carried through the Southern Atlantic, around the Southern Ocean and up the west coast of South America where it surfaces, to travel across the Pacific through the Indonesian Archipelago into the Indian Ocean, around the tip of South Africa and back up the Atlantic (*see* Atlantic Ocean, Volume 1). Breakdown of this current appears to have occurred previously from the incursion of large amounts of cold, fresh water in the Northern Atlantic – a possibility with melting of the northern ice stores and changes in Arctic hydrology and sea ice abundance under global warming. Such an event would, paradoxically, cool Western Europe by several degrees, but in the longer term may enhance atmospheric accumulation of greenhouse gases because of reduced CO_2 sequestration by the oceans. The impacts of such a breakdown on marine ecosystems would be immense but are difficult to predict with current knowledge. Even lesser changes may reduce the ow of oxygen-rich water to the deeper ocean resulting in anoxic conditions that will negatively affect those ecosystems.

There is growing evidence that marine mammal and seabird populations are strongly affected by climate variations, and thus there is growing concern in regard to the impacts of climate change. Whales and related species (Cetaceans) may be particularly vulnerable. Sea ice extent and thickness has already declined in the Arctic and Southern Ocean over the past few decades with, for example, Arctic sea ice thickness declining 40% and areal extent declining 6–9% over the past 20 years. These changes are projected to continue. The sea ice is particularly important because it can affect the abundance of krill, which is the main food source of cetaceans. Declines in sea ice may reduce food sources for krill during the winter, reduce suitable habitat, reduce protection from predators and increase competition from macrozooplankton. Reduced ice-edge foraging zones may also affect cetaceans. Commercial or other whaling operations, if continued, will add to any such pressures on populations. Similar negative impacts may also occur to polar bears and changes in sea ice may affect many other Arctic species with resultant impacts throughout the food chain.

Future changes in fisheries will be a combination of the impacts of climate change and fishing management practices, as well as other factors such as degradation of marine habitats. Observed declines in fish stocks and catches that have previously been ascribed to overfishing (e.g., the collapse of the Peruvian anchovy fishery from the mid-1970s to the mid-1980s) are now considered to have a climatic component. This sensitivity to climate emphasizes the possibility that climate change may have significant impacts on fishery productivity. Globally, sea surface temperatures

have increased along with air temperatures over the past century and increased ocean temperatures are considered one of the most likely future changes. Sea temperature is an important regulator of fish behaviors and is well known to affect species such as sardines, bluefin tuna, skipjack tuna, North Sea cod and also seabirds, sea mammals and zooplankton through impacts on spawning, larval and juvenile survival and growth. In some cases, changes in temperature can adversely affect one species in a location whilst favoring another. In other cases, adverse effects are seen for whole ecosystems. Hence, climate change is likely to result in both positive and negative impacts and these may vary substantially between species. Regional-scale predictions are not feasible at present. Where negative impacts occur, these may be particularly harmful if there is no corresponding reduction in fishery activity. Hence, more informed fisheries management will become even more critical.

Seabirds are an important component of marine ecosystems. For example, in some systems they can consume up to 20–30% of annual fish production. Their mobility and capacity to adapt to different feeding sources often allows them to avoid the consequences of short-term variations in climate. However, the observed sensitivity of seabird populations to longer-scale climate–ocean variability and, in particular, the associated changes in fish abundance, suggest that there could be considerable impacts of climate change on sea birds.

The ENSO system is an important feature of the global climate and oceans. The possibility of the intensity or frequency of ENSO changing with global warming has raised concerns. There remains considerable uncertainty about these changes as they can be in uenced by the strength of trade winds, latitudinal temperature gradients and decadal scale climate–ocean variations. It is becoming increasingly apparent that decadal-scale variations in ocean–atmosphere patterns occur, and that these have major impacts on the location and abundance of marine life. There is a need to better understand these patterns and how they may change with global warming.

See also: Ocean Circulation, Volume 1; Ocean Conveyor Belt, Volume 1; Thermohaline Circulation, Volume 1.

Coastal Zones and Small Islands

The coastal margins of the oceans are very extensive (approximately 1 million km) and densely populated, with about 40% of the world's population living within 100 km of this margin. They are also biologically highly diverse and very productive. However, coastal zones and small islands are highly vulnerable to the effects of climate change through sea level rise, exposure to storm events and impacts from terrestrial catchments. Coral reefs, in particular, may be affected by sea surface warming and impacts, as well as increased CO_2 concentrations.

Global sea level is already increasing with average rises of between 1 and 2.5 mm year^{-1} although in some areas the effective sea level is falling due to geological uplift. This rate is anticipated to increase during this century, leading to sea level rises of 0.2–0.92 m by the year 2100. Continued rises are expected for several hundred years thereafter due to thermal expansion and ice sheet reduction even if atmospheric greenhouse gas levels are stabilized. Sea level rise will inundate coastal lowlands and wetlands, increase rates of coastal erosion and change coastal sedimentation patterns, increase saltwater intrusion into estuaries and fresh water aquifers, increase storm surge ooding risks, alter tidal ranges in estuaries and temporarily change the turbidity of the water, thus affecting light-sensitive organisms and degrading near-shore aquatic breeding areas and nurseries. Coastal ecosystems have the capacity to respond to some of these changes through increased growth or accretion rates for corals and mangrove communities where the rate of sea level rise does not exceed the response rate.

Coral reefs are very important biological systems, providing habitat for 25% of marine fish, and with more classes of organisms than rainforests. The high biological productivity of reefs and the physical protection they provide mean that they are important for adjacent ecosystems as well as for human use. Many reefs (58%) are currently considered to be threatened from existing human activities such as poor fishing practices, inappropriate tourism management, pollution and increased sedimentation and land reclamation. These stresses can increase the susceptibility of reefs to diseases and predators such as the crown of thorns starfish. Healthy coral reefs may be able to grow at up to 10 mm year^{-1} which will probably allow them to keep pace with sea level rise. However, these rates of growth may not be able to be achieved given the above stresses, which may predispose corals to damage from possible increase in coral bleaching events, changes in storm patterns and CO_2 increase. Coral bleaching is a normal event in reef ecosystems and occurs because many coral reefs exist at or close to temperature tolerance thresholds. It is enhanced by high levels of solar radiation and by in uxes of freshwater and riverine sediment. However, the frequency, extent and severity of bleaching events appear to be increasing already in response to general increases in ocean temperatures. The 1998 bleaching event was unprecedented historically and was associated with highest sea surface temperatures on record. Tropical oceans are expected to warm substantially and if 1998 was an indication of the possible future environment in tropical locations, then climate change may have significant biodiversity implications for coral reefs. This is because some coral groups and species such as the branching corals are more adversely affected than others, like massive brain corals. Corals affected by bleaching are more prone to attack from diseases and predators. The combination of these stresses is thought to potentially result in major

changes in community composition and function, such as reversion to soft coral growth forms, algae and sponges. Additional stresses may arise from increased rainfall intensity increasing freshwater out ow events from rivers and also from increases of atmospheric CO_2 potentially reducing reef calcification rates (by 17–35% with doubling of CO_2) as a result of increased sea water acidity. The combination of these factors suggests that in some areas, coral reefs may undergo significant changes in the future.

Mangrove and saltmarsh areas are also important, highly productive ecosystems that are critical for reproduction and growth of marine and estuary species as well as providing a buffer to adjacent terrestrial ecosystems. They occur between mean tide and high tide levels on coastlines where wave energy is low. In many regions, they are already under threat from economic development. These ecosystems will be directly affected by increases in sea level. Where there is low sediment supply and limited rates of organic matter accumulation, there may be landward retreat of the community, which if restricted by steep backshores, will result in reductions in community size and species diversity. In other areas where sediment supply is adequate and/or where organic matter accumulation is high, there may be opportunity for both seaward and landward expansions. Even where such conditions arise, increased storm activity and erosion rates from climate change may limit expansion.

Sea level rise and groundwater extractions may combine to significantly threaten underground fresh water storage in coastal and island environments. This could have significant impacts on the plant communities that grow in these areas, both from salinization in the root zone and from subsidence, which may increase exposure to inundation and wave activity. Similar increases in exposure and erosion may occur with other coastal ecosystems such as those existing on dunes and in deltas. In some cases, human infrastructure may increase impacts by not allowing movement of systems such as dunes or by reducing the energy-absorptive capacity of the shoreline, thus increasing erosive potential.

Aquaculture is a rapidly growing coastal zone industry approximately tripling in production from 1990 to the present. Further increases in demand are anticipated. These fisheries use large amounts of fishmeal and fish oil and the supply of these could be reduced from climate change impacts on the fish species used. Further impacts may occur through higher water and air temperatures in the feeding pens. In cooler climates there may be a lengthening of the growing season, increasing growth rates and feed conversion efficiency. However, warmer water and associated reductions in dissolved oxygen levels may increase the intensity and frequency of disease outbreaks and the occurrence of toxic algal blooms which, along with the increased nutrient loads may have adverse impacts on adjacent ecosystems. As with other coastal infrastructure there may be increased damage from climatic events.

See also: Aquaculture and Environment: Global View from the Tropics to High Latitudes, Volume 3.

Tropical cyclones have significant impacts on coastal zones. Recent analyses suggest little change in frequency or initial locations of storms, although the maximum intensities may increase by 10–20%. Increased ocean temperatures may allow the storms to intrude further into the mid-latitude belt and persist longer. Such storms will result in increased intensity of rainfall events in coastal catchments, which may increase sediment loads. Whilst this may reduce coastal erosion rates and in some areas increase coastal accretion, it could also reduce ecosystem productivity through reducing light penetration in coastal waters. This would particularly affect communities such as sea grasses, which are a significant habitat for many species. The impacts of such reductions in habitat quality from climate changes are likely to be enhanced by human activities such as the transfer of exotic organisms in bilge water, pollution and increased disturbance.

The fossil record suggests that poleward migration of the component species of terrestrial ecosystems has occurred in past global warming episodes. Migration may be feasible for some species over the next centuries. However, southerly or northerly coastal zones represent major migration barriers to many species. Thus, there could conceivably be a concentration of organisms along such coasts. This could present a significant challenge for biodiversity management given the array of new interactions that would happen in such cases and the high level of disturbance and fragmentation in coastal areas due to the generally high human population densities found in such environs. Regions particularly prone to such issues include the megadiverse Cape region of South Africa and southwest Western Australia.

See also: Coral Reefs: an Ecosystem Subject to Multiple Environmental Threats, Volume 2; Intertidal Zones, Volume 2.

Freshwater and Wetland Systems

Freshwater ecosystems such as streams, lakes and inland wetlands, cover only a relative small area of the Earth's land surface (wetlands are 2.3%). They account for only 0.009% of the total volume of water on Earth, with groundwater approximately 66 times more abundant than surface freshwater (*see* Marshes, Anthropogenic Changes, Volume 3). They have an enormous diversity of fish, amphibians, aquatic plants, invertebrates and microorganisms, as well as being an important refuge, feeding and breeding ground for birds and many other terrestrial fauna. They provide a fundamental resource for humans, and have significantly in uenced economic activity and the distribution of human

settlements. Human activities have profoundly impacted on these ecosystems through extractions, drainage, diversions, dams, pollution and species introductions.

Freshwater ecosystems are potentially very sensitive to climate change because of the strong linkages between climate and regional hydrology. Many of these linkages are well understood, but despite this, projections of climate change for freshwater systems remain very uncertain. This is primarily due to the coarse spatial resolution of most current general circulation models and the disparity between the timescales for which information is often provided and those that impact on hydrology (sub-daily). As a result, it is difficult to make broad generalizations about the impact of climate change on freshwater ecosystems without specific reference to particular studies, locations or ecosystem types.

Hydrological regimes are highly sensitive to changes in rainfall. Reductions in precipitation disproportionately reduce mean flows (e.g., a 1% reduction in rainfall results in a 3% reduction in flow in the Gambia River, Africa), increase frequency of low-flow events and reduce frequency of floods. Similarly, increases in rainfall tend to disproportionately increase the frequency of floods. Thus, a critical impact of climate change will be altering the variability of the flow regime and the impacts that this will have on the biology, biogeochemistry and hydrological functions of the associated ecosystems. The sensitivity of the system to such changes tends to be greater in dry environments than in wetter ones. Changes in the behavior of large-scale systems that cause climate variability, such as ENSO, will further affect hydrology. As noted previously, human activities have profoundly affected riverine systems through the construction of large in-stream water storages, small off-stream water storages (e.g., farm dams) and extractions for irrigation. These activities are expected to increase. For example, all the surface water of India will be needed for irrigated crops to just meet growing domestic food demands by 2040. These impacts will interact with any alterations in hydrology caused by climate changes.

The degree to which extreme events, such as floods and drought, impact upon a region is determined by the physical and biological characteristics of the catchment and the degree of human-induced change and not simply climate alone. For this reason, there is a need to consider the impact of climate change on the source/sink relationship between terrestrial and freshwater ecosystems, in order to gain a more holistic understanding of the full implications of climate change. Predictions for terrestrial ecosystems suggest that vegetation composition and distribution will be influenced by changes in temperature and rainfall. If terrestrial production increases, there will be an associated increase in the supply of detritus to adjacent aquatic systems via an effective riparian filter system, which may in turn increase aquatic productivity. Alternatively, decreases in terrestrial productivity caused by a loss of forest cover may degrade the capacity of riparian vegetation to filter out sediment and nutrients, increasing turbidity levels and creating nutrient imbalances that might result in a decrease in aquatic productivity or eutrophication and an increase in toxic algal bloom frequency.

Changes in water temperature and the thermal structure of water columns caused by climate warming are likely to have a major influence on the productivity, persistence and diversity of the species that inhabit these systems. This change might simply be caused by an increase in ambient air temperature, or alternatively, through a shift in the proportions of rain and snow in mountainous regions, resulting in a change in the timing and temperature of runoff. These changes are expected to be more intense at higher latitudes, but nevertheless, extinction of species may occur anywhere that dispersal requirements are prevented by physical or temperature-induced isolation within streams and lakes.

Climate change will affect the cycling of carbon within freshwater wetlands. Some wetlands currently acting as a CO_2 sink are likely to change to a CO_2 source in response to increased temperatures and/or a lowering of water tables. These changes may already be occurring in some high latitude regions. The magnitude and timing of methane emissions will change in response to changes in current wetting and drying regimes, with depletion of wetlands decreasing these emissions but with construction of more dams increasing them.

The distribution and area of wetlands are expected to change with greenhouse warming, but the direction and magnitude of this change are unclear. Any climate impacts will be an added stress as these systems are already under serious threat from human activity. For instance, half of the world's wetlands are estimated to have been destroyed since the beginning of the 20th century. This destruction is expected to increase, particularly in developing countries, and therefore any potential impact resulting from climate change must be placed within the context of further anthropogenic change.

See also: Mangrove Ecosystems, Volume 2; Marshes, Anthropogenic Changes, Volume 3.

ADAPTATION STRATEGIES

In almost all of the systems discussed above, climate change will occur in conjunction with a multitude of existing and growing pressures from human activities. Consequently, in many ecosystems, a prerequisite for adaptation to climate change is the development of more sustainable management regimes. Development of these regimes, as well as adaptation options to global change, need to take into account the institutional, legal, socio-economic and cultural context in which they will be applied. These are significant challenges. For example, climate change will possibly

make less relevant many of the world's existing biosphere conservation systems. Lack of connectivity between these areas means that, at best, they may become refugia for species that no longer are adapted to the local climate. Locations such as mountain tops, islands and coastal zones on poleward sides of continents will be particularly prone to becoming refugia. Over time, greater inputs will be needed to maintain the resident species against various threatening processes and there may be depletion of a diverse gene pool. Difficult decisions will arise over whether to transplant species from other refugia and what that would do to the existing resident species. The alternative of developing a substantial off-reserve conservation system is becoming more difficult as landuse intensifies, with competing priorities from different land users (e.g., tourism versus farmers with respect to the elephants in the Serengeti), with the continuing introduction of exotic plant and animal species, with growing political and institutional instability in some nations and with conversion of existing land-tenure arrangements.

Ecosystems will adapt to climate change and related factors. Actions can be taken that facilitate this adaptation (i.e., providing connectivity across landscapes for migration of species or emplacement of fish ladders across dam walls). There is also the possibility of more actively assisting these changes, such as transplanting species or attempting to ameliorate factors that may prevent some species from adapting (e.g., provision of environmental flows to wetlands). However, this can sometimes have unforeseen consequences (e.g., this may increase the threat to organisms which have adapted to the previous regime). These will only be understood and resolved by research into these ecosystems linked with appropriate monitoring and adaptive management.

Scenario analysis, which investigates impacts in, say, the year 2050 or 2100 has often resulted in climate change not being taken into account in conservation planning as it is beyond the planning horizon. However, in many ecosystems there is growing evidence that observed changes in climate are already having biological impacts that fit with *a priori* predictions of those under future climate change. At this stage it is not possible to attribute these climatic changes to human activities even though in many cases they are in accord with predictions of such changes. An argument can be made that these changes need to be considered in conservation planning regardless of whether they are due to longer-term variability or climate change. Furthermore, there are other factors, which are attributable to human activities (increased CO_2 levels, nitrogen deposition, tropospheric ozone and UV-B) that already are likely to be having impacts. Hence, global change should already be considered in the management of natural ecosystems both on the basis of the current changes listed above as well as anticipatory adaptation that addresses existing pressures, which will reduce vulnerability associated with climate change.

ACKNOWLEDGMENTS

Thanks to Dr Steve Crimp (Queensland Natural Resources) and Mike Doherty (CSIRO Sustainable Ecosystems) for their constructive comments on the draft manuscript.

FURTHER READING

Hughes, L (2000) Biological Consequences of Global Warming: is the Signal Already Apparent? *Trends Rev. Ecol. Evol.*, 15, 56–61.

Margules, C R and Pressey, R L (2000) Systematic Conservation Planning, *Nature*, 405, 243–253.

Sala, O E, Chapin, F S, Armesto, J J, Berlow, E, Bloomfield, J, Dirzo, R, Huber-Sanwald, E, Huenneke, L F, Jackson, R B, Kinzig, A, Leemans, T R, Lodge, D M, Mooney, H A, Oesterheld, M, Poff, N L, Sykes, M T, Walker, B H, Walker, M, and Wall, D H (2000) Global Biodiversity Scenarios for the Year 2100, *Science*, 287, 1770–1774.

Watson, R T, Zinyowera, M C, Moss, R H, and Dokken, D J (1996) Climate Change 1995: *Impacts, Adaptations and Mitigation of Climate Change: Scientific-Technical Analyses*, Cambridge University Press, Cambridge, 1–878.

Watson, R T, Zinyowera, M C, Moss, R H, and Dokken, D J (1998) *The Regional Impacts of Climate Change: an Assessment of Vulnerability*, Cambridge University Press, Cambridge, 1–517.

Plant Dispersal and Migration

JAMES S CLARK, B BECKAGE, J HILLERISLAMBERS, I IBANEZ, S LADEAU, J MCLACHLAN, J MOHAN, AND M ROCCA

Duke University, Durham, NC, USA

Although Americans may have first seen kudzu (Pueraria lobata) *on US soil at the Philadelphia Centennial Exposition of 1876, broad distribution throughout the Southeast did not come until decades later. Kudzu is known for its rapid growth, a potential to bury forest canopies in dense foliage in the course of a single growing season. It is natural to attribute its success as an alien invader to an aggressive habit that co-opts the light supply of native vegetation. But kudzu's inherent invasive potential may be overrated. The modern distribution is the result of a determined campaign to broadly disseminate the species (reproduction is almost entirely vegetative). Even retail (including mail order) sales of the curious ornamental during the early decades of this century might not have been enough to establish it as the dominant alien it has become. That dominance dates from the 1930s, when the Soil Conservation Service paid farmers $8.00 an acre to stabilize their fallow fields, and the Civilian Conservation Corps distributed kudzu widely to control erosion (Miller and Edwards, 1982). In the absence of detailed records of planting during the Depression and additional establishment since, history provides few insights regarding its potential for spread in the absence of intentional propagation by humans.*

Worldwide invasions of exotic plants, together with environmental changes that threaten native species, challenge ecologists to understand population spread in the context of dispersal potential. Many countries now support resident populations of hundreds to thousands of invasive exotic plant species. Exotics represent greater than 50% of the flora on many remote islands (Vitousek et al.*, 1996). Biological invasions by exotic species change the composition and community structure of invaded areas. In addition to the obvious effects of competitive exclusion, such as occurred with yellow star thistle's* (Centaurea solstitalis) *invasion of northern California grasslands, there are less direct effects. Some of the more devastating consequences of cheatgrass'* (Bromus tectorum) *invasion of the intermountain West came from increased fire frequency, which results in a loss of shrubs (Whisenant, 1989). Introduced cordgrasses* (Spartina *spp*) *on tidal flats around the globe have devastated shorebird populations that depend on tidal flats for food (Daehler and Strong, 1996). There is increasing evidence that alien invaders can alter properties of whole ecosystems, including productivity, nutrient cycling, and hydrology (Vitousek, 1990).*

INTRODUCTION

The globe's 250 000 extant species of seed plants have unique distributions and migration histories. *Plant migration* is defined here as a change in the distribution of mature plants that spans more than a generation. Distributions can be continuous or discontinuous. Changes in distributions include both expansions and contractions, but we focus here on expansions. Geographic distributions represent a balance between seed dispersal and vegetative spread, both of which tend to extend the range, and environmental constraints on survival that keep it in check. Population frontiers are static when dispersal does not bring propagules to areas outside the range where they currently survive. Migration occurs when there is a change in dispersal, the environment, or both.

Dispersal and the potential for population spread have dual implications for populations confronted by global change. On the one hand, migration potential, typically aided by human transport, has brought epidemics of exotic aliens with attendant costs in biodiversity loss, restructuring of ecosystems, and the burdens of control measures. Potential forage species, such as alien grasses introduced in the American West (Mack, 1986) and prickly pear (*Opuntia stricta*) in Australia (Parker, 1977) have had severe economic consequences. Aquatic ornamentals, such as water hyacinth (*Eichhornia crassipes*), clog waterways in the southeastern US. Pimentel *et al.* (2000) estimate that 35 billion dollars are spent annually on control of invasive plants in the US. Eradicating or delaying the spread produces unintended consequences, and success rate is low (Daehler and Strong, 1996). Educational activities aimed at

alerting the public to potentially dangerous exotics date to the 19th century, and include bulletins that would aid identifications of seed contaminants by farmers (Mack, 1986). Efforts to increase public awareness might help stiffen existing laws and policies aimed at reducing invasions (Vitousek *et al.*, 1996), but effective legislation requires an understanding of how migrations occur.

On the other hand, a potential to migrate has saved species from extinction during repeated climate changes of the past. The legacy of prehistoric and contemporary plant migrations includes many examples of environmental change, both natural and human-caused, that have led to continental-scale changes in species distributions. The migration potential required to survive past environmental change will be needed again as global change accelerates. Recognition of species threatened by rapid climate change requires an understanding of migration potential.

The history of kudzu spread in the southeastern US captures many of the frustrations of determining how dispersal figures in plant migration potential. Differential ability to invade new environments begs the question of just how important is dispersal. Of the 25 000 alien plant species in Florida, an impressive 900 have become established in native vegetation (Pimentel *et al.*, 2000). The flip side of this calculation is the $>24\,000$ species that can grow there but have not escaped to become components of unmanaged vegetation. Seed dispersal does not appear to contribute importantly to contemporary spread of kudzu. We do know that the modern dominance of this perennial vine began with a widespread planting campaign. Clearly, dispersal potential is only one of many factors that might ultimately predict the potential for and rate of spread.

Although we focus here on how dispersal affects the potential for migration, most of the examples we consider involve widespread changes in the environment, including climate, land cover, or trophic structure (e.g., introductions of other aliens or absence of natural enemies in novel environments). We begin with a brief summary of what we have learned from models of migration, all of which have dispersal at their core. We then summarize the state of our knowledge on dispersal biology, with special emphasis on long-distance events and how well we can estimate them. We then turn to past plant migrations, both prehistoric and contemporary, for the light they can shed on migration potential. Finally, we synthesize the overriding themes that emerge from theory, dispersal biology, and the prehistoric and historic record as a basis for some general guidelines that may help us forecast future migrations and areas needing further research.

THE ELEMENTS OF PLANT MIGRATION

The Dispersal Vectors that Account for Spread

Analyses of factors that might control migration potential identify as key variables total seed production and extreme dispersal events. These two variables go hand in hand, because extreme events become increasingly probable with increased seed production.

The scatter of seed about the parent plant is termed the *seed shadow*. The shape of the seed shadow is determined by the dispersal kernel, which describes the probability that a seed released from the parent plant will travel a given distance. The density of seed expected to settle a given distance from the parent is found by multiplying the dispersal kernel by total seed production, or fecundity. Dispersal kernels for seeds are termed *fat-tailed*, the tail referring to the few long-distance dispersal events. Although these long-distance events are rare, they can still dominate the rate of spread (Figure 1). One rare long-distance event can

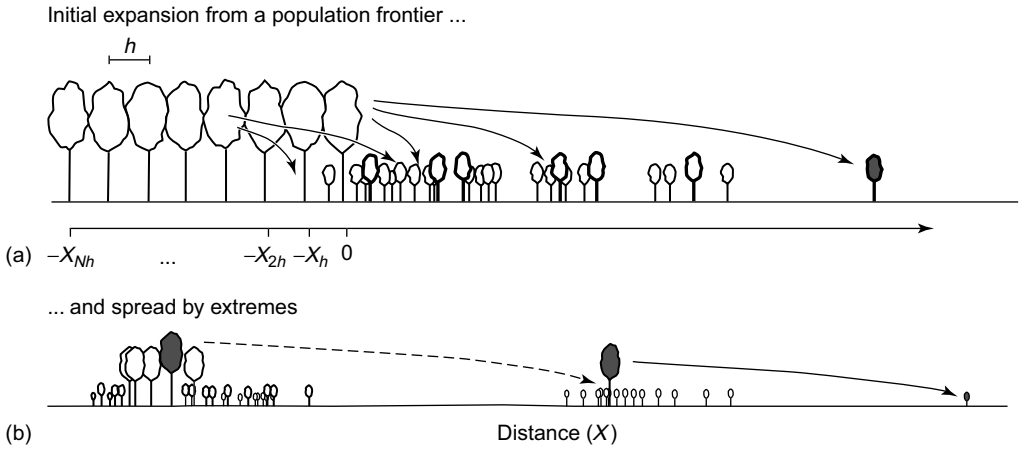


Figure 1 A plant migration includes both diffusive spread that results from local dispersal as well as rare, extreme events that can result in fast and erratic invasion. Spread from a coherent population frontier can proceed rapidly due to combined seed production of large numbers of adults (a), whereas total seed availability diminishes far from the population frontier (b). (Reproduced with permission of the University of Chicago Press from Clark *et al.*, 2001)

overshadow the contributions of the bulk of seeds that land near the parent.

To gain a rough idea of how fecundity interacts with the shape of the seed shadow, compare the velocities of spread predicted by simple models of diffusion and fat-tailed spread. Diffusion refers to a traveling wave of advance that results in the absence of rare, long-distance events. The velocity of this traveling wave is proportional to the square root of the log of lifetime seed production R_0, or net reproduction rate (Kolmogorov *et al.*, 1937; Skellam, 1951). Thus, the rate of spread is only weakly affected by total seed production. By contrast, the velocity of spread expected from dispersal kernels fitted to trees is proportional to the square root of lifetime seed production (Clark *et al.*, 2001).

These two results diverge as R_0 becomes large, because the tail of the dispersal kernel translates at least some of the increased seed production into long-distance events. Diffusive spread occurs when such events cannot occur, regardless of fecundity.

Despite a large dispersal biology literature, seed production and dispersal are poorly quantified for most species in most settings. In general, dispersal data tell us that plants tend to have fat-tailed dispersal kernels (Portnoy and Willson, 1993; Clark *et al.*, 1999a) and dispersal distance depends on aerodynamic properties of seed (Figure 2). Models tell us that when the kernel is fat-tailed, extreme dispersal events control the rate of spread (Kot *et al.*, 1996; Lewis, 1997; Clark, 1998), and total seed production has strong impact on the extremes (Clark *et al.*, 2001a,b). Especially high fecundity is a clue that extreme events can be particularly important. Thus, predicting migration potential by any dispersal mechanism depends on our ability to estimate extreme events and seed production.

Figure 2 The samaras of *Acer rubrum* seed are an example of specialized structures for wind dispersal

Dispersal by Wind

Long-distance dispersal by wind is poorly quantified, but seed trap studies combined with observational data suggest that it can range over many kilometers. Wind dispersal depends on aerodynamic properties of seed, which vary among species, and on wind fields, which vary in space and time. Many species have seeds that are highly specialized for wind dispersal. Plumed fruits and the parachutes of *Taraxacum* and other Liguliflorae (Compositae) provide for broad dissemination in wind velocities of less than $1\,\mathrm{m\,sec^{-1}}$ (Weaver and Clements, 1938). The spread of tumbleweed (including *Salsola kali*) through the Dakotas near the turn of the century was facilitated by saltation of dead plants that scatter seed in transit. Larger seeds have shorter extremes. The well-developed samaras of *Acer rubrum* (Figure 2) increase drag and, thus, reduce fall velocity. The few studies that have measured wind dispersal for multiple years at a single site report significant interannual (Houle, 1998; Clark *et al.*, 1999b) and stand (Clark *et al.*, 1999b) variability in seed production, but shape of the seed shadow is relatively consistent (Clark *et al.*, 1999a). Dispersal distances for temperate deciduous, mixed conifer, and lowland tropical forest species average less than 40 m, with highest mean values for species that produce especially small seeds (e.g., *Betula*) (Gladstone, 1979; Ribbens *et al.*, 1994; Clark *et al.*, 1998, 1999a). Wind-dispersal is more efficient in areas of limited vegetation cover, which reduces wind velocity near the ground.

Long distance dispersal is erratic and hard to predict. Extreme dispersal is not quantified, but fitted dispersal models are fat-tailed at distances of $<100\,\mathrm{m}$ (Clark *et al.*, 1999a), and there are examples of population spread due to seed movement in open environments that exceed 10^3 m. For example, colonization of Glacier Bay at rates of $300-400\,\mathrm{m\,year^{-1}}$ (Fastie, 1995) is comparable with rates inferred from fossil pollen data in the early Holocene (Ritchie and MacDonald, 1986). Storms can transport objects larger than seeds $>100\,\mathrm{km}$ (Snow *et al.*, 1995).

The importance of long distance dispersal for spread is evident on isolated islands, which cannot be reached by diffusive spread. For example, 30% of the seed-bearing plants on Krakatau ($\sim$30 km from Java and 12 km from the

nearest small island) 15 years after the 1883 eruption were brought by wind (Ernst, 1908). Although long-distance dispersal by wind was important, wind-dispersed taxa were underrepresented.

Extreme dispersal by wind is poorly quantified, but there is sufficient evidence to conclude that it is a plausible mechanism for rapid spread of relatively small-seeded species that release seeds in the canopy. It is a less plausible mechanism for especially large seed and for seeds released in dense understories.

Dispersal by Water

Seed transport by water (*hydrochory*) permits spread to islands, along coastlines, across large water bodies, and along river systems. Potential for long-distance dispersal by water is more readily estimated than for wind, because the direction of transport is more predictable, as are advection rates (flow velocities in rivers or ocean currents) and residence times in the water column, which depend on seed properties. For example, the tropical Atlantic shores of South America and Africa are linked by equatorial currents, and they share a common mangrove flora. This flora is distinct from the East Africa, India, and Malaya mangrove flora (Guppy, 1917). Lack of a connection between east and west African coasts, despite their proximity, explains their lack of common species. The efficacy of water transport is apparent from observational studies (e.g., Weaver and Clements, 1938; Craddock and Huenneke, 1997), and it is the only explanation for the presence of some species in remote sites (Godley, 1967).

Colonization of remote islands and spread along coastlines requires adaptations for flotation and salt tolerance. Transport is governed by ocean circulation and by local tidal currents (Huiskes *et al.*, 1995). *Cocos nucifera* (Arecaceae) with its corky husk withstands some exposure to salt water (Schimper, 1891). Small hairs on the seeds of *Salicornia*, a salt marsh chenopod, trap air and facilitate dispersal with tides (Dalby, 1963). Other adaptations involve enlargement of the diaspore surface (e.g., *Matricaria maritima*, Ridley, 1930) and buoyant structures that remain attached to seed, including carpels (Guppy, 1917), perianth bracts (König, 1960), and fragments of the inflorescence (Dalby, 1963). Many mangrove species (involving several angiosperm families) are viviparous, whereby seedlings develop while still attached to the parent plant. Seedlings photosynthesize and obtain nutrients and water while borne on ocean currents and can remain viable for a year or more (Davis, 1940). Seedlings of red mangrove (*Rhizophora mangle*) have a buoyant hypocotyl, whereas other species have buoyant pericarps, cotyledons, or entire embryos (Rabinowitz, 1978).

Extended viability in ocean currents can result in extreme long-distance colonization. The isolated estuaries of the Pacific coast represent islands of suitable habitat for introduced cordgrasses (*Spartina*) that are invaded by propagules borne on tidal currents (Daehler and Strong, 1996). Some mangrove species are capable of voyages spanning the largest oceans (including the Indian, western Pacific, and Atlantic, Sauer, 1988). Accumulation of species on remote oceanic islands is probably limited, in part, by water dispersal (Ernst, 1908).

Rivers can provide corridors for rapid migration. In riparian zones, some trees time fruiting to coincide with seasonal floods (Schneider and Sharitz, 1988; Kubitzki and Ziburski, 1994). Thebaud and Debussche (1991) suggest that water transport during autumn floods facilitates the spread of the typically wind-dispersed species *Fraxinus ornus*. Seeds dispersed by water first colonize littoral (Murray, 1986) or riparian zones, but a spread to upland areas can follow. North-flowing rivers of Eurasia might have accelerated plant migrations following the last ice age (Huntley and Birks, 1983).

Water dispersal can result in rapid spread, but it is limited by the distribution and velocity of water currents. For example, Eurasia contains many northward flowing rivers that might facilitate post-Glacial spread to high latitudes, whereas eastern North America does not.

Animals

Many of the hypotheses that concern rapid plant migration involve vertebrate vectors (Reid, 1899; Johnson and Adkisson, 1985). Vertebrates disperse seed by ingesting fruits and passing or regurgitating seeds intact (endozoochory), by caching (dyszoochory), and by carrying seeds that adhere to fur or feathers (epizoochory) (Howe and Westley, 1997). The most important vertebrate dispersal vectors are: (i) hoarding mammals and birds; (ii) arboreal frugivorous mammals that consume fruits with arils or pulps containing protein, sugar, or starch; (iii) bats that feed on pulps rich in lipids or proteins; and (iv) frugivorous birds (Howe and Westley, 1997). Birds and mammals are the most common vertebrate dispersal agents, but other vertebrate vectors include fish (Goulding, 1980), turtles (Moll and Janzen, 1995), and lizards (Nogales *et al.*, 1998). Ants disperse large quantities of seed, but dispersal distances are typically small, averaging perhaps one meter and ranging to 10 m (Hughes and Westoby, 1992). Vertebrates can process a large portion (e.g., $>70\%$) of a plant's seed or fruit production (Herrera *et al.*, 1994; Vander Wall, 1994) and, therefore, may have a large impact on dispersal. Dispersal is further influenced by birds that prey on other vertebrates that have consumed seed (Reid, 1899; Nogales *et al.*, 1998).

Plant adaptations that facilitate spread by vertebrates have probably contributed to past invasions. Darwin (1859) reports >80 seeds of several species in dried mud taken from a partridge leg. Elaborate adaptations to attract frugivores include thickened seed coats surrounded by brilliant, multicolored fleshy pericarp (Howe and Westley, 1997). Reproductive structures that adhere to animal fur may have

facilitated 20th century invasion of the intermountain West by cheatgrass. Barbs of *Xanthium* and *Bidens* fruits allow for similar mobility. Remote sites may be especially dependent on vertebrate vectors. Birds may have brought >75% of the flora to the Galapagos (Porter, 1976) and smaller fractions of total floras to islands that are less remote (Ernst, 1908). Vertebrates as large as elephants are important dispersers for some plant species today (Chapman *et al.*, 1992), and loss of Pleistocene megafauna may have subsequently limited spread of these species (Janzen and Martin, 1982).

The benefits of animal dispersal can be overemphasized, because losses are typically large (Schupp, 1993). Spread is facilitated not only by transport (Janzen, 1971; Harper, 1977) but also by scarification (pregerminative influences that make seed coats permeable to water and gases) (Fenner, 1985; Chapman *et al.*, 1992) and provision of concentrated resources, such as in animal dung high in nitrogen and phosphorus (Bazzaz, 1991; Chapman *et al.*, 1992). But most vertebrate vectors are seed predators. As little as 1% of seed handled by vertebrate seed predators may escape eventual consumption (Cahalane, 1942), although it can range as high as 15 to 75% (Forget, 1990; Steele and Smallwood, 1994), particularly in years of high seed production. Even frugivores, which pass seed through the intestinal tract intact, can lower viability (Murray, 1988).

Vertebrate dispersal can be more directed than wind dispersal in ways that might have particular significance for migration. Seeds are moved to microsites or habitats that may be especially suited to seedling survival (Hulme, 1997; Hoshizaki *et al.*, 1997; Vander Wall, 1994). Seed burial within specific habitats (Yasunda *et al.*, 2000) can promote establishment success (Forget, 1990, 1991). In one study, common ravens dispersed numerous seeds and placed 75% in habitats favorable for germination (Nogales *et al.*, 1999). Blue jays preferentially cache seeds in regenerating woodland and edge habitats, which may have facilitated spread in the past (Johnson *et al.*, 1997; Kollmann and Pirl, 1995). Herrera *et al.* (1994) found that birds tend to disperse seeds to forest–gap interfaces. Transport to favorable habitats not only increases establishment success, but can also affect dispersal distance, depending on the distribution of land cover.

Dispersal distances vary greatly among vertebrate vectors. Rodents move seeds 10^0 to 10^1 m (Forget, 1990; Kollmann and Schill, 1996) with maximum recorded distances ranging from 20–70 m (Forget, 1990; Tamura and Shibasaki, 1996; Yasunda *et al.*, 2000). Rodent seed caches are repeatedly moved and sometimes robbed by other rodents, potentially increasing the distance seeds are dispersed beyond a single individual's home range (Tamura and Shibasaki, 1996; Vander Wall, 1994). Foxes and bears, which consume fruits, may travel 10 km in a day (Storm and Montgomery, 1975) suggesting potential for long-distance dispersal. Old World frugivorous bats have the potential to disperse small seeds hundreds of kilometers (Shilton *et al.*, 1999). Primate seed dispersal in a Columbian forest ranged from 218 ± 82 to 354 ± 199 m (depending on species) with maximum values exceeding 600 m (Yumoto *et al.*, 1999). Birds commonly move seeds 10^2 to 10^4 m (Kollmann and Schill, 1996; Webb, 1986; Murray, 1988; Johnson and Adkisson, 1985; Vander Wall and Balda, 1977).

Despite obvious potential for long distance dispersal by large vertebrates, there are few quantitative data on extremes. As with other vectors, the most important dispersal events, i.e., the long distance ones, are most likely to be missed. Most of reported extreme distances by birds and large vertebrates are less than 10 km. But storms can disperse birds over large areas resulting in extremes that are difficult to quantify.

LESSONS FROM THE PAST

Recent global warming does not represent the first time that survival of plant species has depended on migration. Ice ages saw repeated and sometimes rapid climate changes. In regions of steep climate gradients and topographic complexity, such as the Atherton Tableland of northern Queensland, declining aridity during the Holocene meant that rainforest species may have migrated only tens of kilometers to replace the sclerophyll vegetation that dominated during Glacial times (Kershaw, 1993; Hopkins *et al.*, 1990). Mountains of southern Europe (Tzedakis, 1993), western North America (Barnosky *et al.*, 1987), South America (Hoogheimstra, 1984), and Africa (Street-Perrott *et al.*, 1997) support, within a given region, most of the same species today that they did during the last glacial maximum (LGM) (*see* **Last Glacial Maximum**, Volume 1). Topographic complexity provides for a range of local climatic conditions that buffer the effects of climate change.

In areas of low relief, such as eastern North America and western Eurasia, suitable habitats for many species shifted across continents. Populations that have survived in these regions migrated repeatedly over successive glacial cycles, perhaps at rates exceeding 10^2 m year^{-1} (Davis, 1986; Huntley and Birks, 1983). Ever since a rough time scale for Pleistocene climate changes became available, ecologists have been impressed with the seemingly impossible dispersal distances required to explain post-glacial spread of plant populations at these continental scales (Reid, 1899; Davis, 1987; Clark *et al.*, 1998b, Figure 8). In eastern North America and western Europe, glacial distributions of temperate species are thought to have been well south of modern distributions. Although the fossil evidence for glacial distributions is spotty, the southern edge of the Laurentide ice sheet in North America represents an extreme northern range limit for most species. Populations that now occupy regions north of the southern ice margin must

have migrated at least that far since the LGM, and most temperate species are thought to have come from refuges much farther south.

Diffusive spread is too slow to account for migration of trees (Skellam, 1951; Clark, 1998), and they are especially inadequate for herbaceous plants. Cain *et al.* (1998) concluded that there is no documented mechanism by which most woodland herbs could have reached their modern geographical ranges since the LGM. For many woodland herbs, ant dispersal is inadequate, and seeds released only centimeters above the forest floor are not susceptible to transport at wind speeds typical of forest understories. Occasional long-distance dispersal events appear to be the only explanation for Holocene colonization of northern temperate forest by woodland herbs.

The fact that dispersal data from modern plant populations do not agree with the rapid spread inferred from the paleo record has been termed *Reid's Paradox* (Clark *et al.*, 1998). Davis (1987) envisioned migration occurring as a series of long-distance leaps by a few propagules that disperse well beyond the typical seed shadow (Figure 1). Birds (Reid, 1899; Webb, 1987; Wilkinson, 1997), wind storms (King and Herstrom, 1997), and water transport (Firbas, 1948) are all cited as potential dispersal vectors. Clark (1998) showed that a small amount of long-distance dispersal is consistent with actual dispersal data that might agree with the rapid spread of some paleo-species.

There is increasing evidence for barriers to migration that limited spread of species or genotypes. Molecular (chloroplast DNA) data from tree populations in western Europe suggest that east–west oriented mountain ranges profoundly affected dispersal and, thus, migration at the end of the last ice age. For example, during the LGM, European beech (*Fagus grandifolia*) was present in both Italy and the Carpathians. But the Italian populations do not appear to have traversed the Alps, and northern European populations originated in the Carpathians (Hewitt, 1999).

Some of these same geographic boundaries bolster the argument that long-distance dispersal contributed to past migrations. Fossil pollen data indicate that the Baltic and North Seas (Kullman, 1996) and the Great Lakes (Webb, 1997; Woods and Davis, 1989) did not represent important obstacles for many species, but oceans did. In contrast to European beech, oak species in western Europe appear to have followed several migrational pathways (Figure 3b). A decline of genetic diversity with distance away from Pleistocene refugia (Ibrahim *et al.*, 1996; LeCorre *et al.*, 1997) and highly structured regional populations (Petit *et al.*, 1997) have been interpreted as consistent with long-distance dispersal events.

Although they represent one of the few plausible alternatives, extreme events, such as those described by fat-tailed dispersal kernels, may not be the complete solution to Reid's Paradox. Clark *et al.* (2001) demonstrate that traditional ways to estimate spread from dispersal data overestimate the potential velocity. A reassessment of migration potential parameterized with modern dispersal data predicts much slower spread than has been interpreted from the paleorecord. No model can rule out the possibility of past long-distance events, but these new analyses suggest that rapid migration is harder to achieve than we previously thought.

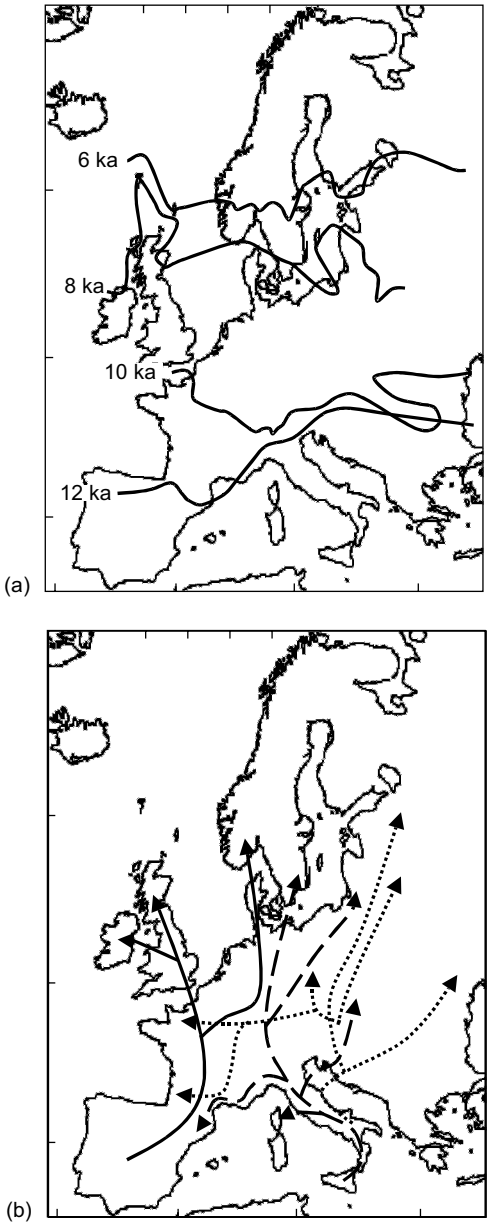


Figure 3 Two perspectives on Holocene oak migration in Europe. (a) The 2% deciduous oak isopols contoured from oak pollen recovered from sediment cores (adapted from Huntley and Birks, 1983). (b) Migration routes reconstructed from cpDNA of modern oaks (adapted from Dumolin-Lepegue *et al.*, 1997)

The reanalysis that revises lower predicted rates of spread suggests that the interpretation of the fossil record could bear further consideration. Moreover, not all interpretations of the paleorecord imply the high rates of spread that seem necessary to explain migration of temperate trees from the southern US to temperate latitudes. Bennett (1985) argues that temperate species may have existed at temperate latitudes (albeit, south of the Laurentide ice margin) before the Holocene. He points out that the fossil pollen data that have been used to interpret spread do not provide concrete evidence for the presence of populations at low density. In North America, two recent studies suggest the presence of temperate hardwoods at least as far north as 35 °N during the LGM (Russell and Sanford, 2000; Jackson *et al.*, 2000), which does not imply migrations rates as high as those needed if populations were much further south. In western Europe the LGM climate was more extreme than in North America, so it is likely that many populations did indeed migrate from southern Europe.

If migration potential represented an important constraint during past climate change, then glacial transitions might be times of extinction. The fossil record does not provide detailed evidence for past extinctions, because fossil pollen resolve most taxa only to family or genus level. Thus, we could detect extinctions only of species that can be resolved in the fossil record. Despite the limitations of the fossil record, several extinctions near the LGM suggest that dispersal limitation during times of rapid warming might be a contributing factor. The rainforest genus *Dacrydium* disappeared from northern Queensland near the LGM when rainforest contracted due to climate change (Kershaw *et al.*, 2000). Fossil cones and needles suggest that *Picea critchfeldii* may have been both abundant and widespread in eastern North America during the LGM, but the fossil evidence disappears at the time of most rapid climate change from 12 to 9 ka. Timing of the event suggests that migration potential might have been a contributing factor (Jackson and Weng, 1999) as may have been the case for *Sequoia*, *Tsuga*, *Carya*, and *Nyssa* during earlier glacial cycles in Europe (West, 1970).

Contemporary Invasions

Because contemporary invasions can be documented, they hold promise for improving our understanding of how dispersal contributes to migration. Most intercontinental invasions begin with human dispersal. Countless examples of non-native invasive species in America have attended globalization of trade, including ornamental plantings (e.g., Brazilian pepper (*Schinus terebinthifolius*), salt cedar (*Tamarix*), velvet leaf (*Miconia calvescens*)), accidental escapes or contaminants (e.g., leafy spurge *Eupohorbia esula*, cogon grass *Imperata cylindrica*, cheat grass *Bromus tectorum*), and erosion control (e.g., mangroves *Rhizophora mangle*, iceplant *Carpobrotus edulis*, kudzu *Pueraria lobata*) (Plant Conservation Alliance, 2000).

Once introduced, aliens have spread at very different rates. The large numbers of recent alien introductions should provide a rich data source concerning the role of dispersal. Unfortunately, the dispersal biology of invading species has seldom been studied in the context of the spatial pattern of spread (Parendes and Jones, 2000), and there are few examples of migrations that can be linked to dispersal data. Most contemporary invasions are known from anecdotal information on time and location of introduction, mode of spread, and impacts on native communities (Mack, 1986). Studies of the invasions themselves tend to focus on the characteristics that make ecosystems prone to invasion (Hobbs and Huenneke, 1992; Lonsdale, 1999) or on characteristics of species that make for successful invaders (Perrins *et al.*, 1992). Several studies that have examined seed dispersal after invasions (Van Wilgen and Siegfried, 1986; Malo and Suarez, 1997) focus on characteristics of seeds rather than on the spatial aspects of spread. While experience with contemporary invasions becomes the basis for assessing invasive potential and for restrictions on the import of exotic species (Ruesink *et al.*, 1995; Ewel *et al.*, 1999), poor understanding of dispersal partly explains our low predictive capability, even *post hoc*.

Several examples where dispersal biology has been reconstructed *post hoc* illustrate the idiosyncrasies of individual invasions. Some species that are absent from suitable habitats, upon introduction, expand rapidly. In 1902, the American Sugar Corporation introduced *Rhizophora mangle*, a Florida mangrove, to Molokai, Hawaii, in an effort to mitigate coastal erosion. The islands had not supported mangrove species prior to this time, presumably due to dispersal limitation. *R mangle* dispersal depends on ocean currents, and those that affect Hawaii flow from the direction of Alaska. However, once introduced, the species has spread by natural drift throughout the archipelago (Wester, 1982). The *R mangle* example illustrates that even a broadly dispersed species can be dispersal-limited, depending on the details of the dispersal process.

High fecundity can interact with novel dispersal vectors to accelerate the spread of alien species. *Purple loosestrife* (*Lythrum salicaria*) came to American shores in the early 1800s in contaminated ship ballast from the European coast (Stuckey, 1980). The plant thrives in marshes, wetlands, and riparian zones, and it spreads rapidly along waterways. The plant produces $>100\,000$ seeds per stem that can disperse short distances by winds and as buoyant seeds and seedling cotyledons in flowing waters and ditches (Balogh, 1986), and along highways in high winds created by trucks (Wilcox, 1989). Once established, purple loosestrife propagates vegetatively and rapidly chokes out native plants. This rapid spread illustrates the importance of high fecundity, the effects of which are amplified by the rare,

extreme dispersal events made possible by water transport and human vectors.

The rate and pattern of spread can be highly variable. Cheatgrass (*Bromus tectorum*) migration throughout the intermountain West represents one of the best-documented alien invasions (Carpenter and Murray, 2000). A reconstruction by Mack (1986) suggests that cheatgrass probably arrived in the 1880s as a seed contaminant from Europe. Before 1900, it was noted from photographs, in collections, and by wheat growers from British Columbia to Utah. An annual with wind dispersed seeds, cheatgrass spread was accelerated by rail transport of livestock, as seeds adhere to fur, and they survive in dung. Following an initial lag phase, cheatgrass range expansion was rapid, but patchy. Centers of abundance spread to several western states in the early 1900s and coalesced over the next few decades to form a relatively continuous range in the Intermountain West. By 1930, cheatgrass had reached its current distribution across the Great Basin. The initial lag has been observed for a number of invasions (e.g., Pitelka *et al.*, 1997). The patchy spread is consistent with fat-tailed dispersal with infilling being more representative of a diffusion process.

The potential to learn from contemporary invasions is unrealized, in part, due to limited study of dispersal. When dispersal is available, it may not be enough to predict rates of spread, because disturbance and species interactions have profound effects on establishment success.

WHAT TO EXPECT: GLOBAL WARMING, MASS INTRODUCTIONS, AND HUMAN DISTURBANCE REGIMES

Dispersal is critical for plant migration, representing the first step in a complex process that involves establishment, growth, and survival. We cannot yet expect to accurately predict many new invasions. Because dispersal is too poorly understood, especially the long-distance events that disproportionately control invasion speed, migrations will be so variable that forecasts will be of limited use for any specific invasion that occurs. Global changes in disturbance, land cover, dispersal vectors, and climate interact with plant migrations, many involving scales of space and time that are not yet well-understood: in addition biotic interactions may play an overriding role in many migrations. These may not be predictable from dispersal potential or from those interactions in their places of origin.

The first two of these concerns deal directly with the relationship between dispersal and migration and are the focus of this review. Too few studies have attempted to quantify dispersal and seed production at necessary spatial and temporal scales (Clark *et al.*, 1999b). Maximum dispersal distances estimated by direct or inverse methods range to several hundred meters, although indirect evidence indicates that propagules sometimes travel much farther. Average dispersal distances are negatively correlated with seed size and positively correlated with fecundity, and fecundity is highly variable. Models lead us to expect erratic migrations with mean rates controlled by extreme dispersal events. High propagule producers have the clear advantage in this process, much more so than would be expected by a process of diffusion. Although dispersal data will allow us to develop estimates of spread potential based on dispersal ability, those estimates will have broad variance. Thus, the rates for any given species in any given environment will deviate from the expectation.

The latter two concerns are not the focus of this review, but they play such a large role in migrations that they will ultimately determine our understanding of population spread. We briefly summarize some of the changes in our biotic and physical environment that will have profound impacts on future plant dispersal and migration.

Interactions with Global Change

Changing chemistry of the atmosphere and climate present new sources of uncertainty for plant migrations. Despite interpretations that past migrations may have occurred rapidly, long distance dispersal may not save many species from the climate change associated with greenhouse warming. The rates of 21st century climate change pose unprecedented challenges to migration, perhaps requiring species ranges to shift at rates of $1-5\,\mathrm{km\,year^{-1}}$ (Davis, 1986; Dyer, 1995; Sykes and Prentice, 1996; King and Herstrom, 1997). Analyses of dispersal consequences for tree migration rates (Clark *et al.*, 2001) suggest that a number of species simply may not disperse far enough. The molecular evidence suggesting the importance of corridors for migration in Europe (Figure 3) raises additional concerns for future migrations.

A changing physical environment is complicated by a shifting biotic one. For predicting migrations, dispersal information from extant ecosystems is not enough. Contemporary invasions have surprised us before, due to unexpectedly high competitive abilities of alien species and novel dispersal vectors. Although we might have been able to quantify dispersal reasonably well in some of these cases, the migration potential often depends on interactions.

Climate change may reorganize food webs in ways that impact migration rates, as some species outrun their mutualists and natural enemies, while others become limited by new biotic settings. Vertebrate dispersal vectors and pollinators may not respond to climate change in step with the plants that would benefit from them (Bethke and Nudds, 1995; Bawa and Dayanandan, 1998; Bazzaz, 1998; Sorenson *et al.*, 1998; Visser *et al.*, 1998; Dunn and Winkler, 1999). For example, droughts that reduce wetlands and the migratory waterfowl that depend on them (Bethke and

Nudds, 1995; Sorenson *et al.*, 1998) will ultimately reduce transport of seed (Ehrlich *et al.*, 1988). Likewise, extreme winter storms reduce passerine densities and their ranges (Mehlman, 1997). Recent global warming in temperate regions has already accelerated some phenological events that affect dispersal. For example, Estonian springs have advanced eight days over the last 80 years, with the last 40 years being particularly warm (Ahas, 1999). In seasonal environments, avian reproduction is successful if timed to food availability. Future increases in spring temperature could result in a mismatch between the timing of egg laying and the availability of food (Visser *et al.*, 1998). Increasing springtime temperatures from 1959 to 1991 have caused North American tree swallows (*Tachycineta bicolor*) to lay eggs nine days earlier (Dunn and Winkler, 1999). The potential decoupling of plant-animal phenologies may be particularly important in the Tropics, where the links among plant reproductive effort, pollination, and vertebrate seed dispersal are strong (Bawa and Dayanandan, 1998; Bazzaz, 1998), and many plant species and their propagules are locally rare (Clark *et al.*, 1999a). If increased temperatures result in increased fire, then not only will fire-adapted species become more abundant (e.g., Starfield and Chapin, 1996), but animal seed dispersers may respond. For example, Johnson *et al.* (1997) have observed that blue jays (*Cyanocitta cristata*) cache more nuts in grasslands following fires.

Given the importance of fecundity for dispersal when kernels are fat-tailed, the effects of changing atmospheric chemistry on seed production could impact on migration potential. Elevated atmospheric carbon dioxide (CO_2) may cause seed production to increase (Farnsworth and Bazzaz, 1995; Schaeppi, 1996; Thomas *et al.*, 1999; LaDeau and Clark, 2001), to decrease (Farnsworth and Bazzaz, 1995; Fischer *et al.*, 1997; Thomas *et al.*, 1999) and/or to occur at an earlier age (Farnsworth *et al.*, 1996; LaDeau and Clark, 2001). Increased nitrogen deposition may have no reproductive impact or it may increase plant fecundity (Gordon *et al.*, 1999), whereas increased tropospheric ozone (O_3) may damage vegetative tissue and reduce reproductive success (Musil *et al.*, 1999; Bergweiler and Manning, 1999). Interactions among climate, CO_2, nitrogen, O_3, and ultraviolet-B radiation could result in surprises.

Land cover change and human disturbance have many indirect consequences for recent migrations in ways that facilitate spread of some and slow spread of others. Disturbance has hastened the spread of many introduced aliens, with temperate agricultural and urban sites being among the most invaded areas (Lonsdale, 1999). In some arid and semiarid lands, land clearance and changes in fire regime have facilitated the expansion of exotic grasses (D'Antonio and Vitousek, 1992). In others, humans have promoted the spread of woody vegetation. For example, 20th century livestock grazing, removal of native Americans, and fire suppression have promoted the spread of shrubs and trees in pygmy woodlands of western North America (Chambers *et al.*, 1999). Damming of the Colorado River reduced water table fluctuations, resulting in replacement of riparian hardwoods by the introduced saltcedar (Vitousek *et al.*, 1996). Eurasian aliens may have flourished in the intermountain West, in part because large, congregating mammals were absent in pre-settlement time. Lack of summer rainfall limited productivity of the warm-season C_4 grasses and, thus, green forage for grazers in summer. The native bunchgrasses did not respond well to introduced livestock and were replaced by Eurasian exotics adapted to both climate and grazers (Mack and Thompson, 1982; Mack, 1986).

On the other hand, human disturbance will hinder spread of many native and exotic species. For many species, land cover change will have the simple consequence of removing much of the area needed for population spread. The alien invader *Lonicera maacki* appears to spread fastest where forest cover is connected; large expanses of agricultural land act as a barrier to dispersal of this species (Hutchinson and Vankat, 1998). Many tree species will experience similar effects. Models incorporating fragmentation often predict delayed invasion due to a simple reduction in available habitat (Higgens and Richardson, 1999). Often the effects are more complex involving alterations in animal vector behavior (Verboom *et al.*, 1991; Pitelka *et al.*, 1997) and presenting novel establishment options, both positive and negative.

The natural dispersal kernels most often studied by ecologists are supplemented by a host of new modes of seed dispersal. Human activities have escalated and spread far beyond those that prevailed in the past. The occurrence of weedy species in charred remains of Neolithic grain stores of central Europe shows that broad, inadvertent dissemination has a history and a prehistory (Behre, 1981; Gluza, 1983). Establishment of towns and trade allowed for introductions as contaminant crop seed, fodder, or as escapes from horticultural plantings (Parker, 1977). Many Eurasian weeds in the western US were introduced as contaminants. Humans disperse propagules in a number of ways, but species producing large numbers of propagules tend to be best represented in two of the most important dispersal pathways in Britain, topsoil transport and cars (Hodkinson and Thompson, 1997). The number of exotic species in nature reserves is correlated with visitation rates (Lonsdale, 1999), which supports the importance of vehicular transport.

Improving Our Understanding

The importance of dispersal for plant migration is sufficiently uncertain that simulations involving future climate change often contrast scenarios of no dispersal with ones

that assume global dispersal (e.g., Sykes and Prentice, 1996; Pitelka *et al.*, 1997). By bracketing reality between extremes, we gain insight into migration potential at scales relevant for some of the most profound environmental challenges to plant populations since the end of the Pleistocene. The simulations that explore these effects demonstrate the importance of dispersal and highlight our inadequate characterization of it.

Understanding plant population spread is a realistic goal that requires realistic expectations. The contribution of dispersal limitation is knowable, but it will require more analysis at the appropriate spatial and temporal scales (Clark *et al.*, 1999b). We know enough now to appreciate its central role in migration potential. We should not expect precise forecasts of spread rate, due to its highly stochastic nature. But we can expect that a better understanding of fecundity and dispersal biology will help us identify species having potential for rapid spread given suitable conditions for growth and survival.

REFERENCES

Ahas, R (1999) Long-term Phyto-, Ornitho- and Ichthyophenological Time-series Analyses in Estonia, *Int. J. Biometeorol.*, **42**, 119–123.

Balogh, G R (1986) Distribution and Seedling Ecology of Purple Loosestrife in Ohio's Lake Erie Marshes, *Ohio J. Sci.*, **86**, 51.

Barnosky, C W, Anderson, P M, and Bartlein, P J (1987) The Northwestern US During Glaciation; Vegetational History and Paleoclimatic Implications, in *North America and Adjacent Oceans During the Last Deglaciation*, eds W F Ruddiman and H E Wright, Jr, Geological Society of America, Boulder, CO, 289–322.

Bawa, K S and Dayanandan, S (1998) Global Climate Change and Tropical Forest Genetic Resources, *Clim. Change*, **39**, 473–485.

Bazzaz, F A (1998) Tropical Forests in a Future Climate: Changes in Biological Diversity and Impact on the Global Carbon Cycle, *Clim. Change*, **39**, 317–336.

Bazzaz, F A (1991) Habitat Selection in Plants, *Am. Nat.*, **137**, supp., 116–130.

Behre, K E (1981) The Interpretation of Anthropogenic Indicators in Pollen Diagrams, *Pollen et Spores*, **23**, 225–245.

Bennett, K D (1985) The Spread of *Fagus grandifolia* Across Eastern North America during the Last 18 000 Years, *J. Biogeogr.*, **12**, 147–164.

Bergweiler, C J and Manning, W J (1999) Inhibition of Flowering and Reproductive Success in Spreading Dogbane (*Apocynum androsaemifolium*) by Exposure to Ambient Ozone, *Environ. Pollut.*, **105**, 333–339.

Bethke, R W and Nudds, T D (1995) Effects of Climate Change and Land Use on Duck Abundance in Canadian Prairie-Parklands, *Ecol. Appl.*, **5**, 588–600.

Cain, M L, Hans, D, and Muir, A (1998) Seed Dispersal and the Holocene Migration of Woodland Herbs, *Ecol. Monogr.*, **68**, 325–347.

Carpenter, A T and Murray, T A (2000) Element Stewardship Abstract for *Bromus tectorum*, The Nature Conservancy. http://tncweeds.ucdavis.edu.

Chambers, J C, Vanderwall, S B, and Schupp, E (1999) Seed and Seedling Ecology of Pin∼on and Juniper Species on the Pygmy Woodlands of Western North America, *Bot. Rev.*, **65**, 1–38.

Clark, J S (1998) Why Trees Migrate so Fast: Confronting Theory with Dispersal Biology and the Paleo Record, *Am. Nat.*, **152**, 204–224.

Clark, J S, Macklin, E, and Wood, L (1998) Stages and Spatial Scales of Recruitment Limitation in Southern Appalachian Forests, *Ecol. Monogr.*, **68**, 213–235.

Clark, J S, Fastie, C, Hurtt, G, Jackson, S T, Johnson, C, King, G, Lewis, M, Lynch, J, Pacala, S, Prentice, I C, Schupp, E W, Webb, III, T, and Wyckoff, P (1998) Reid's Paradox of Rapid Plant Migration, *BioScience*, **48**, 13–24.

Clark, J S, Silman, M, Kern, R, Macklin, E, and HilleRisLambers, J (1999a) Seed Dispersal Near and Far: Generalized Patterns Across Temperate and Tropical Forests, *Ecology*, **80**, 1475–1494.

Clark, J S, Beckage, B, Camill, P, Cleveland, B, HilleRisLambers, J, Lichter, J, MacLachlan, J, Mohan, J, and Wyckoff, P (1999b) Interpreting Recruitment Limitation in Forests, *Am. J. Bot.*, **86**, 1–16.

Clark, J S, Lewis, M, and Horvath, L (2001a) Invasion by Extremes: Variation in Dispersal and Reproduction Retards Population Spread, *Am. Nat.*, **157**, 537–554.

Clark, J S, Horvath, L, and Lewis, M (2001b) On the Estimation of Spread for a Biological Population, *Stat. Probability Lett.*, **51**, 225–234.

Clark, J S, Lewis, M, and Horvath, L (2001) Invasion by Extremes: Variation in Dispersal and Reproduction Retards Population Spread, *Am. Nat.*, in press.

Craddock, C L and Huenneke, L F (1997) Aquatic Seed Dispersal and its Implications in *Cirsium vinaceum*, a Threatened Endemic Thistle of New Mexico, *Am. Midl. Nat.*, **138**, 215–219.

Daehler, C C and Strong, D R (1996) Status, Prediction and Prevention of Introduced Cordgrass *Spartina* spp Invasions in Pacific Estuaries, USA, *Biol. Conserv.*, **78**, 51–58.

Dalby, D H (1963) Seed Dispersal in *Salicornia pusilla*, *Nature*, **199**, 197–198.

D'Antonio, C M and Vitousek, P M (1992) Biological Invasions by Exotic Grasses, the Grass/Fire Cycle and Global Change, *Ann. Rev. Ecol. Syst.*, **23**, 63–87.

Davis, M B (1986) Climatic Instability, Time Lags, and Community Disequilibrium, in *Community Ecology*, eds J Diamond and T J Case, Harper and Row, NY, 269–284.

Davis, M B (1987) Invasion of Forest Communities During the Holocene: Beech and Hemlock in the Great Lakes Region, in *Colonization, Succession, and Stability*, eds A J Gray, M J Crawley, and P J Edwards, Blackwell Scientific, Oxford, 373–393.

Davis, Jr, J H (1940) The Ecology and Geologic Role of Mangroves in Florida, *Carnegie Inst. Wash. Publ.*, (papers from Tortugas Lab. 32) **517**, 303–412.

Dunn, P O and Winkler, D W (1999) Climate Change has Affec ted the Breeding Date of Tree Swallows Throughout North America, *Proc. R. Soc. Lond.*, **266B**, 2487–2490.

Dyer, J M (1995) Assessment of Climatic Warming Using a Model of Forest Species Migration, *Ecol. Model.*, **79**(1–3), 199–219.

Ehrlich, P R, Dobkin, D S, and Wheye, D (1988) *The Birder's Handbook*, Simon and Schuster, New York.

Ernst, A (1908) *The New Flora of the Volcanic Islands of Krakatau*, Cambridge University Press, Cambridge.

Ewel, J J, O'Dowd, D J, Bergelson, J, Daehler, C C, and D'Antonio, C M (1999) Deliberate Introductions of Species: Research Needs, *Bioscience*, **49**, 619–630.

Farnsworth, E J and Bazzaz, F A (1995) Inter- and Intra-generic Differences in Growth, Reproduction, and Fitness of Nine Herbaceous Annual Species Grown in Elevated CO_2 Environments, *Oecologia*, **104**, 454–466.

Farnsworth, E J, Ellison, A M, and Gong, W K (1996) Elevated CO_2 Alters Anatomy, Physiology, Growth and Reproduction of Red Mangrove (*Rhizophora mangle* L.), *Oecologia*, **108**, 599–609.

Fastie, C L (1995) Causes and Ecosystem Consequences of Multiple Pathways of Primary Succession at Glacier Bay, Alaska, *Ecology*, **76**, 1899–1916.

Fenner, M (1985) *Seed Ecology*, Chapman and Hall, London.

Firbas, F (1949) *Spat- und nacheiszeitliche Waldgeschichte Mitteleuropas nordlich der Alpen*, Fischer, Jena, Germany.

Fischer, M, Matthies, D, and Schmid, B (1997) Responses of Rare Calcareous Grassland Plants to Elevated CO_2: a Field Experiment with *Gentianella germanica* and *Gentiana cruciata*, *J. Ecol.*, **85**, 681–691.

Forget, P M (1990) Seed-dispersal of *Voucapoua amricana* (Caesalpiniaceae) by Caviomorph Rodents in French Guiana, *J. Trop. Ecol.*, **6**, 459–468.

Forget, P M (1991) Comparative Recruitment Patterns of Two Non-pioneer Canopy Tree Species in French Guiana, *Oecologia*, **85**, 434–439.

Gluza, I (1983) Neolithic Cereals and Weeds from the Locality of the Langyel Culture at Nowa Huta-Mogila near Cracow, *Acta Palaeobotanica*, **23**, 123–184.

Gladstone, D E (1979) Description of a Seed-shadow of a Wind-dispersed Tropical Tree, *Brenesia*, **16**, 81–86.

Goulding, M (1980) *The Fishes and the Forest: Explorations in Amazonian Natural History*, University of California Press, Berkeley, CA.

Guppy, H B (1917) *Plants, Seeds and Currents in the West Indies and the Azores*, MacMillan, London.

Harper, J L (1977) *Population Biology of Plants*, Academic Press, London.

Herrera, C M, Jordano, P, López-Soria, L, and Amat, J A (1994) Recruitment of Mast-fruiting, Bird-dispersed Tree: Bridging Frugivore Activity and Seedling Establishment, *Ecol. Monogr.*, **64**, 315–344.

Hewitt, G M (1999) Post-glacial Re-colonization of European Biota, *Biol. J. Linnean Soc.*, **68**, 87–112.

Higgins, S I and Richardson, D M (1999) Predicting Plant Migration Rates in a Changing World: the Role of Long-distance Dispersal, *Am. Nat.*, **153**, 464–475.

Hobbs, R J and Huenneke, L F (1992) Disturbance, Diversity, and Invasion: Implications for Conservation, *Conserv. Biol.*, **6**, 324–337.

Hooghiemstra, H (1984) *Vegetational and Climatic History of the High Plain of Bogota, Colombia: a Continuous Record of the Last 3.5 Million Years*, Cramer, Vaduz, Germany.

Hopkins, M S, Graham, A W, Hewett, R, Ash, J, and Head, J (1990) Evidence of Late Pleistocene Fires and Eucalypt Forest from a North-Queensland Humid Tropical Rainforest Site, *Aust. J. Ecol.*, **15**, 345–347.

Hoshizaki, K, Suzuki, W, and Sasaki, S (1997) Impacts of Secondary Seed Dispersal and Herbivory on Seedling Survival in *Aesculus turbinate*, *J. Veg. Sci.*, **8**, 735–742.

Houle, G (1998) Seed Dispersal and Seedling Recruitment of *Betula alleghaniensis*: Spatial Inconsistency in Time, *Ecology*, **79**, 807–818.

Howe, H F and Westley, L C (1997) Ecology of Pollination and Seed Dispersal, in *Plant Ecology*, ed M J Crawley, Blackwell Science, Oxford.

Huiskes, A H, Koutsttl, B P, Herman, P M, Beeftink, W G, Mar kusse, M M, and De Munk, W (1995) Seed Dispersal of Halophytes in Tidal Salt Marshes, *J. Ecol.*, **83**, 559–567.

Hughes, L and Westoby, M (1992) Fate of Seeds Adapted for Dispersal by Ants in Australian Sclerophyll Vegetation, *Ecology*, **73**(4), 1285–1299.

Hulme, P E (1997) Post-dispersal Seed Predation and the Establishment of Vertebrate Dispersed Plants in Mediterranean Shrublands, *Oecologia*, **111**, 91–98.

Huntley, B and Birks, H J (1983) *An Atlas of Past and Present Pollen Maps for Europe 0–13 000 Years Ago*, Cambridge University Press, Cambridge.

Hutchinson, T F and Vankat, J L (1998) Landscape Structure and Spread of the Exotic Shrub *Lonicera maackii* in Southwestern Ohio Forests, *Am. Midl. Nat.*, **139**, 383–390.

Ibrahim, K M, Nichols, R A, and Hewitt, G M (1996) Spatial Patterns of Genetic Variation Generated by Different Forms of Dispersal During Range Expansion, *Heredity*, **77**, 282–291.

Jackson, S T and Weng, C (1999) Late Quaternary Extinction of a Tree Species in Eastern North America, *Proc. Nat. Acad. Sci.*, **96**, 13 847–13 852.

Janzen, D H (1971) *Cassia grandis* L. Beans and their Escape from Predators: a Study in Tropical Predator Satiation, *Ecology*, **52**, 964–979.

Janzen, D H and Martin, P S (1982) Neotropical Anachronisms: the Fruits the Gomphotheres Ate, *Science*, **215**, 19–27.

Johnson, W C and Adkisson, C S (1985) Dispersal of Beech Nuts by Blue Jays in Fragmented Landscapes, *Am. Midl. Nat.*, **113**, 319–324.

Johnson, W C, Adkisson, C S, Crow, T R, and Dixon, M D (1997) Nut Caching by Blue Jays (*Cyanocitta cristata* L.): Implications for Tree Demography, *Am. Midl. Nat.*, **138**, 357–370.

Kershaw, A P (1993) The Development of Rainforest–Savanna Boundaries in Tropical Australia, in *Nature and Dynamics of Forest Savanna Boundaries*, eds P A Farley, J Proctors, and J A Ratter, Chapman and Hall, London, 255–271.

Kershaw, A P, Clark, J S, and Gill, A M (2000) A History of Fire in Australia, in *Flammable Australia: the Fire Regimes of a Continent*, eds R Bradstock and A M Gill, in press.

King, G A and Herstrom, A A (1997) Holocene Tree Migration Rates Objectively Determined from Fossil Pollen Data, in *Past and Future Rapid Environmental Changes: the Spatial and*

Evolutionary Responses of Terrestrial Biota, eds B Huntley, W Cramer, A V Morgan, H C Prentice, and J R Allen, Springer-Verlag, New York, 91–101.

Kollmann, J and Pirl, M (1995) Spatial Pattern of Seed Rain of Fleshy-Fruited Plants in a Scrubland–Grassland Transition, *Acta Oecol.*, **16**, 313–329.

Kollmann, J and Schill, H P (1996) Spatial Patterns of Dispersal, Seed Predation and Germination during Colonization of Abandoned Grassland by *Quercus petraea* and *Corylus avellana*, *Vegetatio*, **125**, 193–205.

Kolmogorov, A, Petrovsky, I, and Piscounov, N (1937) Étude de l'équation de la Diffusion avec Croissance de la Quantité de Matière et son Application à un Problèma Biologigue, *Moscow Univ. Bull. Ser. Internat. Sect. A*, **1**, 1–25.

König, D (1960) Beiträge zur kenntnis der deutschen Salicornien. Mitteilungendel Florstisch-Soziologische Anbeitsgemeinschaft, *Neue Falge*, **8**, 5–58.

Kot, M, Lewis, M A, and van den Driessche, P (1996) Dispersal Data and the Spread of Invading Organisms, *Ecology*, **77**, 2027–2042.

Kubitzki, K and Ziburski, A (1994) Seed Dispersal in Flood Plain Forest of Amazonia, *Biotropica*, **26**, 30–43.

Kullman, L (1996) Norway spruce present in the Scandes Mountains, Sweden at 8000 yr BP: New Light on Holocene Tree Spread, *Global Ecol. Biogeogr. Lett.*, **5**, 94–101.

Le Corre, V, Machon, N, Petit, R J, and Kremer, A (1997) Colonization With Long-distance Seed Dispersal and Genetic Structure of Maternally Inherited Genes in Forest Trees: a Simulation Study, *Genet. Res. Cambridge*, **69**, 117–125.

LaDeau, S and Clark, J S (2001) Rising CO_2 and the fecundity of forest trees, *Science*, **292**, 95–98.

Lewis, M A (1997) Variability, Patchiness, and Jump Dispersal in the Spread of an Invading Population, in *Spatial Ecology*, eds D Tilman and P Kareiva, Princeton University Press, Princeton, NJ, 46–69.

Lonsdale, W M (1999) Global Patterns of Plant Invasions and the Concept of Invisibility, *Ecology*, **80**, 1522–1536.

Mack, R N (1986) Alien Plant Invasion into the Intermountain West: a Case History, in *Ecology of Biological Invasions of North America and Hawaii*, eds H A Mooney and J A Drake, Springer-Verlag, New York, 191–213.

Malo, J E and Suarez, F (1997) Dispersal Mechanism and Transcontinental Naturalization Proneness Among Mediterranean Herbaceous Species, *J. Biogeogr.*, **24**, 391–394.

Mehlman, D W (1997) Change in Avian Abundance Across the Geographic Range in Response to Environmental Change, *Ecol. Appl.*, **7**, 614–624.

Moll, D and Jansen, K P (1995) Evidence For a Role in Seed Dispersal by Two Tropical Herbivorous Turtles, *Biotropica*, **27**, 121–127.

Murray, D (1986) *Seed Dispersal*, Academic Press, Sydney.

Murray, K G (1988) Avian Seed Dispersal of Three Neotropical Gap-Dependent Plants, *Ecol. Monogr.*, **58**, 271–298.

Musil, C F, Rutherford, M C, Powrie, L W, Bjorn, L O, and McDonald, D J (1999) Spatial and Temporal Changes in South African Solar Ultraviolet-B Exposure: Implications for Threatened Taxa, *Ambio*, **28**, 450–456.

Nogales, M, Delgado, J D, and Medina, F M (1998) Shrikes, Lizards and *Lycium intricatum* (Solanaceae) Fruits: a Case of Indirect Seed Dispersal on an Oceanic Island (Alegranza, Canary Islands), *J. Ecol.*, **86**, 866–871.

Nogales, M, Hernández, E, and Valdés, F (1999) Seed Dispersal by Common Ravens *Corvus cenax* Among Island Habitats (Canarian Archipelago), *Ecoscience*, **6**, 51–61.

Parendes, L A and Jones, J A (2000) Role of Light Availability and Dispersal in Exotic Plant Invasion Along Roads and Streams in the HJ Andrews Experimental Forest, Oregon, *Conserv. Biol.*, **14**, 64–75.

Parker, C (1977) Prediction of New Weed Problems, Especially in the Developing World, in *Origins of Pest, Parasite, Disease, and Weed Problems*, eds J M Cherett and G R Sagar, Blackwell Scientific, Oxford, 249–264.

Perrins, J, Williamson, M, and Fitter, A (1992) A Survey of Differing Views of Weed Classification: Implications for Regulation of Introductions, *Biol. Conserv.*, **60**, 47–56.

Petit, R J, Pineau, E, Demesure, B, Bacilieri, R, Ducousso, A, and Kremer, A (1997) Chloroplast DNA Footprints of Postglacial Recolonization by Oaks, *Proc. Nat. Acad. Sci.*, **94**, 9996–10 001.

Pimentel, D, Lach, L, Zuniga, R, and Morrison, D (2000) Environmental and Economic Costs of Nonindigenous Species in the United States, *BioScience*, **50**, 53–65.

Pitelka, L F, Ash, J, Berry, S, Bradshaw, R H, Brubaker, L, Clark, J S, Davis, M B, Dyer, J M, Gardner, R H, Gitay, H, Hope, G, Hengeveld, R, Huntley, B, King, G A, Lavorel, S, Mack, R N, Malanson, G P, McGlone, M, Noble, I R, Prentice, I C, Rejmanek, M, Saunders, A, Solomon, A M, Sugita, S, and Sykes, M T (1997) Plant Migration and Climate Change, *Am. Sci.*, **85**, 464–473.

Plant Conservation Alliance (2000) Weeds Gone Wild: Alien Plant Invaders of Natural Areas, www.nps.gov/plants/alien.

Porter, D M (1976) Geography and Dispersal of Galapagos Island Plants, *Nature*, **264**, 745–746.

Portnoy, S and Willson, M F (1993) Seed Dispersal Curves: Behavior of the Tail of the Distribution, *Evol. Biol.*, **7**, 25–44.

Rabinowitz, D (1978) Dispersal Properties of Mangrove Propagules, *Biotropica*, **10**, 47–57.

Reid, C (1899) *The Origin of the British Flora*, Dulau, London.

Ribbens, E, Silander, J A, and Pacala, S W (1994) Seedling Recruitment in Forests: Calibrating Models to Predict Patterns of Tree Seedling Dispersion, *Ecology*, **75**, 1794–1806.

Ridley, H N (1930) *The Dispersal of Plants Throughout the World*, Reeve and Co., Ashford.

Ritchie, J C and MacDonald, G M (1986) The Patterns of Postglacial Spread of White Spruce, *J. Biogeogr.*, **13**, 527–540.

Ruesink, J L, Parker, I M, Groom, M J, and Kareiva, P M (1995) Reducing the Risks of Nonindigenous Species Introductions: Guilty Until Proven Innocent, *Bioscience*, **45**, 465–477.

Russell, E W and Sanford, S D (2000) Late-Glacial Environmental Changes South of the Wisconsinan Terminal Moraine in the Eastern United States, *Quat. Res.*, **53**, 105–113.

Sauer, J D (1988) *Plant Migration*, University of California Press, Los Angeles, CA.

Schaeppi, B (1996) Growth Dynamics and Population Development in an Alpine Grassland Under Elevated CO_2, *Oecologia*, **106**, 93–99.

Schimper, A F (1891) *Die Indo-Malaïsche Strandflora*, Fischer, Jena.

Schneider, R L and Sharitz, R R (1988) Hydrochory and Regeneration in a Bold Cypress-water Tupelo Swamp Forest, *Ecology*, **69**, 1055–1063.

Schupp, E W (1993) Quantity, Quality and the Effectiveness of Seed Dispersal by Animals, *Vegetatio*, **107/108**, 15–29.

Shilton, L A, Altringham, J D, Compton, S G, and Whittaker, R S (1999) Old World Fruit Bats can be Long-Distance Seed Dispersers Through Extended Retention of Viable Seeds in the Gut, *Proc. R. Soc. Lond. Series B: Biol. Sci.*, **2661**, 219–223.

Skellam, J G (1951) Random Dispersal in Theoretical Populations, *Biometrika*, **38**, 196–218.

Snow, J T, Wyatt, A L, McCarthy, A K, and Bishop, E K (1995) Fallout of Debris from Tornadic Thunderstorms: a Historical Perspective and Two Examples from VORTEX, *Bull. Am. Meteorol. Soc.*, **76**, 1777–1790.

Sorenson, L G, Goldberg, R, Root, T L, and Anderson, M G (1998) Potential Effects of Global Warming on Waterfowl Populations Breeding in the Northern Great Plains, *Clim. Change*, **40**, 343–369.

Starfield, A M and Chapin, III, S F (1996) Model of Transient Changes in Arctic and Boreal Vegetation in Response to Climate and Land use Change, *Ecol. Appl.*, **6**, 842–864.

Steele, M and Smallwood, P (1994) What are the Squirrels Hiding? *Natural History*, **10**, 40–45.

Storm, G L and Montgomery, G G (1975) Dispersal and Social Contact Among Red Foxes: Results from Telemetry and Computer Simulation, in *The Wild Canids: their Systematics, Behavioral Ecology, and Evolution*, ed M W Fox, Van Nostrand Reinhold, New York, 237–246.

Street-Perrott, F A, Huang, Y, Perrott, R A, Eglinton, G, Barker, P, Khelifa, L B, Harkness, D D, and Olago, D O (1997) Impact of Lower Atmospheric Carbon Dioxide on Tropical Mountain Ecosystems, *Science*, **278**, 1422–1426.

Stuckey, R L (1980) Distributional History of *Lythrum salicaria* (Purple Loosestrife) in North America, *Bartonia*, **47**, 3–20.

Sykes, M T and Prentice, I C (1996) Climate Change, Tree Species Distributions and Forest Dynamics: A Case Study in the Mixed Conifers/Northern Hardwoods Zone of Northern Europe, *Clim. Change*, **34**, 161–177.

Tamura, N and Shibasaki, E (1996) Fate of Walnut Seeds, *Juglans airanthifolia*, Hoarded by Japanese Squirrels, Sciurus lis, *J. For. Res.*, **1**, 219–222.

Thebaud, C and Debussche, M (1991) Rapid invasion of *Fraxinus ornus* L along the Herault River System in Southern France – the Importance of Seed Dispersal by Water, *J. Biogeogr.*, **18**, 7–12.

Thomas, S C, Jasienski, M, and Bazzaz, F A (1999) Early Vs. Asymptotic Growth Responses of Herbaceous Plants to Elevated CO_2, *Ecology*, **80**, 1552–1567.

Tzedakis, P C (1993) Long-term Tree Populations in Northwest Greece Through Multiple Quaternary Climatic Cycles, *Nature*, **364**, 437–440.

Vander Wall, S B (1994) Seed Fate Pathways of Antelope Bitterbrush: Dispersal by Seed-caching Yellow Pine Chipmunks, *Ecology*, **75**, 1911–1926.

Vanderwall, S B and Balda, R B (1977) Coadaptations of the Clark's Nutcracker and the Pinon Pine for Efficient Seed Harvest and Dispersal, *Ecol. Monogr.*, **47**, 89–111.

Van Wilgen, B W and Siegfried, W R (1986) Seed Dispersal Properties of Three Pine Species as a Determinant of Invasive Potential, *S. African J. Bot.*, **52**, 546–548.

Verboom, J, Lankester, K, and Metz, J A (1991) Linking Local and Regional Dynamics in Stochastic Metapopulation Models, in *Metapopulation Dynamics: Empirical and Theoretical Investigations*, eds M Gilpin and I Hanski, Academic Press, London, 39–55.

Visser, M E, Van Noordwij, A J, Tinbergen, J M, and Lessells, C M (1998) Warmer Springs Lead to Mistimed Reproduction in Great Tits (*Parus major*), *Proc. R. Soc. Lond.*, **65**, 1867–1870.

Vitousek, P M, D'Antonio, C M, Loope, L L, and Westbrooks, R (1996) Biological Invasions as Global Environmental Change, *Am. Sci.*, **84**, 468–478.

Weaver, J E and Clements, F E (1938) *Plant Ecology*, McGraw-Hill, New York.

Webb, S L (1987) Beech Range Extension and Vegetation History: Pollen Stratigraphy of Two Wisconsin Lakes, *Ecology*, **68**, 1993–2005.

West, R G (1970) Pleistocene History of the British Flora, in *Studies in the Vegetational History of the British Isles*, eds D Walker and R G West, Cambridge University Press, Cambridge, 1–12.

Wester, L (1982) Introduction and Spread of Mangroves in the Hawaiian Islands, *Assoc. Pacific Coast Geographers Yearbook*, **43**, 125–137.

Whitlow, S and Zielinski, G A (1997) The Holocene-Younger Dryas Transition Recorded at Summit, Greenland, *Science*, **278**, 825–827.

Wilcox, D A (1989) Migration and Control of Purple Loosestrife Along Highway Corridors, *Environ. Manage.*, **13**, 365–70.

Wilkinson, D M (1997) Plant Colonization: are Wind Dispersed Seeds Dispersed by Birds at Larger Spatial and Temporal Scales? *J. Biogeogr.*, **24**, 61–65.

Woods, K D and Davis, M B (1989) Paleoecology of Range Limits: Beech in the Upper Peninsula of Michigan, *Ecology*, **70**, 681–696.

Yasuda, M, Miura, S, and Hussein, N A (2000) Evidence for Food Hoarding Behaviour in Terrestrial Rodents in Pasoh Forest Reserve, a Malaysian Lowland Rain Forest, *J. Trop. For. Sci.*, **12**, 164–173.

Yumoto, T, Kimura, K, and Nishimura, A (1999) Estimation of Retention Times and Distances of Seed Dispersal by Two Monkey Species, *Aouatta seniculus* and *Legothrix lagotricha*, in a Columbia Forest, *Ecol. Res.*, **12**, 179–191.

Plants – from Cells to Ecosystems: Impacts of Global Environmental Change

F A BAZZAZ AND S CATOVSKY

Harvard University, Cambridge, MA, USA

Human activities are leading to dramatic changes in the physical and chemical structure of the Earth's atmosphere, which will in turn, affect the functioning of the biosphere. As a result of complex feedbacks between atmospheric processes and natural ecosystems, an integrated perspective on future global dynamics should explicitly consider such bidirectional interactions. Many novel human-induced perturbations involve changes in resources and environmental conditions that directly affect basic plant processes. As the behavior of individual plants ultimately determines many key ecosystem characteristics, it is critical that we examine how major environmental changes will in uence plant function. In this review, we discuss how one of the most immediate and certain global environmental changes, rising atmospheric carbon dioxide (CO_2), will affect plant physiological processes. We focus particularly on processes that directly feedback to in uence atmospheric conditions.

Elevated CO_2 is a primary substrate for photosynthesis, and has many far-reaching effects on plant function. Carbon uptake rates for plants exposed to increased CO_2 concentrations invariably increase, although the degree of enhancement varies between plants with different photosynthetic pathways. This effect is often accompanied by a suite of other leaf-level changes that may modify the exchange of carbon and water between plants and the atmosphere, e.g., respiratory carbon losses, reduced transpiration, and increased water use ef ciency. The degree of photosynthetic enhancement is variable, and is commonly correlated with the extent to which photosynthesis acclimates to rising CO_2 levels. Excess carbohydrates that accumulate in leaf mesophyll cells inhibit manufacturing of the small subunit of ribulose-1,5-bisphosphate carboxylase (Rubisco), a major photosynthetic enzyme, and thus lead to reduced photosynthetic capacity. Despite a good physiological understanding of controls on photosynthesis, we have been unable to relate these mechanisms to the large variation in CO_2 responsiveness observed across species and environmental conditions.

Leaf-level activities do not take place in isolation, but rather depend on processes in other parts of the plant. A full understanding of the effects of elevated CO_2 on plant function should incorporate physiological processes that take place at the whole-plant level. As with photosynthesis, growth responses are generally enhanced by elevated CO_2. The degree of enhancement, however, can be very variable, showing dependence on both species identity and environmental growth conditions. Other whole-plant traits may also be modi ed under elevated CO_2, e.g., development, architecture, timing of growth events, but general trends in these responses are not yet well established. Of particular interest are the CO_2-induced changes in plant traits that feedback to in uence ecosystem-level nutrient availability. Understanding the responses of this suite of interacting traits is critical for our ability to predict long-term ecosystem responses to elevated CO_2, and thus to ascertain the nature of feedback between ecosystem function and atmospheric processes. Studies rarely focus on simultaneous changes in many potential positive and negative feedbacks. Thus, there is still uncertainty on whether processes leading to increased or decreased nutrient availability will predominate. For example, increased foliar carbon/nitrogen (C/N) ratios will act to slow decomposition rates, and thus represent a negative feedback to CO_2 enhancements of productivity. Increased root biomass, increased mycorrhizal infection, and increased soil microbial activity (from increased loss of carbon from roots into the soil) will likely increase nutrient availability to plants, and thus act as a positive feedback. The relative importance of these processes in different systems and under different environmental conditions must now be established.

PLANTS AND THE GLOBAL CARBON CYCLE

Plants represent the primary pathway for the transfer of carbon from the atmosphere into the terrestrial biosphere. By utilizing light energy from the sun, plants are able to reduce atmospheric carbon dioxide (CO_2) into energy-rich carbohydrates and other metabolic by-products. This process forms the basis for all life on Earth, but also represents an important feedback on the global carbon cycle. Interactions between plant processes and global biogeochemical cycles are of particular concern, because anthropogenic perturbations are causing extensive modification of natural ecosystems. Such changes have heightened interest in understanding current and future controls on biosphere activity (Mooney, 1996). Terrestrial ecosystems have the potential to slow future increases in atmospheric CO_2 (Woodwell *et al.*, 1998), but inevitable changes in the structure and function of such ecosystems make it difficult to predict their future role in the global carbon cycle. Increases in the magnitude of seasonal changes in global CO_2 concentrations highlight both the importance of terrestrial vegetation in regulating atmospheric CO_2 and dynamic changes in the functioning of the biosphere (Keeling *et al.*, 1996).

Realization of the tight coupling between human activities and large-scale biogeochemical processes has led to a large increase in studies focusing on the impacts of global environmental changes on natural ecosystems (Koch and Mooney, 1996; Walker *et al.*, 1999). There are close connections between anthropogenic perturbations and ecological processes, as many global changes directly in uence environmental conditions and availability of resources essential to basic plant functioning. The best-documented recent global environmental change is the annual increase in atmospheric CO_2 concentrations (Vitousek, 1994). Measurements of CO_2 levels from air bubbles trapped in ice cores and from direct atmospheric sampling since 1958 clearly demonstrate that atmospheric CO_2 concentrations have increased by almost 30% in the past 100 years (from $280\,\mu l\,l^{-1}$ at the beginning of the industrial revolution to current concentrations of $360\,\mu l\,l^{-1}$) (Schimel *et al.*, 1996). CO_2 is a primary substrate for photosynthesis, and thus changes in atmospheric CO_2 concentrations are likely to have profound impacts on plant function.

CO_2 is also a radiatively active gas that plays an important role in global climate patterns. As a result, elevated CO_2 levels have been predicted to cause increases in the mean and variability of global temperatures, and alterations to global rainfall patterns (Houghton *et al.*, 1996). The complexity of the global climate system, however, increases the uncertainty in such predictions. Nevertheless, recent work has found that the correlation between CO_2 levels and predicted temperature perturbations has increased through this century, suggesting that increasing CO_2 concentrations are having an increasingly important role in climate patterns (Wigley *et al.*, 1998). Many plant processes are sensitive to external temperatures, and so these predicted climate changes might significantly alter plant performance. Changes in rainfall patterns, along with increased atmospheric deposition of pollutants (e.g., sulfur, nitrogen) (Vitousek *et al.*, 1997) may also affect plant behavior within ecosystems, as water and nutrients are both essential requirements for basic plant function.

The multiple, simultaneous environmental perturbations currently facing natural ecosystems make it challenging to predict the state of the biosphere in the future. We believe that, rather than becoming overwhelmed with the complexity of the situation, we can make a valuable contribution to the development of our predictive capabilities by addressing tractable aspects of global change in turn. In this paper, we primarily focus on the impacts of elevated CO_2 on plants, as increases in atmospheric CO_2 concentrations are one of the most direct, immediate and certain changes that we face in the future. We consider direct effects of elevated CO_2 on plant function, highlighting the current state-of-knowledge on the subject. We address how our understanding of CO_2 effects on plants has developed since some recent, prominent reviews on the subject (Eamus and Jarvis, 1989; Bazzaz, 1990; Woodward *et al.*, 1991; Ceulemans and Mousseau, 1994; Amthor, 1995). We discuss effects of CO_2 on both leaf-level and whole-plant physiology, emphasizing the importance of understanding down-regulation of photosynthesis and feedback to the carbon cycle for a more complete view of potential impacts of elevated CO_2 on natural ecosystems in the future. Towards the end of the paper, we discuss how other anthropogenic perturbations may interact with elevated CO_2 to in uence plant performance in the future.

ELEVATED CO_2 AND LEAF-LEVEL PROCESSES

Leaves are the primary point of exchange of energy and matter between plants and the atmosphere, and thus changes in atmospheric CO_2 concentrations are likely to affect leaf-level processes most directly. Elevated CO_2 may in uence both the rate of CO_2 uptake (photosynthesis and respiration), as well as the rate of water exchange and leaf energy balance through effects on stomatal aperture.

Photosynthesis

From a basic understanding of the physiology of photosynthesis, we can make predictions about how increasing CO_2 levels might in uence carbon uptake rates (Pearcy and Bjorkman, 1983). CO_2 plays a number of critical roles in the process of carbon uptake.

1. CO_2 is a primary substrate for Rubisco. At low CO_2 concentrations, the enzyme ribulose-1,5-bisphosphate

(RuBP) is in excess and carbon fixation depends almost completely on CO_2 concentration. As CO_2 concentration increases, however, other processes become limiting, and carbon fixation becomes relatively insensitive to increases in CO_2 concentration. Initially, light-dependent reactions are unable to produce enough adenosine tri-phosphate (ATP) and the reduced form of nicotinamide adenine dinucleotide phosphate (NADPH) for RuBP regeneration. This process may often be quite transient, as substantial Rubisco deactivation follows shortly after this situation is reached (Sage *et al.*, 1988). As CO_2 levels increase further, carbon fixation becomes limited by the capacity of cytosolic enzymes to utilize triose phosphates for sucrose and starch synthesis (Sharkey, 1985).

2. CO_2 competes with O_2 for active sites on Rubisco. Carbon fixation is a relatively inefficient process, as Rubisco may bind O_2 and catalyze photorespiratory reactions that lead to the net release of CO_2 (Lorimer, 1981). This competitive inhibition depends on substrate concentrations, and is reduced as CO_2 concentrations increase (Sharkey, 1988).
3. CO_2 is required for Rubisco activation. Along with Mg^{2+} ions, additional CO_2 is required for photosynthesis to activate Rubisco (Portis, 1995).

These considerations highlight the dependence of photosynthetic activity on CO_2 concentration and suggest that increasing CO_2 above 350 $\mu l\, l^{-1}$ will increase carbon uptake rates until CO_2 ceases to limit photosynthesis. This point varies between species with different photosynthetic pathways (Figure 1). The *Michaelis rate constant* (K_m) for CO_2 fixation is much lower in C_4 plants than in C_3 plants, as C_4 plants have a CO_2 concentrating mechanism. Thus, for C_4 plants, photosynthesis saturates at a CO_2 concentration slightly above current atmospheric levels, while C_3 photosynthesis continues to increase until much higher CO_2 concentrations (Figure 1). Empirical evidence strongly supports these basic predictions (Carlson and Bazzaz, 1980; Bazzaz, 1990; Curtis, 1996) at least in the short-term (see the section on Down-Regulation of Photosynthesis). Photosynthesis of C_3 plants almost invariably increases in elevated CO_2 to some degree, while that of C_4 plants rarely shows much enhancement (Poorter *et al.*, 1996) (*see* C_3 and C_4 Photosynthesis, Volume 2).

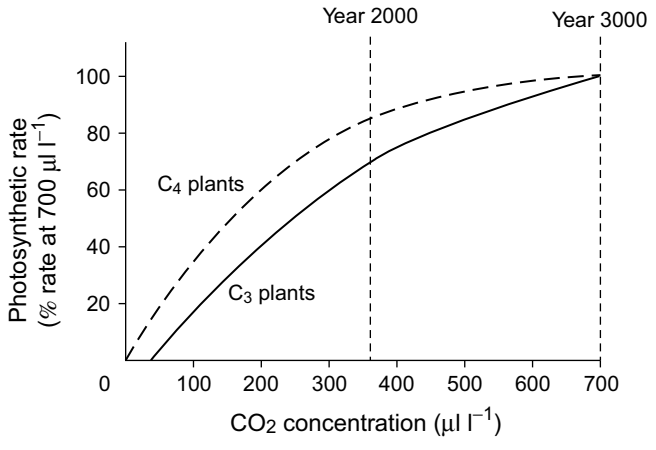


Figure 1 Dependence of CO_2 assimilation rate on atmospheric CO_2 concentration for both C_3 and C_4 plants. Vertical lines mark current and future CO_2 levels. (Redrawn from Woodward *et al.*, 1991)

Respiration

CO_2 induced increases in carbon uptake rates may be offset if elevated CO_2 causes concomitant changes in leaf respiration rates. CO_2 effects on respiration may additionally in uence plant growth and development, as energy from respiration is used to support these processes. Atmospheric CO_2 concentrations may have direct effects on plant respiration through inhibition of mitochondrial enzymes, particularly cytochrome *c* oxidase and succinate dehydrogenase (Drake *et al.*, 1999). These direct effects may be observed in relatively short-term CO_2 enrichment studies but often become a minor component of the response in the longer term as indirect effects come to dominate (Gonzàlez-Meler *et al.*, 1996). Plant respiration is a highly regulated and coordinated process, with multiple feedback controls determined proximally by ATP and NAD(P)H levels, and ultimately by plant carbon status. The complexity of the process makes it difficult to generate *a priori* predictions about the in uence of elevated CO_2 on respiration. To date, results have been equivocal, with both CO_2-induced stimulation and inhibition of respiration documented (Amthor, 1995). It is likely that both the size and composition of carbohydrate pools within plants determine these responses.

Stomatal Conductance and Water Use Efficiency

Elevated CO_2 may not only directly affect leaf metabolic processes, but it may also alter the pathway for gas exchange. Guard cell activity, and hence stomatal conductance, may be sensitive to atmospheric CO_2 concentrations (Mott, 1990). Optimality models suggest that, with rising CO_2 levels, stomatal conductance should decline to maintain a constant water loss for a given amount of carbon fixed (Cowan and Farquhar, 1977). Many studies do indeed find this predicted decline in conductance (Drake *et al.*, 1997; Morison, 1998) and even observe reductions in stomatal density in the longer-term (Woodward and Bazzaz, 1988; Woodward and Kelly, 1995). A recent meta-analysis by Curtis (1996), however, found no significant CO_2 effect on stomatal conductance in tree species. It is now important to repeat this statistically rigorous analysis for all plant species.

These changes in stomatal aperture and density may both serve to reduce water loss from plants. Increase in photosynthetic capacity and reduction in water loss combine to increase plant water use efficiency (Eamus and Jarvis, 1989; Tyree and Alexander, 1993). Whole-plant water use may not always be reduced, as plants often develop a higher leaf area in elevated CO_2 due to their larger size (Field *et al.*, 1995; Keeling *et al.*, 1996). In general, however, increased levels of atmospheric CO_2 do improve plant water relations, especially under water stress conditions (Morse *et al.*, 1993; Jackson *et al.*, 1994). This effect may be an important mechanism by which plants become better able to tolerate water stress conditions (Polley *et al.*, 1996; Catovsky and Bazzaz, 1999).

DOWN-REGULATION OF PHOTOSYNTHESIS: WHAT ARE THE MECHANISMS?

Global change research has focused particularly on effects of CO_2 on carbon uptake rates, both because of the sensitivity of photosynthesis to atmospheric CO_2 concentrations and because of the critical role of photosynthesis in regulating future carbon uxes through terrestrial ecosystems. Recognition of potential feedbacks between the atmosphere and biosphere has increased interest in determining how much of the extra CO_2 released into the atmosphere as a result of human activities will be sequestered by terrestrial vegetation. One of the principal debates in CO_2 research is the degree to which carbon fixation will be stimulated by elevated CO_2 (Idso, 1991; Fajer and Bazzaz, 1992). Although enhancement of photosynthesis is almost invariably observed in CO_2 experiments, the magnitude and duration of enhancement is very variable, and the mechanisms underpinning responses uncertain.

After long-term exposure to elevated CO_2, plants often show acclimation responses to elevated CO_2, such that photosynthetic enhancement declines over time (DeLucia *et al.*, 1985). Photosynthetic capacity for plants grown in elevated CO_2 is typically lower than for plants in ambient CO_2 (Sage, 1994). This is known as *down-regulation*. This phenomenon is one of the most critical for predicting the degree of CO_2 stimulation of terrestrial vegetation, and there is still considerable debate about its importance in determining future ecosystem productivity. Recently, it has been suggested that the degree of down-regulation in trees growing under elevated CO_2 in the field is less than once thought (Ceulemans *et al.*, 1999; Norby *et al.*, 1999). It is now essential that we consider the key factors determining down-regulation so that we can better predict how terrestrial ecosystems will respond to changing atmospheric CO_2 concentrations.

Much evidence points to the importance of sinks within the plant for substantially reducing photosynthetic down-regulation (Arp, 1991; Stitt, 1991). Sinks are non-photosynthetic tissues that act as a constant draw on photosynthetic products produced in the leaves, and thus prevent end-product inhibition of carbon fixation. The relationship between sink strength and plant responses to elevated CO_2 was recently demonstrated with seven different *Brassica* cultivars (Reekie, 1996). Two of the cultivars had large in orescence sinks, two had major carbon sinks in the stem, one of the cultivars had a root sink, and the remaining two had no substantial carbon sinks. Early in the experiment, growth of all cultivars was significantly enhanced in elevated CO_2. As the plants developed, however, only those cultivars with a substantial carbon sink continued to show significant CO_2 growth enhancements, presumably because down-regulation of photosynthesis was avoided. Sink size is clearly a more important control on photosynthetic responses to elevated CO_2 than sink location.

Photosynthesis is a highly regulated process with controls at many different stages. How do sinks in uence down-regulation? In the short-term, it is thought that acclimation occurs as a result of carbohydrate accumulation in leaves (Stitt, 1991). When there are not sufficient sinks within the plant for complete utilization of the products of photosynthesis, starch may build up in chloroplasts of photosynthetic cells (Woodrow and Berry, 1988). If the processes of carbon fixation and sucrose synthesis are not in balance, end-products of photosynthesis, such as triose phosphates, will accumulate in the chloroplast and lead to starch formation through up-regulation of the enzyme adenosine diphosphate (ADP) glucose pyrophosphorylase (Preiss, 1982). Build-up of starch in the chloroplast of photosynthetic cells clearly inhibits subsequent carbon fixation (Signora *et al.*, 1998) although the exact molecular mechanisms are currently unclear. Large starch grains may disrupt chloroplast membrane processes, or direct inhibition of photosynthesis may occur as light-dependent reactions become limited by the availability of molecular precursors (Stitt, 1991).

In the longer term, down-regulation of photosynthesis in elevated CO_2 is likely the result of a decline in Rubisco concentrations in chloroplasts. Transcripts of nuclear DNA coding the small subunit of Rubisco (rbcS) have been shown to decline in leaves exposed to elevated CO_2 (Webber *et al.*, 1994; Van Oosten and Besford, 1996; Cheng *et al.*, 1998). This effect is particularly pronounced when leaves are detached from the plant and thus any carbohydrate sinks removed, and these changes can be mimicked by supplying sucrose or glucose to leaves (Van Oosten *et al.*, 1994). Recent work suggests that these responses are directly regulated by the concentration of 6−carbon sugars (the hexoses, glucose and fructose) in photosynthetic cells (Sheen, 1994). There is now good evidence that plant cells can detect internal sugar concentrations and that such information closely regulates photosynthetic processes through the activity of

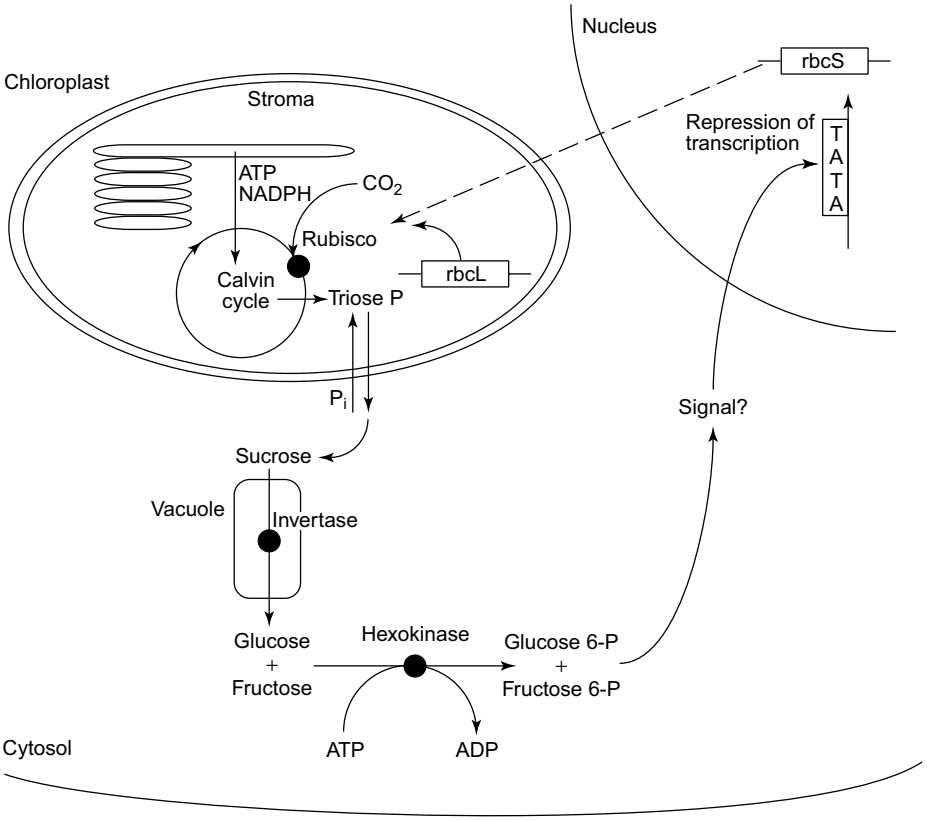


Figure 2 Proposed mechanism for long-term photosynthetic down-regulation in elevated CO_2. Components of pathway are discussed in text. (Redrawn from Van Oosten and Besford, 1996)

hexokinase enzymes (Jang and Sheen, 1997; Moore *et al.*, 1999). When sucrose accumulates in photosynthetic cells, invertase enzymes catalyze its break-down to glucose and fructose (Figure 2). Hexokinases then phosphorylate these metabolic by-products, which may then act to regulate future Rubisco transcription rates.

Now that much of the molecular biology behind photosynthetic down-regulation in elevated CO_2 is well understood, it is critical that we ascertain the significance of these processes in more natural situations and determine the sensitivity of such processes to environmental conditions. The key role of terrestrial ecosystem carbon uptake in the current global carbon cycle makes it particularly important that we address the degree to which photosynthesis is enhanced in elevated CO_2. To date, field-based studies generally support the model for carbohydrate control of photosynthesis (Figure 2), although links between different stages of regulation are less tightly coupled (e.g., Nie *et al.*, 1995). It is now well established that both low nutrient availability and lack of adequate carbon sinks may limit plant responses to elevated CO_2 (Arp, 1991; Pettersson and McDonald, 1994). In the future, we must focus on the underlying molecular mechanisms behind these empirical observations.

ELEVATED CO_2 AND WHOLE-PLANT PHYSIOLOGY

Leaf-level processes represent the primary site of activity for increases in CO_2 concentrations. The short-term physiological changes we have described, however, may not directly relate to whole-plant responses to elevated CO_2. Processes that involve the transport and partitioning of carbon, nutrients and water between different parts of the plant may decouple leaf and plant level activities (Wolfe *et al.*, 1998). The time scale for events at the leaf level and for whole-plant processes may also be different, and complex feedbacks may alter the relationship between processes at multiple scales. In this section, we summarize the major changes that plants undergo when grown in elevated CO_2, emphasizing perturbation of whole-plant physiological processes.

Growth: Dependence on Environmental Conditions and Species Identity

Changes in photosynthesis and respiration generally lead to some degree of growth enhancement in elevated CO_2, as extra assimilate can be allocated to structural or functional tissues. Numerous studies have recorded CO_2-induced

growth increases, with average enhancements of 25–50% for tree seedlings (Eamus and Jarvis, 1989; Celemans and Mousseau, 1994; Wullschleger *et al.*, 1995), 35% for C_3 wild herbaceous species (Poorter *et al.*, 1998) and 40–60% for C_3 crop species (Cure and Acock, 1986; Poorter *et al.*, 1996). We should be cautious, however, of using these average enhancement ratios for predicting ecosystem responses to elevated CO_2. These summary data are based on many studies with methodological problems, especially pseudoreplication (Jas enski *et al.*, 1998). In addition, the time dependence of enhancement ratio calculations makes it very difficult to extrapolate from generally short-term experiments to growth responses throughout an individual's life span, especially when the plants are perennial and long-lived (Thomas *et al.*, 1999). We have demonstrated that, even within the first three years of a tree's life, CO_2 responsiveness generally declines with time and in a species-specific manner (Figure 3) (Bazzaz *et al.*, 1993).

There is huge variation in plant growth responses between experiments, and as a result, average enhancement ratios do not adequately represent the complete picture. It is now well established that species differ significantly in their responses to elevated CO_2 (Bazzaz, 1990). A common goal in much CO_2 research is determining if we can identify species' traits that are well correlated with CO_2 responsiveness. In this way, we could make predictions about how particular communities will respond to future atmospheric CO_2 concentrations without growing every species in an elevated CO_2 environment. To date, it has only been possible to identify very broad functional response groupings of species (Catovsky, 1998) in relation to CO_2 concentration. There is good evidence for differences in CO_2 growth responses between species with different photosynthetic pathways. C_3 plants have consistently higher CO_2 growth increases (47%) than either C_4 (10%) or crassulace acid metabolism (CAM) (19%) species (Poorter *et al.*, 1996). Other CO_2 response groups have been more difficult to distinguish, with analyses producing contradictory findings. Ceulemans and Mousseau (1994) presented evidence that broad leaved tree species respond more strongly to elevated CO_2 than do coniferous species (63% vs. 38%), while Saxe *et al.* (1998) find the opposite result in a more recent review (49% vs. 130%). Similarly, life history strategy does not provide a clear axis on which to distinguish species' CO_2 responsiveness. Substantial evidence suggests that faster growing, early successional species are more responsive to elevated CO_2 than slow growing, late successional species (Hunt *et al.*, 1991; Bazzaz and Miao, 1993; Poorter, 1998). Again, however, this result does not always hold true (Bazzaz *et al.*, 1990; Stirling *et al.*, 1997), and an alternative hypothesis has recently been suggested (Kerstiens, 1998). Research from our lab suggests that there is an interaction between leaf habit and shade tolerance in determining CO_2 responsiveness (S Catovsky, C C Muth and F A Bazzaz, unpublished data). Growth enhancement in elevated CO_2 increased with increasing shade tolerance for seedlings of coniferous species, while CO_2 growth

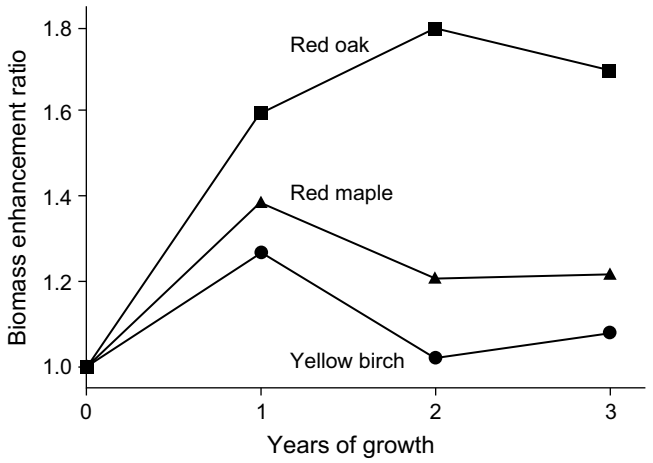


Figure 3 Change in total biomass enhancement ratio (final seedling biomass at 700 µl l^{-1}/biomass at 350 µl l^{-1}) over three consecutive growing seasons for three temperate tree species. Points represent means for species across four combinations of light and nutrient availability. (Redrawn from Bazzaz *et al.*, 1993)

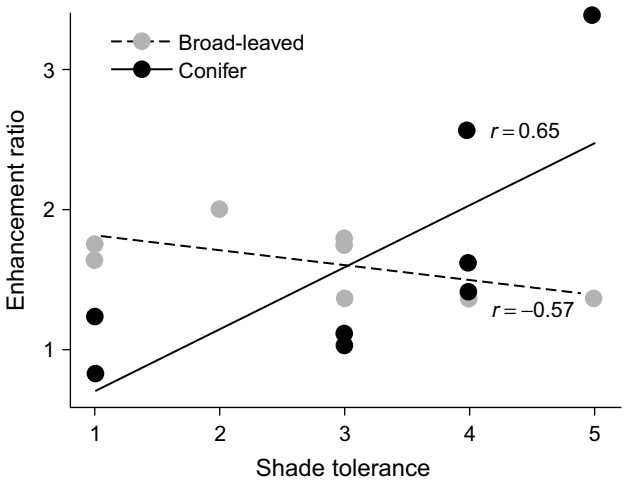


Figure 4 Relationship between species' tolerance ranking (based on Baker's Table) and CO_2 responsiveness (measured as growth enhancement ratio: mean final seedling biomass at 700 µl l^{-1}/biomass at 375 µl l^{-1}) for coniferous and broad-leaved tree species after one growing season. Product-Moment correlation coefficients (r) are shown for each relationship separately (coniferous: solid line; broad-leaved: dashed line). Species included in the study (in order of increasing tolerance): *Larix laricinia*, *Pinus banksiana*, *Pinus rigida*, *Pinus strobus*, *Picea rubens*, *Thuja occidentalis*, *Picea mariana*, *Abies balsamea* (coniferous); *Betula populifolia*, *Populus tremuloides*, *Betula papyrifera*, *Quercus rubra*, *Betula alleghaniensis*, *Quercus bicolor*, *Acer rubrum*, *Acer saccharum* (broad-leaved). (Unpublished data from S Catovsky, C C Muth, and F A Bazzaz)

enhancement decreased with increasing shade tolerance for seedlings of broad-leaved species (Figure 4).

The complex task of identifying functional CO_2 response groups is often made more challenging by substantial intra-group variation in CO_2 responsiveness. For example, we have observed signi cant differences in the responses of C_4 grasses to elevated CO_2 (Kellog *et al.*, 1999). This variation was present both between and within different genera, and between and within different C_4 lineages (Figure 5). C_4 species may be divided into three subtypes based on the acceptor molecule used during carbon x-ation: nicotinamide adenine dinucleotide phosphate-malic enzyme (NADP-ME), nicotinamide adenine dinucleotide-malic enzyme (NAD-ME), and (PCK). This difference in biochemistry may explain some of the variation in CO_2 responsiveness between C_4 species (LeCain and Morgan, 1998), but did not correlate with responsiveness in our experiment. To date, even less attention has been paid to intra-speci c variation in CO_2 responsiveness, which could also be substantial (Korner and Bazzaz, 1996). For example, Thomas and Jas enski (1996) showed that variation in CO_2 responsiveness between different genotypes of *Abutilon theophrasti* equaled the level of variation observed for tree seedlings and saplings in 153 studies.

Ecosystem responses to elevated CO_2 will not only be determined by the species present, but also by the environmental conditions that the plants experience. Growth conditions have very signi cant effects on an individual s response to atmospheric CO_2 concentration. Because

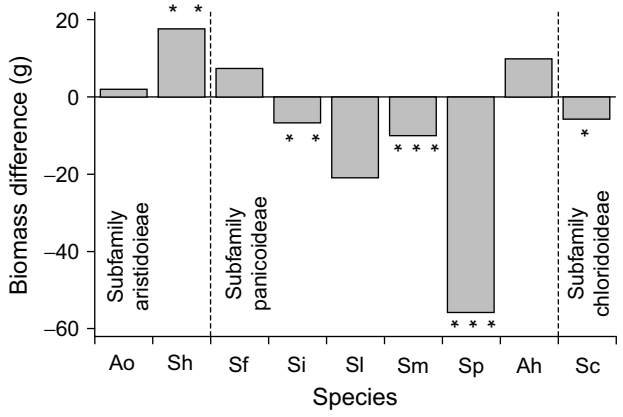


Figure 5 Differences in total biomass for grass species grown under ambient and elevated CO_2 after one growing season (total biomass at elevated CO_2 minus biomass at ambient CO_2). Significance of each difference was determined with nested Anova post-hoc comparisons: $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. Dashed vertical lines delimit separate C_4 lineages. Study species include: *Aristida oligantha* (Ao), *Stipagrostis hirtiglumis* (Sh), *Setaria faberi* (Sf), *Setaria italica* (Si), *Setaria lutescens* (Sl), *Setaria macrostachya* (Sm), *Setaria pumila* (Sp), *Arundinella hirta* (Ah), *Sporobolus cryptandrus* (Sc). (Redrawn from Kellogg *et al.*, 1999)

of the tight coupling and interdependence of above- and below-ground processes, much work has focused on the in uence of soil nutrient availability on plant responses to elevated CO_2. Many empirical studies have demonstrated that low nutrient availability constrains CO_2 responsiveness (Bazzaz, 1990; Poorter, 1998). Photosynthetic acclimation to elevated CO_2 (as discussed in the section on Down-regulation of Photosynthesis) usually appears earlier and is more marked in plants growing under low nutrient conditions (Pettersson and McDonald, 1994). In their meta-analysis of CO_2 effects on tree species, Curtis and Wang (1998) observed that low soil nutrient availability reduced plant responses to elevated CO_2 by 50%. Surprisingly, the mechanistic basis for this limitation is not well understood (Lloyd and Farquhar, 1996). As metabolic processes are reduced under reduced nutrient supply, low nutrient availability may reduce sink strength and increase accumulation of sugars within the plant, leading to a greater degree of photosynthetic down-regulation (Stitt and Krapp, 1999). Alternatively, changes in the allocation of assimilate between different plant parts may be driving these effects. Plants often develop a lower leaf area ratio in elevated CO_2 (Poorter *et al.*, 1996), as more carbon is allocated below ground. This proportional decrease in amount of leaf tissue may become more pronounced in low nutrient soils, causing relative growth increases in elevated CO_2 to be lower. This hypothesis is supported by experiments where plants are given nutrients in relation to their size (Ingestad method), and hence where above- and below-ground processes are more in balance. In such experiments, interactions between nutrient availability and CO_2 responsiveness are not observed (Pettersson *et al.*, 1993; Farage *et al.*, 1998). This suggests that it is the imbalance between above- and below-ground processes that underlies the lack of CO_2 responsiveness with low nutrient availability in most studies.

In contrast to soil nutrient availability, plants seem to respond to elevated CO_2 more strongly at low levels of light availability than at higher levels. For example, for tree seedlings, CO_2 responsiveness was almost 70% greater for plants in low vs. high light (Curtis and Wang, 1998). We do not currently understand the mechanistic basis for these effects, although a number of hypotheses have been proposed, which must now be explicitly tested. Low-light plants may suffer less photosynthetic down-regulation than high-light plants, as carbon assimilation is lower and thus is less likely to exceed demand (sucrose synthesis and export) to such a great a degree. Alternatively, elevated CO_2 could directly affect the light harvesting machinery in plants. Some studies have found that elevated CO_2 can produce decreases in the *light compensation point* (Kubiske and Pregitzer, 1996) or increases in apparent quantum yield (Lewis *et al.*, 1999). These shifts could be a function of nitrogen re-allocation from Rubisco to

light harvesting proteins. In support of this hypothesis, Kubiske and Pregitzer (1997) found a significant increase in photosynthetic nitrogen use efficiency for *Acer rubrum* saplings grown in elevated CO_2.

Water availability is another important resource that varies substantially between and within ecosystems that may in uence plant responses to elevated CO_2. As with light, evidence suggests that CO_2 responsiveness is enhanced at lower soil moisture availability to a greater degree than at high soil moisture (Morison, 1993). Large CO_2 effects under low water conditions are likely brought about through the combined effects of CO_2 on both photosynthesis and water use efficiency (see Stomatal Conductance and Water Use Efficiency section). The issue of contrasting effects of CO_2 on leaf-level instantaneous water use efficiency and whole-plant integrated water use, however, must still be resolved.

Allocation, Architecture, Phenology and Development

The effects of CO_2 concentration on growth may be accompanied by a suite of other changes in whole-plant processes. One of the most common trends observed is increased allocation to roots (Rogers *et al.*, 1994), which fits with resource allocation theory suggesting that plants allocate carbon to tissues such that all resources are equally limiting (Bloom *et al.*, 1985). As CO_2 levels increase, nutrient uptake rates must increase to balance the increased amounts of carbohydrate within the plant. To achieve this goal, plants increase allocation to nutrient acquiring organs at the expense of light capturing organs. Long-term increases in productivity in relation to elevated CO_2 may be constrained under low soil nutrient conditions (see Whole-plant Feedback on the Carbon Cycle section). Increased root to shoot ratios are particularly pronounced in soils of low nutrient availability (Norby, 1994; Curtis and Wang, 1998), again as might be expected from optimality models. Effects of elevated CO_2 on specific root uptake rates of important nutrients, such as (NH_4^+) and (NO_3^-), however, have been variable to date (Stitt and Krapp, 1999). These responses may be related to changes in nutrient concentrations in the soil solution and subsequent feedback mechanisms. Recently, the importance of plant–soil interactions in controlling long-term plant responses to elevated CO_2 has been recognized. In the subsequent section (Whole-plant Feedback on the Carbon Cycle), we discuss the current state of knowledge on these responses, and the implications for whole-ecosystem responses to rising CO_2 levels.

Changes in plant allocation patterns can clearly affect the size and activity of resource capturing parts. In addition to these changes, CO_2 concentration may in uence other whole-plant processes that themselves may determine long-term CO_2 responsiveness. Elevated CO_2 could directly in uence plant developmental patterns, and these changes, in turn, could affect spatial and temporal patterns of growth and tissue maturation. Our knowledge of such changes is still rather limited, with general trends difficult to establish. For example, for short day *Petunia hybrida* individuals, elevated CO_2 decreased branch number and number of days until owering, while plant height was increased (Reekie, 1996). In contrast, other studies have found that elevated CO_2 increases branching or tiller production (Tissue and Oechel, 1987; Rochefort and Bazzaz, 1992). An understanding of the basic physiological processes that regulate plant development is essential for making predictions about CO_2 induced changes in plant developmental patterns. CO_2 effects on plant carbohydrate status could cause major disruption to basic allometric growth patterns, although the nature of the change is currently unclear. Thomas and Bazzaz (1996) found that, under elevated CO_2, leaves of *Taraxacum of cinale* plants exhibited more deeply incised leaf margins and more slender leaf laminae than plants under ambient CO_2. They attributed these changes to effects of carbohydrate status on heteroblastic leaf development, but were unable to establish the exact nature of the link between these different processes.

Changes in developmental patterns, such as those described, could in uence the way plant parts are displayed in space, i.e., architecture. These effects may also ultimately determine long-term plant responses to CO_2 concentration. Elevated CO_2 could alter architectural patterns of both above and below ground parts. For example, Reekie and Bazzaz (1989) found that five tropical tree species showed only moderate biomass responses to elevated CO_2, even when supplied with ample resources, but, in contrast, the species underwent quite substantial architectural responses. In particular, height differences between the species became exaggerated in elevated CO_2 through differential patterns of leaf display and shoot growth. CO_2 concentrations could additionally in uence below-ground architectural patterns. Studies with crop and wild plant species have found that elevated CO_2 increases total root length, branch number and depth of root penetration through the soil (Rogers *et al.*, 1994). Architectural changes could in uence basic plant responses to elevated CO_2, e.g., through increases in total below-ground resource capture (Berntson, 1994), as well as affecting competitive interactions.

Direct effects of elevated CO_2 on plant development may in uence temporal patterns of plant part production, as well as spatial components (architecture). Many studies have recorded changes in the timing of ower, fruit and leaf production, although the direction and magnitude of change is highly variable (Garbutt and Bazzaz, 1984; Reekie and Bazzaz, 1991). Even plants within the same community may show very different phenological responses (Rusterholz and Erhardt, 1998), which, in turn, could in uence competitive interactions, as well as plant-animal

interactions. The mechanistic basis for phenological change is not yet well established, but should be actively addressed in future research if we wish to predict whole-community responses to global environmental change.

WHOLE-PLANT FEEDBACKS ON THE CARBON CYCLE

Plants may directly in uence soil ecosystem processes through litter production and through dynamic interactions between root and soil functioning. Elevated CO_2 may affect numerous whole-plant processes that feedback to in u-ence soil processes (Bazzaz, 1996). Understanding this plant–soil feedback is critical for making long-term predictions about whole-ecosystem responses to elevated CO_2, and, through global biogeochemical feedbacks, how vegetation may affect future levels of atmospheric CO_2. Changes in litter quality and perturbations to plant–soil interactions may both significantly in uence nutrient availability at a site, and thus how plants respond to elevated CO_2 in the long-term (see Growth section). In this section, we highlight the importance of both kinds of feedback for ecosystem CO_2 responsiveness (Figure 6).

Carbon and Nitrogen Partitioning

Rising atmospheric CO_2 concentrations are a direct perturbation on plant function, and, as we have discussed, can lead to multiple changes in leaf and whole-plant level processes. If sufficient sinks exist, elevated CO_2 increases the amount of carbohydrate within the plant, and the concentration of sugar and starch in leaves

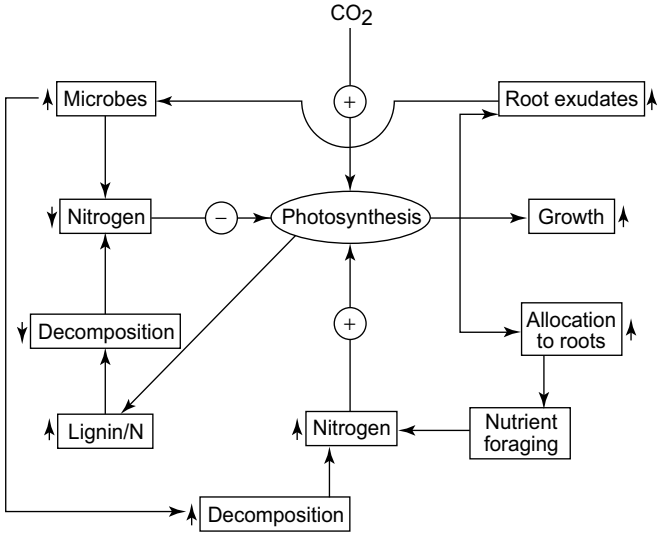


Figure 6 Schematic outlining positive and negative plant–soil feedback processes that determine long-term plant responses to elevated CO_2. (Modified from Bazzaz, 1996)

(see Down-regulation of Photosynthesis section). Nutrient uptake rates rarely increase in direct proportion to the amount of extra carbon taken up by plants in elevated CO_2, and thus tissue carbon/nutrient ratios invariably increase (especially nitrogen) (Lindroth *et al.*, 1993; Traw *et al.*, 1996). A recent review of 75 published studies demonstrated that elevated CO_2 led to an average reduction in tissue nitrogen concentration of 14% (Cotrufo *et al.*, 1998), with the magnitude of decrease varying among species. What mechanisms determine the degree of enhancement of C/N ratio between species? Recently, Bazzaz (2000) proposed that the degree of down-regulation might directly in uence C/N ratios, with plants that show little down-regulation exhibiting the highest degree of nitrogen dilution.

The significance of these changes in tissue chemistry for decomposition processes was recognized early in global change research. Strain and Bazzaz (1983) originally proposed that increased tissue C/N ratios could translate into slower rates of tissue decomposition, and thus act as a negative feedback on CO_2-induced enhancement of productivity. The composition of the additional carbon incorporated into plant tissues will be an important determinant of this relationship. Increases in structural carbohydrates (especially lignin) are more likely to slow decomposition rates than are increases in total non-structural carbon (Melillo *et al.*, 1982).

It is still unclear, however, whether or not increased C/N ratios in living tissues translate into altered chemistry of senesced leaf litter (O'Neill and Norby, 1996). Species have variable patterns of nutrient retranslocation prior to senescence, and rising CO_2 levels could further alter these dynamics. Such changes could decouple the relationship between tissue and litter nutrient concentrations, although they are unlikely to completely reverse trends observed for living tissues. Overall, many studies do find that litter from plants grown in elevated CO_2 decomposes more slowly than litter produced in ambient CO_2 (Cotrufo and Ineson, 1996). In field decomposition studies, we have found species-specific declines in the decomposition rate constant (K) for litter of tree seedlings grown in elevated vs. ambient CO_2 (Figure 7, J M Melillo and F A Bazzaz, unpublished data). Recently, however, results from whole-ecosystem manipulations of atmospheric CO_2 concentration have found only marginal effects of elevated CO_2 on litter decomposition and thus ecologists are beginning to question the generality of the original hypothesis (Norby and Cotrufo, 1998; Couteaux *et al.*, 1999).

These changes in C/N ratio have implications for higher order trophic interactions in ecosystems as well. Nitrogen is often the single most important limiting resource for phytophagous insects, and a reduction in the quality of host plant food (through nitrogen dilution) could significantly affect the dynamics of insect herbivore populations, as well

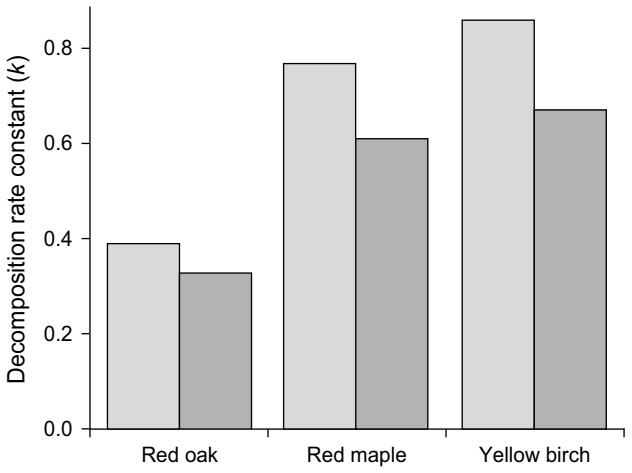


Figure 7 Influence of CO_2 concentration (350 µl l^{-1} = light grey, 700 µl l^{-1} = dark grey) on litter decomposition rate constant ($k = -\ln$ (% litter mass loss)/time) for seedlings of three temperate forest species grown in controlled environment conditions. Leaves were allowed to senesce naturally and then were transferred to litter decomposition bags in Harvard Forest. (Unpublished data from J M Mellilo and F A Bazzaz)

as species further up the food web. From the studies to date, it has been difficult to determine general patterns in herbivore responses to altered host plant quality. Part of the difficulty lies with the wide range of herbivore responses observed, but also with the narrow range of taxonomic and functional diversity sampled in the experiments. Most of the studies have been carried out on leaf-chewing insect herbivores, and for this guild of insects, at least, it is clear that insects fed on leaves from plants grown in elevated CO_2 grow more slowly, and consume more of the host plant. They also take longer to reach maturity, and have lower survival than insects fed on ambient CO_2-grown plants (Bezemer and Jones, 1998; Coviella and Trumble, 1999). Even within this one feeding guild, insect responses vary substantially with host plant species identity and insect developmental stage and sex (Fajer *et al.*, 1989; Traw *et al.*, 1996).

Plant–Soil Interactions

Part of the dif culty with understanding the feedbacks between plant responses to CO_2 concentration and ecosystem-level nutrient availability arises because decomposition rates not only depend on litter chemistry, but also involve a whole series of complex plant soil processes (Figure 6) (Berntson and Bazzaz, 1996). Together, such interactions may substantially alter any potential negative feedback through tissue C/N ratios. Recently, research has focused on CO_2 effects on a number of key plant soil interactions that may signi cantly in uence long-term plant responses to elevated CO_2. Direct effects of elevated CO_2 on plant function may lead to a whole suite of changes in below-ground activities, which themselves in uence processes within the soil itself. It is now well established that elevated CO_2 not only increases the amount of structural carbohydrate allocated to root growth (see Allocation section), but also increases the amounts of labile carbon that pass through the roots. These more transient changes are necessarily harder to detect and measure, as much of the carbon eventually enters the soil organic matter. Nevertheless, such changes play a vital role in determining ecosystem-level feedback to CO_2 growth enhancements and in in uencing carbon ux through soils, which could be a highly signi cant component of total ecosystem carbon gain.

The importance of such changes in labile carbon pools was recognized by Norby *et al.* (1992) in a 3-year eld CO_2 experiment with *Liriodendron tulipifera*. Norby *et al.*, observed signi cant enhancement of leaf-level carbon uptake rates in elevated CO_2, but no concomitant increase in plant biomass. Further exploration suggested that the extra assimilate generated by increased photosynthesis in elevated CO_2 was allocated to ne-root production and turnover (Figure 8). These static measurements of below-ground productivity may, however, substantially underestimate total root turnover rates. Berntson and Bazzaz (1997a) observed increases in both gross ne-root production and loss for *Betula papyrifera* seedlings in elevated CO_2. Effects of CO_2 on these processes were seasonally dependent, with CO_2 enhancement of ne-root production occurring early in the season, while CO_2 enhancement of ne-root loss occurred only in the last half of the growing season. These changes led to a CO_2-induced transient increase in standing root length early in the season that disappeared by the end of the season.

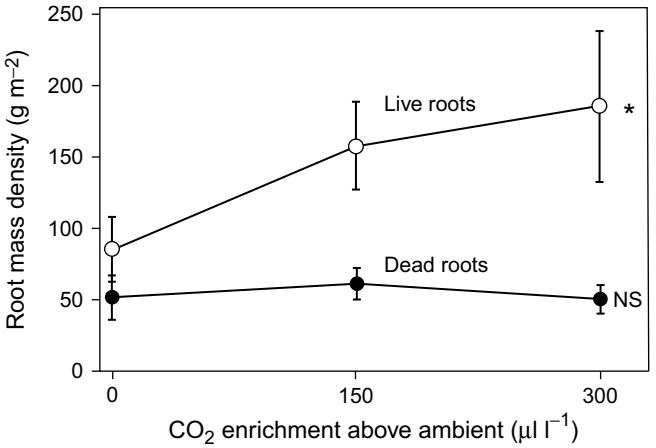


Figure 8 Changes in live and dead root mass density with different levels of CO_2 enrichment for *Liriodendron tulipifera* after three growing seasons (means ± SEM from two replicate open-top chambers). Significance of CO_2 concentration effect shown (NS $P > 0.05$, $^{*}P < 0.05$). (Data taken from Norby *et al.*, 1992)

Elevated CO_2 may also lead to increased losses of carbon from roots into the soil (Cheng, 1999), which together with increases in fine-root turnover, could have important consequences for the dynamics of soil microbial populations (O'Neill, 1994). An increase in the amount of labile carbon entering the rhizosphere is likely to directly stimulate microbial activity and other soil fauna (Jones *et al.*, 1998). As microorganisms play a key role in soil nutrient dynamics, such changes could have significant implications for feedback on plant responses to CO_2. Soil microbe populations are often enhanced following plant exposure to elevated CO_2, although the degree of enhancement may vary substantially with the type of microorganism concerned. For example, elevated CO_2 causes a greater stimulation of root colonization for ectomycorrhizal fungi than for vesicular arbuscular mycorrhizae (O'Neill, 1994; Staddon and Fitter, 1998). Even within one mycorrhizal type, there may be substantial variation in response. Godbold and Berntson (1997) recorded significant changes in the composition of ectomycorrhizal assemblages on *Betula papyrifera* seedlings (Figure 9), with a noticeable shift towards morphotypes with high hyphal density.

Although CO_2-induced changes in microbial abundance and composition have now been recorded in many studies, the functional significance of such changes for plant nutrient dynamics has not been adequately evaluated. This gap in our understanding is well illustrated by considering the opposing hypotheses that have been put forward to link microbial dynamics with soil nutrient processes (Berntson and Bazzaz, 1996). In an open top experiment with *Populus grandidentata* saplings, Zak *et al.* (1993) found that elevated CO_2 increased biomass of soil microorganisms and potential mineralization rates, and hypothesized that increased microbial activity could act as a direct positive feedback on plant responses to elevated CO_2. A contrasting hypothesis was put forward by D az *et al.* (1993) based on an experiment exposing model chalk grassland ecosystems to different CO_2 concentrations. They observed marginal enhancement of above ground biomass growth, finding that much of the additional carbon was entering the soil and stimulating microbial growth. Marked signs of nutrient deficiency in plants grown in elevated CO_2, even following fertilization, suggested that the expanded microbe populations were competing with the plants for soil nitrogen and leading to net nitrogen immobilization.

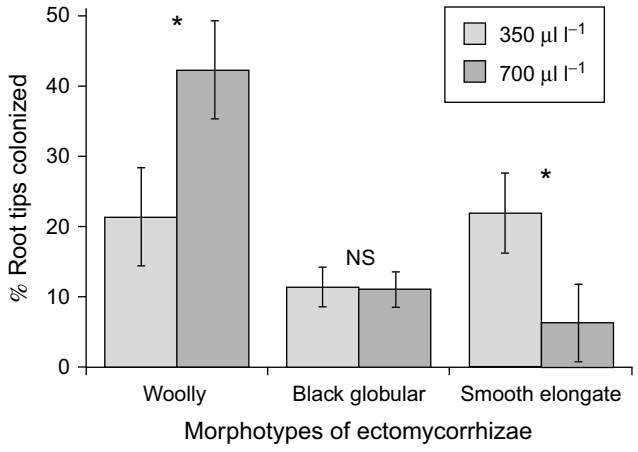


Figure 9 Degree of *Betula papyrifera* root colonization (mean ± SEM) by three different ectomycorrhizal morphotypes under different CO_2 concentrations. (Redrawn from Godbold and Bazzaz, 1997)

Plant responses to elevated CO_2 can clearly lead to simultaneous positive and negative feedback on the global carbon cycle. How do we disentangle these different processes? Now that many of the key processes have been identified, we need to take a mechanistic approach to the issue of plant–soil feedback and determine how the relative importance of different processes changes with ecosystem type and environmental conditions. In two parallel experiments, Berntson and Bazzaz (1997b, 1998) grew stands of *Betula alleghaniensis* and *Betula papyrifera* (planted and natural regeneration from the soil seed bank) and made simultaneous measures of many below-ground processes relating to nutrient availability. They found that, overall, plant nutrient uptake rates were reduced in elevated CO_2. The increases in potential uptake from greater root biomass and increased mycorrhizal colonization (positive feedback) were not sufficient to compensate for reduced mineralization rates (a negative feedback through increased litter C/N). We are likely to reach different conclusions in other systems, and thus we need to determine what ecosystem characteristics determine the relative importance of different feedback components.

INTERACTIONS BETWEEN GLOBAL ENVIRONMENTAL CHANGES

Human activities now have far-reaching consequences for fundamental atmospheric, geological and biological processes. For most of this paper, we have considered plant responses to one of the best documented, most certain global environmental changes that will likely in uence natural ecosystems in the coming decades, i.e., rising atmospheric CO_2 concentrations. These effects, however, will not take place in isolation, as other environmental perturbations will also have substantial impacts on terrestrial ecosystems. In this section, we discuss how two other major environmental changes (nitrogen deposition and climate change) may modify effects of elevated CO_2 on plant function.

Nitrogen Deposition

Substantial quantities of inorganic nitrogen are being deposited in terrestrial ecosystems across much of the industrialized world, as a result of fertilizer production, legume

cultivation and fossil fuel burning (Galloway *et al.*, 1995). As nitrogen is an essential nutrient for plant function and one that may be limiting productivity in many terrestrial ecosystems (Vitousek and Howarth, 1991), this additional nitrogen may increase plant growth in natural systems. Nitrogen fertilization commonly increases foliar nitrogen concentrations (Boxman *et al.*, 1998), which leads to increased rates of photosynthesis through increases in leaf chlorophyll and Rubisco (Field and Mooney, 1986). Nitrogen additions, however, may not consistently increase ecosystem productivity, as high levels of nitrogen deposition could detrimentally in uence plant function through creation of nutrient imbalances within plants (Schulze, 1989; Aber *et al.*, 1998).

Nitrogen deposition and elevated CO_2 may act synergistically to increase ecosystem-level net primary production (Lloyd, 1999), as nitrogen limitation often constrains plant and ecosystem responses to increasing atmospheric CO_2 (see Growth' section). Recently, spatially explicit global carbon models have combined three-dimensional chemical transport models with an analysis of C/N ratios in different plant and soil compartments to estimate the contribution of nitrogen deposition to terrestrial carbon storage (Holland *et al.*, 1997). Many such models suggest that nitrogen deposition may be responsible for over half of the global carbon sink in terrestrial ecosystems ($1\,Pg\,C\,yr^{-1}$) (Norby, 1998), although recent empirical work throws these estimates into question (Nadelhoffer *et al.*, 1999). These models provide good preliminary evidence for potential synergisms between elevated CO_2 and nitrogen deposition. On the whole, however, the models lack an underlying mechanistic basis, and still rely on some broad generalizations. We need to improve our physiological understanding of interactions between carbon and nitrogen dynamics within plants to provide a better framework for future global-scale models.

See also: Nitrogen Cycle, Volume 2; Nitrogen Deposition on Forests, Volume 2.

Climate Change

Anthropogenic perturbations to the physical and chemical structure of the atmosphere are likely to bring about associated changes in large-scale climatic patterns. As we discussed earlier, the precise nature of such changes are difficult to predict accurately. It is likely, however, that elevated CO_2 will increase mean global temperatures, as well as increasing the frequency of extreme weather events (Houghton *et al.*, 1996). Despite its potential importance, we know relatively little about plant responses to the combination of elevated atmospheric CO_2 and higher temperature.

Heat is not a resource that is directly used by plants, rather it is a controller that regulates the rate and direction of different physiological processes within plants (Bazzaz,

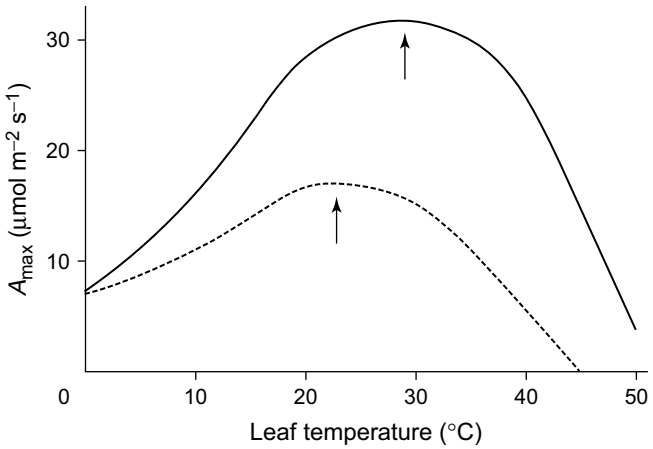


Figure 10 Predicted light-saturated photosynthetic rates (A_{max}) as a function of leaf temperature and CO_2 concentration ($350\,\mu l\,l^{-1}$ = dashed line; $700\,\mu l\,l^{-1}$ = solid line), based on a Farquhar-type leaf photosynthesis model. Arrows mark temperature optima for each CO_2 concentration. (Redrawn from Long, 1991)

1996). Temperature may act directly at the leaf-level to regulate carbon uptake rates. Increased temperature reduces the solubility of CO_2 in the cellular matrix and reduces Rubisco specificity for CO_2 vs. O_2 (Jordan and Ogren, 1984). Together, these processes reduce the efficiency of photosynthesis, although these changes may be countered to some extent by temperature-related increases in the rate of Calvin cycle reactions (Berry and Bjorkman, 1980). Theoretical photosynthesis models suggest that there is some degree of synergism between CO_2 and temperature perturbations (Long, 1991). Elevated CO_2 causes a greater degree of photosynthetic enhancement at higher temperatures and the temperature optimum for photosynthesis increases with increasing CO_2 concentration (Figure 10). This synergism arises because elevated CO_2 counters photorespiratory carbon losses (see Photosynthesis' section), which increase with increasing temperature. Experiments exclusively addressing short-term gas exchange responses to elevated CO_2 and temperature do support the synergistic effects predicted by the models (Morison and Lawlor, 1999).

A suite of other leaf-level and whole plant processes are in uenced by temperature, and these additional temperature effects may uncouple the synergism between temperature and CO_2 suggested by the leaf-level model. For example, increased temperature invariably increases both dark respiration rates and transpiration rates (because of higher vapor pressure deficit), and as a result decreases whole-plant water use efficiency (Eamus, 1991). Temperature may also in uence patterns of carbon allocation within plants (Farrar and Williams, 1991), as well as advancing the timing of many developmental events, such as seedling emergence (Farnsworth *et al.*, 1995), bud break (Ackerly *et al.*, 1992)

and owering (Price and Waser, 1998). All these effects add to the mechanistic complexity of CO_2-temperature interactions. Although some studies have found that elevated CO_2 causes a greater proportional enhancement of growth at higher temperatures as predicted by the model, overall there is little consensus amongst studies that address whole-plant responses to CO_2 and temperature (Morison and Lawlor, 1999).

Our inability to detect clear trends in the way that CO_2 concentration and temperature interact to in uence plant growth and development stems from both conceptual and methodological difficulties. We currently lack mechanistic models that adequately address the complexity of the interactions between these perturbations at the whole-plant level. To achieve this goal, we need to incorporate an understanding of plant developmental patterns, nutrient use patterns and carbon partitioning into current models of leaf-level photosynthesis. Empirical support for these models may be difficult to achieve because of a lack of consensus on temperature regimes to be used in experiments. Elevated CO_2 studies tend to use standard ambient ($350-375\,\mu l\,l^{-1}$) and twice-ambient ($700-800\,\mu l\,l^{-1}$) treatments, while ranges in temperature studies vary from 2 to 40 °C. The reference temperature used is also highly variable, and does not always relate to the natural conditions experienced by the species across its range. We clearly need to resolve these methodological issues before we can begin to validate theoretical predictions about CO_2-temperature interactions.

An increase in global mean temperature will clearly have some far-reaching consequences for plant function in natural ecosystems. Another area of concern relates to extreme temperature events, which may increase in frequency as a consequence of climatic change, e.g., El Nino years (Timmermann *et al.*, 1999). Short periods of extreme cold or heat could have a lasting impact on plant growth and development, and could ultimately in uence the regional distribution and abundance of species. Through its effects on plant function, elevated CO_2 could in uence plant responses to extreme temperature events. The limited evidence we have to date suggests that elevated CO_2 may improve frost hardiness and mitigate freezing stress (Repo *et al.*, 1996; Wayne *et al.*, 1998), while it may sometimes exacerbate heat stress, depending on the species (Coleman *et al.*, 1991; Bassow *et al.*, 1994). Mechanisms for these observations are currently not well established, although some hypotheses have been suggested. CO_2-induced freezing tolerance may arise from synthesis of additional protective carbohydrates during the growing season, while elevated CO_2 may increase the frequency of heat stress, because the cooling benefits of transpiration are reduced as stomatal conductance declines.

See also: CO_2 Enrichment: Effects on Ecosystems, Volume 2; Natural Systems: Impacts of Climate Change, Volume 2; Plant Competition in an Elevated CO_2 World, Volume 2; Plant Growth at Elevated CO_2, Volume 2; Terrestrial and Freshwater Ecosystems: Impacts of Global Change, Volume 2.

CONCLUSIONS

The widespread, rapid and dramatic changes that natural ecosystems are undergoing as a result of human activities make it essential that we understand the responses of ecological systems to novel perturbations. As the behavior of individual plants will ultimately determine many of the key ecological processes that take place within ecosystems, it is critical that we examine how major environmental changes will in uence plant function. We have observed that many human-induced perturbations directly affect basic plant processes, and we can learn much by addressing the impact of global change on underlying physiological mechanisms within plants. Research in this area has progressed by developing hypotheses about future effects of environmental perturbations from first-hand physiological principles, then testing these hypotheses with experimental work (normally at the level of individual tissues, e.g., leaves, fine roots, followed by whole-plant, integrated research), and then re-evaluating the hypotheses based on the degree of agreement between theory and empirical evidence.

Due to the importance of rising levels of atmospheric CO_2 for natural ecosystems in the future, we have focused most of this review on the effects of elevated CO_2 on plants. Increased CO_2 levels are a particularly significant novel perturbation because of potential feedback between CO_2-induced enhancements of ecosystem productivity and the slowing of atmospheric CO_2 increases through uptake by vegetation. Many of the direct and indirect effects of elevated CO_2 on plants have implications for these feedback processes. Developing an understanding of the mechanisms underlying these feedback will allow us to better predict the future role of terrestrial vegetation in the global carbon cycle.

One of the most critical areas of current global change research concerns the degree to which elevated CO_2 will stimulate leaf-level photosynthesis. The magnitude of photosynthetic enhancement is largely determined by the extent to which down-regulation occurs. Plants that undergo strong acclimation responses to elevated CO_2 typically show small CO_2 growth enhancements. We now have a relatively good understanding of the physiological mechanisms underlying down-regulation. To date, however, we have been unable to link this mechanistic understanding with explaining the large variation in CO_2 responsiveness observed across species and environmental conditions. Establishing these connections should be a priority in future CO_2 research.

Interactions between plant and soil processes also form a critical component of potential biosphere-atmosphere feedback. Nutrient availability often constrains plant responses to elevated CO_2, but could, itself, be influenced by the activity of plants within the ecosystem. Elevated CO_2 influences a number of whole-plant processes that themselves determine patterns of ecosystem-level nutrient cycling. Again, developing a mechanistic understanding of the suite of CO_2-induced changes in plant physiological processes is essential for improving our ability to predict long-term ecosystem responses to rising atmospheric CO_2. Different plant processes lead to both positive and negative feedback on ecosystem nutrient availability. The relative importance of these processes in different systems and under different environmental conditions must now be established. Examining these feedbacks and connections becomes particularly important when we consider the suite of other global environmental perturbations that are also likely to affect terrestrial ecosystem functioning. For example, nitrogen deposition may interact synergistically with elevated CO_2 to enhance plant growth and development by countering these negative feedbacks on plant productivity. Depending on future rates of deposition, however, additional nitrogen in natural ecosystems could have a detrimental impact prior to potential beneficial effects in conjunction with elevated CO_2.

Our understanding of CO_2 effects on plants has advanced considerably since the field emerged over two decades ago. We are still hindered, however, by the difficulty of connecting theoretical predictions with empirical evidence and of considering many interacting processes simultaneously. We must now make a concerted effort to improve the gaps in our knowledge, particularly by focusing on processes that relate directly to interactions between the atmosphere and natural ecosystems.

REFERENCES

Aber, J, McDowell, W, Nadelhoffer, K, Magill, A, Berntson, G, Kamakea, M, McNulty, S, Currie, W, Rustad, L, and Fernandez, I (1998) Nitrogen Saturation in Temperate Forest Ecosystems, *BioScience*, **48**, 921 934.

Ackerly, D D, Coleman, J S, Morse, S R, and Bazzaz, F A (1992) CO_2 and Temperature Effects on Leaf Area Production in Two Annual Plant Species, *Ecology*, **73**, 1260 1269.

Amthor, J S (1995) Terrestrial Higher-plant Response to Increasing Atmospheric [CO_2] in Relation to the Global Carbon Cycle, *Global Change Biol.*, **1**, 243 274.

Arp, W J (1991) Effects of Source-sink Relations on Photosynthetic Acclimation to Elevated CO_2, *Plant, Cell Environ.*, **14**, 869 875.

Bassow, S L, McConnaughay, K D M, and Bazzaz, F A (1994) The Response of Temperate Tree Seedlings Grown in Elevated CO_2 to Extreme Temperature Events, *Ecol. Appl.*, **4**, 593 603.

Bazzaz, F A (1990) The Response of Natural Ecosystems to the Rising Global CO_2 Levels, *Annu. Rev. Ecol. Syst.*, **21**, 167 196.

Bazzaz, F A (1996) *Plants in Changing Environments: Linking Physiological, Population, and Community Ecology*, Cambridge University Press, Cambridge.

Bazzaz, F A (2000) Feedbacks on the Carbon Cycle: Down-regulation of Photosynthesis and Changes in Litter Chemistry, *Plant Soil*, in press.

Bazzaz, F A, Coleman, J S, and Morse, S R (1990) Growth Responses of Seven Major Co-occurring Tree Species of the Northeastern United States to Elevated CO_2, *Can. J. For. Res.*, **20**, 1479 1484.

Bazzaz, F A and Miao, S L (1993) Successional Status, Seed Size, and Responses of Tree Seedlings to CO_2, Light, and Nutrients, *Ecology*, **74**, 104 112.

Bazzaz, F A, Miao, S L, and Wayne, P M (1993) CO_2-induced Growth Enhancements of Co-occurring Tree Species Decline at Different Rates, *Oecologia*, **96**, 478 482.

Berntson, G M (1994) Modelling Root Architecture: are there Tradeoffs between Efficiency and Potential of Resource Acquisition? *New Phytol.*, **127**, 483 493.

Berntson, G M and Bazzaz, F A (1996) Below Ground Positive and Negative Feedback on CO_2 Growth Enhancement, *Plant Soil*, **187**, 119 131.

Berntson, G M and Bazzaz, F A (1997a) Elevated CO_2 and the Magnitude and Seasonal Dynamics of Root Production and Loss in *Betula papyrifera*, *Plant Soil*, **190**, 211 216.

Berntson, G M and Bazzaz, F A (1997b) Nitrogen Cycling in Microcosms of Yellow Birch Exposed to Elevated CO_2: Simultaneous Positive and Negative Below-ground Feedback, *Global Change Biol.*, **3**, 247 258.

Berntson, G M and Bazzaz, F A (1998) Regenerating Temperate Forest Mesocosms in Elevated CO_2: below Ground Growth and Nitrogen Cycling, *Oecologia*, **113**, 115 125.

Berry, J A and Bjorkman, O (1980) Photosynthetic Response and Adaptation to Temperature in Higher Olants, *Annu. Rev. Plant Physiol.*, **31**, 491 543.

Bezemer, T M and Jones, T H (1998) Plant Insect Herbivore Interactions in Elevated Atmospheric CO_2: Quantitative Analyses and Guild Effects, *Oikos*, **82**, 212 222.

Bloom, A J, Chapin, III, F S, and Mooney, H A (1985) Resource Limitation in Plants an Economic Analogy, *Annu. Rev. Ecol. Syst.*, **16**, 363 392.

Boxman, A W, Blanck, K, Brandrud, T-E, Emmett, B A, Gundersen, P, Hogervorst, R F, Kjonaas, O J, Persson, H, and Timmermann, V (1998) Vegetation and Soil Biota Response to Experimentally-changed Nitrogen Inputs in Coniferous Forest Ecosystems of the NITREX Project, *For. Ecol. Manage.*, **101**, 65 79.

Carlson, R W and Bazzaz, F A (1980) The Effects of Elevated CO_2 Concentrations on Growth, Photosynthesis, Transpiration, and Water use Efficiency in Plants, in: *Environmental and Climatic Impact of Coal Utilization*, eds J J Singh and A Deepak, Academic Press, New York, 609 623.

Catovsky, S (1998) Functional Groups: Clarifying Our use of the Term, *Bull. Ecol. Soc. Am.*, **79**, 126 127.

Catovsky, S and Bazzaz, F A (1999) Elevated CO_2 Influences the Responses of Two Birch Species to Soil Moisture: Implications

for Forest Community Structure, *Global Change Biol.*, in press.

Ceulemans, R, Janssens, I A, and Jach, M E (1999) Effects of CO_2 Enrichment on Trees and Forests: Lessons to be Learned in View of Future Ecosystem Studies, *Ann. Bot.*, 84, 577–590.

Ceulemans, R and Mousseau, M (1994) Effects of Elevated Atmospheric CO_2 on Woody Plants, *New Phytol.*, 127, 425–446.

Cheng, S H, Moore, B D, and Seemann, J R (1998) Effects of Short- and Long-term Elevated CO_2 on the Expression of Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase Genes and Carbohydrate Accumulation in Leaves of *Arabidopsis thaliana* (L.) Heynh, *Plant Physiol.*, 116, 715–723.

Cheng, W (1999) Rhizosphere Feedback in Elevated CO_2, *Tree Physiol.*, 19, 313–320.

Coleman, J S, Rochefort, L, Bazzaz, F A, and Woodward, F I (1991) Atmospheric CO_2, Plant Nitrogen Status and the Susceptibility of Plants to an Acute Increase in Temperature, *Plant, Cell Environ.*, 14, 667–674.

Cotrufo, M F and Ineson, P (1996) Elevated CO_2 Reduces Field Decomposition Rates of *Betula pendula* (Roth.) Leaf Litter, *Oecologia*, 106, 525–530.

Cotrufo, M F, Ineson, P, and Scott, A (1998) Elevated CO_2 Reduces the Nitrogen Concentration of Plant Tissues, *Global Change Biol.*, 4, 43–54.

Couteaux, M-M, Kurz, C, Bottner, P, and Raschi, A (1999) In uence of Increased Atmospheric CO_2 Concentration on Quality of Plant Material and Litter Decomposition, *Tree Physiol.*, 19, 301–311.

Coviella, C E and Trumble, J T (1999) Effects of Elevated Atmospheric Carbon Dioxide on Insect–Plant Interactions, *Conserv. Biol.*, 13, 700–712.

Cowan, I R and Farquhar, G D (1977) Stomatal Function in Relation to Leaf Metabolism and Environment, *Symp. Soc. Exp. Biol.*, 31, 471–505.

Cure, J D and Acock, B (1986) Crop Responses to Carbon Dioxide Doubling: A Literature Survey, *Agri. For. Meteorol.*, 38, 145.

Curtis, P S (1996) A Meta-analysis of Leaf Gas Exchange and Nitrogen in Trees Grown under Elevated Carbon Dioxide, *Plant, Cell Environ.*, 19, 127–137.

Curtis, P S and Wang, X (1998) A Meta-analysis of Elevated CO_2 Effects on Woody Plant Mass, form, and Physiology, *Oecologia*, 113, 299–313.

DeLucia, E H, Sasek, T W, and Strain, B R (1985) Photosynthetic Inhibition after Long-term Exposure to Elevated Atmospheric Carbon Dioxide, *Photosynth. Res.*, 7, 175–184.

D az, S, Grime, J P, Harris, J, and McPherson, E (1993) Evidence of a Feedback Mechanism Limiting Plant Response to Elevated Carbon Dioxide, *Nature*, 364, 616–617.

Drake, B G, Azcon-Bieto, J, Berry, J, Bunce, J, Dijkstra, P, Farrar, J, Gifford, R M, Gonzalez-Meler, M A, Koch, G, Lambers, H, Siedow, J, and Wullschleger, S (1999) Does Elevated Atmospheric CO_2 Concentration Inhibit Mitochondrial Respiration in Green Plants? *Plant, Cell Environ.*, 22, 649–657.

Drake, B G, Gonzalez-Meler, M A, and Long, S P (1997) More Efficient Plants: A Consequence of Rising Atmospheric CO_2? *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, 48, 609–639.

Eamus, D (1991) The Interaction of Rising CO_2 and Temperatures With Water use Efficiency, *Plant Cell Environ.*, 14, 843–852.

Eamus, D and Jarvis, P G (1989) The Direct Effects of Increase in the Global Atmospheric CO_2 Concentration on Natural and Commercial Temperate Trees and Forests, *Adv. Ecol. Res.*, 19, 1–55.

Fajer, E D and Bazzaz, F A (1992) Is Carbon Dioxide a Good' Greenhouse Gas? *Global Environ. Change*, 2, 301–310.

Fajer, E D, Bowers, M D, and Bazzaz, F A (1989) The Effects of Enriched Carbon Dioxide Atmospheres on Plant-Insect Herbivore Interactions, *Science*, 243, 1198–1200.

Farage, P K, McKee, I F, and Long, S P (1998) Does a Low Nitrogen Supply Necessarily Lead to Acclimation of Photosynthesis to Elevated CO_2? *Plant Physiol.*, 118, 573–580.

Farnsworth, E J, Nunez-Farfan, J, Careaga, S A, and Bazzaz, F A (1995) Phenology and Growth of Three Temperate Forest Life Forms in Response to Artificial Soil Warming, *J. Ecol.*, 83, 967–977.

Farrar, J F and Williams, M L (1991) The Effects of Increased Atmospheric Carbon Dioxide and Temperature on Carbon Partitioning, Source-sink Relations and Respiration, *Plant Cell Environ.*, 14, 819–830.

Field, C and Mooney, H A (1986) The Photosynthesis-Nitrogen Relationship in Wild Plants, in *On the Economy of Plant form and Function*, ed T J Givnish, Cambridge University Press, Cambridge, 25–55.

Field, C B, Jackson, R B, and Mooney, H A (1995) Stomatal Responses to Increased CO_2: Implications from the Plant to the Global Scale, *Plant Cell Environ.*, 18, 1214–1225.

Galloway, J N, Schlesinger, W H, Levy, II, H, Michaels, A, and Schnoor, J L (1995) Nitrogen Fixation: Anthropogenic Enhancement – Environmental Response, *Global Biogeochem. Cycles*, 9, 235–252.

Garbutt, K and Bazzaz, F A (1984) The Effects of Elevated CO_2 on Plants: III. Flower, Fruit and Seed Production and Abortion, *New Phytol.*, 98, 433–446.

Godbold, D L and Berntson, G M (1997) Elevated Atmospheric CO_2 Concentration Changes Ectomycorrhizal Morphotype Assemblages in *Betula papyrifera*, *Tree Physiol.*, 17, 347–350.

Gonzàlez-Meler, M A, Ribas Carbo, M, Siedow, J N, and Drake, B G (1996) Direct Inhibition of Plant Mitochondrial Respiration by Elevated CO_2, *Plant Physiol.*, 112, 1349–1355.

Holland, E A, Braswell, B H, Lamarque, J-F, Townsend, A, Sulzman, J, Muller, J-F, Dentener, F, Brasseur, G, Levy, II, H, Penner, J E, and Roelofs, G (1997) Variations in the Predicted Spatial Distribution of Atmospheric Nitrogen Deposition and their Impact on Carbon Uptake by Terrestrial Ecosystems, *J. Geophys. Res. (Atmos.)*, 102, 15 849–15 866.

Houghton, J T, Meira Filho, L G, Callander, B A, Harris, N, Kattenberg, A, and Maskell, K, eds (1996) *Climate Change 1995: the Science of Climate Change*, Cambridge University Press: Cambridge.

Hunt, R, Hand, D W, Hannah, M A, and Neal, A M (1991) Response to CO_2 Enrichment in 27 Herbaceous Species, *Func. Ecol.*, 5, 410–421.

Idso, S B (1991) Carbon Dioxide and the Fate of Earth, *Global Environ. Change*, 1, 178–182.

Jackson, R B, Sala, O E, Field, C B, and Mooney, H A (1994) CO_2 Alters Water use, Carbon Gain, and Yield for Dominant Species in a Natural Grassland, *Oecologia*, 98, 257–262.

Jang, J C and Sheen, J (1997) Sugar Sensing in Higher Plants, *Trends Plant Sci.*, 2, 208–214.

Jas enski, M, Thomas, S C, and Bazzaz, F A (1998) Blaming the Trees: a Critique of Research on Forest Responses to High CO_2, *Trends Ecol. Evol.*, 13, 427.

Jones, T H, Thompson, L J, Lawton, J H, Bezemer, T M, Bardgett, R D, Blackburn, T M, Bruce, K D, Cannon, P F, Hall, G S, Mowson, G, Jones, C G, Kampichler, C, Kandeler, E, and Ritchie, D (1998) Impacts of Rising Atmospheric Carbon Dioxide on Model Terrestrial Ecosystems, *Science*, 280, 441–443.

Jordan, D B and Ogren, W L (1984) The CO_2/O_2 Specificity of Ribulose-1,5-Bisphosphate Concentration, pH and Temperature, *Planta*, 161, 308–313.

Keeling, C D, Chin, J F S, and Whorf, T P (1996) Increased Activity of Northern Vegetation Inferred from Atmospheric CO_2 Measurements, *Nature*, 382, 146–149.

Kellogg, E A, Farnsworth, E J, Russo, E T, and Bazzaz, F A (1999) Growth Responses of C_4 Grasses of Contrasting Origin to Elevated CO_2, *Ann. Bot.*, 84, 279–288.

Kerstiens, G (1998) Shade-Tolerance as a Predictor of Responses to Elevated CO_2 in Trees, *Physiol. Plant.*, 102, 472–480.

Koch, G W and Mooney, H A, eds (1996) *Carbon Dioxide and Terrestrial Ecosystems*, Academic Press, San Diego, CA.

Korner, C and Bazzaz, F A, eds (1996) *Carbon Dioxide, Populations, and Communities*, Academic Press, San Diego, CA.

Kubiske, M E and Pregitzer, K S (1996) Effects of Elevated CO_2 and Light Availability on the Photosynthetic Light Response of Trees of Contrasting Shade Tolerance, *Tree Physiol.*, 16, 351–358.

Kubiske, M E and Pregitzer, K S (1997) Ecophysiological Responses to Simulated Canopy Gaps of Two Tree Species of Contrasting Shade Tolerance in Elevated CO_2, *Func. Ecol.*, 11, 24–32.

LeCain, D R and Morgan, J A (1998) Growth, Gas Exchange, Leaf Nitrogen and Carbohydrate Concentrations in NAD-ME and NADP-ME C_4 Grasses Grown in Elevated CO_2, *Physiol. Plant.*, 102, 297–306.

Lewis, J D, Olszyk, D, and Tingey, D T (1999) Seasonal Patterns of Photosynthetic Light Response in Douglas-Fir Seedlings Subjected to Elevated Atmospheric CO_2 and Temperature, *Tree Physiol.*, 19, 243–252.

Lindroth, R L, Kinney, K K, and Platz, C L (1993) Responses of Forest Trees to Elevated Atmospheric CO_2: Productivity, Phytochemistry, and Insect Performance, *Ecology*, 74, 763–777.

Lloyd, J (1999) The CO_2 Dependence of Photosynthesis, Plant Growth Responses to Elevated Atmospheric CO_2 Concentrations and their Interaction with Soil Nutrient Status: II. Temperate and Boreal Forest Productivity and the Combined Effects of Increasing CO_2 Concentrations and Increased Nitrogen Deposition at a Global Scale, *Func. Ecol.*, 13, 439–459.

Lloyd, J and Farquhar, G D (1996) The CO_2 Dependence of Photosynthesis, Plant Growth Responses to Elevated Atmospheric CO_2 Concentrations and their Interaction with Soil Nutrients: I. General Principles and Forest Ecosystems, *Func. Ecol.*, 10, 4–32.

Long, S P (1991) Modification of the Response of Photosynthetic Productivity to Rising Temperature by Atmospheric CO_2 Concentrations: Has its Importance been Underestimated? *Plant, Cell Environ.*, 14, 729–739.

Lorimer, G H (1981) The Carboxylation and Oxygenation of Ribulose-1,5-Bisphosphosphate: The Primary Events in Photosynthesis and Photorespiration, *Annu. Rev. Plant Physiol.*, 32, 349–383.

Melillo, J M, Aber, J D, and Muratore, J F (1982) Nitrogen and Lignin Control of Hardwood Leaf Litter Decomposition Dynamics, *Ecology*, 16, 621–626.

Mooney, H A (1996) Ecosystem Physiology, in *Global Change and Terrestrial Ecosystems*, eds B Walker and W Steffen, Cambridge University Press, Cambridge, 13–19.

Moore, B D, Cheng, S-H, Sims, D, and Seemann, J R (1999) The Biochemical and Molecular Basis for Photosynthetic Acclimation to Elevated Atmospheric CO_2, *Plant, Cell Environ.*, 22, 567–582.

Morison, J I L (1993) Response of Plants to CO_2 under Water Limited Conditions, *Vegetatio*, 104–105, 193–209.

Morison, J I L (1998) Stomatal Response to Increased CO_2 Concentration, *J. Exp. Bot.*, 49, 443–452.

Morison, J I L and Lawlor, D W (1999) Interactions between Increasing CO_2 Concentration and Temperature on Plant Growth, *Plant Cell Environ.*, 22, 659–682.

Morse, S R, Wayne, P, Miao, S L, and Bazzaz, F A (1993) Elevated CO_2 and Drought Alter Tissue Water Relations of Birch (*Betula populifolia*) Seedlings, *Oecologia*, 95, 599–602.

Mott, K A (1990) Sensing of Atmospheric CO_2 by Plants, *Plant, Cell Environ.*, 13, 731–737.

Nadelhoffer, K J, Emmett, B A, Gundersen, P, Kj naas, O J, Koopmans, C J, Schleppi, P, Tietema, A, and Wright, R F (1999) Nitrogen Deposition Makes a Minor Contribution to Carbon Sequestration in Temperate Forests, *Nature*, 389, 145–148.

Nie, G Y, Hendrix, D L, Webber, A N, Kimball, B A, and Long, S P (1995) Increased Accumulation of Carbohydrates and Decreased Photosynthetic Gene Transcript Levels in Wheat Grown at Elevated CO_2 in the Field, *Plant Physiol.*, 108, 975–983.

Norby, R J (1994) Issues and Perspectives for Investigating Root Responses to Elevated Atmospheric Carbon Dioxide, *Plant- Soil*, 165, 9–20.

Norby, R J (1998) Nitrogen Deposition: a Component of Global Change Analyses, *New Phytol.*, 139, 189–200.

Norby, R J and Cotrufo, M F (1998) A Question of Litter Quality, *Nature*, 396, 17–18.

Norby, R J, Gunderson, C A, Wullschleger, S D, O'Neill, E G, and McCracken, M K (1992) Productivity and Compensatory Responses of Yellow-poplar Trees in Elevated CO_2, *Nature*, 357, 322–324.

Norby, R J, Wullschleger, S D, Gunderson, C A, Johnson, D W, and Ceulemans, R (1999) Tree Responses to Rising CO_2 in Field Experiments: Implications for the Future Forest, *Plant, Cell Environ.*, 22, 683–714.

O'Neill, E G (1994) Responses of Soil Biota to Elevated Atmospheric Carbon Dioxide, *Plant- Soil*, 165, 55–65.

O'Neill, E G and Norby, R J (1996) Litter Quality and Decomposition Rates of Foliar Litter Produced under CO_2 Enrichment,

in *Carbon Dioxide and terrestrial Ecosystems*, eds G W Koch and H A Mooney, Academic Press, San Diego, CA, 87–103.

Pearcy, R W and Bjorkman, O (1983) Physiological Effects, in *CO_2 and Plants*, ed E R Lemon, Westview Press, Boulder, CO, 65–105.

Pettersson, R and McDonald, A J S (1994) Effects of Nitrogen Supply on the Acclimation of Photosynthesis to Elevated CO_2, *Photosynth. Res.*, 39, 389–400.

Pettersson, R, McDonald, A J S, and Stadenberg, I (1993) Response of Small Birch Plants (*Betula pendula* Roth.) to Elevated CO_2 and Nutrient Supply, *Plant Cell Environ.*, 16, 1115–1121.

Polley, H W, Johnson, H B, Mayeux, H S, Tischler, C R, and Brown, D A (1996) Carbon Dioxide Enrichment Improves Growth, Water Relations and Survival of Droughted Honey Mesquite (*Prosopis glandulosa*) Seedlings, *Tree Physiol.*, 16, 817–823.

Poorter, H (1998) Do Slow-growing Species and Nutrient-stressed Plants Respond Relatively Strongly to Elevated CO_2? *Global Change Biol.*, 4, 693–697.

Poorter, H, Roumet, C, and Campbell, B D (1996) Interspecific Variation in the Growth Response of Plants to Elevated CO_2: A Search for Functional Types, in *Carbon Dioxide, Populations, and Communities*, eds C Korner and F A Bazzaz, Academic Press, San Diego, CA, 375–412.

Portis, A R (1995) The Regulation of Rubisco by Rubisco Activase, *J. Exp. Bot.*, 46, 1285–1291.

Preiss, J (1982) Regulation of the Biosynthesis and Degradation of Starch, *Annu. Rev. Plant Physiol.*, 33, 431–454.

Price, M V and Waser, N M (1998) Effects of Experimental Warming on Plant Reproductive Phenology in a Subalpine Meadow, *Ecology*, 79, 1261–1271.

Reekie, E G (1996) The Effect of Elevated CO_2 on Developmental Processes and its Implications for Plant–Plant Interactions, in *Carbon Dioxide, Populations, and Communities*, eds C Korner and F A Bazzaz, Academic Press, San Diego, CA, 333–346.

Reekie, E G and Bazzaz, F A (1989) Competition and Patterns of Resource use Among Seedlings of Five Tropical Trees Grown at Ambient and Elevated CO_2, *Oecologia*, 79, 212–222.

Reekie, E G and Bazzaz, F A (1991) Phenology and Growth in Four Annual Species Grown in Ambient and Elevated CO_2, *Can. J. Bot.*, 69, 2475–2481.

Repo, T, Hanninen, H, and Kellomaki, S (1996) The Effects of Long-term Elevation of Air Temperature and CO_2 on the Frost Hardiness of Scots Pine, *Plant Cell Environ.*, 19, 209–216.

Rochefort, L and Bazzaz, F A (1992) Growth Response to Elevated CO_2 in Seedlings of Four Co-occurring Birch Species, *Can. J. For. Res.*, 22, 1583–1587.

Rogers, H H, Runion, G B, and Krupa, S V (1994) Plant Responses to Atmospheric CO_2 Enrichment with Emphasis on Roots and the Rhizosphere, *Plant Cell Environ.*, 83, 155–189.

Rusterholz, H P and Erhardt, A (1998) Effects of Elevated CO_2 on Flowering Phenology and Nectar Production of Nectar Plants Important For Butter ies of Calcareous Grasslands, *Oecologia*, 113, 341–349.

Sage, R F (1994) Acclimation of Photosynthesis to Increasing Atmospheric CO_2: The Gas Exchange Perspective, *Photosynth. Res.*, 39, 351–368.

Sage, R F, Sharkey, T D, and Seemann, J R (1988) The *in vivo* Response of the Ribulose-1,5-Bisphosphate Carboxylase Activation State and the Pool Sizes of Photosynthetic Metabolites to Elevated CO_2 in *Phaseolus vulgaris* L., *Planta*, 174, 407–416.

Saxe, H, Ellsworth, D S, and Heath, J (1998) Tree and Forest Functioning in an Enriched CO_2 Atmosphere, *New Phytol.*, 139, 395–436.

Schimel, D, Alves, D, Enting, I, Heimann, M, Joos, F, Raynaud, D, Wigley, T, Prather, M, Derwent, R, Ehhalt, D, Fraser, P, Sauhueza, E, Zhou, X, Jonas, P, Charlson, R, Rodhe, H, Sadasivan, S, Shine, K P, Fouquart, Y, Ramaswamy, V, Soloman, S, Srinivasan, J, Albritton, D, Derwent, R, Isaksen, I, Lal, M, and Wuebbles, D (1996) Radiative Forcing of Climate change, in *Climate Change 1995: The Science of Climate Change*, eds J T Houghton, L G Meira Filho, B A Callander, N Harris, A Kattenberg, and K Maskell, Cambridge University Press, Cambridge, 65–131.

Schulze, E-D (1989) Air Pollution and Forest Decline in a Spruce (*Picea abies*) Forest, *Science*, 224, 776–783.

Sharkey, T D (1985) Photosynthesis in Intact Leaves of C_3 Plants: Physics, Physiology and Rate Limitation, *Bot. Rev.*, 51, 53–105.

Sharkey, T D (1988) Estimating the Rate of Photorespiration in Leaves, *Physiol. Plant.*, 73, 147–152.

Sheen, J (1994) Feedback Control of Gene Expression, *Photosynth. Res.*, 39, 427–438.

Signora, L, Galtier, N, Skot, L, Lucas, H, and Foyer, C H (1998) Over-Expression of Sucrose Phosphate Synthease in *Arabidopsis thaliana* Results in Increased Foliar Sucrose/Starch Ratios and Favours Decreased Foliar Carbohydrate Accumulation in Plants after Prolonged Growth in CO_2 Enrichment, *J. Exp. Bot.*, 49, 669–680.

Staddon, P L and Fitter, A H (1998) Does Elevated Atmospheric Carbon Dioxide Affect Arbuscular Mycorrhizas? *Trends Ecol. Evol.*, 13, 455–458.

Stirling, C M, Davey, P A, and Williams, T G (1997) Acclimation of Photosynthesis to Elevated CO_2 and Temperature in Five British Native Species of Contrasting Functional type, *Global Change Biol.*, 3, 237–246.

Stitt, M (1991) Rising CO_2 Levels and their Potential Significance for carbon Flow in Photosynthetic cells, *Plant Cell Environ.*, 14, 741–762.

Stitt, M and Krapp, A (1999) The Interaction between Elevated Carbon Dioxide and Nitrogen Nutrition: the Physiological and Molecular Background, *Plant Cell Environ.*, 22, 583–621.

Strain, B R and Bazzaz, F A (1983) Terrestrial Plant Communities, in *CO_2 and Plants: The Response of Plants to Rising Levels of Atmospheric Carbon Dioxide*, ed E R Lemon, Westview Press, Boulder, CO, 177–222.

Thomas, S C and Bazzaz, F A (1996) Elevated CO_2 and Leaf Shape: are Dandelions Getting toothier? *Am. J. Bot.*, 83, 106–111.

Thomas, S C and Jas`enski, M (1996) Genetic Variability and the Nature of Microevolutionary Responses to Elevated CO_2, in *Carbon Dioxide, Populations, and Communities*, eds C Kokrner and F A Bazzaz, Academic Press, San Diego, CA, 51–81.

Thomas, S C, Jasienski, M, and Bazzaz, F A (1999) Early vs. Asymptotic Growth Responses of Herbaceous Plants to Elevated CO_2, *Ecology*, 80, 1552–1567.

Timmermann, A, Oberhuber, J, Bacher, A, Esch, M, Latif, M, and Roeckner, E (1999) Increased El Nino Frequency in a Climate Model Forced by Future Greenhouse Warming, *Nature*, 398, 694–697.

Tissue, D T and Oechel, W C (1987) Response of *Eriophorum vaginatum* to Elevated CO_2 and Temperature in the Alaskan Tussock Tundra, *Ecology*, 68, 401–410.

Traw, M B, Lindroth, R L, and Bazzaz, F A (1996) Decline in Gypsy Moth (*Lymantria dispar*) Performance in an Elevated CO_2 Atmosphere Depends Upon Host Plant Species, *Oecologia*, 118, 113–120.

Tyree, M T and Alexander, J D (1993) Plant Water Relations and the Effects of Elevated CO_2: A Review and Suggestions for Future Research, *Vegetatio*, 104–105, 47–62.

Van Oosten, J-J and Besford, R T (1996) Acclimation of Photosynthesis to Elevated CO_2 Through Feedback Regulation of Gene Expression: Climate of Opinion, *Photosynth. Res.*, 48, 353–365.

Van Oosten, J J, Wilkins, D, and Besford, R T (1994) Regulation of the Expression of Photosynthetic Nuclear Genes by High CO_2 is Mimicked by Carbohydrates: A Mechanism for the Acclimation of Photosynthesis to High CO_2? *Plant Cell Environ.*, 17, 913–923.

Vitousek, P M (1994) Beyond Global Warming: Ecology and Global change, *Ecology*, 75, 1861–1876.

Vitousek, P M, Aber, J D, Howarth, R W, Likens, G E, Matson, P A, Schindler, D W, Schlesinger, W H, and Tilman, D (1997) Human Alteration of the Global Nitrogen Cycle: Sources and Consequences, *Ecol. App.*, 7, 737–750.

Vitousek, P M and Howarth, R W (1991) Nitrogen Limitation on Land and in the Sea: How can it occur? *Biogeochemistry*, 13, 87–115.

Walker, B, Steffen, W, Canadell, J, and Ingram, J, eds (1999) *The terrestrial Biosphere and Global Change: Implications for Natural and Managed Ecosystems*, Cambridge University Press, Cambridge.

Wayne, P M, Reekie, E G, and Bazzaz, F A (1998) Elevated CO_2 Ameliorates Birch Response to High temperature and Frost Stress: Implications for Modeling Climate-induced Geographic Range Shifts, *Oecologia*, 114, 335–342.

Webber, A N, Nie, G-Y, and Long, S P (1994) Acclimation of Photosynthetic Proteins to Rising Atmospheric CO_2, *Photosynth. Res.*, 39, 413–425.

Wigley, T M L, Smith, R L, and Santer, B D (1998) Anthropogenic In uence on the Autocorrelation Structure of Hemispheric-Mean Temperatures, *Science*, 282, 1676–1679.

Wolfe, D W, Gifford, R M, Hilbert, D, and Luo, Y (1998) Integration of Photosynthetic Acclimation to CO_2 at the Whole-plant Level, *Global Change Biol.*, 4, 879–893.

Woodrow, I E and Berry, J A (1988) Enzymatic Regulation of Photosynthetic CO_2 Fixation in C_3 Plants, *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, 39, 533–594.

Woodward, F I and Bazzaz, F A (1988) The Responses of Stomatal Density to CO_2, *J. Exper. Bot.*, 39, 1771–1781.

Woodward, F I and Kelly, C K (1995) The In uence of CO_2 Concentration on Stomatal Density, *New Phytol.*, 131, 311–327.

Woodward, F I, Thompson, G B, and McKee, I F (1991) The Effects of Elevated Concentrations of Carbon Dioxide on Individual Plants, Populations, Communities and Ecosystems, *Ann. Bot.*, 67, S23–S38.

Woodwell, G M, Mackenzie, F T, Houghton, R A, Apps, M, Gorham, E, and Davidson, E (1998) Biotic Feedback in the Warming of the Earth, *Clim. Change*, 40, 495–518.

Wullschleger, S D, Post, W M, and King, A W (1995) On the Potential for a CO_2 Fertilization Effect in Forests: Estimates of the Biotic Growth Factor based on 58 Controlled-Exposure Studies, in *Biotic Feedback in the Global climatic System: Will the Warming Feed the Warming*, eds G M Woodwell and F T Mackenzie, Oxford University Press, New York, 86–107.

Zak, D R, Pregitzer, K S, Curtis, P S, Teeri, J A, Fogel, R, and Randlett, D L (1993) Elevated Atmospheric CO_2 and Feedback Between Carbon and Nitrogen Cycles, *Plant- Soil*, 51, 105–117.

Sedimentary Records of Long-term Ecological Change

JOHN DEARING

University of Liverpool, Liverpool, UK

The geological record has long been a rich source of information about ancient environments. Fossil remains of plant and animal species and the nature of sediments can give a remarkable view of past ecosystems and climate. Such records provided the stimulus for Darwin's theory of evolution and they continue to drive research into evolutionary trends and the causes of extinctions. In the 19th century, geologists interpreted the ancient conditions by making comparisons with modern sediments and processes. But during the 20th century, these so-called palaeorecords have been increasingly used to provide a long time perspective on global environmental change that may actually help us to understand modern environments and to manage future ones. The palaeoecological method stems from an increasing need to understand how human activities and climate may alter processes in the natural environment, especially over periods of time that are too long to have been monitored. For this human dimension, scientists have turned to the soft and unconsolidated deposits of mud and peat, and to ice cores, spanning the past few thousand years. In many parts of the world, these sedimentary archives are rich in information about the effects our ancestors had on the environment, the way in which climate affected vegetation and hydrology and the trends of environmental processes right up to the present day. Other important environmental archives include the annual and seasonal growth increments in tree-rings, coral and speleothems that are providing climate records of high resolution. Without this package of archives we have only a limited timespan of recorded knowledge, often no more than a few decades, about climate, hydrology and even pollution. When used in conjunction with historical documents, archaeological records, instrumental records of climate, and mathematical models, sedimentary archives may provide not only a historical perspective but also a fuller understanding of how the environments around us function and react to impacts.

PALAEOECOLOGICAL METHODS

The strength of palaeoecological research lies in the fact that natural environments are open systems that transport energy across their boundaries in the form of water, nutrients, heat and particles. Particles transported in the air and owing water are trapped in several types of environmental sinks – lakes, reservoirs, peat surfaces, river oodplains, river deltas, etc. (Oldfield, 1977). Thus, peat bogs may preserve the partially decomposed, but identifiable, remains of vegetation growing at the bog surface. Similarly, the microscopic organic remains of aquatic organisms that accumulate at lake beds and river oodplains may show layers of silt brought down by successive oods within the surrounding catchment. Some sediment contents, like microscopic algal fossils, may tell us not just about which species were living in the lake in the past but may be used to infer past lake conditions. For example, some algae (diatoms) are tolerant of highly acid water and their presence in lake mud tells us that the lake water must have been acidic when they were living. This indirect process of reconstruction uses the sediment components as environmental indicators or *proxies* of past conditions (see Figure 1). In some studies, the proxy in modern sediments is calibrated against modern conditions to produce a transfer function that may then be applied to earlier sediments. This approach has allowed diatom counts to be translated into lake pH values and pollen frequencies into landscape types. *See also* **Natural Records of Climate Change**, Volume 1.

The reconstruction of past environments is never perfect. Some outputs from systems are altered before or after they are laid down, some proxies such as airborne pollen grains, are not always easily linked to the place of their origin, and some processes simply leave no direct trace. But the ingenuity of palaeoecologists has led to a surprisingly large number of environmental states and processes that may be reconstructed (Lowe and Walker, 1997). A thimble full of lake sediment can easily contain the

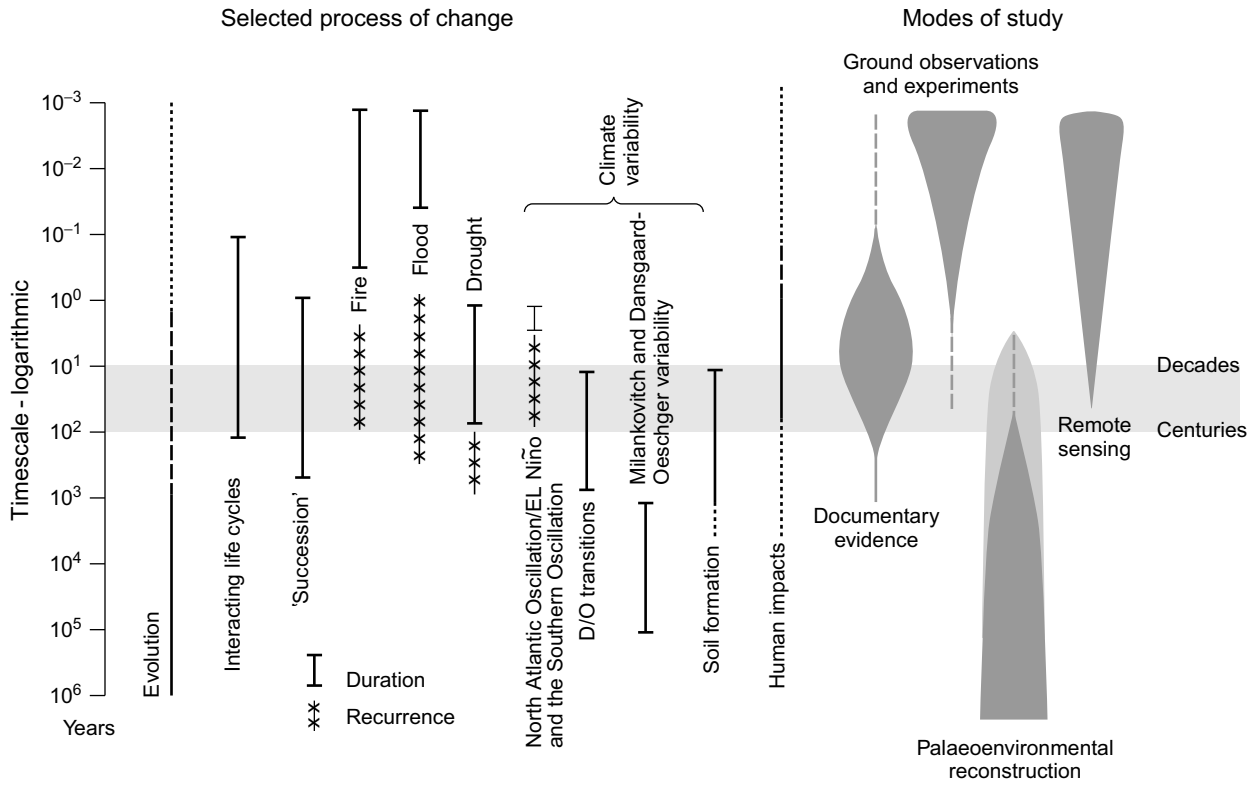


Figure 1 Timescales of environmental change and alternative approaches to their study used by the International Geosphere Biosphere Programme's Past Global Changes (PAGES) core project. Sediments represent a unique means for reconstructing environmental processes and conditions beyond the periods represented by conventional modes of study, monitoring, remote sensing and historical documentation. (Reproduced from Oldfield, 2001)

microfossils linked to more than 100 species of aquatic organisms and terrestrial plants, and the signatures of 20 or so environmental processes like river flooding, eroding soils, volcanic eruptions and industrial pollution (Table 1).

The key to the whole approach is sediment dating. Typically, around 5–15 m of mud have accumulated at the bottom of lakes during the Holocene, the geological period covering the past 11 000 years. A similar thickness of peat has developed in temperate environments. In some lakes the mud is deposited with a seasonal cycle that gives rise to visible layers known as *laminations* which afford a means of dating – one can count back the seasonal layers from the present to find the sediment age for any particular depth. But in the majority of sedimentary archives, dating relies on analytical procedures such as ^{14}C dating of organic remains (c. 40 000–300 years ago), and ^{210}Pb dating that gives sediment ages over the past 100–150 years. Other dating techniques utilize pollution markers like the radioactive fallout from Chernobyl (1986) and early bomb tests (1963), or other markers in the sediment that have a known and often local date, such as pollen linked to the introduction of an exotic plant species.

NATURAL ENVIRONMENTAL VARIABILITY

The first palaeoecologists were particularly interested in reconstructing past vegetation and climate. More than 50 years ago, Johannes Iversen showed through pollen diagrams the major shifts in northern European vegetation through glacial and interglacial periods, and used the climatic range of species such as holly, ivy and mistletoe to calibrate pollen records in terms of annual temperatures (Iversen, 1944). In rain-fed peat bogs and blanket mires, stratigraphic changes in the degree of organic decomposition were interpreted to have been caused by shifts in humidity. Since these early studies, improved methods and investigations at hundreds of sites have produced climate and vegetation records for many regions of the world. Generally, the results show that climate variability in many regions during the past 10 000 years is probably greater than previously believed. There is now a consensus of evidence for a thermal optimum 9000–5500 years before present (BP) that was followed by cooler and, in middle latitudes at least, wetter conditions. Studies of peat stratigraphy, lake sediment laminations and tree-rings also confirm the later fluctuations in climate related to the Medieval Warm Period, the Little Ice Age and

Table 1 Properties of sediments used to reconstruct past ecosystems and environmental processes

Environmental state or process	Example of sediment property, signature or proxy in peat (P), floodplains (F) or lake sediments (L)
Terrestrial ecosystem	
Vegetation type	Pollen (P, F and L) Organic chemistry (P, F and L) Stable isotope chemistry (P and L)
Soil type	Geochemistry (L, F) Mineral magnetism (L, F) Organic chemistry (L, F)
Fire	Charcoal (P, F and L) Mineral magnetism (P, F and L)
Aquatic ecosystem	
Organisms	Diatoms (L) Ostracods (L) Mollusca (L) Plant macrofossils (L)
Lake trophic status	Geochemistry (L) Organic pigments (L) Calcium carbonate (L) Diatom silica (L) Organic chemistry (L)
Lake salinity and acidity	Diatoms (L)
Climate	
Temperature	Pollen (P and L) Coleoptera (L and F) Chironomids (L) Cladocera (L) Stable isotope chemistry (L)
Humidity	Pollen (P) Testate amoebae (P) Humification (P) Stratigraphy (P) Stable isotope chemistry (L)
Hydrology	
Groundwater River discharge	Lake level (L) Sediment layers (L and F) Particle sizes (L and F) Paleochannel forms (F)
Surface water chemistry	Diatoms (L) Geochemistry (L)
Soil erosion	Sediment accumulation rate (L and F) Lamination thickness (L and F)
Volcanic eruptions	Tephra layers (P and L) Mineral magnetism (P and L) Geochemistry (P and L)
Pollution	
Particles	Heavy metals (P and L) Soot/inorganic particles (P and L) Magnetic spheres (P and L) Polycyclic aromatic hydrocarbons (P and L) Radioisotopes (P and L)
Gases (acid rain)	Diatoms algae (L)

global warming since the late 19th century, that have also been recorded in historical documents and instrumental records (Figure 2a) (*see* Little Ice Age, Volume 1; Medieval Climatic Optimum, Volume 1). The effects of climate on vegetation is often seen most strongly across steep environmental gradients, like mountain slopes or the narrow boundary (ecotone) between the grassland and forest zones of North America (*see* Ecotones, Volume 2).

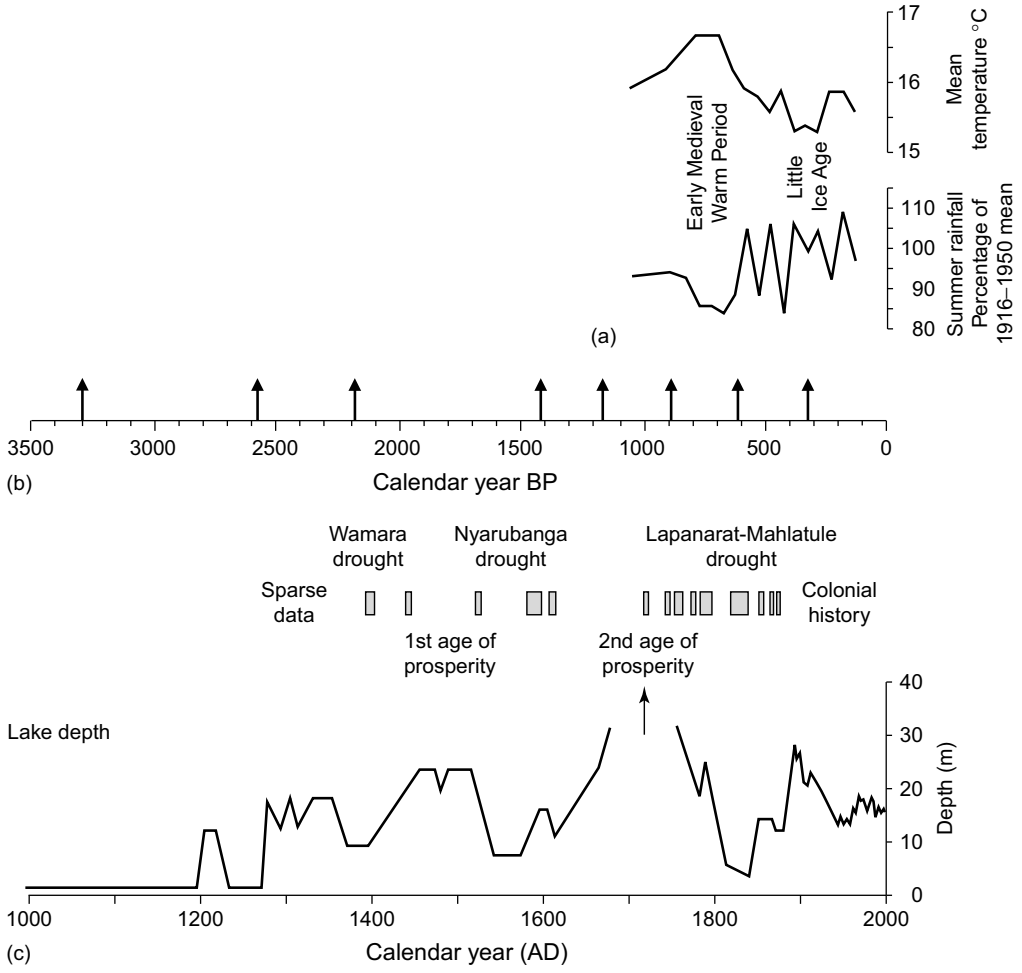


Figure 2 Climate variability recorded in documentary and sedimentary archives: (a) summer rainfall and mean annual temperature in England derived from historical documents in the 900 calibrated (calendar) years BP (Lamb, 1977); (b) mean dates of increased wetness/coldness (arrows) inferred from studies of British and Irish peat in the 3500 cal. year BP (Ellis and Tallis, 2000); (c) reconstructed records of lake water levels in Crescent Island Crater lake, Kenya, for the past 1000 years, where low levels are associated with drought and times of historically recorded poor prosperity and political upheaval (grey bars) (Verschuren *et al.*, 2000)

In these areas, the results from multiple investigations are used to map the changes in the altitude of natural tree-lines or the ebb and ow of ecosystems in response to climate change. In some regions, there have been particularly strong responses to Holocene climate. Lakes that are essentially closed to surface streams are sensitive to climate-controlled groundwater and display evidence for uctuating water levels in the sediments. A recent review of these records in North Africa shows that the monsoonal rains reached a maximum between 9000 and 5000 years ago, and since then the whole region has become progressively drier – showing that recent desertification during the 20th century may be part of a long-term trend. There is strong evidence that arid phases in Kenya caused a worsening agricultural economy and political unrest (Figure 2b). In other regions, the impact of quite modest climate variability has been significant. For example, in the Mississippi catchment, the size and shape of ood sediments and abandoned river channels have been used to estimate the magnitude of extreme oods in the past. The results of James Knox's work shows that the oods were extremely sensitive to small changes in past climate – changes that are actually smaller than those anticipated for the 21st century (Knox, 2000). Fire has had an impact on the large majority of global ecosystems. Counts of charcoal preserved in sediments provide historical records of the frequency and sometimes the intensity of fires in these regions. Some analyses have shown just how common natural fires have been, and in some areas have led to changes in nature conservation policy. Nowadays small wildfires at ground level are encouraged and managed in order to mimic the natural condition and to avoid the devastating effects of large but rare canopy fires.

HUMAN IMPACT ON TERRESTRIAL ECOSYSTEMS

There are few areas on Earth that show no evidence of past human impact on ecosystems. Human impact on vegetation, especially, is recorded widely in pollen diagrams. For example, the cyclical nature of slash and burn methods used in shifting cultivation is recorded as a recognizable sequence of pollen assemblages. The first permanent farmers typically cleared native woodland or scrub, introduced cultivated crops and inadvertently aided the success of weed species, effects that are all recorded. We are able to deduce not only the farming methods but identify the start and subsequent history of the impact in

many parts of the world (Figure 3). For example, the advent of farming in the Near East (>10 000 year BP); the ancient civilizations of the Egyptians, Greeks and Maya; the successive episodes of deforestation stretching across northwest Europe from c. 3500 year BP; the introduction of European agricultural techniques to North America and Australia in the past 250 years; and the recent commercial logging of tropical forests. Recently, the use of pollen data has been extended to recreate aspects of past vegetation cover and land use. For example, in Denmark calibration of modern pollen assemblages to landscapes shows that the major features of land cover patterns have been in existence for more than 3000 years. When set within an archaeological context or when compared with complementary information about climate, these records can help us to understand the links between human activities and climate, and even the feedback on subsequent farming generations. There are several well-established examples of poor farming practices or deteriorating climate (often towards a drier climate) leading to unsustainable agriculture, social unrest, war or migration (Redman, 1999).

Palaeorecords also help us to understand the reasons behind the loss of certain ecosystems and changes in biodiversity. In doing so, we are able to place present trends of species loss in a long time perspective. It is clear from pollen evidence that, perhaps contrary to instinct, human impact often causes biodiversity to increase due to the creation of new habitats. We also now know that many of our wildernesses are in fact the product of human activities or climate, and sometimes a combination of the two. For example, the upland heath ecosystems of northwest Europe are at least 3000 years old, but are the result of the loss of native woodland. As another example palaeoecological evidence from once-thought examples of pristine montane forest in east Africa shows that these environments have often been disturbed over the past 2000 years. We can be sure that future debates about how we manage wildernesses, biodiversity hotspots and National Parks will involve an interrogation of their histories to a far greater extent than before.

HUMAN IMPACT ON SOILS

Farming destabilizes soil structure and may alter soil fertility. Flood layers, increased sedimentation rates and changes in water quality (Table 1) all provide evidence for accelerated soil erosion and nutrient losses, and some generalizations about long-term trends are now possible (Dearing, 1994). What becomes very clear is the close link between erosion and human habitation (Figure 4). For pre-European North America, erosion values are low or declining under limited human impact. The rise and fall of the Classic Maya Civilization in Guatemala is mirrored in the shape of the erosion curve and is also linked to dramatic shifts in the loss of nutrients, especially phosphorus. Erosion values in lowland Europe between c. 10 000 and 3500 years ago are also low, but from c. 3500 BP they tend to show a step-wise pattern of increasing soil losses, where each step is often linked to technological innovation or inward migration. If we focus on a particular step, erosion rates initially rise dramatically but then often decline to a new balanced state, but at a higher level than before the whole transition in

erosion rates typically taking several decades longer than the period of land use change.

Taking the long-term view suggests that in the absence of soil conservation measures, modern erosion rates in some parts of North Europe and North America may be some of the highest recorded. The rates may approach those calculated for the same landscapes at the end of the last Ice Age when there was a bare tundra environment. More than 10-fold increases in erosion over the figure for undisturbed environments are not uncommon. A few studies in Europe, Australia and the US have compared the lake records with recent monitored records of river sediment discharge and source fingerprints in the lake catchment. The results suggest that the eroded sediment found in oodplains and lake muds may come from both river channels and field surfaces. Farming may have had as much impact on runoff, ooding and river channel erosion as it has had on soil erosion. By comparing erosion records from lakes with catchments of different sizes, we can begin to explore the dependence of erosion responses on a spatial scale. It seems that deforestation in small catchments causes not just the largest erosion rates per unit area but also a much larger increase of erosion over the pre-disturbance figure. These findings have clear implications for the impact of modern forest management on reservoir sedimentation, the sustainability of oodplains and on sediment concentrations in coastal ecosystems. What we can also say with some certainty, is that rising curves of erosion associated with some intensive Western agriculture and tropical farming practises suggest strongly that modern soil losses are often unsustainable.

AIR AND WATER POLLUTION

There are now numerous sedimentary records that have helped identify the magnitude, timing, source and causes of pollution. Atmospheric pollution is not new, as demonstrated by analyses of lead (and its isotopes) trapped in cores of glacier ice from Greenland. The results provide a Northern Hemisphere record of atmospheric lead deposition from the first significant rise during the Roman Empire to the switch from leaded to non-leaded petrol in the late 20th century (Boutron *et al.*, 1991). But the majority of pollution studies have focused on pollution in the last two centuries, particularly the causes and rates of change up to the present. The impact of pollution on water quality has been studied using the assemblages of diatom algae, preserved in lake sediments, that are particularly responsive to acidity and dissolved nutrients. Through transfer functions, they allow historical changes in pH and nutrients to be quantified. In the 1980s, one of the hottest environmental debates was the cause of lake acidification, a process that was linked to losses of freshwater fish stocks, rapid changes in aquatic ecosystem communities and increased concentrations of dissolved metals, such as aluminum. Rick Battarbee and his team used diatom transfer functions to show that lakes in Galloway, southwest Scotland had acidified only from the late 19th or early 20th century (Battarbee, 1985), irrespective of the type of vegetation cover in the catchment (Figure 5). Additional evidence from geochemical analyses of the sediments showed that the start of acidification usually coincided with the first significant deposition of heavy metals, such as copper and lead, and soot particles – all indicative of industrial pollution sources. These studies and others in Scandinavia, northwest Europe and North America quickly provided a set of compelling arguments to refute claims that long-term natural processes of acidification or certain types of land use had caused acidification. In fact, some reconstructions showed that far from air pollution accelerating a natural acidification process, long-term natural trends in lake pH were towards increasing alkalinity. Air-borne pollutants often transported long distances from major centres of industrialization, probably aided by the construction of tall chimneys, were clearly implicated as the main causes of lake acidification in many parts of the world. The results had a significant impact on government policy and helped accelerate the reductions in pollution emissions, particularly from large coal-burning power generation plants.

At many sites world-wide, there is clear evidence that major increases in aquatic biomass and shifts in ecological communities in lakes, rivers and coastal zones have been responses to the release of soil nutrients following deforestation or the increased ow of nutrients from fertilizers, sewage and urban runoff. Eutrophication of lakes by nutrients is, in contrast to lake acidification, more often linked to local catchment conditions, but may also cause damage to fisheries and even constitute a health hazard to recreational users and farm animals. Some lake sediment studies have now succeeded in reconstructing not just the impact on aquatic organisms but also the amount of nutrients, particularly phosphorus, that have owed into the lake. The ability to calculate past loading of acidity and nutrients has allowed management programs to estimate the critical loads to a particular site that should not be exceeded.

FROM THE PAST TO THE FUTURE

Over the last decade, the idea of a rapidly changing global environment has not only become the major focus of environmental research, but has also become a permanent part of the public consciousness. Concepts of new weather regimes, sustainable agriculture, and the causes of natural hazards are now commonly discussed by the media. The questions that are now being asked by society will challenge scientists to inform about the future to a far more accurate degree. How will landscapes respond to new weather conditions? Will extreme ood events become more frequent? If we have the opportunity to adopt alternative agricultural

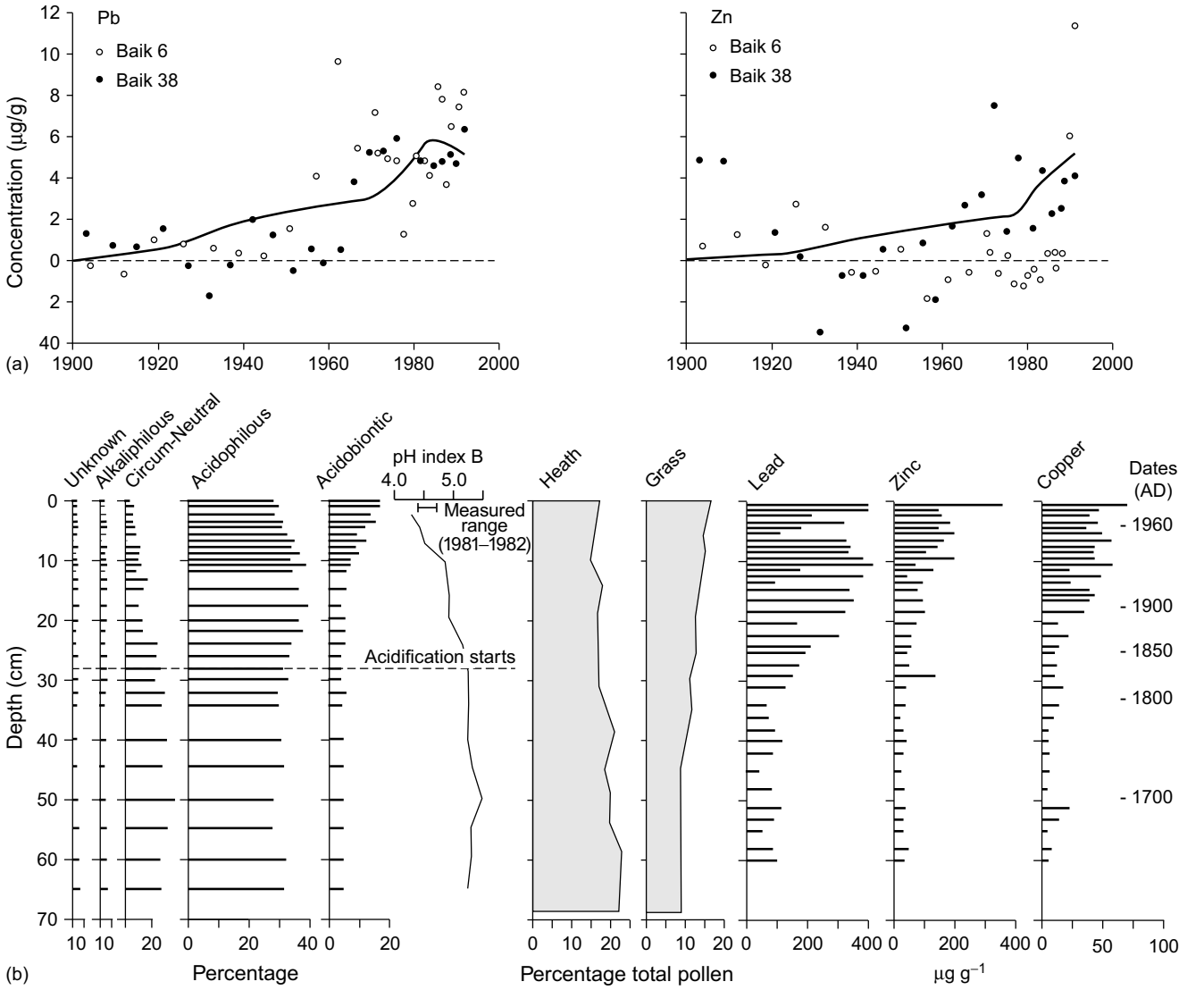


Figure 5 Recent records of pollution derived from lake sediments: (a) reconstructed records in two sediment cores (Baik 6 and 38) of lead (Pb) and zinc (Zn) supply to the southern basin of Lake Baikal, Siberia, with trend line showing evidence for enrichment from local industrial sources after 1940–1950 and again after c. 1970 (Boyle *et al.*, 1998) (*see* **Lake Baikal**, Volume 3); (b) diatom-based reconstruction of lake pH (solid line) in northwest Scotland showing initial decline in pH beginning after 1840 caused by atmospheric emissions, supported by pollen evidence for unchanging heathland species, and correlations with enhanced sediment concentrations of atmospherically-derived heavy metals. The labels *unknown*, *alkaliphilous*, *circum-neutral*, *acidophilous* and *acidobiontic* describe pH classes of diatoms (Battarbee *et al.*, 1985)

responses, how do we judge which is the most appropriate? If we do invest in pollution control, how long will it take to see the benefits? The previous sections have provided a glimpse of the wealth of information that now exists about past environments and the implications for their future, but much more can be learned from the past using a number of alternative approaches. One approach will be to use palaeorecords to test the accuracy of predictive models. Mathematical models that predict the effects of pollution controls on long-term acidification of surface waters and the soil erosion response to climate have already been retrospectively tested using lake sediment records. The models can be modified until they match the historical records, and then be used with much more confidence to project future conditions, or scenarios. For surface water acidification, the MAGIC model (Battarbee, 1994) accurately simulates the acidification that has taken place since the 1850s and provides predictions for future lake pH status under different levels of sulfur emissions. We may increasingly wish to identify systems whose behavior is very difficult to predict, and some new work in this area is based on the idea of comparing palaeorecords with theoretical views

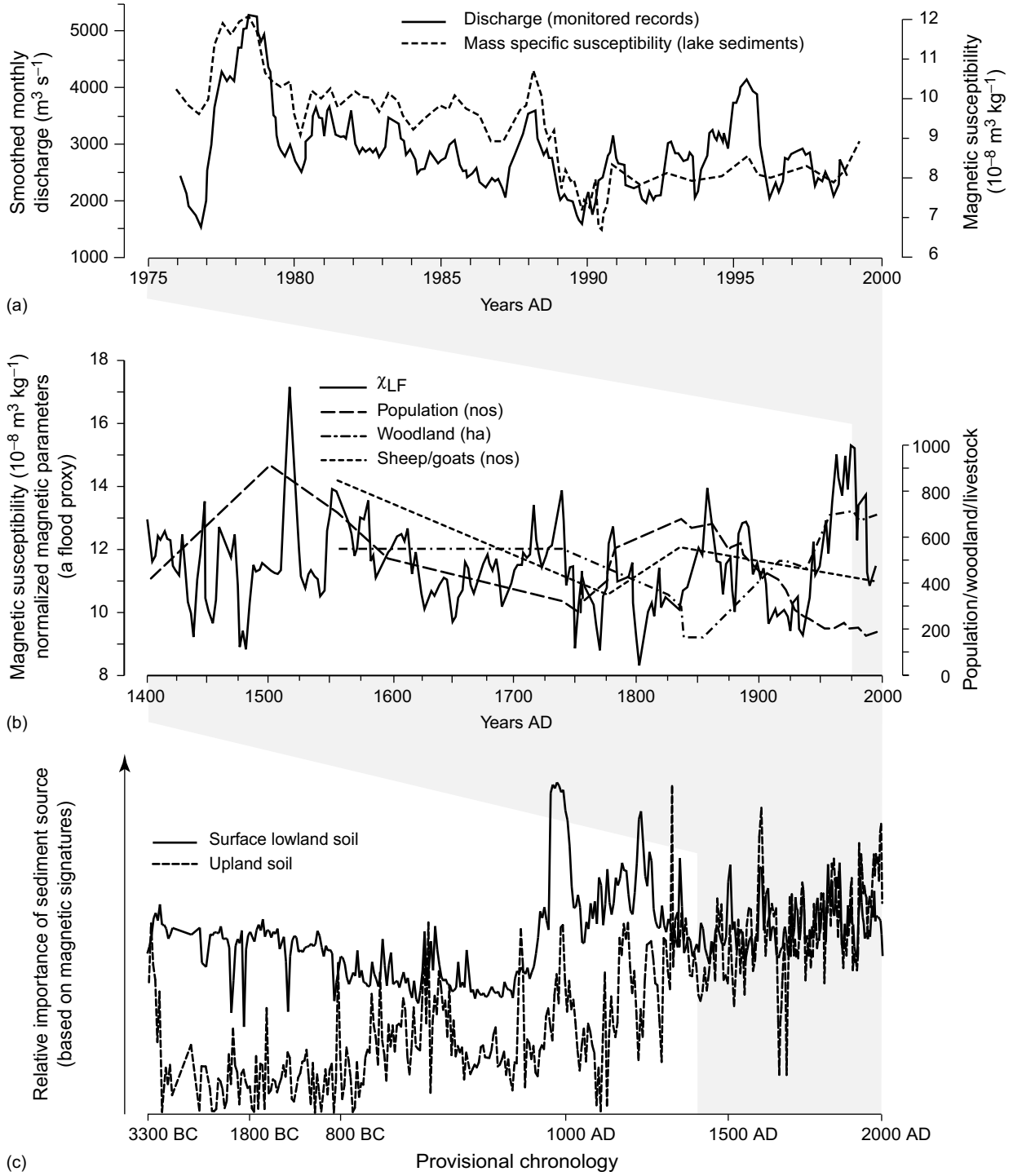


Figure 6 Reconstructing the interactions between human activities and climate on modern and historical flooding at Lac d' Annecy, French Alps, based on a combination of lake sediments, monitored process data and historical documents (author's unpublished data): (a) the magnetic properties of lake sediment deposited 1970–1998 are broadly correlated with monitored records of monthly river discharge, giving a magnetic proxy for flooding in older sediments; (b) the magnetic flood proxy for sediment spanning the past 600 years plotted against historical records of land use and human population showing complex associations between land use change and flooding; (c) magnetic signatures for the transport of surface lowland soil and upland soil from the catchment to the lake since c. 5500 BP showing that major changes, which occurred c. 1000 years ago in response to deforestation, may still exert a significant control on the progressive erosion of upland soil and the modern flood responses to climate and human impact

on how systems in general change. Theories about chaotic behavior and self-organized systems imply that some features of today's environment are not directly linked to easily definable external forces or may be caused by the internal organization of the system (*see* Monitoring in Support of Policy: an Adaptive Ecosystem Approach, Volume 4). Thus, the occurrence of a catastrophic landslide may be part of a long-term underlying system property, and not simply a consequence of a certain set of weather and land use conditions that we can measure. So another approach is to use palaeorecords in conjunction with historical documents and instrumental records in order to learn about the long term relationships between human actions, climate and environmental processes. One obstacle in this kind of study is the tightly woven roles of climate and human activities in controlling the structure and functioning of environmental systems. A recent example from Lac d'Annecy in the French Alps (Figure 6) shows that while modern ood events obviously re ect extreme rainfall events, their magnitude has changed in complex ways partly as a result of land use. In particular, the study traces the modern sensitivity of ood-water generation, erosion, and, possibly, avalanches in the high montane zone back to high altitude forest clearance and grazing that started 1000 years ago. This study, and others, suggest that the way in which present and future environments will respond to, or recover from, a changing climate or a new land use is at least partly conditioned by their history. An extension of this approach is to find historical examples, or analogues, of present day environmental change and observe the responses in the system that have already occurred. As we have seen, the Mayan populations cleared large tracts of tropical forest and caused significant erosion. The lake sediment record showed that erosion levels did eventually decline, but over several centuries. We may propose that erosion under today's logged forest will also take a similar timescale to recover – assuming that, as in the past, clearance and farming virtually cease.

Timescales of natural change and the speed of environmental responses and recoveries to impacts vary with different landscapes and with different environmental processes. The length of this history can vary from a few years to millennia. Managing our global environments so that the worst, destructive aspects of change are limited remains a major challenge, and one that will be increasingly dependent upon the reconstruction and detailed analysis of past ecosystems.

ACKNOWLEDGMENTS

Thanks to Darren Crook, Gez Foster, Richard Jones, Frank Oldfield and David Siddle for permission to show their data in Figures 1 and 6.

REFERENCES

Alverson, K, Oldfield, F, and Bradley, R (2000) Past Global Changes and Their Significance for the Future, *Quatern. Sci. Rev.*, 19, 1–5.

Battarbee, R W, Flower, R J, Stevenson, A C, and Rippey, B (1985) Lake Acidification in Galloway: a Palaeoecological Test of Competing Hypotheses, *Nature*, 314, 350–352.

Battarbee, R W (1994) Surface Water Acidification, in *The Changing Global Environment*, ed N Roberts, Blackwell, Oxford.

Berglund, B E (1991) The Cultural Landscape During 6000 Years in Southern Sweden, *Ecol. Bull.*, 41, 495.

Berglund, B E (1994) Methods for Quantifying Prehistoric Deforestation, *Paläeoklimaforschung*, 12, 5–11.

Boutron, C F, Gorlach, U Candelone, J-P, Bolshov, M A, and Delmas, R J (1991) Decrease in anthropogenic lead, cadmium and zinc in Greenland snows since the late 1960s, *Nature*, 353, 153–156.

Boyle, J F, Mackay, A W, Rose, N L, Flower, R J, and Appleby, P G (1998) Sediment Heavy Metal Record in Lake Baikal: Natural and Anthropogenic Sources, *J. Paleolimnol.*, 20, 135–150.

Dearing, J A (1994) Reconstructing the History of Soil Erosion, in *The Changing Global Environment*, ed N Roberts, Blackwell, Oxford, 242–261.

Ellis, C J and Tallis, J H (2000) Climatic Control of Blanket Mire Development at Kentra Moss, North-west Scotland, *J. Ecol.*, 88, 869–889.

Inversen, J (1944) Viscum, Hedera, and Ilex as climate indicators. A contribution to the study of the post-glacial temperate climate, *Geol. Föreningens I Stockholm Förhandlingar*, 66, 463–483.

Knox, J C (2000) Sensitivity of Modern and Holocene Floods to Climate Change, *Quatern. Sci. Rev.*, 19, 439–457.

Lamb, H H (1965) The Early Medieval Warm Epoch and its Sequel, *Palaeogeogr., Palaeoclimatol., Palaeoecol.*, 1, 13–37.

Lamb, H H (1977) *Climate: Past, Present and Future, Vol. 2, Climatic History and Future*, Methuen, London.

Lowe, J J and Walker, M J C (1997) *Reconstructing Quaternary Environments*, 2nd edition, Longman, Harlow.

Oldfield, F (1977) Lakes and their Drainage Basins as Units of Sediment-based Ecological Study, *Prog. Phys. Geogr.*, 1, 460–504.

Oldfield, F (2001) http://www.pages.unibe.ch/.

Oldfield, F and Dearing, J A (2001) *The Role of Human Activities in Past Environmental Change, International Geosphere-Biosphere PAGES Synthesis*, Springer-Verlag, New York.

Redman, C L (1999) *Human Impact on Ancient Environments*, The University of Arizona Press, Tuscon, AZ.

Roberts, N (1998) *The Holocene. An Environmental History*, 2nd edition, Blackwell, Oxford.

Verschuren, D, Laiord, K R, and Cumming, B F (2000) Rainfall and Drought in Equatorial East Africa During the Past 1000 years, *Nature*, 403, 410–414.

Terrestrial and Freshwater Ecosystems: Impacts of Global Change

BRUCE A HUNGATE AND JANE C MARKS

Northern Arizona University, Flagstaff, AZ, USA

Humans are altering the environment in many ways. While local environmental damage (land lls, oil spills, urban smog) is still prevalent, we now also realize that human activities are altering the Earth System as a whole, that our environmental crisis has become truly global. For example:

- *Our growing use of automobiles, air conditioners, jet planes, and other amenities of modern industrial society has caused the composition and chemistry of the atmosphere to change.*
- *The increasing concentration of heat-trapping gases in the atmosphere (carbon dioxide (CO_2), nitrous oxide (N_2O), methane (CH_4) and others) is already increasing global temperatures, and the rate of change is likely to increase in the future.*
- *The increasing concentration of ozone-depleting gases (chloro uorocarbons and nitrous oxide) in the stratosphere is weakening the ozone shield, allowing more and more damaging ultraviolet (UV-B) radiation to reach the Earth s surface.*
- *Growing food demand, greater use of irrigation, fertilizer and pesticides in agriculture, and the expansion of food production into previously uncultivated lands has increased the inputs of sediments, nutrients, and other pollutants into watersheds and the inputs of greenhouse and chemically active gases into the atmosphere.*
- *Of all the available freshwater on the Earth, humans now appropriate 23% for industrial, municipal, and agricultural uses, thereby altering natural hydrologic regimes and threatening freshwater ecosystems and the services those ecosystems provide.*
- *The spread of exotic species into terrestrial and freshwater ecosystems poses a growing threat to biodiversity conservation and to the services that ecosystems provide to humans.*

Such changes in the Earth s environment are collectively referred to as global change. The nature and magnitude of global change has raised the question, how does global change affect the ecosystems upon which we depend? Ecosystems supply humanity s food, clean water, and clean air and also are important sources of recreation and aesthetic value. Global changes threaten the ability of ecosystems to continue to provide these services. Our understanding of these threats will provide a foundation for determining the in uence of our activities on the Earth System: and for future decisions we must change our lifestyles and use of resources. This article provides an overview of the effects of global change on terrestrial and freshwater ecosystems in coming decades. Many such effects have been postulated or observed in manipulative experiments, but one of the more serious ecological concerns is the possibility of marked shifts in plant species composition due to differing responses of individual members of ecosystems to major changes in carbon dioxide concentrations, temperature, nitrogen emissions and other elements of the global environment.

ECOSYSTEM PROCESSES AND EFFECTS OF GLOBAL CHANGE: AN OVERVIEW

Considering the effects of global change from an ecosystem perspective warrants an overview of the terms and concepts of *ecosystem* ecology. An ecosystem comprises all the organisms in an area and the physical environment with which they interact. *Ecosystem structure* describes the spatial distribution of organisms, elements, and materials in ecosystems, including, for example, above- and below-ground biomass of plant species, soil carbon and nutrient mass, leaf area index in terrestrial ecosystems, or total

biomass of macrophytes and epiphytes and nutrient stores in freshwater ecosystems. By contrast, *ecosystem processes* are the uxes of materials and energy among ecosystem components and between ecosystems and the surrounding environment. For example, rates of photosynthesis and respiration, predation and herbivory, and nutrient turnover and loss are all ecosystem processes (*see* **Ecosystem Structure and Function**, Volume 2).

Ecosystems are not static; rather, disturbances, such as fires or insect outbreaks in forests, or scouring oods in rivers, occur in all ecosystems. *Disturbance* in ecology is defined as the catastrophic removal of biomass that creates space available for recolonization. Succession follows disturbance and involves species progressively recolonizing the disturbed area and being excluded from it (primarily through competition for resources by other species), resulting in a pattern of species replacements and changing species composition through time. Global change can alter the frequency and intensity of disturbances (increasing ood frequency and intensity, for example), but few global changes (excepting perhaps some land-use changes) constitute ecological disturbances in the strict sense. In ecological parlance, global changes are anthropogenic perturbations to ecosystems, human-caused changes in the conditions (temperature, pH, etc.) or resources (nutrients, carbon dioxide, light, water, etc.) that in uence ecosystem structure and processes.

Central to ecosystem ecology is the cycling of biologically essential elements, beginning with carbon. Carbon enters the biosphere through the process of photosynthesis (primary production) by plants and algae, converting atmospheric carbon dioxide to organic forms of carbon. This organic carbon resides in ecosystem compartments, both living (living tissue of plants, animals, and microorganisms) and non-living (as detritus and humus in terrestrial ecosystems and as dissolved or particulate organic matter in aquatic ecosystems). Within ecosystems, carbon is transferred between compartments through trophic interactions (herbivory, predation), senescence, and death. Through respiration by plants, animals, and particularly decomposer microorganisms, organic carbon is returned to the atmosphere as carbon dioxide. *Net primary production* (NPP) is the amount of carbon dioxide taken up by plants through photosynthesis minus respiration by plants, whereas *net ecosystem production* (NEP) is the amount of carbon dioxide taken up by plants through photosynthesis minus the respiration of all organisms in the ecosystem. Ecosystem carbon balance is equivalent to NEP, except when disturbances cause carbon losses by means other than decomposition (fire, for example, which converts organic carbon to carbon dioxide through the process of combustion). As described in detail below, global change can affect all of these important ecosystem processes; carbon gain through photosynthesis, loss through decomposition, and rates of transfers among ecosystem compartments. Changes in these process rates can alter components of ecosystem structure that depend directly on carbon cycling, such as stores of carbon in organic matter and vegetation.

Transfers of carbon dioxide between ecosystems and the atmosphere through photosynthesis and respiration are large enough, for example, that during the Northern Hemisphere summer, the seasonal peak in photosynthesis reduces the temperate zone carbon dioxide concentration by several parts per million, whereas when respiration becomes a relatively larger ux in the winter, atmospheric carbon dioxide concentration increases. These oscillations are mirrored (but more damped because of the smaller land masses) in the Southern Hemisphere, and are much less apparent in tropical regions, where the activities of photosynthesis and respiration are more synchronous throughout the year. These seasonal changes in atmospheric carbon dioxide concentration re ect the breathing of the biosphere and underscore the importance of biological control over atmospheric carbon dioxide concentrations. Because carbon dioxide is a greenhouse gas (increasing in the atmosphere due to fossil fuel burning, cement manufacture, and deforestation, and the major cause of global warming) there has been tremendous interest in understanding how global changes will alter carbon cycling in terrestrial ecosystems and, ultimately, the balance between carbon dioxide release to the atmosphere versus carbon dioxide uptake and storage in ecosystems. Such changes could strongly in uence the trajectory of rising carbon dioxide and associated global warming over the next few centuries.

In terrestrial plants, carbon uptake through photosynthesis is coupled to *transpiration*, the evaporation of water from the internal surfaces of leaves to the atmosphere and the major ux in the water cycle directly mediated by biota. Together, transpiration and the evaporation of water from exposed surfaces constitute *evapotranspiration*, the total ux of water from ecosystems to the atmosphere (*see also* **PET (Potential Evapotranspiration)**, Volume 2). Water not lost through evapotranspiration can be stored in soil and vegetation or lost as surface runoff to aquatic ecosystems and as percolation to groundwater. Precipitation returns water in the atmosphere to the Earth's surface, completing the water cycle. Climate change can alter water cycling through climate-driven shifts in the amount and distribution of precipitation, subsequently affecting ecosystem processes sensitive to water availability, including evapotranspiration, thereby altering surface runoff and percolation. Global changes such as rising carbon dioxide, warming, nitrogen deposition, and increased UV-B radiation can also alter water cycling through their effects on plant transpiration.

Evapotranspiration is regulated by the difference in water vapor pressure between the atmosphere and air above ecosystems and the resistance to water vapor ux between

the ecosystem and the atmosphere. Evapotranspiration occurs when solar radiation striking an ecosystem drives water from liquid to gas, an energy conversion (called *latent heat flux*, λE) that dissipates solar energy that might otherwise have increased the surface temperature of the ecosystem. *Convection* (the upward transfer of heat, also called *sensible heat flux*, C) is also an important way that heat is transferred from ecosystems to the atmosphere. The transfer can also occur in the other direction, if the surface of the ecosystem is colder than the surrounding atmosphere, as is often the case at night. The sum of sensible and latent heat fluxes, along with conduction into the ground (also termed storage, G), balances the net input of radiation (R_n) into terrestrial ecosystems: $R_n = C + \lambda E + G$. Structural and physiological features of ecosystems, including leaf area, stomatal conductance, canopy roughness, and rooting depth, strongly influence the relative importance of these terms, and thus the way in which energy is partitioned in ecosystems. By altering ecosystem structure or physiology through changes in growth or species composition, global change can alter energy partitioning. Such changes have important implications for ecosystem processes sensitive to temperature and moisture availability, as well as for regional climate patterns.

Global change can also alter the cycling of mineral nutrients (such as nitrogen, phosphorus, and others) upon which plant and ecosystem production depend. Indeed, global changes such as nitrogen deposition involve direct changes in the amounts of nutrients added to ecosystems from the atmosphere. Nutrients are made available to primary producers by decomposer organisms in the processes of *mineralization* (release of mineral nutrients through decomposition of dead organic matter); *weathering* (release of nutrients from parent material through chemical and physical reactions); *biological fixation* in the case of nitrogen (conversion of atmospheric N_2 to NH_3, a form usable by plants and microorganisms), and *atmospheric deposition* (in which readily available forms of nutrients enter ecosystems through rainfall or dry deposition). Nutrients are removed from this available pool through microbial immobilization or uptake by plants, chemical sorption reactions (formation of covalent bonds with mineral particles or precipitation with other minerals into unavailable forms), leaching and biotic and abiotic transformations to gaseous forms.

Gaseous fluxes of nitrogen from ecosystems to the atmosphere represent not only losses of an essential nutrient but also inputs to the atmosphere of reactive (nitric oxide and nitrogen dioxide) and radiatively active (nitrous oxide) trace gases. Fluxes of other trace gases, such as methane (CH_4), carbon monoxide (CO), and carbonyl sulfide (COS), while relatively small compared to carbon dioxide fluxes in photosynthesis and respiration, for example, are critical in the chemistry and radiation balance of the atmosphere. Through changes in trace gas fluxes, the responses of ecosystems to global change create feedbacks that can further amplify or mitigate global changes in the atmosphere.

ASSESSING THE EFFECTS OF GLOBAL CHANGE

The scientific community has taken three complementary approaches to assessing the effects of global change on ecosystems: experiments, observations, and models. (1) *Manipulative experiments* compare a set of test plots experiencing current normal conditions to another set where one aspect of the environment is modified to simulate a particular global change factor (e.g., global warming, nitrogen deposition, or the presence of an invading species). The advantage of the experimental approach is that it explicitly identifies the global changes that cause particular ecosystem responses. More complicated experiments involve manipulating two or more global change factors singly and in combination, allowing investigators to test whether knowledge of responses to each factor alone can predict their combined effects, or whether factors interact in surprising ways. Experiments examining full combinations of many factors can address very complex interactions and are extremely valuable in assessing effects of global change. However, such experiments are large and expensive and thus few and far between. Global change experiments are typically long term (1–10 years); they measure responses of ecosystems to a step change, an instantaneous doubling, for example, of atmospheric carbon dioxide concentrations (from 365 to $730\,\mu l\,l^{-1}$). This approach is necessary to be able to discern changes in ecosystem properties against a background of high natural variation. Actual global changes, however, occur more gradually and over longer periods of time (e.g., $3\,\mu l\,l^{-1}\,year^{-1}$, perhaps not reaching $730\,\mu l\,l^{-1}$ until the year 2100). Results from experiments must therefore be interpreted within a framework that accounts for these differences in time scale.

(2) *Observations* allow the impact of actual global changes on ecosystems to be observed. This may be carried out by reconstructing global changes that occurred in the past and examining evidence of their effects on ecosystems, through comparing ecosystems that occur along natural climatic or other gradients in space, or by monitoring ecosystems in the present as they are increasingly affected by global change. For example, the concentration and isotopic composition of gases trapped in air bubbles within ice near the poles provide records of atmospheric composition and temperature, and shows how the accumulation of greenhouse gases in the atmosphere in the past has coincided with increases in global temperatures. By combining such records of past temperatures and atmospheric composition with analyses of pollen in lake sediments, one can understand how plant species shifted in distribution

when, or shortly after, the climate changed. This approach has the considerable advantage of revealing how species composition in an area (an important component of ecosystem structure) has changed over hundreds to thousands of years in response to changes in climate. The gradient approach essentially substitutes space for time using, for example, a transect across the landscape along which mean annual temperature changes systematically, such that observations comparing the colder and warmer sites serve as an analog for ecosystem responses to global warming. Another approach is to monitor current ecosystems directly, to observe their responses to year-to-year variations in climate and to ongoing global changes. This type of approach is central to efforts to document global changes *per se* (e.g., rising global temperatures, atmospheric carbon dioxide concentrations, sea levels, rates of nitrogen and acid deposition and glacial retreat, etc.) and is also becoming more important for understanding ecosystem responses to these changes. One of the challenges associated with this observation approach is that, because many kinds of global change are occurring simultaneously, it is difficult to discern which particular change is causing a particular ecosystem response (though this can be overcome with the gradient approach, for example, if only one variable changes along the gradient). Another challenge is that while current rates of global change are unprecedented, they are slow for even a long-term (10–50 years) campaign of observations. This makes it difficult to distinguish the signal caused by global change from the noise associated with natural variation in ecosystem processes in space and time.

(3) *Mathematical models* are an essential component of efforts to understand the effects of global change on ecosystems. Models provide access to responses of ecosystems to global change that neither manipulative experiments nor observations allow. For example, a manipulative experiment to explore the interactive effects of warming, altered precipitation, increased nitrogen deposition, increased atmospheric carbon dioxide concentrations, and enhanced UV-B radiation on forest ecosystems in the field would be prohibitively expensive, yet models that incorporate information from single-factor manipulative experiments can explore such interactive effects. Similarly, models allow the study of responses over temporal and spatial scales that are impractical with manipulative experiments or observations. Ecosystem models incorporate state-of-the-art scientific understanding of controls over ecosystem processes, and as such, their simulations and predictions are limited by this understanding. Nevertheless, models are one of the more powerful tools available to global change scientists for taking complex and variable processes and translating them to larger spatial and temporal scales.

Efforts to understand ecosystem responses to global change through experiments, observations, or models shed light on the problem, but limitations are associated with each approach. For this reason, where possible, integrated assessments that take advantage of all three approaches are more powerful than any one approach alone.

ELEVATED ATMOSPHERIC CARBON DIOXIDE

The concentration of CO_2 in the atmosphere has increased by 30% since the mid-1800s and is perhaps the most certain component of atmospheric change. Direct measurements of atmospheric carbon dioxide over the past 50 years show this increase clearly (Keeling and Whorf, 2000), and measurements of the concentrations of carbon dioxide in gas bubbles trapped in ice provide a robust record of carbon dioxide concentrations for the past 420 000 years (Petit *et al.*, 1999). Carbon dioxide is the entry point for carbon into the biosphere because it is the substrate for primary production by plants, algae, and some bacteria. Rising carbon dioxide thus represents an increase in resource availability for primary producers, a change with effects that can cascade through ecosystems.

Effects on Terrestrial Ecosystems (*see also* Plant Growth at Elevated CO_2, Volume 2)

Plants usually respond to elevated carbon dioxide by increasing photosynthesis and growth, decreasing transpiration, and by altering patterns of carbon and nutrient use and allocation, patterns that are most clear in short-term laboratory or greenhouse studies. While these physiological changes often lead to responses at the community and ecosystem levels, feedbacks and environmental constraints strongly shape these responses. Indeed, one of the more important conclusions from the past several decades of research on this topic is that, while the short-term physiological responses of plants to elevated carbon dioxide can sometimes guide our predictions of larger-scale responses, in many cases they are misleading about the magnitude (and even the direction) of these responses at the ecosystem scale. For example, in the field, the photosynthetic response to elevated carbon dioxide and the fate of assimilated carbon are sensitive to temperature, water stress, and nutrient availability, and thus the growth responses of plants to elevated carbon dioxide ranges from none to modest to quite large.

Plants grown in elevated carbon dioxide usually have lower nitrogen concentrations, both because the requirement for rubisco and other enzymes (major nitrogen containing compounds) is reduced in elevated carbon dioxide, and because elevated carbon dioxide often causes accumulation of starch and sugars, diluting the amount of nitrogen they contain. Similarly, plants collected over the past several hundred years and stored in herbaria show a consistent decline in nitrogen concentrations in concert with increasing

atmospheric carbon dioxide since the industrial revolution. Reduced nitrogen concentration in litter usually slows the rate at which that litter decomposes and releases nitrogen in plant-available forms, so the commonly observed reduction in green-leaf nitrogen concentration was expected to reduce decomposition rate and nitrogen availability. This effect is only sometimes observed: litter produced by plants grown in elevated carbon dioxide does not usually decompose more slowly, contrary to previous predictions based on the physiological responses of plants (Norby and Cotrufo, 1998).

However, there are other mechanisms through which elevated carbon dioxide can affect nutrient cycling. By increasing root growth and the ow of labile organic compounds into the soil, elevated carbon dioxide can stimulate microbial nutrient immobilization, thereby exacerbating plant nutrient deficiencies, and decreasing nutrient availability for other microbial processes as well, such as nitrification. By increasing plant growth, elevated carbon dioxide can increase nutrient uptake and nutrient accumulation in biomass and soils, eventually increasing nutrient pools in unavailable, organic forms. On the other hand, elevated carbon dioxide can also increase nutrient availability to plants, both through changes in internal cycling rates and through changes in nutrient inputs and losses. For example, by stimulating microbial turnover, elevated carbon dioxide can increase soil nitrogen mineralization. Elevated carbon dioxide also often favors the growth of plants in association with nitrogen-fixing bacteria, as long as other factors (such as low phosphorus availability) do not limit nitrogen fixation. Nitrogen inputs to ecosystems can therefore increase with enhanced growth of nitrogen-fixing species. Elevated carbon dioxide has been shown to increase plant exudation of phosphatase, enzymes that convert soil phosphorus into a plant available form.

Carbon uptake through photosynthesis ultimately feeds all trophic levels in ecosystems, so it is not surprising that changes in carbon uptake in response to elevated carbon dioxide can alter trophic interactions. For example, the reduction in leaf nitrogen concentrations alters herbivore behavior and survival. Insect herbivores growing on leaves produced by plants grown at high carbon dioxide concentration levels usually consume more leaf tissue, but have slower rates of growth and development because of the lower quality of that tissue. Changes in the performance of herbivores can also alter trophic interactions that amplify the detrimental effects of elevated carbon dioxide on insect herbivores. In one case, elevated carbon dioxide caused a two-fold increase in mortality of insect herbivores because of reduced food quality, but also caused a four-fold increase in mortality because these herbivores are more vulnerable to attack by predators and parasitoids (Stiling *et al.*, 1999). In this case, the indirect effect mediated by predators substantially amplified the direct effect of reduced forage quality on herbivore mortality. Again, simple extrapolations of responses in the laboratory may differ substantially from responses observed in the field. Elevated carbon dioxide can also affect ruminant herbivores, and studies in tall grass prairie predict that forage produced in elevated carbon dioxide will cause a reduction in ruminant growth rate (Owensby *et al.*, 1996). In contrast to insects, ruminant tissue consumption declines with forage quality, so wild ruminants will likely suffer reduced growth and reproduction as carbon dioxide continues to rise (Owensby, 1996), with possible implications for higher trophic levels. Increased carbon inputs below ground can also alter soil food webs, changing the activity and abundances of bacteria and fungi, their protozoan and nematode predators, and mites and other organisms occupying higher trophic levels as well. Such changes in soil trophic structure could in uence nutrient cycling and decomposition rates.

As carbon dioxide diffuses from the atmosphere into leaves through stomata, water vapor diffuses from the leaves to the atmosphere, the process of transpiration. Increasing the concentration of atmospheric carbon dioxide strengthens the air-to-leaf concentration gradient of carbon dioxide, thereby increasing carbon dioxide supply to leaves, allowing stomata to close to save water, and resulting in lower rates of transpiration. While herbaceous species and tree seedlings often show strong reductions in stomatal conductance and transpiration in response to elevated carbon dioxide, responses in mature trees are more variable, and in some species there is no response to elevated carbon dioxide. Differences among species is at least partly related to growth form, as the stomatal response of coniferous trees to elevated carbon dioxide tends to be smaller than the response of herbs, and deciduous trees tend to be intermediate (Saxe *et al.*, 1998). The reduction in plant transpiration in response to elevated carbon dioxide can cause a reduction in evapotranspiration (the ux of water from the ecosystem to the atmosphere). Reduced evapotranspiration (latent heat ux) is usually compensated by increased canopy temperature and thus increased sensible heat ux. The ecosystem level consequences of reduced evapotranspiration include increased soil water storage, leaching, and/or runoff, and these changes, in turn, can alter a number of ecosystem processes sensitive to soil moisture, including plant growth, nutrient mineralization and trace gas uxes.

Elevated carbon dioxide usually increases photosynthesis at the canopy scale in both managed (e.g., a wheat field) and unmanaged (e.g., an alpine grassland) ecosystems, leading to the suggestion that terrestrial ecosystems could sequester some of the carbon dioxide being added to the atmosphere from fossil fuel burning, cement manufacturing and deforestation. However, observing carbon accumulation in ecosystem components in quantities

that match carbon uptake inferred by measurements of canopy photosynthesis has so far proven elusive. To date, there have been no publications in which these independent measures of carbon uptake can be balanced from a field experiment, but convergence of these approaches will provide greater confidence in our understanding of carbon dioxide effects on carbon uptake. Carbon uptake in elevated carbon dioxide experiments also illustrates the problem of scaling step-change experiments to the more gradual reality of global change. In response to a sudden increase in atmospheric carbon dioxide concentration, plant production can increase immediately, increasing NEP or carbon uptake. However, because the rate of carbon dioxide release from soil through decomposition increases as the amount of carbon in soil increases, carbon dioxide release through decomposition will eventually catch up, though with a lag period as the carbon is transferred from plants to soil pools of varying turnover times. All ecosystem carbon cycling models agree on this point. The ability of ecosystems to hold carbon over long periods requires large carbon reservoirs in pools that cycle slowly, such as wood and soil. Greater distribution to these pools in response to elevated carbon dioxide, and thus increased carbon storage, will be apparent in carbon dioxide doubling experiments if there is a relatively long lag period between increased carbon dioxide uptake and carbon dioxide release. If this lag period is shorter than expected, the extra carbon taken up in elevated carbon dioxide environments is being preferentially distributed to carbon pools with rapid turnover that quickly return carbon dioxide to the atmosphere. Few experiments have tried to determine the patterns of carbon distribution between carbon pools for short vs. long turnover times, but several have shown that elevated carbon dioxide tends to favor carbon distribution to rapidly cycling pools, limiting potential increases in carbon storage.

Effects on Freshwater Ecosystems

Freshwater ecosystems are frequently supersaturated with carbon dioxide, making direct effects of increasing atmospheric carbon dioxide on these ecosystems unlikely. However, freshwater ecosystems receive water, the dissolved nutrients and organic matter contained therein, and dead leaves, stems, and other plant parts from terrestrial ecosystems. Because terrestrial and aquatic ecosystems are linked in this way, effects of elevated carbon dioxide on production, water use, and nutrient cycling in terrestrial systems can have important consequences for freshwater ecosystems. For example, reductions in transpiration due to elevated carbon dioxide could increase stream flow and water yield from watersheds (though increases in evapotranspiration in response to warming will more than compensate for this in many cases). Increased plant production in response to elevated carbon dioxide could translate into greater inputs of plant litter into aquatic ecosystems, and changes in the chemical composition of that litter (e.g., concentrations of phenolic and other secondary compounds) could affect decomposition rates in aquatic ecosystems. Changes in the balance of microbial mineralization and immobilization of nitrogen in soils, along with changes in nitrification and plant uptake, could alter the export of inorganic nitrogen to streams and lakes. Such effects are only beginning to be documented in global change research.

EUTROPHICATION AND ACIDIFICATION (*see also* Eutrophication, Volume 2; Nitrogen Cycle, Volume 2; Phosphorus Cycle, Volume 2)

Human production of fertilizer nitrogen, cultivation of nitrogen-fixing crops, livestock husbandry, and fossil fuel burning have dramatically increased the inputs of fixed nitrogen (available for plant and microbial uptake) to terrestrial and aquatic ecosystems. These changes in land-use in watersheds have greatly increased inputs of phosphorus into aquatic ecosystems. These inputs constitute the phenomenon of eutrophication, a global change that alters the balance of nutrients that support ecosystem productivity. Atmospheric deposition of nitric and sulfuric acid is responsible for the acidification of terrestrial and freshwater ecosystems, often with deleterious effects.

Effects on Terrestrial Ecosystems

Current global inputs of anthropogenic nitrogen to terrestrial ecosystems match or exceed, natural preindustrial inputs through biological nitrogen fixation and lightning. Human-caused changes in the global nitrogen cycle have profound and widespread effects on ecosystems (Vitousek *et al.*, 1997). Of all the resources that plants require, nitrogen most commonly limits plant growth and NPP in terrestrial ecosystems. For this reason, increased inputs of nitrogen often stimulate NPP and carbon uptake by forest, grassland, and tundra ecosystems, as shown by a number of experiments. By stimulating NEP in these ecosystems, nitrogen deposition has likely contributed to the uptake of carbon dioxide derived from fossil fuels.

In ecosystems where nitrogen availability is naturally low, nitrogen deposition can favor the establishment, growth, and reproduction of introduced species adapted to high nitrogen levels over native species adapted to low nitrogen, causing marked shifts in plant transpiration composition, plant species often leading to biodiversity losses and local extinctions. For example, in the Netherlands, where rates of nitrogen deposition are the highest in the world, nitrogen deposition has converted

species-rich heathlands into species-poor grasslands, a complete change in the structure of the ecosystem (Aerts and Berendse, 1988). Nitrogen deposition can also affect an ecosystem's ability to withstand and recover from perturbations. For example, perennial grasslands in the US subjected to nitrogen addition had higher productivity in favorable years, but were affected more severely and recovered more slowly from drought than control plots without nitrogen addition (Tilman *et al.*, 1994).

Continuous additions of nitrogen will eventually cause some other resource to become more limiting to photosynthesis and growth, so that these processes will no longer show a positive response to nitrogen additions. At this point, the ecosystem is considered to be nitrogen-saturated. Nitrogen added to a nitrogen-saturated ecosystem is often matched by an equivalent amount of nitrogen leaving the system through leaching and gaseous pathways. In many other cases, though, not all nitrogen added through deposition (experimental or otherwise) to nitrogen-saturated ecosystems can be accounted for in increased biomass production, leaching, or trace gas losses. It appears that a substantial amount remains in the soil, immobilized either biotically or abiotically, and unavailable for plant or microbial uptake. This suggests that not all the nitrogen added through deposition will support extra plant growth, but it also suggests that some soils can absorb more ammonium than previously thought.

Not only will the productivity of nitrogen-saturated ecosystems no longer increase with nitrogen additions, but nitrogen saturation can also actually lead to system decline through several mechanisms. First, acid deposition, either as nitric acid (HNO_3) or sulfuric acid (H_2SO_4), directly leads to soil acidification. Even when nitrogen deposition occurs as ammonium (NH_4^+), the NH_4^+ is often quickly converted to nitrate (NO_3^-) via nitrification, a process releasing hydrogen ions and thus increasing soil acidity. Aluminum is biologically available at low pH, and can build up to toxic levels when soils are acidified. Second, nitrogen-saturated ecosystems lose added nitrogen through nitrate leaching and in gaseous forms through microbial transformations (nitrification and denitrification). Leaching losses of nitrate can substantially reduce soil fertility, because positively charged base cations (potassium, calcium, and magnesium) are carried away with the negatively charged nitrate as it leaches. Gaseous losses of nitrogen through microbial processes occur as dinitrogen gas (N_2), which is inert in the atmosphere, but also as nitric oxide (NO), nitrogen dioxide (NO_2), and nitrous oxide (N_2O). Nitric oxide and nitrogen dioxide are reactive trace gases that contribute to photochemical smog and acid rain. Nitrous oxide, by contrast, is a very stable radiatively active greenhouse gas, 200 times more effective at trapping heat than carbon dioxide. Through increased nitrous oxide ef ux from soils (a major source of nitrous oxide to the atmosphere), ecosystems effectively translate one facet of global change, nitrogen deposition, to another, global warming.

Effects on Freshwater Ecosystems

Runoff from urban and agricultural areas transports nutrients (nitrogen and phosphorus) from terrestrial to aquatic ecosystems, leading to *eutrophication*, by many accounts the most widespread and severe type of global change affecting lakes (Lodge, 2001) (*see* **Eutrophication**, Volume 2). Phosphorus is the nutrient most commonly limiting to NPP in freshwaters, particularly in temperate latitudes, so added phosphorus generally increases algal productivity. Nitrogen is often the next most limiting nutrient, so freshwater ecosystems already affected by increased phosphorus inputs are likely to also be very responsive to increased inputs of nitrogen from the atmosphere. Eutrophication, in addition to increasing NPP, can also alter algal species composition, favoring those species better able to take advantage of the higher nutrient levels.

Increased algal productivity in response to moderate eutrophication can stimulate productivity of higher trophic levels in aquatic ecosystems, but more often eutrophication favors particular species of algae that may be inedible and/or toxic, causing noxious *algal blooms*; extremely dense mats of algae that decrease light penetration and alter thermal structure. When these algae die, the input of organic matter fuels decomposition and depletes oxygen concentrations, killing fish and other organisms (Carpenter *et al.*, 1998).

Deposition of sulfuric and nitric acid to lakes causes acidification, particularly in lakes which are low in alkalinity (the chemical capacity to buffer changes in pH in response to acid inputs), such as those in the northeastern part of North America and in Scandinavia (Schindler, 1998). Most organisms are sensitive to changes in pH, so the effects of acidification on species distributions and abundances are often quite direct. For example, crustaceans and mollusks are particularly sensitive to the direct effects of acidification, because it reduces the availability of the calcium and carbonates required to build their exoskeletons. Acidification can reduce algal productivity and alter species composition, indirectly affecting higher trophic levels. Acidification of lakes also increases water clarity and reduces concentrations of dissolved organic carbon (DOC) (Schindler, 1998). Finally, the effects of acidification can be indirect as well. In lakes of the Laurentian Shield, with low-buffering capacity, for example, the increase in hydrogen ion concentration leads to mobilization from the sediments of toxic heavy metals such as mercury. These subsequently accumulate in fish, which then fail to reproduce.

GLOBAL WARMING

Effects on Terrestrial Ecosystems

Global warming will likely alter many ecosystem processes in terrestrial ecosystems, including increasing decomposition, increasing evapotranspiration, altering NPP, and shifting species composition in ways that can substantially modulate the initial direct responses. Potential effects of global warming on terrestrial ecosystems have been investigated in the field by experimentally warming the air around vegetation (using greenhouses or infrared heaters), by warming the soil using heat tape, or by studying natural gradients in temperature across the landscape.

As with elevated carbon dioxide and nitrogen deposition, understanding how global warming will alter the carbon balance of terrestrial ecosystems is a major focus of global change research. In laboratory incubations, microbial respiration responds exponentially to increasing temperature, and litter decomposition in the field shows a similar trend. By contrast, photosynthesis has a saturating relationship with temperature for most plant species, with optimal temperatures close to those typical of the environment in which the plant is found. Warming will likely increase rates of litter decomposition, stimulating a release of carbon dioxide from ecosystems to the atmosphere, and across 30 different ecosystem warming experiments, the ux of carbon dioxide from soil to the atmosphere increased (Rustad *et al.*, 2000). But photosynthesis can increase as well: in tundra, warming did not change the carbon balance; the effects of increased decomposition were counterbalanced by increased photosynthesis (Hobbie and Chapin, 1998). Increased carbon dioxide ef ux from arctic and alpine tundra in response to recent warming trends and to experimental manipulations appears to be due to warmer temperatures causing soil drying. In high latitudes, drying of waterlogged peat soils could release large amounts of carbon dioxide to the atmosphere. Increased decomposition should also increase nutrient mineralization and nutrient availability to plants, an effect that could ultimately enhance carbon uptake, as plants sequester nutrients in higher carbon-to-nutrient ratios than those in soils. However, this effect does not seem to have strongly in uenced results from warming experiments to date.

Global circulation models predict that warming will be most pronounced in upper latitudes. In tundra ecosystems, the indirect effects of increased temperature may more strongly determine the responses of tundra ecosystems to climate change than the direct effects. For example, the temperature controls depth of thaw, soil nutrient availability, and growing season length, all of which exert greater control over tundra plant growth than does temperature directly (Chapin, 1983). Growth responses to temperature vary among tundra species, and increased abundance of shrubs and decreases in mosses and other non-vascular species in response to experimental warming (and that observed over the last decade, a period when mean annual temperatures have also increased) suggest that shifts in species composition in tundra ecosystems will be one of the critical drivers of changes in ecosystem processes in response to global warming (Chapin and Shaver, 1996). In lower latitudes, too, warming could cause shifts in species composition that alter ecosystem processes. For example, warming could favor biological invasions by weedy C4 grasses in hotter climates, because of their higher temperature optimum for photosynthesis and their greater drought tolerance compared to C3 and C4 plants (Dukes and Mooney, 1999; *see* C_3 and C_4 Photosynthesis, Volume 2).

Experimental warming usually results in drying as well, and often the latter is the more critical driver of changes in ecosystem structure and processes, particularly in arid and semi-arid ecosystems. One of the challenges in this research area is to disentangle direct and indirect effects. Climate models predict that changes in precipitation caused by global warming will be regionally specific, increasing in some areas and decreasing in others, and thus are not always well-represented by the drying that occurs with experimental warming.

Effects on Freshwater Ecosystems

Global warming will almost certainly alter the hydrologic cycle, directly altering the amount and distribution of water on the terrestrial surface and thus the extent of freshwater ecosystems. While warming will increase evapotranspiration globally, changes in precipitation will vary from region to region. Thus, regionally-specific climate change scenarios are needed to predict where increases and decreases in the extent of freshwaters are likely to occur. For example, warming is predicted to reduce runoff and lake water levels in the Great Lakes region (Mortsch and Quinn, 1996), but to increase runoff and possibly ood intensity in the Amazon Basin (Neilson and Marks, 1994). Overall, however, evaporation will increase more than precipitation, reducing the extent of freshwater ecosystems globally. In arid and semi-arid regions, as well as in high latitudes, some freshwater habitats could disappear (Schindler *et al.*, 1996).

The paleo record indicates that changes in global temperatures and regional precipitation patterns have strongly affected ood magnitude and frequency (Ely *et al.*, 1993). Global warming is thus likely to alter the *disturbance regime*, an important regulator of ecosystem structure and function in many streams and rivers. Increased temperatures will also likely increase the extent and duration of thermal stratification in lakes (*see also* Disturbance, Volume 2).

Changes in the distribution and abundance of aquatic organisms will likely accompany global warming, leading to increases in the ranges of organisms where corridors between habitats allow migration; and to extinctions where

corridors to habitats with appropriate thermal regimes do not exist. Direct effects of warming on organisms are likely to be greatest in high latitudes, where warming is predicted to be greater and where native biota are more sensitive to increased temperatures. For example, model simulations suggest that warming will increase food requirements of lake trout in arctic lakes, but long-term observations over a recent warming period revealed no changes in primary production to fuel this greater food demand, thus threatening the top trophic level in arctic lakes (McDonald *et al.*, 1996). Such changes are likely to have strong impacts on ecosystem processes, because top predators often are major determinants of nutrient cycling, water clarity, and net ecosystem productivity in freshwater ecosystems (Schindler *et al.*, 1996).

Changes in runoff and in terrestrial ecosystem processes caused by warming could alter the amounts and forms of nutrient and plant litter inputs to freshwaters, affecting production, decomposition, and the food webs these processes fuel in streams, rivers, and lakes. As mentioned above, warming can reduce DOC supply to lakes and streams from terrestrial ecosystems (Schindler *et al.*, 1996), but could cause the opposite effect in northern latitudes, where melting permafrost and increased decomposition of old peat (e.g., Oechel *et al.*, 1993) could enhance DOC transport from tundra to arctic lakes and streams.

ENHANCED UV-B RADIATION

As a direct consequence of human production of ozone-depleting chemicals, such as chloro uorocarbons and nitrous oxide (N_2O), concentrations of ozone (O_3) in the stratosphere (the ozone shield) are currently at their lowest recorded values. Lower levels of stratospheric ozone allows greater penetration of UV-B radiation (280–315 nm) to the Earth's surface where it can damage deoxyribonucleic acid (DNA), proteins, lipids, and other UV-B absorbing compounds in living tissues, alter tissue composition and growth, and cause photolysis of non-living organic matter, all of which have important consequences for terrestrial and freshwater ecosystems.

Effects on Terrestrial Ecosystems

In terrestrial ecosystems, inhibition of photosynthesis by UV-B reduces leaf area and stem growth in many species, potentially reducing NPP. Water stress can mitigate these effects, because water-stressed plants often produce higher concentrations of UV-B absorbing compounds. The effects of enhanced UV-B radiation on plant growth can also interact with elevated carbon dioxide concentrations.

Some plants tolerate UV-B more than others, such that shifts in species composition are a likely consequence of enhanced UV-B radiation. In some cases, these changes can buffer reductions in NPP, if UV-B induced reductions in plant growth lead to the acquisition and growth of more tolerant species (Caldwell *et al.*, 1998).

Organisms vary in their natural abilities to cope with enhanced UV-B radiation. Some, particularly microbes and insects in early stages of development (early instars), are not well protected against UV-B, having little ability to adjust concentrations of UV-B absorbing compounds, and are therefore quite sensitive to UV-B damage. This susceptibility is exacerbated when the natural habitats of those organisms, such as the surfaces of leaves, soils or biological crusts in the desert, involve direct exposure to UV-B radiation. Plants often alter their chemical composition in response to enhanced UV-B radiation, producing greater amounts of phenolic compounds (tannins, lignins, and other refractory materials) or antioxidants (e.g., avenoids) that can absorb some of the damaging radiation and thereby protect the plant. Plants can also alter morphological properties, such as leaf thickness, which offers greater resistance to UV-B damage.

Changes in plant chemistry and morphology can in turn alter rates of herbivory and decomposition processes that are very sensitive to leaf thickness and chemical composition. UV-B can also increase litter decomposition directly, through photolysis, when solar radiation breaks chemical bonds. In these ways, enhanced UV-B radiation can alter the cycling of carbon and nutrients in ecosystems, but the direction of these effects is difficult to predict. For example, in some cases, litter produced under enhanced UV-B decomposes more slowly because the UV-B treatment alters the chemical composition of litter. But in other cases, UV-B increases decomposition, possibly because of the direct effects of photolysis. Enhanced UV-B radiation can also affect nitrogen cycling by inhibiting biological nitrogen fixation.

Effects on Freshwater Ecosystems

In contrast to terrestrial ecosystems, where UV-B radiation is attenuated only by the atmosphere, UV-B penetration into freshwater ecosystems declines with water depth and with increasing concentrations of DOC, particularly DOC high in UV-B absorbing chemical structures (phenols, humins, and other aromatics). Much of this DOC is of terrestrial origin–byproducts of the decomposition of terrestrial plant litter. The water-soluble portion of this material can be transported into freshwater ecosystems and there serve an important role in absorbing UV-B radiation. Through photolysis, UV-B radiation breaks down this DOC into smaller organic compounds that are usually more susceptible to microbial decomposition. Increased photolysis and decomposition in response to enhanced UV-B allow greater penetration of UV-B into the water column and further DOC photolysis, creating a positive feedback cycle. Freshwaters

naturally vary in DOC concentrations, so the penetration of UV-B into the water column, and the potential for detrimental effects of enhanced UV-B on ecosystems, also vary among habitats.

Increased processing of DOC and altered rates of microbial metabolism can in turn affect nutrient cycling and primary production in freshwater ecosystems. For example, photolysis can release nitrogen compounds (ammonium and amino acids) that stimulate microbial activity and turnover, and thus the rates of nitrogen cycling. Many of the enzymes involved in nitrogen transformations, however, such as nitrogenase (nitrogen fixation) and glutamine synthetase (ammonium uptake) are inhibited by UV-B radiation, such that the overall effects of UV-B on rates of nitrogen cycling are difficult to predict. Changes in algal production in response to enhanced UV-B radiation, likewise, will re ect the direct negative impacts of enhanced UV-B on photosynthesis and growth, as well as any indirect effects on the availability of nutrients limiting to production. UV-B also affects zooplankton, both directly, by damaging tissues, and indirectly, by affecting the amount and quality of primary production upon which these grazers depend.

For both phytoplankton and zooplankton, enhanced UV-B often results in a shift in community composition, because species differ in sensitivity to UV-B damage. In some ecosystems, entire trophic levels vary in sensitivity to UV-B damage. For example, algal production increased in response to enhanced UV-B in a stream habitat (Hader *et al.*, 1998). Although the algae in isolation were negatively affected by UV-B, the dominant grazers (larval chironomids) were even more sensitive, and largely eliminated by the UV-B treatment, releasing the algae from grazing pressure. These observations highlight the danger in predicting ecosystem responses to global change by examining any part of the system in isolation.

Because of the central role of DOC in attenuating UV-B penetration into aquatic ecosystems, any global change factor that in uences DOC concentrations will interact with enhanced UV-B radiation. A number of such interactions have been identified and observed through both experiments and observations in the Experimental Lakes Area in Canada. Here, observations over two decades, along with a number of long-term experiments showed that increases in lake temperatures and lake acidification occurred in concert with reductions in DOC concentrations, allowing greater penetration of UV-B into the water column. The reductions in DOC were caused by reduced inputs of DOC through runoff (particularly during drought years), increased microbial processing of DOC, decreased primary productivity, and greater lake stratification, which reduced the transfer of UV-B absorbing compounds from deeper water into the upper layers.

LAND-USE CHANGE

More and more land is used for livestock grazing, agriculture, forestry, and settlement, and these changes in land use and land cover are projected to increase. Land-use change is currently the most severe global change affecting terrestrial ecosystems, and it is likely to remain the dominant one in tropical regions over the next several decades.

Land-use change alters both ecosystem structure and ecosystem processes. For example, land-clearing removes vegetation, and such clearing is often accomplished by burning, so the carbon stored in the plants is thereby released to the atmosphere as carbon dioxide. Similarly, tillage in agriculture disrupts soil associations that stabilize soil organic matter, stimulating decomposition and releasing carbon dioxide to the atmosphere. Largely due to these processes, past land-use changes have already substantially increased carbon losses from ecosystems to the atmosphere, accounting for nearly half of the observed increase in atmospheric carbon dioxide concentrations between 1750 and the present (Houghton, 1994).

Greater rates of nutrient losses often occur after land clearing or conversion to agriculture, due to both higher microbial activity and lower plant nutrient uptake. When land is cleared for agriculture, higher rates of nutrient losses are usually compensated by application of fertilizers and lime, allowing such ecosystems to maintain equal or even higher rates of NPP (particularly where agricultural fields are also irrigated). Nutrient cycling in agricultural ecosystems is thus fundamentally different from that in the natural terrestrial ecosystems they replaced. Typical rates of fertilizer application in intensive agriculture far exceed natural rates of nutrient importation (e.g., through natural nitrogen fixation), and losses of fertilizer also greatly exceed natural rates of nutrient loss (e.g., through leaching and gaseous pathways). For these reasons, managed ecosystems (particularly agricultural ones) are said to save *open nutrient cycles*, where inputs to and losses from the ecosystem are the dominant transfers. By contrast, nutrient cycles in natural terrestrial ecosystems are *closed*, in the sense that inputs and losses are far smaller than exchanges within ecosystems (e.g., mineralization of organic to inorganic forms). Nutrient losses from agricultural ecosystems contribute to a number of global changes: 1) losses of gaseous nitrous oxide contribute to global warming, as nitrous oxide has nearly 200 times more warming potential than carbon dioxide, molecule for molecule; 2) losses of gaseous nitric oxide and nitrogen dioxide contribute to photochemical smog and acidic nitrogen deposition; and 3) leaching losses of nutrients (e.g., NO_3^- and PO_4^{2-}) from agriculture contribute substantially to eutrophication of aquatic ecosystems.

Other changes in ecosystem structure and processes accompany land-use changes. For example, converting a native forest to a pasture reduces canopy height and rooting depth, thereby altering water and energy exchange with

the atmosphere (reducing latent heat flux and increasing sensible heat flux, for example). Over large enough scales, such changes in energy partitioning can change regional and potentially even global climate (Shukla *et al.*, 1990). Land-use changes also alter connections between habitats, converting, for example, a large homogenous forest stand to a patchwork mosaic of managed ecosystems and remnant forest. Such changes in ecosystem structure can threaten animal species that require large ranges and can facilitate species invasions.

Rates of soil erosion driven by both wind and water are also typically much higher as a result of land-use change. Losses of soil are not easily replaced, so increasing soil erosion caused by land-use change remains a serious global problem, affecting the long-term fertility of agricultural ecosystems, as well as the integrity of ecosystems downstream and downwind.

HYDROLOGIC ALTERATIONS

Direct human changes in the global water cycle (through impoundments (dams), surface water diversions, and groundwater extraction) have dramatically altered freshwater ecosystems. Globally, nearly a million dams interrupt natural river flows, appropriating more than half of the available global runoff for human use (Postel *et al.*, 1996).

Water impoundment through dam building converts upstream rivers to reservoirs, and dampens the natural variation in flow regimes downstream, both of which constitute fundamental changes in the structure of river ecosystems. Upstream, converting a river to a reservoir introduces thermal stratification, drastically reduces water flow, and changes the light environment, leading to a complete change in the structure of the producer community with associated changes in food web structure. Additionally, buried terrestrial vegetation and soil organic matter decomposes in reservoirs, releasing carbon dioxide to the atmosphere, and also methane (CH_4), a greenhouse gas produced when decomposition proceeds in the absence of oxygen (St. Louis *et al.*, 2000). Downstream, dams largely eliminate natural flooding regimes, seasonal fluctuations in water temperature, and sediment loads, often endangering or causing the extinction of the native species that were adapted to such fundamental characteristics of the river, including species important to fisheries (Postel, 1998). Such changes in water flow regimes can detrimentally affect important habitat, including floodplains, riparian zones, wetlands, and estuaries, all of which are particularly sensitive to changes in water flow. Reduced water flow can also impair water quality, as pollutants become less and less dilute. The dam itself also constitutes a formidable barrier to species migration, leading to genetic isolation and potentially to extinction (Pringle, 1997). Nutrient delivery (particularly of silicates) is interrupted by impoundments, with detrimental effects on marine algae and associated food webs. For a larger-scale example, see Humborg *et al.* (1997), who reported on the impacts of a recently constructed dam on the Danube River between Hungary and Slovakia, on ecosystem structure in the Black Sea.

INVASIVE SPECIES

Through intentional introductions and through accidental hitchhikers in our global transportation systems, humans have dramatically redistributed the species on earth. Some of these species introductions are relatively benign, but other invasive species detrimentally affect native species and ecosystems. Invasive species are first transported to a new geographic area, they survive and persist in this new area, and eventually they thrive, reproducing successfully, spreading in areal extent, and outcompeting native species, not infrequently driving some native species to extinction. The species that make up an ecosystem strongly influence that ecosystem's structure and processes, so the changes in species composition caused by invasions can have ecosystem-scale effects. In general, invasive species can alter ecosystem processes by changing trophic structure, resource availability, or disturbance frequency or intensity (Vitousek, 1989).

Nile perch, introduced into Africa's Lake Victoria, severely altered the trophic structure of the lake by literally eating the species beneath it on the food chain, leading to the elimination of many of the lake's 500 endemic species of cichlid fishes. Other invasions alter the nutrient or water cycles, such as the nitrogen-fixing tree, *Myrica faya*, which invaded the Hawaiian islands and caused a 90-fold increase in nitrogen inputs. A number of tree species that have invaded the South African Cape Province fynbos ecosystem whose high rates of water extraction have strongly reduced, and in some cases even eliminated, river flow. Australian paperbark has invaded south Florida, covering nearly 200 000 acres, and promotes more intense fires, altering the natural disturbance cycle. Many other invasive species have similar effects, particularly grasses which can cause a buildup of dead flammable material, leading to fires that favor the expansion of the invading grass (D Antonio and Vitousek, 1992). Given their global extent and strong effects on ecosystems, biological invasions are a critical component of human caused global change (Mack *et al.*, 2000).

See also: **Biological Invasions**, Volume 2; **Plant Dispersal and Migration**, Volume 2.

INTEGRATION

Most global change research addresses how one particular type of global change (rising carbon dioxide, nitrogen deposition, or invasion of a particular species) will affect

one (or a few) characteristics of a particular ecosystem. Yet, in reality, global changes are occurring together and they affect many (if not all) aspects of ecosystems, not just those of a particular researcher's specialty. It is thus critical to attempt more integrative assessments of responses to global change. For example, warming or nitrogen deposition can favor the success of an invasive plant that alters the fire cycle. Similarly, global warming or rising carbon dioxide can alter the export of DOC from terrestrial to freshwater ecosystems, fundamentally changing how these aquatic systems respond to enhanced UV-B radiation. Land-use changes can alter albedo and sensible heat flux, causing changes in climate. Elevated carbon dioxide, through a number of indirect pathways, can affect microbial processes in soil that produce other trace gases, including the greenhouse gases nitrous oxide and methane and the reactive gases nitric oxide and nitrogen dioxide. Global changes can thus interact in surprising ways: the response of a given ecosystem to one component of global change can affect its response to another, a neighboring ecosystem's responses, and even the extent of other global changes. Understanding these interactions is a formidable challenge.

CONCLUSIONS

Human activities are clearly changing Earth's terrestrial and freshwater ecosystems, changing both their structure and the processes they perform. No longer can one find pristine habitats free from human influence (Vitousek *et al.*, 1997): rising atmospheric carbon dioxide concentration is a global phenomenon, and pollutants are evident everywhere that researchers have looked, in the high arctic and in remote tropical islands, a clear example of humanity's global shadow.

ACKNOWLEDGMENTS

Thanks to Pep Canadell, Connie Alice Hungate, George Koch, Ted Munn, and Adam Langley for comments on the manuscript and to the Merriam-Powell Center for Environmental Research for financial support.

REFERENCES

Aerts, R and Berendse, F (1988) The Effects of Increased Nutrient Availability on Vegetation Dynamics in wet Heathlands, *Vegetatio*, 76, 63–69.

Caldwell, M M, Bjorn, L O, Bornman, J F, Flint, S D, Kulandaivelu, G, Teramura, A H, and Tevini, M (1998) Effects of Increased Solar Ultraviolet Radiation on Terrestrial Ecosystems, *J. Photochem. Photobiol. B*, 46, 40–52.

Carpenter, S R, Caraco, N F, Correll, D L, Howarth, R W, Sharpley, A N, and Smith, V H (1998) Nonpoint Pollution of Surface Waters with Phosphorus and Nitrogen, *Ecol. Appl.*, 8, 559–568.

Chapin, III, F S (1983) Direct and Indirect Effects of Temperature on Arctic Plants, *Polar Biol.*, 2, 47–52.

Chapin, III, F S and Shaver, G R (1996) Physiological and Growth Responses of Arctic Plants to a Field Experiment Simulating Climatic Change, *Ecology*, 77, 822–840.

D'Antonio, C M and Vitousek, P M (1992) Biological Invasion by Exotic Grasses, the Grass/Fire Cycle and Global Change, *Annu. Rev. Ecol. Syst.*, 23, 63–87.

Dukes, J S and Mooney, H A (1999) Does Global Change Increase the Success of Biological Invaders? *Trends Ecol. Evol.*, 14, 135–139.

Ely, L L, Enzel Y, Baker, V R, and Cayan, D R (1993) A 5000-Years Record of Extreme Floods and Climate Change in the Southwestern United States, *Science*, 262, 410–412.

Hader, D P, Kumar, H D, Smith, R C, and Worrest, R C (1998) Effects on Aquatic Ecosystems, *J. Photochem. Photobiol. B*, 46, 53–68.

Hobbie, S E and Chapin, III, F S (1998) The Response of Tundra Plant Biomass, Aboveground Production, Nitrogen, and CO_2 Flux to Experimental Warming, *Ecology*, 79, 1526–1544.

Houghton, R A (1994) Balancing the Global Carbon Cycle with Terrestrial Ecosystems, in *The Role of Non-living Organic Matter in the Earth's Carbon Cycle*, eds R G Zepp and Ch Sontag, John Wiley & Sons, Chichester, 133–154.

Humborg, C, Ittekkot, V, Coclasu, A, and van Bodungern, B (1997) Effect of Danube River Dam on Black Sea Biogeochemistry and Ecosystem Structure, *Nature*, 386, 385–388.

Keeling, C D and Whorf, T P (2000) Atmospheric CO_2 Records from Sites in the SIO Air Sampling Network, in *Trends: A Compendium of Data on Global Change*, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, TN.

Lodge, D M (2001) Lakes, in: *Future Scenarios of Global Biodiversity*, eds F S Chapin, III, O E Sala, and E Huber-Sannwald, Springer-Verlag, New York.

Mack, R N, Simberloff, D, Lonsdale, W M, Evans, H, Clout, M, and Bazzaz, F A (2000) Biotic Invasions: Causes, Epidemiology, Global Consequences, and Control, *Ecol. Appl.*, 10, 689–710.

McDonald, M E, Hershey, A E, and Miller, M C (1996) Global Warming Impacts on Lake Trout in Arctic Lakes, *Limnol. Oceanogr.*, 41, 1102–1108.

Mortsch, L S and Quinn, F H (1996) Climate Change Scenarios for Great Lakes Basin Ecosystem Studies, *Limnol. Oceanogr.*, 41, 903–911.

Neilson, R P and Marks, D (1994) A Global Perspective of Regional Vegetation and Hydrologic Sensitivities from Climatic Change, *J. Vegetation Sci.*, 5, 715–730.

Norby, R J and Cotrufo, M F (1998) A Question of Litter Quality, *Nature*, 396, 17–18.

Oechel, W C, Hastings, S J, Vourlitis, G, Jenkins, M, Reichers, G, and Grulke, N (1993) Recent Change of Arctic Tundra Ecosystems from a Net Carbon Dioxide Sink to a Source, *Nature*, 361, 520–523.

Owensby, C E, Cochran, R C, and Auen, L M (1996) Effects of Elevated Carbon Dioxide on Forage Quality for Ruminants, in *Carbon Dioxide, Populations, and Communities*, eds

Ch Korner and F A Bazzaz, Academic Press, San Diego, CA, 363–371.

Petit, J R, Jouzel, J, Raynaud, D, Barkov, N I, Barnola, J-M, Basile, I, Bender, M, Chappellaz, J, Davis, M, Delaygue, G, Delmotte, M, Kotlyakov, V M, Legrand, M, Lipenkov, V Y, Lorius, C, Pepin, L, Ritz, C, Saltzman, E, and Stievenard, M (1999) Climate and Atmospheric History of the Past 420 000 years from the Vostok Ice Core, Antarctica, *Nature*, 399, 429–436.

Postel, S L (1998) Water for Food Production: Will There be Enough in 2025? *BioScience*, 48, 629–637.

Postel, S L, Daily, G C, and Ehrlich, P R (1996) Human Appropriation of Renewable Fresh Water, *Science*, 271, 785–788.

Pringle, C M (1997) Exploring how Disturbance is Transmitted Upstream: Going Against the Flow, *J. N. Am. Benthol. Soc.*, 16, 425–438.

Rustad, L E, Campbell, J, Marion, G M, Norby, R J, Mitchell, M J, Hartley, A E, Cornelissen, J H C, and Gurevitch, J (2000) GCTE-News [Network of Ecosystem Warming Studies], A Meta-analysis of the Response of Soil Respiration, Net N Mineralization, and Aboveground Plant Growth to Experimental Ecosystem Warming, *Oecologia*, 126(4), 543–562.

Saxe, H, Ellsworth, D S, and Heath, J (1998) Tansley Review No. 98: Tree and Forest Functioning in an Enriched CO_2 Atmosphere, *New Phytol.*, 139, 395–436.

Schinder, D W (1998) A Dim Future for Boreal Waters and Landscapes: Cumulative Effects of Climatic Warming, Stratospheric Ozone Depletion, Acid Precipitation, and other Human Activities, *BioScience*, 48, 157–164.

Schindler, D W, Curtis, P J, Pareker, B R, and Stainton, M P (1996) Consequences of Climate Warming and Lake Acidification for UV-B Penetration in North American Boreal Lakes, *Nature*, 379, 705–708.

Shukla, J, Nobre, C, and Sellers, P (1990) Amazon Deforestation and Climate Change, *Science*, 247, 1322–1325.

St Louis, V L, Kelly, C A, Duchemin, E, Rudd, J W M, and Rosenberg, D M (2000) Reservoir Surfaces as Sources of Greenhouse Gases to the Atmosphere: A Global Estimate, *BioScience*, 50, 766–775.

Stiling, P, Rossi, A M, Hungate, B A, Dijkstra, P, Hinkle, C R, Knott, III, W M, and Drake, B G (1999) Decreased Leaf-miner Abundance in Elevated CO_2: Reduced Leaf Quality and Increased Parasitoid Attack, *Ecol. Appl.*, 9, 240–244.

Tilman, D, Wedin, D, and Knops, J (1994) Productivity and Sustainability In uenced by Biodiversity in Grassland Ecosystems, *Nature*, 379, 718–720.

Vitousek, P M, Aber, J D, Howarth, R W, Likens, G E, Matson, P A, Schindler, D W, Schlesinger, W H, and Tilman, D G (1997) Human Alteration of the Global Nitrogen Cycle: Sources and Consequences, *Ecol. Appl.*, 7, 737–750.

A

Acidification, Lakes and Soils

see Nitrogen Cycle (**Volume 2**)

Allometric

Allometric is a term most commonly used with reference to growth, usually of plants. *Allometric growth* describes a situation in which the rate of growth of one part of an organism is in a constant ratio with the rate of growth of another part. It may be expressed as the *allometric coefficient* (k), which is the ratio of relative growth rates: $k = (\mathrm{d}\ln x/\mathrm{dt})/(\mathrm{d}\ln y/\mathrm{dt})$. When $k = 1$ the growth of the two parts is said to be *isometric*; when $k > 1$ or $k < 1$ the situations are described as positive or negative allometry, respectively.

Allometric growth of plant organs or parts strongly suggests that correlative effects control one or the other part of the plant, or both. The nature of such effects is often unclear, but is usually internal (e.g., genetic or hormonal). However, environmental effects may affect or disrupt allometric growth. For example, under ideal growing conditions the root/shoot ratio of some plants is essentially constant over considerable periods of the plant's life. This is almost certainly an internal correlative effect that maintains the balance between root and shoot. Environmental factors that may disrupt this allometry include excess soil humidity, which tends to slow root growth, or reduced soil water content, which tends to stimulate it. An understanding of allometric growth may assist agricultural efforts to achieve maximum yield consistent with optimal growth of, for example, fruit or root crops.

R G S BIDWELL *Canada*

Angiosperm/Gymnosperm

Higher plants, called *Spermatophytes* (seed bearing) are vascular plants characterized by the production of seeds. More primitive non-vascular plants such as fungi and mosses and vascular plants such as ferns reproduce by spores. Spermatophytes are divided into two classes: the *gymnosperms* (bearing naked seeds) and *angiosperms* (bearing enclosed seeds). The gymnosperms (*Pinophyta*), the more primitive group, are woody or perennial plants that are commonly evergreen; that is, they retain leaves (frequently needles) for two to several years so that the plant is leafy throughout the year. The angiosperms (*Magnoliophyta*) constitute the largest and most diverse group of plants, which has invaded and adapted to some of the world's most inhospitable habitats. Their habit of growth ranges from annual to perennial; many of the woody perennials are deciduous, losing all their leaves in winter or drought seasons.

Gymnosperms normally produce gametes in the axes of nodes clustered densely on a shoot with shortened internodes called a *strobilus* or *cone*. The gametes may be protected by bracts or scales, but the seeds (the female gametes or ovules) are not enclosed by the structure that bears them, nor in an ovary, and are therefore said to be naked even though they may be protected by the bracts of a cone. Angiosperms produce seeds that are normally born in structures called fruits, and the ovule is enclosed in an ovary; angiosperm seeds are therefore longer lasting after they have been shed. This evolutionary development appears to be largely responsible for the rapid and effective spread of angiosperms throughout the accessible habitats in the world. However, among woody plants, particularly trees, gymnosperms may better survive periods of extreme drought and cold (the most important consequence of intense cold is drought caused by freezing of available water) because of modifications that control water loss from their leaves.

R G S BIDWELL *Canada*

Animal Physiology and Global Environmental Change

Karen Martin[1] and Ken Nagy[2]
[1]Pepperdine University, Malibu, CA, USA
[2]University of California, Los Angles, CA, USA

Animals accommodate to environmental change in several ways, depending on the time scale involved. Over short periods of time, such as hours or days, animals faced with an increase in environmental temperature may use behavioral means to escape into cooler local microhabitats. Alternatively, they may use physiological or morphological means, such as increased water evaporation or shunting circulating uids around the body to keep cool, or the formation of heat shock proteins to protect enzyme systems, but the animals retain their original thermal set points. Over longer periods, such as weeks or seasons, animals can acclimatize by changing their physiological attributes and set points, for example, increasing preferred body temperature by ectothermic animals such as reptiles and adjusting insulation (pelage or plumage or circulation to the skin) by endothermic animals (birds and mammals). Some animals may also use seasonal migration to nd equable thermal conditions. Over even longer periods, such as generations, such within-phenotype adjustments fade in signi cance, and genetic adjustments, that is, evolutionary changes, become important. For example, insects may have higher thermal tolerances after several generations of living in hot conditions. Finally, over centuries or millennia, changes in species' geographic ranges occur when animals are not able to evolve rapidly enough or thoroughly enough to adjust to the new climatic conditions in their original species range, and when adjacent habitats newly having the historically preferred conditions of the species become available.

Much is known about the physiological adjustments animals make on an hourly, daily, and seasonal basis to changes in ambient temperature and water availability. However, because global changes are expected to be rather slow and long-term events, the main concern is with the adaptive changes in physiology and associated attributes that occur over generations and centuries. Unfortunately, it has not been possible yet to measure the physiology of animals that are extinct (Martin and Nagy, 1997). Therefore, limited factual information is available about the evolution of physiological traits, so much of the following must be speculative.

We assume that global environmental change will take the form of gradual warming, which may also be accompanied by more frequent storms and higher variability in climate. Local climatic warming often is accompanied by increasing aridity in speci c terrestrial habitats, as in recent changes in the Sahel in northern Africa, but on a global scale, rainfall may actually increase as a result of global warming. Where warranted below, we discuss the connection between higher temperatures and water balance, and the additional challenges faced by terrestrial animals.

TERRESTRIAL SYSTEMS

In a terrestrial system, increased temperatures would have several effects on animals. Most directly, heat balance and water balance would be affected. Temperature in most terrestrial animals is regulated, either physiologically by endotherms such as mammals and birds, or behaviorally by ectotherms such as reptiles and amphibians. The activity temperature of many animals is only a few degrees below a critically high, lethal or damaging temperature for them. Animals are more sensitive to high temperatures than plants because of the presence of a nervous system. Lethal effects of high temperature are caused by dysfunction of the central nervous system, resulting in loss of coordination and physiological controls within the organism.

Animals could compensate for increased temperatures, in part, by increasing their evaporative cooling, such as by exposing more wet skin or selecting windier perches for most amphibians, or by panting, uttering the gular area of the throat or sweating as in birds and mammals. However, problems would occur if water were not available in sufficient quantity. As temperature increases, relative humidity decreases if water content in the atmosphere remains constant. Thus, animals would lose water by evaporation faster, and have increased water requirements. If increased environmental temperature were accompanied by more frequent droughts, this would further reduce habitat quality for animals. Decreased dry season mist, related to increased warming, is blamed for the extinction of 20 out of 50 species of frogs and toads in the Monteverde Cloud Forests of Costa Rica after a population crash in 1987 (Pounds *et al.*, 1999).

Alternatively, endotherms could evolve higher body temperature set points along with higher thermal optima for metabolic enzymes. Interestingly, there is little evidence that endotherms are able to accomplish this. There are differences in regulated body temperatures among endothermic taxa; birds average about 41 °C, placental mammals average about 38 °C, and marsupial mammals (with pouches) average about 36 °C (Willmer *et al.*, 2000). However, mammals that live in warm habitats do not have higher body temperatures than mammals that live in cold habitats.

Ectothermic animals, such as most amphibians and reptiles, as well as most invertebrates, have limited control over the body temperature and must rely on outside sources for heating and cooling. Many of these animals are able to be active only within a restricted range of temperatures found in their biogeographic range. These body temperatures are obtained by behavioral choices made by moving within microhabitats at appropriate times of day. With increased air temperatures, ectothermic animals living in historically warm habitats would be forced to spend more time in refugia against extremely high temperatures, for example, in burrows or other shaded areas. Excessively high temperatures over extended periods of time could prevent these animals from foraging or otherwise utilizing the habitat, for example, for seeking mates. Of 151 populations of Edith's checkerspot butter y studied, populations found in Mexico were four times more likely to have become extinct than populations in Canada (Parmesan, 1996). In Europe, 63% of 35 non-migratory butter y species have shown northward shifts in range over this past century, by as much as 35–240 km (Parmesan *et al.*, 1999). In extreme situations, as in some deserts, some amphibians and reptiles remain inactive in burrows for nine or more months of the year. During drought years, desert tortoises *Gopherus agassizii* in California reduce their activity, consume much less food and water over the year, and lose weight (Henen *et al.*, 1998). In historically cool habitats, such as at higher latitudes, ectotherms may enjoy increased opportunities for activity through the year, as a result of global warming. This would improve their food consumption, favoring more rapid growth and reproduction, provided that increased aridity did not accompany the warming trend. In Great Britain, the frog *Rana kl esculenta* and the newt *Triturus vulgaris* both spawn 9 10 days earlier today than 25 years ago, following a 1 °C increase in maximum temperature (Beebee, 1995). Among insects, in the Netherlands some Microlepidoptera reach their peak ight date 11.6 days earlier today than two decades ago (Ellis *et al.*, 1997) and ve species of aphids in England now reach their ight period three to six days earlier than 25 years ago (Fleming and Tatchell, 1995).

Increased ambient temperatures affect metabolic processes of ectotherms as well, since their enzyme systems are in uenced by body temperature. Thus, even while at rest, an increase in body temperature increases energy expenditure in an ectothermic animal and requires it to obtain additional food. If additional food is not available, the animal is forced to compensate in other ways, by limiting growth or restricting reproductive output. On the other hand, enzyme thermal optima apparently are evolutionarily exible, especially in ectotherms (Johnston and Bennett, 1996). There is a trend of gradual replacement of ectothermic animals eating a given diet, with endotherms eating the same diet as latitude increases from the equator to the poles. This is exempli ed by the competition between seed-eating ants and rodents in North America (Brown and Davidson, 1977), and it indicates that the nature of ectotherm physiology is competitively superior in warm climates to that of endotherms. This suggests that global warming will be detrimental to endotherms in general, from the perspective of competition with ectotherms for food. A further detriment is beginning to be seen in the form of expanded ranges for disease-bearing insects carrying malaria, dengue, and yellow fever (Epstein *et al.*, 1998) (*see* **Infectious Diseases**, Volume 2; **Malnutrition, Infectious Diseases and Global Environmental Change**, Volume 3).

Endotherms have greater internal control of body temperature, but species still have geographic distributions that are limited in part by ambient temperatures. Every endothermic animal has a range of temperatures, considered its thermoneutral zone, where basal metabolism is suf cient to maintain the body and no excess energy is necessary for heating or cooling. An animal s thermoneutral zone is affected by its size, insulation, and surrounding microhabitat. Endothermic animals in ambient temperatures above the thermoneutral zone must expend energy and effort to reduce the body temperature, in order to prevent body damage. Evaporative cooling, often in the form of panting or sweating, is the only way of getting rid of excess heat in a hot environment. Muscular activity generates heat, so endothermic animals are less able to be active when in high ambient temperatures, especially above the thermoneutral zone. Thus, with additional energy required for cooling, and decreased inclination and ability to be active in the heat, endotherms would be less able to actively forage. Therefore, population size would probably decrease in the short term, and in the long term, the population would have to migrate out of an area with unsuitably high temperatures or evolve greater thermal tolerance. Range expansions into higher latitudes and altitudes have been seen over the past 20 years in 59 species of British birds (Thomas and Lennon, 1999), in 14 of 24 species of birds in the western US (Johnson, 1994), and in 19 species of small mammals in the US (Davis and Callahan, 1992). Also, of 225 bird species examined in the UK from 1971 to 1995, 31% are laying eggs earlier, by an average of 8.8 days (Crick and Sparks, 1999).

MARINE SYSTEMS

In the oceans, temperatures uctuate much less than in air. The high heat capacity of water mitigates air temperature uctuations and insures that marine organisms are far less subject to daily, or even seasonal, changes in ambient temperature than are terrestrial animals in air.

The disadvantage of this is that many marine animals are restricted to a very narrow range of temperatures, and cannot tolerate changes of more than a few degrees Celsius. In addition, the stability of the ocean temperatures over vast distances makes the use of microhabitats as temperature refuges impossible. Thus, most ectothermic marine animals do not regulate their body temperature by behavioral choices within a habitat, but instead by the habitat preference itself.

A lack of options to avoid high temperatures in the ocean within a habitat implies that an animal's only option when temperature is intolerable (aside from death, which is hardly a strategy for survival) is migration out of the area. Thus, global warming effects could take longer to be perceived in the ocean, because of its thermal stability. However, once temperatures increase above tolerable limits, the resulting effects could be much more rapid and devastating than those in terrestrial systems, where animals may obtain some temporary relief from temperature stress at night or in shelters.

Effects of emigration could produce a shift of species to higher latitudes or greater depths in the ocean. In the English Channel, plankton abundance changes in response to sea surface temperatures and barnacle distributions have shifted up to 30 km (Southward *et al.*, 1995). Over the past century, intertidal invertebrates in California have shifted northward in their distributions, and southern species have become more numerous while northern species have declined (Barry *et al.*, 1995). However, range shifts are limited by the restriction of plant life to the photic zone at the surface of the ocean, which further restricts herbivorous animals and their predators to surface waters. Thus, migration to depth seems unlikely for these species. Therefore, surface migrations could leave tropical waters less populated and increase population pressures and competition at higher latitudes. From 1951 to 1993, the biomass of macrozooplankton in southern California waters decreased by 80% (Roemmich and McGowan, 1995). Temperate water fishes and invertebrates would be forced to ever higher latitudes as their temperature tolerances were exceeded, and entire food webs would be disrupted or destroyed. Over the past 25 years, species richness of reef fishes in California waters has declined by 15–25%, with more southern species dominant to the detriment of more northern types, which have declined in abundance (Holbrook *et al.*, 1997).

To some extent, a preview of global warming effects on the temperate ocean can be seen during El Nino–southern oscillation events in the eastern Pacific Ocean. At these times, ocean currents reverse and coastal waters that are usually cool become several degrees warmer, for several months or more. During these events, warm water fishes can be found at high latitudes but are not as common in their usual habitats. Marine animals along the coast, particularly intertidal animals, are severely depleted and also recruit poorly during these events. Schooling fishes such as anchovies and sardines move offshore to cooler waters, removing a critical link of the coastal food chain. Seals and sea lions are unable to forage adequately, resulting in numerous deaths of older individuals and pups. Ectothermic fishes not only cannot find enough food, but their need for food is increased as they also have increased metabolic demands because of increased ambient temperatures. A dramatic drop of 90% of the population of sooty shearwaters, the dominant seabird species off the coast of California, occurred between 1987 and 1994 (Veit *et al.*, 1997) in three different, widely separated transects.

Sessile marine organisms that live attached to the ocean oor are in even worse danger. Since they are unable to migrate out of a territory, rapid environmental change can lead to local extinction. Of great concern currently, coral reefs appear to be affected by increased ocean temperatures of only a few degrees Celsius. These reefs show signs of poor health and decreased growth and reproduction (Wilkinson *et al.*, 1999). Because coral reefs are tropical, one might consider the possibility that global warming could produce increased habitat for these living ecosystems. However, the length of time required to establish a healthy reef as a result of niche expansion may be longer than the period of time available if change is rapid (Hoegh-Guldberg, 1999). In addition, the requirements for a healthy reef include shallow, clear water with stable salinity, conditions that may not be readily available along most temperate coastlines.

Although oceans are very large, their high productivity is restricted to only a very small proportion of the available area. Plants only grow in the surface and shallow waters of the photic zone, where light can penetrate, and so the vast majority of the ocean is dependent upon its thin upper layers. In addition, the most productive areas of the ocean are in cooler water, which is enriched by upwelling nutrients along coasts. Increasing oceanic temperatures will likely reduce productivity of the oceans as a whole, not only of the plants but also of the fish, seabirds, whales, and other animals that depend directly or indirectly on marine plants. Warming in the Antarctic Ocean has reduced sea ice cover, which is correlated with an increased number of penguins (Fraser *et al.*, 1992) but decreased amounts of krill (Loeb *et al.*, 1997).

SUMMARY

In short, global warming is likely to have profound and far-reaching effects on animals. Warming should have direct effects on temperature regulation and water balance in terrestrial animals, probably increasing water requirements

and indirectly increasing food requirements, and probably decreasing population sizes. Species may extend ranges into higher latitudes or higher altitudes but may move away from traditional ranges, potentially leading to increased isolation, local extinctions, and decreased genetic diversity. Alternatively, species may evolve greater thermal tolerances if global warming occurs slowly enough. Changes in distribution of both plants and animals will undoubtedly lead to dietary shifts and changes in food web dynamics, with possible de-coupling of co-evolved species interactions (Hughes, 2000). In marine animals, global warming should lead to migration to higher latitudes in most cases, or to greater depths for cooler temperatures. Sessile animals unable to migrate may face local extinction and range contraction. Reproductive outputs may decline as coastal waters become less productive, and food webs will be altered by changes in latitudinal distributions according to temperature tolerances. Change in the oceans' productivity probably will reduce humans' ability to harvest food from the sea.

See also: Animals: Impacts of Global Environmental Change, Volume 2.

REFERENCES

Barry, J P, Baxter, C H, Sagarin, R D, and Gilman, S E (1995) Climate-related Long-term Faunal Changes in a California Rocky Intertidal Community, *Science*, 267, 672–675.

Beebee, T J C (1995) Amphibian Breeding and Climate, *Nature*, 374, 219–220.

Brown, J H and Davidson, D W (1977) Competition between Seed-eating Rodents and Ants in Desert Ecosystems, *Science*, 196, 880–882.

Crick, H Q P and Sparks, T H (1999) Climate Change Related to Egg-laying Trends, *Nature*, 399, 423–424.

Davis, R and Callahan, J R (1992) Post-pleistocene Dispersal in the Mexican Vole (*Microtus mexicanus*), an Example of an Apparent Trend in the Distribution of Southwestern Mammals, *Great Basin Nat.*, 52, 262–268.

Ellis, W N, Donner, J H, and Kuchlein, J H (1997) Recent Shifts in Phenology of Microlepidoptera, Related to Climate Change (Lepidoptera), *Entmol. Berl. Amst.*, 57, 66–72.

Epstein, P R, Diaz, H F, Elias, S, Grabherr, S, Graham, N E, Martens, W J, Mosley-Thompson, E, and Susskind, J (1998) Biological and Physical Signs of Climate Change: Focus on Mosquito-borne Diseases, *Bull. Am. Meteorol. Soc.*, 79, 401–417.

Fleming, R A and Tatchell, G M (1995) Shifts in the Flight Periods of British Aphids: a Response to Climate Warming? in *Insects in a Changing Environment*, eds R Harrington and N Stork, Academic Press, London, 505–508.

Fraser, W R, Trivelpiece, W Z, Ainley, D G, and Trivelpiece, S G (1992) Increase in Antarctic Penguin Populations: Reduced Competition with Whales or a Loss of Sea Ice due to Environmental Warming? *Polar Biol.*, 11, 525–531.

Henen, B T, Peterson, C C, Wallis, I R, Berry, K H, and Nagy, K A (1998) Effects of Climatic Variation on Field Metabolism and Water Relations of Desert Tortoises, *Oecologia*, 117, 365–373.

Hoegh-Guldberg, O (1999) Climate Change, Coral Bleaching and the Future of the World's Coral Reefs, *Mar. Freshwater Res.*, 50, 839–866.

Holbrook, S J, Schmitt, R J, and Stevens, J A (1997) Changes in an Assemblage of Temperate Reef Fishes Associated with a Climate Shift, *Ecol. Appl.*, 7, 1299–1310.

Hughes, L (2000) Biological Consequences of Global Warming: is the Signal Already, *TREE*, 15, 56–61.

Johnson, N K (1994) Pioneering and Natural Expansion of Breeding Distributions in Western North American Birds, in *A Century of Avifaunal Change in Western North America*, eds J R Jehl and N K Johnson, Proceedings from the Centennial Meeting of the Cooper Ornithological Society, 27–44.

Johnston, I A and Bennett, A F, eds (1996) *Animals and Temperature: Phenotypic and Evolutionary Adaptation*, Cambridge University Press, Cambridge.

Loeb, V, Siegel, V, Holm-Hansen, O, Hewitt, R, Fraser, W, Trivelpiece, W, and Trivelpiece, R S (1997) Effects of Sea-ice Extent and Krill or Salp Dominance on the Antarctic Food Web, *Nature*, 387, 879–900.

Martin, K L M and Nagy, K A (1997) Water Balance and the Physiology of the Amniote Transition, in *Amniote Origins: Completing the Transition to Land*, eds S S Sumida and K L M Martin, Academic Press, San Diego, CA, 399–423.

Parmesan, C (1996) Climate and Species' Range, *Nature*, 382, 765–766.

Parmesan, C, Ryrholm, N, Stefanescu, C, Hill, J K, Thomas, C D, Descimons, H, Huntley, B, Kaila, L, Kullberg, J, Tammaru, T, Tennent, W J, Thomas, J A, and Warren, M (1999) Poleward Shifts in Geographic Ranges of Butter y Species Associated with Regional Warming, *Nature*, 399, 579–583.

Pounds, J A, Fogden, M P L, and Campbell, J H (1999) Biological Responses to Climate Change on a Tropical Mountain, *Nature*, 398, 611–615.

Roemmich, D and McGowan, J (1995) Climatic Warming and the Decline in Zooplankton in the California Current, *Science*, 267, 1324–1326.

Southward, A J, Hawkins, S J, and Burrows, M T (1995) Seventy Years' Observations of Changes in Distribution and Abundance of Zooplankton and Intertidal Organisms in the Western English Channel in Relation to Rising Sea Temperature, *J. Thermal Biol.*, 20, 127–155.

Thomas, C D and Lennon, J J (1999) Birds Extend their Ranges Northwards, *Nature*, 399, 213.

Veit, R R, McGowan, J A, Ainley, D J, Wahls, T R, and Pyle, P (1997) Apex Marine Predator Declines Ninety Percent in Association with Changing Oceanic Climate, *Global Change Biol.*, 3, 23–28.

Wilkinson, C, Linden, O, Cesar, H, Hodgson, G, Rubens, J, and Strong, A E (1999) Ecological and Socioeconomic Impacts of 1998 Coral Mortality in the Indian Ocean: an ENSO Impact and a Warning of Future Change? *Ambio*, 28, 188–196.

Willmer, P, Stone, G, and Johnston, I (2000) *Environmental Physiology of Animals*, Blackwell Science, Oxford.

Anthropogenic Impacts on Atmospheric Oxygen

Ray Langenfelds[1,2]

[1] *CSIRO Atmospheric Research, Victoria, Australia*
[2] *University of Tasmania, Hobart, Australia*

Molecular oxygen (O_2) is the second most abundant gas in the Earth's atmosphere after nitrogen (N_2), with a mole fraction of almost 21%. It is biogeochemically active, most notably as a key agent of biological and combustion processes. Natural, biological cycling of O_2 involves respiratory consumption by animals and plants, combustion through biomass burning and replenishment of the atmosphere by photosynthetic production.

O_2 trends in the contemporary atmosphere are dominated by exchange with land biota and combustion of fossil fuels. These processes provide an intimate link with exchange of carbon dioxide (CO_2), the main contributor to anthropogenic forcing of the Earth's radiation budget. Because of this link, O_2 is emerging as a key tracer of the global carbon cycle. Human-induced changes in atmospheric O_2 levels represent just a tiny fraction of its vast atmospheric reservoir. No direct environmental consequences of these changes have yet been demonstrated. However, their precise quanti cation is helping to elucidate the roles of oceans and the land biosphere in slowing accumulation of CO_2 in the atmosphere. Improved understanding of the global carbon cycle is leading to better forecasts of future atmospheric CO_2 levels and of subsequent effects on Earth's climate.

It has long been suspected that atmospheric oxygen (O_2) levels are being altered by human activities. A declining trend is expected from the anthropogenic perturbation to rates of combustion of biogenic material. Fossil fuels (coal, oil, natural gas) have been the primary energy source over the period of industrialization. Rates of land clearing have increased in tandem with world population growth and expansion of agriculture. Earlier, though as yet unfounded, concerns existed over diminution of O_2 through other processes, such as a possible decline in phytoplankton stocks due to pollution of the oceans. A few decades ago, there was speculation that O_2 losses might be large enough to have detrimental environmental consequences. Extreme predictions warned of a catastrophic O_2 decline that could threaten the habitability of the planet. These fears were demonstrated to be unwarranted by theoretical calculations of oxygen cycle processes (Broecker, 1970) and subsequent measurement of actual changes in the atmosphere.

Before the 1980s, there were no ongoing O_2 monitoring programs. Measurements were sporadic and obtained by different researchers at various times and geographical locations. In a review of measurements reported to 1970, Machta and Hughes (1970) concluded that O_2 had been stable at 20.95% between 1910 and 1970, with any trend too small to be resolved by available data. Since then, advances in measurement technologies have achieved much higher precision, revealing small trends that would previously have escaped detection. Reported observations of O_2 trends from contemporary monitoring programs are listed in Table 1. Air is collected in glass asks from remote sites (usually in the marine boundary layer) and returned to a central laboratory for analysis. Other studies have extended records back as far as 1977 using air extracted from Antarctic firn (the

Table 1 Reported observations of trends in atmospheric O_2

Reference	Experimental method	Mean O_2 trend (ppm year^{-1})[a]	Time period	Inferred net carbon flux (PgC year^{-1})[b]	
				Oceans	Land biosphere
Keeling *et al.* (1996)	Flask sampling	-2.9 ± 0.4	1991–1994[c]	1.9 ± 0.5	1.8 ± 0.7
Battle *et al.* (2000)	Flask sampling	-3.4 ± 0.2	1991–1997	2.0 ± 0.6	1.4 ± 1.0
Battle *et al.* (1996)	South Pole firn	-3.6 ± 0.4[d]	1977–1985	3.0 ± 1.1	-0.4 ± 1.1
Langenfelds *et al.* (1999)	Cape Grim air archive	-3.5 ± 0.1	1978–1997	2.3 ± 0.7	0.2 ± 0.9

[a] O_2 changes are described here in terms of changes in atmospheric mole fraction in units of ppm; however, measurements are of O_2/N_2 ratio and are generally reported elsewhere as relative changes in this ratio, in units of per meg as $\delta(O_2/N_2) = [(O_2/N_2)_{sample}/(O_2/N_2)_{reference} - 1] \times 10^6$ where 1 per meg = 0.001‰ = 0.0001%. Assuming a mole fraction of 20.95% (Machta and Hughes, 1970), $\Delta O_2(ppm) = 0.2095 \times \Delta\delta(O_2/N_2)$ (per meg).
[b] Positive values indicate removal from the atmosphere.
[c] The results quoted here are obtained from measurements at three, globally distributed sites. The record at one of these sites (La Jolla, CA) extends back to 1989 and yields similar values to the global case for the O_2 trend and carbon fluxes.
[d] In this case, the atmospheric O_2 trend is not directly measured, but is inferred from measurements of the ratio-depth profile of O_2/N_2 in the firn. Measurements are corrected for modification of the O_2/N_2 ratio by diffusive, gravitational and bubble exclusion processes.

permeable upper layer of the ice sheet) and from archived air collected at the Cape Grim Baseline Air Pollution Station in Tasmania.

Results confirm that the rate of O_2 decline is minute by comparison to the vast amount of O_2 present in the atmosphere. Of considerable value, however, is the precise definition of these trends, which allows partitioning of fossil carbon dioxide (CO_2) uptake between oceans and the land biosphere, a subject of vigorous debate over recent decades. The value of O_2 measurements for this purpose stems from the fact that unlike CO_2, O_2 has low solubility in seawater. Although there are large seasonal, air-sea uxes of O_2 driven by cycling of marine biological production and ventilation of near-surface waters, there is generally assumed to be negligible change in the oceanic O_2 inventory on interannual time scales. Thus, trends in atmospheric O_2 can be attributed to two processes, fossil fuel combustion and exchange with the land biosphere. The global average stoichiometry of O_2/CO_2 exchange can be estimated for both processes, as can the rate of CO_2 emission from fossil fuel use. Thus, the O_2 trend directly constrains the rate of change of the size of the land biosphere. Together with the concurrent global trend in CO_2, the system can also be solved for the rate of oceanic CO_2 uptake.

Evaluation of global carbon budgets using measured O_2 trends suggests that before the early 1990s, the land biosphere was close to balanced. In other words, the total amount of carbon stored in terrestrial biomass and soils remained constant. The fact that it did so, despite ongoing deforestation, implies that loss of biomass through deforestation was compensated by growth of vegetation elsewhere. In the early 1990s the land biosphere became a net carbon sink, implying even faster growth. Increased terrestrial uptake may re ect either reforestation of previously cleared land or fertilization of existing plants by higher atmospheric CO_2 levels, higher rates of nitrogen deposition or climatic factors (Houghton *et al.*, 1998). The O_2-based budgets also indicate steady oceanic uptake of about $2\,\mathrm{PgC\,year^{-1}}$, in close agreement with predictions from ocean carbon cycle models.

The knowledge gained from recent decades can be used to model the change in O_2 over the period of industrialization (Figure 1). This calculation uses the land/ocean partitioning of CO_2 uptake determined by Trudinger *et al.* (1999) based on a deconvolution of the icecore/ rn CO_2 record from Law Dome, Antarctica (Etheridge *et al.*, 1996). They determine oceanic CO_2 uptake using a box-diffusion model, calibrated against ^{14}C and assessed against the Law Dome $\delta^{13}C$ record of Francey *et al.* (1999). The modeled rate of CO_2 uptake at the end of the record is in close agreement with budgets derived from measured O_2 trends. Records of fossil fuel use and land use change allow determination of net terrestrial exchange or, as displayed in Figure 1, exchange with land biota not including land use change. Corresponding changes in O_2 can be calculated for each of these terms. Depletion of the atmosphere by 151 parts per million (ppm) O_2 is implied for the period 1850 1990. Small, additional changes must also have been incurred before 1850, however early rates of land use change are not well known. An assumed history, involving linear extrapolation of the 1850 estimate back to $0.2\,\mathrm{PgC\,year^{-1}}$ at 1765, would imply pre-1850 O_2 loss of ≥ 12 ppm. Comparable depletion in earlier centuries is unlikely, as the Law Dome CO_2 record shows no signi cant trend between 1000 and 1800 with variations spanning a range of ± 5 ppm. Any CO_2 increase due to land use change is masked by natural variability, for example, as observed during the Little Ice Age (Etheridge *et al.*, 1996). Estimates of early fossil fuel use back to 1750 account for a further loss of 1 ppm. Recent observations indicate a trend of approximately $-3.5\,\mathrm{ppm\,year^{-1}}$ during the 1990s (Table 1). Not included in Figure 1 are effects of other processes that may have partially offset O_2 consumption. They include thermal expansion of oceans due to higher global temperatures, increased storage of organic carbon in oceans from input of anthropogenic nitrogen, and industrial reduction of metal oxide ores (Keeling, 1988). The estimated sum of their cumulative contributions carries large uncertainty, with a best estimate in the range 5 10 ppm.

Thus, by the end of the 20th century, atmospheric O_2 has fallen by an estimated 190 ppm. Over the same time, CO_2 has increased by some 90 ppm. The growth in CO_2 represents a large relative change in concentration of 30% compared to pre-industrial levels. This has signi cantly changed the radiative properties of the atmosphere and has major implications for Earth s climate. By contrast, the relative change in O_2 is just -0.1% of its atmospheric burden (approximately 209 500 ppm). A change as small as this is unlikely to be of any direct signi cance. No environmental impacts have yet been observed. To put the magnitude of this change in context, we may regularly encounter much larger variations as part of our daily lives. O_2 levels can be heavily modi ed on local scales, close to sources of fossil fuel combustion or respiration. Notable examples are urban areas that typically carry large volumes of vehicular traf c, indoors in the presence of people, or in forests where O_2 can be depleted by plant respiration, especially under stable, nocturnal boundary layer conditions.

It is pertinent to consider how O_2 levels may change in the future. Over coming years to decades, further O_2 decline is inevitable. Rates of both fossil fuel combustion and land clearing continue to increase and there is no indication that these trends will be reversed in the near future. Although the land biosphere has partly offset anthropogenic O_2 consumption in recent years, its capacity

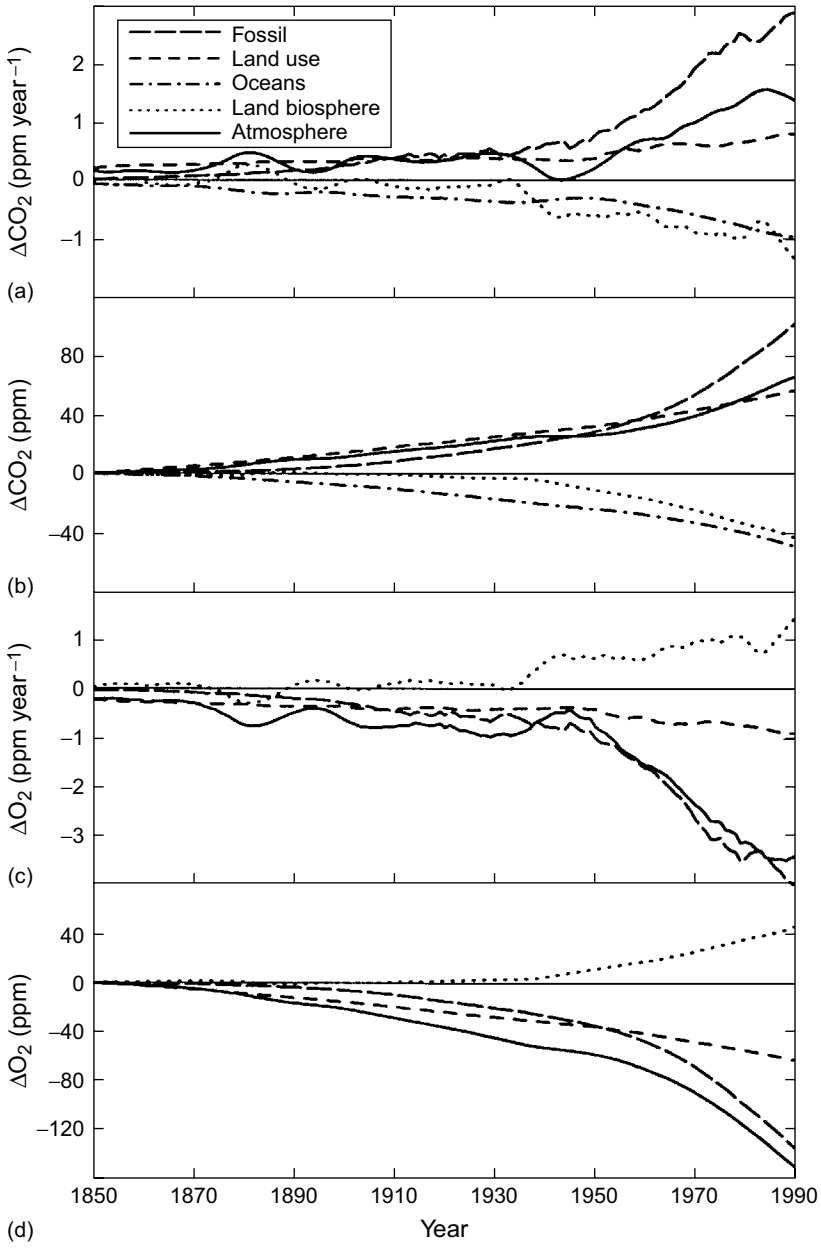


Figure 1 Changes in atmospheric CO_2 and O_2 over the period 1850–1990. CO_2 changes are described in terms of annual (a) and cumulative (b) fluxes. They represent observations of the atmospheric increase from air in Law Dome firn/ice, independent estimates of emissions due to fossil fuels and land use change, modeled oceanic CO_2 uptake and inferred land biospheric exchange not including land use change. Corresponding annual (c) and cumulative (d) O_2 fluxes are calculated assuming no change in oceanic O_2 inventory and known stoichiometric relationships of O_2/CO_2 exchange for terrestrial vegetation and for different fuel types (Keeling, 1988)

to sustain this growth is uncertain. A hypothetical calculation is possible of the maximum depletion that would result from oxidation of all accessible biogenic material. At the end of 1998, proven world fossil fuel reserves amounted to potential O_2 consumption of 600 ppm; however this figure will continue to grow as fuel reserves are boosted by future exploration. Estimates of total carbon stocks (biomass and soils) among four terrestrial biosphere models lie in the range 1150–2300 PgC (Kicklighter *et al.*, 1999), representing a further potential O_2 sink of 600–1200 ppm. If all of the material in both reservoirs were oxidized, atmospheric O_2 would be reduced by upwards of 1200 ppm. This is equivalent to reducing the atmospheric mole fraction from 20.95 to below 20.83%. However, the extent to which the potential loss is realized depends overwhelmingly on energy and land use practices of future generations.

REFERENCES

Battle, M, Bender, M, Sowers, T, Tans, P P, Butler, J H, Elkins, J W, Ellis, J T, Conway, T, Zhang, N, Lang, P, and Clarke, A D (1996) Atmospheric Gas Concentrations over the Past Century Measured in Air from Firn at the South Pole, *Nature*, 383, 231–235.

Battle, M, Bender, M, Tans, P P, White, J W C, Ellis, J T, Conway, T, and Francey, R J (2000) Uptake of Carbon by the Ocean and the Terrestrial Biosphere Inferred from Atmospheric O_2 and $\delta^{13}C$, *Science*, 287, 2467–2470.

Broecker, W S (1970) Man's Oxygen Reserves, *Science*, 168, 1537–1538.

Etheridge, D M, Steele, L P, Langenfelds, R L, Francey, R J, Barnola, J-M, and Morgan, V I (1996) Natural and Anthropogenic Changes in Atmospheric CO_2 over the Last 1000 Years from Air in Antarctic Ice and Firn, *J. Geophys. Res.*, 101, 4115–4128.

Francey, R J, Allison, C E, Etheridge, D M, Trudinger, C M, Enting, I G, Leuenberger, M, Langenfelds, R L, Michel, E, and Steele, L P (1999) A 1000-Year High Precision Record of $\delta^{13}C$ in Atmospheric CO_2, *Tellus*, 51B, 170–193.

Houghton, R A, Davidson, E A, and Woodwell, G M (1998) Missing Sinks, Feedbacks and Understanding the Role of Terrestrial Ecosystems in the Global Carbon Balance, *Global Biogeochem. Cycles*, 12, 25–34.

Keeling, R F (1988) *Development of an Interferometric Oxygen Analyzer For Precise Measurement of the Atmospheric O_2 Mole Fraction*, Ph.D. Thesis, Harvard University.

Keeling, R F, Piper, S C, and Heimann, M (1996) Global and Hemispheric CO_2 Sinks Deduced from Changes in Atmospheric O_2 Concentration, *Nature*, 381, 218–221.

Kicklighter, D W, Bruno, M, Donges, S, Esser, G, Heimann, M, Helfrich, J, Ift, F, Joos, F, Kaduk, J, Kohlmaier, G H, McGuire, A D, Melillo, J M, Meyer, R, Moore, III, B, Nadler, A, Prentice, I C, Sauf, W, Schloss, A L, Sitch, S, Wittenberg, U, and Wurth, G (1999) A First-order Analysis of the Potential Role of CO_2 Fertilization to Affect the Global Carbon Budget: a Comparison of four Terrestrial Biosphere Models, *Tellus*, 51B, 343–366.

Langenfelds, R L, Francey, R J, Steele, L P, Battle, M, Keeling, R F, and Budd, W F (1999) Partitioning of the Global Fossil CO_2 Sink using a 19-Year Trend in Atmospheric O_2, *Geophys. Res. Lett.*, 26, 1897–1900.

Machta, L and Hughes, E (1970) Atmospheric Oxygen in 1967 to 1970, *Science*, 168, 1582–1584.

Trudinger, C M, Enting, I G, Francey, R J, Etheridge, D M, and Rayner, P J (1999) Long-term Variability in the Global Carbon Cycle Inferred from a High-precision CO_2 and $\delta^{13}C$ Ice-core Record, *Tellus*, 51B, 233–248.

Argentine Pampas

see Temperate Grasslands (**Volume 2**)

Aspartate Formers

see Photosynthesis (**Volume 2**)

Atmospheric Oxygen, Anthropogenic Impacts

see Anthropogenic Impacts on Atmospheric Oxygen (**Volume 2**)

Autotroph/Heterotroph

Autotrophs (literally self-nourishing) require only simple inorganic nutrients and manufacture their own basic food requirements using an external energy source. *Heterotrophs* (other, different or outside sources of nourishment) are unable to manufacture the organic molecules they need for energy and for making cellular components and must absorb them from their environment.

Autotrophic organisms may be classified as *photoautotrophs* or *chemoautotrophs*. Photoautotrophs, all photosynthetic plants and bacteria, derive their energy through the absorption of quanta of light. The energy so trapped is converted into chemical potential, which is available for manufacturing the organic nutrients required by the organism. Chemoautotrophs, mainly bacteria, derive their energy requirements from high energy or reduced inorganic chemicals in their environment that are not in themselves essential for the nourishment of the organism, but which are oxidized or otherwise chemically altered by the organism to a lower energy level. The energy so released is stored temporarily as chemical potential so that it may be used for the conversion of primary inorganic foodstuff (usually carbon dioxide) into the nutritional requirements of the organism. This group includes few species, but may be ecologically very important because of the large numbers of organisms and their often major impact on the environment, e.g., sulfur bacteria that oxidize hydrogen sulfide to elemental sulfur.

Heterotrophs, comprising non-photosynthetic plants and the majority of bacteria, are unable to acquire energy directly from such external sources but must absorb organic chemicals. These serve the double purpose of being substrates of respiration, which releases energy that is temporarily stored as chemical potential, and substrates for the synthesis of nutritionally required compounds and cellular materials using the stored chemical potential.

The term autotrophic may also be applied to an ecosystem in which the rate of photosynthesis exceeds that of respiration ($P > R$). Examples include communities that are heavily fertilized (e.g., agricultural or estuarine systems). In a heterotrophic ecosystem the rate of respiration exceeds that of photosynthesis ($P < R$), as may occur in areas of heavy pollution or high sewage deposition (e.g., swamps, cold or polluted streams). Where photosynthesis equals respiration, the ecosystem is said to be stabilized. Most ecosystems are stable (i.e., $P/R = 1$) regardless of their overall rates of photosynthetic production and respiration. A stable ecosystem may be autotrophic in summer and heterotrophic in winter, but the differences balance each other.

R G S BIDWELL *Canada*

AVHRR

see Remote Sensing, Terrestrial Systems (**Volume 2**)

B

BAHC (Biospheric Aspects of the Hydrological Cycle)

see IGBP Core Projects (*Opening essay*, Volume 2)

Benthic/Pelagic

The terms *benthic/pelagic* refer to organisms or situations that are located beyond the littoral or beach area of the sea or a lake. *Benthic organisms* are those that live on or close to the bottom of a lake or sea from the high water mark to the depths; *pelagic organisms* occupy the whole water column from the bottom to the top. Benthic organisms are frequently attached to the bottom (e.g., attached marine algae, mature shellfish), while pelagic organisms are free swimming (e.g., whales, fishes or marine unicellular plants). Pelagic animals are collectively called *nekton*, and pelagic plants (largely unicellular algae or dynoflagellates) are termed *plankton*.

R G S BIDWELL *Canada*

Biocenosis

Biocenosis is a community of interacting organisms (plants, animals, microorganisms, etc.) sharing the same habitat (terrestrial or aquatic). This term was invented by Möbius (1877), a German zoologist, for a community of organisms growing on an oyster reef. Biocenosis is formed as a result of intra- and inter-species competition and natural selection. It is shaped under certain average environmental conditions and any change in this average would lead to transformation of the biocenosis. In addition, the biocenosis would be changed if for some natural or anthropogenic reasons the number of individuals of particular species increased or diminished or one species disappeared, or a new species entered the community. The concept of biocenosis has been widely applied to various communities of living organisms co-existing in one place. The biocenosis is primary if it is formed without human influence (e.g., primary forest, natural grassland) or secondary if humans have altered it (e.g., regrowth after clear-cutting the forest, aquatic community living in artificially created water reservoir). Even an agricultural plantation can be considered as a special case of biocenosis even though it is fully managed by humans. The concept of biocenosis was expanded to biogeocenosis as an independently derived synonym for ecosystem by the Russian plant ecologist Sukachev (1945).

REFERENCES

Möbius, K A (1877) *Die Auster and die Austernwirtschaft*, Verlag von Wiegandt, Berlin.

Sukachev, V N (1945) Biogeocenology and Phytocenology, *Dokl. Akad. Nauk USSR*, **47**, 447–449.

GALINA CHURKINA *Germany*

Biocomplexity

Biocomplexity is a term used to describe the dynamic web of often surprising interrelationships that arise when living things at all levels (from molecular structures to genes to ecosystems) interact with their environment.

The prefix bio refers to life. *The American Heritage Dictionary of the English Language* (1992) defines complex as "consisting of interconnected or interwoven parts". As most life scientists are used to looking at the small picture (cellular interactions, manipulation of a single variable in an experiment, behaviour of a particular species at one site) it may take perspectives from other disciplines, notably the earth sciences and the social sciences, to best explain these interrelationships. "At all scales, from microscopic to megascopic, from atoms to galaxies, natural processes are interconnected", wrote Palmer and Zen (2000). These connections are not necessarily straightforward and linear. Some of them are non-linear and chaotic: especially connections that are related directly to the activities of the human species. These ribbons of interconnections, some patterned, some not, some predictable, some not yet predictable, are the complexities we propose to study in this new century and that we could not study in the last century.

The definition of biocomplexity seems similar to that for ecology: "... the study of the relation of organisms or groups of organisms to their environment ..." (Odum, 1959, p. 4). The study of biocomplexity, however, is greater in scope than the study of ecology and ecosystems. It includes vertically integrated relationships, such as the relationship between the soil, bacteria, plants, herbivores, carnivores, societal systems, geological factors, and climate. Biocomplexity also includes relationships over time, such as events that recur on a regular or even irregular basis. It is a term that includes heretofore-undescribed (perhaps even heretofore-indescribable) relationships that we only now are beginning to have the tools to discover.

REFERENCES

Odum, E P (1959) *Fundamentals of Ecology*, W B Saunders, Philadelphia, PA.

Palmer, A R and Zen, E (2000) Toward a Stewardship of the Global Commons: Engaging "My Neighbor" in the Issue of Sustainability, Part 5. Earth Systems: the Connectedness of Everything, *GSA Today*, **10**(5), 10 [http://bcn.boulder.co.us/basin/local/sustain5.htm].

RITA R COLWELL *USA*

Biodiversity, Functional

see Functional Biodiversity (*Opening essay*, **Volume 2**)

Biodiversity in Freshwaters

Christian Lévêque
CNRS-Programme Environment, Meudon, France

Inland water ecosystems are among the world's most fragile, scarce and threatened ecosystems. Probably more than any other area on the planet, inland waters are the focus of conflicting human uses and resultant pressures. Inland waters have been exposed to a variety of increasing stress such as water extraction for domestic, agricultural and industrial uses, pollution (organic and inorganic), fishing, introduction of alien species, habitat alterations in relation to water management, etc. Even in pristine lakes far from human influences, long-range atmospheric transport may sometimes cause deposition of acidic and potentially toxic substances, e.g., sulfates and polychlorinated biphenyls, respectively, into the lakes. This may cause a long-term decline in biodiversity. These stressors have affected, and continue to affect life in inland water ecosystems all over the world.

BROAD CHARACTERISTICS OF FRESHWATER BIODIVERSITY

Research on inland waters biodiversity offers special challenges:

1. River drainages or lake catchments have been compared to terrestrial islands isolated from one another. So for taxonomic groups unable to cross geographic terrestrial barriers such as fish, mollusc, macrocrustacea, etc., the distribution is patchy. Most inland water fish, for example, can only colonize adjacent basins episodically when climatic or geological conditions allow connections between tributaries. This has a number of implications:

    - inland water biodiversity is usually highly localized, with high levels of endemism, particularly evident in ancient lakes, which have been isolated for millions of years such as the great east African Lakes (Tanganyika, Malawi, Victoria), Lake Baikal, Lake Biwa, Lake Ohrid, etc., (Martens *et al.*, 1994);
    - this patchy distribution results in great genetic variability between isolated sub-populations;
    - many inland water species must survive *in situ* or will risk extinction when exposed to climatic and ecological changes, or when threatened by the impact of human activities. So they have to adapt to the changing environment by speciation and/or

physiological adaptation, or to move when possible to more suitable habitats or refuge zones, or to perish if there is no other possibility.

2. Inland waters contain a tremendous diversity of vertebrates, microorganisms, invertebrates and aquatic plants which are among the most poorly understood organisms on Earth and are now seriously threatened. There is little doubt that inland water fishes represent the most threatened set of vertebrates. Some 25–30% of all vertebrate diversity is concentrated into some 0.01% of the planet's water or about 1–2% of the land surface area. This includes the 10 000 identified species of freshwater fish (compared to 15 000 marine fish) (Nelson, 1994) and other vertebrates including many amphibians, reptiles, birds, mammals, etc., which closely rely on the existence of inland water habitats to implement their life cycle.
3. Freshwater habitats, and particularly shallow habitats, are widely considered to be transient in time and space. Climate and geomorphology determine the global distribution and diversity of freshwaters. The great majority of existing lakes, of which around 10 000 exceed 1 km^2 in extent, are geologically very young and occupy basins formed by ice masses or glacial erosion during recent ice ages (Gorthner, 1994). Many of them occur in the temperate zone of the Northern Hemisphere, and date from the retreat of continental ice-sheets some 10 000 years BP (before present). Only about 10 existing lakes are known to be much older which occupy large subsidence basins, such as Lake Baikal or lakes of the East African Rift. River systems can change course radically as a result of deposition and erosion of their channel, but large rivers rarely disappear.

The present biodiversity in freshwater ecosystems may be viewed dynamically as an heritage of the past climatic changes and evolution processes, in a water body constrained by present climate and geomorphology, and whose future rely on the impact of several natural or anthropogenic factors (Figure 1).

1. Inland water systems tend to be the first habitats to experience degradation since humans have historically congregated around water sources. As a consequence of this anthropogenic pressure, the habitats are subjected to a considerable range of stressors, including damming and canalization, reclamation of floodplains, water abstraction, pollution, loading with organic matter and detritus. In addition, activities pursued within the drainage basin (e.g., land use or deforestation) may have consequences on the aquatic ecosystem. It should also be noted that aquatic environments are recipients of virtually every form of human waste, resulting in rapid and continuous degradation. The increasing demand for inland water resources generated by population growth will most likely result in the further decline of inland water biota unless action is taken urgently.

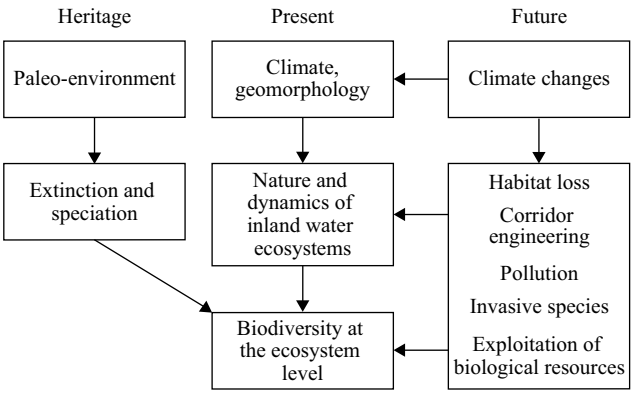


Figure 1 General factors in uencing the dynamics of freshwater diversity

INVENTORY OF FRESHWATER BIODIVERSITY

The global number of species-level biota, animals, plants and microorganisms that occur in fresh water is not precisely known. Knowledge is more complete for vertebrates than for invertebrates, while a few groups of invertebrates (molluscs, crustacea) have been more completely investigated than others. The knowledge is also more complete within the temperate zone than in the tropics.

About 100 000 species of organisms associated with freshwater sediments have been described (Table 1) but the true number is much higher than this (Palmer *et al.*, 1997). The number of species in most taxa can scarcely be estimated and global estimate of microbial diversity remains controversial. For example, some specialists estimate that there are hundred of thousands of aquatic nematodes and only a few percent of these have been described.

The best-known aquatic animals are fish. They exhibit enormous diversity in size, shape, and biology. There are descriptions of an estimated 24 600 valid species of fishes compared with 23 550 tetrapods and the eventual number of extant fish species may be close to 28 500 (Nelson, 1994). Presently, some 10 000 species of fish are found only in freshwaters, and a further more than 500 marine species also use freshwater. The freshwaters are therefore disproportionately rich in species of fish on the basis of volume and area when compared to oceans. That could be viewed as the result of the patchy nature of inland waters. Due partly to their behavior and partly to history, freshwater fishes tend to become isolated within drainage basins, resulting in distinctive populations and subspecies. This isolation

Table 1 Species richness of the freshwater sediment biota for several habitat types. Numbers are very approximate and derived from many sources. (Data collected by Palmer *et al.*, 1997)

Taxon	No species described	Probable number of species	Range of local species richness
Bacteria	>10 000	Unknown	>1000
Algae	14 000	>20 000	0- 1000
Fungi	600	1000- 10 000	0- 300
Protozoa	<10 000	10- 20 000	20- 800
Plants	1000	Unknown	0- 100
Invertebrates	70 000	>100 000	10- 1000
Aschelminthes	4000	>10 000	5- 500
Annelida	1000	>1500	2- 50
Mollusca	4000	5000	0- 50
Acarii	5000	>7500	0- 100
Crustacea	8000	>10 000	5- 300
Insecta	45 000	>50 000	0- 500
Others	1400	>2000	0- 100

favors genetic diversity; and the evolutionary potential of particular phylum such as cyprinodonts, cyprinids or cichlids also explains the high degree of endemicity observed in freshwater fishes.

In contrast with fish that are highly endemic, many flagellates, ciliates and rhizopods appear to be cosmopolitan, at least when considered at the morphospecies level. At higher trophic levels, macrozooplankton, especially many rotiferan and small cladoceran species, resemble phytoplankton in potentially occurring continentally or globally within bands of latitude (Green, 1994).

STATUS OF FRESHWATER BIODIVERSITY

Due to a general lack of data, it is difficult to assess the status of inland water biodiversity. However, evidence of biological impoverishment is pervasive in aquatic systems and when available, the data are very disturbing. For example, in the USA, where the status of aquatic biodiversity is well documented in comparison with other areas, 34% of fish, 65% of crayfishes and 75% of unionid mussels are classed rare to extinct (Master, 1990; Williams *et al.*, 1992). Of 214 stocks of Pacific salmon, 74% have a high or moderate risk of extinction (Nehlsen *et al.*, 1991). As a whole, the native fishes of North America are in serious decline and currently 364 taxa are listed by the American Fisheries Society as endangered, threatened or of special concern (Williams *et al.*, 1989). In California, Moyle (1995) reports a rising trend of extinction and endangerment in the region's native fishes from 24%, in 1988, to 43%, in 1992. Nearly half of Mexico is arid or semiarid with scarce waters. At least 92 springs and 2500 km of river have dried in this area. There are nearly 200 species of freshwater fishes, 120 under some threat, 15 extinct through human impact. As of 1985, an average of 68% of species was eradicated in local fish faunas (Contreras and Lozano, 1994).

Whatever the causes, inland water species are especially vulnerable to global changes since many of them are confined to individual watersheds and cannot readily disperse to undisturbed areas. Distinct species assemblages, highly adapted and specialized, have evolved in a wide array of habitats now threatened, and it has been estimated that 20% of the world's freshwater fish are either endangered or recently extinct (Moyle and Leidy, 1992).

MAJOR THREATS

The various anthropogenic factors that impact upon freshwater systems can be classified according to spatial scale and the location of effects (Boon, 1992) (Table 2). However in most cases, wherever impacts have been investigated and changes in biological diversity demonstrated, multiple factors are involved.

HABITAT ALTERATION

Habitat alteration is one of the major causes of loss of the diversity of aquatic life, and degradation or destruction of habitats is particularly threatening in rivers. Flow regulation occurs on almost every large river. Man-made lakes associated with hydroelectric dams or built for irrigation purposes prevent fish migrations, and alter the flow pattern downstream, one of the major consequences being the disappearance of floodplains and spawning grounds for different species. Canalization of the streambed reduces diversity of habitats. Catchment changes from large-scale land-use practices associated with deforestation, give rise to erosion, and increase of sediment load in river waters, as do

Table 2 Scale and source of factors impacting freshwater biodiversity in rivers (Boon, 1992)

Spatial scale	Source of impact
Supra catchment effects	Acid rain
	Inter-basin water transfers
Catchment land-use change	Deforestation, and afforestation
	Urbanization
	Agricultural development
	Land drainage
Corridor engineering	Flood protection
	Flow regulation; dams, weirs,
	Channelization
	Riparian vegetation removal
	Dredging, mining
Instream impacts	Organic and inorganic pollution
	Thermal pollution
	Abstraction
	Navigation
	Exploitation of native species
	Introduction of alien species

mining-related activities. In this context, fish communities are affected not only by events occurring in the channel, but also by several external influences occurring in the catchment.

Deforestation severely affects aquatic communities because massive quantities of sediment eroded from clear-cut watersheds are eventually discharged into freshwater systems, where they can be extremely destructive to aquatic organisms. Excess sediment loading in standing water bodies reduces light penetration and thereby photosynthetic rates, both planktonic and benthic. Herbivorous fish may be directly affected through reduced foraging efficiency, while zooplanktivores may also be indirectly affected by a decrease in the herbivorous zooplanktonic organisms on which they prey.

As an example, in Lake Tanganyika, 40–60% of the original woodland or forest land has been cleared in the central portions of the lake's drainage basins, and almost 100% around the northernmost portion of the lake. Steep slopes, heavy rainfall, generalized cultivation without protecting the land and slopes, and the abundance of torrential mountain streams are all factors generating rapid headwater erosion and stream incision which generate a massive increase of suspended sediment and sedimentation rates in the nearshore region of the lake. The rate of outbuilding of the Ruzizi River Delta, the major drainage pathway in the northern end of the lake, has probably increased by an order of magnitude in the past 20 years over its rate prior to deforestation. There are many fewer species of diatoms, ostracods and fish, in highly disturbed portions of Lake Tanganyika, than in moderately disturbed or undisturbed areas.

Water diversion may also be responsible for drastic changes in aquatic habitats. Among well-known examples is Lake Aral fed by the Amu Daria and Syr Daria rivers, which was the fourth largest inland body of water, just behind Lake Superior. To grow cotton, an immense irrigation system was built, diverting water from the rivers. The consequence is that the lake itself has shrunk to 25% of its 1960 volume, and its salinity is higher than the ocean leading to the extirpation of most of the 24 fish species (Létolle and Mainguet, 1993) (*see* **Aral Sea**, Volume 4).

INTRODUCTION OF EXOTIC SPECIES

Species introduction into aquatic systems have been, and are, commonplace. Welcomme (1988) listed 237 inland animal species which have been introduced into 140 countries worldwide, but the total number is much higher.

Many exotic species have been voluntarily or accidentally introduced in aquatic systems for a variety of purposes. The main goals of voluntary introductions, most usually decided by fishery officers, were initially to improve sport fisheries, artisanal fisheries and aquaculture, or to develop biological control of aquatic diseases, insects, and plants. More recently, transfers of ornamental plants and fish for the aquarium trade have also sharply increased as well as accidental introductions of these species into freshwater ecosystems.

Aquarium trade seems to be responsible for the introduction of many of the aquatic weeds in the world. In most tropical areas, *Eicchornia crassipes*, the water hyacinth native from South America, has been deliberately introduced in many countries as an ornamental plant for garden ponds. The species as well as the aquatic fern *Salvinia molesta*, also from South America, are now widely distributed through the tropical and subtropical regions of the globe where they are considered as major aquatic pests. Both are free floating plants with a high degree of morphological plasticity, competitive ability and tolerance of a wide range of environmental conditions. In temperate countries, the Canadian species *Elodea canadensis* has invaded canals in Europe and has had a significant effect on the aquatic ecosystems, preventing the full development of the native species.

Fish introductions are usually better documented than other aquatic taxa. The first introductions are believed to date from Roman times, and at least 260 species have been introduced in European waters since 1850 (Cowx, 1997). Approximately one-third of these species have been introduced in the 1960s and 1970s. In North America, the introduction of fish began in the late 1600s and the establishment of exotic species in US freshwaters is rising sharply: in 1920 only six species of exotic origin were established; by 1945, just three more have been added; the big boom of introductions occurred after 1950, and by 1980,

35 exotics had become established, and approximately 50 others had been recorded (Allan and Flecker, 1993). The future promises a continuing spread of exotic species, and managers of aquatic ecosystems will be confronted increasingly with a shifting mix of native and non-native species. Fish introductions are now an added route to invaders, and may be viewed against the background and concomitant increase of massive global changes in type and extent, and alteration in quality of freshwater habitats caused by human activities worldwide.

Invading species have been responsible for major ecological changes. These changes include extinction of native species which are more likely to occur when the successful invader is a top carnivore (Lévêque, 1997a). Direct competition for space or food with the indigenous fauna, as well as the impact of predation of introduced fish species, has been relatively well documented. Moyle *et al.* (1986) called it the Frankenstein effect because the impacts of introduced species on the native indigenous fauna tend to be negative in unpredictable ways. A well-known example is the introduction of *Lates niloticus* (the Nile perch) to lakes Victoria and Kyoga during the 1950s and early 1960s to improve the fisheries. The establishment of the Nile perch was initially supported by the large quantities of native haplochromines which were their main prey (Lévêque, 1997b). With the increasing *Lates* stock, the haplochromines declined rapidly and were drastically reduced.

Transport through ballast water is probably one the most important pathway for alien species invasions in several places, including the North American Great Lakes (Mills *et al.*, 1993). That is the case for the zebra mussel introduced into the Great Lakes, apparently in 1985 or 1986, which spread dramatically throughout the waterways of both Canada and the US with serious economical and ecological consequences. The recent finding of individual mitten crabs (*Eriocheir sinensis*) and the establishment of the alien gastropod *Potamopyrgus antipodarum* in Lake Ontario in 1995 demonstrate that the process of invasion is still going on at a fast rate.

One of the major problems in freshwater species introductions is their irreversibility, at least on scale of a human's lifetime. Once introduced and established, it is impossible, given current technology, to eradicate a fish, mollusc or plant species from a large natural water body. As a consequence, we are likely to see a continued reduction in native aquatic biodiversity and an increased homogenization of the world's freshwater biotas.

CONSEQUENCES OF CLIMATE WARMING FOR INLAND WATER ECOSYSTEMS

The Earth's average temperature will likely increase by 1–3 °C over the next century. Climate changes is likely to affect key processes (precipitation, evaporation, water vapor transport) that determine the amount and distribution of freshwater ecosystems. The information that is important to predicting future changes comes from the detailed study of climatic and biotic changes of the relatively recent past of the quaternary in which the geographical and biological setting is most closely similar to the present. Over the past 25 000 years, freshwater ecosystems have undergone massive changes in spatial extent correlated with trends in regional climate. Future climate change is likely to produce more or less comparable changes in the supply and distribution of freshwaters. However, while the climate has always been changing, current changes are unique in that they are superimposed upon a landscape, which has already been greatly altered by human activities. It is widely recognized that climate change will exert some additional stress on ecosystems that are already threatened by other human activities, such as increasing resource demands, unsustainable management practices and pollution.

The scientific review for the Intergovernmental Panel on Climate Change (IPCC) Second Assessment Report concluded that climate change will lead to an intensification of the global hydrological cycle. However, it is difficult to provide a general overview of how climate will change on a regional basis. We can expect:

- evaporation will lead to a drop in lake levels, and some wetlands will disappear while some of the smaller lakes may become wetlands;
- changes in the total amount, frequency and intensity of precipitation directly affecting the magnitude and timing of runoff, the intensity of floods and droughts, and the accumulation of nutrients and contaminants, etc.; prediction of variability is still more uncertain than predictions of mean values;
- increased runoff in high latitude regions due to increased precipitation, whereas runoff may decrease at lower latitudes due to the combined effects of increased evapotranspiration, and decreased precipitation;
- an increase in sediment runoff as a result of changes in the terrestrial ecosystems surrounding the rivers;
- an increase in the probability of dry days and the length of dry spells (consecutive days without precipitations) in some areas. Where mean precipitation decreases, the likelihood of drought increases.

Rivers will undergo changes in annual and seasonal runoff, and runoff variability. They may expand or contract. The Colorado river may see its water flow halved in the future and some rivers in Western Australia might have a 45% reduced flow, whereas the Ganges and Brahmaputra rivers in Bangladesh could increase by 50% or more. Changes in hydrological factors, such as increased or decreased frequency and magnitude, and changes in timing of high and low discharge events, will affect the ecology of streams and rivers in numerous ways. Current velocity

and turbidity of streams will change and if that leads to altered flow during spawning periods, spawning and nursery areas of migratory fishes may be unavailable. Warming may result also in faster drying out of temporary waters, increasing the challenge they present to aquatic life. With a water temperature increase, a decrease of dissolved oxygen content is expected, and this will have an immediate and direct effect on fish, particularly those living in shallow warm waters. In lakes, climate warming could result in higher surface temperature, longer ice-free periods, and an extended interval of thermal stratification.

In order to evaluate the anticipated effects of climate change on freshwater fish and their habitats, Regier and Meisner (1990) sketch out an iterative assessment process that uses water temperature, water quantity, and water quality variables. Temperature affects all vital processes, including activity, feeding, growth and reproduction. According to recent studies (Hill and Magnuson, 1990), growth of fish is expected to increase with climate warming if other factors now limiting growth also change with climate. But climate warming is also expected to increase productivity at all aquatic trophic levels by about 10–20% per 1 °C increase in temperature (Regier *et al.*, 1990). Another likely long-term consequence of global climate change is the modification of the genetic composition of populations and species. Selection for heat tolerance can quickly lead to the evolution of tolerant genotypes that may have altered life-history traits.

CONCLUSION

Research on the ecological aspects of global change should contribute to basic ecological understanding of processes regulating the Earth's biota. But the ultimate goal for ecologists is to be in a position to assist decision-makers in devising policies to anticipate, ameliorate or respond to global environmental change. Although much of the literature of aquatic ecology is relevant to global change, little work has focused explicitly on the consequences of global change for freshwater ecosystems (Carpenter *et al.*, 1992). Opinions regarding the necessary responses to global change range from those economists who think mankind should simply use his technology to adapt to changes as they occur to those who advocate the expensive option of reducing carbon dioxide (CO_2) emissions and thus preventing of further global change.

Actually, freshwater systems response to global change will depend on a complex array of physical and biological factors. These changes are likely to directly affect the survival, reproduction and growth of organisms, as well as the distribution, persistence and diversity of species:

- Global change may affect climate principally through temperature and precipitation regimes (amount, annual distribution, etc.). Both may have consequences on the overall distribution of freshwater biota (Tonn, 1990), as well as on the physiology and biology of individual species (Regier *et al.*, 1990).
- The conjunction between climate change and anthropogenic impacts could threaten more species than either factor alone. Habitat destruction and artificial barriers will obviously prevent many fish from colonizing new habitats when their native one is threatened. There is a general view that biological communities will change and shift in complex and unpredictable ways as the geographical distributions of species are altered individually rather than in community units. The rate of species invasion and extinction is likely to accelerate further, bringing about complex changes in species compositions and interactions.

Most of the concepts and strategies for biodiversity conservation were designed for terrestrial ecosystems and little has been done for the conservation of aquatic ecosystems. Rivers, particularly, raise specific problems due to their longitudinal structure and it is difficult to establish protected areas in a system for which the ultimate determinant of the quality and quantity of water is the catchment from which it originates. Habitat protection and restoration through a holistic ecosystem approach are the only major long-term means through which successful conservation of aquatic biodiversity will be achieved. Maintaining the integrity of aquatic systems contributes to providing harvestable fish populations, maintaining biological diversity, and helping to control the flux of nutrients from the land to lakes and seas. *See also* **Freshwater Fisheries**, Volume 2.

REFERENCES

Allan, J D and Flecker, A S (1993) Biodiversity Conservation in Running Waters, *BioScience*, **43**, 32–43.

Boon, P J (1992) Essential Elements in the Case for River Conservation, in *River Conservation and Management*, eds P J Boon, P Calow, and G E Petts, John Wiley & Sons, London, 11–33.

Carpenter, S R, Fisher, S G, Grimm, N B, and Kitchell, J F (1992) Global Change and Freshwater Ecosystems, *Annu. Rev. Ecol. Syst.*, **23**, 119–139.

Contreras, B S and Lozano, V M L (1994) Water, Endangered Fishes and Development Perspectives in Arid Lands of Mexico, *Conserv. Biol.*, **8**, 379–387.

Cowx, I G (1997) L'introduction D'espèces de Poissons dans les Eaux Douces Européennes: Succès Économiques ou désastres Écologiques?, *Bull. Fr. Peche Piscic.*, **344/345**, 57–77.

Gorthner, A (1994) What is an Ancient Lake? *Speciation in Ancient Lakes*, eds K Martens, B G oddeeris, and G Coulter, *Arch. fur Hydrobiol.*, **44**, 97–100.

Green, J (1994) The Temperate-Tropical Gradient of Planktonic Protozoa and Rotifera, *Hydrobiologia*, **272**, 13–26.

Hill, D K and Magnuson, J J (1990) Potential Effect of Climate Warming on the Growth and Prey Consumption of Great Lakes Fishes, *Trans. Am. Fish. Soc.*, **119**, 265–275.

Létolle, R and Mainguet, M (1993) *Aral*, Springer-Verlag, Paris.

Lévêque, C (1997a) *Biodiversity Dynamics and Conservation, The Freshwater Fish of Tropical Africa*, Cambridge University Press, Cambridge.

Lévêque, C (1997b) Fish species Introductions in African Freshwaters, in *Stocking and Introduction of Fish*, ed I G Cowx, Fishing New Books, Oxford, 234–257.

Martens, K, Goddeeris, B, and Coulter, G, eds (1994) Speciation in Ancient Lakes, *Adv. Limnol.*, **44**, 1–508.

Master, L (1990) The Imperilled Status of North American Aquatic Animals, *Biodivers. Network News*, **3**(1–2), 7–8.

Mills, E L, Leach, J H, Carlton, J T, and Secor, C L (1993) Exotic Species in the Great Lakes: a History of Biotic Crises and Anthropogenic Introductions, *J. Great Lakes Res.*, **19**, 1–54.

Moyle, P B, Li, H W, and Barton, B A (1986) The Frankenstein Effect: Impact of Introduced Fishes on Native Fishes in North America, in *Fish Culture in Fisheries Management*, ed R H Stroud, American Fisheries Society, Bethesda, MD, 415–426.

Moyle, P B (1995) Conservation of Native Freshwater Fishes in the Mediterranean-type Climate of California, USA; a review, *Biol. Conserv.*, **72**, 271–279.

Moyle, P B and Leidy, R A (1992) Loss of Biodiversity in Aquatic Systems: Evidence from Fish Faunas, in *Conservation Biology: The Theory And Practice of Nature Conservation, Preservation, and Management*, eds P L Fiedler and S K Jain, Chapman and Hall, New York, 127–169.

Nehlsen, W, Williams, J E, and Lichatowichn, J A (1991) Pacific Salmon at the Cross Roads; Stocks at Risk from California, Oregon, Idaho and Washington, *Fisheries*, **16**, 4–21.

Nelson, J S (1994) *Fishes of the World*, 3rd edition, John Wiley & Sons, New York, 1–600.

Palmer, M A, Covich, A P, Finlay, B J, Gibert, J, Hyde, K D, Johnson, R K, Kairesalo, T, Lake, S, Lovell, C R, Naiman, R J, Ricci, C, Sabater, F, and Strayer, D (1997) Biodiversity and Ecosystem Processes in Freshwater Sediments, *Ambio*, **26**, 571–577.

Regier, H A and Meisner, J D (1990) Anticipated effects of climate change on freshwater fishes and their habitat, *Fisheries*, **19**(6), 10–15.

Regier, H A, Holmes, J A, and Pauly, D (1990) Influence of Temperature Changes on Aquatic Ecosystems: an Interpretation of Empirical Data, *Trans. Am. Fish. Soc.*, **119**, 374–389.

Tonn, W M (1990) Climate Change and Fish Communities: a Conceptual Framework, *Trans. Am. Fish. Soc.*, **119**, 337–352.

Welcomme, R L (1988) *International Introductions of Inland Aquatic Species*, FAO Fisheries Technical, Rome, Italy, Paper 294, 318.

Williams, J D, Warren, M L, Cummings, K S, Harris, J L, and Neves, R J (1992) Conservation Status of Freshwater Mussels of the US and Canada, *Fisheries (Bethesda)*, **18**, 6–22.

Williams, J E, Johnson, J E, Hendrickson, D A, Cintreras-Balderas, S, Williams, J D, Navarro Medoza, M, McCallister, D E, and Deacon, J E (1989) Fishes of North America: Endangered, Threatened, or of Special Concern, *Fisheries*, **14**, 2–20.

Biodiversity in Soils and Sediments: Potential Effects of Global Change

Gina A Adams and Diana H Wall
Colorado State University, Fort Collins, CO, USA

Activities of soil and sediment organisms, from their feeding on plant roots, their decomposition of plant and animal detritus and release of nutrients, and their movement through soil and sediment layers, regulate vital processes that yield goods and services (e.g., water purification, renewal of soil fertility) essential to the continued functioning of ecosystems and society. The biological diversity of the organisms in soils and sediments is estimated to be orders of magnitude greater than that above the surface, but is very poorly known. The significance of this vast below-surface biodiversity for ecosystem functioning is also unknown.

Recent assessments indicate that global change is reducing species biodiversity on landscape and global scales, and may cause the extinction of up to two-thirds of the world's species (Pimm and Brooks, 2000). In terrestrial and freshwater systems, habitat loss is the primary cause of these extinctions, but changing atmospheric composition and invasions of non-native species will become increasingly significant. However, most of these assessments are based on biodiversity above the surface of soils and sediments. If we are to accurately predict biodiversity change with global change and consequences for ecosystem functioning, we must rapidly increase our knowledge of distribution and functioning of species within soils and sediments.

Species richness is a surrogate measure of biodiversity that is commonly used for above-surface organisms. However, in below-surface habitats fewer studies have identified organisms to the species level, because of difficulties associated with their microscopic size and high abundance. Therefore, assessments of below-surface biodiversity consider additional measures, including diversity of functional groups (assemblages of species with similar ecological roles) and abundance of individuals (which influences genetic diversity).

This contribution outlines potential effects of global change on the species and functional group abundance and diversity of organisms that dwell in soils and freshwater sediments. It includes organisms that spend their entire life cycle below the surface and those that migrate out of soils or sediments to forage for food (e.g., ants) or as part of their life cycle (e.g., planktonic larval stages of freshwater bivalves). It reveals the significant gaps in our knowledge and offers some recommendations for research priorities.

BIODIVERSITY OF SOILS AND FRESHWATER SEDIMENTS

The dark and opaque domains of soils and freshwater sediments remain largely unexplored by science. While an estimated 13% of species across all habitats have been described (Hammond *et al.*, 1995), most of these are from above-surface domains. Although approximately 50% of larger soil invertebrates such as termites and ants have been described (Wall *et al.*, 2001) most smaller organisms remain undescribed. It is estimated that just 0.025–5% of nematode species have been described, and the majority of those are animal parasites, the relatively small proportion of nematode species that dwell above surface (Baldwin *et al.*, 2000). Similarly, although a single gram of soil may contain 20 000–40 000 bacterial species (Brussaard *et al.*, 1997) only approximately 13 000 bacterial species (Torsvik *et al.*, 1994) have been described, the majority of which are not soil or sediment dwelling. Table 1 shows the number of species described for some of the major phyla within soils and sediments.

Obtaining a clear picture of the diversity of organisms within soils and sediments is a daunting task. Each sample of soil or sediments collected can contain thousands of species of bacteria and fungi and hundreds of species of protozoa, and other small invertebrates (e.g., rotifers, tardigrades, nematodes) many of which are new, undescribed forms. However, new technologies, including molecular techniques and 3D imaging, are helping to advance our understanding by enabling increasingly rapid species identification.

Table 1 The approximate number of species described for some of the major taxonomic groups in soils and freshwater sediments (Palmer *et al.*, 1997; Wall *et al.*, 2001)

	Soils	Freshwater sediments
Bacteria		>10 000
by molecular analysis:	13 000	
by culture techniques:	1718	>10 000
Fungi	18 000- 35 000	600
Algae	Unknown	14 000
Protozoa	1500	10 000
Nematoda	5000	Unknown
Insecta	e.g., ants 8800 e.g., termites 2000	45 000
Annelida	e.g., earthworms 3600	1000
Mollusca	Unknown	4000

Table 2 Ecosystem services regulated by organisms within soil and sediments (modified from Daily, 1997)

- Removal of organic detritus through decomposition
- Release of nutrients for plants and maintenance of soil and sediment fertility through decomposition
- Detoxification of wastes and pollutants by microbes with specialized metabolisms
- Generation and maintenance of soil and sediment structure, e.g., by the burrowing activities of large invertebrates and particle-aggregating properties of microbes
- Filtration and cleansing of water by direct removal of particles (e.g., by filter feeders) and by maintaining the soil and sediment structure that filters water
- Modification of the hydrological cycle
- Erosion and sedimentation control
- Mitigation of floods and droughts
- Control of pests of agriculture and fisheries through predation and competition
- Regulation of concentrations of atmospheric gas (e.g., nitrogen, carbon dioxide, methane, oxygen) through consumption and production activities such as photosynthesis, aerobic and anaerobic respiration and specialized functions (e.g., nitrification and nitrogen fixation)
- Regulation of major biogeochemical cycles such as carbon and nitrogen cycles, through consumption and storage of organic matter and release of its inorganic constituents

ACTIVITIES OF SOIL AND SEDIMENT ORGANISMS IN ECOSYSTEMS

Differences in the types, rate and magnitude of the activities of organisms in soils and sediments contribute to the variety of processes that sustain ecosystems and will influence an ecosystem's response to global change (Table 2). Examining how individual species influence ecosystem functioning is a monumental challenge. One approach has been to reduce the diversity by grouping species with similar morphologies, physiologies and food sources into functional groups. This allows analysis of how biodiversity influences ecosystem functioning, but does not allow effects of individual species to be determined.

Evidence from a few studies indicates that species differences within functional groups can be important determinants of ecosystem processes and response to change. In a Puerto Rican stream, for example, Covich *et al.* (1999) found that two species of benthic macroinvertebrates (shrimp) have specific and complementary roles in the processing of organic matter, and that carbon retention in the headwater of streams is most efficient when both species are present. Where processes are thus regulated by specific species, loss of a single species may have a dramatic effect (e.g., formation of soil porosity by an earthworm species, certain transformations within the nitrogen cycle by a specialized bacterium). However, where numerous species serve identical functions in ecosystem processes, the loss of a species will have an undetectable effect on the process and there will be considerable species redundancy (e.g., formation of soil structure at the microscopic scale

by bacteria and fungi; microbial decomposition of common organic compounds). Improved knowledge of species-level effects on ecosystem processes is vital if we are to predict which ecosystems will be most vulnerable to reduced biodiversity associated with global change.

GLOBAL CHANGE AND SOIL AND SEDIMENT BIODIVERSITY

Drivers of global change (land use change, atmospheric change, climate change and biotic invasions) will influence the primary determinants of biodiversity within soils and sediments, namely:

1. physico-chemical properties (e.g., texture, salinity, pH);
2. resource supply (e.g., vegetation, organic matter and nutrients);
3. biotic interactions (e.g., herbivory, parasitism, predation);
4. microclimate (e.g., temperature, moisture) (Wall *et al.*, 2001);
5. disturbance regime (changes in rate and magnitude of disturbance that determine whether an ecosystem is in a non-equilibrium state populated by colonizing species, or in equilibrium populated by a climax community of species).

The performance of some species will be enhanced and others impaired and the overall response of species diversity will depend on the balance of the differential species responses.

Land use change has the most extensive and pernicious impacts on the determinants of biodiversity (Wall *et al.*, 2001). The impacts of atmospheric change (elevated carbon dioxide, nitrogen deposition, UV-B and ozone) and associated climate change are predicted to rise in the next 50–100 years. Biotic introductions of non-native species into novel habitats will also increase as exchange of materials increases with global travel and commerce.

The effect of the global change drivers on determinants of soil or sediment biodiversity will be direct and indirect (Table 3). Direct effects include altered soil structure through land use change (urbanization and cultivation), acidification and enhanced UV-B radiation through atmospheric change, altered temperature and moisture content of soils and sediments through climate change and altered biotic interactions such as predation and competition from species invasions. Indirect effects of global change drivers on soil and sediment biodiversity result from changes to above-surface vegetation and thus altered amount, quality, timing and diversity of plant inputs to soils and sediments and altered microclimate (through altered shading and precipitation interception). Currently the largest influence on vegetation is land use change, including deforestation, intensive agriculture and urbanization, but large effects as a result of altered atmospheric composition and climate are predicted.

The impact of global change drivers on soil and sediment biodiversity will vary according to biome type, geographical location, and spatial and temporal scales. Understanding and predicting this variability is a major research challenge, but some generalizations can be made:

- Species with short generation times such as microbes will adapt faster to change than longer-lived, larger organisms (macrofauna).
- Species that can migrate between moist and dry patches

Table 3 Direct and indirect mechanisms by which the major drivers of global change affect, or are predicted to affect in the future, biodiversity in soils and sediments

	Soils	Freshwater sediments
Land use change		
Direct:		
Loss of habitat	+	+
Disturbance	+	+
Pollution	+	+
Indirect:		
Removal/alteration of plants and animals above-surface	+	+
Atmospheric change		
Direct:		
Acidi cation	+	+
UV-radiation	+	+
Indirect:		
Altered physiology of plant communities	+	+/?[a]
Altered composition of plant communities	+	+/?[a]
Climate change		
Direct:		
Loss of habitat (ooding of soils, drying of sediments)	+	+
Moisture	+	+
Temperature	+	+
Increased disturbance by extreme climatic events	+	+
Indirect:		
Altered physiology of plant communities	+	+/?[a]
Altered composition of plant communities	+	+
Invasive species		
Direct:		
Altered biotic interactions	+	+

[a] +/?, atmospheric change, including elevated CO_2 and nitrogen deposition, and climate change will alter the physiology and/or composition of vegetation in watersheds with consequences for freshwater sediment biota. However, studies have not examined whether the physiology and composition of submerged aquatic vegetation will be altered.

of soil will be less susceptible to soil drying with global warming.

- Species with large-scale migration or dispersal mechanisms will be better able to find new habitats in response to land use or climate change. Currently, our knowledge of soil and sediment biodiversity is confined to local or patch scales. Until knowledge of biogeographical distributions and migration abilities is increased, our ability to predict large-scale shifts in species distribution with global change will remain severely constrained.
- Currently, species may be most impacted in tropical regions where land use change is greater than temperate regions, but future global warming is predicted to be greatest at high latitudes.
- At the global scale, biodiversity will be reduced due to habitat loss and homogenization of landscapes. At the local patch scale, biodiversity responses will vary depending on site-specific conditions.

LAND USE CHANGE

Species diversity of all soil taxa decreases precipitously with land use change, as a result of altered soil texture, microclimate and timing and amount of vegetation inputs (Wall *et al.*, 2001). Organisms that tolerate a limited range of soil texture (e.g., termites, earthworms and enchytraeids), vegetation quality (e.g., species-specific herbivores or symbionts such as some mycorrhizae), or climate may be most affected, while those with broad tolerances (e.g., some bacteria) may be less affected.

In developing countries forest is rapidly being converted to agricultural land. Deforestation and replacement by pasture can reduce species diversity of soil macroinvertebrates by three-quarters. Conversion to agriculture can reduce soil biodiversity through reduced plant diversity, increased chemical inputs, increased pests and invasive species, depletion of soil organic content, soil compaction and erosion. The agricultural management practices used (e.g., slash and burn, rotations, conventional vs. no-till, minimal vs. high chemical inputs) strongly influence the impacts on soil biodiversity. For example, reduced tillage enhances soil fungal, microarthropod and earthworm populations compared with conventional tillage.

Freshwater sediment biota are also dramatically impacted by land use change. Humans currently appropriate 54% of the Earth's accessible freshwater runoff and this could reach 70% by 2025 (Lake *et al.*, 2000). As a result, the direction and amount of water flow is altered and freshwater habitats including their sediments, especially wetlands, are being rapidly lost.

Agriculture and urbanization practices in watersheds also impact sediment biota through sedimentation, acidification, salinization, contamination and altered timing, amount and quality of plant litter and dissolved organic carbon inputs. Watershed practices, such as clear-cutting, that alter the structure and composition of riparian plant communities, can have major consequences for sediment faunal composition. Reduced log and debris inputs to streams eliminate natural log dams that are a major site of invertebrate reproduction, reducing invertebrate populations and their processing of organic material with impacts on other decomposer species. Altered timing and amount of leaf inputs to streams alters the species composition of communities that shred leaf material in the first stages of decomposition.

Inefficient fertilizer application to agricultural lands decreases soil biodiversity and, through non-point source pollution and eutrophication, reduces biodiversity of freshwater sediments. Eutrophication increases phytoplankton density (but not diversity) and decomposition of this increased biomass consumes available oxygen, leading to anaerobic conditions. The diverse aerobic sediment fauna is replaced by a low diversity community consisting of anaerobic microbes and anoxia-tolerant macrofauna (Lake *et al.*, 2000).

Increased sedimentation, from soil erosion through agriculture, forestry, and urbanization reduces diversity and abundance of sediment biota, since it blocks the spaces between streambed particles that they inhabit (Lake *et al.*, 2000). It also blocks the transport of nutrients from groundwaters to sediments and transport of oxygenated water to the groundwater and hyporheic zone (sediments below running waters). The sediments of the hyporheic zone provide nursery beds for many insect and fish species and refugia for species following disturbance and its deoxygenation can severely reduce species abundance and diversity.

ATMOSPHERIC CHANGE

Direct effects of increasing atmospheric CO_2 concentrations on soil and sediment biodiversity are predicted to be small, because of the large concentrations already below-surface (Wolters *et al.*, 2000). However, indirect effects of increased atmospheric CO_2 concentrations on soil and sediment species may be large, as a result of plant responses to CO_2 enrichment. Elevated CO_2 concentrations can increase plant growth, carbon : nitrogen ratio, root : shoot ratio and water use efficiency. Differential responses of plant species and plant functional groups to elevated CO_2 are predicted to alter composition of plant communities. The altered plant physiology and community composition may affect soil and sediment species by altering organic resources, microclimate, habitat complexity and occurrence of plant host species. Data on the response of soil biota to altered plant community composition under elevated CO_2 are lacking, since most CO_2 enrichment studies are short-term and therefore

measure responses to altered plant physiology rather than community composition. However, there is strong circumstantial evidence for large effects on soil species. For example, plant species differ in their palatability to soil decomposer organisms; and relationships between soil symbiotic mutualists and their plant hosts can be highly specific.

The response of soil biota to elevated CO_2 has been generally considered for functional groups rather than species. Abundance of individuals within decomposer functional groups may increase or show no response to elevated CO_2 (Wall *et al.*, 2001). Fungal biomass can increase more than bacterial biomass, with corresponding increases in fungal-feeding microarthropods (Rillig *et al.*, 1999). Diversity of microbial communities has been altered under elevated CO_2 in some studies (e.g., beneath Mediterranean plant communities (Dhillion and Roy, 1997)) but not others (e.g., beneath grass (Hodge *et al.*, 1998)). The few studies that have examined species level responses of soil biota to elevated CO_2 provide evidence for differential species responses, e.g., with fungal and collembola (selective grazers on fungi) communities (Jones *et al.*, 1998).

Effects of elevated CO_2 on the biodiversity of freshwater sediments are as yet unknown. Altered physiology and composition of riparian and watershed vegetation may influence sediment biodiversity through similar mechanisms as for soil biodiversity. Submerged macrophytes may only show a response to elevated CO_2 in freshwaters where CO_2 is currently limited, such as soft waters.

Atmospheric sulfur, nitrogen and nitrous oxides from industry, agriculture and vehicles can acidify and leach cations (e.g., Mg^{2+}, Ca^{2+}) from soils and sediments. Diversity and composition of soil and sediment species is influenced by both the increased acidity and reduced cation availability. Aquatic organisms, including those that occupy the water film around soil particles and organisms that depend on calcium availability for shell development (e.g., crustaceans and mollusks) may be particularly affected. Following long-term acidification of the Tatra Mountains of the Slovak Republic, abundance and distribution of 'acidophilic' collembollan species increased, while some 'calciophilic' species disappeared. Across the landscape, collembolan species composition became more homogenous and less diverse (Rusek, 1993). Acidification can alter the dominance of foodwebs, for example, shifting the dominance of the soil microflora from bacterial to fungal and, correspondingly, reducing abundance and species diversity of omnivorous and predatory nematodes more than that of root-feeding and fungal-feeding nematodes. In acidified freshwaters, macrophytes are commonly lost and phytoplankton and zooplankton diversity is reduced. Impacts on plant-feeding and decomposer organisms within sediments are likely but have not been quantified.

Atmospheric nitrogen deposition also fertilizes ecosystems. The fertilization effect is modified by the production potential of the vegetation; in ecosystems adapted to low nitrogen availability, plant productivity is increased and plant diversity usually reduced and reductions in soil biodiversity may be expected. However, following long-term nitrogen deposition to the Tatra Mountains, sphagnum bogs adapted to extremely low nitrogen inputs were converted to more productive and species-rich grasslands, with a corresponding increase in abundance and species diversity of the soil collembolan community (J Rusek, personal communication).

Reduced concentrations of stratospheric ozone are increasing UV-B radiation to the Earth's surface. The effects of UV-B on soil and sediment biota are less studied than the effects of elevated CO_2, nitrogen deposition or acidification but enhanced UV-B radiation is likely to influence soil and sediment biodiversity indirectly through altered timing, quantity and quality of plant growth. UV-B radiation does not penetrate far into soils or sediments, so its greatest direct effects will be on those organisms that live at the surface. Enhanced UV-B radiation was shown to kill earthworm populations in approximately seven hours (Sicken *et al.*, 1999). In a freshwater lake, abundance of herbivorous invertebrate chironomid larvae declined under enhanced UV-B radiation (Bothwell *et al.*, 1994). Algae increased under the reduced grazing pressure with potential for effects along the food chain.

CLIMATE CHANGE

Predictions of future climate include a mean global warming of $1-3.5\,^{\circ}C$, increased precipitation with regional variation, and increased severity and frequency of catastrophic climatic events. Climate change will affect soil and sediment biota directly by influencing temperature, soil moisture, wet-drying and freeze-thawing cycles and disturbance regimes, and indirectly via altered vegetation. It is important to remember that the distribution of many soil and sediment species is limited by their cold tolerance and thus an increase in minimum temperatures, as well as maximum temperatures, may have strong effects.

Soil biota show considerable variability in their responses to warming, but less is known about the response of sediment species. In soils, the interaction between temperature and moisture, and the organisms' ability to withstand desiccation or flooding may be the key determinant of their response to warming. In Arctic regions where moisture was not limiting (tundra), collembola and oribatid mites were both resistant to warming of $2-5\,^{\circ}C$ over 3 years. However, where moisture was limiting (semipolar desert), the abundance of collembola, but not oribatid mites, declined under the elevated temperatures (Coulson *et al.*, 1996).

In the dry regions, species within these two groups also showed differential responses to warming. In another study, Briones (1997) showed that only the enchytraeid species that were able to migrate to moist patches or tolerate dry conditions withstood extreme summer temperature conditions at an artificially warmed site. These studies suggest that less mobile or less desiccation-tolerant taxa will be reduced or lost from soils that experience drying with global warming, affecting faunal composition and reducing diversity.

Migration and dispersal abilities will also influence the response of sediment species to warming on a landscape scale. In streams and rivers with intact north-south connectivity and unhindered migration paths, cold water species in high latitude and high altitude systems may be replaced by warm water species, resulting in new species assemblages. In streams and rivers with east-west drainage and in isolated ponds and lakes, thermal refuges may not be available and while some genetic adaptation may occur, there will be local extinction of species.

Increased severity and frequency of extreme climatic events, such as floods and droughts, will lead to localized extinctions of soil and sediment species. Droughts and floods comprise the disturbance regime in streams, rivers and floodplains; these control colonization by invaders and regulate nutrient and organic matter input. Increased flood severity can remove refuges that are vital for retaining reservoirs of species for recolonization. Increased flood frequency can impede successful recolonization. A study in a Californian stream demonstrated that reduced flooding can also reduce diversity (Power, 1995). When droughts reduced disturbance from flooding, grazing invertebrates persisted, suppressing spring algal blooms. As a result, a less complex foodweb was supported. Reduced waterflow in droughts also reduces hydrologic linkages between streams and groundwater zones, reducing nutrient and oxygen turnover and significantly stressing biota in groundwater and hyporheic zones. Droughts also reduce habitat area and fragment streams, which can impair migration and recolonization of species.

Climate change is predicted to alter physiology, composition and distribution of vegetation. Northward expansion of vegetation activity is occurring and may be accompanied by local extinctions of plant species at low latitudes. For plant species that cannot migrate, widespread extinctions will occur, as the fossil record shows occurred with global warming at the end of the Pleistocene. Migration and extinction of plant species with climate change may influence soil biodiversity in a number of ways. Differences in migratory rates of plants and soil biota, including plant symbionts (such as mycorrhizae, nitrogen-fixing bacteria and parasites) and free-living organisms, will disrupt biotic interactions such as herbivory, parasitisms, mutualisms, and decomposition. Whether or not this decoupling leads to extinction of soil and sediment species, or formation of new community assemblages will depend on the specificity of the relationship. Below-surface species that are closely co-evolved with above-surface species, e.g., some associations between plants, fungi and pathogenic nematodes and viruses may become locally extinct.

INVASIVE SPECIES

Animals, plants and microbes invading novel ecosystems have no natural competitors or predators and populations can explode, displacing native species and disrupting ecosystem functioning within soils and sediments. Plant invasions result from intentional planting in agriculture and horticulture, or through unintentional transport. Invasive soil species can be transported in soil attached to plants, animals or agricultural and construction machinery and through soil erosion with wind, irrigation and flooding. Invasive sediment species can be imported through flooding, in ballast water of ships or as parasites and gut inhabitants of species in aquaculture.

In the USA, invasion of the forest understory plant species *Berberis thunbergii* and *Microstegium viminium* altered soil organic and nutrient status, encouraging higher populations of earthworms (Kourtev *et al.*, 1998). Invasions of macrofauna (e.g., ants and earthworms) can alter soil organic matter, aeration, water infiltration and chemistry (Wall *et al.*, 2001) and reduce diversity of native invertebrates. The Argentine ant (*Linepithema humile*) is now invasive on all continents except Antarctica and eliminates native ant species and dramatically reduces diversity and alters composition of non-ant invertebrates (Wall *et al.*, 2001). In Northern Europe, invasions of Australasian planarian flatworms, *Artioposthia triangulata* are decimating earthworm populations on which they prey, with potentially devastating effects on soil porosity and hydrology (Boag *et al.*, 1997).

One of the most notorious invasive species is the zebra mussel *Dreissina polymorpha*, which has stripped phytoplankton from several North American lakes and rivers, disrupted foodwebs, and shifted the frequency and composition of algal blooms and nutrient cycling (Strayer *et al.*, 1999). Populations of native bivalves drop sharply following zebra mussel invasion but populations of other sediment invertebrates may increase as a result of more structurally complex habitat and increased organic matter deposition. Another invasive freshwater species, the crayfish *Orconectes rusticus* has reduced the diversity of sediment gastropods in Wisconsin lakes. It overgrazed macrophyte beds, removing this important refuge for gastropod species. Only gastropod species with thick shells or a burrowing ability to escape predators tended to survive (Covich *et al.*, 1999).

INTERACTIVE EFFECTS

Soil and sediment species will be exposed to multiple drivers of global change simultaneously, whose effects may be additive, synergistic or antagonistic. For practical reasons, most studies have examined single drivers but analysis of their interactions is an urgent research need, since multiple global change drivers can interact to alter the intensity of the change. For example, Gorham (1996) describes how acidification, climate change or removal of wetlands can reduce colored dissolved organic compounds in the water, increasing UV-B penetration to lake sediments. Studies need to be expanded to examine the effects of simultaneous global changes on the biota.

CONCLUSIONS

Global change will have varied effects on soil and sediment biota, but generally will result in reduced abundance and diversity. Differential responses of functional groups and their species will alter competitive relationships and will likely result in large shifts in community composition.

In general we cannot yet predict when and where global change will have minor vs. catastrophic impacts on soil and sediment biodiversity and ecosystem functioning. To advance this goal, research priorities should include:

- identifying those ecosystem processes that are performed by a few species and are thus most vulnerable to species loss;
- determining the rate, magnitude and variability of global change to which these keystone species or functional groups can adapt;
- routinely considering linkages between soils and sediments, since neither exists in isolation and indeed significant areas of the Earth's surface are an interface between the two;
- studying multiple effects of global change drivers to analyze how responses are modified by the presence of other species in a community;
- studying the interactive effects of different global change drivers.

REFERENCES

Baldwin, J G, Nadler, S A, and Freckman, D W (2000) Nematodes – Pervading the Earth and Linking All Life, in *Nature and Human Society: The Quest for a Sustainable World*, ed P R Raven, National Academy Press, Washington, DC, 176–191.

Boag, B, Jones, H D, and Neilson, R (1997) The Spread of the New Zealand Flatworm (*Artioposthia triangulata*) within Great Britain, *Eur. J. Soil. Biol.*, **33**, 53–56.

Bothwell, M L, Sherbot, D M J, and Pollock, C M (1994) Ecosystem Response to Solar Ultraviolet-B Radiation: Influence of Trophic-Level Interactions, *Science*, **265**, 97–100.

Briones, M J I (1997) Effects of Climate Change on Soil Fauna; Responses of Enchytraeids, Diptera Larvae and Tardigrades in a Transplant Experiment, *Appl. Ecol.*, **6**, 110–117.

Brussaard, L, Behan-Pelletier, V M, Bignell, D E, Brown, V K, Didden, W, Folgarait, P, Fragoso, C, Freckman, D W, Gupta, V, Hattori, T, Hawksworth, D L, Klopatek, C, Lavelle, P, Malloch, D W, Rusek, J, Soderstrom, B, Tiedje, J M, and Virginia, R A (1997) Biodiversity and Ecosystem Functioning in Soil, *Ambio*, **26**, 563–570.

Coulson, S J, Hodkinson, I D, Webb, N R, Block, W, Bale, J S, Strathdee, A J, Worland, M R, and Wooley, C (1996) Effects of Experimental Temperature Elevation on High-Arctic Soil Microarthropod Populations, *Polar Biol.*, **16**, 147–153.

Covich, A P, Palmer, M A, and Crowl, T A (1999) The Role of Benthic Invertebrate Species in Freshwater Ecosystems: Zoobenthic Species Influence Energy Flows and Nutrient Cycling, *Bioscience*, **49**, 119–127.

Daily, G C (1997) Nature's Services *Societal Dependence on Natural Ecosystems*, Island Press, Washington, DC.

Dhillion, S S and Roy, J (1997) Soil Microbial Functional Diversity Responds to Plant Diversity and Clevated CO_2, *Bull. Ecol. Soc. Am.*, **78**, 240.

Gorham, E (1996) Lakes Under a Three-Pronged Attack, *Nature*, **381**, 109–110.

Hammond, P M, Hawksworth, D L, and Kalin-Arroyo, M T (1995) Magnitude and Distribution of Biodiversity: 3.1. The Current Magnitude of Biodiversity, in *Global Biodiversity Assessment*, ed V H Heywood, Cambridge University Press, Cambridge, 113–138.

Hodge, A, Paterson, E, Grayston, S J, Campbell, C D, Ord, B G, and Killham, K (1998) Characterization and Microbial Utilization of Exudate Material from the Rhizosphere of *Lolium perenne* Grown Under CO_2 Enrichment, *Soil Biol. Biochem.*, **30**, 1033–1043.

Jones, T H, Thompson, L J, Lawton, J H, Bezemer, T M, Bardgett, R D, Blackburn, T M, Bruce, K D, Cannon, P F, Hall, G S, Hartley, S E, Howson, G, Jones, C G, Kampichler, C, Kandeler, E, and Richie, D A (1998) Impacts of Rising Atmospheric Carbon Dioxide on Model Terrestrial Ecosystems, *Science*, **280**, 441–443.

Kourtev, P S, Ehrenfeld, J G, and Huang, W Z (1998) Effects of Exotic Plant Species on Soil Properties in Hardwood Forests of New Jersey, *Water Air Soil Pollut.*, **105**, 493–501.

Lake, P S, Palmer, M A, Biro, P, Cole, J, Covich, A P, Dahm, C, Gibert, J, Goedkoop, W, Martens, K, and Verhoeven, J (2000) Global Change and the Biodiversity of Freshwater Ecosystems: Impacts on Linkages between Above-Sediment and Sediment Biota. *BioScience*, **50**(12), 1099–1107.

Palmer, M A, Covich, A P, Finlay, B L, Gilbert, J, Hyde, K D, Johnson, R K, Kairesalo, T, Lake, P S, Lovell, C R, Naiman, R J, Ricci, C, Sabater, F F, and Strayer, D L (1997) Biodiversity and Ecosystem Processes in Freshwater Sediments, *Ambio*, **26**(8), 571–577.

Pimm, S L and Brooks, T M (2000) The Sixth Extinction: How Large, When and Why? in *Nature and Human Society.*

The Quest for a Sustainable World, ed P H Raven, National Academy Press, Washington, DC, 46–62.

Power, M E (1995). Floods, Food Chains, and Ecosystem Processes in Rivers, in *Linking Species and Ecosystems*, eds C G Jones and J H Lawton, Chapman and Hall, New York, 52–60.

Rillig, M C, Field, C B, and Allen, M F (1999) Soil Biota Responses to Long-Term Atmospheric CO_2 Enrichment in Two California Annual Grasslands, *Oecologia*, **119**, 572–577.

Rusek, J (1993) Air-Pollution-Mediated Changes in Alpine Ecosystems and Ecotones, *Ecol. Appl.*, **3**, 409–416.

Sicken, O, Tiemann, M, Schmelz, R M, Kestler, P, and Westheide, W (1999) Lethal Effects of UV-radiation on Lumbricid Earthworms (*Eisenia fetida* and *Lumbricus terrestris*), *Pedobiologia*, **43**, 874–879.

Strayer, D L, Caraco, N F, Cole, J J, Findlay, S, and Pace, M L (1999) Transformation of Freshwater Ecosystems by Bivalves: A Case Study of Zebra Mussels in the Hudson River, *Bioscience*, **49**, 19–27.

Torsvik, V, Goksoyr, J, Daae, F L, Sorheim, R, Michelsen, J, and Salte, K (1994) Use of DNA Analysis to Determine the Diversity of Microbial Communities, in *Beyond the Biomass*, eds K Ritz, J Dighton, and K E Giller, Wiley-Sayce Publication, 39–48.

Wall, D H, Adams, G, and Parsons, A N (2001) Soil Biodiversity Under Global Change Scenarios in *Future Scenarios of Global Biodiversity*, eds F S Chapin, III and O E Sala, Springer-Verlag, in press.

Wolters, V, Silver, W L, Bignell, D E, Coleman, D C, Lavelle, P, van der Putten, W H, de Ruiter, P, Rusek, J, Wall, D H, Wardle, D A, Brussaard, L, Dangerfield, J M, Brown, V K, Giller, K E, Hooper, D U, Sala, O, Tiedje, J, and van Veen, J A (2000) Effects of Global Changes on Above- and Belowground Biodiversity in Terrestrial Ecosystems: Implications for Ecosystem Functioning, *BioScience*, **50**(12), 1089–1098.

Biodiversity: The UNEP De nition

Biodiversity is the variability among living organisms from all sources including, *inter alia*, terrestrial, marine and other aquatic ecosystems and the ecological complexes of which they are part; this includes diversity within species, between species and of ecosystems.

FURTHER READING

Article 2: Use of Terms, Convention on Biological Diversity, UNEP, Nairobi, see http://www.biodiv.org/convention/articles.asp.

R E MUNN *Canada*

Biogeochemical Cycle

Biogeochemical cycle describes the cyclic passage through the biotic, atmospheric, aqueous and lithic (rock, soil) compartments of the environment of inorganic nutrient chemicals that enter and pass through the biotic component. Biogeochemical cycles are described as *gaseous* (i.e., carbon dioxide (CO_2), nitrogen) or *sedimentary* (i.e., nutrients that are largely present in rocks and soils). This distinction is not absolute: CO_2 has major lithic as well as gaseous components, and sulfur, primarily sedimentary, has minor gaseous forms. The cycling of water, the *hydrological cycle*, is usually considered independently from biogeochemical cycles.

Nutrient cycling is important for two main reasons: first, nutrients are important to organisms, their importance varying with their abundance, accessibility and the level of their requirement, and second, nutrients may have roles in ecosystem behavior quite unrelated to their obvious biological significance. For example, CO_2 is primarily important to the biotic component of an ecosystem as the substrate of photosynthesis, but it is also involved in environmental pH regulation, particularly in aquatic habitats. In addition, its role as a greenhouse gas may have major environmental and ecological consequences. These different roles interact to modify ecosystem behavior, and hence the behavior of the CO_2 cycle. The fact that major global environmental concerns (e.g., stratospheric ozone depletion, the greenhouse effect, acid rain, nitrogen and phosphorus overfertilization) are closely interrelated through the operation of biogeochemical cycles has greatly increased current interest in the behavior of these cycles.

Biogeochemical cycling of carbon involves its removal from the air or water by photosynthesis and its return by respiration/decomposition of plants, animals and bacteria. Carbon may be stored temporarily in soils or water, and more permanently in carbonaceous rocks such as limestone. Its transfer between water and air is complex because it is affected by many environmental factors (e.g., temperature, pH, water circulation and flow, winds and air movement). Large amounts of carbon are also transferred erratically to the air by fires and volcanism.

Nitrogen cycling is largely biological, involving the bacterial fixation of nitrogen gas, its conversion to organic (reduced) nitrogen in plants and thence animals, and its subsequent loss by bacterial decay to nitrates or nitrogen gas. Of growing concern is the increasing production of nitrogenous compounds to the atmosphere by automobiles, industrial pollution and power plants, and the large contribution to the environment from agricultural fertilizer and wastes. Additional nitrogen also enters the cycle from the production of nitrates or nitrites by lightning and volcanism.

The cycling of lithic nutrient chemicals is less complex, although the sequestering of sedimentary materials

(as well as nitrogen) in litter adds a highly variable biotic component. Cycles of sedimentary nutrients may also be affected by mass transfer throughout ocean ecosystems and their ability to be sequestered or released by lithic components (e.g., phosphorus in rocks, earth, marine deposits, bones, etc.).

See also: **Hydrologic Cycle**, Volume 1; **Carbon Cycle**, Volume 2; **Nitrogen Cycle**, Volume 2; **Phosphorus Cycle**, Volume 2.

R G S BIDWELL *Canada*

Bioindicators

Lesley Hughes
Macquarie University, New South Wales, Australia

Hotter summers, heavier rain, and rising seas – just a few of the predictions as to how the world will change over the coming decades. Indeed, evidence is accumulating that global mean temperatures have already risen by over half a degree since 1860 and that the distribution of rain and snow is also changing. Shrinking sea ice, melting glaciers and warming oceans also provide evidence that steadily increasing concentrations of greenhouse gases are starting to have measurable impacts.

These atmospheric and climatic changes are expected to have profound effects on species and communities in the future. But has the modest warming recorded so far already had an impact on natural systems? Recent analyses of long-term datasets indicate that the answer to this question is a cautious yes. Although natural variation, or non-climatic factors, may be responsible for some of the published trends, the proposition that human-induced atmospheric change is the cause is convincing for some ecologists. The data come from such diverse sources as the isotope ratios of whale baleen, the emergence dates of butterflies, and sophisticated satellite imagery. Some examples are more convincing than others due to the quantity of data, strength of the trend, or a lack of more plausible explanations. No single study can be interpreted as unequivocal evidence for human-induced change, as none represents a controlled experiment and alternate hypotheses cannot be completely discounted. But what started as a trickle of published trends a decade ago has now built to a modest flood.

This article reviews evidence that some species are already responding to the anomalous climate of the last few decades. These responses range from changes in physiology and growth to movements of range boundaries and alterations in the timing of life cycles. A few species even appear to have undergone genetic adaptation as a response to the changing climate. The wait to assess the impact of the enhanced greenhouse effect may be over sooner than we thought.

WHAT DO WE EXPECT?

Ways in which enhanced greenhouse warming can affect the biosphere can be broadly summarized into four categories (Figure 1).

1. *Adaptation*: species with short generation times and rapid population growth rates may undergo micro-evolutionary change at a sufficient pace to keep up with climatic shifts.
2. *Effects on physiology*: changes in atmospheric carbon dioxide (CO_2) concentration, temperature, or precipitation will directly affect processes such as photosynthesis, respiration, growth and tissue composition in plants, and metabolic and developmental rates in many animals.
3. *Effects on distributions*: a 3 °C change in mean annual temperature corresponds to a shift in temperature zones of approximately 300–400 km in a north–south direction (in the temperate zone) or 500 m in elevation. Some species are therefore expected to move upward in elevation or toward the poles as the climate changes.
4. *Effects on phenology*: temperature frequently acts as a trigger for life cycle events such as flowering and fruiting in plants, and hatching in insects and birds. Warmer spring temperatures, in particular, may result in some events occurring earlier in the season.

Any of the above changes to an individual species will inevitably alter its interactions with other species (Figure 1) and eventually lead to changes in the composition and structure of communities. It seems likely that at least some species will become extinct, either as a direct result of physiological stress, or via interactions with other species.

EVIDENCE FOR MICRO-EVOLUTIONARY CHANGE

The anticipated rate of climate change over the next few decades is far greater than any observed in the past. While laboratory experiments have shown that some bacteria populations rapidly adapt to changes in temperature by selection on new mutations, it seems unlikely that most higher eukaryotes, with smaller population sizes and longer generation times, can evolve rapidly enough to track climate change. There is evidence, however, that at least two insect species may have already undergone micro-evolutionary

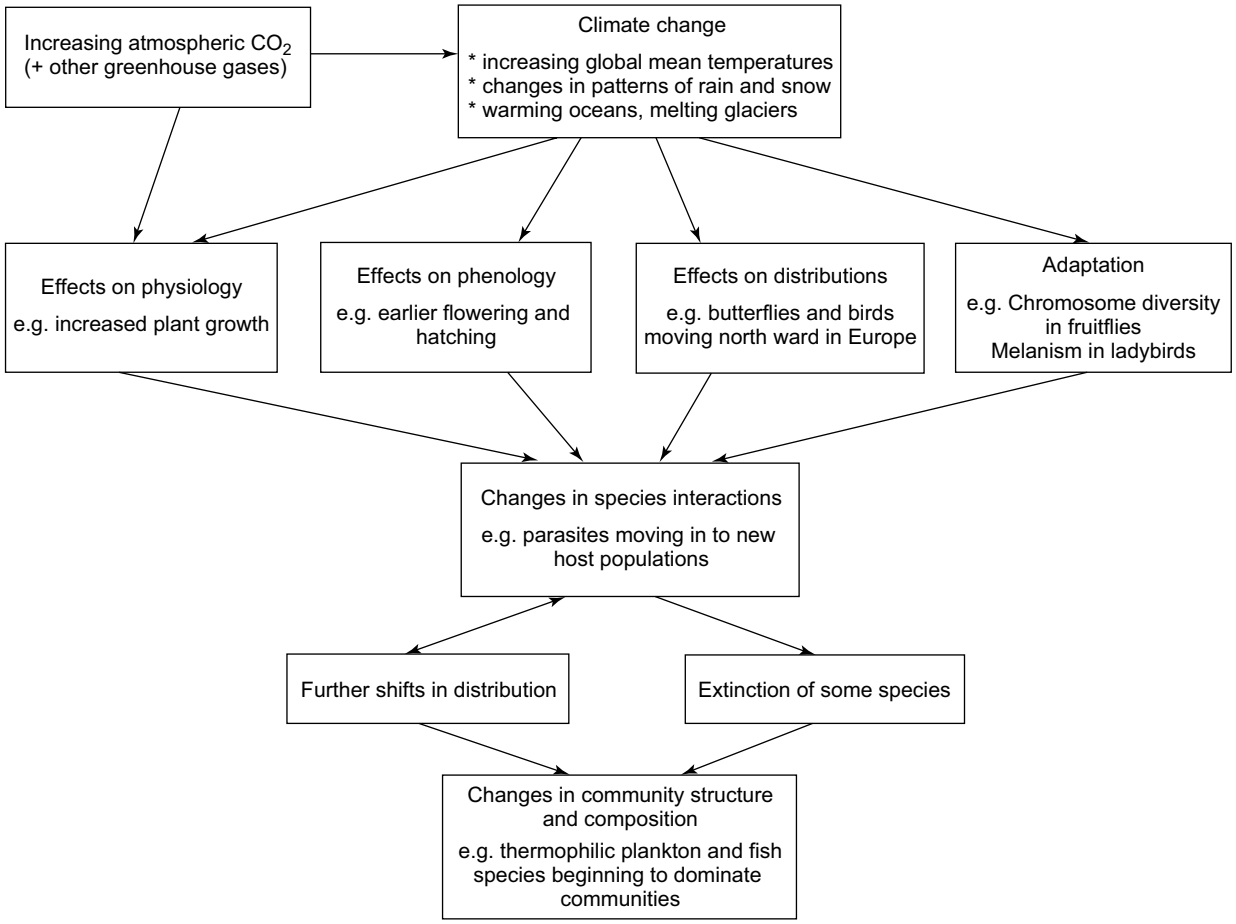


Figure 1 Potential consequences of the enhanced greenhouse effect for species and communities. (Reproduced from Hughes, 2000)

adaptation to the relatively modest warming experienced so far.

The fruitfly, *Drosophila subobscura*, has been sampled over a 16-year period (1976–1980 and 1988–1991) (>80 generations) at a site in Spain that has experienced sustained warming since the mid 1970s. Of particular interest is the *O* chromosome of these flies, which has been linked to thermo-tolerance related traits, and contains a gene family that exhibits direct responses to selection on heat tolerance in the laboratory. A significant reduction (18.3%) in chromosomal diversity in the flies has been detected over the sampling period, with particularly rapid changes in several chromosome inversion polymorphisms. These results are in accord with what would be expected from known latitudinal and seasonal variation in these genetic components (Rodríguez-Trelles and Rodríguez, 1998).

The two-spot ladybird *Adalia bipunctata* has several color morphs that can be broadly classified as predominantly red with black spots (melanics) and predominantly black with red spots (non-melanics). Laboratory experiments show that the melanic forms have a thermal advantage over non-melanics at low temperatures. Over the last two decades there has been a decline in the frequencies of the melanic form in ladybird populations across the Netherlands associated with warmer spring temperatures (de Jong and Brakefield, 1998).

EVIDENCE FOR CHANGES IN PHYSIOLOGY, PRODUCTIVITY AND GROWTH

Both temperature and atmospheric CO_2 concentration directly affect photosynthesis, and hence plant growth and productivity. Plant growth in the Northern Hemisphere has apparently responded to recent trends in warming and CO_2, although it is difficult to separate the relative contributions of the two factors.

Atmospheric CO_2 concentration has been increasing since the mid-1800s. CO_2 concentration rises in winter and declines in summer in response to the seasonal growth of terrestrial vegetation. Since the early 1960s,

the amplitude of this oscillation has increased by 20% in Hawaii and by 40% in the Arctic (Keeling *et al.*, 1996). Increasing assimilation of CO_2 by land plants is one explanation for this trend, and this idea is supported by satellite data showing increased plant growth and a lengthening of the active growing season (1981–1991) in the Northern Hemisphere between 45–70 °N (Myneni *et al.*, 1997). Further support that CO_2-fertilization has already affected tree growth comes from studies of tree-rings at sites in both hemispheres. Increasing CO_2 levels may also have affected the anatomy of leaves. Several studies have shown that the densities of stomates in plants collected recently are significantly lower than in herbarium specimens of the same species collected 70–200 years ago. These trends are consistent with experimental studies using pre-industrial CO_2 concentrations, and with the fossil record.

Some of the recent changes in terrestrial vegetation have important implications for the rate of future atmospheric and climate changes. For example, atmospheric changes and warming are thought to be responsible for the increasing yields of conifer plantations recorded since the mid-1800s in parts of Europe, and for the acceleration of turnover rates and biomass of tropical trees since the 1950s. If CO_2 fertilization is already promoting forest growth, which in turn is sequestering carbon, future climate change might be partially mitigated. A recent report however, that the tundra may have recently changed from being a net carbon sink to a source, is a potential countertrend that may accelerate future change (Serreze *et al.*, 2000).

Recent climate changes may also be affecting productivity in the oceans. Large changes in phytoplankton abundance have been recorded since the late 1940s in the northeast Atlantic and North Sea. Phytoplankton season length and abundance south of 59 °N have increased, while the opposite trend is seen to the north of this latitude. The northern decline may be a result of unusually cold waters spreading from the Arctic where increased freshwater export from melting ice and permafrost is occurring. Further evidence for a decline in phytoplankton productivity at far north latitudes comes from carbon isotope ratios in the baleen plates of bowhead whales in the Bering Sea. Isotope ratios in the zooplankton on which the whales feed, and, by extension, the phytoplankton, are incorporated into keratin deposited into the growing baleen plate. Baleen plates from whales harvested by Alaskan Inupiats from 1947–1997 indicate that seasonal productivity has been declining since the 1970s, with average $\delta^{13}C$ ratios decreasing by over 2.7%, inferring a drop in seasonal carbon fixation of ~30–40% (Schell, 2000). This implies that the carrying capacity of these oceans has been progressively declining, a trend correlated with declines in the top consumers of the region such as sea birds, fur seals and harbor seals.

EVIDENCE FOR CHANGES IN SPECIES DISTRIBUTION AND ABUNDANCE

Human activity over the last millennium has altered the distributions and/or abundance of most species, with most changes being attributed to habitat loss or alteration. Some recent shifts in species distribution toward the poles, or upward in elevation, however, are more convincingly explained by recent climatic trends. Not surprisingly, most of these examples come from species such as alpine and arctic plants whose distributions are most obviously limited by climate, and from organisms that are highly mobile at some stage of their life cycle such as flying insects, birds, and marine invertebrates.

Arctic and Alpine Plants

A substantial warming trend in Antarctica since the 1960s has been associated with dramatic increases in abundance of the only two native vascular plants in the region, *Colobanthus quitensis* (Antarctic pearlwort) and *Deschampsia antarctica* (Antarctic hair grass). At Galindez Island, for example, *D antarctica* increased from 500 individuals in 1964, to 12 030 in 1990 as a result of increased seed germination and successful establishment of seedlings. Similar patterns for these species have been recorded at many other sites (Smith, 1994).

Alpine plants appear to be moving upward, with young trees establishing at elevations or latitudes beyond the current treeline in many regions. Most locations in western North America, for example, show an upward expansion of the forest margin after 1890. Upward colonization of a range of alpine species in the European Alps also appears responsible for increasing plant species richness measured in the early 1990s, compared with historical records.

Invertebrates

Butterflies in both Europe and North America are apparently expanding their ranges northward. In Europe, these expansions range from 35–240 km. For many species, the southern boundaries of the distribution have remained stable, resulting in an overall range expansion. In North America, a survey of over 150 previously recorded populations of Edith's checkerspot butterfly, *Euphydryas editha*, found that locations where the butterflies had successfully persisted were, on average, 2° farther north or at higher elevations than those where populations had become extinct (Parmesan, 1996).

Recent range shifts in other arthropods have serious implications for human health. Increases in mosquito-borne

diseases have been reported in the highlands of Asia, Central Africa and Latin America. Dengue fever, previously limited to about 1000 m in elevation in the tropics, has appeared at 1700 m in Mexico, and *Aedes aegypti*, a vector of dengue fever and yellow fever viruses, has recently been reported at 2200 m in Colombia. *Plasmodium falciparum* malaria is a growing public health threat in the New Guinea highlands, and in 1997 malaria was reported for the first time up to 2100 m in the highlands of Irian Jaya and Papua New Guinea, with similar changes reported from Tanzania and Kenya (Epstein *et al.*, 1998) (*see* **Infectious Diseases**, Volume 2; **Viral Diseases and the Influence of Climate Change**, Volume 3).

Shifts in the distribution of disease and parasites are not limited to the tropics. The northern distribution of the European tick *Ixodes ricinus* has apparently shifted northward in Sweden between the early 1980s and 1990s in response to milder winters (Lindgren *et al.*, 2000).

Marine Species

The rapid, and sometimes dramatic, responses of marine species to the short-term sea surface temperature changes accompanying El Niño/Southern Oscillation (ENSO) events indicate that these organisms respond sensitively to ocean warming. During the past 5000 years, El Niño events have typically occurred at a frequency of one to two per decade but, since the mid-1970s, have occurred more often, been more intense, and persisted longer. The impact of these climate events on marine species is clearly evident among corals, which are known to bleach (by expelling symbiotic algae) in response to a range of environmental stresses. The highest sea surface temperatures associated with an ENSO event ever recorded in the tropics occurred in 1998, topping a 50-year trend for some tropical oceans. This event was associated with the most geographically extensive and severe coral bleaching in recorded history, causing significant mortality worldwide. The coral community in Belize in the Caribbean, for example, collapsed completely. Cores extracted from the Belizean reefs showed that these events were unprecedented over at least the past 3000 years (Aronson *et al.*, 2000) (*see* **Coral Reefs: an Ecosystem Subject to Multiple Environmental Threats**, Volume 2).

The almost 25-year trend of warming winter temperatures on the east coast of the US may also have facilitated the spread of two oyster diseases caused by the protozoan parasite *Perkinsus marinus*. Warmer winters appear to have reduced parasite mortality, favoring the northward expansion of *P marinus* into new, susceptible host populations (Harvell *et al.*, 1999).

Changes in the distribution and abundance of several taxa off the coast of California have been particularly well documented over the past few decades. The surface waters of the California current warmed 1.2–1.6 °C between 1951 and 1993. Data from the Californian Cooperative Oceanic Fisheries Investigations indicate that this warming was accompanied by a 70% decline in zooplankton abundance, possibly because increased surface temperatures have reduced the upwelling of cold, nutrient-rich waters to the surface. One of the top predators in this system, the sooty shearwater *Puffinus griseus*, suffered a 90% decline in abundance off western North America between 1987 and 1994 with a nine month lag in response time to changing surface temperatures (Veit *et al.*, 1997). Whilst other influences, such as gill-net mortality and pollution, cannot be discounted, the coincidence in space and time of an oceanic temperature increase and the decrease in zooplankton abundance suggest a direct causal relationship.

Two further studies of marine organisms associated with the Californian coast also provide evidence of recent climate-induced changes in community composition. Two surveys of invertebrates at the same intertidal site conducted 60 years apart documented that of the 45 species first surveyed in 1931–1933, eight out of nine southern species had increased, while five out of eight northern species had declined, with no trend being evident for cosmopolitan species. Over the same period, annual mean shoreline ocean temperatures increased by 0.75 °C and mean summer maximum temperatures by 2.2 °C (Barry *et al.*, 1995). Analogous changes in community composition have occurred in a Californian reef fish assemblage surveyed over a 20 year period (1974–1993) where the proportion of northern, colder affinity species declined from approximately 50% to about 33%, and the proportion of warmer affinity, southern species increased from about 25 to 35%. These changes were accompanied by substantial (up to 92%) declines in the abundance of most species, with the northern species suffering the greatest reductions (Holbrook *et al.*, 1997). Concurrent declines of a similar magnitude were observed for several trophic levels of the benthic ecosystem farther north, with biomass of understory macroalgae decreasing by about 80%.

Warming oceans may also be responsible for observed changes to the species composition of fish and plankton communities at northern high latitudes. Since the beginning of the 20th century, 16 non-indigenous plankton species have become an integral part of the pelagic system of the North Sea. Ten of these species are classified as *thermophilic*, i.e., species usually found in more southerly and warmer regions, and all were recorded for the first time during the last decade (Nehring, 1998). Similar northward range expansions of warm water species of fish and benthic organisms have been observed in the Adriatic, Mediterranean and the English Channel over the last two to three decades.

Terrestrial Vertebrates

Marine organisms may not be the only species affected by ocean warming. Enhanced evaporation from warm surface waters releases large amounts of water vapor. The latent heat released as this moisture condenses accelerates atmospheric warming and does so proportionately more at higher elevations. In the cloudforests of Monteverde, Costa Rica, this process appears to have resulted in a lifting cloud base and a decline in the frequency of mist days. These trends have been strongly associated with synchronous population declines of birds, reptiles and amphibians on plots at 1540 m. Accompanying the declines has been an increase in colonization of cloud–forest intolerant bird species from lower altitudes (Pounds *et al.*, 1999).

Poleward range expansions have been reported for birds in both Europe and the US. The northern margins of 59 bird species with distributions in the south of Britain, for example, have moved farther north by an average of nearly 19 km over a 20-year period (1988–1991 compared to 1968–1972) (Thomas and Lennon, 1999). A survey of small mammals in the southwest US reported similar northward range expansions for 19 species from a variety of habitats (Davis and Callahan, 1992).

EVIDENCE FOR CHANGES IN TIMING OF LIFE CYCLES

The life cycles of many organisms are strongly influenced by temperature and precipitation (*see* **Phenology**, Volume 2). Warmer conditions in the future are generally expected to advance events such as flowering and fruiting in plants, and to hasten development time in some species. Monitoring of seasonal plant and animal activity (phenology) has a long tradition in some countries, particularly in Europe. Many of the long-term datasets compiled by enthusiastic naturalists are now providing some of the most compelling evidence yet for the impact of human-induced climate change.

Analyses of phenological records of events such as flowering, fruiting and leaf unfolding in Northern Hemisphere plants all point to the same conclusions: spring is getting earlier, and the main growing season is lengthening. Data gathered since the late 1950s from the International Phenological Gardens, a Europe-wide network covering a large latitudinal (42–69 °N) and longitudinal (10 °W–27 °E) range that has genetically identical clones of trees and shrubs, show that spring events such as leaf unfolding have advanced by six days, whereas autumn events such as leaf coloring, have been delayed by nearly five days. This means that the average growing season has lengthened by nearly 11 days over the last four decades (Menzel and Fabian, 1999).

The phenological trends in Europe are mirrored by similar ones in the US and Canada. In central and eastern US, for example, first leaf spring index dates became earlier at a rate of about one day per year from 1978–1990. In Edmonton, Canada, there has been a general advance in the timing of spring by eight days over the last six decades. Over a longer period, between 1936 and 1990, the data of first bloom of 10 spring flower species recorded by the Leopold family on a farm in Wisconsin, became earlier by an average of 13 days (Bradley *et al.*, 1999). These reported changes in phenology were accompanied by increases in temperature during March–May of approximately 1–2 °C across central, southern and western Europe and similar or larger rises in temperature during December–February in the central and eastern US. For some individual species, the phenological advances have been dramatic. In Estonia, for example, the pollination of hazel has advanced by 20 days over the last 48 years while in Canada, there has been a 26 day shift to earlier blooming in *Populus tremuloides* over the last century (1900–1997) (Beaubien and Freeland, 2000).

Evidence that recent warming is affecting animal life cycles comes mainly from insects and birds. Butterflies and moths monitored over the last few decades in Britain (since the mid 1970s) and Europe (since the mid 1950s) appear to be emerging as adults earlier. In the Netherlands, for example, the date of flight peak of over 100 species advanced by an average of 12 days between 1975 and 1994 (Ellis *et al.*, 1997). Similar, but less dramatic trends have been found in the dates of aphid flights with five species recorded by a suction trap network across the UK showing an advance of three to six days over the past 25 years (Fleming and Tatchell, 1995).

The enthusiasm of amateur bird watchers for recording their observations has provided a goldmine of data, showing how warming has affected various aspects of bird ecology. Earlier egg-laying, larger clutch sizes and more rapid chick development have been recorded for many British, European and North American birds, and these trends are strongly associated with increases in spring temperatures over the last few decades. For example, an analysis of the laying dates of 65 British bird species surveyed from 1971–1995 showed significant trends toward earlier egg laying in 20 species (31%), with only one species laying significantly later. The shift toward earlier laying averaged 8.8 days over the 24-year period (range 4–17 days). A subsequent, more extensive analysis of the annual median laying date of 36 British bird species over 57 years (1939–1995) found that 19 (53%) species show long-term trends with laying dates becoming later in the 1960s and 1970s, and then earlier in the 1980s and 1990s (Crick and Sparks, 1999). In the US, the date of first nest in the Mexican jay in southeastern Arizona has advanced by nearly 11 days and mean date of first clutch by 10 days

from 1971–1998 (Brown *et al.*, 1999) and the laying date recorded for North American tree swallows advanced by up to nine days from 1959 to 1991 (Dunn and Winkler, 1999). Parallel trends in earlier reproduction have been reported for several British amphibians over the period 1978–1994, with every 1 °C increase in maximum temperatures corresponding to an advance in spawning date of nine to ten days (Beebee, 1995).

For many species, advances in one stage of the life cycle may result in a lack of synchronization with the availability of important resources. The apparent growing disjunction between low- and high-altitude phenology in the Rocky Mountains provides an example. American robins are arriving 14 days earlier at the Rocky Mountains Biological Station than in 1981, apparently in response to warmer air temperatures at lower altitudes where the spring migration is initiated. But at the higher altitudes where the robins arrive to feed, the timing of snowmelt has not changed and as a consequence, the interval between the birds' arrival and the first date of bare ground has lengthened by 18 days. The birds have therefore faced increasingly longer periods of snow-covered ground before the summer growing season begins. Yellow-bellied marmots in the same area may also have experienced similar problems. These mammals are emerging from hibernation 38 days earlier than 23 years ago (Inouye *et al.*, 2000).

WHAT NEXT?

Many of the trends reviewed here show accelerated rates of change over the last two to three decades and if these studies are indeed early indicators of future change, we will expect to see the following situations becoming increasingly apparent:

- Extensions of species' geographic range boundaries toward the poles or to higher elevations by progressive establishment of new local populations with concomitant extinctions of populations along boundaries at lower latitudes or elevations.
- Increasing invasion by opportunistic, weedy, mobile and/or thermophilic species, especially into sites where local populations of existing species are declining.
- Progressive mismatching of established species interactions (e.g., plants and pollinators, parasites and hosts) due to lack of synchronization between life cycles.

Altogether, such changes at the level of individual species will inevitably, and irrevocably, alter the structure and composition of whole communities. In several decades, some of the trends reviewed here may be reinterpreted as simply due to natural variation. But even if only a fraction are indeed attributable to the influence of humans on the atmosphere, we are left with the sobering thought that they have occurred in response to a warming that is probably only a fifth of that expected over the century ahead.

REFERENCES

Aronson, R B, Precht, W F, Macintyre, I G, and Murdoch, T J T (2000) Coral Bleach-out in Belize, *Nature*, **405**, 36.

Barry, J P, Baxter, C H, Sagarin, R D, and Gilman, S E (1995) Climate-related Long-term Faunal Changes in a California Rocky Intertidal Community, *Science*, **267**, 672–675.

Beaubien, E G and Freeland, H J (2000) Spring Phenology Trends in Alberta, Canada: Links to Ocean Temperature, *Int. J. Biometeorol.*, **33**, 53–59.

Beebee, T J C (1995) Amphibian Breeding and Climate, *Nature*, **374**, 219–220.

Bradley, N L, Leopold, A C, Ross, J, and Huffaker, W (1999) Phenological Changes Reflect Climate Change in Wisconsin, *Proc. Nat. Acad. Sci. USA*, **96**, 9701–9704.

Brown, J L, Li, S H, and Bhagabati, N (1999) Long-term Trend Toward Earlier Breeding in an American Bird: A Response to Global Warming? *Proc. Nat. Acad. Sci. USA*, **96**, 5565–5569.

Crick, H Q P and Sparks, T H (1999) Climate Change Related to Egg-laying Trends, *Nature*, **399**, 423–424.

Davis, R and Callahan, J R (1992) Post-pleistocene Dispersal in the Mexican Vole (*Microtus mexicanus*), an Example of an Apparent Trend in the Distribution of Southwestern Mammals, *Great Basin Nat.*, **52**, 262–268.

de Jong, P W and Brakefield, P M (1998) Climate and Change in Clines for Melanism in the Two-spot Ladybird, *Adalia bipunctata* (Coleoptera: Coccinellidae), *Proc. R. Soc. Lond. Series B*, **265**, 39–43.

Dunn, P O and Winkler, D W (1999) Climate Change has Affected the Breeding Date of Tree Swallows throughout North America, *Proc. R. Soc. Lond. Series B*, **266**, 2487–2490.

Ellis, W N, Donner, J H, and Kuchlein, J H (1997) Recent Shifts in Phenology of Microlepidoptera, Related to Climatic Change (Lepidoptera), *Entomol. Ber.*, (Amsterdam), **57**, 66–72.

Epstein, P R, Diaz, H F, Elias, S, Gradherr, G, Graham, N E, Martens, W J M, Mosley-Thompson, E, and Susskind, J (1998) Biological and Physical Signs of Climate Change: Focus on Mosquito-borne Diseases, *Bull. Am. Meteorol. Soc.*, **79**, 409–417.

Fleming, R A and Tatchell, G M (1995) Shifts in the Flight Periods of British Aphids: a Response to Climate Warming? in *Insects in a Changing Environment*, eds R Harrington and N Stork, Academic Press, London, 505–508.

Harvell, C D, Kim, K, Burkholder, J M, Colwell, R R, Epstein, P R, Grimes, D J, Hofmann, E E, Lipp, E K, Osterhaus, A D, Overstreet, R M, Porter, J W, Smith, G W, and Vasta, G R (1999) Emerging Marine Diseases – Climate Links and Anthropogenic Factors, *Science*, **285**, 1505–1510.

Holbrook, S J, Schmitt, R J, and Stephens, Jr, J S (1997) Changes in an Assemblage of Temperate Reef Fishes Associated with a Climate Shift, *Ecol. Appl.*, **7**, 1299–1310.

Hughes, L (2000) Biological Consequences of Global Warming: is the Signal Already Apparent? *Trends Ecol. Evol.*, **15**, 56–61.

Inouye, D W, Barr, B, Armitage, K B, and Inouye, B D (2000) Climate Change is Affecting Altitudinal Migrants and Hibernating Species, *Proc. Nat. Acad. Sci. USA*, **97**, 1630–1633.

Keeling, C D, Chin, J F S, and Whorf, T P (1996) Increased Activity of Northern Vegetation Inferred from Atmospheric CO_2 Measurements, *Nature*, **382**, 146–149.

Lindgren, E, Tälleklint, L, and Polfeldt, T (2000) Impact of Climatic Change on the Northern Latitude Limit and Population Density of the Disease-transmitting European Tick *Ixodes ricinus*, *Environ. Health Perspec.*, **108**, 119–123.

Menzel, A and Fabian, P (1999) Growing Season Extended in Europe, *Nature*, **397**, 659.

Myneni, R B, Keeling, C D, Tucker, C J, Asrar, G, and Nemani, R R (1997) Increased Plant Growth in the Northern High Latitudes from 1981–1991, *Nature*, **386**, 698–701.

Nehring, S (1998) Establishment of Thermophilic Phytoplankton Species in the North Sea: Biological Indicators of Climatic Changes? *ICES J. Mar. Sci.*, **55**, 818–823.

Parmesan, C (1996) Climate and Species' Range, *Nature*, **382**, 765–766.

Pounds, J A, Fogden, P L, and Campbell, J H (1999) Biological Responses to Climate Change on a Tropical Mountain, *Nature*, **398**, 611–615.

Rodríguez-Trelles, F and Rodríguez, M A (1998) Rapid Micro-evolution and Loss of Chromosomal Diversity in Drosophila in Response to Climate Warming, *Evol. Ecol.*, **12**, 829–838.

Schell, D M (2000) Declining Carrying Capacity in the Bering Sea: Isotopic Evidence from Whale Baleen, *Limnol. Oceanogr.*, **45**, 459–462.

Serreze, M C, Walsh, J E, Chapin, III, F S, Osterkamp, T, Dyurgerov, M, Romanovsky, V, Oechel, W C, Morison, J, Zhang, T, and Barry, R G (2000) Observational Evidence of Recent Change in the Northern High-latitude Environment, *Clim. Change*, **46**, 159–207.

Smith, R I L (1994) Vascular Plants as Bioindicators of Regional Warming in Antarctica, *Oecologia*, **99**, 322–328.

Thomas, C D and Lennon, J J (1999) Birds Extend their Ranges Northwards, *Nature*, **399**, 213.

Veit, R R, McGowan, J A, Ainley, D G, Wahls, T R, and Pyle, P (1997) Apex Marine Predator Declines Ninety Percent in Association with Changing Oceanic Climate, *Global Change Biol.*, **3**, 23–28.

Biological Invasions

see Biological Invasions (*Opening essay*, Volume 2)

Biological Pump in World Oceans

see Iron Cycle (Volume 2)

Biome

Widely separated geographic regions with similar environmental conditions typically contain similar types of biota. This feature allows the partitioning of the biosphere into very large regions, called *biomes*, in which similar geology and climate are home to species with similar adaptations and life history strategies. Biomes are often named after their dominant kind of vegetation (e.g., grassland or shrubland) and, as the geographically largest biotic units, they include all ecosystems and successional stages of ecosystems located within their borders.

The boreal forest is an example of a terrestrial biome. It occurs in Asia, Europe, and North America where the climate and geology combine to produce suitable environmental conditions. The boreal forest biome has the same general kinds of ecosystems everywhere, but the individual constituent species differ from one boreal forest region to another (partly because these regions differ in their ecological, geological, and climatic histories). Nonetheless, the different boreal forests contain organisms that are ecologically similar, i.e., the species differ but they correspond in the sense that they have adapted to similar environments and have developed similar life cycles.

FURTHER READING

Odum, E P (1971) *Fundamentals of Ecology*, W B Saunders, Philadelphia, PA.

RICHARD A FLEMING *Canada*

Biome Models

Wolfgang Cramer

Potsdam Institute for Climate Impact Research, Potsdam, Germany

Biomes may be defined as "... the world's major (plant and animal) communities, classified according to the predominant vegetation and characterized by adaptations of organisms to that particular environment" (Campbell, 1996). A synonym for biome is major life zone. Typical biomes are, e.g., coniferous forest (taiga), desert, grassland, Mediterranean shrubland, rain forest, savanna, temperate forest, tundra, coastal waters, coral reef, freshwater, and open

ocean; but the number of subdivisions in any biome list is arbitrary and depends on the list's specific purpose. Due to their broad geographic extent, often covering broad continent spanning zones, biomes normally consist of multiple types of ecosystems, which reflect local environmental gradients and interactions. For example, the same taiga biome contains wet oceanic cold rain forests and cold adapted continental deciduous conifer forests.

On land, biomes are usually characterized by their dominant vegetation expressed in broad categories. Just like in the more narrowly defined ecosystems, animals and soil biota are equally important elements in their function and therefore also for their structure. Biome definitions usually do not relate to the kind of human land use that takes place within their boundaries. Biomes may, however, be considered as indicators for the agricultural or forest practice that can be carried out potentially in a given region.

CLIMATIC FACTORS THAT DETERMINE BIOME DISTRIBUTIONS

Biomes are well determined by macroclimatic and soil characteristics. Indeed, before the advent of large-scale climate maps based on weather stations, both geographers and climatologists have often considered biomes as indicators for climate in areas with insufficient weather stations (Köppen, 1884; Cramer and Leemans, 1993). Climatic factors may be used to identify environmental correlates or even causes for the distributional limits of either biomes or species. Classically, the factors used were easily accessible longer-term averages of standard variables, such as annual or monthly averages of temperature and precipitation. By visual map comparison, coinciding climatic and biome limits could be used to further characterize the biome and its typical environment.

Increasingly, so-called *bioclimatic* indicators are being developed from climatic boundaries that are based on specific processes of plant growth or survival. Ultimately, the goal of such indicators is to be able to predict the location, and the possible change, of biome distribution limits in response to past or future climate change. Examples are growing degree days or heat sums (i.e., the sum of all temperatures above a certain threshold), absolute cold tolerance limits or estimates of typical maximum drought stress based on a simple water balance calculation. The life zone classification by Holdridge (1947, 1967) was the first attempt to provide a complete biome (life zone) system defined by only two such bioclimatic variables (*see* **Holdridge Life Zone Classification**, Volume 2). The Holdridge system could be applied globally and thereby provided a first simple global biome model that could be used to assess the sensitivity of global vegetation distribution to climate change (Emanuel *et al.*, 1985).

BIOME MODELING APPROACHES

The arrival of digital computers with sufficient capacity to store and manipulate large global climatic databases, as well as maps of other environmental variables, triggered the development of more elaborate mathematical models of global biome distribution. Such models are global in the sense of being applicable at any (ice-free) land surface, if given a description of environmental conditions at this location. The models have no implicit spatial resolution, and there is no lateral interaction between neighboring locations.

Depending on their purpose, biome models have been developed for the characterization of biome distributions as a function of climate and soils. Some related models have also been used to assess the dynamics of carbon and water fluxes, using a map of biome types. A third, more recent group of models combines these two aspects. The most recently developed class of models adds transient dynamics of vegetation structure and carbon pools to the picture. All four classes of models are briefly discussed in the following subsections.

Equilibrium Models of Biome Distribution

For the first generation of the biome models, biome distribution is assumed to be in equilibrium with climate. Box (1981) developed such a model, combining the concepts of life form (Raunkiær, 1907) and life zone (Holdridge, 1967). To capture the essential life forms of all higher plants, Box defined 90 different plant types, as well as bioclimatic envelopes (lower and upper boundaries) for eight variables (listed in Table 1), which all could be mapped on the basis of available climatic normals. Woodward (1987) was the first to show that a more appropriate analysis of biome distribution required the representation of the physiological processes that control survival and performance of plants in a given ecosystem. His analysis and the resulting model focussed primarily upon cold tolerance and the water

Table 1 Bioclimatic indices used by the Box model

T_{max}	Mean temperature of the warmest month (°C)
T_{min}	Mean temperature of the coldest month (°C)
D_T	Range between T_{min} and T_{max} (°C)
P	Mean total annual precipitation (mm)
MI	Moisture index, de ned as the ratio between P and annual potential evapotranspiration (Thornthwaite and Mather, 1957)
P_{max}	Mean total precipitation of the wettest month (mm)
P_{min}	Mean total precipitation of the driest month (mm)
$P_{T_{max}}$	Mean total precipitation of the warmest month (mm)

Source: Box, 1981.

needs of plants, both of which could not easily be derived from existing climate data sets.

The development of a biome distribution model to be applied under different climate regimes (such as those of earlier periods in geological history, or for scenarios of future developments) required the incorporation of mechanistic formulations for all critical processes that could limit plant distributions, as well as the development of suitable data sets for model application worldwide. For the parameterization of process formulations, it was found necessary to group plant species into so called functional groups. Each functional group describes plant types with similar behavior for a given process. A widely used model of this type is the BIOME model (Prentice *et al.*, 1992). It involves four major bioclimatic tolerance/requirement factors (Table 2) which are characterized independently for 13 plant functional types (PFTs). Several PFTs may co-occur at any location (Table 3). Since precipitation provides no direct indication of moisture stress, a process-based water balance model (Cramer and Prentice, 1988; Prentice *et al.*, 1993) is used to assess regions with significant differences between actual and potential evapotranspiration. The development of BIOME was possible since the climatic limitations could be mapped more precisely than earlier using a topography-sensitive global climate data base developed for this purpose (Leemans and Cramer, 1991). A significantly updated version of this data base is available free of charge from http://www.pik-potsdam.de/~cramer/climate.htm.

Applications of BIOME cover a broad range of questions. Since the model gives an appropriate characterization of broad land surface types (in the absence of direct human impact), it has been found to be a suitable replacement for land cover maps that are used in atmospheric general circulation models. Asynchronously coupled to the atmosphere (i.e., with permitted change in biome distribution only every five years or so), the model has allowed one to address critical issues in atmosphere/biosphere stability on the paleoecological time scale (Claussen, 1994, 1996).

Equilibrium Models of Biogeochemical Fluxes

The necessity to better quantify the role of the land biosphere for the overall balance of carbon fluxes was recognized many years ago (Bolin, 1977). As a result of the International Biological Programme (IBP), Lieth (1975) quantified the relationship between net primary productivity (NPP) and climate (temperature, precipitation) in a regression model consisting of one globally applicable equation. The model was revolutionary for its time, but the limitations of this regression for the assessment of future fluxes are nevertheless significant. Most fundamentally, inspection of the Earth's vegetation shows that a single monotonous equation is likely to fail near the boundaries of major biomes, such as forest–tundra boundaries, where the presence of a major life form affects both the local environment and thereby also productivity.

Moreover, the observations naturally cover only the recent range of atmospheric carbon dioxide concentrations, while it is likely that higher carbon dioxide concentrations will cause changes in plant productivity. Finally, NPP is not the variable of direct interest for assessments of global change impacts; net biome production (NBP), i.e., the remaining net flux after deducting all losses due to respiration and disturbance, is more important and only weakly related to NPP (IGBP Terrestrial Carbon Working Group, 1998).

To overcome these limitations, several groups have developed process-oriented biogeochemical models, e.g., TEM (Raich *et al.*, 1991), CARAIB (Nemry *et al.*, 1996) and FBM (Lüdeke *et al.*, 1994). A detailed review of the state of the art for these models may be found in Cramer *et al.* (1999). Most of these models share the biome concept

Table 2 Derivation of bioclimatic indices from interpolated climatic data in the BIOME model

Tolerance/requirement	Ecophysiological mechanism	Bioclimatic index	Climatic variable (monthly means)
Cold tolerance	Killing temperature during coldest period of the year	T_{min} (temperature of the coldest month, lower limit)	Temperature
Chilling requirement	Winter chilling period required for budburst of woody plants	T_{min} (temperature of the coldest month, upper limit)	Temperature
Heat requirement	Annual growth respiration requirement	GDD (growing degree days above 0 °C and 5 °C)	Temperature
Moisture requirement/drought tolerance	Soil moisture availability	AET/PET (annual actual evapotranspiration/ annual potential evapotranspiration)	Temperature, precipitation, cloudiness

Source: Prentice *et al.*, 1992.

Table 3 Environmental constraints for each PFT in the BIOME model

PFT		T_{min} min	T_{min} max	GDD_0 min	GDD_5 min	T_{max} min	AET/PET min	AET/PET max
Trees	Tropical evergreen	15.5					0.80	
	Tropical raingreen	15.5					0.45	0.95
	Warm temperate evergreen	5.0					0.65	
	Temperate summergreen	−15.0	15.5		1200		0.65	
	Cool temperate conifer	−19.0	5.0		900		0.65	
	Boreal evergreen conifer	−35.0	−2.0		350		0.75	
	Boreal summergreen		5.0		350		0.65	
Non-trees	Sclerophyll/succulent	5.0					0.28	
	Warm grass/shrub					22.0	0.18	
	Cool grass/shrub				500		0.33	
	Cold grass/shrub			100			0.33	
	Hot desert shrub					22.0		
	Cold desert shrub			100				
No plants	(Dummy type)							

Source: Prentice *et al.*, 1992.

as a central feature, which is implemented in a static way, i.e., either by using a global vegetation map or by using BIOME as input. Mechanistic formulations of photosynthesis, plant and soil respiration and other processes are parameterized for each biome, and the resulting pools and fluxes of carbon are a function of both biome distribution and climate. Biogeochemical models that make use of remote sensing data are not included here, since these models are not capable of forward-looking simulations and therefore are outside the scope of biome models *per se*.

Coupled Biome Distribution/Biogeochemistry Models

Some equilibrium biome models simulate biogeochemical processes (fluxes) and pattern (vegetation type and structure) simultaneously. Usually, the determination of the vegetation types follows process-optimization rules, e.g., by maximization of NPP according to soils and climate, or by maximization of the leaf area index (LAI) to satisfy the annual moisture and carbon balances. Examples are DOLY (Woodward *et al.*, 1995) and BIOME3 (Haxeltine and Prentice, 1996). In both models, the distribution of PFT (Plant Functional Type), and hence the resulting biomes, is a function of climate and soils, but (different from BIOME) this function is contained in modules that describe both biogeochemical fluxes and resulting pool sizes quantitatively within the basic bioclimatic limits for PFT survival. Photosynthesis, respiration and water use are simulated depending on climate, as well as on atmospheric carbon dioxide concentration, using process descriptions from field and laboratory studies. This allows one to expect a greater degree of realism even for extrapolations into future conditions, as opposed to regression type approaches.

The most recent model of this class is BIOME4 (Kaplan, personal communication). BIOME4 was developed from BIOME3 with the aim of better covering the diversity of biome types by adding additional PFT and updating some process formulations and parameters. It is based on 13 PFTs, each of which has assigned bioclimatic limits that determine whether or not NPP is calculated for a given location. A coupled carbon and water flux scheme is used to determine the LAI that maximizes NPP for each PFT. The PFT with the highest NPP is chosen to represent vegetation for the biogeochemical fluxes and pools. For savanna vegetation, where the interaction between trees and grasses is particularly important, separate rules are defined to estimate the combined effect of co-existing PFTs.

Dynamic Global Vegetation Models

Interest in biome models comes to a large extent from the need to estimate likely changes in carbon stores in the terrestrial biosphere, as a consequence of atmospheric carbon dioxide increase and the associated changing climate. In periods with rapidly changing climate, the equilibrium assumption in classical biome models is invalid, since vegetation structure and carbon stores in above- and below-ground biomass often change only over decades to centuries, except for disturbance type events such as fires. To realistically simulate NBP, growth of PFTs needs to be considered as well as the major processes affecting soil carbon pools.

Dynamic global vegetation models (DGVMs) are specifically developed for the assessment of transient changes in vegetation structure as well as NBP, driven by climate change. Their development is based on the advances made in equilibrium biome models, as well as in forest succession models. DGVMs, such as HYBRID (Friend *et al.*, 1997) or

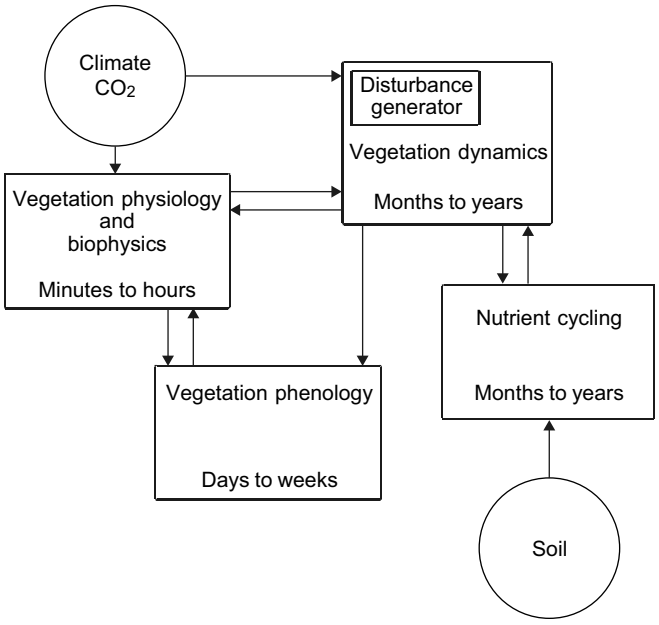


Figure 1 Basic structural elements of a DGVM and their associated time scales (Cramer *et al.*, 2001)

LPJ (Sitch, 2000) contain dynamic formulations for slow and fast processes of carbon and water fluxes (Figure 1), and they maintain the principal pools of carbon over long time scales. A more detailed description of current DGVMs, and an application are to be found in Cramer *et al.* (2001).

Ultimately, the quantitative and transient analysis of biospheric structure, as made in DGVMs, abandons the biome concept for all but the diagnostic description. All PFTs have more or less smooth responses to environmental gradients and they do not usually react abruptly, except for the case of rapidly deteriorating conditions. Most biomes have broad transition zones between each other. Environmental changes, such as those that occurred after the deglaciation of Northern Europe, usually led to smooth transitions from a biome dominated by some species into one dominated by another. It is nevertheless useful for analysis and communication, if the resulting distribution of PFTs is shown as a distinct biome map. Otherwise, the reader would have difficulty grasping the distribution of multiple plant types across the globe, using a single map.

BIOMES AS PART OF THE FULL EARTH SYSTEM

Plant types, ecosystems and biomes are not only distributed as a function of climate, they also influence atmospheric processes in a way that differs between biome types. Important characteristics for this influence are the roughness of the canopy (which influences the characteristics of the surface boundary layer), its albedo (which affects the energy balance) and its LAI (which affects carbon and water fluxes). These effects have long been known at the local scale (Köppen and Geiger, 1936), but were studied at the global scale only much later (Dickinson, 1983). Now, many model experiments have shown that large-scale changes of biome distribution, e.g., due to human land use, affect atmospheric circulation (Salati, 1986). Inevitably, these effects produce feedbacks between atmosphere and biosphere which may even lead to multiple stable states of the coupled system (Claussen, 1998).

The consequence of the notion of bi-directional feedbacks between atmosphere and biosphere (expressed as biomes) is that all assessments of future conditions of the system depend on understanding all of the major elements of the Earth System, which is currently destabilized primarily by human disruption of the carbon balance. What we need to know is whether the Earth's vegetation might amplify or dampen the impacts of this disruption. Present estimates seem to indicate that physiological processes in some major biomes (e.g., the boreal forest) lead to increased uptake and storage of carbon by ecosystems, due to the increased carbon dioxide concentration in the atmosphere (Cramer *et al.*, 2001). However, the resulting changes in temperature may well lead to increased drought and fire risk in some areas, notably the drier variants of forested biomes (White *et al.*, 2000). This would in turn influence the atmosphere due to the resulting release of carbon from dying forests and mineralized soil carbon.

We cannot presently determine the exact trajectory of these changes. But model results are sufficiently plausible to raise concern over the stability of the overall system. Some biomes, such as tropical forests, have also been identified as being crucial for a sustainable Earth from other perspectives, such as the maintenance of biodiversity.

REFERENCES

Bolin, B B (1977) Changes of Land Biota and their Importance to the Carbon Cycle, *Science*, **196**, 613–615.

Box, E O (1981) *Macroclimate and Plant Forms: an Introduction to Predictive Modeling in Phytogeography*, Dr W Junk Publishers, The Hague.

Campbell, N A (1996) *Biology*, 4th edition, The Benjamin/Cummings Publishing Company, Menlo Park, CA.

Claussen, M (1994) On Coupling Global Biome Models with Climate Models, *Clim. Res.*, **4**, 203–221.

Claussen, M (1996) Variability of Global Biome Patterns as a Function of Initial and Boundary Conditions in a Climate Model, *Clim. Dyn.*, **12**, 371–379.

Claussen, M (1998) On Multiple Solutions of the Atmosphere–Vegetation System in Present-day Climate, *Global Change Biol.*, **4**, 549–560.

Cramer, W, Bondeau, A, Woodward, F I, Prentice, I C, Betts, R A, Brovkin, V, Cox, P M, Fisher, V, Foley, J, Friend, A D, Kucharik, C, Lomas, M R, Ramankutty, N, Bitch, S, Smith, B, White, A, and Young-Molling, C (2001) Global Response of

Terrestrial Ecosystem Structure and Function to CO_2 and Climate Change: Results from Six DGVMs, *Global Change Biol.*, in press.

Cramer, W, Kicklighter, D W, Bondeau, A, Moore, III, B, Churkina, G, Nemry, B, Ruimy, A, Schloss, A L, and participants of the Potsdam NPP Model Intercomparison (1999) Comparing Global Models of Terrestrial Net Primary Productivity (NPP): Overview and Key Results, *Global Change Biol.*, **5**(1), 1–15.

Cramer, W and Leemans, R (1993) Assessing Impacts of Climate Change on Vegetation using Climate Classification Systems, in *Vegetation Dynamics and Global Change*, eds A M Solomon and H H Shugart, Chapman and Hall, New York, 190–217.

Cramer, W and Prentice, I C (1988) Simulation of Soil Moisture Deficits on a European Scale, *Norsk Geografisk Tidskrift*, **42**, 149–151.

Dickinson, R E (1983) Land Surface Processes and Climate, Surface Albedos and Energy Balance, *Adv. Geophys.*, **25**, 305–353.

Emanuel, W R, Shugart, H H, and Stevenson, M P (1985) Climatic Change and the Broad-scale Distribution of Terrestrial Ecosystems Complexes, *Climatic Change*, **7**, 29–43.

Friend, A D, Stevens, A K, Knox, R G, and Cannell, M G R (1997) A Process-based, Terrestrial Biosphere Model of Ecosystem Dynamics (Hybrid v3.0), *Ecol. Model*, **95**, 249–287.

Haxeltine, A and Prentice, I C (1996) BIOME3: An Equilibrium Biosphere Model Based on Ecophysiological Constraints, Resource Availability and Competition Among Plant Functional Types, *Global Biogeochem. Cycles*, **10**, 693–709.

Holdridge, L R (1947) Determination of World Plant Formations From Simple Climatic Data, *Science*, **105**, 367–368.

Holdridge, L R (1967) *Life Zone Ecology*, revised edition, Tropical Science Center, San José, Costa Rica.

IGBP Terrestrial Carbon Working Group (1998) The Terrestrial Carbon Cycle: Implications for the Kyoto Protocol, *Science*, **280**, 1393–1394.

Köppen, W, (1884) Die Wärmezonen der Erde, nach der Dauer der Heissen, Gemässigten und Kalten Zeit und nach der Wirkung der Wärme auf die Organische Welt Betrachtet, *Meteorologische Zeitschrift*, **1**, 215–226 (+map).

Köppen, W and Geiger, R (1936) Handbuch der Klimatologie, Gebrüder Bornträger, Berlin.

Leemans, R and Cramer, W (1991) *The IIASA Database for Mean Monthly Values of Temperature, Precipitation and Cloudiness of a Global Terrestrial Grid*, Research Report RR–91–18, International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria.

Lieth, H (1975) Primary Production of the Major Vegetation Units of the World, in *Primary Productivity of the Biosphere*, eds H Lieth and R H Whittaker, Springer-Verlag, Berlin, 203–215.

Lüdeke, M K B, Badeck, F-W, Otto, R D, Häger, C, Dünges, S, Kindermann, J, Würth, G, Lang, T, Jäkel, U, Klaudius, A, Ramge, P, Habermehl, S, and Kohlmaier, G H (1994) The Frankfurt Biosphere Model. A Global Process Oriented Model for the Seasonal and Long-term CO_2 Exchange Between Terrestrial Ecosystems and the Atmosphere, I: Model Description and Illustrative Results for Cold Deciduous and Boreal Forests, *Clim. Res.*, **4**, 143–166.

Nemry, B, François, L, Warnant, P, Robinet, F, and Gérard, J-C (1996) The Seasonality of the CO_2 Exchange Between the Atmosphere and the Land Biosphere: A Study with a Global Mechanistic Vegetation Model, *J. Geophys. Res.*, **101**, 7111–7125.

Prentice, I C, Cramer, W, Harrison, S P, Leemans, R, Monserud, R A, and Solomon, A M (1992) A Global Biome Model Based on Plant Physiology and Dominance, Soil Properties and Climate, *J. Biogeogr.*, **19**, 117–134.

Prentice, I C, Sykes, M T, and Cramer, W (1993) A Simulation Model for the Transient Effects of Climate Change on Forest Landscapes, *Ecol. Model.*, **65**, 51–70.

Raich, J W, Rastetter, E B, Melillo, J M, Kicklighter, D W, Steudler, P A, Peterson, B J, Grace, A L, Moore, III, B, and Vörüsmarty, C J (1991) Potential Net Primary Productivity in South America: Application of a Global Model, *Ecol. Appl.*, **1**, 399–429.

Raunkiær, C (1907) *Planterigets Livsformer*. Gyldendalske Boghandel and Nordisk Forlag, Copenhagen/Kristiania.

Salati, E (1986) Amazon: Forest and Hydrological Cycle, in *Climate–Vegetation Interactions*, eds C Rosenzweig and R E Dickinson, UCAR-OIES, NASA/Goddard Space Flight Center, Greenbelt, MD, January 27–29, 110–112.

Sitch, S (2000) *The Role of Vegetation Dynamics in the Control of Atmospheric CO_2 Content*, PhD Lund University, Lund, Sweden.

Thornthwaite, C W and Mather, J R (1957) *Instructions and Tables for Computing Potential Evapotranspiration and the Water Balance*, Drexel Institute of Technology, Laboratory of Climatology.

White, A, Cannell, M G R, and Friend, A D (2000) CO_2 Stabilization, Climate Change and the Terrestrial Carbon Sink, *Global Change Biol.*, **6**, 817–833.

Woodward, F I (1987) *Climate and Plant Distribution*, Cambridge University Press, Cambridge.

Woodward, F I, Smith, T M, and Emanuel, W R (1995) A Global Land Primary Productivity and Phytogeography Model, *Global Biogeochem. Cycles*, **9**, 471–490.

Biome- BGC Ecosystem Model

Lars Pierce

California State University, Monterey Bay, CA, USA

As human population increases, so do the influences that we as humans have on the biosphere. Human-induced disturbances such as land use change and environmental pollution all impact the Earth system yet our understanding of how ecosystems will respond to these disturbances is limited. We rely upon both field measurements and computer

modeling to describe the response of ecosystems to global environmental changes. Computer models that describe the flow of energy and matter through ecosystems are referred to as ecosystem models. Ecosystem models allow us to explore the potential interactions between environmental disturbance and ecosystem structure and function.

Biome-BGC is an ecosystem model that was first developed in the coniferous forests of the western US, and has since been adapted to simulate the biogeochemical cycling (BGC) of most terrestrial ecosystems. Biome-BGC is designed to simulate the fluxes of carbon (C), nitrogen (N), and water through plants and soils over space and time (Running and Hunt, 1993). The model has been utilized to simulate ecosystem processes from the plot scale up to the global scale (Hunt et al.*, 1996). Biome-BGC is similar in concept and design to several other ecosystem models (VEMAP Members, 1995). In terms of complexity, these models occupy a middle ground by choosing to represent only those ecosystem processes critical for making regional-scale predictions across a range of ecosystems and climates. These models generally do not simulate processes at sub-daily resolutions, and instead operate from daily to monthly timescales. These models differ, however, in their conceptual representation of ecosystem water, carbon, and nitrogen cycling which somewhat reflects the type of ecosystem for which they were initially developed.*

MODEL STRUCTURE AND COMPUTATIONS

Biome-BGC has retained much of the same structure of its predecessor Forest-BGC (Running and Coughlan, 1988), and includes one canopy layer, one soil layer, five vegetation carbon and nitrogen pools, and two soil carbon and nitrogen pools. The model is driven by routinely available daily climate data (air temperature, solar radiation, humidity and precipitation) and the definition of several key state variables; vegetation, soil, and site conditions (Figure 1). Biome-BGC is a big-leaf model; the plant canopy is represented as a single layer whose thickness is defined by the leaf area index (LAI) of the ecosystem. Leaf-level water and carbon fluxes are scaled to the canopy according to LAI, a key variable describing ecosystem structure, which can be obtained at regional scales from satellite imagery (Pierce *et al.*, 1993).

The Biome-BGC model simulates ecosystem processes in both plants and soils across a range of timescales (Figure 1). Processes in the plant compartment occur at shorter timescales (i.e., daily), whereas processes in the soil compartment occur at longer timescales (i.e., monthly to yearly). Within the plant compartment, daily precipitation (PPT) is intercepted by the canopy. Throughfall (T_f) is routed to soil water (H_2O) storage, the capacity of which is a function of plant rooting depth and soil texture. Any water in excess of the soil water-holding

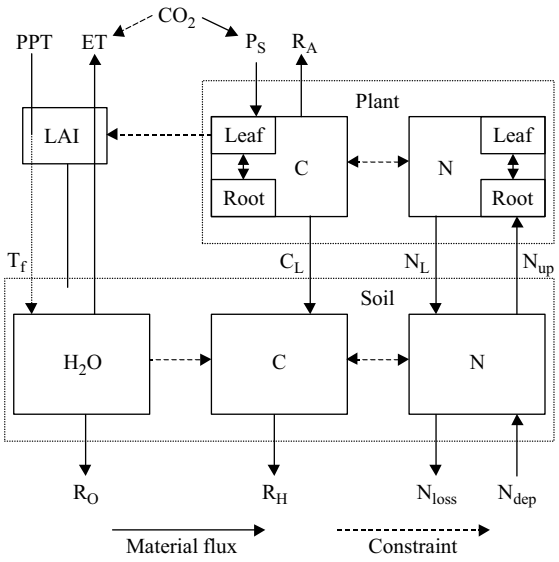


Figure 1 Schematic owchart of the BIOME-BGC Ecosystem Model. Abbreviations and coef cients are de ned in the text

capacity is assumed to leave the ecosystem as runoff (R_O). Soil moisture controls soil evaporation (E) and leaf-level transpiration (T), which are calculated using the Penman-Montieth equation. Photosynthesis (P_S) is calculated using a formulation of the Farquhar model (Hunt *et al.*, 1996). Plant maintenance respiration (R_A) is calculated as a function of biomass, tissue nitrogen concentration, and temperature. Nitrogen (N) is taken up by the plant, (N_{up}), from the soil mineral nitrogen pool at a rate proportional to root biomass, soil temperature, and soil moisture. Carbon and nitrogen are then allocated to leaf and root based upon the relative balance between photosynthesis and nitrogen uptake. Excess carbon uptake via leaves relative to nitrogen uptake via roots favors allocation of resources to acquiring more nitrogen through additional root growth. Excess nitrogen uptake relative to carbon favors allocation of resources to acquiring more carbon through additional leaf growth. Allocation to woody biomass (stem and coarse root) is a function of leaf allocation. Growth respiration is calculated as a constant proportion of allocated carbon.

Within the soil compartment, turnover of live plant carbon and nitrogen is controlled by a balance between maintenance respiration and photosynthetic carbon uptake (Mooney and Gulmon, 1982), such that allocation cannot occur until maintenance respiration requirements have first been met. If maintenance respiration requirements are not satisfied, then live tissue carbon and nitrogen must turn over from plants until plant maintenance respiration requirements are met by carbon uptake. Litterfall (C_L, N_L) is routed to the litter and soil organic pools, where carbon is released through heterotrophic

respiration (R_H) and nitrogen is mineralized into the soil mineral nitrogen pool at a rate controlled by the litter carbon–nitrogen ratio, and soil temperature and water constraints. Nitrogen inputs through atmospheric nitrogen deposition (N_{dep}) and nitrogen losses (N_{loss}) through leaching are also calculated.

Within Biome-BGC, soil moisture and photosynthesis calculations in the plant compartment control processes in the soil compartment. Temperature and water controls, as well as plant production, are calculated in the plant compartment and passed to the soil compartment where this information is used to control litter and soil turnover. Changes in decomposition and nitrogen availability calculated in the soil compartment feed back to control carbon and nitrogen uptake and water use in the plant compartment. In this way, the plant and soil compartments influence one another, and overall model behavior. One of the key feedbacks in Biome-BGC is the influence of soil decomposition processes and nitrogen availability on leaf nitrogen and photosynthesis.

ECOSYSTEM RESPONSES TO ELEVATED ATMOSPHERIC CARBON DIOXIDE

As an example of how an ecosystem model can be used to examine the response of ecosystems to disturbance, such as increases in atmospheric carbon dioxide (CO_2) concentration, we provide an example from the Jasper Ridge CO_2 Experiment (Field *et al.*, 1996). The Biome-BGC ecosystem model is used to simulate the response of a Mediterranean annual grassland community to elevated atmospheric CO_2 under three different rainfall regimes (Field *et al.*, 1997). The results of the simulations can then be compared to measured data to determine the utility of the ecosystem model in estimating the response of an ecosystem to changes in atmospheric CO_2. The grassland community was grown during the winter and spring of 1994–1995 in open-top chambers at the Jasper Ridge Biological Reserve on the Stanford University campus in central coastal California. To closely represent natural environmental conditions, each open-top chamber is exposed to ambient temperature and sunlight. Precipitation inputs to the chambers were controlled to represent dry, average, and wet years at Jasper Ridge. Natural sandstone soil profiles were reconstructed within the chambers, and the chambers were seeded to mimic the species composition and structure of the local annual grassland community. One half of the open-top chambers received ambient CO_2 (350 ppmv), while the other half of the open-top chambers received double-ambient CO_2 (700 ppmv). Total above-ground biomass was then harvested and recorded for each of the six treatments (two CO_2 treatments × three rainfall treatments) at the end of the growing season.

The Biome-BGC ecosystem model was then run to simulate total above-ground biomass for each of the six water and CO_2 treatments. Plant physiological parameters within the model were defined for a Californian annual grassland, according to Jackson *et al.* (1994), who measured a 45% reduction in leaf conductance under doubled atmospheric CO_2. Soil texture and nitrogen content were set according to Field *et al.* (1996). The daily rainfall data collected at Jasper Ridge for the 1994–1995 growing season was modified to include water manipulations representing the dry, average, and wet rainfall treatments. Each of the three water treatments was then simulated using Biome-BGC for ambient and twice-ambient atmospheric CO_2 concentrations. Total above-ground biomass harvested at the end of the growing season (Thayer and Field, unpublished data) was used for comparison to simulated total above-ground biomass production through the same date (Figure 2). Simulated biomass compared favorably with measured biomass across the wide range in rainfall and atmospheric CO_2 ($y = 1.13x - 82.27$, $r^2 = 0.95$). Within Biome-BGC, increases in atmospheric CO_2 enhance leaf-level carbon uptake and reduce stomatal conductance and leaf-level transpiration. More carbon is therefore taken up by the leaf, and less water is used, such that increased atmospheric CO_2 tends to enhance water availability in an annual grassland. Increases in productivity under elevated CO_2 can occur both because of increases in carbon uptake, as well as increases in plant water-use efficiency.

As part of the CO_2 Model-Experiment Activity for improved Links (CMEAL), we have used a suite of ecosystem models, including Biome-BGC, to examine the short- and long-term responses of ecosystems to elevated atmospheric CO_2. Simulations utilizing Biome-BGC (as well as the other ecosystem models) suggest that the response of ecosystem net primary production (NPP) to an instantaneous doubling

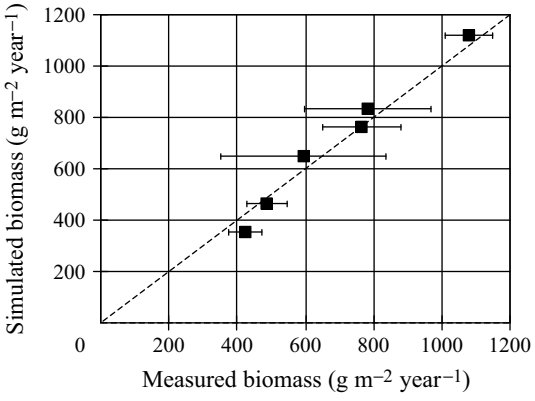


Figure 2 Measured above-ground biomass compared to the above-ground biomass simulated using BIOME-BGC for the water × CO_2 experiment in the Jasper Ridge CO_2 Experiment (Field *et al.*, 1997)

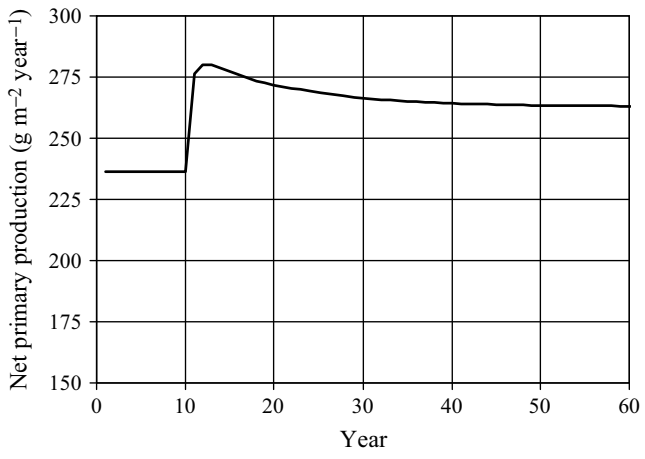


Figure 3 The response of annual NPP for the Jasper Ridge annual grassland to a step increase in atmospheric CO_2 (in year 10). Climate data for 1993 were used repetitively to drive the 60 year simulation

in atmospheric CO_2 is controlled by resource availability at a variety of timescales (Figure 3). Following a step increase in atmospheric CO_2 within the models, short-term increases in NPP (one to three years) are controlled primarily by physiological increases in carbon uptake due to increased atmospheric CO_2. At intermediate timescales (3–20 years), NPP often peaks and then declines. Intermediate timescale changes in NPP are controlled by soil nitrogen availability (plant uptake vs. decomposition) and the plasticity of plant carbon–nitrogen ratios. At longer timescales (>20 years), NPP usually reaches a new equilibrium as soil nitrogen availability comes into balance with plant demand for nitrogen. Long-term NPP responses are controlled by the net amount of nitrogen added (via atmospheric deposition) or removed (via leaching) from the ecosystem.

The CMEAL model intercomparison shows that increased ecosystem nitrogen use efficiency, either through, (a) increases in plant or soil carbon–nitrogen ratios or, (b) increased partitioning of nitrogen from soils to plants, is predicted to be an important mechanism for controlling the response of ecosystem productivity to elevated CO_2. The utility of ecosystem models is that they allow us to test a variety of hypotheses regarding the response of ecosystems to changes in the environment, as well as to extend the results of short-term experiments to the longer-term. An understanding of these interactions has helped experimentalists to identify specific, shorter-term, field measurements that can indicate the longer-term response of ecosystems to change. New measurements, in turn, help to improve the models, and it is this cyclic exchange of information between measurements and models that increases our understanding of how ecosystems respond to global environmental change.

See also: **Biome Models**, Volume 2.

REFERENCES

Field, C B, Chapin, III, F S, Chiariello, N R, Holland, E A, and Mooney, H A (1996) The Jasper Ridge CO_2 Experiment: Design and Motivation, in *Carbon Dioxide and Terrestrial Ecosystems*, eds G W Koch and H A Mooney, Academic Press, San Diego, CA.

Field, C B, Lund, C P, Chiariello, N R, and Mortimer, B E (1997) CO_2 Effects on the Water Budget of Grassland Microcosm Communities, *Global Change Biol.*, **3**, 197–206.

Hunt, Jr, E R, Piper, S C, Nemani, R R, Keeling, C D, Otto, R D, and Running, S W (1996) Global Net Carbon Exchange and Intra-Annual Atmospheric Carbon Dioxide (CO_2) Concentration Predicted by an Ecosystem Process Model and Three-Dimensional Atmospheric Transport Model, *Global Biogeochem. Cycles*, **10**(3), 431–456.

Jackson, R B, Sala, O S, Field, C B, and Mooney, H A (1994) CO_2 Alters Water use, Carbon Gain and Yield for the Dominant Species in a Natural Grassland, *Oecologia*, **98**, 257–262.

Mooney, H A and Gulmon, S L (1982) Constraints on Leaf Structure and Function in Reference to Herbivory, *BioScience*, **32**(3), 198–206.

Pierce, L L, Walker, J, Dowling, T, McVicar, T, Hatton, T, Running, S W, and Coughlan, J C (1993) Hydro-Ecological Changes in the Murray-Darling Basin, Part Three: a Simulation of Regional Hydrologic Changes, *J. Appl. Ecol.*, **30**, 283–294.

Running, S W and Coughlan, J C (1988) A General Model of Forest Ecosystem Processes for Regional Applications, I. Hydrologic Balance, Canopy Gas Exchange and Primary Production Processes, *Ecol. Model.*, **42**, 125–154.

Running, S W and Hunt, Jr, E R (1993) *Generalization of a Forest Ecosystem Process Model for other Biomes, Biome-BGC, and an Application for Global-Scale Models, Scaling Physiological Processes: Leaf to Globe*, eds J R Ehleringer and C B Field, Academic Press, San Diego, CA.

VEMAP Members (1995) Vegetation/Ecosystem Modeling and Analysis Project: Comparing Biogeography and Biochemistry Models in a Continental-Scale Study of Terrestrial Ecosystem Responses to Climate Change and CO_2 Doubling, *Global Biogeochem. Cycles*, **9**(4), 407–437.

Biosphere

The *biosphere* is the life-supporting layer surrounding the Earth's surface. This layer extends from the biologically active depths of the soil on land, and from the deep sea vents in the oceans, to a few kilometers into the atmosphere.

The biosphere constitutes the global ecosystem, including all life and the inert materials which provide the nutrition

and energy necessary to sustain life. This global ecosystem is largely powered by solar energy which drives the constant, characteristic cycling of matter that supports the self-reproduction of various large molecules and cells. Water is the key predisposing material for all life on Earth. In addition, various chemical elements (e.g., carbon, hydrogen, nitrogen, oxygen, phosphorus, and sulfur) combine into complex organic molecules during their cycles and these provide the fuel and building blocks to sustain populations of living organisms.

Human impact on the biosphere has increased enormously over the last 10 000 years. Through the development and global spread of animal husbandry and agriculture, humans now control about 40% of all solar energy captured by terrestrial organisms. In addition, the world's human population is growing at about 80 million per year and may reach 10 billion during the 21st century. In the context of simultaneous, global, economic growth, the increasing human populations threaten ever greater impacts on the biosphere in the future. Perhaps the most serious of these impacts is the release of greenhouse gases (including carbon dioxide, methane, and chlorofluorocarbons) into the atmosphere.

FURTHER READING

Kaukman, D G and Franz, C M (1993) *Biosphere 2000: Protecting our Global Environment*, Harper Collins, New York.

RICHARD A FLEMING *Canada*

Biosphere 2

see Controlled Environment Facilities in Global Change Research (Volume 2)

Biosphere Enhancement Ratio (BER)

Biosphere enhancement ratio (BER) is the biomass of a vegetative cover at a given elevated carbon dioxide (CO_2) concentration, divided by the biomass at some control (e.g., present day) level of CO_2 (g per g) (*see* **Plant Growth at Elevated CO_2**, Volume 2).

R E MUNN *Canada*

Biosphere Reserves

Jeffrey A McNeely
The World Conservation Union, Gland, Switzerland

Biosphere Reserves are areas of terrestrial and coastal/marine ecosystems, which are internationally recognized under United Nations Economic and Social Cooperation Organisation's (UNESCO's) Man and the Biosphere (MAB) Programme. They are designed to promote and demonstrate a balanced relationship between people and nature. They are nominated by national governments and remain under the sovereign jurisdiction of the states where they are located. As of 1999, 88 countries had established a total of 357 biosphere reserves. Several countries have extensive networks of Biosphere Reserves, such as the USA (29), Russia (19), Bulgaria (17), Spain (15), China (14), the UK (13), and Australia (12); but more than half of the participating countries have only one or two Biosphere Reserves. Emerging from UNESCO's MAB Programme as a concept in 1967, the first Biosphere Reserves were formally designated in 1976 (Batisse, 1997).

Biosphere Reserves must meet agreed criteria and adhere to a minimum set of management agreements before being admitted to the worldwide network. Each Biosphere Reserve is expected to perform three complementary functions.

- A biodiversity conservation function (with a focus on conserving a representative sample of major types of ecosystems).
- A development function (with a focus on humans in the biosphere, emphasizing an integrative role for local communities).
- A logistical function (combining conservation, research, education, training and monitoring).

The United Nations Environment and Social Cooperation Organisation (UNESCO) promotes Biosphere Reserves as illustrations of an open system that is able to interact with the surrounding landscapes much more extensively than can traditional protected areas with distinct boundaries. By their very nature as legally established units of land management, conventional protected areas have boundaries. Yet nature knows no boundaries, and recent advances in conservation biology are showing that most protected areas are too small to effectively conserve the large mammals or trees that they are designed to preserve. The boundary is too often also a psychological boundary, suggesting that since nature is taken care of by the protected area, people can abuse the surrounding lands, isolating the protected area as an island

of habitat which is subject to the usual increased threats that go with insularity.

Not only are the boundaries of protected areas difficult to defend even when they are well demarcated, but they also make it very difficult to adapt to change. Recent scientific conclusions about climate change make it very apparent that the distribution of biological systems are virtually certain to change as temperature, rainfall, and seasonality change in the coming years. Protected areas with hard boundaries are going to find it difficult to adapt to these changes.

The multiple-zone approach taken by Biosphere Reserves, on the other hand, is far more flexible and it should be reasonably easy to adjust the relatively soft boundaries between the core, buffer, and transition zones. Thus biosphere reserves may find it far easier to adapt to changing climatic (and other) conditions than will protected areas without buffers and with less flexible approaches to boundaries.

As designed by UNESCO, Biosphere Reserves are a special kind of conservation area with a nested set of zones having different management objectives. These include one or more core areas where human interventions are minimal, a surrounding buffer zone, and a flexible transition area, with the latter two designed to include people within the overall conservation framework.

The core area is devoted to long term protection, non-destructive research, and low-impact uses (such as education and some tourism), while resource management activities compatible with the conservation objectives are permitted in the buffer zone. In the transition zone, such management activities are promoted and developed with the active participation of local communities, management agencies, scientists, non-governmental organizations, cultural groups, economic interests and other stakeholders. Originally envisaged as a series of concentric rings, the three zones have been implemented in many different ways to respond to local needs and conditions.

The Biosphere Reserve approach has generated several variations, including *cluster reserves* (where several non-contiguous protected areas are linked in a network of collaboration on research and management, with the objective of fostering better conservation throughout the entire region). There are also transboundary protected areas (where adjoining protected areas from two or three countries are linked through various forms of cooperation on resource management, tourism, research, and public information) (Fall, 1999).

A few countries have enacted legislation specifically to establish Biosphere Reserves, but in most countries, the core areas and buffer zones typically are designated as part of other types of protected areas under national law, such as national parks. These core and buffer areas typically are public land while the transition area often is under private or communal forms of ownership.

Biosphere Reserves are designed explicitly to support social and economic development in the surrounding lands. Such benefits can include some access to wild plant resources, improved flow of water, and improved tourism. In many cases, local communities are also enabled to participate in decision making about the Biosphere Reserve. However, many of these benefits take time to develop, and Biosphere Reserves are perhaps best seen as the beginning of a socio-economic experiment that is designed to build trust and commitment between conservation managers and local human populations. While biodiversity conservation is a central feature of Biosphere Reserves, conservation is not necessarily a function of all Biosphere Reserve zones; conservation may be absent as an objective within the transition zone where the emphasis may be on the sustainable use of biological resources.

In 1983, the first International Biosphere Reserve Congress was held in Minsk (Bylorus, but then in the Soviet Union), which gave rise to an action plan for Biosphere Reserves that laid out a framework for action under a series of objectives. The action plan provided fairly broad guidance rather than detailed advice. A second International Biosphere Reserves Congress was held in Seville, Spain, in 1995. This event resulted in the Seville Strategy for Biosphere Reserves, and a Statutory Framework of the World Network of Biosphere Reserves (UNESCO, 1996). The statutory framework was formulated with the objectives of enhancing the effectiveness of individual Biosphere Reserves and strengthening common understanding, communication and cooperation at regional and international levels. It sets out clear criteria for qualifying as a Biosphere Reserve as well as a procedure for designating these areas by the International Coordinating Council of UNESCO's MAB Programme.

The statutory framework states that in order to qualify for designation as a Biosphere Reserve, an area should:

1. encompass a mosaic of ecological systems representative of major biogeographic regions, including a gradation of human interventions;
2. be of significance for biological diversity conservation;
3. provide an opportunity to explore and demonstrate approaches to sustainable development on a regional scale;
4. have an appropriate size to serve the three functions of Biosphere Reserves through appropriate zonation that includes a core area, a buffer zone, and an outer transition area.

In addition, organizational arrangements should be provided for participation of a suitable range of groups interested in the design and carrying out of the Biosphere Reserve, including public authorities, local communities, and private interests. In managing the Biosphere Reserve, provisions should be made for managing human use in the buffer zone; a management policy or plan for the area as a

Biosphere Reserve; a designated authority or mechanism to implement the management plan for the site; and programs for research, monitoring, education and training.

The statutory framework also called for the status of each Biosphere Reserve to be reviewed every 10 years, to ensure that it still meets the criteria for participating in this international program.

UNESCO expects Biosphere Reserves in the 21st century to be living examples of sustainable development, incorporating care of the environment and greater social equity, including respect for rural communities and their accumulated wisdom. They remain based on solid science, with research guiding management and feedback from monitoring systems helping to ensure that Biosphere Reserves are able to adapt to changing conditions. Ideally, Biosphere Reserves should be a means for the people who live and work within and around them to live in a balanced relationship with the natural world, while also contributing to the needs of society as a whole by demonstrating the way to a more sustainable future.

REFERENCES

Batisse, M (1997) Biosphere Reserves: A Challenge for Biodiversity Conservation and Regional Development, *Environment*, **39**(5), 8–33.

Fall, J J (1999) Transboundary Biosphere Reserves: a New Framework for Cooperation, *Environ. Conserv.*, **26**(4), 252–255.

UNESCO (1996) *Biosphere Reserves: The Seville Strategy and the Statutory Framework of the World Network*, UNESCO, Paris, http://www.unesco.org/.

Biotron

see Controlled Environment Facilities in Global Change Research (Volume 2)

Boreal Ecosystem-Atmosphere Study (BOREAS)

Ray Desjardins

Agriculture and Agri-Food Canada, Ottawa, Canada

BOREAS was a joint international experiment which took place from 1994 to 1996. It was undertaken to clarify the potential impact of the boreal forest on the concentration of greenhouse gases in the atmosphere and to determine how this ecosystem would respond to climate change. This study, which was carried out in Northern Canada, covered an area 1000 × 1000 km. It involved 75 scientific teams from university, government and industry. They quantified the fluxes of sensible and latent heat as well as carbon dioxide, methane and other important trace gases over a wide range of scales with the best technology available. The results are now published in scientific journals and most of the data collected are available on CD. More details on BOREAS are also available on the internet.

The boreal forest, which is located north of 48 °latitude, occupies 21% of the world's forested land surface. It contains about one-sixth of the carbon in the terrestrial biosphere. It has the potential to be an important sink for the carbon dioxide (CO_2) produced from the use of fossil fuel. One of the predictions of the new dynamic global vegetation models is that by 2100, boreal forests will occupy a substantial portion of the land that is now covered by tundra vegetation (Melillo, 1999). This northward extension of boreal forest is expected to cause an additional warming of approximately 4 °C in spring and about 1 °C in the other seasons (Foley *et al.*, 1994). Understanding the potential impacts of climate change for natural ecosystems is essential if we are going to manage our environment to minimize the negative consequences of climate change and maximize the opportunities that it may offer (Melillo, 1999). BOREAS was undertaken to clarify the role of this ecosystem with respect to global environmental changes (Sellers *et al.*, 1997). BOREAS was a joint international experiment, sponsored primarily by groups from the United States and Canada with the involvement of scientists from Britain, France, Japan and Russia. The region selected for BOREAS covered an area of 1000 × 1000 km in Manitoba and Saskatchewan Canada, and fieldwork took place between the years 1994 and 1996. The project involved 75 science teams from universities, governments, and industry. These teams were grouped into six disciplines: airborne flux and meteorology, tower flux, terrestrial ecology, hydrology, trace gas biogeochemistry, and remote sensing science.

To study differences in temperature and moisture, two areas, located 500 km apart, were selected within the BOREAS study region, each measuring approximately 50 × 50 km. The northern study area (NSA) was located near Thompson, Manitoba, and the southern study area (SSA) was located near Prince Albert, Saskatchewan. Flux towers were used for stand-scale measurements within each study area (five in the NSA, and six in the SSA). The tower flux data, which rely on the eddy covariance technique, provided a complete picture of the fluxes of water vapor, heat, CO_2 over days, seasons and even longer, for some of

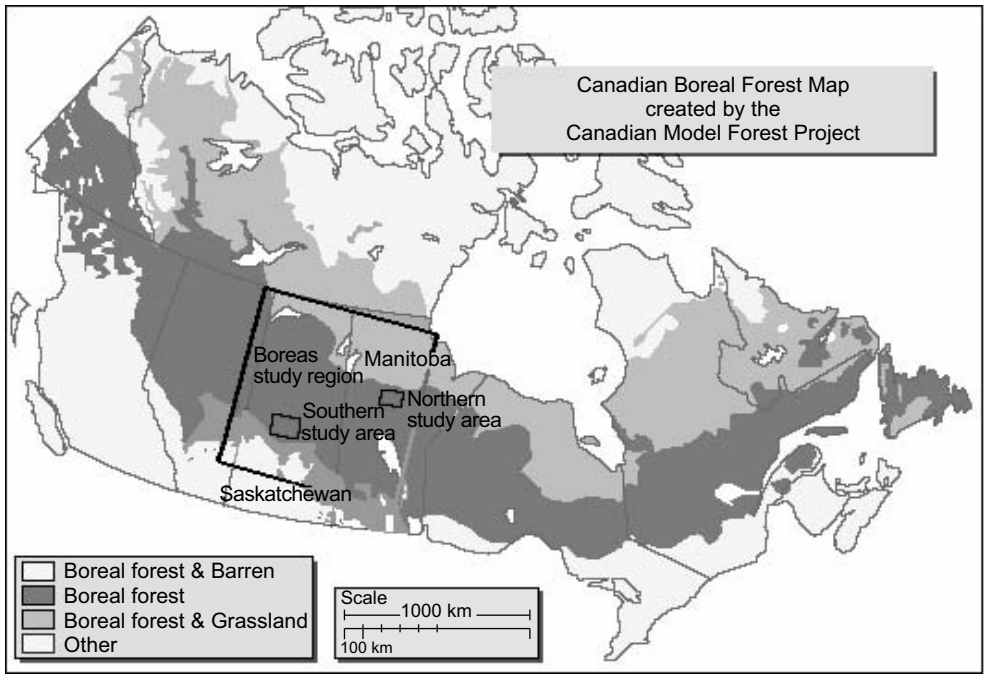


Figure 1 Map showing the BOREAS study region as well as the southern and northern study areas

the major species (Black *et al.*, 1996). Aircraft data were used to supplement and expand the data collected from the tower-based systems (Desjardins *et al.*, 1997). Transects and grid patterns were flown over both study areas (Figure 1). Satellite observations of reflected and emitted radiation were used to expand the sampling area to the entire BOREAS study region. These measurements were collected in conjunction with various ecological, meteorological and soil variables. These data are now being used to improve and validate models for obtaining regional flux estimates (Frolking, 1997).

BOREAS revealed several important features about the relationship between the boreal forest and the atmosphere. Firstly, it was not possible to measure the CO_2 flux over the entire boreal forest accurately enough to determine whether or not the boreal forest is the missing carbon sink in the global carbon cycle (Sellers *et al.*, 1997). Secondly, the atmosphere above the boreal forest is similar to that above a desert; it contains a deep planetary boundary layer and very dry air. Only small amounts of water enter the atmosphere from the boreal forest due to the low evapotranspiration rates of boreal conifers, and the presence of cold, large lakes that act as heat sinks rather than water evaporators (Sellers *et al.*, 1997). Thirdly, results from BOREAS are providing basic information to improve weather forecasts. For example, by providing a more realistic estimate of the albedo, in the presence of snow, BOREAS data allowed the European Center for Medium Range Forecasts to correct a bias in their surface temperature forecast over the boreal forest by up to 10 °C (Viterbo and Betts, 1999).

More details on these results can be found in the special issue of the *Journal of Geophysical Research*, Vol. 102, December 1997. Further information on BOREAS can be found on the Internet at http://boreas.gsfc.nasa.gov/BOREAS/BOREAS_home.html.

REFERENCES

Black, T A, den Hartog, G, Neumann, H H, Blanken, P D, Yang, P C, Russell, C, Nesic, Z, Lee, X, Chen, S G, Staebler, R, and Novak, M D (1996) Annual Cycles of Water Vapour and Carbon Dioxide Fluxes in and Above a Boreal Aspen Forest, *Global Change Biol.*, **2**, 219–229.

Desjardins, R L, MacPherson, J I, Mahrt, L, Schuepp, P, Pattey, E, Neumann, H, Baldocchi, D, Wofsy, S, Fitzjarrald, D, McCaughey, H, and Joiner, D W (1997) Scaling Up Flux Measurements for the Boreal Forest Using Aircraft-Tower Combinations, *J. Geophys. Res.*, **102**, 129 125–129 133.

Foley, J A, Kutzback, J E, Coe, M T, and Levis, S (1994) Feedbacks between Climate and Boreal Forests During the Holocene Epoch, *Nature*, **371**, 52–54.

Frolking, S (1997) Sensitivity of Spruce/Moss Boreal Forest Net Ecosystem Productivity to Seasonal Anomalies in Weather, *J. Geophys. Res.*, **102**, 29 053–29 064.

Melillo, J M (1999) Warm, Warm on the Range, *Science*, **283**, 184–185.

Sellers, P J, Hall, F G, Kelly, R D, Black, A, Baldocchi, D, Berry, J, Ryan, M, Ranson, K J, Crill, P M, Lettenmaier, D P, Margolis, H, Cihlar, J, Newcomer, J, Fitzjarrald, D, Jarvis, P G, Gower, S T, Halliwell, D, Williams, D, Goodison, B, Wickland, D E, and Guertin, F E (1997) BOREAS in

1997: Experiment Overview, Scientific Results, and Future Directions, *J. Geophys. Res.*, **102**, 28 731–28 769.
Viterbo, P and Betts, A K (1999) The Impact of ECMWF Forecasts of Changes to the Albedo of the Boreal Forests in the Presence of Snow, *J. Geophys. Res.*, **104**, 27 803–27 810.

Boreal Forest

E H (Ted) Hogg
Canadian Forest Service, Edmonton, Alberta, Canada

The boreal forest, or taiga, is one of the world's most extensive ecosystems, occupying vast areas of the northern landscape in North America and Eurasia. Natural disturbances such as fire and insect outbreaks, in combination with variation in climate, topography and soils, have led to a diverse patchwork of forest types, often interspersed with numerous lakes and wetlands.

Winter snow cover typically lasts for six months or more, and average year round temperatures are near or below freezing, resulting in widespread permafrost in the coldest portions of the boreal forest.

Although the boreal forest has been inhabited by indigenous peoples for thousands of years, the impact of human activity on the landscape has been minimal until relatively recently. Major human-induced changes began with intensive agricultural land use along the southern edge of the boreal forest, and then expanded into more remote areas with resource development for forestry, oil and gas, mining, and hydroelectric power during the latter half of the 20th century.

The influence of industrial activity has been significant in some areas, but global climate change could lead to more dramatic future changes in the boreal forest, where the climate has already been warming at a much faster rate than in most other parts of the world. Some of the changes noted to date include extensive thawing of permafrost and an increase in forest fire during the 1980s and 1990s, when temperatures were warmer than those recorded in any previous decades.

These changes are expected to accelerate during the 21st century, as levels of carbon dioxide (CO_2) in the atmosphere continue to rise. In certain areas of the boreal forest, where moisture and soil nutrients are in abundant supply, tree growth might increase in response to longer growing seasons and the effects of higher CO_2 levels. However, if the warming leads to drier climatic conditions, the combined effects of drought, fire and insect outbreaks could lead to losses of forest cover, especially in the southern boreal forest.

The results of large international experiments and computer modeling have begun to demonstrate the importance of the boreal forest to global cycles of carbon, water and energy. It has been shown, for example, that carbon uptake by the boreal forest caused a significant slowing of the rate of increase in atmospheric CO_2 levels over the past century, thus indirectly acting to slow the rate of global warming. However, scientists are divided as to whether northern forests will continue in this role as the climate warms: there are already indications that large areas of boreal forest have become sources of CO_2 release to the atmosphere, largely because of recent increases in fire and other disturbances. (See also ***Boreal Forest Carbon Flux and it's Role in the Implementation of the Kyoto Protocol Under a Warming Climate***, *Volume 4.)*

The boreal forest occupies a total area of about 12 million km^2, or about 8% of the earth's land surface. Its distribution extends around the North Pole, across 20 time zones and a distance of more than 15 000 km. Viewed from space, the overall shape of the boreal forest resembles a somewhat jagged doughnut, broken into two pieces. The largest piece is dominated by the boreal forests of Russia, including Siberia, as well as Finland, northern Sweden and Norway. The second piece includes the boreal forests of Canada and Alaska, and also takes in the northern part of Minnesota in the continental US (Figure 1).

The immense size of the boreal forest is partly an accident of geography. The zone of latitudes between 50–70 °North, where the boreal forest is located, has more than twice its share of land when compared to the earth as a whole. In contrast, the latitudinal zone between 50–70 °South is almost exclusively dominated by ocean, so that nearly all of the world's cold-climate forests are located in the northern hemisphere.

The name *boreal forest* comes from Boreas, the ancient Greek personification of the north wind. The Russian word taiga is sometimes also used as a synonym for boreal forest, but more commonly, *taiga* refers specifically to the more northerly portions of the boreal forest that are dominated by slow-growing, coniferous trees. The boreal forest is bounded to the north by subarctic woodlands and arctic tundra, and to the south by temperate forests and grasslands.

More than anywhere else in the world, the boreal forest is a land of extreme seasonal changes. Summers are short, but in combination with the long days, warmth and moisture are sufficient to allow tree growth. Winters are long and cold, with snow on the ground for four to eight months each year, and at the latitudes north of the Arctic Circle, include a mid-winter period of continuous night. A comparison of the climates of three representative locations in the boreal forest shows that Siberia has, by far, the coldest winters – the

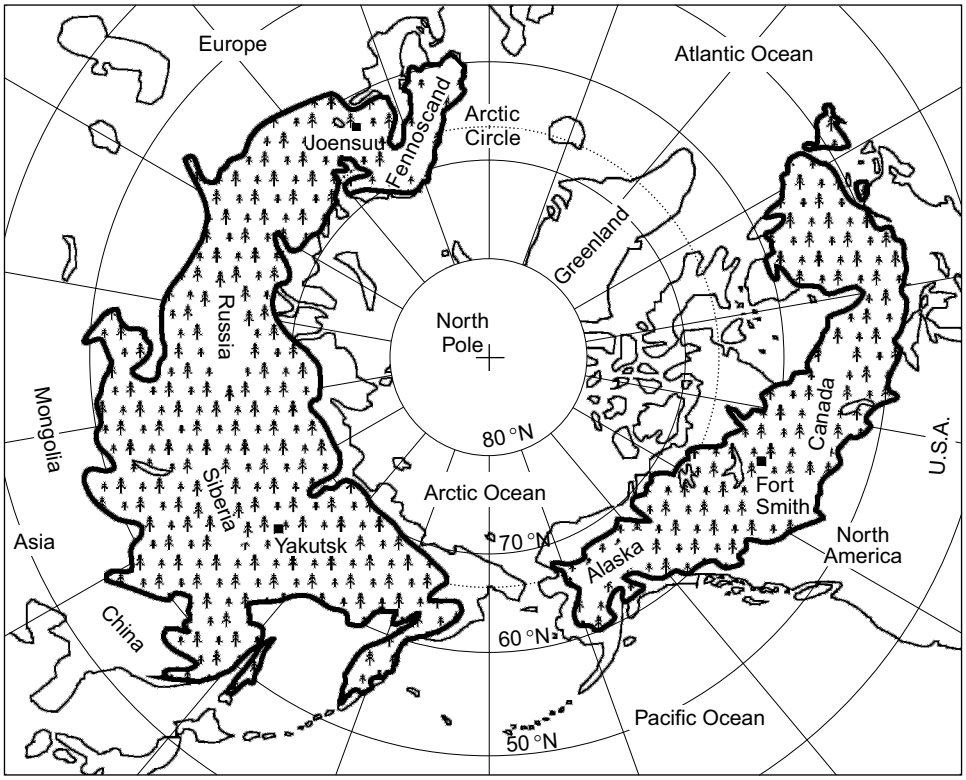


Figure 1 Global distribution of the boreal forest

average temperature in January at Yakutsk is −43 °C, and the lowest recorded temperature in the Siberian boreal forest (−68 °C) is colder than that recorded anywhere else in the world except Antarctica. Winters in the North American boreal forest are generally milder than in eastern Siberia; the average January temperature at Fort Smith, Canada (−25 °C) is typical across much of the region. Northern Europe is balmy in comparison, with an average January temperature of −12 °C at Joensuu, Finland (Figure 2).

Throughout the boreal forest, average annual temperatures range from about +3 °C to −10 °C. Deep layers of soil situated beneath insulating surface layers of moss and lichens are very slow in responding to summer warmth. Thus, they remain permanently frozen in areas where the year-round average temperature is less than 1–2 degrees below freezing. The continuously frozen moisture in the soil is referred to as *permafrost*, while the depth of soil near the surface that thaws each summer is called the *active layer*, where the roots of trees and other plants can obtain liquid moisture and nutrients. Moving north from the southern edge of the boreal forest, permafrost becomes increasingly common – first as isolated patches, but eventually forming a nearly continuous frozen layer of soil in the coldest regions of boreal forest (*see also* **Permafrost**, Volume 1).

Permafrost has profound effects on northern landscapes, as well as posing a major engineering challenge for construction and development. Houses in the most northerly areas of boreal forest are often perched above the ground on poles to prevent their heat from melting the underlying layers of permafrost that support them. In natural ecosystems, the presence of permafrost normally inhibits tree growth, because roots cannot penetrate very far into the soil for water and nutrients. On flat, waterlogged landscapes, however, trees are restricted to growing on patches of ground that are slightly elevated and supported above the water by the underlying permafrost. In these situations, thawing of permafrost can result in a drunken forest – where trees start leaning and eventually fall over into pools formed as the melting mixture of ice and soil collapses.

The boreal forest is dominated by evergreen coniferous trees, primarily spruce, pine and fir. However, deciduous forests of aspen or birch predominate over large areas of the southern boreal forest, both in North America and Eurasia. Another notable exception is in the extreme environment of northeastern Siberia, where the forests consist almost exclusively of larch. Larch is a cone bearing, needle-leaved tree like spruce and pine, but its needles fall off each autumn. Despite the immense area of the boreal forest, there are only 14 boreal tree species in Eurasia and a similar number in North America – none of these trees occurs naturally on both continents. Most regions within the boreal forest have only a few tree species.

Due to the low diversity of trees in the boreal forest, insect and disease outbreaks can produce dramatic changes

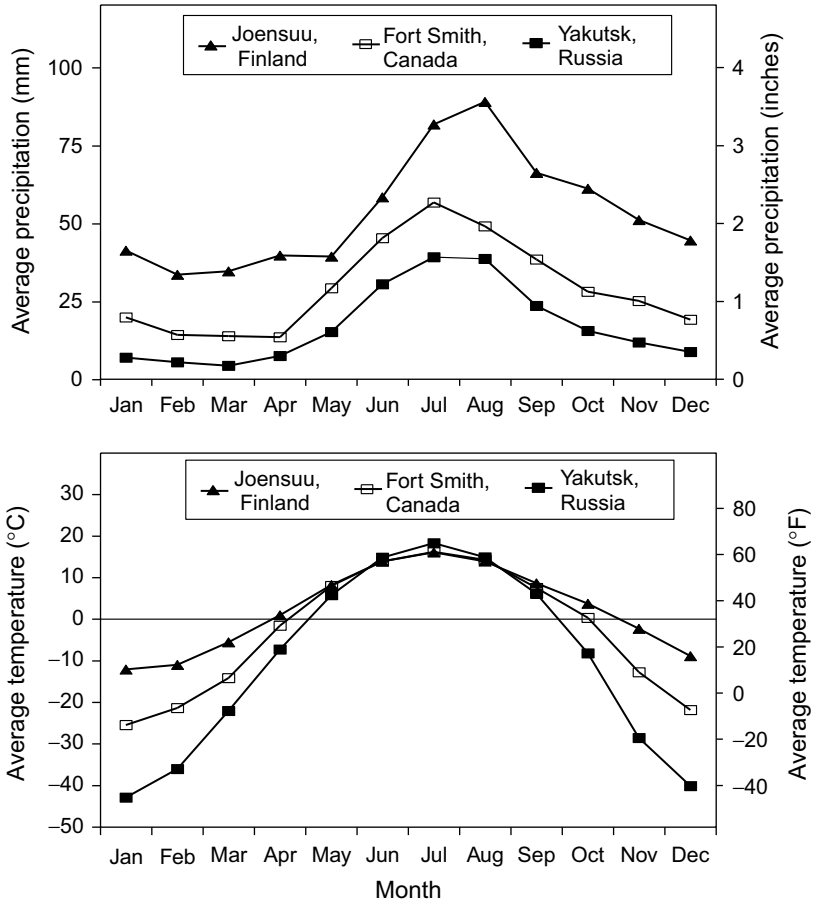


Figure 2 Average monthly climate for three representative climate stations in the boreal forest

on the landscape – even if only a single tree species is affected. In western Canada, for example, periodic population explosions of forest tent caterpillars feed voraciously on the leaves of aspen-dominated forests across areas as large as several million hectares, although the trees usually recover after the outbreak collapses. Coniferous evergreens, on the other hand, are often killed during major insect outbreaks. Examples include the spruce budworm in eastern and central Canada, spruce bark beetle in Alaska, and silkworms in Siberia; each of these insects has caused widespread death of trees over vast areas.

Fire is also an integral part of the boreal forest ecosystem. In remote areas, most forest fires are ignited by lightning and spread rapidly during periods of drought – not uncommonly, a single fire burns more than 100 000 hectares. In the Canadian boreal forest, there are an average of nearly 10 000 fires each year, burning a total of between one and eight million hectares annually. Over the centuries, fire has left its legacy on the boreal landscape in the form of a rich mosaic of boreal forest communities of different ages – mainly determined by the time that has elapsed since the last fire. Even with the recent expansion of the forest industry, fire continues to burn a much larger proportion of the Canadian boreal forest than that harvested by humans.

Another distinctive feature of the boreal forest is the abundance of peatlands, including bogs and fens, where peat has accumulated over thousands of years in waterlogged areas. Peat is formed by the deposition of moss and other carbon-rich plant material below the water table, where the low oxygen levels and cold temperatures greatly reduce their subsequent decay. Peatlands may be either treeless or support a stunted growth of spruce, pine or larch. Altogether, they occupy about 20% of the area of the boreal forest and are estimated to contain 450 billion tons of carbon. This represents about one-fifth of the total carbon stored in the world's vegetation and soils, and is nearly twice as much carbon as the total amount released to the atmosphere by human fossil fuel use since the industrial revolution.

Until recently, most of the boreal forest was a pristine, natural ecosystem that was sparsely populated by indigenous people who lived off the land. Although large areas remain as remote wilderness, significant changes in the landscape have occurred in response to human industrial activity. These include, for example, large reservoirs

created for hydroelectric power, roads, pipelines, and seismic lines that were cut through the forest for oil and gas exploration. Emissions of heavy metals and other pollutants from smelters have led to localized devastation of boreal forest, especially in Russia.

Acid rain, as well as acidic snowfall, dry particles, and deposition of gases occur even in remote areas, as pollutants such as sulfur dioxide and nitrous oxides can be carried long distances by the wind. Acid rain can have major impacts; for example, it has been implicated in the disappearance of peat-forming mosses in some parts of Europe. Paradoxically, however, acid rain may also act as a nitrogen fertilizer that can stimulate forest growth in areas where soil nitrogen concentrations are naturally low. The impacts are greatest on the Precambrian shield, where the soils are naturally acidic.

Global climate change could become the greatest single environmental threat to the boreal forest during the 21st century. Computer models of the earth's climate system indicate that a future doubling of CO_2 in the atmosphere over the next 100 years could lead to an average global warming in the range of 1.5–5 °C. However, the rate of warming is expected to be much greater, 3–8 °C, in the northern continental areas that include much of the world's boreal forest.

Although many uncertainties in these predictions remain, the climate record shows that large areas of the boreal forest have already experienced a warming of about 1.5 °C over the past century. Most of this temperature increase has been attributed to human activities, especially the global emissions of CO_2 from burning fossil fuels – coal, oil and natural gas. The warming has already caused the thawing of permafrost over extensive areas in the boreal forest; it is especially evident along the valleys of northern rivers, where the frozen slopes are transformed into mud, producing massive landslides.

Since the 1970s, the average area burned by forest fires has doubled, when compared to the previous 50 years, despite improvements in technology for fire detection and suppression. Because of the vast size of the boreal forest, natural forces – especially lightning and hot, dry weather – tend to dictate the number and size of the largest fires. The effectiveness of human efforts to stop fire in the boreal forest will continue to be limited, except in small, intensively managed areas. Therefore, it is likely that fire will continue to increase under the warmer and drier climate that has been predicted for many areas of the boreal forest.

Because the climate of the boreal forest is characteristically cold, it might be expected that warming would be beneficial to the growth of trees and other vegetation in the boreal forest. Global warming would lead to a lengthening of the summer growing season as well as warmer soils, where more nutrients are made available by the increased rate of soil decomposition. These positive effects on tree growth are likely to be greatest in the coldest parts of the boreal forest. Warming could also trigger the gradual colonization of the arctic tundra by boreal tree species; there is some evidence that this is already occurring in some areas. Another potential benefit of global change for the growth of trees and other vegetation includes the fertilizing effect of increasing CO_2 levels on photosynthesis – the process by which plants capture the sun's energy to produce the sugars needed for growth. There is considerable controversy, however, as to whether these benefits will be significant, especially in forested areas where growth is currently limited by moisture and nutrients.

From a human perspective, future changes in the southernmost regions in the boreal forest can be expected to have an especially major impact. These regions are most heavily utilized by people, for a wide range of forest-based activities, including forestry and recreation. Based on most model predictions, future changes in the southern boreal forest can be expected to differ dramatically from one region to another.

The boreal forest in eastern North America and Europe is climatically moist, and is bounded to the south by another forest type – the temperate forest. These forests are usually dominated by a high diversity of deciduous hardwoods, especially in North America where they consist of more than 50 tree species. It is thought that many of these hardwood trees are absent from the boreal forest because they are not adapted to survive temperatures lower than about −40 °C. Therefore, many of these species might be capable of colonizing the southern boreal forest as the winters become milder. The result could be a curious mixture of temperate and boreal organisms with some unexpected consequences for ecosystem functioning, although growth and diversity of trees might be expected to increase.

In western Canada and Siberia, the situation is very different. The climate is drier, so that the boreal forest is bounded to the south by grasslands: the Great Plains of North America and the steppes of central Asia. The forest–grassland boundary marks a dramatic change in both the appearance and ecosystem functioning of the landscape. This change is linked, both directly and indirectly, to moisture. In the boreal forest, the amount of rain and melting snow is usually more than enough to recharge the soil with moisture. The excess water drains off the land as runoff that maintains streams and rivers, thus ensuring that water levels in the numerous lakes remain relatively constant from year to year. The moist climate also allows the formation of peatlands, including bogs and fens. Peatlands depend on a constant water level in the soil close to the surface to maintain the oxygen-free conditions that severely retard the breakdown of organic matter. In the grasslands, virtually all aspects of the ecosystem are changed because of the dry climate. There is little surplus moisture, so that

there are very few streams, while lakes are often saline and intermittent – they dry up completely during periods of drought. Peatlands are generally absent. If wetlands are present, they are usually marshes, where little organic matter accumulates because it is repeatedly exposed to oxygen as water levels drop. Coniferous trees are also absent, probably because the establishment of seedlings requires a sustained period when soils are moist – a rare occurrence in the arid grasslands. In western Canada, stunted patches of aspen occur in the heavily cultivated, northern part of the prairies in a vegetation zone called the aspen parkland. They persist in this dry climate because they have the ability to sprout from their roots after drought or grass fires kill the aboveground stems.

If the future climate in boreal regions becomes drier, as predicted by some of the models, the impacts could be severe. For example, drought, in combination with outbreaks of insects or fungal pathogens, could lead to large-scale forest dieback. Forest cover might eventually be eliminated from some areas, as drought not only leads to more fire, but also prevents conifers from regenerating. Adopting innovative management practices could reduce adverse effects on forestry, but it could become increasingly difficult to maintain healthy forests in a natural state within parks and other protected areas.

We are now learning that changes in the earth's ecosystems can have a major effect on the amount of CO_2 in the atmosphere. Only about half of the extra CO_2 released by fossil fuel burning and land use changes can be accounted for by the observed rate of increase in atmospheric CO_2 concentration. This means that over the past century, the earth's oceans and vegetation must have been absorbing CO_2 at a faster than normal rate. Other studies point to northern forests as being important in slowing the rate of CO_2 increase. Potentially, this could be explained by increases in growth rates of these forests, but it is difficult to demonstrate that this has occurred. Another explanation that has emerged from Canadian research is that during the period between 1920 and 1970, the boreal forest experienced relatively little disturbance by fire and insects, which caused an increase in carbon storage in the trees and soil. The research indicates, however, that because of recent increases in disturbance, mainly fire and insects, the Canadian boreal forest may have already become an overall source of CO_2 to the atmosphere. Furthermore, there is concern that under a warmer and drier future climate, lowered water tables in peatlands could lead to much greater releases of CO_2: the peat would decompose more quickly, and the risk of deep-burning peat fires could increase dramatically.

The results of large international field experiments, together with computer simulation models, are beginning to show that the boreal forest has other, more direct effects on the earth's atmosphere and climate patterns. In one of these experiments, the boreal ecosystem–atmosphere study (BOREAS) (*see* **Boreal Ecosystem–Atmosphere Study (BOREAS)**, Volume 2), instruments mounted on towers (Figure 3), aircraft and satellites were used to measure the exchange of CO_2, water vapor and energy between the forest and the atmosphere. The information gained from this study provided insights into the ways that the boreal forest may be influencing its own climate. For example, measurements showed that spruce and pine forests produce a dark landscape that absorbs most of the incoming energy from the sun, even when snow is on the ground. The solar energy absorbed by the trees heats the air, which results in a climate that is warmer than it would be without the trees, because open landscapes reflect more of the sunlight back into space, especially when snow is on the ground.

The boreal forest also plays a role in returning moisture to the atmosphere, mainly through the process of water uptake

Figure 3 Tower in a black spruce (*Picea mariana*) stand in western Canada, used to measure the exchange of carbon dioxide, water vapor and energy between the boreal forest and the atmosphere

by roots and its release as water vapor from the leaves (transpiration). Such recycling of moisture could be important in stimulating summer rainfall across the boreal forest, most of which occupies continental areas where mountain ranges block the supply of moisture from the oceans. Even the type of forest could have a significant influence on regional weather and climate patterns: measurements from BOREAS showed that when green leaves are present, the release of water vapor from aspen forests is nearly twice as great as from nutrient-poor coniferous forests.

Ecologists and meteorologists alike have begun to recognize that the interaction between vegetation and climate is a two-way street: Not only does climate change affect vegetation, but large-scale changes in vegetation can also cause significant feedbacks on regional climate. Thus, the ability to forecast future global climate change and its impacts could ultimately depend on the depth of our understanding of these feedbacks, especially over large continental areas such as the boreal forest.

FURTHER READING

Apps, M J and Price, D T, eds (1996) *Forest Ecosystems, Forest Management and the Global Carbon Cycle*, NATO Advanced Science Institutes Series 1, Global Environmental Change, Vol. 40, Springer-Verlag, Berlin, 1–452.

Apps, M J, Price, D T, and Wisniewski, J, eds (1995) *Boreal forests and Global Change*, Kluwer, Dordrecht, 1–548.

Barber, V A, Juday, G P, and Finney, B P (2000) Reduced growth of Alaskan White Spruce from Temperature-induced Drought Stress, *Nature*, **405**, 668–673.

Bazzaz, F A and Fajer, E D (1992) Plant Life in a CO_2-rich World, *Sci. Am.*, **266**(1), 68–74.

Flannigan, M D, Bergeron, Y, Engelmark, O, and Wotton, B M (1998) Future Wildfire in Circumboreal Forests in Relation to Global Warming, *J. Vegetation Sci.*, **9**, 469–476.

Hogg, E H, Price, D T, and Black, T A (2000) Postulated Feedbacks of Deciduous Forest Phenology on Seasonal Climate Patterns in the Western Canadian Interior, *J. Clim.*, **13**, 4229–4243.

IGBP Terrestrial Carbon Working Group (1998) The Terrestrial Carbon Cycle: Implications for the Kyoto Protocol, *Science*, **280**, 1393–1394.

Kauppi, P E, Mielikainen, K, and Kuusela, K (1992) Biomass and Carbon Budget of European Forests, 1971 to 1990, *Science*, **256**, 70–74.

Kobak, K I, Turchinovich, I Y E, Kondrasheva, N Y, Schulze, E D, Schulze, W, Koch, H, and Vygodskanya, N N (1996) Vulnerability and Adaptation of the Larch Forest in Eastern Siberia to Climate Change, *Water, Air, Soil Pollut.*, **92**, 119–127.

Kurz, W A and Apps, M J (1999) A 70-year Retrospective Analysis of Carbon Fluxes in the Canadian Forest Sector, *Ecol. Appl.*, **9**, 526–547.

Pielke, R A and Vidale, P L (1995) The Boreal Forest and the Polar Front, *J. Geophys. Res.*, **100**, 25 755–25 758.

Pielke, R A, Avissar, R, Raupach, M, Dolman, H, Zeng, X, and Denning, S (1998) Interactions between the Atmosphere and Terrestrial Ecosystems: Influence on Weather and Climate, *Global Change Biol.*, **4**, 461–475.

Schulze, E D, Lloyd, J, Kelliher, F, Wirth, C, Rebmann, C, Lühker, B, Mund, M, Knohl, A, Milykova, I, Schulze, W, Ziegler, W, Varlagin, A, Sogachov, A, Valentini, R, Dore, S, Grigoriev, S, Kolle, O, Tchebakova, N, and Vygodskaya, N N (1999) Productivity of Forests in the Eurosiberian Boreal Region and their Potential to Act as a Carbon Sink – a Synthesis, *Global Change Biol.*, **5**, 703–722.

Sellers, P J, Hall, F G, Kelly, R D, Black, A, Baldocchi, D, Berry, J, Ryan, M, Ranson, K J, Crill, P M, Lettenmaier, D P, Margolis, H, Cihlar, J, Newcomer, J, Fitzjarrald, D, Jarvis, P G, Gower, S T, Halliwell, D, Williams, D, Goodison, B, Wickland, D E, and Guertin, F E (1997) BOREAS in 1997: Experiment Overview, Scientific Results, and Future Directions, *J. Geophys. Res.*, **102**, 28 731–28 769.

Shugart, H H, Leemans, R, and Bonan, G B, eds (1992) *A Systems Analysis of the Global Boreal Forest*, Cambridge University Press, Cambridge, 1–565.

Skinner, W R, Flannigan, M D, Stocks, B J, Martell, D L, Wotton, B M, Todd, J B, Mason, J A, Logan, K A, and Bosch, E M (2002) A 500 mb synoptic wildland fire climatology for large Canadian forest fires, 1959–1996, *Theor. Appl. Climatol.* (in press).

Buffering Capacity

Buffers are mechanisms in natural systems that serve to maintain stability (or resistance) against external physical or chemical stresses that would otherwise cause damage or malfunction. There are many buffering mechanisms in the environment. From a chemical viewpoint, one of the most important is the inherent capacity in soils to stabilize the pH of water filtering through it, thus protecting terrestrial and aquatic flora and fauna from large fluctuations in acidity.

Soils containing aluminosilicate clays play a vital role in neutralizing acidic inputs, particularly in areas subjected to high levels of acid rain. Clay soil particles contain negative charges on their surfaces to which base cations are electrostatically attracted. There are four common environmental base cations: sodium (Na^+), potassium (K^+), magnesium (Mg^{2+}), and calcium (Ca^{2+}). The neutralization reaction involves an exchange on the clay surface of a base cation by hydrogen ion (H^+):

$$\text{clay}\}\text{-}Na^+ + H^+(\text{soil solution}) \longrightarrow \text{clay}\}\text{-}H^+ + Na^+(\text{soil solution})$$

Thus, H^+ is removed from the soil solution and replaced by Na^+. Plants that take up soil water, and lakes receiving

drainage from the soils, are spared from acidification. The total H^+ that can be neutralized by a clay soil depends upon the density of negative charges on the clay surface and the availability of base cations. Together they define the soil's buffering capacity, which can be expressed in units of $kg\,H^+\,ha^{-1}$.

At the peak intensity of acid rain in Central Europe in 1980, deposition of H^+ on forest soils averaged about $3\,kg\,ha^{-1}\,year^{-1}$. The soils receiving this input had a buffering capacity of around $100\,kg\,H^+\,ha^{-1}$. If the acid rain had persisted at the 1980 levels in the succeeding decades, the soil's buffering capacity would have been exhausted in about 33 years.

When buffering in a clay soil is depleted, the pH of the soil solution rapidly descends below $pH = 4.2$. At the lower pH, aluminum, a highly toxic metal, is mobilized via dissolution in the soil solution. Aluminum mobilization in areas strongly affected by acid rain has been implicated in forest dieback and fish extinctions in lakes receiving the aluminum-contaminated waters. Thus, maintaining the buffering capacity in soils should be a high priority in environmental management of soils. This is an example of a chemical time bomb (*see* **Contaminated Lands and Sediments: Chemical Time Bombs?**, Volume 3).

WILLIAM M STIGLIANI *USA*

C

C_3 and C_4 Photosynthesis

James R Ehleringer and Thure E Cerling
University of Utah, Salt Lake City, UT, USA

Atmospheric carbon dioxide is reduced to organic forms through photosynthesis. Among terrestrial and aquatic autotrophs, there are three photosynthetic pathways. Here we discuss the ecological and evolutionary aspects of C_3 and C_4 photosynthesis, the two most widely distributed pathways.

Three photosynthetic pathways exist among terrestrial plants: C_3, C_4, and *crassulacean acid metabolism* (CAM) photosynthesis. C_3 photosynthesis is the ancestral pathway for carbon fixation and occurs in all taxonomic plant groups. The term C_3 photosynthesis is based on the observation that the first product of photosynthesis is a 3-carbon molecule. In C_4 photosynthesis, the initial photosynthetic product is a 4-carbon molecule. C_4 photosynthesis occurs in the more advanced plant taxa and is especially common among monocots, such as grasses and sedges, but not very common among dicots (most trees and shrubs). CAM photosynthesis, in honor of the plant family in which this pathway was first documented, occurs in many epiphytes and succulents from very arid regions. However, CAM photosynthesis is sufficiently limited in distribution that CAM plants are not an appreciable component of the global carbon cycle. This section focuses on the factors influencing the dynamics of C_3 and C_4 dominated ecosystems.

C_3 and C_4 photosynthesis are relevant to global change studies. These two photosynthetic pathways respond quite differently to changes in atmospheric carbon dioxide (CO_2) concentration and to changes in temperature. From a global change perspective, the kind of photosynthetic pathway present influences the magnitude of carbon fixation by the ecosystem, the quality of the plant food resource available to animals, and the isotopic composition of CO_2 released to the atmosphere.

C_3 photosynthesis is a multi-step process in which the carbon from CO_2 is fixed into stable organic products; it occurs in virtually all leaf mesophyll cells (Figure 1). In the first step, ribulose bisphosphate (RuBP) carboxylase-oxygenase (Rubisco) combines RuBP (a 5C molecule) with CO_2 to form two molecules of phosphoglycerate (3C molecule). However, Rubisco is an enzyme capable of catalyzing two distinct reactions: one leading to the formation of two molecules of phosphoglycerate when CO_2 is the substrate and the other resulting in one molecule each of phosphoglycerate and phosphoglycolate (2C molecule) when oxygen (O_2) is the substrate. The latter oxygenase reaction results in less net carbon fixation and eventually leads to the production of CO_2 in a process known as *photorespiration*.

The proportion of the time that Rubisco catalyzes CO_2 versus O_2 is dependent on the $[CO_2]/[O_2]$ ratio; the reaction is also temperature dependent, with oxygenase activity increasing with temperature. This dependence of Rubisco on the $[CO_2]/[O_2]$ ratio establishes a firm link between current atmospheric conditions and photosynthetic activity. As a consequence of Rubisco sensitivity to O_2, the efficiency of the C_3 pathway decreases as atmospheric CO_2 decreases.

C_4 photosynthesis represents a biochemical and morphological modification of C_3 photosynthesis to reduce Rubisco oxygenase activity and thereby increase photosynthetic rate in low CO_2 environments such as we have today (Figure 3). In C_4 plants, the C_3 cycle of the photosynthetic pathway is restricted to interior cells within the leaf (usually the bundle sheath cells). Surrounding the bundle sheath cells are mesophyll cells in which a much more active enzyme, phosphoenolpyruvate (PEP) carboxylase, fixes CO_2 (but as HCO_3) into oxaloacetate, a C_4 acid. The C_4 acid diffuses to the bundle sheath cell, where it is decarboxylated and refixed in the normal C_3 pathway. As a result of the higher activity of PEP carboxylase, CO_2 is effectively concentrated in the regions where Rubisco is located and this results in a high CO_2/O_2 ratio and limited photorespiratory activity. The additional cost of C_4 photosynthesis is the adenosine triphosphate (ATP) requirement associated with the regeneration of PEP from pyruvate.

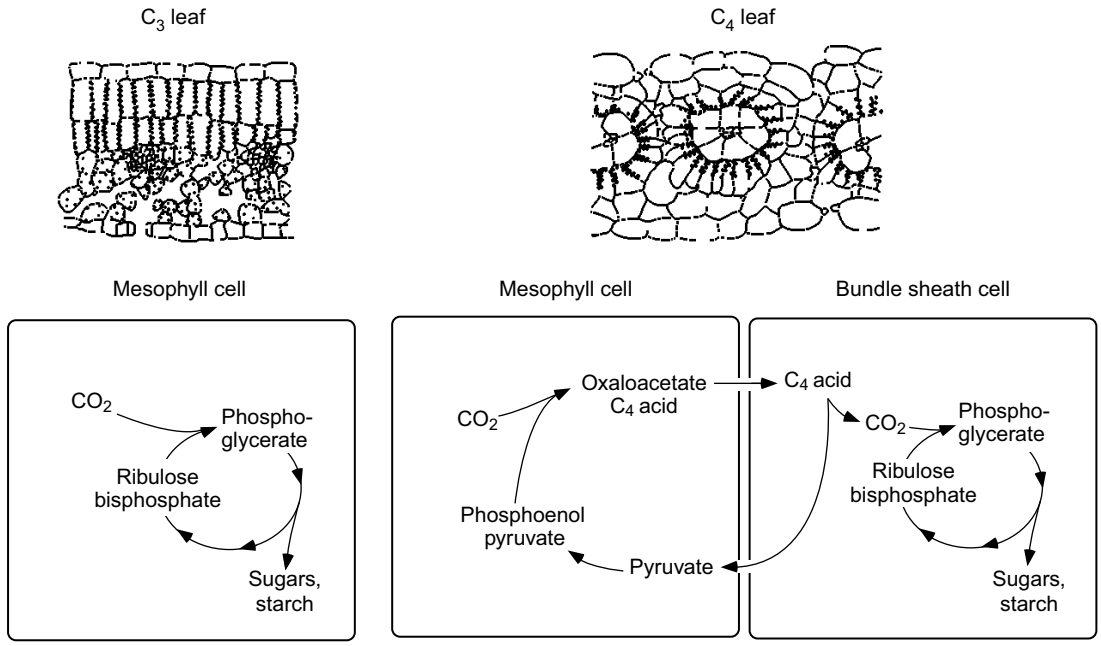


Figure 1 Leaf anatomy and basic photosynthetic biochemistry of C_3 and C_4 photosynthesis

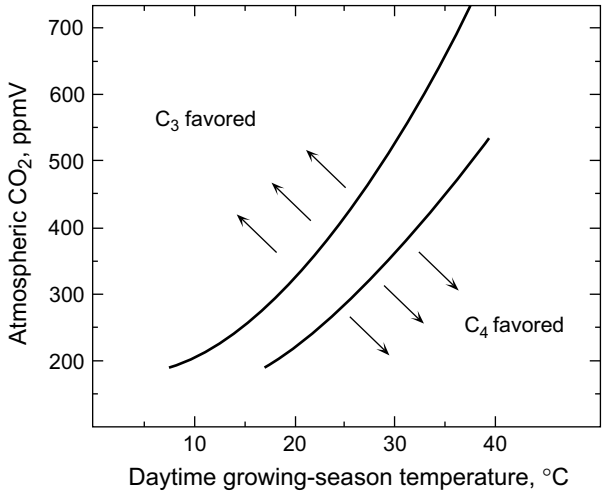


Figure 2 Modeled crossover temperatures of the photosynthetic light-use ef ciency (quantum yield) for C_3 and C_4 plants as a function of atmospheric CO_2 concentrations. The crossover-temperature is de ned as the temperature (for a particular atmospheric CO_2 concentration, in parts per million by volume, ppmV) at which the photosynthetic light-use ef ciencies are equivalent for both the C_3 and the C_4 plant. (Figure is modi ed from Ehleringer *et al.*, 1997)

C_4 photosynthesis is advantageous under low atmospheric CO_2 and/or high temperatures. The advantages of C_4 photosynthesis occur in lower CO_2 environments and/or high temperature environments, where photorespiration rates are relatively high in C_3 plants. Under these conditions, the efficiency of C_4 photosynthesis is greater than that of C_3 photosynthesis. However, under elevated CO_2 environments or at cool temperatures, the efficiency of photosynthesis is greater in C_3 photosynthesis because photorespiration is reduced and the additional ATP cost of C_4 photosynthesis makes it less efficient. We present these trade-offs graphically in Figure 2. From these light-use efficiency model predictions, it is clear that C_4 plants are not expected in environments where atmospheric CO_2 is greater than ≈600 parts per million (ppm). As atmospheric CO_2 decreases, C_4 plants should become most common first in the warmest environments, than in progressively cooler environments as CO_2 levels then continue to decrease.

The recent history of Earth has been one of decreasing atmospheric CO_2 levels. The atmospheric CO_2 levels are thought to have been higher in the Cretaceous than today (Figure 3). Some time following the Cretaceous (perhaps during the late Miocene), CO_2 levels decreased to about 500 ppm. During recent glacial–interglacial cycles, atmospheric CO_2 has fluctuated between 180 and 280 ppm. Since the dawn of the Industrial Revolution, atmospheric CO_2 levels have risen and these increases have been most dramatic since the 1950s (Figure 3).

C_4 photosynthesis occurs primarily within monocotyledonous plants. The flowering plants are classified as monocotyledons or dicotyledons. Approximately 6000 of the 15 000 monocotyledonous plants (primarily grasses and sedges) possess C_4 photosynthesis. In contrast, only about 1600 of the 300 000 dicotyledonous plants possess C_4 photosynthesis. In terms of taxonomic diversity, C_4 photosynthesis occurs in 401 monocotyledonous genera and 86 dicotyledonous genera.

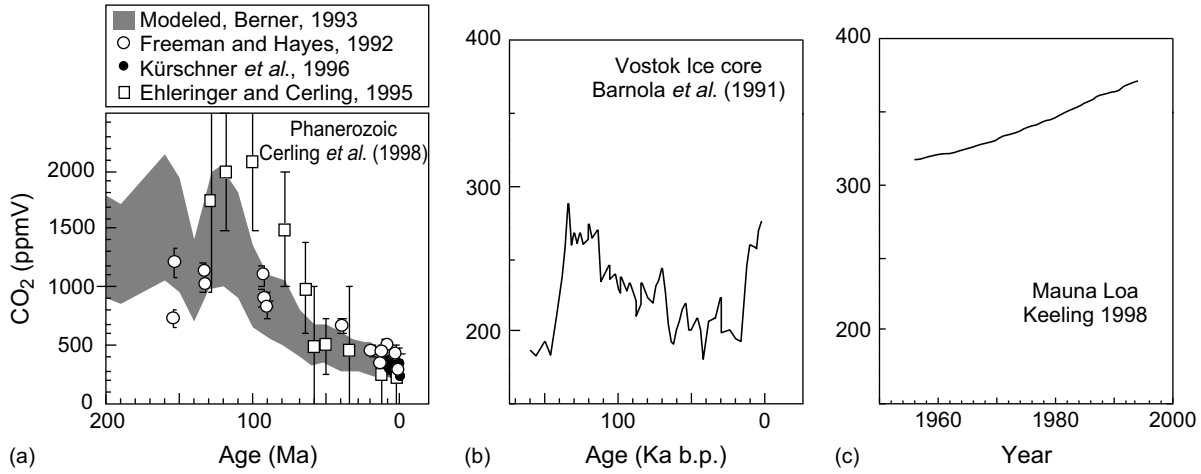


Figure 3 Patterns of atmospheric CO_2 concentrations through time. (a) Reconstruction of paleo CO_2 levels between 200 million years (Ma) ago and present; (adapted from Cerling *et al.*, 1998). (b) Reconstruction of atmospheric CO_2 from ice cores for the past 160 000 years; (adapted from Barnola *et al.*, 1991) (before present, b.p.). (c) Atmospheric CO_2 concentrations recorded at Mauna Loa, Hawaii since 1958; (adapted from Keeling and Whorf, 2000)

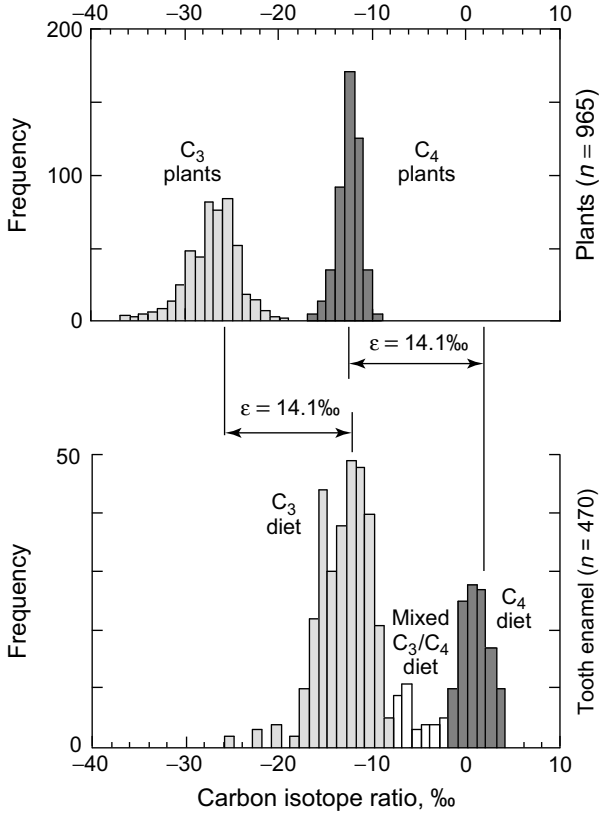


Figure 4 Histograms of the carbon isotope ratios of modern grasses and modern tooth enamel; (adapted from Cerling *et al.*, 1997)

C_4 grasslands emerged globally as an important ecosystem 6–8 Ma ago. Carbon isotope ratios are distinct and different between C_3 and C_4 plants (Figure 4). Variations in the carbon isotope ratios within a pathway reflect changes in environmental conditions and genetic differences among plants within a pathway type. The diet of animals (tissues in extant animals and tooth enamel in fossils) is reflected in their carbon isotopic composition. The offset of 14‰ between carbonate in tooth enamel and the C_3/C_4 food diet as shown in Figure 4 reflects a fractionation associated with apatite (calcium phosphate carbonate) formation in the tooth. Since animal fossils, such as teeth, are much more common in semi-arid and arid ecosystems, we can use the carbon isotopes in tooth enamel to reconstruct the presence of C_4-dominated ecosystems through time.

Between 8 and 6 Ma there was a global expansion of C_4 ecosystems (Figure 5). There is no conclusive evidence for the presence of C_4 biomass in the diets of mammals before 8 Ma, although the presence of small amounts of C_4 biomass is not excluded because of the uncertainty in the $\delta^{13}C$ end member for C_3 plants. By 6 Ma there is abundant evidence for significant C_4 biomass in Asia, Africa, North America, and South America, but not in Europe. Figure 5 documents several different ecosystem-type changes as recorded in mammalian tooth enamel. While each of these regions appears to have been dominated by C_3 ecosystems earlier in the Miocene, the C_3 Pakistani ecosystem was almost completely replaced by a C_4 ecosystem; African, North American, and South American ecosystems retained both C_3 and C_4 components. European and northern portions of North American ecosystems did not show any change in the fraction of C_3 biomass, remaining at virtually 100% C_3 ecosystems. The mixture of both C_3 and C_4 components within a grazing ecosystem can be achieved in one of two ways: a temporal separation with C_3 grasses active in winter–spring and C_4 grasses active in

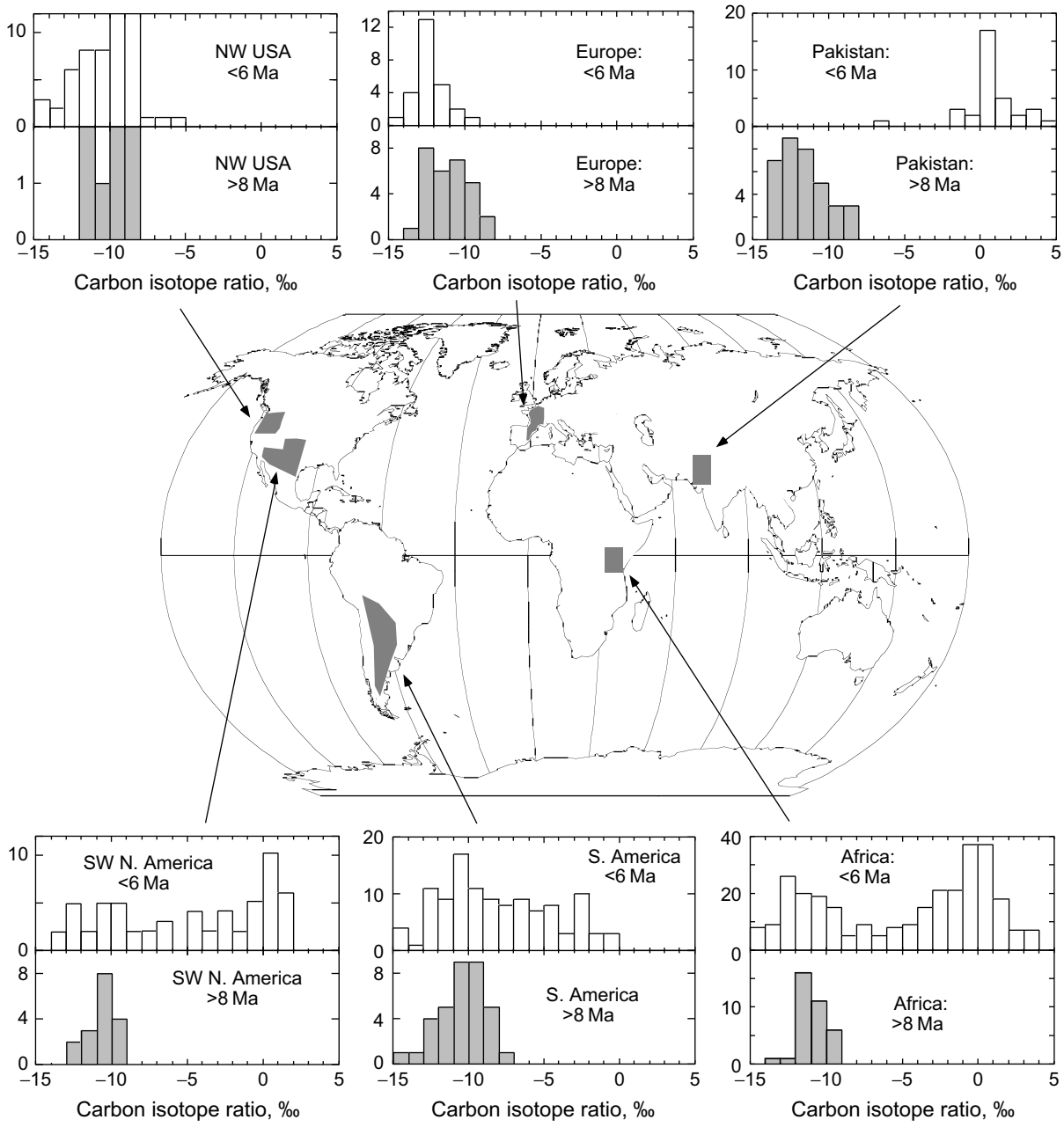


Figure 5 Histograms comparing the carbon isotope ratio values for fossil tooth enamel older than 8 Ma (lower charts) with those that are younger than 6 Ma for six regions of Earth; (adapted from Cerling *et al.*, 1998)

summer or by a monsoonal system with C_4 grasses and C_3 woody vegetation.

The isotopic evidence in tooth enamel indicates clearly that the expansion of C_4 ecosystems was a global phenomenon, persisting until today. The C_3/C_4 changes were accompanied by significant faunal changes in many parts of the world. It is unlikely that the global expansion of C_4 biomass in the late Miocene was due solely to higher temperatures or to the development of arid regions. There have always been regions of Earth with hot, dry climates. To explain the simultaneous global expansion of C_4 plants requires a global process. The light-use efficiency model (Figure 2) suggests that changes in atmospheric CO_2 are a strong possibility for this global mechanism. The supporting evidence indicates that the global expansion of C_4 ecosystems appears to have originated in warmer, equatorial regions and then spread to cooler regions, consistent with the temperature sensitivity predictions of the quantum yield model. Cerling *et al.* (1997) documented that within both modern and fossil horses (equids), the distributions

of isotope ratios strongly support a decrease in abundance of C_4 photosynthesis in moving from warm equatorial to cooler temperate latitudes.

C_4 grasslands are thought to have a wider distribution during glacial periods than they do today. The model in Figure 2 predicts greater global proportions of C_4 biomass during Pleistocene glacial, than interglacial periods. The published literature of organic $\delta^{13}C$ values in peat bogs and lakes from Central and Eastern Africa in regions (areas currently dominated by rain forest ecosystems) strongly suggest extensive C_4 expansion during the last full glacial period. Within western portions of North American, soil carbonate data also indicate that C_4 ecosystems were more extensive during the last glacial period than they are today. The model in Figure 2 suggests that mechanistically C_4 grasses were much more common during the glacial period when C_3 vegetation would have been CO_2 starved. Following deglaciation, the decline in C_4 abundances appears to be correlated with increases in atmospheric CO_2 levels.

What of the future? It is anticipated that atmospheric CO_2 levels will be double the current values by the end of this century. Until mankind's thirst for fossil fuels is quenched, it is likely that atmospheric CO_2 will continue to rise beyond levels experienced in the recent history of this planet. The quantum yield model predicts that as CO_2 levels rise, the atmosphere concentrations will once again cross the CO_2-threshold where C_4 plants do not have a competitive advantage over C_3 plants from the standpoint of reduced photorespiration and enhanced light-use efficiency. Will C_4 plants disappear in the future? That answer is unclear, but it appears that they will not have a competitive advantage. Certainly humans will continue to plant C_4 crops since many of today's most prominent crops are C_4 plants (e.g., corn and sorghum). Regardless of whether or not C_4 plants are as common among subtropical and tropical ecosystems, changes in atmospheric CO_2 will have continued impacts on the distributions of C_4 taxa.

REFERENCES

Barnola, J-M, Pimienta, P, Raynaud, D, and Korotkevich, Y S (1991) CO_2 – Climate Relationship as Deduced from the Vostok Ice Core: a Re-examination based on New Measurements and on a Re-evaluation of the Air Dating, *Tellus*, **43**(B), 83–90.

Cerling, T E, Ehleringer, J R, and Harris, J M (1998) Carbon Dioxide Starvation, the Development of C_4 Ecosystems, and Mammalian Evolution, *Proc. R. Soc. London*, **353**, 159–171.

Cerling, T E, Harris, J M, MacFadden, B J, Leakey, M G, Quade, J, Eisemann, V, and Ehleringer, J R (1997) Global Vegetation Change Through the Miocene–Pliocene Boundary, *Nature*, **389**, 153–158.

Ehleringer, J R, Cerling, T E, and Helliker, B R (1997) C_4 Photosynthesis, Atmospheric CO_2, and Climate, *Oecologia*, **112**, 285–299.

Keeling, C D and Whorf, T P (2000) *Atmospheric CO_2 Concentrations – Mauna Loa Observatory, Hawaii, 1958–1999*, (revised August 2000), http://cdiac.esd.ornl.gov/ndps/nd-p001.html.

Petit, J R, Jouzel, J, Raynaud, D, Barkov, N I, Barnola, J M, Basile, I, Benders, M, Chappellaz, J, Davis, M, Delaygue, G, Delmotte, M, Kotlyakov, V M, Legrand, M, Lipenkov, V Y, Lorius, C, Pepin, L, Ritz, C, Saltzman, E, and Stievenard, M (1999) Climate and atmospheric history of the past 420 000 years from the Vostok ice core, Antarctica, *Nature*, **399**, 429–436.

FURTHER READING

Jouzel, J, Loriu, C, Petit, J R, Genthon, C, Barkov, N I, Kotlyakov, V M, and Petrov, V M (1987) Vostok Ice Core: a Continuous Isotope Temperature Record Over the Last Climatic Cycle (160 000 years), *Nature*, **329**, 403–408.

CAM (Crassulacean Acid Metabolism)

see Photosynthesis (Volume 2)

Carbon and Energy: Terrestrial Stores and Fluxes

Dennis Baldocchi

University of California, Berkeley, CA, USA

The exchanges of solar energy, carbon dioxide, and water vapor between vegetation and the atmosphere are tightly coupled. Vegetation must attain energy to sustain the work that is needed to assimilate carbon dioxide, for biosynthesis, to evaporate water, and to transport nutrients from the soil to the plant. Concurrently, these activities require flows of substrate material, which are obtained from the atmosphere and soil. In this article, the physical, biological and chemical principles that govern carbon, water and energy exchange between the terrestrial biosphere and atmosphere are described for leaf, canopy, biome and global scales.

A cycle of birth, life, death and decay is sustained on Earth by the cycling of carbon and energy between the atmosphere, biosphere and geosphere. Carbon plays a major role in the formation of living matter because of its unique chemical property. Carbon has four electrons in its outer atomic shell. This property facilitates covalent bonding, thereby enabling carbon to form a myriad of long chain macromolecules. Examples of carbon-containing building blocks in living organisms include lignin, cellulose, and nucleic acid.

The cycle of carbon and energy begins with plants. They harvest solar energy through their ability to perform photosynthesis. In doing so, they assimilate CO_2 from the atmosphere and reduce it into a storable and useable chemical energy form. The subsequent respiration (oxidation) of high-energy carbon compounds, by plants, their consumers and the consumers of their consumers, taps this energy for work. Without chemical energy, originating from photosynthesis, living organisms would not be able to maintain their metabolism, build and repair their structure, and reproduce.

This essay discusses fluxes and stores of carbon and energy in the terrestrial biosphere. With regard to contemporary environmental issues, an understanding of this topic is needed to quantify: (1) how on-going perturbations in the Earth's climate and the chemical composition of its atmosphere will impact on the functioning of the biosphere; and (2) how potential changes in the functioning of the biosphere may feedback on the atmosphere and its climate system. The nature of this topic is complicated by the numerous biological, physical and chemical processes, which operate at length scales spanning from cells to the globe, on time scales from fractions of seconds to centuries and have non-linear responses to biotic and abiotic forcings.

CARBON BALANCE CONCEPTS

Carbon dioxide is the most common form of carbon in the atmosphere, and it is the primary source of carbon in organic matter. The mean mole fraction of CO_2 in the contemporary atmosphere (*circa* 1999) is near 365 parts per million. However, this value is increasing at a rate of about 2.4% per year, due to fossil fuel combustion, cement production and deforestation. Carbon dioxide ranks fifth in order of abundance in the atmosphere, after nitrogen (N_2), oxygen (O_2), argon (Ar) and water vapor (H_2O).

Natural fluxes of carbon into and out of the atmosphere cause the CO_2 concentration in the atmosphere to vary on diurnal and annual time scales. The daily variation is the greatest. Daytime photosynthesis depletes atmospheric CO_2 and nighttime respiration supplies it. The daily growth and descent of the planetary boundary layer affects the size of the atmospheric reservoir that is coupled with the surface. Minimum values of CO_2 concentration occur after midday when photosynthesis is greatest and the planetary boundary layer is fully grown and well mixed. At night, CO_2 concentrations become significantly elevated with respect to the global mean concentration. Nocturnal CO_2 concentrations are elevated because the boundary layer is quite shallow (a few hundred meters) and the ecosystem is respiring. It is quite common to observe CO_2 concentrations exceeding 500 ppm a few meters above the ground.

The annual cycle of atmospheric CO_2 results from net carbon assimilation during the growing season and net ecosystem respiration during the dormant period. The amplitude of the annual variation of CO_2 is more conservative than the daily cycle. At specific terrestrial sites, the annual variation is on the order of about 10–15 ppm. In contrast, the mean annual meridional difference in atmospheric CO_2 concentration between North and South Pole is about 4 ppm. The mixing of the atmosphere through its global circulation reduces meridional differences.

The major carbon pools in the terrestrial biosphere are generally differentiated as stores in the leaves, stems, and roots of plants, plant detritus and the soil. The amount of carbon stored in each biosphere pool depends on the difference between the rate at which carbon enters and leaves the biosphere. At steady state (or at equilibrium), the flows of carbon into and out of a specific carbon pool are equal and the amount of carbon in the pool remains unchanged.

The rate that carbon is assimilated by the biosphere is called *gross primary productivity* (GPP). In principle, the assimilation of CO_2 is a consequence of a balance between a leaf's demand and the atmosphere's supply of CO_2. Environmental factors that affect a leaf's demand for CO_2 include sunlight, temperature, humidity, soil moisture and available soil nutrients, such as nitrate and ammonium. Biotic factors affecting photosynthesis include growth form (herbaceous or woody, evergreen or deciduous), leaf type (broad-leaved or needle-leaved, dicot or monocot), photosynthetic pathway (C_3, C_4, Crassulacean acid metabolism (CAM)) and longevity (annual or perennial). The supply of CO_2 to a leaf is determined by the concentration of CO_2 in the atmosphere and by the rate of diffusion through the leaf boundary layer and stomatal pores.

Plants cannot acquire carbon without an expense. This expenditure comes in the form of respiration. Autotrophic respiration is the oxidation of organic compounds to CO_2 and water. Autotrophic respiration is responsible for producing biochemical energy compounds called adenosine triphosphate (ATP) and the reduced from of nicotinamide adenine dinucleotide phosphate (NADPH). These compounds are the biochemical equivalents to currency in an economy. Sustainable autotrophic respiration is constrained by the rate of photosynthesis. On a whole plant

basis, autotrophic respiration is generally 40–60% of gross photosynthesis.

The difference between gross photosynthesis and autotropic respiration of plants is called *net primary productivity* (NPP). The net annual production or expansion of leaves, stems, twigs, branches and roots is a result of NPP. To a first degree, NPP scales with available solar energy. The efficiency of this conversion is rather low for several reasons. The spectral range of photosynthetically active radiation (sunlight with wavelengths in the range between 0.4 and 0.7 μm) spans half of the solar spectrum. Between 5 and 10% of this radiation is reflected by leaves. Once a leaf absorbs a photon, chemical limitations and thermodynamic inefficiencies restrict the direct correlation between absorbed photons and molecules of carbon dioxide that are fixed. In the laboratory, eight photons are required for photosynthesis to fix one molecule of CO_2, whereas 12–20 photons are needed to fix a molecule of CO_2 in the field. Together, these factors yield a 2–3% conversion efficiency of solar radiation to photosynthesis. This conversion efficiency is even lower when plants are subject to soil moisture deficits, high vapor pressure deficits and extremely low or high air temperatures.

The net loss of carbon from the biosphere is a function of the amount of carbon stored in the system (C) and the average turnover time (τ). Two turnover times are considered when quantifying respiratory losses of carbon. One turnover time (τ_r) is associated with heterotrophic respiration (R_h), the respiration by fungi, bacteria and soil fauna and flora. The difference between NPP and heterotrophic respiration is called net ecosystem production (NEP):

$$\mathrm{NEP} = \mathrm{NPP} - \frac{C}{\tau_r} \tag{1}$$

NEP represents the net carbon flux between plants and the soil. The rate that carbon is lost from the soil system is accelerated as a soil warms and is restricted when a soil is too dry or wet.

The other carbon turnover time of interest is related to the frequency of disturbance, as in the cases of fire, harvesting, erosion or insect/pathogen infestation (τ_f). Net biosphere production (NBP) is defined as the difference between NEP and losses of carbon compounds from the biosphere via disturbance:

$$\mathrm{NBP} = \mathrm{NPP} - \frac{C}{\tau_r} - \frac{C}{\tau_f} \tag{2}$$

NBP is a process that occurs on time scales of decades to centuries in natural ecosystems, as most fields and forests do not burn or die from insect or pathogen infestation every year.

The time rate of change of carbon in a given pool, dC_i/dt, is a function of the fraction of net primary productivity that is allocated to each pool, a_i, and τ_i is the turnover time of each pool.

$$\frac{dC_i}{dt} = a_i\mathrm{NPP} - \frac{C_i}{\tau_i} \tag{3}$$

The carbon content in individual pools scales with available nitrogen. In general, ratios between carbon and nitrogen vary 5 : 1 and 10 : 1 for soil microbes and between 10 : 1 and 20 : 1 for leaves and roots. The carbon–nitrogen ratio for tree trunks is on the order of about 500 : 1. Fine roots possess the fastest turnover time, which is less than 1 year. Leaves possess turnover times on the scale of 1–3 years. Typical turnover times for stems, boles and soil organic matter can range from decades to centuries.

Soil is a medium for the anchorage of roots and the venue from where plants obtain water and nutrients. The soil is a reservoir of carbon that is continually undergoing losses as roots respire and microbes decompose litter, fine roots and fallen stems. The decomposition of soil organic matter is a function of temperature, soil moisture and litter amount and quality. Ecological data suggest that soil respiration and net primary productivity are positively correlated with one another. High net primary productivity produces more litter. More litter encourages higher decomposition, via soil respiration. This sequence of events promotes mineralization, which produces more nitrogen for plant growth. Because soil respiration involves enzymatic processes, it is highly dependent upon temperature, provided ample moisture is available. Respiration rates from moist but unsaturated soils double with every 10 °C increase in temperature.

The carbon cycle affects the formation and physical and chemical properties of soil in two ways. The dissolution of carbon dioxide in water forms carbonic acid, a weak acid with a pH near 5.6. Weathering of rock is caused by a geochemical reaction between mineral rock and carbonic acid. On annual time scales, the input of plant detritus (leaves and roots) into the soil leads to an accumulation of organic matter and nitrogen, thereby affecting the texture and chemical composition of soil.

ENERGY BALANCE CONCEPTS

The ultimate source of energy for photosynthesis, evaporation and warmth is the sun. With a radiant temperature of about 6000 K, the sun emits most of its radiation in the ultraviolet, visible and near infrared wavebands (0.35–2.7 μm), in accordance with Planck's Law. The Earth's atmosphere is mostly opaque to high-energy ultraviolet radiation, but it is highly transparent to visible and near infrared radiation. On a clear day about 70–80% of shortwave radiation that is incident at the top of the atmosphere is transmitted to the surface.

Atmospheric CO_2 affects energy exchange and life in many direct and indirect ways. Carbon dioxide is a greenhouse gas. It absorbs infrared radiation in the 2.67, 2.77, 4.25 and 15 μm wave bands. CO_2 absorbs and re-radiates infrared energy to the Earth's surface allowing the Earth's mean surface temperature to remain 15 °C above freezing (288 K). The maintenance of liquid water is crucial for the existence and functioning of life. Liquid water is a constituent of photosynthesis, it hydrates cells, and it is the medium by which nutrients are dissolved and exchanged between organs and storage pools. Water vapor is also a greenhouse gas. Together with CO_2 and other greenhouse gases, our atmosphere remains, on average, 51 K warmer than the temperature that the planet would have without an atmosphere.

The net radiation balance at the Earth's surface (R_n) is a function of the flux densities of incoming solar (R_g) and terrestrial radiation ($L\downarrow$) and the surface's albedo (α) and temperature (T_s):

$$R_n = (1-\alpha)R_g + (1-\varepsilon)L\downarrow - \sigma\varepsilon T_s^4 \quad (4)$$

The albedo of the surface defines what portion of incoming solar radiation is reflected. The albedos of conifer and broad-leaved forests, crops and snow are about 0.10, 0.15, 0.20 and 0.7, respectively. The Earth emits infrared radiation in proportion to the fourth power of its surface temperature. The other terms in Equation (4) represent the Stefan–Boltzmann constant (σ) and the surface emissivity (ε), which is near one for a black body radiator.

The net radiation balance at the Earth's surface is used to drive photosynthesis (P); evaporate water, and melt snow and ice (lambda E), heat the air (convection) (H); and heat the soil (G) and plants (S) (conduction);

$$R_n = H + \lambda E + G + S + P \quad (5)$$

In Equation (5), λ is the latent heat of evaporation and fusion and E is the flux density of evaporation.

There is a close linkage between a plant stand's carbon and energy fluxes through the canopy's ability to intercept solar radiation, its net primary productivity, its site water balance and the amount of leaf area the stand can sustain (Figure 1).

The amount of leaf area a landscape can support is related to its water balance, the difference between the precipitation and evaporation. Plant stands, which form closed canopies (whose leaf area index exceeds three), exist where the water balance is positive (precipitation exceeds evaporation). Plant stands with sparse vegetation inhabit regions with a negative water balance.

A dense, closed canopy will intercept most incident sunlight. Hence, it will experience relatively high rates of photosynthesis and a high surface conductance. This occurrence will translate into relatively higher rates of evaporation and relatively low rates of sensible heat transfer.

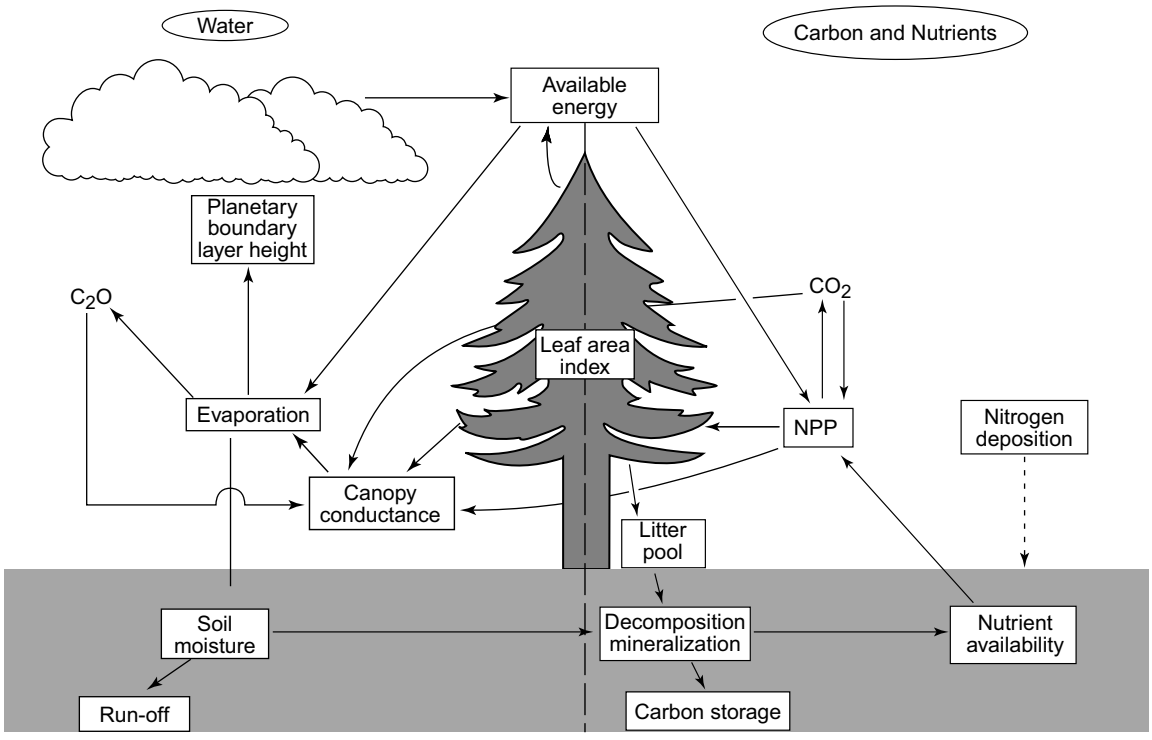


Figure 1 Diagram showing the feedbacks between energy, carbon, water and nutrient fluxes between a plant canopy and the atmosphere

These linkages can limit the height to which the planetary boundary layer develops over the course of a day. A shallow planetary boundary layer is moister than a deep boundary layer. Humidification of the boundary layer also imposes a negative feedback on further evaporation over the course of a day by reducing the atmosphere's vapor pressure deficit. On an annual time scale, a dense canopy is able to provide greater inputs of biomass and nutrients into the soil system. This effect will promote decomposition and mineralization and supply enough nutrients to support high rates of net primary productivity that sustain the leaf area of the closed canopy.

At the opposite extreme, an ecosystem that maintains a low leaf area index (less than three) will intercept less sunlight than a dense stand. Hence, it will experience relatively low rates of photosynthesis per unit ground area, it will possess a low canopy conductance and it will experience relatively low rates of evaporation and high rates of sensible heat transfer. A deeper and drier planetary boundary layer will develop over a sparse, unproductive region than would otherwise occur over a moister and more productive region. On yearly time scales, a sparse canopy provides low inputs of biomass and nutrients to the soil system and its soil is more likely to become dry. These effects will ultimately set limits on decomposition, net primary productivity, leaf area and evaporation.

CO_2 AND WATER VAPOR EXCHANGE AT VARIOUS SPATIAL SCALES

The net exchange of carbon and energy between the biosphere and atmosphere is associated with processes that operate across a spectrum of scales. In this section we discuss CO_2 and water vapor exchange at the cellular, leaf, plant, landscape and global scales.

The Cell

The chloroplast, a cell organelle, is the fundamental scale at which carbon exchange occurs. This is the venue where the suite of biochemical reactions associated with photosynthesis proceeds. A leaf is the fundamental scale at which photosynthesis is sustained as an integral process. The structure of a leaf is needed to support the array of chloroplasts, which harvest sunlight. A leaf also houses stomata, which regulate the diffusion of CO_2 between the atmosphere and the mesophyll cells in the leaf interior, and supports the veins that are tied to the plant's central plumbing, the xylem and phloem.

Leaf photosynthesis involves two separate sets of chemical reactions. One set occurs in sunlight and the other occurs in the dark. The light reactions involve the harvesting of sunlight by chlorophyll and the transfer of electrons to the produce of biochemical energy compounds ATP and NADPH via the process called photophosphorylation. The energy contained in molecules of ATP and NADPH is subsequently used in the dark reactions to drive the photosynthetic carbon reduction cycle, or Calvin Cycle. The functional relationship between leaf photosynthesis and sunlight can be quantified with a rectangular hyperbola, a diminishing returns function.

Three metabolic pathways are possible for the reduction of CO_2. These include the C_3 and C_4 photosynthetic pathways and the CAM (*see* **Photosynthesis**, Volume 2). The C_3 photosynthetic pathway is associated with 95% of plants. In this pathway, a three-carbon compound (phosphoglyceric acid) by a reaction between ribulose bisphosphate (RuBP, a five-carbon compound) and atmospheric CO_2. The C_3 pathway is less efficient than the C_4 pathway because the enzyme, Rubisco (ribulose bisphosphate carboxylase/oxygenase), has a dual and competing affinity with divalent oxygen and CO_2. The competitive reaction between O_2 and Rubisco initiates a biochemical cycle called photorespiration. This auxiliary cycle causes a fraction of CO_2, which is initially assimilated, to be lost. Photorespiration is a function of temperature, oxygen and CO_2 concentrations. The rate of photorespiration diminishes with lower O_2, higher CO_2 and lower temperature.

The C_4 pathway is used by many warm season grasses. The carboxylation of CO_2, by the C_4 photosynthetic pathway, begins with a reaction that is catalyzed by the enzyme PEP carboxylase. It forms an intermediate, four-carbon compound, malate. A key feature of C_4 photosynthesis is its association with a special leaf anatomy, called bundle sheaths. Once formed, the four-carbon compound, malate, diffuses deep within the interior of the leaf, where oxygen is nil. At this locale, malate is decarboxylated. CO_2 that is released by the reaction is, thereby, re-carboxylated via the C_3 pathway (photosynthetic carbon reduction reactions). The combination of the C_4 leaf's unique morphology and this suite of reactions causes this photosynthetic pathway to bypass photorespiration. Furthermore, C_4 plants have a greater quantum yield and higher water use efficiency than C_3 plants and are better adapted to warmer climates.

The CAM pathway is used by cacti and succulent plants in desert and arid regions. These plants keep their stomata closed during the day, but are able to harvest light energy. Then at night, they open their stomata, when compensatory water loss is less, and assimilate CO_2.

Dark respiration produces energy and substrate, in the form of ATP and carbon skeletons, respectively, that are used by growth and maintenance respiration. Dark respiration involves four linked processes. These processes are glycolysis, the oxidative pentose phosphate cycle, the tricarboxylic acid (or Krebs) cycle and oxidative phosphorylation. To produce ATP, these processes transfer electrons, with oxygen being the terminal electron acceptor. From a

stochiometric viewpoint, the respiration of glucose yields six CO_2 molecules, six oxygen molecules and 32 units of ATP. Basal rates of leaf respiration scale positively with leaf nitrogen content and negatively with leaf longevity. Short-term rates are an exponential function of temperature. Typically, respiration rates double with a 10 °C increase in temperature.

The Leaf

The diffusion of CO_2 into a leaf and the transpiration of water vapor from a leaf are regulated by stomata. Stomata are active pores in the epidermis of leaves that occupy less than 5% of a leaf's surface area. Pairs of specialized guard cells characterize stomata. Their opening and closing depends on changes in turgor pressure. Stomatal opening correlates positively with photosynthesis and sunlight. Increasing atmospheric humidity deficits, soil moisture deficits and CO_2 concentration induce declines in stomatal conductance. In general, stomata will open or close in response to environmental perturbations to keep the intercellular CO_2 concentration (C_i) at a relatively constant fraction of the atmospheric concentration (C_a). Typically, C_i/C_a is on the order of 0.7 for C_3 plants and 0.4 for C_4 plants. Theoreticians argue that stomata evolved this behavior to optimize the gain of CO_2 with respect to the loss of water.

Stomatal limitations upon photosynthesis also restrict transpiration since water and carbon dioxide molecules must diffuse through stomata. One consequence of a reduction in transpiration is an increase in sensible heat transfer and leaf temperature. This change forces a negative feedback upon net photosynthesis, as respiration increases exponentially in response to rising leaf temperatures.

Plant Canopy

At the plant canopy scale, the spatial distribution and orientation of leaves affect the performance of the canopy. In general, a plant canopy consists of an assemblage of plants, whose leaves and plants may be arrayed in a random, clumped or regular pattern and whose leaf angles may be erect, flat, or aligned in conjunction with the surface of a sphere. Inside a plant canopy some leaves are fully sunlit, others are exposed to a fluctuating regions of sunflecks, and the remainder exist in deep, but punctuated, shade. The age and the vertical and horizontal position of leaves, their angular relationship to earth–sun geometry, their exposure to current environmental conditions and the availability of soil moisture and nutrients affect how well leaves assimilate and respire carbon dioxide.

Leaves in full sunlight, and normal to the sun's beam, generally have higher photosynthetic capacities than shaded leaves. However, these leaves often experience saturated rates of photosynthesis, as biochemical constraints, such as high temperature and CO_2, limiting their ability to utilize the available solar energy. Furthermore, sunlit leaves are often warmer than air temperature. This situation promotes dark respiration, decreases the solubility of CO_2 relative to O_2 and decreases the Rubisco specificity factor, thereby restraining the maximal rate of photosynthesis attained by sunlit leaves.

The photosynthetic response to light of leaves in sunflecks is not the same as encountered by those exposed to steady light conditions. The history of the preceding sunflecks and the duration and intensity of the current sunflecks affect how the photosynthetic apparatus responds to changes in available light energy. As the light exposure of a leaf changes from dark to bright conditions, a dynamic response, known as induction, occurs if the previous dark exposure period was prolonged and de-activated the enzyme Rubisco. The consequence of this enzymatic de-activation is a delay in the attainment and a reduction in the magnitude of the next steady-state level of photosynthesis experienced by the leaf. The duration of this delayed response can exceed 20 min, but this time response is diminished if the leaf is exposed to repeated light flecks. Post-illumination carbon fixation is another important dynamic response experienced by leaves in fluctuating light. When the view of the sun is eclipsed by an upper leaf, rates of carbon fixation can be sustained for a spell as accumulated pools of metabolites are consumed. In fluctuating light environments, the occurrence of post-illumination photosynthesis enhances assimilation rates in comparison to rates that would occur otherwise under steady conditions with the same mean level of light exposure.

Leaves deep in a canopy adapt to shade. They have lower photosynthetic capacities than leaves positioned at the top of the canopy since less Rubisco, a nitrogen-rich compound, is allocated to them. This effect causes their photosynthesis rates to saturate at lower light levels. Furthermore, the photosynthesis system of these leaves tends to operate at a non-induced level since sunflecks are infrequent.

The functional relationship between light and photosynthesis at the canopy scale is a stark contrast with the behavior exhibited at the leaf scale. Canopy CO_2 flux densities over closed crops are a quasi-linear function of available photon flux density (Q_p), while over forests and sparse vegetation, non-linear response between canopy CO_2 uptake and absorbed photosynthetically active sunlight is observed. The relationship between canopy photosynthesis and available sunlight is affected by whether the sky is clear or cloudy. Rates of net ecosystem CO_2 exchange during clear conditions are diminished by more than half when compared with those observed under cloudy skies and similar photon flux densities. The effect is manifested

through the effect of sky conditions on the distribution of light through the canopy and on the energy budget of leaves. Under cloudy skies, incoming sunlight is more isotropic. The process allows more light to reach deeper into the canopy and illuminate typically sunlight-deficient leaves. On sunny days, the photosynthetic rates of upper sunlit leaves tend to be light-saturated. They also experience a higher heat load, which enhances respiration, so it lowers net photosynthesis.

The light-saturated response of canopy photosynthesis to changes in temperature is very plastic and parabolic. The temperature optimum for canopy CO_2 exchange rates of crops and forests, growing in temperate continental climates, is on the order of $20-30\,^{\circ}C$. The temperature optimum, however, can vary with species, ecotype, site and time of year. A diminution in leaf photosynthesis occurs at supra-optimal temperatures ($T_a > 35\,^{\circ}C$). This reduction occurs from a decrease in the relative solubility of CO_2 compared with O_2, a decrease in the specificity factor of Rubisco, an exponential increase in dark respiration rates and an accumulation of carbohydrates. Soil respiration rates increase exponentially with soil temperature and offset canopy photosynthesis rates.

There are coordinated, whole-plant responses to the environment that also affect photosynthesis. These include a response to temperature, nitrogen availability and soil moisture. Freezing and chilling injury are temperature-related phenomena that affect canopy photosynthesis in a negative manner. Chilling can cause membrane damage, while freezing can burst cells, causing mechanical damage, or dehydrate them. Low soil temperatures can also reduce photosynthesis through effects on water balance and stomatal conductance.

Leaf photosynthetic capacity increases with leaf nitrogen content, as nitrogen is a major component of Rubisco. Native plant stands tend to distribute leaf nitrogen in an optimal or coordinated manner through the canopy in proportion to the extinction of mean absorbed sunlight. Leaf nitrogen content will also vary during the growing season, thereby causing photosynthetic capacity to change.

Soil moisture deficits and high temperature stress have several negative impacts on physiological processes of plants. Soil moisture deficits can cause reductions of cell expansion, leaf area development, photosynthesis and transpiration. Non-stomatal and stomatal factors can limit carbon assimilation when soil moisture is lacking. Non-stomatal limitations of photosynthesis occur when photochemical conversion efficiency or the mesophyll conductance to CO_2 diffusion decreases.

With regard to stomatal limitations, there is ample evidence that a hormonal signal (abscisic acid, ABA) sent from the roots can induce stomatal closure. When ABA signals spawn stomatal closure, CO_2 concentration at the chloroplast is drawn down. Stomatal-induced reductions in chloroplast CO_2 concentration reduce carboxylation rates and alter the kinetic properties of the enzyme, Rubisco. An increase in the rate of photorespiration (relative to carboxylation) and a lowering of the optimum temperature for photosynthesis are two outcomes of changing Rubisco properties.

Canopy respiration involves two activities, growth and maintenance respiration. Canopy respiration is a relatively conservative fraction of canopy photosynthesis. A growing body of empirical evidence shows that this ratio is on the order of 40% and it is independent of temperature and plant life form.

BIOME CARBON AND ENERGY FLUXES AND CARBON STORES

Net CO_2 exchange and carbon content varies by biome (Table 1) due to climatic and genetic factors. Colder climates (e.g., boreal evergreen conifer forests and wetlands) and grasslands tend to store more carbon in the soil and warmer climates tend to store more carbon in above-ground biomass, as is evident in tropical forests and the great redwoods of the Pacific Northwest.

The carbon exchange rates of the world's biomes will vary markedly with time as photosynthetic capacity, the availability of solar radiation and soil moisture, and air and soil temperature vary over the course of the growing season. With regard to forests, deciduous forests lose carbon during the dormant period and gain carbon during the growing season between the period of moderate leaf expansion and leaf senescence. Evergreen forests (conifers and broad-leaved) are capable of acquiring carbon year round, as long as temperature exceeds freezing. Year to year changes on net carbon uptake of broad-leaf and conifer forests are affected markedly by the length of the growing season. Drier systems, like grasslands, savannas and boreal forests, experience numerous days during the heart of the growing season when they lose carbon to the atmosphere.

Several methods exist for measuring and evaluating biosphere–atmosphere carbon fluxes. At the global scale, the model inversion technique is used. This method treats the Earth as a giant cuvette. It combines wind information, from general circulation models, and CO_2 concentration data, from a global flask sampling network, to infer regional sources and sinks of carbon dioxide. The spatial resolution of this method is poor due to the mathematics of the inversion method and the poor resolution of the flask sampling network. Finer resolution information can be deduced by combining satellite imagery with ecosystem models. These carbon flux estimates, however, are discontinuous in time as they depend on images obtained during relatively cloud free conditions. At the stand-level, micrometeorological methods provide a direct and quasi-continuous measure of

Table 1 Estimates of net primary productivity, heterotrophic respiration and net ecosystem CO_2 exchange by plant functional type (Hunt *et al.*, 1996)

Land cover type	Land area (10^6 km^2)	NPP (10^{15} gC year^{-1})	R_h (10^{15} gC year^{-1})	NEE (10^{15} gC year^{-1})
Carbon fluxes				
C_3 grassland	34	9.6	10.7	−1.1
C_4 grassland	20	8.0	11.9	−3.9
Evergreen needle leaf	12	6.6	5.8	0.8
Deciduous needle leaf	2.5	0.5	0.3	0.2
Evergreen broad leaf	17	17.5	22.4	−4.9
Deciduous broad leaf	14	4.6	5.8	−1.2
Shrubland	18	5.2	8.7	−3.4
Total	117	52.0	65.7	−13.6
		Leaf C (10^{15} gC)	Stem C (10^{15} gC)	Soil C (10^{15} gC)
Carbon pools				
C_3 grassland		2.5	0	351
C_4 grassland		1.3	0	159
Evergreen needle leaf		4.3	102	190
Deciduous needle leaf		0.25	10.1	32
Evergreen broad leaf		4.3	224	169
Deciduous broad leaf		1.5	111	167
Shrubland		1.7	10.9	130
Total		15.9	458	1198

canopy–atmosphere CO_2 exchange. In practice, the method is restricted to relatively flat terrain. On the other hand, it gives good information on processes and controls of canopy carbon exchange. Biomass and soil sampling surveys can give long-term integrated measures of carbon fluxes. Yet, long intervals (years to decades) are needed between samples to resolve differences in growth or carbon accumulation in the soil. So this method is unable to provide much information on processes.

GLOBAL CARBON BALANCE

On the global scale, NEP is close to zero and its sign is uncertain (±2 Gt). In contrast, its constituent fluxes are quite large. GPP is about 100 Gt C year^{-1}, while autotrophic and heterotrophic respiration account for about 50 Gt C year^{-1}, each. Year to year variations in the global carbon balance can arise from an array of sources. Potential explanations include land use change (deforestation and burning), a response in biospheric respiration and photosynthesis due to changes in temperature and soil moisture during El Niño periods and a fertilization effect due to rising CO_2 concentration and nutrient deposition.

The major pools of carbon in the Earth's system include the atmosphere, plants, soil and oceans. On land, soil is the largest store of carbon. Although plants and trees are highly visible, more carbon is stored in the transparent atmosphere than in vegetation (Table 2).

Table 2 Carbon stores (Schimel, 1995)

Pool	Gt C
Atmosphere	750
Vegetation	610
Soil and detritus	2190
Surface ocean	1020
Deep ocean	38 100
Marine biota	3
Dissolved organic C	<700
Sediments	150

REFERENCES

Hunt, E R, Piper, S C, Nemani, R, Keeling, C D, Otto, R D, and Running, R W (1996) Global Net Carbon Exchange and Intra-annual Atmospheric CO_2 Concentrations Predicted by an Ecosystem Model and Three-dimensional Atmospheric Transport Model, *Global Biogeochem. Cycles*, **10**, 431–456.

Schimel, D S (1995) Terrestrial Ecosystems and the Carbon Cycle, *Global Change Biol.*, **1**, 77–91.

FURTHER READING

Amthor, J S (1995) Terrestrial Higher-plant Response to Increasing Atmospheric [CO_2] in Relation to The Global Carbon Cycle, *Global Change Biol.*, **1**, 243–274.

Field, C B (1991) Ecological Scaling of Carbon Gain to Stress and Resource Availability, in *Response of Plants to Multiple Stresses*, ed H A Mooney, Academic Press, London, 35–65.

Gifford, R M (1994) The Global Carbon Cycle: a Viewpoint on The Missing Sink, *Aust. J. Plant Physiol.*, **21**, 1–15.

Post, W M, Peng, T, Emanuel, W R, King, A W, Dale, V H, and deAngelis, D L (1990) The Global Carbon Cycle, *Am. Sci.*, **78**, 310–326.

Carbon Cycle

L Danny Harvey
University of Toronto, Toronto, Canada

The term carbon cycle refers to the flow or cycling of carbon from various compartments or components in the climate system to other compartments and eventually back to the original compartment. The atmosphere is one of many compartments in which carbon can occur, primarily as CO_2 and CH_4 – both of which are greenhouse gases. The carbon cycle is critically important to climate because it regulates the amount of these two important greenhouse gases in the atmosphere. The carbon cycle can be represented by a series of carbon storehouses or reservoirs that are interconnected by carbon flows or fluxes. Unlike the amount of water vapour in the atmosphere, which is directly and instantly driven by changes in the climate, the CO_2 concentration in the atmosphere responds to changes in the climate on time scales ranging from a few months to thousands of years.

Figure 1 shows the major carbon reservoirs. The number in each box is the estimated amount of carbon in the reservoir in units of gigatons (Gt, or billions of tonnes) prior to human disturbance. Carbon occurs in the atmosphere primarily as CO_2 and to a much smaller extent as CH_4; in the living biota and soils as organic matter; and in the oceans primarily as dissolved CO_2, HCO_3^- (bicarbonate ion), and CO_3^{2-} (carbonate ion). The CO_2, HCO_3^-, and CO_3^{2-} are collectively referred to as *dissolved inorganic carbon* (DIC). The ocean can be divided into a surface layer and the remaining subsurface water. The surface layer is referred to as the mixed layer because it is well mixed by winds and the effects of winter cooling. Water parcels in the mixed layer are frequently brought to the ocean–air interface and can exchange CO_2 with the atmosphere.

From Figure 1 it can be seen that: (1) the amount of carbon in the land biota is roughly comparable with the amount of carbon in the atmosphere; (2) the amount of soil carbon and detritus is about two times the amount in either the atmosphere or above-ground land biota; (3) the amount of carbon in the ocean mixed layer, which interacts directly with the atmosphere, is comparable with the amount of carbon in the atmosphere itself; and (4) the overwhelming majority of the total carbon in the biosphere + atmosphere + ocean system is in the deep ocean. The ocean can be thought of as a giant flywheel in the carbon cycle, while the atmosphere contains a comparatively small amount of carbon. Thus, a small percentage loss of carbon from the ocean to the atmosphere would translate into a large increase in the amount of carbon in the atmosphere.

Also shown in Figure 1, as numbers next to the arrows between the boxes, are the carbon fluxes in units of Gt year^{-1}. Carbon dioxide is incorporated into living plant material through photosynthesis and released to the atmosphere through respiration (the process by which organic matter is burned to provide energy, either by the plant itself, or by organisms that consume and thereby decompose organic litter). Some of the biospheric carbon is returned to the atmosphere as CH_4 rather than CO_2. This is represented by the flux of methane (M) in Figure 1, which is probably on the order of 0.2 Gt C year^{-1}. This flux is broken down into various biospheric sources in Table 1, where it is seen that the major sources are emissions from wetlands and termites. The CH_4 is largely oxidized to CO_2 within the atmosphere after about 10 years on average. Some of the biospheric carbon is also delivered to the oceans by

Table 1 Estimated emissions of methane from the terrestrial biosphere and oceans (natural emissions), taken from Lelieveld *et al.* (1998)

Source	Emission (Tg year^{-1})
Wetlands	145 ± 30
Termites	20 ± 20
Oceans	10 ± 5
Wild ruminants	5 ± 5
Freshwater	5 ± 5
CH_4 from sediments	5 ± 5
Total natural	190 ± 70

rivers (the riverine flux (R) in Figure 1), where it is either oxidized and returned to the atmosphere as CO_2, or sinks into the deep ocean. This flux is estimated to be on the order of 0.5 Gt C year^{-1}. Carbon is transferred between the atmosphere and the ocean mixed layer through diffusion of gaseous CO_2 across the air–sea interface. When the carbon cycle is in a balanced state, the total flux into each reservoir will equal the total flux out of the reservoir. Photosynthesis, respiration, and related processes also occur in the ocean. There is a small net flux into the deep ocean that, for a balanced system, will be equal to the rate of burial of carbon. Over geological time scales, the burial flux will be balanced by a variety of processes, including volcanic out-gassing of CO_2. These processes are not shown in Figure 1.

ATMOSPHERE-OCEAN MIXED LAYER-DEEP OCEAN EXCHANGE PROCESSES

When CO_2 enters sea water, the following chemical reactions take place

$$\underset{\text{(gas)}}{CO_2} + \underset{\text{(liquid)}}{H_2O} \longrightarrow \underset{\text{(aq) (carbonic acid)}}{H_2CO_3} \quad (1)$$

$$H_2CO_3 \longrightarrow H^+ + HCO_3^- \quad (2)$$

$$CO_3^{2-} + H^+ \longrightarrow HCO_3^- \quad (3)$$

giving the net reaction

$$H_2O + CO_2 + CO_3^{2-} \longrightarrow 2HCO_3^- \quad (4)$$

About 90% of the inorganic carbon in the oceans is in the form of HCO_3^-, about 10% is in the form of CO_3^{2-}, and less than 1% is in the form of CO_2. The transfer of CO_2 between the atmosphere and ocean is driven by the difference in CO_2 partial pressures (pCO_2). If, for example, the pCO_2 in the atmosphere is greater than that of the ocean surface, there will be a net flow of CO_2 into the ocean. The fact that only 1% of oceanic carbon is in the form of CO_2 means that the pCO_2 in the surface water is much smaller than it would be otherwise. This in turn allows the ocean to hold a large amount of carbon without creating a large back pressure that would drive CO_2 back into the atmosphere.

In the surface layer of the ocean, two biologically driven processes of importance occur: the construction of soft organic tissues by photosynthesis, which can be represented by the reaction

$$CO_2 + H_2O \longrightarrow CH_2O + O_2 \quad (5)$$

and the construction of calcareous skeletons, through the net reaction

$$Ca^{2+} + 2HCO^{-3} \longrightarrow CaCO_3 + H_2O + CO_2 \quad (6)$$

Photosynthesis in the oceans is estimated to take up about 50 Gt C year^{-1}. Most of this goes into short lived microorganisms, and is rapidly returned to ocean surface water when these organisms die or are eaten and burned by larger organisms. However, some of the soft tissue and skeletal material produced in the mixed layer ends up in the deep ocean through sinking dead microorganisms. The combined flux is crudely estimated to be about 10 Gt C year^{-1}, and is referred to as the biological pump. About 60–80% of this flux is due to falling soft tissue, and the remainder due to falling $CaCO_3$ particles. Most of the organic tissue decomposes in the upper 1 km of the ocean, while most of $CaCO_3$ dissolves below a depth of 3–4 km. The net result is to deplete the surface water in nutrients (primarily phosphate and nitrate) and DIC, to create maximum nutrient concentrations around a depth of 1 km, and to create a maximum DIC concentration at a somewhat greater depth in the ocean. Because of the sharp increase in the DIC concentration with increasing depth in the upper ocean, turbulent mixing (diffusion) tends to transfer DIC (and nutrients) upward. This is the main process that balances the downward transfer of DIC by the biological pump.

There are thus two reasons why the atmospheric CO_2 concentration is so low, and why most of the carbon is in the oceans: The first reason is the fact that most of the oceanic carbon does not occur as dissolved CO_2, due to the chemistry of DIC. The second reason is because the DIC concentration is comparatively low in the surface mixed layer but high in the deeper ocean, combined with the fact that the atmosphere is in direct contact only with the mixed layer but is largely isolated from the deep ocean. The mixing processes within the mixed layer are quite intense, but the mixing or exchange of DIC between the mixed layer and the deeper water is much slower. This is why the biological pump is able to maintain a low DIC concentration in the mixed layer in spite of 10% greater concentrations immediately below. The low DIC concentration in the surface water combined with the chemistry of DIC results in a low pCO_2 of surface water, and since the atmospheric pCO_2 tends to adjust to the pCO_2 of the water with which it is in contact, the atmospheric pCO_2 is also low.

WEATHERING AND FORMATION OF CARBONATE SEDIMENTS

The weathering of rocks on land removes carbon dioxide from the atmosphere. Rocks can be divided into two broad categories based on chemical composition: silicates, which contain minerals built around the elements silicon (Si) and oxygen (O), and carbonates, which contain minerals that are built around CO_3^{2-}. Most chemical weathering at the Earth's surface occurs through reaction with carbonic acid (H_2CO_3), which forms when CO_2 dissolves in H_2O. The net weathering reactions can be represented schematically as

$$CaCO_3 + H_2O + CO_2 \longrightarrow Ca_2^{+} + 2HCO^{3-} \qquad (7)$$

for carbonate rocks, and

$$CaSiO_3 + 2CO_2 + H_2O \longrightarrow Ca^{2+} + 2HCO^{3-} + SiO_2 \qquad (8)$$

for silicate rocks. Two CO_2 molecules are removed from the atmosphere for each Ca^{2+} released by the weathering of silicates, and one CO_2 is removed for each Ca^{2+} released by the weathering of carbonates. All of the Ca^{2+} delivered to the oceans, whether from the weathering of carbonates or silicates, is available to form $CaCO_3$ through the reverse of reaction Equation (7). This reaction adds only one CO_2 to the atmosphere + ocean system for each Ca^{2+} that is precipitated. If the Ca^{2+} came from the weathering of a carbonate rock, then the permanent burial of $CaCO_3$ has no net effect on the amount of carbon in the atmosphere + ocean system. When weathering of silicate rocks is followed by the burial of $CaCO_3$, there has been a net conversion from silicate to carbonate rocks and a net removal of CO_2 from the atmosphere + ocean system. However, as ocean sediments enter subduction zones through plate tectonic movements, $CaCO_3$ sediments are transformed back to silicate rocks and the missing CO_2 is produced. This CO_2 eventually finds its way back to the atmosphere through volcanic eruptions and hydrothermal vents (e.g., geysers), which are often collectively referred to as the volcanic source. Since the weathering and volcanic degassing fluxes are about $0.2\,\mathrm{Gt\,C\,year^{-1}}$, the removal of CO_2 by weathering + $CaCO_3$ sedimentation, and its restoration from volcanic degassing, completely replaces the atmospheric CO_2 content once every 3000 years or so.

The solubility of sea water with respect to $CaCO_3$ increases with increasing pressure and hence depth. As a result, the deepest ocean (below a depth of 1–4 km, depending on location) is unsaturated with respect to calcite (the most common form of $CaCO_3$), and $CaCO_3$ sediments are restricted to the ocean floor shallower than this depth.

RESPONSE TIME SCALES

Considerable insight into the behaviour of the carbon cycle can be gained by looking at the ratio of the size of a given reservoir to the total rate of flow of carbon into or out of the reservoir. This ratio is called the residence time (τ). It is the average length of time that a carbon atom spends in a reservoir before being transferred to another reservoir; it is also the length of time required, on average, to completely replace all the carbon atoms in a given reservoir. From Figure 1 it is seen that, based on the fluxes between the atmosphere, land biota, and soil carbon and detritus, the atmospheric reservoir of carbon is completely replaced about once every 5 years on average ($\tau_{atm} = 5$ years). The fluxes between the atmosphere and ocean mixed layer are also sufficient to completely replace the atmospheric CO_2 once every 5 years. Photosynthesis is sufficient to completely replace the land biota once every 6 years on average, giving $\tau_{biota} = 6$ years. About 60 Gt carbon enter and leave the soil and detritus pool every year, and since its size is about 1200 Gt, the mean residence time for this reservoir, τ_{soil}, is about 20 years. The oceanic mixed layer exchanges carbon with both the atmosphere and deep ocean; exchange with the atmosphere is sufficient to completely replace the surface layer carbon about once every 6 years, whereas exchange with the deep ocean is sufficient to completely replace surface layer carbon about once every 50 years. On the other hand, the deep ocean carbon is completely replaced only once every 3000 years or so.

The residence time also gives the characteristic time scale with which the carbon content of a given reservoir will adjust to a change in any of the fluxes between the reservoirs. Based on the residence times estimated above, we see that the carbon content of the biota, atmosphere, ocean, mixed layer, and – to a lessor extent – soils, will respond rather quickly to changes in the inter-box fluxes, but that the carbon content of the deep ocean is much slower to fully adjust. The atmosphere + biota + soils + mixed layer components thus form a tightly coupled subsystem which slowly exchanges carbon with the deep ocean.

When fossil fuel carbon is added to the atmosphere, the relevant response time scale is not given by the residence time of atmospheric carbon based on the exchange with the other reservoirs with which the atmosphere rapidly interacts. Rather, the response to fossil fuel carbon is given by the residence time of carbon in the coupled atmosphere–biosphere–mixed layer subsystem. This is because the rapid transfer of carbon from the atmosphere to the biota or mixed layer is quickly followed by a return flow to the atmosphere. The residence time of coupled atmosphere + biota + soils + mixed layer subsystem is given by the total mass of carbon in the subsystem (about 3100 Gt) divided by the rate of exchange with the deep ocean (about 10 Gt C but

highly uncertain). The result is a residence time of about 300 years. This is a very crude representation of highly complex processes, which are outlined in more detail below, but serves to illustrate in an intuitively simple manner how two very different response time scales for atmospheric CO_2 (5–6 years and 300 years) can arise and how we can get a feeling for what the magnitude of the response time scales should be.

STEPS IN THE RESPONSE OF THE CARBON CYCLE TO ANTHROPOGENIC EMISSIONS OF CO_2

The direct effect of an increase in atmospheric concentration is to stimulate photosynthesis by plants on land, and to induce a flow of carbon from the atmosphere to the ocean. Both responses are discussed in considerable detail by Harvey (2000), but are outlined here.

The products of increased terrestrial photosynthesis – more plant matter and litter – will eventually decompose, causing the carbon to return to the atmosphere. If the products of the increased photosynthesis make their way into long-lived soil components, then it will take longer for the return flow of carbon by respiration to balance the increased photosynthesis, and there will be a larger net storage of carbon on land. Climatic change, if it induces a greater incidence of forest dieback and forest fires, could eventually undo some or much of the initial carbon storage.

Although the flow of carbon into the ocean leads to an increase in the concentration of CO_2 dissolved in sea water, there is almost no stimulatory effect on marine photosynthesis – in contrast to the response on land. Recall that, today, less than 1% of DIC is in the form of CO_2. As CO_2 enters the ocean, reaction (4) proceeds to the right. However, once a new chemical equilibrium is achieved, the carbon does not distribute itself among CO_2, HCO_3^-, and CO_3^{2-} in the same proportions as existed before. Rather, the new carbon remains disproportionately in the form of CO_2, such that the increase in the CO_2 concentration is about ten times faster than the increase in DIC concentration (this factor of ten is referred to as the *buffer factor*, and will slowly increase as more CO_2 enters the ocean).

Since the oceanic pCO_2 is directly tied to the increase in CO_2 alone, this causes the back-pressure of oceanic CO_2 on the atmosphere to increase ten times faster than the relative increase in DIC as CO_2 enters the oceanic mixed layer, thereby limiting the ability of the mixed layer to absorb CO_2. On the other hand, the uptake of CO_2 by the mixed layer is much larger than if there were no chemical reactions at all (Harvey, 2000). One might expect, based on Figure 1 and our discussion of response times above, that the mixed layer would adjust to a pulse of CO_2 into the atmosphere on a time scale of 6 years. However, because of the buffer factor, the *turnover time* does not equal the *adjustment time* of the mixed layer to a perturbation. Rather, the mixed layer adjustment time is given by the turnover time divided by the buffer factor. Thus, the adjustment time is less than 1 year.

As DIC accumulates in the mixed layer and the oceanic pCO_2 increases, subsequent uptake of atmospheric CO_2 is tied to the slow downward mixing (diffusion) of DIC into the deeper ocean. In most of the ocean, the time scale to mix down to the 500 m depth is on the order of 100 years. Because this is much slower than the response of the mixed layer to changes in atmospheric CO_2 (about 1 year), downward diffusion serves as a bottleneck on the oceanic uptake of anthropogenic CO_2. Complete mixing of the CO_2 that would be released from combustion of a large part of the recoverable fossil fuel resource (which is estimated to be around 4000 Gt C) through the entire volume of the ocean would dilute the surface DIC concentration enough to allow the oceans to absorb about 75–80% of the carbon added to the atmosphere. However, on the order of 1000 years would be required for most of the absorption to occur.

Reaction (4) implies that, as ocean water absorbs CO_2, the concentration of CO_3^{2-} will decrease. This decreases the degree of supersaturation of water in the upper ocean with respect to $CaCO_3$. As CO_3^{2-}-depleted water penetrates into the deep ocean, the boundary between supersaturated and unsaturated water will rise, allowing the deepest $CaCO_3$ sediments to begin dissolving. As $CaCO_3$ dissolves, CO_3^{2-} is restored to the surrounding water. As this water mixes back up to the surface, the ocean can absorb more CO_2 from the atmosphere. The amount of $CaCO_3$ sediment that can dissolve is limited by the fact that, as dissolution occurs, the non-carbonate sediment is left behind, and this sediment eventually forms a capping layer thick enough to prevent further dissolution of carbonate sediments. Nevertheless, it is expected that enough carbonate sediment can dissolve to allow the oceans to absorb about half of the remaining 20–25% of anthropogenic CO_2, so that the oceans ultimately absorb 85–90% of CO_2 emitted into the atmosphere. However, several thousand years will be required for this to happen. Enhanced chemical weathering on land would take up the remaining 10–15%, but over a time period of about 200 000 years (Archer *et al.*, 1997).

THE CURRENT CARBON BUDGET

The natural carbon cycle, shown in Figure 1, was close to perfectly balanced over time periods of a decade or longer. Humans have upset this balance by adding new emission sources (fossil fuel combustion and deforestation), and the carbon cycle has responded with the oceans and terrestrial biosphere absorbing some of the remaining carbon (the added carbon (as explained above), creating carbon sinks). The difference between sources and sinks has accumulated in the atmosphere. Table 2 summarizes the current carbon

Table 2 The estimated current carbon budget, based on information in Harvey (2000). 1 Gt = 1 gigatonne = 1 billion metric tonnes

Term	Magnitude (Gt C year^{-1})
Emissions	
Fossil fuel emission	5.5 ± 0.5
Tropical deforestation	1.6 ± 1.0
Regrowth of temperate forests	$-0.5 \pm$
Net anthropogenic emission	6.6 ± 2.0
Sinks or accumulation in atmosphere	
Oceanic sink	2.0 ± 1.0
Terrestrial biosphere sink	1.5 ± 1.0
Accumulation in atmosphere	3.3 ± 0.2

budget. There is considerable uncertainty associated with the size of the source due to deforestation, in the partitioning of the total sink between the oceans and the terrestrial biosphere, and in the geographical distribution of the terrestrial biosphere sink.

REFERENCES

Archer, D, Kheshgi, H, and Maier-Reimer, E (1997) Multiple Timescales for Neutralization of Fossil Fuel CO_2, *Geophys. Res. Lett.*, **24**, 405–408.

Harvey, L D D (2000) *Global Warming: The Hard Science*, Prentice Hall, Harlow, UK, 336.

Lelieveld, J, Crutzen, P J, and Dentener, F J (1998) Changing Concentration, Lifetime and Climate Forcing of Atmospheric Methane, *Tellus*, **50B**, 128–150.

Carbon Monoxide

see Carbon Monoxide (**Volume 1**)

Carnegie-Ames-Stanford Approach (CASA)

James T Randerson

California Institute of Technology, Pasadena, CA, USA

The Carnegie–Ames–Stanford Approach (CASA) terrestrial biogeochemical model was specifically designed to take advantage of global satellite observations. In applications over the last decade, the CASA modeling framework has provided several insights about changes in ecosystem function at the scale of biomes and continents.

STRUCTURE

CASA calculates the net carbon balance of an ecosystem as the difference between plant growth (or net primary production (NPP)) and microbial decay (or heterotrophic respiration) (Potter *et al.*, 1993). NPP is estimated at each model time interval, t, and grid cell, x, as the product of absorbed photosynthetically active radiation (APAR; with units of MJ per unit time) and a light use efficiency that reflects the capability of plants to convert light energy into carbon compounds for growth and reproduction (ε; with units of g C/MJ):

$$\mathrm{NPP}(x, t) = \mathrm{APAR}(x, t) \times \varepsilon(x, t) \quad (1)$$

where APAR is derived from satellite observations of *normalized difference vegetation index* (NDVI) (*see* **Remote Sensing**, Volume 2), which is constructed from reflected visible and near infra-red light, and surface radiation in the visible part of the spectrum (specifically known as *photosynthetically active radiation* or PAR, *see* **PAR (Photosynthetically Active Radiation)**, Volume 2). Multiple National Aeronautics and Space Administration (NASA) and National Oceanic and Atmospheric Administration satellites in polar earth orbits provide continuous global coverage of NDVI. NDVI captures seasonal and interannual variation in canopy development and thus the fraction of PAR (FPAR; unitless) absorbed by plants. To obtain APAR, FPAR is multiplied in CASA by estimates of PAR at the top of the canopy. PAR is obtained from a separate set of satellite-derived cloud and aerosol measurements that are used to attenuate incoming solar radiation and thus calculate the light reaching the surface of the Earth. While most spatial and temporal variation in CASA NPP arises from light availability (PAR) and canopy development (FPAR), a small amount of additional variation is introduced by the light-use efficiency term (ε) that responds to locally varying soil moisture and temperature stress.

The simplicity and success of Equation (1) is based on the principle that plant investment in light harvesting (and thus satellite-detectable APAR) is optimized so that all other resources (e.g., below-ground levels of soil nitrogen and phosphorous) are equally limiting to growth (Field *et al.*, 1995).

NPP in CASA is allocated to leaves, wood, and fine roots following an allocation scheme that is sensitive to light limitation, atmospheric carbon dioxide (CO_2), and water stress. The lifetime of leaves, wood, and fine roots are

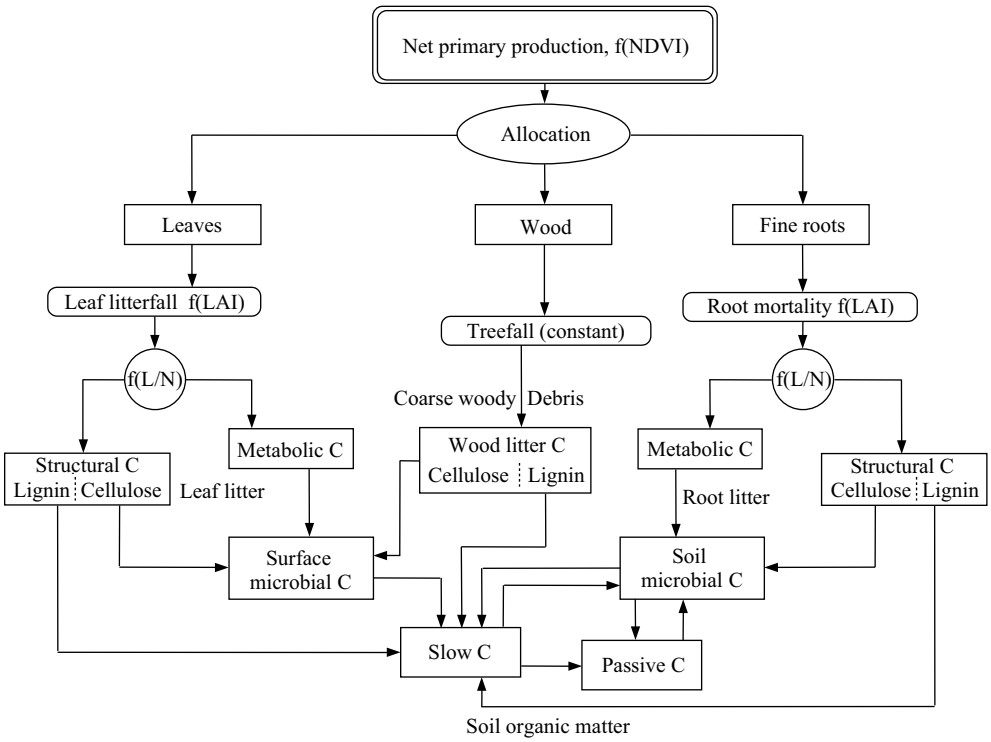


Figure 1 NPP in the CASA terrestrial biogeochemical model is driven by satellite NDVI observations. Carbon transfers in the model follow the arrows during each time interval. Each transfer from litter and soil organic matter pools is microbially-mediated and involves a heterotrophic respiration ux to the atmosphere. At the global scale, the model typically operates with $1 \times 1^{\circ}$ grid cells and a monthly timestep

prescribed as a function of biome type. The mortality (or shedding) of leaves and fine roots and subsequent access by the heterotrophic community is based on satellite-derived changes in leaf area (again derived from NDVI). Decomposition of plant litter and the turnover of soil organic matter depend on soil moisture, temperatures, and the chemical composition (the carbon to nitrogen ratio and lignin content) of incoming plant litter and coarse woody debris following the structure of the CENTURY biogeochemical model (Figure 1) (*see* **CENTURY Ecosystem Model**, Volume 2). The heterotrophic respiration flux (R_h) associated with decomposition has a positive non-linear response to monthly air temperature (described by a Q10 value of 1.5 in the following equation):

$$R_h(x,t,i) = k(x,t,i) \times C(x,t,i) \times \mathrm{Q10}^{(T-30)/10} \quad (2)$$

where $k(x,t,i)$ is the rate constant for decomposition of carbon in pool $C(x,t,i) \cdot k(x,t,i)$ which is a function of the lignin-to-nitrogen ratio of incoming plant litter and the local soil moisture.

CASA model development occurred during the 1990s under the auspices of a NASA Earth Observing System Interdisciplinary Science team led by P J Sellers, H A Mooney, and I Y Fung.

THE WORLD ACCORDING TO CASA

A primary achievement of the terrestrial biogeochemical modeling community during the 1990s was in exploring and defining the boundaries of perturbations to the contemporary carbon cycle arising from elevated levels of atmospheric CO_2, nitrogen deposition, variability in climate, and land use. CASA was able to contribute to advances in our understanding on many of these fronts, with its simplicity in design and dependence on satellite data sometimes serving as an asset and at other times serving as a limitation.

Applications of the CASA model included the analysis of interannual variation in NPP (Malmstrom *et al.*, 1997), spatial and temporal patterns of nitrogen trace gas emission (Potter *et al.*, 1995), the sensitivity of carbon storage to changes in plant litter quality and plant allocation (Potter and Klooster, 1997), and changes in NPP driven by changes in land use (DeFries *et al.*, 1999).

CASA has also guided the interpretation of atmospheric CO_2 data in several ways, including providing seasonally varying (but annually balanced) carbon fluxes required for assessments of the spatial distribution of land and ocean carbon sinks (Fan *et al.*, 1998). Specifically, fluxes from the terrestrial biosphere may co-vary on

seasonal timescales with vertical and horizontal changes in atmospheric circulation, leading to a mean annual build-up of atmospheric CO_2 over the surface of continents, even though the mean annual biosphere flux is zero. This effect is of critical interest, as it must be estimated and removed from the CO_2 observations prior to an inversion calculation of regional sources and sinks. CASA also has been used to estimate the $\delta^{13}C$ of respiration, and the isotopic $\delta^{13}C$ disequilibria of terrestrial vegetation (Fung *et al.*, 1997). The $\delta^{13}C$ disequilibrium is critical for another type of atmospheric inversion that simultaneously uses rates of change in atmospheric CO_2 and $\delta^{13}C$ to partition land and ocean fluxes.

Listed below are a few specific examples of insight into global ecosystem function that may or may not have been gained by the analysis of satellite data and CO_2 observations with CASA.

Northern ecosystem carbon uptake increased during spring months in recent decades. By measuring atmospheric CO_2 from Point Barrow, AK, and Mauna Loa, HI, C D Keeling at Scripps Institution of Oceanography was able show that, in recent decades, the seasonal amplitude of atmospheric CO_2 increased in the Northern Hemisphere and that the phase progressed (Keeling *et al.*, 1996). With a combination of CASA and the Goddard Institute for Space Studies (GISS) atmosphere model, it was possible to demonstrate that the observed changes in the seasonal cycle of atmospheric CO_2 were not consistent with the hypothesis that NPP increases evenly across the growing season. Rather, increased NPP and ecosystem uptake during spring and early summer provide the best fit with the atmospheric data. This is evidence for the existence of a class of mechanisms that would alter the seasonal distribution of NPP (including shifts in the growing season or a changing disturbance regime). Other mechanisms that may have less impact on the seasonal distribution of NPP, such NPP increases from nitrogen deposition or CO_2 fertilization, are less consistent with the atmospheric observations. In additional analyses involving CASA, the spatial pattern of CASA NPP was used to calculate a biosphere-weighted monthly temperature for high northern latitudes. A strong interannual correlation between the biosphere-weighted spring temperatures and ecosystem carbon uptake inferred from atmospheric CO_2 measurements may further implicate a short-term photosynthetic response to temperature.

The potential of an ecosystem to sequester atmospheric CO_2 depends equally on NPP and the distribution of carbon turnover times in plant, litter, and soil pools. Thompson *et al.* (1996) demonstrated that the potential of an ecosystem to serve as a source or sink of atmospheric CO_2 depends on the product of the initial NPP (NPP_0), the initial distribution of carbon turnover times in the ecosystem (τ_0), and the rate at which either NPP or the distribution of turnover times is perturbed (r).

$$\frac{dC}{dt} = NPP_0 \tau_0 r \tag{3}$$

Using CASA to examine spatial patterns of this sink potential has provided evidence that boreal forests and tropical forests have the highest intrinsic potential to alter atmospheric CO_2 levels. Tropical forests have high levels of NPP, and allocation to woody biomass ensures large carbon residence times even though temperatures and soil moisture are favorable to decomposition. In contrast, in boreal forests, NPP is much lower, but in Equation (2), this is compensated for by allocation to wood, chemical recalcitrance of leaf litter, and climate conditions that are unfavorable for decomposition during fall, winter, and spring.

Stimulation of NPP from elevated levels of atmospheric CO_2 may contribute only marginally to the contemporary terrestrial carbon sink. The perturbation to NPP from elevated levels of atmospheric CO_2 (r in Equation 2) is relatively well known. Specifically, the trajectory of CO_2 increase over the last several centuries from industrial activities is relatively well characterized from ice core measurements. Additional information on the response on NPP to levels of CO_2 comes from a number of free air CO_2 enrichment (FACE) studies, in which natural and managed ecosystems were exposed to elevated levels of CO_2 for a period of months to years. With r constrained from the information described above, several global assessments were made using CASA of the magnitude of the carbon sink driven by elevated levels of CO_2. The initial distribution of NPP (NPP_0) was obtained by using the contemporary spatial distribution obtained from satellite data (Equation 1), and scaling it back to a pre-industrial level using a uniformly applied relationship between CO_2 level and NPP obtained from the FACE studies. The other unknown in Equation (2), the distribution of carbon residence times (τ_0), was automatically determined by the structure and parameters of the CASA model. In these simulations, the magnitude of the terrestrial carbon sink is only one-third to one-half of contemporary estimates based on isotopes and oxygen data. This is because CASA estimates of carbon turnover are rapid as compared with other biogeochemical models. Some of the most sensitive parameters in CASA that affect the turnover time of carbon in ecosystems include allocation of NPP to wood, fine roots and leaves, and biome-prescribed turnover times of living vegetation (Figure 1). At the global scale, the spatial co-variation of NPP and carbon turnover emerges as another critical factor in regulating the response of the biosphere to changes in NPP. These model simulations highlight that our greatest limitation in predicting ecosystem carbon sequestration over the next century comes from uncertainties in our knowledge of turnover, carbon

stocks, and allocation, and not NPP or the response of NPP to external forcing from elevated CO_2 or some other factor.

A single light use efficiency (ε) captures global patterns of NPP. Variability in observed NPP across biomes (from Arctic tundra to broadleaf tropical rainforest in the Amazon Basin) can be successfully represented with a single plant light use efficiency (ε) that is not biome-dependent, but responds to local changes in soil moisture and temperature. With high resolution maps of satellite NDVI and PAR that became available in the early 1990s, it was also possible to show with CASA that a maximum (unstressed) ε of $0.4-0.5\,\mathrm{g\,C\,MJ^{-1}}$ PAR was all that was required to explain variability in the global distribution of NPP. This ε for unmanaged ecosystems is several-fold lower than the mean of observations obtained from fertilized and irrigated agricultural systems.

Tropical savanna fluxes are more seasonal than boreal forest fluxes. The largest *growing season net flux* (GSNF) occurs in tropical savannas and not in northern boreal forests. GSNF is defined as the integral of carbon uptake by an ecosystem over the growing season, and is approximately balanced by respiration releases (and fire) outside of the growing season in most areas. Seasonal APAR distributions from CASA provide strong evidence that NPP in tropical savannas is concentrated during the wet season whereas heterotrophic respiration occurs more evenly throughout the year, albeit at much lower rates when soil moisture is limiting. While boreal forests show a similar offset from heterotrophic respiration continuing under snow during fall, winter, and spring months, the total magnitude of the gross one-way flux (i.e., NPP) in boreal systems is only about one-third (on a per area basis) of what occurs in fast growing C_4 savannas across Africa, South America, and Southeast Asia. Boreal and tundra ecosystems, however, can induce large seasonal oscillations in atmospheric CO_2 because in high latitude regions atmospheric convection is small and so seasonal variations in carbon fluxes are trapped near the surface where they have a relatively large impact. In contrast, strong vertical convection in equatorial regions erases the large seasonal variation in fluxes imposed by tropical ecosystems.

Changes in tropical NPP are difficult to detect. It follows from the second point above that even very large changes in tropical NPP will leave almost no trace on the seasonal phase or amplitude of atmospheric CO_2 at pole-ward surface observation stations because of strong vertical mixing. This result, demonstrated by coupling CASA to the GISS atmospheric tracer model (Randerson *et al.*, 1997), provides an excellent example of the challenges involved with assessing changes in tropical ecosystem carbon balance using just surface CO_2 observations.

FUTURE DIRECTIONS

CASA is currently being modified to take advantage of new satellite-derived fire and land use products. The new model structure will include a distribution of plant functional types and vegetation ages within each grid cell. These modifications will provide a critical framework for assessing the impacts of human- and climate-induced changes in disturbance regime on the carbon balance of the terrestrial biosphere.

REFERENCES

DeFries, R S, Field, C B, Fung, I, Collatz, G J, and Bounoua, L (1999) Combining Satellite Data and Biogeochemical Models to Estimate Global Effects of Human-induced Land Cover Change on Carbon Emissions and Primary Productivity, *Global Biogeochem. Cycles*, **13**, 803–815.

Fan, S, Gloor, M, Mahlman, J, Pacala, S, Sarmiento, S, Takahashi, T, and Tans, P (1998) A Large Terrestrial Carbon Sink in North America Implied by Atmospheric and Oceanic Carbon Dioxide Data and Models, *Science*, **282**, 442–446.

Field, C B, Randerson, J T, and Malmström, C M (1995) Global Net Primary Production: Combining Ecology and Remote Sensing, *Remote Sens. Environ.*, **51**, 74–88.

Fung, I Y, Berry, J A, Field, C B, Thompson, M V, Randerson, J T, Malmstrom, C M, Vitousek, P M, Collatz, G J, Sellers, P, Randall, D, Badeck, F, and John, J (1997) Carbon 13 Exchanges Between the Atmosphere and Biosphere, *Global Biogeochem. Cycles*, **11**, 507–533.

Keeling, C D, Chin, J F S, and Whorf, T P (1996) Increased Activity of Northern Vegetation Inferred from Atmospheric CO_2 Measurements, *Nature*, **382**, 146–149.

Malmstrom, C M, Thompson, M V, Juday, G, Randerson, J T, and Field, C B (1997) Interannual Variation in Global Scale Net Primary Production: Testing Model Estimates, *Global Biogeochem. Cycles*, **11**, 367–392.

Potter, C S, Matson, P A, Vitousek, P M, and Davidson, E A (1995) Process Modeling of Controls on Nitrogen Trace Gas Emissions from Soils Worldwide, *J. Geophys. Res.*, **101**(D1), 1361–1377.

Potter, C S and Klooster, S A (1997) Global Model Estimates of Carbon and Nitrogen Storage in Litter and Soil Pools: Response to Change in Vegetation Quality and Biomass Allocation, *Tellus*, **49B**, 1–17.

Potter, C S, Randerson, J T, Field, C B, Matson, P A, Vitousek, P M, Mooney, H A, and Klooster, S A (1993) Terrestrial Ecosystem Production: a Process Model based on Global Satellite and Surface Data, *Global Biogeochem. Cycles*, **7**, 811–841.

Randerson, J T, Thompson, M V, Conway, T J, Fung, I Y, and Field, C B (1997) The Contribution of Terrestrial Sources and Sinks to Trends in the Seasonal Cycle of Atmospheric Carbon Dioxide, *Global Biogeochem. Cycles*, **11**, 535–560.

Thompson, M V, Randerson, J T, Malmstrom, C M, and Field, C B (1996) Change in Net Primary Production and Heterotrophic Respiration: How Much is Necessary to Sustain the Terrestrial Carbon Sink? *Global Biogeochem. Cycles*, **10**, 711–726.

CENTURY Ecosystem Model

William Parton
Colorado State University, Fort Collins, CO, USA

The CENTURY generalized ecosystem model simulates the dynamics of carbon (C), nitrogen (N), phosphorus (P), and sulfur (S) for plant-soil systems. The model is being used extensively to simulate the impact of climatic and land use changes on natural and managed ecosystems at the site, regional, and global level.

GENERAL DESCRIPTION OF THE MODEL

The model represents plant growth, nutrient cycling, and soil organic matter (SOM) dynamics for grassland, agricultural, forest, and savanna systems. The savanna system simulates the growth of trees and grasses (crop growth can also be represented) separately and includes competition for light, nutrients and water. The grass/crop and forest systems have different plant production submodels that are linked to common soil organic and nutrient cycling submodels. The model was developed with the bias that growth of cropland, grassland and forest systems can be increased by adding soil nutrients (Owensby *et al.*, 1970; Lauenroth *et al.*, 1978). The model structure reflects this bias with the soil nutrient cycling and SOM dynamics been represented in great detail, while plant growth is represented using relatively simple submodels.

The SOM and nutrient submodels represent the flow of C, N, P and S in plant litter and different organic and inorganic soil pools, with mineralization of soil nutrients primarily resulting from turnover of SOM pools. The plant production model calculates plant production and allocation of nutrients to above-ground and below-ground compartments (living) as a function of climatic factors and available soil nutrients. The model uses a monthly time step, while the major input variables include monthly precipitation, monthly average maximum and minimum air temperature, soil texture, lignin (L), N, S, and P content of plant material and soil and atmospheric N inputs. This article presents a description of the model, the method used to test and validate the model, and a summary of the application of the model for environmental impact assessment.

Figure 1 shows that the major structural components of the CENTURY model are the plant production, SOM, and the soil water and temperature submodels. The plant production submodel calculates potential plant production and nutrient demand as a function of monthly average soil temperature (STEMP) and precipitation, reduces plant production based on available soil nutrients and allocates new C, N, and P to the different live plant compartments. The monthly soil water flow model calculates water balance, soil water storage, soil water drainage and stream flow, while monthly average STEMP is calculated as a function of above ground plant biomass. Monthly precipitation, stored soil water and STEMP control the rate of decomposition of the SOM pools and release of nutrients from the SOM pools. The SOM submodel simulates the dynamics of C and soil nutrients for the different SOM pools. Decomposition of the SOM pools results in the release of soil nutrients from the SOM pools, which is then available for plant uptake. Dead plant material from the plant production submodel flows into the surface and below-ground litter pools, which are inputs into the SOM model.

Soil Organic Matter (SOM) Model

The SOM model simulates SOM dynamics for soil active, slow and passive pools, while dead litter material is represented using above-ground and below-ground structural and metabolic pools (Figure 1). The active pool (approximately 2% of the total SOM pool) includes soil microbes and microbial products with short turnover times (1–3 months). The slow SOM pool (45–60% of total SOM)

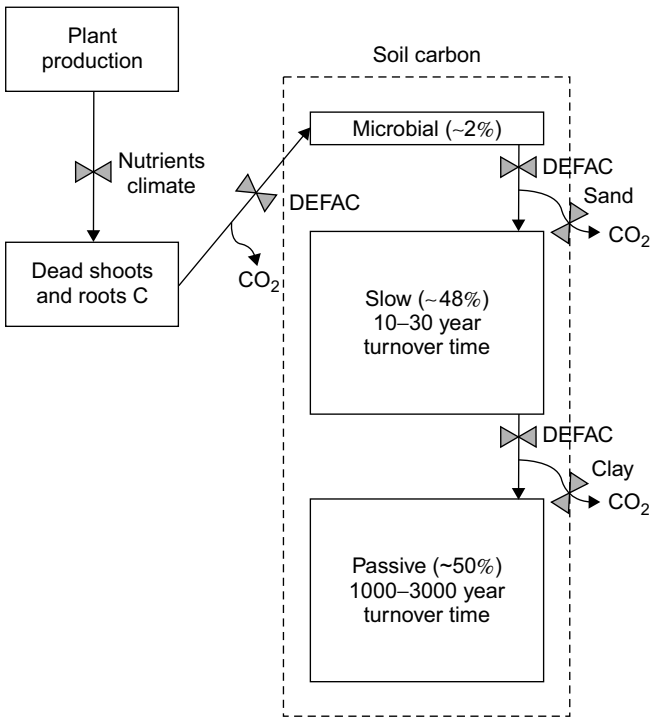


Figure 1 Major structural components of the CENTURY ecosystem model (following Parton *et al.*, 1987). DEFAC, abiotic decomposition factor

includes resistant plant material derived from structural plant material and soil stabilized soil microbial products that have turnover times ranging from 10 to 50 years depending on the climate. The passive pool (45 to 60% of total SOM) includes physically and chemically stabilized SOM that is very resistant to decomposition (turnover times from 400 to 4000 years). The structural material includes the cellulose, hemi-cellulose and lignin fraction of plant material (resistant to decomposition), while the metabolic material is readily decomposable. Plant litter material is split into structural and metabolic material as a function of the lignin to nitrogen ratio (L : N) of the litter (more structural with higher L : N ratio).

The CENTURY model assumes that decomposition of plant residues and the SOM pools are microbially-mediated with an associated microbial respiration carbon dioxide (CO_2) loss. Microbial respiration losses from decomposition of active SOM increase with the soil sand content (from 30 to 80% as sand content increases to 90%), while microbial respiration losses are approximately 50% for decomposition of all of the other litter and SOM pools. The litter and SOM pools have pool specific maximum decomposition rates with the maximum rate being reduced by an abiotic soil decomposition factor that is controlled by the soil moisture and STEMP. The STEMP function increases exponentially with increasing temperature, while the soil moisture function increases as the ratio of stored water plus current rainfall to potential evapotranspiration (PET) increases (the curve is most sensitive to changes in the ratio below 0.6). The decomposition rate of structural litter is also a function of the fraction of the structural material that is lignin (lower for higher fractions) and the lignin fraction of plant material is assumed to flow directly to slow SOM as plant structural material decomposes. The model also assumes that the fraction of the passive pool formed during decomposition of active and slow SOM increases with clay content. The net effect of the soil-texture controls on decomposition of active and slow SOM is to increase soil C stabilization for soils with low sand content and high clay content. A thorough description of the CENTURY SOM model and justification for the approach used in the model are presented by Parton *et al.* (1994).

The CENTURY model has N, P, and S pools that are analogous to all of the soil C pools. The amount of N, P, and S flowing out of a particular pool is equal to the product of the C flow out of the pool times the C-to-element ratio of the pool. A similar approach is used to calculate the flow of different elements into a specific pool but the C-to-element ratio of the receiving pool is a function of the labile inorganic mineral nutrient concentration. Low levels of available nutrients result in high C-to-element ratios for the different pools. Each pool has an allowable C-to-element ratio based on a conceptual model presented by McGill and Cole (1981). The C : N ratios of the SOM pool are narrow (5–20) compared to the C : P and C : S ratios (100 to 400). Mineralization (release of nutrients from SOM) or immobilization of N, P and S (uptake of nutrients by SOM) occurs as a result of decomposition of dead plant material and the SOM fractions. Immobilization of nutrients into SOM generally occurs during the decomposition of structural plant material (high C-to-element ratio material), while mineralization of nutrients occurs as result of decomposition of active and slow SOM (low C-to-element ratio material). The nutrient content of structural material is quite low and nutrients are immobilized into microbial biomass during decomposition of structural material, while slow and active SOM have high nutrient contents and release nutrients (mineralize) while they are being decomposed. A complete description of the soil nutrient model is presented by Parton *et al.* (1988).

Soil Water and Temperature Model

The CENTURY model uses a simplified water budget model to calculate monthly bare-soil evaporation, interception and transpiration water loss, stored soil water, snow water content, stream flow and saturated water flow between soil layers. Interception and bare-soil water loss are calculated as fractions of the monthly precipitation and are subtracted from monthly precipitation before the water is added to the soil. Because each CENTURY simulation is completed for a single cell assumed to have flat topography, the water budget does not take into account precipitation redistribution on uneven terrain by overland flow. Bare-soil water loss is a function of above-ground biomass (decreasing with increasing biomass), while interception water loss increases with increasing above-ground biomass. Transpiration water increases as a function of live leaf biomass. Water loss occurs first as interception followed by bare-soil water loss and transpiration with the sum not exceeding the PET water loss (PET is calculated as a function of maximum and minimum air temperature). Precipitation in excess of PET is stored in soil water layers by adding the water to the top layer. If average air temperature is less than 0 °C then monthly precipitation occurs as snow. Sublimation from snow pack occurs at the PET rate (corrected for the heat of fusion) while snow melt is a linear function of the average air temperature for temperatures > -6 °C.

Near-surface average STEMP is used to calculate the abiotic decomposition rate and the temperature effect on plant growth. STEMP is calculated using equations where the maximum STEMP is a function of maximum air temperature and the canopy biomass (lower for high biomass) and the minimum STEMP is a function of minimum air temperature and canopy biomass (higher for high biomass). It

is important to note that both the soil water balance and STEMP models include the effect of simulated live and dead above-ground plant biomass on STEMP and soil water balance.

Plant Production and Management Model

The CENTURY model is set up to simulate the dynamics of forests, grasslands, agricultural crops and savanna systems. The grassland/crop submodel simulates growth of different crops (corn, wheat, potatoes, sugarcane, etc.) natural plant communities (temperate warm and cool season grasslands, tropical grasslands, etc.), and managed grassland systems (alfalfa, clover, and improved grasslands). The forest submodel simulates the growth of evergreen (pine and fir systems and evergreen tropical systems), temperate deciduous, and drought deciduous systems. The savanna system simulates a tree-grass system by simultaneously running the tree and grassland/crop submodels with the submodels interacting through shading effects and N competition. Both submodels assume that monthly maximum production is controlled by soil moisture and temperature with maximum rates decreased if the soil nutrients supply is insufficient (the most limiting nutrient controls production). The grassland/crop model also includes the effect of shading by standing dead vegetation, while the forest model includes the effect of live leaf area on plant production. Potential production is a function of the maximum growth rate for each grassland/crop or forest system and reduced by 0–1 scalars depending on the factors that limit production. Plant nutrient uptake is a function of living root biomass with uptake increasing as living root biomass increases up to 0.3 $g\,m^{-2}$. As mentioned earlier, in most forest and grassland/crop systems, plant production will increase with the addition of nutrients.

The forest and grassland/crop models are generic plant growth models that can be parameterized to represent a large variety of crop, grassland, and forest systems by altering crop and forest specific parameters. The grassland/crop model includes living shoots and roots, and standing dead plant parts, while the forest system includes living shoots and fine roots, large wood, fine branches and coarse roots. The effect of grazing and fire on the grassland/crop system is represented in the model with the major effect of fire being the increase in root-to-shoot ratio, increase in the C : N ratio of roots and shoots, removal of vegetation and return of nutrients from the fire. Grazing removes live and dead vegetation, alters the root-to-shoot ratio, increases the N content of live shoots and roots and returns nutrient to the soil. All of the natural fire effects and standard forest management practices can be represented by the model (selective logging, clear cutting, etc.). A complete description of the parameterization of the model for different plant systems and the use of the different management options is presented in the CENTURY user manual (Metherell *et al.*, 1993).

Use and Testing of the CENTURY model

Extensive data sets from long-term agricultural experiments and grasslands have been used to test the CENTURY model. We used the observed data to test the model and as a tool for integrating and interpreting the data sets. Plant production and soils data from extensive tropical and temperate grassland around the world (Parton *et al.*, 1993) show that the model correctly simulates the effects of burning, irrigation, fertilization, and grazing on plant production and the seasonal patterns for live and dead biomass. The model has been used to simulate the long-term (30–60 years) dynamics of SOM and plant production for corn, and winter and spring wheat systems. The model was used to correctly simulate the impact of adding different amounts and types of organic matter (pea vine, sawdust, straw, green manure and manure), straw burning, the use of different fertilizer levels, different tillage practices (stubble mulch, conventional plow, and no-till) and wheat pasture rotations on STEMP, soil water dynamics, soil C and N levels, plant production, soil NO_3 leaching and soil N mineralization (Paustian *et al.*, 1992, Probert *et al.*, 1995). The forest model has been evaluated for tropical, temperate and boreal systems and used to simulate the response of forests to different natural disturbances and management practices (Peng *et al.*, 1998).

The CENTURY model has been used extensively to simulate the effect of environmental changes and management practices on natural and managed ecosystems at the site, regional and global level. The grassland model has been used to simulate the impact of climate change and increased atmospheric CO_2 levels on grasslands around the world (e.g., Parton *et al.*, 1995) with a detailed analysis for the US Great Plains region (Schimel *et al.*, 1990). The combined effect of future environmental change and improved land use practices on soil C storage and plant production has been evaluated for the US Corn Belt (Donigian *et al.*, 1995), while Paustian *et al.* (1996) have used CENTURY to evaluate the soil C storage in the US resulting from the conservation reserve program. The Vegetation/Ecosystem Modeling and Assessment program (VEMAP) (VEMAP, 1995) has used the CENTURY model to evaluate the impact of climate change and increased CO_2 levels on the natural ecosystems in the US using a $0.5 \times 0.5^{\circ}$ resolution and has compared model results with two other biogeochemistry models. The model has also been used to simulate ecosystem dynamics at the $0.5 \times 0.5^{\circ}$ scale for global ecosystems (Schimel *et al.*, 1997). We are currently developing a daily version of the model (Kelly *et al.*, 2000) which simulates all of the ecosystem dynamics using more mechanistic soil water and temperature submodels and

also simulates daily trace gas fluxes (N_2O, N_2, NO_x and CH_4).

REFERENCES

Donigian, Jr, A S, Patwardham, A S, Jackson, R B, IV, Barnwell, T O, Weinrich, K B, and Rowell, A L (1995) Modelling the Impacts of Agricultural Management Practices on Soil Carbon in the Central US, in *Soil Management and Greenhouse Effect. Advances in Soil Science*, eds R Lal, J Kimble, E Levine, and B A Stewart, CRC Press, Boca Raton, FL, 121–145.

Kelly, R H, Parton, W J, Hartman, M D, Stretch, L K, Ojima, D S, and Schimel, D S (2000) Intra- and interannual variability of ecosystem processes in shortgrass-steppe: new model, verification, simulations, *J. Geophys. Res.*, **105**(D15), 20,093–20,1000.

Lauenroth, W K, Dodd, J L, and Sims, P L (1978) The Effects of Water and Nitrogen Induced Stresses on Plant Community Structure in a Semi-arid Grassland, *Oecologia*, **36**, 211–222.

McGill, W B and Cole, C V (1981) Comparative Aspects of Cycling of Organic C, N, S and P through Soil Organic Matter, *Geoderma*, **26**, 267–286.

Metherell, A K, Harding, L A, Cole, C V, and Parton, W J (1993) CENTURY Soil Organic Matter Model Environment, Technical Documentation, Agroecosystem version 4.0, *Great Plains System Research Unit Technical Report No. 4*, USDA-ARS, Fort Collins, CO.

Owensby, C E, Hyde, R M, and Anderson, K L (1970) Effects of Clipping and Supplemental Nitrogen and Water on Loamy Upland Range, *J. Range Manage.*, **23**, 341–346.

Parton, W J, Schimel, D S, Cole, C V, and Ojima, D S (1987) Analysis of Factors Controlling Soil Organic Matter Levels in Great Plains Grasslands, *Soil Sci. Soc. Am. J.*, **51**, 1173–1179.

Parton, W J, Stewart, J W B, and Cole, C V (1988) Dynamics of C, N, P and S in Grassland Soils: a Model, *Biogeochemistry*, **5**, 109–131.

Parton, W J, Scurlock, J M O, Ojima, D S, Gilmanov, T G, Scholes, R J, Schimel, D S, Kirchner, T, Menaut, J-C, Seastedt, T, Garcia Moya, E, Kamnalrut, A, and Kinyamario, J I (1993) Observations and Modeling of Biomass and Soil Organic Matter Dynamics for the Grassland Biome Worldwide, *Global Biogeochem. Cycles*, **7**, 785–809.

Parton, W J, Schimel, D S, Ojima, D S, and Cole, C V (1994) A General Model for Soil Organic Matter Dynamics: Sensitivity to Litter Chemistry, Texture and Management, in *Quantitative Modeling of Soil Forming Processes*, eds R B Bryant and R W Arnold, SSSA Special Publication 39, ASA, CSSA and SSA, Madison, WI, 137–167.

Parton, W J, Scurlock, J M O, Ojima, D S, Hall, D O, and SCOPE GRAM Group Members (1995) Impact of Climate Change on Grassland Production and Soil C Worldwide, *Global Change Biol.*, **1**, 13–22.

Paustian, K, Elliot, E T, Peterson, G A, and Killian, K (1996) Modeling Climate, Carbon Dioxide and Management Impacts on Soil Carbon in Semi-arid Agroecosystems, *Plant Soil*, **187**, 351–365.

Paustian, K, Parton, W J, and Persson, J (1992) Modeling Soil Organic Matter in Organic-amended and Nitrogen-Fertilized Long-term Plots, *Soil Sci. Soc. Am. J.*, **56**, 476–488.

Peng, C, Apps, M J, Price, D T, Nalder, I A, and Halliwell, D H (1998) Simulating Carbon Dynamics Along the Boreal Forest Transect Case Study (BFTCS) in Central Canada, *Global Biogeochem. Cycles*, **12**, 381–392.

Probert, M E, Keating, B A, Thompson, J P, and Parton, W J (1995) Modeling Water, Nitrogen, and Crop Yield for a Long-term Fallow Management Experiment, *Aus. J. Exp Agric.*, **35**, 941–950.

Schimel, D S, Emanuel, W, Rizzo, B, Smith, T, Woodward, F I, Fisher, H, Kittel, T G F, McKeown, R, Painter, T, Rosenbloom, N, Ojima, D S, Parton, W J, Kicklighter, D W, McGuire, A D, Melillo, J M, Pan, Y, Haxeltine, A, Prentice, C, Sitch, S, Hibbard, K, Nemani, R, Pierce, L, Running, S, Borchers, J, Chaney, J, Neilson, R, and Braswell, B H (1997) Continental scale variability in ecosystem processes: models, data, and the role of disturbance, *Ecol. Monogr.*, **67**(2), 251–271.

VEMAP (1995) Vegetation/Ecosystem Modeling and Analysis Project: Comparing Biogeography and Biogeochemistry Models in a Continental-Scale Study of Terrestrial Ecosystem Responses to Climate Change and CO_2 Doubling, *Global Biogeochem. Cycles*, **9**, 407–437.

FURTHER READING

Sanford, Jr, R L, Parton, W J, Ojima, D S, and Lodge, D J (1991) Hurricane Effects on Soil Organic Matter Dynamics and Forest Production in the Luquillo Experimental Forest, Puerto Rico: Results of Simulation Modeling, *Biotropica*, **23**, 364–372.

Chaos and Cycles

Bruce E Kendall

University of California, Santa Barbara, CA, USA

Regular fluctuations in population density, both periodic and chaotic, are found in a variety of species in nature, including boreal mammals, insect herbivores, grouse and partridge, and salmon. Although the exact causes of these cycles remain undetermined for most species, they generally involve strongly nonlinear trophic and competitive interactions among a small number of species. Because of these nonlinearities, it is difficult to extrapolate the effects of global environmental change on the dynamics of any particular population. However, if communities become less diverse and conditions change so as to increase many

species' intrinsic growth rates, then cyclic and chaotic population dynamics may become more common.

INTRODUCTION

The study of complex population dynamics is nearly as old as population ecology. In the 1920s, Alfred Lotka and Vito Volterra independently developed a simple model of interacting species that still bears their joint names. The predator–prey version of this linear model displayed neutrally stable cycles. Experimentalists such as G F Gause and A J Nicholson produced cyclic dynamics in laboratory populations of protozoa and insects (Figure 1). In the meantime, Charles Elton had become fascinated by cyclic fluctuations in field populations of voles (small rodents; Figure 2) and in the numbers of furs of Canadian lynx and other large mammals traded by the Hudson's Bay Company. Foresters have long been dismayed by the periodic outbreaks of certain leaf-feeding insects (Figure 2), which can completely defoliate a forest when at peak density. However, it was not until Michael Rosenzweig, as a graduate student with Robert MacArthur in the 1960s, added density dependence and predator behavior to the Lotka–Volterra equations that an ecological model capable of displaying true nonlinear limit cycles was developed.

In the early and mid 1970s, physicist Robert May took theoretical ecology by storm with a series of papers demonstrating that complex dynamics (cycles and chaos) could be generated by the simplest ecological models (e.g., May, 1976), including the logistic map, an extremely simple model of density-dependent population growth with discrete generations:

$$N(t+1) = rN(t)\left[\frac{1-N(t)}{K}\right] \qquad (1)$$

The dynamics of this model are described in **Non-linear Systems**, Volume 2. Theoreticians rapidly embraced the new science of nonlinear dynamics, hoping that chaos could explain some of the erratic fluctuations that were hitherto called noise. Empirical ecologists were rather more skeptical, and never embraced the claims that chaos was important in the real world; but they gradually accepted the idea that nonlinear dynamics in general were important. This acceptance was indicated by a 1990 symposium sponsored by the Ecological Society of America entitled "The Shift from an Equilibrium to a Non-Equilibrium Paradigm in Ecology".

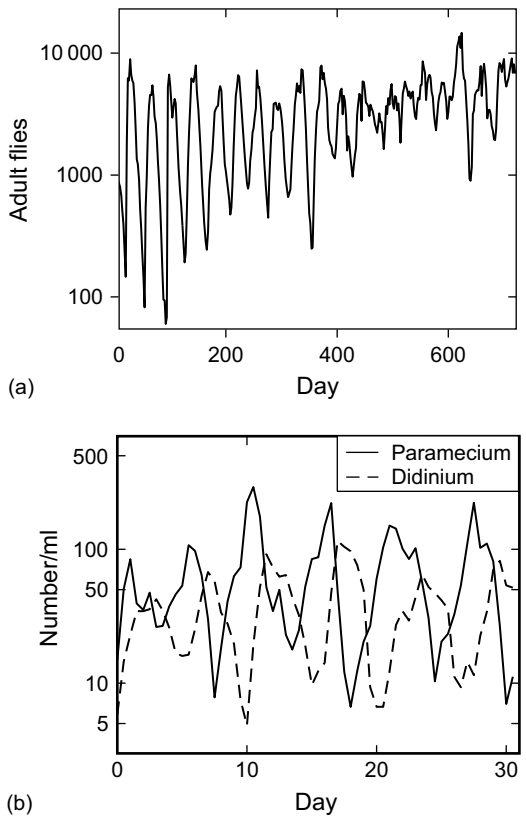


Figure 1 Examples of nonlinear dynamics in laboratory populations. (a) Blow ies (*Lucilia cuprina*). The insects were provided a limited supply of protein, which they require to produce eggs; all other food was *ad lib*, so the cycles were caused by density-dependent fecundity. By the end of the experiment the ies had evolved an increased ability to reproduce at high density, at the expense of their formerly high fecundity at low densities; there were about two generations per cycle. Although the dynamics shift abruptly, the fecundity parameters appear to have been changing continuously, suggesting that the ies evolved through a bifurcation. (b) The protozoa *Didinium nasutum* (predator) and *Paramecium aurelia* (prey). These are pure predator- prey cycles, as the food supply for *Paramecium*, along with all other environmental conditions, was constant. (Redrawn from Nicholson, 1957 and Luckinbill, 1973)

MATHEMATICAL BACKGROUND

Chaos was first knowingly observed in a mathematical model by meteorologist Edward Lorenz in the early 1960s. Not only did he observe deterministic aperiodic fluctuations, but he also discovered sensitive dependence on initial conditions: restarting a simulation model without bothering to copy all of the significant digits, Lorenz observed a sequence of numbers that soon started to diverge from the original simulation. Within the context of turbulence studies, the mathematical and computational strands of chaos were woven together in the 1970s and 1980s. In 1971 mathematicians David Ruelle and Floris Takens coined an evocative term *strange attractor* to describe the complex structures of chaos.

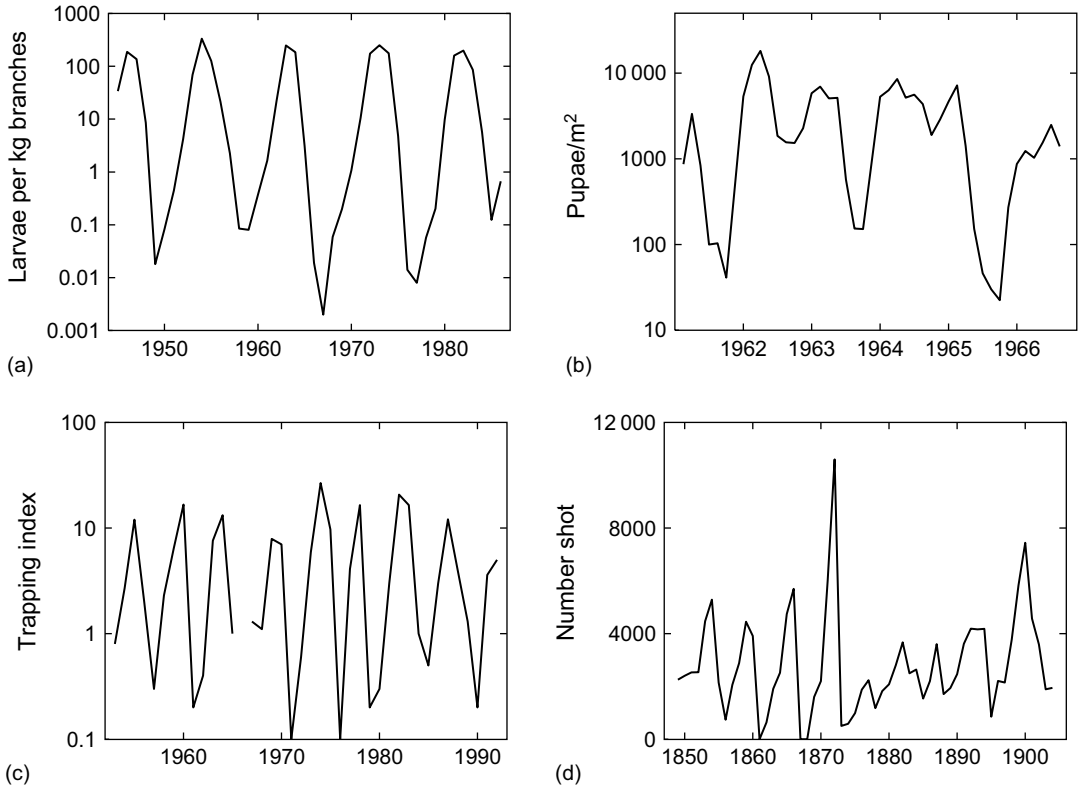


Figure 2 Examples of nonlinear dynamics in eld populations. (a) Larch budmoth (*Zeiraphera diniana*) in Switzerland. These cycles are probably caused by interactions with parasitoids, although there may also be feedback through changes in nutritional quality of the larch needles. Tree-ring data indicate that this extraordinarily regular cycle has persisted for at least 150 years. (b) Coffee leaf-miners (*Leucoptera* spp.) in Tanzania. This also appears to be a host- parasitoid system, coupled with seasonal variation in growth rates (there are eight generations per year). (c) Voles (*Microtus* and *Clethrionomys*) in Finland. This is probably driven by predation by least weasels and stoats, which specialize on small rodents. (d) Red Grouse (*Lagopus lagopus scotius*) in Scotland. Both intraspeci c competition and interactions with an internal parasite have been proposed as mechanisms for these cycles. (Redrawn from Baltensweiler and Fischlin, 1988; Bigger, 1973; Hanski *et al.*, 1993; Middleton, 1934)

The mathematics of deterministic dynamical systems is well understood (*see* **Non-linear Systems**, Volume 2). Four qualitative types of dynamic attractors are observed: *equilibrium*, *periodic* cycles (also called limit cycles), *quasiperiodic* cycles, and *chaos* (Figure 3). Chaos is characterized by a positive *Lyapunov exponent*, which can be thought of as a measure of the long-term unpredictability of the system (see Box 1); equilibrium and periodic attractors have a negative exponent, whereas a quasiperiodic attractor has an exponent of zero. As the system's parameters are changed, the dynamics may undergo an abrupt, qualitative change, known as a *bifurcation* (*see* **Non-linear Systems**, Volume 2).

Environmental Variability and Nonlinear Dynamics

Dynamical systems theory was developed in the context of controlled physical experiments. In ecology, the assumption of a constant, noise-free environment is untenable, and we have to ask whether any of the results for deterministic systems carry over. There are two important ways that these assumptions could be violated: periodic variation in parameter values (e.g., seasonality) or noise (random perturbations to the parameters or state variables due to environmental variability).

Seasonal variation can induce complex dynamics in systems that would be stable in a constant environment. For example, one of the standard epidemiological models of childhood infectious disease has a stable equilibrium in a constant environment. Allowing the contact rate parameter to fluctuate through the year (to simulate the effect of school terms) induces, not surprisingly, a limit cycle with a one-year period. However, further increases in the magnitude of seasonality lead to a periodic or chaotic modulation of the basic annual cycle.

A small amount of noise merely blurs the attractor. The fractal structure of a chaotic attractor is destroyed, but the basic dynamical characteristics of the attractors, such as

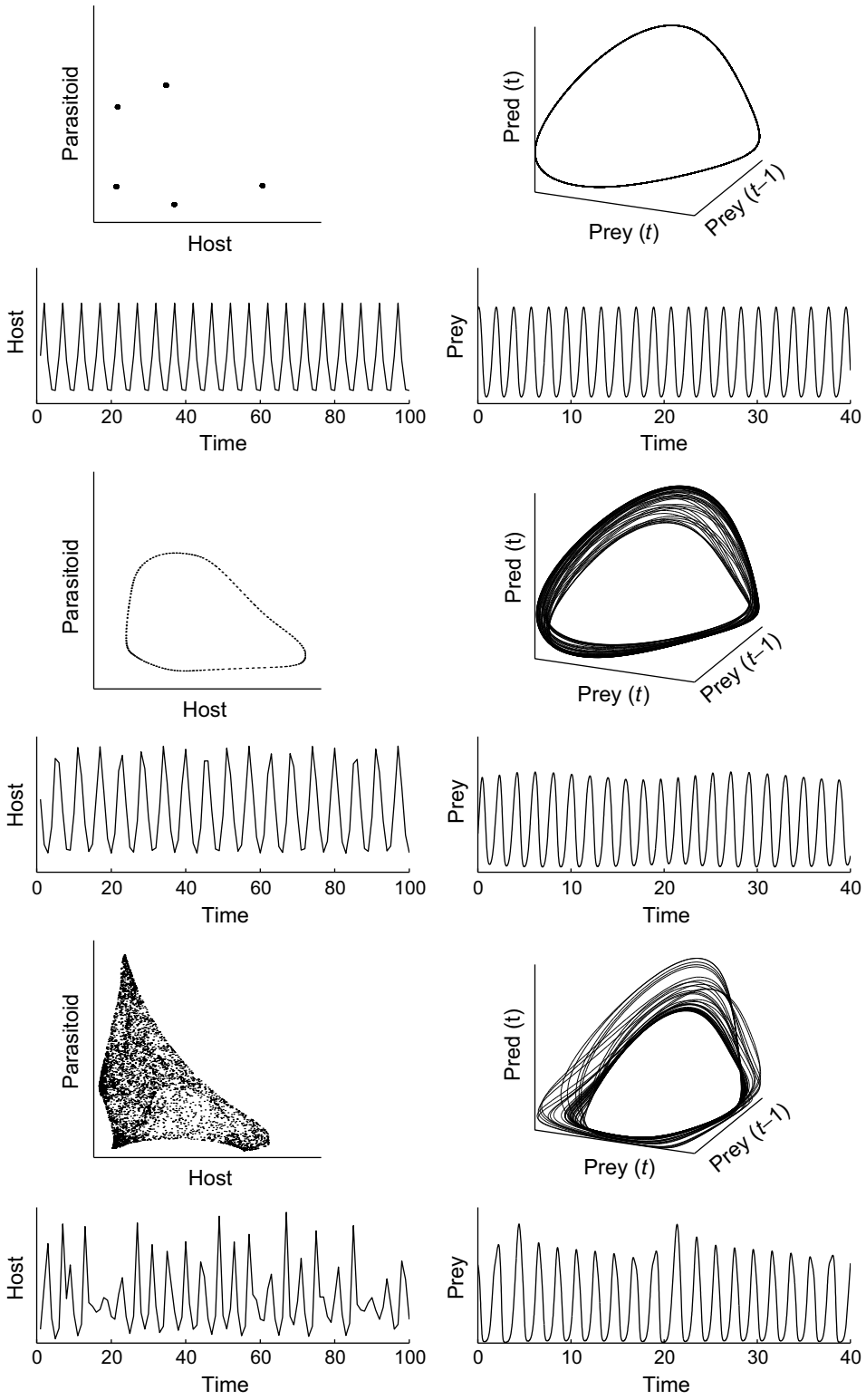


Figure 3 Time series and phase space representations of a discrete-time host- parasitoid model (left panes) and a continuous-time predator- prey model with seasonal variation in the prey carrying capacity (right panes). Top rows: limit cycles. Middle rows: quasiperiodic cycles. Bottom rows: chaos. For the host- parasitoid model the parasitoid attack rate increases from top to bottom. For the predator- prey model the strength of seasonality increases from top to bottom (there is no seasonality in the top pane); notice that the period of the predominant oscillation remains longer than one year. (Models originally published in Beddington, 1976; Rinaldi *et al.*, 1993)

Box 1 Lyapunov exponents

Lyapunov exponents are properties of attractors, and quantify an important aspect of the dynamics associated with the attractors described above: if a system is perturbed slightly, will it return to the trajectory it would have followed in the absence of the perturbation, or will the effect of the perturbation grow through time? If the largest Lyapunov exponent is negative, then the effects of the perturbation will die out; this is associated with stable equilibria and limit cycles. If the largest Lyapunov exponent is positive, then perturbations will tend to grow; this is a property of chaotic attractors. Quasiperiodic attractors have a largest Lyapunov exponent of exactly zero: perturbations neither grow nor decay.

the Lyapunov exponent, are unchanged. In contrast, a large noise variance can have surprising effects on the dynamics. If there are multiple coexisting attractors in the system, then the noise can bounce the trajectory among them, leading to episodes of qualitatively different dynamics. In addition, many systems contain complex transient dynamics as the trajectory returns to the attractor from a distant part of the state space, even when the attractor is a limit cycle or equilibrium. A trajectory following such transients can have all the characteristics of chaos – indeed, the transients are often the ghosts of chaotic attractors from other parameter values. If large perturbations are frequent enough, then the trajectory will spend more time on the transients than on the deterministic attractor.

CHAOS AND CYCLES IN ECOLOGICAL DYNAMICS

Empirical investigations of nonlinear population dynamics have generally proceeded along three directions: statistical analysis of time series data, experimental manipulations of whole populations, and mechanistic models built on known processes.

Time Series Analysis

Spectral analysis is a statistical technique that identifies periodicity by decomposing the time series into sinusoidal components, and asking whether the most important component is larger than would be expected by chance if the fluctuations were random. Even chaotic time series have a strong spectral peak. A recent analysis of 700 long population time series revealed that fully 30% of them displayed statistically significant periodicity (Kendall *et al.*, 1998).

Early approaches for estimating the Lyapunov exponent and the attractor dimension were developed by physicists for use on long (hundreds of points), noise-free time series. Olsen and Schaffer (1990) applied these techniques to data on childhood diseases, and concluded that the chickenpox data displayed noisy periodicity whereas measles epidemics were chaotic. However, these techniques are unsuited to the short, noisy time series that characterize most ecological systems.

Modern techniques for estimating the Lyapunov exponent fit a statistical time series model to the data and then analyze the model directly. A difficult problem that remains unsolved is how to select the best model that fully captures the underlying dynamics without fitting the noise. Ellner and Turchin (1995), using a selection scheme that was biased towards simpler dynamics, analyzed a variety of laboratory and field population time series, and found evidence for chaos in a few, but not many, of the populations. Thus, the answer to the question: "Is Mother Nature a strange attractor?" (Hastings *et al.*, 1993) seems to be occasionally, but not often.

Experiments

There have been numerous laboratory populations raised with the intent of reproducing the predator–prey cycles predicted by the Lotka–Volterra model. Most of these experiments were performed with interacting species (predators and prey) of protozoa or mites (Figure 1b). Along with blowfly populations raised by A J Nicholson (Figure 1a), they demonstrated that population fluctuations could occur even in a constant environment.

A powerful tool for uncovering mechanisms is to modify demographic parameters or interaction rates and observe the resulting changes in dynamics. For example, Constantino *et al.* (1995) induced transitions between equilibrium and cyclic dynamics in laboratory populations of the flour beetle (*Tribolium castaneum*) by artificially increasing the adult mortality rate. In contrast, Nicholson stabilized the blowfly cycles by increasing the juvenile death rate. In the field, Krebs *et al.* (1995) added food and reduced predation on snowshoe hares in 1 km^2 enclosures during the course of a cycle; each manipulation increased the peak density (and delayed the peak). Density manipulations have been used to prolong the cycle peak of Red Grouse (Moss *et al.*, 1996). The results of these manipulations are often initially counter-intuitive, as most of us have intuition based on linear situations. Thus, these results can be best understood by making analogous manipulations to a nonlinear model.

Mechanistic Models

Time series analysis does not provide any information about the processes that govern nonlinear dynamics. Using verbal models to generate predictions leaves the results of experimental manipulations less than sharp. Both of these empirical approaches greatly benefit from the thoughtful use of mechanistic models. For example, the conclusions about smallpox and measles were strengthened by the analysis

of a nonlinear epidemiological model that supported the time-series analysis: with parameters appropriate to smallpox, the model produces stable limit cycles, whereas with measles parameters, it produces chaos (Olsen and Schaffer, 1990). Kendall *et al.* (1999) demonstrated techniques whereby alternative mechanistic models could be fit to a time series, and then time-series analysis could be used to discriminate which model better reproduced the patterns in the data. The *Tribolium* study used a mechanistic model to make predictions (which were sometimes confirmed) about the qualitative changes in dynamics as the mortality rate was varied (Constantino *et al.*, 1995).

The ideal study to uncover the underlying causes of nonlinear dynamics in a population would start with a set of mechanisms that are consistent with the existing data, in the sense of reproducing the dynamic patterns when entered into models. It would then use these models to find the critical experiments that would conclusively distinguish among the competing mechanisms. Finally, the outcomes of these experiments would be analyzed rigorously using time-series statistics. Such an ideal study has never been performed, so we still lack rigorously defensible explanations for the causes of cycles in any field population.

GLOBAL CHANGE AND POPULATION DYNAMICS

Only one experiment has been performed to examine the effects of global environmental change on population fluctuations, but it is possible to make some speculations based on both model behavior and the biology of cyclic populations. A feature common to most population models is that an increase in the intrinsic growth rate of the population makes cycles and chaos more likely. Thus, changing environmental conditions may induce cycles in some populations that do not currently exhibit these dynamics. For example, increased temperature may reduce the generation time in multivoltine insects (such as aphids), allowing their population growth to outstrip the factors that currently regulate their density. Bazzaz *et al.* (1992) demonstrated experimentally that raising the carbon dioxide (CO_2) concentration to 700 $\mu l\, l^{-1}$ increased the fecundity of the annual plant *Abutilon theophrasti* to the point that the plant would exhibit chaotic fluctuations in abundance due to overcompensating density dependence. These oscillations might not be observed in the field (*Abutilon* is an early successional species and persists only for a few years at a given site), but this experiment suggests that cycles in plants (which are presently rare) may become more common with rising CO_2.

Most of the cyclic herbivorous insect species are associated with plant stands that are either naturally or artificially low in diversity (e.g., pine plantations, alpine larch forests). The herbivores are often specialists, and have specialist parasitoids. As plantations replace diverse tropical forests, the potential for such cyclic insect outbreaks increases. There are very few long-term datasets on tropical insect populations, but a study of leaf-miners infesting coffee plantations in Tanzania shows a prominent two-year cycle (Figure 2b).

In temperate forests, many of the cyclic insect herbivores rely on young foliage to meet their nutritional needs for rapid growth. Thus, a key feature of the population dynamics depends on the synchrony between bud burst and hatching of the larvae. If these two processes rely on different environmental cues (as seems likely, for the synchrony often breaks down in years with unusual weather), then changes in temperature and rainfall regimes may break the synchrony and reduce the prevalence of such cycles, at least until the insects can evolve new emergence cues.

One of the hypotheses for the pronounced cycles in boreal mammals is that the simpler communities, and especially the absence of the generalist predators that live farther south, allow classic predator–prey cycles to arise. If this is true, and if there is warming of these environments, then invasion by generalist predators (and their alternative prey) from farther south may lead to these cycles disappearing.

The bifurcation diagrams suggest that we might see abrupt changes in population dynamics as the environment changes. However, the noise that characterizes ecological systems will smooth these transitions. Indeed, these shifts may be accompanied by transient dynamics that mimic the old attractor, so that it may be years or decades before the new attractor is revealed.

See also: **Chaos and Predictability**, Volume 1; **Nonlinear Systems**, Volume 2; **Monitoring in Support of Policy: an Adaptive Ecosystem Approach**, Volume 4.

REFERENCES

Baltensweiler, W and Fischlin, A (1988) The Larch Budmoth in the Alps, in *Dynamics of Forest Insect Populations*, ed A A Berryman, Plenum Press, New York, 331–351.

Bazzaz, F A, Ackerly, D D, Woodward, F I, and Rochefort, L (1992) CO_2 Enrichment and Dependence of Reproduction of Density in an Annual Plant and a Simulation of its Population Dynamics, *J. Ecol.*, **80**, 643–651.

Beddington, J R, Free, C A, and Lawton, J H (1976) Concepts of Stability and Resilience in Predator–Prey Models, *Nature*, **225**, 58–60.

Bigger, M (1973) An Investigation by Fourier Analysis into the Interaction between Coffee Leaf-miners and their Larval Parasites, *J. Anim. Ecol.*, **42**, 417–434.

Constantino, R, F, Cushing, J M, Dennis, B, and Desharnais, R A (1995) Experimentally Induced Transitions in the Dynamic Behaviour of Insect Populations, *Nature*, **375**, 227–230.

Ellner, S and Turchin, P (1995) Chaos in a Noisy World: New Methods and Evidence from Time-Series Analysis, *Am. Nat.*, **145**, 343–375.

Hanski, I, Turchin, P, Korpimäki, E, and Henttonen, H (1993) Population Oscillations by Rodents: Regulation by Mustelid Predators Leads to Chaos, *Nature*, **364**, 232–235.

Hastings, A, Hom, C L, Ellner, S, Turchin, P, and Godfrey, H C J (1993) Chaos in Ecology: is Mother Nature a Strange Attractor? *Annu. Rev. Ecol. Syst.*, **24**, 1–33.

Kendall, B E, Briggs, C J, Murdoch, W W, Turchin, P, Ellner, S P, McCauley, E, Nisbet, R M, and Wood, S N (1999) Why do Populations Cycle? A Synthesis of Statistical and Mechanistic Modeling Approaches, *Ecology*, **80**, 1789–1805.

Kendall, B E, Prendergast, J, and Bjørnstad, O N (1998) The Macroecology of Population Dynamics: Taxonomic and Biogeographic Patterns in Population Cycles, *Ecol. Lett.*, **1**, 160–164.

Krebs, C J, Boutin, S, Boonstra, R, Sinclair, A R E, Smith, J N M, Dale, M R T, Martin, K, and Turkington, R (1995) Impact of Food and Predation on the Snowshoe Hare Cycle, *Science*, **269**, 1112–1115.

Luckinbill, L S (1973) Coexistence in Laboratory Populations of Paramecium Aurelia and its Predator Didinium Nasutum, *Ecology*, **54**, 1320–1327.

May, R M (1976) Simple Mathematical Models with Very Complicated Dynamics, *Nature*, **261**, 459–467.

Middleton, A D (1934) Periodic Fluctuations in British Game Populations, *J. Anim. Ecol.*, **3**, 231–249.

Moss, R, Watson, A, and Parr, R (1996) Experimental Prevention of a Population Cycle in Red Grouse, *Ecology*, **77**, 1512–1530.

Nicholson, A J (1957) Cold Spring Harbor Symposia on Quantitative Biology, No 22, 153–173.

Olsen, L F and Schaffer, W M (1990) Chaos Versus Noisy Periodicity: Alternative Hypotheses for Childhood Epidemics, *Science*, **249**, 499–504.

Rinaldi, S, Muratori, S, and Kuznetsov, Y (1993) Multiple Attractors, Catastrophes, and Chaos in Seasonally Perturbed Predator–Prey Communities, *Bull. Math. Biol.*, **55**, 15–35.

and out-compete potentially invading species in the climax environment. Over long time scales, however, environmental shifts due to climatic change or ecological and evolutionary processes produce changes in the species composition of the climax stage.

In environments with relatively low frequencies of major disturbance, succession has enough time between disturbances to run its course and reach the climax stage. This usually produces a stable assemblage of species dominated by a few prominent ones. Where major disturbances are frequent, succession may never reach the climax stage, resulting in diverse mixes of species with no consistently dominant species.

Global environmental change is likely to reduce the occurrence of climax vegetation in two ways. First, in a lot of areas, disturbances are likely to occur more frequently so succession may no longer have enough time between disturbances to reach its climax. Consequently, climax vegetation and climax species assemblages are expected to become less common. Secondly, the unidirectional nature of much of global environmental change is likely to alter the species composition of the climax vegetation. Where the environment changes relatively quickly compared to the time it takes the species composition of the climax stage to evolve to new environments, the species composition of this climax vegetation may itself be constantly changing. In this sense, true climax vegetation may be unrealizable in these places until the environment there has stabilized (*see also* **Disturbance**, Volume 2; **Succession**, Volume 2).

FURTHER READING

Odum, E P (1969) The Strategy of Ecosystem Development, *Science*, **164**, 262–270.

RICHARD A FLEMING *Canada*

Cheatgrass - Alien Invader

see Plant Dispersal and Migration (*Opening essay*, **Volume 2**)

Climax Vegetation

Climax vegetation is the vegetation present at the ultimate stage (climax) of the ecological succession of a species assemblage under the predominant physical environmental conditions occurring on a given site. Left undisturbed, the species composition of the climax remains relatively stable because these species successfully reproduce themselves

CO_2 Enrichment: Effects on Ecosystems

Ch Körner
University of Basel, Basel, Switzerland

We live in a carbon world, which means that almost half of all dried organismic tissue consists of the element carbon, C. If organismic tissue or debris is oxidized (decomposed), carbon is emitted as carbon dioxide, CO_2. Although CO_2 is called a trace gas, because only ca. 360–370 ppm of

the atmosphere is currently represented by CO_2, it is the major substrate which green plants absorb by photosynthesis. Photosynthesis, in essence, reduces CO_2 to sugar, utilizing solar energy (release of oxygen, O_2). Non-green living parts of plants, green parts at night, all animals and nearly all microbes obtain their energy requirements from re-oxidizing carbon compounds (respiration), and thus consume oxygen and emit CO_2. Photosynthetic binding of CO_2 and respiratory release of CO_2 are the two major components of the biological carbon cycle of the globe, which turns over approximately 100 Gt (i.e., billions of tons) of carbon year^{-1}.

Whenever the magnitude of the two processes is unequal, carbon either accumulates in the atmosphere or in the biosphere (including the pedosphere, i.e., the soil humus, and organic ocean sediments). Net biospheric carbon accumulation during the last 0.5 billion years has created enormous fossil carbon reservoirs below the ground, a large part of which mankind is releasing within ca. 300 years, starting around the year 1800, with dramatic increases of emission in recent decades, further enhanced by deforestation (the release of current biosphere carbon). This is why the atmospheric CO_2 concentration is increasing so rapidly and thereby is changing the biosphere's diet at an unprecedented speed. Atmospheric CO_2 is plant food, but at the same time, it absorbs long wave (infra-red) radiation and, together with other gases, contributes to an enhancement of the global greenhouse effect. The latter phenomenon may create a warmer climate, for which the greenhouse gas CO_2 is best known, but this synthesis aims at explaining the direct biological consequences of CO_2 enrichment, irrespective of any indirect influences on organisms via climatic change. Elevated CO_2 contributes to the biodiversity crises in a slow, invisible manner. It is one of the prime needs of global change science to visualize such possible effects.

ARE CURRENT ATMOSPHERIC CARBON DIOXIDE (CO_2) CONCENTRATIONS UNUSUAL FOR PLANTS?

There is no doubt that human activities are responsible for the current increment of atmospheric CO_2 concentration, which is unusual indeed. Ice cores sampled from the Antarctic ice shield permit a reconstruction of atmospheric CO_2 concentrations over the last 0.4 million years (Figure 1). Air bubbles trapped in fossil ice tell us that CO_2 concentrations oscillated between 180 and 290 ppm over this period of several ice ages and interglacial warm periods such as the current one. Nearly all plant species existing today lived and survived these long periods of low CO_2 concentrations (during peak glaciation, as low as half the current concentration). The mean ambient atmospheric CO_2 concentrations are currently close to 370 ppm (http://cdiac.esd.ornl.gov/ftp/ndp001/maunaloa.co2), nearly one-third above the recorded interglacial maxima, and this has occurred in only about the last 250 years. This is a rapid, very unusual change of the life conditions for plants.

The best evidence that atmospheric CO_2 enrichment is man-made and reflects burning of biotic (both recent and fossil) carbon (C), comes from studies of stable carbon isotopes in the atmosphere. CO_2 containing the heavier ^{13}C isotope (ca. 1.1% of all carbon on the globe) is slightly discriminated by plants during photosynthesis as compared with $^{12}CO_2$. Consequently, all carbon compounds of biological origin, including fossil ones, contain less ^{13}C; and emissions of CO_2 from biotic carbon sources (lighter carbon) will always dilute the ^{13}C concentration in the atmosphere, exactly what records show (Figure 2). Over nearly four decades biologists explored the influences of CO_2 enrichment on plants, and research during the last two decades have produced about 3000 publications on this topic (Saxe *et al.*, 1998; Körner, 2000).

WHY DOES CO_2 ENRICHMENT INFLUENCE PLANT LIFE DIRECTLY?

Just like any other chemical reaction, the enzymatic binding of CO_2 in the chloroplasts of photosynthesizing organs depends on substrate concentration (in this case CO_2 concentration). CO_2 must enter the leaf interior through tiny, but variable pores (the stomata) through which plants also lose water vapor. The more these pores are closed, the more is the uptake of CO_2 restricted. When soil and atmospheric moisture are high, these pores are wide open and thus allow CO_2 concentrations beneath the pores (in the intercellular air spaces of photosynthesizing tissue) to be maintained at 70–80% of that outside the leaf, while photosynthesis inside the leaf consumes CO_2. When stomata are closed during the daytime (e.g., because of drought), the internal CO_2 concentration can fall to less than 10% of that outside the leaf. CO_2 concentration is a rate-limiting factor for photosynthesis.

Although specialized to utilize very low concentrations of CO_2, the photosynthetic machinery, in essence the CO_2-binding enzyme called RubisCO, is able to process CO_2 at much faster rates if concentrations at the reaction sites are increased up to the point where the process becomes substrate saturated (commonly around three times current ambient concentrations). Such non-linear responses to environmental changes can be found for almost any plant trait, but are particularly obvious for the leaf photosynthetic response to CO_2, a curve which is the basis of almost all predictive models, even if not explicitly stated. During geological periods when CO_2 concentrations became very low for the first time (during the Cretaceous period) a number of plant families evolved species which adopt a trick to rate-saturate RubisCO naturally, by actively pumping

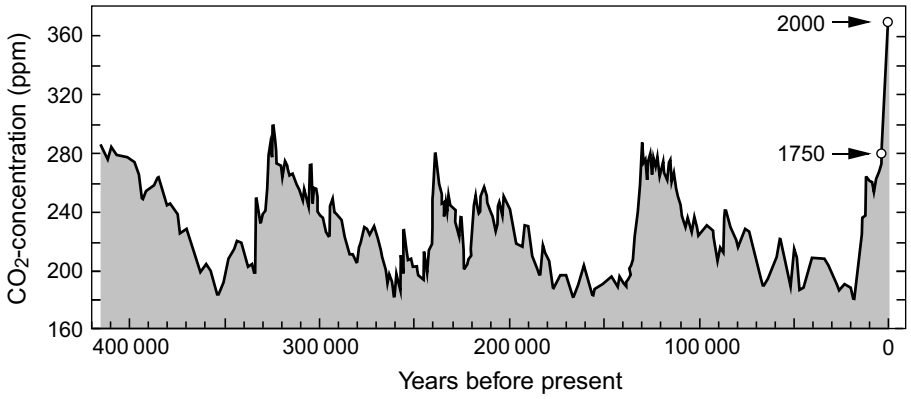


Figure 1 A 420 000 years record of CO_2 concentration in air bubbles trapped in the Antarctic ice indicates that the biosphere operated between 180 and 290 ppm during this period, except for the most recent time, when CO_2 rose to the current mean of ca. 368 ppm (Petit *et al.*, 1999)

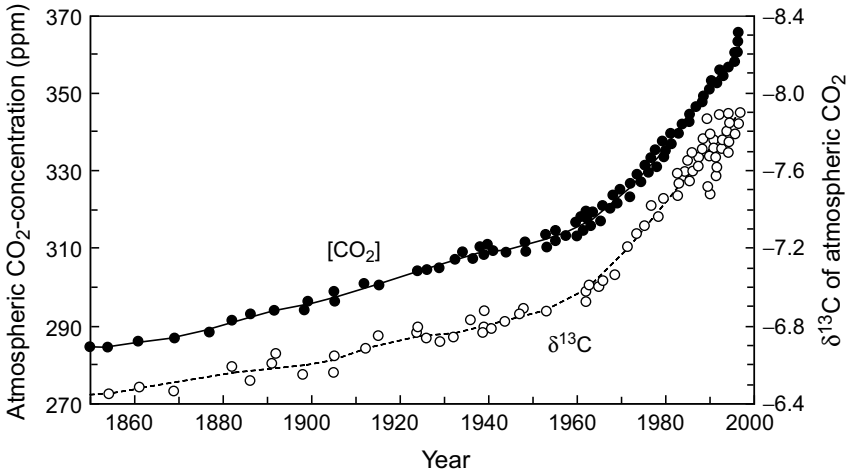


Figure 2 Atmospheric CO_2 enrichment and $^{13}CO_2$ depletion go hand in hand, which underlines the anthropogenic cause of both, namely biogenic carbon release (Joos *et al.*, 1999)

CO_2 to the reaction sites using an enzyme called phosphoenalpyruvat (PEP) carboxylase and a special anatomical arrangement of cells (called kranz anatomy). The first intermediate product of this process is a compound with four carbon atoms, the reason why these specialists are called C4 plants (in contrast to the majority of species which operate without this expensive pump, the so-called C3 species). *See* **C_3 and C_4 Photosynthesis**, Volume 2. The C4 pathway has been and still is very successful in periodically dry and hot environments, where plants need to save water and thus, need to restrict vapor loss, which at the same time restricts CO_2 diffusion into the leaf (among crop plants, maize, sorghum and sugar cane belong to this group of species). It is obvious that photosynthesis of C4 and C3 plant species will not be equally sensitive to atmospheric CO_2 enrichment (Figure 3). Under high soil moisture, C3 plants have been found to be much more responsive. In summary, there are good biochemical reasons why higher atmospheric CO_2 concentrations will influence plants.

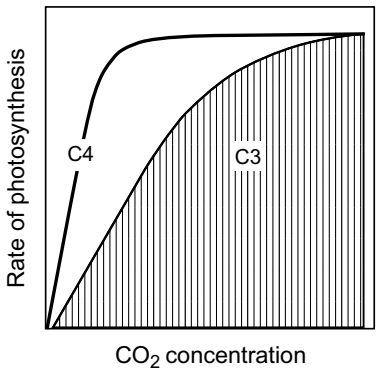


Figure 3 The instantaneous response of photosynthesis to increasing CO_2 concentrations on a unit leaf area basis for C3 and C4 plant species (schematic)

HOW WILL PLANTS RESPOND TO A NEW CARBON DIET?

The principles of photosynthetic responses to environmental changes, including CO_2 enrichment, are well understood and documented. Other responses such as those of growth and development of plants are much more difficult to understand and predict, because these depend on many internal and external drivers in addition to those influencing photosynthesis directly. In many cases it is not even clear whether CO_2 supply represents the bottleneck for biomass production. One can compare the significance of these other determinants of biomass production with the conditions which control automobile speed in a metropolitan area, which rarely depends on fuel availability or the power of the engine. The most important of these external factors for plant growth responses to CO_2 enrichment are mineral nutrient availability, water, light and disturbance (e.g., browsing animals). However, this is not a classical minimum factor limitation situation, because CO_2 enrichment alters the relative demand of these resources, and the sensitivity to these other growth determinants. Commonly, plants can cope with less nutrients, water and light if provided with more CO_2 (Drake *et al.*, 1997). The situation becomes more complex if one wishes to account for interactions between plants (e.g., competition for resources under an altered CO_2 supply) and between plants and other organisms (herbivores, symbionts, soil microorganisms). It becomes impossible to predict plant growth responses to CO_2 enrichment from first principles only (the photosynthetic response), when effects of plant age and plant development in general (including reproduction) come into play. These are the fields where empirical work has improved our capability to develop likelihood scenarios. Experimental sciences (together with mathematical simulation) are now able to provide projections of what is likely and what is less likely to occur, but the long-term adjustment of the whole biosphere to a doubled CO_2 world cannot be predicted with engineering certainty, something decision makers need to bear in mind. What follows here, is a brief account of what is most likely to happen as CO_2 concentration continues to increase, and how observational data must be weighted in order to arrive at a realistic picture.

THE FOUR MOST IMPORTANT CRITERIA FOR UNDERSTANDING PLANT AND ECOSYSTEM RESPONSES TO ELEVATED CO_2

Firstly, it is important that *absolute* (e.g., increment in grams) and *relative responses* (% change) are distinguished. Very poor growth conditions including those at the edge of survival, may in fact facilitate substantial relative responses to elevated CO_2 at otherwise negligible absolute gains. For plants in the deepest shade of a forest, for instance, CO_2 enrichment may permit net carbon gains to increase from zero to a minute positive value, which mathematically counts as an infinite relative stimulation (e.g., Figure 3). Ecologically such relative responses may be very important, but they would not permit much additional carbon binding. On the other hand, a sustained absolute annual increase of the current net carbon stock in the biomass of the forests of the world, of the size of only *ca.* 6 Gt C or ca. 1% of the current global biomass C-pool (ca. 10–15% of annual terrestrial net primary production not becoming recycled by respiration) would remove nearly all annually emitted fossil carbon from the atmosphere/ocean system. To put this number in perspective, one needs to bear in mind that a forest with a 100-year rotation has a mean net carbon fixation of 1% of its final carbon stock per year, hence one additional percent per year means doubling its growth rate or halving its rotation time (see the second criteria below). When CO_2 responses reported in the literature are compared, this distinction between absolute and relative responses is essential. Relative responses are often more relevant in an ecological context (e.g., for biodiversity); absolute responses are of greater interest when carbon binding is discussed.

Secondly, another important distinction is the one between *fluxes* and *pools* of carbon. A growth rate indicates a flux of carbon from the atmosphere to a plant or to the whole ecosystem per unit of time (often termed *productivity*). Live and dead plant mass and soil humus represent the biological carbon inventory per unit land area, a carbon pool of a certain, but limited lifetime. Growth rates and carbon stocks commonly are not coupled, in fact, their interrelationship often is an inverse one. The greatest carbon stocks are found in old forests which have little annual growth (little annual net carbon uptake), whereas highly productive plantations commonly represent much smaller carbon pools but grow quickly. What matters in terms of biospheric carbon binding is the amount of carbon per unit of land area, and not the rate of carbon cycling through these terrestrial pools. Hence, a stimulation of growth represents net carbon binding only if it is not balanced by plant death, burning etc. elsewhere. Since all organisms ultimately die, it is a matter of their age demography whether their populations fix net amounts of carbon. If life cycles are artificially synchronized (e.g., in a forest plantation) a carbon release wave is exported into the future, when these plants reach their life span or harvest size. Hence, a short-term growth stimulation at one time and one place must not be confused with carbon sequestration (IGBP Terrestrial Carbon Working Group, 1998). However, such transitory carbon binding may be seen as a means of buying time, i.e., postponing the atmospheric CO_2 enrichment by a few years, at least partly, at the price of burdening the distant future with carbon recycling. From this it is obvious that the most effective long term management option for biological carbon binding is

to increase the aerial extent of carbon-rich forests and preserve the old ones still existing (cf. IGBP Terrestrial Carbon Working Group, 1998). A second option is to increase the total amount of forest products used in construction. A very effective fossil carbon saving strategy and the only sustainable biological measure to mitigate CO_2 accumulation in the atmosphere, is the substitution of fossil carbon by bio-carbon (Marland and Schlamadinger, 1997).

Thirdly, CO_2 enrichment causes non linear responses and responses change with the duration of exposure and plant age. This makes it extremely difficult to project experimental findings derived from exposing young plants to a step change in CO_2 concentration into the future. As can be seen from Figure 3, the relative response of photosynthesis gets smaller as CO_2 concentration rises, a dampening phenomenon, which becomes enhanced if photosynthetic acclimation takes place. It is plausible to assume that biomass responses to each ppm of atmospheric CO_2 enrichment were greatest when CO_2 concentrations started to rise 250 years ago, are smaller today, and may already (or soon) approach zero in some natural ecosystems (ecosystem CO_2 saturation). Furthermore, a step increase of CO_2 often causes a strong initial stimulation of growth (for a couple of years) which gradually declines with time and/or plant age, so that the duration of the actual observation/treatment period becomes a codeterminant of results (Figure 4). There are possibly not more than five published studies worldwide which extended over more than five years of experimental CO_2 enrichment, and so far none of these included trees, which are responsible for nearly 90% of all biomass carbon on the globe. The only long-term observation suggesting that such age dependency may in fact exist, comes from trees growing around natural (geological) CO_2 springs (Figure 4).

The fourth criterion to be considered is that soil quality determines plant responses to CO_2. One of the lessons of a large number of CO_2 enrichment experiments is that there is no common response, but one which depends on experimental conditions. When amply supplied with essential resources other than CO_2, CO_2 enrichment can greatly enhance plant growth, both in relative but even more so, in absolute terms. Under such conditions the mean seasonal stimulation of biomass accumulation in the temperate zone by a doubling of CO_2 supply is around 30%. This is why CO_2 enrichment became economically important in greenhouse agronomy (Wittwer, 1984). Under less optimal growth conditions, responses depend on the relative abundance of soil resources. In many cases little stimulation of growth has been seen under more natural growth conditions. Most remarkably, but hardly investigated, is the fact that plants can even respond in opposite directions to CO_2 enrichment, when grown in communities under close to natural conditions, but on two different soil types (Figure 5). Another example is given by legumes such as

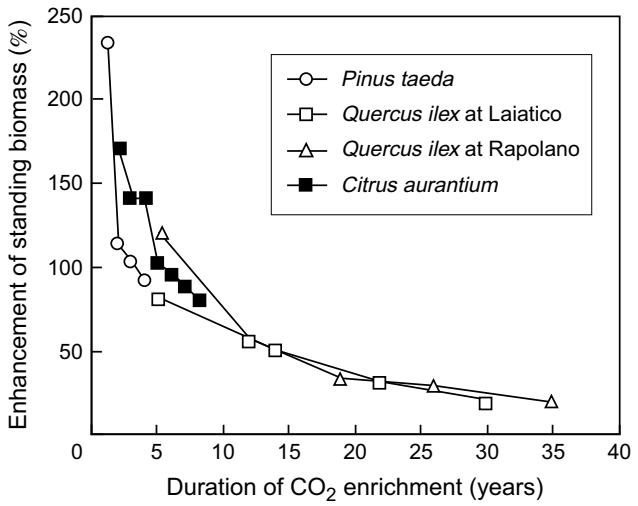


Figure 4 The percent enhancement of standing tree biomass produced by a sustained ca. 300 ppm increase in atmospheric CO_2 concentration for three tree species versus the duration of exposure (Idso, 1999 and references therein)

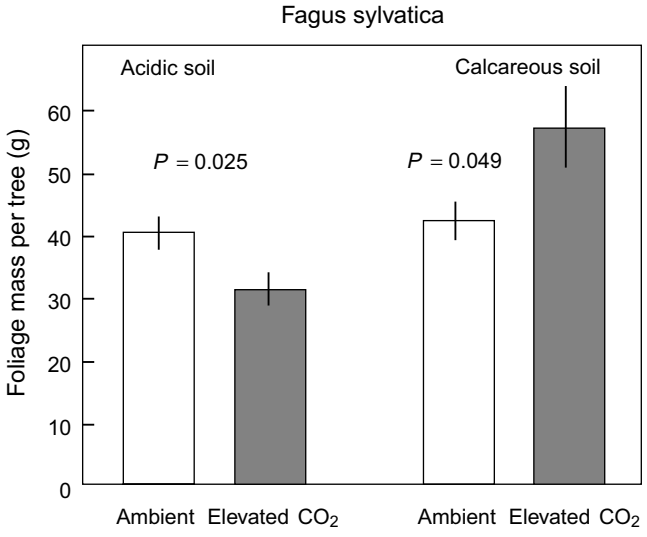


Figure 5 European beech, grown in 600 instead of 360 ppm CO_2 on forest soil in a mixed model community with spruce and understorey plants, either gains or losses in elevated CO_2 in terms of leaf production, when tested on calcareous or on acidic soil (after Egli *et al.*, in Körner, 2000)

clover which live in symbiosis with N_2-fixing bacteria. N_2 fixation is a costly process and can be expected to profit from CO_2 enrichment. This is indeed the case, but only under conditions of adequate phosphate availability, not a situation very common in nature.

All reports on responses of plants and vegetation to atmospheric CO_2 enrichment must be viewed with these four key criteria in mind.

GROWTH AND PRIMARY PRODUCTION

Given that most of the experimental data available today are for a step change of CO_2 (doubling) and for growth conditions under which CO_2 may indeed be *the* limiting resource (e.g., isolated, young, well fertilized plants), it is very difficult to estimate how natural vegetation might respond to the gradual enhancement of CO_2 over the long run. There is no safe way to distill such projections from statistical treatments of published data, because these would always reflect the frequency distribution of experimental conditions, despite attempts to categorize them. Mathematical models are limited to the extent that their parameterization depends on experimental data, and those which start from physiological first principles are likely to come up with overestimates of carbon binding. Approaches assuming stoichiometric proportionality, i.e., the need for certain amounts of nitrogen (N) and phosphate (P) for the biological fixation of one unit of carbon, are more promising, but the central question, whether soils will release the needed amounts of nitrogen and phosphate, and the consequences of atmospheric nitrogen deposition for the carbon cycle, as discussed by Nadelhoffer *et al.* (1999), are still controversial (error terms as big as the current missing carbon of ca. 1.5 Gt C year^{-1}). The nitrogen cycle (and often linked to it, the phosphate cycle) is most likely driving both, the terrestrial as well as the oceanic biotic carbon cycle (for oceans see Falkowski *et al.*, 1998).

A reasonable estimate of the long-term upper limit of CO_2 responsiveness of seasonal biomass production based on a handful of field data obtained under as natural as possible experimental growth conditions may be a 5–15% enhancement, by the time that atmospheric CO_2 concentration will have doubled (second half of 21st century), compared with pre-industrial concentrations. As mentioned above, such seasonal growth stimulations must not be confused with the size of carbon stocks, which also depend on concurrent or follow-up biomass breakdown, mortality and disturbance. In woody plants, such seasonal gains can accumulate and may accelerate forest turnover. If such carbon gains would translate into greater soil humus stocks, the transfer of biomass carbon (C/N ratio 40–80) into humus carbon (C/N ratio 10–20) would almost quadruple nitrogen demand – an unlikely sustainable and unwanted effect, given that it would reduce nutrient availability for plant growth. Due to these constraints, long-term net carbon binding (net ecosystem productivity, NEP) is unlikely to be of the magnitude to keep the global carbon budget in balance (ca. 4–5 Gt C year^{-1}, in addition to the current amount). Currently a biospheric net carbon fixation of around 1.5 Gt C year^{-1} is assumed, a number not derived from biological carbon accounting, but obtained by default from comparing anthropogenic carbon release data with the atmospheric and dissolved surface ocean pools. Carbon release by forest destruction, losses of carbon from soil humus due to intense agriculture or by erosion and carbon release from polar carbon stocks because of the current thawing of the permafrost in arctic ecosystems, all together are likely to balance or exceed CO_2-induced carbon binding (CO_2 fertilization) on a global scale.

Other, more indirect effects of atmospheric CO_2 enrichment on ecosystems, seem to be at least as, if not more influential, than CO_2-induced carbon binding as such. Three of such interactions are of particular interest:

- CO_2 enrichment and water;
- CO_2 enrichment and plant quality;
- CO_2 enrichment and biodiversity.

CO_2 AND WATER RELATIONS

Because plants tend to optimize the CO_2 uptake/water vapor-loss ratio by stomatal adjustments, higher concentrations of CO_2 around a leaf commonly cause a reduction of the stomatal pore width (a reduction of stomatal diffusive conductance). In cases where this happens, this has a multitude of consequences: At given ambient air humidity, leaf transpiration will be reduced. This in turn will cause leaf temperature to slightly increase (reduced evaporative cooling). Reduced transpiration will also cause less humidification of surrounding air which, together with leaf warming, will steepen the leaf–air vapor gradient, which will reduce the net CO_2 effect on transpiration, and make it hard to predict. The effect is further dampened by aerodynamic constraints to gas diffusion, which are not influenced by stomata, but are resistors in series with the stomatal one. The few field data on evapotranspiration of grassland under >500 ppm CO_2 suggest an instantaneous net reduction of only a few percent. However, such small but continuous water savings may accumulate into significant changes of the hydrological balance: during rainless periods, soil water will be less rapidly depleted, drought will be delayed and nutrient availability will be enhanced. In contrast, during more humid periods, runoff and nutrient losses will be enhanced, because soils will remain near full moisture storage capacity for longer periods of time. Several years of experimental field trials indicate that these indirect, moisture-driven effects of CO_2 enrichment are likely to be more significant for growth than direct CO_2 effects (Figure 6). This also explains the puzzling observation that growth of C4-plants is stimulated by CO_2 enrichment (not to be expected from Figure 3). Just like C3 plants, they profit primarily from a moister rhizosphere as a result of stomatal responses to CO_2. Stomatal responses to elevated CO_2 were found to be species specific, hence savings by one species may also be beneficial to neighboring species,

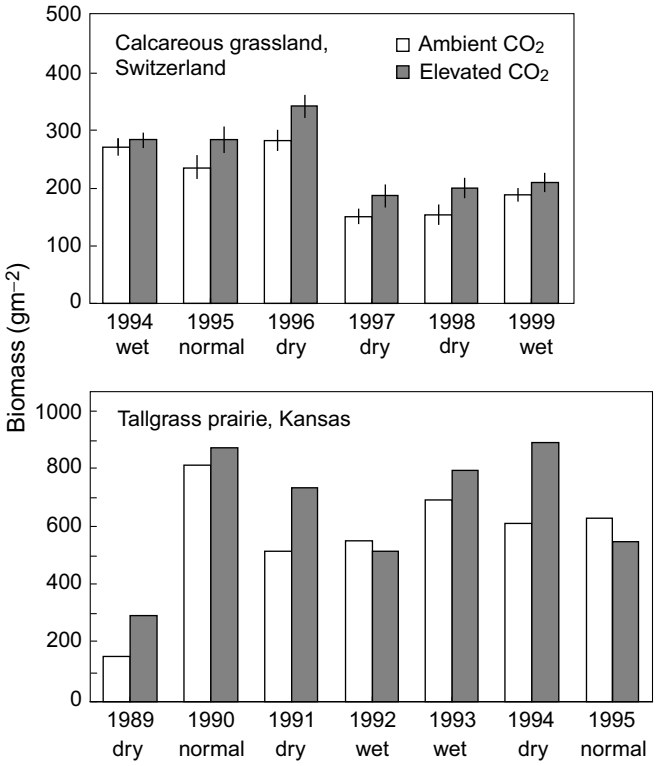


Figure 6 Due to CO_2-driven soil moisture savings, CO_2 effects on biomass production are commonly more pronounced in dry seasons as compared with very wet seasons (examples from Kansas, Owensby *et al.*, 1997, and from Switzerland by P Niklaus and Ch Körner, unpublished)

inducing changes in community structure in favor of water wasters.

While stomata of tree seedlings have often been found to be similarly responsive to CO_2 as are herbaceous plants, the evidence for adult trees is not unequivocal. To date it seems safest to conclude that stomatal responses in tall trees are much smaller – in a number of cases the response was close to zero. This difference between low stature plants in dense herbaceous canopies as compared with tall trees may be associated with better aerodynamic coupling of the latter (smaller boundary-layer resistance to gas diffusion, causing small changes in stomatal aperture to become more effective).

In summary, water relations will be affected, but possibly less in forests than grasslands, and instantaneous responses to evapotranspiration, if they occur, will be very small ($<10\%$). Whether they will become important, is a question of the duration of rainless periods (cumulative effects). The great unknown is whether atmospheric feedback at larger scales or a changing climate will diminish or enhance these effects. Overall, semi-arid regions are likely to profit (more biomass per unit land area) and very humid regions will be unaffected or may suffer (enhanced runoff and flood risk).

CO_2 AND PLANT TISSUE QUALITY

Irrespective of any growth response, plant exposure to elevated CO_2 alters tissue composition. Leaves have almost always been found to contain less protein and more mobile carbohydrates when exposed to higher CO_2 concentrations. The result is a wider C/N ratio, with a multitude of ecosystem consequences (Cotrufo *et al.*, 1998). The increase of the concentration of total non-structural carbohydrates (TNC) e.g., starch, sugars) is an immediate response (new steady state within a few days), and persists over very long times, as has been documented for plants from natural CO_2 springs. The response appears to be largely independent of sink activity, vigor of growth, nutrient supply, size or age of plant and occurs even in deep shade and in the canopy of tall forests, hence it seems to represent a direct response to ambient CO_2 concentration (references in Körner, 2000). Reduced leaf nitrogen concentration often does not go beyond a dilution effect by enhanced TNC on a dry weight basis, but if it does, it is associated with either partial photosynthetic downward adjustment (less demand) or reduced nitrogen availability in the root environment (less supply). CO_2 enrichment appears to stimulate carbon export by roots, which in turn causes microbial and fungal competition for nitrogen in the soil to increase (e.g., Diaz *et al.*, 1993). Nitrogen concentrations in CO_2-enriched leaves also depend on phosphate availability. Secondary metabolites are affected as well, with phenolics most commonly becoming enhanced (e.g., Poorter *et al.*, 1997).

The most immediate consequence of reduced leaf tissue quality is that on herbivores, which (1) may either consume more (compensatory feeding), (2) consume less and suffer, or (3) switch food plants, unless they are host specific (e.g., Lindroth in Körner and Bazzaz, 1996). Litter quality is commonly not (or less) affected (Norby *et al.*, 1999). A most important but hardly studied question is the effect of elevated CO_2 on wood quality. It seems CO_2 enrichment enhances wood density (cf. Körner, 2000), but the mechanical consequences are still unclear (greater/less risk of wind break?). Seed nitrogen concentration, including that of cereals, is also reduced under elevated CO_2. Among all CO_2 responses, the responses of tissues appear to be the most consistent, and are surprisingly robust for a variety of experimental conditions. The prediction that a CO_2 rich world will be inhabited by plants with tissues richer in carbon seems well founded.

CO_2 AND BIODIVERSITY

For every single type of response to elevated CO_2 discussed above, different plant species were found to differ, which

suggests that CO_2 enrichment will affect species abundance and the composition of plant communities (Körner and Bazzaz, 1996). When all available data are considered, conventional functional types of plants do not appear to exhibit common responsiveness, perhaps with the exception that legumes grown in phosphate-saturated substrates are more responsive and C4 plants in very moist substrates are less responsive than other plants (both very artificial situations). Within herbaceous, non-nutrient limited plants growing in isolation and unlimited moisture, C3 species with a high relative growth rate (RGR), were found to be more responsive to CO_2 enrichment, than low RGR-species, but no systematic difference was found between either monocotyledonous and dicotyledonous species or wild versus crop species (Körner and Bazzaz, 1996). Neither do seedlings or saplings of woody plants respond significantly differently from herbaceous species, nor do evergreen differ from deciduous ones in a systematic way. However, within each of these groups highly responsive and hardly responsive species occur, but whether a species is responsive is often codetermined by the genotype, nutrient regime, light regime or neighboring species. The micro-habitat's light regime is a critical determinant of the ranking of CO_2 responsiveness among tree species (Figure 7).

Mycorrhization is mostly stimulated by elevated CO_2 and may codetermine CO_2 responsiveness of species (references in Staddon and Fitter, 1998). More carbon is commonly allocated to soil fungi than to soil bacteria (e.g., Rillig *et al.*, 1999). There is sporadic evidence that growth of non mycorrhizal species is particularly responsive. For instance, several laboratories independently found that *Carex flacca*, a non-mycorrhizal slow growing sedge, out-ranged the growth responsiveness of any other (mycorrhizal) calcareous grassland species tested (references in Körner, 2000).

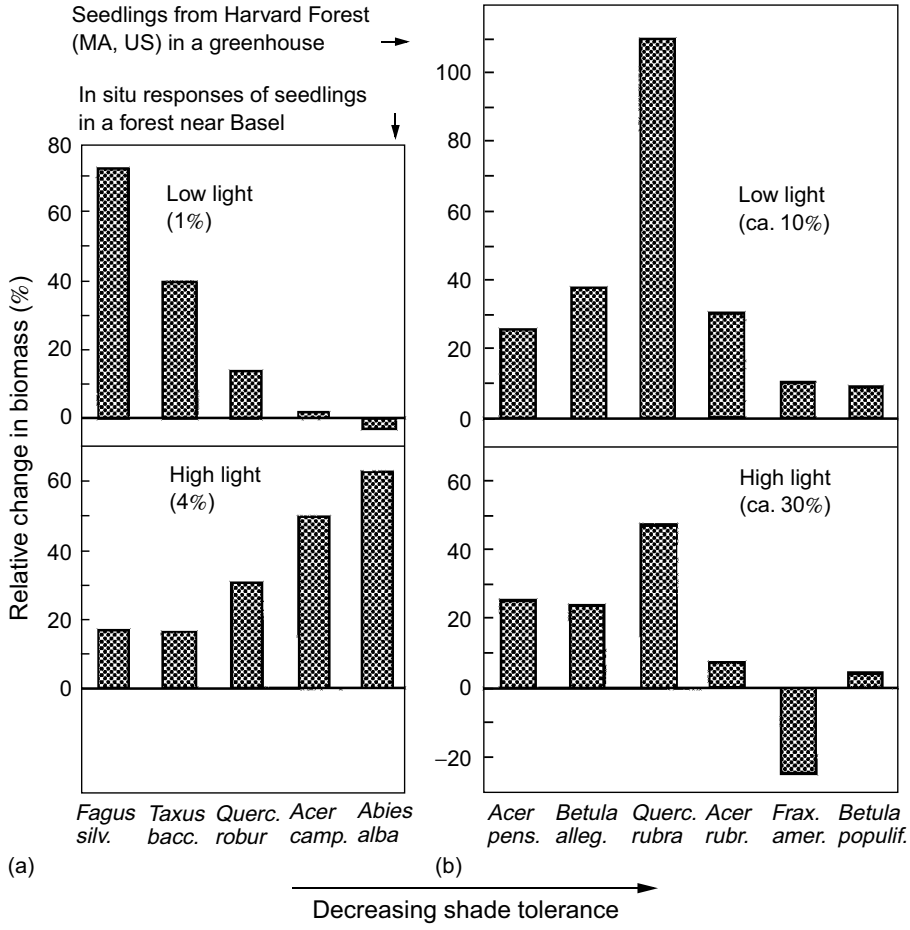


Figure 7 Different plant species respond very differently to atmospheric CO_2 enrichment, but how they respond depends on the concurrent availability of other resources. In the example (a), the availability of either 1 or 4% of full sunlight in the understory of a dense deciduous forest in Switzerland reversed the responsiveness ranking among seedlings of native tree species. Also greenhouse tests (b) document such species and light dependent CO_2 effects. The speci c responses of such species will codetermine future forest species composition (Bazzaz and Miao, 1993; Hättenschwiler and Körner, 2000)

THE FUTURE BIOSPHERE AND THE GLOBAL CARBON BALANCE

Carbon Stocking

As CO_2 rises, the biosphere will be affected in one way or another. However, by far the most significant influence on the globe's biota carbon stock is and will remain human land use. Unless current trends change fundamentally, land use will determine biological carbon pools and biodiversity to such an extent that other influences such as those of elevated CO_2 will remain marginal. The biological impact of elevated CO_2, as presented here, needs to be seen in this larger context, and – apart from controlling emissions – efforts clearly need to be prioritized on achieving sustainable land use and the preservation of a large terrestrial carbon pool. Several authors have recently modified earlier simplistic predictions of a greening earth in view of the past and current trends in atmospheric CO_2 enrichment (Schimel, 1995; Joos *et al.*, 1999). In the face of the magnitude of the land use problem, society should not expect a significant mitigation of atmospheric CO_2 enrichment through CO_2-enhanced biospheric carbon binding. Given that part of the current missing carbon of ca. 1.5 Gt C year^{-1} is explained by recent temperate zone forest re-growth, the pure global carbon fertilization component of this (if there is one) is most likely less than 1 Gt C year^{-1} for the recent dramatic 30% increase of CO_2 concentration above pre-industrial concentrations.

There are only three, quantitatively significant ways for dampening the rate of current atmospheric CO_2 increase through bio-management: (1) a halt to the destruction of carbon-rich pristine forests as well as humus-rich soils, (2) an extension of forests and, within agricultural land, those areas of low energy input agriculture, and (3) the sustainable use of forest- and agro-products in replacement of fossil carbon sources (cf. Marland and Schlamadinger, 1997). Models suggest an upper limit (!) of a CO_2-induced stimulation of carbon binding in biota by the time atmospheric CO_2 concentrations will have doubled (compared with the pre-industrial level) of around 3 Gt year^{-1} or 30 g C m^{-2} a^{-1} on 100 million km^2 of vegetation-covered land, with a mean 15 kg C m^{-2} background value in humus carbon. The total accumulative carbon binding by this time may have reached ca. 200 Gt, not accounting for any effects of land use (ref. in Körner, 2000). By that time humans will have released nearly 1000 Gt of fossil carbon. This upper limit of carbon binding would only be reached if proportional nitrogen and phosphate supplies for the formation of additional biomass and humus became available (this may be so in some polluted, industrialized areas). The narrow C/N ratio of humus will unlikely be converted into a larger biomass carbon pool by way of its wider C/N ratio, as was one of the now revised earlier presumptions (stoichiometric constraints, Nadelhoffer *et al.*, 1999). Hence, the actual CO_2-induced carbon binding will be less than this upper limit (<20% of emissions), and is likely to trap only a small fraction of fossil and land use derived atmospheric carbon. As was explained above, growth stimulation as revealed by experimental CO_2 enrichment, must not be confused with a net landscape carbon stocking effect (IGBP Terrestrial Carbon Working Group, 1998). The biological carbon binding of the ocean, which, in the very long run (many thousands to millions of years), ensures a steady atmospheric oxygen concentration, is estimated to amount to ca. 0.1 Gt C year^{-1} (Reeburgh, 1997), an irrelevant carbon sink with respect to enhanced carbon binding in the light of current global carbon emission rates of ca. 7 Gt C year^{-1} over the periods relevant to society.

CO_2 and the Biodiversity Crisis

Instead of expecting substantial carbon fertilization effects on biospheric carbon stocking, society, scientists and their funding agencies should be concerned about the differential consequences of the rapid dietary change imposed on organisms and their communities via increased CO_2 concentrations. Subtle species-specific responses and effects on organismic interactions represent a substantial biodiversity risk on top of that caused by land use pressure. The rapidity of CO_2 enrichment (Figure 1) leaves little leeway, if any, for evolutionary adjustment, and affects ecosystems at great distance from the prime sources of fossil CO_2. Biodiversity consequences of CO_2 enrichment operate via differential interactions with species' water and nutrient relations, plant–plant, plant–animal, plant–microbe and plant–fungus interactions as was discussed above.

Elevated CO_2 contributes to the biodiversity crises in a slow, invisible manner. It is one of the prime needs of global change science to visualize such possible effects. Many researchers have demonstrated that plants and their microbial symbionts as well as dependent animals respond to CO_2 enrichment in unpredictable ways. Responses depend so much on experimental conditions (Figure 7) and biotic interactions that field-oriented work in non-agro systems should be given a high priority. This sort of ecosystem research is most often more demanding, more expensive and takes much longer to arrive at a conclusive picture, than the study of standardized and simple test systems on disturbed substrates, which have their own place in agro-research (Rosenzweig and Hillel, 1998). This is the price that must be paid to achieve credible answers about the biological consequences of the ever rising CO_2 concentrations. The greatest current limitation of predictions is the unknown response of diverse natural forests, which form the single largest biomass carbon pool and represent a large, if not the largest fraction of global species diversity.

See also: **Plant Growth at Elevated CO_2**, Volume 2.

REFERENCES

Bazzaz, F A and Miao, S L (1993) Successional Status, Seed Size, and Responses of Tree Seedlings to CO_2, Light, and Nutrients, *Ecology*, **74**, 104–112.

Cotrufo, M F, Ineson, P, and Scott, A (1998) Elevated CO_2 Reduces the Nitrogen Concentration of Plant Tissues, *Glob. Change Biol.*, **4**, 43–54.

Diaz, S, Grime, J P, Harris, J, and McPherson, E (1993) Evidence of a Feedback Mechanism Limiting Plant Response to Elevated Carbon Dioxide, *Nature*, **364**, 616–617.

Drake, B G, Gonzalez-Meler, M A, and Long, S P (1997) More Efficient Plants: a Consequence of Rising Atmospheric CO_2? *Annu. Rev. Plant Physiol.*, **48**, 609–639.

Falkowski, P G, Barber, R T, and Smetacek, V (1998) Biogeochemical Controls and Feedbacks on Ocean Primary Production, *Science*, **281**, 200–206.

Hättenschwiler, S and Körner, Ch (2000) Tree Seedling Responses to *In Situ* CO_2-enrichment Differ Among Species and Depend on Understorey Light Availability, *Glob. Change Biol.*, **6**, 213–226.

Idso, S B (1999) The Long-term Response of Trees to Atmospheric CO_2 Enrichment, *Glob. Change Biol.*, **5**, 493–495.

IGBP Terrestrial Carbon Working Group (1998) The Terrestrial Carbon Cycle: Implications for the Kyoto Protocol, *Science*, **280**, 1393–1394.

Joos, F, Meyer, R, Bruno, M, and Leuenberger, M (1999) The Variability in the Carbon Sinks as Reconstructed for the Last 1000 years, *Geophys. Res. Lett.*, **26**, 1437–1440.

Körner, Ch and Bazzaz, F A, eds (1996) *Carbon Dioxide, Populations, and Communities*, Academic Press, San Diego, CA.

Körner, Ch (2000) Biosphere Responses to CO_2-enrichment, *Ecol. Appl.*, **10**, 1590–1619.

Marland, G and Schlamadinger, B (1997) Forests for Carbon Sequestration or Fossil Fuel Substitution? A Sensitivity Analysis, *Biomass Bioenergy*, **13**, 389–397.

Nadelhoffer, K J, Emmett, B A, Gundersen, P, Kjonaas, O J, Koopmans, C J, Schleppi, P, Tietema, A, and Wright, R F (1999) Nitrogen Deposition Makes a Minor Distribution to Carbon Sequestration in Temperate Forests, *Nature*, **398**, 145–148.

Norby, R J, Wullschleger, S D, Gunderson, C A, Johnson, D W, and Ceulemans, R (1999) Tree Responses to Rising CO_2 in Field Experiments: Implications for the Future Forest, *Plant Cell Environ.*, **22**, 683–714.

Owensby, C E, Ham, J M, Knapp, A K, Bremer, D, and Auen, L M (1997) Water Vapor Fluxes and their Impact Under Elevated CO_2 in a C4-tallgrass prairie, *Glob. Change Biol.*, **3**, 189–195.

Petit, J R, Raynaud, D, Barkov, N I, Barnola, J M, Basile, I, Bender, M, Chappellaz, J, Davis, M, Delaygue, G, Delmotte, M, Kotlyakov, V M, Legrand, M, Lipenkov, V Y, Lorius, C, Pepin, L, Ritz, C, Saltzman, E, and Stievenard, M (1999) Climate and Atmospheric History of the Past 420 000 years from the Vostok Ice Core, Antarctica, *Nature*, **399**, 429–436.

Poorter, H, VanBerkel, Y, Baxter, R, DenHertog, J, Dijkstra, P, Gifford, R M, Griffin, K L, Roumet, C, Roy, J, and Wong, S C (1997) The Effect of Elevated CO_2 on the Chemical Composition and Construction Costs of Leaves of 27 C-3 Species, *Plant Cell Environ.*, **20**, 472–482.

Reeburgh, W S (1997) Figures Summarizing the Global Cycles of Biogeochemically Important Elements, *Bull. Ecol. Soc. Am.*, **October**, 260–267.

Rillig, M C, Field, C B, and Allen, M F (1999) Soil Biota Responses to Long-term Atmospheric CO_2 Enrichment in Two California Annual Grasslands, *Oecologia*, **119**, 572–577.

Rosenzweig, C and Hillel, D (1998) *Climate Change and the Global Harvest: Potential Impacts of the Greenhouse Effect on Agriculture*, Oxford University Press, New York.

Saxe, H, Ellsworth, D S, and Heath, J (1998) Tansley Review No. 98 Tree and Forest Functioning in an Enriched CO_2 Atmosphere, *N. Phytol.*, **139**, 395–436.

Schimel, D S (1995) Terrestrial Ecosystems and the Carbon Cycle, *Glob. Change Biol.*, **1**, 77–91.

Staddon, P L and Fitter, A H (1998) Does Elevated Atmospheric Carbon Dioxide Affect Arbuscular Mycorrhizas? *Trend. Ecol. Evolut.*, **13**, 455–458.

Wittwer, S H (1984) Carbon Dioxide Levels in the Biosphere: Effects on Plant Productivity, *CRC Crit. Rev. Plant Sci.*, **2**, 171–198.

Community Richness

Paul H Harvey

University of Oxford, Oxford, UK

If species are accepted as equivalent and independent entities, then there are two intuitive components of biodiversity at a site: species richness and species evenness. Species richness is a measure of how many species there are, and species evenness measures the relative abundance of different species. If species evenness is low, in the extreme a single species predominates and the site is effectively a monoculture. If species evenness is high, in the extreme all species are equally abundant. There are an infinite number of measures of biodiversity based on species richness and evenness, depending, for example, on the relative weights given to each of the two components.

HOTSPOTS AND COLDSPOTS OF SPECIES RICHNESS

Accepting for the moment that richness is simply a measure of the number of species at a site, a reasonable question to ask is what determines that value? In other words, why are some sites richer in species than others? As scientists, we

should hope to produce generalizations, so that the answer to the question for one taxon would be framed in a way that would apply to other taxa.

A landmark study by Prendergast *et al.* (1993) asked if locations in Britain that are species rich for one taxon are also species rich for other taxa? Butterflies, dragonflies, liverworts, aquatic (angiosperm) plants, and breeding birds were used as representative taxa, and the representative scale was a grid of 10×10 km squares superimposed over the British mainland. Despite the fact that "the fine scale and high intensity of recording in Britain produces distributional data sets that are at least as good as ... those available elsewhere", hotspots (species rich grid squares) and coldspots (grid squares with few species) for different taxa did not coincide. Furthermore, from the perspective of conservation biology, grid squares containing rare species, which might be focussed upon for conservation efforts, did not coincide with hotspots or coldspots.

The failure to generalize species richness from one taxon to those of another was emphasized in a study of endangered animal and plant taxa occurring in the various counties of the United States of America (and therefore including Hawaii). The aim of Dobson's *et al.* (1997) analysis was to determine whether hotspots for endangered species coincide across taxa, and to help identify those areas in need of conservation resources. Using a complementarity algorithm, they first identified the county with most endangered species, then the county with the most endangered species that are not included in the first county, and so on. Most endangered species occur in Hawaii, Southern California, the southeastern coastal states and Southern Appalachia. Their complementarity algorithm also selected counties in those same regions. Overall, between 1 and 5% of total land area delimited those counties, which were required to sample more than 50% of endangered species for particular taxa. However, the counties identified for one taxon were not necessarily those identified for other taxa. For example, they applied the complementarity algorithm to birds to identify the minimum number of counties required to locate all endangered species. That area contained only 30 to 40% of endangered reptiles, amphibians, mammal and plant species, 10% of arthropods and fish, and 2% of molluscs.

Would it, then, be possible to find a taxon that is a much better predictor than birds of the presence of endangered species from other taxa? At first sight the answer appears to be plants. By identifying the minimum area required to conserve all plant species, we would also be conserving 94% of birds, and between 39 and 76% of species in other taxa. The downside to this is that the area containing all the plants is almost 10% of the area of the United States, whereas that for birds is less than 2%.

DETERMINANTS OF SPECIES RICHNESS

The fact that patterns in species richness do not generally coincide across taxa is a disappointment for those seeking general patterns, but this should not detract from an extensive ecological literature, which reveals understanding of the reasons for patterns in species richness for particular taxa or communities. Ultimately, three factors seem to predominate: the number of available niches at a site, the capacity of species to invade those niches, and the time the niches have been available. The time dimension is problematical if species are viewed as independent immutable entities. Of course, when habitats and their component niches first appear, it will take time for species from the local or regional pool to invade. However, occupancy of some niches requires evolutionary adaptation by the new inhabitants, and ultimately speciation may ensue.

The Evolutionary Time Scale

Species richness viewed from an evolutionary rather than a purely ecological perspective provides an historical dimension that is necessary for a comprehensive understanding of the causes of spatial patterns in species richness. What is more, the evolutionary context reveals that all species are not necessarily equivalent. The further back that two species last shared a common ancestor, the more genetically divergent they will have become. If purely confined to phenotypes, the concept of richness does not properly include genetic variation. Closely related species may have diverged phenotypically but be very similar genetically and, alternatively, genetically divergent species may be phenotypically similar.

Vane-Wright *et al.* (1991) pointed out that, when choices have to be made concerning the allocation of limited resources to conservation, the preservation of representatives of two distantly related species would result in the retention of more genetic diversity than the preservation of two genetically similar species. If we assume that genetic divergence increases in a clock-like manner then, using the purely genetic criterion, maximizing the conservation of total branch length on an evolutionary tree will result in the conservation of the maximum amount of genetic diversity, or richness. How important is this message for conservation science? Of course, other things being equal, we should wish to maximize the amount of conserved genetic diversity. Nee and May (1997) tackled this question by simulating evolutionary or phylogenetic trees on a computer and then taking samples of species at random and comparing the amount of genetic diversity in the sample with that in a sample chosen on the criterion of maximizing conserved branch length. Their conclusion was that "80% of the underlying tree of life can survive even when approximately 95% of species are lost, and that algorithms that maximize the amount of evolutionary history preserved are

not much better than choosing the survivors at random". The reason for this is that trees are bushy with a preponderance of recent short branches, and fewer longer branches. It is those longer branches that need to be preserved. So long as one representative taxon from each main phylogenetic group is preserved, the longest branches in the tree of life are all represented and only the short branches have been pruned.

The capacity of species to invade niches depends in large part on the species' phenotypes. On occasion, so-called key innovations evolve which allow species to exploit hitherto unoccupied niches and adaptive radiations may result. Cichlid fishes provide a classic example. Liem (1973, 1980) realized that a small shift in the position of a muscle attachment allowed the pharyngeal bones of cichlids to manipulate their prey items while still holding them. The result was that the mandibular and premaxillary jaws were freed to evolve along new routes, which allowed cichlids to evolve a new diversity of feeding mechanisms. This, Liem argued, led to the adaptive radiations of cichlids in African lakes, despite competition from other fish families. That example is a *post hoc* explanation based on a single innovation. Such explanations are strengthened when the same key innovation is found in independently evolving taxa that have speciated at a higher rate than their close relatives that have not evolved the innovation. Such sister taxon comparisons automatically control for the many differences between sister lineages, because any particular innovation other than the one under test is less likely to be partitioned in the same way among sister taxa.

An example of a repeated key innovation that has resulted in many independent sister taxon comparisons, which reveal high species richness in the taxon with the innovation is provided by phytophagous insects. Half the world's insect species are phytophagous (i.e., they feed on vascular plants), but they are restricted to only nine of the 30 orders of insects. Southwood (1973) imagined that there were considerable barriers to an insect group evolving phytophagy: behavioral and morphological adaptations would be required: (i) to reduce the risk of desiccation; (ii) to remain attached to the host; and (iii) to deal with low nutrient food. Once those problems were overcome, Strong *et al.* (1984) suggested that the diversification of phytophages would be promoted by the great diversity of plants species and diverse plant parts as well as the absence of competitors. As a test of the idea that phytophagy promotes insect diversification, Mitter *et al.* (1988) were able to perform 13 sister taxon comparisons between the diversity of a phytophagous clade and its sister clade. Since sister clades originated at the same time, they have had identical times for diversification. In 11 of the 13 comparisons, there were more species in the phytophagous clade than in the sister group, and in each of those cases the difference was greater than twofold.

A second example of sister group comparisons being used to test ideas for the occurrence of differences in species richness goes back to Darwin. Darwin (1871) suggested that sexual selection by female choice might increase the rate of reproductive divergence between populations and thereby increase the speciation rate. Barraclough *et al.* (1995) tested this idea using passerine birds as a case study. It is generally accepted that mate choice is responsible for the evolution of sexual dichromatism (i.e., where the sexes of a species differ in their plumage coloration) in passerine birds, and so Barraclough *et al.* (1995) used dichromatism as an indirect measure of the importance of mate choice for a species. Passerine species were scored as dichromatic or not. On significantly more than 50% of occasions, the clade with a higher proportion of sexually dichromatic species was more speciose than its sister taxon, thereby supporting the idea that mate choice can increase speciation rate.

The demonstration that, coincident with the occurrence of the same character in different evolutionary lineages, we find increases in speciation rate and thereby high contemporary species richness poses the question: how predictable is evolution? If, by chance, a species invades a new location, to what extent is its evolutionary future predictable, particularly in terms of the number of descendent species it will give rise to? Work by Jonathan Losos on lizards in the Caribbean has tackled that question. There are about 150 species of Anolis lizards in the Caribbean, but only six ecomorphs: phenotypically similar species adapted to particular ecological niches. The ecomorphs are mainly described by the habitats which they exploit: twig, crown-giant, trunk-crown, trunk, trunk-ground, grass-bush. Morphological and ecological comparisons show that the ecomorphs are a valid classificatory tool because they tend to form clusters in multidimensional space when data are summarized by principal component analysis. However, the same ecomorphs have often evolved independently by parallel or convergent evolution, so that different ecomorphs from the same island may be more closely related to each other than the same ecomorph from different islands.

Evolutionary trees for the two independent radiations on Puerto Rico and Jamaica are reasonably well established. When Losos mapped ecomorphs on to phylogenetic tree structure, the two islands revealed very similar pictures (Losos, 1992). The first speciation event resulted in a generalist arboreal lizard and a twig species. The generalist then produced daughter species, a crown-giant and a trunk-ground ecomorph. The trunk-crown ecomorph was the next to evolve (on Puerto Rico this evolved from the trunk-ground species and on Jamaica it evolved from the crown species). The evolution of the trunk-crown species, on both islands, was followed by speciation into a larger and a smaller trunk-crown species. Finally, on Puerto Rico but not on Jamaica, the trunk-ground form spawned a daughter species that lives on grass and bushes independently of

trees. A statistical test demonstrated that the chances of two such similar sequences occurring by chance alone is very small indeed. Instead, the patterns point to a common process at work which Losos interprets as niche partitioning (the tree) with the occasional invasion of new niches (such as the ground, grass and bushes).

The supposed role of niche partitioning in these adaptive radiations implies that interspecific competition may have been an important mechanism forcing species to occupy narrower niches, and the differences between the ecomorphs suggest that particular morphological and behavioral changes should occur over the generations when species invade new habitats.

The Ecological Time Scale

Many species do not need to evolve in order to invade available niches; they simply need to physically migrate. Humans, of course, have frequently changed the scale of such migrations by introducing exotic species to many habitats, sometimes with disastrous consequences including the competitive displacement or predation of the native species. Such extirpations lower global species richness, but generally also have the consequence of lowering species richness locally. Human interference apart, most migrations are local and we have already seen that, at such scales, locations of high diversity for different taxa tend not to coincide. Nevertheless, at a global scale there are very distinct trends in species richness which are repeated, albeit crudely, across taxa, and therefore call for general explanations.

The two most obvious global patterns are marine versus terrestrial differences in species richness, and latitudinal gradients in species richness. Astonishingly for such gross patterns, ecologists are divided over the reasons that the patterns exist. The sea covers much more of the global surface than the land, yet while twice as many phyla are represented in the sea than on land, fewer than 15% of species are found there; this latter result is not simply a result of under-sampling marine environments. May (1994) listed a variety of possible reasons for these differences. If life originated in marine habitats, then, other things being equal, one would expect to find more higher taxa there. May (1994) also suggested the possibility that the evolutionary perspective is likewise responsible for the lower species richness in the sea than on land because marine habitats are less heterogeneous and provide fewer niches for speciation. However, May also pointed to other factors that could be involved, including differences in herbivory and body size distributions between marine and terrestrial ecosystems. In short, we do not yet know the causes of these major differences in both higher taxon and species diversities between the two environments.

The second major and repeated general pattern is that of latitudinal richness gradients. At the extreme, there are many fewer species at the north and south poles than at the equator. This should not surprise us since primary productivity is virtually nil at the poles, and therefore viable plant and animal communities cannot exist there. We might, therefore, conclude that the underlying reason for latitudinal diversity gradients is a consequence of differences in primary productivity: the higher the productivity, the greater the species richness. As we move from polar through temperate to tropical environments, the length of the growing season increases, species richness in plant communities increases, and so therefore does the number of niches available to animal species. However, that level of general explanation is by no means generally accepted. In a review of the literature, Gaston and Williams (1996) quoted reasons for latitudinal gradients as including "the effects of competition, mutualism, patchiness, environmental stability, environmental predictability, productivity, area, number of habitats, ecological time, evolutionary time and solar energy!" Of course, many of those factors are not independent of each other but the depth of the confusion is characterized by quotations from Rohde (1992) that "greater species diversity is due to greater effective evolutionary time (evolutionary speed) in the tropics, probably as a result of shorter generation times, faster mutation rates and faster selection at greater temperatures." Rosenzweig (1992) states that "latitudinal gradients arise because the tropics cover more area than any other zone. Their greater area stimulates speciation and inhibits extinction." Of course, for any patterns there are exceptions, and the field of latitudinal diversity gradients is one where exceptions abound; and have probably been unduly focussed upon, thereby obfuscating the generalities (for example, among insects for mainly unknown reasons, aphids, sawflies and bees are either species poor or not unusually species rich in tropical habitats). There remains a need for good, clear-headed analyses of the massive available databases pertaining to latitudinal diversity gradients, not only to distinguish among alternative explanations, but also to evaluate the relative importance of each.

Given the overriding influence of latitude and its numerous climatic correlates, which frequently co-vary, other gradients have been sought which in various ways control for latitude. The most obvious is longitudinal diversity gradients, though Gaston and Williams (1996) also review altitudinal and depth gradients in species richness. In short, though repeated patterns exist, e.g., mid-altitudinal peaks in species richness among insects, explanations are generally found wanting. There is no doubt that high plant species diversity is generally associated with high animal species diversity, which begs the question of what environmental factors are responsible for high plant species diversity? Until we know the answer to that question, we shall not be

able to predict the ecological time frame changes in species richness that global change has in store for us.

REFERENCES

Barraclough, T G, Harvey, P H, and Nee, S (1995) Sexual Selection as a Cause of Taxonomic Diversity in Passerine Birds, *Proc. R. Soc. Lond., Ser. B: Biol. Sci.*, **259**, 211–215.

Darwin, C (1871) *The Descent of Man, and Selection in Relation to Sex*, John Murray, London.

Dobson, A P, Rodriguez, J P, Roberts, W M, and Wilcove, D S (1997) Geographic Distribution of Endangered Species in the United States, *Science*, **275**, 550–553.

Gaston, K J and Williams, P H (1996) *Biodiversity: a Biology of Numbers and Difference*, ed K J Gaston, Blackwell Science, Oxford, 202–229.

Liem, K F (1973) Evolutionary Strategies and Morphological Innovations: Cichlid Pharyngeal Jaws, *System Zool.*, **22**, 425–441.

Liem, K F (1980) Adaptive Significance of Intra- and Interspecific Differences in the Feeding Repertoires of Cichlid Fishes, *Am. Zool.*, **20**, 295–314.

Losos, J (1992) The Evolution of Convergent Structure in Caribbean Anolis Communities, *System Biol.*, **401**, 403–420.

May, R M (1994) Biological Diversity – Differences between Land and Sea, *Philos. Trans. R. Soc. Lond. B*, **343**, 105–111.

Mitter, C, Farrell, B, and Wiegmann, B (1988) The Phylogenetic Study of Adaptive Zones: has Phytophagy Promoted Diversification? *Am. Nat.*, **132**, 107–128.

Nee, S and May, R M (1997) Extinction and the Loss of Evolutionary History, *Science*, **278**, 692–694.

Prendergast, J R, Quinn, R M, Lawton, J H, Eversham, B C, and Gibbons, D W (1993) Rare Species, the Coincidence of Diversity Hotspots and Conservation Strategies, *Nature*, **365**, 335–337.

Rohde, K (1992) Latitudinal Gradients in Species Diversity: the Search for the Primary Cause, *Oikos*, **65**, 514–527.

Rosenzweig, M L (1992) Species Diversity Gradients: We know More and Less than We Thought, *J. Mammal.*, **73**, 715–730.

Southwood, T R E (1973) The Insect/Plant Relationship – an Evolutionary Perspective, *Symp. R. Entomol. Soc. Lond.*, **6**, 3–30.

Strong, D R, Lawton, J H, and Southwood, T R E (1984) *Insects on Plants: Community Patterns and Mechanisms*, Harvard University Press Cambridge, MA.

Vane-Wright, R I, Humphries, C J, and Williams, P H (1991) What to Protect – Systematics and the Agony of Choice, *Biol. Conserv.*, **55**, 235–254.

Composting

see Soil Mineralization (Volume 2)

Conservation Biology

Conservation biology is concerned with the preservation of the environment. Earlier, the main objective was to ensure that mankind would be able to enjoy the best possible lifestyle. However, it has become clear that the present enjoyment of the good life is not necessarily consistent with the long-term stability of the environmental circumstances that would foster human comfort. Economics and the desire of humans for an ever-improving lifestyle have severe environmental consequences, which, without intense study in many scientific disciplines, would lead to massive environmental breakdown.

Examples of the need for biological conservation study and action include all aspects of farming, dam construction for water conservation or power generation, the use of fossil fuels, soil conservation (particularly in relation to farming), urban expansion, the mining and lumber industries, transportation (roadways, exhaust wastes). Apart from the more obvious approaches suggested by these areas of concern, the use of indicator plants that are sensitive to or exceptionally tolerant of specific problems shows great promise for detecting and surveying ecosystems for signs of stress. Virtually all human activity impacts ecosystem stability, often negatively. The effective practice of conservation biology will hopefully mitigate some of the problems, both past and future.

See also: **Metapopulations**, Volume 2.

R G S BIDWELL *Canada*

Conservation Priorities Amongst Habitat Patches

see Metapopulations (Volume 2)

Controlled Environment Facilities in Global Change Research

James F Reynolds
Duke University, Durham, NC, USA

Controlled environment facilities (CEFs) are experimental systems in which biotic and/or abiotic variables and factors

are controlled by the investigator. CEFs play a key role in experimental research on global change by providing researchers with tools to aid in disentangling the complexities of natural ecosystems. While direct observation of ecosystem responses in manipulative field experiments is an important approach to the study of global environmental change, the sheer number of environmental variables that simultaneously affect ecosystems poses serious constraints on such studies. To adequately test null hypotheses – and to successfully tease out a signal (effect) against a background of noise (random variation) – it is necessary to have adequate replication and control, something that is often impossible to do in nature. CEFs offer an alternative – but complementary – approach to study nature.

Typical CEFs for the study of terrestrial ecosystems range from the sophisticated complexes that house multiple plant growth chambers, such as the National Phytotron at Duke University, to the construction of model systems of terrestrial ecosystems, such as in the Ecotron (Figure 1). There are many advantages that make CEFs attractive. They can be used to reproduce different types of environments, ranging from the arctic to the desert to the tropics. Researchers are able to maintain precise control over environmental conditions (e.g., soil type, air temperature, light levels, carbon dioxide (CO_2) concentration, relative humidity, nutrients) and the organisms under study (e.g., mixtures of plant and animal species and their spatial arrangement). Hence, experiments can be designed to explore interactions among multiple stressors, e.g., temperature, carbon dioxide and air pollutants. This provides the opportunity to gain increased understanding of the complex interactions between plants and their environment. Such rigor and control are virtually impossible to achieve under natural field conditions. Other advantages of using CEFs include the ability to study without disturbing the organism, the ease of replication and manipulation, and relatively low research costs.

Limitations of CEFs include their small size and their biological and structural simplicity. No controlled environment experiment can, of course, represent the full complexity of nature. For example, the whole community experiments being conducted at the Ecotron cannot replace field observation and experiments. This is true of even the grandest attempt to date to bottle nature: the immense multi-ecosystem network in Biosphere 2 (Table 1), in which synthetic communities of rainforest, desert, savanna, marsh, ocean and agroforestry are encased in a glass and metal shell. Herein lies a paradox: it is the very complexity of the natural world, where experimental conditions are outside the control of the experimenter, which makes simple model systems so attractive. CEFs provide a model or analog of the real world, both biological and abiotic, which fills the gap between complex field ecosystems and the simplicity of laboratory or greenhouse experiments. CEFs

(a)

(b)

Figure 1 (a) View of tall, walk-in controlled environment growth chambers at the National Phytotron at Duke University. Each chamber can be divided into two independent sub-chambers. (b) Inside an Ecotron chamber, showing model ecosystem. (Courtesy of Center for Population Biology, Imperial College, Silwood Park)

are not intended to mimic the full complexity of nature; rather, they provide models of ecological communities by including many but not all of the biological processes and interactions that occur in the field.

One benchmark of the ability of researchers to understand and predict the behavior of ecological systems is to build their own ecosystems and see if they work as expected. As Lawton (1996) argues, if we cannot understand these simpler communities, how can we expect to understand the full complexity of nature? Because of the diversity of questions and the different levels of biological organization

Table 1 Examples of terrestrial CEFs. Typical abiotic variables controlled include temperature, humidity, irradiance (intensity and quality), day length, gas concentrations (such as carbon dioxide and ozone), water, and nutrients

Facility	Brief description	Web address
Biosphere 2, Tucson, AZ, USA	Covers 1.27 hectares, with synthetic communities of rainforest, desert, savanna, marsh, ocean and agroforestry, all of which are encased in a glass and metal shell. Can be operated in a variety of con gurations including open and closed modes	http://www.bio2.edu/Research/res_entry.htm
Biotron, Swedish University of Agricultural Sciences, Alnarp, Sweden	Thirty- ve chambers, with either natural or arti cial irradiance. Developing a comprehensive experimental database to document performance of different plant species under various controlled conditions; goal is to identify and quantify internal and external factors that limit plant growth	http://www.biotron.slu.se/bio1.htm
Controlled Environment Facility, CSIRO, Brisbane, Australia	Specializes in the growth of tall (up to 8 m) tropical plants, such as sugarcane	http://www.tag.csiro.au/1d_CEF.htm
EcoCELLs, Frits Went Laboratory, Desert Research Institute, Reno, Nevada, USA	EcoCells are sunlit, tempered glass growth chambers (ca. 184 m^3) attached to soil lysimeter; system located inside outer glasshouse; enables studies of material exchange between soil, plants and atmosphere	http://www.maxey.dri.edu/BSC/GBERL/ecocell.html
Ecotron, Imperial College, Silwood Park, UK	Sixteen 5 m^3 replicate terrestrial ecosystems, or microcosms, sustaining up to 40 species of plants and animals (as well as soil microbial organisms) in four trophic (feeding) levels, interacting over several generations	http://forest.bio.ic.ac.uk/cpb/cpb/ecotron.html
Envirotron, University of Georgia, Grif n, Georgia, USA	Nine growth chambers (six 6.7 m^2 and three 3.3 m^2) and eight glasshouses; three movable sunlit chambers (4 m^2) to measure plant CO_2 uptake or loss and water loss	http://www.grif n.peachnet.edu/envirotron.html
National Phytotron at Duke University, Durham, NC, USA	Forty-eight plant growth chambers- over 450 m^2 of growth space (including eight 3.6 m tall chambers, each ca. 15 m^3) - and six CO_2 and temperature-controlled glasshouses (each ca. 200 m^3)	http://www.biology.duke.edu/phytotron/

involved in global change research, a suite of complementary CEFs and methodologies is needed (Table 1). Because of their unique advantages, the use of CEFs in global change research will increase dramatically in the future.

See also: **Biome–BGC Ecosystem Model**, Volume 2.

REFERENCE

Lawton, J H (1996) The Ecotron Facility at Silwood Park: The Value of Big Bottle Experiments, *Ecology*, **77**, 665–669.

FURTHER READING

Fraser, L H and Keddy, P (1997) The Role of Experimental Microcosms in Ecological Research, *Trends Ecol. Evol.*, **12**, 478–481.

Kramer, P J, Hellmers, H, and Downs, R J (1970) SEPEL: New Phytotrons for Environmental Research, *BioScience*, **20**, 1201–1208.

Lawton, J H (1995) Ecological Experiments with Model Systems, *Science*, **269**, 328–331.

Reynolds, J F, Acock, B, and Whitney, R (1993) Linking CO_2 Experiments and Modeling, in *Design and Execution of Experiments on CO_2 Enrichment*, eds E D Schulze and H A Mooney, Ecosystems Research Report, Commission of the European Communities, Brussels, 93–106.

Coral Bleaching (1997-1998)

Barbara E Brown
University of Newcastle, Newcastle-upon-Tyne, UK

During 1997–1998, reports of coral bleaching from all the major tropical oceans suggested that this time period had seen the most geographically widespread bleaching ever recorded, with some areas (e.g., Singapore and the Andaman Islands) witnessing extensive bleaching for the first time in recent history. Coral bleaching has been documented in at least 32 countries and island nations in 1997–1998; with reports from sites in the Pacific, Indian Ocean, Red Sea, Persian Gulf, Mediterranean and Caribbean recording widespread bleaching (Hoegh-Guldberg, 1999). The bleaching response represents a paling of the coral such that it may become entirely white with the white skeleton being reflected through the transparent tissues above. The loss of coloration is due to the release of symbiotic algae and/or their pigments. The symbiotic algae are essential to the health of the coral in that they supply it with carbon (via photosynthesis), sunscreen protection and also accelerated calcification of the skeleton.

Paling of some coral species is an observed seasonal phenomenon in the Pacific, Indian Ocean, and also the Caribbean. Where bleaching is seasonal, or less severe, the likelihood of full recovery of pigmentation is high (Brown *et al.*, 1999). In the 1997–1998 bleaching episode, the response was exceptionally severe with a large number of corals turning completely white and subsequently dying.

Mass bleaching normally occurs when seasonally maximal sea surface temperatures (SST) are exceeded. The likely triggers of bleaching are elevated SST and solar radiation. Research has indicated that these factors act in combination, rather than alone. Additional causes of bleaching, such as extreme low tides and reduced salinities, were also implicated at some sites in 1998. Bleaching was recorded around the world also in 1983, 1987, and 1991 but generally these events were neither as extensive nor as severe as in 1997–1998.

The occurrence of bleaching at specific locations in 1997–1998 was patchy with more severe bleaching recorded in shallow waters than at deeper offshore sites, e.g., bleaching occurred extensively on the inner Great Barrier Reef, Australia but was minimal on the outer reefs (Berkelmans and Oliver, 1999). Not only hard and soft corals, but also sea anemones, zoanthids, giant clams, foraminifera and many other zooxanthellate invertebrates are affected by the loss of their symbiotic algae. Corals can recover from bleaching but death may result if environmental stressors are extreme and/or prolonged. In the Indo-Pacific, fast-growing, branching corals are more susceptible to bleaching than slow-growing boulder corals, leading to a high mortality rate in the former. Recovery of boulder corals was frequently recorded in 1998, often within 1–2 months of initial bleaching. In the Caribbean, however, greater bleaching-related mortality occurred in boulder and plate-like corals than in branching species, which had already suffered extensive mortality from storms, diseases and terrestrial run-off. In the Indo-Pacific, the susceptibility of different corals to bleaching can significantly affect coral community structure and diversity, depress the rate at which the reef builds up, and reduce habitat availability for other reef species. Previous cases of bleaching-induced mortality from 1983 in the Pulau Seribu (Java Sea) and from 1995 in the Similan Islands (eastern Indian Ocean) have provided examples of community change (Brown, 1997). At both sites, shallow parts of the reef were temporarily transformed from being a mixture of branching and boulder corals to areas in which virtually only the boulder corals survived. During the 1998 bleaching, one reef on the Australian Great Barrier Reef was so severely affected that even many of the robust boulder corals (one of them dated as over 700 years of age) were badly damaged or died. Complete

recovery of reefs following severe bleaching is dependent on growth and fragmentation of remaining corals, and on recruitment from stocks in the area. Evidence shows that restoration of the reef to its former state may be slow or, if interrupted by man-made change, may even be halted altogether.

Links have been made between the widespread incidence of coral bleaching in 1997–1998 with one of the strongest El Niños of the 20th century (Stone *et al.*, 1999). While past and present coral bleaching events in many parts of the Pacific appear to be closely matched to El Niño–induced seawater warming, the connections are not clear-cut for all locations in the Pacific. Nor can bleaching in Indian and Atlantic waters be directly linked to only El Niño phenomena. Factors responsible for elevated SSTs in the Indian Ocean, e.g., are likely to be the result of a complex product of El Niño-related, monsoonal and local oceanographic factors which are superimposed on interdecadal patterns in climate variability.

Links between the apparent changing nature and frequency of the El Niño phenomena and global climate change have also been made but are the subject of controversy among climatologists with many suggesting that present patterns reflect the natural variability of the system rather than the effects of greenhouse gas-induced warming. Concerns about the potential effect of global change on future frequencies of severe bleaching events are based on the narrow upper margin of environmental tolerance in corals. While corals display impressive acclimation processes to changes in some environmental parameters, it is not known whether they are able to adapt or acclimatize at rates which match the projected rates of background seawater temperature increase.

The 1997–1998 episode of worldwide bleaching is a major cause for concern. Extensive mortality of corals (80–90%) has been recorded on shallow water reefs throughout the Indian Ocean and although there are signs of recruitment of juvenile corals on some of these reefs, the key question is how frequent and severe will bleaching events be during the next decade? Seawater temperature is rising significantly in parts of the Pacific and Indian Oceans, and also in the Caribbean at a rate of at least 0.13 °C per decade (as measured over the last 50 years). If seawater temperatures were to continue to rise, either as a result of greenhouse gas emissions or natural variability in the ocean/atmosphere system, then we might expect the incidence and severity of coral bleaching to increase yet further with the possibility of substantial changes to the coral reef community structure. It is anticipated that shallow reef flats should be able to keep up with projected sea-level rise over the next few decades but the increased incidence of bleaching may compromise this ability by reducing coral growth rates. The future for corals in a warming world is bleak, with many presently luxuriant coral reefs being transformed into substrates covered by algae (*see* **Coral Reefs: an Ecosystem Subject to Multiple Environmental Threats**, Volume 2).

REFERENCES

Berkelmans, R and Oliver, J K (1999) Large-scale Bleaching of Corals on the Great Barrier Reef, *Coral Reefs*, **18**, 55–60.

Brown, B E (1997) Coral Bleaching: Causes and Consequences, *Coral Reefs*, **16**, S129–S138.

Brown, B E, Dunne, R P, Ambarsari, I, Le Tissier, M D A, and Satapoomin, U (1999) Seasonal Fluctuations in Environmental Factors and their Influence on Photophysiological Parameters in four Indo-Pacific Coral Species from the Andaman Sea, Indian Ocean, *Mar. Ecol. Prog. Ser.*, **191**, 53–69.

Hoegh-Guldberg, O (1999) Climate Change, Coral Bleaching and the Future of the World's Coral Reefs, *Mar. Freshwat. Res.*, **50**, 839–866.

Stone, L, Huppert, A, Rajagopalan, B, Bhasin, H, and Loya, Y (1999) Mass Reef Bleaching: A Recent Outcome of Increased El Niño Activity? *Ecol. Lett.*, **2**, 325–330.

Coral Reefs: an Ecosystem Subject to Multiple Environmental Threats

Jean-Pierre Gattuso[1] and
Robert W Buddemeier[2]

[1]*Laboratoire d'Océanographie de Villefranche, Villefranche-sur-mer, France*
[2]*Kansas Geological Survey, Lawrence, KS, USA*

Although coral reefs may seem exotic to many residents of inland or temperate regions, the ecosystem plays a major role in the environment and economies of many areas. Where reef communities exist, they often play major roles in both defining the environment and providing resources upon which humans depend, both consciously and implicitly (Wilkinson and Buddemeier, 1994). Many developing country populations inhabit the same tropical and subtropical coastal zones that are home to most of the world's reefs. The demise or deterioration of reef communities would therefore have profound social, political, and economic implications for the world, as demonstrated by acute episodes of reef degradation (Wilkinson et al.*, 1999). Costanza* et al. *(1997) have attempted to estimate the economic value of ecosystems and their services; on a global basis, they developed an average value of US$375 $\times$ 10^9 per year for coral reefs*

(out of a total coastal zone value of US$12 568 $\times 10^9$ per year). Although their approach to valuation is not universally accepted, the magnitude of the estimate places reef values in an economic context.

Large-scale and increasingly well documented degradation and die-off of reef ecosystems or key reef organisms has brought them into focus as a prime example of complex ecosystem responses to various combinations of a wide range of environmental stresses. Local destruction can be easily recognized and understood, and the negative effects of some aspects of global change (such as ocean warming and rising carbon dioxide levels) can be either recognized or confidently predicted. However, there remain important symptoms of ecosystem decline, such as the die-off of Acropora in the Caribbean, and the increased frequency and virulence of disease outbreaks, that cannot be directly attributed to a single factor. The situation thus challenges both scientific understanding of the environmental controls on the ecosystem, and management of those factors that we do understand.

The importance of coral reef ecosystems and their responses to environmental change goes beyond the immediate and obvious issues such as coastal protection, tourism, and fisheries. It is widely recognized that reef and related communities are one of the largest storehouses of marine biodiversity, with untapped resources of genetic and biochemical materials, and of scientific knowledge. Further, the living reef communities are not isolated dots on a map, but part of an interacting mosaic of oceanic and terrestrial habitats and communities, and many reef organisms and processes are important in a much wider sphere of related environments. In many important ways, the condition and responses of coral reef communities can be seen as diagnostic of the condition of the world's low-latitude coastal oceans; a truly global signal.

INTRODUCTION

Coral reef ecosystems occur in shallow waters of the tropics and subtropics. They span approximately 600 000 km^2 of coastline of 115 countries or island states. Coral reefs are characterized by a thin veneer of organisms living on top of massive calcium carbonate structures deposited by their predecessors: scleractinian corals living symbiotically with intracellular microscopic unicellular algae (dinoflagellates), coralline algae, and mollusks, among others. Coral reefs are very diverse ecosystems: 93 000 species have been described, which may represent as little as 10% of the total number of reef species (Reaka-Kudla, 1997).

Approximately 350 million people live in areas with a large component of coral reefs, and 450 million within 60 km of a coral reef (Wilkinson, 1996). Coral reefs play a major role in local communities. Their integrated economic and social value, estimated from recent legal actions, ranges from US$1800–2800 m^{-2} (reviewed by Birkeland, 1997), comparable to the flow value of US$6075 ha^{-1} year^{-1} estimated by Costanza *et al.* (1997). They protect the coastline from wave action, and have traditionally provided building material and food resources. Coral reef fisheries presently account for more than 10% of the world fisheries catch. The aesthetic value of reefs and surrounding areas has led to a recent surge in tourism, of considerable economic significance. The tourism industry of the Great Barrier Reef generates US$505 million per year and the contribution to gross national products is proportionally much greater in developing than in developed countries.

Coral reefs provide a most ingenious paradox: they are geologically robust ecosystems built by organisms that appear biologically fragile (Buddemeier and Smith, 1999). Reef structures dominated by scleractinian corals first occurred 200 million years ago and have continued to reoccur episodically since then despite extreme changes of global environmental conditions. During the last glacial–interglacial cycle, tropical sea surface temperature (SST) exhibited changes of 3–6 °C, and sea level changed by more than 120 m due to melting and accumulation of glaciers and ice sheets. Cycles of such changes over geologic time have induced periods of apparent extinction in the geological record, but coral reefs have remained recurrent features in tropical coastal areas.

In spite of their evolutionary survival, there is accumulating evidence that reef organisms are increasingly subject to stresses at the local, regional and global scales. It has been claimed that up to 70% of the world's reef could be lost in four decades (Wilkinson, 1992). One prominent example of stress is the increased frequency of mass *coral- bleaching* events (loss of endosymbionts and/or photodestruction of algal pigments) since the early 1980s. Bleaching is a stress response that does not necessarily lead to mortality. In less severe events, corals often recover their algal population and normal pigment concentration while several severe bleaching events of the last two decades have resulted in considerable mortality. Another disturbing and poorly understood feature of the recent decline in reefs and corals has been the increased frequency, virulence, and variety of lethal epizootic episodes among reef corals (Harvell *et al.*, 1999). Within the past decade there has been a sharp rise in reports of diseases caused by bacterial, fungal, and/or viral agents, especially in the Caribbean and Florida. The large-scale coral loss and community shifts associated with these diseases seem to be unprecedented in Holocene and Late Pleistocene time (e.g., Aronson *et al.*, 1998). Introduction of new pathogens into the marine environment, alteration of environmental conditions that favors the propagation of existing disease-producing agents, and the general debilitation of coral health that makes coral more vulnerable to infection all appear to play a role in this phenomenon. Although the relative importance of these factors is not yet

Table 1 Classi cation matrix for stress-inducing variables: scale (G = global, R = regional, L = local), pathway (P = primary route of stress delivery, S = secondary), interactions with other variables, and relationship to climatic or direct human in uences[a]

	Changing variable	Scale	Atmospheric	Hydrographic	Hydrologic	Interaction with variable #	Climate factor	Human in uence
1	**CO_2 (Ω)**	**G (R)**	**P**	**S**			**Indirect**	**Direct**
2	**Temperature**	**G R**	**P**	**S**		**4, 8**	**Direct**	
3	**Sea level**	**G R L**		**P**		**7, 12**	**Direct**	**Indirect**
4	Light (PAR)	R (G,L)	P	S	S		Indirect	Indirect
5	Ocean currents	R G		P	(S)		Indirect	
6	Storms/waves	R L	P			7, 3	Direct	
7	**UV**	**R G**	**P**	**S**		**2**	**Indirect**	**Indirect**
8	Freshwater	L R		S	P	9, 12, 6	Direct	Direct
9	**Nutrients**	**L R**	**(S)**	**S**	**P**	**11**	**Indirect**	**Direct**
10	Toxics	L R			P			Direct
11	Turbidity	L R		S	P	4, 8, 9, 10, 11		Indirect
12	**Sedimentation**	**L**		**S**	**P**	**9, 11, 13**	**Indirect**	**Direct**
13	**Resource use (physical)**	**L (R)**						**Direct**
14	**Resource use (biological)**	**L R**						**Direct**

[a] The effects of the variables shown in bold typeface are addressed in this contribution. Ω is the calcium carbonate saturation state.

known, the extent and novelty of the epizootic phenomenon clearly implicates significant environmental change at the regional or global scale, and offers little basis to expect that conditions will improve.

Coral reefs are sites of very active metabolic processes. The average gross primary production of the shallow water communities inhabiting the reef flats is 247 mol carbon m^{-2} $year^{-1}$ (Gattuso *et al.*, 1998b). This very high rate often gives the false perception that reefs are highly productive ecosystems. However, the average rate of respiration is only a few percent lower, so net community production is only slightly above zero, which is similar to the oligotrophic ocean waters that surround most reefs (Crossland *et al.*, 1991). Reef communities exhibit the largest average rate of net calcification of all marine ecosystems, with areas of moderate to high coral–algal cover depositing 4–10 kg $CaCO_3$ m^{-2} $year^{-1}$. The major feature of whole-reef ecosystems (which include lagoons and surrounding slopes) is high gross and low net organic carbon production, with high gross and net calcification. Such metabolic features make coral reefs small sources of carbon dioxide to the atmosphere (Gattuso *et al.*, 1999b) and place limits on sustainable fisheries harvests.

The structure, function, and geographic distribution of coral reef communities are controlled, to varying degrees, by temperature, carbon dioxide partial pressure, light, salinity and nutrient concentration (Kleypas *et al.*, 1999b), as well as by a range of natural and human disturbances (Table 1). This paper examines the present and future response of coral reefs to changes of global-scale (temperature, pCO_2, sea level and ultraviolet radiation, UVR) as well as to environmental changes of regional or local significance (e.g., nutrients and human use). Further information, as well as references not specifically attributed in this article, can be found in Smith (1992) and Wood (1998) as well as in the following special issues: *Global Change Biology* (Vol. 2(6); 1996) and *American Zoologist* (Vol. 39(1); 1999).

GLOBAL CHANGES

Temperature

At present, reefs are most widespread and abundant in areas subject to weekly minimum and maximum temperatures of, respectively, 18–31.5 °C, but high and low temperature reefs extend to regions with weekly extreme values ranging from 16.0–34.4 °C (Kleypas *et al.*, 1999b). The reefs experiencing the largest range of temperatures are located in the Northern Persian/Arab Gulf. When average annual temperatures are considered, the high and low temperature environments for coral reefs are approximately delineated by the 29 and 18 °C isotherms, respectively (Figure 1).

Sea surface temperature (SST) is expected to rise 1–3 °C over the coming century, with a non-uniform distribution of increases (Pittock, 1999). The rate of increase will probably be somewhat greater than the warming trend of the past several decades, which has culminated in widespread temperature stress that leads to disruption of the coral–algal symbiosis, termed bleaching.

The term *bleaching* is commonly applied to the loss of some or all of a coral colony's algal symbionts or algal pigments (Glynn, 1996). Although it occurs to some degree in healthy corals in natural settings, it is also a general stress response that can result in mortality. Bleaching is a complex phenomenon that depends on the microenvironment and taxonomy of the zooxanthellae as well as on the taxon and the external environment of the host coral. This phenomenon has been suggested as an adaptive and acclimative mechanism in nature (Buddemeier and Fautin, 1993), but there are clearly limits to this potential, and any role in the recent cycle of bleaching events remains undemonstrated (Hoegh-Guldberg, 1999).

The number of reports of extensive and damaging bleaching events has been increasing since the early 1980s, when a shift in the El Niño Southern Oscillation (ENSO) pattern (*see* **El Niño and La Niña: Causes and Global Consequences**, Volume 1) resulted in greatly increased interannual SST variability in many parts of the world. Many of the most widespread and severe episodes of bleaching have been associated with periods of unusually high temperatures associated with clear, calm weather, resulting in combined temperature–irradiation stress. These are often but not always associated with El Niño conditions, and the most widespread and devastating occurrence was in 1998, when areas of the Indo-Pacific that had not previously experienced major bleaching suffered severe damage (Wilkinson *et al.*, 1999) (*see* **Coral Bleaching (1997–1998)**, Volume 2). While the relationship of El Niño conditions to general global warming is still under discussion, the existence of a global warming trend is well demonstrated, and can be expected to result in increased frequencies and magnitudes of high temperature events.

The increase in mean temperature might be considered a net neutral change, disadvantaging reefs on the high side of the temperature optima, but potentially advantageous to those on the low temperature margins. However, the areas subject to amelioration are much smaller and more marginal than those subject to degradation, and the positive effect of average change is overwhelmed by the negative impacts of increasing frequency and magnitude of high temperature events (Figure 1). At global and regional scales, the question is not "Will it be bad?", but "How bad will it be?" The answer depends on the scenario adopted. If the high frequency SST variation of the past 20 years continues and is superimposed on a general warming trend, the acceleration of bleaching occurrence would result in expanding degradation and near complete destruction in

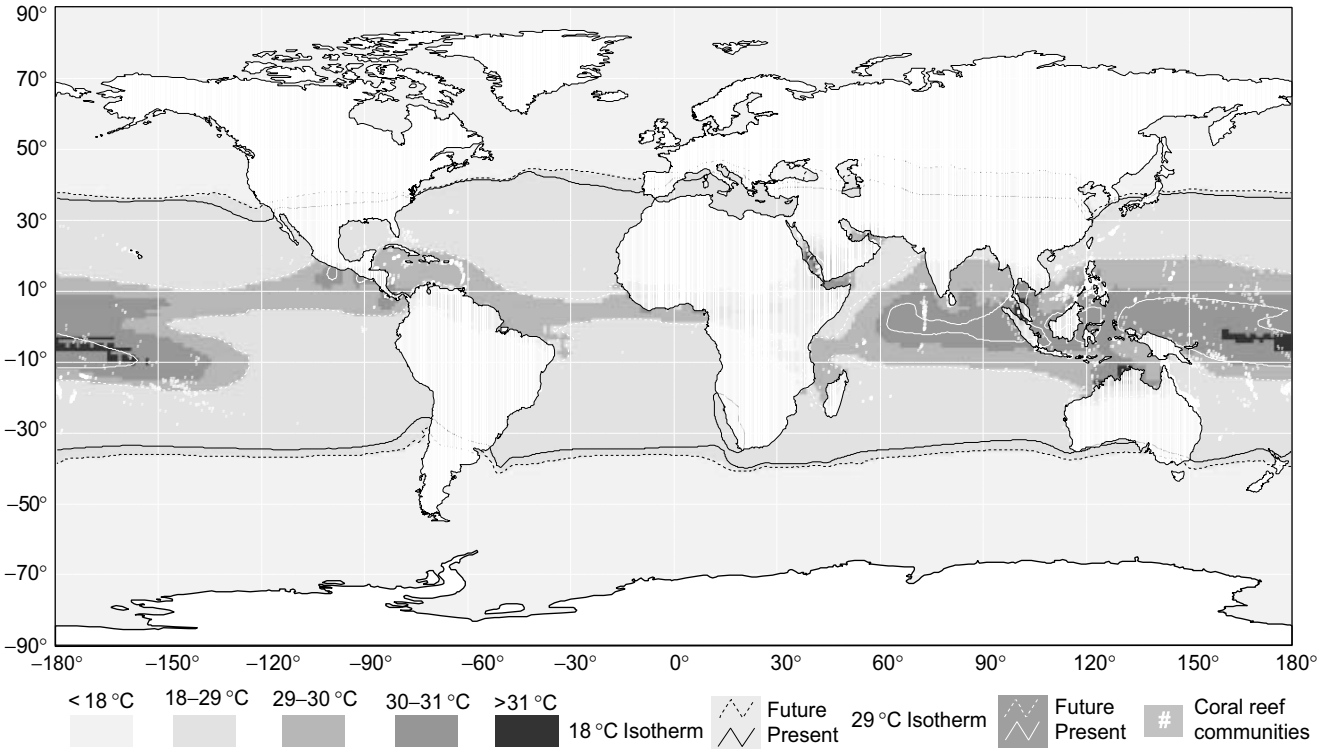


Figure 1 Predicted average annual SST for the years 2040- 2060, with indications of present coral community distribution and of present and future 18 and 29 °C isotherms. Predictions are based on the Canadian Global Coupled Model (CGCM1) with forcing by 1% carbon dioxide increase per year, including the effect of sulfate aerosols (http://www.cccma.bc.ec.gc.ca/data/cgcm1/cgcm1.html). Current SST is from the Reynolds optimal interpolation data set (1982- 1998; http://dss.ucar.edu/datasets/ds277.0/), and reef community locations are from ReefBase (http://www.reefbase.org)

many areas (Hoegh-Guldberg, 1999). Treating the extreme variations as a random phenomenon suggests a possibility of greater intervals between extreme events, but in no case does this translate to a return to previous conditions: the question is whether the next devastating high temperature episode will occur in a few years, or in a few decades.

Frequency and Intensity of Storms and Cyclones

Whether or not global change, and in particular sea surface warming, will result in changes in the locations and intensities of cyclones and tropical storms is an unresolved but significant question. Pittock (1999) reviewed some of the evidence for and potential impacts of such changes. Because coral reef ecosystems are commonly considered disturbance-adapted, changes in storm frequency and intensity would be important, but the detailed effects would be site specific, probabilistic, and difficult to predict. Storm damage is an acute stress that can have critical interactions with long-term chronic stresses such as eutrophication.

Carbon Dioxide Partial Pressure

The marine inorganic carbon system is important to coral reef function because it provides substrates for both photosynthesis and calcification. Total carbon dioxide, or dissolved inorganic carbon (DIC) is the sum of carbon dioxide (aq), HCO_3^-, and CO_3^{2-}. A key factor in the role of coral reefs in the carbon cycle is that photosynthesis is based primarily on the abundant bicarbonate ion, whereas calcification utilizes the relatively scarce carbonate ion.

It is now well established that calcification is controlled by the saturation of seawater with respect to calcium carbonate (Gattuso *et al.*, 1998a; Gattuso *et al.*, 1999a; Langdon *et al.*, 2000; Leclercq *et al.*, 2000):

$$\Omega = \frac{[Ca^{2+}][CO_3^{2-}]}{K'_{sp}} \quad (1)$$

where Ω is the calcium carbonate content and K'_{sp} is the stoichiometric solubility product of the $CaCO_3$ mineral considered (calcite, aragonite or high magnesium calcite). It depends on salinity, pressure and temperature. Since Ca^{2+}

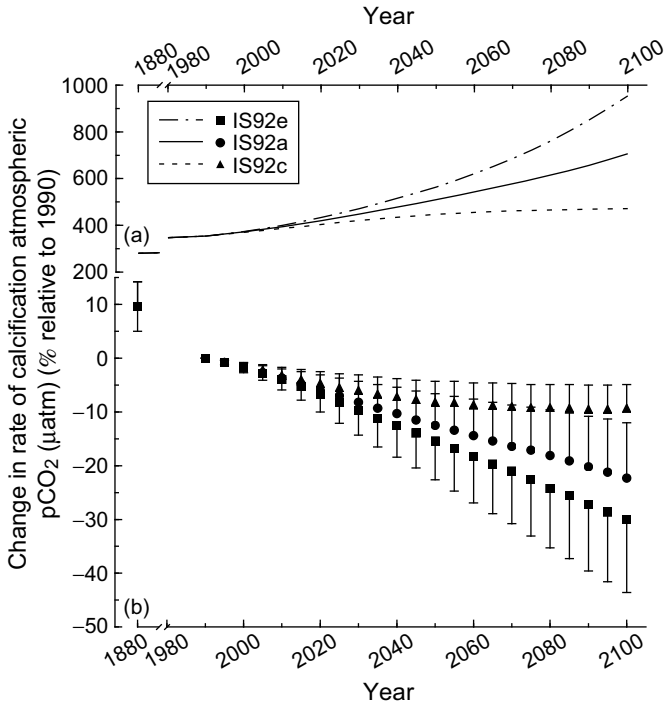


Figure 2 (a) Past and future changes in atmospheric pCO_2. The following scenarios of the Intergovernmental Panel on Climate Change (updated IS92 scenarios; Houghton *et al.*, 1996) are shown: high (IS92e), mid (IS92a) and low (IS92c). (b) Past and future changes in rate of $CaCO_3$ deposition (relative to 1990) of various marine photosynthetic organisms and communities. Mean ± standard error of the mean. (Reproduced by permission of Society for Integrative and Comparative Biology, from Gattuso *et al.*, 1999a)

is about 100 times that of CO_3^{2-} and is a near conservative element of seawater over 10^4 years, Ω is largely determined by CO_3^{2-}. If $\Omega < 1$, undersaturation exists and is favorable to dissolution of $CaCO_3$. A value higher than one indicates supersaturation and is favorable to calcification. Present-day coral reefs occur within an area of the ocean bounded by average $\Omega_{arag} = 3.3–4.1$ (Kleypas *et al.*, 1999b).

Seawater pCO_2 is the most important parameter controlling Ω. The surface ocean tends to equilibrate with the atmosphere; increasing carbon dioxide combines with carbonate ions leading to a decrease in the concentration of CO_3^{2-}, one of the building block of calcium carbonate:

$$CO_2 + CO_3^{2-} + H_2O \longrightarrow 2HCO_3^- \qquad (2)$$

Carbon dioxide partial pressure has increased from 280–368 µatm since the end of the 18th century as a result of human activities (Figure 2). Its present rate of increase is approximately 1.5 µatm $year^{-1}$ (0.4% $year^{-1}$) and, according the average scenario of the Intergovernmental Panel on Climate Change (IPCC; Houghton *et al.*, 1996), pCO_2 could reach 705 µatm in 2100. That would represent a 2.5 fold increase compared to its pre-industrial value. This increase is thought to be beneficial at least in some terrestrial ecosystems that may respond with increased primary production. Very little information is available on the response of reef primary production to increased pCO_2. The first experiments, carried out in a mesocosm, suggest that primary production does not change under elevated pCO_2 (Leclercq *et al.*, 2000). However, the effects on calcification as a result of CO_3^{2-} change are more significant and predictable. The average Ω for the surface ocean between 30 °S–30 °N, was 4.6 in 1880, is currently 4.0, and is projected to drop to 3.1 and 2.8, respectively by the years 2065 and 2100 (Kleypas *et al.*, 1999a). Using data on the effect of Ω on the calcification of temperate and tropical calcifying organisms, Gattuso *et al.* (1999a) predicted the resulting decrease in calcification from mid, low, and high carbon dioxide emission scenarios (Figure 2). It was assumed that the surface seawater and the atmosphere are in equilibrium with respect to carbon dioxide and that the increase in seawater temperature will be identical to the predicted increase in atmospheric temperature. These calculations suggest that $CaCO_3$ deposition by photosynthetic and calcifying marine organisms and communities may have been 10% higher in 1880 than in 1990, and may decrease further by 9–30% (mid estimate: 22%) between 1990 and 2100 (Figure 2).

On the basis of a geographical model, Kleypas *et al.* (1999a) also concluded that the total pre-industrial to 2100 calcification decrease could be as high as 17–35%. They suggested that the reefs likely to experience the largest

absolute decrease in calcification are those presently in high saturation conditions, in the Red Sea, West Central Pacific, and Caribbean. However, the most significant impacts may be due to high relative changes in coral reefs that presently exhibit low rates of calcification and an almost balanced $CaCO_3$ budget (net carbonate loss close to gross calcification). These include high latitude reefs (Bermuda), reefs in upwelling regions (Gálapagos), and reefs subject to intense anthropogenic stresses. Reduced calcification rate is a major impact on the overall fitness of organisms and communities subject to other stresses. Indeed, a decreased rate of calcification may lead to reduced skeletal density and increased vulnerability to physical damage and bioerosion, to reduced skeletal extension that would impair the ability of calcifying organisms to compete for space, or to both.

Sea Level

Because of their dependence on photosynthesis, reef-building corals and coralline algae occur only within a few tens of meters of the ocean surface, yet they are marine organisms tolerating only occasional aerial exposure on low tides. Sea level, and the rate of sea level change, is thus an important control on reef occurrence and growth. However, the protracted stable high-stand of the Late Holocene that humans consider normal is atypical in recent geologic history; the average sea level condition of the Quaternary (the past 1.8 million years) is lower and more rapidly changing than during the past 5000–6000 years.

Eustatic sea level is expected to rise by 10–90 cm during the 21st century, like temperature, a rate similar to but somewhat greater than that observed in the recent past. However, the direct effects, assessed at a global scale, are likely to be somewhat positive for near sea level reefs, many of which have developed large areas that are sea level limited due to the protracted Holocene still-stand. Sea level rise may open up reef flats and other areas for colonization, in locales where there is a ready source of propagules. Increasing sea level will be a disadvantage for extremely marginal deeper reefs; those near the bottom of the depth range of viability, and calcifying or accumulating very slowly.

Locally, sea level rise may have indirect negative effects if increasing coastal erosion mobilizes sediments that are subsequently deposited on nearshore reef areas or coral communities. Issues relating to erosion and sedimentation interact strongly with land use and management practices, and therefore depend on human activities. The increased inundation that may benefit shallow reef communities is likely to be problematic for coastal human communities inshore from reefs, since it will reduce the physical protection afforded by the reefs, resulting in increased coastal erosion and storm damage in addition to the inundation due to sea level rise.

Ultraviolet Radiation (UVR)

More UVR, mostly ultraviolet-B (UVB) in the 300–320 nm waveband, penetrates seawater in the tropics than at higher latitudes because the stratospheric ozone layer is generally thinner (*see* **Ultraviolet Radiation**, Volume 1) and seawater contains less UV-absorbing materials, including phytoplankton. UVR can penetrate to depths of 20 m or more (see Shick *et al.*, 1996). The decrease of stratospheric ozone resulting from anthropogenic inputs of chlorinated fluorocarbons increases the level of UVR reaching the Earth's surface (Madronich *et al.*, 1995). The downward trend of stratospheric ozone observed over the tropics during the period 1979–1994 (2% per decade) is within the background variability. However, the downward trend is significant at slightly higher latitudes (20–30 °N), where subtropical corals exist. The forecast increase in UVR during the next decades, before a projected recovery period during the second half of this century, has led to concerns about the response of coral reef organisms.

Reef organisms have developed a suite of powerful mechanisms and behavior to protect themselves from the deleterious effects of UVR. They produce water soluble compounds, mycosporine-like amino acids (MAAs), that absorb radiation in the UVR waveband, with peak absorption maxima ranging from 309 to 360 nm (Dunlap and Shick, 1998). Corals may contain as many as 10 different MAAs that act as a broadband UVB photoprotective filter. The concentration of MAAs is higher in shallow than in deep specimens and is also higher in exposed versus shaded portions of the same colony. In addition to direct damaging effects, UVR also contributes to the generation of reactive oxygen species that can induce photooxidative stresses (damage of deoxyribonucleic acid (DNA), enzymes, membrane proteins and lipids, and photosystem components). However, algal–invertebrate symbioses possess antioxidant defense mechanisms and the extent of photooxidative stresses in natural conditions is not clearly understood.

Laboratory manipulations and field transplantation experiments demonstrate that although an increase in ambient UVB level stimulates the accumulation of MAAs (e.g., Shick *et al.*, 1999), environmentally relevant increases in UVR have several damaging effects on coral reef organisms (Shick *et al.*, 1996). Exposure to UVR can be fatal both for adult coral colonies and larvae. Natural levels of UVR depress coral skeletal growth as well as photosynthesis of intact corals and isolated zooxanthellae. They are also suspected of interacting with increased temperature and lead to coral bleaching, particularly during doldrum conditions when seawater is especially clear. A recent review concludes that the small increase in UVR will have sublethal effects but that the impact on the distribution of corals and coral reefs will be limited (Shick *et al.*, 1996). The major effect might be a lowering of the minimum depth of occurrence of reef organisms.

REGIONAL AND LOCAL CHANGES

Nutrients

Coral reefs are typically distributed in oligotrophic waters, with undetectable modal values of nitrate and phosphate (Kleypas *et al.*, 1999b), although reefs may tolerate relatively high nutrient conditions (e.g., in the Galápagos Islands). Their net nutrient requirements are low because net ecosystem production is low and similar to that of the nearby oligotrophic ocean (Crossland *et al.*, 1991). Several mechanisms enable coral reef ecosystems to thrive in such conditions (Dubinsky and Stambler, 1996).

1. Many reef invertebrates harbor endosymbiotic microalgae (e.g., corals and mollusks), these mutualistic associations have developed very efficient mechanisms for nutrient conservation and recycling.
2. Elemental C : N : P ratios of reef organisms are generally much higher than in their oceanic counterparts, enabling a higher photosynthetic carbon dioxide fixation per unit of nutrient.
3. Nitrogen limitation is overcome by the capacity of several reef organisms to fix inorganic nitrogen.
4. Very large volumes of oceanic water are advected across many reef flats. Although typically depleted in nitrogen relative to phosphorus, the water is a low concentration, high volume source of phosphorus to the system.
5. Advected planktonic organisms can also represent a significant source of organically fixed nutrients.
6. For some reef locations and processes, sediment-derived nutrient fluxes may be significant. These can include recycling of endogenously fixed nutrients, or geothermal convection of nutrient-rich deep water through the carbonate matrix.

With the probable exception of oceanic atolls, coral reefs are increasingly subject to elevated levels of nutrients due to land management and agricultural practices (use of fertilizers and forest clearing) as well as increased coastal development and general human modification of the global nutrient cycles. The increased nutrient load is delivered through river and groundwater inputs, terrestrial runoff and sewage discharge. Detection of increased nutrient concentration at the regional scale, such as the Great Barrier Reef, has been the source of considerable controversy. The increase at a local scale is nevertheless detectable and significant. High levels of eutrophication can have dramatic direct effects on coral physiology. It decreases skeletal growth and stimulates the growth of symbiotic algae, which retain a larger portion of photosynthates and transfer less to the animal host. More significant in most cases are the indirect consequences of increased nutrient levels, which greatly favor the growth of macroalgae that become more efficient than corals in the competition for space. Eutrophication also increases the phytoplankton standing stock, which increases turbidity and can lead to light limitation of photosynthetic organisms. The decrease in the rate of net carbonate mineral accumulation associated with the stimulation of the activity of bio-eroders also threatens the delicate balance between calcium carbonate precipitation and erosion.

Resource Use

A major factor in the past deterioration and future prospects of reefs is human resource use. This may include various kinds of direct use of the reefs or their products:

1. Overharvesting, whether for food or for sale of organisms in the souvenir or aquarium trades, can easily exceed the limited sustainable harvest capacity of the ecosystem, and more importantly can disrupt ecosystem function by removing key classes of organisms, such as large herbivores (Hughes, 1994).
2. Destructive fishing methods, such as the use of explosives or cyanide on shallow reefs or bottom trawling in areas of deep coral communities, can result in long-term loss of habitat and organisms, far in excess of the actual harvest.
3. Use of the reef materials or area for destructive non-reef development, such as dredging, mining of construction materials, or landfill (e.g., to build airports).
4. Even apparently non-consumptive uses such as tourism can have significant concomitant stresses; high visitor loads can result in damage by boats and divers, incidental contamination, and eutrophication from local tourist development, etc.

Indirect human damage is somewhat more difficult to deal with, since it typically results from activities undertaken for reasons that do not involve consideration of reef effects. Such factors include:

1. Land use and management, even at substantial distances from the coast. Deforestation and agricultural activities in many areas have greatly increased the loads of sediment and nutrients reaching coastal areas. The resultant turbidity, sedimentation, and eutrophication is a major factor in local reef stress.
2. Discharge of waste, including sewage (which contributes to eutrophication) and toxic chemicals.
3. Alteration of the hydrologic cycle (e.g., by urbanization) in ways that increase the likelihood of nearshore salinity stress.

With growing coastal populations and development pressure, these effects are among the most urgently threatening to reefs in spite of the fact that they are the most readily identified and controlled.

CONCLUSION AND PERSPECTIVE

Local direct anthropogenic stresses are presently those that have the greatest short-term damaging impact on coral reefs and must be dealt with urgently. The effects of global change are, by definition, more widespread and longer-term. However, it is clear that both are significant threats to the health and survival of coral reef ecosystems, and will synergistically interact. In general, the reefs at greatest risk are those closest to human populations and major land masses (Smith and Buddemeier, 1992), but the predicted changes in ocean temperature and carbon dioxide levels mean that even the most remote and well situated reefs will experience a significantly deteriorating environment. High levels of concern and research and conservation efforts will be of little value unless the problems of overpopulation and unsustainable development are seriously addressed.

ACKNOWLEDGMENTS

Thanks are due to J A Kleypas for stimulating interaction, to J A Kleypas and J M Shick for reviewing an earlier draft of this paper, and to J Bartley for his invaluable help with producing Figure 1.

REFERENCES

Aronson, R B, Precht, W F, and Macintyre, I G (1998) Extrinsic Control of Species Replacement on a Holocene Reef in Belize: The Role of Coral Disease, *Coral Reefs*, **17**, 223–230.

Birkeland, C (1997) Introduction, in *Life and Death of Coral Reefs*, ed C Birkeland, Chapman and Hall, New York, 1–12.

Buddemeier, R W and Fautin, D G (1993) Coral Bleaching as an Adaptive Mechanism – A Testable Hypothesis, *BioScience*, **43**, 320–326.

Buddemeier, R W and Smith, S V (1999) Coral Adaptation and Acclimatization: a Most Ingenious Paradox, *Am. Zool.*, **39**, 1–9.

Costanza, R, d'Arge, R, de Groot, R, Farber, S, Grasso, M, Hannon, B, Limburg, K, Naeem, S, O'Neill, R V, Paruelo, J, Raskin, R G, Sutton, P, and van den Belt, M (1997) The Value of the World's Ecosystem Services and Natural Capital, *Nature*, **387**, 253–259.

Crossland, C J, Hatcher, B G, and Smith, S V (1991) Role of Coral Reefs in Global Ocean Production, *Coral Reefs*, **10**, 55–64.

Dubinsky, Z and Stambler, N (1996) Marine Pollution and Coral Reefs, *Global Change Biol.*, **2**, 511–526.

Dunlap, W C and Shick, J M (1998) Ultraviolet Radiation Absorbing Mycosporine-like Amino Acids in Coral Reef Organisms: a Biochemical and Environmental Perspective, *J. Phycol.*, **34**, 418–430.

Gattuso, J-P, Allemand, D, and Frankignoulle, M (1999a) Photosynthesis and Calcification at Cellular, Organismal and Community Levels in Coral Reefs: A Review on Interactions and Control by Carbonate Chemistry, *Am. Zool.*, **39**, 160–183.

Gattuso, J-P, Frankignoulle, M, Bourge, I, Romaine, S, and Buddemeier, R W (1998a) Effect of Calcium Carbonate Saturation of Seawater on Coral Calcification, *Global Planet. Change*, **18**, 37–46.

Gattuso, J-P, Frankignoulle, M, and Smith, S V (1999b) Measurement of Community Metabolism and Significance of Coral Reefs in the CO_2 Source–Sink Debate, *Proc. Natl. Acad. Sci. USA*, **96**, 13 017–13 022.

Gattuso, J-P, Frankignoulle, M, and Wollast, R (1998b) Carbon and Carbonate Metabolism in Coastal Aquatic Ecosystems, *Annu. Rev. Ecol. Syst.*, **29**, 405–434.

Glynn, P W (1996) Coral Reef Bleaching: Facts, Hypotheses and Implications *Global Change Biol.*, **2**, 495–509.

Harvell, C D, Kim, K, Burkholder, J M, Colwell, R R, Epstein, P R, Grimes, D J, Hofmann, E E, Lipp, E K, Osterhaus, A D M E, Overstreet, R M, Porter, J W, Smith, G W, and Vasta, G R (1999) Emerging Marine Diseases – Climate Links and Anthropogenic Factors, *Science*, **285**, 1505–1510.

Hoegh-Guldberg, O (1999) Climate Change, Coral Bleaching and the Future of the World's Coral Reefs, Greenpeace Australia, Sydney.

Houghton, J T, Meira Filho, L G, Callander, B A, Harris, N, Kattenberg, A, and Maskell, K (1996) *Climate Change 1995: The Science of Climate Change*, Cambridge University Press, Cambridge.

Hughes, T P (1994) Catastrophes, Phase Shifts, and Large-scale Degradation of a Caribbean Coral Reef, *Science*, **265**, 1547–1551.

Kleypas, J A, Buddemeier, R W, Archer, D, Gattuso, J-P, Langdon, C, and Opdyke, B N (1999a) Geochemical Consequences of Increased Atmospheric CO_2 on Coral Reefs, *Science*, **284**, 118–120.

Kleypas, J A, McManus, J W, and Meñez, L A B (1999b) Environmental Limits to Coral Reef Development: Where do we Draw the Line? *Am. Zool.*, **39**, 146–159.

Langdon, C, Takahashi, T, Marubini, F, Atkinson, M, Sweeney, C, Aceves, H, Barnett, H, Chipman, D, and Goddard, J (2000) Effect of Calcium Carbonate Saturation State on the Rate of Calcification of an Experimental Coral Reef, *Global Biogeochem. Cycles*, **14**, 639–654.

Leclercq, N, Gattuso, J-P, and Jaubert, J (2000) CO_2 Partial Pressure Controls the Calcification Rate of a Coral Community, *Global Change Biol.*, **6**, 329–334.

Madronich, S, McKenzie, R L, Caldwell, M M, and Björn, L O (1995) Changes in Ultraviolet Radiation Reaching the Earth's Surface, *Ambio*, **24**, 143–152.

Pittock, A B (1999) Coral Reefs and Environmental Change: Adaptation to what? *Am. Zool.*, **39**, 10–29.

Reaka-Kudla, M L (1997) The Global Biodiversity of Coral Reefs: a Comparison with Rain Forests, in *Biodiversity II: Understanding and Protecting our Biological Resources*, eds M L Reaka-Kudla, D E Wilson, and E O Wilson, Joseph Henry Press, Washington, DC, 83–108.

Shick, J M, Lesser, M P, and Jokiel, P L (1996) Effects of Ultraviolet Radiation on Corals and Other Coral Reef Organisms, *Global Change Biol.*, **2**, 527–545.

Shick, J M, Romaine-Lioud, S, Ferrier-Pagès, C, and Gattuso, J-P (1999) Ultraviolet-B Radiation Stimulates Shikimate Pathway-dependent Accumulation of Mycosporine-like Amino

Acids in the Coral Stylophora pistillata Despite Decreases in its Population of Symbiotic Dinoflagellates *Limnol. Oceanogr.*, **44**, 1667–1682.
Smith, S V and Buddemeier, R W (1992) Global Change and Coral Reef Ecosystems, *Annu. Rev. Ecol. Syst.*, **23**, 89–118.
Wilkinson, C, Linden, O, Cesar, H, Hodgson, G, Rubens, J, and Strong, A E (1999) Ecological and Socioeconomic Impacts of 1998 Coral Mortality in the Indian Ocean: an ENSO Impact and a Warning of Future Change? *Ambio*, **28**, 188–196.
Wilkinson, C R (1992) *Coral Reefs are Facing Widespread Devastation: Can we Prevent this Through Sustainable Management Practices?* Proceedings of the 7th International Coral Reef Symposium, Guam, 11–21.
Wilkinson, C R (1996) Global Change and Coral Reefs: Impacts on Reefs, Economies and Human Cultures, *Global Change Biol.*, **2**, 547–558.
Wilkinson, C R and Buddemeier, R W (1994) *Global Climate Change and Coral Reefs: Implications for People and Reefs, Report of the UNEP–IOC–ASPEI–IUCN Global Task Team on the Implications of Climate Change on Coral Reefs*, IUCN, Gland, Switzerland.
Wilkinson, C, Linden, O, Cesar, H, Hodgson, G, Rubens, J, and Strong, A E (1999) Ecological and Socioeconomic Impacts of 1998 Coral Mortality in the Indian Ocean: an ENSO Impact and a Warning of Future Change? *Ambio*, **28**, 188–196.
Wood, R (1998) The Ecological Evolution of Reefs, *Annu. Rev. Ecol. Syst.*, **29**, 179–206.

Cretaceous-Tertiary (KT) Extinction Event

see Extinctions and Biodiversity in the Fossil Record (Volume 2)

Crop Models

J Goudriaan
Wageningen University, Wageningen, The Netherlands

The total agricultural production in the second half of the 20th century more than doubled. This increase would not have been possible without strong improvement of farming techniques. In particular, increased application of fertilizers has been essential. At the end of the 20th century, the large-scale application of the industrial process of nitrogen fixation (Smil, 1997) has exceeded the total rate of natural nitrogen fixation. There is no reason why the output of the industrial nitrogen fixation process could not go up further, but other constraints to crop production are beginning to be felt. In particular, in South-East Asia the cultivation of rice is already at a level at which further intensification will be difficult. Therefore, it is questionable whether the fast rate of increase of food production can be maintained during the course of the 21st century to the extent that it can keep pace with the growth of the human population. Global climate change could be decisive to the balance between supply and demand, dependent on where the effects will be beneficial and where they will turn out to be negative to agricultural production.

Crop modeling serves as a powerful intellectual tool to study the importance of different constraints to crop growth. Its results enable us to make prognoses and also it helps to tell which changes in farming techniques could be most effective in a warmer climate.

HISTORY OF CROP MODELING

The classic report of De Wit (1965) on photosynthesis of leaf canopies described a numerical model (see Box 1) to calculate daily photosynthesis of a leaf canopy from leaf characteristics and meteorological conditions. In modern leaf-based crop growth simulation models, basically the same approach is still followed (see Table 6.2 in Walker *et al.*, 1999).

Bouman *et al.* (1996) give a historical overview of how a whole suite of models (see Box 2) evolved from this basic concept, most notably the Simple and Universal CROp growth Simulator (SUCROS) group of crop growth simulation models, including ORYZA for rice. Other groups of models such as the CERES models (Ritchie *et al.*, 1991), SOYbean GROwth simulation model (SOYGRO) models (Boote *et al.*, 1996), AFRC–Wheat model (Porter, 1993) were developed independently.

An important simplification to the leaf-based models was the emergence of the whole-canopy light capture models, pioneered by Monteith and his coworkers (Gallagher and

Box 1 What is a computer simulation model?

Nowadays, physical models are rarely used except for demonstration purposes, and by the term *model* we usually mean a computer model. Computer models are extremely exible as they permit us to formulate almost any relationship between input and output. These relationships are quantitative and may consist of numerous smaller submodels. The formulation of the relationships between input and output may be just empirical, but it may also be based on underlying mechanisms, with the objective of making the model more general. The specifying term *simulation* is added when the model is not just a simple input- output mapping, but when underlying processes are crucial for understanding the model behavior.

Box 2 What is a crop model and why do we use it?

Any prediction about the effect of some farming action (for instance of nitrogen fertilization) on crop yield involves the application of a rule, no matter how simple. The rule for such a prediction is based on experience, and this may well be suf cient for normal circumstances. In a crop model, much more experience is included than what would be needed for just the local situation. A model is a generalized but also simpli ed representation of a real world crop and its environment. The information in such a model is normally quantitative and it is processed mechanistically. A model makes it possible to include information from scienti c sources that would otherwise not of been used.

Biscoe, 1978; Monteith *et al.*, 1994). In these models there is no consideration of individual leaves, or of the diurnal fluctuation of radiation, and the daily radiation intercepted by a plant canopy is immediately converted into a net daily growth of dry matter. The success of this bold move has been remarkable, although the value of the radiation conversion factor remains a subject of numerous studies.

Some of the initial simplicity is lost when provisions are made for the effect of temperature, of radiation level, of ageing, of drought, and of many other factors on the value of the radiation conversion factor.

The model SIMRIW (Horie, 1987, 1993) is an example of such a model applied to the growth of rice.

A model makes it possible to include information from scientific sources that would otherwise not have been used.

LIMITATIONS OF CROP GROWTH MODELS

Crop models embody a number of simplifications (see Box 3). Weeds, diseases, and insect pests are often assumed to be controlled; soil conditions are assumed to be optimal; and extreme weather events such as tornadoes are supposed not to occur. However, even in normal situations, crops usually suffer from shortages of water or nutrients (see Box 4; Figure 1). Weeds that capture resources in competition with the crop exacerbate the effect of such shortages. Also, the effects of diseases and pests reduce the crop yields.

Box 3 Simpli cation: a weakness and a strength

A model is a simpli ed representation of reality (De Wit, 1970). A model will therefore fail to capture all the richness of details of the real world. This absence of details is not a weakness, but a strength, as it makes it possible to see a fundamental structure that would otherwise be blurred. There is, however, a subjectivity in the choice of which details are neglected. This subjectivity is characteristic for model building, and it can only be partly overcome by subjecting the model to a range of rigorous tests on model performance.

Box 4 Production situations and limiting factors

Potential yield is the yield of a crop when the crop has been free from pests and diseases and when it has always had an optimal root environment so that there have been no shortages of water or nutrients (Rabbinge, 1993).

Attainable yield is lower than the potential yield, because of limitations due to a shortage of water or nutrients, but not due to pests, diseases or weeds.

Actual yield also includes the effects of weeds, pests and diseases.

For the most important crops of temperate regions, such as wheat and potato, the models have been used and tested extensively, in contrast to tropical crops such as millet and cassava. Estimates of potential yields of such crops, which may respond differently to both climate change and increases in carbon dioxide (CO_2), are needed for better assessment of climate change impacts in tropical and subtropical regions.

There are contrasting reports on the performance of crop growth models in comparison to observed yields. In experimental situations, where many factors are accurately monitored, the models usually do better than for larger scale applications at regional levels. De Koning *et al.* (1993) found that crop growth simulations did not significantly improve the predictions at the regional level in the European Community. The most likely explanation is that too many other factors, unknown to the modelers, masked the observed results. On the other hand, crop growth models do reasonably well on the regional scale (Casanova, 1998), and according to Supit (2000), a major difficulty is to know the actual cropped area for a certain type of crop such as winter wheat.

IN WHAT WAY DO CROPS REACT TO CLIMATE? (SEE BOX 5)

When it is warmer, a plant develops faster; that means that it goes through its normal life cycle faster. The most common expression of this speeding up is an acceleration of flowering, but other reactions such as increased rate of leaf production or stem elongation can also be noticed. In models, this reaction is often represented by the concept of minimal heat sum requirement to reach a certain stage of development. *Heat sum* (*degree days*) is calculated by daily addition of the number of °C by which the average daily temperature exceeds a certain threshold temperature (also called the base temperature).

For instance, to reach flowering in wheat, a heat sum of about 800 degree days, with a base temperature of 0 °C is needed. For other crops, this requirement can be very different. Soybean only begins to develop above a temperature of about 10 °C, and so it is not grown in cool temperate climates such as in northwestern Europe.

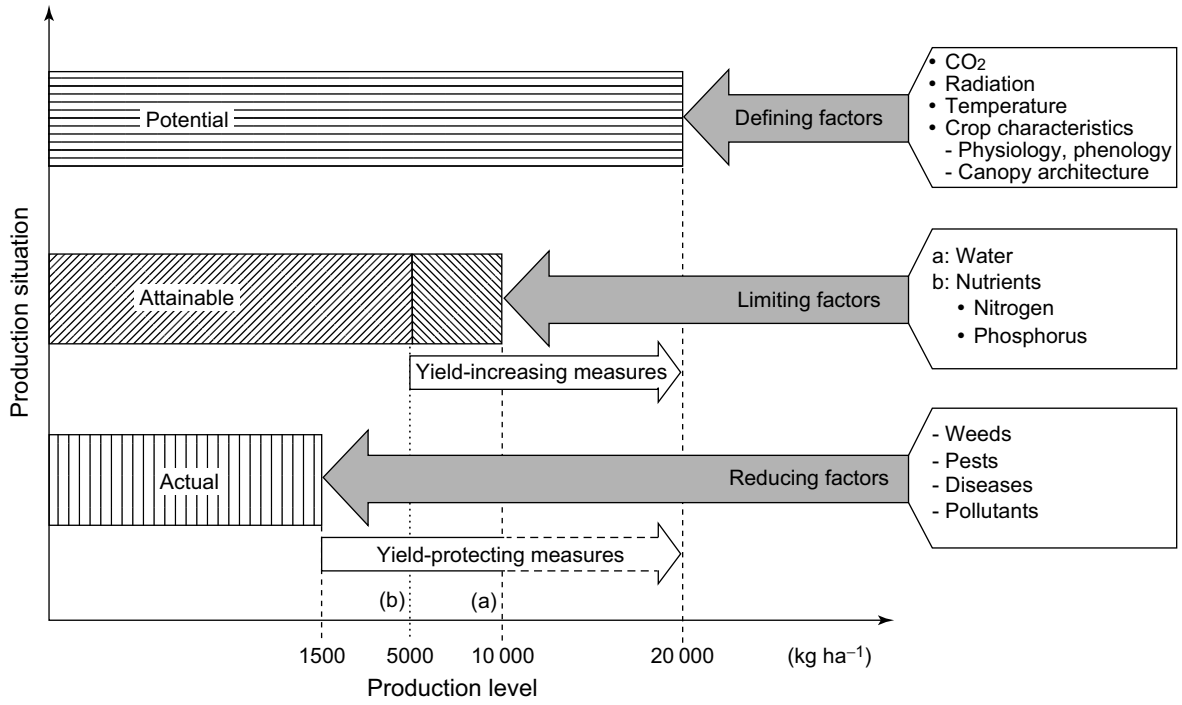


Figure 1 Production situations and their corresponding yield levels. (Reproduced by permission of John Wiley & Sons in Rabbinge, 1993)

Box 5 Crops and their environment

Environmental constraints are comprised of weather, soil, pests and diseases. Soil conditions are important, and encompass soil type and composition. Water and nutrient content may vary strongly during the growing period of the crop. This variation may either be given as input or be calculated within the model itself.

The climatic environment of the crop determines photosynthesis, transpiration and other processes that eventually lead to owering and fruiting. Within any one season, the weather will vary strongly, even from hour to hour. A mechanistic crop growth model allows for these variations in environmental in uences.

How the crop reacts to the environment depends strongly on the type of crop. Therefore, there are different models for different crop types such as wheat, rice, maize, potato, and soybean. A certain crop type is characterized by its genetic coef cients, which tell the model how to calculate the rate of crop development, crop growth, and formation of new organs.

Once the climate of a region is known, its climatic suitability for any given crop type can be calculated, expressed as the potential yield of that crop type. Mapping the boundaries of suitability classes is called agro-ecological zonation.

When transferred to a warmer climate, a crop goes through its lifecycle faster, which means that there will be less time for the formation of harvestable organs. The yield is then smaller. However, crop types are not the same in different agro-ecological zones, as they have been selected for optimal performance in their local climatic conditions.

Development and growth in size are different things. In some conditions, a plant species may be very small and already flower, but in other conditions it may grow in size and remain vegetative. The explanation for such large differences in rate of development is often day length. Day-length sensitivity is a genetic trait that can be changed relatively easily by breeding, and therefore different varieties of one species may have very different day-length sensitivities. This difference in sensitivity explains why the success of a variety is so much linked to a geographical region. Obviously, crop models need to have a correct representation of the day-length sensitivity of the variety for which they are used.

Another important process is leaf area formation. Over the life cycle of a crop, the leaf area grows, reaches a maximum, and then usually declines as the crop senesces. The total amount of radiation intercepted depends on the time integral of incoming radiation multiplied by the fraction of radiation that the leaves can intercept. The fraction of interception is directly linked to the *leaf area index*, the ratio of leaf area and ground area. As a rule of thumb, interception is complete (more than 90%) when the leaf area index exceeds a value of 3.

In simple crop growth models, the daily amount of radiation intercepted is directly converted into a daily growth rate of the crop. The conversion factor is rather stable, of the order of 1 g of dry matter MJ^{-1} of total radiation intercepted, although some variation exists with plant type.

APPLICATIONS TO CLIMATIC CHANGE

Sinclair and Seligman (1996) have given an excellent account of the historical development of crop growth modeling. Since the 1970s, there has been a gradual evolution in studies on the impact of climate change on crop growth from statistical approaches towards crop modeling approaches. The tendency at first was to rely on extrapolations of statistical relationships, with little understanding of the processes that relate the cause to the supposed effects (quote from Parry, 1990). The National Defense University in Washington, DC (1980) conducted such a statistical study in view of disastrous effects of a potential future persistent drought in the major grain-growing regions of the USA and the Union of the Soviet Socialist Republics. There was, however, a growing awareness of the multitude of influencing factors, which makes it difficult to extract significant statistical relationships. The statistical approach was successful in provoking crop growth modelers to come up with explanations as to why some climatic effects predicted by the crop models could not be found in the real world. On the other hand, some strong effects are confirmed, such as the negative effect of higher temperature, which is largely due to accelerated development of the crop, thus shortening the growing period.

The most robust results of crop modeling were not in yield prognoses, but in *agroclimatic zonation* work (see Box 6). The predicted poleward shifts of the limits of crop growing zones are based on rather simple relationships between crop development and heat requirement (Parry, 1990).

MODEL COMPARISON EXERCISES

Each modeling group has worked on the development of its own crop growth model. This is understandable (Sinclair and Seligman, 1996), but it has led to a confusingly large number of crop models, in varying degrees of complexity, robustness, and maturity. This situation called for model comparison exercises that have taken place a few times since the Lunteren meeting in 1993 (Goudriaan, 1996). The Lunteren meeting showed a surprising divergence in the results of wheat models, most likely due to the large environmental range of growing conditions in wheat that presumably had their effects on the implicit model assumptions. In contrast to wheat models, a later comparison of rice models showed that their predictions for potential production were much more convergent, and also quite close to the observed values. This better result is partly because of the more homogeneous conditions for the cultivation of rice than wheat, but the lessons from the preceding wheat workshop may also have played a positive role.

The experience with the wheat and rice comparison exercises has led to an institutionalization of networks, under the auspices of Global Change of Terrestrial Ecosystems/International Geosphere–Biosphere Programme (GCTE/IGBP) (Walker *et al.*, 1999). It has led to quality control of experimental work, of modeling work, of data collection, and of model evaluation. Scientists have learnt to appreciate the usefulness of such networks. At the price of having to sacrifice some independence, the scientist gains access to the work of others, which enables him/her to extend his database. Also, inspection of the data by others has proven to be invaluable to improve the data quality.

Box 6 Major conclusions from the models

Typically models predict an increase of yield and biomass of about 30% as a result of doubling of ambient CO_2 alone. This is well in line with empirical results as far as the mean value is concerned. However, the large variability of the experimental results reported in the literature (ranging from 0 to 100% increase) cannot be explained by models. This means that models are still weak in allowing for interaction with other factors, and also do not suf ciently include other factors such as soil-borne diseases, induction of changes in morphology and other delicate effects in internal balances inside the plant.

The models do predict the general observation of enhanced CO_2 effects under water shortage, and of reduced CO_2 effects under nutrient shortage. Also, the models successfully predict the effect of rising temperature on crop development and its effect on yield.

CROPS AND OTHER PLANTS

In the case of crops and vegetation, it is important to model the physical and biological constraints within the environment, and also those within the plant organisms themselves. Crop models and ecological vegetation models have much in common in terms of physiology. But crops differ, in that they are planted or sown, the soil is managed, weeds, pests and diseases are removed or prevented as much as possible. Finally they are harvested, and the soil is prepared for another crop or season. These human activities remove some of the natural selection pressure and competition effects. Depending on the intensity of management, some pressure of weeds and pests remains.

The major crops of the world are the small grains (mainly rice and wheat), the coarse grains (sorghum and millet), the sugar crops (sugar cane and sugar beet), the tuber crops (potato and cassava), the leguminous crops (soybean, and other beans and peas), and the oil crops (sunflower, oil palm).

Rice and wheat together produce more than half of the total caloric need of the world human population, but apart from these major crops there has been a historic adaptation of the types of crops cultivated, to the large diversity of climates, social cultures and soil types. In spite

of this variation, the unity of physiological mechanisms provides the possibility of a common base for all crop growth models. They all contain the elements of dry matter formation, plant development and uptake of water and nutrients.

In fact, this unity also helps to better understand the ways in which weeds cause lower yields or sometimes even complete loss of a crop. Crop growth modeling offers good possibilities to unravel the interaction of weed damage and climate change, as well as to investigate means to improve weed control.

THE USE OF MODELS FOR ANSWERING ''WHAT-IF'' QUESTIONS

The use of models has become indispensable for future scenario studies (Rabbinge *et al.*, 1994). Scenario studies are not meant to be predictions. Rather, they investigate options for future development, by looking at different possible materializations of scenarios for population, economy, urbanization and climate, in combination with different alternative possibilities for policy measures (see Box 7). In the case of crop models, the scenario studies integrate the best available knowledge in the fields of production ecology, climate, and, for instance, soil science. They use scientific knowledge, both in terms of proven facts and in terms of remaining uncertainty, to find optimal paths for future development. In doing so, scientists do not intend to take over the responsibility of policy makers, but rather to reduce some of the vagueness that often surrounds the promised outcome of policy decisions. Not all uncertainty can be removed, though. Critical consideration of scientific results shows that, often, large error bars have to be attached to observations and predictions. Here it is useful to make the distinction between prediction and exploration. The uncertainty of prediction does not play a role in an explorative study into consequences of different future scenarios. Awareness of a possible disastrous effect may provoke sufficient responses to prevent such an outcome. Yet uncertainties in the model and its parameterizations remain. This is a fact of life and it will require scientific education of the public, including policy makers.

Box 7 Input- output models

Economic input- output models directly specify the empirical relationship between inputs and yield. Such models exist not only for the farm level but also for the level of a region or even for a whole country. Usually these models also allow for technological improvement. Such models are not well suited to predict effects of global change, except for those factors that enter explicitly as input factors.

REFERENCES

Boote, K J, Jones, J W, and Pickering, N B (1996) Potential Uses and Limitations of Crop Models, *Agron. J.*, **88**, 704–716.

Bouman, B A M, van Keulen, H, van Laar, H H, and Rabbinge, R (1996) The School of de Wit Crop Growth Simulation Models: a Pedigree and Historical Overview, *Agric. Syst.*, **52**(2/3), 171–198.

Casanova, D (1998) Quantifying the Effects of Land Conditions on Rice Growth. A Case Study in the Ebro Delta (Spain) Using Remote Sensing, PhD Thesis, Wageningen Agricultural University, The Netherlands, 1–219, ISBN: 90-5485-867-2.

De Koning, G H J, Jansen, M J W, Boons-Prins, E R, Diepen, C A, and Penning de Vries, F W T (1993) *Crop Growth Simulation and Statistical Validation for Regional Yield Forecasting Across the European Community*, Simulation reports CABO-TT, No. 31, AB-DLO, Wageningen, The Netherlands.

De Wit, C T (1965) Photosynthesis of Leaf Canopies, *Agric. Res. Rep.*, 663, Pudoc, Wageningen, 1–57.

De Wit, C T (1970) Dynamic Concepts in Biology, in *Prediction and measurement of Photosynthetic Productivity*, ed I Setlik, Pudoc, Wageningen, 17–23.

Gallagher, J N and Biscoe, P V (1978) Radiation Absorption, Growth and Yield of Cereals, *J. Agric. Sci.*, **91**, 47–60.

Goudriaan, J (1996) Predicting Crop Yields Under Global Change, in *Global Change and Terrestrial Ecosystems*, eds B Walker and W Steffen, Cambridge University Press, 260–274.

Horie, T (1987) A Model for Evaluating Climatic Productivity and Water Balance of Irrigated Rice, and its Application to Southeast Asia, *Southeast Asian Stud.*, **25**, 62–74.

Horie, T (1993) Predicting the Effects of Climatic Variation and Elevated CO_2 on Rice Yield in Japan, *J. Agric. Meteorol.*, **48**(5), 567–574.

Monteith, J L, Scott, R K, and Unsworth, M H, eds (1994) *Resource Capture by Crops*, Nottingham University Press, Nottingham, 1–469.

Parry, M (1990) *Climate Change and World Agriculture*, Earthscan Publications Ltd., London, 1–157.

Porter, J R (1993) AFRCWHEAT2: A Model of the Growth and Development in Wheat Incorporating Responses to Water and Nitrogen, *Eur. J. Agron.*, **2**, 69–82.

Rabbinge, R, Van Diepen, C A, Dijsselbloem, J, De Koning, G J H, Van Latesteijn, H C, Woltjer, E, and Van Zijl, J (1994) Ground for Choices: a Scenario Study on the Perspectives for Rural Areas in the European Community, in *The Future of the Land: Mobilising and Integrating Knowledge for Land Use Options*, eds L O Fresco, L S Troosnijder, J Bouma, and H Van Keulen, John Wiley & Sons, New York, 95–121.

Rabbinge, R (1993) The Ecological Background of Food Production, in *Crop Protection and Sustainable Agriculture*, Vol. 177, eds D J Chadwick and J Marsh, John Wiley & Sons, Baffins Lane, Chichester, Ciba Foundation Symposia, 2–29.

Ritchie, J T, Singh, U, Godwin, D, and Hunt, L (1991) *A Users Guide to CERES-maize – V2.10*, 2nd edition, Muscle Shoals, International Fertilizer Development Center, Alabama.

Sinclair, T R and Seligman, N G (1996) Crop Modeling: From Infancy to Maturity, *Agron. J.*, **88**(5), 698–704.

Smil, V (1997) Cycles of Life, Scientific American Library, New York, 1–221.

Supit, I (2000) An Exploratory Study to Improve Predictive Capacity of the Crop Growth Monitoring System as Applied by the European Commission, Wageningen University, Wageningen, The Netherlands, ISBN 90-804443-5-9, 1–180.

Walker, B H, Steffen, W L, Canadell, J, and Ingram, J, eds (1999) The Terrestrial Biosphere and Global Change, IGBP Book Series No. 4, Cambridge University Press, Cambridge, 1–439.

D

Daisyworld

Daisyworld is a model constructed by Watson and Lovelock to demonstrate the possibility of self-regulation in a system in which the variation of biological populations and the variation of a selective environment are mutually coupled. In the original model, the populations were daisies of two types (black or white) and the environmental variable was planetary surface temperature, externally driven by a solar radiative output that was postulated to increase with time. The relative proportions of black and white daisies determine planetary albedo, which in turn affects the surface temperature. The authors postulated a world in which the growth of the white daisies would increase as the temperature increased. This created a negative feedback that stabilized the temperature against the increasing radiative input. This stabilization occurred because, when solar output was low, the black daisies were assumed to thrive while, when the solar output increased, they are displaced by the white variety. Planetary albedo was thereby modulated and thermal homeostasis achieved as a system property. The model was important as a demonstration that environmental regulation by the biota, central to Lovelock's concept of Gaia, was conceptually possible without assumptions of purpose or foresight. However, many biologists argued that the model failed to reconcile the Gaia hypothesis with Darwinian theory because the biota of the model (daisies) were endowed by the modeler with response properties (optima for growth versus temperature) which, in real life, would be arrived at by natural selection of genetically determined variants.

See also: **Lovelock, James**, Volume 1; **Gaia Hypothesis**, Volume 5.

FURTHER READING

Watson, A J and Lovelock, J E (1983) Biological Homeostasis of the Global Environment: the Parable of Daisyworld, *Tellus*, **35B**, 284.

G R WILLIAMS *Canada*

Darwin, Charles Robert

(1809–1882)

Charles Robert Darwin was an English naturalist and evolutionist. He was the author of the *Origin of Species* (1859) in which he proposed the controversial idea of evolution by natural selection. He dedicated his life to providing evidence to substantiate it, including important writings on plants and the ancestry of human beings. Before this he traveled around the world on the HMS Beagle collecting natural history specimens, described in *Journal of Researches* (1839). Darwin was plagued by ill health, and consequently welcomed his friends, such as Thomas Henry Huxley, defending evolutionary theory. The *Origin of Species* roused violent public debate that spilled far beyond the domain of science. The concept of evolution by natural selection was only gradually accepted. In retrospect, the proposal can be seen as a major shift in environmental and biological thought that helped create the modern world.

Charles Darwin was born in Shrewsbury, the son of a prosperous medical doctor. After a brief period studying medicine at Edinburgh University, he went to Cambridge with a view to becoming a parson. This never came about. After graduation he joined Captain Robert FitzRoy on the Beagle voyage (1831–1836). The voyage was the key event in Darwin's life, greatly expanding the way he thought about the natural world. The ship visited Tahiti, the Galapagos Islands, New Zealand, and Brazil. He spent several years in South American waters, especially Chile and Tierra del Fuego. Darwin traveled and collected extensively, afterwards writing an account that is today recognized as an important record of natural environments. He learned from Charles Lyell's *Principles of Geology* that the environment is constantly changing. He applied this idea fruitfully to the origin of mountain ranges, coral reefs and other natural

history questions. On his return Darwin became converted to evolution by applying the same idea to biology. The Galapagos finches were very significant but only when properly identified. He took the idea of competitive struggle and survival rates from Malthus's *Principle of Population*. While writing for publication, he received from Alfred Russel Wallace a short essay containing near-identical ideas. A joint publication was arranged at the Linnean Society of London, July 1858, and Darwin published the *Origin of Species* in November 1859. In this he proposed that every animal and plant is variable. Individuals that are best adapted to the surrounding conditions will be the ones that survive and reproduce. Over eons of time, and with gradually changing conditions, organisms evolve. He included many instances of environmental change. Although not the first to propose evolution, Darwin's work aroused intense controversy since Victorians found it hard to accept gradual organic change, and equally hard to remove God from the creative process. The *Origin of Species* was often perceived as a manifesto for social and political reform and as a dangerously atheistic tract. It was vigorously defended by T H Huxley – Darwin's bulldog. In later life, Darwin was revered as a grand old man of science. He was buried in Westminster Abbey.

Photo: Charles Darwin at the time he was writing *On the Origin of the Species*. Photographed by the London firm of Maull and Polyblank around 1857. Reproduced by permission of The Wellcome Library.

JANET BROWNE *UK*

Data Banks: Archiving Ecological Data and Information

Jonathan M O Scurlock, Richard J Olson, Raymond A McCord[1], and William K Michener[2]
[1]*Oak Ridge National Laboratory, Oak Ridge, TN, USA*
[2]*University of New Mexico, Albuquerque, NM, USA*

Scientific data gathering has a relatively long history, but mechanisms for data distribution and exchange are more recent. Technological developments have made it possible to manipulate environmental data and then to store them. During the 1980s and 1990s, it became routine to transfer environmental data through computer networks; more recently, the Internet has opened up the possibility of browsing multiple large data collections anywhere in the world.

Traditionally, science involves making systematic observations that can be replicated. Over the past 20 years, the environmental sciences have shifted from studies of isolated populations, communities, watersheds and ecosystems towards a broader scale approach including modeling, synthesis, and assessment. In response to recent, dramatic increases in the capacity for access, processing and storage of information, scientists are now collecting data over the Internet in addition to field and laboratory work. As science moves toward regional and global interdisciplinary studies, mechanisms are needed to share data among many disciplines (e.g., ecology, meteorology, hydrology, soil science, agriculture). Consequently, submitting data to banks or archives and acquiring data from archives have become integral to today's scientific process.

Data banks or archives are permanent collections of datasets stored with accompanying metadata (information describing the data) such that users can readily acquire, understand, and use the data. A contemporary bank or archive preserves data and metadata in an electronic form that will continue to be accessible as technology changes. It also provides complete metadata so that secondary users can understand the inherent limitations associated with new applications of the archived data. It should also facilitate data sharing and foster broader scientific discoveries. Throughout this article, we use the terms data bank and data archive more or less interchangeably since both are frequently found, although as a verb archiving is more common than banking. Data center is less prevalent in our experience.

HISTORIC EXAMPLES

Certain well-known archived datasets have contributed greatly to the advancement of scientific knowledge through serendipitous applications not foreseen at the original time of data collection. For example, recent studies of the global carbon cycle have made much use of the measurements of atmospheric carbon dioxide initiated over 40 years ago at Mauna Loa, Hawaii by C D Keeling (Pales and Keeling, 1965; Keeling *et al.*, 1982; Figure 1). The now-famous pulse, representing the interaction between biosphere and atmosphere in response to global climate was later combined with other atmospheric CO_2 data from flask sampling worldwide and air sampling covering the Pacific Ocean from north to south (Keeling, 1998; WMO-WDCGG, 1999). Through the techniques of deconvolution and inversion modeling, global patterns of carbon dioxide sources and sinks were analyzed in the combined dataset, and the notorious terrestrial carbon sink was first identified by Tans and co-workers (Tans *et al.*, 1990).

Another example is provided by records of the net primary productivity (NPP) of vegetation. Data collection began in earnest in the late 19th century as accounts of

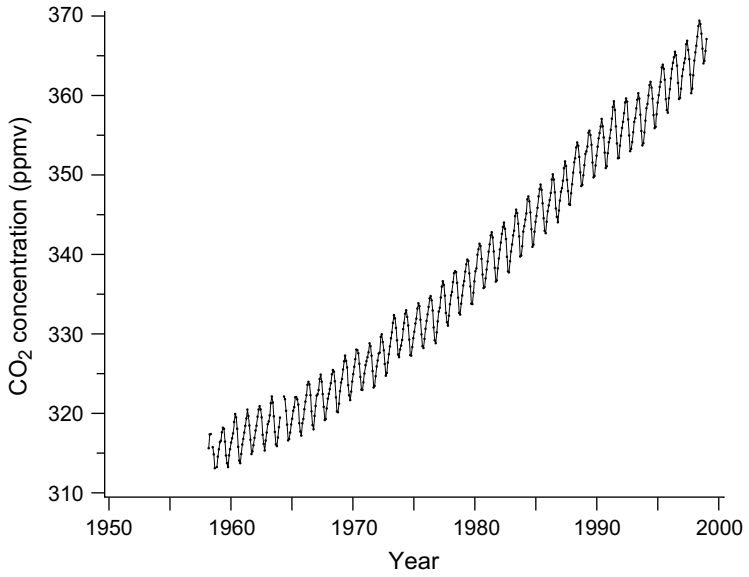


Figure 1 Atmospheric CO_2 record at Mauna Loa, Hawaii, initiated in 1958. (Thoning *et al.*, 1994; reproduced by permission of CDIAC)

the economically valuable production of crops, forage, and forest products. Some of the long-term datasets on NPP originated from such commercial agricultural interests (e.g., the Park Grass experiment at Rothamstead, UK; Silvertown *et al.*, 1994; rangeland forage production at the Central Plains Experimental Range, Colorado, USA, Lauenroth and Sala, 1992). However, long-term studies where the productivity of natural vegetation was explicitly monitored did not begin until the 1950s; coordinated efforts since that time have included such well-known scientific projects as the International Biological Program (IBP) and the US Long-term Ecological Research (LTER) program (Golley, 1993; Franklin *et al.*, 1990). Today, NPP data from many study sites worldwide are scattered throughout the peer-reviewed literature as well as in government reports and other printed matter of limited circulation, but comprehensive summaries of NPP data are rare (e.g., DeAngelis *et al.*, 1981; Cannell, 1982). More recently, the need for data to validate ecosystem models of NPP in response to increasing CO_2 and climate change, coupled with the recent growth of access to high-speed personal computers and data networks, has stimulated the formation of a world NPP data archive (Scurlock *et al.*, 1999; Scurlock and Olson, 2000).

DATA BANKS TODAY

What is a Data Bank?

Data banks comprise more than just data and metadata. Other key components are the data storage system (the media on which the data are stored, including security and back-up), the information system (which describes the data), the network connections, promotional and advertising material, optional data analysis tools, staff to maintain the archive, and, most importantly, staff to provide user services. Typically, these staff are a mixture of systems specialists, database administrators, user interface specialists, information specialists, and scientists.

The goal of a data bank is to support data use in the future, not merely to stockpile data (Kanciruk, 1997). To meet this principal objective, a data bank must be able to incorporate new data, provide user support, and maintain the integrity and accessibility of stored data (for computer systems this includes future upgrades of software, hardware, storage media, and network connectivity). See also Figure 2.

Examples of Data Banks

Most well-developed environmental data banks have limited themes and contain large amounts of data collected over long time spans at many places. For example, the network of World Data Centers (WDCs) include several data banks that have an environmental focus. The WDC system was first established through data management plans to serve various scientific disciplines in the International Geophysical Year of 1957–1958 (which itself was modeled on the International Polar Years of 1882–1883, the first large-scale international scientific enterprises). WDCs in the United States include the US National Climatic Data Center (NCDC), the Carbon Dioxide Information Analysis Center (CDIAC; for atmospheric trace gases) and the EROS Data Center (remotely sensed land data). The NCDC, for instance, which is operated by the US National

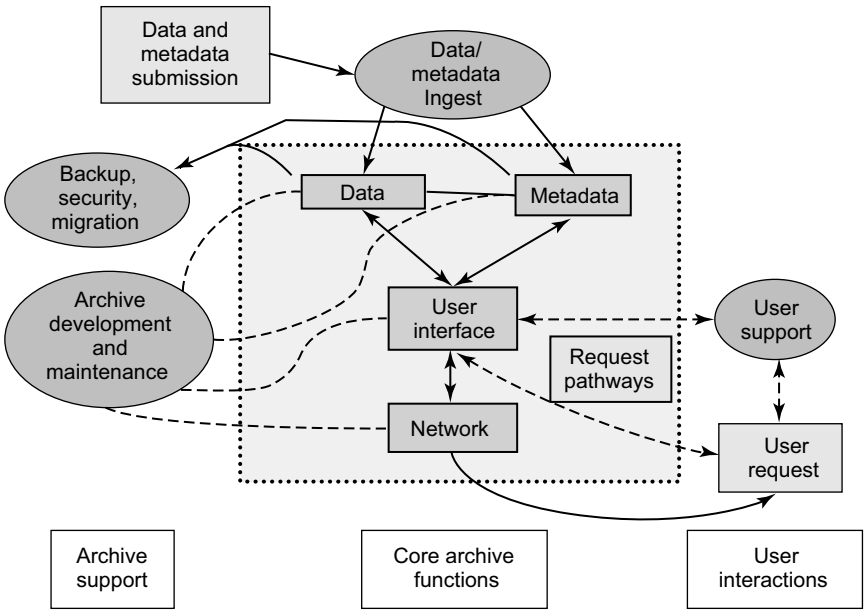


Figure 2 The ow of data and metadata from data submission to users' requests showing the internal archive and system functions. Solid lines indicate the primary ow while dashed lines indicate actions that periodically occur as needed or requested (Olson and McCord, 2000, reproduced by permission of Blackwell Science Publishers, Oxford)

Oceanographic and Atmospheric Administration (NOAA), was established in 1951, contains data collected since the 1800s), and is among the largest in terms of data volume (55 gigabytes added each day) and numbers of requests (over 170 000 data requests per year). Another relevant WDC is the International Soil Reference and Information Center (ISRIC) in the Netherlands, which stores broad-scale soils information in its Soil Information System.

Oak Ridge National Laboratory (ORNL) in Tennessee maintains two comprehensive environmental data banks: CDIAC, funded by the US Department of Energy (DOE), and the ORNL Distributed Active Archive Center (DAAC) for Biogeochemical Dynamics, funded by the US National Aeronautics and Space Administration (NASA) (Table 1). CDIAC already has a long-standing reputation for distributing data that are of high quality with known accuracy and reliability (NRC, 1995). It has produced data packages since 1982 that include extensive data description, detailed format information, extensive quality-assurance (QA) checks, examples of data applications, and copies of key associated literature. More recently, it has moved from producing annual reports and publications to providing data online. CDIAC produces value-added products for the global change research community and uses a process by which they are certified as valid by those that collected the data and carried out the quality-control checks (NRC, 1997).

The Earth Observing System (EOS) is the centerpiece of NASA's Earth Science Enterprise (ESE) (see http://eospso.gsfc.nasa.gov/), consisting of a science component and a data system supporting a coordinated series of satellites for long-term global observations – a major contribution to the US Global Change Research Program. The EOS Data and Information System component (EOSDIS) is a distributed system to support archiving and distribution of data at multiple data centers or DAACs. These centers are connected by an Information Management System (IMS) which provides an interface for one-stop shopping for earth science data allowing users to search for and order data from multiple data centers. While most of the DAACs primarily archive imagery data from the EOS satellites, the ORNL DAAC maintains data related to biogeochemical dynamics to support the validation and interpretation of satellite products. The ORNL DAAC originally grew out of the need to provide a data bank for large-scale NASA field experiments such as FIFE (First ISLSCP Field Experiment) ISLSCP (International Satellite Land-Surface Climatology Project) and BOREAS (Boreal Ecosystem-Atmosphere Study). These involved pulling together and synthesizing data from multiple sources (ground level, carbon dioxide flux towers, aircraft, satellites, etc., Sellers *et al.*, 1988; Sellers *et al.*, 1997).

Many other data facilities and research networks store and distribute environmental data, although they may represent different stages of development toward a comprehensive data bank (Table 1). Many of these systems are evolving rapidly because of significant changes in technology such as internet access and reduced costs for storage devices. Typically, their datasets contain a relatively small number (tens) of measurement types generated by a few types of instruments, methods, and sensors.

Table 1 Selected environmental data storage and distribution activities worldwide. Data centers vary in their mode of operation, including broad versus narrow thematic scopes; controlled versus public access; and short- versus long-term commitment to archive

Data Center	Web URL	Supporting Agency[a]
Atmospheric Radiation Measurement (ARM) Archive	http://www.archive.arm.gov	DOE
Base de Dados Tropicales (Brazilian tropical database)	http://www.bdt.org.br/	André Tosello Foundation, Brazil
Breeding Bird Survey (BBS)	http://www.mbr.nbs.gov/bbs/bbs.html	FWS
Carbon Dioxide Information Analysis Center (CDIAC)	http://cdiac.esd.ornl.gov	DOE
Databases at the Swedish Museum of Natural History	http://www.nrm.se/databas.html.en	Swedish Museum of Natural History
Ecological Monitoring and Assessment Network (EMAN)	http://www.cciw.ca/eman-temp/tools/datasets/intro.html	Environment Canada
Environmental Information Centre (Institute for Terrestrial Ecology, Monk's Wood, UK)	http://www.nmw.ac.uk/ITE/eic.html	NERC
National Climatic Data Center (NCDC)	http://www.ncdc.noaa.gov/wdcamet.html	NOAA
Forest Inventory and Analysis (FIA)	http://www.srs a.usfs.msstate.edu/	USFS
Global Population Dynamics Database	http://cpbnts1.bio.ic.ac.uk/gpdd/	Center for Population Biology (CPB) and NSF
ISRIC Soil Information System	http://www.isric.nl	ISRIC
Long Term Ecological Research Network (LTER)	http://lternet.edu/	NSF
EROS Data Center	http://edcwww.cr.usgs.gov/doc/edchome/world/wdcguide.html	USGS
National Environmental Satellite, Data, and Information Service (NESDIS)	http://ns.noaa.gov/NESDIS/NESDIS_Home.html	NOAA
National Geophysical Data Center (NGDC)	http://www.ngdc.noaa.gov/	NOAA
National Ocean Data Center (NODC)	http://www.nodc.noaa.gov/	NOAA
National Soils Data Access Facility (NSDAF)	http://www.statlab.iastate.edu/soils/nsdaf/	NRCS
National Water Information System (NWIS)	http://waterdata.usgs.gov/nwis-w/US/	USGS
Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC) for biogeochemical dynamics	http://www-eosdis.ornl.gov	NASA
Threatened and Endangered Species	http://www.fws.gov/~r9endspp/endspp.html	FWS
World Conservation Monitoring Centre	http://www.wcmc.org.uk/	IUCN/WWF/UNEP
World Lakes Database	http://www.ilec.or.jp/database/database.html	ILEC, UNEP, Japanese Environment Agency

[a] Agencies: DOE, US Department of Energy; FWS, US Fish and Wildlife Service; ILEC, International Lake Environment Committee Foundation; ISRIC, International Soil Reference and Information Center; IUCN, International Union for the Conservation of Nature/World Conservation Union; NASA, US National Aeronautics and Space Administration; NOAA, US National Oceanic and Atmospheric Administration, NERC, UK Natural Environment Research Council; NSF, US National Science Foundation; NRCS, US Natural Resources Conservation Service; UNEP, United Nations Environment Program; USFS, US Forest Service; USGS, US Geological Survey; WWF, Worldwide Fund for Nature/World Wildlife Fund.

A number of data directories provide information about available data (metadata) but do not store the actual data (Table 2). For example, the US National Environment Data Index (NEDI) provides direct access to environmental data and information descriptions, and thereby improves awareness of and facilitates access to data and information holdings. NASA's Global Change Master Directory (GCMD) is a comprehensive directory of descriptions of

Table 2 Directories of metadata which point to other data banks

Metadata directory	Web URL	Supporting agency[a]
Center for International Earth Science Information Network (CIESIN)	http://www.ciesin.org	Various
European Topic Center on Catalogue of Data Sources (CDS)	http://www.mu.niedersachsen.de/cds/	EEA
International Geosphere Biosphere Program Data and Information System (IGBP-DIS)	http://www.cnrm.meteo.fr/igbp/index.html	Many
Global Change Master Directory (GCMD)	http://gcmd.gsfc.nasa.gov/	NASA
National Environmental Data Index (NEDI)	http://www.nedi.gov/	
NERC Data Centres	http://www.nerc.ac.uk	NERC

[a] Agencies: EEA, European Environment Agency; NASA, US National Aeronautics and Space Administration; NERC, UK Natural Environment Research Council.

Table 3 Other selected sources of environmental data on global environmental change

Title	Web URL
BP Amoco Statistical Review of World Energy Use	http://www.bp.com/worldenergy/
FAO Yearbooks, FAOSTAT data on CD-ROM and online subscription services	http://www.fao.org/waicent/faostat/help-copyright/subscription.htm
Forestworld (private/public sector resource for international information on timber production, wood characteristics, sustainability, etc.)	http://www.forestworld.com
World Bank World Development Report	http://publications.worldbank.org/ecommerce/
Worldwatch Institute - ''State of the World'' annual reports	http://www.worldwatch.org

datasets relevant to global change research, including those covering climate change, the biosphere, hydrosphere and oceans, geology, geography, and human dimensions of global change (Figure 3). The European Environmental Agency maintains a Catalogue of Data Sources (CDS), a locator system on environmental information. CDS supplies information on who has what information in Europe, in which form and where the data exist, as well as how to gain access. (See Table 3 for a selected list of other sources of environmental data on global environmental change.)

DATA ARCHIVE FUNCTIONS

Preparing Data For Archive

Olson and McCord (2000) describe the process of preparing data for submission to a data bank. Projects that synthesize and model ecological processes at multiple spatial and temporal scales often involve bringing together and sharing data from numerous sources. The resulting collections of data from such projects are ideally suited for storage and distribution through an archive. An information system model is described by Olson *et al.* (1999) that facilitates assembling, managing, and sharing diverse data from multiple disciplines, scales, and sites to support integrated ecological studies. In this model, working groups from particular scientific domains coordinate the development of data associated with their domain, including decisions about data requirements, data to be compiled, data formats, derived data products, and schedules across multiple study sites. The data and information system is Web-based, comprising nodes for each working group plus a central node that provides data access, project information, data query, and other functionalities. This approach incorporates scientists and computer experts in the working groups and provides incentives for individuals to submit documented data for archiving.

Experience with managing environmental datasets over the past 30 years suggests that, at minimum, basic guidelines must be observed to avoid future problems with data and metadata (Slagel, 1994; Olson and McCord, 2000). For example:

1. each measurement event should be represented by only one value in the dataset i.e., there are no duplicates (such as may occur when the same values are entered with different punctuation, case, string length, etc.);
2. each value is associated with a unique place name or location (e.g., latitude/longitude);
3. data transported between organizations are accompanied by summary statistics to confirm a correct transfer, including number of records, mean, minimum, maximum, number of missing values.

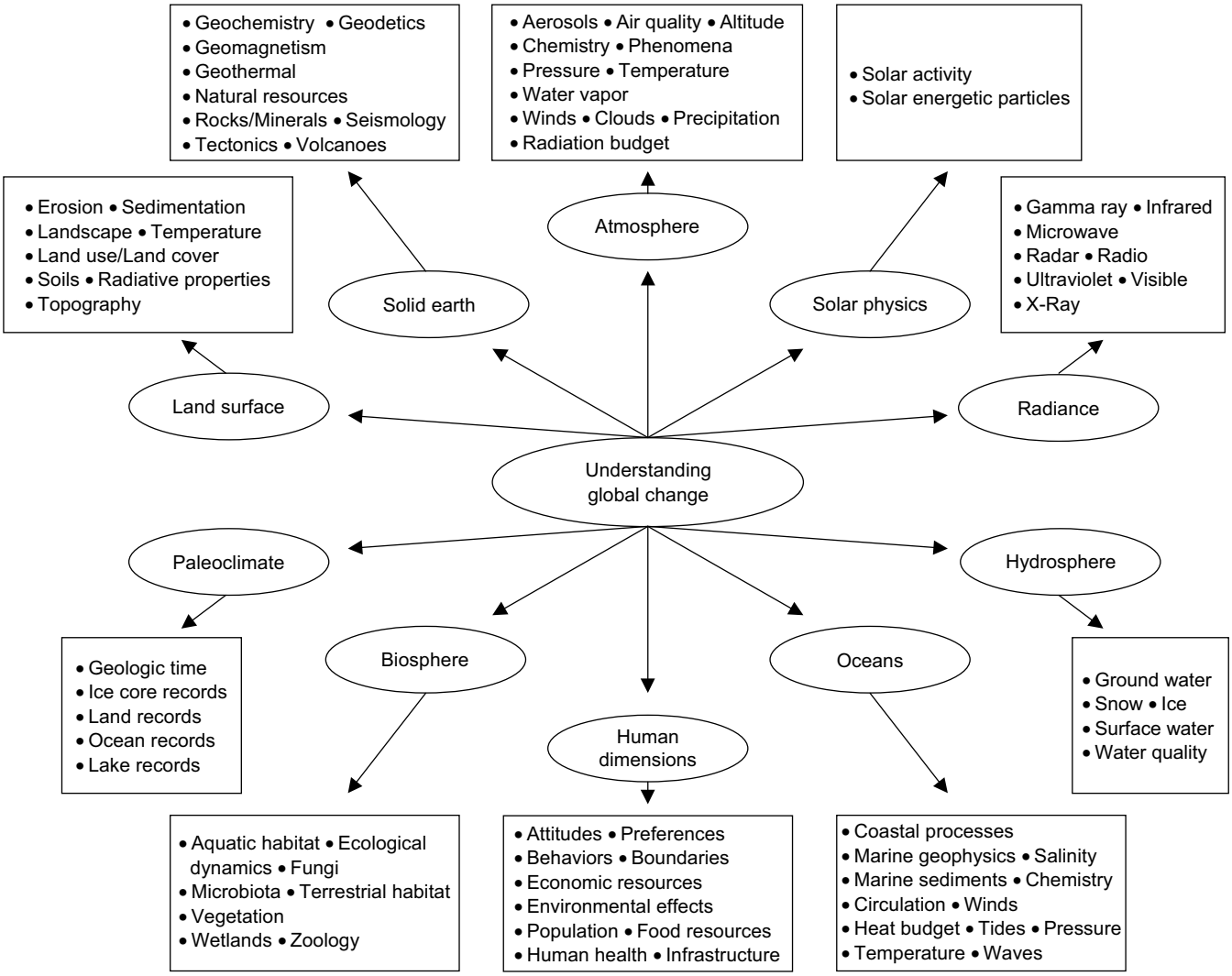


Figure 3 NASA's Global Change Master Directory (GCMD) covers a wide range of datasets including climate change, the biosphere, hydrosphere and oceans, geology, geography, and human dimensions of global change

Adding Data

Although there may be many variations on the overall process, an initial step for archiving is to connect a contributor with the appropriate data archive to add new data. Most archives continually work to identify user needs, availability of new data, and establish priorities to acquire new data. Often a data archive is associated with a specific program or has a thematic orientation and will actively seek selected datasets. The data archive provides guidelines to the contributor for formatting and submitting data and metadata. The contributor prepares data and metadata as completely as possible and submits them to the data bank, following established guidelines.

Archive staff review data and metadata and may reformat them to achieve consistency and completeness, making sure that the metadata include citation information so that the original data contributors can be correctly acknowledged. Staff also review the quality assurance that was performed on the data as documented in the metadata. They may also select keywords based on the metadata to be used in search and order functions. Archive staff and contributors work together to resolve questions and review changes. After consensus is reached, metadata and data are entered into the archive for public access and long-term storage.

Data Preservation

One of the prime responsibilities for a data archive is to ensure that the data will be accessible beyond the life of projects and programs. There are many examples of lost data, such as the ecological data from the IBP that were stored on early forms of electronic media in the early 1970s with the promise of long-term availability but are no longer

accessible (Gross *et al.*, 1995). Storing data on electronic media is no guarantee that the data archive problem is solved! Obsolescence of older computer technology seems to be the price paid for continual improvements. The lessons are twofold – maintaining paper copies of valuable datasets is a worthwhile practice for key datasets; and planning and resources are required in order to continually upgrade electronic storage, such as transferring from floppy diskettes to CD-ROMs, to DVDs (Digital Versatile Disks), and future media.

Rothenberg (1995) reviews hardware and software considerations to ensure that future generations can successfully read present-day digital files, such as avoiding data compression, avoiding proprietary storage formats and devising bootstrap standards as an input to future programs. Libraries have not had to address this problem with paper archives such as books, but periodic upgrading of storage media requires both long-term planning and funding. Life cycles seem to be about 3–5 years for current technology and around 10 years for available technology. The reliability of extending the functional life of archives beyond these limits should be carefully considered. The *20 year rule* is a practical guideline for archiving data and metadata. The aim is for the data to be fully comprehensible and employable to a user unfamiliar with the data or how they were obtained, 20 years after the data were archived, based solely on the documentation provided (NRC, 1991).

Back-up and security are important for insuring the longevity of the archive. Even the most carefully managed information systems risk information loss from mistakes, accidents, and malicious actions. Back-up copies should be stored on media that are independent of the primary storage system; ideally these should be physically located in another building as a guard against catastrophic loss. The procedure for retrieving the data from back-up should be tested routinely, and updated whenever the archive itself is updated. Limited access to the back-up copy contributes to the security of the information (security procedures should limit and record any transactions that can change the data or metadata, covering all networking and database software).

Digital Formats

There is no universal way to preserve data and metadata in a digital format. With evolving computer technology, both media and formats have changed and will continue to change. Currently ASCII flat files appear to be the most stable, neutral format for both data and metadata. Avoidance of proprietary software packages is recommended, since these may become obsolete or lack forwards- or backwards-compatibility. Most commercial packages have built-in functions to output ASCII files, such as comma-separated variables (CSV) from spreadsheets.

A drawback of ASCII formats is that a tedious, error-prone process is required to read a ASCII file into a software package (e.g., a spreadsheet or relational database management system) to use the data. It may also be difficult to detect format errors and there is no way to confirm the completeness of ASCII documentation. Several self-documenting, semi-neutral formats are emerging, including network common data form (NetCDF) and hierarchical data format (HDF). *Self-documenting* means that the files includes information about the data they contain.

NetCDF is an interface for scientific data access that implements a machine-independent, self-describing, extendible file format. It was developed for atmospheric and other data types by UCAR (University Corporation for Atmospheric Research, Boulder, Colorado, USA). It is a widely used, robust binary format with a variety of database functions and subroutines. The UCAR website at http://www.unidata.ucar.edu/packages/netcdf/ has more information.

HDF was developed by NCSA (US National Center for Supercomputing Applications, Champaign, Illinois) to assist users in the transfer and manipulation of scientific data across diverse operating systems and machines. HDF is a self-defining file format that supports a variety of data types: scientific data arrays, tables, and text annotations, as well as several types of raster images and their associated color palettes (*raster images* are patterns of closely spaced rows of dots).

Further information can be found at http://hdf.ncsa.uiuc.edu/.

A User's Guide (v.4.1r2) is available from the NCSA FTP (File Transfer Protocol) server (ftp.ncsa.uiuc.edu) in the directory: /HDF/Documentation/HDF4.1r2/Users_Guide/.

Locating Data

Representing the contents of the archive to the potential user community is a significant objective. It is essential that scientists be involved in the archive operations, where they may play a critical role in the organization and presentation of data, quality-assurance and quality-checking (QA/QC) review of data and metadata, and development of value-added products. Success of the communications between the archive and scientists depends on the proper selection of a documentation structure (how are datasets categorized) and language (what keywords are used). Input from scientists about the suitability of documentation is very important. In addition, a scientific advisory group can be a valuable link between users and sponsors.

Index System

The computing technology associated with the index or catalog is typically a database. This database contains key

fields and words that allow for tabulation of data attributes (e.g., How many datasets or records of what type? From what place? From what time?) and querying of dataset records that match user-specific criteria. The database technology required for this function of the archive includes relational databases and report generation utilities or more integrated data analysis software (e.g., Statistical Analysis Software (SAS), Statistical Package for Social Sciences (SPSS)). The database software may include functions that respond to users' queries with either the display of selected data records or summary reports. The database may also be required to record user attributes (name, e-mail address, phone number, etc.) and file access processing (FTP logs, mass storage staging status). Secondary analysis tools should also be included in the database functionality. Statistical and graphical summaries of data stored or data accessed are helpful for describing the contents of the archive, identifying data of interest, and justifying its continued operations.

User Interface

The user interface may be the most visible component of the computer technology associated with an archive. The interface provides descriptions of the archive contents to potential data users. These descriptions can be organized in hierarchical lists (e.g., sample locations within field sites within eco-regions, measurement types within environmental discipline within temporal or spatial scales). In some cases, the structural relationship of the datasets is too complex or numerous to be represented as a hierarchy and the descriptions are presented after a data base search on keywords, time intervals, or geographic regions. The user interface must be able to process the following major functions:

- The implementation of links between different levels of a hierarchy. (HTML protocols support this processing in web technology.)
- Entry and display of database parameters. The entry function provides a mechanism for the interface to record user preferences for query criteria. The display function processes the output of query results. This output includes simple statistics (number of matching records), record display (titles of descriptions), and bulk text (sections of documentation may originate from the database).
- Display of multimedia information that includes maps, graphs, text, sound, animation, and video. The documentation of environmental data frequently includes these non-text information types.
- Transmission of information by way of the Internet. This information includes the text and graphics associated with data documentation (i.e., by way of web technology). The information exchange also includes the transfer of data files by way of FTP.

The user interface is typically implemented with locally developed software that incorporates logic provided by applications software such as database servers, forms generation utilities, and image display tools. The web protocols are a common mechanism for implementing the functions of the user interface.

Subsetting

Archives that store information in databases may include logical functions that allow users to fully specify a desired data subset (e.g., by spatial, temporal, or thematic criteria), although archives that comprise collections of files or datasets (groups of loosely related files) are unlikely to generate subsets with consistent and complete structures. The end product of subsetting is customized extraction of information for the user (i.e., only those data records that match the time period, locations, parameters measured, and measurement range). The subsetting function is an important feature for archives that contain very large data objects (e.g., satellite images) or very long sequences of similar information (e.g., weather or stream flow records). The US NCDC and NWIS are examples of archives that provide subsetting from large data collections (Table 1).

Data Delivery

After the user has identified data of interest, the archive is responsible for delivering the data. Options for delivery include network, tape, and diskette. Network delivery is possible by Web transfer or FTP, but if the network connection and reliability are insufficient for the quantity of data, then tape or diskette options should considered. Data delivery on tape may be appropriate when quantities of data are large and immediate data access is not essential. However, it is important to confirm compatibility of tape hardware and formats between the archive and the user, since several tape media types are currently available, each of which has performance options that may be proprietary for the tape drive hardware or computer operating system. When networking is not available, small quantities of data may be provided on diskette, including floppy disks, disk cartridges, and CD-ROM, each of which have a limited number of format options.

Supporting Users

Other functions of the archive that may enhance these basic functions include conducting QA on the data and metadata beyond what the original data compiler performed, and combining archive data with other data. Often synthesis and modeling projects may contribute to an archive an integrated collection of data (a value-added data product) that is of particular interest to other users. Therefore, processing performed by the archive to promote the consistency and completeness of data (e.g., converting to SI metric units)

enhances the value of the data. An archive may also be able to generate summary statistics on data users (e.g., by country or Internet domain), and may be able to contact users when updates and additions to data are available.

In addition to maintaining a long-term, secure data archive, the archive staff also provide post-project support, such as answering user questions, identifying sources of reference information, informing users of updates and additions, and maintaining user statistics. Archives can perpetuate their growth and value of their data holdings by including a strategy for incorporating data updates, value-added products (especially from synthesis and modeling applications), and user feedback. Staff can collaborate with scientists to determine useful enhancements to datasets (e.g., add common variables, aggregate to common units, or calculate uncertainty) based on user needs. It is also crucial for the staff to promote the availability of the data by interacting with the user community, attending professional meetings (and staffing exhibition booths), and utilizing other marketing techniques.

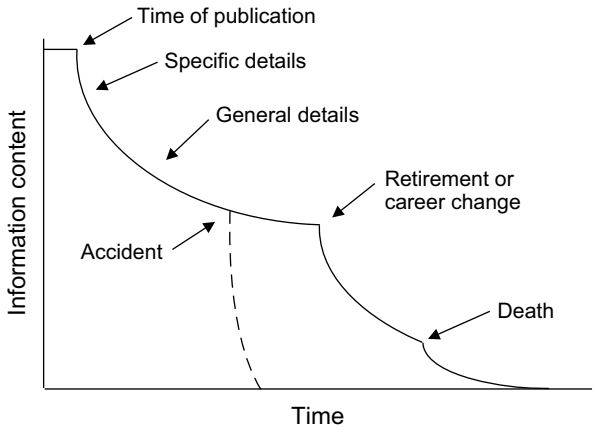


Figure 4 Example of the normal degradation of information content associated with data and metadata over time (''information entropy''). Accidents or changes in storage technology (dashed line) may eliminate access to remaining raw data and metadata at any time. (After Michener *et al.*, 1997, by permission of the Ecological Society of America)

DATA SHARING, INCENTIVES FOR ARCHIVING, AND DATA RESCUE

Data Sharing

Data may be collected as part of a project with integral data planning and funding resources for an archive, or they may follow a more haphazard path before finally arriving at an archive or data bank. In general, data flow from a principal investigator (PI) or group of PIs (those who collect, process, analyze, and publish data) towards the sharing, integrating, further analyzing and re-publishing of data. The maturing of data associated with large field projects is described by Strebel *et al.* (1994) as moving from initial infancy through adolescence to maturity. In the infant stage, data are collected and processed, while in the adolescent stage colleagues may peer review the data, which may be revised before moving to the adult stage as a free-standing, independent, published dataset. An alternative perspective is provided by Michener *et al.* (1997), who describe information entropy as the long-term degradation of information content associated with data as an investigator's personal knowledge of the data is forgotten or data or metadata are accidentally lost (Figure 4). Data archiving is one way to promote the sharing of data, to minimize the entropy of data, and to accumulate equity from research investments.

Typically, the PI has the greatest and most immediate interest in analyzing and publishing the data that he or she just collected. Only limited data sharing may occur before these initial interests have been fulfilled, such as a graduate student completing a dissertation. Although the data may thus be processed before any metadata are forgotten, data archiving may fail when PIs fear that providing immediate access to their data could result in others scooping the analysis or applications. Some funding agencies have attempted to circumvent this by specifying a specific exclusive time period for PIs.

In large collaborative multi-investigator research programs, project priorities such as research synthesis or model development may override those of individual PIs. Newly collected data may be unique, providing a fresh perspective on a science problem. Sometimes PIs may share specific information or less-sensitive data summaries while continuing to analyze their more detailed data. Prior to publication, project co-investigators and colleagues may form a wider user group around the PI. Typically the members of this group know the PI, have equivalent technical expertise and can recognize data limitations. They may be collaborators on a large project, requesting the data directly from the investigator and not requiring an independent source of extensive documentation in order to use the data.

Once data are described in a publication, an additional set of users may become interested. The next generation of data users may be from disciplines that are very different from that of the original investigator, and will require more documentation to understand the limitations associated with using the data for a particular application (Strebel *et al.*, 1994). By this time, the original scientist may have moved on to another project and may not have the resources or interest to provide additional technical support to these secondary users.

Several factors may promote a high level of interest in and demand for a particular dataset. As more publications cite a particular study and its data, there may be positive feedback in user demand for the data. However, if it requires months or years to get data into an archive for

access by others, then the archiving process may become a constraint to conducting science involving synthesis and modeling. Secondary users such as modelers will use whatever is available, or may generate their own estimates.

Incentives for Data Archiving

Project leaders, sponsors, and science managers can provide the following incentives for investigators to archive data, complementary to those incentives associated with conventional publications:

- establish a citation policy which gives credit to those who contribute data (in some cases there may be multiple contributors) as well as those who process and integrate datasets;
- provide adequate resources, guidance, and planning for data management to investigators;
- involve data personnel in the initial project planning;
- provide guidelines and training for metadata preparation;
- produce high visibility data products (e.g., CD-ROMs or hardcopy data products);
- give credit (i.e., include in promotion and salary review) to those that produce well-documented datasets.

In order to facilitate free and open exchange of data among ecologists and to create the reward structure necessary to encourage ecologists to share and exchange data, the Ecological Society of America has launched a new electronic publication (Ecological Archives) that enables the publication of peer-reviewed data papers as well as data-rich Appendices and supplements to print journals: http://esa.sdsc.edu/esapubs/Archive_main.htm.

Similarly, the British Ecological Society (BES) provides a venue for publishing ecological data as part of the *Journal of Ecology* data archives: http://www.demon.co.uk/bes/jearcmen.htm.

The BES archive contains material in HTML files, such as data, complex methods, etc., referred to in papers printed in the *Journal of Ecology*. In this case, authors are solely responsible for the accuracy of their archive material and no peer review is conducted.

Data Rescue

Sometimes datasets may enter an extended period (possibly several decades long) of low interest to be followed by renewed interest when they become part of a larger collection over time or space. Long-term data are especially valuable to detect rare events, subtle processes, and complex phenomena (Gross *et al.*, 1995). Often these datasets are orphaned because the investigator and documentation may not be readily available or because the data were recorded in an electronic format (or on punched cards) with no currently available technology to read the old files. The challenge is to be able to identify potential orphan data with legacy value that may become a key to understanding environmental trends and patterns.

An example of a data compilation and rescue effort is the hosting by the British Museum of Natural History of the Web-based Caterpillar Hostplants Database: http://www.nhm.ac.uk/entomology/hostplants/index.html.

This project offers a central facility to pool and permanently record rearing data of all Lepidoptera groups from anywhere in the world. Records of caterpillar host plants are scattered through published and manuscript sources and are difficult to retrieve. Many rearing records are never published and so are not accessible to other entomologists. Collected host plant records in this database form a valuable scientific resource that can be used to answer broader biological questions about how Lepidoptera and plants interact.

Other examples of data rescue are known from datasets contained in the NPP database at the ORNL DAAC data center: http://daac.ornl.gov/NPP/npp_home.html.

One collection of over 700 estimates of global terrestrial NPP was compiled in the early 1980s but was never made available to the scientific community. These data (the Osnabrück NPP dataset) were reviewed, documented, reorganized, and published in a citeable form (Esser *et al.*, 1997). Less auspicious is the case of the IBP Woodlands Dataset (DeAngelis *et al.*, 1981) and other datasets from the IBP, which were originally stored on magnetic media in the 1970s, but became unusable and had to be re-entered laboriously from faded computer printouts and/or printed summaries (see also “Data Preservation” above).

FUTURE DIRECTIONS

Research related to global environmental change is likely to be driven by sustainable production, protection of biodiversity, land use change, and other interdisciplinary issues. It has been suggested that this requires the establishment of not only traditional data archives, but information analysis centers that can integrate data across disciplines and disparate spatial and temporal scales (Kanciruk and Farrell, 1990). Attempts to scale research across regions, continents and the globe require an unprecedented level of collaboration among scientists, sharing of data across national and disciplinary borders, and ready access to high-quality, well-documented data from archives. Widespread use of the Internet and a gradual change in scientific philosophy to a less hierarchical and more cooperative paradigm is helping to bring this about.

However, data pertaining to different environmental resources may be incompatible, uncoordinated or stored in obscure locations. To alleviate this problem, a US government advisory committee known as PCAST (the President’s

Council of Advisors on Science and Technology) proposed the development of a federated information infrastructure (FII) whereby very large volumes (terabytes) of data from satellite, field and laboratory sources could be searched efficiently and compiled for re-analysis and re-synthesis in issue-driven research (PCAST, 1998). The FII would operate seamlessly, so that data layers such as soils, land-use and satellite imagery could be accessed from multiple databases, regardless of where these are physically located. Functions such as subsetting, merging or *concatenating* (linking) datasets would be one-click operations. The current challenges and barriers to achieving this utopian level of functionality are technical (although computer networking, communications, processing and storage are advancing rapidly), semantic (in terms of different standards and formats between multiple data sources) and social (in terms of modifying the current scientific reward structures to encourage data sharing and documenting).

Recommendations

Sharing and archiving of data can be more efficient if the following general principles are considered in the overall project planning and operations:

1. establish the flow of data from investigator to a long-term data archive as part of the work plan;
2. process data to achieve consistency and completeness of data and metadata;
3. institute policies to give the data producers adequate credit for their data archival efforts.

To more fully share data, we suggest that the environmental research community take the following actions:

- provide incentives for sharing and archiving data;
- recognize datasets with metadata as valuable research products;
- establish a universal citation policy for data;
- establish guidelines or expectations for metadata;
- develop data distribution and archive centers (including information analysis centers);
- ensure long-term financial and institutional support.

As the marginal cost of data distribution approaches zero for most types of data, and since most publicly-funded research data have already been paid for (by the public, who make up many of the users), a strong case may be made that data banks should distribute data for free (Kanciruk, 1997).

The use of environmental data will continue to expand during 21st-century synthesis and assessment studies that address long-term global and regional issues (Olson and McCord, 2000). Many future discoveries will depend on the intensive analysis of very large aggregations of data. Further reductions in cost per unit of electronic storage will continue to enable opportunities for expanded archives (storage capacity per unit hardware cost has decreased one thousand-fold over the past 10 years, and this trend is expected to continue). Exponential increases in the size of data holdings – in conjunction with increased demand for the information and knowledge necessary for informed decision-making – will require additional advances in data mining and knowledge discovery tools, as well as new technologies that can facilitate data processing (Michener, 2000). However, a minimal level of consistent information will probably still be required in order to reduce uncertainties and make data mining an effective technique. Likewise, more value-added products – such as integrated datasets for modeling – will also become available, and tools for documenting data will be used routinely.

ACKNOWLEDGMENTS

This research was sponsored by the Terrestrial Ecology Program, Office of Earth Science, US NASA, under Interagency Agreement No. 2013-I096-A1 with the US Department of Energy, and under University of Tennessee-Battelle LLC contract DE-AC05-00OR22725 with the US Department of Energy. W K Michener was supported by the US National Science Foundation. The authors also wish to thank Paul Kanciruk of Oak Ridge National Laboratory for critical comments on the manuscript.

REFERENCES

Cannell, M G R (1982) *World Forest Biomass and Primary Production Data*, Academic Press, London.

DeAngelis, D L, Gardner, R H, and Shugart, H H (1981) Productivity of Forest Ecosystems Studied During the IBP: the Woodlands Data Set, in *Dynamics of Forest Ecosystems*, IBP 23, ed D E Reichle, Cambridge University Press, 567–672.

Esser, G, Lieth, H F H, Scurlock, J M O, and Olson, R J (1997) *Worldwide Estimates and Bibliography of Net Primary Productivity Derived from Pre-1982 Publications*, ORNL Technical Memorandum TM-13485, Oak Ridge National Laboratory, Oak Ridge, TN, 1–122.

Franklin, J F, Bledsoe, C S, and Callahan, J T (1990) Contributions of the Long-term Ecological Research Program, *Bioscience*, **40**, 509–523.

Golley, F B (1993) *A History of the Ecosystem Concept in Ecology*, Yale University Press, 1–254.

Gross, K L, Pake, C E, and the FLED Committee Members (1995) *Final Report of the Ecological Society of America Committee on the Future of Long-term Ecological Data (FLED): Volume 1: Text of the Report*, Ecological Society of America, Washington, DC.

Kanciruk, P and Farrell, M P (1990) The Case for Issue-oriented Information Analysis Centers in Support of the Global Change Program, Appendix VII, in *Global Climate Trends and*

Greenhouse Gas Data: Federal Activities in Data Collection, Archiving and Dissemination, Report to the Congress of the United States, US Department of Energy, Office of Environmental Analysis, Washington, DC. DOE/PE-0094P.

Kanciruk, P (1997) Pricing Policy for Federal Research Data, *Bull. Am. Meteorol. Soc.*, **78**(4), 691–692.

Keeling, C D, Bacastow, R B, and Whorf, T P (1982) Measurements of the Concentration of Carbon Dioxide at Mauna Loa Observatory, Hawaii, in *Carbon Dioxide Review:* 1982, ed W C Clark, Oxford University Press, New York, 377–385.

Keeling, C D (1998) Rewards and Penalties of Monitoring the Earth, *Annu. Rev. Energy Environ.*, **23**, 25–82.

Lauenroth, W K and Sala, O E (1992) Long-term Forage Production of North American Shortgrass Steppe, *Ecol. Appl.*, **2**, 397–403.

Michener, W K (2000) Transforming Data into Information and Knowledge, in *Ecological Data: Design Management and Processing. Methods in Ecology*, eds W K Michener and J W Brundt, Blackwell Science, Oxford, 142–161.

Michener, W K, Brunt, J W, Helly, J, Kirchner, T B, and Stafford, S G (1997) Non-Geospatial Metadata for the Ecological Sciences, *Ecol. Appl.*, **7**, 330–342.

National Research Council (1991) *Solving the Global Change Puzzle, a US Strategy for Managing Data and Information*, Report by the Committee on Geophysical Data Commission on Biosciences, Environment and Resources. National Academy Press Washington, DC.

National Research Council (1995) *Finding the Forest in the Trees: The Challenge of Combining Diverse Environmental Data*, National Academy Press, Washington, DC, 1–129.

National Research Council (1997) *Bits of Power: Issues in Global Access to Scientific Data*, National Academy Press, Washington, DC, 1–235.

Olson, R J, Briggs, J M, Porter, J H, Mah, G R, and Stafford, S G (1999) Managing Data From Multiple Disciplines, Scales, and Sites to Support Synthesis and Modeling, *Remote Sens. Environ.*, **79**, 99–107.

Olson, R J and McCord, R A (2000) Archiving Ecological Data and Information, in *Ecological Data: Design Management and Processing. Methods in Ecology*, eds W K Michener and J W Brundt, Blackwell Science, Oxford, 117–141.

Pales, J C and Keeling, C D (1965) The Concentration of Atmospheric Carbon Dioxide in Hawaii, *J. Geophys. Res.*, **24**, 6053–6076.

PCAST (1998) *Teaming with Life: Investing in Science to Understand and Use America's Living Capital*, Office of Science and Technology Policy, Washington, DC (Available through: http://www.whitehouse.gov).

Rothenberg, J (1995) Ensuring the Longevity of Digital Documents, *Sci. Am.*, 42–47.

Scurlock, J M O and Olson, R J (2000) Terrestrial Net Primary Productivity: A Worldwide Database (in review: submitted to *Ecol. Appl.*).

Scurlock, J M O, Cramer, W, Olson, R J, Parton, W J, and Prince, S D (1999) Terrestrial NPP: Towards a Consistent Data Set for Global Model Evaluation, *Ecol. Appl.*, **9**(3), 913–919.

Sellers, P J, Hall, F G, Asrar, G, Strebel, D E, and Murphy, R E (1988) The First ISLSCP Field Experiment (FIFE), *Bull. Am. Meteorol. Soc.*, **69**(1), 22–27.

Sellers, P J, Hall, F G, Kelly, R D, Black, A, Baldocchi, D, Berry, J, Ryan, M, Ranson, K J, Crill, P M, Lettenmaier, D P, Margolis, H, Cihlar, J, Newcomer, J, Fitzjarrald, D, Jarvis, P G, Gower, S T, Halliwell, D, Williams, D, Goodison, B, Wickland, D E, and Guertin, F E (1997) BOREAS in 1997: Experiment Overview, Scientific Results and Future Directions, *J. Geophys. Res.*, **102**(D24), 28 731–28 770.

Silvertown, J, Dodd, M, McConway, K, Potts, J, and Crawley, M (1994) Rainfall, Biomass Variation, and Community Composition in the Park Grass Experiment, *Ecology*, **75**, 2430–2437.

Slagel, R L (1994) Standards for Integration of Multisource and Cross-media Environmental Data, in *Environmental Information Management and Analysis: Ecosystem to Global Scales*, eds W K Michener, J W Brunt, and S G Stafford, Taylor and Francis, London, 222–223.

Strebel, D E, Meeson, B W, and Nelson, A K (1994) Scientific Information Systems: a Conceptual framework, in *Environmental Information Management and Analysis: Ecosystem to Global Scales*, eds W K Michener, J W Brunt, and S G Stafford, Taylor and Francis, London, 59–85.

Tans, P P, Fung, I Y, and Takahashi, T (1990) Observational Constraints on the Global Atmospheric CO_2 Budget, *Science*, **247**, 1431–1438.

Thoning, K W, Tans, P P, and Waterman, L S (1994) Atmospheric CO_2 records from Sites in the NOAA/CMDL Continuous Monitoring Network, in *Trends: a Compendium of Data on Global Change*, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN.

WMO-WDCGG (1999) *WMO World Data Center for Greenhouse Gases, Data Report WDCGG No. 20* (data accepted January 1999–June 1999; Covering the Period January 1968–April 1999), Japan Meteorological Agency in Cooperation with the World Meteorological Organization, Tokyo.

De Wit, C T

(1924–1993)

Kees (Cornelis) De Wit grew up in a rural village in the eastern part of the Netherlands. As a consequence of enrollment at the Landbouwhoogeschool (Agricultural College, the predecessor of Wageningen Agricultural University), he spent most of World War II as a farm laborer. This greatly increased his interest in the complexities of farming and in farmers and was to guide his professional career. He completed his studies after the war with a PhD thesis (*A Physical Theory on Placement of Fertilizers*, 1953).

Subsequently, as a Dutch development worker, he joined the Ministry of National Planning of the Union of Burma and established the Agricultural Research Station in Gyogon. That period widened his interest in agriculture to developing countries.

After his return in 1956, De Wit remained in Wageningen. He was employed at the Institute for Biological and Chemical Research on Field Crops and Herbage (IBS) and its successor, Centre for Agrobiological Research (CABO), and produced some of his most influential papers:

- *Transpiration and Crop Yields* (1958); a major reinterpretation of crop water use data, based on a physical analysis of canopy processes aiming at understanding and generalizing.
- *On Competition* (1960) describing in physical and mathematical terms the interaction between plants of different species.
- *Ionic Balance and Growth of Plants* (1963, with Dr Wim Dijkshoorn), elaborating on the chemical composition of plants and the crucial role of anion/cation balance.
- *Photosynthesis of Leaf Canopies* (1965) introducing computer modeling in crop physiology.
- *A Dynamic Model of the Vegetative Growth of Crops* (1968, with Dr Rienk Brouwer), the first publication on crop growth simulation.

These publications are still widely quoted and constitute significant steps in the progress in agricultural science.

De Wit was appointed professor at Wageningen Agricultural University in 1968 to create the Department of Theoretical Production Ecology. Through its research and teaching (De Wit (co-)supervised 32 PhD theses), under his strong leadership, systems analysis and modeling gained a firm footing in the agricultural research community. "No simulation without experimentation," and later also the reverse, were among his most vivid expressions, always presented with characteristic conviction. He was the initiator of the series *Simulation Monographs* (Publisher PUDOC, Wageningen), that in the 1970s and 1980s formed the major outlet for publication of models of agricultural production systems.

In the 1980s, De Wit was member of the Dutch Scientific Council for Government Policy (WRR) and of the Technical Advisory Committee (TAC) of the Consultative Group on International Agricultural Research (CGIAR), and either initiated or took a very active part in discussions on environment, sustainability and development. For him, these appointments were challenges to combine biophysical and socio-economic sciences at a scientific and an operational level. His innovative efforts yielded a new approach to multi-stakeholder planning of land use and agricultural development (1988), an unconventional view of *Resource Use Efficiency in Agriculture* (1990) and transparant priority setting for international research (1992).

De Wit traveled extensively, had excellent professional and personal contacts around the world, and held several honorary appointments. Numerous illustrious visiting scientists profited from his insight and inspiration. His scientific qualities, his keen interest in human beings, deep feelings of justice and equality, his informality and his very sharp and systematic mind made him welcome, unavoidable and outstanding in any meeting. He received the Wolf prize of the State of Israel, the Nobel Prize for Agriculture, in 1975.

His approach to science has been adopted by many in the international science community, both within and outside agriculture. In Wageningen, The C T De Wit Graduate School for Production Ecology bears his name as a lasting testimony to his achievements.

FRITS PENNING DE VRIES AND HERMAN VAN KEULEN
Thailand

Degraded Ecosystems, Restoration of

see Restoration, Ecosystem (Volume 2)

Denitri cation

see Nitrogen Cycle (Volume 2)

Deposition, Dry

Dry deposition is the term used to describe, collectively, the processes by which atmosphere gases and particles are removed from the atmosphere and deposited at the surface of the earth, which do not involve precipitation. The term is also used to describe the flux (i.e., mass per unit time per unit area) of such material to the surface. Individual processes which contribute to dry deposition include gravitational sedimentation (primarily of larger particles), the turbulent transfer of small particles and gases to the near surface region, uptake of small particles at the surface by impaction and retention, adsorption, dissolution, chemical reaction, and biological uptake (e.g., through stomata). The effectiveness of dry deposition depends on a number of atmospheric and surface conditions: the atmospheric

concentration of constituents, their physical characteristics and chemical properties; the intensity of vertical exchange processes in the atmosphere which bring material to the near surface region; and the characteristics of the surface itself i.e., wetness, roughness, reactivity, and chemical, physical and biological properties.

Dry deposition is important because it is an effective mechanism for removing gases and particles (e.g., pollutants) from the atmosphere, and of transferring these same constituents to surface ecosystems where they may be beneficial (e.g., as nutrients; nitrates, iron) or harmful (e.g., as contaminants; sulfate in acid rain, organic chemicals). Dry deposition and wet deposition are important pathways in the biogeochemical cycles of some chemicals, transferring material from the atmosphere to the marine and terrestrial reservoirs. Dry deposition has played an important role in a number of environmental issues: lead deposition in urban areas, the transfer of semi-volatile organic species from tropical source regions to the Arctic, and the deposition of toxic metals and organic chemicals.

The relative importance of wet and dry deposition is species and location dependent. Close to sources dry deposition tends to exceed wet deposition, whereas far from source regions, wet deposition becomes dominant. For soluble and insoluble species, wet and dry deposition dominate, respectively. In the case of sulfur, for example, wet and dry deposition are of comparable magnitude on a global basis.

The quantity of dry deposition is very difficult to measure directly in most cases, and is currently done more often in a research mode than systematically in networks. Instruments to measure dry deposition are not standardized, are often very sophisticated, and tend to be species specific.

See also: **Deposition, Wet**, Volume 2.

DOUGLAS WHELPDALE *Canada*

Deposition, Wet

Wet deposition is the term used to describe, collectively, the processes by which atmospheric gases and particles are removed from the atmosphere and deposited at the surface of the earth by rain, snow or other forms of precipitation. The term is also used to describe the flux (i.e., the mass per unit area per unit time) of such material to the surface. Individual processes which contribute to wet deposition include nucleation of cloud droplets, scavenging within and below clouds, and subsequent transport to the surface in precipitation. The effectiveness of wet deposition depends on the type and intensity of precipitation, and on the physical (e.g., size, shape) and chemical properties (e.g., solubility, reactivity) of the atmospheric constituents.

Wet deposition is important because it is an effective mechanism for removing gases and particles (e.g., pollutants) from the atmosphere, and of transferring these same constituents to surface ecosystems where they may be beneficial (e.g., as nutrients; nitrates, iron) or harmful (e.g., as contaminants; sulfate in acid rain, organic chemicals). Wet deposition and dry deposition are important pathways in the biogeochemical cycles of some chemicals, transferring material from the atmosphere to the marine and terrestrial reservoirs. Wet deposition has been a key element of several environmental issues: removal of radionuclides from nuclear testing, acid rain, excess nitrate fertilization, deposition of toxic metals and organic chemicals.

The relative importance of wet and dry deposition is species and location dependent. Close to sources where atmospheric concentrations are relatively high, dry deposition tends to exceed wet deposition; whereas far from source regions, wet deposition becomes dominant. For soluble and insoluble species, wet and dry deposition dominate, respectively. In the case of sulfur, for example, wet and dry deposition, are of comparable magnitude on a global basis.

The quantity of wet deposition can be reliably measured. This is done using standardized instruments, methods and quality control procedures. Several national and international networks (such as the Global Atmosphere Watch) operate around the world to measure wet deposition systematically and the resulting data are stored in accessible data systems.

See also: **Deposition, Dry**, Volume 2.

DOUGLAS WHELPDALE *Canada*

Dispersal and Migration: Plant

see Plant Dispersal and Migration (*Opening essay*, Volume 2)

Disturbance

Virginia Dale
Oak Ridge National Laboratory, Oak Ridge, TN, USA

A disturbance can be defined as any relatively discrete event in time that disrupts ecosystem, community, or population

structure and changes resources, substrate availability, or the physical environment (White and Pickett, 1985). Disturbances are an integral part of most ecological systems. Rather than being catastrophic agents of destruction, many disturbances are normal, perhaps even integral, parts of long-term system dynamics. The composition, structure, organization, and developmental and trophic dynamics of these systems are, partially, the products of disturbances. In fact, some ecological systems contain species (e.g., lodgepole pine (Pinus contorta)) that have evolved in response to disturbances. Disturbances shape landscapes, and the disturbance-generated mosaic has important influences on biotic structure and ecosystem processes. Understanding the nature of landscape patterns and the factors controlling them is essential for understanding and predicting ecosystem dynamics and vegetation development.

Disturbances are typically characterized by their size, spatial distribution, frequency or return interval, predictability, and magnitude, including both severity and intensity. Severity describes the effects of disturbance on the biota, and intensity refers to the energy released or the force exerted by the disturbance. Severity and intensity are related. Disturbance severity is important in its own right, e.g., hurricanes of the same spatial extent may have highly variable severities, which strongly influence ecological responses to these different large-scale events. The relationship between severity and disturbance size (i.e., area impacted) is complex. The heterogeneity created by disturbances refers to the spatial distribution of disturbance severities across the system.

Disturbances span a broad range of sizes and frequencies. Small, more frequent disturbances include individual tree falls; small fires and ice storms; annual floods; and small, patchy insect outbreaks. Large, infrequent disturbances include volcanic eruptions, crown fires, and extreme floods or storms. The latter are ecologically important because the imprint they leave on the ecosystem is large in area and may persist for a very long time (Turner and Dale, 1998). Indeed, a large disturbance may even become the dominant force structuring the system, creating the template upon which subsequent ecological processes and interactions among species occur. Large crown fires in boreal forests, for example, create a mosaic of stands of varying ages that may persist for several centuries. Volcanic eruptions create a spatial structure of vegetation patterns that may endure for millennia. Large, intense disturbances can set an ecological system back to an early stage in its developmental sequence or initiate a different successional pathway. Typically, large disturbances do not result in extensive areas of uniform impact. Rather, large disturbances create complex patterns across the landscape in which the disturbance may affect some locations more intensely than others. For example, even very large fires leave some stands unscathed because of wind shifts or natural fire breaks.

In assessing the ecological effects of disturbances, residuals that remain after the disturbance are critical. Ecological residuals of disturbance have both biological and physical components. Biotic residuals, or legacies, refer to the types, quantities, and patterns of organisms and biotic structures that persist from the pre-disturbance ecosystem. Residuals may include surviving individuals, standing dead trees, vegetative tissue that can regenerate, seeds, litter, carcasses, and components of the microbial and fungal soil community. Abiotic legacies are physical modifications of the environment that may result from the disturbance, e.g., mass movements like mudslides or slope failures, lava flows, boulder movements during floods in high-gradient streams, channel and bank movement, and silt deposition. Abiotic legacies may restructure the system. The spatial distribution of both biotic and abiotic legacies is important to the recovery of affected ecosystems.

Some species and even ecosystems are adapted to either frequent or large disturbances. For example, lodgepole pines (Pinus contorta) have cones that only open subsequent to exposure to intense heat such as occurs in a forest fire. The evolution of attributes that enable a species to respond to a disturbance relates to the frequency of the disturbance events, relative to the life span of the organism of concern.

Disturbances may serve as catalysts for sudden and unexpected changes in ecological systems – there may be surprises or threshold effects that are only manifested following extreme events (Paine et al.*, 1998). For example, disturbances may lead to the unanticipated recruitment of species or may increase the likelihood of alternative successional pathways developing at a site. Ecologists need to understand the forces that may lead to such unanticipated changes across a landscape. Global environmental changes will likely both include alterations in disturbance regimes and be affected by disturbances.*

NATURAL AND ANTHROPOGENIC CAUSES OF DISTURBANCE

The origin of disturbances can be considered as a continuum from entirely natural to entirely anthropogenic (Dale *et al.*, 1998). Some disturbances are caused solely by natural processes. El Niños, hurricanes, and volcanic eruptions are examples of such disturbances. An important feature of naturally induced disturbances is that they may have complex origins.

Several different disturbances, each with its own characteristics, can interact synergistically to produce different post-disturbance conditions. For example, the 1980 eruption of Mt St Helens in Washington State, USA produced a complex array of disturbance agents (e.g., pyroclastic flow, debris avalanche, mudflows, ash deposits, and blowdown). At specific sites, several agents interacted to affect the degree of damage and the course of recovery (e.g.,

snow cover protected under-story vegetation and small trees; some sites were protected by topography from the northward directed blast). Additionally, combinations of post-disturbance site characteristics can influence the biota. For example, the 1985 tornado blowdown in Northwest Pennsylvania, USA created a variety of physical habitat types, each of which affected survival, recruitment, and interaction of organisms that could contribute to the forest recovery.

Disturbances that are totally anthropogenic constitute the other extreme. These disturbances can be accidental or purposeful. Oil spills are examples of accidental, anthropogenic events that have no analog in nature. Some anthropogenic events, however, are similar to natural disturbances. The 1996 floods created in the Grand Canyon, USA, are an example of an intentional human-caused disturbance that was meant to approximate natural floods. In some cases, unintentional human-caused disturbances mimic nature, such as fires that are set by humans but have gotten out of control.

Land use can be considered a long-term, broad-scale, chronic disturbance by humans. The diversity of land uses ranges from actions that wholly transform the system into a unique ecological situation (e.g., farming creates agriculture systems) to actions that mimic natural disturbances (e.g., the practice of silviculture is meant to simulate how natural disturbances affect trees). Land use can also unintentionally replicate the effects of natural disturbances, as is the case with military training at Fort McCoy in Northern Wisconsin, USA. The training activities produce effects similar to fires that historically burned these areas; both create oak savannahs which are requisite for the obligate host plant of the endangered Karner blue butterfly (*Lycaeides melissa samuelis*), which thrives on the military reservation.

Many disturbances are cascading (Dale *et al.*, 2001). For example, drought often weakens tree vigor, leading to insect infestations, disease, or fire. Insect infestations and disease promote future fires by increasing fuel loads, and fires promote future infestations by compromising tree defenses. Increased fire intensity or extent enhances the potential for landslides. Also, changes in land use, forest management, and atmospheric chemistry can interact with natural disturbances. For example, an increase in landslide frequency may result from harvest and road establishment in landslide-prone areas coupled with increased precipitation (that might occur with global climate change). Establishment of ski slopes in the Alps has increased the frequency of mudslides. In the southern Appalachians of the USA ozone exposure coupled with infestations of exotic insects (*Adelges piceae*) and climate change appears to be increasing Fraser fir (*Abies fraseri*) mortality and red spruce (*Picea rubens*) stress. In some cases, however, a combination of disturbances may ameliorate impacts. Under droughty conditions, stomata tend to close, reducing the effects of high ozone exposure.

Nevertheless, when ecosystems experience more than one disturbance, the multiple effects can lead to new domains. A new domain is entered when the system has not recovered from the first disturbance before a second perturbation occurs, leading the system to a new long-term condition. For instance, the combination of overgrazing with severe drought can lead to rapid desertification. Furthermore, the combination of climatically driven wildfires, fragmentation caused by agriculture, and logging in the boreal forest has resulted in significant and unprecedented changes in forest composition with a decline in the species that take a long time to reach maturity, have serotinuous cones, or the potential to reproduce vegetatively (Paine *et al.*, 1998, Cushman and Wallin, 2000). Under climate change these multiple interactions may be unprecedented and unpredictable. They are likely to appear slowly and be difficult to detect.

Such perturbations can sometimes yield ecological surprises. For example, in the winter of 1982–1983, kelp forests off the coast of Alta and Baja California, Mexico were devastated by the combination of 4–5 °C warm-water anomalies (instigated by El Niño-Southern Oscillation) coupled with storms (common winter events to which they are well-adapted) (Paine *et al.*, 1998). Similarly, the combination of the introduction of an exotic mollusc (*Potamorcubula amurensis*) and drought followed by extreme floods has initiated sweeping changes in ecosystem structure in the San Francisco Bay, USA. Thus, understanding synergistic effects of disturbances is critical.

Sometimes the factors instigating a disturbance can be detected, but often the prime causal force cannot be identified. In these cases the historical influences may be subtle, and their effects may be indirect. For example, insect disturbances to forests appear to be more frequent now either because many pest insects have been introduced by human activity or because humans have altered the state of some ecological systems (e.g., via air pollution) and, thereby, increased their vulnerability to attack. Climate change can affect forests indirectly by altering the frequency, intensity, duration, and timing of fire, drought, impacts by non-native species, insect and pathogen outbreaks, hurricanes, windstorms, ice storms, or landslides (Dale *et al.*, 2001).

The constraints and context of a disturbance may be influenced by human manipulation of disturbance events. For example, flood control has created systems that are now susceptible to large-scale, infrequent flooding rather than smaller scale, more frequent natural flooding regime. Similarly, the buildup of fuel on the forest floor as a result of fire suppression means that a subsequent fire might be particularly large and devastating. As another example, the historical patterns of farming in an east Tennessee, USA,

watershed have influenced subsequent patterns of insect outbreaks. Clearly, humans can both instigate disturbances and influence their impact.

Because humans can alter the susceptibility of systems to disturbances, they can also alter disturbance regimes. The frequency, intensity, rate of spread, and size of a disturbance are subject to human modification for some types of disturbances. Fire may be the best example of this phenomenon. In Southern Ethiopia, fire suppression leads to expansion of bush cover and a decline in range conditions (Oba *et al.*, 2000). At Eglin Air Force Base in Northern Florida, USA, decades of fire control led to the establishment of fire-prone forests dominated by sand pines (*Pinus clausa*). As biologists set controlled fires to gradually return the system to chronically burned forests, the sand pines are now being replaced by the naturally occurring longleaf pine (*Pinus palustris*) forests. One of the implications of recent changes in climate may be alterations in the frequency and intensity of natural disturbances.

Changes in disturbance regimes can alter the landscape. For example, human-managed systems tend to have more linear features (e.g., roads and straight lot lines) than systems dominated by natural patterns of disturbances. Disturbance regimes can affect plant distribution, and the size, shape, and distribution of patches are influenced by disturbance frequency and size. For instance, the substitution of a widespread, relatively uniform anthropogenic disturbance regime (e.g., clear cutting) for the patchy disturbances of the pre-settlement forests of Wisconsin, USA, has resulted in a more homogeneous forest (White and Mladenoff, 1994).

DISTURBANCES AND GLOBAL CHANGE

The impacts of global change are moderated by interactions between climate and disturbances. Therefore, it is important to understand how particular disturbances influence ecological systems and are affected by climate change. Furthermore, the frequency, size, intensity, seasonality, duration and type of the disturbances often depend on weather and climate in addition to site characteristics. As a long-term phenomenon, climate change is likely to interact with other types of disturbances.

A challenge to understanding effects of climate change on ecological systems is not only obtaining information on the impacts of temperature and precipitation but also acquiring knowledge about how climate change affects disturbance and how disturbances affect the ecological system. The impacts of disturbances on ecosystems are often not easy to predict (e.g., there are at least 12 possible outcomes of hurricanes in tropical forests (Lugo, 2000)). Furthermore, new disturbances, such as the spread of introduced species, can cause unforeseen effects (Simberloff, 2000).

The link between climate change and disturbances needs to be formally integrated into our understanding of climate change. There are relatively few field studies on how climate change interacts with disturbances. Fires and droughts are the cases where such interactions have been explored with field experiments (Flannigan *et al.*, 2000; Hanson and Weltzin, 2000).

Such interactions between climate, disturbances, and ecological systems may be critical in determining how global change expresses its effects. For example, changes in species diversity are most quickly apparent after a disturbance has cleared the landscape and species adapted to the new climate become established. Because forests are so long-lived, it may be that many climate change effects on forests are most easily observed in places where the successional pathway has been disrupted by a disturbance.

MANAGEMENT IMPLICATIONS

The realization that disturbances play an integral role in ecological processes influences management of disturbance-prone systems. Some disturbances are readily incorporated into planning and management. For example, forest thinning is designed to replicate small disturbances in forests.

Large, infrequent disturbances are difficult to plan for. When such disturbances do occur, they may trigger departures from average conditions. Even so, the costs and benefits of management actions are assessed at relatively small temporal and spatial scales. For example, even in hurricane-prone coastal areas, flood insurance is generally available to homeowners except in the days immediately preceding an impending hurricane. Furthermore, planning for response to large disturbances does not typically consider the length of time required for repair to occur naturally. In the same sense planners do not always recognize that human developments in the vicinity of the disturbance may have broad-scale implications. For example, although it may be cost-effective for an individual locality to build a levee to protect it against floods, the levee can increase the chance of downstream floods. Failure to anticipate decadal increases in the costs of flood damages in the US is partially caused by this short-term perspective. Traditional economic analyses do not anticipate such broad-scale costs that result from suppressing disturbances on a local scale.

Infrequent disturbances may not be included in planning because people's perceptions are most influenced by what has occurred during their life times. The temporal and spatial scale of large, infrequent disturbances is so immense that our impressions are dominated by the lack of such disturbances. Even actuaries struggle with how to include catastrophes. Institutional memories are often no better than individual memories in dealing with disturbances. Thus, management goals typically do not consider what aims are appropriate in the face of disturbance.

A key challenge of dealing with disturbances is developing and employing management strategies to cope with the changes. For management purposes the conditions that lead to or result from disturbances need to be understood so that manipulations can be developed, when possible. Management plans may focus on the potential for managing (a) the system prior to the disturbance, (b) the disturbance itself, (c) the system immediately after the disturbance, (d) the recovery process, or (e) a combination thereof (Dale *et al.*, 1998).

MANAGING THE SYSTEM PRIOR TO THE DISTURBANCE

Preparatory actions in anticipation of a disturbance are designed to manipulate the system and be consistent with management goals. The system can be manipulated to alter its vulnerability, its resistance or its response to a disturbance. An example of changing a system's vulnerability is altering fuel loads in a forest so that fire is more or less intense when it occurs. Resistance can be increased at some location in the system so that the extent of the disturbance is limited (e.g., through fire breaks, spraying for insects or planting windbreaks). Then the disturbing force can still occur with its full intensity, but the system will not be altered to the extent it would have been without the management intervention. For example, restoration of the Kissimmee River in FL, USA, involves reconnecting the river and the floodplain through channel restructuring so that flood events improve dissolved oxygen conditions in the stream and enhance wetland communities. Greater attention is needed on the potential for managing ecosystems so that they can better withstand a disturbance. Tools for accomplishing this goal include intercropping, silvicultural techniques, species selection, genetic engineering and experimenting with new species combinations. For example, novel combinations of native species, naturalized species and non-invasive, non-native species can create a system that is more resistant to disturbances. An example of a system made more resistant is the historical development of vegetation in New England, USA: as the effects of disturbances such as hurricanes, winter storms, and massive deforestation, agriculture, and land abandonment have taken their toll, the types and spatial patterns of forest communities that prevail in the New England landscape have changed dramatically, as has the area's susceptibility to disturbances. In some cases the landscape impacts of changing the resistance of systems may be severe. For example, in the aftermath of the massive 1988 fires at Yellowstone National Park, USA, it was recognized that the large scale of those fires was, in part, the result of previous fire-control actions that connected patches of fire-prone forests.

MANAGING A DISTURBANCE

The disturbance itself can be managed in one of three ways: no action, prevention of the disturbance or manipulation of its effects. Note that "no action" is not a default but a conscious management decision based on knowledge about the potential disturbance, the system and the management goals. "No action" is taken when nothing can be done about the disturbance (as in the case of most tornados, volcanoes or hurricanes) or when one can accept the consequences of the event. For instance, the 1988 fires in Yellowstone National Park were initially not fought because it was felt that the effects would be acceptable. As another example, with today's technology it is prohibitively expensive to prevent the wide-scale destruction that could result from meteors crashing into the Earth, i.e., by launching nuclear bombs on rockets to intercept the meteors, explode, and thus, change the meteors' courses.

Prevention is usually motivated by the desire to satisfy a human concern. Managers attempt to prevent disturbance when the costs of such efforts are perceived to be balanced or exceeded by the anticipated benefits. Thus, dams, levees, reservoirs and coastal barriers are built to prevent disturbances caused by water action.

Disturbance manipulation requires human intervention with the causal forces which, when possible, is always costly and difficult. Disturbances can be manipulated to either reduce their effects or to increase their impacts in an attempt to recreate more natural conditions. The disturbances most amenable to manipulation may be the ones that result from anthropogenic activities, such as some fires. However, other attempts have been made to create disturbances that mimic natural events. For example, in March 1996, the water flow of the Colorado River was increased by releasing flood-level volumes from Glen Canyon Dam. The goal of this anthropogenic flood was to instigate geomorphic and biotic conditions similar to those of the natural flooding regime of the Colorado River that had been kept in check by dams for many years. The result of the flood was an increase in suitable habitat for a variety of native fish species and a reduction in the density of non-native riparian species (such as the naturalized and highly invasive plant, tamarisk).

MANAGING THE SYSTEM AFTER THE DISTURBANCE

An inability to anticipate disturbances is often paired with a lack of recognition of the potential for recovery. Recovery typically is a slow, step-by-step process by which individual organisms become established and, eventually, trophic interactions and other ecosystem processes become re-established. Because the potential for recovery is seldom recognized by either the public or the political system

and because public attention is heightened after a large disturbance, managers are often pressured to do something. In more than one instance, such actions have not only been very costly but also detrimental to the natural recovery process. For example, leaving the system alone made more sense than some of the intensive recovery actions taken after Hurricane Hugo, the eruption of Mount St. Helens and the Exxon Valdez oil spill.

After Hurricane Hugo passed over Puerto Rico in 1989, stream channels in the Luquillo Experimental Forest were packed with logs from trees knocked down by the high winds. Some of the logjams near bridges or recreation areas were safety hazards and were removed. However, in other locations, logjams were allowed to remain in the streams because such structures provide cover and food for stream organisms and retain valuable nutrients, sediments, and organic matter in place (Covich and Crowl, 1990). As a result, the recovery of the stream community following a flood is faster with the logjams than without them. This realization allowed forest managers to focus their limited resources on safety hazards and other priorities rather than clearing all log jams, actions that would have been counterproductive or neutral in terms of sustaining natural-system functions.

In contrast, the use of non-native seeds in an attempt to reduce erosion in the aftermath of the 1980 eruption of Mount St. Helens is an example of money poorly spent in recovery actions (Dale, 1991). Petitions from scientists and a resolution from the International Congress of Systematic and Evolutionary Biology curtailed the wide-scale distribution of non-native seeds; nevertheless $50 000 was spent distributing seeds by helicopter onto 32 000 ha of the new volcanic material. Subsequent studies showed that the non-native seeds had no effect on the erosion problem. In fact, areas containing the non-native species had a higher mortality of native conifer seedlings and a lower diversity of native species. A similar experience occurred in landslide areas in the Luquillo Experimental Forest after Hurricane Hugo. Non-native species planted on the landslides failed to grow and were rapidly overwhelmed by native vegetation.

Steam cleaning birds after the 1989 Exxon Valdez oil spill may also have been unnecessary (Wiens, 1996). Although more than 30 000 oiled bird carcasses were retrieved following the spill, little evidence remained of the event four years later, even in oiled environments. The demonstrated resilience of these communities raises the possibility that the effectiveness of expensive cleaning activities immediately after the spill requires more analysis and study.

MANAGING THE RECOVERY PROCESS

Because managing recovery can be thought of as managing succession, knowledge of natural succession processes increases the likelihood that the management plan will be achievable. The first step in managing recovery is to evaluate the site potential. This activity includes the evaluation of the spatial pattern of the effects, the biological and physical residuals of the disturbance, site environmental conditions, the potential for naturally dispersed seeds to aid in the recovery, and the influence of structural heterogeneity on the re-establishment of plants and animals.

Following a disturbance, the site may be changed in many ways, but one important set of characteristics of the disturbance is what remains, i.e., the residuals. Residuals include organisms, their seeds and other reproductive parts, abiotic site characteristics and organic matter. These are soon joined by migrants whose rate of arrival can be characterized by their distance from the site and their ability to spread. These organisms may interact in myriad ways by providing shade, moisture, nutrients or future propagules.

A common goal of recovery management is to shorten the process of succession or to halt the process of succession when a state has been reached considered particularly desirable for human purposes. Thus, following some disturbances, the recovery plans for the area may include altering the patterns of landscape connectivity or suppressing or removing undesirable residuals, under the assumption that such action is cost-effective. In some cases after a significant disturbance, non-native plant species may occupy the recently open site to the detriment of the native species that one is attempting to re-establish. The negative influences of the non-native cheat grass (*Bromus tectorum*) on sagebrush community recovery following wildfire are well documented (Mack, 1981). A classic tropical example of this phenomenon is the invasion of grasses (e.g., *Imperata cylindrica*) after a disturbance which subsequently inhibits forest recovery – a situation that can be avoided by site preparation or planting of native trees (Otsamo *et al.*, 1995). Thus, humans can act as agents of migration for a variety of desirable species, thereby shortening the process of succession and establishing species that prevent invasion of undesirable forms. The establishment process can be hastened in a variety of ways: by fertilizing, providing shade, providing mulch, adding mycorrhizae or undertaking other activities that allow desirable species to rapidly establish and reproduce. Managing the interactions of species and of landscape attributes is a bit more difficult and costly. Often so little is known about these interactions that the only alternative is to establish a management situation in which natural processes are likely to occur and then let the community develop on its own.

Sometimes, the effects of a disturbance or the combined effects of a disturbance and human modifications of sites (e.g., past land uses) alter the direction of succession. This situation may occur when sites are highly degraded after the disturbance. For example, forest recovery in abandoned pastures in Puerto Rico with non-native species (10 of 112

species listed) experienced a different rate of succession than sites recovering only from hurricanes (Aide *et al.*, 1996). Under the former conditions, species that are usually considered undesirable because they are invasive or non-native become desirable because they enhance repair of site conditions and the eventual restoration of native species. Similar trends appear in temperate successions and should be assessed in management scenarios.

A disturbance may be so large that the only practical approach to manage the system immediately after the disturbance is simply to allow the area to regenerate naturally with some limited surgical management to ensure that undesirable, outside forces do not influence the trajectory of recovery. For example, species may be planted to create shade under which many weedy plants cannot flourish. In other cases extensive management activities may be necessary to attain the desired goal. The balance between cost and desirability of a return to a given endpoint must be weighed in making these decisions.

In developing recovery plans, managers should focus on critical stages of the process: migration, establishment, biotic interactions (such as competition, predation, or mutualism) and the reaction of the systems to the changes that occur over successional time. For each successional stage a management prescription can be developed that offers three alterative interventions: add something to the system, kill or remove a biological component, or alter the structure of the system. For example, during the migration process a manager could add seeds, eggs or attractants for seed dispersers (e.g., posts that birds might rest upon). Alternatively, one could physically remove existing propagules from the system by events, such as an intentional flooding that removes zebra mussels from rivers. On the other hand, a manager could frighten away or exclude animals that might eat the propagules or could remove the dispersal agent of undesirable seeds. Structural additions to the system might include bird perches and alterations in landscape structure so that seed input is enhanced.

Successional trajectories after a disturbance event can follow a variety of directions, but in many cases they culminate in functional communities similar in structure and composition to the original ones. Alternatively, the successional trajectory can involve novel or unfamiliar successional communities if the magnitude and low-frequency of the disturbance sets succession back to early or uncommon stages. For example, after Hurricane Hugo, stands in the Luquillo Experimental Forest in Puerto Rico were invaded by vines and herbaceous species not observed at that site in more than 50 years (Scatena *et al.*, 1996). Yet, this early stage in the recovery of the rain forest quickly gave way to more familiar successional trees such as *Cecropia peltata.* Within seven years after the hurricane, vines and herbaceous vegetation had assumed their previous low abundance in specialized locations, such as in the canopy gaps of tall forests. However, their dominance after the hurricane was critical to immobilizing nutrients released by the disturbance and in re-establishing forest conditions. In this case non-arboreal vegetation did not inhibit forest regeneration. The traditional management practice of eliminating vines and weeds when they occur in tree plantations would have been counterproductive in the context of managing a hurricane disturbance. Anyone who witnessed a damaged stand covered by vines (e.g., Kudzu (*Pueraria thunbergiana*)) might not have imagined that the trees covered by native members of this life form would eventually grow and replace them. This lack of foresight underscores the need to base disturbance management on ecological principles and not on headlines, personal memory, or ideas about what is attractive.

As the Earth's environment continues to change, it is necessary to understand how disturbances are affected by these changes. It is also necessary to improve the scientific underpinnings of management actions that can be implemented in the face of these changes.

ACKNOWLEDGMENTS

This manuscript has been authored by a contractor of the US Government under contract No. DE-AC05-96OR22464. Accordingly the US Government retains a non-exclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for US Government purposes.

REFERENCES

Aide, T M, Zimmerman, J K, Rosario, M, and Mercado, H (1996) Forest Recovery in Abandoned Cattle Pastures Along an Elevational Gradient in Northeastern Puerto Rico, *Biotropica*, **28**, 537–548.

Covich, A P and Crowl, T A (1990) Effects of hurricane storm flow on transport of woody debris in a rain forest stream (Luquillo Experimental Forest, PuertoRico), in Tropical hydrology and Caribbean water resources, eds J H Krishna, V Q Aponte, and F Gomez-Gomez, American Water Resources Association, San Juan, PR, 197–205.

Cushman, S A and Wallin, D O (2000) Rates and Patterns of Landscape Change in the Central Sikhote-alin Mountains, Russian Far East, *Landscape Ecol.*, **15**, 643–659.

Dale, V H, Lugo, A, MacMahon, J, and Pickett, S T A (1998) Ecosystem Management in the Context of Large, Infrequent Disturbances, *Ecosystems*, **1**, 546–557.

Dale, V H, Joyce, L A, McNulty, S, Neilson, R P, Ayres, M P, Flannigan, M D, Hanson, P J, Irland, L C, Lugo, A E, Peterson, C J, Simberloff, D, Swanson, F J, Stocks, B J, and Wotton, B M (2001) Forest Disturbances and Climate Change, *BioScience*, in press.

Dale, V H (1991) The debris avalanche at Mount St. Helens: Vegetation establishment in the ten years since the eruption, *Natl. Geogr. Res. Explor.*, **7**, 328–341.

Flannigan, M D, Stocks, B J, and Wotton, B M (2000) Climate Change and Forest Fires, *Sci. Total Environ.*, **252**, 221–230.

Hanson, P and Weltzin, J (2000) Drought Disturbance from Climate Change: Response of United States Forests, *Sci. Total Environ.*, **252**, 205–220.

Lugo, A E (2000) Effects and Outcomes of Hurricanes in a Climate Change Scenario, *Sci. Total Environ.*, **262**, 243–252.

Mack, R N (1981) The Invasion of *Bromus Tectorum* L. into Western North America, *Agroecosystems*, **7**, 145–165.

Oba, G, Post, E, Syvertsen, P O, and Stenseth, N C (2000) Bush Cover and Range Conditions in Relation to Landscape and Grazing in Southern Ethiopia, *Landscape Ecol.*, **15**, 535–546.

Otsamo, A, Ådjers, T, Hadi, T S, Kuusipalo, J, Tuomela, K, and Vuokko, R (1995) Effect of Site Preparation and Initial Fertilization on the Establishment and Growth of Four Plantation Tree Species Used in Reforestation of *Imperata cylindrica* (L.) Beauv. Dominated Grasslands, *For. Ecol. Manage.*, **73**, 271–277.

Paine, R T, Tegner, M J, and Johnson, E A (1998) Compounded Perturbations Yield Ecological Surprises, *Ecosystems*, **1**, 535–545.

Scatena, F N, Moya, S, Estrada, C, and Chinea, J D (1996) The First Five Years in the Reorganization of Aboveground Biomass and Nutrient use following Hurricane Hugo in the Bisley Experimental Watersheds, Luquillo Experimental Forest, Puerto Rico, *Biotropica*, **28**, 424–440.

Simberloff, D (2000). Global Climate Change and Introduced Species in Forests, *Sci. Total Environ.*, **252**, 253–262.

Turner, M G and Dale, V H (1998) What Have we Learned from Large, Infrequent Disturbances? *Ecosystems*, **1**, 493–496.

White, P S and Pickett, S T A (1985) Natural Disturbance and Patch Dynamics: an Introduction, in *The Ecology of Natural Disturbance and Patch Dynamics*, eds S T A Pickett and P W White, Academic Press, New York, 3–13.

White, M A and Mladenoff, D J (1994) Old-Growth Forest Landscape Transitions from Pre-European Settlement to Present, *Landscape Ecol.*, **9**, 191–205.

Wiens, J A (1996) Oil, Seabirds and Science: the Effects of the Exxon Valdez Oil Spill, *BioScience*, **46**, 587–597.

DIVERSITAS

Ghillean T Prance
University of Reading, Reading, UK

A description is given of the programs, administration and structure of DIVERSITAS an international program to promote biodiversity science. It is sponsored by six international scientific organizations and functions through five Core Programmes (CPEs) and five special target areas of research (STARs). CPEs: (1); the effect of biodiversity on ecosystem function; (2) origins, maintenance and change of biodiversity; (3) systematics, inventory and classification of biodiversity; (4) monitoring of biodiversity; (5) conservation, restoration and sustainable use of biodiversity. STAR: (1); soil and sediment biodiversity; (2) marine biodiversity; (3) microbial biodiversity; (4) freshwater biodiversity; (5) human dimension of biodiversity. During 2001–2002 DIVERSITAS is also organizing the International Biodiversity Observation Year (IBOY).

DIVERSITAS is an international program of biodiversity science. It was created in 1991 to bring together under one umbrella, scientists of the many disciplines that study biodiversity. DIVERSITAS was the name selected by the original founders of the organization and is not an acronym. It is co-sponsored by six international scientific organizations: the International Union of Biological Sciences, the Scientific Committee on Problems of the Environment, the United Nations Educational, Scientific and Cultural Organisation, the International Union of Microbiological Societies (IUMS), the International Council for Science (ICSU), and the International Geosphere–Biosphere Programme. In 1966, DIVERSITAS was re-organized and a new Scientific Steering Committee of experts from many areas of research was set up to coordinate the work and set its operational plan. This work is divided into ten program elements, each focused on a fundamental scientific question about the diversity of life on Earth. Five of these elements are considered the CPEs and five are STARs. The CPEs are:

1. *The effect of biodiversity on ecosystem functioning*: which addresses the issues of how biodiversity influences ecosystem stability and resilience and how global climate change affects these relationships. It also studies the role that biodiversity plays in ecosystem processes (e.g., productivity, decomposition and nutrient cycling). This includes programs on soil and sediment biodiversity, the impact of global change on the biology of soils, the study of global change and ecological complexity.
2. *Origins, maintenance and change of biodiversity*: this element supports research at the intraspecific genetic, population and community levels, and research on the processes and mechanisms of species extinction and community organization. It seeks to understand the population and genetic processes leading to biotic diversification as well as loss, and will use this knowledge to develop effective strategies for maintaining biodiversity. The focus is on the integrating population genetics and ecological studies and integrating levels or organization from an individual to the community. Central to this element is the study of interactions among species and of habitat fragmentation as well as the ecology and genetics of rare species.

3. *Systematics: inventory and classification of biodiversity*: which has held a series of international meetings to develop consensus on the methods and priorities for a systematic inventory of the world's species, and analyze and synthesize the information obtained into predictive classifications that reflect the evolutionary history of life. A 1998 workshop discussed the conceptual and methodological framework and tools for establishing and evaluating systematic inventories. Workshops held in Crete in 1997 and in Darwin, Australia in 1998 have drawn attention to the pivotal role of systematics to all biodiversity science. Affiliated to this is the Species 2000 programme, which is gradually forming a link between taxonomic databases around the world. The ultimate goal of Species 2000 is to provide a worldwide service of an index to all known animals, plants, fungi and microorganisms. This index will form a baseline reference system for communications about biodiversity.
4. *Monitoring of biodiversity*: this element seeks to develop efficient and standardized methodologies and priorities for monitoring changes in the current and future state of biodiversity. This is essential in order to obtain more reliable and predictive comparisons between sites around the world. It is obviously closely linked and dependent upon Programme Element 3 on Systematics. The monitoring of species of all organisms and their interactions with other species and with the abiotic environment is essential to the understanding and management of biodiversity. Decisions for investigating and maintaining global biodiversity, considering its change, including the actual manipulation of this change, must be based on scientifically informed data and analyses.
5. *Conservation, restoration and sustainable use of biodiversity*: the conservation and the sustainable use of biodiversity are both crucial to the maintenance of global stability. This element focuses on the conservation of genetic diversity in wild species, especially the wild relatives of cultivated plants and domesticated animals, also on strategies for restoring degraded ecosystems and landscapes, and how to determine how different land use and land management practices could sustainably maximize production for human consumption, while simultaneously conserving ecosystem services and biodiversity. This element is also closely involved with conservation policy legislation and the Convention on Biological Diversity (CBD) and is working on how the dynamics of biodiversity affects conservation policy development.

The STARs are themes which cut across the boundaries of the Core Programme Element themes and focus on problems of special concern within biodiversity science that are often neglected or receive only limited attention elsewhere.

1. *Soil and sediment biodiversity*: are poorly known yet play a central role in ecosystem processes, including the control of the transformation of nutrients and the generation of greenhouse gases. The oxygenation of soil and sediments and formation of organic matter are other dynamic processes performed by soil biota that contribute to the sustainability of the planet. DIVERSITAS seeks to facilitate and coordinate information on how soil and sediment species composition and community structure influence ecosystem function, and to determine how human intervention and global change impact this relationship and influence feedback to global processes. The understanding of what differences exist in soils and marine and freshwater sediments and how they control particular ecosystem processes, the vulnerability of their species to disturbance, their distribution, abundance and interactions are questions that cut across many international scientific programs.
2. *Marine biodiversity*: the conservation of marine ecosystems, which are often considered as a global commons, requires expanded and improved understanding of the patterns and processes that control the diversity of life in the sea. The effects of biodiversity loss at sea are poorly understood. This STAR seeks to investigate how marine biodiversity is affected by fisheries operations, eutrophication, alteration of the physical habitat, exotic species invasion, and other human impacts on this vital ecosystem. This element of DIVERSITAS concentrates on the building of networks among interested institutions as well as addressing aspects of marine biodiversity origin, maintenance and ecosystem function.
3. *Microbial biodiversity*: plays a critical role as the basic component of many food chains and biogeochemical cycles. Despite their importance, less than 5% of the world's microorganisms have been described. New molecular techniques allow the much more rapid investigation of this area. This element of DIVERSITAS seeks to stimulate the development of innovative methods and techniques to accelerate the discovery and functional characterization of microbial biodiversity, and develop a database mechanism to collect and exchange information on the biological characteristics of microorganisms. The microbial diversity STAR is integrated throughout the other elements of DIVERSITAS and is fundamental to its mission since microbial diversity is important in soil fertility, nitrogen fixation, agricultural production, water safety and public health. In the field of microbial conservation, DIVERSITAS seeks to stimulate the

better integration of extant culture collections in many countries around the world. This STAR interacts closely with the IUMS one of the co-sponsors of DIVERSITAS.

4. *Freshwater biodiversity*: the investigation of freshwater biodiversity poses special scientific and conservation challenges because of the variety of habitats, their isolated bodies of water and their extremely rich flora and fauna. Research in this area also involves studies of water supply for various human uses, physical alteration of habitats, chemical pollution and eutrophication, over-exploitation of fish and other fishery organisms, invasion of exotic species, global change, including the effects of acid rain, potentially rising temperatures and desertification and also changes to lake circulation and nutrient supply.
5. *The human dimensions of biodiversity*: this STAR studies the relationship between biodiversity and human well-being and is acting as a catalyst in generating and synthesizing scientific information for understanding linkups between the cultural attributes of societies and their relevance for building a scientific understanding of biodiversity, evaluating the effects on human well-being of the relationships between population pressure and biodiversity dynamics; analyzing the highly complex ways in which traditional and modern human societies have managed and are managing biodiversity; and evaluating strategies for ensuring people's participation in conserving biodiversity and sustainable managing of biotic resources. The major objectives of this STAR are to acquire a better understanding of the complex relationships among the ecological, social, economic, cultural and behavioral dimensions involving humans, and also to utilize this information for managing biodiversity in a sustainable way to fulfill human needs. This STAR is involved in capacity building, and the education of a new breed of scientists who will be able to co-ordinate their research efforts with local communities and policy planners at all levels.

THE ADMINISTRATION OF DIVERSITAS

A Scientific Steering Committee manages DIVERSITAS, and is composed of experts in particular research areas of biodiversity, in consideration of the program's overall scientific balance and goals. An independent DIVERSITAS Secretariat has been established to provide overall coordination for, and access to, the program and its expertise in biodiversity science, as well as to assure relationships among current and emerging policy issues. The DIVERSITAS Homepage on the Web contains up-to-date information of its activities, at www.lemp.jussieu.fr/icsu/DIVERSITAS/.

Linkages with the Convention on Biological Diversity (CBD)

DIVERSITAS, through its access to ongoing integrated research activities, has signed a Memorandum of Cooperation with the CBD Secretariat. DIVERSITAS has been officially cited by the Conference of Parties (Bratislava, 1998) to be used as a mechanism to provide advice to the Subsidiary Body on Scientific, Technical and Technological Advice.

The Formation of DIVERSITAS National Committees

National Committees are indispensable to the DIVERSITAS program, in order to adapt to national concerns and priorities: biodiversity implies a diversity of organisms and ecological systems, a diversity of human-related impacts, as well as a diversity of approaches.

National Committees

- Can create targeted national programs to feed into DIVERSITAS program Elements. This will also help in the consolidation of different scientific disciplines within the country and coordinate with other national initiatives (systematics, genetic resources, conservation and restoration);
- involve the scientific community in the international scientific area. Science is thought to transcend all boundaries but this is not always the case. Very often, nations operate on a bilateral system, but rarely on a multi-national system;
- stimulation of international initiatives such as workshops both within and outside the country;
- facilitate the obtaining of funds. There are many national ministries or donor organizations that receive requests for funds for various projects within the country. Forming National Committees can organize and consolidate the activities, creating a unified front with the best possible science guaranteed;
- aid in the CBD process. Ensure that the information channel is open from the bottom up as well as top down;
- at present the following countries have established national committees: Brazil, France, Norway, Germany, Indonesia, Malaysia, Peru and Russia.

DIVERSITAS WESTERN PACIFIC AND ASIA (DIWPA)

One of the most active programs of DIVERSITAS is DIWPA, which has formed an international network among the countries of the Western Pacific and Asia to promote cooperative studies and information exchange on

biodiversity. The main activities are: (1) the promotion of inter-regional research projects; (2) the facilitation of an international citizens' program; and (3) the promotion of governmental and non-governmental activities for the conservation and sustainable utilization of biodiversity. DIWPA has held several international meetings and issues a regular newsletter. The office is at DIWPA, Center for Ecological Research, Kyoto University, Shimosakamoto 4, Otsu 520, Japan.

INTERNATIONAL BIODIVERSITY OBSERVATION YEAR

As part of its program, DIVERSITAS coordinated an initiative for 2001/2002 entitled the IBOY (IBOY now extends into 2002). This is an important step in the effort to know and better understand the biological diversity of our planet, and to bring biodiversity science to the forefront of the scientific agendas of nations and international organizations interested in this subject. One of the most important objectives will be to educate the general public.

There is an urgent need to fill in the gaps of information about biodiversity, at every level from the genetic variability of individual species to the workings of ecosystems – and to add that to data about how human societies interact with, and depend on, natural systems. IBOY will cover the levels of biological diversity (genetic, species and ecosystems), surveying the field for the most obvious lacks in our knowledge of global patterns of biodiversity, focusing on those areas most amenable to a short term, intensive year of international cooperation in data collection. It is hoped that the results of the integrated products of this program will be of value towards managing the Earth System and providing tools for stemming the tide of diversity losses.

IBOY is inspired by the 1957 International Geophysical Year (IGY), when scientists worked together across national and academic boundaries to advance knowledge about the earth, oceans and atmosphere. It brought unprecedented cooperation among disciplines as well as added momentum to these disciplines. Similarly, it is hoped that IBOY will generate unprecedented long-term interest and momentum for research on, and conservation of, biodiversity. Like the IGY, DIVERSITAS/IBOY will focus on projects amenable to international cooperation in data collection, with room for scientists from developing nations to participate, however modestly. It is expected to integrate what is already known and to add new data to advance the state of knowledge about biodiversity and earth's living systems.

For further information, see http://www.icsu.org/diversitas/.

Down Regulation, of Photosynthetic Capacity

see Plants - from Cells to Ecosystems: Impacts of Global Environmental Change (*Opening essay*, Volume 2)

E

Ecosystem Integrity

see Ecosystem Integrity (Volume 4)

Ecosystem Services and Costing

Steven Cork
CSIRO Sustainable Ecosystems, Canberra, Australia

Ecosystem services are the processes and conditions by which natural ecosystems, and the species that make them up, sustain and fulfil human life. Ecosystems are suites of living organisms interacting with one another and with their non-living environment. Ecosystem services emerge from these interactions and we expect the delivery of the services to decline when components of ecosystems are removed or degraded.

Just as we know we have received a professional service when we feel fulfilled personally (e.g., after attending a performance by an accomplished musician) or have received a product that we value (e.g., a piece of clothing). We can tell we have received an ecosystem service when natural ecosystems have fulfilled or sustained our lives or have provided us with goods that benefit us. Examples of ecosystem services include: maintenance of a breathable atmosphere, livable climates, and drinkable water; protection from pests and diseases; mitigation of damage from floods and other severe weather; maintenance of genetic resources from which we obtain foods and useful medicinal products; and provision of opportunities for recreation, and spiritual and cultural fulfillment. The maintenance of other species by ecosystem processes is also an ecosystem service, because it is fulfilling to humans to know that other species are surviving even if we don't use or benefit from them directly.

The concept of ecosystem services is important for assessing the effects of global change because it focuses on a limited set of outputs from the environment that encompass what is important to, and can be recognized by, most people. To put the concept into practice, we need more information on what services come from which ecosystems, how human activities are affecting these services, what the impacts of declining services will be for human welfare, and how we get the balance right between ecosystem services and technological alternatives or supplements.

Much of the literature on ecosystem services has focused on economic valuation. A major reason for a decline in many ecosystem services is that they do not pass directly through conventional markets and hence do not acquire economic value like other goods and services. Various methods have been developed to assess what people are willing to pay for ecosystem services or what it will cost to replace them, but these measures depend on how well people are informed and what technological alternatives to ecosystem services are available. This means that the economic value of ecosystem services can change through time for reasons other than global environmental change.

This suggests that economic valuation of ecosystem services is not a reliable way to assess the effects of global change. However, as economic valuation improves and moves closer to its ultimate objective of measuring overall human welfare, a measure of how well life-fulfilling and life-sustaining services are being delivered by ecosystems and technology in combination could be a powerful way to assess the impacts of global change on humans.

ECOSYSTEM SERVICES: THE CONCEPT

The terms *nature's benefits*, *nature's services*, and *ecosystem services* have been coined to describe "the conditions and processes through which natural ecosystems, and the species that make them up, sustain and fulfil human life" (Daily, 1997). This definition is not meant to imply that the needs of humans are more important than those of other species. Rather it seeks to characterize the outputs from natural ecosystems in terms of a limited set of services that all humans are capable of recognizing as important, rather than focussing on complex concepts and processes that only specialist ecologists or economists

can fully appreciate. In this way the approach potentially makes the concept of sustainable management tangible to a wide range of people. It is important to note that fulfillment of human life includes altruism towards other humans and other species just as much as it includes individual self-interest.

Table 1 lists a typical set of ecosystem services. Figure 1 presents a slightly different way of categorizing ecosystem services. It depicts three broad types of service:

1. services that transform natural assets into goods or other products of value to humans (including everything from food to life-fulfilling feelings);
2. services that assimilate the residues and by-products from production back into natural assets;
3. services that maintain natural assets via processes within ecosystems.

In our everyday lives we know we have received a professional service when one of two types of transformations takes place. Firstly, we recognize services when we are transformed (or fulfilled) personally, such as when a hairdresser styles our hair, making us feel better, when a singer or an actor's performance gives us enjoyment, or when the hands of a masseuse relieve pain from sore muscles. Secondly, we recognize services when we receive goods

Table 1 A classi cation and examples of ecosystem services (adapted from Daily, 1999)

Broad category	Sub-category	Examples
Production of goods	Food	Terrestrial animal and plant products, forage, seafood, spice
	Pharmaceuticals	Medicines, precursors to synthetic drugs
	Durable materials	Natural bre, timber
	Energy	Biomass fuels, low-sediment water for hydropower
	Industrial products	Waxes, oils, fragrances, dyes, latex, rubber, precursors to many synthetic products
	Genetic resources	The basis for the production of other goods
Regeneration processes	Cycling and ltration	Detoxi cation and decomposition of wastes, renewal of soil fertility, puri cation of air and water
	Translocation	Dispersal of seeds necessary for revegetation, pollination of crops and native vegetation
Stabilizing processes		Coastal and river channel stability, compensation and substitution of one species for another when environments vary, control of the majority of potential pest species, moderation of weather extremes (such as temperature and wind), partial stabilization of climate, regulation of the hydrological cycle (mitigation of oods, droughts)
Life-ful lling functions		Aesthetic beauty, cultural, intellectual, and spiritual inspiration, existence value, scienti c discovery, serenity
Preservation of options		Maintenance of ecological components and systems needed for the future, supply of goods and services awaiting discovery

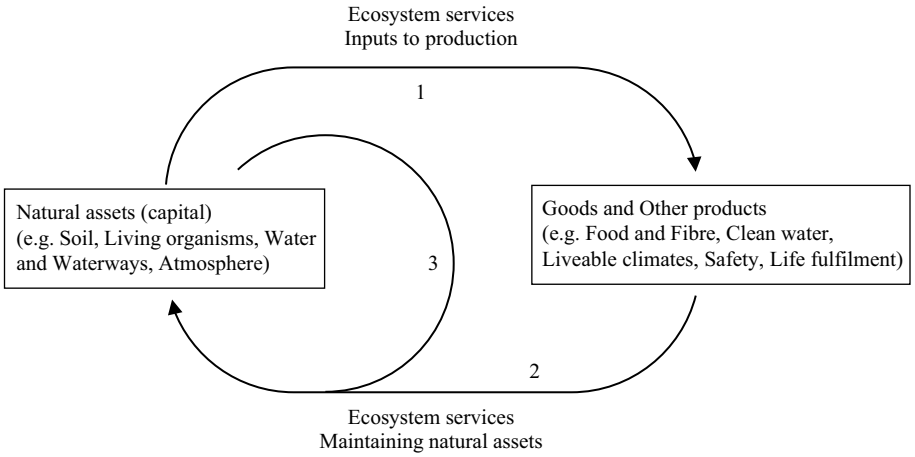


Figure 1 Schematic classi cation of ecosystem services into three broad categories: (1) services that transform natural assets into goods or products of value to humans; (2) services that assimilate the residues from production of products back into natural assets; (3) services that maintain natural assets via processes within ecosystems

of value to us that have been produced by someone else transforming materials of lesser value; for example, the production of a table from timber, bread from wheat, or a floral arrangement from a collection of individual flowers.

Similarly, ecosystem services can be recognized when people are transformed (fulfilled) by their use, or knowledge of, natural assets, or when the natural assets themselves are transformed by ecosystem processes to provide products that are either valuable to humans or help to maintain the natural assets upon which humans depend. Natural assets include the various non-living components of the atmosphere, water and soil, as well as the living organisms (plants, animals, fungi, bacteria and others) that live in, use, and help to maintain the non-living components (Figure 1).

A major strength of the concept of ecosystem services is its ability to bridge gaps in understanding and language between scientists and economists and between these disciplines and other members of the community. The concept provides a framework and a language within which the complexities of ecological and social sciences (including economics) can be simplified into ideas and language that allow debate about the key issues across disciplines and among people not formally trained in these disciplines. Lack of clarity and simplicity in defining the terms ecosystem, natural, and services can hinder this objective, not only because some people are unclear about what the terms mean but also because others have strongly-held but diverse views on what the terms mean.

Ecosystems are suites of living organisms interacting with one another and with their non-living environment. Ecosystem services are those services that emerge from these interactions. They are thus properties of ecosystems that amount to more than the sum of the properties of individual species. We can debate what degree of human influence distinguishes between a natural and an unnatural ecosystem. This begs the question of how much we should consider humans and their activities a part of nature. This debate can be very unproductive. What is important is that organisms and interactions among organisms have evolved with varying influence from natural selection; and benefits have come to humans from that evolution. Many of the benefits continue to be delivered from city parks, agricultural fields, and urban gardens, even though these have been changed greatly under the influence of humans.

There is a degree of inconsistency in the literature about the distinction between ecosystem services and ecosystem processes. Again this is an area where semantic arguments can be unproductive. Processes that transform natural assets occur at all levels of organization within ecosystems. Species, genes and chemical reactions provide products used by one another. In practice, it is useful to acknowledge a hierarchy of processes as illustrated in Table 2. As we move down the hierarchy, it becomes more difficult for most people to recognize and appreciate the significance of the services or processes. For example, while most people are aware that fruits and vegetables come from nature, they are surprised to learn that the services of pollination, natural pest control and maintenance of soil structure and fertility support the production of food.

Table 2 A hierarchical classi cation of ecosystem services

High-level services	Lower-level ''supporting'' services
Production or maintenance of: food, fuel, bre, medicines, building materials, liveable climates and atmosphere, human safety from natural disaster, ltered and puri ed water, biodiversity, aesthetic, cultural, intellectual value, raw material for scienti c discovery options, for the future	Detoxi cation of human wastes Dispersal of seeds, spores, and other propagules Maintenance of ecosystem resilience Maintenance of soil fertility and structure Pollination of crops and native vegetation Regulation of the hydrological cycle Control of pests and diseases Stabilization of waterways

HOW CAN THE CONCEPT OF ECOSYSTEM SERVICES BE USED TO ASSESS THE EFFECTS OF GLOBAL CHANGE?

The services that people need from nature are not just ecological but also economic and social (Table 1). The concept of ecosystem services integrates these diverse needs into a small number of major services, the importance of which we could estimate with the right additional information. Because the concept focuses on a level of complexity that most people can understand, it offers a way to communicate the significance of global change to a wide range of people and to engage those people in debate about solutions.

Investigating the impacts of global change on ecosystem function and the interactions with social and economic processes is difficult or impossible to do experimentally, so researchers rely heavily on models (see the next section).

UNDERSTANDING AND ASSESSING ECOSYSTEM SERVICES

In a recent report to the president of the US (PCAST, 1998; Daily, 1999), a group of eminent scientists, industry representatives and bankers identified a set of questions that need to be addressed to allow an efficient, effective and equitable balance between economic development and

ongoing delivery of ecosystem services. These questions are summarized in the following list:

1. What ecosystems provide which life support services?
2. Who benefits and over what scales of time and space?
3. What are the impacts of humans upon the supply of services?
4. How is the supply of services related to the condition of ecosystems?
5. How much damage has been done already? What is needed to repair damaged ecosystems?
6. Where are the problems geographically?
7. How interdependent are ecosystem services?
8. How reliant are ecosystem services on biological diversity?
9. How much can technology substitute for ecosystem services?
10. Given likely future technology, what area of natural ecosystems will be needed to support human life into the future?

Subsets of these questions have formed the basis for various attempts in the last decade to assess the nature of services coming from natural ecosystems around the world and to estimate their importance to humans. This work builds on research over a similar time frame that seeks to understand how biodiversity affects the way ecosystems function.

Assessments of the contributions of ecosystem services to human welfare have been done with varying degrees of rigor and with varying numbers of assumptions and expert judgements included to fill the usually considerable gaps in data. Most assessments to date have focussed on the magnitude and value of services delivered under current or past land management regimes. This is important as it informs people of what benefits they are getting now.

For example, a lower-bound estimate of the contribution of natural nitrogen fertilization to global crop production can be obtained using average estimates of net primary production and the nitrogen concentration in the crops produced (Daily *et al.*, 1997). The economic value of this contribution can be estimated as the cost of fertilizer to replace the naturally fixed nitrogen (US$45 billion per year), taking into account that fertilizer nitrogen is used much less efficiently than naturally fixed nitrogen. A value for the contribution to production of all land plants can be estimated similarly (US$320 billion).

Most major decisions relating to sustainable land management, however, require information on how the delivery of ecosystem services might change under various scenarios. This demands a more complex modeling approach linking ecology and economics dynamically, which pushes current knowledge and modeling approaches to their limits (Bockstael *et al.*, 1995).

If we want to estimate changes in delivery of fertilization services, we must take account of the effects of changed management on many influencing variables, including hydrological processes (surface and subsurface water flows), erosion, tillage and harvesting regimes, fire, climate and the many variables that in turn affect them. A typical result from simulation modeling suggests that reductions in ecosystem services by clearing vegetation in the US and Europe is leading to losses of soil, water-holding capacity, soil organisms, and nutrients that are causing a 1–8% decrease in crop productivity per year (Pimentel *et al.*, 1995). Declines in other parts of the world are higher. Such predictions can be converted readily into estimates of economic, social and broader ecological impacts.

One of the best examples of a scenario-based approach is the work of Higgins *et al.* (1997), who developed a dynamic ecological–economic model of ecosystem services provided by the mountain Fynbos Ecosystems of South Africa. The model investigated the cost-effectiveness of a set of alternative scenarios that differed with respect to management of fire, alien plants, wildflower harvesting, and controlling hikers and ecotourists to control invasion by alien plants. The model comprised five interactive submodels of hydrology, fire behavior, plant ecology, management regimes and economic valuation. For each scenario, two levels of economic benefit were considered for wildflower harvesting, hiking and ecotourist visitation, water production and future use of genetic resources stored in endemic plants. This was done to encompass uncertainty associated with estimating values. It was concluded that water production and genetic storage were the most valuable ecosystem services and that the costs of clearing alien plants would be a tiny proportion of the overall value of the ecosystem services protected.

In Australia, an attempt is being made to assess what services come from a range of ecosystems around the country and the condition of the natural assets that deliver the services. This is being accompanied by development of models that allow decision makers to ask "what if ... agriculture were practiced differently, cities were managed in other ways, or innovative new market structures, economic incentives, social policies or institutional structures were put in place?" (Cork and Shelton, 2000).

Ecological processes associated with water are frequently the focus for integrating ecosystem services and hence the scale of study is frequently the watershed or catchment. A study by a Swedish group (Rockström *et al.*, 1999) pointed out that most analyses of water use and movements in catchments fail to consider the requirements of natural ecosystems for water and that diverting water from these ecosystems for human consumption or use in irrigation potentially means declines in the services coming from the ecosystems.

The concept of ecosystem services relates strongly to the concept of the *ecological footprint*, which considers the relationship between humans and the natural environment in terms of both the impacts and needs of towns and cities (*see* **Ecological Footprint**, Volume 3). A major study linking these concepts (Folke *et al.*, 1997) estimated that on average around 1000 km^2 of forests, arable land, wetlands, and inland and marine waters are required to provide natural resources and assimilate carbon, nitrogen and phosphorus wastes per km^2 of city around the Baltic Sea. Generalizing the analyses to the 744 largest cities in the world, it was estimated that the world's needs for waste assimilation already outstrip capacity substantially (Folke *et al.*, 1997).

All models such as these represent trade-offs among realism, generality and precision. The balance between these will be determined by the objectives of the modeling. In the case of modeling ecosystem services, a prime objective will be to reduce the complexity of multiple interactions among ecosystem processes sufficiently for decision makers and stakeholders to understand what has been done, while making the uncertainty associated with predictions both clear and within acceptable limits.

Another emerging approach to investigating the complex and changing interactions between ecological and social systems is the use of complex adaptive systems models (*see* **Integrated Assessment Models of Global Change**, Volume 4; **Modeling Human Dimensions of Global Environmental Change**, Volume 5). This approach provides a framework for simplifying the multiple interactions and understanding the organization of social–ecological systems. It recognizes the grouping of processes at a limited number of scales, the partial autonomy of the groupings, and the key interactions among them. Unlike mainstream ecology and economics, it addresses non-linear change (Abel *et al.*, 2000). It acknowledges that when people are doing research and making decisions about how to manage ecosystems, these decisions will affect the way that ecosystems adapt. In turn, the response of the ecosystems will affect the thinking and actions of the decision-makers in ways that can be complex and difficult to predict (Carpenter *et al.*, 1999).

KEY ECOLOGICAL ISSUES STILL TO BE ADDRESSED

When considering ecological measurements, we need to know not only how much service is being delivered but also how that level meets people's needs and expectations. The key to understanding what will happen to the services in the future is to understand how the components of ecosystems interact and what the effects of changing the balance between components might be. We need also to know when to start worrying about change; that means thinking about what the critical levels of service are, how to tell when we are getting near these levels, and how long it might take to get there or to come back.

In ecological terms we need to look for points where the relationship between ecosystem change and delivery of a service is approaching a threshold or a situation where a small change in the ecosystem could bring about a large change in the level of services (Figures 2 and 3). The more rapidly that change is likely to occur, the more fragile the ecosystem or landscape (Figure 2). If we want to repair an ecosystem, we need to know whether the rate of recovery of ecosystem services will be the same as the rate of decline. Another key question is how the delivery of ecosystem services, is influenced by the numbers and functional types of species and genes in an ecosystem, and how this functional diversity affects the resilience of ecosystems (their ability to maintain function in the face of disturbance). For example, if a system has may species

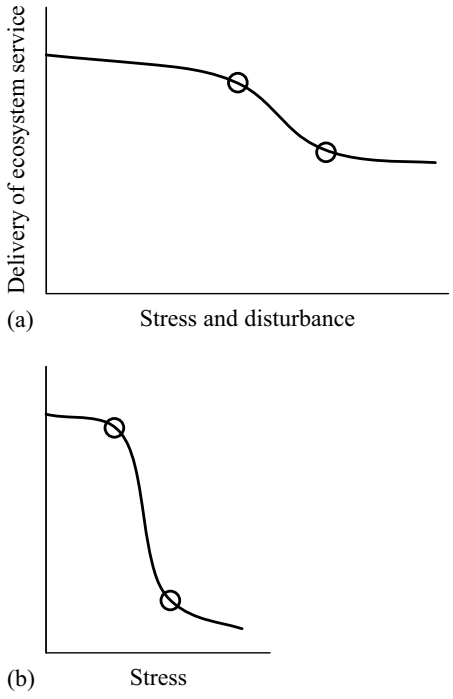


Figure 2 Hypothetical examples of how the delivery of ecosystem services might change with stress or disturbances to ecosystems within a landscape (adapted from Tongway and Hindley, 2000). (a) Represents a robust landscape which is relatively insensitive to stress and disturbance, whereas (b) represents a fragile landscape, which deteriorates rapidly with only small changes in stress and disturbance. The left-hand (upper) circle in each case indicates the point where an increasing rate of deterioration begins. We need to know how close we are to this point, and standards and management should be set to keep landscapes comfortably to the left. The right-hand (lower) circle is suggested to be the point beyond which landscapes will not return to their original state in an appropriate time frame once stress and disturbance are reduced

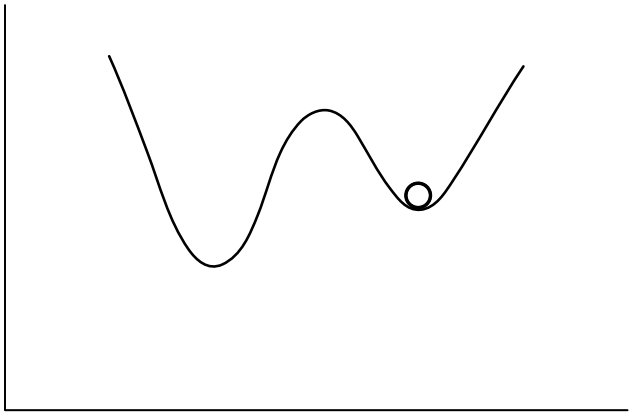


Figure 3 A ball-and-topography analogy for multiple stable states in ecosystems (Krebs, 1972). The two valleys represent alternative stable states. The different states could deliver a different set of ecosystem services. Once an ecosystem is in one of these states, external forces (e.g. ecological disturbance) is required to move it to the alternative, just as force is needed to move a ball against gravity. If the ball (or ecosystem) is only half way up the valley wall, then a small amount of extra impact might not change its state greatly and removing any external forces will allow it to return rapidly to its original state. However, if the ball (or ecosystem) is near a peak, then a small amount of additional force could cause a rapid change in state, and hence a major change in delivery of ecosystem services

performing similar functions, is it more likely to continue to deliver ecosystem services when climate or management change than a system with fewer replicates of functional types? (Walker *et al.*, 1999).

There are few data to use to construct these kinds of relationships for ecosystem services. Such information tends to emerge only once the services start to decline. For example, around the world the understanding of water chemistry and hydrology has increased greatly in recent years due primarily to the establishment of water quality regulations by governments in response to declines in natural water purification services as well as increased input of contaminants into waterways. Similarly, increasing attention is being paid to how we measure the ecosystem service of carbon sequestration since it has been recognized that concentrations of carbon in the atmosphere are reaching alarming levels and governments are discussing regulations to control emissions. We might expect to see similar attention given to the role of forests in stabilizing soil and protecting against extreme weather in the near future given the increases in damage and loss of life from floods, avalanches and mud slides in the last decade. In some places concern is rising about the decline in native pollinators for crops and other vegetation, and researchers are looking for ways to assess trends in the effectiveness of pollinators and pollination processes. For examples of these and other ecosystem services see Daily (1997).

KEY SOCIAL ISSUES TO BE ADDRESSED

The social dimensions of the ecosystem services concept have only recently been addressed in their own right. There are many questions about how people perceive their relationship with ecosystems and how that determines the importance they attach to environmental issues. When considering what our critical requirements are for ecosystem services, there is as much a social as an ecological element. Taking an ecological approach, we might consider the views of scientific experts on where ecological thresholds might exist. A social and economic approach might consider the relationships between delivery of ecosystem services and the costs of healthcare or legal liabilities. For example, standards for water quality around the world are based at least partly on considerations of human health, what levels of regulation will be tolerated by industry, and what risks of disease will be tolerated by society. When physical scientists can provide the understanding and methods to measure delivery of a service, even without identifying the critical levels, it allows regulations and standards to be developed and enforced. This is a critical interaction between the physical and social sciences (Salzman, 1997).

Another critical issue for social scientists is how we design better formal and informal institutions and organizations to identify and deal with changes in ecosystem services. These institutions need to be flexible, adaptable and able to deal with uncertainty. Some researchers have suggested that one service that ecosystems could provide to humans is a model for how to design these institutions (Berkes and Folke, 1997).

ECONOMIC VALUATION OF ECOSYSTEM SERVICES AS A MEANS OF ASSESSING THE EFFECTS OF GLOBAL CHANGE

The literature on ecosystem services has focussed strongly on methods for economic valuation (e.g., see Bingham *et al.*, 1995; Heal, 2000). The rationale for this focus is that many ecosystem services are undervalued in decision-making, leading to inequitable, inefficient and ineffective use of natural resources worldwide. Texts have been written on the concept of *value*, and expert panels have been convened to seek a unifying definition. The consensus seems to be that it is not possible to develop a single unifying definition of *value* and that it is more constructive to recognize and understand the different perceptions within society and how they relate to one another. Economic value is seen by some as just one of many values that nature can have, while others argue that properly carried out economic valuations encapsulate all societal values under the banner of social welfare. Some commentators suggest that applying economic valuations to nature enhances rather than slows its degradation, while others argue that both the causes

and solutions to environmental degradation lie in economic assessments. With respect to ecosystem services, the choice of which services to value is in itself a value judgement usually made on the basis of economic and social values. A pragmatic approach is to focus on the full range of welfare-enhancing benefits of ecosystem services, accepting that economic value is often considered in decision-making but is rarely the sole or primary determinant.

Three broad reasons are proposed for the failure of economic structures to capture the full value of ecosystem services. Firstly, people are ignorant of the benefits they obtain from ecosystem services, a problem that can be addressed by obtaining and communicating better information. Secondly, many of the benefits from ecosystems are not privately owned and consequently cannot be sold or bought directly in markets. Thirdly the costs incurred by degradation of ecosystem services are often incurred outside the market and so do not affect the price of traded goods and services.

Approaches to Valuation

Markets are a mechanism to reveal how much people will pay for tradable goods and services. Some ecosystem services contribute directly to the productions of goods that have an established market value (e.g., food, fiber, fuel, pharmaceuticals). The goods themselves can be valued by conventional techniques, but valuing the ecological processes that contributed to the goods requires detailed information on the role of each process. Often, information is not available in sufficient detail.

Indirect valuation approaches are applied to ecosystem services that do not pass through markets. These approaches include assessing the extra amounts people pay for real estate in a natural setting (*hedonic pricing method*) or for travel to natural places (*travel cost method*). The choices people make about tradeoffs between non-market and market goods and services can be analyzed (*choice modeling*) to reveal the value they place on the environmental components of the choice. Alternatively, people can be asked directly what they will give up to keep the service or would accept as compensation to lose it. This contingent valuation has been used most commonly to estimate the value of intangible services like the value of knowing that natural systems and the species within them exist.

In recent years there has been considerable attention paid to the cost of replacing ecosystem services, or the cost avoided by conserving the services, as a measure of their value. This approach requires us to consider, on the one hand, the capital costs and ongoing maintenance costs of possible technological alternatives, appropriately discounted to give net present value (*see* **Discounting**, Volume 5), and on the other hand, our ability to repair and protect ecosystem processes and to predict the long-term flows of those services. This can be difficult to assess (Salzman, 1997), but the exercise leads to many possibilities for different approaches to managing natural resources.

For example, the estimates of the contribution of natural soil fertilization services given in a previous section ('How can the concept of ecosystem services be used to assess the effects of global change?') were costs of replacement. In a similar way, the cost of replacing the role of soil in supporting plants physically and the nutrient translocating services provided by the living and non-living components of soil have been estimated at over US$900 000 ha^{-1}. The services of water filtration and purification provided by ecosystems in the catchment for New York City were recently estimated to be worth at least US$8 billion, which was the difference between the cost of repairing the ecosystems and building artificial filtration facilities to replace the degraded capacity of the ecosystem services (Chichilnisky and Heal, 1998). A study of the water catchment for the city of Melbourne in Australia suggested that the catchment was worth more if managed on long rather than short rotations of timber harvesting because of the increased amount and quality of water produced when the catchment contained mostly mature forests.

Costs avoided by conserving ecosystem services are not necessarily the same as costs of replacement, since the degradation of ecosystem services can have far reaching social and economic implications that should be taken into account when considering the effects of global change. Considering these social costs is called *shadow pricing* (Starrett, 2000).

Pollination services provide an extreme example of how very large numbers can be obtained if considering the value of ecosystem services as replacement costs. About 92% of plants worldwide are pollinated by animals, and in about half of these reproduction is determined more by the numbers of pollinators visiting them than by weather, soil fertility, diseases, parasites or animals that eat their flowers (Nabhan and Buchmann, 1997). There are more than 1200 vertebrate species and between 100 000 and 200 000 invertebrates involved in pollination of flowering plants, yet the main pollinating species of crops around the world is the European honey bee (Nabhan and Buchmann, 1997). Pollination by honey bees is estimated to be responsible for agricultural production worth US$2–8 billion annually in the US alone (Nabhan and Buchmann, 1997).

Concern has been raised about this heavy reliance on one pollinator species. A pollination crisis has been predicted in the US due to a major decline in honey bees caused by a range of diseases and mite infestations and the low numbers and capacity of native pollinators remaining after land conversion for agriculture. For some small enterprises, replacing bee pollination by manual pollen transfer by people or machines is conceivable, but the technology and expense of replacing insect pollination over large geographic scales is inconceivable at present. Some economists

have argued that the true value of many ecosystem services, measured in terms of what it would cost to replace them or what people would be prepared to accept as compensation to lose them, is infinite.

While estimates of total replacement costs get the attention of decision-makers due to the large numbers, the more relevant question for day-to-day decision-making (and for estimating the effects of global change at local scales) is whether a small, or marginal, change in ecosystems would affect the delivery of ecosystem services and their economic consequences. The ability of ecologists to make these sorts of predictions about marginal changes in ecosystem services is limited. For example, although we are now confident that concentrations of atmospheric carbon dioxide need to be reduced, we cannot predict confidently what changes to climate will occur in response to the sorts of changes we can bring about within a few years or decades.

The results of research in progress in Australia suggest that improved management of soil structure and biodiversity can be achieved over time frames of 5–10 years that could halve the amount of water, pesticides and fertilizers necessary in orchards (Matt Colloff, personal communication).

Two of the most often discussed issues relating to valuing ecosystem services are the public versus private value dichotomy and the issue of discounting.

Private Versus Public Ownership of Ecosystem Goods and Services

Privately owned ecosystem goods and services (e.g., food produced on a farm or the soil used to grow crops there) should in theory be protected by the owner in his/her own interest. On the other hand, it is often unclear who should take responsibility for ecosystem goods and services that are not owned privately or by a specific group. Examples include wildlife, water filtration services, pollination services from native insects and birds, and flood mitigation services. When impacts of declining services are external to the decision making of the people causing the impact (e.g., the impacts of removing trees on water tables, salinity and carbon sequestration), there are no traditional economic forces to encourage better practices (*see* **Commons, Tragedy of the**, Volume 5).

Discounting

In both economic and non-economic decisions, humans tend to put more value on having goods and services now than later. The primary economic rationale is that a future benefit, had it been received today, could have been invested and earned interest. Economists conventionally discount the future value of the benefit at the annual rate at which it could have earned interest. Future costs are likewise discounted because it is assumed that whoever will pay the cost is meanwhile earning interest on it. Any positive discount rate works against sustainable management of ecosystem services (*see* **Discounting**, Volume 5). Many economists argue that applying discounting is ethically indefensible because future generations will pay the full cost, and receive the full benefit. The discounted values are only from the perspective of current generations making an investment now. Discounting is valid and rational only from that perspective. A practical solution to the dilemma is to estimate benefits and costs with and without discounting, then make the investment decision based on both ethical and economic considerations. Arbitrary reductions of discount rates are not helpful because they mask the actual benefits and costs from the perspectives of both current and future generations.

Consumer or Ecosystem Sovereignty?

Some economists recently argued that while applying the criterion of economic efficiency based on the preferences of individuals is relevant when considering many day-to-day, short-term decisions, achieving longer-term sustainability requires integrated assessment of ecological, social and economic systems and consensus-building among stakeholders, to reflect societal rather than individual preferences and a wider range of objectives than economic efficiency. Whereas democratic principles and traditional economic analyses give equal weight to the preferences of all individuals, it is argued that achieving sustainable delivery of ecosystem services requires weighting for certain key life support services. In the language of economics, this means a change in focus from consumer sovereignty to ecosystem sovereignty (Costanza and Folke, 1997). Such a change almost certainly requires changes in our institutions, which in general are not sufficiently flexible and adaptive and do not work at the scales of time and space needed to fit well with ecosystem processes (Berkes and Folke, 1997).

Is Economic Valuation Useful for Assessing the Effects of Global Change?

The issues discussed in the previous sections suggest that economic valuation should be interpreted cautiously when assessing the effects of global change. On the other hand, it could be argued that a practical and realistic assessment of the impacts of global change should include an economic assessment integrated with social and ecological assessments. The challenge is to make this assessment in ways that avoid distortions that are not necessarily related to human welfare.

Such distortions can emerge in several ways. The relationship between rarity and economic value means that increasing economic value of ecosystem services is undesirable. Clean water, which is vital to human existence, has

long been worth very little economically because it has been so abundant relative to human needs but is now taking on increasing importance and value due to increasing human demand and decreasing supply. Similarly, the prices paid for properties with views of water or other natural settings are huge in most countries because of the relative rarity of such properties and the importance that humans place on natural scenery.

Mainstream economists argue that these are examples of markets working as they should and that increases in price will encourage us to move our attention away from over-exploited natural assets towards finding alternative ways to obtain the necessary services. *Ecological economics* (a new branch of economics) questions this confidence in markets, suggesting that the fundamental requirement for efficient markets (that people are fully aware of the costs and benefits of the services) is unlikely to be satisfied, and that those markets that do develop will do so too slowly to regulate ecosystem services within acceptable limits (*see* **Ecological Economics**, Volume 5).

Economic value also is affected strongly by government regulations and policies. Imposing standards for water quality or atmospheric carbon dioxide emissions, for example, imposes a regulated scarcity because the ecosystem assets and processes often have to be enhanced to meet the standards.

As pointed out earlier, economic value can be affected by how well people are informed about the benefits that they are likely to receive. Similarly, a change in the cost and feasibility of technological alternatives to ecosystem services can change the economic value of the services without any change in the state of the ecosystem. For example, the high value of New York's water comes largely from the high cost of the technological alternative. A decrease in the cost of that alternative, due for example to improved technology, would devalue the water filtration services even if the ecosystem did not change at all.

Most of these problems would be solved if individual countries and the world had an accounting system that could track changes in the biological status of ecosystems, and the services they provide, and measure the importance of changes in terms that could be considered along with changes in other economically and socially important assets and flows. There has been major discussion nationally and internationally about including natural capital in national accounts, but there remain large problems to overcome. Many economists suggest that there is a need for better ways to assess welfare at community and societal scales, and there is spirited debate about how this could be achieved. If approaches can be developed that allow changes in delivery of ecosystem services to be assessed in terms of overall human welfare, including both market and non-market values and societal as well as individual needs and preferences, then this might ultimately be a powerful way to assess the impacts of global change.

See also: **Natural Capital**, Volume 2.

REFERENCES

Abel, N, Ive, J, Langston, A, Tatnell, B, Tongway, D, Walker, B, and Walker, P (2000) Resilience of NSW Rangelands: A framework for Analysing a Complex Adaptive System, in *Management for Sustainable Ecosystems*, eds P Hale, A Petrie, D Moloney, and P Sattler, The Centre for Conservation Biology, Brisbane, 59–71.

Berkes, F and Folke, C (1997) *Linking Social and Ecological Systems: Management Practices and Social Mechanisms for Building Resilience*, Cambridge University Press, Cambridge.

Bingham, G, Bishop, R, Brody, M, Bromley, D, Clark, E, Cooper, W, Costanza, R, Hale, T, Hayden, G, Kellert, S, Norgaard, R, Norton, B, Payne, J, Russell, C, and Suter, G (1995) Issues in Ecosystem Valuation: Improving Information for Decision Making, *Ecol. Econ.*, **14**, 73–90.

Bockstael, N, Costanza, R, Strand, I, Boynton, W, Bell, K, and Wainger, L (1995) Ecological Economic Modeling and Valuation of Ecosystems, *Ecol. Econ.*, **14**, 143–159.

Carpenter, S, Brock, W, and Hanson, P (1999) Ecological and Social Dynamics in Simple Models of Ecosystem Management, *Conserv. Ecol.*, **3**(4) online.

Chichilnisky, G and Heal, G (1998) Economic Returns from the Biosphere, *Nature*, **391**, 629–630.

Cork, S J and Shelton, D (2000) The Nature and Value of Australia's Ecosystem Services: A Framework for Sustainable Environmental Solutions, in *Sustainable Environmental Solutions for Industry and Government, Proceedings of the 3rd Queensland Environmental Conference*, Environmental Engineering Society Queensland Chapter, The Institution of Engineers Australia Queensland Division, and Queensland Chamber of Commerce and Industry, Brisbane, 151–159.

Costanza, R and Folke, C (1997) Valuing Ecosystem Services with Efficiency, Fairness, and Sustainability Goals, in *Nature's Services – Societal Dependence on Natural Ecosystems*, ed G E Daily, Island Press, Washington, DC, 49–65.

Daily, G C (1999) Developing a Scientific Basis for Managing Earth's Life Support Systems, *Conserv. Ecol.*, **3**(14), online.

Daily, G C, Matson, P A, and Vitousek, P M (1997) Ecosystem Services Supplied by Soil, in *Nature's Services – Societal Dependence on Natural Ecosystems*, ed G E Daily, Island Press, Washington, DC, 113–132.

Daily, G E (1997) *Nature's Services – Societal Dependence on Natural Ecosystems*, Island Press, Washington, DC.

Folke, C, Jansson, Å, Larsson, J, and Costanza, R (1997) Ecosystem Appropriation by Cities, *Ambio*, **26**, 167–172.

Heal, G (2000) Valuing Ecosystem Services, *Ecosystems*, **3**, 24–30.

Higgins, S I, Turpie, J K, Costanza, R, Cowling, R M, Le Maitre, D C, Marais, C, and Midgley, G F (1997) An Ecological Economic Simulation Model of Mountain Fynbos Ecosystems – Dynamics Valuation and Management, *Ecol. Econ.*, **22**, 155–169.

Krebs, L J (1972) *Ecology: The Experimental Analysis of Distribution and Abundance*, Harper and Row, New York.

Nabhan, G P and Buchmann, S L (1997) Services Provided by Pollinators, in *Nature's Services – Societal Dependence on Natural Ecosystems*, ed G E Daily, Island Press, Washington, DC, 133–150.

PCAST Panel on Biodiversity and Ecosystems (1998) *Teaming with Life: Investing in Science to Understand and Use America's Living Capital*, President's Committee of Advisors for Science and Technology, Washington, DC.

Pimentel, D, Harvey, C, Resosudarmo, P, Sinclair, K, Kurz, D, McNair, M, Crist, S, Shpritz, L, Fitton, L, Saffouri, R, and Blair, R (1995) Environmental and Economic Costs of Soil Erosion and Conservation Benefits, *Science*, **267**, 1117–1123.

Rockström, J, Gordon, L, Folke, C, Falkenmark, M, and Engwall, M (1999) Linkages among Water Vapor Flows Food Production and Terrestrial Ecosystem Services, *Conserv. Ecol.*, **3**(5), online.

Salzman, J (1997) Valuing Ecosystem Services, *Ecol. Law Q.*, **24**, 887–903.

Starrett, D A (2000) Shadow Pricing in Economics, *Ecosystems*, **3**, 16–20.

Tongway, D and Hindley, N (2000) Assessing and Monitoring Desertification with Soil Indicators, in *Rangeland Desertification*, eds O Arnalds and S Archer, Kluwer Academic Publishers, Dordrecht, 89–98.

Walker, B, Kinzig, A, and Langridge, J (1999) Plant Attributes, Diversity, Resilience, and Ecosystem Function: The Nature and Significance of Dominant and Minor Species, *Ecosystems*, **2**, 95–113.

FURTHER READING

Bastan, Y (1997) *The Work of Nature*, Island Press, Washington, DC.

Ecosystem Stability

Ecosystem stability is the ability of the ecosystem to resist change or to recover from change. Stable ecosystems are presumed to have mechanisms that facilitate adaptation and internal structures (i.e., interconnections and relationships among species and among other components of diversity) which tend to dampen perturbations.

The notion of stability pertains to at least three different facets of ecosystems: their component parts, their structure, and their dynamic processes. Different measures of stability correspond to each facet. For instance, measures of the viability of populations of individual species in the ecosystem correspond to component stability; measures of the persistence of species composition or food web organization reflect on the ecosystem's structural stability; measures of the sustainability of net ecosystem productivity gauge process stability. The stability of each facet is generally linked to that of the others, and for each, substantial fluctuations may occur at scales ranging from those of local ecosystems to landscapes (e.g., as a result of cycles of succession following periodic disturbance) without destabilizing the ecosystem.

There are two different but interrelated aspects to ecosystem stability and this has resulted in much confusion in terminology. *Resistance* (or local stability or *engineering resilience*) is the capacity of an ecosystem to remain constant and predictable in the face of minor disturbance (*see* **Resilience**, Volume 2). Resistance mechanisms are those characteristics of an ecosystem's current steady state that allow it to absorb a small perturbation with little discernable impact so that the perturbation does not grow into a large one and throw the ecosystem into the domain of an entirely different steady state. Resistance is often measured by the rate of return to the original steady state following perturbation. When resistance is overcome, an ecosystem must rely on its resilience to remain stable.

Ecosystem resilience (or ecological resilience) is the ability of an ecosystem to return to its original steady state after a large perturbation has thrown it into an entirely different one (*see* **Resilience**, Volume 2). Resilience is measured by the magnitude of disturbance that can be absorbed before the ecosystem loses its capacity to return to its original steady state. The process of succession after an intense periodic disturbance (e.g., severe wildfire in mature boreal forest) exemplifies resilience. Species composition may change substantially (putting the ecosystem into an early successional state) but ensuing succession eventually returns the ecosystem to mature forest, its original steady state (with the same productivity, nutrient cycling, and species composition it had before disturbance). Thus resilience confers persistence in the face of severe disturbance.

Evolutionary theory suggests that species evolve mechanisms to deal with the disturbances they historically experience in an ecosystem, and that these adapted mechanisms confer both resistance and resilience on the ecosystem as a whole. Instability occurs if the ecosystem is thrown over a threshold from which recovery to a former state is either impossible (e.g., extinction) or, if possible, occurs only over very long time periods, or only with outside subsidies of energy and matter (e.g., fertilizer to replace lost topsoil). Destabilizing perturbations tend to differ from those experienced historically (e.g., industrial pollution, climate change, alien species, fire in tropical, but not in boreal, forests).

FURTHER READING

Holling, C S (1973) Resilience and Stability of Ecological Systems, *Annu. Rev. Ecol. Systematics*, **4**, 1–23.

Perry, D A (1994) *Forest Ecosystems*, Johns Hopkins Univerity Press, Baltimore, MD, 1–649.

Peterson, G, Allen, C R, and Holling, C S (1998) Ecological Resilience, Biodiversity and Scale, *Ecosystems*, **1**, 6–18.

RICHARD A FLEMING *Canada*

Ecosystem Structure and Function

For purposes of focused study and practice, conventional sciences of all kinds have long classified particular perceptions of phenomena as relating to either form or function. Different terms may be used in different sciences for this distinction, e.g., statics and dynamics, fixed and variable, goods and services, and constraints and freedoms. In the literature on ecosystems, the contrast is frequently stated in terms of *structure* and *function*.

Some versions of a systems approach focus mostly on either one or the other of these complementary kinds of perceptions. Frequently both perceptions are interlinked, especially within system dynamics approaches, say. Modernist reductionists may use the differential and integral calculus to inter-link perceptions of structure and function while post-modern synthesists may use relational algorithms, as Berlinski (2000) has pointed out.

Within an evolutionary systems mindscape – as in contemporary cosmology, geology, biology, ecology and sociology – nothing may be perceived to be permanent for all time and space. Anything that we perceive from this evolutionary perspective has some rate of change appropriate to it. Change may occur continuously and smoothly or episodically and discontinuously, or both together. A statistical concept like *half-life* may be invoked in this respect; it is the length of time required for half the items within a homogeneous set of phenomena to change into other kinds of phenomena. The longer the half-life, the more permanent a structure is perceived to be and the slower the relevant rate of change or frequency of transformation. With humans, a related concept is called the average lifespan or span of a generation.

Implicitly or explicitly, evolutionary scientists do have in mind some time-related concept when using the first of any of the pairs of terms mentioned above, e.g., structure, fixed, etc. In turn, the second term of each pair (e.g., process, variable, etc.) relates directly to a perception of change and also has some rate of change appropriate to it. Here too a change may occur gradually or in a step-wise manner. The rate of change with a function or process is comparatively rapid, and the manifestation of any temporary structure-like phenomenon within a process has a short half-life or short average life, and may therefore be ignored for purposes of approximation.

Consider an example. Some ecologists study the wave spectrum of heat radiated from the earth's surface to outer space, say while the sky is cloudless. They note that forested patches are, on balance, comparatively cool while urban areas with roofs, pavement and bare ground are hot. A simplistic analytical schema may be used for this complex phenomenon. The structure part of such a schema may consist of the type of surface. The function part may include the intensities and spectral compositions of the in-coming and out-going radiation as well as what happens to light and heat during the transformation at the surface, i.e., within the biospheric shell at that site.

As another example, a park warden with a particular interest in trees may perceive them to be crucial long-lived habitat structures for many animals and other kinds of organisms with shorter lives. Small, short-lived creatures may be lumped together and taken to be part of the decomposition and nutrient-recharging process for the benefit of the trees, say. In turn, a commercial forester may perceive the trees as contributing to a dynamic pool of renewable resources of a flow type, to be accumulated naturally and harvested at sustainable levels at optimum time intervals. Such a commercial forester may perceive tree growth to be a process and the whole forest as a semi-permanent structure, the goal being to achieve sustainability of both structure and process to be sustained in a productive state.

The time rates of change of different features of living things differ in part because of relevant gene-based information that operates through enzyme and hormone complexes, stated simplistically. A particular enzyme, say, generally manifests a set of rate parameters that are sensitive to temperature, e.g., a minimum temperature at which the enzyme begins to act, a rate that rises with temperature to an optimum temperature, then rises somewhat further to a temperature at which the rate is maximal, and then falls rapidly to a maximum temperature at which the enzyme may be denatured or destroyed, perhaps together with the living thing. For natural rates, whether slow as with phenomena that we may call structures or fast as with processes, the rate constant for the intermediate, rising range of the function described above generally has an exponential form first inferred by Svante Arrhenius a century ago (Lin and Regier, 1995).

This Arrhenius empirical relationship has been invoked in studies of the effects of climate warming on both structures and processes in natural and cultural features of large ecosystems. What may be perceived to be structures have Arrhenius coefficients with comparatively small absolute values; in turn, processes have coefficients with larger absolute values. The values of such coefficients are positive where growth occurs and negative where decomposition or erosion occurs. In effect, climate-warming scales up the

effects modeled by these coefficients and thus acts to shift all these phenomena towards a more dynamic state, in terms of conventional time.

Within living things, matter-based structures may be perceived to have slower rates of change than energy-based flows. (Of course, Albert Einstein showed that matter and energy are ultimately inter-convertible, but we are dealing here with matter and energy as distinct, as at the usual temperatures in ecosystems.) But at the level of enzymes operating under direction by the gene complex, these matter-related and energy-related rates must meld. At this level, the information content of individual genes, operating under poorly understood guidance by the whole genetic complex or genome, may have long permanence. Some of the genes in humans, say, appear to be the linear descendants of genes of primitive organisms that first emerged a billion years ago. So the information of genes, that guides the flows of matter and energy through the structures of a living system, may be the most permanent or structural part of that living system. We use the term *heritage* for such structures, as in our efforts to preserve biodiversity.

An ecosystem, whether in a natural, a cultural or a mixed state (*see* **Phase Shifts or Flip-flops in Complex Systems**, Volume 5), may be perceived to consist of semi-autonomous smaller systems nested within semi-autonomous larger systems. Each of these constrained self-organizing systems is called a *holon*. The whole nested structural complex of holons, with the many functional interactions among holons of all levels, is called a *holonarchy*. The latter connotes a kind of hierarchy in which each element or holon acts to maintain structural and functional integrity consistent with its information canon. That canon in turn contains nearly permanent gene-based structural information complemented by other kinds of information including temporary experience-based process learning at the level of an organism. With such a complex, use of the terms *structure* and *function* depends strongly on the reasons that motivate a researcher's study and where that study's primary focus falls within some larger systemic complex.

All the above is cast within an evolutionary perspective. Nowadays, science contains no alternative perspective to rival it. Structure and function are complementary conveniences for perceiving something in a system, and together the concepts help to clarify evolving systems. It should be apparent why science does not nourish a hope that it, i.e., science, will eventually provide a full understanding of such phenomena.

REFERENCES

Berlinski, D (2000) *The Advent of the Algorithm, the Idea that Rules the World*, Harcourt, New York.

Lin, P and Regier, H A (1995) Use of Arrhenius Models to Describe Temperature Dependence of Organismal Rates in Fish, in *Climate Change and Northern Fish Populations*, ed R J Beamish, *Can. Spec. Publ. Fish. Aquat. Sci.*, **121**, 211–225.

FURTHER READING

Keller, D R and Golley, F B (2000) *The Philosophy of Ecology: from Science to Synthesis*, University of Georgia Press, Athens, GA.

Regier, H A, Holmes, J A, and Pauly, D (1990) Influence of Temperature Changes on Aquatic Ecosystems: an Interpretation of Empirical Data, *Trans. Am. Fish. Soc.*, **119**, 374–389.

HENRY A REGIER *Canada*

Ecotones

Paul G Risser
Oregon State University, Corvallis, OR, USA

Ecotones are transitional areas between adjacent habitats or types of vegetation. For example, the fringe area between a forest and an adjacent field is called an ecotone. At a larger spatial scale, the timberline area between a conifer forest and the open alpine tundra on a mountaintop is an example of an ecotone. At the continental scale, the savanna zone between the eastern deciduous forest and the central plains grasslands in North America can be called an ecotone. Thus, ecotones are defined at various spatial scales.

Ecotones have always been of interest to scientists who study the relationships between climate and the distribution of plants and animals. At a mountain timberline, the climate and soil conditions are marginal for trees to survive. By studying the environmental conditions in this ecotone, it is frequently possible to understand the most significant factors controlling plant distribution. At the continental scale, scientists can study the ecotone between forests and grasslands, observing the behavior of various species in these regions during periods of drought and increased precipitation. These observations about ecotones and changing climate conditions have become particularly important recently as we learn about global climate change, and attempt to predict the biological consequences of changes in climate.

Within the last 20 years, scientists have learned that ecotones sometimes influence the movement of water, nutrients and soil particles across the landscape. For example, strips

of forest along streams are a type of ecotone called gallery forests or riparian forests. These ecotonal forests play several roles. For example, they may capture nutrients and sediments that would otherwise flow from adjacent fields into the stream or river. In other situations, these riparian forests produce large trees that eventually die, and during storms, fall into streams thus contributing crucial habitat for some fish species.

In summary, ecotones are transitional areas between different types of habitat or vegetation. Ecotones are scale-independent and can be defined at small spatial scales, such as local habitats, or at broad spatial scales across the continent. Ecotones may provide important habitats, serve as indicators of climate change, and influence the movement of plants, animals, water and materials across the landscape.

DEVELOPMENT OF THE CONCEPT OF ECOTONES

The concept of an ecotone as a transitional area between habitat types or vegetation types began in the first half of this century from three different perspectives over a range of spatial scales (Risser, 1995a). In the 1930s, ecotones were described locally and used in a very practical way, mostly by wildlife managers. Habitat descriptions of both mammals and birds frequently noted that certain species preferred areas that contained edges of some vegetation types or mixtures of habitats (Leopold, 1933). For example, habitat descriptions of white-tailed deer (*Odocoileus virginianus*) in North America or the fieldfare (*Turdus pilaris*) in Europe frequently mention a preference for the ecotones between forests and open fields. In most of these instances, animals select ecotones as preferred habitat because of the combination of food, cover and breeding conditions available within these transitional areas.

During the same time period, scientists began to examine the transition between the high-altitude forested areas and the adjacent alpine tundras of mountains (Griggs, 1938). This interest arose because it appeared obvious that trees were at their limit of physiological existence, and conditions of weather, soils, and topography restrained their growth farther up the mountain. Thus, the ecotone between forests and tundra offered an opportunity to study the environmental factors controlling plant distribution.

In studying these montane ecotones, several instances were noted where dead trees were found above the current tree line. This indicated that in previous times, trees had survived in habitats now occupied by tundra, that is, in areas where trees were not growing today. Among the possible explanations were that disturbances such as fire had influenced the distribution of vegetation, and/or that the climate had changed so trees were no longer able to grow at the higher elevations. A more detailed description of tree-line responses to climate change is presented later in this discussion.

At the continental spatial scale, scientists were interested in the transitions between vegetation types, such as between the eastern deciduous forest and the more western grassland biomes in central North America (Weaver and Albertson, 1956). This interest increased during and after the droughts of the 1930s and 1950s. Under drought conditions trees at the edge of the grassland died, and in effect, the grasslands moved farther eastward. The more drought-tolerant grasses of the central Great Plains became more conspicuous in the normally more humid grasslands of the eastern Great Plains. In other words, the ecotones between the forest and grasslands, and between different kinds of grasslands in the Great Plains, were not stable but rather moved depending upon the climatic conditions. Similar considerations of continental-scale ecotones have been considered elsewhere, for example, the productive zone at the edge of the receding ice edge of the Antarctic continent (Smith and Nelson, 1985).

Ecotones frequently support relatively high levels of biodiversity because they contain habitat characteristics of both adjacent ecological units and have unique characteristics of the ecotone itself. Within ecotones at all spatial scales, the ranges of plants and animals frequently overlap and thus species from two or more biomes may occur together. This combination of species distributions and the unique habitat characteristics found in ecotones frequently support a distinctive flora and fauna leading to higher biological diversity. Despite this intuitively logical explanation, evidence of high biodiversity in ecotones, while available is not plentiful (Delcourt and Delcourt, 1992). Moreover, because habitats are usually fragmented within broad ecotones, within-population genetic variation may be constrained because of isolation or small local habitat size (Holsinger, 1993).

At local and landscape spatial scales, ecotones frequently control or modulate the movement and migration of plants and animals. An example is found along the Garonne River in southwestern France (Decamps, 1993). Harvey's groundsel (*Senecio harveyanus*), introduced from South Africa, was found along the Garonne River in 1936. During the next 50 years, the species migrated along the ecotonal forests of the Garonne and Adour rivers, both upstream and downstream (Figure 1). The groundsel is a weedy species and the periodic flooding of the rivers provided a sequence of disturbed habitats along the ecotone for colonizing and migration.

Ecotones can also influence the movement of water, nutrients and other materials. These controls occur at the local or landscape levels, frequently in the ecotonal forests along streams and rivers. Peterjohn and Correll (1984) studied a landscape in the eastern United States that consisted of agricultural fields with ecotonal riparian forests along

the streams. The flow of nitrogen and phosphorous from the fields into the streams was significantly modified by these riparian ecotones. Much of the phosphorous that was attached to soil particles in erosional surface flow from the fields was trapped in the riparian forests. Similarly, a significant amount of the nitrogen washed from the fields was captured by the ecotones. Nitrogen was taken up by the trees from the soil water or lost from the forest floor by denitrification. There are numerous examples of riparian forests capturing soil particles from erosion and thus reducing the sedimentation in streams.

The concept of the ecotone has developed over the past 50 years or so from a relatively static, habitat descriptor to one of an active participant in the ecological processes occurring at several spatial scales. With the observations that ecotones apparently moved under different climate conditions, it became obvious that ecotones are dynamic and responsive to environmental conditions. Studies demonstrated that ecotones can influence the movement of plants, animals, water and various materials. Ecotones were also found to support relatively high levels of biodiversity at the community and species levels, although diversity at the genetic level has not been conclusively characterized. As described below, our understanding of ecotones continues to unfold and to become more complex. The relationships between climate change and ecotones are complicated but promise to be instructive for anticipating possible changes in ecological systems in the forthcoming decades.

Based on these studies conducted earlier this century, it is now possible to define ecotones more clearly. First, ecotones are not static, but are responsive to climate change. Second, ecotones are spatially scale-independent, that is, they are defined at many spatial scales ranging from local habitats to biomes. Third, ecotones serve as specific habitats for some organisms, and they mediate the flow of organisms and materials across the landscape.

Gosz (1993) has provided a useful framework for a hierarchical organization of ecotones based on their spatial scale (Table 1). This framework also postulates the possible major controls over the structure and function of ecotones. Global climate change would be expected to influence ecotones directly at the larger spatial scales, i.e., at the scales of biomes or landscape. Controls or constraints on ecotones at smaller spatial scales would be caused by the effect of macroclimate on the microclimate of smaller patches of vegetation or of individual plants. Gosz suggests that there are more recognizable controls at the smaller spatial scales.

RESPONSES OF ECOTONES TO CLIMATE CHANGE

Early studies of ecotones and their response to climate change were relatively simple, such as the observations made by Weaver and Albertson (1956) who noted that the boundary between the North American eastern deciduous forest and the adjacent Great Plains grassland shifted

Table 1 Ecotone hierarchy for a biome transition area. Each level in the ecotone hierarchy has a range of constraints and interactions between the constraints: × symbolizes interactions between constraints. (Reprinted from Gosz, 1993)

Ecotone hierarchy	Probable constraints
Biome ecotone	Climate (weather) × topography
Landscape ecotone	Weather × topography × soil characteristics
Patch ecotone	Soil characteristics × biological vectors × species interactions × microclimatology
Population ecotone (plant pattern)	Interspecies interactions × intraspecies interactions × physiological controls × population genetics × microtopography × microclimatology

during the droughts of the first half of the 20th century. More recent studies have begun to demonstrate additional complexity in the relationships between ecotones and climate change.

At the biome level, water balance and cold winter temperatures usually control the ecotone position (Neilson, 1993), and both of these processes are affected by global climate change. Delcourt and Delcourt (1992) propose that the following are most likely to influence the response of ecotones to climate change: rate and magnitude of environmental change; the physiological tolerance of the dominant species; the proximity, competitive abilities and dispersal capabilities of the invading species or pathogens; differences in soil characteristics; and the ability of organisms to withstand a different suite of environmental conditions. These processes are similar to those suggested by Gosz (1993) at the finer biome scales.

In the Chihuahuan desert in southwestern United States, there is a broad ecotone between desert grassland and shrubland (Grover and Musick, 1990). In general, the shrubland invades the more semiarid grassland under drier climatic conditions. This invasion was examined using photographic records taken over a 50 year period. The shrubland vegetation invaded the grassland and within the ecotone itself, the patches of shrubs expanded in size. The shrubland near the ecotone became more like the interior of the shrubland, thus not only was there a movement of the major vegetation type, but the size of the patches of vegetation changed during the response of the ecotone (Gosz, 1993). As more studies are conducted, it has become more obvious that the response of ecotones to climate change does not simply result in the movement of a coherent zone across the landscape (Hobbs, 1994).

Various authors have attempted to construct models to predict the response of ecotones to climate change. These predictions have involved both the behavior of individual species (e.g., Peters and Darling, 1985; Payette and Filion, 1985) and entire vegetation types (e.g., Emanuel *et al.*, 1985; Davis and Zabinski, 1992). On mountain tops at the tree line, it would be expected that increased concentrations of carbon dioxide and higher ambient temperatures would allow lower altitude plants to move into the colder regions at higher elevations. However, at the montane tree line wind stress is greater so lower-elevation species may not be able to expand their range (Woodward, 1993). Using a different model for mountain ecotones at lower altitudes, Noble (1993) has postulated that fires in the higher elevation grasslands would reduce the likelihood of trees moving upward with global warming.

In a particularly useful study, Lloyd and Graumlich (1997) reconstructed the tree line in the southern Sierra Nevada mountains in California, USA. During the 3500 year period of study, the tree line was higher than now, but there were two notable declines. The first decline in the elevation of the tree line occurred during a relatively warm period. However, these warmer temperatures were accompanied by drought conditions which caused mortality of adult trees. The second decline occurred during a 400-year period of colder temperatures which reduced reproduction and also caused adult mortality.

Throughout the 3500-year period under study, the abundance of the trees changed continuously, both increasing and decreasing. The tree line, however, remained static for long periods of time, and then moved only in brief periods. Thus, the high-elevation tree line itself may not be as responsive to climate change as the absolute and relative abundance of individual species. In fact, the changes in tree abundance preceded the changes in tree line by about 400 years (Lloyd and Graumlich, 1997). These results are consistent with those of Payette and Filion (1985) who found that climate warming led to changes in forest density rather than changes in position of the ecotone.

Iverson and Prasad (1998) took a different approach to the question of how vegetation pattern would be expected to respond to climate change. They combined the habitat conditions of 80 tree species with the predicted conditions with a doubling of ambient CO_2 in the atmosphere. Depending upon the actual climate change scenario, 30 tree species would expand their range or become more important in a community by at least 10%. Nearly half of the tree species (36/80) would have their range shifted by at least 100 km, and the range of seven species would move more than 250 km (*see* **Plant Dispersal and Migration**, Volume 2).

In reconstructing species changes in the Holocene (*see* **Holocene**, Volume 1), most species changed their range,

but the rates and directions were largely individualistic. Species assemblages changed and at these large spatial scales it was difficult to discover whether competitive and dispersal mechanisms played an important role. However, these changes occurred over thousands of years and through landscapes largely devoid of human-constructed barriers. Responses to global change will occur with faster climate changes and over fragmented landscapes (Iverson and Prasad, 1998).

Neilson (1993) has offered a generalized model to predict the ways in which biodiversity and habitat structure of ecotones might respond to climate change. If global warming is accompanied by water stress, the entire landscape should become fragmented because every variation in topography and soil becomes important to water balance and survival. Under these conditions, many different survival strategies operate and diversity remains high in the ecotone with an eventual equilibrium occurring at a new ecotone. If global warming is not accompanied by water stress, then fragmentation should not occur and organisms will simply move in a more or less homogeneous pattern in response to changing temperature. Thus, in Neilson's model, the degree to which an ecotone is an effective indicator of climate change will depend upon interactions among climatic and habitat factors. With more habitat heterogeneity, it is less likely that the ecotone will respond in a uniform manner.

Ecotones may have characteristics that are closely tied to changes in the global climate. Decamps (1993) examined the behavior of ecotonal riparian and floodplain forests along four rivers (Figure 2). Denitrification occurred in the saturated soils of the riparian forests and the floodplain forests, just as in the previously cited study in North America (Peterjohn and Correll, 1984). If climate change affects the flows of rainfall and river flows, there would be direct connections with the amount of denitrification and loss of nitrogen to the atmosphere. Thus, the relationship between ecotones and climate change is not limited to the distribution of organisms, it also includes the ecological roles of ecotones on the landscape (Pinay and Decamps, 1988).

Because organisms are at the edge of their ranges at ecotones, it would be expected that ecotones would represent ideal locations to monitor the effects of global climate change. However, as noted in the above empirical and modeling studies, the relationship may not be a simple one. First, there may be countervailing forces (such as drought, wind, fire or pathogens) that reduce the ability of organisms to respond directly to climate change. Second, ecotones are frequently fragmented and their spatial pattern is altered in response to climate change. As a result, measurement of ecotonal responses may require careful analysis of spatial patterns (Hansen and diCastri, 1992). Third, during glacial periods individual species have responded as individuals (Davis, 1989; Davis and Zabinski, 1992) so there is little reason to expect that intact ecotones would move in response to climate change even if the rate of climate change is significantly faster than during glacial times. Fourth, the abundance of species may be more responsive than the position of the ecotone to climate changes. Thus, it is likely that individual species may be a more effective target for measuring responses to climate change rather than the ecotone itself. However, ecotones probably represent ideal locations for identification of organisms for these studies (Risser, 1995b).

Ecotones that might be suitable for monitoring global climate change are probably those that occupy steep environmental gradients and those that recover rapidly from disturbances. Because these steep environmental gradients result in minimal fragmentation, they would be expected to differentiate species and ecological processes in finer spatial scales. The rapid recovery of organisms and processes from disturbance suggests that responses to climate change are less likely to be obscured by the impacts of disturbance.

In summary, the definition and understanding of ecotones has progressed significantly from the early attention to wildlife habitat and descriptions of boundaries among vegetation types. Ecotones represent an information-rich ecological system in which to understand responses to climate

change. These responses are complex, but there is sufficient research information on which to base observational and experimental studies and to draw initial conclusions.

REFERENCES

Davis, M B (1989) Insights from Paleoecology on Global Climate Change, *Bull. Ecol. Soc. Am.*, **70**, 222–228.

Davis, M B and Zabinski, C (1992) Changes in Geographical Range Resulting from Greenhouse Warming: Effects on Biodiversity in Forests, in *Global Warming and Biological Diversity*, eds R L Petersand and T E Lovejoy, Yale University Press, New Haven, CT, 297–308.

Decamps, H (1993) River Margins and Environmental Change, *Ecol. Appl.*, **3**, 441–445.

Delcourt, P A and Delcourt, H R (1992) Ecotone Dynamics in Space and Time, in *Landscape Boundaries: Consequences for Biotic Diversity and Landscape Flows. Ecological Studies 92*, eds A J Hansen and F diCastri, Verlag, New York, 19–54.

Emanuel, W R, Shugart, H H, and Stevenson, M P (1985) Climate Change and the Broad-scale Distribution of Terrestrial Ecosystem Complexes, *Climate Change*, **7**, 29–43.

Gosz, J R (1993) Ecotone Hierarchies, *Ecol. Appl.*, **3**, 369–376.

Griggs, R F (1938) Timberlines in the Northern Rocky Mountains, *Ecology*, **19**, 548–564.

Grover, H D and Musick, H B (1990) Shrubland Encroachment in Southern New Mexico, USA: an Analysis of Desertification Processes in the American Southwest, *Climate Change*, **12**, 165–190.

Hansen, A J and diCastri, F, eds (1992) *Landscape Boundaries: Consequences for Biotic Diversity and Ecological Flows. Ecological Studies 92*, Verlag, New York.

Hobbs, R J (1994) Dynamics of Vegetation Mosaics: Can We Predict Responses to Global Change? *Ecoscience*, **1**, 346–356.

Holsinger, K E (1993) The Evolutionary Dynamics of Fragmented Plant Populations, in *Biotic Interactions and Global Change*, eds P M Kareiva and J G Kingsolver, Sinauer Associates, Sunderland, MA, 198–216.

Iverson, L R and Prasad, A M (1998) Predicting Abundance of 80 Tree Species Following Climate Change in Eastern United States, *Ecol. Monogr.*, **68**, 465–486.

Leopold, A (1933) *Game Management*, Charles Scribner, New York.

Lloyd, A H and Graumlich, L J (1997) Holocene Dynamics of Tree-line Forests in the Sierra Nevada, *Ecology*, **78**, 1199–1210.

Neilson, R P (1993) Transient Ecotone Response to Climatic Change: Some Conceptual and Modeling Approaches, *Ecol. Appl.*, **3**, 385–395.

Noble, I R (1993) A Model of the Responses of Ecotones to Climate Change, *Ecol. Appl.*, **3**, 396–403.

Payette, S and Filion, L (1985) White Spruce Expansion at the Tree Line and Recent Climate Change, *Can. J. For. Res.*, **15**, 241–251.

Peterjohn, W T and Correll, D L (1984) Nutrient Dynamics in an Agricultural Watershed: Observations on the Role of a Riparian Forest, *Ecology*, **65**, 1466–1475.

Peters, R L and Darling, J D (1985) The Greenhouse Effect and Nature Reserves, *BioScience*, **35**, 707–717.

Pinay, G and Decamps, H (1988) The Role of Riparian Woods in Regulating Nitrogen Fluxes Between the Alluvial Aquifer and Surface Water: a Conceptual Model, *Regul. Rivers Res. Manage.*, **2**, 507–516.

Risser, P R (1995a) Recommendations for Managing Ecotones, *Ecol. Int.*, **22**, 95–102.

Risser, P R (1995b) Ecotones: the Status of the Science, *BioScience*, **45**(5), 318–325.

Smith, W O and Nelson, D M (1985) Phytoplankton Bloom Produced by a Receding Ice Edge in the Ross Sea: Spatial Coherence with the Density Field, *Science*, **227**, 163–166.

Weaver, J E and Albertson, F W (1956) *Grasslands of the Great Plains: Their Nature and Use*, Johnsen, Lincoln, NE.

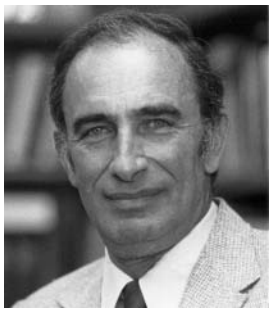
Woodward, F I (1993) The Lowland-to-upland Transition-modeling Plant Responses to Environmental Changes, *Ecolog. Appl.*, **3**, 404–408.

Ecotron

see Controlled Environment Facilities in Global Change Research (**Volume 2**)

Ecotype

see Tundra (**Volume 2**)



Ehrlich, Paul

(1932–)

Paul Ralph Ehrlich, American ecologist, evolutionist, and environmental scientist is internationally renowned for the breadth, originality, and importance of his scientific insights. His far reaching contributions shaped profoundly the way people approach problems in fields ranging from biological and cultural evolution to conservation biology, global change and ecological economics. Ehrlich is also distinguished for his social leadership on issues of environment and society. His tireless efforts greatly elevated public awareness worldwide of the importance of: population

growth and over-consumption as factors in environmental impacts; the meaning of biodiversity loss to humanity; the value of ecosystems as capital assets, delivering services to society; the relationship between socioeconomic equity and environmental sustainability; and numerous issues in evolution, such as resistance to antibiotics and genetic determinism.

He received his BA from the University of Pennsylvania (1953) and his doctorate from the University of Kansas (1957). His early work on the higher classification of butterflies remains the standard to this day, but after taking part in the numerical taxonomy revolution of the 1950s, he joined the faculty of Stanford University in 1959 and turned his major attention to studies of the biology of butterfly populations. His continuing work on checkerspot butterflies, integrating studies of ecology, evolution, and behavior, is considered a model of long-term research on natural populations. In the 1960s he (along with P H Raven) founded the field of *coevolution*, and he was a pioneer in establishing *conservation biology*. Ehrlich his best known to the public for his work in *human ecology*. In the 1960s and 1970s, with his wife Anne and colleague John P Holdren, he mapped out the framework for modern environmental science. More recently he has focused his efforts on developing the interface between the natural and social sciences, especially economics, to help the transition of society to an environmentally sustainable course.

Ehrlich has done fieldwork on every continent, and from alpine zones to the sea floor, carrying out research on plants, invertebrates, reef fishes, reptiles, birds, and mammals. He is the author of several dozen books and some 800 scientific papers and articles. In recognition of his scientific distinction and long crusade for environmental sanity and social justice, he has received virtually every prize and honor the scientific and environmental communities can bestow, including the Craford Prize of the Swedish Academy of Sciences, an explicit substitute for the Nobel Prize in fields where the Nobel is not offered.

FURTHER READING

Ehrlich, P R (2000) *Human Natures: Genes, Cultures, and the Human Prospect*, Island Press, Washington, DC.

Ehrlich, P R, Ehrlich, A H, and Daily, G C (1995) *The Stork and the Plow: The Equity Answer to the Human Dilemma*, Putnam, New York.

Ehrlich, P R, Harte, J, Harwell, M A, Raven, P H, Sagan, C, Woodwell, G M *et al.* (1983) Long-term Biological Consequences of Nuclear War, *Science*, **222**, 1293–1300.

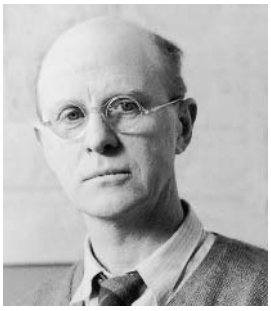
Ehrlich, P R and Holdren, J (1971) Impact of Population Growth, *Science*, **171**, 1212–1217.

Ehrlich, P R and Raven, P H (1964) Butterflies and Plants: a Study in Coevolution, *Evolution*, **18**, 586–608.

GRETCHEN C DAILY *USA*

El Niño Ecological Impacts

see El Niño and La Niña: Causes and Global Consequences (**Volume 1**)



Elton, C S

(1900–1991)

Charles Sutherland Elton is one of the founders of Animal Ecology. His major contributions stemmed from several books, all of which he wrote quickly and with crystal clarity. The first of these was *Animal Ecology* (1927). This set out the key features of animal communities in terms of their species composition, species richness and zonation. He defined the concept of a species' niche as the animal's role in the community, analogous to a person's occupation in human society and discussed the segregation of species by their niches described in terms of time, weather and their size. He recognized that the relative numbers at each level of a food chain constitute a "pyramid of numbers". He also explored the main concepts in population ecology; here, stimulated largely by his own experience in the arctic, he pointed to the cycling of natural populations. He highlighted the role of natural enemies in the regulation of many populations, but suggested that for those at the top of a food chain, territory should be the limiting factor. There is almost no area of modern ecology that is not introduced in this work; he records that he wrote it in 85 days. Between 1930 and 1933 he published three small books that expanded on or summarized topics in Animal Ecology. His concerns about nature conservation led to another book that was truly seminal: *The Ecology of Invasions by Animals and Plants* (1958). It was derived from a series of three broadcasts he had given the previous year. In this work he contrasted the likely outcome of an invasion of a continent by an introduced species with the impact on an island and analyzes the reasons why invasions fail or succeed. The book ends with two chapters devoted to the ideas of conservation. He also published two large and more detailed books that stemmed from his own research: *Voles, Mice and Lemmings* (1942) and *The Pattern of Animal Communities* (1966); both are mines of information but have not had the lasting impact of his other works.

He published relatively few substantial research papers. As was appropriate for someone who considered ecology to be "scientific natural history", he authored many short notes on natural history observations. He founded and for many years edited the *Journal of Animal Ecology*.

Elton was born in Manchester, UK, but for most of his childhood his father was Professor of English Literature at Liverpool University. A large part of his vacations was spent in the country and, a naturalist from an early age, these allowed him to develop his interests. By the time he entered Oxford in October 1919 he was a very knowledgeable natural historian. The course in comparative anatomy failed to arouse his enthusiasm and he was only saved from relative failure through the influence of his tutor, Julian Huxley. In the summer of 1921 he accompanied Huxley and others on an expedition to Spitzbergen; he wrote a brilliant report on his observations there and on the strength of this continued at Oxford, eventually (1936) being appointed Reader in Animal Ecology. He founded the Bureau of Animal Population there and was its Director until his retirement in 1967. He was elected a Fellow of the Royal Society in 1953 and received the Tyler Award in 1976. Very quietly spoken, he had an unpretentious and gentle manner, but these qualities belied the strength with which he held his views.

RICHARD SOUTHWOOD *UK*

ENRICH (European Network for Research in Global Change)

see European Network for Research in Global Change (ENRICH) (Volume 2)

Environment

The term *environment* refers to the physical, chemical, and biotic factors that affect an organism or an ecosystem and ultimately determine its form or structure and survival. All organisms living in a particular area, including humans and anything that influences or is influenced by those organisms are part of the environment.

Constituents of the environment can be classified as either *abiotic* (including climate, minerals, soil, sunlight, water, and all other nonliving elements) or *biotic* (i.e., living organisms).

The flow of energy and the cycling of nutrients link these constituents together. The major components of the Earth's physical environment include the atmosphere, climate, weather, continental landforms, hydrosphere, cryosphere and oceans. The relationship between the principal physical components of the environment, and the major ecosystems of the Earth is mediated through the biosphere.

Human interference in the global environment is widespread and accelerating. Most of this interference derives from three basic contributing factors: human population growth, pollution, and misuse of resources and natural ecosystems. Because these three factors are interrelated, environmental problems are extremely complicated. Another factor in global environmental change is that the world is changing at an increasing rate, with internationally coordinated environmental stewardship seeming to fall ever further behind social and economic development. Environmental gains from better policies and improved technology are being outstripped by the pace and scale of human population growth and economic development. Major environmental issues include climatic change, biodiversity loss, stratospheric ozone depletion, nitrogen overloading, air and water pollution, and deforestation.

RICHARD A FLEMING *Canada*

EROS (Remotely Sensed Land Data)

see Data Banks: Archiving Ecological Data and Information (Volume 2)

European Network for Research in Global Change (ENRICH)

Pierre Mathy
European Commission, Brussels, Belgium

ENRICH is an initiative developed by the European Commission, within the European Framework Programme for Research, Technological Development and Demonstration (RTD) activities.

The mission of ENRICH can be characterized by three broad aims:

- *to improve and facilitate research networking and capacity building between European researchers and the wider international research community in the area of global change;*
- *to improve the flow and exchange of information on global change research with a particular emphasis on the relevance and value of research investment to European policy, society and economic activity; and*
- *to provide a focal point for European collaboration with complementary international initiatives such as the Inter-American Institute and the Asia Pacific Network.*

DEFINITION OF GLOBAL CHANGE

Global change research seeks to understand the integrated Earth system in order to: (i) identify, explain and predict natural and anthropogenic changes in the global environment; (ii) assess the potential regional and local impacts of those changes on natural and human systems; and (iii) provide a scientific basis for the development of appropriate technical, economic and societal mitigation/adaptation strategies. In this context, global change encompasses climate change; changes in atmospheric physics and chemistry (e.g., ozone depletion, tropospheric pollution); changes in land use and cover; variability in marine environments and ice cover; changes in biogeochemical cycles; impacts of these changes on natural balances (ecosystems), on natural resources (freshwater, forest products, crops, marine resources), and on socio-economic balances (sustainable development, industrial development, human health, quality of life); biodiversity related issues.

IMPLEMENTATION: PROJECTS SELECTION

ENRICH was initially conceived as a joint initiative primarily based on three specific programs of RTD: Environment and Climate; Marine Science and Technologies; and Cooperation with Third World Countries and International Organizations, part of the 4th Framework Programme (FP4) for RTD activities. ENRICH was implemented through two calls for proposals followed by evaluation and selection procedures applied within FP4.

Those calls sought proposals for additional catalytic activities, accompanying measures and supporting initiatives falling in the remits of the RTD programs indicated above. Proposals were sought to fill gaps that cannot be supported by individual research projects alone and that:

- add value to ongoing research activities carried out in the area of global change through the promotion of voluntary scientific coordination including data compatibility schemes and the corresponding information systems and information sharing/dissemination arrangements; or
- prepare the ground for more coherent and integrated research undertakings based on sound and functional research partnerships that might be considered for implementation at a later stage.

ACTION TYPES

The proposals falling within the remit of ENRICH had to include in a scientifically coherent and operationally feasible way, one or more of the following types of action:

- Development of cooperative links and networking actions, in particular with a view both to strengthening integrative and interdisciplinary approaches in the European Union (EU) and associated states and also promoting partnerships between their researchers and those in other countries.
- Improving visibility and exchange of data and scientific information; collation of disparate datasets; strengthening harmonized data collection, distribution and modeling strategies; integration and consolidation of research results; development of information systems and information sharing arrangements.
- Improving research capacity to address the issues of regional implications of global change (impacts on natural resources, implications for natural resources management, etc.) and for policy analysis relating to the global implications of regional changes in natural and socio-economic resources.
- Identification of regional gaps such as the need for data, experimental settings, methodological improvements and regional research priorities to improve the integrated knowledge of global change processes and impacts and their policy implications.
- Contributions to the development of science agendas or science plans for interdisciplinary research on global change of mutual interest to the EU and its partners, with particular attention to Eastern and Central European regions; Mediterranean Regions; Arctic and Boreal Regions.
- Promoting and improving the packaging and diffusion of research results on global change in a manner responsive to the specific needs of users, such as policy makers; natural resources managers; industrial, economic and societal actors.
- Projects selected continued until 2000. A new call for proposals was launched in November 1999.

More information may be obtained from the author: pierre.mathy@dg12.cec.be.

Eutrophication

Eutrophication is the enrichment of streams, lakes, ponds, bays and estuaries with nitrogen and/or phosphorus, the two macronutrients that normally limit the photosynthesis of phytoplankton. The nutrients come from leaching of inorganic fertilizers, from manures, human wastes, industrial processes and from atmospheric deposition. Water transparency and color are the most obvious indicators of eutrophication: transparent oligotrophic (nutrient poor) waters support low plant productivity and appear either blue or brown (when stained in peaty regions). Eutrophic waters have high primary productivity as large amounts of phytoplankton make them turbid and limit their transparency to less than 50 cm.

Advanced eutrophication is marked by blooms of cyanobacteria and siliceous, scum forming, and also potentially toxic, algae. Eventual decomposition of this phytomass creates hypoxic or anoxic conditions near the bottom. The most worrisome consequences range from offensive taste and odor of drinking water and formation of trihalomethanes during water chlorination to serious health hazards to livestock and people ingesting soluble neurotoxins and hepatotoxins released by decomposing algal blooms. Reduced fish yields, extensive summer fish kills and changed composition of fish species in affected water bodies are common. Another particularly offensive consequence of eutrophication includes the growth of thick coats of algae on any submerged substrates, be they aquatic plants, stones, docks or boats.

Coastal eutrophication now threatens the Great Barrier Reef, the world's largest coral formation, by smothering them with algae as well as by promoting the survival and growth of the larvae of *Acanthaster planci*, the crown-of-thorns starfish, that has recently destroyed large areas of the reef. And as the Baltic unfortunately demonstrates, even a whole sea can become eutrophic: a third of the sea bottom of the Baltic proper (the southern part of the sea) is now intermittently deprived of oxygen. The worst affected offshore zone in North America is a large region of the Gulf of Mexico where the nitrogen load brought by the Mississippi and Atchafalya rivers creates every spring a large hypoxic zone that kills many bottom dwelling species and drives away fish.

VACLAV SMIL *Canada*

Evolutionary Processes in Ecosystems

Michel Loreau
Ecole Normale Supérieure, Paris, France

Environmental changes are major drivers of evolutionary changes. Although evolution is typically a slow process, there is increasing evidence that the time-scale of evolution can be shorter than previously believed under strong selective pressures.

The prediction and understanding of these evolutionary changes and their implications, however, require identification of the proper context of constraints within which natural selection operates. The ecosystem provides such a context, because an organism's environment is made up of interacting biotic and abiotic components which constitute a local ecosystem. Ecosystem processes can channel selection in different directions from those expected in the absence of these constraints. In particular, nutrient cycling is a key process which transmits predictable indirect effects in ecosystems. These indirect effects can be so strong that they prevail over direct effects and exert effective selective pressures on the species involved, provided that there is sufficient spatial heterogeneity in the system or trade-offs between traits associated with the direct and indirect effects.

Thus, the merging of evolutionary and ecosystem perspectives, which have been increasingly separated in modern ecology, is fundamental to predict the responses of ecological systems to environmental changes.

ENVIRONMENTAL CHANGES AND EVOLUTIONARY PROCESSES

All organisms experience environmental changes continuously. Many of these changes, however, are regular or frequent enough so that organisms can adapt to them. Diurnal or seasonal cycles, for instance, are so regular and predictable for most organisms that they are perceived as an integral part of a constant environment. Other changes are catastrophic or directional, and impose new environmental conditions on organisms, either suddenly or gradually. Most environmental changes generated by human activities – habitat destruction, biological invasions, changing atmospheric composition, global warming, and pollution – are of that kind. What distinguishes them from similar changes in the past is their rhythm and scale, which are, in many respects, unprecedented.

Faced with such changes, organisms have three basic alternatives: to migrate, adapt, or die. Catastrophic changes, such as habitat destruction, often result in

massive mortality, population extinctions, and eventually, species extinctions. This is the unfortunate fate of an increasing number of species today. More gradual changes, such as climatic changes, often result in progressive migrations and shifts in species distributions, as testified by palaeoecological records. A less obvious but widespread outcome is the adaptation to changed conditions through the evolution of species traits.

Evolution is ultimately the result of three basic processes: mutation, which generates genetic variability within populations; genetic drift, which increases genetic variability among populations; and selection, which reduces genetic variability selectively within populations. There is increasing evidence that environmental stress can directly affect mutation frequencies or the expression of accumulated *silent* mutations in organisms as different as bacteria or fruit flies (Pennisi, 1998; Rutherford and Lindquist, 1998), thereby increasing the genetic variability allowing rapid evolution. It is also known that habitat fragmentation and population subdivision favor genetic drift, and ultimately, speciation; this is the basis of the widely accepted model of allopatric speciation, in which new species are formed in geographically distinct areas. Finally, environmental changes modify the adaptive landscape of species and the selective pressures that are exerted on them; they determine the direction in which natural selection drives evolution. Thus, environmental changes are major drivers of evolutionary changes, and ultimately, of the generation of species diversity. This shows that they are not intrinsically detrimental to biodiversity. The current (and unfortunately, mostly future) biodiversity crisis is fundamentally a problem of the rhythm and scale of environmental changes. Massive numbers of species are currently driven to extinction without being replaced by new species because evolution simply cannot keep pace with the current environmental changes for most organisms.

Evolution is typically a slow process because it is based on changes in gene frequencies between generations. However, there is increasing evidence that the time-scale of evolution can be shorter than previously believed under strong selective pressures i.e., in the order of a few generations (Thompson, 1998). This is still long in *human* time-scales for long-living organisms such as trees, but it may be very fast for short-lived organisms such as bacteria. Even in vertebrates, significant evolutionary changes have been documented in the field over a period of a few years, using long-term experiments or observations of populations under changing conditions (Grant and Grant, 1995; Reznick *et al.*, 1997; Losos *et al.*, 1997). Small organisms with short generation times and high population abundances can even adapt to genuine environmental catastrophes. The most dramatic examples of such rapid evolutions include the resistance of bacteria to antibiotics and the resistance of pests to insecticides. These examples point to the limits of our ability to control nature, even when we orchestrate large-scale chemical or biochemical catastrophes to destroy the organisms in question. Thus, the potential for, and implications of, evolutionary changes resulting from environmental changes cannot be ignored.

THE ECOSYSTEM AS THE PROPER CONTEXT OF NATURAL SELECTION

Predicting and understanding these evolutionary changes and their implications, however, requires identification of the proper context of the constraints within which natural selection operates. Traditionally, evolutionists have considered constraints to be internal to the organisms, such as from allocations among competing needs. In the classical view, the environment is regarded as external to the organism and constant (Figure 1A). Although most evolutionary biologists today would probably agree that this view is an oversimplification of reality, it has been, and still is, widely used in theoretical evolutionary biology as an implicit conceptual framework because of its simplicity. The modern view of natural selection recognizes that organisms modify and interact with their environment, which generates an organism–environment feedback in the operation of natural selection (Figure 1B; Lewontin, 1983). However, in order to understand the full implications of this feedback, it is further necessary to resolve an organism's environment into its real physical, chemical and biological constituents and their interactions. This is what I call the ecosystem

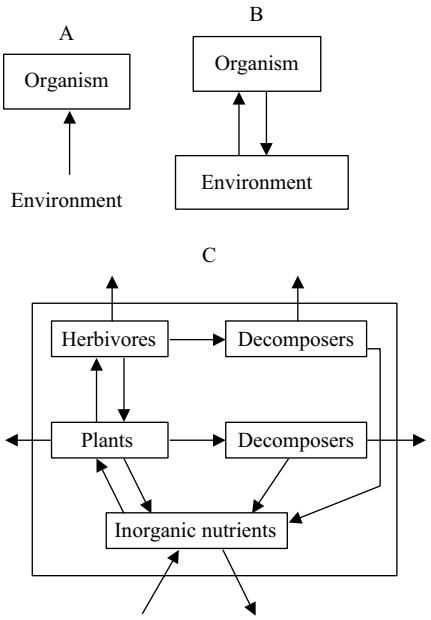


Figure 1 Three views on how natural selection operates: (A), the 'classical' view; (B), the 'modern' view; and (C), the 'ecosystem' view

view of natural selection, for an ecosystem is precisely a local system of interacting biotic and abiotic components (Figure 1C). Since each organism's environment is constituted by other organisms or components, the ecosystem concept contains both the organisms and their environments, and thus, transcends the duality between organism and environment. Recognizing the ecosystem as the proper context within which natural selection, and hence evolution, operates is a major challenge for ecology today, with important implications in both basic science and applied areas, such as conservation biology and ecosystem management. This challenge emphasizes the need to overcome the barrier that has increasingly separated population and evolutionary ecology on one hand, and ecosystem ecology on the other hand.

Feedbacks via ecosystem processes can channel selection in different directions to those expected in the absence of these constraints. As an example of a simple organism–environment feedback that has this property, consider the effect of changing soil nutrient supply on the evolution of plant antiherbivore defense. The classical resource availability hypothesis (Coley *et al.*, 1985) predicts that low resource availability favors plants with inherently slow growth rates, which in turn favors large investments in antiherbivore defense. This hypothesis has been influential and attractive because it seemed to explain the patterns of plant defense and herbivory in a wide range of ecosystems. It hinges, however, on a simple theoretical argument with a number of simplifying assumptions, in particular, the assumption that the amount of available resources is unaffected by plants. This implicit assumption of a constant environment led Coley *et al.* (1985) and subsequent authors to measure plant fitness by what they called the plant *realized growth rate*, which, in effect, is a potential growth rate which ignores the feedback generated by plant resource consumption. This fitness measure may make sense for pioneer species colonizing temporary environments, but is inappropriate for species competing for limited resources in more stable environments. Whenever plants have accumulated enough biomass to affect the amount of resources in their environment, they compete for these resources, and their growth hinges on their ability to tolerate low concentrations of the resource that is limiting. If the environment is homogeneous, fitness is determined by the ability to deplete the limiting resource (Tilman, 1982). If the environment is spatially structured, fitness is determined by the basic reproductive rate (Loreau, 1998).

Taking into account this plant–resource feedback explicitly changes the theoretical predictions altogether (Loreau and de Mazancourt, 1999). If the environment is homogeneous, with all plants having equal access to the limiting nutrient, and the plant–nutrient system is allowed to reach an ecological equilibrium, the evolutionary optimal defense investment is determined by features of the plant–herbivore interaction, but is independent of both plant maximum growth rate and nutrient availability. If the environment is spatially structured and competition obeys a competitive lottery for vacant sites, the outcome is strongly dependent on the factor that determines a plant's ability to produce successful propagule establishment in vacant sites, which, in itself, determines fitness. The optimal defense investment may then either increase, stay constant or decrease with nutrient supply (Figure 2). Thus, the classical view that lower resource availability favors evolution towards larger investment in antiherbivore defense has little theoretical support once the interaction between plants and their limiting resources is taken into account.

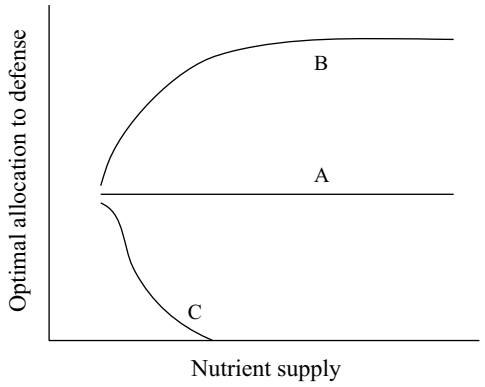


Figure 2 Evolutionary optimal plant allocation to defense as a function of nutrient supply. Optimal defense investment is: (A) independent of nutrient supply in most cases i.e., when either the environment is homogeneous, or the environment is spatially structured and plant tness is proportional to the part of biomass that is mobilized for growth or to total productivity; (B), increases with nutrient supply when plant tness is proportional to total biomass; and (C), decreases with increasing nutrient supply when plant tness is proportional to the part of productivity that is allocated to growth. (Loreau and de Mazancourt, 1999)

EVOLUTION OF INDIRECT INTERACTIONS IN ECOSYSTEMS

Many indirect interactions, and hence, of indirect organism–environment feedbacks, are likely to occur among organisms because of the complexity of ecosystems (Puccia and Levins, 1985; Wootton, 1994). These indirect effects can be weak or unpredictable (Yodzis, 1988), but some can be strong and predictable. In particular, material cycling is a key ecosystem process that drives a circular causal chain in ecosystems, thus transmitting predictable indirect ecological effects and evolutionary constraints to their component species. Thus, plants (*autotrophs*) and decomposers (*heterotrophs*) are involved in an indirect mutualistic interaction through the cycling of limiting nutrients, such as nitrogen or phosphorus. The ecological and evolutionary

dynamics of this indirect interaction can explain broad-scale successional and evolutionary trends in the functional properties of ecosystems, in particular those towards increased productivity and biomass, a decreased productivity/biomass ratio, and tighter nutrient cycles (Loreau, 1998).

A more controversial, but illuminating case of the mutual dependency between autotrophs and heterotrophs generated by material cycling is provided by plant–herbivore interactions. These interactions have been traditionally regarded as antagonistic because herbivores have a negative direct effect on plants through biomass consumption. This view, however, has been challenged by the *grazing optimization hypothesis*, which states that primary productivity, or even plant fitness, is maximized at an intermediate rate of herbivory (Owen and Wiegert, 1976; McNaughton, 1979; Hilbert *et al.*, 1981). One mechanism capable of producing grazing optimization is nutrient cycling, which mediates a positive indirect effect of herbivores on plants. Although this hypothesis has been hotly debated (e.g., Silvertown, 1982; Belsky, 1986; Paige and Whitham, 1987; Bergelson and Crawley, 1992; Belsky *et al.*, 1993; Lennartsson *et al.*, 1997), recent theoretical work has been able to identify general rules that govern grazing optimization in the form of increased primary production through cycling of a limiting nutrient (Loreau, 1995; de Mazancourt *et al.*, 1998). Qualitatively, this requires that: (1), the proportion of nutrient lost while flowing along the herbivore recycling pathway is sufficiently smaller than the proportion of nutrient lost while flowing in the rest of the ecosystem; and (2), nutrient inputs into the system exceed a threshold value, which depends on the sensitivity of the plant uptake rate to soil mineral nutrient. These results are very general; they do not depend on the structure of the ecosystem or on the functional form of herbivore consumption. They are also potentially relevant to natural ecosystems (de Mazancourt *et al.*, 1999).

Does this imply that ecosystem-level constraints make the plant–herbivore interaction actually mutualistic, not antagonistic? The evolutionary consequences of grazing optimization, and of ecological indirect interactions in general, are complex, for two main reasons. First, increased plant productivity does not necessarily translate into increased plant fitness. Second, when it does, it is not absolute, but relative fitness that counts. If two plant types (species or genotypes) are mixed, one of them being tolerant (*mutualistic*) and the other resistant (*antagonistic*) to herbivory, the resistant type is expected to outcompete the tolerant type because it benefits from the positive indirect effect of increased nutrient cycling, but does not suffer the negative direct effect of herbivore consumption. As a result, tolerance should not evolve, even though it is indirectly beneficial.

Two factors, however, counteract this advantage of antiherbivore defense. First, the spatial structure of the plant–herbivore system can generate spatially heterogeneous nutrient cycling (de Mazancourt and Loreau, 2000). If herbivores recycle nutrients in the vicinity of the grazed plants, or plants from the same type are aggregated, herbivores tend to recycle proportionally more nutrient on the plants that are more heavily grazed, thus augmenting the indirect benefit of grazing for the grazed plants. Evolution is then governed by the balance between two conflicting levels of selection (Wilson, 1980): (1), individual selection within patches, which favors the resistant type over the tolerant one because it has a higher relative fitness; and (2), group selection between patches, which favors patches with a higher proportion of the tolerant type because they have a higher average absolute fitness. The outcome of evolution depends on the strength of spatial aggregation and patch size; tolerance to grazing will evolve provided that the spatial aggregation is strong enough or the patch size is small enough (de Mazancourt and Loreau, 2000).

The second factor that counteracts the advantage of antiherbivore defense is its cost in terms of nutrient investment, which generates a trade-off between defense and nutrient uptake in plants. A theoretical study of plant adaptive dynamics (Dieckmann, 1997) in a spatially structured model ecosystem shows that, for most ecologically plausible trade-offs, plant evolution then leads to a single *continuously stable strategy* i.e., a strategy to which evolution converges and which cannot be invaded by any other close strategy (de Mazancourt *et al.*, in preparation). This evolutionary strategy has complex relationships with the strategies that maximize plant production or plant biomass, depending on ecosystem parameters. Due to this complexity, different ecological and evolutionary scenarios of herbivore addition or removal are possible, which highlight the ambiguity of the notion of 'mutualism'. It is useful to distinguish two types of mutualism: an ecological mutualism, in which each species gains a benefit from the presence of its partner in the absence of any evolutionary change, as revealed e.g., by an ecological press perturbation; and an evolutionary mutualism, in which the mutual benefit persists even after evolution has occurred (de Mazancourt *et al.*, in preparation). The conditions for an evolutionary mutualism are more stringent than those for an ecological mutualism because interacting species may have evolved a mutual dependence, so that the removal of one species may have a negative impact on the other in the short term, but this negative impact may disappear after each species has had the opportunity to evolve and adapt to the new conditions created by the absence of its partner (Douglas and Smith, 1989; Law and Dieckmann, 1998).

When a plant's reproductive ability is determined by its productivity, herbivory is indeed capable of improving plant performance on both an ecological and an

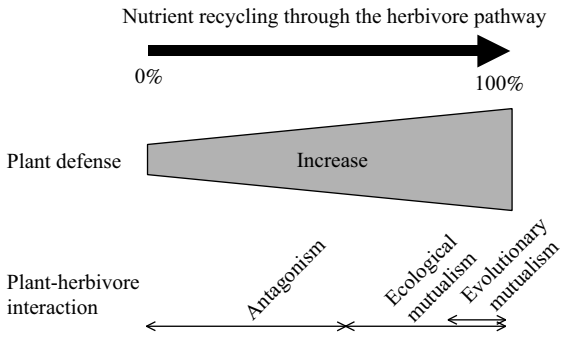


Figure 3 Changes in the plant-herbivore interaction along a gradient of increasing herbivore nutrient recycling efficiency, under the assumption that plant reproductive ability is proportional to plant productivity. Plant defense increases as herbivores are more efficient at recycling the nutrient, but at the same time, the interaction becomes increasingly mutualistic. (de Mazancourt *et al.*, in preparation)

evolutionary time-scale, provided that the herbivore recycling efficiency is sufficiently greater than plant recycling efficiency, thus generating a plant–herbivore mutualistic interaction. As expected, the requirements on herbivore recycling efficiency are more stringent for an evolutionary mutualism than for an ecological mutualism. Surprisingly, however, as herbivore recycling efficiency is increased, the plant–herbivore interaction becomes increasingly mutualistic (first ecologically, then evolutionarily), but at the same time, plants evolve to increase their level of antiherbivore defense because they gain a higher benefit from not being consumed relative to less defended plants (Figure 3). Thus, mutualism can go hand in hand with increased conflict between partners. Although paradoxical at first sight, such evolutionary conflicts are also known in other mutualistic interactions (Anstett *et al.*, 1997; Law and Dieckmann, 1998).

CONCLUSION

Evolutionary processes are capable of leading to a significant transformation in species ecological properties over relatively short time-scales. However, the evolutionary dynamics of species are embedded in the dynamics of ecosystems as complex systems of interactions among species. Species traits and the evolution of species traits are ultimately constrained by ecosystem processes, just as ecosystem properties are constrained by the ecological and evolutionary history of interacting species. Indirect effects mediated by ecosystem processes can be so strong as to prevail over direct effects and exert effective selective pressures on the species involved, provided that there is sufficient spatial heterogeneity in the system or trade-offs between traits associated with the direct and indirect effects. They are even able to change the nature of species interactions, both in an ecological and an evolutionary sense. Thus, merging the evolutionary and ecosystem perspectives, which have been increasingly separated in modern ecology, is fundamental to predict the responses of ecological systems to environmental changes.

REFERENCES

Anstett, M C, Hossaert-McKey, M, and Kjellberg, F (1997) Figs and Fig Pollinators: Evolutionary Conflicts in a Coevolved Mutualism, *Trends Ecol. Evol.*, **12**, 94–99.

Belsky, A J (1986) Does Herbivory Benefit Plants? A Review of the Evidence, *Am. Nat.*, **127**, 870–892.

Belsky, A J, Carson, W P, Jense, C L, and Fox, G A (1993) Overcompensation by Plants: Herbivore Optimization or Red Herring? *Evol. Ecol.*, **7**, 109–121.

Bergelson, J and Crawley, M J (1992) Herbivory and *Ipomopsis aggregata*: the Disadvantages of Being Eaten, *Am. Nat.*, **139**, 870–882.

Coley, P D, Bryant, J P, and Chapin, III, F S (1985) Resource Availability and Plant Antiherbivore Defense, *Science*, **230**, 895–899.

de Mazancourt, C and Loreau, M (2000) Grazing Optimization, Nutrient Cycling and Spatial Heterogeneity of Plant–Herbivore Interactions: Should a Palatable Plant Evolve? *Evolution*, **54**, 81–92.

de Mazancourt, C, Loreau, M, and Abbadie, L (1998) Grazing Optimization and Nutrient Cycling: When do Herbivores Enhance Plant Production? *Ecology*, **79**, 2242–2252.

de Mazancourt, C, Loreau, M, and Abbadie, L (1999) Grazing Optimization and Nutrient Cycling: Potential Impact of Large Herbivores in a Savanna System, *Ecol. Appl.*, **9**, 784–797.

Dieckmann, U (1997) Can Adaptive Dynamics Invade? *Trends Ecol. Evol.*, **12**, 128–131.

Douglas, A E and Smith, D C (1989) Are Endosymbioses Mutualistic? *Trends Ecol. Evol.*, **4**, 350–352.

Grant, P R and Grant, B R (1995) Predicting Microevolutionary Responses to Directional Selection on Heritable Variation, *Evolution*, **49**, 241–251.

Hilbert, D W, Swift, D M, Detling, J K, and Dyer, M I (1981) Relative Growth Rates and the Grazing Optimization Hypothesis, *Oecologia*, **51**, 14–18.

Law, R and Dieckmann, U (1998) Symbiosis through Exploitation and the Merger of Lineages in Evolution, *Proc. R. Soc. Lond., Ser. B*, **265**, 1245–1253.

Lennartsson, T, Tuomi, J, and Nilsson, P (1997) Evidence for an Evolutionary History of Overcompensation in the Grassland Biennial *Gentianella campestris* (Gentianaceae), *Am. Nat.*, **149**, 1147–1155.

Lewontin, R C (1983) The Organism as the Subject and Object of Evolution, *Scientia*, **118**, 65–81.

Loreau, M (1995) Consumers as Maximizers of Matter and Energy Flow in Ecosystems, *Am. Nat.*, **145**, 22–42.

Loreau, M (1998) Ecosystem Development Explained by Competition within and Between Material Cycles, *Proc. R. Soc. Lond., Ser. B*, **265**, 33–38.

Loreau, M and de Mazancourt, C (1999) Should Plants in Resource-poor Environments Invest more in Antiherbivore Defence? *Oikos*, **87**, 195–200.
Losos, J B, Warheit, K I, and Schoener, T W (1997) Adaptive Differentiation Following Experimental Island Colonization in *Anolis* Lizards, *Nature*, **387**, 70–73.
McNaughton, S J (1979) Grazing as an Optimization Process: Grass–Ungulate Relationships in the Serengeti, *Am. Nat.*, **113**, 691–703.
Owen, D F and Wiegert, R G (1976) Do Consumers Maximize Plant Fitness? *Oikos*, **27**, 488–492.
Paige, K N and Whitham, T G (1987) Overcompensation in Response to Mammalian Herbivory: The Advantage of Being Eaten, *Am. Nat.*, **129**, 407–416.
Pennisi, E (1998) How the Genome Readies Itself for Evolution, *Science*, **281**, 1131–1134.
Puccia, C J and Levins, R (1985) *Qualitative Modeling of Complex Systems*, Harvard University Press, Cambridge, MA.
Reznick, D N, Shaw, F H, Rodd, F, Helen, and Shaw, R G (1997) Evaluation of the Rate of Evolution in Natural Populations of Guppies (*Poecilia reticulata*), *Science*, **275**, 1934–1937.
Rutherford, S L and Lindquist, S (1998) Hsp90 as a Capacitor for Morphological Evolution, *Nature*, **396**, 336–342.
Silvertown, J W (1982) No Evolved Mutualism between Grasses and Grazers, *Oikos*, **38**, 253–259.
Thompson, J N (1998) Rapid Evolution as an Ecological Process, *Trends Ecol. Evol.*, **13**, 329–332.
Tilman, D (1982) *Resource Competition and Community Structure*, Princeton University Press, Princeton, NJ.
Wilson, D S (1980) *The Natural Selection of Populations and Communities*, Benjamin/Cummings, Menlo Park, CA.
Wootton, J T (1994) The Nature and Consequences of Indirect Effects in Ecological Communities, *Annu. Rev. Ecol. Syst.*, **25**, 443–466.
Yodzis, P (1988) The Indeterminacy of Ecological Interactions as Perceived through Perturbation Experiments, *Ecology*, **69**, 508–515.

Extinctions and Biodiversity in the Fossil Record

Ricard V Solé and Mark Newman
Santa Fe Institute, Santa Fe, NM, USA

Life has existed on Earth for more than three billion years. Until the Cambrian explosion about 540 million years ago, however, life was restricted mostly to single-celled microorganisms that were, for the most part, poorly preserved in the fossil record. From the Cambrian explosion onwards, by contrast, we have a substantial fossil record of life's development, which shows a number of clear patterns, including a steady increase in biodiversity towards the present, punctuated by a number of large extinction events that wiped out a significant fraction of the species on the planet, and in some cases, caused major reorganizations amongst the dominant groups of organisms in the ecosphere.

The history of life on the Earth begins in the oceans during the *Achaean eon*, somewhere around 3.8 billion years ago, probably with the appearance first of self-reproducing RNA molecules and subsequently of prokaryotic single-celled organisms. At the start of the *Proterozoic eon*, around 2.5 billion years ago, the atmosphere changed from reducing to oxidizing, as a result of the depletion of stocks of elemental iron in the Earth's crust, and oxygen-breathing life became possible. Eukaryotes, multicellular life, and sexual reproduction all appeared for the first time during the Proterozoic, although the exact order and dates are still in dispute, since the fossil record of this period is poor. The earliest firm evidence of multicellularity dates from about 575 million years ago.

About 540 million years ago, for unknown reasons, an enormous increase in the diversity of multicellular animals took place. This event, known as the *Cambrian explosion*, produced all the major body plans of animals seen today, as well as a number of others that have since become extinct. The first land-dwelling plants appeared during the *Silurian period*, about 430 million years ago, followed shortly afterwards by the first land-dwelling animals, which were insects, and then the first land-dwelling vertebrates during the *Devonian*. About 250 million years ago, at the end of the *Permian* period, the largest mass extinction of all time took place, killing at least 90% of all species on the Earth, and ending the eon named the *Paleozoic*.

The *Mesozoic*, colloquially known as *the age of the dinosaurs*, followed the Paleozoic. In addition to dinosaurs, the Mesozoic also saw the first appearance of mammals and of flowering plants. It ended about 65 million years ago with the *Cretaceous–Tertiary (KT) extinction event*, which wiped out the dinosaurs along with about 70% of all other species then alive. The interval from the KT event until the present, known as the *Cenozoic eon*, saw the radiation of the mammals to fill many of the dominant land-dwelling niches and, eventually, the evolution of mankind.

This information comes from geological studies and from the substantial fossil record of extinct lifeforms. The currently known fossil record includes about one-quarter of a million species, mostly dating from the time interval between the Cambrian explosion and the present. There are numerous biases in the fossil record, which make accurate quantitative investigations difficult, including the following.

1. Older fossils are harder to find because they are typically buried in deeper rocks than more recent ones.
2. Accurate dating of fossils is difficult. Radiocarbon dating, for example, is not useful for rocks that are hundreds of millions of years old. Radioactive isotopes other than carbon with longer half-lives are used for most of the geologic time-scale, but resolution of dates using these isotopes can be poor.
3. Particularly rich fossil beds, or particularly zealous investigators, may produce very complete records for some time periods or groups of organisms, while other periods or groups may be comparatively poorly researched.
4. Prehistoric environmental disturbances can upset the deposition processes by which fossils are formed and give rise to time periods in which the fossil record is poor.
5. Marine organisms tend to be much better preserved than land-dwelling ones, because deposition is much more uniform and reliable in the oceans than it is on land.

Despite these biases, a number of trends are clear from the fossil record.

Figure 1 shows a plot of the best estimate of the number of living families of marine organisms, as a function of time since the start of the Cambrian. As the plot shows, diversity appears to increase substantially over time, and this is believed to be a real trend. On average, there have been more families, and indeed more species, alive in recent times than earlier. Over the course of the plot, diversity appears to increase by a factor of about five, although some of this increase is an artifact of the greater availability of fossils of recent species in more easily accessible rocks.

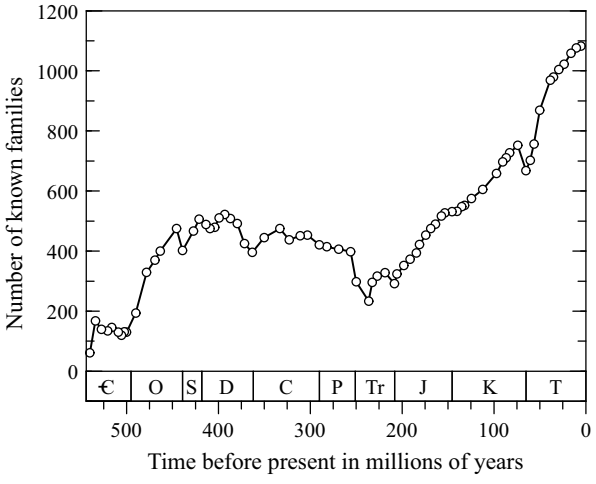


Figure 1 Number of known marine families alive over the time interval from the Cambrian to the present. (The data are taken from Sepkoski, 1992)

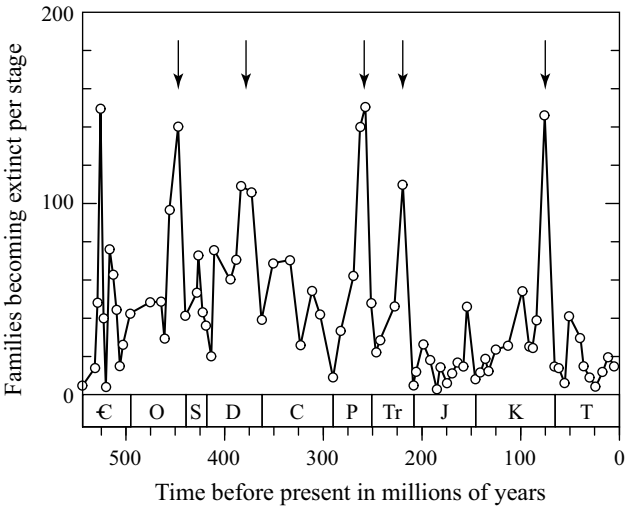


Figure 2 Estimated extinction of marine animals in families per stratigraphic stage since the Cambrian. The arrows indicate the positions of the big five mass extinction events discussed in the text

Figure 2 shows the extinctions, again of marine organisms, as a function of time over the same interval. The graph gives the number of families becoming extinct per stratigraphic stage. Stages are uneven intervals of time based on recognizable geological and paleobiological markers, with typical length about 7 million years. As the figure shows, there has been considerable variation in the intensity of extinction over prehistoric time. Of particular note are the five large peaks in extinction marked with arrows. These are the big five mass extinction events that marked the ends of the Ordovician, Devonian, Permian, Triassic, and Cretaceous periods. A sixth peak in the Cambrian is also visible, but this is thought probably to be an artifact of poor fossil preservation during that period rather than a real extinction event. The basic features of the big five extinctions are as follows (see also Table 1).

The end-Ordovician event about 440 million years ago appears to have occurred in two bursts, separated by about

Table 1 Extinction intensities at the genus and species level for the big five mass extinctions of the Phanerozoic. Estimates of genus extinction are obtained from directed analysis of the fossil record while species loss is inferred using a statistical technique called reverse rarefaction. (Figures are taken from Jabionski, 1991)

Extinction	Genus loss (%) (observed)	Species loss (%) (estimated)
End Ordovician	60	85
Late Devonian	57	83
Late Permian	82	95
End Triassic	53	80
End Cretaceous	47	76

1 million years, which between them wiped out about 85% of then-living species. The event was confined to marine species, since multicellular life had not yet colonized the land. Particularly affected were brachiopods, bivalves, echinoderms, bryozoans, and corals. The immediate cause of extinction appears to have been the continental drift of a significant landmass into the south polar region, causing a global temperature drop, glaciation, and consequent lowering of the sea level, which destroyed species habitats around the continental shelves. The sea level rose again with the end of the glacial interval about a million years later and caused a second burst of extinction.

The late-Devonian extinction around 360 million years ago is complex and rather poorly understood. It is probably, in fact, composed of a number of separate events – as many as seven – spread over about 25 million years, including particularly notable extinctions at the ends of the Givetian, Frasnian, and Famennian stages. Overall, about 80% of living species died out in the late Devonian. Particularly hard hit were corals, brachiopods, bryozoans, ammonoids, and fish. The causes of these extinctions are unclear. The leading theories suggest that changes in sea level and ocean anoxia, possibly triggered by global cooling or oceanic volcanism, were most likely responsible, although the impact of an extraterrestrial body such as a comet has also been considered.

The late-Permian extinction around 250 million years ago was the largest extinction event of all time, killing some 95% of marine species and about 70% of land-dwelling species. Like the end-Ordovician event, it seems to have been composed of two bursts, separated by an interval of about 10 million years, the second being the larger of the two. Notable extinction happened again amongst brachiopods, ammonoids, and corals, as well as gastropods and, unusually, insects. Despite an enormous amount of research on the subject, the causes of the late-Permian event are still a subject of debate. It is clear, however, that the sea level rose during this period, levels of oxygen in the oceans were low, and carbon dioxide (CO_2) levels were high. There is some suggestion that a cometary impact may have been involved, or a shift in ocean circulation driven by climate change, or CO_2 and sulfur release following large-scale volcanic activity. The late-Permian event had a profound effect on the terrestrial ecosystem, which is still being felt today, a quarter of a billion years later. A particularly notable example amongst marine faunas is that of the bivalves, a relatively minor group during the Paleozoic that took advantage of the ecological vacuum left by the extinction to establish a solid grip on shallow-water environments, leading to their dominance over the previously very successful brachiopods and gastropods.

The end-Triassic extinction around 210 million years ago is probably the most poorly understood of the big five extinction events. It appears to have killed about 80% of species then living, either in one burst or possibly in two, separated by about 20 million years. Major extinction is observed particularly amongst ammonoids, bivalves, gastropods, and brachiopods. Leading theories of the causes of the end-Triassic event are ocean anoxia, massive volcanism, or possibly an asteroid impact.

The end-Cretaceous event, usually called the *KT event*, has attracted the most popular interest of any extinction because it saw the death of those perennial movie stars, the dinosaurs, but it was in fact the smallest by quite a wide margin of all the big five. The KT event appears to have been a single pulse of extinction around 65 million years ago, which wiped out about 70% of all species then living. As well as the dinosaurs, it extinguished many other land-dwelling vertebrates, especially large-bodied ones, along with large numbers of (marine) bivalves, gastropods, and foraminifera. The proximal cause of the KT event was, almost certainly, the impact of a large comet or meteor near the present site of the town of Puerto Chicxulub on the Yucatán peninsula in eastern Mexico, with an associated drop in sea level and possibly short-term cooling or heating, or acid rain.

The fossil record can give us valuable insight into the nature of extinction and the effects of large-scale environmental change. It also indicates that recovery from extinction is a slow process by human standards, typically taking on the order of 5 or 10 million years. Thus, it is important to fend off such extinctions before they happen, rather than hoping that the ecosystem will prove robust enough to take care of itself.

Comparison between fossil and present-day extinction is not straightforward since, as mentioned above, the fossil record consists largely of marine organisms, whereas interest in contemporary extinction focuses mostly on land-dwelling organisms. Also recently extinct or currently endangered species tend to be rare, whereas the fossil record primarily reflects the most abundant and numerous biotas. Still, there are a number of general patterns in the fossil extinction record that may help us in the conservation of modern-day biodiversity.

First, we note that habitat loss, such as the destruction of shallow-water environments on the continental shelf as a result of changes in sea level, appears to have been an important cause of extinction again and again. A rise in sea level as a result of global warming over the next century, for example, could be devastating for reef communities. In this context, human-driven habitat loss is leading to extensive habitat fragmentation, i.e., to the generation of spatially isolated subareas. In many cases the subdivision of a large population into many weakly connected subpopulations substantially increases extinction risk. Again, the fossil record is very useful in providing a source of comparison. Habitat fragmentation enhanced the extinction of large

mammals during the Pleistocene (75% of them are extinct). The area reduction during glacial cycles led to widespread habitat fragmentation and eventually to the extinction of many species. A considerable number of current endemisms in plant species are due to the confinement to small areas during the Ice Age.

Second, most past extinction events appear to have been selective to some extent. The end-Ordovician event, which was associated with a period of global cooling and glaciation, particularly favored species that were well adapted to cold-water conditions, and was particularly harsh on those who were not, for obvious reasons. The KT boundary event, as mentioned previously, appears to have come down especially hard on large-bodied animals. A number of explanations for this latter effect have been put forward. Large-bodied animals have smaller populations and greater area requirements, and are thus more sensitive to habitat loss or fragmentation. Moreover, their trophic requirements and low rates of population growth make them slow to recover from environmental change. It has been suggested that the loss of large herbivorous faunas could trigger major changes in biogeographic vegetation patterns that could in turn trigger further extinctions.

This last observation leads us to an additional question: to what extent do interactions between species affect the response of an ecosystem to environmental stress? Ecological interdependence between species may have heightened the impact of some of the mass extinctions; some studies, for example, have suggested that the collapse of marine food chains at the end-Cretaceous contributed to the KT extinction event. The KT event had a rapid effect on most biotas and a subsequent long-term effect perhaps related to a decrease in primary productivity. This seems particular likely in the case of marine biotas, which show a marked dip in the rate of accumulation of carbonates, especially calcium carbonate, following the KT boundary, indicating a decrease in productivity. The resulting decrease in food supply would then produce extinctions at higher trophic levels. Similar mechanisms may also have been at work during the human-driven end-Pleistocene extinctions of mastodons and mammoths in North America, which were associated with widespread changes in vegetation patterns and the disappearance of many other species. One may well ask whether the current extinction of large herbivorous species in African ecosystems will result in similar concurrent extinctions. It is also believed that symbiotic and parasitic species are particularly vulnerable to extinction, as a result of their dependence upon a partner or host. Many living species may, in fact, already be doomed to extinction, because of the loss of a partner in an essential symbiotic relationship (Figure 3). The situation with parasites may be even worse, since most parasites are specific to a single species of host, and extinction amongst

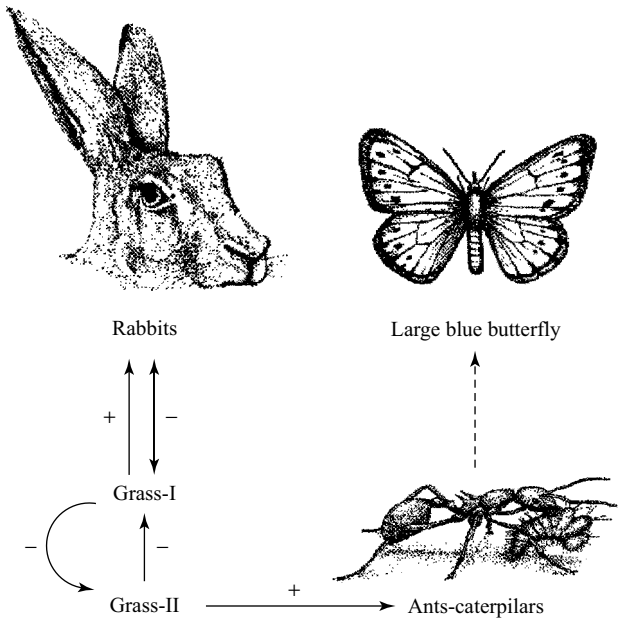


Figure 3 An example of coextinction: the introduction of the myxoma virus to rabbit warrens in England triggered the extinction of the Large Blue butter y (*Maculina arion*), an already endangered, endemic species. Myxomatosis caused a dramatic decline in rabbit populations and a consequent surge in the abundance of a species of grass on which the rabbits previously grazed (Grass-I). Another previously dominant species of grass (Grass-II) was unable to compete and declined substantially in its abundance, giving rise to a shortage of nesting material for certain variety of ant (*Mynnica*). As a result of a symbiotic dependence between the butter y *Maculina* and this species of ant, the butter y then became extinct

parasites often goes unnoticed, even though their importance in maintaining diversity has been stressed by many studies. Since all known animals and plants have some parasitic load, extinctions of parasites must be widespread in evolutionary history. The eggs of what are believed to be ectoparasitic mites have been discovered on fossilized dinosaur remains, indicating that coevolutionary parasitism, and hence coextinction (the coupled extinction of ecological partners) is an ancient phenomenon. Estimating extinction rates for parasites is unfortunately difficult, since they are poorly represented or difficult to identify in the fossil record, and internal parasites are basically not preserved at all.

Overall, recent extinction rates in most plant and animal taxa are relatively low, but for some groups they do approach the levels associated with prehistoric mass extinction events (Table 2). These observations have led some researchers to suggest that the biosphere may be on the verge of another such event. Some claim that the world has already entered a sixth period of mass extinction, driven primarily by the human population explosion. Theoretical studies of long-term ecological responses to habitat

Table 2 Comparison between different features of past mass extinction events (as reported from the fossil record) and present-day, human-driven mass extinction. Here the effects of habitat loss are widespread but different mechanisms were at work: changes in sea levels and continental breakup (fossil record) and human-driven habitat destruction/fragmentation (present day) respectively. (Estimations of species loss are from Wilson, 1992)

	Fossil record	Present day
Time resolution	$\sim 10^5$ - 10^6 years	1 year
Most affected biotas	Tropical biotas	Coral reefs, rainforests
Selectivity on size	Large-sized species	Large-sized species
Loss of endemics	Not well known	Widespread
Effects of habitat loss	Widespread	Widespread
Time scale of recovery	~5- 10 Myr	Not known
Direct effects on food webs	Very important	Very important
Indirect effects on food webs	Not known	Very important
Extinction rates	75- 80%	$\sim 10^4$ species year^{-1}

destruction suggest that steady increases in extinction rates are to be expected in the near future. Paleontological studies indicate that rare, localized, or specialized species that have evolved to survive in particular niches are the mostly likely to become extinct, while widespread or adaptable species, or opportunistic colonizers are likely to prevail. The fossil record is thus more than a cautionary tale.

As paleontologist David Jablonski has put it:

> The lessons from the past are inevitably blurry at a coarse scale. At the present stage of knowledge, the fossil record is more revealing for potential long-term consequences than for immediate solutions. However, the history of life of Earth provides an array of worst-case scenarios ... that are sufficiently spectacular to militate against inaction.

REFERENCES

Jablonski, D (1991) Extinctions: A Paleontological Perspective, *Science*, **253**, 754–757.

Sepkoski, Jr, J J (1992) *A Compendium of Fossil Marine Animal Families, Milwaukee Public Museum Contributions in Biology and Geology, 83*, 2nd edition, Milwaukee Public Museum, Milwaukee, WI.

Wilson, E O (1992) *The Diversity of Life*, Harvard University Press, Cambridge, MA.

FURTHER READING

Lawton, J H and May, R M, eds (1995) *Extinction Rates*, Oxford University Press, Oxford.

Leakey, R and Lewin, R (1995) *The Sixth Extinction*, Doubleday, New York.

Mckinney, M and Drake, J, eds (1998) *Biodiversity Dynamics*, Columbia University Press, New York.

Extinctions (Contemporary and Future)

Thomas M Brooks

Center for Applied Biodiversity Science, Conservation International, Washington, DC, USA

The current documented rate of species extinctions is known to be maybe a hundred times higher than natural. Estimates of the overall rate converge at thousands of times higher. It is clear, then, that the planet faces a loss of species comparable to the five great extinction events of the geological past (Raup, 1986). In this essay, I begin by examining the causes of this loss of species. I then review the evidence that documents the size of the current extinction crisis, and assess how we can estimate the overall extinction rate. I conclude by highlighting the key geographical priorities for action to avert this crisis.

*I must give four points of reference as to what this essay is and what it is not. First, as the title suggests, I discuss contemporary (since approximately 1500) and future extinctions, not prehistoric nor even historic extinctions, which are amply treated elsewhere (*see ***Extinctions and Biodiversity in the Fossil Record****, Volume 2). Second, I discuss global species extinctions, not extirpations (the local extinctions of populations), although such extirpations are extremely important and, taken together, eventually lead to global extinction (Hughes* et al.*, 1997) nor the selective loss of deeper evolutionary units (McKinney, 1997). Third, I discuss only the mechanics, magnitude and geography of extinction, not what we should do to prevent it nor indeed why we should care. Fourth, I discuss, primarily, bird*

extinctions, not just because I am an ornithologist, but also because birds are one of the few groups well enough studied for us to understand species extinctions in reasonable detail.

What predisposes a species towards extinction? Three factors stand out: species that have low abundances, small range sizes (often making them endemic to specific areas) and specialized ecological niches tend to have a high probability of extinction. In combination, these factors comprise the seven forms of rarity (Rabinowitz *et al.*, 1986). Beyond these, a wide variety of biological correlates of extinction proneness have been identified including, for example, large body size, slow reproductive rate (K-selection) and high population variability in space or time (Gaston, 1994). However, these factors do not cause extinction *per se*, only increase its probability if threats are operating. What are these threats?

Diamond (1984) identified four factors (*ultimate causes*) by which people cause species to decline towards extinction:

(1) direct overkill, which primarily affects large, tasty species, and those that people catch because of their attractive colors or songs;
(2) the introduction of species into environments where they were not previously found (often oceanic islands), allowing them to run rampant and drive native species to extinction through predation or competition;
(3) habitat destruction, whether through deforestation, wetland drainage, conversion of grassland to agriculture, or damage to aquatic environments;
(4) indirect knock-on effects within the ecological community, whereby impacts on one species are transmitted through the food web to drive the extinction of another.

While these factors cause species to decline towards extinction, another set of factors (*proximate causes*) is likely to deliver the *coup de grace* to the last individuals: (Simberloff, 1986).

(1) Demographic stochasticity leads to extinction when no individuals of a species can reproduce, maybe because they are all of the same sex, or because they cannot find any mates.
(2) Genetic deterioration of a species occurs when a population becomes so small that inbreeding and subsequent loss of individual fitness occurs, or when a population is swamped by hybridization with a closely related species.
(3) Social dysfunction, prevalent when species with evolutionary adaptations to abundance (such as colonial breeding or group foraging) decline to scarcity in short periods of time.
(4) A wide range of extrinsic forces – unfortunate natural disasters like fires, storms and drought, and human disasters like war – can kill off the last individuals of a species.

The ultimate and proximate causes of extinction are usually hard to disentangle. Pimm (1996) illustrates with the case of Hawaii how few extinctions can be pinned onto any particular cause. This is partly because the ultimate and proximate mechanisms driving extinction appear to be inherently synergistic, and partly because species that become extinct tend to be rare and so, almost by definition, we know very little about them.

This paucity of knowledge has led to great problems in the documentation of contemporary extinctions. Only maybe a million and a half species have been formally described, while the actual number of species on the planet may well be ten or even a hundred times that number (May, 1988). As a result, we must be causing the extinctions of huge numbers of species for which we do not even have names. The most notorious example of this is the case of the Centinelan extinctions (Dodson and Gentry, 1991). In 1975, the flora of Centinela forest in western Ecuador was surveyed for the first time and found to hold about 100 plant species not known anywhere else in the world; by 1988, the forest had been completely cleared. Were it not for that single 1975 survey, those 100 species would have disappeared unnamed and unmourned.

There are clearly enormous gaps in what we know of those species that have names, too. I illustrate this problem in Figure 1 by plotting the numbers of known contemporary extinctions listed on the International Union for the Conservation of Nature and Natural Resources (IUCN) Red Lists (Baillie and Groombridge, 1996; *see also* **World Conservation Union (IUCN) Red List of Threatened Species**, Volume 4) and approximate numbers of described species (Smith *et al.*, 1993) for seven major taxa. The numbers of species are plotted on a logarithmic scale to allow comparison between the two sets of values. The proportion of extinctions among birds and mammals is an order of magnitude higher than it is among less well-known vertebrates and plants, and two orders of magnitude higher than among those invertebrates for which we have names.

Even for vertebrates, our knowledge of the loss of individual species is astonishingly poor. MacPhee and Flemming (1999) recently compiled rigorous documentation for contemporary mammal extinctions and found only 49% compatibility between their list and that given by Baillie and Groombridge (1996). The situation for birds appears likely to be only a little better. Most dramatically of all, a similar survey by Harrison and Stiassny (1999) could identify only three freshwater fish species for which there is conclusive evidence of extinction. This compares with no less than 93

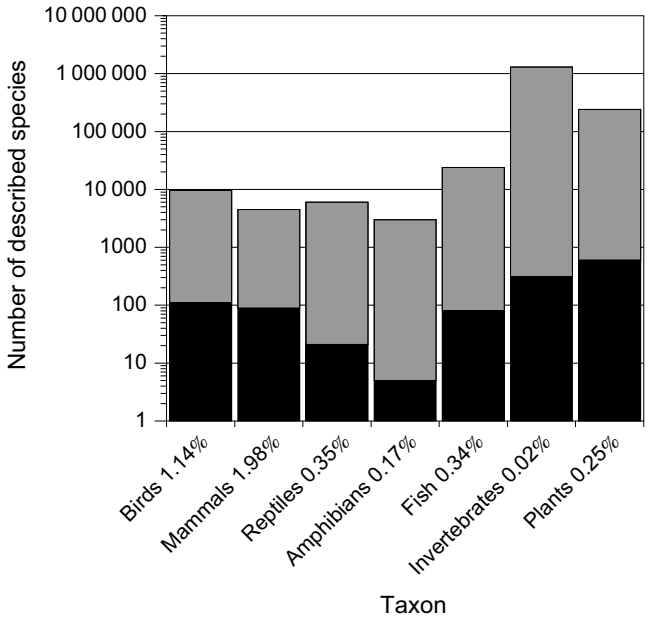


Figure 1 Numbers of known contemporary extinctions listed on the IUCN Red Lists (Baillie and Groombridge, 1996) and approximate numbers of described species (Smith *et al.*, 1993) for seven major taxa

species listed as extinct by Baillie and Groombridge (1996), of which more than half are cichlids from Lake Victoria, recently demonstrated to be of uncertain conservation status (Seehausen *et al.*, 1997).

What could explain these discrepancies? Partly, they are due to Baillie and Groombridge's (1996) use of a more precautionary principle in listing species as extinct unless proven extant (Diamond, 1987). Probably, however, they are due to the fact that the Red Lists are (extremely powerful) conservation tools aimed at documenting threats to species that can still be saved, and so little concerted effort has been invested in documenting species that have already crossed the threshold into extinction. In an attempt to rectify this, an independent committee on recently extinct organisms (http://creo.amnh.org) has recently been formed; their work is currently in progress.

Since extinctions are a natural part of the history of life, absolute numbers of species lost would be rather meaningless for comparative purposes, even if we had such numbers. The contrast between the current crisis and the natural background is one of extinction rate, that is, number of extinctions per unit time. Further, since, as I have shown, we have no idea how many species exist in total, it is sensible to compare extinction rates per unit species – extinction rates per, say, every million species. We term this measure *extinctions per million species years* E/MSY (Pimm *et al.*, 1995). The background extinction rate across diverse fossil taxa is 0.1–1 E/MSY (May *et al.*, 1995). Figure 1 plots extinctions documented over the last 500 years, and so if we trust the numbers of extinctions for birds, at least, we can see that the contemporary extinction rate is $100/(9500 \times 500) = \approx 20$ E/MSY. Most of these extinctions have, in fact, been in the last century, and so the contemporary documented bird extinction rate is a bit higher than this, maybe two orders of magnitude greater than the background rate.

How can we look forward to assess likely extinction rates in the near future? A number of techniques have been used to do this (Table 1). Maybe the simplest of these is to return to the Red Lists (Baillie and Groombridge, 1996) and assume that, for well-known taxa, trends over the last 500 years of human activity will continue. Thus for the 1215 threatened and extinct species of birds, this gives a very conservative extinction rate of ≈250 E/MSY. Using the new IUCN criteria for threatened status on the Red Lists, it is also possible to sum extinction probabilities across species, giving a higher estimated extinction rate of

Table 1 Estimates of global extinction rates, measured as number of extinctions per million species-years

Source	Method	Extinction rate
May *et al.* (1995)	Background rate (13 studies)	0.1- 1
Baillie and Groombridge (1996)	Bird species extinct since 1500	20
Baillie and Groombridge (1996)	Assuming extinction of all Red List bird species	300
Ehrlich (1994)	Increasing human energy consumption	10 000
Mace (1994)	Extinction probabilities from vertebrate Red List categories	1100- 2200
Myers (1979)	Extrapolation of exponentially increasing numbers of extinctions	4000
Pimm and Brooks (1999)	Half-life of 50 years for threatened birds	1000
Reid (1992)	Species-area relationship from deforestation rates (nine studies)	1000- 11 000
Smith *et al.* (1993)	Movement of birds and mammals through Red List categories	1400- 2000

1100–2200 E/MSY for various vertebrates (Mace, 1994). A twist to this technique is to apply calculated "times to extinction" for birds in recently fragmented forest to the threatened species on the Red List, which yields a rate of 1000 E/MSY (Pimm and Brooks, 1999).

An alternative method is to extrapolate the exponentially increasing number of extinctions of the last few centuries forward in time. Assuming that there are 10 million species worldwide, this gives an estimated contemporary extinction rate of 4000 E/MSY (Myers, 1979). A more detailed extrapolation technique uses the rates at which bird and mammal species listed as threatened in the Red Lists are upgraded to higher categories of threat, to produce an extinction rate of 2000 E/MSY (Smith *et al.*, 1993). This method is flawed by the fact that changes in the Red List categories are more a product of changing knowledge than of changing conservation status (Cuarón, 1993), but that it falls within the same order of magnitude as the estimate from simple extrapolation is striking.

These methods are limited, of course, by our lack of knowledge of the conservation status of nearly all species. One way to avoid this problem is to use the species-area relationship, $S = cA^z$, well-known for predicting species numbers in areas of different sizes over space, and instead adapt it to predict surviving species numbers in shrinking habitat areas over time (Reid, 1992). Basing calculations on the assumption that most species – approximately two-thirds of the world's total (Raven, 1988) – are found in the tropical forests, on a conservative estimate of the total number of species – approximately 10 million worldwide (May, 1988), and on known rates of tropical deforestation – approximately 1% $year^{-1}$ (Whitmore, 1997) – this technique estimates global extinction rates to lie between 1000 and 11 000 E/MSY (Reid, 1992). The species-area approach to estimating extinction rates has been criticized, largely because it is so difficult to prove (Budiansky, 1994). However, it has been successfully applied to predict the number of extinctions following historical habitat destruction in North America for both mammals (Dial, 1994) and birds (Pimm and Askins, 1995).

The broadest but most mechanistic technique used to estimate extinction rates is that of Ehrlich (1994), who shows how increases in extinction rates are tightly correlated with increases in human energy appropriation. Since global human energy use has increased approximately 10 000-fold since the beginnings of agriculture 10 000 years ago, it seems reasonable to assume that extinction rates will have been elevated by a similar factor, to maybe 10 000 E/MSY. While this calculation is likely to be an upper estimate, it falls into the same range as rates calculated by other methods.

Estimates of overall extinction rates in the immediate future are, therefore, hundreds of times higher than the contemporary extinction rates documented to date. Why this difference? For one thing, tropical deforestation continues apace, and so near-future extinction rates must be increasing (Whitmore, 1997). Over and above this acceleration of habitat loss, though, there is a time lag between habitat loss and extinction. Areas where human impacts have been severe for several centuries have typically lost as many species as would be predicted from the loss of their habitat, as in eastern North America (Pimm and Askins, 1996); or more, as in Hawaii (Pimm, 1996). In contrast, there have been far fewer extinctions in the recently-deforested continental tropics than we would predict, based on the extent of habitat loss. However, if we include currently threatened as well as extinct species, we find that both in Indonesia and the Philippines (Brooks *et al.*, 1997) and in South America's Atlantic forests (Brooks *et al.*, 1999), as many species are on the verge of extinction as we would predict given habitat destruction.

This test can be more generally applied by examining those concentrations of rarity and threat, where species extinctions are happening most rapidly. A number of recent efforts have mapped these concentrations, of which three stand out. Mittermeier *et al.* (1999) of Conservation International delimited 25 hotspots to each of which 0.5% of the global flora are endemic and within each of which <30% of the original habitat remains. Olson and Dinerstein (1998) of the World Wildlife Fund mapped the global 200 most biological valuable ecoregions, the most outstanding examples of each of the world's habitat types. Stattersfield *et al.* (1998) of Bird Life International identified 218 endemic bird areas that hold two or more bird species with ranges restricted to <50 000 km^2. The convergence between these prioritization systems is high (Figure 2).

Looking within each of these areas of convergence, we can use data on their proportion of habitat lost and number of avian endemics (Mittermeier *et al.*, 1999) to make predictions of bird extinctions based on habitat loss (Pimm *et al.*, 1995). We can then compare these to the numbers of endemic birds within each hotspot that are actually threatened (Collar *et al.*, 1994) or extinct (Baillie and Groombridge, 1996) by simply plotting the two

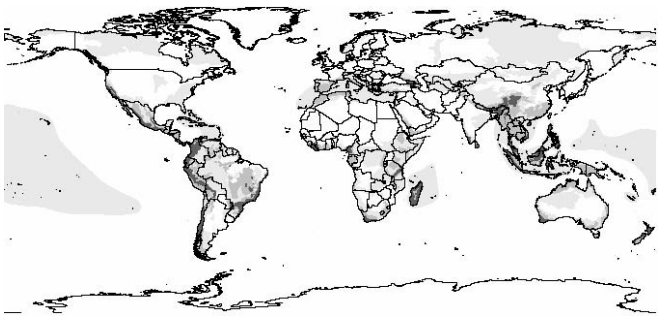


Figure 2 Hotspots (Mittermeier *et al.*, 1999), Global 200 ecoregions (Olson and Dinerstein, 1998) and Endemic Bird Areas (Statters eld *et al.*, 1998)

Table 2 Hotspots (Mittermeier *et al.*, 1999) and their habitat loss, avian endemism, predicted extinctions (Pimm *et al.*, 1995), and threat (Collar *et al.*, 1994) plus extinction (WCMC, 1992)[a]

Hotspot	Habitat remaining %	Avian endemics	Predicted extinction	IUCN Red List of threatened + extinctions
Tropical Andes	25.0	677	198	105
Mesoamerica	20.0	251	83	20
Caribbean	11.3	148	62	34
Chocó-Darién-Western Ecuador	24.2	85	25	10
Atlantic Forest Region	7.5	181[a]	86	60
Brazilian Cerrado	20.0	29	10	10
Central Chile	30.0	4	1	0
California Floristic Province	24.7	8	2	1
Madagascar and Indian Ocean Islands	9.9	199	87	74
Eastern Arc Mountains and Coastal Forests of Kenya and Tanzania	6.7	22	11	15
Cape Floristic Province	24.3	6	2	0
Succulent Karoo	27.0	1	0	1
Guinean Forests of West Africa	10.0	90	39	29
Mediterranean Basin	4.7	47	25	6
Caucasus	10.0	3	1	0
Sundaland	7.8	139	66	23
Wallacea	15.0	249	94	35
Philippines	8.0	183	86	73
Indo-Burma	4.9	140	74	35
Mountains of South-Central China	8.0	36	17	20
Western Ghats and Sri Lanka	6.8	40	20	7
New Caledonia	28.0	22	6	8
New Zealand	22.0	68	21	40
Polynesia/Micronesia	21.8	174	55	81
Southwest Australia	10.8	19	8	4

[a] The number of avian endemics (including non-forest species) in the Atlantic Forest Region is an update (Brooks *et al.*, 1999).

sets of numbers against each other (Table 2, Figure 3). We accurately predict bird extinction and threat in areas that have already been highly degraded, such as the Atlantic forests (Brooks *et al.*, 1999) and Philippines (Brooks *et al.*, 1997). Habitat loss overestimates threat and extinction in the tropical mountains of South and Central America, maybe because montane endemic species tend to be abundant within their (albeit tiny) ranges (Manne *et al.*, 1999). Conversely, the extent of habitat loss underestimates threat and extinction on islands, notably in the Pacific, where the effects of species introduced into these naïve ecosystems have been so severe (Pimm, 1996). Overall, though, we can see that the extent of habitat loss within these hotspots tends to predict rather well the number of bird species endemic to that hotspot that are either threatened or already extinct. Conservation must now focus within these areas to prevent species extinctions while still possible.

To conclude, I will summarize the key messages of this essay:

1. A number of factors increase the probability of a species' extinction, and these interact with a range of threats to actually cause extinctions. However, these interactions are complex, and only rarely do we have the information to identify a single cause of extinction.
2. This lack of information is reflected in how little we know about biodiversity overall, and translates into almost insurmountable difficulty in documenting extinctions for most groups of species. For the few vertebrate groups for which we have any confidence that data on extinctions are comprehensive, contemporary documented extinction rates appear to be about two orders of magnitude higher than natural rates.
3. A variety of techniques can be used to predict near-future extinction rates, some based on extrapolation of contemporary extinction and threat, others on the relationships between human impacts and species loss. These techniques produce estimates of near-future extinction rates converging at about four orders of magnitude higher than natural rates.
4. This difference between documented and overall extinction rates is largely due to the time-lag between

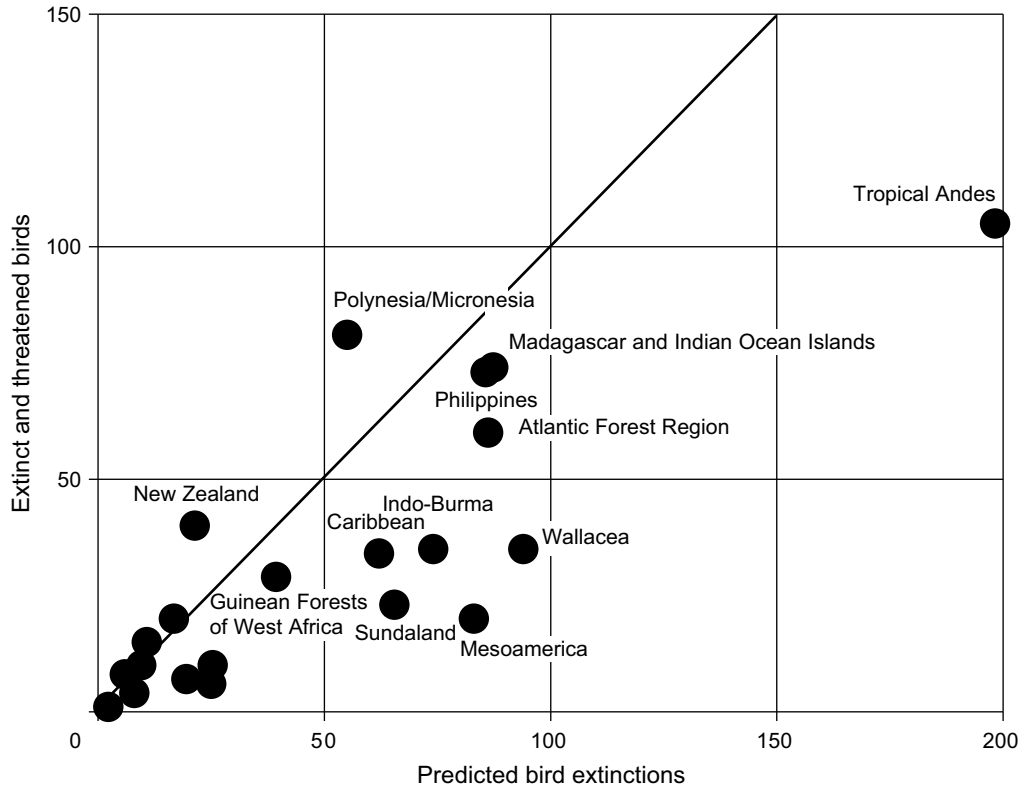


Figure 3 Comparing numbers of predicted bird extinctions in the hotspots, calculated using data on habitat loss and bird endemism (Mittermeier *et al.*, 1999) and the species-area relationship (Pimm *et al.*, 1995) with numbers of threatened (Collar *et al.*, 1994) and extinct (Baillie and Groombridge, 1996) species

habitat loss and extinction. This can be demonstrated by using habitat loss to predict the number of birds likely to become extinct in the world's hotspots of threat: these predictions match actual threat rather closely. Conservation must now target these areas to prevent further species extinctions.

REFERENCES

Baillie, J and Groombridge, B (1996) *1996 IUCN Red List of Threatened Animals*, IUCN–The World Conservation Union, Gland.

Brooks, T M, Pimm, S L, and Collar, N J (1997) Deforestation Predicts the Number of Threatened Birds in Insular South–East Asia, *Conserv. Biol.*, **11**, 382–394.

Brooks, T, Tobias, J, and Balmford, A (1999) Deforestation and Bird Extinctions in the Atlantic Forest, *Animal Conserv.*, **2**, 211–222.

Budiansky, S (1994) Extinction or Miscalculation? *Nature*, **370**, 105.

Collar, N J, Crosby, M J, and Stattersfield, A J (1994) *Birds to Watch 2, The World List of Threatened Birds*, BirdLife Conservation Series No. 4, BirdLife International, Cambridge.

Cuarón, A D (1993) Extinction Rate Estimates, *Nature*, **366**, 118.

Dial, R (1994) Extinction or Miscalculation? *Nature*, **370**, 104–105.

Diamond, J M (1984) Historic Extinctions: A Rosetta Stone for understanding Prehistoric Extinctions, in *Quaternary Extinctions*, eds P S Martin and R G Klein, University of Arizona Press, Tuscon, AZ 824–862.

Diamond, J M (1987) Extant Unless Proven Extinct? Or, Extinct Unless Proven Extant? *Conserv. Biol.*, **1**, 77–79.

Dodson, C H and Gentry, A H (1991) Biological Extinction in Western Ecuador, *Ann. Missouri Botanical Garden*, **78**, 273–295.

Ehrlich, P R (1994) Energy Use and Biodiversity Loss, *Philos. Trans. R. Soc. Lond.*, B, **344**, 99–104.

Gaston, K J (1994) *Rarity*, Chapman and Hall, London.

Harrison, I J and Stiassny, M L J (1999) The Quiet Crisis: A Preliminary List of the Freshwater Fishes of the World that are Extinct or "Missing in Action", in *Extinctions in Near Time: Causes, Contexts, and Consequences*, ed R D E MacPhee, Kluwer Academic, New York, 271–331.

Hughes, J B, Daily, G C, and Ehrlich, P R (1997) Population Diversity: its Extent and Extinction, *Science*, **278**, 689–692.

Mace, G M (1994) Classifying Threatened Species: Means and Ends, *Philos. Trans. R. Soc. Lon.*, B, **344**, 91–97.

MacPhee, R D E and Flemming, C (1999) Requiem Æternam: the Last Five Hundred Years of Mammalian Species Extinctions, in *Extinctions in Near Time: Causes, Contexts, and Consequences*, ed R D E MacPhee, Kluwer Academic, New York, 333–371.

Manne, L L, Brooks, T M, and Pimm, S L (1999) Relative Risk of Extinction of Passerine Birds on Continents and Islands, *Nature*, **399**, 258–261.

May, R M (1988) How Many Species Are There on Earth? *Science*, **241**, 1441–1449.

May, R M, Lawton, J H, and Stork, N E (1995) Assessing Extinction Rates, in *Extinction Rates*, eds J H Lawton and R M May, Oxford University Press, Oxford, 1–24.

McKinney, M L (1997) Extinction Vulnerability and Selectivity: Combining Ecological and Paleontological Views, *Ann. Rev. Ecol. System.*, **28**, 495–516.

Mittermeier, R A, Myers, N, Goettsch Mittermeier, C, and Robles Gil, P (1999) *Hotspots: Earth's Biologically Richest and most Endangered Terrestrial Ecoregions*, CEMEX, Mexico City.

Myers, N (1979) *The Sinking Ark: A New Look at the Problem of Disappearing Species*, Pergamon Press, Oxford.

Olson, D M and Dinerstein, E (1998) The Global 200: a Representation Approach to Conserving the Earth's Most Biologically Valuable Ecoregions, *Conserv. Biol.*, **12**, 502–515.

Pimm, S L (1996) Lessons From a Kill, *Biodivers. Conserv.*, **5**, 1059–1067.

Pimm, S L, Russell, G J, Gittleman, J L, and Brooks, T M (1995) The Future of Biodiversity, *Science*, **269**, 347–350.

Pimm, S L and Askins, R A (1995) Forest Losses Predict Bird Extinction in Eastern North America, *Proc. Natl. Acad. Sci. USA*, **92**, 9343–9347.

Pimm, S L and Brooks, T M (1999) The Sixth Extinction: How Large, How Soon, and Where? in *Nature and Human Society: the Quest for a Sustainable World*, ed P Raven, Proceedings of the 1997 Forum on Biodiversity, National Academy Press, Washington, DC, 38–54.

Rabinowitz, D, Cairns, S, and Dillon, T (1986) Seven Forms of Rarity and Their Frequency in the Flora of the British Isles, in *Conservation Biology: the Science of Scarcity and Diversity*, ed M E Soulé Sinauer Associates, Sunderland, MA, 182–204.

Raup, D M (1986) Biological Extinction in Earth History, *Science*, **231**, 1528–1533.

Raven, P H (1988) Our Diminishing Tropical Forests, in *BioDiversity*, ed E O Wilson, National Academy Press, Washington, DC 119–122.

Reid, W V (1992) How Many Species Will There Be? in *Tropical Deforestation and Species Extinction*, eds T C Whitmore and J A Sayer, Chapman and Hall, London, 55–73.

Seehausen, O, Witte, F, Katunzi, E F, Smuts, J, and Bouton, N (1997) Patterns of the Remnant Cichlid Fauna in Southern Lake Victoria, *Conserv. Biol.*, **11**, 890–904.

Simberloff, D (1986) The Proximate Causes of Extinction, in *Patterns and Processes in the History of Life*, eds D M Raup and D Jablonski, Springer-Verlag, Berlin, 259–276.

Smith, F M, May, R M, Pellew, R, Johnson, T H, and Walter, K S (1993) Estimating Extinction Rates, *Nature*, **364**, 494–496.

Stattersfield, A J, Crosby, M J, Long, A J, and Wege, D C (1998) *Endemic Bird Areas of the World*, BirdLife Conservation Series No. 7, BirdLife International, Cambridge.

Whitmore, T C (1997) Tropical Forest Disturbance, Disappearance, and Species Loss, in *Tropical Forest Remnants*, eds W L Laurence and R O Bierregaard, Jr, Chicago University Press, Chicago, IL, 3–12.

F

Fisheries: Effects of Climate Change on the Life Cycles of Salmon

see Fisheries: Effects of Climate Change on the Life Cycles of Salmon (**Volume 3**)

Fixed Nitrogen

Molecular nitrogen may be assimilated by various microorganisms and reduced to ammonium. The process is known as *nitrogen fixation*, and the product is *fixed nitrogen*. This process is of greatest significance to the growth of plants because nitrogen is their most important nutrient and because they are unable to utilize molecular nitrogen themselves. Therefore unless they are agriculturally fertilized, they must rely on decaying organic matter (*see* **Nitrogen Mineralization**, Volume 2) or nitrogen fixation. Nitrogen is lost from soils by mineralization and leaching (*see* **Nitrate Leaching**, Volume 3); therefore nitrogen fixation is essential for the continued fertility of natural ecosystems.

Nitrogen fixation is accomplished mainly by free living bacteria such as *Azotobacter* (aerobic) or *Clostridium* (anaerobic) and by several species of blue-green algae, both free living and in associations with such plants as ferns and cycads. But by far the most important nitrogen fixing system is the symbiotic association of *Rhizobium* bacteria with legumes. The bacteria infect the roots and induce the formation of nodules in which they live and fix nitrogen using energy derived from metabolites from the host plants. Considerable energy is required: to reduce one mole of molecular nitrogen requires over 80 kcal, which is derived ultimately from host photosynthesis.

The enzyme of nitrogen fixation, nitrogenase, involves both iron and molybdenum cofactors, and these elements must be available in the soil (or the oceans). In addition, oxygen required for bacterial metabolism is transported into the nodules by leghemoglobin, which is similar to animal hemoglobin. Leghemoglobin is a truly symbiotic product: the two parts of the molecule are synthesized independently in the bacterial and host cells and assembled in the bacterial environment within the host tissue.

Much effort has been spent in attempts to develop or create symbiotic associations between nitrogen fixing microorganisms and major crop plants such as maize. While it is true that photosynthetic production (i.e., crop yield) would be decreased because of the energy drain for fixing nitrogen, the overall economic balance would be strongly positive because of the high cost of nitrogen fertilizer. The ecological benefits would be much greater. Even with careful management a substantial portion of nitrogen fertilizer is leached out of fields. Not only is it wasted, it contributes to eutrophication and environment degradation. While efforts to apply these ideas have been unsuccessful, the agricultural application of *Rhizobium* to legume plantations (to the seed or in the field) has been very successful in increasing production while reducing fertilizer requirement.

R G S BIDWELL *Canada*

Fluxnet

see IGBP Core Projects (*Opening essay*, **Volume 2**)

Food Webs

Gary A Polis[1] and Gary R Huxel
University of California, Davis, CA, USA

Food webs occupy a central position in community ecology. Charles Darwin (Darwin, 1859) introduced the concept of

[1] Deceased March 27, 2000.

an entangled bank in which he envisioned many kinds of species interdependent upon each other in a complex manner governed by laws acting around us. In the simplest context, food webs incorporate the two factors that, a priori, one would consider most fundamental to the success or failure of a species: resources and enemies. All species must acquire resources (food or nutrients) and suffer energy losses or mortality from enemies (predators, parasites, pathogens). The abundance and success of any species is thus a product of these feeding interactions. This inclusion of such bottom-up (productivity and resources) with top-down (consumption) factors largely determines the distribution and abundance of almost every species on the planet.

Knowledge of food web structure and dynamics is central to our understanding of almost all aspects of population and community ecology. By their very nature of representing feeding relationships between species, food webs have the capacity to embody the rich complexity of natural systems. In fact, most important interactions (e.g., competition, predation, mutualism) cannot be isolated from a food web context. Many important advances have arisen from those analyses that concurrently incorporate more than one interaction in a food web: keystone predation and herbivory, the intermediate predation and disturbance hypotheses, the size efficiency hypothesis, trophic cascades, intraguild predation, apparent competition, and recognition of the importance of indirect effects. The outcome of virtually all interactions within a community can be modified, directly and indirectly, by other members of the food web. This insight penetrates all areas of community ecology. For example, the results of experiments must be interpreted carefully for at least two reasons. First, indirect effects, moderated by other species in the web, may exert large and sometimes contradictory effects to the direct effects of the manipulation. Thus, under some food web configurations, removal of a predator may directly increase the level of its prey or may actually cause the prey to decrease because of indirect interactions. Second, changes in species dynamics claimed to be caused by one factor may actually be a product of a second process.

TYPES OF FOOD WEBS

Food web research has grown at a tremendous rate and taken a diversity of forms. Not surprisingly, ecologists have diverged in their methods, emphasis and approaches. Nevertheless, trophic relationships in communities can be delineated in three basic ways. Paine (1980) and Polis (1991) distinguished three types of food webs that evolved from ecological studies (Figure 1). The first is the classic food

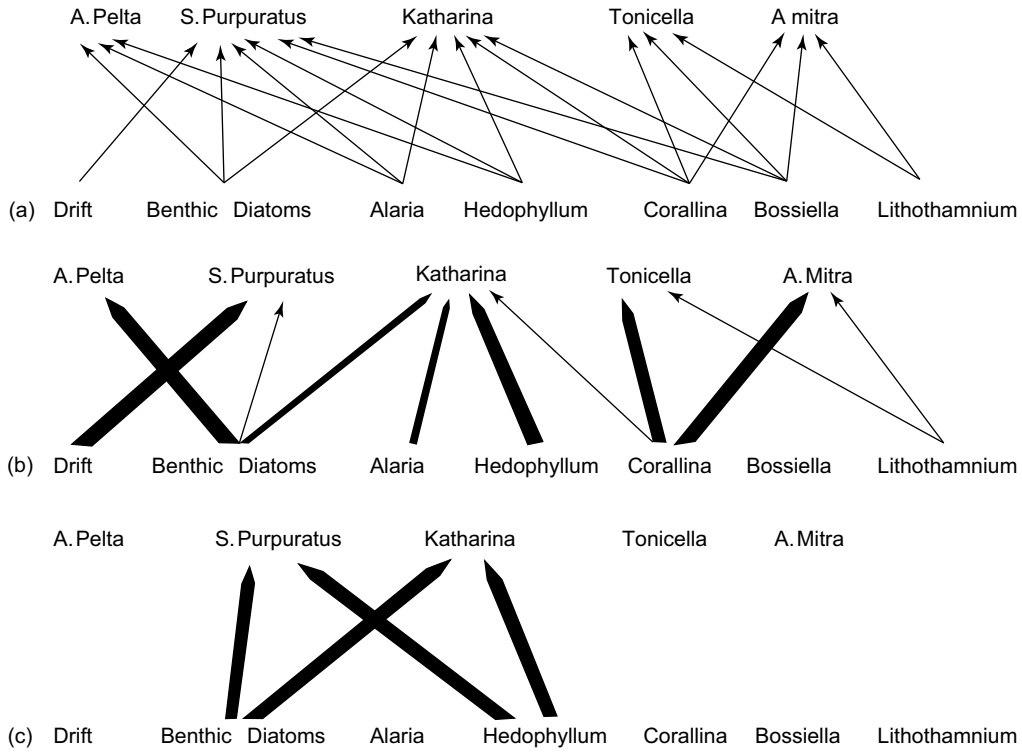


Figure 1 Three conceptually and historically different approaches to depicting trophic relationships, illustrated for the same set of species. The connectedness web (a) is based on observation, the energy ow one (b) on some measurement and literature values, and the functional web (c) on controlled manipulation. (Reproduced by permission of Blackwell Science Ltd., from Paine, 1980)

web, a schematic description of connectivity specifying feeding links. Such connectivity food webs simply demonstrate feeding relationships. Examples of these are the early food webs of Forbes (1887) and Summerhayes and Elton (1924) (Figure 1). The second web type is also descriptive, quantifying the flow of energy and matter through the community. These energetic food webs quantify the flow of energy (and/or materials) between trophically connected species. The third type use experiments to dissect communities to identify strong links and dynamically important species. Such interactive or functional food webs demonstrate the most important connections in an ecosystem (Figure 1). These food webs depict the importance of species in maintaining the integrity and stability of a community as reflected in its influence on the growth rates of other species. They require experimental manipulations of the community (by either removal or addition of particular species). In the next section, we discuss the strengths and weaknesses of each approach. Of the three, only the last two have contributed substantially to our understanding of natural systems.

Connectivity Food Webs

Connectivity food webs are representations of who eats whom without inference about the strength or type of interaction and energy flow (Figure 1). Early food webs were constructed for essentially two reasons: first, to depict the interconnectivity of natural systems, second, to examine issues of the balance of nature, i.e., to analyze how harmony is maintained through complex predatory and competitive interactions within communities (Forbes, 1887). Such an approach was applied to agricultural systems to examine pests and possible food web manipulations to control pests. As early as the 1880s, beetles were introduced into the US to control agricultural pests (following the success of *Vedalia*, a coccinellid ladybird beetle, in controlling Cottony-cushion scale in California in 1888, about 50 more coccinellids were introduced in the 1890s).

The knowledge required to construct connectivity food webs is straightforward: a rough, qualitative knowledge of who eats whom is all that is necessary to produce a simple food web whereas experimental manipulations or quantitative measurements are necessary to construct webs of interaction or energy flow. Consequently, connectivity webs most frequently represent trophic interactions in communities and have received the most attention. Hundreds of such webs slowly accumulated over a century. They were useful to illustrate, in a totally non-quantitative manner, the feeding interactions within a specific community. Different scientists constructed webs of different diversity, complexity and resolution, depending on their knowledge of the system and bias or understanding of particular groups. For example, some may emphasize birds, and represent all insects as one group. Others will divide the insects into scores of groups and represent one to two bird species (see Figure 2).

There are many problems with descriptive webs, of which only a few are mentioned here:

1. Most vastly under-represent the species diversity in natural communities. Most communities have $100\,s - 1000\,s$ of species, yet these webs represent $<10-30$ species on the average. As a consequence, most connectivity webs have severe problems with lumping species and taxonomic biases. Some trophic levels are distinguished by species (e.g., birds or fish) while other groups suffer a high degree of aggregation, e.g., all species of insect or annual plants are represented as one super-species, insects or plants.
2. Most vastly under-represent the true number of feeding connections within a community. Species eat and are eaten by $10\,s - 100\,s$ of species, yet these depictions show averages of two to four links with resources and enemies. It is an arduous task to represent the actual diet of even one to two species. Doing this for a complete community is truly a Herculean task that likely will never be accomplished without some technical breakthrough that analyzes a consumer's body tissue to determine the diversity of its foods.
3. Most species are highly omnivorous, feeding on many resources and prey. Each species has a distinct trophic history, and is often at a different trophic level than other species. Because diet is so difficult to delineate, most connectivity webs greatly under-represent the true nature of omnivory. This poses several fundamental problems (see the section on omnivory and the structure of food webs).
4. Connectivity webs only offer a static view of the world. Food webs are usually idealized representations that show all linkages that occur over large spatial and temporal scales. Therefore much of the important variability and changes due to local environmental conditions and great natural heterogeneity are lost.

The consensus today is that the connectivity webs are highly over-simplified and probably speak more about the pattern of the human mind and constraints on edited publications than of real biological communities. They provide quick glimpses of a community but there is little value in their analysis. The hope is that someday, the actual food web of a few communities can be delineated completely. Analyses of a few such communities could offer great insight into the structure of complex natural communities.

Energetic Food Webs

Starting with the classic studies of Summerhayes and Elton and Lindemann (1942), food web studies turned towards

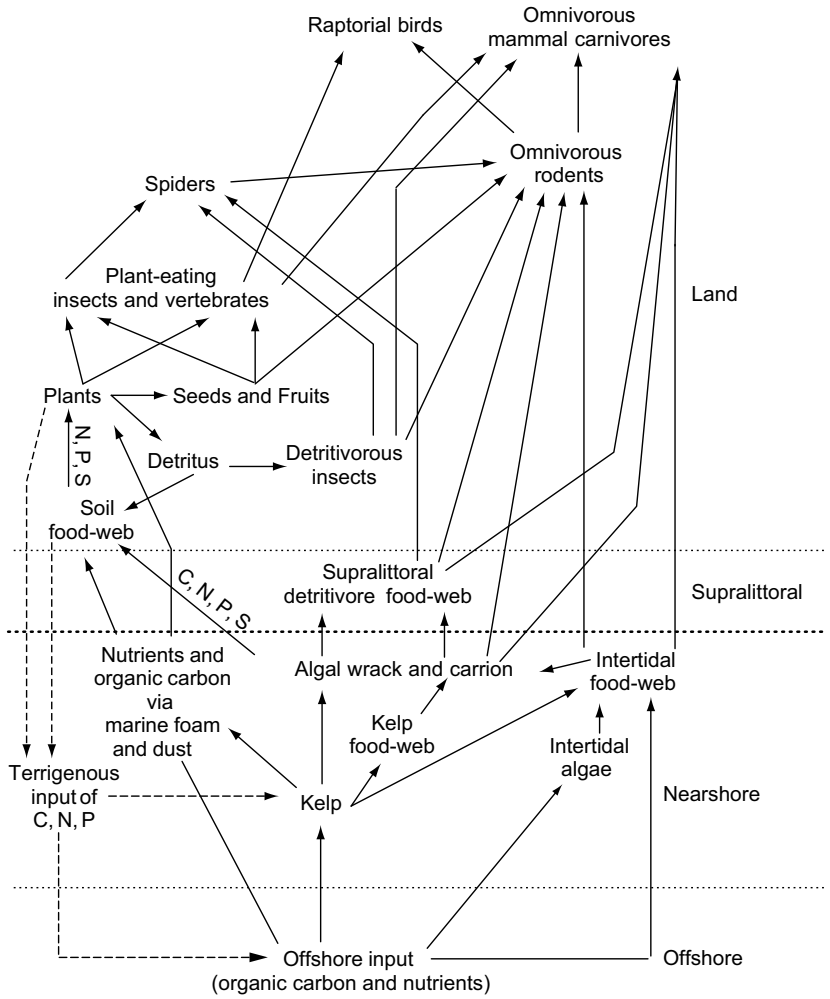


Figure 2 Food web

quantifying flows of energy and nutrients in ecosystems and the biological processes that regulate these flows (e.g., Figure 2). This approach is an alternative to connectivity food webs to describe trophic connectedness within communities. This process-functional approach explicitly incorporates producers, consumers, detritus, abiotic factors, flow out of a system and the biogeochemical recycling of nutrients. It views food webs as dynamic systems in time and space. Such an approach necessitates analyzing energy and material fluxes in order to understand the behavior of ecosystems. Thus, a typical analysis would quantify the amount of energy or matter as it travels along different pathways (e.g., plants → consumers → detritus → decomposers → soil). For example, the tracking of energy and dichloro-diphenyl-trichloromethane (DDT) through a food web in a Long Island estuary enabled researchers to study bioaccumulation effects on top predators.

This approach has provided a rich understanding of the natural world and allowed us to understand a great deal about ecosystems. Several important processes are included in energetic webs. First, they quantify energy and material pathways and key species or processes that facilitate or impede such flows. Second, they include an explicit recognition of the great importance of detritus, a subject virtually ignored in connectivity webs. (10–99% of all primary productivity from different habitats immediately becomes dead organic detritus rather than being eaten by herbivores; see the section on detritus). Third, this approach recognizes that a great amount of energy, nutrients and prey originates outside the focal habitat, a key insight to understanding natural communities. These resources are also known as spatial-subsidies. In general, nutrient inputs (nitrogen (N), phosphorus (P), trace elements) increase primary productivity; detrital and prey inputs produce population increases in their consumers. Thus, energetic webs show how ecosystems function and which species dominate biomass and energy.

The analysis of energy and matter flow is also necessary and central to understanding the dynamics of populations and communities. The success of a population is always

strongly related to the energy and biomass available to it. Consequently, it is difficult or impossible to understand the dynamics and structure of food webs and interacting populations without incorporating energy flow from below. However, this energetic approach *per se*, although necessary, is not sufficient by itself to understand the dynamics of communities because energy flow and biomass production are themselves functions of interactions among populations within the food web. The transfer of energy and matter becomes complicated as these resources pass through the many consumers that populate community food webs. For example, increasing the amount of nutrients to plants may increase the biomass of each consumer in the web or may just increase the biomass of a subset of consumers (e.g., only the plants; plants, herbivores, and predators; only the herbivores; only the plants and predators), depending on the relationship between consumers and their resources.

In order to understand the dynamics of energy and material transfer, pathways must be placed in the context of functional food webs to determine how organisms modify and regulate such transfers.

Functional or Interactive Food Webs

These webs use experiments to determine the dynamics within a community. They are also called interaction or dynamical webs. Starting with Connell (1961) and Paine (1966), empiricists began to examine the species to species interactions within communities and food webs to discover which species or interaction most influenced population and community dynamics. They did so with experiments, whereby they manipulated species that natural history or energetic analyses suggested were important. They used either press (continual) or pulse (singular) experiments to manipulate populations of single species and follow the response of other species within the food web and community. The philosophy of these studies was to simplify the complexity of natural systems with the recognition (belief) that many species and links between species were unimportant dynamically. Paine (1980) tested this belief and found that indeed many links between species were weak (essentially zero), but a few very strong species often characterized a community.

Experimental analyses of food webs are designed to identify species and feeding links that most influence population and community dynamics. These alone are placed into an interaction web that, in theory, encompasses all the elements that most influence the distribution and abundance of member species. However, unlike connectivity webs, key species are identified through experiments rather than diet frequency or energy transfer. The initial process of choosing certain species and interactions for experiments and excluding others is subjective, optimally based on strong intuition and a rich understanding of the natural history. As the researcher learns more, some elements are discarded and others are subject to further experimentation. Eventually, the community is distilled into an interaction web, a subset including only species that dominate biomass and/or regulate the flow of energy and matter.

This approach has been used by experimental and theoretical ecologists to produce a rich understanding of the processes that most influence their communities. They have been remarkably fruitful and have introduced many food web paradigms right to the center of ecology: e.g., keystones species, the intermediate disturbance or predation hypothesis, the size-efficiency hypothesis, top down and bottom up control, trophic cascades, apparent competition.

However, this approach is not without limitations. Three major problems stand out. First, many statistical shortcomings can beset experimental manipulation of food webs. For example, replications are commonly difficult (time consuming, expensive) and so experiments often lack the degree of replication and the statistical power necessary to avoid Type 2 statistical errors (significant biological differences exist among treatments but low sample size precludes their detection statistically). Second, the number of possible experiments is almost infinite. Which ones should be conducted, which species should be manipulated?

The third and perhaps most troublesome problem is that experiments isolate a subset of species and links from the community food web, largely ignoring how manipulations interact with the remainder of the community. Thus, unobserved indirect or higher order interactions may exert important effects on the dynamics of experimental species and, in theory, make the outcome of experiments indeterminate. For example, predators are thought typically to suppress their prey. However, if a predator is omnivorous, eating not only the prey but also consuming a more efficient predator of the same prey (i.e., it is an intraguild predator), it may actually relax the predation load on their shared prey, thus increasing the shared prey's abundance. To illustrate: guilds of biological control agents must be carefully structured because some species eat not only the host but also other predators/parasitoids and thus their presence decreases the number of control agents and increases target pest populations. Many other cases exist where consumers, via such intraguild predation, may indirectly facilitate its prey while concurrently exploiting it via direct consumption. Another example of indirect effects mediated by species other than the studied focal species is shown by the interaction between Australian bell miners and their homopteran (an order of plant feeding insects) food. After these birds were removed experimentally, the insects first increased greatly in number and then vanished when other bird species invaded the now

undefended miner bird territories. Thus, the apparent effect of leaf miners on lerp insects (here, suppression or facilitation) depends on when the insects were surveyed. Such complications have undoubtedly interfered with seemingly clear interpretation of many experiments. The caveat is seemingly clear: experiments can be indeterminate, producing contradictory, counterintuitive or no results, depending on the relative strengths of the direct and indirect effects.

These problems can be anticipated and partially negated with the application of good intuition of the natural history of the system and important mechanisms. Such intuition is a product of intimate empirical knowledge gained through observation and guided by a conceptual awareness of which of the interactions are potentially important. Initially, this process is essential to design the appropriate experiments and identify which species and trophic links may be dynamically important. At the end, experiments must be interpreted in a food web context to assess possible indirect and higher order effects. Experimental results must be complemented with good descriptive, mechanistic and comparative data to produce a deep understanding of the system. This is one role for energetic and dietary data. Experiments in the absence of natural history often do not succeed and may mislead.

In summary, this section conveys many important messages. The complex food webs of natural communities can be simplified and understood by isolating key species and links into interaction webs. Experiments are absolutely necessary for this process. Experiments must be designed and interpreted with sound intuition based on natural history, the community's food web and theory.

CURRENT TOPICS/TRENDS IN FOOD WEB STUDIES

Here relatively under-studied aspects of food webs perceived to be central to understanding populations, communities, and ecosystems are identified. Some of the topics are now major focal points for food web research, both empirical and theoretical.

Omnivory and the Structure of Food Webs

Consumers are classified in terms of resource specialization (i.e., number of species eaten within one group, e.g., plants) or trophic specialization (i.e., number of different types eaten, e.g., plants, detritus, arthropods). Species vary from resource specialists that eat a few species of the same resource type to trophic generalists (omnivores) that feed on several food types. Empirical research and logic have shown that the vast majority of consumers on this planet are quite omnivorous, feeding on many types of food throughout the entire food web. This is not to say that all species are so catholic in their diets. Specialists abound, e.g., many herbivores or parasites consume only specific plants or hosts. However, these form a minority of consumers. The ubiquity of omnivory carries many implications for our efforts to produce theory and models to understand how food webs function in nature.

Omnivory occurs ubiquitously when consumers eat prey from general classes of prey (e.g., plankton, arthropods). The existence of multiple trophic types within these classes causes consumers to feed on species from many trophic levels. To illustrate: arthropodivores such as many birds or lizards eat whatever properly sized arthropods are available (e.g., predaceous spiders and insects; insect parasitoids, herbivores and detritivores) without pausing to discriminate amongst their prey according to trophic status. For example, in the Coachella Valley desert delineated by Polis (1991) over 10 years of study, predaceous and parasitoid arthropods formed 41% of the diet of vertebrate and 51.5% of invertebrate arthropodivores, the remainder of the diet being herbivorous and detritivorous prey. Similarly, inspection of diet data of planktivores, piscivores, insectivores, carnivores, or benthic feeders reveals that such omnivory is almost universal with the exception of those few taxa that specialize on one or a few species of prey.

Another important type of omnivory occurs when consumers eat whatever resources are available or abundant at a particular time or place, regardless of their trophic history. The diet of a single species usually shows great differences through time (e.g., seasonally) and space (patches or habitats). For example, many, or perhaps most, vertebrates opportunistically switch from plant to animal foods with season. For example, granivorous birds, rodents and ants primarily eat seeds but normally feed on arthropods (i.e., insects from all trophic levels and spiders) when they are abundant in spring. Alternately, many omnivorous, arthropodivorous and carnivorous species consume significant quantities of seed or fruit. In the Coachella Valley, 79% of 24 primary carnivores eat arthropods and/or plants. For example, coyotes eat mammals (rabbits, rodents, gophers, antelope, ground squirrels, kit foxes, other coyotes), birds (including eggs and nestlings of roadrunners, doves, quails), snakes, lizards, young tortoises as well as scorpions, insects and fruit. Overall, it appears that most consumers eat whatever is available that they can catch.

Life history omnivory describes the great range of foods eaten during growth and development by most species. Such omnivory includes abrupt diet changes in species undergoing metamorphosis (e.g., many marine invertebrates, amphibians, and insects that undergo complete metamorphosis – also called holometabolous insects) and gradually changing diets in slowly growing species (e.g., reptiles, fish, arachnids, and insects that do not undergo complete metamorphosis – also called hemimetabolous insects). Changes at metamorphosis can be great; e.g., 22% of the insect families in the Coachella

Valley desert community undergo radical change in diet: larvae are predators or parasitoids and adults are herbivores. Although not as dramatic, significant changes characterize slowly growing species so that differences in body size and resource use among age classes are often equivalent to or greater than differences among most biological species.

Incidental omnivory occurs when consumers eat foods in which other consumers live. Thus, scavengers and detritivores not only eat carrion or organic matter but the trophically complex array of microbes and macroorganisms that live within these foods. Frugivores and granivores commonly eat insects associated with fruits and seeds. Predators eat not only their prey but the array of parasites living within the prey. In each case, consumers automatically feed on at least two trophic levels.

These types of omnivory are widespread and common. Their ubiquity poses many questions. First, how does omnivory affect food web structure? Most obviously, it increases complexity and connectivity. Second, can we ignore omnivory in the analyses of food webs? By its very nature, omnivory causes consumers to have a great number of links, each of which may be numerically rather unimportant in the diet. For many reasons delineated below, we cannot arbitrarily ignore apparently minor diet links if we hope to understand dynamics (see next section).

Food Webs as Open Systems

Virtually all natural systems are open and can exhibit tremendous spatial heterogeneity. Nutrients and organisms ubiquitously move among habitats exerting substantial effects on the food webs (e.g., increased productivity and a subsequent increase in consumers) in which they are deposited. Trophic linkage between habitats depends upon the degree of differentiation in the structure and species composition. Systems that have moderate differences tend to have broader transition zones and greatly overlap in species composition and include grassland–forest, littoral–sublittoral, and benthic–pelagic zonations. Habitats that have significant and abrupt changes in structural and species composition include aquatic–terrestrial and photic–aphotic zones. Photic zones are those in which phytosynthesis by phytoplankton is the primary source of productivity; whereas aphotic zones are those areas where photosynthesis does not occur and the energy sources are dead organisms sinking from the photic zones.

Moving resources (energetic or nutrients) can be utilized by different trophic levels and the organisms that move across boundaries also may be of different trophic levels (e.g., predators and prey). Studies of communities on island systems by Gary Polis and his research group have shown that most of the allochthonous resources (i.e., inputs from another habitat) from the ocean are available to detritivores and predator/scavengers. Other systems whose consumers are heavily dependent on such inputs include: caves; mountaintops; snowfields; recent volcanic areas; deserts; soil communities; marine filter feeding communities in unidirectional currents or advective areas; and shaded headwater streams that receive most energy from leaf litter inputs. The movement of herbivores such as hippopotamuses and geese across habitat boundaries to feed can also result in large energetic and nutrient flows across habitats.

Allochthonous inputs into the top level include carrion or carcasses, the movement of prey species into the habitat and movement of predators across habitats. For example, the *Allen paradox* describes cases where secondary production within streams is insufficient to support levels of fish production found in them. Similarly, studies of coyote populations along the coast in Baja California demonstrate that they are highly subsidized by inputs from the ocean (47.8% of their diet) and are able to maintain a three to ten times higher density than in adjacent inland areas. Predators moving along the interface between ecosystems (i.e., shorelines, river banks, benthic and pelagic systems) can utilize resources across habitat. In addition, allochthonous inputs may enter at multiple trophic levels. The river continuum concept argues that allochthonous resources entering into small headwater streams provide much of the productivity for organisms downstream in larger order streams. These allochthonous resources include prey, dissolved and particulate organic matter, and litter fall. This type of pattern is also seen in estuarine systems in which rivers carry allochthonous inputs into estuaries. Similarly, run-off from terrestrial systems into aquatic systems (and *vice-versa* in the case of marine to terrestrial) provides litter, dissolved and particulate organic matter, and prey.

Spatial coupling can be key to dynamics (Polis and Strong, 1996). Abundant detrital kelp from the sublittoral zone promotes dense intertidal limpet and urchin populations, that then graze on non-coralline algae. Allochthonous subsidies commonly influence stream systems: leaf fall subsidizes herbivores, who in turn, depress algae. Many species that live along the coasts of streams, rivers, lakes or the ocean are often very dense because they feed on aquatic insects. Subsidies allow consumers to be more abundant than if supported solely by *in situ* resources with consequent suppression of *in situ* resources decoupled from *in situ* productivity.

Detritus

Little of the energy fixed by plants passes directly into the grazing food chain – herbivores eat plants and are eaten by carnivores. Most primary productivity is uneaten by herbivores (median $>80\%$ on land, $\sim 50\%$ in water). Uneaten plants (and animals) enter the detrital web where

they are processed by microbes and some Metazoa. Most ecosystems are not net accumulators of undigested biomass, and most plant biomass is thus soon digested by detritivores, with nutrients and energy passing through functional compartments comprised of diverse microbes and animals where several factors regulate the flow and control the availability of detritus and detritivores to consumers. What happens to this dominant chunk of the world's productivity? Is the detrital web a self-contained sink internally recycling energy and nutrients, or a link that affects the population dynamics of the larger species?

Detrital energy and nutrients reenter grazing chains when detritivores are eaten by predators that also eat herbivores. Such detrital shunts are common, interweaving energetics and dynamics of *biophages* (i.e., consumers of living organisms) and *saprophages* (i.e., consumers of dead organic matter). Bypassing herbivores, this linkage can affect herbivore regulation in a manner analogous to the spatial subsidies to consumers discussed above. Predator populations, subsidized by detritivorous prey, can increase and suppress other predators or herbivores.

The exact effect of detrital shunts depends on the relative benefits for each species and where detritus reenters (to producers, herbivores, intermediate or higher consumers). For example, nutrients from detritus greatly influence plant productivity; models show that a 10% reduction in detritus can cause a 50% reduction of plant biomass. The dynamics of consumer control within the detrital web and those produced by infusion of detritivores into the grazing web are crucial to community structure and dynamics.

Age Structure Effects in Food Webs

Almost all species display complex life cycles, marked by moderate to radical changes in diet and habitat; such life histories fundamentally must affect every species with which they interact. Yet, our understanding of how age and stage-structured processes affect food webs and communities is embryonic.

Life history omnivory describes shifts in diet during development. Such shifts are often accompanied by ontogenetic changes in habitat. Diet can change substantially either discontinuously (e.g., at metamorphosis) or slowly with growth. Such life histories are widespread; an estimated 80% of all animal species undergo metamorphosis. Changes in resource use can be dramatic (e.g., predaceous juveniles – plant-feeding adults in parasitoids and many other insects; herbivorous tadpoles and predaceous frogs and toads) with prey size variation as great as three to four orders of magnitude. Even among non-metamorphic species, diets change greatly with age, with diet differences among age classes often more distinct than those among most species.

Life history omnivory can either impede consumer control or amplify resource suppression via dynamics similar to those of spatial subsidy or detrital shunts. Overall, complex life histories and age structure omnivory can exert diverse effects on food chains, from thwarting or amplifying top-down control of resources by consumers.

The Roles of Nutrients and Stoichiometry

Animals require both energy and a variety of nutritional requisites to grow, complete their life cycle, and reproduce. Important nutrients include nitrogen, phosphorus, some trace elements, fatty acids, and vitamins. Nitrogen is an integral component of many essential compounds – it is a major part of amino acids, the building blocks of protein, including the enzymes that control virtually all cellular processes. Other nitrogen compounds include nucleic acids and chlorophyll. Phosphorus is used for adenosine triphosphate (ATP – the energy currency of all cells), nucleic acids (deoxyribonnucleic acid (DNA), ribonucleic acid (RNA)), and phospholipids, particularly in cell membranes.

The availability of nutritional requisites constrains growth and reproduction in virtually every species. Nitrogen and phosphorus are particularly important. The ratio of carbon (C) to nitrogen (C–N) in plants ranges from 10/1–30/1 in legumes and young green leaves to as high as 600/1 in some wood. The C–N ratios in animals and microbes are much lower, ordinarily falling between 5/1 and 10/1. Such differences in C–N ratios between plants and their consumers lower the rate of decomposition by microbes. There is ample evidence that heterotrophs chronically lack adequate nitrogen to grow or reproduce optimally. The importance of nutritional restriction is reinforced by the foraging literature that clearly shows that herbivores choose their foods based on nutrient as well as energy content.

In many cases, phosphorus availability constrains herbivore success. The Redfield ratio describes the approximate stoichiometric mix (110 C–250 H–75 O–16 N–1 P) of elements essential to organisms (*see* **Redfield Ratio**, Volume 2). In particular, the N–P ratio crucially determines productivity and species composition. Thus, energy (C–C bonds) and nitrogen could be abundant, but neither individuals nor populations grow maximally because phosphorus is insufficient. As phosphorus is essential to cell division (and thus reproduction), a high N–P ratio especially limits the growth of organisms that have high potential r-max (high reproductive ability), such as most herbivores and detritivores. These organisms are key to the potential regulation of plant biomass (and detritus). Evidence now suggests that high N–P ratios can impede trophic cascades. For example, *Daphnia*, a key to many lake cascades, respond sufficiently rapidly to phytoplankton productivity to depress plant biomass. In lakes with inadequate phosphorus, slower

growing copepods replace *Daphnia*; these copepods do not have the reproductive capacity to depress phytoplankton biomass.

Ecologists are beginning to understand how stoichiometry and nutritional balance affect population and food web dynamics. Nevertheless it is extremely likely that herbivore growth is often less than maximal solely because its environment does not provide sufficient quantities of all key nutritional requisites. In fact, the greatest disparity in biochemical, elemental, and stoichiometric composition in the entire food web occurs at the link where herbivores convert plant material into animal tissue. The implication is clear: even in a world full of green energy, many or most herbivores cannot obtain enough requisite resources to grow, survive, or reproduce at high rates. Nutritional shortages regulate herbivore numbers and often limit their effects on plant biomass.

RECOGNITION

Professor Gary Allen Polis died in a boating accident on March 27th, 2000. The loss of Dr Polis is a great tragedy to those of us who knew Gary professionally and personally – we will miss him greatly.

REFERENCES

Connell, J H (1961) The Influence of Interspecific Competition and Other Factors on the Distribution of the Barnacle Chthamalus Stellatus, *Ecology*, **42**, 710–723.

Darwin, C (1859) *The Origin of Species*, John Murray, London.

Forbes, S (1887) The Lake as a Microcosm, *Bull. Ill. State Nat. Hist. Surv.*, **15**(1925), 537–550.

Lindeman, R (1942) The Trophic-dynamic Aspect of Ecology, *Ecology*, **23**, 399–418.

Paine, R T (1966) Food Web Complexity and Species Diversity, *Am. Nat.*, **100**, 65–75.

Paine, R T (1980) Food Webs: Linkage, Interaction Strength and Community Infrastructure, *J. Anim. Ecol.*, **49**, 667–685.

Polis, G A (1991) Complex Trophic Interactions in Deserts: an Empirical Critique of Food Web Theory, *Am. Nat.*, **138**, 123–155.

Polis, G A and Strong, D R (1996). Food Web Complexity and Community Dynamics, *Am. Nat.*, **147**, 813–846.

FURTHER READING

Hairston, Sr, N G, Smith, F E, and Slobodkin, L B (1960) Community Structure, Population Control and Competition, *Am. Nat.*, **94**, 421–425.

May, R (1973) *Stability and Complexity in Model Ecosystems*, Princeton University Press, Princeton, NJ.

Oksanen, L, Fretwell, S, Arruda, J, and Niemela, P (1981) Exploitation Ecosystems in Gradients of Primary Production, *Am. Nat.*, **118**, 240–261.

Forest Gap Models

Herman H Shugart
University of Virginia, Charlottesville, VA, USA

Computer models designed to simulate the dynamics of a forest by following the fates of each individual tree in a forest stand were created initially in the mid-1960s. The earliest such model was built by Newnham (1964), followed by similar developments in several Schools of Forestry in the US. These early models involved using a digital computer to dynamically change a map of the sizes and positions of each tree in a simulated forest. Importantly, these initial models were as complex and as detailed as their present day descendants. With increased computing power becoming cheap and available, applications using these models have become increasingly frequent (Shugart and West, 1980).

Gap models simulate the changes in a small patch of forest on the order of 0.1 ha in size. The models project the annual growth of each individual tree on the patch as well as the death of each tree and the regeneration of new individual trees. Gap models are important because of their widespread applications in assessing the effects of environmental change on forests. The models have been applied to develop theories about the overall functioning of forested landscape and have also been used to reconstruct the composition of forests of the past and under future conditions.

Gap models are a special case of the more general individual-based models. They have been used particularly for forest ecosystems. Because gap models have been applied in a very wide range of ecosystems, often the term, gap model, is used interchangeably with other individual-based models when applied to mixed age and multi-species ecosystems. The models simulate landscapes as dynamically changing mosaics with each element of the mosaic having an interactive population of trees of different species and sizes. Output from the models resembles field data as sampled from a standard forest survey unit: a list of the diameters of each tree (and its species) alive on the patch for each year. These model data can be transformed to estimates of biomass (and hence, net production), leaf area profiles through the canopy, and tree ring histories for each tree.

INDIVIDUAL BASED MODELS IN FOREST ECOLOGY

Since gap models are a relatively widely used subset of individual-based models, it is appropriate to initially describe this class of ecological model and the associated modeling philosophy. Since the mid-1960s, a broad range of ecologists (including animal behaviorists, population

biologists, and ecosystem ecologists) have independently created a diverse array of such models (Huston *et al.*, 1988). The catalyst for this development was undoubtedly the increased availability and power of digital computers, the necessary tools needed in order to use most of these models. The earliest such models were developed by population ecologists (Holling, 1964; Rohlf and Davenport, 1969) interested in incorporating animal behavior into population models and foresters who were interested in relating tree growth to forest yield.

Most ecologists are interested in variations in individuals (a basis for the theory of evolution and a frequently measured aspect of plants and animals) and appreciate spatial variation as being quite important. Increasingly, models that simulate the dynamics of ecological systems by accounting for changes in each of a large number of individuals in the system have been developed and applied in population and ecosystems ecology (Huston *et al.*, 1988; DeAngelis and Gross, 1992). One advantage of such models is that the following two implicit assumptions associated with the more traditional approaches used in ecological population modeling are not necessary:

1. The unique features of individuals are sufficiently unimportant to the degree that individuals are assumed to be identical.
2. The population is perfectly mixed so that there are no local spatial interactions of any important magnitude.

Individual-based models spring from a rejection of these assumptions. Indeed, these assumptions seem particularly inappropriate for trees which are sessile and which vary greatly in size over their life span. This may be one of the reasons that tree-based forest models are among the earliest and most widely elaborated of this genre of models.

A HISTORY OF THE DEVELOPMENT OF THE GAP MODELING CONCEPT

Among the earliest individual-based models in ecology were models of forests based on the growth of the individual trees that comprise the forest stand. These models were developed by quantitatively oriented foresters and were focused from their inception toward practical issues in production forestry. That trees were an initial focus for individual-based models springs naturally from the ecology and biology of trees. An emphasis on sizes of individual trees in the understanding of forest change has been traditional in forestry and forest ecology for over 200 years. This, associated with an interest in the changes in sizes of trees with changes in numbers (from birth and death processes), lead naturally to an interest in simulating the changes in individual trees as a basis for understanding forest dynamics. Some of the first individual-based models used in ecology were for trees.

Individual-based models in forestry were focused on solving several general problems involving *stand yield tables*. Modern forestry uses stand yield tables as a scientific basis for making decisions about the expected productivity of forest stands under different conditions to plan harvest and thinning schedules. Yield tables use the site index (the expected height growth of trees at a given age) to calibrate the expected changes in tree sizes and densities based on study plots remeasured over decades or centuries. Stand yield tables are calibrated by using data on the performance (height and diameter growth) of individual trees from study plots, with different densities of trees per unit area and grown under different conditions. Some of the European yield tables are based on continuous studies of such plots for as long as 200 years. However, the large detailed data sets required to develop a yield table represent a weakness in the approach. If one is interested in the growth and yield of plantations of an exotic, such as Radiata Pine (*Pinus radiata*), a species that is being planted in dozens of tropical and sub-tropical countries as a source of softwood, then how does one develop a yield table in the absence of a long data calibration set? Similar predicaments arise in evaluating altered forestry practices such as forest fertilization; in projecting response of forests affected by air pollution; in predicting yield in forests with genetically improved species. Indeed, any condition that alters the performance of individual trees from the conditions for which a yield table is calibrated is a potential source of difficulty. This very early interest in projecting forest performance under altered conditions presaged the use of this class of models in problems involving the prediction of climate change effects on forests.

GAP MODELS

An important sub-category of individual organism based tree models that have been widely used in ecological (as opposed to traditional forestry applications) are the so-called gap models. Shugart and West (1980) coined the term, *gap model*, to emphasize the similarity of the simulation unit in the models (a ca. 1/10 ha plot) and the venerable concept of mosaic dynamics as a basic feature of the dynamic change in forests (Watt, 1947). The first such model was the JABOWA model (Botkin *et al.*, 1972) developed for a forest in New England. Several gap models were subsequently developed (see Shugart, 1998 for a more detailed review) for a wide range of forests; and eventually, for grasslands and savannas (Coffin and Lauenroth, 1989; Menaut *et al.*, 1990). The earlier versions were used for inter-ecosystem comparison (Shugart, 1984) and adhered relatively closely to the original JABOWA formulation. More recent models have diverged, particularly in the addition of spatial competition among trees, in a range of modifications to allow simulation of a wider range

of environmental conditions, and in additional biological and/or physiological mechanisms. These latter models are essentially at the same level of detail and spatial resolution as the first individual based forest models developed in the 1960s.

STRUCTURE OF GAP MODELS

In a gap model, as with other individual-based models, each tree is simulated as an independent entity. The model structure emphasizes two features important to a dynamic description of a vegetation pattern:

1. The response of the individual tree to the prevailing environmental conditions.
2. How the individual tree modifies those environmental conditions.

The models are hierarchical in that the higher level patterns observed (population, community, and ecosystem) are the integration of plant responses to the environmental constraints defined at the level of the individual.

The models include random processes with the birth and death of trees. To implement the models, one typically averages the responses of 100 or so independent simulated plots. The models can assemble forests by producing ecological succession from bare ground. The models can also be initiated from forest survey plots to simulate a parcel of land from the time it was sampled. The richness of the output data allows a number of direct comparisons with forest survey data and for this reason the models have been tested extensively in a number of contexts, including the reconstruction of past forests and associated climates.

The parameters in the models are essentially of two sorts: First, the individual tree parameters are used in species-specific equations for the growth, relation between diameter and height, and other biological and ecological descriptors. Second, parameters for equations in the model that characterize the interactions among trees (for example, shading of light through different levels of the canopy, water availability) or that simulate the overall condition on the plot (depth of leaf litter, soil temperature, etc.).

Growth

The growth of an individual is calculated using a species-specific equation to predict the expected diameter increase for each tree under optimal conditions (Figure 1). This increase in diameter is functionally related to the tree's current diameter. This optimum increment is then modified by the environmental response functions and the realized increment is added to the tree. There are several different

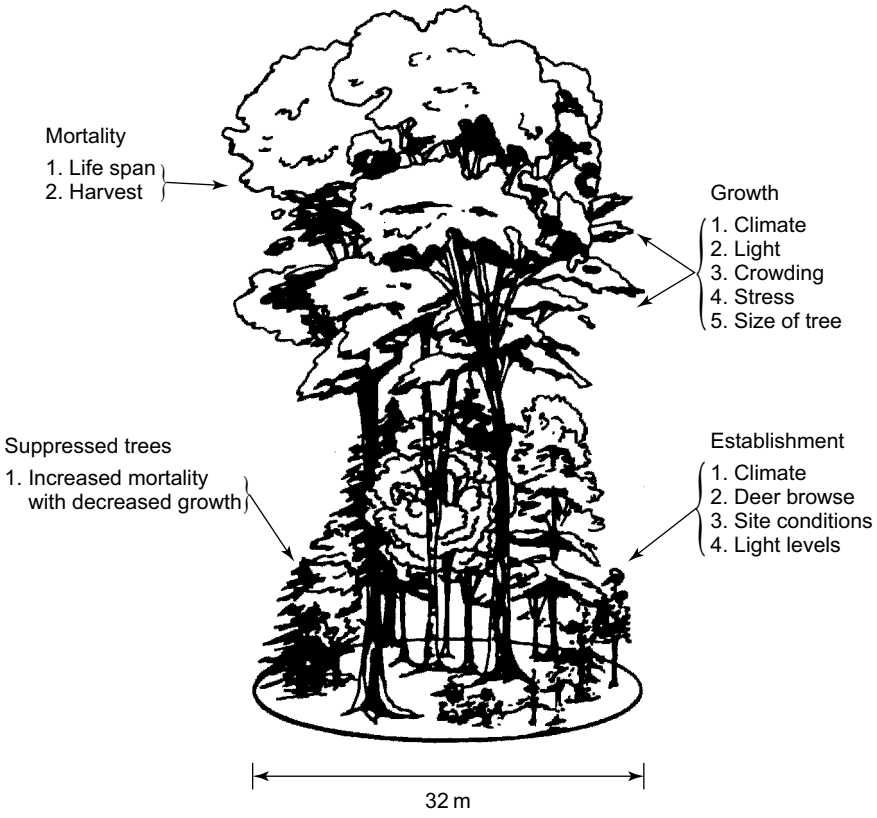


Figure 1 Schematic diagram of the function of a typical gap model

growth equations used in the set of gap models developed over the past three decades. Usually they are developed with an assumption that growth in trees is the consequence of two opposite processes:

1. A positive process based on the argument that as trees get larger, the leaf area of the tree increases, which increases potential growth. The magnitude of this positive growth rate depends on the net photosynthetic rate of the tree per unit of leaves.
2. Opposing this positive rate is a negative rate that is associated with the energetic cost of maintaining a given volume of living tissue. Presumably, as a tree increases in size, these negative costs increase.

The difference in these two rates is the actual growth, which becomes smaller as trees become larger. Botkin *et al.* (1972) included these effects in the initial JABOWA model using a simple derivation based on arguments on the relations among diameters, leaf areas, growth, height and longevity of trees. A principal advantage of this initial formulation was that the parameters for the models could be obtained from available ecological descriptions of tree species silvics.

Moore (1989) developed what is probably a better tree growth equation, if such an equation is based on the assumptions stated in the initial formulation. Moore noted that the initial formulation used in the JABOWA model implied that the losses associated with large size in trees increased at a rate greater than the increase in the tree's volume. He developed an alternate formulation based on a simplification of a carbon metabolism-based representation of the constructive versus destructive metabolism of the individual tree. This formulation currently is used in the FORCLIM model (Bugman and Solomon, 1995). Leemans and Prentice (1987) derived an equation for tree volume increment that included the effects of vertical canopy geometry. Pacala *et al.* (1993) argued that gap models should have the parameters in the models tied to direct empirical observations. Their model developed toward this goal uses several regression relations to improve the response of an individual-based model intended as an extension of the existing gap models. Growth is driven by an index of local light availability that integrates annual light condition (a complex function of changing sun angles, ratios of diffuse and direct radiation, and the canopy pattern).

Environmental Constraints and Resource Competition among Trees

Gap models simulate each individual tree's response to light availability at height intervals on the plot. Even in the early gap models, the vertical structure of the canopy is modeled explicitly. The sizes of individuals (height and leaf area predicted by regressions on each tree's diameter) are used to construct a vertical leaf area profile. An equation for the extinction of light through the canopy is computed to capture the effect of trees shading one another.

Other resources are incorporated to varying degrees in different gap models. These include soil moisture, fertility (often available nitrogen, specifically), temperature, as well as disturbances such as fires, hurricanes, floods, and wind throw. In most of the models, the environmental responses are modeled via a constrained potential paradigm; a tree has a maximum potential behavior under optimal conditions (i.e., maximum diameter increment, survivorship, or establishment rate). This optimum is then reduced according to the environmental context of the plot (e.g., shading, drought), to yield the realized behavior under ambient conditions. The curves describing species response to environmental resources tend to be generic curves that scale between 0.0–1.0, and species are often categorized into a small number of functional types.

The limitation effects of resource shortages in the models were initially computed in a multiplicative fashion (Botkin *et al.*, 1972; Shugart, 1984). This would mean that if, due to shading, the lack of light reduced the optimal growth of a tree to 90% of the optimal and water availability was 80% of optimal, then the growth would be 72% of the optimal ($0.90 \times 0.80 = 0.72$). Some later models use other approaches, notably assigning the growth reduction to that of the most limiting factor (Pastor and Post, 1986).

Competition in the model depends on the relative performance of different trees under the environmental conditions on the model plot. These environmental conditions may be influenced by the trees themselves (e.g., a tree's leaf area influences light available beneath it), or may be modeled as extrinsic and not influenced by the trees (e.g., temperature).

The tree that has the best performance, relative only to other trees on the plot, is the most successful. Competitive success depends on the environmental conditions on the plot, which species are present, and the relative sizes of the trees. Each of these varies through time in each simulation from a gap model. Competition for light is asymmetric and exploitative: a tree at a given height absorbs light and reduces the resource available to trees at lower positions in the canopy. In most gap models, competition for belowground resources (water and nutrients) is symmetric. Each tree experiences a resource level common to the plot. This approach to modeling the competition process has the result that competitive ability depends strictly on the context of the modeled gap.

Mortality

The death of individuals is modeled as a stochastic process. Most gap models have two components of mortality: age related and stress-induced (Figure 1). The age-related component applies equally to all individuals of a species

and depends on the maximum expected longevity for the species; this age is typically on the order of 300 years, yielding an annual survivorship rate of around 1–2%. The probability of mortality of each individual tree is determined as a stochastic function. Each tree is assumed to have an intrinsic mortality rate such that, under normal conditions, 1% of the individuals in a cohort could be expected to live long enough to attain their maximum height and diameter.

The mortality of plants, in general, and certainly of trees is a relatively poorly known aspect in plant ecology and the formulations in many of the gap models reflect the state of knowledge. The tendency in most of the existing gap models has been to use relatively straightforward descriptions of mortality in the anticipation of additional information. Trees that are growing slowly have an increased probability of death. In the earlier gap models, trees with a minimum diameter increment of only 1 mm per year are subjected to additional mortality that has the effect of allowing only a 1% chance that they will survive 10 years. In more recent gap models, stress is defined with respect to a minimal diameter increment (typically 10% of optimum growth for a given size of tree) and individuals failing to meet this minimal condition are subjected to an elevated mortality rate.

Establishment

The complexity of the regeneration process in trees and its stochastic nature makes it nearly impossible to predict the success of an individual tree seedling. This problem is compounded when spatial effects are also considered. Most gap models are designed to treat regeneration in trees from a pragmatic view that the factors influencing the establishment of seedlings can be usefully grouped in broad classes (Figure 1). Tree establishment and regeneration are largely stochastic, with maximum potential establishment rates constrained by the same environmental factors that modify tree growth. Each simulation year, a pool of potential recruits is filtered through the environmental context of the plot, and a few new individuals are established.

Malanson (1996) has applied a spatially explicit gap model to investigate the effects of dispersal probability on forest dynamics. He found that relatively rare instances of a given species of tree entering the patch and establishing itself could have profound influences on the forest dynamics and diversity. Unfortunately, such events are also very difficult to observe, much less quantify, under field conditions.

TESTS AND APPLICATIONS OF GAP MODELS

Gap models are well suited for examining vegetation response to changing environmental conditions because the response to the environment is not limited to reproducing present day vegetation patterns. The individual-based approach also provides a useful link between more detailed physiological models and larger-scale responses. The models have also been tested in terms of their ability to reproduce important features of a wide variety of forests and other ecosystems (Table 1). Simulation models, in general, and individual-based gap models, in particular, have several features that incline scientists to expect a high standard of model testing. The models seem to lack the parsimony that is expected in models; mathematically, the models are not open to formal analysis as are many simpler population models; the high level of detail in model output implies a concomitant degree of detail in testing of models. Since the models are often used in application domains outside the range of conditions used in model development, good scientific practice requires continual model testing and assessment.

Simulations involving forest response to complex gradients and the applications of the models to simulate prehistoric forests (Table 1) are germane to the use of the models in assessment of global environmental change. While gap models have been tested to a relatively great degree compared to most ecological models, the models are under a continual revision as new phenomena are incorporated and should be continually tested. Further, any model is an abstraction and as such may not incorporate all the relevant phenomena for a given assessment. Any model extrapolations of future vegetation should be regarded with caution and compared with other equivalent evaluations from other bases.

Gap models have been applied to examine the response of forested systems to climate changes in both predictive and reconstructive modes. The models have been used in the reconstruction of prehistoric Quaternary forests (for example, Solomon and Webb, 1985). These reconstructions serve to test whether or not the species-level information in the models appears to be consistent with paleoecological data given a climate condition thought to have been the norm in the past. These reconstructive applications have inspired model projections of possible consequences of future climate changes. These model-based predictions are for changes in species composition, vegetation structure, productivity and standing biomass.

As an example of the application of a gap model to assess the effects of a regional climate change, the effects of several different climate change cases was inspected for forests growing on sites in the North American boreal forest at a site near Fairbanks, Alaska (US). The model was developed and tested on its ability to simulate the stand composition, recovery after fire and response to slope and aspect of the site. Fairbanks is in the zone of discontinuous permafrost. North facing slopes and poorly drained sites have permafrost, a persistent presence of an ice layer in the soil, and are dominated by Black Spruce (*Picea marina*), one of the few tree species than can grow under such

Table 1 Examples of the types of model tests used on various gap models; details can be found in Chapter 8 of Shugart (1998)

Type of model test	Example cases
With a priori parameter estimation for species, predict forest level features (total biomass, leaf area, stem density, average tree diameter, etc.)	Prediction of biomass for oodplain forests of the Mississippi River Predict effects of timber harvest on nutrient regimes in northern (US) hardwood forests Similar predictions for a diverse array of systems: Estonian conifer forest, New Zealand southern beech forests, Paci c Coast Douglas- r forests, West African Savanna: etc.
Run model for a long period of time, test results against mature forests preserved in a region	Predict structure and composition of relic mature forests in Sweden Predict old growth forest in montane spruce re zone in Southern Appalachians, in Mississippi River oodplain, upland forests in Arkansas and in Chinese mixed forests
Calibrate model on stands of a given age; predict stand structure on stands of a different age	Using a model that predicts the structure and composition of old growth forests in the Great Smoky Mountains National Park (US), test by independently predicting composition and structure of forest clear cut 40 and 70 years before Predictions of the composition and structure of Canadian forests, diameter structure of Puerto Rican rain forests stands of different ages, Australian rain forest at different ages after harvest
Use introduction of disease or change in conditions (e.g., re frequency, etc.) as a natural experiment	Predict the composition of mature forests in East Tennessee (US) before the introduced chestnut blight eliminated one of the major tree species Predict composition or structure of forest at different re frequencies in mixed conifer forests, for Puerto Rican montane rain forests under different hurricane conditions
Predict independent tree diameter increment data	Predict the growth and diameter increment by species and diameters for subtropical rain forest Predict diameter frequency distributions for Puerto Rican montane rain plots of different ages
Predict forestry yield tables	Test model on its ability to reproduce Swedish yield tables Predict Eucalyptus yield tables for montane Australian forests, for Loblolly Pine, for Swiss forests and for Chinese mixed forest
Predict forest composition change to single environmental gradients	Predict change from deciduous to coniferous forest in the mountains of New England Predict changes in forest composition as a function of ood frequency, at a range of locations in the boreal zone, along latitudinal gradients in China
Predict forest composition response to multiple environmental gradients	Predict composition and structure of boreal forests on north and south facing slopes and for different ages since wild re in Fairbanks, Alaska region Tests against re and altitude variables in mountains of California and Australia. Tests in complex terrain (slopes, altitudes, soils) in Switzerland
Reconstruct composition of vegetation under past climates (Paleo-reconstruction)	Reproduce 16 000 year record of forest change based on fossil pollen chronology from East Tennessee Reproduce forest composition under full glacial and other conditions

conditions. South facing slopes have no permafrost and the recovery of the system after a wildfire features a succession from birch and aspen (*Betula* and *Populus*) to a forest dominated by White Spruce (*Picea glauca*).

Using the boreal forest gap model to assess climate change effects, Bonan *et al.* (1990) and Smith *et al.* (1995) investigated the responses of the forest by taking several different examples of climate change predictions from general circulation models of the atmosphere (an example of two such cases are shown in Figure 2). The climate change was administered by running the models for several hundred years on 100 simulated plots for conditions associated with north facing and south facing slopes. The climate conditions used in the models were then transitioned to those associated with a greenhouse warming over a 50 year period (see monthly temperature and precipitation information on Figure 2). The results of this model experiment were to produce conditions in which the cold forests of Black Spruce growing on north facing slopes was largely unaffected by the climatic warming, but the warmer, White Spruce forests of the south facing slopes were strongly impacted by the change. Conditions on the south facing slopes were outside the ecological conditions under which the common tree species near Fairbanks are known to survive. For White Spruce the limiting condition appeared to be moisture stress brought on by the temperature change. Significantly, tree ring studies a decade later, which investigated the effects of the current run of warmer-than-usual

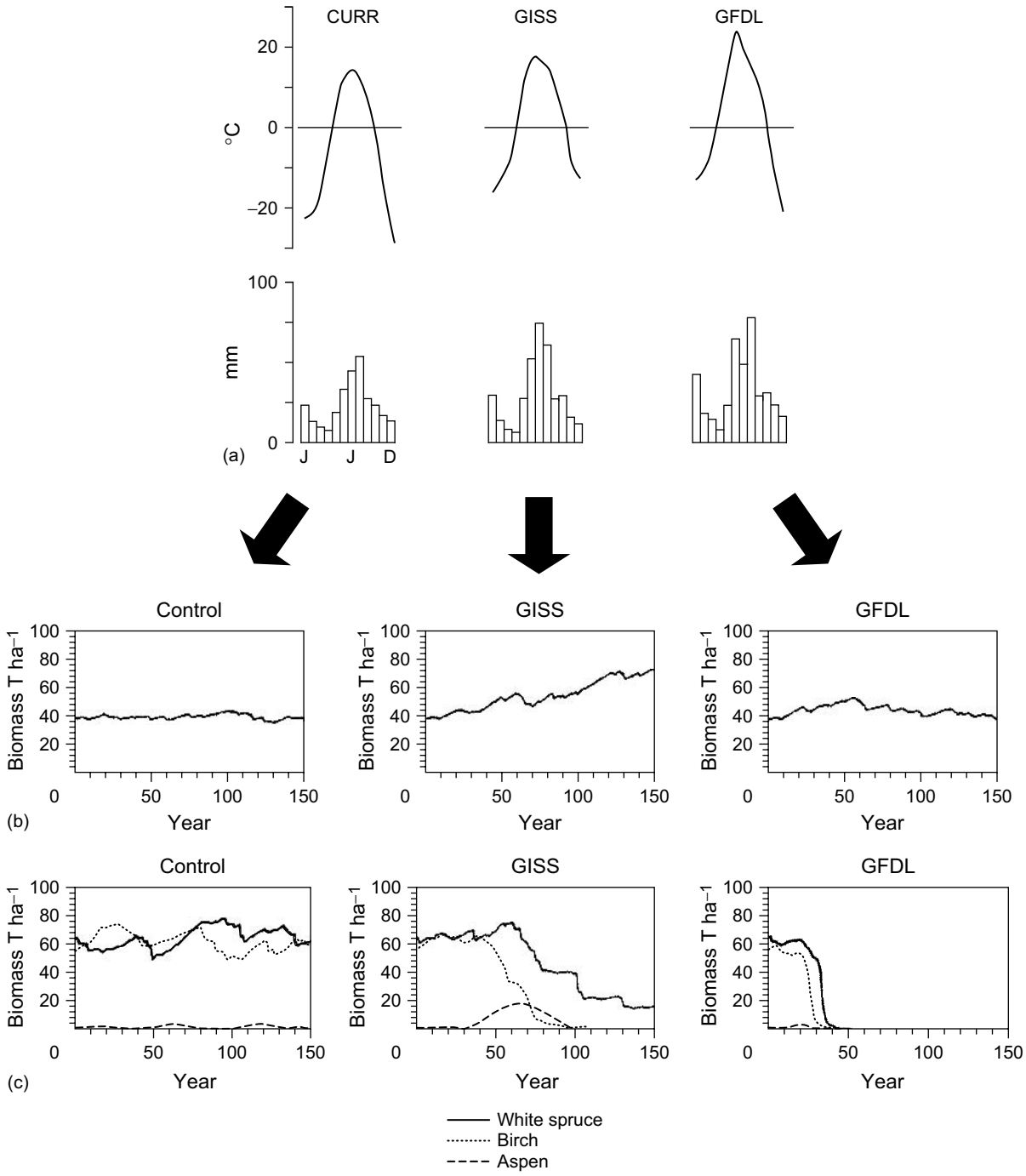


Figure 2 Application of a gap model to predict the dynamic response of forests associated with the current climate and two different climate change scenarios for forests growing in the vicinity of Fairbanks, Alaska. (a) The monthly temperature and precipitation for Fairbanks and equivalent information from the GISS Hansen *et al.* (1988) and UKMO Mitchell (1983) climate simulation models. Note the elevated winter temperatures in both models. (b) Average of 100 gap models simulating forest change over 150 years on north facing slopes. The transition from the current climate to the new climate occurred incrementally over the first 50 years. The only tree species that occurs in these conditions is Black Spruce (*Picea marina*). (c) Average of 100 gap models simulating forest change over 150 years on south facing slopes. The three species that normally occur in these settings are eliminated from these sites

decades on White Spruce near Fairbanks, have found evidence for the same moisture stress-mediated effect predicted by the gap model a decade before (Barber *et al.*, 2000).

The field validation of *a priori* predictions from a gap model is reassuring for those who would apply the models to assessing climate change. However, this example also points to a valuable attribute of gap models; the potential ability to predict change across landscapes including the heterogeneous nature of these changes. The prediction of the models for a warmer Fairbanks, Alaska are for a landscape featuring Black Spruce forests in north-facing slopes and other colder sites. A novel vegetation type, without trees, is expected over the rest of the landscape. The condition might be something like the boreal forest transition into grassland (the Taiga-Steppe) found south of Krasnoyarsk, Siberia, Russia. Other climate evaluations using gap models have also indicated a complex landscape-level response to climate change. An example (predicted as a response of a warming for forests in Minnesota, US by Pastor and Post (1988)) would be the occurrence of a landscape mosaic of both deciduous and boreal forests; each of these forest types would be stable and would have very different nitrogen cycles.

REFERENCES

Barber, V A, Juday, G P, and Finney, B P (2000) Reduced Growth of Alaskan White Spruce in the 20th Century from Temperature-induced Drought Stress, *Nature*, **405**, 668–673.

Bonan, G B, Shugart, H H, and Urban, D L (1990) The Sensitivity of Some High Latitude Boreal Forests to Climatic Parameters, *Clim. Change*, **16**, 9–31.

Botkin, D B, Janak, J F, and Wallis, J R (1972) Some Ecological Consequences of a Computer Model of Forest Growth, *J. Ecol.*, **60**, 849–872.

Bugman, H K M and Solomon, A M (1995) The Use of a European Forest Model in North America: A Study of Ecosystem Response to Climate Gradients, *J. Biogeogr.*, **22**, 477–484.

Coffin, D P and Lauenroth, W K (1989) A Gap Dynamics Simulation Model of Succession in a Semiarid Grassland, *Ecol. Model.*, **49**, 229–266.

DeAngelis, D L and Gross, L J (1992) *Individual based Models and Approaches in Ecology: Populations, Communities and Ecosystems*, Chapman and Hall, New York.

Hansen, J, Fung, I, Lacia, A, Lebedeff, S, Rind, D, Ruedy, R, and Russell, G (1988) Global Climate Changes as Forecast by the GISS 3-D Model, *J. Geophys. Res.*, **93**, 9341–9364.

Holling, C S (1964) The Analysis of Complex Population Processes, *Can. Entomol.*, **96**, 335–347.

Huston, M, DeAngelis, D L, and Post, W M (1988) New Computer Models Unify Ecological Theory, *BioScience*, **38**, 682–691.

Leemans, R and Prentice, I C (1987) Description and Simulation of Tree-layer Composition and Size Distributions in a Primeval *Picea-Pinus* Forest, *Vegetatio*, **69**, 147–156.

Malanson, G P (1996) Effects of Dispersal and Mortality on Diversity in a Forest Stand Model, *Ecol. Model.*, **87**, 103–110.

Menaut, J C, Gignoux, J, Prado, C, and Clobert, J (1990) Tree Community Dynamics in a Humid Savanna of the Côte-d'Ivoire: Modeling the Effects of Fire and Competition with Grass and Neighbors, *J. Biogeogr.*, **17**, 471–481.

Mitchell, J B F (1983) The seasonal response of a feneral circulation model to changes in CO_2 and sea temperatures, *Q. J. R. Meterol. Soc.*, **109**, 113–152.

Moore, A D (1989) On the Maximum Growth Equation used in Gap Simulation Models, *Ecol. Model.*, **45**, 63–67.

Newnham, R M (1964) *The Development of a Stand Model for Douglas-Fir*, PhD Thesis, University of British Columbia, Vancouver.

Pacala, S W, Canham, C D, and Silander, Jr, J A (1993) Forest Models Defined by Field Measurements: I, The Design of a Northeastern Forest Simulator, *Can. J. For. Res.*, **23**, 1980–1988.

Pastor, J and Post, W M (1986) Influences of Climate, Soil Moisture, and Succession on Forest Carbon and Nitrogen Cycles, *Biogeochemistry*, **2**, 3–27.

Pastor, J and Post, W M (1988) Response of Northern Forests to CO_2-induced Climate Change, *Nature*, **334**, 55–58.

Rohlf, F J and Davenport, D (1969) Simulation of Simple Models of Animal Behavior with a Digital Computer, *J. Theor. Biol.*, **23**, 400–424.

Shugart, H H (l984) *A Theory of Forest Dynamics: The Ecological Implications of Forest Succession Models*, Springer-Verlag, New York.

Shugart, H H (1998) *Terrestrial Ecosystems in Changing Environments*, Cambridge University Press, Cambridge.

Shugart, H H and West, D C (1980) Forest Succession Models, *BioScience*, **30**, 308–313.

Smith, T M, Halpin, P N, Shugart, H H, and Secrett, C M (1995) Global Forests, in *If Climate Changes: International Impacts of Climate Change*, eds K M Strzepek and J B Smith, Cambridge University Press, Cambridge, 146–179.

Solomon, A M and Webb, III, T (1985) Computer-aided Reconstruction of Late Quaternary Landscape Dynamics, *Ann. Rev. Ecol. System.*, **16**, 63–84.

Forest Stand

The term *forest stand* is used in many contexts, from industrial forestry to the conservation of biological diversity. A forest stand is a group of contiguous trees, which are similar enough in condition, arrangement of age classes, and species composition to be considered a homogeneous unit. The smallest spatial unit for which forest management can feasibly plan cuts is typically a group of stands, which are administered as an integrated whole. Ideally, management uses ecological and silvicultural principles in planning the sequence and the methods by which individual stands are harvested to reach the target wood volume.

Problems can arise during planning, however, if large groups of dissimilar stands are managed as though they each matched the overall average conditions of the management unit. Nonetheless, somewhat arbitrary decisions are always needed about how small a stand can be considered as a separate entity during planning. The number and size of stands distinguished depends on the variation in site conditions, the degree of management intervention, and stand value. Stands as small as one hectare may be distinguished in areas where intensive forestry is conducted; a similar forest might be separated into management units of at least several hundred hectares where crude, extensive forestry is practiced.

Stand structure refers to the distributions of tree sizes and species. It results from the species' inherent growth characteristics as modified by management interventions and environmental conditions during stand development. Two structures are commonly distinguished: *even aged* and *uneven aged*. Trees in even-aged stands are relatively consistent in height and belong to a single age class (which typically extends over 10–20 years). Thus a natural stand which seeds over several years may still be considered even aged. Even-aged stands develop when environmental conditions or disturbances allow trees to become established over a single, brief, reproductive period. In uneven-aged stands, the trees originate almost continuously and therefore vary in size and age class from germinating seedlings to the over-mature.

RICHARD A FLEMING *Canada*

Forest: The FAO De nition

The FAO has an agreed definition of a forest: land of more than 0.5 hectares, with a tree canopy cover of more than 10%, which is not primarily under agricultural or urban land use.

The motivation for an internationally accepted definition is the need for consistency across national borders in forest cover inventories.

Forests are defined both by the presence of trees and the absence of other predominant land uses. The trees should be able to reach a maximum height of 5 m *in situ*. Areas under reforestation which have yet to reach a crown density of 10% or tree height of 5 m are included, as are temporarily unstocked areas, resulting from human intervention or natural causes, that are expected to regenerate (see Figure 1).

The FAO recognizes that this definition is understandably a subject of debate, for at least a couple of reasons:

1. some feel that a threshold of 10% canopy cover is too generous, although the definition does exclude areas where other land uses dominate;

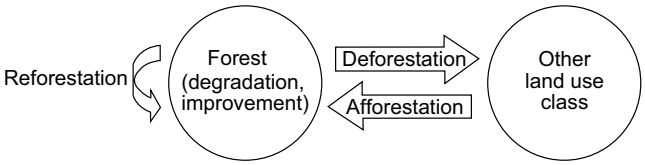


Figure 1 Relationships between forest change terms. Forest degradation and forest improvement occur within forests that continuously stay above the 10% canopy threshold. Reforestation occurs when forests regrow after temporarily having had below 10% canopy cover, but were considered as forests throughout the time. Deforestation and afforestation represent the transfers between forest and other land use classes

2. the classification of particular vegetation types is sometimes difficult. Trees in dry tropical areas may present a particular problem as no distinct boundaries may exist between areas of varying canopy cover.

Further information on the FAO definition may be obtained from robert.davis@ fao.org or peter.holmgren@fao.org.

Other definitions relating to forests include:

Reforestation, deforestation, afforestation: see Figure 1 (FAO, 2000);
Closed forests: forests where 40% or more of tree crowns are interlocked (Singh *et al.*, 2001);
Old growth forest: any forest that has evolved its own complexity under benign conditions over extensive periods of time; *see* **Phase Shifts or Flip-flops in Complex Systems**, Volume 5.

FURTHER READING

FAO (2000) *On Definitions of Forest and Forest Change*, FRA2000, Forest Resources Assessment Programme, Working Paper 33, FAO, Rome.

Singh, A, Shi, H, Zhu, Z, and Forestman, T (2001) An Assessment of the Status of the World's Remaining Closed Forests, see www.na.unep.net/pub/closed forest/.

R E MUNN *Canada*

Fragmentation and Corridors

Richard J Hobbs
Murdoch University, Murdoch, Australia

Land transformation constitutes one of the major global changes, affecting over half the Earth's land surface.

Fragmentation is one component of land transformation, and refers to the process whereby existing land cover types are removed or altered by agriculture, forestry, urbanization or other human land uses. The process of removal leaves patches of the pre-existing cover types, which are of a variety of sizes and are isolated to a greater or lesser extent by a modified matrix. Remaining patches are subject to the impacts of local and landscape-scale changes in environmental characteristics, including altered microclimate and hydrology and invasion by non-native organisms.

Biota within remaining habitat patches are affected not only by these influences but also by the overall reduction in habitat area, isolation of individual patches and reduced ability to move around the landscape. The extent to which the connectivity of the landscape is affected by fragmentation depends on the mobility of the organisms involved, the extent to which the intervening matrix acts as a barrier to movement, and the extent of connection between patches by either corridors or other connecting habitat. Management of fragmented habitats requires consideration of the habitat and movement requirements of species of conservation concern, and current and potential habitat configurations in relation to the likely viability of populations. Management of not only habitat patches but of the overall matrix is required, leading to notions of reserve networks and integrated management of conservation and production lands.

Fragmentation of natural ecosystems has been called one of the most pervasive changes in terrestrial ecosystems across the Earth. It occurs wherever land transformation results in the removal of the pre-existing land cover and its replacement with other cover types, be it urban, agriculture, production forestry or other human-dominated land uses. Such activities may remove most of the pre-existing cover types, transforming landscapes and regions in relatively short periods. Examples include deforestation in tropical areas such as Amazonia, and clearance for agriculture in the mid-west of the United States and in Australia (Hobbs and Saunders, 1993; Laurance and Bierregaard, 1997; Schwartz, 1997). In these and many other cases, the result of such activities is that the pre-existing cover types are left as patches in a modified matrix of production land. These patches may be of varying sizes and may be isolated from other such patches to varying degrees.

IMPACT OF FRAGMENTATION ON BIOTA

Considerations of fragmented landscapes have focussed on factors which influence the likelihood of persistence of the biota retained in the remaining fragments of habitat. Obviously, the major impact of transforming one ecosystem type to another is the dramatic reduction in the amount of habitat left for those species dependent on the pre-existing ecosystem type. The actual area of many ecosystem types is reduced to a few percent of their original extents, and this not only affects their representation in the landscape and in any reserve system that is developed, but also affects the ability of individual species to survive. Species vary greatly in their habitat requirements, and it is likely that in fragmented systems, populations of some species are limited simply by the amount of habitat available.

In addition to the simple amount of habitat remaining, the spatial arrangement of that habitat may also be important. Considerable debate centered for many years on whether fragmented systems would perform better in terms of retaining species if they consisted of a few large fragments versus lots of small patches. The so-called single large or several small (SLOSS) debate stemmed from ideas generated from the theory of island biogeography, which related species numbers on islands to the island size and isolation. There is no real resolution to this debate, since different types of organism will probably require different amounts and distributions of habitat. For instance, large mammalian predators probably require large uninterrupted tracts of habitat, whereas insects and plants can probably survive well in much smaller areas.

Other factors may also limit populations, however. For instance, some species may be limited, not by the areal extent of habitat, but by the availability of a particular resource on which they depend. An example may be nectivorous birds, which depend on a continuous supply of nectar throughout the year. If the fragmentation process removes a particular plant community or species, which provided nectar resources at a critical time of year, nectivorous birds could no longer survive in the area.

CONNECTIVITY AND CORRIDORS

A further factor influencing population viability may be the ability of species to move around the landscape. Species need to move for numerous reasons, such as to fulfil resource requirements, to disperse to new territories, or to migrate. Some species may move only short distances and hence be able to exist entirely within an individual habitat patch, but others may need to move either between different habitat types or across the landscape between different fragments. Others may need to move across entire continents as they follow migration routes.

Population survival in fragmented landscapes has often been discussed in terms of metapopulation dynamics (McCulloch, 1996; Hanski and Gilpin, 1997). The idea of metapopulations proposes that species populations can exist as a series of interlinked sub-populations or metapopulations, which persist in separate habitat patches but may be linked by inter-patch movement. Populations are hence spatially structured into assemblages of breeding populations and movement among the local populations

may have some effect on local dynamics, including the possibility of population re-establishment following local extinction. Hence occupancy of a particular patch by a species may be intermittent and dependent on movement of that species from one patch to another.

In an unfragmented landscape, the ability of species to move depends entirely on the permeability of the different ecosystem types present, i.e., the extent to which different patch types either aided or hindered movement. The permeability of different landscape elements is likely also to vary from species to species. Some species will only move in thickly vegetated cover types, whereas other species favor more open country.

In a fragmented landscape, dense wooded cover is often replaced with low, open crop or pasture. In this situation we might expect that the permeability of the landscape would decrease for species which require wooded cover. The exact impact of the fragmentation depends both on its extent and pattern and also on how particular species view the landscape. Using simple geometric models, it can be shown that the extent to which the various habitat patches in the landscape remain connected declines as the extent of habitat removal increases. However, there is also evidence that the relationship is not linear, and that there is a threshold in connectivity at about 60% habitat removal, at which the level of connectivity declines sharply.

The question of landscape connectivity is a complex one and a frequent component of discussions on the topic revolves around the question of corridors. Using the standard dichotomy of habitat/non-habitat or patch/matrix, it has frequently been espoused that a major way to retain or improve landscape connectivity in fragmented landscapes is by the provision of connections between habitat patches. These connections can be thought of as linear strips of habitat which either act fully as habitat, allowing fauna to live in them, or provide enough habitat characteristics to allow species to move along them and hence move between habitat patches. Many fragmented landscapes do, in fact, retain extensive networks of linear strips, for instance along road verges or hedgerows, and there is evidence that these strips play an important role as habitat for some species and a conduit for movement of others (Hobbs, 1992; Bennett, 1999). Nevertheless, the importance of corridors in conservation networks remains controversial, and it is difficult to obtain conclusive evidence for their role in aiding species movement. Despite this, corridors are a central component of plans for conservation networks, and linear habitat elements often make up a significant proportion of the remaining habitat in highly fragmented landscapes.

So far we have considered fragmented landscapes using a simple dichotomy between habitat and non-habitat, but this model ignores the possibility that the matrix retains some habitat characteristics or is at least still semi-permeable and allows species to move within it. An example might be where scattered trees are retained in an otherwise deforested landscape or where the pre-existing ecosystem type is utilized for grazing rather than actually being cleared. Such landscapes have been termed variegated, and they may function quite differently from what have been classically described as fragmented landscapes (McIntyre and Hobbs, 1999). This emphasizes the conclusion that the degree of fragmentation and the type of management practiced in the matrix may both influence the impacts on the biota. Also of importance will be the pattern of fragmentation, and factors which may be important here will include the number and size of remaining patches, and their spatial distribution (whether they are clumped together or dispersed). Also of importance is the condition of the patches, as discussed next.

EDGE EFFECTS AND LANDSCAPE PROCESSES

In addition to the actual removal of the pre-existing ecosystem type and the creation of the patch/matrix type landscape discussed above, ecosystem modification can also occur, whereby the pre-existing ecosystem type is not removed entirely, but is altered by human use, such as by grazing, timber harvesting and so on. The two processes of fragmentation and modification can occur simultaneously, and often fragments remaining in a modified matrix are also severely impacted by modifying influences (McIntyre and Hobbs, 1999; Hobbs, 1993). Thus, often it is important to consider not only the amount of a particular ecosystem which is left, but also the condition of the remaining patches.

Factors which can influence the condition of patches include both factors operating within individual patches and factors arising from the surrounding landscape. An important feature of fragmented ecosystems is that the pre-existing disturbance regime can no longer operate (Hobbs, 1987). For instance, fire frequency and intensity may increase or decrease in isolated fragments, and this in turn affects the survival and regeneration of the plant communities in the fragments. Also, if movement of faunal species around the landscape becomes restricted, seed dispersal, and hence gene flow, of plant species may be reduced or cease. As species are lost from fragmented systems, so too are any functions that these species performed. For instance, if burrowing mammals disappear from fragmented systems, soil disturbance and turnover are reduced, and hence nutrient cycling and water infiltration to the soil may be affected.

The isolation of fragments in a modified matrix opens the fragments up to a range of microclimatic, hydrological and biotic changes (Saunders *et al.*, 1991). A primary result of this isolation is the creation of an edge between the pre-existing ecosystem and the modified matrix. At this edge, the conditions prevalent in the matrix impinge into the fragment, and a series of edge effects can occur (Murcia, 1995). These effects can include dramatic differences in

radiation input, temperature, frost frequencies, moisture relations, and impact of wind. Such changes can result in considerable mortality of the species up to 100 m from the fragment edge (Laurance *et al.*, 1997), and this impact can be particularly severe in small fragments where virtually the entire fragment becomes edge rather than interior habitat.

Other edge effects can include the input of nutrients or pesticides, either in spray drift from adjacent agricultural operations or in soil blown from adjacent fields. An array of biotic edge effects can also occur; in particular, weed invasion and nest predation may increase greatly at fragment edges.

In addition to the localized edge effects, the surrounding landscape can influence fragments more generally. For instance, significant changes in hydrology can result from the process of replacing one cover type with another. These large-scale changes influence not only the production matrix but also the ecosystems remaining in fragments (George *et al.*, 1995). Also, animals and plants prevalent in the management matrix may exert profound influences on the biota in fragments. In particular, domestic stock can dramatically alter the structure and composition of fragmented systems by grazing and trampling. Similarly, feral predators can be responsible for the decline of faunal species, which would otherwise survive in the fragmented system. Finally, invasive plant species may disperse into fragmented systems from the surrounding matrix, and again these have the potential to significantly alter the structure and dynamics of the fragmented system. Hence fragmentation results not only in the formation of small, isolated fragments, but also in landscape-scale system changes.

REFERENCES

Bennett, A F (1999) *Linkages in the Landscape: The Role of Corridors and Connectivity in Wildlife Conservation*, IUCN, Gland.

George, R J, McFarlane, D J, and Speed, R J (1995) *Nature Conservation 4: The Role of Networks*, eds D A Saunders, J L Craig, and M E Mattiske, Surrey Beatty, Chipping Norton, NSW, 9–22.

Hanski, I A and Gilpin, M E (1997) *Metapopulation Biology: Ecology, Genetics, and Evolution*, Academic Press, New York.

Hobbs, R J (1987) *Nature Conservation: The Role of Remnants of Native Vegetation*, eds D A Saunders, J L Craig, and M E Mattiske, Surrey Beatty, Chipping Norton, NSW, 233–240.

Hobbs, R J (1992) Corridors for Conservation: Solution or Bandwagon? *Trends Ecol. Evol.*, **7,** 389–392.

Hobbs, R J (1993) Effects of Landscape Fragmentation on Ecosystem Processes in the Western Australian Wheat Belt, *Biol. Conserv.*, **64,** 193–201.

Hobbs, R J and Saunders, D A (1993) *Reintegrating Fragmented Landscapes: Towards Sustainable Production and Conservation*, Springer-Verlag, New York.

Laurance, W F and Bierregaard, R O, eds (1997) *Tropical Forest Remnants: Ecology, Conservation and Management of Fragmented Communities*, University of Chicago Press, Chicago, IL.

Laurance, W F, Laurance, S G, Ferreira, L V, Rankin-de Merona, J M, Gascon, C, and Lovejoy, T E (1997) Biomass Collapse in Amazonian Forest Fragments, *Science*, **278**, 1117–1118.

McCulloch, D R (1996) *Metapopulations and Wildlife Conservation*, Island Press, Washington, DC.

McIntyre, S and Hobbs, R J (1999) A Framework for Conceptualizing Human Impacts on Landscapes and its Relevance to Management and Research, *Conserv. Biol.*, **13**, 1282–1292.

Murcia, C (1995) Edge Effects in Fragmented Forests: Implications for Conservation, *Trends Ecol. Evol.*, **10**, 58–62.

Saunders, D A, Hobbs, R J, and Margules, C R (1991) Biological Consequences of Ecosystem Fragmentation: A Review, *Conserv. Biol.*, **5**, 18–32.

Schwartz, M W (1997) *Conservation in Highly Fragmented Landscapes*, Chapman and Hall, New York.

Freshwater Biodiversity

see Biodiversity in Freshwaters (Volume 2)

Freshwater Fisheries

Robin L Welcomme[1] and Henry A Regier[2]

[1]*Imperial College, London, UK*

[2]*University of Toronto, Toronto, Canada*

Water is essential to living organisms. Its various chemical properties have come to be exploited by organisms in the course of evolutionary time, sometimes as opportunities and sometimes as constraints. But water's desirable features can also create difficulties for the organism. Water as atmospheric precipitation carries many substances downward from the atmosphere which are then harmful to organisms immersed in the receiving surface water. Also, the water dissolves, corrodes and erodes many substances in cities and on land and transports these to the streams and lakes either as surface runoff or ground water discharge. Humans drain or spill many kinds of polluted wastes directly into surface waters; currents in these waters then help to disperse the wastes downstream. All freshwater fish in the world are now subject to some adverse effects of the discharge of human-made wastes into the atmosphere, the soil and the water anywhere in the Biosphere. Some waters are so fouled that no fish at all can survive in them.

A FISH IN ITS HABITAT

Molecular water, often complexed with many other compounds, contributes most of the body mass of organisms. In the cells of these organisms the water contains dissolved substances of many kinds which create osmotic pressure within the cells that leads purer water from outside a cell to pass through a cell's membrane into a cell. Generally fresh water, but not sea water, has less dissolved matter in it than does water in a fish's cells, hence fresh water tends to migrate from the outside into a fish. To prevent physiological flooding, a fish in fresh water must try to keep such unwanted water out of its cells and to pump out excess water from them. Such pumping is difficult to do biochemically and requires that the cell and organism do work to accomplish it.

Oxygen is used in the body of creatures to combine with organic substances, that contain stored chemical energy, in order to liberate some of that energy in another form in which it can be used for the work of the cell and of the creature's body. Oxygen is only slightly soluble in water, whether in the cell or in the habitat. Hence organisms generally have evolved some special oxygen carrying devices such as blood corpuscles and complex organs such as gills to arrange for exchange and transport of oxygen from the habitat to the cells where it is needed. Again the process is biochemically difficult and requires work by the fish.

The organism's biochemical processes involved in managing water and oxygen within itself, and many other more complex processes, such as growth and reproduction, are all strongly affected by temperature. For any particular process there is a window of vitality for the various environmental factors. With respect to temperature, this window consists of an interval of temperature below which or above which a particular physiological process does not work at all, or works only in a pathological way. Generally the more complicated a physiological or behavioral process, the narrower is the relevant window with respect to the range in temperature included within it.

Within the window of vitality, most biochemical and physiological rates are strongly dependent on temperature, with an optimum rate often a few degrees below the upper limit of the window. Fish have some behavioral capabilities to select habitat at a temperature near their optimum if such appropriate habitat at that temperature is available, but have practically no physiological capability to control their bodies' temperature. To do so would be very costly of energy in any case, because the water in which fish are immersed has a high heat capacity and would thus rapidly cool any fish that is warmer than its habitat. (Fish have never evolved the capability to create their own insulation from the environment such as sub-cutaneous blubber or external feathers and fur.)

Since the temperature of fish varies with its habitat, which in turn varies daily and seasonally, fish must cope with different muscular, neural and endocrine rates and capabilities at different times of the day and year. Fish have some genetic capability to adapt physiologically and behaviorally to an annual cycle of habitat temperature. Thus both limits of the window of vitality may shift gradually toward a higher temperature during spring and summer, and to a lower temperature in fall and winter. Tasks such as actively rearing young and predatory feeding, for example, may require vigorous and complex activity. Fish tend to schedule such complex and vigorous activities in seasons or daily periods when the habitat temperature is near optimal for the species. A particular family of species tends to have its own general adaptation pattern to these fundamental parameters of fish life, with a finer scale of differences between species.

The concept of a window of vitality can also be used for other environmental factors, such as chemical substances dissolved in water or complexed with food substances. Thus there are lower and upper limits for the concentration in the body of fish for many substances that have been in the fishes' habitat throughout evolutionary time. Such substances include the hydrogen, copper, sodium, chloride and nitrate ions, for example. But some recent man-made chemicals, such as some of the organohalides and polycyclic aromatic hydrocarbons, are new to fish and some are harmful as genetic toxins and disruptive of endocrine functions. For such substances, the concept of a window of vitality appears to have little meaning, since any amount at all may be disabling and lethal to at least some individual embryos and older individuals of a population.

Because of the properties of water chemically and hydrologically, and the fact that fish have no choice but to relate intimately with that water, fish must cope with consequences of many effects of human activity in the upstream watershed and in the atmosphere upwind of the watershed. In effect, such fish are integrative indicators of cultural stress effects, interacting with the natural stress regime in the watershed. Fish have long been used as such indicators in human attempts to control and correct abuses upstream and upwind of their habitats.

DIFFERENT KINDS OF HABITAT - FISH COMPLEXES

All surface waters have flows and currents, both in horizontal and vertical dimensions. The velocities of liquid water are somewhere between those of gaseous water as part of the atmosphere, and those of the solid water of slowly flowing glaciers covering an even more slowly flowing lithosphere at high latitudes and altitudes.

Any currents in open environments have self-organizing capabilities and dynamics that are unpredictable in part.

Some features emerge unpredictably for fish and also for the most expert humans. Fish find opportunities and challenges in liquid habitats to which they must adapt continuously. For a particular species, and more generally for genera or sub-families of species, windows of vitality can be inferred with respect to salient features of water currents. Some fish taxa have evolved to thrive in torrential habitats while others need placid, slowly flowing waters. If humans alter water bodies with respect to such habitat features, then the fish species association changes adaptively in those locales.

Researchers use abstractions with respect to flow regimes, e.g., continuous flow and pulse flow. As an ideal notion, continuous flow relates to currents in rivers and along lake shores that do not fluctuate markedly in volume and velocity over time, as between seasons in a year or within a period of years. With pulse flow, much of the flow occurs over a short period of time, during which the volume and velocity is much higher than at other times. The current regimes of most freshwater surface waters of the biosphere fall somewhere between these two extremes.

Consider that a watershed includes the groundwater-fed headwater lakes and streams as tributaries, the main stem of the river to transport much water, and the branching distributaries that self-organize to spread the river flow into the receiving coastal sea or ocean. The association of fish species within such a watershed includes some that have become adapted through evolution to a continuous current regime, some to a pulse regime, and some to a combination of the two. Further, a particular species may have evolved to exploit pulse features for part of a life cycle and continuous features for another part. Such a complex life history may involve migration of a population from one part of the watershed to another part, to find the appropriate current regime seasonally (*see* **Fisheries: Pollution and Habitat Degradation in Tropical Asian Rivers**, Volume 3).

Currents in lakes are complex and are caused by the action and interaction of many forces. These include the inertial force of inflowing river water, surface forcing by winds, Coriolis forces due to the Earth's rotation, gravity forces acting vertically on waters of different densities, etc.

The shores and bottom of a lake and river act to constrain the direction of a current in a lake, so currents are often strong near shore and in a long-shore direction. The currents of rivers and lakes carve channels and shorelines to create patterns for which empirical generalizations have been inferred. Generally the currents tend to bend in a counter clockwise direction north of the Equator and clockwise south of it. Because of the current regime, lakes are sometimes referred to as one-sided rivers. Near shore in a lake, where currents are strongest, the association of fish and other biota resembles that in the mainstem of the inflowing rivers, in part because of the similarity of current regimes.

Pulses occur with seasonal or occasional flooding due to rapid spring snow melt, to intense rainy seasons such as monsoons, or to massive precipitation events such as those accompanying hurricanes. These pulses tend to have their largest ecological effects on floodplains and coastal zones of downstream rivers, lakes and estuaries with their distributaries. Permanent wetlands in flood plains and coastal zones usually owe their existence to such pulses: the bigger the pulses, the larger the areas of wetlands generally. After a pulse event, waters of wetlands or slowly moving mainstem reaches may become relatively quiescent, with an increase in temperature in summer and a reduction of oxygen concentration due to decomposition of organic materials carried into these waters by the pulse currents.

From the perspective of a watershed's current regime, as some seasonally varying mix of continuous and pulse features, the species of fish in a watershed can be classified, very approximately, into three types: white, gray and black. Again, the colors used to code these types are abstractions. Here white fish tend to be light colored or silvery, partly translucent and reflective of light due to soft and transparent scales. Gray fish often have mottling or colored patches with thicker scales and spines and a relatively opaque appearance. Black fish may have dark leathery skin. There are many anatomical, physiological and behavioral features associated with these color differences.

White fish, as defined here, prefer clear and clean currents of waters with much oxygen and often of cool to cold temperatures. If the current regime in a locale of a watershed fluctuates between continuous and pulse phases, then the whitefish may move longitudinally from locale to locale, including the distributor waters in a downstream sea or ocean.

Gray fish prefer near shore waters where their habitats include large aquatic plants, rocky coasts and reefs, woody debris, etc. Where and when large current pulses occur, gray fish species may move into the flood plain and flooded coastal zone, but then retreat back to the river proper when these flood waters recede.

Black fish prefer shallow waters with low currents, including the relatively stagnant waters of wetlands after flood waters have receded. Because these waters may become anoxic, some black fish species have adapted anatomically by creating primitive lung-like structures, which are filled with air periodically by gulping at the surface. In addition, a species may have adapted physiologically so that it can obtain energy from its food through anaerobic metabolism. In extreme pulse situations, where the water in a wetland dries completely on an annual basis, some fish species adapt by creating moisture-holding cocoons of mud in the bottom, which then dries out.

With each of these three functional classes of fish, there are some species in which individuals become large and

old and may therefore be considered to be K-selected in that they thrive in closely organized natural settings. Individuals of other species remain small, reproduce at a young age and all die within a few years of birth. These are r-selected and thrive where an aquatic ecosystem is markedly unstable either for natural reasons or because of destabilizing activities by humans. At early stages of a pulse event, for example, the r-selected species of all three types may erupt which are then preyed upon by K-selected species of the three types during the less turbulent, waning stages of the pulse event (*see* **r–K Strategies**, Volume 2).

Generalizations like those above should be interpreted heuristically and not as deterministic inferences to apply to individual fish or populations; ecosystem life is much too complex to be described fully in such terms.

DIRECT USE AND INDIRECT ABUSE OF FISH

Many fish species of all three types (white, gray and black) are preferred foods for humans; some larger, older K-selected species are generally, but not always, valued more highly than are younger, smaller r-selected species. Recreational fishers value K-selected white and gray species in particular. The ornamental pet fish trade prefers colorful, r-selected gray fish. Bait fisheries in North America prefer r-selected white fish.

Direct harvesting by commercial and recreational fishers generally selects and thus suppresses K-selected species of all three types, to the advantage of r-selected species. Ornamental and bait fishers complement this effect by harvesting r-selected species where they are abundant because predation by K-selected species has been relaxed. The overall effect on the fish species association is a shift toward small, young fish in which r-selected species can reproduce and thrive but not so with K-selected species.

Humans alter the watery parts of a basin with respect to its current regime, shores, flood plains, toxic burden, oxygen concentration, temperature, etc., all in ways to make the habitat more turbulent, less predictable and more stressful for fish. Overall there is usually a shift in current regime toward the pulse type, combined with a more erratic timing and magnitude of the pulses. If allowed to progress toward extreme modification due to careless human activities, the aquatic environment becomes a disorderly mess. Only the most tolerant of opportunistic r-selected species may be found in such locales, and then perhaps only during seasons when the complex of stresses is least intense.

DIFFERENT KINDS OF FISHERIES AND GOVERNANCE CONVENTIONS

Organized fisheries have existed for millennia, all over the world. Until about five centuries ago fisheries were generally small and local. With the technical innovations that are associated historically with modernization, some individual fishery enterprises have extended their operations from the local to the regional and eventually to the global. But local versions, often using less advanced technology, have persisted with difficulty in many places.

The generalization above relates to each of what are usually termed commercial food fisheries, recreational fisheries as a service industry, the ornamental pet fish trade, and aquaculture enterprises.

Small-scale local fisheries initiatives occur all over the world. Sometimes these provide opportunities as an occupation of last resort, for the poorest of the poor people of a region. The available fish resources are not of sufficient quality or quantity to make it worth an entrepreneur's effort to pre-empt them coercively for external markets.

Water and aquatic habitats are usually treated as property held in common by the people of a locale. Complex governance regimes, with controls on access and harvest for example, are usually in place to share these resources among fishers and the community served by the fishers. A widespread myth, sometimes attributed to Garret Hardin, that common property resources under local governance are inevitably over-harvested is simplistic if not wrong, effective governance has occurred in many places for centuries (*see* **Commons, Tragedy of the**, Volume 5).

Modernized fisheries, serving markets distant from a locale in which the fishing occurs, usually operate under a quite different governance regime than do local artisanal fisheries. Modern fishers often shared implicitly a cornucopeian view that the fish were abundant and resilient to harvesting activities and that the fisher's capabilities to capture preferred fish were generally severely limited. When difficulties arose with respect to collapses of the most preferred species and competitive animosities between individual fishers or fishing fleets, only the least constraining kinds of regulations were imposed. Thus the amount of *fishing effort* permitted by a group of fishermen was set legally. But the effort concept could not be defined rigorously in practice, and fishers generally adapted by making their allocated effort more effective in capturing fish. To make a long story short, the end result of modern fishing practices and modern fishery regulation, is an over-fished fish association, an uneconomic fishery, and a politically unstable fishing community. (Note that this too is a simplistic version, and is a wrong analysis for what happened in some fisheries.)

Late in the 20th century a growing interest emerged in the earlier governance regimes, some of which still persist with respect to artisanal and small-scale local fisheries. Perhaps effective regimes will self-organize to transcend both historical approaches and incorporate good features of each.

CURRENT STATUS, DEGRADATION WORLDWIDE

Human settlements are generally located on large bodies of fresh water, either rivers or lakes. At the beginning of the 21st century, people on all continents continue to migrate to such locales, and especially to coastal zones that contain supplies of fresh water.

For centuries, modernization of fish harvesting, processing and marketing when combined with modernization of economic development on the adjoining terrestrial parts of a watershed has caused the degradation of fish associations and their habitats. The closer a locale is downstream from a large human settlement, the more intense the degradation. Even large coastal seas downstream from large industrialized conurbations have been exhibiting increasingly intense debasement.

Some degraded fresh waters and their fish associations have been a focus for costly efforts toward rehabilitation. The Thames, Rhine and St. Lawrence/Great Lakes watersheds are examples in which some rehabilitative success is now apparent. But with each of these, the improved quality of the fish and their habitat does not imply a return to an earlier pristine state. These aquatic systems have been changed in many ways, and are continuing to be changed permanently in new ways. One of the latter includes the introduction of non-native species, for example, through release of ballast water containing live organisms from ocean-going ships. Thus what kind of fish association to expect with respect to particular species in a particular watershed will remain unpredictable.

Generally, the goals of such rehabilitation appear to include a mix of healthy populations of desirable K-selected species of each of the white, gray and black ecological types. While some watersheds of badly industrialized parts of the world seem to be trending in that direction, many other watersheds in industrializing watersheds in other parts of the world are now degrading and are losing their valued K-selected species of all three types.

Global climate warming scenarios generally forecast much larger temperature increases at higher than at lower latitudes. At mid-latitudes the severity of effects is expected to be spotty, both longitudinally and latitudinally, with some regions more severely affected than others. With fish, both the temperature and the volume of water are important ecologically. In mid-latitude regions of the various continents, decreases in the amount of surface waters and thus of aquatic habitat would result in reductions in the abundance and diversity of fish, with extinctions of local fish populations where the waters dry out seasonally. (In some cases a species may exist only as one population, and thus the whole species would be rendered extinct.) Temperature increases will also occur to various extents regionally at mid-latitudes and more generally at high latitudes. Where aquatic habitats and interconnections will persist, especially as in rivers flowing towards regions of colder temperatures, e.g., down the MacKenzie toward the Arctic Ocean, fish that are genetically programmed for warm water will likely migrate and there displace the species that are limited to cold water habitats. At high latitudes, new cold water habitat may emerge with melting of permafrost and longer open water seasons. Where north–south connections of aquatic habitats will not exist, local extinctions of species can be expected to occur at lower latitudes, with species associations of low diversity at higher latitudes above such habitat discontinuities. Humans may intervene to transport fish species across such discontinuities, as a mitigative measure.

G

Gaia

Gaia: The Earth seen as (or the view of the Earth as) a system that possesses at least some emergent properties typical of living organisms. The defining characteristics of Gaian organization (which could, in principle, be found elsewhere in the solar system or in other planetary systems) are postulated to be.

- Stabilization of the global environment by the biota, analogous to the self-regulatory processes that lead to homeostasis of the internal environment in many animal species.
- profound modification of the chemical composition of the planetary atmosphere away from the composition predicted by equilibrium thermodynamics.

This latter feature provides a potential basis for the detection of Gaian systems by space exploration. According to the Gaia theory, both of these properties arise because of the existence of tightly coupled mutual interactions, both within the collectivity of the global biota (symbioses) and between the global biota and their inorganic environment. Gaia (Mother Earth in Greek mythology) was proposed as a name for the Earth-as-organism by James Lovelock based on a suggestion from Nobel Prize winning novelist William Golding.

See also: Gaia Hypothesis, Volume 5.

G R WILLIAMS *Canada*

GAIM (Global Analysis, Interpretation and Modeling) Program

see GAIM (Global Analysis, Interpretation and Modeling) Program (**Volume 1**)

GCMD (Global Change Master Directory)

see Data Banks: Archiving Ecological Data and Information (**Volume 2**)

GCTE (Global Change and Terrestrial Ecosystems)

see IGBP Core Projects (*Opening essay*, **Volume 2**)

Generic Decomposition and Yield (G'DAY): A Model of Global Change Impacts on Plant and Soil

Ross E McMurtrie

University of New South Wales, Sydney, Australia

The effects of climate change on plant growth are complex, including direct instantaneous effects on numerous plant physiological processes, and indirect plant and soil feedbacks that act on time scales ranging from a few days to centuries. Consequently, predicting the responses of vegetation to environmental change presents a formidable scienti c challenge. One approach is through experiments that manipulate environmental conditions. However, this approach has practical limitations for predicting long-term responses. For long-term predictions the primary scienti c

tools are ecosystem models based on an understanding of the processes involved. One such model is G'DAY (Comins and McMurtrie, 1993).

G'DAY simulates the cycling of carbon, nitrogen and water in plant and soil. The plant sub-model simulates the carbon and nitrogen contents of leaves, fine roots, and wood. The soil sub-model predicts carbon and nitrogen in plant litter and soil organic matter pools (as in the CENTURY model; see *CENTURY Ecosystem Model, Volume 2) and water storage in the rooting zone. Processes represented include plant photosynthesis and respiration, plant water and nitrogen uptake, tissue growth and senescence, litter and soil decomposition, net soil nitrogen mineralization, nitrogen input by atmospheric deposition and biological fixation, and nitrogen loss by leaching and gaseous emission. Photosynthetic rates depend on [CO_2] and temperature, respiration depends on temperature, and decomposition depends on soil temperature and moisture content. G'DAY has been used to investigate effects of altered climate and land use on forest ecosystems in tropical, temperate, Mediterranean and boreal environments and on temperate grasses.*

G'DAY can be operated using either observed daily weather data or constant, mean annual weather. The daily version is used where the purpose is to simulate relatively short-term plant experiments, while the version with mean annual climate is used for investigating longer-term global change impacts.

DAILY VERSION OF G'DAY

G'DAY simulates the cycling of carbon (C), nitrogen (N) and water in plant and soil. The model's general structure is shown in Figure 1. A simulation with daily weather data is shown in Figure 2 for a water-limited stand of *Pinus radiata* growing near Canberra, Australia. The figure illustrates simulations of soil water storage over a four-year period at ambient (350 μmol mol^{-1}) and double ambient (700 μmol mol^{-1}) concentrations of atmospheric carbon dioxide. The simulation at 350 μmol mol^{-1} agrees closely with measured soil moisture (circles in Figure 2). For this water-limited site, G'DAY predicts wetter soils at elevated [CO_2]. This prediction is a consequence of improved water use efficiency at high [CO_2]. Wetter soils enhance plant growth because of an extended growing season and increased soil nitrogen availability.

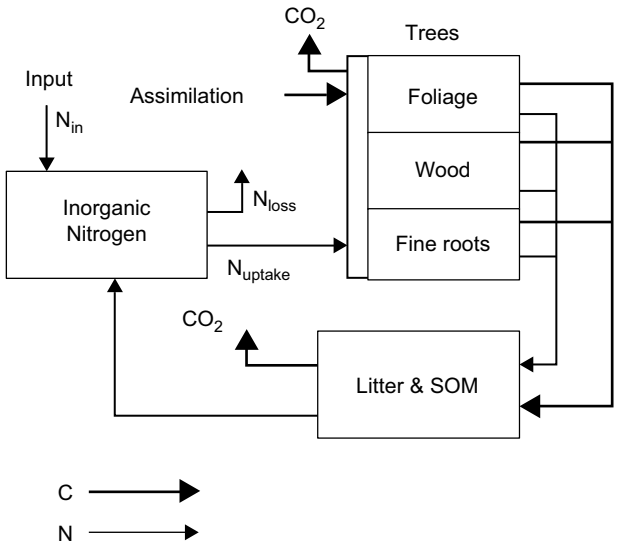


Figure 1 Pools and fluxes of carbon and nitrogen in G'DAY

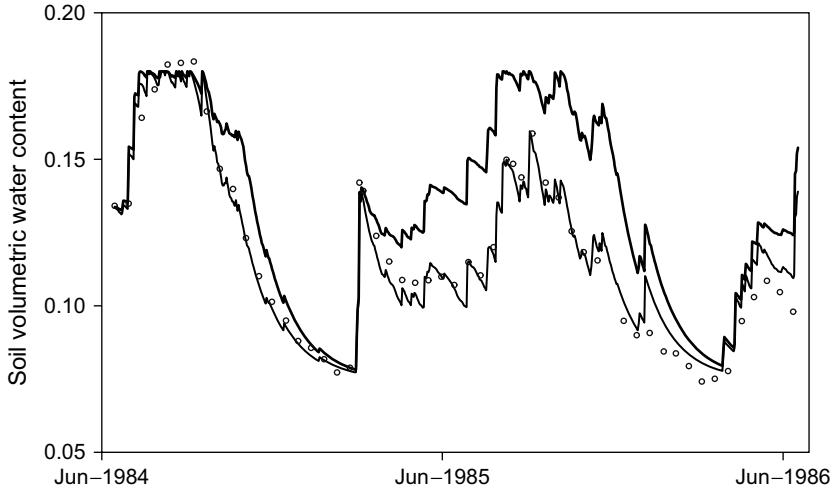


Figure 2 Simulated soil moisture contents to a depth of 2 m for a stand of *Pinus radiata* at ambient (350 μmol mol^{-1}) and double ambient (700 μmol mol^{-1}) [CO_2] concentrations (thin and thick lines, respectively). Measured soil moisture contents at current [CO_2] are also shown (circles)

G'DAY WITH SIMPLE CLIMATE

The daily version of G'DAY is appropriate for applications to short-duration experiments and for investigating short-term response to environmental change. But the daily model has limitations for investigating longer-term impacts. One limitation is that large daily and seasonal fluctuations in model output can make it difficult to ascertain why the model responds in a particular way. In simulations with a simple climate, that task is much easier, as illustrated in Figure 3 where G'DAY is applied to simulate the impact of gradual climate change on a stand of Norway spruce at Flakaliden, northern Sweden. This site is nitrogen-limited, but not water-limited. All simulations were initiated by running G'DAY to equilibrium under the baseline climate at Flakaliden, and then imposing gradual increases in temperature ($+0.03\,^{\circ}\mathrm{C\,year^{-1}}$) and $[CO_2]$ ($+3\,\mu\mathrm{mol\,mol^{-1}\,year^{-1}}$) which, when applied together, approximate the Intergovernmental Panel on Climate Change (IPCC) climate change scenario IS92a (IPCC – WG 1, Vol. 5). Model outputs shown in Figure 3 are net primary production (NPP), and net carbon dioxide exchange by the ecosystem (= NPP minus decomposition of litter and soil organic matter), which is known as the C sink or *net ecosystem production* (NEP). Under increasing $[CO_2]$ NPP increases but tends to decelerate. The carbon sink reaches a peak of approximately $9.3\,\mathrm{g\,C\,m^{-2}\,year^{-1}}$ after 20 years followed by a gradual decline. The carbon sink declines for a variety of reasons, including saturation of the photosynthetic carbon dioxide response, and increased decomposition because litter and soil carbon gradually accumulate. However, according to G'DAY the rate of decline is highly sensitive to parameters describing the flexibility of soil

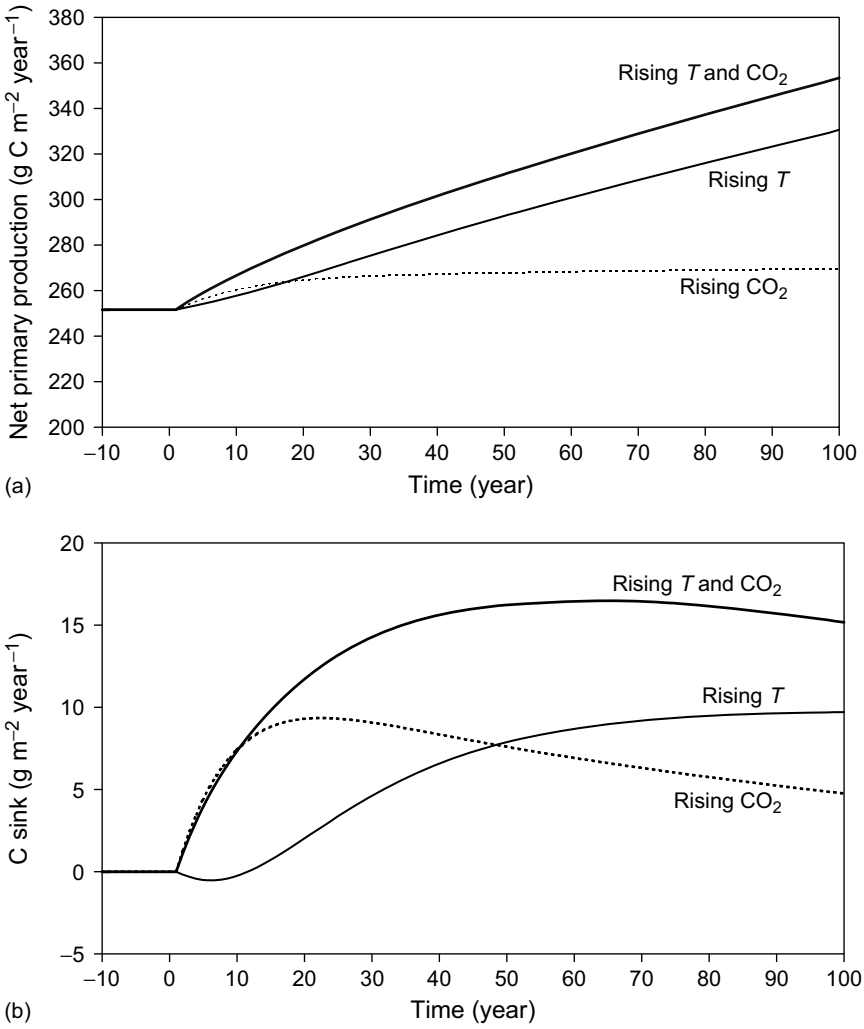


Figure 3 Simulated responses of G'DAY to gradual increases in $[CO_2]$ of $+3\,\mu\mathrm{mol\,mol^{-1}\,year^{-1}}$ alone (dashes), temperature of $+0.03\,^{\circ}\mathrm{C\,year^{-1}}$ alone (thin solid line) and both $[CO_2]$ and temperature (thick solid line). (a) NPP and (b) the ecosystem carbon sink. Each simulation was initiated by running G'DAY to equilibrium at current T

nitrogen: carbon ratios at high [CO_2], and the primary control over the carbon sink is soil nitrogen availability. The carbon sink declines under rising [CO_2] due to nitrogen immobilization in accumulating soil organic matter.

Figure 3 also illustrates simulations of NPP and the carbon sink with temperature increasing gradually at a constant rate of $0.03\,^{\circ}\mathrm{C\,year^{-1}}$. NPP increases steadily throughout the simulation. The carbon sink is negative for an initial transient period, after which it increases steadily to a peak value of approximately 9.7 g carbon $\mathrm{m^{-2}\,year^{-1}}$. The explanation for this temperature response is that warming stimulates soil decomposition causing a release of soil nitrogen, which promotes NPP. Unlike the high carbon dioxide simulation, there is no tendency for the carbon sink to decline over time. The simulation with gradual increases of both temperature and [CO_2] indicates a sustained positive carbon sink over the next 100 years. G'DAY's prediction of a strong and continuing carbon sink under increasing [CO_2] and temperature contrasts with claims that saturation of the carbon dioxide response together with increasing soil decomposition due to rising temperatures will cause a reduction in the sink (e.g., Pearce, 1999). Those claims are based largely on short-term experimental evidence that the temperature response of soil decomposition is steeper than the temperature response of photosynthesis. However, this argument ignores plant–soil feedbacks that are not evident in short-term experimental data (McMurtrie *et al.*, 2001).

MATHEMATICAL ANALYSIS OF G'DAY

Much of the research with G'DAY has been based on a mathematical technique called *two-timing* or quasi-equilibrium analysis. That analysis exploits the fact that ecosystems are made up of plant and soil pools with widely differing turnover times. For instance, leaves and fine roots have faster turnover times than stemwood. Thus, on a specified time scale, the system may be separated into fast and slow turnover pools, where fast pools are in approximate (or quasi-) equilibrium and slow pools are effectively constant. At quasi-equilibrium the total ux of carbon into the fast pools is equal to the total ef ux of carbon (and similarly for nitrogen). The carbon and nitrogen ux equilibria of the fast pools lead to an enormous simplification of G'DAY which gives powerful insight into the model's behavior. The quasi-equilibrium analysis can be represented graphically in terms of two relationships between NPP and leaf nitrogen : carbon ratio: (1) the photosynthetic constraint to production (corresponding to the quasi-equilibrium carbon balance of the fast pools); and (2) the nitrogen availability constraint to production (corresponding to the quasi-equilibrium nitrogen balance of the fast pools). NPP on the specified time scale is then represented by the intersection of the two constraint curves (Comins and McMurtrie, 1993). Responses to rising [CO_2] and temperature can be evaluated from how the intersection shifts on different time scales. Because the analysis leads to explicit equations for the two constraint functions as functions of [CO_2] and temperature, it can explain in graphical fashion how the response to climate change depends on all model parameters. The analysis has been applied to evaluate the response of NPP, the carbon sink, and ecosystem carbon storage, to rising [CO_2] and temperature for both nitrogen- and water-limited ecosystems, and has been used to identify key model parameters in u-encing responses on differing time scales (e.g., McMurtrie and Comins, 1996; Medlyn *et al.*, 2000; McMurtrie *et al.*, 2000).

REFERENCES

Comins, H N and McMurtrie, R E (1993) Long-term Response of Nutrient-limited Forests to CO_2 Enrichment; Equilibrium Behavior of Plant–Soil Models, *Ecol. Appl.*, 3, 666–681.

McMurtrie, R E and Comins, H N (1996) The Temporal Response of Forest Ecosystems to Doubled Atmospheric CO_2 Concentration, *Global Change Biol.*, 2, 49–57.

McMurtrie, R E, Dewar, R C, Medlyn, B E, and Jeffreys, M P (2000) Effects of Elevated CO_2 on Forest Growth and Carbon Storage: a Modeling Analysis of the Consequences of Changes in Litter Quality/Quantity and Root Exudation, *Plant Soil*, 224, 135–152.

McMurtrie, R E, Medlyn, B E, and Dewar, R C (2001) Increased Understanding of Nutrient Immobilization in Soil Organic Matter is Critical for Predicting the Carbon Sink Strength of Forest Ecosystems over the next 100 Years, *Tree Physiol.*, (in press).

Medlyn, B E, McMurtrie, R E, Dewar, R C, and Jeffreys, M P (2000) Soil Processes Dominate the long-term Response of Forest Net Primary Productivity to Increased Temperature and Atmospheric CO_2 Concentration, *Can. J. For. Res./Rev. Can. Rech. For.*, 30, 873–888.

Pearce, F (1999) That Sinking Feeling, *New Scientist*, 164, 20–21.

Genotype

see Phenotype/Genotype (**Volume 2**)

Geographical Information Systems (GIS)

see Geographic(al) Information Systems (GIS) (**Volume 4**)

GIS (Geographical Information Systems)

see Geographic(al) Information Systems (GIS) (**Volume 4**)

Global Ecology

see Biological and Ecological Dimensions of Global Environmental Change (*Introductory essay*, **Volume 2**)

Global Ocean Observing System (GOOS)

Colin Summerhayes
UNESCO, Paris, France

GOOS is designed to provide information to a wide range of users about the present and future states of the ocean and its resources on both short and long time scales. It will help to improve forecasts of weather and climate, and to enable the safe and sustainable management of coastal seas and their living resources.

Safe navigation, the sustainable exploitation of marine resources, the safeguarding of marine environments, and the forecasting of climate for the suppliers of food, water and energy increasingly depend on a greatly enhanced ow of information about seas and oceans and about their roles in climate change. Provision of this operationally useful information about the present state of the sea, and about its future states for as far ahead as possible, for a wide range of users and customers, comes from operational oceanography, comprising systematic and routine observations, forecasts, services and products to meet users' needs.

Because even local patches of ocean are affected by the transport of heat and energy from far away by currents, waves, winds and tides, the predictability of local conditions is improved by the use of regional and global observations and forecasting models. Operational oceanography is being developed at the regional and global scale through GOOS, which is sponsored by three United Nations (UN) agencies including the Intergovernmental Oceanographic Commission (IOC) of the UN Educational, Scientific and Cultural Organisation, the World Meteorological Organisation and the UN Environment Programme along with the International Council for Science. GOOS is not solely operational, but includes work to convert research understanding into operational tools. It will enable nations to convert research results into useful products to meet societal needs.

A good example of GOOS operating at the basin-scale is the Pacific Ocean system for observing El Nino events. It integrates a network of buoys along the equator measuring surface winds and temperature, and subsurface temperature and salinity, sea-level measurements made on Pacific coasts and islands, subsurface temperature measurements made from cargo vessels on trade routes, measurements of current speed and direction made by drifting buoys, and measurements from satellites of the height and temperature of the sea surface. Numerical models use these data to forecast El Nino events for as much as 9–12 months ahead, helping countries to plan to reduce the impact of those events, which recur every four years or so, causing ooding in some places and droughts in other places, as well as damaging the fishing industry.

GOOS implementation will be incremental, following tests of the system through pilot projects, and should be complete by 2010–2020. Much of its success will depend on remote sensing from outer space, key requirements being the continuity of observations from one satellite mission to the next, as well as a sufficient number of similar sensors being own at the same time. These must be matched by a fully global network of data from within the body of the ocean. To provide comprehensive global coverage of upper ocean properties for the first time, GOOS scientists are planning the Argo Pilot Project, which will seed the ocean with 3000 profiling oats to collect profiles of temperature and salinity globally over a long time period. Added to satellite data from the ocean surface, these profiles will underpin models of ocean behavior and of climate.

Ultimately, the development of GOOS will depend on commitments to GOOS plans and funding from national governments.

For further information, see the GOOS website at http://ioc.unesco.org/goos.

Global Soil Erosion Network

see IGBP Core Projects (*Opening essay*, **Volume 2**)

Global Terrestrial Observing System (GTOS)

James R Gosz
University of New Mexico, Albuquerque, NM, USA

GTOS is a dedicated long-term, integrated, user-driven observing system for monitoring the extent, form and function of terrestrial ecosystems. This includes managed lands and the detection and measurement of alterations in them resulting from climate changes, changes in human activities such as land use and industrial development, and changes from other causes such as tectonic activities (http://www.fao.org/gtos). This is accomplished through the development of partnerships between data generators and users that meet both the short-term needs of national governments and the longer-term needs of the global research and policy community.

Within the last decade, with rapid evolution of environmental monitoring technologies and political acceptance of the importance of global environmental problems, a set of Global Observing Systems (Global Ocean Observing System, Global Climate Observing System, and GTOS), integrating *in situ* and remotely sensed data, have been established, each focusing on a major component of the biosphere. Each of these observing systems relies on existing global data centers and networks to provide a global picture of the state of the environment of the planet Earth. The observing systems are sponsored by a number of international organizations – the Food and Agriculture Organization (FAO), the International Council for Science, the United Nations Educational, Scientific and Cultural Organisation, the United Nations Environmental Programme and the World Meteorological Organization.

GTOS addresses the total terrestrial ecosystem complex, including coastal systems, and thus considers biological, chemical and physical properties as well as terrestrial, oceanic, hydrological and atmospheric processes.

GTOS has five key global change issues that are its priority action areas:

- land-use and land-cover change, including land degradation and the sustainability of managed ecosystems;
- water resources management;
- loss of biodiversity;
- climate change;
- pollution and toxicity.

The issues are along thematic lines because of a focus on generating products aimed at specific users and their needs. In general, the issues are concerned with anthropogenic impacts on terrestrial ecosystems. Thus, GTOS is helping to obtain better understanding of some of the basic issues confronting sustainable human development: food and food availability; freshwater supply and demand; changes in terrestrial ecosystems and their life support capacities; changing biological diversity; and exposure of people and other life forms to toxic and hazardous substances.

The emphasis throughout GTOS is on the production of high quality, reliable information and data obtained in a harmonized cost-effective manner. The products are intended to answer specific questions posed by users such as experts in modeling as well as senior policy makers who participate in negotiations related to the implementation of global environmental agreements.

The Global Terrestrial Observing Network seeks to link existing terrestrial observing networks in a hierarchical structure in order to improve data exchange and access to information. These networks are a combination of national networks and thematic ones (e.g., permafrost, hydrology). The Terrestrial Carbon Initiative is a global effort that coordinates a wide array of scientific partners to obtain a better understanding of the pools and uxes of terrestrial carbon that are a central part of the Framework Convention on Climate Change and the Kyoto Protocol.

A number of projects have been initiated that will utilize the information from sites/networks to develop global patterns of plant productivity from satellite imagery, a global terrestrial carbon cycle, and relationships between ecosystem processes and biodiversity. GTOS is working closely with the earth observation and *in situ* scientific communities, both for purposes of validation of regional and global products, as well as the generation of new data and information. The Integrated Global Observing Strategy Partners is an active forum where this coordination takes place.

A metadata base – Terrestrial Ecosystem Monitoring Sites – is available on the internet and allows access to information on the terrestrial observation and monitoring stations (http://www.fao.org/gtos/pages/tems.htm). A secretariat for GTOS is provided by FAO and located in Viale delle Terme di Caracalla, 1-00100 Rome, Italy.

GLOBEC (Global Ocean Ecosystem Dynamics)

see IGBP Core Projects (*Opening essay*, **Volume 2**)

GPP (Gross Primary Productivity)

see Productivity (**Volume 2**)

GTOS (Global Terrestrial Observing System)

see Global Terrestrial Observing System (GTOS) (**Volume 2**)

Gymnosperms

see Angiosperm/Gymnosperm (**Volume 2**)

H

Habitat

Habitat describes the physical and environmental conditions within which a given species or group is able to live (e.g., northern boreal forest, swamp, grassland, etc., although narrower definitions may be appropriate). The habitat of an organism should not be confused with its niche (*see* **Niche**, Volume 2) which describes not only the physical and environmental habitat but also the intimate physical and biological relationships of the species with all other kinds of organisms that also occupy its habitat, and which may limit its range within its habitat.

R G S BIDWELL *Canada*

Habitat Connectivity

see Metapopulations (Volume 2)

Halophyte/Halophobe

Halophytes are plants that can survive in or benefit from an environment with a high level of salt (i.e., sodium chloride), as in saline soils and seawater. *Halophobes* cannot endure such high levels of salt, which has serious deleterious effects on protoplasmal organization. The Greek root suggests that halophytes love salt, but the situation is not so simple. Halophytes are often merely salt-tolerant plants, and would grow better in a low salt environment. However, they do not compete well with halophobic species in low salt conditions, so they can survive and grow only in saline regions where halophobes cannot grow.

While salt (or rather, sodium) is an essential nutrient in very small quantities for most species of plants, some halophytes require and benefit from a high salt environment. Those plants that absorb and store large amounts of salt have a much lower osmotic potential in their cells (i.e., higher osmotic pressure) which enables them to absorb water readily from a saline environment, and also to withstand more extreme cold. Such plants include species that tolerate desert conditions. Other halophytes survive in salty conditions because they have avoidance mechanisms that prevent the uptake of salt or transport it rapidly to glands in the leaves where it is stored out of harm's way or excreted, so that it does not disrupt the physiological functions of the plant.

Halophytic species, either natural or produced by genetic manipulation, are most important in increasing agricultural production by making possible the cultivation of saline soils and the use of salt water for irrigation.

R G S BIDWELL *Canada*

Heterotrophic

see Oligotrophic/Heterotrophic/Eutrophic (Volume 2)

Heterotrophs

see Autotroph/Heterotroph (Volume 2)

Holdridge Life Zone Classi cation

Many ecologists are interested in the effect that climate warming might have on shifting the positions of the world's vegetation zones. One approach is to seek an

empirical relation between the current positions of the main vegetation belts and the current mean values of the main climatic elements (temperature and precipitation) in these belts. A shift in temperature and/or precipitation in a climate-warmed Earth could then be used to estimate the shifts that might take place in the positions of the world's vegetation zones. The search for such relations is assisted by taking account of latitude, elevation above sea level and proximity to bodies of water, although these factors are reflected in the temperature and precipitation values themselves.

A number of climate/vegetation classification schemes have been proposed beginning at the time of Pythagoras in the 6th century BC (Sanderson, 1999) and continuing more recently through Köppen (1918), Holdridge (1947), Penman (1948), to Thornthwaite (1948). One of the more popular approaches with ecologists has been that of Holdridge (1947), perhaps because he was a botanist and he had actually undertaken field studies of the rich variety of vegetation types in a mountainous area (Haiti).

The Holdridge classification is in the form of quite a simple nomogram. The horizontal axis is scaled in millimeters of mean annual precipitation, ranging from less than 125 mm (desert) on the left to more than 8000 mm (rain forest) on the right. The vertical axis is scaled in terms of mean annual temperature, ranging from the tropics to the poles (on the left) and from sea level to alpine and mountain top (on the right). This leads to a differentiation of world vegetation into 100 types. Holdridge also sketched in some tentative lines of equal evaporation but he remarks that "No data are available to support his suggestions regarding evaporation".

The Holdridge nomogram has been used by ecologists (see, for example, Solomon and Leemans (1990) and Yates *et al.* (2000)) to provide a first estimate of how the world's vegetation belts would respond to climate warming simulations as obtained from the atmospheric general circulation models.

See also: **Phenology**, Volume 2.

REFERENCES

Holdridge, L R (1947) Determination of World Plant Formations from Simple Climatic Data, *Science*, **105**, 367–368.

Köppen, W (1918) Klassification der Klimate nach Temperatur, Niederschlag und Jahreslauf, *Petermanns Mieteilungen*, **64**, 193–203, 243–248.

Penman, H L (1948) Natural Evapotranspiration from Open Water, Bare Soils and Grass, *Proc. R. Soc. Ser. A*, **193**, 120–145.

Sanderson, M (1999) The Classification of Climates from Pythagoras to Köppen, *Bull. Am. Meteorol. Soc.*, **80**, 669–673.

Solomon, A M and Leemans, R (1990) Climatic Change and Landscape-Ecological Response: Issues and Analysis, in *Landscape–Ecological Impact of Climatic Change*, eds M W Boer and R de Groot, IOS Press, Amsterdam, 293–316.

Yates, D N, Kittel, T G F, and Cannon, R F (2000) Comparing the Correlative Holdridge Model to Mechanistic Biogeographical Models for Assessing Vegetation Distribution Response to Climatic Change, *Clim. Change*, **44**, 59–87.

R E MUNN *Canada*

Holling, C S

(1930–)

Crawford (Buzz) Holling, ecologist, Canadian citizen, recipient of the Austrian Cross of Honour for Arts and Science, Fellow of the Royal Society of Canada, Director of the International Institute for Applied Systems Analysis (1981–1984), and occupant of the University of Florida's Arthur R. Marshall Jr Chair in Ecological Sciences (now retired), has furthered human understanding of, and approaches to policy development for, global environmental change (Jansson, 1996).

Shortly after his PhD (1957, University of British Columbia (UBC)), Holling (1963) worked with the Canadian Forest Service in Sault Ste. Marie developing Experimental Component Analysis. Through this approach, the variability of a complex population process is explained by dissecting the process into its components and investigating how those components occurring universally interact with the variation in subsidiary components. This approach was particularly effective in relating predator hunting success to prey density and resulted in the Holling functional response equations which are ubiquitously employed in ecological population models involving predation.

Later, while directing UBC's Institute of Resource Ecology, Holling (1973) published his concept of (ecological) resilience, a holistic system property which determines how much disruption a system can experience before losing its integrity. By contrast, stability is a property of an equilibrium condition which determines how quickly the system recovers after a small temporary disturbance. Holling (1987) reasoned that real ecosystems are sustained by renewal cycles encompassing multiple equilibria and that some instability in the form of temporal and spatial variability is needed to maintain the cycles and hence the system's resilience.

Holling cautioned that humans typically reduce variability (and inadvertently, ecosystem resilience) to sustain high yields. With his colleagues (Holling, 1978), he proposed an alternative approach, adaptive management, which uses monitoring and large scale experiments involving abrupt changes in human impacts to learn about, and thus develop policies respectful of, an ecosystem's renewal cycles.

Holling also published a more mechanistic underpinning for resilience (Holling, 1992). He demonstrated that the ecological world is lumpy: body sizes, patch geometries, and ecosystem dynamics occur at discrete time and space scales determined by a few key processes. It is the diversity of functional and species relationships within and across scales that maintain renewal cycles and thus generate ecosystem resilience.

REFERENCES

Holling, C S (1963) Experimental component analysis of population processes, *Mem. Entomol. Soc. Can.*, **32**, 22–32.

Holling, C S (1973) Resilience and stability of ecological systems, *Annu. Rev. Ecol. Syst.*, **4**, 1–23.

Holling, C S (1978) *Adaptive Environmental Assessment and Management*, John Wiley & Sons, London.

Holling, C S (1987) Simplifying the complex: the paradigms of ecological function and structure, *Eur. J. Oper. Res.*, **30**, 139–146.

Holling, C S (1992) Cross-scale morphology, geometry and dynamics of ecosystems, *Ecol. Monogr.*, **62**(4), 447–502.

Jansson, B O (1996) Condensed review of the scientific milestones of Dr C S Holling, Unpublished Mimeograph.

RICHARD A FLEMING *Canada*

Holling Figure-eight

see Succession (Volume 2)

Human Ecology

see Human Disturbance of the Earth System: Dynamics and Complexities (*Introductory essay*, Volume 3)

Humi cation

see Soil Mineralization (Volume 2)

Hutchinson, G Evelyn

(1903–1991)

Evelyn Hutchinson was renowned for his syntheses in the fields of ecology, biogeochemistry, and limnology. He was one of the pioneers in the introduction of quantitative methods in ecology, in an era which still relied primarily on verbal descriptions. Quantification of the regular changes in various ecosystems allowed him to discover the methane cycle. It also permitted a rigorous approach to the long-standing question of why there are so many species. The concept of the *niche* (*see* **Niche**, Volume 2) as a hyper-dimensional space unique to each species is largely associated with his name, and this led to the development of a host of studies on species interactions and distributions.

His major research falls into three distinct categories: studies of the biogeochemistry of atmospheric and aquatic systems; biogeography and systematics of invertebrates, particularly Hemiptera; chemical stratification, metabolism and morphology of lakes; and (in the early part of his life) studies of lakes at high altitudes. The first of these culminated in the creation of a distinct field of science, for which his monumental treatise on biogeochemistry still serves as the principal text, while his systematic and biogeographic work, both directly and through the activities of his students, added whole new families to the zoological nomenclature. His limnologic work was original in its conception of lakes as parts of interacting systems, and continues to excite research to the present day.

What made Hutchinson stand apart from his contemporaries, however, was not merely his own research, but his embodiment of the Renaissance Man. His encyclopedic knowledge and breadth, is nowhere better revealed than in the column, *Marginalia*, which was for many years a mainstay of the prestigious journal, *American Scientist* and in which he anticipated such major findings as the significance of bacterial transformation, or the importance of elements such as aluminum to the biogeochemistry of the earth. His other publications spanned subjects from art to psychiatry, although his single most important work is his treatise on biogeochemistry.

While the measure of a scientist's importance may be thought to be the number of articles published, to academicians what counts more is the number of prominent

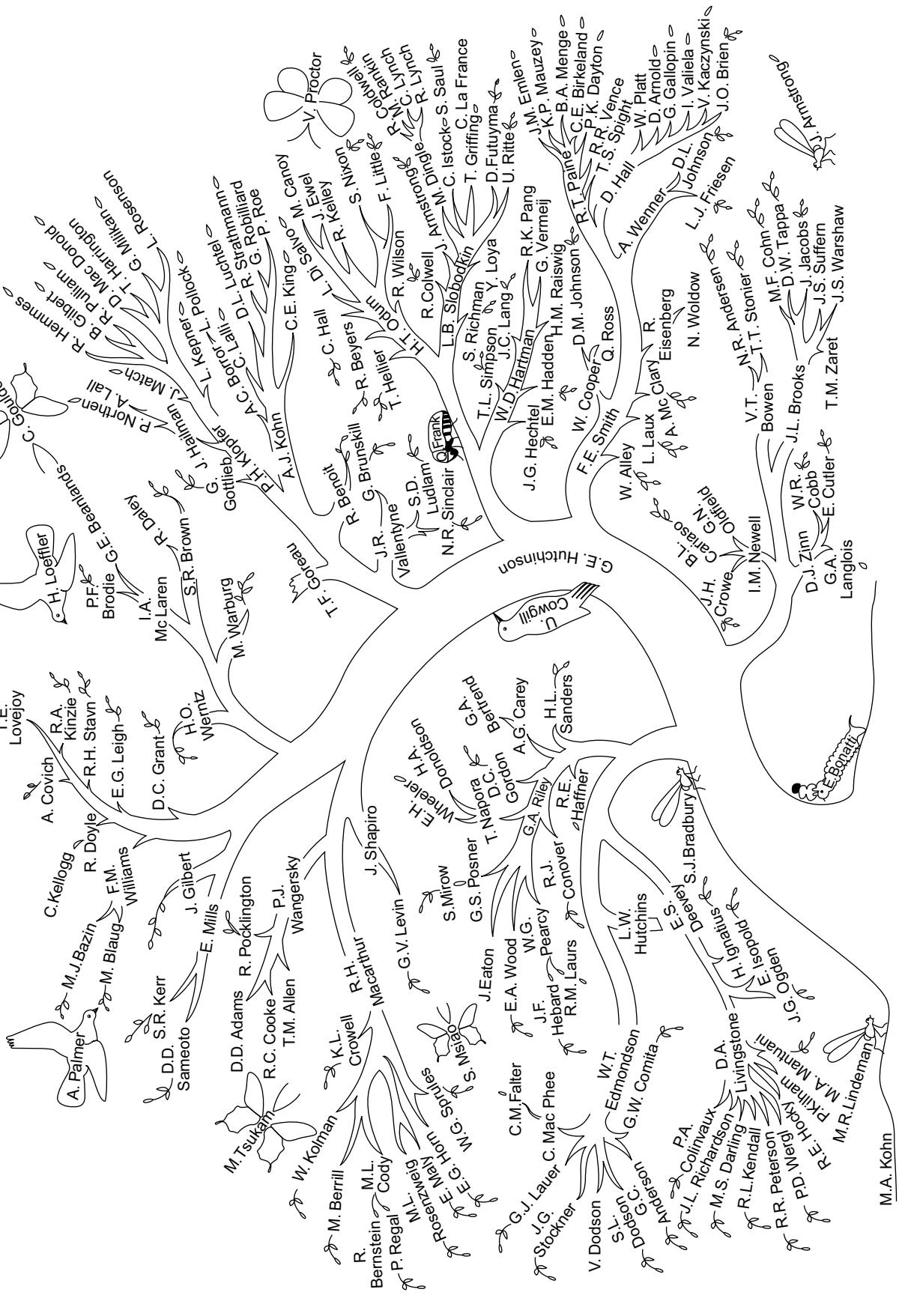


Figure 1 The GEH tree, original drawing by Yvette Edmondson. (Reproduced from Kohn (1971) by permission of the American Society of Limnology and Oceanography)

colleagues who trace their intellectual origins to that single source. In 1971, a celebratory issue of *Limnology and Oceanography* (vol. 16) in honor of Hutchinson appeared, with an analysis of his "phylogeny and biogeography" (Kohn, 197l). Already by that date, it was apparent that the ideas of a majority of contemporary ecologists could be traced to Hutchinson. This trend has continued at an accelerated pace. More significant is the prominence within ecology and behavioral ecology of these offspring: it is a fair estimate that over half of the tenured professors in these fields trace their intellectual origins to Hutchinson (Figure 1). They are responsible for major studies of pleistocene climatic changes in Africa, the history and dynamics of lakes in all parts of the globe, impressive systematic studies of molluscs and arthropods, current theories of biogeography, and a great variety of other studies ranging from animal behavior to zoological history.

Hutchinson's 1979 memoir, *The Kindly Fruits of the Earth*, describes much of his early life, which commenced with his birth in Cambridge, England, in 1903. The family were scientifically inclined, his father a mineralogist at Cambridge University. It was at Emmanuel College, Cambridge, that Hutchinson received his formal training. After a short stint at the zoological station in Naples, he took his first position in South Africa at the University of Witwatersrand. From there, through misadventure he came to the Osborne Zoology Labs at Yale University. Several versions of this tale are in circulation: the one current during my student years was that he blandly cabled Yale that he was available, but whatever the details, he was appointed to the faculty, eventually becoming Sterling Professor. He had no advanced degree at that time, other than a Masters, which Cambridge more or less automatically granted all its graduates after a 2- or 3-year interval on the sound theory that a Cambridge BA would certainly in that time have produced work worthy of an MA. In this case, the assumption was valid.

His first marriage to the renowned zoologist, Grace Evelyn Pickford, was contracted, both claimed, in order that Grace might engage in fieldwork, which was forbidden to unwed women. They were amicably divorced soon after, and remained colleagues and friends. Soon after, in 1933, he married Margaret Seal, and this more romantic relationship lasted until her death half a century later. He married once more, to Anne Twitty Goldsby, who assisted him until her death in 1990. He himself died soon after, in London, 17, May, 1991, aged 88.

REFERENCE

Kohn, A (1971) Phylogeny and Biogeography of Hutchinsonia: G E Hutchinson's Influence Through his Doctoral Students, G Evelyn Hutchinson Celebratory Issue, *Limnol. Oceanogr.*, **16**.

FURTHER READING

Deevey, Jr, E S (1972) Growth by Intussusception. Ecological Essays in Honor of G Evelyn Hutchinson, *Trans. CT Acad. Arts Sci.*, **44**, 1–443.

Edmondson, Y H, ed (1971) G Evelyn Hutchinson Celebratory Issue, *Limnol. Oceanogr.*, **16**, 157–477.

Klopfer, P H (1999) *The Politics and People of Ethnology: Personal Reflections on the Study of Animal Behavior*, Bucknell University Press, Lewisburg, PA.

PETER KLOPFER *USA*

Hydrology

John Roberts
Centre for Ecology and Hydrology, Wallingford, UK

The hydrological sciences deal with the study of the waters of the earth. In its broadest sense, hydrology includes the study of the occurrence, the movement, and the physical and chemical characteristics of water in all its forms within the hydrosphere of the earth. Hydrology is primarily concerned with water on or near the land surface, dealing with that part of the hydrological cycle from the arrival of water at the land surface as precipitation to its eventual loss from the land either by evaporation or transpiration back into the atmosphere or by flow at the surface or subsurface to the sea. There are very strong links between the land surface hydrology and studies of the water in the atmosphere and in the oceans. These are the subject of atmospheric sciences and oceanography, respectively.

The hydrologic cycle is the most fundamental aspect of hydrology. Water evaporates from the oceans and the land surface, is carried over the earth in atmospheric circulation as water vapor, precipitates again as rain or snow, is intercepted by trees and vegetation, provides runoff on the land surface, infiltrates into soils (from where it can be exploited by plants), recharges groundwater, discharges into streams, and ultimately, flows into the oceans, from which it will eventually evaporate again. This immense engine fuelled by solar energy, driven by gravity, proceeds endlessly with or without the activity of humans. Through evaporation and precipitation, water transfers much of the heat energy received by the earth from the tropics to the poles. Annual availability of water on land is the single

most important factor that determines the growth of plants (Kramer, 1982).

GLOBAL HYDROLOGY

The amounts of water in the global hydrologic cycle are so large that it is normal to quantify the compartments and transfers between them in units of km^3 ($1\,km^3 = 10^{12}$ liters $= 10^{15}$ g; see Figure 1). Not surprisingly, the oceans form the dominant component in the global water cycle. Seawater composes 97% of all the water at the surface of the earth. The next biggest store is the water in the polar ice caps and in glaciers. Soils contain 121 800 km^3 of water, of which about 58 100 km^3 is in the root zone of plants (Webb *et al.*, 1993). The poorest estimate of one of the components in the global water cycle is that of the groundwater, but without the influence of human activities, that water would interact little with the biosphere. In relation to other parts of the hydrological cycle, the amount of water in the atmosphere is very small, equivalent to about 30 mm if it were spread evenly over the globe, but enormous quantities of water move through the atmosphere each year (see Table 1).

Evaporation removes about 425 000 km^3 of water from the world's oceans each year. Thus, the mean residence time of ocean water with respect to the atmosphere is about 3100 years. Only about 385 000 $km^3\,year^{-1}$ of the water returns to the oceans as rainfall; the rest contributes to precipitation on land which amounts to 111 000 $km^3\,year^{-1}$. Plant transpiration and evaporation from soils return 71 000 $km^3\,year^{-1}$ to the atmosphere. Thus, with respect to precipitation inputs or evapotranspiration losses, the mean residence time of soil water is about one year. Owing to the excess of precipitation over evapotranspiration on land, about 40 000 $km^3\,year^{-1}$ becomes runoff.

Global averages obscure large regional differences in the water cycle. Evaporation from the oceans is not uniform, but ranges from $4\,mm\,day^{-1}$ in tropical latitudes to $<1\,mm\,day^{-1}$ at the poles (Mitchell, 1983). Although much precipitation falls at tropical latitudes, an excess of evaporation over precipitation in the tropics provides a net regional flux of water vapor to the atmosphere. Net evaporative loss accounts for the high salinity in tropical oceans, and the movement of water vapor in the atmosphere carries latent heat to polar regions (Vonder Haar and Oort, 1973).

On land, the relative balance of precipitation and evaporation differs strongly between regions. In tropical rainforests, precipitation may greatly exceed evapotranspiration. Shuttleworth (1988) calculates that half of the rainfall becomes runoff in the forested region of the Amazon basin. In desert regions, precipitation and evapotranspiration are more or less equal with limited recharge of groundwater.

The sources of water contributing to precipitation also differ greatly in different regions of the Earth. Nearly all the rainfall over the oceans is derived from the oceans. On land in maritime and monsoonal regions, much of the rainfall is also derived from evaporation from the sea. In contrast, up to 50% of the water falling in the Amazon Basin is derived from evaporation within the basin itself, the remainder coming from long-distance atmospheric transport (Salati and Vose, 1984; Eltahir and Bras, 1994).

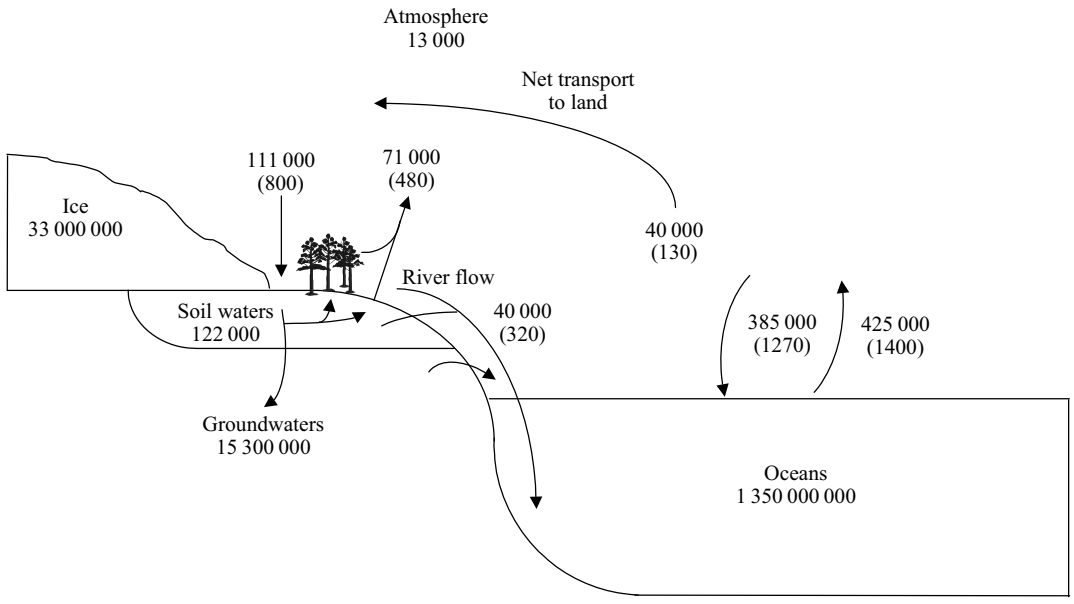


Figure 1 The global water cycle. Values shown are pools (km^3) and fluxes ($km^3\,year^{-1}$). (Reproduced by permission of Academic Press in Schlesinger, 1997)

Table 1 Storage and turnover in the major hydrological stores

Store	Best estimate volume (10^3 km^3)	Area (10^6 km^2)	Per cent total	Per cent freshwater	Turnover rate years
Oceans	1 350 000	360	97.40	0	3000
Ice caps/glaciers	27 500	16	1.98	76.4	8000- 15 000
Terrestrial waters	8477.8	134	0.61		
- groundwater	8200		0.59	22.8	>5000 deep <330 active
- inland seas	105		0.008		
- lakes	100		0.007	0.313	10
- soil moisture	70		0.005	0.219	0.038- 0.96
- rivers	1.7		0.0001	0.005	0.038
- biota	1.1		0.0001	0.003	0.077
Atmosphere	13	510	0.001	0.04	0.027
Man-made stores					
- reservoirs	5		0.0004	0.016	
- irrigated soil	2		0.0001	0.006	
Total global	1 385 990.8				
freshwater	36 000		2.6		

Source: Reproduced by permission of Longmans in Jones (1998).

Most recent estimates assume global river flow is about 40 000 km^3 year^{-1} (Schlesinger, 1997) but the distribution of flow among rivers is highly skewed. The 50 largest rivers carry about 43% of the total river flow; therefore calculations of global transport of dissolved and suspended materials can be estimated from data on a few large rivers. Because of the positions of the continents and their surface topography relative to global climatic patterns, there are large differences regionally in runoff to the sea and associated transport of materials. For example, there is a large contrast between the runoff from North America ($\sim$320 mm year^{-1}) compared to Australia; with a considerable area of dry regions and internal drainage, the annual runoff from Australia is $\sim$40 mm year^{-1} (Tamrazyn, 1989). In the Northern Hemisphere, 77% of the water flow to the oceans comes from rivers that have been regulated by dams or manipulated in some way by human agency. This will have a major impact on the transport of sediment to the oceans. For example, Humborg *et al.* (1997) showed that damming of the River Danube, and subsequent reduction of silicate flow in to the Black Sea, has had an impact on the composition of phytoplankton species in the sea. The shift has been from siliceous diatoms to non-siliceous coccolithophores and flagellates. It is calculated (Postel *et al.*, 1996) that over 50% of the volume of rivers is now used by human activities, particularly irrigated agriculture. Although there is interchange of waters between different oceans, differences in salinity do occur. Despite the considerable input of fresh water from the Amazon basin, the greater evaporation from the Atlantic Ocean than fresh water input means that it is more saline than other oceans.

RIVER BASIN HYDROLOGY

The drainage basin (the term *drainage basin* is synonymous with *catchment* or *watershed* which is commonly used in the US) provides the natural unit for the generation of river flow. As such, it is widely recognized also as the logical unit for water management and for the scientific study of hydrological processes and responses. Within a catchment, runoff is determined by the overall water balance of the catchment:

$$\text{Runoff} = \text{Precipitation} - \text{Evaporation} \pm \Delta\text{Storage}$$

where evaporation includes rainfall intercepted by vegetation canopies and evaporated back into the atmosphere, evaporation from bare soil surfaces and transpiration from plants. ΔStorage is the change in the amount of water stored in soils or rocks. Traditionally, water balances are calculated over 12 month periods but evaluations over shorter periods are equally valuable, but constrained by the resolution of the different components of the water balance especially over short time periods. The water balance may also be calculated for a fraction of the catchment or for a plot.

Precipitation (rain, snow, hail and dew) is an input to the catchment, hillside or plot, but traditional methods only make point measurements. Apart from the problems of acquiring representative values because of the spatial variation in precipitation, particularly in hilly areas, there are technical problems associated with precipitation measurement, which include air turbulence around the rain gauge, evaporation prior to measurement and wetting errors. There are additional problems in estimating snow deposition in

a representative way. A full discussion of the pitfalls associated with rainfall measurement is given by Jones (1997).

In some situations, plant, rock or soil surfaces can collect water directly from the air, bypassing the normal precipitation process and in some cases this may be a very important contribution of water to the soil and subsequently the stream or ground water. There are two processes involved: impaction of cloud or mist droplets on obstructing surfaces, tree crowns are particularly effective in this respect; and radiational cooling of surfaces at night, which might cause condensation of water vapor as dew. The best examples of cloud water deposition are to be found in the cloud forests of tropical mountains where there is substantial evidence that up to nearly half of the stream flow might be a result of trees scavenging water from the clouds.

Evaporation from the land surface involves three main processes: (i) evaporation of rainfall (or snow) intercepted by the vegetation canopies; (ii) transpiration, water taken up from the soil by plants and lost as vapor from pores (the stomata) in their leaves; and (iii) evaporation from the soil surface.

Leaves and branches intercept precipitation before it reaches the ground, and a certain amount can be retained on the canopy from where it is evaporated. This evaporation process, leading to the rainfall interception loss, is most effective in tall, rough canopies, well ventilated by wind, e.g., woodlands and forests. In extreme cases, the rainfall interception losses from forests can reach up to 40% of gross rainfall. An important feature of vegetation predominantly composed of trees and shrubs is that rainfall reaching the ground as throughfall can, in some cases, be preferentially concentrated close to the trunk because of *stemflow*. This is rainfall that is caught on the canopy and directed downwards and inwards towards the trunk by the funneling action of the branches.

Transpiration is controlled by many variables because it depends upon the many dynamic aspects of plant activity. (1) Different plant species can transpire very different amounts depending on the amount of foliage and the stomatal conductance, the sizes and changes of opening of pores in the leaf surfaces through which water vapor passes but which plants can close to control water loss; (2) seasonal changes include variations in radiation, temperature and humidity that influence evaporative demand, as well as phenological changes in the vegetation that control the amount of foliage cover; (3) time of day changes the radiation balance, photosynthesis and growth rates and stomatal activity; (4) meteorological factors such as light levels, wind and humidity deficit can control stomatal opening and influence transpiration; (5) the amount and distribution of water in soil influences its availability for plant roots which affects plant transpiration.

Soil evaporation occurs most readily when the soil surface has been recently wetted by rain and sufficient energy reaches the soil surface to sustain evaporation. Therefore, the highest amounts of soil evaporation will be observed in situations where vegetation cover is absent or very sparse and rainfall events frequently wet the soil surface. Evaporation decreases sharply after rainfall as the surface water content falls, until, with a dry surface, it becomes zero. Without plant roots, the upward movement of water to the soil surface from the wet layers beneath is controlled by capillarity and will, therefore, tend to differ with the texture of the soil. In fine-textured soils, with small pores between the soil particles, capillary movement may only occur over a meter or so but this distance will be much less, $\sim$ a few centimeters, in coarse textured soils.

When water from a rainstorm or melting snow reaches the ground, some or all of it will infiltrate into the soil. The rate of infiltration depends on the intensity of the input, the initial moisture content of the surface layers of the soil and the hydraulic characteristics of the soil. Infiltration will be retarded by compaction of the soil layers by any means and enhanced by the presence of small or large pores such as worm burrows or channels previously occupied by roots. Water in excess of the infiltration capacity of the soil will flow overland as surface runoff once the minor undulations in the soil surface (the depression storage) have been filled. Apart from built and other impermeable surfaces, such runoff occurs most frequently on bare soils and in areas subject to high rainfall intensities. In many environments, rainfall intensities rarely exceed the infiltration capacities of vegetated soil surfaces. The occurrence of surface runoff is then more likely to occur when rain falls on soil that is completely saturated.

The soil provides a major reservoir for water within a catchment. Soil moisture levels increase when there is sufficient rainfall to exceed losses by evaporation or drainage to streams or groundwater. Soils are depleted of water in seasons when evaporation is high and perhaps rainfall is also low or absent. Levels of soil moisture are important for plant and crop growth, soil erosion, and stability of slopes. The rate at which water flows through the soil is dependent on the gradient of hydraulic potential and the physical properties of the soil, expressed in terms of hydraulic conductivity, which varies with soil moisture in a non-linear way.

Runoff is the downward movement of surface water under gravity, in channels ranging in size from small rills to large rivers. Flow in channels can be *ephemeral*, flowing intermittently after periods of rainfall or snow melt (*see also* **Wadi**, Volume 3), while some watercourses are *perennial*, i.e., flowing all the time. Surface waters provide the majority of the water utilized by humans. Stream discharge is normally expressed in units of volume per unit time (e.g.,

cubic meters per second). There are a number of techniques for measuring stream discharge. Measurements of velocities using current meters or ultrasonic devices can be multiplied by the cross-sectional area of the stream channel. Weirs of different types are frequently used to measure discharge. These are constructed so as to give a unique relationship between water level and stream discharge. Water levels can then be measured continuously, usually with a float recorder to construct a record of discharge over time – the *stream hydrograph*. Analysis of the response of the stream hydrograph to rainfall can reveal a great deal about the nature and responses of the catchment and its hydrological functioning.

Many of the components of the water balance can be quantified from measurements made at a wide variety of scales, ranging, for example, from measurements of transpiration from individual plant leaves or changes in moisture content at single points in the soil to measurements made at much larger scales, such as the whole catchment. Details of measurement techniques relevant to different components of the water balance and for different scales will be found in texts such as Maidment (1992) and American Society of Civil Engineers (1996).

Measuring evaporation on a routine basis is not yet a practicable option other than for detailed research studies. For routine operational purposes, e.g., for scheduling irrigation, evaporation is usually estimated by formulae that utilize routinely measured climatic variables. Methods based on only a few variables, e.g., temperature alone are usually highly empirical. These formulae typically make no distinction between transpiration and interception losses from the vegetation. Formulae containing the most realistic physical and physiological descriptions of the evaporation processes require a large amount of information about the environment and the vegetation. Therefore, these formulae are unlikely to be used in practical operational applications, e.g., irrigation scheduling.

Modeling of evaporation losses from vegetation is only one aspect of the physical modeling of the hydrological behavior of a catchment. Comprehensive modeling of all aspects of the hydrological functioning of a catchment of typical complexity is an ideal that is never likely to be achieved. A major problem is to incorporate all the variation due to variations in slope, aspect, vegetation, soils and geology. Not only will it be unlikely that all key descriptions of these elements of the catchment will be available but also that some processes may still be inadequately understood. A description of some of the approaches to the hydrological modeling of rainfall-runoff relationships of catchments is given by Jones (1997). One of the most commonly used approaches is a statistical one, which relates rainfall to runoff, based on data records available over a sufficiently long period. This approach plays a key role in the forecasting of floods, which represent an important form of extreme event in rainfall–runoff relationships.

Extremes in hydrology take many forms, from extremes of magnitude to extremes of duration or both. By definition, an extreme event must be rare and unusual, but there is no sudden break or gap between extreme events and normal ones, and all hydrological events must be viewed as part of a continuous series. The fact that floods and droughts are different from common events by degree rather than of nature is of considerable importance for understanding and predicting their occurrence. Gumbel's (1941) extreme value analysis uses the concept that extreme events can be estimated from observing the magnitude-frequency distribution of more ordinary and not so severe events.

In the past, water planning and management relied on the assumption that future climatic conditions would be well represented by past conditions, and all water supply systems were designed on this basis. The design of dams and reservoirs are based on existing flows in rivers, and the size and frequency of expected floods and droughts. Reservoirs are operated for a variety of purposes using the past hydrologic record to guide decisions. Irrigation systems are designed using historical information on climatic conditions, water availability and soil water requirements. Because of climate warming expected in the next 100 years, however, this reliance on the past record may lead to faulty and possibly dangerous or expensive decisions. Because of those threats, it would be prudent to try both to understand what will be the consequences of climatic change for water resources and to initiate plans to accommodate those changes.

LAND USE AND WATER RESOURCES

Important differences in evaporation from different vegetation covers affect the runoff (stream flow) from catchment areas. Both the transpiration and interception components of evaporation contribute to differences in evaporation. Deep-rooted vegetation, especially forests, can sustain transpiration during long dry periods during which deeper soil water reserves are available to forests but not to shallow-rooted vegetation such as cereals and grass. There will also be greater interception losses from forests because their taller rougher canopies facilitate drying during and after rainstorms. Removal of forests can increase the amount of stream flow, but also there is the prospect that the timing and intensity of peak flows may be altered also. Forest logging often involves road construction and use of heavy machinery that can compact soils stimulating more rapid runoff. There will also be major implications for water quality.

Apart from the now well-documented cases of cloud forests producing rain, the notion that forests might cause

an increase in rainfall needs to be treated cautiously. General circulation models that incorporate substantial land cover changes over considerable areas, e.g., removal of the Amazon forest or major reductions in tree cover in the Sahel, have shown links to rainfall. However, at much smaller scales, the evidence that changes in land cover, such as afforestation, have an effect on rainfall is virtually non-existent. Nevertheless, there is evidence from studies in St Louis, US that the development of extensive urbanization and the heat island effect can be related to increased rainfall (Chagnon *et al.*, 1991).

WATER QUALITY

Water quality is a term covering a wide range of aspects of water and depends on the use to which the water will be put. Common aspects of water quality are physical characteristics such as color, temperature, taste and smell, as well as chemical characteristics such as acidity, hardness and the concentration of various other chemicals such as nitrates, sulfates, phosphates and dissolved oxygen and man-made pollutants including pesticides and herbicides. Deteriorating water quality is a threat to water resources and wildlife in many parts of the world. Much of the problem has been generated by industrialization and urbanization, from atmospheric and water pollution. More recently, agriculture has become a major pollutant source. Attempts to control agricultural sources require basin-wide attention. Particularly, there is concern about the quantities of nitrogen being deposited into the environment as fertilizer and as gases from industry and motor vehicles. Some of the consequences of overfertilization with nitrogen are (Munn *et al.*, 1999):

1. eutrophication of inland lakes and rivers, caused by run-off of nitrogen compounds;
2. development of coastal algal blooms;
3. creation of mineral deficiencies in forest soils caused by leaching;
4. loss of biodiversity in prairie grasslands.

Although some early agricultural activities, such as forest clearance, may have changed water quality, the major sources of pollution in the early industrial era were sewage and chemical pollution especially in urban areas, i.e., point source pollution. However, since World War II there has been a rapid rise in non-point sources of pollution, which enters groundwater and stream/river networks over a broad area.

An issue with the assessment of water quality is the restricted value of simple chemical analyses. In-stream devices are now available that can provide frequent measurements of key parameters of water quality. These might provide valuable information for the purposes of evaluating drinking water quality, for example, but have drawbacks in the evaluation of a suitable habitat for diverse stream biota. Measurements of the chemical properties of stream water provide no guidance towards defining the suitability of a habitat for animals and plants. Biological indicators are being developed to assess the quality of running waters; key indicators are the macroinvertebrates that inhabit streams. Their diversity and numbers in a particular stream are compared with the population in pristine waters in the same region.

See also: **Hydrologic Cycle**, Volume 1; **Lakes and Rivers**, Volume 2; **Water Use: Future Trends, and Environmental and Social Impacts**, Volume 3; **Circulating Freshwater: Crucial Link between Climate, Land, Ecosystems, and Humanity**, Volume 5.

REFERENCES

American Society of Civil Engineers (1996) *Hydrology Handbook*, Manuals and Reports on Engineering Practice Number 28, American Society of Civil Engineers, New York.

Chagnon, S A, Shealy, R T, and Scott, R W (1991) Precipitation Changes in Fall, Winter, and Spring caused by St Louis, *J. Appl. Meteorol.*, **30**, 126–134.

Eltahir, E A B and Bras, R L (1994) Precipitation Recycling in the Amazon Basin, *Quart. J. R. Meteorol. Soc.*, **120**, 861–880.

Gumbel, E J (1941) The Return Period of Flood Flows, *Ann. Math. Stat.*, **12**, 163–190.

Humborg, C, Ittekkot, V, Coclasu, A, and van Bodungen, B (1997) Effect of Danube River Dam on Black Sea Biogeochemistry and Ecosystem Structure, *Nature*, **386**, 385–388.

Jones, J A A (1997) *Global Hydrology*, Longmans, London, 1–399.

Kramer, P J (1982) Water and Plant Productivity of Yield, in *Handbook of Agricultural Productivity*, ed M Recheigl, CRC Press, Boca Raton, FL, 41–47.

Maidment, D R (1992) *Handbook of Hydrology*, McGraw-Hill, New York.

Mitchell, J F B (1983) The Hydrological Cycle as Simulated by an Atmospheric General Circulation Model, in *Variations in the Global Water Budget*, eds A Street-Perrott and M Beran, Reidel, Dordrecht, 429–446.

Munn, R E, Whyte, A, and Timmerman, P (1999) Emerging Environmental Issues: a Global Perspective of SCOPE, *Ambio*, **28**, 464–471.

Postel, S L, Daily, G C, and Ehrlich, P R (1996) Human Appropriation of Renewable Fresh Water, *Science*, **271**, 785–788.

Salati, E and Vose, P B (1984) Amazon Basin: a System in Equilibrium. *Science*, **225**, 129–138.

Schlesinger, W H (1997) *Biogeochemistry: an Analysis of Global Change*. Academic Press, San Diego, CA, 1–588.

Shuttleworth, W J (1988) Evaporation from Amazonian Rainforest, *Proc. R. Soc. Lond.*, **B233**, 321–346.

Tamrazyn, G P (1989) Global Peculiarities and Tendencies in River Discharge and Wash-down of the Suspended

Sediments – the Earth as a Whole, *J. Hydrol.*, **107**, 337–341.

Vonder Haar, T H and Oort, A H (1973) New Estimate of Annual Poleward Energy Transport by Northern Hemisphere Oceans, *J. Phys. Oceanogr.*, **3**, 169–172.

Webb, R S, Rosenzweig, C E, and Levine, E R (1993) Specifying Land Surface Characteristics in General Circulation Models: Soil Profile Data Set and Derived Water Holding Capacities, *Global Biogeochem. Cycles*, **7**, 97–108.

I

IAIGC (Inter American Institute for Global Change)

see Inter American Institute for Global Change (IAIGC) (Volume 2)

IGAC (International Global Atmospheric Chemistry)

see IGAC (International Global Atmospheric Chemistry) (Volume 1)

IGBP Core Projects

see IGBP Core Projects (*Opening essay*, Volume 2)

IGBP (International Geosphere-Biosphere Programme)

This Programme, founded in 1986 under the auspices of the International Council for Science (ICSU), has as its objective:

> to describe and understand the interactive physical, chemical and biological processes that regulate the total Earth system, the unique environment that it provides for life, the changes that are occurring in this system, and the manner in which they are influenced by human actions.

Priority is placed on those areas that deal with key interactions and significant changes on time scales of decades to centuries, that most affect the biosphere, that are most susceptible to human perturbation, and which will most likely lead to a practical predictive capacity.

More than 70 countries have national IGBP committees, which facilitate the national coordination of research projects that are consistent with the IGBP research plan. These bodies are important in garnering national funding both for IGBP-related research and the international coordinating function of the program.

Since its inception, the IGBP has evolved a set of research projects and framework activities on data, modeling and regional research facilities. Research emphasis has been focused on six questions, and from these a set of research projects has been initiated. These include the International Global Atmospheric Chemistry Project (IGAC), Global Change and Terrestrial Ecosystems (GCTE), Land-Use/Land Cover Change (LUCC), Biospheric Aspects of the Hydrological Cycle (BAHC), Land–Ocean Interactions in the Coastal Zone (LOICZ), Joint Ocean Flux Study (JGOFS), Global Ocean Ecosystem Dynamics (GLOBEC), and Past Global Changes (PAGES). Two of these projects are overseen jointly with other international bodies. JGOFS planning and administration is shared with ICSU's Scientific Committee on Oceanic Research. Similarly, LUCC planning and administration is shared with the International Human Dimensions Programme (IHDP) on Global Environmental Change.

The crosscutting framework activities of IGBP include a Data and Information System (DIS), a task force on Global Analysis, Interpretation and Modeling (GAIM), and a Global System for Analysis, Research, and Training (START). This last activity addresses regional research initiatives and needs, and is co-sponsored with the IHDP and the World Climate Research Programme.

A scientific committee governs the planning and implementation of IGBP. Members include the chairs of each of the projects and framework activities and additional members appointed by ICSU. Every third year a scientific advisory council is convened to review the progress of the IGBP with representatives of national committees and various related ICSU bodies.

Since its inception, the Royal Swedish Academy of Sciences has hosted the secretariat for the IGBP, with generous

support from both the Academy and the government of Sweden. Similarly, the coordinating offices for the various projects are hosted by individual nations.

Science and implementation plans for IGBP projects are published in an IGBP Report Series. In addition, the secretariat publishes a quarterly newsletter. While most of the research products of IGBP projects appear regularly in relevant scientific journals, an IGBP Book Series covers advances in synthesis, some of which arise from IGBP scientific symposia.

Further information on the IGBP can be obtained from: IGBP Secretariat, The Royal Swedish Academy of Sciences, Lilla Frescativägen 4, Box 50005, S-104 05 Stockholm Sweden, Telephone (+46-8) 16 64 48, Fax 16 64 05, http//www.igbp.kva.se.

JAMES McCARTHY *USA*

IGBP-DIS (International Geosphere-Biosphere Program Data and Information System)

The role of the IGBP–DIS project is to assist IGBP core projects in the development of their individual data system plans; help provide an overall data system plan for IGBP; carry out activities leading directly to the generation of data sets; ensure the development of effective data management systems and act, where appropriate, to ensure the meeting of the data and information needs of IGBP through international and national organizations and agencies.

The need for a crosscutting, data-focused activity within IGBP was recognized during the programme's early planning phase. The subsequent development and implementation of IGBP–DIS has been achieved through workshops, scientific collaborations, and other activities of the IGBP–DIS standing committee and its office. Most workshops have been held jointly with one or more IGBP core projects, or with the GAIM task force.

The main aim of IGBP–DIS is to improve the supply, management and use of the data and information that are needed to attain IGBP's scientific goals.

Component objectives are to:

- carry out activities directly leading to the generation of relevant data sets and other data system components for IGBP;
- ensure the development of effective data and information management systems for IGBP, including the preparation and implementation of a programme-wide data system plan, and assisting IGBP core projects in the development of their individual data system plans;
- assist in meeting the data and information needs of IGBP through partnerships and collaborations with other organizations and agencies.

A suite of activities and tasks has been defined to achieve these objectives. An IGBP-wide data system plan is currently being developed that will deal with data management and exchange policies; partnerships and collaborations; and interoperability and standards.

Further information may be obtained from IGBP–DIS Office, CNRM – 42 Av G Coriolis, 31057 Toulouse cedex, France, Tel.: +(33) 5 61 07 85 81, Fax: +(33) 5 61 07 85 89.

See also: **IGBP Core Projects**, Volume 2.

JOHN S PERRY *USA*

IGBP Terrestrial Transects

Will Steffen

IGBP Secretariat Royal Swedish Academy of Sciences, Stockholm, Sweden

Global change research has prompted the development of new scientific research tools and the modification of existing ones. Ecologists have long used gradients of various types and scales to understand the influence of underlying environment factors in key system processes and characteristics. Global change research has adapted this approach to address some key questions about the interaction of the Earth's land surface with global change phenomena. The result has been the formation of a set of continental scale transect studies, the International Geosphere–Biosphere Program (IGBP) Terrestrial Transects (Koch, 1995b). The original aim of the transects was to determine how underlying 'global change gradients'–changes in temperature, precipitation and land use – affect terrestrial biogeochemical cycling, and what will be the consequent feedbacks to the Earth system. The original concept has been expanding beyond biogeochemistry to including other important global processes, such as the hydrological cycle, especially water and energy exchange between the land surface and the atmosphere.

DESCRIPTION

The IGBP Terrestrial Transects are best conceptualized when a linear or quasi-linear variation in an underlying environmental variable provides the organizing principle for the study. For example, the moisture gradient in northern Australia from the coast to the interior decreases virtually linearly for 1500 km, providing an ideal gradient on which to base a transect study. The temperature gradient from north to south in the Arctic-boreal region provides another such gradient. In addition to these biophysical gradients, some of the transects are organized around land use change 'intensity' as a gradient. These transects are more difficult to visualize in physical space, as they consist of a network of sites scattered through a biome (e.g., humid tropical forests) and organized into a gradient in an analytical sense rather than a physical one.

In scale, the transects are normally of order 1000–2000 km in length and a few hundred kilometres wide. This is important to facilitate the scale transitions discussed below. Also, the transects should be long enough to encompass a major ecotone between biomes – e.g., the tundra-boreal forest ecotone or the temperate forest – grassland ecotone – as well as the biomes on either side.

The generic transect design is shown in Figure 1. The backbone of the transect infrastructure would be a small number (normally 3–5) of field research stations where intensive observational or experimental work would be carried out. For example, eddy correlation observations of water, energy and carbon exchange would be established at such stations. In addition, a number of less intensive sampling sites would be established to explore sub-gradients, such as soil type or land use change intensity (for transects where a biophysical gradient is the predominant one) or to explore various stages in the successional dynamics of the ecosystem.

RATIONALE

Global Change Gradients

The fundamental rationale for the IGBP Terrestrial Transects is that key ecosystem properties and processes are controlled by environmental factors, which will change

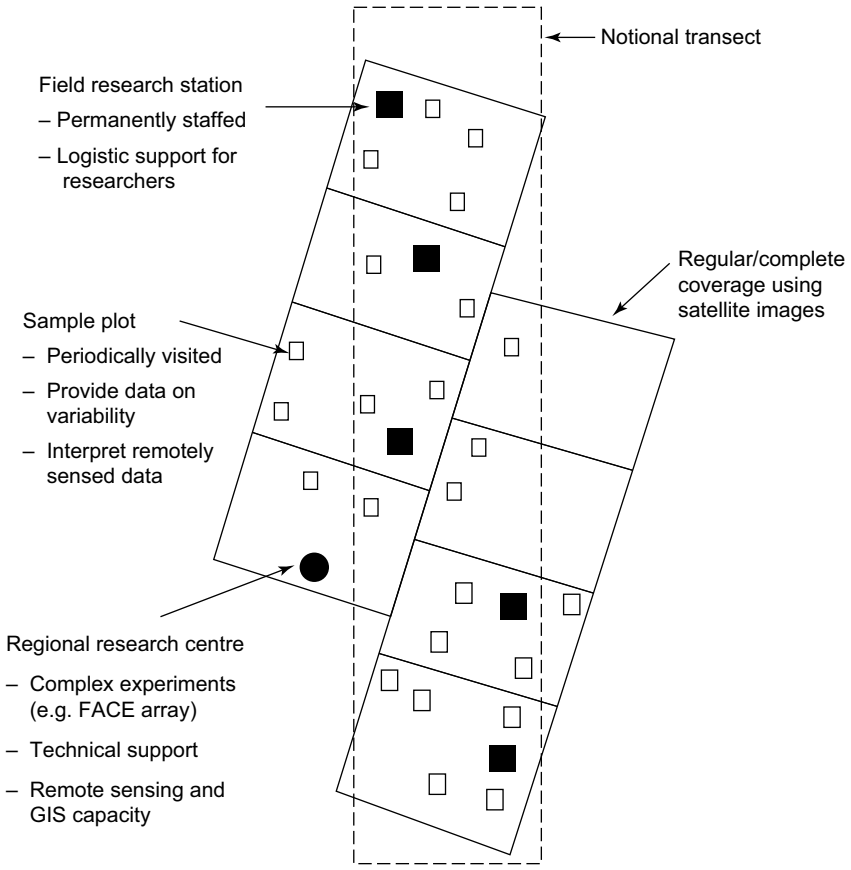


Figure 1 Idealized design of an IGBP Terrestrial Transect. The notional transect is about 1000 km long and a few hundred km wide (Koch *et al.*, 1995a, The IGBP Terrestrial Transects: Science Plan. IGBP Report No. 36. IGBP Secretariat: Stockholm, Sweden, 61)

under global change, and understanding how ecosystems might respond to such changes is best studied using a gradient approach. Examples of key ecosystem properties which are correlated to environmental gradients are: (1) leaf-area index and above – ground biomass related to precipitation in semi-arid savannas; (2) maximum rooting depth and soil moisture in grasslands. Examples of gradient-driven processes include (3) net primary production and temperature in the high latitudes; (4) flow of water through a catchment (gravity-driven); (5) movement of cattle and people along a moisture gradient in West Africa (Menaut *et al.*, 1995).

Thresholds along a Continuum

Transects are probably the only feasible approach, apart from modeling, that can identify thresholds along a continuum. Simple networks cannot do this. The threshold concept is particularly important for global change research, in which an increasingly important paradigm is that of a stress on a system reaching a critical threshold, causing a change of state of the system, rather than a simple cause-and-effect (or dose-response) approach. Examples of thresholds in global change ecological research include (a) an abrupt change in leaf area index along the Northern Australia Tropical Transect (NATT) in response to a smooth gradient in precipitation, and (b) an abrupt change from nutrient – rich to nutrient – poor tree leaves on savanna trees in response to a smooth precipitation gradient on the Kalahari Transect in southern Africa (Steffen *et al.*, 1999).

Facilitation of Interdisciplinary Research

Although not one of the original rationales for the establishment of the IGBP Terrestrial Transects, the enhancement of collaboration across disciplines has grown in importance. For example, the emission of trace gases from soils (atmospheric chemistry community), the exchange of water and energy between the land surface, both soil and vegetation, and the atmosphere (climatologists and hydrologists), the uptake and metabolism of carbon dioxide (CO_2), water and soil nutrients (ecosystem physiologists), and changes to the structure and composition of ecosystems (ecosystem dynamics) are all interlocking and closely related processes. In addition, human activities, especially land use and land cover change, are increasingly affecting these basic ecological processes. Encouraging the co-location of these studies at the same sites along the same gradient is a good way to facilitate the difficult process of building truly interdisciplinary research.

Scale Transitions

Global change research is often confronted with the challenge of scaling up results and understanding from process studies at small scales up to the global. This is particularly true for terrestrial ecosystems, where much process – level work is done on small, homogeneous patches of an ecosystem. The key question is: when can such process–level understanding be scaled up directly to the globe, and when must processes which emerge at intervening scales (e.g., the landscape scale) be taken into account in the extrapolation process. The transects are designed so that the transition from patch to landscape to region can be made taking appropriate account of scale-emergent properties and processes. In this regard, remote sensing plays a key role in the scale-up process, and all of the transects employ remote sensing techniques to a greater or lesser extent in the scale-up process.

DESCRIPTION OF INDIVIDUAL TRANSECTS

The locations of the current suite of IGBP Terrestrial Transects are shown in Figure 2. More details of the transects can be found in Steffen *et al.* (1999) and Walker *et al.* (1999) and references therein.

High Latitude

The high latitude transects are based on a north–south temperature gradient and encompass the tundra and boreal forest biomes and the transition between them. The long-term carbon dynamics of these systems is a crucial research and policy question.

Alaska

This interdisciplinary study builds on the work of the United States Arctic System Science Study, and is focused largely on two long-term ecological research stations north of the Brooks Range. The work includes studies on ecosystem physiology, hydrology, trace gas emissions and biodiversity. A Geographic Information System (GIS) database for the region is well established and many types of modeling tools are employed.

Canada

The Boreal–Forest Transect Case Study (BFTCS) extends along a southwest to northeast axis from the wheat belt of the Canadian prairies through extensive stands of boreal forest to a small area of tundra along Hudson Bay. Work in BFTCS concentrates on the long-term changes to the structure and carbon dynamics of the forests due to logging and natural disturbances – fire and insect outbreaks. Work along the BFTCS and across Canada as a whole has demonstrated the important influence of disturbance in long-term carbon dynamics (*see* **Boreal Ecosystem–Atmosphere Study (BOREAS)**, Volume 2) (Kurz *et al.*, 1995).

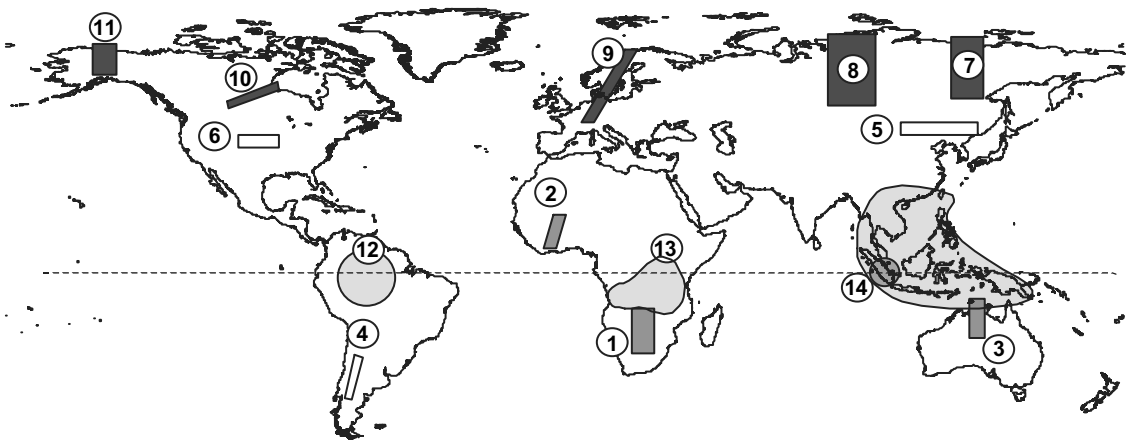


Figure 2 Locations of the existing IGBP Terrestrial Transects and those in the advanced planning phase (from Steffen *et al.*, 1999)

Siberia

Two transects have been established as part of the broader IGBP Northern Eurasia Study (Steffen and Shvidenko, 1996). The Central Siberian transect basically follows the Yenesei River from the Arctic Ocean south to near Krasnoyarsk. The river forms a boundary between the low-lying forests and wetlands to the west, dominated by pine and spruce species, and the higher, colder, drier forests to the east, dominated by larch. Considerable work, ranging from ongoing flux measurement stations to tree-ring studies, has already been established along the transect. The Far East Siberian Transect follows the Lena River basin from Tiksi in the north through Yakutsk and then extends south to the Khabarovsk region. Larch is the dominant tree species of the extensive forests along much of this transect, and the interplay of fire and permafrost in controlling the dynamics of these forests is a key research question.

Scandinavia

A recent initiative called DART (Dynamics of the Arctic Treeline), supported by the European Union (EU), could form the backbone of a Scandinavian high latitude transect. The present research design is focused on ecosystem dynamics but physiological studies co-located at the DART sites could enhance the value of this developing transect (B Huntley, personal communication). There has also been some planning under the SCANTRAN (Scandinavian Transect) initiative (OW Heal, personal communication).

Mid-latitude

These transects cover the temperate forest and grassland biomes and are heavily dominated by land use change. Two of the world's major agricultural areas – the USA grain belt in the middle of the continent and the north China agricultural belt – are encompassed in these transects. The underlying biophysical gradient is moisture.

NECT (North East China Transect)

This transect runs along an east–west axis from the temperate conifer – broadleaf mixed forests along the east coast, through meadow steppes and agricultural land in the middle, to steppes and desert grasslands in the west. The transect has recently been extended into Mongolia. The major issue along the NECT is land use change. Food demand is growing strongly in north China, and a combination of intensification and extensification of agriculture is likely. Previous extension of agriculture has led to large-scale land degradation in some areas of the transect. Avoiding such problems in future is a high priority.

USA Prairies

As for NECT, the USA Prairies transect is an east–west gradient dominated by land use change. In fact, the temperate forests which potentially occur at the eastern end of the transect (the states of Illinois and Indiana) have been almost completely cleared and converted to agriculture. A

large amount of research has already been carried out on this transect in various studies, but a co-ordinated effort is required to assess the relevant data and adapt it to the transect approach.

Argentina

This transect follows the moisture gradient through the western areas of Argentina. An interesting ecological question is why the entire transect is covered by grassland, whereas precipitation is sufficient on the northern end of the transect to support forest. Much of the work to date on the transect has focused on soil structure and dynamics, including estimates of rooting depth.

Savannas

Like the mid-latitude transects, these are based on moisture gradients. However, the evaporative demand is much higher in these areas, and the ecosystems are dominated by savannas, grading into dry or humid tropical forests at one end and into desert at the other. The contrast among the three transects described below in terms of the importance of land use change is striking, and offers an excellent opportunity to use the suite of transects in a comparative way to quantify this important biogeochemistry factor.

NATT (Northern Australia Tropical Transect)

The NATT extends from the dry tropical forests around Darwin (1500 mm annual precipitation) to the arid woodlands and desert near Alice Springs (300 mm annual precipitation). Much of the region is lightly grazed, population density is low, and tourism is a growing industry. Much of the vegetation is close to its potential structure, and its dynamics is strongly influenced by disturbance regime (fire) and grazing pressure. Currently much of the work along NATT is oriented towards ecosystem dynamics, but the transect is also ideally placed to support research on trace gas emissions and water and energy exchange.

SALT (Savannas in the Long-term)

A companion transect to NATT, SALT is organized along a West African gradient extending north from the edge of the humid coastal forest for 2000 km into the semi-desert on the boundary between the Sahel and the Sahara Desert. These West African savannas have been largely modified by intense human pressure, which is likely to increase even more. SALT is a well established transect which has produced a large amount of understanding of the human-natural linked dynamics that operate in this region (e.g., Valentin, 1996). One of the early land-surface experiments – HAPEX (hydrologic atmospheric pilot experiment) – SAHEL – was located at a site along the SALT transect.

The Kalahari Transect

The Kalahari Transect is midway between NATT and SALT in terms of the dominance of human pressure on the structure and functioning of the ecosystems along the transect. In addition, the transect is largely located on the Kalahari sandy soils, and thus largely eliminates soil type as another environmental variable. Current work on this transect is focused both on the ecosystem interaction with the strong biophysical gradient and on the growing likelihood of significant human influence from population growth and economic change. The current dominant land use is cattle grazing, but this is under pressure from conversion to more intensive agriculture (i.e., cultivation) in some places and broadening to multi-species systems in others.

Humid Tropics

In these regions the dominant global change driver is land use change, and the transects are ‘conceptual’ ones, in that the sites are not organized along a line in physical space but rather along a conceptual gradient of land use change intensity (see Steffen *et al.*, 1999).

LBA (Large-scale Biosphere- Atmosphere experiment in Amazonia)

This is perhaps the largest of the regional global change studies yet undertaken, and aims to determine how Amazonia as a region functions as part of the Earth system and how human modification of Amazonia might affect that functioning (LBA Science Planning Group, 1996). LBA is organized along two sets of gradients, one a biophysical moisture gradient from humid tropical forests through dry tropical forests and into savanna; and the other a conceptual land use change intensity gradient, as described above. The study has been carefully planned around an integrated set of climate, hydrological, biogeochemical, ecological and land use change components (*see* **Large Scale Biosphere–Atmosphere Experiment in Amazonia (LBA)**, Volume 2).

Southeast Asia

The land use change intensity/biogeochemistry work is part of a larger study, the Southeast Asian Integrated Study (Lebel and Steffen, 1998), which has a strong focus on the human dimensions of global environmental change. Southeast Asia provides an interesting contrast to Amazonia. Whereas the latter is a large continental land mass in which land use change can significantly affect the regional climate, Southeast Asia is an archipelago of islands and peninsulas whose climate is dominated by the surrounding oceans, and the Asian monsoon. In addition, land use change patterns and driving forces contrast sharply between the two regions.

Central Africa

This transect is in the very early discussion and development phase. There is an existing regional global change study in central/southern Africa, the Miombo Woodlands Network, which could also be adapted to a transect approach based on a land use change intensity gradient. The focus of the Miombo work is on dry tropical forests.

Altitudinal Gradients

A series of altitudinal gradient studies to be undertaken as part of a mountains research initiative associated with IGBP are complementary to the set of transects described. Insofar as these altitudinal gradient studies are based on temperature gradients, they will provide an interesting comparison to the high-latitude transects.

THE TRANSECTS IN ACTION

Although much remains to be done to reach the full potential of the transects as a global change research tool, there have already been some notable successes in terms of application of the transect approach to global change problems. For example, the concept of a land use intensity gradient has been successfully applied to a study of deforestation in Jambi Province, Sumatra, Indonesia (M van Noordwijk, personal communication). In that work sites have been stratified according to their degree of agricultural management, in effect, land use change intensity. Interestingly, the degree of ecological disturbance is strongly correlated with this gradient, and the loss of biodiversity is inversely related to the gradient. In additional to these biophysical parameters, a number of socio-economic variables also correlate with the land use intensity gradient when plotted against total net output of the economic system in US$/ha. These include the 'actors', who benefit from the economic activity (e.g., small – hold farmers or government supported trans-migrants); the current development trajectory with time; and the possible new variations of development trajectories.

A second example of the gradient approach, with more direct management applications, comes from the SALT transect in West Africa. Along much of that transect the vegetation grows in a banded pattern, with strips of woody vegetation interspersed with barren areas. This pattern is formed by a complex interaction of rainfall, runoff and evapotranspiration. The pattern can only be understood by considering two opposing gradients, over a long spatial scale. Early attempts to manage this system, based on local understanding only, were spectacular failures and led to significant degradation of the system as a whole. It was only when the transect approach was adopted, and the complex dynamics of the banded vegetation pattern understood, that the management regime was modified to maintain the productivity of the woody strips (Valentin, 1996).

The most comprehensive attempt to use the IGBP Terrestrial Transects as a global set was the regional – scale analysis of global change impacts undertaken as part of the GCTE (Global Change and Terrestrial Ecosystems) synthesis (Walker *et al.*, 1999). In that study two Dynamic Global Vegetation Models (DGVMs) were applied to two transects in each set-high latitude, mid-latitude, and semi-arid tropical savannas. In addition, a global vegetation model that incorporates projections of land use change was also applied to the transects in a 100 year run into the future. Furthermore, 'expert knowledge' was obtained from the transect co-ordinators; this latter approach was also applied to two of the humid tropical land use change intensity transects. The results were fascinating in that much of the 'expert knowledge' did not agree with the projections of the global models. There could be many reasons for this, two of which are that (1) the global models might be missing key processes, and (2) the 'expert knowledge' is simplistic and is missing critical interactions and feedbacks that can be simulated in a mathematical modeling approach. An urgent challenge is therefore to develop a suite of finer scaled models of ecosystem dynamics and ecosystem processes for each transect which can then be compared more directly with the global models.

FUTURE CHALLENGES

The transects described above all have regional objectives and priorities as their primary focus. This has strengths in that it enhances the prospects for national and regional funding and promotes the application of results from global change research to more immediate management goals. From the perspective of global change research, however, the growing emphasis on the regional perspective is presenting a considerable challenge to using the transects as a co-ordinated international set to answer global questions. To reach the full potential of the IGBP Terrestrial Transects as a global change research tool, the following key activities have been identified (Steffen *et al.*, 1999):

- Creation of consistent basic GIS datasets for each transect, incorporating information such as soil type and characteristics, vegetation composition, climate (present and future scenarios) and land cover and use.
- Agreement on common experimental protocols and observational regimes to facilitate truly comparative research across the transects.
- Intercomparison of experimental results and model simulations, and periodic synthesis and integration activities.

For further information please refer to the IGBP website www.igbp.kva.se.

REFERENCES

Koch, G W, Scholes, R J, Steffen, W L, Vitousek, P M, and Walker, B H (1995a) The IGBP Terrestrial Transects: Science Plan. IGBP Report No. 36, IGBP Secretariat, Stockholm, Sweden, 61.

Koch, G W, Vitousek, P M, Steffen, W L, and Walker, B H (1995b) Terrestrial Transects for Global Change Research, *Vegitatio*, **121**, 53–65.

Kurz, W A, Apps, M J, Beukema, S J, and Lekstrum, T (1995) 20th century Carbon Budget of Canadian Forests, *Tellus*, **47B**, 170–177.

LBA Science Planning Group (1996) The Large Scale Biosphere–Atmosphere Experiment in Amazonia (LBA). Concise Experimental Plan. LBA Co ordinating Office, Centro de Previsão de Tempo e Estudos Climáticos, Instituto Nacional de Pespquisas Espaciais, Cachoeira Paulista, SP, Brazil.

Lebel, L and Steffen, W, eds (1998) Global Environmental Change and Sustainable Development: Science Plan for a SARCS Integrated Study. Southeast Asian Regional Committee for START (SARCS), Bangkok, Thailand, 139.

Menaut, J C, Flouzat, G, and Valentin, C (1995) Modélisation, cartographie et télédétection. Programme SALT (CNRS, ORSTOM, CNES). In Ecosystèmes & Changements Globaux, Les dossiers de l'Environnement de l'INRA, No. 8, 261–279.

Steffen, W L and Shirenko, A Z, eds (1996) IGBP Northern Eurasia Study: Prospectus for an Integrated Global Change Research Project, IGBP Report No. 37, The International Geosphere–Biosphere Programme, Stockholm, 1–95.

Steffen, W L, Scholes, R J, Valentin, C, Zhang, X, Menaut, J C, and Schulze, E D (1999) The IGBP Terrestrial Transects, *The Terrestrial Biosphere and Global Change. Implications for Natural and Managed Ecosystems*, eds B Walker, W L Steffen, J Canadell, and J Ingram, IGBP Book Series No. 4, Cambridge University Press, Cambridge, 66–87.

Valentin, C (1996) Soil Erosion Under Global Change, *Global Change and Terrestrial Ecosystems*, eds B H Walker and W L Steffen, IGBP Book Series No. 2, Cambridge University Press, Cambridge, 317–338.

Walker, B H, Steffen, W L, and Langridge, J (1999) Interactive and Integrated Effects of Global Change on Terrestrial Ecosystems, *The Terrestrial Biosphere and Global Change. Implications for Natural and Managed Ecosystems*, eds B H Walker, W L Steffen, J Canadel, and J Ingram, IGBP Book Series No. 4, Cambridge University Press, Cambridge, 329–375.

Impacts of Climate Change on Natural Systems

see Natural Systems: Impacts of Climate Change (*Opening essay*, **Volume 2**)

Impacts of Global Change on Animals

see Animals: Impacts of Global Environmental Change (*Opening essay*, **Volume 2**)

Impacts of Global Change on Terrestrial and Freshwater Ecosystems

see Terrestrial and Freshwater Ecosystems: Impacts of Global Change (*Opening essay*, **Volume 2**)

Impacts of Global Environmental Change on Plants: from Cells to Ecosystems

see Plants - from Cells to Ecosystems: Impacts of Global Environmental Change (*Opening essay*, **Volume 2**)

Infectious Diseases

Paul R Epstein
Harvard Medical School, Boston, MA, USA

Climate constrains the range of infectious diseases, while weather affects the timing and intensity of them. Global warming has helped extend the geographic range of several infections (Epstein et al.*, 1998) and extreme weather events (EWEs) spawn clusters of diseases (Epstein, 1999). Synergistic impacts of warming and increased climate variability have also produced several surprises. While economic inequities and ecological disruption are key determinants of disease susceptibility and transmission, an unstable climate is playing an ever-increasing role in driving the global emergence, resurgence and redistribution of infectious disease (Patz* et al.*, 1996; McMichael* et al.*, 1996).*

Since 1976, over 30 diseases have emerged that are new to medicine (World Health Organization, 1996). Of equal concern is the resurgence of old diseases once thought under control. Some infections – like tuberculosis – are transmitted person to person, and their resurgence primarily reflects

social disruption and poor public health infrastructures. Infectious diseases involving two or more species – mosquitoes, ticks, deer, rodents and humans – however, also respond to environmental factors.

HISTORICAL BACKGROUND

Pandemics – epidemics spanning multiple continents – have often been associated with social transitions and have stimulated social change. Declining living conditions and overtaxed infrastructures set the stage for two plague pandemics: the Plague of Justinian of 541 AD (apparently *Yersinia pestis*, by clinical description), and the Black Death (also bubonic plague), emerging from the overcrowded urban conditions of the Middle Ages (1346 AD). These two pandemics, in turn, altered settlement and labor patterns for the centuries after.

Another upsurge of epidemic disease, more akin to today's, had a more positive outcome. In the early 1800s, growth of national wealth in Europe, and decline in infectious disease, was matched by deterioration in the quality of life. By the 1830s, water and sanitation, workhouse and factory conditions reached critical thresholds, the Dickensian conditions proving ideal breeding grounds for cholera, smallpox and tuberculosis.

This time underlying conditions and growing contagion sparked protests and constructive responses. In England, the Sanitary and Environmental Reform Movements were born (1842–1854), and epidemiology (beginning with disease mapping) ushered in modern public health principles and later a national health program. The epidemics abated over several decades, before the discoveries of Pasteur and Koch, and three-quarters of a century before the advent of anti-microbials.

In the mid-1970s, a similar threshold was reached. Today, the exhaustion of finite resources, generation of wastes, and climatic change compound increasing poverty. Together social and environmental instability are creating conditions vulnerable to the global emergence, resurgence and redistribution of infectious disease.

Infections involve a disease agent (a pathogen), host(s) and the environment, and some require vectors (or carriers). Climate can influence pathogen and vector populations, alter host defenses, change the migration patterns and food sources, and can radically alter habitat.

Mosquitoes, in particular, are highly sensitive to climate, being hot weather insects with fixed thresholds for survival. The relationship between mosquito-borne diseases and climate was described as early as 1920 (Gill, 1920) and the dependency on temperature quantified in the 1950s (McArthur, 1972). *Anopheline* spp. mosquitoes can transmit falciparum malaria parasites, for example, when temperatures remain above 16 °C (Martens *et al.*, 1997; Patz *et al.*, 1996), while the mosquitoes that transmit dengue fever, *Aedes aegypti*, are restricted by the 10 °C winter isotherm, as freezing kills the eggs, larvae and adults (Martens *et al.*, 1997). Within their survivable range, warmth accelerates biting and parasite and viral maturation within mosquitoes. Malarial protozoa, for example, mature in 26 days at 20 °C and just 13 days at 25 °C (McArthur, 1972). *Anopheline* mosquitoes live but several weeks, thus warmer temperatures accelerate development allowing for two crucial blood meals: one from an infected person and a second for transmission.

Warm nights and warm winters, i.e., high minimum temperatures (TMINs) are particularly favorable to the insect activity. Fossils from the end of the last Ice Age (about 10 000 years ago) demonstrate that changes in TMINs were correlated with geographic shifts of beetles. Insects are thus superb paleothermometers, shifting latitude and altitude in response to temperature regime shifts ahead of grasses, shrubs, trees and large animals.

Natural Regulatory Systems

Antibodies that stun or kill pathogen invaders and phagocytes that consume them are part of the arsenal of animal defense systems. Natural systems have also coevolved functional groups like predators and prey, competitors and recyclers, that help regulate the populations of nuisance organisms. Wide swings in weather and other disturbances can disarm bodily defenses, and several features of global environmental change may harm predators disproportionately, releasing prey from natural biological controls. Among these are habitat loss and monocultures, excessive chemical use (harming friendly insects, birds and amphibians), increased ultraviolet radiation, and changes in seasonal climate.

CLIMATE AND INFECTIOUS DISEASES

Data on global warming are well established (Mann *et al.*, 1998), and several physical aspects of climate change are especially pertinent to the biological responses. Global warming is not uniform; warming is occurring disproportionately at high latitudes, high elevations, and during the nighttime and the winter (IPCC, 1996). Since 1950, springs begin almost a week earlier, on average, and fall a week later, and TMINs over land areas have risen twice as fast as overall warming (Easterling *et al.*, 1997). Less precipitation falls as winter snow and many biological systems on land and at sea are responding to the combined changes. Growing zones are shifting, but not as rapidly as the pests that prey upon nutritional staples – corn, soybean, wheat – and cotton.

Extreme events (droughts, floods, storms, severe winds) are intensifying (Karl *et al.*, 1995). A warmer atmosphere

also holds more vapor (6% more for each 1 °C warming), enhancing evaporation and the hydrological cycle. As the ocean warms, sea ice melts and atmospheric water vapor rises. Heat accumulating in the ocean (Levitus *et al.*, 2000) thus increases clouds, humidity and heat indices, contributes to warm nights and tropical-like downpours, fuels hurricanes, and alters ocean and air currents that influence weather patterns globally.

VECTOR-BORNE DISEASES

Hot, humid weather and absent relief at nighttime, directly challenge human (and livestock) health. These conditions also promote mosquito-borne diseases.

One-half of the world's population is exposed to malaria on a daily basis. In the 1990s, worldwide incidence of malaria increased dramatically. While poverty, deforestation and drug resistance all contributed, a warmer climate provided conditions conducive to increased transmission in temperate zones and highland regions. Models based on climate parameters and transmission dynamics project that the warming accompanying doubling of atmospheric carbon dioxide (CO_2) will increase vectoral capacity 100-fold in temperate zones, and extend the area capable of sustaining transmission from that containing 45% of the world's population of 60% (Martens *et al.*, 1997). A statistical model projects less of a change (Rogers and Randolph, 2000). But all the analyses to date are based on projected changes in average temperatures, rather than the more rapid changes in minimum temperatures being observed; and thus may underestimate the biological responses.

Evidence suggests the projected changes are underway. *Anopheline* mosquitoes are present in the US, and, earlier this century, malaria was prevalent, and control went along with advances in social development. In the 1990s, however, small outbreaks of locally transmitted malaria (not surrounding airports or imported) (see Zucker, 1996) have occurred during hot spells in Texas in 1994, Michigan in 1995, Georgia in 1996, Florida, New Jersey in 1995 (Zucker, 1996), New York and Toronto (see Epstein *et al.*, 1998), and in California, where transmission occurred in the 1980s. Malaria has extended its latitudinal range in South Africa, while dengue fever has spread into Northern Australia and into Argentina. Swedish research has demonstrated that warm winters facilitate northern migration of ticks carrying Lyme disease (Lindgren and Gustafson, 2001) – the most important vector-borne diseases in the US.

These changes are consistent with projections, but high altitude changes are most telling because of the emerging global pattern and the consistency with data from other fields.

Changes at High Elevations

Insects and insect-borne diseases (including malaria and dengue fever) are today being reported at high elevations in Africa, Asia, and Latin America (see Epstein *et al.*, 1998). Malaria is now prevalent in Central African highlands, affecting high elevation urban centers like Nairobi, and rural highland areas of Papua New Guinea. Once limited by temperature to 1000 m, *Aedes aegypti* (vectors of dengue and yellow fevers) are found at 2200 m in the Colombian Andes and India. Displacement of plants to higher elevations has also been documented in the European Alps, and reported in Alaska, the US Sierra Nevada and in New Zealand. (A plant shift of 500 m in elevation is equivalent to a 300 km shift in latitude (Peters, 1991).)

These insect and botanical trends, indicative of gradual, systematic warming, have been accompanied by another widespread physical phenomenon: Montane glaciers are in retreat in Argentina, Peru, Alaska, Iceland, Norway, the Swiss Alps, Kenya, the Himalayas, Indonesia, and New Zealand. Some may soon disappear.

Polar researchers suspect that melting at the base of the Greenland ice sheet may be sculpting fault lines that could diminish its stability (Krabill *et al.*, 1999). Shrinking of Earth's ice cover (cryosphere) has implications for water (agriculture, hydropower, and for albedo [reflectivity]), thus climate stability.

Finally, since 1970, the level at which freezing occurs has climbed about 160 m higher, corresponding to a 1 °C warming (Diaz and Graham, 1996). Thus, mountain regions – where biological, glacial and isotherm shifts are especially apparent – provide sentinel sites for monitoring the consequences of climate change.

EXTREME WEATHER EVENTS (EWEs) AND EPIDEMICS

A significant trend linked to heat buildup in the atmosphere and oceans is the alteration of EWEs (Intergovernmental Panel on Climate Change, 2001). Shifts in climatic norms alter the frequency and intensity of individual events, as well as their temporal and spatial patterns (Karl *et al.*, 1995). Storms and other extreme events are exacting a severe toll on public health and public health infrastructures worldwide (Epstein, 1999).

Since the 1970s, EWEs have increased in the continental US and elsewhere. Periods of drought have become longer as warming dries soils, and bursts of tropical-like precipitation (>5 cm over 24 hours) occur more frequently (Karl *et al.*, 1995).

Extreme disturbances favor nuisance organisms. Prolonged droughts, that fuel fires, haze and respiratory ailments, also support population explosions of insects and rodents, when punctuated by heavy rains. Floods foster

fungal growth, can create insect breeding sites, and flush microorganisms, chemicals and nutrients into waterways, triggering harmful algal blooms.

Intense rains and flash flooding, plus animal husbandry practices, can breach sanitation and water systems, and precipitate water-borne disease outbreaks of *Cryptosporidium parvum* protozoa and others.

The relationships between El Niño/southern oscillation (ENSO) events and disease outbreaks have been well studied (Bouma *et al.*, 1997). Analyses of multiple time series demonstrate associations between these climatic events, the accompanying anomalous weather, and outbreaks of vector-, water- and rodent- borne disease. Other climate signals – the North Atlantic Oscillation and regional sea surface temperatures (as with Rift Valley fever in West Africa) – are also drivers of conditions conducive to disease outbreaks.

The 1990s: A Decade of Extremes

In the mid-1970s the pattern of ENSO events quickened from the 4.2-year periodicity since 1877 (Trenberth, 1999). In the 1990s, the pattern changed again, each year of the decade experiencing an El Niño or a La Niña event. Build up of heat in the ocean may be altering the natural ENSO mode (Timmermann *et al.*, 1999), carrying profound implications for disease patterns. The largest El Niño of the past century – 1997/1998 – and the La Niña of 1998/2000 – spawned weather anomalies and associated disease clusters across the globe (Epstein, 1999).

EWEs have a profound impact on public health, particularly in developing countries, and the aftershocks can ripple through economies (International Federation of Red Cross and Red Crescent Societies, 1999). Hurricane Mitch, for example, nourished by a warmed Caribbean, hit Central America in November 1998, killing over 11 000 people and causing damages exceeding $5 billion. The intense precipitation and flooding associated with the hurricane spawned a cluster of disease outbreaks, including cholera (30 000 cases), and malaria and dengue fever (30 000 and 1000 cases, respectively) (Epstein, 1999).

In the summer of 1999, the drought affecting US Atlantic coastal states brought several health-related surprises to New York State. On July 6th, the power demands of the heat wave caused a blackout on New York City's upper west side, and, when backup diesel generators at two Columbia Medical School buildings failed, years of biomedical research were lost. The event affected the medical, scientific and financial world, highlighting the need for off the grid energy sources, like fuel cells, that are nearly 100% reliable.

At summer's end (August 26th), torrential rains drove farm waste into an upstate underground aquifer supplying the Washington County Fair. Over 1000 people fell ill from toxic *Escherichia coli*, some with hemolytic uremic syndrome and there were several deaths.

The third surprise for New York in 1999 was the appearance of the mosquito-borne West Nile Virus (WNV) (discussed later).

EWEs (winds, storms, droughts and fires) and sea level rise are taking an increasing toll on habitat itself, with implications for species diversity and disease. Habitat modification from abrupt climate change is considered the leading cause of four of the five mass extinctions of the past 440 million years. In Honduras, drought-sustained wildfires consumed 11 000 km^2 of forest during the summer preceding Mitch, widening the ongoing deforestation and magnifying the flooding and devastation (Epstein, 1999).

Sequential Extremes, Synergies and Surprises

In 1992, the US Institute of Medicine, in the first major report published on emerging infectious diseases, predicted that conditions in the US were ripe for the emergence of a new disease. What they did not foresee was that climate was to play a significant role in the emergence, amplification and transmission of two of this decade's surprises: the hantavirus pulmonary syndrome (HPS) in the Southwest and the WNV in New York City. Sudden weather changes and sequential extremes can decouple ecological rhythms and yield biological surprises.

HPS

The diversity of rodent predators provides insurance when some decline. The prolonged California and US southwest drought from 1987 to 1992 reduced raptors, coyotes and snakes. And when drought yielded to intense rains in 1993 (the year of the Mississippi flood), grasshoppers and piñon nuts on which rodents feed flourished. The resulting 10-fold increase in mice populations by June 1993, led to the emergence of a new disease: HPS (Epstein, 1995). The virus may have been present, but dormant. Altered predation and habitat associated with changing meteorological conditions provided the conditions allowing its amplification and transmission. By summer's end predators returned and the outbreak abated, demonstrating the resilience ecosystems retain. Subsequent episodes in the US have been limited with the aid of early warnings, but HPS has appeared in several Latin American nations, with about 50% mortality and several cases of person to person transmission.

WNV

While the introduction of the WNV into the US remains a mystery (through migrating birds, humans or airline-transported mosquitoes), conditions favoring diseases that

cycle among birds, urban mosquitoes and humans – like St. Louis encephalitis (SLE) and WNV – are well known. They are warm winters and hot, dry summers, conditions associated with climate change in the Northeast US The first recorded outbreak of SLE in the US occurred around St. Louis in 1933 during the dust bowl and the hottest summer to date, for the century. Most urban SLE outbreaks from 1933 to 1974 were preceded by multimonth droughts (unpublished data).

The precise mechanisms are matters for conjecture. Mild winters and dry summers favor breeding of city-dwelling mosquitoes (*Culex pipiens*), while their predators decline with drought. Birds may have congregated about shrinking water sites during the July 1999 heat wave, favoring viral amplification among birds and mosquitoes; the heat accelerating viral maturation. Finally, torrential rains in late August unleashed a crop of *Aedes* spp. mosquitoes that may have acted as an additional bridge vector to humans (Epstein, 2000).

Against this is the background of generalists – crows, Canada Geese and gulls – out-competing specialist avians, the specialists losing ecological niches and refuges. Generalists, with wide ranging diets, may, in general, be more tolerant of disease agents, thus more apt to maintain blood levels that favor disease transmission. In this case, WNV may have recently evolved in strength, as it took an unusual toll on birds. The importation of WNV highlights the potential spread of agents evolving in association with social and environmental change anywhere on the globe.

COSTS

The economic impacts of diseases of humans, livestock, and food crops can be far-reaching (Daszak *et al.*, 2000). White flies, whose numbers swell when there is drought, are injecting at least 18 types of Geminiviruses into tomato, squash, and bean plants throughout Latin America, threatening regional food security (Anderson and Morales, 1993).

The 1991 cholera epidemic cost Peru over $1 billion in seafood exports and tourist income, while international airline and hotel industries lost from $2 to $5 billion from the 1994 Indian plague, and cruise boats avoid islands affected by dengue fever, coral diseases and harmful algal blooms (Harvell *et al.*, 1999).

Prolonged anomalous conditions of the sort that applied in 1990–1995 can have cumulative biological consequences. In New Orleans, for example, five years without a killing frost fostered an explosion of mosquitoes, cockroaches and termites in summer 1995. Termites persisted inside trees during subsequent cold winters and threaten stands of New Orleans' mighty oaks.

If more frequent and intense EWEs continue to be a primary manifestation of climate change, we have begun to experience the true social and economic costs of an unstable climate. Viewed as a symptom, the resurgence of infectious disease across a wide taxonomic range suggests that we may be vastly underestimating the true costs of business-as-usual.

The International Federation of the Red Cross and Red Crescent Societies (1999) reports that developing nations are experiencing 96% of the impacts of climate change. In the last several years, however, Hurricane Floyd in North Carolina, fires in Florida and California, WNV in New York City, and the December 1999 $4 billion wind storms across Europe indicate the costs of climate change will borne by all.

CONCLUSIONS

We are changing the chemistry of the air, and in the process altering the Earth's heat budget. Global environmental change, climate change, stratospheric ozone depletion, excessive use of chemicals, urbanization and land cover change constitute a destabilizing array of forcing factors. The resurgence of infectious disease at the close of the 20th century is one of several symptoms of global change, which includes the amphibian decline on six continents and the proliferation of harmful algal blooms along coastlines globally. The decline in one type of specialist – pollinators, such as bees, birds, bats, butterflies, and beetles – threatens flowering plants, including crops.

There may be positive impacts, such as a drop in schistosomiasis in regions too warm to sustain snail hosts and lower winter mortality (Martens *et al.*, 1997). However, the respiratory, and injury, impacts of wide swings in weather (e.g., warm winter rains interrupted by sudden cold snaps) may overshadow the health gains of warmer winters.

Global warming is not linear. North Polar and Greenland ice sheet melting in the Northern Atlantic, plus cold upwelling in the Pacific (La Niña conditions), may usher in cold weather across Europe. Thus the greatest threat of climate change lies, not with individual diseases, but with the impacts on ecosystems, water quality and quantity, soils, crops and marine life – the life support systems that underpin our health and well-being.

RECOMMENDATIONS

Solutions may be divided into three levels. First-level solutions are improved surveillance and response capability, including vaccine and drug development, and more widespread support for public health infrastructures.

The second level – ecological epidemiology – involves integrating health surveillance and environmental monitoring. Greater use of satellite imaging and climate forecasts

can complement field work to produce health early warning systems and facilitate timely, less costly, environmentally-friendly public health interventions (Colwell, 1996; Epstein, 1999).

The third level is prevention: addressing the underlying social and environmental practices. Population stabilization is essential and World Bank figures confirm that a society's ability to stabilize its population is directly related to its income equity.

Addressing the economic, ecological and climatic conditions converge upon a central question: the mode of development and how it is powered. Use of non-polluting, clean energy sources is essential as is the maintenance of forests and wetlands that buffer climatic extremes and control the proliferation of opportunists and diseases they carry.

Addressing all levels requires resources and financial incentives. Adoption of the Montreal Protocol on stratospheric ozone-depleting chemicals, also driven by health concerns, depended upon establishing a fund for technology development and transfer.

Funds are needed for basic public health measures, electricity (for refrigeration) and clean water, and these will require clean energy sources. Indeed the looming water crisis may help drive the energy transition, for clean sources to pump and purify water, power irrigation and desalinate the rising seas. The Fund for HIV/AIDS, TB and malaria is a major step forward and an important precedent.

The global environment is most at risk from climate change certain to accompany fossil fuel combustion, compounding the direct health hazards of air pollution and the ecological damages due to their mining, refining and transport.

Systems can be restabilized. Just as the body heals after illness and lungs can recover with cessation of smoking (if the assault is time limited and has surpassed a threshold of irreparable damage), resilience allows ecosystems to re-establish new (dynamically) stable states. The climate system can also be restabilized, but only if the rate of forcing factors is slowed. Greenhouse gas emissions must be reduced by 60% according to the Intergovernmental Panel on Climate Change (1996) to stabilize atmospheric concentrations. Implementing the United Nations-sponsored Framework Convention on Climate is an essential first step.

Fortunately, values can also change rapidly, sometimes more rapidly than the natural systems we depend upon. Perhaps we are underestimating the economic opportunities waiting if we use resources efficiently, generate energy cleanly, and restore essential functions of the natural environment. Curbing our unhealthy addiction to fossil fuels may be the lever that opens the portal to a productive and healthy future.

REFERENCES

Albritton, D L, Allen, M R, Baede, A P M *et al.* (2001) IPCC Working Group I Summary for Policy Makers, Third Assessment Report: Climate Change 2001: The Scientific Basis (available at http://www.ipcc.ch).

Anderson, P and Morales, F J (1993) The Emergence of New Plant Diseases, in *Disease in Evolution*, eds M E Wilson, R Levins, and A Spielman, New York Academy of Sciences, NY, 181–194.

Bouma, M J, Kovats, S, Cox, J, Goubet, S, and Haines, A (1997) A Global Assessment of El Niño's Disaster Burden, *Lancet*, **350**, 1435–1438.

Colwell, R R (1996) Global Climate and Infectious Disease: the Cholera Paradigm, *Science*, **274**, 2025–2031.

Daszak, P, Cunningham, A A, and Hyatt, A D (2000) Emerging Infectious Diseases of Wildlife – Threats to Biodiversity and Human Health, *Science*, **287**, 443–449.

Diaz, H F and Graham, N E (1996) Recent Changes in Tropical Freezing Heights and the Role of Sea Surface Temperature, *Nature*, **383**, 152–155.

Easterling, D R, Horton, B, Jones, P D, Peterson, T C, Karl, T R, Parker, D E, Salinger, M J, Razuvayev, V, Plummer, N, Jamason, P, and Folland, C K (1997) Maximum and Minimum Temperature Trends for the Globe, *Science*, **277**, 363–367.

Epstein, P R (1995) Emerging Diseases and Ecosystem Instability: New Threats to Public Health, *Am. J. Public Health*, **85**(2), 168–172.

Epstein, P R (1999) Climate and Health, *Science*, **285**, 347–348.

Epstein, P R (2000) Is Global Warming Harmful to Health? *Sci. Am.*, **August**, 50–57.

Epstein, P R, Diaz, H F, Elias, S, Grabherr, G, Graham, N E, Martens, W J M, Mosley-Thompson, E, and Susskind, E J (1998) Biological and Physical Signs of Climate Change: Focus on Mosquito-borne Disease, *Bull. Am. Meteorol. Soc.*, **78**, 409–417.

Gill, C A (1920) The Relationship between Malaria and Rainfall, *Indian J. Med. Res.*, **8**, 618–632.

Harvell, C D, Kim, K, Burkholder, J M, Colwell, R R, Epstein, P R, Grimes, J, Hofmann, E E, Lipp, E, Osterhaus, A D M E, Overstreet, R, Porter, J W, Smith, G W, and Vasta, G (1999) Diseases in the Ocean: Emerging Pathogens, Climate Links, and Anthropogenic Factors, *Science*, **285**, 1505–1510.

International Federation of Red Cross and Red Crescent Societies (IFRC and RCS) (1999) *World Disasters Report*, Oxford University Press, NY.

Karl, T R, Knight, R W, and Plummer, N (1995) Trends in High-frequency Climate Variability in the Twentieth Century, *Nature*, **377**, 217–220.

Krabill, W, Frederick, E, Manizade, S, Martin, C, Sonntag, J, Swift, R, Thomas, R, Wright, W, and Yungel, J (1999) Rapid Thinning of Parts of the Southern Greenland Ice Sheet, *Science*, **283**, 1522–1524.

Levitus, S, Antonov, J I, Boyer, T P, and Stephens, C (2000) Warming of the World Ocean, *Science*, **287**, 2225–2229.

Lindgren, E and Gustafson, R (2001) Tick-borne Encephalitis in Sweden and Climate Change, *Lancet*, **358**, 16–18.

Mann, M E, Bradley, R S, and Hughes, M K (1998) Global-scale Temperature Patterns and Climate Forcing Over the Past Six Centuries, *Nature*, **392**, 779–787.

Martens, W J M, Jetten, T H, and Focks, D (1997) Sensitivity of Malaria, Schistosomiasis and Dengue to Global Warming, *Clim. Change*, **35**, 145–156.

McArthur, R H (1972) *Geographical Ecology*, Harper and Row, NY.

McMichael, A J, Haines, A, Slooff, R, and Kovats, S, eds (1996) *Climate Change and Human Health*, World Health Organization, World Meteorological Organization, United Nations Environmental Programme, Geneva.

Patz, J A, Epstein, P R, Burke, T A, and Balbus, J M (1996) Global Climate Change and Emerging Infectious Diseases, *J. Am. Med. Assoc.*, **275**, 217–223.

Peters, R L (1991) Consequences of Global Warming for Biological Diversity, in *Global Climate Change and Life on Earth*, ed R L Wyman, Routledge, Chapman and Hall, NY.

Rogers, D J and Randolph, S E (2000) The Global Spread of Malaria in a Future, Warmer World, *Science*, **289**, 1763–1766.

Timmermann, A, Oberhuber, J, Bacher, A, Esch, M, Latif, M, and Roeckner, E, (1999) Increased El Niño Frequency in a Climate Model Forced by Future Greenhouse Warming, *Nature*, **398**, 694–697.

Trenberth, K E (1999) The Extreme Weather Events of 1997 and 1998, *Consequences*, **5**, 3–15.

World Health Organization (1996) *The World Health Report 1996*: *Fighting Disease, Fostering Development*, WHO, Geneva.

Zucker, J R (1996) Changing Patterns of Autochthonous Malaria Transmission in the United States: a Review of Recent Outbreaks, *Emerg. Infect. Dis.*, **2**, 37.

Inter American Institute for Global Change (IAIGC)

Diana Liverman
University of Arizona, Tucson, AZ, USA

The Inter American Institute for Global Change (IAI) is an intergovernmental institution dedicated to the better understanding of global environmental change (e.g., climate variability, comparative studies of ecosystems, water and land use, air pollution, integrated assessment and human dimensions) in North and South America through collaborative research, networks of scientists, education, training, and free and open exchange of relevant data.

The IAI is an intergovernmental institution, established in 1992, dedicated to the better understanding of global environmental change in North and South America through collaborative research, networks of scientists, education, training, and free and open exchange of relevant data. As of 2000, the IAI had 18 member countries and had provided almost $20 m in grants for workshops, projects, networks, and training to scientists in the region. The majority of these grants are targeted at activities involving scientists from at least four member countries and with substantial education and training components that would build scientific capacity, especially in Latin America.

The IAI is managed by a small directorate, based in São José dos Campos, Brazil, advised by a 10-member scientific advisory committee, and reporting to an executive council and conference of parties consisting of representatives of member governments. Financial support has included contributions from member governments and national science foundations, and a training grant from the Global Environment Facility. The current members of the IAI are Argentina, Bolivia, Brazil, Canada, Chile, Colombia, Costa Rica, Cuba, Ecuador, Guatemala, Jamaica, Mexico, Panama, Paraguay, Peru, Dominican Republic, United States, Uruguay, and Venezuela.

The IAI science agenda focuses on major regional environmental issues including: (1) understanding climate variability associated with ocean–land–atmosphere interactions; (2) comparative studies of ecosystems, biodiversity, land use and water resources from the tropics to temperate and cold latitudes, including terrestrial, coastal and oceanic environments; (3) changes in the composition of the atmosphere including greenhouse gas emissions, ozone depletion and regional air pollution; and (4) integrated assessments, human dimensions and applications to decision making and policy. For example, several projects and networks of scientists are focusing on understanding and predicting climate variability associated with El Niño, and others seek to understand the impacts of land use change on biogeochemistry, ecosystems, and water quality. Other networks are reconstructing the environmental history of the Americas through dendrochronology, monitoring and analyzing ultraviolet radiation and its impacts, and focusing on understanding the climate of specific regions such as the Caribbean and South Atlantic. Of direct relevance to policy are projects that examine vulnerability and responses to climatic disasters, climate and health, fisheries dynamics, and the socioeconomic drivers of land use change.

Challenges facing the IAI include the establishment of a sustainable and diverse financial base, enhancing capacity in countries with smaller scientific communities, and the more effective integration of social science and policy into activities and projects. The IAI is making progress in establishing networks of researchers conducting multinational, collaborative research and in building a regional community of scientists who can contribute to international assessments of global change and advise policy makers in individual countries.

The IAI web site (http://www.iai.int) provides further information and updates on the IAI, including copies of

newsletters, lists of projects and publications, and contact information.

See also: **IAI (Inter-American Institute for Global Change Research)**, Volume 4.

International Biosphere Program (IBP) - a Foundation for Global Studies

Frank B Golley
University of Georgia, Athens, GA, USA

IBP was launched in 1964, with the aim of contributing to the optimal exploitation, on a global basis, of the biological resources on which humans are dependent for food and other products. The IBP was sponsored by the International Council for Science, and planned to terminate in 1972 but continued until 1974.

While the IBP initially had broad intentions, it developed into a largely ecological project. Its scientific director, Barton Worthington (1975, pg. 64), commented "it is the ecosystem approach which distinguishes much of the IBP research from what dominated ecology before …. The application of system analysis to biological systems has been one of the major innovations developed during the IBP. Some would go further and say that it has been one of the IBP's major accomplishments".

Given the knowledge of ecology at the time, several steps were required to organize a global project. First, the Earth's terrestrial surface was divided into geographical units, such as habitats or biomes. Next, sites representing the average condition of these biomes were identified and studied intensively. The intensive studies proceeded in two directions. First, one could treat the site as an ecosystem and measure the productivity of the vegetation and herbivore, carnivore and decomposer subsystems. The performances of these linked units were integrated using ecosystem models and were extrapolated to the entire biome. Alternatively, one could begin with the biological species encountered at the representative site, study their feeding relationships and organize them into food webs. Ultimately the food webs were assembled into subsystems or representative ecosystems. Both approaches were employed in IBP.

The ecosystem studies were most highly developed in the US where teams of researchers studied biological productivity and factors promoting or controlling production at representative locations. Variation across the biomes was determined by taking measurements at other sites. The biome program combined terrestrial and aquatic studies and used ecological modeling. Biome studies were carried out in the tundra, taiga, grasslands, coniferous forests, deciduous forests and deserts. A tropical biome was never successfully organized, although tropical projects were carried out in Malaysia, India, Ivory Coast and Senegal.

The species-focused IBP projects were classical biological investigations and tapped into the wealth of information on such groups as small mammals, social insects, granivorous birds and soil organisms. These studies were located at study sites representing different habitats, Traditional biologists felt more comfortable in carrying out comparative studies across habitats, using familiar biological categories. Eventually the species-oriented studies were assembled into whole systems so that both approaches were brought together into what was called the IBP biome program. However, these were not the only IBP projects. Other IBP efforts considered conservation, human adaptation to extreme environments, islands, marine upwellings and freshwater ecology.

The IBP was extended for two years and ended in 1974. Over 2000 scientists were involved. They produced an enormous literature, which has never been fully catalogued. In most countries IBP work disappeared when IBP ended. However, in a few countries, such as the US, IBP was reorganized and research was continued. In the US, biome studies were relabeled as ecosystem studies, and were formally made a program of the Division of Environmental Biology of the US National Science Foundation (NSF). In 1979 a Long-Term Ecological Research Program was initiated by NSF, which incorporated several IBP sites into a national program of ecosystem research.

Summarizing: IBP resulted in a substantial increase in our understanding of the geographical pattern of ecological productivity. It also developed a global network of scientists experienced in cooperative international work. IBP studies contributed to the development of hierarchy theory, landscape ecology, global biogeochemistry, and long-term ecological research.

See also: **Long-term Ecological Research (LTER) System**, Volume 2.

REFERENCE

Barton Worthington, E (1975) *The Evolution of the International Biosphere Program*, Cambridge University Press, Cambridge.

International Geosphere-Biosphere Programme (IGBP)

see IGBP (International Geosphere- Biosphere Programme) (Volume 2)

International Geosphere-Biosphere Programme Data and Information System (IGBP-DIS)

see IGBP-DIS (International Geosphere- Biosphere Program Data and Information System) (Volume 2)

Intertidal Zones

Virginia Burkett
US Geological Survey, Lafayette, LA, USA

The intertidal zone occurs within the range of high and low tides along the shorelines of coastal landforms and estuaries. The intertidal zone is a physically dynamic zone that is exposed to alternate flooding and drying, heat from the sun, wind, waves and currents. Organisms that inhabit intertidal systems are uniquely adapted to environmental conditions that occur along the energy, salinity and moisture gradients that extend from the subtidal region to the top of the intertidal zone. In contrast to terrestrial systems, which have physical gradients that can stretch over 10s or 1000s of km, intertidal gradients are relatively short, particularly along rocky shores (Raffaelli et al.*, 1991). The intensely specialized and often crowded intertidal environment is considered more fragile than many other ecosystem types.*

The intertidal zone may include coastal marshes, sand and gravel beaches, estuaries, mangrove swamps, algal beds, unvegetated mud and sand flats, and rocky shorelines that are periodically flooded by tidal waters. Coral and oyster reefs that are subtidal, irregularly exposed and regularly or irregularly flooded, are also generally considered intertidal ecosystems. Environmental gradients that typify intertidal zones are responsible for their high biodiversity, due to different species tolerances for the physical forces that structure intertidal plant and animal communities (Day et al.*, 1989). In most regions, intertidal zones are subject to bi-directional flooding about twice every day, which is one of the most striking physical attributes of these systems, and one that influences both their biotic and abiotic structure (Wiegert and Freeman, 1990). At high latitudes, ice cover may last for several months, creating relatively harsh intertidal environments with unique faunal assemblages.*

There are many classifications of intertidal zones based on their biota, substrate and ecosystem type (bay and estuary, open coast). At the simplest level, however, intertidal zones are generally classified as upper, middle and lower.

UPPER

The upper part of an intertidal ecosystem is the least exposed to water. It is only inundated by the highest tides or it may receive moisture primarily from sea spray or wave splash. Some classification schemes include beaches in this zone, but generally the upper part includes only the more frequently wetted regions that lie between the mean flood of the highest tides and slightly below mean sea level. In general, the upper intertidal zone is above the mussel beds that are common on many coasts. Organisms that inhabit the upper zone include barnacles and other animals accustomed to tolerating more air than water (Ricketts *et al.*, 1968).

The upper intertidal zone on rocky shores is often call the *littorinid* zone because it is commonly populated with snails of the genus *Littorina* (Latin for shore-dweller) and a wide variety of other gastropods, flatworms and crustaceans that stay close enough to the sea to receive the occasional moisture needed for their survival (Ricketts *et al.*, 1968). Adaptations that prevent desiccation of animals living in this zone include thick mucous coverings, shells and hard exoskeletons.

Algal species in the upper intertidal zone may be perennial or seasonal, responding to ever-changing tide levels. Mangroves (halophytic trees and shrubs) and salt marsh grasses (halophytic grasses, rushes and sedges) commonly occur in the upper to upper–middle intertidal zones. Tidal fresh marshes are considered upper intertidal plant communities. Mangroves dominate intertidal subtropical and tropical estuaries between 25 °N and 25 °S latitude, and salt marshes dominate the mid to upper intertidal areas of more temperate coastlines (Keannish, 1986). Black lichens are common in upper intertidal sandy shores in some regions (Raffaelli *et al.*, 1991).

MIDDLE

The middle intertidal zone is usually inundated during high tides and exposed during low tides. On the west coast of North America this zone is typically populated with mussels

that are either attached to the substrate or to each other by abyssal threads. The mussels themselves create a unique substrate that provides food and shelter for other organisms. Many species of crabs, barnacles and tiny crustaceans can be found in the middle intertidal zone, depending upon wave energy, substrate, temperature and other factors. In rocky substrates, sea anemones are common, as well as many barnacle species. Numerous species of limpets, seastars, hermit crabs, nudibranchs (shell-less snails), chitons, hydroids, and roundworms are also common among middle intertidal rocks and pools in warm to temperate waters (Ricketts *et al.*, 1968).

Tidal flats along coastlines dominated by unconsolidated sediments are commonly classified within the middle intertidal zone. They support rich benthic communities and serve as a nursery ground for many species of fish and shellfish. These unvegetated portions of sounds, lagoons, estuaries and river mouths, and shorelines that lie between the high and low tide mark are also critical habitats for many bird species. Tidal flats receive and transform relatively large amounts of organic detritus from the land and sea, which accounts for their high productivity and important function within the coastal food chain (Reise, 1985). Microalgal mats, filamentous algae, bacteria and fungi of tidal flats play an important role in the dynamics of estuarine systems by converting energy bound up in detritus to forms that are usable to consumer organisms (Peterson and Peterson, 1979).

Macrophtyes may also be found in the middle intertidal zone, including mangroves and the more salt tolerant marsh grasses. Along some low energy coastlines, eelgrass (*Zostera*) flats may be frequently exposed during low tide.

LOWER

The lower intertidal zone, covered by tides nearly all of the time, often supports large brown algae in temperate regions, and red algae in warm–temperate regions (Raffaelli *et al.*, 1991). Few of the organisms found in this zone can tolerate long exposure out of water. They are generally bared only by lower-than-average low tides (*minus tides*) and therefore infrequently if not rarely exposed to wave action, direct solar radiation and wind. Low intertidal pools of the US Pacific coastline are often populated by molting dungeness crabs (in Oregon and Washington) and spiny lobsters (in California). Oyster and coral reefs that are exposed during the lowest tides are also considered intertidal. Although seagrasses are generally considered subtidal, their light requirements place them at the deep edge of the lower intertidal zone.

Fish that are commonly found in this zone include mature forms as well as a wide range of juvenile pelagic species that seek shelter, food and other accommodations found in this environment. Bottom dwellers, such as the Atlantic flounder, are common in low intertidal waters of the Gulf of Mexico and temperate to warm Atlantic coastlines.

Occurring along the narrow border between inland and marine ecosystems, the intertidal zone is considered highly vulnerable to the effects of human development activities, sea level rise and changes in climate. Although coastal areas comprise only about 5% of the world's surface, it is estimated that 50–70% of the global human population lives in the coastal zone (Intergovernmental Panel on Climate Change (IPCC), 1995). For example, approximately 53% of the total population of the US lives in 17% of the land area considered coastal (Culliton, 1998). Along heavily developed shorelines, seawalls, jetties, levees and navigation channels have disrupted water circulation patterns; altered salinity levels; and changed freshwater, sediment and nutrient delivery. The loss of plant and animal communities due to pollution, overharvesting and a variety of other stresses is also common along developed coastlines.

Over 90% of the intertidal wetlands in California have been lost as a result of human development activity (Watzin and Gosselink, 1992). Large areas of intertidal marsh and mangroves swamps have been impounded and converted to shrimp farms in Honduras, Bangladesh, and Mexico. Human-induced environmental change is subtler in areas that are impacted by discharges from sewage and industrial plants, agricultural runoff, and invasive species. An estimated 40% of US estuaries are impaired by some form of pollution or habitat degradation (EPA, 1996). Nutrient enrichment associated with agricultural runoff can result in estuarine hypoxia, harmful algal blooms and fish kills.

Climate variability and change can also alter the structure and ecological function of intertidal systems. Rapid changes can occur when storms produce catastrophic waves that reshape the sea/land boundary. Natural multi-year climate/ocean regime shifts such as the El Niño/Southern Oscillation (ENSO) and the Pacific Decadal Oscillation can also modify intertidal ecosystems. On the Pacific coast of North America, the prolonged El Niño of 1982–1983 and 1997–1998 brought winter storms farther south, causing extensive coastal erosion and much higher than average tide levels (Field *et al.*, in press).

Longer-term climatic change associated with global warming could also have significant, and possibly serious, effects on most intertidal systems. The IPCC Third Assessment Report (McLean *et al.*, in review) suggests key changes to the ocean environment that portend large-scale changes of the world's intertidal zones:

- increases in sea-surface temperature, sea level and coastal inundation;
- decreases in sea-ice cover;
- changes in salinity, alkalinity, wave action and circulation.

Sea-level rise is regarded as one of the more certain consequences of increased global temperature. Average global sea level rose between 4–9 inches (10–25 cm) during the past 100 years (IPCC, 1995) and this average rate is projected to accelerate two to five-fold over the next 100 years (IPCC, 1998). Increased sea level rise can lead to permanent inundation of coastal marshes and other intertidal habitats that cannot migrate inland or accrete sediment vertically at a rate that equals or exceeds sea level rise (Cahoon *et al.*, 1998). Low-lying wetlands that are subsiding, receiving inadequate sediment or freshwater supply, or stressed by salt-water intrusion (another impact associated with sea level rise) are particularly vulnerable (Burkett and Kusler, 2000).

Man's development activities can also affect the survival of intertidal systems as sea level rises (Figure 1). Levees, seawalls, roads and other development can prevent the inland migration of seagrass, beaches, wetlands and other coastal features as sea level rises.

Countries that have established various forms of managed retreat requirements to prevent intertidal and other coastal ecosystems from being squeezed between development and the advancing sea include Barbados, Sri Lanka and the UK (McLean *et al.*, in review).

Sea-level rise will increase tidal flushing in estuaries and will also increase storm surges over coastal landforms. Both average and peak salinity levels will increase in estuaries and adjacent habitats, thereby altering the zonation of vegetation and other biota. Increased storm surge and mean tide levels will also alter the disturbance regime in shallow coastal waters, thereby influencing the composition and productivity of seagrasses (Short and Neckles, 1999) and benthic fauna that are vulnerable to changes in sedimentation patterns, current velocity, and turbidity. The production of estuarine-dependent fish and shellfish will likely decline in regions where wetlands, seagrasses and other elements of the intertidal food chain are adversely impacted.

An analysis of approximately five million ocean temperature profiles taken over the past 55 years by Levitus *et al.* (2000) indicates that the mean temperature of the oceans between 0–300 m increased by 0.31 °C during that same period. Although this increase cannot be linked conclusively with global climate change, the increase is in line with land-based global temperature trends. Increased ocean temperature, most evident in shallow waters, can change the structure and productivity of intertidal ecosystems by directly altering the metabolic rates and phenology

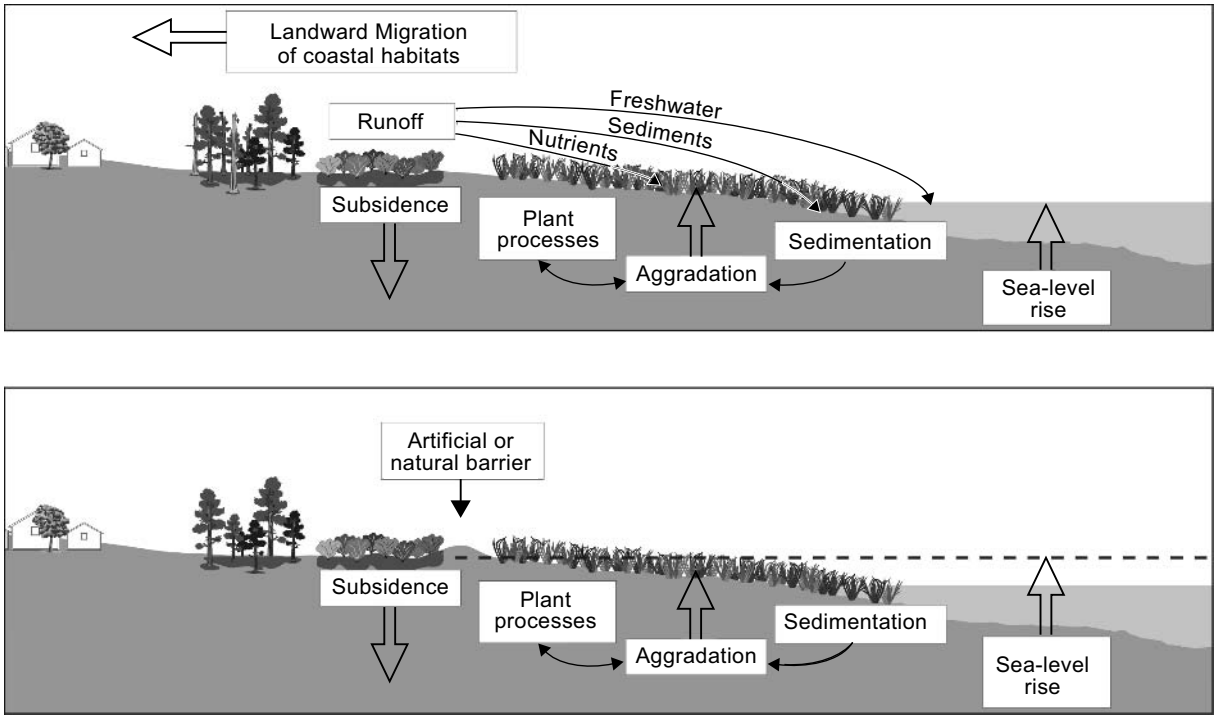


Figure 1 Through the accretion of organic and inorganic sediments, many vegetated coastal wetlands have the capacity to maintain their elevation as sea-level rises. Thresholds at which sea-level rise results in wetland loss varies among coastal regions due to differences in sediment accretion rates, subsidence and other factors (e.g., tectonism and isostacy) that affect the elevation of the land surface. If the rate of vertical marsh movement does not keep pace with sea-level rise, coastal wetlands will either be submerged or displaced landward. Natural or arti cial barriers can prevent the landward migration of coastal habitats

of organisms. Indirectly, increased ocean temperature can impact intertidal ecosystem productivity by altering water column stratification, circulation and mixing patterns, nutrient cycling and many other coastal processes.

Increases in hurricane wind strength could result from future elevated sea surface temperatures over the next 50–100 years. A recent model investigation shows that increases in hurricane wind strength of 5–10% are possible with a sea surface warming of 2.2 °C (Knutson, 1998). Other research supports the possibility that tropical cyclones could become more intense or more frequent (Raper, 1993; Kerr, 1999). For a moderate hurricane, a 5–10% increase in wind strength would translate into approximately a 25% increase in the destructive power of the winds; the resulting increase in wave action and storm surge would be greater than the percentage increase in wind speed. Similarly, wave height and storm surge would have a greater percentage increase than wind speed, potentially yielding increased impacts to coasts (Field *et al.*, in press).

Some recent global climate model investigations have shown that ENSO extremes could become more frequent with increasing greenhouse gas concentrations. For example, Timmerman *et al.* (1999) suggest that the tropical Pacific is likely to shift to a state corresponding to present-day El Niño conditions. Since fewer hurricanes occur in the Atlantic during El Niño years, Atlantic hurricanes will likely decrease in frequency. However, while prolonged El Niño conditions could decrease the occurrence of hurricanes in the Atlantic, they would likely enhance coastal impacts by winter storms on the Pacific coasts of Latin and North America.

Hydrological cycles are fundamental components of climate, and climate change is likely to affect freshwater runoff to the intertidal environment. In general, the consensus is that the hydrologic cycle will become more intense, with average precipitation increasing, especially at high latitudes (Field *et al.*, in press). The effects of these alterations in runoff to coastal zones and estuaries may include changes in freshwater and sediment inflows, water quality (particularly salinity and nutrient loading) and frequency of coastal flooding.

See also: **Tides, Oceanic**, Volume 1; **Mangrove Ecosystems**, Volume 2.

REFERENCES

Burkett, V and Kusler, J (2000) Climate Change: Potential Impacts and Interactions in Wetlands of the United States, *J. Am. Water Res. Assoc.*, **36**(2), 313–320.

Cahoon, D R, Day, J W, Reed, D J, and Young, R S (1998) Global Climate Change and Sea-level Rise: Estimating the Potential for Submergence of Coastal Wetlands, in *Vulnerability of Coastal Wetlands in the Southeastern United States: Climate Change Research Results*, US Geological Survey, Biological Science Report – USGS/BRD/BSR 1998–0002, 21–34.

Culliton, T J (1998) *Population: Distribution, Density and Growth*, NOAA State of the Coast Report, Silver Spring, MD: NOAA, URL: http://state-of-coast.noaa.gov/topics/html/pressure.html.

Day, J, Hall, C A S, and Kemp, W M (1989) *Estuarine Ecology*, John Wiley & Sons, New York, 1–558.

Field, J C, Boesch, D F, Scavia, D, Buddemeier, R, Burkett, V R, Cayan, D, Fogarty, M, Harwell, M, Howarth, R, Mason, C, Park, R A, Pietrafesa, L F, Reed, D, Royer, T, Sallenger, A, Spranger, M, and Titus, J G (2001) Potential Consequences of Climate Variability and Change on Coastal Areas and Marine Resources, in *Climate Change Impacts in the United States: Foundation Document*, US Global Change Research Program, Washington, DC, in press.

IPCC (1995) *Climate Change 1995: Impacts, Adaptations and Mitigation of Climate Change: Scientific–Technical Analysis*, Cambridge University Press, Cambridge, 1–878.

IPCC (1998) The Regional Impacts of Climate Change: an Assessment of Vulnerability, Cambridge University Press, Cambridge, 1–514.

Keannish, M J (1986) *Ecology of Estuaries*, CRC Press, Boca Raton, FL, 1–254.

Kerr, E A (1999) Thermodynamic Control of Hurricane Intensity, *Nature*, **401**, 665–669.

Knutson, R R, Tuleya, R E, and Kurihara, Y (1998) Simulated Increase of Hurricane Intensities in a CO_2 Warmed Climate, *Science*, **279**, 1018–1020.

Levitus, S, Antonov, J I, Boyer, T P, and Stephans, C (2000) Warming of the World Ocean, *Science*, **287**, 2225–2229.

McLean, R F, Tsyban, A, Burkett, V, Codignotto, J O, Forbes, D L, Ittekkot, V, Mimura, N, and Beamish, R J (in review) Coastal Zones and Marine Ecosystems, in *Intergovernmental Panel on Climate Change (IPCC) Third Assessment Report*.

Peterson, C H and Peterson, N (1979) *The Ecology of Intertidal Flats of North Carolina: a Community Profile*, US Department of Interior, Fish and Wildlife Service, Washington, DC, FWS/OBS-79/39, 1–73.

Raffaelli, D, Karakassis, I, and Galloway, A (1991) Zonation Schemes on Sandy Shores: a Multivariate Approach, *J. Exp. Mar. Biol. Ecol.*, **148**, 241–253.

Raper, S (1993) Observational Data on the Relationship between Climatic Change and the Frequency and Magnitude of Severe Storms, in *Climate and Sea Level Change: Observations, Projections and Implications*, Cambridge University Press, Cambridge, 192–212.

Reise, K (1985) *Tidal Flat Ecology*, Springer-Verlag, New York, 1–191.

Ricketts, E F, Calvin, J, and Hedgepeth, J (1968) *Between Pacific Tides*, Stanford University Press, Stanford, CA, 1–614.

Short, F T and Neckles, H A (1999) The Effects of Global Climate Change on Seagrasses, *Aquat. Bot.*, **63**, 169–196.

Snead, R (1982) *Coastal Landforms and Surface Features*, Hutchinson Ross Publishing Company, Stroudsburg, PA, 1–247.

Timmerman, A J, Oberhuber, J, Bacher, A, Esch, M, Latif, M, and Roeckner, E (1999) Increased El Nino Frequency in a Climate Model Forced by Future Greenhouse Warming, *Nature*, **398**, 694–696.

US Environmental Protection Agency (1996) *Municipal Solid Waste Characterization Report*, EPA/530-R-97–015.

Watzin, M C and Gosselink, J G (1992) *The Fragile Fringe: Coastal Wetlands of the United States*, Louisiana Sea Grant College Program, Louisiana State University, Baton Rouge, LA, 1–16.

Wiegert, R G and Freeman, B J (1990) *Tidal Salt Marshes of the Southeast Atlantic Coast: a Community Profile*, US Department of Interior, Fish and Wildlife Service, Washington, DC, Biological Report 85(7.29).

IPCC (Intergovernmental Panel on Climate Change)

see IPCC (Intergovernmental Panel on Climate Change) (**Volume 1**)

Iron Cycle

Keith A Hunter

University of Otago, Dunedin, New Zealand

The importance of iron as an essential chemical element in a wide range of biological processes is well known. Iron is the key component of many enzymes, as well as being the central ingredient of the oxygen-carrying protein, hemoglobin, in human blood. Yet despite being one of the most abundant chemical elements in the Earth's crust, iron is present at only trace levels in the oceans, the original birthplace of all biological systems.

There is now mounting scientific evidence that low iron availability in the global oceans is one of the most important factors controlling the growth rate of oceanic phytoplankton. These primary organisms not only represent the base of the marine food chain, but also act as an important control on the distribution of the greenhouse gas, carbon dioxide (CO_2), between the atmosphere and the ocean. Since the ocean is the dominant reservoir of carbon, changes in the availability of iron to phytoplankton within it may be quickly reflected in the Earth's climate. Indeed, there is compelling indirect evidence that during glacial times, an increased supply of iron to the ocean surface from increased weathering and transport of terrestrial dust (a rich source of iron) may have stimulated additional uptake of CO_2 by phytoplankton. This, in turn, would have the effect of bringing down atmospheric CO_2 levels, thus diminishing the greenhouse effect and contributing to the colder temperatures characteristic of glacial periods.

Not surprisingly, there has been vigorous scientific activity in this interesting field over the last two decades. However, many of the key features of the role of iron in controlling climate remain elusive, mainly because of the complexity of both its chemical behavior and the biological processes in which it is involved. Nonetheless, the evidence assembled so far points towards iron being a key component in the complex network of controls on global climate contained in the Earth's ocean–atmosphere system.

THE CLIMATE CHEMISTRY OF THE OCEAN

The importance of iron in global climate change derives from its ability to influence and control key processes in the global carbon (C) cycle (Figure 1). The level of atmospheric carbon dioxide (CO_2) determines, through greenhouse warming, the net radiation balance and therefore temperature of the atmosphere. However, as Figure 1 demonstrates, atmospheric CO_2 is dynamically linked, by rapid air–sea exchange, to a much greater reservoir of carbon dissolved in the oceans. As a result, long-term changes in atmospheric CO_2 are thought to result primarily from changes in the CO_2 chemistry of the oceans.

These changes are, however, slow. Most of the ocean's volume (about 95%) consists of very cold, deep water that spends on average 1000–2000 years circulating in the deep ocean basins out of contact with the atmosphere. As a result, the comparatively rapid introduction of fossil fuel carbon into the system means there has not yet been enough time for the extra CO_2 to be absorbed into this deep ocean reservoir, causing atmospheric CO_2 levels to rise over the last 200 years.

Lying above most of the deep ocean is a thin stable layer of warm water that is heated by the atmosphere. Only at the highest latitudes does the cold, deep water come into direct contact with the atmosphere. In these surface waters, the primary organisms of the marine biota, known as phytoplankton, convert CO_2 into organic tissue using sunlight energy in the process of photosynthesis. As well as requiring CO_2 and sunlight, photosynthesis also requires a whole raft of essential chemical ingredients that must be present in the seawater. These include nitrate and other forms of nitrogen (used, for example, in amino acids), phosphate (used for phosphate esters such as adenosine triphosphate), silicon (used for structural components of diatoms), iron (various iron-based enzymes) and other metals. These ingredients are usually termed nutrients.

Figure 1 shows that of the 50 Gt year^{-1} of carbon (1 Gt = 1 gigaton or 1 billion metric tons of carbon) taken up by photosynthesis into the marine biota, 40 Gt or 80% is put back directly into the surface waters by respiration. This process, which is just the reversal of photosynthesis,

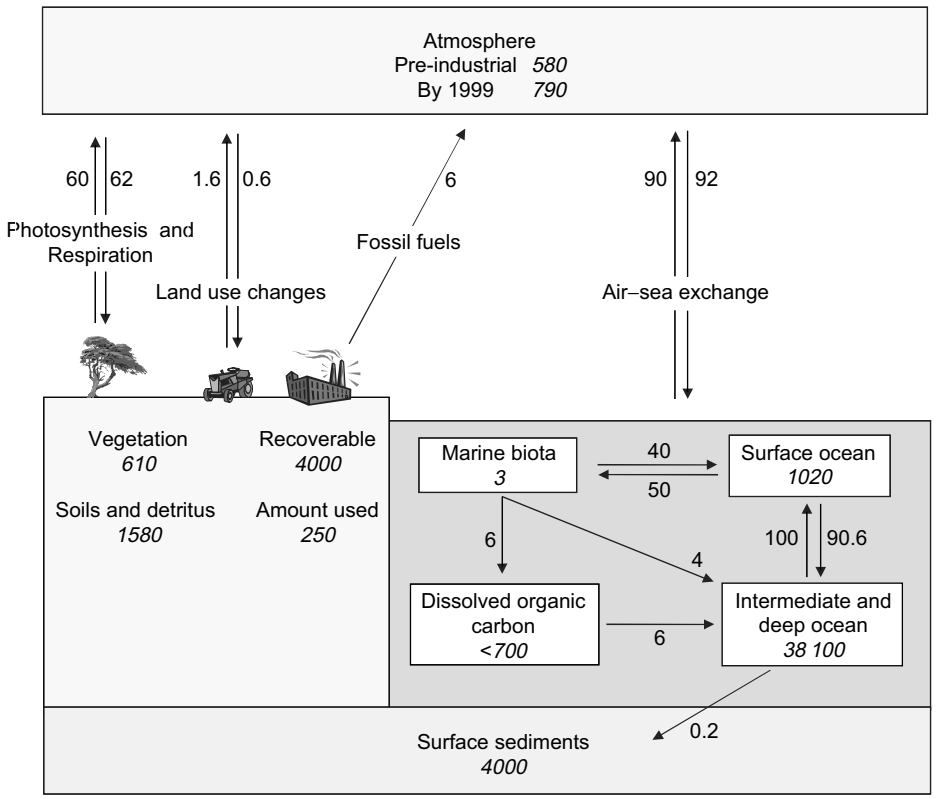


Figure 1 Global cycle of CO_2 in the ocean-atmosphere system representative of the present day. Fluxes between reservoirs are expressed in terms of gigatonnes of carbon per year (Gt C year^{-1}, 1 Gt being 1×10^9 tonnes or 1×10^{15} g of CO_2 expressed as carbon). Quantities in each reservoir are shown in Gt C as italicized numbers. Adapted from the IPCC document (Houghton *et al.*, 1996) by adding additional estimates of fossil fuel reservoirs and consumption, and by augmenting the original atmospheric inventory from 750 to 790 Gt assuming an annual growth of 3.3 Gt year^{-1} since the mid-1980s. Fluxes of CO_2 within the ocean have been modified from IPCC estimates to give balanced fluxes and a better match to measurements of the CO_2 increase in surface waters

occurs mainly through consumption by microscopic animals known as zooplankton, and through bacteria that feed on organic tissue. As a result of respiration, 80% of the nutrients needed for further photosynthesis are recycled within the surface waters.

THE BIOLOGICAL PUMP

The remaining 20% of carbon is transferred into the deep ocean as rapidly-settling particles comprising dead cells, fecal matter and other biological detritus. There, it eventually undergoes respiration, both directly and through the intermediate formation of breakdown products known as dissolved organic carbon. However, the slow circulation of the deep ocean means that recycling of the carbon and nutrients back to the surface takes thousands of years.

This process represents a biological pump through which deep water becomes enriched in carbon and nutrients and surface water becomes depleted. The overall distribution of the different chemical ingredients depends on the balance between the biological pump and the physical pump represented by the circulation system. Figure 1 shows that for carbon, the physical pump brings into the deep layer 90.6 Gt year^{-1}, or about 90% of the 100 Gt year^{-1} it pumps back to the surface. The balance of 10% is the carbon supplied by biological pumping.

This differential pumping has important consequences for atmospheric CO_2. Because of it, the total carbon content of the surface ocean is depleted by about 10%. In addition, because of the pH changes associated with photosynthesis, its pH is about 0.3 units higher than that of deep waters. The combined effect of these two factors means that if the biological pump were to cease, atmospheric CO_2 levels would increase to 2–3 times greater than they are now, precipitating a super-greenhouse effect. This underlines how the Earth's climate system is dynamically controlled, over short time scales, by the biological pump.

WHAT CONTROLS THE BIOLOGICAL PUMP?

Most of the plankton growth in the surface ocean is driven by locally recycled carbon and nutrients. However, the

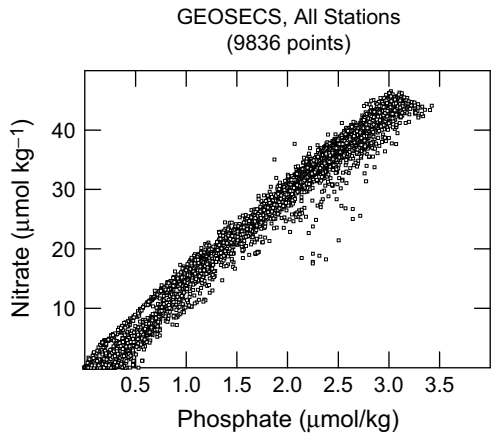


Figure 2 Relationship between measured concentrations of nitrate and phosphate in the global ocean. Drawn using data gathered during the GEOSECS programme

biological pump is driven only by the carbon and nutrients recycled into the surface layer from the deep ocean. This represents only a fraction of the total growth rate, and is often termed new production. Although there is agreement on what the concept of new production means, a variety of conflicting methods have been used to measure it, the details of which are beyond the scope of this article.

Since the surface water concentration of carbon is only depleted by about 10%, photosynthesis is not limited by the supply of carbon. In regions of high latitude, a lack of sunlight may be a limiting factor for part of the year, but throughout much of the global ocean, new production seems to be co-limited by the availability of the nutrients nitrate and phosphate. Throughout the mid-latitude range, between approximately 45 °S and 45 °N in the main ocean basins, levels of both phosphate and nitrate in surface waters are essentially zero compared to deep waters. More importantly, the concentrations of these two substances are closely correlated with each other throughout the ocean, as shown in Figure 2. The remarkable consistency of this correlation points to closely regulated, parallel controls on the availability of both nutrients by phytoplankton. This implies that the biological pump has stripped surface waters of nitrate and phosphate until both are depleted to near zero levels. At this point, new production will be small and limited by the resupply of these substances from deep waters. This nitrate-phosphate limitation is one of the central paradigms of oceanography.

ANOMALOUS REGIONS OF THE OCEAN

However, some regions of the global ocean do not conform to this picture. These are where high residual levels of nutrients remain at the surface, but phytoplankton growth rates remain low, as evidenced by low concentrations of the plant pigment chlorophyll, which is characteristic of the abundance of photosynthesizing organisms. The advantage of using chlorophyll is that its concentration can be indirectly sensed using a satellite image of ocean color. Such areas are termed high-nitrate, low-chlorophyll (HNLC) areas and they defy the nitrate–phosphate paradigm.

The main HNLC regions are located in the equatorial Pacific Ocean, in the Pacific Ocean at high latitudes (sub-Arctic Pacific), and in the Southern Ocean at latitudes greater than 50 °S. All are recognized as areas of upwelling, where the nutrient-rich deep waters spill upwards to reach the surface. Therefore, they should be particularly favored regions for nitrate-supported new production.

Clearly, another key ingredient for photosynthesis is missing in HNLC waters. Satellite images of ocean color reveal that in these regions, the highest levels of chlorophyll pigments are consistently found in areas close to landmasses, or in the downstream influence of landmasses. This implies that the missing nutrient is one that is particularly abundant in the terrestrial environment.

IRON - IS IT THE MISSING INGREDIENT?

From the 1980s, mainly through the pioneering efforts of the late John Martin of Moss Landing Laboratories, CA, and amid much scientific controversy, the element iron has become recognized as an important nutrient responsible for most, if not all, the control on phytoplankton new productivity in HNLC areas.

Early experiments involved inoculating small bottles (1–30 l) of seawater taken from HNLC regions with a small amount of iron, and observing the effects of iron addition on growth rates and nutrient uptake. These were carried out on board ship immediately after collection of the water. The main finding was a dramatic increase in the growth rate of primary organisms, particularly phytoplankton having larger cells such as diatoms. The stimulation of production was not immediate, but became apparent only after a time lag of several days. A change in ecosystem structure was also observed, from primary producers utilizing largely ammonium as a nitrogen source to one with an increased demand for nitrate. Consequently, the originally high nitrate levels in the water decreased as production developed.

These bottle incubation experiments were strongly criticized because they do not accurately reproduce the ecological conditions in the actual surface ocean. Specifically, taking a small volume of seawater does not include a statistically significant concentration of the much less abundant zooplankton that graze on phytoplankton. Thus, the observed increases in growth rate may have resulted from a lack of grazing pressure. Another criticism is that the

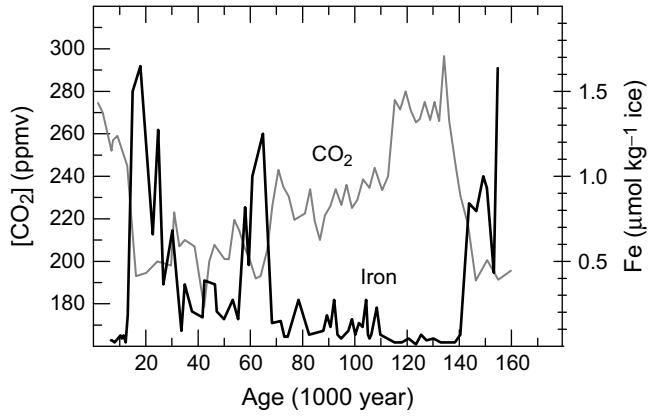


Figure 3 Comparison between CO_2 concentrations measured in air bubbles trapped in layers of ice taken from the Vostok core, East Antarctica (Barnola *et al.*, 1987) with iron concentrations derived from measurements of dust concentration in the same core (De Angelis *et al.*, 1987). The horizontal age scale is based on ice stratigraphy

walls of the bottle may provide an ecological niche that artificially enhances the growth rates of some species of phytoplankton.

In spite of these objections, the concept that iron may control phytoplankton growth in HNLC regions has intriguing implications for global climate change. John Martin was quick to point out that the periods of lowest CO_2 concentration during the last glacial period corresponded to periods of very high dust input because of larger arid regions (Figure 3). He suggested that iron-containing dust fertilized the glacial ocean, enhancing the uptake of carbon by the biological pump. However, this view is still hotly debated.

IRON CHEMISTRY AND ITS UPTAKE BY PHYTOPLANKTON

Iron is essential for the growth and metabolism of all organisms. It occurs in cytochromes and in proteins involved in key metabolic processes, e.g., photosynthetic and respiratory electron transport, nitrate and nitrite reduction, N_2 fixation, and sulfate reduction. Iron, therefore, has a wide variety of potentially important roles in marine phytoplankton, most of which are catalytic in nature, rather than being structural components of the cell. As a result, low iron availability should have its greatest influence on rates of metabolism and cell growth.

Phytoplankton that assimilate nitrate as their nitrogen source, such as diatoms, require nitrate reductase, an iron-containing enzyme. This explains in part why diatoms dominate after iron is added to iron-depleted waters. By contrast, the smaller organisms that are capable of assimilating reduced forms of nitrogen, such as ammonia, are more viable than diatoms under low iron conditions.

The chemistry of iron in seawater poses major problems for organisms requiring it for growth. Most iron in seawater is present in the oxidation state Fe (III), which in the alkaline conditions of seawater should readily form insoluble iron oxides. Calculations of the solubility of iron oxides suggest that the maximum concentration of Fe (III) should be less than 0.2 nM ($1\,nM = 10^{-9}\,mol\,l^{-1}$). In fact, throughout most of the ocean, measured Fe (III) concentrations are significantly higher (0.3–0.6 nM, with higher values in coastal waters). We now know that this superabundance is caused by the binding of Fe (III) to highly specific organic compounds that appear to be of biological origin. Indeed, we can speculate that if it were not for this organic binding, all of the ocean would be iron-limited. Thus, phytoplankton may have actually evolved to generate this iron-binding organic matter to create the conditions they require for growth!

It is not well understood how phytoplankton take up this organically bound iron from seawater. It seems clear that the bound iron is not itself biologically available, but first has to be released from its binding agent. This release may be the factor limiting the rate of iron uptake, and hence the rate of metabolism in iron-depleted regions. One mechanism for release is the action of sunlight on the iron-organic complex, which is known to generate Fe (II), which although unstable in seawater, may exist long enough to be assimilated by cells. An alternative mechanism involves the use of binding agents on the cell surface that assist in removing the iron from its complex. Although much is known about these mechanisms in other environments, e.g., soils, the very low levels of iron in the ocean make the study of its uptake extremely difficult. Therefore, much of our understanding at present is based on analogy rather than hard scientific evidence.

Experiments in the laboratory suggest that even coastal phytoplankton, which are usually not iron-limited, appear to be assimilating iron at optimal rates, given the constraints mentioned in the previous paragraph. Oceanic phytoplankton have, therefore, had to evolve strategies for achieving the same metabolic rates with a reduced iron supply. However, their flexibility is demonstrated by the fact that they can take advantage of variations in the iron supply, and go on an iron spending spree. This implies that they have alternative genetic pathways for synthesizing enzymes that depend on the local iron availability.

LARGE-SCALE IRON FERTILIZATION EXPERIMENTS

Due to the many objections to bottle-scale iron fertilization experiments mentioned earlier, much larger-scale experiments were advocated, in which a patch of surface water of the order of 10×10 km in size would be inoculated with iron. Over this distance scale, both grazing and

wall-effect anomalies would be absent. What made these experiments possible was the development of methods that would allow the location of a fertilized patch of water to be traced over a period of many days. This involved the addition of sulfur hexafluoride (SF_6), a harmless but readily-measured substance along with the iron. To date, three of these (very expensive!) experiments have been conducted.

The first two experiments, termed IRONEX I and II, took place in the equatorial Pacific Ocean (Martin *et al.*, 1994; Coale *et al.*, 1996). In both cases, a ~10 km-sized region of surface water was dosed with iron and the changes inside this region compared with control stations outside it for a period of 1–2 weeks. In IRONEX I, iron was added in a single addition, while in IRONEX II a similar total amount was added in three separate additions at three-day intervals with the aim of more nearly simulating a continuous supply from a natural source.

In both experiments, a significant increase in light energy consumption by the phytoplankton was measured within hours of iron addition, indicating new photosynthesis stimulated by the extra iron. Over the first few days, chlorophyll levels also increased significantly, and CO_2 levels began to decline. At the same time, significant levels of dimethylsulfide (DMS), an indirect product of phytoplankton growth, were observed (*see* **Dimethylsulfide (DMS)**, Volume 1).

However, after three days, the courses of the two experiments diverged. In IRONEX I, CO_2 levels in the fertilized patch did not decline further after the first two days. This shows that carbon was no longer being exported from the patch by the biological pump, implying that the grazing zooplankton responded rapidly to consume the enhanced phytoplankton growth. Moreover, organisms with larger cells such as diatoms, which are not as efficiently removed by grazing, did not have long enough to become dominant because the enhanced iron levels had decreased rapidly over the three-day period.

However, in IRONEX II, the same amount of iron was added in three separate doses, each three days apart, allowing the larger-celled diatom species to become dominant and eventually to form a bloom. This resulted in much more dramatic, and prolonged, decreases in CO_2 levels within the patch, indicating carbon export from the surface layer. Again, nitrate levels decreased significantly through this period, but did not reach zero. Instead, the ultimate limiting factor of the iron-stimulated patch appeared to be the diminished availability of silicate needed for diatom growth.

The third iron fertilization experiment (Southern Ocean Iron Release Experiment, SOIREE) was undertaken by a joint group from New Zealand, the United Kingdom and the Netherlands in early 1999 in the Southern Ocean, with additional dramatic results (Boyd *et al.*, 2000). A multiple dose of iron was used, as in IRONEX II, resulting in similar effects in terms of stimulated growth rates, nitrate uptake, decrease in CO_2 and DMS production. It was found, with great surprise, that although the added iron decreased in concentration over the first few days, afterwards it remained quite high throughout the experiment, contrary to expectations based on the known chemistry of iron. More significantly, satellite imagery indicated that the resultant bloom persisted for at least 55 days! In this case, it appears that the bloom did not run out of silicate because of the nature of the water movements adjacent to the patch that provided a re-supply of silicate as it was consumed by the diatom bloom.

IS IRON FERTILIZATION A FEASIBLE SOLUTION TO GLOBAL WARMING?

The realization that large areas of the global ocean were iron-limited quickly gave rise to the notion that it might be possible to fertilize HNLC areas with iron in order to bring down the concentration of CO_2 in the atmosphere. Calculations based on the amount of iron needed to stimulate new production of the required amount of carbon suggested that it might be economically sound. Iron is vastly abundant in the Earth's crust, and one of the cheapest mineral materials.

However, prudent examination suggests that such schemes would probably not work. The problem is to keep iron in the surface layer long enough to stimulate a reasonable amount of new production before it precipitates as iron oxide particles, but ensure that the newly-produced carbon sinks out of the surface layer. Clearly these two requirements are mutually exclusive. Indeed, detailed calculations suggest that the practical outcomes of iron fertilization would not be worth the trouble and expense in terms of the total amount of CO_2 that could be removed.

More importantly, it is probably extremely unwise to risk all of the unknown ecological and climate changes that might result from large-scale iron fertilization of the ocean (Chisholm, 2000). In terms of human influences on the global climate system, this might be the time to say enough is enough.

CONCLUSIONS

The last 10 years have seen the emergence of a fascinating new paradigm involving the importance of iron as a factor controlling the productivity of the global ocean. This involves the intriguing notion that phytoplankton have created the chemical conditions in the ocean that best suit them, in the process taking control of the global carbon cycle, with all that this implies for global climate change. There are many gaps in the picture, and future research will no doubt reveal even more subtle detail in this fascinating story.

REFERENCES

Barnola, J M, Raynaud, D, Korotkevich, Y S, and Lorius, C (1987) Vostok Ice Core Provides 160 000-year Record of Atmospheric CO_2, *Nature*, **329**, 408–414.

Boyd, P W, Watson, A, Law, C S, Abraham, E, Trull, T, Murdoch, R *et al.* (2000) Phytoplankton Bloom Upon Mesoscale Iron Fertilization of Polar Southern Ocean Waters, *Nature*, **407**, 695–702.

Chisholm, S W (2000) Stirring Times in the Southern Ocean, *Nature*, **407**, 685–687.

Coale, K H, Johnson, K S, Fitzwater, S E, Gordon, R M, Tanner, S, Chavez, F P *et al.* (1996) A Massive Phytoplankton Bloom Induced by an Ecosystem-scale Iron Fertilization Experiment in the Equatorial Pacific Ocean, *Nature*, **383**, 495–501.

De Angelis, M, Barkov, N I, and Petrov, V N (1987) Aerosol Concentrations Over the Last Climatic Cycle (160 kyr) From an Antarctic Ice Core, *Nature*, **325**, 318–321.

Houghton, J, Meira Filho, L, Callander, B, Harris, N, Kattenberg, A, and Maskell, K (1996) *Climate Change 1995: the IPCC Scientific Assessment*, Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge.

Martin, J H, Coale, K H, Johnson, K S, Fitzwater, S E, Gordon, R M, Tanner, S J *et al.* (1994) Testing the Iron Hypothesis in the Ecosystems of the Equatorial Pacific Ocean, *Nature*, **371**, 123–129.

ISRIC (International Soil Reference and Information Center, The Netherlands)

see Data Banks: Archiving Ecological Data and Information (Volume 2)

ITEX (International Tundra Experiment)

see Temperature Increase: Effects on Terrestrial Ecosystems (Volume 2); Tundra (Volume 2)

JKL

JGOFS (Joint Global Ocean Flux Study)

see IGBP Core Projects (*Opening essay*, Volume 2)

Keystone Species

As the name implies, *keystone species* play such important and unique roles in the structure and functioning of an ecosystem that the ecosystem would be very different without them. Some keystone species help to maintain local diversity within an ecosystem by controlling populations of species that would otherwise dominate. For instance, the predatory starfish *Pisaster ochraceus* is considered a keystone species in the rocky marine intertidal of North America's Northwest Coast. By feeding on and thus suppressing populations of the mussel *Mytilus californianus*, the starfish maintains much of the local species diversity. Removal of the starfish in small-scale experiments has allowed mussel populations to increase so rapidly in size and spread so densely over the intertidal rocks within experimental enclosures that other species could find no space to gain a foothold. The starfish, however, has little overall effect on some other ecosystems in which it occurs and this shows that keystone status depends at least partly on the nature of the ecosystem.

Other keystone species help to maintain local species diversity by providing critical resources for a wide range of species. For instance, beaver are often thought of as keystone species in North America because the ponds behind the dams they build provide unique habitat for many amphibians, aquatic species, and water birds. Particular species (e.g., nodulated, nitrogen fixing plants) that carry out a crucial function for the rest of the ecosystem are also considered to be keystone species. Another example is the fig trees in the rain forests of South America and Borneo because they often bear fruit all year and thus provide a reliable food source for many mammals and birds during periods of scarcity.

FURTHER READING

Mills, L S, Soule, M E, and Doak, D F (1993) The Keystone-species Concept in Ecology and Conservation, *Bioscience*, **43**, 219–224.

Power, M E, Tilman, D, Estes, J A, Menge, B A, Bond, W J, Mills, L S, Daily, G, Castilla, J C, Lubchenco, J, and Paine, R (1996) Challenges in the Quest for Keystones, *Bioscience*, **46**, 609–620.

RICHARD A FLEMING *Canada*

KT Events

see Extinctions and Biodiversity in the Fossil Record (Volume 2)

Kudzu - Alien Invader

see Plant Dispersal and Migration (*Opening essay*, Volume 2)

Lakes and Rivers

John J Magnuson
University of Wisconsin-Madison, Madison, WI, USA

Lakes and rivers cover a tiny fraction of the earth's surface, but these ecosystems and the resources they contain are essential to human survival. Many of the ecosystem services they provide are under-valued. As ecosystems, lakes

and rivers cannot be understood, managed or conserved without including the interactions with their shorelines, their watersheds and their climates. Thus, their functioning in terms of biogeochemical cycling, productivity, and biodiversity habitat are influenced by anthropogenic changes not only to the waters themselves, but also to the land, the air and the climate.

Inland aquatic environments are relatively rare environments. Water covers about 75% of the earth's surface, but inland waters only constitute 0.0171% of the earth's surface waters by volume (0.0091% are freshwater). This contrasts with the 0.5% for groundwater, 2.2% for ice and the massive 97.3% for oceans (Cole, 1994).

Lakes and rivers are ecosystems – interacting physical and biological systems somewhat independent from adjacent lands and atmosphere. Rivers and streams are the arteries of our continents; and lakes, rivers and wetlands integrate the activities and events in their watersheds (Naiman *et al.*, 1995). They have their own biota, physics, and chemistry; their own processes and behavior, yet both are so dependent on their surrounding catchments or watersheds, their airsheds, and climate, that their behavior can be neither understood nor predicted without considering these external influences. The link to the watershed is so strong that an inland water ecosystem often is defined to be the water body plus its drainage system.

The science of inland waters is called *limnology*, the sister science to oceanography or marine science. Limnology includes the biology, physics, chemistry and ecology of all waters on continents and islands. Thus, limnology includes wetlands, groundwater, ponds and springs. Not all who study inland waters call themselves limnologists; some think of themselves, e.g., as aquatic ecologists, hydrologists, water chemists or wetland scientists. Somewhat counter to intuition, limnology is not the science of freshwaters, *per se*; some inland waters are more saline than the ocean, e.g., the Caspian Sea, the Dead Sea in the Middle East and Salt Creek in Death Valley, CA.

Lakes (Figure 1) and streams (Figure 2) provide many products and services for humankind (Naiman *et al.*, 1995; Postel and Carpenter, 1997). These include the direct uses of inland waters, products harvested from them, and a long list of often unappreciated or under-valued services (Table 1).

Recent general books on limnology include Cole (1994); Horn and Goldman (1994); Lampert and Sommer (1997); Brönmark and Hansson (1998); Giller and Malmqvist (1998); Moss (1998); Boulton and Brock (1999).

LAKE AND RIVER TYPOLOGIES

The classic distinction between lakes and rivers is that lakes are *lentic* (still waters) while rivers and streams are *lotic* (moving waters). Rivers have a shorter water residence time than lakes. All gradations between the two occur

Figure 1 Lago de Nicaragua, Central America. (Photograph by J J Magnuson, May 1985)

Figure 2 The lower Wisconsin River, North America. (Photograph by J J Magnuson, July 1997)

Table 1 Goods and services from freshwater lakes and streams (Naiman *et al.*, 1995; Postel and Carpenter, 1997)

Goods other than water
Fish
Waterfowl
Clams and mussels
Pelts
Streambed minerals (gravel, etc.)
Riparian forests (timber, etc.)
Plants (wild rice, cranberries, etc.)
Services
Recreation (sh, hunt, swim, and boat)
Flood control
Transportation
Natural waste puri cation (human, livestock)
Pollution dilution
Freshwater storage (ice and watersheds)
Biodiversity habitat
Soil fertilization
Enhanced property values
Aesthetics and mental health
Non-user values
Biogeochemical cycling
Water supply
Domestic uses
Livestock
Industrial and fossil fuel power production
Irrigation (agriculture and urban)
Aquaculture
Fire ghting

and we often refer to wide sections of rivers as lakes, e.g., Lake Pepin as part of the Mississippi River. Streams and rivers tend to have a dominating, unidirectional or downstream flow and lakes do not. Lakes, however, are not still waters; wind-driven currents, waves, and mixing typify their physical behavior.

No single or exclusive set of designations characterizes types of lakes or rivers. They can be classified by location, size, chemical and physical features, productivity, shoreline or watershed features, use or modification by humans, position in the landscape, or even the geological processes of formation. Many types are largely self-evident but useful in the eye of the beholder – large or small, tropical or arctic, urban or forested, clear or dark water, bass or trout, etc.

Some classification requires more explanation. *Eutrophic* lakes, owing to high supply rates of essential nutrients, are productive and green with anoxic (without dissolved oxygen) deep waters; *oligotrophic* lakes, owing to low supply rates of at least one major nutrient such as phosphorus (P) or nitrogen (N), are unproductive and clear (Horn and Goldman, 1994). *Mesotrophic* lakes are intermediate. *Dystrophic* lakes are not productive and have stained brown waters owing to the dissolved organic materials from decaying aquatic vegetation in the watershed. Streams have similar differences, but suspended solids often reduce clarity as well.

Rivers are typically viewed in the context of their flow path from headwater streams to the floodplains in the lowlands nearer the ocean (Hauer and Lamberti, 1996). These are often classified into stream orders where the headwater stream is first-order; a stream formed from two first-order streams is second-order; one formed from two second-order streams is third-order. The Mississippi at New Orleans is by this system a 13th order stream. Lakes can also be classified in a similar way (Magnuson and Kratz, 2000).

HABITATS IN LAKES AND RIVERS

Geomorphology provides the basic basin structure for lakes and rivers. Additionally, there is an internal structure based on water temperature and light intensity as well as water flow for streams.

The immediate shoreline, the *riparian* zone, is proportionately longer for streams than lakes when considered relative to the surface water area. Additionally, the *floodplain* is considered part of the riparian zone for streams (Jones and Mulholland, 2000); lakes typically have little floodplain compared to large rivers. Not surprisingly, while important to both lakes and streams, ecological interactions between the riparian and a water body are given more emphasis in stream ecology than in lake ecology. For example, living trees can shade much of a small stream, not so for lakes. Coarse woody debris and fallen dead trees provide structure and habitat for aquatic life but are relatively more significant in streams than lakes.

Lakes contain a number of habitats. The *littoral zone* of lakes is the interface between the shoreline and the open water. It extends from the splash zone on shore, offshore to the depth at which light is still sufficient for growth of rooted aquatic plants and bottom-living algae (Wetzel, 1975). Emergent plants such as reeds and cattails live nearest to the shore; submergent plants live farther offshore; and bottom living algae are at the greatest depths of the littoral zone. The size of the littoral is determined by basin shape as well as by suspended sediments, dissolved color, and dense algal populations in the water column that reduce light penetration and thus the maximum depth of plant growth. Offshore are the open water or *pelagic zone* and the *profundal* or deep bottom water habitat. *Benthic* habitats are the bottom substrate environments in all locations.

Lakes tend to thermally stratify while rivers tend not to. Turbulent flow usually mixes rivers from top to bottom; the work of the wind is responsible for mixing in lakes but often mixing does not extend to the bottom. In summer many lakes thermally stratify into an upper warmer, less dense layer or *epilimnion*; a *thermocline layer* or *metalimnion*; and a deeper cooler, denser layer or *hypolimnion*. A rule of thumb for defining the depth of the thermocline is where temperature changes $>1\,^{\circ}\mathrm{C\,m^{-1}}$ depth in the water column. If a lake becomes ice covered in winter, stratification again occurs but in this case the coldest water floats over less cold water. Winter stratification takes place because the maximum density of fresh water is $4\,^{\circ}\mathrm{C}$. Between summer and winter stratification, lakes mix from top to bottom (turnover). This seasonal pattern of stratification is for a temperate-zone lake; tropical lakes tend to remain stratified all year; Arctic lakes tend not to stratify in summer.

Stream drainage networks are hierarchically nested into valley segments, stream reaches, and channel geomorphic units (Hauer and Lamberti, 1996). Valley segments include colluvial, alluvial and bedrock. Valley sediments in colluvial segments derive from organic and inorganic materials eroded from the adjacent hillslopes. In alluvial valleys, sediments derive from upstream sources and currents are insufficient to scour the valley to bedrock. Bedrock valleys usually have confined valleys and little fill materials. The channel is V-shaped in relatively young mountain catchments with considerable erosive influence while stream channels in the alluvial or deposited terraces of floodplains are complex. Channel reaches in alluvial valleys can be cascades, braids, step pools, plane beds, pool riffles and low gradient meanders. Channel geomorphic units are initially classified by fast water (turbulent and non-turbulent) and slow water (scour pools and dammed pools). Many sub-units exist such as falls, cascade, rapids, riffle and chute for turbulent fast water; sheet and run for non-turbulent fast water; eddy, trench, mid-channel, convergence, lateral and plunge for scour pools; and debris, beaver, landslide, backwater and abandoned channel for dammed pools.

In streams, and lakes as well, a distinctive biota lives in the zone below and adjacent to open waters where mixtures of ground and surface water occur (Cole, 1994; Hauer and Lamberti, 1996). For streams, this hyporheic zone is an important habitat for organisms and for hydrologic interactions with the surface stream with complex influences on nutrient cycling. The psammolittoral zone of lakes exists in the interstitial waters within sand beaches.

BIOTA IN LAKES AND RIVERS

Biota of lakes and rivers occur in and on the bottom or benthic zone, in the water column, floating on the surface, and in the hyporheic and psammolittoral zones. Primary producers make new organic carbon; they include phytoplankton suspended in the water column, macrophytes and benthic algae in the littoral zone, and chemosynthetic and photosynthetic bacteria. Many bacterial groups are involved in biochemical transformations of nitrogen, sulfur (S), and other elements (*see* **Biogeochemical Cycle**, Volume 2). Primary consumers include many invertebrate taxa, such as microscopic crustaceans like water fleas, immature aquatic insects and snails, plus some vertebrates. Secondary consumers are dominated by fishes, but include other vertebrates such as birds and many invertebrates such as immature and adult aquatic insects. Tertiary consumers are typically large piscivorous fishes, but include birds and mammals such as bears. Humans sit at the top of many inland water food webs.

Aquatic organisms often are treated as functional groups or guilds owing not only to their role in ecosystems, but also to their high species diversity, the taxonomic complexity of food webs, and an inability to identify and count all players. In forest streams, animals are considered by

their source of food and feeding mode. Scrapers – scrape algae and other organisms from the surfaces of rocks and dead wood; shredders – shred fallen tree leaves, rich with bacteria and fungi, leaving only the veins of the leaf and fine particulate material; gatherers – gather the fine particulates, and predators eat the scrapers, shredders and gatherers. In the open waters of lakes and ponds, organisms are grouped by life style and strength of swimming. *Phytoplankton* are small unicellular plants suspended in the water column; *zooplankton* are small weak-swimming animals carried by the currents; *nekton* are the strong swimming fishes, birds, and reptiles; and *neuston* are the diverse organisms living on the water surface film. Fishes can be *herbivores*, *zooplanktivores*, *zoobenthivores*, and *piscivores*. Nitrogen-fixing algae convert dissolved nitrogen gas (N_2) into usable nitrates; denitrifying bacteria convert nitrate back to nitrogen gas. Aquatic plants are grouped into emergent, submergent, and floating groups, each with different sets of adaptations and ecological roles.

CONCEPTS OF LAKE AND RIVER STRUCTURE AND FUNCTION

Biogeochemical Cycling

Biogeochemical cycling includes the major elements such as the important nutrients, phosphorus and nitrogen, and industrial chemicals such as polychlorinated biphenyls (PCBs), and biocides such as dichlordiphenyltrichloromethane (DDT). Compounds are transformed by biological and chemical processes, transported by movement of water and biota, deposited or stored in sediments. Some such as DDT and PCBs accumulate in food webs with harmful biological effects.

Major biogeochemical cycles are greatly influenced by whether dissolved oxygen is present and by pH or acidity. These conditions affect solubility of compounds and thus biotic availability; they also determine which biochemical processes can occur. Below, we consider phosphorus, nitrogen, and mercury (Hg).

In anoxic sediments, hyporheic zones, or hypolimnia, iron phosphates are soluble and the phosphorus can be recycled back into the biota. If dissolved oxygen is present, then the iron phosphates are insoluble, precipitate, and accumulate in the sediments. This process can intensify the negative influence of excess production, because excess production leads to anoxic conditions and the release of phosphorus that, in turn, can lead to more production.

Cycling of nitrogen by bacteria goes from nitrite to nitrate in oxygenated conditions, but from nitrate to nitrite in anoxic conditions and ultimately to nitrogen gas. In river floodplains, water covered sediments become anoxic and nitrate is transformed to nitrogen gas and released to the atmosphere. If the sediments are dry or oxygenated, nitrates persist and are available for transport downstream and increased biological productivity during a flood event. This process contributes to the movement of nitrate from major floodplain rivers to coastal marine waters where excess production and anoxia can occur owing to these transported nutrients (Naiman *et al.*, 1995; Turner and Rabalais, 1994).

Biogeochemical cycling of nutrients is conceptualized as nutrient cycling in lakes and nutrient spiraling in streams (Wetzel, 1975; Allen, 1995; Essington and Carpenter, 2000). In lakes, nutrients cycle between the biota and the water and can accumulate in bottom sediments. In streams, nutrients cycle from the biota to the water, but in so doing move progressively downstream with each exchange; thus there is a visual analogy of a nutrient spiraling downstream. Differences in the export rates of a nutrient to the sediments of lakes or downstream in streams helps conceptualize the importance of export in determining productivity in the two systems. In lakes the number of times a nutrient molecule is sequestered by producers before export is governed by processes that promote mineralization (*see also* **Nitrogen Mineralization**, Volume 2) and retard particulate sedimentation; in streams it is governed by processes that promote uptake and retard downstream loss of dissolved nutrients.

Mercury is deposited on the surface of the waters by atmospheric deposition (Porcella *et al.*, 1995). About half of this mercury is from the burning of coal. The form of mercury taken up by the food web is methyl mercury (CH_3Hg^+). Bacterial processes that form methyl mercury thrive in more acidic waters of pristine headwater lakes where contamination of fishes reaches levels unhealthy to humans.

Contaminants such as PCBs and DDT that are soluble in lipids bioaccumulate in food webs by about a factor of 10 for each step upwards. Concentrations in the apex predators such as fishes are thus greatly biomagnified over levels in water (Stow *et al.*, 1995).

Productivity

Productivity of lakes and streams has long been of interest because people use these systems for food and fiber on the one hand and because high productivity produces limnological conditions judged to be unfavorable to humans on the other hand. Freshwater and inland fisheries yielded 7.7 million metric tonnes of fishes, amphibians, crustaceans, and mollusks in 1997 (FAO, 1999). Another 17.7 million tonnes came from aquaculture, much of which also comes from inland waters. Eutrophic lakes have noxious algal growths that not only are unsightly and smell bad, but also result in the loss of dissolved oxygen needed for the survival of coldwater fishes in the hypolimnia of lakes. Excessive nutrient from sewage, or agriculture and industrial inputs to rivers also produce putrefaction, loss of dissolved oxygen, and loss of fishes.

Productivity varies greatly among waters, depending on morphometric (size and depth), edaphic (watershed soil and nutrients), climatic (precipitation and temperature), and anthropogenic (fertilization and runoff) factors. Primary production in lakes generally increases with nutrient input or loading and decreases with increased depth (Horn and Goldman, 1994). With the same input of nutrients, deeper lakes tend to be more oligotrophic in character than shallower lakes. This is because dilution is greater in larger lakes, and because a greater store of oxygen exists in the large hypolimnion that is less likely to be depleted by the decomposition of phytoplankton as they sink and die.

In inland waters, primary productivity is usually limited by phosphorus (Horn and Goldman, 1994). However, in some regions where phosphorus is naturally abundant or where phosphorus is a major pollutant, limitation by nitrogen is common. In a lake, silicon (Si) usually becomes limiting for diatom growth by mid-summer because sinking diatoms with their heavy siliceous frustules have sunk into the hypolimnion carrying their silicon with them. In some systems even rare elements like molybdenum can limit productivity through their influence on nitrogen fixation by plants in the watershed. Production of benthic algae in some streams can be patchy owing to the patchy nature of nutrient upwelling from the hyporheic zone (Jones and Mulholland, 2000).

Consumer Driven Dynamics

While food webs and productivity are driven by nutrient and food resources, the influence of consumer-driven processes is equally important (Carpenter, 1988; Carpenter and Kitchell, 1988; Allen, 1995). In lakes, about half of the interannual variability in phytoplankton production is attributable to variation in nutrient availability and the other half to variability in the abundance of consumers. Herbivorous zooplankton can graze on and deplete the phytoplankton. Zooplanktivorous fishes can feed on and deplete the zooplankton. As a consequence, the presence of zooplanktivorous fishes in floodplain ponds can result in green water, the absence of fish to clear water. Alewife in small lakes can result in the loss of large zooplankton and the replacement with small zooplankton that are less effective herbivores on the phytoplankton. The visible influence of consumers on their prey is lessened in streams where prey immigration rates are high.

Consumer-driven changes are perhaps most apparent when a large change in abundance of a consumer is brought about by stocking of game fishes, the invasion of an exotic species, or by the catastrophic loss of fishes such as with summerkill or winterkill events resulting from low dissolved oxygen. Biomanipulation (Hansson *et al.*, 1998) is the developing management practice of reducing excess phytoplankton by stocking predaceous fishes that reduce the abundance of zooplanktivorous fishes that, in turn, increases the abundance of zooplankton that eat phytoplankton. Similarly, invasions or introductions of exotic crayfish can forage on and eliminate many macrophytes; invading zebra mussels can filter massive quantities of water and reduce the abundance of small phytoplankton and indirectly cause noxious blue/green algal blooms; predaceous fishes can eliminate smaller species of fishes. The perceived value of such changes ranges from positive to negatively catastrophic.

Commercial, artisanal, and recreational fishers can overfish lake and river populations. Over-fishing as a consumer-driven process, results in a reduction in the abundance of targeted fish species and a reduction in their average body size with subsequent reverberations down the food web. Additionally, fisheries increase the probability of introductions of harvested or exotic species by fish managers and escapees from aquaculture, bait production, and the aquarium trade.

Climate

The balance between precipitation and evapotranspiration determines the very existence of inland water ecosystems. Where evaporation greatly exceeds precipitation, saline waters form in lowland areas. Dilute waters characterize upland areas when precipitation exceeds evapotranspiration. The flow of water in streams, the annual flood and drought, and contributions of ground water are determined by the climate acting on the local landscape and geomorphology.

Lakes and streams are warmed by solar radiation and warmed and cooled by heat transfer by thermal conduction, evaporation, and hydrologic flow. Winds drive internal dynamics of physical mixing working against the density gradients resulting from heating near the surface. Cloudiness can shade the waters from direct sun.

Lakes and streams are highly responsive to climatic variability and warming (Arnell *et al.*, 1996; Cushing, 1997). As a class of ecosystems, they are vulnerable to climate change (Gitay *et al.*, 2001). The most vulnerable components are the loss of lake and river ice, habitat loss for cold-water fishes, increases in poleward invasions and subsequent extinctions of biota, and the exacerbation of existing problems with eutrophication, toxic pollutants, and water use.

Regional Ecology

Lakes appear as patches, and streams as dendritic patterns on the landscape. The heterogeneity among lakes and among stream reaches is organized across the landscape in respect to the hydrologic flow path. The basic idea, and a tenet from landscape ecology, is that position matters.

For streams, the stream continuum concept portrays changes in the dominance of particular ecological processes

from the headwaters downstream to the large floodplain rivers (Vannote *et al.*, 1980; Johnson *et al.*, 1995). In headwaters of a forested stream, for example, structure and function are dominated by groundwater inputs and shading by riparian trees; the food web is driven by organic inputs of leaves and branches from the riparian vegetation. Thus, upstream waters are cooler than lower reaches, terrestrial primary production controls stream productivity, and a food web dominated by shredders and gatherers exists. In mid-reaches, streams widen and the shallow flowing waters are increasingly energized directly by the sun as the canopy opens. Thus, waters warm and instream primary production from benthic algae and macrophytes and consumption by herbivores dominate. In the lower reaches of large rivers, fine organic particulates and sediments enter from the floodplain. The waters warm; sunlight for instream photosynthesis is low because the stream is deeper and the light is absorbed and scattered by suspended materials. Gatherers and filterers dominate the primary consumption in the food web. Thus, there is a continuum of energy inputs and a continuum of energy processing from *heterotrophic* in the headwaters, to *autotrophic* in mid-reaches, to heterotrophic again in the lower reaches.

Viewing a river ecosystem in the continuum context provides a concept of the whole and helps eliminate a road bridge view of how rivers function. Additionally, it emphasizes the importance of upstream and lateral influences.

For lakes, the concept of position in the landscape portrays a sequence of structure and function with clear analogies to the stream continuum (Magnuson and Kratz, 2000; Riera *et al.*, 2000). Headwater lakes tend to be small and have greater inputs of rainwater relative to groundwater or overland flow, while lakes lower in the flow system tend to be large and have more inputs from ground water and tributary streams. Thus, upland lakes tend to have lower concentrations of dissolved ions and lower species richness of fishes. In some regions, many other characteristics map on this template such as human dwellings, lake shape, invertebrate distribution, chlorophyll concentrations, etc.

Other landscape considerations important to lakes and streams include interaction with the watershed, the flood pulse concept, ecological interactions that cross the boundary between lakes and streams, and island biogeography (Magnuson and Kratz, 2000). Considering a lake or stream interacting with its catchment is clearly a landscape approach that recognizes the strong interaction between two major dissimilar patches on the landscape – water body and adjacent land. Owing to changes in land use, this is a particularly significant consideration in any discussion of future global environmental changes of lakes and streams.

In flood plain rivers, the flood pulse onto the land is incorporated into the functioning of river ecosystems and constitutes an important land-water interaction (Bayley, 1995). In this sense, the flood is not a disturbance to a river ecosystem, but a process that contributes to productivity in the river and its floodplain. Organisms and biogeochemical cycles are adapted to the progressive advance and retreat of the waters. For example, young and adult fishes disperse onto the floodplain to utilize rich foraging areas where a progression of nutrient inputs, production of aquatic and terrestrial plants, and decomposition of this production occurs over the flood pulse. River channels are dynamic, and primary succession and agriculture occur on recently deposited sediments.

Biodiversity

Biodiversity of inland waters is high. For example, of the more than 28 000 known fish species, 41% are freshwater and 58% marine (Moyle and Cech, 1996), even though the oceans are much larger – about 70 times by area and 10 000 times by volume. In lakes formed in ancient times, species flocks of fishes, algae, and many invertebrates have evolved with many endemic species unique to that particular water. Individual rift valley lakes in East Africa contain species flocks of almost 250 cichlid species. In Lake Baikal in Russia 35% of the plants and 65% of the animals are endemic (Burgis and Morris, 1987). Individual streams have high species diversity; the Breitenbach, a small German stream has at least 1044 invertebrate species, including 476 dipteran species, 125 nematodes, and 106 rotatorians (Allen, 1995). As many as 250 species of diatoms are known from individual species-rich streams in the US.

Isolation of lakes and headwater streams is a primary explanation of the high diversity both in individual waters and across a region. Species composition depends on a range of regional to local factors (Tonn *et al.*, 1990). The number of species increases with the area of a lake (Barbour and Brown, 1974) and the drainage basin area of streams (Sepkoski and Rex, 1974), largely owing to the increase in habitat diversity in larger systems. The number of species is difficult to determine and numbers found is greatly dependent on sample size (Arnott *et al.*, 1998).

Because humans have increased both invasion and extinction rates, the species turnover rates of inland waters have apparently increased (Magnuson, 1976). Species composition and richness in more isolated waters may be more controlled by adaptive radiation and local extinction than by invasion. The more isolated a water body is from invading colonists, the more endemism will result for older lakes and streams. More relic flora and fauna will persist in more isolated waters as dispersal brings warmer-loving colonists poleward. In more open or connected waters, species composition and richness may be controlled more by invaders and their interactions with native organisms. Invasion rates, owing to transplantation by humans, have been occurring at a rapid rate and floras and faunas are tending towards spatial homogeneity (Rahel, 2000).

THREATS IN A CHANGING GLOBE

Both the functioning of lake and stream ecosystems and the supply of human products and services are sensitive to anthropogenic change. Responses will not only come from direct changes to the waters themselves, but also from changes to the riparian zone and the watersheds as well as the air and climate. Effects of global environmental change on land will be accentuated downward across the landscape in lakes and streams. Many of these systems are already damaged with present levels of population and development.

Changes of concern to the function and value of lakes and streams include increases or changes in land use, climate (e.g., falling lake levels), biological invasion rates, toxic and nutrient release or redistribution, food production and aquaculture, hydrologic engineering structures for water regulation, and energy production and use. Each of these general topics includes a rich array of concerns. Land use, for example, includes changes in agricultural lands and practices, urbanization and suburbanization, shoreline modification, and altered vegetative cover. Hydrologic engineering structures include dams, dykes, channelization, instream flow regulation structures, bulkheads, pilings and walls.

Change will occur. Not all will negatively influence lakes and streams, many will. A key is to understand how these changes will influence the functioning of inland aquatic ecosystems and the products and services they provide. There is also an increasing awareness that restoration or rehabilitation of ecosystem functions can help maintain and recover the function and the goods and services that inland waters can provide.

REFERENCES

Allen, J D (1995) *Stream Ecology, Structure and Function of Running Waters*, Chapman and Hall, London, 388.

Arnell, N, Bates, B, Land, H, Magnuson, J J, Mullholland, P I, Fischer, S, Lui, C, McKnight, D, Starosolov, O, and Taylor, M (1996) Hydrology and Freshwater Ecology, Climate Change 1995 – Impacts, Adaptations and Mitigations of Climate Change: Scientific and Technical Analysis. *Contributions of Working Group 2 to the Second Assessment Report of the IPCC Working Group 2*, Chapter 10, eds R T Watson, M C Zinyowera, and R H Moss, Cambridge University Press, Cambridge, 325–363.

Arnott, S E, Magnuson, J J, and Yan, N D (1998) Crustacean Zooplankton Species Richness: Single and Multiple Year Estimates, *Can. J. Fish. Aquat. Sci.*, **55**, 1573–1582.

Bayley, P B (1995) Understanding Large River-Floodplain Ecosystems, *BioScience*, **45**, 153–158.

Barbour, C D and Brown, J H (1974) Fish Species Diversity in Lakes, *Am. Nat.*, **108**, 473–489.

Boulton, A J and Brock, M A (1999) *Australian Freshwater Ecology, Processes and Management*, Gleneagles, Australia, 300.

Brönmark, C and Hansson, L-A (1998) *The Biology of Lakes and Ponds*, Oxford University Press, Oxford, 216.

Burgis, M J and Morris, P (1987) *The Natural History of Lakes*, Cambridge University Press, Cambridge, 218.

Carpenter, S R, ed (1988) *Complex Interactions in Lake Communities*, Springer, New York, 283.

Carpenter, S R and Kitchell, J F (1988) Consumer Control of Lake Productivity, *BioScience*, **38**, 764–796.

Cole, G A (1994) *Textbook of Limnology*, 4th edition, Waveland Press, Prospect Heights, IL, 412.

Cushing, C E, ed (1997) *Freshwater Ecosystems and Climate Change in North America: A Regional Assessment*, John Wiley & Sons, New York, 262.

Essington, T E and Carpenter, S R (2000) Nutrient Cycling in Lakes and Streams: Insights From a Comparative Analysis, *Ecosystems*, **3**, 131–143.

FAO (1999) *Fisheries Statistics – Capture Production*, 1997, Vol. 84, FAO Fisheries Series No. 52, FAO Statistics Series No. 147, Rome.

Giller, P S and Malmqvist, B (1998) *The Biology of Streams And Rivers*, Oxford University Press, Oxford, 296.

Gitay, H, Brown, S, Easterling, W, Jallow, B, with Antle, J, Apps, M, Beamish, R, Chapin, T, Cramer, W, Frangio, J, Laine, J, Magnuson, J, Noble, I, Price, J, Prowse, T, Sirotenko, O, Root, T, Schulze, E-D, Sohngen, B, and Soussana, J-F (2001) *Ecosystems and Their Goods And Services. Climate Change: Impacts, Adaptation, and Vulnerability*, Chapter 5, IPCC Working Groups II's Contribution to the Third Assessment Report.

Hansson, L-A, Annadotter, H, Bergan, E, Hamrin, S F, Jeppesen, E, Kairesalo, T, Luokkanen, E, Nilsson, P-A, Søndergaard, M, and Strand, J (1998) Biomanipulation as an Application of Food-Chain Theory: Constraints, Synthesis, and Recommendations for Temperate Lakes, *Ecosystems*, **1**, 558–574.

Hauer, F R and Lamberti, G A, eds (1996) *Methods in Stream Ecology*, Academic Press, San Diego, CA, 673.

Horn, A J and Goldman, C R (1994) *Limnology*, 2nd edition, McGraw-Hill, New York, 576.

Johnson, B J, Richardson, W B, and Maimo, T J (1995) Past, Present, and Future, Concepts in Large River Ecology, *BioScience*, **45**, 134–141.

Jones, J B and Mulholland, P J (2000) *Streams and Ground Waters*, Academic Press, New York, 425.

Lampert, W and Sommer, U (1997) *Limnoecology, the Ecology of Lakes and Streams*, Translated by J F Haney, Oxford University Press, Oxford, 382.

Magnuson, J J (1976) Managing with Exotics – a Game of Chance, *Trans. Am. Fish. Soc.*, **105**, 1–9.

Magnuson, J J and Kratz, T K (2000) Lakes in the Landscape, Approaches to Regional Limnology, *Verh. Int. Ver. Limnol.*, **27**, 74–87.

Moss, B (1998) *Ecology of Fresh Waters, Man and Medium, Past and Future*, 3rd edition, Blackwell Science, Oxford, 557.

Moyle, P B and Cech, Jr, J J (1996) *Fishes, an Introduction to Ichthyology*, Prentice-Hall, Upper Saddle River, NJ, 590.

Naiman, R J, Magnuson, J J, McKnight, D M, and Stanford, J A (1995) *The Freshwater Imperative, a Research Agenda*, Island Press, Washington, DC, 165.

Porcella, D, Huckabee, J, and Wheatley, B (1995) Mercury as a Global Pollutant, *Water, Air, Soil Pollut.*, **80**(nos. 1–4).
Postel, S and Carpenter, S (1997) Freshwater Ecosystem Services, in *Natures Services, Societal Dependence on Natural Ecosystems*, ed G C Daily, Island Press, Washington, DC, 195–214.
Rahel, F J (2000) Homogenization of Fish Faunas across the US, *Science*, **288**, 854–856.
Riera, J L, Magnuson, J J, Kratz, T K, and Webster, K E (2000) A Geomorphic Template for the Analysis of Lake Districts Applied to the Northern Highland Lake District, Wisconsin, USA, *Freshwater Biol.*, **43**, 301–318.
Sepkowski, Jr, J J and Rex, M A (1974) Distribution of Freshwater Mussels: Coastal Rivers as Biogeographic Islands, *Syst. Zool.*, **23**, 165–188.
Stow, C A, Carpenter, S R, Madenjian, C P, Ebby, L A, and Jackson, L J (1995) Fisheries Management to Reduce Contaminant Consumption, *BioScience*, **45**, 752–758.
Tonn, W M, Magnuson, J J, Rask, M, and Toivonen, J (1990) Intercontinental Comparison of Small Lake Fish Assemblages: the Balance Between Local and Regional Processes, *Am. Nat.*, **136**, 345–375.
Turner, R E and Rabalais, N N (1994) Coastal Eutrophication near the Mississippi River Delta, *Nature*, **368**, 619–621.
Vannote, R L, Minshall, G W, Cummins, K W, Sedell, J R, and Cushing, C E (1980) The River Continuum Concept, *Can. J Fish. Aquat. Sci.*, **37**, 130–137.
Wetzel, R G (1975) *Limnology*, W B Saunders, Philadelphia, PA, 743.

Land-Ocean Interactions in the Coastal Zone (LOICZ)

see LOICZ (Land- Ocean Interactions in the Coastal Zone) (**Volume 2**)

LANDSAT

see Remote Sensing, Terrestrial Systems (**Volume 2**)

Landscape Ecology

Traditionally, ecologists have focused on the study of individual ecosystems. *Landscape ecology* is the study of the ecology of a large geographic area (or landscape) as a unit. The focus is on the spatial arrangement of different kinds of ecosystems (e.g., swamps, fields, ponds, roads, and woods) and how these arrangements affect the movement of animals, plants, nutrients, and energy across landscapes. Landscape ecologists look at the spatial structure of diverse ecosystems, their function and interaction, and how they change over time.

Landscape ecology moves beyond measures that represent averages over the many different patches in an area and considers where patches are located relative to one another and how particular variables differ in value from one specific location to another. Landscapes are distinguished by properties of their constituent patches (i.e., patch quality) and by how their constituent patches interact. These interactions depend not only on the spatial arrangement of the patches, but also on the behavioral responses of organisms to landscape structure, or the properties of spatial flows of materials and spread of disturbances (e.g., fire, insect outbreaks, disease epidemics) through landscapes. Both landscape pattern and the reaction of organisms or physical processes to it depend on spatial scale, so spatial scale is important. The appropriate scales at which to study landscapes depend on the causal mechanisms behind the reactions of organisms or systems to landscape structure. Natural resource management has usually concentrated on particular blocks of habitat or species of interest, and management practices have been conducted on the hectare-to-kilometer scales corresponding to human land use. Landscape ecology stresses the need to consider entire landscape mosaics over a broad range of scales to gain a fuller understanding of how spatial patterns and processes interact to produce ecological dynamics.

RICHARD A FLEMING *Canada*

Large Scale Biosphere-Atmosphere Experiment in Amazonia (LBA)

Diane E Wickland[1], Michael Keller[2], Carlos Nobre[3], and Pavel Kabat[4]

[1] *National Aeronautics and Space Administration, Washington, DC, USA*
[2] *United States Forest Service, USA*
[3] *Brazilian Space Research Institute, Brazil*
[4] *Wageningen University and Research Centre, The Netherlands*

The Brazil-led LBA is an international, interdisciplinary research program designed to address the climatological, ecological, biogeochemical, and hydrological functioning of

Amazonia, the impact of land use change on these functions, and the interactions between Amazonia and the entire Earth System. LBA's key questions are:

- *How does Amazonia currently function as a regional entity?*
- *How will changes in land use and climate affect the biological, chemical and physical functions of Amazonia, including the sustainability of development in the region and the influence of Amazonia on global climate?*

Tropical forests in Amazonia are being cleared and developed for agriculture at unprecedented rates. Road building, followed by immigration of farmers and fiscal incentives in Brazil for cattle ranching, resulted in this sharp increase in deforestation. Deforestation and agricultural land use are changing the cycling of water, energy, carbon, and nutrients and are expected to have climatic and environmental consequences at local to global scales. Further, some agricultural practices have not succeeded, and the long-term sustainability of development in Amazonia is a concern. To understand the consequences, to plan mitigation efforts, and to predict future conditions and the fate of agricultural development, enhanced knowledge is needed of Amazonian system functioning.

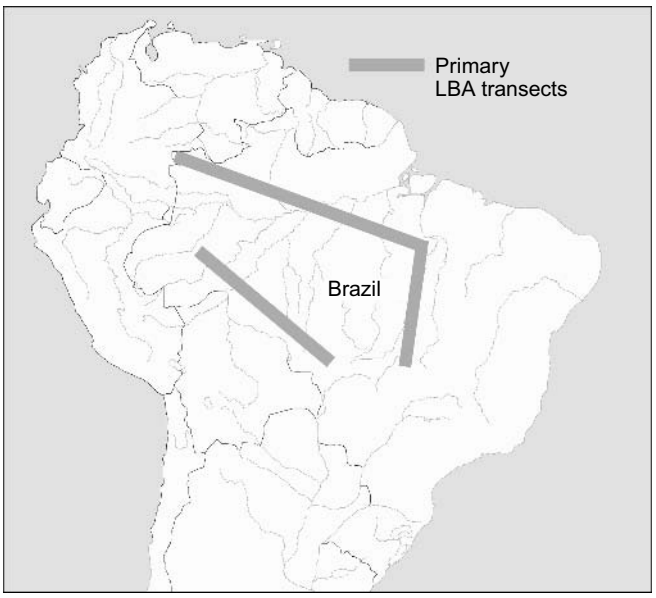


Figure 1 The two LBA transects are delineated on this map of Northern South America. Amazonia lies predominantly within Brazil, under the direct in uence of the Amazon and Tocantins Rivers, but includes portions of Bolivia, Peru, Ecuador, Columbia, Venezuela, Guyana, Surinam, and French Guiana

The LBA's biological and ecological research focuses on the effects of tropical forest conversion, regrowth, and selective logging on carbon storage, nutrient dynamics, trace gas fluxes, and the prospect for sustainable land use in Amazonia (Houghton, 1997; Fernandes *et al.*, 1997; Richey *et al.*, 1997; Keller *et al.*, 1997; Nepstad *et al.*, 1997). Conversion of undisturbed tropical forests to agriculture or to regrowth, combined with the effects of fire and selective logging, are studied in relation to their effects on biogeochemical cycles. Quantifying Amazonia's carbon balance, nutrient pools and trace gas fluxes is emphasized. Land use is studied in the context of both its biophysical and its socioeconomic causes, and land cover changes throughout Amazonia are documented.

LBA employs a multi-scale approach that considers the effects of forest conversion at scales ranging from the local site-level scale, to that of the entire Amazonian region. It started in 1998 and will continue till 2003, with observations running for 3–5 years to capture seasonal and interannual variations. Its field research is carried out along two transects (Figure 1), spanning gradients from wet, aseasonal forests to relatively dry and seasonal forests or savannas. Continuous or periodic observations of a core set of measurements (e.g., carbon dioxide, water, and trace gas fluxes; trace gas concentrations; micrometeorological conditions; vegetation and soil properties) are made along the transects in predominant land uses for that location. Time series of remote sensing observations are compiled for past and current satellite sensors. The local-scale measurements and results of process studies are combined with the remotely sensed data and other geographically extensive data sets to derive regional-scale understanding. Models play a key role in testing understanding, extending it across scales, and predicting future responses.

If the past is a guide, pressures to provide for a growing human population and plans for expanded commerce will spur future development that will lead to increasing deforestation and a continuing search for options for sustainable land use. LBA has a substantial education and training component dedicated to building local capacity, fully engaging Amazonian scientists in LBA research, and training people within the region to be prepared to address such global change challenges.

See also: **IGBP Terrestrial Transects**, Volume 2.

REFERENCES

Fernandes, E C M, Biot, Y, Castilla, C, do C Canto, A, Matos, J C, Garcia, S, Perin, R, and Wanderli, E (1997) The Impact of Selective Logging and Forest Conversion for Subsistence Agriculture and Pastures on Terrestrial Nutrient Dynamics in the Amazon, *Ciênc. Cult.*, **49**, 34–47.

Houghton, R A (1997) Terrestrial Carbon Storage: Global Lessons for Amazonian Research, *Ciênc. Cult.*, **49**, 58–72.

Keller, M, Melillo, J, and de Mello, W Z (1997) Trace Gas Emissions from Ecosystems of the Amazon Basin, *Ciênc. Cult.*, **49**, 87–97.

Nepstad, D C, Klink, C, Uhl, C, Vieira, I C, Lefebvre, P, Pedlowski, M, Matricardi, E, Negreiros, G, Brown, I F, Amaral, E, Homma, A, and Walker, R (1997) Land-use in Amazonia and Cerrado of Brazil, *Ciênc. Cult.*, **49**, 73–86.

Richey, J E, Wilhelm, S R, McClain, M E, Victoria, R L, Melack, J M, and Araujo-Lima, C (1997) Organic Matter and Nutrient Dynamics in River Corridors of the Amazon Basin and their Response to Anthropogenic Change, *Ciênc. Cult.*, **49**, 98–110.

Life Cycles

Richard A Fleming,[1] Robert L Fleming,[1] and Ian A Fleming[2]

[1] *Great Lakes Forest Research Centre, Sault Ste. Marie, Canada*

[2] *Norwegian Institute for Nature Research, Trondheim, Norway and Coastal Oregon Marine Experiment Station, Oregon State University, Newport, OR, USA*

Although no generally agreed definition of life exists, various processes associated with living organisms often help distinguish them from inert matter. These processes include growth, or its potential; reaction to stimuli in the immediate environment; passive or active movement; metabolism through which matter and energy are transformed into the organization of organic compounds unique to each individual; and reproduction.

As organisms grow and age, most go through various developmental stages in which some of these processes, or certain aspects of them, are particularly emphasized. For instance, a butterfly began life as an egg and later emerged as a larva, the eating-for-growth stage. The larva eventually pupated and underwent metamorphosis before emerging as a butterfly. Only as a butterfly can the insect mate, reproduce, and thus complete its life cycle. A species' life cycle is thus the sequence of developmental stages that individuals pass through, going from the start of one stage to the start of that same stage in the next generation.

Animals and many simple organisms complete their life cycle in a single generation. For instance, individuals of various species of protists (microorganisms which have both plant and animal characteristics) and bacteria originate from the fission of a single existing individual, grow to maturity, and then complete their life cycle by dividing into two new individuals. In more 'advanced' animal species, the individual originates from the fusion of male and female sex cells (gametes), grows to reproductive maturity, produces gametes, and then completes the life cycle through successful mating.

Most plants, and some fungi and protists, have life cycles with alternating generations. Individuals originate from germinating spores, which grow into gamete-producing organisms. Fertilized gametes can grow into spore-producing organisms. Spore germination completes the life cycle.

Life cycles provide frameworks for exploring how different species might cope with global change. Global change is occurring because the scale of human activities now exceeds that of all other species and affects biospheric fluxes of matter and energy. Global change encompasses the direct and indirect effects of changes in atmospheric composition, biological diversity, land use, and climate. Since different stages of a life cycle often have different functions (e.g., growth, reproduction), and requirements (e.g., temperature, habitat, food) to fulfil those functions, different aspects of global change will likely have different, possibly even compensatory, impacts over a complete life cycle. Thus, one needs to consider how each component of global change might affect each life stage in forecasting how a species might respond to global change as a whole. This approach is illustrated with three very different life forms: the spruce budworm, a naturally outbreaking forest insect; wild salmon, a group of fishes which spend parts of their life cycle in freshwater and parts in saltwater; and black spruce, an ubiquitous tree species of North America's boreal forest.

RESPONSES OF THE SPRUCE BUDWORM TO GLOBAL CHANGE

The spruce budworm, *Choristoneura fumiferana* (Lepidoptera: Tortricidae), the dominant insect defoliator of North America's coniferous forests, has been outbreaking on a roughly 35 year cycle since before the first Europeans arrived on the continent. Its outbreaks last 10–15 years and during outbreaks spruce (*Picea* species) and balsam fir (*Abies balsamea*) are often killed over vast areas. In its life cycle (Figure 1), the spruce budworm has one generation a year. In mid-August, 2–3 weeks after the eggs are laid and most adult moths have died, the first instar (stage immediately prior to shedding the first skin) larvae hatch and move to overwintering sites on the branches. They overwinter as small, second instar larva and become active again in early May. Then they pass through four heavily feeding instars before pupating in late June. In July, emerged moths mate, disperse, and lay eggs. The spruce budworm's potential for causing tree mortality stems from its rapid growth and high fecundity (realized fertility) of 170–200 eggs per female. A mature sixth instar larva weighs approximately 100 mg (fresh weight) and is roughly 20–30 mm long. This is about 1500 times larger than it was as a second instar larva just 6–9 weeks earlier.

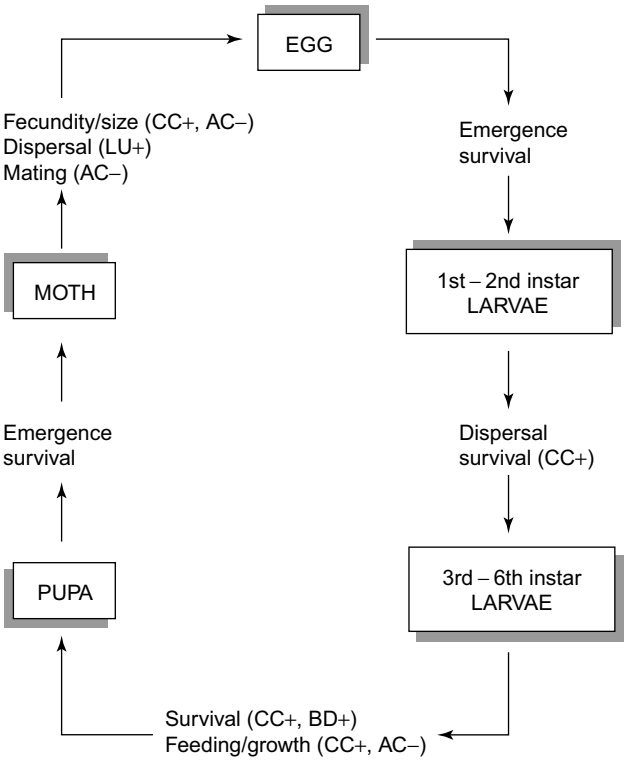


Figure 1 Generalized life cycle of the spruce budworm. Boxes enclose the major stages. The principal functions and processes needed to progress from stage to stage are designated by phrases interrupting the directing arrows. Letters in parentheses indicate components of global change: AC, atmospheric chemistry; BD, biodiversity; CC, climate change; or LU, land use, which the text describes as potentially in uencing the function or process. A plus or minus indicates whether the in uence is expected to be largely positive or negative for the insect

Global change, particularly climate change, may affect insect populations directly through their *per capita* growth rates or indirectly through interactions and feedbacks with other species and abiotic components of the environment. These interactions and feedbacks often depend on delicately tuned phenological relationships. For instance, the spruce budworm, like many other herbivores, has synchronized its greatest nutritional demand (the heavily feeding larval stages) with the time when developing, rather then mature, host plant foliage is most available. Developing is often better for herbivores than mature plant tissue because it is lower in fibre (which can limit digestibility), higher in nitrogen (nutritional value), and lower in secondary metabolites (defensive chemicals). In a warmer climate, foliage is generally expected to develop more quickly, thus diminishing the time interval when herbivores can find the best foliage. Since cold-blooded animals such as insects also develop faster and feed more actively at higher temperatures, however, the net effect is uncertain. Many insect herbivores may remain well synchronized in phenological development with that of their host plants' foliage, even in changing climates. On the other hand, increased atmospheric CO_2 concentrations may allow host plant foliage to accumulate secondary metabolites fast enough that herbivory becomes inhibited relatively earlier in development, and this may work against some insects.

Under climate change, the southwestern part of the spruce budworm's range can expect drier conditions with the increasing likelihood of drought. Warmer, drier conditions directly increase spruce budworm reproductivity and larval survival. Indirect effects are also important. The vulnerable larval stages can escape many natural enemies (e.g., the fungal pathogen *Erynia (Zoophthora) radicans*) because larvae develop faster at high temperatures than at normal temperatures, and this reduces the length of their exposure. Furthermore, larvae develop more quickly at high temperatures than many of their natural enemies. As many of these natural enemies have alternate hosts, they are less closely synchronized phenologically with the spruce budworm than is the budworm with its host trees. Hence, some phenological de-synchronization of the spruce budworm with its natural enemies is expected as climate change progresses. In summary, warmer, drier conditions tend to weaken the regulating effects of natural enemies and simultaneously increase reproductivity and larval survival. Thus, the expectation for longer or more frequent outbreaks in warmer, drier climates.

Protracted trends toward warmer climates will allow many successive generations of directed natural selection, particularly for organisms like some insects (e.g., aphids) with many generations per year. (Increased climatic variability may disrupt this process frequently for short periods but seems unlikely to forestall it over the long term). Consequently, genotypes best suited to warmer environments are expected to become increasingly common, resulting in genetically adapted populations. Experimental and observational evidence suggests that such adaptations may produce increases in the rates of phenological development; increases in the number of generations per year; shifts in geographical and also possibly host range; changes in morphology, physiology, and reproductive strategy; and possibly even speciation or extinction.

Other aspects of global change could affect spruce budworm outbreak patterns but these are more speculative. For instance, biodiversity loss manifested as a lack of suitable plant habitat or lack of alternate hosts could have serious consequences for certain natural enemies. Up to a point, this loss may be compensated by other natural enemies so that there is little overall effect on the spruce budworm's outbreak cycle. In central parts of the range, land use changes, manifested as increased wildfire control and harvesting, allow the spruce budworm's host tree species to capture ever increasing proportions of the forest. The resulting increase in feeding and egg-laying sites

lends itself toward larger scale, and possibly more intense, outbreaks. In contrast, experiments suggest that elevated atmospheric CO_2 concentrations can impede the chemical communication systems that many insects use for finding mates and food sources. The prevalence and importance of this impedance in the wild is unknown.

Figure 1 summarizes this discussion. Climate change has a more pervasive influence throughout the life cycle of the spruce budworm than other components of global change, and will likely benefit spruce budworm populations. Biodiversity loss and land use change are also represented as having net positive influences for the spruce budworm. The only negative aspects of global change diagrammed in Figure 1 are associated with elevated atmospheric CO_2 concentrations. The large scale consequences of other changes in atmospheric composition are even less certain.

RESPONSES OF SALMON TO GLOBAL CHANGE

Wild salmon populations, *Oncorhynchus* and *Salmo* species, are declining globally: populations of Atlantic salmon are declining or have been extirpated in two-thirds of their native range, hundreds of Pacific salmon populations have gone extinct during the past century and hundreds more are at risk. The cause of these declines is multifaceted, and the relative contributions of human-induced changes, natural climatic events and other abiotic and biotic factors have not been firmly established. Unquestionably, however, human impact has been substantial.

The spatial scale at which human-induced changes affect wild salmon varies through the course of their life history, with localized impacts predominating in fresh water and global impacts predominating in oceanic environments (Figure 2). In fresh water, habitat degradation, pollution and introductions of exotic species represent the main human-induced impacts. Logging, agricultural, urban and resource extraction activities may all degrade habitat (e.g., sedimentation, channelization, flow conditions, removal of bank vegetation and fragmentation) and water quality (e.g., temperature, dissolved oxygen, nutrients and pollution). The effects of pollution also occur at large scales: acid rain is blamed for losses of Atlantic salmon populations in Europe and North America. Introduction of cultured salmon, exotic fishes and other aquatic organisms generally have localized impacts in fresh water. Escape of aquacultured salmon and stocking of hatchery salmon pose ecological (e.g., resource competition), genetic (e.g., loss of diversity), and disease threats to wild populations. In Norway, such practices have introduced and spread the parasite *Gyrodactulus salaris* which has almost eradicated 40 salmon populations.

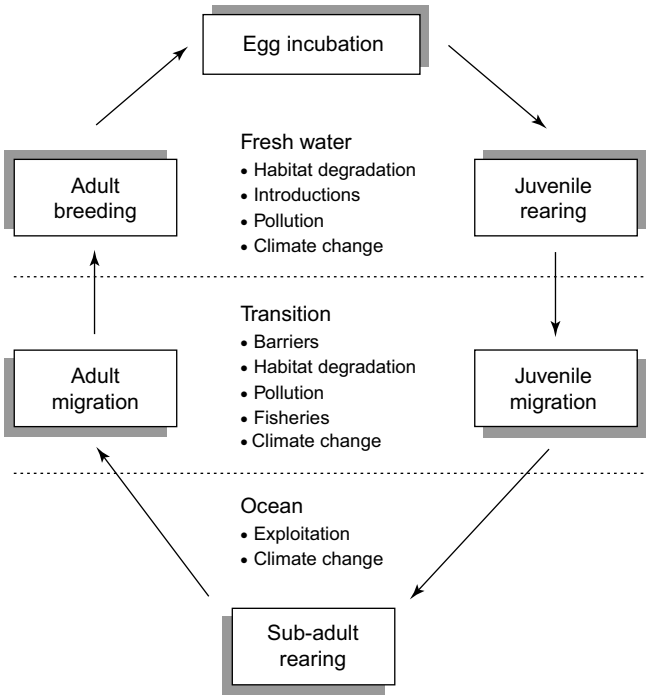


Figure 2 Principal impacts of global change during the life cycle of anadromous salmon, from egg incubation through juvenile freshwater and sub-adult oceanic rearing to adult freshwater breeding. Three critical stages are depicted: (1) fresh water; (2) the transition between fresh and salt water; and (3) ocean

The transition stage is also affected at predominately local scales. The major cause of salmon extirpations is blockage of their upstream migration routes, particularly by hydropower dams. Fortunately, the importance of this is declining. Effective fish passage mechanisms have been developed, inefficient dams removed, and with few hydropower opportunities remaining, dam construction is falling. Habitat degradation and pollution affect the energetic costs, susceptibility to predation, physiological state and ultimately, survival of salmon during migration. Over-fishing can also severely depress target populations and contribute to their risk of extirpation.

Oceanic impacts on wild salmon occur mainly at large scales, reflecting the vastness of this domain. Particularly important are climatic changes (*see* **Fisheries: Effects of Climate Change on the Life Cycles of Salmon**, Volume 3) and exploitation. It has been predicted that modest rises in sea surface temperatures due to climate change could result in salmon disappearing from large areas because of their narrow thermal limits during winter and early spring. Warmer sea surface temperatures may also interact with currents to alter the location and availability of food for salmon, which current evidence suggests will cause a reduction in salmon body size. Exploitation effects occur both directly (from salmon fisheries

[discussed above] and incidental catches during other fisheries), and indirectly (e.g., industrial fisheries competition of salmon with forage fish such as capelin and sand lance).

This analysis shows that most human impacts on Atlantic salmon are intensifying throughout the lifecycle, that they are predominantly negative, and that they are interacting in complex ways. Sustainability of wild salmon thus depends on forecasting and minimizing these impacts. The latter depends on maintaining a diversity of gene pools and population structures to provide evolutionary flexibility for salmon in response to an uncertain future.

RESPONSES OF BLACK SPRUCE TO GLOBAL CHANGE

Trees are amongst the longest-lived organisms on earth; several decades are usually required to reach sexual maturity or full fecundity, and generation times are often 50–200 years. Adaptations and changes in forest tree composition in response to global change will largely reflect the competitive abilities and ecological tolerances of existing ecotypes rather than the rapid evolution of new ecotypes. As the main structural component of forested landscapes, trees are the primary providers of habitat for many terrestrial organisms, and changes in the overstory resulting from global change will have wide-ranging repercussions.

Black spruce, *Picea mariana*, is an abundant, evergreen conifer found throughout the North American boreal forest. It has the most northerly range of any tree species on the continent, extending to the tree line, and is commonly found as far south as the temperate forests of the Great Lakes region. This slow-growing tree reaches heights of 15–25 m and ages exceeding 200 years on productive sites, but assumes a more shrub-like stature on very poor sites or in extreme climates. Over much of its southern and central range, it forms dense stands of considerable standing biomass, either by itself on wetter sites, or in combination with other species on well-drained sites. In northern locations, stands are more open and growth is reduced; at its southern limits, this species is confined to cold, wet, or shallow-soiled sites.

Black spruce seed and pollen cones emerge in the spring, and fertilized seeds mature by fall of the same year. Cones continue to disperse seed for several years after ripening. Seed production occurs regularly in stands over 25 years old, and peaks in stands between 50 and 150 years of age. Seed fall is accelerated for 2–3 years after wildfire. Seeds are commonly dispersed 100–200 m from stand edges, but individual seeds can achieve much greater dispersal distances, particularly during high winds or on crusted snow. Black spruce also reproduces vegetatively from lower living branches, which become partially buried by organic debris. This form of reproduction is common in open-grown stands, particularly in the far north where successful pollination and subsequent seed germination are restricted.

Climate change could have large direct impacts on the range, abundance and vigour of virtually all forest trees, including black spruce. Warmer temperatures should improve seed production and germination at higher latitudes and may allow considerable extension of black spruce beyond the current tree line. In contrast, the forecasted northerly extension of the western Canadian prairies would greatly reduce the southern extent of the species' range in western Canada. The direct effects of changes in precipitation patterns are less predictable. Black spruce is well adapted to drier upland sites as well as wetter lowland sites, both in the colder continental climates of northwestern North America, and in the cool, more humid climates to the east. However, drier conditions combined with warmer temperatures may restrict germination, limit seedling establishment, and increase drought stress in established stands. Increased temperatures and CO_2 levels will improve tree growth, but may put black spruce at a competitive disadvantage on sites which provide suitable habitat for competitors such as white pine (*Pinus strobus*), jack pine (*Pinus banksiana*) and oak (*Quercus*) species which have greater temperature or CO_2 growth responses.

In the southeastern portion of its range, climate change effects on black spruce populations may be manifested largely through alterations in competitive relationships with other species. Longevity, conservative resource requirements and broad ecological tolerances suggest that established black spruce trees and stands may not be greatly affected by the forecasted changes in climate. Increased competition from faster growing and/or more shade tolerant competitors, however, is likely to reduce seedling establishment and sapling vigour.

Climate change influences on the frequency and intensity of rarely occurring, catastrophic events (e.g., wildfire, insect outbreaks, storms, severe frosts) are equally as important to forest trees as the direct effects of incremental changes in climate averages. Like most boreal species, black spruce is adapted to and dependent upon infrequent wildfires, which destroy existing stands but produce dense, vigorous new stands of the same species. Increased fire frequencies (i.e., within every 30–40 years) could reduce black spruce abundance by burning many stands before most trees reach full fecundity. In contrast, substantial reductions in fire frequency could permit the incursion of faster-growing or more shade-tolerant competitors which are less fire-adapted but better resource competitors than black spruce. Feedback between changes in stand composition and fire frequency are also likely. Large increases in broad-leaved species will reduce fuel availability and perhaps fire frequency.

Insect epidemics affect individual tree species directly through herbivory, and indirectly by altering competitive

relationships with other plants. Outbreaks of the major insect pest of black spruce, the spruce budworm, reduce stand vigour and destroy current seed production, but cause little long-term damage to black spruce. This reflects the phenological development of feeding larvae, which is synchronized with vegetative bud flush and foliar development in balsam fir and white spruce, the spruce budworm's principal host species. Vegetative bud flush in black spruce usually occurs 10–14 days later. Differential effects of climate change on herbivore-host phenology could either reduce or greatly increase defoliation of black spruce by altering the synchrony of bud flush and budworm larval feeding. Climate-induced changes in the frequency and severity of insect outbreaks on associated species (e.g., forest tent caterpillar on trembling aspen and spruce budworm on balsam fir) could alter the competitive position of black spruce.

Other facets of global change which have widespread implications for forest trees include logging, deforestation and afforestation. Most of the land base currently occupied by black spruce has little agricultural value because of climate and soil limitations, and more land suitable for black spruce is currently reverting from agriculture to forests than *vice versa*. In contrast, logging activities have had large and sustained negative impacts on the representation of black spruce on the landscape. Black spruce is highly valued for pulpwood and saw timber, and has been logged extensively throughout the southern boreal. Typically, on peatland sites, black spruce remains the predominant species in second growth stands, although there may be regeneration delays and reductions in stand densities. Without planting and stand tending, many areas on upland sites revert from black spruce-dominated to mixed-wood forests with black spruce a minor component. The widespread establishment of black spruce plantations has maintained this species on other parts of the landscape, but often at reduced densities and with different stand compositions and structures than are found in stands of wildfire origin.

CONCLUSION

The life cycles of three very different species have provided analytic frameworks for developing a qualitative understanding of how each species is likely to react to global change. Global change will affect some stages of a life cycle more than others, and certain aspects of global change, either individually or interactively, are more important than others at these particularly vulnerable stages. To establish in a new area or persist in its current habitat, a species must be able to complete its life cycle. By identifying the most vulnerable stages, life cycle analyses can provide a focus for management efforts ranging from repelling invading species to protecting endangered ones (*see* **Integrated Pest Management (IPM) in an Era of Global Environmental Change**, Volume 4).

FURTHER READING

Ayres, M P and Lombardero, M J (2000) Assessing the Consequences of Global Change for Forest Disturbance from Herbivores and Pathogens, *Sci. Total Environ.*, **262**, 263–286.

Black, R A and Bliss, L C (1980) Reproductive Ecology of *Picea mariana* (Mill.) BSP., at Tree Line Near Inuvik, Northwest Territories, Canada, *Ecol. Monogr*, **50**, 331–354.

Farrar, J L (1995) *Trees in Canada*, Fitzhenry and Whiteside, Markham, Ontario, 1–502.

Fleming, R A (2000) Climate Change and Insect Disturbance Regimes in Canada's Boreal Forests, *World Resour. Rev.*, **12**(3), 520–555.

Johnson, E A (1992) *Fire and Vegetation Dynamics: Studies from the North American Boreal Forest*, Cambridge University Press, Cambridge.

Loehle, C and LeBlanc, D (1996) Model-based Assessments of Climate Change Effects on Forests: a Critical Review, *Ecol. Model.*, **90**, 1–31.

Mather, M E, Parrish, D L, and Flot, C L, eds (1998) Integrating Across Scales: Predicting Patterns of Change in Atlantic Salmon, *Can. J. Fish. Aquat. Sci.*, **55**(1), 1–323.

Nelhsen, W, William, J E, and Lichatowich, J A (1991) Pacific Salmon at the Crossroads: Stocks at Risk from California, Oregon, Idaho and Washington, *Fisheries*, **16**(2), 4–21.

Slaney, T L, Hyatt, K D, Northcote, T G, and Fielden, R J (1996) Status of Anadromous Salmon and Trout in British Columbia and Yukon, *Fisheries*, **21**(10), 20–35.

United Nations Environment Program (1999) *Global Environment Outlook 2000*, Earthscan, London, 1–398.

Viereck, L A and Johnston, W F (1990) *Picea Mariana* Mill.) B.S.P., in *Silvics of North America*, eds R M Burns and B H Honkala, Handbook 654, US Department of Agriculture, Forest Service, Washington, DC, 227–237, Vol. 1.

Limnology

see Lakes and Rivers (Volume 2)

Littoral Zone/Riparian Zone

see Lakes and Rivers (Volume 2)

LOICZ (Land-Ocean Interactions in the Coastal Zone)

The LOICZ project is one of the program elements of the International Geosphere–Biosphere Programme (IGBP). As

its name implies, LOICZ focuses on the area of the Earth's surface where land, ocean and atmosphere meet and interact. The overall goal of this project is to: determine the nature of that dynamic interaction at regional and global scales; determine how changes in various compartments of the Earth system are affecting coastal zones and altering their role in global cycles; assess how future changes in these areas will affect their use by people; and provide a sound scientific basis for future integrated management of coastal areas on a sustainable basis.

Unlike many of the other core projects of the IGBP, LOICZ deals with a specific, spatially heterogeneous domain rather than a process. To achieve the overall goals and objectives, a global network of coastal scientists must be developed. While LOICZ does not directly address coastal zone management, it does seek to provide a sound scientific basis for the future sustainable use and integrated management of coastal environments under conditions of global change.

LOICZ includes three classes of research projects: core, regional, and relevant. These classes are distinguished on the basis of geographic extent, scientific contribution to LOICZ global research, and commitment to the project.

Reports and data collections are being produced, and a newsletter is regularly published.

Further information may be obtained from: LOICZ International Project Office, Netherlands Institute for Sea Research (NIOZ), PO Box 59, 1790 AB Den Burg, Texel, The Netherlands, Tel.: +31-222 369404, Fax: +31-222 369430, E-mail: loicz@nioz.nl, WWW home page: http://www.nioz.nl/loicz/.

See also: **IGBP Core Projects**, Volume 2.

JOHN S PERRY *USA*

Long-term Ecological Research (LTER) System

James R Gosz
University of New Mexico, Albuquerque, NM, USA

The concept of long-term research sites dedicated to collaborative research was developed in the US during a series of three workshops sponsored by the National Science Foundation (NSF) from 1977–1979. At these workshops, the philosophy of collaborative research was emphasized and a centralized working hypothesis approach proposed to facilitate collaboration. Five core areas of research were defined to orient LTER projects toward question/hypothesis formulation and resolution. The core research areas common to all LTER sites in the US are:

- *pattern and control of primary production;*
- *spatial and temporal distribution of populations selected to represent trophic structure;*
- *pattern and control of organic matter accumulation in surface layers and sediments;*
- *patterns of inorganic inputs and movements of nutrients through soils, groundwater and surface waters;*
- *patterns and frequency of site disturbances.*

These areas are basic to all ecosystems and are required to understand patterns, processes, and relationship to change. They also require participation by different disciplines that form the basis for collaborative interactions.

General Details

The first six site awards were initiated in 1980 and as of 2000, 24 sites had LTER awards based on competitions that evaluated research excellence, quality and duration of existing data sets, and strength of the commitment to long-term research and site security (Figure 1). As part of their commitment to the LTER program, each site conducts a series of measurements and experiments directed towards the understanding of the five core areas and studies addressing ecological issues specific to the site. The most common scientific approaches include observation, experimentation,

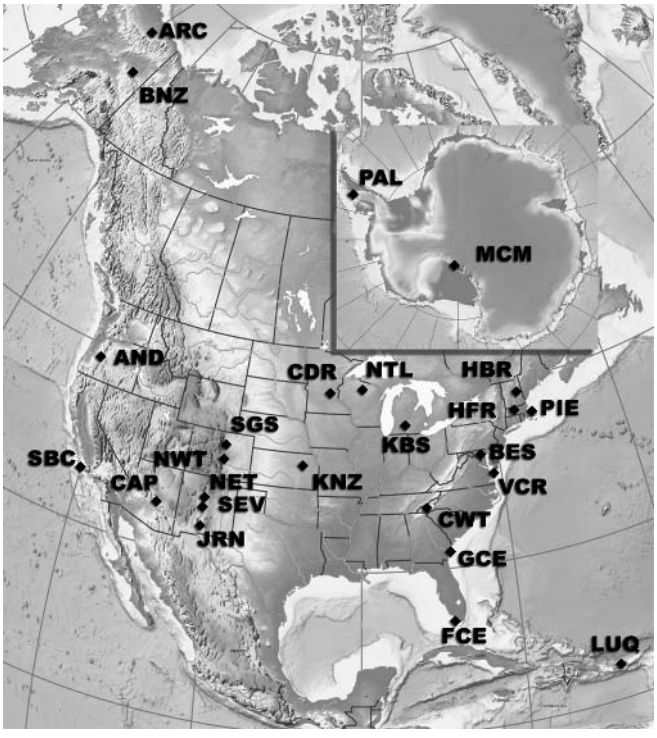


Figure 1 Locations of the LTER sites in the United States

Table 1 US LTER site names and locations

H.J. Andrews Experimental Forest, OR (AND)
Arctic Tundra, Toolik Lake, AL (ARC)
Baltimore Urban Ecosystem Study, MD (BES)
Bonanza Creek Experimental Forest, AL (BNZ)
Central Arizona- Phoenix Urban Ecosystem Study, AZ (CAP)
Cedar Creek Natural History Area, MN (CDR)
Coweeta Hydrologic Laboratory, NC (CWT)
Florida Coastal Everglades, FL (FCE)
Georgia Coastal, GA (GCE)
Harvard Forest, MA (HFR)
Hubbard Brook Experimental Forest, NH (HBR)
Jornada Experimental Range, NM (JRN)
Kellogg Biological Station, MI (KBS)
Konza Prairie Natural Research Area, KS (KNZ)
Luquillo Experimental forest, PR (LUQ)
McMurdo Dry Valleys, Antarctica (MCM)
North Temperate Lakes, WI (NTL)
Niwot Ridge- Green Lakes Valley, CO (NWT)
Palmer Station, Antarctica (PAL)
Plum Island Ecosystem, MA (PIE)
Santa Barbara Coastal Ecosystem, CA (SBC)
Sevilleta National Wildlife Refuge, NM (SEV)
Short Grass Steppe, CO (SGS)
Virginia Coast Reserve, VA (VCR)
LTER Network Of ce, Albuquerque, NM (NET)

comparative analysis, retrospective study and modeling, although emphases differ among sites.

The LTER sites comprise a broad array of ecosystems ranging from tropical rain forest in Puerto Rico to polar desert oases at McMurdo Station, Antarctica (Table 1). Collectively, the sites provide opportunities to contrast marine, coastal, and continental regions, the full range of climatic gradients existing in North America, and aquatic and terrestrial habitats in a range of ecosystem types. All sites are sufficiently large to incorporate moderate to large landscape mosaics, and most sites include human-manipulated as well as natural ecosystems. Most sites embody considerable within-site variability in habitats and ecosystem processes and attempt to characterize this variability in the context of broad regional gradients covering hundreds of kilometers.

LTER Network

The 24 sites are now organized into a LTER Network representing a wide variety of research emphases and approaches with over 1200 scientists and graduate students involved. The mission of the LTER Network is to conduct and nurture ecological research by:

1. understanding general ecological phenomena that occur over longer temporal and large spatial scales;
2. creating a legacy of well designed and documented long-term experiments and observations for the use of future generations;
3. conducting major synthetic and theoretical efforts; and
4. providing information for the identification and solution of societal problems.

Over time (i.e., since 1980) the focus of research efforts has shifted from a concentration on individual site research to a broader synthetic viewpoint. Expectations for results from the research programs now include not only the individual and site-based products, but also cross-site, network-wide, and international collaborative studies. There are many examples of the value of information developed at LTER sites. Long-term research is crucial in understanding issues like climate change, effects of pollution, and biodiversity and in contributing to management decisions.

The range of locations provides an understanding of the diversity of regional patterns of the effects of climate change. For example, McMurdo Dry Valley in the Antarctic has demonstrated an increase in glacier mass since 1993 which coincides with a period of cooler than normal summers and more than average snowfall. The larger glaciers seem to be advancing, primarily a legacy of past climate, roughly 103 years before present. In contrast, the Bonanza Creek and Toolik Lake sites in Alaska have recorded a warming period unprecedented in the past 200 years. Warmer air and soil temperatures and drought are associated with increased melting of ice-rich permafrost, fire, insect outbreaks, reduced growth and seed production in dominant tree species, extensive grassland development following disturbance, reduction in moss abundance and increased shrub and birch densities in the far north. Increased warming is associated with changed dominance of grass species in the Short Grass Steppe in Colorado while the Sevilleta site in central New Mexico had increased tree growth during the wet period of the 1980s–early 1990s, unprecedented in the past 500 years.

Nitrogen deposition effects have been identified through long-term watershed studies on biogeochemical cycles, especially at Coweeta Hydrological Laboratory, Harvard Forest and the Hubbard Brook Experimental Forest, producing evidence of nitrogen saturation associated with forest maturation and increased nitrogen deposition over time. Forests influenced by management and natural disturbance (e.g., hurricanes, insect outbreaks) exhibit more advanced stages of nitrogen saturation. At Cedar Creek, increased nitrogen deposition is linked to increased carbon sequestration, but also loss of plant diversity and increased growth of weedy, non-native plant species.

Biodiversity research results range from the effects of climate change and pollution as described above to the influence of management. Research at the Konza tall grass prairie site has demonstrated that re-establishing grazing by native ungulates (bison) in combination with burning enhances productivity and biodiversity. Agricultural research at the Kellogg Biological Station shows that although agricultural management decreases plant diversity

in cropping systems, different management strategies can be developed that promote different assemblages of species which have important implications for ecosystem services such as protection from insect pests and nutrient conservation associated with species considered to be weeds. Results show that rotation of crops in conventionally managed systems does not increase the diversity or composition of the weed community that emerges, whereas under biological management, the composition of the weed community changes markedly with crop rotation. These sorts of results are key to creating designer weed communities that provide ecosystem services without competing with the crop for limited resources.

In aquatic systems, such as the Northern Temperate Lakes site, long-term data have revealed time lags in effects of invaders and disturbance on lake communities. Species such as the Lake Cisco went extinct 16 years after a smelt invasion and in Trout Lake the loss in aquatic plant diversity was delayed as long as 20 years after a Rusty Crawfish invasion. Biological recovery from an effect such as lake acidification was shown to lag behind chemical recovery. These results are critical to developing management programs.

Perhaps the best way of documenting the value of LTER research is by the way it has influenced management. In Puerto Rico, the Luquillo LTER demonstrated that dams obstruct migrations of fish and freshwater shrimp; and water abstraction at low flows can completely obliterate downstream migration of juveniles and damage estuaries below by removing all incoming freshwater. As a result local water authorities changed the design of a new water filtration plant to maintain minimum stream flows. At Toolik Lake, long-term experiments of removing top predators from the food web of lakes showed dramatic alterations of lake populations of small fish and zooplankton. As a result, the Alaska Fish and Game Department has instituted a catch-and-release law in northern Alaska to preserve the population structure of lakes. At the Virginia Coast Reserve, repeated surveys have documented major island-scale geomorphic adjustments as result of storms. Hundreds of meters of shoreline retreat and hundreds of square meters of destruction of major dune environments can occur as evidenced by the major storms of early 1998. Studies of these island adjustments are critical in human occupation of coastal areas. Research on the Short Grass Steppe in Colorado has demonstrated that livestock grazing has only limited impacts on the Short Grass Steppe ecosystem structure and function, likely due to a long evolutionary history of grazing by bison and other large herbivores. Research demonstrated that the plant community has a high degree of tolerance to both grazing pressure and yearly fluctuations in precipitation. Cultivation in this system, however, was shown to have strong impacts on ecosystem structure and function and a strong forcing function on regional climate. Regional land use patterns appear to have a larger impact on regional climate than does the signal from global greenhouse-induced climate change. In New Mexico, LTER research on small mammal populations is successfully predicting rodent increases and the potential for increased zoonotic diseases such as Hantavirus and bubonic plague. This ability to forecast based on El Niño predictions is being used to increase scientific awareness and public health awareness through media communication with the public. In Oregon, the Andrews Forest LTER program has had long, strong links with natural resource policy and management. Basic understanding of forest–stream interactions, characteristics of old-growth forests, roles of woody debris in temperate forest ecosystems, invertebrate biodiversity and ecosystem function have been incorporated in management guidelines, plans and regulations for public and private lands throughout the Pacific Northwest. For example, the riparian management guidelines for the Willamette National Forest, where the Andrews Forest is located, became the cornerstone concepts of forest–stream interactions used in President Clinton's Northwest Forest Plan for the region. Later they were used in similar ways for the state of Minnesota and translated into Japanese to be used to increase environmental sensitivity in management of rivers in Japan.

Many other examples of the value of LTER research are available. Access to the science at all of the sites is available through the LTER web site (http://www.lternet.edu).

Information Management

Information management is crucial to the success of the long-term research programs and the development of comparable data sets, and a Network Information System (NIS) to set standards in information management is actively facilitated by information management professionals at each site. There is a continuous effort to develop standards for equipment and methodology for commonly collected data (e.g., climate measurements) and to develop tools to access and analyze data sets existing in the distributed Network environment. The NIS sets standards in information management sufficient to achieve network-wide data integration. The goal of NIS is to facilitate seamless data exchange and synthesis within the LTER Network and to make data collected at LTER sites available to scientists worldwide. The goal is accomplished through the use of the latest advances in World Wide Web database interfaces, that provide access to network-side data sets through a single point of entry.

Management

The management of the US LTER Network is led by a coordinating committee comprised of representatives of all 24

sites. This group meets twice a year to plan LTER activities. An Executive Committee meets more frequently and is empowered to act on behalf of the Network in routine matters. Eight standing committees help to coordinate specific LTER activities; collection and synthesis of climate data, education, social and economic science collaboration, publications, information management, graduate student participation, research synthesis and technology. The members of these committees are drawn from the LTER community at large. The activities of the standing committees include the development of network-wide policies and initiatives, preparation of proposals and interactions with private and public institutions as well as other segments of the scientific community. In 1983, NSF established the LTER Network Office under the direction of the LTER Coordinating Committee Chair and Executive Committees. The goals of the Network Office are:

- to facilitate communication among the LTER sites and between the LTER program and other scientific communities;
- to support the planning and conduct of collaborative research efforts, including provision of some technical support services;
- to facilitate intersite scientific activities;
- to provide a focal point and collective representation of the LTER Network in its external relationships;
- to develop linkages with other relevant long-term research programs, site networks, and science and technology centers.

US Partnerships/Collaboration Among Networks

Strong partnerships have developed between the academic research programs that represent LTER awards and the US federal and state agency research programs at many sites. Most of the LTER sites are on federal property (e.g., Forest Service Experimental Forests, Agriculture Research Service Experiment Stations, National Wildlife Refuges, National Parks) and involve collaborative efforts between federal and academic researchers. The LTER Network continues to develop interactions with federal network programs such as those of National Aeronautics and Space Administration (NASA) (e.g., Mission to Planet Earth, AErosol RObotic NETwork (AERONET), Oak Ridge Distributed Active Archive Center), National Resource and Conservation Service (e.g., soil mapping), United States Department of Agriculture (USDA) (National UVB Radiation Monitoring National Network), and National Oceanic and Atmospheric Administration (NOAA) (e.g., National Estuarine Research Reserves). Personnel from the LTER Network Office actively cooperate with several national agency/interagency efforts to assure interoperability between the LTER Network and the greater scientific community (e.g., National Biological Information Infrastructure (NBII) metamaker project, US National Committee for Committee on Data for Science and Technology (CODATA).

International Networking

Since the first exploratory discussions with other countries in 1993, LTER programs have developed rapidly, based on experiences in the US. At that time an International LTER (ILTER) Network was formed with a mission to facilitate international cooperation among scientists engaged in LTER. Subsequently, the ILTER Network Committee established the following mission statements:

- to promote and enhance the understanding of long-term ecological phenomena across national and regional boundaries;
- to promote comparative analysis and synthesis across sites;
- to facilitate interaction among participating scientists across disciplines and sites;
- to promote comparability of observations and experiments, integration of research and monitoring, and encourage data exchange;
- to enhance training and education in comparative LTER and its relevant technologies;
- to contribute to the scientific basis for ecosystem management;
- to facilitate international collaboration among comprehensive, site-based, long-term, ecological research programs;
- to facilitate development of such programs in regions where they currently do not exist.

Each country must assess its own needs and resources if it wishes to develop a national LTER program and involve itself in an ILTER program. Each nation will have a unique set of opportunities and limitations that are best evaluated by the scientists and policy makers of that country. The typical procedure is for the scientists and funding agencies of that country to decide whether to endorse the premise that ecology and environmental management are significantly benefited by long-term, integrated studies over broad spatial scales. A plan is then developed that establishes the context and mission for such studies, and sites and programs are identified that will contribute to this mission. Typically, support is obtained from within that country and/or international organizations for implementation and continued maintenance. The models developed by most countries are similar to that in the US; however, the core areas that focus research efforts across sites can vary. Many countries add biodiversity, agriculture, and anthropogenic effects to the core research areas. The requirement for joining the ILTER Network is formal recognition by a national-level organization that a national network of long-term research sites has been established for that country. It is anticipated that

Table 2 Countries participating in the International LTER Network (2000)

Australia
Brazil
Canada
People's Republic of China
Taiwan, China
Colombia
Costa Rica
Czech Republic
Hungary
Israel
Korea
Mongolia
Namibia
Poland
Slovakia
Switzerland
Ukraine
UK
US
Uruguay
Venezuela
We anticipate the following countries to be added by the end of 2001
France
Mexico
South Africa
Other countries are in the process of developing national LTER programs

each country's program will be part of a global network of scientists and of scientific information that will advance our understanding of not only local and regional, but also global issues and provide solutions to environmental problems at these scales.

Currently (2000), there are 21 countries with LTER programs (Table 2) and additional countries in the process of establishing such programs. A more recent development is the formation of Regional LTER Networks. Because neighboring countries often have similar issues, there are increased opportunities for collaboration and support. At this point, regional organizations exist for the East Asian Regional LTER Network, the North American Regional LTER Network, the Central and Eastern Europe Regional Network, the Middle East Regional Network, and the Latin American Regional Network. The various sites within a national network to those in regional national networks to the entire international LTER Network all contribute to the vision of integration and collaboration needed for modern, global science.

LTER sites in the countries of the ILTER Network can provide unparalleled opportunities for cross-site and comparative research efforts on many of the world's ecosystems. Just as with focused site research described earlier for the US, these global LTER sites lead to interdisciplinary research at broader scales that can provide the scientific basis for management and policy decisions. The ILTER Network is now well positioned to interact with other international activities such as the International Geosphere Biosphere Programme (IGBP), the Global Terrestrial Observing System (GTOS), and the United Nations Educational, Scientific and Cultural Organization (UNESCO) Man and the Biosphere (MAB) program. We anticipate rapid progress in these interactions in addition to continued expansion of the ILTER Network.

For further information on LTER, see http://www.lternet.edu.

LUCC (Land-use/Cover Change)

see IGBP Core Projects (*Opening essay*, Volume 2)

Lyapunov Exponent

see Chaos and Cycles (Volume 2)

M

Malate formers

see Photosynthesis (**Volume 2**)

Mangrove Ecosystems

Robert R Twilley

University of Louisiana at Lafayette, Lafayette, LA, USA

Mangroves are forested wetlands that are salt tolerant and dominate the land- sea interface of warm tropical environments. These tropical forests are unique communities of diverse biological organization that offer scientists and resource managers insights into how ecosystems can respond to a variety of ever-changing environmental conditions. The plant physiologist nds it remarkable that a tree that has its roots surrounded by salt water (which can be considered an osmotic desert) can transport a parcel of water to a leaf that is up to 40 m above the ground (over 120 ft). Population and community ecologists marvel at the biodiversity and associated mutualistic interactions that result in a most productive ecosystem even along a salinity spectrum from nearly freshwater (oligohaline) to marine (euhaline) conditions. The ecosystem ecologist and coastal geomorphologist can together catalogue mangroves into schemes that describe nearly every type of coastal formation from riverine deltas to oceanic reefs - another example of the tremendous biodiversity of mangrove ecosystems. And the coevolution of geological and biological processes of these ecosystems over the last several interglacial periods has produced an intriguing ecological landscape.

Mangroves truly represent an ecosystem that we can use to monitor several dimensions of global environmental change (Figure 1). Mangroves also demonstrate the importance of understanding the relative impacts of global environmental change on ecosystem goods and services along with socioeconomic decisions of regional land use development (Figure 2). The coupling of mangroves to coastal waters is considered to be the most important link in the sustainability of subsistence, commercial and recreational fisheries associated with estuaries and related nearshore marine habitats. In addition, mangrove plantations have historically promoted sustainable use of forest products such as timber, fuelwood, tannins, pulpwood, and charcoal throughout the tropics. Mangroves also play a major role in maintaining water quality and shoreline stability, by controlling nutrient and sediment distributions in estuarine waters. However, there are several examples of how regional land use decisions have resulted in negative feedback effects on sustaining these ecosystem goods and services. Recent forms of direct exploitation include the destruction of biodiversity by land uses such as aquaculture (shrimp ponds), agriculture (rice and salt ponds), urban development, and forest clearfelling. Indirect loss of mangrove ecosystems has resulted from human alterations of upland watersheds causing diversion of freshwater (dams and canals) and deterioration of water quality (nutrients, heavy metals, oil spills, pesticides). Projections of how mangroves will respond to global environmental change will have to be considered along with these human impacts to the coastal landscape, coupled with impacts to upland tropical watersheds (Figure 2).

The environmental settings of mangroves are a complex behavior of regional climate, river discharge, tides, wind, and oceanographic currents (Figure 1). There are about $240 \times 10^3\,\mathrm{km}^2$ of mangroves that dominate tropical continental margins from river deltas, lagoons and estuarine complexes, developed from terrigenous sediments to carbonate deposits. In addition, mangroves colonize shorelines and islands in oceanic formations (non-continental) formed principally from carbonate sediments. These geomorphologically distinct landscapes have local variations in topography and hydrology that result in the evolution of distinct ecological types of mangroves such as riverine, fringe, inland, and scrub forests. These ecological types of mangroves have specific structural characteristics and ecological

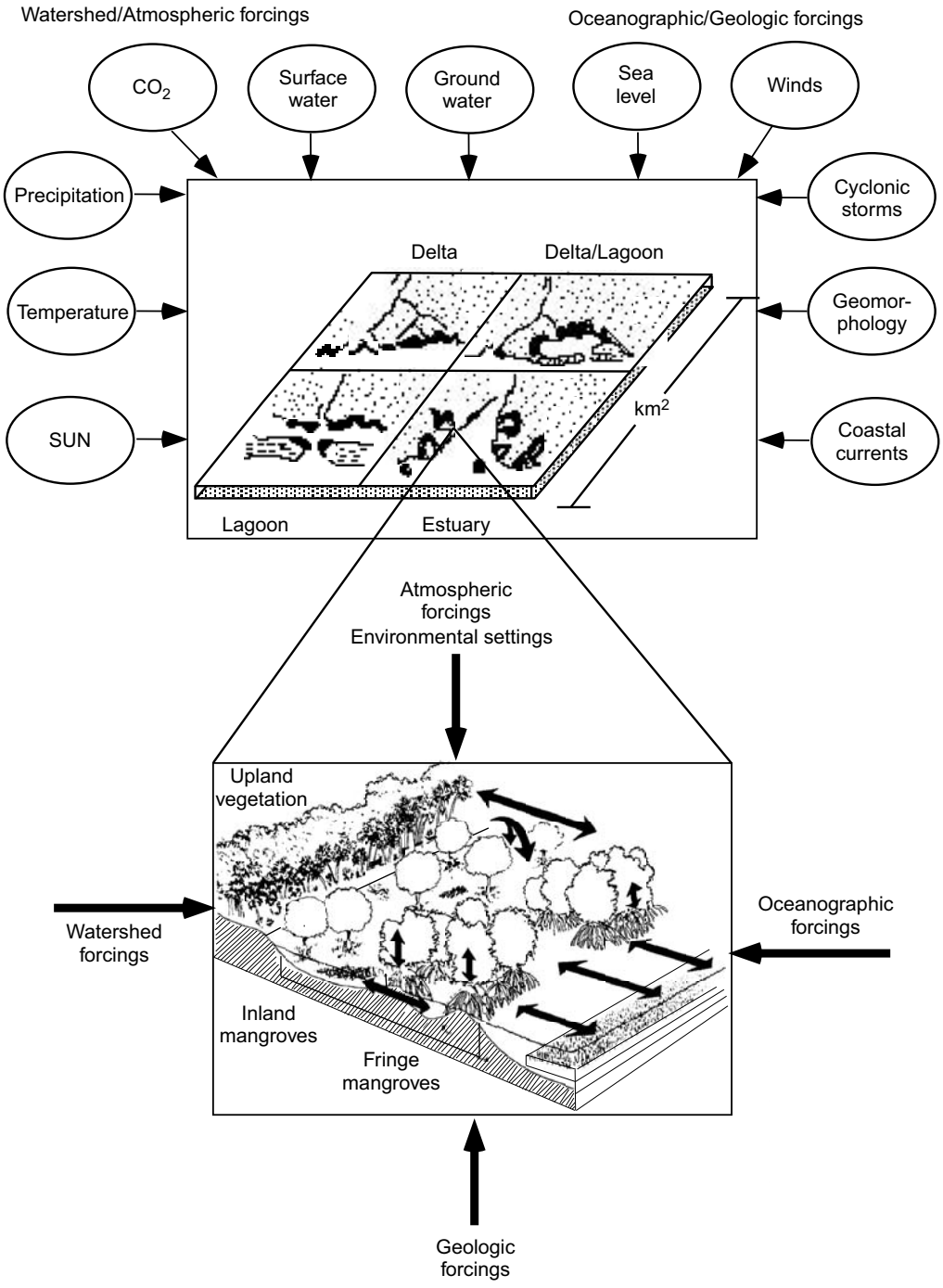


Figure 1 Classification system to describe how the relative strength of different forcings across different types of environmental settings control the spatial distribution of various ecological types of mangroves across the coastal landscape

functions. Although there are relatively few species of mangrove trees (54 true species), mangrove ecosystems are nevertheless diverse because mangroves in the coastal environment support a disproportionately rich diversity of fauna. The unique location of these tropical forests along the sea combines the diversity of terrestrial and marine ecosystems into one of the more diverse ecosystems among coastal habitats.

There is evidence that mangrove biodiversity is intricately linked to ecosystem function. Mangrove ecosystems

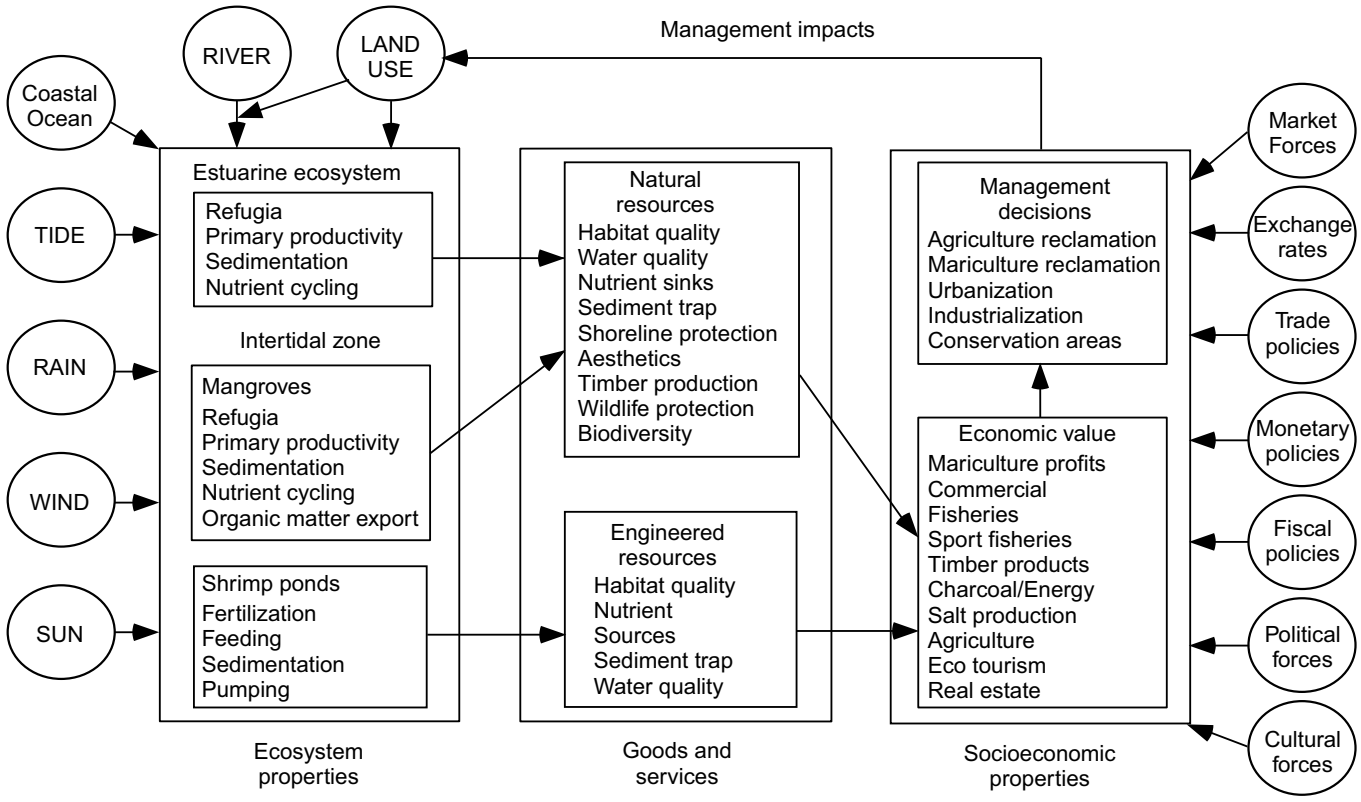


Figure 2 Diagram of how both environmental forcings (left side of the diagram) and social constraints (right side of diagram) control ecosystem and socioeconomic properties of mangrove ecosystems; including feedback effects of human utilization on ecosystem properties

support a variety of marine and estuarine food webs involving an extraordinarily large number of animal species and complex heterotrophic microbial food webs (Figure 3). Extensive surveys of marine fauna have found 26–114 species of fish and more than 100 species of macroinvertebrates that exert control of community dynamics in fringe mangrove zones. One of the best examples is the mangrove crab fauna that is considered a keystone guild in these forested wetlands. The genus *Sesarma* accounts for over 60 species of crabs predominantly associated with mangroves with many functions, such as contributing to plant zonation by differential seed predation, to litter dynamics by transporting leaves into burrows, and to nutrient dynamics by aerating otherwise anaerobic soils. Sponges, tunicates, and a variety of other forms of epibionts on prop roots are equally diverse and may be sources of nutrition for higher-level predators, as well as providing nutrients to fringe mangrove forests.

In addition to the marine and estuarine food webs and associated species, there are a relatively large number and variety of animals from inland habitats, such as insects and birds, that live in and/or that feed directly on mangrove vegetation. These include arboreal feeders such as foliovores and frugivores, and ground-level seed predators. Over 200 species of insects have been documented in mangroves such as in the Caribbean. In Belize, wood boring insects in mangroves are important to the formation of gaps and the specific architecture of mangrove trees. Also, the diverse avifauna, mainly those that use mangroves for rookeries, control nutrient levels and foliage dynamics through nesting and the associated herbivory. These keystone guilds have disproportionate effects on ecological processes of mangrove ecosystems, and represent the importance of biodiversity in ecosystem functioning.

Mangroves occupy a transitional zone from land to ocean, and such ecotones (*see* Ecotones, Volume 2) illustrate ecosystems that are particularly sensitive to environmental change. There is often a distinct elevation gradient from the water's edge (lower intertidal zone) to the terrestrial transition zone (supratidal zone). Mangroves are taller in the lower intertidal zone, where they are inundated nearly daily by tides, known as *fringe mangroves*. Lower intertidal zones with more in uence from rivers than tides have a greater supply of sediments and nutrients, and lower salinities. Mangroves in this zone are even more robust and are known as *riverine mangroves*. Farther inland to the mid and upper intertidal zone, there

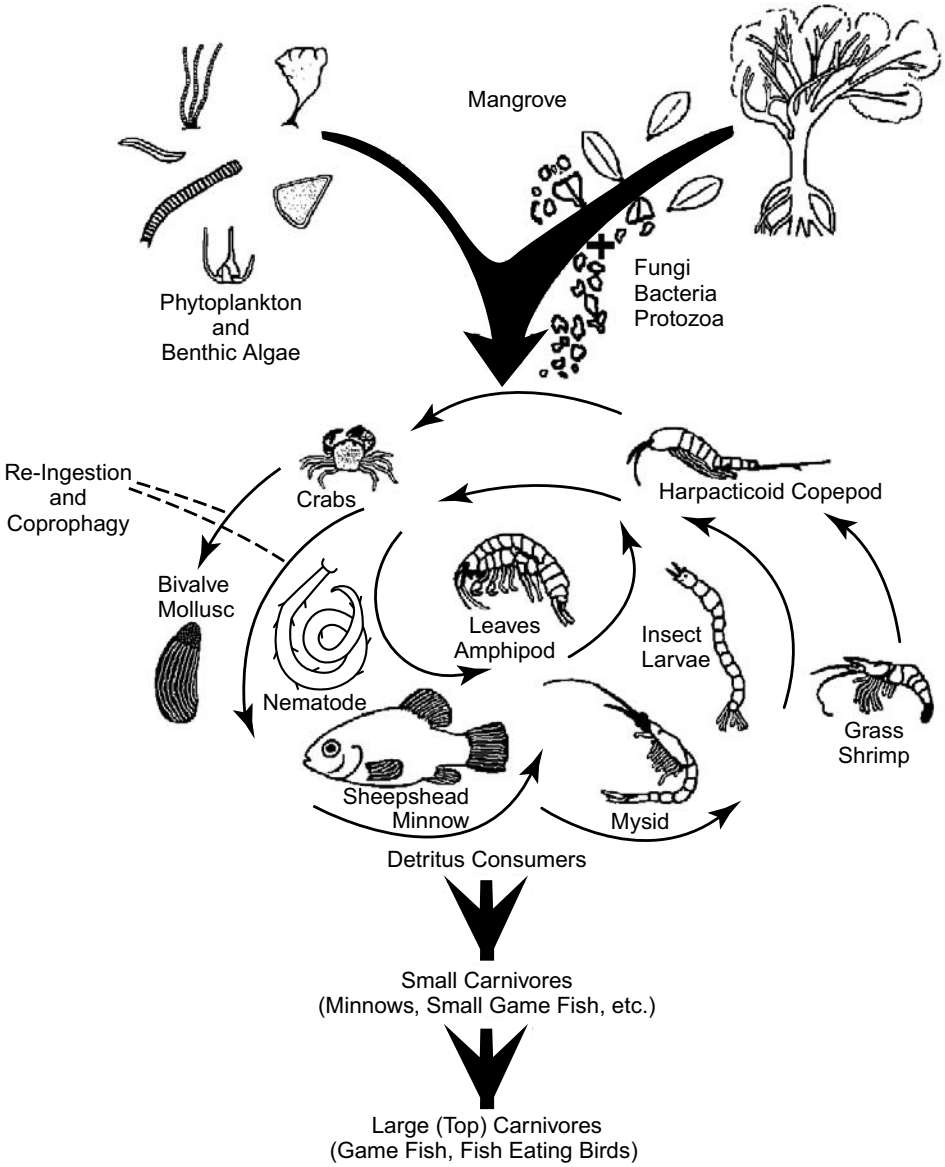


Figure 3 Conceptual diagram describing the fate of mangrove leaf litter in the food chain of an estuary in south Florida determined by William Odum and Eric Heald

is unique *inland mangrove* vegetation whose stature is more dependent on local climate and upland freshwater discharge. Even farther away from shore in the supratidal region, mangroves are at the edge of optimum growth conditions, and together with competition from upland vegetation, they usually have restricted growth and are known as *scrub mangroves*.

The diverse patterns in species composition of mangrove trees in these different zones (riverine, fringe, inland and scrub), along with distinct patterns of forest structure and productivity, reveal the response of mangroves to a variety of environmental gradients in resources and regulators. These gradients are associated with the relative strength of environmental drivers (e.g., temperature, tidal amplitude and river ow) across the coastal landscape (Figure 1). These spatial patterns in response to environmental gradients may also provide insights as to how these ecosystems will respond to global environmental change in time.

The predominate forcings from the ocean include tides and sea level change, oceanographic currents, and the source of salinity that represents one of the major stressors in mangrove habitats. Atmospheric forcings include wind (including cyclonic disturbances), precipitation, relative humidity, and temperature; the latter two components determine regional loss of water by

evapotranspiration. The watershed provides both surface and subsurface freshwater discharge that potentially transports sediments and nutrients to the coast. This forcing can determine the type and fertility of the estuarine environment. The magnitude and distribution of these forcings across the landscape topography determines the patterns of hydroperiod, soil nutrients, and pore water salinity that largely control zones of mangrove vegetation. Environmental changes associated with either the oceanic, atmospheric, or watershed forcings will change the species composition, structure, and function of mangrove ecosystems.

Accelerated rates in sea level rise have been speculated as the most critical environmental change affecting the continued existence of mangrove ecosystems. Numerous processes contribute to vertical accretion of mangroves at a rate that balances the increase in regional sea level rise. Critical rates in sea level rise have been estimated above which there is a projected collapse of mangrove ecosystems. While some speculation suggests that mangroves cannot sustain existence at sea level rise $>1.2-2.3\,\mathrm{mm\,year^{-1}}$, there is evidence that mangroves existed through periods of accelerated rates of sea level rise above these critical rates in particular environmental settings. Mangroves in the south Alligator tidal river in Australia can keep pace with changes in sea level rise with rates ranging from $0.2-6\,\mathrm{mm\,year^{-1}}$. Mangrove forests in many estuaries in northern Australia tolerated sea level rise of $8-10\,\mathrm{mm\,year^{-1}}$ in the early Holocene. These rates are higher than many projections of accelerated rates of sea level rise associated with global climate change. Many of these mangroves receive terrigenous sediments and exist in macrotidal environments, with critical rates that are much different than for mangroves in microtidal and oligotrophic (carbonate) environments. However, given projected rates of accelerated sea level rise of $10-20\,\mathrm{mm\,year^{-1}}$, most mangrove wetlands will collapse as viable coastal ecosystems.

Mangrove areas can be sustained along the coastline by migrating inland under conditions of increased sea level rise. However, this inland migration will depend on whether suitable inshore landscapes are available. Coastal landscapes with steep topographic gradients resulting from regional geologic and oceanographic processes may limit inland migrations; but the most significant recent restriction to mangrove colonization is human land use of available landscapes. At the oceanic edge of mangroves, mineral accumulation is enhanced by mangrove prop roots and pneumatophores, which trap sediments suspended in the water column during low-velocity flooding. Coastal forested wetlands thereby enrich intertidal soils and retard the forces of erosion along the shoreline. This process has led to a significant expansion of mangrove shorelines in several regions and is considered a very important service in protecting shorelines from erosion. Accelerated sea level rise is projected to increase coastal erosion and destabilize shorelines presently colonized by mangroves.

The tropical nature of mangrove forests constrains their latitudinal distribution depending on the frequency, duration and/or severity of freezing temperatures. Red mangrove seedlings are more severely impacted by low temperatures among three Caribbean mangrove species due to their inability to resprout following foliar damage from freezing temperatures. The greater resprouting ability of black mangroves results in better recovery from freeze damage, resulting in the distribution of this species into Texas, Louisiana and Mississippi. Global warming will result in potential spread of mangroves into higher latitudes. On many coastlines, however, the latitudinal extension of mangroves is either restricted by cold oceanographic currents (such as along the Pacific coast of South America), or has expanded into subtropical latitudes by warm water flowing away from the equator (Atlantic coast of North and South America). Changes in ocean temperatures and currents will influence the latitudinal distribution of heat along tropical coastlines and influence the latitudinal distribution of selected mangrove species.

Warmer ocean temperatures could increase the intensity and/or the frequency of cyclonic storms. For those mangroves distributed in latitudes where the frequency of hurricanes is high (such as the Caribbean), the frequency of cyclonic disturbances can restrict forest development (i.e., forest structure, species composition, etc.) and community dynamics. Damage by hurricanes is more severe on large trees, where tree mortality of 36% can represent 68% of the basal area of a preexisting stand. Species attributes and availability of propagules are important factors, along with the severity of storm and sediment disturbance, in projecting recovery patterns. Frequent storm disturbance tends to favor species capable of constant or timely flowering, abundant seedling or sprouting, fast growth in open conditions, and early reproductive maturity. Although mangrove trees show these traits, it is important to consider the cumulative impact of human activities on these ecosystems, in conjunction with the complex natural cycle of regeneration and growth of mangrove forests. Cyclonic disturbance in areas with higher rates of sea level rise have been demonstrated to cause sediment collapse (drop in surface elevation), which reduces the ability of mangroves to recolonize disturbed areas. The loss of above-ground biomass following disturbance reduces below-ground productivity, which is critical to continued soil formation and accretion. Thus, the increased frequency of tropical disturbances, together with accelerated sea level rise, could expand the loss of mangrove ecosystems along coastlines susceptible to both types of environmental change.

Atmospheric forcings in the upper intertidal and supratidal ecotones, where tides are less frequent, control the moisture content of intertidal soils by the balance of rainfall and evapotranspiration. Evapotranspiration is linked to seasonal patterns of temperature and relative humidity, and together with rainfall, determine dry or humid life zones. In dry climates with little freshwater discharge from the estuarine watershed, an imbalance of soil moisture in these zones, even though infrequently ooded by tides, can result in salinities over three times that of the ocean. As a result, scrub mangroves occupy these hypersaline soils and *salinas* (zones void of vegetation) are common in the supratidal region. Scrub mangroves may also occur in the upper intertidal zone of moist environments in response to oligotrophic soils or long hydroperiods rather than high salinity. In these moist environments, the scrub zone is inhabited by different species composition competing with freshwater marsh or forested wetland vegetation. In dry coastal tropical regions where precipitation deficits occur, limits to mangrove zonation and forest structure can be negated by the presence of freshwater inputs (surface or subsurface) to the coast from upland watersheds. Thus, the amount of moisture in mangrove soils is a complex interaction of regional climate and hydrology, including upland freshwater input. Some mangroves may be located in a dry climatic zone, but they may have the characteristic zonation of mangroves in humid coastal environmental settings due to the presence of freshwater inputs. These freshwater subsides from upland watersheds increase the moisture of mangrove soils and in uence the functional ecology of mangroves.

River (and surface runoff) diversions that deprive tropical coastal deltas of freshwater and silt result in losses of mangrove species diversity and organic production, and alter the terrestrial and aquatic food webs that mangrove ecosystems support. Freshwater diversion of the Indus River to agriculture in Sind Province over the last several hundred years has reduced the once species-rich Indus River delta to a sparse community dominated by *Avicennia marina* (*see* Indus Basin: a Case Study in Water Management, Volume 3). It is also responsible for causing significant erosion of the seafront due to sediment starvation and the silting-in of the abandoned spill rivers. A similar phenomenon has been observed in southwestern Bangladesh following natural changes in river channels of the Ganges and the construction of the Farakka barrage that reduced the dry season ow of freshwater into the mangrove-dominated western Sundarbans. Freshwater starvation, both natural and human-induced, has had negative impacts on the biodiversity of mangroves in the Ganges River delta, as well along the dry coastal life zone of Colombia (the C enaga Grande de Santa Marta lagoon). The effects of human land use change of tropical watersheds are not only important regionally, but they have significant impact on biodiversity and ecosystem goods and services of coastal systems down stream. Mangroves and estuaries are important nurseries because of the dependence of marine organisms on estuarine conditions during the juvenile or adult stages of their life cycle. Many organisms migrate from either freshwater habitats upstream or marine habitats offshore to mangrove refugia during critical stages of their life cycle. The fragmentation of mangrove-dominated landscapes may create the same types of problems for estuarine-dependent organisms that are associated with the fragmentation of upland tropical forests.

Another significant type of global environmental change is the recent rise in atmospheric concentrations of carbon dioxide (CO_2), which along with other greenhouse gases, is the precursor to many of the environmental changes described above. Mangroves may respond directly to this increase in atmospheric CO_2 by increasing rates of carbon assimilation, growth and biomass allocation. Greenhouse studies have shown an increase in the amount of root productivity relative to allocation to above-ground biomass, which may be important with respect to soil formation and accretion relative to accelerated sea level rise. These responses depend on local conditions of soil salinity, which will control species-specific responses to increased CO_2 concentrations. However, in those environments where water use efficiency may be limiting mangrove growth, increased CO_2 will most likely enhance mangrove productivity.

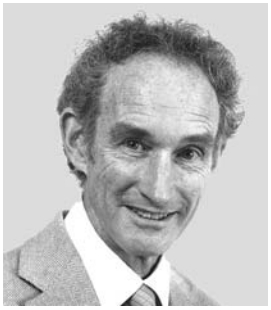
Mangroves are arguably an excellent indicator of how ecosystems will respond to the manifold impacts of global environmental change. The unique position of these forested wetlands along the continental margins of tropical environments integrates the combined effects of oceanic, atmospheric, and watershed forcings on ecosystem structure and function. Some scenarios of global environmental change will restrict mangrove ecosystems, such as increased erosion along the ocean margin, increased frequency and magnitude of cyclonic disturbances, and decreased moisture along the terrestrial ecotone. Other scenarios may expand mangrove-type habitats such as warmer environments at the latitudinal edges of mangrove global distribution, and increased CO_2 concentration. All of these environmental changes have to be considered in concert with human land use change both on the coast and in upland watersheds. The balance of energy and water between the ocean, atmosphere and watershed components of the Earth System will change several forcings that control the distribution, biodiversity, and function of mangrove ecosystems. Given present patterns, the combined effects of climate and land-use change will be noticeably evident in reduced goods and services to human systems throughout the tropics in this century.

Margalef, Ramon

(1919-)

Ramon Margalef, born in Barcelona, Spain, in 1919, an early self-taught naturalist, had to serve 6 years in the Army due to the Spanish Civil War. In 1944, his work on freshwater phytoplankton with a self-made microscope, while working as a clerk in an insurance company, attracted the interest of some naturalists. He received a grant and, in just two years, completed his secondary education with the highest score, then a degree in Natural Sciences at the University of Barcelona and a PhD (1951) with a thesis on *Temperature and Morphology of Living Things*. In 1952, he obtained a post as a researcher of the Consejo Superior de Investigaciones Cient ficas (CSIC) at the Instituto de Investigaciones Pesqueras (IIP, now Instituto de Ciencias del Mar), Barcelona, where he served as Director 1965–1967, launching many international contacts and programs. He became the first full professor in Ecology in Spain, at the University of Barcelona, but continued his research at the IIP until 1977. Since the late 1950s, he has contributed to theoretical developments in ecology, and his work has had an international impact. The most relevant of his contributions is in the application of Information Theory to the analysis of diversity in natural communities, and in the application of thermodynamic general principles to the understanding of succession and of ecosystems organization. As a limnologist, he recognized his debt to the ideas of A Thieneman and his friendship with V Tonolli and L Tonolli. He had many contacts within the ecology community, such as G E Hutchinson, R H MacArthur, E Odum and H Odum. His ideas on relationships between ecology and evolution, organizational patterns in the biosphere, the use of macroscopic parameters to study ecosystems and the need for a physical, rather than a mathematical approach to ecological understanding, have had a strong in uence on ecological theorists. Only a few of his synthesis books have been published in English (for instance, *Perspectives in Ecological Theory*, 1968; *Our Biosphere*, 1997). *Ecolog´a* (1974) and *Limnolog´a* (1983) are two monumental works that have been essential reading for Spanish ecologists, and the same is true for *Biosfera, Entre la Termodinámica y el Juego* (1980) and *Teor´a de los Sistemas Ecológicos* (1991), two beautiful theoretical synthesis. But he has also published an enormous amount of specific research, around 400 scientific papers and books mainly on oceanography and limnology. Even in retirement, he is still active. For some years, he has been the most quoted Spanish scientist in any topic. His work has been recognized with many national and international prizes (including the Huntsman Oceanography Award 1980, Premio S. Ramon y Cajal 1984, Italgas Prize 1989, Humboldt Award 1990, Naumann-Thieneman Medal 1989, Ecology Institute Prize 1995, several *Honoris Causa* doctorates and honorable memberships of the British Ecological Society and the Ecological Society of America, among others). Married in 1952, Professor Margalef has two sons and two daughters.

JAUME TERRADAS *Spain*



May, Robert, M

(1936-)

Robert McCredie May is a mathematical biologist. No aspect of global change is more prominent than the wholesale transformation of once species-rich ecological communities to depauperate ecosystems, and no single individual has contributed more to understanding this change than Sir Robert May. These ubiquitous changes in the earth's biota consist of extinctions and changes in species composition caused either by the substitution of native communities for managed ones such as farms or plantations or by the introduction of exotic species. Before the 1930s, the focus of ecology was on documenting natural history and chronicling patterns in the distribution and abundance of organisms. A J Lotka and V Volterra then provided a mathematical framework for understanding how such patterns may be generated by dynamic ecological processes. R H MacArthur dramatically extended their approach, providing elegant models that highlighted the ecological importance of structure and composition in communities. MacArthur in uenced many ecological and environmental biologists, including May, who led the way in restructuring ecology from a science that primarily described the earth's biota to one that predicts the causes and consequences of its change.

May's contributions are fivefold, all of which have and continue to contribute to every major area in global change ecology. These are the relationship between the

stability of ecological systems and its biological complexity (May, 1974), patterns in distribution and abundance (May, 1975), non-linear or chaotic dynamics in ecological systems (May, 1976), disease dynamics (May, 1977), and patterns in global extinctions (May, 1988). May has written hundreds of scientific and popular publications both individually and in collaboration with leading ecologists. The central theme of May's extraordinarily broad and diverse work is that the dynamics of the earth's biota, although seemingly complex beyond comprehension, are fundamentally understandable. Even the apparently chaotic dynamics and sometimes rapidly shifting compositions of biological communities we sometimes observe in nature can be understood by the theory May developed. This work provides the foundation for quantifying, modeling, and predicting the potential environmental consequences of anthropogenic impacts on populations, communities, and ecosystems.

His in uential contributions have been internationally recognized. May was elected to the American Academy of Arts and Sciences (1977), the Royal Society (1979), the Australian Academy of Sciences (1991), the US National Academy of Sciences (foreign member 1992) and to the Academia Europaea (1994). His contributions have earned him numerous honorary degrees around the globe, including England, Scotland, Sweden, the United States, and Australia. He was awarded the Royal Swedish Academy of Science's Crafoord Prize (1996) and the 1998 Balzan Prize for his contributions to the mathematical analysis of biodiversity. He was knighted in 1996 and awarded the Companion of the Order of Australia in 1998. Numerous awards preceded these honors, including the MacArthur Award by the Ecological Society of America (1984), the Croonian Lectureship of the Royal Society of London (1985), the Linnean Society's Zoological Medal (1991), and the Zoological Society's first Marsh Award for Conservation Science (1992) and its Frink Medal (1996).

May has and continues to serve on numerous key scientific and governmental boards. These positions have included those with the Natural History Museum, London; the Nuffield Foundation; the Royal Botanic Gardens, Kew; the Joint Nature Conservancy Councils; and as Chief Scientific Advisor to the UK Government and Head of UK Office of Science and Technology. Most recently he has been elected President of the United Kingdom's Royal Society. His insightful scientific contributions, of theory informed by data, continue to inspire the field of global environmental change.

REFERENCES

May, R M (1974) *Stability and Complexity in Model Ecosystems*, Princeton University Press, Princeton, NJ.

May, R M (1975) Patterns of Species Abundance and Diversity, in *Ecology and Evolution of Communities*, eds M L Cody and J M Diamond, Belknap Press of Harvard University, Cambridge, MA, 81–120.

May, R M (1976) Simple Mathematical Models with Very Complicated Dynamics, *Nature*, 261, 459–467.

May, R M (1977) Dynamical Aspects of Host-parasite Association: Crofton's Model Revisited, *Parasitology*, 75, 259–276.

May, R M (1988) How Many Species are There on Earth? *Science*, 241, 1441–1449.

SHAHID NAEEM AND SARA F TJOSSEM *USA*

Mercury in the Environment

Arne Jernelöv

International Institute for Applied Systems Analysis, Laxenburg, Austria

Environmental problems associated with mercury can be said to consist of several only weakly connected parts.

The one that was rst identi ed to be of environmental concern was the use of methyl mercury as a seed-dressing agent. It caused mortality and population decline among seed-eating and predatory birds in Sweden in the 1950s and 1960s. Human consumption of methyl mercury (CH_3Hg^+) dressed seed also resulted in the largest mercury-related catastrophe hitherto in Iraq, 1971- 1972. Its use ended in the 1970s.

Industrial use of mercury (Hg) and handicraft applications have for a long time been an occupational health concern. In the 1960s, it was also recognized that industrial discharge of mercury to nearby water courses resulted after bio-conversion in locally elevated areas of methyl mercury in larger predatory shes and a toxic risk to sh consumers. The Minamata catastrophe in the 1950s is the largest tragedy to humans from this type of mercury use. Industrial use and release of mercury has been drastically reduced in the last three decades following environmental concerns and governmental restrictions.

A technology with its roots in Roman times and one that was heavily used by the conquistadors in South America came back in large scale in the mid-1980s. Amalgamation as an extraction method of gold and silver replaced gold panning and resulted in extensive use of mercury in small-scale gold mining from the early 1980s to the late 1990s, predominantly in tropical countries. The mercury came to a large extent from inventories of the chemical industry in the

Organization for Economic Cooperation and Development (OECD) countries and the former Soviet Union. Following the fall in gold price, this practice has almost disappeared during the last few years.

A very large source of anthropogenic mercury is releases to the atmosphere from coal combustion and increased evaporation following land use changes. Those emissions have, together with atmospheric discharges from industrial uses, doubled mercury concentrations in the air and may have contributed to elevated concentrations in marine top predators. The current and forecast future trend for mercury release from these sources is upward.

Mercury as an environmental pollutant interacts with many other contaminants, e.g., eutrophication, acidi cation, and ground-level ozone. Thus acidi cation partly caused by sulfur emissions from coal combustion has led to an increase in methyl mercury concentrations in fresh water sh.

PHYSICAL PROPERTIES OF MERCURY AND GENERAL OCCURRENCE

Mercury (Hg) belongs to the metal group of basic elements. Within the group it is unique in that it is in liquid form and has a vapor pressure that results in evaporation at room temperature. Mercury has a high density ($13.7\,g\,ml^{-1}$) and surface tension and rare dielectric properties. From the elementary (metal) stage it can be oxidized to mono- or divalent form of which the latter is most common and it forms ions and salts (Turner, 1987).

Mercury in its divalent form binds very strongly to sulfide and selenide and forms salts with extremely low solubility (10^{-53} for mercuric sulfide). It also forms complex ions like HgS_2^{2-} and $HgCl_4^{2-}$.

In organisms, divalent mercury tends to bind to sulfur groups in proteins. Most of its toxic properties are associated with protein denaturation following such bindings. Higher organisms like mammalians have barrier-like blood/brain and placenta barriers that prevent divalent mercury from reaching the most sensitive places, the central nervous system and the fetus. Mercury also forms organic compounds through covalent bindings to carbon (C). Methyl mercury (CH_3Hg^+) is formed biochemically in nature and that compound, as well as many other organo-mercurials, is synthesized. Organo-mercurials like methyl mercury readily pass the protective barriers of the mammalian body and can thereby cause brain damage and harm to the fetus.

Dimethyl mercury ($(CH_3)_2Hg$) is highly volatile. In nature, mercury occurs predominantly as a mineral in the form of oxide (cinnabar) or sulfide.

Metallic mercury forms mutual solutions with many other metals like zinc or gold. These are called amalgams.

MERCURY USAGE FROM ROMAN DAYS TO CURRENT PRACTICE

The many unique properties of mercury have resulted in a very wide range of applications. Already, 2000 years ago, it was a well-known element and around the Mediterranean there were several established mercury mines. One of them – Almaden in Spain – has survived to the present day.

Those mines were based on cinnabar deposits where a proportion of the mercury occurred in elementary/metallic forms and frequently would bleed out of the rocks in liquid form. Consequently, the air in the mines would contain high concentrations of mercury from evaporation. High mortality from mercury poisoning among slave miners contributed to the relatively high cost of mercury in Roman days. The dominating uses at the time were two: metallurgic and pharmaceutical. In metallurgy, mercury was used to extract gold and silver from grained minerals in the form of amalgams. Later the amalgams were heated and the gold or silver retrieved. This technique was also used by the conquistadors in Latin America and again came into use in small-scale gold mining in the late 20th century.

The toxic properties of mercury were used early on in various pharmaceutical applications against a large number of diseases and illnesses. After the discovery of the Americas and the introduction of syphilis in Europe, mercury became a well-known standard treatment of that disease. Some scholars argue that some of the classical symptoms of syphilis are very much like those of mercury poisoning and could indeed be the result of the treatment rather than of the disease.

Mercury as mercuric chloride ($C_7H_{14}ClHg$) or sublimate was also a popular poison for many countries as humans cannot smell or taste it and it therefore was regarded as undetectable. Among widespread historic uses was also the mixing of mercury salts into starch for uses in wet weather or under damp conditions, e.g., on top hats where the starched clothing would otherwise quickly mould. Mercury poisoning therefore became a widespread occupational risk for people in that trade and the mad hatter as in *Alice in Wonderland* became a well known phenomenon.

With industrialization and electrification, mercury found a large number of new applications. The specific electric properties gave rise to many functions in electrical applications such as switches, batteries, light bulbs and tubes. For some of these, it has been especially difficult to find substitutes with similar performance. Other uses were as catalyst in industrial syntheses such as formaldehyde (CH_2O) and acetaldehyde (C_2H_4O) or as electrolyte in the chlor-alkali process when common salt (NaCl) is converted to alkali (NaOH) and chlorine gas (Cl_2).

The mercury inventory in chlor-alkali plants has frequently been very high, often in the range of several hundred tons in a single room. Spillage of mercury often impregnated oors and walls and resulted in high ambient

air concentrations with subsequent exposure of workers. Medical symptoms of mercury intoxication were common in chlor-alkali plants for many decades.

Industrially produced organo-mercurials found uses as seed-dressing agents for fungi-control (methyl-mercury) or as a slimecide in pulp and paper industry (phenyl-mercury).

Mercury is mined predominately from two types of deposits:

1. Cinnabar ores where mercury is the direct target of the mining process; and
2. sulfide ores with a variety of metals where some other – often copper – is the main product and mercury only the by-product.

Global mercury production peaked at some 10 000 tons annually around 1970 and has since fallen sharply as a consequence of the environmental and health concerns. Cinnabar-based mercury mines have been closed or nearly so and it is only the sulfide ore-based mines that continue to yield mercury at a rate of some 500 tons year^{-1}. As other forms of mercury mobilization by humans – most notably coal burning – release mercury at the rate of some 10 000 tons annually, mining is no longer a major source of anthropogenic mercury (Lindqvist *et al.*, 1984).

ENVIRONMENTAL CYCLING AND (BIO)CHEMICAL TRANSFORMATIONS

Mercury has several environmental cycles that are partly inter-linked and lack of which is based on biochemical or chemical transformations. One of the cycles is global or semi-global in nature and involves evaporation from land or sea of elementary mercury oxidation in the atmosphere to divalent mercury (H^{++}) and subsequent wash-out with precipitation. From both land and sea, mercury can then be either quickly re-emitted to the atmosphere or more permanently bound to organic or inorganic receptors. There is also an atmospheric cycle involving emissions of dimethyl mercury mostly formed in biochemical processes in the sea. In this case the main atmospheric conversion process is probably one involving ultraviolet (UV) light and/or acids (H^{++}) resulting in water soluble mono-methyl mercury (CH_3Hg^+) and subsequent washout with rainwater. In both cases, there is also dry deposition and re-emittance of unknown magnitude.

In aquatic systems, mercury also undergoes a cycling process. After introduction in a water course, mercury in divalent inorganic form binds to particles and through this attachment settle in sediment. While suspended in the water or placed in the upper part of the sediment, different types of microorganisms, both bacteria and fungi, may methylate mercury (Jensen and Jernelov, 1969). Depending on pH for example, the end product can be either mono- or dimethyl mercury. The process of biological methylation is especially fast in the low oxygen concentration range where sulfate-reducing bacteria thrive.

Some biochemical mechanisms for biological methylation of mercury have been described. Among them are methylcobalamin (B_{12}) mediated methylation in methanogenic bacteria and erroneous synthesis of methionin from homocystein when mercury binds to the target sulfur (S) group in Neurospora (Landner, 1971).

There are also bacteria that convert methyl mercury to inorganic mercury, as do seals and fish-eating birds. Dimethyl mercury largely evaporates into the atmosphere while mono-methyl mercury is readily picked up by living organisms. Through the food chain, a substantial fraction of the methyl mercury in for example a lake will find its way to fishes.

Although fishes have fairly efficient excretion mechanisms for inorganic mercury, they can neither excrete methyl mercury nor degrade it. Methyl mercury concentrations in individual fishes thus mostly increase with size and age. As a consequence, methyl mercury constitutes a large fraction of the total mercury body load of fishes – mostly over 90% (Bloom, 1992). In fish, most of the methyl mercury is bound to sulfur groups in muscle-proteins. Certain organs, especially the parapharyngeal or pseudo-bronchial one show extremely high methyl mercury concentrations such as two orders of magnitude more than the whole fish average.

Only in extreme situations with methyl mercury concentrations exceeding 20 parts per million (ppm) have fishes demonstrated symptoms of poisoning. Fish consumers are at risk much earlier. Most countries have health standards banning sale at 0.5 or 1 ppm. Fish disappear from acidifying lakes because of difficulties in reproduction. In some lakes whole age classes are missing.

HUMAN MERCURY TOXICOLOGY AND MASS MORTALITY FOLLOWING EXPOSURE TO METHYL MERCURY

The effects of mercury on humans as well as other mammalians or birds are highly dependent on the form of mercury and route of exposure. As noted above, mercury in the body binds predominantly to sulfur-groups in proteins; and effects, e.g., on enzymes, are most often associated with protein denaturalization. Inorganic divalent mercury does not pass the blood brain or placenta barriers. The cause of death when mercuric chloride was used as a poison in, for example, renaissance Venice, was due to damage to liver and kidney functions. Inorganic divalent mercury is excreted with urine with a half-life in the body of weeks.

The lungs effectively take up elementary mercury when inhaled while mostly passing through the intestinal tract if swallowed. After lung uptake, elementary mercury is

oxidized in the body to divalent mercury and thereafter behaves as such. Prior to oxidation, though, elementary mercury may pass the blood-brain barrier. Thus damage to the central nervous system is part of the symptom package associated with exposure to elementary mercury in air.

Methyl mercury is readily absorbed in the intestinal tract or through the skin if applied in solution, and via the lungs if present (as dimethyl mercury) in air. In methylated form, mercury easily passes all types of biological membranes and readily reaches the brain or fetus. In modeling methyl mercury, the body can be treated as one compartment due to the fast equilibration (Cox *et al.*, 1989).

Symptoms of methyl mercury poisoning are associated with damage to the central nervous system, as the effects here occur at much lower concentrations of mercury than those of other organs in the body. Especially sensitive is the brain of the fetus.

Methyl mercury is excreted with the bile but reabsorbed in the intestine. Excretion also occurs via milk in lactating females and through irreversible incorporation in hair and nails. The half-life in the human body is typically some months.

The first environmental catastrophe involving mercury and mass mortality among humans occurred in the 1950s in Minamata in Japan. Among the fishing population in the small town and around the bay, severe cases of central nervous system dysfunctioning including loss of balance, ataxia, cramps and blindness frequently resulting in death occurred from the mid-1950s. From observations of similar symptoms among domestic cats, it became clear that fish were the vehicles for the causative agent, but it took until 1959 before it was concluded that methyl mercury was the cause of the poisoning.

Several hundred people were killed or permanently damaged. Some of them still live at a special local hospital. The mercury that contaminated fish and shellfish in Minamata Bay came from a chemical factory making acetaldehyde with mercury as a catalyst. Some of the mercury was chemically methylated in a holding pond where different ef uents were mixed before discharge to the bay. The exact level of mercury and methyl mercury in aquatic life in the bay at the time has been much debated since the colorimetric analytical method available at the time was not reliable. Concentrations well above 10 ppm were in all likelihood common. A similar case occurred at the same time in Niigata in Japan involving fewer cases.

In 1971/1972, the by far largest mercury catastrophe up until that time occurred in Iraq. After some years of bad harvests Iraq, in 1971, decided to import Green Revolution wheat from Mexico. To protect that seed from fungal attacks, it was dressed with methyl mercury. The distribution to the farmers took longer than planned so when the pink colored grain arrived, the sowing season was already over. The farmers were told not to eat the seed but the previous year's bad harvest and a general mistrust of governmental instructions caused many of them to try to feed chicken and sheep and sometimes cattle. When no short-term effects were seen, they progressed to feeding expendable family members like grandparents and soon whole families were eating pink bread.

By the time that humans started to show symptoms, many food products on the markets like meat, milk, cheese, chicken and egg were also contaminated. A strict governmental order was issued that all pink grain within a fortnight should be handed in to the storehouses where the farmers had obtained it. Failure to comply would be punished by death. But farmers only got the information after the death penalty had already come into force. Thus methyl mercury dressed seed was dumped in irrigation canals and along roadsides, which resulted in contamination also of fish and wild birds, which suddenly became easy to catch.

In total some 10 000 people died and maybe 100 000 were permanently injured. A secretive governmental policy lead to very limited publication of this catastrophe compared to the one in Minamata.

MERCURY IN TERRESTRIAL AND AQUATIC FOOD CHAINS

In the 1950s and early 1960s, methyl mercury was extensively used as a seed-dressing agent in Sweden. Environmental concerns arose when populations of yellow sparrow, pheasant and other seed-eating birds declined and individuals of the species were found unable to y, moving in circles or paralyzed. Predatory birds were also found dead without physical injury and populations were steadily shrinking. The cause was methyl mercury uptake from the dressed seed in the seed-eating birds and further food chain magnification. Similar patterns were found in a variety of aquatic environments from small lakes and rivers to marine bays and semi-enclosed seas to open ocean systems. In fact, methyl mercury in aquatic systems is one of the very best examples of bio-magnifications in food chains, as fish lack excretion mechanisms for methyl mercury. This is also the reason why other naturally occurring mercury forms do not show the same pattern.

The original studies of mercury in the environment were carried out in temperate systems in Scandinavia, Japan and North America, in the 1960s and 1970s. When mercury use and release started to expand in tropical and subtropical regions in connection with the use in small-scale gold mining, some reported analytical results did not match the established picture. An International Council of Scientific Unions/Scientific Committee on Problems of the Environment (ICSU/SCOPE) study was launched to look at the fate and effects of mercury in tropical systems. With regard to food chains, one structural difference between tropical and temperate ecosystems that significantly affect mercury

bio-magnification is that tropical predators frequently are specialists while those in temperate areas often are generalists. The latter results in relatively small differences in methyl mercury concentrations between individuals and between closely related species. In tropical systems with geographically specialized individuals and prey-specialized species, the variations in (methyl) mercury concentrations are much larger with consequences for required sample sizes for trend analyses and for establishment of statistical differences (Jernelov and Ramel, 1995).

In open ocean systems, large and long-lived top predators like swordfish, tuna and sharks have high concentrations of methyl mercury that frequently reach several ppm. Mediterranean fish show higher concentrations than their relatives from the Atlantic Ocean, which may re ect both a different size/age relationship depending on the more oligotrophic conditions in the Mediterranean and mercury rich geological strata in the region.

Seals also have high concentrations of mercury but practically all of it is in inorganic and selenium-bound form, despite the fact that the fish on which seals feed contain predominantly methyl mercury.

As already noted, methyl mercury differs significantly from inorganic forms of mercury with regard to bioaccumulation and food chain magnification. However, in fish where methyl mercury is the dominating form of mercury, current methods of mercury analysis (ameless atomic absorption) cannot distinguish between inorganic and methylated mercury. In other situations the analytical method used may blur the results and complicate interpretation as the proportion of methyl mercury in different samples remains unknown.

An illustration of the importance of the food chain is given by the situation in Scandinavian acidified lakes. As a rule of thumb it can be said that in a lake with pH below six, a 1 kg pike will have a mercury level above 1 ppm. Many factors of both chemical and biological nature contribute to this. The single most significant factor is that the food web is significantly altered when roach disappear as they are more acid sensitive, and pike start to feed mainly on perch, which through their feeding habits and growth rate have higher methyl mercury concentration than roach.

MERCURY IN SMALL-SCALE GOLD MINING

In the early 1980s, mercury had replaced the panning process for density-dependent separation in small-scale gold mining. Amalgamation allows capture of gold-containing particles in the dust – the very fine particles – that could never be gravimetrically separated in a pan, and thus the mercury process gives a substantially higher yield than panning. Sometimes, depending on local conditions, the increase in yields may be as high as ten-fold. Mercury and amalgamated gold is then separated by putting the mixture in a cloth and squeezing the mercury droplets out. The amalgam is then heated which evaporates the mercury leaving the gold (Kato and Pfeiffer, 1996).

Occupational mercury exposure occurs throughout the process but is especially pronounced when scraping gravel off the mercury bed or with a modified technique when gravel and mercury are mixed. Most of the immediate intoxication occurs in connection with heating the amalgam and evaporating the mercury as this is frequently done indoors while hiding to avoid robbery.

Environmental contamination occurs both through mercury spillage on the ground and in water courses, as riverbeds frequently are favored places for gravel extraction, and to the atmosphere through evaporation.

The practice started in the Amazon region in Brazil and quickly grew with regard to the number of practitioners and spread to neighboring countries in the region. In the late 1980s, several hundred thousand people in the Amazon region were using the technique.

In the early 1990s, the mercury technique spread to Africa, in particular Tanzania and Ghana. A little later, a boom came to South Africa when people re-extracted old tailings from gold mines with the technique.

The global spread continued to Asia, first to Vietnam and Indonesia and also reaching Papua, New Guinea.

In the mid-1990s, an estimated ten million people were earning their livelihood from small-scale gold extraction using the mercury method.

As gold production or mercury use figures from these mostly unauthorized activities is non-existent or non-reliable, the SCOPE group used an indirect method to access mercury use. Small-scale gold miners tend to get 15–25% of the world market price when they sell their unrefined gold. To cover their costs of living and support families they would in the Amazon region be required to produce one gram of gold per garrimpero per day. As the mercury requirement for gold production is 1.5–1, each miner would release 0.5 mg of mercury per year. Thus the mercury used in the trade globally is estimated to amount to 5000 tons annually in the mid-1990s.

In the late 1990s, the gold market collapsed and the small-scale miners were hit by a price drop of frequently 50%. This put most of them out of business and a rough guess is that today their number does not much exceed one million – most of them in Asia.

Another question that the SCOPE group studied was the source of the 20 000 tons or so of mercury that were used from the early 1980s to the late 1990s. The conclusion was that it came from industrial mercury inventories, first from OECD countries when, for example, their chlor-alkali plants shifted from mercury to other techniques for environmental reasons. Later the bulk of the mercury came from the inventories of Soviet industry through private profiteers (Jernelov and Ramel, 1997).

CURRENT SCIENTIFIC CONTROVERSIES

Effects on Humans of Long-term, Low-level Exposure from Marine Food

In connection with the mercury poisoning catastrophe in Iraq, it was observed that schoolchildren exposed to methyl mercury showed no apparent physical symptoms but they performed badly in school in the years following exposure. This and other indications together with more theoretical considerations based on mercury action in nerve cells have led to concern about developmental effects in children from long-term, low-level, methyl mercury exposure. A special focus has since been placed on populations with a high general intake of marine food products containing significant amounts of methyl mercury. Two large such studies have yielded con icting results.

One study was carried out in the Seychelles where tuna with methyl mercury concentrations of 1–3 ppm is the staple food for half the year. This study involved a large number of children who were followed from birth. Developmental indicators like age of walking were noted and the children were tested using a series of standard methods. No development effect significantly correlated to mercury exposure was found (Davidson *et al.*, 1995).

The other study was undertaken on the Faeroe Islands. Here traditional whaling in the bays provides a substantial part of the local diet, with most of the meat having high mercury concentrations. Also here, a group of children was followed from birth. Developmental indicators were noted and the children tested with especially developed or modified tests. This study reported that delayed development was correlated to mercury exposure (Budtz-Jorgensen *et al.*, 2000; Grandjean *et al.*, 1997).

The debate has focused on the sensitivity and reliability of the methods used to test the children. Much less attention has been given to the difference between tuna and whalemeat. The underlying sensitive question – both from an economic and a cultural point of view – is whether mercury concentrations in those traditional marine foods render them unsuitable as staples.

Resident Time of Mercury in the Atmosphere

The annual calculated anthropogenic mercury emission to the atmosphere plus the estimated natural degassing from the earth more or less correspond to the total amount of mercury present in the atmosphere. Thus the apparent resident time of mercury in the atmosphere should be in the order of one year. From this it follows that mercury in the atmosphere stays there long enough to have at least a hemispheric distribution bordering on a global one. Like carbon dioxide (CO_2), locations of emissions should thus be of little relevance for local and regional fall-out (Fitzgerald, 1986; Fitzgerald *et al.*, 1991).

Many studies over the years have on a smaller scale yielded results that fit poorly into this picture. Atmospheric emissions from chlor-alkali plants in Sweden were shown to contribute to locally elevated deposition. Mercury in rainwater falling on lakes was calculated to be re-emitted very quickly to a large extent. After closure of a major East German chemical industry after re-unification, mercury fall-out in Southern Sweden markedly decreased.

To reconcile those calculations and studies, a *ping pong ball model* was proposed. In this model, after emission to the atmosphere, mercury will have a resident-time much shorter than one year but will have a high probability of re-emission (to bounce back up) upon deposition. The picture prepared was one of ping pong balls on a stone oor with some soft carpets here and there. The calculated one-year residence-time would then be the average time the balls kept bouncing before coming to rest rather than the average time that the mercury on the average stayed in the air before hitting the oor (Jernelov, 2000).

The soft carpets in the real world would be, for example, acid lakes or arctic areas where chemistry and temperature reduce the probability of re-emission. They would accumulate mercury not because of a high fall-out but because of low re-emission. The two models obviously lead to very different results when assessing local and regional consequences of mercury discharges to the atmosphere, e.g., when preparing a formal environmental impact assessment (EIA). In addition, detailed analyses of wet and dry deposition and of (re)emission show that there are other measurable differences between the two models, sometimes referred to as the balloon (stay up a long time) and the ping pong ball (bounce back up):

- A much larger part of atmospheric mercury must consist of water soluble or reactive forms in the ping pong ball than in the balloon case.
- Reduced mercury use and emissions in recent decades should result in a steeper trend of decreasing mercury concentrations in fish in lakes close to industrial centers than far away in the ping pong ball case with little or no difference if the balloon model is valid.

Mercury Toxicity Risks with Dental Amalgam

Amalgam has been widely used for dental fillings on a large scale for more than a century. Numerous studies by national health authorities and international organizations have concluded that this practice is safe and that side effects are small and rare. During recent decades a growing number of dental patient associations have claimed the opposite, with the support of some scientists.

Recently the Swedish Counsel for Planning and Coordination of Research (FRN) was commissioned by the Swedish government to review the question once more (FRN, 1998). Among the conclusions from that study were

that documented individual variations, both in mercury leakage from amalgam-fillings (caused for example by saliva pH and grinding jaw movements) and sensitivity to mercury were larger than previously accounted for, and that the possibility of existence of a fairly large group of people with medical symptoms caused by amalgam fillings could not be ruled out. The Council also concluded that very different risk philosophies rule in the environmental domain where mercury bound to sulfur or as amalgam is not considered safe for storage above ground (also when cast in concrete) and the dental domain where amalgam in the mouth of most people is considered safe. Clearly from an environmental point of view, mercury emissions from crematoriums are an important anthropogenic mercury source.

Short Distance Air Transport

Several studies in the Amazon region have indicated or yielded results that have been interpreted to indicate a substantial degassing of mercury from rain forest soil and a subsequent trapping of that mercury in the foliage of the trees.

There are no supporting studies from other areas and the existence of a significant short air cycle is debated.

MERCURY TURNOVER AND GLOBAL ENVIRONMENTAL CHANGE

Mercury undergoes several environmental cycles involving various chemical transformations between elemental and oxidized stages and between inorganic methylated forms.

Obviously temperature is a rate determining factor for many of these processes. However, the complexity is such that no simple conclusions can be drawn. In very general terms, it can be said that a given amount of mercury tends to give rise to higher methyl mercury values in predatory fishes in cold and temperate climates than in tropical ones but compounding factors are many, e.g., nutrition and growth rate.

Globally, mercury cycling is thought to involve transport towards the polar regions. Warming could increase re-emission to the atmosphere of some of that mercury. Would it then just move a little farther north (and south) or would it expose ecosystems globally? Atmospheric/oceanic mercury exchange has been demonstrated to involve large local uxes, but these go both ways and upwelling may be a more important factor than temperature. Thus, if global change involves ocean circulation on a large scale it will in uence mercury cycling too, but it is too early to say how.

Acidification clearly has a strong effect on mercury turnover and resulting mercury levels in fish. More acidification means higher methyl mercury levels (in less fish). Acidification is caused by sulfur emissions, and is associated with coal burning; this may be further emphasized as coal burning today is a major source of anthropogenic mercury emissions. This could be the most direct link between other aspects of global environmental change and mercury. Increased emissions from coal burning is the causative agent behind many of them.

High mercury levels in open ocean top predators like swordfish, tuna and shark have long been seen as a natural phenomenon. Computations within the framework of the ICSU/SCOPE study indicate that anthropogenic emissions may have contributed by a factor two, not only to atmospheric concentrations but also to that of the upper layer of the oceans. Declining anthropogenic mercury emissions from industrial sources may, with a time delay of a decade or two, result in lower concentrations of methyl mercury in pelagic marine organisms – unless other emissions notably from coal burning increase correspondingly.

Besides the generally higher speed of biological and chemical processes associated with higher temperatures and the other factors discussed above, global climate change may affect mercury cycles in the following ways:

- Mercury methylation rates are frequently especially high in newly ooded areas.
- Increased ooding may therefore contribute to formation of more methyl mercury and thus to higher levels in fish. Increased UV-radiation may contribute to the oxidization of elementary mercury in the atmosphere thus shortening the residence time and speeding up washout.

See also: Environmental Cadmium in the Food Chain, Volume 3.

REFERENCES

Bloom, N S (1992) On the Chemical form of Mercury in Edible Fish and Marine Evertebrate Tissues, *Can. J. Fish. Aquat. Sci.*, 49, 1010–1017.

Budtz-Jorgensen, E, Grandjean, P, Keiding, N, White, R, and Weicke, P (2000) Benchmark Dose Calculations of Methyl Mercury – Associated Neurobehavioral Deficits, *Toxicol. Lett.*, 112–113, 193–199.

Cox, C, Clarkson, T W, Marsh, D O, Amin-Zaki, L, Tikriti, S, and Meyers, G G (1989) Dose Response Analysis of Infants Prenatally Exposed to Methyl Mercury: an Application of a Single Compartment Model to Single Strand Hair Analysis, *Environ. Res.*, 49, 318–322.

Davidson, P W, Myers, G, Cox, C, Shamlaye, C, Marsh, D, Tanner, M, Berlin, M, Sloane-Reeves, J, Cernichiari, E, Choisy, O, Choi, A, and Clarkson, T (1995) Longitudinal Neurodevelopment Study of Seychellois Children Following In Utero Exposure to Methyl Mercury from Maternal Fish Ingestion: Outcome at 19 and 20 months, *Neurotoxicology*, 16, 677–688.

Fitzgerald, W F (1986) Cycling of Mercury between the Atmosphere and the Oceans, in *NATO Advanced Science Institutions Series C*, Reidel, Dordrecht, 363–408.

Fitzgerald, W F, Mason, R P, and Vandal, G M (1991) Atmospheric Cycling and Air-Water Exchange of Mercury over Mid-Continental Lacustrine Regions, *Water, Air and Soil Pollution*, 56, 745–68.

FRN (1998) *Amalgam and Health: Risks in a New Light*, Report 98 : 22, IBSN 0348: 3991.

Grandjean, P, Wecke, P, White, R F, Debes, F, Arak, S, Yokoyama, K, Murata, K, Sorensen, N, Dahl, R, and Jorgensen, P (1997) Cognitive Deficit in Seven Year Old Children with Prenatal Exposure to Methyl Mercury, *Neurotoxicol. Teratol.*, 20, 417–428.

Jensen, S and Jernelov, A (1969) Biological Methylation of Mercury in Aquatic Organisms, *Nature*, 223, 5207.

Jernelov, A (2000) Some Comments on Biological Methylation, Global Emissions and Trade and on Atmospheric Residence Time of Mercury, *J. Environ. Sci.*, 12, 102–107.

Jernelov, A and Ramel, C (1995) SCOPE Project; Evaluation of the Role and Distribution of Mercury in Ecosystems with Special Emphasis on Tropical Regions, *Ambio*, 24(5), 319–320.

Jernelov, A and Ramel, C (1997) The Siberian Mercury Anomaly and a Tentative Answer to the Missing Mercury Problem, *Ambio*, 27(2), 155.

Kato, H and Pfeiffer, W C, eds (1996) Proceedings of the International Workshop on; Environmental Mercury Pollution and its Health Effects in the Amazon River Basin, Rio de Janeiro, Brazil.

Landner, L (1971) Biochemical Models of the Biological Methalation of Mercury Suggested from Methylation Studies in Vivo with Neurospora Crossa, *Nature*, 230, 452–453.

Lindqvist, O, Jernelov, A, Johnansson, U, and Rodhe, H (1984) *Mercury in the Swedish Environment - Global and Local Sources*, SNU Report No. 1816, IBSN 91 7590 154 4.

Turner, D R (1987) Specialisation and Cycling of Arsenic, Cadmium, Lead and Mercury in Natural Waters, in *SCOPE 31: Lead, Mercury, Cadmium and Arsenic in the Environment*, 175–186.

Metadata

Jonathan M O Scurlock[1], Paul Kanciruk[1], Raymond A McCord[1], Richard J Olson[1], and William K Michener[2]

[1] *Oak Ridge National Laboratory, Oak Ridge, TN, USA*
[2] *University of New Mexico, Albuquerque, NM, USA*

Metadata (also written sometimes as meta-data or meta data) is information that enables data users to locate, acquire, understand, and apply or manipulate data. In other words, it documents data characteristics so that users can determine the tness of the data set for their purpose. A brief de nition of metadata is therefore data about data, or information about data. However, this encompasses a wide range of possibilities ranging from human-generated textual description of a resource to machine-generated information that may be useful only to computer software applications. Metadata describes the content, quality, condition, lineage, and other characteristics of data. Metadata makes it possible for data users to search, retrieve, and evaluate data set information from networks of databases by providing standardized data descriptions.

The term *metadata* may be interpreted in a number of ways. In its simplest and most restricted interpretation, metadata is the information necessary to locate data. This may be as simple as a unique resource locator (URL, or web address) where data may be found on the Internet, or it may comprise more detailed search and order criteria (e.g. spatial or temporal attributes, measurement characteristics or types, keywords). An analogy of this interpretation is the information contained in a library's card catalogue of books: author, title, subject, year published, publisher, etc. The generalized metadata standards developed by the Dublin Core Metadata Initiative are such an example. But although this information is effective in identifying the data of interest, it would be only marginally useful in understanding or interpreting the data. A more comprehensive interpretation of the term *metadata* includes all the information necessary not only to find the data of interest, but also to fully understand and effectively use these data. These include, for example, details of methods used to collect data, data formats and pointers to further information (references, etc.).

Metadata serve to describe data sets, databases, or information products (such as maps or technical reports) by documenting such things as subject matter, data set contents, context, quality, structure and accessibility. This includes the basic who, what, when, how, where (and sometimes why) of the data, their accuracy, and availability and distribution information. Metadata occur in layers with some parts being unique, and very intimately and specifically related to a measurement (e.g., the exact time and place of a single measurement). Other parts are more general and less directly connected to an individual measurement (e.g., a technical discussion of how the data-collecting instrument works). Other metadata may have intermittent and irregular interactions with the data (e.g., the effects of corrective maintenance activities).

Metadata is a term of relatively recent origin, probably first used during the 1980s. Over the past 15 years its usage has increased, and has become particularly common with the rising popularity of the World Wide Web. Until the mid-1990s, it was a term most prevalently used by communities involved with the management and interoperability of geospatial data, and with data management and systems design and maintenance in general. For these communities, *metadata* refers to a suite of industry or disciplinary

standards, as well as additional internal and external documentation and other data necessary for the identification, representation, interoperability, technical management, performance, and use of data contained in an information system. However, the underlying concepts have been in use for as long as collections of information have been organized. Library catalogs represent a well-established variety of metadata that have served for decades as collection management and resource discovery tools.

METADATA FOR GLOBAL CHANGE SCIENCE

Environmental monitoring and research programs depend on being able to store, retrieve, and understand data that are collected over long periods. Long-term accessibility of high-quality data is hindered by data entropy, the natural tendency for data to degrade in information content over time (*see* Data Banks: Archiving Ecological Data and Information, Volume 2 and references therein). In order to minimize data entropy, data require comprehensive documentation (i.e. metadata), as well as storage in secure data archives.

Investing adequate time and energy in metadata development is beneficial for several other reasons. Data sharing is facilitated; data previously collected for a different purpose can be mined to answer new questions. Descriptions of sampling and analytical procedures, data quality and data set structure, facilitate subsequent analyses and interpretations. Many of the questions now being addressed by environmental scientists require more data than could be realistically collected, managed and analyzed by a single individual.

Various groups of computerized data users are developing metadata standards to increase interoperability and efficiency within and between different information technology organizations, software products, etc. Industry groups, international standards organizations and new associations such as the World Wide Web Consortium (W3C) all have a stake in the setting of metadata standards, as well as those charged with the management of scientific, and, more specifically, environmental data (see Table 1).

Considerable resources have been devoted to developing, adopting and implementing spatial metadata standards. In 1994, the US Federal Geographic Data Committee published standards which were developed as part of the US National Biological Information Infrastructure, to standardize geographical data collected by the federal government. However, metadata standards for non-geospatial data typically do not exist in any accepted format beyond individual studies, projects or organizations. Environmental studies often require large amounts of diverse data related to the physical and chemical environment, as well as the organisms, populations, communities and ecosystems comprising the biotic portion of the environment. It is unlikely that a single metadata standard, no matter how comprehensive could encompass all environmental data: but some metadata guidelines have nevertheless been developed recently, as well as computer software tools for compiling and submitting metadata. Certain special metadata may be required which are peculiar to certain kinds of environmental data, e.g., consistent taxonomy and nomenclature for species, families, etc., whilst other environmental data will share metadata standards with other kinds of geospatial and technical data.

Metadata for environmental data include a particularly wide range of context information that helps to define the meaning of an environmental parameter. An example is provided here by temperature: the contextual information needed to correctly interpret this common parameter includes metadata about:

- what media (air, water, soil, etc.);
- what units (C, K, or F);
- where (latitude, longitude, height above or below the Earth's surface);
- who (name of observer, institution, storage location, etc.);
- when (integration interval, date and time);
- how (what measuring device, what precision);
- how good (methods of quality assurance, what precision, etc.).

This contextual information can be extended to include more detailed information about the primary metadata values (e.g., how was the location determined, what units are used for the coordinate system, whether time is given as local time, Greenwich Mean Time (GMT), Coordinated Universal Time (UCT) or possibly the proposed Universal Date and Time standard, what other parameters were measured at the same time and place, what other programs

Table 1 Examples of metadata descriptions and standards available online

Name	Web URL
Dublin Core Metadata Initiative	http://purl.org/dc/
Getty Research Institute (USA) Standards Program	http://www.getty.edu/gri/standard/intrometadata/
US Federal Geographic Data Committee Biological Data Working Group	http://biology.usgs.gov/fgdc.bio/index.html
US National Biological Information Infrastructure	http://www.nbii.gov
World Wide Web Consortium	http://www.w3.org

are related to this research project, etc.). Context may be further expanded if the data represent results taken from a sample. In this case, information about sampling scheme, sample size, and sample preparation are also part of the metadata. Some parameters may even require other measurements as part of their metadata (e.g., dissolved oxygen is more meaningful when water temperature is also measured). The data themselves may even become metadata for some applications. For example, cloud cover is a characteristic of LANDSAT satellite images, but this information becomes a valid metadata variable itself when users want to select only those images that are cloud free.

The adequacy of metadata is determined ultimately by future data users. If they are familiar with the specific scientific discipline and/or the research project or investigator, the users can successfully guess the correct values for incompletely specified metadata. If the users are unfamiliar with the discipline, incomplete metadata will significantly limit the usefulness of the results. Good metadata will minimize user problems with undocumented or misunderstood data, incorrect linkage of data and metadata, as well as the increased effort required to make later corrections, and consequently wasted time in the user community. It will result in efficient data processing and communication with the data user, and improved information longevity. As information technology rapidly increases the ability to store, share, and process environmental information, all scientists that gather and record data should more rigorously challenge the sufficiency of their own metadata.

Metapopulations

Pablo A Marquet
P. Universidad Católica de Chile, Santiago, Chile

A metapopulation is an ensemble of interacting populations connected by dispersal of individuals among them. Many species exist as metapopulations, and more are expected to do so, considering that landscapes are becoming increasingly patchy through habitat loss, degradation, and fragmentation, one of the most important of the many interacting components of global change affecting ecological systems.

In this essay a broad overview of metapopulation models and theory, from an ecological point of view, is presented, with emphasis on those concepts and approaches most important to understand the consequences of changes in the global environment. Global change through fragmentation can increase metapopulation extinction by: (1) destroying key habitat patches, which albeit their low quality (sinks), may be essential to maintain the connectivity of the system or by destroying those of high quality (sources) that are important sources of new colonists; and (2) by altering the natural dynamic regimes of habitat patches within landscapes. The effects of these changes may be subtle, lagged in time, and usually with sharp threshold transitions. Hence it is essential to establish long-term landscape monitoring programs to ensure metapopulation persistence in the face of a global change in the environment.

INTRODUCTION

Individuals and the habitats they occupy are not homogeneously distributed in space. Population ecologists have recognized the importance of spatial heterogeneity for a long time (e.g., Gause, 1935). However, the study of temporal uctuations in the number of individuals within populations has historically received most attention from both theoretical and empirical ecologists, likely as a result of the complexity involved in the consideration of space in theoretical models and experiments. Researchers following this non-spatial approach pictured populations as closed systems that changed in size as a consequence of births and deaths that occurred locally. However, the observation that populations can go extinct at particular locations and later reappear owing to colonization events, or can persist even if additions of new individuals through births do not compensate for losses through death, pointed to the importance of considering the exchange of individuals among populations in order to account for their dynamics and persistence. Under this view, populations are open systems, not independent from nearby populations with which they interact through the exchange of individuals. This ensemble of interacting populations (Figure 1) is what is called a *metapopulation* or *population of populations* as epitomized by Richard Levins in two foundational papers (Levins, 1969, 1970). Under the metapopulation paradigm, in addition to the importance of local populations, there is an added emphasis on the dynamics of the ensemble of local populations and the way they interact through migration.

Many species exist as metapopulations, and more are expected to do so, considering that the distribution of habitats within landscapes is becoming increasingly patchy through habitat loss, degradation, and fragmentation. Fragmentation in particular is the largest, and one of the most important, of the many interacting components of global change affecting ecological systems (Vitousek, 1994). Recent analyses of global databases on human disturbance of natural habitats point out that the effects of these degradation processes are acute in different biomes (Groom and Schumaker, 1993; Hannah *et al.*, 1995). Major ecosystems have already been reduced to fragments and numerous species and genetically distinct populations have been lost in the process (Ehrlich and Wilson, 1991). The effects

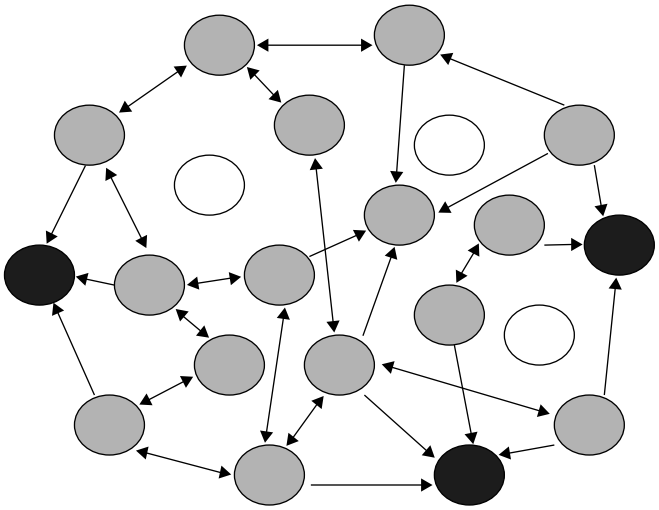


Figure 1 An ideal metapopulation composed of several local populations (or subpopulations) interacting through migration of individuals among them. Empty habitat patches, susceptible to be colonized from nearby local populations, are shown in white. Local populations whose persistence depends on immigration from nearby local populations (sink populations) are shown in black, while sources or immigrants (source populations) are shown in gray

of habitat loss and fragmentation are pervasive, featuring as the primary causes of endangerment of two-thirds of the world's vertebrates categorized as at risk of extinction (Prescott-Allen and Prescott-Allen, 1978) and of 82% of the endangered bird species (Temple, 1986).

Landscapes subjected to change through fragmentation represent one of the real world situations where metapopulation theory and models have been shown to be particularly suited to understand species persistence and population dynamics. The process of habitat fragmentation entails the creation of discrete habitat patches whose effect is that local populations interact with each other through the exchange of dispersing individuals, thus behaving as a metapopulation system (Hanski and Simberloff, 1997). In this essay I will present an overview of metapopulation models and theory from an ecological point of view, emphasizing those concepts and approaches deemed most useful to understand the consequences of changes in the global environment. A more in-depth treatment of metapopulation theory may be found in four recent books on the topic (Gilpin and Hanski, 1991; McCullough, 1996; Hanski and Gilpin, 1997; Hanski, 1999). It should be noted that population geneticists have been analyzing microevolutionary processes within metapopulations or spatially structured populations for a long time (Wright, 1931). The interested reader may refer to Hasting and Harrison (1994) and Harrison and Hastings (1996) for reviews.

METAPOPULATION MODELS AND THEORY

The pioneer of metapopulation models is that proposed by Levins (1969). This simple model assumes that N, the total number of available patches, is a constant. Let U and O denote the number of unoccupied and occupied patches, respectively. Levins' model assumes that instantly upon colonization of an empty patch, the organisms achieve their carrying capacity, thus reaching their demographic equilibrium within each patch. Assume that at this equilibrium, each individual in the patch produces a total of β propagules per unit time. Therefore, βO represents the total number of propagules produced by all the individuals in the occupied patches. These propagules find unoccupied patches at a rate proportional to their frequency U/N, thus unoccupied patches are lost to colonization at a rate $\beta OU/N$ per unit time, and occupied patches increase by the same number per unit time. If we assume that occupied patches become extinct at a rate e, then eO is the number of occupied patches that become extinct per unit time. Furthermore, this model assumes that occupied patches become unoccupied and immediately available for colonization at the same rate at which they become extinct, implying a closed system without an independent patch dynamics. The Equations that govern this system are:

$$\frac{d}{dt}O = \beta O\frac{U}{N} - eO$$
$$\frac{d}{dt}U = -\beta O\frac{U}{N} + eO$$

Dividing both Equations by N, defining $O/N = p$, and noting that $U/N = 1 - p$, then the Equations reduce to the Levins metapopulation model (Equation 1):

$$\frac{dp}{dt} = \beta p(1-p) - ep \qquad (1)$$

Because all patches and colonizing individuals are identical, one can dynamically follow the proportion of occupied patches instead of their actual numbers, and characterize the whole dynamics with two parameters: β and e. At equilibrium, the proportion of occupied patches is given by $p^* = 1 - e/\beta$, which is globally asymptotically stable, such that perturbations of this equilibrium as well as all initial conditions converge to this point.

In Levins' model, a positive proportion of occupied patches at equilibrium will exist whenever $\beta/e > 1$ (Figure 2). Taking into account that the average lifetime of an occupied patch is $1/e$, the above expression represents the total number of secondary colonizations produced by an occupied patch during its lifetime. This threshold parameter corresponds to the basic reproductive number used in epidemiology (usually denoted by the symbol R_0) which is a measure of the number of secondary infections that a single infectious individual produces when introduced

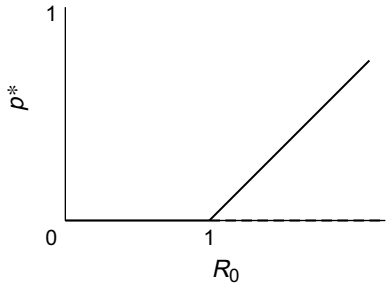


Figure 2 Bifurcation diagram for Levins' metapopulation model. For values of $R_0 < 1$ only the steady-state $p^* = 0$ exists and is stable. For $R_0 > 1$, a second steady-state appears. The other equilibrium still exists but is unstable. Note that the equilibrium proportion of occupied patches increases as R_0 increases. (Reproduced by permission of Sociedad de Biología de Chile in Marquet and Velasco-Hernández, 1997)

in a completely susceptible population (Diekman *et al.*, 1990; Hernandez-Suarez *et al.*, 1999). If this number is higher than one, the disease spreads in the host population. Otherwise, no epidemic outbreak ensues and the disease dies out. The basic reproductive number is therefore an invasion criterion: it determines if a pathogen will be able to survive in a host population once it is introduced. In a metapopulation context, it determines if a landscape composed of a set of empty patches will be successfully colonized, and also determines its long-term persistence (Marquet and Velasco-Hernandez, 1997). To appreciate the importance of R_0 in affecting metapopulation persistence, we can re-scale time in Equation (1) by taking as a unit the average time to extinction $1/e$. With this re-scaling, Levin's original model becomes (Hernandez-Suarez *et al.*, 1999):

$$\frac{dp}{d\tau} = R_0 p(1-p) - p \qquad (2)$$

where τ stands for the new re-scaled time. It is clear that if $R_0 < 1$, $p \to 0$ and if $R_0 > 1$, $p \to 1$.

This simple patch-occupancy metapopulation model provides a simple and fruitful way to understand the basic dynamical properties of metapopulations. Its success is re ected in its many subsequent modifications and applications to describe single species, two species, and multispecies interactions (Hanski, 1999). Although the many assumptions made by Levins' model limits its application to understand real world metapopulations, it has been of paramount importance to unveil the existence of important processes affecting species persistence in patchy environments. In retrospect, Levins' simple model is the ancestor of a plethora of more complicated models, some of which are described in Table 1. Among them, spatially explicit models have become very important in metapopulation theory (e.g., Keymer *et al.*, 1998) and well grounded in ecology (Durrett and Levin, 1994; Tilman and Kareiva, 1997; Dieckmann *et al.*, 2000). After all, metapopulations as well as ecological systems in general, are spatially extended systems whose dynamics are highly dependent on their topological arrangement, and neighborhood interactions. Two typical spatially explicit metapopulation models are presented in Figure 3. In the first model (Figure 3a) space is included as a regular lattice of patches that can be occupied by a species. However, most landscapes have a more complex spatial distribution of patches, which have different shapes and are located at different distances from each other. This complexity can be captured in spatially realistic models where a real landscape (in the form of a Geographical Information Systems (GIS) layer or a classified satellite image), instead of an arbitrary lattice, is used to run the model. Examples of this approach are the RAMAS-GIS model developed by Akcakaya (1995), and the PATCH model developed by Schumaker (1998). Figure 3(b) shows the metapopulation structure of the California spotted owl (*Strix occidentalis occidentalis*) in the Sierra Nevada and

Table 1 Different types of metapopulation models

Patch occupancy models or patch models
Models where the environment consists of an array of patches in two possible states, occupied or empty. They ignore attributes of local populations such as density and in most cases assume that all patches are equal. This type of model dynamically follows the proportion of patches in each state. Levins' model (Levins, 1969, 1970) is a patch-occupancy model
Structured metapopulation models
Structured metapopulation models explicitly include within-patch dynamics (modeling changes in local population sizes) and allow for the existence of differences in patch quality (reviewed by Gyllenberg *et al.*, 1997)
Spatially implicit metapopulation models
Models that ignore the spatial geometry or arrangement of patches, assuming that all are equally accessible and connected. Levins' model is spatially implicit
Spatially explicit metapopulation models
Models that include space explicitly, usually as a regular lattice of patches. Dispersal is restricted such that the dynamics of each patch in the lattice is a function of the state of the patches in its neighborhood. In this category are Cellular Automata models (e.g., Keymer *et al.*, 1998), coupled map lattice models (Hassell *et al.*, 1991), and interactive particle system models (Durrett and Levin, 1994)
Spatially realistic metapopulation models
These models are spatially explicit, where the lattice of patches is a real landscape (a GIS layer or a remote sensing image). This lattice preserves the relative position of patches, areas and other attributes of the real landscape (Keitt *et al.*, 1997; Schumaker, 1998)

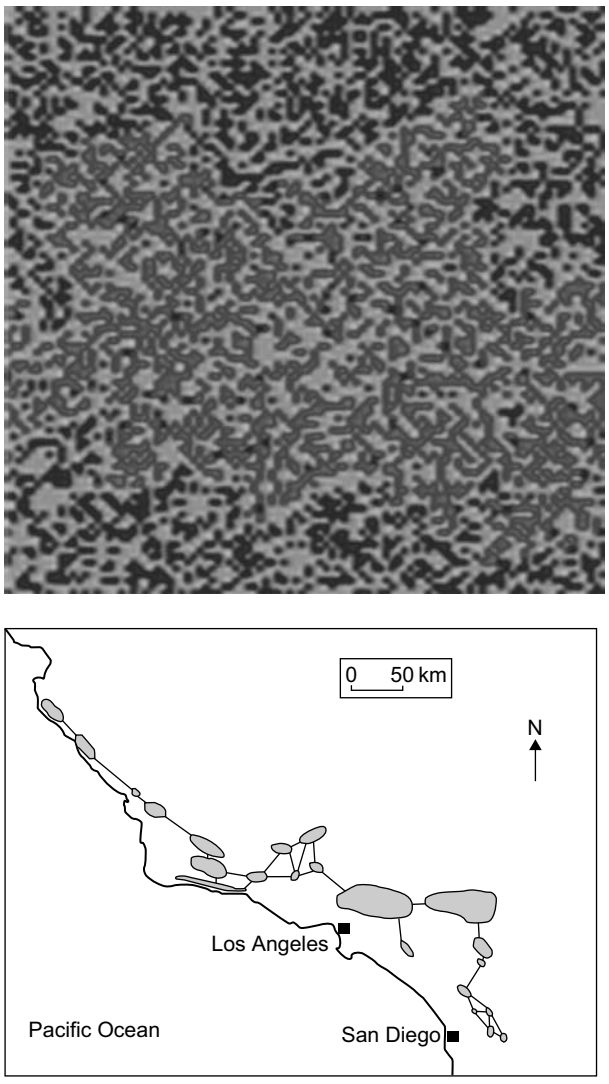


Figure 3 Two types of spatially explicit metapopulation models. (a) Space enters into the model as a regular lattice where each cell can be in a different state. (b) The metapopulation structure of the California spotted owl (*Strix occidentalis occidentalis*) is shown in the Sierra Nevada and in several isolated local populations in the mountains of southern California. (Reproduced with the permission of the British Ecological Society in Lahaye *et al.*, 1994)

in several isolated local populations in the mountains of southern California. A metapopulation model run on this patch network would be spatially realistic to the extent that the topology of the real system is preserved and affects its dynamics.

In Levins' patch occupancy model, the dynamics of the metapopulation is the result of the rate at which patches go extinct and the rate at which empty patches are colonized. However, by relaxing some of the assumptions about the source of the colonists, and the effect of migration between extant local populations, more complex and realistic scenarios can be analyzed.

SOME KEY METAPOPULATION CONCEPTS

Levins' model assumes that all colonists come from local populations within the system. However, there are other possible formulations (e.g., MacArthur and Wilson, 1967) with colonists coming from outside the system. In this scenario there is a propagule rain (a continuous source of migrants that could potentially colonize an empty site) which implies that the colonization rate depends, linearly on the fraction of empty patches only (Gotelli, 1991):

$$\frac{dp}{dt} = \beta(1-p) - ep \tag{3}$$

Examples of situations in which the assumption of propagule rain may be appropriate include:

1. a collection of forest fragments separated from a larger expanse of forest which serves as a source of colonizers;
2. an archipelago of islands near a continental source of propagules;
3. intertidal habitats with organisms that are sedentary as adults but have widely dispersed pelagic larvae.

Among metapopulation models that include the colonization process as a propagule rain are the so-called mainland-island metapopulations.

Individuals or propagules produced by a local population may either land in an empty habitat patch, an event that is called *colonization*, or in an occupied patch where there are already *conspeci c* individuals (i.e., that belong to the same species), in which case this event is called *immigration*. Immigration can be of great importance for the persistence of local populations, especially if the population receiving the immigrants is close to extinction. In this case, the genetic and demographic contribution of immigrant individuals may potentially rescue the population from extinction by increasing its population size, thus lowering the chances of disappearing because of demographic or genetic stochasticity. This concept named *rescue effect*, or the effect of immigration on extinction, was elegantly developed by Brown and Kodric-Brown (1977) in the context of the equilibrium theory of island biogeography (MacArthur and Wilson, 1967). It implicitly emphasizes the importance of connectivity among local populations and how this may potentially enhance metapopulation persistence. Clearly, this is of great importance to applied questions such as the maintenance of habitat corridors across landscapes.

Hanski (1982) was the first to include the rescue effect in a metapopulation model. This author reasoned that if

rescue effects are operating, then the extinction rate should decrease as the proportion of occupied habitat patches increases. This is because the probability of a local population being rescued from extinction increases as the numbers of potential sources of immigrants to any given local population increases. A simple way to include the rescue effect into Levin's model is (Hanski, 1982):

$$\frac{dp}{dt} = \beta p(1-p) - ep(1-p) \tag{4}$$

which reduces to:

$$\frac{dp}{dt} = \lambda p(1-p) \tag{5}$$

where $\lambda = \beta - e$. This model, which has a stable equilibrium at $p^* = 1$, gives rise to a bimodal distribution of occupied patches when λ is considered to be a random variable. In a multi-species case, this model predicts the existence of two types of species: those found in most of the habitat patches and those found in very few. This is what is known as the *core- satellite species hypothesis* (see Hanski, 1999 for a review and alternative models).

Source–Sink Dynamics

Landscapes are heterogeneous in time and space. Spatial heterogeneity is manifested, among other things, in differences in habitat quality such that the demographic rates of local populations are different in different patches. Extreme cases across a continuum of habitat-specific demographic rates are represented by source and sink populations (Pulliam, 1988, 1996). A *source population* is one where births exceed deaths and emigration exceeds immigration (Pulliam, 1988; see also Roughgarden and Iwasa, 1986). In other words, source populations are net exporters of individuals. In *sink populations*, on the other hand, deaths exceed births and immigration exceeds emigration. By definition, a sink population would not persist if immigration were impeded, because it has a negative rate of population increase (deaths outnumber births). Thus, rescue effects are essential for the persistence of sink populations. Clearly, sink populations will be found to the extent that there are source populations subsidizing them and engaged in source–sink dynamics. According to Pulliam's (1988) definition, real metapopulations should be composed of a mosaic of source and sink local populations. In practical terms, source populations are essential to metapopulation persistence and should be of great conservation concern. But why would some individuals immigrate into low quality patches inhabited by sink populations? There are at least two likely explanations (Dias, 1996). Individuals can occupy low-quality habitats as a result of interference competition whereby juvenile or subordinate individuals are forced to leave high quality (source) habitat patches, or because of passive dispersal as in plants and sessile invertebrates in intertidal habitats. A classical example of the first alternative is Carl's (1971) study on Arctic ground squirrels (*Spermophilus undulatus*). This species lives in breeding colonies where burrowing sites are limited and hence actively defended against conspecifics. Those individuals forced to ee the main breeding colonies occupy low quality (sink) habitat where mortality is high due to predation and environmental perturbations. Virtually all individuals in this population are immigrants in a low quality habitat. An example of a sink population maintained by passive dispersal is that described in Kadmon's (1993) study of the demography of the desert annual herb *Stipa capensis* in three habitats (slopes, depressions, and dry water course–*wadis*). Kadmon estimated that 75–99% of the seeds were produced in the depression and wadis, even though these two habitats represent less than 10% of the area occupied by the species. This evidence and previous results, which indicated that seeds produced in the wadis accounted for more than 90% of the abundance of *Stipa capensis* in the slope habitat, suggest that slope populations are maintained by immigration and correspond to sink populations. Typically sink populations are common at the border of geographic ranges of species. As noted by Lawton (1996), this fact may underlie the observed low success rate in species reintroduction programs at the edge of a species range in comparison with reintroductions in central areas (Griffith *et al.*, 1989).

Source–sink dynamics is an important metapopulation process with profound implications for species conservation in the face of global change. In the first place, it makes clear that the presence of a species in a given habitat is not proof that the habitat is suitable for its persistence or able to support a locally breeding population in the absence of immigration. Secondly, it suggests that the density of a species may be a misleading indicator of habitat suitability (Pulliam, 1996) because some species can maintain large populations in unsuitable (sink) habitats (Pulliam, 1988).

The previous discussion emphasized the importance of source populations for species persistence. However, sink populations can also be important on ecological and evolutionary grounds. Under some circumstances, sink populations can stabilize interactions between species, can foster coexistence in ecological systems (e.g., Holt, 1985; Loreau and DeAngelis, 1997), and can increase metapopulation persistence (Howe *et al.*, 1991; but see Marquet and Velasco-Hernandez, 1997) especially in variable environments (Jansen and Yoshimura, 1998). Sink populations in the periphery of the geographic ranges of species tend to be genetically divergent from central populations and potentially important for future speciation, especially if mutations of large effects on individual fitness occur in

these sink populations. If the mutations are large enough, they can overcome the selection bias toward increasing adaptation to source habitats, rather than to sinks (Holt and Gaines, 1992). Thus, under some circumstances peripheral sink populations can be of conservation value (Lesica and Allendorf, 1995), by keeping genetic variability that may be useful for future adaptation, especially in a climate change scenario.

GLOBAL ENVIRONMENTAL CHANGE AND METAPOPULATION EXTINCTION

Most species are spatially structured as metapopulations within their geographic ranges, and their global extinction is usually mediated through changes in metapopulation dynamics, as a consequence of habitat loss associated with human encroachment of natural habitats. How these events affect metapopulation extinction depends on three factors: (1) how many local populations or habitat patches are lost; (2) the quality of the remaining patches; and (3) the resulting changes in metapopulation connectivity.

The effect of habitat loss or destruction depends on how many local populations are required to ensure the long-term persistence of the metapopulation or its *minimum viable metapopulation size*. This is equivalent to one of the key concept in population and conservation biology – the *minimum viable population size*, i.e., the minimum number of individuals within an isolated population necessary to guarantee its long-term persistence (Schaffer, 1981). Until recently, this concept remained almost unexplored in a metapopulation context. Following the lead of Lande (1987), Nee and May (1992) proposed a general model that attempted to understand the effect of habitat destruction upon species interactions and persistence in patchy landscapes. By focusing on patch destruction, this model allows one to estimate the minimum number or fraction of habitable patches required for metapopulation persistence, also known as the *extinction threshold* (Lande, 1987; Lawton *et al.*, 1994; Hanski *et al.*, 1996; Hernandez-Suarez *et al.*, 1999). Conceptually, the extinction of a metapopulation following patch destruction is equivalent to the collapse of a disease epidemic following the removal of susceptible hosts (Lawton *et al.*, 1994). The process depends on the *eradication threshold* (Anderson and May, 1991), that is, on the minimum number of susceptible individuals that will enable the disease to persist.

To understand the effect of habitat destruction on metapopulation persistence we can use a simple model. Imagine a metapopulation composed of a finite number (N) of patches and suppose that D patches are destroyed. Now consider the question of how many patches can be destroyed without driving the situation to extinction? In this situation the number of suitable patches is $N - D$,

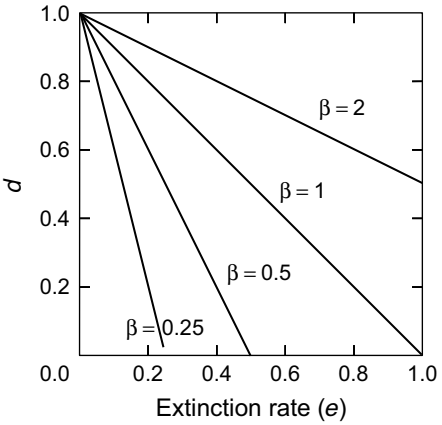


Figure 4 Changes in the critical amount of habitat destruction (d) needed to drive a metapopulation to extinction as a function of the species' extinction rate (e) and for species with different colonization rates (β). Notice that for the same extinction rate, species with higher colonization potential can withstand a higher proportion of habitat destruction before going extinct, and that this critical threshold decreases as extinction rates increases

which can be either occupied (O), or empty (U), such that $N - D = O + U$. Dividing by N, defining the proportion of occupied patches as $O/N = p$, the proportion of destroyed patches as $D/N = d$, the proportion of empty patches as $U/N = 1 - d - p$, and then replacing these terms in Equation (1), we arrive at a model equivalent to that proposed by Levins (1969) but now including habitat destruction:

$$\frac{dp}{dt} = \beta p(1 - d - p) - ep \tag{6}$$

In this model the proportion of occupied sites at equilibrium in the face of destruction (p_d^*) is:

$$p_d^* = 1 - \frac{e}{\beta} - d \tag{7}$$

Equation (7) has several implications. First, metapopulation extinction will occur ($p_d^* = 0$) if a critical proportion of habitat (d) is destroyed. This threshold corresponds to

$$d = 1 - \frac{e}{\beta} \tag{8}$$

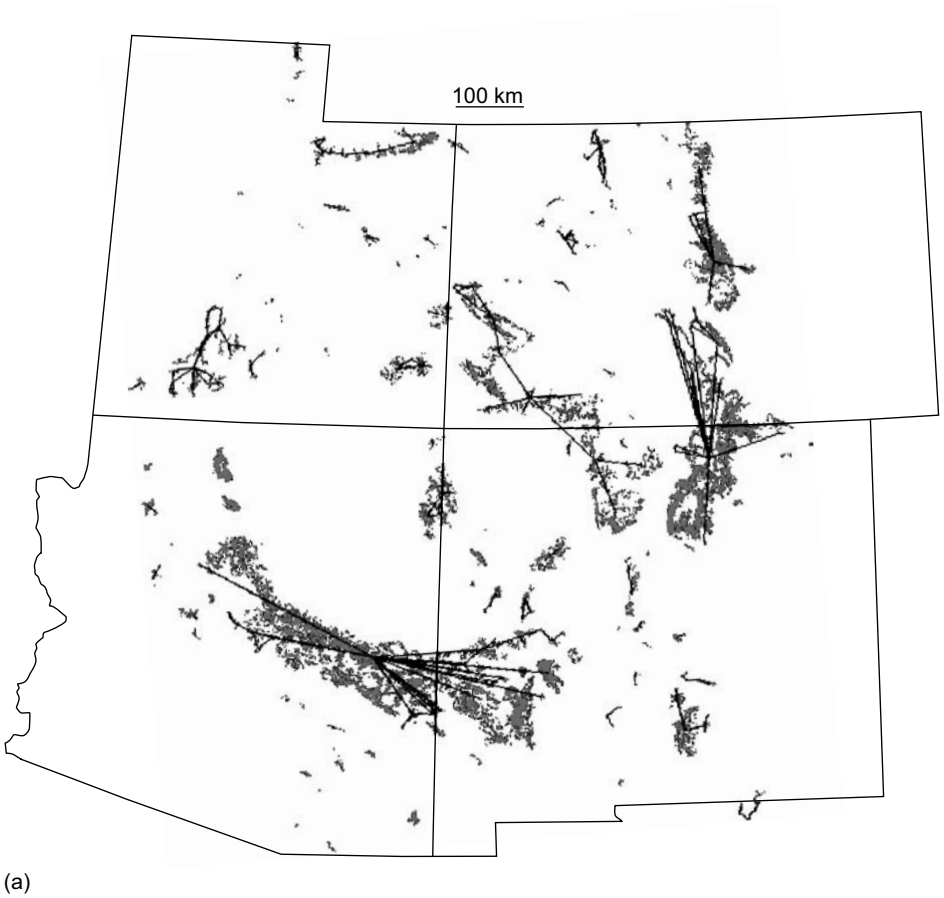
This implies that species with higher colonization rates or lower extinction rates can sustain higher amounts of habitat destruction (Figure 4). Interestingly, as noted by Tilman *et al.* (1997), this critical threshold corresponds to the equilibrium proportion of occupied patches in the absence of destruction such that a rule of thumb for metapopulation persistence is: A necessary and sufficient condition for metapopulation persistence is that the proportion of destroyed habitat patches should be less than the proportion originally occupied by the species in the

intact system. Thus, if a species in a pristine habitat occupied 25% of the available patches, then the random destruction of >25% of the patches will result in its extinction. Notice also that Equation (7) can be rearranged as

$$1 - d - p_d^* = \frac{e}{\beta} \tag{9}$$

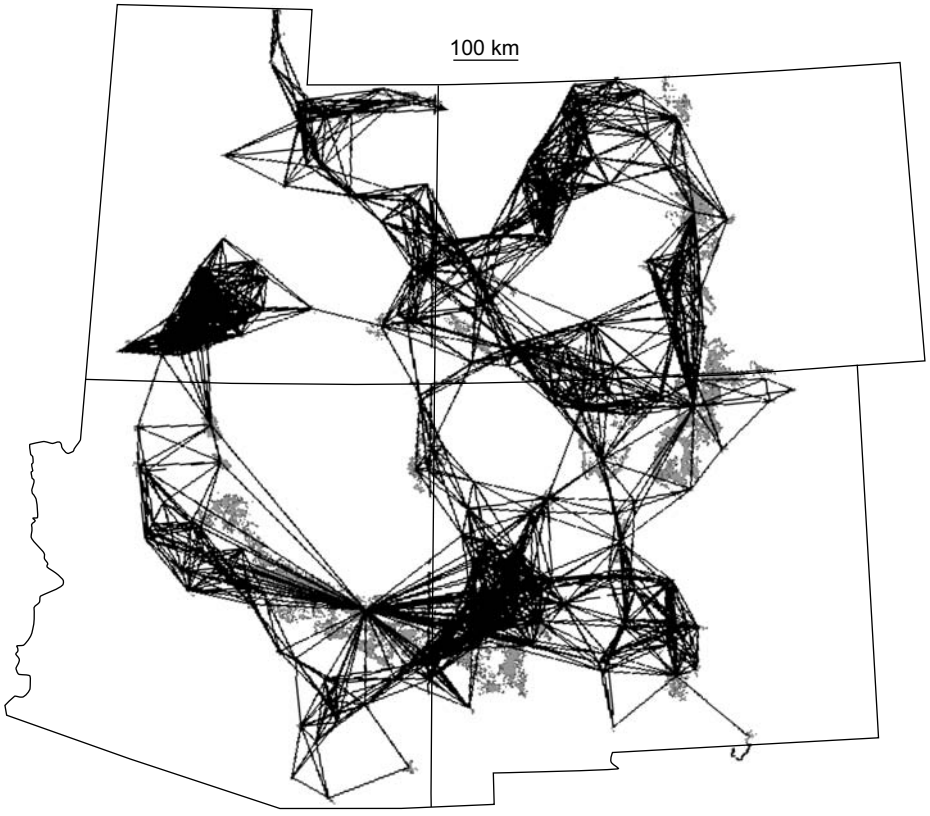
Noticing that $1 - d$ corresponds to the total amount of suitable habitat left after destruction, an equivalent rule of thumb (see Lawton *et al.*, 1994; Hanski *et al.*, 1996) can be formulated as: A metapopulation will persist if the proportion of suitable habitat patches after destruction exceeds the proportion of empty but suitable habitat patches in the intact system (i.e., $1 - d > e/\beta$). However, there are many other factors not captured by this simple model that can have an effect on metapopulation extinction.

A metapopulation may go extinct well before habitat destruction seriously threatens its persistence, as derived in the models above, because of stochasticity linked to a reduced number of local populations, such that a metapopulations can go extinct simply because all local populations happen to become extinct at the same time. This effect has been termed *extinction-colonization stochasticity* (Hanski, 1991) and is equivalent to the concept of demographic stochasticity for closed populations (May, 1973). But even if there is no habitat destruction and the metapopulation is large enough to escape extinction-colonization stochasticity, large-scale changes in the environment may tend to increase the correlation in local population dynamics, thus increasing the probability that all local populations become sinks, which can lead to metapopulation extinction. This effect has been termed *regional stochasticity* (Hanski, 1991) and is equivalent to the notion of environmental stochasticity in closed populations (May, 1973). To the extent that local populations are not correlated in their local



(a)

Figure 5 Digital cover map of the southwestern United States showing the spatial distribution of mixed-conifer and ponderosa pine habitat patches and the potential connection among them as a function of the scale at which organisms interact with the landscape pattern. In (a) black lines represent paths along which habitat patches are at a distance less than or equal to a threshold of 10 km, and thus organisms with a dispersal capability <10 km will perceive a highly fragmented landscape with many isolated patches. On the contrary, an organism capable of long distance dispersal [100 km as shown in (b)] will perceive a highly connected landscape. (Reproduced with permission of the Ecological Society of America in Keitt *et al.*, 1997)



(b)

Figure 5 (*Continued*)

dynamics, there will always be some local populations acting as sources (and rescue effects operating) to counteract local extinctions.

Finally, the quality of patches lost to destruction can make a big difference in terms of metapopulation persistence. Traditionally, the quality of patches has been assessed in relation to their suitability for population growth (with source and sink being extreme cases). In this context the destruction of a source patch has a higher impact on persistence than the destruction of a sink. However, patches of low quality for population growth can be of paramount importance in maintaining the connectivity of the system by acting as stepping stones between high quality patches; consequently their loss can have a tremendous impact on persistence. Thus, in terms of connectivity, source patches can be dependent on sink patches. Recently, Keitt *et al.* (1997) proposed a method to quantify habitat connectivity and to assign conservation priority to habitat patches based on their contribution to connectivity. Because connectivity of landscapes depends not only on the spatial configuration of habitat patches but also on the scale at which organisms interact with landscape pattern, e.g., Gardner *et al.* (1989), Keitt *et al.* (1997) quantified connectivity at different scales (Figure 5). Then, for each map, the authors removed patches (one at a time) and recorded the changes in landscape connectivity. This patch removal experiment allowed them to assess the importance of each patch in terms of landscape connectivity. Interestingly, after controlling by area, small patches had a very high contribution to connectivity, which suggest that they may be of high priority for conservation.

This is an important result, provided that small patches are more sensitive to environmental changes and may be the first to disappear. If this is the case, current environmental change may have a strong negative impact on metapopulation persistence even if source patches are not affected. The disappearance of sink populations may trigger a cascade of events such as disruption in connectivity, increase in isolation, limited rescue effects, and loss of useful genetic variability, thus increasing metapopulation extinction risk. Because these effects may be subtle, lagged in time (Tilman *et al.*, 1994), and usually with sharp threshold transitions (Bascompte and Sole, 1996; Keitt *et al.*, 1997), it is essential to establish long-term multiscale landscape monitoring programs to ensure metapopulation persistence in the face of a global change in the environment.

PATCH DYNAMICS

So far we have discussed the role of landscape pattern and the importance of particular patches for metapopulation persistence. However, Keymer *et al.* (2000) have shown that in addition to the pattern of *patch distribution*, *patch dynamics* (i.e., the creation of extinction and patches) greatly affects metapopulation persistence. Keymer *et al.* (2000) propose that metapopulation persistence is the result of the interaction between landscape pattern, patch dynamics, and species life history. The inclusion of patch dynamics in metapopulation theory underscores the importance of landscape monitoring, and the need for an adequate characterization of patch dynamic and disturbances regimes in order to gain insight and be able to predict species extinction. In particular, the establishment of monitoring program inside and outside protected areas would be desirable to understand the interaction between direct human impacts and environmental changes upon landscape pattern and dynamics, and how these affects metapopulation extinction.

The effects of global environmental changes upon ecological systems are complex and involve changes in landscape patterns and dynamics. In this context, metapopulation theory provides the theoretical and empirical framework needed to understand and foresee the potential consequences of these changes in the global environment.

ACKNOWLEDGMENTS

The author acknowledges support from grant FONDECYT 1990144. Special thanks to J Keymer, J Mena and J X Velasco-Hernandez, who provided stimulating discussions on metapopulations in recent years. M Lima and F Jaksic provided helpful comments on the manuscript. Figure 3(a) was kindly provided by J Keymer.

REFERENCES

Akcakaya, H R (1995) *RAMAS/GIS: Linking Landscape Data with Population Viability Analysis*, Applied Mathematics, Setauket, NY.

Anderson, R M and May, R M (1991) *Infectious Diseases of Humans*, Oxford Science Publications, Oxford.

Bascompte, J and Sole, R V (1996) Habitat Fragmentation and Extinction Thresholds in Spatially Explicit Models, *J. Anim. Ecol.*, 65, 465–473.

Brown, J H and Kodric-Brown, A (1977) Turnover Rates in Insular Biogeography: Effect of Immigration on Extinction, *Ecology*, 58, 445–449.

Carl, E A (1971) Population Control in Arctic Ground Squirrels, *Ecology*, 52, 395–413.

Dias, P C (1996) Sources and Sinks in Population Biology, *Trends Ecol. Evol.*, 11, 326–330.

Diekman, O, Hesterbeek, J A P, and Metz, J A J (1990) On the Definition and Computation of the basic Reproductive Ratio R_0 in Models for Infectious Diseases in Heterogeneous Populations, *J. Math. Biol.*, 28, 365–382.

Dieckmann, U, Law, R, and Metz, J A J (2000) *The Geometry of Ecological Interactions. Simplifying Spatial Complexity*, Cambridge University Press, Cambridge.

Durrett, R and Levin, S (1994) The Importance of being Discrete (and Spatial), *Theor. Popul. Biol.*, 46, 363–394.

Ehrlich, P R and Wilson, E O (1991) Biodiversity Studies: Science and Policy, *Science*, 253, 758–762.

Gardner, R H, O'Neill, R V, Turner, M G, and Dale, V H (1989) Quantifying Scale-dependent Effects of Animals Movements with Simple Percolation Models, *Landscape Ecology*, 3, 217–227.

Gause, G F (1935) *The Struggle for Existence*, William and Wilkins, Baltimore, MD.

Gilpin, M and Hanski, I (1991) *Metapopulation Dynamics: Empirical and Theoretical Investigations*, Academic Press, London.

Gotelli, N J (1991) Metapopulation Models: the Rescue Effect, Propagule Rain, and the Core-satellite Hypothesis, *Am. Nat.*, 138, 768–776.

Griffith, B, Scott, J M, Carpenter, J W, and Reed, C (1989) Translocation as a Species Conservation Tool: Status and Strategy, *Science*, 245, 477–480.

Groom, M J and Schumaker, N (1993) Evaluating Landscape Change: Patterns of Worldwide Deforestation and Local Fragmentation, in *Biotic Interactions and Global Change*, eds P M Kareiva, J G Kingsolver, and R B Huey, Sinauer, Sunderland, MA, 24–44.

Gyllenberg, M, Hanski, I, and Hastings, A (1997) Structured metapopulation models, in *Metapopulation Biology*, eds I Hanski and M E Gilpin, Academic Press, San Diego, CA, 93–122.

Hannah, L, Carr, J L, and Lankerani, A (1995) Human Disturbance and Natural Habitat: A Biome Level Analysis of a Global Data Set, *Biodivers. Conserv.*, 4, 128–155.

Hanski, I (1982) Dynamics of Regional Distribution: the Core and Satellite Hypothesis, *Oikos*, 38, 210–221.

Hanski, I (1991) Single-species Metapopulation Dynamics: Concepts, Models and Observation, *Biol. J. Linn. Soc.*, 42, 17–38.

Hanski, I (1999) *Metapopulation Ecology*, Oxford University Press, New York.

Hanski, I and Gilpin, M E (1997) *Metapopulation Biology*, Academic Press, San Diego, CA.

Hanski, I, Moilanen, A, and Gyllenberg, M (1996) Minimum Viable Metapopulation Size, *Am. Nat.*, 147, 527–541.

Hanski, I and Simberloff, D (1997) The Metapopulation Approach, its History, Conceptual Domain, and Application to Conservation Biology, in *Metapopulation Biology*, eds I Hanski and M E Gilpin, Academic Press, San Diego, CA, 5–26.

Harrison, S and Hastings, A (1996) Genetic and Evolutionary Consequences of Metapopulation Structure, *Trends Ecol. Evol.*, 11, 180–183.

Hassel, M P, Comins, H N, and May, R M (1991) Spatial Structure and Chaos in Insect Population Dynamics, *Nature*, 353, 255–258.

Hastings, A and Harrison, S (1994) Metapopulation Dynamics and Genetics, *Annu. Rev. Ecol. Syst.*, 25, 197–188.

Hernandez-Suarez, C M, Marquet, P A, and Velasco-Hernandez, J X (1999) On Threshold Parameters and Metapopulation Persistence, *Bull. Math. Biol.*, 60, 1–14.

Holt, R D (1985) Population Dynamics in Two Patch Environments: Some Anomalous Consequences of an Optimal Habitat Distribution, *Theor. Popul. Biol.*, 28, 181–208.

Holt, R D and Gaines, M S (1992) Analysis of Adaptation in Heterogeneous Landscapes: Implications for the Evolution of Fundamental Niches, *Evol. Ecol.*, 6, 433–447.

Howe, R W, Davis, G J, and Mosca, V (1991) The Demographic Significance of Sink Populations, *Biol. Conserv.*, 57, 239–255.

Jansen, V A A and Yoshimura, J (1998) Populations can Persist in an Environment Consisting of Sink Habitats Only, *Proc. Natl. Acad. Sci.*, 95, 3696–3698.

Kadmon, R (1993) Population Dynamic Consequences of Habitat Heterogeneity: An Experimental Study, *Ecology*, 74, 816–825.

Keitt, T H, Urban, D L, and Milne, B T (1997) Detecting Critical Scales in Fragmented Landscapes, *Conservation Ecology* [online], 1, 4, available from the internet, URL: http://www.consecol.org/Journal/vol1/iss1/art4.

Keymer, J E, Marquet, P A, and Johnson, A R (1998) Pattern Formation in a Patch Occupancy Metapopulation Model: A Cellular Automata Approach, *J. Theor. Biol.*, 194, 79–90.

Keymer, J E, Marquet, P A, Velasco-Hernandez, J X, and Levin, S A (2000) Extinction Thresholds and Metapopulation Persistence in Dynamics Landscapes, *Am. Nat.*, 156, 478–494.

Lahaye, W S, Gutierrez, R J, and Akcakaya, H R (1994) Spotted Owl Metapopulation Dynamics in Southern California, *J. Anim. Ecol.*, 63, 775–785.

Lande, R (1987) Extinction Thresholds in Demographic Models of Territorial Populations, *Am. Nat.*, 130, 624–635.

Lawton, J H (1996) Population Abundances, Geographic Ranges and Conservation: 1994 Witherby Lecture, *Bird Study*, 43, 3–19.

Lawton, J H, Nee, S, Letcher, A J, and Harvey, P H (1994) Animal distributions: patterns and processes, in *Large-scale Ecology and Conservation Biology*, eds P J Edwards, R M May, and N R Webb, Blackwell Science, London, 41–58.

Lesica, P and Allendorf, F W (1995) When are Area Peripheral Populations Valuable for Conservation? *Conserv. Biol.*, 9, 753–760.

Levins, R (1969) Some Demographic and Genetic Consequences of Environmental Heterogeneity for Biological Control, *Bull. Entomol. Soc. Am.*, 15, 237–240.

Levins, R (1970) Extinction, in *Some Mathematical Problems in Biology*, ed M Gesternhaber, American Mathematical Society, Rhode Island, 77–107.

Loreau, M and DeAngelis, D L (1997) Source–sink Dynamics and the Coexistence of Species on a Single Resource, *Theor. Popul. Biol.*, 51, 79–93.

MacArthur, R H and Wilson, E O (1967) *The Theory of Island Biogeography*, Princeton University Press, Princeton, NJ.

Marquet, P A and Velasco-Hernandez, J X (1997) A Source–sink Patch Occupancy Metapopulation Model, *Revista Chilena de Historia Natural*, 70, 371–380.

May, R M (1973) *Stability and Complexity in Model Ecosystems*, Princeton University Press, Princeton, NJ.

McCullough, D R (1996) *Metapopulations and Wildlife Conservation*, Island Press, Washington, DC.

Nee, S and May, R M (1992) Dynamics of Metapopulations: Habitat Destruction and Competitive Coexistence, *J. Anim. Ecol.*, 61, 37–40.

Prescott-Allen, R and Prescott-Allen, C (1978) *Sourcebook for a World Conservation Strategy: Threatened Vertebrates*, International Union for the Conservation of Nature and Natural Resources, Gland.

Pulliam, H R (1988) Sources, Sinks and Population Regulation, *Am. Nat.*, 132, 267–274.

Pulliam, H R (1996) Sources and Sinks: Empirical Evidence and Population Consequences, in *Population Dynamics in Ecological Space and Time*, eds O E Rhodes, Jr, R K Chesser, and M H Smith, The University of Chicago Press, Chicago, IL, 45–69.

Roughgarden, J and Iwasa, Y (1986) Dynamics of a Metapopulation with Space-limited Sub-populations, *Theor. Popul. Biol.*, 29, 235–261.

Shaffer, M L (1981) Minimum Population Sizes for Species Conservation, *BioScience*, 31, 131–134.

Schumaker, N (1998) A Users Guide to the PATCH Model, EPA/600/R-98/135, US Environmental Protection Agency, Environmental Research Laboratory, Corvallis, OR.

Temple, S A (1986) The Problem of Avian Extinctions, *Current Ornithology*, 3, 453–485.

Tilman, D and Kareiva, P (1997) *Spatial Ecology. The Role of Space in Population Dynamics and Interspeci c Interactions*, Princeton University Press, Princeton, NJ.

Tilman, D, Lehman, C L, and Yin, C (1997) Habitat Destruction, Dispersal, and Deterministic Extinction in Competitive Communities, *Am. Nat.*, 149, 407–435.

Tilman, D, May, R M, Lehman, C L, and Nowak, M A (1994) Habitat Destruction and the Extinction Debt, *Nature*, 371, 65–66.

Vitousek, P M (1994) Beyond Global Warming: Ecology and Global Change, *Ecology*, 75, 1861–1876.

Wright, S (1931) Evolution in Mendelian Populations, *Genetics*, 16, 97–169.

Methane

see Volatile Organic Compounds, Biogenic (VOCs) (**Volume 2**)

Methanol

see Volatile Organic Compounds, Biogenic (VOCs) (**Volume 2**)

Microbial Diversity

James T Staley[1] and Jed Fuhrman[2]
[1]University of Washington, Seattle, WA, USA
[2]University of Southern California, Los Angeles, CA, USA

Microbial life has ourished on Earth for at least 3.8 Ga (Ga = 10^9 years) of its 4.5 Ga existence. In contrast, plant and animal life has been here for less than 600 million years. The early evolution of microorganisms (bacteria, archaea and eucaryotic microorganisms such as protists and fungi) and their long tenure on the planet have resulted in their occupation of every conceivable environment that is habitable. Thus, in addition to living in all ecosystems along with plants and animals, some live in boiling hot springs, sub-freezing sea ice, acidic mine waters, saturated salt waters and radioactive wastes that cannot be inhabited by plants and animals. During their long evolutionary history, they have evolved many biochemical capabilities that allow them to use a large variety of energy sources that are oxidized by unique mechanisms of respiration. For example, methane-producing (methanogenic) archaea cannot grow or respire in the presence of oxygen. Yet they are able to oxidize hydrogen gas as an energy source and respire with carbon dioxide that is xed into organic cell materials and methane gas.

Microorganisms formed the rst biosphere and continue to carry out a multitude of essential geochemical activities that prepared Earth for the eventual evolution of macroscopic life. Furthermore, microorganisms live in associations with all plants and animals and are responsible for their well-being. Thus, almost all animals have an intestinal tract lled with resident bacteria that help digest their food.

What is known about microbial life only scratches the surface of what still remains largely unexplored. This contribution discusses the diversity of microbial life, provides some examples of the important roles they play in ecosystem functions and introduces some of their major habitats.

MICROBIAL LIFE AND BIODIVERSITY

Until the last 10 to 15 years, biologists had difficulty in comprehending how plants and animals were related to microorganisms. Most biologists knew that microorganisms were simpler in their morphology, suggesting they had evolved early; however, there was no scientific information that would allow them to be satisfactorily classified with plants and animals which were more complex and whose evolution was partly known from the fossil record. Indeed, until the 1960s, fossils of microorganisms had not yet been discovered (Barghoorn and Tyler, 1965).

All of this has changed. Now, fossils of microorganisms are well known and some have been dated to about 3.5 Ga (Ga = 10^9 years) ago (Schopf, 1993). Likewise, geochemical evidence from early sedimentary deposits indicates that microbial life has existed on Earth for the past 3.8 Ga (Schidlowski, 1988).

But the fossil and chemical evidence provides little information on which microorganisms came first and how microorganisms are related to one another and to plants and animals. A major breakthrough in biology occurred when Carl Woese compared the sequence of a highly conserved ribosomal Ribonucleic acid (RNA) macromolecule found in all cellular organisms. From this he unveiled for the first time, a truly scientific *tree of life* containing the major groups of all living cellular organisms (Figure 1). It should be noted that viruses, which are not independent cellular organisms, do not fit on this tree because they do not have their own cellular machinery and rely on cellular hosts to reproduce.

The tree of life has three major branches, called *domains*, of organisms. One domain is called the *Bacteria*, another the *Archaea* and the final one the *Eucarya*. The most striking difference that distinguishes the eucarya from the bacteria and archaea is that their cells possess a membrane around their nucleus. Also, in contrast to the eucarya, the bacteria and archaea are *haploid*, that is, they have only one set of genes on an unpaired chromosome. The Domain Bacteria, contains the organisms we normally think of when we think of bacteria. With the exception of the planctomycetales group, the bacteria all contain a cell wall material called peptidoglycan consisting of amino sugars and short peptide chains, that makes them potentially sensitive to the antibiotic penicillin. None of the archaea contains peptidoglycan. Bacteria live in soils and aquatic environments and a few of them cause plant and animal diseases. It is interesting to note that photosynthesis appears to have evolved in the Domain Bacteria because several of the phyla have photosynthetic members including the proteobacteria, cyanobacteria, green sulfur bacteria, green filamentous bacteria and one of the gram positive phyla.

The members of the Domain Archaea are unusual in that some of them live in hot springs, others in saturated salt environments and others are responsible for production of methane gas. None of them is known to cause a plant or animal disease. Although we think of the archaea as being unusual in the habitats they occupy, many are also found in typical soil and aquatic habitats along with representatives of the bacteria and microbial eucarya.

The Domain Eucarya is also largely comprised of microorganisms. The cell of the eucarya is more complex than that of the bacteria and archaea. In fact, the eucaryotic cell

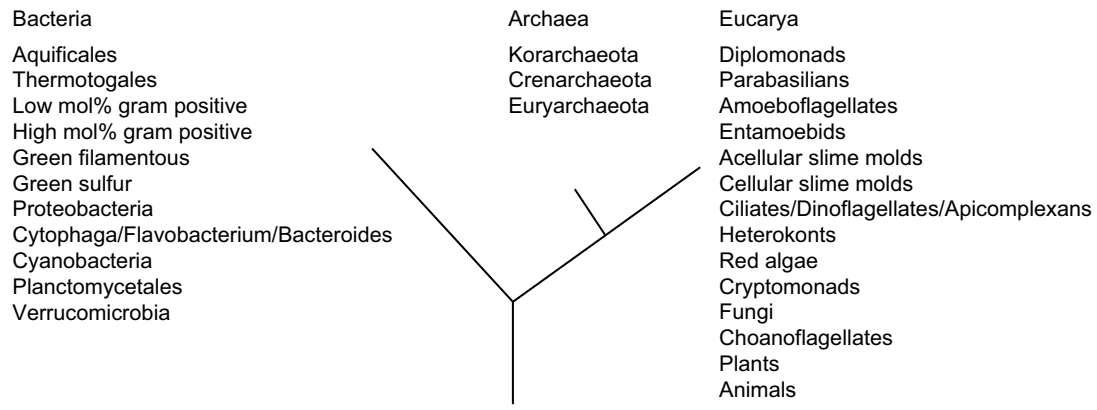


Figure 1 A diagram showing a rooted Tree of Life with its three major branches (domains), the Bacteria, Archaea and Eucarya based upon sequence analyses of the small sub-unit RNA from ribosomes. Under each domain is a listing of the major phylogenetic groups. Branching patterns are provided in some published trees (e.g., Woese *et al.*, 1990). Also, it should be noted that in addition to those sub-domain groups listed, the Bacteria are now known to contain a total of at least twice the number of major groups shown (more than 30 total), many from which no representatives have yet been cultivated (Hugenholz *et al.*, 1998)

actually contains organelles such as the mitochondrion that have evolved from intracellular symbiotic bacteria. Thus, the mitochondrion, which is responsible for aerobic respiration, has evolved from one of the members of the proteobacteria. Likewise, the chloroplast found in all photosynthetic members of the eucarya has evolved from the cyanobacteria. Most believe that the eucarya evolved from the bacteria and archaea, however; the details of how this has happened are still not clear. The eucarya group contains several microbial groups of single-celled organisms that are called *protists*, in addition to the fungi. And, finally in addition, it contains plants and animals.

What is abundantly clear from the tree of life is that most of the major branches and therefore most of the diversity of life on Earth is microbiological. To amplify on this point, it should be noted that many more major phylogenetic groups of microorganisms, particularly in the Domain Bacteria, are still being discovered. The tree shown in Figure 1 can now be updated to include a total of more than 30 major phyla of bacteria which are equivalent to the Animal or Plant Kingdoms by the earlier classifications (Hugenholz *et al.*, 1998). This view of diversity dramatically illustrates the genetic and evolutionary depth and variety of microbial life forms. Although there are only about 5000 different species of bacteria and archaea with formal Latin names, and these are biased towards those of medical or industrial significance, it is believed that the true number of different species probably falls in the millions. The large discrepancy derives from the general lack of morphological distinguishing features and the difficulty and cost in formally describing and naming a microorganism on the basis of biochemical and genetic characteristics.

Viruses

Viruses do not fit onto the tree of life because they are not independent organisms. Viruses consist of a nucleic acid genome (deoxyribonucleic acid (DNA) or RNA, single or double stranded) within a protein coat. They are obligate parasites that live by invading cells and using the cell's own machinery to reproduce, typically killing the host cell when viral progeny are released. Although it is not known for sure, it is thought that virtually all kinds of organisms (bacteria, archaea, and eucarya) are susceptible to viral infection. In general, a particular kind of virus usually can infect only one species or a closely related species (with a few well-known exceptions). Also, for the systems that have been examined closely (humans, farm animals, bacteria like *Escherichia coli*), one species of host may have dozens or hundreds of different viruses. This suggests that viruses are probably extremely diverse, more so than all other organisms combined.

FUNCTIONAL DIVERSITY OF MICROORGANISMS

What the tree of life does not show are the crucial roles that microorganisms play in the biosphere. They play critical roles in all ecosystems and all biogeochemical cycles.

Photosynthesis and Chemosynthesis in Carbon Dioxide Fixation

The oxygen we breathe is produced by photosynthetic microorganisms and plants. Interestingly, oxygen did not

exist on Earth when it was first formed because there was no photosynthesis. Indeed, until about 2.0–2.5 Ga ago, there was no oxygen in the atmosphere at all. As mentioned previously, many of the bacteria contain photosynthetic members. However, the first photosynthetic bacteria, which appear to have been members of the proteobacteria, did not produce oxygen in photosynthesis (Xiong *et al.*, 2000). Instead, they used reduced sulfur compounds such as hydrogen sulfide (H_2S) as the reductant in photosynthesis. Thus, carbon dioxide was reduced to organic carbon and H_2S was oxidized to sulfur and sulfate. However, one group of photosynthetic bacteria, the cyanobacteria, evolved to use water in place of H_2S and when this happened, oxygen was generated. The ubiquitous presence of water (especially as compared to H_2S) made this an incredibly valuable physiological trait, and the cyanobacteria became enormously successful. As that process evolved, oxygen began to oxidize reduced chemicals and eventually began to accumulate in the atmosphere. Because oxygen is essential for macroscopic life forms, its production provided conditions for the evolution of the Plant and Animal kingdoms. Moreover, the oxygen in the atmosphere enabled the development of a stratospheric ozone layer through photochemical reactions with oxygen. The ozone layer absorbs ultraviolet light, which, as a mutagenic agent, prevented the evolution of life on land. This example illustrates how important microbial life forms have been in setting the stage for the evolution and maintenance of plant and animal life. It also illustrates the power of microorganisms in affecting global change.

It is also noteworthy that some bacteria can gain energy for growth from the oxidation of reduced chemicals such as hydrogen gas (one example is the methanogens mentioned earlier), sulfur, ammonia and iron. Some of these organisms are able to carry out the fixation of carbon dioxide, just like plants, but in the absence of light. This process is referred to as *chemosynthesis*. Only certain members of the bacteria and archaea are able to do this, none of the eucarya. This, then is another example of how microbial life is capable of carrying out important, unique and remarkable biogeochemical processes that affect the carbon cycle, the sulfur cycle, the nitrogen cycle and the iron cycle.

Another Biogeochemical Cycle: Nitrogen

The nitrogen cycle also illustrates many unique roles of bacteria. The original atmosphere on Earth is thought to have contained reduced forms of nitrogen including nitrogen gas, N_2, which is quite inert chemically, and ammonia. The current nitrogen cycle consists of other inorganic forms including nitrite, nitrate as well as organic nitrogen produced by organisms. The most remarkable process in the nitrogen cycle is *nitrogen fixation*, the reduction of N_2 in the atmosphere by certain members of the archaea and the bacteria to form the amino group of nitrogen in their cells. Many of the bacteria which carry out this process live as free individuals or colonies in soil and water, but some grow symbiotically with plants such as peas, beans and other legumes as well as alder trees. Thus, the enzymes for nitrogen fixation, including the unique enzyme nitrogenase, are produced only by members of the bacteria and archaea (*see* Nitrogen Fixers, Volume 2).

Two other activities in the nitrogen cycle are carried out exclusively by bacteria. One is *nitrification*, the two-step oxidation of ammonia to nitrite (NO_2^{-1}) and nitrite to nitrate (NO_3^{-1}) by chemosynthetic bacteria. Ammonia, nitrite, and nitrate are fixed forms of nitrogen that can be used as nutrients for plants. The other bacterial process is *denitrification*, which as the name suggests, is simply the reverse of nitrification. Thus, nitrate or nitrite is used as an electron acceptor and reduced to produce, in some bacteria, ammonia and in others nitrogen gases including N_2O and N_2. This latter process is important agriculturally because it removes an essential fixed and therefore utilizable form of nitrogen for plants from soils and aquatic environments. If it were not for microbial nitrogen fixation, the loss of fixed nitrogen by denitrification would cause the biomass of life on Earth (dependent on nitrogen for protein) to dwindle without replacement. Therefore, it is widely thought that on pre-industrial Earth, nitrogen fixation generally balanced the loss of nitrogen by N_2 producing denitrification, on geological time scales. Man has altered this balance by artificially fixing large quantities of nitrogen for the production of fertilizer, among other potential perturbations. It should also be noted that N_2O is one of the potent greenhouse gases of major concern, and microbial processes are responsible for much of its production as well as removal.

These examples provide some sense of the types of processes that are carried out exclusively by microbial life forms, in particular by members of the bacteria and archaea domains. Many other examples exist and many processes are yet only poorly understood. *See also* Nitrogen Cycle, Volume 2.

SYMBIOTIC MICROORGANISMS AND VIRUSES

Many of us may be unaware of the importance of microorganisms to our personal wellbeing. Indeed, we carry billions of microorganisms, particularly bacteria, around with us every day. What is true for us is true for virtually all other animals: microorganisms live in our intestinal tract where they help us degrade the foods that we eat. In so doing, they help provide us with the amino acids and vitamins that are essential in our nutrition. Woe to the

person that is on extended antibiotic therapy, who has lost this essential group of microorganisms. It can be life threatening unless action is taken to replace the microbiota quickly.

Our skin and all of the orifices to our body are other places where microorganisms live. For the most part, they are benevolent organisms that live as *commensals*, neither harming nor necessarily helping us in our own lives. In most instances, they serve us well by preventing the colonization of other microorganisms that could be harmful.

What is true for animals is also true for plants. Microorganisms that grow profusely in the rhizosphere in association with plant roots are responsible for helping plants obtain nutrients for growth. Some, such as the nitrogen-fixing bacteria are confined to specialized tissues in the plants. Others have a less obligate type of symbiosis, but nonetheless still serve vital roles.

It has only been in recent years that it was recognized that viruses might play important roles in natural ecosystems. They are extremely abundant in nature, typically 10 times the numbers of bacteria, and they have relatively rapid turnover rates. In marine systems where they are best studied in this regard (see review by Fuhrman, 1999), it is thought that most viruses infect bacteria and are responsible for about 10–40% of the total bacterial mortality. However, they also infect algae and are thought to sometimes cause the crash of algal blooms. Because of their specificity, viruses may also be important agents in fostering biodiversity. This is because viral infection rates depend on the density of the host organisms. Therefore, when an organism becomes abundant due to superior competition, it also becomes susceptible to a viral epidemic.

MICROBES AND GLOBAL ENVIRONMENTAL CHANGE

Despite the diminutive size of individual microbial cells, the high abundance of microorganisms in virtually all habitats ensures that they contribute significantly to global biomass. It has been estimated recently that the carbon biomass of prokaryotic microbes is 60–100% of that thought to occur in plants worldwide, and microbial nitrogen and phosphorus biomass exceeds that of all other organisms (Whitman *et al.*, 1998). Because microbes are known to have potentially high metabolic and growth rates (in line with the general observation that physiological rates per unit mass tend to increase with decreasing size), it follows that microbial participation in global carbon, nitrogen, and phosphorus cycles is quite large. This includes a major component of global photosynthesis and respiration rates.

A big unanswered question is how microorganisms might respond to global environmental change. Because of their major roles in biogeochemical cycles, changes in the functioning of microbes can have global consequences. Will they buffer some of the changes (by negative feedback) or accentuate them? These are difficult questions to answer, especially given the great diversity of microorganisms and their geochemical functions. A simplistic view might examine a process such as microbial respiration (oxidation of organic carbon), which typically increases with increasing temperature. In a global warming scenario, one might imagine a transient burst of increased microbial respiration, especially in polar climates where permafrost can melt and suddenly make organic matter available for microbial consumption. This would cause a transient increase in CO_2, lasting until frozen substrates get depleted. However, it is only to the extent that respiration uses old organic matter that it would cause a net increase in CO_2, because respiration of newly fixed carbon does not lead to a net release of CO_2.

The ultimate effects of global change on microbial activities, and *vice versa*, often relate to the complex interplay between physical, chemical, and biological factors. For example, in the sea, increased temperatures could cause changes in circulation patterns and an increase in aquatic stratification that might lead to a decrease in aquatic production due to the reduced availability of nitrogen and phosphorus nutrients. This could lead to the rain of particulate organic matter from surface waters to the deep sea, the so-called biological pump of carbon. Thus, less carbon might be removed from the atmosphere by this mechanism, and this can be a positive feedback with regard to global warming. However, would increased winds or storms cause deeper mixing and reduction of the stratification? This could have the opposite effect.

Microorganisms carry out especially important roles in the biogeochemical cycling of two other greenhouse gases besides CO_2, namely methane and nitrous oxide. The biological production of these substances is microbial, with the methanogenic archaea producing methane and the denitrifying bacteria producing nitrous oxide. Both gases also have significant non-biological or anthropogenic sources as well. Interestingly, the biological removal of both of these gases is also microbiological. Methane is oxidized by specialized bacteria (called *methanotrophs*), organisms that live where they have a supply of methane (most often from an anoxic zone) and oxygen. Some methanotrophs live as symbionts in invertebrates near natural methane sources, such as methane seeps. Nitrous oxide can be reduced by denitrifying bacteria to N_2 gas under anoxic conditions. Although it is known that both methane and nitrous oxide are increasing in the atmosphere, there is still considerable uncertainty in the accounting of supply and removal processes and uncertainty about how much of the change

is due to biological processes. Therefore, it is difficult to make mechanistic predictions about how microorganisms might lead to or respond to future changes in these gases.

Another microbial process of note in this regard is the formation and removal of dimethyl sulfide (DMS) (*see* Dimethylsul de (DMS), Volume 1). This gas is an important regulator of climate because it leads to the production of sulfate aerosols that form nuclei necessary for cloud formation over much of the planet. Clouds affect re ection of sunlight into space and also the distribution of precipitation. DMS is formed from dimethylsulfonyl proprionate (DMSP), which is an osmotic regulator in certain groups of microscopic marine algae. DMSP is converted to DMS biologically by a variety of organisms (microbes and small herbivores). Its production and release to the atmosphere is controlled by the combined activities of these microbes and animals, as well as physical mixing and gas exchange (Kiene and Bates, 1990). We know very little about the details of what controls the biological component of this production. Therefore, it is extremely difficult to predict how global change will affect this important gas.

Ocean Iron Fertilization

The primary production of much of the open ocean is thought to be limited by the availability of iron. The connection is that iron is needed for the enzymes that are used to reduce nitrate (a common marine nitrogen source) and also for bacteria that fix N_2 gas. Iron arrives on airborne dust from remote arid regions, and a change in world climate and wind patterns can change how iron in dust is distributed, and therefore affect marine productivity (Falkowski *et al.*, 1998; Falkowski *et al.*, 2000).

Because iron limits primary productivity in some marine environments, proposals have been made to deliberately fertilize large tracts of the ocean with iron, in an attempt to stimulate production and draw CO_2 from the atmosphere. This approach would result in the human engineering of a planetary geochemical cycle, via microbes, on a massive scale. A few pilot research projects have shown that one can indeed stimulate productivity with iron additions (Boyd *et al.*, 2000). However, the long-term removal of CO_2 from the atmosphere is not a clear result from these projects, and there are major concerns that large-scale fertilization may cause greater warming or other problems in the long run from unplanned geochemical consequences of fertilization (Fuhrman and Capone, 1991). For example, because the idea is to permit the use of available nitrogen and therefore drive the nitrogen cycle faster, there are concerns that microbial nitrification and denitrification associated with the fertilization may cause the release of enough N_2O to cause greenhouse warming by itself. Similarly, the deep sea microbial respiration stimulated by iron fertilization may make extensive tracts of the deep sea anaerobic, killing most sea life in the area and altering geochemical cycles unpredictably. Effects on methane and DMS are also hard to predict. Therefore, the scientific basis of iron fertilization as a partial solution to global warming is lacking. Indeed, such fertilization might make matters much worse (*see also* Iron Cycle, Volume 2).

REFERENCES

Barghoorn, E S and Tyler, S A (1965) Microorganisms from the Gun int Chert, *Science*, 147, 563–577.

Boyd, P W, Watson, A J, Law, C S, Abraham, E R, Trull, T, Murdoch, R, Bakker, D C E, Bowie, A R, Buesseler, K O, Chang, H, Charette, M, Croot, P, Downing, K, Frew, R, Gall, M, Hadfield, M, Hall, J, Harvey, M, Jameson, G, Laroche, J, Liddicoat, M, Ling, R, Maldonado, M T, Mckay, R, Nodder, S, Pickmere, S, Pridmore, R, Rintoul, S, Safi, K, Sutton, P, Strzepek, R, Tanneberger, K, Turner, S, Waite, A, and Zeldis, J (2000) A Mesoscale Phytoplankton Bloom in the Polar Southern Ocean Stimulated by Iron Fertilization (In Process Citation), *Nature*, 407(6805), 695–702.

Falkowski, P, Scholes, R J, Boyle, J, Canadell, J, Canfield, D, Elser, J, Gruber, N, Hibbard, K, Hogberg, P, Linder, S, MacKenzie, F T, and Moore, III, B (2000) The Global Carbon Cycle: a Test of Our Knowledge of Earth as a System (In Process Citation), *Science*, 290(5490), 291–296.

Falkowski, P G, Barber, R T, and Smetacek, V (1998) Biogeochemical Controls and Feedbacks on Ocean Primary Production, *Science*, 281(5374), 200–207.

Fuhrman, J A (1999) Marine Viruses: Biogeochemical and Ecological Effects, *Nature*, 399, 541–548.

Fuhrman, J A and Capone, D G (1991) Possible Biogeochemical Consequences of Ocean Fertilization, *Limnol. Oceanogr.*, 36(8), 1951–1959.

Hugenholz, P, Goegel, B M, and Pace, N R (1998) Impact of Culture Independent Studies on the Emerging Phylogenetic View of Bacterial Diversity, *J. Bacteriol.*, 180, 4765–4774.

Kiene, R P and Bates, T S (1990) Biological Removal of Dimethylsulphide from Seawater, *Nature*, 345(6277), 702–705.

Schidlowski, M (1988) A 3800 Million Year Isotopic Record of Life from Carbon in Sedimentary Rocks, *Nature*, 333, 313–318.

Schopf, J W (1993) Microfossils of Early Archean Apex Chert: New Evidence of the Antiquity of Life, *Science*, 260, 640–646.

Whitman, W B, Coleman, D C, and Wiebe, W J (1998) Prokaryotes: the Unseen Majority, *Proc. Natl. Acad. Sci. USA*, 95(12), 6578–6583.

Woese, C R, Kandler, O, and Wheelis, M C (1990) Towards a Natural System of Organisms: Proposal for the Domains Archaea, Bacteria and Eucarya, *Proc. Natl. Acad. Sci. USA*, 87, 4576–4579.

Xiong, J, Fischer, W, Inoue, K, Nakahara, M, and Bauer, C E (2000) Molecular Evidence for the Early Evolution of Photosynthesis, *Science*, 289, 1724–1730.

Migration

see Plant Dispersal and Migration (*Opening essay*, **Volume 2**)

MIOMBO Network

see IGBP Core Projects (*Opening essay*, **Volume 2**)

Monsi, Masami

(1914- 1997)

Masami Monsi, Japanese plant ecologist and environmental scientist, Professor of the University of Tokyo 1954–1975, started his career with studies on the regulation of stomatal opening and plant transpiration, but his most important and well known contribution to plant ecology is the functional analysis of canopy structure (Monsi and Saeki, 1953). Monsi developed his idea from P. Boysen Jensen's classical work *Die Stoffproduktion der P anzen* (1932). He paid particular attention to the light climate as a factor to determine the structure and function of plant canopies. He considered dry matter production to be the most important function of plant canopies and invented the stratified clipping method to determine the productive structure of a plant stand, which describes the vertical distribution of light climate, biomass and leaf area through the canopy. He first formulated the relationship between light climate (I) and cumulative leaf area index (F) as $I = I_0 e^{-kF}$, on which he built a mathematical model of canopy photosynthesis. The model predicted the optimal canopy architecture (leaf area index, leaf size and angle, and leaf distribution) to maximize canopy photosynthesis, which was later successfully applied to breed rice varieties to increase agricultural production. He further applied his theory for the analysis of plant succession after gap formation, with which he pioneered model simulation of vegetation dynamics (Monsi and Oshima, 1955). In 1960 he presented a comprehensive scheme to model growth and reproduction of plants with different life histories, based on photosynthesis in leaves and matter allocation to other organs (Monsi, 1960). From 1964 he was involved in organizing international programs such as the International Biological Program (IBP, International Council for Science (ICSU)) and Man and the Biosphere (MAB, United Nations Educational, Scientific and Cultural Organization (UNESCO)). He led the Research Group of Production Processes of IBP where his theory and model on dry matter production and plant growth were further developed and accepted internationally (Monsi, 1968; Monsi and Murata, 1971).

REFERENCES

Monsi, M and Saeki, T (1953) Uber den Lichtfaktor in den P anzengesellschaften und seine Bedeutung fur die Stoffproduktion, *Jap. J. Bot.*, 14, 22–52.

Monsi, M and Oshima, Y (1955) A Theoretical Analysis of the Succession of Plant Community based on the Production of Matter, *Jap. J. Bot.*, 175, 60–82.

Monsi, M (1960) Dry-matter Reproduction in Plants 1. Schemata of Dry-matter Reproduction, *Bot. Mag. (Tokyo)*, 73, 81–90.

Monsi, M (1968) Mathematical Models of Plant Communities, in *Functioning of Terrestrial Ecosystems at the Primary Production Level*, ed F Eckardt, UNESCO, Paris, 131–149.

Monsi, M and Murata, Y (1971) Development of Photosynthetic Systems as In uenced by Distribution of Matter, in *Prediction and Measurement of Photosynthetic Productivity*, ed I Setlik, Pudoc, Wageningen, 115–129.

TADAKI HIROSE *Japan*

Mooney, Harold A

(1932-)

Harold A Mooney holds the Paul S Achilles Professorship in Environmental Biology at Stanford University. Born in 1932 in California, he received his PhD from Duke University in 1960. Mooney has made major contributions to plant and ecosystem ecology and environmental biology. A renowned international expert, he has been invited to advise governments on a variety of ecological issues such as the effects of global change on the terrestrial biosphere.

His research on the carbon balance of plants has provided a major theoretical framework for ecophysiological studies, and has been instrumental in the incorporation of physiological understanding to studies of ecosystem processes. This work has also led to several lines of research on the nature of interactions of plants with their biotic environment, and has provided an objective measure for evaluating many of the current theories of plant–animal interaction. He has demonstrated that convergent evolution takes place

in the properties of different ecosystems that are subject to comparable climates, and has pioneered in the study of the allocation of resources in plants. He has worked in many of Earth's diverse ecosystems, including the arctic alpine, the Mediterranean climate scrub and grasslands, tropical wet and dry forests, and the deserts of the world.

Mooney has been engaged for over a decade in the study of the impacts of global change on terrestrial ecosystems, with an emphasis in understanding processes that control uxes and their quantities in a changing environment. A second major emphasis has been on understanding the role of biotic assemblages and how they are affected by multiple stressors, and examining the factors that promote the invasions of non-indigenous plant species.

Mooney has been instrumental in developing the new global ecology as we know it today by being involved in the organization of international activities through which he has brought together scientists from many diverse disciplines to address topics that promise to contribute substantially to the advancement and integration of ecology. Mooney was deeply involved in the International Biosphere Programme (IBP), and he was instrumental in the development of the International Geosphere Biosphere Programme (IGBP) in the early 1990s, specifically with its core project Global Change and Terrestrial Ecosystems (GCTE) in which he served as a scientific adviser for many years. Most recently, Mooney was instrumental in the development of the new international program, the Global Invasive Strategies Program (GISP). He is also a force behind the International Biodiversity Observation Year (IBOY) to take place in 2001–2002. Currently he is serving as a panel co-chair for the Millennium Ecosystem Assessment, a program devoted to strengthening our capacity to manage ecosystems sustainably for human wellbeing. He is also Secretary General of the International Council of Scientific Unions (ICSU), now called International Council for Science, and was earlier Secretary General of SCOPE. Through these efforts and his lengthy publication record of over 400 scientific books, papers, and articles, he has developed bridges between physiological ecology and other areas of ecology, and he has explored the contributions that ecologists can make to resolving the growing problems of global habitability.

Mooney has served on many editorial boards for ecological journals, and on advisory committees of many funding agencies, universities, and national and international agencies. He served as President of the Ecological Society of America 1988–1989, President of the American Institute of Biological Sciences in 1993, Secretary General of Scientific Committee on Problems of the Environment, 1987–1988, and is currently Secretary General of the International Council for Science. He is an editor of *Flora*, *Trends in Ecology and Evolution*, *Tree Physiology*, *Environment and Pollution*, *Global Change Biology*, *Global Environmental Research*, *Ecosystems*, *Journal of Mediterranean Ecology*, and *Biological Invasions*, as well as series editor of *Physiological Ecology* (Academic Press), and *Ecological Studies* (Springer-Verlag).

Among his many honors, he was elected to the National Academy of Sciences, the American Academy of Arts and Sciences, and the American Philosophical Society. He has received the Eminent Ecologist Award and the Mercer Award of the Ecological Society of America, a Humboldt Senior Distinguished US Scientist Award, the Max Planck Research Award, the Ecology Institute Prize for Terrestrial Ecology and the Nevada Medal. He is a Fellow of the American Association for the Advancement of Science and a foreign member of the Russian Academy of Sciences. He worked in Chile and France as a National Science Foundation Senior Postdoctoral Fellow and in Australia and Africa as a Guggenheim Fellow.

Mooney is a person of exceptional modesty, enthusiasm, and never ending energy. Combined with his skills as an outstanding speaker and sincere commitment to science, he has inspired hundreds of scientists in the US and worldwide, and in uenced policy agencies and governments to shape a more environmentally sound and sustainable world.

JOSEP G CANADELL *Australia* AND ANN SCHRAM *USA*

Mutualism

see Symbiosis (**Volume 2**)

N

Natural Capital

Natural capital is what nature provides: air to breathe, water to drink, food to eat, fiber, living space, waste disposal and other things. Traditionally, the word *capital* is thought of as a stock of materials such as tools, machinery or buildings that generate a flow of services more commonly known as profits. The tools, machinery or buildings are known as the manufactured capital. But there are other types of capital such as human capital, and natural capital. The essential feature of all forms of capital is that capital provides a capacity for excess production, that is, a capacity for growth. It is this feature that links the traditional forms of capital with human capital, cultural capital and natural capital.

Natural capital refers to that stock of materials provided to us by nature free of charge: ecosystems, soils, trees, air and water. The goods and ecosystem services provided by natural capital are vast in scale, and fundamental to human survival on the planet. As an example, consider the naturally fixed nitrogen in the soil and the ecosystem service of providing this critical component for plant growth. The cost of replacing the natural capital with artificial fertilizers is prohibitive. As yet, these ecosystem services are not fully captured in commercial markets or adequately quantified, so as they are often given too little weight in policy decisions involving natural resources and the environment. By attempting to price services such as water purification, pest control or climate amelioration, we can make quantitative estimates of what is lost when we degrade the free services that nature provides. Tailored to specific regions or watersheds, such information greatly improves our ability to take explicit account of the tradeoffs we make in the course of decisions about land use and development.

Economists have been trying for many years to put an economic value on natural capital. Costanza *et al.* (1997) estimated that the total value of all the Earth's ecosystem services stemming from natural capital totaled $33 trillion per year, a figure that exceeds the sum of the world's gross national products. Others have called the $33 trillion per year a serious underestimate of infinity, as without the services from natural capital, humans would not exist.

An alternative approach to valuing the ecosystem goods and services is to bring an ecosystem's natural capital to the market by some mechanism. One method is to evaluate the local ecosystems capacity for growth and to issue tradable permits for economic development appropriate to that growth potential. The marketplace will then value the reduction of the particular ecosystem's goods and services by the changing price of the tradable permits.

See also: **Ecosystem Services and Costing**, Volume 2.

REFERENCES

Costanza, R, d'Arge, R, de Groot, R, Farber, S, Grasso, M, Hannon, B, Naeem, S, Limburg, K, Paruelo, J, O'Neill, R V, Raskin, R, Sutton, P, and van den Belt, M (1997) The Value of the World's Ecosystem Services and Natural Capital, *Nature*, **387**, 253–260.

Prugh, T (1995) Natural Capital and Human Economic Survival, ISEE Press, White River Junction, VT.

ADAM FENECH AND ROGER HANSELL *Canada*

NBP (Net Biome Productivity)

see Productivity (**Volume 2**)

NDVI (Normalized Ratio of the Near-infrared and the Red Spectral Bands

see Remote Sensing (**Volume 2**)

NEP (Net Ecosystem Productivity)

see Productivity (**Volume 2**)

Net Primary Productivity (NPP)

see Productivity (**Volume 2**)

NEWS (GCTE Network of Ecosystem Warming Study Sites)

see Temperature Increase: Effects on Terrestrial Ecosystems (**Volume 2**)

Niche

Each species in a community is a specific part of that community, and it occupies a specific place, or *niche*, in the structure of that community. The niche is far more than simply the physical or spatial location of each species in the ecosystem; in addition it expresses the concept of the functional role that each species has in the totality of the community. A species has a *fundamental niche*, which is the total environmental range that it is adapted to occupy, and a *realized niche*, which is that reduced portion of its fundamental niche that it is forced into by competition.

The niche occupied by some species may be very broad, encompassing a variety of habitats and food sources. Species occupying such broad niches are usually strongly dominant in the ecosystem. Other species may occupy a narrow or highly specialized niche. In either case the evolution of successful adaptations enables the species to compete successfully for the occupation of the niche. A niche is never occupied by more than one species, although some overlapping of ecologically adjacent niches may occur. As is the case for the diversity of species, the diversity of niches in an ecosystem is directly correlated with its stability and its ability to withstand environmental or climatic insult. A large number of narrow niches fosters community sharing rather than competition for its resources, which increases the efficiency of resource use. Agriculture, by introducing a small number of strongly dominant species occupying broad niches, greatly reduces the stability of an ecosystem, decreases its environmental tolerance, and ultimately leads to its deterioration. It may be of interest to note that the human species has evolved to occupy a broader niche with a greater degree of dominance than has ever occurred previously in the world ecosystem.

R G S BIDWELL *Canada*

Nitri cation

see Nitrogen Cycle (**Volume 2**)

Nitrogen Cycle

Robert W Howarth
Environmental Defense, Boston, MA, USA

Nitrogen (N_2) is an element essential for all life and is the fourth most abundant element in living biomass, after hydrogen, carbon, and oxygen. Nitrogen is of central importance to the biogeochemistry of the earth, since compared to the needs of plants, it is frequently the most scarce of the essential elements. Thus, primary production is limited by nitrogen in most temperate and arctic terrestrial ecosystems and in most estuaries and coastal oceans (Vitousek and Howarth, 1991). Nitrogen, together with phosphorus, also co-limit primary production in many of the open ocean areas of the planet (National Research Council, 2000). Further, nitrogen availability is a key factor regulating the rate of decomposition of organic matter, so the nitrogen cycle is intimately tied to the sequestration of carbon dioxide (CO_2) in both the oceans and terrestrial ecosystems. Indeed, changes in the nitrogen cycle over the past decades may be responsible for increasing the storage of carbon in the world's forests to a globally significant extent. Nitrogen is therefore a major regulator of the CO_2 concentration in the atmosphere and of the rate of global warming (Vitousek et al.*, 1997).*

Of the major elements of biogeochemical importance, nitrogen is undergoing the most rapid change globally. Most nitrogen on Earth is present as gaseous, elemental N_2 that is largely unavailable to plants. A century or more ago, the only major mechanism whereby this molecular N_2 became available to most organisms was by bacterial nitrogen fixation, the conversion of element nitrogen into more reactive and biologically available forms However, several types of human activity also convert N_2 to more reactive and biologically available forms. Over the past few decades, these processes under human control have increased to the point that they have roughly doubled the biological availability of nitrogen on the land surfaces of the planet. That is, since the 1960s human activity has changed the nitrogen cycle from one that was primarily controlled by biological processes to one that is now dominated by human-generated pollution. Although nitrogen is essential to life, too much nitrogen is deleterious to life, just as consumption of too much food by humans leads to obesity.

The acceleration of the nitrogen cycle is leading to a large variety of very serious environmental consequences (Vitousek et al.*, 1997). Nitrogen pollution is now the major source of acid rain in developed countries, and perhaps globally. Nitrogen pollution in the atmosphere also contributes markedly to smog and ozone formation in the lower atmosphere, to the depletion of ozone in the ozone layer of the stratosphere, and to the formation of fine particles that diminish visibility and pose a serious public health hazard. One gaseous form of nitrogen pollution in the atmosphere contributes to greenhouse warming. Nitrogen is now the largest pollution problem in the coastal oceans of the US and perhaps worldwide, causing dead zones devoid of oxygen, increased blooms of toxic algae, and a loss of biodiversity. More than half of the coastal rivers and bays in the US are degraded from nitrogen pollution (National Research Council, 2000). Accelerated nitrogen cycling on land is also a major contributor to the loss of biotic diversity in terrestrial ecosystems, and in extreme cases, can lead to the dieback of forests (Vitousek* et al.*, 1997; Tartowski and Howarth, 2000).*

The vast majority of nitrogen on earth is present as elemental nitrogen (N_2), with most of this being in the atmosphere (3.9×10^{21} g) and some dissolved in the oceans. N_2 is the most abundant gas in the atmosphere and makes up 78% of its volume. Of all nitrogen on Earth, only 0.002% is present in living tissues and detrital organic matter on land and in the oceans (Schlesinger, 1997). Nonetheless, this small amount of organic nitrogen is critical to life, as its recycling supplies most of the nitrogen that supports plant growth. Elemental N_2 is quite unreactive because of the extreme strength of the chemical N−N bond. However, under appropriate conditions, N_2 can react with other elements such as oxygen, hydrogen and carbon to form a variety of compounds that are both more chemically reactive and more biologically available. The process of this conversion is called *nitrogen fixation* (*see* **Nitrogen Fixers**, Volume 2).

Prior to the disruption of the global nitrogen cycle by human activity, most nitrogen fixation on earth was biologically mediated through an enzymatic process catalyzed by bacteria that produces the reactive form ammonia (NH_3). Many different types of bacteria are capable of this process, including photosynthetic cyanobacteria and symbionts of some plants, but no higher forms of life are capable of nitrogen fixation without bacterial symbionts. Some nitrogen is also fixed naturally by lightning and by volcanic action; in this case, the element N_2 is converted to nitrate (NO_3^-), another biological form of nitrogen, as extreme heat catalyzes the reaction with atmospheric oxygen. Prior to large-scale alteration of the landscape by humans, bacterial nitrogen fixation on the land surface was probably in the range of 90–195 Tg N year^{-1} (Vitousek *et al.*, 1997; Cleveland *et al.*, 1999). An additional 200–300 Tg N year^{-1} is probably fixed by cyanobacteria in the oceans of the world, although there is much uncertainty associated with this estimate (Karl *et al.*, 2002). Thus, before humans began to alter the nitrogen cycle, nitrogen fixation by bacteria globally was probably in the range of 300–500 Tg N year^{-1}. Nitrogen fixation by lightning is much smaller, probably less than 10 Tg N year^{-1} globally. Given the tremendously large pool of elemental N_2 in the atmosphere and oceans, the conversion of N_2 to reactive forms through nitrogen fixation yields a turnover time of N_2 of approximately 10 million years.

Human activity has greatly altered the nitrogen cycle globally (Figure 1), with the single largest change resulting from the increased use of synthetic inorganic fertilizers in agriculture (Table 1). The process for making synthetic nitrogen fertilizer was invented during World War I by German chemist Fritz Haber, but the synthetic fertilizer was not used on a large scale until the 1950s. Use has increased greatly since then and is in large part responsible for the green revolution in agriculture that has reduced starvation

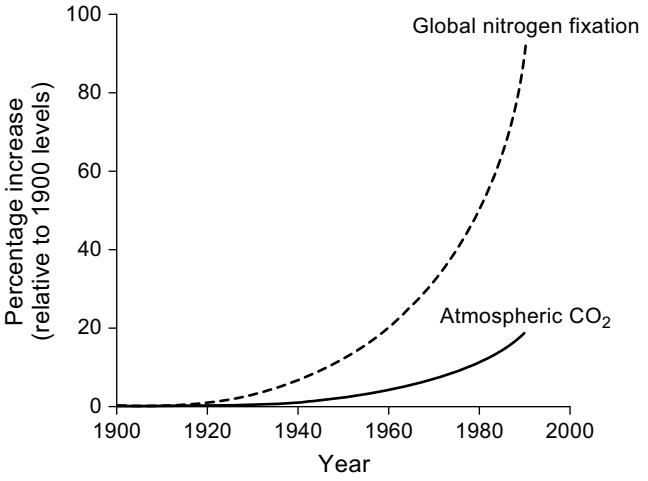


Figure 1 The relative change in the rate at which human activities have accelerated global nitrogen xation in comparison to the increase in atmospheric carbon dioxide since 1900. (Modi ed from Vitousek *et al.*, 1997)

Table 1 Best mean estimates for global rates of nitrogen xation on land (Tg N year^{-1})

	Before industrial revolution	Current
Natural terrestrial ecosystems	140	140
Lightning	5	5
Agricultural crops	0	40
Fertilizer production	0	83
Fossil fuel combustion	0	20
Total	145	288

and allowed the human population to grow so dramatically over the past few decades (Smil, 1997; *see* **Green Revolution**, Volume 3). As of 1996, the global use of inorganic nitrogen fertilizer was approximately 83 Tg N year^{-1}. The rate of growth in fertilizer use is so great that half of the inorganic nitrogen fertilizer that has ever been used on Earth has been applied during the last 15 years (National Research Council, 2000).

The application of nitrogen fertilizer is not uniform over the surface of the Earth. Use grew first in the US and other developed countries, but has been relatively constant since 1980. Globally, nitrogen fertilizer use decreased somewhat for a five-year period beginning in the later 1980s (Figure 2). This was a result of the collapse of the former Soviet Union, which led to widespread disruption of economic activity, including agriculture, in Russia and elsewhere in eastern Europe. The increase in fertilizer use in recent years has been driven largely by use in Asia. Today, China uses almost one-third of all inorganic nitrogen fertilizer produced on the planet.

Production of nitrogen fertilizer is the largest process whereby human activity has altered the global nitrogen cycle. However, other human-controlled processes, such as growing nitrogen-fixing crops in agriculture and the combustion of fossil fuels have also resulted in increased nitrogen fixation. Many agricultural crops such as soybeans, alfalfa, peas, and peanuts have symbiotic bacteria that fix nitrogen. Globally, this probably contributes some 40 Tg N year^{-1} of reactive, biologically available nitrogen (Vitousek *et al.*, 1997). The combustion of fossil fuels contributes another 20 Tg N year^{-1} to the global cycle

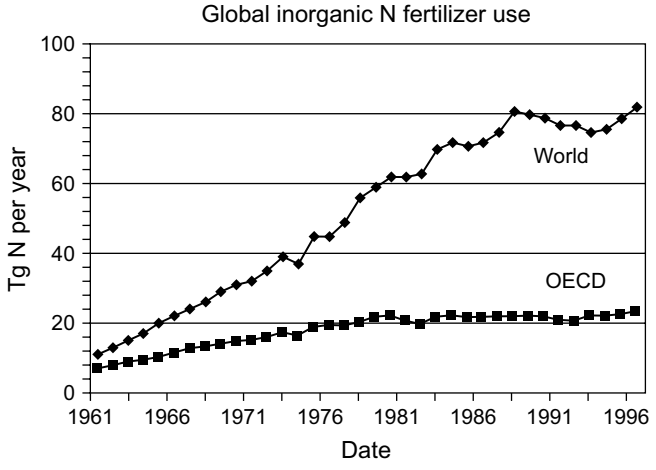


Figure 2 The global use of inorganic, synthetic nitrogen fertilizer since 1961 in the world as a whole and in the developed nations of the world (Organization for Economic Co-operation and Development - OECD), which include the US, Canada, western Europe, Japan and Korea. (Data are from the Food and Agricultural Organization [unpublished])

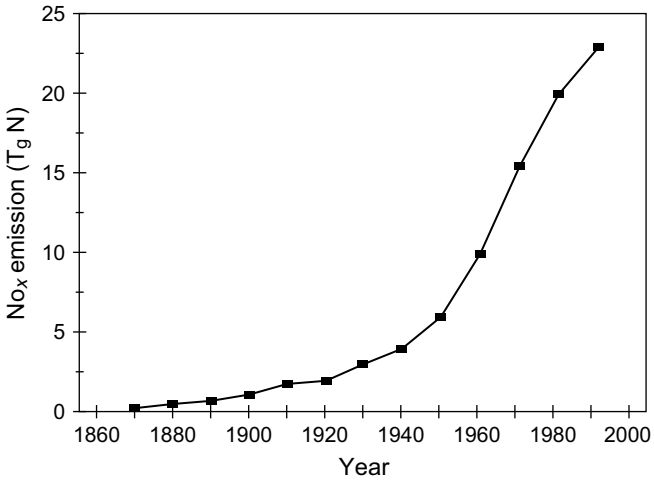


Figure 3 The global emission of NO_x to the atmosphere from human activities since 1860. (Estimates are from the World Meteorological Organization [unpublished])

of reactive nitrogen, and this too is increasing rapidly (Figure 3; Galloway *et al.*, 1995). Some of the nitrogen released from burning fossil fuels represents nitrogen contained in the fuels themselves, and some comes from the heat and pressure of combustion catalyzing the reaction between O_2 and N_2 in the atmosphere.

Collectively, the nitrogen fixation under control by humans now contributes some 140 Tg N year^{-1} world-wide. Given the estimates of natural nitrogen fixation by bacteria on the land surface of the Earth (90–195 Tg N year^{-1}, with a mean value of 140 Tg N year^{-1}), human activity has roughly doubled the availability of nitrogen to the terrestrial ecosystems of the planet. This nitrogen fixation through human activity increased by some two- to three-fold over the three decades between 1960 and 1990, and it continues to grow (Galloway *et al.*, 1995).

Although humans have altered the nitrogen cycle globally, the intensity of the alteration varies greatly among regions. This is because both combustion of fossil fuels and agricultural activity are most intense in developed countries and in Asia, and because most reactive nitrogen is not transported very far from where it is released. So far, the acceleration of the nitrogen cycle has been more pronounced in north temperate regions than in the tropics, since most industrial activity and agriculture have developed in the north temperate regions (Holland *et al.*, 1999). Further, the natural rates of biological nitrogen fixation on land are far greater in the tropics than in temperate-zone ecosystems (Cleveland *et al.*, 1999). One measure of how humans have altered the nitrogen cycle is the rate at which reactive forms of nitrogen are deposited onto the land surface from the atmosphere in rain, snow, and dry deposition (the process whereby gases and small particles leave the atmosphere). On average worldwide, nitrogen deposition

from the atmosphere has probably doubled in the tropics and increased more than six-fold in the north temperate zone because of human acceleration of nitrogen cycling (Holland *et al.*, 1999). On the other hand, human activity has had little effect on nitrogen deposition in the temperate zone of the Southern Hemisphere. The flux of nitrogen in large rivers provides another measure of how humans have altered the nitrogen cycle in particular regions. As of 1990, human activity had increased the nitrogen flux in the watershed of the North Sea in Europe by more than 10-fold, in the rivers of the northeastern US by some eight-fold, down the Mississippi River by more than four-fold, and not at all in the rivers in northern Canada (Figure 4; Howarth *et al.*, 1996). The flux of nitrogen from the watersheds of large regions is proportional to the use of nitrogen by humans in the landscape (Figure 5).

Despite the huge reservoir of elemental N_2 on Earth, the process of nitrogen fixation would eventually – over a time period of 10 million years – deplete this pool if it were not for the process of *denitrification*. Denitrification is the bacterial conversion of nitrate to elemental N_2. In this process, the bacteria use nitrate as an electron acceptor

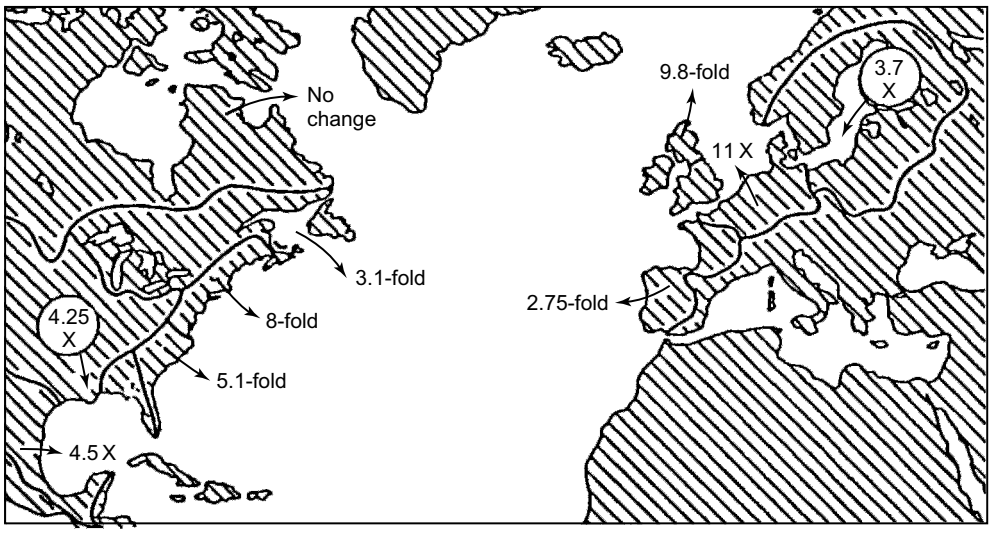


Figure 4 The extent to which human activity has altered the ux of nitrogen from the landscape to coastal waters. (Reprinted from Howarth *et al.*, 2000)

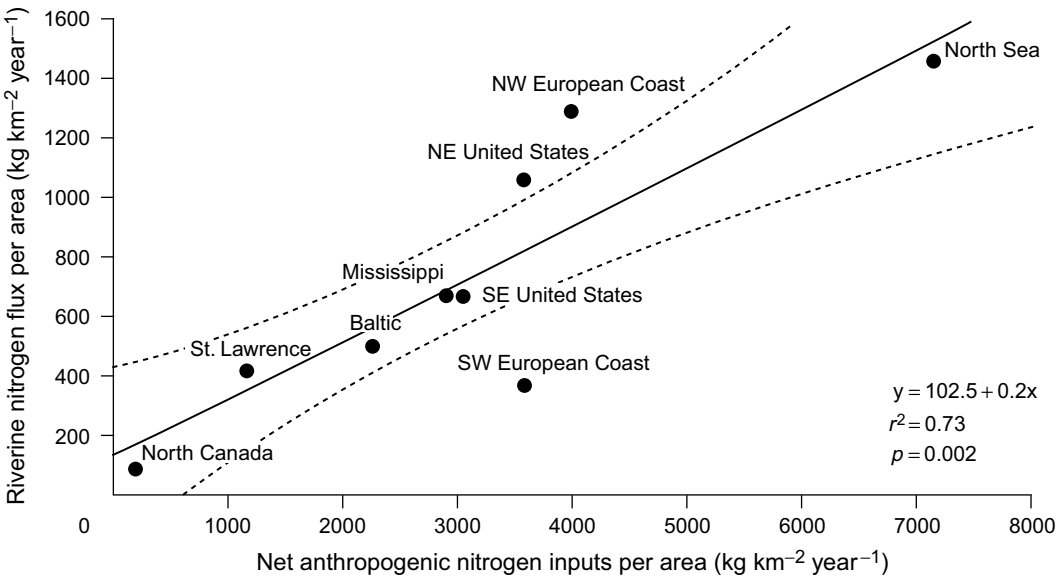


Figure 5 The export of nitrogen per area of watershed for different large regions surrounding the North Atlantic Ocean as a function of the inputs of nitrogen to those regions from various human activities. (Reprinted from Howarth *et al.*, 1996)

to support their respiration. Since this yields less energy to the bacteria than if they used O_2 as the electron acceptor, denitrification tends to occur only in environments where O_2 is absent, or where concentrations are low, such as in the sediments of wetlands, lakes, rivers, and coastal oceans. As human activity has increased the flux of reactive nitrogen through the landscape, rates of denitrification have increased in direct proportion (Nixon *et al.*, 1996; Howarth *et al.*, 1996).

Even to understand the nitrogen cycle in its simplest terms, we need to consider a few other processes (Figure 6). Plants and algae assimilate most of the nitrogen they need for growth as NH_3 or as nitrate. They convert this to the organic nitrogen that comprises their tissues. As this material is eaten by animals or decomposed by microorganisms, most of the organic nitrogen is recycled to the environment as ammonia. In the presence of O_2, certain types of bacteria

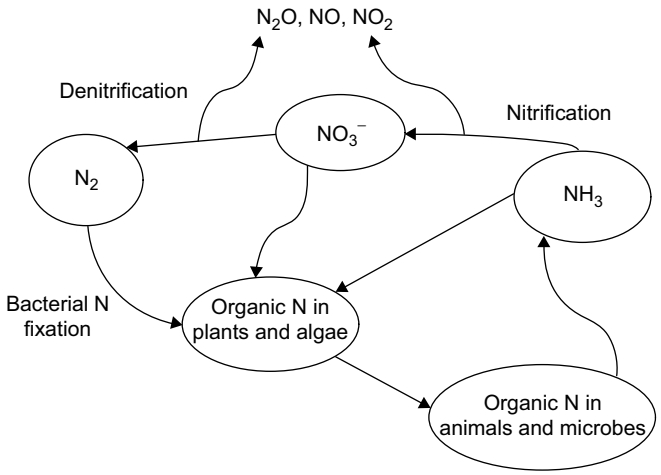


Figure 6 A simpli ed representation of the natural nitrogen cycle. Note that the atmospheric gases N_2O, NO, and NO_2 come from leakages from both the processes of nitri cation and denitri cation. A human acceleration of the nitrogen cycle increases these processes, resulting in greater uxes of these pollutant gases to the atmosphere

can oxidize the NH_3 to nitrate, and thereby gain energy to support their growth, in a process called *nitrification*. In many sediments, zones that are oxygen-free coexist in close proximity to zones that have oxygen, allowing NH_3 to be oxidized to nitrate and denitrified to N_2 in coupled processes in close proximity.

As nitrogen is converted by bacteria in the processes of nitrification and denitrification, some leaks to the environment as nitrous oxide (N_2O), and most of this probably diffuses to the atmosphere. Globally, the current flux of N_2O to the atmosphere is approximately 16 Tg N year^{-1} (Table 2; Schlesinger, 1997). The mean residence time of this gas in the atmosphere is relatively long, 120 years. This is of global significance, as N_2O is a major contributor to the destruction of ozone in the ozone layer of the stratosphere and is a potent greenhouse gas, some 300-fold more so than is CO_2 per molecule (Schlesinger, 1997). The mean concentration of N_2O in the atmosphere as of the 1990s was approximately 310 ppb, or some 1000-fold less than the concentration of CO_2. Currently, the concentration of N_2O in the atmosphere is increasing at a rate of 0.3% per year. Data from ice cores show that the concentration was lower and stable for the thousand or so years before the start of the Industrial Revolution, at approximately 280 ppb (Schlesinger, 1997). The increase is largely the result of the general human acceleration of the nitrogen cycle, which is increasing the processes of denitrification and nitrification in many different ecosystems globally, although some of the nitrous oxide comes directly as a byproduct of industrial activity, such as the synthesis of nylon (Vitousek *et al.*, 1997). Although many different types of ecosystems contribute to the flux of N_2O to the atmosphere, the fluxes are particularly high from coastal oceans and from tropical soils. The fluxes from tropical soils increase markedly in areas where forests are converted to agricultural lands and are fertilized (Schlesinger, 1997).

In addition to N_2O, the other major forms of reactive nitrogen gases in the atmosphere are nitric oxide (NO), nitrogen dioxide (NO_2), and NH_3. NO and NO_2 are

Table 2 Nitrogen gases in the atmosphere: uxes to the atmosphere under current and pristine conditions, and concentrations and residence times in the atmosphere[a]

	Flux before Industrial Revolution (Tg N year^{-1})	Current ux (Tg N year^{-1})	Concentration (ppb)	Residence time (years)
N_2[b]	−140	−283	780×10^6	10×10^6
N_2O	10	16	310	120
NO_x	12	36	<10	0.003
NH_3	15	46	<10	0.003

ppb, parts per billion.
[a] For nitrogen xation, the ux is from the atmosphere to the land surface, and so is indicated here as a negative ux.
[b] Considering only uptake by nitrogen xation on the land surface.

collectively referred to as NO_x. In contrast to N_2O, the concentrations for NO_x in the atmosphere are small, only 0.02–10 ppb, and the residence times are short, approximately one day (Schlesinger, 1997). NO_x gases are typical of polluted air, and they are instrumental in catalyzing the production of ozone pollution and smog in the troposphere. NO_x is oxidized to nitric acid in the atmosphere, and this nitric acid is a major contributor to acid rain. The other major contributor, sulfuric acid, has been more effectively controlled in developed countries over the past two decades, so that NO_x is increasingly becoming the dominant contributor to acid rain in the US and western Europe. The best current estimate is that the global flux of NO_x to the atmosphere is 36 Tg N year^{-1}, although there is considerable uncertainty associated with this estimate (Holland *et al.*, 1999). Most of this (20 Tg N year^{-1}) comes from the combustion of fossil fuels, with electric-power generating plants, cars, trucks, buses, and off-road vehicles all contributing. Other sources are lightning, biomass burning, and leakage from the microbial processes of nitrification and denitrification. Human activity has probably increased the total flux of NO_x to the atmosphere by some three-fold, with the largest increases being in the eastern US, in Europe, and increasingly, in east Asia (Holland *et al.*, 1999). The change has been recent, and rapid (Figure 3).

Human activity has also increased the flux of NH_3 to the atmosphere by three-fold or more globally (Table 2), although the emissions prior to human acceleration of the nitrogen cycle are poorly known. Currently, the flux is estimated as 45 Tg N year^{-1} or more (Holland *et al.*, 1999). As with NO_x, concentrations in the atmosphere are low, and the residence time is short. NH_3 is a base, and so unlike NO_x, it does not directly contribute to acid rain. However, if the NH_3 that is deposited onto the land or into a lake is nitrified to nitrate, 2 moles of protons are released for every mole of nitrogen oxidized. Thus, per atom of nitrogen, NH_3 actually has twice the capacity to contribute to lake and soil acidification as does nitric acid. NH_3 is also a major contributor to the creation of fine aerosol particles in the atmosphere, which reduce visibility and pose a public health threat to humans.

The flux of NH_3 from domestic animal wastes is the single largest source to the atmosphere, contributing roughly half of the global total. Other major sources are volatilization from inorganic nitrogen fertilizers, biomass burning, and the natural flux from upwelling zones of the ocean (Holland *et al.*, 1999). The past few decades have seen a remarkable increase in the production of domestic animal wastes, as the world population has grown and the *per capita* consumption of meat has increased (Figure 7; National Research Council, 2000). In developed countries, most crop production is fed to animals, and so large quantities of the inorganic nitrogen fertilizer applied to fields are processed through animal wastes. In North America, 15% of

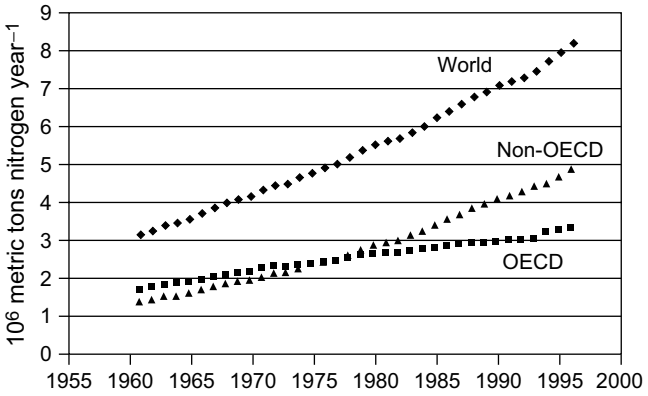


Figure 7 The increase in meat protein (expressed in terms of the nitrogen content of the meat) since 1961 globally, in the developed countries of the world (OECD), and in the developing countries (non-OECD). (Data are from the Food and Agricultural Organization [unpublished])

the inorganic nitrogen fertilizer that is used eventually ends up volatilized to the atmosphere as NH_3 (National Research Council, 2000). Globally, the figure is perhaps twice this high, due in part to the use of particularly volatile fertilizers in China (Holland *et al.*, 1999; National Research Council, 2000). The huge flux of NH_3 to the atmosphere is a very recent phenomenon, a direct result of the green revolution.

See also: **Nitrogen Deposition on Forests**, Volume 2; **Nitrogen Fixers**, Volume 2; **Nitrate Protection Zones in Europe**, Volume 3; **Nitrogen: Agricultural Uses and their Impacts**, Volume 3; **Trends in Global Emissions: Carbon, Sulfur, and Nitrogen**, Volume 3.

REFERENCES

Cleveland, C C, Townsend, A R, Schimel, D S, Fisher, H, Howarth, R W, Hedin, L O, Perakis, S S, Latty, E R, von Fischer, J C, Elseroad, A, and Wasson, M F (1991) Global Patterns of Terrestrial Biological Nitrogen (N_2) Fixation in Natural Ecosystems, *Global Biogeochem. Cycles*, **13**, 623–645.

Galloway, J N, Schlesinger, W H, Levy, H, Michaels, A, and Schnoor, J L (1995) Nitrogen Fixation: Atmospheric Enhancement-environmental Response, *Global Biogeochem. Cycles*, **9**, 235–252.

Holland, E A, Dentener, F J, Braswell, B H, and Sulzman, J M (1999) Contemporary and Pre-industrial Global Reactive Nitrogen Budgets, *Biogeochemistry*, **46**, 7–43.

Howarth, R W, Billen, G, Swaney, D, Townsend, A, Jaworski, N, Lajtha, K, Downing, J A, Elmgren, R, Caraco, N, Jordan, T, Berendse, F, Freney, J, Kudeyarov, V, Murdoch, P, and Zhu, Z L (1996) Regional Nitrogen Budgets and Riverine N and P Fluxes for the Drainages to the North Atlantic Ocean: Natural and Human Influences, *Biogeochemistry*, **35**, 75–139.

Howarth, R W, Anderson, D, Cloern, J, Elfring, C, Hopkinson, C, Lapointe, B, Malone, T, Marcus, N, McGlathery, K,

Sharpley, A, and Walker, D (2000) Nutrient Pollution of Coastal Rivers, Bays, and Seas, *Issues Ecol.*, **7**, 1–15.

Karl, D, Michaels, A, Bergman, B, Carpenter, E, Letelier, R, Lipschults, F, Paerl, H, Sigman, D, and Stal, L (2002) Dinitrogen Fixation in the World's Oceans, *Biogeochemistry*, in press.

National Research Council (2000) *Clean Coastal Waters: Understanding and Reducing the Effects of Nutrient Pollution*, National Academy Press, Washington, DC.

Nixon, S W, Ammerman, J W, Atkinson, L P, Berounsky, V M, Billen, G B, Boichourt, W C, Boynton, W R, Church, T M, DiToro, D M, Elmgren, R, Garber, J H, Giblin, A E, Jahnke, R A, Owens, N J P, Pilson, M E Q, and Seitzinger, S P (1996) The Fate of Nitrogen and Phosphorus and the Land–Sea Margin of the North Atlantic Ocean, *Biogeochemistry*, **35**, 141–180.

Schlesinger, W H (1997) *Biogeochemistry: an Analysis of Global Change*, Academic Press, San Diego, CA.

Smil, V (1997) Global Population and the Nitrogen Cycle, *Sci. Am.*, July, 76–81.

Tartowski, S L and Howarth, R W (2000) Nitrogen Cycling in *Encyclopedia of Biodiversity*, in press.

Vitousek, P M and Howarth, R W (1991) Nitrogen Limitation on Land and in the Sea: How Can it Occur? *Biogeochemistry*, **13**, 87–115.

Vitousek, P M, Aber, J, Howarth, R W, Likens, G E, Matson, P A, Schindler, D W, Schlesinger, W H, and Tilman, G D (1997) Human Alteration of the Global Nitrogen Cycle: Causes and Consequences, *Issues Ecol.*, **1**, 1–15.

Nitrogen Deposition on Forests

Peter Högberg

Department of Forest Ecology, Umea, Sweden

Industrial fixation of nitrogen (N) in the production of fertilizer for agriculture and fixation of nitrogen in combustion processes in energy production, industry and transport now exceed natural biological N_2 fixation. Hence, the input of nitrogen to the Earth's ecosystem has doubled in a century. Simultaneously, nitrogen has turned from being a growth-limiting nutrient to a problematic pollutant. Problems of excess nitrogen are found in northern temperate forests close to major agricultural and industrial areas, in particular in central Europe and northeastern North America. At low levels, nitrogen deposition increases tree growth in initially nitrogen-limited forests. However, more sensitive plant species and other organisms can become negatively affected already at low levels of nitrogen deposition. Ultimately, high continuous supply of nitrogen leads to soil acidification, pollution of ground and surface waters by nitrate and inorganic aluminum (Al), changes in species composition, and can contribute to forest decline. Some of these adverse effects reverse rapidly if the nitrogen load is removed, while others will have a more lasting impact. It is difficult to reduce nitrogen emissions because of the tight coupling between the large nitrogen inputs to agriculture and the demand for food by a growing population, and also because of the increasing use of cars worldwide. It can be predicted that the problem of nitrogen deposition will increase for these reasons in the developing world, especially in East and Southeast Asia. This will be particularly serious in the tropics and sub-tropics, where natural ecosystems are often phosphorus (P) limited, and where available nitrogen is thus already present in excess of ecosystem demand.

INPUTS OF NITROGEN TO THE ECOSYSTEMS HAVE DOUBLED IN 100 YEARS!

Nitrogen (N) is the nutrient needed in largest quantities by plants, and the rate of nitrogen supply frequently limits plant production, especially in temperate ecosystems and in agriculture worldwide. Unlike other major nutrients, nitrogen is rarely released by weathering of minerals at rates needed to match plant demand (but see Dahlgren, 1994). The larger portion of the nitrogen in the biosphere was originally bound by N_2-fixing prokaryotic organisms, but some atmospheric N_2 has also been fixed in connection with lightning. The last hundred years the Haber–Bosch process (industrial N_2 fixation), which is used in the production of fertilizers for agriculture, in particular, increased use of N_2-fixing legumes in agriculture, and fixation of N_2 into oxidized forms, NO_x, in industrial combustion processes and in combustion engines in transport vehicles, have together doubled the input of nitrogen to the ecosystems of the Earth (Galloway *et al.*, 1995; Vitousek *et al.*, 1997). The rate of global natural N_2 fixation is estimated at $90–130\ \mathrm{Tg\ N\ year^{-1}}$, which compares with anthropogenic contributions from energy production, fertilizer production and cultivation of N_2-fixing crops, of 20, 80 and $40\ \mathrm{Tg\ N\ year^{-1}}$, respectively (Galloway *et al.*, 1995). In and close to densely populated areas of North America, Central and Northern Europe, and East Asia, nitrogen is added to ecosystems in excess of plant demand, and turns from being a good thing in great demand to a problematic pollutant. This article focuses on this new role of nitrogen in forest ecosystems, which has been a major concern the last 10–15 years (Nihlgård, 1985; Aber *et al.*, 1989, 1998).

Nitrogen, mainly as ammonium from agriculture or nitrate from energy production, industry and transport is a component of the acid rain affecting these forests. Sulfur (S), the other major component of acid rain, is largely

emitted from relatively few industrial and energy producing point sources burning fossil fuels, and also from transport vehicles, and it has been technically possible to greatly reduce the emissions by using oil products with lower sulfur content, catalytic processes, etc. In Europe, the emissions of sulfur went down by 50% between 1980 and 1995. To reduce the emissions of nitrogen is more problematic. Catalytic burners remove some NO_x from car exhausts, but nitrogen as ammonia (NH_3) is emitted from the activity of millions of animals, and from nitrogen-rich decaying plant residues on arable lands. Consequently, the emissions of nitrogen dioxide (NO_2) and NH_3 over Europe have decreased by not more than 10% between 1980 and 1995. Similarly, in the USA, the implementation of environmental policies has substantially reduced sulfur emissions, but has reduced nitrogen emissions only marginally (Aber *et al.*, 1998). Efforts continue to be made to reduce nitrogen emissions further, but in addition to technical problems, these efforts are also hampered by many socio-economic constraints. In the developing world, nitrogen emissions are increasing and can be expected to continue to do so because of increasing number of cars, and increased crop and animal production. It is expected that the rapid development in East and Southeast Asia will lead to large emissions of nitrogen. In fact, Galloway *et al.* (1995) predicted that by 2020, Asia will account for more than half of the global anthropogenic N_2 fixation.

NITROGEN DOSE- ECOSYSTEM RESPONSE RELATIONS

In Europe, where the nitrogen load over forests regionally is higher than elsewhere, the nitrogen load varies from a few $kg\,ha^{-1}\,year^{-1}$ in the far north, to between $50-100\,kg\,ha^{-1}\,year^{-1}$ in the worst affected parts in Central Europe, e.g., in the Netherlands (Dise and Wright, 1995). Deposition levels are lower in North America, but may reach $25\,kg\,N\,ha^{-1}\,year^{-1}$, e.g., in the densely populated areas of northeast and eastern central USA and in downwind areas in Canada (Fenn *et al.*, 1998). This compares with, and adds to, annual natural cycle rates between soils and trees, which are commonly of the order of $20-150\,kg\,N\,ha^{-1}\,year^{-1}$.

Nitrogen additions to initially nitrogen-limited forests first make the trees and other plants grow faster. Under continuous high nitrogen load, a point is ultimately reached, at which nitrogen additions do not increase tree growth any more. At this stage the biological demand for nitrogen is fulfilled and the biota is nitrogen saturated. Definitions of ecosystem nitrogen saturation vary (Skeffington and Wilson, 1988), a common being the one above. This definition is, however, not consistent with the use of the term *saturation* in chemistry, which more strictly would refer to a situation when losses of nitrogen from the ecosystem equal nitrogen inputs. Ultimately, under conditions of high nitrogen load, also this definition will become satisfied. However, long before this stage is reached, several processes and organism groups become affected. Aber *et al.* (1989, 1998) described four stages of forest ecosystem responses to increasing nitrogen supply:

1. nitrogen limiting of;
2. initial effects of chronic nitrogen deposition;
3. nitrogen saturation;
4. forest decline;

Nitrogen-Limited Forests

This is the normal condition in most temperate forests, while, e.g., many tropical lowland rain forests are phosphorus (P) limited (Vitousek, 1984). Nitrogen limitation is demonstrated in experiments, in which additions of nitrogen increase plant growth (Figure 1). As a consequence, fertilization with nitrogen is often a profitable practice in forestry. For example, in most of Sweden a single addition of $150\,kg\,N\,ha^{-1}$ increases forest production by around 30% over a period of five years (Binkley and Högberg, 1997). Typical of a nitrogen-limited ecosystem is low nitrogen concentrations of plant tissues and soils, and consequently a high carbon (C) to nitrogen (C/N) ratio of plants and soil organic matter. Nitrogen limitation typically occurs in young ecosystems (e.g., sand dunes, young volcanic soils and soils left bare after glacial recession). It also occurs in older ecosystems, where rates of nitrogen accretion (by, e.g., N_2 fixation) are low in relation to rates of nitrogen loss. This is typical for many northern temperate forests and for tropical savannas, in which considerable losses of nitrogen in wildfires are a recurrent phenomenon. Characteristic of forests in cold climates is also that the cycling rates of nitrogen and other elements are low due to the low temperatures. Nitrogen limitation controls not only plant growth, but also many organism interactions, notably interactions between trophic levels in ecosystems. In nitrogen-limited forests, leaching losses of nitrogen are minimal, and often, organic nitrogen compounds are a large component of the leachate reaching ground and surface waters. Mineralization of nitrogen (*see* **Nitrogen Mineralization**, Volume 2) results in immediate uptake by roots, often mycorrhizal roots, or uptake by microorganisms other than mycorrhizal fungi. It is, in fact, possible that organic compounds like amino acids are an important nitrogen source for plants, as they can be taken up directly without any prior mineralization, which may be especially important in cold climates (Näsholm *et al.*, 1998). Hence, concentrations of inorganic nitrogen in the soil are kept minimal, and especially nitrification is strongly limited by a low supply of ammonium and a low soil pH. Thus, there is little risk of production

Figure 1 The effect of fertilizer nitrogen addition on the growth of Norway spruce (*Picea abies*): control plots to the left, treated plots to the right (photo by P Högberg 1993). In this experiment at Strasan, Sweden (Tamm, 1991), which comprises 56 plots, tree growth was primarily limited by nitrogen. Once this limitation was removed by additions of nitrogen, there were secondary limitations imposed by the supply of phosphorus and other elements. The NPK treatment increased tree growth several-fold although it acidi ed the soil

(through nitrification) and leaching of the most mobile nitrogen species, nitrate.

Initial Effects of Chronic Nitrogen Deposition, and Dose Responses

The concentration of nitrogen in plants often increases a little, but more importantly, the biomass of foliage increases more significantly (Figure 1), when an initially nitrogen-limited forest is subjected to increased nitrogen inputs. Thus, the tree canopy will capture more light energy and will contain more of the CO_2-fixing enzyme RuBisCo, which leads to increased photosynthesis and, thereafter, increased growth. A portion of the nitrogen deposited is sequestered into wood. Despite the low concentration of nitrogen in wood, budget calculations do suggest that it can be regarded as an important sink for nitrogen at low rates of nitrogen deposition (Johnson, 1992).

With age, and especially under high continuous nitrogen supply, however, the tree canopy closes in the end. Fully closed tree stands with a sufficiently high foliar nitrogen concentration will not respond to further nitrogen additions, and are, therefore, in a biological sense, nitrogen saturated. Before this occurs, the closing canopy reduces the amount of light received by understory species, which will become light limited and display high nitrogen concentrations (Högbom and Högberg, 1991). Under the course of nitrogen loading, the trees will not show drastically elevated nitrogen concentrations to begin with because the nitrogen they take up will be diluted by growth. However, in more slow growing species, e.g., many ground mosses, like *Sphagnum* and *Racomitrium*, the increased nitrogen load has caused considerable increases in nitrogen concentrations of tissues even in open habitats (Baddeley *et al.*, 1994). *Sphagnum* is capable of assimilating nitrate to some extent, but continuous high loading leads to loss of plant nitrate reductase activity (Woodin *et al.*, 1985; Figure 2) and in the end degradation of the moss cover, and even erosion of peatlands.

Nitrogen loading not only promotes directly the growth of the potentially more fast growing plant species, but may also affect the plant species composition indirectly through complex plant–animal interactions. One of the best described examples of such interactions is the effect of nitrogen deposition on grazing by the heather beetle (*Lochmaea suturalis*) on heather (*Calluna vulgaris*) in Dutch heathlands. In the past this insect had outbreaks at return intervals of *ca.* 20 years, but since 1980, the return interval has been less than eight years (Bobbink *et al.*, 1998). Evidently, grazing by this beetle on heather correlates positively with the concentration of nitrogen in plant shoot tissue. Grazing leads to increased light interception by understory grasses like *Deschampsia flexuosa* and *Molinia caerulea*, which may thus become dominant. In boreal forest, a single addition of ammonium nitrate at 25 kg N ha^{-1} resulted in significantly increased damage

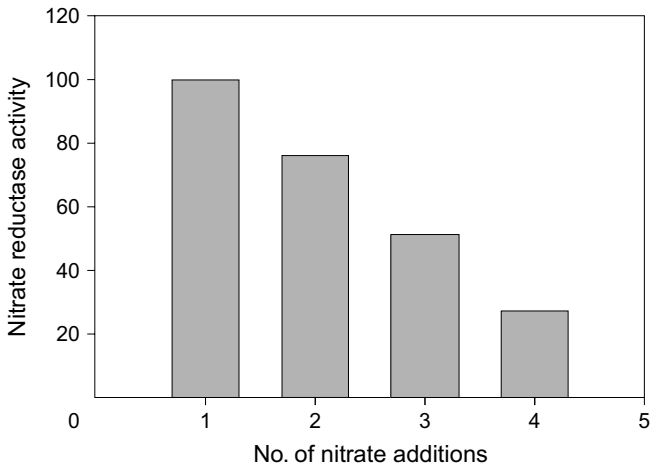


Figure 2 Effect of repeated additions of nitrate on nitrate reductase activity of the moss *Sphagnum fuscum* (calculated from data provided by Woodin *et al.*, 1985)

within two months, by two parasitic fungi and herbivory by insect larvae, on shoots of the dwarf shrub *Vaccinium myrtillus* (Nordin *et al.*, 1998). This effect of nitrogen can help to explain the increased growth of the grass *Deschampsia flexuosa*, which occurs together with *V. myrtillus* in this type of forest. Thus, deposition of nitrogen may influence the competition between plant species both directly, by increasing the growth of the more nitrogen-demanding species, and indirectly, by increasing grazing, fungal attack, frost damage, etc. on some species. In many northern temperate ecosystems this results in a loss of slow growing ericaceous dwarf shrubs and a dominance of faster growing grasses (Figure 3).

Despite the drastic local effects of air pollution, sometimes in combination with abrupt weather changes, notably in Central European mountains, where thousands of hectares of forests died in a few years in the late 1970s and early 1980s (the classical *Waldsterben* or forest decline, *see* **Waldsterben**, Volume 5), the overall trend of forest growth in Europe is increased productivity (Kauppi *et al.*, 1992; Spiecker *et al.*, 1996). Several factors could cause increased productivity, e.g., nitrogen deposition, increased concentration of atmospheric CO_2, higher temperatures (Cannell *et al.*, 1998), improved silvicultural practices, etc. This general increase in productivity, which has occurred despite increased soil acidity, and the growth responses in experimental acidifying nitrogen loading treatments, refute the idea that the acidity of the soil (expressed, e.g., as the calcium (Ca) to aluminum (Al) (Ca/Al) ratio) exerts an immediate control on tree growth (Binkley and Högberg, 1997). This is only likely under conditions of nitrogen saturation (see below). Clearly, in nitrogen-limited forests, the supply of nitrogen exerts a proximal control on tree growth.

The nitrogen input versus nitrogen output dose responses are relatively well established for northern temperate forest (Dise and Wright, 1995; cf. Figure 4; Fenn *et al.*, 1998). At deposition levels below $9\,kg\,N\,ha^{-1}\,year^{-1}$, the leaching of nitrogen is only a few $kg\,ha^{-1}\,year^{-1}$, deposited nitrogen is strongly retained by the system, and the ecosystem nitrogen capital is increasing. At intermediate levels of deposition, i.e., $9–25\,kg\,N\,ha^{-1}\,year^{-1}$, most forest ecosystems still have a tight nitrogen cycle with only

Figure 3 Grasses on the way to out-compete the dwarf shrub bilberry, *V. myrtillus*, in an area receiving high inputs of nitrogen and sulfur. Forest at Nacetin, Ore Mts., Czech Republic (photo by P Högberg, 1995)

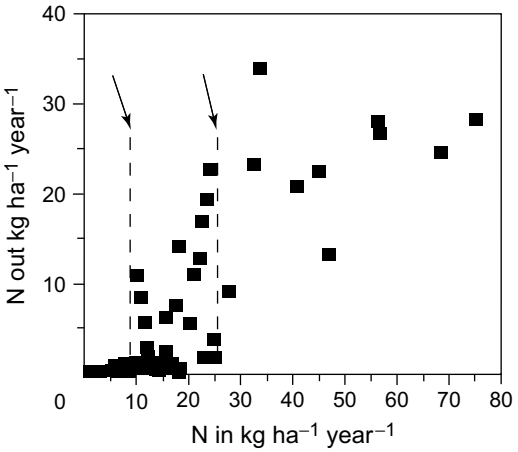


Figure 4 The relation between leaching of inorganic nitrogen (nitrogen out) and nitrogen deposition (nitrogen in) across forest ecosystems in Europe (Dise and Wright, 1995). (Reproduced by permission of Elsevier Science Ltd. from Dise and Wright, 1995)

small losses, but there are systems with considerable (>5 kg N ha^{-1} year^{-1}) nitrogen leakage. At a deposition of nitrogen above 25 kg ha^{-1} year^{-1}, virtually all forests leak significant amounts of nitrogen, especially as nitrate, and in some cases outputs rival inputs, and the systems are nitrogen saturated by any definition.

As for trees and other plants, the allocation of carbon to the root system is high under nitrogen limitation, but decreases when the nitrogen supply increases (Ericsson, 1995). The same dose response occurs for phosphorus and sulfur, but possibly not for potassium (K) and magnesium (Mg). Hence, nitrogen deposition could result in a lower biomass of active roots. There is also a mechanism of down-regulation of the root uptake capacity for, e.g., ammonium, when the supply exceeds the demand. It is thus possible that trees in nitrogen-loaded areas down-regulate both their root biomass and their specific uptake capacity for at least ammonium. However, they cannot effectively exclude all excess nitrogen by down-regulation of root uptake. Locally, there may be considerable foliar uptake of gaseous ammonia (Pearson and Stewart, 1993) from farmed animals, or of NO_x from traffic. Ammonia in high concentrations may even kill trees directly downwind of larger animal farms. Ammonia may also react with long-range transported sulfuric acid to form ammonium sulfate, leading to exceptionally high nitrogen and sulfur co-deposition locally (van Breemen *et al.*, 1982). Very high deposition is also encountered at forest edges, especially when located downwind of larger urban, industrial or agricultural areas. In conifers, excess nitrogen is often stored as the amino acid *arginine*, which thus is a good indicator of incipient nitrogen saturation (Näsholm *et al.*, 1997). As regards soil microorganisms, their capacity to retain nitrogen is strongly and directly related to the supply of available carbon. Moreover, long-term nitrogen retention is also strongly dependent on the overall presence of carbon compounds in the soil. Therefore, the soil C/N ratio is very often a useful indicator of the nitrogen retention capacity of soils (Dise *et al.*, 1998; Gundersen *et al.*, 1998). Below a C/N ratio of 15–20, the capacity of the soil to store additional nitrogen is usually marginal, and there is a greater risk of leaching of nitrate.

As northern temperate forests often are nitrogen-limited, it could be argued that, additions of nitrogen would increase the sequestration of carbon, and hence would reduce the greenhouse effect of atmospheric carbon dioxide (Schimel, 1995). It has thus been suggested that nitrogen deposition contributes to the missing sink for atmospheric carbon dioxide, which has been located in the terrestrial biomes of the north. However, most of the nitrogen deposition is on forests that are already, or will soon be, nitrogen saturated, e.g., forests in Central Europe. In theory, nitrogen fertilization of more nitrogen-limited forests could be used as a strategy to increase the carbon sink strength of forests, but this would not be a long-term solution given the trend of nitrogen eutrophication and associated problems.

Nitrogen Saturation and Forest Decline

When the combined biological demand for nitrogen by trees and other plants and microorganisms is exceeded by the nitrogen supply, the losses by leaching of nitrogen, primarily in the form of nitrate, will start to increase, and will ultimately reach the same level as the nitrogen inputs. This may cause serious nitrogen pollution of ground and surface waters, and may render them unsuitable for human consumption. Along with the nitrate anion, cations like calcium, magnesium and potassium, base cations in the first place, and subsequent inorganic aluminum species, are leached out. Some fish species are sensitive to inorganic aluminum, and soil acidification may thus cause fish death. The loss of base cations from the soil and the high nitrogen supply can cause nutrient imbalances in trees. The forest decline in the German Fichelgebirge was argued to be linked to a high N/Mg ratio in tree foliage (Schulze, 1989). Excess nitrogen in soils will also increase rates of denitrification, and hence the production of the greenhouse gas N_2O.

The ultimate stage of nitrogen saturation was described (Aber *et al.*, 1989, 1998) as forest decline. However, although excess nitrogen may certainly have been a contributing factor to forest decline in Central Europe, damaged forests occurred in areas with high concentrations also of sulfur dioxide (SO_2) and other gaseous pollutants, and it is difficult to identify the major cause of damage. When forest decline was first observed, the impact of SO_2 directly on the canopy was initially thought to be

the major cause. Later on, the effects of soil acidification, especially aluminum toxicity effects on tree roots was a major focus (Hüttermann and Ulrich, 1984). However, sensitivity to inorganic aluminum varies considerably amongst plants, and forest tree species like spruce and beech are among the more tolerant. It has, nevertheless, been argued that indicators of acidity such as soil solution Ca/Al ratios or base cation to aluminum ratios correlate negatively with tree growth (Cronan and Grigal, 1995; Sverdrup *et al.*, 1992). Critical points have been raised about the use of such ratios (e.g., Lökke *et al.*, 1996), and at present there is no consensus as regards their usefulness. Clearly, the Ca/Al ratio cannot have proximal control on tree growth when nitrogen is limiting.

How much nitrogen load can a forest then tolerate? This is probably highly variable and dependent on the availability of other nutrient elements. In areas with low background deposition levels, forests have been subjected to experimental additions of around $100\,\mathrm{kg\,N\,ha^{-1}\,year^{-1}}$ for 20 years without a significant increase in tree mortality despite serious soil acidification (Tamm, 1991; Binkley and Högberg, 1997). The treatment caused nutrient imbalances and deficiencies in the micronutrient boron (B), though. In other areas, with older and more leached soils, the forest may become negatively affected by additional nitrogen at lower rates (Aber *et al.*, 1998). The experimental evidence at hand does not point at a generally applicable lethal dose of nitrogen for trees. Many other species and ecosystem functions become affected long before serious damage is obvious in tree stands.

ARE THERE ANY REMEDIES?

As discussed below, reductions in nitrogen emissions lead to some immediate positive effects, e.g., improved quality of ground and surface waters. Reductions of emissions are, of course, also the necessary and ultimate effective long-term solution of the problem of excess nitrogen. However, the results of attempts to control nitrogen emissions are so far meagre.

Awaiting major reductions in emissions, a policy has been to reduce harmful effects of acid deposition. Nutrient imbalances in trees and soil acidification can be counteracted by so-called *vitality fertilization* or by liming. The conventional remedy against soil acidification, liming, may, however, have clear disadvantages, e.g., it may cause increased nitrification, and hence, increased leaching of nitrate along with negative effects on several groups of soil organisms (Kreutzer, 1995). Many forest ecosystems have naturally acid soils due to the presence of organic acids, and liming results in a larger perturbation of soil biota than when used in agriculture. There are cases however, where severely damaged forests have recovered after liming, e.g., in the French Ardennes (Bonneau *et al.*, 1990). In Germany, liming was tested experimentally many decades ago and before the forest decline (Waldsterben) in the 1980s; a reanalysis of those experiments showed positive results with regard to forest growth notably where dolomitic lime, containing both calcium and magnesium, was used, and on sites where a magnesium deficiency could have developed (Hüttl and Zöttl, 1993). In many cases, forest growth was not affected by liming.

Vitality fertilization aims at correcting nutrient imbalances in trees, which, for example, occur as a result of an excessive supply of nitrogen, rather than to counteract soil acidification *per se*. Commonly vitality fertilizers contain phosphorus, calcium, potassium, magnesium and some micronutrients. It is believed that such targeted additions will have less negative side effects than conventional liming.

WHAT HAPPENS IF THE NITROGEN LOAD IS SUSPENDED?

This question has been addressed most explicitly in experiments, where: (1) nitrogen-saturated forests in areas of high-nitrogen deposition have been sheltered by roof constructions (Figure 5) under which nitrogen-depleted rain water has been added (Boxman *et al.*, 1998); or where (2) continuous experimental nitrogen additions (started by C O Tamm, see bibliography) have been suspended on nitrogen-saturated plots (Quist *et al.*, 1999). These different approaches yield strikingly similar and most positive results; i.e., in both cases levels of available inorganic nitrogen decrease rapidly in the soil. In the case of roof experiments, the effect on leaching losses of inorganic nitrogen was a remarkable reduction from a rate of $50\,\mathrm{kg\,N\,ha^{-1}\,year^{-1}}$ to a few $\mathrm{kg\,N\,ha^{-1}\,year^{-1}}$ in a couple of years. The lesson from these experiments is that high concentrations of inorganic nitrogen in the soil cannot be maintained without high inputs. However, the negative effects of the previous nitrogen deposition on the base saturation of the soil may be much more lasting. Also the changes in plant community composition may last for many more years, because it is difficult for plants to invade closed vegetation communities.

CONCLUSIONS

High levels of nitrogen deposition are going to continue to be a major environmental problem, and will most likely soon be evident in areas of intense agricultural and industrial development other than densely populated areas of Europe and North America, e.g., in East and Southeast Asia. Although much of the focus in this context has been on the growth of forest trees, many other organisms are

Figure 5 The roof experiment at Klosterhede, Denmark (photo P Högberg, 1992). Throughfall deposition containing acidifying sulfur and nitrogen compounds is prevented from reaching the soil. Water with a salt composition of pre-industrial rain water is sprinkled under the roof

more sensitive, and we will continue to witness ecosystem nitrogen eutrophication resulting in shifts in species composition, along with soil acidification and aluminum and nitrate pollution of ground and surface waters. Thus, it is most important that emissions from transport, industry and agriculture be reduced substantially. This would result in an immediate reduction in the availability of inorganic nitrogen in forest soils, and hence in decreased leaching of nitrate and aluminum. As regards other ecosystem functions and species composition, the time needed to recover is not known.

See also: **Nitrogen Cycle**, Volume 2.

REFERENCES

Aber, J D, Nadelhoffer, K J, Steudler, P, and Melillo, J M (1989) Nitrogen Saturation in Northern Forest Ecosystems, *BioScience*, **39**, 378–386.

Aber, J D, McDowell, W, Nadelhoffer, K, Magill, A, Berntson, G, Kamakea, M, McNulty, S, Currie, W, Rustad, L, and Fernandez, I (1998) Nitrogen Saturation in Temperate Forest Ecosystems: Hypotheses Revisited, *BioScience*, **48**, 921–934.

Baddeley, J A, Thompson, D B A, and Lee, J A (1994) Regional and Historical Variation in the Nitrogen Content of *Racomitrium lanuginosum* in Britain in Relation to Atmospheric Nitrogen Deposition, *Environ. Pollut.*, **84**, 189–196.

Binkley, D and Högberg, P (1997) Does Atmospheric Deposition of Nitrogen Threaten Swedish Forests?, *For. Ecol. Manage.*, **92**, 119–152.

Bobbink, R, Hornung, M, and Roelofs, J G M (1998) The Effects of Air-borne Nitrogen Pollutants on Species Diversity in Natural and Semi-natural European Vegetation, *J. Ecol.*, **86**, 717–738.

Bonneau, M, Landmann, G, and Nys, C (1990) Fertilization of Declining Conifer Stands in the Vosgues and in the French Ardennes, *Water, Air, Soil Pollut.*, **54**, 577–594.

Boxman, A W, van der Ven, P J M, and Roelofs, J G M (1998) Ecosystem Recovery after a Decrease in Nitrogen Input to a Scots Pine Stand at Ysselsteyn, the Netherlands, *For. Ecol. Manage.*, **101**, 155–163.

Cannell, M G R, Thornley, J H M, Mobbs, D C, and Friend, A D (1998) UK Conifer Forests may be Growing Faster in Response to Increased N Deposition, Atmospheric CO_2 and Temperature, *Forestry*, **71**, 277–296.

Cronan, C S and Grigal, D F (1995) Use of Calcium/Aluminum Ratios as Indicators of Stress in Forest Ecosystems, *J. Environ. Qual.*, **24**, 209–226.

Dahlgren, R A (1994) Soil Acidification and Nitrogen Saturation from Weathering of Ammonium-bearing Rock, *Nature*, **368**, 838–841.

Dise, N B, Matzner, E, and Forsius, M (1998) Evaluation of Organic Horizon C : N Ratio as an Indicator of Nitrate Leaching in Conifer Forest Across Europe, *Environ. Pollut.*, **102**(S1), 453–456.

Dise, N B and Wright, R F (1995) Nitrogen Leaching in European Forests in Relation to Nitrogen Deposition, *For. Ecol. Manage.*, **71**, 153–161.

Ericsson, T (1995) Growth and Shoot : Root Ratio of Seedlings in Relation to Nutrient Availability, *Plant Soil*, **168–169**, 205–214.

Fenn, M E, Poth, M A, Aber, J D, Baron, J S, Bormann, B T, Johnson, D W, Lemley, D A, McNulty, S G, Ryan, D F, and Stottlemyer, R (1998) Nitrogen Excess in North American

Ecosystems: Predisposing Factors, Ecosystem Responses, and Management Strategies, *Ecol. Appl.*, **8**, 706–733.

Galloway, J N, Schlesinger, W H, Levy, II, H L, Michaels, A, and Schnoor, J L (1995) Nitrogen Fixation: Anthropogenic Enhancement – Environmental Response, *Global Biogeochem. Cycles*, **9**, 235–252.

Gundersen, P, Callesen, I, and de Vries, W (1998) Nitrate Leaching in Forest Ecosystems is Related to Forest Floor C/N Ratios, *Environ. Pollut.*, **102**, S1, 403–407.

Högbom, L and Högberg, P (1991) Nitrate Nutrition of *Deschampsia flexuosa* (L.) Trin. in Relation to Nitrogen Deposition in Sweden, *Oecologia*, **87**, 488–494.

Hüttermann, A and Ulrich, B (1984) Solid Phase Solution Root Interaction in Soils Subjected to Acid Deposition, *Philosophical Transactions of the Royal Society London*, **B305**, 352–368.

Hüttl, R F and Zöttl, H W (1993) Liming as a Mitigation Tool in Germany's Declining Forests – Reviewing Results from Former and Recent Trials, *For. Ecol. Manage.*, **61**, 325–338.

Johnson, D W (1992) Nitrogen Retention in Forest Soils, *J. Environ. Qual.*, **21**, 1–12.

Kauppi, P E, Mielikäinen, K, and Kuusela, K (1992) Biomass and Carbon Budget of European Forests, 1971–1990, *Science*, **256**, 70–74.

Kreutzer, K (1995) Effects of Forest Liming on Soil Processes, *Plant Soil*, **168–169**, 447–470.

Lökke, H, Bak, J, Falkengren-Grerup, U, Finlay, R D, Ilvesniemi, H, Nygaard, P H, and Starr, M (1996) Critical Loads of Acidic Deposition for Forest Soils: is the Current Approach Adequate?, *Ambio*, **25**, 510–516.

Näsholm, T, Ekblad, A, Nordin, A, Giesler, R, Högberg, M, and Högberg, P (1998) Boreal Forest Plants take up Organic Nitrogen, *Nature*, **392**, 914–916.

Näsholm, T, Nordin, A, Edfast, A-B, and Högberg, P (1997) Identification of Coniferous Forest with Incipient Nitrogen Saturation Through Analysis of Arginine and N–15 Abundance of Trees, *J. Environ. Qual.*, **26**, 302–309.

Nihlgård, B (1985) The Ammonium Hypothesis – an Additional Explanation to the Forest Dieback in Europe, *Ambio*, **14**, 2–8.

Nordin, A, Näsholm, T, and Ericson, L (1998) Effects of Simulated N Deposition on Understory Vegetation of a Boreal Coniferous Forest, *Funct. Ecol.*, **12**, 691–699.

Pearson, J and Stewart, G R (1993) The Deposition of Atmospheric Ammonia and its Effects on Plants, *New Phytol.*, **125**, 283–305.

Quist, M E, Näsholm, T, Lindeberg, J, Johannisson, C, Högbom, L, and Högberg, P (1999) Responses of a Nitrogen-Saturated Forest to a Sharp Decrease in Nitrogen Input, *J. Environ. Qual.*, **28**, 1970–1977.

Schimel, D S (1995) Terrestrial Ecosystems and the Carbon Cycle, *Global Change Biol.*, **1**, 77–91.

Schulze, E-D (1989) Air Pollution and Forest Decline in a Spruce (*Picea abies*) Forest, *Science*, **244**, 776–783.

Skeffington, R A and Wilson, E J (1988) Excess Nitrogen Deposition: Issues for Consideration, *Environ. Pollut.*, **54**, 159–184.

Spiecker, H, Mielikäinen, K, Köhl, M, and Skovsgaard, J P (1996) *Growth Trends in European Forests*, Springer Verlag, Berlin.

Sverdrup, H, Warvinge, P, and Rosén, K (1992) A Model for the Impact of Soils Solution Ca : Al Ratio, Soil Moisture and Temperature on Tree Base Cation Uptake, *Water, Air, Soil Pollut.*, **61**, 365–383.

Tamm, C O (1991) *Nitrogen in Terrestrial Ecosystems. Ecological Studies No. 81*, Springer Verlag, Berlin.

van Breemen, N, Burrough, P A, Velthorst, E J, van Dobben, H F, de Wit, T, Ridder, T B, and Reijnders, H F R (1982) Soil Acidification from Atmospheric Ammonium Sulfate in Forest Canopy Throughfall, *Nature*, **299**, 548–550.

Vitousek, P M (1984) Litterfall, Nutrient Cycling and Nutrient Limitation in Tropical Forests, *Ecology*, **65**, 285–298.

Vitousek, P M, Aber, J D, Howarth, R W, Likens, G E, Matson, P A, Schindler, D W, Schlesinger, W H, and Tilman, D G (1997) Human Alteration of the Global Nitrogen Cycle: Sources and Consequences, *Ecol. Appl.*, **7**, 737–750.

Woodin, S J, Press, M C, and Lee, J A (1985) Nitrate Reductase Activity in *Sphagnum fuscum* in Relation to Wet Deposition of Nitrate from the Atmosphere, *New Phytol.*, **99**, 381–388.

Nitrogen Fixers

Jean-Francois Soussana
INRA, Clermont-Ferrand, France

Nitrogen (N_2) is critical for the growth of plants and animals and can be assimilated only in the form of ammonia. Biological nitrogen fixation, the reduction of N_2 to ammonia, is a process unique to bacteria, which occurs both in free-living autotrophs and heterotrophs and in associative or symbiotic organisms. The most efficient systems for N_2 fixation are those where the microorganisms are associated with a plant in differentiated organs (i.e., root nodules) to benefit crop productivity. Currently, biological nitrogen fixation would account for approximately two-thirds of the total inputs of combined nitrogen to the biosphere (Stewart, 1977), that is 160 million tons nitrogen per year. Pre-industrial biological nitrogen fixation on land is difficult to estimate but has been placed in the range of 90–130 million tons nitrogen per year (Galloway et al.*, 1995).*

Nitrogen limitation to production is a strong constraint on carbon storage in the biosphere and increased production under elevated carbon dioxide (CO_2) may decrease over time as nitrogen becomes more limiting due to increased growth. In this context, the activity

of nitrogen-fixing microorganisms may offer the unique possibility to give plants access to an unlimited nitrogen source from the atmosphere. Due to the intimate relationships between the carbon and nitrogen cycles in ecosystems, biological nitrogen fixation could account for a possible, CO_2 induced, increased nitrogen demand thereby favoring the sequestration of additional carbon into terrestrial ecosystems (Gifford et al.*, 1996).*

POSSIBLE EFFECTS OF ELEVATED ATMOSPHERIC CARBON DIOXIDE (CO_2) CONCENTRATIONS ON FREE-LIVING AND ASSOCIATIVE NITROGEN FIXERS

There are approximately 100 bacterial genus which are known to contain at least one nitrogen (N_2)-fixing species (Young *et al.*, 1992). With the exception of the nitrogen-fixing photosynthetic bacteria and cyanobacteria, most of the free-living nitrogen-fixing bacteria are heterotrophs requiring habitats like soils, capable of providing a source of utilizable carbon (C) substrates necessary to meet the energy demands for N_2 fixation. In general, in most natural habitats, N_2 fixation is limited by the availability of carbohydrates, which helps explain why typical soil free-living bacteria like *Azotobacter* spp. contribute less than 0.5 kg nitrogen $ha^{-1}\,yr^{-1}$ (Evans and Burris, 1992).

Despite a wealth of information on root associated nitrogen-fixing bacteria, there are so far no convincing data on significant nitrogen transfer to root colonized plants. This not only follows from extensive studies with ^{15}N isotope and nitrogen balance techniques but also can be deduced from the observation that none of the listed bacteria emits significant amounts of ammonia during nitrogen-fixing growth. For *Azospirillum*, the plant growth-promoting effect is clearly demonstrated but is mostly attributed to the bacterial synthesis of phytohormones (Vanderleyden, 1998).

The effects of elevated CO_2 on free-living and associative nitrogen-fixing bacteria would be mediated through plants, primarily by altering rhizodeposition (the release of cells, exudates, mucilages and other compounds into the rhizosphere by plant roots). Plants grown under elevated CO_2 generally increase the allocation of photosynthates to roots which increases the capacity and/or activity of below-ground carbon sinks, enhancing in some but not all systems root turnover and deposition of organic carbon into the soil (Rogers *et al.*, 1994). However, no experimental demonstration of an increased nitrogen-fixation by free-living or associative N_2 fixers has yet been reported for plant–soil systems grown under an elevated atmospheric CO_2 concentration.

DIRECT EFFECTS OF ELEVATED CO_2 ON SYMBIOTIC N_2 FIXATION

Two types of symbioses between nitrogen-fixing bacteria and higher plants are known where the bacteria are hosted in special organs, root nodules, that are formed by the host plants upon interaction with the microsymbionts. In one of these symbioses, the microsymbionts are rhizobia that interact with plants of the legume family or with one non-legume, *Parasponia* (*Ulmaceae*). In the other one, actinomycetes of the genus *Frankia* interact with plant species from eight different families, mostly woody shrubs.

It has long been known that elevated levels of atmospheric CO_2 concentration stimulate legume growth and N_2 fixation (Wilson *et al.*, 1933). In a wide range of plants, elevated CO_2 leads to an increase in nitrogenase activity within several days. The longer term increase of total nitrogenase activity under high CO_2 would be due to an increase in the number and in the individual weight of the nodules, but the nitrogenase activity per unit nodule dry weight is not greater than in the control (Norby *et al.*, 1987; Ryle *et al.*, 1992). Nevertheless, with *Gliricidia sepium*, a fast-growing woody legume, greater rates of carbon supply to nodules under elevated CO_2 did not affect nodule mass per plant but did increase specific nitrogenase activity (Tissue *et al.*, 1997). The CO_2 enhanced plant growth may result in an increased nitrogen sink strength which is detectable within a few days. As a result, nitrogenase activity is up-regulated to meet the increased nitrogen demand, a concept supported by some physiological evidence (Oti-Boateng *et al.*, 1994).

In annual legumes, up to one-third of photosynthates may be used for energy and carbon skeletons for nitrogen fixation and assimilation. Dark CO_2 fixation by nodule PEPC (phosphoenol–pyruvate carboxylase) may provide an additional source of carbon for nodule function and maintenance. With soybean, about half of the CO_2 efflux inside the nodule is recycled through the PEPC activity. Low rhizosphere (below 100 ppm) CO_2 concentrations result in a significant decline in nitrogenase (acetylene reduction) activity of some legume nodules (Bethenod *et al.*, 1984). However, increasing the partial pressure of CO_2 in the rhizosphere above ambient does not stimulate the nitrogenase activity of soybean (O Bethenod, personal communication). Therefore, direct effects of an increased rhizosphere CO_2 partial pressure on nitrogenase activity from nodulated legume roots are unlikely.

During biological N_2 fixation, H^+ and N_2 are simultaneously reduced by nitrogenase. The proportion of total electron flow through nitrogenase that is used in N_2 fixation to ammonia (NH_3) is usually not modified significantly under elevated CO_2 (Crush, 1993). Therefore, although the evidence available at present is rather limited, it seems that neither the nitrogenase activity per unit nodule dry weight, nor the efficiency or the carbon cost of the nitrogen-fixing

process within legumes, will be directly affected by an increase in the atmospheric CO_2 concentration.

Nitrogen-fixing activity is reduced in proportion to the amount of the available soil mineral nitrogen. In the presence of mineral nitrogen, the long-term inhibition of the symbiosis involves a decline in the number and in the total weight of nodules per plant, whereas the short-term (during a few days) decline in the nitrogenase activity is associated with a decreased nodule O_2 permeability and is thus reversible. With pea, the short-term increase in the atmospheric CO_2 concentration from 320 to 1200 ppm does not prevent the inhibition of nitrogenase activity by root medium nitrate, although it increases the supply of photosynthates to the nodule (Chen and Phillips, 1977). In the same way, with white clover (Zanetti *et al.*, 1998), in the presence of a high level of mineral nitrogen, elevated pCO_2 *per se* did not alter the percentage nitrogen derived from symbiosis. As long as the mineral nitrogen supply satisfies the CO_2 induced increase in the plant's nitrogen sink strength, the relative contribution of symbiotically fixed nitrogen as opposed to mineral nitrogen, would therefore be unaffected by the CO_2 enrichment.

CAN SYMBIOTIC N_2 FIXATION COMPENSATE FOR AN INCREASED ECOSYSTEM DEMAND FOR NITROGEN IN ELEVATED CO_2?

Symbiotically fixed nitrogen can be transferred to non-fixing species by an above-ground route via herbivore excreta and by below-ground transfer via either direct excretion of nitrogen compounds from legume roots or decomposition of legume nodule, root and senescing root material.

Whether elevated CO_2 can affect the above-ground transfer of fixed nitrogen is not known, due to a lack of grazing experiments in elevated CO_2. Concerning the below-ground transfers, data from model grassland ecosystems on fertile soils support the concept of an increased below-ground transfer of fixed nitrogen in elevated CO_2. In grass–clover mixtures, the proportion of legume increased significantly under elevated CO_2 (Hebeisen *et al.*, 1997), a conclusion also reached by a number of experimental studies with temperate and rather fertile managed grasslands. It appears that the entire additional amount of nitrogen assimilated in clover under elevated CO_2 came through symbiotic N_2 fixation (Soussana and Hartwig, 1996). The below-ground transfer of fixed nitrogen was estimated in these experiments with a ^{15}N isotope technique. This transfer of fixed nitrogen contributed up to 60% of the nitrogen nutrition of the mixed grass and was usually significantly greater in elevated than in ambient CO_2 during two growing seasons (Soussana and Hartwig, 1996).

An increased N_2 demand (Soussana *et al.*, 1996), both in the legume and in the grassland ecosystem as a whole, could account for these observations of an increase in nitrogen fixation and below-ground nitrogen transfer in elevated CO_2. The extent of competition for soil-available nitrogen by associated non-legumes is one factor that determines the proportion of nitrogen derived from symbiotic N_2 fixation in mixtures. This proportion will also increase if soil nitrogen availability is reduced by any other mechanism (leaching, denitrification, low mineralization, nitrogen immobilization).

These results suggest a direct physiological link between symbiotic N_2 fixation and the nitrogen demand of the whole ecosystem. Under elevated CO_2, N_2 fixation is likely to respond positively to the increased nitrogen demand of the ecosystem and this will, presumably, buffer the imbalance between the carbon and nitrogen cycles.

POSSIBLE LONG-TERM ELEVATED CO_2 EFFECTS ON THE POPULATION DYNAMICS OF LEGUMES

From the model by Schwinning and Parsons (1996), any CO_2 induced advantage of legumes in mixed communities, by increasing the soil nitrogen availability, will stimulate in the longer term the growth of non-fixing species. It would therefore be clearly over-simplistic to assume that the rise in atmospheric CO_2 concentration will lead to a steady increase in the legume content of the vegetation.

Moreover, for nitrogen fixation to track ecosystem carbon fixation in the long-term, more phosphorus must enter the biological cycle. Some experimental evidence shows that plants in elevated CO_2 may have the capacity to mobilize more phosphorus from so-called unavailable sources using mechanisms involving exudation of organic acids and phosphatases (Gifford, 1996). The interactive effects of elevated CO_2, supply of phosphorus and legume presence were studied in model communities of calcareous grassland by Stöcklin and Körner (1999). Legumes were significantly stimulated by P supply and their CO_2 response was much larger when P was added. Also, in another study with intact monoliths from calcareous grassland, the difference in biomass production between elevated and ambient CO_2 was higher when P was added, especially in legumes (Stöcklin *et al.*, 1998). Therefore, nitrogen-fixing legumes appear to be crucial for the community response, but legume responsiveness is largely controlled by P availability.

CAN A LARGER (FIXED) NITROGEN INPUT RESULT IN THE SEQUESTRATION OF ADDITIONAL CARBON UNDER ELEVATED CO_2?

The decline in leaf nitrogen concentration under elevated CO_2, which affects litter decomposition processes, is likely

to be restricted mostly to the non-nitrogen-fixing species (Hartwig *et al.*, 2000). Therefore, in elevated CO_2, the decomposition of the legume material could supply the soil and the ecosystem with a large nitrogen input. This result can be expected as long as the nitrogen-fixing symbiosis is not hampered by other factors, like low temperature, low phosphorus or water stress.

With monoliths of an established ryegrass/white clover mixture, net primary productivity below-ground and soil CO_2–C production were enhanced, resulting in a more rapid turnover of carbon at elevated CO_2. However, no detectable change in carbon pool sizes was observed in elevated CO_2 (Ross *et al.*, 1995). Therefore, with some soils, the positive effects of the nitrogen-rich legume material on the decomposition process might be dominant and lead to increased soil respiration rates, with no supplemental below-ground carbon storage. Nevertheless, in some other soils (acidic, calcic, or with large contents of clay) the high degree of protection of the soil organic matter (SOM) might favor an increased below-ground carbon sequestration under elevated CO_2.

A model of the interacting global carbon and nitrogen cycles (CQUESTN) was developed by Gifford *et al.* (1996). With increasing atmospheric CO_2 concentrations stimulating photosynthesis, and inputs of litter with increased carbon–nitrogen ratio, the model assumes that the increased chemical energy supply will enable the ecosystem to gradually acquire nitrogen either by biological nitrogen fixation or from the nitrogen that would otherwise flow to atmosphere or rivers. In CQUESTN, this is modeled as a tendency for ecosystems to shift the carbon–nitrogen ratio of humus back to an equilibrium value. This model also assumes that nitrogen acquisition from the atmosphere responds to demand, with its suppression by a high soil–nitrogen status and its stimulation by a low soil nitrogen status. With the nitrogen fixation response turned off, a large increase in the carbon–nitrogen ratios of the litter was predicted, as well as a slow increase in the carbon–nitrogen ratio of the total SOM. By contrast, when it was assumed that 5% of the nitrogen deficit in SOM below the pre-industrial value was recovered per year by fixation, the global carbon sequestration increased from 1.8–2.6 Gt C yr^{-1}. While this model is heuristic not predictive, it shows the major role that the nitrogen fixation process can play for the global carbon sequestration in terrestrial ecosystems.

REFERENCES

Bethenod, O, Prioul, J L, and Deroche, M E (1984) Short-term Inhibition of Acetylene Reduction Activity by Low CO_2 Concentrations Around the Nodulated Roots of Various Legumes, *Physiol. Vég.*, **22**, 565–570.

Chen, P C and Philips, D A (1977) Induction of Root Nodule Senescence by Combined Nitrogen in *Pisum sativum*, *Plant Physiol.*, **59**, 440–442.

Crush, J R (1993) Hydrogen Evolution from Root Nodules of *Trifolium repens* and *Medicago sativa* Plants Grown under Elevated Atmospheric CO_2, *N. Z. J. Agr. Res.*, **36**, 177–183.

Evans, H J and Burris, R H (1992) Highlights in Biological Nitrogen Fixation During the Last 50 years, in *Biological Nitrogen Fixation*, eds G Stacey, R H Burris, and H J Evans, Chapman and Hall, New York, 1–42.

Galloway, J N, Schlesinger, W H, Levy, II, H, Michaels, A, and Spencer, J L (1995) Nitrogen Fixation: Anthropogenic Enhancement–Environmental Response, *Global Biochem. Cycles*, **9**, 235–252.

Gifford, R M, Lutze, J L, and Barrett, D (1996) Global Atmospheric Change Effects on Terrestrial Carbon Sequestration: Exploration with a Global C and Nitrogen Cycle model, *Plant Soil*, **187**, 369–387.

Hartwig, U A, Lüscher, A, Daepp, M, Blum, H, Soussana, J F, and Nösberger, J (2000) Due to Symbiotic N_2 Fixation, Elevated Atmospheric pCO_2 May Have No Effect on Litter Nitrogen Concentration in a Grassland Ecosystem, *Plant Soil*, **224**, 43–50.

Hebeisen, T, Lüscher, A, Zanetti, S, Fischer, B U, Hartwig, U A, Frehner, M, Hendrey, G R, Blum, H, and Nösberger, J (1997) Growth Response of *Trifolium repens* L. and *Lolium perenne* L. as Monocultures and Bi-species Mixture to Free Air CO_2 Enrichment and Management, *Global Change Biol.*, **3**, 149–160.

Norby, R J (1987) Nodulation and Nitrogenase Activity in Nitrogen Fixing Woody Plants Stimulated by CO_2 Enrichment of the Atmosphere, *Physiol. Plant.*, **71**, 77–82.

Oti-Boateng, C, Wallace, W, and Silsbury, J H (1994) The Effect of the Accumulation of Carbohydrate and Organic Nitrogen on Nitrogen Fixation (Acetylene Reduction) of Faba Bean cv, Fiord, *Ann. Bot.*, **73**, 143–149.

Rogers, H H, Runion, G B, and Krupa, S V (1994) Plant Responses to Atmospheric CO_2 Enrichment with Emphasis on Roots and the Rhizosphere, *Environ. Pollut.*, **83**, 155–189.

Ross, D J, Tate, K R, and Newton, P C D (1995) Elevated CO_2 and Temperature Effects on Soil Carbon and Nitrogen Cycling in Ryegrass/white Clover Curves of an Endoquaept Soil, *Plant Soil*, **176**, 37–49.

Ryle, G J A, Powell, C E, and Davidson, I A (1992) Growth of White Clover, Dependent on N_2 Fixation, in Elevated CO_2 and Temperature, *Ann. Bot.*, **70**, 221–228.

Schwinning, S and Parsons, A J (1996) Analysis of the Coexistence Mechanisms for Grasses and Legumes in Grazing Systems, *J. Ecol.*, **84**, 799–813.

Soussana, J F, Casella, E, and Loiseau, P (1996) Long-term Effects of CO_2 Enrichment and Temperature Increase on a Temperate Grass Sward. 2. Plant Nitrogen Budgets and Root Fraction, *Plant Soil*, **182**, 101–114.

Soussana, J F and Hartwig, U (1996) The Effects of Elevated CO_2 on Symbiotic N_2 Fixation. A Link Between the C and Nitrogen Cycles in Grassland Ecosystems, *Plant Soil*, **187**, 321–332.

Stewart, W P D (1977) Present Day Nitrogen Fixing Plants, *Ambio*, **6**, 167–173.

Stöcklin, J and Körner, C (1999) Interactive Effects of Elevated CO_2, P Availability and Legume Presence on Calcareous Grassland: Results of a Glasshouse Experiment, *Funct. Ecol.*, **13**, 200–209.

Stöcklin, J, Schweizer, K, and Körner, C (1998) Effects of Elevated CO_2 and Phosphorus Addition on Productivity and Community Composition of Intact Monoliths from Calcareous Grassland, *Oecologia*, **116**, 50–56.

Tissue, D T, Megonigal, J P, and Thomas, R B (1997) Nitrogenase Activity and N_2 Fixation are Stimulated by Elevated CO_2 in a Tropical N_2 Fixing Tree, *Oecologia*, **109**, 28–33.

Vanderleyden, J (1998) Root Associated Nitrogen Fixing Bacteria in Retrospective and Perspective, in *Biological Nitrogen Fixation for the 21st Century*, ed C Emelrich, Kluwer, Dordrecht, 373–374.

Wilson, P W, Fred, E B, and Salmon, M R (1933) Relation Between Carbon Dioxide and Elemental Nitrogen Assimilation in Leguminous Plants, *Soil Sci.*, **35**, 145–165.

Young, J P W (1992) Phylogenetic Classification of Nitrogen Fixing Organisms, in *Biological Nitrogen Fixation*, eds G Stacey, R H Burris, and H J Evans, Chapman and Hall, New York, 43–86.

Zanetti, S, Hartwig, U A, and Nösberger, J (1998) Elevated Atmospheric CO_2 Does Not Affect Per Se the Preference for Symbiotic Nitrogen as Opposed to Mineral Nitrogen of Trifolium repens L., *Plant, Cell Environ.*, **21**, 623–630.

Nitrogen Mineralization

Nitrogen mineralization is an aspect of soil mineralization (*see* **Soil Mineralization**, Volume 2) and is the conversion of organic nitrogen in the soil to the inorganic state. Nitrogen is probably the most important plant nutrient. It exists in soil in a number of mineralized forms (molecular nitrogen, ammonium, nitrite and nitrate) and each of these may be absorbed by either plants or microorganisms for use in the formation of amino acids, proteins and other organic nitrogenous compounds. In addition, many microorganisms are able to extract energy by oxidizing ammonium nitrogen to nitrite, and others by oxidizing nitrite to nitrate. This process is called *nitrification*. Still other microorganisms convert nitrate or nitrite to molecular nitrogen by the process of *denitrification*. Some denitrifying bacteria also convert nitrate to ammonia.

Higher plants are unable to absorb molecular nitrogen and most are not able to absorb organic nitrogen compounds to any significant degree. Therefore they must rely on added nitrogen fertilizer, or nitrogen fixation, or the mineralization of organic nitrogen that enters the soil as organic detritus. However, the presence of excess organic material in soils leads to very high rates of microbial growth and this may demineralize available nitrogen, even to the extent of using up a large proportion of agriculturally applied fertilizer. On the other hand, if excess mineralization occurs, there is a danger that the mineralized nitrogen will be permanently lost from the ecosystem by leaching (*see* **Nitrate Leaching**, Volume 3). It is thus very important, particularly in agricultural systems, that the soil nitrogen cycle of mineralization and demineralization be maintained in balance.

R G S BIDWELL *Canada*

Non-equilibrium Ecology

James F Reynolds
Duke University, Durham, NC, USA

One of the most controversial and well-known theories associated with global change is the Gaia hypothesis. Originated by biologist James Lovelock, this hypothesis states that over geologic time, biotic life on Earth has evolved to control the physical forces of nature, resulting in a bio-geosphere that is self-regulatory, or homeostatic. Thus, acting like the thermostat in your house, any deviation from this balanced condition or set-point will result in a corrective response in order to maintain the stable equilibrium. Do ecosystems behave this way? If so, predicting their response to global change would be relatively easy. However, notions of stable equilibria – as espoused by the Gaia hypothesis – are centered around the concept of the balance of nature, which has dictated western views of science since antiquity. It is ironic that the Gaia hypothesis is receiving such popular attention at a time when ecologists are eschewing the equilibrium-centered view per se in favor of a non-equilibrium one. During the past several decades, ecology has experienced a paradigm shift: from the balance-of-nature (equilibrium/stability) to non-equilibrium. This paradigm shift, which embraces spatiotemporal dynamics of ecosystems, is essential to our ability to develop a predictive understanding of the dynamics of ecosystems.

Many pioneer naturalists adhered to the belief that nature reflected a balance of destructive and constructive forces and would therefore maintain or achieve some degree of permanent condition or equilibrium if left undisturbed (Wu and Loucks, 1995). This *balance-of-nature paradigm* long dictated Western views of science and has greatly influenced our interpretation of natural phenomena. Frederic Clements (1874–1945) was particularly influential in shaping this attitude. Clements proposed that *natural succession*,

which is the directional change in community structure over time, always ends in a single type of self-perpetuating community within a given climatic region (a mono-climax). For example, if we were to observe an abandoned tobacco field in the Piedmont of North Carolina, USA over many years, we would initially see it colonized by various grasses and herbaceous vegetation, followed in time by woody plants, shrubs and small trees, and eventually by a forest of mixed hardwood species (the climax vegetation). The equilibrium view of Clementsian succession is that this balance of species over time is achieved via the mechanisms of tight biotic interactions and feedbacks (mainly competition and density dependent forces) within the constraints of a relatively constant environment. Once achieved, this predetermined climax community remains more or less stable through time unless disturbed by external forces. Even in the face of disturbance, the system persists, either remaining at its equilibrium (measured as resistance) or eventually returning to it (measured as resilience or return time). While Clements recognized that disturbance was omnipresent and that "... even the most stable association is never in complete equilibrium ...," he considered disturbance as simply a mechanism that reset the system on its inexorable march towards equilibrium (Pickett and White, 1985). This deterministic thinking dominated ecology for years and led to the development of numerous concepts related to the development and maintenance of stable equilibria in populations, communities, and ecosystems.

Today, however, the balance-of-nature paradigm has been supplanted with what might be more aptly termed the dynamics of nature. This viewpoint, commonly labeled *non-equilibrium ecology*, sees nature as being uncertain and constantly changing. Whereas equilibrium theory is only concerned with the behavior of a system at or near its equilibrium point, non-equilibrium ecology significantly broadens this view. Non-equilibrium ecology embraces transient dynamics, including behaviors like chaos, thresholds, multiple equilibria, and limit cycles. These dynamics are often in phase with naturally occurring disturbances, e.g., fires, floods, drought, windstorms, and insect outbreaks, which can have profound effects on ecosystems. They also result from the ever-changing local rules of species interactions, which vary over time and space during ecosystems development (Levin, 1998). Certain opportunistic species thrive while others disappear, often in ways that are hard to predict. The non-equilibrium view of succession sees historical effects, spatial heterogeneity, stochastic factors, and variability as key agents that affect the rates and directions of ecological dynamics.

Famous historical debates originated from the equilibrium and non-equilibrium points of view of nature, including the role of density-dependent versus density-independent population regulation, how competition structures ecological communities, and theories on ecosystem resistance, resilience, and stability (Wu and Loucks, 1995). In the context of current concerns about global change, a more germane question is, "Can we understand and predict the behavior of non-equilibrium ecosystems?" This question raises a plethora of issues that challenge our basic understanding of ecosystem structure and function. One key issue is the overwhelming importance of time and spatial scales in understanding and interpreting ecosystem dynamics. Given that the probability of observing an extinction in a population increases directly with temporal scale and inversely with spatial scale (Berryman, 1987; DeAngelis and Waterhouse, 1987), the following examples illustrate the importance of spatiotemporal scales in governing our perception of ecosystem equilibria and dynamical states.

At the temporal scale of the last 100 years, it has been a popular notion that aspen forests in Yellowstone National Park (Wyoming, USA) are in equilibrium. It's easy to see how these forests could be considered so since they appear to have always been there. It has also been noticed in recent years that aspen trees are failing to regenerate, something that has widely been attributed to the direct effects of fire suppression and/or excessive elk numbers. However, following the extensive Yellowstone fires in 1988, Romme *et al.* (1995) found that regeneration remained low. In sharp contrast to the equilibrium model, they concluded that aspen forests are actually temporary features of the Yellowstone landscape and are destined to disappear. In fact, the very existence of the present-day aspen stands is a result of circumstances that occurred during the period 1870–1890, where the numbers of elk, beaver and moose were low, climate was moist, extensive fires had recently occurred, and wolves were present – a unique and episodic set of conditions that favored aspen establishment. Similarly, the importance of spatial scale in interpreting ecosystem dynamics is illustrated in a study by DeAngelis and Waterhouse (1987), who describe the maintenance of rare species in areas where habitat is shrinking due to human encroachment. Using computer simulation, they partitioned a formerly contiguous forest into quasi-isolated patches resembling a checkerboard, where no forest patch could hold all of the species of plants, birds, mammals, arthropods, etc., that the original forest contained. The population dynamics of the small patches, unlike the original forest, were highly transient and often included species extinction (non-equilibrium properties). However, they showed that if the rates of plant migration and propagule transport between the forest patches were large enough, a suitably large and closely spaced cluster of forest patches would have characteristics of a stable equilibrium population. Hence, stable equilibrium is a attribute that can emerge asymptotically by extrapolating to sufficiently large spatial scales. These examples show that ecosystems may display elements of both equilibrium and non-equilibrium characteristics depending on the spatial and temporal scale in which

they are viewed. The non-equilibrium paradigm covers this spectrum of conditions, where at one end are stable equilibria dynamics structured primarily by biotic forces and few stochastic effects (i.e., self-regulating population growth, compensatory interactions, intense competition) and at the other are non-equilibria dynamics, characterized by weak biotic couplings, independence of species, abiotic controls, and large stochastic effects.

To illustrate the interplay between equilibrium and non-equilibrium factors in controlling how ecosystems may respond to forces of global change, I use an example of changing rainfall variability in semiarid rangelands. I select rangeland ecosystems because they are ecologically important, covering about 2/5 of the land surface of the globe and home to 20% of the human population, and because there has been much controversy regarding their behavior as either equilibrium or non-equilibrium systems. This debate has been influential in shaping scientific discourse, research approaches, and management priorities for dealing with land degradation, over-grazing by domestic herbivores, climate change, and famine (in pastoral regions of Africa) (Illius and O'Connor, 1999). Arid and semiarid rangelands are particularly vulnerable to climate variability, of which precipitation is the most important component. A slight shift in seasonal precipitation and/or frequency of extreme rainfall could potentially lead to major ecological and biogeochemical impacts, but the impacts may be influenced by local grazing history and management, current stocking rates, soil types, species composition, etc. In Figure 1, a hypothetical rangeland site is depicted in terms

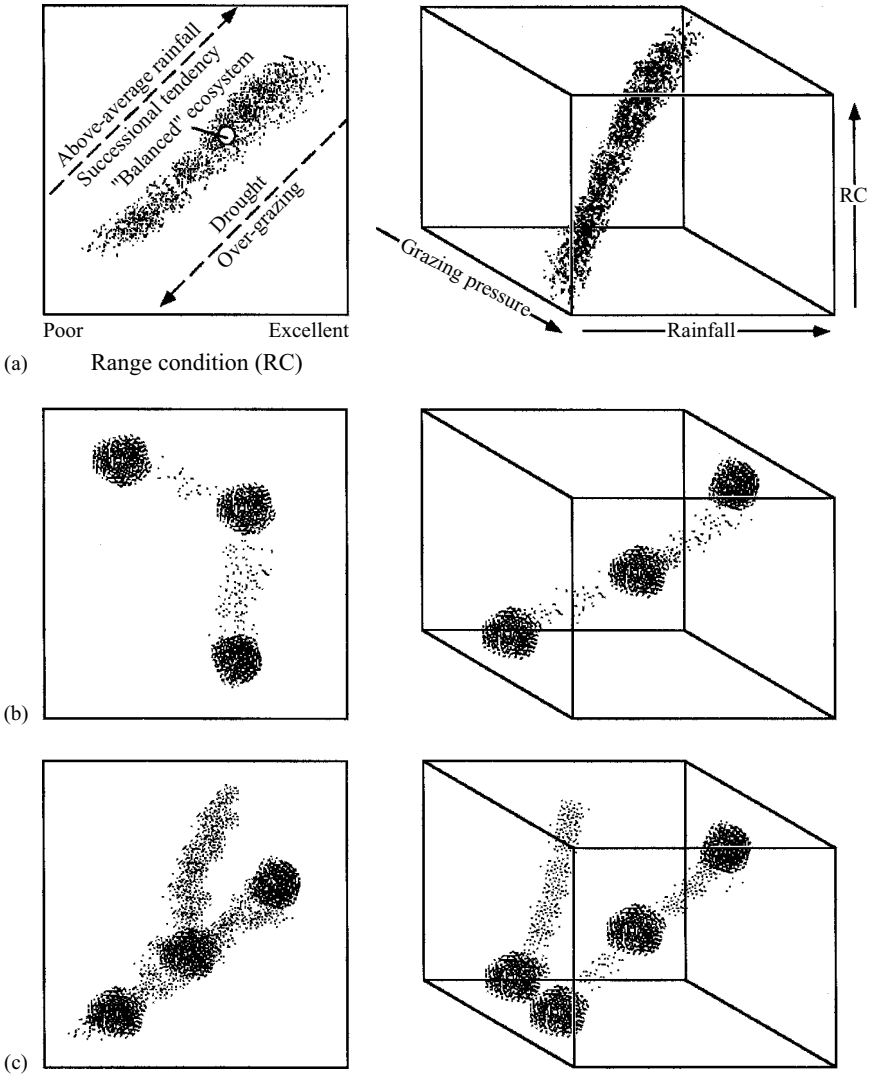


Figure 1 Probability distributions of a hypothetical rangeland site shown in two-(left) and three-(right) dimensional descriptive space. (a) Equilibrium Clementsian model. (b- c) Non-equilibrium model. (Based on Westoby *et al.*, 1989 and Stafford Smith and Pickup, 1993)

of how it might vary over thousands of years (based on Stafford Smith and Pickup, 1993). Over this long period of time, many site characteristics would change, including species composition, soil types, plant biomass, herbivore population densities, soil arthropods diversity, soil nutrient status, etc. If, for each year, we plotted the location of this site within the n-dimensional space defined by these site characteristics, a plot would emerge where the density of points at a particular location would be proportional to the probability that this site occupied that location. This is shown in Figure 1 but simplified for only two or three site characteristics. In Figure 1(a), we illustrate the range condition model, where different states of vegetation are portrayed along a linear gradient (following Westoby *et al.*, 1989). For over 50 years, this equilibrium model has been the dominant model for assessing and managing these systems. It assumes that rangelands ecosystems have a single persistent state: the Clementsian endpoint (equivalent to excellent range condition). Practices such as over-grazing drive the ecosystem in the opposite direction of its natural march towards equilibrium. As herbivory increases, perennial grasses decline and are replaced by unpalatable forbs and other weedy species (equivalent to poor range condition); a reduction or removal of the grazing pressure results in a recovery of the system towards its climax. Drought is considered to affect vegetation in a similar way to grazing whereas above-average years of rainfall accelerate the succession tendency. Consequently, typical development policies and intervention strategies in rangelands based on the range condition model consider over-stocking and over-grazing in pastoral rangelands in Africa as inevitable outcomes of practices by local pastoralists (Ellis and Swift, 1988) and, in principle, a land manager can employ a rather obvious strategy to achieve a sustainable yield of livestock products: select a stocking rate that establishes a long-term balanced ecosystem (Figure 1a) between the pressure of grazing and the natural succession tendency of the system (Westoby *et al.*, 1989). Presumably, shifts in rainfall variability would simply move the system along the linear continuum: upwards in wet years and downwards in drought years. In recent years the equilibrium model has been heavily criticized because, for most rangelands, it simply fails to explain actual dynamics. For example, removing a disturbance such as grazing does not result in an orderly resumption of succession towards some persistent endpoint nor is the cumulative effects of rainfall on the ecosystem linear. The equilibrium-based range condition model fails to capture the dynamic of nature.

As alternatives to the Clementsian model, various closely related non-equilibrium models have been proposed. Based on observations made in rangelands of Australia, Westoby *et al.* (1989) proposed state and transition models. Vegetation states at a particular rangeland site may change over time depending on both exogenous and endogenous causal factors, e.g., the consequences of initial conditions on subsequent patterns of development, how plants modify the soil environment, spatial heterogeneity in soil characteristics, temporal and spatial variability in weather, and opportunistic growth by different plant species. Transitions between those states are triggered by natural events (e.g., fires, drought, large rainfall events) or by management (burning, changes in stocking rates). The result is an assorted collection of vegetation states in which the adjustment or removal of a single factor, e.g., rainfall or grazing, does not necessarily result in a shift along a linear continuum, e.g., from a degraded to less degraded condition, as in the equilibrium model. For our hypothetical rangeland site, such dynamics are depicted in Figure 1(b), which shows the site oscillating between three different domains of attraction or states and where the transitions between these states are relatively rapid (i.e., the density of points between the clusters is relatively light). An even more complex possibility is shown in Figure 1(c). In this scheme, alternative succession pathways are shown for the site, each with different numbers of probable states. Note that based on just two descriptors they appear to intersect but when the third descriptor is added, separate paths are evident, which could represent a case where erosion processes result in a permanent modification of the site.

Does the non-equilibrium model suggest that it will be impossible to predict how ecosystems may respond to global change? Not necessarily. As Illius and O'Connor (1999) point out, previous attempts to draw a distinction between equilibrium and non-equilibrium behavior *per se* in terms of identifying the underlying mechanisms that control ecosystem function have not been very successful. Rather, the valuable lesson that emerges from this is that we must develop a robust framework for coupling pattern with process at different scales. We must develop an understanding of how ecosystems function both far from, and close to, equilibrium conditions, including the importance of initial conditions, what processes lead to steady-states (stable equilibria), and which ones lead to thresholds or breakpoints (unstable equilibria). Such understanding will only come from studies and models that explicitly focus on the spatiotemporal dynamics of ecosystems and embrace the dynamic of nature paradigm as a conspicuous and integral part of normal ecosystem function.

See also: **Chaos and Predictability**, Volume 1; **Chaos and Cycles**, Volume 2; **Non-linear Systems**, Volume 2.

ACKNOWLEDGMENTS

The author acknowledges support from the Humboldt Stiftung, CIS-HDGC (SBR-9521914) and Jornada LTER (DEB 94-11971) and comments by J. Wu.

REFERENCES

Berryman, A A (1987) Equilibrium or Non-equilibrium: Is that the Question? *Bull. Ecol. Soc. Am.*, **68**, 500–502.

DeAngelis, D L and Waterhouse, J C (1987) Equilibrium and Non-equilibrium Concepts in Ecological Models, *Ecol. Monog.*, **57**, 1–21.

Ellis, J E and Swift, D M (1988) Stability of African Pastoral Ecosystems: Alternative Paradigms and Implications for Development, *J. Range Manage.*, **41**, 450–459.

Illius, A W and O'Connor, T G (1999) On the Relevance of Non-equilibrium Concepts to Arid and Semiarid Grazing Systems, *Ecol. Appl.*, **9**, 798–813.

Levin, S A (1998) Ecosystems and the Biosphere as Complex Adaptive Systems, *Ecosystems*, **1**, 431–436.

Pickett, S T A and White, P S, eds (1985) *The Ecology of Natural Disturbance and Patch Dynamics*, Academic Press, San Diego, CA.

Romme, W H, Turner, M G, Wallace, L L, and Walker, J S (1995) Aspen, Elk, and Fire in Northern Yellowstone Park, *Ecology*, **76**, 2097–2106.

Stafford Smith, D M and Pickup, G (1993) Out of Africa, Looking In: Understanding Vegetation Change, in *Range Ecology at Disequilibrium: New Models of Natural Variability and Pastoral Adaptation in African Savannas*, eds R H Behnke, I Scoones, and C Kerven, Overseas Development Institute, London, 196–244.

Westoby, M, Walker, B, and Noy-Meir, I (1989) Range Management on the Basis of a Model which does not Seek to Establish Equilibrium, *J. Arid Environ.*, **17**, 235–239.

Wu, J G and Loucks, O L (1995) From Balance of Nature to Hierarchical Patch Dynamics: a Paradigm Shift in Ecology, *Q. Rev. Biol.*, **70**, 439–466.

Non-linear Systems

David W Hilbert

Tropical Forest Research Center, Atherton, Australia

Not only in research, but also in the everyday world of politics and economics, we would be better off if more people realized that simple non-linear systems do not necessarily possess simple dynamical properties.

Sir Robert May (1976)

In a world of rapid global change, we would also be better off if more people realized that ecosystems are non-linear systems that can display surprising and unexpected changes in response to continuous and gradual changes in the environment. Will the storage of carbon by ecosystems change in direct proportion to changes in atmospheric carbon dioxide (CO_2), temperature and precipitation, or will there be discontinuous shifts where areas that now take up carbon suddenly release it to the atmosphere? Does gradual, continuous clearing of forests incrementally reduce biodiversity or are there thresholds beyond which extinction of species is much more rapid? Will the population dynamics of insect pests change qualitatively as global temperatures rise, with species that now have unpredictable outbreaks becoming common in all years and species that are now always rare entering into cycles of high and low abundance? The answers to these questions are largely unknown but it is known that non-linear systems display these kinds of thresholds and qualitative changes in behavior so an understanding of the dynamics of non-linear systems is an essential part of understanding how the biosphere will be modified by global change.

When the interactions among a system's components are not simply proportional, the system is said to be non-linear. This is the rule rather than the exception in biological systems and is a pervasive feature of most systems, whether physical or biological. Linear systems respond smoothly to gradual changes in their environment and have simple dynamics that lead to simple, constant states when the environment is unchanging. Non-linear systems, on the other hand, display many qualitatively different behaviors including the possibility of several different constant states, regular cycles, complex cycles, and seemingly random chaos. Non-linearity produces these complex behaviors in all but the simplest systems, but its greatest significance is that it can produce surprises in the form of sudden, discontinuous, and qualitative changes in a system's behavior in response to gradual changes in the external environment. So, while global changes in the environment may be incremental, the responses of ecosystems to these changes may not be gradual and continuous.

The word *system* refers to a collection of things that interact strongly with each other. For example, an ecosystem can be thought of as a set of components – soil, plants, animals, etc. – that affect one another and, in so doing, produce changes in the amounts of each component. For example, when grazing animals eat leaves, they decrease the amount of plants while increasing the amount of animal matter with perhaps some indirect affects on the soil. Of course, everything can interact and the entire Earth is a system. But many interactions are weak or indirect, so – to keep things manageable – one can define a system as just the set of components that interact strongly and are important for some specific interest. The study of the ways systems change through time often uses mathematical models where each component is represented by one or more variables (called *state variables*) that quantifies its essential properties. The values of the whole set of these variables describe the properties or "state" of the system at any particular moment. The real system is modeled by a set of equations that describe how each of

the system's components affects both itself and the other components. Of course, no system is isolated from the rest of the world so models of systems also need to include variables that represent the larger environment. These variables affect the system but are not strongly affected by the system itself. Rainfall, for instance, affects the growth of plants but individual plants don't have much direct effect on when or how much it rains. This concept of an interactive system imbedded in a larger environment is illustrated in Figure 1.

The behavior of any system depends on the responses of its components to the environment and to the interactions and feedbacks among the state variables. So, to understand and predict the impacts of global change on ecosystems, we need to know how their components respond to external environmental variables such as atmospheric CO_2 concentration, temperature and precipitation and the way these responses affect the interactions within the system. While we have less than complete knowledge of these factors, we do know that the interactions among variables in most real systems are non-linear. The defining feature of non-linear systems is that changes in a variable create non-linear changes in the rate of change of that variable and/or other variables in the system. This non-linearity has deep implications for the study of global change impacts on biological systems. Smooth, continuous changes in temperature, precipitation, atmospheric concentrations of CO_2, and other "external" variables, may have discontinuous impacts on the qualitative behavior of ecological systems. Since ecological systems are non-linear and their responses to the external environment are non-linear, we could even say that surprises in the sense of discontinuous changes from one dynamic equilibrium to another or from one form of behavior to another must be expected. Luckily, despite the bewildering variety of systems and models of systems, there is only a limited range of possible behaviors and these have been well described by dynamical systems theory. One focus of this theory is the study of how the dynamics of a system can change *qualitatively* as a parameter (including external variables like climate) changes smoothly. In other words, the field of non-linear systems dynamics is primarily about how surprises occur in the behavior of systems (see Ekeland, 1988).

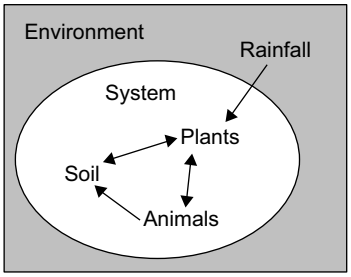


Figure 1 A schematic representation of a simple system and its environment. The system is de ned by three interacting variables (plants, animals, and soils) with their interactions indicated by arrows. These variables are separated conceptually from environmental variables, rainfall in this example, that affect the system in some way but are not themselves directly affected by the system. If any of the interaction among the system's variables is non-linear, then it is a non-linear system

The study of change in systems often uses the concept of *state space* developed by the French mathematician Henri Poincaré (1854–1912). A list of the values of all the variables that define the system at some moment in time describes the "state" of the system at that instant. For example, the population density of foxes and hares might describe the state of a predator–prey system. The state space for this system is the universe of all possible combinations of the system variables – in this case, a plane with the two axes representing the number of foxes and hares. So the state of the system at any particular time can be represented by its location in the state space; and the dynamics of the system is thought of as the trajectory of this point through the state space. For every point in the state space, there is a single vector defined by the direction and rates of change of each of the system variables. Often, there are especially important points in the state space where the rates of change of all the systems variables are zero. These are called *critical points*. If all the system's vectors near such a point converge toward it, the point is a stable critical point or equilibrium because, after any small perturbation, the system will automatically move back to that location. This is the state of the system we would expect to see most frequently. In this sense, a stable point is an *attractor* for the system since after starting the system at one of many places in the state space, it will be attracted to the stable critical point (*see also* **Chaos and Cycles**, Volume 2). Non-linear systems can also have more complex attractors such as closed curves (called *stable limit cycles*) where the system displays regular cycles that will be returned to after any perturbation. An analogous attractor in three dimensions is a torus (like the surface of a doughnut), which can produce very complex cycles called *quasi-periodic behavior*. The system follows a trajectory that continually spirals around the surface of the torus.

Box 1 Non-linearity

In non-linear systems, the rate of change of one or more of the state variables are non-linear functions of one or more of the state variables. For example, if a system has two state variables (x and y) and the rate of change of 'x' (dx/dt) is proportional to both x and y ($dx/dt = ax - by$) and the same is true for the rate of change of y ($dy/dt = cy - dx$), then the system is linear since the responses of the system are simply proportional to the values of the state variables. On the other hand, all more complex functions are non-linear, for example $dx/dt = ax - bxy$ or $dx/dt = ax^2 - b/y$.

Finally, non-linear systems may be chaotic – seemingly random but nonetheless having deterministic dynamics – in which case the attractor is called *strange* since it has a fractal, non-integer dimension (see Gleick, 1987 for an introduction to chaos). Logan and Allen (1992) provide an excellent overview of chaos in the context of population dynamics.

Lorenz's (1963) famous butterfly is the strange attractor of a model describing convection. Figure 2 shows part of the system's trajectory through state space for Lorenz's model. As the system continues to evolve through time, the trajectory will trace out more and more of the attractor, which looks somewhat like a butterfly's wings. Looking at the values of any one of the system's variables through time (see inset) the dynamics appears random. But, unlike truly random behavior, there is a very distinct, although strangely shaped, structure that attracts the system and produces a kind of order that keeps the system within clearly defined bounds. The dynamics are precisely determined by the interactions among the system's variables so, if we knew the exact state of the system at any moment, we could predict all of its future accurately. Unfortunately, chaotic systems are extremely sensitive to the starting point in the state space. Initial conditions that are unimaginably close to each other will eventually follow very different trajectories in the state space, although all will lie somewhere on the attractor. So, while chaotic, non-linear systems are deterministic, they are not predictable in practice since we can never measure their condition at any moment of time with perfect accuracy. A classic example of a chaotic system is the weather, which results from convection and turbulence in the atmosphere. Precise, long-term prediction of the weather will never be possible, because the atmosphere is a non-linear, chaotic system. Nonetheless, reasonably good, large-scale and general predictions can be made for weeks to months in advance since the weather/climate system is not truly random. Whether populations of organisms or ecosystems are sometimes chaotic is still being debated. One difficulty here is that ecological systems are strongly affected by weather, whose chaotic behavior may hide the true, underlying dynamics of the ecological system. But ecological systems are non-linear

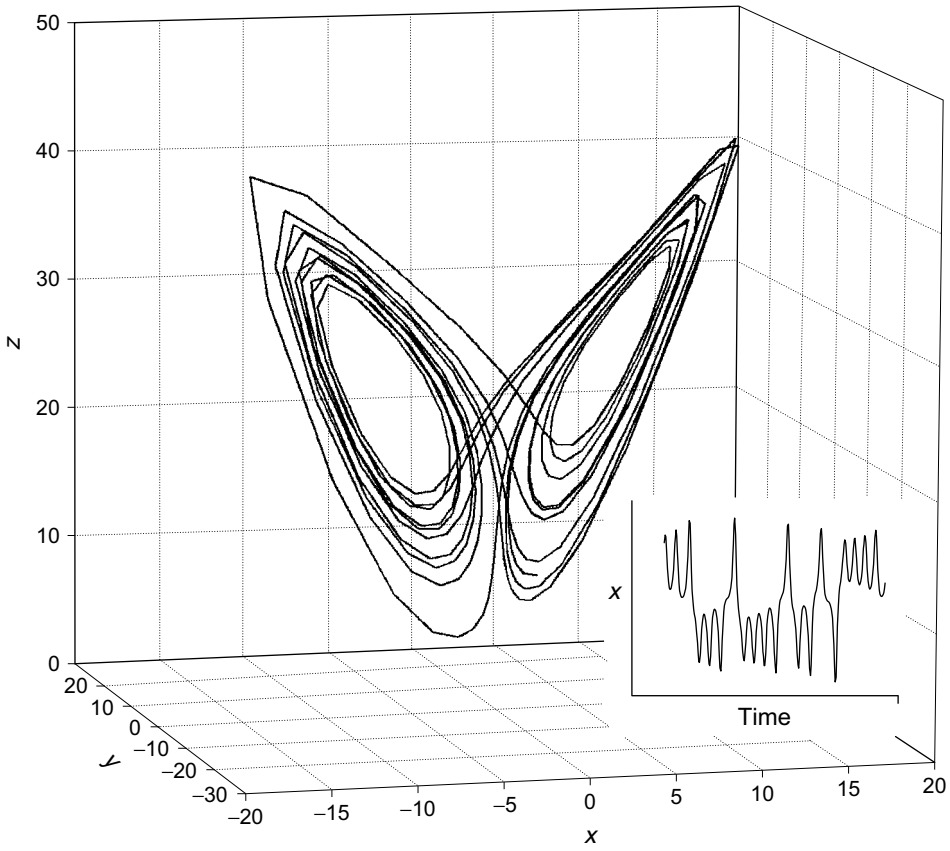


Figure 2 Lorenz's butter y attractor. The curve is part of a trajectory in the three-dimensional state space of a simpli ed model of convection described by the equations $dx/dt = 10(y - x)$, $dy/dt = -xz + 28x - y$, and $dz/dt = xy - (8/3)z$. The presence of the products xz and xy makes this a non-linear system. Starting from anywhere in the state space, the system's trajectory will be drawn to the attractor with the shape that is partially described in the gure by the butter y-shaped trajectory

and thus they could display chaos, at least under certain conditions. Most importantly, non-linear systems can display all these kinds of dynamics, from one or more simple point attractors, to cycles and chaos, depending on the parameters or levels of external variables like climate.

Non-linear systems can also have relatively simple dynamics (similar to linear systems) where the attractor is a stable point in the state space. In this case the system has one equilibrium to which it is always drawn. Even if the system is continually pushed away from this point – perhaps by droughts, floods or fires – it will move back to the original, stable condition. This is essentially the "balance of nature" concept that once dominated the study of ecology and natural history. The location of this equilibrium normally depends on environmental conditions like climate and the richness of the soil. So, e.g., an ecosystem's storage of carbon will depend on the mean annual temperature, annual rainfall, and nutrient content of the soil. If ecosystems were linear, then a change in climate or nitrogen added to the system would change the location of the attractor in some smooth and regular way. A somewhat warmer climate, for example, might slightly alter the equilibrium amount of plants the ecosystem can support. However, a typical feature of non-linear systems is the possibility of more than one possible equilibrium in at least some environments. An example is shown in Figure 3 for a hypothetical system made up of plants and the herbivores that eat them. In a cool climate there might be a single attractor with a relatively large amount of plant biomass and relatively few herbivores. The system will always move back to that equilibrium after any sudden change or perturbation like a fire that alters the state of the system. In a more moderate climate, however, the non-linear interactions between plants and herbivores might produce two simultaneously possible equilibria. Then, parts of the state space will lead to an attractor with lots of plants but few herbivores and other parts of the state space lead to an equilibrium with much less plant biomass but more herbivores. In this case, if the system was near the upper attractor and a fire reduced the plants to a very low-level, then the ecosystem would not return to its previous state. Instead, it would be attracted to the equilibrium with lower plant biomass. Situations like this may occur when fish populations are reduced to very low levels through over-fishing. The boundary between two equilibria may have been crossed so that even stopping the fishing will not allow the population to recover to previous levels. Returning to the illustration in Figure 3, one of the two co-existing attractors may cease to exist in a warmer climate where there is only one, low equilibrium for plant biomass. This property of non-linear systems – multiple attractors and the way they appear and disappear with small changes in the environment – is of obvious importance for how biological systems respond to global change. For some time, the affect of climate change may be smooth and regular until some threshold is passed when the number and qualities of equilibria suddenly change.

Hilbert *et al.* (2000), e.g., demonstrated this possibility for peatlands. Their simple model views peatlands, such as bogs and mires, as just two state variables; the depth of accumulated dead organic matter (peat) and where the water table is found beneath the surface of the peat. These system variables interact non-linearly, primarily because it can be either too wet or too dry for the growth of mosses. Under an intermediate range of precipitation, peatlands may have two possible equilibria with different amounts of accumulated peat and different depth to the water table. At lower or higher precipitation regimes, only one steady state exists. Thus, as climate changes there could be a discontinuous change in the behavior of a peatland as one equilibrium vanishes and another, possibly very different one appears. Similar cases of multiple equilibria have been described for grazing systems (Noy-Meir, 1975). Discontinuous changes in equilibria may also explain discontinuous changes in the characteristics of ecosystems along spatial gradients where environmental conditions vary gradually (Wilson and Nisbet, 1997).

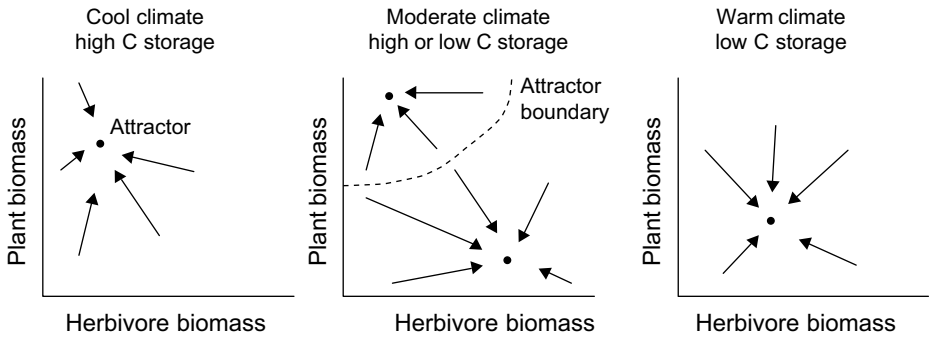


Figure 3 A hypothetical example illustrating how the number and location of equilibria (point attractors) can change as the external environment is altered. The arrows indicate the direction of the system's vectors when the system is at the tail of the arrow. When two simultaneous equilibria occur, there is a boundary between them that separates points that are attracted to one or the other equilibrium

The effect of a gradual change in the environment may be even more profound than in the example of multiple attractors because, as described previously, non-linear systems can have several qualitatively different types of behavior. A simple and generic example of this is the discrete logistic equation. This equation says that the population of an organism next year is a function of both the current population size and the square of its size. The presence of the square term in the equation makes the model non-linear and time is thought of as occurring in discrete steps (year 1, 2, 3, and so on). The effect of increasing the single parameter 'a' (see Box 2) on the population's behavior is illustrated in Figure 4. For each value of 'a' the figure shows the value of the population (x) that the model predicts for each of 300 years after the model ran for an initial 200 years. The inset figures, for 'a' = 2.8, 3.2, 3.5, and 3.9, plot the value of 'x' through time. When 'a' is below 3.0, the population comes to a single equilibrium that becomes higher as 'a' increases. When 'a' = 3.0, however, there is a qualitative change in the population's behavior. Suddenly, it cycles back and forth between two values as illustrated in an inset for 'a' = 3.2. As 'a' increases further, there is a quantitative change in that the difference between the high and low

Box 2 Discrete logistic

In this model of a seasonally breeding population with no overlap of generations, there is only one parameter (a), which determines how fast the population can grow. Discrete, non-linear equations like this one, show all the complex behavior of more complicated systems because they implicitly include a time lag in the feedback between the state of a system and its rate of change. May (1974, 1976) studied how this model's behavior changes as 'a' is varied. [If 'a' was a function of climate, he could just as well have been studying how population behavior could be affected by gradual, global climate change.] Surprisingly, this equation displays thresholds, a special kind of bifurcation, where the behavior changes from a single steady state, to simple cycles, to more complex cycles, and nally to chaos.

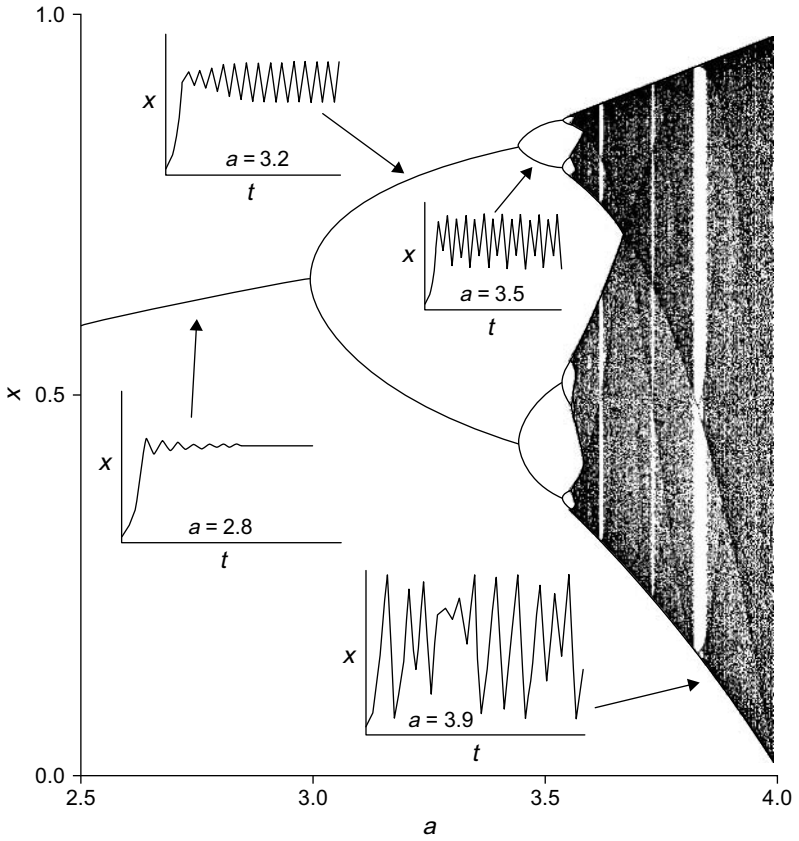


Figure 4 Behavior of the discrete logistic model of a population (x) changing in time (t), $x_{t+1} = ax_t(1 - x_t)$, as the intrinsic rate of increase (a) is varied. The graph shows the states of the population calculated from time 200 to 500 for values of 'a' from 2.5 to 4.0. When 'a' is below 3.0, the population reaches a single equilibrium (see inset for $a = 2.8$) so the graph of population states for time 200 to 500 is a single point. But as 'a' is increased, there are many thresholds, a type of bifurcation, where the behavior of the population changes qualitatively. Passing one such threshold, at $a = 3.57$, switches the system to chaotic dynamics. In this case the population changes erratically so it will be observed in many possible states, thus the dark bands lling in a range of 'x' values

values of the cycle become farther apart. At approximately 'a' = 3.4, there is another qualitative change where the population switches to a cycle with four different values (shown in the inset for 'a' = 3.5). Again there is a period of quantitative change until another qualitative switch occurs where the cycle has eight values. Eventually, at 'a' = 3.57, the dynamics become chaotic, illustrated by the inset for 'a' = 3.9. Surprisingly, there are islands of stable cycles within the potentially chaotic region. These appear as the largely white stripes in the figure for some values of 'a' above 3.57. This bifurcation diagram shows the great sensitivity of non-linear systems to changes in their parameters and many features of it are now known to be applicable to a wide range of more complex and realistic models. In the context of global change, the diagram illustrates that populations, and other kinds of systems, may suddenly change their qualitative behavior as parameters like temperature and precipitation change. Simple models of interacting predators and their prey have revealed that enriching the system, as might happen through increased nitrogen deposition or shorter winters, can cause a shift from a stable equilibrium to large cycles that are likely to cause extinction of both the predator and the prey (Rosenzweig, 1971). Even plant physiological processes may display sharp transitions from one type of behavior to another. The photosynthetic rate of a plant, e.g., has been observed to suddenly switch from stable oscillations to chaos as temperature is increased slightly (Luttge and Beck, 1992).

Traditionally, ecologists have assumed that ecosystems have relatively simple dynamics with a single equilibrium state. So, after a disturbance such as a fire, the ecosystem will go through a regular series of changes as the original ecosystem is gradually re-established. If climate were to change slightly, then the qualities of the ecosystem – species composition, total biomass, etc. – would change gradually and smoothly. However, this is only one way that non-linear systems can behave and we know that ecosystems are non-linear. So, the traditional wisdom is now less widely accepted and there is an increasing understanding that ecological systems can have much more complex dynamics including sudden quantitative or qualitative changes in response to gradual and smooth variation of the environment. The general conclusion is that non-linear systems can display abrupt changes from one equilibrium to another or from one mode of behavior to another (simple equilibrium to periodic, periodic to quasi-periodic, quasi-periodic to chaotic). Thus, system non-linearity implies that we cannot extrapolate one behavior, measured under a restricted range of external variables, to the full range. Instead we may be surprised that the behavior may change suddenly and qualitatively.

See also: **Chaos and Predictability**, Volume 1; **Chaos and Cycles**, Volume 2; **Monitoring in Support of Policy: an Adaptive Ecosystem Approach**, Volume 4.

REFERENCES

Ekeland, I (1988) *Mathematics and the Unexpected*, The University of Chicago Press, Chicago, IL.

Gleick, J (1987) *Chaos: Making a New Science*, Viking Penguin, New York.

Hilbert, D W, Roulet, N, and Moore, T (2000) Modeling and Analysis of Peatlands as Dynamical Systems, *J. Ecol.*, **88**, 230–242.

Logan, J A and Allen, J C (1992) Non-linear Dynamics and Chaos in Insect Populations, *Ann. Rev. Entomol.*, **37**, 455–477.

Lorenz, E N (1963) Deterministic Non-periodic Flow, *J. Atmos. Sci.*, **20**, 130–141.

Luttge, U and Beck, F (1992) Endogenous Rhythms and Chaos in Crassulacean Acid Metabolism, *Planta*, **188**, 28–38.

May, R M (1974) Biological Populations with Non-overlapping Generations: Stable Points, Stable Cycles, and Chaos, *Science*, **186**, 645–647.

May, R M (1976) Simple Mathematical Models with very Complicated Dynamics, *Nature*, **261**, 459–467.

Noy-Meir, I (1975) Stability of Grazing Systems: an Application of Predator–prey Graphs, *J. Ecol.*, **63**, 459–481.

Rosenzweig, M L (1971) Paradox of Enrichment: Destabilization of Exploitation Ecosystems in Ecological Time, *Science*, **171**, 385–387.

Wilson, W G and Nisbet, R M (1997) Co-operation and Competition Along Smooth Environmental Gradients, *Ecology*, **78**, 2004–2017.

FURTHER READING

Crawley, M J (1983) *Herbivory: The Dynamics of Animal–Plant Interactions*, University of California Press, Berkeley, CA.

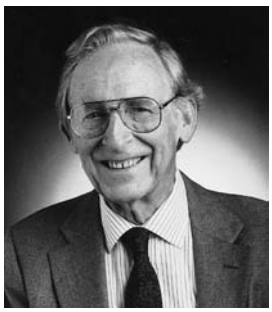
NPP (Net Primary Productivity)

see Productivity (Volume 2)

O

Obligate

see Symbiosis (**Volume 2**)



Odum, Eugene P

(1913-)

For more than 60 years, Eugene P Odum has taught both academicians and the general public the value of nature's goods and services, emphasizing that natural ecosystems provide the life support system for human societies.

Odum received his PhD with Charles Kendeigh at the University of Illinois, US, in 1939. After being turned down for about 30 college faculty positions, he took a job at the Huyck Preserve in upstate New York. A year later, he accepted a position at the University of Georgia, US, where he has remained and became an institution builder, setting up the Savannah River Ecology Laboratory in 1951, what is now called the Marine Institute at Sapelo Island in 1954, and the Institute of Ecology in 1961.

At a faculty curriculum committee meeting in 1945, when he proposed ecology as a required course for zoology majors and was asked What is ecology, anyway?, he decided to write a textbook. When a publisher balked, saying there were no general ecology courses and so no need for the book, his response was Let's create a market. In 1953, he published *Fundamentals of Ecology*. Soon, general ecology was taught at virtually every major college and university, and the book is now in its 5th edition. In addition to over 200 papers in scientific journals, he has written many popular articles and essays, and his latest publication, *Ecological Vignettes: Ecological Approaches to Dealing with Human Predicaments* (1998), is a book for the general reader.

In the field of biology, which has traditionally tried to learn about life by breaking it down into its parts, Odum has emphasized the importance of looking at how the parts work together to make a functioning whole. Many of Odum's earliest publications were fairly traditional studies of avian physiology and energetics, but by the early 1950s he began to study the physiology and energetics of whole ecosystems. Ever seeking ways of measuring ecosystem processes above the species level, he pioneered the measurement of the total rates of primary and secondary production, decomposition, and nutrient ow in aquatic and terrestrial ecosystems. He argued that ecosystems, like individuals, have homeostatic mechanisms that regulate energy ow and the recycling of nutrients. In *The Strategy of Ecosystem Development*, published in *Science* in 1969, he drew parallels between organismal development and ecological succession, pointing out the regularity of change in production and respiration over time. He also argued that, like individuals and species, ecosystems interact with one another, showing, for example, that the export of detritus from salt marshes is important to ocean productivity. Finally, drawing on examples of symbiosis in the plant and animal world, he argued that human ecology and economics are interrelated and that humans should recognize the value of nature's free goods and services and strive for a more mutualistic relationship with nature.

Formal recognition of Odum's contributions has been voluminous. He was elected to the National Academy of Sciences in 1970, he won the La Institute de la Vie Prize (jointly with his brother, H T Odum) in 1975 and the Tyler Award in 1977. In 1987 he received the Crafoord Prize from the Royal Swedish Academy of Science. This is often considered an equivalent to the Nobel Prize, which is not given in ecology. Other honors he has received are Georgia Scientist of the Year in 1967, Eminent Ecologist Award from the Ecological Society of America in 1974, and Educator of the Year from the National Wildlife Federation in 1983, plus over 45 teaching, service, and scientific awards and honorary degrees given by governments, clubs, conservation organizations, and universities.

I first met Eugene Odum in 1965. I was a sophomore in college, and he was already known as the Father of Modern Ecology. Although I was only a teenager, Professor Odum took me to lunch, shared his ideas with me, and listened to my ideas. Now at age 87, Gene looks more like a stately grandfather but he still has the youthful energy and keen intellect that have characterized his lifetime of contributions to science and society. About once a week I still go over to Gene's office in the Institute of Ecology to have lunch with this extraordinary man. He is a bit less sure-footed and much more bent over than he was in 1965, but his mind is as steady and straight as ever, and he is still excited about ecology.

RON PULLIAM *USA*

OH–Radical: is the Cleansing Capacity of the Atmosphere Changing?

Guy Brasseur[1] and Ron Prinn[2]

[1]Max Plank Institute for Meteorology, Hamburg, Germany
[2]Massachusetts Institute of Technology, Cambridge, MA, USA

Many people consider the atmosphere's role in global change to be primarily associated with climate. However, human activities are also profoundly affecting the chemistry of the atmosphere in its own right, irrespective of the greenhouse properties of trace gases. One of the most important features of the atmosphere is its cleansing ability, its capability to transform a wide range of gases emitted from the Earth's surface. The primary chemical species responsible for this cleansing ability is the hydroxyl radical, OH.

The atmosphere is an oxidizing medium. Almost all hydrogen containing gases of natural and anthropogenic origin are oxidized by the hydroxyl (OH) free radical, a species often referred to as the detergent of the atmosphere. Other oxidation processes include reactions with ozone (O_3), hydrogen peroxide (H_2O_2), the nitrogen trioxide free radical (NO_3), and halogen species. Hydrogen peroxide, for example, converts sulfur dioxide to sulfuric acid in cloud and rain droplets at low pH, and hence contributes to the formation of acid precipitation. The nitrogen trioxide free radical plays a significant role at night, primarily in polluted areas.

The rate at which oxidation occurs varies widely from species to species. For example, the oxidation of methane typically takes nine years, while that of isoprene takes less than a day. Thus, even though the magnitude of the surface emissions of these two hydrocarbons is similar, their atmospheric burdens differ by several orders of magnitude.

The rate at which chemical compounds are eliminated from the atmosphere is very often not determined by the precipitation rate, but by the rate at which they are oxidized into a water-soluble compound. This oxidation rate therefore plays a central role in many global environmental problems. An important issue is to quantify any potential change in the oxidizing power of the atmosphere that may have resulted from human activities (Thompson, 1992).

Over the past decade and century, the oxidizing power of the atmosphere may have changed. For example, the increasing penetration of solar ultraviolet light in response to stratospheric ozone depletion and the increased emissions of NO_X and non-methane hydrocarbons from combustion tend to increase ozone and OH concentrations in the troposphere. Conversely, the increase in carbon monoxide emissions, which have also resulted from increased biomass and fossil fuel burning since the pre-industrial era, tend to reduce the ozone and OH abundances. Similarly, increases in the methane abundance due to increases in cattle, rice, and natural gas usage should have contributed to a decrease in the global OH concentration and therefore in the oxidizing power of the atmosphere.

Since there is presently no instrumentation for regional to global-scale observations of tropospheric OH, the change in the oxidation power of the atmosphere cannot be directly monitored. Estimates must therefore rely primarily either on model calculations or deductions using measurements of compounds destroyed by OH. Model calculations suggest that global tropospheric levels of OH have probably decreased significantly since the pre-industrial era. The quantitative value of this decrease varies from model to model, however, with figures ranging from 3% in the model of Lelieveld and Van Dorland (1995) to 22% in the model of Roelofs *et al.* (1997). Intermediate values of 9% are provided by Wang and Jacob (1998) and 10–15% by Crutzen and Zimmermann (1991). Earlier estimates based on one-dimensional models suggested a decrease, ranging from 10–30% (Thompson, 1992). The general consensus from modeling studies is that, despite the large increase in the emissions of carbon monoxide, hydrocarbons and nitrogen oxides, and the stratospheric ozone depletion during the 20th century, the partially offsetting effects of these factors implies that the globally averaged concentration of OH has remained relatively constant. Decreases probably do not exceed 20% (Wang and Jacob, 2000).

The analysis of the methylchloroform (CH_3CCl_3) observations performed since 1978 at several surface monitoring stations of the Atmospheric Lifetime Experiment/Global Atmospheric Gases Experiment/Advanced Global Atmospheric Gases Experiment (ALE/GAGE/AGAGE) (Prinn *et al.*, 2000 and references therein) provides an indirect method to derive trends in the global concentration of tropospheric OH. Methylchloroform's sources, which are purely anthropogenic, are believed to be very accurately accounted for. By combining the estimates of these sources with the methylchloroform measurements, little change in the OH density was deduced by Prinn *et al.* (1995) for the period 1978–1994 (specifically $0.0 \pm 0.2\%$ per year, 1s error range). A similar study by Krol *et al.* (1998) suggested a small increase in the OH density ($0.46 \pm 0.6\%$ per year, full error range) from 1978–1993.

In summary, recent analyses suggest that the oxidizing power of the atmosphere has remained relatively constant since the pre-industrial era, with a decrease in the OH concentration, and hence an increase in the lifetime of major primary pollutants (or greenhouse gases like methane) that have not exceeded 20%. On shorter time scales, the abundance of OH may have changed rather substantially, however; for example, in response to rapid ozone depletion in the stratosphere, uctuations in tropospheric water vapor, or in the emissions of nitrogen oxides or carbon monoxide (see, e.g., Madronich and Granier, 1992; Granier *et al.*, 1996). Future changes in the oxidizing power of the atmosphere will depend on future anthropogenic emissions and hence on future economic development and population growth.

REFERENCES

Crutzen, P J and Zimmermann, P H (1991) The Changing Photochemistry of the Troposphere, *Tellus*, 43(AB), 136–151.

Granier, C, Muller, J F, Madronich, S, and Brasseur, G P (1996) Possible Causes for the 1990–1993 Decrease in Global Tropospheric CO_2 Abundances: A Three-dimensional Sensitivity Study, *Atmos. Environ.*, 30, 1673–1682.

Krol, M, van Leeuwen, P J, and Lelieveld, J (1998) Global OH Trend Inferred from Methylchloroform Measurements, *J. Geophys. Res.*, 103, 10 697–10 711.

Lelieveld, J and Van Dorland, R (1995) Ozone Chemistry Changes in the Troposphere and Consequent Radiative Forcing of Climate, in *Atmospheric Ozone as a Climate Gas*, eds W C Wang and I S A Isaksen, Springer-Verlag, New York, 227–258.

Madronich, S and Granier, C (1992) Impact of Recent Total Ozone Changes on Tropospheric Ozone Photo Dissociation, Hydroxyl Radicals, and Methane Trends, *Geophys. Res. Lett.*, 19, 465–467.

Prinn, R G, Weiss, R F, Miller, B R, Huang, J, Alyea, F N, Cunnold, J M, Fraser, J P, Hartley, D E, and Simmonds, P G (1995) Atmospheric Trends and Lifetime of CH_3CCl_3 and Global OH Concentrations, *Science*, 269, 187 192.

Prinn, R G, Weiss, R F, Fraser, P J, Simmonds, P G, Cunnold, D M, Alyea, F N, O'Doherty, S, Salameh, P, Miller, R B, Huang, J, Wang, H R J, Hartley, D E, Harth, C, Steele, L P, Sturrock, G, Midgeley, P, and McCulloch, A (2000) A History of Chemically and Radiatively Important Gases in Air Deduced From ALE/GAGE/AGAGE, *J. Geophys. Res.*, 105, 17 751–17 793.

Roelofs, G-J, Lelieveld, J, and Van Dorland, R (1997) A Three-dimensional Chemistry/General Circulation Model Simulation of Anthropogenically Derived Ozone in the Troposphere and its Radiative Forcing, *J. Geophys. Res.*, 102, 23 389–23 401.

Thompson, A (1992) The Oxidizing Capacity of the Earth's Atmosphere: Probable Past and Future Changes, *Science*, 256, 115–1165.

Wang, Y and Jacob, D J (1998) Anthropogenic Forcing on Global Tropospheric Ozone and OH Since Pre-industrial Times, *J. Geophys. Res.*, 103, 31 123–31 136.

Oligotrophic/Heterotrophic/Eutrophic

Oligotrophic/heterotrophic/eutrophic are terms that describe stages in the aging of aquatic ecosystems. They are usually applied to lakes and ponds but also to marine locations that are sufficiently sequestered to allow some degree of independence from the main body of the ocean. Freshwater aquatic ecosystems typically start as oligotrophic lakes, which have low nutrient levels and are consequently relatively unproductive. Oligotrophic lakes are characteristically pristine, have high oxygen levels and only small populations of plankton and fish. An input of organic materials (humus or pollution) may lead to a heterotrophic ecosystem, where a somewhat larger biotic component is supported by the external source of organic nutrition. Continuous enrichment of a lake with inorganic nutrients leads to the process of eutrophication. Eutrophic systems are highly productive, being characterized by heavy populations of plankton and of those fishes that thrive in polluted waters. Eutrophic lakes have much lower oxygen levels, and may be anoxic in their deeper waters; they have high levels of inorganic nutrients. Phosphate levels are especially helpful in characterizing their status: high or low phosphate contents respectively indicate the relatively stable oligotrophic or eutrophic condition; intermediate levels suggest an unstable system that is undergoing eutrophication.

The process of eutrophication is normal and natural, causing virtually all aquatic ecosystems slowly but inevitably to be converted into swamps and eventually to

dry land. But eutrophication is greatly speeded up by many human activities, so that pristine lakes are rapidly converted to ecosystems characterized by algal blooms, pollution and anoxia. Such an ecological succession involves major or almost complete replacement of the ora and fauna (including the species of both commercial and sport fishing) of the system, and results in very large increases in the rates of deposition of both organic and inorganic material.

R G S BIDWELL *Canada*

Ontogeny

see Plant Ontogeny (**Volume 2**)

Oxygen, Anthropogenic Impacts

see Anthropogenic Impacts on Atmospheric Oxygen (**Volume 2**)

P

PAGES (Past Global Changes)

see IGBP Core Projects (*Opening essay*, **Volume 2**)

PAR (Photosynthetically Active Radiation)

PAR is light of the specific wavelengths that are absorbed by the photosynthetic pigments of plants or bacteria and whose energy is trapped and used to drive photosynthesis. In most green plants the chlorophylls are the primary PAR absorbing pigments of photosynthesis. They absorb mainly in the red (620–720 nm) and blue (380–480 nm) regions of the spectrum. In red and brown algae and many photosynthetic bacteria, the phycobilin pigments (absorbing in the range 500–650 nm) are also photosynthetically active. Carotenes (420–540 nm) may transfer some of their absorbed energy to the photosynthetic pigments, but much of the light energy absorbed in the 500 nm region is not available for photosynthesis.

In most land plants, the important PAR is light at the blue and red end of the spectrum (absorbed by chlorophylls). Marine algae are faced with a problem in that red light is much more strongly absorbed by water than light of shorter wavelengths, and these plants have adapted in varying degree to overcome this difficulty. The green algae, which live mainly near the surface, rely primarily on chlorophyll for photosynthetic light absorption. Brown algae, whose range of PAR is substantially shifted toward the blue end of the spectrum, tend to live in deeper water. The red algae, whose pigments absorb more strongly in the blue region, live in the greatest depths of water, down to 200 m.

Not all light absorbed by plants is PAR. Higher plants, in particular, contain small concentrations of a range of pigments that are sensitive to daylength or that stimulate the hormone mechanisms that control morphology (e.g., phototropism). Other non-photosynthetic pigments whose precise physiological roles are often difficult to determine may be present in large amounts (e.g., carotenes, anthocyanins). They may create spectacular colors in plants or their parts, but they are not involved in photosynthesis.

R G S BIDWELL *Canada*

Passerine

Passerine refers to birds of the order Passeres or Passeriformes (perching birds). This large order includes about half of the known bird species, including most of the common land birds. The term is sometimes used to refer to sparrows, which are members of the Passeriformes, and is occasionally applied (descriptively rather than scientifically) to birds that resemble a sparrow or are approximately sparrow size (e.g., Passerine Owl, Passerine Parrot).

R G S BIDWELL *Canada*

Patch Dynamics

see Metapopulations (**Volume 2**)

Patchy Landscapes

see Metapopulations (**Volume 2**)

Pelagic

see Benthic/Pelagic (**Volume 2**)

Permafrost Thawing: Ecological Effects

see Permafrost (Volume 1)

REFERENCES

Budyko, M I (1956) *The Heat Balance of the Earth's Surface*, (translated from the Russian by N I Stepanova, 1958), Weather Bureau, Washington, DC.

Penman, H L (1963) *Vegetation and Hydrology*, Commonwealth Bureau of Soils, Tech. Comm. No. 53, Commonwealth Bureau of Soils, Harpenden, UK, 1–125.

Thornthwaite, C W (1948) An Approach Toward a Rational Classification of Climate, *Geogr. Rev.*, **38**, 55–94.

MARIE SANDERSON *Canada*

PET (Potential Evapotranspiration)

The term PET was introduced into the climatic literature in the mid-20th century by three scientists working independently: Thornthwaite in the United States, Penman in England and Budyko in the USSR. They recognized that two aspects of evapotranspiration must be identified. *Actual evapotranspiration* is the water that evaporates and transpires under existing soil moisture conditions, while *potential evapotranspiration* (PET) represents the amount of water that would be lost by evaporation and transpiration from a large vegetation-covered surface if an unlimited supply of water were to exist. All three realized that PET depends primarily on the evaporation equivalent of the available net radiation. Penman and Budyko derived formulae for estimating PET using data on net radiation, humidity and wind speed. However for most locations throughout the world, data on these parameters are not available.

Thornthwaite in the United States was aware of this problem and devised a method of estimating PET from data on temperature and length of day only. His formula was based on the use of water by irrigated crops in different parts of the United States, and monthly values of PET could be readily calculated from published tables and nomograms. PET was the central concept in his water budget model in which PET was compared with water supply (precipitation plus soil moisture) to derive estimates of actual evapotranspiration, water surplus and water deficiency. As computer programs of the water budget became available, it has been used worldwide by hydrologists, engineers, botanists and geographers to estimate actual evapotranspiration, water surplus and water deficiency.

PET can be measured directly by an evapotranspirometer, a Thornthwaite-designed instrument in which a block of soil with vegetation-covered surface is kept supplied with water by underground irrigation. Results from evapotranspirometers installed in various regions of the world indicate good agreement with computed PET in most areas.

Phenology

Martin J Lechowicz
McGill University, Montréal, Canada

Phenology is the study of the seasonal timing of events in nature: when flowers bloom, trees leaf out, birds migrate, animals hibernate, fish spawn, phytoplankton blooms, lakes freeze, and the like. Phenology is not concerned with events such as the solstices and equinoxes that are determined by astrophysical laws – these vary only slowly and predictably over millennia. Events that are influenced by climatic factors, on the other hand, are far more variable and less predictable from year to year. Organisms able to survive in a given climate must respond effectively to the vagaries of day to day weather during the seasonal cycle. The advance of spring warming is an expected and normal part of high-latitude climates, but no one expects the end of winter to begin on a particular date year after year or to follow a specific sequence of ever-warmer days until summer arrives. Phenological records thus can provide an integrative index of weather through the seasons, a capability that is becoming increasingly valuable in gauging trends indicative of changing climate.

Long before scientific interest in phenology arose in the 19th century, people had paid close attention to the seasonal timing of natural events. Some phenological events associated with cultural and religious festivals have been recorded for a very long time; annual records for the flowering of cherry trees in Kyoto, Japan date back to the 9th century. Before weather records and weather forecasting existed, the everyday lives of people depended on closely observing the timing of natural events. The return of migratory birds or the leafing of a particular species of forest tree could provide a reliable signal of when best to plant a frost-sensitive crop. Calendars of the seasons, collating this sort

of folk knowledge, began to appear in the 18th century and at about the same time, individuals and governments began to formally observe and record phenological events. The Marsham family recorded daily phenological observations at their estate in Norfolk, England from 1736 through 1925! Many, shorter series of phenological records are available, especially for sites in the north temperate zone.

The scientific study of phenology initially focused on the relationship between weather and the timing of events. Can we predict the timing of events from our knowledge of the response of an organism to a progression of weather conditions? Yes, to a considerable degree we can, usually with an accuracy of a few days. For example, it is possible to predict the timing of leafing out or flowering in most trees within 2–5 days from place to place and year to year. We also have some understanding of the reasons why different species in the same locality do not all have exactly the same phenology. Due to differences in functional organization conserved in evolution, each species at a locality responds to changing climatic conditions somewhat distinctly.

More recently, the scientific study of phenology has shifted to questions of how phenology will be affected by global change, and what consequences any changes may have for species distribution and ecosystem function. An important impetus to this shift was the recent discovery that modest changes in the length of the season when plants are actively photosynthesizing can significantly alter the storage of carbon. A forest at high latitudes can be a net sink or a net source of atmospheric carbon dioxide by virtue of a week more or less in the length of the growing season. In other words, there are feedbacks between phenology and factors affecting the global carbon balance that may act to accelerate or decelerate rates of climate change. The study of phenology therefore has taken on new urgency and scientific significance as we struggle to understand the nature and consequences of climate change.

BRIEF HISTORY OF PHENOLOGY

The modern roots of phenology as a science are in the careful observations of naturalists in the 18th and 19th centuries, Gilbert White and Henry David Thoreau being perhaps the best known among them. In these times, in fact, many people engaged in regular observation of phenological events. The daily records of the Marsham family, noting the seasonal activity of birds and plants in central England (Sparks and Carey, 1995), span the period between 1736 and 1925. Thomas Mikesell, a farmer in Ohio, made meticulous, twice-daily records of weather and phenology between 1883 and 1912 (Lechowicz, 1995). Long-term data sets of this sort exist in moderate amounts, usually for only single localities. Naturalists have continued a tradition of phenological observation into this century (Fitter *et al.*, 1995; Bradley *et al.*, 1999), and their efforts have been augmented by the collection of phenological data in various scientific and government networks with greater spatial coverage (Ahas, 1999; Bruns, 1996). For example, the Royal Meteorological Society (UK) published phenology data regularly for at least 75 years, terminating about 1945.

FACTORS AFFECTING PHENOLOGY

The precise timing of events in the natural cycle of the seasons often determines the success or failure of individual plants or animals. Many migratory birds must lay eggs in advance of the availability of ample food for their young, but neither so early that hatchlings want for caterpillars nor so late that the young are not ready to migrate in advance of winter. Deciduous trees burst bud only when the risk of late frosts is low, but not so long into summer that their annual period of photosynthetic activity is much diminished. In the seasonal tropics, many trees flower during the dry season with their fruits maturing in time for seedlings to establish themselves with the onset of the rains. In all these instances, the timing of phenological events must respond to environmental cues that presage future conditions, usually with allowance for the costs and benefits of alternative timing of an event. An exception is in events such as blooms of oceanic phytoplankton, which have very short life cycles and simply respond rapidly to periods of high irradiance in less cloudy weather when water temperatures are adequate (Townsend *et al.*, 1994).

Although the diversity and complexity of phenological responses are impressively great, the environmental cues on which phenological events depend are relatively few. Phenological cues fall into two basic classes: photoperiodic and climatic. Temperature influences predominate among the climatic cues, although precipitation and irradiance regimes sometimes also play a role. The photoperiodic and climatic signals differ in their nature and in their utility for gauging environmental conditions later in the annual cycle.

Photoperiod, the length of the day relative to the night, changes at different latitudes in a very predictable cycle through the year. The tilt of the Earth's axis of rotation and its orbital path around the sun determine the seasonal progression of photoperiod. As a consequence, photoperiodic cues in a locality are essentially invariant from year to year. Day length will begin to increase after the winter solstice, reaching 12 hours on the vernal equinox and getting steadily longer until the summer solstice; then the cycle reverses with shortening days until the winter solstice passing through 12 h of daylight on the autumnal equinox. The annual amplitude of variation in photoperiod does vary from place to place, being near zero in equatorial regions and increasingly greater toward the poles. Organisms in a given locality thus can use photoperiod to reliably gauge the time of year.

The problem with photoperiod alone as a phenological cue is that climate is nowhere entirely predictable within or among years. Just knowing the time of the year is insufficient to evaluate the likely environmental conditions in the weeks or months ahead. Lengthening days can cue the onset of spring, but not whether the spring will advance quickly or slowly, or be especially warm or cool. Organisms thus use temperature, and sometimes water, regimes to better estimate the progression of seasons in a given year and locality. These climatic signals are useful cues for phenological events, but are also not entirely reliable. Changes in atmospheric circulation may indeed bring an earlier than usual onset to spring warming, but not necessarily preclude a late freeze. Nonetheless, there is enough coherence in weather over weeks and months in a given year that recent events can help anticipate future conditions.

Many organisms outside the tropics seem to depend on short-term (2–12 week) temperature trends as the most reliable aspect of climate for assessing seasonal changes and triggering phenological responses. Phenological events in spring, for example, often can be predicted from heat sums, which are a measure of the accumulation of warmth above a base temperature. Phenological events occur when a certain heat sum is attained. Thermal summation may begin after some requisite degree of chilling obtains, or when a photoperiodic cue is satisfied. In some instances, the heat sum required to trigger the event is contingent on the degree of prior chilling or the length of the photoperiod. For example, longer chilling of trees can reduce the heat sum required for flowering. Similarly, some oak trees will leaf out at a lower heat sum as the days become longer. Photoperiod becomes a more dominant cue later in the season but some influence of warming or chilling trends persists. The formation of flower buds is often set by quite precise photoperiodic signals, but flowering itself depends on the temperature regime (Fitter *et al.*, 1995). Similarly, egg-laying by birds is set by a combination of photoperiodic and climatic cues, including indirect effects of temperature regime on the availability of insects to feed hatchlings (Visser *et al.*, 1998). The physiological means by which organisms actually monitor and respond to these sorts of direct and indirect influences of chilling, warming and photoperiod are poorly known. We can use photoperiod and weather data to predict phenological events, but without any sure understanding of underlying mechanisms (Hunter and Lechowicz, 1992; Chuine *et al.*, 1999).

In tropical regions, seasonal variations in photoperiod and temperature are less marked than at higher latitudes, but precipitation can be strongly seasonal in some regions (Corlett and LaFrankie, 1998). Although most tropical trees have consistent and fairly well defined flowering times during the year, the environmental cues that trigger flowering are not well known. Some trees flower and fruit when there is a recharge of tissue water either through the onset of rains or when the dropping of leaves in the dry season places less transpirational demand on the tree, but this is not a universal phenomenon (Wright and Calderon, 1995). Flowering and fruiting typically occur throughout the year on one species or another in tropical ecosystems, with the adaptive value of timing of flowering for each species obscure. The phenology of tropical species is too little studied to support broad generalizations (Corlett and LaFrankie, 1998).

PHENOLOGICAL VARIATION WITHIN COMMUNITIES

Species of generally similar organisms in a locality do not usually share exactly the same phenology. This is because phenology is only one element in each species suite of adaptations to the local environment. Take as an example, the warm- versus cool-season grasses that co-occur throughout the North American Great Plains (White *et al.*, 1997). The cool-season grasses are green and actively growing in the cool and moist weather of fall and spring, but brown and dormant in summer. The warm-season grasses conversely only begin greening up in late spring and are actively growing during the warm, dry summer months. The warm-season grasses have C_4 photosynthetic metabolism, well suited to warm, dry conditions but sensitive to chilling. The cool-season grasses have C_3 photosynthetic metabolism, well suited to cool, moist conditions but sensitive to drought. The contrasting phenology of the warm- and cool-season grasses is part of an integrated suite of traits that defines their functional ecology. There are, of course, also phenological differences among species within each functional group, but these are relatively minor variations within the basic dichotomy. The two groups of grasses together lengthen the effective period for photosynthetic production in the prairie ecosystem.

The phenological differences among co-occurring species sometimes can become more diverse and less easy to characterize by simple dichotomies. Consider, for example, the situation in a mixed wood at mid-latitudes where two functional groups of trees share dominance in the community. Deciduous species are leafless over winter, with individual leaves having longevity measured in months. Evergreen species keep leaves over the winter, with individual leaves having longevity measured in years. The two groups are apparently well defined by the sharp distinction in their overwintering leaf habit, but, in fact, they fall along a continuum of variation in leaf longevity that organizes their production ecology (Reich *et al.*, 1997). Evergreen and deciduous species overlap in their timing of production of new leaves in spring and early summer, both within and between the two apparent functional groups. Such differences in foliar phenology among co-occurring tree species are linked to evolutionary and paleoecological as well as

functional considerations unique to each individual species (Lechowicz, 1984). This individualistic nature of species phenology complicates the analysis of the impact of changing climate on community composition and ecosystem function. As climate shifts, we can expect that relationships among species differing in phenology may also shift, both in terms of competitive and trophic interactions (Harrington *et al.*, 1999).

PREDICTING THE EFFECTS OF CHANGING PHENOLOGY

We can imagine three basic responses of species to a reasonably gradual climate change. First, species may migrate to remain in climatic regimes to which they are phenologically well adapted. There is good evidence that such migrations have occurred during deglaciation. As glaciers retreat and climatic zonation shifts to higher latitudes, the formerly glaciated landscape has been colonized by species better suited to north temperate, boreal and arctic regions that have migrated from glacial refugia now subject to increasingly warm and less strongly seasonal climates. Second, species subject to a changing climate may not migrate to track their favored climatic regime, but rather stay in place and adjust to changing conditions. In terms of phenology, there is always a certain degree of variation among individuals within populations in their responses to environmental cues for seasonality. From an evolutionary point of view, we expect the maintenance of such variability among individuals as a form of bet hedging against normal patterns of year to year climatic variability. We, therefore, can expect the average timing of a phenological event in a particular species to track climate change provided the ongoing changes do not exceed the previous range of climatic variability. A simple shift in the balance of phenologically early or late individuals in a population can then accommodate the shift in climate; over sufficient time the shifting advantage of different phenological variants within the population may affect a genetically determined adaptive shift in the phenology of the species. Finally, species may simply fail to accommodate effectively to changing climate, perhaps becoming much less dominant in a locality or even going locally extinct. It is difficult to know with any certainty what effects these diverse changes at the species-level may have at the community and ecosystem levels.

RECENT SHIFTS IN PHENOLOGY ASSOCIATED WITH CHANGING CLIMATE

Since phenology is so closely linked to weather conditions, trends in phenology may serve as natural indicators of an ongoing climate change. Where we have suitable records over reasonably long periods of time, we can directly evaluate the degree to which recent climatic trends have influenced phenological events. Various phenological observations do suggest a climatic warming trend in recent decades in some parts of the world. This is the case for leafing of trees, timing of flowering, and egg laying by birds, and spring migration by both birds and butterflies. Modeling (White *et al.*, 1999), ground (Menzel and Fabian, 1999) and satellite observations (Myeni *et al.*, 1997) all support a recent increase of about 8–10 days in the period when mid-latitude forests of Europe and North America are in leaf. This lengthening of the growing season results primarily from earlier springs, but also a later onset of winter. The increasing earliness of spring has also been manifest in advances in the leafing and flowering of many individual plant species (Fitter *et al.*, 1995; Sparks and Carey, 1995; Bradley *et al.*, 1999). Egg laying in 36 species of birds in the United Kingdom has become earlier in recent decades (Crick and Sparks, 1999) as has the spring arrival of migrant species (Sparks and Carey, 1995; Sparks, 1999). These results suggest that recent climate warming may be having widespread impact on the phenology of many species, which can potentially alter biotic interactions at the community and ecosystem levels. While there are some observable changes in ecosystem function (Myeni *et al.*, 1997; Menzel and Fabian, 1999) and some butterfly species have shifted their ranges poleward (Parmesan *et al.*, 1999), there is no evidence at this time that global systems are being disrupted by widespread species migration or extinction associated with climate change. The real question is whether or not climate change has begun that in time will lead to such reorganization of biota and major shifts in global ecosystems.

PREDICTING LONG-TERM PHENOLOGICAL RESPONSES TO CLIMATE CHANGE

Climate may change in two respects, quantitatively in its regional expression, or qualitatively in its basic seasonal structure. Each has quite different potential consequences for species and ecosystems. A quantitative shift in local climate, provided it is neither too rapid nor too great, may have only negligible influence on species distribution and ecosystem function. Species may simply respond to their usual cues, perhaps prospering by a longer growing season or shifting their range to track their preferred climatic regime. It is unlikely that modest quantitative shifts in climate will dramatically alter community composition or ecosystem function. This sort of modest quantitative change seems to characterize our present situation. On the other hand, qualitative shifts in the nature of the seasonal cycle have greater potential to disrupt phenological responses in a community in ways that more rapidly disrupt ecosystem function. If the basic nature or stability of the seasonal cycle is altered by climate change, then cues evolved in

an earlier climatic regime may no longer lead to effective phenological responses. Even if the cues remain consistent with evolved phenological responses, any increase in the variance of climatic regimes poses a serious risk. Extreme events such as an early winter storm can extirpate populations adapted to a reliable progression of fall weather. There is evidence that global patterns of seasonality are reasonably stable from year to year, at least at the coarse temporal resolution afforded by satellite data (Potter and Brooks, 1998). We do not know whether more precise data would reveal trends in seasonality with adverse implication for the adaptive value of contemporary phenological responses. The greatest challenge of global change to contemporary biological systems may not be a gradual shift in average climate over time, but rather the possibility of qualitative shifts in the nature and stability of seasonality.

REFERENCES

Ahas, R (1999) Long-Term Phyto-, Ornitho- and Ichtyophenological Time-Series Analyses in Estonia, *Int. J. Biometeorol.*, **42**, 119–123.

Bradley, N L, Leopold, A C, Ross, J, and Huffaker, W (1999) Phenological Changes Reflect Climate Change in Wisconsin, *Proc. Nat. Acad. Sci. USA*, **96**, 9701–9704.

Bruns, E (1996) Observation Program of the International Phenological Network (Ipn) – Prepared by the Deutscher Wetterdienst (DWD), *Phenol. Seasonality*, **1**, 25–31.

Chuine, I, Cour, P, and Rousseau, D D (1999) Selecting Models to Predict the Timing of Flowering of Temperate Trees: Implications for Tree Phenology Modeling, *Plant, Cell Environ.*, **22**, 1–13.

Corlett, R T and LaFrankie, Jr, J V (1998) Potential Impacts of Climate Change on Tropical Asian Forests Through an Influence on Phenology, *Clim. Change*, **39**, 439–453.

Crick, H Q P and Sparks, T H (1999) Climate Change Related to Egg-Laying Trends, *Nature*, **399**, 423–424.

Fitter, A H, Fitter, R S R, Harris, I T B, and Williamson, M H (1995) Relationships between first Flowering Date and Temperature in the Flora of a Locality in Central England, *Funct. Ecol.*, **9**, 55–60.

Harrington, R, Woiwod, I, and Sparks, T (1999) Climate Change and Trophic Interactions, *Trends Ecol. Evol.*, **14**, 146–150.

Hunter, A F and Lechowicz, M J (1992) Predicting the Timing of Budburst in Temperate Trees, *J. Appl. Ecol.*, **29**, 597–604.

Lechowicz, M J (1984) Why do Temperate Trees Leaf Out at Different Times? Adaptation and Ecology of Forest Communities, *Am. Nat.*, **124**, 821–842.

Lechowicz, M J (1995) Seasonality of Flowering and Fruiting in Temperate Forest Trees, *Can. J. Bot.*, **73**, 175–183.

Menzel, A and Fabian, P (1999) Growing Season Extended in Europe, *Nature*, **397**, 659.

Myeni, R B, Keeling, C D, Tucker, C J, Asrar, G, and Nemani, R R (1997) Increased Plant Growth in the Northern High Latitudes from 1981 to 1991, *Nature*, **386**, 698–701.

Parmesan, C, Ryrholm, N, Stefanescu, C, Hill, J K, Thomas, C D, Descimon, H, Huntley, B, Kalla, L, Kullberg, J, Tammaru, T, Tennent, W J, Thomas, J E, and Warren, M (1999) Poleward Shifts in Geographical Ranges of Butterfly Species Associated with Regional Warming, *Nature*, **399**, 579–583.

Potter, C S and Brooks, V (1998) Global Analysis of Empirical Relations between Annual Climate and Seasonality of NDVI, *Int. J. Remote Sens.*, **19**, 2921–2948.

Reich, P B, Walters, M B, and Ellsworth, D S (1997) From Tropics to Tundra: Global Convergence in Plant Functioning, *Proc. Nat. Acad. Sci., USA*, **94**, 13 730–13 735.

Sparks, T H (1999) Phenology and the Changing Pattern of Bird Migration in Britain, *Int. J. Biometeorol.*, **42**, 134–138.

Sparks, T H and Carey, P D (1995) The Responses of Species to Climate Over Two Centuries: an Analysis of the Marsham Phenological Record, 1736–1947, *J. Ecol.*, **83**, 321–329.

Townsend, D W, Cammen, L M, Holligan, P M, Campbell, D E, and Pettigrew, N R (1994) Causes and Consequences of Variability in the Timing of Spring Phytoplankton Blooms, *Deep-Sea Res.*, **41**, 747–765.

van Vliet, A and der Groot, R (2001) *Phenology and Climate: Evidence of Change*, CHANGE, RIVM, Box 1, 3720 BA Bilthoven, The Netherlands, 12–14, Vol. 56.

Visser, M E, van Noordwijk, A J, Tinbergen, J M, and Lessells, C M (1998) Warmer Springs Lead to Mistimed Reproduction in Great Tits (Parus major), *Proc. Royal Soc., Lond., Ser. B*, **265**, 1867–1870.

White, M A, Running, S W, and Thornton, P E (1999) The Impact of Growing-season Length Variability on Carbon Assimilation and Evapotranspiration Over 88 Years in the Eastern US Deciduous Forest, *Int. J. Biometeorol.*, **42**, 139–145.

White, M A, Thornton, P E, and Running, S W (1997) A Continental Phenology Model for Monitoring Vegetation Responses to Interannual Climatic Variability, *Global Biogeochem. Cycles*, **11**, 217–234.

Wright, S J and Calderon, O (1995) Phylogenetic Patterns Among Tropical Flowering Phenologies, *J. Ecol.*, **83**, 937–948.

Phenotype/Genotype

The *phenotype* is the sum of the characteristics that are expressed by an organism; that is, its appearance. The *genotype* is the genetic constitution, the sum of all the genes present in an organism; that is, its heredity.

The phenotype of an organism is dependent to a large extent on its genes. Indeed, the phenotype of an organism changes with its development, and it may be affected by environmental factors. However, the development of the organism and the manner in which it reacts to environmental stimuli are programmed by the genotype. The phenotype is thus a product of the interaction of the genotype with the environmental factors it encounters during its lifetime. The

resulting variability in the appearance of organisms with the same genotype is referred to as *phenotype plasticity*.

The genotype of an organism is constant, or nearly so, during the organism's life, and is normally received at fertilization as a contribution from both its parents. Some organisms reproduce asexually (e.g., bacteria, primitive plants) or by apomyxis (in some higher plants, e.g., certain *Rosaceae* and *Compositae*, pollination takes place but fertilization does not, and the embryo develops from the egg cell only) or by natural cloning; these receive the unchanged genotype of only one parent. Changes in genome during the life of an individual normally occur only by genetic mechanisms involving actual changes in the genes (e.g., mutations) or their reassortment (e.g., by recombination).

It is possible for organisms to have the same genotype but different phenotypes because of developmental or environmental factors. It is also possible for organisms to have the same phenotype but different genotypes. This could occur, for example, if organisms were either homozygous or heterozygous for a dominant allele, which would be expressed equally in either case. But these possibilities do not alter the fact that the phenotype of an organism depends largely on its genotype.

R G S BIDWELL *Canada*

Phosphorus Cycle

H Tiessen and John W B Stewart
University of Saskatchewan, Saskatoon, Canada

Phosphorus (P) is an essential component of all organisms, with both genetic information and energy metabolism depending on phosphorus. Nitrogen (N) and phosphorus are the principal nutrients limiting biological production, but the fixation of atmospheric nitrogen into biomass depends on an adequate phosphorus availability, so that phosphorus can be said to be the single limiting determinant of ecosystem productivity and function. Phosphorus is liberated from primary rocks, is relatively abundant in young landscapes and becomes scarce with the weathering of soils. Organisms contain phosphorus at concentrations much higher than those present in the environment and have therefore evolved mechanisms of acquiring, concentrating and recycling phosphorus. The need for efficient use of phosphorus has also prompted human populations to evolve elaborate management for maintaining phosphorus fertility in agriculture, recycle nutrients and limit phosphorus oversupply from wastes, particularly in aquatic ecosystems. Improvements in management are still needed if increasing populations are to manage their environments sustainably.

Phosphorus is the 11th most common element with an average abundance of 1200 mg kg^{-1} in the Earth's crust. Phosphorus is released from rocks to soils during weathering, and in natural ecosystems, phosphorus is cycled between inorganic and organic compounds (Figure 1) so that phosphorus concentration in the soil solutions (and therefore phosphorus availability to plants) is controlled both by physical chemical reactions such as sorption/desorption and precipitation/dissolution and by biological processes such as immobilization (uptake by plants and microorganisms) and by mineralization (decomposition of residues). Inorganic phosphorus forms are associated with calcareous compounds and amorphous and crystalline iron and aluminum oxides. Overall, soil pH is the main property controlling inorganic phosphorus forms with Ca compounds dominating phosphorus sorption in neutral and alkaline soils. In acid soils, amorphous and hydrous oxide forms of aluminium (Al), iron (Fe), and manganese (Mn) dominate phosphorus sorption processes. Organic forms include relatively labile biochemical compounds and more resistant humified materials. The forms generalized in Figure 1 are not discrete entities, as intergrades and dynamic transformations continuously occur among them.

Thus, if phosphorus were not removed from natural ecosystems, these cycling processes would ensure that phosphorus remained in the system for reuse. However, as the earth's crust weathers to form soils, over time both total phosphorus and solution phosphorus contents decrease. This change in total phosphorus is brought about mainly by a combination of plant phosphorus removal, phosphorus associated with erosion and run-off, and leaching so that extremely weathered and eroded soils may contain as little as 50 mg kg^{-1}. More commonly the range in soils is between 200 and 800 mg kg^{-1}. There is no direct correlation between total phosphorus and solution phosphorus as the latter property depends on several soil properties as illustrated in Figure 1.

At higher latitudes, soils have been regularly rejuvenated by the mixing-in of unweathered rocks during glacial periods with the most recent ice age being approximately 10–12 000 years ago. As a result, soils in these regions are more fertile than tropical soils, and continue to be highly productive despite the fact that they have supported agriculture for long periods.

On the stable land surfaces of the tropics, weathering has further advanced, and total phosphorus content has been continuously depleted by erosion and leaching, resulting in low phosphorus availability to plants. Since phosphorus limits nitrogen fixation and the uptake of other nutrients, such phosphorus-limited ecosystems are in general more fragile. Gardner (1990) estimates that release

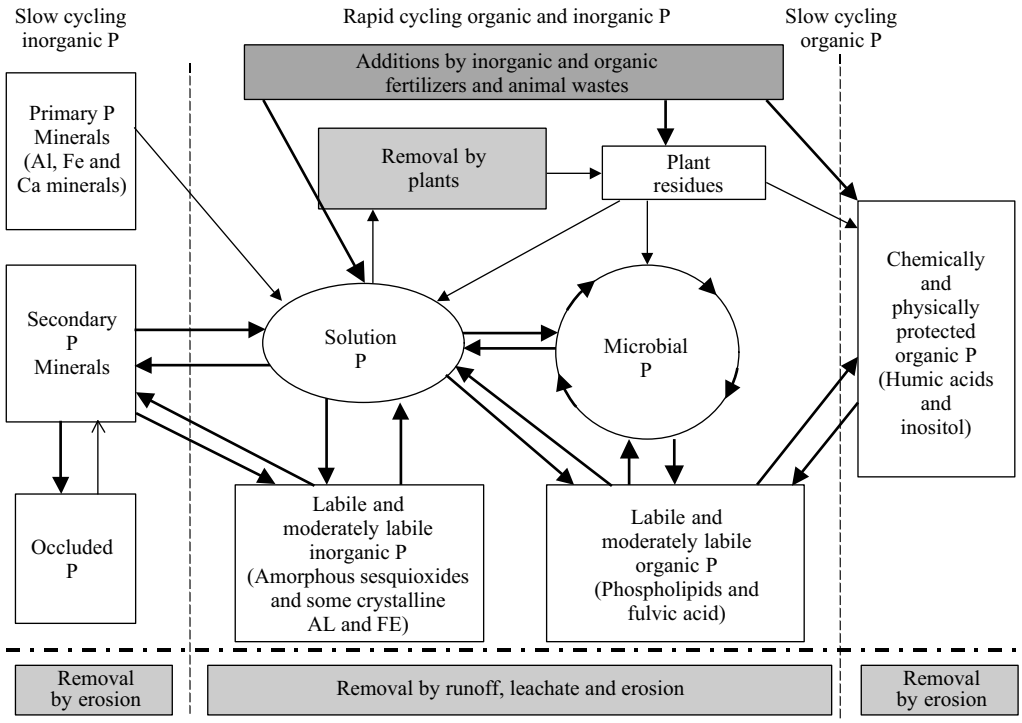


Figure 1 The soil phosphorus cycle: its components, forms and flows. (Adapted from Stewart and Tiessen 1987)

of phosphorus from rock weathering is in a range from 0–1600 moles km^2 $year^{-1}$ ranging from 163 for a temperate US site to 242 for the low-fertility Rio Negro basin and to 457 for the Amazon basin. Inputs into this ecosystem from aerial deposition is approximately 800 moles km^2 $year^{-1}$, and export from the three watersheds is 256, 820 and 4490 moles km^2 $year^{-1}$ respectively. This shows the large net export from the Amazon basin. Export is dominated by particles, with only a small portion of the phosphorus present in solution. Many of the enduring civilizations of warmer regions are associated with valleys, in which the soil resource is frequently renewed by sediments.

Farther downstream of such valleys, coastal seas receive large amounts of sediment-bound and dissolved phosphorus. Globally some 33 Tg P $year^{-1}$ are estimated to be discharged into oceans, of which some 15–20 Tg P $year^{-1}$ are carried by rivers and the remainder is coastal runoff (Howarth *et al.*, 1995). The bio-available portion of phosphorus in fluvial transport is estimated to be approximately 2 Tg $year^{-1}$. This quantity reaches estuaries and coastal seas, where it is responsible for the high productivity of near-shore waters. Water of the high oceans contains only 0.01 mg l^{-1} of phosphorus, and phosphorus is critically limiting productivity of these waters.

In addition to the chemical weathering and transport of phosphorus from land to eventually the seas, phosphorus enters into biological transformations. Biological materials concentrate phosphorus above average environmental levels, so that seeds or grains contain 400 mg kg^{-1} and mammals 25 000 mg kg^{-1} (on a dry weight basis). An example of the ability of organisms to accumulate phosphorus against a concentration gradient from their environment is the uptake of phosphorus by plants. Phosphorus concentrations in soil solution are near 0.1 mg l^{-1}, phosphorus concentrations in xylem sap, which transports nutrients from plant roots to tops, are near 100 mg l^{-1}, a concentration factor of 1000. High levels of phosphorus in the environment are therefore commonly indicators of present or past biological activity and the accumulation of biological residues. Some of the highest natural phosphorus concentrations are found in rock phosphate, which has an average phosphorus content of 144 000 mg kg^{-1}. Most deposits of rock phosphate are the product of the sedimentation and accumulation of marine organisms.

Plant and animal production therefore creates hot spots of phosphorus concentration and accumulation, and phosphorus loads in the environment are closely correlated with human occupation. A long-term record of diatom populations in the sediments of an English lake has shown the effects of 6000 years of agricultural and urban development (Anderson, 1997). The shift from slash/burn agriculture of the Neolithic period to more sedentary agricultural production in the early Bronze Age increased phosphorus loads in the lake water from 33–220 µg l^{-1} $year^{-1}$. A similar increase was observed in an Austrian lake caused by deforestation during the early Bronze Age, when phosphorus

concentrations increased from less than $10 -> 100\,\mu g\,l^{-1}$. The levels in the English lake increased further during the medieval period due to the development of towns with their effluent point sources of phosphorus. In the Canadian province of British Columbia, diatom records in lake sediments show a significant phosphorus enrichment since 1850, the time of European settlement. The advent of phosphorus fertilizers and concentrated livestock production in some areas has greatly increased phosphorus loads in surface waters.

Phosphorus inputs into fresh waters can accelerate eutrophication and impair water quality by restricting water use. Due to increased growth of algae and aquatic weeds and resulting oxygen shortages due to their decomposition, fresh water can be rendered non-drinkable and inland lakes made unusable for fisheries, recreation, and industry. Remedial action in such cases has focused on the role of phosphorus although nitrogen and carbon also are essential for the growth of aquatic biota. This is due to the difficulty in controlling the exchange of nitrogen and carbon between atmosphere and water and fixation of atmospheric nitrogen by blue–green algae (Sharpley and Rekolainen, 1997). An example of phosphorus flux in one Swedish lake watershed (Ryding *et al.*, 1990) illustrates the different components of an environmental phosphorus budget: inputs are up to 62% from fertilizer, 30% from manure, 5% from sewage, 2% atmospheric deposition and 0.3% from the natural weathering of rocks. Outputs are 93% crop exports, 6% due to arable erosion and 0.6% ground water export. Total inputs are three times greater than outputs, i.e., the watershed shows a net phosphorus retention. This is typical of many watersheds. Another example illustrates that since 1945, phosphorus has increased in rural Irish lakes by $1–2\,\mu g\,l^{-1}$ each year. This is related to increases in soil phosphorus levels by $10\,kg\,ha^{-1}$ each year from both inorganic fertilizer addition and an extensive use of animal manure. Since the mid-1970s when most point sources of sewage input to rivers were eliminated, dissolved reactive phosphorus, the most bio-available form, has increased in lakes by $1.5\,\mu g\,l^{-1}$ each year (Foy *et al.*, 1995; Withers *et al.*, 2000). This has been attributed to high livestock concentrations in neighboring farmland and extensive use of fertilizer and animal manure.

At the global level, some $10 \times 10^6\,Mg\,year^{-1}$ of phosphorus were traded and transported between world regions in the late 1980s (Beaton *et al.*, 1995), and a comparable amount was used as fertilizer. World trade movements of phosphorus were estimated to be 81% in fertilizer, rock or phosphoric acid, 15% in plant and 1% in livestock commodities. Since trade flows are uneven between world regions and since much of the traded phosphorus originated from mined deposits, these figures imply a significant enrichment of the earth's surface with phosphorus and also an important regional concentration of that phosphorus.

Soils commonly contain less phosphorus than they can hold, i.e., they contain active phosphorus sorption sites and are normally deficient in phosphorus relative to the requirements of crops. This phosphorus deficiency varies with soil type, ranging from $<1\,kg\,ha^{-1}$ in aridisols to $200\,kg\,ha^{-1}$ in oxisols and $300\,kg\,ha^{-1}$ in andosols. Oxisols cover large areas in the tropics and subtropics and account for 50% of the world phosphorus requirements for crop production. In many phosphorus-deficient tropical regions too little phosphorus is used for optimum crop production, because the high phosphorus-fixation and high phosphorus requirements together with low purchasing power of farmers in developing countries severely limit investment in fertilizers. On the other hand, too much phosphorus accumulates in regions with large and wealthy populations, which import both fertilizers and other phosphorus-containing commodities. Although most watersheds will show a net phosphorus retention even at substantial phosphorus inputs, they are leaky, and rates of phosphorus export are higher after phosphorus application than the natural levels. Exports of phosphorus from watersheds are thus related to population density: A regression performed on major rivers and their watersheds by Caraco (1995) showed loss rates of $0.3\,kg\,P\,km^{-2}$ at a population density of $0.1\,km^{-2}$, rising to $30\,kg\,P\,km^{-2}$ at $300\,km^{-2}$.

Some 10% of phosphorus export from land occurs by leaching and ground water transport, while 90% is transported by overland flow as sediment or dissolved phosphorus. Despite the small proportion of leached phosphorus, it has a greater effect on eutrophication of receiving waters because it is soluble and therefore easily available to biota. Phosphorus losses by over-land transport range from 0.1 to $10\,kg\,ha^{-1}\,year^{-1}$, or more on highly erosive sites. Not all site losses are exported from a watershed, because part of the sediment may be redeposited and part of the dissolved phosphorus may be resorbed within the watershed. On average, surface runoff waters carry $10\,\mu g\,l^{-1}$ of dissolved and $1000\,\mu g\,l^{-1}$ of sediment phosphorus (Melak, 1995). Often, sediment is phosphorus enriched relative to the source soil. While phosphorus is transported downstream, sedimentation and resuspension occur. The concomitant movement and recycling of phosphorus results in a spiraling of phosphorus as it moves down rivers. Spiraling varies along the stream and it is therefore difficult to quantify phosphorus fluxes, just as it is difficult to quantify phosphorus loads in a lake where phosphorus is released from the sediment when solution phosphorus is diminished. Best estimates (Howarth *et al.*, 1995) of river input of phosphorus to the ocean ranges from $21–30\,Tg\,year^{-1}$. The ultimate fate of most phosphorus entering the ocean is burial in sediments although a small amount of the total phosphorus is harvested in fish catch. Filippelli and Delaney (1994) used sediment core analysis to show a large decrease in the rate of phosphorus deposition in Pacific sediments over the past

five million years. The changes in sediment phosphorus were attributed to redistribution of phosphorus between high and low latitudes, to shifts between oceans and to changes in sea level. Phosphorus accumulation rates ranged from 5–60 μmoles P cm^{-2} 1000 $year^{-1}$. These figures could not be reconciled with continental weathering rates.

During these transformations and transfers, phosphorus activates biological production and biological cycles of other elements. The capacity of many ecosystems to sequester extra carbon dioxide (CO_2) depends on an adequate supply of phosphorus, and phosphorus is a key requirement for nitrogen fixation from the atmosphere by legumes. The key processes involved in the cycling of carbon, nitrogen, phosphorus and S have been the subject of many investigations (Bolin and Cook, 1983), continue to be a major part of ongoing International Geosphere–Biosphere Program (IGBP) investigations of ecosystem and atmospheric chemistry, and have been incorporated into predictive ecosystem models. Because of the interacting nature of these processes and controls, models of organic matter change and equilibrium under disturbance and management such as CENTURY (*see* **CENTURY Ecosystem Model**, Volume 2) are being applied worldwide to evaluate different management strategies.

ANTHROPOGENIC ENVIRONMENTAL CHANGE AFFECTS NATURAL PHOSPHORUS TRANSFORMATIONS AND TRANSFERS

The accelerating changes to the Earth's environment are being driven by growth in the human population, by the increasing level of resource consumption by human societies and by changes in technology and socio-political organizations (IGBP, 1997; Vitousek *et al.*, 1997). Four aspects were considered by IGBP under the term global change: (1) changes in land use and land cover; (2) the worldwide decline in biodiversity; (3) changes in atmospheric composition especially the rise in CO_2 concentration, and (4) changes in climate. From the perspective of the phosphorus cycle we know most about impact of changes in land use and cover. These can be summarized as follows:

1. Increased food demand from old, weathered, often tropical soils has reduced the fertility of these soils beyond the traditional capacity for regeneration (such as under shifting cultivation) and land degradation is evident. Phosphorus availability in these soils is very low and will not sustain a larger population without extra phosphorus inputs.
2. Fossil sedimentary phosphorus is mined and reintroduced into terrestrial production systems (principally high-input agriculture), where it activates biological systems and their element cycles.
3. Phosphorus-containing agricultural and industrial products are transported and concentrated in industrialized regions. The impact on water quality in these regions is greatest and already there are signs that several agro-ecosystems are unable to contain all the added phosphorus.
4. High-input agriculture increases phosphorus demands and speeds up phosphorus transformations and transfers.
5. Runoff and transported sedimentary phosphorus and nitrogen are reflecting increasing anthropogenic loads and are causing eutrophication of some near-shore waters.

One of the key issues for the future is the availability of phosphate ore reserves to meet the amounts of fertilizer needed to replace the phosphorus used by agricultural production. Tiessen (1995) combined available estimates and arrived at a range from <100 years to many hundreds of years of potential supply although economic and political constraints mean that at present, phosphorus remains in short supply over large parts of the globe. Soil phosphorus availability will remain a major constraint on food production. Management of fertilizer phosphorus inputs together with an understanding of organic matter cycling should allow for the development of more sustainable agriculture practices.

Coastal waters are increasingly suffering from high nutrient loads. Phosphorus is only part of the problem and management will be complex. The sensitivity of coastal regions means that prevention should be favored over damage management.

Inland lakes are affected by past fertilizer and animal waste management. More dialogue between freshwater and land use practitioners around the transfers of nutrients between systems is needed to develop regional plans to prevent phosphorus loss.

REFERENCES

Anderson, N J (1997) Reconstructing Historical Phosphorus Concentrations in Rural Lakes Using Diatom Models, in *Phosphorus Losses from Soil to Water*, eds H Tunney, O T Carton, P C Brookes, and A E Johnston, CAB International, Wallingford, 95–118.

Beaton, J D, Roberts, T L, Halstead, E H, and Cowell, L E (1995) Global Transfers of P in Fertilizer Materials and Agriculture Commodities, in *Phosphorus in the Global Environment – Transfers, Cycles and Management*, ed H Tiessen, SCOPE 54, John Wiley & Sons, Chichester, 7–26.

Bolin, B and Cook, R B, eds (1983) *Major Biogeochemical Cycles and Their Interactions*, SCOPE 21, John Wiley & Sons, Chichester.

Caraco, N F (1995) Influence of Human Populations on Phosphorus Transfers to Aquatic Systems: a regional case study using large rivers, in *Phosphorus in the Global Environment – Transfers, Cycles and Management*, ed

H Tiessen, SCOPE 54, John Wiley & Sons, Chichester, 235–244.

Filippelli, G M and Delaney, M L (1994) The Oceanic Phosphorus Cycle and Continental Weathering During the Neocene, *Paleoceanography*, **9**, 643–652.

Foy, R H, Smith, R V, Jordan, C, and Lennox, S D (1995) Upward Trend in Soluble Phosphorus Loadings to Lough Neagh Despite Phosphorus Reduction at Sewage Treatment Plants, *Water Res.*, **29**, 1051–1063.

Gardner, L R (1990) The Role of Rock Weathering in the Phosphorus Budget of Terrestrial Watersheds, *Biogeochemistry*, **11**, 97–110.

Howarth, R H, Jensen, H S, Mariano, R, and Postma, H (1995) Transport to and processing of P in Near-shore and Oceanic Waters, in *Phosphorus in the Global Environment – Transfers, Cycles and Management*, ed H Tiessen, SCOPE 54, John Wiley & Sons, Chichester, 321–345.

IGBP (1997) *The Terrestrial Biosphere and Global Change: Implications for Natural and Managed Ecosystems. A Synthesis of GCTE and Related Research*, eds B Walker and W Steffen, IGBP, Stockholm.

Melak, J M (1995) Transport and Transformations of P, Fluvial and Lacustrine Ecosystems, in *Phosphorus in the Global Environment – Transfers, Cycles and Management*, ed H Tiessen, SCOPE 54, John Wiley & Sons, Chichester, 245–254.

Ryding, S O, Enell, M, and Wennberg, L (1990) Swedish Agricultural Non-point Source Pollution: a Summary of Research and Findings, *Lakes Reserv. Manage.*, **6**, 207–217.

Sharpley, A N and Rekolainen, S (1997) Phosphorus in Agriculture and its Agricultural Implications, in *Phosphorus Loss from Soil to Water*, eds H Tunney, O T Carton, P C Brookes, and A E Johnston, CAB Int., Cambridge, 1–54.

Stewart, J W B and Tiessen, H (1987) Dynamics of Soil Organic Phosphorus, *Biogeochemistry*, **4**, 41–60.

Tiessen, H, ed (1995) *Phosphorus in the Global Environment – Transfers, Cycles and Management*, SCOPE 54, John Wiley & Sons, Chichester.

Vitousek, P M, Mooney, H A, Lubchenco, J, and Melillo, J M (1997) Human Domination of Earth's Ecosystems, *Science*, **277**, 494–499.

Withers, P J A, Davidson, I A, and Foy, R H (2000) Prospects for Controlling Non-point Phosphorus Loss to Water: a UK perspective, *J. Environ. Qual.*, **29**, 167–175.

Photosynthesis

Photosynthesis is the use of energy from light to fix carbon dioxide (CO_2), reduce it, and convert it to organic carbon compounds. The process has three basic steps: the trapping of light energy and its conversion to chemical potential (the light reaction), the fixation of CO_2, and the biochemical reaction sequences that reduce the fixed carbon (the dark reactions).

In the light reaction, photons are absorbed by the photosynthetic pigments, the most important of which are the chlorophylls, although other pigments are involved. Energy from the absorbed photons is channeled by the pigments to oxidize water to oxygen and manufacture high-energy intermediate compounds. These products of the light reaction are used in the dark reactions to power a complex cyclic reaction sequence, called C_3 photosynthesis because the first product of fixation is a C_3 acid, which is ultimately converted to sugars, the normal end products of photosynthesis.

This reaction sequence has a serious problem: it evolved at a time when atmospheric oxygen levels were much lower than today and CO_2 was much higher. Ribulose bisphosphate carboxylase/oxygenase (Rubisco), the carboxylating enzyme of the C_3 cycle, is unable to distinguish adequately between oxygen and carbon dioxide. Oxygen acts competitively, and the oxygenase reaction leads to metabolically useless products that must be recycled and in part converted back to CO_2, a wasteful process, by the reactions of photorespiration. In today's conditions of high (20%) oxygen and low (0.04%) CO_2 the oxygenase function is sufficiently active to be strongly disadvantageous and greatly reduces photosynthetic capacity of the so-called C_3 plants.

An obvious experimental solution to this problem is to grow plants in a low oxygen, high CO_2 atmosphere. However, a natural solution is provided by an alternative reaction cycle called C_4 photosynthesis because the first product of carboxylation is a C_4 acid. In this cycle CO_2 (as bicarbonate) is fixed to a C_3 compound in the mesophyll cells (near the leaf surface) by the enzyme PEP carboxylase. The product of this reaction is converted to the C_4 acids aspartate (by transamination) or malate (by reduction), which are transported to an inner region of the leaf, the bundle sheath. There they are decarboxylated to produce CO_2, which is fixed by the normal C_3 cycle. The C_3 compound left after the decarboxylation is shunted back to the mesophyll, using energy from products of the light reaction. The C_4 cycle functions in effect to pump CO_2 to an inner compartment of the leaf. This raises the CO_2 concentration at the internal site of C_3 fixation where oxygen (which penetrates only slowly by diffusion) is well below atmospheric levels. The result is that the carboxylase function of Rubisco is greatly enhanced and the oxygenase function reduced, with a corresponding increase in the rate of photosynthetic production and decrease in losses due to photorespiration.

There are further consequences of C_4 photosynthesis. PEP carboxylase has a much higher affinity for CO_2 (as bicarbonate) than Rubisco, so C_4 plants can function well at much lower CO_2 concentrations. An important ecological result is that in the event of water shortage C_4 plants can continue photosynthesis with nearly closed stomata (through which water and CO_2 exchange of the leaf

occurs). This allows continued active growth with greatly reduced water loss, so C_4 plants survive well in hot, dry climates. The downside of C_4 photosynthesis is that substantial light energy is used to run the C_4 cycle, which in itself is unproductive. However, at high illumination, when C_3 plants are often embarrassed by excess light because not enough CO_2 can be fixed, C_4 plants can operate at full speed. The advantage of C_4 photosynthesis is not, therefore, that it is more efficient, but that it goes faster and continues under drought conditions when C_3 plants would be shut down to conserve water. C_4 plants that transport aspartate (aspartate transporters, e.g., *Atriplex* spp.) as the C_4 acid are slightly more energy efficient than malate transporters (e.g., sorghum, maize), but require a substantially increased metabolism and supply of nitrogen and therefore, under high light conditions, are at an ecological disadvantage.

Crassulacean acid metabolism (CAM) is an interesting variation of C_4 photosynthesis in which the C_3 and C_4 carboxylations are separated in time rather than in location. Many desert succulent plants and cacti use CAM to fix CO_2 into acids during the night (a process that does not require light energy). During the day the C_4 acids are decarboxylated and the released CO_2 is fixed by normal C_3 photosynthesis. Thus, CAM plants absorb CO_2 through their stomata at night (when the air is more humid and water loss not so serious) and carry on photosynthesis behind closed stomata during the day, when open stomata would result in deadly water loss. In this way they can survive and grow in extreme drought conditions.

C_4 plants tend to thrive best in dry tropical or subtropical grasslands. They include some major crops (e.g., sorghum, maize) as well as some pestilent weeds (crabgrass, goosefoot). They are not usually competitive where light intensities are lower and water plentiful as in temperate regions, although some successful alpine and tundra plants are C_4, perhaps because of the reduced availability of water due to freezing. The consequences of global warming – higher temperatures and greater areas of drought – suggest that efforts in selection, breeding or genetic modification to create new C_4 crops may be valuable, provided that cloud cover does not reduce light below levels where C_4 photosynthesis can increase production.

See also: **C_3 and C_4 Photosynthesis**, Volume 2.

R G S BIDWELL *Canada*

Photosynthetically Active Radiation (PAR)

see PAR (Photosynthetically Active Radiation) (**Volume 2**)

Phytophagy

Phytophagy means the eating of plants. Phytophagous animals (including many insects, seed-eating birds, and mammals such as most ungulates and rodents) are vegetarians. Their importance in communities has many facets, from controlling plant growth to seed dispersal and beyond. The ecological niches (*see* **Niche**, Volume 2) of phytophagous animals and insects are strongly controlled by the ecological relationships of their preferred food, and their niches vary in breadth in proportion to the niche breadth of their diet. Darwin's finches probably provide the best-known example of the relationship between phytophagous food preference and evolutionary adaptation in the development of ecological niches.

It is apparent that many phytophagous insects have coevolved with their specific hosts to the point where remarkably high degrees of specificity exist in some animal–host relationships. As with species diversity in general, the absolute frequency of such highly specific relationships is greatest in relatively stable tropical ecosystems and decreases toward the boreal regions.

R G S BIDWELL *Canada*

Phytotron

see Controlled Environment Facilities in Global Change Research (**Volume 2**)

Plant Competition in an Elevated CO_2 World

S Catovsky and F A Bazzaz
Harvard University, Cambridge, MA, USA

Elevated carbon dioxide (CO_2) influences the way plants interact with one another and these effects can have major consequences for natural ecosystems in the future. Plant competition, in particular, is a prevalent force structuring many processes within terrestrial ecosystems. In this paper, we consider how rising CO_2 levels may influence competitive interactions between plants and how these changes may scale up to affect the dynamics of populations, communities and ecosystems.

Elevated CO_2 can affect interactions between individuals of the same plant species (intraspecific). We address how elevated CO_2 influences the development of dense stands of plants and their future dynamics. Contrary to expectations from the addition of other resources to plant populations, elevated CO_2 seems to favor growth of smaller plants within a stand and thus reduce overall size differences between dominant and subordinate plants. These changes in the performance of different sized individuals within a population may also have implications for changes in the genetic composition of a population. Elevated CO_2 increases overall plant growth and hence tends to speed up competition between neighboring plants. As plants compete more strongly with one another, there is stronger selection on individual genotypes and the potential for evolutionary change is likely to increase. Based on current evidence, however, elevated CO_2 appears not to favor those plants whose individual growth is most strongly increased as CO_2 levels rise, but rather those that are better overall competitors in the stand. In addition, these CO_2 induced changes in stand development can have further consequences for population dynamics in subsequent generations. To date, however, our current knowledge about these future patterns of population development is rather limited.

Understanding competitive interactions between plants of different species (interspecific) can additionally refine our ability to predict future changes in community species composition. As we found for populations, information from individually grown plants may be inappropriate for accurately predicting dynamics at higher levels of organization. Many studies have now shown that plants grown individually often respond to elevated CO_2 differently than plants grown in mixed species stands. In addition, the direction of change is not consistent, with some studies finding that competition reduces species' responses to elevated CO_2 and others finding an enhancement of response. Our capability for predicting future community change will improve as we develop a mechanistic understanding of the nature of competition between different plant species. Plant competition can even influence CO_2 effects on ecosystem productivity. To date, all the evidence suggests that the expected amount of carbon taken up by terrestrial ecosystems is less if we use predictions from plant competition experiments than if we use CO_2 responses of individually grown plants.

The importance of plant competition at many levels of biological organization clearly warrants its incorporation into models of potential ecosystem responses to elevated CO_2. As we improve our understanding of the mechanisms underlying plant competitive interactions, we should be able to develop more complex and more accurate models of our future world.

IMPORTANCE OF PLANT COMPETITION

The central role of competition in structuring many ecological processes has long been recognized. There has been much debate about the precise meaning of plant competition but most agree on the basic tenants of the process. In this paper, we take plant competition to refer to an interaction in which individuals are drawing from a shared resource, the result of which limits their performance. Plant competition is an important component in our understanding of ecological interactions, because the sessile habit of plants means that they may spend much of their life next to other plants. Neighbors must continually compete with each other for limiting resources, and these competitive interactions influence processes at many levels of biological organization. Plants modify the distribution and abundance of local resources, which often affect the growth, survival, and reproduction of neighboring individuals. Alteration of individual plant function may then scale up to influence the dynamics of populations, communities, and ecosystems.

Given the importance of plant competitive interactions at many ecological levels, it is critical that we incorporate competition into our understanding of how global environmental change may influence terrestrial ecosystem structure and function. Rising levels of atmospheric CO_2 are likely to be the most direct environmental perturbation affecting terrestrial ecosystems in the future (*see* **Plants – from Cells to Ecosystems: Impacts of Global Environmental Change**, Volume 2). As a result, much ecological research has concentrated on determining potential influences of elevated atmospheric CO_2 on plant and ecosystem function. In this paper, we consider how elevated CO_2 influences plant competitive interactions, and what consequences these effects have for the dynamics of plant populations, communities, and ecosystems (Figure 1). We argue that incorporating competition into conceptual models of the effects of global change on ecological processes may change many predictions derived from information on individually grown plants. We focus our discussion on elevated CO_2 but consider other environmental perturbations (e.g., rising temperature, increased nitrogen deposition) when differences in response arise. Unfortunately, relatively little information currently exists on effects of other global changes on plant competition.

WHY MIGHT ELEVATED CO_2 INFLUENCE COMPETITIVE INTERACTIONS?

Increased atmospheric CO_2 may alter the way neighbors interact with one another by differentially affecting neighbor performance. CO_2-induced changes in individual growth rates will clearly alter competitive interactions, but so will changes in patterns of allocation and the timing of growth processes. It is now well established that elevated CO_2 often differentially influences the performances

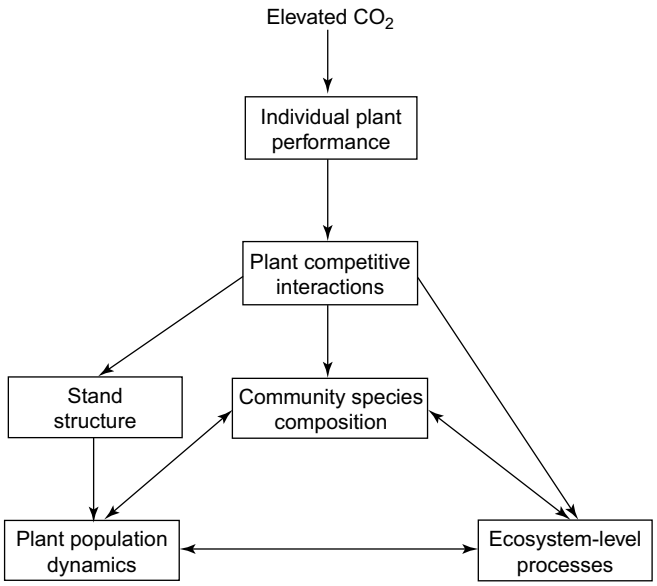


Figure 1 Schematic diagram for the effect of global environmental change on plant competitive interactions and their consequences for higher levels of organization

of genotypes within a population, and species within a community (Bazzaz, 1990). Alteration of patterns of growth and development may modify the way different genotypes or different species compete with one another, as different combinations of limiting resources (nutrients, water, and light) now affect plant performance. In addition, the presence of neighbors changes local conditions around an individual. These changes may directly influence plant responses to elevated CO_2, because the effects of elevated CO_2 on plants are commonly contingent on growth conditions (*see* **Plants – from Cells to Ecosystems: Impacts of Global Environmental Change**, Volume 2). When limiting resources are directional, and thus preemptable, competitive interactions are usually asymmetric. In these situations, small CO_2 induced changes in plant performance may be further magnified through time, potentially creating large differences in growth and development between genotypes or species.

ELEVATED CO_2 AND INTRASPECIFIC INTERACTIONS IN PLANT POPULATIONS

Competitive interactions between individuals within a population (intraspecific) often determine the nature of stand development, and ultimately, the future dynamics of the population. Many species that regenerate following a large disturbance (e.g., early successional species) often form dense, largely monospecific stands in the early stages of their life cycles. Consequently, early stand dynamics are frequently dominated by intense intraspecific competition. It is critical to understand how plant competitive interactions influence the development of these stands, as the outcome of competition determines which individuals reach reproductive maturity and thus contribute to subsequent generations. Ultimately, these early intraspecific interactions may influence the nature of selective forces shaping the population and the direction of evolutionary change. If rising atmospheric CO_2 concentrations influence intraspecific competitive interactions, then current patterns of population development will change in the future. We consider how elevated CO_2 may affect three distinct aspects of plant populations: stand size structure, population genetic structure, and intergenerational population dynamics.

Stand Size Structure

Many factors may contribute to the development of size hierarchies in plant populations, including patterns of seed dispersal, timing of germination, genetic variation in seed size and initial growth rates, fine-scale heterogeneity, and intraspecific competitive interactions. The importance of these different factors in stand development will vary depending on the species involved and the conditions in which the stand is growing. To date, the few studies that have addressed the impact of global changes on stand size structure have carried out the experiments in relatively controlled conditions. By their nature, such experiments emphasize more predictable processes (e.g., plant competition) over more random processes (e.g., small-scale heterogeneity and population variation). From basic ecological principles, we might predict that global environmental changes stimulating plant growth would increase the magnitude of size hierarchies in plant populations by intensifying intraspecific competitive interactions. The limited evidence we have so far, however, suggests that the influence of global changes on population size structure depends on how these changes affect plant responses to their local environment.

Wayne and Bazzaz (1997) grew competitive stands of yellow birch (*Betula alleghaniensis*) for two years and investigated the effects of atmospheric CO_2 concentration on stand size structure. They found that elevated CO_2 slightly, but consistently, reduced overall stand size inequality (Figure 2). As elevated CO_2 commonly stimulates plant growth, their results contrast with our original predictions. Wayne and Bazzaz provide evidence that size inequality was reduced in these stands because elevated CO_2 altered the way seedlings competed with one another for light. They suggest that these changes were, in part, brought about through greater CO_2 responsiveness in subordinate plants than in canopy dominants, perhaps because subordinates experienced less degree of photosynthetic acclimation in elevated CO_2 than did dominants. This acclimation is often observed in elevated CO_2 experiments, and is thought to occur because of starch accumulation in leaves and/or

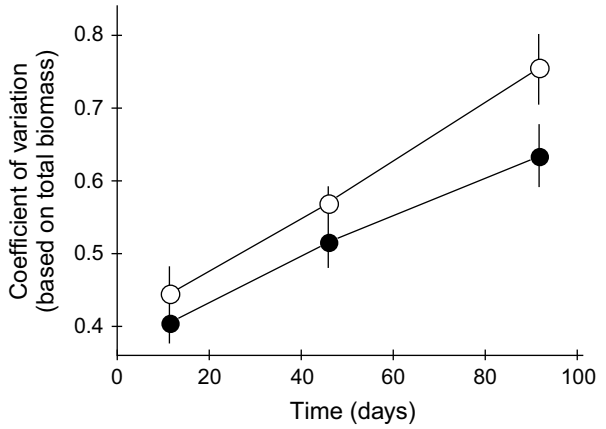


Figure 2 Stand size inequality (measured by coef cient of variation) for *Betula alleghaniensis* seedlings grown at ambient (○, 350 μl l^{-1}) and elevated (●, 700 μl l^{-1}) CO_2 at three times through the growing season (modi ed from Wayne and Bazzaz, 1997)

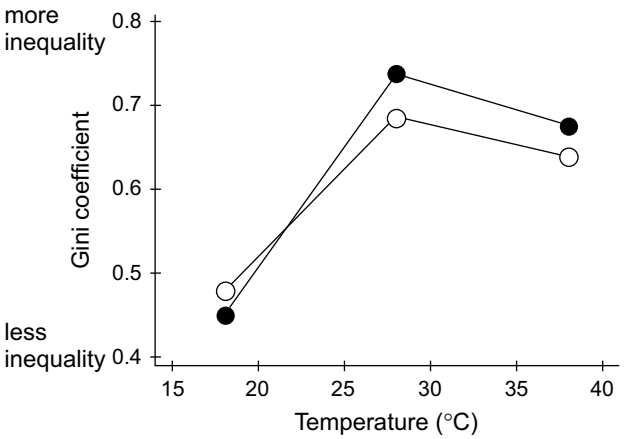


Figure 3 Stand size inequality measured by Gini coef cient, which measures the degree of size inequality in a population. It equals 0 when all individuals in a population are of equivalent biomass, and approaches 1 as the difference in biomass between individuals increases) for *Amaranthus retro exus* populations after 80 days of growth at ambient (○, 400 μl l^{-1}) and elevated (●, 700 μl l^{-1}) CO_2 and under three temperature regimes (data from Morse and Bazzaz, 1994)

direct inhibition of photosynthesis by low levels of inorganic phosphate (*see* **Plants – from Cells to Ecosystems: Impacts of Global Environmental Change**, Volume 2).

The only other study that has addressed the consequences of rising CO_2 levels for stand size structure found little effect of elevated CO_2 on population size hierarchies in two annual plant species (Figure 3; only *Amaranthus retroflexus* is shown) (Morse and Bazzaz, 1994). In contrast, they found very pronounced effects of temperature on size structure. For both species, stand size inequality increased significantly with increasing temperature, reaching a peak at 28 °C. By following stands through their development, Morse and Bazzaz ascertained that size hierarchies were magnified with increasing temperature, because greater initial recruitment into the population (higher germination rates) and faster growth rates intensified intraspecific competitive interactions. The increase in the strength of competition between individuals led to larger differentials in growth and survival between canopy dominants and subordinates. This experiment highlights how different global environmental changes might have contrasting effects on plant competition. These differences likely arise from the contrasting ways global change factors influence plant responses to other resources. For example, elevated CO_2 differentially improves plant growth at lower light levels, reducing the ability of dominants to outcompete subordinate plants for light, while elevated temperatures act to speed up competitive processes and resource preemption. Effects of nitrogen deposition on stand development are also relatively poorly understood, but from early agricultural experiments adding fertilizer to monospecific stands of crop species, it is clear that supplementing below-ground resources intensifies asymmetric competition for light above-ground and results in heightened stand size inequality.

Population Genetic Structure

The underlying genetic component in the formation of size hierarchies suggests that the dynamics of stand development may have evolutionary consequences. If we wish to ascertain whether elevated CO_2 will act as a future selective force for plant populations, we need to determine how rising CO_2 levels will influence the genetic structure of plant populations through time. As genetic variation for CO_2 responsiveness has been recorded in a number of experimental studies, the potential for evolutionary responses to elevated CO_2 certainly exists (Körner and Bazzaz, 1996). The extent to which variation for fitness-correlated traits under elevated CO_2 is heritable is, however, still unclear. In addition, this potential for evolutionary change might not be reached in natural plant populations, where biotic interactions and other abiotic factors may alter plants' responses to elevated CO_2. Thus, selection may not necessarily act to promote increased CO_2 responsiveness in the future. Competitive interactions, in particular, are likely to influence genotypic responses to the environment.

To date, only two studies (Bazzaz *et al.*, 1995; Wayne and Bazzaz, 1995) have compared the genetic changes that plant populations might undergo when exposed to elevated CO_2 as individuals and under crowded stand conditions. Despite using very different plant species, these two studies reached very similar conclusions, suggesting that the results might be more widely applicable. Wayne and Bazzaz (1995) tested their hypotheses on stands of *Betula alleghaniensis*, a mid-successional, deciduous

tree species, while Bazzaz *et al.* (1995) compared these *Betula* responses with those of *Abutilon theophrasti*, an old-field annual species. In both these studies, increasing the intensity of intraspecific competition by manipulating plant density led to greater variance among genotypes in CO_2 responsiveness (eight-fold increase for *Abutilon*, three-fold for *Betula*). This result suggests that intraspecific competition increases the potential for natural selection to act on responsiveness to elevated CO_2 in the future. This potential, however, is unlikely to be reached in stands of natural populations, as individuals favored in elevated CO_2 in high density stands were not those that showed inherently greater CO_2 responsiveness when grown at low density. In *Betula*, there was a complete reversal in ranking of genotype-specific growth enhancements in elevated CO_2 (Figure 4). Similarly, for *Abutilon* populations, there was a poor correlation between genotype-specific growth responses to elevated CO_2 at the two densities.

These results demonstrate that, although there may be strong potential for evolutionary responses to rising CO_2 levels, selection for increased CO_2 responsiveness in the future may be weak, especially in crowded populations. As competitive interactions intensify with elevated CO_2, selection may favor traits that enhance an individual's competitive ability in dense stands rather than its inherent CO_2 responsiveness, *per se*. These traits are likely to include those that improve an individual's capability to obtain other limiting resources in their environment, e.g., increased allocation to stem biomass when competing for light, and increased amount and activity of fine root biomass when competing for soil nutrients.

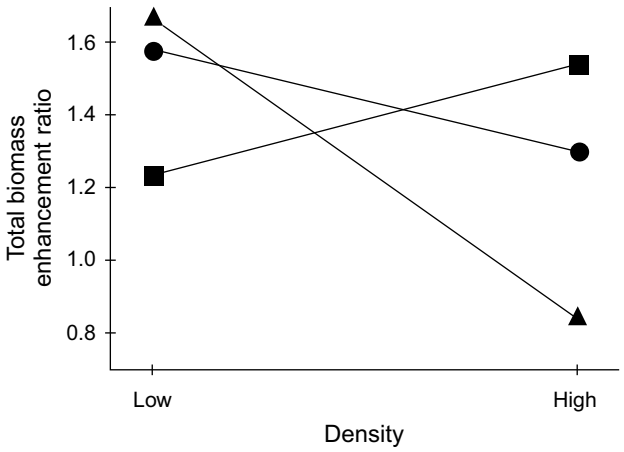


Figure 4 Total biomass enhancement ratio (biomass at 700 µl l^{-1}/biomass at 350 µl l^{-1}) for three maternal families (●, ▲, ■) of *Betula alleghaniensis* seedlings grown together in low and high density stands (data from Wayne and Bazzaz, 1995)

Inter-generational Population Dynamics

Patterns of stand development may also influence the future dynamics of plant populations. Despite the importance of this longer-term perspective, most studies use total biomass as their response variable, while few studies have directly considered the influence of elevated CO_2 on many population-level processes, such as recruitment, survival and reproduction. Although there is a reasonably good positive relationship between plant biomass and total reproduction, there is significant variation between individuals due to genotypic differences and phenotypic plasticity. This relationship may become further decoupled in the future as rising CO_2 concentrations alter patterns of allocation to flowers, fruits and seeds. In a survey of eight CO_2 studies, Ackerly and Bazzaz (1995) found that there was an overall positive relationship between enhancement of total and reproductive biomass, but that this relationship only explained 32% of the variation in response. In one-third of cases, even the direction of change in elevated CO_2 differed between total and reproductive biomass.

Of the few studies that do directly address the influence of elevated CO_2 on reproductive output, even fewer have considered interactions between CO_2 concentration and density. Given the ubiquity of competitive interactions in natural systems, understanding these effects is critical for predicting future patterns of population change. In another study on *Abutilon theophrasti*, Bazzaz *et al.* (1992) recorded significant effects of elevated CO_2 on several population level parameters, and they found that these CO_2 effects were highly dependent on population density. Elevated CO_2 increased population survivorship and the probability of flowering for surviving individuals, but only in the high-density treatments. Using a simple population dynamics model, Bazzaz *et al.* (1992) showed that these interactions between CO_2 and density effects would have significant impacts on the future dynamics of these populations. The higher fecundity of high-density populations grown at elevated CO_2 increased the magnitude of future population cycles. From this single study addressing effects of interactions between CO_2 and density on population parameters, it is challenging to make any generalizations. We can clearly see, however, that there is the potential for substantial effects of elevated CO_2 on population dynamics, especially under competitive conditions. These changes may be magnified over subsequent generations, as effects accumulate.

ELEVATED CO_2 AND INTERSPECIFIC INTERACTIONS IN PLANT COMMUNITIES

Scaling from Individual Responses to the Community

Understanding the influence of elevated CO_2 on ecosystem structure and function is a central goal of global

change biology. Community species composition is a key component of ecosystem structure and, as a result, has been the focus of much research. A large proportion of such research has focused on how species differ in their responses to rising CO_2 levels, and how differential responsiveness may lead to altered community composition. The vast numbers of species that exist on the Earth make it unfeasible to directly assess how each species may respond to elevated CO_2 in the future. Instead, ecologists are attempting to distinguish groups of species (and their related traits) that respond in similar ways to elevated CO_2, i.e., functional response groups. Only very few such groups that respond in predictable ways to rising CO_2 levels have been identified, e.g., species that differ from one another in photosynthetic pathway (e.g., C_3 vs. C_4 plants). In general, fully characterizing species' CO_2 responsiveness based on a limited set of traits has been difficult, and to date, we do not have good predictive abilities for determining species' responses in the future (*see* **Plants – from Cells to Ecosystems: Impacts of Global Environmental Change**, Volume 2).

As CO_2 research technology has developed, our ability to conduct larger scale, longer term, and more complex experiments has improved. Over the past decade, there has been a concerted effort to identify key global ecosystems and determine how they might be influenced by rising CO_2 levels. As a result, community level changes in response to elevated CO_2 have been documented in numerous studies (Körner and Bazzaz, 1996). These experiments have produced a wide range of results, from relatively little alteration of community species composition (e.g., Navas *et al.*, 1995), to extensive change in relative species abundance within the community (e.g., Potvin and Vasseur, 1997). Much of the research is still phenomenological, simply documenting the nature of change, rather than investigating potential underlying mechanisms. In a number of these experiments, community level change did not follow predictions based on functional groupings of species (see chapters in Körner and Bazzaz, 1996), but the overriding reasons leading to this divergence from predicted patterns have been unclear to date. Some progress in elucidating mechanisms of community change has been made by taking a resource-based approach. Studies that have come closest to explaining the underlying mechanisms for CO_2-induced changes in community structure have been those monitoring changes in resource availability under different CO_2 scenarios and documenting differential resource use patterns for different species, e.g., water availability in a Californian annual grassland (Chiariello and Field, 1996), and nitrogen uptake in disturbed seasonal grasslands from southwestern USA (Berntson *et al.*, 1998).

What information do we need to make predictions about future patterns of community change? Clearly, we need to understand the underlying physiological mechanisms behind individual species' responses to elevated CO_2. Even with this comprehensive information, we still face difficulties in scaling individual responses to community-level phenomena. Mechanisms for community change involve many interactions beyond individual species' responses that need to be incorporated into our conceptual models of global change. In this section, we focus on the importance of interspecific competitive interactions for determining future community patterns with rising CO_2 levels.

Importance of Interspecific Interactions

Competition between plant species (interspecific) is a ubiquitous process within most natural ecosystems and plays a central role in shaping the structure and dynamics of plant communities. The outcome of interspecific competition, however, is often difficult to predict *a priori*. The preemptive nature of plant competitive interactions means that quite small differences between individuals can influence which individual achieves dominance. Even when differential species' responsiveness to elevated CO_2 is clear, experiments with plants grown under competitive conditions may yield unexpected results. For example, Bazzaz *et al.* (1989) grew plants of *Abutilon theophrasti* (a C_3 annual) and *Amaranthus retroflexus* (a C_4 annual), both individually and in competition with one another. They found that, contrary to expectations, elevated CO_2 shifted the competitive balance toward the C_4 species in mixed stands. Growth analysis of the plants through the experiment suggested that *Amaranthus* responded earlier to elevated CO_2, and that this difference in the timing of response enabled the C_4 plant to compensate for the initial smaller size of its seedlings through increased allocation to leaves (higher leaf area ratio) in elevated CO_2.

The multitude of processes involved in interspecific interactions makes it challenging to predict community responses to elevated CO_2. Scaling from individual plant responses may only be possible with extensive information on behavior at many stages of a plant's lifecycle. Even then, plant responses to elevated CO_2 are rarely the same when grown individually and in competitive stands. As yet, we cannot determine the direction of change, as the range of responses are highly varied. In one of the first studies addressing this phenomenon, Williams *et al.* (1988) observed that, for herbaceous communities from California serpentine grassland, competition significantly damped the effects of elevated CO_2. They found that, as individuals, two species (*Agoseris heterophylla* and *Layia platyglossa*) showed significant CO_2 induced growth increases, while there were no such increases for any species in competitive stands. Similarly, when grown individually, sweetgum (*Liquidambar styraciflua*) seedlings responded

strongly to elevated CO_2 under dry soil conditions, while loblolly pine (*Pinus taeda*) seedlings showed no such increases. When grown together in mixed species, dry soil stands, however, seedlings of both species showed no substantial CO_2 induced growth enhancements (Groninger *et al.*, 1995). Competition treatments, however, do not always restrain CO_2 growth enhancements. For tropical tree saplings, Reekie and Bazzaz (1989) observed much larger differences in CO_2 responsiveness for plants grown in competitive stands than for individually grown plants

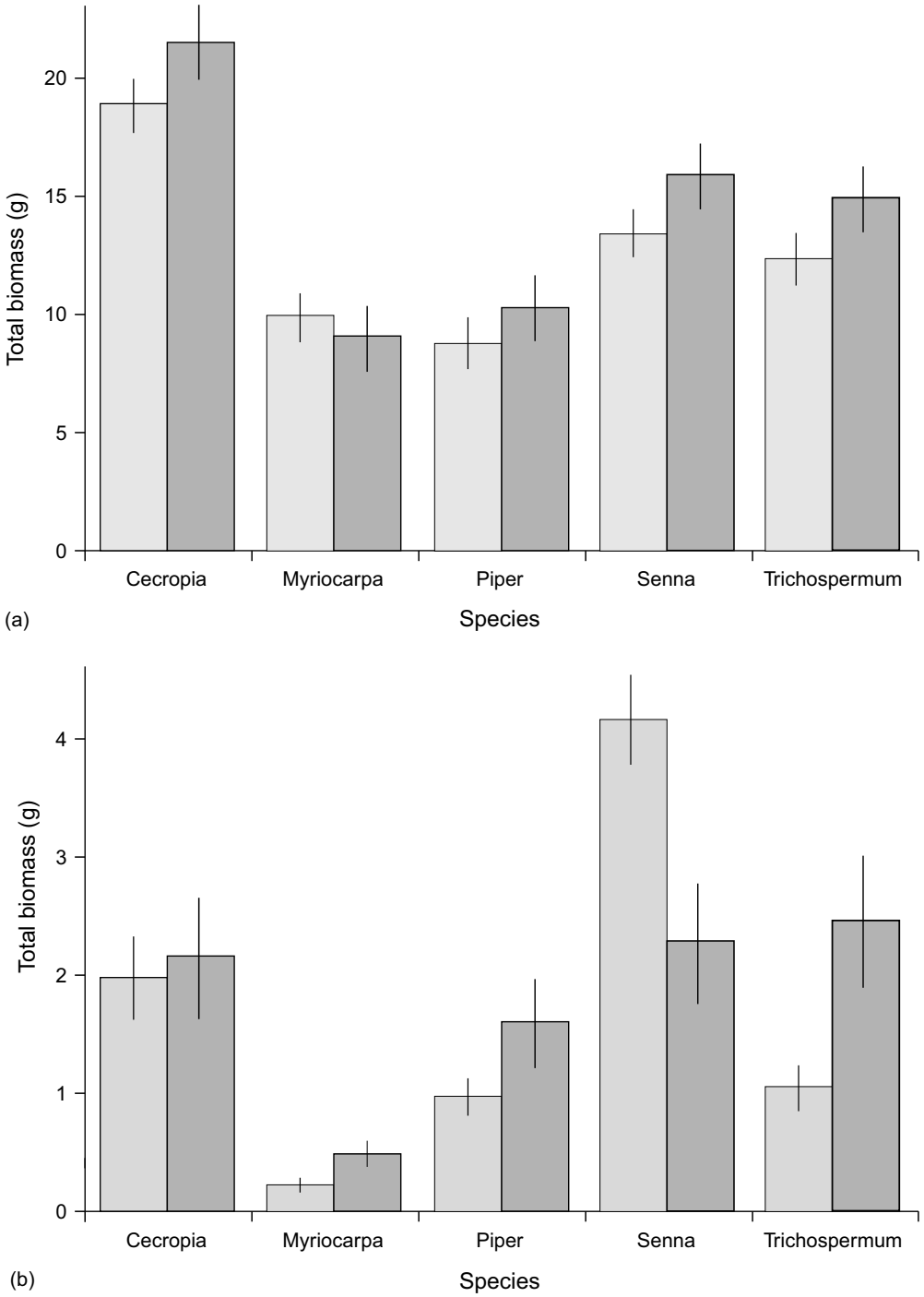


Figure 5 Total biomass for ve tropical tree saplings after one growing season at ambient (lighter box, 350 µl l^{-1}) and elevated (darker box, 700 µl l^{-1}) CO_2 and under two different competition regimes, i.e., (a) individually grown and (b) competitively grown (modi ed from Reekie and Bazzaz, 1989)

(Figure 5). In particular, in mixed species stands grown in elevated CO_2, they recorded substantial increases in *Trichospermum mexicanum*, *Piper auritum* and *Myriocarpa longipes*, mostly at the expense of *Senna multijuga*, while no significant CO_2 effects were found for individually grown plants. More recently, for two temperate forest birch species (*Betula alleghaniensis* and *Betula papyrifera*), Catovsky and Bazzaz (1999) found that competition magnified species' responses to CO_2 and water treatments, when compared with individually grown plants (Figure 6). *Betula papyrifera* showed substantially larger CO_2 growth enhancements in dry soil treatments, and these effects were more pronounced when the seedlings were grown in stands than when they were grown individually. Similarly, *Betula alleghaniensis* showed larger CO_2 growth enhancements on moister soils, but particularly when grown in competition.

Towards a Mechanistic Understanding of Interspeci c Competition

Why do we often find such disparate CO_2 responses between individually and competitively grown plants? Within populations, we have already seen that elevated CO_2 favors different phenotypic traits for plants grown individually and in competitive stands. In a similar way, for interspecific interactions, we suggest that species' traits correlated with improved performance in elevated CO_2 will differ for individually and competitively grown plants. In the competition experiment with tropical saplings described above, Reekie and Bazzaz (1989) found that competitive success in elevated CO_2 was highly correlated with increased canopy height (through increased allocation of biomass to stems). They hypothesized that rising CO_2 levels favored species that were better competitors for light in this system. Understanding which resources are most limiting within a stand may be the key to predicting competitive outcomes. Conceptual models from resource allocation theory may improve our predictive abilities for competitive mixtures (Reynolds, 1996). For example, when soil fertility is high, elevated CO_2 may intensify competition for light and favor plants with higher allocation to and activity of above-ground parts. Alternatively, at lower soil resource availability, root competition may be more intense, and plants dominating in elevated CO_2 may be those that are better competitors for below-ground resources. Naturally, many of these predictions will be highly density dependent.

As yet, few studies have addressed the interplay between root and shoot competition in elevated CO_2, and so we have few direct experimental tests of these resource-based models. Two recent studies have focused on elevated CO_2 effects on root competition and root growth in different resource combinations. Jifon *et al.* (1995) grew monocultures and mixed stands of loblolly pine (*Pinus taeda*) and sweetgum (*Liquidambar styraciflua*) under different CO_2–water–nitrogen combinations, and investigated variation in root length density (RLD) in these stands. In monocultures, they found that elevated CO_2 increased RLD most under low water, low nitrogen conditions, while in mixed species stands, the largest CO_2-induced increases in RLD were observed under the high nitrogen, high water treatment (Figure 7). This finding suggests that, contrary to our predictions, competition for

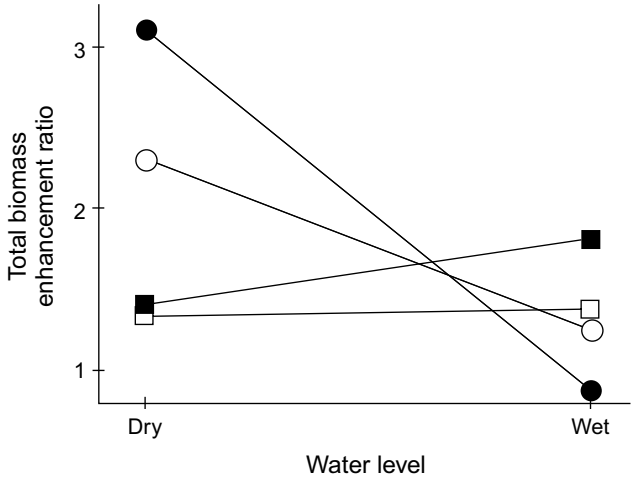


Figure 6 Total biomass enhancement ratio (biomass at 700 µl l^{-1}/biomass at 375 µl l^{-1}) for *Betula papyrifera* (○, ●) and *Betula alleghaniensis* (□, ■) seedlings grown individually (○, □) and in mixed competitive stands (●, ■) for one growing season (data from Catovsky and Bazzaz, 1999)

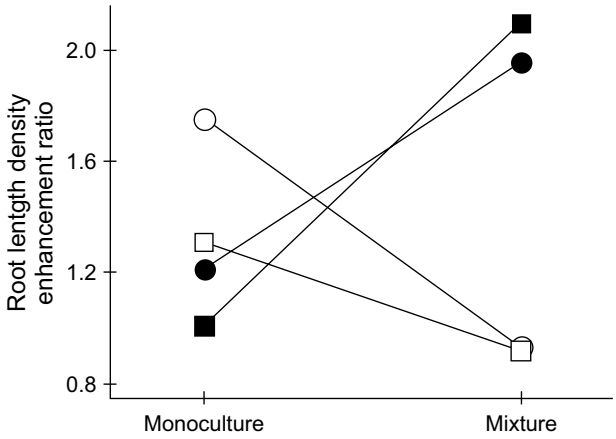


Figure 7 RLD enhancement ratio (RLD at 750 µl l^{-1}/RLD at 350 µl l^{-1}) for *Liquidambar styraci ua* (○, ●) and *Pinus taeda* (□, ■) seedlings grown with low (○, □) or high (●, ■) soil resources under two competition regimes (data from Jifon *et al.*, 1995)

light is not always favored in mixed stands growing in rich soil resource conditions. We should note, however, that Jifon *et al.* (1995) did not measure competitive interactions directly and had no individually grown plants in their experiment, which would provide a control for our hypotheses. In contrast, Arnone and Kestenholz (1997) did not vary resource levels, but did directly manipulate below-ground interspecific competitive interactions. They investigated competitive interactions between the annual ruderal species, *Linum usitatissimum*, and the annual weed, *Silene cretica*. Arnone and Kestenholz found that in competitive mixtures, *Silene* responded strongly to elevated CO_2, while *Linum* biomass was not significantly increased. When vertical soil partitions were placed between individuals in these mixtures, however, *Silene* no longer inhibited *Linum* responses to elevated CO_2 (Figure 8). This experiment demonstrates the potential importance of root competition in influencing species' responses to rising CO_2 levels, but we still do not know how the relative importance of root versus shoot competition would change under different resource scenarios. This information is critical for making accurate predictions about the outcome of competition in future elevated CO_2 concentrations.

Even if we were able to use species' traits to predict the outcome of competition under particular conditions, we might not be correct, as direct CO_2 effects on plant development may further complicate the result. In many of the experiments discussed so far, morphological responses to elevated CO_2 commonly determine the nature of competitive interactions among species. Our knowledge of changes (*see* **Allometric**, Volume 2) in the patterns of plant growth throughout their plant life may allow us to forecast the outcome of these interactions. If, however, elevated CO_2 differentially affects certain specific developmental pathways, as has been demonstrated in a small number of studies (*see* **Plants – from Cells to Ecosystems: Impacts of Global Environmental Change**, Volume 2), predicting the exact outcome of competition will be very difficult.

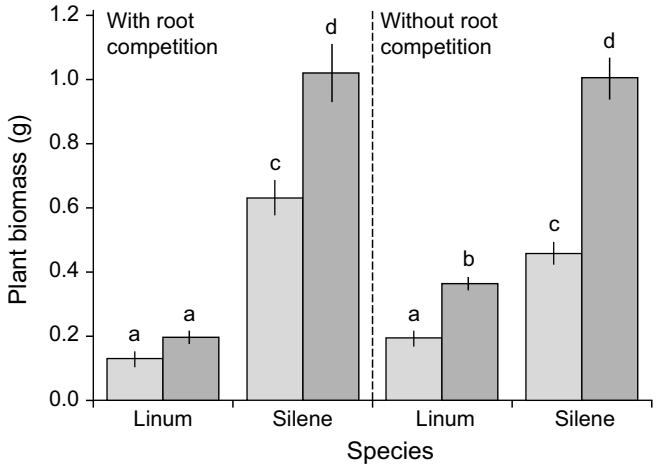


Figure 8 Total biomass of *Linum usitatissimum* and *Silene cretica* individuals growing at ambient (lighter box, 320 $\mu l\, l^{-1}$) and elevated (darker box, 600 $\mu l\, l^{-1}$) CO_2 and under two different competition regimes (with/without root competition) for 30 days (modified from Arnone and Kestenholz, 1997)

Competition in Natural Plant Communities

Elevated CO_2 will not only affect competition between plant species in natural ecological communities, but will also influence a suite of other interspecific interactions, including plant soil, plant herbivore, and plant pollinator. Experiments that have been carried out in more realistic field situations than those described above are often not able to uncouple the separate effects of elevated CO_2 on these individual processes. They are, however, able to place the results of controlled environment studies into context. For example, Leadley *et al.* (1999) found that, as with many glasshouse studies, individual plant responses to elevated CO_2 did not necessarily predict grassland community responses in the field. Interactions between plants and other organisms (e.g., herbivores, soil organisms) are likely to decouple plant–plant competitive interactions observed in controlled studies (Bezemer and Jones, 1998; Zaller and Arnone, 1999). More detailed examination of higher level interactions is necessary for a fuller understanding of elevated CO_2 effects on natural ecological communities.

ELEVATED CO_2 AND ECOSYSTEM FUNCTION

It is now well established that elevated CO_2 will have numerous, direct effects on a variety of ecosystem processes in the future. In particular, there is considerable focus on understanding factors that regulate carbon dynamics in terrestrial ecosystems, because of potential feedbacks to the global carbon cycle. Many studies have extrapolated from experiments on individually grown plants to make predictions about the impact of elevated CO_2 on future ecosystem productivity. This approach, however, may not provide an accurate representation of the future carbon gain capabilities of terrestrial ecosystems. Many of these studies may be misleading, as they rarely consider the impact of plant competitive interactions during stand development. The influence of plant competition on ecological processes may extend beyond plant population and community dynamics to whole ecosystem properties. A number of recent studies comparing plants grown

Table 1 Differences in enhancement ratio (biomass at 700 $\mu l\, l^{-1}$/biomass at 350 $\mu l\, l^{-1}$) for individually and competitively grown plants (low vs. high density)

	Enhancement ratio		
Species	Low density	High density	Reference
Abutilon theophrasti	1.16	1.04	Bazzaz *et al.* (1995)
Betula alleghaniensis	1.49	1.14	Wayne and Bazzaz (1995)
Brassica kaber (early)	2.41	1.59	Wayne *et al.* (1999)
Brassica kaber (late)	1.28	1.30	Wayne *et al.* (1999)
Sinapis alba	1.86	1.15	Retuerto *et al.* (1996)

individually and in monospecific stands have demonstrated that competition often reduces growth enhancement under elevated CO_2 (Bazzaz *et al.*, 1995; Wayne and Bazzaz, 1995; Retuerto *et al.*, 1996) (Table 1). In addition, enhancement effects appear to change though time during stand development (Wayne *et al.*, 1999). In a survey of studies investigating plant responses to CO_2 gradients, Ackerly and Bazzaz (1995) found that plants grown individually usually exhibited positive growth responses due to increasing CO_2 concentrations (61% of the time), while those in competition with neighbors only showed positive responses 19% of the time. The physiological and developmental mechanisms underlying this reduced enhancement with increasing density are currently unclear and certainly need to be investigated further to understand the consequences for terrestrial ecosystem carbon sequestration.

CONCLUSIONS

Plant competitive interactions may have substantial impacts on the responses of populations, communities, and ecosystems to rising CO_2 levels. Experiments comparing responses of plants grown individually and in competition demonstrate that using individuals to extrapolate to higher levels of organization may be misleading. Contrary to our expectations from individually grown plants, selection under future elevated CO_2 concentrations may favor traits that confer improved ability to compete for resources, rather than traits that promote individual CO_2 responsiveness. Similarly, species-specific growth responses to elevated CO_2 may not be useful for predicting future patterns of community change, as competitive interactions may either enhance or depress species' CO_2 responsiveness in mixed species stands. In addition, growing plants in competition consistently reduces estimates of future ecosystem productivity based on CO_2-induced growth enhancement of plants without neighbors.

Because of the ubiquity of plant competition in terrestrial ecosystems, predictions from individually grown plants cannot be extended to the majority of natural settings and are limited to a very small set of environmental conditions. It is clear that models of ecosystem responses to global environmental change should account for potential effects of plant competitive interactions – otherwise, they will be of limited use. As our mechanistic understanding of CO_2 and density interactions grows, we should be better able to develop models that account for the more complex underlying dynamics.

REFERENCES

Ackerly, D D and Bazzaz, F A (1995) Plant Growth and Reproduction Along CO_2 Gradients: Non-linear Responses and Implications for Community Change, *Global Change Biol.*, **1**, 199–207.

Arnone, III, J A and Kestenholz, C (1997) Root Competition and Elevated CO_2: Effects on Seedling Growth in Linum usitatissimum Populations and Linum-Silene cretica Mixtures, *Funct. Ecol.*, **11**, 209–214.

Bazzaz, F A (1990) The Response of Natural Ecosystems to the Rising Global CO_2 levels, *Ann. Rev. Ecol. Syst.*, **21**, 167–196.

Bazzaz, F A, Ackerly, D D, Woodward, F I, and Rochefort, L (1992) CO_2 Enrichment and Dependence of Reproduction on Density in an Annual Plant and a Simulation of its Population Dynamics, *J. Ecol.*, **80**, 643–651.

Bazzaz, F A, Garbutt, K, Reekie, E G, and Williams, W E (1989) Using Growth Analysis to Interpret Competition Between a C_3 and a C_4 Annual Under Ambient and Elevated CO_2, *Oecologia*, **79**, 223–235.

Bazzaz, F A, Jasíenski, M, Thomas, S C, and Wayne, P (1995) Microevolutionary Responses in Experimental Populations of Plants to CO_2 Enriched Environments: Parallel Results from Two Model Systems, *Proc. Natl. Acad. Sci. USA*, **92**, 8161–8165.

Berntson, G M, Rajakaruna, N, and Bazzaz, F A (1998) Growth and Nitrogen Uptake in an Experimental Community of Annuals Exposed to Elevated Atmospheric CO_2, *Global Change Biol.*, **4**, 607–626.

Bezemer, T M and Jones, T H (1998) Plant Insect Herbivore Interactions in Elevated Atmospheric CO_2: Quantitative Analyses and Guild Effects, *Oikos*, **82**, 212–222.

Catovsky, S and Bazzaz, F A (1999) Elevated CO_2 Influences the Responses of Two Birch Species to Soil Moisture: Implications for Forest Community Structure, *Global Change Biol.*, **5**, 507–518.

Chiariello, N R and Field, C B (1996) Annual Grassland Responses to Elevated CO_2 in Multiyear Community Microcosms, in *Carbon Dioxide, Populations, and Communities*, eds C Körner and F A Bazzaz, Academic Press, San Diego, CA, 139–157.

Groninger, J W, Seiler, J R, Zedaker, S M, and Berrang, P C (1995) Effects of Elevated CO_2, Water Stress, and Nitrogen Level on Competitive Interactions of Simulated Loblolly Pine and Sweetgum Stands, *Can. J. For. Res.*, **25**, 1077–1083.

Jifon, J L, Friend, A L, and Berrang, P C (1995) Species Mixture and Soil Resource Availability Affect the Root Growth Response of Tree Seedlings to Elevated Atmospheric CO_2, *Can. J. For. Res.*, **25**, 824–832.

Körner, C and Bazzaz, F A, eds (1996) *Carbon Dioxide, Populations, and Communities*, Academic Press, San Diego, CA.

Leadley, P W, Niklaus, P A, Stocker, R, and Körner, C (1999) A Field Study of the Effects of Elevated CO_2 on Plant Biomass and Community Structure in a Calcareous Grassland, *Oecologia*, **118**, 39–49.

Morse, S R and Bazzaz, F A (1994) Elevated CO_2 and Temperature Alter Recruitment and Size Hierarchies in C_3 and C_4 annuals, *Ecology*, **75**, 966–975.

Navas, M-L, Guillerm, J-L, Fabreguettes, J, and Roy, J (1995) The Influence of Elevated CO_2 on Community Structure, Biomass and Carbon Balance of Mediterranean Old-field Microcosms, *Global Change Biol.*, **1**, 325–335.

Potvin, C and Vasseur, L (1997) Long-term CO_2 Enrichment of a Pasture Community: Species Richness, Dominance, and Succession, *Ecology*, **78**, 666–677.

Reekie, E G and Bazzaz, F A (1989) Competition and Patterns of Resource use among Seedlings of Five Tropical Trees Grown at Ambient and Elevated CO_2, *Oecologia*, **79**, 212–222.

Retuerto, R, Rochefort, L, and Woodward, F I (1996) The Influence of Plant Density on the Responses of *Sinapsis alba* to CO_2 and Windspeed, *Oecologia*, **108**, 241–251.

Reynolds, H L (1996) Effects of Elevated CO_2 on Plants Grown in Competition, in *Carbon Dioxide, Populations, and Communities*, eds C Körner and F A Bazzaz, Academic Press, San Diego, CA, 273–286.

Wayne, P M and Bazzaz, F A (1995) Seedling Density Modifies the Growth Responses of Yellow Birch Maternal Families to Elevated Carbon Dioxide, *Global Change Biol.*, **1**, 315–324.

Wayne, P M and Bazzaz, F A (1997) Light Acquisition and Growth by Competing Individuals in CO_2-enriched Atmospheres: Consequences for Size Structure in Regenerating Birch Stands, *J. Ecol.*, **85**, 29–42.

Wayne, P M, Carnelli, A L, Connolly, J, and Bazzaz, F A (1999) The Density Dependence of Plant Responses to Elevated CO_2, *J. Ecol.*, **87**, 183–192.

Williams, W E, Garbutt, K, and Bazzaz, F A (1988) The Response of Plants to Elevated CO_2: V. Performance of an Assemblage of Serpentine Grassland Herbs, *Environ. Exp. Bot.*, **28**, 123–130.

Zaller, J G and Arnone, III, J A (1999) Interactions between Plant Species and Earthworm Casts in a Calcareous Grassland Under Elevated CO_2, *Ecology*, **80**, 873–881.

Plant Functional Types

Sandra Lavorel

Centre d'Ecologie Fonctionnelle et Evolutive CNRS, Montpellier, France

The search for schemes of species classification that move beyond analyses of particular patterns or phenomena has been a recurring theme in the study of ecological systems since Theophrastus' first recognition of the meaningfulness of plant life forms. This theme has been revived recently by the necessity to predict the effects of human-originated changes in climate, atmospheric composition and land use on ecosystem structure and functioning. Plant functional types (PFTs) can be broadly defined as groups of species with a similar response to one or several environmental factors, in conjunction or not with a shared physiognomy and biology. They are seen as tools that make it possible: (1) to describe the dynamics of plant communities and ecosystems in a generic way, (2) to use this general scheme to generate predictions of ecosystem change under changed environmental conditions, and (3) to develop management strategies that move beyond a focus on particular species. They can be used in a variety of situations, including cases where detailed information is limited. In order to be successful, such schemes need to be based on robust ecological patterns.

Recent research on the prediction of the effects of global change, in particular within the Global Change and Terrestrial Ecosystems (GCTE) program, has significantly advanced the conceptualization of plant functional groups (Woodward and Cramer, 1996; Smith et al.*, 1997; Lavorel and Cramer, 1999). PFTs were intended to be the building blocks of dynamic global vegetation models (DGVMs) that can predict shifts of vegetation types with changing climate and atmospheric composition. Though the initial aim of building a single, globally applicable key of PFTs was rejected, recent research has also demonstrated the usefulness of the PFT concept for approaches at a variety of scales, from the community to the landscape and the globe.*

A CLASSIFICATION OF PLANT FUNCTIONAL CLASSIFICATIONS

The idea of grouping functionally similar species has been used for a long time in biology and has been applied to plants, animals and microbes. Global change research has stimulated the development of plant functional classifications. Though these studies have focussed on plants, the conceptual and methodological advances presented in this article are now widely applied to other kingdoms, including

functional classifications for birds, fish and amphibians, ants or soil microbes.

Published approaches involving the analysis of sets of biological attributes fall into four main types of functional classifications of plant species (Lavorel *et al.*, 1997). In order of increasing specificity of objective, these are: (1) emergent groups – groups of species that reflect natural correlation of biological attributes; (2) strategies – groups of species sharing a set of biological attributes that are interpreted as adaptations to particular patterns of resource use; (3) functional types – species with similar roles in ecosystem processes by responding in similar ways to multiple environmental factors; and (4) specific response groups – species which respond in similar ways to specific environmental factors. The two latter categories represent substantial recent conceptual advances stimulated by research aimed at predicting the effects of global change on vegetation dynamics.

Emergent Groups

Emergent groups reflect the combination of adaptive responses and of evolutionary constraints appearing as sets of correlated traits. They are identified in an inductive manner, using multivariate analyses of usually large sets of traits – covering life history, morphology, physiology, phenology and regeneration biology – expected to determine species behavior in the ecosystem. Such classifications tend to not address any ecosystem function explicitly or, when they attempt to, lack relationship to particular mechanisms. They tend to produce sets of traits essentially corresponding to main life forms (trees, shrubs, grasses and forbs), although the degree of detailed subdivision depends on the breadth of the spectrum taken from global, to regional and to a particular vegetation type. Classifications into emergent groups have, however, been useful to identify broad correlation patterns, such as relationships between plant size or seed mass and important functional traits such as growth rate. Since these correlation patterns may sometimes only reflect the fact that taxa sharing common evolutionary ancestors (phylogenetically-related taxa) can inherit common groups of attributes regardless of function, it is essential to use analyses which compare pairs of phylogenetically independent taxa.

Trait correlation analyses have also pointed out that adult and regeneration traits can show incongruent patterns of correlation, as highlighted by the recognition of distinct regeneration strategies (Grime, 1979). This discrepancy is highly pertinent to the construction of functional classifications because it implies that in the analysis of traits relating to specific functions, adult and juvenile traits need to be considered independently.

Strategies - Adaptive Patterns Relating to Resource Use and Availability

The identification of adaptive strategies is a more directed exercise in linking species traits to variation in resource availability in habitats, and their consequences for plant growth and reproduction. Strategies are identified as "major recurring axes of adaptive specialization in life history and in the physiology of the adult (established) organism [which] appear to be associated with variation in the duration and quality of the opportunities which habitats provide for resource capture, growth and reproduction" (Grime, 1979).

The r-K strategy scheme is the oldest and perhaps best-known classification across subdisciplines. It distinguishes between colonizer species with a high intrinsic population growth rate, but poorly competitive, and species with a slower population growth rate and a strong ability to maintain themselves in competitive situations. These strategies are associated with biological attributes relating to life cycle, reproduction rates and allocation into individual offsprings.

The best known and most debated example in plant ecology is the plant strategy (or competition stress disturbance (CSR)) scheme (Grime, 1979) where species are classified according to their abilities to cope with competition (C), abiotic stress (S) and disturbance (R). These axes of environmental specialization are related to proposed sets of biological attributes, whose relevance has been fully demonstrated only recently in the case of the vegetation of Northwestern Europe (Grime *et al.*, 1997). These attributes can be grouped to define syndromes (i.e., groups of co-occurring attributes). However, the difficulty of using such predetermined syndromes is that they may be meaningful only in vegetation for which they were developed. Instead, the use of three simple plant traits, specific leaf area (SLA) (L), plant height (H), and seed mass (S) has been proposed to represent the C, S and R axes, respectively (leaf height seed (LHS) scheme, Westoby, 1998). These traits present the advantage of being easy to measure using a standardized methodology, while showing strong evidence for correlation with plant performance in competitive, stressful, or disturbed habitats, respectively.

The CSR scheme has been successfully applied in studies of the distribution and dynamics of many communities. However, these studies have tended to focus on the environmental classification as an interpretative tool rather than identifying attributes that are correlated to the same end of an environmental gradient. Although the scheme has been much criticized, it is important to recognize strategy classifications for what they are – general schemes. Their value is to provide a broad framework and to identify traits that may be used as a starting point in more explicitly targeted studies.

Functional Groups - Species with Similar Ecological Roles

Plant functional types (PFTs) are defined as groupings of species which perform similarly in an ecosystem based on a set of common biological attributes. These groupings are independent of whether species are taxonomically related or not. They can be defined in relation to either the contribution of species to ecosystem processes (such as carbon or water cycling) or the response of species to changes in environmental variables (such as climatic variables or disturbance). The search for functional groups usable in models of global vegetation dynamics needs to provide for species groupings that may differ according to the function or group of functions considered. For example, PFTs associated with climatic factors differ from those relevant to species migration. Classifications using *a priori* selected traits under present factor combinations, e.g., present climate, will also not necessarily hold under changing conditions, such as new combinations of temperature and precipitation regimes. These two concerns can be addressed by explicitly linking function to response mechanisms (Woodward and Cramer, 1996). This is possible for functions for which adaptive syndromes are established from direct structural manifestations, but ultimately functionality of any attribute should be assessed experimentally. For example, drought tolerance can be related to patterns of allocation to leaves versus roots, leaf characteristics and metabolic adjustments (e.g., C3, C4, CAM). In other cases, however, the relationship between form and function is known only from correlation patterns, as is the case for plant size and relative growth rate (RGR). Some essential functional attributes (e.g., photosynthetic pathway) do not translate to morphology. Finally, classifications cannot be expected to be universal, yet their development should aim at reaching greater generality by addressing mechanisms.

In practice, the construction of PFTs has involved a combination of subjective pre-selection of traits based on expert knowledge and correlation analyses applied to environmental gradients and experimental manipulations. Individual studies may combine emergent and function-oriented classifications, although not always explicitly. For example, classifications used to model global or regional vegetation shifts associated with climatic change are based on life form (Prentice *et al.*, 1992). That decision reflects both the outcome of emergent classifications at large scales and the relevance of life form to primary ecosystem functions such as carbon assimilation. Indeed subjective life form classifications have been repeatedly found to reflect broad ecosystem functions along regional environmental gradients. However, additional information on traits for which function has been explicitly demonstrated may be of primary importance to develop accurate, mechanistic predictions of vegetation response to global change drivers, as well as to understand plant distribution and dynamics along resource gradients within regions or landscapes.

Specific Response Groups

Refined analyses of vegetation structure and dynamics at landscape or community scales require different sets of traits, which explore the details of interactions between the plant and its abiotic and biotic environment. Searches for plant functional classifications at these finer scales have explicitly considered traits relating to response to competition (as described later in the section on PFTs along resource gradients) and other biotic interactions such as herbivory or litter decomposition, and to response to disturbance (as described later in the section on response groups to disturbance). These classifications are based on a detailed understanding of the effects of a limited set of traits on a specific plant function. For example, the identification of response groups and corresponding plant attributes has been a recurring theme in the grazing literature that predates the formalization of PFT approaches. Similarly, ecophysiological research has been focusing on physiological traits determining plant resource acquisition patterns. These traits are of interest for the detailed analysis of plant response groups for competition or different stress factors, but have only recently started to be included in plant functional classification studies.

METHODS FOR PLANT FUNCTIONAL CLASSIFICATIONS

General Approaches to Plant Functional Classification

Functional classifications can be achieved by following three types of broad approaches (Gitay and Noble, 1997). A *subjective* classification assumes the existence of functional types based on *a priori* knowledge of one or several ecosystems. It is not based on the knowledge of a particular process or of specific traits relevant to that process but solely on expert knowledge of the ecological dynamics. The classification of early-, mid- and late-successional species provides an example for such a classification. An *inductive* classification identifies functional types using a set of traits that are presumed important for a given function or set of functions. Species are then grouped according to common sets of attributes for these traits, and it is assumed that they will as a result perform similar functions. For example, models of vegetation dynamics following climatic and atmospheric changes have assumed that life form, phenology (*see* **Phenology**, Volume 2) and photosynthetic pathway determined species response and could hence be used to build a global vegetation

classification. Finally a *deductive* classification uses *a priori* knowledge of important ecological processes or functions and then deduces the set of potential functional types. A well-known application of this approach is presented by the Vital Attributes Scheme, which proposes a functional classification based on species survival and regeneration mode after disturbance and on their tolerance to competition for recruitment (Noble and Slatyer, 1980). Specific studies often use a mix of inductive and deductive approaches. Ideally, robust plant functional classifications should be based on the reconciliation of a trait-based classification (inductive) with a classification based on response for processes of interest (deductive). This approach involves four steps: (1) the selection of a list of traits to be analyzed; (2) the analysis of response of trait attributes for the process of interest; (3) the analysis of plant response for the process of interest and the identification of groups of species with a similar response; and (4) the reconciliation of steps (2) and (3) to build species functional types characterized by a list of determining attributes. In the following we present the main methods and issues for each of these four steps.

Trait Selection (Step 1)

Assuming that a specific function or set of functions have been identified, it is possible to use expert knowledge to build a list of potentially relevant traits. The selection of a final list of traits for analysis is an exercise in compromise between several constraints. Since traits should be closely linked to actual function, the choice should favor traits that are known to directly determine plant physiology or other aspects of plant behavior (e.g., germination requirements). Such traits, termed hard or functional traits, provide robust information but also present two main difficulties. First, for many functions the exact determining trait or set of traits is unknown. Second, if known, these traits can be difficult or costly to measure, a constraint which can often be incompatible with the need to carry out measurements on a large number of individuals and species, often in the field rather than in the laboratory. As a compromise, soft or structural traits may be chosen. These traits have morphological manifestations that can be measured directly on an individual plant, and are known or assumed to correlate with a harder trait which is the actual functional trait. For example, an accumulating body of evidence suggests that specific leaf area (SLA) provides a robust surrogate to assess the ability of plants to capture resources. The second advantage of soft traits is that they can be easily measured on large numbers of individuals, using relatively simple and inexpensive techniques. An additional constraint for trait selection is the requirement for traits to be robust to variation in other environmental or biotic factors than those considered for analysis. Finally, parsimony is an important quality of a trait list for the analysis of PFTs. The first reason for this constraint is practical but it is not the only one. Using partially redundant or irrelevant traits can also reduce the power of the analysis. In general, classifications are likely to include larger numbers of traits when scaling down from very general schemes applicable to the globe or broad regions, to more and more specific schemes at the regional, landscape and community scale.

Statistical Analysis of Trait Response (Step 2)

Vegetation sampling provides information on species distribution across samples taken under a set of environmental conditions of interest. This information can be represented by a matrix of sample $\times$ species records. For each species, attributes are measured for each of the traits in the selected list. This information can be summarized in a second matrix of species $\times$ attribute records. It is then possible to calculate the distribution of attributes across samples by a simple product between these two matrices. The resulting matrix of attributes contains the representation of each attribute across samples distributed across environmental conditions. Response of plant attributes to one or several environmental factors can then be carried out by analyzing the frequency distribution of particular attributes across factor levels. Traits significantly associated with a given set of environmental factors can then be termed functional traits, and the attributes represented under given environmental conditions are functional attributes. While providing direct tests of hypotheses regarding specific attributes, such individual testing does not provide insights into correlation patterns among traits, i.e., how specific attributes may be grouped into syndromes represented in actual species. The univariate method also lacks information about the relative importance of potential mechanisms. These shortcomings can be overcome through the use of multivariate analyses carried out on the attributes matrix. These can be either indirect methods (e.g., detrended correspondence analysis, DCA), where samples are ordinated and their position then interpreted to represent environmental factors, or direct methods (e.g., canonical correspondence analysis, CCA), where correlations with environmental factors are explicitly integrated into the analysis. Multivariate methods thus identify sets of correlated attributes represented for each environmental factor level. These then represent the functional syndromes. Multivariate methods also provide a ranking of importance of contribution of different traits, which can be used to generate hypotheses about the mechanisms of response.

Statistical Analysis of Plant Response (Step 3)

Groups of plants with a similar response to environmental factors of interest can be easily identified using multivariate analyses or classification techniques applied to the sample $\times$ species matrix. Species are grouped according to similarity in frequency distribution across samples. Direct ordination methods like CCA have the advantage

of making an explicit connection between species groups and environmental factor levels, and are recommended to achieve functional classifications. The groups of species identified through these classifications are species functional groups. At this stage they are not explicitly characterized in terms of biological attributes.

Construction of Functional Types (Step 4)

The construction of PFTs as formally defined implies that the species functional groups be linked with corresponding functional syndromes identified for the same environmental factor levels. This methodological stage has presented difficulties to date due to the lack of objective methods to reconcile what are effectively two classifications, each with a subjective component, in the delimitation of classes. Recent research has proposed direct ordination methods that can circumvent this difficulty (Pillar, 1999).

Robustness of Functional Classi cations

Functional classifications can be carried out at varying levels of specificity. Generality and applicability of PFT classifications outside of the specific ecosystem where they were developed, may hold under certain conditions. It is essential to distinguish different levels at which classifications may be repeatable (finding the same PFTs for several places or several dates using a common trait set) or congruent (finding the same PFTs based on different trait sets). Important factors that may influence the degree of robustness include general abiotic conditions, e.g., productivity, and evolutionary history of the flora.

PFTs ALONG RESOURCE GRADIENTS

Validation of the Plant Strategy Scheme

Several recent studies have carried out direct tests of the relevance of the plant strategy scheme to the distribution of plant species along productivity gradients. They evaluated the relevance of the proposed functional trait list to species distribution in specific regions. A comprehensive screening of traits in common plants of the British flora showed strong evidence for the organization of plant biology along a gradient from stressful (growth of all species limited by scarce resources) to competitive (growth limited by uneven access to abundant resources between species) conditions (Grime *et al.*, 1997). This gradient was associated with a suite of traits reflecting plant resource acquisition capacity. Similar patterns have been uncovered in analyses of functional types and syndromes in a range of habitats including wetlands, grasslands and heathlands.

Response to Resource Gradients

Beyond the competition-stress-disturbance (CSR) scheme and functional classifications *per se*, targeted studies have investigated changes in biological traits along the stress–competition gradient assumed to be associated with increasing resource availability. Results from this research bring important elements for the construction of PFTs in response to resource gradients.

Stress Tolerance

The relevance of a suite of traits to plant tolerance to nutrient stress has been examined. Results have first confirmed the relevance of several traits included as part of the stress–competition gradient, such as growth rate, plant height (or shoot mass), shoot : root ratio, loss of leaf tissue to herbivores, and mycorrhizal infection. In addition, seedling tolerance to stress has been shown to be correlated with seed mass as a reflection of embryo reserves. Further studies have concentrated on moving towards true functional traits reflecting physiological mechanisms. They confirmed the significance of traits relating to nutrient acquisition (e.g., SLA, tissue nitrogen concentration, specific root length) but also pointed out a trade-off between resource acquisition and resource conservation traits (Chapin *et al.*, 1993). Resource conservation traits include leaf life span and carbon-based defenses.

Competition: a Still Unresolved Debate

Several studies have attempted to predict plant response to competition. To date, this effort has not resulted in a clear-cut view of which functional traits are responsible for the outcome of competitive interactions. Competitive ability has been generally assumed to be strongly related to plant size, and more specifically plant height (e.g., Westoby, 1998). However, detailed studies have shown that the situation may not be quite so simple. While this trait seems to be a good predictor of plant competitive effect (i.e., the capacity for a plant to suppress its competitors), it may not be related to competitive response (i.e., the ability of a plant to cope with neighbor competition). In addition, in communities with a large number of short-lived species, seed size may be a better predictor of competitive effect than plant size. In perennial-dominated communities, competitive effect has been generally related to a number of root and shoot traits determining foraging for patchy resources. These include RGR, plant height, ability of lateral spread (e.g., phalanx vs. guerilla strategy in clonal perennials) and a trade-off between scale and precision in nutrient foraging.

RESPONSE GROUPS TO DISTURBANCE

At regional or landscape scales, natural and human disturbance regimes are primary determinants of vegetation

dynamics. Climate and land use affect both natural and human disturbance regimes. For example, fire regimes are directly affected by climate and human activities, at the levels of ignition and spread. Fire spread is affected by landscape heterogeneity. Landscape vegetation patterns are determined by the interaction between natural processes of colonization and growth, driven by climate and possibly atmospheric composition, and land use.

The identification of disturbance response groups has been developed, building on knowledge acquired from the development of the more general classifications.

Elements for the Identification of Disturbance Response Groups

In the CSR scheme, the axis describing disturbance has been subjected to much criticism because it potentially incorporates different types (e.g., soil disturbance, grazing, fire) and variation in intensity and frequency. The relevance of this so-called ruderal axis was, however, confirmed by the screening of traits for British plants (Grime *et al.*, 1997) as it has been for successional gradients in herbaceous and heathland communities.

Broad PFT analyses of species response to disturbance invariably converge on a growth-form based classification. For example, in herbaceous-dominated vegetation (e.g., old-fields, grasslands, herbaceous stratum of shrublands and woodlands) broad disturbance PFTs can be identified with four life forms: annuals and short-lived perennial forbs, perennial forbs, grasses (with possibly subdivisions between annuals and perennials), and shrubs if present. In further refinements, response-specific attributes have been used to split life form groups for particular disturbances, e.g., using palatability for response to grazing, or resprouting capacity for response to fire. Recent work to characterize response groups to specific disturbances in a given vegetation type, has shown the usefulness of classifications which explicitly operate in a hierarchical manner within life forms (Lavorel *et al.*, 1997). While the relevance of life form remains primary, and its effects are accounted for as a first step, further analyses within each life form are carried out to examine how other traits contribute. Different sets of traits can be examined for each life form. The degree of splitting within each life form and within life histories depends on the number of species within each group. Within each life form, traits are distributed into three categories designed and analyzed independently: life history, adult morphology, and regeneration. Studies in a given floristic and environmental context may involve removing and adding specific traits to a core list defined for a particular disturbance type.

Grazing and Small-scale Soil Disturbance in Herbaceous-dominated Vegetation

While the ruderal strategy is useful to describe broad disturbance functional types, e.g., at regional scales, it appears inappropriate for dealing with local disturbance effects, in particular by grazing. A redesigned scheme for the inclusion of grazing can account for disturbances of a different frequency and intensity. Specific attributes of grazing tolerance incorporate elements of either *S* or *R* strategies (e.g., low stature, tough and narrow leaves, low shoot : root ratio). The response of herbaceous-dominated vegetation to grazing and soil disturbance, the two most common disturbances in many grassland areas, has been related to a number of life history, adult structural traits and regeneration traits.

1. Life history – annuals or short-lived perennials are usually favored by disturbances that open gaps in the usually grassy and mainly perennial-dominated matrix.
2. Adult structural traits – tend to be well correlated with grassland response to disturbance. Plant size or height has often been found to decrease with increasing disturbance intensity. Presumably, this pattern results from the correlation between plant size and several physiological attributes that provide adaptation to open environments, and from the direct relationship between plant size and accessibility to grazers. Alternatively, synthetic descriptors like Raunkiaer's life form classification (based on the vertical distribution of perennating meristems during unfavorable periods, e.g., winter or dry season) can be used to analyze morphological changes along gradients of increasing disturbance intensity. The patterns obtained in analyses considering association of several morphological traits (e.g., height, lateral spread, leaf morphology) ultimately reflect life form. Finally, in the case of grazing, a number of traits relating to physical (e.g., spines) and chemical defense (e.g., secondary compounds) have been considered, though their relevance may strongly vary across regions and communities.
3. Regeneration traits – although generally inferred to be key factors determining plant population response to disturbance, they have received little detailed investigation. Synthetic classifications designed under the schemes of Grime's regeneration strategies and the Vital Attributes Scheme have attempted to capture species persistence mechanism (seed bank, dispersal or vegetative regeneration) and the spatio–temporal distribution of recruitment. When dealing with specific floras, the list of relevant regeneration traits depends on dominant life forms. In vegetation with a large annual component, seed persistence and germination phenology determine post-disturbance dynamics. When perennials dominate, seed size, resprouting ability and

clonal spread need to be considered. Finally, although its relationship to timing of disturbance is essential, species phenology has often been neglected.

Fire

Unlike the predominantly inductive approach described for grassland response to grazing and soil disturbance, response to fire has been analyzed through the examination of deductive functional types. A well-known example is that of the Vital Attributes Scheme (Noble and Slatyer, 1980). Functional types are based around a limited set of functional traits: species tolerance to fire as established individuals, species regeneration mode (seeding vs. resprouting), tolerance of competition and other abiotic constraints during regeneration, and species life schedules relative to fire return intervals.

GLOBAL CHANGE: TOWARDS A MORE COMPLEX VIEW OF PFTs?

At the scale of the globe or of very large regions, vegetation models have used for a long time functional classifications such as the Holdridge or the Köppen classifications, which are variations around the combination of life form and phenology (*see* **Holdridge Life Zone Classification**, Volume 2). These can reliably predict present distributions and capture important functional processes governing biome distribution. They can also be adequate to examine ecosystem functioning globally, e.g., geographical patterns of net primary productivity, and their changes under given climate change scenarios. However, the understanding of vegetation change in specific ecosystems or regions, and predictions of changes in ecosystem processes appear to require a greater level of detail in plant functional traits. At this level, the proposition of a global PFT key has been rejected (Woodward and Cramer, 1996; Smith *et al.*, 1997). In particular, a common classification for response to different components of global change (climate, atmospheric composition, land use and biological diversity) may not be feasible because non-overlapping classifications are expected when different components involve biologically independent traits. In particular, adult traits involved in the response to abiotic factors such as temperature, CO_2 concentration, or nitrogen availability are often decoupled from regeneration traits involved in response to land use disturbance. Previous sections have provided the elements for an examination of possible plant functional classifications in relation to changes in productivity (brought on by CO_2 enrichment, nitrogen deposition or climate change) and land use changes (through changes in disturbance regimes). Additional elements need to be included to account for response to changes in increasing atmospheric CO_2 concentration, specific climatic parameters and exotic species invasions. Table 1 presents a synthesis of different functional classifications that have been tested with respect to response to different global change factors, at a range of scales.

Response to Increasing Atmospheric CO_2

Studies of plant response to elevated CO_2 initially concentrated on the response of subjective groups, based on basic knowledge of plant physiology. These included: C3 vs. C4 grasses, nitrogen fixers (e.g., legumes) vs. non-nitrogen fixers, monocotyledons vs. dicotyledons and annuals vs. perennials. Overall, these *a priori* PFTs do not appear to be truly functional, as they are poor predictors of plant growth response to elevated CO_2. Tests of response among plant strategies (C, S, R) also showed a strong heterogeneity of response within strategies, invalidating the prediction that fast-growing, competitive species should show a strong stimulation. More focussed analyses were then concentrated on an inductive approach linking observed growth response to species traits. They have consistently shown the relevance of a few traits relating to carbon assimilation including SLA, RGR and stomata density. In addition, several reviews have pointed out the necessity to consider traits relating to degree and nature of plant interactions with soil organisms (in particular symbiosis and mycorrhizae), but few significant results are available at this stage. Finally, it is important to remember that traits affecting seed germination and seedling recruitment may be disconnected from those determining growth of established plants and that their effects on plant establishment under elevated CO_2 would also need to be considered.

Response to Changing Climate

Beyond the effects of increased numbers of growing days, taken into account at the level of life form and phenology, analyses of PFTs for response to climate change also need to consider changes in extreme temperatures. For example, frost resistance appears to be positively correlated with genome size (estimated by DNA content), while genotypes with small genomes are expected to show a greater growth response than those with larger genomes. As a result, expansion of small genome species may be limited by the distribution of frosts rather than by growing temperatures. This example points towards the difficulty of identifying PFTs that will reflect the possible complexities in response to temperature regimes rather than to mean temperatures. Similarly, analyses of PFTs relating to response to changed precipitation regimes have pointed out the importance of response to water stress. In semi-arid grasslands and shrublands, savannas, or rainforests, life form and species distributions

Table 1 Summary of functional classi cations whose relevance for different global change factors has been tested empirically or using models of vegetation distribution[a]

	Global change factor						
	Climate[b]		Atmosphere	Land use			Biodiversity
Classi cation	Global	Local	CO_2	Fire	Soil disturbance	Grazing	Invasions[c]
Life form[d]	Y	N	Y	N	Y	Y	N
Life cycle[e]	Y	Y?	Y	Y	Y	Y	Y
Taxonomy[f]	N	N	N	N	N	Y	N
Growth phenology[g]	Y	Y	N	N	N	Y	Y
Plant size[h]	Y	N	?	N	Y	Y	Y
Rooting depth	Y	Y	N	N?	N?	Y?	Y
Leaf form	Y	Y?	?	?	?	Y	N?
RGR[i]	Y	Y?	Y	N	Y	Y	Y
Photosyn- thetic pathway[j]	Y	Y	Y	N	N	N?	N
Nitrogen xation	N	N	Y?	N	N	(Y)	Y
Secondary compounds	N	N	Y?	N	N	Y	N
Resprouting ability	N	N	N	Y	Y	Y	N
Seed size[k]	N	N	N	Y	Y	Y?	N
Seed dispersal[l]	Y?	N?	N?	Y?	Y?	Y?	Y?
Germination phenology	Y?	Y	N	Y	Y	Y	Y
Plant strategies (CSR)	N	Y	Y?	N	Y	N	Y

[a] Y, classi cation with consistent relevance; N, classi cation with no consistent effect; ?, hypothetical effect or incompletely proven.
[b] Effects on global or continental scale distributions over time scale assumed to allow for species migration are separated from effects within communities of particular regions that take place over a few years within the local species pool.
[c] Group of species affecting community resistance to invasion.
[d] Tree/shrub/herbaceous or sub-classi cations within each of these; the degree of re nement depends on the scale addressed.
[e] Annual, biennial, short-lived perennial, long-lived perennial.
[f] Monocotyledon vs. dicotyledon; particular families.
[g] Sempervirent vs. deciduous; summer vs. winter-growing.
[h] Height or height × lateral spread.
[i] RGR, rate of growth per unit of biomass; estimated based in soft surrogates such as SLA (SLA = leaf mass/leaf area) and leaf water content.
[j] C3, C4 or CAM.
[k] Measured as seed dimensions or preferably seed mass; correlates with seedling establishment capacity.
[l] Capacity resulting from the combination of fecundity and dispersal distance. This has been tested in only few speci c cases.

appear determined by differential capacities of plants to grow and survive during drought periods, either seasonally predictable, or unpredictable across years. These result from both water stress tolerance (related to root : shoot ratio, leaf characteristics, photosynthetic pathway for established plants and seed size for seedlings) and avoidance (through deep roots). Finally, in some biomes such as Mediterranean-type ecosystems, savannas, and arctic tundra, climate change is expected to have strong effects on fire regimes. In this case, a relevant functional classification would need to include features of fire response groups.

Interactive Effects of CO_2, Climate and Nitrogen Deposition

Global change components do not impact ecosystems in isolation but interactively. Potential growth response to CO_2 can be strongly modified by water or nitrogen availability. Similarly, the improvement of growth under higher temperatures and/or moister conditions is likely to be enhanced by increases in inorganic nitrogen availability resulting from atmospheric deposition. These types of interactions are those which pose the greatest challenges to PFT research because they involve responses to multiple factors, each involving different sets of traits.

Response to Changing Land Use

Plant functional classifications relevant to changing land use can be based on disturbance PFTs. However, because these are by essence specific to a single disturbance type (response groups), the challenge lies in combining different types of disturbances when new human disturbances (e.g., pollution, soil removal) are introduced.

Response to Exotic Invasions

Invasion by plant, animal or pathogen species that originated from exotic regions can have large impacts on plant community composition, ecosystem functioning, and disturbance regimes. At this point, very little is known about the susceptibility of different receiving communities as a function of their functional composition. Current knowledge suggests that functional classifications of invaded communities would have to be specific to the type of invader (i.e., response groups). In particular, in the case of plant invasions, it has been proposed that susceptibility of a resident community may depend on the relationship between the invader's functional traits and those of the invaded community.

CONCLUSION

PFTs, identified at broad scales, may conceal the more subtle mechanisms that underlie the response of species within life forms to a specified global change factor. While the feasibility of a global plant classification has been rejected, developing a global scheme of PFTs, based on major disturbances (e.g., fire, grazing) specific to different biomes and regions, is crucial for further developments of models of global vegetation dynamics. The aim should not be a universal classification but rather to identify a hierarchy of generality in functional attributes ranging from global to regional. Success will depend on setting a contextual framework, along with a standardized methodology, for comparing studies across regions with different climates, evolutionary histories and current land use (Lavorel and Cramer, 1999).

REFERENCES

Chapin, III, S T A, Autumn, K, and Pugnaire, F (1993) Evolution of Suites of Traits in Response to Environmental Stress, *Am. Nat.*, **142**, 578–592.

Gitay, H and Noble, I R (1997) What are Functional Types and How Should We Seek Them? in *Plant Functional Types: Their Relevance to Ecosystem Properties and Global Change*, eds T M Smith, H H Shugart, and F I Woodward, Cambridge University Press, Cambridge, 3–19.

Grime, J P (1979) *Plant Strategies and Vegetation Processes*, John Wiley & Sons, Chichester.

Grime, J P, Thompson, K, Hunt, R, Hodgson, J G, Cornelissen, J H C, Rorison, I H, Hendry, G A F, Ashenden, T W, Askew, A P, Band, S R, Booth, R E, Bossard, C C, Campbell, B D, Cooper, J E L, Davison, A W, Gupta, P L, Hall, W, Hand, D W, Hannah, M A, Hillier, S H, Hodkinson, D J, Jalili, A, Liu, Z, Mackey, J M L, Matthews, N, Mowforth, M A, Neal, A M, Reader, R J, Reiling, K, Ross-Fraser, W, Spencer, R E, Sutton, F, Tasker, D E, Thorpe, P C, and Whitehouse, J (1997) Integrated Screening Validates Primary Axis of Specialisation in Plants, *Oikos*, **79**, 259–281.

Lavorel, S and Cramer, W, eds (1999) Plant Functional Types and Disturbance Dynamics, *J. Veg. Sci.*, **10**, 603–730.

Lavorel, S, McIntyre, S, Landsberg, J, and Forbes, T D A (1997) Plant Functional Classification: from General Groups to Specific Groups Based on Response to Disturbance, *Trends. Ecol. Evol.*, **12**, 474–478.

Noble, I R and Slatyer, R O (1980) The Use of Vital Attributes to Predict Successional Changes in Plant Communities Subject to Recurrent Disturbances, *Vegetatio*, **43**, 5–21.

Pillar, V D (1999) On the Identification of Optimal Plant Functional Types, *J. Vegetat. Sci.*, **10**, 631–640.

Prentice, I C, Cramer, W, Harrison, S P, Leemans, R, Monserud, R A, and Solomon, A M (1992) A Global Biome Model Based on Plant Physiology and Dominance, Soil Properties and Climate, *J. Biogeogr.*, **19**, 117–134.

Smith, T M, Shugart, H H, and Woodward, F I, eds (1997) *Plant Functional Types. Their Relevance to Ecosystem Properties and Global Change*, Cambridge University Press, Cambridge.

Westoby, M (1998) A Leaf-height-seed (LHS) Plant Ecology Strategy Scheme, *Plant Soil*, **199**, 213–227.

Woodward, F I and Cramer, W, eds (1996) Plant Functional Types and Climatic Changes, *J. Veg. Sci.*, **7**, 306–308.

Plant Growth at Elevated CO_2

Hendrik Poorter[1] and Marta Pérez-Soba[2]

[1]*Utrecht University, Utrecht, The Netherlands*
[2]*Plant Research International, Wageningen, The Netherlands*

Elevated carbon dioxide (CO_2) concentrations stimulate the rate of photosynthesis of most plant species. Consequently, it is expected that prolonged growth of plants at increased atmospheric CO_2 concentrations will enhance their biomass. This prediction is in agreement with most experimental results performed at optimal growth conditions. However, plant growth is stimulated less than expected on the basis of the increased rate of photosynthesis per unit leaf area. This is due to the fact that plants at elevated CO_2 have reduced leaf area on display per unit biomass, thereby decreasing the relative amount of machinery. The reduction goes with increased accumulation of non-structural carbohydrates.

The magnitude of the growth response to elevated CO_2 depends on the plant species. Species like maize and millet, which have a special type of photosynthesis, show a marginal growth response to elevated CO_2 (on average 11%). However, most plant species respond more strongly, with an average stimulation in biomass of 47% for a

doubling of the present atmospheric CO_2 level. Inherently slow-growing species are generally less responsive than fast-growing species. The growth response also depends on the level of other environmental factors. In nutrient-poor and cold environments the proportional growth stimulation by elevated CO_2 will be less. There is a slight increase in response when levels of light and water are decreased. The strongest interaction is with ozone, with plants showing a much stronger CO_2 response when ozone concentrations are higher.

Elevated atmospheric carbon dioxide (CO_2) concentrations generally have two direct and momentary physiological effects on plants. First, they increase the rate of photosynthesis of the leaves, because of higher CO_2 concentrations at the sites where enzymes fix CO_2. Second, they cause stomata to partly close, thereby reducing the water loss due to transpiration. Thereafter, secondary effects may occur, when plants experience elevated CO_2 concentrations for a longer time. Respiration rates may decrease or increase, chemical composition of the plant may change, and so may leaf morphology and anatomy. For a review of the effect of CO_2 on these physiological processes (*see* **Plants – Carbon Balance and Growth**, Volume 2). In this contribution, we analyze how the myriad of direct and indirect changes affects the growth of individual plants in the vegetative stage. First, plant growth responses under relatively optimal conditions will be discussed. Thereafter, we will consider CO_2 effects under conditions limiting for plant growth. Responses are analyzed by compiling literature data for a large range of published experiments on individually growing whole plants, grown either in growth chambers, glasshouses or open-top chambers.

GROWTH STIMULATION DOES NOT ONLY DEPEND ON THE STIMULATION OF PHOTOSYNTHESIS

The growth of a plant over time can be adequately described by the relative growth rate (RGR; see Box 1 for more information). Mathematically, this variable is the product of three biologically meaningful growth parameters: ULR, SLA and LMF. The ULR is the biomass increase per unit leaf area and time, and is a composite parameter that includes the daily rate of photosynthesis per unit leaf area, the fraction of fixed carbon that is respired the same day and the carbon content of the plant material. SLA is the amount of leaf area formed per unit leaf biomass, a parameter linked to the light-intercepting capacity of a leaf. The LMF is the relative amount of total plant biomass invested in leaves, and thus in the process of carbon fixation.

How are RGR and its components affected by elevated CO_2? We analyzed 130 observations on the growth of species with the most common type of photosynthesis (C_3 species) from 50 published experiments (see Poorter and Nagel, 2000). For each growth parameter we calculated the ratio between the value at elevated CO_2 (approximately 70 Pa) and that at current levels (approximately 35 Pa). The distribution of these ratios is shown in Figure 1. On average, the RGR is stimulated to a small extent (9%). The increase can be attributed to an increase in ULR (23% on average). Although both respiration and carbon concentration may change to some extent (Bunce, 1994; Poorter *et al.*, 1997), it is mainly the stimulation in the rate of photosynthesis per unit leaf area that increases ULR. At the same time, the SLA decreases (13% on average), which implies that plants grown at elevated CO_2 have less leaf area on display per gram of leaf. This decrease in SLA counteracts the increased rate of photosynthesis per unit leaf area, the net result being that growth is stimulated substantially less than was expected on the basis of the stimulation in the rate of photosynthesis alone. The decrease is at least partly linked to the accumulation of starch and other non-structural carbohydrates (Körner *et al.*, 1995). The most likely reason for the accumulation is that plants have increased photosynthesis to the extent that they cannot usefully deploy the newly fixed carbohydrates into new growth. They are then accumulated as non-structural carbohydrates in the leaves, causing the SLA to decrease. The last parameter involved, the LMF, hardly changes (1% increase on average), indicating that biomass allocation is generally not affected by elevated CO_2 (Stulen and Den Hertog, 1993). In the end, it is the

Box 1 Terminology used. All biomass data presented in this article are on a dry weight basis

Abbreviation	Full term	De nition	Units
BER	Biomass enhancement ratio	Biomass at elevated CO_2/biomass at control levels	$g\,g^{-1}$
LMF	Leaf mass fraction	Fraction of biomass allocated to leaves	$g\,g^{-1}$
RGR	Relative growth rate	Increase in biomass per unit biomass already present and per unit time	$m\,g\,g^{-1}\,day^{-1}$
SLA	Speci c leaf area	Leaf area per unit leaf biomass	$m^2\,kg^{-1}$
ULR	Unit leaf rate	Increase in biomass per unit leaf area and time	$g\,m^{-2}\,day^{-1}$

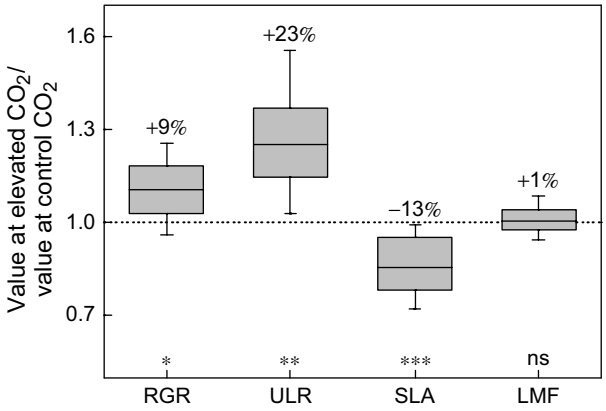


Figure 1 Effect of elevated CO_2 on RGR and its components ULR (g dry mass increase m^{-2} leaf area and day^{-1}), SLA (m^2 leaf area per kilogram leaf mass) and LMF (g leaf per gram total plant). The effect is calculated by dividing the value at elevated CO_2 by the value at ambient levels of CO_2. Results from 130 growth analyses (compiled in Poorter and Nagel, 2000) are shown as boxplots. The numbers above the boxplots are the average percentage increases due to elevated CO_2 (10%-trimmed means). Boxplots indicate the distribution of a range of observations. The lower part of the box shows the 25th percentile, the value below which 25% of the observations are found. The highest part of the box gives the 75th percentile, and the line in between the median (50th percentile). The error bars give the 10th (lower) and 90th (higher) percentile. Signi cance levels for the H_0-hypothesis that the ratio equals 1 are indicated above the *x*-axis: ns, non-signi cant; $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$

net balance of all these physiological and morphological changes that determines the overall growth response of plants to elevated CO_2, and not only the response of leaf photosynthesis.

GROWTH STIMULATION IS TIME-DEPENDENT

Given the possible feedbacks discussed above (changes in physiological processes, biomass partitioning between plant parts and chemical composition), it is difficult to forecast the growth response of plants from short-term effects on the rate of photosynthesis and transpiration alone. This problem becomes even larger as the growth stimulation turns out to be time-dependent. In quite a number of studies, RGR is increased by elevated CO_2 only at the beginning of the experiment, the growth stimulation lasting *circa* 5–15 days. Thereafter, growth rate becomes more or less equal to that of plants at ambient CO_2. In the rest of this contribution, we will use the ratio of plant biomass at elevated and at ambient levels of CO_2 as a measure of the growth response of plants. We will call this the BER (*biomass enhancement ratio*). If RGR were only stimulated during the first 10 days of CO_2

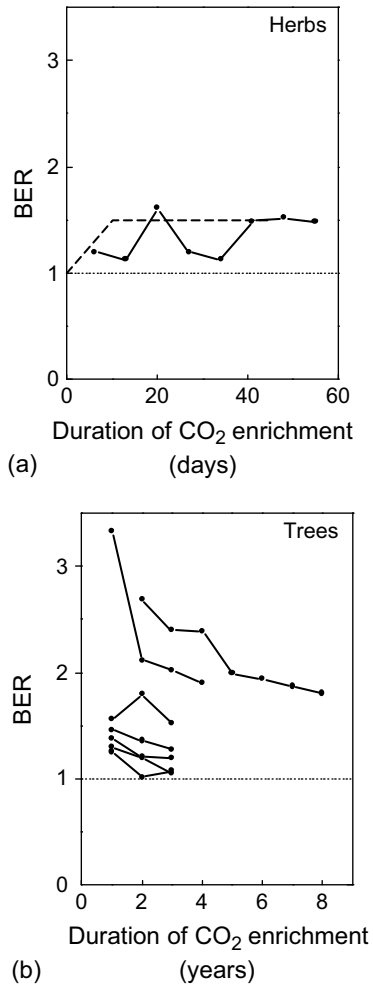


Figure 2 Time course of the BER due to elevated CO_2. (a) An example of an idealized response (bold line) and the response of a herbaceous species. (b) The observation for some tree species

enrichment, BER would start at a value of 1 at the beginning of the experiment, increase up to a certain level until day 10, and then remain constant (Figure 2a, broken line). Particularly in longer-term experiments with tree seedlings, it has been reported that BER even decreases after the first year of the experiment (Figure 2b). In this contribution we will not discuss possible differences in duration between experiments, the reader is referred to reviews like Poorter (1993) and Curtis and Wang (1998).

GROWTH STIMULATION DIFFERS BETWEEN GROUPS OF SPECIES

While keeping in mind that the growth stimulation may be time-dependent, it would be useful to condense the information from the numerous experiments on elevated CO_2 that have been carried out with a wide range of plant species. We

did so by screening the literature, restricting our analysis to experiments where plants were grown at elevated CO_2 levels between 60 and 80 Pa, and at control levels of CO_2 between 30 and 40 Pa. For each experiment and each species where whole plant dry mass was determined, we calculated the BER, arriving at a total of 600 observations (for more information see Poorter, 2001). These were first averaged per species, and subsequently species were classified into a limited number of groups that we thought to have important characteristics in common. Such a separation into functional types may increase insight into the differential response of groups of species and enables generalization of the results. The first distinction we made was on the basis of the type of photosynthesis. Most species fix CO_2 during the day, and experience relatively low levels of CO_2 at the actual site of fixation (so-called *C_3 plants*) (*see* **C_3 and C_4 Photosynthesis**, Volume 2). A number of species, like maize and millet, have a very efficient way to increase the CO_2 concentrations to high levels at the place in the leaf where carbon is fixed (so-called *C_4 plants*), even at low ambient concentrations of CO_2. Purely on that basis they can be expected to respond less to an increase of CO_2 in the atmosphere than C_3 species. Other species, like cacti, take up CO_2 from the air during the night but actually incorporate that carbon into carbohydrates during the day (crassulacean acid metabolism (*CAM*) species). The C_3 species respond most strongly to elevated CO_2, with an average increase in biomass of 47% (Figure 3). Data on CAM species are scarce; the mean response for the 6 species investigated being 21%. Rather unexpectedly, C_4 species also respond to increased CO_2 levels (11%). This might be due to increased rates of carbon fixation at elevated CO_2 in some C_4 species. Another explanation might be that in some experiments plants were thought to grow under optimal conditions, when in fact they were suffering from water stress to some extent. Then, the partial closure of the stomata caused by elevated CO_2 was advantageous for their water balance and would have enhanced growth by alleviating stress (see next section). Such an effect may play a role in experiments with C_3 species as well.

The second distinction we made was within the group of C_3 species. On average, crop species show the highest biomass enhancement (59%) and wild herbaceous plants the lowest (41%). Most of the experiments with woody species have been carried out with seedlings, thus covering only a small part of the life cycle of these plants. Their growth stimulation is on average 49%. Crop species are generally fast growing. The herbaceous wild plants consist of a wide mix of species, some growing as fast as crop species, others almost as slow as tree seedlings. It has been suggested that inherent variation in RGR may account for part of the interspecific variation between C_3 species (Poorter, 1993). Indeed, inherently slow-growing species respond generally less (27% on average) than the

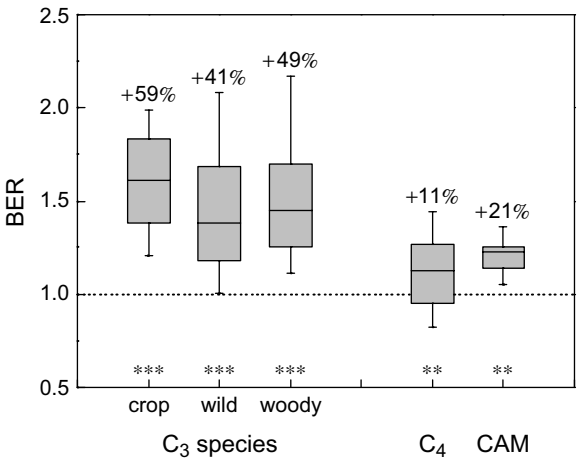


Figure 3 Distribution of BER for several functional types of species. Distributions are based on 540 observations from 280 C_3 species; 55 observations from 30 C_4 species; 13 observation from six CAM species. C_3 species were separated into three groups. There were 110 observations from 24 crop species; 190 from 115 wild herbaceous species; and 230 from 135 woody species. (Literature data compiled in Poorter, 2001)

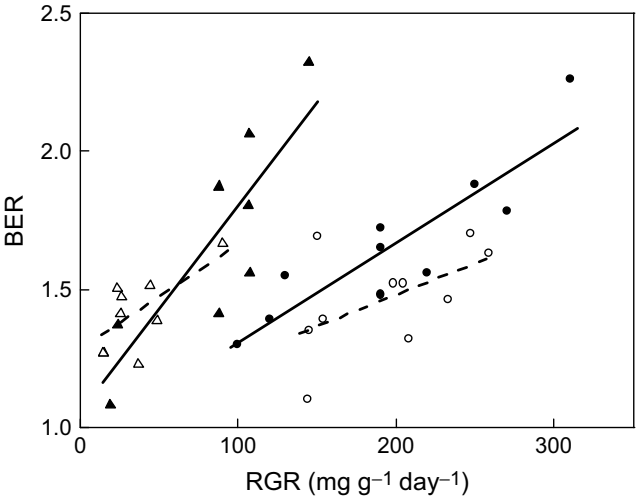


Figure 4 BERs of different species as a function of their RGR at ambient CO_2. All data pertain to C_3 species. Open triangles: temperate woody seedlings (Cornelissen *et al.*, 1999); closed triangles: tropical woody seedlings (Winter and Lovelock, 1999); closed circles: herbaceous crop and weed species (Bunce, 1997); open triangles: herbaceous wild species (Poorter, 1993). Continuous lines are regression lines that differ signi cantly from zero ($P < 0.05$), broken lines are non-signi cantly different

category of fast-growing species (60% on average). This generalization is confirmed in most experiments where fast- and slow-growing species are compared for their BER. Two examples for herbaceous species and two for woody species are shown in Figure 4. Although the correlation

is not always statistically significant, in all research on the response of a range of species, the slope of the line that fits the BER versus RGR data is found to be positive (Poorter, 2001). Most likely, it is the balance between the carbohydrates supplied by photosynthesis and those utilized in structural growth and/or specialized storage organs that determines the degree of response to elevated CO_2. This may explain why fast-growing species, or species with storage sinks as tubers are among the species showing the stronger responses (Reekie *et al.*, 1998).

GROWTH STIMULATION DEPENDS ON THE LEVEL OF PRIMARY RESOURCES

The former analysis was for plants that were grown under laboratory conditions at levels of nutrients, light and water that are generally optimal or close to optimal for growth. What will happen if one of the primary resources is in short supply and causes a growth limitation? Given that both control as well as CO_2 enriched plants would be hampered in growth, one may expect that the absolute growth stimulation by elevated CO_2 will decrease. This is indeed the case. But what will happen to the relative growth stimulation? We compiled experiments that have been carried out at ambient and elevated levels of CO_2, and with high and low levels of the three primary resources (nutrients, light and water) (Poorter and Pérez-Soba, 2001). The severity of the stress was expressed by calculating how much the biomass of sub-optimally grown plants at ambient levels of CO_2 was reduced relative to the biomass of the optimally grown plants at the same CO_2 level. We then calculated the BER at sub-optimal and optimal conditions, and subsequently the slope of the line that connects the two BER values. If the slope is positive, it implies that BER increases when a primary resource becomes growth limiting. An example is given in Box 2.

The distribution of the slopes for a range of experiments is represented by the boxplots of Figure 6. On average, the slope is negative in the case of nutrients. This implies that a decrease in nutrient availability decreases the relative response of plants to elevated CO_2. The pattern of responses was similar between experiments where nitrogen, phosphorus, or all nutrients together were varied. It seems that at ambient CO_2 conditions the growth of plants at low nutrient levels is not restricted by carbon availability, since high concentrations of starch and other non-structural carbohydrates are found in nutrient-stressed

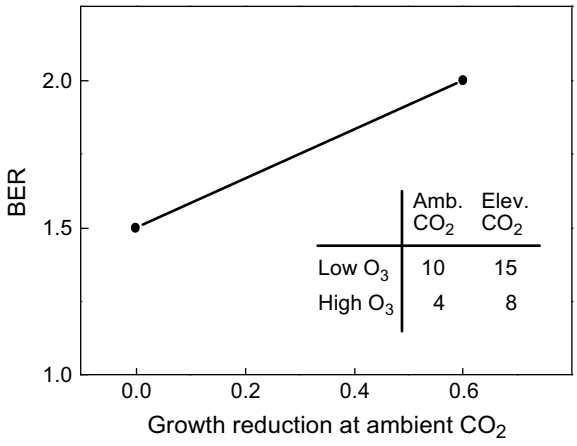


Figure 5 The CO_2 response of plants that grow under a low concentration of ozone (at $x = 0$) or at a high concentration. The insert gives the hypothetical biomass data on which the graph is based

Box 2 Calculating the effect of limiting factors on the BER

To analyze how elevated CO_2 interacts with another environmental factor at the level of plant growth, a minimum of four observations is required: the mass of plants at ambient and elevated CO_2 levels at both a low and high level of the other environmental factor. In Figure 5, an example is given for the interaction with ozone. Assuming that the mass of plants is as given in the insert, then the BERs are 1.5 and 2 at low and high ozone concentrations, respectively. It is dif cult to assess the severity of the ozone stress just from a description of the experimental design. Therefore, as a measure of the stress experienced by the plants, we use the growth reduction shown by the high-ozone plants at ambient CO_2, scaled relatively to the mass of the high-ozone plants at ambient CO_2. If that fraction is close to 0, the stress was minor; if the fraction is much higher, then the ozone stress was apparently more severe. The BER value of the low-ozone plants is plotted at $x = 0$. The BER value of the high ozone plants is plotted at $x = 0.6$, this value being the growth reduction due to the high ozone concentration. After plotting these ratios, the slope of the line connecting the points is calculated. A positive slope indicates that at high ozone concentrations, the relative increase of plant biomass by elevated CO_2 is larger. Note that in absolute terms, high-ozone plants at elevated CO_2 are still smaller than low-ozone plants at elevated CO_2.

The advantage of this approach is that the same procedure can be followed for other environmental variables, just replacing the low ozone treatment with, for example, low water supply or high SO_2. In the case of temperature there is an optimum rather than a linear response. Here we selected the treatment with the highest biomass as the optimal one. A disadvantage of the approach followed is that the analysis implicitly assumes that the BER changes linearly across the range of another environmental factor from sub-optimal to optimal. For the calculations, all BER values have to be transformed prior to any mathematical analysis, as ratios are log-normally distributed. The slope in this case is 0.48.

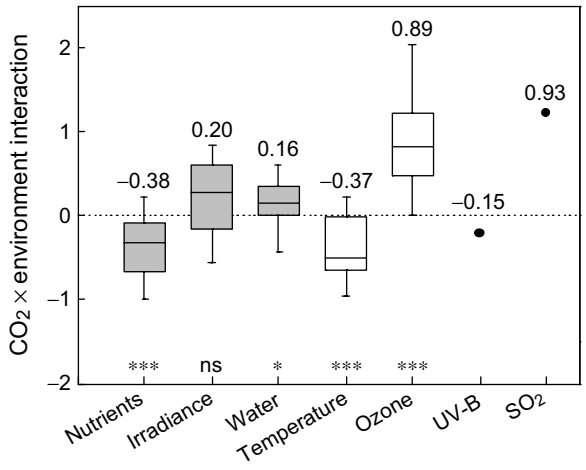


Figure 6 Interaction between elevated CO_2 and various environmental factors on plant growth. Boxplots give the distribution of the slopes, obtained as outlined in Box 2. Values below zero indicate that the relative stimulation of plant growth by elevated CO_2 is smaller at growth-limiting levels of the various factors under investigation. Values above zero indicate that the BER is larger. Data are based on a literature review (Poorter and Pérez-Soba, 2001) on factorial experiments with combinations of elevated CO_2 and nutrients ($n = 102$ observations), irradiance ($n = 19$), water ($n = 29$), temperature ($n = 56$), O_3 ($n = 24$), UV-B ($n = 7$) or SO_2 ($n = 3$). Due to the low number of observations for UV-B and SO_2 we only calculated the average values. Signi cance levels are given for the v-hypothesis that the slope equals zero

plants. Consequently, an increase in the carbohydrate pool due to an increase in photosynthesis by elevated CO_2 concentrations is not expected to stimulate growth to a large extent.

The slope is positive for the factor irradiance, but variability is high and the mean does not differ significantly from zero. A positive slope would imply that at low irradiance the response to elevated CO_2 is relatively stronger than at higher irradiance. It has been suggested that elevated CO_2 decreases the light compensation point of photosynthesis (*see* **Plants – Carbon Balance and Growth**, Volume 2). However, that would only hold at very low irradiance levels (Kimball, 1986a). Elevated CO_2 causes a partial closure of stomata, which in turn reduces plant water use. Therefore, plants growing at limited water availability might thrive longer on a given water supply. The relative growth response is expected to be higher, and this is indeed what has been observed (Figure 6). However, on average the interaction with water is small.

It should be noted that, in general, interactions can be expected between these primary resources: at low water supply, nutrients will be less mobile, and at high irradiance, plants will grow faster, with a higher chance of nutrient or water depletion around the roots.

CO_2 INTERACTS WITH OTHER GLOBAL CHANGE FACTORS

Global change is not restricted to an increase in the atmospheric CO_2 concentration. Increases in temperature due to the accumulation of greenhouse gases such as CO_2, methane and chlorofluorocarbons are anticipated as well. Moreover, because of anthropogenic activities, atmospheric ground-level concentrations of ozone (O_3), nitrogen oxide (NO_x) and sulfur dioxide (SO_2) are rising, and so is the UV-B radiation. How will these factors influence plant growth responses to rising atmospheric CO_2 concentrations? We followed the same approach as in the previous section, analyzing factorial experiments in which both CO_2 and one of these other environmental factors were investigated. Once more, we calculated the growth reduction under suboptimal conditions relative to the optimal one and determined the slope of the line connecting the two BER values.

The interaction of CO_2 with supraoptimal temperatures has hardly been investigated. On average there is a smaller growth response to elevated CO_2 at temperatures suboptimal for growth (Figure 6; cf. Idso and Idso, 1994). One explanation is that elevated CO_2 stimulates photosynthesis relatively less at lower temperatures. Another is that at suboptimal temperatures, growth is inhibited more than photosynthesis, causing non-structural carbohydrates to accumulate. As with low nutrient availability, this may preclude a strong response to increased CO_2. For those plant species that currently grow under sub-optimal temperatures, one may expect a greater CO_2 response if a global warming of the climate takes place.

Ozone shows the strongest positive interactions with CO_2 (Figure 6). For almost all observations, the slope is positive, implying that CO_2 strongly ameliorates the effect of ozone. There is good evidence that in plants in which stomatal conductance is reduced by CO_2 enrichment, O_3 flux into the leaf interior is reduced and this contributes to an amelioration of the detrimental effects of O_3 on plant growth and physiology (Turcsányi *et al.*, 2000). It is yet not clear to what extent other mechanisms induced by elevated CO_2 may play a role in amelioration, such as detoxification or repair processes. A second point of concern is the strength of the combined effect of elevated CO_2 and O_3 on the growth and productivity of species in which the stomata are less responsive to CO_2 enrichment, such as many conifers. Data indicate that for those species, there may be similar effects of O_3 at ambient and elevated CO_2, or at least much less amelioration of ozone damage than that observed in herbaceous species (Pérez-Soba *et al.*, 1995).

Experiments with $CO_2 \times$ UV-B interactions are scarce, the data hinting at the fact that elevated CO_2 could not compensate for the harmful effect of UV-B (Figure 6). However, these results obtained under artificial environmental conditions could result in responses different from

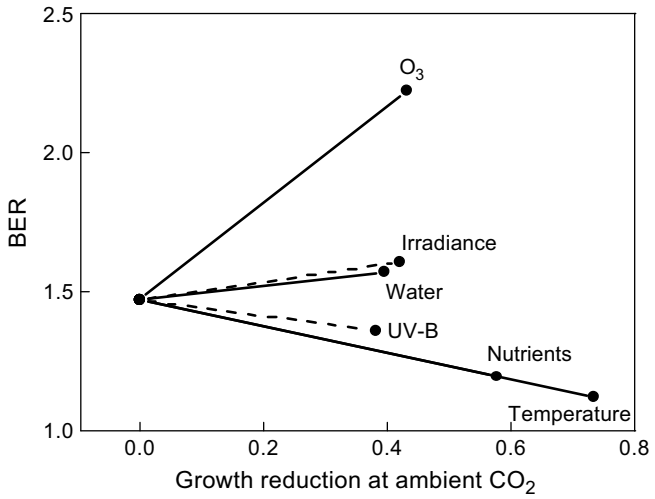


Figure 7 The average growth response of plants when elevated CO_2 is combined with sub-optimal levels of other environmental factors. Responses were calculated using a BER of 1.47 for plants grown under optimal conditions (Poorter, 1993). The average slope was calculated from the data of Figure 6 and the average reduction in growth at 35 Pa CO_2 as calculated in the compiled literature

the field situation. First, the UV-B levels used in the experiments are generally very high. Second, leaves developed under full sunlight have morphological and physiological adaptations that afford some protection from UV-B (Teramura and Murali, 1987). Consequently, plants in growth chambers, in which the visible light levels (*photosynthetically active radiation*) (*see* **PAR (Photosynthetically Active Radiation)**, Volume 2) are about two times lower than in field conditions may be more sensitive to UV-B than those in the field. The information on combined effects of elevated CO_2 and SO_2 or nitrogen (N) deposition is so limited that it precludes any generalized statement.

To summarize the effects of elevated CO_2 at sub-optimal growth conditions, we plotted the average BER values against the average growth reduction observed in the compiled literature (Figure 7). The growth enhancement by elevated CO_2 is severely reduced at low temperatures or poor nutrient supply. On the contrary, it is increased at low soil water availability, but the rise in BER is only small. The interaction between elevated CO_2 and O_3 shows the most dramatic response. The loss of biomass due to elevated ozone is, on average, more than compensated in the presence of elevated CO_2.

AN OUTLOOK

Up to now we have focused on experiments that have been carried out with individual plants in the vegetative stage. When extrapolating these results to the field situation, the following considerations should be taken into account. First, we have focused on average responses based on underlying data that are quite variable, partly due to the different growth conditions between experiments. Second, part of the increase in biomass is due to the accumulation of high levels of non-structural carbohydrates. This part of the extra-achieved biomass does not necessarily add to the performance of the plant and therefore may be regarded as non-functional in a number of cases. Third, generative biomass (seeds, fruits) is often stimulated to a lesser extent than vegetative biomass, at least in crop species (26% on average, Kimball, 1986b). In addition, it is not necessarily true that the species stimulated strongly in the vegetative phase will also be those that are stimulated relatively stronger in generative biomass (Ackerly and Bazzaz, 1995). Fourth, plants growing in monoculture have been shown to respond less than individually grown plants (Wayne and Bazzaz, 1995). Finally, most research was carried out with plants that have not been able to adapt genetically to high CO_2 levels.

How do growth responses of laboratory-grown individual plants, grown at optimal conditions, compare to those in the field? The above-ground biomass of monocultures of plants grown under agricultural field conditions increases by 20–40% due to elevated CO_2. This increase is somewhat smaller than that found in Figure 3. On the basis of the strong interaction with nutrients (Figure 7), it can be expected that natural types of vegetation, where plants are generally growing under far less fertile conditions, will show even smaller responses in biomass production. Current information shows that this is indeed the case, above-ground biomass in CO_2 enriched grasslands being increased by 14% on average (Mooney *et al.*, 1999). A discussion on the effects of elevated CO_2 on ecosystem functioning is given by Körner (*see* **CO_2 Enrichment: Effects on Ecosystems**, Volume 2).

ACKNOWLEDGMENTS

We thank Hans Cornelissen for providing part of the data compiled here and Jeremy Barnes, Ton Gorissen, Paskal Niklaus and Ineke Stulen for advice during the preparation of the manuscript.

REFERENCES

Ackerly, D D and Bazzaz, F A (1995) Plant Growth and Reproduction Along CO_2 Gradients: Non-linear Responses and Implications for Community Change, *Glob. Change Biol.*, **1**, 199–207.

Bunce, J A (1994) Responses of Respiration to Increasing Atmospheric Carbon Dioxide Concentrations, *Physiol. Plant.*, **90**, 427–430.

Bunce, J A (1997) Variation on Growth Stimulation by Elevated Carbon Dioxide in Seedlings of some C3 Crop and Weed Species, *Glob. Change Biol.*, **3**, 61–66.

Cornelissen, J H C, Carnell, A L, and Callaghan, T V (1999) Generalities in the Growth, Allocation and Leaf Quality Responses to Elevated CO_2 in Eight Woody Species, *New Phytol.*, **141**, 401–409.

Curtis, P S and Wang, X (1998) A Meta-analysis of Elevated CO_2 Effects on Woody Plant Mass, Form, and Physiology, *Oecologia*, **113**, 299–313.

Idso, K E and Idso, S B (1994) Plant Responses to Atmospheric CO_2 Enrichment in the Face of Environmental Constraints: a Review of the past 10 years' Research, *Agric. For. Meteorl.*, **69**, 153–203.

Kimball, B A (1986a) CO_2 Stimulation of Growth and Yield Under Environmental Restraints, in *Carbon Dioxide Enrichment of Greenhouse Crops Vol. II. Physiology Yield and Economics*, eds H Z Enoch and B A Kimball, CRC Press, Boca Raton, FL, 53–67.

Kimball, B A (1986b) Influence of Elevated CO_2 on Crop Yield, in *Carbon Dioxide Enrichment of Greenhouse Crops Vol. II. Physiology Yield and Economics*, eds H Z Enoch and B A Kimball, CRC Press, Boca Raton, FL, 105–115.

Körner, C, Pelaez-Riedl, S, and Van Bel, A J E (1995) CO_2 Responsiveness of Plants: a possible link to Phloem Loading, *Plant, Cell Environ.*, **18**, 595–600.

Mooney, H A, Canadell, J, Chapin, F S, Ehleringer, J R, Körner, C, McMurtrie, R E, Parton, W J, Pitelka, L F, and Schulze, E D (1999) Ecosystem Physiology Responses to Global Change, in *The Terrestrial Biosphere and Global Change*, eds B Walker, W L Steffen, J Canadell, and J S I Ingram, Cambridge University Press, Cambridge.

Pérez-Soba, M, Dueck, T A, Puppi, P, and Kuiper, P J C (1995) Interactions of Elevated CO_2, NH_3 and O_3 on Mycorrhizal Infection, Gas Exchange and N Metabolism in Saplings of Scots Pine, *Plant Soil*, **176**, 107–116.

Poorter, H (1993) Interspecific Variation in the Growth Response of Plants to an Elevated Ambient CO_2 Concentration, *Vegetatio*, **104/105**, 77–97.

Poorter, H (2001) What Functional Groups may add in Understanding the Growth Response of Plants to Elevated CO_2, *New Phytol.*, in preparation.

Poorter, H and Nagel, O (2000) The Role of Biomass Allocation in the Growth Response of Plants to Different Levels of Light, CO_2, Nutrients and Water: a Quantitative Review, *Austr. J. Plant Physiol.*, **27**, 595–607.

Poorter, H, Van Berkel, Y, Baxter, B, Bel, M, Den Hertog, J, Dijkstra, P, Gifford, R M, Griffin, K L, Roumet, C, and Wong, S C (1997) The Effect of Elevated CO_2 on the Chemical Composition and Construction Costs of Leaves of 27 C_3 Species, *Plant Cell Environ.*, **20**, 472–482.

Poorter, H and Pérez-Soba, M (2001) The Growth Response of Plants to Elevated CO_2 under Non-optimal Environmental Conditions, *Oecologia*, submitted for publication.

Reekie, E G, MacDougall, G, Wong, I, and Hicklenton, P R (1998) Effect of Sink Size on the Growth Response to Elevated Atmospheric CO_2 within the Genus Brassica, *Can. J. Bot.*, **76**, 829–835.

Stulen, I and Den Hertog, J (1993) Root Growth and Functioning under CO_2 Enrichment, *Vegetatio*, **104/105**, 99–115.

Teramura, A H and Murali, N S (1987) Intraspecific Differences in Growth and Yield of Soybean Exposed to Ultraviolet-B Radiation under Greenhouse and Field Conditions, *Environ. Exp. Bot.*, **26**, 89–95.

Turcsányi, E, Cardoso-Vilhena, J, Daymond, J, Gillespie, C, Balaguer, L, Ollerenshaw, J H, and Barnes, J D (2000) Impacts of Tropospheric Ozone: Past, Present and Likely Future, in *Trace Gas Emissions and Plants*, ed S N Singh, Kluwer Academic Publishers, The Netherlands, 249–272.

Wayne, P M and Bazzaz, F A (1995) Seedling Density Modifies the Growth Response of Yellow Birch Maternal Families to Elevated Carbon Dioxide, *Glob. Change Biol.*, **1**, 315–324.

Winter, K and Lovelock, C E (1999) Growth Responses of Seedlings of Early and Late Succesional Tropical Forest Trees to Elevated Atmospheric CO_2, *Flora*, **194**, 221–227.

Plant Migration

see Plant Dispersal and Migration (*Opening essay*, Volume 2)

Plant Ontogeny

Plant ontogeny describes the whole course of development of an individual plant. It represents the realization of the phenotype of the organism under the influence of its genotype (*see* **Phenotype/Genotype**, Volume 2). The ontogeny of a typical higher plant begins with fertilization and the determination by hormonal fields or external environmental factors of the polarity of the zygote. Differentiation and morphogenesis (the development of form and structure) proceeds in the shoot and root by terminal cell division and growth, and by enlargement through further cell division and maturation. At intervals (nodes), shoots, branches or leaves arise in specific patterns. At a time determined by the organism's genotype, and by environmental stimuli, the growth pattern of shoots undergoes massive changes with the induction of flowering. After flowering and seed production, the plant usually senesces, severs its leaves and dies – all actively controlled stages of the plant's ontogeny.

The control of all the events and stages of ontogeny – of plant development in space and time – is vastly complex. It stems from the plant's genotype, and operates largely through hormone interactions, metabolic gradients, and the influence of external stimuli such as photoperiodism, temperature and gravity. The patterns of ontogeny in lower plants are less complex, but are nevertheless governed by essentially similar controls, the complexity of

which depends only in small part on the morphogenetic complexity of the organism.

Ontogeny can be interrupted by accident and usually, if the damage is not too severe, reverts to an earlier stage and starts over. The ontogeny of plants, unlike that of animal early development, does not appear to reflect their phylogeny (evolutionary history). Plant ontogeny, in fact, is remarkable for the underlying or basic similarity of its patterns and control mechanisms in a very wide range of plant groups.

R G S BIDWELL *Canada*

Plants - Carbon Balance and Growth

Robert W Pearcy
University of California, Davis, CA, USA

Photosynthetic carbon gain is the process by which plants sequester both the energy used for growth and the carbon necessary for actually constructing the plant. The balance of the gains in carbon and the losses of carbon due to the respiration involved in constructing and maintaining the plant body and in the acquisition of other resources such as water and nutrients is referred to as the carbon balance. Carbon balances can be measured for leaves, for branches or for whole plants. When determined for leaves over their life span, it expresses the net profit (measured in carbon units) derived from the leaf. The net profit is the carbon that, when allocated within the plant, can be used for growth, maintenance processes and reproduction. Carbon balances for whole plants are directly related to their growth rate.

*The carbon gain of a leaf is a function of its photosynthetic capacity and the effects of the environment that limit photosynthesis. Photosynthetic capacity is determined principally by the concentrations of photosynthetic enzymes in the leaf and the limitations imposed by the stomata to diffusion of carbon dioxide (CO_2) into the leaf. Environmental limitations such as temperatures that are too hot or too cold, light that is too low, or stresses such as drought can greatly reduce the realized photosynthetic rate. Plants also differ in their photosynthetic pathway for carbon fixation and this has important implications for global change. Photosynthesis of plants with the C_3 photosynthetic pathway increases much more strongly with increasing atmospheric CO_2 concentrations than does photosynthesis in plants with the C_4 pathway. This difference is projected to make C_4 species less competitive relative to C_3 species and lead to a future expansion of C_3 species (*see ***C_3 and C_4 Photosynthesis**, *Volume 2).*

The growth rate of a plant is a function of the carbon balance of its leaves and pattern by which this carbon is allocated to new structures. Investment in increasing the leaf area of the plant drives growth because it leads to more carbon gain. Increasing atmospheric CO_2 increases growth of C_3 plants via its stimulation of photosynthesis but this stimulation is partially offset by a lower allocation to leaf area production. Growth of C_4 plants exhibits only a small stimulation by high CO_2 because their photosynthetic carbon gain is only slightly stimulated. Carbon balance provides a conceptual framework for understanding the response of plant growth to climate change as well as a tool for assessing its quantitative impacts.

CARBON GAIN BY LEAVES

The photosynthetic capacity (A_{max}) of a leaf is defined as its net CO_2 assimilation rate under conditions where light is saturating, the temperature is optimum, the relative humidity is high and CO_2 and O_2 concentrations are within the normal range encountered during growth. It is the best photosynthetic performance a plant can achieve given its genetic potential and the conditions under which it has grown. Photosynthetic capacity can be expressed either per unit leaf mass or per unit leaf area. On a per unit mass basis it expresses the rate of return on investment in the leaf. However, the solar radiation that drives photosynthesis is measured on an area basis so that for many purposes such as scaling the values up to whole canopy photosynthesis, expression of photosynthetic rates on an area basis is advantageous. For either basis of expression, photosynthetic capacities among terrestrial higher plants vary over an approximately 100-fold range. Generally, species from resource-rich environments have higher photosynthetic capacities than those from resource-poor environments where water, nutrients or other environmental factors may be strongly limiting.

Leaf photosynthetic capacity is determined primarily by the concentrations and catalytic activities of enzymes involved in photosynthesis and to a lesser extent by the limitations on diffusion of CO_2 to the site of photosynthesis. These physical diffusion limitations are imposed by laminar boundary layer over the leaf, the stomata in the leaf surface and by the anatomical architecture of the leaf mesophyll. For leaves having the C_3 photosynthetic pathway the relationship between photosynthetic rate and the CO_2 concentration in the intercellular air spaces (c_i) typically exhibits a strong, almost linear increase up to about 200–250 ppm CO_2 followed by a gradually decreasing slope at higher CO_2 (Figure 1). This relationship is often referred to as an A/c_i curve. Elegant studies by von Caemmerer and Farquhar (1981) established that this

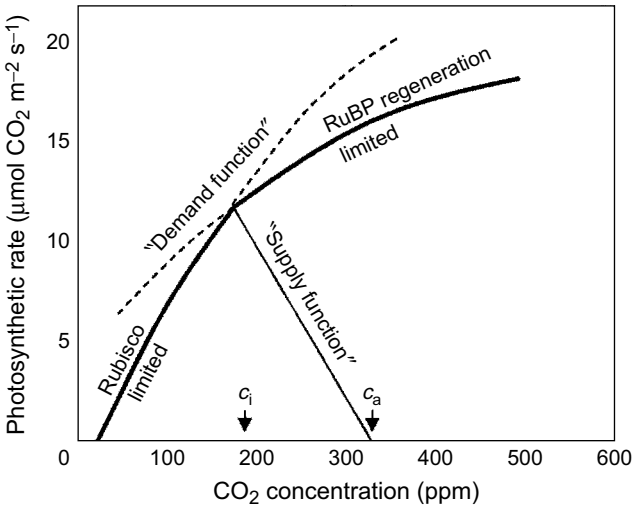


Figure 1 The dependence of photosynthetic rate on CO_2 concentration in the intercellular air spaces (bold line). This is the demand function for CO_2 while the supply function as determined by the stomatal conductance is shown by the thin diagonal line. Also shown are the Rubisco limited and RuBP-regeneration limited parts of the response

curve could be understood in terms of the kinetics of the primary CO_2 fixing enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) which catalyzes the condensation of ribulose bisphosphate (RuBP) with CO_2 to yield the first products of CO_2 assimilation. The carboxylation rate in a leaf depends on the amount of active enzyme present and any limitations by the substrates CO_2 and RuBP. In addition, oxygen in the atmosphere competitively inhibits carboxylation by Rubisco, leading to photorespiration.

At low c_i values, photosynthesis is limited by the CO_2 concentration itself while at high c_i it is limited primarily by the rate at which the acceptor molecule RuBP can be regenerated. RuBP regeneration is thought to be limited primarily by photosynthetic electron transport capacity that is responsible for generating the energy and reducing power required for carbon metabolism. The combined curve, reflecting limitations by Rubisco activity and RuBP regeneration activity, represents the demand function for CO_2 established by photosynthetic metabolism. The supply function is set mostly by the stomata and is represented by the diagonal line connecting the ambient CO_2 concentration (c_a) with c_i. The negative slope of this diagonal line is proportional to the stomatal conductance to CO_2. As the stomatal conductance decreases, the line rotates to the left and the new photosynthetic rate of a leaf is given by the intersection with the A/c_i curve. For unstressed plants, c_i is often close to the transition from strictly CO_2-limited to RuBP regeneration-limited photosynthesis. At this point, the investment in the photosynthetic apparatus is thought to be optimal because Rubisco and RuBP regeneration capacity are co-limiting rather than one or the other being present in excess of requirements. Increasing the Rubisco activity in the leaf increases the initial slope of the curve giving a higher photosynthetic rate at low c_i values. Achieving an increase at high c_i values additionally requires an increase in RuBP regeneration capacity. Although a small increment in photosynthetic rate will result even if the stomatal conductance does not change, realizing the full potential increase in photosynthetic rate also requires a concomitant increase in stomatal conductance. Indeed interspecific comparisons reveal a strong correlation between stomatal conductance and photosynthetic capacity (Körner *et al.*, 1979).

The variation in photosynthetic capacity between plants with different leaf characteristics has been shown to have a consistent global pattern related to investments in biomass for a given leaf area, leaf nitrogen contents and leaf longevity (Reich *et al.*, 1997). Long-lived leaves tend to be thick with substantial investments in cell wall material, giving a low specific leaf area (SLA, leaf area per unit of leaf mass) and also to have low nitrogen concentrations. At the other end of the spectrum, short-lived leaves tend to have high photosynthetic capacities, high leaf nitrogen contents per unit leaf mass and high SLA. Indeed, when photosynthetic capacity is plotted versus leaf nitrogen and SLA, a curved plane circumscribing the full range of variation results (Figure 2). The different leaf types then fall out on different parts of this plane. These patterns were essentially the same for all ecosystems or biome of origin suggesting that the same tradeoffs are involved despite large differences in climate. There were no species with thin short-lived leaves and low nitrogen, or with thick leaves with high nitrogen, in either a wide range of species measured by Reich and his colleagues or from published results from other researchers. Photosynthetic capacity and leaf nitrogen concentration are highly correlated because of the investments of nitrogen required in synthesizing photosynthetic enzymes. The correlation between photosynthetic capacity and SLA may be driven by several factors. Long-lived leaves may require substantial investment in cell walls and other structural components to withstand stresses and to resist herbivory. These thick dense leaves may cause internal shading within the leaf, limiting the light reaching some of the chloroplasts and they may also have greater internal diffusion limitations, limiting the supply of CO_2 to the chloroplasts. Selective pressures for increasing photosynthetic capacity by increasing photosynthetic enzyme concentrations would run up against these limitations. It could also be that thick leaves with high nitrogen would be subjected to a greater probability of herbivory over their long life, limiting their realized longevity and selecting against their evolution. On the other hand, short-lived leaves with a low photosynthetic capacity may be disadvantageous because of a relatively low payback over their life span.

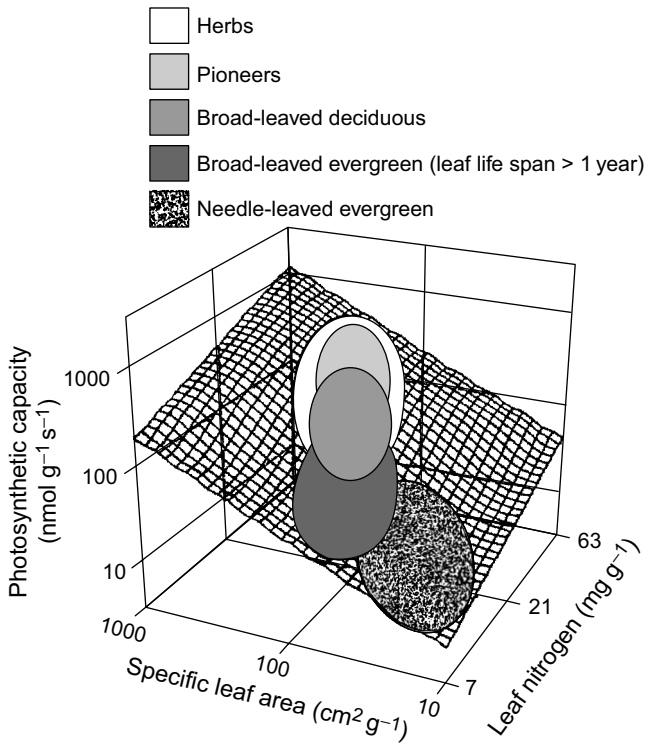


Figure 2 The dependence of photosynthetic capacity per unit leaf mass on the speci c leaf area and leaf nitrogen content. The shaded areas circumscribe the ranges of values for each leaf type independent of the biome or ecosystem of origin from the arctic tundra to tropical forests. Note that all axes are $\log_{10}$ scale [from Reich *et al.* (1997) reproduced by the permission of The National Academy of Sciences of the United States of America]

On a leaf area basis, photosynthetic capacity and leaf nitrogen per unit leaf area are positively correlated, and leaf life-span and leaf photosynthetic capacity are generally negatively correlated. But when expressed on this basis, there is a much greater divergence among leaf types and across biomes in the scaling of the relationships.

The photosynthetic capacity of a leaf can be altered by the growing conditions such as the prevailing temperatures or light. Leaves developing in the shade have a low photosynthetic capacity per unit leaf area as compared with sun-developing leaves. This acclimation of photosynthesis to the light environment appears to involve a tradeoff between maximizing light capture and maximizing photosynthetic capacity. Thinner leaves, but with greater chlorophyll contents per unit leaf mass allows for greater total leaf area on the plant, maximizing light capture in shaded environments. In high light, thicker leaves having a higher concentration of photosynthetic enzymes per unit area may maximize photosynthesis. In many cases the photosynthetic capacity per unit leaf mass is similar in sun and shade-acclimated leaves. This pattern is somewhat different than for the cross-species comparison presented above but the two- to four-fold variation in photosynthetic capacity between sun and shade leaves of a species is but a small fraction of the interspecific variation observed globally.

Acclimation can also occur in response to an increase in the atmospheric CO_2 concentration. In many species, growth in elevated as compared with normal ambient CO_2 concentrations results in a downward shift in the A/c_i curve (Drake *et al.*, 1997). This shift results in lower photosynthetic rates at the normal ambient CO_2 concentration and less of a stimulation in photosynthetic rates at the elevated CO_2 concentration than if no acclimation occurred. The decrease in the A/c_i curve is due to decreased investments in Rubisco and RuBP regeneration capacity. The resulting lower leaf nitrogen also has the additional effect of making the leaves less nutritious for herbivores. In addition, stomata often respond to elevated CO_2 concentrations, resulting in a reduced stomatal conductance, further limiting photosynthetic rates. Acclimation may function to balance the supply of carbohydrates from photosynthesis to the capacity to utilize them in growth. Not all species exhibit this acclimation response and some show a slight stimulation. Acclimation to CO_2 is usually greater when there are additional limitations to growth such as low nitrogen or water availability (Drake *et al.*, 1997). However the effect can be quite complicated since acclimation also in part relieves the demand for these resources.

Limitations in the environment often cause the realized photosynthetic rate of a leaf to be substantially lower than the photosynthetic capacity. These limitations occur because the light is too low, temperature is too high or too low for optimum photosynthesis, or the humidity is too low, causing stomatal closure. In addition, stresses such as drought, salinity, etc. can strongly limit the photosynthetic rate. The nature and extent of these limitations varies widely between ecosystems. In temperate deciduous forests low light due to cloudiness, short day lengths in the spring and fall and cool temperatures during this time can limit the annual leaf photosynthesis to 35–45% of its potential. In the desert, where water stress and high temperatures are the principal limitations, annual leaf photosynthesis may only be 20–30% of its potential.

ALTERNATE PHOTOSYNTHETIC PATHWAYS

Terrestrial plants can be grouped according to the three photosynthetic pathways known to occur in them. The C_3 pathway occurs in the greatest proportion of species while the C_4 pathway is found in grasses with an evolutionary origin in the tropics, many annual herbs native to warm environments and some arid habitat shrub species. The C_4 pathway is found in evolutionarily advanced genera and perhaps evolved multiple times as the angiosperms diversified. The third, crassulacean acid metabolism (CAM) is found in many succulent species including succulent

epiphytes in tropical forests and desert succulents. CAM plants are therefore only a minor component of the world's vegetation. C_3 plants account for about 85% of the world's photosynthesis. Crops such as wheat, barley and soybeans have C_3 photosynthesis as do virtually all tree species. C_4 grass species make especially important contributions in tropical grasslands and savannas whereas northern grasslands are dominated by C_3 grasses. The important crops, maize, sorghum and sugar cane as well as all tropical forage grasses are C_4 species (*see* **C_3 and C_4 Photosynthesis**, Volume 2).

In C_3 photosynthesis, all mesophyll cells have essentially the same complement of photosynthetic enzymes including Rubisco and the enzymes necessary for RuBP regeneration (Figure 3). In C_4 photosynthesis, two cell types are present with the mesophyll cells containing the enzyme PEP carboxylase that first fixes CO_2 into a C_4 acid. A cycle then operates between the mesophyll and bundle sheath that acts as a CO_2 pump, resulting in high CO_2 concentrations (>2000 ppm) in the bundle sheath cells where Rubisco is localized in C_4 plants. These high CO_2 concentrations cause Rubisco to operate efficiently and cause its oxygenase activity and therefore photorespiration to be suppressed.

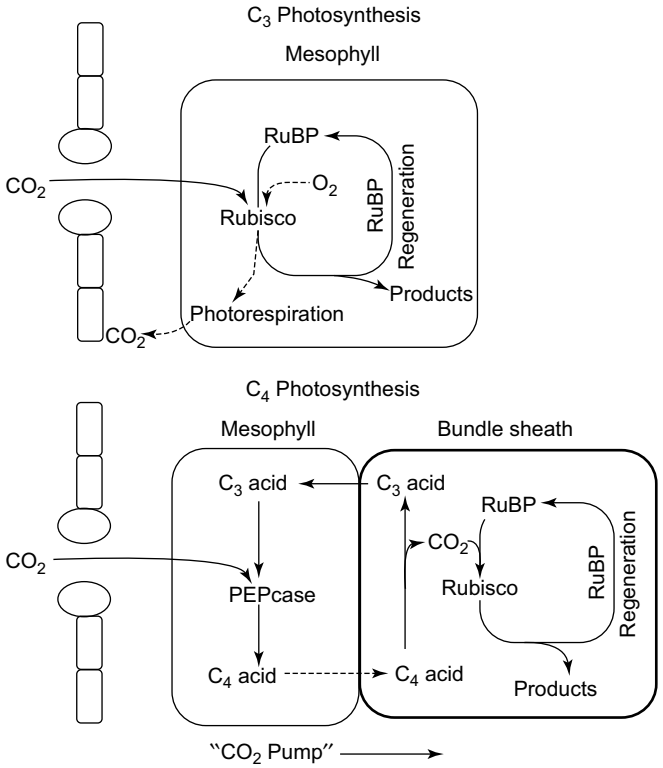


Figure 3 The pathways for CO_2 uptake and assimilation in C_3 and C_4 photosynthesis. Diffusion through the stomata is shown on the left side. In C_4 photosynthesis CO_2 is first fixed via the enzyme PEP carboxylase in the mesophyll. The C_4 cycle operating between the mesophyll and the bundle sheath acts as a CO_2 pump

C_4 plants can efficiently utilize low CO_2 concentrations but they tend to saturate at moderate CO_2 concentrations. Thus, in comparison to C_3 plants, C_4 plants exhibit a lower stimulation of photosynthetic rates with rising CO_2 concentrations in the atmosphere. The ability to efficiently utilize low CO_2 concentrations means that C_4 plants can maintain high photosynthetic rates with lower stomatal conductances than C_3 plants, giving them higher water use efficiency (ratio of photosynthesis to transpiration). Since increasing temperature stimulates photorespiration in C_3 plants, C_4 plants generally have an advantage in terms of photosynthetic performance in warmer climates. Indeed C_4 grasses reach their greatest importance in environments where most of the precipitation comes in the warm season. In these environments the high water use efficiency coupled with the photosynthetic advantage at high temperatures strongly favors C_4 over C_3 grasses.

The different responses to atmospheric CO_2 has important implications for the response of C_3 and C_4 plants to climate change scenarios. In C_3 plants, both temperature and CO_2 are important factors because they interact to determine the ratio of carbon flow through the wasteful process of photorespiration as compared with the assimilatory process of photosynthesis (Figure 4). As a result of these differences, C_4-dominated plant communities are hypothesized to have expanded during glacial periods in the Pleistocene when atmospheric CO_2 concentrations may have been only 190 ppm (Ehleringer *et al.*, 1991). The currently increasing atmospheric CO_2 should favor C_3 over

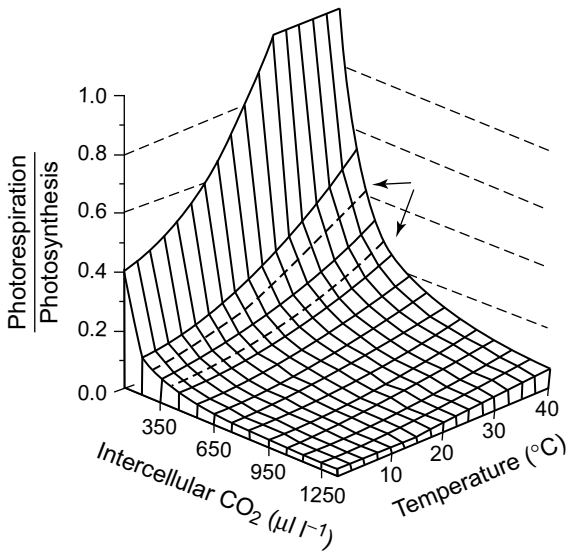


Figure 4 The dependence of the ratio of photorespiration to photosynthesis on temperature and intercellular CO_2 concentration. The arrows and dashed lines delineate the range of intercellular CO_2 concentrations typically occurring in leaves of C_3 species. (Reprinted from Ehleringer *et al.* (1991) © 1991 with permission from Elsevier Science)

C_4 species. This advantage for C_3 photosynthesis will also depend on any global warming. Moreover, local or regional changes in precipitation and especially its seasonality could be expected to further influence the shifts in distribution and abundance of C_4 versus C_3 species.

THE CARBON BALANCE OF LEAVES

Leaves must return more carbon from their photosynthetic activities than was required to construct them in order to provide any benefit to the plant. Initially during development, photosynthetic capacities are low and the leaf imports carbon for construction (Figure 5). Then photosynthetic capacity increases up to a plateau where it remains until senescence occurs. The realized daily leaf CO_2 exchange is in part a function of the photosynthetic capacity but it also depends on how favorable the environment is. Cloudiness, too warm or too cool temperatures, or periods of drought can significantly reduce it. Ultimately sufficient carbon is gained to meet the leaf construction costs and thereafter further carbon gain is net profit that can be used for growth. This can occur within a few days to a few weeks for herbaceous plants and deciduous trees, which have relatively high photosynthetic capacities and low construction cost leaves, but it can take considerably longer for evergreen species with thick, long-lived leaves. Evergreen leaves such as the needles of coniferous species gain little or no carbon over the winter months, but evergreen species do have the advantage of already having a canopy in the spring and of maintaining it in the fall. These are times when deciduous species may either be leafless or just developing or senescing leaves and therefore incapable of significant carbon gain. Also, long-lived evergreen leaves can continue to

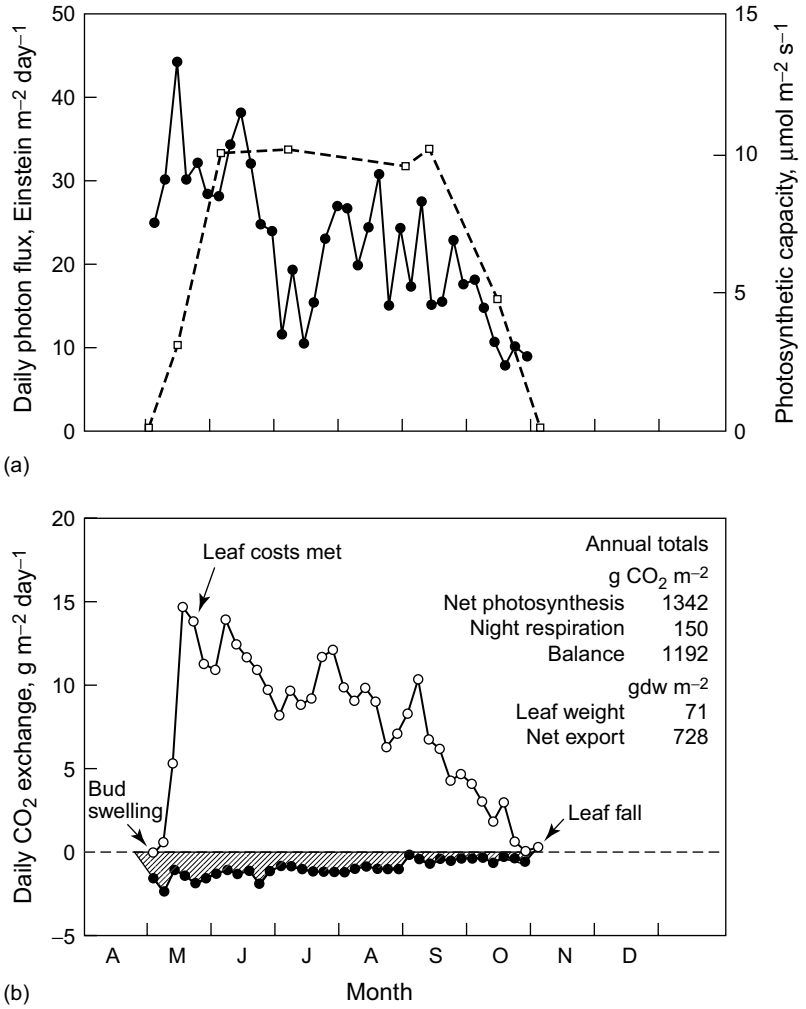


Figure 5 (a) Annual course of photosynthetically active radiation (●) and leaf photosynthetic capacity (□) for the deciduous tree, *Acer campestre* during 1980. (b) the daily CO_2 uptake (○) and nighttime respiration (●) for this species. Also shown are the annual total balances for net carbon gain and the resulting net export (pro t) from an *A. campestre* leaf [from Pearcy *et al.* (1987) as adapted from data of Kuppers (1984, 1985)]

return a profit over several years, albeit at a decreasing rate because of age-related decreases in photosynthetic capacity and because the leaves are typically increasingly shaded by new growth. At the other extreme are leaves of fast-growing weedy species and herbaceous crops with life spans as short as only 1–2 months. Leaves on these species are rapidly overtopped and shaded by newly produced leaves and thus the period when they return a high carbon gain is short. Indeed leaves held past the point where they become so shaded that net photosynthesis is negative begin to return a loss rather than a profit. In addition to minimizing losses, a rapid leaf turnover is also advantageous in these rapidly growing species because during senescence 40–60% of the nitrogen in the leaf is mobilized and exported. This nitrogen can then be utilized again in the upper newly expanding leaves where it will return the greatest benefit in terms of useful photosynthetic capacity.

Respiratory losses required for growth and maintenance of the leaf offset a portion of the carbon gain of leaves, reducing the carbon balance. In addition, there is an alternative pathway of respiration that appears wasteful in terms of energy but may be important for controlling levels of carbohydrates or balancing the needs for energy versus carbon compounds required to construct the plant. Early in their development, leaves exhibit elevated respiration rates because of the metabolic activities required for their growth. In addition, respiration is required throughout the leaf's life span to provide the metabolic energy needed to maintain the leaf cells, to synthesize sucrose and other sugars, and to load the phloem so that these sugars can be exported to other parts of the plant. The respiration rate of leaves is typically only 5–10% of its photosynthetic capacity, but under environmental conditions where the photosynthetic rate is strongly restricted, such as in the shade, it can be a much higher fraction. Under shaded conditions, mechanisms that reduce the respiration rate can be at least as important for the leaf carbon balance as those which enhance the photosynthetic rate in low light. In the short term, respiration rates increase with temperature, exhibiting an approximate doubling every 10 °C but in the longer term, physiological adjustment results in homeostasis of rates. Thus, global warming may have relatively little direct impact on respiration rates. Both increases and decreases in leaf respiration rates in response to increasing atmospheric CO_2 have been observed but these appear to be mostly indirect effects (Drake *et al.*, 1997). For example, when photosynthesis is stimulated by increased atmospheric CO_2, extra carbohydrate processing and phloem loading may be required, raising leaf respiratory rates.

PHOTOSYNTHESIS AND GROWTH

Understanding how the carbon balance of leaves scales up to whole plant carbon gain requires knowledge of the total leaf area and how light and hence photosynthesis is distributed on this leaf area in order to obtain a total for the whole plant. In addition, knowledge is required of the respiratory losses of the leaves and of the non-photosynthetic plant parts such as stems and roots. These processes are dynamic and continually change during development, with each process changing at a different rate. Leaf photosynthetic capacity changes with age and as new leaves develop, older leaves may become shaded, modifying their environment and reducing their photosynthetic rate. Ultimately these leaves will be shed, modifying the leaf area. Moreover, changes in allocation of carbon and nutrients can occur within the plant during its growth, increasing or reducing the rate at which new leaf area is produced. Reinvestment in leaves is especially important since this drives the tendency for plants to increase exponentially in size. Competing with this reinvestment is the need for sufficient investment in shoots and roots, which must supply adequate water, nutrients and support. Monsi (1960) demonstrated that because of the exponential tendency of the growth of plants, relatively small changes in either the photosynthetic rate of the leaves or in the allocation of carbon to leaf area production can result over time in large differences in plant size.

The complexity of determining whole-plant carbon balances has meant that only a few species have been studied in this way, especially over longer time frames. Instead most knowledge of how the growth of plants responds to the environment and differs between species has come from growth analysis. In growth analysis, plants from a uniform population are harvested at sequential times and the net whole-plant carbon gain is inferred from the increase in biomass. Expressed in dimensions of biomass produced per unit leaf area per unit time, it is referred to as the net assimilation rate (NAR). During growth analysis, the partitioning of biomass into leaves, stems and roots is also determined. Of most importance is the leaf area ratio (LAR) in dimensions of leaf area per unit biomass, which is a measure of how biomass is partitioned to produce the leaf area on the plant. It is essentially a measure of the leafiness of the plant. The product of LAR and NAR is the relative growth rate (RGR) of the plant, in dimensions of mass of plant produced per unit mass of plant per unit time. LAR and NAR are not completely independent variables, however. The major determinant of LAR is the SLA, which, as discussed above (Figure 2), also is correlated with photosynthetic capacity.

Most growth analysis studies have concentrated on the seedling stage because at this time all the leaves are more or less functionally equivalent with little self-shading within the canopy. Measured under conditions of adequate light, water, nutrients, and optimal temperatures the RGR of seedlings is taken as the maximum potential RGR for the species. Later, self-shading reduces the NAR; and a change in biomass partitioning towards greater investments

in stems and roots tends to reduce LAR. Thus, RGR declines.

Comparisons of fast and slow growing species of similar growth form (for example, herbaceous C_3 photosynthetic species) show that the fast growing species typically have a higher LAR and especially a higher SLA than the slow growing species (Lambers and Poorter, 1992). In contrast, there is in most cases little or no correlation between RGR and NAR for plants grown in the same optimal environment. Differences in NAR are, however, often more explanatory of the differences in RGR of plants of a single species grown in different environments. Thus, the growth stimulation in elevated CO_2 can generally be accounted for by an increase in NAR through the CO_2 effects on photosynthesis. However, the environment may also influence LAR and especially SLA so the effects are often complex. For example, shade-grown plants produce leaves with a high SLA, which partially compensates for the strong reductions in NAR in the shade as compared with the sun. As a result, the RGRs of shade and sun plants exhibit a smaller difference than would be predicted by the differences in NAR or photosynthetic rate per unit leaf area alone. Comparisons of herbaceous C_3 species grown under projected elevated atmospheric CO_2 concentrations reveal on average 1.47 larger dry mass at final harvest for those grown at elevated atmospheric CO_2 (Poorter *et al.*, 1996). This was accounted for by a modest (7%) increase in RGR, a 22% increase in NAR, which was in line with expectations for the stimulation of C_3 photosynthesis by elevated CO_2, and an 11% decline in LAR.

Ultimately, part of the carbon fixed in photosynthesis is partitioned into reproduction. This fraction varies widely, approaching 60% of the biomass in some annuals while being as little as 1–10% in some trees. Increased atmospheric CO_2 has been shown to increase reproductive yields in some C_3 pathway annual crops such as wheat, and cotton (Pinter *et al.*, 1996). On the other hand, increased temperatures have been shown to significantly decrease reproductive yields in rice, and therefore to partially offset or even eliminate the stimulation due to CO_2 (Baker *et al.*, 1996). Other factors such as nutrient and water supply are also important influences on the extent of CO_2 stimulation of reproductive yields in these crops. Thus, whether or not the potential for increased reproductive output is realized depends on complex interactions with other environmental variables. It is also likely to be highly species specific and depend on the phenological and developmental effects of increased CO_2 or temperature. Even less is known about the response of reproductive output in perennial plants such as trees and shrubs to increased CO_2 though it could be expected that additional carbohydrates would favor enhanced reproductive output. As with annuals, interactions with nutrient and water supply are undoubtedly important but little understood for these plants.

REFERENCES

Baker, J T, Allen, L H, Boote, K J, and Pickering, N B (1996) Assessment of Rice Responses to Global Climate Change: CO_2 and Temperature, in *Carbon Dioxide and Terrestrial Ecosystems*, eds G W Koch and H A Mooney, Academic Press, San Diego, CA, 265–282.

Drake, B G, Gonzalez-Meler, M A, and Long, S P (1997) More Efficient Plants: a Consequence of Rising Atmospheric CO_2? *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **48**, 609–639.

Ehleringer, J R, Sage, R F, Flanagan, L B, and Pearcy, R W (1991) Climate Change and the Evolution of C_4 Photosynthesis, *Trends Ecol. Evol.*, **6**, 95–99.

Körner, C, Scheel, J A, and Baur, H (1979) Maximum Leaf Diffusive Conductance in Vascular Plants, *Photosynthetica*, **13**, 45–82.

Kuppers, M (1984) Carbon Relations and Competition Between Woody Species in a Central European Hedgerow. 1. Photosynthetic characteristics, *Oecologia*, **64**, 332–343.

Kuppers, M (1985) Carbon Relations and Competition Between Woody Species in a Central European Hedgerow. 3. Carbon and Water Balance on the Leaf Level, *Oecologia*, **65**, 94–100.

Lambers, H and Poorter, H (1992) Inherent Variation in Growth Rate Between Higher Plants: a Search for Physiological Causes and Ecological Consequences, *Adv. Ecol. Res.*, **23**, 188–261.

Monsi, M (1960) Mathematical Models of Plant Communities, in *Functioning of Terrestrial Ecosystems at the Primary Production Level*, ed F E Eckhardt, UNESCO, Paris, 418–473.

Pearcy, R W, Bjorkman, O, Caldwell, M, Keeley, J E, Monson, R L, and Strain, B (1987) Carbon Gain by Plants in Natural Environments, *BioScience*, **37**, 21–29.

Pinter, P J, Kimball, B A, Garcia, R L, Wall, G L, Hunsaker, D J, and LaMorte, R L (1996) Free-air CO_2 Enrichment: Responses of Cotton and Wheat Crops, in *Carbon Dioxide and Terrestrial Ecosystems*, eds G W Koch and H A Mooney, Academic Press, San Diego, CA, 215–249.

Poorter, H, Roumet, K, and Campbell, B D (1996) Interspecific Variation in the Growth Response to Elevated CO_2: a Search for Functional Types, in *Carbon dioxide, populations, and communities*, eds C Körner and F A Bazzaz, Academic Press, San Diego, CA, 375–412.

Reich, P B, Walters, M B, and Ellsworth, D S (1997) From Tropics to Tundra: Global Convergence in Plant Functioning, *Proc. Natl. Acad. Sci.*, **94**, 13 730–13 734.

von Caemmerer, S and Farquhar, G D (1981) Some Relationships Between the Biochemistry of Photosynthesis and the Gas Exchange of Leaves, *Planta*, **153**, 376–387.

Pollen as Climate Proxy

see Natural Records of Climate Change (Volume 1)

Population Sizes, Changes

Esa Ranta[1] and Veijo Kaitala[2]
[1] University of Helsinki, Helsinki, Finland
[2] University of Jyväskylä, Jyväskylä, Finland

Changes in population size are due to four separate processes, births, deaths, immigration and emigration, often expressed as per capita rates. The rates may (or may not) be independent from each other. Births and immigration increase population size, while deaths and emigration reduce it. Births and deaths are local processes whereas immigration and emigration are regional processes that connect population units to other similar units. Very few populations live in complete isolation. The relative importance of local and regional processes makes the local population either a source or sink population. Source populations send more individuals to nearby populations than they receive, and the fluctuation in source population size is governed by local population renewal. Sink populations behave in the opposite manner.

The net difference between birth and death rates is known as the per capita growth rate of the population. Often the growth rate of a local population is some function of the population size, know as density dependence. Density dependence may affect population growth rate directly or with delays. With delayed density dependence, the past population sizes affect the renewal process with lagged time. The upper limit for a population size is often given by local availability of resources. When this limit, the carrying capacity, is approached density dependency starts to reduce population growth rate due to increased competition on available resources.

The size of any given population is seldom constant but fluctuates due to various superimposed factors. Demographic stochasticity, most effective in small populations, is due to chance events in individual life histories. Environmental stochasticity, in turn, is seen as random (or temporally correlated) changes affecting either the carrying capacity of the population or the growth rate, or both. Stochasticity may cause local populations to become extinct even though the population would have survived in an undisturbed situation. Other causes of extinction include local catastrophes and population over exploitation, either by natural predators or by humans.

Populations are not homogeneously distributed over all areas suitable for independent renewal. Rather, they form networks where local units are connected with immigrating and emigrating individuals. Thus, dispersal of individuals may potentially rescue an extinct local population in an adjacent area. If dispersal is reduced, e.g., due to human-induced habitat fragmentation, the extinction rate of local populations is enhanced and this may lead to a loss of the entire population. Dispersal linkages also synchronize fluctuations of several local population units. The level of synchrony is often a function of proximity between the population sub-units. Nearby populations fluctuate in step, while more distantly located sub-units become increasingly out of phase. Another agent synchronizing fluctuations of populations is external disturbance. For example, adverse weather may cause all local populations to reduce in size at the same time. It is likely that synchronizing agents act in concert yielding a high level of region-wide synchronicity in dynamics of populations.

INTRODUCTION

Population ecology is interested in temporal changes in population size, known as *population dynamics* (Figure 1). Documented temporal changes are rich in their appearance ranging from cyclic and periodic dynamics to chaotic fluctuations, random walks and decreasing or increasing trends. Causes for such changes may be manifold. Even for such a limited data set as displayed in Figure 1, the list may include biotic interactions (e.g., predator and prey dynamics), local perturbations, large-scale environmental changes, human-induced modifications in habitat structure and over-harvesting. The challenge for ecology and population management is to understand what makes various populations fluctuate temporally the way they do.

The essence of population dynamics lies in two processes: births B and deaths D (Box 1). These processes may operate on current population density as well as on past population densities with time lags (Royama, 1992). Local populations change in size subject to demographic stochasticity (Figure 2b,c), hazards in life expectancy and breeding success of individual females (May, 1974; Hilborn and Mangel, 1997). In the classical definition, populations are understood as isolated units of interbreeding individuals (Emlen, 1977). However, populations rarely lie in entire isolation. The current view is (e.g., Tillman and Kareiva, 1997; Bascompte & Solé, 1998) that the landscape in which individuals live, is composed of a network of habitable areas. These local units are considered large enough to obey population renewal of their own, but they are also connected to other units via dispersing individuals. Dispersal can be decomposed into two opposing components, emigration, departure of individuals from their natal patch, and immigration, arrival of individuals from surrounding population sub-units to the focal unit.

Changes in population size from year to year can also be modulated with external noise (Figure 2e) and trends, which can be composed of regional and local elements (Box 1). Candidates for such modulators are many, but most often they are associated with local weather and regional, or global climate. The North Atlantic Oscillation

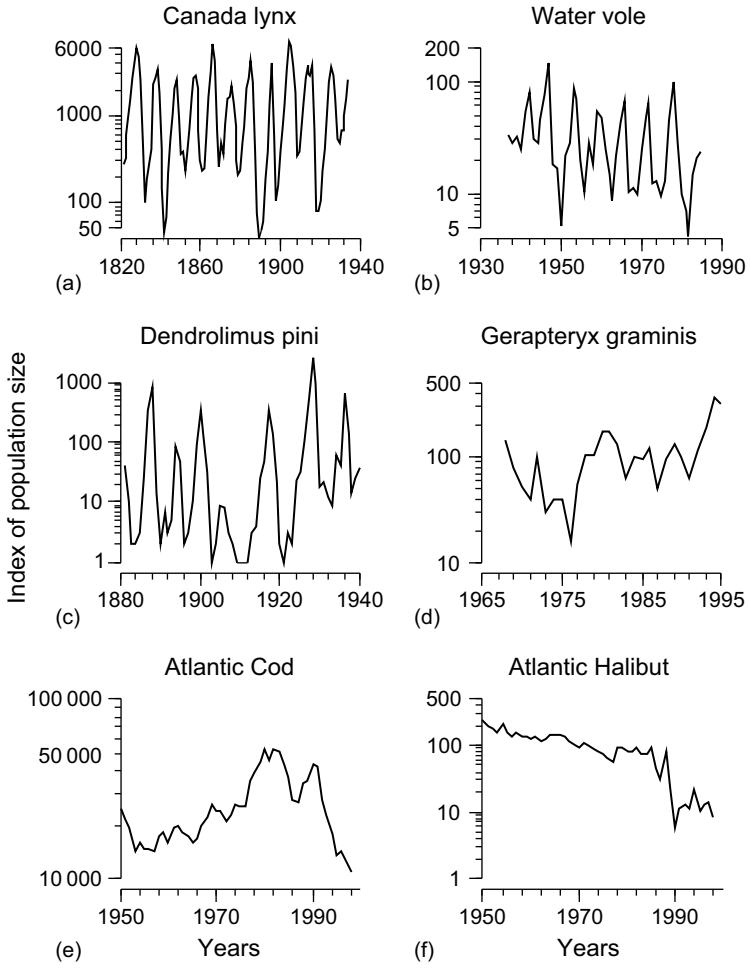


Figure 1 Examples of long-term population changes in mammals (a,b), insects (c,d) and sh (e,f). The Canadian lynx (trappers pelt returns) displays pronounced cyclic dynamics, periodicity is less clearly pronounced in the water vole dynamics (snap trap data). *Dendrolimus pini*, a forest pest, displays more or less irregular outbreaks while in *Gerapteryx graminis* there is no clear pattern of population uctuations (both trap data). The two Atlantic sh species are both targets of heavy sheries, and the past decade displays declining landing statistics (metric tonnes) in New England sh harbors. (Data in (a- d) from Ginzburg and Inchausti (1997), in (e) and (f) from http://www.st.nmfs.gov/st1/commercial/)

(NAO) is a good candidate for a large-scale modulator (*see* **North Atlantic Oscillation**, Volume 1). Furthermore, the anticipated global warming is a long-term change, which most certainly will affect all extant populations. This sets the stage for our endeavor to determine the significance of internal and external effects causing changes in population size.

Our view is that changes in population sizes are not due to population-specific intrinsic (density-dependent) processes only, but also to the changes in nearby populations (acting via density dependence in the focal unit), as well as to density-independent, or external processes of various scales. These factors modulate temporal changes in the size of local populations, even to the extent that their dynamics cannot be understood without referring to large-scale population dynamics. However, these facts are often rather hard to confirm with data on extant populations. Thus, computer simulations are used to demonstrate how population linkages, via dispersing individuals and external disturbances, or noise, may entirely change the intrinsic population dynamics.

SPACE, DISPERSAL LINKAGE AND DYNAMICS OF POPULATIONS

Two Examples

The argument of contemporary population ecology is that populations only rarely live in entire isolation (e.g., Bascompte and Solé, 1997). Rather, they are connected via dispersing individuals. Thus, individuals depart from and enter to local population sub units. This interplay keeps

Box 1 A few elementary models on population growth

For a change in population size, we write:

$$\Delta N = B + I - D - E, \tag{B1.1}$$

i.e., the change ΔN is the balance between births (B) and immigration (I) and deaths (D) and emigration (E). In discrete time t, the model can be expressed as:

$$\Delta N_t = N_{t+1} - N_t,$$

or

$$r_t = \frac{\Delta N_t}{N_t} = \frac{N_{t+1} - N_t}{N_t} = \frac{B + I - D - E}{N_t}. \tag{B1.2}$$

Here r_t is the *per capita* rate of increase (when $r > 0$ the population increases, with $r < 0$ the population decreases). Assuming that r is constant over all time intervals, a model for exponential growth of a population (Figure 2a) becomes:

$$N_{t+1} = N_t + rN_t \tag{B1.3}$$

One method to add demographic stochasticity into the equation is to take N'_{t+1} from random numbers obeying a Poisson distribution with the parameter N_{t+1}. An example of the effects of demographic stochasticity on exponential growth is shown in Figure 2(b,c).

No population can grow *ad in nitum* without any upper bounds. A way to add an upper limit for the population size is to assume a carrying capacity K for the population. One possible expression for bounded population dynamics is:

$$N_{t+1} = N_t + r\left(1 - \frac{N_t}{K}\right)N_t \tag{B1.4}$$

which is referred to as the logistic model for population growth. The behaviour of the logistic growth model is explored in Figure 2(d). The effect of environmental stochasticity, superimposed on K, on the emerging dynamics is illustrated in Figure 2(e).

A third model is the Ricker model (Ricker, 1954):

$$N_{t+1} = N_t \exp\left[r\left(1 - \frac{N_t}{K}\right)\right] \tag{B1.5}$$

The behaviour of the Ricker model is governed by a single parameter r and its dynamics are well-known (May, 1974). For this reason, the Ricker model has gained a bench-mark status in population ecology. The model has a stable equilibrium point when $r < 2$, from $2 < r < 2.526$ the model yields a two-point cycle, followed by more complex limit cycles (4-point, 8-point, etc.) and when $r > 2.692$ the dynamics assume chaotic behaviour.

The nal example is the Moran - Ricker model (Turchin, 1990), which - unlike the previous ones - acknowledges the in uence of a time-lag affecting the predicted population size:

$$N_{t+1} = N_t \exp(r + a_1 N_t + a_2 N_{t-1}) \tag{B1.6}$$

The parameters a_1 and a_2 take care of the direct and delayed density-dependency. Even this model can yield rich dynamics (Royama, 1992) ranging from stable dynamics to damped dynamics, to periodic and cyclic dynamics and to chaotic uctuations.

All population renewal models with one time lag can be written in a general form as:

$$N_{t+1} = f(N_t, \varepsilon_t) \tag{B1.7}$$

where f is a function mapping the population density at time t to time $t+1$, and ε represents environmental stochasticity, the Moran effect (e.g., a random deviate with zero mean). The spatial structure of a population can be written as:

$$N_{i,t+1} = (1 - m)f(N_{i,t}, \varepsilon_t) + \sum_{s,s\neq i} M_{si,t} \tag{B1.8}$$

where m is the fraction of individuals dispersing from the population sub-unit i. The summation term indicates how many immigrants the ith units have received from neighbouring units. Often the dispersal success is taken as $\exp(-cd_{si})$, d_{si} being the distance between the units s and i.

the semi-independent local units connected to the regional dynamics. This will be demonstrated here with two examples based on simulations. The argument is that dispersal linkage and spatial structure together may transform the local population dynamics such that they are qualitatively entirely different when compared to similar renewal of populations in total isolation.

First Example

Assume two habitable patches with matching carrying capacity are proximate enough to enable dispersing individuals freely to reach each of them. In every generation, a fraction m of the resident individuals leaves the natal population to disperse into the other population. Let the Ricker dynamics (Box 1; $K = 1$ for both populations) suffice as the renewal process (both populations have been initiated with independent random numbers from a uniform distribution from 0–1). The renewal dispersal process is allowed to run for 500 generations. Its effect on the dynamics of the two populations is evaluated by scoring the temporal match in the fluctuations between the two populations for the final 50 time steps. For this purpose, the cross correlation coefficient is used with zero time lag (Chatfield, 1984). Large positive values of the coefficient indicate a high level of temporal match, large negative values mean that the population pair fluctuates in the opposite phase, and values around zero indicate that the pair does

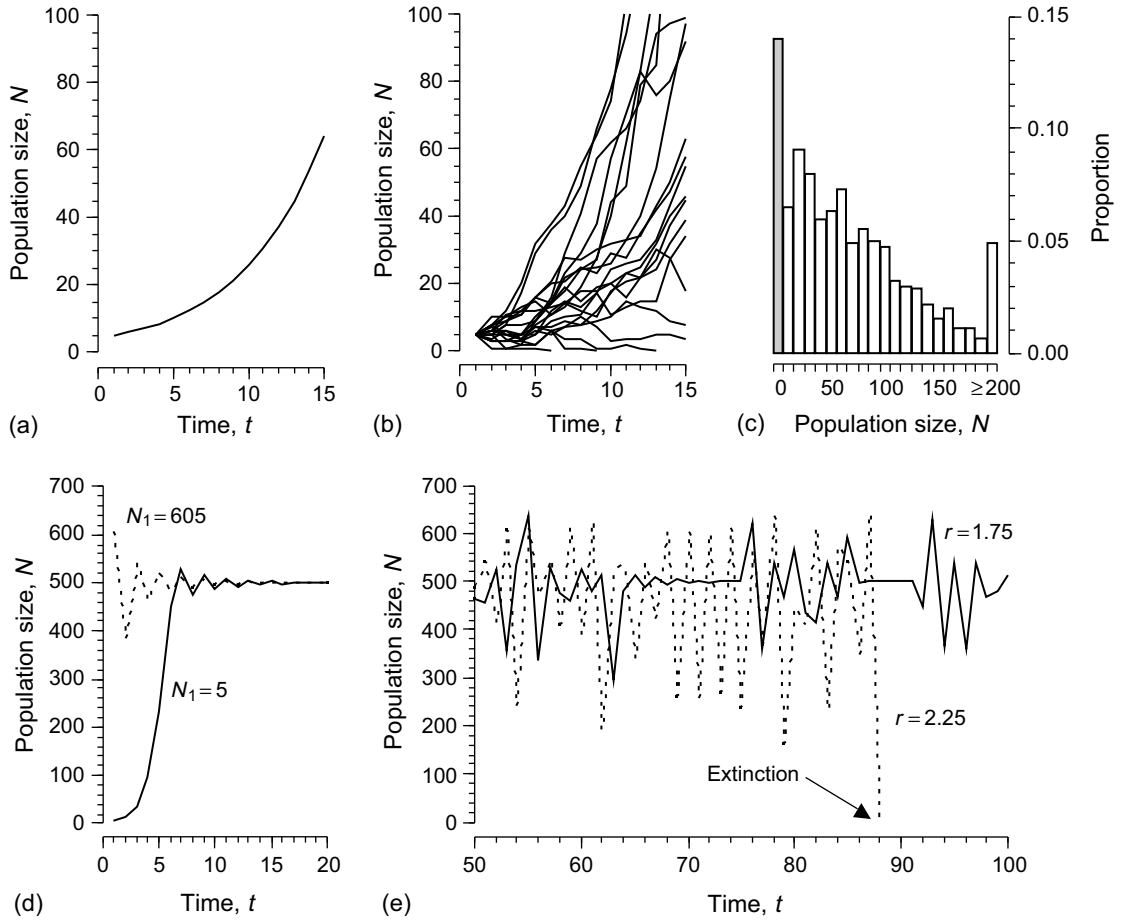


Figure 2 Behavior of selected models on population dynamics. (a) Deterministic exponential growth, Equation (B1.3), $r = 0.2$, (b) as in (a) but now with demographic stochasticity, (c) as (b) but the frequency distribution is the final population size at $t = 15$. Note that a modest fraction (13.8% hatched bar) of the iterations of exponential growth with demographic stochasticity leads to extinction. (d) Logistic growth (parameter values inserted) with two initial population sizes: when $N_1 = 5$ the model yields an exponential growth, which levels off at the carrying capacity (K) with some bouncing back and forth (due to the intrinsic density dependence with the selected r value). When $N_1 = 605$, density-dependent feedback in the logistic equation also brings the population size back to K. (e) Two (r values inserted) population size trajectories of the logistic growth with environmental stochasticity (Moran effect) affecting K with probability of $P = 0.33$. If an external disturbance takes place, K will be multiplied by uniform random numbers drawn from 0.8- 1.2. Note that this high perturbation rate quite often leads to population extinction when r is high enough

not have much in common in their phase of temporal fluctuations.

We examined the effect of dispersal as the synchronizing agent for the dynamics of the two populations over the periodic and chaotic range of the Ricker (1954) dynamics ($2 \leq r \leq 4$, Box 1). If the proportion m of individuals moving per generation is low, then the two populations do not share many individuals in common in affecting their dynamics via the density-dependent feedback (Box 1). However, even a small dispersal rate is capable of synchronizing dynamics of the two populations in the periodic range of r (Figure 3a). Further, even a small amount of dispersal will modulate the chaotic Ricker dynamics into a two-point cycle (Figure 3b). When the value of m is increased sufficiently, dispersal can perfectly synchronize the dynamics of the two populations (Figure 3a).

Second Example

To demonstrate further how space and dispersal may modulate local dynamics we shall replace the Ricker dynamics with the Moran–Ricker dynamics (Box 1), and will have 25 populations in randomized locations of a 25×25 grid (Ranta *et al.*, 1997a). In each generation 10% of the current population will take leave from the natal patch and re-distribute themselves among the remaining patches. Here the populations are in an explicit space, and distance will matter in the dispersal process. The dispersal is taken to be negatively distance dependent (Box 1): reaching

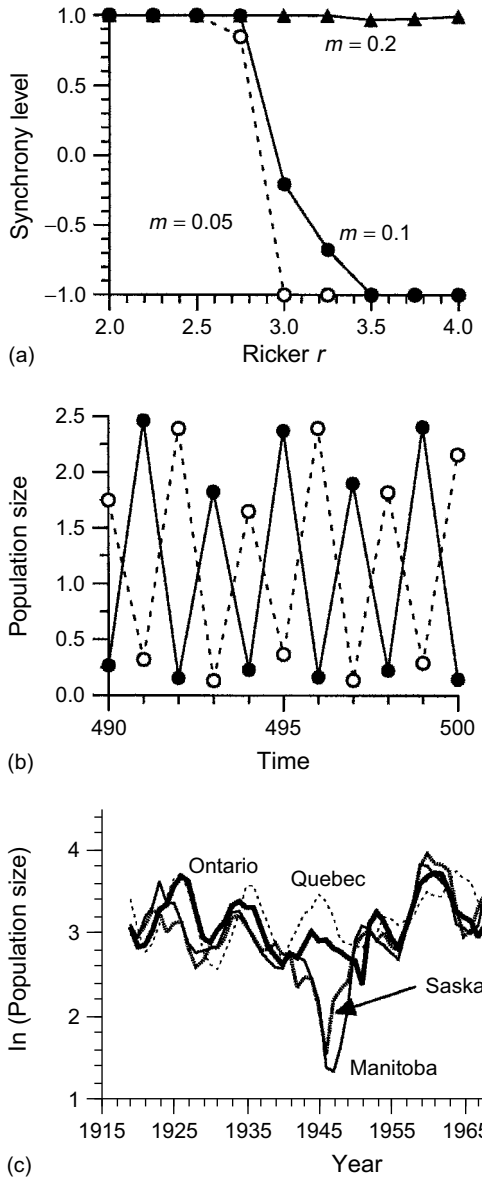


Figure 3 (a- b) Synchronicity of dynamics of two populations (a) obeying Ricker dynamics (only the range of r from periodic to chaotic dynamics is shown here) and sharing dispersing individuals (m gives the fraction of individuals dispersing in each generation), (b) with $m = 0.05$ and $r = 3$, the two populations do not become synchronized but uctuate out of phase. (c- e) Dynamics of the Canadian lynx in six Canadian provinces and two Territories (c,d). (e) Synchrony of the Canadian lynx uctuations are calculated in pairs between all provinces and Territories (28 comparisons) for the data in (c) and (d) using a 15-year sliding time window which is put through the 67 year lynx series with a time step of one year. Though the lynx population highs and lows roughly match, there are violent temporal uctuations in the level of synchronicity, e.g., between Yukon versus Ontario and Northwestern Territories (NWT) and Quebec. (Graph modi ed from Ranta *et al.*, 1997a.) See next page for Figures 3d and 3e

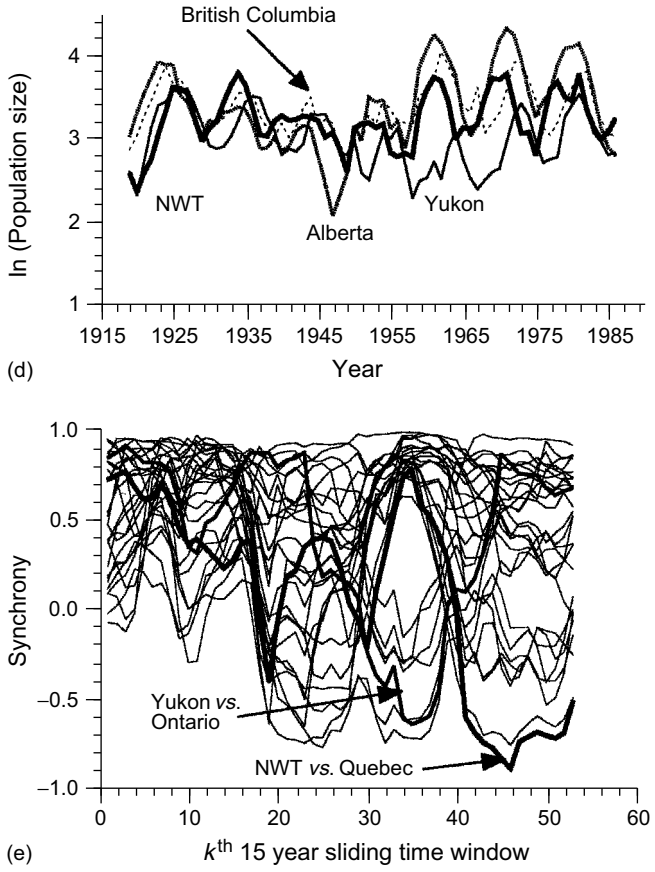


Figure 3 (*Continued*)

a nearby population is much easier than travelling over long distances. The populations were initiated in phase and left to renew after the Moran–Ricker dynamics with 9–10 year periodicity (Ranta *et al.*, 1997a) and the system was allowed to run for 1000 generations.

The outcome, exemplified by temporary fluctuations of a typical pair of populations, is illustrative. In such a system, there can be periods of high temporal match in dynamics for a substantially long time. This is not all, however. There can be times when synchronously fluctuating populations drift out of phase, drastic changes in amplitude may emerge, and even disappearance and re-appearance of the local cyclic dynamics can be witnessed (Ranta *et al.*, 1997a).

Canadian Lynx and Finnish Voles in Time and Space

An interesting question rises: do we find in nature such population dynamics as described above? The answer is yes! A particularly good and thoroughly studied example is the countrywide dynamics of the Canadian lynx (Elton and Nicholson, 1942; Royama, 1992; Ranta *et al.*, 1997a; Stenseth *et al.*, 1999). First, it is known that the dynamics

of the Canadian lynx obeys cyclic fluctuations with a 9–11 year period, as indicated with Canadian provincial and territorial data (Figure 3c,d). At times, two provinces may display high level of synchrony in fluctuations of lynx numbers. At other times, lynx records from the very same pair of provinces may be in opposite phase, to fluctuate later in step.

This feature of the Canadian lynx dynamics becomes more apparent when a sliding time window allowed to run through the entire data, and for each step, the level of synchrony is assessed for all pairs of time series. One finds that there are violent temporal fluctuations in the level of synchrony (Figure 3e). This has been interpreted in terms of travelling waves in population dynamics (Ranta *et al.*, 1997a).

Other, and perhaps more convincing examples come from voles in Finland (Ranta and Kaitala, 1997) and in Scotland (Lambin *et al.*, 1998). The travelling wave is actually not a flush of individuals passing by. Rather, one should understand travelling waves as spatially correlated fluctuations in population numbers. Thus, at any time one finds large regions with high population densities but with increasing distance these densities fade away to give large areas of low population densities. In particular, if the dynamics of the target species are periodic, like in the Fennoscandian voles (Hansson and Henttonen, 1985), the pattern in space repeats itself approximately with the cycle period length. Such a repetition of the spatial pattern gives an impression of travelling waves (Ranta *et al.*, 1997a; Ranta and Kaitala, 1997, 1998).

Habitat Loss and Dynamics of Populations

A particularly interesting question to be dealt with is how the loss of habitable area may affect the dynamics of dispersal-coupled local populations. For this purpose, we shall make use of cyclic dynamics. We choose to use Finnish grouse as an example (Lindén, 1981; Lindström *et al.*, 1995). Periodically oscillating populations have been one of the favorite subjects of research in population ecology since their first appearance in the literature three quarters of a century ago (Elton, 1924; Lindström *et al.*, 2001). Several hypotheses have been put forward to explain the remarkable pattern of self repeatability of cyclic dynamics as so well documented with the Canadian lynx (Elton and Nicholson, 1942; Royama, 1992) and especially with Fennoscandian voles (Stenseth, 1999). Despite the efforts by numerous research groups there is no consensus, not even with the Canadian lynx, as to what maintains the cycles (Stenseth *et al.*, 1999). A recent intriguing feature with the cyclic dynamics of the Fennoscandian voles is that in the 1980s the regular four year cycle waned to mere irregular variation (Hansen *et al.*, 1999). A similar waning has also been observed with the Finnish grouse, which have been showing a six-year periodicity in their cyclic dynamics at least from the beginning of this century (Lindström *et al.*, 1995; Figure 4a).

One starts to wonder: Where have all the cycles gone? An obvious explanation would be that something has changed, either these populations or in their environment. A harder answer is to find a candidate for the changed something. One option we can identify is landscape management, which has changed in Finland due to recent forestry practice. Forested landscape has become more fragmented due to cuttings and the fraction of old forest has declined both in Sweden and in Finland during the past 20 years.

To address the effect of habitat loss on population changes, we shall assume that our focal species obeys Moran–Ricker dynamics (Box 1) in its renewal process. By selecting the parameters properly, one can achieve cyclic dynamics with the period length of six years (Kaitala *et al.*, 1996). The landscape will consist of habitable areas coupled with dispersing individuals. We shall assume that migration among habitat patches close to each other is far easier than dispersing long distances. A further refinement is that, in the increasing phase of a cycle, a substantial proportion of the population consists of young individuals prone to disperse (Lindén, 1981). As the population declines the proportion of young individuals is lower, cutting down the dispersal rate.

For simplicity, we shall assume that the landscape is comprised of n habitable units ($n = 10$ in this example) arranged in one-dimensional space. In the homogeneous landscape the distances between the adjacent habitat subunits are equal (1 unit), and the distance dependency of dispersal is selected so that migrating individuals do not have any problems successfully crossing the short inter-patch distances. The landscape with habitat loss is composed of a gradient of habitable patches with increasing inter-patch distances from the dense end to the sparse end of the fragmentation gradient. In our example, the consequent inter-patch distances are 1, 1, 1, 1, 2, 4, 6, 8, 10, and 12 units from the dense end to the sparse end.

In the simulations, the populations in the local habitat units were initiated with independent random numbers drawn from a uniform distribution (between 1 and 20). The system was allowed to run for 2000 generations. We shall sample only the final 100 generations for calculating our descriptive statistic: the coefficient of variation in population size. Large values of this coefficient indicate that the periodicity of the cyclic dynamics is clearly pronounced, while small values suggest the dying out of the cyclic dynamics.

Our principal finding with the data on black grouse is that the cyclic dynamics has transformed into weak annual fluctuations around the long-term mean (Figure 4a). Our principal finding, with modeling of habitat loss and differential

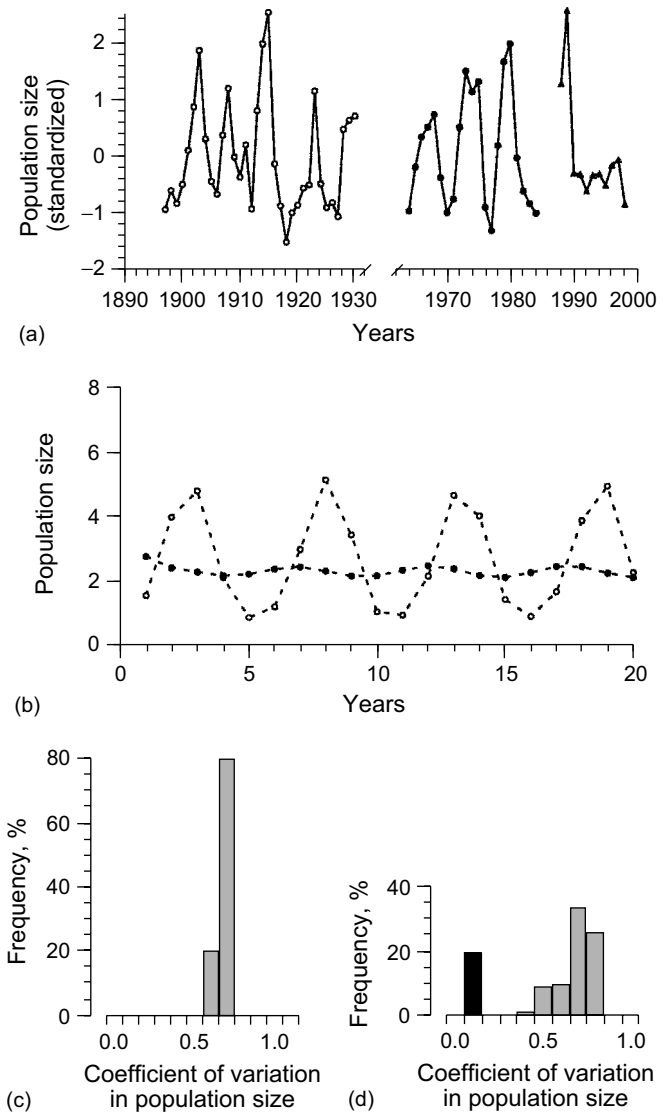


Figure 4 (a) Black Grouse dynamics (population size standardized to zero mean and unit variance for the three periods) in southwest Finland. Note the pronounced periodicity in the two older time series and the disappearance of the cyclic dynamics in the most recent data set. (b) Population trajectories of a Moran- Ricker model yielding 6- 7 year periodic dynamics ($r = 1.45$, $a_1 = -0.1$, $a_2 = -0.45$) in a fragmented landscape. The open symbol refers to populations in the dense end of the fragmentation gradient while the closed symbol is for the population in the sparsest end of the fragmentation gradient. (c) Coef cient of variation of populations in a system with no habitat loss (all values are relatively high, thus all populations uctuate from low to high values), and (d) with habitat loss. Some values are small indicating that there are populations with small temporal variability, some values are high, indicating high temporal uctuations in population size. (Modi ed from Ranta *et al.* in prep.)

dispersal in the increasing and decreasing phases of the cycle, is that the cycle disappears showing only minor variations around the mean population size (Figure 4b,d). It is worth noting that with uniformly structured habitat there are no signs that the cycle in the population dynamics will vanish (Figure 4c).

There are, of course, many other ways to demonstrate how the spatial linkage among habitable units affects population dynamics both at local and regional levels. However, the few examples mentioned here underline the fact that without acknowledging the significance of immigration and emigration (the links between local populations and regional dynamics) one can easily be misled when analyzing local population dynamics.

MORAN EFFECT: EXTERNAL PERTURBATION AND DYNAMICS OF POPULATIONS

Elton (1924) was arguably the first to draw attention to the large-scale synchrony in the dynamics of the Norwegian lemming and the snowshoe hare. He (Elton and Nicholson, 1942, 239–240) was very explicit about the synchronicity in the case of the Canadian lynx (Figure 3):

> The [lynx] cycle covers the whole northern forest zone of Canada, from Labrador to British Columbia to Yukon .... The most extraordinary feature of this cycle is that it operates sufficiently in line over several million square miles of country not to get seriously out of phase in any part of it .... (This evidence) makes it certain that some overriding process maintains the cycle in line over the whole extent of Canada.

Moran (1953a,b) proposed the idea of how two (or more) populations could become synchronized due to stochastic density-independent but correlated processes. His assumption was that population regulation may be density dependent, and that populations may be divided into smaller independent sub-populations.

Suppose (Moran 1953a,b) that the dynamics of two populations $N_{1,t}$ and $N_{2,t}$ in time t can be described as:

$$N_{1,t+1} = aN_{1,t} + bN_{1,t-1} + \varepsilon_t \quad (1)$$

$$N_{2,t+1} = aN_{2,t} + bN_{2,t-1} + \eta_t \quad (2)$$

so that the a and b are identical for N_1 and N_2. The random elements ε_t and η_t are different but correlated. The correlation arising between N_1 and N_2 will then equal the correlation between ε_t and η_t. If the density-independent series ε_t and η_t are caused by, or just correlated with regional disturbances, this provides an explanation for the synchronous dynamics of the two populations N_1 and N_2. The density-dependent coefficients a and b do not necessarily need to be perfectly matched between the two populations. Synchronous dynamics is still achievable via external disturbances but the correlation between N_1 and N_2 will not equal that between ε_t and η_t. This reasoning later became known as the *Moran theorem* (Royama, 1992)

or the *Moran effect* (Ranta *et al.*, 1997a). Contemporary population ecologists emphasize the importance of the interplay of density-dependent and density-independent factors (Kaitala *et al.*, 1996; Grenfell *et al.*, 1998; Hieno *et al.*, 1997), thus echoing Moran's (1953a,b) seminal ideas. A central finding of current research suggests that populations tend to fluctuate in step over large areas and that the degree of synchrony levels off with increasing distance among the population units compared (Ranta *et al.*, 1997a; Lambin *et al.*, 1998; Bjørnstad *et al.*, 1999). This finding is substantiated also by increasing evidence obtained from long-term spatial time series data sets from different animal taxa (as in Figure 5).

As an example, we take the data on 1964–1984 dynamics of Finnish Black Grouse in 11 provinces (Figure 6). Besides cyclic behavior (see also Figure 4a) there is also a high level of synchrony over large areas. Note, however, that the nation-wide synchronicity is far from perfect, and also that the level of synchrony tends to go down with increasing distance between the time series compared (Figure 5a). To propose possible causes for the observed patterns, we have made use of a variety of theoretical models (Ranta *et al.*, 1997a). The populations are assumed to consist of a number of habitable units, which are subjected to local environmental noise. To explore the effect of dispersal and the Moran effect, the units are either independent or linked by dispersal. Further, they are or are not subjected to global environmental disturbance.

The results show that the Moran effect is capable of synchronizing population dynamics, as indicated by the mean of the cross-correlation coefficients deviating significantly from zero (Figure 5). The Moran effect alone does not show (as expected) any trend with increasing

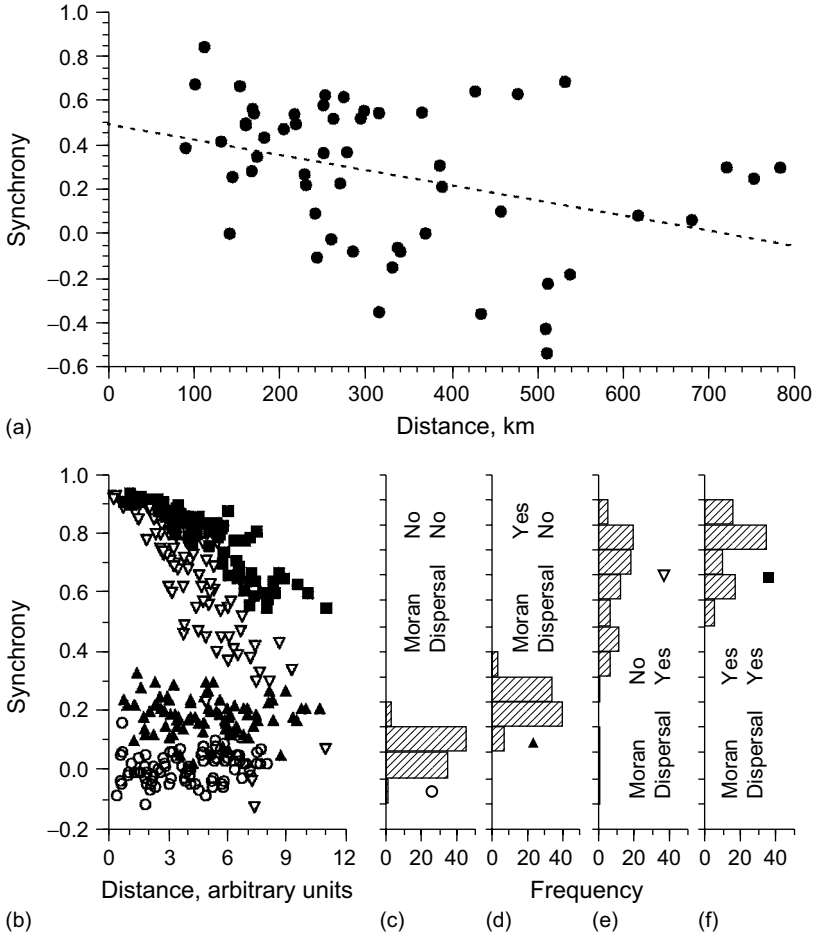


Figure 5 (a) The level of synchronous uctuations in Finnish Black Grouse dynamics (the 11 provinces in Figure 6 compared in pairs) is reasonable (mean of the cross-correlation = 0.265) but it levels off with increasing distance among the provinces compared. (b) Synchrony levels against distance in population dynamics of 10 populations in explicit 10×10 space obeying Black Grouse-like dynamics. The different symbols are for different combinations of dispersal and the Moran effect, indicated in panels (c- f) giving the corresponding marginal distributions of the synchrony measure. (Modi ed after Ranta *et al.*, 1997b)

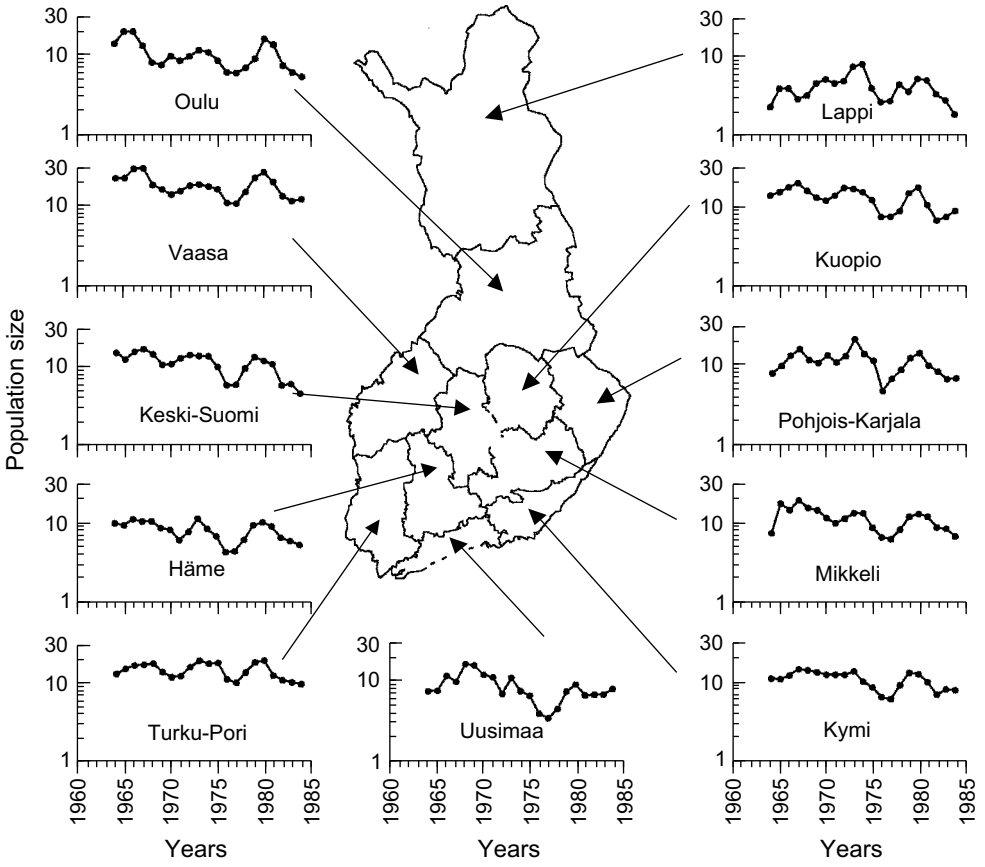


Figure 6 Black Grouse population (individuals km^{-2}) uctuations in different parts of Finland. (Data courtesy of Game Division of the Finnish Game and Fisheries Research Institute Lindén, 1981)

distance among the compared populations. On the other hand, dispersal among the populations alone is powerful enough to cause the populations to fluctuate in synchrony (see also Figure 3d), and the level of synchrony levels off with distance among the populations (Figure 5e). Superimposing Moran's effect on the dispersal does not improve much the level of synchrony, nor does it affect dramatically the negative correlation between synchrony and distance (Figure 5f). As a whole, we are facing the difficulty of deciding whether the negative correlation between the level of synchrony and distance among the compared natural populations is due to the effect of dispersal alone, or both dispersal and the global disturbance acting together.

According to one interpretation, the Moran effect (Moran, 1953a,b; Royama, 1992) can be understood as a global disturbance influencing the renewal processes similarly and simultaneously in different sub-populations or localities. From the above, we know that this effect is capable of synchronizing the dynamics of otherwise independent populations. However, in a process as described exactly by the Moran's theorem, the level of synchrony does not decrease with increasing distance. This prediction does not correspond to empirical observations. Consequently, two different alternatives have been proposed to develop Moran's idea further. First, the dispersal of individuals linking population sub-units has been put forward to explain this feature (Ranta *et al.*, 1997a,b; Kaitala and Ranta, 1998). Second, Ranta *et al.* (1999) applied an approach incorporating both spatially autocorrelated Moran effect and dispersal. With several different model classes of population renewal, they show that a spatially autocorrelated disturbance alone, especially in the presence of a population linkage by dispersal, is capable of producing synchronous dynamics with the synchrony leveling off with distance. There is one more complication, as emphasized by Kaitala *et al.* (1996), the actual agent causing the stochastic perturbation in nature is complex. It does not have to be the same every time, nor does it have to be the same for all species in a given year, supposing they are simultaneous enough.

CONCLUDING WORDS

Long-term population data collected simultaneously in more than two localities provide not only the means

to elaborate the relative significance of local and global processes but also permit evaluation of the significance of external perturbations, the Moran effect, on the dynamics of populations. There are few (if any) doubts that external disturbances affect populations on various geographical scales. Research on the dynamics of various animal species suggests that population fluctuations can be explained (at least partially) by weather and climate factors (e.g., Arditi, 1979; Slagswold and Grasaas, 1979; Steen *et al.*, 1988). This, in fact, dates back to the speculated causes of the regular dynamics of the snowshoe hare (Elton, 1924).

No population stays constant over many years, population sizes are prone to changes for several reasons. We have treated some of these cases above. In particular, our emphasis has been in the intrinsic population processes, that is, births and deaths. One has to acknowledge the fact that populations seldom live in perfect isolation. Emigrating individuals disperse elsewhere and immigrating individuals arrive from other localities. This process ties regional and local processes together. There is good theoretical and empirical evidence to indicate that proper understanding calls for research on both local and regional levels. A particularly good example is the recent study on the fluctuations of the Canadian lynx (Stenseth *et al.*, 1999) which shows that there are clear regional differences in the kind of dynamics the lynx displays in areas exposed to differing ecological conditions.

In this approach, the rationale is to hunt for weather and climate variables that best correlate with long-term data on population changes. This is because of a temptation to find biologically rational explanations for the effect of almost any weather-derived variable on population fluctuations via survival and reproduction, the key elements of population persistence. The origin of the idea is in the climatic control theory (Andrewartha and Birch, 1954).

Despite well articulated caveats (Royama, 1992) the climatic control paradigm continues to arise now and then. Recent concern about global warming has revived the search for correlations between observed ecological time series with climate indices such as the North Atlantic Oscillation (NAO) (*see* **North Atlantic Oscillation**, Volume 1), North Pacific Index, El Niño Southern Oscillation, or Palmer Drought Severity Index. These mostly deal with the NAO and aquatic taxa, e.g., North Atlantic phytoplankton density (Reid *et al.*, 1999), toxic plankton algae along the Swedish west coast (Belgrano *et al.*, 1999), dynamics of two calanoid copepod species in the Northeastern Atlantic (Fromentin and Planque, 1996), and annual landing statistics of European herring and sardines (Alheit and Hagen, 1997). There are also some examples from terrestrial ecosystems: flowering phenology and reproductive traits in some plant species (Galen and Stanton, 1991), population dynamics of moose, white-tailed deer (Post and Stenseth, 1998) and red deer (Forchhammer *et al.*, 1998a) and the breeding biology of migratory birds in the UK (Forchhammer *et al.*, 1998b).

When attributing long-term population changes to external factors, there is a caveat which is often forgotten. We shall illustrate the problem here with the annual salmon catch in one of the Norwegian rivers, the Namsen (Figure 7a). The data cover an exceptionally long period, 117 years (1876–1992) to serve as an excellent example of the central problem when trying to find an explanation for the long-term fluctuations in a population. For the same period of time we also have the standardized NAO index values (Figure 7b). Our rationale is to search for any covariance between these two time series. The cross-correlation coefficient with various time lags is a perfect tool for this purpose. However, prior to the analysis one has to remove any long-term trends. One now finds that with a lag of 10 years there is a positive correlation ($r_{10} = 0.385$, $P < 0.01$). An explanation for this long time lag is that the maturation of Atlantic salmon takes 4–6 years, and it may take two generations for the effect to cascade into the catchable adult biomass. *Ad hoc* explanations like this abound in the literature on climatic control theory.

With 117 years of data one can examine whether the aggregated correlation between the lagged time series remains if the covariance structure between the salmon catch data and the NAO index is studied in finer detail. A 20 year sliding time window was allowed to pass through the time series and we calculated the cross-correlation coefficient r_{10} for all subsequent windows. The results are not very encouraging (Figure 7c); cross-correlations obtained in this way fluctuate from statistically significant negative ones to statistically significant positive ones. Most of the time, however, there is no significant covariance. Based on this analysis, one must use caution in drawing conclusions about the covariance between population data and indices like the NAO based on short time series.

Without exception, populations change in size and spatial patterns through time. Both temporal and spatial patterns may be extremely rich and complicated. Two basic approaches are studied in population ecology to unravel the order behind population fluctuations. The first approach attempts to resolve the puzzle by postulating that the changes are due to an intrinsic factor in the populations, that is, density dependence. The second approach assumes that environmental fluctuations are diffused into the population dynamics. The current, commonly shared view emphasizes the synthesis of these two approaches: population fluctuations are a result of both factors. We should also add the spatial dimension of the population dynamics, which alone can act as a strong modifying factor in population dynamics. Unraveling such joint effects appears to

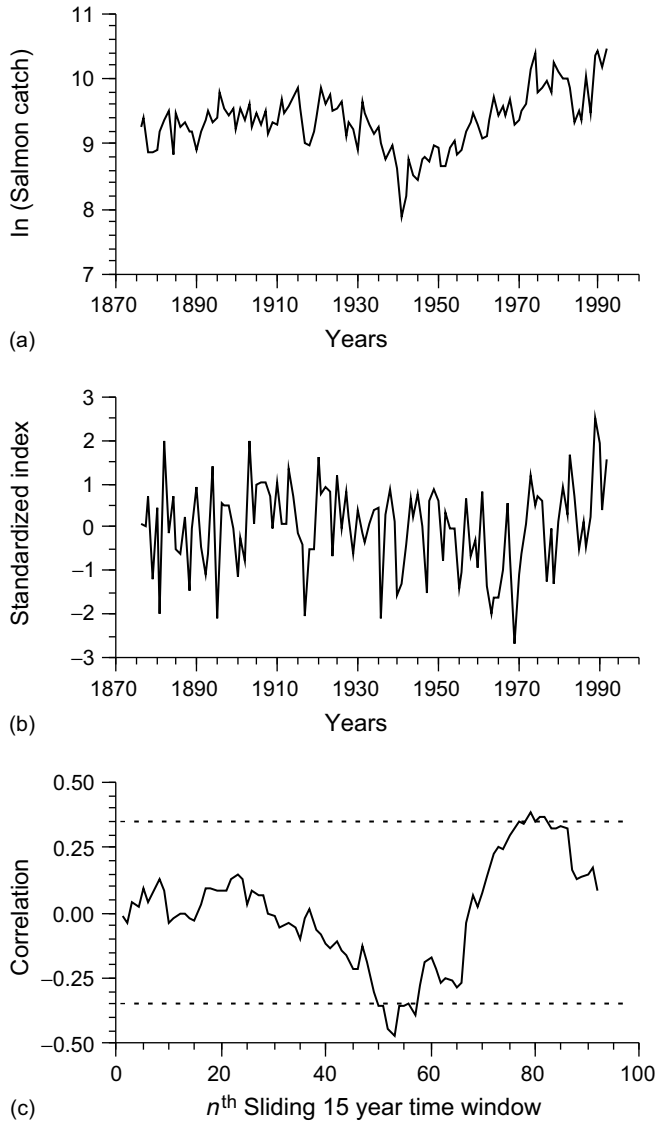


Figure 7 (a) Annual salmon catch in the Namsen river, Norway from 1876- 1992. (b) Standardized (zero mean, unit variance) NAO index for the corresponding period. (c) Temporal covariance structure between the salmon data and the NAO series (15-year sliding window is passed though the data with a time step of one year and the level of synchrony between the resulting time series is assessed). The broken lines indicate the 95% con dence limits of the correlation coef cient

be extremely demanding. The effects of combining all these factors, in population dynamics results commonly in complex and counterintuitive consequences where local consequences may differ from global ones. Thus, predicting the effects of a changing environment on the qualitative properties of population dynamics may remain a challenge for population ecologists for a long time to come.

REFERENCES

Alheit, J and Hagen, E (1997) Long-term Climate Forcing of European Herring and Sardine Populations, *Fish. Oceanogr.*, **6**, 130–139.

Andrewartha, H G and Birch, L C (1954) *The Distribution and Abundance of Animals*, The University of Chicago Press, Chicago, IL.

Arditi, R (1979) Relation of the Canadian Lynx Cycle to a Combination of Weather Variables: a Stepwise Multiple Regression Analysis, *Oecologia*, **41**, 219–233.

Bascompte, J and Solé, R V (1997) *Modeling Spatiotemporal Dynamics in Ecology*, Springer-Verlag, Berlin.

Belgrano, A, Lindahl, O, and Hernroth, B (1999) North Atlantic Oscillation Primary Productivity and Toxic Phytoplankton in the Gullmar Fjord, Sweden (1985–1996), *Proc. R. Soc. Lond., Ser. B: Biol. Sci.*, **266**, 425–430.

Bjørnstad, O N, Ims, R A, and Lambin, X (1999a) Spatial Population Dynamics: Analyzing Patterns and Processes of Population Synchrony, *Trends Ecol. Evol.*, **14**, 427–432.

Chatfield, C (1984) *The Analysis of Time Series: an Introduction*, Chapman and Hall, NY.

Elton, C S (1924) Periodic Fluctuations in the Numbers of Animals: their Causes and Effects, *Br. J. Exp. Biol.*, **2**, 119–163.

Elton, C S and Nicholson, M (1942) The Ten-year Cycle in Numbers of the Lynx in Canada, *J. Anim. Ecol.*, **11**, 215–244.

Emlen, J M (1984) *Population Biology: the Coevolution of Population Dynamics and Behavior*, Macmillan, NY.

Forchhammer, M C, Post, E, and Stenseth, N C (1998b) Breeding Phenology and Climate, *Nature*, **391**, 29–30.

Forchhammer, M C, Stenseth, N C, Post, E, and Langvatn, R (1998) Population Dynamics of Norwegian Red Deer, Density dependence and Climatic Variation, *Proc. R. Soc. Lond., Ser. B: Biol. Sci.*, **265**, 341–350.

Fromentin, J-M and Planque, B (1996) Calanus and Environment in the Eastern North Atlantic, 2, Influence of the North Atlantic Oscillation on C. finmarchicus and C. helgolandicus, *Mar. Ecol. Prog. Ser.*, **134**, 111–118.

Galen, C and Stanton, M L (1991) Consequences of Emergence Phenology for Reproductive Success in Ranunculus Adoneus (Ranunculaceae), *Am. J. Bot.*, **78**, 978–988.

Ginzburg, L R and Inchausti, P (1997) Asymmetry of Population Cycles: Abundance-growth Representation of Hidden Causes of Ecological Dynamics, *Oikos*, **80**, 435–447.

Grenfell, B T, Wilson, K, Finkenstädt, B F, Coulson, T N, Murray, S, Albon, S D, Pemberton, J M, Clutton-Brock, T H, and Crawley, M J (1998) Noise and Determinism in Synchronized Sheep Dynamics, *Nature*, **394**, 674–677.

Hansen, T F, Stenseth, N C, and Henttonen, H (1999) Multiannual Vole Cycles and Population Regulation During Long Winters: An Analysis of Seasonal Density Dependence, *Am. Nat.*, **154**, 129–139.

Hansson, L and Henttonen, H (1985) Gradients in Density Variations of Small Rodents: the Importance of Latitude and Snow Cover, *Oecologia*, **67**, 394–402.

Heino, M, Kaitala, V, Ranta, E, and Lindström, J (1997) Synchrony and Extinction Rates in Spatially Structured Populations, *Proc. R. Soc. Lond., Ser. B: Biol. Sci.*, **264**, 481–486.

Hilborn, R and Mangel, M (1997) *Ecological Detective*, Princeton University Press, Princeton, NJ.

Kaitala, V and Ranta, E (1998) Travelling Wave Dynamics and Self-Organization in a Spatio-temporally Structured Population, *Ecol. Lett.*, **1**, 186–192.

Kaitala, V, Ranta, E, and Lindström, J (1996) Cyclic Population Dynamics and Random Perturbations, *J. Anim. Ecol.*, **65**, 249–251.

Lambin, X, Elston, D A, Petty, S J, and MacKinnon, J L (1998) Spatial Asynchrony and Periodic Travelling Waves in Cyclic Populations of Field Voles, *Proc. R. Soc. Lond., Ser. B: Biol. Sci.*, **265**, 1491–1496.

Lindén, H (1981) Changes in Finnish Tetraonid Populations and Some Factors Influencing Mortality, *Finnish Game Res.*, **39**, 3–11.

Lindström, J, Ranta, E, Kaitala, V, and Lindén, H (1995) The Clockwork of Finnish Tetraonid Population Dynamics, *Oikos*, **74**, 185–194.

Lindström, J, Ranta, E, Kokko, H, Lundberg, P, and Kaitala, V (2001) From Arctic Lemmings to Adaptive Dynamics: Charles Elton's Legacy in Population Ecology, *Biol. Rev.*, **76**, 129–158.

Lundberg, P, Ranta, E, Ripa, J, and Kaitala, V (2000) Population Variability in Space and Time, *TREE*, **15**, 460–464.

May, R M (1974) Biological Populations with Nonoverlapping Generations: Stable Points, Stable Cycles and Chaos, *Science*, **186**, 645–647.

Moran, P A P (1953a) The Statistical Analysis of the Canadian Lynx Cycle, I, Structure and Prediction, *Aust. J. Zool.*, **1**, 163–173.

Moran, P A P (1953b) The Statistical Analysis of the Canadian Lynx Cycle, II, Synchronization and Meteorology, *Aust. J. Zool.*, **1**, 291–298.

Post, E and Stenseth, N C (1998) Large-scale Climatic Fluctuation and Population Dynamics of Moose and White-tailed Deer, *J. Anim. Ecol.*, **67**, 537–543.

Ranta, E and Kaitala, V (1997) Travelling Waves in Vole Population Dynamics, *Nature*, **390**, 456.

Ranta, E, Kaitala, V, and Lundberg, P (1997a) Population Variability in Space and Time: the Dynamics of Synchronous Population Fluctuations, *Science*, **278**, 1621–1623.

Ranta, E, Kaitala, V, and Lindström, J (1997b) Spatial Dynamics of Populations, in *Modeling Spatiotemporal Dynamics in Ecology*, eds J Bascompte and R V Solé, Springer-Verlag, Berlin, 45–60.

Ranta, E, Kaitala, V, and Lindström, J (1999) Spatially Autocorrelated Disturbances and Patterns in Population Synchrony, *Proc. R. Soc. Lond. Ser. B: Biol. Sci.*, **266**, 1851–1856.

Reid, P C, Edwards, M E, Hunt, H, and Warner, A E (1999) Phytoplankton Change in the North Atlantic, *Nature*, **391**, 546.

Ricker, W E (1954) Stock and Recruitment, *J. Fish. Res. Bd. Can.*, **11**, 559–623.

Royama, T (1992) *Analytical Population Dynamics*, Chapman and Hall, London.

Slagswold, T and Grasaas, T (1979) Autumn Population Size of the Capercaillie Tetrao Urogallus in Relation to Weather, *Ornis Scandinavica*, **10**, 37–41.

Steen, J B, Steen, H, Stensethm, N C, Myrberget, S, and Marcström, V (1988) Microtine Density and Weather as Predictors of Chick Production in Willow Ptarmigan Lagopus l. lagopus, *Oikos*, **51**, 367–373.

Stenseth, N C, Chan, K-S, Tong, H, Boonstra, R, Boutin, S, Krebs, C J, Post, E, O'Donoghue, M, Yoccoz, N G, Forchhammer, M C, and Hurrell, J W (1999) Common Dynamic Structure of Canada Lynx Populations Within Three Climatic Regions, *Science*, **285**, 1071–1073.

Stenseth, N C (1999) Population Cycles in Voles and Lemming: Density Dependence and Phase Dependency in a Stochastic World, *Oikos*, **87**, 427–461.

Tilman, D and Kareiva, P (1997) *Spatial Ecology: the Role of Space in Population Dynamics and Interspecific Interactions*, Princeton University Press, Princeton, NJ.

Turchin, P (1990) Rarity of Density Dependence or Population Regulation with Lags? *Nature*, **344**, 660–663.

Potential Evapotranspiration (PET)

see PET (Potential Evapotranspiration) (Volume 2)

Productivity

Most energy flow through an ecosystem begins with the fixation of sunlight and assimilation of carbon by green plants. Primary productivity, the rate of conversion of radiant energy and carbon by plants into organic substances which serve as food materials, refers to the first and basic form of organic matter production in an ecosystem. *Gross primary productivity* (GPP) is the total rate of this production including the organic matter used up in respiration. Like other organisms, plants consume stored organic matter through the process of respiration to fuel their reproduction and maintenance. *Net primary productivity* (NPP), the rate of total plant growth, is the total rate of storage of organic matter after subtracting the rate at which this stored organic matter is consumed through respiration (i.e., NPP = GPP − plant respiration). Net primary production can build up over time as plant biomass: while some is lost through decomposition, much of the rest may be retained as living material.

Plant productivity plays a crucial role in the carbon cycle, and hence in considerations of the greenhouse effect and global climatic change. For instance, in North American boreal forests, almost 50% of the carbon assimilated through photosynthesis (i.e., GPP) is consumed in plant respiration, leaving only about half for plant growth (i.e., NPP). About 90% of this plant growth is eventually returned to the soil as litter and is subsequently mineralized by soil organisms during their respiration. (The

resulting mineral nutrients can subsequently be taken up during future plant growth.) The remaining 10% of NPP is known as *net ecosystem productivity* (NEP) (i.e., NEP = NPP − decomposition respiration). Almost 90% of NEP is lost through disturbances such as timber harvest, fire, severe weather, or insect outbreaks. Because many of these disturbances have severe effects but occur very infrequently in a given area, realistic estimates of their average impacts must be made from data taken at very large (e.g., biome) scales.

Net biome productivity (NBP) is the 10% of NEP left in the ecosystem after subtracting losses to disturbance. Much of NBP exists as slowly decomposable organic material (stable humus) or long-lived carbon (charcoal). Carbon losses due to major disturbances (e.g., forest fire, insects) at landscape or larger scales, over long time periods (decades to centuries, or even conceivably millennia) should, in principle, balance net ecosystem uptake due to growth minus respiration and decomposition. Hence, in general NBP is much less than NEP which, in turn, is usually much less than NPP, and in at least some ecosystems, NBP should be negligible.

FURTHER READING

Breymeyer, A I, Hall, D O, Melillo, J M, and Agren, G I, eds (1996) *Global Change: Effects on Forests and Grasslands*, John Wiley & Sons, Chichester, 1–459.

Odum, E P (1971) *Fundamentals of Ecology*, W B Saunders, Philadelphia, PA, 1–574.

Waring, R H, Landsberg, J, and Williams, M (1998) Net Primary Production of Forests: A Constant Fraction of Gross Primary Production? *Tree Physiol.*, **18**, 129–134.

Waring, R H and Schlesinger, W H (1985) *Forest Ecosystems: Concepts and Management*, Academic Press, New York, 1–340.

RICHARD A FLEMING *Canada*

Productivity of Terrestrial Ecosystems

Stith T Gower
University of Wisconsin, Madison, WI, USA

Net primary productivity (NPP) of terrestrial ecosystems is a key determinant of the net exchange of carbon dioxide (CO_2) between terrestrial ecosystems and the atmosphere, and hence atmospheric carbon dioxide concentration. NPP varies by 12-fold among terrestrial ecosystems, with deserts having the lowest NPP and temperate broad-leaf evergreen and tropical forests having the highest. Water and nutrient availability strongly influence NPP; the basis for this relationship is that greater resource availability increases leaf area, which in turn increases the amount of photosynthetic active radiation (PAR) absorbed by the vegetation. Humans have altered the productivity of terrestrial ecosystems by changing the climate, atmospheric chemistry, land use, and disturbance regime of terrestrial ecosystems.

INTRODUCTION

Net primary production (NPP) is a useful measure of the metabolism of an ecosystem and the potential of the vegetation to produce food and fiber for human use. Detritus, which derives from the death and shedding of components of NPP, replenishes the soil with organic matter and nutrients that are essential to maintaining long-term soil fertility. Climate, soil fertility, and ecological variables influence the NPP of natural terrestrial ecosystems; thus it is not surprising that NPP varies greatly among major biomes. Few terrestrial ecosystems can be considered natural anymore because they have been directly or indirectly affected by global change. The three components of global change, land use and cover change, changes in atmospheric chemistry, and climate change, all influence NPP. Humans have changed land use and management practices, the chemical composition of the gases comprising the atmosphere, atmospheric nitrogen deposition, and climate. The potential of the biosphere to continue to meet the human demand for products of NPP is influenced by all the components of global change. Quantifying the effects of global change on the global carbon budget is limited by an incomplete understanding of how global change affects NPP. Food and fiber yield is directly proportional to NPP, and humans use almost half of total global NPP for food, fiber, and energy. Humans have noticeably changed or transformed 39–50% of the land surface to meet their needs for food and fiber (Vitousek *et al.*, 1997), and these changes directly and indirectly influence NPP. The rationale for converting natural ecosystems to managed forest plantations and agroecosystems is to increase yield, and economic profit. But shortened fallow periods and removal of a greater fraction of NPP from managed ecosystems may result in a long-term decrease in soil fertility because of the reduced input of detritus to the soil (Landsberg and Gower, 1997).

Humans have increased the atmospheric concentration of carbon dioxide (CO_2) (a greenhouse gas) by 30% since the pre-industrial era, and the increase will continue for the foreseeable future. The elevated atmospheric carbon dioxide concentration contributes to climate warming and may enhance the growth of plants in some regions (IPCC, 1996). Increased atmospheric nitrogen deposition may initially increase the productivity of nitrogen-limited terrestrial

ecosystems, but long-term chronic deposition may lead to permanent decreases in NPP of the terrestrial ecosystems and adjacent aquatic ecosystems (Aber, 1992).

There is general agreement that climate warming and greater interannual variation in climate are occurring, and that the magnitudes of changes differ among terrestrial ecosystems (IPCC, 1996). Predicting how these changes will affect NPP, and the availability of food and fiber, is problematic because of the complex interactive effects of changes in climate and chemistry of the atmosphere (e.g., ozone, nitrogen deposition, and carbon dioxide concentrations).

The objective of this article is to review the production budget, briefly summarize methods used to measure NPP, compare the NPP budgets of the major terrestrial biomes, and examine the major environmental and ecological factors, especially those affected by global change, that influence production budgets. The focus is on NPP because these data are more abundant than the other components of production.

For a summary of the influence of the physical environment on leaf-level carbon dioxide exchange, see **Carbon and Energy: Terrestrial Stores and Fluxes**, Volume 2. The focus here is on stand-level constraints on the productivity of terrestrial ecosystems.

DEFINITIONS AND COMPONENTS OF TERRESTRIAL CARBON BUDGETS (TABLE 1)

Photosynthesis is the assimilation of carbon dioxide by plants. GPP is the total amount of carbon dioxide assimilated by all vegetation life forms (i.e., overstory, shrub, herbs and bryophytes) (Figure 1, arrow 1). A fraction of the carbon dioxide assimilated by vegetation is used to construct new tissue (growth or construction respiration) and repair and maintain existing tissues (maintenance respiration). The sum of maintenance and growth respiration is referred to as *autotrophic respiration* (R_A) and results in a loss of carbon dioxide from the vegetation to the atmosphere (Figure 1, arrow 2).

NPP (Figure 1, arrow 3) is the balance between GPP and R_A (Equation 1)

$$NPP = GPP - R_A \quad (1)$$

NPP is expressed on a dry organic matter or carbon basis, per unit area per year. Values reported on a dry mass can be converted to a carbon basis by assuming carbon:organic matter ratios; the ratio for foliage, herbaceous vegetation and fine roots is approximately 0.45 and the ratio for woody tissue (e.g., stem wood and bark, branches and coarse roots) is 0.50 (Gower *et al.*, 1999). In practice, it is extremely difficult to measure NPP as the difference between GPP and R_A because GPP is difficult to measure; most estimates of GPP are derived from ecosystem process models. Consequently, NPP is calculated as the sum of the annual production (P) of all tissues (i) for all vegetation strata (v) (Equation 2)

$$NPP = (P_{vi}) + H \quad (2)$$

All tissues (e.g., stem, branch, foliage, coarse roots fine roots and mycorrhizae, reproduction) for all plant growth forms (overstory, understory, and ground cover) should be included. The loss of biomass to herbivory (H) should be

Table 1 Carbon budget symbols and de nition

Symbol	De nition
P_{vi}	Biomass production of the i^{th} tissue and v^{th} vegetation strata
C	Carbon
D	Detritus
GPP	Gross primary production
H	Herbivory
NEB	Net biome production
NEP	Net ecosystem production
NPP	Net primary production
NPP_A	Net primary production, aboveground
NPP_B	Net primary production, belowground
R_A	Respiration, autotrophic
R_H	Respiration, heterotrophic

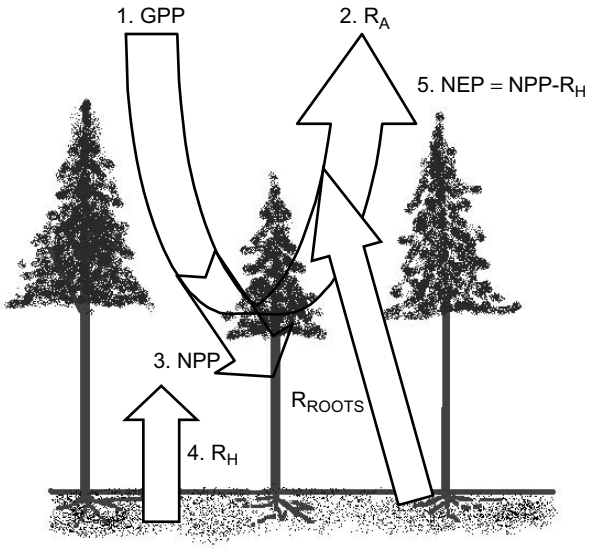


Figure 1 Conceptual diagram of the carbon budget of a forest ecosystem. The de nition of each ux is provided in the text. Processes such as R_A, NPP, soil surface carbon dioxide ux, detritus production, and heterotrophic respiration (R_H) can be measured separately, and the annual budgets are summed to estimate NEP. NEP can also be measured continuously using micrometeorological methods. GPP is dif cult to measure directly, and is usually estimated indirectly or modeled using an ecosystem process model

added to avoid underestimating NPP. During non-outbreak conditions, vertebrates and invertebrates consume <10% of NPP (Landsberg and Gower, 1997), suggesting that ignoring this small loss of tissue will result in only a small underestimation of NPP for forest ecosystems. This assumption is less valid for non-forested ecosystems and for forest ecosystems when herbivore populations are high.

The net exchange rate of carbon, in the form of carbon dioxide between the atmosphere and terrestrial ecosystems, including the vegetation and soil, is referred to as NEP (Figure 1, arrow 4). NEP is of great interest because it defines if the ecosystem is a carbon sink (accumulating carbon) or a source (losing carbon). NEP is the balance between two major processes: NPP and *heterotrophic* respiration, the efflux of carbon dioxide from the soil surface resulting from the oxidation of soil organic matter by heterotrophic organisms (Figure 1, arrow 5).

MEASURING NPP

Equation (2) can be used to calculate NPP for any ecosystem, although the measurement approach varies among ecosystems because of differing vegetation structure, spatial heterogeneity, etc. (Gower *et al.*, 1999). Two common approaches used to estimate NPP aboveground (NPP_A) are *area harvest* and *allometry*. Area harvest entails periodic destructive harvests of all aboveground live biomass in randomly located plots during the growing season. The harvested organic matter is dried (or sub-sampled and dried); then statistically significant increments, or production, of organic matter between each measurement period are summed to estimate NPP_A. This approach is best suited for short stature ecosystems such as grasslands, agricultural crops, tundra, etc. The area-harvest approach is inappropriate for most forests because the annual production of organic matter is small relative to the spatial variability of standing organic matter. The alternative is to establish permanent plots, record the diameter and species of each tree or shrub in the plot, and estimate the annual production of each biomass component from annual radial increment data and allometric equations. Allometric equations of the form $Y = a + \exp^{bX}$ where X is diameter at breast height are used (Gower *et al.*, 1999). The annual production of each tissue is calculated by measuring annual stem diameter increment for each tree in a fixed plot and applying the appropriate allometric equation to estimate the change in biomass for each tissue for one or more years. The theory and measurement protocol for estimating NPP is summarized by Gower *et al.* (1999).

There are far fewer data for belowground than aboveground NPP because estimating belowground NPP (NPP_B) remains controversial and all the methods used are laborious and costly (Gower *et al.*, 1999). NPP_B is comprised of two components: coarse- and fine-roots + mycorrhizae. There is no consistent definition for diameter size class to distinguish between coarse- and fine-roots, but fine-roots are generally considered to have diameters <2–3 mm. Coarse-root biomass far exceeds fine-root biomass, but fine-root NPP exceeds coarse-root NPP. Coarse-root biomass and NPP are estimated using allometric equations and Equation (2). Historically, fine-root NPP was assumed to comprise a constant fraction of NPP_A and much of the fine-root NPP data are based on this method. Approaches to directly estimating fine-root NPP include: max–min live fine-root biomass, sequential coring fine-root biomass, ingrowth cores, minirhizotron, nitrogen mass balance and soil carbon balance. More information about the different methods, and their limitation, to estimate fine-root NPP are given in Fahey *et al.* (1998).

NPP OF MAJOR TERRESTRIAL BIOMES

A *biome* is a group of ecosystems with similar climate, physical appearance, or physiognomy, and leaf habit (evergreen versus deciduous), and is the coarsest level of classification used for terrestrial vegetation. Figure 2 summarizes the average NPP for the major terrestrial biomes. Tropical broad-leaf evergreen and temperate broad-leaf evergreen forests have the highest NPP. Boreal forests have the lowest NPP of the forest biomes because of the extremely short growing season, and the cold soils restrict decomposition, and hence nutrient availability (Landsberg and Gower, 1997). The NPP of marshes can be extremely high because

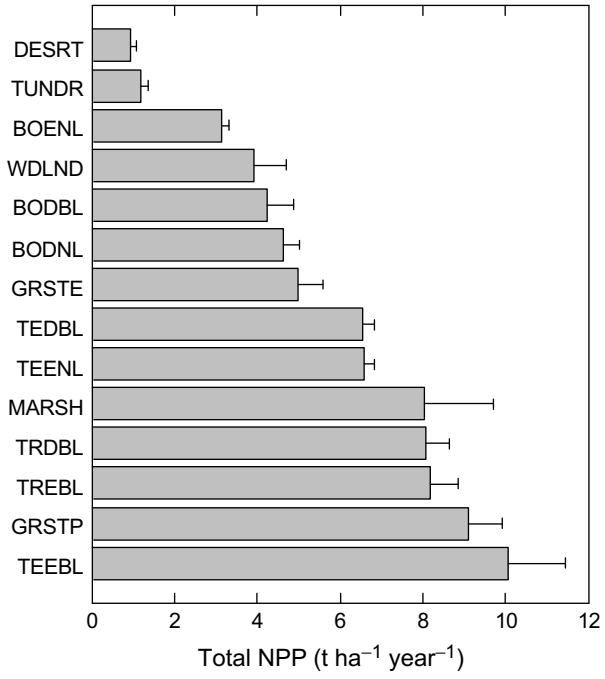


Figure 2 Average NPP for the major terrestrial biomes of the world. (Adapted from Gower *et al.*, 2001)

of the abundant moisture and high fertility resulting from the frequent deposition of nutrients and organic matter. The NPP of tropical grasslands is also very high because these biomes commonly occur on fertile soil, the climate is favorable for year-round growth, and water stress is moderate during the growing season. Deserts and tundra have the lowest NPP, but the environmental constraints on NPP differ between the two biomes. The low annual amounts of precipitation, the irregular distribution of precipitation, and large evaporative losses back to the atmosphere are the primary constraints to plant growth in deserts. On the other hand, the extremely short growing season (30–90 days), caused by extreme cold, and infertile soils are the primary constraints on NPP of tundra. Because the distribution, structure and NPP of terrestrial biomes are controlled by different environmental constraints, it is unlikely that the effect of climate change on NPP will be of similar magnitude or direction (IPCC, 1996).

ENVIRONMENTAL FACTORS INFLUENCING NPP

NPP of terrestrial ecosystems is influenced by a myriad of environmental factors, some of which are interrelated. For example, decomposition and nutrient mineralization, processes that influence nutrient availability to plants, are influenced by water availability.

Plants are photoautotrophs and require PAR for photosynthesis to occur (*see* **PAR (Photosynthetically Active Radiation)**, Volume 2). NPP is positively correlated to the annual fraction of intercepted, or more appropriately, absorbed photosynthetic active radiation (F_{APAR}); the relationship has a sound physiological basis (Landesberg and Gower, 1997). Photosynthesis by leaves is proportional to absorbed visible quanta or PAR. Although the light–photosynthesis relationship is non-linear for individual leaves, it is linear for most canopies, reducing the stand-level canopy photosynthesis to two primary factors: incident PAR and leaf area index (LAI). Incident PAR is strongly influenced by geographic location, aspect, and cloud cover. The fraction of incident PAR that is usable by plants is controlled by environmental factors that influence stomatal conductance (*see* **Carbon and Energy: Terrestrial Stores and Fluxes**, Volume 2; Landsberg and Gower, 1997). Extreme cold temperatures, high vapor pressure deficits, inadequate soil moisture, and nutrient deficiencies restrict canopy conductance, preventing the plant from using the absorbed PAR in the process of photosynthesis.

The large range in NPP among the terrestrial biomes is caused by environmental constraints on the amount of photosynthetic surface area, or LAI of vegetation and the diurnal-to-season patterns of carbon dioxide assimilation. If environmental constraints on canopy photosynthesis are accounted for, the ratio of $NPP : F_{APAR}$ is relatively conservative among biomes (Gower *et al.*, 1999). There is increasing evidence that plants optimize carbon allocation to equalize multiple resource limitations, thereby maximizing carbon gain. LAI and NPP_A are positively correlated to annual precipitation at the biome level (Figure 3a,b). The tropical broad leaf evergreen forests biome appear to occur below the general pattern, and may be explained by the fact that large fractions of the NPP data for this biome were for native *Metrosideros* spp. forests that occurred on extremely infertile soils in Hawaii. The basis for the strong positive relationship is optimization of carbon balance. The simultaneous processes of photosynthesis and transpiration, or carbon assimilation and water loss (*see* **Carbon and Energy: Terrestrial Stores and Fluxes**, Volume 2) couple the carbon and water budgets of plants. Plants can ill afford supporting large amounts of leaf area that will assimilate carbon dioxide for only a small portion of the year because respiration costs (Figure 1, arrow 2) continue, even though carbon dioxide uptake is limited by water

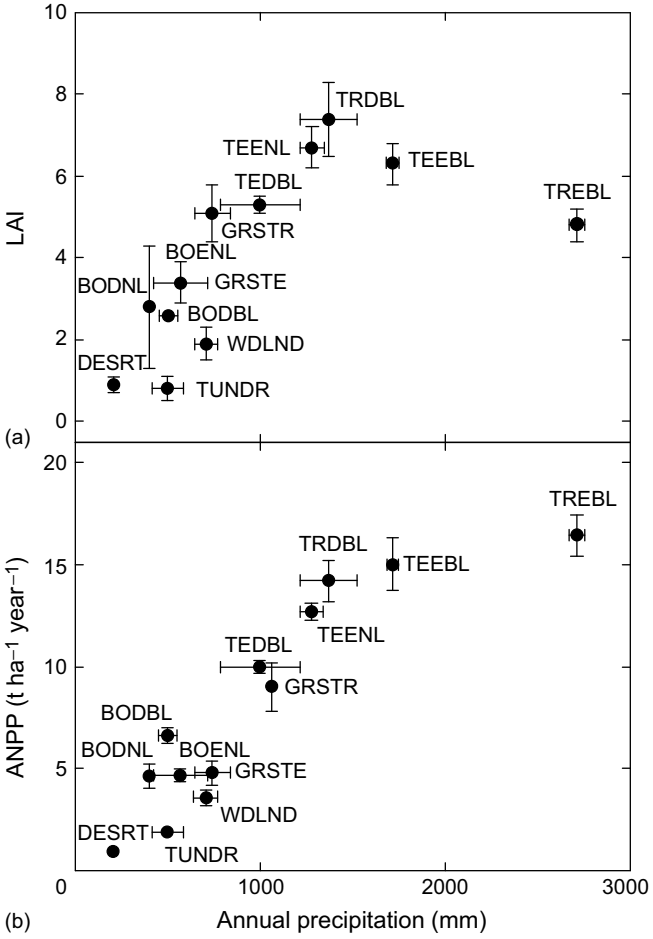


Figure 3 Relationship between annual precipitation and (a) LAI and (b) aboveground NPP for the major terrestrial biomes of the world. (Values are the biome average from Gower *et al.*, 2001)

and/or nutrient availability. An extremely short growing season limits the amount of leaf area, but evergreen forests have somewhat overcome this constraint by retaining needles longer. Thus, the structure and distribution of biomes appears to be in quasi-equilibrium with the environment in which they occur. Water and nutrient availability determine the maximum LAI of a terrestrial ecosystem. Grier and Running (1977) first demonstrated the strong positive relationship between site water balance, LAI, and NPP, and this relationship has been substantiated for temperate forests in the Midwestern US (Fassnacht and Gower, 1997). The physiologically-based interrelationships among water availability, LAI, and NPP suggest that changes in water balance, caused by either changes in precipitation or evaporative demand, will affect the structure and productivity of terrestrial ecosystems.

The availability of nutrients also influences the allocation of NPP to biomass components (Landsberg and Gower, 1997). Foliage has the highest nutrient concentrations of any major plant tissues (stem, branch, roots), thus large amounts of nutrients are required to construct a canopy. The allocation of carbon to the various plant components is determined by the availability of the most limiting resource (e.g., water, nutrients). Trees allocate less biomass to foliage production and more to fine roots and mycorrhizae production on infertile soils, and *vice versa* on fertile soils (Landsberg and Gower, 1997). Prescribed fire, which temporarily stimulates nutrient availability, promotes greater LAI of native grasslands (Aber and Melillo, 1991).

NPP is the balance between GPP and R_A (Equation 1). Views on the sensitivity of the fraction of GPP consumed by R_A to climate change have changed in recent years. The exponential relationship between respiration of individual tissues and temperature, and the asymptotic relationship between photosynthesis and temperature led early ecologists to assume that the fraction of GPP allocated to R_A would increase from cold to warm climates, and would increase with climate warming. However, the ratio of R_A : GPP for forests in contrasting climates is surprisingly stable, averaging around 0.50 (Landsberg and Gower, 1977). The stable proportion of GPP allocated to R_A can be explained by adjustments in carbon allocation to various biomass components and physiological acclimation of respiration to temperature (Ryan *et al.*, 1994).

Land-use change and disturbance history (i.e., frequency and intensity) are important components of global change and have a profound influence on NPP and NEP. Although NEP data do not exist for a disturbance sequence, a plausible scenario can be hypothesized based on a process-based understanding of the biophysical and ecological controls on NPP and heterotrophic respiration. Immediately following disturbance or land use change, NPP is small or close to zero. Heterotrophic respiration, however, is large.

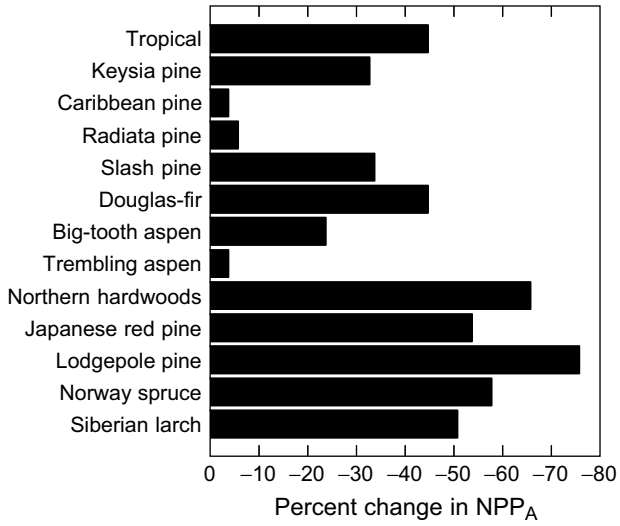


Figure 4 The relative change in NPP_A for forest age sequences in contrasting climates. A negative value denotes that NPP_A reached a maximum early during stand development and then decreased. (Data were adapted from Gower *et al.*, 1996)

The removal of the canopy increases soil temperature and moisture, which stimulate decomposition. NEP approaches zero and becomes positive as LAI and NPP increase and reach a maximum. Typically, aboveground NPP reaches a maximum during the early life of the stand, often corresponding to maximum LAI, and then declines by 20–60% as the stand ages (Figure 4). The physiological cause(s) for this age-related NPP decline are not well understood, but it has been observed in almost every forest biome of the world (Gower *et al.*, 1996). The two most plausible mechanisms responsible for the age-related NPP decline are nutrient constraints and hydraulic limitations (Gower *et al.*, 1996). The well documented age-related NPP decline and assumed relatively constant heterotrophic respiration in aging stands would imply that NEP would gradually decrease until it approaches zero.

In conclusion, NPP is an important component of the global carbon budget, and varies widely among terrestrial biomes, land use and management practices, disturbance regime.

REFERENCES

Aber, J A (1992) Nitrogen Cycling and Nitrogen Saturation in Temperate Forest Ecosystems, *Trends Ecol. Evol.*, **7**, 220–224.

Aber, J D and Melillo, J M (1991) *Terrestrial Ecosystems*, Saunders College Publishing, Philadelphia, PA.

Fahey, T, Bledsoe, C, Day, F, Ruess, R, and Smucker, A (1998) *Fine Root Production and Demography*, CRC Press, Boca Raton, FL.

Fassnacht, K S and Gower, S T (1997) Interrelationships Among Edaphic and Stand Characteristics, Leaf Area Index, and Aboveground Net Primary Productivity for Upland Forest Ecosystems in North Central Wisconsin, *Can. J. For. Res.*, **27**, 1058–1067.

Gower, S T, McMurtrie, R E, and Murty, D (1996) Aboveground Net Primary Production Decline with Stand Age: Potential Causes, *Trends Ecol. Evol.*, **11**, 378–382.

Gower, S T, Kucharik, C J, and Norman, J M (1999) Direct and Indirect Estimation of Leaf Area Index, F_{APAR}, and Net Primary Production of Terrestrial Ecosystems, *Remote Sens. Environ.*, **70**, 29–51.

Gower, S T, Feldkirchner, D C, and Olson, R J (2001) Net Primary Production of Terrestrial Ecosystems of the World, *Nature*, in press.

Grier, C C and Running, S W (1977) Leaf Area of Mature Northwestern Coniferous Forests: Relation to Site Water Balance, *Ecology*, **58**, 893–899.

Intergovernment Panel on Climate Change (1996) *Climate Change 1995: The Science of Climate Change*, Cambridge University Press, Cambridge.

Landsberg, J J and Gower, S T (1997) *Applications of Physiological Ecology to Forest Management*, Academic Press, San Diego, CA.

Ryan, M G, Linder, S, Vose, J M, and Hubbard, R M (1994) Respiration of Pine Forests, *Ecol. Bull.*, **43**, 50–63.

Vitousek, P M, Mooney, H A, Lubchenco, J, and Melillo, J M (1997) Human Domination of Earth's Ecosystems, *Science*, **277**, 494–499.

FURTHER READING

Canadell, J G, Mooney, H A, Baldocchi, D D, Berry, J A, Ehleringer, J R, Field, C B, Gower, S T, Hollinger, D Y, Hunt, J E, Jackson, R B, Running, S W, Shaver, G R, Steffen, W, Trumbore, S E, Valentini, R, and Bond, B Y (2000) Carbon Metabolism of the Terrestrial Biosphere: a Multi-technique Approach for Improved Understanding, *Ecosystems*, **3**, 115–130.

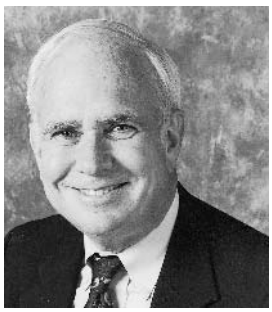
Purple Loosestrife - Alien Invader

see Plant Dispersal and Migration (*Opening essay*, Volume 2)

R

Rainforest

see Tropical Forests (**Volume 2**)



Raven, Peter

(1936–)

Peter Hamilton Raven, US botanist and environmentalist, a native Californian with a lifelong interest in natural history, has built on his science to achieve international prominence in the fields of botany and conservation. After obtaining his undergraduate degree at the University of California, Berkeley (AB, 1957) and carrying on graduate studies at the University of California, Los Angeles (PhD, 1960), Raven pursued multidisciplinary and often collaborative studies of the plant family Onagraceae, the evening primrose family, throughout the world, making it one of the best-known plant groups. He made significant contributions to biogeography, many of them with D I Axelrod, and was among the first to link the distribution of organisms in the Southern Hemisphere securely with the developing concepts of plate tectonics. In the 1960s, while on the faculty at Stanford University, he developed (with P R Ehrlich) the concept of co-evolution, and acquired a lifelong interest in conservation and sustainability that has formed the theme of many of his activities subsequently. After assuming the directorship of the Missouri Botanical Garden in 1971, Raven has become a leading exponent of the central importance of biodiversity in the US, especially through his role in several key governmental and non-governmental advisory bodies, including the US National Academy of Sciences, the National Science Board, and the President's Committee of Advisors in Science and Technology. In addition, he has devoted a major effort to lecturing throughout the world and working with governmental and non-governmental agencies in many developing countries to promote the understanding of sustainable use of biological resources. He has been especially active in Russia and the former Soviet Union generally, raising funds for the repair of the Herbarium at the Komarov Botanical Institute in St. Petersburg and serving on major granting agencies concerned with building science in those countries. Building on his role as co-editor of the Flora of China project, he has interacted widely with scientists and scientific bodies in China and helped to develop international understanding there. His work is well expressed through the development of the Missouri Botanical Garden (since 1971) into one of the major institutions dealing with plant diversity throughout the world. Its activities are concentrated mainly in Latin America and Africa. He has authored numerous influential scientific papers and books, and is co-author of *The Biology of Plants*, now in its 6th edition, a world-leading botanical text. For these efforts, he has received numerous prizes and honorary degrees, and is a member or foreign member of more than 20 academies of science throughout the world. A tireless spokesman for biodiversity, sustainable development, and conservation, Raven has contributed much to our appreciation of the significance of the loss of biodiversity to global change, and its potentially severe impact on human futures.

Photo: by Patti Gabriel.

PAUL R EHRLICH *USA*

Red eld Ratio

In 1934 Alfred C Redfield published a paper entitled *On the Proportions of Organic Derivatives in Sea Water and their Relation to the Composition of Plankton* (Redfield, 1934). In this paper he suggested that the ratio of

the nutrient elements, carbon, nitrogen and phosphorus, in sub-surface sea water is established by the decomposition of planktonic material and must reflect the ratio of these elements in the surface planktonic population. This ratio, which subsequently became known as the *Redfield ratio* has become of cardinal importance in the study of biogeochemical cycles. This is because the cycles of the nutrient elements are meshed by the processes of photosynthetic assimilation into biomass and by the catabolic regeneration of these elements. For marine plankton, an atomic ratio of P : N : C of 1 : 16 : 106 is generally accepted. This ratio does not possess the constancy characteristic of a defined chemical compound. The stoichiometry observed at any specific site will vary somewhat with the species composition of the planktonic population and with the nutritional status of the individual organisms. However, the metabolic and structural requirements of algal biochemistry will place stringent limits on the variability of the Redfield ratio. The situation becomes more complex when large multicellular organisms are involved. For instance, mature trees will contain a high proportion of carbon stored as wood. Nonetheless, Redfield's pioneering insight is crucial to our understanding of the dynamics of the biosphere. The relationship between rates of flow of nutrient elements in the biogeochemical cycles of all ecosystems is determined by the approximately constant chemistry of the metabolically active fraction of biomass.

REFERENCES

Pahlow, M and Riebesell, U (2000) Temporal Trends in Deep Ocean Redfield Ratios, *Science*, **287**, 831–833.

Redfield, A C (1934) On the Proportions of Organic Derivatives in Sea Water and their Relation to the Composition of Plankton, in *James Johnstone Memorial Volume*, University Press of Liverpool, 176–192.

G R WILLIAMS *Canada*

Redox Potential

The chemical definition of an oxidation–reduction reaction (nicknamed *redox reaction*) is one that involves the complete or partial transfer of one or more electrons from one chemical species to another. In the biosphere, redox reactions control one of the fundamental requirements of life – the need to generate biochemical energy attained through the oxidation of organic carbon (CH_2O) to carbon dioxide (CO_2). The respiration reaction with molecular oxygen (O_2) yields considerable energy:

$$CH_2O + O_2 \longrightarrow CO_2 + H_2O + 502\,\text{kJ mol}^{-1}\ CH_2O$$

Plentiful microorganisms in the aquatic and terrestrial biospheres utilize O_2 in respiration. In soils, waters, and sediments, however, this reaction is limited because of relatively low concentrations of O_2. However, life does not stop in anaerobic environments. Aerobic organisms die off, but others arise that utilize oxidizing agents that substitute for O_2 in the respiration reaction shown above. For example, nitrate (NO_3^-) can serve the same purpose, albeit with less energy per molecule of organic carbon consumed:

$$CH_2O + \frac{4}{5}NO_3^- + \frac{4}{5}H^+ \longrightarrow CO_2 + \frac{2}{5}N_2 + \frac{7}{5}H_2O + 475\,\text{kJ mol}^{-1}\ CH_2O$$

There are in fact six molecules (including O_2) that control the oxidizing power of the biosphere. These are: (1) O_2; (2) NO_3^-; (3) manganese dioxide (MnO_2); (4) ferric hydroxide ($Fe(OH)_3$); (5) sulfate (SO_4^{2-}); and (6) CO_2. As illustrated in Figure 1, the choice of oxidizer proceeds via a preordained ordering, rather than by random utilization of all available oxidants. This ordering results from the relative oxidation potential of each molecule in its reaction with CH_2O, or, in chemical terms, its redox potential. The lower the redox potential, the lower the energy obtained from the reaction.

In an environment with a mix of oxidants, microbes preferentially select the one that yields the most energy. When that oxidant is used up, microbes use the oxidant with the

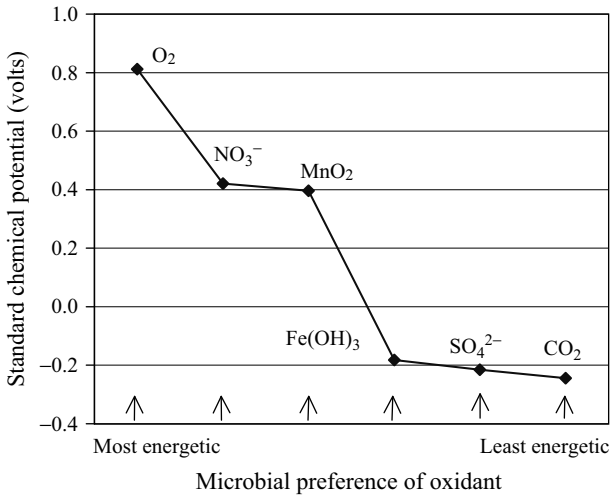


Figure 1 Six common oxidants in the biosphere and the sequence in which they are utilized by microbes

next highest yield of energy. Thus, one can think of redox potential as a kind of chemical switch that determines the observed sequence by which microorganisms utilize oxidants. In clean lakes with low amounts of CH_2O, microbes use O_2 as the preferred oxidant. When the amount of organic carbon exceeds about $3.4\,mg\,l^{-1}$ water, the O_2 will be depleted and nitrate becomes the preferred oxidant. When all of the nitrate is depleted, then MnO_2 becomes the preferred oxidant, and so on.

WILLIAM M STIGLIANI *USA*

Remote Sensing

Susan Ustin
University of California, Davis, CA, USA

Remote sensing imaging technologies provide a basis for monitoring, analyzing, and quantifying the information needed to characterize natural and human-induced environmental changes across a wide range of space and time scales. Imaging systems span spatial resolutions from submeters to kilometers and are available on a wide range of platforms, ranging from hand-held digital cameras to satellites. Many satellites and airborne imaging sensors are suitable for synoptic mapping and monitoring. These instruments provide, first, repeated measurements of the earth in a referenced spatial context, allowing explicit analysis of change detection, and second, in the case of satellite sensors, repeated measurements from as short a time as 30 minutes to years. There is, however, a general trade-off between frequent coverage by satellites and high spatial resolution.

Remote sensing instruments measure a physical property, electromagnetic (EM) radiation in specific spectral regions, thereby providing information about the Earth's surface that can be interpreted in terms of environmental conditions, and allow the user to spatially visualize this information. The past decade has seen extensive use of this technology for global change research and a transition from reliance on visual patterns and empirical relationships to ones based on physical processes and principles. These interpretative advances have depended on greatly improved detector technologies that became available during the 1990s, which are capable of measuring additional regions of the EM spectrum beyond the visible and near-infrared (NIR) wavelengths.

The development of spatial analysis tools using geographic information systems, and locational mapping using global positioning satellite systems occurred simultaneously with new remote sensing capabilities. More powerful computer capabilities and new quantitative analysis tools developed in the 1990s have contributed a suite of new tools to characterize site conditions in their spatially explicit contexts. While measurements at all scales are relevant to global change research, satellites are of particular interest due to the acquisition of images at regular repeated qintervals, making it more practical to quantify environmental changes.

SATELLITE AND AIRBORNE SENSORS USEFUL FOR ASSESSING GLOBAL ENVIRONMENTAL CHANGE

Global and Continental Scale Processes

An initial consideration of environmental assessments is to start by considering the scale(s) of the processes for which information is desired, the frequency with which repeat measurements need to be acquired to determine changes, and the type of environmental properties for which EM characteristics have potential to provide information. At the global scale, the weather satellites (metsats) provide the best source of environmental change information for processes manifested at regional to global scales. Many countries operate metsats, with programs going back to the 1960s. Data are of two types, the polar orbiting sun-synchronous satellites and the geostationary satellites. A range of meteorological data is acquired from multiple instruments but the most widely used for environmental change studies are from the imaging systems. These systems have relatively coarse ground resolution (kilometers) but may be acquired over much of the Earth at very frequent time intervals (0.5 hours to daily). The most widely used sensor for assessing global change has been the National Oceanic and Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR), available since 1983, which acquires data in visible ($0.4-0.7\,\mu m$), NIR ($0.7-1.0\,\mu m$), and thermal infrared (TIR) ($3-12\,\mu m$) spectral regions.

One of the first significant papers to examine feedbacks between the atmosphere and the terrestrial biosphere using satellite data was that by Tucker *et al.* (1986), who showed that AVHRR satellite measurements of planetary greenness, a property correlated with the amount of terrestrial vegetation and the absorption of photosynthetically active radiation (*see* **PAR (Photosynthetically Active Radiation)**, Volume 2), followed the cycle of atmospheric carbon dioxide (CO_2) concentrations measured by the global network of ground-based recording stations. A four-year data record showed that the monthly averaged greenness measurements followed the seasonal and interannual cycles and the latitudinal differences that had been reported by the CO_2 network. This study demonstrated that satellite data responded to terrestrial photosynthesis and respiration processes and provided unprecedented information about

temporal and spatial patterns in biospheric activity. Since then, numerous studies have used satellite data to establish carbon and water fluxes and surface energy budgets. For example, a 10-year satellite observation record was used to show that the arctic growing season is lengthening due to planetary warming and atmospheric carbon dynamics (Myneni *et al.*, 1997).

In other cases, increasing desertification due to unwise land uses and loss of vegetation following extensive regional droughts has been demonstrated using the AVHRR records. For example, the onset and extent of the devastating drought in the African Sahel in the early 1980s was demonstrated using AVHRR data (Tucker and Choudhury, 1987) and numerous other examples have been investigated since that time. Deforestation has also been examined using these data (e.g., Malingreau and Tucker, 1998), although due to the smaller scales of forestry practice, it has been less effective. Weather satellite data have contributed to improved global estimates of wildfires and the extent of biomass burning (French *et al.*, 1996), and have provided the most accurate maps of regional (Goward *et al.*, 1986) and global vegetation patterns (DeFries and Townshend, 1999). Satellite data have also been used to assess suitable habitat for disease vectors, such as Rift Valley Fever (Linthicum *et al.*, 1987).

REGIONAL ASSESSMENTS

The first non-military satellite to provide highly detailed images of the Earth was Landsat, launched in 1972, with four visible and NIR bands on the Multispectral Scanner (MSS). Landsat 4 was launched in 1982 and also carried a new seven-band sensor, the Thematic Mapper (TM). Beginning with the launch of Landsat-7 in 1999, an enhanced version of the TM (ETM+) became the only instrument on this satellite. The ETM+ has 30 m ground resolution pixels with six bands in the visible and reflected infrared regions, and one 60-m resolution TIR band. It also has a 15 m panchromatic sharpening band, which, when fused with the multispectral bands, produces a sharper visual image. The wavelength position of Landsat's spectral bands was chosen to provide information about plant pigments and water content, and structure of plant canopies, for accurate discrimination of environmental conditions. These data provide a bridge between local measurements at the scale of field experiments and the coarser resolution of the weather satellites. The multispectral data from Landsat contribute information that can be interpreted based on the general classes of materials composing the landscape (e.g., soils, vegetation, geologic outcrops, water, etc.). Landsat initiated a wide and growing range of environmental applications and provided the incentive to develop and launch many other land-observing satellites. Stoney (1997) identified 30 satellite sensors that are planned for launch by governments and private companies between 1999–2002. While some satellite programs have not been successful, it is clear we are at a new stage in the availability and quality of remotely sensed information. The Environmental Remote Sensing Center (ERSC) web site (http://rs320h.ersc.wisc.edu/ERSC/Resources/EOSF.html) provides a description of these satellites.

Among all classes of satellites there is a clear trend toward increasing the number of spectral bands and decreasing the minimum spatial resolution. For example, the National Aeronautics and Space Administration's (NASA's) Terra satellite launched in 1999 included two multi-band instruments, the Moderate Resolution Imaging Spectrometer (MODIS), with 36 spectral bands and spatial resolution up to 250 m pixels, and the Advance Spaceborne Thermal Emission and Reflection Radiometer (ASTER), with 14 bands and up to 15 m pixel resolution. It is now possible to obtain high spatial and temporal resolution data for virtually any type of spectral resolution sensor; therefore the problem now is to identify the most suitable data to address specific global change problems.

Radar Remote Sensing Systems

Radar satellites like the European Remote Sensing Satellite, the Japanese Earth Resources Satellite and the Canadian Radarsat are single-channel radar imagers with minimum spatial resolutions from about 100 m^2 to 10 km^2 and repeat cycles of a few days to about two months. At radar frequencies, clouds and water vapor are transparent, making these satellites almost the only option for viewing the ground for terrain with persistent cloud cover, e.g., in the tropics and the polar regions. Radar backscatter increases with the dielectric constant of natural material, and water is the only common material with a high dielectric constant. Therefore, image patterns generally reflect the water content of vegetation and soils, and produce images that resemble those detected by optical sensors that measure light absorption by photosynthetic pigments. Stereo radar data pairs produce interoferometric images that are used to generate high vertical relief digital elevation maps. NASA's Space Shuttle Topographic Mapping Mission in 2000 was designed to produce a detailed high-resolution digital elevation map for about 70% of the land surface of the Earth.

Multispectral Remote Sensing Systems

The Landsat MSS and TM are examples of multispectral satellites, while the French SPOT (Système pour l'Observation de la Terre) satellite and the Indian Resource Satellite are examples of commercialized multispectral satellites. Multispectral Satellites typically have between four to eight spectral bands and moderate spatial resolutions

(10–30 m), providing local to regional area coverage. There are plans for an additional 10 satellites in the early 2000s that will have Landsat-like multispectral capabilities. Most will have at least one short-wave infrared (SWIR) band (in the 1.5–2.5 μm range), a TIR band, and a higher resolution (1–15 m) panchromatic band. These sensors are designed to provide detailed spatial mapping and analysis of land cover and land use classes at the landscape scale. Band placements allow roads and developed infrastructure, water bodies, and general vegetation patterns to be mapped. Land use patterns, e.g., from agriculture and forestry, are routinely defined, as are boundaries between communities with different physiognomic forms, like grasslands and forests. The sensors provide information about vegetative cover and spatial patterns related to fragmentation and natural disturbances, e.g., wildfires, floods, droughts, defoliation, and severe storms. Coregistration of satellite images from different years provides detailed information on changing patterns of land use over time and makes it possible to assess rates of land conversion.

Regional environmental processes are better studied using the spatial detail that Landsat type data allow. Figure 1 depicts three years of MSS data over a 20-year period, centered over Lake Tahoe (largest lake, shown in black), on the border between California and Nevada. Each image (center panel) shows the midsummer greenness for the region (whitest pixels), covering an area approximately 34 225 km^2. Numerous small finger lakes are seen in the Sierra Nevada Mountains. The enlarged boxes shown above the regional images illustrate: (1) the temporary drying and loss of Washoe Lake (northeast of Lake Tahoe) due to a regional drought, which was in its seventh year in 1992 (above); and (2) the increasing number of clear cut logging areas and their extent in the Sierra Nevada Mountains (below). The logging activity changes forest connectivity in the region, creating habitat fragmentation and changes in species composition and age class structure. These changes lead to further impacts on the wildfire regime, erosion and runoff, and on the fluxes of gases and exchange of energy. These consequences are localized and the impacts depend on specific site conditions.

Hyperspectral Imaging Systems

These instruments represent the newest and most advanced remote sensing technology. Hyperspectral imaging spectrometers have large numbers (50–1000 s) of adjacent narrow wavelength spectral bands so that the band sequence produces a high resolution spectrum for each pixel. Due to the large number of coregistered spectral

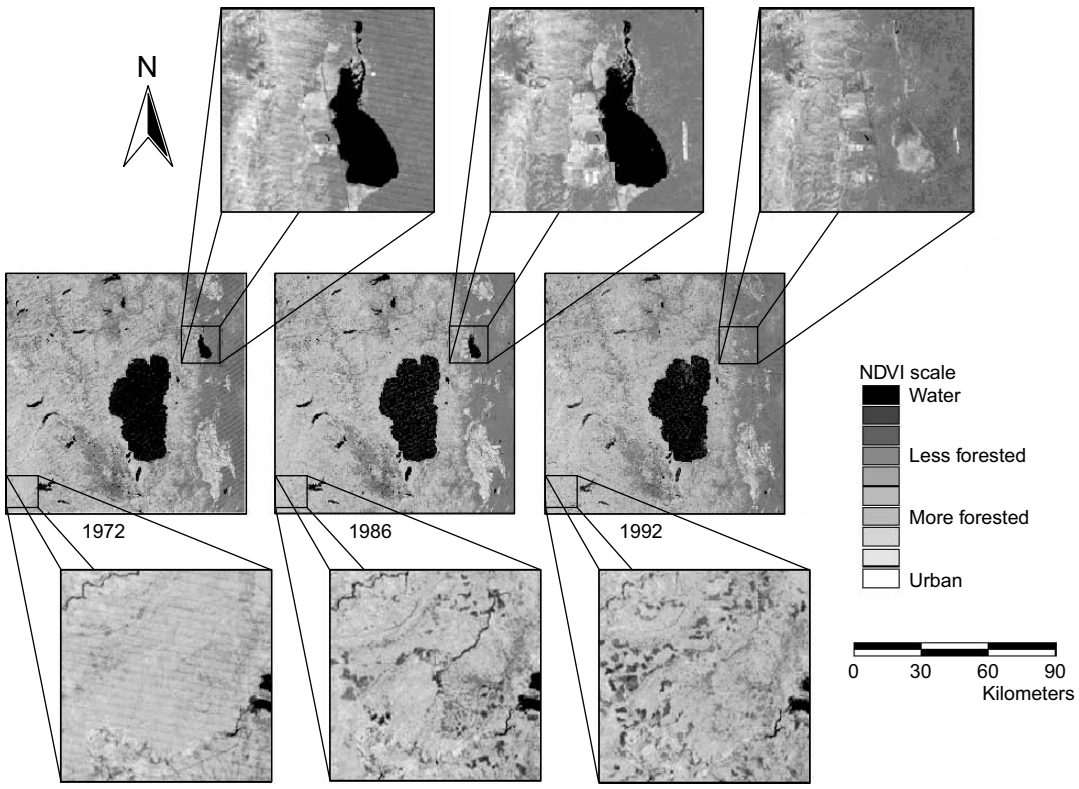


Figure 1 MSS normalized difference vegetation index (NDVI) images of Lake Tahoe region (California-Nevada, USA) in 1972, 1986, and 1992. NDVI is a normalized ratio of the NIR and red spectral bands, where brighter values indicate greater plant cover and darker values indicate lower plant cover. Diagonal striping in the 1972 image is an artifact

bands, a dataset is often described as an image cube. A way to visualize this is to imagine that all the images (created by displaying all the pixel values for each spectral band) are displayed as pages in a book. The first page in the book is the shortest wavelength band and the last page is the longest wavelength band. The data set that a multispectral satellite produces is analogous to a 4–8 page book compared to a 150–500 page book produced by a hyperspectral satellite.

The first hyperspectral satellite, Hyperion, was launched on NASA's EO-1 platform in November 2000, to be followed by three commercial hyperspectral sensors that are proposed for the 2000–2005 time frame. Each sensor has 150–220 spectral bands coverage covering the 0.40–2.50 μm spectral region and pixel resolutions of 30 m or better. More than 30 airborne hyperspectral sensors have been operated in different countries (see http://rst.gsfc.nasa.gov/sect13/is_list.html) and it is likely that more will be available in the future.

These spectrometers can be used to identify the surface expression of specific biogeochemical conditions in plants, soils, and geologic materials. Many minerals, clay minerals, carbonates, organic matter, and iron can be identified in soil (Clark, 1999; Ben-Dor *et al.*, 1999). In green foliage, photosynthetic pigments determine absorption of energy in the visible spectrum and water in the NIR and SWIR. These compounds determine the overall spectral absorption features of green foliage and plant canopies. For example, in dry leaves and woody tissues, structural carbon compounds like cellulose and lignin (Ustin *et al.*, 1999) and nitrogen compounds can be detected (Curran, 1989). From these spectral building blocks, it is possible to infer information regarding ecosystem health and possibly, the presence of stress agents, like contaminants. Most of the spectroscopy literature has been based on laboratory studies of environmental materials; however, some studies using airborne imaging spectrometer data demonstrate the potential to extract biogeochemical information from remotely sensed data. It is possible to map the distribution of canopy water content and pigments.

Hyperspatial Imaging Sensors

High spatial resolution image data (1–5 m pixels) are useful to identify localized environmental impacts and to evaluate the representativity of field data collected at specific locations, with respect to the larger region. There are potentially 12 commercial satellites that may acquire very high spatial resolution imagery from space (Stoney, 1997), focused largely on traditional markets of the photogrammetry industry. The Ikonos (Space Imaging Corporation), has a 1 m resolution panchromatic sensor and a 4 m resolution four-band (blue, green, red, and NIR) multispectral sensor on their satellite. Due to the narrow ground swath, global coverage will require periods from four months to two years to acquire a repeat measurement (Stoney, 1997). Off-nadir pointing capability will increase the frequency of repeating site measurements. Virtually any application where aerial photography is currently used is a candidate for using this type of sensor data. Clearly, these data provide a mechanism to scale measurements from field samples to landscapes. Many uses related to global issues are relevant for detailed site characterization at these scales, including mapping community composition in ecologically sensitive habitats, monitoring rare and endangered species, monitoring habitats threatened by invasion of noxious weeds, and characterization of habitat structure and successional states.

See also: Table 6 in **Biomass Burning**, Volume 3.

REFERENCES

Ben-Dor, E, Irons, J R, and Epema, G (1999) Soil Reflectance, in *Manual of Remote Sensing, Vol. 3, Remote Sensing for the Earth Sciences*, ed A Rencz, John Wiley & Sons, New York, 111–188.

Clark, R N (1999) Spectroscopy of Rocks and Minerals and Principles of Spectroscopy, in *Manual of Remote Sensing, Vol. 3, Remote Sensing for the Earth Sciences*, ed A Rencz, John Wiley & Sons, New York, **3**, 3–58.

Curran, P J (1989) Remote Sensing of Foliar Chemistry, *Remote Sens. Environ.*, **30**, 271–278.

DeFries, R S and Townshend, J R G (1999) Global Land Cover Characterization From Satellite Data: from Research to Operational Implementation? *Global Ecol. Biogeogr.*, **8**, 367–379.

French, N H F, Kasischke, E, Johnson, R D, Bourgeau-Chavez, L L, Frick, A L, and Ustin, S L (1996) Estimating Fire-related Carbon Flux in Alaska Boreal Forests using Multisensor Remote Sensing Data, in *Biomass Burning and Climate Change*, ed J S Levine, MIT Press, Cambridge, MA, **2**, 808–826.

Goward, S N, Tucker, C J, and Dye, D G (1986) North American Vegetation Patterns Observed with the NOAA-7 Advanced Very High Resolution Radiometer, *Vegetatio*, **64**, 3–14.

Linthicum, K J, Bailey, C L, Davies, F G, and Tucker, C J (1987) Detection of Rift Valley Fever Viral Activity in Kenya by Satellite Remote Sensing Imagery, *Science*, **235**, 1656–1659.

Malingreau, J P and Tucker, C J (1988) Large-scale Deforestation in the Southeastern Amazon Basin of Brazil, *Ambio*, **17**, 49–55.

Myneni, R B, Keeling, C D, Tucker, C J, Asrar, G, and Nemani, R R (1997) Increased Plant Growth in the Northern High Latitudes From 1981–1991, *Nature*, **386**, 698–702.

Stoney, W E (1997) Land Sensing Satellites in the Year 2000, *Proc. Int. Geosci. and Remote Sensing Symposium.*, Singapore, Aug. 7, 1997 (http://geo.arc.nasa.gov/sge/landsat/pecora.html).

Tucker, C J, Fung, I Y, Keeling, C D, and Gammon, R H (1986) Relationship between Atmospheric Carbon Dioxide Variations and a Satellite-derived Vegetation Index, *Nature*, **319**, 195–199.

Tucker, C J and Choudhury, B J (1987) Satellite Remote Sensing of Drought Conditions, *Remote Sens. Environ.*, **23**, 243–252.

Ustin, S L, Smith, M O, Jacquemoud, S, Verstraete, M M, and Govaerts, Y (1999) Geobotany: Vegetation Mapping in Earth

Sciences, in *Manual of Remote Sensing, Vol. 3, Remote Sensing for the Earth Sciences*, ed A Rencz and R Ryerson, John Wiley & Sons, New York, **3**, 189–248.

Remote Sensing, Terrestrial Systems

Anthony C Janetos
World Resources Institute, Washington, DC, USA

Field and forest, valley and mountain, wherever one goes, the land surface varies. The landscape is so varied that it has taken the work of many thousands of scientists to try to document its characteristics by going from place to place, compiling inventories of land characteristics and vegetation type. While much was accomplished and many detailed characteristics can be documented, an important problem was trying to ensure uniformity in the classification schemes, both because different scientists were making the observations and because observations were made at different times.

In contrast to the difficulties of surface measurements, satellite-based remote sensing can provide global coverage on a daily or near-daily basis. Such efforts were begun in the mid-1970s, and assembled data sets now provide rapid and relatively accurate sampling for some types of ecosystem status and characteristics. For a few variables, there is an approximately 25 year time series of reasonably well-calibrated data with nearly global coverage. In addition, because the costs of analyzing imagery are low, remote sensing offers essentially the only affordable means of sampling for measurements over very large spatial areas. Even though the capital costs of developing and launching satellite missions are extremely large, operational costs for satellite systems are not consistently higher than the personnel costs of large in situ monitoring efforts, and for whole Earth coverage, satellites become extremely cost effective.

LONG-TERM SATELLITE IMAGERY

Two types of satellite imagery are widely used by ecologists: (1) Landsat thematic mapper (TM), and (2) advanced very high resolution radiometer (AVHRR). Both cover the majority of the Earth's surface, with TM imagery providing 30 m resolution and AVHRR imagery providing 1 km resolution. Each system records the intensity of radiation reflected from the Earth's atmosphere-surface system in 5–7 spectral bands in the visible, near-infrared (IR), and thermal IR parts of the spectrum, and then uses this information to extract surface images.

The Landsat data record starts in 1972. Over this period, three primary instruments have been used. Video images, some of which are still archived, provided the initial record of surface conditions. The multi-spectral scanner, with approximately 80 m spatial resolution, was used along with the video system and very quickly became the preferred instrument for quantitative analysis. Since 1982, the primary instrument has been the TM, which measures six spectral bands in the visible and shortwave-near IR, all of which have 30 m spatial resolution, and one thermal band with 120 m spatial resolution. A Landsat scene is approximately 185 km on a side, and the combination of swath width and orbital characteristics of the satellite means that each location on Earth is revisited at about 16–18 day intervals. There have now been seven Landsat missions launched, although Landsat 6 failed to achieve orbit.

The US Geological Survey's Eros Data Center archives nearly global data from the Landsat system for the years 1972–1980, although the particular dates of cloud-free acquisitions vary from place to place on the globe. Starting in 1980, the US policy of commercializing the Landsat system resulted in a less comprehensive coverage being held in US archives (although the record for the US itself remains relatively complete). However, there have been 10–17 international ground stations operating during the period of the Landsat missions, many of which have maintained accessible data archives, so global data sets can be derived for the 1980s and early 1990s.

Landsat 7 was launched in early 1999, carrying an enhanced TM plus (ETM+), and this has improved the availability of data dramatically. The ETM+ observes the same spectral bands as previous Landsats, thus building on the already long time series, but also has a 15 m black and white band that is useful both in its own right and for sharpening the imagery from the 30 m bands. The ETM+ also has enhanced calibration capabilities compared to previous Landsats. Most importantly, however, the system used on Landsat 7 is designed to update the US-held global archive on approximately a seasonal basis, in addition to adding to the archives held by international ground stations (Goward and Williams, 1997). Data are also more affordable than was earlier the case.

The second primary data set on terrestrial ecosystems is derived from the AVHRR, which is a standard instrument on the National Oceanic and Atmospheric Administration polar-orbiting meteorological satellites. Although it was originally intended for use as a meteorological instrument, it has several spectral bands in the visible, near IR, and IR parts of the spectrum that are useful for monitoring terrestrial ecosystems. However, the AVHRR collects data with very different spatial characteristics than does Landsat, reflecting its origin as a meteorological instrument.

Global area coverage data, which are resampled on-board the satellite before being transmitted, have nominally 4 km spatial resolution. Local area coverage data, which are only directly transmitted to the ground, are nominally 1 km spatial resolution. The combination of the swath width (ca. 1000 km) of the sensor and the orbital characteristics of the satellite platform result in between daily and twice daily global coverage. However, unlike the Landsat TM or ETM+, whose radiometric properties are known and monitored with a high degree of accuracy, the AVHRR has no provision for on-orbit calibration; in addition, the various satellite orbits have drifted substantially over the years. Thus, the data include some noise because of the poorly known calibration of the instruments and poorly quantified intercalibration of instruments on different platforms. Therefore, although roughly decadal-length time series of truly global data exist, they have to be extensively preprocessed before they can be used for quantitative analyses of ecosystem properties and processes.

NEW SERIES OF SATELLITE IMAGERY

While efforts are being made to continue the long-term series, new, more detailed data sets are being generated. For example, the Earth observing system AM platform, known as *Terra*, was launched in late 1999. Terra carries five instruments, of which the moderate resolution imaging spectroradiometer (MODIS) is of particular interest from an ecological perspective. MODIS has 32 spectral channels, with stringently defined radiometric calibration. The majority of the spectral channels have 1 km spatial resolution, although a few have 250–500 m resolution. Global data sets are being acquired roughly every few days. Very importantly, the combination of spectral and spatial characteristics enables a more accurate calculation of a variety of vegetation indices, land-cover products, fire detection products, and land-cover change indices.

An even newer instrument will be the vegetation canopy lidar (VCL), an instrument capable of measuring the vertical foliage height profile everywhere on Earth. The technology has already been tested on the space shuttle, and the imager itself is currently being carried aboard a research aircraft. The satellite-mounted VCL will allow monitoring of harvesting and regrowth of vegetation, aboveground biomass, and structural habitats important to animal species.

The commercial sector is also increasingly active in providing remote sensing information. The IKONOS satellite, for example, provides very high spatial resolution data: 1 m resolution in black and white, and 4 m resolution multi-spectral information. Ecological and other scientific analyses of these data will almost certainly grow in number beyond this early stage. Analyses of recently declassified data with similar characteristics indicate that they will have great utility for vegetation classification and analysis of land-cover change. The recently declassified data themselves also provide the opportunity for analyzing time series of land-cover back into the 1960s in some locations.

INFORMATION HANDLING AND CALIBRATION

The use of remote sensing information for ecological analyses is still very much in its infancy, although tremendous strides have been made over the past five or six years. The ready availability of time series of satellite measurements with well-understood precision and accuracy has been an important issue in the use of remotely sensed data. In addition to the technical issues of sensor calibration and intercalibration of instruments on different satellite platforms, data access and affordability have also had to be considered because of the high capital costs of developing satellite systems and in part because of their commercial potential.

For many years, the technological challenges involved in simply handling and processing the data were so large that they inhibited the use of the systems by any but the most sophisticated laboratories. The rapid improvement, both in terms of cost and performance, of computer hardware and software is removing many of the technical impediments for the use of remote sensing. Two of the most important remaining issues, however, are the care, maintenance, and accessibility of data archives, and the intercalibration of the remote sensing instrumentation itself.

The latter is a serious challenge for the providers of remote sensing information, but is absolutely necessary for the information to be useful. Any instrumentation that is used to measure the long-term variation of ecosystems (or of any quantity) must be calibrated sufficiently that there can be confidence in ascribing changes in measurement to the ecosystem being measured and not to instrument drift. While this seems a simple requirement, achieving calibration precise enough for quantitative scientific measurements is tremendously difficult in practice. Very delicate hardware can be calibrated and tested on the ground, but it is subjected to tremendous vibrational stresses during launch, and then exposed to the thermal and radiation stresses of low-Earth orbit. Maintaining instrument calibration is thus a non-trivial task, although it can be achieved, as the Landsat data record shows. As important as calibrating each individual instrument, however, is ensuring that calibration is maintained for those data sets that are intrinsically long-term. This requires ensuring that data set calibration be achieved which lasts longer than the lifetime of any one instrument. This is best achieved by ensuring that successive instruments are flown for a period of overlap. This is especially important as both the hardware and software of

the instrumentation evolve so that trends due to degradation of instrument performance and offsets due to changing orbital geometry or new technology are identified and quantified before any analyses are attempted.

See also: **Earth Observing Systems**, Volume 1.

REFERENCE

Goward, S N and Williams, D L (1997) Landsat and Earth System Science: Development of Terrestrial Monitoring, *Photogram. Eng. Remote Sensing*, **LXIII**, 887–900.

Resilience

Lance Gunderson[1], C S Holling[2], L Pritchard[1], and G D Peterson[3]

[1]*Emory University, Atlanta, GA, USA*
[2]*University of Florida, Gainesville, FL, USA*
[3]*University of Wisconsin, Madison, WI, USA*

Resilience has two meanings in the ecological literature, both related to system state and disturbance. Engineering resilience is the time of return to a global equilibrium following a disturbance. Ecological resilience is the amount of disturbance that a system can absorb before it changes state. Ecological resilience is based on the demonstrated property of alternative stable states in ecological systems, while engineering resilience implies only one stable state (and global equilibrium). The alternative meanings of resilience have significant implications for application of the concept to understanding and managing complex systems.

Resilience has been defined in two different ways in the ecological literature, each reflecting different aspects of stability. One definition focuses on efficiency, constancy and predictability – all attributes of engineers' desire for fail-safe design. The other focuses on persistence, change and unpredictability – all attributes embraced and celebrated by evolutionary biologists and by those who search for safe-fail designs. Holling (1973) first emphasized these different aspects of stability to draw attention to the tensions between efficiency and persistence, between constancy and change, and between predictability and unpredictability.

The more common definition, which we term *engineering resilience* (Holling, 1996), considers ecological systems to exist close to a stable steady-state. Resilience is the ability to return to the steady state following a perturbation (Pimm, 1984; O'Neill *et al.*, 1986; Tilman and Downing, 1994; Tilman, 1996). This idea of disturbance away from and return to a stable state is at the center of economic theory as well (Varian, 1992; Kamien and Schwartz, 1991).

The second definition, which we term *ecological resilience* (Walker *et al.*, 1969; Holling, 1996), emphasizes conditions far from any stable steady-state, where instabilities can flip a system into another regime of behavior, i.e., to another stability domain (Holling, 1973). In this case, resilience is measured by the magnitude of disturbance that can be absorbed before the system redefines its structure by changing the variables and processes that control behavior.

The differences between these two aspects of stability – essentially between a focus on maintaining efficiency of function (engineering resilience) versus a focus on maintaining existence of function (ecological resilience) – are so fundamental that they can become alternative paradigms whose devotees reflect traditions of a discipline or of an attitude more than a reality of nature. Those using the concept of engineering resilience tend to explore system behavior near a known stable state, while those examining ecological resilience tend to search for alternative stable states, and the properties of the boundaries between states.

Those who explore engineering resilience and the near equilibrium behavior of ecosystems operate in the primarily deductive tradition of mathematical theory (e.g., Pimm, 1984) that imagines simplified, untouched ecological systems, or they draw upon the traditions of engineering, that are motivated by the need to design systems with a single operating objective (Waide and Webster, 1976; De Angelis, 1980; O'Neill *et al.*, 1986). These approaches simplify the mathematics and accommodate the engineer's drive to develop optimal designs. However, there is an implicit assumption that ecosystems exhibit only one equilibrium steady state, or, if other operating states exist, that those states should be avoided.

On the other hand, those who emphasize ecological resilience come from traditions of applied mathematics and applied resource ecology at the scale of ecosystems, e.g., of the dynamics and management of fresh water systems (Fiering, 1982), of forests (Holling *et al.*, 1977), of fisheries (Walters, 1986), of semi-arid grasslands (Walker *et al.*, 1969) and of interacting populations in nature (Sinclair *et al.*, 1990; Dublin *et al.*, 1990). Since these studies are rooted in inductive rather than deductive theory formulation and in experience with the impacts of large scale management actions, it becomes clear that the variability of critical variables forms and maintains the stability landscape, and that when this variability is reduced, an ecosystem can flip from one state to another.

In economics, there has also been a focus on the single stable state. The history of economics has been to rapidly move from establishing the existence of a general equilibrium to issues of equilibrium uniqueness or stability. If multiple equilibria are shown to theoretically exist, then the

challenge that is usually taken up is to theoretically eliminate stable states by proposing individuals' strategic expectations and predetermined normative and social institutions. This approach does not examine or explain the conditions that can cause a system to move from one stability domain to another. However, the identification of multi-stable states due to path dependency (Arthur *et al.*, 1987), chreodic development (Clark and Juma, 1987) and production nonconvexities such as increasing returns to scale (David, 1985) has reintroduced multiple stable states to economics.

The existence, or at least importance, of multiple or a single stable state determines the appropriateness of an engineering or ecological approach to resilience. If it is assumed that only one stable state exists or can be designed to exist, then the only possible definition and measures for resilience are near equilibrium ones – such as characteristic return time. That is certainly consistent with the engineer's desire to make things work, not to intentionally make things that break down or suddenly shift their behavior. However, nature and human society are different.

See also: **Ecosystem Services and Costing**, Volume 2.

REFERENCES

Arthur, W B, Ermoliev, Y M, and Kaniovski, Y M (1987) Path-dependent Processes and the Emergence of Macro-structure, *Eur. J. Oper. Res.*, **30**, 294–303.

Clark, N and Juma, C (1987) *Long-run Economics: an Evolutionary Approach to Economics Growth*, Pinter, London, 1–206.

David, P A (1985) Clio and the Economics of QWERTY, *Am. Econ. Rev.*, **75**(2), 332–337.

De Angelis, D L (1980) Energy Flow, Nutrient Cycling and Ecosystem Resilience, *Ecology*, **61**, 764–771.

Dublin, H T, Sinclair, A R E, and McGlade, J (1990) Elephants and Fire as Causes of Multiple Stable States in the Serengeti-mara Woodlands, *J. Anim. Ecol.*, **59**, 1147–1164.

Fiering, M B (1982) Alternative Indices of Resilience, *Water Resour. Res.*, **18**, 33–39.

Holling, C S (1973) Resilience and Stability of Ecological Systems, *Ann. Rev. Ecol. Syst.*, **4**, 1–23.

Holling, C S (1996) *Engineering Resilience Versus Ecological Resilience. Engineering Within Ecological Constraints*, P C Schulze, National Academy Press, Washington, DC, 31–43.

Holling, C S, Jones, D D, and Clark, W C (1977) Ecological Policy Design: a Case Study of Forest and Pest Management, in *Proceedings of a Conference on Pest Management, October 1976*, eds G A Norton and C S Holling, IIASA CP-77-6, 13–90, Laxenburg, Austria.

Kamien, M I and Schwartz, N L (1991) *Dynamic Optimization: the Calculus of Variations and Optimal Control in Economics and Management*, Elsevier Science, North-Holland, Amsterdam, 1–377.

O'Neill, R V, De Angelis, D L, Waide, J B, and Allen, T F H (1986) *A Hierarchical Concept of Ecosystems*, Princeton University Press, Princeton, NJ.

Pimm, S L (1984) *The Balance of Nature*, University of Chicago Press, Chicago, IL.

Sinclair, A R E, Olsen, P D, and Redhead, T D (1990) Can Predators Regulate Small Mammal Populations? Evidence From House Mouse Outbreaks in Australia, *Oikos*, **59**, 382–392.

Tilman, D and Downing, J A (1994) Biodiversity and Stability in Grasslands, *Nature*, **367**, 363–365.

Tilman, D (1996) Biodiversity: Population Versus Ecosystem Stability, *Ecology*, **77**, 350–363.

Varian, H R (1992) *Microeconomic Analysis*, Norton, New York, 1–506.

Waide, J B and Webster, J R (1976) Engineering Systems Analysis: Applicability to Ecosystems, in *Systems Analysis and Simulation in Ecology*, Vol. IV, ed B C Patten, Academic Press, New York, 329–371.

Walker, B H, Ludwig, D, Holling, C S, and Peterman, R M (1969) Stability of Semi-arid Savanna Grazing Systems, *Ecology*, **69**, 473–498.

Walters, C J (1986) *Adaptive Management of Renewable Resources*, Macmillan Press, New York.

Respiration

Respiration in plants is the oxidation of organic substrates by organisms to release useful energy, usually accompanied by the uptake of oxygen and the release of carbon dioxide (CO_2). All organisms rely on respiration to release, for metabolic purposes, the potential chemical energy stored in respiratory substrates (characteristically reduced organic compounds such as sugars and sugar polymers). Respiration is normally accomplished by a complex series of oxidative reactions, some of the energy so released being captured and temporarily stored in high-energy compounds such as adenosine triphosphate and reduced pyridine nucleotides. These high-energy compounds are then available to drive synthetic reactions, to power cellular or intercellular transport mechanisms, or to drive other energy-requiring activities such as the contraction of muscle-like proteins found in many plant cells.

The end product of respiration is CO_2, and normal plant respiration requires an approximately equal uptake of oxygen. Some respiratory activity can go on anaerobically by fermentation, in which the initial steps of respiratory metabolism permit the production of high-energy compounds at the expense of reactions that convert sugars to alcohol or other similar end products with the production of CO_2. However, fermentation releases only a fraction of the energy from the substrate compared with normal aerobic respiration.

Respiration of a plant is not a constant of that plant, but varies under the influence of many external and internal factors and controls. It is temperature sensitive and, without internal control, doubles on a 10 °C rise in temperature.

Internal controls include a series of biochemical and metabolic feedback mechanisms that prevent the organism from using up its resources merely because they are there, but increase or decrease the rate according to need. In the event that the normal substrates of respiration (sugars and starch in most higher plants) become seriously depleted, such mechanisms will allow the use of alternate (but metabolically or energetically less desirable) substrates such as proteins and amino acids or lipids.

Respiration and fermentation are characterized by the production of CO_2, and of organic end products such as alcohols in the case of fermentation. Also, because the respiratory process is only approximately 30% efficient in trapping the chemical potential of substrate molecules in high-energy intermediates, a substantial amount of heat is produced in respiration (ca 400 kcal $mole^{-1}$ glucose). This is not usually important in leaves, for example, but may raise the temperature of sequestered or insulated locations with a high biotic content (particularly of microorganisms) to the point of spontaneous combustion, as in damp hay bales.

The main ecological impact of respiration is CO_2 production, and respiration is one of the most important components of the biogeochemical cycling (*see* **Biogeochemical Cycle**, Volume 2) of carbon. Diurnal variation of plant respiration and photosynthesis as well as the annual variation in plant respiratory activity due to climatic temperature variations add very considerably to the complexity of the global carbon cycle. While plants do indeed remove large amounts of CO_2 from the atmosphere by photosynthesis, they return much of it in respiration. The difference (the amount of carbon that is actually sequestered) represents plant growth and litter production. Ultimately even that will be returned to the atmosphere (by microbial respiration) unless over even longer times it is sequestered as carbonate rock or coal, oil or gas.

R G S BIDWELL *Canada*

Restoration, Ecosystem

John P McCarty[1] and Joy B Zedler[2]

[1] *Environmental Protection Agency, Washington, DC, USA*
[2] *University of Wisconsin, Madison, WI, USA*

Ongoing changes to the earth's climate, disruptions of nutrient cycles, and changes in land use are expected to result in extensive disruptions to natural communities and ecosystems. These disruptions will likely result in the extinction of valued species of plants and animals and the loss of important ecological services. At the same time, the potential strategies for minimizing the effects of these changes on ecosystems at the regional scale are extremely limited.

Restoration of degraded ecosystems is one promising strategy for reducing the negative impacts of global change. Ecological restoration attempts to promote recovery of ecosystems damaged by human activities. Restoration is increasingly attempted in freshwater, wetland, grassland, and forest ecosystems. The results of these efforts have been mixed. Many restoration projects involve high inputs of time and money for each hectare restored and ongoing maintenance is frequently required. In many cases, the species that make up a community can be established but healthy ecosystem functioning is more difficult to restore.

In the future, restoration ecology will likely play an important role in speeding the recovery of ecosystems damaged by global change. While some of the difficulties facing ongoing restoration projects are likely to be overcome with time and effort, it is probable that future ecological restoration will be unable to restore degraded natural communities completely.

Restoration and related land-management projects offer potential tools for mitigating some of the effects of global change on ecosystems. According to the Society for Ecological Restoration, "Ecological restoration is the process of assisting the recovery and management of ecological integrity. Ecological integrity includes a critical range of variability in biodiversity, ecological processes and structures, regional and historical context, and sustainable cultural practices." This approach presents an attractive basis for addressing some of the effects anticipated from changes in climate, nutrient cycles, and land use patterns.

Around the globe, restoration efforts are underway to reverse the damages of historical land use to restore ecosystem services and natural diversity. Most projects are very small, but a few large regional programs are underway (e.g., restoring the hydrology of the Everglades (Florida, US) and restoring parts of the San Francisco Bay Delta, US). The ultimate goal of many restoration projects is to reestablish a community that existed prior to human disturbance. For the purposes of this article, we also include *rehabilitation* projects, where it is recognized that past conditions cannot be recreated. Instead, establishing some alternative ecosystem is the target. We also consider *reclamation* projects that attempt to create an ecosystem that will provide specific ecological services, such as constructing a wetland to treat runoff, without regard for the species composition or overall ecological integrity.

The typical goal of restoration ecology is to restore a past state. Yet the very nature of environmental changes at the

global scale, especially global climate change, means that conditions in a given area are changing in such a way that they will no longer be appropriate for the communities of recent history. Instead, restoration will provide tools and experience to smooth 'the recovery and management of ecological integrity'. For a given area, the target ecosystem will be one consistent with the new and changing climatic conditions. Restoration may also be valuable in helping to preserve currently existing, distinct communities by establishing them in new areas where climate has become favorable.

In this article, we first review how restoration ecology is being applied in different types of ecosystems. For each ecosystem, we consider how global change is expected to affect species and communities and how restoration ecology might address problems that emerge. We focus on climate change, rather than other aspects of global change, because changes in climate are expected to affect virtually all ecosystems (Intergovernmental Panel on Climate Change, 1996), and because prevention measures that might reduce impacts of changing climate have not been widely adopted. The lack of prevention strategies increases the importance of approaches such as restoration that focus on minimizing or reversing negative impacts on ecosystems. Restoration has proven to be a useful tool for addressing ecosystem degradation due to non-climate components of global change such as acid or nutrient deposition and changes in land use. Next, we discuss how ongoing restoration efforts might be modified to prepare for future climate change. Finally, we discuss the limitations and advantages of restoration ecology as a strategy for helping ecosystems adapt to climate change. While restoration has many advantages over other possible strategies for mitigating the effects of climate change on ecosystems, it is clear that none of the currently available tools will be sufficient to avoid the negative effects of climate change.

ECOLOGICAL RESTORATION AND POTENTIAL APPLICATIONS TO CLIMATE CHANGE

Traditionally, much of the focus of environmental science has been on removing sources of stress from the environment, such as removing toxic chemicals or reducing inputs of nutrients. Global climate change challenges this traditional approach in part because elimination of warming associated with global climate change presents formidable technological, social, and political obstacles. Because it is not likely that we will be able to reverse climate change in time to prevent widespread disruptions of ecosystems, it becomes necessary to find other means of reducing the negative impacts of climate change. One approach will be to speed the natural processes responsible for ecosystem recovery. We suggest that restoration, especially large regional projects, can help address these changes in the future. It is also possible for ongoing projects to incorporate actions that will help buffer ecosystems against the negative impacts of climate change.

Ecosystems will differ in how they respond to climate change, which would have far-reaching effects on natural ecosystems (Intergovernmental Panel on Climate Change, 1996). In general, the effects of climate change will act on populations and species. The tolerances of each species for changing climatic conditions will differ, but it is expected that many species will no longer be able to persist in part or all of their current geographic ranges (Intergovernmental Panel on Climate Change, 1996). Even species that can tolerate wide ranges of climatic conditions are likely to be impacted as their predators, competitors, and prey respond to climate change. In response, species might shift their geographic range to suitable areas, change their behavior so they are active at cooler times of year, or evolve to tolerate the new conditions. Species that fail to adjust to new conditions rapidly enough will become locally extinct.

Here we describe progress in ecological restoration in freshwater, wetland, grassland, and forest ecosystems. For each ecosystem type, we briefly outline special challenges they will likely face from changes in climate and outline how restoration might be applied to meet those challenges.

Freshwater Ecosystems

Under generally accepted climate change scenarios, aquatic ecosystems face possible changes in both temperature and the amount of water available (Intergovernmental Panel on Climate Change, 1996). Freshwater systems are already severely challenged and contain large numbers of threatened and endangered species (Dobson *et al.*, 1997). Many aquatic organisms are sensitive to water temperature and will face increasing stress from changes in thermal regimes.

Efforts to rehabilitate degraded lake ecosystems date back many decades. Most early efforts were responses to excessive nutrient inputs and resulting eutrophication. Noteworthy examples include Lake Washington in Seattle, US and Lake Erie in the US and Canada, where reductions in nitrogen and phosphorus inputs produced dramatic improvements in water quality. Similar approaches will be most relevant where climate change exacerbates existing problems, such as increased nutrient loading due to increased precipitation or decreases in dissolved oxygen concentration due to higher water temperature.

Acid deposition is one component of global environmental change, and the scale of the efforts devoted to treating impacted areas suggests that substantial resources can be dedicated to regional restoration efforts. One approach to restoring lake ecosystems affected by acid deposition involves application of lime to neutralize acid (Schindler, 1997). In Sweden alone, 7500 lakes have been repeatedly

limed as part of efforts to restore acidified lakes (Appleberg, 1998). Regular application of lime can moderate acidity but additional restoration, such as restocking of fish, is usually needed to restore communities fully (Schindler, 1997; Appleberg, 1998).

Changes in thermal characteristics of habitats will likely exacerbate the already high rates of endangerment of aquatic species (Dobson *et al.*, 1997). Many endangered and threatened aquatic species have geographic ranges that encompass a narrow range of thermal conditions and have limited abilities to disperse to new habitats. Rising water temperatures will tend to exclude fish species from areas where they now exist, while appropriate thermal habitats will occur to the north of current ranges. Restoration ecology could play a role in providing new habitats for those species in thermally appropriate areas. The slow rate of dispersal to new habitats, especially for species in isolated lakes, will severely limit the ability of fish to shift their geographic range in response to climate change. Translocations could speed the adjustment of dispersal-limited species. However, many of the existing problems in aquatic systems result from introductions of species by humans to new habitats. Introducing species to cooler waters would need to follow thorough research on the potential impacts to species already in place and acknowledgment of the ecological risks involved. Creation of new habitats in enclosed basins (reservoirs) may be one approach to minimize the risks from such assisted range extensions.

Of particular importance to future efforts will be improving techniques for dealing with invasions by nuisance species. Changes in water temperature will probably allow further expansion of the ranges of aggressive exotic species, exacerbating existing problems. Experience with aggressive invaders such as rusty crayfish *Orconectes rusticus* and zebra mussel *Dreissena polymorpha* as well as numerous aquatic plant and fish species shows the difficulties managers face in attempting to limit ecosystem damage caused by invaders. Introductions of predatory game fish have disrupted many unique communities, but significant restoration can be initiated if the fish can be eliminated (McNaught *et al.*, 1999). Unfortunately, attempts to control the species composition of aquatic communities, especially smaller, non-game fish, algae, and higher plants, have had limited success (Welch and Cooke, 1987) (*see* **Lakes and Rivers**, Volume 2).

Wetland Ecosystems

Climate change threatens to stress wetland species through direct effects of temperature and carbon dioxide (CO_2) concentration and through changes in hydrology due to precipitation, evaporation and sea level changes (Michener *et al.*, 1997). Temperature increases are likely to damage wetlands that store carbon in the form of organic matter. Canadian peatlands and the Arctic tundra could release major quantities of carbon to the atmosphere through increased decomposition rates as a result of warming or drying (Intergovernmental Panel on Climate Change, 1996). Experiments with multiple factors and many types of wetlands are needed to make long-term predictions of community change, however. The hydrologic restoration of drained peatlands in advance of climate change would help buffer the globe against these impacts.

Small ponded wetlands and dry-end wetlands will be very sensitive to altered hydrology (Sorenson *et al.*, 1998). Because half of North America's waterfowl are reared in prairie potholes, and because duck production is directly correlated with the number of small ponds, there would be measurable impacts to waterfowl following changes in precipitation and evaporation (Sorenson *et al.*, 1998). The acceleration of current efforts to restore drained potholes throughout the Midwest could mitigate some of the negative impacts in advance, but biodiversity would still likely decline, as not all species return with renewed ponding (Galatowitsch and van der Valk, 1996).

Many wetlands have lost their natural water regimes, especially flood pulsing and the accompanying mechanical disturbances and nutrient influxes that invigorate floodplain wetlands (Middleton, 1999). In some cases, restoration may be accomplished simply by removing impediments to water flow (Gilbert and Anderson, 1998). Where dams permanently reduce flooding of wetlands, significant improvements can be achieved by periodic water releases that mimic seasonal floods (Vaselaar, 1997; Middleton, 1999).

Hydrological changes may be greatest for coastal wetlands, where sea levels in many areas may rise more than 0.5 m by 2100 (Intergovernmental Panel on Climate Change, 1996). Salt marshes will be lost where sea walls obstruct their migration inland and/or where sedimentation cannot keep pace with inundation. Restoration efforts would help mitigate future losses due to sea level rise. Methods include reestablishing tidal flow through removal of dikes, reintroducing species, and excavating new wetlands (Simenstad and Thom, 1996; Zedler, 1996a; Williams and Watford, 1997). Inundation problems can be combated by encouraging sedimentation and marsh building (Smit *et al.*, 1997) or accommodated by 'managed retreat'. In the latter process, levees can be breached to allow salt marsh to reestablish in low lying areas (Packham and Willis, 1997; Gilbert and Anderson, 1998) or broad inland buffers can be set aside to allow salt marsh migration up slopes.

One of the most damaging effects of climate change on coastal wetlands may be from increased storm frequencies and magnitudes. Extreme high tides can carry salts inland on to intolerant vegetation and soils. Salty soils will favor the inland migration of halophytes, which will be necessary to maintain salt marshes where sediment accretion cannot keep up with inundation (Michener *et al.*, 1997).

Species will likely migrate into newly inundated areas at different rates and be affected by different substrates and competitors inland, so some species may need to be translocated. The high-intertidal marsh of Southern California, US, is a good example of an assemblage that may need to be translocated in anticipation of sea-level rise. Three high-marsh perennials are already rare and they rarely reproduce from seed. *Frankenia palmeri* has only one natural occurrence in the US, but in Mexico it occurs in both salt marsh and upland areas. *Salicornia subterminalis* (= *Arthrocnemum subterminale*) and *Monanthochloe littoralis* are more widespread in Southern California but their populations have been greatly reduced by trails, roads, and other disturbances. These species show promise for experimental planting in upland buffers. Plants should be grown from seed to provide genetic diversity, so that selection could operate after tides begin to inundate the populations. The benefits (enhanced genetic diversity) and risks (contamination of local gene pools) of bringing in seed from Mexico would first need to be weighed.

Much of the motivation for restoring wetlands is to provide ecological services, such as nutrient removal and retention. Restoring specific functions is usually more difficult than simply reestablishing plant cover (Mitsch and Wilson, 1996; Simenstad and Thom, 1996; Zedler, 1996b), and even state of the art techniques cannot yet restore a heavily disturbed ecosystem to a specific, self sustaining state.

The most restorable sites are those with some remnant of natural topography, hydrology, or species composition. The return of the channeled Kissimmee River to its historical riverbed rapidly restored native wetland plant cover and attracted populations of desirable invertebrates, fishes, and birds in the pilot demonstration project (Toth, 1996). More challenging is the plan to restore water flows to the Everglades, as water supplies are finite and Southern Florida, US, metropolitan areas permanently divert much of the water that used to form the 'river of grass'. On the positive side, the US National Park System sustains a large area of native vegetation, and providing more water should help it persist. Least restorable is the San Francisco Bay Delta, a huge complex system of river channels (mostly leveed for flood control), water diversion canals (designed to move water to Southern California), and islands (leveed, because farming has caused substantial subsidence). Although the main ecosystem restoration goal is to enhance native species and reduce exotics (especially fishes), there is little natural habitat left and water demands preclude full hydrologic restoration. Restoration to historical conditions is not possible, but this is just the type of project that could include large experiments to create climate-buffered ecosystems. Some portion of the project could be designated as an experimental climate change reserve (*see* **Marshes, Anthropogenic Changes**, Volume 3).

Grassland Ecosystems

Grasslands and prairies occur where conditions are wet enough for grasses but too dry for shrubs and trees (or too disturbed by fire or grazing). Because moisture and fire are key factors, grassland will be especially sensitive to changes in precipitation and in climate variability.

Grasslands and prairies have undergone extensive restoration over the last 65 years to recreate the aesthetic structure of prairies, to reintroduce fire to rehabilitate prairie remnants, and to create functioning ecosystems on abandoned farmland (Kline and Howell, 1987; Muller *et al.*, 1998). Restoring biodiversity requires continual planting and continual management of exotic invasions. Some species are notoriously difficult to maintain, while many invasive exotics are notoriously difficult to eradicate (Cottam, 1987; Kline and Howell, 1987). A good understanding of the ecology of the organisms involved and ongoing management, are needed to maintain restored habitats (Cottam, 1987). Difficulties are compounded when the original soil is badly degraded and a diverse seed bank absent. Success of restoration attempts of species-rich grasslands in France has been slowest and least predictable in areas where intensive agriculture or engineering processes have caused degradation of soil quality and loss of seed banks (Muller *et al.*, 1998).

In North America, prairie restoration tools include the use of herbicides, cultivation, and planting (Cottam, 1987; Kline and Howell, 1987). Labor intensive seed gathering and planting has been successful on small-scale plots (Cottam, 1987; Packard and Mutel, 1997). Larger scale efforts have adapted agricultural practices to the harvest and planting of native species.

Perhaps the most important lesson from work on grassland restoration is the usefulness of controlling fire and other disturbances as a means of altering or maintaining community composition. Many grassland communities depend on the periodic disturbance caused by fire to prevent woody species from invading (Packard and Mutel, 1997; Davison and Kindscher, 1999). The manipulation of disturbance in the form of fire or grazing can be used to restore native communities. Compared to many restoration activities, controlled burning or grazing can impact large areas of habitat, as will be needed to mitigate effects of global change.

Prairie restoration has focused almost exclusively on the establishment of plant communities, assuming that animals will find their way to suitable habitats (Kline and Howell, 1987). While many prairie specialist birds and small mammals do become established on restored prairies, many others will not do so without active intervention. For example, mound building ants play an important role in natural prairie ecosystems by cultivating the soil and providing open space where forbs can become established. Experience with the restored prairies at the University

of Wisconsin Arboretum, US, found that even 50 years after plant communities were in place, ants were absent from much of the available habitat (Kline and Howell, 1987).

Because most restored prairies require continual maintenance (time and money), opportunities for mitigating the effects of climate change exist primarily at small spatial scales. These efforts could provide important source populations for the eventual natural spread of prairie species. Even relatively small patches of habitat may prove valuable in this role. For example, the 24 ha Curtis Prairie at the University of Wisconsin Arboretum, US, has maintained populations of about 170 native prairie plant species, even though it is surrounded by non-prairie habitats (Cottam, 1987). Manipulation of disturbance will be possible at larger scales, provided grassland species are already present (*see* **Temperate Grasslands**, Volume 2; **Tropical Savannas**, Volume 2).

Forest Ecosystems

Forest ecosystems experience ongoing degradation and destruction through harvesting and clearing activities. Remaining fragments are subject to multiple stresses such as air pollution, invasive species, and new pests and pathogens. Changes in temperature and moisture will be a source of added stress, which will likely alter forest productivity and ecosystem functioning, while the distributions of most temperate zone trees will shift poleward, changing species composition (Intergovernmental Panel on Climate Change, 1996).

From the long history of tree planting, we know that many species can thrive when planted far outside their natural ranges (Ashby, 1987). Hence, some tree species should be able to survive despite climatic changes, while others should be able to grow when planted in favorable sites. While this is encouraging, forests also provide some of the clearest examples of the need to plant locally adapted populations or ecotypes (Ashby, 1987). The ease with which trees are established in artificial settings does not mean that planted populations will be self-sustaining or that other forest species of plants and animals will establish a functioning ecosystem (Ashby, 1987). One of the more difficult factors to account for in future restoration of forests will be the tight symbiosis between many plants and mycorrhizal fungi. Most plants depend on mycorrhizae for efficient uptake of water and nutrients, but the abilities of soil organisms to adapt to climate change are unknown. The ability of restoration to promote the rapid reestablishment of forest functioning will depend in large part on our ability to establish healthy populations of soil organisms (Haselwandter, 1997).

Longleaf pine (*Pinus palustris*) forest ecosystems provide a good case history for restoration of forest landscapes. Longleaf pine forests once covered 60% of the Southeastern US coastal plain, stretching from North Carolina to Texas. Longleaf pine forests host a surprising diversity of plant and animal species, including a large number of threatened and endangered species. Critical to the maintenance of these systems are periodic fires carried by understory plants that prevent invasion by deciduous oaks. Today, less than 10% of the original forests remain intact (Johnson and Gjerstad, 1998). This loss of a valuable ecosystem has led to an increasing interest in restoring longleaf pine communities by reintroducing fire and establishing populations of indigenous plants and animals. Some of the largest restoration projects ever undertaken are now in progress. The US Department of Defense military bases alone manage approximately 400 000 ha of longleaf-type vegetation (Johnson and Gjerstad, 1998).

Wiregrass (*Aristida* spp.) forms the dominant groundcover in most natural longleaf communities and provides fuel to carry fires (Means, 1997). Attempts to create a wiregrass understory in longleaf pine restorations provide an informative example of the limitations of restoration efforts. While wiregrass populations are present in many stands undergoing restoration, and plantings can be established with relative ease, these populations have typically failed to spread (Means, 1997; Seamon, 1998). Thus, while the species is present in the community, it fails to expand to serve a necessary function in the ecosystem (e.g., promoting the spread of beneficial fires).

Large-scale restoration of tropical forests has been initiated in both Latin America, Asia, and Australia. In Costa Rica, restoration efforts have focused on facilitating the natural recovery of ecosystems by promoting regeneration of forest trees and by attempting to control disturbance due to fire (Janzen, 1988; Holl, 1998). An alternative approach is to increase the ecological integrity of the large areas of tropical forest that have been converted to timber plantations. These areas may be improved, though not restored to a pristine state, by promoting the use of native species, planting mixtures of species, and altering landscape patterns of forests to provide buffers (Lamb, 1998). The experience gained through landscape-scale forest restorations is highly valuable for evaluating impacts and adaptation to global change, as well as our ability to compensate through restoration (*see* **Boreal Forest**, Volume 2; **Nitrogen Deposition on Forests**, Volume 2; **Temperate Coniferous Forests**, Volume 2; **Temperate Deciduous Forests**, Volume 2; **Rainforest**, Volume 2).

BUILDING A CLIMATE CHANGE BUFFER INTO RESTORATION PROJECTS

Significant uncertainties about future climatic conditions exist. This is especially true at the relatively small scale

of ecological processes and for precipitation and climatic variability. In order for current restoration efforts to prepare for the negative impacts of climate change, we need to understand the nature of climatic changes anticipated for a given site, and how land-management projects can best be planned to accommodate them. Significant advances in both climatic modeling and in understanding how ecosystems respond to changing climate will be needed before this is possible.

All sizes of restoration projects can build in some buffer against the negative aspects of climate change, but the greatest potential lies with the larger projects, such as the longleaf pine forest, Kissimmee River, Everglades, and San Francisco Bay Delta projects in the US. Large size restoration projects, like large preserves, can sustain more species and more genotypes. They can also support more effective exotic-species control programs, as they are more likely to have smaller perimeter/interior ratios, thus reducing chances for reinvasion. Finally, they would have the space needed to include well-planned translocation experiments.

Until more specific information is available, recommendations for buffering restoration sites are general, and they focus on increasing the resilience of communities in the face of environmental variability. These actions include:

1. Reintroduction of the full range of biodiversity (plants, animals, microbes) that was thought to occur prior to site degradation.
2. Maintenance of topographic heterogeneity; too many restoration sites have artificial landforms that are unnaturally smooth, due to modification by bulldozers.
3. Reintroduction of diverse gene pools. Growing plants from seed in preference to vegetative propagation. If the latter is necessary, ramets should be from multiple clones.
4. Eradication of exotic species and monitoring range expansion by new invasive species.
5. Allowing native species to remain on site. Some exceptions might be highly aggressive natives that might lower overall biodiversity.
6. Carefully monitored, experimental translocations of declining native species at restoration sites outside these species' natural range. This recommendation is consistent with biodiversity conservation efforts involving gardens, zoos, and seed banks but must be carefully planned to avoid adverse effects from introducing non-native species.

These suggestions are not specific to any climate scenario but are meant to increase the resilience of natural communities faced with a variety of anthropogenic stresses. Thus, they could be applied before reliable site-specific information on climate change becomes available.

THE STATUS OF THE SCIENCE OF RESTORATION ECOLOGY

Several scientific challenges will need to be met before climate-buffered ecosystems can be designed.

1. Foremost, we will need reliable predictions of when, where, and how environmental conditions will change.
2. Next, we need to predict how species will respond to these changes, especially which species are most likely to go extinct as a result of climate change and which ecosystem functions will be impaired.
3. We need to be able to select appropriate tools to sustain biodiversity.
4. We need to understand the relationship between ecosystem structure and functioning. More attention will need to be directed towards the interactions among species that govern processes such as nutrient and water cycles, community stability, and the regulation of primary productivity.
5. We need to document the current status of reference sites. A major constraint on restoring historical habitats is our lack of knowledge of their structure and functioning. But looking ahead 50 to 150 years, we will have no excuse for not making detailed records of conditions such as species composition, productivity, and nutrient dynamics. With new technology, such as remote sensing, we can plot the precise locations of rare species populations and the geographical boundaries of species with various traits, for which we hypothesize likely expansions, compressions, range shifts, or no change. This will provide future biodiversity managers a basis for assessing change and considering restoration actions.

Climate change may affect critical components of the ecosystem that are yet unknown. The processes that go on below ground, for example, are at the frontier of ecology. The tight symbiosis between plants and mycorrhizal fungi has not received the attention it deserves. Owing to their poor dispersal and habitat specificity, mycorrhizae need to be added to highly degraded restoration sites (Haselwandter, 1997). While this is a common tool, we do not know the abilities of mycorrhizae and other soil organisms to adapt to climate change. Further complications may involve other microbes.

Research will need to be done before we can rely on restoration tools to ameliorate the effects of climate change at the large scale. Some encouragement comes from fire and water management, however, as both have been used to restore large expanses of habitats. The release of flood waters from Glen Canyon provides a positive model (Vaselaar, 1997), although the magnitude of planned flooding can never match historical extremes. We

encourage replication of flooding experiments over time and large-scale experimentation in general.

CONCLUSIONS

It is unlikely that restoration tools can be used to prevent negative impacts on ecosystems caused by climate change. However, global change can and should be considered when designing restoration projects. Detailed adaptation of ongoing ecological restorations is hampered by the uncertainties about the exact nature of future changes in climate and about ecosystem responses at specific sites. Until we have reliable, site-specific predictions, it is premature to suggest translocations into natural habitats, as we know too little about how the receiving site might be affected. A possible exception would be restoration sites designed to sustain rare plants and animals beyond their natural range (climate change reserves).

In general, restoration tools will be more useful after natural systems are noticeably degraded by climate change. If ecosystem degradation due to changing climate cannot be prevented, then ecological restoration may be one of the few strategies available to ameliorate the damage. Once again, restoration is unlikely to compensate fully for the loss of species and the reduction in ecosystem functioning that will likely result from significant climate change.

We expect that most efforts to ameliorate climate change will be of small spatial scale. Nevertheless, they will be important. Small patches of habitat provide refugia for sedentary species and sources of propagules and colonists for future natural expansion. Preservation of genetic diversity will offer greater potential for natural selection in the face of changing climate. The result, especially if the restoration site is large, should tend toward a climate-buffered ecosystem.

Progress in developing restoration tools provides both a reason for hope and sometimes a sobering lesson in the complex task that awaits us. Additional research advances will be needed to understand how ecosystems function and how humans can intervene in disturbed communities to promote diverse and healthy ecosystems.

ACKNOWLEDGMENTS

We thank the American Association for the Advancement of Science (AAAS) and the US Environmental Protection Agency's National Center for Environment Assessment (for support of JPM), and the National Science Foundation (DEB 96-19875 for support of JBZ) during preparation of this manuscript. L L Wolfenbarger provided valuable input on earlier versions of this paper. The views are those of the authors and do not reflect those of the USEPA.

REFERENCES

Appleberg, M (1998) Restructuring of Fish Assemblages in Swedish Lakes following Amelioration of Acid Stress through Liming, *Restor. Ecol.*, **6**, 343–352.

Ashby, W C (1987) Forests, in *Restoration Ecology: a Synthetic Approach to Ecological Research*, eds W R Jordan, III, M E Gilpin, and J D Aber, Cambridge University Press, New York, 89–108.

Cottam, G (1987) Community Dynamics on an Artificial Prairie, in *Restoration Ecology: a Synthetic Approach to Ecological Research*, eds W R Jordan, III, M E Gilpin, and J D Aber, Cambridge University Press, New York, 257–270.

Davison, C and Kindscher, K (1999) Tools for Diversity: Fire, Grazing and Mowing on Tallgrass Prairies, *Ecol. Restor.*, **17**, 136–143.

Dobson, A P, Rodriguez, J P, Roberts, W M, and Wilcove, D S (1997) Geographic Distribution of Endangered Species in the United States, *Science*, **275**, 550–553.

Galatowitsch, S M and van der Valk, A (1996) The Vegetation of Restored and Natural Prairie Wetlands, *Ecol. Appl.*, **6**, 102–112.

Gilbert, O L and Anderson, P (1998) *Habitat Creation and Repair*, Oxford University Press, New York.

Haselwandter, K (1997) Soil Micro-organisms, Mycorrhiza, and Restoration Ecology, in *Restoration Ecology and Sustainable Development*, eds K M Urbanska, N R Webb, and P J Edwards, Cambridge University Press, New York, 65–80.

Holl, K D (1998) Tropical Moist Forest Restoration on Agricultural Land in Latin America, in *Damaged Ecosystems and Restoration*, ed B C Rana, World Scientific, New Jersey, 25–41.

Intergovernmental Panel on Climate Change (IPCC) (1996) *Climate Change 1995: Impacts, Adaptations and Mitigation of Climate Change: Scientific-technical Analyses*, eds R T Watson, M C Zinyowera, R H Moss, and D J Dokken, Cambridge University Press, New York.

Janzen, D H (1988) Guanacaste National Park: Tropical Ecological and Cultural Restoration, in *Rehabilitating Damaged Ecosystems*, ed J Cairns, Jr, CRC Press, Boca Raton, FL, 143–192.

Johnson, R and Gjerstad, D (1998) Landscape-scale Restoration of the Longleaf Pine Ecosystem, *Restor. Manage. Notes*, **16**, 41–45.

Kline, V M and Howell, E A (1987) Prairies, in *Restoration Ecology: a Synthetic Approach to Ecological Research*, eds W R Jordan, III, M E Gilpin, and J D Aber, Cambridge University Press, New York, 75–87.

Lamb, D (1998) Large-scale Ecological Restoration of Degraded Tropical Forest Lands: the Potential Role of Timber Plantations, *Restor. Ecol.*, **6**, 271–279.

McNaught, A S, Schindler, D W, Parker, B R, Paul, A J, Anderson, R S, Donald, D B, and Agbeti, M (1999) Restoration of the Food Web of an Alpine lake following Fish Stocking, *Limnol. Oceanogr.*, **44**, 127–136.

Means, D B (1997) Wiregrass Restoration: Probable Shading Effects in a Slash Pine Plantation, *Restor. Manage. Notes*, **15**, 52–55.

Michener, W K, Blood, E R, Bildstein, K L, Brinson, M M, and Gardner, L R (1997) Climate Change, Hurricanes and Tropical

Storms, and Rising Sea Level in Coastal Wetlands, *Ecol. Appl.*, **7**, 770–801.
Middleton, B (1999) *Wetland Restoration, Flood Pulsing, and Disturbance Dynamics*, Wiley, New York.
Mitsch, W J and Wilson, R F (1996) Improving the Success of Wetland Creation and Restoration with Know-how, Time, and Self-design, *Ecol. Appl.*, **6**, 77–83.
Muller, S, Dutoit, T, Alard, D, and Grévilliot, F (1998) Restoration and Rehabilitation of Species-rich Grassland Ecosystems in France: a Review, *Restor. Ecol.*, **6**, 94–101.
Packham, J R and Willis, A J (1997) *Ecology of Dunes, Salt Marsh and Shingle*, Chapman and Hall, London.
Packard, S and Mutel, C F (1997) *The Tallgrass Restoration Handbook*, Island Press, Washington, DC.
Schindler, D W (1997) Liming to Restore Acidified Lakes and Streams: a Typical Approach to Restoring Damaged Ecosystems? *Restor. Ecol.*, **5**, 1–6.
Seamon, G (1998) A Longleaf Pine Sandhill Restoration in Northwest Florida, *Restor. Manage. Notes*, **16**, 46–50.
Simenstad, C A and Thom, R M (1996) Functional Equivalency Trajectories of the Restored Gog-Le-Hi-Te Estuarine Wetland, *Ecol. Appl.*, **6**, 38–56.
Smit, H, van der Velde, G, Smits, R, and Coops, H (1997) Ecosystem Responses in the Rhine-Meuse Delta during Two Decades after Enclosure and Steps Toward Estuary Restoration, *Estuaries*, **20**, 504–520.
Sorenson, L G, Goldberg, R, Root, T L, and Anderson, M G (1998) Potential Effects of Global Warming on Waterfowl Populations Breeding in the Northern Great Plains, *Clim. Change*, **40**, 343–369.
Toth, L A (1996) Restoring the Hydrogeomorphology of the Kissimmee River, in *River Channel Restoration: Guiding Principles for Sustainable Projects*, eds A Brookes and F D Shields, Jr, Wiley, Chichester, 369–383.
Vaselaar, R T (1997) Opening the Flood Gates: the 1996 Glen Canyon Dam Experiment, *Restor. Manage. Notes*, **15**, 119–125.
Welch, E B and Cooke, G D (1987) Lakes, in *Restoration Ecology: a Synthetic Approach to Ecological Research*, eds W R Jordan, III, M E Gilpin, and J D Aber, Cambridge University Press, New York, 109–129.
Williams, R J and Watford, F A (1997) Identification of Structures Restricting Tidal Flow in New South Wales, Australia, *Wetlands Ecol. Manage.*, **5**, 87–97.
Zedler, J B (1996a) Coastal Mitigation in Southern California: the need for a Regional Restoration Strategy, *Ecol. Appl.*, **6**, 84–93.
Zedler, J B (1996b) Ecological Function and Sustainability in Created Wetlands, in *Restoring Diversity: Strategies for Reintroduction of Endangered Plants*, eds D A Falk, C I Millar, and M Olwell, Island Press, Washington, DC, 331–342.

Rhizosphere

Rhizosphere is the volume of soil in which the microbial populations and the physical properties of the soil are influenced by the proximity of a plant root. It constitutes, in effect, the environment of the root. The major components of the rhizosphere are the soil and its nutrient minerals, soil water, and the microorganisms that live in the soil. The root and the rhizosphere, including all its components, exist in a dynamic reciprocal relationship. The size of the rhizosphere depends to some extent on its water content and the ease of diffusion of nutrients through the soil. It also depends on the availability of water for a specifically biological reason: plants react to drought stress by growing much larger and more extensive roots, which thereby increases the volume of the rhizosphere in drier areas.

The most important component of the rhizosphere is the microbiological population. The presence of roots leads to a substantial increase in the variety and a very great increase (as high as 100–2000 fold) in the numbers of bacteria. Plant roots are the major source of substrates and energy for the soil microflora, through sloughed off root cells (from root caps and root hairs), by the exudation of organic compounds (sugars, organic acids, amino acids) and by the production of mucilaginous gels that allow the root to pass easily through the soil. In addition, many bacteria or fungi may be closely associated with or attached to the root (i.e. on the *rhizoplane*, the root surface), and others may penetrate into the root where they live symbiotically (*see* **Symbiosis**, Volume 2). Pathogenic microorganisms are also found in far greater numbers in the rhizosphere.

Interactions may be even more complex. The rhizosphere microbial population may influence nutritional uptake by roots in various ways: reducing it by competition or by damaging roots, or stimulating it by stimulating uptake mechanisms or by acting to dissolve minerals that would otherwise be unavailable to the plant. Some rhizosphere organisms produce substances (e.g., gibberellins) that are powerful plant growth hormones, and so stimulate root growth. Others may cause a buildup of inhibitory substances: many fungi metabolize lignins in the rhizosphere litter, releasing phenolic acids which inhibit ion uptake by roots.

It is well known that certain rhizosphere and rhizoplane (and also symbiotic) microorganisms, both bacteria and fungi, fix nitrogen which greatly enhances the growth of plants. These microorganisms obtain the energy for their metabolism and nitrogen fixation from the plants, even in cases where there is no physical association of roots and microbes. Other beneficial consequences of rhizosphere bacteria and fungi are less well defined, but evidence has been presented that adding appropriate microbes much improves the survival of planted tree seedlings and results in stronger growth and increased production in some agricultural crops, particularly potatoes. The mycorrhizal fungi (symbiotically associated with tree roots) may transport nutrients to the roots and in turn be nourished by organic nutrients from the tree roots.

The rhizosphere is one of the most important aspects of the environment of plants. Its health and stability are essential to the successful growth and survival of plants.

R G S BIDWELL *Canada*

Riparian Zone/Littoral Zone

see Lakes and Rivers (Volume 2)

r- K Strategies

In theory, the life history strategy of a species is the result of eons of natural selection acting to optimize the allocation of energy within each individual among growth, maintenance, and reproduction. According to evolutionary theory, individuals of a species are expected to use energy for maintenance and growth in order to maximize their reproductive contribution to future generations (growth is important because fecundity and survival often increase with size).

The r- and K-strategies represent the extremes in a continuum of realized life history strategies. (The letters, r and K, correspond to the parameters of an equation used to describe population growth in very simple ecosystems). The *r-strategists* produce large numbers of seeds, eggs, or offspring, which have poor survival rates; they include many weeds and insects. In the life histories of r-strategists, natural selection has sacrificed longevity and survival in order to maximize reproduction. These species are opportunists, dispersing widely to invade ephemeral habitats quickly and exploit the relatively uncompetitive situations there while they last. The mate-finding tactics of r-strategists tend to be very efficient during the periods of very low population densities, which they often experience. Although many local populations may go extinct as their habitats change, species of r-strategists as a whole tend to be very resilient. They are typically the colonizing species of ecological succession and make dangerous biological invaders.

Compared to r-strategists, *K-strategists* tend to grow into bigger, longer lived, more competitive individuals, which produce fewer offspring but provide more care for them. K-strategists are the climax species of ecological succession and they tend to have more stable population sizes than those of r-strategists. Generally, populations of K-strategists are less resilient than those of r-strategists when reduced to low densities, and if substantially depressed throughout the entire range of the species, extinction becomes a real threat. Consequently, K-strategists rather than r-strategists generally merit the greatest concern of conservationists. According to the fossil record, individuals in many lines of animals have increased steadily in size as their populations have gone extinct. Natural selection fine-tuned these lines to become ever more extreme K-strategists, closely adapted to a specialized habitat, which eventually shrank and disappeared as its environment changed.

See also: **Life Cycles**, Volume 2; **Succession, Definition of**, Volume 2.

FURTHER READING

Pianka, E R (1970) On r- and K-selection, *Am. Naturalist*, **104**, 592–597.

Southwood, T R E (1976) Bionomic Strategies and Population Parameters, in *Theoretical Ecology: Principles and Applications*, eds R M May and W B Saunders, Philadelphia, PA.

RICHARD A FLEMING *Canada*

Rodin, Leonid E

(1907–1990)

Leonid E Rodin was a famous Russian ecologist, botanist, geographer and traveler. He graduated from the Leningrad State University, Department of Biology. Since 1933, until the end of his life, he worked at the Department of Geobotany of the Komarov Botanical Institute. From 1970 until 1980 he was head of this department. Earlier he was head of the Laboratory of Arid Zone Vegetation. Most of his life was devoted to study of the vegetative cover of Middle (Central) Asia, especially the deserts of Turkmenistan. He has written many papers on this problem (Rodin, 1955, 1956, 1958, 1961, 1963), and has prepared many maps of vegetation and pasture zones. From the early 1950s in collaboration with soil scientists, Rodin studied the process of biological turnover in desert biogeocoenosis and he made his first generalizations on this subject (Rodin and Bazilevich, 1968; Rodin *et al.*, 1968).

From the very beginning of the creation of the International Biological Program (IBP), L E Rodin was one of

the heads of the Soviet National Committee and a head of the Committee on Productivity of Terrestrial Communities. For many years he coordinated the IBP program in some interdisciplinary stations in Middle Asia and Kazakhstan and he was editor of numerous publications on the results of these investigations (Rodin, 1972, 1975, 1977, 1978). At that time Rodin worked at the interdisciplinary station Repetek in southeastern Kara-Kumes, where under his leadership the most complete complex of biological units and communities was studied. Moreover, Rodin took part in a geobotanical study of Algerian pastures. As a result, recommendations on the rational use of arid areas were given, geobotanical articles were written and vegetation maps were prepared (Rodin, 1970). The last large project of Rodin for the Program of Man and the Biosphere (MAB) of the United Nations Educational, Scientific and Cultural Organization (UNESCO), was the project on Phytocoenosis of the Main Ecosystems in the Badkhyz: Structure and Functioning (1976–1986) (Rodin, 1978). In the last years of his life Rodin became very interested in the problem of the influence of fire on desert vegetation. Unfortunately, a book that he was writing on the subject was not completed.

L E Rodin was an author of many popular-science books and botanical journeys around Brazil, India, and the Near East. He expounded ideas and works of the Russian geneticist and botanist–geographer N Vavilov. For many years L E Rodin was a main scientific secretary of the Russian Geographical Society.

REFERENCES

Rodin, L E (1955) Relief, Soils, Vegetation, Map of Pasture, in *Articles About the Nature of the Kara-Kumes*, (Ocherki prirody Kara-Kumov), Akad. Nauk SSSR, Moscow.

Rodin, L E (1956) *Vegetative Cover of USSR*, (Rastitelny pokrov SSSR.T.2.), Akad. Nauk SSSR, Moscow-Leningrad.

Rodin, L E (1958) *Middle Asia*, (Srednaya Asia), Akad. Nauk SSSR, Moscow.

Rodin, L E (1961) *Dynamics of Vegetation of Desert*' (Dinamika rastitelnosti pustyn), Akad. Nauk SSSR, Moscow-Leningrad, 1–227.

Rodin, L E (1963) *Vegetation of the Western Turkmenistan Deserts* (Rastitelnost, pustyn, Zapadnoi Turkmenii), Akad. Nauk SSSR, Moscow-Leningrad, 1–309.

Rodin, L E and Bazilevich, N I, eds (1968) *Production and Mineral Cycling in Terrestrial Vegetation*, Oliver and Boyd, Edinburg.

Rodin L E, Remezov, S U, and Bazilevich, N I (1968) *Instructional Methods of Study of the Dynamics and Biological Turnover in Phytocoenosis*, Nauka, Leningrad, 1–143.

Rodin, L E (1970) *Medea*, Algerie.

Rodin, L E (1972) *Eco Physiological Productivity in the Arid Zone*, International Symposium, Leningrad, USSR, 2.

Rodin, L E (1975) *Resources of the Biosphere*, Synthesis of the Soviet Studies for the International Biological Program, Leningrad, USSR.

Rodin, L E (1977) *Productivity of Vegetation in the Arid Zone of Asia*, Synthesis of Soviet Studies for the IBP, 1965–1974, Leningrad, USSR.

Rodin, L E (1978) *Improvement and Rational Use of Vegetation of the Arid Zone of Asia*, Synthesis of Soviet Studies for the IBP, 1965–1974, Leningrad, USSR.

FURTHER READING

Rodin, L E (1990) *Pistachio-tree in the Badkhyz*, Komarov Botanical Institute, Leningrad, USSR.

Rodin, L E and Bazilevich, N I (1965) *Dynamics of Organic Matter and Biological Turnover of Ash Elements and Nitrogen in the Main Types of World Vegetation*, Academy of Sciences of the USSR and Komarov Botanical Institute, Nauka, Moscow-Leningrad, 254.

BORIS K GANNIBAL *Russian Federation*

Rosswall, Per Thomas

(1941–)

Per Thomas Rosswall was born and educated in Sweden at the University of Stockholm and Uppsala. His research has focused on biogeochemical cycling, and has ranged from microbial ecology to global interactions. He is the author of over 100 papers and editor of more than 10 scientific books, as well as numerous articles in the popular press. Rosswall has had many national and international appointments and has held such major positions as secretary-general of the Scientific Committee on Problems of the Environment (SCOPE) (1986–1988), director of the International Global System for Analysis, Research and Training Secretariat (START) (1992–1993), and executive director of the International Geosphere–Biosphere Program (IGBP) (1987–1994). His leadership is characterized by a consultative and consensus style that is credited with fostering interdisciplinary collaboration among scientists from different countries, and forging effective international networks in both hemispheres.

Rosswall's first major research project was connected with the International Biological Program (1964–1974), in which he investigated carbon and nitrogen flows through tundra communities, and served for part of the time as

secretary to the International Steering Committee. Charged by the Swedish Natural Science Research Council with stimulating the field of microbial ecology in Sweden, he organized the first international symposium on microbial ecology in 1972, established a series of Nordic Summer Schools on the topic, and played a key role in the Swedish Coniferous Forest Research Project, in which his work on carbon and nitrogen cycling achieved international recognition. From 1978 to 1980 he was director of the SCOPE/United Nations (UN) Environment Programme International Nitrogen Unit, served on the executive committee of SCOPE for many years, and was its secretary-general from 1986 to 1988. From 1980 to 1983 he was secretary and chairman of the International Commission on Microbial Ecology. At the same time, he was instrumental in promoting microbiology in developing countries as member and secretary (1979–1983) of the UN European Office/UN Educational, Scientific and Cultural Organization/International Cell Research Office Panel on Microbiology.

Rosswall became well known for his participation in the Swedish University of Agricultural Sciences project *Ecology of Arable Land*. He contributed to the understanding of nitrogen cycling with his quantification of denitrification losses and the regulatory role of water and cropping systems on gaseous soil losses.

In 1984, Rosswall was appointed professor at the University of Linköping where his research focused on land and water use management in the tropics, and the effect of microbial processes on metal mobility in terrestrial ecosystems. In 1987 he became the first executive director of the IGBP. Under his leadership the program was defined and successfully initiated. Taking a temporary leave of absence from IGBP from 1992 to 1993, he accepted an appointment as the first director of the International START Secretariat in Washington, DC, where he began to organize regional networks to strengthen the abilities of developing countries to take part in international research collaboration dealing with global issues. Rosswall continues to play a crucial role in developing links among different scientific disciplines, between the natural and social sciences, and between scientists and policy makers. Since 1994 he has served as the rector of the Swedish University of Agricultural Sciences in Uppsala.

JAMES McCARTHY *USA*

S

Salinization

Terrestrial salinity refers to the presence of soluble salts, such as sodium, magnesium and potassium, in water and soil. It is measured in various ways, including parts per million (ppm) and electrical conductivity (EC) units (microSiemens cm^{-1}). Salinization is the accumulation of salts to levels that result in the degradation of the water and soil, either naturally or as a result of human activities. The salts come from precipitation, rocks and aeolian deposits. They are dissolved and mobilized by water, and brought to or near the surface by rising groundwater levels and capillary action. Salts then become concentrated as a result of the water movement, both over the ground surface and within the soil profile. Where water movement is impeded, evaporation leads to increasing concentrations, resulting in the salinization of the land. Over large areas, water containing salts, applied to the land naturally by precipitation or artificially by irrigation in excess of that needed by growing plants, infiltrates down to the watertable. In time, the watertable rises to or near the ground surface, producing the same salinization impacts. Salinization is particularly associated with irrigation in arid and semi-arid areas where there are high rates of evapotranspiration and inadequate or non-existent drainage. In particular, it can result from the excessive and wrongly timed application of water. In coastal districts, salinization can occur where seawater moves into aquifers as a result of over use of the groundwater. Dryland salinization occurs where deep-rooted native vegetation has been cleared and replaced by shallow-rooted crops and pastures that allow too much rainfall to pass into the groundwaters. Throughout agricultural history, salinization has resulted in reduced agricultural productivity and, ultimately, the sterilization of the land (for all but the most salt-tolerant plants). Its costs are enormous, from the collapse of early societies to the decline of many modern irrigation developments and the emerging problems of dryland salinization in parts of Australia, Canada and the United States.

FURTHER READING

Ghassemi, F, Jakeman, A J, and Nix, H A (1995) *Salinisation of Land and Water Resources*, University of New South Wales Press, Sydney, and CAB International, Wallingford.

Murray-Darling Basin Commission (1999) *The Salinity Audit of the Murray-Darling Basin: A 100-year Perspective*, Murray-Darling Basin Ministerial Council, Canberra.

PETER CRABB *Australia*

Sedimentary Records of Long-term Ecological Change

see Sedimentary Records of Long-term Ecological Change (*Opening essay*, **Volume 2**)

Seed Shadow

see Plant Dispersal and Migration (*Opening essay*, **Volume 2**)

Soil Mineralization

Soil mineralization is the process of the breakdown of organic material in soil and the conversion of organic complexes of nutrient chemical elements in the soil to inorganic

states. Organically bound sulfur, nitrogen, phosphorus, manganese, copper, iron and various trace elements may be mineralized, e.g., converted respectively to sulfides or sulfates, to ammonium, nitrite, nitrate or molecular nitrogen, to phosphates, and to various metallic ions and salts. In addition, organic carbon is mineralized to carbon dioxide.

Mineralization of soil often occurs following *humi cation*, in which organic detritus is converted into humus by microbial digestion or by earthworms. Mineralization takes place mostly by the action of soil bacteria (e.g., *Bacillus* and *Pseudomonas*) or fungi (e.g., *Aspergillus* and *Penecillium*). Mineralization goes on naturally in all soils, and is a normal aspect of the soil cycling of nutrients. Some nutrients must be mineralized before they can be absorbed by plants (e.g., nitrogen), while other nutrients are no longer available following mineralization because they cannot be absorbed (e.g., some phosphates), or become insoluble (e.g., iron). In addition, many mineralized nutrients, being soluble, are readily leached from the soil. Nutrients are also *demineralized* in the soil cycle by being incorporated into microorganisms during growth, and these are then unavailable for plant uptake. If the soil cycling process becomes unbalanced as a result of reduced input of organic material (e.g., by routine removal of crops) then the soil is in danger of becoming strongly or fully demineralized and its fertility, as well as its physical structure, may be seriously compromised.

While mineralization is a natural process, it can be harnessed for ecologically and agriculturally useful purposes in the practice of *composting*. In this process, waste organic materials are allowed to decompose under somewhat anaerobic conditions, which reduces nitrogen loss and oxidation of minerals to salts that may be unavailable to plants. The composting site is usually protected from rain to prevent leaching, and at least somewhat insulated so that the heat generated by biological activity will speed the process. The initial step is humification, and the product may be used as humus for gardens and fields. If composting is allowed to continue, mineralization occurs and a lower volume of richer fertilizer results.

R G S BIDWELL *Canada*

Soils and Sediments Biodiversity

see Biodiversity in Soils and Sediments: Potential Effects of Global Change (**Volume 2**)

Stable Isotopes

James R Ehleringer and Thure E Cerling
University of Utah, Salt Lake City, UT, USA

Stable isotopes *indicate those isotopes of an element which are stable and that do not decay through radioactive processes over time. Most elements consist of more than one stable isotope. For instance, the element carbon (C) exists as two stable isotopes,* ^{12}C *and* ^{13}C*, while the element oxygen (O) exists as three stable isotopes,* ^{16}O*,* ^{17}O*, and* ^{18}O*. As is the case with both carbon and oxygen, the more common stable isotope species typically contains the fewest number of neutrons for that element. Stable isotopes should not be confused with* radioactive isotopes *of an element. Radioactive isotopes have limited life times and undergo a decay to form a different element, although the time required for this decay may vary widely ranging from nanoseconds to thousands of years. For instance, carbon has six radioactive isotopes (*^{9}C*,* ^{10}C*,* ^{11}C*,* ^{14}C*,* ^{15}C*, and* ^{16}C*) of which* ^{14}C *is perhaps best known because of its utility in dating biological materials and as a tracer in metabolic studies.*

HOW COMMON ARE STABLE ISOTOPES?

A brief listing of the stable isotopes and their abundances for the elements most commonly used in global change research can be seen in Table 1.

HOW ARE STABLE ISOTOPE ABUNDANCES TYPICALLY MEASURED?

Of particular interest for global change studies are variations in the isotopic abundances of hydrogen, carbon, nitrogen, and oxygen isotopes. These relatively light elements are typically measured using a gas isotope rationing mass spectrometer. The mass spectrometer consists of a source to ionize the gas, a ight tube with a magnet to de ect the path of the ionized gas, and a detector system at the end of the ight tube to measure the different isotopic species. First, the element of interest must be converted to a gaseous form for introduction into the mass spectrometer. The most commonly used approaches involve introducing hydrogen as H_2, carbon as CO_2, nitrogen as N_2, and oxygen as CO_2. As the gas is introduced into the mass spectrometer, it is ionized by removal of an electron as the gas is bombarded by a source. Then as the ionized gas travels down the ight tube (under vacuum), the paths of light and heavy isotopic species are de ected differently by the magnet. Detectors are positioned at the end of the ight tube to

Table 1 A brief listing of the stable isotopes and their abundances for the elements most commonly used in global change research

Element	Isotope	Abundance (%)
Hydrogen	^{1}H	99.985
	^{2}H	0.015
Carbon	^{12}C	98.89
	^{13}C	1.11
Nitrogen	^{14}N	99.63
	^{15}N	0.37
Oxygen	^{16}O	99.759
	^{17}O	0.037
	^{18}O	0.204
Sulfur	^{32}S	95.00
	^{33}S	0.76
	^{34}S	4.22
	^{36}S	0.014
Strontium	^{84}Sr	0.56
	^{86}Sr	9.86
	^{87}Sr	7.02
	^{88}Sr	82.56

measure the abundance ratios of the heavy and light isotopic species.

EXPRESSION OF STABLE ISOTOPE ABUNDANCES

The abundance of stable isotopes is typically presented in delta notation (δ), in which the stable isotope abundance is expressed relative to a standard (Equation 1):

$$\delta = \left(\frac{R_{\text{sample}}}{R_{\text{standard}}} - 1 \right) \times 1000 \qquad (1)$$

where R is the molar ratio of the heavy to light isotopes, e.g., Equation (2):

$$R = \frac{^{13}C}{^{12}C} \text{ or } \frac{D}{H} \text{ or } \frac{^{18}C}{^{16}O} \qquad (2)$$

For the most commonly used light isotopes, the internationally recognized standards are shown in Table 2.

Table 2 Internationally recognized standards for the most commonly used light isotopes

H	SMOW[a]	$R = 0.0001558$
C	PDB[b]	$R = 0.0112372$
N	AIR[c]	$R = 0.0036765$
O	SMOW	$R = 0.0020052$
S	CDT[d]	$R = 0.0450045$

[a] Standard Mean Ocean Water (SMOW).
[b] Pee Dee Belemnite (PDB).
[c] Atmospheric air (AIR).
[d] Canyon Diablo Triolite (CDT).

HOW DOES NATURAL VARIATION IN STABLE ISOTOPES COME ABOUT?

Variations in the abundances of stable isotopes among different compounds arise because the chemical bonding is stronger in molecules containing heavier isotopic forms, making it more difficult to break up the molecule in a chemical reaction (often termed *kinetic fractionation*), or because of differences in the physical properties of molecules containing heavier isotopic forms (often termed *diffusive and equilibrium fractionation*). With kinetic fractionation, the rate of an enzymatic reaction is faster with substrates that contain the lighter isotopic form than in reactions involving the heavier isotopic form. As a consequence, there will be differences in the abundances of the stable isotopes between substrate and product. Such differences will occur unless, of course, all of the substrate were consumed, in which case there would be no difference in the isotopic composition of substrate and product. Expression of a significant kinetic fractionation in most biological reactions involves substrates at branch points in metabolism, such as the initial fixation of carbon dioxide in photosynthesis. Equilibrium fractionation events re ect the observation that during equilibrium reactions, such as the equilibration of liquid and gaseous water, molecules with the heavier isotopic species are typically more abundant in the lower energy state phase. Diffusive fractionation events re ect the observation that heavier isotopic forms diffuse more slowly than lighter isotopic forms.

WHAT IS THE NATURAL RANGE OF ISOTOPIC VARIATION IN NATURE?

The natural variations in isotopic abundance can be large. Figure 1 illustrates the natural variations found for materials frequently of interest in global change studies: waters, greenhouse gases, and biological materials. As a starting point, note that some atmospheric gases, such as CO_2, N_2, and O_2, exhibit limited variation, while N_2O and CH_4 exhibit wide isotopic variation. The larger isotopic ranges in the latter two gases re ect both significant isotopic fractionation by microbes as well as different biological substrates which are used to produce these gases.

Oceans, the largest volume of water on Earth, exhibit only small variation in isotopic abundance and most of this is associated with changes in salinity. However, once water evaporates from the oceans and later recondenses as precipitation, there are large isotopic variations that are dependent on both cloud temperature and the amount of residual moisture remaining within the cloud mass. Lakes and rivers re ect precipitation input values, but are often further enriched by evaporative processes, which favor the movement of lighter isotopic forms of water into the vapor phase.

There can be significant biological fractionation against carbon dioxide during photosynthesis, which results in

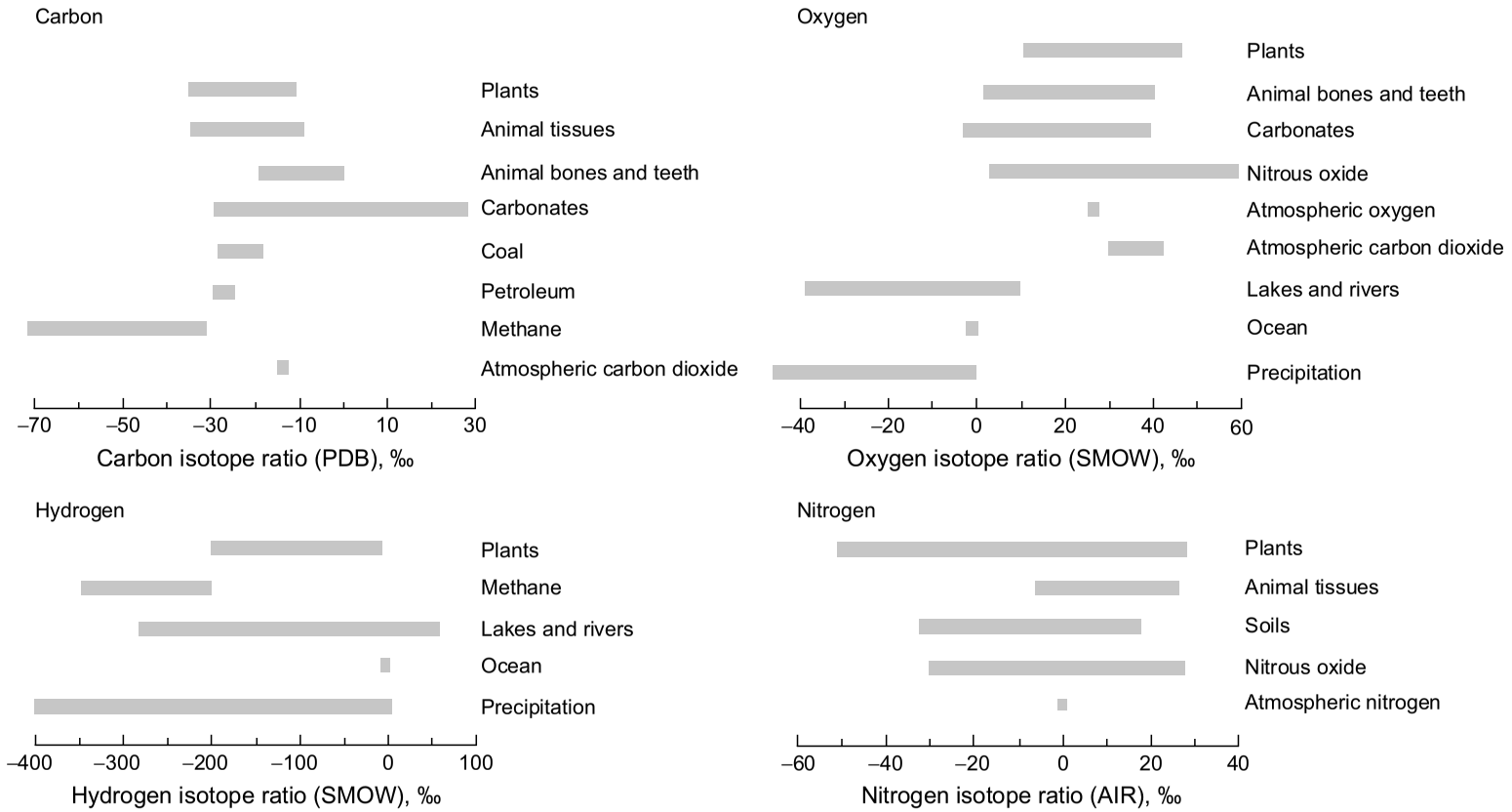


Figure 1 Variations in the carbon, oxygen, hydrogen, and nitrogen isotope ratios of materials of interest in global environmental change studies. (Data were collected from a very wide ranging series of studies)

plants being isotopically depleted relative to the carbon dioxide substrate. Animals largely re ect the carbon isotopic composition of their food resources, a useful tool for food web studies. However, the carbon in animal bones and teeth consists of carbonates, which are isotopically enriched relatively to the original carbon dioxide substrate (in this case respired food resources or animal tissues). In general, carbonates are both ^{13}C and ^{18}O enriched, relative to their substrates, carbon dioxide and water.

There is very limited isotopic variation in diatomic nitrogen. The microbial process of nitrogen fixation ($N_2 \rightarrow NH_4^+$) exhibits little isotopic fractionation and thus these products have a similar nitrogen isotope ratio as the atmosphere. Yet subsequent nitrogen transformation reactions exhibit strong isotopic fractionations. In general, the isotopic composition of animals within a trophic level are enriched by ~ 3 relative to their food substrate. The wide variations in nitrogen dioxide re ect the large differences associated with aerobic versus anaerobic processes and industrial versus stratospheric processes.

HOW ARE STABLE ISOTOPES USED IN GLOBAL CHANGE RESEARCH?

Stable isotopes are useful for several important aspects of global change research. One major area is to use stable isotope ratios to reconstruct past environmental conditions, which is possible when a compound is laid down and preserved over time. Some common examples would include ice profiles, the shells of mollusks, and tree rings. Another major application of stable isotopes is to reconstruct diets of animals and ecosystem structure. The isotopic composition of an animal re ects its food source; while soft tissues and hair are lost after animals die, the bones and teeth can be used to reconstruct past diets. Similarly for ecosystems, the isotopic composition of soil organic matter re ects the plant material that occupied the surface in the past. Another very common application of stable isotopes is as a tracer to identify where materials had originated. This is a particularly powerful approach in atmospheric studies where the interest is in identifying which terrestrial surface (e.g., ocean versus land) or which terrestrial ecosystem (e.g., grassland versus forest) is contributing or removing a gas species from the atmosphere.

Below we provide several examples to illustrate how stable isotopes are used in global change research. However, because of space limitations, we will only consider examples involving hydrogen, carbon, and oxygen isotopes.

Reconstruction of Past Air Temperatures using Hydrogen Isotopes of Water in Ice Cores

The hydrogen and oxygen isotope ratios of water have been used to reconstruct temperature patterns. The basis for this observation is that the isotopic fractionation associated with cloud vapor to precipitation formation is temperature dependent. Scientists have capitalized on this observation by analyzing deep ice profiles, such as the ice in Antarctica. Figure 2 shows the initial depth profile of hydrogen isotope ratios of ice water found in the Vostok ice core. The data shown reveal the isotopic oscillations observed during a glacial–interglacial cycle. When these isotopic values are converted to equivalent temperature differences, we see that the temperature at this location varied by $10\,^{\circ}C$ between glacial and interglacial periods. Of particular importance also are the smaller temperature cycles, which re ect oscillations in climate. Note also how rapid the change is from one temperature regime to another. The bottom of the Vostok ice core extends back to 420 000 years before present. By analyzing the ice records from different locations, scientists are able to reconstruct the spatial patterns of temperature changes and the extent to which climate oscillations were of regional versus global extent.

Carbon Isotopes in Fossils and Fossil Soils are used to Reconstruct Past Terrestrial Ecosystems

The carbon isotope difference between C_3 and C_4 vegetation can be exploited to determine ecological change in the past. C_3 and C_4 plants (*see* C_3 and C_4 Photosynthesis, Volume 2) have $\delta^{13}C$ values averaging approximately -27 and -12 , respectively. The ratio of C_3 to C_4 plants in the ecosystem or in the diet of herbivores is recorded in fossil soils as organic carbon or as soil carbonate, and in mammalian tooth enamel. Organic matter preserved in soils has about the same $\delta^{13}C$ value as in the original soil, whereas soil carbonate and mammalian tooth enamel are both enriched by about 14 compared to the vegetation.

The study of fossil soils and fossil tooth enamel from the Siwalik sediments in Pakistan illustrates how isotopes can be used to reconstruct past ecosystems. In Figure 3,

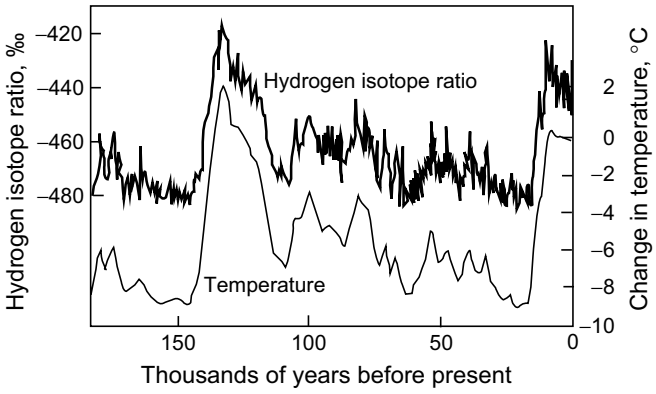


Figure 2 The hydrogen isotope ratios of waters from segments of ice from the Vostok ice core in Antarctica and the calculated temperatures associated with those time periods for the past 200 000 years. (Adapted from Jouzel *et al.*, 1990)

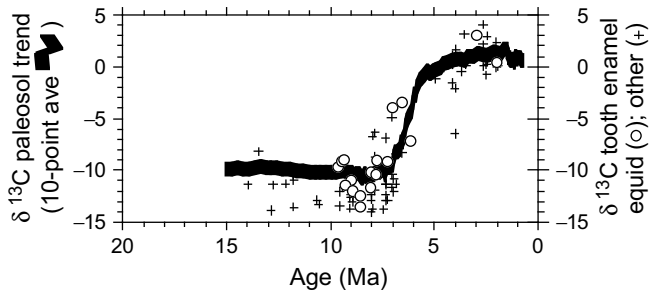


Figure 3 $\delta^{13}C$ values for soil carbonates from the Siwaliks in Pakistan shown as a 10 point running mean. $\delta^{13}C$ values for equids show that they had a C_3-dominated diet before 8 million years ago, but changed to a C_4 diet as the ecosystem underwent a change from C_3 dominated to C_4 dominated. (Adapted from Cerling *et al.*, 1998)

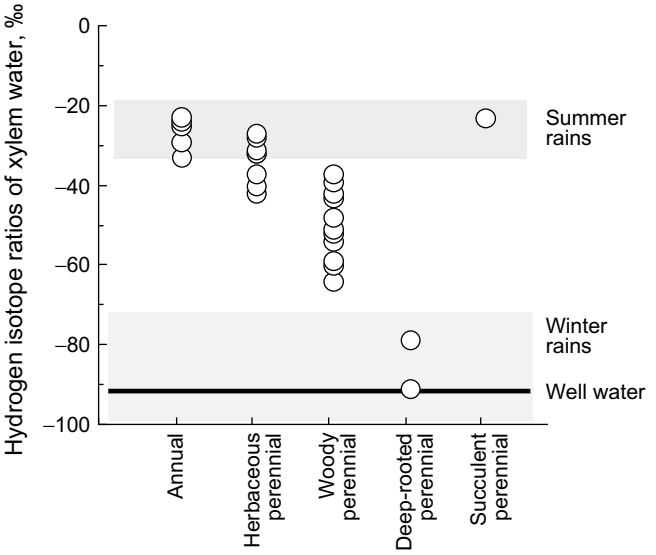


Figure 4 Hydrogen isotope ratios of water found in plants growing in the same arid land location in southern Utah following an intense summer rain event. Shown also are the observed variations in the isotopic composition of summer rains (upper depth soil moisture source) and winter rains (lower depth soil moisture sources). (Adapted from Ehleringer *et al.*, 1991)

the fossil soil record shows that before eight million years ago, the local ecosystem was dominated by C_3 plants, and that they were gradually replaced by C_4 plants so that by five million years ago the ecosystem was dominated by C_4 plants. Fossil tooth enamel from the same region shows that equids and other mammals had a C_3-dominated diet before eight million years ago, but changed to a C_4-dominated diet as the ecosystem changed to become C_4-dominated. This change in vegetation is accompanied by a significant change in the $\delta^{18}O$ of local water of atmospheric origin indicating that the monsoon system may have undergone a significant change at this time. Recent studies have also shown that this was a time of major faunal turnover in the region (Cerling *et al.*, 1998).

Hydrogen and oxygen isotope ratios of water are used to determine current sources of plant water from within the soil, while those isotopes in cellulose of tree rings provide historical patterns. For plants, the primary sources of hydrogen and oxygen are from waters of atmospheric origins. Since the isotopic composition of precipitation will vary both seasonally and geographically, largely re ecting cloud condensation temperature, we have a tool for addressing ecological and global change questions. In plants, the isotopic composition of tree rings will in part re ect the isotopic composition of plant water sources, providing a tool for now understanding how precipitation patterns

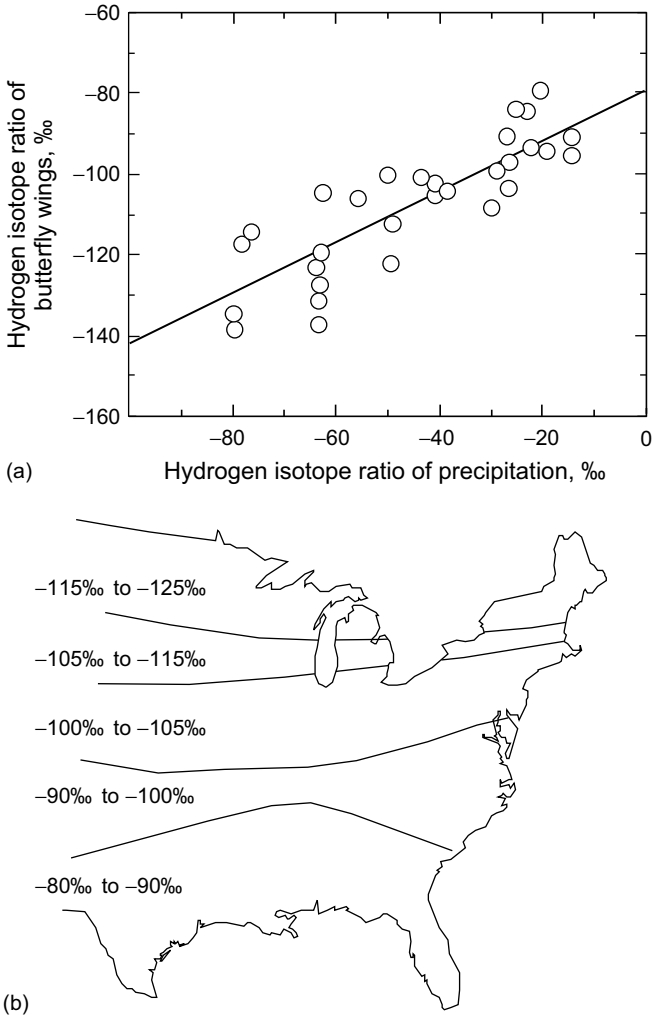


Figure 5 (a) The relationship between hydrogen isotope ratios of wings of field-raised monarch butterflies and of weighted mean growing-season precipitation. (b) Geographic patterns of hydrogen isotope ratios of precipitation across the eastern US. (Adapted from Hobson *et al.*, 1999)

as re ected in tree rings may have changed during a plant's lifetime. For long-lived organisms, such as redwood trees, this provides a valuable tool for reconstructing climatic variations over periods of hundreds to thousands of years.

A second area in which isotopes of water are used in plant global change studies is for understanding the dynamics of water use by different species within a community, particularly for evaluating which species may or may not be competing for water. Figure 4 illustrates the hydrogen isotope ratios of different species at an arid site in western North America following intensive summer rains, which saturate the upper soil layers while the remaining winter moisture exists only in deeper soil layers. Shown also are isotopic ratios expected for rains falling in summer versus winter months. Note that species utilize different moisture sources, suggesting a strong sensitivity of each species to whether rain falls in winter or summer. Strong interannual variations in precipitation are associated with El Nino events and this will in uence the growth and competitive dynamics of plants in those regions of the globe impacted by El Nino Southern Oscillation (ENSO). While analyses of water within stems of modern plants can be used to immediately assess the impact of ENSO on plant performance, tree ring analyses can be used to reconstruct the periodicity and extent of the impact over historical periods.

Reconstruction of Animal Migration Routes using Hydrogen and Oxygen Isotope Ratios of Feathers

In global change studies, it is often useful to know if animals have changed their migration routes, re ecting some change in the quality of an environment. Or for animals that congregate in a common winter location, such as the monarch butter ies that winter in central Mexico, where did these individuals originate? The isotopic composition of butter y wings and of bird feathers have been shown to re ect the isotopic composition of water at the location where that tissue was initially formed. Thus, isotopes provide a powerful tool for identifying where an individual might have originated from. In the case of the monarch butter y, Figure 5(a) illustrates the strong correlation between the hydrogen isotope ratios of precipitation at a location and the subsequent hydrogen isotope ratios observed for organic matter in the butter y wings. Since virtually all of the hydrogen in organisms (both plants and animals) originates from precipitation, the strong correlation is expected. The offset from a 1 : 1 relationship re ects biological fractionation processes during synthesis of the wing. The basis for geographical variations in isotopic composition of precipitation re ects temperature differences in the clouds at the time of precipitation. Colder storms are found at higher latitudes. As shown on Figure 5(b), there is a variation of ~45% in hydrogen isotopes across the eastern portions of the US allowing sufficient variation for scientists to identify regions from which a monarch butter y might have come.

Dissecting the Global Carbon Cycle Today with Carbon and Oxygen Isotopes in Carbon Dioxide

The carbon isotope ratios of fossil fuels are depleted in ^{13}C relative to the atmosphere (Figure 1). Since the increases in atmospheric CO_2 are largely associated with fossil fuel emissions, there has been a long-term trend of progressive decreases in the average carbon isotope ratio of atmospheric carbon dioxide from −6.5 to −8 today.

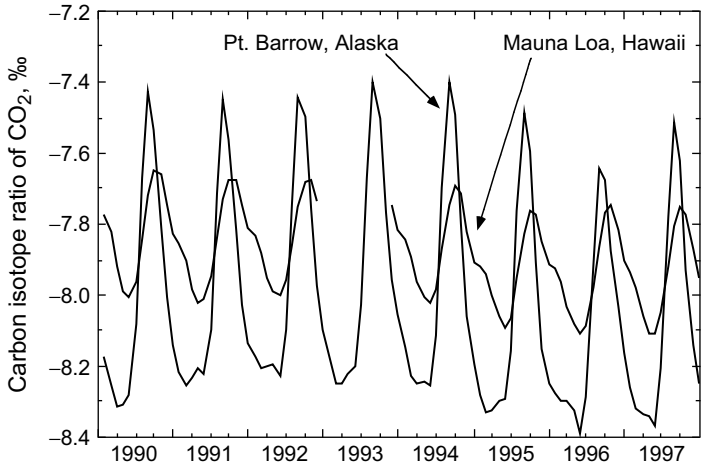


Figure 6 Seasonal changes in the carbon isotope ratios of atmospheric CO_2 at Point Barrow (Alaska) and Mauna Loa (Hawaii) between 1990 and 1997. The seasonal changes are driven by differences in the relative rates of photosynthesis and respiration by ecosystems at these latitudes. The enhanced amplitude at northerly latitudes reflects the shorter growing season. (Adapted from the National Oceanic and Atmospheric Administration – Climate Monitoring and Diagnostics Laboratory Monitoring Program; data courtesy of Jim White and Pieter Tans)

On an annual basis, variations in the carbon isotope ratio analyses of atmospheric carbon dioxide are used to understand seasonal dynamics on the land surfaces. In Figure 6, the monthly average carbon isotope ratio, for atmospheric carbon dioxide at Point Barrow, Alaska (latitude 71 °N) and Mauna Loa, Hawaii (latitude 19 °N) are shown for the period 1990–1997. The strong seasonal cycles re ect the addition of carbon dioxide during periods when ecosystem respiration exceeds photosynthesis, resulting in a net loss of carbon dioxide from the ecosystem. In contrast, during periods when photosynthetic activity is high, the carbon isotope ratio of the atmosphere increases as photosynthetic processes discriminate against the heavier $^{13}CO_2$. By analyzing these time trends, it is possible to detect changes in the length of the growing season over time and to address carbon cycle questions, such as the extent to which terrestrial versus oceanic surfaces are responsible for absorbing anthropogenically produced carbon dioxide.

REFERENCES

Cerling, T E, Harris, J M, MacFadden, B J, Leakey, M G, Quade, J, Eisenmann, V, and Ehleringer, J R (1998) Miocene/Pliocene Shift: One Step or Several? Reply, *Nature*, 393, 127.

Ehleringer, J R, Phillips, S L, Schuster, W F S, and Sandquist, D R (1991) Differential Utilization of Summer Rains by Desert Plants: Implications for Competition and Climate Change, *Oecologia*, 88, 430–434.

Hobson, K A, Wassenaar, L I, and Taylor, O R (1999) Stable isotopes (δD and δ^{13}C) are geographic Indicators of Natal Origins of Monarch Butter ies in eastern North America, *Oecologia*, 120, 397–404.

Jouzel, J, Petit, J R, and Raynaud, D (1990) Palaeoclimatic Information from Ice Cores: The Vostok Records, *Trans. R. Soc. Edinb.: Earth Sci.*, 81, 349–355.

FURTHER READING

Cerling, T E, Harris, J M, MacFadden, B J, Leakey, M G, Quade, J, Eisenmann, V, and Ehleringer, J R (1997) Global Vegetation Change Through the Miocene/Pliocene Boundary, *Nature*, 389, 153–158.

START (Global Change System for Analysis, Research and Training)

see IGBP Core Projects (*Opening essay*, **Volume 2**)

State Variable

see Non-linear Systems (**Volume 2**)

Strange Attractor

see Chaos and Predictability (**Volume 1**); Chaos and Cycles (**Volume 2**); Non-linear Systems (**Volume 2**); Monitoring in Support of Policy: an Adaptive Ecosystem Approach (*Opening essay*, **Volume 4**)

Succession, Definition of

The assemblage of species in a given location changes continuously. One cause of this change is disturbance, which ranges from local events such as the creation of a gap in the forest canopy by a fallen tree, to events with broader and more severe impacts such as large forest fires. By removing biomass and opening space, each disturbance provides an opportunity for new species to colonize that location. Often the most successful colonizing species quickly dominate the disturbed location, but after a while they are succeeded by other species. This succession occurs because the most abundant of the early colonizers are too short lived to maintain their dominance or because they alter the local environment so that it begins to favor even newer species. In the altered environment, these newer (latter successional) species eventually exclude the colonizers by out-competing them for resources. These newer species may themselves be succeeded by even newer species for the same reasons. Thus, succession can produce consistent patterns of changing species composition and species replacement after disturbance.

Two types of succession are widely recognized. *Primary succession* occurs in areas where the soils initially support no life (e.g., new lava ows). *Secondary succession* is much more common; it occurs where the soils initially contain remnants of previous species assemblages (e.g., fire resistant seeds).

At least three stages of succession are commonly distinguished. In the *colonizing stage*, ecosystems typically have few species and short food chains. In this stage, ecosystems are very productive (accumulating organic matter much faster than they break it down) but they are unstable. In the *mature stage*, ecosystems tend to be more complex, more diversified, and more stable. Many ecosystems in the final (*climax*) stage of succession are quite stable and exhibit

large species diversity and complex food webs. Climax ecosystems tend to be unproductive because almost all captured solar energy is needed to maintain the large amounts of organic matter already accumulated, leaving little for the production of more organic matter.

See also: Climax Vegetation, Volume 2; Disturbance, Volume 2; Succession, Volume 2.

FURTHER READING

Moore, P D (2001) A Never-ending Story, *Nature*, 409, 565.
West, D C, Shugart, H H, and Botkin, D B, eds (1981) *Forest Succession*, Springer-Verlag, New York.

RICHARD A FLEMING *Canada*

Succession

Virginia Dale
Oak Ridge National Laboratory, Oak Ridge, TN, USA

The process of succession refers to the gradual replacement of one species population by another that occurs over time after a site has been disturbed. The concept can be extended to the replacement of one kind of community by another kind, the progressive changes in vegetation and in animal life, which may culminate in dominance by a community that is stable until the next disturbance. Succession refers to changes that occur over 1- 500 years and not to seasonal changes in populations.

*A typical successional sequence, or sere, is initiated by a disturbance that removes much or all of the vegetation. The process is called primary succession if it begins on an entirely denuded site such as in the aftermath of a lava ow or glacier retreat. Studies of succession in Glacier Bay, Alaska US, show that a newly exposed area is rst occupied by mosses and lichens, and then by dwarf willows (*Salix *spp). After about 50 years a dense thicket of alders (*Alnus crispa*) occupies the area, and this community is gradually replaced by Sitka spruce (*Picea sitchensis*) and hemlock (*Tsuga heterophylla *and* T mertensiana*) to form a taiga community after about 200 years. This pattern of changes in species over time has been documented in new volcanic soils in such diverse locations as New Zealand, Iceland, Indonesia, and Mexico.*

Secondary succession occurs when remnants of the prior community remain. For example, old- eld succession in the piedmont region of North Carolina, US, occurs after agricultural activities are abandoned. The site is rst occupied by annual and then perennial herbaceous species. Pine seedlings that mature and eventually die out as they succeed to oaks and hickories gradually replace this herbaceous community. Wildlife composition and number of species also change with each successional stage. The most apparent changes in bird species occur as new life forms become established (e.g., the herbaceous community is replaced by pines and later by hardwoods), for birds relate more closely to the structure of the vegetation than to the species. No one plant or bird species is dominant over the entire sere; instead each species thrives best at some point in the successional period.

The original paradigm that successional communities eventually mature to a persistent climax state has been replaced by the realization that disturbances are characteristic of almost all ecosystems, and thus a nal persistent endpoint is rarely achieved. Furthermore, the distinction between primary and secondary succession is now recognized as being too simplistic to capture the true complexity of the variability in successional patterns and processes. Instead, primary and secondary succession are treated as extremes along a continuum of abundance of biotic and abiotic residuals left by a disturbance and a continuum of soil development.

A HISTORY OF THE CONCEPT OF SUCCESSION

McIntosh (1999) provides a thorough overview of the developing concepts about succession. Thoreau coined the term succession in 1860. However it was not used again until Cowles (1899) revived the concept and emphasized its dynamic nature, sought to describe patterns of succession, and produced a physiographic theory based on his classic studies of the Lake Michigan sand dunes. He emphasized the variability in the process and noted that it was not a step-by-step linear process to one particular endpoint.

Clements (1916), in contrast, described succession as an orderly, predictable, progressive and convergent phenomenon and even assigned superorganism properties to it. Although the concept of the superorganism was rejected fairly quickly, the orderly and predictable nature of succession was accepted by a host of ecologists. Gleason's (1917) recognition of the individualistic concept of succession as a mobile and unfixed process ew in the face of these ideas but were largely ignored or denied.

Nevertheless, Cowles' ideas were revived by Whittaker's (1953) acknowledgment of succession as involving highly variable and irregular changes in populations over time. Olson (1958) revisited the Michigan dune sites using new techniques and observations to quantify Cowles' focus on the variable nature of succession. Subsequently, Odum (1969) presented a list of trends to be expected under

Table 1 Trends expected during succession (based on Odum, 1983 and Bazzaz, 1979)

Species and community structure
Species composition changes
The number of species increases
The distribution of species becomes more even
r-strategists replaced by K-strategists
Life cycles increase in length and complexity
Size of organisms and propagules increases
Mutualistic symbiosis increases
Physiological characteristics of dominant species
Seed dispersal distance declines
Seed germination less enhanced by high light levels, fluctuating temperatures, and high NO_3
Seed germination less inhibited by far-red light
Light saturation intensity declines
Light compensation lower
More efficient at low light
Lower photosynthetic rates
Lower respiration rates
Lower transpiration rates
Higher stomatal and mesophyll resistances
Higher resistance to water transport
Slower recovery from resource limitation
Energetics
Biomass and organic detritus increase
Gross primary production increases; secondary production changes little
Net production decreases
Respiration increases
The ratio between gross production and respiration moves toward one
The ratio between biomass and gross production increases
Nutrient cycling
Element cycles increasingly closed
Turnover time and storage of essential elements increase
Cycling ratio increases
Nutrient retention and conservation increase
Stability
Resistance increases
Resilience decreases
Overall strategy
Efficiency of energy and nutrient use increases

succession which are still debated by some but endorsed by others (a subset of Table 1).

The use of mathematic analysis and computers has led to a new era of successional studies. Many computer models are now available that build upon the transition model approach of Horn (1975) and the JABOWA model developed by Botkin *et al.* (1972) that uses life history characteristics of individual tree species to simulate establishment, growth and survival in tree gaps. Yet models do not exist for succession in areas with complex spatial heterogeneity. The advantage of the modeling format is that it quantifies theories of succession and makes testable hypotheses more apparent.

Today, analysis of succession often combines field and modeling components. Together these approaches provide documentation of changes and a way to formalize the concepts and extrapolate them over the long time over which succession typically occurs. Succession remains as a key concept in ecology, yet its complexity still defeats scientists' desire for order and simplicity.

CHARACTERISTICS OF SUCCESSION

The causes of succession are more realistically viewed as a continuum ranging from changing external geophysical forces (*allogenic* succession) to biological processes modifying conditions and resources (*autogenic* succession). Under allogenic succession, physical forces override biological causes. For example, a regional drying trend over the past 100 years in the Southwestern US may be partially responsible for the replacement of desert grassland with desert scrub. Accelerated cycles of erosion and deposition during the Dust Bowl in the central US changed vegetation patterns. Alterations in the meandering of rivers over time change riparian vegetation and create new sites for succession. Autogenic succession occurs as a result of the organisms' effect on their environment. Plants can create shade, add to the litter, moderate temperature oscillations, increase humidity, and alter soil structure and chemistry. Autogenic succession is the focus of this article.

Succession occurs within a physically changing environment as changes in biological composition, structure and function over time affect ecological processes and *vice versa*. As global environmental changes occur, they create a new backdrop for ecological succession. Odum (1969) hypothesized that the major traits changing over a sere include biomass, net primary production, plant-species diversity, species life-history traits, propagule-dispersal modes, propagule longevity, and nutrient storage (Table 1). Many later studies have documented that these traits do change over successional time; however, less clear are when these traits change and how changes in the spatial heterogeneity of the physical and biological conditions in uence the successional process.

The species dominating early and late successional communities typically have a common set of characteristics. Early successional communities are characterized by species with light, wind-dispersed seeds that germinate and grow rapidly under high light conditions, whereas late successional communities are dominated by species with heavy seeds that germinate and grow more readily under low light conditions. These groups of species are referred to as *r-strategists* and *K-strategists*, respectively, depending on which part of the logistic growth curve (Equation 1) is

most responsible for the increase in population size:

$$\frac{dN(t)}{dt} = \frac{rN(t)[K - N(t)]}{K}$$

or

$$N(t) = \frac{K}{1 + e^{-rt}} \quad (1)$$

where r is the rate of growth of the population, K is the asymptote of the S-shaped curve, and $N(t)$ is the population size at time t. Thus, dN/dt is the rate of population growth. In uncrowded environments, selection favors species with high rates of reproductive potential, and thus r-strategists are more common. In contrast, crowded environments favor species with lower growth potential but better abilities to compete (K-strategists) (*see* r–K Strategies, Volume 2).

Life history traits of surviving species relate to resistance to or tolerance of disturbance and influence the species composition of the biotic residuals. For example, aquatic plants tend to survive floods by growing upward into the light zone as waters rise. After the 1993 midwestern floods in the US, a rare species of false aster (*Boltonia decurrens*) became abundant, for it requires fresh mud flats for seed germination (Smith *et al.*, 1998). Furthermore, surviving seeds can greatly influence reestablishment as evidenced by the dramatic recruitment that occurs from buried seed after a fire. Similarly, seed survival in serotinous cones or fruits permits rapid establishment after a fire. The nitrogen-fixing capability of lupines that survived the 1980 eruption of Mount St Helens, US, ameliorated soil nitrogen conditions, and thus facilitated establishment by other plants. Where deposition of volcanic ash from Mount St Helens is only a few decimeters, surviving pocket gophers (*Thomomys talpoides*) mix the original soil into the ash, and thus both expose buried seeds ripe for germination and improve nutrient conditions of the new ash (Anderson and MacMahon, 1985).

Early stages of succession are influenced largely by substrate characteristics (soil texture, moisture, and nutrient condition), distance to a seed source, and seed morphology, whereas later plant establishment is largely determined by environmental changes caused by earlier immigrants. *Facilitation* is the process of early successional species altering conditions or the availability of resources in a habitat in a way that favors later successional species. For example, the first species to become established create shade, alter the soil moisture properties, and ameliorate soil texture and nutrient conditions via decomposition of their parts. Often nitrogen is a limiting factor in the early stages, and the presence of plants with the ability to fix nitrogen through association with nitrogen-fixing bacteria enhances soil development. In the Pacific Northwest of the US, the early successional tree, alder (*Alnus* spp) is a key facilitating species because it (1) fixes nitrogen; (2) has light, wind-dispersed seeds that germinate readily; (3) grows rapidly; and (4) reproduces at an age as young as three years.

In contrast to facilitation, the process of *inhibition* may actually arrest successional development. This process occurs when a resource is so intensely used by a species that it is not available to other species. Space is the resource that is typically constrained. For example, following a mudslide off a steep slope of Mount Rainier in Washington State, US, the Kautz Creek area was quickly inundated by an almost continuous mat of mosses and lichens. Germinating tree seedlings could not penetrate the mat to reach the mineral soil, thus tree establishment was inhibited for decades.

Both facilitation and inhibition can be important in any one successional sequence. For example, early successional species typically alter the physical environment to make it more suitable for other species. But as these plants grow, they may inhibit other species by shading them and competing with them for moisture.

Life history characteristics can thus influence the time over which succession occurs. The early establishment of a particular species can inhibit some species and thus shorten the time to dominance by a late-successional species. In the Northeastern US, abundant establishment of cherry (*Prunus pensylvanica*) inhibits the growth of other mid-successional species. On the other hand, the peak of forb and graminoid seedling recruitment three years after the 1988 fires in Yellowstone National Park and the scarcity of conifer seedlings may have lengthened the time required for forest development or may have precluded it entirely. Thus, ecosystem structure, composition, and processes may reflect for centuries the presence or absence of residuals following a disturbance.

SUCCESSION AS A CYCLE

Succession can be viewed as a cycle that moves the ecological system between the functions of exploitation and conservation (Figure 1) (Gunderson and Holling, 2000). In early stages of succession, *exploitative* processes dominate and lead to an organization of the resources by means of binding of nutrients, accumulation of biomass and modification of the environment. Later, *conservative* forces, such as competition, dominate and increase organization via trophic and competitive connections. Forces of changes also affect the ecological system: organisms die and are replaced, thus releasing previously bound resources. Holling (1986) refers to this stage as *creative destruction*, for it contains both destructive features and creative ones. The final stage is one of *renewal*. In this stage, the system has carbon, nutrients and energy available for use by component organisms.

The cycle is presented as a pattern in time, but it also reflects patterns in space (Holling, 1986). The more rapidly a system moves through the cycle, the more likely it is to create a patchy system. Whereas those systems that move slowly through the cycle tend to be uniform in space as a

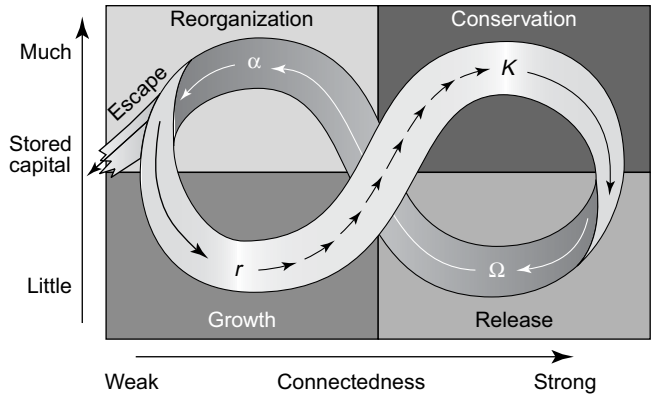


Figure 1 The dynamics of a system as it is dominated by each of the four processes: rapid growth (r), conservation (K), release (Ω), and reorganization (α). The arrows indicate the speed of the cycle. The short, closely spaced arrows indicate a slow change, while the long arrows indicate rapid change. The cycle reflects systemic change in the amount of accumulated capital (nutrients, resources) stored by the dominant structuring process in each phase, and the degree of connectedness within the system. The exit from the cycle at the left of the figure indicates the time at which a systemic reorganization into a less or more productive and organized system is most likely to occur. (Adapted from Gunderson and Holling, 2000)

result of discontinuous changes that move as a wave across space (like an insect outbreak in a forest).

SUCCESSION IN OTHER ENVIRONMENTS

Succession is just as common in aquatic habitats as terrestrial ones. However, succession in ponds, lakes, and streams are particularly in uenced by anthropogenic inputs of materials and energy that speed up, slow, or even stop community development.

Studies in intertidal communities have been particularly useful in developing successional theory, for in these environments the entire sere occurs over a decade where it may take centuries to a millennium in forest ecosystems. Wave action is the most common disrupting force in the intertidal. Newly exposed substrates are colonized by a series of algae that eventually cover 60–90% of the rock surface. However, the presence of the early succession green algae, *Ulva*, can inhibit further succession. In fact, the succession is characterized by a series of species, each of which inhibits its successor. The early pioneers are more resistant to the rigors of the physical environment but also more susceptible to grazers in the system (e.g., crabs). As grazing pressure increases, the early successional species are removed, making way for later successional species.

Degradation is one type of succession in which dead organic material is decomposed. Over time, a series of species of microorganisms and detritivorous animals will exploit and decompose the organic matter until it is completely metabolized and mineralized (*see* Intertidal Zones, Volume 2).

PERSPECTIVES ON SUCCESSIONAL ENDPOINTS

As mentioned previously, the idea that all seres end in stable climax states is no longer accepted. Instead, late successional communities are thought to be in a state of dynamic equilibrium. The equilibrium perspective is based upon the importance of interspecific interactions in shaping the community. There may be some species turnover in a system in a state of dynamic equilibrium, but the addition of new colonists is balanced by extinctions. Predation, competition, and symbiosis become more common in the late successional stages, and this biotic saturation thwarts the introduction of new species.

Alternatively, some communities reach cyclic successional stages on a local scale. In this situation, the juveniles of the late successional species may not be able to establish themselves in the presence of the adults. Shrub-dominated communities often undergo cyclic succession. For example, open areas within the Texan desert scrub in the US are typically invaded by the widely-dispersed creosote bush (*Larrea tridentata*); these areas subsequently succeed to a cactus, the Christmas tree cholla (*Opuntia leptocaulis*), and eventually revert back to bare ground, only to renew the cycle (Yeaton, 1978). The cactus establishes itself in the shade of the creosote bush and over time outcompetes it for the limited supply of moisture. Yet when the protection of the creosote bush is gone, the cactus is subject to erosion and burrowing rodents, which weaken its root system. The cycle begins anew on the barren soil.

A perspective contrary to the dynamic equilibrium views communities as continually being in a state of ux. At all stages in the *sere* (the successional sequence), the identities and numbers of species change. This non-equilibrium model assumes that unpredictable factors such as dispersal and disturbance dominate development of community structure and composition. In this situation, the course of succession may vary with the particular species that become established first. Severe and frequent disturbances may prevent equilibrium from ever being achieved.

Multiple stable states are alternative successional outcomes that exist in one environment, yet do not converge over time to a single state. Multiple stable states are rarely observed but may be more common under intense human activities or natural disturbances. For example, livestock grazing and clear-cutting can lead to new stable conditions. Where grazing pressure in intense and a major source of deforestation, such as in the Brazilian Amazon, grazing can result in a downward spiral toward unsustainability of the land (Serrao and Toledo, 1990). Alternatively, extreme

natural floods in 1991 may have produced multiple stable states now present in riparian vegetation in Southwestern Colorado, US (Baker and Walford, 1995).

Most communities occur in a complex mosaic reflecting the diversity of past disturbances or the topography. For example, fires occur with a variety of intensities across a landscape, which leaves a distinct pattern of vegetation in different states of succession. The progressive retreat of a glacier leaves bands of vegetation in different stages of succession. The matrix of different stages of succession is called a *chronosequence*. When the mosaic reflects topographic differences such as might occur along an elevational gradient, it forms a *toposequence*.

SUCCESSION FROM LARGE, INFREQUENT NATURAL DISTURBANCES

The spatial variability in disturbance intensity can lead to a heterogenous pattern of surviving organisms (Turner *et al.*, 1998). The abundance and distribution of these biological legacies or residuals influence successional dynamics. These surviving organisms dictate the patterns of survival and highly influence the initial stages of reestablishment. They provide some of the earliest seeds as well as the competitive environment to which the new seedlings are exposed. Residual plants often grow larger vegetatively, compared to the newly established seedlings. For example, after a tornado created a 400-ha windthrow in Northwestern Pennsylvania, US, thickets of surviving beech (*Fagus grandifolia*) and maple (*Acer pensylvanicum*) had a clear size advantage over new plants, and that dominance has been maintained (Peterson and Pickett, 1995).

Physical and abiotic legacies are also important to reestablishment after a disturbance. The presence of large rocks or downed trees can provide a safe site where soil moisture and temperature may be more suitable for seedling establishment and growth. The physical presence of plants or even spider webs can trap seeds and retain enough moisture to enhance germination.

Disturbance intensity and size influence those processes that affect succession (Turner *et al.*, 1998). For high intensity disturbances, reestablishment is almost entirely dependent upon propagules being transported from outside the disturbed area. Long-distance colonization by propagules is important especially in the middle of disturbed sites. Typically, the density of propagules dispersed into a disturbed area declines with distance from the edge. However, this pattern of declining density of propagules does not hold as one moves toward the interior of very large disturbances. Disturbance size also interacts with the life history characteristics of potentially colonizing species. For example, *ruderal* (growing on wasteland) species are often common in large disturbed areas. Fireweed (*Epilobium angustifolium*) reestablishes via its light plumed seeds even in large burned areas in the Western US. Yet late-successional species that typically have large seeds are rare. The rate of species migration is also influenced by the size of the disturbance. Furthermore, disturbance size can influence biotic interactions such as herbivory, which can accelerate, impede, or redirect succession. Herbivores are noticeably rare subsequent to large disturbances.

For illustrative purposes, consider one of the largest natural disturbances of the 20th century: the eruption of Mount St Helens in Washington State, US. Plant succession in the aftermath of the 1980 eruption has similarities to recovery after other volcanic eruptions around the world, but, in contrast to other natural disturbances, is much slower and more heterogeneous, probably because of its large size and diversity of impacts (Turner *et al.*, 1997). The 1980 eruption created a variety of disturbances: a 15-km^2 pyroclastic flow, 550 km^2 of trees that had been blown down bordered by 96 km^2 of scorched trees, a 60-km^2 debris avalanche, and massive mud flows. Each of these disturbances resulted in a unique pattern of surviving plants, depending on the life form of species present before the eruption and the characteristics of the disturbance. For example, the Mount St Helens avalanche is the largest landslide in recorded history, and few species or plants survived. There has been a gradual increase in diversity and cover, but only 61% of the species present before the eruption were reestablished in the first 15 years. At that time, plant cover averaged 36% but was extremely patchy. Because part of the avalanche was deliberately seeded by nonnative plant species to reduce erosion, the site offers the opportunity to compare successional processes occurring with and without nonnative species.

Fifteen years after the eruption and 10 years since the seeding by introduced species, sites with dominance by nonnative species have greater vegetation cover and more native plant diversity than plots that were not seeded. Significantly greater mortality of conifers occurred in the plots dominated by nonnative species shortly after seeding, but no difference in conifer mortality has occurred since then. Even so, the plots dominated by nonnative species still have fewer conifer trees. Thus, the short-term pulse of conifer mortality after the seeding of introduced species may have long-term effects on the recovery of the dominant vegetation. Long-term recovery trends and effects of nonnative species on succession are important to understand since typical management practices rely on nonnative species for enhancing vegetation recovery of denuded sites along roadsides, strip mines or other human-generated clearing.

QUESTIONS STILL REMAINING

It is difficult to sort out the relative role of pattern versus processes in the early stages of succession because spatial heterogeneity, nonnative species, long-distance dispersal,

herbivory, and many other factors in uence the path and rate of the seral process. Unfortunately, there is no single instance in which all of these factors have been examined. Current theory of early succession does not adequately consider the broad range of biotic and abiotic variables controlling the early stages of plant establishment and succession. Yet, it is in understanding the early stages of succession that ecological theory may provide the greatest insight into the recovery of disturbed landscapes.

Ecologists are still challenged to predict the role of spatial heterogeneity on vegetation recovery even though the gradual replacement of species through succession is well described. It is not known how the early successional process is shaped by the spatial distribution of resources; the location and diversity of reestablishing species; and the frequency, severity, and extent of stochastic environmental events. There is a need for better understanding of the links between landscape patterns and ecosystem processes to enable better understanding of in uences on plant establishment and biodiversity. These in uences include nonnative species, herbivores, subsequent disturbances, dispersal, and spatial heterogeneity. Appropriate modeling of species invasion is a critical need for forest succession models.

The introduction of nonnative species into established vegetation areas can have far-reaching and long-term impacts. The significant short-term effects of competition with introduced species during early stages of succession are already known. However, how long the introduction of nonnative species during early seral stages will have an effect is not predictable. Alternative stable states of succession may be achieved, or later species in the sere may not be affected by competition between native and nonnative species during the early successional stages. Better understanding of such long-term effects would provide information necessary to develop appropriate management strategies for dealing with nonnative and invasive species.

Seed dispersal typically declines with increasing distance from the parent plant. The pattern of short-range dispersal is well predicted by current theoretical methods. Over larger distances, fewer seeds are dispersed, but their presence has a disproportionate effect on invasion. Seed dispersal over long distances may best be explained by seed morphology and the physical conditions of the surrounding landscape (e.g., microtopography, deposition zones, or wind patterns) (*see* Plant Dispersal and Migration, Volume 2).

Herbivory selectively impacts certain species. Yet the effects of herbivory on early succession are difficult to predict in variable environments. For example, on the debris avalanche at Mount St Helens in Northwestern US, plant diversity and cover are higher within elk (*Cervus elaphus*) exclosures than outside. However, a possible plant response to herbivory is to increase reproduction. The effects of these responses on succession are not clear.

Although succession is a much-studied phenomenon central to ecology, many questions remain about how succession occurs in heterogeneous environments. As global environmental changes alter the Earth, understanding how succession occurs under changing conditions will become even more important. For example, with the growth in human population and subsequent land use changes, there is an expansion in the number and size of areas experiencing early successional conditions. Also, it is not clear how the rate of succession may alter with changing regimes of temperature and precipitation. Furthermore, changes in disturbance regimes can affect patterns of succession.

See also: Succession, De nition of, Volume 2.

REFERENCES

Anderson, D C and MacMahon, J A (1985) Plant Succession Following the Mount St Helens Volcanic Eruption: Facilitation by a Burrowing Rodent, *Thomomys talpoides*, *Am. Midl. Nat.*, 114, 62–69.

Baker, W L and Walford, G M (1995) Multiple Stable States and Models of Riparian Vegetation Succession on the Animas River, Colorado, *Annu. Assoc. Am. Geogr.*, 85, 320–338.

Bazzaz, F A (1979) The Physiological Ecology of Plant Succession, *Annu. Rev. Ecol. Syst.*, 10, 351–372.

Botkin, D B, Janak, J F, and Wallis, J R (1972) Some Ecological Consequences of a Computer Model of Forest Growth, *J. Ecol.*, 60, 849–873.

Clements, F E (1916) Plant Succession: an Analysis of the Development of Vegetation, *Carnegie Institution of Washington*, Publication Number 242.

Cowles, H C (1899) The Ecological Relation of Vegetation on the Sand Dunes of Lake Michigan, *Bot. Gazette*, 27, 95–117, 167–202, 281–308, 361–391.

Gleason, H A (1917) The Structure and Development of the Plant Association, *Bull. Torrey Bot. Club*, 43, 463–481.

Gunderson, L H and Holling, C S, eds (2000) *Theories for Sustainable Futures - Understanding and Managing for Resilience in Human-Ecological Systems*, Island Press, in press.

Holling, C S (1986) The Resilience of Terrestrial Ecosystems: Local Surprise and Global Change, in *Sustainable Development of the Biosphere*, eds W C Clark and R E Munn, Cambridge University Press, New York.

Horn, H S (1975) Markovian Properties of Forest Succession, in *Ecology and Evolution of Communities*, eds M L Cody and J M Diamond, Harvard University Press, Cambridge, MA, 196–211.

McIntosh, R P (1999) The Succession of Succession: a Lexical Chronology, *Bull. Ecol. Soc. Am.*, 80, 256–265.

Moore, P D (2001) A Never Ending Story, *Nature*, 409, 565.

Odum, E P (1969) The Strategy of Ecosystem Development, *Science*, 164, 262–270.

Odum, E P (1983) *Basic Ecology*, Saunders College Publishing, New York.

Olson, J S (1958) Rates of Succession and Soil Changes on Southern Lake Michigan Sand Dunes, *Bot. Gazette*, 119, 125–169.

Peterson, C J and Pickett, S T A (1995) Forest Reorganization: a Case Study in an Old-growth Forest Catastrophic Blowdown, *Ecology*, 76, 763–774.

Serrao, E A and Toledo, J M (1990) The Search for Sustainability in Amazonian Pastures, in *Alternatives to Deforestation*, ed A B Anderson, Columbia University Press, New York, 195–214.

Smith, M, Keevin, T, Mettler-McClure, P, and Barkau, R (1998) Effect of the Flood of 1993 on *Boltonia decurrens*, a Rare Floodplain Plant, *Regul. Rivers Res. Manage.*, 14, 191–202.

Turner, M G, Dale, V H, and Everham, E H (1997) Crown Fires, Hurricanes and Volcanoes: a Comparison Among Large-scale Disturbances, *BioScience*, 47, 758–768.

Turner, M G, Baker, W L, Peterson, C J, and Peet, R K (1998) Factors In uencing Succession: Lessons from Large, Infrequent Natural Disturbances, *Ecosystems*, 1, 511–523.

Whittaker, R H (1953) A Consideration of Climax Theory: the Climax as a Population and Pattern, *Ecol. Monogr.*, 23, 41–78.

Yeaton, R I (1978) A cyclic relationship between *Larrea tridentata* and *Opuntia leptocaulis* in the Northern Chihuahuan Desert, *J. Ecol.*, 66, 65–656.

Suess Effect

The *Suess Effect* is the dilution of radioactive carbon in the atmosphere by the combustion of fossil fuels. Atmospheric carbon dioxide contains a relatively constant content of the radioactive isotope of carbon, ^{14}C. When carbon dioxide is used by photosynthetic organisms, ^{14}C is incorporated into cellular material, for example the cellulose and lignin of growing plants. ^{14}C has a half-life of 5700 years and therefore the ^{14}C content of preserved biological material decreases with time. Thus measurement of ^{14}C in biologically derived objects provides an important method of dating such remains. Fossil fuels are many times older than 5700 years and the radioactive carbon of the original plant material from which they are derived has decayed to unmeasurable levels. Consequently, when such fuels are burned, the extra carbon dioxide liberated into the atmosphere contains no ^{14}C and the radioactive carbon of the atmosphere will be diluted. It is to be expected that the ^{14}C content of the biological products of photosynthesis from this atmosphere containing anthropogenically derived carbon dioxide will be correspondingly low in ^{14}C. This dilution effect is clearly shown in tree rings dated between 1820 and 1950, an observation that was first reported by Suess (1955), after whom the effect is named.

The dilution of atmospheric ^{14}C by fossil carbon provides a key piece of evidence for the impingement of human activity on the global carbon cycle. Unfortunately, this important record was totally obscured by the major release of ^{14}C into the atmosphere as a byproduct of nuclear weapons testing.

REFERENCE

Suess, H E (1955) Radiocarbon Concentration in Modern Wood, *Science*, 122, 415–417.

G R WILLIAMS *Canada*

Symbiosis

Symbiosis describes the situation where two distinct species are closely associated so that they live together. In the broadest definition, symbiosis describes all possible types of consortium between dissimilar organisms, whatever the outcome. In practice, the term is more often applied to relationships where the outcome is beneficial or neutral. When the result is negative or lethal to one member the term *parasitism* is applied. The outcome of parasitism is normally disease (e.g., the growth of mistletoe on poplar trees, or any disease-causing bacterial infection).

Symbiotic association can be *facultative*, where the partners may live either alone or in association, or *obligate*, where one or both of the partners is unable to survive independently. The best-known examples of obligate symbiosis are the lichens, where algal cells within the thallus of a fungus supply organic nutrition by photosynthesis to their host, and are in turn provided with inorganic nutrition and an adequate environment. In the mycorrhizal symbiosis (*see* Rhizosphere, Volume 2) fungi infect roots, usually of trees, and absorb nutrients from the soil, which are transported to the host trees. In return, the trees exude or transport organic nutrients to the fungi. This provides a specific advantage to trees because the fungi may be better able to absorb inorganic nutrients than tree roots. Also, it costs much less, in terms of mass production, to penetrate and mine a given volume of soil by growing fungal hyphae than by growing roots. Both these associations are examples of mutualism, in which both partners benefit from the association.

There are many examples of plant–plant and innumerable examples of animal–animal, bacteria–plant or bacteria–animal associations, but only a few of plant–animal symbiosis. Algal cells living within the cells or protoplasts of animals such as clams, sea anemones or hydra result in the apparent anomaly of photosynthetic animals. Such associations may be long term, with the algal and animal cells dividing in synchrony so that the number of algal cells in

a host cell remains constant. While such associations may not be mutualistic, the animals clearly benefit, and mobile forms will move to seek illumination.

One of the most important symbioses, in terms of agriculture, is the association of nitrogen-fixing bacteria (mostly *Rhizobium* spp.) with the roots of legumes. The process of nitrogen fixation, powered ultimately by the host's photosynthetic metabolism, is estimated to fix upwards of 10^8 tons of nitrogen year^{-1} worldwide. It greatly outweighs the amount of nitrogen applied as fertilizer even in developed countries, and its production is pollution-free.

There are many other nitrogen-fixing symbioses, including the Actinomycete–alder association that can effectively return run-down pastures to productivity. A very interesting example is the association of a cycad (a primitive tree) and a blue-green alga, in which the alga causes a reversal of the normal positive *geotropism* (i.e., downward growth) of the host roots so that they grow up through the soil surface and the algae are able to intercept light for photosynthesis. But this, and the many forms of insect–plant association in which the insect induces the plant to grow an elaborate structure or gall in which it can live, are really examples of non-lethal parasitism.

R G S BIDWELL *Canada*

Symbiotic Concurrent Extinctions

see Extinctions and Biodiversity in the Fossil Record (**Volume 2**)

Systematics

see Taxonomy/Systematics (**Volume 2**)

T

Taiga

see Tundra (**Volume 2**); Boreal Forest (**Volume 2**)

Tamm, Carl Olof

(1919–)

Tamm, Carl Olof, Swedish forest ecologist, is Professor of Forest Ecology at the Forestry Faculty of the Swedish University of Agricultural Sciences, and one of its forerunners, the Royal College of Forestry (1957–1984). Carl Olof Tamm is one of several prominent members of the family Tamm; his father Olof Tamm was professor of Forest Soils at the Royal College of Forestry in Stockholm. Young Tamm received much inspiration from his father's work and also from the senior colleagues of his father at the Forest Research Institute like ecologist/soil biologist Lars-Gunnar Romell and the plant ecologist Carl Malmström. After studies in Lund and Stockholm, Tamm obtained a PhD in botany in Stockholm. During his time in Lund, he demonstrated that a typical so-called "nitrate-plant", fireweed (*Chamaenerion (Epilobium) angustifolium*), could exist on ammonium as the sole nitrogen source, which suggested that the nitrogen ecology of this plant was a question of competition for nitrogen rather than a physiological demand for nitrate. His knowledge and talent were rapidly recognized, and he obtained his Professorship shortly after his PhD. Tamm energetically organized a number of thorough large-scale and long-term experiments with nutrient additions to forests. The larger of these are still unique in an international perspective in terms of the number of treatments, the number of replicate plots and the size of the experimental plots. His work clearly confirmed that the growth of boreal forests was limited by the supply of nitrogen; the experiments also demonstrated that phosphorus, boron, potassium and magnesium might impose secondary limitations on plant growth. When acid rain became a matter of concern in the late 1960s, he was quick to add soil acidification (additions of sulfuric acid or elemental sulfur) and liming treatments to his trials. These treatments, as well as the long-term nitrogen additions, have provided most important insights into the effects of nitrogen and sulfur deposition on forests, and on the effects of remedial actions like liming. Because these trials are located in relatively unpolluted environments, they are ideally suited to assess the effects of soil acidity alone (without any impact of gases like sulfur dioxide), and the effects of nitrogen alone or in combination with other elements, on tree growth. Tamm's interests have been and are wider than this. He stimulated the seminal work on plant nutrition by Torsten Ingestad, played a leading role in the Swedish International Biological Program (IBP) Coniferous Forest Project, explored the ecological role of ground mosses, studied the long-term dynamics of a forest herb, and re-analyzed the soil pH of his father's plots to study acidification, etc. He has set a lasting example with his careful analytical and strongly experimental approach. The scientific legacy of Carl Olof Tamm is found in his 1991 book on terrestrial nitrogen cycling, published seven years after his retirement (his successor as Professor of Forest Ecology was Sune Linder).

REFERENCES

Tamm, C O and Östlund, G (1960) Radiocarbon Dating of Soil Humus, *Nature*, **185**, 706–707.

Tamm, C O (1964) Determination of Nutrient Requirements of Forest Stands, *Int. Rev. For. Res.*, **1**, 115–170.

Tamm, C O and Hallbäcken, L (1988) Changes in Soil Acidity in Two Forest Areas with Different Acid Deposition: 1920s to 1980s, *Ambio*, **17**, 56–61.

Tamm, C O (1991) *Nitrogen in Terrestrial Ecosystems*. *Ecological Studies No. 81*, Springer-Verlag, Berlin.

PETER HÖGBERG *Sweden*

Taxon

Taxonomy arranges organisms into groups with certain degrees of similarity called *taxa* (singular *taxon*). The term *taxon* applies to any taxonomic group regardless of its size or rank. A taxon may thus be very large (e.g., a *kingdom*, which is subdivided into many smaller taxa) or small (e.g., a *genus*, which contains *species*). The taxonomic hierarchy of plant classification may be summarized as follows: the *kingdom Planta*, which contains all plants, is subdivided into divisions (e.g., Magnoliophyta, the flowering plants), which are further divided into classes (e.g., Magnoliopsida, the dicotyledonous plants). The next subdivision is *orders* (e.g., Rosales, the roses and their relatives), further divided into families (e.g., Rosaceae, the rose family), the *divisions* of which are *genera* (e.g., the genus *Prunus*, the plums and cherries), which contains species (e.g., *Prunus domestica*, the plum). Each of these levels of the taxonomic hierarchy is a taxon. It should be noted that various systems of classification, containing more or fewer taxa and levels in the hierarchy of taxa, have been proposed and are currently in use. The one described above for plants, shown in Box 1, is widely but by no means universally accepted.

Box 1 Hierarchy of a plant taxonomic system, with examples

Kingdom	Planta	The plants
Division	Magnoliophyta	Flowering plants
Class	Magnoliopsida	Dicotyledons
Order	Rosales	Roses and related families
Family	Rosaceae	Rose family
Genus	*Prunus*	Plums and cherries
Species	*Prunus domestica*	Cultivated plum

The exact size of taxa and the degree in which they should be grouped or subdivided are still matters of opinion and debate (*see* **Taxonomy/Systematics**, Volume 2).

R G S BIDWELL *Canada*

Taxonomy/Systematics

Taxonomy is the study of the classification of organisms. Classification means fitting organisms into groups or taxa (singular taxon; *see* **Taxon**, Volume 2) based on specific shared characteristics. Various methods of grouping organisms have been developed. The earlier ones were largely artificial and static, i.e., they relied on obvious, usually visual characteristics (e.g., flower color) and regarded all organisms as fixed (the products of creation). The Swedish botanist Linnaeus understood the value of *natural* categories, i.e., categories based on biological characteristics of the organisms, but he was aware that the basic biological understanding of organisms at his time was inadequate. Darwin's theory of evolution provided a much more biologically sound basis for the classification of organisms based on the relationships of their evolutionary ancestry, and this became known as *systematics*. Today the terms *systematics* and *taxonomy* are used virtually interchangeably. Modern taxonomy relies more and more heavily on evolutionary relationships, but many less natural conventions still prevail where adequate genetic evidence of relationships has not yet been developed.

The fundamental unit of taxonomy is the *species*. Species are usually defined as groups of individuals that are similar or identical in appearance and other characteristics and that can interbreed, geographic conditions permitting. However, it is now clear that the species is not a static unit, as organisms continue to evolve. The result is that the definition of any given species is frequently blurred and often quite arbitrary, i.e., to suit the conventions of the taxonomist. Furthermore, species are frequently subdivided into *subspecies* and *races* in order to place organisms that appear to belong to a single species into smaller groups on the basis of consistent, but minor, differences in certain characteristics.

Taxonomy, once considered a purely descriptive science, is now developing on a wide experimental base. The systematic approach includes classification on the basis of genetics, cytology, biochemistry and chemistry. In addition, the structure of the taxonomic system depends more and more on genetic and evolutionary analyses. Nevertheless, the earlier requirements for observation, description, nomenclature and documentation (i.e., collection) of species and specimens are still essential in the science of taxonomy.

R G S BIDWELL *Canada*

Temperate Coniferous Forests

Richard H Waring
Oregon State University, Eugene, OR, USA

Temperate coniferous forests include representatives of the most massive forms of terrestrial life on Earth. In the

coastal redwood (Sequoia sempervirens) forests of northern California, trees may reach heights >100 m and accumulate biomass in excess of 3000 t ha^{-1}. In the more extensive Douglas fir (Pseudotsuga menziesii) forests of the Pacific Northwest, USA and Canada, biomass in older stands frequently reaches 1000 t ha^{-1}, with an additional 500 t ha^{-1} of standing dead and fallen trees. Above-ground biomass ranges widely, however, as does above-ground growth. Temperate coniferous forests produce an average of about 5 t ha^{-1} year^{-1} of above-ground biomass but in climatically diverse areas production may range from <1 to >20 t ha^{-1} year^{-1} (Runyon et al.*, 1994).*

Because of their high growth capabilities, favorable wood structural properties and columnar form, temperate coniferous forests play a disproportionate role in meeting global demands for lumber and fiber. As a consequence, the geographic range of temperate conifers has been extended beyond the native ranges of many species.

In this contribution we focus on the attributes of temperate evergreen coniferous forests that distinguish them from their major competitors, deciduous hardwoods. This differentiation should help us anticipate the future role and importance of temperate coniferous evergreen in relation to rising levels of air pollution, atmospheric carbon dioxide (CO_2), and changing climate.

EVOLUTION

Conifers bear naked seeds, unlike angiosperms, and differ from other gymnosperms by having the seeds arranged inside cones. During the Carboniferous period (300 million years ago), conifers represented a minor group of gymnosperms, but they evolved rapidly as conditions became drier and colder during the Permian (286–240 million years ago). Temperate coniferous forests were most widely distributed during the Tertiary period (90–15 million years ago), with coastal redwood (*Sequoia sempervirens*) and dawn redwood (*Metasequoia*) distributed throughout both Eurasia and North America. Today, conifers are the dominant representatives of gymnosperms with about 50 genera and 550 species, predominantly in the Northern Hemisphere (Raven *et al.*, 1986).

Conifers are well adapted to arid and cold conditions. The surfaces of needle-shaped leaves are protected by waxes and sunken stomata and a large proportion of their stems consists of sapwood that serves as a storage reservoir, as well as a conduit for transport of water from roots to leaves.

GEOGRAPHIC DISTRIBUTION

Today, temperate evergreen coniferous forests cover approximately 2.4×10^6 km^2 (Melillo *et al.*, 1993; Landsberg and Gower, 1997). Conifers dominate the montane forests in North America, Europe, and China. Smaller areas of temperate conifers are located in montane regions of Korea, Japan, Mexico, Nicaragua, and Guatemala. In Europe, the distribution of native and introduced conifers has been expanded into areas more climatically favorable to hardwoods. Pines (*Pinus*), represented by more than 90 species, are the most widely distributed conifer and their range has increased by widespread planting in the Southern Hemisphere. Other important genera include the firs (*Abies*), spruces (*Picea*), hemlocks (*Tsuga*), false-hemlocks (*Pseudotsuga*), larches (*Larix*), cypresses (*Cupressus*), cedars (*Chamaecyparis*, *Thuja*, *Libocedrus*) and juniper (*Juniperus*). In the yews (*Taxus*), the seeds are not borne in cones but are enclosed within a fleshy structure. Additional conifer genera (*Araucaria*, *Fitzroya*, *Austrocedrus*) are present in the Southern Hemisphere, where they were once more widely distributed (Axelrod *et al.*, 1991).

ECOLOGICAL DISTRIBUTION

The potential ecological niche of temperate conifers falls within regions of >250 mm in annual precipitation that experience subfreezing conditions, down to, but not below −45 °C, which is the limit for supercooling of water. Only boreal tree species, which include some pines and spruces, are adapted to temperatures below this limit (Waring and Running, 1998). In regions where precipitation is well distributed throughout the growing season, deciduous hardwoods usually dominate over temperate conifers. Following major disturbances, however, conifers can become established and achieve temporary dominance (Landsberg and Gower, 1997).

In the Pacific Northwest region, where summer drought is common, the situation is reversed and many long-lived conifers replace earlier established hardwoods (Waring and Franklin, 1979). To explain these shifts in dominance requires an appreciation of how resources are captured by temperate evergreen conifers throughout the year.

PHYSIOLOGICAL AND MORPHOLOGICAL ADAPTATIONS OF CONIFERS

Conifer leaves are more clumped and narrower in cross-section than the foliage of most angiosperms, and the branch structure is much less extensive. These properties allow light to penetrate more deeply through conifer canopies than through hardwood forests with comparable leaf area. As a result, the maximum surface area of conifer foliage, often expressed as m^2 of (projected) leaf area per m^2 of ground surface, may reach 10–12, while hardwood forests rarely exceed half these values. In addition, conifer canopies reflect only about 5–10% of intercepted solar radiation, whereas hardwoods reflect 15–25%. In combination,

these differences in canopy properties permit coniferous forests to absorb a greater proportion of incoming solar radiation than forests composed of hardwoods (Jones 1992).

Although conifers absorb a larger fraction of intercepted radiation than hardwoods, they are less prone to suffer damage from elevated temperatures because the needle-shaped foliage dissipates heat efficiently, even under unventilated conditions. Consequently, conifers conserve water under conditions where broad-leaf species must transpire at higher rates, wilt, or shed foliage to prevent temperatures from rising above 45 °C, when enzymes begin to become unstable.

Another foliage characteristic that distinguishes most conifers from other evergreen vegetation is the duration that leaves remain functional. Conifers generally hold some foliage for 3–5 years, although a few species maintain functional leaves for 30–40 years (Landsberg and Gower, 1997). The fact that temperate evergreen conifers can maintain dense, needle-leaf canopies, composed of multiple-age classes of foliage, affects many ecosystem processes including the capture of water, nutrients, and pollutants from the air, and the rates of decomposition, nutrient release, and soil organic matter accumulation.

Anatomically, long-lived foliage is characterized by an epidermis covered with a thick waxy cuticle, above one or more layers of compactly arranged, thick-walled cells, which in turn surround a structurally reinforced vascular system. The result is that nutrient concentrations in conifer needles may be only half those in deciduous hardwood leaves of equivalent mass. Lower nutrient concentrations in individual leaves translate into lower demand, yet long-lived foliage provides a greater store of nutrients available for transfer during peak growth periods than is available in deciduous hardwoods.

In addition, in environments where mineral weathering rates are slow and nutrient availability limited, evergreen conifers can take up nutrients from the soil as long as conditions remain above freezing. Photosynthesis can also continue under these conditions. As a result of nutrient uptake and photosynthesis during the dormant season, stored reserves (starch, sugars, amino acids) are available to activate symbiotic fungi on tree roots that efficiently scavenge nutrients. In contrast, deciduous trees, which also have symbiotic root fungi, are restricted to taking up nutrients when foliage is present.

The evolutionary advantages that conifers have on infertile soils can place them at a disadvantage on more fertile sites. On infertile soils, most available nitrogen (N) is in the form of ammonium (NH_4^{+1}), which conifers take up selectively in preference to nitrate–nitrogen (NO_3^{-1}). As nitrate–nitrogen becomes more available, conifers are disadvantaged in competition with hardwoods (Smirnoff and Stewart, 1985).

ECOSYSTEM RESPONSES OF TEMPERATE CONIFEROUS FORESTS

Decomposition

Leaf litter, wood, and root materials produced by evergreen conifers usually contain twice the amount of carbon (C) in relation to nitrogen found in corresponding materials produced by deciduous angiosperms. As a result, the decomposition of coniferous litter is usually 3–4 times slower than hardwood litter, leading to a greater accumulation of forest floor litter under conifers (Figure 1).

With time, as litter decays, soils under coniferous forests maintain high C–N ratios and serve as storage sites for amounts of carbon that far exceed above-ground biomass, and have turnover times of centuries and millennia. This property of coniferous forest ecosystems, combined with their potential rapid growth, makes them an attractive vegetation type to manage for the sequestration of large amounts of atmospheric carbon dioxide (CO_2).

Hydrology

The potential of evergreen coniferous forests to maintain dense, well ventilated canopies throughout the year significantly increases the annual water vapor transfer from

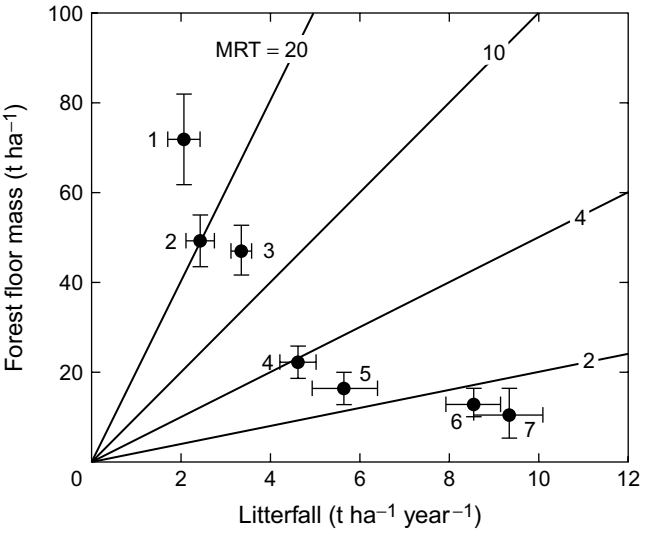


Figure 1 Average forest oor mass plotted against litter-fall for (1) boreal needle-leaved evergreens, (2) boreal broad-leaved deciduous, (3) temperate needle-leaved evergreen, (4) temperate broad-leaved evergreen, (5) temperate broad-leaved deciduous, (6) tropical broad-leaved evergreen, and (7) tropical broad-leaved deciduous trees. Assuming the forest oor is at steady state, the average mean residence time (MRT, years) can be calculated as forest oor mass/litterfall mass. The lines indicate that forest oor litter of temperate needle-leaved evergreens take about 15 years to turnover whereas other temperate hardwoods turnover in <4 years. (Reproduced from Landsberg and Gower, 1997)

their canopies, relative to deciduous forests. As a result, evergreen coniferous forests reduce stream flow by an average of 15–25% in comparison to deciduous hardwoods in similar environments (Waring and Schlesinger, 1985). The dense, finely dissected canopies of evergreen conifers also have the propensity to condense water (and pollutants) from clouds, leading to fog drip, which can increase the effective precipitation at high elevations substantially in windy environments (Harr, 1982).

Water stored in the sapwood of conifer trees can serve as a temporary buffer against drought. During the day, this reservoir may contribute up to a third of the water transpired, with refilling taking place during the night. In general, the ratio of sapwood mass to leaf area increases in conifers with harshness of the environment (Waring and Running, 1998); the total volume of sapwood determines the reservoir of water. Thus, a 50 m tall Douglas fir forest with about a quarter of the stem composed of sapwood may hold 30 mm of stored water whereas a 15 m tall pine forest, with stems composed largely of sapwood, can hold 20 mm of stored water (Waring and Running, 1978; Waring *et al.*, 1979). In a few species, such as redwood, the voluminous heartwood also plays an important role as a reservoir.

RESPONSE OF TEMPERATE CONIFEROUS FORESTS TO AIR POLLUTION

In clean environments, nitrogen deposition is usually $<5\,kg\,ha^{-1}\ year^{-1}$ whereas in heavily polluted areas, nitrogen additions exceed $50-70\,kg\,ha^{-1}\ year^{-1}$. Much of this excess nitrogen ends up as nitrate, which is easily leached in soil solution. The negatively charged nitrate ions extract calcium, magnesium, and other positively charged nutrient from soil cation (+) exchange sites. Because weathering rates are slow, the lost nutrients are often replaced with positively charged acid radicals (H^{+1}) and toxic forms of aluminum (Al^{+3}). As a result, a nutrient imbalances may arise where the availability of nitrogen is in excess to that of other elements, creating conditions where plants become more nutritious, and less resistant to attack from a variety of organisms (Waring and Running, 1998).

In areas where hydrocarbons released by the combustion of fossil fuels interact with sunlight, ozone ($O_3{}^{-3}$) is formed. Ozone is a highly reactive oxidizing agent that diffuses through stomatal pores on leaf surfaces when they are open and injures the photosynthetic machinery inside. Over a single growing season, conifer foliage suffers less damage than deciduous hardwoods because gas exchange rates through stomata are only about half those of deciduous hardwoods for an equivalent surface area (Reich and Amundson, 1985). With continued exposure, however, older foliage on conifers sustains accumulated damage and is shed prematurely (McLaughlin, 1985), leading to reduction in tree vigor and a greater likelihood of disease and insect outbreaks (Waring and Running, 1998).

RESPONSE OF TEMPERATE CONIFEROUS FORESTS TO CLIMATIC CHANGE AND INCREASING ATMOSPHERIC CO_2

We might expect temperate coniferous forests to be favored over deciduous hardwoods if climatic conditions during the dormant season become more favorable and those during the growing season less so. Thus, if minimum temperatures rise during the winter, the constraints exerted by subfreezing and suboptimal temperatures on photosynthesis will be reduced (Runyon *et al.*, 1994; Landsberg and Waring, 1997). The relative importance of warmer conditions during the winter can be assessed by considering the seasonal shifts in potential radiation available at different latitudes, aspects, and slopes to evergreen conifers (Running *et al.*, 1987; Waring and Running, 1998). If warming is combined with more frequent storms, without necessarily any increase in total precipitation, the associated increase in cloud cover may substantially reduce solar energy available to melt snow or drive photosynthesis.

On the other hand, if climatic warming extends throughout the year, the growing season will be lengthened. A higher frequency of storms in this case could moderate summer temperatures by allowing vegetation to maintain transpiration at near maximum rates and reduce the fraction of energy going into heating the air (Figure 2). More frequent wetting of surface litter and soil would also foster more rapid decomposition and release of nutrients, while reducing the likelihood of forest fires. This climatic scenario would favor replacement of conifers by hardwoods.

To factor in the response of vegetation to rising atmospheric CO_2 concentrations is more of a challenge. In general, increases in CO_2 should result in elevated rates of photosynthesis because carbon dioxide is a limiting resource for all tree species. Moreover, a rise in CO_2 should reduce photorespiration and increase the efficiency of photosynthesis, particularly if conditions become warmer. Because stomatal adjustments in most tree species maintain CO_2 concentrations within their leaves at about 75% of the value in ambient air, we might conclude that rising atmospheric CO_2 will result in a net reduction in transpiration per unit leaf area, and thus increased carbon gain per unit of water lost. This may be generally true for many hardwood species, but there is less evidence that conifer stomata respond similarly (Norby *et al.*, 1999). In fact, in one field experiment where ambient CO_2 levels were artificially elevated for more than two years, no shift in pine stomata behavior was recorded (Ellsworth, 1999).

If net photosynthesis were to increase significantly, canopy leaf area index would reach a higher value, assuming sufficient nutrients were available. An increase in leaf

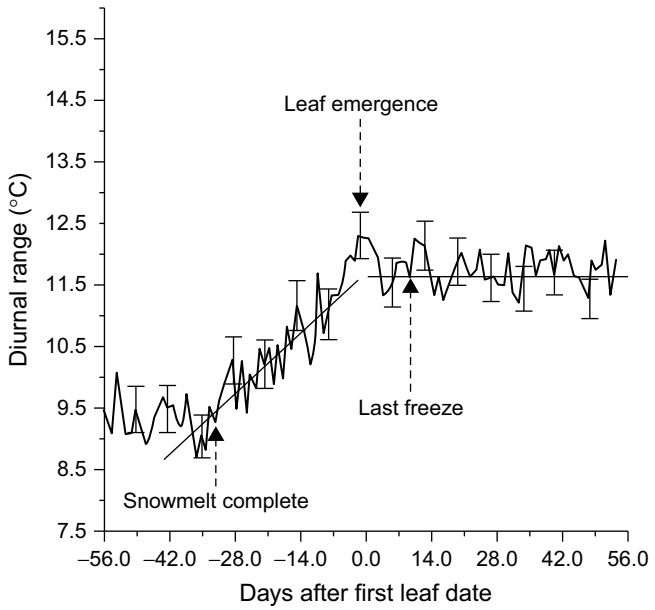


Figure 2 Average daily surface temperature amplitude (maximum- minimum), measured before and after spring leaf emergence of local vegetation for 12 sites across the north-central United States. The rapidly warming spring temperatures generate progressively larger temperature amplitudes during the 2- 3 week period prior to lea ng out; the amplitude is dampened, however, by leaf emergence as partitioning of incoming solar radiation is shifted as a result of cooling by rapidly transpiring vegetation. (Reproduced from Schwartz and Karl, 1990; Waring and Running, 1998)

area would affect the water balance by increasing canopy interception, evaporation, and transpiration under a stable climate. On the other hand, if climatic conditions were to warm without a corresponding increase in precipitation, drought could reduce photosynthesis, growth, and canopy leaf area.

In summary, although changing patterns of climate and atmospheric CO_2 concentration together are likely to have subtle effects on tissue biochemistry and the allocation of resources, forest growth and composition could be dramatically altered as a consequence of disturbance from outbreaks of fire, diseases, and insects.

FOREST MANAGEMENT OPTIONS

With increasing demand for forest products, and increasing appreciation of other values derived from forests, there is considerable controversy and concern over how forests, on both public and private lands, are to be managed. Temperate coniferous forests have tremendous value to human societies, so there is justified concern about their management. Ancient coniferous forests have long been appreciated by many cultures as places of beauty, often with religious significance. Many groves have been designated as sacred and others set aside as natural wonders to be preserved in parks. The value of forests for maintaining clear running streams and high quality drinking water has also been long appreciated. Recently, modifying the structural characteristics of young coniferous forests to create properties more typical of mature stands has gained acceptance as a means of creating more favorable habitat for rare and endangered species commonly associated with extremely old forests. At the same time, ways of increasing the amount of carbon stored in biomass and soils are being developed. These carbon-sequestering techniques affect site preparation and harvests as well as the design, use, and recycling of forest products.

It is clear that young plantations will play an increasing role in providing lumber and fiber: whether these plantations replace native species or reclaim marginal farmland, however, will make a difference in the extent that plant and animal biodiversities are affected. Similarly, the fraction of the landscape occupied by plantations will alter the likelihood of fire, water runoff, and the potential for outbreaks of insects and diseases.

The challenge associated with more active management for wood, water, biodiversity and carbon storage, is the need to anticipate the extent to which disturbances may occur across broad landscape and ownership patterns. The introductions of exotic organisms into native forests is a particular concern because exotic insects and pathogens can result in replacement of native trees and open the environment to invasion by noxious weeds and other organisms.

General models that employ computer and satellite technology to compare projected changes with those that can be observed across broad landscapes and regions can provide a valuable aid to decision making. Remote sensing technology by itself permits us to monitor the state of the Earth's vegetation and the extent of change in vegetative cover over the last three to four decades.

REFERENCES

Axelrod, D I, Kalin-Arroyo, M T, and Raven, P H (1991) History of Temperate Vegetation in the Americas, *Rev. Chil. Hist. Nat.*, **64**, 413–446.

Ellsworth, D S (1999) CO_2 Enrichment in a Maturing Pine Forest: are CO_2 Exchange and Water Status in the Canopy Affected? *Plant, Cell Environ.*, **22**, 461–472.

Harr, R D (1982) Fog Drip in the Bull Run Municipal Watershed, Oregon, *Water Resour. Bull.*, **18**, 785–789.

Jones, H G (1992) *Plants and Microclimate*, 2nd edition, Cambridge University Press, Cambridge.

Landsberg, J J and Gower, S T (1997) *Applications of Physiological Ecology to Forest Management*, Academic Press, San Diego, CA.

Landsberg, J J and Waring, R H (1997) A Generalized Model of Forest Productivity Using Simplified Concepts of Radiation-use Efficiency, Carbon Balance and Partitioning, *For. Ecol. Manage.*, **95**, 209–228.

McLaughlin, S B (1985) Effects of Air Pollution on Forests, a Critical Review, *J. Air Pollut. Ctrl. Assoc.*, **35**, 513–534.

Melillo, J M, McGuire, A D, Kicklighter, D W, Moore, III, B, Vorsmarty, C J, and Schloss, A L (1993) *Global Climate Change and Terrestrial Net Primary Production*, *Nature*, **363**, 234–240.

Norby, R J, Wullschleger, S D, Gunderson, C A, Johnson, D W, and Ceulemans, R (1999) Tree Responses to Rising CO_2 in Field Experiments: Implications for Future Forests, *Plant, Cell Environ.*, **22**, 683–714.

Raven, P M, Evert, R F, and Eichhorn, S E (1986) *Biology of Plants*, 4th edition, Worth, New York.

Reich, P B and Amundson, R G (1985) Ambient Levels of Ozone Reduce Net Photosynthesis in Tree and Crop Species, *Science*, **230**, 566–570.

Running, S W, Nemani, R R, and Hungerford, R D (1987) Extrapolation of Synoptic Meteorological Data in Mountainous Terrain and its Use for Simulating Forest Evapotranspiration and Photosynthesis, *Can. J. For. Res.*, **17**, 472–483.

Running, S W, Loveland, T R, Pierce, L L, and Hunt, E R, Jr (1995) A Remote Sensing Based Vegetation Classification Logic for Land Cover Analysis, *Remote Sens. Environ.*, **51**, 39–48.

Runyon, J, Waring, R H, Goward, S N, and Welles, J M (1994) Environmental Limits on Net Primary Production and Light-use Efficiency across the Oregon Transect, *Ecol. Appl.*, **4**, 226–237.

Smirnoff, N and Stewart, G R (1985) Nitrate Assimilation and Translocation by Higher Plants: Comparative Physiology and Ecological Consequences, *Physiol. Plant*, **64**, 133–140.

Schwartz, M D and Karl, T R (1990) Spring Phenology: Nature's Experiment to Detect the Effect on "green-up" on Surface Maximum Temperatures, *Mon. Weather Rev.*, **118**, 883–890.

Waring, R H and Franklin, J F (1979) Evergreen Coniferous Forests of the Pacific Northwest, *Science*, **204**, 1380–1386.

Waring, R H and Running, S W (1978) Sapwood Water Storage: its Contribution to Transpiration and Effect Upon Water Conductance Through the Stems of Old-growth Douglas-fir, *Plant, Cell Environ.*, **1**, 131–140.

Waring, R H, Whitehead, D, and Jarvis, P G (1979) The Contribution of Stored Water to Transpiration in Scots Pine, *Plant, Cell Environ.*, **2**, 309–317.

Waring, R H and Schlesinger, W H (1985) *Forest Ecosystems: Concepts and Management*, Academic Press, Orlando, FL.

Waring, R H and Running, S W (1998) *Forest Ecosystems: Analysis at Multiple Scales*, 2nd edition, Academic Press, San Diego, CA.

FURTHER READING

Townshend, J, Justice, C, Li, W, Gurney, C, and McManus, J (1991) Global Land Cover Classification by Remote Sensing: Present Capabilities and Future Possibilities, *Remote Sens. Environ.*, **35**, 243–255.

VEMAP members (1995) Vegetation/ecosystem Modeling and Analysis Project (VEMAP): Comparing Biogeography and Biogeochemistry Models in a Continental-scale Study of Terrestrial Ecosystem Responses to Climate Change and CO_2 Doubling, *Global Biogeochem. Cycles*, **9**, 407–437.

Temperate Deciduous Forests

Peter B Reich and Lee Frelich
University of Minnesota, St. Paul, MN, USA

Temperate deciduous forests are among the major biomes on earth. They are also among the biomes that have been most heavily impacted by global change factors. Although there is the perception that the greatest concern should be held for tropical rain forests, temperate deciduous forests have a smaller fraction of original vegetation remaining than boreal or tropical forests, and also have been more severely impacted by land use change and air pollution. Temperate deciduous forests are most common in Europe, eastern North America, Japan and China, and in all these regions humans have used them heavily as a resource. Logging, grazing, conversion to agriculture, fire suppression and manipulation of wildlife populations, have caused massive alterations of forest composition, structure and diversity. Temperate biomes in Europe and North America bore the brunt of the industrial revolution, providing wood for fuel, building materials, and raw material in industrial processing. Boreal biomes have been spared widespread permanent conversion in vegetation type due to their inappropriateness for intensive agriculture. Conversion from native vegetation to agricultural landcovers has become widespread in the tropics only late in the 20th century. Thus, temperate forests have experienced more severe impacts over broader areas than any other large forest biome (Williams, 1989).

Additionally, because they are present in some of the most heavily populated and industrialized regions on earth, temperate deciduous forests have also been subject to biological invasions and to elevated levels of air pollution, including ozone, sulfur dioxide (i.e., acid deposition) and nitrogen deposition. Nonetheless, the land covered by temperate deciduous forests (as opposed to land once converted to agriculture) generally increased late in the 20th century. In this essay, we will describe the extent and characteristics of temperate deciduous forests, and the major global change agents that impact them.

NATURE AND EXTENT OF TEMPERATE DECIDUOUS FORESTS

Temperate deciduous forests occupy a central position among the world's biomes, occurring in regions of intermediate temperature, precipitation and latitude. They are bracketed geographically by biomes that are more extreme in climate, productivity and diversity: boreal forest and tundra to the colder high latitudes; temperate coniferous forests

and grasslands toward the drier climate zones; and tropical forests in warmer regions (Reich and Bolstad, 2001). Forest vegetation is classified by broad climate zones (tropical, temperate, boreal) and by leaf form and habit (broad versus needle-leafed forest, evergreen versus deciduous). Temperate forests are considered one of the five dominant forest types, along with boreal forest, temperate mixed or coniferous forests, subtropical/tropical wet forests, and subtropical/tropical dry forests. Generally, a forest is considered temperate if it is found in a climate with a warm summer and a markedly cool or cold winter. Typically, temperate forests are found between 3 and 18 °C mean annual temperature, with midwinter mean monthly temperatures generally below 8 °C and midsummer mean monthly temperatures generally above 18 °C.

Temperate forests occupy much of North America, Europe, Asia, Australia, New Zealand, and a small part of South America, but those in the southern hemisphere are almost exclusively evergreen, so temperate deciduous forests are generally restricted to mid-latitudinal belts in the Northern Hemisphere. Although needle-leafed deciduous forests occur, the most common of these are Larix-dominated and typically a component of boreal forests and these will be largely ignored in this essay. Thus, we will focus on broad-leaved forests.

Broad-leafed deciduous forests typically occupy colder and/or moister parts of the temperate zone than broad-leafed evergreen forests, which span the ecotones between warm temperate, subtropical, and Mediterranean biomes. For instance, in Japan, moving from north to south within a continuously moist climate, native forests are dominated by mixed conifer–hardwood forests in the cold temperate north, by broad-leafed hardwoods in the central region, and in the warm temperate south, by mixed deciduous–evergreen or evergreen broad-leafed forests.

MAJOR GLOBAL CHANGE ELEMENTS, MECHANISMS OF IMPACT, AND MAGNITUDE OF EFFECTS

There are a number of different global change factors that collectively influence temperate deciduous forests of the world. These include land use history, current land management, climate change, fire suppression, changes in atmospheric chemistry, and invasions of exotic species. These human impacts on temperate forests have been both direct and indirect. Direct impacts include site conversion to other land covers and land uses, including inadvertent effects on the native species composition. Indirect impacts include the introduction or elimination of herbivorous or competing species, fire suppression and attendant changes in disturbance regimes, atmospheric pollution, particularly carbon dioxide (CO_2), ozone, sulfur, and nitrogen compounds, changes in the age structure of forest through cutting patterns and practices, and changes in stand age demographics on longer time scales related to historical changes in local, national, and global economies. Although their effects tend to interact, for simplicity we will discuss each in turn.

Land Use History

Almost all temperate deciduous forests have been severely impacted by human use. In Europe and North America, less than 1% of all temperate deciduous forests remain in undisturbed state, free of logging, grazing, deforestation or other intensive use. The largest direct impacts on temperate deciduous forests stem from conversion to other land uses. Pre-industrial forest extent is unknown, but conversion to other landcovers has been profound. Native forests have been all but eliminated in some countries, such as Ireland and Britain. In general, roughly half the original temperate deciduous forestland has been converted to other uses, primarily agriculture. Thus, forest conversion to agriculture has been of critical importance. In Europe, massive deforestation occurred centuries ago. In eastern North America, deforestation migrated westward with agricultural settlement from the 1600s to the mid 1800s. For instance, lands originally almost entirely forested, such as Vermont, were more than 80% deforested and converted to agriculture in the 19th century. However, many temperate zone economies have shifted from rural, agricultural to urban manufacturing and technologically driven economies. This has led to large-scale agricultural abandonment and the reversion of agricultural lands back to forests. Recent studies indicate an increase in forest area and volumes in both North America and Europe (Kauppi *et al.*, 1992; Houghton, 1995).

Land Management and Fire Suppression

Although fire was not historically as severe a disturbance in temperate deciduous forests as in boreal or temperate coniferous forests, it nonetheless played an important historical role (Pyne, 1982). In eastern North America, Native American burning and lightning resulted in relatively frequent fire in the temperate deciduous forests. This was important to drier regions such as near the prairie–forest border. Forest fire had a tremendous impact on the composition and age structure of the forest, since certain species have adaptations such as thick bark, ability to sprout, and rapid post-fire colonization that enable them to thrive under such conditions (Reich *et al.*, 1990; Abrams, 1992; Kruger and Reich, 1997; Peterson and Reich, 2001).

Fire frequency regulated the balance between late successional species such as sugar maple, beech, and linden and shade-intolerant early successional species of oak and aspen. Oak and aspen were most abundant along the prairie

forest border and areas with sandy soil where fires were most frequent (Grimm, 1984; Abrams, 1992).

With increasing development following European settlement and expansion, fires in temperate deciduous forests became much less frequent. This was due to cessation of intentional burning, direct suppression of fires, and land use changes that disrupted the contiguity of burnable vegetation across the landscape. Hence, temperate deciduous forests have gradually become increasingly dominated by shade-tolerant species such as maple, beech, ash, and linden, with decreased abundance of oaks (Crow, 1988; Abrams, 1998). In the absence of fire, oaks do not establish well in either shaded understorys or sunlit openings, because they are neither shade tolerant nor fast growing (Reich *et al.*, 1990; Abrams, 1992; Kruger and Reich, 1997). Hence, a major change in temperate deciduous forests of North America has resulted from the ascendency of fire-intolerant species to a dominant position in these regions. However, oaks are expanding their range westward into the prairie–woodland region, as a consequence of fire suppression. Fire suppression has likely had similar effects in Europe and Asia, but the greater time since active fire regimes makes it more difficult to be specific about the changes that have occurred.

Along with fire management, silvicultural management has also altered the temperate deciduous forest regions of Europe and North America (Parry, 2000). In Europe, almost all native temperate deciduous forest has been extirpated and of the remaining area still forested, or again forested, the majority has been planted to monocultures or mixtures not necessarily representative of the native vegetation. Much native temperate deciduous forestland has been planted with Scots pine or Norway spruce monocultures, or deciduous mixtures dominated by Fagus. In North America, many temperate deciduous forests have been replaced by pine plantations or mixed forests, especially in the southeast. Conversely, forests dominated by pines in the northern range of temperate deciduous forests have often shifted to aspen dominance following logging.

Species Effects: Local and Invasions

Local species extirpations (e.g., *Castor canadensis*, the North American beaver) and introductions (e.g., *Lymantra dispar*, the European gypsy moth to North America) can have large aggregate impacts on species composition and nutrient and mass cycling rates. The exotic Chestnut blight and Dutch elm disease have caused the virtual disappearance of native Chestnut (*Castanea*) and American elm (*Ulmus*) from the temperate deciduous forests of North America. Uncharacteristically low and high region-wide populations of beaver and white-tailed deer, respectively, have caused shifts in the relative abundance of tree species whose populations or habitat are affected by browsing of these animal species. Although temperate deciduous forests in northeastern North America have not been invaded by a successful tree canopy dominant from overseas, the Asian species *Paulownia* is a successful invader in the southeastern US and two small trees, *Lonicera* spp and *Rhamnus cathartica*, have displaced native shrubs and prevented tree regeneration in many suburban regions of the eastern US. Moreover, the nitrogen-fixing tree, *Robinia pseudoacacia* has greatly expanded its range in eastern North America in the past several centuries. In Europe, several North American species, including white pine, black locust, and red oak, were planted and have become naturalized in the temperate deciduous forest region.

Impacts on Temperate Deciduous Forests Due to Changes in Atmospheric Chemistry

The potential indirect effects of elevated CO_2 on climate warming, and resultant consequences, are addressed elsewhere (*see* various papers in volume 1, and **Intergovernmental Panel on Climate Change (IPCC): an Historical Review**, Volume 4). Additionally, due to its role as substrate for photosynthesis, elevated atmospheric CO_2 levels may directly increase forest productivity and cause shifts in species composition (Parry, 2000). Numerous hypotheses have been proposed regarding which kinds of species will be favored under enriched CO_2 conditions, with faster-growing, less shade-tolerant species often considered potentially more responsive. However, the evidence is extremely inconclusive, leaving us with high uncertainty regarding how elevated CO_2 levels will change future forests (*see* **Nitrogen Cycle**, Volume 2; **Nitrogen Deposition on Forests**, Volume 2).

Elevated levels of nitrogen deposition and tropospheric ozone are also common in temperate deciduous forests regions and known to have significant effects on forests. Atmospheric nitrogen inputs have increased significantly over the last 30 years (Aber *et al.*, 1989), with large increases in per hectare nitrogen inputs reported, particularly in Europe (Kauppi *et al.*, 1992). Nitrogen deposition may be having substantial effects on tree production and patterns of carbon and nitrogen cycling (Townsend *et al.*, 1996; Magill *et al.*, 1997). Changes in nitrogen availability could have large impacts on species composition (toward more nitrogen-demanding taxa), litter quality and quantity, decomposer populations, and nutrient and mass cycling rates (*see also* **CO_2 Enrichment: Effects on Ecosystems**, Volume 2; **Plant Growth at Elevated CO_2**, Volume 2). Ozone is present at elevated levels during the growing season in the European, North American, and probably Asian regions of temperate deciduous forests. These pollution levels are sufficient to cause reduced productivity at leaf, tree, and stand-scales (Reich, 1987; Ollinger *et al.*, 1997). Whether ozone pollution causes shifts in forest composition is still unclear. Increases in atmospheric greenhouse gases

may also have large impacts through climatic alterations (see below).

Climate

The climate in temperate deciduous forest zones has been fairly stable for the last 1000–2000 years. The impact of warming during the 20th century has not yet manifested itself in forest change, although it is likely that direct human disturbance in temperate deciduous forests would mask any change due to climate that may have occurred.

Temperate deciduous forests are heavily impacted by wind disturbance, including tornadoes and thunderstorm downbursts in central North America and hurricanes (or typhoons) and severe extra-tropical low-pressure systems for forests near the oceans. The frequency of such stand-leveling winds is expected to increase under a warmer climate, so that fewer stands would reach old-growth stages of development. Warmer climates would also cause mixed effects due to greater drought stress and fire frequency in some regions, but fewer in other regions.

POTENTIAL TEMPERATE FOREST RESPONSES TO GLOBAL CHANGE

How will changing environmental factors (climate, CO_2, nitrogen deposition, land use) collectively affect temperate forest net primary production (NPP) and associated processes? Although it is nearly impossible to address these questions empirically at the present time (due to the lack of data), these issues may be addressed from a conceptual and quantitative modeling framework (Parry, 2000; Reich and Bolstad, 2001).

Acknowledging the shortcomings of current models, what do they suggest about temperate forest responses to global change issues such as land use change, elevated CO_2, nitrogen deposition, ozone pollution, climate and other issues? Running and Nemani (1991) modeled forest ecosystem response to various climate change factors, including a doubling of atmospheric CO_2, a 4 °C temperature rise, and 10% increased precipitation, and found predicted changes which depended on the forest type and current climatic conditions. Model simulations by Aber *et al.* (1995) suggested that neither an unrealistically large (6 °C) temperature warming, a 15% decrease in mean annual precipitation, nor a doubling of atmospheric CO_2 had substantial consistent effects on NPP of temperate forests in the northeastern US. However, the occurrence of all three changes simultaneously resulted in model predictions of increased productivity because increased water use efficiency associated with elevated atmospheric CO_2 would enable sustained higher canopy photosynthesis during the summer and an extended growing season without development of severe soil water deficits. For the same region of the northeastern US, model predictions of NPP responses to ambient ozone pollution levels that occurred from 1987 to 1992 (Ollinger *et al.*, 1997) suggested that forest productivity is currently reduced by 5–10% annually across the region.

Hudson *et al.* (1994) developed a linked terrestrial/atmospheric global carbon cycling model to investigate global carbon pools, and to aid in illuminating discrepancies in the measured global carbon balance. They report concurrence with the notion of temperate forests being largely nitrogen limited, and identify increased atmospheric nitrogen deposition, originating from fossil fuel combustion and other anthropogenic sources, as contributing to large increases in carbon immobilization in terrestrial vegetation.

Using the BIOME3 model, simulations suggest that temperate deciduous forests in Europe may substantially increase in productivity under combined global change scenarios during the next 80 years (Parry, 2000). Moreover, the optimal type shifts from the more northern mixed forest to temperate deciduous forest in roughly one-third of the European temperate zone under this simulation.

SUMMARY

Temperate deciduous forests are important to regional economies as a source of wood, they provide many regional ecological services, and they play a major role in the global carbon cycle. As illustrated above, we can characterize temperate deciduous forest changes, and are beginning to understand the collective impacts of a variety of global change factors on the processes that combine to control composition and ecosystem function in these forests. As scientists, we still face enormous challenges in understanding those complex interactions and in measuring forest patterns and processes at relevant regional to global scales. In this century, temperate forests will continue to play an expanding role as demands for fiber, ecological services (such as water supply and stream water quality, carbon storage, and others), other ecological benefits (wildlife, other biodiversity) and other human values (recreation, tourism, aesthetics, spirituality) increase with burgeoning global populations. Given that temperate deciduous forests are located in regions that are highly developed and already highly disturbed by human use, their future extent, composition, health, and productivity may well signal whether human civilization at high population density is compatible with the sustainability of ecological and economic health and productivity in any ecosystem type or biome.

REFERENCES

Aber, J D, Nadelhoffer, K J, Steudler, P, and Melillo, J M (1989) Nitrogen Saturation in Northern Forest Ecosystems, *BioScicience*, **39**, 378–386.

Aber, J D, Ollinger, S V, Federer, C A, Reich, P B, Goulden, M, Kicklighter, D W, Melillo, J M, Lathrop, R G, and Ellis, J M (1995) Predicting the Effects of Climate Change on Water Yield and Forest Production in the Northeastern United States, *Clim. Res.*, **5**, 207–222.

Abrams, M D (1998) The Red Maple Paradox, *BioScience*, **48**, 355–364.

Abrams, M D (1992) Fire and the Development of Oak Forests, *BioScience*, **42**, 346–353.

Crow, T R (1988) Reproductive Mode and Mechanisms for Self-replacement of Northern Red Oak (Quercus rubra) – a Review, *For. Sci.*, **34**, 19–40.

Goulden, M L, Munger, J W, Fan, S M, Daube, B C, and Wofsy, S C (1996) Exchange of Carbon Dioxide by a Deciduous Forest: Response to Interannual Climate Variability, *Science*, **271**, 1576–1578.

Grimm, E C (1984) Fire and Other Factors Controlling the Big Woods Vegetation of Minnesota in the Mid-nineteenth Century, *Ecol. Monogr.*, **54**, 291–311.

Houghton, R A (1995) Changes in the Storage of Terrestrial Carbon Since 1850, in *Soils and Global Change*, eds R Lal, J Kimble, E Levine, and B A Stewart, CRC Press, Inc., Lewis Publishers.

Hudson, R J M, Gherini, S A, and Goldstein, R A (1994) Modeling the Global Carbon Cycle: Nitrogen Fertilization of the Terrestrial Biosphere and the Missing CO_2 Sink, *Global Biogeochem. Cycles*, **8**, 307–333.

Kauppi, P, Mielikäinen, K, and Kuusela, K (1992) Biomass and Carbon Budget of European Forests, 1971 to 1990, *Science*, **256**, 70–74.

Kruger, E L and Reich, P B (1997) Responses of Hardwood Regeneration to Fire in Mesic Forest Openings. I. Post-fire Community Dynamics, *Can. J. For. Res.*, **27**, 1822–1831.

Magill, A H, Aber, J D, Hendricks, J J, Bowden, R D, Melillo, J M, and Steudler, P A (1997) Biogeochemical Response of Forest Ecosystems to Simulated Chronic Nitrogen Deposition, *Ecol. Appl.*, **7**, 402–415.

Melillo, J M, Kicklighter, D W, McGuire, A D, Moore, B, Vorosmarlty, C J, and Grace, A L (1993) Global Climate Change and Terrestrial Net Primary Production, *Nature*, **363**, 234–240.

Ollinger, S V, Aber, J D, and Reich, P B (1997) Simulating Ozone Effects on Forest Productivity: Interactions Among Leaf-, Canopy-, and Stand-level Processes, *Ecol. Appl.*, **7**, 1237–1251.

Parry, M L, ed (2000) *Assessment of Potential Effects and Adaptations for Climate Change in Europe*: the Europe ACACIA Project, Jackson Environment Institute, University of East Anglia, Norwich, 1–320.

Pastor, J, Dewey, B, Naiman, R J, McInnes, P F, and Cohen, Y (1993) Moose Browsing and Soil Fertility in the Boreal Forests of the Isle Royale National Park, *Ecology*, **74**, 467–480.

Peterson, D and Reich, P B (2001) Fire Frequency and Stand Dynamics in an Oak Savanna–Woodland Ecosystem, *Ecol. Appl.*, (in press).

Pyne, S J (1982) *Fire in America: a Cultural History of Wildland and Rural Fire*, Princeton University Press, Princeton, NJ.

Reich, P B (1987) Quantifying Plant Response to Ozone: a Unifying Theory, *Tree Physiol.*, **3**, 63–91.

Reich, P B, Abrams, M D, Ellsworth, D S, Kruger, E L, and Tabone, T J (1990) Fire Affects Ecophysiology and Community Dynamics of Central Wisconsin Oak Forest Regeneration, *Ecology*, **71**, 2179–2190.

Reich, P B and Bolstad, P (2001) Temperate Forests: Evergreen and Deciduous, in *Global Terrestrial Productivity*, eds H Mooney, J Roy, and B Saugier, Academic Press (in press).

Running, S W and Nemani, R R (1991) Regional Hydrologic and Carbon Balance Responses of Forests Resulting from Potential Climate Change, *Clim. Change*, **19**, 349–368.

Schimel, D S (1995) Terrestrial Ecosystems and the Carbon Cycle, *Global Change Biol.*, **1**, 77–91.

Townsend, A R, Braswell, B H, Holland, E A, and Penner, J E (1996) Spatial and Temporal Patterns in Terrestrial Carbon Storage due to Deposition of Fossil Fuel Nitrogen, *Ecol. Appl.*, **6**, 806–814.

Williams, M (1989) *Americans and Their Forests*, Cambridge University Press, Cambridge.

Temperate Grasslands

José M Paruelo

Universidad de Buenos Aires, Buenos Aires, Argentina

Temperate grasslands have been the biome most affected by one of the main dimensions of the global change: land cover modification. The replacement of grasslands by agricultural fields have local, regional, and global consequences, including loss of soil fertility, soil erosion, reduction of biological diversity, hydrological changes, climate alteration and modification of atmospheric composition. Field and laboratory experiments are beginning to show a positive response, both in total biomass production and water use efficiency (WUE) for the dominant species of temperate grasslands to the increase of atmospheric carbon dioxide (CO_2). Simulation analyses have indicated that climate change by itself has increased net primary production in most of the grassland regions of the world and has reduced soil carbon stocks everywhere. The combined effect of climate change and elevated CO_2 has increased net primary production in all the sites studied and has reduced carbon losses by half.

Temperate grasslands are an important biome in Eurasia (Paleartic realm, 7.5×10^6 km^2), North America (Neartic realm, 2.7×10^6 km^2) and southern South America (Neotropical realm, 1.1×10^6 km^2), also occupying small areas in Africa, Australia and New Zealand. At least three physiognomic types are considered grasslands in temperate regions:

meadows, prairies, and steppes. Species of the Poaceae family are dominant or co-dominant in these physiognomic types. Grasslands occur in between forest and deserts. The boundaries among these biomes are not clear, and transitional areas develop among them; scattered trees or shrubs are then a common feature in grassland areas. The relative importance of the woody components in grassland is often the consequence of the disturbance regime.

Grasslands have played an important role in human history because the centers of origin of the most important cultivated plants (wheat, barley, peas, etc.) and livestock (goats, sheep, cattle) are located in them. Humans dramatically modified grassland areas and the sub-humid portion of this biome has been almost completely transformed by agriculture.

GENERAL CHARACTERISTICS OF TEMPERATE GRASSLANDS

Climate and Soils

Grasslands occur over a broad range of temperatures. Mean annual temperature (MAT) of grassland areas range between $-3\,^{\circ}C$ in the Mongolian and Russian steppes and $20\,^{\circ}C$ in the tallgrass prairies of Texas (USA) and the desert grasslands of Northern Mexico. A large proportion of grasslands has a continental climate, showing much larger seasonal temperature differences than areas under maritime influence. The range of mean annual precipitation (MAP) where grasslands can be found varies from 150–1500 mm.

Grasslands are not evenly distributed in the climatic space defined by the ranges of MAT and MAP presented above. Grasses are not dominant where temperatures are high and precipitation low (close to the $20\,^{\circ}C$ and the 150 mm corner of a MAT–MAP graph (see Figure 1)). At this extreme, grasslands become replaced by shrublands or deserts. At the coldest and wettest extreme, forests are the dominant biome. The distribution of grasslands in the climatic space is related to the magnitude of the water deficit. Grasslands show a ratio of mean annual precipitation and potential evapotranspiration (PET) lower than 1. The MAP/PET ratio is close to 0.3 for dry steppes and to 0.6 for sub-humid prairies (*see* **PET (Potential Evapotranspiration)**, Volume 2).

Grassland soils are among the most productive under agricultural use. High productivity derives mainly from the high content of organic matter. Mollisols is the dominant soil order in this biome.

Biotic Characteristics and their Environmental Controls

An important structural characteristic of temperate grassland is that most of the carbon is stored below ground, in soil organic matter and plant roots. This large accumulation of carbon results from low decomposition rates relative to net primary production. The total amount of soil organic carbon decreases with MAT and increases with MAP (Burke *et al.*, 1989). Fine textured soils retain more organic matter than coarse textured soils (Burke *et al.*, 1989).

Despite the dominance of just one family, Poacea, grasslands show marked differences in species composition. Over continental scales, a structural comparison of grasslands based on species composition is precluded by the differences in floras among the different biogeographical realms. Such comparison is possible, though, at the plant functional level (PFT) (*see* **Plant Functional Types**, Volume 2). Five main PFTs can be identified in grassland areas: C_3 grasses, C_4 grasses, shrubs, forbs and succulents. The first three account for most of the biomass and show a clear pattern across environmental gradients. The relative abundance of C_4 grasses increases with MAP, MAT, and the proportion of precipitation falling in summer. C_3 grasses are more abundant as MAT decreases and in areas with a high proportion of precipitation falling in winter. The relative abundance of shrubs increases in dry areas with most of the precipitation falling in winter.

Above-ground net primary production (ANPP) is a key attribute of grassland ecosystems and the main entrance of energy to the ecosystem. Characterization of the climatic controls of ANPP is essential to understanding the impact of global changes on this biome. A positive, linear relationship between ANPP and MAP has been documented for many grasslands around the world (McNaughton *et al.*, 1993). The slope of the relationship between ANPP and MAP in grassland ecosystems ranges between 0.48 and $0.64\,g\,C\,m^{-2}\,mm^{-1}$. Epstein *et al.* (1997) showed a negative relationship between ANPP and MAT when MAP was held constant. Soil texture, through its effect on soil water-holding capacity, also showed a significant relationship with ANPP (Sala *et al.*, 1988; Epstein *et al.*, 1997). Below-ground net primary production is equivalent or slightly greater than ANPP. Annual net primary production (below and above ground) may vary then, from less than $100\,g\,m^{-2}$ in a semi-arid shortgrass steppe to almost $1000\,g\,m^{-2}$ in a tallgrass prairie (Paruelo *et al.*, 1999). Compared to other biomes, grasslands' above-ground biomass is low. Shrublands may have 10 times more biomass than grasslands and evergreen forest up to 100 times more (Chapin, 1993). Differences in ANPP among biomes, however, are not so extreme as in the case of biomass (Chapin, 1993). As a consequence, the ratio ANPP/biomass is higher in grasslands than in any other biome.

McNaughton *et al.* (1993) showed that biomass and consumption of wild herbivores were positively and exponentially related to ANPP. The exponential form of these

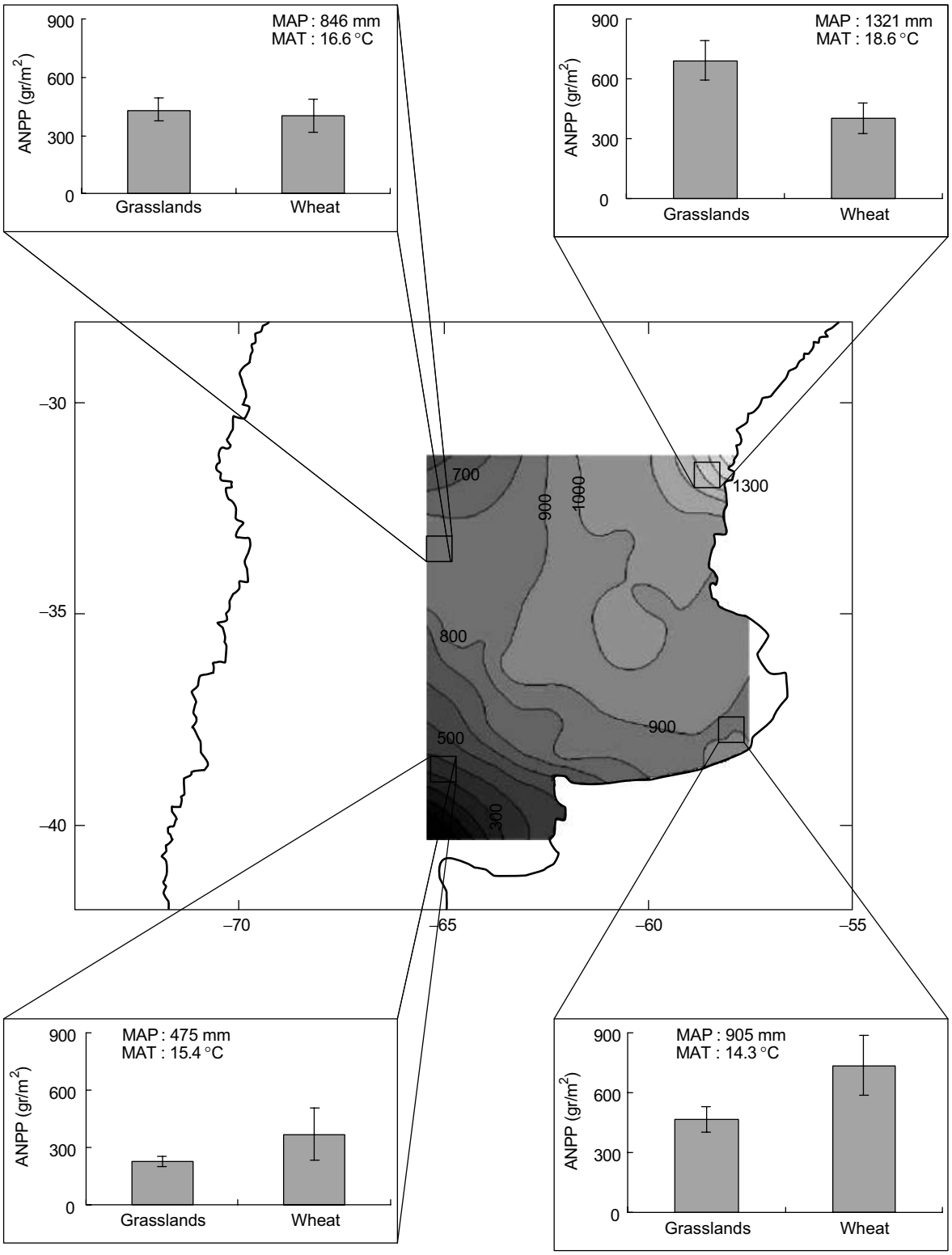


Figure 1 ANPP for wheat and grasslands sites located in the climatic extremes of the Pampas grasslands of Argentina. For wheat fields ANPP was estimated from grain yields (Argentine agricultural census) and harvest index. Wheat ANPPs are the average of 10 years. Small bars showed the temporal standard deviation. Grassland ANPPs are the average of estimates derived from four different models for the relationship ANPP- MAP (Sala *et al*., 1988; McNaughton *et al*., 1993; Paruelo *et al*., 1999 (two models)). Small bars corresponded to the standard deviation of the estimates derived from the different models. Isopleths correspond to mean annual precipitation

relationships suggests that the proportion of the ANPP consumed by herbivores increases with ANPP. The efficiency of energy transfer between trophic levels (animal production/net primary production) is higher in grasslands than in forests.

The disturbance regime affects carbon gains in grasslands. Fire reduces ANPP in the driest portion of a precipitation gradient (Oesterheld *et al.*, 1999). In subhumid grasslands, fire generally increases ANPP, except in dry years (Knapp *et al.*, 1998). In these grasslands, fire removes large amounts of above-ground dead material, increasing light and soil nutrient availability (Knapp and Seastedt, 1986). In dry years, this effect is less important than the reduction in evapotranspiration promoted by the detritus layer (Knapp *et al.*, 1998). Grazing may increase or reduce ANPP depending on the evolutionary history of grazing, MAP, and levels of consumption (Milchunas and Lauenroth, 1993). Over a broad range of grasslands, Oesterheld *et al.* (1999) found that, on average, grazing by large vertebrates slightly reduced ANPP.

Grassland biogeochemistry is largely controlled by precipitation, temperature and soil texture. Decomposition rates are more affected than ANPP by temperature. Given the same amount of precipitation, the consequence of the effect of temperature on decomposition is a higher accumulation of organic carbon in cold than in warm areas. Species composition, through litter quality, has a significant influence on decomposition rates and nutrient immobilization.

THE IMPACT OF GLOBAL CHANGE ON TEMPERATE GRASSLANDS

Land Use

Land cover changes associated with agriculture and grazing by domestic herbivores have had an enormous impact on the structure and functioning of grasslands. Agricultural practices have reduced soil carbon stocks by 35% in the US Central Grasslands during the last 50 years (Burke *et al.*, 1989). Based on carbon isotopic analyses, Wilson (1978) estimated that pioneer agriculture (1860–1890) in temperate areas of Australia, New Zealand, North and South America contributed one and a half times the amount of carbon dioxide (CO_2) produced by all the fossil fuels burnt before 1950. This reduction of soil carbon stocks results from an increase in decomposition rates due to plowing and a reduction of carbon inputs to the soil due to the low below-ground biomass of annual crops and to exports of above-ground biomass.

Land-use pattern has a critical influence on mesoscale atmospheric processes, and hence on local climate (Pielke *et al.*, 1997). Mosier *et al.* (1991) reported significant changes in trace gas fluxes associated with cultivation of a semi-arid grassland: agriculture increases the release of methane (CH_4) and nitrous oxide (N_2O), two greenhouse gases. These additions of CO_2 and other trace gases to the atmosphere may represent significant feedbacks to global climate change.

It is not clear to what extent agricultural practices modify total carbon gains. In the Pampas of Argentina, agriculture may increase or decrease above-ground net primary production relative to the native grasslands depending on environmental conditions (Figure 1). The difference in primary production between grasslands and wheat fields is higher in the wettest portion of the Pampas. In warm areas, grasslands are more productive than wheat fields. The opposite happens in the coolest part of the region.

Total carbon gains should be lower for agricultural fields than for native grasslands because below-ground production is one order of magnitude lower for crops than for grasslands. Land-use may modify not only total carbon gains but also the seasonal dynamics of leaf area index and intercepted radiation. Figure 2 shows the idealized course of the fraction of radiation intercepted by the canopy for different land covers. Curves were drawn from data provided by spectral indices derived from meteorological satellites (Advanced Very High Resolution Radiometer – National Oceanographic and Atmospheric Agency (NOAA/AVHRR)). Land use modifies both the intra-annual range of the fraction of intercepted radiation and the timing of maximum interception. These phenological changes modify the partitioning between latent and sensible heat and, consequently, they may affect the dynamics of the boundary layer of the atmosphere.

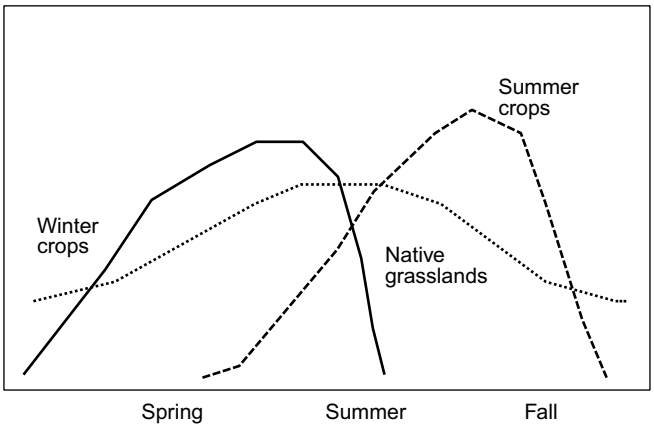


Figure 2 Seasonal dynamics of the fraction of radiation intercepted by the canopy from three land cover categories in the Argentine Pampas. Intercepted radiation is an estimator of ANPP. Curves were derived from normalized difference vegetation index data from NOAA/AVHRR satellites

Land use has increased the rate of species extinction not only by replacing natural ecosystems, but also by changing the disturbance regime. Leach and Givnish (1996) showed that 8–60% of the original plant species of remnant prairie sites were lost after 50 years due to landscape fragmentation or fire suppression. Changes in species diversity and composition have direct effects on the functioning of grasslands, affecting their productivity and stability.

The introduction of domestic herbivores promotes structural and functional changes in grasslands. A common effect of overgrazing in grasslands is shrub encroachment (Archer, 1994). The increase in shrub cover in grass steppes of Patagonia due to overgrazing (León and Aguiar, 1985) altered water dynamics, increased albedo, and reduced primary production and herbivore biomass (Aguiar *et al.*, 1996). Bryant *et al.* (1990) and Nasrallah *et al.* (1994) reported changes in surface temperature associated with overgrazing in southwestern USA.

Atmospheric Composition

Early experiments and photosynthetic theory suggested a differential response of C_3 and C_4 grasses to changes in atmospheric concentration of CO_2. C_3 and C_4 plants differ in their metabolic pathways of CO_2 fixation. In C_3 plants the first product of the photosynthetic CO_2 fixation is a phosphorylated three-carbon compound. In C_4 plants the first product is a four-carbon acid. C_3 and C_4 plants differ also in their leaf anatomy. Two different groups of photosynthetic cells are found in C_4 plants: mesophyllic and bundle sheaths cells. The latter surround the leaf vascular bundles and separate them from the mesophyllic cells. C_4 grasses present a CO_2 concentration mechanism in the bundle sheath cells that increases the effective concentration of CO_2 at the carboxylation site. Such a CO_2 pumping mechanism would saturate photosynthesis at current CO_2 levels. Therefore, it has been suggested that this plant functional group would not show a significant growth response to increases in atmospheric CO_2 concentrations. However, a meta-analysis by Wand *et al.* (1999) indicates that the response of wild C_3 and C_4 grasses to elevated CO_2 concentrations was surprisingly similar (*see also* **C_3 and C_4 Photosynthesis**, Volume 2).

At the leaf level, an increase in atmospheric CO_2 reduces stomatal conductance and increases net photosynthesis. A consequence of these changes is an increase in water use efficiency (WUE). Specific leaf area and nitrogen concentration are, in general, lower under elevated CO_2 concentrations. At the individual level these changes result in an increase in both above and below-ground biomass and in the number of tillers (Wand *et al.*, 1999). The mentioned analysis showed an overall increase in total biomass of 44% in C_3 species and of 33% in C_4 species.

Studies using open-top chambers allow one to explore changes at the population, community, and ecosystem level. These changes result not only from the effect of elevated CO_2 on plant production but also from major changes in soil water dynamics associated with changes in water use efficiency. Jackson *et al.* (1994) working in Mediterranean grasslands in California showed increases in density, seed production, and survival of the dominant grass *Avena barbata* under elevated CO_2. Changes in phenology associated with a delayed senescence have been reported in C_4 tallgrasses (Owensby *et al.*, 1997). In the tallgrass prairie of eastern Kansas (USA) an eight year long open-top chamber experiment provided no evidence for a replacement of the dominant C_4 grasses by C_3 grasses or forbs under elevated CO_2 concentrations (Owensby *et al.*, 1993). In these grasslands, the balance between C_3 and C_4 grasses seems to be more sensitive to the main disturbances (fire and grazing) than to CO_2 concentration. Elevated CO_2 decreased dominance in sub-humid temperate grasslands (Potvin and Vasseur, 1997) who showed a higher effect of elevated CO_2 on early successional species growth, which prevented the dominance of late successional ones.

Combining field data and simulation models, Jackson *et al.* (1998) showed that the direct effect of CO_2 on water use efficiency and total biomass production on Mediterranean grasslands affected water dynamics at the ecosystem level. Under elevated CO_2, plant transpiration was lower. Low transpiration at the beginning of the growing season increased soil water content and deep percolation. Studies based both on field experiments and on model simulations suggest that the observed increase in WUE at the ecosystem level would compensate for increases in potential evapotranspiration associated with climatic changes driven by greenhouse gases. Biogeochemistry models (BIOME-BGC, Century and Terrestrial Ecosystem Model) simulated net primary production responses to doubled atmospheric CO_2 for grasslands that range from 1–20% increase. *See also* **Biome–BGC Ecosystem Model**, Volume 2; **CO_2 Enrichment: Effects on Ecosystems**, Volume 2; **Plant Competition in an Elevated CO_2 World**, Volume 2.

Climate

Changes in climate are difficult to isolate from changes in CO_2 concentrations. Climate modifications not linked to the increase of greenhouse gases may occur, however, as a result of land-use changes (Nasrallah *et al.*, 1994; Pielke *et al.*, 1997). General circulation models simulate changes in temperature in temperate grasslands ranging between +2.3 and +6.2 for a doubled CO_2 scenario (Houghton *et al.*, 1992). Changes in mean annual precipitation may range from −2.6% up to 17.5%, depending on the GCM and type of grassland considered. For the Central Grasslands of the USA, an increase of 1° Celsius in mean annual temperature represents an increase in potential evapotranspiration of 50 mm.

Two main approaches have been used to explore the effect of climate on grasslands: warming experiments and simulations using climate change scenarios. Experimentally-heated plots in montane meadows showed significant differences in the timing and magnitude of net carbon fluxes (Saleska *et al.*, 1999). The reduction of primary production resulted from a combined effect of a reduction in soil moisture and a shift in plant community composition. Alward *et al.* (1999) showed that the observed increase in nocturnal minimum temperatures in a semi-arid steppe was correlated with a reduction in the net primary production of the dominant grass species, and with increased abundance and production of exotic species and C_3 forbs.

Model simulations indicated that climate change, by itself, increased net primary production in most of the grassland regions of the world and reduced soil carbon stocks everywhere. The combined effect of climate change and elevated CO_2 increased net primary production in all the sites studied and reduced carbon losses by half (Parton *et al.*, 1995).

REFERENCES

Aguiar, M R, Paruelo, J M, Sala, O E, and Lauenroth, W K (1996) Ecosystem Consequences of Plant Functional Types Changes in a Semiarid Grassland, *J. Vegetation Sci.*, **7**, 381–390.

Alward, R, Detling, J, and Milchunas, D (1999) Grassland Vegetation Changes and Nocturnal Global Warming, *Science*, **283**, 229–231.

Burke, I C, Yonker, C M, Parton, W J, Cole, C V, Flach, K, and Schimel, D S (1989) Texture, Climate, and Cultivation Effects on Soil Organic Content in US Grassland Soils, *Soil Sci. Soc. Am. J.*, **53**, 800–805.

Chapin, F S (1993) Functional Role of Growth forms in Ecosystem and Global Processes, in *Scaling Physiological Processes*, eds J R Ehleringer and C B Field, Academic Press, San Diego, CA, 287–312.

Epstein, H E, Lauenroth, W K, and Burke, I C (1997) Effects of Temperature and Soil Texture on ANPP in the US Great Plains, *Ecology*, **78**, 2628–2631.

Jackson, R B, Sala, O E, Paruelo, J M, and Mooney, H A (1998) Ecosystem Water Fluxes for Two Grasslands in Elevated CO_2: a Modeling Analysis, *Oecologia*, **113**, 537–546.

Jackson, R B, Sala, O E, Field, C B, and Mooney, H A (1994) CO_2 Alters Water use, Carbon Gain, and Yield from the Dominant Species in a Natural Grassland, *Oecologia*, **98**, 257–262.

Knapp, A K, Conard, S L, and Blair, J M (1998) Determinants of Soil CO_2 Flux From a Sub-humid Grassland: Effect of Fire and Fire History, *Ecol. Appl.*, **8**, 760–770.

Knapp, A K and Seastedt, T R (1986) Detritus Accumulation Limits Productivity of Tallgrass Prairie, *Bioscience*, **36**, 662–668.

Leach, M K and Givnish, T J (1996) Ecological Determinants of Species Loss in Remnant Prairies, *Science*, **273**, 1555–1558.

McNaughton, S J, Sala, O E, and Oesterheld, M (1993) Comparative Ecology of African and South American Arid to Subhumid Ecosystems, in *Biological Relationships between Africa and South America*, ed P Goldbaltt, Yale University Press, New Haven, CT, 548–567.

Milchunas, D G and Lauenroth, W K (1993) Quantitative Effects of Grazing on Vegetation and Soils Over a Global Range of Environments, *Ecol. Monogr.*, **63**, 327–366.

Mosier, A, Schimel, D, Valentine, D, Bronson, K, and Parton, W (1991) Methane and Nitrous Oxide Fluxes in Native, Fertilized and Cultivated Grasslands, *Nature*, **350**, 330–332.

Nasrallah, H A and Balling, R C (1994) The Effect of Overgrazing on Historical Temperature Trends, *Agri. Forest Meteorol.*, **71**, 425–430.

Oesterheld, M, Loreti, J, Semmartin, M, and Paruelo, J M (1999) Grazing, Fire, and Climate Effects on Primary Productivity of Grasslands and Savannas, in *Ecosystems of Disturbed Ground*, ed L R Walker, Elsevier Science, Oxford.

Owensby, C E, Coyne, P I, Ham, J M, Auen, L M, and Knapp, A K (1993) Biomass Production in a Tallgrass Prairie Ecosystem Exposed to Ambient and Elevated CO_2, *Ecol. Appl.*, **3**, 644–653.

Owensby, C E, Ham, J M, Knapp, A K, and Bremer, D (1997) Water Vapor Fluxes and their Impact Under Elevated CO_2 in a C_4- Tallgrass Prairie, *Global Change Biol.*, **3**, 189–195.

Parton, W J, Scurlock, J M O, Ojima, D S, Schimel, D S, and Hall, D O (1995) Impact of Climate Change on Grassland Production and Soil Carbon Worldwide, *Global Change Biol.*, **1**, 13–22.

Paruelo, J M, Lauenroth, W K, Burke, I C, and Sala, O E (1999) Grassland Precipitation use Efficiency Varies Across a Resource Gradient, *Ecosystems*, **2**, 64–68.

Pielke, R A, Lee, T J, Copeland, J H, Eastman, J L, Ziegler, C L, and Finley, C A (1997) Use of USGS-provided Data to Improve Weather and Climate Simulations, *Ecol. Appl.*, **7**, 3–21.

Potvin, C and Vasseur, L (1997) Long-term CO_2 Enrichment of a Pasture Community: Species Richness, Dominance, and Succession, *Ecology*, **78**, 666–677.

Sala, O E, Parton, W J, Joyce, L A, and Lauenroth, W K (1988) Primary Production of the Central Grassland Region of the United States, *Ecology*, **69**, 40–45.

Saleska, S R, Harte, J, and Torn, M S (1999) The Effect of Experimental Ecosystem Warming on CO_2 Fluxes in a Montane Meadow, *Global Change Biol.*, **5**, 125–141.

Walter, H (1939) Grassland, Savanne und Busch der arideren Teile Afrikas in Ihrer Okologischen Bedingtheit, *Jb. Wiss Bot.*, **87**, 750–860.

Wand, S, Midgley, G, and Jones, M (1999) Responses of Wild C_4 and C_3 Grass (Poaceae) Species to Elevated Atmospheric CO_2 Concentration: a Meta-analytic Test of Current Theories and Perceptions, *Global Change Biol.*, **5**, 723–741.

Wilson, A T (1978) Pioneer Agriculture Explosion and CO_2 Levels in the Atmosphere, *Nature*, **273**, 40–41.

Temperature Increase: Effects on Terrestrial Ecosystems

Lindsey E Rustad[1] and Richard J Norby[2]
[1]USDA Forest Service, Durham, NH, USA
[2]Oak Ridge National Laboratory, Oak Ridge, TN, USA

Temperature is a fundamental factor that affects the rates of most biological and chemical reactions in ecosystems, and thus plays an important role in regulating the flow of energy, water and nutrients within and through ecosystems. Indeed, with the exception perhaps of water, temperature influences a greater number of terrestrial ecosystem processes than any other element of global change, including elevated atmospheric carbon dioxide, atmospheric nitrogen and sulfur deposition, tropospheric ozone, and increased ultraviolet-B radiation. Although short-term responses to changes in temperature of individual ecosystem processes, such as photosynthesis and respiration, have been studied for well over a century, a comprehensive understanding of whole-ecosystem response to global warming remains elusive. This is because of the complex array of direct and indirect responses to changes in temperature, the variable time scale of response of different ecosystem processes and components to changes in temperature, and the importance of initial conditions (Shaver et al.*, 2000).*

These complexities are apparent when we try to predict the effect of warming on ecosystem carbon balances. The net carbon balance of an ecosystem, also known as net ecosystem productivity (NEP), represents the total amount of carbon fixed by an ecosystem minus the total amount of carbon respired or lost by an ecosystem. NEP is calculated as net primary production (NPP, or the total amount of CO_2-C fixed photosynthetically by plants each year minus the total amount of C respired by plants each year) minus heterotrophic respiration (R_h, or the total amount of carbon respired by animals and decomposers each year). Both NPP and R_h generally increase as temperatures increase, with R_h increasing somewhat more in the short term than NPP (Woodwell, 1995). Increases in temperature can, however, also indirectly affect NPP and R_h via changes in soil moisture, nutrient availability (particularly nitrogen and phosphorus), species composition, and interactions with other stress factors in complex ways that are not yet fully understood. Ultimately, a warming-induced change in NEP will depend on the relative sensitivity of NPP and R_h to changes in temperature. Understanding temperature effects on NEP is important because of the potential feedbacks to the climate system: an increase in NEP will remove carbon from the atmosphere, sequestering carbon in plant biomass, and thereby helping to ameliorate further increases in atmospheric carbon dioxide concentrations, whereas a decrease in NEP will exacerbate the already escalating increases in atmospheric carbon dioxide.

GLOBAL WARMING SCENARIOS FOR THE 21ST CENTURY

Recent evidence, including reports of warmer mean ocean temperatures, decreasing thickness of arctic sea ice, retreat of glaciers, earlier ice breakup on lakes and ponds, and analysis of long-term temperature records, suggest that mean global temperature has increased by 0.3–0.6 °C over the last 100 years, with the rate of warming showing the greatest increase over the last 20 years (Karl *et al.*, 2000). Current predictions based on General Circulation Models indicate that, unless greenhouse gas emissions are significantly curtailed, mean global temperature will increase by an additional 1–3.5 °C in the 21st century (Intergovernmental Panel on Climate Change (IPCC), 1996). The spatial and temporal variability in this increase is expected to be large. This is due in part to natural fluctuations in climate that occur on decadal time scales, and in part due to identifiable factors such as the thermal inertia of oceans. Specific predictions associated with climate change are that:

1. surface warming will be greater over the land than the seas in winter;
2. surface warming will be greater at higher latitudes than at the equator in winter;
3. little surface warming will occur over the arctic in summer;
4. nighttime minimum temperatures will increase proportionately more than daytime maxima;
5. there will be an increase in the occurrence of extremely hot days and a decrease in the occurrence of extremely cold days (Intergovernmental Panel on Climate Change (IPCC), 1996) (*see* **Projection of Future Changes in Climate**, Volume 1).

A CONCEPTUAL FRAMEWORK FOR EVALUATING ECOSYSTEM RESPONSE TO GLOBAL WARMING

Direct Effects of Warming on Ecosystems

Temperature directly influences the rates at which a wide range of chemical and biological processes take place. The increase in the reaction rate of biological processes per 10 °C increase in temperature is referred to as the biological Q_{10}, and is typically around two, or a doubling of biological reaction rates for every 10 °C increase in temperature from 0–~35 °C. Above ~35 °C, reaction rates typically decline rapidly due to the thermal denaturation of

proteins and cell membranes. Exceptions are found only in a few thermophilic, sulfur oxidizing bacteria, which can grow at temperatures of 80 °C or more.

All biological processes have a temperature optimum, with reaction rates increasing with increasing temperature from ~0 °C to the temperature optimum, and then declining rapidly with further increases in temperature. Thus, in order to understand and predict the biological response to an increase in temperature, it is important to know the temperature optima for a particular process. A 1–2 °C increase in temperature may increase rates for a specific process when temperatures are below the temperature optimum, but the same 1–2 °C increase in temperature could decrease rates if the temperature optimum for that process is exceeded. This is further complicated by the fact that biological organisms are typically adapted to regional climates, and thus temperature optimal vary regionally for the same process. For example, arctic and alpine plants typically have a lower optimum temperature (15–20 °C) for shoot and root growth than temperate species (25–30 °C). Furthermore, plants can acclimate to changing temperature by shifting their temperature optimum, as discussed below.

Temperature has profound direct influences on plant physiology, growth, and development through all life stages. Critical processes that are sensitive to temperature include seed germination, cell division and expansion, root and shoot growth, carbohydrate translocation, plant development and phenology, and flowering. Photosynthesis is sensitive to temperature primarily through the membrane-bound photochemical reactions, but the carboxylation efficiency of the primary photosynthetic enzyme is also sensitive to temperature. Respiration typically increases exponentially with increasing temperature. Increases in temperature beyond the thermal tolerance limit can cause irreversible damage to the photosynthetic membranes. Mitochondria (organelles found in plant cells responsible for aerobic respiration), in contrast, are more stable, so respiration continues to increase as photosynthesis declines (Weiss and Berry, 1988). Volatilization of hydrocarbons from plants increases with temperature, representing a loss of carbon. The net effect of the many different influences on carbon gains and losses can be difficult to predict, but generally NPP increases with modest increases in temperature.

At higher latitudes and higher elevation, plant growth is constrained to a large extent by low winter temperatures and the length of the growing season. Global warming can be expected to lengthen the growing season in temperate zones. Indeed, several lines of evidence suggest that the start of the growing season in the Northern Hemisphere has already advanced by seven days since 1958 (Myneni *et al.*, 1997). The length of the growing season is an important determinant of annual NPP, accounting for much of the inter-annual variation in NPP for a given site, as well as for NPP differences between sites at different latitudes. Phenological events such as bud break in spring and senescence and dormancy in fall are controlled by a complex suite of environmental signals including not only temperature but also photoperiod, which is not subject to anthropogenic perturbation. Climatic warming may create disjunction between the temperature and photoperiodic influences on phenology, and surprises such as insufficient winter chilling or untimely frosts could sometimes decrease NPP (Norby *et al.*, 2000) (*see* **Phenology**, Volume 2).

Temperature also directly influences processes mediated by soil micro- and macro-organisms. Soil micro- and macro-organisms play a dual role in terrestrial ecosystems. (1) They serve as a sink for carbon and nutrients during periods of population expansions, and a source of carbon and nutrients during periods when populations are declining; and (2) they are the agents of many biochemical processes fundamental to nutrient cycling. These processes include *decomposition* of plant and animal residues (thereby releasing carbon back to the atmosphere largely in the form of carbon dioxide and releasing nutrients back to soils and surface waters), *ammonification* (the transformation of organic N to NH_4), *nitrification* (the transformation of NH_4 to NO_3), *denitrification* (the transformation of NO_3 to NO and N_2O), and *methane reduction and oxidation*. In general, these processes speed up with increasing temperature. As with net ecosystem carbon balance, temperature effects on soil organic matter (SOM) will reflect the balance between increased litter inputs (as a result of increased NPP) and increased rates of litter decomposition. In general, SOM declines with increasing temperature, as evidenced by the documented global trends of declining SOM with decreases in latitude or altitude. SOM is important because of its contribution to soil nutrient status, soil water holding capacity, and soil texture.

Indirect Effects of Warming on Ecosystems

Temperature can also influence ecosystems and ecosystem processes through multiple, indirect pathways, particularly those involving temperature-induced changes in water, nutrients, species composition, and the activity of herbivores and pathogens. For example, temperature and water are closely linked, and many of the effects of temperature on ecosystems occur through changes in evapotranspiration and soil moisture. Although the water-holding capacity of the atmosphere increases with temperature, the vapor pressure gradient from leaf or soil to air also increases. Hence, plant transpiration and direct evaporation from soil are higher at warmer temperatures. Increased drought is a prominent prediction of climate warming models (Intergovernmental Panel on Climate Change (IPCC), 1996), and drought is an important determinant of NPP. Soil moisture strongly controls microbial processes, including litter

decomposition and nutrient cycling, and is an important mediator of the diffusion of ions to root surfaces. The indirect effects of temperature increases via soil moisture deficits on plant and soil metabolism are opposite in sign from many of the direct effects of temperature, and they could be the dominant signal in many ecosystems. For example, tree ring analysis of Alaskan white spruce (Barber *et al.*, 2000) suggests that despite the apparent increase in the length of the growing season at high northern latitudes, radial growth has declined in response to increasing temperature and temperature-induced drought stress.

Temperature-induced changes in nutrient availability can also have a large, albeit indirect, impact on ecosystems. Increasing temperature can increase nutrient availability by: (1) increased rates of litter decomposition, and thereby more rapid release of nutrients from decomposing organic matter; (2) increased rates of nitrogen mineralization and nitrification; and (3) increased rates of chemical diffusion to the root surface.

Direct or indirect effects of temperature on the populations and/or activity of herbivores and pathogens could have profound effects on ecosystem structure and function by changing the amount and timing of detrital inputs to the soil, and by dramatically altering species competitive interactions. Changes in species composition can, in turn, affect a multitude of other ecosystem properties, such as plant growth rates, litter quantity and quality (and consequently microbial dynamics and litter decomposition), canopy structure and architecture, shading, and rooting depth, to name a few.

Variable Time Scale and Direction of Response

The time scale of ecosystem response to warming ranges from the millisecond to the millennium. At the very short time scale (seconds to days), warming directly influences the rates of chemical and biological reactions, as discussed above. At the month-to-seasonal time scale, warming can influence ecosystem carbon and nutrient allocation, nutrient mobilization and immobilization, and microbial species composition. At the annual-to-decadal time scale, warming can influence plant species competitive interactions, litterfall quantity and quality, and herbivore and pathogen dynamics. Over time scales ranging from several decades to several centuries, sustained increases in temperature can result in shifts in species composition (including migration of species to higher latitudes and higher elevations, and invasion of new, exotic species), changes in soil and SOM development, altered susceptibility to herbivores and pathogens, and changes in ecosystem disturbance regimes, particularly susceptibility to fire, as ecosystem fuel loads increase or decrease.

Changes in both the direction and the magnitude of response over time are possible, representing a changing sequence of controls. For example, plant growth may initially increase with warming due to higher rates of photosynthesis and longer growing seasons. Over time, however, this increased growth may not be sustained due to depletion of available soil nutrients, and growth rates may return to previous or even below previous levels. Similarly, the short-term (days–months) rates of microbial respiration typically increase with increasing temperature. This can continue as long as labile (i.e., chemically unstable) carbon sources are available. As these labile carbon sources are depleted, rates of microbial respiration may decline, despite higher temperatures. Litter decomposition also generally increases with increasing temperature if moisture remains adequate. However, changes in plant growth rates or changes in species composition can result in changes in the quality of litter over time. If litter quality declines, rates of litter decomposition may decline despite increased temperature.

Finally, the response of individual species to sustained increases in temperature may change over time as plants and soil organisms become acclimated to changing conditions. Because of this acclimation, much of our understanding of ecosystem response to changing temperature comes from time scales that may not be completely relevant to global change analyses. Over the short-term (hours), for example, autotrophic respiration responds to temperature by the Arrhenius function, but as a plant acclimates to a higher temperature, its respiration rate is reduced at a given temperature. Global change predictions that assume a Q_{10} of two for plant respiration will almost certainly overestimate the increase in respiration resulting from increases in temperature occurring over decades. Similarly, photosynthesis acclimates to temperature, but here the analysis is more complicated. Patterns of photosynthetic acclimation to temperature include a shift in the optimum temperature or an adjustment in the rate at all temperatures (Gunderson *et al.*, 1999). Membrane-bound processes such as ion uptake acclimate to changes in temperature through changes in the lipid composition of the membranes, so short-term responses of ion uptake to temperature probably are not relevant to changes in mean annual temperature occurring over decades.

Adjustments in soil processes to temperature changes will have different characteristics from adjustment processes in plants because of the rapid turnover of microbial populations. There have been suggestions that R_h (i.e., animal and decomposer respiration) acclimates to increased temperature and that Q_{10} representations are not appropriate (Giardina and Ryan, 2000). However, it can be difficult to separate the effects of temperature and soil moisture, which also has a large effect on soil microbial activity. On a global scale, fine root turnover (an important source of carbon to soil) is higher at higher temperatures, but this pattern is not apparent at specific sites subjected to warming (Gill and Jackson, 2000). As with other plant growth processes, there is evidence that root production acclimates to higher

temperature (Fitter *et al.*, 1999), and other factors such as stimulation of aboveground production or indirect effects on nutrient availability are likely to be more important controls on root production than temperature alone.

Since short-term temperature manipulations may provide misleading answers to longer-term responses, it is useful to study patterns of temperature response across the landscape, analyzing the relationship between fully acclimated populations of plants and the prevailing temperature regime. Such studies generally reveal broad patterns of NPP versus mean annual temperature, which might provide a basis for predictions of response to climatic change; an approach known as substituting space for time. But this approach also requires caution. Much of the basis for increased NPP at higher temperature is attributable not to a primary response of the instantaneous balance between photosynthesis and respiration, but to the increased length of the growing season at lower latitudes under current climate, and the length of the growing season is controlled in part by day length; one global factor that will not be affected by human activity. Another confounding factor with application of natural temperature gradients as an analog for climatic warming is that plant populations adapt to the prevailing climate via genetic change, a process that is not relevant to long-lived organisms (e.g., trees) over the several decade period of relevance to policy makers.

Importance of Initial Conditions

Overlaid on the aforementioned complexities in determining the direction and magnitude of ecosystem response to warming is the importance of initial conditions. For example, a $1-2\,^{\circ}C$ increase in temperature could increase productivity and plant growth in a moist, temperate forest ecosystem, but the same increase in temperature could result in significant declines in productivity in a hot, dry desert ecosystem due to additional moisture and high temperature stress. Or an increase in temperature could increase productivity in an ecosystem with high initial stocks and turnover rates of available nitrogen, but could have no effect on productivity in a nitrogen-limited ecosystem. Or an increase in temperature could result in a better competitive advantage for a species at the northern limit of its range, whereas the same increase in temperature could reduce the competitive advantage of the same species already at the southern limit of its range. Shaver *et al.* (2000) suggested that the following list of initial conditions should be considered when assessing ecosystem responses to warming:

1. stocks and initial turnover rates of labile soil carbon;
2. stocks and initial turnover rates of labile soil nitrogen;
3. relative size of the carbon pools of plants and soil;
4. dominant form of available nitrogen in the soil (e.g., organic nitrogen, ammonium, nitrate);
5. soil water and precipitation regime;
6. chemical composition and turnover rates of live plant tissues and litter;
7. longevity of individuals and population turnover rates of dominant plant species.

The time scale of response and the major controls on response will thus vary in different ecosystems. For example, in ecosystems where the dominant vegetation is long-lived trees, changes in soil nutrient availability may have a greater impact on total ecosystem productivity on the annual to decadal scale, whereas in ecosystems where the dominant vegetation is short-lived grasses, changes in species composition may be more important.

Interactions With Other Elements of Global Environmental Change

Ultimately, the effects of elevated temperature on terrestrial ecosystems must be understood in the context of concomitant changes in other elements of global environmental change, including higher atmospheric carbon dioxide concentrations, regionally variable atmospheric nitrogen inputs, and both increases and decreases in precipitation. Elevated atmospheric carbon dioxide concentrations, for example, are expected to stimulate plant productivity and increase the flux of carbon to soil. This could partially or completely offset the carbon loss from soil resulting from concurrent increases in temperature. For nitrogen, which is generally considered to be one of the key limiting nutrients in terrestrial ecosystems (Vitousek *et al.*, 1997), a warming-induced increase in the internal production of inorganic nitrogen (nitrogen mineralization, *see* **Nitrogen Mineralization**, Volume 2) could stimulate plant growth and NEP at nitrogen limited sites. This would effectively transfer nitrogen from soil pools with typically low ratios of carbon to nitrogen (carbon/nitrogen of 30 or less) to plant biomass pools, with typically higher ratios (carbon/nitrogen ratios >30). Because of these higher carbon/nitrogen ratios, the same amount of nitrogen can support a greater net storage of carbon in an ecosystem. However, in nitrogen-sufficient or nitrogen-rich environments, particularly those in the Northern Hemisphere which are already impacted by elevated atmospheric nitrogen deposition, a warming induced increase in internal nitrogen production could lead to, or further exacerbate, conditions of nitrogen saturation, where the input of nitrogen equals or exceeds the ability of an ecosystem to assimilate the added nitrogen (Aber *et al.*, 1998). Symptoms of nitrogen saturation include nutrient imbalances in plant tissues, forest decline, increased gaseous loss of nitrogen, and increased nitrogen leaching from soils with consequent declines in associated surface water quality (Vitousek *et al.*, 1997) (*see also* **Nitrogen**

Cycle, Volume 2; **Nitrogen Deposition on Forests**, Volume 2). Changes in precipitation regimes could dramatically alter responses to warming alone because of the many interactions between temperature and water, as discussed previously. Increased precipitation could offset the increased water loss due to warming-induced increases in evapotranspiration, thereby preventing low-moisture stress, and resulting in higher NPP. Reduced precipitation, on the other hand, could exacerbate low-moisture stress, resulting in declines in ecosystem productivity.

ECOSYSTEM WARMING EXPERIMENTS: EVIDENCE OF RESPONSE TO WARMING

Over the last 5–10 years, as global warming predictions have become more widely accepted, the need for information on the response of ecosystems to changing temperature has been addressed by a growing number of ecosystem warming experiments that have been initiated around the world. Although these studies do not perfectly simulate the expected increase in temperature which will occur on the decadal-to-century time scale, and few of these studies have evaluated interactions between elevated temperature and other elements of global change, these experimental manipulations do provide a useful tool to (a) test conceptual and quantitative models of ecosystem response to warming, and (b) begin to evaluate what kinds of ecosystems and which ecosystem processes are most sensitive to warming.

Consistent with the conceptual model of multiple, interacting direct and indirect responses to warming occurring over variable time scales, results from this first decade of ecosystem warming experiments have been highly variable in time and space, making general response patterns difficult to identify. For example, at different sites, experimental warming has resulted in increases, decreases, and no response in ecosystem processes such as soil respiration, litter decomposition, methane production and oxidation, nitrogen mineralization, and plant productivity. To address these apparently different responses to experimental warming, Rustad *et al.* (2001) synthesized results on the response of soil respiration, net nitrogen mineralization, and aboveground plant productivity to experimental warming at sites associates with the Global Change in Terrestrial Ecosystem's Network of Ecosystem Warming Study Sites (GCTE-NEWS). Arft *et al.* (2000) synthesized results on the response of plant productivity and reproductive success to experimental warming at sites associated with the International Tundra Experiment (ITEX), as summarized below. Both syntheses used *meta-analysis* to evaluate responses to warming. Meta-analysis is a technique developed specifically for the statistical synthesis of independent experiments (*see* **Metadata**, Volume 2).

The GCTE-NEWS Synthesis

The GCTE-NEWS includes 32 experimental warming sites, representing four broadly defined biomes: high tundra, low tundra (including a temperate peat bog), forest, and grassland. Warming methods included electrical heat resistance ground cables, greenhouses, open top chambers, climate controlled chambers, and passive night time warming. The duration of these studies ranged from two to nine years, and the experimental increases in soil or air temperature (whichever was the target) ranged from 0.3 to 6.0 °C, with a mean increase of 2.4 °C across all 32 studies.

Data on the response of soil respiration (which represents the combined respiration of roots and soil micro- and macro-organisms) to experimental warming were available at 17 of the 32 GCTE-NEWS sites. Across these 17 sites, experimental warming significantly increased rates of soil respiration by a mean of 20% (with a 95% confidence interval of 18–22%). Such an increase in soil respiration would correspond to an increased release of $\sim$14–20 Pg C year^{-1} from the soil to the atmosphere, based on estimates of annual global soil respiration ranging from 68 Pg C year^{-1} (Raich and Schlesinger, 1992) to 100 Pg C year^{-1} (Musselman and Fox, 1991). This increase is two to three times the estimated $\sim$7 Pg C year^{-1} of carbon dioxide currently released to the atmosphere via combined fossil fuel combustion and land-use changes (Lal *et al.*, 1995), and could significantly exacerbate already escalating atmospheric carbon dioxide concentrations and further disrupt the global carbon balance and global climate system. Caution must be exercised, however, in extrapolating results from experimental warming studies to a global basis, particularly as it is possible that the observed increases in soil respiration are, at least in part, a transient response, reflecting the oxidation of only the most labile soil carbon compounds. Once the labile soil carbon fractions are decomposed, the response of soil respiration to experimental warming may decline. The magnitude and duration of the response of soil respiration to warming will also likely be influenced by changes in the quantity and distribution of precipitation, or warming induced declines in soil moisture, as discussed previously. Site specific studies have shown a negative response of soil respiration to elevated temperature under situations where soil moisture becomes limiting (e.g., Harte and Shaw, 1995).

Data on the response of net nitrogen mineralization (*see* **Nitrogen Mineralization**, Volume 2) in the upper organic soil horizons to experimental warming were available at 12 of the 32 GCTE-NEWS sites. Across these 12 sites, experimental warming significantly increased rates of upper organic soil horizon net nitrogen mineralization by a mean of 46% (with a 95% confidence interval of 30–64%). Gross rates of nitrogen mineralization were probably higher than the values reported here for net mineralization, because

a large percentage of total mineralized nitrogen is typically quickly taken up and immobilized by the microbial community, particularly at nutrient-limited sites. For the 12 sites in the GCTE-NEWS synthesis, a 46% increase in net nitrogen mineralization corresponded to an increased internal production of 4.8 kg N ha^{-1} $year^{-1}$. This compares to atmospheric wet plus dry deposition inputs ranging from <5.0 kg N ha^{-1} $year^{-1}$ in relatively pristine areas to 50 kg N ha^{-1} $year^{-1}$ or more in the heavily polluted areas of the northeastern US and Europe (Vitousek *et al.*, 1997).

Data on aboveground plant productivity response to experimental warming were available at 20 of the 32 GCTE-NEWS sites. Across these 20 sites, experimental warming significantly increased rates of aboveground plant productivity by a mean of 19% (with a 95% confidence interval of 15–22%). Increases in plant productivity would result in a greater storage of carbon in plant biomass (especially in systems with a significant woody component) and increased flux of carbon to soils in leaf and root litter. This could at least partially offset the warming-induced increase in carbon flux from soils to the atmosphere via soil respiration.

The ITEX Synthesis

The ITEX is a collaborative, multi-site experiment using a common open top chamber experimental design to evaluate variability in species response to warming in tundra ecosystems. Arft *et al.* (2000) examined the response of plant phenology, growth, and reproduction to 1–4 years of experimental warming at 13 circumpolar sites. Results showed different pathways of response in time and space. For example, although key phenological events such as leaf bud burst, flowering, and seed dispersal generally occurred earlier in warmed plots compared to control plots throughout the four years of the study, plant productivity was greater in the warmed plots than the control plots during the first 2–3 years of the experiments, whereas reproductive effort and success increased in later years, particularly in year four. Arft *et al.* (2000) attributed the decrease with time in plant productivity response to warming in part to a depletion of belowground resources. They attributed the delayed reproductive response to the fact that flower buds are typically formed from one to several seasons before flowering. Arft *et al.* (2000) further showed that: (1) the vegetative response to warming differed among plant life forms, with the response being generally greater in herbaceous than woody species; (2) warmer, low arctic sites showed the strongest growth response to warming; and (3) colder, high arctic sites showed the strongest reproductive response to warming. Overall, these differential responses to warming observed among sites, life forms, and over time are consistent with the conceptual model of ecosystem response to warming reviewed above, and underscore the need to evaluate ecosystem response to warming at different temporal and spatial scales.

CONCLUSIONS

There is no question that concentrations of atmospheric carbon dioxide and other greenhouse gases have increased dramatically over the last century due to human activities. The evidence continues to accumulate that this increase in greenhouse gas emissions has been and will continue to be accompanied by increases in mean global temperature. Because temperature is a fundamental driver of most chemical and biological reactions, these human-induced changes in thermal regimes will have a multitude of direct and indirect effects on terrestrial ecosystems at multiple spatial and temporal scales. Predicting the response of whole ecosystems to increases in temperature will therefore be difficult, and will require continued refinement and integration of conceptual and quantitative models of short- and longer-term ecosystem response to increased temperature.

Completed and ongoing ecosystem warming experiments have proven useful tools to explore the response of ecosystems and ecosystem components to experimental warming, and have shown conclusively that increases in temperature in the range of 0.3–6.0 °C can have significant impacts on fundamental ecosystem processes, including soil respiration, nitrogen mineralization, and plant growth, phenology, and reproductive success. Future ecosystem warming experiments need to: (1) more fully integrate the experiments at the ecosystem level, including both aboveground and belowground processes. (2) Investigate interactions with other vectors of environmental change, especially water, carbon dioxide, nitrogen deposition, and tropospheric O_3; and (3) expand the spatial and temporal scales of analysis, including additional experiments in under-represented biomes, such as tropical, desert, and Mediterranean ecosystems, and over longer time periods.

REFERENCES

Aber, J, McDowell, W, Nadelhoffer, K, McGill, A, Berntson, G, Kamaka, M, McNulty, S, Currie, W, Rustad, L, and Fernandez, I (1998) Nitrogen Saturation in Forest Ecosystems: Hypotheses Revisited, *BioScience*, **48**, 921–934.

Arft, A M, Walker, M D, Gurevitch, J, Alatalo, J M, Bret-Harte, M S, Dale, M, Diemer, M, Gugerli, F, Henry, G H R, Jones, M H, Hollister, R, Jónsdóttir, I S, Laine, K, Lévesque, E, Marion, G M, Molau, U, Mølgaard, P, Nordenhäll, U, Raszhivin, V, Robinson, C H, Starr, G, Stenström, A, Stenström, M, Totland, Ø, Turner, L, Walker, L, Webber, P, Welker, J M, and Wookey, P A (2000) Response Patterns of Tundra Plant Species to Experimental Warming: a Meta-analysis of the International Tundra Experiment, *Ecol. Monogr.*, **69**, 491–512.

Barber, V A, Juday, G P, and Finney, B P (2000) Reduced Growth of Alaskan White Spruce in the 20th Century from Temperature-induced Drought Stress, *Nature*, **405**, 668–673.

Fitter, A H, Self, G K, Brown, T K, Bogie, D S, Graves, J D, Benham, D, and Ineson, P (1999) Root Production and Turnover in an Upland Grassland Subjected to Artificial Soil Warming Respond to Radiation Flux and Nutrients, Not Temperature, *Oecologia*, **120**, 575–581.

Giardina, C P and Ryan, M G (2000) Evidence that Decomposition Rates of Organic Carbon in Mineral Soil do not vary with Temperature, *Nature*, **404**, 858–861.

Gill, R A and Jackson, R B (2000) Global Patterns of Root Turnover for Terrestrial Ecosystems, *New Phytol.*, **147**, 13–31.

Gunderson, C A, Norby, R J, and Wullschleger, S D (2000) Acclimation of Photosynthesis and Respiration to Simulated Climatic Warming in Northern and Southern Populations of *Acer saccharum*: Laboratory and Field Evidence, *Tree Physiol.*, **20**, 87–96.

Harte, J and Shaw, M R (1995) Shifting Dominance within a Montane Vegetation Community: Results from a Climate-warming Experiment, *Science*, **267**, 876–880.

Intergovernmental Panel on Climate Change (IPCC) (1996) *Climate Change 1995: The Science of Climate Change*, University Press, Cambridge.

Karl, T R, Knight, R W, and Baker, B (2000) The Record Breaking Global Temperature of 1997 and 1998: Evidence for an Increase in the Rate of Global Warming? *Geophys. Res. Let.*, **27**, 719.

Lal, R, Kimble, J, Levine, E, and Stewart, B A (1995) *Soil Management and the Greenhouse Effect*, CRC Press, London.

Myneni, R B, Keeling, C D, Tucker, C J, Asrar, G, and Nemani, R R (1997) Increased Plant Growth in the Northern High Latitudes from 1981–1991, *Nature*, **386**, 98–702.

Musselman, R C and Fox, D G (1991) A Review of the Role of Temperate Forests in the Global carbon dioxide Balance, *J. Air Waste Manage. Assoc.*, **41**, 798–807.

Norby, R J, Long, J T M, Hartz-Rubin, J S, and O'Neill, E G (2000) Nitrogen Resorption in Senescing Tree Leaves in a Warmer, Carbon Dioxide-enriched Atmosphere, *Plant Soil*, **224**, 15–29.

Raich, J W and Schlesinger, W H (1992) The Global Carbon Dioxide Flux in Soil Respiration and its Relationship to Vegetation and Climate, *Tellus*, **44**, 81–89.

Rustad, L E, Campbell, J L, Marion, G M, Norby, R J, Mitchell, M J, Hartley, A E, Cornelissen, J H C, Gurevitch, J, and GCTE-NEWS (2001) A Meta-analysis of the Response of Soil Respiration, Net N Mineralization, and Aboveground Plant Growth to Experimental Ecosystem Warming, *Oecologia*, **126**, 543–562.

Shaver, G R, Canadell, J, Chapin, III, F S, Gurevitch, J, Harte, J, Henry, G, Ineson, I, Jonasson, S, Melillo, J, Pitelka, L, and Rustad, L (2000) Global Warming and Terrestrial Ecosystems: a Conceptual Framework for Analysis, *BioScience*, **50**, 871–882.

Vitousek, P M, Aber, J, Howarth, R W, Likens, G E, Matson, P A, Schindler, D W, Schlesinger, W H, and Tilman, G D (1997) Human Alteration of the Global Nitrogen Cycle: Causes and Consequences, *Ecol. Appl.*, **7**, 737–750.

Weis, E and Berry, J A (1988) Plants and High Temperature Stress, in *Plants and Temperature*, eds S P Long and F I Woodward, Society for Experimental Biology, Cambridge, 329–346.

Woodwell, G M (1995) Biotic Feedbacks from the Warming of the Earth, in *Biotic Feedbacks in the Global Climatic System: will the Warming Feed the Warming*? eds G M Woodwell and F T McKenzie, Oxford University Press, Oxford, 3–21.

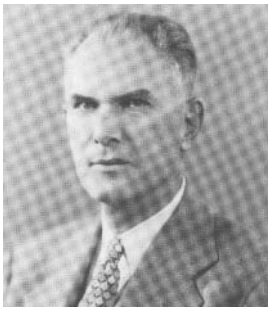
Terpenes

see Volatile Organic Compounds, Biogenic (VOCs) (Volume 2)

Terrestrial Transects

see IGBP Terrestrial Transects (Volume 2)

Thornthwaite, Charles Warren



(1899–1963)

Charles Warren Thornthwaite changed the course of climatology in the 20th century by stressing that climatology was not solely the realm of the geographer, but belonged also to the botanist, hydrologist and engineer. He was the founder of the field of applied climatology. His water budget model has been used by hydrologists as a means of quantifying the hydrologic cycle, to estimate surface runoff where no data exist, and to extend the runoff record back in time. It provides data on soil moisture and evapotranspiratiion of benefit to agriculture and forest scientists, and estimates of water deficiency useful to crop scientists and irrigation engineers. In recent years, it has been widely applied in studies of the impact of climate change on water resources.

Charles Warren Thornthwaite was born in 1899 on a farm in central Michigan (USA). After graduating from Central Michigan University he became a high school teacher, but at the age of 25 was persuaded by friend John Leighly to take graduate work in geography at the University of

California. After receiving his PhD from UCLA he became a professor of geography at the University of Oklahoma. Until this time he had shown little interest in climatology, but living in the dry state of Oklahoma made him realize the importance of water. In his first climatic publication (in 1931) *The Climates of North America According to a New Classification* he used evaporation data to derive precipitation/evaporation indices.

During the depression years of the 1930s President Roosevelt established the Soil Erosion (later Conservation) Service in the Department of Agriculture in Washington and in 1935, Thornthwaite was appointed head of the Climatic and Physiographic Research Branch of this Service. The appointment marked a change in Thornthwaite's life and the beginning of his career in climatic research. He obtained large grants for his research in the loss of water from land to air, first in western Oklahoma and later in Arizona and Ohio. Much of this research was published in the 1941 Yearbook of Agriculture *Climate and Man*.

During the World War II years, Thornthwaite continued his climatic research as a consultant in irrigation to the Mexican Government. He revised his 1931 climate classification paper using actual water use data. *An Approach Toward a Rational Classification of Climate* published in 1948 was undoubtedly the most important of his writings. The central concept of the revision was *potential evapotranspiration*, defined as the amount of water that would be lost from a vegetated surface by evaporation and transpiration if an unlimited supply existed. He formulated a water budget model in which the need for water (potential evapotranspiration) was compared with the water supply (precipitation plus soil moisture) to derive quantitative estimates of actual evapotranspiration, water surplus and water deficiency.

Thornthwaite was also interested in practical applications of his research and in 1946 combined theory and practice by becoming a consulting climatologist to Seabrook Farms in New Jersey, one of the world's largest vegetable growers. He advised Seabrook on scheduling the planting and harvesting of peas, and a method of disposing of wastewater by irrigation. He also realized a life-long dream of a climatic research institute by forming a Laboratory of Climatology to conduct research and, in collaboration with The Johns Hopkins University, to train students in climatology. Research at the laboratory included the measurement of vertical gradients of wind, temperature and humidity, and of potential evapotranspiration by evapotranspirometer tanks. The laboratory also designed and manufactured instruments such as net radiometers and wind profile systems.

During the 1950s and 1960s many famous climatologists visited the laboratory: Anders Ångström and C C Wallén from Sweden, Rudolf Geiger from Germany, C E P Brooks and Howard Penman from England, Ken Hare from Canada and Helmut Landsberg and Heinz Lettau from the United States. Thornthwaite received many honors. In 1951 he was elected president of the Commission of Climatology of the newly established World Meteorological Organization. He was elected president of the American Association of Geographers and was awarded the Cullum medal by the American Geographic Society.

FURTHER READING

Mather, J R and Sanderson, M (1996) *The Genius of C. Warren Thornthwaite, Climatologist – Geographer*, University of Oklahoma Press, Norman, OK.

MARIE SANDERSON *Canada*

Transport and Species Dispersal

see Transport and Species Dispersal (Volume 3)

Tree Rings

see Natural Records of Climate Change (Volume 1)

Tropical Forests

Ghillean T Prance
University of Reading, Reading, UK

Tropical forests are renowned for their species diversity and harbour more than 50% of the world's species. Tropical forests are the source of many well-known products of economic importance such as rubber, cacao and coffee. Concern is expressed about the amount of tropical deforestation in recent years and its effect on species loss and worldwide climate patterns.

The type and structure of tropical forest vary principally with altitude, soil, flooding regime, and rainfall patterns. The three main blocks of tropical forests occur in Amazonia, West and Central Africa, and in Southeast Asia, but tropical forest also occurs on many tropical islands in the Caribbean and the Pacific.

A large part of the tropical belt is covered by a variety of forest types. These vary mainly according to the climate, soil, topography and flooding regime. The major types of tropical forest are given in Table 1, but each of these categories can be subdivided in many different ways. Floristically, tropical forests divide into three main regions, each with its own group of species of plants and animals.

Table 1 Principal types of tropical forest[a]

1 Tropical evergreen forest (rainforest)

A seasonal forest with mostly evergreen species that have little or no bud protection against cold or drought

(a) Lowland rainforest: multilayered, many trees exceeding 30 m in height, closed canopy with sparse undergrowth, large lianas frequent, epiphytes relatively scarce

(b) Montane rainforest: few trees exceeding 30 m, palms and tree ferns common in undergrowth, ground layer rich in herbs and mosses, epiphytes common; often subdivided into lower montane and upper montane rainforest

(c) Cloud forest: high altitude closed forest with many gaps, trees rarely over 20 m, often gnarled, numerous lianas, tree ferns, herbs, shrubs and abundant epiphytes

(d) Alluvial forest: forest growing in seasonally inundated areas along river banks, multilayered, closed forest with numerous gaps, buttresses and stilt roots common, palms, herbaceous undergrowth and lianas common, epiphytes common

(e) Swamp forest: forest on permanently wet areas, buttresses, stilt roots and pneumatophores common, palms, ferns and herbs abundant

(f) Peat forest: forest over deep peat deposits in nutrient poor soils mainly in tropical Asia. Rarely above 20 m in height, ground cover mostly of ferns

2 Tropical monsoon forest

Forest with high, but strongly seasonal rainfall, largely evergreen, but more leaf shedding in dry season, trees of canopy moderately drought resistant, tree ferns absent in montane forest, but climbers and evergreen shrubs common, dense canopy with little undergrowth. Often divided into lowland and montane monsoon forest

3 Tropical semi-deciduous forest

Trees of upper canopy shed leaves in strong dry season, small trees and shrubs usually evergreen, open with shrubs, herb and tree saplings abundant, epiphytes rare

4 Mangrove forest

Single layered forests to 30 m height growing in intertidal zones in salt water throughout tropics and subtropics, evergreen, stilt roots and pneumatophores common, little ground vegetation, few epiphytes

[a] Adapted from Longman and Jenik (1987).

The largest area of tropical forest occurs in Central and South America and accounts for 54% of the total. West and Central Africa have about 18% and the East Asian tropics have 27%. Small areas also occur in various islands in the Caribbean and Pacific regions. Although the flora and fauna in each region are different, the structure and appearance of the various types of tropical forest are quite similar on the different continents.

The term *rainforest* was first used by Schimper (1903), one of the founding fathers of plant geography. He defined it as "evergreen, hygrophilous in character, at least 30 m high, rich in thick-stemmed lianas and in woody as well as herbaceous epiphytes". The term should be confined to the aseasonal, evergreen, multi-layered forest types. The term *tropical moist forest* extends beyond this to other closed canopy tropical forests where the annual rainfall is high (over 2000 mm) but seasonal such as in areas influenced by monsoons.

Lowland rainforest is characterized by its division vertically into three or four relatively distinct strata. The canopy level varies between 30 and 60 m. It tends to be taller in the forests of tropical Asia that are dominated by species of the plant family Dipterocarpaceae, than in the American forests. Above the canopy are a number of emergent trees that can reach 70–80 m in Asia (e.g., *Koompassia excelsa*). Emergents in the Amazon rainforest include the Brazil nut, *Bertholletia excelsa* and the enormous legume tree *Dinizia excelsa*. The strata are not always very distinct, but below the canopy, there are often layers of middle sized trees (10–25 m) and of lower trees (5–10 m). This layer is often of saplings of the larger trees, but each layer has characteristic species of trees that reach only to that height. Below the tree layers are the shrub and herb layers. Very little light reaches ground level (about 3% of what falls on the canopy), and so species of the rainforest are adapted to growing in conditions of extremely low light intensity. This is why some of them (e.g., *Philodendron* and *Calathea* species) make good houseplants since they can grow in shady rooms. The shrub layer is usually poorly developed in dense rainforest. The lowland forests consist largely of broad-leaved evergreen tree species. Large woody vines (lianas) are relatively common especially in Amazonian forests, but are much rarer in the lowland forests of tropical Asia and the Pacific.

At higher elevation, on tropical mountains, the climate is cooler and often wetter and the trees are smaller, the canopy more open and there are many more understory plants. Vegetation ecologists recognize a gradual progression of forest types from *lower montane forest* to *upper montane forest*, and finally, *cloud forest* at the highest elevations to sustain forest. The latter type is low in stature, with gnarled trees and an extremely rich epiphyte flora (plants not rooted in soil). Epiphytes are plants that live on the branches of other plants, but unlike parasites, do not draw nutrients

from their host. In its most reduced form, cloud forest is sometimes called *elphin forest*. There is a tendency towards smaller leaves as altitude increases. Lower montane forests are lower and less species rich than lowland forest and usually have only two strata. There are more epiphytes in this forest. Upper montane forests have even lower canopies at less than 20 m. There are no climbers, but epiphytes are abundant. In some areas single species dominate montane forests.

In regions where the rainfall is more seasonal, such as on the southern borders of the Amazon forest, in parts of Central America and in the monsoon areas of southern India and Myanmar (formerly Burma), the forests are progressively more deciduous, depending upon the length of the dry season. Tropical monsoon forests are mainly evergreen, but the canopy trees show drought-resistant adaptations. In tropical semi-deciduous forests, which occur in areas with a long dry season, the trees of the upper canopy shed their leaves in the dry season, but many of the small trees and shrubs remain evergreen.

Areas that are liable to flooding have different types of forest adapted to the flooding regime. They are less species diverse in terms of trees than the lowland forests on non-flooded ground. *Alluvial forests* exist in areas liable to flooding by the rising river levels of the rainy season, by far the greatest area occuring in Amazonia. The term used in Amazonia for areas flooded by white water rivers is *várzea* forest, whereas areas flooded by black and clear water rivers such as the Rio Negro (black water) and the Rio Xingu (clear water) is called *igapó* in Brazil. The flooded forests of Amazonia were divided into seven categories by Prance (1979). In the Amazon delta region, the river flow is blocked by tidal movements and there is a considerable area of forest which is subjected to twice daily inundation and is known as *tidal várzea*. Structurally and floristically, it is similar to the seasonally flooded várzea but has a greater quantity of palms. Swamp forest is forest on poorly drained areas that are permanently waterlogged. The tree flora is less diverse, palms are abundant and such adaptations as buttressed trunks, stilt roots and pneumatophores (aerial breathing roots) are frequent. Ferns are often abundant in the ground cover. In Asia, especially in Borneo, there is another type of swamp forest over peat. The peat forests have nutrient-poor soils and as a result are rarely over 20 m tall and are less species diverse than the drier forest types. Some peat swamp is dominated by palms, and other areas by broad-leaved trees.

Mangrove forests occur in intertidal zones throughout the tropics and subtropics. They are characterized by the stilt roots of several species and the presence of pneumatophores in many. This forest type is evergreen and single layered and adapted to growth in salt water. Mangroves play an essential role in the stabilization of tropical coast lines. They also support a wide range of wildlife and are equally important as a source of fish and crustaceans. Their rapid destruction is causing loss of income for local fishermen and serious problems of erosion (*see* **Mangrove Ecosystems**, Volume 2).

The largest and still the most intact tract of tropical forest is that of the Amazon Basin, which contains almost 50% of the total, but other important areas of tropical forest also occur in the Americas. Tropical forests reach as far north as Mexico, where they are mainly semi-deciduous and extend in patches throughout Central America. West of the Andes, the coastal region of Colombia (the Chocó) and northern Ecuador is the wettest rainforest in the world (precipitation up to 6000 mm $year^{-1}$), and it contains a large number of endemic species of plants, especially in the shrub and herbaceous layers. The other important neotropical rainforest area is a narrow belt along the Atlantic coast of Brazil from the State of Ceará in the north to São Paulo in the south. Unfortunately, less than 6% of the *Mata Atlantica* remains intact. It contains many endemic and threatened species of plants and animals such as the golden lion tamarin (*Leontopithecus rosalia rosalia*).

The African tropical forests mainly occur in West and Central Africa with a few small, scattered patches extending into East Africa. The forests of West Africa have largely been destroyed and the remaining tropical forests are mainly in the Congo Basin in the Congo Republic, Cameroon, Equatorial Guinea, Gabon and Congo. Only a few small isolated patches of the former forests of West Africa remain intact. Other rainforest areas that have been severely depleted are the unique forests of eastern Madagascar where most species of plants and animals are endemic.

The Indo-Malesian forests have a much more disjointed distribution owing to the island nature of the region. These forests extend from the western Ghats in India, Sri Lanka, through Assam, Myanmar, Thailand, Indo China and the Malay Peninsula to the Philippines, the Indonesian archipelago, and east as far as Fiji in the Pacific, and into the Queensland Peninsula of Australia where it reaches its southernmost limit and blends into the subtropical forests of New South Wales. The main difference between the Malesian forest and those of the rest of the tropics is the dominance of a single plant family the Dipterocarpaceae (genera *Dipterocarpus*, *Dryobalanops*, *Hopea* and *Shorea*).

Most tropical forests are over the soil types *oxisols* and *ultisols*, both of which are acid, low in organic matter content and extremely low in available plant nutrients. Most of the nutrients are in the decomposing leaves and branches on the soil surface and are rapidly recycled back into the vegetation. Ultisols are prone to erosion and to the formation of laterite consisting of iron-rich clay. One of the consequences of the poor soils is that most attempts to fell tropical forest and replace it with agriculture do not work. This can be easily seen from the abandoned areas of cattle pasture along the trans-Amazon highway or the vast areas

of unusable coarse grassland which has replaced much of the rainforests of Madagascar and Malesia, where *Imperata cylindrica* grassland prevents the recolonization by woody species.

Lowland tropical forests are renowned for their diversity of plant and animal species. Table 2 gives some statistics for the diversity of rainforest trees. It is estimated that about 50% of the world's species occur in tropical rainforests, which occupy only 7% of the land surface. This diversity and the local distribution of so many species is, in itself, a powerful argument for the preservation of tropical forests. In some areas of the Neotropical rainforest 300 tree species of over 10 cm in trunk diameter occur on a single hectare. The 100 km square Reserva Florestal Adolfo Ducke near Manaus, Brazil contains 2125 species of flowering plants in comparison to the 1500 in the entire British flora. Barro Colorado Island in Panama has 1316 species of plants in its 15.6 km^2, and the Río Palenque Science Center in Ecuador contains 1100 species in 167 hectares, and Brunei Darussalam, an area of 5767 km^2 has 4000 species of vascular plants. Table 2 shows that the non-flooded forests tend to be more diverse than flooded ones.

Table 2 Selected examples of a number of tree species ≥10 cm diameter ha^{-1} of lowland forest

Location and forest type	No of tree species/ ha ≥10 cm	≥5 cm
Cuyabeno, Ecuador; Terra rme	307	473
Serra Grande, Bahia, Brazil; Terra rme	300	
Yanomono, Peru; Terra rme	283 (+17 lianas)	
Añangu, Ecuador; Terra rme	228	
Johore, Malaysia; Terra rme	227	
Mulu, Sarawak; Terra rme	223	
Oveng, Gabon; Terra rme	131	
Belém, Brazil; Terra rme	153	
Rio Negro, Brazil Igapó		
Lightly flooded	137	
Medium flooding	103	
Long flooding	44	
Rio Japurá, Brazil; Flooded forest	119	
Alto Ivon, Bolivia; Terra rme	94	
Belém, Brazil; Flooded várzea forest	60	
Rio Xingu, Brazil (0.5 ha); Flooded várzea forest	40	
Pantanal, Mato Grosso, Brazil; Semi-deciduous forest	35	
Rio Tapajós, Brazil; Igapó	21-30	

The diversity of plants in tropical forests enables them to sustain an enormous diversity of animals, especially insects, many of which are adapted to the canopy of a particular species of tree. Animals have adapted to the resources in the different levels of the forest. Birds and bats hunt above the canopy, and there is a vast selection of birds, fruit bats and mammals, including primates feeding in the canopy. In the middle zones there are both flying animals and animals whose habitat is the tree trunks, such as many well-disguised lizards and chameleons. At ground level, the animals vary from large mammals such as pigs, tapirs and deer to small rodents and mixed flocks of birds which follow army ant colonies feeding on the insects which the ants disturb. In Australia and New Guinea the large flightless cassowaries occupy the fruit-eating niche on the forest floor that is occupied by mammals in other tropical forests. There are many cases of different groups of animals performing the same function on the different tropical continents. A striking example is the similarity between the Neotropical toucans and the Old World hornbills.

Until relatively recently, most tropical forests were not seriously disturbed. Indigenous populations have lived in and used tropical forests for many generations, but their activities had little effect on the distribution of forest or the species composition. A few notable species became extinct in earlier times such as some of the megafauna of South America and various island species such as the dodo on Mauritius. From about 1600, as western civilization began to colonize the world and introduce many new crops, large areas of forest were cleared. Coffee and cacao plantations in Brazil, rubber in Malaysia, tea in India, sugar in the Caribbean and northeast Brazil all took their toll. However, it has been the last 50 years, as the world population has soared and tropical countries have developed, that a vast change has taken place and deforestation has been rapid.

Deforestation has occurred for many reasons, but one of the principal ones has been to provide land for increasing populations. Transmigration projects to occupy forested areas have been sponsored by the governments of Brazil and Indonesia. The need for fuel wood by poor people and for tropical timber by the developed countries has caused much deforestation, the latter particularly in tropical Asia and West Africa. Roads, hydroelectric dams, the expansion of cities, and mining activities have all taken their toll. Today just over half of the world's tropical forests have been cleared.

An assessment of the status of the world's remaining closed forests has recently been published by UNEP (Singh *et al.*, 2001). See www.na.unep.net/pub/closedforest/.

Table 3 Areas of closed forest in nine of the countries that contribute most to the world's closed forest cover. (Temperate zone forests included for comparison)

	Country size of area of closed forest (billions of hectares)	Percent of country's land area
Russia	669.6	40
Canada	368.6	37
Brazil	361.5	42
USA	236.6	25
Rep. Congo	116.2	49
China	115.5	12
Indonesia	92.7	49
Mexico	60.1	30
Peru	59.3	45

Closed forests are defined as forests where 40% or more of tree crowns are interlocked (Singh *et al.*, 2001). See Table 3 for summary.

Some of the most interesting results are that:

Of the planet's remaining closed forests, 80.6% are located in only 15 countries – Russia, Canada, Brazil, the United States, the Democratic Republic of the Congo, China, Indonesia, Mexico, Peru, Columbia, Bolivia, Venezuela, India, Australia and Papua New Guinea. Four are industrialized countries and 11 are in the developing world. Ten are in the tropics, or partly so.

Population pressures vary between countries. Overall, 88% of the world's closed forests in these 15 countries have low or non-existent population densities.

Overall, an estimated 88% of these forests have low levels of people living in and around them, giving conservation efforts a real chance of success.

Amongst the 15 key countries, Venezuela enjoys the highest level of protection, with 63% of the closed forest designated as protected. This is followed by Bolivia with 30%, Columbia 25% and Indonesia 20%. In contrast, the lowest level of protection is only 2% in Russia while the figures are 6.7% for the United States and 7.4% for Canada.

The loss of tropical forests is contributing to the global increase in atmospheric CO_2 both through the release of carbon from burning the forest and from removal of this important sink of carbon. Brazilian scientists Salati and Vose (1984) showed that deforestation in Amazonia could lead to a vast change in the rainfall pattern of the entire region, because 75% of the rainfall is recycled through evapotranspiration and only 25% flows out through the river system. The loss of tropical forest has many other implications such as the loss of species and of the potential uses to which they might be put. In addition deforestation is leading to a vast loss of the nutrients stored in the vegetation and to serious erosion in many places. Often the increase in sediment washed out to sea is having an adverse effect on marine ecosystems, especially the species-diverse coral reefs. The tropical forests of the world are being destroyed largely for short-term and unsustainable uses. One of the great challenges facing the 21st century is to halt the destruction of tropical forest, and introduce sustainable land use systems in the areas where the forest has already been destroyed.

An excellent summary of tropical rainforests can be found in Richards (1996).

See also: **Large Scale Biosphere–Atmosphere Experiment in Amazonia (LBA)**, Volume 2; **Mangrove Ecosystems**, Volume 2; **Afforestation: Environmental Impacts**, Volume 3; **Deforestation and Habitat Fragmentation in the Amazon Basin**, Volume 3; **Deforestation, Tropical: Global Impacts**, Volume 3.

REFERENCES

Groombridge, B, ed (1992) *Global Biodiversity: Status of the Earth's Living Resources*, Report Compiled by the World Conservation Monitoring Center, Chapman and Hall, London.

Longman, K A and Jenik, J (1987) *Tropical Forest and its Environment*, 2nd edition, Longman Scientific and Technical, Harlow.

Prance, G T (1979) Notes on the Vegetation of Amazonia III. The Terminology of Amazonian Forest Types Subject to Inundation, *Brittonia*, **31**, 26–38.

Richards, P W (1996) *The Tropical Rainforest, an Ecological Study*, 2nd edition, Cambridge University Press, Cambridge.

Salati, E and Vose, P B (1984) Amazon Basin: a System in Equilibrium, *Science*, **225**, 129–138.

Schimper, A F W (1903) *Plant-geography Upon a Physiological Basis*, Translation by W R Fisher, Clarendon Press, Oxford.

Singh, Ashbindu, Shi, H, Zhu, Z, and Foresman, T (2001) An Assessment of f the World's Remaining Closed Forests, see www.na.unep.net/pub/closedforest/.

Tropical Savannas

Ernesto Medina

Instituto Venezolano de Investigaciones Cient´ cas, Caracas, Venezuela

Tropical savannas are ecosystems characterized by a continuous herbaceous cover of sun loving (heliophilous), perennial bunch grasses and sedges, almost exclusively with the carbon-4 (C4) photosynthetic pathway, that show a clear seasonality related to water stress. Woody species (shrubs, trees and palms) occur in varying density depending on soil depth and water availability. Legumes are commonly found in savanna vegetation as herbaceous, shrubby and tree species; particularly in South American savannas they can

reach high specific diversity. Most shrub and tree species in tropical savannas of Western Africa, Australia, and South America are evergreen, and produce new leaves towards the end of the dry season.

Physiognomic and functionally similar systems extend into the subtropics in Africa, Australia and South America. However, at higher latitudes frost occurrence becomes an additional selecting factor. The main difference between savannas and temperate prairies is the occurrence in the latter of frost-resistant herbaceous and woody species.

Figures on the total extent of savannas are variable. Estimates range from 21 to 24 $\times 10^6$ km^2 from which about 15 $\times 10^6$ km^2 correspond to tropical latitudes. Savanna ecosystems cover large areas of the terrestrial surface of Africa (65%), Australia (60%) and South America (45%).

Savannas have been heavily modified throughout human history. The cradle of humankind is located in the African savannas (about 500 000 years ago), while human occupation is only 40 000 years old in Australia and around 15 000 years old in South America. From the very beginning human modification of savannas operated through the utilization of fire for hunting, and the domestication of herbivores. However, massive modification is relatively recent, beginning with the arrival of European settlers. Most modifications are related to the agricultural exploitation of grasslands, resulting in pronounced variations in species diversity and disruption of ecosystem functioning such as productivity and related processes of water and nutrient cycling (particularly nitrogen).

Currently savannas are burned regularly, mostly because of human influences. There is little doubt that the present extent of savannas is to a significant degree the result of human induced fires. However, the occurrence of fire probably regulated the distribution and species composition of savannas long ago, particularly in tropical South America. Due to recurrent fires, the current carbon balance of savanna areas is probably at equilibrium, organic matter accumulated during the rainy season being returned to the atmosphere during the dry season. However, precise, large-scale measurements of the carbon balance are lacking.

Most tropical savannas are comparatively humid, with annual rainfall above 800 mm, have acid, dystrophic to slightly mesotrophic soils, and are burned on an annual or biannual basis. Grasslands in the African highlands characterized by the occurrence of brown, eutrophic soils of alkaline pH are not considered here, because of their contrasting nutritional conditions.

VEGETATION AND FLORISTIC DIVERSITY IN SAVANNAS

The large-scale geographical distribution of savannas in the tropics appears to be primarily correlated with the presence of a pronounced seasonal rainfall regime, but their regional distribution pattern is more strongly influenced by local edaphic and biotic conditions. Different savanna types occupy in some cases wide and continuous areas with a relatively homogeneous climatic regime, as for example in Northern and Central South America, central and Western Africa, Southern India, and Northern Australia. On the other hand, isolated savannas occurring in small patches interspersed in large forested regions, as for example in the Amazon basin or in parts of Southeast Asia, are mainly the result of peculiar edaphic or historical situations.

Generally, the density of the ligneous component is used as a differential criterion for the definition of a particular savanna type. Thus, savanna vegetation types vary from open grasslands, where the woody component, if present, does not surpass the height of the dominant grass layer, through shrub savannas, tree savannas and savanna woodlands or forests dominated by tree species with hard, leather-textured, leaves (*sclerophyllous*).

In African savannas the herb layer shows various degrees of differentiation into tall grass savannas and low grass savannas, probably in response to the selective impact produced by the much higher density and diversity of large herbivores.

Contrary to tropical humid forest ecosystems, where the predominance of certain plant families is rather the exception, savannas are invariably dominated in their herbaceous layer by only two large families, namely the grasses (*Poaceae*) and the sedges (*Cyperaceae*). In most savannas of the world, these two families account not only for the largest number of species, but they also constitute the main biomass of the ecosystem. Third in floristic importance are the legumes (*Fabaceae*). Other important herb families are *Asteraceae*, *Rubiaceae*, *Euphorbiaceae*, *Labiatae*, *Convolvulaceae*, *Scrophulariaceae*, *Gentianaceae*, and *Sterculiaceae*.

The ligneous compartment consists of a variable number of species belonging to many families. Of these, four families are particularly important in the neotropical shrub savannas: *Dilleniaceae*, with the widespread but almost exclusive savanna treelet *Curatella americana*; Malphigiaceae, with the large genus *Byrsonima*; *Proteaceae* (*Roupala*), and *Arecaceae* (palms) with numerous genera and species. Furthermore, each savanna region of South America has an additional set of typical woody species. Examples are *Salvertia convallariodora* (Vochysiaceae) and *Hancornia speciosa* (Apocynaceae) in the central Brazilian cerrados, *Antonia ovata* (Loganiaceae) in the Northern Brazilian Rio Branco savannas, and *Heteropteris laurifolia* (Malphigicaeae) in the Bolivian savannas. In Australian savannas, species of the genus *Eucalyptus* dominate the ligneous compartment in shrub and tree savannas, while in Africa the legumes genera *Brachystegia*, *Julbernardia*, *Burkea*, *Colophospermum*, and *Acacia* dominate the ligneous stratum of the largest tracts of savanna ecosystems.

LIFE FORMS AND THEIR FUNCTIONAL SIGNIFICANCE

Well-developed savanna ecosystems contain almost the entire spectrum of life forms in the classical sense of Raunkiaer (Danish ecologist who classified higher plants according to the protection of their growing points against environmental stress), such as phanerophytes, chamaephytes, hemicryptophytes, geophytes, and therophytes. However, dominant life forms in savannas are those of perennial species and within them two groups contain the largest number of species: the herbaceous graminoid and broad-leaved species and the sub-ligneous and ligneous hemicryptophytes.

The graminoid life form presents two major growth forms, i.e., the tussock or bunch grass growth form, and the solitary growth form. The former consists of a dense tuft of up to one meter or more in diameter of perennial grasses or sedges, all arising from a common rootstock. This is by far the predominant growth form in the herbaceous layer of most savannas, as for example *Trachypogon plumosus* (Poaceae) or *Scleria cyperina* (Cyperaceae) in many Neotropical savannas. The solitary growth form is more commonly found in annual grass or sedge species, although some perennial grasses may also exhibit this feature.

Tree and herbaceous components in savannas exploit water and nutrient resources from different soil layers. In neotropical savannas, trees appear not to be limited by water supply, because they are able to exploit soil layers influenced by the water table, while grasses show strong seasonal variations in water potential and a large degree of osmotic adjustment. However, both density and height of the tree-root layer are positively related to soil water availability during the dry season. The establishment of tree seedlings depends on the occurrence of wet years, that help them to withstand root competition from perennial grasses during early stages of growth.

Soil fertility also influences actual tree/grass ratios, as shown in the physiognomic sequence from savanna woodland to the open shrubby grasslands in central Brazil. The reduction in the number of tree stems and species per hectare (ha) observed in this sequence may be attributed both to a reduction in nutrient availability and increased aluminum saturation of the soil exchange complex. However, these nutritional gradients might be the result and not the cause of vegetation composition, especially when only the chemical composition of the upper soil layers is taken into account.

Differences in density of the tree layer may also result from variations in frequency and intensity of the fire regime. Fire suppression for more than 20 years in several savanna sites in different continents led to significant increases in diversity of both tree and herbaceous layers, and the immigration of tree species from surrounding forests.

Resistance of savanna ecosystems to environmental stresses such as drought, herbivory, and fire is associated with the development of underground organs. Most savanna ecosystems are characterized by large below-ground/above-ground ratios. Allocation of biomass to underground organs conserves carbon, energy and nutrients allowing the recovery of grass, shrub, and tree layers after regular, or even catastrophic events of drought, herbivory, and fire.

BIODIVERSITY AND THE EFFICIENCY OF RESOURCE UTILIZATION

Photosynthetic Types and Water-use Ef ciency

Combinations of rainfall and soils texture regulate moisture availability, determining gradients from wet to dry savanna environments. Savanna ecosystems extend from semi-arid regions to seasonally flooded areas because of the combination of high rainfall and heavy soils. Savannas are accordingly classified as *dry*, *intermediate*, *wet* and *swampy*, or *slightly seasonal*, *seasonal* and *hyperseasonal* (Table 1). The graminoid layer of the different savanna types falls into a well-separated taxonomic and physiologic pattern.

A few tribes of grasses have strong dominance in savanna environments: Panicae (e.g., *Axonopus*, *Digitaria*, *Panicum*, *Paspalum*, *Setaria*), Andropogonae (*Andropogon*, *Elionurus*, *Gymnopogon*, *Heteropogon*, *Hyparrhenia*, *Imperata*,

Table 1 Savanna types and predominant grass groups (based on Sarmiento, 1984, and Johnson and Tothill, 1985). Moisture regimes derive from a combination of rainfall, potential evaporation and soil texture

Moisture regime and soil texture	Savanna type	Dominant grass tribes
Wet (1000- 2000 mm), medium to heavy	Swamp grasslands Hyperseasonal Savannas	Panicoid- Oryzoid Panicoid- Andropogonoid
Seasonal (500- 2000 mm), medium to sandy	Savanna woodland Savanna parkland Savanna grassland	Panicoid- Andropogonoid
Dry (250- 800 mm), heavy to sandy	Low tree and shrub grasslands	Eragrostoid- Aristidoid- Chloridoid

Themeda, *Trachypogon*), Aristideae (*Aristida*), *Eragrostidae* (*Eragrostis*, *Dactyloctenium*, *Eleusine*) and Chlorideae (*Chloris*, *Microchloa*). The Andropogonae dominate the core of the savanna ecosystems, while the semi-arid savannas are dominated by representatives of the Chloridoid, Aristidoid and Eragrostoid tribes. The proportion of Panicoid increases strongly towards wetter savannas and the Oryzoids dominate in the swampy savannas, or esteros of South America. However, the Panicoid have the largest ecological range, and this tribe is heavily represented in all seasonal and hyperseasonal savannas.

The variation in dominance of different grass tribes from dry savannas to wet and swampy, hyperseasonal savannas, is associated to the photosynthetic metabolism. Carbon-4 (C4) grasses dominate the herbaceous layer of semi-arid and seasonal humid savannas, while in swampy savannas frequently C3 grasses reach the absolute dominance level, or share the niche with flood-tolerant C4 grasses (*see* **C_3 and C_4 Photosynthesis**, Volume 2).

Within the C4 photosynthetic type further subdivisions can be made according to the first stable acid synthesized after carbon dioxide (CO_2) fixation in the mesophyll, as malate formers, and aspartate formers. Aspartate formers predominate in dry savannas, while the malate former is the dominant photosynthetic type in humid savannas growing on dystrophic soils. This distribution is probably associated with water use efficiency, higher in C4 than in C3 grasses, and with higher nutritional requirements of C3 species and of the aspartate formers among the C4 species.

The predominance of the C4 pathway of photosynthesis within the single most important family in savanna environments suggests that the higher water use efficiency and productivity associated with this photosynthetic pathway provides competitive advantage in a hot, seasonally dry climate. Similar distribution patterns of photosynthetic pathways can be detected within the Cyperaceae (C4 types in seasonal and dry savannas, C3 types predominating in hyperseasonal savannas). However, among the dicotyledonous families, the C3 photosynthetic type predominates, C4 species being found only within the genera *Euphorbia* (Euphorbiaceae) and *Pectis* (Asteraceae). The Leguminosae, the most characteristic plant group of tropical savannas beside the grasses, belong homogeneously to the C3 photosynthetic type. When grown under optimal conditions, legume species always have lower photosynthetic capacity than C4 grasses. The variety and frequency of legumes shows that water stress is not the only determinant driving force structuring the herbaceous layer of savanna ecosystems.

Phenological Spread as a Means of Resource Partition

According to their periods of flowering, exponential growth, and maximum biomass accumulation, savanna grasses can be separated into groups of precocious, early, intermediate, and late growing species. The periods of exponential growth of early, intermediate, and late species do not overlap, and intermediate and late growers accumulate larger above-ground biomass than early growers do. The below-ground to above-ground biomass ratios decrease from the early growers, to intermediate, and late growers. Early growers have higher water use efficiency than late growers and are able to exploit the early pulse of nutrient supply resulting from the mineralization of organic matter at the beginning of the rainy season. However, they produce less organic matter than the late growers. Phenological spread has also implications for the variability of the grass layer in tropical savanna. The full set of phenological behaviors can be found in humid tropical savannas with relatively short dry seasons. In this vegetation the delay in the onset of rain results in decreasing production of the early growers, while a shorter duration of the rainy season affects production and reproduction of the late species.

Physiological Differentiation in Nutrient-poor Soils

Most neotropical seasonal savannas are dystrophic or hyperdystrophic, as characterized by a sum of exchangeable bases ranging from 1–5 and less than 5 meq per 100 g soil, respectively. Therefore, the herbaceous species are expected to be very sensitive to changes in nutrient availability, and have lower nutritional requirements for normal growth than species from grassland on eutrophic soils.

The nutritional factor may contribute to the grasses-sedges dominance observed in all savanna regions of the world, because in general these families have lower nutritional requirements than dicots, particularly the Leguminosae. In natural seasonal savannas in Venezuela, legumes are deficient in phosphorus (P) and potassium (K), while nitrogen (N) reaches similar values to those expected from cultivated forage legumes. By contrast, nutritional levels of the above-ground biomass of grasses at the peak of their development are only slightly below the level expected for cultivated forage grasses. Fertilization experiments conducted in savanna areas of Africa, Australia and South America show that additions of nitrogen, or nitrogen and phosphorus together, promote organic matter production of natural grass swards. Legume growth, on the other hand, is strongly stimulated by increases of the phosphorus supply, or amendments of the soil pH through the addition of calcium carbonate ($CaCO_3$). Higher phosphorus uptake increases growth because of improved photosynthesis and nodulation. Patches of higher pH, lower phosphorus fixation after organic matter mineralization, and a higher sum of bases favor legumes over grasses and sedges in the herbaceous layer. Nutrient patches, therefore, account for much of the diversity observed in different savannas.

Photosynthetic groups within the Poaceae show differences in nutrient requirements. Aspartate-former species

of the Andropogonae are favored against malate-formers by increased nitrogen availability, a response consistent with the hypothesis that aspartate formers have relatively higher nutritional requirements. This may also explain the widespread dominance of malate formers in seasonal savannas with dystrophic soils.

BIOLOGICAL TRAITS REGULATING NUTRIENT AVAILABILITY

Grass- tree Interactions

Trees and shrubs generally exploit deeper soil layers compared to grasses and other herbaceous species. Scattered trees and shrubs constitute areas of nutrient enrichment, at least in the upper soil layers. This may account partially for the differences in floristic composition, density, and productivity of the vegetation underneath woody perennials in several savanna areas. The presence of trees within a grassland also results in changes in light conditions, temperature, and evapotranspiration, factors powerful enough to explain the variations observed in the herbaceous vegetation. The richer nutrient environment underneath the tree components may allow the establishment of other woody species with higher nutrient requirements leading to the development of tree groups. In South American savannas, deciduous trees, presumably more nutrient demanding, are restricted to forest groves, while less demanding evergreen trees are found scattered within the grass layer of the savannas. However, this distribution may also result from differences in fire tolerance.

Mound-building Termites and Leaf-cutting Ants

Mound-building termites are common in all savanna regions of the world, although they are most significant in Northern Australia and Western Africa. During mound building, termites modify the physicochemical characteristics of the soils, in texture by selecting finer particles more appropriate for mound building, and in nutrient content because of the mixing with food residues. Termite species are efficient foragers of organic debris (nearly $1000\,kg\,ha^{-1}\,a^{-1}$), therefore they can deprive the surroundings of organic carbon input and the nutrients contained in the plant litter. In Australia, *Amitermes* and *Nasutitermes* mineralize approximately $250\,kg\,ha^{-1}\,a^{-1}$ of organic matter, and this may be an important contribution to nutrient recycling in nutrient-poor soils.

Leaf-cutting ants play a significant role in nutrient cycling in some Brazilian savannas. Burying large amounts of nutrient enriched litter may provide a source of nutrients for deep-rooted woody plants. This foraging activity sequesters significant amounts of nutrients from aboveground, avoiding their volatilization through fire at the end of the rainy season.

Rhizospheric Associations and Mutualistic Symbiosis

The predominance of soils with intrinsic low nutrient availability constitute the ecological background for the evolution of several symbiotic or quasi-symbiotic mutualistic associations in tropical savannas: (a) legume–*Rhizobia* symbiosis mediating nitrogen-fixation; (b) mycorrhizal symbiosis, improving phosphorus and water uptake through expanded soil exploration, and (c) rhizospheric microflora, enhancing nutrient availability through solubilization, or direct nitrogen fixation.

A large number of savanna leguminous species, particularly within the Fabaceae, are nodulated in the field. The nitrogen fixation activity of these nodules, however, is frequently limited by widespread phosphorus deficiency of savanna soils.

Phosphorus requirements of native legumes under natural conditions can be met through the formation of endo-mycorrhizal symbiosis. Savanna legumes develop frequently tripartite symbiosis with native *Rhizobia* strains, and mycorrhizogenous fungi under natural conditions. Enhancement of phosphorus uptake through mycorrhiza increases nodulation by native or introduced Rhizobia and improves nitrogen-fixation.

Rhizospheric associations with microorganisms of the type of *Azospirillum* are widespread in tropical savannas and they seem to be responsible for a significant amount of nitrogen supply to savanna soils.

Fixation of atmospheric nitrogen in savannas is also mediated by blue-green algae. In many savanna sites algal mats develop at the soil surface during the rainy season, with rates of nitrogen-fixation amounting to several $kg\,ha^{-1}$. Nitrogen flow into savanna soils mediated by algae and rhizospheric organisms seems to be large enough to offset partially the nitrogen losses caused by annual grass burning.

SUMMARIZING REMARKS

The following characteristics can be uniquely associated with tropical savanna vegetation, either as separate categories or as integrated properties:

1. Grasses with the C4 photosynthetic pathway dominate both the physiognomy and the production of organic matter in all seasonal savannas of the world. Within the grasses, an additional differentiation is observed regarding the dominance of aspartate formers in drier savannas, and malate formers in humid, seasonal savannas with soils of low nutrient availability. The relationships between photosynthetic pathways and nutritional requirements have not been experimentally established, but long-term field experiments show clearly that fertilization with nitrogen favors the production of organic

matter of aspartate-formers and reduces the competitiveness of malate formers in grass swards where both types coexist.

2. The typical savanna structure includes the coexistence of trees and grasses. The competition for resource acquisition, particularly water, between these two contrasting life forms is reduced by the differences in root distribution in the soil. Establishment of tree seedlings, however, might be limited to wet years, to allow the root of the tree seedling to cross the soil layer dominated by grass roots.
3. All savannas are characterized by the predominance of perennial species with strong development of underground organs, the root/shoot ratio being generally above one. This biomass allocation pattern is very well related to the tolerance of savanna vegetation to stresses such as drought, fire, and herbivory.
4. Diversity of developmental patterns of savanna species and the differentiation in nutritional requirements allow segregation in space and time of the demand for resources. Grasses in general appear to be more dependent on nitrogen supply, while legumes, and other dicots, are more limited by phosphorus supply.
5. Savanna environments being nutrient limited, emphasize the significance of symbiosis for nutrient acquisition. Legume–rhizobium–mycorrhizal symbioses, and rhizospheric associations with free-living nitrogen fixing bacteria, are very widespread, but their actual activity in nutrient supply at the ecosystem level has been measured only in a few cases. The activity of root symbioses together with the ubiquitous presence of blue-green algae, provide the basis for stabilizing nutrient budgets in savannas, particularly the recovery of nitrogen losses induced by extensive burning.

REFERENCES

Johnson, R W and Tothill, J C (1985) Definition and Broad Geographic Outline of Savanna Lands, in *Ecology and Management of the World's Savannas*, eds J C Tothill and J Mott, The Australian Academy of Science, Canberra in conjunction with C.A.B., Farnham Royal, Bucks, 1–13.

Sarmiento, G (1984) *The Ecology of Neotropical Savannas*, Harvard University Press, Cambridge.

FURTHER READING

Abbadie, L, Mariotti, A, and Menaut, J-C (1992) Independence of Savanna Grasses from Soil Organic Matter for their Nitrogen Supply, *Ecology*, **73**, 608–613.

Arias, I, Koomen, I, and Dodd, J C *et al.* (1991) Growth Responses of Mycorrhizal and Non-mycorrhizal Tropical Forrage Species to Different Levels of Soil Phosphate, *Plant Soil*, **132**, 253–260.

Balandreau, J, N'Dri Allou, R, and Villemin, G *et al.* (1976) Fixation Rhizosphérique de l'azote en Savane de Lamto, *Revue d' Ecologie et Biologie du Sol.*, **13**, 529–544.

Barrios, S and González, V (1971) Rhizobial Symbiosis in Venezuelan Savannas, *Plant Soil*, **34**, 707–719.

Belsky, A J, Amundson, R G, Duxbury, J M, Riha, S J, Ali, A R, and Mwonga, S M (1989) The Effects of Trees on their Physical, Chemical, and Biological Environments in a Semi-arid Savanna in Kenya, *J. Appl. Ecol.*, **26**, 1005–1024.

Boddey, R M and Döbereiner, J (1995) Nitrogen Fixation Associated with Grasses and Cereals: Recent Progress and Perspectives for the Future, *Fertil. Res.*, **42**, 241–250.

Bourliére, F, ed (1983) *Tropical Savannas: Ecosystems of the World*, Elsevier, Amsterdam.

Chapman, G P, ed (1992) *Grass Evolution and Domestication*, Cambridge University Press, Cambridge.

Cole, M M (1986) *The Savannas: Biogeography and Geobotany*, Academic Press, Harcourt Brace Jovanovich, London.

Coutinho, L M (1984) Aspectos Ecológicos Da Saúva no Cerrado.- A Saúva, as Qeuimadas e sua Possível Relação na Ciclagem de Nutrientes Minerais, *Bol. Zool., Univ. Sao Paulo*, **8**, 1–9.

Donaldson, C H, Rootman, G, and Grossman, D (1984) Long Term Nitrogen and Phosphorus Application to the Veld, *J. Grassland Soc. S. Afr.*, **1**, 27–32.

Eiten, G (1972) The Cerrado Vegetation of Brazil, *Bot. Rev.*, **38**, 201–241.

Frost, P, Medina, E, Menaut, J-C, Solbrig, O, Swift, M, and Walker, B (1986) Responses of Savannas To Stress and Disturbance, *Biol. Int., Spec. Iss.*, **10**, 1–82.

Goldstein, G, Sarmiento, G, and Meinzer, F (1986) Patrones Diarios y Estacionales en las Relaciones Hídricas de árboles Siempreverdes De la Sabana Tropical, *Acta Oecol., Oecol. Plant.*, **7**, 107–119.

Goodland, R and Pollard, R (1973) The Brazilian Cerrado Vegetation: a Fertility Gradient, *J. Ecol.*, **61**, 229–234.

Högberg, P (1986) Nitrogen-fixation and Nutrient Relations in Savanna Woodland Trees (Tanzania), *J. Appl. Ecol.*, **23**, 675–688.

Holt, J A and Coventry, R J (1990) Nutrient Cycling in Australian Savannas, *J. Biogeog.*, **17**, 427–432.

Huntley, B J and Walker, B H, eds (1982) *Ecology of Tropical Savannas. Ecological Studies 42*, Springer Verlag, Berlin.

Isichei, A O (1980) Nitrogen Fixation by Blue-green Algal Crusts in a Nigerian Savanna, in *Nitrogen Cycling in West-African Ecosystems*, ed T Rosswall, SCOPE-UNEP International Nitrogen Unit. Royal Swedish Academy of Sciences, Stockholm, 191–198.

Kellman, M (1979) Soil Enrichment by Neotropical Savanna Trees, *J. Ecol.*, **67**, 565–577.

Ludlow, M and Wilson, G L (1971) Photosynthesis of Tropical Pastures. I. Illuminance, CO_2 Concentration, Leaf Temperature and Leaf-air Vapour Pressure Difference, *Aust. J. Biol. Sci.*, **24**, 449–470.

Maasdorp, B V (1987) Contribution of Associative N_2-fixation (Acetylene Reduction) in some Grassland Ecosystems in Zimbabwe, *Soil Biol. Biochem.*, **19**, 7–12.

Medina, E and Silva, J F (1990) Savannas of Northern South America: a Steady State Regulated by Water-fire Interactions

on a Background of Low Nutrient Availability, *J. Biogeog.*, **17**, 403–413.

Muoghalu, J I and Isichei, A O (1991) Effect of Tree Canopy Cover on The Yield, Crude Protein and Fiber Content of Four Species in Nigerian Guinea Savanna, *Vegetatio*, **95**, 167–175.

Newman, E I, Child, R D, and Patrick, C M (1986) Mycorrhizal Infection in Grasses of Kenyan Savanna, *J. Ecol.*, **74**, 1179–1183.

Pomeroy, D E (1983) Some Effects of Mound-building Termites on The Soils of a Semi-arid Area in Kenya, *J. Soil Sci.*, **34**, 555–570.

Saif, S R (1986) Vesicular-arbuscular Mycorhizae in Tropical Forrage Species as Influenced By Season, Soil Texture, Fertilizers, Host Species and Ecotypes, *Angew. Bot.*, **60**, 125–139.

Sanford, W W, Usman, S, Obot, E O, Isichei, A O, and Wari, M (1982) Relationship of Woody Plants To Herbaceous Production in Nigerian Savanna, *Trop. Agric. (Trinidad)*, **59**, 315–318.

Sicardi De Mallorca, M and Izaguirre-Mayoral, M L (1993). A Comparative Evaluation of The Symbiotic N_2-fixation and Physiological Performance of Thirty-six Native Legume Species Collected in a Tropical Savanna During The Rainy and Dry Seasons, *Symbiosis*, **16**, 225–247.

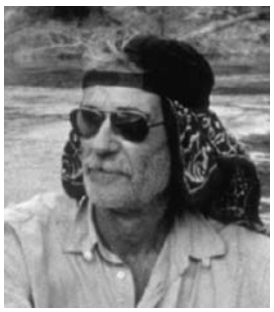
Solbrig, O T, Medina, E, and Silva, J, eds (1996) *Biodiversity and Savanna Ecosystem Processes, Ecological Studies 121*, Springer Verlag, Berlin.

Sprent, J I, Geoghegan, I E, Whitty, P W, and James, E K (1996) Natural Abundance of ^{15}N and ^{13}C in Nodulated Legumes and Other Plants in The Cerrado and Neighbouring Regions of Brazil, *Oecologia*, **105**, 440–446.

Tothill, J C and Mott, J, eds (1995) *Ecology and Management of the World's Savannas*, The Australian Academy of Science, Canberra in Conjunction with CAB, Farnham Royal, Bucks.

Walker, B H, ed (1987) *Determinant of Savannas*, IUBS Monographs Series No. 3, Chapter 3, IRL Press, Oxford.

Walter, H (1973) *Die Vegetation Der Erde in Öko-physiologischer Betrachtung. Band I. Die Tropischen und Subtropischen Zonen*, VEB Gustav Fischer Verlag, Jena.



Tucker, Jim

(1946–)

Compton James Tucker III, born 15 April, 1946 in Carlsbad, NM, received his BS in Biology (1969), MS (1973) and PhD (1975) from the College of Forestry of the University of Colorado, Ft. Collins. He has been affiliated with NASA's Goddard Space Flight Center in Maryland since 1975, first as a National Academy of Sciences post-doctoral fellow, and since 1977, as a physical scientist. His awards include the Exceptional Scientific Achievement Award (1986), Exceptional Scientific Achievement Medal (1987) and William Nordberg Memorial Award for Earth Sciences (1995), all of the National Aeronautics and Space Administration; the Henry Shaw Medal (1992) of the Missouri Botanical Garden, the National Air and Space Museum Trophy (1993), and the William T. Pecora Award of the US Geological Survey. He is also the first American co-recipient (with Mary L Cleave) of the Mongolian Friendship Medal, and a Goddard Senior Fellow.

Dr Tucker is a key pioneer in the use of satellite observations to study vegetation dynamics and its changes through time. He led the development and application of satellite observations of the normalized difference vegetation index (NDVI) for land surface studies. Because of the intrinsic difference in spectral properties of green leaves and of soils, the NDVI, expressed as the difference divided by the sum of the near infrared and visible reflectances, is a good correlate of photosynthetic activity of the land surface. The utility of the NDVI was first demonstrated by Dr Tucker with ground-based data collection and analyses using spectrometers and hand-held radiometers. He is instrumental in processing visible and near-infrared reflectance data from the advanced very high resolution radiometer (AVHRR) on board the polar-orbiting weather satellites (the NOAA-n series), and the production of the first continental and global observations of photosynthetic activity. These large-scale observations of the terrestrial biosphere were remarkable not only because they revealed a coherent structure over scales much larger than would be expected from the heterogeneity of the landscape, but also because the satellite instruments were not specifically designed for biospheric observations. The global multi-temporal observations of the NDVI have been used to classify and estimate areal extents of vegetation types, to study the photosynthetic drawdown of carbon dioxide (CO_2) from the atmosphere, to monitor the expansion and contraction of deserts and other xeromorphic vegetation, and to detect changes in the growing season at high-latitudes. It is now a routine observation of the terrestrial biosphere.

In addition to the AVHRR, Dr Tucker has also spearheaded the use of Landsat satellite data for studying land vegetation dynamics including deforestation, habitat fragmentation, and desert boundary determination.

By the 1990s, NDVI observations of the terrestrial biosphere had become routine, and widely used in large-scale studies of the terrestrial biosphere and its changes. These observations are no longer derived only from the AVHRR on weather satellites, but also from satellite instruments designed for biospheric observations. These include the sea-viewing wide field of view sensor (SeaWIFS) on board

the Seastar spacecraft launched in August 1997 and the moderate-resolution imaging spectroradiometer (MODIS) instrument on board Terra satellite launched in December 1999.

INEZ FUNG *USA*

Tundra

Philip A Wookey
Uppsala University, Uppsala, Sweden

Lying beyond the poleward or altitudinal limit for the growth of trees, tundra ecosystems exhibit much greater diversity in community type, soil characteristics, geomorphology and topography than is commonly perceived. They also occupy a broad environmental space (range of climatic conditions, expressed in terms of temperature and precipitation, under which tundra systems develop), from extreme polar deserts of, for example, the Russian Arctic archipelagos, northern Canada and Greenland to the closed-canopy tundra of the Low Arctic, and the alpine tundra of mountain regions. Although net primary productivity (NPP) is generally low, there is a very large range of mean values (from 1–375 g m^{-2} $year^{-1}$ total above- and below-ground NPP for high arctic polar deserts and low arctic low shrub tundra, respectively), again reflecting the broad heterogeneity of tundra. In spite of low NPP, tundra systems are key feeding and breeding grounds for large populations of grazing mammals (e.g., reindeer in Eurasia and caribou in North America and Greenland, muskoxen, lemmings and voles) and birds (e.g., snow and barnacle geese) together with their associated predators and scavengers (e.g., wolves, foxes and raptors).

Tundra ecosystems are of global biogeochemical significance since, on balance, they have been removing carbon dioxide (CO_2) from the atmosphere during the last 10 000 years and storing it as soil organic matter (although recent research suggests that this situation may now be changing), while releasing methane (CH_4). They also influence the global energy budget by reflecting significantly more solar energy back to outer space than the boreal forests to the south. Tundra is thus coupled to the global climate system and also to atmospheric composition (CO_2 and CH_4 concentrations, in particular). Ecosystem responses to environmental change in tundra therefore have potential to feed back on further change through alterations in net fluxes of radiatively forcing (greenhouse) gases or through changes in surface roughness and reflectance (albedo).

General circulation models (GCMs) suggest that climatic change in Arctic tundra regions could be of greater magnitude than the average predicted changes for the Earth as a whole. Many ecologists also hypothesize that any given change in temperature will have a greater biological impact in cold systems than in warmer ones, due to the disproportionate impact on heat sum (degree days). It has also been suggested that, since some tundra species may be surviving at or near the geographical limits of their range, they may therefore be particularly sensitive to an improvement or deterioration in environmental conditions. Climatic changes, together with other key facets of global environmental change (rising atmospheric CO_2 concentrations, increasing deposition of airborne nitrogen (N)-containing pollutants, and possible increased fluxes of ultraviolet (UV)-B radiation to the surface) all have the potential, both singly and in combination, to alter the structure and functioning of tundra systems.

BACKGROUND

The term *tundra* is derived from the Sámi word *tundar* (in Finnish *tunturi*), which refers to a treeless hill, or elevated ground, in northern Fennoscandia. The high latitude tundras (*Arctic tundra*) of the Northern Hemisphere are by far the largest in the tundra biome, covering around 5.6 million km^2. At lower latitudes, tundra ecosystems are found at higher elevations associated with mountain regions (*alpine tundra*) and in equatorial regions they occur above the tree line at altitudes higher than 4000 m. Alpine tundra occurs across a very great range of latitudes and geological settings and their environments (photoperiod, net receipt of solar radiation, diurnal temperature ranges, length of growing season) are correspondingly diverse. A notable feature of alpine tundra is that it is often geographically highly fragmented and occurs as 'islands' separated by warmer ecosystems of lower elevation. Topographic controls often result in better-drained soils in alpine tundra than in arctic systems, and alpine tundra often also lacks underlying permafrost. Tundra ecosystems are not, therefore, homogeneous (Figure 1).

For arctic tundra, several distinct vegetation zones have been recognized, with forest-tundra representing the (often diffuse) transition between the boreal forest (*taiga*) and tundra biomes. Bliss and Matveyeva (1992) have sought to unify the North American and Russian classification schemes. With increasing climatic severity, forest–tundra grades into several contrasting Low Arctic systems including *shrub tundras* with an open canopy of low shrubs some 40–60 cm high (comprising species such as dwarf birch, *Betula nana*, shrub willows and alders) and an almost complete ground cover of dwarf heath shrubs, abundant foliose lichens (notably *Cladonia* spp., *Cladina* spp., and *Cetraria* spp.), mosses, grasses and forbs. Such Low Arctic shrub

communities represent the most productive tundra types and they occur in favorable environments just north of the tree line. At the other end of the spectrum, and found in areas with a thaw period as short as six weeks per year with average temperatures for the warmest month not exceeding 2–6 °C, are the High Arctic polar deserts and semi-deserts with their incomplete and diminutive vegetation cover of cushion or mat-forming plants (e.g., mountain avens, *Dryas octopetala*, and purple saxifrage, *Saxifraga oppositifolia*) creeping dwarf shrubs (e.g., *Salix arctica*), rosette-forming plants, mosses, graminoids (grasses and sedges), lichen and cyanobacterial crusts. The extreme High Arctic ecosystems have very low precipitation totals (often with a pronounced summer peak) and winter snow cover is, paradoxically, often thin (compared with much deeper snow in many Low Arctic tundras farther south): surface drying during the thaw period can create significant water stress here, particularly if substrates are freely draining.

Although rates of NPP are generally low in tundra ecosystems compared with the other major biomes, the reservoirs of organic material in tundra soils are disproportionately high, especially in poorly drained regions of the Arctic tundra (Figure 2). Tundra soils worldwide are estimated to contain between 121–191 Gt of carbon (C) (ca. 11–14% of the global total soil carbon reservoir, although tundra systems cover ca. 5–6% of the land area).

Figure 2 Arctic tundra, North Slope of Alaska (68 °N). The picture is taken in the foothills of the Brooks Range and shows tussock tundra, where hare's tail cotton sedge (*Eriophorum vaginatum*) and dwarf birch (*Betula nana*) are important elements in the vegetation community. The region is underlain by continuous permafrost and the soils are poorly drained. The combination of cold and wet conditions restricts decomposition processes so that soil organic matter contents are high. The mean temperature of the warmest month is around 10 °C (July)

GLOBAL ENVIRONMENTAL CHANGE AND TUNDRA

Tundra systems have been receiving considerable attention from researchers interested in the potential impacts of global environmental change. The output from GCMs points to a greater than average warming at high northern latitudes, especially during the winter months. These predictions, together with the potential increases in UV-B fluxes to the surface (associated with the already observed springtime depletion of stratospheric ozone at high latitudes) and the increased deposition of, in particular, nitrogen(N)-containing pollutant compounds originating from industrial regions, are all facets of global environmental change that have implications for the structure and functioning of tundra systems. The scenarios of environmental change in tundra regions are complex and there is considerable uncertainty regarding the magnitude and even, for some parameters (e.g., precipitation quantities and form – rain or snow) the direction of change. The key scientific issues, however, are introduced through reference to major research topics and case studies. Tundra ecosystems are also directly impacted by mineral and fossil fuel extraction, transport infrastructure, animal husbandry (e.g., reindeer herding) and tourism: the effects of global environmental change must be viewed within the context of these local and regional disturbances, which under some circumstances will strongly override, or interact with, the impacts of global change (see Crawford (1997) and articles therein).

Experiments with Ecosystems

In tundra it is possible to conduct experiments in plots of manageable size (a few square meters or even smaller) in which the environment can be manipulated to simulate the impacts of global environmental change on well-defined community types. Some notable experiments that have sought to understand the potential responsiveness and resilience of tundra plant species, vegetation communities and soils to climatic warming include studies at Toolik Lake on the North Slope of Alaska (Chapin *et al.*, 1995), the International Tundra Experiment (ITEX) (Henry and Molau, 1997; Arft *et al.*, 1999), and the comparative studies on European Arctic (Swedish Lapland and northwest Spitzbergen) ecosystems by Malcolm Press and colleagues (see Press *et al.* (1998) and references therein). In all of these examples, air (and to a lesser extent soil) temperature during the thaw period has been manipulated quite simply using either plastic greenhouses or hexagonal open-topped chambers to give a warming of 1–3.5 °C above ambient temperatures. The work at Toolik was initiated in 1981, and provides the longest time series, while the ITEX and other European Arctic studies began in the early 1990s but cover a broader range of arctic and alpine sites.

These experiments have together highlighted some key aspects of tundra response to warming: (i) plant species respond individualistically to warming (and, as shown in the Toolik experiments, fertilizer additions and shading treatments), making generalizations about growth form (e.g., deciduous or evergreen dwarf shrubs, forbs or graminoids) difficult, and extrapolating conclusions from one tundra community type to another potentially misleading; and (ii) responses to warming generally decline over time as some other resource (other than thermal energy) becomes limiting (e.g., soil nutrients). The Toolik studies, in addition, indicate that the individualistic responses of plant species to alterations in temperature (and also nutrient availability) generally cancel one another out and result in little overall change in community NPP and biomass (Terry Chapin and colleagues suggest, accordingly, that the system is buffered against change: Chapin *et al.*, 1995). How general this phenomenon is throughout the tundra biome is hard to judge without comparable data from other systems, but it does suggest a degree of resilience to environmental variability, at least in these Low Arctic systems with their complete vegetation cover (such observations were also supported by the work in Swedish Lapland).

Experiments being conducted by the ITEX team and in the European study, although focused principally on individual plant species of broad circumpolar arctic and alpine distribution, have allowed geographical comparisons to be made, and have also illustrated the detailed basis of individual species responses to warming. What can be concluded at this stage from the work is that (i) plant reproductive development is accelerated and seed output is increased markedly with warming, especially in the High Arctic; and (ii) vegetative growth at most sites was increased in the early years of the experiment, but this response declines with time and may not be sustained in the longer term (due, perhaps, to nutrient limitations). For many vascular plant species in Arctic Tundra (most of which are perennials) the ability to respond rapidly to warming, in terms of reproductive development, is important because flowering, pollination and seed-set must be completed rapidly before the onset of winter. Vegetative growth, by contrast, can be spread over many years and may be as much constrained by nutrient availability as growing-season duration and temperature. Increased success of seed set in response to warming may have particularly profound implications for tundra ecosystems in the High Arctic, where large areas of incomplete vegetation cover may provide opportunities for seedling establishment and colonization of bare ground. In one experimental study in a polar semi-desert community in northwestern Spitzbergen, the author and colleagues (1995, cited in Press *et al.*, 1998) observed a seven-fold increase in the output of viable seed of mountain avens (*D. octopetala*) in response to a warming of 3.5 °C above ambient (5.4 °C) summer temperatures. Whether such increased production

of viable seed could actually translate into colonization of bare ground in the High Arctic will, however, depend on the availability of suitable microsites (so-called *safe sites*) for seedling germination and establishment, and suitable moisture, nutrient and temperature regimes for plant survival in the longer term. A rapid greening of High Arctic tundra areas in response to global warming is possible, although restricted nutrient availability may, again, play a key role in constraining this process.

Experiments designed to assess the impacts of increased fluxes of UV-B radiation (Lars Olof Björn and colleagues, Chapter 11 in Woodin and Marquiss, 1997), and rising atmospheric CO_2 concentrations (Oechel *et al.*, 1994; Körner *et al.*, 1997) on tundra have generally been limited to locations with good infrastructure and electrical power supplies (e.g., the Abisko area in Swedish Lapland, Toolik Lake field station in Alaska and Furka Pass in Switzerland). For elevated CO_2 concentrations, there are consistent indications that plants rapidly acclimate and rates of community NPP down-regulate to levels comparable with plants growing at ambient CO_2 concentrations: Such homeostatic adjustment is thought to reflect limitations in sink strength for photosynthates caused by nutrient limitations to plant growth, and has been seen in Alaskan tussock tundra by Walter Oechel and colleagues (Oechel and colleagues, Chapter 12 in Woodin and Marquiss, 1997) and in an alpine grassland in Switzerland by Christian Körner's research group (Körner *et al.*, 1997). In the sense that mineral nutrients may play a key role in modulating tundra responses to elevated CO_2 concentrations, there are parallels with the nutrient limitations to sustained temperature responsiveness. For UV impacts, it is much more difficult, at this stage, to gain a clear picture of potential ecosystem responses because measured effects depend greatly on the duration of the study (for UV-B they are often cumulative, and therefore potentially more serious for perennial and evergreen plant species than for annuals and deciduous species) and they are also species-specific. Plant responses to changing CO_2 concentrations and UV-B fluxes can also, of course, have knock-on effects at other trophic levels: changes in plant tissue chemistry and physical structure in response to these factors could, for example, influence decomposition processes (and hence nutrient recycling) and patterns of both vertebrate (e.g., caribou, reindeer, geese, lemmings and microtines in Arctic tundra regions) (Figure 3) and invertebrate herbivory (e.g., by aphids or woolly bear caterpillars). It is too early to make reliable predictions regarding the significance of these indirect effects for tundra ecosystems.

Alterations in timing of plant growth and flowering patterns in response to all facets of global environmental change in tundra could also have repercussions at other trophic levels: the activities of pollinators and herbivores may be particularly sensitive to these changes in tundra ecosystems due to the highly seasonal environments and short growing seasons. In one experimental study on Spitsbergen, for example, Andy Strathdee and colleagues noted that artificial warming (of 2.8 °C above ambient temperature) produced a dramatic 11-fold increase in the egg density of a sap-sucking aphid (*Acyrthosiphon svalbardicum*) on the host plant *D. octopetala*, partly because the aphid has a flexible life cycle and was able to complete three sexual generations instead of two in response to the additional 215 degrees days of thermal energy produced in the warming treatment (Bale and colleagues, Chapter 6 in Woodin and Marquiss, 1997).

Figure 3 Tundra ecosystems are important feeding and breeding grounds for both mammalian and avian herbivores. Here, a barnacle goose (*Branta leucopsis*) sits on the nest on Storholmen Island, Kongsfjord, Spitsbergen. The small island provides protection from the arctic foxes (*Alopex lagopus*) which exert a severe impact upon breeding groups on the nearby Br gger Peninsula. The Spitsbergen barnacle geese over-winter in the British Isles

In general much more work is needed to improve our ability to predict the implications of global environmental change on tundra ecosystems at all trophic levels. Experimental studies to date have focused principally on net primary producers and on single facets of environmental change or simple combinations. Clearly, single environmental factors do not change in isolation, and thus ecosystem ecologists need to consider more complex interactions and scenarios of change. Additionally, the time-scales of system response (Figure 4) are of key significance and short-term studies have only limited potential to improve our predictive ability.

Tundra Carbon Fluxes and Links with Global Biogeochemistry

Carbon Dioxide

On balance, tundra soils have been a net sink for carbon (as soil organic matter) during at least the Holocene (the post-glacial epoch covering the last 10 000 years), capturing carbon at a rate of between 0.1 and 0.3 Gt C year^{-1}.

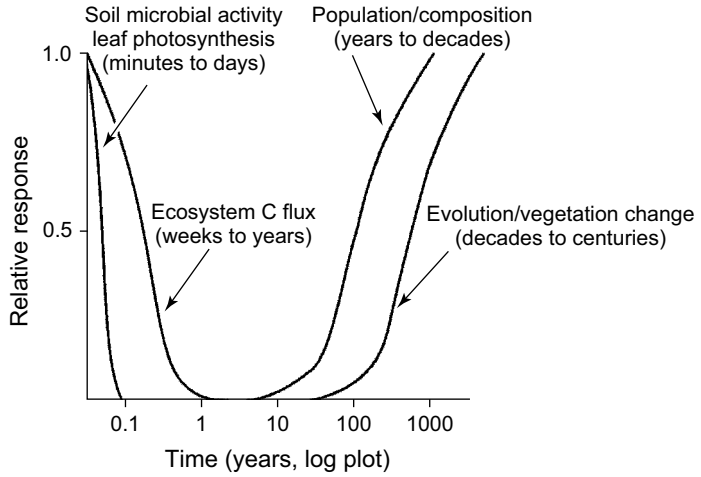


Figure 4 Hypothetical relative response times for adjustment to new atmospheric CO_2 concentrations and climatic change, for changes at the level of leaf photosynthesis and soil microbial activity, ecosystem net CO_2 ux, population structure, community composition, vegetation and evolution. (Reprinted from Oechel and Vourlitis, 1994 with permission from Elsevier Science and from the authors)

The combination of waterlogging (in many areas drainage is impeded by the presence of permafrost beneath the seasonally thawed active layer), low temperatures and poor litter quality reduces the activity of decomposer organisms in soils and has resulted in the accumulation of poorly decomposed peaty organic material, particularly in tundra wetlands and mires. Carbon accumulates when rates of NPP exceed rates of decomposition, and this pattern has been a feature of ecosystem development in wetter tundra areas. One hypothesis that was advanced by Dwight Billings and colleagues (Billings *et al.*, 1982), early in the debate on global environmental change and ecosystem carbon balances, was that decomposition in cold, wet and largely anaerobic soils would be markedly responsive to an increase in temperature and/or a decrease in the depth of the water table (drying). Warmer and drier conditions should favor more rapid metabolism in the decomposer community (see Oechel and colleagues, Chapter 12 in Woodin and Marquiss, 1997). Since decomposition processes release CO_2 back to the atmosphere, it was hypothesized that climate warming could result in many tundra and boreal ecosystems switching from being a net sink for atmospheric CO_2 to being a source, thus contributing, as a positive feedback effect, to the anthropogenic increases in atmospheric CO_2 concentrations, and thereby to further global warming.

This topic remains of key interest but there is considerable controversy, still, regarding (i) the timescales over which changes in the direction of CO_2 flux can be sustained; (ii) the actual magnitudes of contemporary tundra soil carbon reservoirs; (iii) the geographical extent of regions currently emitting carbon; (iv) the potential influence of elevated atmospheric CO_2 concentrations themselves on ecosystem processes controlling carbon fluxes; and (v) even the methodologies used to measure net fluxes. Major research work casting light on this problem has been undertaken by Walter Oechel and colleagues, and field measurement campaigns of carbon flux (particularly on the North Slope region of Alaska) have revealed significant net losses of carbon from tundra to the atmosphere in recent years. In 1993 Oechel and his associates reported rates of carbon loss amounting to 156 and $34\,g\,C\,m^{-2}\,year^{-1}$ for tussock tundra (Figure 2) and wet coastal tundra, respectively, based on measurements during the thaw period, but suggested that losses could be even greater due to continuing soil respiration in the autumn and early winter in unfrozen soils beneath the insulation provided by a snow-pack (Oechel *et al.*, 1993). Subsequent winter CO_2 flux measurements, both in Arctic and alpine tundra regions, have confirmed the importance of continuing soil respiration, even beneath the snow, which cannot be ignored in tundra carbon budgeting. Oechel *et al.* (1993) hypothesized that the cause of the switch from sink to source of CO_2 on the North Slope may relate to a regional warming (of $2-4\,^{\circ}C$) over the last few decades, resulting in increased evapotranspiration and surface drying of soils: such drying could increase penetration of oxygen into the soil profile and accelerate aerobic metabolic activity in the decomposer community.

The debate on feedbacks between climate change and tundra carbon fluxes again illustrates the importance of time-scales and geographical coverage. Several lines of evidence suggest that substantial net losses of carbon from tundra ecosystems to the atmosphere as a result of warming and drying may be transitory (see Figure 5): with increased carbon loss there should, in the medium to longer term (years to decades) be an associated increase in the rate of mineral nutrient (e.g., nitrogen and phosphorous) release through the process of mineralization (although this simple

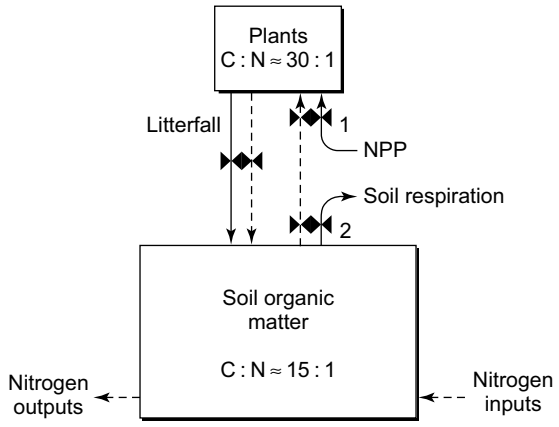


Figure 5 A conceptual model (Shaver *et al.*, 1992) of carbon/nutrient interactions in terrestrial ecosystems, using carbon and nitrogen as an example. There are two major pools of organic matter in the ecosystem (plants and soil organic material). Carbon uxes into and out of these pools are indicated by solid lines and nitrogen uxes by dashed lines. Bow ties represent links between the carbon and nitrogen uxes. Bow tie no. 1 implies that net carbon uptake by plants (NPP) is constrained by the ability of the plant to take up nitrogen, and vice versa. The suggestion made by Shaver *et al.* (1992) is that in nutrient-limited ecosystems such as tundra, the great majority of plant nitrogen supply comes from decomposition (mineralization) of soil organic matter, and that nitrogen mineralization is linked with losses of carbon due to soil respiration (bow tie no. 2). The overall carbon balance of the system is largely determined, therefore, by carbon gains associated with nitrogen uptake and losses associated with nitrogen release. In an environmental change scenario, therefore, any factors in uencing the uxes of nitrogen or carbon (e.g., changes in soil temperature or moisture regimes, atmospheric CO_2 concentrations and temperature, deposition of airborne nitrogen-containing compounds) can have implications for net carbon balance. (Reprinted from Shaver *et al.*, 1992, with permission from the authors)

linkage has not been demonstrated experimentally) (*see* **Nitrogen Mineralization**, Volume 2); in a system where such nutrients are limiting for the rate of photosynthesis, this increased availability to plants should stimulate carbon fixation and thus NPP; such responses could bring the system back from being a temporary source to a sink for carbon, as hypothesized by Shaver *et al.* (1992). Evidence in support of this can be found in the paleoecological record of organic matter accumulation rates, while ecosystem modeling activities (see McKane *et al.*, 1997), simulating controls on carbon and nitrogen fluxes in tussock tundra, also suggest that large losses of carbon may be transitory in nature. Indeed, an update by Oechel *et al.* (2000) reported that continued climate warming and soil drying in the North Slope region of Alaska has diminished the strength of CO_2 emissions (compared with those reported in 1993), although they note that the systems studied currently remain a net source of CO_2 to the atmosphere. Further field flux measurements in contrasting ecosystems also emphasize the importance of studying a range of tundra community types: extrapolations temporally or spatially have substantial risks attached.

Methane

Although carbon fluxes between tundra and the atmosphere are quantitatively dominated by CO_2, and they have received major attention from researchers, the net release of CH_4 from wetter tundra ecosystems to the atmosphere is another process that strongly couples tundra systems to atmospheric chemistry. Methane is present in the atmosphere at concentrations more than two orders of magnitude lower than CO_2, but it is increasingly of concern as a powerful greenhouse gas because, molecule for molecule, CH_4 is 21 times more radiatively forcing than CO_2. Best estimates of CH_4 flux from wet tundra range between $18-35\,\mathrm{Tg\,year^{-1}}$, but these are based upon data from restricted sample points. The potential impacts of global environmental change on net CH_4 fluxes between tundra ecosystems and the atmosphere are receiving growing attention, because atmospheric CH_4 concentrations are increasing by about 1% per year and CH_4 emissions from tundra represent almost 25% of the total global emissions from natural sources (e.g., including such sources as ruminant animals and termites). One concern is that climate warming could significantly increase rates of the anaerobic bacterial process of methanogenesis in wet tundra and mire ecosystems. This process is thought to be highly temperature-dependent and could result in significantly greater fluxes of CH_4 from tundra to the atmosphere, thus acting as a positive feedback on further environmental change. Predicting the effect of environmental change on net CH_4 fluxes is, however, highly complicated due to several interacting factors (see Figure 6). For example (i) net fluxes of CH_4 are not only temperature-dependent, but are also strongly influenced by soil hydrology and oxygen status (these factors interact because waterlogging severely restricts O_2 diffusion to depth in soils); and (ii) in aerobic soils there is also a group of bacteria called methanotrophs, which actually consume CH_4 and use it as an energy and carbon source. Thus, warming without any increase in precipitation could result in soil drying, increased O_2 availability at depth, and both increased CO_2 emissions (due to increased aerobic decomposition: as described earlier) and reduced CH_4 emissions (due both to reductions in the rate of methanogenesis and increases in the rate of methane consumption by methanotrophs, Figure 6). Assessing the overall implications for radiative forcing due to greenhouse gases is challenging, therefore, because tundra CO_2 and CH_4 fluxes can respond quite differently to the same environmental factors. Warming combined with increased precipitation is, however, likely to increase net emissions of CH_4 from tundra to the

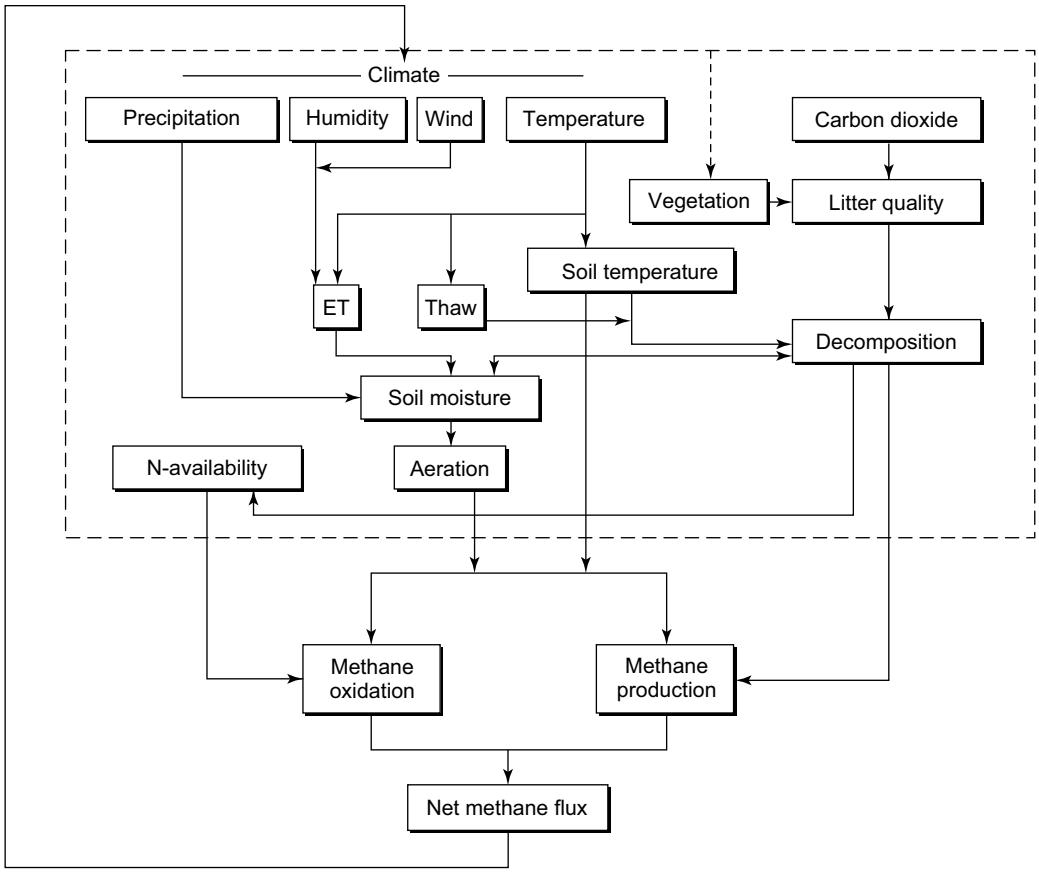


Figure 6 Some of the ecological links between climatic change and methane production and consumption in natural ecosystems. The dashed box indicates the many factors in uencing vegetation; ET, evapotranspiration. Note the suite of environmental variables that interact to in uence soil temperature and moisture regimes, and thus also aeration status. The aerobic processes of methane oxidation (methanotrophy) and the anaerobic process of methane production (methanogenesis) are strongly in uenced by soil moisture and aeration status, which in turn are in uenced by general climate. The balance between these two bacterial processes in uences net CH_4 ux, as does the effect of decomposition (supplying methanogenic bacteria with raw materials for CH_4 production, and consuming O_2 in the soil). (Reprinted from Torn and Chapin, 1993, with permission from Elsevier Science and from the authors)

atmosphere, as described in a modeling study by Christensen and Cox (1995). As yet it is impossible to predict the impacts of increasing atmospheric CO_2 concentrations and UV-B fluxes at the surface on net CH_4 fluxes, although experimental studies in, for example, temperate mire systems suggest that elevated atmospheric CO_2 concentrations could influence methanogenesis through alterations in the amount and chemistry of materials exuded from plant root systems (methanogenesis requires a supply of organic materials and decomposition products).

Nitrogen Compounds

In tundra ecosystems generally, one facet of global environmental change that has not received adequate attention is the global increase in deposition of nitrogen-containing airborne compounds: this is a striking omission, particularly since ecosystem modeling studies by Melillo *et al.* (1993) support the view that climate change impacts on NPP in northern ecosystems will be modulated via the effects of temperature on nitrogen availability (e.g., though rates of nitrogen mineralization in soils). Evidence from ice core analysis also indicates that there has been a 100% increase in nitrogen deposition at remote stations (e.g., central Greenland) since the 1950s: clearly the cumulative inputs of this biologically available nitrogen to tundra ecosystems have considerable potential to influence soil nitrogen reservoirs, NPP, net carbon fluxes and species composition (see Woodin and Marquiss, 1997 for a review).

Surface Energy Exchange and Links with the Global Radiation Budget

The energy budget of tundra ecosystems is of global environmental significance. Arctic tundra ecosystems, in particular, have a cooling influence globally, because when snow-covered, they reflect incoming solar radiation strongly

back to space. This contributes significantly to the 30% of incident solar radiation that is reflected (e.g., by snow, clouds and atmospheric particulates) without causing surface warming. Just how important tundra ecosystems are (and have been during the Holocene) to the global radiation budget and for regional climate has been demonstrated convincingly by a climate modeling study of Foley *et al.* (1994). In this study, a GCM was used to simulate the climate of the high northern latitudes (between 60–90 °N) during the mid-Holocene (about 6000 years ago), a period known to be warmer in these regions than the present. First the team ran the computer simulation as a control, with input parameters (in terms of solar energy receipts) reflecting contemporary orbital characteristics of the Earth. This provided a baseline for comparison with two experimental simulations. Exp-A was a simulation in which the orbital parameters (eccentricity – or orbital stretch – around the sun, degree of tilt of the Earth's axis, and the seasonal timing of the perihelion – the point in the orbit at which Earth is closest to the sun) were set for 6000 years ago: The changes result in an increased receipt of solar energy during the northern summer and a delay in the formation of sea ice in autumn and winter, resulting in year-round warming (relative to the control) of 1.8 °C.

Exp-B, by contrast, not only altered the orbital parameters, but also simulated an expansion of boreal forest onto what is currently tundra, in line with paleobotanical evidence for a northward extension of the tree line during the middle Holocene. This expansion of forest onto tundra produced an additional annual average warming of 1.6 °C (leading to a total of 3.4 °C warming). The explanation for this is that the modest warming resulting from changes in orbital parameters allowed the northern limit of the boreal forest to expand onto tundra: this expansion resulted in a change in vegetation canopy structure and a reduction in reflectivity (albedo), thus allowing more solar radiation to be absorbed, especially in the springtime (with lengthening days but a complete snow-cover on the adjacent tundra). The replacement of southern tundra with boreal forest thus accentuated warming caused by orbital factors. The significance of the vegetation forcing of climate, compared with the orbital forcing, is shown in Figure 7, with the springtime peak clearly evident. In the future, a significant northward expansion of boreal forest onto tundra, together with the associated changes in albedo, and possibly reinforced by reduced snow-cover in springtime on remaining open tundra, could act as a strong positive feedback on further global warming.

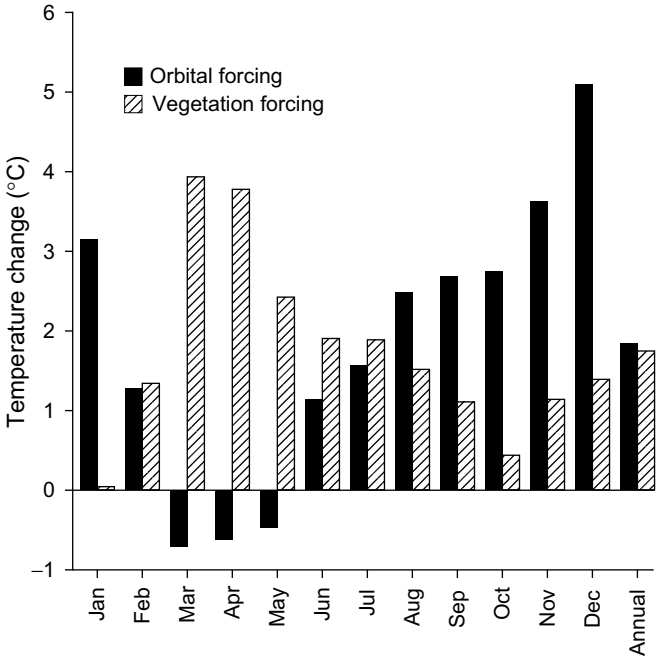


Figure 7 Changes in zonal average (over land from 60 to 90 °N) monthly surface temperatures due to orbital and vegetation forcing. The vegetation forcing component illustrates the role of boreal forest encroachment onto tundra during the mid-Holocene (6000 years before present) in reinforcing warming caused by orbital changes. The forest acts to lower albedo and re ect less solar energy to space compared with snow-covered tundra: the result is more solar energy absorbed and greater warming. (Reprinted by permission from Foley *et al.*, 1994, with permission from the authors and Macmillan Magazines)

Issues Relating to Biodiversity and System Stability

Tundra systems are frequently said to be *fragile*, particularly in terms of the impacts of physical disturbance. The general validity of this assertion in relation to global environmental change is, however, open to question, particularly in the light of major changes during the Quaternary and the ongoing dynamics of tundra grazing systems and physical disturbance associated with cold-region geomorphological processes (Figure 8).

Interspecific biodiversity in tundra ecosystems is low compared with many other major biomes, but within-species (intraspecific) diversity of vascular plants can be high, potentially pre-adapting plants to environmental change. For several key circumpolar arctic and alpine tundra species (e.g., *D. octopetala*, and *Saxifraga oppositifolia*), there are distinct *ecotypes* (ecotype – a locally adapted population of a widespread species – Oxford Dictionary of Botany, 1992), occupying highly contrasting micro-environments (often topographically controlled) such as snow-beds or ridge-tops. The classic study on ecotypes of *D. octopetala* by McGraw and Antonovics (1983) has demonstrated that although an ecotype growing on exposed ridge-tops (*D. octopetala octopetala*) is able to hybridize with *D. octopetala alaskensis*, occupying nearby sheltered snow-bed habitats, this may be very rare in nature due

Figure 8 A high arctic environment, Br gger Peninsula, Spitsbergen (79°N). The mantle of rock debris is actively churned by freezing and thawing processes (cryoturbation), resulting in the sorting of different sized rock material to produce patterned ground. Here primary productivity is very low and physical disturbance is of major signi cance. The ne-grained materials (darker tone, in foreground) support a cyanobacterial crust, which provides a source of inorganic nitrogen (through biological xation) and may help to stabilize the surface. If the cryoturbation is not too severe, some mosses and vascular plants (such as purple saxifrage - *Saxifraga oppositifolia*) may be able to establish themselves. The mean temperature of the warmest month is around 4.5°C

to separation in the flowering period of the ecotypes (on ridges flowering is earlier due to rapid melting of the thinner snow cover). Morphological differences in the ecotypes are maintained when they are grown under the same (common garden) conditions, demonstrating a genetic basis for the differences. In a scenario of environmental change, the presence of the two ecotypes should confer survival potential through shifts in the relative abundance of one ecotype compared with the other. Thus with greater winter/early spring snowfall, there may be stronger selection for *D. octopetala alaskensis*, the snow-bed ecotype, while a longer growing season, or thinner winter snow-cover may benefit *D. octopetala octopetala*.

Those tundra ecosystems currently characterized by incomplete vegetation cover and much bare ground (e.g., many of the High Arctic and Alpine tundras), are of interest in relation to global change scenarios and the biodiversity debate because they provide the space for new species to move in from farther south or from lower elevations (in an alpine setting). Competition may be less intense in these extreme environments, allowing immigration and, at least temporarily, an increase in species richness. This process of immigration has been noted by Grabherr *et al.* (1994) for several mountain summits higher than 3000 m in the central Alps (in western Austria and eastern Switzerland), based on contemporary census work, and historical data (from 1895–1918 and 1947–1953). Temperature records indicate a warming of 0.7°C in the region over the last 70–90 years. Although species richness has increased, for some plant species at higher elevations in mountain regions, there may be nowhere left to migrate, and increasing interspecific competition from less cold-tolerant species may pose a threat in future.

CONCLUSION

Tundra ecosystems are of major interest and importance in relation to global environmental change. These systems are strongly coupled to the global climate system via fluxes of radiatively forcing trace gases and through their influence on surface energy budgets. If current scenarios of high latitude climate change are well founded, then the aerial extent of the tundra biome is likely to decrease substantially over the coming centuries (possibly even by 32%), and Low Arctic tundra types will be forced northwards. In the High Arctic the immigration of species from farther south, and the increased colonization of bare ground, may substantially alter NPP, stocks of organic matter in biomass and soils, biodiversity and the surface energy balance. The rate of change, however, and the degree to which it could be constrained by lack of mineral nutrients, or of unsuitable surfaces for colonization and establishment, is unknown, although the limited paleoecological information from the High Arctic suggests this to be a slow process (with many systems still possibly readjusting to warming following termination of the Little Ice Age).

Invasion of Low Arctic tundra systems by tall shrubs and trees could be a much more rapid process due to infilling of tundra by expansion of boreal forests currently found in sheltered river-valleys and extending well into the tundra: such processes could be particularly rapid in the very broad forest–tundra ecotone. In organic rich soils of the Low Arctic, any temperature-induced increase in rates of nutrient mineralization could also translate within a few years or decades into increased NPP and a switch to more nutrient-demanding shrub and tree species.

Dynamic responses of tundra ecosystems to single and multiple facets of environmental change will have impacts on net fluxes of important trace gases. What is particularly difficult to judge at present is how long such readjustments take, and how persistent imbalances in, for example, net CO_2 fluxes, will last. Tundra systems have experienced rapid environmental change in the past, and they continue to be subject to marked fluctuations in physical environment (e.g., associated with flooding, ice encasement and summer snowfalls) and in numbers and distribution patterns of secondary producers (invertebrate and vertebrate herbivores) and consumers. The combination of drivers of global environmental change, can, however, be expected to result in adjustments in the structure, function and aerial extent of

tundra ecosystems globally in the future: these effects can also, arguably, manifest themselves more clearly in tundra ecosystems – in which direct human impacts (e.g., associated with land-use change) remain less intense – than in many other biomes.

REFERENCES

Arft, A M, Walker, M D, Gurevitch, J, Alatalo, J M, Bret-Harte, M S, Dale, M, Diemer, M, Gugerli, F, Henry, G H R, Jones, M H, Hollister, R D, Jónsdóttir, I S, Laine, K, Lévesque, E, Marion, G M, Molau, U, Mølgaard, P, Nordenhäll, U, Raszhivin, V, Robinson, C H, Starr, G, Stenström, A, Stenström, M, Totland, Ø, Turner, P L, Walker, L J, Webber, P J, Welker, J M, and Wookey, P A (1999) Responses of Tundra Plants to Experimental Warming: Meta-Analysis of the International Tundra Experiment, *Ecol. Monogr.*, **69**, 491–511.

Billings, W D, Luken, J O, Mortensen, D A, and Peterson, K M (1982) Arctic Tundra: a Sink or Source for Atmospheric Carbon Dioxide in a Changing Environment? *Oecologia*, **53**, 7–11.

Bliss, L C and Matveyeva, N (1992) Circumpolar Arctic Vegetation, in *Arctic Ecosystems in a Changing Climate – an Ecophysiological Perspective*, eds F S Chapin, III, R L Jefferies, J F Reynolds, G R Shaver, and J Svoboda, Academic Press, San Diego, CA, 59–89.

Chapin, III, F S and Shaver, G R (1985) Arctic, in *Physiological Ecology of North American Plant Communities*, eds B F Chabot and H A Mooney, 16–40, Chapman and Hall, New York.

Chapin, III, F S, Shaver, G R, Giblin, A E, Nadelhoffer, K J, and Laundre, J A (1995) Responses of Arctic Tundra to Experimental and Observed Changes in Climate, *Ecology*, **76**, 694–711.

Crawford, R M M, ed (1997) *Disturbance and Recovery in Arctic Lands: an Ecological Perspective*, Kluwer Academic Publishers, Dordrecht.

Christensen, T R and Cox, P (1995) Response of Methane Emission From Arctic Tundra to Climatic Change: Results from a Model Simulation, *Tellus*, **47B**, 301–309.

Foley, J A, Kutzbach, J E, Coe, M T, and Levis, S (1994) Feedbacks between Climate and Boreal Forest During the Holocene Epoch, *Nature*, **371**, 52–54.

Grabherr, G, Gottfried, M, and Pauli, H (1994) Climate Effects on Mountain Plants, *Nature*, **369**, 448.

Henry, G H R and Molau, U (1997) Tundra Plants and Climate Change: The International Tundra Experiment (ITEX), *Global Change Biol.*, **3** (Supplement 1), 1–9.

Körner, C, Diemer, M, Schäppi, B, Niklaus, P, and Arnone, III, J (1997) The Responses of Alpine Grassland to Four Seasons of CO_2 Enrichment: a Synthesis, *Acta Œcol.*, **18**, 165–175.

McGraw, J B and Antonovics, J (1983) Experimental Ecology of *Dryas octopetala* Ecotypes. I. Ecotypic Differentiation and Lifecycle Stages of Selection, *J. Ecol.*, **71**, 879–897.

McKane, R B, Rastetter, E B, Shaver, G R, Nadelhoffer, K J, Giblin, A E, Laundre, J A, and Chapin, III, F S (1997) Reconstruction and Analysis of Historical Changes in Carbon Storage in Arctic Tundra, *Ecology*, **78**, 1188–1198.

Melillo, J M, McGuire, A D, Kicklighter, D W, Moore, III, B, Vorosmarty, C J, and Schloss, A L (1993) Global Climate Change and Terrestrial Net Primary Production, *Nature*, **363**, 234–240.

Oechel, W C, Hastings, S J, Vourlitis, G L, Jenkins, M, Riechers, G, and Grulke, N (1993) Recent Change of Arctic Tundra Ecosystems from a Net Carbon Dioxide Sink to a Source, *Nature*, **361**, 520–523.

Oechel, W C and Vourlitis, G L (1994) The Effects of Climate Change on Land–Atmosphere Feedbacks in Arctic Tundra Regions, in *Trends Ecol. Evol.*, Vol. 9, 324–329.

Oechel, W C, Vourlitis, G L, Hastings, S J, Zulueta, R C, Hinzman, L, and Kane, D (2000) Acclimation of Ecosystem CO_2 Exchange in the Alaskan Arctic in Response to Decadal Climate Warming, *Nature*, **406**, 978–981.

Press, M C, Callaghan, T V, and Lee, J A (1998) How Will European Arctic Ecosystems Respond to Projected Global Environmental Change? *Ambio*, **27**, 306–311.

Shaver, G R, Billings, W D, Chapin, III, F S, Giblin, A E, Nadelhoffer, K J, Oechel, W C, and Rastetter, E B (1992) Global Change and the Carbon Balance of Arctic Ecosystems, *BioScience*, **42**, 433–441.

Torn, M S and Chapin, III, F S (1993) Environmental and Biotic Controls Over Methane Flux From Arctic Tundra, *Chemosphere*, **26**, 357–368.

Woodin, S J (1997) Effects of Acid Deposition on Arctic Vegetation, Chapter 10 in *Ecology of Arctic Environments*, eds S J Woodin and M Marquiss, Blackwell Science, Oxford, 219–239.

V

Value of an Ecosystem

see Ecosystem Services and Costing (**Volume 2**)

Vegetation Ecosystem Model and Analysis Project (VEMAP)

William J Parton and Dennis S Ojima
Colorado State University, Fort Collins, CO, USA

VEMAP was initiated to meet the needs of the ecological community to evaluate different hypotheses of climate and carbon dioxide (CO_2) perturbations on ecological systems. The project was designed to evaluate the ecological response to changes in climate derived from current climate conditions and different levels of carbon dioxide using vegetation change and biogeochemistry models. The models used to study the vegetation response to climate changes include the mapped atmosphere-plant-soil (MAPSS), BIOME and the dynamic global phytogeography model (DOLY) (VEMAP, 1995). At the same time biogeochemical responses were evaluated using the Biome-BGC (Biome-biogeochemical cycles) model, Century, and TEM (terrestrial ecosystem model) (VEMAP, 1995). Analysis of the impact of changes in vegetation distribution on biogeochemical dynamics was also investigated.

The objective of VEMAP was to intercompare the biogeochemical response among a set of models driven by the same environmental inputs under current climate and projected climate and atmospheric carbon dioxide concentrations. Data sets representing the existing variability of climate and soils over the continental region of the US (Kittel *et al.*, 1995; VEMAP, 1995; Schimel *et al.*, 1997) were constructed as part of the project. In the first phase of VEMAP, equilibrium calculations were made for the two sets of models (biogeochemical and biogeographical models). The equilibrium calculations were derived by simulating the ecosystem and vegetation responses by running each model with monthly and daily climate files over the US, gridded into 50 km cells. Atmospheric carbon dioxide levels were represented at current levels and at a double carbon dioxide level consistent with the climate projections from various general circulation models. The climate information used for the simulations included minimum and maximum surface temperatures, precipitation, total incident solar radiation, surface humidity, and surface wind speeds. The model results indicated that the biogeography and biogeochemistry models were able to simulate similar contemporary conditions of net primary production, total system carbon, and other characteristics of the ecosystems simulated; however under the different climate scenarios and double carbon dioxide levels, the biogeochemical models displayed differential sensitivity to these altered conditions (VEMAP, 1995). The range of responses indicates that the various hypotheses of the effect of temperature, moisture, and carbon dioxide levels in uence carbon uptake, decomposition, water balance and nutrient controls in these different models.

In our recent analysis, the transient dynamics of uptake of carbon was modeled from 1895 to 1993 (Schimel *et al.*, 2001). Over this period, the uxes were highly variable as the weather varied from one year to the next, providing an indication of the sensitivity of biological processes controlling carbon uxes. The models also show that the amount of ecosystem uptake is greatly variable and thus an ecosystem can act as a source or sink depending on inter-annual and seasonal weather patterns. Land use management also emerged as an important factor contributing to transient carbon uxes from US ecosystems. In particular, the regrowth of forests on abandoned farmland (a major process in the US) and in previously harvested forests replaces carbon lost during previous decades. The VEMAP result suggests that, for the US at least, forest regeneration is between 100 and 200% as significant as carbon dioxide as fertilization (Schimel *et al.*, 2001).

This is important for the prediction of future carbon dioxide levels because while carbon dioxide fertilization will continue for the foreseeable future, regrowth of trees will slow or cease when forests reach maturity or are harvested.

See also: Biome–BGC Ecosystem Model, Volume 2; CENTURY Ecosystem Model, Volume 2.

REFERENCES

Kittel, T G F, Rosenbloom, N A, Painter, T H, Schimel, D S, and VEMAP Modeling Participants (1995) The VEMAP Integrated Database for Modeling United States Ecosystem/Vegetation Sensitivity to Climate Change, *Biogeography*, 22, 857–862.

Schimel, D S, VEMAP Participants, and Braswell, B H (1997) Continental Scale Variability in Ecosystem Processes: Models, Data, and the Role of Disturbance, *Ecol. Monogr.*, 67, 251–271.

Schimel, D S, Melillo, J, Tian, H, McGuire, A D, Kicklighter, D, Kittel, T, Rosenbloom, N, Running, S, Thornton, P, Ojima, D, Parton, W, Kelly, R, Sykes, M, Neilson, R, Rizzo, B, and Pitelka, L (2001) Carbon Storage by Natural and Agricultural Ecosystems of the US 1980–1993, *Science*, in press.

VEMAP Members (1995) Vegetation/Ecosystem Modeling and Analysis Project: Comparing Biogeography and Biogeochemistry Models in a Continental-scale Study of Terrestrial Ecosystem Responses to Climate Change and CO_2 Doubling, *Glob. Biogeochem. Cycles*, 9, 407–437.

Vernadsky, Vladimir

(1863- 1945)

Vladimir Vernadsky, Russian natural scientist, philosopher was founder of the Russian School of Earth Sciences which included genetic mineralogy, biogeochemistry, radiology, and biosphere study. In his book *The Biosphere* (Vernadsky, 1926), he was the first to define the biosphere from a very modern thermodynamic and biogeochemical perspective by conceiving of it as an envelope of the Earth encompassing the range from atmosphere to soils where any form of life exist naturally. In contrast to the previous definitions of biosphere by Jean-Baptiste Lamarck and Eduart Suess, which included only living organisms, Vernadsky's concept of biosphere integrated all living and non-living matters which have been affected in any way by life. Furthermore in the books *The Biosphere* (Vernadsky, 1926) and *Chemical Content of the Earth's Biosphere and its Surroundings* (Vernadsky, 1965), Vernadsky presented the holistic concept of the biosphere and the study of biogeochemical cycles of the biosphere.

Vernadsky was one of the first natural scientists to describe environmental changes caused by human activity (*History of the Natural Waters* (1933–1934) Vernadsky, 1960). In this manuscript, he emphasized the in uence of humans on the geographical distribution of aquifers (disappearance of natural rivers and lakes and formation of artificial ones) and water quality. During the most tragic time in Russian history, when many friends and colleagues of Vernadsky were put into prison by Stalin's government, Vernadsky worked on the book *Scienti c Thought as a Planetary Phenomenon* (1937–1938) (Vernadsky, 1991). In the book, he emphasized the increasing role of humans in the evolution of the biosphere and the necessity for integrated studies of the Earth as a whole (including integration of biogeochemistry with social sciences). He considered humans as part of the biosphere, and human-caused environmental changes as not an accidental, but an inevitable step in the evolution of biosphere. He suggested that human activity would result in a transformation of the biosphere into the *noosphere* – the sphere of reason. In the noosphere, human thought and activity will be the major transforming forces of the environment (*see* Noosphere, Volume 5).

Vernadsky was far ahead of his time when promoting the necessity to integrate the study of the Earth system by including living and non-living matters as well as human activities and scientific thought for future investigation of the biosphere's evolution. A unification of scientific knowledge and close scientific cooperation between different countries were among the necessary conditions proposed for the success of this integration.

REFERENCES

Vernadsky, V I (1926) *The Biosphere*, Nauka, Leningrad.

Vernadsky, V I (1960) *History of the Natural Waters. Selected Works*, Publisher of the Academy of Science USSR, Moscow, Vol. 4.

Vernadsky, V I (1965) *Chemical Content of the Earth's Biosphere and its Surroundings*, Nauka, Moscow.

Vernadsky, V I (1991) *Scienti c Thought as a Planetary Phenomenon*, Nauka, Moscow.

Vernadsky, V I (1998) *The Biosphere*, Copernicus Springer Verlag, New York.

FURTHER READING

Vernadsky, V I (1945) The Biosphere and the Noosphere, *Am. Sci.*, 33(1), 1–12.

GALINA CHURKINA *Germany*

Vitality Fertilization

see Nitrogen Deposition on Forests (**Volume 2**)

VOCs (Volatile Organic Compounds)

see Volatile Organic Compounds, Biogenic (VOCs) (**Volume 2**)

Volatile Organic Compounds, Biogenic (VOCs)

Manuel Lerdau
State University of New York, Stony Brook, NY, USA

The production and consumption of volatile organic compounds (VOCs) by plants and microbes is one of the crucial ways in which ecosystems in uence the atmosphere. VOCs are organic molecules that have a low enough boiling point that they change from the liquid to the gas phase at a high rate at physiologically realistic temperatures, 10- 45 °C. Once these molecules are in the gas phase, they can diffuse rapidly from the plant or microbe to the atmosphere. Plants and microbes together produce approximately 2×10^{15} *g of VOC each year, and this quantity constitutes about 80% of the chemically reactive VOCs added each year to the atmosphere (with most of the remainder coming from fossil fuel combustion). These VOCs are primary determinants of the oxidative capacity of the atmosphere (a measure of the atmosphere's potential to remove pollutant chemicals) of the lower atmosphere, the dynamics of carbon monoxide and tropospheric ozone, and the abundance of atmospheric aerosols. Because these VOCs are produced and consumed through complex physiological processes, their rate of emission to the atmosphere is highly sensitive to changing environmental conditions such as temperature, precipitation, and land use practice. Biological exchange of VOCs thus has the potential to be a vital feedback link between global environmental change and ecosystem processes.*

INTRODUCTION

The range of VOCs produced by organisms is as large as their atmospheric impact. The most important VOCs range in size from the one-carbon methane (CH_4) through the 15-carbon sesquiterpenes ($C_{15}H_{24}$), with approximately 20 000 different chemical species found in this size range. Despite this nearly overwhelming diversity, certain patterns exist that allow one to group these VOCs by their chemical families, their impacts on the atmosphere, their biosynthetic pathways, or the taxonomic positions of the organisms that produce them. Most of the major VOCs are either alkanes such as methane (saturated hydrocarbons), alkenes (hydrocarbons with one or more carbon–carbon double bonds), or aldehydes, ketones, and organic acids (molecules with unsaturated carbon bonds and at least one oxygen).

Because they exist in low concentration in the atmosphere, most VOCs are difficult to measure. Thus our understanding of VOC emissions are often limited by development of techniques to measure them. Most VOCs are measured using gas chromatography to separate the chemical of interest from a sample of air and then a detection method such as ame ionization to quantify the compound. The identification of a particular compound is usually determined by using a gas chromatograph that is linked to a mass spectrometer. For a few of the VOCs that are particularly important in global climate and atmospheric chemistry, rapid measurement methods have been developed. For example, tunable diode lasers have been developed that allow determination of methane concentrations in nearly real time. In the case of isoprene (a compound discussed below), a chemical reaction and uorescence technique allows measurements on the time-scale of 10 Hz.

MAJOR CATEGORIES OF VOCs

Methane

Methane is the single most abundant VOC in the atmosphere, with an annual ux rate of approximately 500 terragrams (Tg, 1 Tg = 10^{12} g) of carbon per year and a standing concentration of approximately 1.8 ppm. Compared with most of the other biogenic VOCs, methane is inert in the atmosphere, reacting slowly with its principal oxidant, the hydroxyl radical (*see* OH–Radical: is the Cleansing Capacity of the Atmosphere Changing?, Volume 2). Calculations suggest that methane's atmospheric lifetime is on the order of 7–10 years. Because the entire atmosphere is mixed on the order of 1 year, methane tends to be homogeneously distributed across the surface of the earth. In the atmosphere, methane's inertness and its ability to absorb in the infrared portion of the spectrum make it one of the most important greenhouse gases (after water vapor and carbon dioxide). The oxidation of methane, although slow, is also important because it contributes to the global burden of carbon monoxide, a greenhouse gas and pollutant.

The two main sources of methane are fossil fuel related and biological. Coal mines and natural gas production both

yield methane as side products, and fossil fuel combustion produces methane as well. On a global basis, these fossil fuel sources are fairly small, combining to contribute only approximately 20% of the world's methane, with the rest derived from biological processes. All of this biogenic methane is produced through anaerobic metabolism in which a carbon source, either acetate or carbon dioxide, is reduced to methane. While acetate cleavage was almost certainly the first methanogenic pathway to evolve, both it and CO_2 reduction are found today. Many ecosystems contain anoxic zones that are sites for methanogenesis. However, those ecosystems that have oxic regions on top of their anoxic ones often have low net methane uxes because of consumption by methanotrophic bacteria living in oxic zones near the surface. These methanotrophs oxidize methane and can reduce the net ux so much that some systems, such as many forests, become net methane sinks.

The major sources of biogenic methane include rice paddies (≈20% of the total biogenic), natural wetlands (≈25%), fermentation in the guts of ruminants such as cattle (≈20%), and waste disposal sites such as landfills and sewage treatment plants (≈20%). The remaining 15% of the biogenic methane comes from a variety of small sources, including termites and freshwater systems such as rivers. These numbers indicate that approximately 60% of the global methane ux results directly from human activities and is thus potentially amenable to change if those activities are modified. For example, there is currently great interest in modifying rice cropping strategies in order to increase yields and minimize methane uxes.

Phytogenic Volatile Organic Compounds

With the exception of methane, as discussed above, all the other major biogenic VOCs are derived from plant physiological processes. The most important compounds, from the perspective of both the plant and the atmosphere, are the isoprenoids, which range in size from the 5-carbon isoprene, C_5H_8, to the 15 carbon sesquiterpenes, $C_{15}H_{24}$. Plants make many larger isoprenoids as well, but these are big enough that they have very low vapor pressures and rarely volatilize. Isoprenoids have played a large role in human affairs since long before the advent of global ecological concern. For example, monoterpenes, $C_{10}H_{16}$, have played a significant role in human history because they are major constituents of many herbs and spices, e.g., frankincense, myrrh, and rosemary. The significance of phytogenic VOCs for the atmosphere was first noted in Soviet Russia in the late 1920s, and then independently rediscovered in the United States and Soviet Georgia in the 1950s.

Isoprene

Of the isoprenoids, the most abundant, as a percentage of photosynthesis and in terms of concentration in the atmosphere, is isoprene. The annual ux of isoprene is equivalent to the ux of methane, but isoprene is much more reactive, with a lifetime ranging from minutes to tens of minutes. This high reactivity means that isoprene has an average concentration of only approximately 1–2 ppb. The high reactivity and low concentration of isoprene precludes any direct role for it in greenhouse warming. On the other hand, it is the dominant reduced compound in the troposphere and plays a major role in the chemistry of ozone and carbon monoxide.

The physiological and biochemical bases of isoprene production have only been understood in the last few years. Isoprene is produced through a complex metabolic pathway that uses two simple products of photosynthesis and respiration, glyceraldehyde phosphate and pyruvate to produce a 5-carbon sugar that is then converted to a compound that serves as the universal precursor for all isoprenoids, dimethylallyl pyrophosphate (DMAPP). When the phosphate groups are removed from DMAPP and nothing added in their stead, isoprene is formed. This step is apparently irreversible. That is, plants do not appear to metabolize isoprene, and they do not store it.

The enzyme catalyzing this final step has been well studied, and its kinetic and thermodynamic properties have been well described. The enzyme is activated by light and magnesium atoms and shows a strong pH dependency. In addition, activity of the enzyme depends strongly on temperature. Detailed studies of both isoprene uxes and enzyme activities have shown that the variability in emission from any one leaf over a time scale of seconds to minutes to hours can be explained by impacts of light and temperature on enzyme activity. What is not yet understood at all, however, is why, at the same light and temperature levels, some species of plants produce high levels of isoprene, others produce low levels, while still others appear to produce none.

There have been numerous suggestions regarding the possible functional significance of isoprene production and emission by plants. The first suggestion, made in the 1950s, was that it served as a by-product of a not-yet-identified second carboxylase in plants. This idea, however, has been completely dismissed. Next, the suggestion was made that isoprene production might serve as a sink for reductant when light harvesting exceeded the sink capacity of the Calvin cycle (the reactions of photosynthesis that consume light energy and convert CO_2 to organic forms). Energetic calculations, however, make this possibility remote. The reductant cost of isoprene synthesis is too small to confer a substantial benefit.

Two possibilities are currently considered most likely. The first is that, as isoprene diffuses from the chloroplast (its site of synthesis) to the intercellular air space, it alters the thermal stability of the membranes it passes through and makes the membranes more stable at higher

temperatures (Sharkey and Singsaas, 1995). This hypothesis is concordant with the strong positive relationship between isoprene biosynthesis and leaf temperature, but the demonstrated increase in thermal stability is very small, and there is not yet a clear biophysical mechanism for this increase in stability.

The second possibility, suggested more recently, is that isoprene production results from inefficiency in an enzyme whose primary function is in another part of the biosynthetic pathway from which isoprene is derived (Lerdau *et al.*, 1997). Studies on the kinetics of isoprene synthase fit with this theory; the enzyme has a remarkably high K_M, which suggests that its primary activity may be to catalyze a different reaction than isoprene synthesis (K_M equals the substrate concentration at which a reaction rate is one-half of its maximal value). There have not, however, yet been any direct tests of this hypothesis. For the moment, the function of isoprene production by plants remains a question open for investigation.

Terpenes

The other primary group of phytogenic VOCs are the mono- and sesquiterpenes. These C10 and C15 molecules serve diverse roles in plants, although for the most part they are involved with defense against pests and pathogens. Once terpenes are in the atmosphere, they tend to be much more involved in aerosol formation and relatively less involved in the dynamics of carbon monoxide and ozone than is isoprene. Recent evidence from studies of tropical clouds has shown that most rain water comes from transpiration and that organic compounds such as terpenes are responsible for most of the condensation nuclei, demonstrating that trees not only transpire the water that falls back down as rain, they also play a major role in cloud formation (Hoffmann *et al.*, 1998).

As with isoprene, the terpenes are made from DMAPP. In the case of the monoterpenes, two molecules of DMAPP are combined to produce the universal terpene precursor, geranyl pyrophosphate (GPP). This molecule then undergoes a series of reactions that lead to the many thousand of monoterpenes observed in plants. When sesquiterpenes are produced, the GPP combines with another DMAPP to produce a 15-carbon precursor that then becomes the final sesquiterpene product. These biosynthetic routes are well understood and have been characterized down to the molecular genetic level for some taxa.

The emission mechanism of these higher molecular weight isoprenoids is fundamentally different from that of isoprene. While isoprene is produced and immediately begins diffusing toward the atmosphere, with plants having no specialized structures or organelles for isoprene storage, the terpenes tend be stored in cavities, ducts, glands, or canals. Their emission depends on the volatility of the individual compound, its concentration in solution within the storage structure, and temperature. Together, these three parameters determine the vapor pressure of the compound and the emission is a diffusion process from the organelle to the air. Thus, unlike the case of isoprene, instantaneous emission rate is almost independent of light intensity and physiology. There have been recent reports that a few plants in the Mediterranean region of Europe emit terpenes in a light-dependent manner similar to isoprene, but it is not yet known whether these emissions are of importance outside of this region (Lerdau *et al.*, 1997). Many terpenes are emitted from tropical forest, but the relative proportions of light-independent and light-dependent emissions are not known.

Oxygenated Compounds

The third main group of phytogenic VOCs are the oxygenated compounds. Unlike the case of the isoprenoids, these compounds are classified together only by all having an oxygen molecule rather than through a common biosynthetic route. As one might expect, the oxygenated species are a highly diverse group, formed through many pathways, emitted through many different mechanisms, and serving many different ecological roles. In general, however, their atmospheric impacts are similar. The oxygenated species tend to have low vapor pressures, to be quite soluble in water, and to play important roles in aerosol formation rather than in carbon monoxide or ozone dynamics. It would be impossible to cover the full range of these compounds in one brief article, so attention will focus on a few that are notable.

One of the most abundant oxygenated VOCs is methanol. Methanol is produced when leaves are expanding as the cell walls grow, and current evidence suggests that methanol is produced as the pectin in the cell wall is de-methylated. The methyl group then becomes hydrated to form methanol. To date, every type of plant that has been examined has been found to emit methanol during leaf expansion. In contrast to methanol, with it ubiquitous distribution, another alcohol, methyl butenol, is found in only one group of pines, the yellow pines (Harley *et al.*, 1998). This compound, whose biosynthesis is incompletely understood, is the dominant VOC emitted by lodgepole and ponderosa pines, two of the most abundant trees of the western United States. It has never been found, however, as a product from any other group of plants.

Other examples of oxygenated compounds emitted by plants include acetone and hexanal, which are produced by leaves that have been cut. These compounds appear to help prevent microbial infection after injury. Also significant, although not well studied, are phytogenic organic acids, which may make up a substantial portion of the proton budget in tropical air that is not contaminated by human activities. A curious mystery of organic acid emissions is that it is not known whether these compounds are actually made by plants or whether they are produced by bacteria living on the surface of leaves.

EFFECT OF GLOBAL CHANGE ON VOLATILE ORGANIC COMPOUND EMISSIONS

There has been little direct experimental research on how global environmental change will affect the production and emission of methane and the phytogenic VOCs to the atmosphere. Certainly, however, the potential for feedback between methane emissions and greenhouse warming and phytogenic VOCs and atmospheric chemistry makes these questions paramount. The following section brie y covers four of the most significant environmental changes that are occurring on a global scale and how they might affect the emission of VOCs from ecosystems to the atmosphere: (1) warming from increased concentrations of greenhouse gases; (2) elevated CO_2 as a phenomenon that directly affects ecosystems; (3) increased deposition of nitrogen as a result of human activities; and (4) land use change as increasing amounts of land are converted to managed ecosystems.

According to most models of global change, increases in global mean temperatures are likely to occur disproportionately in high latitudes. These high latitude ecosystems are often characterized by having large amounts of accumulated organic matter, often in the form of peat. This peat, currently frozen for much or all of the year, could thaw and become available to microbes for decomposition. Much of this decomposition is likely to be anaerobic and lead to large amounts of methane production. This increase in methane could well act as a positive feedback on global temperature through methane's role as a greenhouse gas. On the other hand, if drying accompanies warming, then the proportion of an ecosystem that is anaerobic (and thus capable of supporting methanogens) may decline. Warming is also likely to affect the emission of the isoprenoid VOCs. Both isoprene and terpenes are emitted in strongly temperature-dependent manner, and the range of temperature increases suggested by most global climate models suggests that emission of these VOCs may increase exponentially in the 21st century. This change could, with the increase in reactive oxides of nitrogen that result from fertilization and fossil fuel combustion, lead to an increase in levels of carbon monoxide and tropospheric ozone, both important pollutants and greenhouse gases (Brown, 1998).

The direct effects of elevated CO_2 on methane production are harder to predict. Elevated CO_2 tends to increase the productivity of terrestrial ecosystems (*see* CO_2 Enrichment: Effects on Ecosystems, Volume 2), and this increase could lead to a rise in the deposition of organic matter to sediments and other anoxic environments. This increase in the substrate available for anaerobic decomposition may lead to an increase in methane production in many ecosystems. Similarly, as elevated CO_2 increases productivity in terrestrial systems, the standing biomass available to produce VOCs will also rise, and emissions may increase. There have been suggestions that elevated CO_2 will increase the relative availability of carbon to nitrogen in ecosystems and this will lead to a proportional increase in the production of carbon-based compounds such as isoprenoids, but experiments are only beginning to examine this possibility (Dacey *et al.*, 1994).

An environmental change that is occurring locally across the globe and may be of great relevance to biogenic VOC production is the increased deposition of reactive nitrogen (*see* Nitrogen Deposition on Forests, Volume 2). This reactive nitrogen could serve to fertilize ecosystems and lead to large increases in productivity. Increased nitrogen deposition may also increase the deposition of sediments and the production of methane. There have been several experiments showing that isoprenoid emission by plants responds strongly to nitrogen fertilization (Litvak *et al.*, 1996). Increases appear to occur both as a result of increased carbon fixation and of increased activity of the enzymes responsible for isoprenoid production. One aspect of nitrogen deposition that may alter VOC uxes is the impact of nitrogen deposition on biological diversity. It is not yet known whether nitrogen deposition will tend to favor VOC-producing species or not, but the effects could be profound.

Perhaps the single most important change in the global environment that is occurring with respect to VOCs is the extensification of agriculture, especially in tropical regions and the resulting conversion of forest to cropland, pasture, and then secondary forest. This conversion often involves the replacement of forests that contain many isoprenoid-emitting taxa with agricultural species such as corn, cassava, beans, or soy, which all happen to have very low levels of isoprenoid emission. The C4 grasses that often occur in tropical pastures also have very low levels of emissions. On the other hand, many secondary forest species, palms are a good example, have very high levels of VOC emissions. While it is difficult to make general predictions as to how VOC emissions will change in response to changes in land use, the highly species-specific nature of the isoprenoid emissions makes it likely that land use change will overwhelm the effects of other global environmental changes on biogenic VOC emissions.

Another important aspect of agricultural extensification that is likely to affect biogenic VOC uxes is the ever increasing area of land devoted to wetland rice production. As noted above, the sediments underlying rice paddies are substantial sources of methane. The large inputs of organic matter from rice production, the anaerobic nature of the sediments, and the many conduits within rice tissue that promote oxygen transport to roots combine to provide abundant substrate, an ideal production environment, and superb conditions for transport to the atmosphere of methane.

REFERENCES

Brown, D A (1998) Gas Production from an Ombrotrophic Bog – Effect of Climate Change on Microbial Ecology, *Clim. Change*, 40, 277–284.

Dacey, J, Drake, B, and Klug, M (1994) Stimulation of Ethane Emission by Carbon Dioxide Enrichment of Marsh Vegetation, *Nature*, 370, 47–49.

Harley, P, Fridd-Stroud, V, Greenberg, J, Guenther, A, and Vasconcellos, P (1998) Emission of 2-methyl-3-buten-2-ol by Pines: a Potentially Large Source of Reactive Carbon to the Atmosphere, *J. Geophys. Res.*, 103, 25 479–25 486.

Hoffmann, T, Bandur, R, Marggraf, U, and Linscheid, M (1998) Molecular Composition of Organic Aerosols Formed in the Alpha-pinene/O-3 Reaction: Implications for New Particle Formation Processes, *J. Geophys. Res.*, 103, 25 569–25 578.

Lerdau, M, Guenther, A, and Monson, R (1997) Plant Production and Emission of Volatile Organic Compounds, *Bioscience*, 47, 373–433.

Litvak, M, Loreto, F, Harley, P, Sharkey, T, and Monson, R (1996) The Response of Isoprene Emission Rate and Photosynthetic Rate to Photon Flux and Nitrogen Supply in Aspen and White Oak Trees, *Plant Cell Environ.*, 19, 549–559.

Sharkey, T and Singsaas, E (1995) Why Plants Emit Isoprene, *Nature*, 374, 769.

von Humboldt, Alexander

(1769- 1859)

Alexander von Humboldt (Baron) was born September 14, 1769 in Berlin, and died May 6, 1859 in Berlin. He was an outstanding research traveler, geographer and forerunner of modern ecological environmentalists such as Vernadsky (*see* Vernadsky, Vladimir, Volume 2).

After extensive study in Frankfurt/Oder, Gottingen, Hamburg and Freiberg/Saxonia (1787–1792), von Humboldt inherited property in 1796 that enabled him to pursue the life of an independent geographical explorer and researcher. On June 5, 1799 he started his famous expedition to South America (Venezuela–Cuba–Columbian Andes–Ecuador–Peru–Chile), and Mexico, returning from there via Cuba and the US (Philadelphia and Washington as a guest of President Jefferson) to Europe on August 3, 1804. Based on extensive observations, measurements, botanical collections (together with the French botanist Aime Bonpland) and physical/chemical experiments, von Humboldt started to publish the main scientific results of his extremely well prepared, meticulously planned and carefully documented research expedition to Central and South America, partly in co-authorship with Aime Bonpland, in 36 volumes entitled *Voyage aux régions équinoxiales du nouveau continent fait en 1799, 1800, 1801, 1802, 1803 et 1804, par Alexander de Humboldt et A Bonpland*. This monumental collection and documentation contains regional descriptions and botanical, geographical, geological, astronomical and zoological observations and analyses, most of it written during his stay in Paris (1807–1827) and published in French between 1805 and 1834. Besides this *chef d'oeuvre*, parts of von Humboldt's work were published separately and have been well received. These books include *Ideen zu einer Geographie der P anzen nebst einem Naturgemälde der Tropenländer* (1807), which basically is an atlas of vegetation, and *Ansichten der Natur* (1808), in which, for the first time, von Humboldt's ideas of a comprehensive and holistic interpretation of natural environments and the close interactions of geosphere, biosphere and atmosphere became apparent. The five volume (and atlas) *Essai politique sur le royaume de la Nouvelle Espagne* published between 1808 and 1811, and his *Essai politique sur l ˆle de Cuba* (1826) are predecessors of regional geographies, with a strong emphasis on the interactions between nature and society while being, at the same time, highly political in their anti-slavery discourses.

Among his later expeditions and explorations, the Russian–Siberian expedition (April 12 – December 28, 1829) is especially worth mentioning, because it was planned as a comparative and complementary endeavour to his great American expedition. However, the resulting publication *Asie Centrale. Recherches sur les chaines de montagnes et la climatologie comparée* (Paris, three volumes, 1843) indicates that the aims and scope of this trip were much more limited.

von Humboldt's great and lasting achievements in regard to a holistic-integrative world view are based on innovative scientific methodologies, with the ultimate goal of discovering connections and interrelationships between the physical and biological spheres at the Earth's surface. The use of sophisticated new instruments, accurate observations and quantitative measurements provided analyses of ora and fauna, topographies and climatic data. Thus, von Humboldt's expeditions solved a great number of geographical riddles (e.g., the connection between the Orinoco and the Amazon catchment areas via the Casiquiare), and introduced new conceptual tools for presentation of his findings. Using isolines for the first time, von Humboldt produced the first map of global isotherms, as well as three-dimensional vegetation patterns in the tropical Andes, revealing their interrelationship with climate and weather, topography and

geology as well as with culture. These are probably his most outstanding and lasting achievements in this regard.

The very essence of von Humboldt's new way of looking at our planet is laid down in his monumental, partly posthumous *Kosmos. Entwurf einer physischen Weltbeschreibung* (five volumes and one atlas; 1845–1862). In particular, volume one of *Kosmos* contains the foundations of von Humboldt's holistic–ecological view of our natural environment and its analysis well in advance of his geographical contemporaries: In der Lehre vom Kosmos wird das Einzelne nur in seinem Verhaltnis zum Ganzen, als Theil der Welterscheinungen betrachtet (volume one, 1845, p 40). The plea to describe, analyse and understand each single part of our natural environment, in relation to the whole and as part of the total cosmos was von Humboldt's main goal. Being well aware of the methodological problems of such an attempt, von Humboldt used the metaphor of producing ein allgemeines Naturgemalde als Ubersicht der Erscheinungen im Kosmos (p XII), a general tableau of nature as a holistic entity of all expressions of the cosmos. Nature, as such, became a scientifically relevant object of analysis, with natural landscapes (*Naturlandschaften*) being their visual and objectively experienceable expression. However, the totalcharakter of a landscape is only the exterior expression of an innere Verkettung des Allgemeinen mit dem Besonderen (ibid.), i.e., the inner combination/interaction of the general with the specific. Thus, the integration of different kinds of phenomena, each, however, in its specific setting, are the basis for von Humboldt's attempt to understand the cosmos and its various regional and ecological differentiations. *Cosmography/Weltbeschreibung* – a term adapted from the philosopher Kant – thus became the study of all that has been created, of all aspects of nature that exist in a geographic context on Earth (things and forces of reality) as well as their simultaneously existing natural totality (*Naturganzes*!). With such an approach, von Humboldt became not only one of the most in uential scientists in the Age of Enlightenment, but also one of the most important forerunners of a new and extremely modern concept for the Earth sciences and of a holistic–integrative approach towards analysis and understanding of the interrelationships of diverse orders in nature.

FURTHER READING

Armstrong, P (1999) Charles Darwin's Image of the World: the In uence of von Humboldt on the Victorian Naturalist, in *Text and Image: Social Constructions of Regional Knowledges*, eds A Buttimer, S D Brunn, and U Wardenga, Institut fur Landerkunde, Leipzig, 46–53.

Beck, H (1959) *von Humboldt. Biographie in zwei Bänden*, Fr Steiner Verlag, Wiesbaden.

Bunkse, E V (1981) Humboldt and an Aesthetic Tradition in Geography, *Geog. Rev.*, 71, 127–146.

Godlewska, A M-Cl (1999) From Enlightenment Vision to Modern Science: Humboldt's Visual Thinking, in *Geography and Enlightenment*, eds D N Livingstone and C W Withers, University of Chicago Press, Chicago, IL, 236–275.

Hard, G (1969) Kosmos und Landschaft. Kosmologische und Landschaftsphysiognomische Denkmotive bei von Humboldt, in *von Humboldt*, ed H Pfeiffer, Werk und Weltgeltung, Munchen, 133–177.

Hard, G (1973) Der Totalcharakter der Landschaft, in *Geographishe Zeitschrift/Erdkundliches Wissen Heft 23, von Humboldt*, F Steiner Verlag, Wiesbaden, 49–73.

Humboldt, A von (1999) *Netzwerke des Wissens, Catalogue of an Exhibition*, edited by Kunst-und Ausstellungshalle der Bundesrepublik Deutschland GmbH, Bonn *et al.* (Comprehensive survey of different aspects of A von Humboldt's life and work).

Kratz, O (1997) *von Humboldt, Wissenschaftler - Weltbürger - Revolutionär*, Callwey, Munchen.

Pfeiffer, G (1960) Ritter, Humboldt und die Moderne Geographie, in *Deutscher Geographentag Berlin, Tagungsbericht und Wissenschaftliche Abhandlungen*, eds E Otremba and H G Gierloff-Emden, Fr Steiner Verlag, Wiesbaden, 69–83.

ECKART EHLERS *Germany*

W

WDC (World Data Center)

see Data Banks: Archiving Ecological Data and Information (**Volume 2**)

Weed

A *weed* is any plant that interferes with some human purpose. The fact that many pretty garden plants (e.g., purple loosestrife, *Lythrum salicaria*), once introduced to new locations around the world, became aggressive weeds emphasizes the subjectivity and context dependence of the term weed. In the context of agriculture and forestry, weeds interfere with the growth of crop plants and a variety of control methods are available. The choice of method depends on its environmental impact, the characteristics of the weed species, and the resources (e.g., personnel, time, budget, equipment) available.

Ecologically, weeds are usually r-strategists and dangerous as biological invaders. They often proliferate shortly after a disturbance (e.g., plowing, harvesting) because of their efficient exploitation of the relatively free access to abundant resources then. There are four principal mechanisms by which weeds establish their large, post-disturbance populations: (1) individuals of some weed species (e.g., bunchberry *Cornus canadensis*) survive the disturbance and grow vigorously under the strong light penetration of the disturbed crop canopy. These weeds may flower abundantly and set many seeds, and temporarily dominate the site. When the canopy closes and competition from neighbouring plants increases again, these weed populations decline. (2) Some weeds (e.g., pin cherry, *Prunus pensylvanica*) persist as large populations of viable seeds, living in a diapause state until disturbance stimulates them to germinate, often by exposing the seeds to sunlight. Many of these seeds grow to maturity in the resource rich conditions and reproduce profusely, leaving another generation of persistent seeds buried in the soil to await the next disturbance. (3) Other weed species (e.g., birches, *Betula papyrifera*, in a forestry context) may survive the disturbance in small numbers, but then seed abundantly and establish a cohort of new individuals. (4) Lastly, and perhaps most commonly, many species (e.g., goldenrods, *Solidago* spp., and fireweed, *Epilobium angustifolium*) invade the disturbed site, usually through wind-dispersed, or sometimes through animal-dispersed seeds.

Once established, weeds compete with crop plants for light, water, and nutrients. On rangelands and pastures, weeds may be unpalatable or even harmful to animals and reduce the quality of animal products. For instance, foxtails (*Alopecurus* spp.) can become stuck in horses' mouths; the price of wool drops when cockleburs (*Xanthium* spp.) are lodged in it.

Many weeds host insect pests or plant disease organisms. For instance, prickly lettuce (*Lactuca scariola*) provides a reservoir for downy mildew.

See also: **r–K Strategies**, Volume 2; **Integrated Pest Management (IPM) in an Era of Global Environmental Change**, Volume 4.

FURTHER READING

Rao, V S (2000) *Principles of Weed Science*, 2nd edition, Science Publishers, Enfield, NH.

RICHARD A FLEMING *Canada*

Whittaker, Robert H

(1920–1980)

Robert H Whittaker, was one of the giants of ecology, who by blending penetrating theoretical insights with landmark empirical studies helped rewrite the dominant theories of how ecological communities are organized. A brilliant and honored researcher, he was an equally dedicated mentor. Among my colleagues at Cornell University, none placed

more importance on training the next cadre of ecologists than did Bob Whittaker; hence, it is not surprising that the students he trained are among the leaders of their generation. His undergraduate text, *Communities and Ecosystems*, remains a gem a quarter of a century after he wrote it, still timely despite the advances in ecology in the intervening years.

Whittaker's great contributions began with his 1948 doctoral thesis (at Illinois), in which he examined the patterns of plant species distributions along altitudinal gradients in the Great Smoky Mountains. This in itself was rather remarkable, since he was a student in the Zoology Department, rather than the Botany Department, which had denied his candidacy because of his insufficient background in botany. Whittaker was a student of Charles Kendeigh, after his original adviser Victor Shelford retired. But it was perhaps Arthur Vestal who had the greatest influence upon him, by stimulating him to question the prevailing dogma, advocated by F E Clements, that plant communities were like super-organisms, collections of species that had coincident distributions and coevolved as units. Whittaker carefully mapped out the distributions of plant species in the Smokies, and showed that this was not the case at all; rather, species were distributed individualistically, as H A Gleason had argued. Whittaker thus demonstrated that communities were fuzzy, albeit useful, concepts, whose borders were arbitrary and membership ephemeral; they could not, hence, represent evolutionary units because they did not maintain sufficient integrity and continuity over time. This finding, which surprised Whittaker, remains influential today, and undergirds the recognition that the homeostatic properties of ecosystems and the biosphere must be recognized as emerging from selection acting on their member species rather than on the system as a whole.

Though his demonstration that species were distributed continuously in space was Robert Whittaker's most enduring contribution, the scope and magnitude of his other work was daunting. He continued his analysis of vegetation with work in the Siskiyous and on the Saguaro cactus of the Santa Catalina Mountains (with W A Niering), winning the Mercer Award of the Ecological Society of America for the latter work. His empirical work led him to pioneer (with H G Gauch, Jr) the application of multivariate ordination methods to species distributions, revolutionizing and dominating the subject. With G M Woodwell he wrote a long series of seminal papers on the effects of gamma irradiation on plant communities, and on patterns of primary production. With G E Likens and F H Bormann he published fundamental work on the Hubbard Brook ecosystem, and with Likens extended this to assemble influential summaries of world distributional patterns of production and biomass. With P P Feeny and T E Eisner he explored the chemical dimensions of ecology; and with S A Levin and R B Root, the interface between population biology and community theory. Indeed, perhaps the most remarkable aspect of Whittaker's career was not his studies of diversity, but the diversity of his studies. His work touched on every dimension of ecology; from physiology to ecosystems, from populations to communities. Few before him or since have shown such breadth of scope and depth of insight.

FURTHER READING

Whittaker, R H (1975) *Communities and Ecosystems*, 2nd edition, Macmillan, New York.

SIMON A LEVIN *USA*

Wilson, Edward Osborne

(1929–)

Edward Osborne Wilson, US biologist, writer, and philosopher, a native of Alabama and graduate of that State's university (BS, 1949; MS, 1950) and of Harvard University (PhD, 1955), has earned great distinction in several fields of scholarship. An international leader in the study of biodiversity and an effective spokesman for its conservation, Wilson has distinguished himself through his seminal contributions to evolutionary biology and biogeography, including the co-development (with R H MacArthur, 1967) of the theory of island biogeography. Wilson's lifelong focus on the study of ants has not only been of central importance for our modern understanding of the evolutionary relationships and systematics of that important group of insects, but led him to carry out important investigations and develop fruitful hypotheses about the nature and evolution of social behavior in insects generally. He has been involved in these studies especially concerned with the role of *pheromones*, active hormones outside of the bodies of the animals that produce them, in determining the behavior of social insects. Wilson's innovative development of the theory of sociobiology and his courageous championing of this important field of investigation has led to a great deal of important scholarship concerning the role of genetics in the determination of behavioral characteristics in

human beings and other animals. These considerations are of central importance in understanding the role of mankind in nature, and thus in the determination of our common future. Wilson's often moving and uniformly eloquently-written books including *On Human Nature* (1978) Pulitzer Prize, 1979; *Biophilia* (1984); *The Ants*, with Bert Holldobler (1990) Pulitzer Prize, 1991; and *Consilience* (1998) have been widely influential.

In the last title, Wilson argues for the unity of all human knowledge and the primacy of a small number of what he considers to be fundamental natural laws that underlie every branch of learning. No one has contributed more to the development of our common awareness of a shared dependency on the living world, or of our appreciation of the disastrous consequences of the rapid changes it is undergoing now.

Photo: John Chase of Harvard.

FURTHER READING

Wilson, E O (1978) *Life on Earth*, 2nd edition, Sinauer Associates, Sunderland, MA.

Wilson, E O (1978) *On Human Nature*, Harvard University Press, Cambridge, MA.

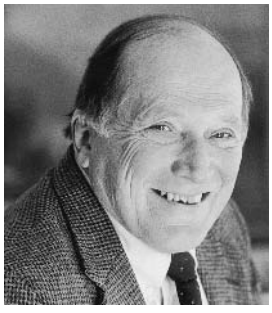
Wilson, E O (1984) *Biophilia*, Harvard University Press, Cambridge, MA.

Wilson, E O (1992) *The Diversity of Life*, Belknap Press of Harvard University Press, Cambridge, MA.

Wilson, E O (1994) *Naturalist*, Island Press/Shearwater Books, Washington, DC.

Wilson, E O (1978) *Consilience*, 1st edition, Knopf (distributed by Random House), New York.

PETER RAVEN *USA*



Woodwell, George Masters

(1928–)

G M Woodwell is a native of Cambridge, MA, a graduate of the Boston Public Latin School, holds degrees from Dartmouth College (AB, 1950) and Duke University (AM, 1956; PhD, 1958), and is the recipient of various honorary degrees. He held a commission in the US Navy between 1950 and 1953 and has held appointments and lectureships at various universities. Between 1957 and 1961 he was Assistant and Associate Professor of Botany at the University of Maine in Orono. He joined the staff of the Biology Department of Brookhaven National Laboratory in 1961 and remained there until 1975, when he founded and became the Director of the Ecosystems Center at the Marine Biological Laboratory (MBL) in Woods Hole. He held various positions including Deputy Director of the MBL, Assistant Director for Education, and Distinguished Scientist. In April 1985 he founded the Woods Hole Research Center, an institute for global environmental research of which he is Director and President. He held an adjunct appointment as Lecturer in Ecology at Yale University for many years.

Dr Woodwell's research has been on the structure and function of natural communities and their role as segments of the biosphere. He has worked extensively in forests and estuaries in North America and has made well-known studies of the ecological effects of ionizing radiation and the circulation and effects of pesticides and other toxins. He has, for many years, studied the biotic interactions associated with the warming of the earth.

Dr Woodwell has published more than 300 papers in ecology and has contributed articles to *Science*, *Scientific American*, *BioScience*, *Ecology* and the *Journal of Ecology*, among others. He has edited books on the effects of nuclear war, the global carbon cycle, biotic impoverishment, and satellite imagery used in measuring the area of forests globally.

Dr Woodwell is a member of the National Academy of Sciences and a Fellow of the American Academy of Arts and Sciences. He was the recipient of the 1996 Heinz Environment Prize.

Dr Woodwell was President of the Ecological Society of America in 1977–1978. He was a founding trustee and is Vice-Chair of the Board of the Natural Resources Defense Council; was a founder and remains an honorary member of the Board of Trustees of the Environmental Defense Fund; was a founding trustee of the World Resources Institute, and was a board member (1970–1984) and Chair (1980–1984) of the Board of Directors and is currently a member of the National Council of the World Wildlife Fund. He was Chair of the 1982 Conference on the Long-term Worldwide Biological Consequences of Nuclear War. He is former Chair of the Ruth Mott Fund of Flint, Michigan, a former member of the Board of Trustees of the Center for Marine Conservation, and currently serves on the advisory board of the Duke University Nicholas School of the Environment.

Photo: © Gabriel Amadeus Cooney.

R A HOUGHTON *USA*

Encyclopedia of
Global Environmental Change

egec

The volumes in this Encyclopedia are:

Volume One
The Earth system: physical and chemical dimensions of global environmental change

Volume Two
The Earth system: biological and ecological dimensions of global environmental change

Volume Three
Causes and consequences of global environmental change

Volume Four
Responding to global environmental change

Volume Five
Social and economic dimensions of global environmental change

Encyclopedia of
Global Environmental Change

Editor-in-Chief

Ted Munn

Institute for Environmental Sciences, University of Toronto, Canada

3

egec

Causes and consequences of global environmental change

Volume Editor

Ian Douglas

University of Manchester, Manchester, UK

JOHN WILEY & SONS, LTD

Copyright © 2002 John Wiley & Sons, Ltd, Chichester
Baffins Lane, Chichester
West Sussex PO19 1UD, UK

National: 01243 779777
International: (• 44) 1243 779777
e-mail (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on http://www.wiley.co.uk
or http://www.wiley.com

Copyright Acknowledgments
A number of articles in the Encyclopedia of Global Environmental Change have been written by government employees in the United Kingdom, Canada and the United States of America. Please contact the publisher for information on the copyright status of such works, if required. In general, Crown copyright material has been reproduced with the permission of the Controller of Her Majesty's Stationery Office. Works written by US government employees and classified as US Government Works are in the public domain in the United States of America.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except under the terms of the Copyright Designs and Patents Act 1988 or under the terms of a license issued by the Copyright Licensing Agency, 90 Tottenham Court Road, London, W1P 0LP, UK, without the prior permission in writing of the publisher.

Other Wiley Editorial Offices

John Wiley & Sons Inc., 605 Third Avenue,
New York, NY 10158-0012, USA

Wiley-VCH Verlag GmbH, Pappelallee 3,
D-69469 Weinheim, Germany

Jacaranda Wiley Ltd, 33 Park Road, Milton,
Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01,
Jin Xing Distripark, Singapore 129809

John Wiley & Sons (Canada) Ltd, 22 Worcester Road,
Rexdale, Ontario M9W 1L1, Canada

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library.

ISBN: 0 471 97796 9

Typeset in 10pt Times by Laser Words Private Limited, Chennai, India
Printed and bound in Great Britain by Antony Rowe, Chippenham, Wiltshire

Editorial Board

Editor-in-Chief

Ted Munn
Institute for Environmental Studies, University of Toronto, Canada

Volume Editors

Volume One ***The Earth system: physical and chemical dimensions of global environmental change***
Dr Michael C MacCracken, *Lawrence Livermore National Laboratory, California, USA* and
Dr John S Perry, *formerly, National Research Council USA*

Volume Two ***The Earth system: biological and ecological dimensions of global environmental change***
Professor Harold A Mooney, *Stanford University, Stanford, USA* and
Dr Josep G Canadell, *GCTE/IGBP, CSIRO Sustainable Ecosystems, Canberra, Australia*

Volume Three ***Causes and consequences of global environmental change***
Professor Ian Douglas, *University of Manchester, Manchester, UK*

Volume Four ***Responding to global environmental change***
Dr Mostafa K Tolba, *International Center for Environment and Development, Cairo, Egypt*

Volume Five ***Social and economic dimensions of global environmental change***
Mr Peter Timmerman, *IFIAS, Toronto, Canada*

International Advisory Board

Dr Joe T Baker
Commissioner for the Environment, ACT, Australia

Professor Francesco di Castri
CNRS, France

Professor Paul Crutzen
Max Planck Institute for Chemistry, Germany

Professor Eckart Ehlers
IHDP/IGU Germany

Professor José Goldemberg
Universida de São Paulo, Brazil

Dr Robert Goodland
World Bank, USA

Professor Hartmut Grassl
WCRP, Switzerland

Professor Ronald Hill
University of Hong Kong, Hong Kong

Dr Yu A Izrael
Institute of Global Climate & Technology, Russia

Professor Roger E Kasperson
Clark University, USA

Professor Peter Liss
University of East Anglia, UK

Professor Jane Lubchenco
Oregon State University, USA

Mr Jeffrey A McNeely
IUCN, Switzerland

Professor Thomas R Odhiambo
Hon. President, African Academy of Sciences and
Managing Trustee, RANDFORUM, Kenya

Sir Ghillean T Prance
University of Reading, UK

Professor Steve Rayner
Columbia University, USA

Contents

Preface to the Encyclopedia of Global Environmental Change xi

Preface to Volume Three xvii

Human Disturbance of the Earth System: Dynamics and Complexities 1
Global land cover and land use trends and changes 13
Global Population Trends 16
Trends in Global Emissions: Carbon, Sulfur, and Nitrogen 35
Anthropogenic Metabolism and Environmental Legacies 54
Industrial and Anthroposystem Metabolism 73
Water Use: Future Trends, and Environmental and Social Impacts 84
Contaminated Lands and Sediments: Chemical Time Bombs? 98
Rice and its Spread: Double Cropping; New Varieties – Environmental Consequences and Methane Gas, Sustainability 116
Environmental Change and Human Health: Extending the Sustainability Agenda 130
Environmental Changes Driven by Civil Conflict and War 146

Acid Sulfate Soils 151
Afforestation: Environmental Impacts 156
Agricultural Intensification in Java 159
Agricultural Intensification in Western Europe 162
Agricultural Subsidies and Environmental Change 168
Agroforestry 172
Aircraft Emissions 178
Animal Production, Feedlots, and Manure Problems in the US (Agriculture, Intensive) 186
Anthropogenic 190
Aquaculture and Environment: Global View from the Tropics to High Latitudes 190
Aquaculture in Asia 196
Aquaculture: Salmon Farming 200

Biomass Burning 205
Biomass Burning in Rural Homes in Tropical Areas 214
Biomass Fuel Power Development 216
Biomass Use for Urban Fuels in Developing Countries 217
Bioremediation 218
Capacity, Assimilative 220
Cattle Grazing: Impacts on Land Cover and Methane Emissions 221
Cereal Cultivation (Agriculture, Extensive) 228
Changes in World Marine Fish Stocks 238
Chernobyl 241
Climatic Extremes 243
Common Agricultural Policy (CAP) 247
Consumption Patterns: Economic and Demographic Change 249
Critical Load 252
Curi, Kriton 253

Deforestation and Habitat Fragmentation in the Amazon Basin 255
Deforestation in Historic Times 259
Deforestation, Tropical: Global Impacts 265
Demographic Change: Indonesian Transmigration 272
Demographic Change: Peopling of the Pacific Islands 273
Demographic Change: the Aging Population 277
Desertification, Definition of 282
Desertification 282
DVI (Dust Veil Index) 290

Earthquake Intensity and Magnitude Scales 293
Earthquakes Triggered by Human Activities 294
Ecological Capacity 296
Ecological Footprint 302
Ecotoxicology 302
Emergy 303
Environmental Cadmium in the Food Chain 303

Fisheries: Effects of Climate Change on the Life Cycles of Salmon 309
Fisheries: Minamata Disease 312
Fisheries: Pacific Coast Salmon 314
Fisheries: Pollution and Habitat Degradation in Tropical Asian Rivers 316
Food Consumption Patterns and their Influence on Greenhouse Gas Emissions 323

Forest Logging Systems in Tropical Countries: Differential Impacts 328
Fugacity 335

Geoengineering: a Way to Stabilize Global Climate? 336
Geography 337
Geomorphological Change for Urbanization and Industry 338
Geomorphological Change: Landscape Modification for Recreation 345
Green Revolution 347
Greenhouse Food Production (Agriculture Intensive) 352
Greening of Cities 356
Groundwater Withdrawal and the Development of the Great Man-made River Project, Libya 362

Harmful Algal Blooms (HABs) 371
Holocene: Climate Changes and Society 372

Indus Basin: a Case Study in Water Management 377
Insect Pests and Global Environmental Change 381
Inter-basin Transfer (IBT) for Water Supplies 387
International Rice Research Institute (IRRI) 391
Irrigation: Environmental Impacts 392
Irrigation: Induced Demise of Wetlands 399

Karst 411
Kovda, V A 411

Lahars 413
Lake Baikal 413
Land Degradation in the Mediterranean 417
Land Reclamation from Seas 424
Land Subsidence 430
Land Transformation in China: 3000 Years of Human Ecological History 430
Low Concentration Methane Bearing Gases 435

Malnutrition, Infectious Diseases and Global Environmental Change 440
Marine Mammal Exploitation: Whales and Whaling 446
Marsh, George Perkins 450
Marshes, Anthropogenic Changes 451
Materials Flows for Mining and Quarrying 454
Methane: Industrial Sources 461
Migrations: The Environmental Challenge of Population Movements 465

National Responsibilities for Greenhouse Gas Emissions 472
Natural Hazards 479
Nitrate Leaching 493
Nitrate Protection Zones in Europe 493
Nitrogen: Agricultural Uses and their Impacts 499
Nuclear Waste: Geological Issues 506
Nuclear Winter 515

Oil and the Arctic Environment 517
Oil Fires: Kuwait 523
Oil Industry: Cultural Impacts and Environmental Risk Tolerance 525
Oil Shales and Tar Sands 528
Organic Farming and the Environment 532

Phosphorus: Global Transfers 536
Phosphorus Water Quality Objectives 542
Prehistoric People and Forest Soils 546

Red Tide 548
Refugees and Human Conflict 548
River Regulation 551

Salinity and Agriculture 559
Shifting Cultivation and Land Degradation 565
Shipping: Harbors and Ports 572
Soft Energy Path 583
Soil Amelioration 583
Soil Deterioration and Loss of Topsoil 587
Swidden 594

Tidal Power Development 596
Tourism and Ecosystems 597
Tourism as a Global Driving Force for Environmental Change 609
Tourism: Climate Change and Tourist Resorts 623
Toxic Wastes: Generation and Disposal: a Case Study of UK Practice and Legislation 628
Transboundary Movements of Hazardous Wastes 631
Transport and Species Dispersal 632
Transport: Global Freight and Passenger Flows 633
Transport Infrastructure 643

Urban Climate and Respiratory Disease 649
Urban Ecosystems 655
Urban Heat Island 660
Urban Population Change 666
Urban Poverty and Environmental Health 672
Urban Sulfurous and Photochemical Smog 678
Urban Wastes 684

VEI (Volcanic Explosivity Index) 689
Viral Diseases and the Influence of Climate Change 690
Volcanic Eruptions: Mt Merapi, Indonesia 694

Volcanoes and Cities 696
Volcanoes and the Environment 699

Wadi 707
Waste Dumps in Megacities: Case Study of Istanbul 707
Waste: the Global Mass and its Management 709
Wastes: Landfill and Land Raising 713
Water Resources: Baltic 720
Water Resources: Great Lakes Case Study 721
Water Resources: Río de la Plata 723
Waterlogging 726
Wave Power Development 726
Weather Extremes and Climate Impacts: a Case Study for the United States 728
Wind Power Development 733
World Energy Consumption Trends 734
World Energy Council 738

Yellow River, China: a Case Study in Water Resources 740

Alphabetical List of Articles 743

Preface to the Encyclopedia of Global Environmental Change

THE ENVIRONMENT

The word *environment*, whose dictionary meaning is simply that which surrounds, has in the last few decades become a buzzword , encompassing an exceedingly diverse array of elements and social issues. Taking the original meaning as a point of departure, it is clear that we humans depend totally on the environment provided by planet Earth for the food we eat, the water we drink, and the air we breathe. Thus changes in this environment must be of vital concern. Will Earth continue to sustain humans in a way that also encourages the ourishing of the other living things with whom we share the planet? This question has loomed ever larger as it has become more evident that human activities have been inducing major changes in all of the compartments of the global environment. We have converted forests and savannas to farms and cities; we have exhumed ancient treasures of fuels and minerals; we have used the rivers and winds as convenient sewers; and we have released entirely new chemical compounds and organisms into the environment. In the 1960s, the scienti c community began to use the word *environment* in this new non-specialist sense. Soon too, Departments of the Environment were created by many governments, and new scienti c journals began publication while others were re-named. For example, the *International Journal of Air Pollution* became *Atmospheric Environment*! In the ensuing decades, the world community has come to see the environment in many different ways, as a life-support system, as a fragile sphere hanging in space, as a problem, a threat and a home.

GLOBAL ENVIRONMENTAL CHANGE

A broader and deeper understanding of the global aspects of environmental concerns emerged in the 1970s and 1980s, and a new phrase *global environmental change* acquired popular currency. Paleoresearch had revealed that environmental change was far from new, and by no means the sole result of the actions of heedless humans. Since the planet s formation, virtually every element of its environment has been undergoing massive changes on all space and time scales. Oxygen waxed and carbon dioxide waned in the atmosphere. Continents moved about the planet s surface like scum on a soup kettle. Great ice sheets grew and shrank. Above all, the force of life (the biosphere) emerged as a dominant driver of planetary change.

However, another vital insight began to emerge about 1980: the inescapably interlinked nature of these many environmental changes. On the very longest time scales, continental drift moved lands into different climates but also changed the climate of the globe itself. Photosynthetic life changed the atmosphere but also made possible more advanced life forms that could take advantage of the new environment. On shorter time scales, atmosphere and ocean often interact to produce the massive changes in the Southern Paci c that we term El Niño and La Niña, whose consequences extend across the planet, and profoundly affect even our socioeconomic systems. Indeed, we have come to realize that human-induced perturbations in the environment are becoming increasingly large, and are potentially coming to dominate the natural workings of the complex and interdependent global system that sustains life on Earth. Humans and their global environment are no longer independent; they are ever-increasingly becoming interdependent components of a single global system.

Thus, the term *global environmental change* has come to encompass a full range of globally signi cant issues relating to both natural and human-induced changes in Earth s environment, as well as their socioeconomic drivers. This implicitly includes concerns for the capacity of the Earth to sustain life that have motivated the development of studies of global change and sustainable development in the last few decades. Analyses of global environmental change therefore demand input from the social sciences as well as the natural sciences (and indeed also from the engineering and health sciences) necessitating an inescapably interdisciplinary approach.

Scientists from many disciplines have been attracted to this growing eld of global environmental change. This is particularly noticeable in the biological sciences through the encouragement of IGBP (International Geosphere Biosphere Programme), which invites ecologists to expand their eld of vision from the plot and landscape scales to the regional, continental and global ones, and to interact with scientists from other disciplines in exploring environmental change at these larger scales. Indeed this trend has encouraged publishers and scienti c societies to introduce new journal titles, and these publications from all accounts appear to be ourishing.

At the same time, human social, economic, and cultural systems are rapidly changing under the in uence of growing globalization. In the economic sphere, for example, today s discourse centers on multinational corporations, the Global

Economy, and the Globalization of Trade. Thus, the agendas of most of the physical, life, and social sciences increasingly focus on global-scale changes.

THE EVER-CHANGING ENVIRONMENT

Since the launching of Earth-viewing satellites, we have been able to see a constantly updated moving picture of our planet. Great weather systems sweep around the globe, waves of green and brown ebb and flow over the continents with the seasons, as do white waves of ice and snow. Nevertheless, a remarkable feature of the Earth system has been its relative stability. Since the dawn of life, the planet s environment has remained within a range of conditions that has supported life. Moreover, with some notable and perhaps cautionary exceptions, changes from year to year, decade to decade, millennium to millennium, have been modest, particularly during the last 10 000 years. Many features of human society exhibit similar behavior. Civilizations and cultures evolve slowly for the most part in response to environmental, economic, and social driving forces over their long lives. Great cities like Rome endure through the ages. A citizen of ancient Babylon probably would have little difficulty understanding the politics of modern Toronto.

But today we appear to be entering an era of change unprecedented since Babylon was a cluster of mud huts. There are many reasons to believe that changes greater than humankind has experienced in its history are in progress and are likely to accelerate.

Virtually every measure of human society numbers of people and automobiles, airplanes, energy consumption, generation of waste is increasing exponentially. While the values of a few indicators are leveling off, the magnitudes of annual increases of others remain immense.

Major changes are becoming evident in many critical elements of the environment increasing carbon dioxide concentrations, stratospheric ozone depletion (not to mention the stratospheric Antarctic ozone hole and the possibility of a similar Arctic event), rising sea level, declining productivity of soils, widespread collapses of fisheries, dramatic declines in biodiversity, destruction of tropical forests, etc. Within living memory, the countryside around large cities has been swallowed up by suburban developments and highways. Humans are stepping on the gas pedal of the planet s environment, and perhaps recklessly breaking the survivable speed limit of global change.

Paleoclimatological studies have shown that the stable environment that we take for granted has not always prevailed. Indeed, the relatively brief period in which human civilization has developed is somewhat unusual in its equable stability. Neolithic hunters in Europe some 13 000 years ago saw their climate shift from temperate to glacial in a single short lifetime. The evolution of human society is punctuated by wars, pestilence, and technological revolutions. Major perhaps even catastrophic change could occur in the future, because we know it has occurred in the past.

Rapidly advancing understanding of both natural and human systems and above all the ability to translate that understanding into quantitative models has enabled us to explore the future of our global society and its global climate with unprecedented realism. Although prediction of the single future path that we will follow is inherently unpredictable, it is possible to map a broad range of future environmental trajectories that we might take, each completely consistent with our understanding of how the system works. Such scenario-building exercises amply confirm our concerns that the changes of the 21st Century could be far greater than experienced in the last several millennia. Business-as-usual for human society appears to imply business-as-highly-unusual for the global environment.

THE INTERLOCKING BIOGEOPHYSICAL – SOCIOECONOMIC SYSTEMS

Recognition came in the 1970s that many of the environmental issues are inter-connected through the biogeochemical cycling of trace substances, especially carbon, sulfur, nitrogen and phosphorus. In fact, in a prescient statement on the main environmental research priority of the 1980s, Mostafa Tolba (then Executive Director of the United Nations Environment Programme) and Gilbert White (then President of the Scientific Committee on Problems of the Environment) drew the attention of both the scientific and the science-policy communities to the need to understand the major global biogeochemical cycles in order to maintain the global life support systems in a healthy state (Tolba and White, 1979). Quoting from that statement,

> We draw attention to the fundamental scientific importance of understanding the biogeochemical cycles which link and unify the major chemical and biological processes of the Earth s surface and atmosphere. The results will have practical significance for all of us who inhabit an Earth with limited resources and who, by our actions, increasingly affect the quality of the human environment.

So it is that many of the global environmental issues acid rain, stratospheric ozone depletion, climate change, nitrogen over-fertilization are inter-related through the global biogeochemical cycles.

One of the interesting results of the study of the Earth s history has been the discovery of the global teleconnections of the Earth system, with some major climatic shifts occurring simultaneously in the two hemispheres. In an analogous way, there has been recognition for a long time that human social, economic, and cultural systems are globally inter-related. That these are connected in turn to the

Earth system was implicitly recognized as early as the 1972 Stockholm Conference on the Human Environment.

To manage the human responses to this enormously varied but at least moderately coupled world system in an era of increasing global change through the diverse array of local, national and international organizations is indeed a daunting challenge. An essential first step is to describe past and present states, then to explain the various phenomena observed, and finally to develop predictive models (or at least a range of scenarios) describing the future behavior of this total Socioeconomic Cultural Environmental system. In this process, environmental scientists must learn how to assimilate better the new information constantly being received, although uncertainties, often large, will continue to exist. Within this mix, rational and effective policies must be developed that will balance the risks and costs of global environmental change in an adaptive way. Only then will we be able to even formulate, much less start to implement, rational and effective policies that cope effectively with global change. The prospect of change tempts us to think in terms of winners and losers. However, such analyses often do not play out in simple ways. For example, a modest increase in rainfall would cause farmers to rejoice while vacationers would despair. However, greater farm production can lead to lower commodity prices, thereby reducing farm incomes while making vacation food purchases less expensive. Human society is pretty well adapted to the present environment, so change is necessarily a challenge.

In the longer-term, much depends on the ability of societies to respond, to adapt. Societies will differ in the resources natural, human, and technological that are available to them. They will also differ in terms of the values and priorities they attach to physical, social, and environmental goals, and in the social and political mechanisms that they employ to seek these goals. These differences in the human world of generations yet unborn may be as great and as significant as the changes in the global environment. A major challenge for our times is to develop frameworks for understanding complex interdisciplinary issues of this complexity.

No discussion of change and the future can be complete without consideration of risk. Projections of the future, however imaginative or soundly based, necessarily center on plausible, surprise-free scenarios population will increase, economies will advance, climate will change all typically slowly and smoothly. However, such projections made at the beginning of the 20th Century would have missed two World Wars, the automobile, aviation, space travel, television, and McDonald s in Beijing. By definition, genuine surprises cannot be predicted. However, an understanding of the impacts of past surprises may help us to make our society and our world somewhat more robust in the face of the unknown surprises that await us.

THE HUMAN DIMENSIONS OF GLOBAL ENVIRONMENTAL CHANGE

The human dimension of global change has been defined as the ways in which individuals and societies contribute to global environmental change, are influenced by it, and adapt to it. Many empirical studies have been undertaken that describe human activities in physical terms, based on various kinds of indicators. Rates of deforestation, urbanization, and changing levels of emissions of greenhouse gases are only some measurable contributors to environmental change. But the dynamics of human activities and global change are much more complex, and they reflect the complexities of the human nature relationship.

While the sheer burden of human activity on the planet is important, we know that the major forces at work are human beings operating together through political systems, corporations, interest groups, and beliefs that sway whole peoples. These raise questions about the nature of choice, about the hopes, dreams, and frustrations that impel people forward, or block them. At the moment, for instance, the dream of the good life which has been a critical element of every religious tradition in the world is increasingly being defined in terms of material possessions and powerful images that are shown on global media. Can this particular version of human happiness be sustained on a limited planet? Some people say that it can; others say that it cannot.

The central questions about the role of human beings in global environmental change revolve around social, cultural, economic, ethical, and even religious issues. These are becoming more and more pressing, and more and more foundational as human beings deliberately or inadvertently modify more and more of the planet. It is also obvious that in this generation the modification of organisms and ecosystems may well extend to the modification of human beings themselves. Among the fundamental questions are: What motivates us towards saving or harming the environment? How do we see ourselves with regard to nature? What is our responsibility to this and future generations? Who do we think we are, and who would we like to become? Appropriately enough, Volume Five of this Encyclopedia wrestles, in many different voices, with these ultimate questions that remain intimately linked with the sweep of physical, chemical, biological, geographical, and institutional changes documented and discussed throughout the earlier four volumes.

The Brundtland philosophy urges us not to reduce options for future generations. Implementation of this idea is, however, difficult. For example, it often seems more compelling to alleviate current poverty than to protect the environment and renewable resources for future generations. Many scientists agree that new approaches are needed to meld the social with the natural sciences in the policy arena. Some of the new methodologies that go beyond the physical sciences

are described in Volumes Four and Five of this Encyclopedia, e.g., post-normal science, integrated environmental assessment, and the precautionary principle.

In a commentary in which he described the 21st Century as the century of the environment, Edward O. Wilson began by referring to the growing *consilience* (the interlocking of causal explanations across disciplines) so that the interfaces between disciplines become as important as the disciplines themselves. He then stated that this interlocking amongst the natural sciences will in the 21st Century also touch the borders of the social sciences and humanities. In the environmental context, environmental scientists in diverse specialities, including human ecology, are more precisely defining the arena in which that species arose, and those parts that must be sustained for human survival (Wilson, 1998).

This is a major challenge for environmental scientists. Already through DNA techniques, for example, it is possible to trace back connections amongst prehistoric peoples through the last Ice Age to modern times.

THIS *ENCYCLOPEDIA OF GLOBAL ENVIRONMENTAL CHANGE*

More than two million papers are published every year in science and medicine, a twentyfold increase since 1940 that taxes the resources of concerned citizens, scientists, university departments, research institutes and libraries. At the same time, many policy-motivated organizations find it difficult to draw together the necessary expertise for resolving the newly emerging environmental issues. The scientific literature relating to the environment is burgeoning. However, research syntheses are in general still scattered across a broad spectrum of journals and books, and information on global environmental change is not readily available in an inter-related way. In particular, it is quite uncommon for contributions from the natural sciences to appear in the same journal or workshop proceedings as contributions from the social sciences and the humanities.

This *Encyclopedia of Global Environmental Change* is a comprehensive and integrated reference in the broad area of global environmental change that will be conveniently accessible and productively usable by students, managers, administrators, legislators, and concerned citizens. The Encyclopedia consists of five volumes of interrelated material:

Volume 1 The Earth system: physical and chemical dimensions of global environmental change (Michael C MacCracken and John S Perry, editors)

Volume 2 The Earth system: biological and ecological dimensions of global environmental change (Harold A Mooney and Josep G Canadell, editors)

Volume 3 Causes and consequences of global environmental change (Ian Douglas, editor)

Volume 4 Responding to global environmental change (Mostafa K Tolba, editor)

Volume 5 Social and economic dimensions of global environmental change (Peter Timmerman, editor)

The first four volumes cover the broad issues of the science and politics of global environmental change. Volume Five adds an often understated but extremely important aspect, linking as it does, global environmental change to the socioeconomic, cultural and ethical dimensions of human societies. Here will be found a rich panoply of writings by people who are not natural scientists but who have thought deeply about the environment. It places global environmental change in a refreshing historical, sociological and cultural context. In many contributions, the time horizons of most interest are the last hundred and the next hundred years. However, some contributions dip backward millions of years. Throughout, the emphasis is upon the dynamics of the various processes discussed how and why did they change? A second recurrent theme is the interconnection of processes and changes What produces change? What is impacted by change? Finally, we attempt to deal even-handedly with natural and human-induced change, and with impacts on both the natural world and human society. From the numerous diverse articles in the Encyclopedia, we believe that the user can obtain a coherent picture of this complex and dynamic system of which we all are a part.

To assist in promoting this coherence, each volume begins with a group of extended essays on major topics that embrace the field covered in that volume. These are intended to provide an introduction to the topic, a convenient road map through cross-references for exploring the Encyclopedia. Then there follows, in alphabetical order, shorter articles on a variety of scientific topics, descriptions of scientific programs, definitions, acronyms, and biographies of leading contributors to the field from Charles Darwin, through the Russian ecologist Vernadski, to the three most recent environmental Nobel Laureates Crutzen, Rowland and Molina. Indeed, these definitions, biographies and acronym definitions are, we believe, a uniquely valuable feature of the Encyclopedia. In the case of acronyms of international organizations and programs, it is no exaggeration to state that a young environmental scientist requires not only a good understanding of science, but also a good knowledge of acronyms, if they are to follow the discussions taking place at many international meetings! Also included in the alphabetic listings are abundant cross-references to related topics in the same or other volumes.

The substantive scientific articles that comprise the meat of the Encyclopedia are original contributions by active scientists from around the world, and thus

represent authoritative and up-to-date summaries of the state of current knowledge, direct from the producers of this knowledge. A number of these articles break new ground in synthesizing and summarizing our understanding from novel viewpoints and in unconventional ways. Thus, readers will find a wide variety of styles and approaches within the articles, reflecting in a unique way the rich diversity of today s world science. The articles have also had the benefit of careful reviews, particularly by our Editors.

The scientific essays and some of the program descriptions begin with a few italicized paragraphs written for non-specialists. These are not intended to be abstracts of the paper to follow, but rather are aimed at providing the reader with an introduction into why the topic is important and where it fits into the broader aspects of global change a kind of encouragement to read on. Reading on should not be too difficult a task, since most of the scientific essays are written at the level of journals such as *Scienti c American* and *AMBIO* that are intended for the non-specialist. The Encyclopedia will be a valuable source of information for everyone with a general interest or a need-to-know in the various environmental fields (the natural sciences, socio-economics, engineering, the health sciences, and policy analyses) particularly as they relate to global-scale environmental change, its drivers, and its consequences. It is also expected that among the audiences for each volume will be practitioners and researchers in the fields covered by the other four volumes. We believe that this rather unusual indeed unique Encyclopedia will be used in a variety of ways.

1. Some people will employ it simply as a convenient source of information on specific topics What has been happening recently to the sea ice in the Arctic Ocean? Who was Roger Revelle? What is soil mineralization? What is ICSU or IGBP? What is the Kyoto Protocol? What is deep ecology?
2. Serious students in the environmental sciences may begin by reading one or all of the introductory essays, which present highly compressed crash courses in a range of central topics.
3. We believe that the substantive specialized articles that constitute the bodies of the volumes are productively and enjoyably readable in their own right.
4. Finally, a case has been made by one of the volume editors (Peter Timmerman) for encyclopedia browsing as an enjoyable and productive pastime (*see* Encyclopedias: Compendia of Global Knowledge, Volume 5).

We believe that these five volumes are in the tradition of the human aspiration towards the compilation of global knowledge that sparked the preparation of the first encyclopedias in the 18th Century.

Our Editors have had the benefit of a marvellous postgraduate education in the course of the Encyclopedia s evolution! We are immensely grateful to the many, many authors in our virtual university who have contributed their wisdom to this project.

ACKNOWLEDGMENT

Many helpful comments and draft paragraphs were received from the Volume Editors during the preparation of this Preface.

REFERENCES

Tolba, M K and White, Gilbert F (1979) Global Life Support Systems, SCOPE Newsletter, No. 7, October. 2 pp.

Wilson, Edward O (1998) Integrated science and the coming century of the environment, SCIENCE 279, 2048 2049.

Ted Munn
Editor-in-Chief

Preface to Volume Three

The character of the Earth s surface is constantly changing as a result of the interactions of the great circulations of energy, water and materials that create the biogeochemical cycles between the atmosphere, hydrosphere, lithosphere, cryosphere and biosphere. These cycles are the sum of the processes that make ecosystems what they are at particular places and times. The processes involved will continue whatever astronomical, geophysical, biological and people-driven events do to the Earth, but their nature is constantly being modified. People meet their needs by altering biogeochemical cycles. They divert substances along new pathways, transforming them to create new goods and services. In doing so they also generate by-products, gaseous and liquid emissions and solid matter, all of which add to modifications of the biogeochemical cycles and thus to global environmental change.

These human actions have to be seen as part of a whole Earth system that is changing and producing feedbacks and responses. Some of these impacts are readily expected, but others are less easily anticipated. Some take place where the changes have occurred, such as subsidence above mine workings, but others occur far away. One example of the latter is the effect of acid rain created by industrial emissions in an industrial region on lakes and rivers hundreds of kilometers away. These connections also have a time aspect. For example, most engineering works, including large structures, dams, pipelines, flood and coastal defences, are usually designed to last 100 years or more, assuming that on average the environment will not change. When the structures are designed, historical records are used to estimate the magnitudes of major floods and storms. It is assumed that the frequency of occurrence of events of a given magnitude will not change for at least 100 years after the structure is built. Global environmental change now tends to make such assumptions invalid. Major weather systems, cyclone tracks and events like El Nino and La Nina are becoming less predictable. Already local environmental changes are affecting floods, as many people living on floodplains have already found. Floods of a given depth have become more frequent because land use changes have altered rates of runoff, especially in small highly urbanized watersheds.

To unravel these interactions, feedbacks and causes, holistic thinking is required. Ecology helps in developing this, by looking at organisms in their habitat and their interactions with other organisms and their environment. The extension of ecology from the study of plants and animals in specific habitats to the study of interactions amongst people, other organisms and their environment requires the inclusion of social processes. The resulting analysis of the interplay of social life, from politics to industry, work to recreation, with the great biogeochemical, water and energy circulations is often termed *human ecology*. That interplay is where the causes and consequences of global environmental change arise. Its analysis requires an integrated science that looks at the world as it really is. The task in this volume of the Encyclopedia is therefore to provide, or report upon, some of the key aspects of human knowledge that are required for that integrated science.

Ian Douglas
Editor of Volume Three

Human Disturbance of the Earth System: Dynamics and Complexities

IAN DOUGLAS

University of Manchester, Manchester, UK

The Earth System is dynamic, constantly changing, even though much of the system changes so slowly compared with human life spans that it appears to be permanent and stable. Other geophysical events, like volcanic eruptions, earthquakes, hurricanes and floods, produce rapid changes. Human activity accelerates natural processes, but also alters ecosystems, either through direct actions such as land clearance, drainage, excavation of minerals or dumping of waste, or through gradual indirect impacts such as the effects of acid rain on forests. In these ways, human action has already seriously disturbed Earth surface systems. Any future greenhouse gas-induced changes will be superimposed on other anthropogenic changes to the planet, so compounding them. Earth s ecosystems have an intrinsic resilience, but in many places that resilience is being outstretched by the magnitude of human activity. In some environmental systems, it is now difficult to be sure what their natural pre-disturbance condition was like. Even remote lakes in Canada and northwest Europe have had their pH levels reduced since the time of the Industrial Revolution, often crossing a critical biological threshold at around pH 5.5 in the process.

Land use and land cover change and global environmental change are part of a complex and interactive system in which human action and decision-making lead to changes in ecosystems and soils which produce environmental feed backs that, in turn, alter other ecosystems and eventually feed back on other human activities. These linkages and interactions occur at different spatial and temporal scales and are therefore sometimes difficult to interrelate. Any loss of soil nutrients, for example, has immediate impacts on land productivity, vegetation changes and soil erosion, then mid-term impacts on landscape fragmentation and land productivity, and possible long-term impacts on climate change.

Environmental change is driven by many factors, including economic growth, population growth and movements, urbanization, transportation, and war. Agricultural intensification, industrialization, and rising energy use stand out in terms of their profound impacts on the physical environment and their enormous potential for influencing human health. In the modern world, these three activities are essential for economic development and greater human well-being. Yet, all lead to pressures on the environment, such as pollutant emissions and resource depletion, which in turn can increase human exposure to threats in the environment.

In turn, regional and global environmental changes impact human activities. More extreme climatic conditions can affect a wide range of aspects of human life including crop yields, rangeland regeneration, water supplies for irrigation and domestic consumption, river flows for navigation and fish stocks, weather in tourist resorts, the breeding of disease vectors and sporting activities. Climatic extremes can aggravate local and regional environmental problems, and thus it is helpful to distinguish global change from global environmental problems. Global changes are those that are inherently worldwide in character, and that have to be studied primarily at a global scale. Global environmental problems include large-scale regional phenomena, usually common to several different parts of the world, such as acid rain in North America, Europe and China (Roberts, 1994).

To understand the causes and consequences of global environmental change, it is necessary to think holistically across different time and space scales. People have to be seen as part of the ecosystems or environments in which they live, while it is also recognized that what happens in one place is often governed by decisions taken far away, perhaps on the other side of the world. If a European supermarket chain decides to purchase prawns from a Central American country, the conversion of the local mangrove swamps for is accelerated, perhaps to the detriment of small-scale

local sheries and charcoal making. If a Japanese vehicle manufacturer decides to invest in a brown eld site in an old industrial zone in western Europe, then the biodiversity of exotic and local plants, insects and birds on the derelict site is lost, new drainage of a huge paved and roofed area is required, and a new node of vehicle emissions is created. Environmental change is about hundreds of millions of decisions from individual households to the boardrooms of trans-national corporations and the cabinet rooms of governments. The dynamics and complexities of human disturbance of the Earth System arise from these multi-level, multi-scale decisions.

THE INTEGRATED VIEW: HOLISM AND HUMAN ECOLOGY

In endeavoring to depict a systems view of the Earth, there is a danger of falling between theory and practice. On the one hand, there is a desire to have a robust account of the system, to be able to quantify the components and the linkages amongst them. On the other hand, there is the complexity encountered in practical situations of environmental management, when there are always many stakeholders to be considered and many outcomes from the decision made.

Robust accounting approaches tend to endeavor to simplify cause and effect relationships, for example, by predicting forest clearance from readily gathered variables such as population increase and road building. In reality, what people do to their land is a complex consequence of multiple factors, some external to a household or farm unit and some internal. To improve understanding, knowledge of the biophysical processes operating in the environment has to be brought together with knowledge of how people interpret the environment and make decisions about their behavior, activities and investments. Many people, from the extremes of wealth to the extremes of poverty, are so preoccupied about a range of things, from how much they

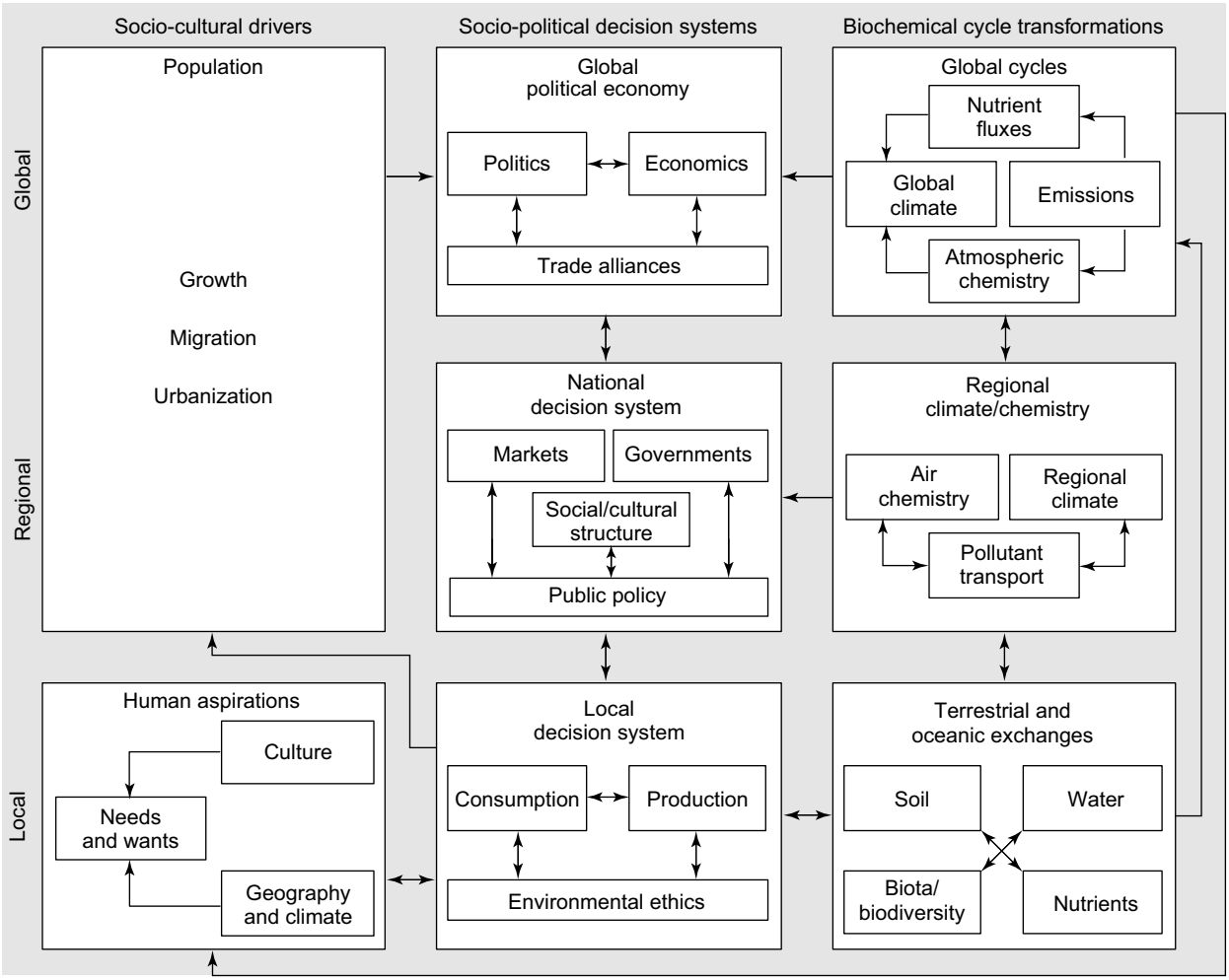


Figure 1 Diagram to illustrate how general human needs and wants are translated into land use changes that in turn affect regional air quality and climate (after Meyer and Turner, 1994)

will lose on the stock exchange in a day to where they are going to find food for their children s next meal, that they do not consciously think about their actions in terms of what they do to ecosystems or the global environment. However, they are all consumers, albeit at different scales, and human consumption, or more strictly diversion and modification of natural biogeochemical cycles, is a major driver of global environmental change.

In Figure 1, human consumption is expressed as population in the large box under studies of social drivers. Human aspirations mediated by culture, and where and how people live, influence human needs and wants. Individual decisions, the micro-decision system, influence land directly and also affect public policy that, in turn, influences the global political economy. The outcomes of decisions at these different levels feed back to the micro-decision system and influence what people do to their land and ecosystems. Thus, land is modified and biogeochemical cycles are altered.

Biogeochemical cycles embrace the turnover of nutrients and other chemicals at the surface of the earth, between the atmosphere, hydrosphere, lithosphere and biosphere. They are the sum of the processes that make ecosystems what they are at a particular place and time. These processes will continue, whatever astronomical, geophysical, biological and people-driven events do to the Earth, but their nature is constantly being modified. People s needs are met by changing those biogeochemical cycles, diverting substances along new pathways, transforming them and creating by-products that are often termed waste, but that have potential uses. That transformation requires energy that, in turn, involves either the extraction of hydrocarbons or coal, or the installation of facilities to exploit renewable energy. This diversion of earth materials, adjustment of ecosystems, and exploitation of past solar energy inputs, produces releases to the atmosphere, hydrosphere and lithosphere that may adversely affect existing ecosystems and cumulatively add up to changes in local, regional and global climate. Gradually, wetlands, lakes and rivers, and eventually coastal and enclosed seas, change their chemical and biotic character. Exploitation of marine resources adds to the global nature of this human impact. Eventually resources that had supported local economies for centuries, such as the cod fisheries of the Newfoundland Grand Banks and the North Sea, become exhausted.

None of these causes of change is isolated from some others. The causes interact with one another. They have to be seen as parts of a whole Earth system that is changing and producing feedbacks and responses; some readily expected and others less anticipated; some taking place where the changes have occurred, such as subsidence above mine workings, others far away, such as the effects of acid rain created by industrial emissions in one place on lakes and rivers hundreds of kilometers away.

To unravel these interactions, feedbacks and responsibilities, holistic thinking is required. Ecology helps in developing this by looking at organisms in their habitat, and their interactions with other organisms and their environment. The extension of ecology from the study of plants and animals in specific habitats to the study of interactions between people, other organisms and their environment requires the inclusion of social processes and introduced the concept of *human ecology*. Early in the 21st century, several varieties of human ecology have emerged. For some, it is primarily concerned with people s interactions with the built environment, while others consider it to be about how the human being, as an organism, responds to environmental stresses and changes and thus suffers disease or retains good health. A third view examines the way society and communities can manage their surroundings to achieve a better life while not over-exploiting natural resources. Whatever the precise definition of human ecology, it is the analysis of the interplay of social life, from politics to industry, football games and lotteries, with the biogeochemical, water and energy circulations that maintain all life on earth. That interplay is where the causes and consequence of global environmental change arise. Its analysis requires an integrated science that looks at the world as it really is. The task in this volume of the encyclopedia is therefore to provide, or report upon, some of the key aspects of human knowledge that are required for that integrated science.

TOOLS FOR THE INTEGRATED VIEW OF ENVIRONMENT AND SOCIETY

Some of the most challenging problems on Earth lie at the interface between the natural system and the human scene (Orme, 1985). Organizing science to tackle this interface is not always easy. Many reports or environmental impact statements dealing with specific people environment issues end up as simply a set of comments on air, water, soils, biota, cultural heritage, employment and society. Matrices, such as the Leopold matrix used in environmental impact assessment (Leopold *et al.*, 1971), provide a way of classifying how one component of an environmental issue affects another. Application of the ranking scheme within the Leopold matrix helps to assess the relative importance of different interactions. However, critics argue that the Leopold matrix is of little value for evaluating overall and cumulative beneficial or harmful impacts or for comparing alternative developments (Barrow, 1997). Later analysts suggested that global and regional environmental changes are the product of cumulative impacts, both beneficial and harmful (Rees, 1995). Analysis of these cumulative changes requires consideration of the sources of cumulative environmental change, the pathways of cumulative change and the cumulative effects. Tables of

Table 1 Types of cumulative effects of human impacts on the environment (after Spaling, 1994)

Type	Main characteristics	Example
Time crowding	Frequent and repetitive impacts on an environmental system	Forest selective logging at a rate exceeding the regeneration rate
Time lags	Delayed impacts	Chemical time bombs
Space crowding	High spatial density of impacts on an environmental system	Transfer of agrochemical residues to groundwater from non-point sources
Trans-boundary movement	Impacts occur downwind, downstream and away from source	Damage to coastal mangroves by polluted sediment washed down by rivers
Fragmentation	Changes in land cover	Isolation of forest remnants; disrupting fish runs by dam and weir construction
Compounding effects	Effects arising from multiple sources or pathways	Complex biogeochemical changes in lakes resulting from inputs by different inflowing rivers
Indirect effects	Secondary impacts	Landslides triggered by earthquakes resulting from the creation of large reservoirs
Triggers and thresholds	Fundamental changes in system behavior or structure	Change of a river from meandering to braided state due to increased flows of water and sediment from major construction sites

different types of cumulative effects (Table 1) help in this process.

Most of the direct impacts on the global environment show the effects of time crowding and space crowding, through repeated actions altering biogeochemical cycles, such as emissions of gases, applications of fertilizer, logging of forests and fishing the same area of ocean repeatedly. Trans-boundary movements detach the causers of local and regional problems from their zones of impact, as with the effects of acid rain on lakes and rivers in countries other than those responsible for the bulk of the release of sulfur to the atmosphere. Time lags and indirect effects can be particularly significant for future generations and are a most important element in the analysis of environmental change.

Energy Fluxes, Materials Flow and Industrial Metabolism

A stimulating approach to the analysis of people environment interactions is provided by the interpretation of ecosystems as open energy systems, using energy circuit diagrams to present energy flows and storages in both ecological and social systems (Odum, 1971, 1972). Such diagrams can be used to illustrate the flows of materials between cities and their hinterlands (Figure 2) and to identify where the biogeochemical cycles involved fail to re-use substances and the relative rates of losses and storage in natural and social systems (Douglas, 1983). The principle behind such analyses is that we should be able to fully account for the paths and sinks of all chemical residues from human activities, particularly industrial processing and subsequent domestic use. It makes sense to process and reuse concentrated materials in wastes, whereas the dilute wastes may be returned to the global biogeochemical cycles in ways that promote productivity and are not toxic (Odum *et al.*, 2000).

Accounting for these changes in energy fluxes and materials flows can be assisted by Odum s concept of emergy (*see* Emergy, Volume 3). Since work is required to concentrate materials, higher concentrations of minerals have greater emergy. For example, in nature, this increase in emergy occurs as minerals are concentrated during the earth movements that create high temperature liquid melts during mountain building events. In industrial ecology (*see* Industrial and Anthroposystem Metabolism, Volume 3), emergy increases as minerals are mined, processed and transformed into other products, e.g., lead into electrical storage batteries. Such an approach has been applied to the analysis of materials flows between the environment and society in Taiwan (Huang, 1998).

Urban metabolism (Wolman, 1965), industrial metabolism (*see* Industrial and Anthroposystem Metabolism, Volume 3), life cycle analysis and materials flow analysis are all techniques for looking at the movement of substances through complex human ecosystems. They help to quantify what happens to materials in the industrial, economic and social sectors, and to facilitate the recognition of

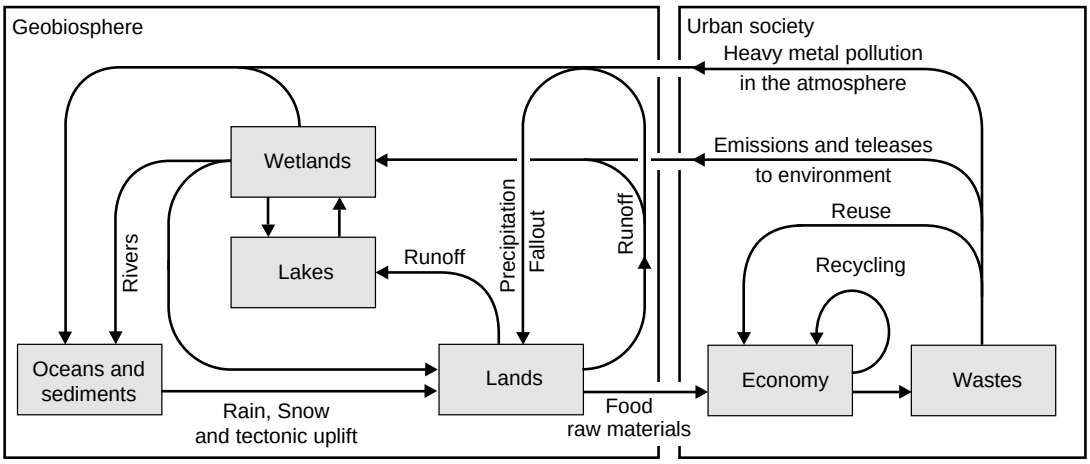


Figure 2 Diagram to show the relationship between the Earth System (geobiosphere, left hand box) and the social system (right hand box). The pathways indicate the extraction of minerals from the Earth and their transfer into the industrial system, the stocking of wastes (tank symbol) and the recycling of residues back to the Earth System, including wetlands, with release of pollutant aerosols to the atmosphere (after Odum *et al.*, 2000)

waste streams that have the potential to be used as resources elsewhere. Development of understanding of sources of emissions and pollutants through these analyses helps bring greater understanding of what might be done to mitigate some of the impacts of the human drivers of global environmental change.

THE HUMAN DRIVERS OF GLOBAL ENVIRONMENTAL CHANGE

Population and human aspirations (Figure 1) encapsulate a range of human driving forces affected by the socio-economic and political decision-making systems that lead to manipulations of the biophysical conditions of the land. These might be clearance of forests to increase grazing land for cattle, excavation of ancient river deposits to nd alluvial gold, or lling in a coastal wetland to extend an airport runway. All actions of this type change the quality and quantity of the local biogeochemical cycles and cumulatively lead to global environmental changes. The magnitude, nature and direction of these changes depend on the pre-existing environmental conditions and ecosystem states.

In broad terms, these driving forces can be clustered in two major groups: those concerned with land cover and land use change (*see* **Global land cover and land use trends and changes**, Volume 3) and those concerned with industrial metabolism (*see* **Industrial and Anthroposystem Metabolism**, Volume 3). These drivers give rise to complex response mechanisms, giving either positive or negative feedback.

Land Cover and Land Use Change

Land cover change and intensi cation of agriculture are essential for producing more food but, when not well managed, create substantial risks, such as exposing workers and communities to toxic pesticides, contaminating groundwater supplies and creating pesticide-resistant pests. Land clearing, irrigation and dams can bring increases in vector-borne diseases such as malaria and schistosomiasis, both of which exact a huge toll in rural areas of the developing world.

Industrial Metabolism

Industrialization fosters economic growth, and with its associated urbanization, usually leads to major improvements in human health. Yet, along with rising standards of living at least for a majority of the population industrialization often means increased exposure to heavy metals, persistent chemicals such as polychlorinated biphenyls, and other toxic chemicals. This is especially true for workers and the poor who often live close to factories. Such exposures are likely to be increasingly pronounced in the developing world, where the most rapid industrialization is occurring.

Rising energy use (http://www.wri.org/wr-98-99/002-ener.htm) is needed to fuel industrial growth but brings many attendant problems. Local air pollution from industrial and vehicle emissions has proved dif cult to manage even in developed economies. Emissions from both fossil fuel use and biomass fuel burning contribute to greenhouse gases (*see* **Biomass Burning in Rural Homes in Tropical Areas**, Volume 3; **Low Concentration Methane Bearing Gases**, Volume 3), with a predicted range of health impacts from severe storms, to drought, to ooding, to an increase in insect-borne diseases such as malaria. Energy demand, which is already huge in the developed countries, is rising fastest in the developing world.

Links Between Land Cover, Land Use and Industrial Metabolism

Trends in the human drivers of global environmental change rarely occur in isolation. Rising energy use, for instance, is part and parcel of industrialization and agriculture. The effects of industrialization are often difficult to disentangle from those of urbanization. Many of the human health effects of these trends are well known and predictable (e.g., increased air pollution that accompanies rising use of fossil fuels, or exposure to toxic chemicals through improper disposal of industrial wastes). Others, however, are far less certain, though potentially large, such as those associated with global climate change and wide-scale ecological disruption (*see* Environmental Change and Human Health: Extending the Sustainability Agenda, Volume 3).

As soon as these impacts of human driving forces on the environment and on the global climate begin to be examined in depth, it becomes clear that changes in the environment can affect health in indirect and often unexpected ways through the disruption of local or global ecosystems. For example, soil erosion stemming from poor agricultural practices can result in reduced crop yields; this could have important consequences for nutrition. Farm animal wastes in the eastern US are believed to cause toxic algal blooms, leading to massive fish kills and potential harm to humans. Note also the recent e-coli outbreak in the Walkerton (Canada) water supply that caused several deaths and many illnesses. Even well-intended development projects can have unexpected outcomes, as occurred in Africa s Senegal River Valley, where the construction of two dams set off a cascade of events that ultimately contributed to nutritional problems for the population and a dramatic increase in schistosomiasis.

CAUSES AND CONSEQUENCES, DECISION SYSTEMS AND CULTURAL ATTITUDES

An alternative way of considering these driving forces is to examine themes in Volumes 3, 4 and 5 of this encyclopedia. Volume 3 is concerned with energy and material flows, food supplies, and the demands and impacts of cities and all other human settlements. Volume 4 looks at environmental management and policy responses, while Volume 5 considers attitudes to the environment, including understanding of how different societies and their governments relate to the environment.

Volume 3 takes the biogeochemical cycles elaborated in Volume 2 and examines how they are modified by human-driven energy and materials flows. Food production and consumption are directly related to land cover, land use and biodiversity issues, and the major impact on the environment throughout the food production, processing, transport, consumption, and waste cycle (*see* Food Consumption Patterns and their Influence on Greenhouse Gas Emissions, Volume 3). Cities are where people are most concentrated and have most modified the surface of the earth, but their demands affect other ecosystems, both nearby and in other parts of the world (*see* Urban Ecosystems, Volume 3). In these cities, in new high technology parks, and among home-workers, evolving technologies transform lives and facilitate decision-making. Modern information technology, from remote sensing to mobile phones, helps both to hasten exploitation of environmental resources and to monitor and police environmental degradation and environmental protection legislation.

Information and Communication as Driving Forces

Developments in information and communication are now major driving forces in societal transformation and have important implications for the global environment. Information technology has three benefits for society and industry by: (1) compressing time; (2) overcoming geography; and (3) altering the structure of relationships, such as bypassing intermediaries. For transnational corporations, this means greater efficiency and effectiveness and facilitates innovation, quality improvements and the penetration of new markets. The manufacturing plants of Volkswagen AG in the Czech Republic, China, Mexico, Spain and elsewhere are linked by satellite connections to the company headquarters in Wolfsburg, Germany, enabling the company to bypass incompatible and costly national telecommunications networks (Malecki, 1997). Taking advantage of the modern telecommunications and information technology, many businesses have shifted research and development and data processing activities to localities where labor and land costs are low. India has a large high technology complex in Bangalore to which firms from the US, Singapore and the UK have sub-contracted work in electronics, software and data processing. Information and communication technology thus alters locations of economic activity, promoting urban development and consequent environmental changes in places where they had not occurred before.

Technology and communications can have both environmental benefits and examples of social inequity. Canon s global center for recycling and refurbishing laser printer tone cartridges is in Dalian, China. While the recycling is good for the global environment, the location of these dirty jobs in China reinforces the spatial division of labor that increasingly puts low paid, less healthy jobs in the poorer countries (Shrivastava, 1995). Another side of this coin is that telecommunications and modern air traffic permit agribusiness to produce seasonal crops in appropriate places around the world at any time of the year, and supply supermarkets in the wealthy countries with commodities like grapes and cherries continuously. The growth of this business has driven many land use changes, particularly in

tropical highlands of countries like Colombia and Kenya. Information and communications are thus facilitating more dispersed, more rapid environmental changes than would otherwise occur.

Macro Decision-making and the Global Political Economy

Decisions by governments, both in individual states and through international agreements and organizations such as the World Trade Organization (*see* WTO (World Trade Organization), Volume 4) and the Intergovernmental Panel on Climate Change (*see* Intergovernmental Panel on Climate Change (IPCC): an Historical Review, Volume 4), can effect changes in driving forces and exert influence over people s relationship with the environment. Policies such as the European Union s (EU s) Water Framework Directive can help reverse environmental trends by ensuring that participating governments take action to (in the Water Framework Directive s case)

> prevent further deterioration and protect and enhance the status of aquatic ecosystems and, with regard to their water needs, terrestrial ecosystems and wetlands directly depending on the aquatic ecosystems; and to promote sustainable water use based on a long-term protection of available water resources.
>
> (European Union, 2000)

Agricultural subsidies are a major influence on land use, changes in national government policies quickly altering inputs of fertilizer and forms of land management (*see* Agricultural Subsidies and Environmental Change, Volume 3).

PATTERNS OF CAUSES AND CONSEQUENCES OF GLOBAL ENVIRONMENTAL CHANGE

Several themes relating to the human drivers of environmental change are developed in this volume. Overall, they derive from population growth and associated changes in human settlements. Conversion of the Earth s surface results from the effects of agriculture, biomass burning, deforestation and logging, deliberate earth moving activities including mining and quarrying, and the growth of tourism and recreation. The use of fossil fuels in industry, machinery and transportation drive a wide range of environment changes. Water resource use creates many changes to river systems and wetlands. Fish production, both by fishing and by aquaculture, introduces major ecosystem changes. Urban expansion, urban materials use, waste dumping and transport exert interrelated impacts on ecosystems and are vulnerable to climate changes. Modifications to the pathways of particular chemicals also require attention.

Land Cover Conversion and Land Use Change

Land cover change to increase cropland dominated ecosystem change from 1700 to the mid-1980s. Cropland increased globally by 392 to 466%, depending on the means of estimation, expanding from an area roughly the size of Argentina to that of the South American continent. It grew at the expense of forest, grassland, and wetlands. Change accelerated globally, in terms of both the conversion of lands to cultivation and of the intensification of agriculture on land already cultivated. The net global area of irrigated cropland has increased by 2400% since 1800, from $80\,000\,km^2$ to $2\,000\,000\,km^2$. Today, cropland conversion and perhaps intensification are most rapid in the less industrialized portions of the world, while the area of cropland has decreased in Europe (*see* Global land cover and land use trends and changes, Volume 3). The total area of all sorts of grassland has changed little sine 1700, perhaps a net global decline of 1%, as its loss to cropland (primarily in Europe, North America, and Southeast Asia) has been compensated by its gain from forested land (mostly in Africa and Latin America). During the next decade and century, agriculture will shift from expansion towards intensification (*see* Green Revolution, Volume 3), while urban areas, particularly along continental margins, will continue to grow. By the end of this century, most of the world s land area will be intensively and formally managed, and open lands will no longer exist. Furthermore, local atmospheric pollution will affect much of world s prime agricultural lands.

Since pre-agricultural times, worldwide forest/woodland/tree cover has diminished by about 15%, but the size of this figure depends on how precisely woody vegetation is defined as forest or woodland. Currently, considerable afforestation is occurring in Europe and North America, as land is taken out of cultivation (*see* Afforestation: Environmental Impacts, Volume 3), while large-scale tropical deforestation is occurring because of logging, cultivation, and livestock uses (*see* Deforestation and Habitat Fragmentation in the Amazon Basin, Volume 3).

Shifts of Earth Surface Materials: Mining, Quarrying and Construction

The materials flows due to mining and urbanization dominate the human transformation of the Earth s surface and directly link industrial activity to the sustainable use of resources. The total mass of material moved is approximately three times the amount of mineral matter actually entering processing streams and globally exceeds the mass of sediment transferred from the continents to the world s oceans by rivers (*see* Materials Flows for Mining and Quarrying, Volume 3). Mining and quarrying create new landforms as they remove hillsides and even mountain ridge crests, create holes in valley floors and amass spoil heaps (*see* Geomorphological Change for Urbanization

and Industry, Volume 3). These landscape changes are sometimes used creatively to provide facilities for tourism, from artificial ski slopes in urban areas, to resorts and golf courses built by landscaping hillsides or reclaiming land from the sea (*see* Geomorphological Change: Landscape Modi cation for Recreation, Volume 3).

Fossil Fuel Use

Global energy production in 1993 reached 338 exajoules (1 exajoule • 10^{18} joules, equivalent to about 163 million barrels of oil), 40% more than in 1973. However, energy use is extremely uneven, the *per capita* consumption in the US is five times the world average, that of western Europe three times, and that of sub-Saharan Africa only one-tenth the world average. This is reflected in the pattern of *per capita* carbon dioxide (CO_2) emissions among the countries of the world (*see* National Responsibilities for Greenhouse Gas Emissions, Volume 3). Nuclear power offers a solution to the greenhouse gas problem, but has major difficulties for radioactive waste disposal and reactor safety (*see* Chernobyl, Volume 3; Nuclear Waste: Geological Issues, Volume 3). Most people in the world s poorest countries depend on wood fuel, with over 80% of forest cutting in many countries being used to produce fuel. Already forests are being cut to make charcoal for consumption in towns and cities hundreds of kilometers from the centers where it is used (*see* Biomass Use for Urban Fuels in Developing Countries, Volume 3). During the 21st century, those sources of wood will largely be used up. Countries will have to aim at alternative fuel sources, as well as encouraging agroforestry to promote cultivation of fuel wood. This biomass burning contributes to greenhouse gases, just as fossil fuel use does. However, future shortages of wood will help drive urban and rural electrification schemes forward, based where people have access to fossil fuel use, or to nuclear energy. Each of these alternatives has large environmental change feedbacks.

Fossil fuel consumption is driven by increase in wealth: the wealthiest people use the most energy. Motor vehicle ownership is one index of wealth. In 1993, private car ownership ranged from 1.48 $1000^{\bullet 1}$ in China to 3.1 in south Asia, and 561 in the US. However, since 1993, passenger car ownership has increased rapidly in many Asian countries, growing 17% $yr^{\bullet 1}$ in India. Predictions are that the number of motor vehicles in the world will have risen from 250 million in 1970 to over 800 million by 2010 (World Resources Institute, 1996). The materials flows involved in making all those cars, disposing of those that are worn out, and supplying fuel for them are major social drivers of global environmental change. Despite improvements in the fuel consumption and energy efficiency of motor vehicles, the sheer growth in their numbers maintains the level of their collective environmental impact, but the global location of growth of that impact is shifting towards the great cities of Asia and Latin America. Indeed, a simulation model has projected that by 2025, 10 35% of the world s grain production could be at risk from increasing frequencies and intensities of ground-level ozone episodes caused by the sheer growth in the motor vehicle population (Chamiedes *et al.*, 1994).

Water Resource Use

Water supplies for agriculture and the demands of irrigators are already critical elements of the land use dynamics in regions such as West Asia and semi-arid Africa. Agriculture accounts for as much as 70 80% of water use worldwide and competition for water from other economic and social activities is growing as countries develop. In some countries, nearly all the cropland depends on irrigation, e.g., 100% of that in Egypt, 78% of that in Pakistan 78%, and 47% of that in China and Iraq 47%, compared to a world average of 16%. Irrigation has major impacts on biogeochemical cycles, on soil conditions and on energy consumption (*see* Irrigation: Environmental Impacts, Volume 3). In some cases, the reservoirs storing irrigation water even cause minor earthquakes (*see* Earthquakes Triggered by Human Activities, Volume 3). Changes to the quantities of water available and to the quality of that water are both significant for future water resource management. Reuse of water along major rivers in every continent is threatened by deteriorating water quality, from agricultural pollution as well as urbanization and industry (*see* Nitrogen: Agricultural Uses and their Impacts, Volume 3). Efforts in old industrial areas, even in Manchester, UK, the home of the industrial revolution, have brought fish back to formerly polluted rivers. However, for every kilometer of river so improved, there are probably tens of kilometers of rivers elsewhere in the world being contaminated by effluents from new industries, inputs of agricultural chemical residues, runoff from parking lots or by the sediment washed into streams as new land is cleared for agriculture or housing. Water is already perceived as a potential flash point for international conflict (Libiszewski, 1999), which, in turn, would cause further environmental damage (*see* Environmental Changes Driven by Civil Conflict and War, Volume 3). Efforts to manage water in the future will have major implications for land use/cover change. Policies affecting the mechanisms of water distribution and use (institutions, water rights and pricing mechanisms) will play a key role. In addition, changes in precipitation regimes and in land cover, especially through deforestation, will aggravate water resource problems. Water issues are primarily those of dealing with the consequence of global environmental change, but the actual ways water is used contribute to local and regional changes, particularly the eutrophication of lakes and wetlands (*see* Eutrophication, Volume 2;

Circulating Freshwater: Crucial Link between Climate, Land, Ecosystems, and Humanity, Volume 5).

Fisheries and Aquaculture: Marine Ecosystems

For most of human history, the oceans were part of the global commons, in the sense that theoretically anyone could fish anywhere. Today, countries have exclusive economic zones, within which fishing may be regulated, but many of the major commercial fish, including cod, herring, salmon and tuna, migrate across zone borders and thus stocks are difficult to manage on a global basis. New technologies, especially gill nets, and large and powerful fishing fleets have resulted in depletion of the total stock, just as excessive logging overexploits the timber stock. The hunting of whales illustrates the issues of exploitation of marine life particularly harshly because of the impacts of technological innovation and the sheer scale of the slaughter around the world. The whaling industry acted as though the oceans were an inexhaustible resource, creating a real tragedy of the commons. International agreements regulating whaling remain difficult to police (*see* **Marine Mammal Exploitation: Whales and Whaling**, Volume 3). Aquaculture aims to replace some of the sh caught in the wild, but it has its own set of environmental and ecological impacts, with some serious implications for the transmission of disease from farmed stocks to wild stocks (*see* **Aquaculture: Salmon Farming**, Volume 3). Conversion of wetlands for aquaculture has had major effects on many mangrove ecosystems (*see* **Aquaculture in Asia**, Volume 3).

Urbanization as a Driver of Environmental Change

The increasing urbanization of the world population has triggered major changes in land use and land cover, and in the consumption of materials and food from the land and the oceans. The social factors linked to urbanization, particularly rural urban migration, feed back to environmental change as people abandon farms, or husbands go off to the city, while wives struggle to both raise children and till the land. In other situations, people who work in cities become commuters from country villages, and farmland becomes much more valuable as building sites. As indicated earlier, urban demands greatly manipulate biogeochemical cycles.

Analysis of biogeochemical cycling can be extended to incorporate an urban, or industrial, system in which the urban fabric is a repository of substances, alongside the lithosphere, biosphere, hydrosphere and atmosphere (Odum, 1971; Douglas, 1983) (*see* **Anthropogenic Metabolism and Environmental Legacies**, Volume 3). Indeed, the urban fabric is constantly being renewed, demolition creating a recycling opportunity for crushed concrete to be used as aggregate. The urban fabric can thus be interpreted as the urban deposit (Wilburn and Goonan, 1998), a form of renewable resource, constantly being restocked with new structures. In this way, people s activities are seen as part of the processes changing the Earth s surface at geological scales.

Changes to the Pathways of Particular Chemicals

Human use of materials has greatly modi ed the ows and stocks of key elements. Intensi cation of agriculture has led to great growth in the application of fertilizers to a world average level of 97 kg ha$^{\bullet 1}$ in 1996 1997 (World Resources Institute, 2000), ranging from extremely low quantities in countries like Namibia and Rwanda to 1475 kg ha$^{\bullet 1}$ in Belgium, 821 kg ha$^{\bullet 1}$ in the Netherlands and 693 kg ha$^{\bullet 1}$ in Korea. Pesticide use involves a far higher mass; data are not available for all countries, but the published range is from 1 kg ha$^{\bullet 1}$ in countries like Afghanistan and Burkino Faso to 19 288 kg ha$^{\bullet 1}$ in Italy and 13 829 kg ha$^{\bullet 1}$ in Korea. These agricultural chemical transfers are, in part, replacing the nutrients removed from the soil during harvesting. Some of those nutrients may be returned to the soil as sewage sludge. However, some are leached out, enter other food chains and contribute to major ecosystem changes and eutrophication of water bodies (*see* **Nitrogen: Agricultural Uses and their Impacts**, Volume 3).

Nitrogen is a key chemical in the agricultural system. Methane emission from agricultural activities is another (*see* **Cattle Grazing: Impacts on Land Cover and Methane Emissions**, Volume 3). Carbon is involved in all human activities, while sulfur has been particularly associated with fossil fuel combustion. Emissions of these substances have changed through time (*see* **Trends in Global Emmisions: Carbon, Sulfur, and Nitrogen**, Volume 3). Their impacts are now pervasive, with acid rain around industrial zones now affecting wide continental areas, and nitrate contamination of groundwater supplies widespread. Attention to sources and pathways of these and other elements can provide the means of identifying their fates and how their impacts may be modi ed. Establishing ow diagrams and developing good predictive models helps to identify where industrial and agricultural chemical use can be modi ed to achieve greater ef ciency, more recycling, less waste and lower levels of environmental impact.

FEEDBACKS IN THE SOCIETY: ENVIRONMENT SYSTEM

Figures 1 and 2 indicate links running in both directions between the earth system, with its biogeochemical cycles, and the social systems. These feedbacks always occur. However, often the relationships between economic activities and biogeochemical circulations are not immediately

evident. Nevertheless, there are important connections from the impacts of both direct and hidden flows of materials on the environment and the subsequent feedback from those impacts to the economy. In Europe, the River Rhine provides several examples of these feedbacks between economic materials flows and geological processes at the surface of the earth. In the Swiss upper reaches, the Jura water connection to improve navigation caused such changes to the stability of the channels that a second set of engineering works had to be undertaken to remedy the impacts of the first. In the Rhine Rift Valley, the river is canalized, but flows so swiftly that it scours its bed. A large, continuous gravel replenishment operation, involving a materials flow of gravel from the floodplain into the rectified channel, is necessary to maintain the equilibrium of the channel. At a smaller scale in Britain, canalization of the River Mersey through the south of Manchester has caused severe bank erosion downstream of the rectified section which has led to large expenditure in relocating electricity pylons to avoid undermining by the river.

Biophysical feedbacks play a critical role in long-term dynamics of land use and land cover, but there is little agreement on their strength or temporal and spatial scales. For example, some say that soil erosion losses and their implications for long-term agricultural productivity and land cover change may have been overestimated. The United Nations estimates of desertification and erosion of global soils may be too high and the economic risks of soil loss may be exaggerated (*see* Desertification, Definition of, Volume 3). Despite observations indicating that overgrazing in parts of Africa would shortly destroy the landscape for that activity, other assessments considering traditional management practices, local soil conditions, and patterns of soil erosion and deposition suggest that widespread land cover degradation and change at the sub-continental scale may be 100 or more years away, although this does not mean that there is no ecological degradation. Scale is important here. Local soil degradation matters greatly to local people, directly affecting their livelihoods, but it may not be extensive enough to cause, for example, a sufficiently large change in surface albedo to contribute significantly to climate change at the global scale.

Many feedbacks occur long after the event that caused them. These delayed impacts, such as the arrival of nitrate in spring water 10 or more years after it was applied on the land surface above the aquifer, or the transfer of salt from the hill slopes to the valley bottoms decades after land clearance, can be insidious and long lasting, being spatially and temporally removed from their causes (*see* Salinity and Agriculture, Volume 3). In many ways such local consequences have a parallel in the ozone holes over Polar Regions; they are delayed stratospheric responses to cumulative actions at the surface of the Earth.

CONCLUSION

The Earth System operates through a series of temporary equilibrium states, in which ecological relationships tend to be stable for several years, decades or centuries. Natural climatic oscillations, particularly those of the Quaternary period, have left legacies that are only marginally stable under present-day conditions. Thus, although sea level is now rising, some areas, such as Scandinavia and the Hudson Bay lowlands, are rising faster than the sea level through isostatic rebound following the retreat of the great ice sheets of the last Ice Age. On the desert margins, fossil sand dunes are held in position by their natural vegetation. Once the precarious vegetation cover is removed, the underlying sands can be remobilized. Thus, some ecosystems, some vegetation soil landform combinations, are more vulnerable to change than others. Thus, in places, once the human driving forces come into play, environmental change is rapid. Other ecosystems have a greater resilience, some sustaining a high level of intensive agriculture for long periods (*see* Agricultural Intensification in Java, Volume 3).

To understand what is happening at present and predict what might happen in the future, understanding of both the biogeochemical cycles on which individual ecosystems depend and the human driving forces altering those ecosystems is required. Both robust accounting systems and holistic thinking are required. Theoretical diagrams and the detailed analysis of case studies help in learning what might happen and what has happened.

The themes outlined in this essay indicate the types of interrelationships, causes, consequences and feedbacks involved in environmental change. The articles in this volume provide frameworks, accounts and examples of what has happened and what might happen. They indicate responsibilities for environmental change and its consequences at the individual, community, national and international levels. They show the contributions science has made to specific issues and the way that environmental policies could be improved or better enforced to alleviate some of the impacts of environmental change. If people see themselves as part of the great Earth system, as actors in the great drama of environmental change, they may appreciate that their individual and collective roles, however small, are as vital a part of planetary human ecology as the ants in the soil, the dung beetle in the forest, the operators of bulldozers and the presidents of transnational corporations and nation states. The challenge is to realize that we are all responsible, and that we can all do something. This encyclopedia sums up what we know, what we need to know and what we must do.

REFERENCES

Barrow, C J (1997) *Environmental and Social Impact Assessment: an Introduction*, Arnold, London.

Chamiedes, W L, Kasibhatla, P S, Yienger, J, and Levy, II, H (1994) Growth of Continental-scale Metro-agroplexes, Regional Ozone Pollution and World Food Production, *Science*, 264, 74 77.

Douglas, I (1983) *The Urban Environment*, Arnold, London.

European Union (2000) Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy, *Of cial J. European. Communities*, 22.12.2000 L327/1-L327/72.

Huang, S-L (1998) Urban Ecosystems, Energetic Hierarchies, and Ecological Economics of Taipei Metropolis, *J. Environ. Manage.*, 52, 39 51.

Leopold, L B, Clarke, E E, Hanshaw, B B, and Balsley, J R (1971) A Procedure For Evaluating Environmental Impact, *US Geological Survey Circular*, 645.

Libiszewski, S (1999) International Conflicts Over Freshwater Resources, in *Ecology, Politics and Violent Con ict*, ed M Suliman, Zed Books, London, 115 138.

Malecki, E J (1997) *Technology and Economic Development*, 2nd edition, Longman, Harlow.

Meyer, W B and Turner, II, B L, eds (1994) *Changes in Land Use and Land Cover: a Global Perspective*, Cambridge University Press, Cambridge.

Odum, H T (1971) *Environment, Power and Society*, Wiley-Interscience, New York.

Odum, H T (1972) *An Energy Circuit Language for Ecological and Social Systems: its Physical Basis*, Systems Analysis and Simulation in Ecology, Vol. 2, Academic Press, New York, 139 211.

Odum, H T, Wójick, W, Pritchard, Jr, L, Ton, S, Delfino, J J, Wójick, M, Leszczynski, S, Patel, J D, Doherty, S J, and Stasik, J (2000) *Heavy Metals in the Environment: Using Wetlands for their Removal*, Lewis, Boca Raton, FL.

Orme, A R (1985) Understanding and Predicting the Physical World, in *The Future of Geography*, ed R J Johnston, Methuen, London, 258 275.

Rees, W (1995) Cumulative Environmental Assessment and Global Change, *Environ. Impact Assess. Rev.*, 15(4), 295 309.

Roberts, N (1994) The Global Environmental Future, in *The Changing Global Environment*, ed N Roberts, Blackwell Science, Oxford, UK, 3 21.

Shrivastava, P (1995) Environmental Technologies and Competitive Advantage, *Strategic Manage. J.*, 16, 183 200.

Spaling, H (1994) Cumulative Effects Assessment: Concepts and Principles, *Impact Assess.*, 12(3), 231 251.

Wilburn, D R and Goonan, T G (1998) Aggregates from Natural and Recycled Sources, *US Geological Survey Circular*, 1176.

Wolman, A (1965) The Metabolism of Cities, *Sci. Am.*, 213(3), 179 190.

World Resources Institute (1996) *World Resources 1996- 1997*, Oxford University Press, Oxford.

World Resources Institute (2000) *World Resources 2000- 2001*, Earthscan, London.

Global land cover and land use trends and changes

IAN DOUGLAS

University of Manchester, Manchester, UK

Land cover and land use changes may be grouped into two broad categories: conversion or modi cation. Conversion refers to change from one cover type to another, such as the conversion of forest to pasture. Modi cation involves the maintenance of a broad cover or use type while changes take place in its structure or function, as happens when a forest is selectively logged and its canopy characteristics are greatly altered.

*For the Earth's land surface, land use and land cover changes are going to continue to be the most important component of global change, at least for several more decades (*see ***IGBP Core Projects****, Volume 2). The world has some 130.5 million km^2 of land. Present land cover is about 34.5 million km^2 of forest (World Resources Institute, 2000), 26.9 million km^2 grassland and 20.7 million km^2 cropland of all forms (Table 1). Thus agricultural activities (de ned as arable, pasture, and marginal land use) currently occupy approximately 37% of the earth's land surface. Urban and mining activities probably cover only around 1%. The remainder includes barren and steep terrain, including glaciers and mountain summits largely devoid of vegetation and unsuitable for agricultural activity.*

Up to and including most of the 20th century, agricultural output was increased mainly by bringing more land into production by converting forested and natural grassland areas to cultivated land. The limits of geographic expansion were reached many years ago in densely populated parts of India, China, Java, Egypt and Western Europe. Nevertheless, the total area of land used for agriculture rose from 45.5 million km^2 in 1966 to 49.3 million km^2 in 1996. This global average 9% increase in agricultural land is unevenly distributed. In the industrialized countries, the agricultural area increased by only 2.9% in those 30 years, but in the developing countries, it grew by 15.4%.

INCREASES IN AGRICULTURAL LAND

Among the countries with over 20 000 km^2 of cropland where the increase in agricultural area exceeded 25% between 1983 and 1993 are Egypt (26.5%), Mali (25.1%), Nicaragua (55.6%), Malaysia (46.6%) and Saudi Arabia (46.6%). The role of irrigation is outstanding in the Egyptian and Saudi Arabian cases, but generally irrigation is used to intensify production on existing land (*see* Irrigation: Environmental Impacts, Volume 3). In Asia particularly, intensification of production, obtaining more output from a given area of agricultural land, has been achieved primarily through producing multiple crops each year in irrigated agroecosystems using new, short duration crop varieties (*see* Agroforestry, Volume 3). The global irrigated area grew by more than 70% from 1966 1996. While irrigation is used on only 5.4% of agricultural land globally, it supplies 35% of cropland in South Asia, 15% in East Asia, and 7% in South East Asia.

Land abundant developed countries still possess the physical capacity to increase food production significantly. However, because of more intensive use of the existing arable land, and public policies that allow farmers to set-aside a proportion of their land, some 40 million hectares of land have been taken out of agricultural production in the US, Western Europe and Oceania over the last three decades. Cropland in the developed world actually fell by 2.2% between 1987 and 1997, with that of the US and Canada combined falling by 3.9% and that of Europe by 2.4%. The highest individual national cropland decreases over the same decade were in Italy (9.5%), the UK (8.3%), Romania (7.4%), Spain (6.0%) and the US (4.7%).

Unlike cropland, forest cover has globally declined, falling by 0.32% per year over the 5 years 1990 1995. The percentage decline is most marked in small countries, such as in Central America and the Caribbean (fall of 1.26% per year), and in countries with few native

Table 1 Estimates of pre-agricultural and late 20th century land areas under major land cover types (pre-agricultural and 1983 after Matthews (1983) and Williams (1997); 1997 estimated from data in World Resources Institute (2000), units • millions km^2

	Pre-agricultural	1983	1997
Forest	46.8	39.3	36.5
Wood	9.7	7.9	7.7
Shrub	16.2	14.8	14.6
Grassland	34	27.4	26.9
Tundra	7.4	7.4	7.4
Desert	15.9	15	15
Cropland	0	17.6	20.6
Urban/mining	0	0.9	1.3
Total land	130	130	130

forests, such as the Middle East and North Africa (fall of 1.00% per year). In contrast Europe (annual gain of 0.06%) and the US and Canada (annual gain of 0.17%) show slight increases in forest cover. Among the countries with more than 50 000 km^2 of forest that experienced an annual decrease in area of over 2% per year are: The Philippines (3.54%), Paraguay (2.65%), Thailand (2.65%), Nicaragua (2.54%) and Malaysia (2.43%). Across the Amazon Basin as a whole in 1986 1992, tropical deforestation was approximately 0.34% per year (*see* **Deforestation and Habitat Fragmentation in the Amazon Basin**, Volume 3).

POSSIBLE FUTURE TRENDS IN LAND USE AND THEIR EFFECTS ON LAND COVER

Total global food demand is expected to increase by 50% by 2015 and by more than 110% by 2050 (UNEP (United Nations Environment Programme), 1997). The combination of the increase in agricultural demands, changes in consumption patterns (less staple foods and more luxury products, which require more inputs), and ongoing technological development in agriculture will lead to a 27% increase of the total agricultural area by 2015 and about a 42% increase by 2050.

In the past, food production rose by both more intensive use of the land and by expansion of the agricultural land area. As in the 1990 2000 decade, the mainstay of agricultural production growth in the future will increase yields through intensi cation, from the application of new varieties and of increasing use of the products of agricultural technology such as fertilizers, irrigation systems and tractors, as well as improvements in management. (Tractor use in the Middle East and North Africa increased by 34% from 1987 1997 and in South America by 14%). The global area of irrigated land increased by over 21% in the ten years to 1997.

Agriculture s share of the total land is projected to increase to 46% by 2015 and to around 50% by 2050, whereas the ratio between pasture and arable lands is expected to remain relatively stable over time. These global changes mask quite distinct regional differences in land use changes. In industrial regions, agricultural land is projected to increase from its 1990 level by about 3% by 2015 and 10% by 2050. In North America, the percentage increase is 4% by 2015 but only 2% by 2050, while in Europe and the former USSR the gures are 4% by 2015 and 18% by 2050.

Developing regions, on the other hand, show a projected increase in agricultural land area of about 33% by 2015 and 45% by 2050 compared with 1990. The lowest increases are expected in Latin America and the Caribbean about 10% by 2015 and 20% by 2050. Because the large populations of South and East Asia have already occupied virtually all the available land suitable for agriculture the increases in Asia and the Paci c are about 28% by 2015 but only 24% by 2050. In Africa and West Asia, increases in agricultural area are projected to be much higher about 55 and 57% by 2015, respectively, and 90 and 95% by 2050. These gures include extensive use of grassland.

The large increases in the area of cultivated land in Africa and West Asia result from the sharp rise in demand already shown by the rate of expansion of cropland in Egypt and Saudi Arabia. The gures assume only moderate developments in agricultural management and technology. In particular, traditional animal husbandry systems, which require considerable grazing land, are expected to be continued. However, the regional expansion in farm mechanisation, indicated by the growth in tractor numbers, may indicate a trend to more elaborate farming systems, with better management but which would lead to far less dramatic land use changes by 2050.

In summary, to meet future food demands, a considerable extension of the area currently used for agriculture is needed, in addition to improvements in yields, unless more drastic changes take place in societies. It is projected that this extension will principally affect developing regions and some parts of North America, Europe, and the former USSR. In all areas, any extension of agricultural area is projected to occur at the expense of remaining natural areas. However, climate change is likely to bring a reduction in harvests in many regions. In the Canadian prairies, for example, drought now occurs on average once every 15 years; but by the end of this century, it could occur once every three years (Paul Slovak, personal communication).

POSSIBLE IMPACTS ON GLOBAL CLIMATE CHANGE

The process of land cover conversion to new uses, such as forest clearance for grazing land or plantation development,

Table 2 Annual net flux of carbon to the atmosphere from land use change: 1850–1990 (after Houghton, 1999), units • Tg C •1 teragram • 10^{12} g • one million tonnes • 1 Mt)

Year	North America	South and Central America	Europe	North Africa and Middle East	Tropical Africa	Former Soviet Union	China	South and South-east Asia	Pacific developed region	Total flux
1850	87.28	42.48	55.04	3.98	5.61	58.56	56.62	85.63	2.05	397.14
1900	158.92	60.36	45.19	17.55	12.56	57.79	47.54	161.44	21.75	583.11
1950	25.97	229.84	24.66	43.16	61.33	13.08	73.72	340.01	34.00	845.77
1990	12.42	577.16	• 18.08	23.24	341.50	20.11	48.69	1094.39	3.92	2103.34

leads to biomass burning and the release of greenhouse gases to the atmosphere. Past trends in such emissions are indicated by careful calculations of the annual net flux of carbon to the atmosphere from land use change (Houghton, 1999) (Table 2). The 18th century agricultural changes and early 19th century rise in European populations produced the highest flux from Europe in 1850. Thereafter the European flux declined until about 1980, when the trend reversed with the forest cover starting to increase and carbon sequestration growing.

The great 19th century agricultural expansion in North America produced a peak flux there by 1900, since when the flux has decreased. In North Africa, China and the Pacific the annual fluxes peaked around 1950. However, in Central and South America, Sub-Saharan Africa, South and South East Asia, the fluxes are increasing, and because of the huge total biomass of their tropical forests, are much greater in magnitude than those of other parts of the world. The predictions of agricultural land expansion suggest most strongly that these annual fluxes will continue to increase at least as rapidly as they did in the period 1950 1990. Thus land cover change will continue to be a major factor in the release of greenhouse gases.

REFERENCES

Houghton, R A (1999) Annual Net flux of Carbon to the Atmosphere from Land use Change: 1850 1990, *Tellus*, 51(B), 298 313.

UNEP (United Nations Environment Programme) (1997) *Global Environmental Outlook Global State of the Environment Report 1997 Chapter 4: Looking to the Future*, http://www.grida.no/geo1/ch/ch4_10.htm.

World Resources Institute (2000) *World Resources 2000 2001*, Earthscan, London.

FURTHER READING

Baulies, X and Szejwach, G (1998) *LUCC Data Requirements Workshop Barcelona*, 11 14 November 1997, LUCC report Series 3, Institut Cartogràfic de Catalunya, Barcelona.

Lambin, E F, Baulies, X, Bockstael, N, Fischer, G, Krug, T, Leemans, R, Moran, R F, Rindfuss, R R, Sato, Y, Skole, D, Turner, II, B L, and Vogel, C (1999) Land Use and Land Cover Change Implementation Strategy, IGBP Report No. 48, and IHDP Report No. 10, Scientific Steering Committee and International Project Office of LUCC, Bonn.

Meyer, W B and Turner, II, B L (1994) *Changes in Land Use and Land Cover: a Global perspective*, Cambridge University Press, Cambridge.

Global Population Trends

WOLFGANG LUTZ AND ANNE GOUJON

International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria

The 21st century started with a demographic paradox. Fertility in most developing countries declined at rates much faster than the decline experienced in the industrialized countries. Today already more than 60 countries with 44% of the world population have fertility below the so-called replacement level of two children per woman. On the other hand, the world population is still expected to grow strongly, most likely by more than two billion within the next 30 years.

*In this article, we summarize the current demographic situation and discuss projected trends in population growth and aging over the next century based on the most recent projections of the International Institute for Applied Systems Analysis (IIASA) (*see ***IIASA (International Institute for Applied Systems Analysis)****, Volume 4). The paper also draws on O'Neill* et al. *(2001). Our projections indicate that the world population will rise signi cantly from its present level of around 6 billion (Lutz, 1996). On the other hand, it is unlikely to double from its present size and a decline may occur during the second half of the 21st century. The distribution of the population will further tilt from more developed countries (MDCs) to less developed countries (LDCs). Finally, the world population will continue to age: the proportion of the 60• age group in the total population will almost certainly double and will likely triple over the next century.*

WORLD DEMOGRAPHIC TRENDS

Extremely rapid population growth is a relatively new phenomenon. For millennia, the strong fluctuations between birth and death rates cancelled each other, leading to slow population growth. It is in the 17th century that the world population took off, and reached 1 billion after 1800. The second billion was added more than a century later, in 1930. It took only 30 years, until 1960, to add the third billion. The fourth billion was added between 1960 and 1975, and the six billion mark was passed in 1999. In fact, despite the apparent acceleration, both the annual growth rate of world population and the number of persons added each year have passed their peaks and are expected to continue falling. The growth rate peaked at 2.1% year^{-1} in the late 1960s and has since fallen to 1.3% (see Table 1). This does not mean that little additional population growth is to be expected. The world population could rise to 10 12 billion by the end of the 21st century.

The statistics are, however, indicative of the tremendous changes that have taken place over the past four decades. The TFR, defined as the number of children a woman would give birth to if through her lifetime she experienced the set of age-specific fertility rates currently observed, declined with a rapidity that was unimaginable in the 1960s. This decline was very pronounced in some regions, i.e., Asia, Latin America, and North America. For example, in Latin America, the TFR fell by 1.4 children from 1955 1960 to 1975 1980 and again by almost 2 children between 1975 1980 at the turn of the century. One exception to this general decline has been Africa, where fertility rates remained well above six children per woman through the late 1980s; since then, the beginning of a fertility decline has become apparent. Meanwhile, regions like Europe that had already achieved very low fertility by 1975 1980 saw these rates persist or fall even further. In 2000, MDCs (More Developed Countries) have, on average, below replacement fertility levels, but even some major developing countries, most prominently China. Inhabitants of the MDCs on average have fewer children than LDCs (Less Developed Countries): TFR is 1.6 in MDCs versus 3.0 in LDCs.

The spread of modern sanitation and medicine caused infant and child mortality rates to fall, resulting in significant reductions in mortality during the 1950s and 1960s. During the period 1955 1960, life expectancy was lowest in Africa (40 years) and Asia (45 years), but had already increased significantly in Latin America (54 years). *Life expectancy* at birth is the average lifespan implied by the observed age-specific mortality rates in a given year. Over the subsequent 20 years, life expectancy increased impressively in all parts of the world. On average, the

Table 1 World and regional demographic trends since 1960

	Total population size (in millions)			Average annual rate of population growth (in %)			TFR			Life expectancy, both sexes combined (in years)			Population density (per km^2)		
	1960	1980	2000	1955–1960	1975–1980	1995–2000	1955–1960	1975–1980	1995–2000	1955–1960	1975–1980	1995–2000	1960	1980	2000
World	3022	4440	6055	1.8	1.7	1.3	4.9	3.9	2.7	49.6	59.8	65.4	22	33	45
MDCs	916	1083	1188	1.2	0.6	0.3	2.8	1.9	1.6	68.5	72.1	74.9	17	20	22
LDCs	2106	3358	4867	2.1	2.1	1.6	6.0	4.7	3.0	44.3	56.8	63.3	25	41	59
Africa	277	467	784	2.3	2.8	2.4	6.7	6.5	5.1	40.0	48.0	51.4	9	15	26
Asia	1702	2641	3683	2.0	1.9	1.4	5.6	4.2	2.6	44.8	58.5	66.3	54	83	116
Europe	605	693	729	1.0	0.5	0.0	2.6	2.0	1.4	68.3	71.2	73.3	26	30	32
Latin America and Caribbean	218	361	519	2.7	2.3	1.6	5.9	4.5	2.7	54.3	63.1	69.2	11	18	25
Northern America	204	255	310	1.8	0.9	0.8	3.7	1.8	1.9	69.8	73.3	76.9	9	12	14
Oceania	16	23	30	2.2	1.1	1.3	4.1	2.8	2.4	63.0	68.2	73.8	2	3	4

TFR, total fertility rate.
Source: UN, 1998.

world population gained 10 years between 1955 and 1975. The gain was over 13 years in Asia over this short period, and eight years in Africa. Improvements continued over the next 20 years to 1995 2000 but at a somewhat slower speed, with developing regions all seeing substantial improvements, especially Asia where the life expectancy improved by eight years. Very recently, some African countries seriously affected by the human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS) epidemic have seen reversals of this trend. The African continent has, on average, a life expectancy hardly above 51 years. The second lowest figure is Asia with 66 years. Inhabitants of the MDCs on average live significantly longer. Life expectancy at birth for both sexes combined is about 75 years, versus 65 in LDCs.

These trends in fertility and mortality resulted in different patterns of population growth in different parts of the world. The combination of high fertility rates together with declining infant and child mortality led to soaring population growth rates in the 1950s and 1960s in less developed regions. As shown in Table 1, growth rates in 1955 1960 were highest in Latin America because high fertility in the region was associated with already lower infant and child mortality. By 1975 1980, however, the population growth rate in Africa surpassed that in Latin America as Africa s mortality rates fell and fertility rates hovered at about 6.5 children on average per woman. African population growth rates have only recently started to decline; it was 2.4% for the 1995 2000 period.

The demographic contrast is significant between the well-off populations of Europe, North America, and Japan and the poorer populations of Asia, Africa, the Middle East, and Latin America. The population of the MDCs is relatively small (about 1.2 billion in 2000) and is expanding very slowly (0.3% year^{-1}). That of the LDCs is large (about 4.9 billion in 2000) and expanding rapidly (1.6% year^{-1}) after increasing by a factor of 2.1 since 1955. There are two main consequences to this trend:

1. the share of today s industrialized countries in the world population decreased from 30% in 1955 to 20% in 2000 and is likely to decrease much more in the future;
2. the population density has been soaring in the developing world. In 1960, MDCs and LDCs territory were, on average, similarly populated: respectively 17 and 25 inhabitants km^{-2}. In 1980, the density of LDCs was already more than twice that of LDCs and the difference has increased further with the past 20 years.

The population density per square kilometer of Asia, and Oceania doubled in the forty-year period between 1960 and the century end. Asia has by far the highest density of all world regions with 116 inhabitants per sq. kilometer. The European population density has been stagnating around thirty inhabitants, and that of Northern American region is less than 15.

The age pyramids in Figure 1 show how varying demographic experiences in different regions have a strong imprint on the age structures of their populations. Low fertility caused the narrowing of population pyramids at the bottom in MDCs. Extended longevity is the reason for the fattening at the top. This phenomenon of population aging is an important social phenomenon in MDCs, especially in relation to the uncertain future of pension and health care systems and implications for labor markets and the work force (Kinsella and Gist, 1995). Aging has already started in LDCs, and the speed of LDC fertility decline will accentuate the aging phenomenon (*see* Demographic Change: the Aging Population, Volume 3; Demographic transition, Volume 5).

The outstanding demographic event of the last 25 years has been the rapid decline in fertility around the world, particularly in LDCs. All research suggests that fertility decline is complex, regionally differentiated, and context dependent. For sure, the adoption of contraception has been the principal source of fertility decline in LDCs. However, it is more difficult to see why and how couples adopt contraceptive practices. It is a function of many influences, not only to do with the presence in a country of a family planning program. The synergism between the various factors national family planning programs combined with socioeconomic development and empowerment of women result in fertility decline. However, there is one indication that may be the key to the lowering of fertility levels in LDCs: female basic education. It is of importance to the three synergetic factors mentioned above (*see* Ecofeminism, Volume 5).

Female education has long been known as the strongest social correlate of fertility, far stronger, for instance, than income, occupation, labor-force participation, and the like (Cochrane, 1979; Cleland and Wilson, 1987; Cleland and Rodriguez, 1988). Even when controlling for socio-economic status, the independent effect of education exceeds that of any other factor (United Nations, 1987, 1993). One part of this impact is an accounting effect: if girls attend secondary school, age at first marriage, one of the principal proximate determinants of fertility, is higher. More important, however, is the effect of education on values and tastes that provide the context for fertility decisions. The strongest intermediate link between female educational attainment and lower fertility is the greater likelihood that married women who were schooled as girls will use contraception. While there are increasing returns (in terms of lowered fertility) as the number of years of schooling rises, women who have had even a few years of primary schooling are observed to be more likely to practice contraception than those who have had none.

Part of the effect of girls education is economic. It increases the cost of children by making girls unavailable

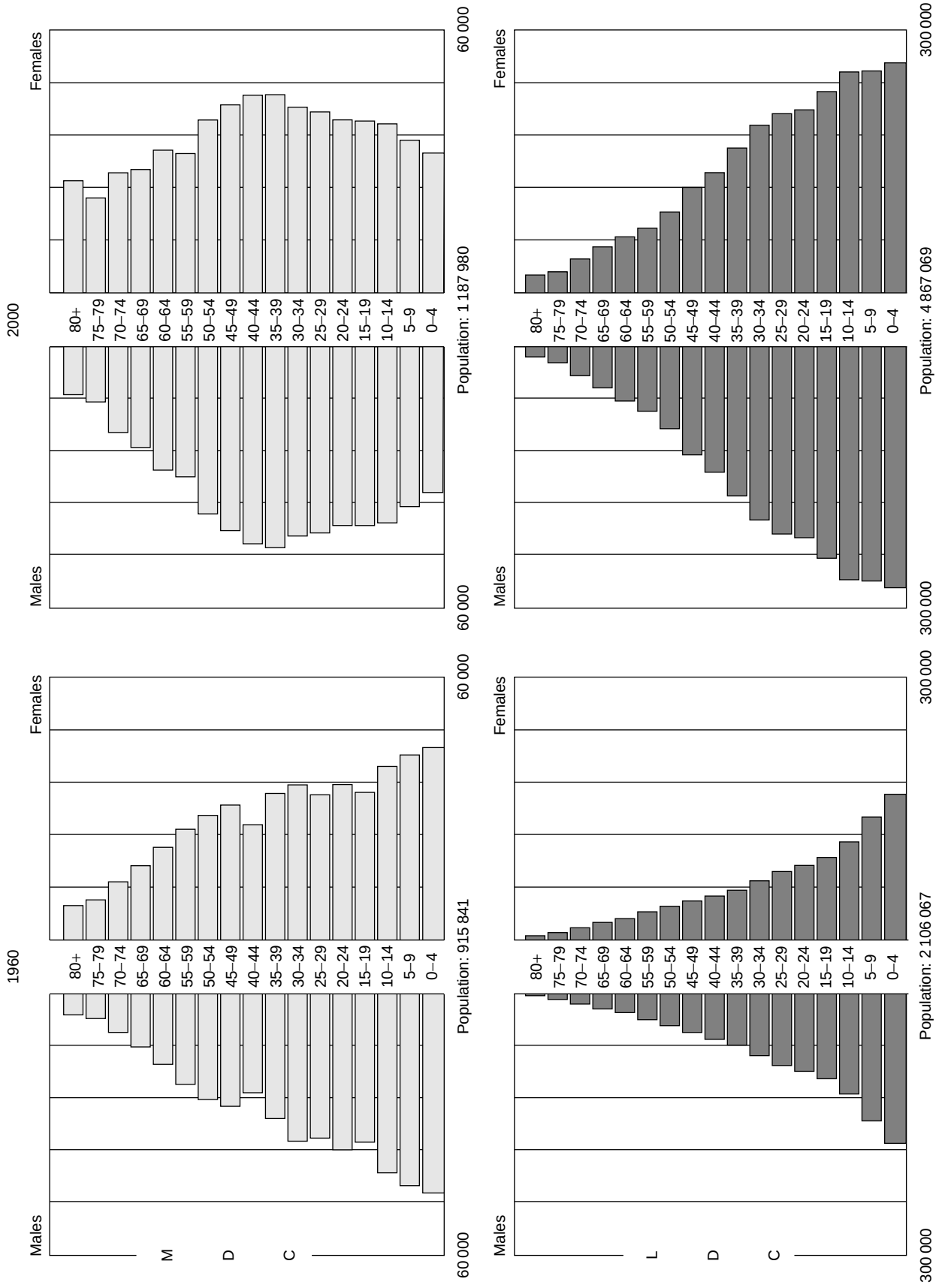


Figure 1 Population pyramids for MDC and LDC in 1960 and 2000

for domestic chores; educated mothers are also more likely to earn income outside the home than uneducated ones, and hence face higher opportunity costs in child rearing. But there is more to education of girls than just shifting economic costs and benefits. Parents decision to endow girls with even a modicum of education, as opposed to none at all, reflects a fundamental shift in the status of women. It sets in motion an irreversible shift in women s perceptions, ideals, and aspirations (Cleland and Wilson, 1987; Kasarda *et al.*, 1986). Women who have been educated are better able to follow their own, usually lower, fertility desires rather than those of their husband or their extended family. Another possible link between women s status and low fertility might be through improved health of children. Closely related to the low status of women are poverty and insecurity, both of which raise the value of children as economic assets (in particular, as sources of old-age security) and impede the attitudinal change necessary for fertility decline (Thomas, 1991). Finally, provision of basic female education promotes a political culture and civil society in which national family planning programs are likely to be established and function effectively.

POPULATION PROJECTIONS

The most-used method for population projections is the so-called *cohort-component model*, which projects the population by age group. This method was introduced by Whelpton in 1936, formalized by Leslie in 1945, and first used to project the world population by Notestein of the same year. The cohort-component method recognizes the strong effects of age structure on population dynamics and growth. *Cohorts* are groups of men and women born during the same time interval and representing a specific age group at any later point in time. Cohort size can only change through mortality and through migration. The size of the youngest age group is determined by age-specific fertility rates applied to the corresponding female cohorts of childbearing age. Hence, the projected size and age structure of the population at any point in the future depends exclusively on the size and age structure in the starting year, combined with assumed age-specific fertility, mortality, and migration rates over the projection period. These three variables determine all future population trends.

Population growth is also characterized by a momentum that arises from the persistence of a population s age structure. The youthful age structure of a high fertility population ensures that it will continue to grow for many years, even if the TFR declines rapidly, since for a generation progressively larger cohorts of young women will be entering their reproductive years. Population momentum has two important implications. First, population size reacts only very slowly to shifts in fertility parameters. For example, if LDC fertility rates were suddenly to decline to replacement levels, population in these countries would still continue to grow for several decades before leveling off. Second, this population inertia allows demographers to project population several decades into the future with a fair degree of confidence, just because the population that will be alive 25 years from now has already been born. Migration is the most volatile parameter since it depends in part on political conditions; fertility and mortality tend to change more slowly over time.

In this section, we summarize the assumptions and results of the population outlook published by IIASA (Lutz, 1996). These projections differ from others, such as United Nations (UN) projections, in three respects: (1) there is extensive discussion of the basis for assumed fertility, mortality, and migration trends; (2) alternative paths of future mortality and migration, in addition to fertility, are incorporated into the projections; and (3) probabilities are attached to alternative population paths.

Methodology

It has become the standard practice of most national and international agencies that prepare population projections to provide the user with a medium variant resulting from a set of assumptions considered most likely, plus low and high variants based on different assumptions and resulting in lower and higher total population sizes, respectively. Although such low and high variants are sometimes misunderstood as providing confidence intervals in a probabilistic sense, they typically are not based on any systematic variation of parameters, but simply represent alternative sets of fertility assumptions. That is the case for example of projections prepared by the UN Population Division (UN, 1998).

IIASA invited a number of experts on fertility, mortality, and migration to write chapters that summarized the existing demographic knowledge relevant for specifying the uncertainty range of future trends. These experts were asked to provide, for 13 world regions in the period 2030 2035, high and low estimates, designed to cover approximately 90% of the possible range, for fertility, mortality, and migration parameters. Central variant assumptions were taken as averages of the high low range. Assumptions for 2030 2035 were extended to 2080 2085 as described below and were held constant thereafter, until 2100.

The Future of Fertility

The level of fertility has in the long run the greatest impact on population growth. Its power derives from the multiplier effect of each additional child born today having additional children in the future.

In Developing Countries

Assumptions about future fertility for each of the world regions must consider present and past fertility. However, birth registration systems are very weak in most developing countries. The most reliable source of fertility related data are the international surveys conducted since the 1970s most important among them the World Fertility Surveys and the Demographic and Health Surveys. Table 2 gives an example of measured TFRs in some developing countries, with estimated desired fertility and the percentage of women using a contraception method modern or traditional during the last 10 years. The table shows that many women have more children than they want, even in countries where contraceptive prevalence is relatively high, e.g., in the Latin American and Caribbean region. Sub-Saharan Africa has by far the highest fertility levels (effective and desired) and the lowest levels of contraceptive prevalence. Within this region, countries with the

Table 2 Actual and desired total fertility and contraceptive prevalence in selected countries with Demographic and Health Surveys (DHS) in the 1990s

Countries	Year	TFR	Desired[a] TFR	Current % use of any method[b]
Sub-Saharan Africa				
Benin	1996	6.3	5.0	16
Burkina Faso	1992	6.9	6.0	8
Cameroon	1991	5.8	5.2	13
Central African Republic	1994	5.1	4.7	15
Comoros	1996	5.1	3.7	21
Cote d'Ivoire	1994	5.7	4.7	11
Ghana	1993	5.5	4.2	20
Kenya	1993	5.4	3.4	33
Madagascar	1992	6.1	5.2	17
Malawi	1992	6.7	5.7	22
Mali	1995	6.7	6.0	7
Namibia	1992	5.4	4.8	29
Niger	1992	7.4	7.1	4
Rwanda	1992	6.2	4.2	21
Senegal	1992	6.0	5.1	7
Tanzania	1991	6.3	5.6	10
Uganda	1995	6.9	5.6	15
Zambia	1992	6.5	5.4	15
Zimbabwe	1994	4.3	3.5	48
North Africa and Middle East				
Egypt	1995	3.6	2.6	48
Jordan	1991	5.6	3.9	35
Morocco	1995	3.3	2.2	50
Yemen	1991	7.7	6.0	10
Asia				
Bangladesh	1993	3.4	2.1	45
India	1992	3.4	2.6	41
Indonesia	1994	2.9	2.4	55
Nepal	1996	4.6	2.9	29
Philippines	1993	4.1	2.9	40
Latin America and Caribbean				
Bolivia	1994	4.8	2.7	45
Brazil	1996	2.5	1.8	77
Colombia	1995	3.0	2.2	72
Dominican Republic	1996	3.2	2.5	64
Guatemala	1995	5.1	4.0	31
Paraguay	1990	4.7	3.5	48
Peru	1996	3.5	2.2	64

[a] The desired fertility is what the fertility would be if all unwanted births were avoided.
[b] Includes use of any modern or traditional method. Excludes prolonged abstinence. Based on currently married women only.
Source: Macro International Inc., Demographic and Health Surveys on web site: (http://www.macroint.com/dhs/indicatr/datasearch.asp).

lowest use of contraceptive methods such as Niger, Mali, and Burkina Faso have the highest fertility rates. However, in all African countries, the desired TFR of women interviewed during the survey are clearly lower than the actual rates. The difference is two children in Kenya and Rwanda, and more than one in Benin, Comoros, Ghana, Uganda, and Zambia. This, together with indications of increasing contraceptive prevalence, leads to the expectation of significant fertility declines in the next decade even in sub-Saharan Africa, where fertility has remained persistently high until recently.

In North Africa, the fertility transition is already much further advanced, and half of the female population employs contraception, considerably more than in the rest of Africa. From the two countries chosen for the Middle East, Jordan shows a relatively high contraceptive prevalence more than one-third of the female population and a desired fertility of 3.9. The second country Yemen has very low current use of contraceptive method as well as very high desired fertility although the difference is almost two children. Further declines are also to be expected in this region unless dramatic events, such as a rise in pronatalist traditional religious values, change the pattern.

Three of the five Asian countries considered here, Bangladesh, India and Indonesia have TFRs between 2.9 and 3.4. The TFRs of Nepal and Philippines were still above four children but in all those countries the desired fertility was below three children, pointing to some further decline in the region.

Contraceptive prevalence is high in many Latin American countries more than 60% in Peru and the Dominican Republic and Colombia, and almost 80% in Brazil. In terms of desired and actual fertility, Latin America has an intermediate position, comparable with that of North Africa. Further fertility declines can be expected in that region.

For the longer-term future, general social and economic development, together with cultural and political factors, will determine the course of fertility decline. Although our understanding of the causal mechanism at work is incomplete, the evidence favors further fertility decline in LDCs. Therefore, even high fertility variants in these regions generally call for fertility to be lower in the period 2030 2035 than it is today. One exception is China, where the high variant assumes that the country s one-child policy is relaxed to allow fertility more in line with desired fertility levels. High-variant fertility assumptions for other regions are that the ongoing fertility transition is retarded or stalled. That this can easily occur is illustrated by the important case of India, where fertility decline was arrested between the late 1970s and the early 1990s at a TFR of about four.

Relying on the concept of *demographic transition*, population forecasters have conventionally assumed that every country will reach replacement level fertility sometime in the future, even for European populations that have already been far below that level for extended periods, implying increases in future fertility for these countries (*see* Demographic transition, Volume 5). This replacement level assumption is mechanistic, however, and is usually not supported by much substantive analysis. Therefore, our low-variant assumptions imply that fertility decline in LDCs, as in MDCs, does not stop at a TFR of 2.1, but continues to decline, carrying countries into the

Table 3 Alternative fertility, mortality, and migration assumptions for world regions in 2030–2035

	Fertility				Mortality				
	TFR	TFR in 2030–2035			Life expectancy 1995		Change in life expectancy per decade (years)		
Region	1995	Low	Central	High	Male	Female	Low	Central	High
North Africa	4.4	2.0	3.0	4.0	62.7	65.3	0.5	2.3	4.0
Sub-Saharan Africa	6.2	2.0	3.0	4.0	50.6	53.9	• 2.0	1.0	4.0
China and CPA[a]	2.0	1.5	2.3	3.0	66.4	70.1	Men 1.0	1.5	2.0
							Women 1.0	2.0	3.0
Pacific Asia	2.9	1.7	2.4	3.0	63.1	67.4	0.0	2.0	4.0
Pacific OECD	1.5	1.3	1.7	2.1	76.1	82.2	1.0	2.0	3.0
Central Asia	3.4	2.0	3.0	4.0	65.1	72.5	1.0	2.0	3.0
Middle East	5.5	2.0	3.0	4.0	65.6	68.0	0.5	2.3	4.0
South Asia	3.8	1.7	2.4	3.0	59.7	59.7	Men 0.0	1.5	3.0
							Women 0.0	2.0	4.0
Eastern Europe	1.7	1.3	1.7	2.1	67.3	75.0	1.0	2.0	3.0
European FSU[b]	1.5	1.3	1.7	2.1	61.1	72.8	Men 0.0	2.0	4.0
							Women 1.0	2.0	3.0
Western Europe	1.7	1.3	1.7	2.1	72.1	78.1	1.0	2.0	3.0
Latin America	3.1	1.7	2.4	3.0	66.3	71.5	1.0	2.0	3.0
North America	2.0	1.4	1.9	2.3	72.3	79.1	1.0	2.0	3.0

[a] Centrally Planned Asia.
[b] European part of the former Soviet Union.

range of sub-replacement fertility. The central assumptions, which represent the most likely case, result in fertility slightly above replacement level in 2030 2035 in most LDC regions substantially so in sub-Saharan Africa (see Table 3 on assumptions).

Fertility assumptions were extended into the future by assuming that in 2080 2085, the TFR in all LDC regions will lie between 1.7 and 2.1, depending on the population density of the region. Although there is some empirical evidence that high population density tends to induce lower fertility after controlling for other factors (Lutz, 1997), this procedure is still somewhat *ad hoc*, although no more so than the common alternative, which is to assume that fertility falls (or rises, as the case may be) to replacement level and then remains fixed.

In MDCs

Making assumptions about future fertility in populations that have already passed the demographic transition is easier than for today s high fertility because the range of possible fertility rates is narrower, i.e., for a modern urban society with high female labor-force participation, it is unlikely that fertility will be significantly above replacement level. It is harder because the direction of change i.e., whether fertility will increase or decrease is unclear. It seems unlikely that fertility will remain permanently below a level that is only around half the replacement level. On an individual level, the desire for children is deeply rooted, and on a societal level, extreme population aging would bring serious problems that, in turn, might lead to measures aimed at increasing fertility.

Figure 2 plots the trends in TFRs for selected European countries since 1950. It shows the baby boom during the 1960s, which was a phenomenon common to all Western industrialized societies. The increase in births was mostly due to economic prosperity that caused a marriage boom, as well as an increase of fertility within marriage. The 1970s were characterized by rapid fertility declines in almost all countries, beginning with the northern and western European countries. The southern European countries lagged somewhat behind, but subsequently experienced an even steeper decline. Greece, Italy and Spain are currently experiencing the lowest fertility levels in Europe, with TFRs as low as 1.2 children per woman. In France, and especially in Scandinavia, fertility has never declined to such low levels. In Sweden, fertility showed a temporary increase to replacement level during the 1980s. This increase seems to have been associated with generous family benefits and excellent

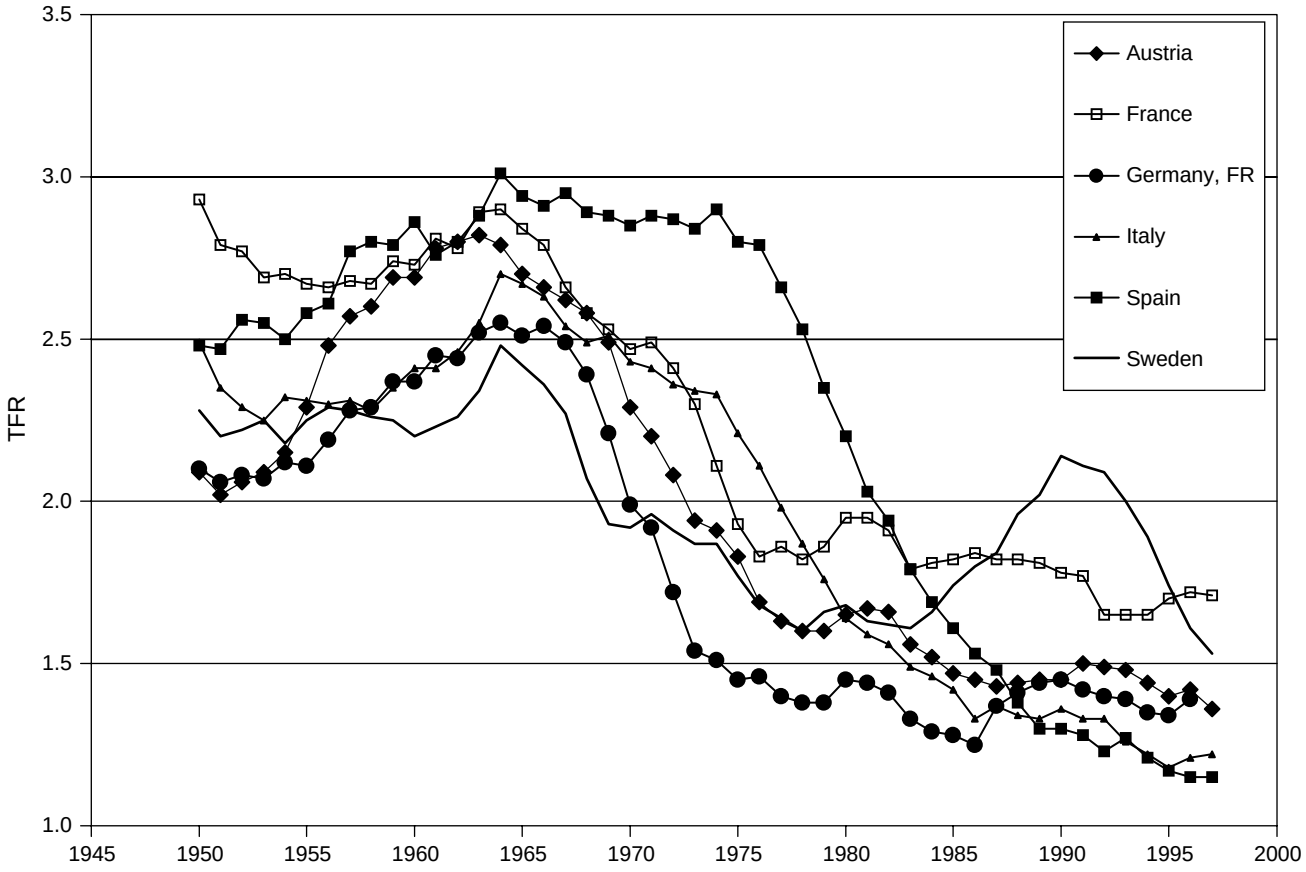


Figure 2 TFRs in selected European countries, 1950–1998

childcare provisions that facilitated the combination of child rearing and work outside the home. In the first half of the 1990s, however, fertility in Sweden came down again.

As the historical record in the developed world is mixed, the direction of future fertility trends is not clear. For their high variant, the IIASA projections assume a return-to-replacement level fertility (or slightly above) by 2030 2035 (see Table 3). The assumptions are slightly higher than replacement fertility levels for North America because of the different ethnic composition. This assumption, common to the central variants of other institutions discussed above, is theoretically supported by the homeostasis argument, which says that individual behavior is a function of the system. Therefore, fertility levels are not so much the sum of individual behavior, as one aspect of the evolution of a system where individual behavior is a function of the status of the system (see Vishnevsky, 1991). Under such a systems approach, the assumption of replacement fertility in the longer run makes sense.

There has been criticism of the homeostatic systems as too much evidence points toward continuing low fertility in industrial countries. The return to replacement fertility has been criticized as an assumed magnetic force (Westoff, 1991) without empirical support, while there are many significant arguments supporting an assumption of further declining fertility levels. These include the weakening of family ties both in terms of declining marriage rates and high divorce rates, the increasing independence and labor-market orientation of women, and the value change toward materialism and consumerism. These forces, together with increasing demands and norms concerning the resources (personal attention, education, money, etc.) to be given to children, is likely to result in fewer couples with more than one or two children, and an increasing number of childless couples.

A further argument suggesting lower fertility is that surveys still show very high proportions of unplanned births, even in societies where contraception is universally available. For example, in Canada, Japan and the US, about 30% of births are estimated to be unintended at the time of conception (Brown and Eisenberg, 1995; Montgomery, 1996; Henshaw, 1998; Alan Guttmacher Institute, 1999). A third of these are unwanted (born to women who did not want any more children) and the rest are mistimed (born to women who did not want to become pregnant until a later time). If contraceptive methods without side effects were developed that required positive action in order to become pregnant instead of forethought to avoid pregnancy, the number of births would most likely decline further.

The bulk of evidence therefore suggests that fertility will stay low or even further decline in industrial countries. How far it can decline is not clear. Golini (1998) recently suggested an absolute lower limit of about 0.7 to 0.8 children per woman based on the assumption that 20 30% of women remain childless and the rest have just one child. The IIASA projections set low values (which assume only a 5% probability of even lower levels) by 2030 2035 of 1.4 in North America and 1.3 in the other developed regions, with central variants obtained by averaging high and low assumptions.

The Future of Mortality

More so than for fertility, changes in the level of mortality are directly related to biomedical aspects of development. They also reflect changes in behavior, through the choice of individual lifestyle (such as smoking), which, at least in low mortality countries, has a significant impact on individual survival probabilities.

The uncertainties about future life expectancies are quite different in today s high and low mortality countries. Both have seen impressive increases in life expectancies. It is no longer uncommon for individuals in MDCs to survive to ages that used to be considered a biological upper limit to the human life span. Assumptions about future improvements depend on whether or not such a limit exists and will soon be reached. In regions that still have much lower life expectancy, this question is irrelevant and future mortality conditions will be determined by the effectiveness of local health services, the spread of both traditional (e.g., malaria) and new diseases (e.g., AIDS), and the general standards of living and education. Given even modest social and economic development, and apart from the AIDS epidemic, it is likely that the gap between life expectancy in MDCs and LDCs will narrow further.

In Developing Countries

Generally, most of developing countries show almost linear improvement of their life expectancies over the past 40 years, with several gaining more than 20 years of life expectancy. Probably the most impressive increase (about 25 years) occurred in China. India also had a very substantial increase from less than 40 years in 1955 to around 60 years in 1995 (Figure 3).

As with fertility, uncertainties about future mortality trends are highest for sub-Saharan Africa, partly due to the lack of reliable data. For example, the stalling of infant and child mortality declines in some countries could be due to poor data quality (Bucht, 1996). There is no doubt that modern health infrastructure developed significantly after 1960, as shown by an increasing number of health personnel *per capita*. Efforts were also made to introduce new vaccines and to build up primary health care strategies throughout the continent. Garenne (1996) concludes a survey of African mortality by stating that if in the past, transfers of technology from the West made most of the difference, then in the

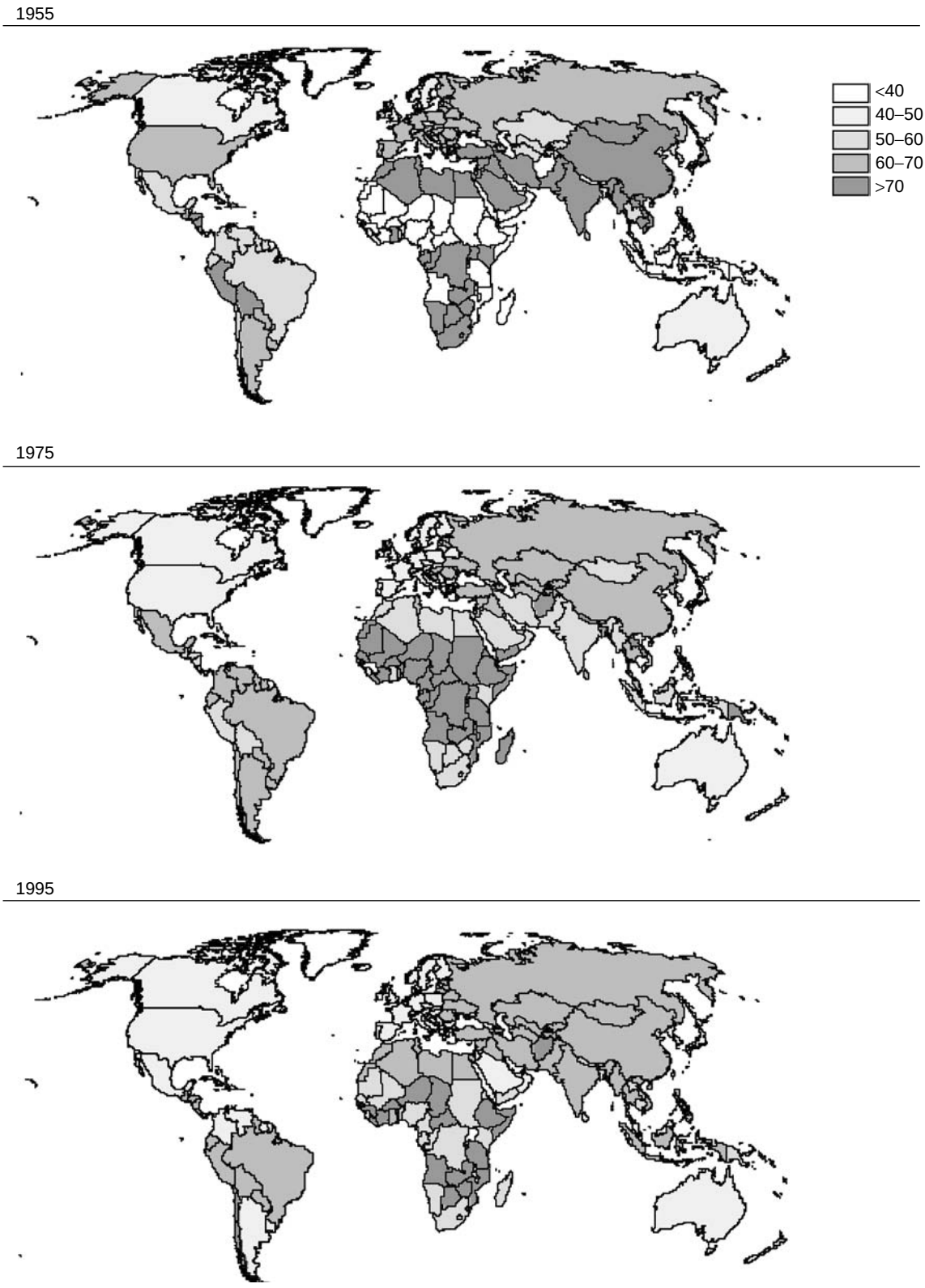


Figure 3 Life expectancy at birth in years in the calendar years

future, public health, nutrition, economic development, and education will be the key determinants of mortality decline. The spread of infectious diseases, especially HIV, as well as problems of basic subsistence, may arrest the declining trend of mortality of the past 30 years.

The sheer number of deaths due to AIDS in Africa, and hence the magnitude of the epidemic on a human and social scale, is immense. By subjecting members of the most economically active and productive groups to a wasting and premature death, AIDS is imposing an enormous economic and social toll on the continent (Ainsworth and Over, 1994). The ultimate impact of AIDS on the population of Africa as a whole will be moderate if, as assumed by Bongaarts (1996) and others, exceptionally high HIV prevalence rates exist only in a few countries, e.g., Botswana and Malawi. If high prevalence rates turn out to be more widespread, AIDS could have a significant impact on African population dynamics. Recent simulations for Botswana (International Institute for Applied Systems Analysis, 2001), which has an HIV prevalence rate estimated to be well above 30% among pregnant women, indicate that total population size may well decline or stabilize due to AIDS. Moreover, the age structure will become severely distorted and rural populations will become more vulnerable with increasing proportions of AIDS orphans and elderly. The macro economic consequences of rising health expenditures, and a worsening health status of the labor force will differ markedly depending on the economic structures of different countries.

In terms of the scenario assumptions, for sub-Saharan Africa the IIASA projections assume a two-year per decade decline in life expectancy as the low value, and a three-year increase as the high value, resulting in a one-year increase per decade for the central scenario. AIDS is also taken into consideration in the mortality assumptions for southern Asia and Southeast Asia, where the low case assumes zero improvements in mortality. The high values assume that AIDS can be largely brought under control and mortality will improve quickly following the path of today s industrialized countries.

In MDCs

Currently, Japan holds the highest life expectancy of any major country: 82 years for women and 77 years for men. Only three decades before, in 1960 1965, life expectancy was 72 for women and 67 for men, implying an average increase of more than three years per decade for Japanese women and men. Eastern Europe, on the other hand, had life expectancies similar to those of Japan during the early 1960s but has had little improvement (or even declines) since then, bringing life expectancy to about 74 for women and 63 for men at present. Western Europe and North America take an intermediate position, with life expectancies that have increased steadily by about two to three years per decade. The recent trend points toward further increases but, with indications of an imminent leveling off (Valkonen, 1991). For example, studies on occupational mortality differentials in the Nordic countries which are more homogeneous than many other countries still show significant inequalities, mostly by social class (Andersen, 1991), which may be taken as an indication of the possibility of further improvements if the higher mortality groups change their lifestyles.

The existence of a fixed biological limit to the human life span is controversial. The more traditional view implies lower rates of increase in life expectancy in the future. According to its defenders, aging is a process intrinsic to all cells of the human body. Recent and possible future mortality improvements are interpreted as a rectangularization of the survival curve through the elimination of premature deaths and resulting concentration of mortality shortly before the maximum age (Duchene and Wunsch, 1991). Lower rates may also be induced by worsening living conditions, health-impairing behaviors, or worsening environmental pollution (Day, 1991). Olshansky *et al.* (1990) also argue that an increase in life expectancy beyond 85 is unlikely because it would require a decline of more than 55% in mortality rates from all causes at all ages, which is improbable. Following these lines of reasoning, the IIASA scenarios assume an increase in life expectancy of only one year per decade (for both men and women) as the low value.

The alternative view sees that under favorable conditions, we could see much higher levels of life expectancies in the future. Aging is seen as a multidimensional process of interaction, in which partial loss of function in one organ can be compensated by other organs, and only total loss of a necessary organ system results in death (Manton, 1991). Manton *et al.* (1991) present impressive evidence from small special populations followed for short periods of time that show average life expectancies of well above 90 years. Further supporting evidence for this can be found by studying changes in very old age mortality over time. Using Swedish data since 1900, Vaupel and Lundstom (1996) show that mortality rates at ages 85, 90, and 95 have substantially declined at an accelerating rate, which over the past decades averaged 1 2% year$^{\bullet 1}$ for women and 0.5% for men. These considerations provided the motivation for the assumption of much more signi cant increases in life expectancy in the IIASA high scenario three years per decade than those assumed by most other national and international population projections. The average of the high and low scenarios results in a central scenario increase in life expectancies of two years per decade.

Assumptions on Future International Migration

International migration is the most dif cult component of population change to project. Short-term political and

economic changes in both the receiving and sending countries can affect migration trends. Net migration always results from the combination of two partly independent migration streams: that entering and that leaving a certain region for whatever reasons. Furthermore, the reliability of statistical data to assess past and present legal and illegal migration levels is meager, and more importantly, migration streams are volatile. For example, during the early 1970s, West Germany had an annual net migration gain of more than 300 000; this had declined to only around 6000 five years later, and even to 3000 during the early 1980s. During the 1985 1990 period, however, the annual net gain increased sharply again to 378 000, 100 times that of the previous period. These ups and downs exist also in the traditional countries of immigration the US, Canada, and Australia. Annual net migration to Australia declined from 112 000 in the early 1970s to 54 000 in the late 1970s. During the 1980s, it increased again to over 100 000. In this sense, labor immigration within Asia has been remarkable over the past decades. For instance, during the late 1980s more than 430 000 workers left the Philippines annually, with around 280 000 of them going to the Middle East. The Republic of Korea lost an average of 170 000 workers each year during the early 1980s; India lost 240 000 and Pakistan 130 000 annually during the same period. The largest proportion of these workers went to oil-rich countries in the Middle East. During the late 1980s these migratory streams within Asia showed significant declines.

There is potential for further increases in interregional migration due to better communications between regions, cheap mass transport, and the persisting gap between the North, which is not only richer but also rapidly aging and in need of young labor, and the South, which has many young people with rising skill levels but low income opportunities in their home countries. In addition to these economic factors, expected environmental changes in many parts of the world may cause further significant pressure to migrate to other regions where conditions are perceived to be better.

Still, the actual extent of future South North migration streams depends not only on pull and push factors, but also on the immigration policies of the receiving countries. If the European Community, for instance, were to enforce a policy of closed outside borders, this might result in a situation of almost no net migration. Hence for all regions, the low scenario value for net migration chosen was zero. This is not the same as assuming that borders will be closed to migrants; it only implies that, on average, the number of in-migrants approximately equals the number of out-migrants.

For the high migration scenarios, annual net migration gains of 2 million in North America, 1 million in Western Europe, and 350 000 in Japan/Australia/New Zealand have been assumed. For Eastern Europe special assumptions were made for the period 1990 2030, with out-migration dominating the first decade.

The distribution of migrants from the assumed sending regions to the receiving regions, and migration patterns among the developing regions, are based mostly on currently observed migratory streams as given in Zlotnik (1996). Annual net migration is assumed to be • 475 000 in northern Africa, • 510 000 in sub-Saharan Africa, • 650 000 in Central America, • 260 000 in South America, • 15 000 in western Asia, • 835 000 in Southern Asia, • 470 000 in China, and • 1 135 000 in Southeast Asia (see Table 4). Model migration schedules by Rogers and Castro (1981) were used to determine the age patterns of migrants.

The IIASA 1996 Population Projections: Results

Using the high, central, and low scenarios for fertility, mortality, and migration in 13 world regions, a number of projections of future population were performed

Table 4 Assumed central values of annual migration flows in thousands

	To				
From	North America	Western Europe	Pacific OECD	Middle East	Total
North Africa	45	125	10	8	188
Sub-Saharan Africa	58	75	20	3	156
China and CPA[a]	135	25	25	–	185
Pacific Asia	200	25	50	5	280
Central Asia	5	15	–	–	20
Middle East	8	15	5	–	28
South Asia	150	50	40	8	248
Eastern Europe	25	50	–	–	75
European FSU[b]	25	75	13	–	113
Latin America	350	45	13	2	410
Total	1000	500	175	25	1700

[a] Centrally Planned Asia.
[b] European part of the former Soviet Union.

over the 1995 2100 period. We summarize results of two distinct approaches. More detailed results can be found in Lutz (1996). A complete set of results is available from the authors and accessible on the IIASA Web site at the address show below; in this set the results are arranged by five-year age groups in five-year time intervals (1995 2100) for all 13 regions and 27 scenarios. See: http://www.iiasa.ac.at/Research/POP/docs/Population_Projections_Results.html. In the first approach, the systematic combinations of the three values (low, central, and high) for all three demographic components (fertility, mortality, and migration) result in 27 scenarios for future population and age structure. A comparison of these scenarios allows conclusions to be drawn regarding the sensitivity of results to the underlying assumptions. In the second approach, the high low range for each vital rate is interpreted as a 90% uncertainty interval and a set of probabilistic projections is performed, resulting in a distribution of outcomes for population size and age that provides new insight into the likelihood of demographic trends over the next 100 years.

Table 5 shows the total population size in 1995, 2020, 2050, and 2100 for the nine scenarios resulting from the combination of the three fertility assumptions (low, central, and high) with the three mortality assumptions. Migration is held at its central level in all of these scenarios. Results are presented at the level of major world regions. The central scenario (combination of central fertility with central mortality assumptions) describes the most likely future population path under our assumptions. If these central assumptions hold for all world regions, then the total world population will increase from 6.0 billion in 2000 to 7.9 billion in 2020, 9.9 billion in 2050. This scenario implies a clear leveling off of world population growth toward the end of the 21st century. The world population reaches an all-time high of 10.6 billion around 2080, and starts to decline somewhat to 10.4 billion by the end of the 21st century (see Figure 4).

Table 5 Projections of total population (in millions) according to alternative fertility and mortality assumptions (and central-migration assumptions) in 2020, 2050, and 2100

		Fertility								
		Low			Central			High		
Mortality	1995	2020	2050	2100	2020	2050	2100	2020	2050	2100
High										
Africa	720	1202	1312	707	1278	1710	1484	1355	2173	2739
Asia-East	1956	2239	1979	994	2412	2601	2308	2589	3369	4650
Asia-West	1445	2073	2210	1241	2181	2741	2432	2290	3352	4321
Europe	808	769	593	291	811	716	546	854	860	961
Latin America	477	644	687	454	686	867	927	729	1080	1700
North America	297	335	321	251	354	386	423	374	463	697
Less developed	4451	6011	6071	3330	6404	7780	7040	6802	9810	13 228
More developed	1251	1250	1032	607	1319	1241	1081	1388	1486	1841
World	5702	7261	7103	3937	7723	9021	8120	8191	11 300	15 070
Central										
Africa	720	1252	1562	1140	1332	2040	2366	1414	2599	4349
Asia-East	1956	2268	2116	1205	2444	2760	2704	2621	3554	5348
Asia-West	1445	2117	2421	1634	2228	2995	3136	2338	3659	5508
Europe	808	782	640	342	825	766	624	868	912	1075
Latin America	477	650	722	530	693	906	1056	737	1125	1910
North America	297	339	340	283	359	406	467	378	483	757
Less developed	4451	6138	6694	4433	6541	8554	9137	6946	10 765	16 915
More developed	1251	1270	1107	700	1340	1319	1216	1410	1568	2033
World	5702	7408	7802	5134	7879	9874	10 350	8356	12 330	18 950
Low										
Africa	720	1298	1799	1662	1381	2344	3348	1467	2983	6058
Asia-East	1956	2296	2257	1440	2473	2918	3090	2652	3733	5960
Asia-West	1445	2158	2624	2073	2270	3235	3847	2383	3939	6627
Europe	808	796	692	402	839	819	705	882	968	1185
Latin America	477	656	756	612	700	945	1180	743	1169	2096
North America	297	344	360	318	363	427	511	383	505	812
Less developed	4451	6257	7301	5701	6665	9285	11 325	7080	11 643	20 524
More developed	1251	1291	1188	806	1360	1402	1355	1431	1654	2214
World	5702	7547	8488	6507	8026	10 690	12 680	8510	13 300	22 740

Source: Lutz, 1996.

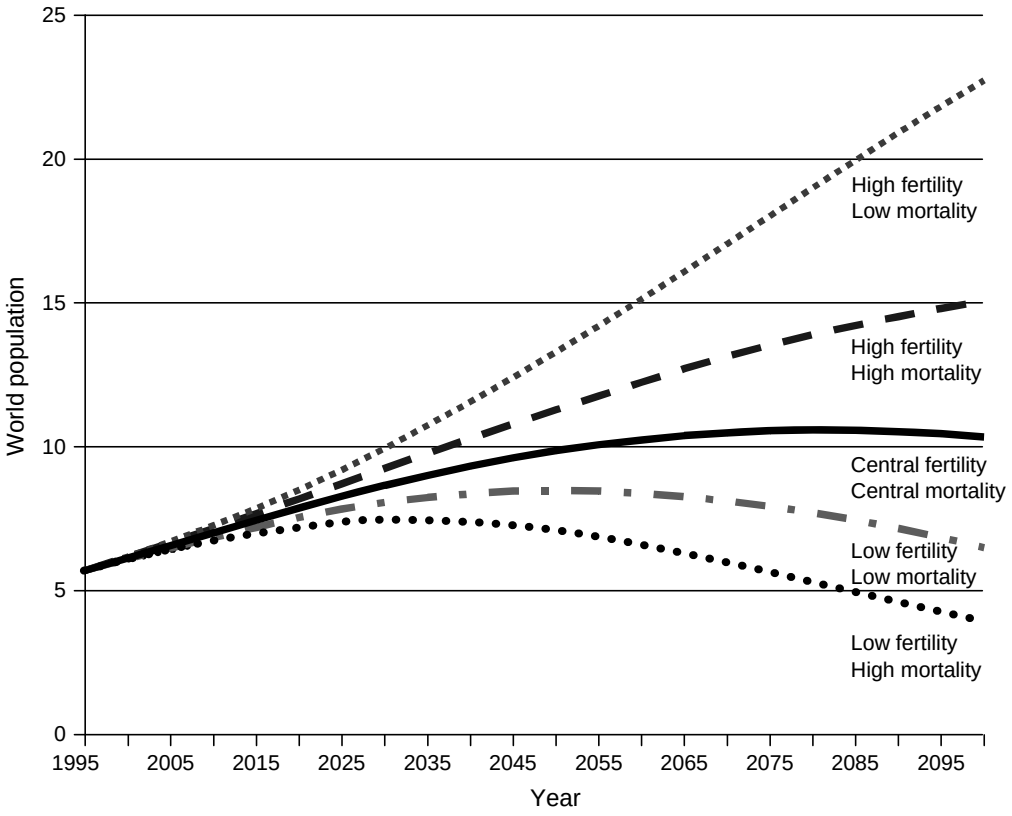


Figure 4 Selected projected paths of world population size (five scenarios, central migration)

The highest population path combines low mortality with high fertility, producing a world population of 13.3 billion by 2050 and 22.7 billion by 2100. The lowest population path results from combining high mortality with low fertility in every world region. In this case, world population levels off during the 2030s, reaches 7.1 billion in 2050, and by 2100 declines to about 4 billion. However, these extremes are considered unlikely. If fertility remains high in LDCs, it is likely to be a result of slow economic and social progress, in which case it is more likely that mortality will be high, not low. Similarly, if rapid change results in low fertility, it is more likely that mortality will be low, not high.

Scenarios corresponding to more plausible combinations of assumptions might be called the rapid demographic transition scenario (low mortality and low fertility) and slow demographic transition scenario (high mortality and high fertility). These less extreme (from the standpoint of population growth) but more meaningful scenarios are given in the upper right and lower left corners of Table 5. In the first case, world population rises to 15.1 billion in 2100; in the second case, it rises to 8.5 billion in 2050 before declining slowly to 6.5 billion in 2100.

These global totals hold only if fertility and mortality trends are perfectly correlated across all world regions (for example, all regions simultaneously follow the high fertility assumptions). In the real world, however, it is much more likely that one region will follow one scenario while another follows an alternative scenario. Why, for example, should the fertility trend in Latin America be correlated to that in Eastern Europe? Table 4 allows the user to define any mix of scenarios in different world regions and aggregate them into global population sizes. All these regionally mixed scenarios will lie between the extremes because diverging trends in different regions will partly offset each other.

The global world population growth results from a great diversity of regional patterns. Under the central scenario, the population on the African continent would grow the most rapidly, almost tripling over the next five decades. Accordingly, the share of the world population that lives in Africa increases from 13% in 1995 to 21% in 2050. The population of East Asia increases by 41% under the central assumption until 2050; it doubles in west Asia during the same period. Latin America currently has 8% of the world s population; it is likely to maintain this proportion because its growth is expected to be close to that of the world average. The proportion of the world population living in Europe falls from the current 14% to 7 8% by 2050. Europe and North America both grow below the world average under all scenarios.

Table 6 Projections of percentages of the population above age 60 according to alternative fertility and mortality assumptions (and central-migration assumptions) for 2020, 2050, and 2100

	Fertility									
		Low			Central			High		
Mortality	1995	2020	2050	2100	2020	2050	2100	2020	2050	2100
High										
Africa	5.0	6.2	12.5	28.5	5.8	9.6	20.0	5.5	7.5	14.8
Asia-East	9.4	16.1	29.7	34.9	15.0	22.6	24.6	14.0	17.4	18.0
Asia-West	6.6	9.9	18.5	34.1	9.4	14.9	24.4	9.0	12.2	18.2
Europe	17.8	25.1	38.1	37.5	23.8	31.6	28.5	22.6	26.3	22.0
Latin America	7.6	12.4	24.4	33.1	11.6	19.4	23.8	10.9	15.5	17.8
North America	16.4	24.5	32.9	33.2	23.1	27.4	26.3	21.9	22.8	20.6
Less developed	7.2	11.2	21.1	32.9	10.5	16.4	23.3	9.9	13.0	17.3
More developed	17.7	25.7	36.9	36.2	24.4	30.7	28.1	23.1	25.7	21.9
World	9.5	13.7	23.4	33.4	12.9	18.4	24.0	12.2	14.7	17.8
Central										
Africa	5.0	6.3	13.1	30.7	6.0	10.0	21.5	5.6	7.9	15.8
Asia-East	9.4	16.6	31.9	40.1	15.4	24.5	28.4	14.3	19.0	20.9
Asia-West	6.6	10.1	19.7	37.7	9.6	16.0	27.2	9.2	13.1	20.3
Europe	17.8	25.9	41.4	44.1	24.6	34.6	34.0	23.4	29.1	26.3
Latin America	7.6	12.7	26.0	37.5	11.9	20.7	27.1	11.2	16.7	20.3
North America	16.4	25.2	35.9	39.1	23.8	30.0	31.2	22.6	25.2	24.6
Less developed	7.2	11.5	22.2	36.4	10.8	17.4	25.9	10.1	13.8	19.2
More developed	17.7	26.5	40.2	42.5	25.1	33.7	33.4	23.9	28.4	26.1
World	9.5	14.0	24.8	37.2	13.2	19.6	26.8	12.5	15.7	19.9
Low										
Africa	5.0	6.5	14.3	35.8	6.1	11.0	25.2	5.8	8.6	18.5
Asia-East	9.4	17.0	34.5	46.6	15.8	26.7	33.3	14.7	20.9	24.3
Asia-West	6.6	10.4	21.3	43.0	9.9	17.3	31.2	9.4	14.2	23.4
Europe	17.8	26.8	44.9	51.1	25.4	38.0	40.0	24.2	32.1	31.1
Latin America	7.6	13.0	27.8	42.8	12.2	22.3	31.1	11.5	18.0	23.2
North America	16.4	25.9	38.9	45.0	24.5	32.9	36.1	23.2	27.8	28.5
Less developed	7.2	11.7	23.8	41.6	11.0	18.7	29.8	10.4	15.0	22.1
More developed	17.7	27.3	43.6	49.1	25.9	36.9	39.0	24.7	31.3	30.7
World	9.5	14.4	26.6	42.5	13.5	21.1	30.8	12.8	17.0	22.9

Source: Lutz, 1996.

Table 6, which gives results for population aging, here indicated by the proportion of the population above age 60, shows that significant population aging is inevitable. Universal population aging is the most consistent and dominant population trend of this century. Even in the slow demographic transition scenario combining high fertility with high mortality (which also tends to reduce aging), the proportion above age 60 increases from the current 9.5% to 17.8%. In the rapid demographic transition scenario combining low fertility with low mortality, the proportion above age 60 increases to 42.5% of the total population. Considered in terms of increases in absolute size of the population above age 60, rather than as increases in the proportion, the rapid and slow demographic transition scenarios, as well as the central scenario, imply a fivefold increase in the 60• age group in the long run. Among the developing regions, China together with its centrally planned economy, is certain to experience the most rapid aging because fertility has already declined dramatically. Within less than four decades China s old age dependency burden will be higher than North America s today.

In MDCs, although the proportion of elderly is already quite high, further and signi cant increases are inevitable. For example, in North America, the population over 60 years of age represented about 16% of the total population in 1995. As a result of the large cohort born during the post-World War II baby boom, the share of the population below age 60 will still increase in the next decade. However, when the baby boomers reach age 60, after 2010, the proportion of elderly will soar, reaching between 23 and 34% by 2030, depending on the fertility, mortality and migration assumptions. The central most likely scenario projects that 28% from North America will be 60 years of age or older by 2030.

Deceleration of population growth is a broadly supported goal of international population policy. The estimates in Tables 5 and 6 reveal, however, two quali cations that should be kept in mind. First, the faster that fertility

declines, the faster that mortality is likely to decline as well, the more population aging is accelerated and intensified. Second, low population growth may be desirable but is associated with the highest mortality, which is hardly a desirable option. On both counts, population stabilization is not a goal that should be pursued blindly.

Fertility assumptions, as was mentioned at the beginning of this section, have a far greater impact on projected long-term population size than do mortality assumptions. Holding fertility and international migration assumptions at their central values and allowing only mortality assumptions to vary, the forecast range for 2100 is between a population of 8.1 billion in the high mortality case and 12.7 billion in the low mortality case. Holding international migration and, this time, mortality assumptions at central values and allowing only fertility to vary, the range of variation is between 5.1 billion in the low fertility scenario and 19.0 billion in the high fertility scenario.

Considered region by region, population size is more sensitive to the future course of fertility in LDCs than in MDCs, where the range of possible variation in behavior is lower and the base is smaller. Fertility trends are the dominant influence on population size, but sensitivity to mortality trends is significant enough to justify alternative mortality assumptions. Fertility is also the most important factor governing population aging. Under central mortality and migration assumptions, the high fertility scenario implies 19.9% of the population over age 60 in the year 2100 and the low fertility scenario implies 37.2% a spread of over 17 percentage points (see Table 6). Under central fertility and migration assumptions, however, the high mortality scenario implies 24% of the population over age 60 and the low mortality scenario implies 30.8%, a spread of only about seven percentage points. Migration assumptions have even less impact on aging than mortality. Taking Western Europe as an example, if mortality and fertility assumptions are held at central levels and migration assumptions are varied, the forecast range for the proportion of the population over age 60 in 2100 is only between 34.1% in the high migration case and 36.5% in the low migration case. In terms of absolute change in the proportion of the population over age 60, both regions are about equally sensitive to the future course of fertility. For example, in 2050 the difference between the low and central fertility scenarios (given central mortality and migration assumptions) is 6.5 percentage points (40.2% as opposed to 33.7%) for MDCs and 4.8 percentage points (22.2% as opposed to 17.4%) for LDCs.

Probabilistic Projections

All the scenarios described above represent more or less combinations of assumptions, which have no probabilistic interpretation. There is no one simple way to apply probabilistic methods to population projections (see Lutz *et al.*, 1999; Lee, 1999). The IIASA projections use an approach that is based on expert judgment. The procedure fits a normal distribution to the three values (high, central, and low) that resulted from the expert discussions, with 90% of the cases lying between the high and low values. Sensitivity analysis showed that the results are relatively insensitive to the chosen value for the confidence intervals. Thus, the high low range may be thought to cover 85 95% of all cases (see Lutz, 1996). Results were derived through a set of 4000 simulations that randomly combined fertility, mortality, and migration paths from the three normal distributions for the 13 world regions. These simulations also considered the possibility that fertility and mortality trends may be correlated within regions (e.g., high fertility in sub-Saharan Africa is more likely to go hand in hand with high mortality than low mortality) and that regional trends may be either independent of each other (e.g., fertility in sub-Saharan Africa uncorrelated with fertility in Latin America) or correlated.

Figure 5 shows the distribution of future population sizes derived from the full set of 4000 projections at five-year intervals to the year 2100. The high and low boundaries give the range into which 95% of all cases fall. The upper line indicates that there is an unlikely possibility of almost linear population growth between 1995 and 2100. The lower line shows that there is also an equally unlikely possibility that the world s population will peak in the middle of the 21st century and fall thereafter to below 6 billion by 2100. The much more probable range of future paths (between the 0.2 and 0.8 fractiles, covering 60% of all cases) is remarkably small. By 2050, this uncertainty range is less than 1.5 billion people, and by 2100 it doubles to about 3 billion people. The figure also shows that in more than 60% of all cases, total world population levels

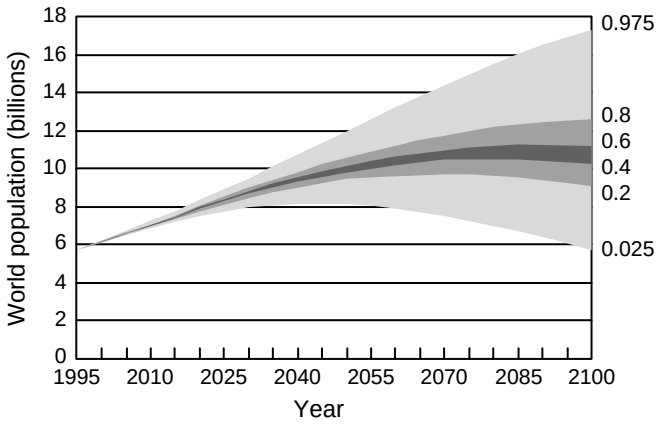


Figure 5 Merged distribution of the size of the world's population (in billions), 1995–2100. The figures on the right-hand side show the proportion of cases in which the percentages falls below the line indicated

off during the second half of this century, or even starts to decline. Given that the world population in mid-1996 was estimated at 5.8 billion, Lutz *et al.* (1996b, 1997) find that the probability of a doubling that is, reaching or surpassing the 11.6 billion mark at any point during the next century is only 33%. In other words, there is a two-thirds probability that world population will not double by the end of the 21st century.

When looking at regional disparities, Sub-Saharan Africa displays the largest range of uncertainty in future population size, with a 95% confidence interval in 2100 ranging between 578 million and 4.345 billion around a mean of 1.9 billion. This large range results from the unusually large uncertainty surrounding trends in fertility and mortality in the region, in addition to the assumption that the two trends will be uncorrelated. Lutz *et al.* (1996a) have shown that if fertility and mortality are instead assumed to be positively correlated within the region, the range of uncertainty becomes significantly smaller. The distribution of the percentage of the global population over age 60 is shown in Figure 6. All the lines are rising, indicating near certainty that the percentage of older people in the population will rise over time. In 2050, the median is 20% (with a 95% uncertainty interval between 15% and 26%), compared with 9.5% in 1995. By 2100, the median increases to 27%, with a 95% uncertainty range of 19% to 41%. In other words, there is a 95% chance that the proportion of the elderly will at least double over the next century. In the most likely case, it will almost triple and it may even more than quadruple. The uncertainty in the percentage above age 60 grows significantly during the second half of the 21st century, due mainly to the uncertainty in future old-age mortality when combined with uncertainty in fertility. To summarize, three near certainties emerge from the range of scenarios and

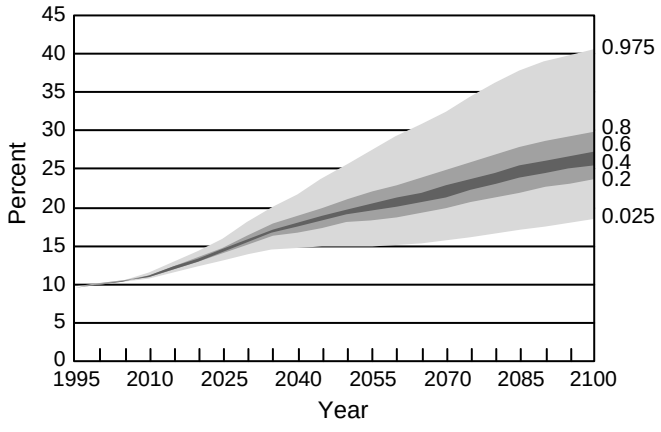


Figure 6 Merged distribution of the world's population in the 60• aged group (in percent), 1995–2100. The figures on the right-hand side show the proportion of cases in which the percentages fall below the line indicated

from the probabilistic projections. World population will increase substantially from its current level. Even in the lowest growth scenario, population increases by close to 2 billion before commencing its decline. However, a doubling of world population has become unlikely in the face of recent fertility declines. The distribution of world population will continue to shift toward LDCs. Even assuming rapid fertility decline and little improvement in mortality, LDCs still account for a rising share of world population. The world population will continue to age. Probabilistic projections show that a doubling of the proportion above age 60 is a near certainty.

CONCLUSION

This article has provided, through a wide range of extreme scenarios as well as a probabilistic population projection exercise, many possible numbers for the world population size for the next 100 years. In a nutshell, it has identified three near certainties in the world population outlook: further population growth, further tilt toward LDCs in the world distribution of population, and significant population aging. These robust findings embody the sectoral expertise of an international panel of researchers. They support the conventional wisdom that the speed of fertility decline in LDCs and, in the longer term, the ultimate fertility level reached there, are the key uncertainties from the standpoint of the future size and age structure of the total human population. Our understanding of the factors influencing fertility is far from complete. However, research supports the view that a combination of effective national family planning programs, socioeconomic development, and empowerment of women all with a special emphasis on education results in fertility decline.

Since it positively contributes to all three, investing in female basic education is likely the policy intervention that most strongly accelerates fertility decline. While less important than fertility, future mortality trends will also have a significant impact on future global population size and age structure.

It seems safe to predict, based on the projections summarized above, that the international population community, as well as policymakers, will gradually shift their concern away from population size and rate of growth to population age structure, specifically, aging. While the MDCs are entering unknown territory in terms of the proportion of the population over age 60, the LDCs are approaching the age structure of today s MDCs. Thus, under the central scenario, the proportion of the population over age 60 in MDCs is expected to rise from 17.7% in 1995 to 33.7% in 2050 while in LDCs it will rise from 7.2% to 17.4%. Based on the sensitivity analyses presented here, the extent of population aging is

equally sensitive to choice of fertility scenario in the two regions.

This expected new emphasis on population age structure, however, should not hide the fact that in some parts of Africa, Asia, and Latin America, populations are still expected to more than double over the coming decades, and many of these population live in fragile ecological systems.

REFERENCES

Ainsworth, M and Over, M (1994) AIDS and African Development, *World Bank Res. Observer*, **9**(2), 203 240.

Alan Guttmacher Institute (1999) *Sharing Responsibility: Women, Society, and Abortion Worldwide*, Alan Guttmacher Institute, NY.

Andersen, O (1991) Occupational Impacts on Mortality Declines in the Nordic Countries, in *Future Demographic Trends in Europe and North America: What Can We Assume Today?* ed W Lutz, Academic Press, London, 1 54.

Bongaarts, J (1996) Global Trends in AIDS Mortality, in *The Future Population of the World: What Can We Assume Today?* ed W Lutz, Earthscan, London, 170 195.

Brown, H S and Eisenberg, H L, eds (1995) *The Best Intentions: Unintended Pregnancy and the Wellbeing of Children and Families*, National Academy Press, Washington, DC.

Bucht, B (1996) Mortality Trends in Developing Countries: a Survey, in *The Future Population of the World: What Can We Assume Today?* ed W Lutz, Earthscan, London, 133 148.

Cleland, J and Rodriguez, G (1988) The Effects of Parental Education on Marital Fertility, *Popul. Stud.*, **42**(3), 419 442.

Cleland, J and Wilson, C (1987) Demand Theories of the Fertility Transition: an Iconoclastic View, *Popul. Stud.*, **41**(1), 5 30.

Cochrane, S (1979) *Fertility and Education: What Do We Really Know?* Johns Hopkins University Press, Baltimore, MD.

Day, L H (1991) Upper-age Longevity in Low-Mortality Countries: a Dissenting View, in *Future Demographic Trends in Europe and North America: What Can We Assume Today?* ed W Lutz, Academic Press, London, 117 128.

Duchene, J and Wunsch, G (1991) Population Aging and the Limits to Human Life, in *Future Demographic Trends in Europe and North America: What Can We Assume Today?* ed W Lutz, Academic Press, London, 27 40.

Garenne, M (1996) Mortality in Sub-Saharan Africa: Trends and Prospects, in *The Future Population of the World: What Can We Assume Today?* ed W Lutz, Earthscan, London, 149 169.

Golini, A (1998) How Low Can Fertility Be? An Empirical Exploration, *Popul. Dev. Rev.*, **24**(1), 59 74.

Henshaw, S K (1998) Unintended Pregnancy in the United States, *Fam. Planning Perspec.*, **30**, 24 29, 46.

International Institute for Applied Systems Analysis (2001) *Botswana s Future, Modeling Population and Sustainable Development Challenges in the Era of HIV/AIDS*, IIASA, Laxenburg, Austria.

Kasarda, J D, Billy, J O G, and West, K (1986) *Status Enhancement and Fertility: Reproductive Responses to Social Mobility and Educational Opportunity*, Academic Press, New York.

Kinsella, K and Gist, Y J (1995) *Older Workers, Retirement, and Pensions: a Comparative International Chartbook*, Report No. IPC/95 2RP, Bureau of the Census, Washington, DC.

Lee, R (1999) Probabilistic Approaches to Population Forecasting, in *Frontiers of Population Forecasting, Population and Development Review, a Supplement to Vol. 24, 1998*, eds W Lutz, J Vaupel, and D A Ahlburg, 42 58.

Leslie, P H (1945) On the use of Matrices in Certain Population Mathematics, *Biometrika*, **33**, 183 212.

Lutz, W, ed (1996) *The Future Population of the World: What Can We Assume Today?* Earthscan, London.

Lutz, W (1997) *Determinants and Consequences of the World s Most Rapid Fertility Decline on the Island of Mauritius*, Paper presented at the meeting of the Population Association of America, 27 29 March, Washington, DC.

Lutz, W, Goldstein, J R, and Prinz, C (1996a) Alternative Approaches to Population Projection, in *The Future Population of the World: What Can We Assume Today?* ed W Lutz, Earthscan, London, 14 44.

Lutz, W, Sanderson, W, and Scherbov, S (1996b) Probabilistic Population Projections Based on Expert Opinion, in *The Future Population of the World: What Can We Assume Today?* ed W Lutz, Earthscan, London, 397 428.

Lutz, W, Sanderson, W, and Scherbov, S (1997) Doubling of World Population Unlikely, *Nature*, **387**, 803 805.

Lutz, W, Goujon, A, and Doblhammer-Reiter, G (1999) Demographic Dimensions in Forecasting: Adding Education to Age and Sex, in *Frontiers of Population Forecasting, Population and Development Review, a Supplement to Vol. 24, 1998*, eds W Lutz, J Vaupel, and D A Ahlburg, Population Council, New York, 42 58.

Manton, K G (1991) New Biotechnologies and the Limits to Life Expectancy, in *Future Demographic Trends in Europe and North America: What Can We Assume Today?* ed W Lutz, Academic Press, London, 97 115.

Manton, K G, Stallard, E, and Tolley, H D (1991) Limits to Human Life Expectancy, *Popul. Dev. Rev.*, **17**(4), 603 637.

Montgomery, M R (1996) Comments on Men, Women, and Unintended Pregnancy, in *Fertility in the United States: New Patterns, New Themes*, eds J B Casterline, R D Lee, and K A Foote, Supplement to *Popul. Dev. Rev.*, **22**, 100 106.

Notestein, F W (1945) Population the Long View, in *Food for the World*, ed T W Schultz, University of Chicago Press, Chicago, IL, 36 57.

Olshansky, S J, Carnes, B A, and Cassel, C (1990) In Search of Methuselah: Estimating the Upper Limits of Human Longevity, *Science*, **2**, 634 640.

Rogers, A and Castro, L (1981) *Model Migration Schedules*, RR-81 30, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Thomas, N (1991) Land, Fertility, and the Population Establishment, *Popul. Stud.*, **45**, 379 397.

United Nations (1987) *Fertility Behavior in the Context of Development, Population Studies*, 100.ST/ESA/SER.A/100, Department of International Economic and Social Affairs, New York.

United Nations (1993) *World Population Prospects: the 1992 Revision*, United Nations, New York.

United Nations (1998) *World Population Prospects: the 1998 Revision*, United Nations, New York.

Valkonen, T (1991) Assumptions about Mortality Trends in Industrialized Countries: a Survey, in *Future Demographic Trends in Europe and North America: What Can We Assume Today?* ed W Lutz, Academic Press, London, 3 25.

Vaupel, J W and Lundstrom, H (1996) The Future Mortality at Older Ages in Developed Countries, in *The Future Population of the World: What Can We Assume Today?* ed W Lutz, Earthscan, London, 278 295.

Vishnevsky, A (1991) Demographic Revolution and the Future of Fertility: a Systems Approach, in *Future Demographic Trends in Europe and North America: What Can We Assume Today?* ed W Lutz, Academic Press, London, 257 270.

Westoff, C (1991) The Return to Replacement Fertility: a Magnetic Force? in *Future Demographic Trends in Europe and North America: What Can We Assume Today?* ed W Lutz, Academic Press, London, 227 233.

Whelpton, P K (1936) An Empirical Method of Calculating Future Population, *J. Am. Stat. Assoc.*, 31, 457 473.

Zlotnik, H (1996) Migration to and From Developing Regions: a Review of Past Trends, in *The Future Population of the World: What Can We Assume Today?* ed W Lutz, Earthscan, London, 299 335.

Trends in Global Emissions: Carbon, Sulfur, and Nitrogen

ARNULF GRÜBLER

International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria

Global anthropogenic emissions of carbon, sulfur and nitrogen are reviewed by major category. Both present and historical emissions are reviewed including respective uncertainty ranges. The article concludes with a brief discussion of how to relate trends in emissions to the evolution of their underlying driving forces using energy-related carbon emissions as an example.

Main conclusions for the three gases are summarized below:

Anthropogenic emissions of carbon are small compared to the sizes of natural carbon reservoirs and the annual ow rates between them. Yet they constitute important perturbations of the natural carbon cycle, well quanti ed for industrial carbon emissions (mostly the burning of fossil fuels), but remaining uncertain for carbon emissions arising from biomass burning and land-use changes. Since the onset of the Industrial Revolution, some 530 billion tons of carbon (Pg C) gross emissions have been released by human activities to the atmosphere, part of which was balanced by (highly uncertain) terrestrial biospheric carbon sinks. Net carbon emissions over the period 1800- 2000 are estimated at 280 Pg C. After accounting for oceanic uptake, these emissions have increased the atmospheric carbon loading by 200 Pg C, or by one-third natural levels. In the absence of stringent climate policies, between 770 and 2540 Pg C could be released to the atmosphere over the next 100 years. A signi cant ampli cation of humankind's "discernible in uence on the climate system" is thus likely.

Anthropogenic emissions of sulfur have surpassed natural ows ever since the rst quarter of the 20th century. Since about 1975 global emissions (mostly from burning of sulfur-rich fossil fuels such as coal and oil) have stayed roughly constant. A continued rise in sulfur emissions in developing countries has been compensated by drastic declines in emissions in Organization for Economic Cooperation and Development (OECD) countries as a result of environmental policies. Similar policies will be required also in developing countries to counter local and regional environmental impacts of high sulfur emissions. A slowdown and ultimately a trend reversal of sulfur emission growth trends also in developing countries could thus be likely over the next few decades.

Emissions of nitrogen take a variety of forms and originate from a wide array of different sources including: nitrogen oxides, principally formed in high temperature combustion (burning of fossil fuels); ammonia, principally arising from animal manure; and nitrous oxide (N_2O), a powerful greenhouse gas, principally arising from soil microbiological process as well as agricultural activities and animal manure. Because of the multitude of gases and sources, emission estimates are poorly understood and uncertain. There is also a paucity of both past and present inventory data and future emission scenarios. More research is therefore needed before considering policy interventions to control these gases beyond well established source categories (e.g., nitrogen oxide emissions from automobiles, or nitrous oxide emissions from adipic acid production).

CARBON

Anthropogenic emissions of carbon arise predominantly from the combustion of fossil fuels, the burning of biomass, as well as oxidation of soil carbon following land-use changes (e.g., deforestation). These anthropogenic emissions are embedded in large fluxes between the (yet larger) carbon reservoirs (atmosphere, oceans, vegetation and soils, as well as deposits of fossil fuels) that characterize the carbon cycle (Houghton and Skole, 1990; Schimmel *et al.*, 1995, *see* Carbon Cycle, Volume 2). Due to the magnitude of the reservoirs and the annual exchange rates between them, that are known only approximately, human alterations to the carbon cycle are comparatively

small, are difficult to quantify and are hence subject to considerable uncertainties, especially for land-use change carbon emissions.

The Intergovernmental Panel on Climate Change (IPCC) summarizes current knowledge on the size of the respective carbon reservoirs and fluxes for the 1980s as follows (Schimmel *et al.*, 1995; Nakicenovic *et al.*, 1996): the largest carbon reservoirs are the oceans (close to 40 000 Pg C [Pg C = Petagrams (10^{15} g) elemental carbon, 1 Pg = 1 Gt (gigatons, or one billion metric tons), 1 Pg = 1000 Tg (teragrams, or 10^{12} g)]), deposits of fossil fuels (coal, conventional and unconventional oil and gas (such as *inter alia* oil shales, tar sands, gas in tight formations, in aquifers, and coal seams), including methane hydrates: 25 000 Pg C), vegetation and soils (2200 Pg C), and finally the atmosphere (750 Pg C). Only the latter carbon reservoir size is directly measurable, and the increasing atmospheric carbon dioxide concentrations are proof that human activities and resulting emissions are indeed altering the carbon cycle, the composition of the atmosphere, and hence the radiative (heat) balance of the planet (global warming). The global carbon cycle involves the annual exchange of some 90 Pg C between the atmosphere and the oceans, and some 60 Pg C between the biosphere (vegetation and soils) and the atmosphere. The inter-annual variations of measured atmospheric carbon dioxide concentrations (5.4 parts per million by volume (ppmv) in 1998 at Mauna Loa, Hawaii (Keeling and Whorf, 1999)) clearly indicate the importance of seasonal vegetation cycles in biospheric carbon fluxes. The inter-annual variation of 5.4 ppmv corresponds to some 11 Pg C. Compared to the size of annual fluxes that characterize the carbon cycle and its inter-annual variations, anthropogenic alterations to the carbon cycle are comparatively small and hence impossible to measure directly. Emission estimates need therefore to be based on inventory data linking socio-economic activity data such as fossil fuel use or land use changes, with specific emission factors in the case of fossil fuel emissions, or micro- and meso-scale models of how soil and vegetation carbon reservoirs, and the exchange rates between them, are affected by human activities.

Currently (AD 2000), human-induced alterations to carbon flows include emissions of some $6.6(\pm 0.6)$ Pg C year^{-1} from industrial activities (mostly the burning of fossil fuels) that constitute a net addition to natural carbon fluxes, albeit not necessarily in the same year. In addition, estimates indicate a net additional flux of about one (uncertainty range: 0 2.8) Pg C year^{-1} from the burning of biomass and changes in soil carbon in conjunction with land use changes. This estimated net biospheric flux is the difference between estimates of a variety of carbon sources and sinks (see discussion below) and is affected by high uncertainty margins. This is illustrated in the following simplified representation of the global annual carbon balance in Pg C for the 1980s (Schimmel *et al.*, 1995; Houghton, 1999).

$$\begin{aligned}\text{atmospheric increase} &= \text{industrial emissions}\\ &\quad + \text{net land-use emissions} - \text{ocean uptake}\\ &\quad - \text{residual (missing sink)}\\ 3.3(\pm 0.2) &= 5.5(\pm 0.5) + 1.1(0 - 2.8) - 2.0(\pm 0.8)\\ &\quad - 1.3(0 - 3.3)\end{aligned}$$

Table 1 summarizes the main, contemporary anthropogenic industrial and biotic carbon emission sources as well as associated uncertainty ranges. The numbers exclude carbon dioxide fluxes arising from the metabolism of human and animal populations for which inventory data are lacking.

Industrial sources of carbon emissions are comparatively well understood and quantified. The dominant source is fossil energy production, conversion, and use, i.e., burning of fossil fuels (6 Pg C). Secondary sources are the manufacture of cement (0.2 Pg C) and the flaring of natural gas (0.06 Pg C) which occurs when gas is produced together with crude oil, but due to the absence of local markets or infrastructures cannot be used as a fuel and is flared (burned) instead. Re-injection into the oil reservoir for repressuring avoids the need for flaring and is practiced in many oil fields (CEDIGAZ, 1998). Yet even the modest costs involved seem not to justify the preservation of natural gas resources for future uses in many locations. Hence flaring continues (albeit declining). Globally some 120 billion m^3 of natural gas are flared, about equal to the global trade in LNG (liquefied natural gas) or the natural gas consumption of Germany, Austria, Switzerland, and Belgium combined (CEDIGAZ, 1998). Minor carbon emission sources include the manufacture of lime and ammonia (Harvey, 1999), the release of carbon dioxide from carbon dioxide rich fossil fuel reservoirs, and the production of carbon dioxide from natural reservoirs for tertiary oil recovery and for the food (beverage) industry. Even if small when compared to natural carbon fluxes, carbon emissions are the largest material flow of the industrial metabolism. The 6 Pg C of industrial carbon emissions compares, for instance, to an annual production volume of 2.5 Pg for the seven most important industrial commodities (in decreasing order of tonnage: cement, steel, paper, fertilizers, glass, aluminum, copper, cf. Grubler, 1998a).

Emission estimates for industrial sources are quite accurate ($\pm 10\%$). As energy is a preferred target for taxation, production, trade, and stock statistics are fairly accurate, at least for commercially traded energy forms such as fossil fuels. (Uncertainties for non-commercial energy use such as the traditional biomass fuels used in developing countries are however substantial (*see* Biomass Burning in Rural Homes in Tropical Areas, Volume 3).) The net balance between production, trade flows and stock changes allows one to determine annual apparent fossil

Table 1 Global carbon emissions as estimated for 1990 per major source category and uncertainty ranges (in Tg C year^{-1})[a]

		Net	Gross	Uncertainty range
	Coal	2424		
	Oil (fuels)	2285		
	Oil (feedstocks)		324	
	Gas	1135		
	Cement	157		
	Gas flaring	60		
Industrial		6061	6385	5800–7000
	Fuelwood[b]		530	
	Traditional biofuels[b]		630	
Biofuels[b]			1160	??–1600
	Savannah fires[c]		1660	
Other biomass				??–1700
	Tropical forests[d]		1100	
	Temperate forests[d]		0	
Land-use change[b]			1100	0–2800[e]
Total		6061	8645	5800–>13 100

[a] Emission categories that are not balanced by (uncertain) biospheric carbon sinks or that are not released to the atmosphere in the same year are listed as gross emissions, all others as net emissions (see text). For land-use change related emissions, the net biospheric flux as estimated by IPCC (1995) for the 1980s (the latest period for which global estimates are available) and the uncertainty range as estimated by Houghton (1999) for the same period are given. (Data source: see text.)
[b] Emissions of biofuels and land-use change not necessarily entirely additive.
[c] Andreae (1991). Not included in total gross emissions.
[d] Estimated net biospheric flows (IPCC, 1995).
[e] Upper range for net biospheric flux due to land-use change (Houghton, 1999), no estimates of gross emissions available.

fuel consumption, and that combined with carbon emission factors makes quantification of corresponding carbon emissions rather straightforward. Emission estimates are made regularly from the energy statistics of the United Nations (Marland *et al.*, 1999), the International Energy Agency (IEA, 1998), or the statistics published by British Petroleum (BP, 1999, see Table 1) drawing on similar methodologies (for the IPCC/IEA guidelines, see Houghton *et al.*, 1996). Aggregate results yield comparable global numbers, albeit differences (i.e., uncertainties) remain in the partition between various fuel types and between regions as a result of statistical differences and uncertainties in the heat content of different fuels, especially for coal. As an approximation, carbon emission factors are practically constant per unit energy content for different qualities of fuelwood, coal, oil, and gas, respectively, reflecting their molecular composition (ratio of hydrogen to carbon atoms). Characteristic emission factors (based on lower heating values) are 29.9 kg C GJ^{-1} for fuelwood (and other biomass fuels), 25.8 kg C GJ^{-1} for coal, 20 kg C GJ^{-1} for oil, and 15.3 kg C GJ^{-1} for gas (Nakicenovic *et al.*, 1996). Hence, uncertainties in heat values of different fuel qualities translate into proportional uncertainties of carbon emissions. Emission estimates for non-energy uses of fossil fuels (feedstocks) are affected by larger uncertainties, as the carbon contained in feedstock fuel (mostly oil) ends up in a wide variety of products (plastics, asphalt, bitumen, lubricants) which are in many instances long lived. Several decades may therefore pass before the carbon contained in these products is ultimately oxidized (e.g., when these products are burned) and released to the atmosphere. Hence, for any given year, it is necessary to distinguish between gross and net emissions. *Gross emissions* result from the transfer of carbon between various reservoirs due to human activity such as converting crude oil extracted from an oilfield where it has remained for geological ages to combustible fuel (e.g., gasoline) or long-lived materials (e.g., plastics, asphalt). *Net emissions* of any given year are those that result from human activities in that year, e.g., burning of gasoline in an internal combustion engine car. For fossil fuel, the differences between gross and net emissions are comparatively minor: some 0.3 Pg C, or 5%. Yet the difference is of similar importance as the secondary industrial sources i.e., cement manufacture and gas flaring.

Another source of uncertainty relates to biomass fuels. When produced at sustainable rates, sources and sinks will balance, albeit not necessarily in a given year. Whenever, biomass regrowth falls short of the rate of biomass burning, sources exceed sinks, giving rise to net emissions. Historically, this was the case for industrialized countries. Recently, biotic sinks have exceeded biotic emissions in the Northern Hemisphere, justifying the classification of biofuels as *carbon neutral*. The current situation in many developing countries is far removed from this carbon

benign state of affairs, resembling that in the industrialized countries some decades to a few centuries ago, characterized by deforestation and unsustainable biomass fuel use. The historian of pre-industrial Europe Carlo Cipolla (1981) characterizes medieval European forestry use as eminently parasitic and extremely wasteful . Biomass fuels are therefore included in the biotic carbon sources given in Table 1 as they usually are accounted for in detailed land-use change carbon inventories (Houghton, 1999). Note however, that the values given for gross emissions of biomass burning (e.g., fuelwood) and those from land-use changes are not necessarily additive because fuelwood gathering in connection with deforestation is usually also accounted for in the respective land-use change related carbon uxes. Thus, part (perhaps up to 0.5 Pg C) of the 1.2 Pg C gross emissions due to biofuel burning may already be included in the 1.1 Pg C gross emissions due to land-use changes reported in Table 1. This uncertainty and the risk of possible double counting has led to the doubtful practice of generally ignoring biofuel related carbon emissions in energy studies and emission inventories.

Differences between gross and net emissions as well as uncertainties are especially pronounced in the case of biotic carbon emissions. In that case, gross emissions, e.g., from biomass burning, need to be balanced with corresponding sinks, i.e., the uptake of carbon dioxide by vegetation in order to determine net emissions. In addition, both sources and sinks operate frequently on different time scales (burning a wood log takes an hour, growth of a tree many decades). To complicate things further, changes in vegetation cover also involve changes in soil carbon. The response rates of top and deep soil layers to a change in vegetation cover (e.g., conversion of forests to agricultural elds) are poorly understood and involve time scales of many decades (Houghton and Skole, 1990; Houghton, 1999).

Consider the burning of agricultural residues either in the eld or as energy for rural households: emissions from biomass grown on an annual rotational cycle are balanced by corresponding vegetation regrowth (carbon uptake) in the following or preceding growing season. A similar case is that of savannah res (natural or induced by human activities). Combined, these two categories involve the annual burning of some 5.7 billion tons dry matter per year and an estimated release of 1.6 Pg C (Andrae, 1991, see Table 1). As sources and sinks are assumed to balance over a time scale of approximately a year, these emission categories are generally excluded from carbon inventories (and are also not shown in the Figures below, not the least because historical estimates are unavailable. An exception is biomass fuel outside fuelwood, for which a conservative estimate is included in the data presented here (based on interpolated values given in UN, 1952; Putnam, 1954; Nakicenovic *et al.*, 1998. As simpli cation, the carbon emission factor for these other traditional biofuels like agricultural residues, dried animal dung, etc., is assumed to be the same as fuelwood).

For a complicated case consider the case of forest clearing: Felling of trees may entail short-term carbon emissions when wood is used as fuel, and a signi cantly delayed response when used as construction material (when a wood structure eventually burns down). In most cases, the dominant source of carbon emission is the decay of woody material after forest clearing (Houghton *et al.*, 2000). Carbon releases from soils depend critically on the ensuing land uses: secondary forest regrowth, or conversion to intensive agriculture. Soil carbon release rates depend therefore on (changing) land uses that follow deforestation and involve a time scale of many decades. Estimating corresponding carbon uxes is thus a formidable challenge.

Large uncertainties thus prevail for biospheric carbon sources and sinks. Assuming high biospheric (gross) emissions implies assuming a large biospheric sink (frequently assumed to be located in northern latitudes, cf. Schimmel *et al.*, 1995) in order to balance the global carbon cycle. Low biospheric emission uxes, especially when combined with upper range estimates of oceanic uptake rates, reduce the size of the residual (missing) sink required to balance the global carbon cycle. The average value of 1.1 Pg C year^{-1} net biospheric carbon emissions for the 1980s suggested in the IPCC review (Schimmel *et al.*, 1995) is derived from adopting a mean value of 1.6($\pm$1) Pg C from tropical latitudes balanced against an estimated uptake rate of 0.5($\pm$0.5) Pg C year^{-1} for northern latitudes. The estimated net biospheric ow of 1.1 Pg C represents a top down estimate, based on balancing the global carbon cycle as well as on carbon cycle model calculations. These models are frequently initialized with a 1980s global biospheric ux of 1.1 Pg C when reproducing the historical record of carbon dioxide concentration changes or when calculating future scenarios (for an overview see Houghton *et al.*, 1997; for a representative model see e.g., Wigley *et al.*, 1994).

Conversely, bottom-up assessments based on detailed inventory models, that yield changes in vegetation and soil carbon in response to land-use changes, typically yield higher estimates of net biospheric carbon ows: 2($\pm$0.8) Pg C year^{-1} for the 1980s, including 1.9($\pm$0.6) Pg C in the tropics and 0.1($\pm$0.5) Pg C in temperate latitudes (Houghton, 1999). Available (uncertain) forest inventory data indicate an average rate of tropical forest clearing of some 13.3 million ha year^{-1} for the 1980s, and an average 12.6 million ha year^{-1} for the period 1990 1995, compared to an average afforestation rate of 1.3 million ha year^{-1} in non-tropical regions for the period 1990 1995 (FAO, 1997; WRI, 1998). Translating these land-use change numbers into estimates of gross biospheric emissions and balancing those against sinks to determine estimates of net emissions

is highly problematic because of uncertainties in secondary forest regrowth, soil carbon, as well as the magnitude of the (undisturbed) forest carbon sink. Recent estimates for the Brazilian Amazon, characterized by high rates of forest clearing and regrowth (*see* **Deforestation and Habitat Fragmentation in the Amazon Basin**, Volume 3), conversion to agricultural uses followed by later abandonment, and wood logging, indicate nonetheless a roughly balanced carbon budget between emissions due to land-use changes and carbon uptake by vegetation regrowth and by natural ecosystems in the region (Houghton *et al.*, 2000). However, considerable variations exist between individual years: the Brazilian Amazon is estimated to have been either a (net) source or sink of about 0.2 Pg C for different individual years over the period 1989 1998.

Whereas bottom-up assessments thus continue to be the best way for estimating current and historical gross biotic carbon emissions, it is not possible yet to accurately determine net biospheric carbon emissions. This is why no estimates of respective current net emissions are given in Table 1. Also in the figures of historical emissions given below, respective biospheric carbon emissions are reported as gross values only.

From a combination of direct atmospheric measurements and analyses of air samples trapped in glacier ice, there is compelling evidence that atmospheric carbon dioxide concentrations have increased from some 280 ppmv (Neftel *et al.*, 1985; Schimmel *et al.*, 1995) in pre-industrial times to some 370 ppmv (estimate for AD 2000): i.e., an increase in the atmospheric carbon reservoir of close to 200 Pg C over the last 200 years. How can this increase in atmospheric concentration be allocated to different carbon emission categories?

Historical emission inventories for fossil fuel use and other industrial sources have a long scientific history starting from the pioneering contributions of Putnam (1954) and Keeling (1973). For land-use related emissions, the main historical data sources are the estimates of Houghton and Skole (1990) and Houghton (1999). A combined historical inventory of industrial and biotic carbon emissions was developed by Grubler and Nakicenovic (1994), an update of which is summarized in Figures 1 and 2. The original inventory spanned the period 1800 1988. Updates for fossil fuels, cement and gas flaring to 1998 are based on BP (1999) (and earlier vols.), CEDIGAZ (1998); and Marland *et al.* (1999). Values for the year 2000 are estimates based on extrapolation of short-term (1995 1998) trends. For biotic emissions, a comparable value as estimated by the IPCC (Schimmel *et al.*, 1995) for the 1980s, i.e., average annual net biotic emissions of 1 Pg C, was assumed for the 1990s, based on FAO (1998) forest inventory data that indicate similar average deforestation rates in the two time periods.

Figure 1 presents an overview of historical carbon emissions by source since 1800. Figure 2 adds these emissions categories, presenting them as cumulative totals. Only emissions from burning of fossil fuels and from the manufacture of cement can be classified clearly as net contributions to the atmosphere. As discussed above, the fraction of gross emissions that is not absorbed by sinks, or that may be released over long time periods (feedstocks) is highly uncertain. Hence these emission categories are shown as

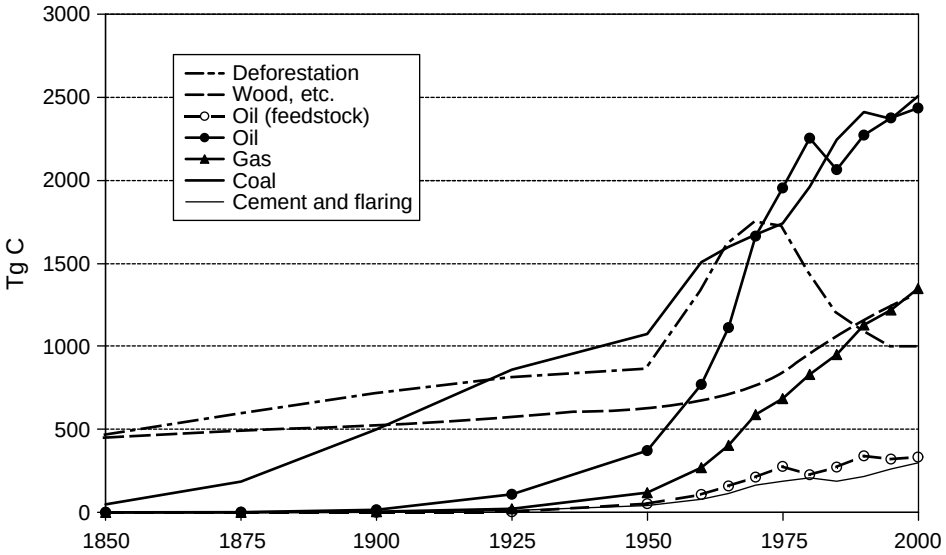


Figure 1 Annual carbon emissions per major source category 1850–2000 (in Tg C year^{-1}). Emission categories that are not balanced by (uncertain) biospheric carbon sinks or that are not released to the atmosphere in the same year are shown as gross emissions (dashed lines), all others as net emissions (solid lines). Values for the year 2000 are preliminary estimates. (Reproduced from Grübler and Nakicenovic, 1994 and Grübler, 1998a [Updated (see text)])

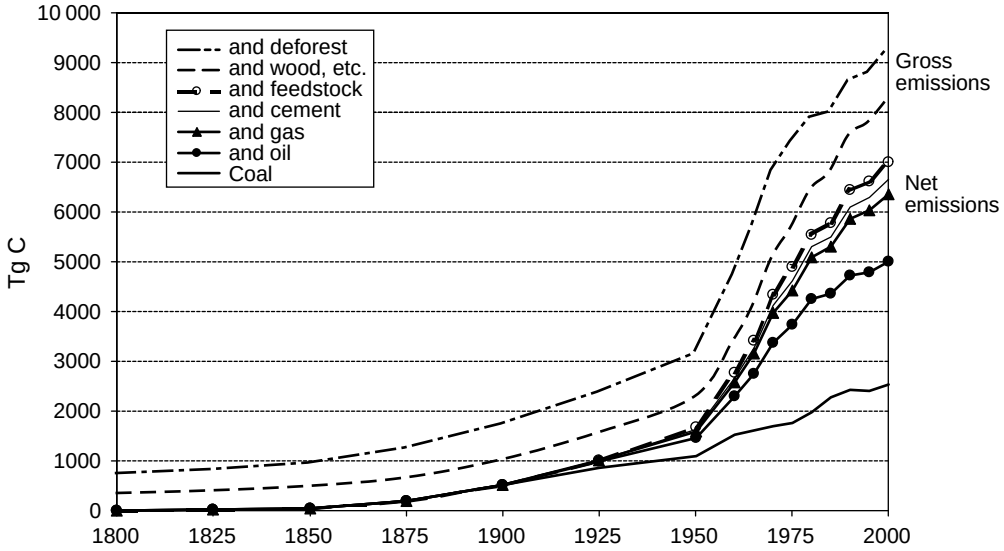


Figure 2 Annual carbon emissions per major source category 1800–2000, cumulative totals (in Tg C year^{-1}). (Reproduced from Figure 1)

gross emissions only in Figure 2. Cumulatively over the period 1800 2000, net carbon emissions total (a minimum) of some 280 Pg C, whereas gross carbon emissions total (a maximum) of 530 Pg C. In explaining the growth of atmospheric carbon, net emissions need to be balanced against the magnitude of the oceanic carbon sink, whereas gross emissions need to be balanced against the size of both the oceanic and the terrestrial carbon sinks.

In order of decreasing importance in terms of cumulative emissions since 1800 (derived from Figures 1 and 2), burning of coal and emissions due to land use changes are the two most important emission categories (148 Pg C net for coal and 155 Pg C gross for land-use changes). Houghton (1999) estimates for the period 1850 1990 that land-use changes (including fuelwood burning and logging) released a total of 373 Pg C to be balanced against an estimated terrestrial sink uptake of 249 Pg C (211 Pg C carbon uptake by vegetation regrowth and 38 Pg C soil carbon recovery). This yields an estimated net biospheric carbon flux of 124 Pg C, a comparable magnitude to the 155 Pg C gross carbon emissions due to land-use changes (excluding fuelwood burning) estimated here for the period 1800 2000. Oil ranks third (91 Pg C net emissions or 103 Pg C when non-energy feedstock uses (12 Pg C), classified here as gross emissions, are added). Burning of fuelwood and other (traditional) biomass fuels for energy purposes is estimated to have released some 77 Pg C (gross) emissions, albeit this quantity is not necessarily entirely additive to the emissions estimated from land-use changes. Burning natural gas as fuel is estimated to have released some 38 Pg C (net) since 1800 (most of it since 1950), whereas the manufacture of cement and the flaring (waste) of natural gas yields a comparatively small number of 5 Pg C (net) emissions since 1800.

Thus, the historic atmospheric increase of some 200 Pg C balances against a range of between 280 and 530 Pg C (net/gross) carbon emissions from all sources and an inferred cumulative uptake by all carbon sinks (oceanic and terrestrial) of between 80 and 330 Pg C over the period 1800 2000.

Most of the historical human-induced perturbations of the global carbon cycle originate from emissions in the industrialized countries. From the estimated cumulative net carbon emissions since 1800 (some 280 Pg C), about three-quarters originate from the industrialized countries, and one quarter from developing countries. For gross emissions, the difference is less pronounced: about 55% originate from the industrialized countries versus 45% from developing countries. Yet, such comparisons of historical emissions ignore important regional disparities in emission drivers such as population, levels of affluence, and levels and efficiency of technologies employed. When making these comparisons, it must be remembered that since 1800 70% of the world s people have lived in the developing countries (Grubler and Nakicenovic, 1994). Therefore when measured on a *per capita* basis the share of developing countries in historical or current emissions is much smaller than the simple numbers above would suggest (for detailed calculations see Fujii, 1990; Grubler and Fujii, 1991).

The division of annual gross and net emissions since 1800 for three macroregions is illustrated in Figure 3. The definition of these three macroregions follows that of the original UN Framework Convention on Climate Change (UNFCCC, 1992). It comprises: the countries of the OECD as of 1990 (Annex II in United Nations Framework Convention on Climate Change (UNFCC)); the countries of Central and Eastern Europe as well as of the

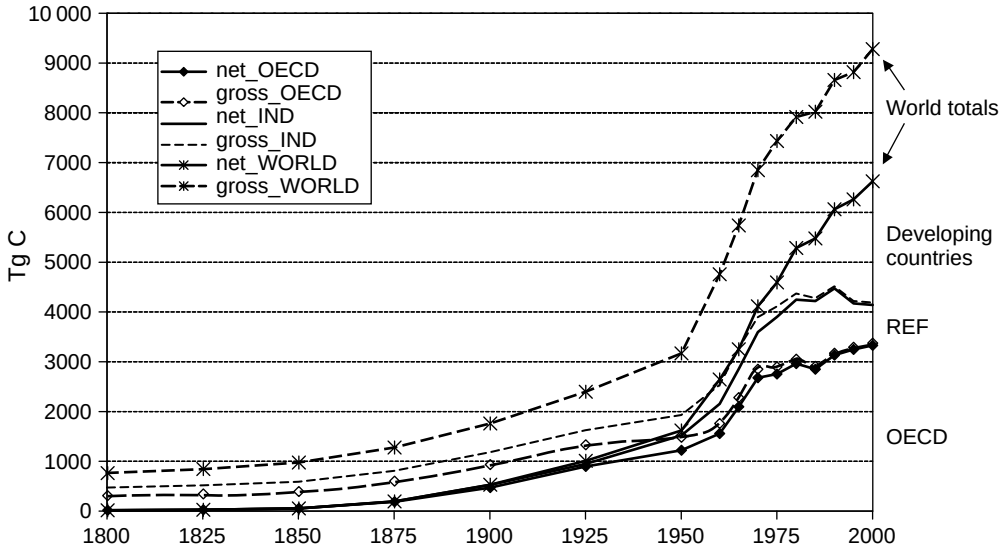


Figure 3 Annual carbon emissions per major emitting regions and given as cumulative totals (in Tg C year^{-1}). The regions are as defined in the 1990 UNFCC (*see* **United Nations Framework Convention on Climate Change and Kyoto Protocol**, Volume 4): OECD countries as of 1990 (OECD, UNFCCC Annex II countries), countries undergoing economic reform (REF, UNFCCC Annex I countries outside Annex II), and developing countries (DEV, UNFCCC non-Annex I countries). OECD and REF combined represent the industrialized countries referred to in the UNFCC as Annex I countries (IND in Table 3). Respective totals for gross and net carbon emissions (all sources) are shown as dashed and solid lines respectively (see text). (Reproduced from Grübler and Nakicenovic, 1994 [updated (see text)])

former Soviet Union undergoing economic reform (Annex I outside Annex II in UNFCC); as well as developing countries (non-Annex I countries in UNFCCC). Combined OECD and economic reform countries constitute the industrialized countries (IND in Figure 3) as defined by Annex I of the 1992 UNFCCC (*see* United Nations Framework Convention on Climate Change and Kyoto Protocol, Volume 4).

It is important to note that differences between gross and net emissions for different regions change over time. For instance, whereas this difference is substantial for the industrialized countries throughout the 19th century and also up to approximately 1950, the difference gets smaller, gradually disappearing after the 1970s. In other words, biotic carbon emissions such as land-use changes and fuelwood use (the main source for the difference between gross and net emissions shown in Figure 3), while being important in the early history of industrialization, are no longer important in the industrialized countries. There, the dominant emission source at present is the burning of fossil fuels. Currently, biotic carbon emissions in industrialized countries are close to zero. Historically, they account for about one quarter of cumulative carbon emissions of the industrialized countries since 1800. For developing countries, the situation resembles that of the industrialized countries more than a century ago (when the industrial countries had similar (low) levels of *per capita* incomes as the developing countries have today): an important share of emissions comes from biotic sources, with fossil fuel shares increasing in importance over time. Industrial emission sources account for around 40% of cumulative emissions of developing countries since 1800; by the year 2000 they already account for two thirds of carbon emissions from developing countries.

A brief discussion of how to relate the historical trends in emissions discussed above to main driving forces is given at the end of this contribution using energy-related carbon dioxide as example (*see* Carbon Dioxide Concentration and Climate Over Geological Times, Volume 1; Carbon Dioxide, Recent Atmospheric Trends, Volume 1; Carbon Cycle, Volume 2; Boreal Forest Carbon Flux and it s Role in the Implementation of the Kyoto Protocol Under a Warming Climate, Volume 4).

Concerning future emission trends, the (vast) literature on this subject cannot be reviewed in detail here. A comprehensive, quantitative overview of the scenario literature is given by Nakicenovic *et al.* (2000), who also present the latest long-term emissions scenarios developed under the auspices of the IPCC. In this scenario set, representative of current understanding of the underlying uncertainties in demographic, economic and technological developments, global carbon emissions could range anywhere between 8 and 27 Pg C by 2050 and between 3 and 37 Pg C by the end of this century. It is also instructive to look at projected levels of cumulative emissions, which are a (rough) first approximation of the possible magnitude of climate change. By 2100 cumulative carbon emissions could range between 770 and 2540 Pg C, or between equal, to more than three

times the amount currently contained in the atmosphere. In the absence of vigorous climate change policy initiatives (the new IPCC scenarios were developed under the mandate of explicitly excluding additional climate policies) even the lowest scenarios suggest that more carbon will be released to the atmosphere over the next 100 years (>770 Pg C) than has cumulatively been released since the onset of the Industrial Revolution (range of 230 530 Pg C net/gross emissions since 1800, respectively). A significant amplification of humankind s discernible influence on the climate system (IPCC, 1995) is thus likely.

SULFUR

Studies of the global sulfur cycle (good reviews are provided by Husar and Husar, 1990; and Smil, 1997) indicate that anthropogenic emissions of sulfur have surpassed natural flows ever since the first quarter of the 20th century. Dominant natural sources include the weathering of rocks and soil, volatile biogenic sulfur emissions from land and the oceans as well as volcanoes. Combined, natural sources are estimated to release between 40 and 60 Tg S (1 Tg S = one million (metric) tons elemental sulfur (MtS). To convert to SO_2, the customary unit of most sulfur studies, multiply by 2 (IPCC, 1995; Husar and Husar, 1990). This compares to total anthropogenic sulfur emissions of between 65 and 90 Tg S in the early 1990s (Benkovitz *et al.*, 1996; Olivier *et al.*, 1996; WMO, 1997). The dominant form of anthropogenic sulfur emissions is airborne emissions of SO_2 with some smaller quantities of SO_3 (hence sulfur emissions are frequently referred to as SO_x).

The ecological impacts of large anthropogenic sulfur emissions arise at three spatial and temporal scales. First, at the local level, high ambient concentrations of SO_2 have well documented (WHO and UNEP, 1993; WMO, 1997) impacts on human health, vegetation, and materials (corrosion, stone cancer of historical sandstone buildings). Second, sulfur emissions are one of the main contributors to acidic deposition (acid rain) that affects ecosystems up to a continental scale (changes in pH of streams and lakes and resulting decline in fish populations, reductions in the vitality of the forest ecosystem. Forest dieback, or Waldsterben however is increasingly recognized as a multiple stress phenomenon that cannot be related simply to a single source of environmental stress such as acidic precipitation) (*see* Waldsterben, Volume 5). Typically, impacts accrue in a highly non-linear fashion (dependent on acidic deposition levels and the buffering capacity of soils) which have led to the formulation of *critical loads* of acidic deposition (see e.g., Amann *et al.*, 1995; Posch *et al.*, 1996) (*see* Critical Load, Volume 3). Thirdly, sulfur emissions also assume global ecological significance; evidence is increasing that sulfate aerosols exert a pronounced cooling effect (counterbalancing greenhouse gas (GHG)-induced warming) in the Northern Hemisphere (IPCC, 1995).

A detailed review of available sulfur emission inventories is given in Grubler (1998b), where available global sulfur emission inventories are also compared for the year 1990. The best-guess value suggested in that review is 76 Tg S global anthropogenic sulfur emissions for 1990.

The dominant anthropogenic sources of sulfur emissions are (as for carbon emissions) the burning of fossil fuels. In the case of sulfur, emissions are dominated by burning of coal (some 53 Tg S in 1990, Lefohn *et al.*, 1999) and to a smaller degree by oil products (12 Tg S). Natural gas is (with a few exceptions that do not assume any global significance) almost sulfur free (small traces of sulfur compounds (H_2S) are added deliberately as a safety measure; a smelly tracer helps to detect gas leakages) and hence not a significant source. The third largest emissions category is metallurgical processes (reduction of sulfide ores during the smelting of copper, lead, and zinc), estimated at some 6 Tg S in 1990 (Lefohn *et al.*, 1999). Smaller additional emissions sources are biomass burning (some 2 Tg S, cf. Pepper *et al.*, 1992), as well as marine bunker fuels (usually high sulfur-containing heavy fuel oil which is generally not allowed to be burned on land) with some 3 Tg S in 1990 (4 Tg S in 1994, cf. Corbett *et al.*, 1999). Finally, sulfur emissions from pulp and paper mills, although globally marginal, may have important local ecological effects (apart from their offensive smells). For the sake of completeness, it also should be mentioned that the elemental sulfur mined (some 25 Tg S, Smil, 1997) as feedstock for the chemical industry (mostly used in the production of sulfuric acid) generally does not result in airborne emissions. This market however, is declining with the increasing availability of elemental sulfur recovered in oil refineries as a byproduct of fuel desulfurization to meet environmental standards (likewise gypsum mining faces increasing competition from gypsum produced from flue gas desulfurization units of coal-fired power stations).

Sulfur emission inventories are developed on a regular basis in a number of regions, including within the EMEP and CORINAIR programs in Europe and the National Acid Precipitation Assessment Program (NAPAP) in North America. Recently more detailed emission inventories have become available for Asia, where emission growth rates are particularly high (Streets *et al.*, 2000), including work in conjunction with the World Bank sulfur project (Foell *et al.*, 1995), and the detailed bottom up estimates of Akimoto and Narita (1994) and Kato (1996). Sulfur emissions in all regions not mentioned above (i.e., Pacific OECD countries like Australia and New Zealand, and all developing countries outside Asia) are much less well studied. Even if presently small when compared to those of Europe, North America and Asia, the sulfur emissions in these regions are likely to grow significantly in the longer-term.

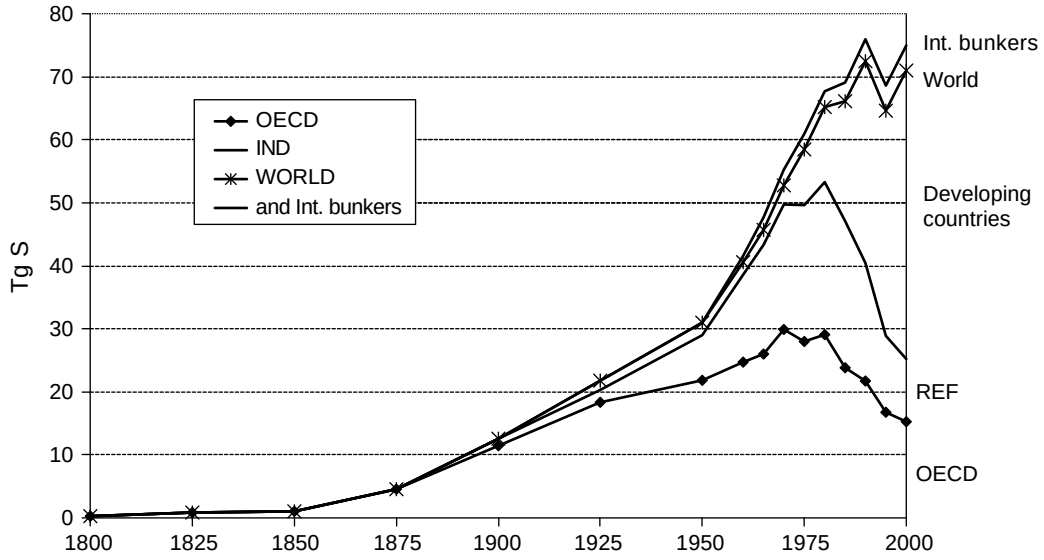


Figure 4 Annual sulfur emissions per major emitting region as defined in Figure 3, 1800–2000, shown as cumulative totals (in Tg S year^{-1}). Emissions from international marine bunker fuels that are usually not accounted for in national/regional emission inventories are shown separately. (Data sources: see text)

At the global level, a number of spatially detailed sulfur emission inventories have been developed. Global, gridded (1×1 degrees) sulfur emissions inventory data are needed for a variety of purposes, most notably for climate modeling purposes. As a rule, the required cycle times for the development of such detailed inventories are rather long, implying that data can be quite outdated. With the exception of the EDGAR data base (Olivier *et al.*, 1996) referring to 1990, global gridded data sets available are quite outdated in view of rapidly changing regional sulfur emission trends. For instance, the Spiro *et al.* (1992) inventory refers to the year 1980; the Global Emissions Inventory Activity (GEIA) (Benkovitz *et al.*, 1996) gridded sulfur emission data are an update of the Spiro *et al.*, data set for 1985 for a number of regions, most of them OECD countries, but retain 1980 values for many countries and regions.

Contrary to carbon emissions, global and regional sulfur emission trends are much more dynamic, particularly since the early 1970s. As a result of sulfur abatement efforts, emissions in the OECD countries, particularly in Europe and Japan have declined drastically. Emissions in Central and Eastern Europe as well as the successor states of the former Soviet Union have also declined, especially in the wake of their economic restructuring. Conversely, with accelerated economic development, the growth of sulfur emissions in many part of Asia has been fast, albeit growth rates have declined recently (Streets *et al.*, 2000). Thus sulfur emission trends must be analyzed regularly and with as recent data as possible.

Historical sulfur emission inventories were first developed for Europe (cf. the formidable historical work of Mylona, 1993, 1996) as well as for the US (Gschwandner *et al.*, 1985; EPA, 1995). Global estimates have been developed by Dignon and Hameed (1989); Stern and Kaufmann (1996) and Lefohn *et al.* (1999). A synthesis of these long-term sulfur emission trends is given in Figure 4.

The long-term pattern of global sulfur emissions is characterized by three phases: First until about the 1920s, sulfur emissions rose very rapidly (at an estimated average annual growth rate of some 4% year^{-1}) with the expanding use of coal, the fossil fuel richest in sulfur. In the period 1925 to ca. 1975, growth in global emissions continued, albeit at slower rates (some 2% year^{-1} on average), as coal increasingly was replaced by oil, generally lower in sulfur content. Since around 1975, global emissions have stayed roughly constant. The continued rise of sulfur emissions in developing countries has been compensated by the drastic declines in emissions in the OECD countries (first Japan, then Europe and North America) as a result of environmental policies leading to fuel substitution, fuel desulfurization as well as stack gas cleaning (scrubbing). Emissions in Central and Eastern Europe, as well as in the successor states of the former Soviet Union, have also declined drastically as a result of replacement of coal by other fuels, as well as the drastic economic recession since the early 1990s (the effects of environmental control measures like flue gas desulfurization have been more limited to date). The industrialized countries accounted with some 50 Tg S for well over 80% of global sulfur emissions in 1975. By 2000, their sulfur emissions were down to some 25 Tg S (33% of global emissions), whereas those of the developing countries (mostly in Asia) have risen to some 50 Tg S (around 67% of global emissions).

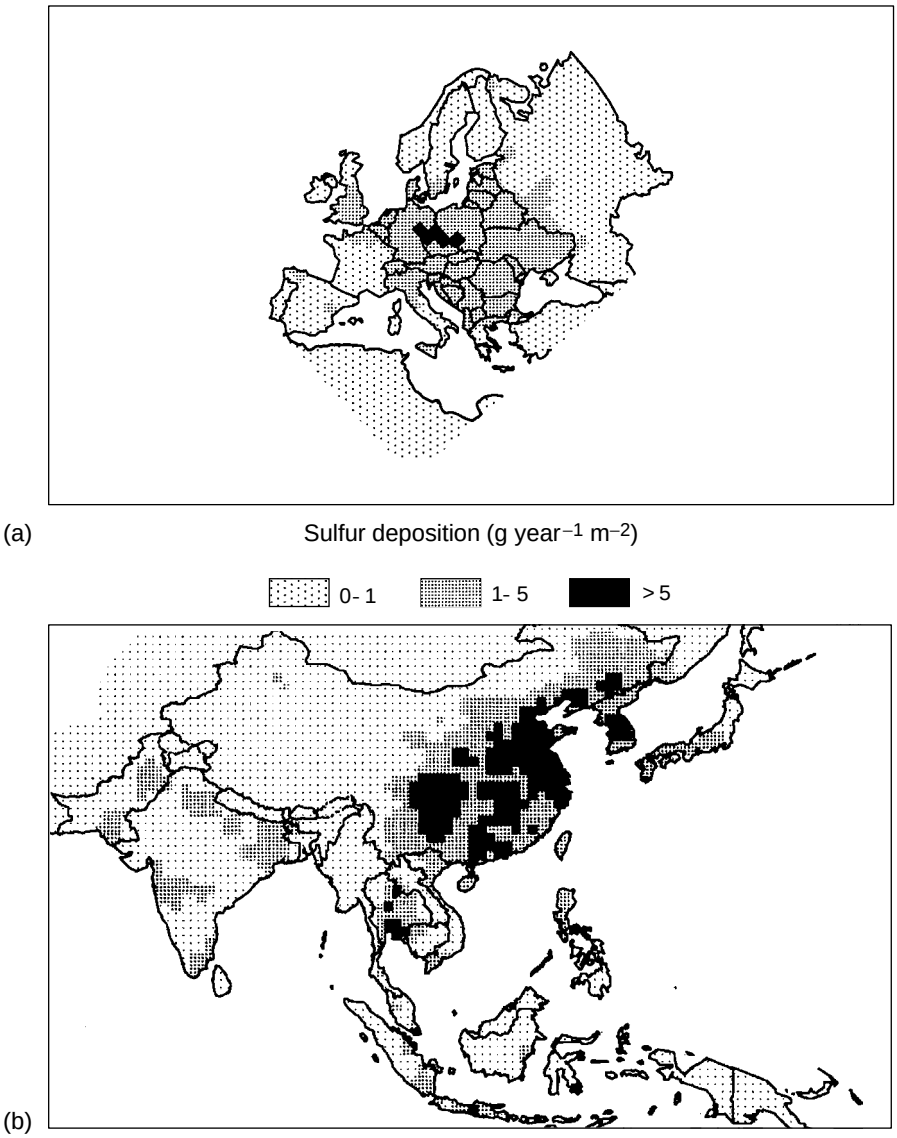


Figure 5 Extent of likely peak sulfur deposition levels: 1990 sulfur deposition in Europe (a) and projection for an (unabated) high-growth scenario for Asia in 2020 (b), in grams sulfur (S) m^{-2}. (Reproduced from Grübler, 1998a based on Amann *et al.*, 1995)

Recent estimates for Asia (Streets *et al.*, 2000) indicate however, that also there, emission growth rates are declining as a result of the introduction of sulfur control legislation (Grubler, 1998b; IEA, 1999; Nakicenovic *et al.*, 2000; Streets *et al.*, 2000). An acceleration of these trends seems highly desirable in view of the projected large-scale impacts on human health, food production, as well as ecosystems of unabated growth in sulfur emissions in the densely populated, coal intensive economies in Asia. A representative result of such projected scenarios (based on Amann *et al.*, 1995) is shown in Figure 5, which contrasts 1990 European sulfur deposition levels with those of Asia by 2050 in a high (unabated) sulfur emission scenario. Typically, in such scenarios, sulfur emissions in Asia alone could surpass current global levels as early as 2020 (Amann *et al.*, 1995; Posch *et al.*, 1996). Sulfur deposition above $5\,g\,m^{-2}\,year^{-1}$ occurred in Europe in 1990 in the area of the borders of the Czech Republic, Poland, and Germany (the former GDR); often referred to as the black triangle . In view of its ecological impacts, it was officially designated by the United Nations Environment Programme (UNEP) as an ecological disaster zone . In the scenario of high sulfur emission growth in Asia illustrated in Figure 5, similar high sulfur deposition would occur by around 2020 over more than half of Eastern China, large parts of southern Korea, and some smaller parts of Thailand and southern Japan.

In order to avoid excessive damages, sulfur controls similar to those used previously in the OECD countries

are required, and indeed are already beginning to be implemented in a number of Asian countries. Doubtless, similar challenges will have to be faced by other developing countries outside Asia over the next two to three decades. The recent long-term scenario literature (reviewed in more detail in Grubler, 1998b; and Nakicenovic *et al.*, 2000) indeed reflects such developments. Characteristic projected future global sulfur emission levels range between 20 and 80 Tg S by 2050 and between 15 and 60 Tg S by 2100. The comparatively low levels of projected future sulfur emissions reflects the assumed phase in of sulfur controls also outside the industrialized countries.

Sulfur reduction policies aiming to protect local and regional populations and ecosystems have one interesting side effect: they lower the cooling effect caused by sulfate aerosols in the atmosphere, thus increasing the warming arising from future greenhouse gas emissions (Rogner and Nakicenovic, 1996; Subak *et al.*, 1997) (*see* Aerosols, Effects on the Climate, Volume 1).

NITROGEN

Emissions of nitrogen compounds take many forms and originate from a wide array of different sources, natural and anthropogenic. Because of this diversity, individual source categories are studied in much less detail and therefore large uncertainties prevail.

Nitrogen emissions occur in three principal molecular forms: as nitrogen oxides (NO and NO_2 that are subsumed generally as NO_x), principally formed in high temperature combustion (burning of fossil fuels, or natural lightning); as ammonia (NH_3), principally arising from animal manure; and as nitrous oxide (N_2O), a powerful greenhouse gas, principally arising from soil microbiological process as well as agricultural activities and animal manure. Both NO_x and NH_3 are characterized by short atmospheric residence times and thus by only regional to continental dispersion from their respective emission sources. Their ecological significance arises principally in connection with acidification impacts (although NO_x has also a significant influence on atmospheric chemistry, including ozone (O_3) destruction, and therefore is indirectly affecting GHG concentrations in the atmosphere). Nitrous oxide is a powerful, long-lived (120 years) greenhouse gas whose atmospheric concentrations are estimated to have been 275 parts per billion by volume (ppbv) in pre-industrial times and have been measured at above 310 pbbv in the mid 1990s (IPCC, 1995). For ease of comparability, all emissions are expressed here in elemental weight of nitrogen per year ($Tg\,N\,year^{-1}$).

NO_x has long been of concern in the context of acidic deposition. In the last decade, however, the question of nitrogen over-fertilization of the biosphere has also been raised. See, for example, Munn *et al.* (1999, 468). Possible long-term consequences include the creation of mineral deficiencies in forest soils due to leaching, and a general decrease in biodiversity, particularly in grasslands. Indeed, some ecologists are concerned that nitrogen overfertilization constitutes a long-term threat to the whole boreal forest system.

Until now, most research has focussed on emissions of nitrogen oxides (NO_x, for a review see Smil, 1990), especially for energy-related emissions source categories since motorized traffic is one of the main sources of NO_x emissions and a major contributor to urban smog. Conversely other source categories, including natural processes have been much less studied and large uncertainties prevail. The first nitrogen inventory covering all three gases and all source categories at a regionally highly desegregated level has been developed by Olivier *et al.* (1998), based on earlier studies of Olivier *et al.* (1996), Lee *et al.* (1997); and Bouwman *et al.* (1997), and the discussion here draws heavily on this study. A summary of main emission categories is given in Table 2.

Nitrogen emissions are dominated by anthropogenic sources: 77(37 113) Tg N compared to 42(16 80) Tg N natural sources. In terms of their contributions to total nitrogen emissions, NO_x and NH_3 dominate with 50(22 81) and 54(23 88) Tg N each, compared to 15(8 24) Tg N in the case of nitrous oxide. For anthropogenic emissions, ammonia (43(20 61) Tg N) surpasses NO_x (31(16 46) Tg N), whereas nitrogen-wise, nitrous oxide is a comparatively small source of nitrogen (3(1 6)) (IPCC, 1995 indicates a slightly higher uncertainty range of 4 8 Tg N for nitrous oxide), cf. Olivier *et al.* (1998).

For emissions of nitrogen oxides (NO_x), the dominant anthropogenic source category is the burning of fossil fuels, most notably of automobile fuel and in the generation of electricity. NO_x emission levels depend both on nitrogen content in fuels as well as on very diverse and variable operating and firing characteristics: NO_x emissions tend to increase with increasing burning temperature. In this case, tradeoffs in improving environmental performance become apparent: improving efficiency of energy conversion and lowering of carbon emissions, e.g., in an electric power plant, requires an increase in firing temperatures (the second law of thermodynamics), which tends to increase NO_x emissions. Reducing the latter requires catalytic reduction, which in turn involves an efficiency penalty, slightly increasing fuel use and hence carbon emissions. For cars de-NO_x catalytic equipment also depends critically on operating temperature: at cold engine startup catalytic converters do not function. Therefore related emission estimates of this category of emissions are highly uncertain. Because of the dominance of industrialized countries in global car ownership and electricity consumption, they account for over 75% of global NO_x emissions from road transport and for 70%

Table 2 Global nitrogen emissions as estimates for 1990 per major source category and flux (NO_x, NH_3, and N_2O) and uncertainty ranges (in Tg N year^{-1})

	NO_x (1990)	+/−	NH_3 (1990)	+/−	N_2O (1990)	+/−
Anthropogenic						
Fossil fuel burning	21.9	13–31	0.1	0–0.2	0.2	0.1–0.5
Industrial processes	1.5		0.2	0.1–0.3	0.3	0.1–0.5
Agriculture:						
Animals			21.6	10–30	1	0–2
Fertilizers			9	4.5–13.5	1	0.3–2.3
Crops and waste			4.1	1.4–5	0.1	
Biomass burning	7.7	3–15	5.4	3–7.7	0.6	0.4–1
Sewage			2.6	1.3–3.9		
Total	31.1	16–46	43	20–61	3.2	0.9–6.3
Natural						
Soils	5.5	4–12	2.4	0–10	5.2	2.6–7.8
Grasslands, wild animals, etc.			0.1	0–1	2.3	1.1–3.5
Oceans			8.2	3–16	3.6	2.8–5.7
Lightning	12.2	2–20				
Atmospheric processes	1.6	0.4–2.6			0.6	0.3–1.2
Total	19.3		10.7	3–27	11.7	6.8–18.2
Grand total	50.4	22–81	53.7	23–88	14.9	7.7–24.5

Source: Olivier *et al.* (1998).

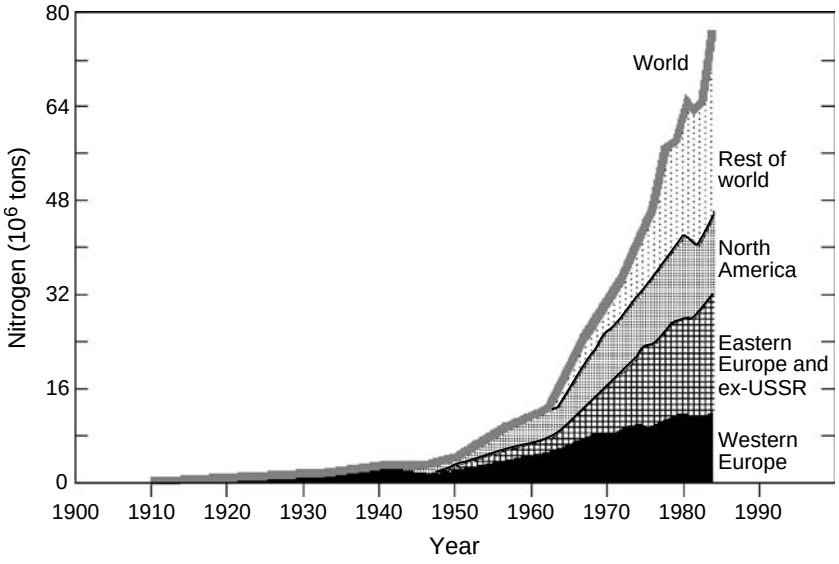


Figure 6 World nitrogen fertilizer use by region (cumulative totals, in Tg N year^{-1}). (Reproduced from Grübler, 1998a)

of all NO_x emissions from fossil fuel burning (Olivier *et al.*, 1998). Anthropogenic NO_x emissions are also the only ones where estimates of long-term historical time series are available. For instance, the estimates of Dignon and Hameed (1989) indicate that global NO_x emissions have increased by almost a factor five over the period 1930 1980.

For ammonia (NH_3), the dominant source of anthropogenic emissions is the agricultural sector, including emissions from animal manure (22(10 30) Tg N) and application of fertilizer (9(4 14) Tg N), that combined account for over 70% of NH_3 emissions globally. The dominance of agricultural sources also explains that contrary to the case of NO_x ammonia emissions predominantly originate in developing countries (a detailed discussion of NO_x, NH_3 and N_2O emissions in India is given by Parashar *et al.*, 1998), which accounted for approximately 70% of agricultural and total ammonia emissions in 1990. Rising animal populations, as well as rising fertilizer use and resulting ammonia (and nitrous oxide, see discussion below) emissions are the other side of the coin of vastly increased agricultural production for ever rising populations (Figure 6). The increase to over 80 Tg N year^{-1} would have been impossible without

the major technological breakthrough of the Haber-Bosch method of ammonia synthesis introduced in 1912.

For emissions of nitrous oxides, the dominant anthropogenic sources are again fertilizer applications and animal husbandry (manure) in the agricultural sector (2(0.3 4.3) Tg N year^{-1} in 1990). The third largest source are industrial processes (production of adipic and nitric acid) with 0.3(0.1 0.5) Tg N estimated for 1990 (Olivier *et al.*, 1998), albeit these emissions are declining rapidly due to implementation of voluntary industry measures due to climate change concerns (Nakicenovic *et al.*, 2000). Yet, even with drastic emission reductions, a long-term legacy of nitrous oxide emissions remains due to its long residence time in the atmosphere. As in the case of ammonia emissions, developing countries, with their high population and food requirements, dominate global emissions, producing about 60% of agricultural and of total nitrous oxide emissions (Olivier *et al.*, 1998). Even though emissions from natural processes, particularly microbiological processes in soils, greatly exceed anthropogenic emissions (3 8 Tg N versus 1 6 Tg N respectively), emissions of nitrous oxide are a good illustration of the impacts of a growing world population and of changing diets (towards higher meat consumption) on NO_x emissions. These in turn require increased food output and changes in agricultural practices and technologies that result in increased emissions, even if the long-term history of ammonia and nitrous oxide emissions remains to be researched.

The literature on long-term nitrogen emissions scenarios is sparse. Due to the multitude of gas species and source categories, no scenarios exist that treat nitrogen emissions comprehensively. The most detailed projections available currently are those developed under the auspices of the IPCC Special Report on Emissions Scenarios (Nakicenovic *et al.*, 2000), but even there, projections for ammonia emissions are lacking as the study focussed on (direct and indirect) greenhouse gases. Nonetheless, the study re ects current best knowledge on uncertainties in future nitrogen emission levels. By 2050, global nitrogen emissions could span a range of 45 110 MtN (N_2O and NO_x) and by 2100 a range of 20 170 MtN (compared to 37.6 MtN excluding NH_3) in 1990 as estimated in Nakicenovic *et al.* (2000). Thus, over the long-term, nitrogen emissions could range between half of current to about four times current levels indicating the large uncertainties in underlying driving forces such as population and economic growth, future developments in the agricultural, energy, and transport sectors as well as in technology.

Because of the multitude of gases and sources, nitrogen emission estimates are poorly understood and uncertain with the paucity in both past and present inventory data and future emission scenarios. More research is therefore needed before considering policy interventions to control these gases beyond well established source categories (e.g., nitrogen oxide emissions from automobiles, or nitrous oxide emissions from adipic acid production). *See* **Nitrogen Cycle**, Volume 2; **Nitrogen Deposition on Forests**, Volume 2.

LINKING EMISSIONS TO DRIVING FORCES: THE CASE OF ENERGY-RELATED CARBON DIOXIDE

Emissions are the result of a large set of interrelated driving variables in the domains of demographics, economics, resources and technology as well as (environmental) policies. Economic, social and technical systems and their interactions are highly complex and this short discussion cannot provide a comprehensive overview.

Instead, the links from demography and the economy to resource use and emissions will be discussed in a simpli ed format using proxy variables for various categories of driving forces. A frequently used analytical approach to describe these linkages is through the so called IPAT identity (Equation 1):

$$\text{Impact} = \text{population} \times \text{af uence} \times \text{technology} \quad (1)$$

where environmental impacts (emissions) are the product of the level of population times the af uence (income per capita) times the level of technology deployed (emissions per unit of income) (*see* **DPSIR (Driving Forces Pressures State Impacts Responses)**, Volume 4). The IPAT identity has been widely used in the analysis of energy related carbon dioxide emissions (for a review see Alcamo *et al.*, 1995; Gaf n, 1998; and Nakicenovic *et al.*, 2000) where it is generally referred to as the KAYA identity (Kaya, 1990) (Equation 2):

$$\begin{aligned}\text{Carbon dioxide emissions} &= \text{population} \\ &\times (\text{GDP/population}) \times (\text{energy/GDP}) \\ &\times (\text{carbon dioxide/energy}) \end{aligned} \quad (2)$$

In this extended IPAT identity, emissions are represented as the product of population times their level of af uence (gross domestic product (GDP)/capita), multiplied with two proxy variables for the level of technology, i.e., the ef ciency of energy use per unit of GDP (Energy/GDP) as well as the carbon intensity of energy used (carbon dioxide/Energy). We classify these latter two variables as proxy variables for technology as they represent highly aggregated indicators that in addition also depend on many other factors, all related to income levels. These include the structure of the economy (e.g., industry versus service orientation), consumption patterns and the extent and nature of environmental policies. Energy price differences also matter, but are generally not related to differences in income levels. Frequently, the poor while unable to afford (high

priced) commercial energy forms such as electricity, rely on traditional fuels gathered largely outside the formal economy. Their costs however can be very high when considering the amount of time spent gathering fuelwood and environmental externalities such as high levels of indoor air pollution associated with burning of traditional fuels in open fire places (Smith, 1993). In turn, with rising incomes, consumer preferences shift to higher priced commercial energy forms (electricity, gas). Cleanliness and convenience of use afford a quality premium (willingness to pay higher prices), including higher energy efficiency and lower emissions (particulates, sulfur, carbon dioxide) per unit of energy. Thus, at the point of final energy use the energy and carbon intensity of the KAYA identity are related to income levels (affluence). This may not necessarily be visible in aggregated data, as the provision of clean energy forms at the point of consumption can entail efficiency losses and proportionately higher emissions at the point of energy production, e.g., when electricity is generated in traditional coal fired power stations.

This multiplicative identity has the advantage that percentage growth rates of its components are additive, thus giving a first approximation of the scale of influence on emissions growth of different driving variables. However, despite the appeal of a simple representation, there are important caveats to bear in mind when using the above identity in interpreting past emissions trends (or future emission scenarios). The first is the issue of spatial heterogeneity. The second one is that it is not possible to treat the growth of individual components independently.

For example, on face value, the KAYA and IPAT identities would suggest that carbon dioxide emissions scale linearly with population; that is, doubling the population means doubling the emissions. However, there are reasons to doubt such a simplified view. There is great heterogeneity among populations with respect to greenhouse gas emissions, so where the population increase takes place is not necessarily the same as where emissions increase. The ratio of *per capita* emissions of the world s richest countries to that of its poorest countries approaches ratios of several hundred (Gaffin, 1998). This problem requires one to consider spatially desegregated population growth. Of course, some level of aggregation is necessary for a global analysis. The biggest correction on emissions takes place when switching from a global to a breakdown of industrialized (IND) versus developing (DEV) countries (Lutz, 1993) (Table 3).

The second caveat relates to the fact that it is not population *per se* that emits carbon dioxide, but the technologies that people employ and the goods and services they consume with their available income. This raises the issue about the relationship between demographics and economic growth (between population and affluence in the IPAT identity) as well as between technology and affluence, which in turn would preclude consideration of a simple linear interpretation of the role of population growth in emissions.

Demographic development interacts in many ways with social and economic development. Fertility and mortality trends depend, amongst other things, on education, income,

Table 3 Major components (population, per capita income, energy and carbon intensities) for the increase of energy-related (gross) carbon emissions (Kaya Identity) and their average annual growth rates, 1800–2000, global (WOR) and regional totals[a]

		1800	*%/yr*	1900	*%/yr*	1950	*%/yr*	1975	*%/yr*	2000	**1800–2000** **%/yr**
POP	IND	246	*0.82*	555	*0.81*	833	*1.10*	1096	*0.77*	1327	**0.85**
million	DEV	737	*0.39*	1083	*0.88*	1681	*2.21*	2901	*2.00*	4764	**0.94**
	WOR	983	*0.51*	1639	*0.86*	2514	*1.87*	3996	*1.70*	6091	**0.92**
GDP/POP	IND	675	*1.26*	2352	*1.39*	4686	*3.35*	10 674	*1.68*	16 188	**1.60**
$/capita	DEV	500	*0.21*	619	*0.69*	874	*2.73*	1715	*2.28*	3013	**0.90**
	WOR	544	*0.80*	1206	*1.15*	2137	*2.71*	4171	*1.39*	5883	**1.20**
ENE/GDP	IND	0.743	*−0.40*	0.497	*−0.20*	0.448	*−0.33*	0.413	*−1.49*	0.284	**−0.48**
kgoe/$	DEV	0.450	*−0.17*	0.379	*−0.01*	0.377	*−0.54*	0.329	*−0.48*	0.292	**−0.22**
	WOR	0.541	*−0.17*	0.457	*−0.13*	0.429	*−0.40*	0.388	*−1.20*	0.287	**−0.32**
C/ENE	IND	1.238	*−0.12*	1.103	*−0.33*	0.937	*−0.64*	0.798	*−0.66*	0.676	**−0.30**
kg C/kgoe	DEV	1.252	*0.00*	1.249	*−0.23*	1.112	*−0.36*	1.017	*−0.40*	0.921	**−0.15**
	WOR	1.246	*−0.09*	1.144	*−0.31*	0.979	*−0.55*	0.853	*−0.38*	0.776	**−0.24**
Carbon	IND	152	*1.56*	715	*1.67*	1638	*3.48*	3851	*0.27*	4123	**1.66**
Tg C	DEV	208	*0.43*	318	*1.33*	616	*4.05*	1664	*3.42*	3858	**1.47**
	WOR	360	*1.06*	1033	*1.57*	2254	*3.64*	5516	*1.49*	7980	**1.56**

[a] The two regions are defined in the UNFCCC, including Annex-I countries (IND, for industrialized countries) and non-Annex-I countries (DEV, for developing countries). For units, see table. Estimates of energy use, energy intensities, and (gross) carbon emissions include non-commercial, traditional biofuels which are particularly uncertain.
Source: Updated (see text) from Grübler and Nakicenovic (1994) and Nakicenovic and Grübler (2000).

social norms and health provisions. In turn, they determine the size and age composition of the population. All factors combined are recognized as important in explaining long-run productivity and economic growth and technological change as well (Barro, 1997). In turn, long-run per capita economic growth is closely linked with advances in knowledge and technological change. In fact, analysis of long-run macroeconomic growth accounts (e.g., Solow, 1956; Denison, 1962, 1985; Maddison, 1995; Barro and Sala-I-Martin, 1995) confirm that advances in knowledge and technology form the single largest source for long-run economic growth, more important than growth in other factors of production like capital and labor. Abramovitz (1993) demonstrates that capital and labor productivity cannot be treated as independent from technological change. In terms of the IPAT identity, it is thus not possible to treat the affluence and technology variables as independent from each other either. In addition, pollution abatement efforts appear to increase with income, a relationship often referred to as the environmental Kuznets curve that seems well established for traditional pollutants such as particulates and sulfur (e.g., World Bank, 1992) but which has not been demonstrated for aggregate greenhouse gas emissions.

Technological, economic and social innovations have long been means by which a greater number of people could live from the same environmental resources. The best known historical examples of major periods of innovation include the Neolithic revolution (beginnings of organized agriculture from around 10 000 years ago); and the Industrial Revolution beginning two centuries ago (Rosenberg and Birsdzell, 1996). In each case, changes in patterns of primary production (food, energy, materials), the efficiency of resources use, as well as changes in environmental impacts have been linked to changes in social organization, institutions, economy and technology (Mumford, 1934; Landes, 1969; Mokyr, 1990). No one of these changes can be considered to be the primary driver, nor as independent from the others: each played a role in an interconnected system.

Before the Neolithic revolution, the world population amounted to perhaps 5 10 million. Over the following millennia complex societies developed based around cities, with highly differentiated social roles and complex political, economic and legal systems. By the birth of the current major religions 1500 2500 years ago, world population had increased to around 300 million. Following a period of cultural, economic and technical stagnation to about 1000 AD with notable exceptions in China and the Islamic world (Mokyr, 1990), further population growth began with the productivity increases resulting from cultural, economic and technological developments associated with the European Renaissance. By the time of Thomas Malthus in the late 18th century, the global population had reached about one billion. The six-fold increase in global population over the following 200 years has been enabled by a continuing revolution in medical, agricultural and industrial technology. However, the most remarkable change in recent decades has been the so-called *demographic transition*, as people increasingly manage their own rate of reproduction, leading to a decline in population growth rates (see Table 3) or to a stabilization of population size in many parts of the world (*see* Population Sizes, Changes, Volume 2; Demographic Change: the Aging Population, Volume 3; Global Population Trends, Volume 3; Urban Population Change, Volume 3; Demographic transition, Volume 5).

This transition has typically been linked to improved education, female empowerment, access to information, improved medical services (and related reductions in infant mortality) and social and economic development facilitated by the international diffusion of new technologies. Therefore, it is not possible to treat changes in demographics, income, and technology as independent from each other.

With the above caveats in mind, let us consider the quantitative evidence since the onset of the Industrial Revolution (ca. 1800), cast in terms of the KAYA identity.

Table 3 summarizes the growth of population, *per capita* income, energy use per unit economic activity and energy related (gross) carbon emissions per unit energy used since 1800. The estimated indicators represent a revised updated data set based on Grubler (1998a); Grubler and Nakicenovic (1994) and Nakicenovic and Grubler (2000). Data with a finer temporal and spatial resolution are available from the author upon request. Because the energy use data include the historically important non-commercial uses of traditional biofuels, the resulting energy use and energy intensities (energy use per unit of GDP) estimates are subject to a high degree of uncertainty; the values given here represent a conservative lower bound estimate and are subject to further research and revision. Total energy use in the 19th century industrialized countries and in developing countries prior to 1975 (see the discussion in Nakicenovic *et al.*, 1998) are thus likely to be underestimated. Resulting historical improvements in energy intensities (including all energy forms) may consequently also be significantly underestimated. For instance the aggregate improvements in energy intensity of 0.5% $year^{-1}$ since 1800 for the industrialized countries (including Eastern Europe and Russia) given in Table 3, need to be contrasted with estimates for individual countries with good historical records that indicate long-term improvement rates of about 1% $year^{-1}$ since the 19th century (Nakicenovic, 1984; Martin, 1988; Grubler, 1991) and much higher rates since the early 1970s (reflected in Table 3). Non-commercial energy use and resulting energy intensity improvement rates for developing countries prior to 1975 are particularly

uncertain, possibly underestimated by up to a factor of three.

In order to simplify, only two macro-regions (industrialized versus developing countries) and trends between five major dates are shown. Interspersed between the main historical dates, the respective annual growth rates of these indicators calculated over their respective time periods are shown in italics. Total period (1800 2000) growth rates are presented in bold (most far right column in Table 3). As mentioned above, the relative magnitude of the growth rates and how they sum up to explain the growth rates in energy related carbon dioxide emissions are indicative only as they do not capture the interdependence among variables; notably between technology and economic growth on the one hand and demographics on the other.

A first observation on Table 3 is that there is no uniform, unambiguous answer about what (proxy) driving force dominates in explaining historical emissions growth. Component growth rates are highly variable across regions and over time. On the positive (i.e., emission increasing) side of the KAYA identity are population growth and increases in *per capita* incomes, on the negative (emission decreasing) side of the identity are improvements in energy intensity (energy use per unit of GDP) as well as decreases in the ratio of (gross) carbon emissions per unit energy used, a trend termed frequently as *decarbonization* (Grubler, 1991; Nakicenovic, 1993).

Demographic and economic development tend to increase emissions, a trend partly compensated by improvements in the two proxy variables representing technology. Globally, an average increase of 1.6% $year^{-1}$ in (gross) energy related carbon emissions since 1800 can be decomposed into roughly equal contributions from population and *per capita* income growth (average 2.1% $year^{-1}$) to be contrasted against an about equal contribution from improvements in technology (energy intensity improvements and decarbonization) of jointly $-0.5\%\,year^{-1}$. However, these numbers vary enormously depending on which time period is being considered, as well as across regions.

As discussed above, the component variables of the KAYA identity are not independent from each other. For instance, the influence of technological change goes far beyond the numbers captured in the percentage rates of change of the energy intensity and decarbonization variables presented in Table 3; this is because technological change is also the main source for productivity growth and hence increases in *per capita* incomes. This illustrates the paradox of technological development (Gray, 1989), where technology is both the source as well as the (partial) remedy of the historical increases in environmental burdens. Prometheus unbound (Landes, 1969) has released the powers of technology that helped humans to expand in numbers at ever higher levels of affluence, thus increasingly burdening the global commons. But technological change has also helped to reduce (part of) these environmental burdens by improved efficiency in the use of energy and the carbon atom. More research is needed to answer the question, if and how fast the paradox of technological development may be resolved in the future, tilting the balance of the equation towards stabilizing emissions, even their long-term decline. The challenge becomes compounded by the inevitable demographic momentum (the mothers of the future are already born today) that makes in increase of global population to some 8 billion almost a certainty (Lutz *et al.*, 1997). There is also a need for continued social and economic development for the larger part of the global population still excluded from the benefits of the use of modern technology.

REFERENCES

Abramovitz, M (1993) The Search for the Sources of Growth: Areas of Ignorance, Old and New, *J. Econ. Hist.*, 52(2), 217 243.

Akimoto, H and Narita, H (1994) Distribution of SO_2, NO_x, and CO_2 Emissions From Fuel Combustion and Industrial Activities in Asia with 1×1 Degree Resolution, *Atmos. Environ.*, 28(2), 213 225.

Alcamo, A, Bouwman, A, Edmonds, J, Grubler, A, Morita, T, and Sugandhy, A (1995) An Evaluation of the IPCC IS92 Emissions Scenarios, in *Climate Change 1994*, ed J T Houghton, Cambridge University Press, Cambridge, 232 304.

Amann, M, Cofala, J, Doerfner, P, Gyarfas, F, and Schoepp, W (1995) *Impacts of Energy Scenarios on Regional Acidi cation*, Report, WEC Project 4 on Environment, Working Group C Local and Regional Energy Related Environmental Issues, World Energy Council, London.

Andrae, M O (1991) Biomass Burning: its History, Use, and Distribution and its Impacts on Environmental Quality and Global Climate, in *Global Biomass Burning*, ed J S Levine, MIT Press, Cambridge, MA, 3 21.

Barro, R J (1997) *Determinants of Economic Growth*, The MIT Press, Cambridge, MA.

Barro, R J and Sala-I-Martin, X (1995) *Economic Growth*, McGraw-Hill, NY.

Benkovitz, C M, Scholtz, M T, Pacyna, J, Tarrason, L, Dignon, J, Voldner, E C, Spiro, P A, Logan, J A, and Graedel, T E (1996) Global Gridded Inventories of Anthropogenic Emissions of Sulfur and Nitrogen, *J. Geophys. Res.*, 101(D22), 29 239 29 253.

Bouwman, A F, Lee, D S, Asman, W A H, Dentener, F J, Van der Hoek, K W, and Olivier, J G J (1997) A Global High Resolution Emission Inventory for ammonia, *Global Biogeochem. Cycles*, 11(4), 561 587.

BP (British Petroleum AMOCO) (1999) *BP AMOCO Statistical Review of World Energy 1999*, BP, London, see also http://www.bpamoco.com/worldenergy/download/index.htm.

CEDIGAZ (Centre International d Information sur le Gaz Naturel et tous Hydrocarbures Gazeux) (1998) *Natural Gas in the World 1998 Survey*, CEDIGAZ, Rueil Malmaison.

Corbett, J J, Fischbeck, P, and Pandis, S N (1999) Global Nitrogen and Sulfur Inventories for Oceangoing Ships, *J. Geophys. Res.*, **104**(D3), 3457 3470.

Denison, E F (1962) *The Sources of Economic Growth in the United States and the Alternatives before Us*, Supplementary Paper No. 13, Committee For Economic Development, New York.

Denison, E F (1985) *Trends in American Economic Growth, 1929 1982*, The Brookings Institution, Washington, DC.

Dignon, J and Hameed, S (1989) Global Emissions of Nitrogen and Sulfur Oxides from 1860 to 1980, *J. Air Waste Manage. Assoc.*, **39**(2), 180 186.

EPA (Environmental Protection Agency) (1995) National Air Pollutant Emission Trends 1900 1994, EPA 454/R 95 011, EPA, Washington, DC, see also http://www.epa.gov/oar/emtrnd94/emtrnd94.html.

FAO (Food and Agriculture Organization of the United Nations) (1997) *State of the World s Forests 1997*, Forest Resources Division, FAO, Rome.

Foell, W, Amann, M, Carmichael, G, Chadwick, M, Hettelingh, J-P, Hordijk, L, and Dianwu, Z (1995) *Rains Asia: an Assessment Model for Air Pollution in Asia*, Report on the World Bank sponsored project Acid Rain and Emission Reductions in Asia , World Bank, Washington, DC.

Fujii, Y (1990) *An Assessment for the Responsibility for the Increase in CO_2 Concentration and Intergenerational Carbon Accounts*, WP 90 55, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Gaf n, S R (1998) World population projections for greenhouse gas emissions scenarios, *Mitigation Adapt. Strat. Global Change*, **3**(2 4), 133 170.

Gray, P E (1989) The Paradox of Technological Development, in *Technology and Environment*, eds J H Ausubel and H E Sladovitch, National Academy Press, Washington, DC, 192 204.

Grubler, A (1991) Energy in the 21st Century: From Resource to Environmental and Lifestyle Constraints, *Entropie*, **164/165**, 29 33.

Grubler, A (1998a) *Technology and Global Change*, Cambridge University Press, Cambridge.

Grubler, A (1998b) A Review of Global and Regional Sulfur Emission Scenarios, *Mitigation Adapt. Strat. Global Change*, **3**(2 4), 383 418.

Grubler, A and Fujii, Y (1991) Inter-generational and Spatial Equity Issues of Carbon Accounts, *Energy Int. J.*, **16**(11 12), 1397 1416.

Grubler, A and Nakicenovic, N (1994) *International Burden Sharing in Greenhouse Gas Reduction*, RR 94 9, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Gschwandtner, G, Gschwandtner, K C, and Eldridge, K (1985) *Historic Emissions of Sulfur and Nitrogen Oxides in the United States from 1900 to 1980*, Vol. I and II, EPA 600/785 009a and 009b, EPA, Washington, DC.

Harvey, D L D (1999) *Global Warming, the Hard Science*, Prentice Hall, Harlow.

Houghton, J T *et al.* (1996) *IPCC Guidelines for National Greenhouse Gas Inventories*, Vols. I III, revised 1996 edition, OECD/IEA, Paris.

Houghton, J T, Meira Filho, L G, Griggs, D J, and Maskell, K (1997) *An Introduction to Simple Climate Models used in the IPCC Second Assessment Report*, Technical Paper II, IPCC, Geneva.

Houghton, R A (1999) The Annual Net Flux of Carbon to the Atmosphere from Changes in Land Use 1850 1990, *Tellus*, **51**(B), 298 313.

Houghton, R A and Skole, D L (1990) Carbon, in *The Earth as Transformed by Human Action*, eds B L Turner, W C Clark, R W Kates, J F Richards, J T Mathews, and W B Meyer, Cambridge University Press, Cambridge, 393 408.

Houghton, R A, Skole, D L, Nobre, C A, Hackler, J L, Lawrence, K T, and Chomentowski, W H (2000) Annual Fluxes of Carbon from Deforestation and Regrowth in the Brazilian Amazon, *Nature*, **430**, 301 304.

Husar, R B and Husar, J D (1990) Sulfur, in *The Earth as Transformed by Human Action*, eds B L Turner, W C Clark, R W Kates, J F Richards, J T Mathews, and W B Meyer, Cambridge University Press, Cambridge, 409 421.

IEA (International Energy Agency) (1998) *CO_2 Emissions from Fuel Combustion*, IEA, Paris, see also, http://www.iea.org/stats/ les/keystats/stats_98.htm.

IEA (International Energy Agency) (1999) *Non-OECD Coal-fired Power Generation Trends in the 1990s*, IEA Coal Research, London.

IPCC (Intergovernmental Panel on Climate Change) (1995) Radiative Forcing of Climate Change, in *Climate Change 1994*, eds J T Houghton *et al.*, Cambridge University Press, Cambridge, 1 231.

Kato, N (1996) Analysis of Structure of Energy Consumption and Dynamics of Emission of Atmospheric Species Related to the Global Environmental Change (SO_x, NO_x, and CO_2) in Asia, *Atmos. Environ.*, **30**(5), 757 785.

Kaya, Y (1990) *Impact of Carbon Dioxide Emission Control on GNP Growth: Interpretation of Proposed Scenarios*, Paper presented to the IPCC Energy and Industry Subgroup, Response Strategies Working Group, Paris (mimeo).

Keeling, C D (1973) Industrial Production of Carbon Dioxide from Fossil Fuels and Limestone, *Tellus*, **25**(2), 174 197.

Keeling, C D and Whorf, T P (1999) Atmospheric CO_2 Records from Sites in the SIO air Sampling Network, in *Trends: a Compendium of Data on Global Change*, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, see also http://cdiac.esd.ornl.gov/trends/co2/sio-mlo.htm.

Landes, D S (1969) *The Unbound Prometheus: Technological Change and Industrial Development in Western Europe from 1750 to the Present*, Cambridge University Press, Cambridge.

Lee, D S, Kohler, I, Grobler, E, Rohrer, E, Sausen, F, Gallardo-Klenner, L, Olivier, J G J, Dentener, F J, and Bouwman, A F (1997) Estimates of Global NO_x Emissions and their Uncertainties, *Atmos. Environ.*, **31**, 1735 1749.

Lefohn, A S, Husar, J D, and Husar, R B (1999) Estimating Historical Anthropogenic Global Sulfur Emission Patterns for the Period 1850 1990, *Atmos. Environ.*, **33**(21), 3435 3444, see also http://ourworld.compuserve.com/homepages/ASL_ASSOCIATES/sulfur.htm.

Lutz, W (1993) Population and Environment: What do we Need More Urgently: Better Data, Better Models, or Better Questions? in *Environment and Population Change*, eds

B Zaba and J Clarke, Liege, Derouaux Ordina Editions, 47 62.

Lutz, W, Sanderson, W, and Scherbov, S (1997) Doubling of World Population Unlikely, *Nature*, 387, 803 805.

Maddison, A (1995) *Monitoring the World Economy 1820 1992*, OECD Development Center Studies, Organization for Economic Co-operation and Development, Paris.

Marland, G, Boden, T A, Andres, R J, Brenkert, A L, and Johnston, C (1999) Global, Regional, and National CO_2 Emissions, in *Trends: a Compendium of Data on Global Change*, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, see also http://cdiac.esd.ornl.gov/trends/emis/tre_glob.htm.

Martin, J-M (1988) L intensité énergétique de l activité economique dans les pays industrialisés: Les evolutions de très longue periode liverent-elles des enseignements utiles? *Economies et Societés*, 4, 9 27.

Mokyr, J (1990) *The Lever of Riches: Technological Creativity and Economic Progress*, Oxford University Press, Oxford.

Mumford, L (1934) *Techniques and Civilization*, Harcourt, New York.

Munn, T, Whyte, A, and Timmerman, P (1999) Emerging environmental issues, *AMBIO*, 28, 464 471.

Mylona, S (1993) *Trends in Sulfur Dioxide Emissions, Air Concentrations and Deposition of Sulfur in Europe Since 1880*, EMEP/MSC W Report 2/1993, Norwegian Meteorological Institute, Oslo.

Mylona, S (1996) Sulfur Dioxide Emissions in Europe 1880 1991 and their Effect on Sulfur Concentrations and Depositions, *Tellus*, 48(B), 662 689.

Nakicenovic, N (1984) *Growth to Limits: Long Waves and the Dynamics of Technology*, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Nakicenovic, N (1993) Long-term Strategies for Mitigating Global Warming, *Energy*, 8, 401 609.

Nakicenovic, N, Grubler, A, Ishitani, H, Johansson, T, Marland, G, Moreira, J R, and Rogner, H-H (1996). Energy Primer, in *Climate Change 1995, Impacts, Adaptations and Mitigation of Climate Change: Scienti c-Technical Analyses*, eds R T Watson, M C Zinyowera, and R H Moss, Cambridge University Press, Cambridge, 75 92.

Nakicenovic, N, Grubler, A, and McDonald, A (1998) *Global Energy Perspectives*, Cambridge University Press, Cambridge.

Nakicenovic, N and Grubler, A (2000) Energy and the Protection of the Atmosphere, *Int. J. Global Energy Issues*, 13(1 3), 4 57.

Nakicenovic, N *et al.* (2000) *Emissions Scenarios*, Report of Working Group III of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge.

Neftel, A, Moor, E, Oeschger, H, and Stauffer, B (1985) Evidence from Polar Ice Cores for the Increase in Atmospheric CO_2 in the Past Two Centuries, *Nature*, 315, 45 47.

Olivier, J G J, Bouwman, A F, van der Maas, C W M, Berdowski, J J M, Veldt, C, Bloos, J P J, Visschedijk, A J H, Zanfeld, P Y J, and Haverlag, J L (1996) *Description of EDGAR Version 2.0: A Set of Global Emission Inventories of Greenhouse Gas Gases and Ozone-depleting Substances for all Anthropogenic and most Natural Sources on a per Country Basis and on a 1 × 1 grid*, RIVM Report No. 771060 002, RIVM, Bilthoven, the Netherlands.

Olivier, J G J, Bouwman, A F, Van der Hoek, K W, and Berdowski, J J M (1998) Global Air Emission Inventories for Anthropogenic Sources of NO_x, NH_3 and N_2O in 1990, *Environ. Pollut.*, 102(S1), 135 148.

Parashar, D C, Kulshrestra, U C, and Sharma, C (1998) Anthropogenic Emissions of NO_x, NH_3 and N_2O in India, *Nutr. Cycling Agroecosyst.*, 52, 255 259.

Pepper, W, Leggett, J, Swart, R, Wasson, J, Edmonds, J, and Mintzer, I (1992) *Emission Scenarios for the IPCC: an Update, Assumptions, Methodology, and Results*, IPCC, Geneva.

Posch, M, Hettelingh, J-P, Alcamo, J, and Krol, M (1996) Integrated Scenarios of Acidification and Climate Change in Asia and Europe, *Global Environ. Change*, 6(4), 375 394.

Putnam, P C (1954) *Energy in the Future*, Macmillan Press, London.

Rogner, H H and Nakicenovic, N (1996) The Role of Sulfur in Climate Change, *Energiewirtschaftliche Tagesfragen*, 46(111), 731 735.

Rosenberg, N and Birdzell, L E (1986) *How the West Grew Rich: the Economic Transformation of the Industrial World*, I B Tauris, London.

Schimmel, D, Enting, I G, Heimann, M, Wigley, T M L, Raynaud, D, Alves, D, and Siegenthaler, U (1995) CO_2 and the Carbon Cycle, in *Climate Change 1994, Part 1: Radiative Forcing of Climate Change*, eds J Y Houghton *et al.*, Cambridge University Press, Cambridge, 35 71.

Smil, V (1990) Nitrogen and Phosphorous, in *The Earth as Transformed by Human Action*, eds B L Turner *et al.*, Cambridge University Press, Cambridge, 423 436.

Smil, V (1997) *Cycles of Life: Civilization and the Biosphere*, Scientific American Library, W H Freeman, New York.

Smith, K R (1993) Fuel Combustion, Air Pollution Exposure, and Health: The Situation in Developing Countries, *Annu. Rev. Energy Environ.*, 30(10), 16 35.

Solow, R (1956) A Contribution to the Theory of Economic Growth, *Q. J. Econ.*, 70, 56 94.

Spiro, P A, Jacob, D J, and Logan, J A (1992) Global Inventory of Sulfur Emissions with 1 × 1 Degree Resolution, *J. Geophys. Res.*, 97, 6023 6036.

Stern, D I and Kaufmann, R K (1996) *Estimates of Global Anthropogenic Sulfate Emissions 1860 1993*, WP 9602, Center for Energy and Environmental Studies, Boston University, Boston, MA.

Streets, D G, Tsai, N Y, Akimoto, H, and Oka, K (2000) Sulfur Dioxide Emissions in Asia in the Period 1985 1997, *Atmos. Environ.*, 34(26), 4413 4424.

Subak, S, Hulme, M, and Bohn, L (1997) *The Implications of FCCC Protocol Proposals for Future Global Temperature: Results Considering Alternative Sulfur Forcing*, CSERGE Working Paper GEC 97 19, CSERGE University of East Anglia, Norwitch.

UNFCCC (United Nations Framework Convention on Climate Change) (1992) *Convention Text*, UNEP/WMO Information Unit on Climate Change, WMO, Geneva.

UN (United Nations) (1952) *World Energy Supplies in Selected Years 1929 1950*, Department of Economic Affairs, Statistical Office of the United Nations, New York.

Wigley, T M L, Solomon, M, and Raper, S C B (1994) *Model for the Assessment of Greenhouse-gas Induced Climate Change. Version 1.2*, Climate Research Unit, University of East Anglia.

WHO (World Health Organization) and UNEP (United Nations Environment Program) (1993) *Urban Air Pollution in Megacities of the World*, 2nd edition, Blackwell Publishers, Oxford.

WMO (World Meteorological Organization) (Whelpdale, D M and Kaiser, M S) (1997) *Global Acid Deposition Assessment*, WMO Global Atmospheric Watch No. 106, WMO, Geneva.

WRI (World Resources Institute) (1998) *World Resources 1998- 1999*, WRI, Washington, DC.

World Bank (1992) *World Development Report 1992: Environment and Development*, Oxford University Press, Oxford.

Anthropogenic Metabolism and Environmental Legacies

PAUL H BRUNNER AND HELMUT RECHBERGER

Vienna University of Technology, Vienna, Austria

For prehistoric Man, the total metabolism (input, output and stock of materials and energy to satisfy all human needs for provisions, housing, transportation, etc.,) was mainly determined by the physiological need for food, for air to breathe, and for shelter. The material turnover of modern Man is 10- 20 times larger. The fraction that is used for food and breathing is small. The so-called anthropogenic metabolism covers not only the physiological metabolism but includes also the thousands of goods and substances necessary to sustain modern life. "Anthropogenic" stands for man-made. The Anthroposphere is the sphere in which human activities take place, sometimes called technosphere or biosphere. Today, the most important man-made material flows are due to activities such as cleaning, transporting, residing and communicating. These activities were of little metabolic significance in prehistoric times.

Even more astonishing than the increase in the material flows is the tremendous material stock, which mankind has been accumulating, and which continues to grow faster than ever. This stock is found everywhere. It begins with mining residues left behind as tailings. It goes on with the material stocks of industry, trade and agriculture, and culminates in the urban stock of private households and of the infrastructure for transport, communication, business and administration. It ends with the comparatively small but growing material wastes in landfills.

The difference in stock accumulation between prehistoric and modern times is striking. Present-day material stocks amount to about 3- 400 tons per urban citizen. We have to manage and maintain this stock. Today's societies have to make far-reaching decisions about the constant renewal of the urban stocks, such as buildings, roads, communication systems, etc. The residence time of materials in the stock ranges up to 100 years. This means that once a material is consolidated into the stock, it will probably not show up quickly in the output of the stock, namely in waste management.

The presence of this large and growing stock has many implications:

1. *as an important reservoir of valuable resources, it holds a tremendous potential for future recycling;*
2. *as a mostly unknown source of materials, the importance of which is not yet in focus, which awaits assessment with respect to its significance as a resource and as a threat to the environment;*
3. *as a long-term source of severe pollutant flows to the environment. An assessment has been made that urban stocks contain more hazardous materials than so-called hazardous waste landfills, which are a focus of environmental protection;*
4. *as a challenge for future planners and engineers to design new urban systems. In the future, the location and amount of materials in city stocks should be known. Materials should be incorporated into the stock in a way which allows easy reuse and environmental control;*
5. *as an economic challenge to maintain high growth rates, building up even larger stocks, and setting aside sufficient resources to maintain this stock properly over long periods of time.*
6. *as a challenge to simulation modelers, who must deal with the complexities of the many processes contributing to urban metabolism, including the influence of long-term global, regional and local environmental, socioeconomic and cultural changes (*see ***Contaminated Lands and Sediments: Chemical Time Bombs?****, Volume 3).*

METABOLISM OF THE ANTHROPOSPHERE

Phenomenology: Growing Stocks and Flows Surpassing those of Nature

The following phenomena are typical for the metabolism of today s affluent societies:

- rates of material consumption and stock accumulation are high and growing;
- for many materials, man-induced material flows exceed natural flows;
- consumption emissions are surpassing production emissions;
- the metabolism of cities is mainly linear (throughput economy), thus cities are heavily dependent on their hinterland as a source for raw materials, as a partner for trade of manufactured goods, and as a sink for dissipation.

High Growth Rates of Material Flows and Stocks in Affluent Societies

As presented in Figure 1 and Table 1, the household consumption of goods has increased from prehistoric to modern times by more than an order of magnitude. The term *good* is used for products with a positive or negative economic value. A good is made from substances, which are defined as chemical elements or compounds consisting of uniform units such as atoms (element) or molecules (compound). The term materials stands for both goods and substances. This growth is even larger if all of the materials used in mining, agriculture, forestry, manufacturing, distribution and consumption are included. The main target of these large material flows is the consumer. With his/her

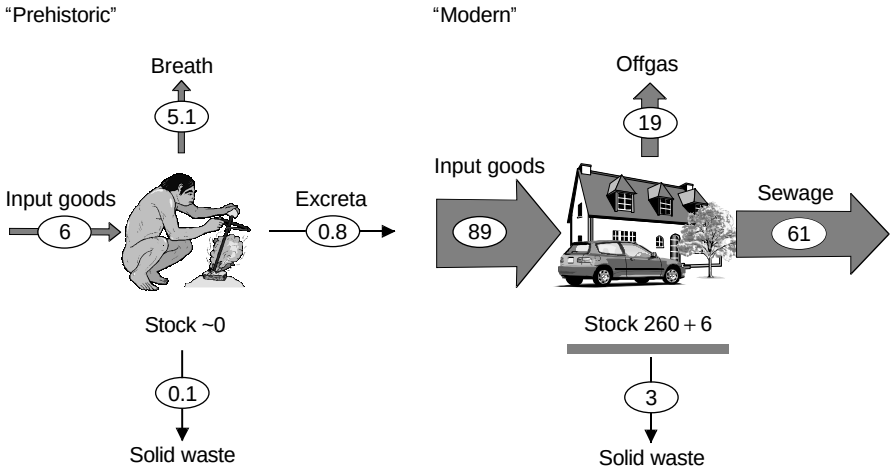


Figure 1 Material turnover of prehistoric and modern Man, in tons per capita per year. Included are all materials used in private households to satisfy the needs for food, shelter, hygiene, transportation, communication, etc., (for details, see Baccini and Brunner, 1991)

Table 1 Comparison of material flows and stocks for selected activities for prehistoric and modern Man[a]. T/c.y stands for tonnes per capita per year

	Material turnover t/c.y		Material stock t/c		Annual change in material stock t/c.y	
Activity	Prehistoric	Modern	Prehistoric	Modern	Prehistoric	Modern
To clean	<0.1	60	0	0.1	0	<0.1
To communicate	0	10	0	160	0	3
To reside	<0.1	10	<0.1	100	0	3
To breathe	4	4	0	0	0	0
To nourish	1.6	1.7	0	<0.1	0	<0.1
Total	6	86	<0.1	260	~0	+6

[a] The most outstanding and unprecedented features of today's economies are the very large stock of materials that have accumulated in the anthroposphere. ''To clean'' comprises water, chemicals and equipment needed for laundry, dishwashing, personal hygiene, toilet etc.; ''to communicate'' takes account of the transport of persons, goods, energy and information due to the needs of private households (including materials for road construction); ''to reside'' consists of the buildings, furniture, appliances etc., needed for living; ''to nourish'' includes all flows of goods associated with the consumption of food within private households, e.g., food, water for cooking, etc.

high purchasing power, he/she is able to pull vigorously at the end of the material chain, buying products larger quantities of consumer goods and energy (Schmidt-Bleek *et al.*, 1996). The terms rucksack and gray characterize the invisible energy, materials and wastes associated with the production of a good during primary production and manufacturing of the constituents and the product itself.

The lifestyle of modern urban Man demands a large amount of materials and energy to sustain his/her metabolism. The reasons are manifold. Economic conditions in the affluent part of the world allow large consumption flows. Resources such as energy and minerals are still abundant and cheap. The early warnings about the limits to growth due to diminishing resources (Meadows *et al.*, 1972) proved to be (numerically, but not in principle!) wrong. Advanced technology enables the development of new products in shorter cycles, making many appliances obsolete before the actual end of their life. Consumption seems to be a value *per se*; many modern consumer products transfer social prestige from the product to the owner. Although there may be small signs of consumer fatigue or even opposition here and there, in general there is no realistic alternative to this growth in material consumption in sight. It remains to be shown that the decoupling of economic progress with the growth of material consumption is economically feasible. It is obvious that developing countries, which still consume at a much lower rate, wish to reach the same material turnover rates as in the most affluent societies.

Compared to the flows of goods, the increase in the flows of many substances, such as lead, is much more dramatic. It has been estimated (Settle and Patterson, 1980) that the mining of lead rose during the last 7000 years from about 1 ton $year^{-1}$ to more than 3 million tons in 1990 (Figure 2). The situation is similar for other metals, and for natural and synthetic organic substances. Due to technological and economic development, there are constant shifts in the use of materials. This is illustrated by the changes in the amounts of materials used for construction in the US (Figure 3). Wood is no longer the key construction material, concrete dominates, and plastic materials are becoming increasingly more important. Metals and organic additives are also growing in importance. In waste management, materials that were used for construction 25 100 years ago, have to be treated now (recycled and disposed of). The nature of these materials will change rapidly in coming decades. There will be more mixtures of more materials reaching waste management sites. The fraction of hazardous waste materials in urban wastes will increase because of the growing consumption of chemicals in the 1970s and 1980s used to stabilize components of manufactured goods. Even if these substances are not used anymore, there exist large reservoirs of banned substances such as chlorofluorocarbons (CFCs) and cadmium (Cd) in the urban stock. As a consequence, waste management must also be capable of handling materials that have not been in use for decades. Prevention is a strategy that has only a slow effect when applied to materials incorporated in the urban stocks with long residence times.

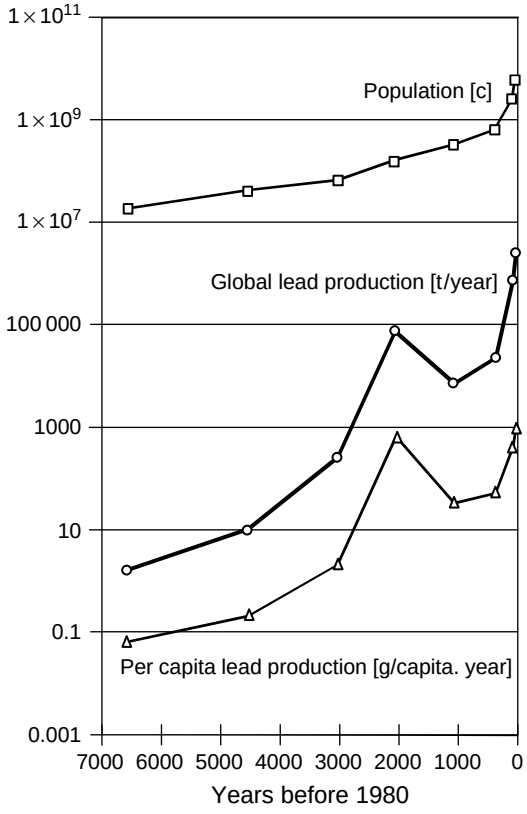


Figure 2 Global production of lead: due to unprecedented progress in mining technology and economic development, the per capita mining of lead increased from 0.1 g $capita^{-1}$ $year^{-1}$ to 1000 g $capita^{-1}$ $year^{-1}$. In combination with the growth of the global population (which increased at a smaller rate than the per capita production), the total amount of lead produced from mining increased from 1 ton $year^{-1}$ to more than 3 million tons $year^{-1}$ (lead data from Settle and Patterson, 1980)

Goods with long residence times are typically incorporated into the infrastructure, such as housing, industry and manufacturing, traffic and communication systems, and so on. A large, and at present mostly unknown mass of substances is stored in this stock. Figure 2 serves as an example. There is little information about today s location of approximately 100 million tons of lead exploited over the last 7000 years. Part of the lead stock is still used as water piping or insulation for telecommunication cables. Another fraction has been dispersed via exhaust gases from cars fueled by leaded gasoline; it remains in the soil and in sediments, and a small fraction has shown up in the Greenland ice caps and other remote areas. Lead has also been land filled with industrial wastes and with municipal solid wastes (MSWs). For the city of Vienna, it has been roughly

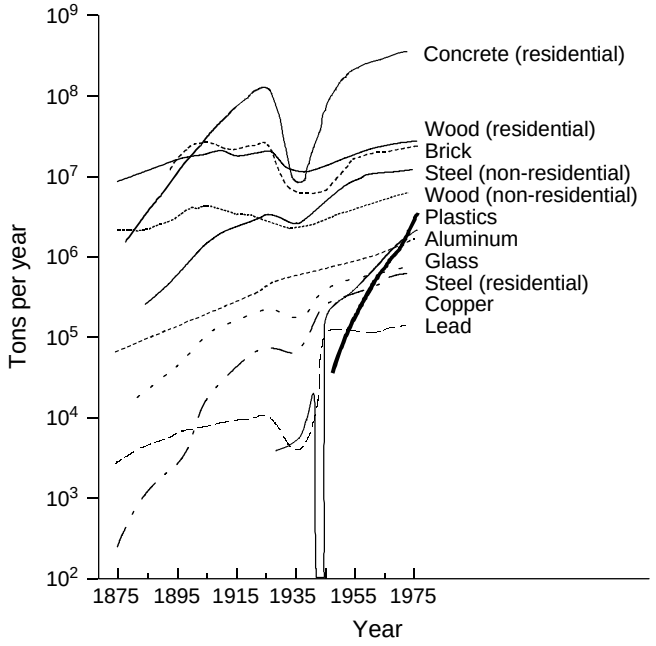


Figure 3 Construction materials used in the US from 1875 to 1975 (adapted from Wilson, 1975). Around 1900, wood was surpassed by concrete as the main construction material. Plastics and aluminum were among the fastest growing construction materials in the 1970s. Change in the consumption of construction materials also alters the construction stock. Future waste management will have to deal with large amounts of plastics and metals such as aluminum and copper

estimated that about 90% of the lead stock is still within the city, either in use or hibernating, and that only 10% is stored in waste landfills (Hendriks *et al.*, 2000). From this point of view, the hazardous substances stored in the so-called hazardous landfills are less than the amounts still within the city anthroposphere. With a few exceptions, the long-term fate and environmental significance of the urban stock of metals and organic materials is unknown (Bergback *et al.*, 1998).

When all stocks associated with the material chain are included, a first assessment reveals the following: waste deposits from primary production (mining wastes such as tailings) comprise about 10% of the total stock; in use and in hibernating materials within the anthroposphere about 80%; and land filled waste products about 10%. Hence, from a mass point of view, the most important future material reservoir is the urban stock, and not landfills and tailings. For an economic evaluation of this stock, the concentrations and speciations of the materials in these stocks may be even more important than the total amounts. The same is true for an environmental assessment. Up to now, the emission rates, chemical species, pathways, and final sinks of substances incorporated in urban stocks have not been investigated in a systematic and comprehensive way. Thus, aside from some specific case studies, it is not yet possible to discuss future effects on the environment, which may occur due to the dissipative material flows from urban stocks.

Anthropogenic Flows Exceed Natural Flows

Due to the large growth rate of the exploitation of essential minerals, man-made flows are approaching and even surpassing the natural flows of many substances. As a consequence, the flows, stocks and concentrations of certain substances such as heavy metals and nutrients are rising, causing the biosphere to change more rapidly and

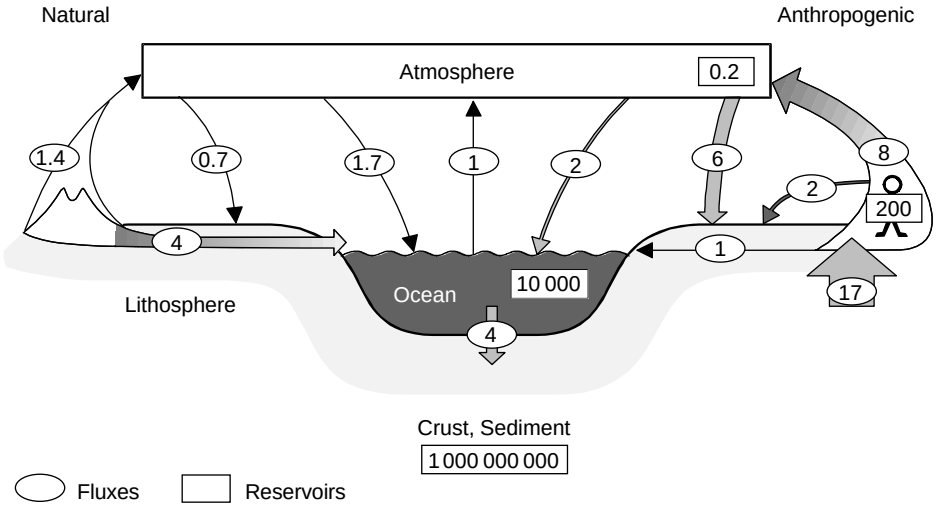


Figure 4 Global anthropogenic and natural Cd flows and stocks in the 1980s, in Kilotons per year (flows), and Kilotons (stocks). The man-made flows surpass the natural flows. Comparatively high atmospheric emissions lead to a significant accumulation of anthropogenic Cd in the soil. The stock of anthropogenic Cd grows by 3% per year. It needs to be managed carefully in the future if long-term environmental impacts are to be avoided (Baccini and Brunner, 1991)

in different directions than without anthropogenic material ows. Cd is used herein as a case study to exemplify the importance of today s man-made ows. In this case there is an abundance of data available. In general, the gap between actual and toxic concentrations in soils is smaller for Cd than for other metals.

In Figure 4, the global natural and anthropogenic ows of Cd during the 1980s are summarized. The total man-made ows of about 17 Kt year^{-1} are more than twice the size of the natural ows of about 6 7 Kt year^{-1} due to erosion, volcanic eruptions, and sea spray. Also, the input into the anthroposphere (17 Kt year^{-1}) is larger than the sum of all outputs (11 Kt year^{-1}) of the anthroposphere. Thus, about 6 Kt year^{-1}, or 30% of the Cd annually mined from the Earth s crust, is accumulated in long-lived goods. The Cd stock in the anthroposphere grows with a rather short doubling time of 30 to 40 years. In the 1980s, the most important pathway for Cd from the anthroposphere to the environment was the atmosphere. This is due to (1) the physical chemical properties of Cd, which has a comparatively high vapor pressure as an element and for most of its compounds; and (2) the lack of appropriate ue gas cleaning technology two decades ago. The man-induced global ows of Cd into the atmosphere were an order of magnitude larger than the natural ows. Today, the emissions into the atmosphere are lower, due to advanced ue gas cleaning technology in metallurgical and waste treatment processes.

The anthropogenic Cd ow by surface waters to the oceans is still smaller than the transport of weathered Cd by the large river systems. The main sink for Cd was and still is the soil, followed by ocean sediments. Regional Cd balances reveal that in densely populated areas, the atmospheric deposition of Cd onto the soil is greater than the output from the soil by leachate, erosion or plant uptake. This situation is indicated in Figure 4 as well. The global input into the soil by atmospheric deposition (6 Kt year^{-1}) and land lling (2 Kt year^{-1}) is much larger than the output by erosion and river systems (1 Kt year^{-1}). Comparing the natural side (ratio of deposition to leaching of 17 : 1) with the anthropogenic side (ratio of deposition to leaching of 8 : 1) in Figure 4, the large accumulation of anthropogenic Cd in the soil becomes evident.

For Cd and other metals, humans are already more important than nature as material transformers. This is true not only at the global level, but also, and more often, at the regional level. Investigations from Sweden (Figure 5) (Bergback, 1992; Azar *et al.*, 1996) show that the weathering rates of several heavy metals are surpassed by the weathering (or emission) rates due to anthropogenic activities.

The consequences for both regional and global levels are the same: the reservoirs in soils, the anthroposphere, and land lls are growing. For some materials, the industrial contribution has already caused the prehistoric concentrations in the reservoir to double, while for others, this will happen in the future, if material management does not change fundamentally. The long residence times of materials in the anthroposphere and, especially in the soil, will result in a long lasting challenge for future generations. Our successors will be given a large source of secondary raw materials, but at the same time they will have to make sure that the emissions from this source are small and under control.

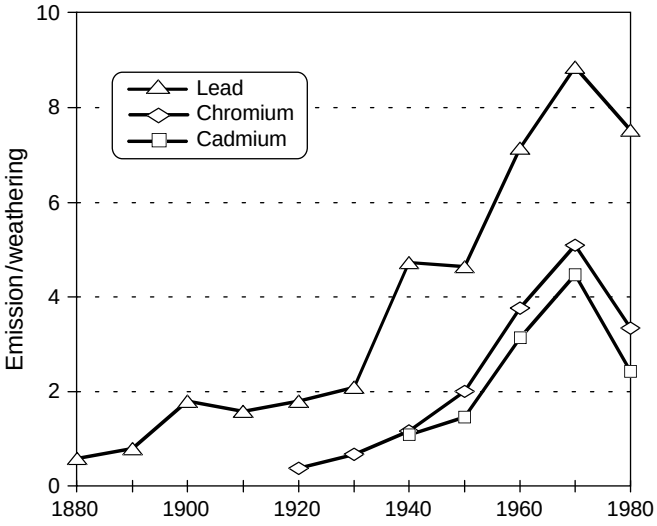


Figure 5 Anthropogenic emission rates of lead, chromium, and cadmium are larger than natural weathering rates (Bergbäck, 1992). Although these examples come from Sweden, similar results have been observed in other countries with advanced economies

Consumption Emissions Surpass Production Emissions

A third phenomenon of the metabolism of advanced societies is that production emissions are decreasing, while certain consumer emissions are increasing. The data (Bergback, 1992) indicate that for heavy metals in Sweden, the change in dominance of industrial sources towards non point sources occurred in the 1970s (Figure 6). This is due to, on the one hand, advanced technology and legislation in the eld of industrial environmental protection. On the other hand, the high and still growing rate of consumption has led to large stocks of materials that have to be maintained and that discharge material that is hardly noticed, but in total is signi cant to water, air, and soil. Hot spots of such non-point sources are cities. Examples of consumer related emission sources are the weathering of surfaces of buildings (zinc, copper, iron, etc.), the emission of carbon dioxide and other greenhouse gases due to space heating and transportation, the corrosion and erosion (chassis, tires, brakes) of vehicles and infrastructure for traf c, and the growing problem of nitrogen overload.

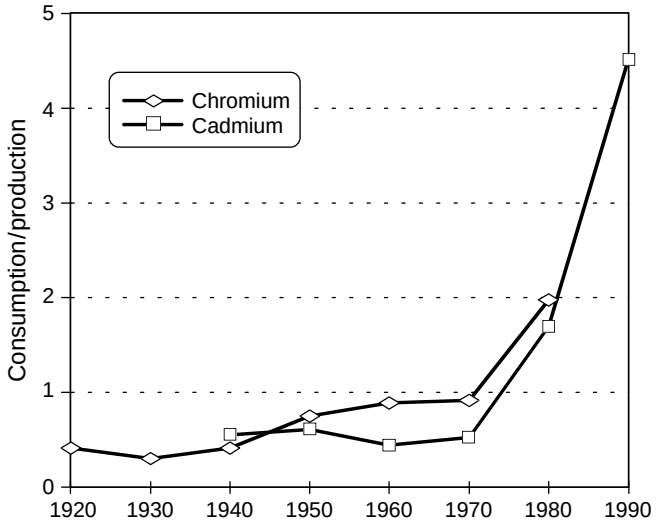


Figure 6 Rates of consumption emissions versus production emissions (Bergbäck, 1992): for modern economies with a high environmental standard, consumption is relatively more important than production. This requires a shift in pollution abatement: future environmental management must concentrate more on non-point sources and on the design of environmentally acceptable products, and has to include consumer behavior and lifestyles

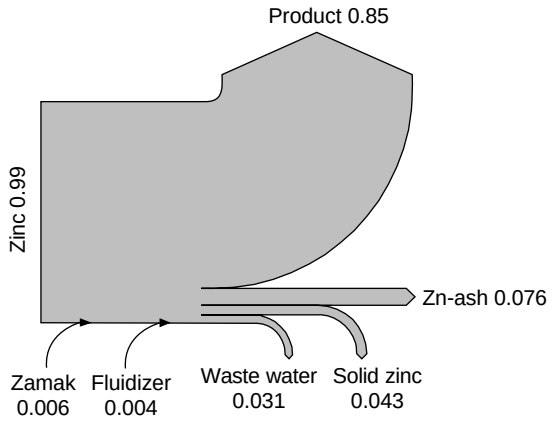


Figure 7 Zinc flow through an advanced electroplating plant: 85% of the zinc used is sold with the product; it protects the product surface against corrosion. Nearly all of the 15% zinc not sold is recycled. Emissions to the environment during production are very small. During use of the zinc-coated product, the major part of the zinc is dispersed in the environment due to corrosion. Hence, despite the effort to protect the environment by pollution prevention measures in the factory, most zinc ends up in the environment

In Figure 7 (Brunner *et al.*, 1996), an example is given to illustrate the significance of consumer emissions versus production emissions. Modern electroplating plants produce only small residues; most of the metals used to protect surfaces leave the factory incorporated in the product. In general, the solid and liquid metal residues are recycled to a large extent, resulting in low losses of metals to the environment. Thus, from a company point of view, the goal of pollution prevention has been achieved, and most of the zinc or chromium leaves the factory as coatings on products. Nevertheless, during the lifetime of the product, processes such as corrosion and weathering damage the surface coating, resulting in losses of heavy metals to the environment. Hence, despite pollution prevention measures during the electroplating process, as much as 85% of the zinc may end up in the environment (Figure 7). Since the corrosion process is slow and the residence time of metals on surfaces can be rather long, the flows of metals to the environment last for years to decades. The legacies of today's protected surfaces are tomorrow's environmental loadings. At present, the significance of such emissions has not been assessed and evaluated. There are no methods available yet to measure and appraise these sources of pollutants to the environment.

To prevent consumer-related emissions is more difficult than to stop industrial pollution: (1) it is a much more demanding challenge for a company to change or discontinue a product than to add an air pollution device to its production facility; (2) the number of (consumer) sources is many orders of magnitude larger than of production facilities; (3) the individual emission may be very small, and only the multiplication by the large number of sources may cause a hazard for the environment. Thus, consumer awareness for the overall relevance of his/her small contribution must be created first, in order to be able to tackle the problem. (4) In general, emissions of industrial sources can be reduced with a much higher efficiency than small-scale household emissions. An example is waste combustion: due to sophisticated air pollution control systems, the specific emissions per tons of waste incinerated has decreased in the last 50 years by several orders of magnitude (reduction of e.g., dust $>10^5$; lead $>10^5$; dioxin $>10^3$). Attempts to decrease emissions by changing consumer behavior (collection of potentially hazardous materials such as lead or chlorinated materials) are not as effective. They reduce the flows by a maximum of one order of magnitude only. So far, the general public does not notice the large cumulative power of consumer emissions. It has not been anticipated yet that the poor practice of burning waste materials in a few home heating systems and wood stoves leads to higher overall loads of dioxins than nationwide state-of-the-art waste incinerators.

The Linear Metabolism of Cities and Their Hinterlands

Today's cities are mainly linear throughput reactors. The material inputs into cities are slightly larger than the outputs, resulting in a growing stock of materials. In Figure 8, the anthropogenic metabolism of Vienna is presented as a

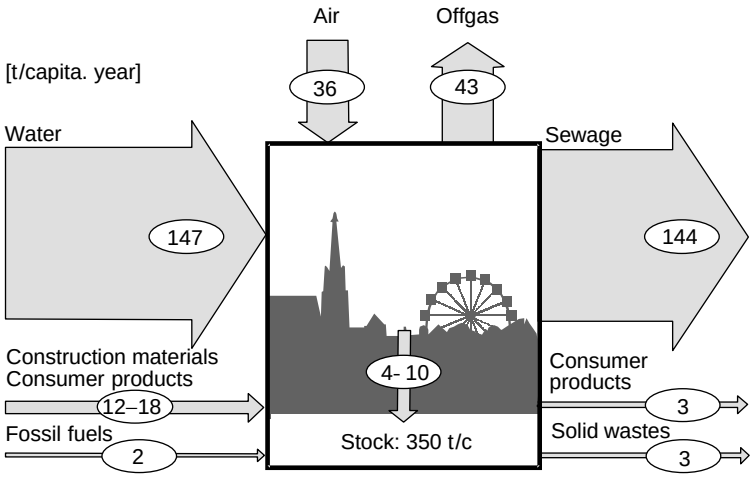


Figure 8 The flows and stocks of materials in the city of Vienna. Inputs (200 t c^{-1} year^{-1}) of materials into the city are larger than outputs, resulting in continuous growth of the already large stock, which doubles in 50 to 100 years. Basically, the modern anthropogenic metabolism can be characterized as a linear throughput reactor, with less than 1% of materials being involved in regional cycles

typical example (Obernosterer *et al.*, 1998). The material stocks in private households, the public and private sectors, and the infrastructure double in about 50 to 100 years. This stock is made up from materials exploited from a local, regional and global hinterland. The hinterland of a city is not a speci c geographical area, but it comprises many scales, depending on the materials discussed. In the case of Vienna, the ow of water, which comprises the largest single ow of a material for all current cities, comes from regional mountain springs, while most construction materials are drawn from local resources. For many metals and nutrients, Vienna depends on a global hinterland. Even if some of the essential materials could still be found locally, it is often more economical to bring in these materials from a globalized world market. Rees calls the impact a city has on the hinterland the *ecological footprint* (Rees, 2001). This footprint comprises materials included in the production of goods, such as water needed to grow food produced in the hinterland but consumed within the city (*see* **Ecological Footprint**, Volume 3).

The large and growing stock of materials in cities represents a resource for the future, if certain conditions are met: the materials must be incorporated into the structure of a city in a way which allows one to nd, collect and separate valuable substances again. The concentrations of the substances must be high enough for economic recovery, or low enough for environmentally safe disposal. At present, this is not the case for most materials. Exceptions are extremely valuable substances such as gold, or highly toxic materials such as plutonium, both of which are managed very carefully. The elevated concentrations of these materials in their nal sinks (soils and sediments) show how dif cult it is to control the ows of materials.

The linear ow through cities means that the hinterland has to cope with the leftovers of urban activities. Hence, the hinterland serves as a supply system for the city on the one hand, but is also a disposal system on the other hand. Actually, the second function is becoming increasingly important. The growth in material ows implies more wastes as well. Due to the build-up of long-term stocks, there is a time lag between material input and output. Future amounts of waste materials will be much larger than nowadays, even if consumption were to remain at present rates. Hence, the city s metabolism and corresponding stock determines tomorrow s emissions and state of the environment in the hinterland!

The importance of the hinterland depends on the waste management system of a city, and the substances involved. For nutrients such as nitrogen, the receiving waters of the hinterland are crucial. While many anthropogenic nitrogen emissions such as NO_x from automobiles or industrial sources can be reduced, the direct nitrogen ows from humans cannot be changed signi cantly; they can only be decreased by expensive wastewater management and treatment methods. Thus, the fast growth of large urban populations can have a severe impact on the receiving water system. In addition, the lifestyle of the urban population is important: if the nitrogen is consumed as animal protein, the hinterland emissions for the production of this protein are very high. If the dietary nitrogen stems from plant sources, the hinterland emissions will be 50% smaller. Thus, in order to protect surface waters as well as adjacent groundwaters, and the estuaries where the rivers join the sea, nutrient management programs on a large scale are necessary, integrating agriculture as the primary source with food industry emissions, traf c emissions and other urban nutrient ows.

While nutrient flows are of concern on a watershed (regional to continental) level, greenhouse gas emissions are of global significance. For every city, the most important hinterland, from a disposal point of view, is the global atmosphere. Without this large conveyor belt and intermediate sink for carbon dioxide, the energy metabolism of cities could not survive a single day! It is the stock of materials in the infrastructure of a city that determines the emissions of greenhouse gases. If this stock includes advanced insulation of residential and commercial buildings, city-wide central heating systems, efficient public transportation, appropriate structures to shorten transportation distances for commuting, shopping and leisure etc., the emissions could be much less than in present cities. Since the residence time of the city stock is long (decades to centuries), a rapid change of materials and structure of the stock is not feasible. Thus, the present stock of materials will determine future emissions of greenhouse gases for quite some while.

Besides the large global and regional emission pathways, there are local sinks for urban wastes, such as landfills. Today, most of the minerals imported into cities are disposed of in sanitary landfills. Although these anthropogenic sediments currently contain not more than about 10% of the stock of metals within a city, they will grow and may become an important source of resources in the future. In particular, if specific wastes are concentrated and accumulated for future reuse and are not diluted with other wastes or additives such as cement, future amounts may become large enough to make re-use economically feasible.

Analysis: How to Assess Material Flows and Stocks?

First attempts to determine the metabolism of urban regions date back to the 1970s. Duvigneaud and Denayeyer-De Smet compared the city of Brussels with a natural ecosystem, and concluded that the knowledge about the material and energy turnover of cities was still very small (Duvigneaud and Denayeyer-De Smet, 1975). They suggested redesigning the city metabolism in a way that allows more recycling, less pollution and better use of resources. Newcombe *et al.*, describe the metabolism of Hong Kong in a similar way. They, too, ask for a better understanding of the key metabolic processes in order to improve future urban design (Newcombe *et al.*, 1978). A comprehensive method to describe the metabolism of anthropogenic systems was developed in the 1980s, and is described by Baccini and Brunner: *material ow analysis* (MFA) is based on the following definitions (Baccini and Brunner, 1991; Brunner *et al.*, 1994): the term *material* stands for both substances and goods. A *substance* consists of uniform units (atoms or molecules) and is a chemical element (e.g., lead, carbon) or a chemical compound (lead chloride, carbon dioxide). Substance flows are measured in mass per time units, substance fluxes in mass per time, and cross section. The cross section can be an entire region, a household, or a person; hence the flux unit might be kg *per capita* and year. The selection of substances for an MFA is a demanding task. In general, for several reasons (data availability, limited resources) the number of investigated substances is small. Therefore, it is necessary to select certain so-called indicator substances. These are elements or compounds that show a behavior typical for a group of substances. For instance, the partitioning of Cd in a combustion process can be regarded as exemplary for volatile elements such as zinc, antimony, and tin. Iron serves as an indicator for nonvolatile metals such as cobalt, chromium, manganese, and nickel. The selection of appropriate indicator elements also depends on the goods and products investigated. Regarding the combustion of coal, for example, arsenic and selenium are appropriate indicators because coal is a major source for these elements (Greenberg *et al.*, 1978). One way to determine indicator elements is to establish the ratio of elemental concentrations in the product to the average in the earth s crust. Elements with high ratios are candidates for the MFA (cf. Table 2). Another criterion for the selection of substances can be emission limits as defined by environmental legislation.

As stated above, a *good* consists of one or many substances, such as a pipe made of lead, or gasoline containing benzene. A good has a negative (e.g., waste) or positive (e.g., car, drinking water) economic value. In the economic sense, goods can also be energy, information, or services. In classic MFA, which is the subject of this article, the term is applied to material goods only.

A *process* is defined as a transport, transformation, or storage (stock) of goods, materials, energy, and information.

Table 2 Ratio of elemental concentrations of oil, coal and MSW (Municipal Solid Wastes) to the average on the earth's crust as a means of selecting indicator elements for an MFA[a]

	Cl	As	Cd	Co	Cr	Cu	Hg	Ni	Pb	Se	V	Zn
Oil/ØEC	*0.7*	*0.4*	0.05	0.04	0.01	0.06	*0.3*	*0.7*	0.02	**10**	**1.5**	0.05
Coal/ØEC	*7.7*	**13**	2.5	0.4	0.2	0.6	*5*	0.4	1.5	**38**	0.3	0.7
MSW/ØEC	**77**	2	**50**	0.2	2	*11*	**100**	1	**62**	6.3	0.03	*17*

[a] Bold figures indicate high relevance, italic figures medium. (Fiedler and Rösler, 1993; Merian, 1984; Tauber, 1988; Zimmermann *et al.*, 1996; Frischknecht *et al.*, 1995).

A transformation often involves a change in the value of a good. There are processes possible on all levels: a car engine may be looked at as a process in the same way as the human body, a private household, a waste incinerator, a branch of a regional economy, or an entire region.

Finally the investigated system comprises the processes that are linked via flows of materials and goods. The system boundary is defined in space and in time. Depending on the objective of the MFA and the availability of data, the boundary in time can be hours, days, or a year.

An *activity* is defined as a set of processes and fluxes of goods, materials, energy, and information serving an essential basic human purpose, such as to nourish, to clean, to reside, or to communicate. Hence, the concept of activities allows one to evaluate the design and management of entire material flow and stock systems with the objective of meeting certain goals such as sustainability. For example, the activity to nourish, comprising the production, upgrading, storage, distribution, preparation, and consumption of food, involves large fluxes of energy, nitrogen and phosphorus, which have an impact on water, air, and soil. An MFA of the entire activity to nourish, including agricultural production, industrial processing, distribution of food, and preparation and consumption of meals reveals those processes and flows of goods which are most important from the point of view of resource conservation and environmental loadings. Thus, MFA is an essential tool for the analysis and design of sustainable systems.

Each flow or flux has a process of origin and a process of destination, and thus is precisely defined. Equally, each process is linked with other processes by means of fluxes or flows. A good, which flows from Process A to Process B (Good 3) is called an output good for Process A and an input good for Process B. Import (Good 1) and export (Good 2) goods are defined as goods which cross the systems boundary. The same terminology applies to the flux of substances. In Figure 9, a process is graphically presented by a square, a good by an oval, and fluxes of goods or materials by arrows. Although these definitions seem tedious, they become important as soon as one attempts to link various data about the flux of materials of individual processes. Often the measurement of the output good X does not yield the same result as the measurement of X as an input good into the next process. For example, the figure for consumption of drinking water varies by 20 30% depending on whether it is measured at the water supply (output good) or the consumer (input good). Daxbeck & Brunner (1993) subdivide an MFA into the following nine steps.

1. Outline of the objectives of the MFA.
2. Definition of the system, including the selection of system boundaries (space and time), processes, goods and substances. First rough assessment of balances of goods and substances, using easily available data in order to identify the relevant goods and processes.

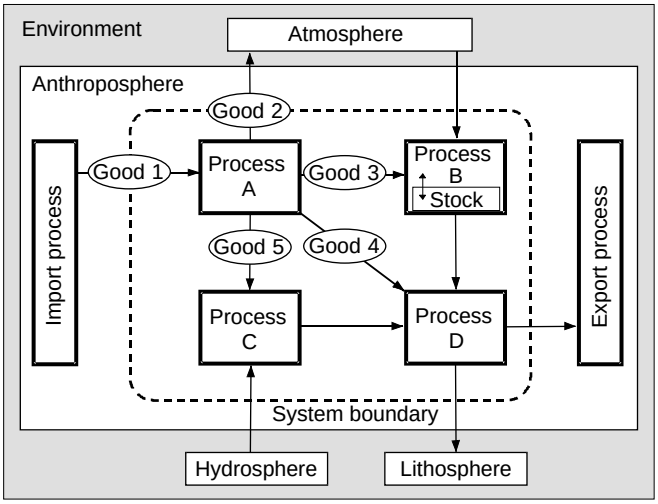


Figure 9 Example of definitions and symbols for system boundaries, processes, and flows of goods as used in MFA methodology

3. Establishing a program for the investigation and measurement of flows of goods and substances, taking into account the necessary accuracy of the final results.
4. Determination of the mass-flows of the goods (m). Mass balance for goods for process A in Figure 9:

$$\sum_{i=1}^{1} m_i = \sum_{i=2}^{5} m_i$$

5. Determination of concentrations (c_j) of the selected substances (j) in the goods.
6. Calculation of the substance flows ($m \times c_j$) including an estimation of the uncertainty of the results. Substance balance for process A in Figure 9:

$$\sum_{i=1}^{1} c_{i,j} \times m_i = \sum_{i=2}^{5} c_{i,j} \times m_i = X_{A,j}$$

7. calculation and optimization of the transfer coefficients (k_j): transfer coefficient of substance j for good 3 of process A in Figure 9:

$$k_{A,3,j} = \frac{c_{3,j} \times m_3}{X_{A,j}}$$

8. popular presentation of the results and interpretations including a detailed report to ensure transparency and reproducibility.

Evaluation: How to Determine the Environmental Significance of the Flows and Stocks?

The results of an MFA are mass balances for goods and several substances. Obviously they can be interpreted separately, as they indicate the most important processes and

goods from which appropriate strategies for an optimized materials management can be derived. In this section, some additional, more or less established, methods for evaluation are briefly described. Initially, these methods were not developed for the urban metabolism, and with the exception of the second approach they have not been applied to the metabolism of cities either. This might be partly due to the lack of data availability, and the practicability and complexity of the methods.

Evaluation by Emission Limits

A rst way to evaluate a material-balance of a system is to compare substance concentrations in various mass- ows with emission standards of related legislation. Usually such limits are laid down for gaseous and aqueous emissions in the various national environmental protection directives. For instance, emissions of factories, power plants, waste treatment facilities and so on, are controlled this way by local authorities. However, non-point (dissipative) emissions and emissions via products and goods can hardly be evaluated by this approach. Therefore, estimates of emissions exclusively by concentrations are not yet suf cient; the total loads discharged into the environment have to be considered as well. The advantage of a complete balance is that the ows of a substance via residues and products that are not regulated by legislation are also evaluated. Nowadays, this aspect is becoming more and more important. For example, as most of today s applied waste treatment technologies obey stringent emission limits, their emissions are negligible and no longer important for decision-making. Nonetheless, due to different process technology, they produce completely different amounts and qualities of residues. A comparison based exclusively on emissions would yield no difference between such technologies, and is therefore not suf cient and is not goal-oriented.

Evaluation by Relating Anthropogenic to Natural Flows and Stocks

This approach investigates the changes of natural ows and stocks through anthropogenic activities (ows and stocks, induced by society). Accordingly, this approach considers the following two important requirements for sustainability (Enquete-Kommission, 1994; SUSTAIN, 1994). First, substance ows from the anthroposphere to the environment should neither exceed local and global assimilation capacity nor surpass the range of natural substance ows. Second, substance ows should be managed in a way to prevent the depletion of useful (resources) substances and the accumulation of harmful substances in the environment. These goals re ect the *precautionary principle* as laid down in modern environmental legislation (*see* **Precautionary Principle**, Volume 4). The principle is derived from the realization that controlling the urban metabolism exclusively via threshold values is not yet possible. First, these values continuously change as a result of enhanced knowledge, and second, they are mostly a result of damage to human health and the environment that has already happened.

The following example demonstrates how the method can be applied. The change in soil by the emission of Cd from a MSW incinerator (anthropogenic-induced material ow) is investigated. The plant processes 200 000 tons of MSW year^{-1}. It is equipped with an advanced ue gas treatment system, which results in an extremely low transfer coef cient of Cd into the off-gas: $k = 0.001$ ($\approx$0.002 mg Cd Nm^{-3}). The mean concentration of Cd in MSW is about 10 mg kg^{-1}. This yields an annual emission of 2 kg Cd to the atmosphere. The plant is situated in an urban area with a population density of 4000 inhabitants km^{-2}. Waste generation rate amounts to about 150 kg per person and year. Thus, the incinerator has a catchment area of roughly 300 km^2. Assuming that the 2 kg of Cd are evenly distributed over this area, a deposition rate of about 7 μg m^{-2} year^{-1} can be calculated. (Actually, due to the aerosol character of the Cd emission, the burdened area would be much larger, and the real deposition essentially lower.) The content of Cd in natural soil is about 70 mg m^{-2}. Therefore, the incinerator would cause a doubling of the Cd content within 10 000 years. Even considering the above simpli cations, such an emission can be judged as environmentally compatible with this approach.

Contrary to the evaluation via emission limits, this approach considers the load of an emission or product. It has been successfully applied to the metabolism of a city by Paumann *et al.* (1997).

Entropy-based Evaluation (MASTER)

MASTER (Method for the Analysis of the Substance concentrations of Treatment End-products and Recycling goods) was initially developed as an evaluation tool for waste management.

The MASTER approach quanti es the power of a system, be it an engine, a factory, or a city, to dilute or concentrate substances with a single measure (Rechberger, 1999). This method can be directly applied to the results of any MFA. While the basis of an MFA is the principle of the conservation of mass (a special case of the First Law of Thermodynamics), MASTER is based on the Second Law (known also as the Entropy Law): the approach utilizes Statistical Entropy to evaluate substance partitioning of a system. It can be expected that entropy is one of several appropriate measures to judge urban metabolism, as it quanti es the dissipation caused by point and non-point emissions as well as of stocks. The advantage of linking MFA and MASTER is that one single measure in the range between [0, 1] per substance and system is derived. Therefore, this coupling is well suited to establish time series for urban material balances.

Life Cycle Assessment (LCA) and Swiss Eco-Points (SEP)

LCA (SETAC, 1996) and SEP (BUWAL, 1998) are methods for evaluating the emissions caused by a product, a process or a system. An LCA consists of the following steps: classification, characterization, and evaluation. First, emissions are identified, inventoried and assigned to predefined stressor categories such as greenhouse effect, ozone depletion, human toxicity, etc., (classification). Second, the contributions of the different emissions are assessed, calibrated and aggregated into relatively homogenous categories (characterization). In the evaluation step, which is optional and not entirely based on scientific principles, but on socio-political (subjective) values, the impacts are weighted and aggregated to one final single index. The SEP method is a one-step process based on the concept of limited environmental absorption capacity for pollutants. Therefore, it uses a weighting system to aggregate substance flows (eco-points) by considering critical loads and actual total loads of selected substances mostly at the national level.

Material Input per Service Unit (MIPS)

MIPS (Schmidt-Bleek, 1994; Schmidt-Bleek *et al.*, 1998) is a measure for the aggregated mass flow due to the consumption of a defined service (e.g., the total material and energy used to transport a good from A to B). In the MIPS concept, all materials including ores, mining wastes, auxiliary goods, water, air, off-gas, etc., which are required to render a certain service, are taken into account for the entire lifetime of the goods associated with the service. The MIPS method defines the procedures to set system boundaries, to generate data and determine how to handle the issue of lack of data. The method is limited to the level of goods and does not allow one to evaluate the substance flows needed for services. It serves well as a method for a first assessment of different options. However, without considering individual substances, the MIPS concept can result in misleading conclusions.

Sustainable Process Index (SPI)

The SPI (Krotscheck, 1995) is based on the assumption that for sustainability, solar *exergy* (available solar energy) is the limiting factor. The utilization and conversion of solar energy into usable exergy by photovoltaic cells, biomass or other means requires a land area. Hence, in an SPI-evaluation, all material and energy flows of the investigated product (or process or system) are transformed by a defined procedure into land area equivalents. The SPI equals the total area calculated for the product, divided by the specific area available for an average person living in the investigated region. Sustainability is achieved if the SPI is <1. The SPI is a so-called holistic approach as it considers not only traditional resources (materials, energy, area) but also attempts to include social aspects such as employment, residential area of employees, etc.

The methods summarized above do not yet allow one to comprehensively evaluate the environmental relevance of the anthropogenic metabolisms. So far, no method exists to appraise the diffusive emissions resulting from the weathering of surfaces in a city, surfaces exposed to the atmosphere (roofs, cars), or buried in the ground (pipes, cables). Also, the long-term fate of materials hibernating in cities has not yet been investigated. Thus, new methods have to be developed in order to supplement existing evaluation methodologies.

Case Studies

The following two case studies illustrate how material flow and stock analysis can be used to assess, evaluate and control the anthropogenic metabolism. Both examples have been developed and used to support decisions in the field of environmental and waste management. They are presented here to point out the qualitative and quantitative importance of material stocks. Efficient decisions regarding environmental protection and resource conservation can be taken only if all flows and stocks of a material and their contributions to the total metabolism are known.

Plastic Materials

The flows and stocks of plastic materials in Austria have been investigated by an MFA (Fehringer and Brunner, 1997). This assessment is summarized in Figure 10. The total per capita consumption of plastics amounts to about 140 kg (8.1 million inhabitants); 50 kg are incorporated into the stock, and 90 kg are collected and treated by waste management systems; 84% of the plastic wastes are land filled, 10% are incinerated, and 6% are recycled. There are two important plastic stocks: the accumulation in the consumption process (buildings, transportation and communication systems, households, industry and trade, etc.), and the waste plastic deposits in landfills. Landfills contain more plastic and grow faster than any other stock.

In addition to polymers such as polyethylene or polyvinylchloride, plastic goods contain additives such as softeners, stabilizers, biocides and fire retardants (Table 3). These additives are much less abundant in packaging materials. The stock of some of the potentially hazardous additives in use and hibernating in Austria amounts to several 10 000 tons. The mass of additives in landfills is even larger. Information on the chemical composition of the plastic flows and stocks is scarce and hard to collect. Hence, an important base for environmental and resource management is lacking at present.

An evaluation of the plastic flows and stocks in view of the goals of the Austrian Waste Management Act (protection of mankind and the environment, conservation of resources materials, energy and space) reveals

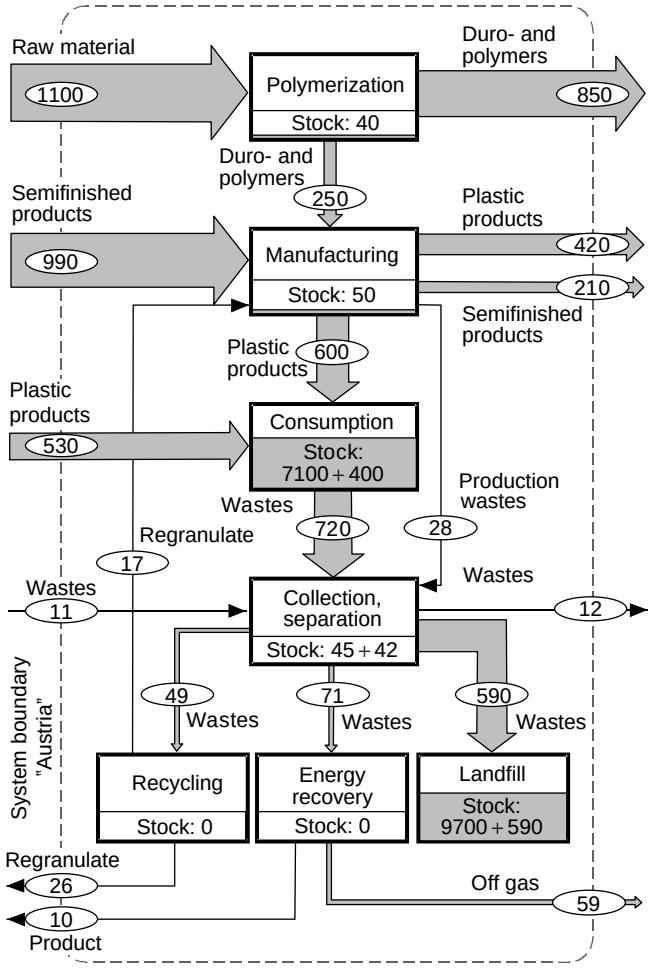


Figure 10 Flows and stocks of plastic materials in Austria in 1994 in kt/year: 40% of the annual consumption of plastic materials is accumulated in the stock. Most of the residual 60% is disposed of in landfills, which are already the largest and fastest growing stocks of plastics in Austria. The federal packaging ordinance, which requires separate collection and recovery of plastic packaging materials, has little effect (<10%) on the overall plastic flows and stocks (Fehringer and Brunner, 1996)

the following. There are hardly any immediate and direct environmental hazards resulting from present plastic waste management. Nevertheless, since the fate of in excess of 10 000 tons of heavy metals and organic substances contained in land filled waste plastic is not known, a long-term risk may exist. This risk needs to be addressed, taking into account slow biogeochemical reactions of polymers and additives, as well as immobilization and mobilization reactions in landfill environments. It must be pointed out that most of today s landfill research is directed towards comparatively fast reactions, which take place within months to years, rarely decades. In the case of polymers, investigations are needed concerning slow reactions, which take place over decades to centuries. First assessments of the behavior of landfill liners indicate that plastic materials are not as stable in landfill environments as anticipated.

Land filling plastic wastes means burying valuable fuel. In Austria, about 500 000 tons of fuel oil equivalents are disposed of in landfills every year. This clearly contradicts the goal of conservation of resources of the Waste Management Act. On the other hand, about 50 000 tons of packaging plastics are recycled at extremely high cost. Recycling of stabilized plastic wastes means keeping the hazardous constituents within the anthroposphere. This cycle either needs to end in an advanced waste treatment plant, where the hazardous materials are extracted, concentrated and immobilized, or the cycle must be improved by new, sophisticated technologies not available yet. In order to integrate products made from plastic wastes, such as artificial wood fences, roof tiles or concrete artifacts in an environmentally sound waste management strategy, concepts for the safe final disposal of these products must be developed as well. Otherwise, the hazardous constituents are dispersed in the environment by burning the artificial wood in household furnaces, etc. The general public cannot be blamed for such practices, since the origin and legacy of recycling products is neither known nor trackable.

Table 3 Characterization of the plastic flows and stocks in Austria. Packaging polymers contain few additives. In contrast, plastic materials with long residence times, e.g., for construction purposes, have to be protected against degradation by light, temperature, mechanical, chemical and biochemical stress. Hence, they often contain large amounts of heavy metals and organic stabilizers. These constituents may jeopardize future recycling. If not properly managed, they will also pollute the environment (data from Fehringer and Brunner, 1996)

Material	Consumption kt 1994	Packaging material kt 1994	Stock kt 1994
Polymers	1100	200	7100
Softeners	14	0.2	140
Ba/Cd-stablizers	0.27	0.0002	2.6
Pb-stablizers	1.8	0.002	18
Fire retardants	2.3	0	22

In the future, decisions regarding the management of plastic materials and wastes should be based on appropriate, comprehensive information, including gures about individual polymers, additives, stabilizers and so on. In order to control the risks and exploit the resource potential arising from large stocks and ows of plastic materials, the management of plastic goods needs to be redesigned. Most polymers are produced from fossil fuels. The largest part of fossil fuels is used for traf c systems, heating, power generation and industrial purposes; plastic manufacturing is of minor importance. The incineration of waste plastics can reduce the overall consumption of fossil fuels and of greenhouse gas emissions. Incineration is especially important for polluted plastic materials, containing hazardous constituents. It is essential that appropriate nal sinks be found for these constituents, such as an ef cient thermal process for organic substances, or a secure land ll for immobilized heavy metals in incineration residues. Recycling cannot solve these problems, unless new technologies are developed to extract the hazardous components from waste plastic mixes. Packaging ordinances cover only a small part of the total ows and stocks of plastics, are inef cient and do not tackle the problem of the hazardous constituents (cf. Table 3). In general, the total pathway and stocks of plastic materials should be taken into account when new plastic goods and management schemes are developed. It may be advantageous to start a new plastic era, where additives are omitted for environmental and recycling bene ts, and where clever chemists nd new polymers with better functions with less heavy metals and hazardous organic compounds.

CFCs

CFCs are a group of highly volatile compounds that contribute to the degradation of the stratospheric ozone layer. They have been widely used in aerosol sprays, in re extinguishing systems, as foaming agents in insulation materials for construction and refrigeration, as cooling agents, and as solvents and cleaning chemicals. In 1987, most CFCs were banned by the Montreal Protocol, due to their stratospheric ozone depletion properties. This treaty had an immediate effect on the consumption of CFCs, and on the emissions of short-term applications such as aerosol sprays. However, the in uence on the stocks of existing CFCs in use and hibernating is long-term. Waste management will have to deal with large amounts of CFCs during the coming decades, and recovery of the ozone layer will be a slow process.

In order to evaluate environmental and waste management strategies in the eld of CFCs, the ows and stocks of these chemicals were comprehensively investigated in a case study for Austria (Obernosterer, 1994). In Figure 11, the results of an MFA are summarized in ozone depletion units (ODP); these units are applied to account for the different effects of individual CFCs on the stratospheric ozone layer. The largest stock consists of insulation materials in construction. Other foams, coolants and re extinguishing agents account for about half of the CFC stock, household refrigerators for less than 3%. The OPD units in the present stock account for about one-third of all CFCs ever used in Austria. Thus, if not properly disposed of, the degradation of the ozone layer by CFCs will continue.

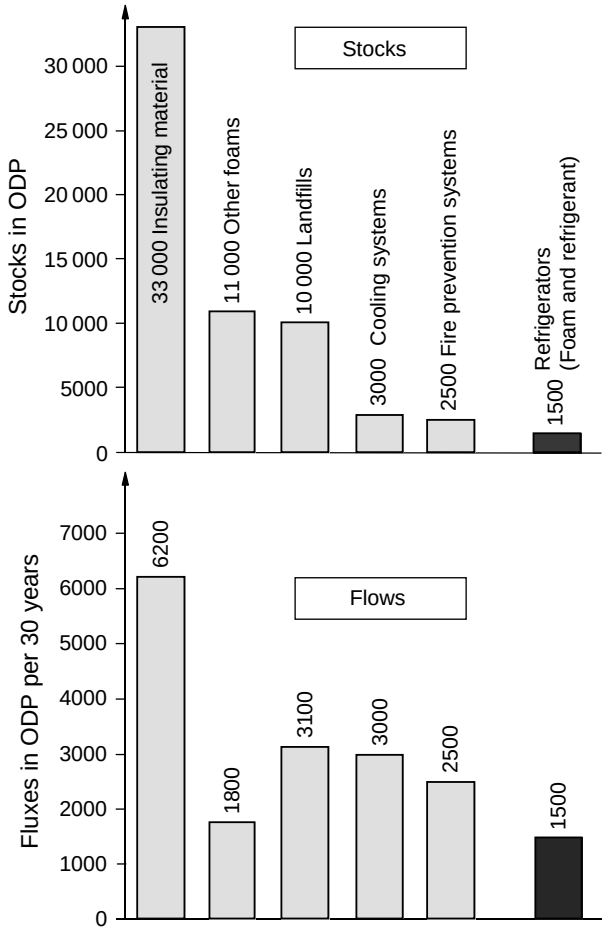


Figure 11 CFC stocks and emissions in Austria over the next 30 years, expressed as ODP. The largest stock and emission is due to insulation materials for construction. At present, waste management regulation focuses mainly on refrigerators, but this comprises less than 10% of the entire CFC stock and emissions in Austria (Obernosterer and Brunner, 1997)

The largest ow (emission) over the next 30 years will come from the stock of insulation materials in buildings. MFA reveals that top priority should be given to the ef cient collection and disposal of construction wastes. At present, legislation in the eld of waste management focuses mainly on refrigerators, which account for less than 10% of the emissions.

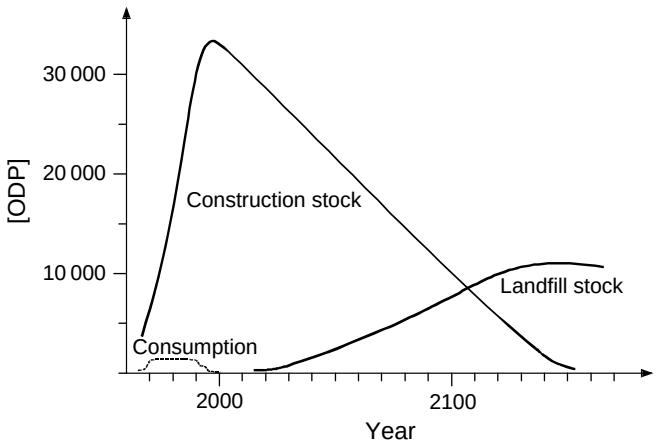


Figure 12 Consumption, buildup and shift of stocks of construction-related CFCs. During a relatively short consumption period of 25 years, a large stock of CFCs in insulation materials has accumulated, which will have to be managed carefully for the next 100 years. According to present waste management practice, this stock will be shifted gradually to landfills. If no counter measures are taken, CFCs will be lost to the stratosphere from both stocks – construction and landfills

The case studies of plastics and CFCs represent typical examples of many modern chemicals (Figure 12). During a relatively short period (approximately two decades), a large amount of a specific material is consumed and accumulated in the anthropogenic stock. If, due to changes in technology or new information about environmental risks, the production of the chemical ceases, there is already a legacy of a large stock of the chemical to be managed in the future. For safe and efficient material management, it is important to assess the significance of a material before large amounts are accumulated in the anthroposphere. In any case, technology needs to be developed to treat and dispose of the large accumulation of stock of materials. For CFCs and plastics in construction materials, this could be energy recovery with state-of-the-art flue gas cleaning, producing mainly carbon dioxide, water, chlorides, fluorides, and concentrated metals, when properly incinerated in a modern municipal waste incinerator (Vehlow *et al.*, 1992, 1994; Rittmeyer *et al.*, 1993, 1994).

LEGACIES DERIVED FROM THE ANTHROPOGENIC METABOLISM

Modern economies depend on a high material turnover. The stocks they build up and maintain amount to several hundred tons of materials *per capita*. This is not only an economic fact, but also a cultural reality. Material growth is incorporated into Western culture; it defines progress like no other single criteria. Future generations will have to deal with this stock; they can use, renew, recycle or dispose of it. The materials contained in the stock can serve as future resources. They also imply future problems in so far as they contain hazardous substances. During and after the use of the stock, substances will be lost, and they will pollute the environment. The maintenance and disposal of the stock will bring about considerable costs for future generations. Will future generations value this stock as a precious asset or as a technology/economy brake?

It is undisputable that if the present metabolic rate is maintained, there will ultimately be constraints for development. These may occur as resource scarcities at the supply side, or as environmental degradation at the disposal side of the anthroposphere. Will these constraints become effective in decades, centuries or millennia? Probably the most important legacy from today's anthropogenic metabolism is the addiction to it. The high growth rates and large stocks became so attractive and convenient to consumers and producers that alternative economic and cultural models do not seem to be noteworthy, and other metabolic options are neither studied nor implemented to a significant degree. The metabolism of an affluent society seems to confirm the Latin saying *plenus venter non studet libenter* (a full stomach does not like to study)!

Cities as Hot Spots

The largest anthropogenic material stocks have been built up in cities. Soon, more than half of the world's population will be living there, thus the urban stock will continue to grow. From a materials management point of view, this is an advantage. Human needs can be fulfilled more efficiently and with less supply of materials and energy in a densely populated area. From an environmental point of view, the same applies. Environmental protection measures are more efficient if many people are connected to a compact urban system than if the same number of persons are joined through a wide and sparse rural system. On the other hand, the material flow density in cities, expressed, for example, in mass per time and area, or in mass per deposition rate, is orders of magnitudes higher than in most rural areas. Modern cities are material hot spots containing more hazardous materials than most hazardous landfills. The case studies of plastics and CFCs exemplify the large stocks of anthropogenic materials. This stock is not managed nor has it been noticed yet. It is built into the urban infrastructure in a way that does not allow one to economically recover the materials without losses. On the contrary, many materials are used so that re-use becomes impossible, leaving a large mass of hibernating materials in cities. In particular, underground networks for the transportation of water, sewage, energy (gas, electricity) and information are not yet recoverable when their lifetime is over. Other examples are materials used to prevent corrosion and weathering (e.g., surface coatings, pigments),

such as zinc and copper. The emissions from these sources surpass the flows from point sources such as production facilities.

Cities are also hot spots for flow-through materials. In particular, nitrogen is of long-term concern, since this is an essential element for human nutrition. Wastewater treatment technology will have to be improved, and costs will rise, if large river systems and estuaries are to be protected from nitrogen overload. Alternative wastewater management concepts may become competitive, if urbanization continues to increase the nitrogen flows to surface waters.

While the dietary nitrogen emissions cannot be tackled by a prevention strategy, carbon dioxide flows, which are mainly produced for transportation and room heating, can. From a technical point of view, these emissions can be avoided by switching from fossil to solar energy sources. Thus, to tackle the hot spot city, a multi-facet strategy is needed, including prevention, new technologies and state-of-the-art waste management.

Limits at the Back End and Not at the Front Side

For most of human history, the anthropogenic metabolism was limited at the supply side. The foremost challenge of prehistoric mankind was to find enough food and shelter to meet essential needs. The development of civilization was accompanied by increasing requirements for supplies beyond basic physiological needs. Since the beginning of history, the search for energy and raw materials has led to many innovations, conquests, and wars. Today, the need for resources is greater than ever, the rate of consumption is large and growing, and a fair distribution of goods seems a much bigger challenge than exploitation of resources. At the beginning of the third millennium, the supply problem has been solved by modern technology, and severe resource limits at the front end are not in sight in the next few decades.

Today, the new question arises at the output side of the metabolism: how can we safely dispose of the very large amounts of waste materials? The need for resources is paralleled by the need for final sinks for these resources. This challenge is new. Never before has mankind had to care about the dissipation of anthropogenic materials on a global level, such as carbon dioxide or CFCs, although industrial activities always relied on the dissipation of residues in air and water. In the 20th century, the anthropogenic energy metabolism depended on the seemingly unlimited possibility of dispersing CO_2, the main product of energy conversion, in the atmosphere. However, CO_2 dispersal is no longer acceptable because of the concern about global climate change. This will require fundamental changes in the future anthropogenic metabolism, since, for example, a solar energy supply will result in significantly different material needs than a fossil fuel supply.

The search for appropriate final sinks is not only important for linear flow-through materials such as carbon and nitrogen compounds. Ultimately, materials in the stock and materials recycled need to be disposed of in final sinks as well. Up to a certain level, materials can be diluted in the environment without negative effects. On a global level, the limits for dissipation are not yet known. The controversy about carbon dioxide emissions causing greenhouse effects and leading to climate change shows how difficult it is to define these limits. Considering the large exploitation of metals displayed in Figure 2, two questions arise: where are these minerals now (in use, hibernation, in landfills, or dissipated in water, air and soil?), and where should they finally be? Neither question has been adequately addressed so far. There are no global inventories of the most important resources yet. While information about mining and primary production is abundant, data about product use and consumption are much more limited. Besides a few exotic examples such as gold and plutonium, there is very little information about the stocks, and there is virtually no systematic information about the fractions of materials already dispersed in soils, sediments, water, air, or landfills.

The Soil as a Long-Term Sink and Source

Sinks for anthropogenic materials either transform organic substances into mineralized natural materials (e.g., organic compounds are transformed to carbon dioxide and water) or store immobile inorganic materials safely for long time periods. They include:

- the atmosphere, where organic emissions may be transformed by reactions such as photolysis;
- the soil, where materials are deposited by atmospheric deposition, by agricultural and industrial practice, and by waste management;
- the sediments, which trap the pollutants imported into lakes and oceans by rivers and atmosphere.

If a sink for metals and inorganic matter is defined as a sphere on the globe where materials have residence times longer than several centuries, the soil may be called the most important sink, either for dispersed materials, or for wastes in landfills.

Material balances such as the cadmium example in Figure 4 reveal for all scales (local, regional, global), that, in general, the anthropogenic input of metals into regions is larger than the output (the sole exceptions are mining areas). The difference between the input and the output is in part built into the anthropospheric stock, in part accumulated in the soil because of atmospheric deposition, direct agricultural input or land filling with wastes. Soil monitoring programs confirm the results of material balances. They show increasing concentrations of heavy metals, organic chemicals, and phosphorous in soils, with

doubling rates for concentrations from decades to centuries. These accumulations have long-term consequences. On the one hand, there is now a phosphorous reservoir in many agricultural soils, which can supply valuable phosphorous. On the other hand, in mountainous areas, this stock is subject to erosion and is gradually transported to surface waters, causing eutrophic lakes for long periods, even if the agricultural application of phosphorous is stopped. The soil is only an intermediate sink for phosphorous, and it is a long-term source as well.

The same legacy corresponds to several metals in soils. According to the soil properties and the metal inputs, cadmium, lead, zinc and other heavy metals accumulate in soils, and may pose a future threat to food production and quality. This is a long-term process. Soil conditions, which are determined by the underlying rock formations, the rate of soil creation, inputs from the atmosphere and farming, and agricultural practice, are in a state of constant change. It is not possible to predict atmospheric deposition, soil use, and agricultural practice for long periods such as centuries. Therefore, since chemical speciation, and thus the availability of metals, depends upon the changing soil characteristics, a truly long-term forecast of the availability of heavy metals is difficult if not impossible. Stigliani talks about chemical time bombs (Stigliani, 1991) (*see* Contaminated Lands and Sediments: Chemical Time Bombs?, Volume 3). This is the reason why some nations have chosen to apply the precautionary principle, and to limit total concentrations of heavy metals in soils. The question remains what should be done once these maximum values are reached, which in densely populated, affluent areas is likely to occur within the next few centuries. A safe strategy for long time periods would be to adjust all inputs into the soil to the outputs, trying to balance soil concentrations continuously. This would require a large effort in air pollution control, cautionary and balanced agricultural practice, and considerable improvements in waste management concepts and practice. Such a challenging strategy does not appear feasible at this time.

Waste Management as a Future Key Process

The goals of waste management are to protect humans and the environment, to conserve resources such as energy, materials and land, and to reduce long-term risks from the disposal of wastes to an acceptable level. Hence, waste management is the paramount means to control the flows from the anthroposphere to the environment in a systematic and efficient way. In particular, land filling has a very important function. It must ensure that the large quantities of materials obtained from the ground find a long-term safe final storage place. A final storage landfill is defined as a (mainly inorganic) landfill with low emissions that do not alter the material flows in the surrounding environment (ground and surface waters, atmosphere) for long time periods (>1000 years). For most wastes, this can be achieved only if the wastes are mineralized and then immobilized.

The amounts of materials under waste management will increase because of rising consumption, and the huge amounts of materials in the stock that has to be renewed. Waste management inherits most of the materials in the stock. The composition of the long-living stock is more hazardous than the short-lived consumer products, which are recycled today, such as packaging materials. Even if recycling is extended, there will be large and rising amounts of metals and organic substances to be treated and disposed of safely. Waste management systematically closes the anthropospheric cycles of materials, giving them back to the Earth s crust in a controlled and environmentally sound manner.

Recycling is only an intermediate and not a nal solution for materials management. It does prolong the residence time in the anthroposphere, but for thermodynamic reasons, recycling cannot prevent the need for an ultimate sink. Recycling can be an excellent means to reduce the overall environmental impact of those materials requiring much energy and producing vast amounts of wastes and emissions during primary production. It also reduces the need for nal storage land ll space by decreasing the amount of primary material needed. However, recycling will not decrease the materials already used, and thus cannot change the future need for the disposal of materials already in the anthroposphere. At present, recycling is looked at primarily from a process technology point of view. In future, recycling has to be integrated into the design of new systems, processes and goods. The stock of a modern urban system can be ef - ciently extracted only if there is information available about the density and location of individual materials. As for geological exploration of natural resources, speci c tools are needed to identify and locate the future urban ores.

FUTURE STRATEGIES

The present metabolism of the anthroposphere is not sustainable for long periods of times. Emissions at the back end, the lack of appropriate sinks, as well as the scarcity of resources at the supply side, set limitations. The majority of the global population does not enjoy an af uent metabolism yet, but it has the same objective, i.e., to reach a similar material turnover as that of wealthier parts of the world. There are no signs yet indicating that in the future the developed world will break its growth path and slow down material consumption. Therefore, it is highly probable that global material ows and stocks will considerably increase in years to come. The rst warning signs are there already, such as the steady growth of the ozone hole, increasing greenhouse gas concentrations, nutrient overloadings in surface and groundwaters and others indicating that it is time

now to take action to redesign the present anthropogenic metabolism.

Today s situation may be compared to other ruptures in the development of society and technology, such as the introduction of industrial fertilizer or the change from biomass to fossil fuels for energy. Both of these changes allowed mankind to overcome severe constraints in the advancement of civilization. In the coming decades, equivalent changes in technology, economy and culture may occur. The anthropogenic metabolism must be redesigned to satisfy the needs for a global population based mainly on renewable resources without impairing the environment.

As a consequence, a new material management strategy is needed. Environmental considerations must become an essential part of material science. Decisions about material alternatives can be taken only if the fate of a material from source to final sink is known. Analysis and modeling of total material flows and stocks over long periods are needed. The availability of a final sink for all materials used is an important design consideration. If no appropriate final sink can be assigned to a material, it should be phased out and replaced. The possibilities and limitations of waste management to recycle and dispose of materials safely should be considered at the design stage. The new goals are that only three kinds of residues should result from the use of goods:

- residues and emissions that can be either completely mineralized or that can be safely dispersed in appropriate final sinks;
- clean residues suitable for multiple recycling;
- residues with final storage quality for which an appropriate final sink is available.

For modern economies with advanced environmental standards, emissions during consumption are relatively more important than emissions during production. This requires a shift of priorities in environmental management. Future measures must concentrate more on non-point sources, on consumer-related sources, and on the design of environmentally acceptable products; they must take into account consumer behavior and lifestyle. Also, a new awareness for materials must be created. At present, questions like Where is the final sink for this material? or How can this complex mixture be dismantled? are rarely asked by consumers and often not answered by designers.

In order to make better use of the stock, MFA must be supplemented by material accounting, i.e., the continuous monitoring of flows and stocks at a few key points of the anthropogenic metabolism. This allows for early identification of the accumulation and depletion of valuable resources, and for assessment of the future burden on the environment. In a competitive material economy, material accounting can result in a comparative advantage, as was the case when financial accounting was introduced.

City mining may eventually replace traditional mining of ores. Since the urban stock of materials grows constantly, cities become attractive concentrations of many minerals. If properly designed and maintained, it will be more economical to exploit the anthropogenic material stock than natural resources. Future urban areas may redevelop into partially industrial cities, taking advantage of their own stocks and taking a large part of their supply from their own domestic sources. Material extraction and transformation results in considerable amounts of waste energy, which in contrast to remote mining areas can be used in cities for heating and industrial purposes. The residues, which have to be land filled are much smaller for urban mining than for traditional mining.

It is not known how much time is left to counteract the negative effects of today s anthropogenic metabolism. In any case, the longer it takes until measures are introduced for better management of material flows and stocks, the larger the metabolic crisis will be. It is time now to take a close look at the way we manage materials, to start crucial research and careful experiments, and to make a next quantum leap towards a new anthropogenic metabolism which fulfils the needs of both mankind and the environment in the third millennium.

In a companion article in this volume (*see* **Contaminated Lands and Sediments: Chemical Time Bombs?**, Volume 3), William Stigliani draws attention to the impacts that climate change (and other types of global change) might have on the mobilization, and sometimes the immobilization, of anthropogenically produced substances in the environment. Sometimes the changes will take place long after the material had stopped entering the biosphere, and seemed to be safely buried in the sediments of a harbor, or locked up in the urban fabric. This behavior presents a major challenge to simulation modelers to model the long-term behavior of urban metabolism.

See also: **Contaminated Lands and Sediments: Chemical Time Bombs?**, Volume 3; **Industrial and Anthroposystem Metabolism**, Volume 3.

REFERENCES

Azar, C, Holmberg, J, and Lindgren, K (1996) Socio-ecological Indicators for Sustainability, *Ecol. Econ.*, **18**, 89 112.

Baccini, P and Brunner, P H (1991) *Metabolism of the Anthroposphere*, Springer, New York.

Bergback, B (1992) *Industrial Metabolism: the Emerging Landscape of Heavy Metal Emission in Sweden*, PhD Thesis, Linkoping University, Sweden.

Bergback, B, Hedbrant, J, Jonsson, A, Lohm, U, Sviden, J, Sorme, L, and Ostrlund, C (1998) Urban Metabolism: Metals in Stockholm, Case Study Report 2, in *Materials Accounting as a Tool for Decision making in Environmental Policy (MacTEmPo)*, Final Report EC Program Environment and Climate 1994 1998 ENV4 CT96 0230.

Brunner, P H, Daxbeck, H, and Baccini, P (1994) Industrial Metabolism at the Regional and Local Level: A Case study on a Swiss Region, in *Industrial Metabolism: Restructuring for Sustainable Development*, eds R U Ayres and U E Simonis, United Nations, University Press, Tokyo, 163 193.

Brunner, P H, Daxbeck, H, Fehringer, R, Merl, A, and Obernosterer, R (1996) Abfall und Stoffwirtschaft, in *Umwelt und Unternehmen: Erfolgreiches Umweltmanagement. Strategien, Lösungen*, eds P H Brunner, T Gutwinsky, H Kroiss, W List, and J Stiegler, Signum, Vienna.

BUWAL (Bundesamt fur Umwelt, Wald und Landschaft) (1998) *Bewertung in Ökobilanzen mit der Methode der ökologischen Knappheit*, Schriftenreihe Umwelt Nr 297, BUWAL, Bern.

Daxbeck, H and Brunner, P H (1993) Stoffflussanalysen als Grundlagen fur Effizienten Umweltschutz, *Österreichische Wasserwirtschaft*, 45, 90 96.

Duvigneaud, P and Denayeyer-De Smet, S (1975) L Ècosysteme Urbs, L Ècosysteme Urbain Bruxellois, in *Productivité biologique en Belgie*, eds P Duvigneaud et and P Kestemont, Travaux de la Séction Belge du Programme Biologique Internationale, 581 597.

Enquete-Kommission Schutz des Menschen und der Umwelt des Deutschen Bundestages (1994) *Die Industriegesellschaft Gestalten*, Economia, Bonn.

Fehringer, R and Brunner, P H (1996) *Kunststoff üsse und Möglichkeiten der Kunststoffverwertung in Österreich*, Federal Environment Agency, Monographien Band 80, Vienna.

Fehringer, R and Brunner, P H (1997) Flows of Plastics and their Possible Reuse in Austria, in *From Paradigm to Practice of Sustainability*, Proceedings of the ConAccount Workshop, eds S Bringezu, M Fisher-Kowalski, R Kleign, and V Palm, Wuppertal Special 4, Wuppertal Institute and Science Center, Leiden, 272 277.

Fiedler, H J and Rosler, H J (1993) *Spurenelemente in der Umwelt*, 2nd edition, ed G Fischer, Jena, Stuttgart.

Frischknecht, R, Hofstetter, P, Knoepfel, I, Dones, R, and Zollinger, E (1995) *Ökoinventare für Energiesysteme: Grundlagen für den ökologischen Vergleich von Energiesystemen und den Einbezug von Energiesystemene in Ökobilanzen für die Schweiz*, Report, Bundesamt fur Energiewirtschaft, Swiss.

Hendriks, C, Obernosterer, R, Muller, D, Kytzia, S, Baccini, P, and Brunner, P H (2000) Material Flow Analysis: a Tool to Support Environmental Policy Decision Making. Case-studies on the City of Vienna and the Swiss lowlands, *Local Environm.*, 5(3), 311 328.

Greenberg, R R, Zoller, W H, and Gordon, G E (1978) Composition and Size Distribution of Particles Released in Refuse Incineration, *Environ. Sci. Technol.*, 12(5), 566 573.

Krotscheck, C (1995) *Prozessbewertung in der nachhaltigen Wirtschaft*, PhD Thesis, Technical University of Graz, Institute for Process Engineering, Graz.

Meadows, D H, Meadows, D L, Randers, J, and Behrens, W W (1972) *The Limits to Growth: a Report for The Club of Rome's Project on the Predicament of Mankind*, Potomac Associates, Universe Books, New York.

Merian, E, ed (1984) *Metalle in der Umwelt: Verteilung, Analytik und biologische Relevanz*, Verlag Chemie, Weinheim, Deerfield Beach, FL.

Newcombe, K, Kalma, I D, and Aston, A R (1978) The Metabolism of a City, the Case of Hong Kong, *AMBIO*, 7(1), 3 16.

Obernosterer, R W (1994) Fluchtige Halogenkohlenwasserstoffe FCKW, CKW, Halone-Stoffflussanalyse Osterreich, Master Thesis, Vienna University of Technology, Institute for Water Quality and Wastes Management, A-1040 Vienna, Austria.

Obernosterer, R and Brunner, P H (1997) Baurestmassen als zukunftige Hauptquelle fur FCKW in der Abfallwirtschaft, *Müll und Abfall*, 2, 89 95.

Obernosterer, R, Brunner, P H, Daxbeck, H, Gagan, T, Glenck, E, Hendriks, C, Morf, L, Paumann, R, and Reiner, I (1998) Urban Metabolism: the City of Vienna, Case Study Report 1, in *Materials Accounting as a Tool for Decision making in Environmental Policy (MacTEmPo)*, Final Report EC Program Environment and Climate 1994 1998 ENV4 CT96 0230.

Paumann, R, Obernosterer, R, and Brunner, P H (1997) *Wechselwirkung Zwischen Anthropogenem und Natürlichem Stoffhaushalt der Stadt Wien am Beispiel von Kohlenstoff, Stickstoff und Blei*, Report, Vienna University of Technology, Institute for Water Quality and Waste Management, Vienna.

Rees, W E (2001) *Human Carrying Capacity: Living with Global Life Support*, Global Environmental Change, Volume III, John Wiley & Sons, Chichester.

Rechberger, H (1999) Entwicklung Einer Methode zur Bewertung von Stoffbilanzen in der Abfallwirtschaft, PhD Thesis, Vienna University of Technology, Institute for Water Quality and Waste Management, *Wiener Mitteilungen*, 158.

Rittmeyer, C and Vehlow, J (1993) Decomposition of Organohalogen Compounds in Municipal Solid Waste Incineration Plants. Part I: Chlorofluorocarbons, *Chemosphere*, 26, 2129 2138.

Rittmeyer, C, Kaese, P, Vehlow, J, and Vilohr, W (1994) Decomposition of Organohalogen Compounds in Municipal Solid Waste Incineration Plants. Part II: Co-combustion of CFC Containing Polyurethane Foams, *Chemosphere*, 28, 1455 1465.

Schmidt-Bleek, F (1998) Das MIPS-Konzept: Weniger Naturverbrauch mehr Lebensqualitat durch Faktor 10, Droemer Knaur, Munchen.

Schmidt-Bleek, F, Bringezu, S, Hinterberger, F, Liedtke, C, Malley, J, Schutz, H, Stiller, H, Tischner, U, Welfens, M J, Behrenmeier, R, Bruggemann, U, Lehmann, H, Manstein, C, Merten, T, Richard-Elsner, C, and Zieschang, H (1996) *MAIA. Einführung in die Material-Intensitäts-Analyse nach dem MIPS-Konzept*, Report, Division for Material Flows and Structural Change, Wuppertal Institute for Climate, Environment and Energy, Wuppertal.

Schmidt-Bleek, F (1994) *Wieviel Umwelt braucht der Mensch? MIPS: Das Ma für ökologisches Wirtschaften*, Birkhauser, Berlin.

SETAC (Society of Environmental Toxicology and Chemistry) (1996) *Towards a Methodology for Life Cycle Impact Assessment*, SETAC, Brussels.

Settle, D M and Patterson, C C (1980) Lead in Albacore: a Guide to Lead Pollution in Americans, *Science*, 207, 1167 1176.

Stigliani, W M (1991) Chemical Time Bombs, in *IIASA Options*, Laxenburg, Austria.

SUSTAIN-Verein zur Koordination von Forschung und Nachhaltigkeit (1994) *Forschungs- und Entwicklungsbedarf für den*

Übergang zu einer nachhaltigen Wirtschaftsweise in Österreich, Report, Technical University of Graz, Institute for Process Engineering, Austria.

Tauber, C (1988) *Spurenelemenet in Flugaschen: Kohle- Kraftwerk- Umwelt*, TUV Rheinland GmbH, Koln, Germany.

Vehlow, J, Jay, K, Rittmeyer, C, Stieglitz, L, and Vielohr, W (1992) Disposal of Rigid PUR Foams in Municipal Waste Incinerators, in *Alternatives to CFCs and Halons*, International Conference, Berlin, 24 26 February, 1992, organized by BMU, Umweltbundesamt, US EPA, Proceedings, 543 555.

Vehlow, J, Rittmeyer, C, Vogg, H, Mark, F, and Hayen, H (1994) Einflu von Kunststoffen auf die Qualitat der Restmullverbrennung, GVC Symposium Abfallwirtschaft, Wurzburg, 17 19 October, 1994, Preprints 203 219.

Wilson, D G (1975) The Resource Potential of Demolition Debris in the US, *Resour. Recovery Conserv.*, 11, 129 140.

Zimmermann, P, Doka, G, Huber, F, Labhardt, A, and Ménard, M (1996) *Ökoinventare von Entsorgungsprozessen: Grundlagen zur Integration der Entsorgung in Ökobilanzen*, Institut fur Entsorgungstechnik, ETH Zurich, ESU-Reihe 1/96, Zurich.

Industrial and Anthroposystem Metabolism

THOMAS E GRAEDEL AND ROBERT J KLEE

Yale School of Forestry and Environmental Studies, New Haven, CT, USA

Metabolism has a central role in ecological research. The word has two quite distinct meanings. One, which is organism-centered, is "the chemical changes in organisms by which energy is provided for the vital processes and activities, and new material is assimilated for growth and repair or replacement." The second definition is substance-centered: "the sum of the processes by which a particular substance is handled in the organism." Studies of the ways in which organisms use energy and the ways in which substances flow through and are transformed in organisms provide substantial insights into the ways in which organisms receive and utilize the nutrients on which they depend.

If one thinks of an industrial factory or group of facilities as an organism, the biological analogy can be used to define an industrial metabolism in which the flows of nutrients, energy, and water can be analyzed. Alternatively, the flows and transformations of a particular technological substance can constitute a study in industrial metabolism. Similarly, if the organism in question is a human system and its technological components operating over a geographical region (such as a city), studies of the flows of nutrients and energy constitute research in anthroposystem metabolism. Both of these new technological metabolic subjects are areas of active and productive research.

In ecology, metabolic studies have as their goal the understanding of the way in which an organism works. That is also the goal in industrial and anthroposystem metabolic studies, but there is an additional goal as well, i.e., the anticipation that a detailed understanding of anthropogenically-created organisms will generate ideas for improved performance - financial, environmental, or otherwise. In industrial and anthroposystem metabolism, we seek not only to understand the organisms under study more completely, but also to change them.

INDUSTRIAL METABOLISM

Industrial metabolism is the integrated collection of physical processes that convert raw materials and energy into finished products and residues. The basic analytical tool of industrial metabolism study is *materials flow analysis*. This tool works by highlighting the primary drivers of the system the flows of materials and energy. A generic materials flows diagram for today s industry is shown in Figure 1, highlighting the linear nature of traditional industrial processes, where little material is recycled. In contrast, nature s cycles and those of the more advanced industrial analogues tend to create loops and niche organisms to handle wastes and re-route essential nutrients throughout the system. Aside from the simple quantification of stocks and flows of materials, the materials flow analysis can direct the scientist or policy maker towards addressing hazardous materials, reducing excessive raw material inputs, controlling dissipative waste streams, and discovering opportunities to reuse or recycle materials or energy. The proscriptive nature of industrial metabolism studies includes attempts to find ways to modify the current structure so that it closely mimics the biological/ecological ideal of total utility of materials and energy, where residues become feedstocks to other metabolic processes.

Metabolism Surrounding a Firm or Industry

The industrial firm is the smallest basic unit of industrial metabolic study that yields useful system-wide information. Today s total industrial ecosystem is largely a disordered agglomeration of independent firm-sized industrial organisms. (As will be discussed in the following sections, tomorrow s total industrial ecosystem has the potential to resemble instead a modestly organized conglomeration of cooperative industrial organisms.) The materials flows that sustain an individual industrial firm and its products, processes, and facilities can be analyzed with the goal of highlighting wasteful materials and energy practices. For example, Figure 2(a) shows the total metabolic process for the creation of a generic automobile in a generic automobile manufacturing industry or firm in the 1990s. If the material

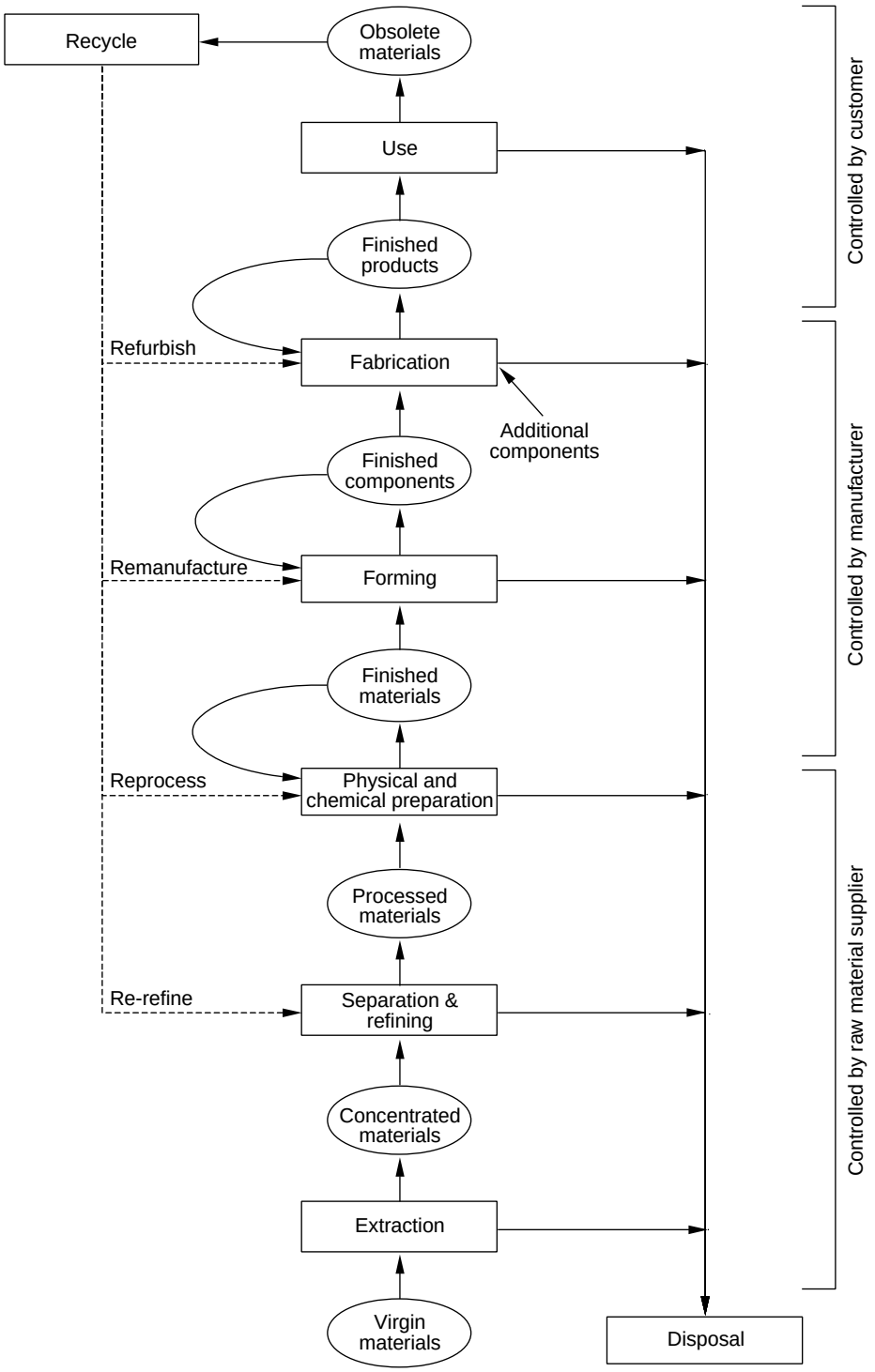


Figure 1 The total industrial metabolism cycle

ows and energy requirements are quanti ed, the rm can attempt to address inef cient system leaks and unrealized materials resources. Hence, Figure 2(b) shows an alternative look at the automobile ow diagram, focusing on the role that a single material in this case copper plays in the creation of a single generic 1990s automobile. An automobile manufacturing rm can use such a metabolic map of copper use per automobile to address points of material loss, as well as the environmental implications of materials choices through various life stages of a rm s product.

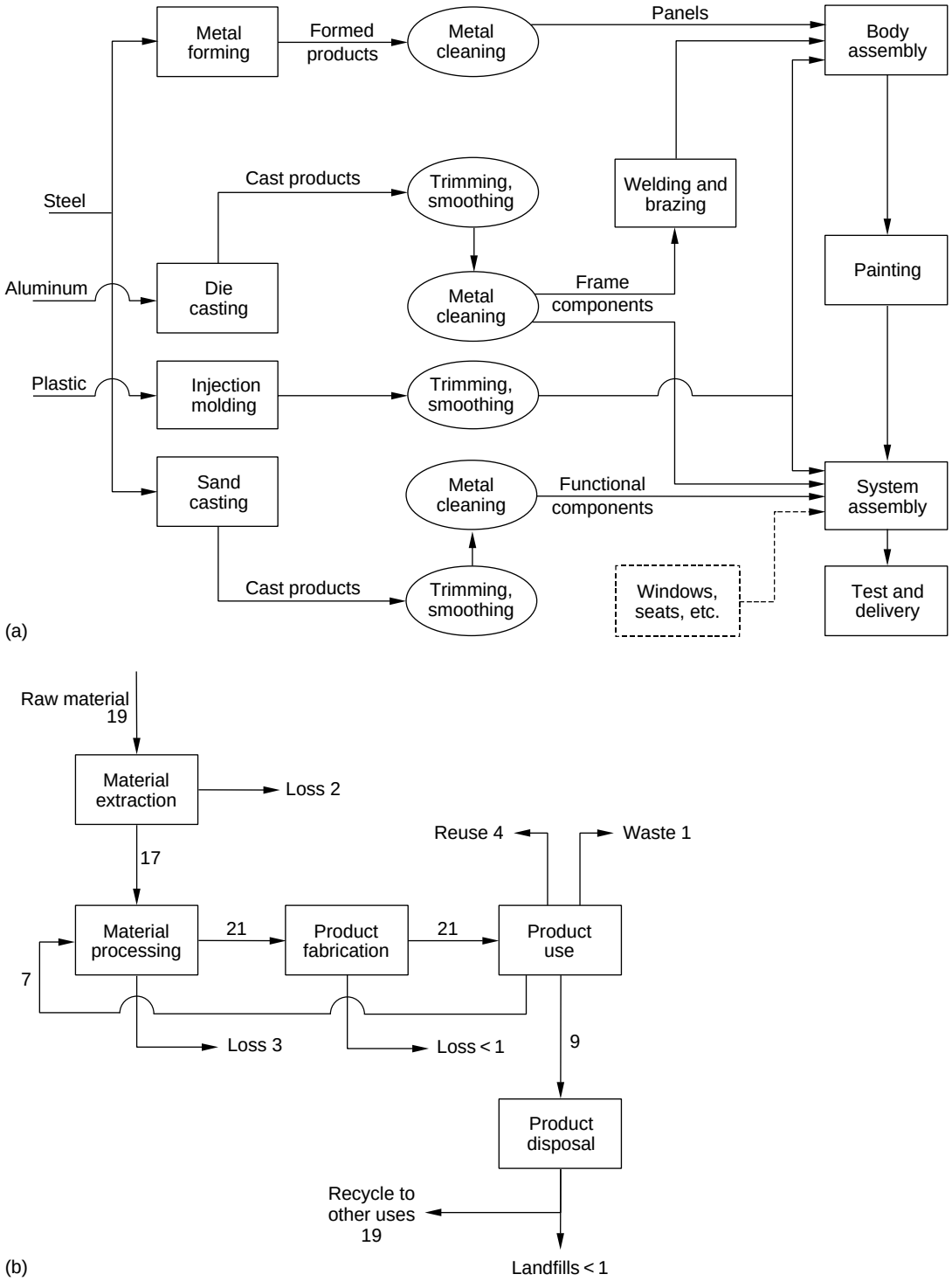


Figure 2 (a) Industrial metabolic processes for the manufacture of a generic automobile in the 1990s. Primary processes appear in rectangles, complementary processes in ovals. The boxes in dashed lines are included to indicate that only the principal metal structural components and their processes are treated, but that a more comprehensive analysis would indicate other components and their processes as well. (Reproduced by permission of Lucent Technologies from Graedel and Allenby, 1998). (b) The flow of copper (kg) in a typical 1990s automobile. (Reproduced by permission of Elsevier Science Limited from Ginley, 1994)

In either the industrial organism or industrial material case, metabolic studies tend to highlight inefficiencies in the industrial ecosystem at the firm or industry level. Firms can recognize under-utilized portions of raw material input that do not end up as part of the final product output, which is an economic and environmental inefficiency. The loss of raw material inputs is especially important when potentially harmful materials (such as lead-based solders, chromium-based plating compounds, and cadmium-based paints, among others) are part of the industrial system. Furthermore, a materials flow diagram can help a firm better understand its reliance on potentially hazardous intermediate process materials (such as aggressive acids or chlorinated solvents) that never end up in final products and almost always pose handling and disposal concerns. Environmental management initiatives can grow out of metabolic studies. For instance, firms might create an inviolate list of materials to assist product and process engineers in keeping hazardous and/or toxic materials out of the industrial metabolism, thereby reducing the potential social and environmental costs from manufacturing, use, and eventual disposal of the product.

Comprehensive metabolic studies would also include materials flows beyond the walls of an individual facility. Long before any processed material enters an individual firm, the extraction of raw materials has had a great potential for environmental harm, from energy use and subsequent air emissions, acid mine drainage, and aggressive chemical purification processes, for instance. Similarly, the use of a product once it is in the hands of a customer may have a great potential for environmental harm from energy use during operation of the product, dispersion of materials, and methods of disposal. Broadening the scope of a firm's vision to consider the upstream activity to support its operations and the downstream ramifications of its products is one of the first key steps in achieving a systems view of industrial activity. Through industrial metabolism analyses, firms can make educated choices that can have large positive influences both upstream and downstream.

In addition to addressing the potential hazards from industrial activity, a metabolic study may uncover a firm's reliance upon materials that in the long run will become increasingly scarce or will be fundamentally unavailable from natural sources. The adequacy of world mineral reserves can be determined by the depletion time of a resource. This measure is an approximation of the number of years before all known global reserves are gone, and is easily calculated by dividing the known reserves of a mineral resource by the rate of consumption of that mineral resource. Reliance on potentially scarce industrial materials is inherently unsustainable in the long term. A few potentially scarce yet commonly used industrial materials include: zinc (19 year depletion time), lead (20 year depletion time), cadmium (27 year depletion time), copper (35 year depletion time), and tin (40 year depletion time) (Kesler, 1994). A firm can address the scarcity issues raised in the metabolic analysis by finding ways to modify products and processes to use more abundant materials. In addition, the firm may uncover underutilized pathways to incorporate pre-consumer or post-consumer recyclable materials into products and processes, thereby addressing both potential scarcity issues and potential environmental hazard issues.

Metabolic results not only highlight potential hazards from a firm's industrial processes, products, and facilities or scarcity issues from materials choices, they may also direct fundamental conceptual shifts in the design process. Designing a product for the environment by excluding ecotoxic compounds or scarce raw materials, or designing a product for eventual recycling, should be considered concurrently with more traditional design considerations such as designing for durability, cost considerations, and ease of manufacture. Modifying the manufacturing process through educated decision making and environmental design has a much greater potential for environmental benefit than do the current end-of-pipe solutions for pollution emanating from industrial facilities.

Metabolism Between Firms

The next logical scale increment of industrial metabolism concerns the potential for creating a system of material and energy interchanges among individual firms. Interfirm industrial metabolic studies usually have a goal of identifying opportunities for reducing wastes and pollution by exploiting low value residues of certain processes (such as heat energy, water, or material) as raw materials for others. This is sometimes called *industrial symbiosis*, a term again borrowed from natural science, in which two or more industrial processes or firms are linked in a mutually beneficial exchange of resources or materials.

Eco–industrial parks (EIP) exemplify industrial symbiosis in action. An EIP may be best defined as a community of manufacturing and service businesses seeking enhanced environmental and economic performance through collaboration in managing environmental and resource issues. By working together, the community of businesses seeks collective environmental benefits greater than the sum of the individual benefits each company would realize if it optimized its individual performance only (such as reduced energy consumption and waste minimization). Economic benefits also flow from these environmental benefits, realized in costs avoided for waste disposal, reductions in raw material purchases, and from shared/centralized materials management services. Not to be overlooked are the social benefits that usually follow from EIP development,

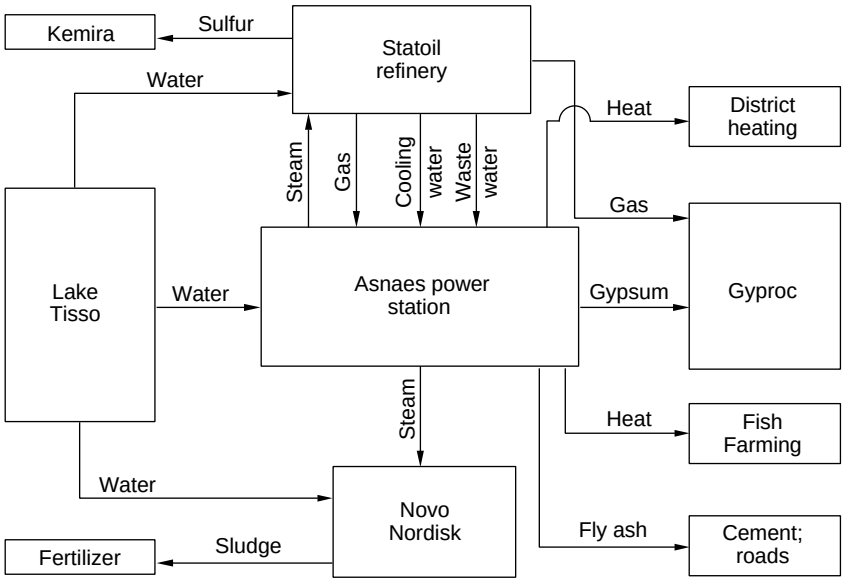


Figure 3 Industrial metabolism governing the town of Kalundborg, Denmark. Kalundborg's industrial network behaves very much like an ecological food web, where industrial organisms consume each other's waste materials and energy, forming bonds of interdependence. The Kalundborg eco–industrial system includes five core partners: the 1500 MW coal-fired Asnaes power station, the Statoil Refinery, the Gyproc plasterboard manufacturer, the Novo Nordisk pharmaceutical company, and the City of Kalundborg. The power station provides waste steam heat to the 20 000 residents of Kalundborg (for home heating) and to the Novo Nordisk pharmaceutical company and Statoil oil refinery for their industrial processes. The power plant uses fresh water for cooling and a small portion of the waste heated water is transferred to nearby aquaculture fish ponds, and uses a portion of surplus refinery gas from Statoil for power generation to supplement coal burning. A local cement company utilizes the de-sulfurized fly ash from the power plant. The power plant influences the sulfur dioxide (SO_2) in the stack gas with calcium carbonate to produce calcium sulfate (gypsum), which is in turn sold to Gyproc. The refinery provides excess gas (that otherwise would be flared off) to Gyproc as a low cost alternative fuel source. The refinery's de-sulfurization operation produces pure liquid sulfur that is sold to a nearby sulfuric acid producer, Kemira. Sludge from the fermenting and other biological processes at Novo Nordisk and from the fish farms are used as fertilizer in nearby farms. Local farmers use surplus yeast from the insulin production as pig food

including fostering a sense of community among businesses and surrounding neighborhoods.

The archetype for the EIP is found in the small town of Kalundborg, Denmark, where bilateral business agreements were developed over 25 years among large industrial firms in the business and social community of Kalundborg. Participants realized economic and environmental benefits through these agreements, and only towards the culmination of the 25 year development did they become cognizant of the industrial symbiosis network they had created. Figure 3 outlines the major flows of materials and energy that drive the system at Kalundborg. The Kalundborg model of industrial development, based on metabolic principles, can serve as a model system for a broad range of new and existing industrial systems to emulate.

The fundamental principle behind inter-firm industrial metabolism is that there is no such thing as waste, there are only under-utilized resources waiting to be exploited by an as yet unidentified industrial firm. For inter-firm industrial metabolism to work, residues must be in a usable form: at appropriate temperatures and pressures, in an economically viable quantity, in a reliable and steady supply, and free of hazardous contaminants. These are the engineering and technical challenges that face future creators of Kalundborg-like systems. Their achievement will require creative thinking and fundamental modifications of the individual firm s industrial metabolism in some of the ways outlined in the previous section.

ANTHROPOSYSTEM METABOLISM

The complexities at the scale of the rm and industrial park have historically obscured the vision of industrial processes and system interactions at regional, national, and global scales. At these scales, the complex interplay of individual rms and industrial systems with human and social constructions can create system problems not readily apparent nor explicitly felt by a single rm. Industrial metabolism, with its focus on products, processes, and facilities, is a term that appropriately addresses the metabolism surrounding individual rms and between rms. However, new thinking is required to directly incorporate human actions within a metabolic framework. The metabolism of a human system and its technological components includes

materials and energy flows, products, industrial processes, and facilities interacting over complex spatial and temporal scales. Although the more common term *urban metabolism* nearly covers all of these interactions, a broader term is more appropriate for the regional, national, and global metabolism, which are not wholly urban. We offer the term *anthroposystem metabolism* to represent the metabolic processes of human technological and human social systems at all spatial scales, broadening the basic principles of urban metabolism to include all of the technical and social constructs that support the modern technologically-related human.

Metabolism of a City or Urban Region

A city is a concentrator of resources and a generator of residues. A generic metabolic image of a city is shown in Figure 4. A typical modern city draws resources from the entire world, including energy, water, food, and a great variety of processed materials such as vehicles, structural steel, and furniture, among many others (*see* **Ecological Footprint**, Volume 3). Its transformation processes include the human food cycle, the combustion of fossil fuel, and the wearing out or gradual obsolescence of manufactured products. Leaving the city are a variety of residues of little perceived value: air pollutants, wastewater, and large, heterogeneous amounts of solid wastes. These various residues are dissipated in a number of ways and over various distance scales, but the dissipation lengths are much shorter than the original acquisition lengths.

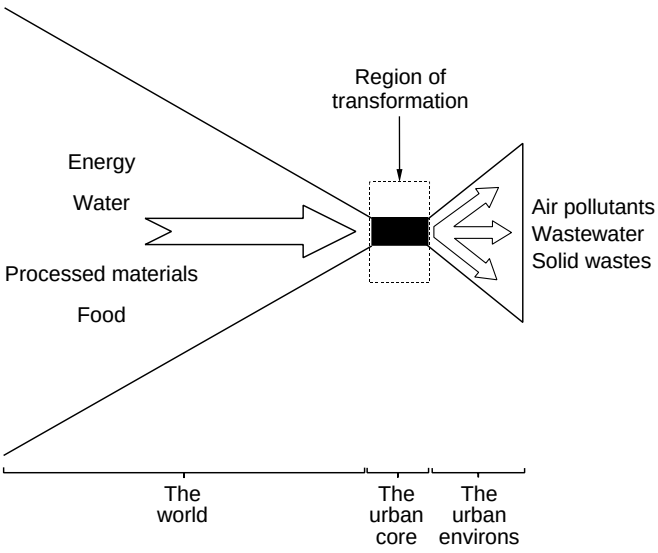


Figure 4 Generalized materials flows of urban metabolism. Materials are transported from throughout the world (a vast spatial scale) into the urban core, are transformed in order to sustain the human population, and are then released as wastes into the urban environs, a significantly smaller region adjacent to the urban core

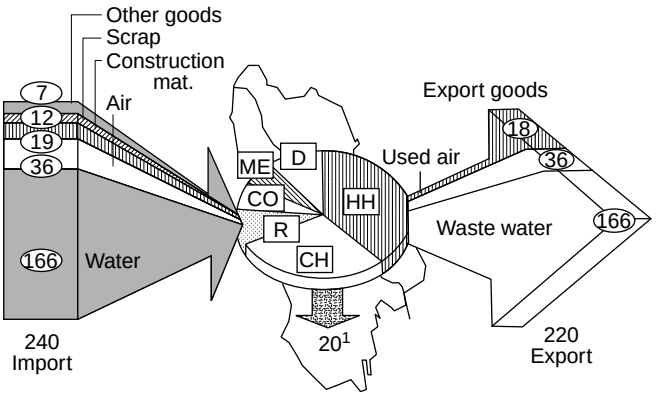


Figure 5 Flux of goods through the urban metabolism of the Bünz Valley, Switzerland, in metric tons capita-year^{-1}. HH, private households; CH, production of chemicals; CO, construction; ME, metal processing; FD, processing of food and drink; R, others. (Reproduced by permission of The United Nations University from Brunner *et al.*, 1994)

An example of an urban metabolic analysis is shown in Figure 5 for the Bunz Valley in Switzerland. Several facets of the analysis deserve comment. The first is the sheer magnitude of the flows: 220 metric tons person^{-1} year^{-1}, on average. Also of note is that the most abundant resource flow, by far, is that of water. The water is not retained; it passes through the system in various ways and leaves as part of the resource outflow. The same is true, though at substantially lower throughput, for air used in combustion processes. The resource flows with more complex fates are those of industrial chemicals and metals; of the 58 tons capita^{-1} year^{-1} that enter the region, 20 (more than a third) do not leave. Rather they are added to the built stock, namely roads, equipment, housing, and so forth (Brunner *et al.*, 1994).

Any environmental impacts attributable to the flows of resources in the urban metabolic system result from the residues. Interest in environmental impacts thus encompasses not only the magnitudes of the residue flows, but also their chemical form and the ecosystems into which they are dissipated. As with industrial metabolism, urban metabolism thus studies reservoirs of materials and the flows among the reservoirs. Unlike the more general industrial metabolism, however, that of urban areas refers to a specific location and thus to that location s logistical and environmental constraints. Where industrial metabolism implies an environmental relationship, urban metabolism explicitly addresses the environmental dimension in space and time.

Metabolic studies of the urban environment will be necessary to solve the frequently encountered problems of failed urban systems, in which the urban environs is under-provided with freshwater, energy, and raw materials, and is overwhelmed by sewage and industrial pollution. Many of the lessons and initiatives used in the industrial

metabolism example described above apply equally to urban metabolism. Most of today s cities were built at a time when resources seemed cheap, abundant, and infinitely available. Now we realize that financial and environmental costs are high, many resources are scarce, and none is infinitely available. Our modern urban areas currently require a high level of materials and energy use to sustain them. Urban metabolic health is largely determined, as with that of any organism, by the adequacy of the supply of sustaining technical nutrients and by the efficient removal of wastes. When the materials movement requirements of urban areas exceed the natural and/or technological capacity, the urban metabolism is inherently unsustainable. Thus, at some critical population density, materials movement systems may be unable to provide a sufficient quality and quantity of raw materials inputs and may be unable to remove the wastes that have the potential to poison the urban organism. Finding ways to turn wastes into useful materials is one notion that could greatly ameliorate this problem.

Land itself has become waste in many urban areas, as a repository for technological materials that leak out of the urban metabolic system. Inefficient metabolic activity (highlighted through materials flow analysis of urban areas) can poison the urban organism through heavy metal pollution, smog, and toxic waste, making these areas unfit for use without costly and time-consuming remediation. In the case of many older cities, established factories and waterfront areas have been abandoned for new industrial sites in the suburbs or in isolated industrial parks, leaving behind unsightly and unusable land. In response, new brownfield regulations allow reclamation and remediation of degraded urban industrial land to levels safe for future industrial activity, even if the soil is not quite fit for children to eat. This recycling of neglected urban sites generates environmental and economic benefits, while solving a problem of decay within the urban organism. As the brownfield story highlights, understanding and locating potential problems through materials flow analysis is the first step towards finding innovative solutions.

Data from a wide array of population studies, environmental science assessments, and human health statistics show that rapidly worsening urban environmental problems may dominate environmental discourse throughout the 21st century. Metabolic studies show the strengths and weaknesses of the structure of the typical urban organism, yet the more pressing question is whether or not the current urban metabolic system can accommodate the inevitable population growth. The data in this regard are truly staggering. In 1950, only 30% of the world s people lived in cities, a global total of 734 million urban inhabitants. By 1995, the world s urban population had more than tripled to 2.5 billion about 45% of the world s population (*see* Urban Population Change, Volume 3).

As is too often the case, the greatest potential for environmental degradation will occur in the poorest countries and in developing countries with the highest economic growth rates. This local environmental pressure will manifest itself in various failures of the metabolic systems that should move materials, energy, and wastes. The symptoms of this metabolic collapse will likely include increased pollution of urban water sources, diminishing supplies of fresh water, increases in numbers of cases of diarrheal diseases, heavy metal loading in urban soils, and general ecosystem damage, among others. Metabolic studies are important first-step tools to help guide urban development policy in answering the difficult questions of how to create a significant number of sustainable new cities, and how to design sustainability into expanding mega-cities. The mega-city, a phenomenon of the late 20th century, has materials and energy requirements at magnitudes well beyond our current rudimentary understanding of how to handle them appropriately. The mega-cities of today already suffer from inefficiencies in moving materials and wastes. The daunting number and complexities of energy and materials pathways in a mega-city have yet to be tackled in any rigorous manner, yet the mega-city is the most likely way in which burgeoning urban populations will be incorporated into the planet s urban fabric. This is the challenge for researchers of the urban metabolism how to sustainably accommodate urban growth in some of the most challenging social, political, and economic environments.

For a study of the metabolism of Vienna, *see* Anthropogenic Metabolism and Environmental Legacies, Volume 3.

Metabolism of a Nation or the Globe

On the national scale, metabolic studies have the potential to show a very broad picture of the materials requirements of the technological society, as well as the individual material responsibility of a person living in our material world. As shown in Figure 6, over 50 kg of material and energy inputs (not including water) are required to satisfy an average American s wants each and every day. At this rate, a person consumes 1600 metric tons over the course of an 80 year life, a mass that would occupy a cube almost 18 m on a side. The lesson from this metabolic analysis is that our current resource use and inevitable waste generation is inherently unsustainable. Furthermore, this analysis also gives a first-order approximation and valuable insights on the level of materials use and waste production to be expected when the entire world strives for a developed world standard of living. Quantification, even if only at order of magnitude accuracy, quickly shows the unsustainable nature of our current industrial metabolism when extrapolated into the future, assuming population

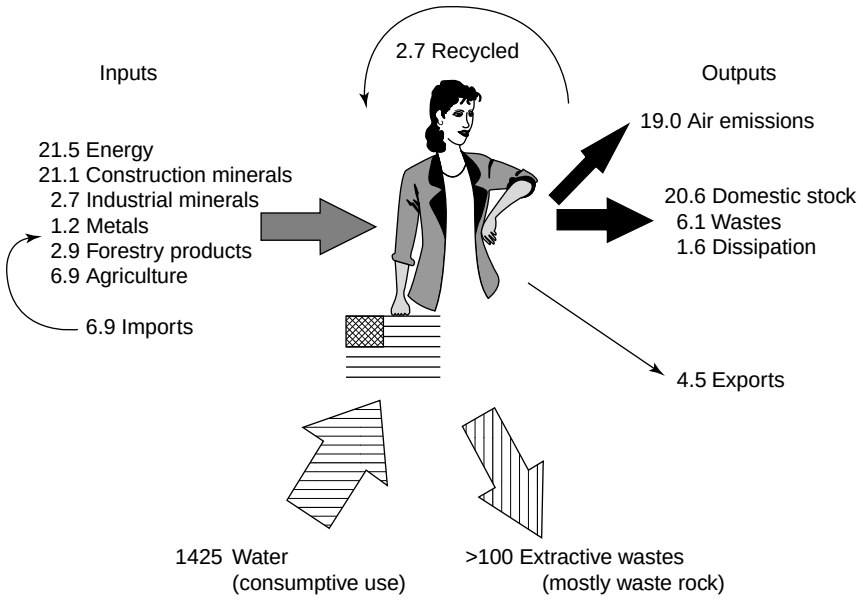


Figure 6 US Material flows, circa 1990, in kg capita-day^{-1}. (Reproduced by permission of Annual Reviews Inc. © 1995 from Wernick and Ausubel, 1995)

growth and greater *per capita* consumption of energy (*see* **Ecological Capacity**, Volume 3).

In addition to understanding the total material movements of a nation, the speci c fate and transport of potentially harmful materials require individual metabolic analyses. One of the most well documented heavy metals released to the environment by industrial processes and products is lead. The dangers of exposure to high levels of lead have been understood for centuries (in fact, acute lead poisoning has been documented for workers at ancient Roman lead smelters in Spain and England). Beginning in the 1970s, epidemiological studies reported adverse effects on child development at extremely low levels of lead exposure. Furthermore, the discovery of lead in ice cores from glaciers in Greenland showed that anthropogenic lead could travel great distances. This led to sweeping reform in industrial use in lead and the eventual removal of lead from gasoline in many countries. Metabolic analyses like that in Figure 7, which pictures the US ows of lead, can help direct researchers and policy makers in controlling releases of potentially harmful substances from a nation s industrial activity. Metabolic study helps differentiate between uses of a substance that are inherently dissipative (such as lead additives in gasoline) and therefore unrecoverable, readily bio-available, and potentially harmful to humans and ecosystems, from those uses of a substance that are inherently concentrated (such as lead in batteries) and therefore nearly in nitely recyclable, not bio-available, and sustainable over the long run with effective control systems.

The sum total of the industrial metabolic activity of all nations comprises the ows that determine the degree to which human activity disrupts global biogeochemical cycles. Just as considerable enlightenment is provided by materials ow analysis at local, regional, and national scales, global budgets and cycles of common industrial materials will yield insights on the sustainability of resource consumption and the dispersion of potential hazards through the world environment at concentrations greater than the natural ows. These materials ow analyses can direct policy and action towards the potential untapped reservoirs of previously exhumed and processed raw materials, e.g., materials in semi-permanent structures (buildings, roads, bridges, and the like), or in industrial and municipal land- lls. Detailed materials ow analyses have the potential to uncover untapped resources that could be recycled back into productive uses, commonly referred to as *mining the landfill*, even though these potential resource stocks are not limited to land lls. Tracking the global stocks and ows of common industrial materials is an area of active and growing research that will be of increased importance as global resources dwindle and the rami cations of systemic disruption of global biogeochemical cycles are better understood.

THE UTILITY OF METABOLIC STUDIES

While it is interesting to think of industrial systems and anthroposystems as organisms and ecosystems, such an activity is more than merely an intellectual exercise. It is rather a framework that permits us to examine the bene ts to be derived from studies of industrial or anthroposystem metabolism. These bene ts may be grouped into three categories, as indicated below.

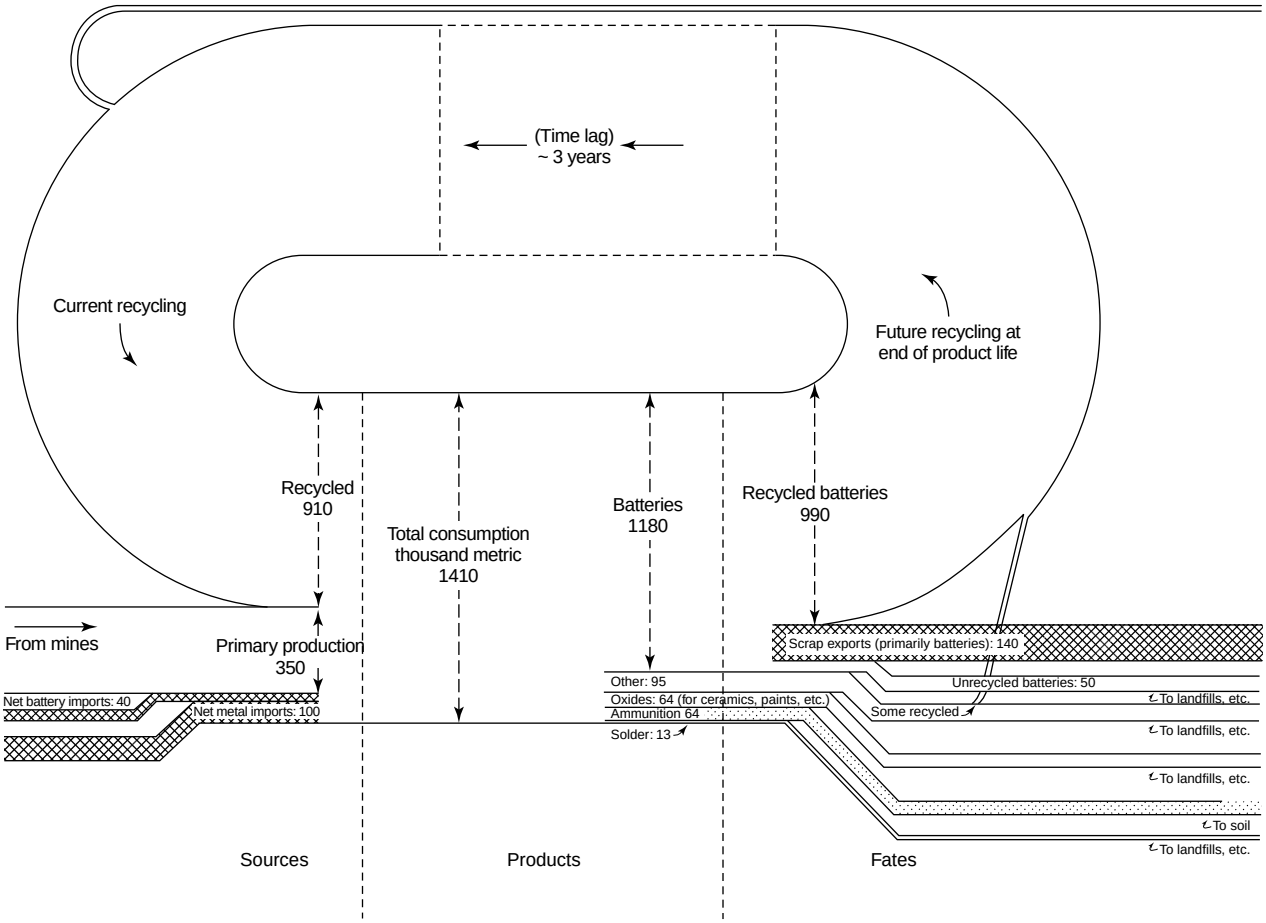


Figure 7 Flows of lead through the US economy in 1993–1994. Data displayed are averages of data for 1993 and 1994, in thousands of metric tons of lead year^{-1}. Other products are bearings, brass, bronze, cable covering, caulking, lead, casting materials, pipes, and sheet lead. Emissions to water or air – the primary health concerns – are not shown because they are too small to be visible at this scale. Also not shown is the roughly 0.6 thousand metric tons of lead used in aviation gasoline (about a billion liters with a lead concentration of 0.4–0.8 g l^{-1}). (Reproduced by permission of Cambridge University Press from Thomas and Spiro, 1994)

Maintaining Human Systems

Industrial and urban complexes require that resources to support them be provided in sufficient magnitude and on a satisfactory time schedule. In the case of small industrial organisms and small urban areas, materials flows are seldom compromised by input or output capacity. As these organisms and ecosystems grow and evolve, limits to efficient connectivity often begin to emerge. In industrial systems, especially those in densely populated urban areas, these limits are often related to infrastructure.

Urban infrastructure roads, power lines, natural gas distribution systems, water supply systems, sewerage, etc. have not traditionally been designed with flexibility and expansion in mind. Furthermore, they are customarily placed out of sight (and out of mind) under streets, along railroad right-of-ways, within building walls so that they are difficult to study and difficult to modernize. A recurring problem in growing cities is that infrastructure is unable to keep pace with growth in population and in industrial and commercial activity. By studying and predicting requirements for material flows, industrial and urban ecologists can aid in new design approaches that have the potential to accommodate evolving needs or desires involving flows of resources of all types.

In a similar vein, industrial and anthroposystem ecologists can evaluate and anticipate limits involving the management of residues. Infrastructure is an issue here as well, but the assimilative capacity of the environments receiving the residues is crucial. Topics for study include the amount and characteristics of wastewater flows, the rate of generation of solid wastes, the disposal of toxic residues, and the like.

It is interesting and unfortunate that we have little in the way of a clear understanding of what the limits might be to resource supply and removal within densely

populated areas. Might we be better able to reuse materials if population density is high? Will high population density be acceptable from a social point of view? From an infrastructure point of view? Can we reasonably hope to make cities more densely populated than Hong Kong (• 2000 people ha• ^{1}), and would we want to?

Industrial and anthroposystem ecologists can respond to these issues not only by suggesting limits to supply and assimilative capacity, but also by working to develop alternative systems for delivering resources, for recycling materials within local metabolic systems rather than exporting them, and by finding ways to reduce flows while maintaining the services those flows now provide. These efforts may immediately determine the feasibility from a materials standpoint of reordering society with the goal of improving long-term sustainability.

Maintaining Environmental Systems

The continued existence of a planet suitable for supporting life is clearly an important thing to think about. It is also clearly a function of the environmental stresses placed upon the planet by human activities. The levels of these stresses now and in the future can be determined by studies of industrial and anthroposystem metabolism, as the following examples show.

The air in the Los Angeles Basin of California was clear and healthy until the middle of the 20th century. In the 1950s, however, a new phenomenon, photochemical smog, began to make the quality of the air progressively worse. Although motor vehicles had been driven around Los Angeles for many years, there were finally enough of them and the tailpipe emissions were voluminous enough that an environmental threshold had been exceeded. Future exceedences of this type may be predictable, and prediction will surely be possible only if the relevant environmental science is understood, the current residue flows have been determined, and future residue flows can be estimated.

A second type of environmental systems benefit from anthroposystem metabolism studies is the potential for determining the relative intensity of sources of residues. In a study of emissions of copper to the Swedish environment (Landner and Lindestrom, 1999), neither industrial activity, nor mining, nor roof runoff proved to be the dominant source of copper, as was anticipated. Surprisingly, residues from automotive brake linings were the source of copper with the highest flow rate to the environment. With this information in hand, Sweden was able to place less emphasis on dealing with unimportant flows of copper to the environment, and more on the crucial ones.

Redesigning Human Systems

Industrial ecology operates from the perspective that an ideal technological society is one in which materials, once extracted from their natural reservoirs, are retained in useful forms as long as possible through reuse and recycling. That is, technological resources should cycle just as do nature s resources. Such an achievement is only possible with considerable effort, i.e., if products are designed to use recycled materials, materials are used in ways easy to recover, and so on.

Industrial ecology further contributes to resource cycling by identifying previously unappreciated or unmeasured reservoirs that may contain reusable material. In a classic study of this type (Kimbrough *et al.*, 1996), David Allen and co-workers, then of the University of California at Los Angeles, studied silver loss to the San Francisco Bay as well as the industrial and societal use of silver in the same area. Two important facts emerged:

- Most of the loss of silver to the Bay occurred during X-ray film developing by small medical and dental offices. Once this silver source was identified, arrangements were made for regular pickup of discarded film developing baths. The result was an improved Bay environment and the recovery of silver for future reuse.
- Approximately 10% of all silver used was retained on developed X-ray film in medical and dental offices. Since the information on those films can now be scanned into digital form and stored electronically, the X-ray film represents a significant stock of silver that can be mined in the future as needed.

Although research such as this has thus far been rare, we can readily imagine that similar work will lead to information on other important but unappreciated reservoirs and on losses important to avoid in order to provide healthy environments.

CONCLUSION

One of the most crucial foundations for understanding the causes and consequences of global environmental change is the ability to anticipate and monitor the path of change. Studies of anthroposystem and industrial metabolism are analytical approaches that fundamentally measure human, industrial, and natural system interactions. Once we know as accurately as possible where we were, where we are, and where we are going, we can begin to make educated and informed decisions concerning how to modify the trajectory of global environmental change. Many of the examples highlighted above show the potential for metabolic analysis when coupled with the concepts of industrial ecology. Industrial ecology requires that industrial and human systems be viewed not in isolation from surrounding systems, but in concert with them, and that industrial and human systems try to emulate the natural ideal of optimized resource cycling. The overriding goal of current and future industrial and anthroposystem metabolism studies

is to ensure our long term survival and prosperity by providing the information we need to make informed decisions.

See also: **Contaminated Lands and Sediments: Chemical Time Bombs?**, Volume 3.

REFERENCES

Ayres, R U and Simonis, U E (1994) *Industrial Metabolism: Restructuring for Sustainable Development*, United Nations University Press, Tokyo.

Brunner, P H, Daxbeck, H, and Baccini, P (1994) Industrial Metabolism at the Regional and Local Level: a Case Study on a Swiss Region, in *Industrial Metabolism: Restructuring for Sustainable Development*, eds R U Ayres and U E Simonis, United Nations University Press, Tokyo, 163 193.

Ginley, D M (1994) Material Flows in the Transport Industry, an Example of Industrial Metabolism, *Resour. Policy*, **20**, 169 181.

Graedel, T E and Allenby, B R (1998) *Industrial Ecology and the Automobile*, Prentice Hall, New Jersey.

Kesler, S E (1994) *Mineral Resources, Economics, and the Environment*, Macmillan, New York.

Kimbrough, D E, Wong, P W, Biscoe, J, and Kim, J (1996) A Critical Review of Photographic and Radiographic Silver Recycling, *J. Solid Waste Technol. Manage.*, **23**(4), 197 207.

Landner, L and Lindestrom, L (1999) *Copper in Society and in the Environment*, Vasteras, Swedish Environmental Research Group, Sweden.

Socolow, R, Andrews, C, Berkhout, F, and Thomas, V (1994) *Industrial Ecology and Global Change*, Cambridge University Press, Cambridge.

Thomas, V and Spiro, T (1994) Emissions and Exposure to Metals: Cadmium and Lead, in *Industrial Ecology and Global Change*, eds R Socolow, C Andrews, F Berkhout, and V Thomas, Cambridge University Press, Cambridge, 297 318.

Wernick, I K and Ausubel, J H (1995) National Materials Flows and the Environment, *Annu. Rev. Energy Environ.*, **20**, 463 492.

Water Use: Future Trends, and Environmental and Social Impacts

BRUCE MITCHELL

University of Waterloo, Waterloo, Canada

By the early 1990s, it was estimated that 22 countries had serious shortages of freshwater, and by the year 2025 it has been calculated that there could be 52 countries with a total population of three billion people facing serious water shortages. The regions most vulnerable to water shortages are South Asia, Africa and the Middle East. In contrast, most areas in North America, Europe and Latin America will be able to meet water needs.

Various factors contribute to growing water demand, with the most important being population growth, increasing urbanization, growing economies which are becoming more industrialized, technological change, climate warming, evolving social tastes and values, and government policies. Furthermore, as initiatives are taken to develop new water supply sources, there can be negative environmental and social impacts. For example, major new dams and reservoirs ood land, which may result in loss of agricultural production or the need to resettle people in urban areas, create habitats which support parasitic and waterborne diseases, and reduce or eliminate indigenous ora and fauna. Social impacts occur due to negative consequences for health, and the potential regional instability and even war that can emerge due to competition for scarce water resources.

In planning to meet future water needs, it is important to consider three types of future - probable, desirable and feasible. Too often, overemphasis is placed on trying to identify the most probable or likely future, and not enough attention is given to considering what might be the most desirable future. Sustainable development is one choice for a desirable future, and its achievement will require a strategy, which includes both provision of new water supply sources and changes in human patterns of water use.

DEFINITION OF WATER USE

Before being able to consider temporal and regional patterns, and future trends, of water use, it is necessary to be clear about the nature of water use. As Gleick (1996) has commented that water use can include many different ideas and it is often misleading and confusing . In his view, water use can be either withdrawal (intake) of water, gross water use (intake plus recirculation plus reuse), or the consumptive use of water. Other distinctions have been made. Kindler (1992) identified three different types of water use: (1) on-site; (2) intake; and (3) instream or flow. On-site use includes activities which deplete water supplies before they reach surface or groundwater systems. This category includes water taken by wetlands, natural vegetation, rain-fed crops, wildlife and evaporation. Intake uses involve situations in which water is physically removed from a supply source for domestic, municipal, industrial or agricultural purposes. Some of the water is returned to the natural system after use, but usually in a modified state. Instream or flow use is represented by water used for generation of hydroelectricity, navigation, fish and other aquatic life, and recreational activity by humans. This category of use usually is not consumptive and the water is subsequently available for other uses.

Kindler (1992) also distinguished between amounts of water withdrawn and consumed. Water becomes consumed when it is incorporated into a product, such as a beverage, or when it is lost to the atmosphere through evapotranspiration. Consumptive uses can sometimes result in the water becoming available for subsequent use, but not always at or near the place from which it was originally taken. Another distinction is made by Ayibotele (1992) when he refers to consumptive and non-consumptive uses (hydro power, navigation, fisheries, recreation).

NEED FOR WATER

It has been suggested that a country or region experiences water stress when supplies are less than

17 000 m^3 person^{-1} year^{-1}. To place this latter figure in perspective, Serageldin (1995) reported that the global average annual supply of renewable freshwater in the early 1990s was approximately 7400 m^3 person^{-1} year^{-1}. This average figure, however, masks the reality that in the early 1990s some 22 countries had renewable water resources of less than 1000 m^3 person^{-1} year^{-1}. The greatest shortages occurred in the Middle East and North Africa, while Latin America was the region with the greatest *per capita* abundance of water. Looking to the future, Serageldin (1995) stated that by 2025 up to 52 countries with a total population of 3 billion people could experience water stress or chronic water scarcity, relative to the threshold noted above.

Another study of 118 countries, completed by the International Water Management Institute, concluded that 1 billion people in northern China, western and southern India, Libya, Saudi Arabia, United Arabic Emirates, Kuwait, Oman, Jordan, Yemen, Israel, Egypt, Tunisia, Iraq, Iran, Syria, Afghanistan, Pakistan, South Africa and Singapore could be expected to face absolute water scarcity by the year 2025 (Seckler *et al.*, 1999). When those countries were combined with other nations expected to experience economic water scarcity , Seckler and his colleagues concluded that some 26% of the 1990 world population would experience severe water shortages. In contrast, countries in North America and Europe were estimated to be in the best position to meet their water needs by 2025.

Gleick (1996) has suggested that attention must be given both to the quantity and quality of water relative to four basic needs: (1) drinking water; (2) sanitation; (3) human hygiene; and (4) preparing food. Further details about these different basic water use needs are presented below.

Drinking Water

Notwithstanding that water needs for survival by humans are influenced by physical conditions and physiological characteristics, the variation in need is not large. Most humans begin to experience thirst after losing only 1% of bodily fluid, and are in danger of death when fluid loss approaches 10%. The minimum water requirements to replace lost fluid for a person in average temperate climate conditions have been calculated to be about 3 liters day^{-1}. If climate and activity levels are different, the daily water requirements change. For example, in a hot climate, it has been calculated that a 70 kg human will perspire between 4 and 6 liters day^{-1} without any significant change in food intake or activity.

Based on a review of estimates by a variety of research organizations, Gleick (1996) concluded that minimum water requirements for a human to survive in a typical temperate climate and involved in normal activities is 3 liters day^{-1}. In tropical or subtropical conditions, this minimum value is increased to about 5 liters person^{-1} day^{-1}, which also can be expressed as just under 2 m^3 person^{-1} year^{-1}.

Sanitation

Humans health usually requires access to safe potable water for consumption, as well as a sanitary way to dispose of human wastes. As discussed in the article in Volume 4 focused on the International Drinking Water Supply and Sanitation Decade (IDWSSD) (*see* IDWSSD (International Drinking Water Supply and Sanitation Decade), Volume 4), many people in the world, especially in developing countries, do not have access to either safe water supply or sanitation systems.

Technologies do exist which allow disposal of waste without use of any water, as shown in Table 1. On the other extreme, the most inefficient toilets may use more than 75 liters person^{-1} day^{-1}. The sanitation technology selected is a function of the development goals and capacity of a country or region, available water, and key cultural and social values. Given that sanitation technology does exist which uses no water at all, and the context-specific nature of water requirements for sanitation, it is difficult to establish one number regarding a minimum amount to meet basic sanitation needs. Technically, it is feasible to set the minimum amount at 0. However, as Gleick (1996) noted, several aspects suggest a number of 0 is not appropriate. First, additional benefits to health can be documented when up to 20 liters person^{-1} day^{-1} of clean water are available. Second, if economic factors are not prohibitive, cultural and social preferences favor water-based sanitation systems. Accessibility to some water for sanitation, along with accompanying education related to water use, can significantly reduce the incidence of diseases. As a result of these considerations, Gleick (1996) recommended that a minimum of 20 liters person^{-1} day^{-1} be used to reflect the benefits of combining waste disposal and associated hygiene, and to accommodate dominant cultural and societal preferences.

Table 1 Sanitation technologies which require no water

VIP
ROEC
Ventilated improved double-pit latrines
DVC
Continuous composting

DVC, Double-vault composting toilets; ROEC, Reed odorless earth closets; VIP, Ventilated improved pits.
Source: Gleick (1996).

Water for Bathing

Beyond direct sanitation requirements related to disposal of human wastes, additional water is used for bathing (showers, baths). Research in North America and Europe indicates that average water use for bathing is approximately 70 liters person^{-1} day^{-1}, with the range extending from 40 100 liters person^{-1} day^{-1}. Comparable data from developing countries are not generally available or are not reliable. From a review of various studies, Gleick (1996) suggested that the minimum water needed for bathing is in the range of 5 15 liters person^{-1} day^{-1}, while that needed for showering is 15 to 25 liters person^{-1} day^{-1}. He has recommended that 15 liters person^{-1} day^{-1} be used as a basic level of service for bathing.

Water for Food Preparation

Studies of residential water use usually have not separated the water needed for cooking or food preparation. However, a review of work in developed countries has suggested that 10 50 liters of water are used for each person per day for food preparation, and that the average use is about 30 liters person^{-1} day^{-1}. Studies that have considered food preparation needs in both developed and developing countries have concluded that an average of 10 20 liters person^{-1} day^{-1} satisfies most regional needs. On that basis, Gleick (1996) concluded that 10 liters person^{-1} day^{-1} would meet most basic needs.

Water for Biophysical Ecosystems

The above discussion of basic needs is anthropocentric in that it focuses only on human requirements. However, other living species, which share the Earth with humans have basic water needs, especially fauna and flora associated with aquatic systems. Minimum flow requirements to meet the needs of natural ecosystem functions and processes are difficult to estimate, although attempts are being made, particularly in North America, Europe and Australia (Born *et al.*, 1998). For example, in 1999, a dam was removed from a New England river in order to allow the flow regime in a river system to function to meet the needs of fish in that aquatic system. In California, under various federal and state statutes, almost 30 000 million m^3 of annual runoff in river systems, or 28% of the total annual average runoff in the state, have been designated for environmental purposes, including fish, waterfowl, and wetland systems. While no specific number can be identified for all situations, it is important that those responsible for making decisions about water use needs consider the requirements of non-human users of ecosystems.

Table 2 Recommended basic water requirements for human needs[a]

Purpose (liters per person per day)	Recommended minimum (liters per person per day)	Range
Drinking water[b]	5	2 to 5
Sanitation services[c]	20 to over 75	
Bathing	15	5 to 70[d]
Cooking and kitchen	10	10 to 50[d]
Total recommended basic		
Water requirement	50	

[a] Excludes water required to grow food.
[b] This number is a true minimum required to sustain human life in moderate climate conditions and with average activity.
[c] An average (not minimum) of 40 liters person^{-1} day^{-1} is estimated to be adequate for direct sanitation hookups in industrialized countries. The upper end reflects very inefficient toilets. In regions in which water is extremely scarce, technology is available for sanitation systems, which use no water, but these are not usually socially accepted.
[d] The higher numbers in the range represent social preferences for moderately industrialized countries. Use in some water-abundant regions may exceed these amounts. The lowest values represent minimum uses in developing countries.
Source: Gleick (1996).

Summary Regarding Basic Water Needs

Following his analysis of basic water needs, Gleick (1996) concluded that international organizations, as well as national and local governments and water suppliers, should adopt a basic water requirement standard for human needs of 50 liters^{-1} person day^{-1}, as shown in Table 2. In his view, unless this basic need is met, large-scale human misery and suffering will continue and grow in the future, contributing to the risk of social and military conflict (Gleick, 1996). The information about minimum water requirements provides the context for the information and narrative provided in this article.

PATTERNS AND TRENDS OF WATER USE

Drivers of Use

Various factors drive water use. Key drivers are population, urbanization, nature of the economy, technology, climate, social values and tastes, and policy decisions, especially pricing. Estimates have to be made regarding each of these, before being able to make estimates about future water use patterns and trends. The uncertainty associated with the calculations for each of these drivers becomes compounded when such estimates become the inputs for calculations of future water use, emphasizing how easy it is for forecasts of water use patterns to be inaccurate.

Population

One of the most important drivers regarding future water use patterns is population growth. Ideally, to identify future trends of water use, information is required regarding the size and composition of the population, including data about size as well as age, sex, fertility rates, migration and death rates. Given the difficulty in assembling such data, attention often focuses only on the size of future populations.

The United Nations (UN) (1991) projected that the growth rates in percent per year between the years 2000 and 2025 could be expected to be 1.54% at a global scale, resulting in an overall population of 8.5 billion, compared to the anticipated 6.3 billion by 2000. In 1990, the developed countries had just under 23% of the global population, with slightly more than 77% in the developing countries. By 2025, if the population forecasts are correct, those percentages will be 16 and 84%, respectively (*see* Global Population Trends, Volume 3).

The implications for water use patterns are significant. As standards of living rise in the developing countries, *per capita* water use will steadily climb upwards. Furthermore, as the economies in developing countries become more industrialized, higher rates of water use can be anticipated. And, finally, with more people, there will be increased pressure to increase food production, which will lead to more initiatives to develop irrigation agriculture, even if that occurs on land not ideally suited to irrigation.

Urbanization

Not only will there be a steady overall population increase continuing into the 21st century, there also will be an increasing proportion of people living in cities (Stren *et al.*, 1992). Niemczynowicz (1996) stated that the world s urban population increased 2.5 times between 1960 and the early 1990s. By the mid 1990s, about half of the world s people lived in cities; by the year 2000, twice as many people were expected to be living in cities compared to rural areas. Furthermore, the number of megacities (cities with more than 10 million people) were expected to increase to 23 by the year 2000, 80% of which would be in developing countries. By the year 2000, Tokyo was expected to have about 30 million people, followed by Mexico City with just under 26 million, Sao Paulo with 24 million and New York City with 18.5 million. In terms of cities with more than 5 million inhabitants, it was expected that by 2000 there would be 60 (*see* Urban Population Change, Volume 3).

Niemczynowicz (1996) noted that the drinking water demand in megacities was growing rapidly, even though at a global scale, urban populations accounted for only about 15% of total water consumption. Much of the most rapid urbanization in developing countries has occurred in squatter areas, without any control or investment in supporting urban infrastructure. To illustrate, he indicated that 80% of the residents of Addis Ababa and 70% of the inhabitants of Calcutta lived in squatter areas and were not provided with access to safe potable water or sanitation facilities. When water supply systems were in place, they often were inefficient, with significant amounts of water lost to leakage. According to Niemczynowicz (1996), water leakage accounted for losses of 33% in Mumbai (Bombay), 36% in Chennai (Madras), 37% in Seoul, 49% in Bangkok, 50% in Cairo and 51% in Manila. Corresponding numbers for cities in developed countries were 12% in Munich, an average of 18% in Canadian cities, and 19% in Stockholm. Thus, increasing urbanization can be expected to contribute to greater pressure on water supply systems to meet basic needs for water use.

Nature of Economy

The evolution of an economy in a country or a region is significant for future patterns of water use. Two major implications arise. First, direct impacts occur as a result of water demands from different uses. Second, indirect impacts occur as a result of growing regional income and higher standards of living, which usually result in higher levels of water use. Thus, it is important to be able to estimate both the size and shape of future regional economies (Landsberg, 1964). Indicators or measures often used are gross national product, individual spending, government investment, and private sector investment.

Estimating the nature of a future economy is fraught with difficulties. Calculations are required regarding the number and mix of industries (manufacturing, service, etc.), the future demand for goods and services, the comparative advantage of a region relative to competing regions, and effects of changes in public policies on the economy. Water is usually a minor cost relative to total production costs, but when real water scarcity occurs, it can be a significant constraint on economic development.

Technology

The role of technology is always difficult to factor in, since it may increase or decrease water use. For example, Sewell and Bower (1968) suggested that technological change can affect products as well as processes used, raw materials involved as inputs, and methods for handling water. Landsberg (1964) highlighted these considerations when noting that industry uses water in many ways. Water may be an ingredient in some products, be used to cleanse materials during production, be a medium for transporting material within a plant and be an agent for many chemical processes. However, by far the greatest proportion of water withdrawn for industrial use is for cooling, accounting for more than 80% of all water consumed in petroleum refining, chemicals and non-ferrous metals, about 70% in the iron

Box 1 Uncertainty created by technology

The large amounts of water used by industry for heating and cooling are at the heart of the problem of estimating industrial demand. The many technical substitution processes available for water conservation in industry all have different costs and performance characteristics (Gupta, 1992).

and steel industry, and 50% in the food products industry (see Box 1).

To reduce water use, industry has adopted many approaches to handle cooling processes. In some instances, air cooling and/or brackish or seawater has been substituted for fresh water. As Landsberg (1964) remarked, Revolutionary technological and scienti c advances could drastically affect the future demand outlook. Conceivably, a type of air cooling that uses no water whatsoever might be substituted for conventional cooling processes . Changing technology relative to irrigation practices also could signi cantly reduce the amount of water required to support agricultural production in many parts of the world. In a similar way, a major innovation in design of toilets could signi cantly reduce domestic water use.

Thus, to estimate future patterns of water use, it is desirable to be able to incorporate the likely effects of technological change on water use. Needless to say, this is not an easy task, since it is not always possible to know which technological innovations will occur, nor which ones will lead to either reductions or increases in water use.

Climate Warming

Scienti c agreement is growing that climate warming is occurring and that it is likely to continue well into the next century. High-latitude countries will experience more warming compared to low-latitude countries. However, even within regions or countries with a large spatial extent, there can be signi cant variations in the implications for hydrological systems. For example, Arnell (1999) prepared four scenarios to the 2050s related to the effect of climate change for hydrological conditions in Europe. He concluded that there would be general reduction in runoff in southern Europe (south of latitude 50°) but increased runoff in northern Europe. Overall, lower summer ows could be expected along with increased drought conditions in western Europe and decreased drought conditions in eastern and northern Europe. This pattern was con rmed by Mimikou *et al.* (1999) in their analysis of expected conditions in northern Greece, and they calculated that reductions in mean annual and summer runoff, along with prolonged droughts, would have negative implications for production of hydroelectricity and agricultural crops. Timmerman and White (1997) identi ed other complications for coastal cities, vulnerable to sea level rise due to melting of glaciers driven by higher temperatures. They argued that, in general, coastal cities would be at risk due to sea level rises through increased likelihood of ooding and impeded drainage, salinization of freshwater supplies, higher water tables leading to reduced safety of building foundations, and increased beach erosion.

Regarding the implications of climate change for water resources in the Great Lakes basin of North America, Clamen (1990) highlights the dif culties in forecasting possible implications of changed water regimes due to climate change because of the different ways in which variables are expected to be affected. Thus, overland evapotranspiration, lake evaporation and overlake precipitation were expected to *increase*, soil moisture and ice cover were expected to *decrease*, and snowmelt, runoff, net basin supply and lake levels were expected to *decrease but with an earlier spring peak* (*see* **Water Resources: Great Lakes Case Study**, Volume 3).

What can be concluded is that some regions will have less natural water available to draw upon to meet their needs due to reduced runoff and increased evapotranspiration. Where this condition occurs in a region in which natural supplies are or will become limited relative to demand, then climate change will contribute to greater pressure and stress.

Social Values and Tastes

As Sewell and Bower (1968) have observed, social values and tastes can exert an important in uence on water use and demand, and can either increase or decrease water use. Growing af uence in developed countries often increases water demands, as people are able to give more attention to cleanliness and personal hygiene. As people include one or more daily showers as part of their normal routine and as households add washing machines and dishwashers, water use rises. When households add private swimming pools on their properties, water use increases, as it also does through excessive watering of ower gardens and lawns.

On the other hand, once a stage is reached in which basic needs for water are satis ed, social values and tastes can also shift in a way to reduce water demand. As people become more environmentally sensitive, some will choose to replace lawns and exotic ower gardens with native fauna, which require less water. Others add rockeries or other landscaping features, which require no water.

As with technological changes, it is not easy to anticipate what shifts may occur in social tastes and values, and what their consequences may be for water use. However, any estimates of future water use patterns need to be sensitive to the possible rami cations of changing social values.

Policies

Governments can in uence water use patterns through policies related to pricing for water, metering, and water quality

standards. In many countries, consumers do not pay the full cost for water services, and as a result there is scope to achieve more efficiencies in water use if charging schemes were to be introduced so that consumers paid the total cost for the water received. It is usually believed that full-cost pricing would lead to more careful use by consumers. At the same time, it also is recognized that, until basic needs are met, introduction of full-cost pricing could make access to safe, potable water beyond the reach of the poorest and most vulnerable groups in a society. As a result, issues of effectiveness, efficiency and equity have to balanced in deciding which policies are most appropriate relative to water use.

Sectoral Uses

At a global scale, the largest use of water is for irrigation. Estimates vary, but there is general agreement that irrigation represents almost 70% of water use. Use by industry is second, at about 23%, with domestic and municipal use a distant third at 7%. Over 40% of all withdrawn water is returned to the hydrological system in some form of wastewater. Industry is the source of most of the returned water, since almost 90% of water withdrawn by industry is converted into wastewater. Attention is given to these three uses in this article, since these uses each change the nature of any water returned to the hydrological system for subsequent use by others.

However, it is important to recognize that demands or uses arise from non-consumptive users, such as hydro power generation, navigation and fisheries. The global capacity of all hydro power generating plants in 1970 was 290 gigawatts (GW), which represented 24% of the capacity of all electricity generating plants (conventional thermal, nuclear, hydro electricity, other). By 1984 the capacity was 542 GW, but in terms of total global capacity for electricity generation its share had fallen to 23% due to other types of generating stations being built. By 1988, it was estimated that hydroelectricity generated 7% of total global energy. In contrast, freshwater fish catches grew by a factor of 6 between 1950 and the early 1990s.

CASE STUDY: AGRICULTURE

Temporal Patterns and Trends

Irrigation water use has increased steadily. Rydzewski and bin Abdullah (1992) reported that irrigation agriculture increased from about 8 million hectares (ha) in 1800 to 48 million ha in 1900, and to 94 million ha by 1950. By the early 1990s, it was estimated that irrigation agriculture had expanded to cover 260 million ha, with two thirds of that area in developing countries, and could be expected to continue to increase but at a slower rate because most land suitable for irrigation had already been developed (*see* Irrigation: Environmental Impacts, Volume 3).

Box 2 Reasons for the decline in irrigation investment in Southeast Asia

Contributing factors include the large public and foreign debt loads carried by most of the agriculturally-based economies in the region; the declining share of unexploited irrigation development potential in many countries in the region; and political resistance from environmental interests and those displaced or otherwise negatively affected by irrigation development. However, recent studies show that the main reasons for declining investment are the decline in world rice prices and the increasing real costs per ha of new irrigation development (Svendsen and Rosegrant, 1994).

Given different spatial patterns of irrigation agriculture, it is a challenge to make global generalizations about future water use patterns for irrigation. The experience in Southeast Asia is used here to illustrate possible patterns of future use. Svendsen and Rosegrant (1994) reported that irrigated areas in Asia consisted of about 143 million ha by the early 1990s, which represented about 55% of the total in the world. More than one third of all cropland in Asia was under irrigation by the early 1990s. The steady growth in irrigation agriculture in the region had been due to development of high-yield rice varieties or rice which required heavy inputs of water and agrochemicals to realize their potential. However, in the last decade the rate of creation of new irrigation projects has diminished.

Many reasons accounted for the decline in the establishment of new irrigation areas or expansion of existing projects (Box 2). Certainly one of the key ones was the decline in the price of rice after 1960. By the mid-1990s, the real price for rice was about one third of what it had been in the early 1960s. Thus rice became both more abundant and accessible to lower income groups than at any previous time, and that new irrigation projects to produce more rice became increasingly difficult to justify.

In terms of future patterns of water use for irrigation, Svendsen and Rosegrant (1994) suggested that four factors technology, economic and political liberalization, competition for water and environmental and health concerns would be the most influential. Regarding technology, they believed that adoption of new irrigation technologies in developing countries has been much slower than expected, given the remarkable advances made in such technology in the West over the past 30 years. The explanation, in their view, was that most of the technological advances have occurred with regard to water application rather than conveyance and distribution of water. In Southeast Asia, the normal farm is much smaller than Western farms, but the irrigation development projects have been much larger. As a result, in Southeast Asia, e.g., there

has been little adoption of sprinkler and drip technologies, the focus of research and innovation in the West. Until technological innovations occur with regard to water conveyance, distribution and allocation, there is likely to be little impact of technological change on reducing water use patterns in Southeast Asia. The potential water savings from use of improved technology are considerable, given that Rydzewski and bin Abdullah (1992) have reported that typical efficiency figures for large projects in Asia are 65% for conveyance, 75% for distribution, and 60% for field application, resulting in an overall efficiency of just under 30%. These (in)efficiencies indicate that about 2500 tn of water may have to be diverted into an irrigation system for each tonne of cereal produced.

Economic Liberalization

The factor labeled as *economic liberalization* has been important, and perhaps will be the most important, in influencing development of new irrigation facilities, and therefore future water use needs. Economic liberalization is characterized by market-oriented economic policies, realistic exchange rates, facilitation of international trade, a central role for the private sector in producing goods and services, reduction of subsidies and transparent economic policy instruments. Specific modifications have been made to restrictive policies protecting local producers, keeping domestic prices above world prices or maintaining overvalued exchange rates. Such changes make some industries less competitive at a global scale, relative to agriculture in some nations in Southeast Asia and could create pressure to expand irrigation agriculture and water-intensive manufacturing while internationally uncompetitive industries close or downsize. At the same time, however, given that agricultural exports will have to compete with similar products from other countries, there will be increased incentive to use more water-efficient irrigation in order to reduce production costs, especially if charges for water use are increased under economic liberalization policies. This latter factor could help to reduce the demand for irrigation water use.

Competition for Water

Competition for water can be expected to increase, given that water is a finite resource and that for many purposes no real alternative substitutes exist. Some uses, such as agriculture, will have to compete for water against other higher value commodities, which may make it difficult for the agricultural sector to obtain rights to water needed to support expanded irrigation crop land. In addition, in countries in which priorities are established for water uses, domestic and municipal water use normally is given highest priority, and industrial use second priority, with agricultural water use having even lower priority. In the words of Svendsen and Rosegrant (1994), this situation is likely to result in irrigation agriculture coming under increasing pressure to reduce its consumption and to pay the economic scarcity cost for water rather than the currently highly subsidized rates . When agriculture cannot pay the real cost for water, its rights or allocation may be transferred to other users, as has happened in the American Southwest where higher value users are receiving water once used by agriculture.

Box 3 Predictions regarding irrigation and industrial water use in the future

The use of water for irrigation is bound to grow, but at rates below those experienced in the past century, as most land suitable for irrigation is already in use as such and stricter water pricing will tend to increase water efficiency in agriculture. Industrial water use for processing, cleaning, cooling and removal of wastes, will increase in the next few decades as the economies of developing countries are making the transition to industrialization (Traoré, 1992).

Environmental and Health Concerns

Environmental and health concerns were identified as the fourth force influencing water use patterns. Salinization of irrigated land in arid regions, combined with chemical contamination of surface and groundwater, and rising water tables, are all environmental impacts, which can be associated with irrigation projects. A rising groundwater table can be either positive if it reduces extraction costs for groundwater, and negative if it contributes to secondary salinization of crop land when the groundwater is saline. Many health concerns can emerge, ranging from an increase in waterborne diseases to chemical contamination of domestic water supplies. In addition, the construction of major capital works (dams, reservoirs) to provide water to meet irrigation needs can modify downstream flow regimes, inundate productive cropland, homes and settlements, and destroy or significantly modify habitat which in turn affects biodiversity.

The end result is likely to be that, while irrigation agriculture will remain the single most important water use on a global scale, its importance can be expected to decline relative to industrial, municipal and domestic uses (see Box 3).

In contrast, industrial use of water can be expected to increase, in absolute and relative terms, especially as developing countries industrialize (Table 3). However, as with agriculture, improved water efficiency technologies have the potential to dampen the effects of increasing water use as a result of urbanization and industrialization. Gupta (1992) noted that estimates reveal up to 20% of the initial capital for manufacturing plants in Nigeria has

Table 3 Temporal patterns of water demand (km^3 $year^{-1}$), 1900–2000[a]

Water users	1900	1940	1950	1960	1970	1980	1990	2000
Irrigated area (Mha)	47.3	75.8	101	142	173	217	272	347
Agriculture								
A	525	893	1130	1550	1850	2290	2680 (68.9)	3250 (62.6)
B	409	679	859	1180	1400	1730	2050 (88.7)	2500 (86.2)
Industry								
A	37.2	124	178	330	540	710	973 (21.4)	1260 (24.7)
B	3.5	9.7	14.5	24.9	38.0	61.9	88.5 (3.1)	117 (4.0)
Municipal supply								
A	16.1	36.3	52.0	82.9	130	200	300 (6.1)	441 (8.5)
B	4.0	9.0	14	20.3	29.2	41.1	52.4 (2.1)	64.5 (2.2)
Reservoirs								
A	0.3	3.7	6.5	23.0	66.0	120	170 (3.6)	220 (4.2)
B	0.3	3.7	6.5	23.0	66.0	120	170 (6.1)	220 (7.6)
Total								
A	579	1060	1360	1990	2590	3320	4130 (100)	5190 (100)
B	417	701	894	1250	1540	1950	2360 (100)	2900 (100)

Source: Ayibotele (1992).
[a] A, total water consumption; B, irretrievable water losses. Percentage figures in parentheses.

been required to compensate for inadequate supply of urban power and water in that country. Technological improvements could assist to both reduce the amount of water used by the industries and to improve the delivery of water within urban water supply systems.

Municipal and domestic use also can be expected to increase, due to population growth, urbanization, and higher standards of living leading to higher *per capita* use of water. Water use per person varies from 10 35 liters each day in rural areas to 40 to over 300 liters day^{-1} in areas with the highest living standards. Improved water supply systems also could significantly improve health conditions for people in developing countries, which also could contribute to increased productivity. As Gupta (1992) has observed, For the urban poor, water is important for improving their productivity and keeping them in good health. Almost 60% of the deaths amongst the slum settlements in various countries are known to be water-related ••• .

REGIONAL PATTERNS AND TRENDS

Gupta (1992) reported that global water use from 1970 2000 was likely to double, with the North American proportion decreasing from 21 15%. In contrast, the water use needs in Asia were estimated to increase from 58 to over 60% during the same time period. Comparable figures for 1970 and 2000 were 4% and 4% for Africa, 3.3% and 4.2% for South America, and 11% and 13% for Europe.

Regarding domestic water use, the poorest service for safe potable water supplies in 1980, prior to the start of the Water Decade, was provided to the rural poor in the nations of West Africa, with 22% of the population serviced, and in Africa south of the Sahara, where 23% received service. For all of Africa, an average of 30% of the population had access to safe portable water. In contrast, the poorest level of rural sanitation coverage was in Southeast Asia and in the eastern Mediterranean where less than 10% of the population was serviced. These figures translate into global figures of 1.65 billion rural people without access to an adequate supply of safe water supply in 1980, and 1.47 billion rural people without access to good sanitation.

By 1990, Traoré (1992) reported that 1 billion people in developing countries were without access to safe water supply, and some 1.75 billion did not have access to satisfactory sanitation. This need existed, despite the significant progress during the Water Decade (*see* IDWSSD (International Drinking Water Supply and Sanitation Decade), Volume 4). As Traoré (1992) noted, much of this achievement can, however, be accounted for by the major programs having being implemented in China and India, two countries with a major part of the world s rural population. In many of the developing countries less significant progress has been made and conditions are still unsatisfactory with levels of coverage under 20% being not uncommon . On a positive note, most countries in South America have adequate potable water, and indeed Gupta (1992) has stated that South American countries are able to provide about 10 times more water *per capita* than do the countries in Africa and Asia. Tables 4 and 5 provide further insight regarding the spatial patterns of water availability.

In terms of *per capita* water use, the ranges are from as little as 20 liters to over 500 liters. However, only some

Table 4 Water availability *per capita*, mid 1980s

Category	*Per capita* availability	Countries (%)
Very low	1000 m^3 a^{-1} or less	14
Low	1000–5000 m^3 a^{-1}	37
Medium	5000–10 000 m^3 a^{-1}	14
High	10 000 m^3 a^{-1} and more	35

Source: Gupta (1992).

Table 5 Average annual availability of freshwater (surface and groundwater) for selected countries, mid-1980s

Countries	Total (km^3)	*Per capita* ($m^3 \cdot 10^3$ per person)
Water-rich countries		
Iceland	170	685.48
New Zealand	397	117.53
Canada	2901	111.74
Norway	405	97.40
Nicaragua	175	49.97
Brazil	5190	36.69
Ecuador	314	31.64
Australia	343	21.30
Cameroon	208	19.93
Indonesia	2530	14.67
US	2478	10.23
Water-poor countries		
Egypt	1.00	0.02
Saudi Arabia	2.20	0.18
Barbados	0.05	0.21
Singapore	0.60	0.23
Kenya	14.80	0.66
Netherlands	10.00	0.68
Poland	49.40	1.31
South Africa	50.00	1.47
Haiti	11.00	1.59
Peru	40.00	1.93
India	1850.00	2.35
China	2800.00	2.58

US, United States.

4% of the global population use water in the range of 300 400 liters person^{-1} day^{-1}. In contrast, two thirds of the world s population, located primarily in Africa and Asia, use less than 50 liters daily (Gupta, 1992). It was estimated that by the year 2000 there would be 17% of the global population using more than 300 liters person^{-1} day^{-1}, with the US, Canada and Switzerland having the highest water uses. At the other end of the spectrum, by 2000 it also was believed that 30% of the global population would use fewer than 50 liters daily.

In terms of irrigation agriculture, Rydzewksi and bin Abdullah (1992) concluded that during the period 1950 1980 the annual rate of increase was 4 to 5 million ha. The relative importance of irrigation agriculture varies

Box 4 Forecasts regarding irrigation potential

The forecasts for ultimate potential for irrigated areas in developing countries indicate that it could be as high as double the present figure, i.e., rising to over 300 million ha, but such estimates are usually based only on availability of land and water resources and they become less optimistic when social and economic factors are introduced (Rydzewski and bin Abdullah, 1992).

signi cantly among different world regions. For example, in relation to total irrigatable land at a global scale, irrigation agriculture accounts for 33% in Asia, 21% in Near East/North Africa, 8.5% in Latin America, and 2.7% for Africa. Two countries (China and India) accounted for 40% of the increase in irrigation agriculture between 1950 and 1980, with India being the global leader in terms of irrigation agriculture. Within regions, the importance of irrigated agriculture is even more striking for some countries. Thus, e.g., 100% of agriculture in Egypt is based on irrigation, and Pakistan has 78% of its agriculture land supported by irrigation (see Box 4) (*see* **Irrigation: Environmental Impacts**, Volume 3).

Since 1980 the annual rate of increase of irrigated land fell by about a half, primarily because the best irrigatable land had already been developed and therefore construction costs had increased. In addition, lower commodity prices were a disincentive for farmers in some countries to invest in irrigation infrastructure.

Given the current and anticipated patterns of water use and the ongoing degradation of water sources, it can be expected that there will be regions in the world in which serious scarcity problems will continue to exist. The most obvious trouble spots, in this regard, are the Middle East, North Africa, and various countries in East Asia. Groundwater depletion also is emerging as a problem as it has been documented that groundwater is being extracted at unsustainable rates in many regions of the US, China and India.

ENVIRONMENTAL IMPACTS

Traore (1992) has argued that over wide areas, the degradation of water resources already constitutes the gravest environmental problem, as for instance in several countries in eastern Asia . Contamination of groundwater threatens the safety of water users. For example, in Bangladesh about 97% of the population is dependent on tube well water for drinking and cooking, and it has been established about 40% of Bangladeshis, or 56 million people of the 1999 population of 140 million, was vulnerable to arsenic poisoning from groundwater sources (Naidu and Skinner, 1999).

Four types of environmental impacts from water use are often recognized (Kindler, 1992). First, there may be impacts on or disruption to both human settlement and human activities, as when a dam and reservoir are constructed for water supply purposes, and people have to be relocated from the area, which is inundated by the reservoir waters. Second, developments to meet water use needs can create habitats supportive of parasitic and waterborne diseases (e.g., schistosomiasis, malaria, filariasis and liver fluke infections). Irrigation projects in both humid tropical and arid regions have led to such impacts. Third, there may be physical or chemical impacts, especially as a result of changes in land use or in surface or groundwater regimes. The causal events in this regard normally are the construction of storage reservoirs and flood levees or dykes, designed to control flows of river systems. Fourth, there can be impacts on indigenous flora and fauna.

Expanding use of water for irrigation agriculture illustrates some of the negative environmental impacts, particularly regarding salinization, water logging and contamination of surface and groundwater systems with agrochemicals. Indeed, the success of the Green Revolution has been predicated on the provision of abundant water and the application of fertilizers to achieve higher yields from the new species of rice and other grains. Without sufficient supplies of water, the capacity to grow food from the new hybrid rice to meet the nutritional needs of growing populations will be inadequate.

The Aral Sea basin is a specific example of some of the negative environmental impacts related to irrigation. As Bedford (1996) noted, the Aral Sea basin is plagued by serious environmental degradation, fundamentally related to water use , and has been described variously as a water management disaster, a quiet Chernobyl and the most serious environmental catastrophe of the planet. Most of the negative environmental impacts have been caused by a drive to increase cotton production in the basin through expanding irrigation. Between 1950 and the late 1980s, the area under irrigation increased from 2.9 million to 7.2 million hectares. Specific negative environmental impacts from the massive increase in water withdrawals have included desiccation of the Aral Sea and its delta wetlands, leading to loss of habitat and species, wind-transported salts from the dry Aral Sea bed onto agricultural lands, and alteration of the micro climate. Negative impacts from the application of the withdrawn water onto what often was marginal agricultural land have included the typical arid region problems associated with irrigation of soil salinization and water logging. Increased rates and amounts of use of agrochemicals have led to pollution of food and drinking water supplies. In turn, the ingestion of agrochemicals contained in or on food products has created health problems (*see* Aral Sea, Volume 4).

SOCIAL IMPACTS

Health (Box 5)

Box 5 Irrigation-related diseases

Two of the most important irrigation-associated diseases are malaria and schistosomiasis. [For malaria] ••• the most dramatic effects are seen in arid and semi-arid zones where irrigated agriculture is introduced. However, even in areas which already have a certain level of malaria endemicity, irrigation development may mean a shift from seasonal to perennial transmission with severe consequences especially for children.

Schistosomiasis is a chronic debilitating disease. Many cases of sharp increases in the prevalence rates of this disease from 9% to 70–80% in the wake of irrigation development have been documented, particularly from Africa (Rydzewski and bin Abdullah, 1992).

Despite the Water Decade, the concern about health relative to water use continues to be a critical concern, given that a strong association exists between four diseases or disease groups, and water supply and sanitation. First, as Traoré (1992) has explained, the classical water-borne diseases are cholera and typhoid fever, as well as infectious hepatitis and shigellosis. Even more common, and in the long term much more deadly, are diarrheal diseases associated with poor hygiene and fecal–oral transmission. Second, regarding water-based parasitic diseases, schistosomiasis and dracunculiasis are the most significant. Third, new reservoirs to serve as sources of water for domestic use, irrigation or hydro-power generation affect malaria prevalence. Regarding sanitation, facilities such as pit latrines and sewage drains have contributed to increased incidences of filariasis. Fourth, a major category of water-related disease has emerged, illnesses associated with nitrates, heavy metals, and pesticides.

In terms of future conditions, Traoré (1992) suggested that by the early 1990s it appeared as if diarrhea was causing between four and five million deaths annually in children younger than five years of age in developing nations. Indeed, the World Health Organization (WHO) (1989) estimated that in the first two years of life as many as 15 of 1000 children were dying from diarrhea.

Regional Instability

Scarcity of water, or drought, as demonstrated during the 1970s in the Sahel region of Africa, can lead to significant changes in land cover and land use, and also to the out-migration of residents from the area in which water use needs cannot be satisfied. In such situations, what are increasingly being referred to as environmental refugees migrate to other areas and often need assistance and support. Because the host areas do not provide the basic requirements for potable water and sanitation to the existing domestic population, the arrival of thousands of new

water users can exacerbate an already difficult situation and often lead to social unrest.

Regional instability also can be created when people in a region face chronic rather than acute water shortages. Some people leave the area but most remain. Relative to accepted indicators of well-being, these people subsist and live below the poverty line. However, they also live in what they consider to be their homeland , so are reluctant to migrate away, unless conditions become life-threatening.

It increasingly is being stated that if during the 20th century some regional conflicts and wars were triggered about concerns over access to fossil fuel resources, such conflicts in the 21st century will at least partially be attributable to fair sharing of scarce water supplies. In that context, the most frequently cited region for such potential instability is the Jordan River basin in the Middle East, especially related to access to water on the West Bank. The Israelis, Jordanians and Palestinians all face constraints on economic development due to water shortages, and there are inequities in the way in which water is presently allocated (Ben-Zvi *et al.*, 1998). In the same region, conflicts also exist regarding access to water in Gaza.

Such water conflicts are not hypothetical possibilities. Wolf (1995) referred to the period of 1964 1967 as one of Water Wars , beginning with an Arab decision in 1964 to construct an all-Arab diversion from the Jordan River headwaters which would preclude the operation of the Israeli National Water Carrier (which takes water by pipeline from Lake Kinneret/Galilee in northern Israel southward to the Negev region. In addition, the National Carrier distributes water to other regions, cities, towns and irrigation districts in Israel between Lake Kinneret and the Negev (Kuffner, 1993)). In 1967, Israel used tanks and air strikes to halt construction of the diversion works.

The potential conflicts are not confined to Israel and her Arabic neighbors. During the 1990s, Syria and Iraq each threatened retaliative action against Turkey if that country proceeded to build dams in the headwaters of rivers, especially the Euphrates River, which later flow into their two countries (Turan, 1993; Wakil, 1993). As a result, considerable multilateral effort has been devoted to trying to resolve the allocation of scarce water supplies to meet the water use needs in the Jordan River system (Biswas and Wolf, 1994; Wolf, 1995).

IMPLICATIONS

Distinction Among Different Futures: Probable, Desirable and Feasible

When thinking about future water use demands, estimates of the most probable patterns of water use are compared against estimates of water supplies. When it is concluded that water use will exceed water supplies, then decisions usually are taken to develop new supply capacity so that future water use needs can be met.

However, as highlighted in the discussion regarding the drivers of water use, the preparation of forecasts always involves making many assumptions, ranging from the rate of population growth and urbanization, *per capita* water use, and the nature of the economy, as well changes in technology, climate, social values and tastes and policies. If forecasts related to any of these drivers are faulty, then these become built into the forecasts of future water use patterns, making them also inaccurate.

There is also the issue of focusing on the most probable future, when such future conditions may not be desirable. Indeed, the Water Decade was predicated on the idea that without intervention, the most probable future regarding provision of safe potable water supplies and sanitation services would not lead toward desirable future conditions for billions of people in the world. Thus, it is important that attention be given to both what is likely, and what is desirable. A third consideration requiring attention is what *can* happen. In other words, we also must think about what is a feasible future, as well as what are probable and desirable futures. Society may not always have the human and financial resources to achieve the most ambitious desirable future. For example, some of the goals of the Water Decade were not achieved because time was not taken to train people in villages to maintain pumps and other equipment related to new wells that were dug. As a result, often within a period of one year, the new wells no longer functioned, and the local people had returned to contaminated sources of water.

Role of Sustainability as an Overriding Vision

It is essential to have a sense of desired end point being sought. The word vision often is used to characterize that sense of direction. As Senge (1990) has stated, a vision is a specific destination, a picture of a desired future , and, at its most basic level, answers the question What do we want to create? .

Stewart (1993) has outlined an 8-step process or method to develop a vision, and it is presented here to stimulate thinking about how a vision can be prepared. In his view, the steps should include:

1. developing a comprehensive list of stakeholders and viewing the future state and the present state through their eyes;
2. developing a broad description of the likely future environment;

3. creating a comprehensive vision of what could be, unrelated to existing barriers, before considering the present state;
4. contrasting the future vision with the present situation;
5. identifying the values needed to guide a society or organization as it seeks to realize its vision;
6. ensuring the vision is articulated as actionable concepts;
7. developing the vision in a participatory manner through involvement of stakeholders; and,
8. avoiding development of a strategy or taking action until the vision and values have been established.

The distinctive features of Stewart s approach are that a conscious decision is taken to separate the characteristics of the desirable future state with the pathway or process to reach it, and to first focus on the future before assessing the present situation.

Sustainability, or sustainable development, has been held out by many as a desirable vision or endpoint for water management (Young *et al.*, 1994). In that context, emphasis has been placed on meeting the needs of people today, while ensuring that the needs of future generations also will be met. As the World Commission on Environment and Development (WCED) (1987) explained, sustainable development as a vision contained the key concept of ••• needs, in particular the essential needs of the world s poor, to which overriding priority should be given . The Water Decade focused on meeting the demands of the World s poorest people regarding basic water use needs, and this should continue to be a global priority. A World Water Commission has been established, and one of its first priorities has been to develop a global vision for water (Serageldin, 2000).

Role of Supply and Demand Management to Meet Needs

The conventional approach to meeting water use needs has been to expand existing sources, or to create new sources. Thus, if water use was expected to exceed supply capacity, the usual approach has been to build another water supply reservoir, construct a pipeline to a lake or similar source, or drill a deeper or new borehole or well. This approach has been labeled as the *supply management approach*, since the emphasis has been upon enhancing or increasing the supply of water. The key component of the supply approach has been structural, or building new water supply facilities.

One of most promising supply management options is desalinization. Dziegielewski (2000) reported that in 1962 the energy requirement to desalinate 1000 gallons of water was 100 kilowatt-hours (kWh), or 26.4 k Wh m$^{\bullet 3}$ whereas by 1998 this requirement had decreased to 11 k Wh 1000$^{\bullet 1}$ gallons or 2.9 k Wh m$^{\bullet 3}$. This reduction in energy requirements is primarily due to the invention of reverse osmosis. In the US, the cost of desalinated water in 1999 was about \$11 000 gallons$^{\bullet 1}$, making it competitive with the cost of importing water into many urban areas. Thus, desalinization is no longer significantly more expensive, as it used to be when contrasted to the historical cost of developing new water supplies. Indeed, the cost of desalinated water continues to decrease, whereas the cost of water from other sources is generally increasing. In Tampa Bay, Florida, the decision was taken to build a 25 million gallons per day desalinization plant which will be completed in 2002 and will provide 10% of the total water supply capacity. This option was chosen because of the environmental impacts (groundwater overdraft and over pumping of lakes) associated with an alternative project to withdraw additional freshwater.

As the most suitable dam and reservoir sites became developed, and as awareness grew about some of the negative environmental and social impacts of major structural works, attention has turned increasingly to what has been labeled as a demand management approach. In this approach, the focus shifts away from providing additional water supplies to meet anticipated use and toward modifying human behavior, and especially to reduce wasteful use of water resources (Shrubsole and Tate, 1994). In situations in which basic human needs already are satisfied, the *demand management approach* is appropriate. Through various pricing schemes and regulations on activities such as watering of gardens and lawns, it is possible to reduce peak demands, which require water supply capacity only used for a small part of the year, and also to reduce overall levels of use. However, in situations in which basic human needs are not met, increasing the price of water, so that consumers pay the total cost of providing it to them, has the potential to make safe drinking water inaccessible to the poorest and most vulnerable people in a society. As a result, caution has to be taken in using some of the tools in the demand management approach, notwithstanding that in most developed countries there is considerable scope to achieve efficiencies in water use patterns, and to reduce waste.

It has never been suggested that the demand management approach should replace the supply management approach. Instead, the argument usually has been that water managers have over-relied on the supply management approach, usually because the key decision makers have been civil engineers, whose training emphasized building things to fix problems, whether they be water, energy or transportation. The emerging view is that supply and demand management approaches should and can be used together, to develop a package custom-designed for any given situation. This view is an appropriate one, and over-reliance on the supply management approach must steadily give way to a more balanced and combined approach in which both supply and demand management methods are used.

REFERENCES

Arnell, N W (1999) The Effect of Climate Change on Hydrological Regimes in Europe: a Continental Perspective, *Global Environ. Change*, **9**(1), 5 23.

Ayibotele, N B (1992) The World s Water: Assessing the Resource, *International Conference on Water and the Environment: Keynote Papers*, ICWE Secretariat, World Meteorological Organization, Geneva, 1.1 1.26.

Bedford, D P (1996) International Water Management in the Aral Sea Basin, *Water Int.*, **21**(2), 63 69.

Ben-Zvi, A, Dlayahu, E, Gottesmann, M, and Passal, A (1998) Evolution of the 1990 1991 Water Crisis in Israel, *Water Int.*, **3**(2), 67 74.

Biswas, A K and Wolf, A T (1994) Middle East Water Commission, *Water Int.*, **19**(1), 3 4.

Born, S M, Genskow, K D, Filbert, T L, Hernandez-Mora, N, Keefer, M L, and White, K A (1998) Socioeconomic and Institutional Dimensions of Dam Removals: the Wisconsin Experience, *Environ. Manage.*, **22**(3), 359 370.

Clamen, M (1990) Implications of Climate Change for Water Resources in the Great Lakes Basin, eds G Wall and M Sanderson, *Climate Change Implications for Water and Ecological Resources*, Occasional Paper No. 11, Department of Geography, University of Waterloo, Waterloo, Ontario, 179 186.

Dzietgielewski, B (1999) The Role of Desalinization in Averting a Global Water Crisis, *Water Int.*, **24**(4), 395 398.

Gleick, P H (1996) Basic Water Requirements for Human Activities: Meeting Basic Needs, *Water Int.*, **21**(2), 83 92.

Gupta, D B (1992) The Importance of Water Resources for Urban Socio-economic Development, *International Conference on Water and the Environment: Keynote Papers*, ICWE Secretariat, World Meteorological Organization, Geneva, 5.1 5.19.

Kindler, J (1992) Water the Environmental and Development Dimensions Striking a Balance, *International Conference on Water and Environment: Keynote Papers*, ICWE Secretariat, World Meteorological Organization, Geneva, 2.1 2.19.

Kuffner, U (1993) Water Transfer and Distribution Schemes, *Water Int.*, **18**(1), 30 34.

Landsberg, H H (1964) *Natural Resources for U.S. Growth: a Look Ahead to the Year 2000*, Johns Hopkins Press, Baltimore, MD.

Mimikou, M A, Kanellopoulouand, S P, and Baltas, E A (1999) Human Implications in the Hydrological Regime due to Climate Change in Northern Greece, *Global Environ. Change*, **9**(2), 139 156.

Naidu, R and Skinner, H C W (1999) Arsenic Poisoning of Rural Groundwater Supplies in Bangladesh and India: Implications for Soils Quality, Animal and Human Health, *Proceedings of the International Conference on Diffuse Pollution*, CSIRO Land and Water, Wembley, Western Australia, 407 418.

Niemczynowicz, J (1996) Megacities from a Water Perspective, *Water Int.*, **21**(4), 198 205.

Rydzewski, J R and bin Abdullah, S (1992) Water for Sustainable Food and Agricultural Production, *International Conference on Water and Environment: Keynote Papers*, ICWE Secretariat, World Meteorological Organization, Geneva, 4.1 4.25.

Seckler, D, Barker, R, and Amarasinghe, U (1999) Water Scarcity in the Twenty- rst century, *Int. J. Water Res. Dev.*, **15**(1/2), 29 42.

Senge, P M (1990) *The Fifth Discipline: The Art & Practice of the Learning Organization*, Doubleday, New York.

Serageldin, I (1995) *Toward Sustainable Management of Water Resources*, The World Bank, Washington, DC.

Serageldin, I (2000) A Report of the World Commission on Water for the 21st century, *Water Int.*, **25**(2), 284 302.

Sewell, W R D and Bower, B T (1968) Problems and Procedures, *Forecasting the Demands for Water*, eds W R D Sewell and B T Bower *et al.*, Queen s Printer and Controller of Stationery, Ottawa, 19 42.

Shrubsole, D and Tate, D, eds (1994) *Every Drop Counts*, Canadian Water Resources Association, Cambridge, Ontario.

Stewart, J M (1993) Future State Visioning a Powerful Leadership Process, *Long Range Plan.*, **26**(6), 89 98.

Stren, R, White, R, and Whitney, J, eds (1992) *Sustainable Cities: Urbanization and the Environment in International Perspective*, Westview Press, Boulder, CO.

Svendsen, M and Rosegrant, M W (1994) Irrigation Development in Southeast Asia Beyond 2000: will the Future be like the Past? *Water Int.*, **19**(1), 25 35.

Timmerman, P and White, R (1997) Megahydropolis: Coastal Cities in the Context of Global Environmental Change, *Global Environ. Change*, **7**(3), 205 234.

Traore, A (1992) Water for the People Community Water Supply and Sanitation, *International Conference on Water and Environment: Keynote Papers*, ICWE Secretariat, World Meteorological Organization, Geneva, 3.1 3.22.

Turan, I (1993) Turkey and the Middle East: Problems and Solutions, *Water Int.*, **18**(1), 23 29.

United Nations (1991) *World Population Prospects*, United Nations, New York.

Wakil, M (1993) Analysis of Future Water Needs for Different Sectors in Syria, *Water Int.*, **18**(1), 18 22.

Wolf, A T (1995) International Water Dispute Resolution: the Middle East Multilateral Working Group on Water Resources, *Water Int.*, **20**(3), 141 150.

World Commission on Environment and Development (1987) *Our Common Future*, Oxford University Press, Oxford.

World Health Organization (1989) *Diarrhea Diseases Control Program: Intervention-related Research Priorities of the Scientific Working Group on Epidemiology and Disease Prevention*, WHO, Geneva.

Young, G J, Dooge, J C I, and Rodda, J C (1994) *Global Water Resource Issues*, Cambridge University Press, Cambridge.

FURTHER READING

Abu-Zeid, M, Shady, A M, and Coulomb, R (1998) The World Water Council in Three Years 1994 1997, *Water Int.*, **23**(2), 97 106.

Beale, B and Fray, P (1990) *The Vanishing Continent: Australia s Degraded Environment*, Hodder and Stoughton, Sydney.

Engleman, R and LeRoy, P (1993) *Sustaining Water: Population and the Future of Renewable Water Supplies*, Population Action International, Washington, DC.

Falkenmark, M and Lindh, G (1974) How can we Cope with the Water Resources Situation by the Year 2015? *Ambio*, **3**(3 4), 114 122.

Frey, G W (1993) The Political Context of Conflict and Cooperation over International River Basins, *Water Int.*, 54 68.

Grover, B and Howarth, D (1991) Evolving International Collaborative Arrangements for Water Supply and Sanitation, *Water Int.*, 16(3), 145 152.

Loucks, D P (2000) Sustainable Water Resources Management, *Water Int.*, 25(1), 3 10.

McCaffrey, S C (1992) A Human Right to Water: Domestic and International Implications, *Georgetown Int. Law Rev.*, V(1), 1 24.

Mitchell, B (1997) *Resource and Environmental Management*, Longman: Harlow, UK.

Postel, S (1984) *Water: Rethinking Management in an Age of Scarcity*, Worldwatch Paper 61, Worldwatch Institute, Washington, DC.

Reisner, M (1987) *Cadillac Desert: The American West and its Disappearing Water*, Penguin Books, New York.

Robinson, J B (1982) Energy Backcasting: a Proposed Method of Policy Analysis, *Energy Policy*, 10(4), 337 344.

Robinson, J B (1988) Unlearning and Backcasting: Rethinking Some of the Questions we Ask about the Future, *Technol. Forecast. Soc. Change*, 33(4), 325 338.

Robinson, J B (1990) Futures Under Glass: a Recipe for People who Hate to predict, *Futures*, 22(8), 820 842.

Semiat, R (2000) Desalination: Present and Future, *Water Int.*, 25(1), 54 65.

Shiklomanov, I A (2000) Appraisal and Assessment of World Water Resources, *Water Int.*, 25(1), 11 32.

United Nations (1995) *World Population Prospects: the 1994 Revision*, Department for Economic and Social Information and Policy Analysis, ST/ESA/SER.A/145, New York.

Vorosmarty, C J, Green, P, Salisbury, J, and Lammers, R (2000) Global water resources: vulnerability from climate change and population growth, *Science*, 289, 284 287.

World Health Organization (1991) *The International Drinking Water Supply and Sanitation Decade: Review of Decade Progress (as at December 1990)*, WHO Division of Environmental Health, Geneva.

Contaminated Lands and Sediments: Chemical Time Bombs?

WILLIAM M STIGLIANI

Center for Energy and Environmental Education, University of Northern Iowa, Cedar Falls, IA, USA

In 1988, investigators reported a trend of increasing mercury content in sh in Swedish lakes, even though there had been a signi cant decrease in mercury emissions from Swedish industries in the previous two decades (Hakanson et al.*, 1988). Commenting on this trend, the authors noted that " old Swedish 'sins' are still causing a lot of problems" . In the 1990s, the addition of phosphate (PO_4^{-3}) fertilizer to fruit orchards in Washington State caused the fruits to become contaminated with arsenic, even though the application of lead arsenate (AsO_4^{-3}) insecticides was discontinued in 1947 (Creger and Peryea, 1994; Peryea, 1991). Today in regions such as the Great Lakes of North America, infants exposed in the womb to dichloro-diphenyl-trichloroethane (DDT) and polychlorinated biphenyls (PCBs) suffer developmental and learning disorders, although these chemicals have been banned or severely restricted since the 1970s (National Research Council, 1999).*

Such occurrences suggest that: (1) environmental effects may be displaced in time from the activities that cause them; and (2) even when reductions in polluting emissions are enacted, environmental effects may still occur owing to the accumulation of polluting materials in the environment over previous decades. Policy makers and the public alike may be surprised by the onset of such observed deterioration in the environment, since prevailing wisdom might suggest that an era of lowered emissions of pollutants to the environment should be accompanied by an improvement in its state, rather than a worsening.

While regulators have focused their attention on highly contaminated waste sites and the means by which to mitigate acute threats to the public, they have largely ignored the potential impacts from the slow accumulation of toxic materials in soils and sediments over the long term. From a public policy viewpoint, this lack of regard supports the assertion by Brooks (1986) that the manifestation of economic bene ts and environmental costs may characteristically differ in two important respects. First, whereas the bene ts of activities that pollute the biosphere are directly proportional to the scale (higher production and use produces more wealth), the environmental costs may vary nonlinearly with scale. Second, bene ts are typically manifested on fast time scales, while the costs become obvious on relatively slow time scales. This mismatch of time scales can be problematic because a technological activity with all its obvious bene ts may become strongly entrenched, and thus dif cult to scale back, before its true environmental costs are known.

DETOXIFICATION AND RETOXIFICATION OF THE BIOSPHERE

As conceptualized by Salomons (1995), one may visualize the transport of heavy metals in the environment as a wheel in which the metal moves with relentless progression from the hub to the periphery (left side of Figure 1). Heavy metals released to the air or aqueous (aq.) systems (center and inner circle of the wheel, respectively) are transferred to soils and sediments (outer circle). The lifetimes of metals in air are on the order of days, and in water they are months to years. In contrast, the residence times of heavy metals in soils and freshwater and estuarine sediments can be much longer, on the order of decades to centuries or more. Eventually, through erosion and resuspension processes the metals are transferred from the land to marine and deep ocean sediments, their ultimate sink. One may view this transport as a natural detoxification process by which air, soils, freshwaters and their sediments purge themselves of pollutants and transfer them for burial, and out of harm s way, in the sediments of the deep oceans. This voyage, however, from the terrestrial biosphere to deposition in ocean sediments, can take thousands of years or longer. The cycle is renewed though crustal uplifting when ocean sediments are transformed into new landforms, but this process takes millions of years. Hence, on time scales

important to human concerns, soils and freshwater and estuarine sediments can be viewed as the ultimate sinks for heavy metals. Typically, well under 1% of the metals released to the environment reside in air and water. The remainder accumulates in soils and sediments. Persistent organic compounds such as DDT and PCBs are another class of harmful chemicals that accumulate in soils and sediments, and are discussed later in this article.

Important sources of heavy metals to the environment are direct industrial discharges, atmospheric deposition, corrosion of building materials and traffic related emissions in urban areas, erosion and storm runoff from agricultural soils and paved urban areas, and leaks from municipal landfills and toxic waste dumps. Emissions from these sources may be in accordance with regulatory standards, and stem from common anthropogenic activities not normally considered as important sources of soil and sediment pollution. The point to stress, however, is that while such sources may add small incremental increases to the environment on an annual basis, their cumulative load integrated over many decades can be substantial (*see* Anthropogenic Metabolism and Environmental Legacies, Volume 3).

It is not total accumulation of a heavy metal in a soil or sediment that is important, but rather its bioavailability. Bioavailability is mostly determined by the water soluble fraction of the accumulated chemical in soil water solution or in pore water in sediments; the adsorbed, insoluble phase is not available for biological uptake. In relatively unpolluted soils and sediments, the soluble phase is typically only a small fraction of total accumulation. The question raised here and illustrated on the right side of Figure 1 is whether, and under what conditions, soils and sediments serving as sinks for toxic heavy metals can reverse their role and become important sources of heavy metal pollution in drinking water, crops and vegetation, and aquatic and terrestrial wildlife? We may think of such a reversal as a retoxification process in the biosphere.

THE CHEMICAL TIME BOMB CONCEPT

The chemical time bomb (CTB) concept addresses the varied mechanisms by which retoxification may occur. It was originally formulated by Stigliani (1988), and resulted in a five-year collaborative effort between the International Institute for Applied Systems Analysis (IIASA) in Laxenburg, Austria, the National Institute of Public Health and Environmental Protection in Bilthoven, the Netherlands, and the Netherlands Ministry of Housing, Physical Planning and Environment. The findings of the ensuing studies are provided in Stigliani (1996), Salomons and Stigliani (1995), Stigliani and Anderberg (1994), Stigliani and Salomons (1993), ter Meulen *et al.* (1993), Hesterberg *et al.* (1992), Stigliani (1991), and Stigliani *et al.* (1991).

A CTB is defined as an unforeseen chain of events resulting in the delayed and sudden occurrence of harmful effects due to the mobilization or chemical transformation of chemicals stored in soils and sediments in response to saturation or alteration in certain environmental conditions (adapted from Stigliani *et al.*, 1991). The idea of chemical saturation of stored pollutants in soils is illustrated in Figure 2(a). It shows that an added pollutant is partitioned between an adsorbed (sorbed) solid state (Q) and a dissolved aq. state (C). At an early stage, most of the pollutant remains in the sorbed phase, illustrated by the large value of the ratio of $\Delta Q_1/\Delta C_1$. As pollution continues over time, the ratio of the sorbed to mobile pollutant diminishes, as shown by the lower value of $\Delta Q_2/\Delta C_2$. Thus, a pollutant added at a constant rate over time will cause the fraction that goes into solution to become larger and larger. The figure also shows a maximum soil sorption capacity . This is the point at which virtually all of the input pollutant goes to the soluble phase because the soil is saturated with respect to its ability to adsorb it.

A second mode of saturation is illustrated in Figure 2(b). This mechanism involves a change in a soil property that

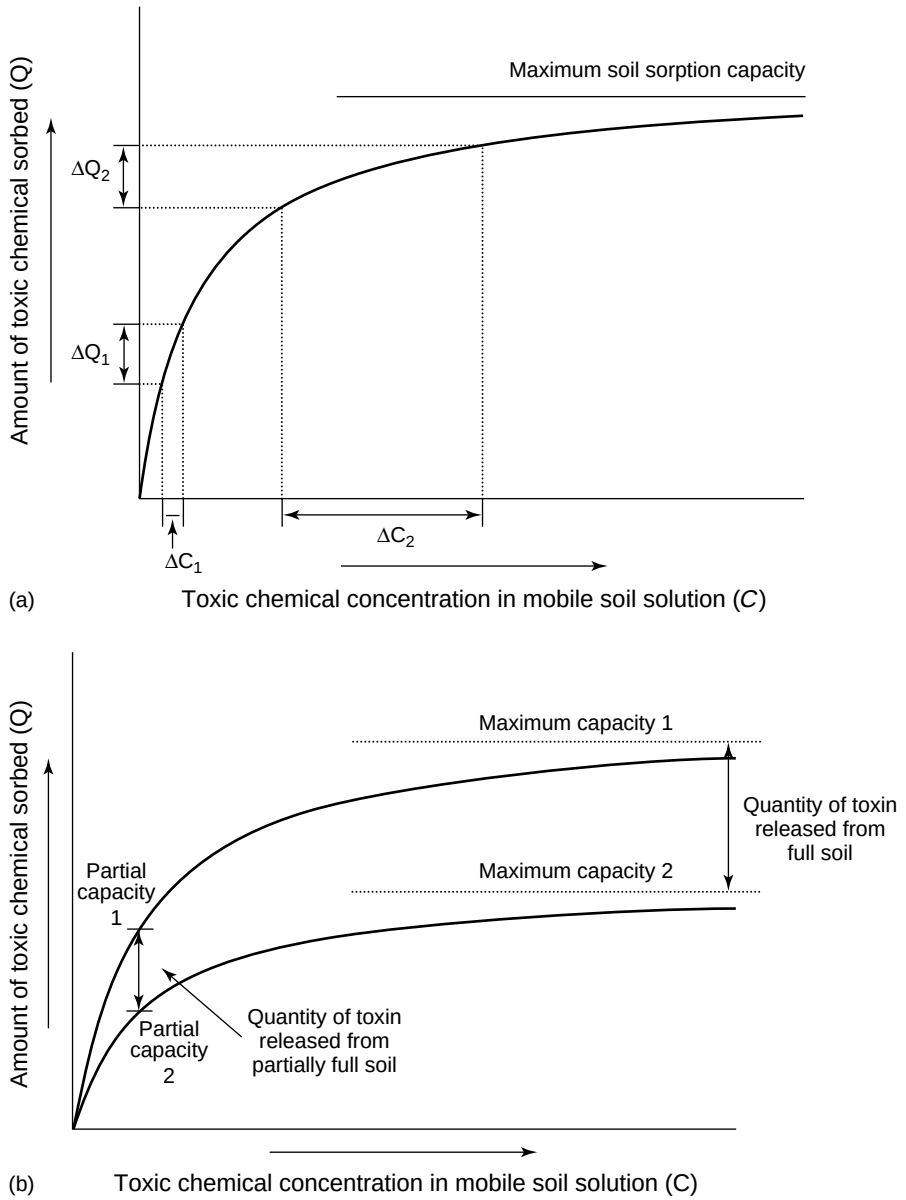


Figure 2 (a) Partitioning of a toxic chemical between its sorbed (Q) and soluble (C) states; (b) effect of changing soil or sediment properties on mobilizing a stored toxic chemical. (Reproduced from Hesterberg *et al.*, 1992)

reduces the maximum capacity from state 1 to state 2. When this occurs, a significant fraction of the sorbed chemical from state 1 (indicated by partial capacity 1 on the upper curve in the figure) is desorbed as the system reaches a new equilibrium defined by state 2 (indicated by partial capacity 2 on the lower curve in the figure). Reaching saturation in soils with high initial sorption capacities (Figure 2a) can take a long time. However, saturation can occur quickly and unexpectedly via the process indicated by Figure 2(b).

To illustrate how the second mode operates, consider the example, cited above, of arsenic contamination of fruit orchards in Washington State. Acid lead arsenate ($PbHAsO_4$) was the preferred pesticide for insect control in deciduous fruit trees before the introduction of DDT in 1947. Frequent high rate applications led to substantial lead and arsenic accumulation in orchard topsoils. AsO_4^{-3} adsorbed to the surfaces of soil particles through ligand exchange with hydroxides (OH^-), hydrous metal oxide minerals (iron/magnesium oxides), phyllosilicate clays, and soil carbonates. In recent decades, fruit growers began applying PO_4^{-3} fertilizer to their orchards to increase yield. This had the unexpected result of mobilizing AsO_4^{-3} with subsequent uptake through the roots and into the fruit. PO_4^{-3} additions mobilized AsO_4^{-3} because it adsorbs more

tightly to soil surfaces, i.e.;

$$\{\text{soil particle surface}\} - AsO_4^{-3} + PO_4^{-3}(\text{aq.}) \longrightarrow$$
$$\{\text{soil particle surface}\} - PO_4^{-3} + AsO_4^{-3}(\text{aq.})$$

Thus, PO_4^{-3} in the aq. phase replaces adsorbed AsO_4^{-3} on the surface of the soil, rendering the AsO_4^{-3} soluble and available for plant uptake.

The mechanism at play here was a rather rapid reduction in the soil s initial capacity to adsorb AsO_4^{-3} and keep it safely locked in the soil. Referring to Figure 2(b), the soil was in a partially full state (marked by partial capacity 1). Addition of PO_4^{-3} effectively shrank the capacity of the soil to adsorb AsO_4^{-3} by taking over the sites at the soil surfaces that had been occupied by AsO_4^{-3}. A new equilibrium was established between sorption and solubility as indicated by the curve marked by maximum capacity 2 . To reach this new stage, the soil released sorbed AsO_4^{-3} to the soluble phase (in an amount marked by the arrow between partial capacity 1 to partial capacity 2) that became available for plant uptake.

With respect to heavy metals, six soil/sediment parameters are particularly important in causing the type of shift shown in Figure 2(b). These parameters, dubbed Capacity Controlling Parameters (CCPs), are given in Table 1.

Thus, reducing the load of mobilizable heavy metals in the biosphere depends not only on reducing the input loads, but also on controlling human activities that tend to affect CCPs. As noted in Table 2 these include land use and intensity of land use, energy production, and the numerous activities that contribute to climate change. It is particularly noteworthy that climate change may be expected to affect all the CCPs to some degree.

Table 1 Important soil/sediment capacity-controlling properties (CCPs) for heavy metals

CCP	Detrimental environmental effect
CEC[a]	Soils with low CECs have correspondingly low capacities to retain heavy metals by sorption; CEC depends on inorganic clay mineral content and type, organic matter content, and soil pH
pH	Lowering pH increases heavy metal solubility, decreases CEC, and alters soil microbial populations
Eh[b]	Decreasing redox potential (*see* **Redox Potential**, Volume 2) dissolves iron and manganese oxides, resulting in the mobilization of sorbed toxic chemicals; in Hg contaminated sediments under low Eh, bacteria can generate highly toxic methylmercury; increasing redox potential mobilizes heavy metals by dissolving insoluble metal sulfides
OM[c] content	Decreasing OM content reduces CEC, soil pH, buffering capacity, sorption capacity for toxic organics, soil water-holding capacity; alters physical structure (e.g., increases soil erodibility), and decreases microbial activity
Salinity	Increasing salinity solubilizes toxic chemicals by altering the ion exchange equilibrium, increasing soluble complexation, and decreasing chemical thermodynamic activities in solution; it can also decrease microbial activity
Microbial activity	Altering microbial activity can affect other CCPs including organic matter, redox potential, pH; sulfate-reducing bacteria synthesize and emit methylmercury

[a] CEC, Cation exchange capacity.
[b] Eh, Redox potential.
[c] OM, Organic matter.

pH AS A CAPACITY-CONTROLLING PROPERTY IN SOILS

Acid Buffering Domains in Soils

Ulrich (1983, 1981) has classified the dominant buffering reactions in soils into four distinct ranges (*see* Buffering Capacity, Volume 2). These are shown in Table 3.

Table 2 Linkages between human activities and changes in CCPs

CCP	Human activities affecting CCP
CEC	Affected indirectly by activities that affect pH, organic matter, and salinity
pH	Atmospheric acid deposition from coal combustion; agricultural practices (liming increases pH, nitrogen fertilizer decreases pH); deforestation decreases pH; climate change altering the nitrogen cycle, and the seasonal patterns of precipitation and evapotranspiration
Eh	Drainage of wetlands or flooding of dry lands either as deliberate land use policy, or as a consequence of climate change
OM content	Deforestation decreases OM; changes in vegetation; agricultural practices (recycling crop residues and manure increases OM, intensive, high input farming without recycling decreases organic matter); climate change can increase or decrease OM
Salinity	Irrigation with salty ground waters; salt water intrusion of ground waters due to sea level rise
Microbial activity	Affected indirectly by activities that affect redox potential and pH, and directly by synergistic effects of multiple toxins, and climate change

Table 3 Buffering ranges in soil solution

Type	Range
Carbonate	$pH \geq 6.2$
Cation exchange	$4.2 \leq pH < 6.2$
Aluminum	$3.8 \leq pH < 4.2$
Iron	$pH < 3.8$

Source: Ulrich (1983, 1981).

Carbonate solids are efficient buffers in the neutral to basic range ($pH \geq 6.2$). Addition of acid in the soil causes the dissolution of solid carbonate and neutralization by the following reaction:

$$MCO_3 + 2H^+ \longrightarrow M^{+2} + H_2CO_3 (M = \text{Ca and/or Mg})$$

When the carbonate solids are distributed in sufficiently fine particles or grain coatings and evenly dispersed in the soil horizon (calcareous soils), the buffering capacity is nearly inexhaustible for any reasonable level of acid inputs.

If carbonate solids are depleted or absent from the soil, the buffering processes are controlled by cation exchange on aluminosilicate clays, a dominant soil type in temperate zones of the world. Clay soil particles possess negative charges on their surfaces that govern the balance between accumulation and solubilization of positively charged ions (cations) present in soil water solution. The number of negative sites available for a given volume of soil depends on soil type, and is measured by the CEC. Soil organic matter also contributes significant CEC owing to the presence in its structure of negative charges originating from phenolic ($-OH$) and organic acid ($-COOH$) constituents. These negative sites attract cations (C^+) that adsorb on the surface of the particle (see Figure 3a). The common natural cations found in abundance in clay soils are the four so-called base cations that include sodium (Na^+), potassium (K^+), magnesium (Mg^{+2}), and calcium (Ca^{+2}). These cations serve two essential roles: (1) they are nutrients for plant and animal life; and, (2) they buffer the acidity of soil solutions. As shown in Figure 3(a), hydrogen ions (H^+) in soil water take the place of base cations at the negatively charged surface sites. This step renders the solubilized base cation bioavailable, while at the same time reducing the soil water s acidity. The exchange is efficient because H^+ adsorbs more strongly to the charged clay site than the base cations. In general, cations are held and displace one another in the sequence:

$$Al^{+3} > H^+ > Ca^{+2} > Mg^{+2} > K^+ > NH_4^+ > Na^+$$

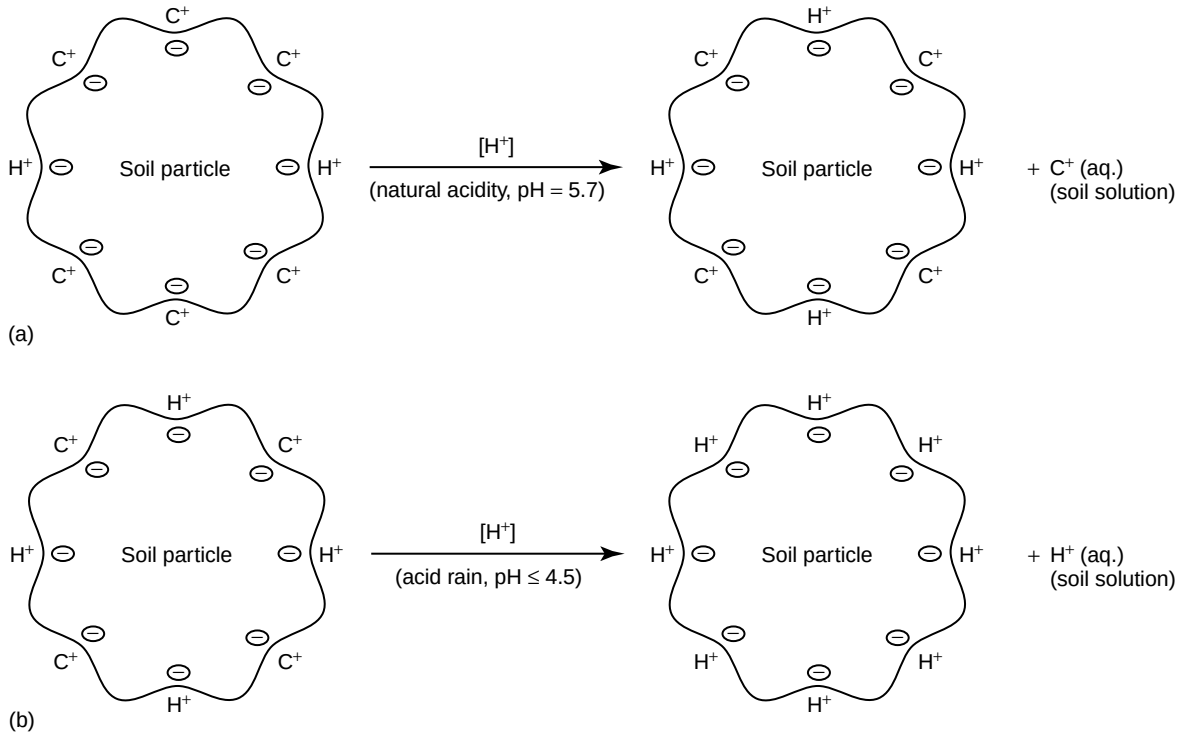


Figure 3 (a) Natural exchange between hydrogen ion (H^+) and a base cation (C^+) in a clay soil with high base saturation (β); (b) exchange between H^+ and C^+ in a clay soil subjected to acid rain

Unpolluted rainfall that drains through clay soils is naturally acidic (pH = 5.7), so the exchange shown in Figure 3(a) is an important step in the cycling of nutrients in the biosphere. Problems arise when polluted rainfall (so-called acid rain with pHs typically 4.5 or lower) percolates through clay soils. As illustrated in Figure 3(b), in this case H^+ is so concentrated that it sweeps away most of the base cations, leaving few additional surface sites available for H^+/base cation exchange. At this point, the clay has reached its saturation point with respect to its ability to buffer the acidic soil solution water, and runoff from these soils contains acid (H^+) rather than base cations. When such soils lie within the watershed of a lake, the unbuffered surface and subsurface runoff causes lake acidi cation.

Soil scientists use the parameter base saturation, signi ed by β, to describe the percentage of exchange sites that are occupied by base cations. For example, a clay soil with a β value of 100% has all of its negative sites occupied by base cations; values of 50 and 0% have one half and none of the sites occupied, respectively. In practice when the value of β declines to about 5%, the soil has essentially lost its ability to buffer acids. The buffering capacity at any given time equals $\beta \times$ CEC.

Soils continually replenish the base cation supply through weathering of bedrock. The supply of base cations as a component of soil (the exchangeable pool) is typically minuscule compared to the total pool that is locked in the parent rock (primary and secondary aluminosilicate minerals). While the exchangeable pool reacts rapidly with soil solution, the pool of base cations contained in mineral rocks is released relatively slowly, owing to the resistance of the parent rock to decomposition and dissolution. Weathering rates vary greatly for different soil types. The most acid-sensitive soils produce enough base cations to neutralize only 0.02 to 0.05 grams of H^+ per square meter per year ($g\,m^{-2}\,year^{-1}$). In areas highly impacted by acid rain, for example Central Europe in the early 1980s, H^+ deposition rates as high as $0.60\,g\,m^{-2}\,year^{-1}$ have been measured.

At the lower end of the cation exchange buffering range (when β declines to approximately 5% and the pH hovers around 4.2), sulfate, the typical anion associated with acid rain, is immobilized as aluminum hydroxysulfate. If sulfate is the only strong acid anion present, the soils do not acidify any further. On the other hand, the presence of another strong acid anion such as nitrate will promote the onset of the aluminum buffering regime, which operates over the pH range from about 4.2 down to 3.8. Nitrate is a common component of acidic deposition and is also generated in natural biochemical processes. Moreover, it cannot be immobilized, except by plant uptake. Thus, when the rate of nitrate inputs to the soil exceeds the rate of plant uptake, aluminum becomes the dominant buffer.

The reaction governing aluminum buffering is:

$$Al(OH)_3(solid) + 3H^+ \rightleftharpoons Al^{+3}(aq.) + 3H_2O$$

At pHs above 4.2, the equilibrium is shifted to the left in the above equation, rendering aluminum mostly in the insoluble form of $Al(OH)_3$. Under strongly acidic conditions below pH = 4.2, the equilibrium is shifted to the right, causing neutralization of H^+ and dissolution of $Al(OH)_3$ to soluble Al^{+3}. Unlike the neutralization reactions in the cation exchange buffering range that generate dissolution of the bene cial base cations, Al^{+3} is extremely toxic. Releases of soluble Al^{+3} from forest soils have been documented in the northeastern US (Johnson *et al.*, 1981), and in Central Europe (Schulze, 1989). Since aluminum compounds are abundant in clay soils, buffer capacity in this range is rarely depleted.

The iron buffer range occurs only at an extreme stage of acidi cation soil solutions with pHs lower than 3.8. Its mechanism is similar to that of aluminum; H^+ is neutralized by dissolution of iron oxides. Soils in this extreme pH range often cannot support fauna and ora because they leach heavy metals and nutrients.

In summary, Figure 4 shows the progression of soil acidi cation that may occur as the buffering regime shifts from carbonate buffering to cation exchange buffering (at time t_1), and nally to aluminum buffering (at time t_2). The pace at which acidi cation occurs varies greatly, depending on both the rate of acidic inputs and speci c soil properties. Calcareous soils may never become acidi ed, but poorly buffered soils on slow weathering siliceous bedrock may acidify quickly.

The Acidification of Big Moose Lake: a Practical Example

The above discussion suggests that the process of reducing a soil s buffering capacity is not instantaneous, but rather occurs over time scales that vary widely, depending upon values of CEC, β, and the base cation replenishment rate, as well as on the rate of input of the pollutant to the soil. Very few studies are available that have monitored the progression of a soil from its natural, unpolluted state up until its loss of ability to buffer acidic inputs. The acidi cation of Big Moose Lake, shown in Figure 5, is one case where such information is available (Stigliani, 1988). The gure shows historical trends in pH, upwind emissions of SO_2, and sh extinctions. Emissions of SO_2, generated mostly by the burning of coal, are relevant because this molecule is converted to sulfuric acid (H_2SO_4) in the atmosphere. It is the dominant ingredient of acid rain globally, although nitric acid (HNO_3) plays a major role as well. The lake is situated in the Adirondack Mountains of New York State. The soils in this region possess relatively low buffering capacities, and, positioned downwind from western Pennsylvania

and the Ohio River Valley, the old industrial heartland of the US, they have experienced among the highest levels of acid deposition in North America (approximately 2.5 grams of sulfur and 0.14 grams of H^+ per square meter per year in the first half of the 20th century).

One may observe that the pH of the lake remained essentially constant at around 5.6 (except for a slight decline due to natural senescence) over the entire period from 1760 to 1950. Then, within the space of 30 years (1950 to 1980), the pH declined more than one whole pH unit down to about 4.5. Interestingly, the onset of the pH decline was not synchronous with the onset of sulfur emissions. Rather, it occurred 70 years after the beginning of SO_2 emissions, and 30 years after the emissions peaked at between three and four million tons of sulfur per year. Apparently, it required this amount of time for the acid inputs to deplete the buffering capacities of the soils in Big Moose Lake s watershed. Thus, beginning around 1950 atmospheric acid deposition

moved through the buffer-depleted soils and percolated into the lake with little or no neutralization. At that point, acid-sensitive fish species such as smallmouth bass, whitefish, and longnose sucker progressively disappeared, followed in the late 1960s by the more acid-resistant lake trout.

From a chemical point of view, one may infer from the shape of the historical pH curve that Big Moose Lake was the subject of an inadvertent titration experiment conducted over four generations of industrial activity. The coal-driven industrialization of the Ohio Valley supplied the acid inputs (mostly as H_2SO_4 formed from SO_2 released during coal combustion), and the soils of the watershed provided the supply of buffering chemicals. The watershed s natural buffering capacity delayed the deleterious environmental effects of coal burning for nearly a century.

Effect of Acidification on Soils Contaminated with Toxic Heavy Metals

Although acidification of soils and freshwaters is, in and of itself, deleterious to biota, the effect is magnified when soils are polluted by toxic heavy metals such as cadmium, copper, nickel, lead, and zinc. As cations, these metals compete with hydrogen and the base cations for cation exchange sites in clay soils. Under conditions of high pH, heavy metals in well-buffered soils are generally retained at the exchange sites, and the concentrations in soil solution are low. But, as illustrated in Figure 6 for the case of cadmium (Cd), mobilization measured as the leaching velocity increases by more than an order of magnitude as the pH declines from 7 to 4. (*Leaching velocity* is the rate at which dissolved cadmium would leach through a soil under standard rainfall conditions in the Rhine Basin. But cadmium in its dissolved form is equally likely to be taken up by crops. Thus, leaching velocity is an indicator of the metal s bioavailability.)

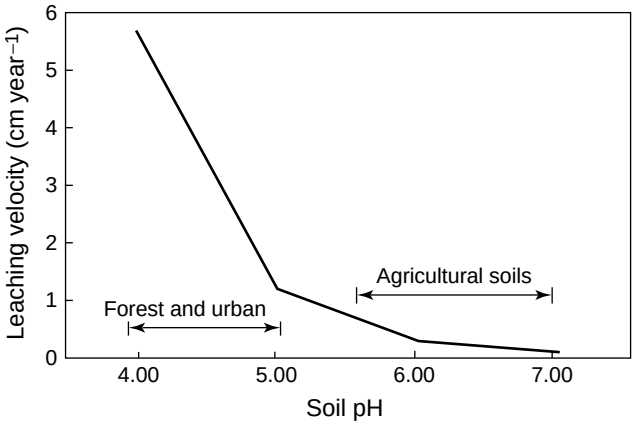


Figure 6 Leaching velocity of cadmium as a function of soil pH. (*Source:* P R Jaffe and W M Stigliani (1993) personal communication, International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria)

Therefore, at near-neutral pHs the soil will tend to store and accumulate cadmium. As the soil acidifies, H^+ increasingly replaces Cd^{+2} at the exchange sites, and the concentration of Cd^{+2} increases in soil solution. Once in the aq. phase, the cadmium is mobilized and biologically active. It is either transported downward to deeper soil horizons and groundwater (if the groundwater table is high), or taken up by vegetation.

Apart from clays, other common soil minerals are efficient adsorbers of heavy metals. Among the most important are the iron oxides (hematite, maghemite, magnetite) and oxyhydroxides (ferrihydrite, goethite, akaganeite, lepidocrocite, feroxyhite), and the manganese oxides (phyllomanganates, birnessite group of minerals).

The impact of pH, iron and manganese oxides, carbonate, and redox potential (Eh) on bioavailability and human health is obvious from Table 4, which compares cadmium contamination in two disparate regions. Shipham, England was the site of zinc mining from the 16th to the 19th centuries. In the Jinzu region, Japan some relatively minor silver and lead mining occurred in the 16th century, but activities greatly escalated in 1905 when a large zinc refining plant came into operation, generating high tonnages of wastes that were routinely dumped into the Jinzu River. Since cadmium is present in trace amounts (up to a few percent) in virtually all zinc ores, zinc mining and refining are almost always associated with cadmium pollution.

Cadmium in the Jinzu soils was more mobile (and thus more bioavailable) than cadmium in the Shipham soils owing to differences in CCPs (see Tables 1 and 4). The lower pH in Jinzu soils reduced their capacity to adsorb and immobilize the cadmium (H^+ bonds more strongly to soil adsorption sites than Cd^{+2}). Both high pH and high concentrations of manganese/iron oxides (which, as noted above, are particularly effective adsorbers of heavy metal cations such as Cd^{+2}) enhanced adsorption of cadmium in the Shipham soils. The difference in redox potential was another important factor. The submerged soils in the

Table 4 Impact of cadmium pollution on human health: a comparison of two regions

Soil parameter	Shipham, England	Jinzu, Japan
Maximum soil Cd concentration	>300 ppm[a]	70 ppm[a]
Soil pH	7.5	5.1
$CaCO_3$ concentration	6–14%	0.4%
Mn/Fe oxide concentration	High	Low
Eh	Moderate	Low
Health effects	Slight	Severe

[a] ppm, parts per million.
Source: adapted from Morgan and Simms (1988).

rice paddy fields provided a low redox environment under which the manganese/iron oxides were reduced to the soluble ions Mn^{+2} and Fe^{+2}. These chemical forms are incapable of adsorbing heavy metals, and, in any case, as soluble ions they are removed from the soils by rapid leaching.

Thus, even though soil cadmium levels exceeded 300 ppm (a thousand times higher than typical rural levels) at the English site, the soil conditions rendered most of the cadmium unavailable for crop uptake. Consequently, health inventories of inhabitants ingesting food grown in this area showed only slight effects attributable to cadmium (Morgan and Simms, 1998). In contrast, soils in the Jinzu region, with cadmium levels much lower at 70 ppm, favored the mobilization of cadmium and uptake into the rice. This resulted in the world s worst episode of cadmium poisoning, in which hundreds of persons became seriously ill from eating rice contaminated by polluted river water used for irrigation (Nogawa *et al.*, 1999).

Are there other similar disasters waiting to happen, and if so, how can they be identified? In order to forecast where heavy metal pollution may occur, two indicators are: 1) areas where high loads of heavy metals are accumulating; and 2) areas where changes in soil/sediment factors (CCPs) could plausibly occur.

Example of Agriculture in the Basin of the Rhine River

Agricultural lands of Europe s Rhine Basin provide one example of an area where toxic metals have accumulated (Stigliani and Anderberg, 1994; Stigliani *et al.*, 1993). This area, comprising about 10 million hectares, is among the most intensively cropped land in the world with a history of exceedingly high agrochemical inputs. One of the major sources of cadmium pollution in the basin has been PO_4^{-3} fertilizer. Like zinc, PO_4^{-3} ores carry cadmium as a trace impurity, and it is retained in the nished fertilizer product. PO_4^{-3} additions to cropland added an estimated 2100 tons of cadmium to the Rhine Basin soils between 1950 and 2000. Moreover, the basin is one of the most highly industrialized regions in the world with a long history of industrialization. Atmospheric emissions from iron and steel manufacturing, non-ferrous metal re ning, coal combustion, and municipal waste incineration have contributed large amounts of cadmium to soils via atmospheric deposition, estimated at about 2400 tons over the same 50 year period.

Figure 7 shows the estimated net buildup of average cadmium concentrations in the basin s agricultural soils, suggesting an overall doubling from around 350 g hectare^{-1} in 1950 to about 750 g hectare^{-1} in 2000. One obvious concern is whether the doubling in concentration poses health risks to humans eating food grown in the basin. The question is complex because, as emphasized in the earlier discussion, the mobilizable fraction of cadmium taken up by the crop will be less than the total amount present in the root zone of the soil.

Numerous publications have indicated soil pH is a major factor controlling both the total and relative crop uptake of cadmium (Kabata-Pendias, 2001). A sensitivity analysis was conducted to determine how plant uptake might change with changes in soil pH. The results are provided in Figure 8, which include the in uences of both changing pH

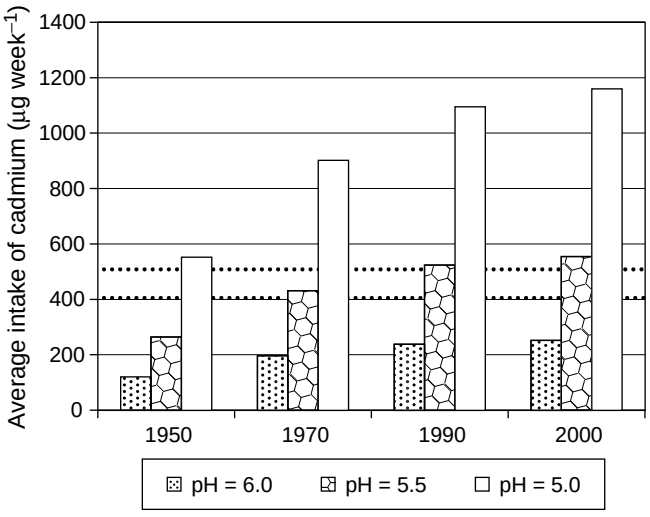


Figure 8 Estimated average dietary intake of cadmium from crops grown in the Rhine Basin as a function of time and changes in soil pH. It is assumed that the total food intake is from crops grown in the basin. Dotted horizontal bar refers to weekly limits of cadmium ingestion defined by the World Health Organization (WHO). (Reproduced from Stigliani and Jaffe, 1993)

and increasing concentrations of soil cadmium over time. Also in the figure is the recommended ingestion limit established by the WHO of 400 500 μg cadmium per week. The baseline of comparison is set at pH = 6.0. This corresponds to a typical pH in a well-limed soil. Holding this pH constant over time, the model estimates that in 1950 a person whose diet included only food grown in the basin ingested about 120 μg cadmium per week, rising to about 250 μg week^{-1} in 2000. Both of these values are well below the WHO limit. If cadmium inputs were to remain constant at the 2000 level, it would take about 100 years for the ingestion rate to reach 400 μg week^{-1}, and 1500 years if the cadmium were removed from the applied PO_4^{-3} fertilizer. Thus, it would appear, over the short term, that cadmium toxicity does not pose a serious threat to those consuming food grown in the basin.

On the other hand, the figure shows that even small declines in the soil s pH would cause relatively large increases in crop uptake. A decrease of just one-half a pH unit to 5.5 in 2000 would increase the cadmium intake to greater than 550 μg week^{-1}, and a decrease of one pH unit to 5.0 would increase cadmium consumption to over 1150 μg week^{-1}. It is noteworthy that in 1950 a one-half unit decrease in pH would not have caused intake to exceed the WHO standard. Even a one whole unit decrease to pH = 5.0 would have just barely exceeded the intake standard (to 552 μg week^{-1}). Thus, the overall effect of the creeping accumulation of cadmium has been to diminish the margin of safety with respect to enhanced crop uptake with declining pH. Moreover, these results illustrate that consistent liming of agricultural soils, particularly near industrialized regions with high cumulative inputs of heavy metals and acid deposition, becomes increasingly important in order to stabilize the soil pH and avoid sharp and sudden increases in crop uptake of trace metals.

Since agricultural lands in the basin are well managed, and probably frequently limed, excessive cadmium intake through diet may not be a pervasive problem. Perhaps of more consequence is the possibility of changing land use. Much of the agricultural land in the basin is situated in upland areas mixed with a high percentage of forests. Because this is not considered prime agricultural land, it may one day be abandoned and converted to forest. A historical example in England (Figure 9) shows that one outcome of converting farmland to forest may be a rather fast and significant decline in soil pH in the plow layer (top 20 cm that contain the roots of crops), particularly if liming were to cease altogether. This occurrence could trigger the mobilization of sizable amounts of cadmium out of the topsoil. Some would be taken up in new vegetation, which could harm wildlife grazing on the new growth. Alternatively, there may be increased leaching of cadmium to groundwater, which could pose a threat for drinking water supplies.

Table 5 Estimated time required for a 40% reduction in cadmium from agricultural soils of the Rhine Basin[a]

Soil pH	Yearly leaching rate (percent)	Years required for 40% reduction
6.0	0.14	365
5.5	0.3	170
5.0	0.6	85
4.5	1.4	36
4.0	2.9	17

[a] Assumes organic matter content of soil to be 3.5%, and that over the time periods indicated, there are no further net inputs of cadmium to the soil.
Source: adapted from Stigliani *et al.* (1993).

In any case, this analysis begs the question: How should the toxic load of stored cadmium be managed in the face of future land use changes? The impact will depend to a great extent on the rate at which the pollutant leaches from the soil, and this rate will depend on the rate at which the pH changes. Table 5 lists the estimated times required for 40% of the soil s cadmium to leach out of the top 40 cm as a function of pH. This corresponds to an average of about 300 g of leached cadmium per hectare. For an agricultural area of 10 million hectares, total leaching would amount to about 3000 tons of cadmium. If the leaching occurred in 17 years (pH = 4.0), then the annual inputs to lower soil horizons and groundwater would be around 176 tons year^{-1}.

In some areas, allowing the lands to acidify and the metals to leach out of the topsoils as quickly as possible may be a viable option. For a topsoil with a cadmium concentration of 720 g ha^{-1} to a depth of 20 cm, the dissolved phase concentration is estimated to increase from 0.0005 to 0.006 mg l^{-1} as the soil pH declines from 6 to 4.5. Given that the primary drinking water standard for

cadmium is $0.01\,mg\,l^{-1}$ and that mobilized cadmium will be further diluted as it moves into the saturated zone, a large-scale deterioration of groundwater quality due to an increase in cadmium concentration may not be expected. On the other hand, it should be noted that significant regional variations in soil cadmium concentrations exist within the basin, and there are undoubtedly hot spots of cadmium accumulations, particularly near intensive industrial areas where inputs from atmospheric deposition and other possible sources were excessive in former times. In these areas the heavily polluted soils that overlie shallow groundwater could contaminate the water. A safer option would be to slow down the rate of leaching by controlling the rate of pH decline. For example, if the pH of the soil were lowered to 4.5 or 5.0 rather than 4.0, the rate of leaching could be reduced by one-half to one-fifth, respectively. This option would provide more assurance that vulnerable areas of the Rhine Basin would not be subjected to unduly high releases of cadmium.

REDOX POTENTIAL AS A CAPACITY-CONTROLLING PROPERTY IN SEDIMENTS AND AQUATIC ECOSYSTEMS

Organisms oxidize organic carbon (CH_2O) to obtain energy for their metabolic needs. Oxidation of CH_2O requires a reactant called an oxidant. Only six such reactants control the oxidizing power of the biosphere. These are, in order of decreasing microbial preference: (1) oxygen (O_2); (2) nitrate (NO_3^-); (3) manganese dioxide (MnO_2); (4) ferric hydroxide ($Fe(OH)_3$); (5) sulfate (SO_4^{-2}); and (6) carbon dioxide (CO_2) (*see* Redox Potential, Volume 2).

As shown in Table 6, oxidants participating in redox reactions generate chemical by-products that affect the environment in different ways. With respect to nitrate, about 5% of the time NO_3^- goes to nitrous oxide (N_2O) rather than the fully reduced product N_2. N_2O is a strong greenhouse gas (200 times more potent than CO_2 on a per molecule basis). It also is the precursor for formation of nitric oxide (NO) in the stratosphere, which plays a key role in the nitrogen chain reactions that regulate, along with chlorine, the formation and destruction of stratospheric ozone.

The oxidized forms of manganese (MnO_2) and iron ($Fe(OH_3)$) are insoluble. They are common environmental minerals that coat the surfaces of soil and sediment particles. They play a useful role in the environment because they efficiently adsorb deleterious pollutants such as PO_4^{-3}, pesticides, and heavy metals. This situation changes, however, under anaerobic conditions and in the absence of nitrate. Microbial reactions using MnO_2 and $Fe(OH)_3$ as oxidants generate the reduced products Mn^{+2} and Fe^{+2}. These are soluble forms and all the pollutants that were adsorbed to the oxidized minerals are released to the water; i.e.,

$$\{MnO_2/Fe(OH)_3\} \Leftarrow \text{(adsorbed pollutants)} \longrightarrow$$
$$\text{[insoluble oxidized state]}$$
$$\{Mn^{2+}/Fe^{2+}\} \Rightarrow \text{(soluble pollutants)}$$
$$\text{[soluble reduced state]}$$

Anaerobic microbes using sulfate (SO_4^{-2}) as an oxidant generate the sulfide ion (S^{-2}) as the reduced product. On the negative side, sulfide reacts with H^+ to form hydrogen sulfide (H_2S) gas, a deadly toxin. Sulfate-reducing bacteria (SRB) are also major players in the generation of methylmercury (Benoit *et al.*, 2001; Benoit *et al.*, 1999), a toxic substance posing high risks to human populations ingesting contaminated fish. On the positive side, other heavy metal ions form insoluble metal sulfides that render them unavailable to biota. Because of the natural abundance of sulfate in seawater, these reactions are particularly relevant in marine ecosystems, and in wetlands and soils that were once under the sea.

Methogenic bacteria are found under anaerobic conditions at extremely low redox potentials, as may be found in wetlands, flooded areas, rice paddies, and the sediments of enclosed lakes. In these conditions, bacteria produce methane (CH_4) as well as CO_2 by the disproportionation of organic carbon. CH_4 is a strong greenhouse gas

Table 6 Environmental consequences of microbial redox reactions

Oxidant	Significant reaction products	Impact
O_2	CO_2	Depends on source of organic carbon[a]
NO_3^-	N_2O	A potent greenhouse gas; player in stratospheric ozone depletion
$MnO_2/Fe(OH)_3$	Mn^{+2}/Fe^{+2}	Mobilizes PO_4^{-3}, pesticides, heavy metals
SO_4^{2-}	H_2S; methylmercury; bound metal sulfides;	H_2S and methylmercury are highly toxic; metal-bound sulfides are immobilized
CO_2	CH_4	A potent greenhouse gas

[a] CO_2 is a greenhouse gas, but will not add to the greenhouse effect if the source of organic carbon from which it is generated is renewable biomass, which is usually the case.

whose potency is about 20 times higher than CO_2 on a per molecule basis.

Sudden changes in redox potential are possible, which can lead to deleterious environmental effects that appear as surprising or unpredicted occurrences. Some examples of this behavior are described in the following discussion.

Anoxia in Coastal Waters

Anoxia may occur in intermediate or deep waters of a gulf or a fjord with restricted circulation with surface waters. Figure 10 illustrates the underlying mechanism. As carbon is gradually added to the deeper waters, O_2 is progressively consumed by aerobic bacterial activity, and since there is little mixing with the surface layers, replenishment of O_2 from the atmosphere does not occur. (Oxygen in seawater is sufficient to oxidize only about 3.4 mg of carbon per liter.) Because seawater is rich in sulfate salts, the favored reaction under anaerobic conditions is sulfate reduction to hydrogen sulfide. Although H_2S is generally confined to the lower layers of seawater, during a storm event rapid mixing of the deeper, anoxic layers with the surface layer may subject marine life to high exposures of H_2S.

In coastal areas of Denmark in 1981 (Figure 11) and again in 1983, there were unprecedented depletions of oxygen and releases of hydrogen sulfide, causing the massive killing of fish. In a comprehensive investigation, the Danish National Agency for Environmental Protection concluded that meteorological conditions had triggered the two episodes, but that the underlying cause was the increasing over-fertilization of the coastal waters with an accompanying over-abundance of oxidizable organic carbon. The source of nutrient inputs was primarily runoff of nitrogen fertilizer from cropland (Schroder, 1985). The same phenomenon was observed in the 1980s in the Gulf of Saronik near Athens, Greece. The cause in this case was not nutrient runoff from cropland, but the direct disposal of raw sewage containing large concentrations of organic carbon.

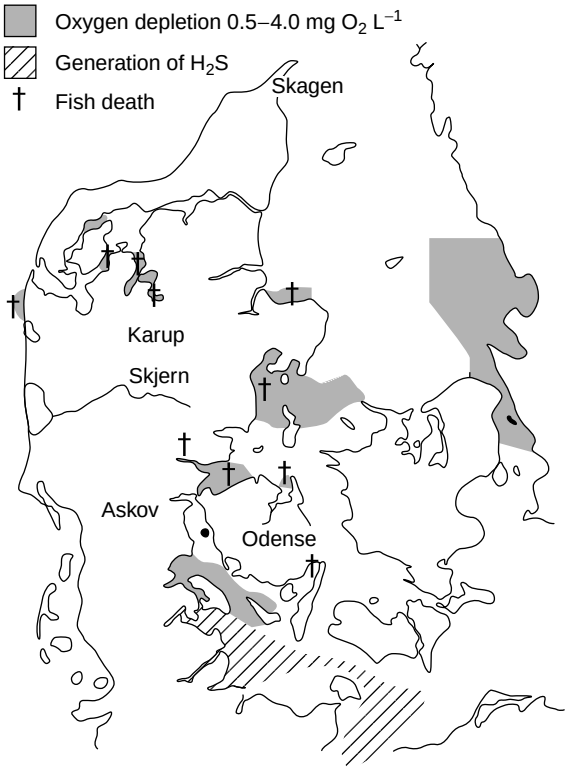


Figure 11 Scandinavian coastal areas where oxygen depletion, fish suffocation, and generation of H_2S were recorded in September 1981. (Reproduced from Miljostyrelsen, 1984)

Wetlands as Chemical Sinks

The episode of fish kills depicted in Figure 11 might have been avoided if the original coastal wetlands had not been drained. Wetlands are typically anoxic with large amounts of organic carbon; they create natural buffer zones for nearby fresh or marine waters by trapping inflowing nitrate (Figure 12a). Nitrate enters wetlands in agricultural runoff, and bacteria use it to oxidize stored carbon via the reduction of nitrate to N_2, which is vented to the atmosphere. By depleting nitrate before it enters the estuary, wetlands limit the excessive growth of biomass and subsequent anoxic conditions in the estuary waters. Furthermore, if wetlands are of marine origin, they are likely to contain high concentrations of sulfur in the form of reduced sulfide minerals such as pyrite. Under the redox/pH conditions prevalent in wetlands, these sulfides are highly insoluble and immobilized.

Draining wetlands radically changes the chemistry of the landscape (compare Figure 12a to Figure 12b). Nitrates enter receiving waters unmitigated, and sulfides, upon exposure to the atmosphere, are oxidized to sulfates. In contrast to sulfides, sulfate salts are appreciably soluble. Thus, draining wetlands can trigger the release of mobilized H_2SO_4, or heavy metal sulfates (e.g., $CdSO_4$, or $ZnSO_4$).

An example of this phenomenon occurred in a coastal area of Sweden near the Gulf of Bothnia, where wetlands were first drained in the early 1900s and more extensively in the 1940s for conversion to agricultural lands. The draining exposed sulfides to atmospheric O_2, resulting in their oxidation to H_2SO_4. Runoff from the drained land during storm events acidified nearby lakes. The pH in one of these lakes, Lake Blamissusjon, dropped from about 5.5 in the 19th century to a current value of around 3. Even though agricultural activities ceased in the 1960s, the lake has not recovered; it is widely known as the most acidic lake in Sweden. Similar problems occur on the global scale (*see* Acid Sulfate Soils, Volume 3).

Methylation of Heavy Metals

Bacterial methylation of metals such as mercury (Hg), arsenic (As), and tin (Sn) is a major concern because of the toxicities of the generated organo-metallic products. Most worrisome have been risks to the environment and human health posed by the methylmercury ion (CH_3Hg^+). Inorganic mercury, in any of its common valence states Hg^0, $Hg_2{}^{+2}$, and Hg^{+2}, is not particularly toxic when ingested; it tends to pass through the digestive system, although Hg^0 is highly toxic when inhaled. But CH_3Hg^+ is extremely toxic, regardless of its route of exposure.

Two CCPs, redox potential and microbial activity (see Table 1), play a crucial role in the generation of CH_3Hg^+ in the environment. The key microbes appear to be the same sulfate reducing bacteria (SRB) that thrive under low redox conditions in the presence of sulfate (Benoit *et al.*, 2001; Benoit *et al.*, 1999). In sediments contaminated with sulfur and mercury, SRBs transform mercury into the complex HgS^0. Because it is a neutral molecule (i.e., without charge), it is lipophilic and readily crosses bacterial cell membranes. Once inside the cell, SRBs synthesize methylmercury via the action of methylcobalamin, a derivative of vitamin B_{12} with a $CH_3{}^-$ anion bound to cobalt. The bacteria then eject the methylated product into the water, where it is absorbed by fish from water passed across their gills or from their food supply. CH_3Hg^+ forms CH_3HgCl in the saline milieu of a fish s biological fluids. This neutral complex crosses biological membranes, distributing itself throughout the fish tissue, in which the chloride is displaced by protein and peptide sulfhydryl groups. Because of mercury s high affinity for

sulfur ligands, CH_3Hg^+ is eliminated only slowly and is therefore subject to bioaccumulation when little fish are eaten by bigger fish. The phenomenon is the same as for DDT and other lipophiles, but the mechanism is different because mercury accumulates in protein-laden tissue (the edible part of the fish) rather than in fat.

Humans ingesting large quantities of fish contaminated with methylmercury suffer serious health effects. These include numbness in the limbs, blurring and even loss of vision, and loss of hearing and muscle coordination all symptoms of brain dysfunction resulting from the ability of methylmercury to cross the blood brain barrier. Likewise methylmercury can pass from mother to fetus, resulting in mental retardation and motor disturbance in newborns.

Minamata: an Exploded CTB

The Minamata catastrophe of the 1950s alerted the world to the gravity of mercury toxification (Harada, 1995). A polyvinylchloride factory near the fishing village of Minamata, Japan used Hg^{+2} as a catalyst, and discharged mercury-laden residues directly into the fishing waters of Minamata Bay. At the time, conventional wisdom viewed mercury in sediments as biologically inert. Given its high density (13.6 times the density of water), it was thought that the metal would sink out of harm s way into the sediments. The generation of methylmercury by bacteria was unknown. In a clear example of a CTB, the chain of events from the dumped inorganic mercury waste to its accumulation in the form of methylmercury in the aquatic food chain was unexpected and not understood until the damage was already done. The fish consumed by the villagers had accumulated methylmercury to levels approaching 100 ppm. Thousands were poisoned, and hundreds died from it (*see* Fisheries: Minamata Disease, Volume 3).

The Amazon Basin: a Ticking CTB

Although it would seem that one Minamata type disaster would be enough to curtail unsafe uses and disposal of mercury, an ongoing CH_3Hg^+ poisoning episode is playing out on a much grander scale in the Basin of the Amazon River in Brazil (Lacerda *et al.*, 1995) (*see* Mercury in the Environment, Volume 2). Over the past 35 years, prospectors have used mercury as an amalgam to extract trace amounts of gold from dredged sediments in several tributaries of the Amazon River. The gold is separated from the amalgam by roasting it in pans in open air (temperatures of 350 600 °C). It is estimated that about 130 tons of mercury each year are released in the Basin from this activity. Figure 13 shows the fate of the mercury in the amalgam; about half escapes to the air and the remainder is lost directly to the rivers in the form of metallic mercury (Hg^0). In the atmosphere it is oxidized to Hg^{+2} and returned to the land by atmospheric deposition. Runoff from the land causes the input of additional mercury into the numerous tributaries that feed into the Amazon.

In recent years, another important and more spatially diffuse source of CH_3Hg^+ has been discovered (Roulet *et al.*, 1998, 1999). The forest soils of the Amazon Basin possess naturally high concentrations of mercury (19 33 mg m^{-2} in the top 20 cm). The top organic layer adsorbs mercury on constituent humic materials, and the deeper mineral layers adsorb it on clays and iron oxides (the same mineral that locks cadmium in the soils of Shipham, England, as noted in Table 4). Conversion of these lands to agriculture by the slash and burn method burns off much of the top organic layer, leaving behind a mobile, mercury-containing residue. Additionally, the newly exposed mineral layer is subject to significantly enhanced rates of erosion, during which combustion residues, fine clay constituents, and associated iron oxides, all of which contain mercury, are preferentially exported from the soil and delivered to rivers via runoff during storms. The preferential leaching of clay and iron oxides acidifies the remaining soil, which further increases the rate of mercury exportation. Roulet *et al.* (1999) estimate that erosion of the top 1 cm of soil from deforested land mobilizes 500 3000 μg Hg m^{-2}. The total estimated deforested land in the Amazon is 270 000 400 000 km^2. This implies that erosion of 1 cm from this land mobilizes 135 1200 tons of mercury. Total soil loss of mercury from the top 10 cm could be ten-fold larger, although not all of the eroded mercury will end up in the basin s aquatic ecosystems.

The mercury load to the rivers has caused elevated concentrations of CH_3Hg^+ in carnivorous fish in the Madeira and Tapajos Rivers (Lacerda *et al.*, 1995; Lebel *et al.*, 1997). Compared to background mercury concentrations in fish of 0.2 μg g^{-1} in unpolluted freshwaters and less than

$0.15\,\mu g\,g^{-1}$ in oceans, carnivorous fish in the Madeira River have an average mercury burden of $0.70\,\mu g\,g^{-1}$, with a range from 0.07 to $2.89\,\mu g\,g^{-1}$; in the Tapajos River it averages $0.55\,\mu g\,g^{-1}$, with a range from 0.04 to $3.77\,\mu g\,g^{-1}$. These numbers are for the most part greater than the $0.5\,\mu g\,g^{-1}$ limit established by Brazilian legislation in 1975.

An overall assessment of the extent of human intake of CH_3Hg^+ and the prevalence of disease symptoms among exposed populations is hampered by lack of data, particularly in the most remote Amazonian populations. Nevertheless, existing and ongoing studies provide a worrisome indication that the sequence of steps from initial input of mercury to the aquatic environment to uptake of methylmercury in the human diet is already in place, at least in some parts of the basin.

Human hair is a good indicator for assessing CH_3Hg^+ contamination in exposed populations as it reflects blood concentration at the moment hair is being formed. Detailed hair mercury studies were conducted in towns and villages near the Mediera and Tapajos Rivers; the results are shown in Table 7. These values may be compared to $6\,\mu g\,g^{-1}$, the limit set by the Brazilian government for indication of mercury stress in an individual. In the Tapajos River Basin, the elevated levels in the upstream towns may reflect their proximity to the gold mining areas, and perhaps differences in diet relative to the downstream towns. The relatively low levels found in gold miners in the Madeira River Basin is probably due to a diet less oriented toward fish.

Research groups from Japan and Canada have independently concluded that Minamata disease has spread to remote fishing villages in the Amazon rainforest. The Japanese investigators examined 50 people from 10 villages around San Luis do Tapajos, near the site of gold mining activities. They found that three among this group suffered from nervous symptoms peculiar to Minamata disease, including fits of trembling (Pearce, 1999). The Canadians studied other remote sites and found evidence for early nervous system dysfunction associated with low level methylmercury exposure (Dolbec *et al.*, 2000; Lebel *et al.*, 1998, 1996). Both teams agreed that the Amazon poisoning appears to be widespread, and that additional victims will likely be found across the basin, particularly if gold mining and deforestation continue unabated.

The story of mercury toxification in the Amazon Basin is a clear example of a CTB. Three CCPs defined in Table 1 (CEC, pH, and OM content) regulate mercury adsorption in the soils, and deforestation affects all three CCPs (see Table 2). Eh, a fourth CCP, changes when mercury mobilization and erosion processes transport the metal from land to rivers. In this new aquatic environment, the soil particles sink to the sediments where anaerobic conditions (i.e., lower redox conditions) prevail. As noted earlier, the iron oxides are reduced to form dissolved Fe^{+2} accompanied by the release of the mercury, which then undergoes microbial methylation to form CH_3Hg^+. This is bioaccumulated in fish, and subsequently taken up in the human diet.

Table 7 Mercury concentrations in hair samples ($\mu g\,g^{-1}$) of populations and gold miners living near the Madeira and Tapajos rivers

	Mean	Range	N[a]
Madeira river			
Riverine population (females)	9.2	0.6–37.2	22
Riverine population (males)	11.8	0.5–71.4	25
Gold miners (males)	5.3	0.2–24.2	30
Tapajos river			
Upstream			
Jacareacanga	25.5	5.7–51.6	10
Brasilia Legal	26.0	4.8–151.2	55
Downstream			
Ponta de Pedras	11.4	6.9–22.3	46
Cameta	9.0	5.5–12.5	68
Santarem	2.7	0.7–6.5	11

[a] Signifies number of samples.
Sources: all data from Lacerda *et al.* (1995), except for Cameta, which is taken from Dolbec *et al.* (2000).

CTBs AND CLIMATE CHANGE

As noted in Table 2, the CCPs that determine CTB behavior by heavy metals in soils and sediments are affected by climatic factors and sea level rise. The climate/CTB pyramid shown in Figure 14 indicates the complex linkages between

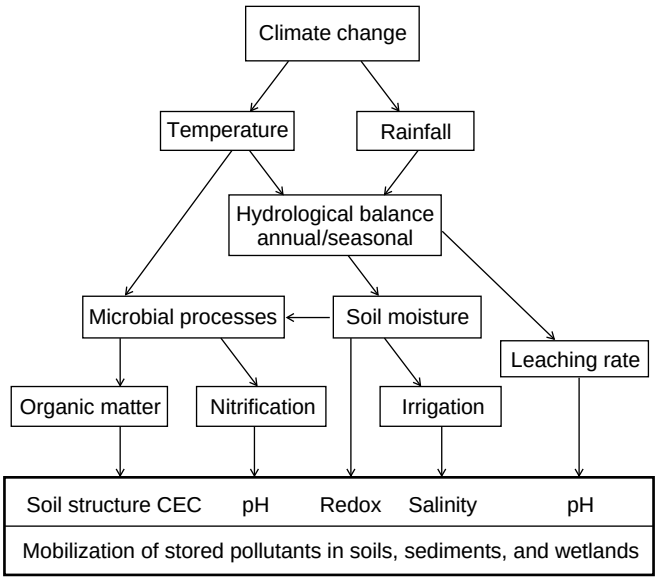


Figure 14 The climate/CTB Pyramid. (Reproduced from W M Stigliani and W Salomons (1993) Personal communication, IIASA, Laxenburg, Austria)

the two. A fundamental question is whether the chemicals stored in the biosphere under current climatic conditions will remain stored under a new climatic regime? Or will the changing climate transform current pollutant sinks into pollutant sources?

The Arctic regions of the world have served as vast sinks for volatile organochlorine chemicals, the most important of which have been PCBs, and the pesticides DDT, hexachlorocyclohexane (HCH), chlordane, and toxaphene. They concentrate in water, ice, and snow, and are taken up by fish, mammals, and humans. This has led to concern for the health impacts on native people who consume traditional foods such as seals and other aquatic species.

These organochlorine chemicals are transported to the Arctic via two mechanisms. One is normal atmospheric transport, which is a fast, episodic process. The second is the so-called *global distillation effect*, which is driven by the temperature gradient between the equator and the poles. By this process the chemicals generated in the Northern Hemisphere gradually make their way northward through evaporation, and then preferentially condense in the Arctic because of the coldness of the air and oceans. For example, it is estimated that the upper 200 meters of the Arctic Ocean contain 8.1 kilotons of HCH (Harner, 1997).

The greatest temperature increases from global warming are expected to occur at the poles. If this happens, the temperature gradient between the equator and the poles will decline, and the global distillation effect may not be as pronounced. As noted by Harner (1997) this may cause reduced polar-bound transport of pollutants, resulting in a redistribution of the organochlorines on a global scale. Moreover, if the warmer temperatures decrease the mass of ice cover, the melting ice would release pollutants long stored inside the mass.

On the other hand, polar warming may have the opposite effect (Harner, 1997). It could increase Arctic water evaporation and precipitation as snowfall, resulting in an accumulation of ice pack, and increased storage of the pollutants. Additionally, global warming is likely to increase ocean temperatures. This could lead to increased volatilization of the organochlorines in the tropical and temperate oceans, and enhanced transport to the colder regions (Iwata *et al.*, 1993). In any case, climate change is likely to redistribute the sources and sinks for these pollutants worldwide, in ways that are currently too uncertain to identify and quantify.

CONCLUSION

In recent decades, industrial societies have become acutely aware of the problems caused by the inappropriate disposal of toxic industrial wastes, the legacy of earlier generations that favored expediency over precaution. Typically these wastes were contained in metal drums and buried haphazardly in dumpsites near large industrial centers. The problem of course is that metal drums corrode over time and eventually leak, causing the stored chemicals to disperse to the wider environment. In actuality the storage, accumulation and eventual release of deleterious chemicals in the environment is far more pervasive than the single issue of toxic waste dumpsites and their management. In a broader sense, the mobilization and spread of chemicals in the biosphere now has global implications. No place on earth has been spared. In the Arctic, PCBs show up in the adipose tissue of polar bears. In the South Pacific, radioactive fallout threatens inhabitants of the Solomon Islands and Guam, while some islands of the Marshalls have been rendered permanently uninhabitable. In the remotest parts of the Amazon rainforest, mercury is poisoning indigenous populations.

The world over, activities that pollute the environment have been condoned and even encouraged in part because they provide instant service or profit without apparent adverse effects. This dynamic seems to work because the biosphere can store and accumulate toxic wastes out of harm s way for a long time. At first glance, this creates the impression of an environment that is infinitely forgiving of our wanton disregard for maintaining pollution within the limits imposed by natural buffering capacities.

But as discussed in this paper, these capacities are exhaustible either through direct saturation, or changing environmental conditions that shrink the ability in soils and sediments to store and retain harmful chemicals. Once pollutant inputs exceed natural limits for their safe assimilation we set in motion a retoxification process that is usually unpredictable and unexpected. Chemicals long stored are released and mobilized, analogous to the way toxic chemicals leak from corroded containers in a toxic waste dump. Leaks in the biosphere, however, are of more concern. Replenishment of depleted buffering capacities takes a long time, and in some cases the damage is irreversible. Moreover, the spatial scale of damage can be far more extensive than that of an industrial toxic waste site. Buffering capacities are valuable and often unappreciated natural resources. Maintaining them should be an important component in the evolution of an economy based on ecological sustainability of the biosphere.

REFERENCES

Benoit, J M, Gilmour, C C, and Mason, R P (2001) The Influence of Sulfide on Solid-phase Mercury Bioavailability for Methylation by Pure Cultures of *Desulfobulbus propionicus* (1 pr 3), *Environ. Sci. Technol.*, 35, 121 126.

Benoit, J M, Gilmour, C C, Mason, R P, and Heyes, A (1999) Sulfide Controls on Mercury Speciation and Bioavailability to Methylating Bacteria in Sediment Pore Waters, *Environ. Sci. Technol.*, 33, 951 957.

Brooks, H (1986) The Typologies of Surprises in Technology, Institutions and Development, in *Sustainable Development of the Biosphere*, eds W C Clark and R E Munn, Cambridge University Press, Cambridge.

Creger, T L and Peryea, F J (1994) Phosphate Fertilizer Enhances Arsenic Uptake by Apricot Liners Grown in Lead-Arsenate-Enriched-Soils, *HortScience*, 29, 88 92.

Dolbec, J, Mergler, D, Sousa Passos, C-J, Sousa de Morais, S, and Lebel, J (2000) Methylmercury Exposure Affects Motor Performance of a Riverine Population of the Tapajos River, Brazilian Amazon, *Int. Arch. Occup. Environ. Health*, 73, 195 203.

Hakanson, L, Nilsson, A, and Andersson, T (1988) Mercury in Fish in Swedish lakes, *Environ. Pollut.*, 49, 145 162.

Harada, M (1995) Minamata Disease: Methylmercury Poisoning in Japan Caused by Environmental Pollution, *Crit. Rev. Toxicol.*, 25, 1 24.

Harner, T (1997) Organochlorine Contamination of the Canadian Arctic, and Speculation on Future Trends, *Int. J. Environ. Pollut.*, 8, 51 73.

Hesterberg, D, Stigliani, W M, and Imeson, A C, eds (1992) *Chemical Time Bombs: Linkages to Socioeconomic Development*, International Institute for Applied Systems Analysis (IIASA) Executive Report 20, Laxenburg, Austria.

Iwata, H, Tanabe, S, Sakai, N, and Tatsukawa, R (1993) Distribution of Persistent Organochlorines in the Oceanic Air and Surface Seawater and the Role of Ocean on their Global Transport and Fate, *Environ. Sci. Technol.*, 27, 1080 1098.

Johnston, A E, Goulding, K W T, and Poulton, P R (1986) Soil Acidification During more than 100 Years under Permanent Grassland and Woodland at Rothamsted, *Soil Use Manage.*, 2, 3 10.

Johnson, N M, Driscoll, C T, Eaton, J S, Likens, G E, and McDowell, W H (1981) Acid Rain Dissolved Aluminum and Chemical Weathering at the Hubbard Brook Experimental Forest, New Hampshire, *Geochim. Cosmochim. Acta*, 45, 1421 1437.

Kabata-Pendias, A (2001) *Trace Elements in Soils and Plants*, third edition, CRC Press, Boca Raton, FL.

Lacerda, L D, Malm, O, Guimaraes, J R D, Salomons, W, and Wilken, R D (1995) Mercury and the New Gold Rush in the South, in *Biogeodynamics of Pollutants in Soils and Sediments, Risk Assessment of Delayed and Non-linear Responses*, eds W Salomons and W M Stigliani, Springer-Verlag, Berlin.

Lebel, J, Mergler, D, Brances, F, Lucotte, M, Larribe, F, and Dolbec, J (1998) Neurotoxic Effects of Low-level Methylmercury Contamination in the Amazonian Basin, *Environ. Res.*, 79, 20 32.

Lebel, J, Mergler, D, Lucotte, M, Amorim, M, Dolbec, J, Miranda, D, Arantes, G, Rheault, I, and Pichet, P (1996) Evidence of Early Nervous System Dysfunction in Amazonian Populations Exposed to Low-levels of Methylmercury, *Neurotoxicology*, 17, 157 168.

Lebel, J, Roulet, M, Mergler, D, Lucotte, M, and Larribe, F (1997) Fish Diet and Mercury Exposure in a Riparian Amazonian Population, *Water, Air, Soil Pollut.*, 97, 31 44.

Miljostyrelsen (1984) *Oxygen Depletion and Fish Kill in 1981 - Extent and Causes*, National Agency of Environmental Protection, Copenhagen.

Morgan, H and Simms, D L (1988) *The Shipham Report: an Investigation into Cadmium Concentration and Its Implications for Human Health*, Elsevier Science Publishers, Amsterdam.

National Research Council (1999) *Hormonally Active Agents in the Environment*, National Academy Press, Washington, DC.

Nogawa, K, Kurachi, M, and Kasuya, M, eds (1999) *Advances in the Prevention of Environmental Cadmium Pollution and Countermeasures*, Eiko Laboratory, Kanazawa, Japan.

Pearce, F (1999) A Nightmare Revisited: an Affliction Last Seen in the 1950s has Struck Again, *New Scientist*, 161(2172), 4.

Peryea, F J (1991) Phosphate-induced Releases of Arsenic from Soils Contaminated with Lead Arsenate, *Soil Sci. Soc. Am. J.*, 55, 1301 1306.

Roulet, M, Lucotte, M, Farella, N, Serique, G, Coelho, H, Sousa Passos, C J, de Jesus da Silva, E, Scavone de Andrade, P, Mergler, D, Guimaraes, J-R D, and Amorim, M (1999) Effects of Recent Human Colonization on the Presence of Mercury in Amazonian Ecosystems, *Water, Air, Soil Pollu.*, 112, 297 313.

Roulet, M, Lucotte, M, Saint-Aubin, A, Tran, S, Rheault, I, Farella, N, de Jesus da Silva, E, Dezencourt, L, Sousa Passos, C-J, Santos Soares, G, Guimaraes, J-R D G, Mergler, D, and Amorim, M (1998) The Geochemistry of Mercury in Central Amazonian Soils Developed on the Alter-do-Chao Formation of the Lower Tapajos River Valley, Para State, Brazil, *Sci. Total Environ.*, 223, 1 24.

Salomons, W (1995) Long-term Strategies for Handling Contaminated Sites and Large-scale Areas, in *Biogeodynamics of Pollutants in Soils and Sediments, Risk Assessment of Delayed and Non-linear Responses*, eds W Salomons and W M Stigliani, Springer-Verlag, Berlin.

Salomons, W and Stigliani, W M, eds (1995) *Biogeodynamics of Pollutants in Soils and Sediments, Risk Assessment of Delayed and Non-linear Responses*, Springer-Verlag, Berlin.

Schroder, H (1985) Nitrogen Losses from Danish Agriculture Trends and Consequences, *Agric., Ecosyst. Environ.*, 14, 279 289.

Schulze, E D (1989) Air Pollution and Forest Decline in a Spruce (*Picea abies*) forest, *Science*, 244, 776 783.

Stigliani, W M (1988) Changes in Valued Capacities of Soils and Sediments as Indicators of Nonlinear and Time-delayed Environmental Effects, *Int. J. Environ. Monit. Assess.*, 10, 245 307.

Stigliani, W M, ed (1991) *Chemical Time Bombs: De nition, Concepts, and Examples*, International Institute for Applied Systems Analysis (IIASA) Executive Report 16, Laxenburg, Austria.

Stigliani, W (1995) Global Perspectives and Risk Assessment, in *Biogeodynamics of Pollutants in Soils and Sediments, Risk Assessment of Delayed and Non-linear Responses*, eds W Salomons and W M Stigliani, Springer-Verlag, Berlin.

Stigliani, W M (1996) Buffering Capacity: its Relevance in Soil and Water Pollution, *N. J. Chem.*, 20, 205 210.

Stigliani, W M and Anderberg, S (1994) Industrial Metabolism at the Regional Level: The Rhine Basin, in *Industrial Metabolism - Restructuring for Sustainable Development*, eds R U Ayres and U E Simonis, United Nations University Press, Tokyo.

Stigliani, W M, Doelman, P, Salomons, W, Schulin, R, Smidt, G R B, and van der Zee, SEAT (1991) Chemical Time Bombs: Predicting the Unpredictable, *Environment*, 33, 4 9/26 30.

Stigliani, W M and Jaffe, P R (1993) *Industrial Metabolism and River Basin Studies: a New Approach for the Analysis of Chemical Pollution*, International Institute for Applied Systems Analysis (IIASA) Research Report 93, Laxenburg, Austria.

Stigliani, W M, Jaffe, P R, and Anderberg, S (1993) Heavy Metal Pollution in the Rhine Basin, *Environ. Sci. Technol.*, 27, 786 793.

Stigliani, W and Salomons, W (1993) Our Father s Toxic Sins, *New Scientist*, 140, 38 42.

Stigliani, W M and Shaw, R W (1990) Energy Use and Acid Deposition: the View from Europe, *Ann. Rev. Energy*, 15, 201 216.

ter Meulen, G R B, Stigliani, W M, Salomons, W, Bridges, E M, and Imeson, A C, eds (1993) *Proceedings of the European State-of-the-Art Conference on Delayed Effects of Chemicals in Soils and Sediments*, 2 5 September 1992, Veldhoven, The Netherlands, The Foundation for Ecodevelopment Stichting Mondiaal Alternatief , Hoofddorp, The Netherlands.

Ulrich, B (1981) Theoretische Betractungen des Ionenkreislaufs in Waldokosystemen, *Zeitschrift P anzenernährung Bodenkunde*, 144, 647 659.

Ulrich, B (1983) Soil Acidity and its Relation to Acid Deposition, in *Effects of Accumulation of Air Pollutants in Forest Ecosystems*, eds B Ulrich and J Pankath, Workshop Proceedings, Gottingen, Federal Republic of Germany, 16 19 May 1982, D Reidel, Dordrecht, The Netherlands.

Rice and its Spread: Double Cropping; New Varieties – Environmental Consequences and Methane Gas, Sustainability

R D HILL

The University of Hong Kong, Hong Kong

*Rice the world s most popular food for man and beast! Rice the word for food in a dozen Asian languages! Amongst the grasses (*Graminae*) rice is pre-eminent. Roughly three billion people and countless animals eat it as a major source of carbohydrate. It also contains fats and proteins, especially in its unpolished, thus more nutritive form. Of the two species of cultivated rice, that of Asian origin,* Oryza sativa*, L., is by far the most widespread. The indigenous West African species* O. glaberrima *has spread little from its homeland and in the continent as a whole,* O. sativa *is now at least as common as its African cousin. (The North American wild rice,* Zinzania aquatica*, is a very distant relative of the cultivated species though belonging to the same botanical tribe, the Oryzeae). Over the last four decades rice has steadily expanded its share of the global grain production area to 21% and to 27% of total grain produced from 18% and 26%, respectively, in the early 1960s.*

Physiologically, rices are swamp-dwelling plants and one of the major consequences of their domestication has been the requirement both to adapt many different kinds of tropical and temperate environments to their needs and to select varieties for a range of conditions. The yields of cultivated rice are highly responsive to increased labor inputs aimed at optimally matching the characteristics of each of the many thousands of varieties of O. sativa *especially to local micro-environments. Thus transplanting, which reduces competition with weeds, close planting, frequent weeding, tight control over water-supply and general good husbandry, each implying higher levels of environmental modification and control, are needed to obtain maximum returns. The result may be an almost total remaking of the habitat, especially in wet-terrace cultivation. The other major response has been to adapt the rice plants physiology to other environments, slightly saline ones along coasts or drier ones on hill slopes, a process probably occurring later than initial domestication though still prehistoric. Such selection has continued, of course, especially since the 1950s when the problem of feeding Asia s rapidly expanding population became pressing. In order to maximize yields, the basic scientific strategy has been to breed high-yielding varieties (HYV s) that require a high level of inputs and precise ordering of the conditions in which they are grown.*

From a homeland in East and Southeast Asia, O. sativa *spread widely, beginning perhaps eight thousand years ago. It reached the Middle East and possibly coastal East Africa two millennia ago, Southern Europe patchily from the fourteenth to the nineteenth century and was well established in the Americas by the eighteenth century.*

Systems of rice production are quite varied, though probably less so now than before the wide adoption of HYVs. Rice is grown commercially in subtropical areas of the United States, Australia and Southern Europe as an element in rotational cropping. It is grown as a mono-crop for the subsistence of millions of Asian families. The strength of this direct support for life is re ected in the fact that only about 4% of rice produced enters international trade. Most is consumed directly as human food but some feeds animals: in Australia sheep lightly graze the growing plants, adding dung to the soil. Rice polishings and poor-quality grain are widely used to feed pigs and poultry. Some is used industrially, about 40% of the US crop for beer, while in Asia rice wine is commonly made both domestically and commercially. A major by-product, rice husk, is highly siliceous and is of little use except as fuel and for cleaning the interior of jet engines.

The total amount of land modi ed for and by the cultivation of rice is not known. Data from the Food and Agricultural Organization of the United Nations (FAO) indicate a current harvested area of about 150 million hectares but this includes an element of double counting since signi cant areas are cropped with rice twice annually, with some, mainly in Java, being used even more intensively. International Rice Research Institute (IRRI) data indicate that about 24% of the rice area is double-cropped with rice while a further unknown fraction also grows rice and another crop. The area currently growing rice under shifting cultivation is not known either but given that the customary bush-fallow is seven to ten years after a year or two of cropping, the area now affected by clearance for this dry form of cultivation must be three to ten times greater than the area actually growing rice at any time. The degree of environmental impact by rice cultivation varies with the intensity of land use, ranging from an almost total remaking of the environment in wet-terrace cultivation such as in Java, North-central Luzon and many parts of Southern and Southwestern China to relatively little change in long-fallow systems of shifting cultivation such as survive in remote areas of Indochina. Lowland systems, where soils are wet and high in organic matter, are a major global source of methane (CH_4), an important greenhouse gas.

ORIGINS

The genus *Oryza* comprises 20 or so species of aquatic grasses that are native to permanently or seasonally flooded sites in tropical and subtropical regions around the world so long as these are not particularly saline or acid. In detail their distributions show remarkable disjunction, the reasons for which are unclear. Notable is the disjunction between the two cultivated species, *O. glaberrima*, whose home is in West Africa and *O. sativa* whose home is a matter of debate but almost certainly in one or more parts of a large region extending from the lower Yangtze valley in central China, through South China, mainland Southeast Asia and the margins of the Bay of Bengal as far west as the Indian state of Orissa. In this zone grow a number of possible wild ancestors, of which the closest relatives of *O. sativa* are *O. nivara*, a wild annual, and *O. perennis*, a perennial. The latter is thought to have contributed genetically to floating varieties of *O. sativa* that have the capacity to elongate their internodes in response to rapid rises in water level.

The environment in which rice was domesticated can be established with some certainty, for all rices are tropical and subtropical swamp grasses. But where were such swamps, given that some authorities argue that such regions naturally have a forest climate? Grassy swamps can have been of very restricted distribution, perhaps occupying lake margins and flood plains where trees could become established only with difficulty by reason of fluctuating water levels and the deposition of flood debris. But Paleolithic peoples almost certainly possessed fire and unquestionably used it directly to assist in hunting and may also have used it, consciously or unconsciously, to modify tree vegetation, for in monsoonal climates, standing forest will burn in dry seasons. Grassy swamps, even with standing water will also burn in the same season. The effects would likely have been to increase carbohydrate sources for gathering and to increase the number of grazing animals for hunting. There is as yet insufficient palynological evidence to confirm such a scenario but in historical times it is known that very extensive grassy swamps existed for instance in the Brahmaputra-Ganges delta and along the Chao Phrya river.

What is certain is that rices cannot have been domesticated in temperate climes, whether in higher latitudes or at higher altitudes in the tropics. Nor can they have been domesticated in the equatorial zone, for traditional varieties of *O. sativa* are highly photosensitive, requiring long days to initiate flowering and the formation of the grain. In the Peninsular Malaysian state of Negeri Sembilan, for instance, experiments in the nineteenth century showed that rice planted out of season simply remained at the initial vegetative stage of growth, flowering only when days lengthened. (The varieties used had probably already been selected for sensitivity to small differences in day-length, at that location barely half an hour). Equally, neither domesticated rice nor agricultural systems based upon it could have arisen on hill-slopes for rice is physiologically a swamp plant so that drought-tolerance is necessarily not a feature of original domesticates. Much the same considerations apply to brackish-water tolerance, a capability clearly bred into rice by selection. However, with some exception in respect of cold-tolerance, just when these characteristics were acquired is simply not known, in part because these questions have been little studied. The Japanese case is instructive. Arriving in the south about 2300 years ago, rice moved slowly north, not reaching Northern Honshu until the seventeenth century and Hokkaido two centuries later, a process clearly dependent upon the development of successively more cold-tolerant varieties and of methods to warm irrigation water using the sun s heat.

The key to domestication was the selection of rice that did not shatter, shedding the grain when ripe. That selection did not necessarily occur by the time that rice came to be cultivated. Oka and Chang (1959) argued, on the basis of experimental study of the effects of human impact upon wild rices, that, ••• the degree to which

genotypes of a population [of wild rices] approaches that of the cultivated form is related to the degree of disturbance by man ••• Should this be so, then within the range of wild progenitors, domestication need not have been a once-for-all occurrence.

However, most prehistorians argue that such once-for-all domestication is the norm rather than the exception. But what is meant by domestication? Loss of shattering is one of the characteristics but this can scarcely be expected to show up in the archeological record. Loss of awns (long hairs on the grain), on the other hand does show up in the record. Several stages of domestication may be envisaged. First, desired rice plants would have been protected but no more than this. Second, grain would have been deliberately sown but with little or no clearing of the vegetation to receive it. This practice survived into the nineteenth century in the Malay Peninsula, as *padi terbuang* literally thrown rice. Yields would have been low because of competition with other plants so that clearance by fire and/or weeding unquestionably would have increased them. This may well have been a third stage. A further step in enhancing yields by reducing competition with weeds would have been tillage, probably at first using the hoe, an implement still used in preference to the animal-drawn plough in many parts of Asia into the 1960s. Other yield-enhancing technological advances include transplanting, fertilizer application and close control of water by irrigation, advances which have been made at very various times and places down to the present time. At some point, probably quite early, *Oryza sativa* came to be treated as an annual, for even today rice stubble will produce a second or *ratoon crop* if permitted to do so though yields are much less than the first time around. A comment in the Chinese *History of the Chin Dynasty (AD 265- 419)*, referring to the Mekong kingdom of Funan, suggests just such a practice. The general picture is thus one of increasingly intensive production, doubtless with regressions and diversions on the way, each involving increased environmental control and biological simplification as rice, once one crop amongst many, has become a highly labor-intensive monoculture.

SPREAD

What do prehistory and the archaeological record have to say about the broad picture presented here? Because of the wide distribution of wild and weedy races of *Oryza* in east and Southeast Asia and because of extensive hybridization, an accurate delimitation of a place of origin of the domesticated plant on the basis of living plants is not feasible. This is so not least because wild rices are common weeds in rice-fields. Thus in Peninsular Malaysia are found *Oryza fatua* and *O. minuta*, both bird rice to indigenes, both, like *O. sativa*, light-lovers, both crossbreeding with each other and with cultivated rice. (A third wild rice, *O. ridleyi*, is a shade-loving dweller of forest margins that has probably played no part in the game).

Theories are complicated for both pragmatic and political reasons. India, mainland Southeast Asia and China have all been touted as earliest homes of domesticated rice in Asia, following a once-for-all interpretation. A serious technical difficulty is that various characteristics of grain shape and size are not necessarily clearly diagnostic of domestication, a conclusion reached by Douglas Yen concerning imprints on pottery found at Ban Chiang, Thailand, dated about 2000 BC and by TT Chang with respect to Philippines material of 1240 BC. Another problem is that the extraction of plant remains from excavated debris at archeological sites, and analysis by C^{14} dating, has not necessarily been carried out, so that claims of great antiquity may not be firmly established. If Chinese claims are accepted, rice was cultivated in the Huai valley c. 8000 9000 years ago and a thousand years later in the Yangtze valley. These areas are on the margin of the generally accepted zone of Asian wild rices rather than at its core as postulated by the noted Russian botanist Vavilov. This apparently was the beginning of at least some of the processes of rice-induced environmental change that still continue. Rice impressions of Neolithic age have been found near Allahabad, India, dating 6000 7000 thousand years ago and from Thailand some 5000 years ago but it is not clear that these are truly cultivated varieties. In Southern China the earliest sites with the rice remains date from 4000 to 5000 years ago but that relatively moist region, like Southeast Asia, is not one that is conducive to the preservation of remains, unless carbonized, and it is possible that future work will turn up evidence of earlier use (Figure 1).

So far as progressive intensification of production and environmental impact is concerned, the evidence of archeology is complex and confused. What are interpreted as stone hoe blades appear early in the record but are not necessarily associated with preserved evidence of rice. Ploughs are certainly later but, like rice itself, appear to have spread at extremely uneven rates, having been introduced or possibly reintroduced, in Sarawak as recently as the nineteenth century. Irrigation works are notoriously difficult to date though Sri Lanka s are claimed to be at least 1000 years old. In the deltaic regions, irrigation works were probably long preceded by drainage works as for example those at Oc-Eo on the lower Mekong dating from the fifth century AD. In such areas the means of raising water from drains to fields may be quite late, the Chinese pedal-powered pump, probably little more than 1500 years old in its homeland, having been used for no more than 200 years in the lower Mekong at best.

Once into historical times, we are on firmer ground in attempting to document the spread of rice, though not necessarily the spread of related engineering and water management strategies. The crop s spread to the northeast

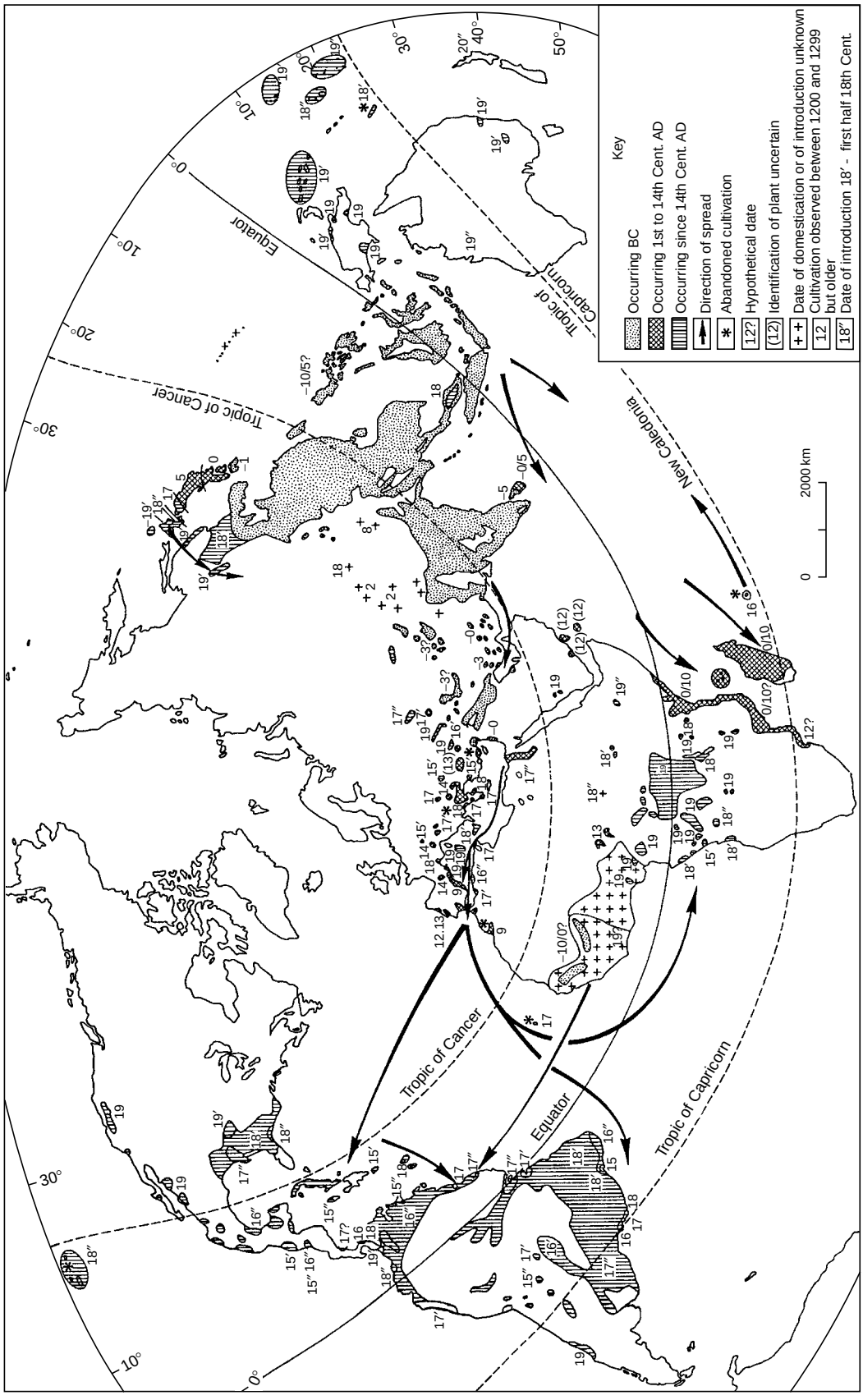


Figure 1 The global spread of rice (inventaire geographique et chronologique pour un atlas d'histoire mondiale (Paris, 1968) updated by R Hill)

has been mentioned already with its arrival in Japan about 300 BC. Its emergence in insular Southeast Asia, in Java, probably dates back to no more than 2000 BC, at the very earliest but its penetration is clearly very uneven. It did not reach Nias and Enggano, islands off Western Sumatra, until colonial times. In upland central peninsular Malaysia, aboriginal peoples have a story of a millet king, who, after a great battle was displaced by a rice king, whilst to the south of the Peninsula, late nineteenth-century aboriginals were growing the American tuber *manioc* rather than rice, the latter not being reported until the 1960s.

No one really knows why rice did not spread from Southeast Asia into the Paci c along with the Austronesian ancestors of present-day Polynesian, Micronesian and other Paci c Island peoples. The migrating Austronesians may not have had it or did but discontinued its cultivation. The latter is more likely for early in the rst millennium AD people from what is now Indonesia crossed the Indian Ocean to colonize Madagascar. Their descendants are the Malagasy people and rice elds are a major feature of the island s cultural landscape. Farther north, rice reached Western India quite early but encountered the barrier of dryness, no longer being capable of being grown in rain-fed conditions. By about the second century BC, both Chinese and Greek sources reported rice cultivation in Persia and Mesopotamia (now mainly Iraq). The spread of rice into Europe was by no means an even east west progression, certainly being in Moorish Spain by the 10th century AD. Its Arabic name *eruz* has entered many European languages, for instance as *arroz*. The earliest cultivation in Italy s Pontine marshes dates only from the 15th century where it was long hindered by the existence of endemic malaria. The growing of rice under unirrigated conditions and/or as part of a crop rotation in Europe seems to have begun in the eighteenth century.

It was from Europe that *O. sativa* spread to tropical and subtropical parts of the Americas, beginning in the sixteenth century. A further route of dispersal was from Europe to parts of central Africa and by the nineteenth century it was also being grown in the West African homeland of *O. glaberrima*. That species, however, has seemingly not spread beyond its zone of domestication though little is known of it generally. The growing of rice in Australia, New Zealand and the Paci c islands is mostly little more than a century old, the Australasian system being particularly interesting because it involves rotational farming and grazing by sheep. (Australia is a signi cant rice exporter).

The discussion thus far has concentrated upon the spread of rice to regions previously lacking it. But even in regions that have possessed rice for millennia, the area under rice has expanded enormously, especially over the last two centuries, partly in response to population growth, partly in response to growth in international trade and the international division of labor. In Asia, for example, the rice-eating population has more than doubled over the last 40 years. While there is good evidence of a per person reduction in rice-consumption in wealthy Asian regions, notably Japan, Singapore and Hong Kong, the general trend has been an expansion of demand and production, at rst largely met by expansion of the cultivated area without much increase in yield. This seems to have been at about 1.0 1.5 tonnes per hectare ($t\,ha^{\bullet\,1}$) for centuries, perhaps millennia.

Much of the expansion in countries that were already rice consuming took place in Southeast Asia as the imperialist powers insisted upon opening markets. Thus Lower Burma, centered on Rangoon (Yangon) and the Chao Phrya plain, centered on Bangkok, saw large and rapid expansions beginning from the mid-nineteenth century. This was followed by similar growth in the Mekong delta as French and Sino-Vietnamese capital owed in at the very end of the nineteenth century and early in the twentieth. In the Chao Phrya and Mekong, the initial stage was canal cutting to ensure adequate drainage, followed by massive deforestation. At rst, cultivation intensity and yields of grain were low but within 5 10 years, the remaining debris rotted away, and tillage and transplanting replaced broadcast sowing. Yields rose for a period and then fell to a stable level around $1\,t\,ha^{\bullet\,1}$ or a little more. These clearances must have involved a major addition to atmospheric carbon dioxide (CO_2) though no estimates have been made and increases would have been masked by simultaneous increases in European and American industrial CO_2 emissions. Ironically, much of the rice produced went to feed workers in just those CO_2 producing industries though some also owed to feed plantation workers in the region, themselves by the late nineteenth century responsible for major forest clearances for perennial crops such as coffee, tea, coconut, and later, rubber and oil palm.

Clearances for rice have decreased markedly since about 1950 though data do not exist. (The standard statistical series are for area harvested and for production.) At the same time cultivation intensity has increased substantially in Southeast Asia especially, though two crops a year have been usual in most of the main production areas of the Indian subcontinent, China, Northern Vietnam, Java and Bali or possibly three for four centuries, sometimes longer. The consequence is that it is impossible to say how much new land is being developed for rice, as it unquestionably is in parts of West Africa especially, or how much is going out of production as some is. In the Southern Chinese province of Guangdong for example, factories now occupy highly productive rice lands. Other areas are planted to less labor demanding crops or simply abandoned as farmers leave the land for better-paying urban

Table 1 Harvested area by major region, mid-1950s, 1976, 1996[a]

Region	1950s: Area (000 ha)	Percent	1976: Area (000 ha)	Percent	2000: Area (000 ha)	Percent
Africa	3000	2.8	4358	3.1	7762	5.0
N.&C. America	1299	1.2	1771	1.2	1892	1.2
South America	2840	2.6	7709	5.4	5728	3.7
East Asia	31 968	29.6	41 452	29.1	33 880	22.0
Southeast Asia	25 576	23.7	33 527	23.6	42 519	27.7
South Asia	42 213	39.1	52 225	36.7	60 077	39.1
Rest of Asia	583	0.5	777	0.5	1139	0.7
Europe	385	0.4	377	0.3	614	0.4
Oceania	30	–	86	0.1	155	0.1
Total	107 894	99.9	142 282	100.0	153 766	99.9

[a] East Asia: China, Japan, N.&S. Korea; Southeast Asia: Myanmar, Thailand, Cambodia, Laos, Vietnam, Indonesia, Malaysia, Philippines; South Asia: India, Nepal, Bangladesh, Sri Lanka, Bhutan.
Source: FAO Production Yearbooks and http://apps1.fao.org accessed May 25, 2001.

jobs Guangdong again, Malaysia and parts of Java are examples.

From the standpoint of environmental effects, the period of large CO_2 additions from land clearance for rice are long past. But CH_4 emissions continue to expand as crop intensities increase. Since the mid-1950s the harvested area, production and grain yield have risen substantially (Table 1) though with the widespread adoption of dwarf HYVs, it is possible that actual phytomass production per unit of area has fallen somewhat.

Since 1950 the global harvested area has increased by 43% but by only 8% in the 24 years to 2000 (Table 2). While the distribution amongst the main rice-growing regions has remained much the same, the once less-important regions Africa, South America, Rest of Asia and Oceania (mainly Australia) have increased their shares substantially, more than doubling their rice areas in four decades. In these the cultivation intensity is low usually a single crop annually, whereas elsewhere, except in North America, at least two-fifths of the rice land grows two crops a year. (A further but unknown proportion grows a summer rice crop and some other crop in winter).

Table 2 Average yields (t ha^{-1}) by major region, mid-1950s, 1976, 1996

Region	1950s	1976	2000
Africa	1.4	1.8	2.2
N.&C. America	2.5	4.0	5.9
South America	1.6	1.8	3.6
East Asia	2.8	3.8	6.2
Southeast Asia	1.4	2.1	3.5
South Asia	1.3	1.7	3.0
Rest of Asia	1.8	3.1	2.8
Europe	4.4	4.4	5.1
Oceania	3.5	5.2	9.2
Global average	1.8	2.4	3.9

Source: As for Table 1.

Yields are another matter. These reflect genetic characteristics, cultivation practices, water supply and control as well as sunlight. It is notable that yields are especially high in sunny regions, even if not tropical, other things being equal. Temperate production areas have longer daylight hours in summer. Some are rich and can afford to make free use of artificial fertilizers, especially nitrogenous ones, herbicides and insecticides. Only in China are comparable yields obtained, there by the heavy use of organic fertilizers, with significant consequences for CH_4 emissions. Globally, yields have doubled in four decades, the most rapid increase being between 1976 and 1986. Below average rates of yield increase occurred in Africa, where yields are low and also in the Rest of Asia where they are moderate. It is unlikely that yields will increase much beyond the 5.5 6.5 t ha^{-1} already achieved in temperate regions and in East Asia, for there is clear evidence of diminishing returns to increased and improved inputs. The United States, however, should be able to achieve Australia s remarkable 9.7 t ha^{-1}.

Overall, in most parts of Asia the spread of rice has slowed dramatically within the last two decades, in its older production areas a century or more ago. One reason is the high cost of development roughly US$20 000 ha^{-1} at which it is scarcely possible to break given current prices. Where the opportunity cost of labor is low, as in much of Africa south of the Sahara and where significant areas of alluvial lowland remain available for development, such as the Niger delta and in Sierra Leone, pioneer clearance for rice may continue. In Southeast Asia, where the opportunity cost of labor may be diminishing somewhat, the concern is that remaining upland, mainly secondary forest areas, may be attacked for rice-based shifting cultivation or that scrub fallows in existing areas where this system is practiced will be further reduced. This would almost certainly lead to enhanced upland erosion and sedimentation in the lowlands.

RICE SELECTION, BREEDING, HIGH YIELDING VARIETIES

As noted earlier, cultivated rice is by origin a plant of seasonal or permanent swamps. Selection has aimed to match varieties with varied environmental characteristics day-length of growing season, depth of standing water (or none on hill slopes), speed of change in water depth, maturation period, drought resistance, fertilizer response, degree of shattering, color, taste, milling characteristics, degree of stickiness and a dozen other factors. The consequence is a huge number of varieties, at least 120 000. Botanically the main distinction in *O. sativa* is between *indica* and *japonica* strains. The former are mostly long-grained and photoperiod sensitive, while the latter have plumper grain and mostly are less photo-period sensitive, maturing at a more or less fixed time after planting. (Some systematists also recognize a *javanica* sub-species). Both *indica* and *japonica* rices have existed since very early times, the latter, despite their name, originating in the general vicinity of the Yangtze valley.

Well-developed panicles, even individual seeds, were probably selected as seed-rice from early times. The use of brine to separate good and bad, the former sinking, was well known in China at least four centuries ago. Purposeful selection has resulted in making it possible to grow rice in a great range of environments, most however, having in common the need for growing-season temperatures to be frost-free and over about 20 °C and below about 35 °C, above which crops may suffer heat stress. Nevertheless, some varieties will still grow, though not well, between 4° and 21 °C or at temperatures over 35 °C so long as there is sufficient water. Prior to the era of scientific rice breeding, it is difficult to establish where and when innovations were made. Initial selection for resistance to the stresses of dryness and cold are clearly prehistoric. The same may not be true of tolerance of brackish water for earliest reports come from thirteenth-century China. Perhaps the most important early adaptation, for a short growing-period, was that of Champa rices named for their homeland in Southern Vietnam. These were known to the Chinese by the first century AD. Reaching Fujian before the 11th century, they were taken north to the Yangtze in 1012. Though at first low yielding, but needing only 90 120 days to reach maturity, further selection raised yields. Thus, instead of being a type to be grown when failure of the rains forced late planting, the descendants of Champa rices permitted twice annual rain-fed cropping, a practice widespread in Southern China and Northern Vietnam by about the 16th century.

Selection for survival in deep-water environments is another important process, though its history is obscure. So-called floating rices characterize such environments. As floodwaters deepen, these varieties elongate at the internodes, the plant sometimes lengthening as much as six or seven meters. When the floods recede and the plants touch the ground again, they rapidly go through the final stage of growth before harvest. Such rices are particularly important in Bangladesh and in the back-swamps of Myanmar, Cambodia, Thailand and Southern Vietnam. Similar rices also occur in Mali, Guinea and Sierra Leone. While such deep-flooding areas are a home of *O. perennis*, the evolution of non-shattering but elongating varieties of *O. sativa* allowed the reclamation of such swamps for rice cultivation.

Modern rice breeding was greatly promoted by the Japanese who spread improved varieties, mainly *japonicas*, amongst farmers. Much of their work was based in Taiwan, one notable high-yielding dwarf variety, Taichung Native No. 1, being widely spread in the region and forming a basis for later HYV development. The major impetus to rice-yield improvement came in the 1960s with the realization that Asia s substantially rice-eating population was growing so fast that it was likely to double within 25 years or less. This somewhat belated realization resulted in a strategy of pushing for maximum yields even at the expense of vulnerability to disease and the need for considerable precision in the application of fertilizers, water control and enhanced requirements for weeding, disease and pest control. In a word, the shift was from low input-low output systems of production to high input-high output systems. The plant s architecture was changed from tall to short with the straw having to be stiffened through breeding to carry the additional weight of the larger panicles, a process that forced a change in harvesting methods, especially in Southeast Asia. Varieties were also bred for photoperiod insensitivity and especially, for response to nitrogen inputs, now substantially from artificial fertilizer. At the same time the early HYVs proved susceptible to pests and diseases. For example, the brown plant-hopper, once a relatively minor pest, is now a major one. Much the same is true of tungro virus. The immediate answers have often lain in the application of toxic chemicals, especially organophosphates, one consequence of which has been a reduction in the production of fish, mollusks and crustaceans, once an important component of rice-field ecosystems. While very substantial yield increases have certainly resulted, these have not been without environmental costs, especially the eutrophication of fresh waters. However, later HYVs have been bred for resistance to many of diseases and pests affecting the earlier ones. But generally, environmental requirements are more stringent than for the older traditional varieties.

PRODUCTION SYSTEMS AND ENVIRONMENTAL IMPACTS

The enormous range of production systems, and thus their associated environmental impacts, makes adequate generalization concerning them is difficult. For instance, a study

of system characteristics in Malaysia in the 1960s, before production became much more uniform as a result of the introduction of HYVs, recognized 36 distinct systems, 31 on naturally flat plains or terraces, one on artificial terraces and four on sloping land. Some have disappeared from Asia generally, including rice-based shifting cultivation in mangroves (Myanmar), shifting cultivation involving virgin forest (Southeast Asia generally), permanent brackish-water cultivation in Hong Kong and wet land sowing without clearing or tillage (Peninsular Malaysia).

While production systems are very varied, most fit into the global classification used at IRRI, described by Greenland (1997) and illustrated in Figure 2. This sets out schematically the kinds of environments and topographical positions in which the crop is grown together with approximate areas under each form of cultivation for the early 1990s. Lowland rain-fed and flood systems occupy about a third of the cultivated area but produce only about a fifth of the grain. This is because the water supply is highly variable. The amounts of sediment deposited by flooding are inconsistent and in the deeply flooded areas, plant toxins form through anaerobic conditions. In addition, largely because of farmers uncertainties about the water regime, limited amounts of fertilizer are used. Genetic enhancement of the rice varieties used in rain-fed and flooded areas is also in its infancy, contributing to low yields. Yields range from as low as $1.0\,t\,ha^{-1}$ to as much as $4.0\,t\,ha^{-1}$ in good seasons and favorable locations.

Rice cultivation on dry, sloping land occupies only about a tenth of the total area, yielding a miniscule 4% of global production. Rice grown in such environments, whether by shifting cultivation or semi-permanently, is even more susceptible to dryness. To this may be added weeds, pest problems, especially birds, and a lack of available soil nutrients, especially nitrogen (N) and phosphorus (P).

Irrigation, permitting control of the timing and amount of water, removes or at least mitigates the effects of plant stress through dryness; its provision at the crucial phases of flowering and panicle formation is particularly important. Just over half of the global rice area is said to be irrigated. (Some sources give a lower proportion of about two-fifths.) Of that figure, roughly half again bears two rice crops a year. Consequently almost three-quarters of total production comes from irrigated land. With irrigation, advantage can be taken of higher insolation intensities in the dry season so that grain yields at that time are generally substantially higher than at other times, in California and Australia regularly reaching $10\,t\,ha^{-1}$. Pest and disease pressures are also lower then. Genetic engineering has also produced HYVs well adapted to these controlled conditions, whereas for all other systems, the need for plants tolerance of a variety of unpredictable water conditions has severely constrained rice breeding.

In terms of environmental impacts, a distinction can readily be drawn between those that are associated with the initial transformation of natural ecosystems by clearance and fire and impacts that are the continuing consequences of the existence and intensification of rice-field ecosystems. Not much is known, in biological terms, of what happens when, as in usual, rice replaces forest, especially so far as microorganisms are concerned. Most obvious is a large fall in the standing crop. Where this has been mature forest the phytomass values fall from somewhere in the range of $80\,t\,ha^{-1}$; for light monsoonal forest to about $550\,t\,ha^{-1}$; for well-grown equatorial rain forest to roughly $5\ 15\,t\,ha^{-1}$ of rice plants and associated weeds.

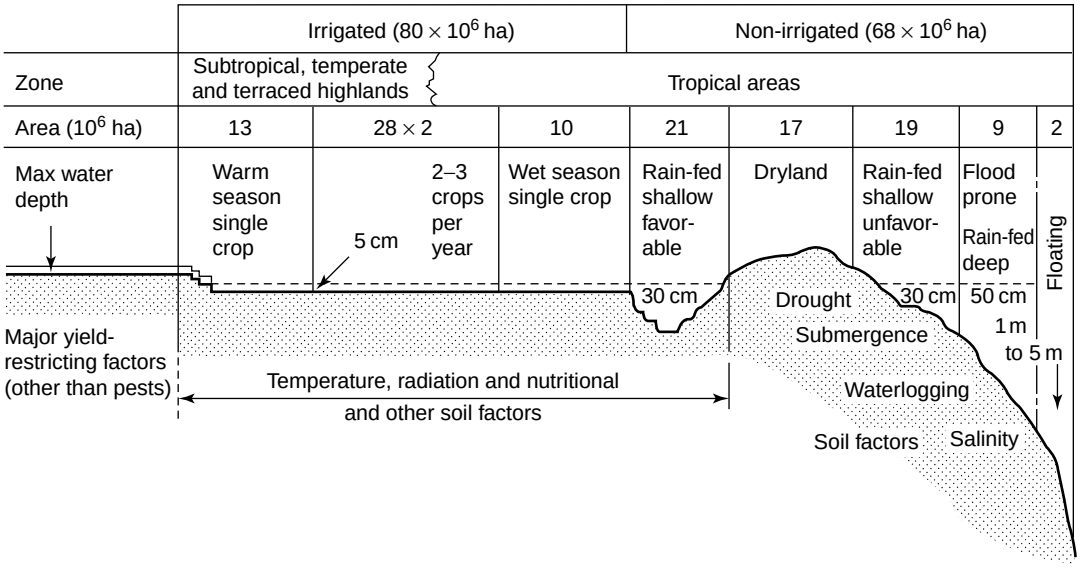


Figure 2 Major rice production systems and areas in 1991 (after Greenland, 1997)

Annual biological production falls less drastically, from 15 30 t ha^{-1} for forest to 5 25 t ha^{-1} for a once-a-year rice crop and its weeds. The amount exported from such systems varies. Where only the grain leaves the eld at each harvest, the range is from 1.0 16.0 t ha^{-1}, though towards the lower end of that span under pioneer conditions as rice ecosystems are being established. The crop residue may remain in place, rot and be recycled, but since this practice tends to lead to a build-up of disease-causing organisms, it is often burnt, releasing CO_2, nitrogen oxides (NO_x), NH_3 and other gases to the atmosphere together with P, K (potassium), (calcium), and Mg (magnesium) to the soil. Where the straw is removed from the eld, similar processes take place in and around farmsteads, rice straw being transformed into combustion products in domestic stoves, into compost as bedding materials for animals or into dung as it is consumed and then recycled onto elds. Increasingly some is used in biogas digesters.

The once-for-all clearance of (usually) forested land for rice unquestionably has added substantial CO_2 and NO_x to the global atmosphere, though since this happened at varying rates in time past it is dif cult to say by how much. Since these gases, bear isotope signatures that permit the identi cation of source, it is possible that fossil air trapped in glaciers and ice caps contains evidence of burning following clearance. This is complicated, however, by the lack of adequate knowledge of the degree to which CO_2 uptake increases in the remaining vegetation. Burning of cleared vegetation certainly releases substantial amounts of nitrogen, estimates for forest ranging from about 100 300 kg ha^{-1}, which may be recycled by washout in rain. Of the other major elements added to the soil, the rice crop takes up a small but vital proportion and larger fractions are lost by eluviation, by volatilization (limited in wet systems) and by sequestration through soil processes. The amounts freed vary in step with the standing crop, up to about 3 t ha^{-1} for mature forest, about half that for well-grown secondary forest and only a few kilograms where the original vegetation is grass.

A particular feature of clearance is a rapid but short-lived ef orescence of soil microorganisms, some of which may be nitrogen- xing. These include actinomycetes, various species of symbiotic and non-symbiotic bacteria, including *Azotobacter* and *Clostridium* spp. as well as the Cyanobacteria, the last being of considerable signi cance where soils are ooded.

Once clearance takes place and a crop or two are taken, the most important consideration is what happens next. On many alluvial soils in rice-growing regions, cultivation is continued, resulting in very different ecosystems from those under natural vegetation. On a few and on slopes, a fallow follows during which natural vegetation regrows and nutrients are restored. This process is more or less rapid depending especially upon the size of the cultivated plot, for regeneration is most rapid at the edges and where there is a plentiful supply of seeds. Regeneration is also in uenced by the period over which rice cropping has taken place since the seeds of some wild plants may survive cropping. What is certain is that *Oryza*, partly because it is a light-loving species, does not long survive the abandonment of cultivation.

Plant succession varies considerably, both from region to region, being in uenced by the regional vegetation and thus the nature of the seed supply and locally depending especially upon the frequency of successive cycles of cultivation. Where the latter is low, woody plants quickly dominate and where a further clearance is long delayed, shade-loving forest species are able to establish themselves. The classical studies of Wyatt-Smith in Peninsular Malaysia in the 1940s showed that within a few years, quick growing trees such as *Mallotus*, *Trema* and *Macaranga* spp. were some 7 10 m high. Beneath them seedlings of true rainforest trees were present, including *Shorea*, *Milletia*, *Durio*, *Koompassia* spp. By contrast, where the cultivation cycle is frequent and fallows short, many invaders are grasses such as *Imperata* and *Paspalum* spp. These are characteristically weeds of cultivation and usually expand their area on abandonment. Should such grass-dominant fallows be burnt, the grasses may form a more or less stable re climax vegetation, returning little to the soil. Where re is kept out, woody species, in Peninsular Malaysia *Macaranga* and *Mallotus* spp. in particular, eventually suppress the grasses, acting as a nursery for more truly forest species.

ECOLOGY

Where permanently cultivated rice- elds are established, the ecosystem is very different. A regular alternation of standing crop, with or without short fallow is established. Over 20 or so years the debris of original clearance nally rots away, a process within ve or so years suf ciently advanced to permit tillage. Where an annual crop is usual, the phytomass increases in a characteristic sigmoid curve from virtually nothing at planting to perhaps 5 25 t ha^{-1} at harvest. It falls away as the crop is cut, rising somewhat as the rice stubble regrows and as weeds invade. In wet systems, as the rice grows, the weeds are suppressed, non-aquatic weeds by ooding and tillage early in the cultivation cycle and aquatic weeds by drying out late in the cycle.

But wet elds are not the only kind to be more or less permanently cultivated. Dryland rice, unlike that grown by shifting cultivation (bush-fallowing), always involves tillage, the crop being grown in unbunded elds depending entirely upon rainfall for moisture. Such cultivation is of particular importance in Latin America and in Africa where, with rice grown by shifting cultivation, dry, upland production accounts for between three-quarters and half of the total rice area. Quite often rice is relay cropped

or intercropped with maize, sorghum, soyabean, peanuts, pulses, manioc, sugarcane, even permanent tree-crops such as coconuts and citrus. (Such systems reduce the incidence of pests and diseases, soil erosion, prevent and provide varied dietary and income sources, reducing vulnerability to drought.) Because of erosion risk and substantially lower grain yields than from wet systems of rice production, permanent dry cultivation is limited to relatively gentle slopes; and where practised without fallows, or added fertilizer. It confined mostly to youthful, nutrient-rich soils especially basic volcanics such as those of Java, Bali and the Philippines.

Not much is known in detail of the ecology of such systems. Certainly the weed-load is much greater than for wet systems, especially as many weeds grow more rapidly than rice in the first two months of the growth cycle. A West African study, for example, showed that weeds reduced grain yield by 33 75% on unweeded wet plots but by 70 100% on unweeded upland plots. The weed communities vary from major region to region, many genera *Echinochloa*, *Ipomoea*, *Cyperus* also being common in lowland wet systems. One major feature however, is the much greater soil loss by erosion on dry, tilled slopes, from which sediment production may be three or four orders of magnitude higher than from lowland or terraced wet systems. Sediment yield is also two or three orders of magnitude higher than from land under the shifting cultivation of rice, where the regeneration of vegetation rapidly reduces erosion unless ineffective grass fallows, such as under *Imperata* spp., follow cultivation.

Little can be said of rice-field communities in general because of a lack of comprehensive ecological studies, though much is known of major plant and animal pests and diseases. Wet rice fields, including those created as artificial swamps on hill-slopes, characteristically contain variable quantities of weeds such as grasses and sedges like *Digitaria*, *Eleusine*, *Cyperus* and *Echinochloa* spp., also wild rice, especially *O. perennis*, and herbs such as *Ipomoea triloba*, in some areas *Chenopodium* and *Amaranthus* spp. themselves furnishing food for man or beast. Floating plants include *Salvinia* and *Nymphaea* spp., and in Asia the American introduction *Eichhornia crassipes*, more important in irrigation systems than in actual fields, as well as free-floating *Azolla* spp., an aquatic fern that in symbiosis with Anabaena fixes valuable nitrogen from the air. In some regions trees may be part of the rice-field ecosystem. From India to Cambodia *Borassus* palms, once a major source of sugar, may grow in the fields, sometimes *Arenga* palms as well. In the fields of Northeastern Thailand, a wide range of trees, both forest survivors and planted ones, is quite common, providing food, medicine, fodder and lumber, all carefully trimmed to avoid excessive shading of the rice crop and consequent depression of yields.

In Northeast Thailand, Heckman noted the presence of 589 animal species, of which 130 can be considered typical rainy season rice-field biota. Another 250 species are common in other habitats but sometimes enter fields. Amongst these are economically important fish, some partly air breathing and capable of surviving the dry season. Common in many lowland parts of Asia are carp (*Cyprinidae*), gouramies (*Trichogaster* spp.), serpent-heads (*Channa* spp.) and catfish (mainly *Clarias* spp.). Yields of such fish vary widely, being very low where toxic pesticides are used. Annual production of 2200 kg $ha^{\bullet 1}$ has been reported from Japan but this seems very exceptional, yields in Taiwan, Malaysia and Tanzania ranging from about 100 kg $ha^{\bullet 1}$ to 300 kg $ha^{\bullet 1}$. Crayfish are locally important, yields of 560 kg $ha^{\bullet 1}$ being reported for Louisiana paddy-fields, while along the Travancore Cochin coast, Southwestern India, prawns are raised during the fallow season, yields ranging from 780 to 2100 kg $ha^{\bullet 1}$ per season.

The most striking remaking of ecosystems for rice cultivation has come about in mountainous areas, notably of South and Southwestern China, especially Sichuan province, in Java and Bali and in the Ifugao region of North-central Luzon, Philippines. In all these regions wet rice terraces are of two kinds. On the one hand, they may be naturally watered by entrapped rainfall, overland flow and to some degree near the bottoms of slopes, by emergent groundwater. On the other hand they receive water from elaborately constructed irrigation systems tapping into the upper courses of small streams and diverting water in a series of carefully controlled cascades falling from one level to the next. Traditionally, both kinds of terrace were confined to lower and middle slopes while the upper slopes acted as gathering grounds for water. Since about 1950, and mainly in South and Southwestern China, the use of mechanical pumps has enabled irrigated terraces to be continued to the tops of slopes, water being recycled several times as a consequence.

The dating of wet terraces is mostly uncertain. Some are thought to be two millennia old. Others are much younger, the early period of Communist rule in China seeing a notable expansion as labor was substituted for capital and local self-sufficiency was encouraged. Construction and maintenance are highly labor-intensive, the former requiring 3000 to 5000 person-days per hectare. Given the recent enhancement of the opportunity cost of labor in wet-terrace areas, there is now virtually no new construction and in some places rice is no longer grown as disintensification takes hold. This may involve the replacement of rice with lychee (*Nephelium litchi*) in upland Guangdong, the growing of sweet potato (*Ipomoea batatas*) rather than rice, or complete abandonment. The last occurred in the 1950s and 1960s in Hong Kong and is now happening in Ifugao, Luzon, where there is concern over the possible loss of a major tourist attraction.

Little is known of the ecology of such artificial vertical swamps, though they are known to contain mollusks and crustaceans, both of some significance as protein sources. Unlike their lowland counterparts, large fish are absent. In addition, there is always the risk of terraces collapsing, despite the usual practice in most areas of facing them with supporting stonework. As in the lowlands, distinctive anthropogenic soils evolve. These are marked (when moist) by a structureless A horizon, often gray in color with reddish mottles, with a marked clay-iron horizon at a depth of about 30 cm. This develops in the course of cultivation and serves to slow infiltration and consequent water loss. Like their lowland counterparts, such rice soils are also high in organic matter, the breakdown of which by various species of Cyanobacteria leads to significant production of CH_4 gas.

METHANOGENESIS

Methanogenesis refers to the making of CH_4, in the case of wet rice fields by a group of Archaebacteria that thrive in the anaerobic environment of the wet fields, as they do in naturally-occurring freshwater swamps. Their significance lies in the fact that CH_4, like CO_2, is a greenhouse gas. CH_4 from various sources may contribute about 15% of the greenhouse effect, compared with roughly 60% from CO_2. When its effects are corrected for decay time (10 years) and for indirect effects, it is said to be 15 times stronger than CO_2 whose decay time is 120 years. Using fossil air trapped in glaciers it has been estimated that global CH_4 levels rose from 0.70 ppm (on a volume basis) in 1787 to 1.68 ppm by 1987, an average rise of roughly 1% a year. Since the mid-1980s, though, the rates have fallen to about 0.5% annual increase (*see* Methane, Volume 1).

Estimates of global CH_4 production and of the proportion contributed by CH_4-producing organisms living in wet rice fields are very varied, reflecting differing methods of analysis, varying assumptions in extrapolating from experimental to global levels and the fact that methanogenesis is highly variable in time and space. A current best estimate is that wet rice fields annually produce $120 \pm 50\ \mathrm{Tg\ yr^{-1}}$ (trillion tons), making up about a fifth of the global production. Stern and Kaufmann indicate a rise of annual production in wet rice fields from 40.1 Tg in 1860 to 101.4 Tg in 1996.

Details of just how certain groups of Archaebacteria produce CH_4 are not entirely clear. In wet rice fields, as elsewhere in nature, a basic prerequisite is the presence in sufficient quantity of appropriate substrates, notably H_2, CO_2 and acetate. The origin of these is not clear – possibly humus, plant residues, autolysates or root exudates. What is clear is that the presence of organic matter increases CH_4 production and that anaerobic conditions are obligatory, for the producers die in the presence of oxygen. The factors responsible for CH_4 production interact in complex and not yet fully understood ways.

Methanogenesis can occur at all temperatures at which rice grows – as well, in nature, at both lower and much higher temperatures. The saturation of the soil with water, giving reducing rather than oxidizing conditions, is essential though a small quantity of CH_4 may be produced in water films around soil particles in some unsaturated soils. Other things being equal, the water regime, both during the period of cropping and during fallow, has major effects. Methanogenesis depends upon a fairly narrow range of negative redox potentials (*see* Redox Potential, Volume 2) and that in turn depends on soil water content. Thus a soil that is wet during fallow is already in a reduced condition whereas a soil initially dry then flooded will not produce CH_4 until several weeks have elapsed. There is some evidence that in such circumstances, even peak production will be lower than in ever-wet conditions. The reasons for this are not altogether clear for not all the CH_4 produced in a field is emitted to the atmosphere. Most is emitted via the rice plants themselves, and probably also by such other emergent aquatic plants as possess gas pathways within their tissues. The remainder that reaches the atmosphere does so by ebullition, bubbling up through the muddy water. But some is consumed, mainly in the root zone, by methanotrophs, organisms using CH_4 to support their metabolism, living in the narrow oxidized zone around every living root. Thus an increase in the number of methanotrophs present reduces emissions so that oxidizing conditions have effects both on CH_4-producing organisms and on CH_4 consumers.

Various management practices clearly affect CH_4 emissions. The growing of forerunner and post rice-harvest crops leads to reduced conditions for shorter periods. Broadcasting seed rice, though leading to lower yields than when seedlings are raised in a nursery and transplanted, is said to reduce the period of inundation required. The application of nitrogen as urea increases CH_4 from some soils but limits its production from others. The use of HCH pesticide (hexachlorocyclohexane) retards CH_4 oxidation, leading to more CH_4 whereas another commonly used rice pesticide, carbofuran, has the opposite effect.

But perhaps the most important consideration, next to water regulation, is organic matter content. Some rice soils, though not good ones, have been developed from naturally organic peats and mucks. Many are developed from alluvium or colluvium. Some, on wet terraces, are developed from sedentary, hill slope soils. Aquatic plants such as *Azolla*, weeds, stubble and even whole rice plants minus their panicles which alone may be harvested, lead to large inputs of organic matter, as much as $8 \pm 10\ \mathrm{t\ ha^{-1}}$ per crop. The bacteria thus have plenty of substrate to live on. In single rice-crop systems, however, post-harvest debris can readily be burnt and though this results in a loss of

nitrogen, the practice leads to less CH_4 output, as does dry fallowing. Ironically, organic farming almost certainly enhances CH_4 output. China s rice production levels are high and are heavily dependent not only upon recycled rice-field phytomass but also on large nutrient subsidies from organics, including pig and human manure as well as compost. It seems likely that China s wet rice fields produce above-average quantities of CH_4 as a consequence.

The prognosis for wet rice-field methanogenesis is circumscribed by data limitations and is thus little better than guesswork. In some places, for example Peninsular Malaysia and Java, wet rice cultivation is being abandoned and not replaced. Where sites remain wet, it seems likely that CH_4 production will continue though at lower levels since vegetation is no longer cut and incorporated into the soil. Where sites dry out, a substantial reduction in methanogenesis is likely. On the other hand, it is clear than in Asia, though not in Africa, clearance of land for new wet rice-fields has substantially ceased. Future increases in rice production, triggered by continuing population growth and possibly also by price rises, may well involve a substantial rise in the area double-cropped with rice. Thus longer periods of flooded, reducing soil conditions and more CH_4 may be expected. This could be partly offset by tighter water control as costs of supply rise and by increased application of artificial nitrogenous fertilizers that may inhibit methanogenesis. Because the decay-period for CH_4 is only about five years, compared with 120 years for CO_2, and because of the much greater global warming potential of CH_4, any reduction in CH_4 production will probably affect the atmosphere quite quickly, diminishing the greenhouse effect.

GLOBAL WARMING AND RICE

Global warming will affect rice production in four ways:

1. Increasing concentrations of carbon dioxide will stimulate plant growth. The effect of this on plant production will depend on whether a species is a photosynthetic C_3 or C_4 type, as well as on the presence of competitors (including weeds) and other factors. This general topic is treated in several articles in Volume 2, although no specific reference is made to rice (*see* C_3 and C_4 Photosynthesis, Volume 2; CO_2 Enrichment: Effects on Ecosystems, Volume 2; Plant Competition in an Elevated CO_2 World, Volume 2; Plant Growth at Elevated CO_2, Volume 2). While early studies, in Japan, indicated increased photosynthesis in rice at levels up to about 500 ppm, above which rates leveled off, later work indicated that at high temperatures, even under low light intensities, increased growth and grain yield occurred at CO_2 levels far beyond what may reasonably be expected globally. Yoshida (1976) suggested that the optimum concentration for growth and yield is in the range 1500 2000 ppm, increases being attributed almost entirely to increased number of panicles. Since this is four to six times current levels it may be concluded that enhanced CO_2 levels would be wholly beneficial. This would certainly be the case for rice yields. But what are not yet clear are the relationships between CO_2 uptake in rice and CO_2 uptake in forests that rice cultivation is still replacing in some parts of the world. The likelihood is that uptake levels for rice are lower than for forest so that land conversion results in enhanced CO_2 levels generally, though matters are much complicated by limited knowledge about uptake by forest species, especially in the tropics, and their response to higher levels of CO_2 in the atmosphere.
2. Warmer temperatures, particularly at night, thus reducing the diurnal temperature cycle. The temperature increases will be highest in polar regions.
3. Changes in the atmospheric general circulation, leading to changes in the strength and timing of ENSO, particularly in the case of Southeast Asia. In its latest scientific assessment (IPCC, 2001), the IPCC is confident that the evidence provided from recent trends and simulation models support the view that the world is warming, and that the warming over the next 100 years will be greater than previously thought (*see* Intergovernmental Panel on Climate Change (IPCC): an Historical Review, Volume 4). (See also several review articles in Volume 1).
4. Rising sea-level: Unfortunately, data relating rice areas and height above sea level seem not exist. Clearly coastal areas would be vulnerable not merely to higher sea levels themselves but also to the predicted greater frequency of storm surges, bringing salt or brackish water onto coastal fields. Mitigation would include the construction of sea walls, an impossibly expensive option for poor rice-growing countries such as Bangladesh and Viet Nam, and the greater use of salt-tolerant varieties. Unfortunately, these produce less grain than normal varieties. Given that pond production of marine animals is already more remunerative than growing rice, though probably environmentally damaging and unsustainable unless very carefully managed, conversion of use away from rice is probably inevitable in such threatened coastal areas.

Overall, warming could lead to expansion of cultivation into areas currently too cool to permit rice to be grown. In some areas where winter coolness prevents growing a second crop parts of Northern India, central China and tropical uplands generally warming would permit double cropping where only one crop is now possible. Such expansion would be conditional upon water supplies being sufficient to meet increased demands and upon basic economic perceptions amongst farmers. In rice-production

areas where temperatures are already high, and these are around 25 35° latitude, not near the equator, even higher temperatures may depress yields by increasing the incidence of spikelet sterility in the plant and/or by increasing water demand to levels that cannot be met.

SUSTAINABILITY

Given that humankind has been growing rice, sustaining life and high civilizations for several thousand years, it may seem odd to question the long-term sustainability of its cultivation. The fact is that for most of that time yields have been variable from season to season around a mean per crop of roughly 1.0 1.5 t ha^{-1}. By heavy use of organic fertilizer and by close control of the amount and timing of water application, traditional production systems doubled that yield. In rice-based systems of shifting cultivation, where bush fallows were 10 40 years long, yields were probably close to those in most rain fed lowland systems and higher than in some, especially flood-prone ones. The substantial shift to high-input high-output production that began in Western temperate areas in the 1930s and spread rapidly, in Asia especially, in the 1960s and 70s represented a one-way quantum jump to high yields. It would be impossible economically to produce the present amount of rice at the yield-levels prevailing through most of its history. There simply is not enough land. At the same time, it is clear that there are both technological and economic limits to intensification. One straw in the wind here is the abandonment of an experiment at IRRI, Los Banos, which succeeded in annually raising 20 t ha^{-1} of grain by multiple-cropping, optimal water management, varietal choice, fertilizer application and pest control. It was neither technically sustainable nor an appropriate model for ordinary farmers.

Giving details of nutrient balances in pre- and post-1960s cropping system, Greenland (1997) suggests a key role for the nutrients contained in sediments in irrigation water, especially P. These are reduced by sedimentation in the storage-based irrigation systems used in dry areas and dry seasons. However, any macro- or micro-nutrient deficiencies can readily be made up by the continued application of artificial fertilizers, the raw materials for which are likely to be available for some centuries yet. Nitrogen, of course, can readily be fixed from the air using appropriate biological nitrogen-fixers. Their use is not necessarily straightforward. *Azolla-Anabaena*, for instance, requires considerable management skill. *Azolla* readily washes off fields and dies when it is cold or dry. Though at present biological nitrogen-fixation may not be financially competitive with petroleum-derived nitrogen such as urea, the position will improve as the real cost of petroleum products increases. Recycling human urine is technically feasible but suffers from the high cost of concentrating very dispersed sources. Recycling sewage has some of the same problems plus the likelihood of increasing CH_4 as an output from wet fields, plus religious unacceptability in some areas, notably India. Increased nutrient applications will unquestionably lead to increased eutrophication of fresh waters, especially lakes, and near-shore waters. The frequency of damaging algal blooms may increase, while for fresh waters, possible toxicity problems such as excessive nitrate may arise.

Loss of water quality for rice production has already been mentioned but of greater concern is the related question of increasing competition for and cost of water as well as the problems of optimizing its management. The additional amount of water needed by a rice crop depends upon the amount, intensity and timing of rainfall, evaporation, crop duration, soil permeability, and physiological need. Wet field systems need additional water for land soaking to soften the soil aggregates hardened by exposure to sun and thus to allow tillage. According to Greenland (1997), 300 500 mm are needed for land preparation and 540 1620 mm for actual crop growth in systems requiring 100 days from transplanting to harvest with final drainage being initiated 10 days before harvest. Little can be done about rainfall or evaporation. On-field water storage, by raising the height of field bunds is rarely feasible as there is then a risk of drowning the crop, for most varieties have limited tolerance of submergence. Shortening the cultivation period reduces water demand and this can be done by selecting varieties with shorter maturation periods and by direct seeding instead of transplanting. But both alternatives may reduce yields, the former because many short-term varieties produce less grain than long-term varieties, other things being equal, the latter because of increased competition from weeds. Machine tillage also reduces the land-soaking requirement since tractors can handle much more cohesive soil conditions than buffaloes or other traction animals. But costs may increase.

Excess water is also a problem. In dry areas, lack of proper soil drainage readily leads to salinization fatal to rice. In deltaic areas, both depth and duration of flooding are crucial, in relation to crop height, for rice is vulnerable to drowning as just explained. Varieties that elongate when deeply flooded are uniformly low yielding at present so that the breeding of elongating HYVs is a strategic priority for such lands.

Pests, including weeds and disease, continue to be problems, some caused by the very success of enhancing yields. (The cases of tungro and brown planthopper were mentioned earlier). One major concern is that varietal diversity, itself a protection, is being reduced by the use of single varieties over large areas, for each variety has a different pattern of vulnerability to pests. Increase in rice-crop frequency may enable significant pest populations to carry over from one crop to the next. There are well-founded reports of farmers reverting to single annual crops for this

reason. Increasing cropping intensity by using some other crop, preferably a legume to increase nitrogen, is helpful, since many pests are crop-specific. But both insecticides and herbicides may cause major environmental problems, effects upon rice-field animals used as a food source being one example. The use of herbicide has been proposed in mechanized cultivation of rice on sloping land where weed control is a serious problem. But environmental consequences may be rather severe. These may include loss of biodiversity though few studies of this have been undertaken in tropical areas. In these areas generally, it is commonly believed that, other things being equal, pesticide and herbicide use is less damaging than in temperate areas because higher temperatures lead to more rapid chemical breakdown.

A further issue is that of land availability. While it is true that there are significant areas of land that could technically be brought into rice production, even in Asia, it is highly doubtful if such a transformation is economically feasible even if potential farmers value their own labor at close to zero. In the remaining lowland areas, there is competition with forestry, nature conservation, fisheries and a host of other potential uses. In addition, many areas contain soils that are difficult to drain and manage. Such include soils with high levels of organic matter and mangrove soils, in many of which, the cat clays (free sulfuric acid forms on drainage) rendering prolonged flushing with fresh water necessary before planting rice (*see* **Acid Sulfate Soils**, Volume 3). Much prime rice land has been lost forever to urban and industrial expansion, notably in China, the southern province of Guangdong especially, and on the Bangkok plain. Other lands are lost by disintensi cation and abandonment as labor costs rise in relation to prices. Ironically, the latter have fallen in real terms since the 1950s, except for a transitory upward blip in the mid-70s, even as global production has more than doubled. When prices begin to rise once again, as surely they eventually must in the face of continued, though slowing population growth, it will probably again become economic to expand the area under rice. Where production is mainly for self-subsistence, as in many West African rice areas, expansion may still take place even if this is uneconomic at the macro-level. For the rest, intensi cation of production is likely to meet future demand for rice as a result of changes in diet, even though the increase in demand is likely to remain somewhat below the rate of increase of rice-eating populations. What, in detail, the environmental consequences of such intensi cation may be are not entirely clear.

See also: **Demographic Change: Peopling of the Paci c Islands**, Volume 3; **Fisheries: Pollution and Habitat Degradation in Tropical Asian Rivers**, Volume 3; **International Rice Research Institute (IRRI)**, Volume 3; **Irrigation: Induced Demise of Wetlands**, Volume 3.

REFERENCES

Greenland, D J (1997) *The Sustainability of Rice Farming*, CAB International, Wallingford and IRRI, Manila.

Oka, H I and Chang, W S (1959) The Impact of Cultivation on Populations of Wild Rice, *Oryza sativa* var. *spontanea*, *Phyton*, **13**, 105 117.

Yoshida, S (1976) Carbon Dioxide and Yield of Rice, *Proceedings of the Symposium on Climate and Rice*, IRRI, Manila, 211 221.

FURTHER READING

Bray, F (1986) *The Rice Economies, Technology and Development in Asian Societies*, University of California Press, Berkeley, CA.

Evenson, R E, Herdt, R W, and Hossain, M, eds (1996) *Rice Research in Asia, Progress and Priorities*, CAB International, Wallingford and IRRI, Manila.

Grist, D H (1975) *Rice*, Longman, London.

Gupta, P C and O Toole, J C (1986) *Upland Rice, a Global Perspective*, IRRI, Manila.

Hazell, P B R and Ramsamy, C, eds (1991) *The Green Revolution Reconsidered*, Johns Hopkins University Press, Baltimore, MD.

International Rice Research Institute (1976) *Proceedings of the Symposium on Climate and Rice*, IRRI, Manila.

International Rice Research Institute (1987) *Weather and Rice*, IRRI, Manila.

Luh Bor, S, ed (1991) *Rice*, Van Nostrand Reinhold, New York.

Matthews, R B, Kropff, M J, Bachelet, O, and van Laar H H, eds (1995) *Modeling the Impact of Climate Change of Rice Production in Asia*, CAB International, Wallingford and IRRI, Manila.

Moormann, F R and van Breemen, N (1978) *Rice: Soil Water, Land*, IRRI, Manila.

Oka, H I (1988) *Origin of Cultivated Rice*, Japan Scienti c Societies Press, Tokyo.

Pearse, A (1980) *Seeds of Plenty, Seeds of Want: Social and Economic Implications of the Green Revolution*, Clarendon Press, Oxford.

Pingali, P L, Hossain, M, and Gerpacio, R V (1997) *Asian Rice Bowls: The Returning Crisis?* CAB International, Wallingford and IRRI, Manila.

Environmental Change and Human Health: Extending the Sustainability Agenda

A J McMICHAEL

London School of Hygiene and Tropical Medicine, London, UK

Human communities have depleted natural resources and degraded local ecosystems over many years. The usual local consequences have been impairment of health, vitality and social viability. Today, this process is beginning to be played out on a much larger scale, and some of the longer-term consequences for the health of human populations could be commensurately more serious. Today's unprecedentedly large human impact on the environment re ects the great, ongoing, increases in population size and in energy-intensive, high-throughput consumerism. Human-induced changes are becoming evident in the composition of the lower and middle atmospheres, in the worldwide depletion of other natural systems (e.g., soil fertility, aquifers, ocean sheries, and biodiversity in general), and in the disturbances to great natural cycles (especially water, nitrogen and sulfur). These large-scale physical, biological and ecological systems are the source of life-support *upon which the sustained health of populations (human and other) depends.*

The risks to health posed by these global environmental changes are, in many ways, rather different from the well known, locally acting, environmental hazards from direct-acting toxic pollutants. Global change processes are likely to alter the geography of infectious diseases (especially vector-borne diseases such as malaria and dengue, sensitive to changes in climatic conditions and land-use patterns), the productivity of food-producing systems on land and at sea, the availability of freshwater, and, through biodiversity loss, to destabilize and weaken the ecosystems that underpin these and other life-support processes. The depletion of stratospheric ozone poses several direct risks, including increases in skin cancer incidence and the occurrence of eye disorders, and it may also affect immune system activity. Climate change related changes in weather extremes, natural disasters and sea-level rise would have various direct and indirect health consequences. The recently identi ed global dissemination of various Persistent Organic Pollutants (POPs) poses another set of environmental health problems, which may contribute to biodiversity losses in general and to changes in human biological functioning (fertility, reproduction, and immune function).

The spectrum and extent of health impacts from global environmental change will vary around the world. Poor and isolated populations will, usually, be the most vulnerable. Environmental changes and stresses may often coincide, compounding the impact on local human populations. For example, the combined impacts of climate change, freshwater shortages, and land degradation may impair agricultural productivity most in subtropical and semi-arid regions where food insecurity is already prevalent.

The scienti c assessment of actual or potential health impacts faces some unusual dif culties in this complex topic area. There are various types and levels of uncertainties. Many of the systems being studied are complex, non-linear and dynamic. An interdisciplinary research effort is therefore needed that is able to accommodate an unusual mix of complexity, uncertainty and predictive modeling.

There is one particularly important dimension to this topic. These global change processes have major implications for the long-term sustainability of human population health: they may weaken, perhaps irreversibly, the natural life-supporting infrastructure. The widespread, dramatic, gains in health and longevity over the past century have been closely related to the processes of urbanization, social modernization, industrialization and increasing material wealth, as well as to technological advances in preventive medicine, medical care, sanitation and hygiene. Those gains have therefore been associated, to some (uncertain) extent, with the depletion and degradation of our external environment, including most recently larger-scale environmental changes. However, there are likely to be thresholds above which the further use of environmental resources causes no improvement

in health (as is broadly evident from the plateauing of national average life expectancy in relation to per capita gross national product (GNP) at higher income levels). The question arises: given, now, the advent of global environmental change, can we maintain parallel increasing trends in consumption, environmental damage and life expectancy?

INTRODUCTION

Humans, like various other large mammals, are patch disturbers (Rees, 2000). Hunter-gatherers by origin, they have long subsisted in nature by intensive exploitation of a local environment, before moving on to an adjoining virgin patch. Slash-and-burn agriculture is a more structured variant of this behavior. Indeed, among agrarians the principle of rotating crops and leaving land fallow for a season is an extension of this strategy. With variable intensity, and one that usually increases over time as culture and technology evolve, human communities tend to deplete natural resources and degrade local ecosystems. The usual, local, consequences have been malnutrition, impairment of health, conflict, and decreased social viability.

Today, this process is beginning to be played out on a much larger scale, and some of the longer-term consequences for the health of human populations could be commensurately more serious. The scale of human impact on the environment has increased rapidly over the past two centuries, as human numbers have expanded and as the intensity of economic activity has increased. Global economic activity increased twenty-fold during the twentieth century. Meanwhile, in absolute terms, the human population has been growing faster than ever in the last quarter-century, capping a remarkable fourfold increase from 1.6 to 6 billion in the 20th century. While we remain uncertain of Earth s human carrying capacity , we now expect that world population will approach nine billion by around 2050 (*see* Ecological Capacity, Volume 3; Global Population Trends, Volume 3).

Over the past century or so, the typical, localized, symptoms of human environmental impact have been industrial air pollution, urban squalor and chemical pollution of waterways. These local health hazards are now being supplemented with the health risks posed by changes to some of the planet s great biophysical systems and biogeochemical cycles. This includes the climate system, the stratosphere, the hydrological cycle, elemental cycles (especially carbon, nitrogen and sulfur), soil fertility and biodiversity. We are beginning, unintentionally and at a global level, to alter the conditions of life on Earth: even though we remain largely uncertain, even ignorant, of the long-term consequences.

The best known of these global environmental changes is the impact that humankind has begun to have on the composition of the lower and middle atmospheres; the former via the accumulation of heat-trapping greenhouse gases in the troposphere (the lowest 10 km of atmosphere), the latter via the accumulation of ozone-destroying gases in the stratosphere (10 50 km altitude). Other signs of planetary overload include an apparent recent plateauing in the productivity of our main terrestrial and marine food producing ecosystems, a widespread loss of biodiversity, and the global dispersion of various persistent non-biodegradable pollutants (McMichael, 1993; UNEP, 1999). The central issue here is that the underpinnings of human health are being perturbed or depleted. The sustained good health of any population, over time, requires a stable and productive natural environment that yields assured supplies of food and fresh water, has a relatively constant climate in which climate-sensitive physical and biological systems can adapt to short-term and long-term variability, and retains its richness of biodiversity (a source of both present and future value). For the human species, a social animal , the texture and stability of the social environment is also important to population health.

Some of these environmental stresses are likely to cause tensions between human communities, leading to conflict; one of the four biblical Horsemen of the Apocalypse known throughout history as a perennial scourge of population health. For example, Ethiopia and the Sudan, upstream of Nile-dependent Egypt, increasingly need the Nile s water for their own crop irrigation. Around the world, many other river systems are shared uneasily between neighbors in unstable regions: including the Ganges, the Mekong, the Jordan, and the Tigris and Euphrates rivers. Two-fifths of the world s population, living in 80 countries, now face some level of water shortage (Gleick, 1998, *see* Water Use: Future Trends, and Environmental and Social Impacts, Volume 3). Therefore, after the most war-scarred and arms-profiteering century on record, international conflict arising over environmental tensions continues to cast a long shadow over the prospects for human health.

These various global environmental change processes coexist, and share common underlying causes (Watson *et al.*, 1998). They often interact with one another, both in their genesis and their impact (Figure 1). Hence, the net impact of environmental changes upon the health of a population will typically reflect a complex configuration of environmental changes and the synergies amongst them.

A Caveat About the Handling of Uncertainties

Before reviewing the actual or potential impacts of these global changes on human health, it is important to underscore the complex and uncertain nature of the

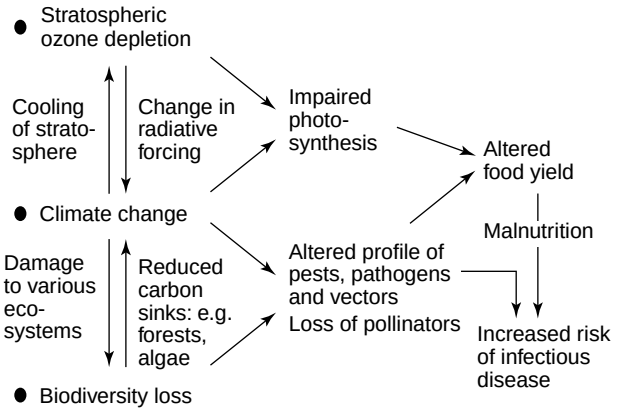


Figure 1 Examples of interactions amongst three global change processes, both in their genesis and in their impacts upon ecosystems and human health

risk assessments being made. The empirical study of single factor toxicity is relatively easy on two counts: first, the scientist can make direct observations of associations (i.e., empiricism), and, second, the assumed causal model is direct-acting and straightforward. Thus, we can calculate the risk of thyroid cancer in children exposed to measured levels of ionizing radiation following the Chernobyl explosion. However, it is a qualitatively different scientific task to assess how the construction of the Three Gorges Dam in China is likely to affect future patterns of vector-borne infectious diseases, since this latter assessment refers to the (non-observable) future and to disturbances of complex, non-linear, dynamic systems and processes.

The task becomes even more complex when the scientist is assuming certain scenarios of climate change for the year 2050 (around which there is inevitably much uncertainty), seeking to model how that will affect (for example) certain species of malaria-transmitting mosquitoes, and then linking that to models of populations with their estimated size, density, age structure, preexisting immunity, nutritional status, and so on. There are various types of uncertainty including, statistical variability (the familiar random variation in a world that is incompletely observed), uncertainties about the sensitivity of the response of particular variables or systems (e.g., mosquito breeding) to a change in an upstream antecedent system (e.g., temperature pattern), and frank ignorance of the identity of all the variables that are actually relevant to the health risk assessment exercise.

GLOBAL CLIMATE CHANGE AND HUMAN HEALTH

The potential impacts of global climate change provides a ready entry point to a more general consideration of the potential health impacts of global environmental change. The Second Assessment Report of the United Nation s (UN s) Intergovernmental Panel on Climate Change (IPCC) (IPCC, 1996) anticipates an increase in average world temperature of approximately 2 3 °C over the coming century (*see* Intergovernmental Panel on Climate Change (IPCC): an Historical Review, Volume 4). This, the report points out, would be more rapid than any temperature increase experienced by humans since the advent of agriculture around 10 000 years ago. Indeed, the most recent analyses of polar ice cores indicate that this would be very rapid warming relative to the global temperature chart over the past 420 000 years (which includes three earlier interglacial periods). Since climatic conditions and climate-sensitive natural processes are an integral part of Earth s life-supporting mechanisms, biologists expect such a change in climate would in uence the prospects for health and survival of many of the species and local populations in affected ecosystems (IPCC, 1996). Likewise, the health of our species, *Homo sapiens*, would be affected.

The various potential health impacts of climate change (direct and indirect, immediate and delayed) are summarized in Table 1 below (McMichael and Haines, 1997). The anticipated effects would be variously mediated via physical, chemical, ecological and social disturbances. Some health outcomes would be bene cial. For example, warmer temperatures would mean milder winters, and this should reduce the seasonal winter peak in deaths that occurs in temperate countries. In currently hot locations, an increase in summer temperatures might reduce the viability of mosquito populations. However, the judgement of scientists engaged in this topic area is that most of the anticipated health impacts of climate change would be adverse (IPCC, 1996). Shifts in climatic means and changes in variability would perturb important physical and biological systems to which human biology, culture and behavior is adapted.

Table 1 Categories of potential health hazards due to global climate change

Thermal stress and weather disasters
Heat waves and cold spells
Floods, storms, droughts, fire
Changes in physical and chemical exposures
Photochemical (and other) air pollutants
Pollens and fungal spores
Freshwater supplies and quality
Changes in ecological systems and relationships
Vector-borne infectious diseases
Water-borne and food-borne infectious diseases
Food yields (land and sea)
Social and demographic disruption
Examples: sea-level rise; impairment/disturbance of agriculture; environmental refugees

The direct health effects of climate change would include changes in mortality and morbidity from heat waves and thermal stress. Climatologists forecast an increase in the frequency of heat waves (because of both the rising mean temperature and a possible increase in weather variability). Some research has indicated that large urban populations in temperate zones, especially in non-coastal locations unbuffered by maritime climates, would experience several thousand additional health related deaths per summer by around 2050 (McMichael *et al.*, 1996). However, relatively little is known about the capacity of human populations to adapt, physiologically and culturally, to such changes across a decadal timeframe. Also, since extremes of heat and cold affect particularly the elderly and the sick, the average life-shortening impact may not be large (Rooney *et al.*, 1998).

Other direct-acting effects would include the respiratory health consequences of altered concentrations of aeroallergens (spores, moulds, etc.) or of air pollutants such as ozone which are produced by temperature sensitive photochemical reactions. The formation of ozone increases markedly as temperatures rise above average summer levels. A change in the tempo of extreme weather events, including storms and floods, would inevitably have a wide range of physical, infectious disease and psychological health consequences.

Of potentially greater consequence are the indirect health effects. Many of these would result from the perturbation of complex ecological systems, including alterations in the range and seasonality of vector-borne infectious diseases; altered transmission of person-to-person infections (including food poisoning and water-borne pathogens); the nutritional and health consequences of regional changes in agricultural productivity; and the various consequences of rising sea-levels. Population health would also be affected by demographic displacements and by regional conflicts over climate-affected shortages of agricultural and water resources.

Vector-borne Infectious Diseases

Changes in patterns of vector-borne infectious diseases would occur because of climatic influences on the vector organisms for diseases such as malaria, dengue fever, the various types of viral encephalitis and schistosomiasis (Martens, 1998). Increased temperatures and altered precipitation would affect the range, proliferation and behavior of vector organisms (e.g., mosquitoes) and intermediate hosts, as well as the viability and maturation rates of the infectious agents within the vector organism. While well-resourced populations in developed countries may be able to maintain public health defenses against an extension (or re-introduction) of vector-borne infections, poorer populations on the margins of endemic areas in tropical and subtropical countries would be at increased risk.

Tropical and Subtropical Regions

Two of the greatest vector-borne disease problems of the tropical and subtropical regions are malaria and dengue fever. Malaria is currently resurging in many countries where it had previously been eliminated or greatly reduced with vector control measures. More than 400 million cases of malaria are estimated to occur annually, including approximately one million (mostly childhood) deaths (World Health Organization, 1999). The *Anopheles* genus of mosquito that transmits malaria has a wide geographic range, that extends through much of the world s temperate zone; as did the disease itself up to the middle of the twentieth century. A malaria crisis is currently emerging in Africa, where poverty is combined with widespread increases in chloroquine resistance in the malarial plasmodium. In India, recent resurgence of malaria has been linked to the combined problems of chloroquine resistance, reduced efficiency of insecticides and the adaptive spread of anopheline mosquitoes into urban areas. The transmission of malaria is increased at warmer temperatures and at higher humidity, and in conditions where surface water is available for mosquito breeding.

Dengue fever is the world s most prevalent vector-borne viral disease, causing an estimated 100 million cases annually in tropical and subtropical countries. Dengue has increased markedly in recent decades, particularly in Central and South America due both to contraction of the earlier mosquito control insecticide spraying programs and to trends in population mobility, urbanization, poverty and regional warming. The *Aedes aegypti* mosquito that is the main vector for dengue is confined geographically by the 10 °C winter isotherm, at around 30° latitude, north and south. Both the intra-mosquito incubation rate of the dengue virus and the biting rates of the *Aedes aegypti* mosquito increase markedly at warmer temperatures.

Other tropical vector-borne infectious diseases that are known or likely to be sensitive to changes in climatic conditions are schistosomiasis (or bilharzia, spread by water-snails), leishmaniasis (spread by a sand fly), onchocerciasis (river blindness, spread by the black fly), and perhaps yellow fever (spread, like dengue fever, by the *Aedes aegypti* mosquito).

Recent analogue studies of malaria in Zimbabwe, Rwanda and Ethiopia have used natural climate variability to foreshadow aspects of future climate change impacts on health (Patz *et al.*, 1996). Those studies indicate that future warming would very probably cause malaria to move to higher altitudes and thereby affect currently unexposed highland populations. Other recent analogue studies have confirmed the sensitivity of both malaria and dengue to inter-annual climatic variations, especially those associated with the El Nino La Nina cycles (Bouma and Dye, 1997; Hales *et al.*, 1999). Interestingly, there have been reports

of recent upwards movement of malaria, dengue fever, or their mosquito vectors in several continents, in conjunction with glacier retreat and alpine plant ascent (Epstein *et al.*, 1998).

Simulation with integrated mathematical modeling, for standard scenarios of climate change, project that the geographic range of potential malaria transmission would undergo a net expansion (Martens, 1999). An additional 4 6% of the world s people (i.e., several hundred million people) would be at risk of malaria, depending on the assumed extent of climate change. Similar mathematically modeled projections have been made for dengue fever and schistosomiasis. While the models show that the geographic range of potential transmission of dengue fever would increase under the climate change scenarios, global warming is likely to lead to a diminution in the geographic range of transmissibility of schistosomiasis because the warming of water would reduce the viability of certain major species of water snail.

It must be emphasized that these global level models necessarily gloss over many regional differences in microclimate, local ecological relationships and biological differences between species of vector mosquitoes or snails. However, in future, with sufficient local detail and with greater computing power, it will be possible to apply this type of mathematical model at higher resolution to local conditions. The models provide estimates of the *potential* transmission of these diseases. The actual incidence of infection would depend on preexisting defenses (such as surveillance systems, environmental management, and rapid case treatment) and on how the populations respond to changes in risk.

Temperate Zone

Various vector-borne infectious diseases that occur in temperate regions are also sensitive to meteorological conditions. Such diseases are likely to undergo changes in patterns of occurrence under conditions of climate change. The diseases include:

1. Tick-borne viral encephalitis: This infection, which occurs in parts of Western Europe and Scandinavia, appears to be sensitive to winter seasonal temperature. There is some evidence that warming over the past two decades has caused this disease to increase its geographic range northwards in Sweden (Lindgren, 1998).
2. Leishmaniasis: This infection, shared by dogs, foxes and humans and spread by infected sand flies, is endemic in the rural Mediterranean region of Europe and in the eastern Mediterranean. Modeling studies indicate that climate change may extend the habitat of the sand fly vector northwards. The disease has been reported, anecdotally, in dogs and foxes at higher altitudes in the European Alps in recent years.
3. Lyme disease: This infection is caused by the spirochete *Borrelia burgdorferi* (McMichael *et al.*, 1996). It is transmitted in Europe and North Eastern USA by the infected tick *Ixodes ricinus*, and has been apparently increasing in incidence in at least the USA in recent decades. Temperature influences the tick s three-stage lifecycle (larvae, nymphs, adults) and, hence, the probability of tick infection by the spirochete. Very high temperatures reduce the survival of ticks. Land-use patterns and climatic conditions influence the density of other mammalian hosts to the tick, such as rodents and deer.

Person-to-person Infectious Disease

Climate change would also influence various directly transmitted infections, especially those due to contamination of drinking water and food, many bacteria and protozoa sensitive to temperature. Further, changes in the pattern of rainfall can disrupt surface water configuration and, hence, drinking water supplies. Hence, the occurrence of infectious diseases such as cryptosporidiosis and giardiasis, spread via contaminated drinking water, would be influenced by a change in climatic conditions. It is therefore relevant to note that rainfall intensity increased in the USA during the twentieth century (Karl *et al.*, 1995).

Research findings suggest, increasingly, that the spread of cholera is facilitated by warmer coastal and estuarine waters and their associated algal blooms. Recognition of this necessitates a revised transmission model of cholera: we must add an ecological dimension. Transmission is not simply person-to-person via the direct fecal contamination of local drinking water. Detailed electron microscopic studies have indicated that the tiny phytoplanktonic (algal) and zooplanktonic organisms (the base of the marine food web) act as a natural environmental reservoir, and amplification system, for the cholera bacterium (Colwell, 1996). There is now evidence indicating that the seasonal or inter-annual warming of coastal water around the coasts of Peru and Bangladesh facilitates the entry of cholera into the coastal communities. Further, this effect is probably heightened by the eutrophication of coastal and estuarine waters that occurs because of the increased loads of human-produced phosphates and nitrates in surface water and wastewater runoff.

Food Production, Hunger and Malnutrition

Yields of food, especially cereal crops, are sensitive to temperature, rainfall, and soil moisture. Climate change would result in regional gains and losses, reflecting the local balance of changes in temperature, soil moisture, carbon dioxide fertilization, and alterations in crop pest and pathogen activity (Rosenzweig and Hillel, 1998). Modeling studies consistently indicate that tropical and subtropical

countries would be negatively affected (Rosenzweig and Hillel, 1998). Further, many communities reliant on traditional agriculture may lack the resources and adaptability to switch to alternative crops and production methods. Although some mid-continental drying in summer months would occur in North America and Europe, high latitude temperate zones (e.g., Canada and much of Siberia) might gain in agricultural production. With longer-term continuation of climate change, however, the net impact upon world agriculture may be negative.

Poor and economically disconnected populations, unable to offset reduced yields via trade, would be most severely affected by reductions in local food production. Hunger and malnutrition increase the risk of infant and child mortality and cause physical and intellectual stunting. Energy levels, work capacity and health status are reduced in adults. One pioneering modeling study estimated the additional number of hungry people attributable to standard projections of climate change (as defined by the IPCC) by the year 2060, within a range of plausible future trajectories of demographic, economic and trade liberalization processes (Parry and Rosenzweig, 1993). The estimate varied, depending on the mix of other assumptions, between an additional 40 million and 300 million relative to a future background total of around 600 million hungry people. Subsequent modeling by Parry and colleagues, using updated climate inputs, indicates a figure of around 70 million by the 2080s, this figure reduces by approximately one half if greenhouse gas emissions are constrained to stabilize at around 550 750 ppm (Parry *et al.*, 1999).

Water supplies, an essential input to agriculture, animal husbandry and personal hygiene, may also be adversely affected by climate change in many regions. Tensions over freshwater shortages, especially in low-to-middle latitude locations where adjoining countries share river basins, would be exacerbated by changes in rainfall. Conflict and public health adversity might then result.

Sea-level Rise: Effects on Coastal Regions and Populations

Sea-level is forecast to rise by approximately 40 cm by 2100 (IPCC, 1996). This rate of rise would be several times faster than over the past century. Its importance derives from the fact that over half of the world s population lives within 60 km of the coast. Sea-level rise may affect various processes and structures upon which our food production, freshwater supplies, economies and public health depend. However, because sea-level rise is a relatively slowly occurring physical process, protective action should be possible in many coastal regions.

Some of the world s coastal arable land would be damaged by sea-level rise, via salination and erosion. Rising seas would salinate freshwater aquifers, particularly under small islands, and may cause sanitation problems for coastal populations. The 10 countries most vulnerable to sea level rise include Bangladesh and Egypt, with huge river delta farming populations, Pakistan, Indonesia and Thailand, all of which have large and relatively poor coastal populations.

A particular public health hazard would arise if small island populations and coastal populations in poor, populous countries with few material resources were displaced. A half meter rise (assuming today s population) would approximately double the number who experience flooding annually, currently around 46 million. Sea-level rise would also affect sewage and wastewater disposal, the physical safety of coastal structures, the viability of coral reefs, mangroves and wetlands (fish nurseries), and the ecology of certain infectious diseases (e.g., malaria and cholera). Sea-level rise, ocean warming, and alteration of currents and nutrient flows would all contribute to changes in marine ecosystems. The health related consequences of these disturbances include algal blooms (implicated in cholera transmission) and the production of toxins, such as ciguatera, in edible fish and shellfish.

STRATOSPHERIC OZONE DEPLETION, ULTRAVIOLET RADIATION AND HEALTH

Stratospheric ozone depletion is essentially a separate phenomenon from greenhouse gas accumulation in the troposphere. It is primarily caused by the buildup of human-made ozone-destroying gases in the stratosphere, such as the chlorofluorocarbons (used in refrigeration, insulated packaging and spray-can propellants). The resultant increase in UltraViolet Radiation (UVR) exposure is expected to peak within the first two decades of the 21st century and then to decline, in response to the substantially successful phasing out of the major ozone-destroying industrial and agricultural chemicals. The concentration of stratospheric ozone, and its ultraviolet shielding effect, should return to normal levels by around 2050.

For at least the first two-thirds of the twenty first century, we can expect the ongoing increase in UVR exposure to augment the severity of sunburn, the incidence of skin cancers, and the incidence of various disorders of the eye (especially cataracts). It could also cause some suppression of immune functioning, thus increasing susceptibility to infectious diseases and perhaps reducing vaccination efficacy (UNEP, 1998). Another potentially important, though indirect, health detriment could arise from ultraviolet-induced impairment of photosynthesis on land (terrestrial plants) and at sea (phytoplankton). Although such an effect could reduce the world s food production, few data are yet available.

Destruction of stratospheric ozone is continuing, and increases in ground level UVR flux are presumed to

have occurred, particularly at higher latitudes and altitudes (UNEP, 1998). The approximate estimates of increases in biologically active radiation (i.e., the UV band which causes skin reddening) in the 1990s, relative to the 1970s, are:

- Latitude 40 50 °N: 7% in winter/spring; 4% in summer/fall;
- Latitude 40 50 °S: 6% on a year round basis;
- Antarctic: 130% in spring;
- Arctic: 22% in the austral spring.

Skin Cancer

Most interest has centered on an expected increase in the risk of skin cancer. Epidemiologists have firmly established that solar radiation exposure increases that risk. However, the relationship differs according to the type of skin cancer.

Non-melanotic skin cancers are of two major histological types: basal cell carcinoma and squamous cell carcinoma. The risk of these cancers has generally been thought to correlate with cumulative lifetime exposure to solar radiation. However, recent evidence suggests that the relationship is more complex, at least for basal cell carcinoma, for which it appears that childhood exposure may be important. It has, accordingly, been estimated that persistence of the level of ozone loss experienced over the past decade or two would cause the subsequent annual incidence of basal cell carcinoma (the dominant type) to increase by 1 2% at low latitude (5°), 14% at 55 65° in the Northern Hemisphere (e.g., the UK), and 25% at that latitude in the Southern Hemisphere. The estimated percentage increases for squamous cell carcinoma would be approximately twice as great.

Malignant melanoma occurs much less often than the other two types, but it is much more likely to be fatal. The relationship between malignant melanoma skin cancer and UVR is complex. Overall, 60 90% of melanomas in fair-skinned populations are estimated to involve sunlight exposure. Repeated sunburn episodes in early life are thought by epidemiologists to be important for the development of melanoma.

The most recent assessment by the United Nations Environment Programme (UNEP) (UNEP, 1998) has updated the above projections for total skin cancer for a European (i.e., fair-skinned) population living around 45° North. Under the amended Montreal Protocol there will be an estimated excess incidence that peaks at around 5% during the third quarter of the 21st century. This entails an extra 100 cases of skin cancer per million population per year due to stratospheric ozone depletion, relative to the current background rate of skin cancer of approximately 2000 cases of skin cancer per million population per year. If the moderate aging of the European population is factored into the modeling, the excess incidence becomes, proportionally, a little higher. These calculations assume, for the sake of simplicity, that behavioral and other risk factors do not change, and that current ozone depletion rates and UVR exposure increases are sustained during the next several decades. Darker skinned populations would be affected less, both because of their lesser susceptibility to skin cancer and because the changes in ambient UVR levels are less at lower latitudes.

Ozone Depletion, Ultraviolet Radiation and Damage to the Eye

Exposure to UVR can also damage the tissues of the eye. The condition most directly linked to UVR exposure is corneal photokeratitis (snow blindness), which is caused by acute exposures and is the ocular equivalent of sunburn. Chronic exposure to UVR also contributes to conjunctival conditions such as pterygium (UNEP, 1998). The role of UVR in the formation of cataracts (opacities in the lens) remains complex and unclear. Some types of cataract appear to be associated with UVR exposure, while others do not. Since cataract accounts for a majority of the tens of millions of cases of blindness in the world, even a marginal impact of increased UVR exposure on their occurrence would be significant.

Ozone Depletion, UVR and the Immune System

There is evidence from both humans and experimental animals that UVR causes both local (i.e., occurring only at the site of irradiation) and systemic (body wide) suppression of immune system activity. The familiar effect of solar exposure on herpes simplex infection (cold sores) attests to the immunosuppressive effect of UVR. The consequences of immunosuppression for human health, particularly for susceptibility to infectious disease in human populations, are uncertain but potentially important.

UVR-induced changes in immune response may also affect autoimmune diseases. Increase in UVR may either suppress or aggravate the disease depending on the type of immune response that underlies the pathology. It has recently been proposed, on the basis of epidemiological and laboratory evidence, that increased levels of ambient UVR exposure are associated with decreased occurrence of multiple sclerosis. This may account for the otherwise unexplained strong positive correlation between latitude and prevalence of multiple sclerosis. UVR is known to aggravate lupus erythematosus, another autoimmune disorder, when used as a medical treatment (UNEP, 1998).

LAND DEGRADATION, OCEAN FISHERIES AND FOOD YIELDS

Land degradation and the loss of topsoil have become a worldwide problem in recent decades. During the 1980s, the

combination of erosion, desiccation and nutrient exhaustion, along with irrigation-induced water-logging and salination, rendered unproductive one-fifteenth of the world s 1.5 billion hectares of readily arable farmland (World Resources Institute, 1998; UNEP, 1999). The world s per person production of cereals, the main source of dietary energy, seems to have faltered a little since the mid-1980s (Kendall and Pimentel, 1994). It is not easy to assess to what extent this reflects a decline in the productive resource base. Other political, economic and consumer trends have also influenced patterns of world agriculture (Dyson, 1999). Nevertheless, the Green Revolution , which fed much of the expanding Third World population during the 1960s to the 1980s, depended greatly on the combination of laboratory-bred high-yield cereal grains, fertilizers, groundwater and arable soils. In retrospect, the productivity gains of the Green Revolution are assessed to have come, in part (and probably substantially) from the exhaustion of natural resource capital, especially depletion of topsoil and groundwater (*see* **Green Revolution**, Volume 3).

Meanwhile, at sea, many of the world s great sheries are now on the brink of being over-exploited (*see* **Changes in World Marine Fish Stocks**, Volume 3). The UN s Food and Agricultural Organization estimates that we have neared the sustainable sh-catch limit: around 100 million tones per year (Food and Agriculture Organization, 1995). Although yields from aquaculture are increasing, it faces some unusual ecological dif culties, with problems of water pollution and infection of the farmed sh. Further, it is much less energy ef cient than agriculture since plant or animal feed stuffs (compared to sunlight!) are needed as the energy inputs.

Today, as we continue on the treadmill of having to extract greater food yields to feed ever more people, almost one-tenth of the world population is malnourished to an extent that impairs health. The *absolute* numbers of malnourished persons in the world, especially children, are still growing. This longstanding problem is today being exacerbated by the rapidly widening disparity between rich and poor: evident in the contrasting trends of widespread continuing hunger and steady increases in the prevalence of obesity in urban populations everywhere.

LOSS OF BIODIVERSITY, AND SPREAD OF INVASIVE SPECIES

An increasingly pervasive problem is the worldwide loss of biodiversity: the loss of species, strains and local populations and, more serious, the weakening or collapse of ecosystems when key species disappear. Through our own species spectacular reproductive success and our energy-intensive economic activities, we have occupied, damaged or eliminated the natural habitat of many other species. Biologists estimate that this fastest-ever mass extinction may cause around one third of all species alive last century to be gone before the end of the twenty- rst century (Soule, 1991; UNEP, 1999). The adverse consequences of the loss or weakening of whole ecosystems would include reduced yield from natural and managed food producing systems and disturbances of the ecology of various vector-borne and other infections.

A continuing loss of biodiversity will forfeit a rich repertoire of genetic and phenotypic material. In particular, our major cultivated food plants are selectively enhanced descendants of wild strains. To maintain their hybrid vigor and environmental resilience, a diversity of wild plants needs to be preserved as a source of genetic additives. Similarly, a high proportion of modern medicinal drugs in western medicine has natural origins, and many defy synthesis in the laboratory. Scientists test many thousands of novel chemicals from nature each year, seeking new drugs to treat HIV, malaria, drug-resistant tuberculosis, cancers, and so on.

About one quarter of all Western medicines and pharmaceuticals come from plants and another quarter from animals and microorganisms. This includes analgesics, antibiotics, tranquilizers, diuretics, steroid contraceptives and anti-cancer drugs. Their annual commercial value is around US$40 billion. Myers has calculated that the usual hit rate in the bioprospecting of tropical plant extracts for new medicinal drugs is of the order of 10^{-5} to 10^{-6} (Myers, 1997). Therefore, the as yet unprospected extracts (estimated at 750 000, each of which could be screened about 500 ways for different properties) should yield at least 375 new drugs. This compares with the 48 in current use (including curare, quinine, codeine, pilocarpine and the life-saving cancer treatment drugs vincristine and vinblastine). This simple arithmetic suggests that ve-sixths of tropical vegetative nature s medicinal goods have yet to be recruited to the bene t of humankind.

But nature s medicine chest is much larger than that. From the animal kingdom come many promising compounds, secreted by frogs (antimicrobials and potent painkillers), sea-urchins, crustaceans, barnacles and many others. Presumably some simple arithmetic could be done here, too, that would underscore the immense option value of these as yet undiscovered, unused, goods.

Closely related to the loss of species is the introduction of invasive (non-indigenous) species into distant, unrelated, ecosystems. This process is accelerating in today s world as a result of chance, ignorance or human folly, combined with the global escalation in human mobility and trade. Invasive species can affect human health via several pathways (McMichael and Bouma, 2000). Changes in the continental range and incidence of human infectious disease is the major, age-old, example. Other introduced species that impair food yields (pests, pathogens, weed competitors, or herbivores such as rabbits in Australia), food storage

(moulds and pests such as rodents and sparrows) or that produce food-borne biotoxins (various species of aquatic algae and toxin producing moulds) also pose risks to human health.

The new fluidity in patterns of human infectious disease that has been evident over the past several decades provides a clear signal that ecological opportunities and physical dissemination pathways for microbes are increasing in the contemporary world (Wilson, 1995). The geographic range of many human infectious diseases has increased with recent changes in patterns of land use and regional climate. Intercontinental trade and rapid long-distance human mobility have introduced various pathogens into distant populations. Some of the more dramatic examples include the introduction of the Asian tiger mosquito (*Aedes albopictus*, which spreads dengue fever and yellow fever) to the Americas, Africa and South East Europe, the reintroduction (after a century of absence) of the cholera bacterium to South America, and the spread of the life-threatening *Escherichia coli 0157* bacterium in meat exports. Meanwhile, antibiotic-resistant bacteria (e.g., tuberculosis, staphylococci, enterococci), a special type of invasive variant, are now becoming a widespread problem for human health (*see* Biological Invasions, Volume 2).

PERSISTENT ORGANIC POLLUTANTS (POPs)

We are by now very familiar with the occurrence of local toxic chemical hazards from environmental pollution of air, water, soil and food. They are the continuing focal point of most public concern about environmental hazards. However, it is becoming clear that there is another, potentially serious, dimension to this problem: many organic chemical pollutants are stable and persistent and, as their entry into the environment builds up, they are tending to disperse more widely. This, of course, has long been the case for non-biodegradable heavy metals, such as lead, and perhaps for long lasting radioactive elements, such as strontium-90. But we have recently learnt that various semi-volatile organic chemicals are readily spread from low to high latitude, via atmospheric global distillation processes, and that they may consequently achieve higher environmental and biotic concentrations at those higher latitudes (Harner, 1997). In recent years the concentrations of bioaccumulated toxic chemicals in fish eaten by the Inuits has increased.

There is therefore, today, a heightened concern over compounds such as dioxins, dichlorodiphenyltrichloroethane (DDT), the polychlorinated biphenyls (PCBs), toxaphene, dieldrin and hexachlorobenzene (Watson *et al.*, 1998; UNEP, 1999). These accumulate in the fatty tissues of many organisms, particularly predators at the top of the food chain. There is suspicion that several of these compounds may cause birth defects or cancer in humans. Some of them are known to interfere with the normal functioning of mammalian, bird and fish organ systems: the reproductive system, the immune system and the central nervous system. Many examples of damage to wildlife have been documented, including declines in bird populations due to eggshell thinning, the poisoning of aquatic organisms, and the apparent suppression of immune function in seals.

It is not yet clear what health effects from these exposures might be occurring in human populations. Insidious effects on intellectual functioning, on fertility, or on immune activity are not easy to detect in epidemiological studies. Nor, yet, has much direct evidence been sought of changes in indices of human biological function in association with exposure to POPs. There is some suggestive evidence of a decline in the density and quality of male sperm over the past half-century, which could reflect the impact of environmental endocrine disrupting chemicals (Sharpe and Skakkebaek, 1993). However, this remains a contentious issue.

TRADE, LAND-USE, ECONOMIC DEVELOPMENT AND INFECTIOUS DISEASES

In the 1950s, René Dubos pointed out that technological innovation, whether industrial, agricultural or medical, disrupts natural ecological balances and unleashes infectious diseases (Dubos, 1959). During the third quarter of the 20th century, as the Western style model of economic development was urged upon all non-communist Third World countries, it became clear that large-scale human interventions in the natural environment often, and usually adversely, affected infectious disease patterns. During the 20th century, large development projects such as dams, irrigation schemes, land reclamation, road construction and population resettlement programs (as is currently occurring in Indonesia) have often potentiated the spread of diseases such as malaria, dengue, schistosomiasis and trypanosomiasis. Such large-scale developments in the Eastern Mediterranean, Africa, South America and Asia have been consistently associated with increases in communicable diseases, especially schistosomiasis and filariasis (which includes elephantiasis). However, because changes in land-use patterns are typically accompanied by changes in population density, population mobility, pesticide usage, and regional climate, it is difficult to assign specific causal explanations. Indeed, many of the health outcomes result from interactions between these various change processes.

Long-distance shipping apparently introduced the cholera bacterium into Peruvian coastal waters in 1991, causing the first outbreak of cholera in that continent for nearly 100 years (Colwell, 1996). The inoculation coincided with unusual warming of coastal waters and extensive algal blooms at the inception of an El Nino event. Similarly, in the 1980s, intercontinental trade in used car tires introduced the East Asian mosquito vector dengue fever, *Aedes*

albopictus, into South America, the Southern USA and Africa. At about the same time, the mosquito vector for Ross River virus disease, *Anopheles camptorhyncus*, probably reached New Zealand on a ship that transports pine wood across the Tasman Sea between New Zealand and Australia.

Modern patterns of human labor, touristic mobility and sexual networks potentiated the spread of HIV from a quiet backwater in central Africa. Other aspects of modern human ecology (the rapid worldwide urbanization, intensification of food production and distribution systems, and various clinical and recreational exchanges of blood and other tissues between human individuals) are all opening new doorways for microbial traffic.

Land Use Patterns and Infectious Disease

Land-use patterns influence patterns of infectious diseases. As population pressures increase, more forest and other virgin land is cleared for agricultural and pastoral purposes. The composition of vector species generally changes following changes in environmental conditions. For example, deforestation in the Indo-Australian region has enabled malaria-transmitting *Anopheles punctulatus* species to become established. In contrast, several *Anopheles* species (including *Anopheles dirus* in Thailand and *Anopheles darlingi* in South America) have disappeared following deforestation that removed the flora and fauna upon which they depended.

Forest clearance in South America during recent decades, to extend agricultural land, has mobilized various viral haemorrhagic fevers that previously circulated quietly (and generally benignly) in wild animal hosts. For example, the Junin virus, which causes Argentine haemorrhagic fever, naturally infects wild rodents (the mouse, *Callomys callosus*). However, extensive conversion of grassland to maize cultivation in recent decades stimulated a proliferation of this species of virus-bearing mouse, thus exposing human farm workers to this new virus. In the past 35 years, the land area carrying this new human disease has expanded seven-fold and the average annual number of infected persons is of the order of several hundred, up to one-third of whom die (WHO, 1997). Likewise, other new viral infectious diseases in rural South American populations have been caused by the Machupo virus, Basia virus, Oropouche fever, and so on. In South America, leishmaniasis (spread by sand flies) is now widely considered an occupational disease of forest workers engaged in land clearance for agriculture, road building, timber extraction, or mining.

Ongoing land clearance in other populous regions of the developing world, and the extension of irrigation, continue to influence the patterns of infectious diseases. In the Sudan, for example, schistosomiasis appeared within several years of the start of the Gezira Scheme to irrigate cotton production; simultaneously, the prevalence of malaria also increased markedly in this region. Dam building and extensions of irrigation also affect the geography of malaria and other mosquito-borne infections. For example, outbreaks of Rift Valley Fever occurred in the Nile Valley in 1977 and in Mauritania in 1987 following the damming of major rivers. In Senegal, the irrigation of rice and sugarcane agriculture that followed the construction of the Diama Dam on the Senegal River caused a massive outbreak of schistosomiasis (WHO, 1997). Similarly, the building of the Aswan dam on the Nile River resulted in a seven-fold increase in schistosomiasis in the local population. Lymphatic filariasis in the southern Nile Delta has increased 20-fold in prevalence since the 1960s, primarily due to the increase in breeding sites for the mosquito *Culex pipiens* following the rise in the water table caused by extensions of irrigation. The situation has been exacerbated by increased pesticide resistance in mosquitoes due to heavy agricultural pesticide use and by rural-to-urban commuting among farm workers.

Other aspects of land use also contribute to the generation of new or amplified infectious disease agents. An example of the production of new variants of infectious agent is that of the influenza virus. This virus has its origins in birds. As the number and density of co-habiting farmers, ducks and pigs in southern China increases, so do the probabilities multiply of creating and launching new genetic recombinants of the influenza virus. Pigs infected by multiple strains of avian influenza virus act as genetic mixing vessels, yielding new recombinant deoxyribose nucleic acid (DNA) viral strains that may then infect the pig-tending humans (*see* Infectious Diseases, Volume 2).

Changes in patterns of rainfall, runoff and surface water movement would have implications for the volume and quality of freshwater supplies. The deforestation of hillsides exacerbates the runoff of rainfall, causing flooding in the short term and reduced water availability in the medium term. In some regions, including current semi-arid regions, rainfall may decline because of climate change; in other regions rainfall is forecast to become more intense. Flooding and overloading of wastewater systems would often contaminate freshwater supplies: as occurred in Tajikistan, in 1994, when unusually heavy rainfall overwhelmed the urban sewage system, causing an outbreak of typhoid fever.

URBANIZATION AND HEALTH

An integral part of globalised changes in the world environment is the spectacular shift of human populations into cities. Humans now appear to be destined to become an urban species. The urbanised proportion of the world s population has grown from around 5 to 50% in the past two centuries, and is still rising. This urban migration reflects,

variously, the advent of industrialization, the contraction of rural employment, the flight from food insecurity and other forms of insecurity, and the search for jobs, amenities or stimulation.

This massive urbanization is occurring in a world that is, today, changing rapidly on many fronts: economic, social and environmental. Today s world is one in which local urban employment opportunity is affected, often abruptly, by global free market forces; in which the sprawling growth of slums and shanty towns is an expression of persistent and markedly widening economic inequalities. As a result the spread of new and resurgent infectious diseases is facilitated by unhygienic conditions, intensified patterns of human contact (sexual and other), regional environmental change and ecological disruption, and rapid human mobility and long-distance trade.

The urban environment, physically, confers various benefits upon human health and wellbeing. There is readier access to health services, education, financial and social services. The urban environment is stimulating, open, and rich with new opportunities. Yet big cities are often unfriendly, impersonal, and sometimes menacing. Noise, traffic congestion and residential crowding are stress inducing. The air quality is poor in cities around the world, particularly in many of today s large Third World cities. An estimated 130 000 deaths and 50 70 million incidents of respiratory illness occur each year due to episodes of urban air pollution, half of them in East Asia (WHO, 1997). An even greater toll of chronic disease is attributable to long-term exposures to air pollution. Third World urban populations, particularly the very poor in slums and shantytowns, also face the hazards of unsafe drinking water, solid waste accumulation and toxic wastes (*see* Urban Poverty and Environmental Health, Volume 3).

Urban Transport

The density of private cars in the world s cities is escalating rapidly. In large part this reflects the unplanned and uncontrolled domination of urban transport systems by roads and cars in a world in which national governments have become increasingly beholden to free-market orthodoxy and in which transnational corporations (including car manufacturers and petrochemical companies) assume increasingly powerful, global, influence. Currently, we appear to have little vision of an urban future that is not dominated by privately owned vehicles. Accordingly, traffic congestion now characterises cities nearly everywhere: although the past decade has seen the beginnings of a swing back to light rail (especially tram) systems in many European and some US cities (Newman and Kenworthy, 1999).

In addition to the fragmentation of neighbourhoods, intrusive noise, and restrictions on physical exercise, there are three broad categories of public health hazard from this burgeoning urban car traffic:

1. Fatal and non-fatal injuries of car occupants, pedestrians and cyclists. Over three-quarters of a million deaths from road crashes occur annually, the great majority of them in developing countries.
2. Exhaust emissions that cause local air pollution, particularly photochemical smog during summer months. Urban air pollution has, in recent decades, become a worldwide public health problem. In China, urban air pollution with particulates and sulfur dioxide is increasing sharply in association with rapid industrialization and the proliferation of automotive vehicles. The main source of pollution is the industrial use of coal, with its relatively high sulfur content. However, emissions from domestic cooking and heating fuels are also increasing. Morbidity and mortality in Chinese cities is expected to increase steeply over the coming two decades.
3. Exhaust emissions that contribute to acid rain and to the global accumulation of the greenhouse gas carbon dioxide, each of which has wider ranging consequences for human health. In developed countries, traffic exhaust gases account for around one quarter of national greenhouse gas emissions.

Heatwaves, Urban Vulnerability and Mortality

Severe heat waves adversely affect health. It is very likely that the frequency, if not the intensity, of heat waves will increase in the 21st century as world temperatures rise. Obversely, some temperate-zone populations will experience an amelioration of winters, with fewer periods of severe cold. The mortality impact of heat waves is typically greatest in the centers of large cities, where not only do temperatures tend to be higher than in the leafier suburbs and the surrounding countryside, but the relief of night-time cooling is lessened. This heat island effect is due to the large heat-retaining structures and treeless asphalt expanses of inner cities and the physical obstruction of cooling breezes (*see* Urban Heat Island, Volume 3).

In July of 1995, in the USA, more than 460 extra deaths were certified as due to the effects of the extreme heat wave in Chicago when temperatures reached 40 °C. The rate of heat-related deaths was much greater in black than in white people, and in persons bedridden or otherwise confined to poorly ventilated inner city apartment blocks. In the correspondingly severe five-day heat wave in England and Wales in July August 1995, a 10% excess of deaths occurred in all age groups, particularly in adults, from respiratory and cerebrovascular disease. In Greater London, where daytime temperatures were higher (and where there was less cooling at night), mortality increased by around 15% (Rooney

et al., 1998). The excess mortality was generally greater in socio-economically deprived groups, who typically lack access to air-conditioning of any kind. (The corollary of this heat-wave-induced rise in mortality was a lessened winter seasonal peak of mortality in an above-average warm year.)

There is a need for studies of the health impacts of extremes of heat and cold in Third World urban populations. It is important, especially in anticipation of global climate change, that urban planning and development should seek to minimize the adverse impacts.

Urban Ecological Footprints

Urbanism can affect human health directly, as a function of the properties of the urban physical and social environment. It also has an influence on a much larger spatial-temporal scale, via its distant impact on the integrity and productivity of the huge environmental penumbra that is the source of materials and services for urban populations. These are the *ecological footprints* of the city (*see* Ecological Footprint, Volume 3).

There are undoubted ecological benefits of city living. Cities confer various economies of scale, of proximity, and of shared use of resources. There are great but largely unrealized possibilities for reuse and recycling. Equally, there are great externalities. Urban populations do not subsist on their own urban land; they depend on food grown elsewhere, on raw materials (timber, metals, fossil fuels, etc.) extracted from elsewhere, and on having their wastes disposed of elsewhere. Thus, an urban population imports raw materials (thereby causing environmental damage at the source and in transit), makes demands on ecological systems on land and sea to produce food, dumps wastes into rivers, lakes and oceans, buries wastes on land, and emits wastes as gases and particulates into the atmosphere.

Urban populations thus depend on the natural resources of ecosystems that, in aggregate, are vastly larger in area than the city itself. The highly urbanized Netherlands draws in resources from a total surface area 15 times larger than itself. A study of the renewable resource appropriations by the cities of the Baltic Sea region showed that the amounts of resources consumed by 29 cities (wood, paper, fibers and food (including seafood)) depend upon a total area 200 times greater than the combined area of those cities (Folke *et al.*, 1996). That figure of 200 comprises 17 units of forest, 50 units of arable land and 133 units of marine ecosystems.

The scale of impact of urban populations is growing, and now includes massive urban contributions to the world s problems of greenhouse gas accumulation, stratospheric ozone depletion, land degradation, coastal zone destruction, and aquifer emptying, Via this scaled-up externalized impact, urbanism is thus jeopardizing the health of current and future generations.

DIFFERENCES IN POPULATION VULNERABILITY TO CLIMATE CHANGE AND OTHER GLOBAL CHANGES

The health impacts from global environmental change will vary around the world. Not only will the resultant environmental stresses vary in type and intensity in different geographic regions, but, in general, poor populations will be the most vulnerable; whether at the regional, national or subnational levels. With respect to climate change, the impacts of heat waves, other extreme weather events (e.g., tropical cyclones and droughts) and gradual coastal inundation would affect such populations most. Much of the anticipated increase in contagious and vector-borne infectious disease would occur in developing countries (McMichael *et al.*, 1998).

In many African, Asian and Latin American countries, the average life expectancy is 20 30 years less than for rich Western countries. Infectious diseases remain the main killer, particularly in children below the age of five. Much of this health deficit in poor Third World countries reflects the widespread poverty, the adverse social consequences of export oriented economic development, and the environmental adversity caused by exploitation of natural resources. In today s increasingly globalized economy, which operates to the general disadvantage of poor countries, the exacerbation of land degradation, rural unemployment, food shortages and urban crowding all contribute to health deficits for the rural dispossessed, the underfed and the slum dweller.

The plight of Sub-Saharan Africa, with its entrenched poverty and marginalization from the global economy, illustrates well these complex relationships. The region includes 33 of the world s 50 poorest countries, and trends in health, education, and living standards have reversed in the past two decades and are continuing to fall. Meanwhile, the region s population growth rates are the highest in the world, land degradation is increasing, and tropical forest clearance continues (World Resources Institute, 1998). A majority of Sub-Saharan Africans live in absolute poverty (Logie and Benatar, 1997). More than half still lack safe water and 70% lack proper sanitation. Infant mortality rates are 55% higher than in the world s other low income developing countries; average life expectancy, at 51 years, is 11 years less. Malaria and tuberculosis are widespread and increasing, while in parts of central, Southern and Eastern Africa one in three pregnant women are HIV positive. There can be no doubt that the two-way relationship between poverty and ill health erodes African economic productivity substantially, and also renders African populations very vulnerable to the adverse impacts of superimposed environmental changes.

There are various social, cultural and political influences on population vulnerability to environmental change. For example, Woodward and colleagues (Woodward *et al.*,

1998) have described poverty, environmentally destructive growth, political rigidity, dependency and isolation as causes of increased vulnerability. Their reasoning is as follows:

1. Impoverished populations are at greater risk because they have fewer choices, less information, and inadequate physical and social infrastructure.
2. Rapid increases in population size, density of settlement and the use of natural resources may increase vulnerability by damaging those ecosystems that buffer against environmental adversity. The devastation caused by Hurricane Mitch in November 1998 was, reportedly, greatly ampli ed by deforestation of large hillsides in the Honduras.
3. Political rigidity may have contributed to recent, greater than expected, impacts of climate-related disasters in parts of Asia.
4. In a globalizing world in which increased mobility of people and goods is connecting geographically remote countries to the international economic system, unconnected remoteness of a country may become a liability.

Changes in the age structure of populations will also affect vulnerability. As average life expectancies increase, so the proportion of elderly persons increases. Further, in developed countries increased longevity generally means an increase in the number of years lived with disabilities. In many developed countries, elderly populations have increased disproportionately in coastal areas and it is these areas that are, in general, more vulnerable to extreme weather events.

The impact of regional changes in food and water availability will be of particular public health importance in potentially vulnerable regions where population growth is pronounced and food insecurity already exists. Interactions between local environmental degradation and larger-scale environmental changes are likely to be important in determining the net impacts on human health and society. For example, local deforestation due to population pressure may directly alter the distribution of vector-borne diseases, and contribute to local increases (*see* **Demographic Change: the Aging Population**, Volume 3) in temperature (on top of any global increase).

In many populations, environmental insecurity is not simply borne of land exhaustion or the exceedance of local carrying capacity. Many traditional communities are under pressure because of the encroachment of increasingly globalized commercial activity. For the moment, con ict may not be evident because of the unequal power relations, but there certainly are widespread tensions. This encompasses, forest tribes in the Amazon (confronted by deforestation, displaced landless peasants, and the biotoxic methods of uncontrolled river bed gold mining), Ladakhi farmers whose traditional sustainable methods of farming are being undercut by the trucking in of government subsidized produce from large-scale intensive farming in India, and traditional rural communities in Zimbabwe where adverse maternal and child health consequences have resulted from the imposition of insecure seasonal employment in large export oriented plantations. Today s increasingly globalized food trade may, at a larger scale, have various adverse effects on whole countries and regions.

Demographic Entrapment: Controversial but Potentially Serious

Because of the combination of poverty, environmental decline and hunger with population growth, some of the world s poorest populations may be becoming demographically entrapped (King, 1990). That is, the weight of their current or projected numbers currently exceeds, or soon will exceed, the carrying capacity of their environment, and, lacking trade and migratory safety valves, they face starvation, disease or fratricide. Rwanda, where population size rst exceeded certain estimates of carrying capacity (around 6 7 million) in the 1980s, may be an example. Demographic, economic and environmental indicators in Malawi look similarly precarious.

Last (1995), too, refers to the notion of demographic entrapment, and describes how the desperate efforts of such populations to provide themselves with food, water, and fuel-wood can degrade an already fragile ecosystem into a desert that may take centuries to recover. This, he says, is happening in parts of Sub-Saharan Africa, in alpine foothills in the Himalayas and the Andes, in crowded small nations such as Haiti and Honduras, and elsewhere in Central and South America. Nevertheless, a reassuring counter-example comes from the experience of the Machakos district in Kenya, where, since the 1930s, a six-fold increase in population has been accompanied by a substantial regeneration of previously seriously depleted local soils and forested areas: albeit helped by an abundance of local rains in that location.

The notion of demographic entrapment remains controversial, and it is of course very dif cult to study. Indeed, debate about the relationship between carrying capacity and demographic entrapment discomforts many demographers. These complex life-sized problems, af icting whole regions or populations, are not reducible to simpli ed linear analysis. We need to develop a more interdisciplinary type of scienti c assessment that enables us to integrate across disciplines and sectors.

HEALTH AS A SUSTAINABLE STATE

The expectation that global environmental change will have various impacts, mostly adverse, on human health raises the unfamiliar issue of the *sustainability* of population

health. We conventionally measure population health cross-sectionally, as an *achieved* entity. This type of measure reflects, in large part, society s recent consumption of natural capital : but it tells us nothing about the possibility of sustaining (or improving upon) those achieved levels in the future. This is analogous to our conventional assessment of society s economic performance, wherein we measure accrued wealth and achieved output rather than some index of sustainable productivity.

To argue, therefore, that environmental conditions must actually be getting better since life expectancies are increasing is to misconstrue this type of environmental hazard. Gains in life expectancy tend to happen in circumstances conducive to rapid population growth that is, when the immediate carrying capacity (supply) of the environment exceeds the number of dependent individuals (demand). These generalized gains in human life expectancy indicate that, currently and recently, the life-supporting capacity of the human-modulated environment has been increasing. But we must ask, at what cost, or at what future risk?

The considerable and widespread gains in health and longevity over the past century have depended primarily on reductions in early-life infectious disease mortality. Basic gains in food security and in sanitation, supplemented by advances in vaccination, antibiotic treatment and oral rehydration therapy, have changed the profile of infectious disease mortality. These technical and social improvements have been closely bound up with the processes of urbanization, industrialization and increasing material wealth. They, and the resultant gains in life expectancy, have therefore (seemingly paradoxically) proceeded in parallel with increasing levels of physical disruption and chemical contamination of our ambient environment.

For how long can we expect to maintain these parallel increasing trends in consumption, life expectancy and environmental impact? At what stage might depletion of the world s ecological and biophysical capital rebound against the health of human populations? As discussed above, many of Earth s vital life-supporting systems are today showing signs of unprecedented systemic stress.

Risks to health will tend to increase so long as human societies depend upon a linear, high-throughput, waste-generating metabolism that is at odds with the circular metabolism of the rest of nature (wherein every output becomes an input). The high consumption lifestyle of Western nations depends greatly on continued access to inexpensive inputs from non-Western nations, resulting in depletion of natural resources or the co-option of traditional agriculture into export crop production. Meanwhile, as developing countries pursue their own economic aspirations, there will be additional strains upon the biosphere: and, therefore, increasing risks to population health in countries everywhere.

After two centuries of hugely increasing fossil fuel combustion by today s rich countries, further rapid increases in energy use in East and Southeast Asia must be anticipated to contribute greatly to the accumulation of greenhouse gases (*see* National Responsibilities for Greenhouse Gas Emissions, Volume 3). Continual forest clearance brings exposure to new and potentially mobile infective organisms such as hemorrhagic fever viruses (e.g., various arenaviruses in Latin America, Ebola virus in Africa). The over-fishing of the ocean and the decline and collapse of some fisheries appears to preclude any further gains in *per capita* supplies of seafood. Continued pressure on agro-ecosystems in vulnerable regions, coupled with land degradation and population growth, will (in the absence of yield-enhancing genetic modifications of foods, and their politically enlightened application in regions of greatest need (Hazell, 1999; Seragedlin, 1999)) increase migratory pressures and their attendant public health risks.

CONCLUSION

In this new millennium we are beginning to understand that large-scale damage to global and regional natural systems endangers the long-term sustainability of population health. We must think about population health and its determinants within a more explicitly ecological, systems-based, framework.

The prevailing economic ethos is shaped by rapid technological change, consumerism, a discounting of distant and deferred environmental impacts, and a pervasive free-market philosophy. Currently, the policy-setting role of national governments is contracting, as trade and financial transactions become globalized and deregulated, as the balance of power between private and public sectors shifts, and as the resultant cost-cutting competitiveness between nations puts a squeeze on social expenditures. Hence, at a time when coordinated, strong and farsighted government is needed to constrain damage to the world s ecological infrastructure, and hence too human health, we are instead entrusting our futures to the myopia of the marketplace.

Today, we face an unfamiliar range of hazards to human health from the various emergent global environmental changes. We must integrate this prospect into our thinking, planning, and preventive policy-making: without allowing it to diminish the importance of dealing with existing public health problems. In fact, both existing and future environmental health problems share many of the same underlying causes, relating to poverty, inequality, and social economic values and practices. The existing insecurity and vulnerability of many Third World populations makes more important the need to slow the environmental change processes while also shoring up the protective and adaptive capacities of populations in tropical countries.

In the near future we can expect to see increased efforts towards achieving a *sustainability transition*: one that embraces the essential idea of sustaining the conditions for healthy human life. This transition would require a transformation of our social and economic structures, values and practices. This topic is taken up in some detail in Volume 5. For public health scientists, the challenge is to forecast the population health impacts of a future world with and without such social and environmental change. The challenge for policy-makers and for the public at large is to incorporate the goal and values of ecological sustainability, and the sustaining of population health, into social policy.

Other articles relating to human health are to be found in: Infectious Diseases, Volume 2; Malnutrition, Infectious Diseases and Global Environmental Change, Volume 3; Policy Responses to Public Health Issues Relating to Global Environmental Change, Volume 4; WHO (World Health Organization), Volume 4; Theories of Health and Environment, Volume 5.

REFERENCES

Bouma, M J and Dye, C (1997) Cycles of Malaria Associated with El Nino in Venezuela, *J. Am. Med. Assoc.*, 278, 1772 1774.

Colwell, R (1996) Global Climate and Infectious Disease: the Cholera Paradigm, *Science*, 274, 2025 2031.

Dubos, R (1959) *Mirage of Health: Utopias, Progress, and Biological Change*, Harper and Row, New York.

Dyson, T (1999) Prospects for Feeding the World, *Br. Med. J.*, 319, 988 991.

Epstein, P R, Diaz, H F, Elias, S, Grabherr, G, Graham, N E, Martens, W J M, Mosley-Thomson, E, and Susskind, J (1998) Biological and Physical Signs of Climate Change: Focus on Mosquito-borne Diseases, *Bull. Am. Meteorol. Soc.*, 78, 409 415.

Folke, C, Larsson, J, and Sweitzer, J (1996) Renewable Resource Appropriation, in *Getting Down to Earth*, eds R Costanza and O Segura, Island Press, Washington, DC.

Food and Agriculture Organization (1995) *State of the World's Fisheries, 1995*, FAO, Rome.

Gleick, P H (1998) *The World's Water. The Biennial Report on Freshwater Resources 1998- 1999*, Island Press, Washington, DC.

Hales, S, Weinstein, P, Souares, Y, and Woodward, A (1999) El Nino and the Dynamics of Vector-borne Disease Transmission, *Environ. Health Perspect.*, 107, 99 102.

Harner, J (1997) Organochloride Contamination of the Canadian Arctic, and Speculation on Future Trends, *Int. J. Environ. Pollution*, 8, 51 73.

Hazell, P B R (1999) *Agricultural Growth, Poverty Alleviation, and Environmental Sustainability: Having it All*, 2020 Brief No. 59, International Food Policy Research Institute, Washington, DC.

Intergovernmental Panel on Climate Change (IPCC) (1996) *Second Assessment Report. Climate Change 1995*, Cambridge University Press, New York, Vols. I, II, III.

Karl, T R and Knight, R W (1995) Trends in High-frequency Climate Variability in the Twentieth Century, *Nature*, 377, 217 220.

Kendall, H W and Pimentel, D (1994) Constraints on the Expansion of the Global Food Supply, *Ambio*, 23(3), 198 205.

King, M (1990) Health is a sustainable state, *Lancet*, 336, 664 667.

Last, J M (1995) Redefining the Unacceptable, *Lancet*, 346, 1642 1643.

Lindgren, E (1998) Climate and Tick-borne Encephalitis in Sweden, *Conserv. Ecol.*, 2, 5 7.

Logie, D E and Benatar, S R (1997) Africa in the 21st Century: Can Despair be Turned to Hope? *Br. Med. J.*, 315, 1444 1446.

Martens, W J M (1998) *Health and Climate Change: Modeling the Impacts of Global Warming and Ozone Depletion*, Earthscan, London.

Martens, W J M, Kovats, R S, Nijhof, S, de Vries, P, Livermore, M T J, Bradley, D, Cox, J, and McMichael, A J (1999) Climate Change and Future Populations at Risk of Malaria, *Global Environ. Change*, 9(Suppl.), S89 107.

McMichael, A J (1997) Integrated Assessment of Potential Health Impact of Global Environmental Change: Prospects and Limitations, *Environ. Model. Assess.*, 2, 129 137.

McMichael, A J (1993) *Planetary Overload: Global Environmental Change and the Health of the Human Species*, Cambridge University Press, Cambridge.

McMichael, A J and Bouma, M (2000) Global Changes, Invasive Species and Human Health, in *The Impact of Global Changes on Invasive Species*, eds H Mooney and R Hobbs, Island Press, Washington, DC, 191 210.

McMichael, A J and Haines, A (1997) Global Climate Change: the Potential Effects on Health, *Br. Med. J.*, 315, 805 809.

McMichael, A J, Haines, A, Slooff, R, and Kovats, S (1996) *Climate Change and Human Health*, WHO/EHG/96.7, WHO, Geneva.

McMichael, A J, Patz, J, and Kovats, R S (1998) Impacts of Global Environmental Change on Future Health and Health Care in Tropical Countries, *Br. Med. Bull.*, 54, 475 488.

Myers, N (1997) Biodiversity s Genetic Library, in *Nature's Services. Societal Development on Natural Ecosystems*, ed G C Daily, Island Press, Washington, DC.

Newman, P and Kenworthy, J (1999) *Cities and Sustainability*, Island Press, Washington, DC.

Parry, M L and Rosenzweig, C (1993) Food Supply and the Risk of Hunger, *Lancet*, 342, 1345 1347.

Parry, M L, Rosenzweig, C, Iglesias, A, Fischer, G, and Livermore, M T J (1999) Climate Change and Global Food Security: a New Assessment, *Global Environ. Change*, 9(Suppl.), S51 68.

Patz, J A, Epstein, P R, Burke, T A, and Balbus, J M (1996) Global climate change and emerging infectious diseases, *JAMA*, 275, 217 223.

Rees, W (2000) Patch Disturbance, EcoFootprints, and Biological Integrity: Revisiting the Limits to Growth (Or Why Industrial Society is Inherently Unsustainable), in *Ecological Integrity, Integrating Environment Conservation, and Health*, eds L Westra, D Pimentel, and R Noss, Island Press, Washington, DC, 139 156.

Rooney, C, McMichael, A J, Kovats, R S, and Coleman, M (1998) Excess mortality in England and Wales during the 1995 heatwave, *J. Epidemiol. Community Health*, 52, 482 486.

Rosenzweig, C and Hillel, D (1998) *Climate Change and the Global Harvest: Potential Impacts of the Greenhouse Effects on Agriculture*, Oxford University Press, New York.

Serageldin, I (1999) Biotechnology and Food Security in the 21st Century, *Science*, 285, 387 389.

Sharpe, R M and Skakkebaek, N (1993) Are Oestrogens Involved in Falling Sperm Counts and Disorders of the Male Reproductive Tract? *Lancet*, 341, 1392 1395.

Soulé, M E (1991) Conservation: Tactics for a Constant Crisis, *Science*, 253, 744 750.

UN Environment Program (1998) *Environmental Effects of Ozone Depletion: 1998 Assessment*, Elsevier, Lausanne.

UN Environment Program (1999) *Global Environmental Outlook 2000*, Earthscan, London.

Watson, R T, Dixon, J A, Hamburg, S P, Janetos, A C, and Moss, R H (1998) *Protecting Our Planet. Securing our Future. Linkages Among Global Environmental Issues and Human Needs*, UNEP, USNASA, World Bank.

Wilson, M E (1995) Infectious Diseases: an Ecological Perspective, *Br. Med. J.*, 311, 1681 1684.

Woodward, A, Hales, S, and Weinstein, P (1998) Climate Change and Human Health in the Asia Pacific Region: Who will be the Most Vulnerable? *Clim. Res.*, 11, 31 38.

World Health Organization (1997) *Health and Environment in Sustainable Development*, WHO/EHG/97.8, WHO, Geneva.

World Health Organization (1998) *The World Health Report, Life in the 21st Century: a Vision for All*, WHO, Geneva.

World Resources Institute (1998) *1998- 1999 World Resources. A Guide to the Global Environment. Environmental Change and Human Health*, Oxford University Press, Oxford.

FURTHER READING

Bouma, M J and van der Kaay, H J (1996) The El Niòo Southern Oscillation and the Historic Malaria Epidemics on the Indian Subcontinent and Sri Lanka: an Early Warning System for Future Epidemics? *Trop. Med. Int. Health*, 1, 86 96.

Kalkstein, L (1993) Direct Impacts in Cities, *Lancet*, 342, 1397 1399.

World Health Organization (1999) *The World Health Report 1999: Making a Difference*, WHO, Geneva.

Environmental Changes Driven by Civil Conflict and War

IAN DOUGLAS

University of Manchester, Manchester, UK

When the human drivers of environmental changes are considered, the devastation caused by civil con ict and war seldom receives adequate attention. However, in many parts of the world legacies of past wars and the impacts of continuing current con icts deny access to agricultural land, poison aquatic food resources and destroy vegetation. The impacts of war upon the environment are not only localized to individual regions, but also have global effects. There is no more environmentally destructive human activity than warfare. Its purpose is to damage or destroy the natural and human environments. Scorched Earth is an expressive term for a particular military tactic, but it captures the essence of the consequences of much of modern warfare.

As the con icts in Kosovo and the Gulf demonstrated, human deaths and the destruction of military targets are not the only immediate consequences of war. Modern weapons rely on toxic chemicals for much of their explosive force and propulsion. They create environmental impacts through their own composition as well as their destructive power.

BOMBS AND LAND MINES

When a heavy bomb goes off, it creates temperatures of around 3000 °C that destroy the local flora and fauna and sterilize and damage the lower layers of soil much more effectively than forest fires. The targets destroyed by the bombs also contribute to the environmental devastation of war. The air raids on Pancevo and Sremicica in the suburbs of Belgrade destroyed a plastics factory and an ammonia production unit in petrochemical plants releasing toxins such as chlorine, ethylene dichloride and vinyl chloride monomer into the atmosphere. These chemicals not only have an immediate and life threatening effect on humans but also have a residual effect on the environment.

The Gulf War produced similar negative environmental impacts through the targeting of oil tankers and oil production facilities. According to the 1993 *Marine Pollution Bulletin*, 6 8 million barrels of crude oil were spilled from sunken vessels (including Iraqi tankers) and from oil transfer facilities such as the Kuwaiti Mina Al-Ahmadi Sea Island terminal and Iraqi Mina Al Bakr loading terminal between January 19 28, 1991. Reports suggest that approximately 30 000 marine birds perished as a result of this incident, and this figure excludes those that were trapped in oil pools in the desert. Furthermore, approximately 20% of mangroves were oiled, 50% of coral reefs affected, and hundreds of square kilometers of sea grass infected. The deliberate burning of Kuwaiti oil facilities released gaseous pollutants, smoke and particulates to the atmosphere (*see* Oil Fires: Kuwait, Volume 3).

War often results in huge movements of people out of conflict zones to areas thought to be safer. This exodus to safer areas places great strains on the resources and environment of those regions. Thus both the war zone and the refuge areas suffer environmental degradation.

Often people cannot return to their homelands because of the unexploded bombs (UXOs) and land mines left there. Land mines are cheap, devastating and increasingly popular in regional conflicts. The International Committee of the Red Cross estimates that they kill or maim between 1000 and 2000 people every month. One hundred and ten million land mines now lie in wait around the world with two million more added each year. Peacetime detonations of land mines cause most victims. The Save the Children Fund has repeatedly stated that children are especially vulnerable to anti-personnel mines. The injury or death of a parent directly affects their welfare; their small size and curiosity increase risk of injury; and childhood occupations such as herding expose them to greater danger. Such injury and loss of family members reduces families ability to manage the land, to protect the soil against erosion and maintain food production.

Between 10 and 20 million land mines prevent people in Angola from making progress with recovery and rebuilding

in many parts of the country. The United Nations (UN) estimates that these silent killers have created 700 000 amputees in Angola. For three decades mines were scattered in Angola s elds, villages, roads, and other unexpected places to intimidate, maim and kill innocent victims. Land mines have a devastating effect upon the environment by restricting the movement of people, deterring farming, disrupting economies, and killing and mutilating many innocent men, women, and children.

It could be argued that by preventing access to elds and forests, land mines stop people altering the soil, cutting down trees, extracting minerals, or dumping chemicals. However, by their very nature, land mines are a people-made pollutant that adversely alters the environment for future generations. For example, in Angola thousands of kilometers of riverbanks, and tens of thousands of hectares of farmland, pastures, and forest are now unusable. In addition, the migration of people from the mined areas to towns and cities has led to other tracts of country being stripped of wood and wild game while water supplies have been depleted and contaminated, leading to increases in reported cases of dysentery, malaria and cholera. All the pressures on the land as refugees try to survive may lead to severe soil degradation and land cover change from forest and woodland to bare or sparsely vegetated soil.

Land mines, inexpensive to produce or buy and easy to distribute, are extremely dif cult to detect and costly to remove. In some places, mines seem to multiply faster than people. Cambodia and the Lao Peoples Democratic Republic (Laos) demonstrate their lasting environmental impacts particularly well. As a 1994 UN report stated, Cambodia has more mines than children; two for every child . After the Paris peace agreements of October 1991, which followed two decades of con ict, Cambodia had to cope with 385 000 returnees from Thailand, 190 000 internally displaced persons and 40 000 people maimed by land mines. Security problems persisted, with more than 5 million land mines still in place. After 30 years of con ict Cambodia is among the most mine- and UXO-affected countries in the world. In 1998, seven years after the 1991 peace agreement, mines and UXOs caused more than 1200 accidents. Most of the mined areas are in the provinces along the Thai Cambodia border where much ghting occurred after 1979. The Eastern provinces are mostly affected by UXOs (though there are also some mined areas) as a result of the Vietnam War.

In 1998 there were estimated to be over 14 000 km^2 of land affected by mines with a further 5300 km^2 suspected to contain land mines. Land mines affect the rural communities living along the Thai Cambodian border in various ways. There is a shortage of land for settlement, for agriculture, it is dif cult to rehabilitate rural infrastructure (schools, road, irrigation systems), there are obstacles to safe travel and to forest gathering activities to earn an income and there is no security for children.

The bomb and land mine legacy of the Indochina wars persists in Laos. Without a formal declaration of war, Laos became one of the most heavily bombed countries of all time. Between 1964 and 1973, the US dropped on Laos the equivalent of one B 52 payload of ordnance every eight minutes for nine years. Nearly 30 years after the cessation of hostilities, the debris of this bombardment continues to deny land to, and in ict casualties on, the largely rural population of Laos.

The physical and social effects of war still play an important role in current problems of environment and resource management in the Nam Ngum watershed north of Vientiane, the capital of Laos. War damage was especially heavy on the Plain of Jars in Xieng Khouang province, but other areas also suffered destruction of forests and villages. Aerial bombardment destroyed agricultural lands and forestry continues to be affected by shrapnel fragments lodged in trees. The northeastern part of the watershed in particular is still affected by bomblets from cluster bombs that still lie unexploded in the soil. No inventory of damage has been made to date, but remaining live munitions continue to cause injury or death to local people working their elds. Large areas remain uncultivable.

The wartime disruptions led many people to move outside the Nam Ngum watershed area, either to lowland areas closer to Vientiane, or to revolutionary strongholds close to the border with Vietnam. This in turn led to further forest clearance for cultivation, but following the war it also prompted renewed environmental destruction within the watershed area, because the returnees rice elds had been destroyed, as had their buffalo and farm implements. Moreover live munitions in the soil made plowing dangerous. For many years after 1975, former sedentary farmers relied on shifting cultivation, often in those areas of forest not badly affected by the war. Now hillsides have been converted from forest to grassland by shifting cultivation operating too intensively and lack of attention to re escaping from cleared areas to adjacent trees and shrubs. The legacy of war is severe land degradation, forest depletion and insuf cient cultivable land for wet rice production.

CHEMICAL DEFOLIANTS IN VIETNAM

Between 1962 and 1971, Operation Ranch Hand sprayed about 72 Megalitres (Ml) (19 million gallons) of herbicide. Forty-one Ml of this total were Agent Orange. The spray fell mostly on the forests of South Vietnam, but some was used in Laos, and some killed crops to deprive Viet Cong and North Vietnamese troops of food. The military purpose for using herbicides on non-cropland was to remove the vegetation cover used by Viet Cong and North Vietnamese forces for concealment. Along roads, canals, railroads,

and other transportation arteries, Ranch Hand cleared a swath several hundred meters wide to make ambushes more difficult. In Laos, the herbicide removed the jungle canopy from the network of roads and trails used for infiltrating men and supplies, making them more vulnerable to attack from the air. Ranch Hand also cleared large areas of forest that hid sanctuaries and bases, thereby forcing the North Vietnamese and Viet Cong to move or risk discovery and attack. In all, Ranch Hand planes sprayed herbicide over about 25 000 km^2, not correcting for multiple coverage.

The herbicides Ranch Hand sprayed were common agricultural chemicals in wide use in the US and other countries at that time. The most common ingredients in the herbicide mixtures were 2,4 D and 2,4,5 T, phenoxy herbicides that act as growth regulators and cause destructive proliferation of tissues in plants when they are in a stage of active growth. Another plant growth regulator used was picloram. Cacodylic acid, an organic arsenic compound, killed crops by causing them to dry out.

Crop destruction struck at the very heart of a rural South Vietnamese farmer s existence, eliminating not only the food supply upon which he and his family depended, but also obliterating in one spray pass the product of many months of his family s labor. The herbicides main effect on trees was to kill their leaves, and there was usually little lasting damage in future growing seasons unless the trees had been sprayed three or more times. Only about 12% of the total area covered by Ranch Hand received triple coverage. Mangrove areas in South Vietnam were an exception, because mangroves were killed by just one dose of spray due to their high sensitivity to herbicides. About 36% of the mangrove forest area in South Vietnam was destroyed and will not return to its natural state for perhaps a century without extensive reseeding.

The herbicides are thought not to have had any lasting effects on nutrients in the soil, with the possible exception of potassium. The more conventional wartime bombing and shelling had a worse effect on inland forests than herbicides. Besides killing trees, shrapnel imbedded in wood made logging both costly, and hazardous. The removal of trees by displaced people in the period immediately following the war probably caused more biomass loss than the defoliants. The collective impact of all these war-driven events has been to deprive much of Vietnam of mature forests. In 2000, few of the forest trees in State Forests in Vietnam were much more than 25 years old.

ENVIRONMENTAL EFFECTS OF THE KOSOVO CONFLICT

The European Commission s preliminary findings on the war in Yugoslavia (The Regional Environmental Center for Central and Eastern Europe, 1999) show a comprehensive range of environmental impacts. Surface waters suffered largely as a result of leakage from damaged industrial plants or pollution from poorly planned refugee centers. Specific impacts include the release of polychlorinated biphenyls (PCBs) from damaged transformer stations; leakage of oil products into the Danube River from the Pancevo industrial center and the refinery at Novi Sad; escape of more than 100 tones of ammonia into the Danube and release of more than 1000 tones of ethylene dichloride spilled from the Pancevo petrochemical complex into the Danube.

Air pollution impacts are thought to have been short-term phenomena. They include radioactive pollution from depleted uranium weapons. Vinyl chloride monomers (VCMs) reached concentrations of 10 600 times more than permitted levels near the Pancevo petrochemical plant. Polluted clouds carried the products of combusted VCMs: phosgene, chlorine, chlorine oxides and nitrogen oxides, Products from incomplete hydrocarbon combustion were released as a result of strikes on oil refineries. Following the Pancevo incidents, a 15 km long cloud of smoke lasted for 10 days. Concentrations of soot, sulfur dioxide (SO_2) and chlorocarbons increased by 4 8 times the allowable limits. Nitrogen oxides were released from jet aircraft and through burning industrial installations. Unusually acid rain was measured in several places in Romania and Bulgaria, the timing of which and the prevailing wind trajectories tend to link the acid rain with attacks on Yugoslav industries.

Much of the air and water pollution will eventually settle into the surrounding soil. Bombing has created deep craters in the humus layer, which will take years to recuperate. A 240 kg bomb makes a 50 m^2 crater. Not only are the craters unusable, but so is some land around the craters. The destruction of the upper layers of the soil means the destruction of its flora and fauna and degradation of habitats. Five national parks were directly affected by the conflict. In Albania some protected areas had refugee camps built within them (Rrushkull, Divjaka) and refugees changed land cover and aquatic ecosystems in nearby areas by felling trees and polluting water sources.

CIVIL CONFLICT AND FOREST RESOURCES

Local civil conflicts and insurrections are often more insidious than international wars. In addition to claiming 60 000 lives, the lengthy ethnic conflict in Sri Lanka has devastated the rich biodiversity of this island nation. Heavy explosives have destroyed natural vegetation, trees and parks in wilderness and urban areas, agricultural land, domestic and wild animals and birds. Since 1990, the separatist Liberation Tigers of Tamil Eelam and the government troops have cut down more than 2.5 million palmyrah palm trees (*Borrasus abellifera*), a staple of the people s diet in the north and east of the country. Palmyrah trees act as wind breaks, protecting the villages from damage by the strong

winds affecting this part of the country. Absence of forest wardens during the conflict has meant that much illegal logging has occurred.

Lack of control of land and biotic resources management during conflicts leads to uncontrolled exploitation, often with the connivance of the warring factions. Whether it is the illegal cutting and export of logs, the smuggling of diamonds or the uncontrolled alluvial mining of gemstones, the result is resource exploitation without any thought of the future. No replanting of trees, no restoration of mined areas and no control of erosion and sedimentation occur.

NUCLEAR WINTER

Despite the real situations described above the scenario of the likely environmental consequences of a major nuclear war would be truly global (Pittock *et al.*, 1989; Harwell and Hutchinson, 1989). A nuclear explosion at or near the surface of the Earth causes an injection of dust into the atmosphere. The fires ignited by the explosion also cause soot to be injected into the atmosphere. Models of the long-term behavior of this dust and smoke, based on those used for the injection of soot and dust from volcanic eruptions, generally assume that the nuclear explosions will occur in the Northern Hemisphere. The reduction of light levels due to the layers of dust and smoke causes severe drops in temperature at the surface of the Earth, this is termed a *Nuclear Winter* (*see* Nuclear Winter, Volume 3).

Although 10 25% of the soot may be washed out of the atmosphere within two weeks, modeling and satellite imagery suggest that the remaining smoke can be transported halfway round the world within two weeks. For example, the Chinese wildfires in May 1987 reduced daytime temperatures over Alaska by 2 6 °C later that month. In the first 1 3 months, the modeling predicts that nuclear war-created dust and smoke reduce solar heating of the land surface, without sufficient compensating infrared heat transfer from the atmosphere. The models predict that a large July smoke injection would produce a 22 °C drop in temperature in mid-latitudes. In humid climates the temperature drops are more moderate at about 10 °C. Over mid-latitudes there would be a 75% decrease in rainfall. In winter, temperature drops of only a few degrees are predicted. However, because the land would be colder to start with, more intense frosts and freezing events would be likely. Light level reductions would range from 0%, at low latitudes, to a 90% reduction in areas of high smoke injection. Ocean surface temperatures could fall by between 2 6 °C. The smoke and oxides of nitrogen injected could produce a long-term ozone depletion of about 50% or more. As a result, the ultraviolet radiation could increase up to 200% of normal levels. These changes would cause immense disruptions of ecosystems and major problems from human health, food supplies and wellbeing.

CONCLUSION

War and civil conflict lead to environmental scarcity, to forced migrations of people, land cover degradation and loss of biodiversity. Increasingly they have impacts on ecosystem biogeochemistry and ultimately the chemical residues of weapons and the damage to chemical processing plants will feed through the life forms and sediments in wetlands, lakes and coastal seas. The extent of war and conflict-induced land cover transformation cannot easily be segregated from vegetation removal through other processes. However, these impacts, of immense significance to the people and ecosystems immediately affected, contribute to the global consequences of changing land surface characteristics.

While war and civil conflict indirectly affect global environmental change, climatic changes will themselves be drivers of future conflicts. Sea level rises for example will induce migrations into other areas (*see* Environmental Refugees, Volume 4). Repeated droughts since 1960 have already been a major factor in causing civil conflicts and wars in the Horn of Africa, as people move from dry environments into better-watered territory already exploited by others. These environmentally related conflicts are not simple responses to climatic variations or changes. They are reflections of political events, changes in food production and marketing systems, traditional tribal land occupation and use, as well as deteriorating natural conditions (Suliman, 1999). Such complications should not be permitted to obscure the global significance of the environmental impacts of war and civil conflict, nor allow the sources of the weapons that make the destruction so pervasive to be forgotten.

REFERENCES

Harwell, M C and Hutchinson, T C (1989) *Environmental Consequences of Nuclear War Vol. II Ecological and Agricultural Effects*, 2nd edition, John Wiley & Sons, Chichester.

Pittock, A B, Ackerman, T P, Crutzen, P J, MacCracken, M C, Shapiro, C S, and Turco, R P (1989) *Environmental Consequences of Nuclear War Vol. I Physical and Atmospheric Effects*, 2nd edition, John Wiley & Sons, Chichester.

Suliman, M (1999) *Ecology, Politics and Violent Con ict*, Zed Books, London.

The Regional Environmental Center for Central and Eastern Europe (1999) *Assessment of the Environmental Impact of Military Activities During the Yugoslavia Con ict Preliminary Findings, June 1999*, European Commission DG XI; Environment, Nuclear Safety and Civil Protection, Brussels, http://www.rec.org/Rec/Announcements/yugo/contents.html.

A

Acid Sulfate Soils

Nazeer Ahmad
The University of the West Indies, St. Augustine, Trinidad and Tobago

Acid sulfate soils are problem soils which are worldwide in distribution, and which develop in marine and coastal inorganic and organic sediments in which iron sul de (pyrite) has accumulated over time. The accumulated plant biomass is also rich in organic sulfur, which on oxidation, adds to the pool of acid forming sulfate. On draining and oxidation, the black-colored pyrite (FeS_2) is oxidized to straw-colored jarosite ($Fe_3(SO_4)_2(OH)_6$). This reacts with soil water forming sulfuric acid and iron oxide, mainly goethite (FeO.OH) and in some instances, haematite (Fe_2O_3). While these processes are taking place, the soil develops extreme acidity (<pH 3.5) due to the formation of sulfuric acid, and the layer in the pro le thus affected is termed a sulfuric horizon. In this condition the soils are known as actual acid sulfate soils.

Potential acid sulfate soils are those rich in pyrite as well as sulfur in the organic residues of saline or brackish swamp vegetation which have accumulated in the soil body but which remained anaerobic due to continuous water-logging in coastal, estuarine, deltaic and riverain locations. The soil layer in which these accumulations occur is known as a sul- dic horizon. The soils have the capacity to become highly acid once oxidation is allowed to occur. The term acid sulfate soils was rst used by Chenery (1954) to describe these soils; although their extreme properties were known before, the soils were called by other names.

SOIL GENESIS

The environmental conditions for the deposition of pyrite and sulfur-rich soil materials (sulfidic materials) are normally met in shallow, saline or brackish swamps and marshes and in estuarine and deltaic locations. Even under the most favorable natural conditions, the rate of accumulation of pyrite is slow, around $10\,kg\,m^{-3}$ of sediment per 100 years in mineral soils (Dent, 1986). In these locations, mangrove (*Rhizophora* spp.), *Avicennia* spp. and other salt tolerant aquatic plants grow. The vegetation has three main effects in the evolutionary process of these soils:

1. acts as wind and water break of the coastal water, thereby encouraging mineral sedimentation of suspended material;
2. provides a source of energy for bacteria, mainly *Thiobacillus ferrooxidans*, which reduce the sulfate to sulfide in the transported sediments and which lead to the precipitation of the sulfide as FeS_2 (pyrite). This precipitate accumulates in the soils and results in a progressive development of a bluish-gray to black color of the soil material. The contributory conditions are a high content of easily decomposable organic matter and abundant soluble sulfate;
3. accumulates sulfate and organic sulfur compounds in the organic residues.

In equilibrium conditions, when pressure is applied on the surface of these inundated soils, gases escape, and the characteristic odor of H_2S is discernible. This odor becomes more distinguishable when the actual soft soil material is smelt and it becomes stronger farther down in the soil profile. Because of this unpleasant odor, these soils have long been known as *cat clay soils* by Dutch soil scientists since they claimed the smell was reminiscent of the excrement of cats.

Under natural conditions, the land would be inundated and the soil material quite soft. Ideally, the soil consists of an intimate admixture of mineral and organic residues. Not uncommonly, there may be layers of well-preserved organic materials alternating with layers of more mineral soil. These organic layers indicate deposition periods when the floodwater was poor in suspended sediments so the bulk of the deposited material originated from the vegetative cover. The entire profile can be predominantly organic for the same reason.

The concept of *soil ripening* is important in the development of these soils from freshly deposited mud, and embraces the physical, chemical and biological processes by which this material is transferred to a dry land soil.

Physical ripening involves an irreversible loss of water and consequent hardening of the soil. Clay and organic-rich sediments deposited in water have a very open structure and consequently a very high water-filled pore space. In this condition, the soil structure is easily deformed and the material has little friction or cohesive strength.

During accretion in a tidal environment, water is lost as a result of consolidation, evaporation and drainage at low tide and by extraction of water by plant roots. Removal of water leads to the partial collapse of the initial very open microstructure, and this is the soil ripening process. The degree of soil ripening can be assessed in the field by squeezing the soil in the hand and observing its behavior. (Smits *et al.*, 1962; Pons and Zonneveld, 1965).

Totally unripe mud is fluid; in predominantly mineral sediments, the water content is more than 80% by weight. Practically unripe mud is very soft, sticks fast to any surface and can be squeezed through the fingers by gentle pressure the water content can be between 70 and 80%. Half-ripe mud is fairly soft and sticky, and can be squeezed through the fingers the water content can be between 65 and 75%. Nearly ripe material is fairly firm; it tends to stick to the hands and can be kneaded and may have moisture content of between 55 and 65%. Ripe material is firm, not particularly sticky and cannot be squeezed through the fingers. The ripening process can be hastened with reclamation and drainage.

In the natural state, the soil is unripe, bluish gray in color and may be essentially mineral, organic or a combination of both. The pH value of the mineral soil in the fresh state is only slightly acidic. This is influenced by the solution of H_2S in the soil environment, by a soil cation exchange capacity that is saturated with Fe^{2+}, which is usually plentiful in the mineral sediments, and by other cations present in sea water. In this state, they are known as *potential acid sulfate soils*. It is only on drainage and oxidation, when the pyrite gets oxidized and mineralization of the organic sulfur progresses, that the acidity develops and the soils can then be described as *actual acid sulfate soils*.

Not all coastal, deltaic and estuarine soils are potential or actual acid sulfate soils. The potential for a soil deposited in a marine environment to so develop is a function of the rate of deposition of inorganic sediments. If this is very rapid, if the level of coastal inundation is too deep to allow vegetation to grow, or if the neutralizing capacity of the sediments is considerable, clay soils result with little pyrite material, or if present, the acidity developed on oxidation is neutralized; such material has little or no capacity to acidify on drainage.

The coastal soils from two geographical areas, i.e., the northern coast of South America and the coast of West Africa are strongly contrasting in this regard. The ocean water on the northern coast of South America is rich in mineral sediments of Amazonian origin and provides an environment for rapid mineral sedimentation. Consequently, good clay soils (Entisols, Inceptisols, Vertisols and Intergrades) predominate. By comparison, the marine-influenced soils on the coast of West Africa were deposited in an environment poorer in mineral sediments and acid; consequently acid sulfate soils are dominant.

SOIL CLASSIFICATION

According to *Soil Taxonomy* (Soil Survey Staff, 1975) the potential acid sulfate soils are classified as Sulfihemists in the Histosols if they are dominantly organic, and Sulfaquents in the Entisols if the soils are mineral and have pyrite material within 50 cm of the soil surface. Actual acid sulfate soils are classified as Sulfohemists in the Histosols if they are dominantly organic or as Sulfaquepts in the Inceptisols if they are mineral soils.

OCCURRENCE

On a world scale, acid sulfate soils are not extensive. In recent coastal plains and inter-tidal swamps, there are an estimated 12 million ha, mostly in the tropics, in which the topsoil will become severely acid or has already done so as a result of land reclamation. There is probably a much greater area of potentially acid sulfate soil material covered by a shallow layer of non-acid peat or alluvium (Dent, 1986). Large areas occur in the Far East in Vietnam, Thailand, Malaysia, Burma, Indonesia and Philippines, in West Africa (Senegal, Gambia, and Guinea Bissau (Marius, 1982) and Sierra Leone) (Bloomfield and Coulter, 1973), the Caribbean and South America (Suriname, Guyana, and Orinoco Delta of Venezuela and Trinidad) (Ahmad and Wilson, 1992). They also occur in temperate climates, i.e., New Zealand (Metson *et al.*, 1977), Australia, Canada, and in eastern United States and in Europe Sweden (Wiklander *et al.*, 1950), Denmark (Larsen, 1978), Finland (Kivinen, 1950), Holland and England (Dent and Turner, 1981). Changes in sea levels over the past 7000 years are related to the distribution patterns of these soils. In the Chao Phraya Delta in Thailand, for example, older sulfidic sediments of the Bangkok Plain are now 1 2 m above sea level and represent an earlier phase of deposition. They are succeeded, seawards, by a belt, about 100 km wide, of more recent marine alluvium that is mostly not potentially acid. Similar effects are present in the Mekong Delta in Vietnam (Dent, 1986).

IDENTIFICATION OF ACID SULFATE AND POTENTIAL ACID SULFATE SOILS IN THE FIELD

Due to the problem of excessive acidity when drained, it is very important to identify the soils in their natural state so

that adequate precautions can be taken if their reclamation and use are to be attempted. Such associated morphological features as soil color, soil consistency, presence of admixed organic material, jarosite mottles, root encrustations and the disagreeable odor of H_2S are always present. Where there has been some oxidation and development of acidity, the water in channels is clear and sediment-free and devoid of any fish life; these water courses are usually free of aquatic weeds but may have colonies of molds, iron bacteria and ochre (Trafford *et al.*, 1973). The water can cause skin irritation due to the content of dissolved acid. However, for positive identification, the pH value of the soil should be measured in the field, then a sample should be taken and treated with hydrogen peroxide or incubated for three months and the pH measured again. The values should decrease from near neutrality in the field to less than 4 in the incubated sample or to less than 2.5 after peroxidation (Soil Survey Staff, 1975; Brinkman and Pons, 1973).

It is obvious that in agricultural use, if these levels of soil acidity are allowed to develop on reclamation, they would cause Al^{3+} toxicity and plant nutrient deficiencies which would be important soil management problems (Ahmad, 1961). In the Caribbean, the Mara, Godineau and Alta Gracia soils (Table 1) are potential acid suppurate soils with an overlay of non-sulfate soil material. The Godineau soil has a buried sulfate peat layer deeper than 95 cm. By comparison, in the San Carlos soil, sulfuric conditions develop from the surface on drying. Note the high soluble sulfate contents throughout the soil profiles, which are associated with the low pH values. The other soil properties show the soils to be typical marine clays with high chemical activity.

USE AND MANAGEMENT

In many parts of the world, these problem soils occur in centers of high population density, so there is great urge to utilize them. In the process, there are many examples in reclamation where the land was over-drained, causing too rapid oxidation of the sulfidic materials, both organic and inorganic, with development of extreme acidity and the project had to be abandoned. In West African countries, the damming of rivers which prevent tidal influences in the flood plains resulted in over-draining of the land, and a transformation of some of the soils from potential acid sulfate (sulfidic) to actual acid sulfate in nature. The same process occurred in the Delta of the Orinoco in Venezuela and elsewhere. This process is not easily reversible (Ahmad, 1991).

Successful reclamation depends greatly on a full understanding of the properties of the soil. The main features to assess are existing or potential acidity, soil lime requirement, soil salinity, soil texture, degree of ripeness, and availability of good quality water for leaching. If acid sulfate conditions exist from the soil surface for instance, it may be advisable to leave the area as a nature reserve rather than attempt reclamation. Those soils that have an overlay of pyrite-free material do offer a much better prospect for successful reclamation. The main principles in attempting to utilize them are that oxidation of the pyrite and organic sulfur material should be controlled by careful water management. The main drainage system should be able to maintain the water table in the reclamation area at a sufficient depth below the surface to permit the leaching of salts and enough soil ripening of the topsoil to support agricultural operations. The field drainage should facilitate leaching of salts and create an aerated rooting zone for dry-land crops and a topsoil firm enough to support farm animals and light machinery. However, at the same time, the sub-soil-rich sulfidic layers must be kept anaerobic.

Continuous rice cultivation, in which the land is kept inundated for most of the year and the water drained off temporarily prior to harvesting of the particular rice crop, is one of the most sustainable uses for these soils. In this case, all operations have to be manual or utilizing only hand-operated light machinery. This use is common in Thailand and Vietnam and other Far East countries. In a few West African countries, where the crop is grown in rain-fed conditions, the land is formed manually in ridges or narrow beds, in which the relatively non-pyritic surface soil is concentrated for the growth of the crop.

In the Far East, advantage is taken of concentrating the non-sulfidic surface layer in raised beds that are then used for the cultivation of crops including vegetables, pineapple, food crops and even perennial crops. Shallow, wide drains of about 2 3 m in width are dug and the material spread on the intervening beds. The width and depth of the drains and the height of the beds are dependent on the crop to be grown and by the depth in the soil at which the potential acid sulfate condition is serious. The drains, in addition to providing the material to build the raised beds, serve to control the level of the water table. The depth of water in the field drains is maintained by a gate, where the drain discharges into the main drain; by opening or closing this gate, the level of water in the drain is maintained; there is also a similar device at the intake end of the drain to allow fresh water to enter. Good water must be available to periodically replace the water in the drains, which can become very acid. With this layout and these precautions, the soils can be used sustainably for production of a wide range of crops.

The other approach to utilization of adjusting the soil pH by use of ground limestone is not practical. Experiments have been conducted in different parts of the world which showed that up to 100 tons ha^{-1} of ground limestone may

Table 1 Properties of representative acid sulfate soils in the Caribbean region (after Ahmad and Wilson, 1992)

Sample depth		Particle size (%)			100 g soil[a]					Base	%		SO_4
(cm)	pH	Sand	Silt	Clay	CEC	Ca	Mg	K	Na	sat. %	C	N	(ppm)
Mara series, Guyana													
8–0	4.4	Organic			41.2	0.4	1.1	0.3	0.2	5	9.8	0.58	420
0–3	4.4	Organic			29.2	0.5	2.4	0.2	0.1	11	7.2	0.29	360
3–10	4.0	2	45	53	28.4	0.5	3.4	0.2	0.1	15	2.1	0.03	360
10–28	4.0	1	46	53	27.0	0.5	4.0	0.3	0.2	18	0.8	0.24	240
28–82	3.4	10	40	50	26.5	0.5	3.2	0.2	0.4	16	0.5	0.05	300
82–101	3.1	1	46	53	30.5	0.5	3.6	0.2	0.4	15	1.4	0.09	720
101–152	3.5	1	48	51	31.1	0.5	2.3	0.4	0.3	11	2.6	0.06	1440
Godineau clay, Trinidad													
0–10	4.0	1	8	92	34.8	13.6	15.8	0.86	3.34	76	7.9	0.12	866
10–25	4.0		6	96	33.4	6.5	8.0	0.42	2.37	37	5.0	0.17	860
25–55	3.9		12	89	37.2	8.6	9.5	0.60	3.17	35	4.2	0.35	1085
55–95	3.7	2	12	78	23.6	5.3	5.8	0.43	1.65	53	2.0		1186
>95	2.3	Buried peat									Peat		
San Carlos Holding, Isla Manamito, Orinoco Delta, Venezuela													
0–30	3.6	15	20	48	27.9	0.9	1.0	0.49	0.60	9	7.2	0.68	1188
30–45	3.3	12	20	64	19.5	1.0	1.1	0.13	0.34	12			790
45–60	3.0	24	16	60	20.5	0.3	0.9	0.12	0.34	7			950
60–135	2.4	43	25	32	15.7	0.4	0.7	0.04	0.17	7			2160
135–180	3.0	27	12	52	12.8	0.6	1.2	0.09	0.20	15			1180
Near Alta Gracia, Isla Manamito, Orinoco Delta, Venezuela													
0–15	4.2	10	13	57	29.3	9.8	4.2	0.33	0.61	50	8.0	0.58	633
15–30	4.1	6	12	68	39.3	3.9	3.7	0.25	0.18	20	3.2	0.26	435
30–60	3.8	13	17	53	33.4	8.5	4.9	0.37	0.50	43			792
60–90	3.0	9	17	62	23.7	1.1	3.6	0.20	0.13	21			1267
90–135	2.4	19	25	49	26.0	0.6	2.5	0.06	0.04	12			2400
135–150	2.1		55	11	25.0		3.7	0.08	0.22	16			3300

[a] m.e., milliequivalents; CEC, cation exchange capacity.

be needed to appreciably increase the soil pH and reduce the level of exchangeable Al^{3+}. This requirement would only cater for the current acidity and not for the reserve acidity represented by the pool of sulfur present in organic form, or the unoxidized pyrite. Therefore, these high dressings of ground limestone may well have to be continued for some time. Phosphate fertilization has always proved to be very bene cial in association with lime, probably counteracting, to some extent, the high exchangeable Al^{+++}. Other fertilizers may also be necessary to achieve good production and this should be determined by experimentation.

Successive ooding with dilute seawater and subsequent leaching with fresh water as a means of reclaiming acid sulfate soils was tried in Guyana for sugar cane cultivation (Evans and Cate, 1962) in Sierra Leone for rice cultivation (Hart *et al.*, 1963) and in Denmark (Larsen, 1978). The treatment resulted in some change of the cation and anion population of the soil and related increases in soil pH in Guyana, but its wide use has severe restrictions, such as the ability to control the water so that other non-problem soils are not ooded in the process, and the availability of adequate amounts of good water for leaching.

Fish and prawn culture has been successfully carried out in countries such as Vietnam and the Philippines on acid sulfate soils. Without specific treatment, the yields are generally poor; however, with treatment consisting of successive flooding and draining after excavation of a new pond until the pH of the flood water in the pond remains stable at 5 or higher, and then incorporating lime, phosphate and chicken manure to the soil material and finally flooding and stocking, yields can greatly increase (Brinkman and Singh, 1982).

OTHER PROBLEMS

Environmental Impact

Where acid sulfate soils are developed, dramatic changes in the chemical properties of the associated water take place and these are exported from the reclaimed area with the drainage waters. Acid floodwaters generated over large areas of acid sulfate soils may adversely affect crops cultivated on adjacent better soils (Brinkman, 1982). Also, drainage from acid sulfate soils may affect aquatic plants, shellfish, crustacea and fish in both tidal and inland waters. Ferric hydroxide is deposited on the sides of streams and drainage channels and can accumulate in such quantities as to partially block them (Trafford *et al.*, 1973). Also, reclamation of these natural wetlands would result in loss of wild-life habitat and recreational facilities.

Engineering Problems

The chemical properties of potential acid sulfate or actual acid sulfate soils are detrimental to regular construction materials such as cement and steel. Several approaches are possible to counteract these problems. Steel can be insulated with bituminous material or covered with a layer of concrete. Protection can also be achieved by mixing lime with the soil layer surrounding the metal or concrete. In the case of concrete structures, sulfate-resistant cement and acid-resistant masonry or hardwood can be used (Dent, 1986). Any urban and industrial development on acid sulfate soils faces the engineering problem of uneven subsidence of unripe soils and corrosion of structural materials if adequate precautions are not taken.

See also: **Intertidal Zones**, Volume 2; **Mangrove Ecosystems**, Volume 2; **Contaminated Lands and Sediments: Chemical Time Bombs?**, Volume 3; **Marshes, Anthropogenic Changes**, Volume 3.

REFERENCES

Ahmad, N (1961) Aluminum Toxicity of Certain Soils on the Coastal Plain of British Guiana and Problems of their Agricultural Utilization, *Trans. VIIth International Congress Soil Science*, **2**, 161 170.

Ahmad, N and Wilson, H W (1992) Caribbean Acid Sulfate Soils, *Soil Sci.*, **153**, 154 164.

Bloomfield, C and Coulter, J K (1973) Genesis and Management of Acid Sulfate Soils, *Adv. Agron.*, **25**, 265 326.

Brinkman, R (1982) Social and Economic Aspects of the Reclamation of Acid Sulfate Soil Areas, in *Proceedings of the Bangkok Symposium on Acid Sulfate Soils*, eds H Dost and N van Breemen, Publication 31, ILRI, Wageningen, 21 36.

Brinkman, R and Pons, L J (1973) Recognition and Prediction of Acid Sulfate Soil Conditions, in *Acid Sulfate Soils. Proceedings of the First International Symposium*, ed H Dost, Vol. 1 and 11, Publication 18, ILRI, Wageningen, 169 203.

Brinkman, R and Singh, V P (1982) Rapid Reclamation of Brackish Water Fishponds in Acid Sulfate Soils, in *Proceedings of the Bangkok Symposium on Acid Sulfate Soils*, eds H Dost and N van Breemen, Publication 31, IRLI, Wageningen, 318 330.

Chenery, E M (1954) Acid Sulfate Soils in Central Africa, in Trans 5th International Congress, ISSS Leopold ville, **4**, 195 198.

Dent, D (1986) *Acid Sulfate Soils, a base line for Research and Development*, Publication 39, ILRI, Wageningen.

Dent, D L and Turner, R K (1981) Acid Sulphate Soils in Broadland Part II: Economic Evaluation BARS 3 Broads Authority, Norwich, England.

Evans, H and Cate, R B (1962) Studies in the Improvement of a Problem Soil in British Guiana, *World Crops*, **14**, 270 373.

Hart, M G R, Carpenter, A J, and Jeffrey, J W O (1963) *Problems in Reclaiming Saline Mangrove Soils in Sierra Leone*, Agron. Trop. (Paris), 800 802.

Kivinen, E (1950) *Sulfate Soils and their Management in Finland*, in Fourth International Congress Soil Science, Amsterdam, **11**, 259 261.

Larsen, V (1978) *Reclamation of Acid Sulfate Soils in Denmark*, in 11th Congress ISSS Edmonton, **1**, 364.

Marius, C (1982) Mangrove Soils of Senegal and Gambia, in *Proceedings of the Bangkok Symposium on Acid Sulfate Soils*, eds H Dost and N van Breemen, Publication 31, ILRI, Wageningen, 103 136.

Metson, A J, Gibson, E J, and Cox, J E (1977) The Problem of Acid Sulfate Soils with Examples from North Auckland, New Zealand, *N. Z. J. Sci.*, **20**, 371 394.

Pons, L J and Zonneveld, J S (1965) *Soil Ripening and Soil Classi cation*, Publication 13, ILRI, Wageningen.

Soil Survey Staff (1975) *Soil Taxonomy*, USDA Handbook 436, Washington, DC.

Smits, A J, van Schreven, D A, and Bosma, W A (1962) *Physical, Chemical and Microbiological Ripening of Soils in the Ijsselmeer Polder* (in Dutch), Van Zetot Land, 32, Zwolle.

Trafford, B P, Bloomfield, C, Kelso, W I, and Pruden, G (1973) Ochre Formation in Field Drains in Pyretic Soils, *J. Soil Sci.*, **24**, 453 460.

Wiklander, L G, Halgren, G, Brink, N, and Jonsson, E (1950) Studies in Gyttja Soils. Some Characteristics of Two Profiles from Northern Sweden, *Ann. Royal Ag. Coll. Sweden*, **17**, 24 36.

Afforestation: Environmental Impacts

Piers Maclaren
Piers Maclaren and Associates, Rangiora, New Zealand

Forests, even if they consist of monocultural plantations, have a number of desirable effects on the environment. There are also a few negative impacts, but (depending on the circumstances) these are usually outweighed by the bene ts. Compared to a vegetation cover of grass or annual crops, forests usually provide greater protection against soil degradation, including soil erosion. They minimize the output of greenhouse gases and, indeed, can absorb large quantities of carbon dioxide during their establishment phase. They maintain water quality and reduce ooding in the headwaters of river systems during small storms. They have greater biodiversity. Their major negative effect is to reduce the annual total water yield: river ow is less, at all times of the year, if the catchment is in tall forest as opposed to short vegetation.

For the purposes of this article, *afforestation* is defined as planting of new forests on lands, which historically, have not contained forests. Given that trees are generally considered to be a climax vegetation (where soils and climate allow), why did some areas have no trees in the historic past? Regeneration of trees may have been prevented by human activities leading to frequent burning, livestock grazing, cultivation, soil nutrient depletion or waterlogging of soil. These practices can be reversed by such means as firebreaks, fences, fertilizer application and drainage. Climatic influences such as frost, or rainfall, can also explain the absence of trees in certain regions, but this can be overcome by planting hardy species and by irrigation.

Afforestation falls into two broad types: land abandonment and the deliberate planting of trees, usually for commercial purposes. Abandonment (called *old eld succession* in the United States) is a slow and haphazard process, which relies on the presence of seed sources from nearby forest remnants. Local inhabitants may come to regard such forests as natural, and seek their protection with the same vigor as they defend virgin old-growth forests. In contrast, commercial plantations often do not have a natural appearance. In order to maximize profitability, there is usually only a single species in each stand, planted in straight lines, with suppression of competing species (weeds). (A *stand* is an area of forested land containing trees of uniform age and species (or species mix), grown on a uniform management regime. It should not be confused with a *forest*, which is a larger unit, consisting of many stands). The chosen timber species may not be native to the region or the country, and is selected because it is relatively cheap and simple to grow and yields large quantities of marketable timber in a short period of time. Commercial forestry plantations can have more in common with an agricultural crop than a natural forest.

Most environmentalists are enthusiastic about natural forests, and seek their protection and enhancement. The environmental benefits of such forests are widely and readily acknowledged. Commercial plantations, however, do not always enjoy similar public support. These forests are often seen as a travesty of the natural model, with the profit motive creating a land-use that is not in the interests of the environment or the local human inhabitants. Such views can be based on emotion and prejudice, rather than on demonstrable fact, but nevertheless are a real and powerful factor in contemporary forestry throughout the world, and need to be taken seriously by foresters.

EFFECT ON SOIL

Soil quality is a subjective term, to a certain extent, because the sustainable productive potential of a soil cannot be discussed without specifying the intended land use. It appears possible, for example, that conifer plantations, on ex-pasture sites can cause a slight decline in the pH of topsoil and a dramatic reduction in earthworm numbers. Soil acidity can be a problem if the any subsequent crop requires a high pH, and if liming is not an economic option. But some crops (including conifers) prefer slightly acid soils. Similarly, earthworm numbers are not an issue unless we wish to have the sort of ecosystem that requires earthworms.

It is not possible, therefore, to state categorically that afforestation (even with monocultural conifer plantations) will cause soil deterioration, without being more specific as to what is meant. Under the influence of vegetation, nutrients on the surface of clay or organic particles can be made more available to plants, or else locked up. Such changes may be quite temporary, and removal of the vegetation can reverse the process. All commentators would agree, however, that soil can be said to have deteriorated if key elements, including nitrogen, phosphorus, potassium and sulfur, are actually removed from the site. This can occur by leaching (i.e., movement into the groundwater, under the action of rain or irrigation), by volatilization (i.e., movement of mainly nitrogen and sulfur into the air), or by extraction in agricultural or forest products.

Forests generally conserve nutrients very well. This is evidenced by the purity of the streams that flow from them. The mycorrhizal network regulates the decomposition, mineralization and uptake of nutrients in the forest floor in a way that minimizes loss to the system. In the absence of fire, there is also little loss by volatilization. Massive

depletion of nutrients caused by erosion is also less likely than in other forms of land cover. It is possible that trees, being deep rooted, can translocate nutrients from considerable depths and return these to the surface by means of litter-fall. There is debate, however, over whether this actually occurs in practice.

With regard to the removal of nutrients in harvest produce, forestry generally has a lower impact than agriculture, especially if the harvest is limited to stem-wood. This is because wood consists predominantly of a carbohydrate (cellulose, hemicellulose and lignin) which is derived from rainwater and carbon dioxide (CO_2), with minimum contribution from soil minerals. Indeed, carbon (C), hydrogen and oxygen comprise more than 99.7% of the oven-dry weight of stem-wood. In contrast, agricultural products can be rich in elements such as nitrogen, calcium, phosphorus, and sulfur, and these need to be replaced with fertilizers (organic or mineral) to avoid mining the soil.

It is somewhat academic to discuss soil degradation in terms of the concentration of essential elements, and their availability to plants, when the threat in many situations is a massive loss of topsoil as a result of erosion. For example, cultivation of fragile soils can cause wind-blown erosion, and removal of forest cover can cause slipping of soils on steep slopes. Topsoil loss must be the ultimate form of soil degradation. Not only can erosion strip a soil down to the parent rock, the ensuing silting of rivers can cause huge problems downstream. Poor forestry practices, particularly those involving roads or harvesting, can cause considerable soil erosion but in most cases this is avoidable.

Some degree of soil erosion occurs naturally, but human activity in many parts of the world has accelerated this beyond the restorative capacity of the ecosystem and its human inhabitants. Afforestation is an effective preventative, and partial cure, for most erosion. Trees reduce wind speeds at ground level. They maintain the soil in a drier state. Most importantly, they bind the soil with their strong, interlocked roots.

EFFECT ON WATER

Myths relating to the interaction of trees and water are widespread. It is not true, for example, that (in some mysterious way) trees attract rain. The energy of the sun causes water to evaporate, mainly from the oceans, and fuels the winds, albeit modified by the Coriolis effect. The forces involved are many orders of magnitude greater than the small influence of trees. Nevertheless, the water evaporated or transpired by trees adds to the humidity of the atmosphere and is likely to enhance rainfall elsewhere.

Trees influence the water cycle in two major ways. Firstly, and most importantly, they prevent rainfall reaching and penetrating the ground. They do this by interception of water on the canopy, and subsequent evaporation. Trees intercept and evaporate more water than short vegetation, not necessarily because they have a greater leaf area, but because there is greater air turbulence (and therefore drying and re-wetting) at heights farther from the ground. The second major influence of trees is due to the fact that, compared to annual crops, they are deep rooted. They continue to transpire water through their leaves long after annual crops have wilted and closed their stomata. In both of these ways, trees tend to dry the soil and prevent water reaching the waterways.

Decreased water flow as a result of afforestation can be expected at all times of the year. It is *not* true that forests act like a sponge, soaking up water during wet periods and releasing it slowly during dry periods. Indeed, even sponges do not work that way. Macropores release their water, within minutes, under the influence of gravity. In micropores, capillary action is stronger than gravity and the water is removed only by evaporation from the surface or by the action of tree roots, using osmosis. In New Zealand, reductions of water yield of 25 50% have been demonstrated, with the higher end of the range being observed on drier sites.

A forest floor has a high infiltration rate. This means that water will penetrate the ground rather than flowing over the surface, as may occur in compacted soils of agricultural systems. Improved infiltration may lengthen the time it takes for rainfall to reach the streams. In this way, together with the rainfall interception that occurs on the forest canopy, forests can act as a buffer against flood peaks. The role of forests in flood prevention, however, is often greatly exaggerated. Large storms overwhelm the limited buffering effect of trees. Also, in larger river systems, the profile of flood peaks is hardly affected by any smoothing that takes place in subsidiary streams.

The lowering of water tables following afforestation has at least one desirable side effect. In many arid regions of the world, deforestation has allowed salt laden water to reach the surface. Salt deposits then occur as a result of evaporation from saturated solutions (*see* Salinization, Volume 2). For this reason, agriculture in large parts of Australia may soon be abandoned. Planting trees in upland areas will lower the water tables and eventually (after many years of flushing by rainfall) the soil may be restored to its former productivity.

Fresh water is vital, and increasingly scarce in many regions of the world. Sometimes, planning authorities afforest key catchment areas in order to reduce soil erosion and maintain water quality. This is admirable, but it usually has the unwanted (and often unanticipated) effect of also reducing annual water yield. Hobson s choice is between abundant, but polluted, supplies or a depleted, but clean, water source.

There is no doubt that forests are the preferred vegetation cover for water quality. Even if water is derived

from deep aquifers, rather than shallow groundwater, such supplies can eventually become polluted by human activity at ground level. Pathogens (viruses, bacteria, plasmodia, etc.) present in agricultural systems can sometimes be prevented from reaching waterways, but nitrogen contamination is an intractable problem. Intensive livestock farming, or arable farming with nitrogenous fertilizers, will result in leaching of this element. This is almost inevitable, because nitrate salts are extremely soluble and can quickly bypass or overwhelm any filter or barrier. Above certain levels, nitrates and nitrites are considered to be a health hazard in drinking water, and are expensive to remove. Phosphates are somewhat less soluble, but are another major cause of water pollution. The end result of nutrient enrichment of waterways is undesirable aquatic growth, such as algal blooms. Toxins are released from these blooms, and (when the algae decay) water is stripped of its oxygen, causing eutrophication.

Other than pathogens and chemicals derived from farming, a main source of water pollution is erosion resulting from deforestation. Sedimentation can block rivers and cause them to overflow their banks (which is the major reason why afforestation is an effective preventative against flooding in larger catchments!). Silt can flow out to sea and block harbors, destroy fisheries, and so forth. Debate over the composition of trace pollutants in streams can become somewhat academic if the water is a thick silty soup.

EFFECT ON AIR

Plants remove carbon from the air by the process of photosynthesis and retain this in the form of wood. Half the dry weight of wood is elemental carbon. Land that supports a forest cover will hold considerably more carbon than land that does not, even if the forest is managed for production and parts are being felled and replanted at any particular time. The high carbon density (tonnes carbon per hectare) of forests is their primary contribution to the mitigation of global warming.

Many people have no difficulty with the concept of deforestation as a cause of increased atmospheric CO_2. The conversion of a high carbon density landscape (such as tropical rainforest) to one of lower carbon density (such as pasture) results in a transfer of carbon from the earth s surface to the atmosphere. Indeed, it has been calculated that global deforestation between 1850 and 1990 resulted in cumulative emissions of 122 ± 40 Gt C, compared to a contribution of approximately 230 Gt C from fossil fuels. In other words, deforestation is responsible for perhaps one-third of the CO_2 emissions over the last 140 years.

Many commentators fail to note that *reforestation* is merely the inverse of *deforestation*. If it were possible to restore the vanished forests in their entirety, this would compensate for the emissions caused by earlier forest removal. Just as forest clearance puts carbon into the air; so the establishing of forests takes carbon out of the air. The distinction between natural steady state forests and rotation forests is one of perception rather than one of substance relevant to carbon sequestration. Averaged over time, or over a whole landscape, both types of forest have a high carbon density.

Wood products do not increase the carbon density of the earth s surface to an appreciable extent. There is not much carbon in wooden buildings compared to the amount of carbon in trees. And while there is tremendous scope for reforesting large areas of the earth s surface and thus taking carbon out of the air, there is very limited potential for increasing carbon storage in the form of wood products. Their main value lies in the fact that they are a substitute for non-wood materials that use considerable fossil fuel and that emit CO_2 as a chemical byproduct, during their manufacture. A tree is a natural, biological solar panel and fuel storage unit. Wood is a sustainable, fossil-fuel-free product of solar energy.

EFFECT ON WILDLIFE AND PEOPLE

Afforestation, especially in tropical regions, creates an ecosystem that is considerably less complex than natural forest. The diversity of plants and animals is markedly inferior. But is this comparison valid? Afforestation is the conversion of a non-forest to that of a forest, and nearly always, this is a positive step in the direction of increased biodiversity. Forests have a vertical dimension that agricultural systems do not, and (despite attempts to control weeds) so-called monocultures can be surprisingly diverse. Also, in providing an alternative timber source, plantations may take the pressure off indigenous forests and thus protect natural habitats.

Local inhabitants can benefit from forestry, because the intrinsic biological productivity of a forest can be higher than the degraded landscape that it replaced. It can provide employment, and essentials such as house timber and firewood. About half of the world s wood is used for firewood, and this important item should not be overlooked. People can starve either because of a shortage of food, or because their staple diet (potatoes, maize, rice, cassava) requires cooking before it is digestible.

Some plantations are owned by large corporations and offer little benefit to local inhabitants. Commercial softwoods may not make good quality, slow burning charcoal, and may not provide the wealth of medicines, fruits, game and other products of the natural forest. Indeed, guards may be placed to keep locals out of the plantation. Criticism of this situation is often misplaced. The fault may lie with the ownership patterns of the community rather than with the land-use *per se*. Conversion of natural forests can be prevented by legislation, as there is (in most countries)

abundant non-forested land that is available and suitable for plantations. There is normally no need to replace complex natural ecosystems (which will usually regenerate well even after intensive logging) with simpli ed pulpwood plantations.

SUMMARY

Afforestation is usually driven by commercial imperatives. Cost minimization and revenue maximization dictate the creation of forests that are conspicuously arti cial. The comparison with natural forests generates a degree of hostility towards commercial plantations. Analysis of the environmental effects of plantation forests, however, shows that most of the bene ts are present in both types of forest. The true contrast is between forest and non-forest, and it is clear that the world would bene t from increasing the area of forest cover, even if this increase takes the form of tree plantations that resemble an agricultural crop.

Political pressure to restrict commercial afforestation will not necessarily result in a greater area of complex indigenous forest. More likely it will result in a continuation of the environmental harm that deforestation and non-wood substitutes has caused, and increased pressure to exploit and clear indigenous remnants.

Almost every sentence in this article is controversial, and therefore requires deeper discussion and source references. For further information, see Maclaren (1996) and **Forest: the FAO De nition**, Volume 2.

REFERENCE

Maclaren, J P (1996) Environmental Effects of Planted Forests in New Zealand, *FRI Bull.*, **198**, 180 (website: http://www.forestresearch.cri.nz).

Agricultural Intensification in Java

Joan Hardjono
Padjadjaran State University, Bandung, Indonesia

Recent environmental changes in Java pose threats to the sustainability of land use systems. Centuries ago, farmers developed irrigation networks and rice cultivation techniques that enabled a growing population to be supported without disrupting environmental equilibrium. Further intensification in rice cultivation occurred when high-yielding seeds were introduced in the 1970s. Despite overall production increases, loss of genetic diversity in seed and the continuous cropping of rice led to pest plagues. Meanwhile, Java s resource base is deteriorating as greater deforestation and the cultivation of vegetables and secondary food crops on steep slopes give rise to wide-spread soil erosion. Sedimentation on the lowlands in turn leads to extensive ooding in the wet season and water shortages in the dry season. Con icts of interest in the use of land and water resources occur increasingly, as the ever-growing need to produce food comes up against the wish to cultivate commercial crops like tobacco for domestic and export purposes.

The island of Java in the Indonesian archipelago illustrates how human activity has led to environmental changes that are now threatening the sustainability of the very land use systems on which survival is dependent (Table 1).

For many centuries environmental equilibrium was maintained in Java, despite constantly greater intensi cation in rice cultivation associated with the need to feed a growing population (Booth, 1988). Farmers long ago learned that careful control of water ow and drainage, in conjunction with the transplanting of seedlings and frequent weeding, would give them high yields. As time passed, they found that they could increase rice output without disrupting nature by terracing lower mountain slopes and developing gravitational irrigation systems to divert water from rivers and springs when rainfall was inadequate. Even when technically controlled irrigation networks were constructed in the nineteenth century, to ensure greater reliability in irrigation for the intensive cultivation of sugarcane in rotation with rice (Geertz, 1963), there were no adverse environmental consequences.

Table 1 Java: land use, 1994

	Million ha	Percentage of total area
Rice land[a]	3.3963	25.7
Forest land[b]	3.0128	22.8
Dry fields	3.0807	23.3
Residential land	1.7325	13.1
Plantations	0.6201	4.7
Wooded land[c]	0.3211	2.4
Fish-ponds	0.1523	1.2
Fallow agricultural land	0.0879	0.7
Grassland	0.0428	0.3
Other land[d]	0.7725	5.8
	13.2190	100.0

[a] Irrigated and rain-fed land used for rice cultivation.
[b] Land managed by the Department of Forestry.
[c] Land covered in timber-producing trees, bushes and bamboo outside the authority of the Department of Forestry.
[d] Includes land used for public facilities and industrial purposes as well as rivers, lakes and reservoirs.

Intensification was taken a step further when the high-yielding rice seeds of the Green Revolution (*see* Green Revolution, Volume 3) were introduced at the end of the 1960s. By 1985, the national output had risen to a point where Indonesia no longer needed to import rice to feed its population of some 164 million in that year. But the new technology had ecologically adverse effects. Farmers found that heavy applications of chemical fertilizers caused a hardening of certain soils, which made dry-season plowing more difficult. More importantly, loss of genetic diversity in seed and the continuous cropping of rice on the same land led to the emergence of pests, in particular the brown plant-hopper (wereng or *Nilaparvata lugens*), the spread of which was encouraged by the use of pesticides that destroyed its natural enemies (Fox, 1991). The development of new varieties of rice provided only temporary solutions to the plant-hopper problem until a ban was imposed in 1986 on the use of most pesticides on rice, though not on other crops (Table 2).

Today, land is still cultivated intensively but monoculture is less common since the adoption in the mid-1980s of a policy of crop diversification in response to the decline that had occurred in the output of important secondary food crops like soy beans and corn. Rice-farmers are now encouraged to plant at least one non-rice crop every year and some are returning to traditional rice varieties in the interests of pest control and sustained productivity. As a result greater quantities of rice have to be imported every year.

Intensive agriculture on lowland Java has not been limited to the growing of rice. In many areas there is a long tradition of intensively cultivated home gardens in which cooking spices and herbs are grown in the midst of taller plants like corn and banana, with perennials in the form of fruit and coconut trees overshadowing and protecting them. Fertilizer is rarely used in these gardens, and insecticides are not needed because the ecological variety in plant species discourages the proliferation of insect pests. Environmentally, the system functions well, because the inter-cropping of species of different heights prevents the erosion of topsoil by heavy rain. Even where vegetables like tomato, chili and eggplant are inter-cropped in open fields in a pattern that involves continuous use of the same land, the careful leveling of fields and control of water prevents soil loss.

With ever-greater population pressure, however, environmental crises have become increasingly frequent in Java in recent decades. Most have their origins in soil erosion, which is causing rapid deterioration in the island s resource base. The need for additional arable land and for more employment in farming has driven the agricultural frontier farther up mountain slopes, causing accelerated change in land use systems. In some places change has involved intensive horticulture and in others expansion of the area under secondary food crops.

Table 2 Java: area and production of rice, 1961–1997

	Harvested area (10^3 ha)[a]	Output (million tons)[b]
1961	3992	9.2
1962	4087	9.9
1963	3647	8.5
1964	3655	8.4
1965	4034	9.6
1966	4117	10.0
1967	4021	9.6
1968	4264	10.7
1969	4294	11.3
1970	4302	13.7
1971	4416	13.9
1972	4336	15.6
1973	4567	13.0
1974	4730	13.9
1975	4653	13.7
1976	4466	14.1
1977	4378	13.7
1978	4750	15.6
1979	4628	15.7
1980	4777	18.4
1981	5046	20.5
1982	4749	20.9
1983	4799	21.7
1984	5212	23.5
1985	5301	24.2
1986	5331	24.5
1987	5185	24.5
1988	5208	25.1
1989	5439	27.1
1990	5419	27.2
1991	5184	26.4
1992	5553	27.3
1993	5515	27.4
1994	5176	25.7
1995	5479	27.2
1996	5475	27.4
1997	5381	27.9

[a] Irrigated and rain-fed fields.
[b] Dry, unhusked rice.

The development that has occurred in high-altitude horticulture can be traced in part to earlier land-use policies. In the late nineteenth century large-scale expansion in plantation agriculture took place at altitudes above 1250 m and on lower land on Java s southern mountain slopes, where poor-quality calcareous soils make the cultivation of annuals economically unrewarding. At the same time horticulture, often in conjunction with dairying, was established on a small scale in upland areas close to major cities to meet urban demands for vegetables.

Since the 1960s local farmers have been moving onto former plantation land and, in imitation of older horticultural practices, have planted crops such as cabbage,

cauliflower, carrot and potato, which bring better returns than perennials like tea, rubber, coffee and cinchona. Cultivators have continued to move upwards, and today are clearing land above 1700 m for vegetable growing. Adopting the intensive cultivation methods of rice growers, they transplant vegetable seedlings, weed frequently and hand-water crops when rainfall is irregular. In addition to animal manure, they use large quantities of chemical fertilizers to maintain soil fertility and apply pesticides to control the insects that thrive when as many as four crops of vegetables are grown annually on the same land. Due to the need to ensure good drainage, land is not terraced despite the readily apparent loss of topsoil from fields, which are often dug out on slopes greater than 45° (Hardjono, 1991).

At the same time, expansion in food-crop cultivation is taking place in the uplands on soils not inherently fertile enough for horticulture. The area under permanent tree cover is declining as farmers clear steep slopes to plant crops like cassava on land that they often neglect when prices are low. Meanwhile, other people remove trees of all kinds for sale as timber or fuel. Reforestation efforts have proved relatively ineffective, because with socio-economic pressures increasing every year, the forestry agency permits cultivators to inter-crop vegetables with newly planted trees, a practice that tends to impede tree growth and to encourage further encroachment on forested land. The consequences can be seen in the runoff of soil and in the landslides and flash floods that occur frequently in the wet season. At the same time, reduced infiltration of water on denuded slopes is affecting the flow from springs used for domestic purposes and farming at lower altitudes. While the replacement of primary forest by perennials a century ago had few adverse effects on the environment, the change from perennials to annuals has led to high rates of erosion in Java s major watersheds.

The consequences of erosion are being increasingly felt at lower altitudes in the sedimentation of rivers, canals and harbors and in the reduced storage capacity of the island s reservoirs, some of which can no longer provide the quantities of water needed to irrigate rice fields or generate electricity in the dry season. Extensive flooding occurs in many lowland areas every year, making the cultivation of any kind of crop impossible, yet pumps are needed to lift water onto fields in the dry season when water levels drop in rivers. Urban residents experience the consequences when floods damage infrastructure like roads, bridges and train lines and inundate low-lying, formerly irrigated farmland that has been converted to human settlements and factory sites.

The changes in aquaculture since 1980 along the northern coast of Java illustrate the way in which a long-established system of land use can have unfavorable environmental impacts when intensified, yet can itself be threatened in terms of sustainability by human activity elsewhere. Traditional fish-farming in brackish water ponds called *tambak* involves a modification of the mangrove forest ecosystem that constitutes the natural habitat of shrimp and milkfish. The economic profitability of these commodities has led to both intensification and expansion in area. Fish-farmers now use modern technology in the form of water aerators, pumps and manufactured fish-feed to produce greater quantities of shrimp per unit of land while at the same time clearing more land to make ponds (*see* Aquaculture and Environment: Global View from the Tropics to High Latitudes, Volume 3; Aquaculture in Asia, Volume 3).

The widespread destruction of mangroves has, however, led to a significant decline in the larvae and fry that fish-farmers need to restock their ponds and has exposed the coast to the abrasive action of the Java Sea. At the same time, deterioration of the environment some distance inland has begun to undermine the viability of the production system. Sedimentation has reduced the capacity of the channels that carry water into ponds and drain them, while the silt that passes into ponds carries with it traces of highly toxic fertilizers and insecticides. Ponds are increasingly affected by flooding, especially in the wet season when drainage canals are unable to carry water away from agricultural land in the interior.

The environment in Java has changed profoundly over the centuries since human settlement began. Intensification can go no further and today the long-term sustainability of agriculture is open to question. Even to maintain the present levels of production, measures are required that will restrict certain forms of human activity and for that reason are not likely to be adopted. Conflicts of interests in the use of land and water resources occur increasingly, as the need to produce food comes up against the demand for commercial crops like sugar, tea and tobacco for domestic and export purposes.

REFERENCES

Booth, A (1988) *Agricultural Development in Indonesia*, Allen & Unwin, Sydney.

Fox, J J (1991) Managing the Ecology of Rice Production in Indonesia, in *Indonesia: Resources, Ecology, and Environment*, ed J Hardjono, Oxford University Press, Singapore, 61 84.

Geertz, C (1963) *Agricultural Involution: The Process of Ecological Change in Indonesia*, University of California Press, Berkeley, CA.

Hardjono, J (1991) Environment or Employment: Vegetable Cultivation in West Java, in *Indonesia: Resources, Ecology, and Environment*, ed J Hardjono, Oxford University Press, Singapore, 133 153.

Agricultural Intensification in Western Europe

Ian Bowler
University of Leicester, Leicester, UK

Western Europe is one of the world's most intensively farmed regions, measuring intensi cation as agricultural inputs or outputs per hectare of farmland. However, there is variation between farm types and farm sizes and consequently intensi cation varies between regions within western Europe. There is some recent evidence to suggest a reversal of intensi cation (i.e., extensi cation) of agriculture in some regions in an era of post-productivism.

Intensi cation has been led by the demand for food from urban consumers, the production and application of new farm technologies, the in uence of food processing and retailing corporations, and intervention by the state. A range of damaging consequences has been created by agricultural intensi cation, including excessive costs of state intervention, the production of food surpluses, and environmental impacts in terms of air and water pollution and natural habitat loss. Most recently social problems associated with food health and safety have emerged, with consumers increasingly demanding food whose origins can be traced and whose quality can be guaranteed.

Remedial action is now underway, including a reduction in state intervention, the introduction of sustainable farming practices, payment for the production of environmental goods and reductions in purchased inputs into farming. However, climate change in western Europe will require further adjustments to production patterns and farming practices, with varied and uncertain impacts on the development trajectories of farming regions.

INTRODUCTION

Western Europe has emerged as one of the world s most intensively farmed regions, measured by either the inputs to, or outputs from, the agricultural sector (Dent and McGregor, 1994). Farming inputs and outputs can be examined as a ratio of the capital or labor employed in agriculture, but this analysis follows the convention of measurement per hectare of farmland. The contribution of livestock to farm production is a second global characteristic of western European agriculture, although the role of livestock, as with the level of intensification, varies between particular farm types and locations (Table 1).

The inputs to contemporary agriculture in western Europe are mainly purchased from the non-farm sector rather than generated on the farm. They include inorganic fertilizers (nitrogen, phosphorous and potassium), agri-chemicals (herbicides, fungicides, pesticides), crop seeds, livestock (pigs, poultry, milk cattle, beef cattle, sheep), plant (crop preparation for marketing, on-farm food processing), machinery (tractors, field cultivation equipment, milking equipment), livestock feed, petroleum and labor. Livestock feed, in particular, is purchased outside Europe, so that the farm sector effectively adds additional hectares to the western European farm, thereby contributing to the ecological footprint of western European agriculture (*see* Ecological Footprint, Volume 3). Land, as an input, is mainly owner-occupied, but with a rented (i.e., tenant-farm) sector; the intensive nature of agriculture is reflected in the high value of farmland, which causes it to be a resource for both production and consumption (i.e., investment).

The outputs from agriculture form the raw materials for the food processing, distribution and retailing sectors, but can be measured in terms of either volume or value. The most intensive outputs per hectare of farmland by volume are produced by crop farming, for example, cereals and field vegetables. Measured by value, however, horticultural, housed pig/poultry, beef feedlot and dairy production units are amongst the most intensive producers of agricultural products (Table 2).

THE CAUSES OF INTENSIFICATION IN AGRICULTURE

Intensification in west European agriculture has increased over time and the causal processes can be simplified to five headings: demand for food, farming technology, non-farm sectors of the economy, economies of scale and state intervention.

1. *Demand for food*: the urban population of western Europe is both numerous and relatively wealthy, expressing its purchasing power through a demand for high quality, high protein, processed foods. Livestock plays an important role in meeting this demand in terms of meat products, milk products and eggs, while the farming systems required to produce such products, especially in the volumes demanded, are intensive rather than extensive in character.
2. *Farming technology*: industrial capital has produced ever more sophisticated technologies (e.g., farm machinery, farming practices, crop varieties, livestock breeds) for raising the productivity of resources employed in the farm sector. The impact can be seen especially in the substitution of technology for farm labor; the latter, historically, has been the most expensive factor of production. This capitalization of agriculture, through the adoption of new farm technology, continues to raise the productivity of the land, displaces labor into the urban-industrial economy, and permits the evolution of intensive agriculture. The current controversy over the

Table 1 Enterprise outputs for farm types within the European Union (EU). (Output of each enterprise as a percentage of the total value of output of the farm type)[a]

	Enterprise							
Farm type	Cereals	Field crops	Fruit	Dairy	Beef	Pigs	Poultry	Total[b]
Arable	36	26	1.7	1.3	3.6	2.6	1.6	72.8
Dairying	3.5	0.8	0	68	19	1.5	0.4	75.2
Dry stock	5.5	1.1	0.1	16	42	1.0	0.6	66.3
Pigs/poultry	4.3	0.9	0.1	0.3	0.8	71	18	95.4
Mixed	15	6.3	0.4	21	14	25	5	86.7
All farms	12	7	4	22	12	10	3	70.0

[a] *Source:* European Commission, 1998.
[b] Totals do not sum to 100% as only a selection of enterprises is recorded.

Table 2 Agricultural production in member states of the EU. (Percentage of each product in the final value of production of each country)[a]

Product	Germany	Greece	Spain	France	Sweden	United Kingdom	EU
Wheat	5	3	3	9	8	11	5
Fresh vegetables	3	17	14	7	4	8	9
Milk	25	12	8	17	33	24	18
Beef	11	3	6	13	10	8	10
Pigs	17	3	13	8	13	9	12
Poultry	3	3	5	7	3	10	5
Total[b]	64	68	49	71	71	70	59

[a] *Source:* European Commission, 1998.
[b] Totals do not sum to 100% as only a selection of products is recorded.

introduction of genetically modified crops is but the most recent manifestation of the long-term technological evolution of an intensive, industrialized agriculture.

3. *Non-farm sectors of the economy*: the food processing and retailing sectors form an increasingly important part of the marketing chain between producers and consumers. Food processing companies and major retailers, by exercising a control over the prices received in the farm sector, have created such nancial pressure on farmers as to require them continually to seek cost-reducing and output-increasing farm technologies (i.e., a price-cost squeeze). This process of induced innovation has ensured the continued application of new technology in agriculture, although historically the result has been to drive down the real value of farm gate prices. In addition, forward contracts between farmers and food producers and retailers, which increasingly characterize the food supply system in western Europe, commonly specify the farm technologies to be employed in agricultural production.
4. *Economies of scale*: in the farm sector, increased output can create economies of scale, whereby the costs of production are spread over more units of output (i.e., tons of wheat or litres of milk). But economies of scale can also be obtained by increasing the concentration of land into ever-larger farm units and by simplifying and specializing farming systems. The forces of competitive capitalism ensure the elimination of the least economically successful farming units and, through the land market, the acquisition of their land by larger, more successful farm businesses. On specialization, by eliminating all but the most nancially important farm enterprises, economies can be obtained in terms of farming and management skills, investment in plant (e.g., milk cooling and storage equipment) and machinery (e.g., potato planters, fruit harvesters), marketing contracts and volume of production. Specialization can be observed, for example, in the wine producing regions of France, Italy and Spain, the intensive horticultural regions of the Netherlands, and the cereals producing regions of northern France and eastern England.
5. *State intervention*: the state has been deeply implicated in these evolutionary trends throughout western Europe, rst through national agricultural policies and subsequently, since the 1960s, through the development of the Common Agricultural Policy (CAP) of the EU (*see* **Common Agricultural Policy (CAP)**, Volume 3). The state has: (a) subsidized research and development in the production of new farm technologies (e.g., experimental farms and research laboratories); (b) funded education and extension services, whereby advice and information on farming technology is diffused through the farm population; (c) supported the prices received

at the farm gate, thereby maintaining farm profits and the capacity to invest in new technology; and (d) subsidized programmes of farm modernization that have raised the intensification of agriculture. Taken together, the state has maintained an economic environment which, by reducing the risks attached to farm investment and food production, has stimulated the specialization of farming systems, financially rewarded investment in larger farm businesses, including raising the asset value of farmland, and encouraged the intensification of agriculture. This has been particularly noticeable in state price support through intervention agencies: not only has the farm sector in western Europe been sheltered from world markets and guaranteed a certain price for its produce, but export subsidies have enabled surplus production to be dispersed onto world markets, albeit at the cost of destabilizing the agricultural exports of other countries, such as the USA, New Zealand, Brazil and Argentina.

THE GEOGRAPHY OF INTENSIFICATION IN AGRICULTURE

Some perspective is needed on these generalizations, for the intensification of agriculture in western Europe is neither uniform nor unidirectional. Firstly, as already noted, the degree of intensification varies by farm type and, since farm type varies by location throughout western Europe, the intensification of agriculture also varies over space (Bowler, 1985). Secondly, intensification varies with farm size: in general, smaller farms have higher inputs and outputs per hectare of their land compared with larger farms. Since farm size also varies significantly by country and region within western Europe, so intensification varies over space. In broad terms, the intensification of agriculture is greatest in the Netherlands and Belgium and decreases outwards towards the regions at the spatial margins of western Europe. There are outlying regions of agricultural intensification, for example, in the Po Valley of northern Italy, Brittany in northern France, southeast England and along Europe s Mediterranean littoral.

Intensification in agriculture, however, is no longer unidirectional. There are now signs in western Europe of a counter-movement towards extensification, but again, the process is spatially variable (Laurent and Bowler, 1997). At issue is the evolution of agriculture: since the early 1980s, farm development has progressed from a productivist to a post-productivist era, led by a revision of state intervention in the farm sector. Rather than an orientation towards maximizing food production, as in the four decades following the Second World War, new objectives for state intervention are being set for the farm sector. These include the staged reduction in state financial support for agriculture under the CAP; the introduction of more liberal trading policies under the World Trade Organization (WTO), so that west European farmers are more open to competition from food producers elsewhere in the world; the development of farming systems that are more sustainable in their use of resources (e.g., integrated farming systems (IFS) and organic agriculture); the production of environmental goods (e.g., particular biotopes and landscapes), in parallel with conventional farm products; and the diversification of incomes within agriculture, so that producing raw materials for the food sector is no longer the sole objective for increasing numbers of farm businesses. To date there is little evidence of a reversal in the process of concentration of agricultural holdings (i.e., increasing average farm size).

Increasing intensification remains the dominant trend, the regions exhibiting the greatest rates of intensification being located in Scotland, southern England and central France (Figure 1), in association with a large-farm structure or the development of cereal farming (Jansen and Hetsen, 1991). Converse tendencies are detectable in other regions of western Europe, for example, extensification (i.e., lower inputs and outputs per hectare of farmland) can be detected in central Greece, northern and southwestern regions of France, Wales, Northern Ireland, northern England and northern Germany.

THE CONSEQUENCES OF INTENSIFICATION IN AGRICULTURE

Three groups of consequences have flowed from the intensification of west European agriculture: economic, social and environmental. In the first group can be placed the excessive costs of state intervention and the surplus farm production that it has generated. In the EU, the costs of the CAP escalated through the 1970s and into the early 1980s, so that by 1984 agriculture was consuming nearly 66% of the EU s total budget of 27 000 million ECU (Bowler, 1985). Most of the funding was spent supporting the prices of a handful of products (i.e., milk, cereals and beef), produced mainly in the northern regions of the EU (Table 3). In effect, technological progress in agricultural intensification had been so successful as to outstrip the more modest increases in the domestic demand for food. Moreover, the costs and prices at which farm production had been secured were so high as to render the surplus production untradable at world market prices without further export subsidies. At the same time as undercutting market prices for traditional agricultural exporting countries, such as the USA, New Zealand, Brazil and Argentina, and protecting European markets against competitive imports, the EU also purchased increasing quantities of animal feed and vegetable oil from all parts of the world, including some from developing countries.

On social consequences, research on the allocation of state subsidies to agriculture during the 1980s found that

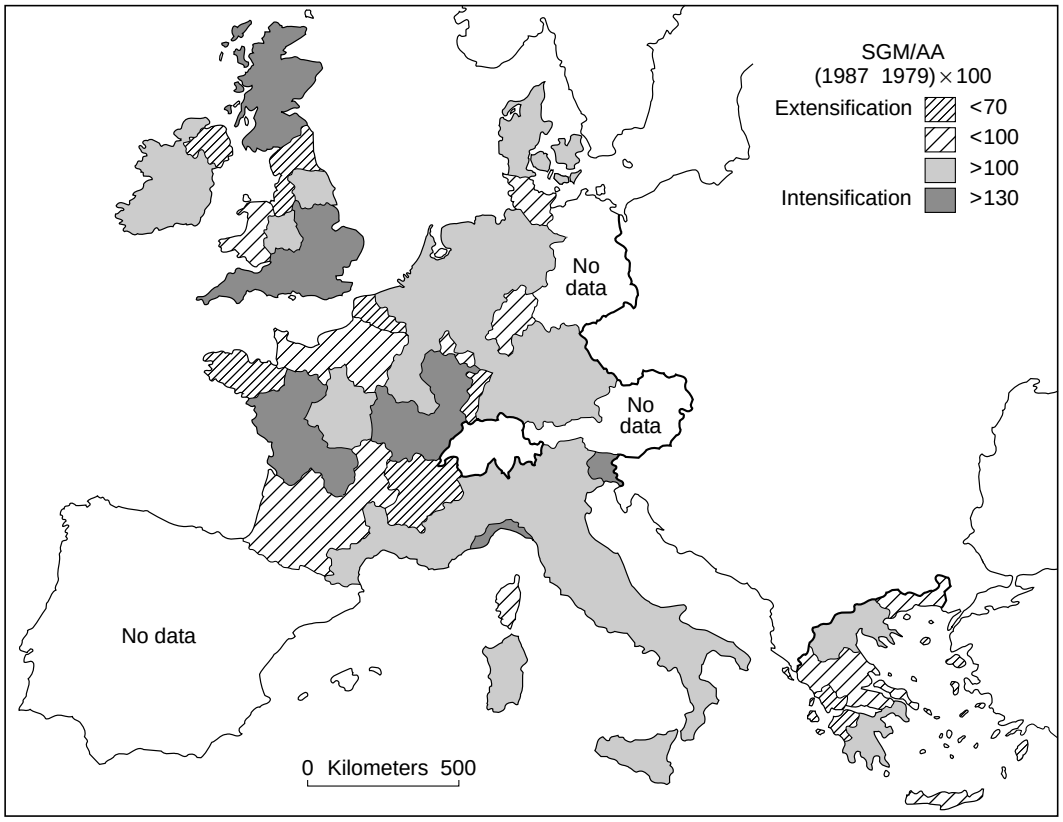


Figure 1 Changing intensification of agriculture within the EU, 1979–1987 (standard gross margin per hectare of agricultural land – SGM/AA). (*Source:* adapted from Laurent and Bowler, 1997)

Table 3 The allocation of expenditure under the CAP

Item	1990	1994	1998[a]
Budget of the EU (m ECU)	44 379	59 909	81 434
Budget of CAP (% EU budget)	64	60	56
Guarantee element (% CAP budget)	93	92	90
Guidance element (% CAP budget)	7	8	10
Plant products (% CAP budget)	38	61	59
Animal products (% CAP budget)	52	27	24

[a] Official estimate. Compiled from data in the European Commission (various years) *Agricultural Situation in the European Union. Annual Report*. The Commission, Brussels.

the major beneficiaries were larger farms, located mainly in northern regions of the EU, rather than small, financially struggling family farms for whom they were intended. Moreover, a welfare transfer payment was taking place from poorer, tax-paying urban families to wealthier farm families, as well as between the regions and member states of the EU. A succession of food scares in the late 1980s and 1990s, notably the link established between bovine spongiform encephalopothy (BSE mad cow disease) and a new variant Creutzfeldt-Jacob disease (nvCJD) in humans, drew attention to the problems of health and safety for food produced by intensive farming methods. In sum, political, taxpayer and consumer support for intensive agriculture diminished in western Europe during the 1990s, as demonstrated by social resistance to the introduction of genetically modified foods at the end of that decade.

Arguably, pollution of atmospheric, water and soil resources, together with destruction of natural habitats, comprise the greatest impacts of agricultural intensification on global environmental change. On air pollution, intensive agriculture contributes to the emission of a number of greenhouse gases (GHG), with consequences for global and regional climate change. The conversion of grassland to intensive crop farming, for example, releases carbon stored in organic matter in the soil; the conversion of forests to crops and grass has released similar carbon fluxes over the last 300 years. Carbon dioxide is also released by the burning of crop residues, as well as the use of petroleum to power agricultural plant (e.g., grain dryers) and machinery (e.g., tractors). Nitrous oxide is emitted from inorganic fertilizers, while methane emissions are associated with enteric fermentation in beef cattle and milk cows and with stored

slurry (*see* Cattle Grazing: Impacts on Land Cover and Methane Emissions, Volume 3; Nitrate Protection Zones in Europe, Volume 3; Nitrogen: Agricultural Uses and their Impacts, Volume 3; Organic Farming and the Environment, Volume 3). An estimate for developed countries suggests that between 6 and 15% of their GHG emissions can be traced to agriculture.

Looking now at water pollution, this impact is related mainly to the increased application of inorganic fertilizers in both crop and grassland farming (Brouwer *et al.*, 1991). The nitrification of groundwater, and the eutrophication of watercourses, from the excessive spreading of fertilizers and animal manures from intensive livestock units, has been recorded across western Europe. In addition, evidence is mounting of the accumulation of pesticide residues in both soils and farm produce, and discharges of livestock slurry into streams and rivers (Kronert *et al.*, 1999). Such evidence can be added to other documented impacts of agricultural intensification, such as the compaction, erosion and salinization of soils and the lowering of water tables through the excessive pumping of water for irrigation.

Destruction of natural habitats, with increases in arable land, has also been reported from surveys across western Europe, together with the loss of diversified land uses (e.g., permanent pasture, rough grazings and woodland) (Kronert *et al.*, 1999). A broad reduction in biodiversity has been recorded both as regards flora and fauna. For example, the loss of biodiversity through the clearance of open olive and cork woodland has been recorded in southern Portugal and Spain; in the UK, a national survey of rough grazings (i.e., uncultivated marshland, moorland, hills and mountains), and their associated diversity of flora and fauna, recorded a decrease of 28% between 1945 and 1990.

ADDRESSING THE CONSEQUENCES OF INTENSIFICATION IN AGRICULTURE

Four broad approaches can be identified for addressing the damaging consequences of intensification in agriculture, namely: adjustment to farm prices, extensification of agriculture, sustainable farming, and agri-environmental measures. Looking first at agricultural prices, under the General Agreement on Tariffs and Trade (GATT, subsequently the WTO) the EU is liberalizing its markets for farm produce by the staged reduction in levels of price support, and the withdrawal of protection against food imports (Arden Clarke, 1992; Robinson and Ilbery, 1993). The anticipated, phased reduction by farmers of their purchased inputs (e.g., fertilizers and agri-chemicals), in the face of falling prices for their produce, is expected to produce significant environmental benefits. However, the relationship between reducing the level of state support for agriculture, and improved environmental conditions is indirect.

Turning now to extensification in agriculture, the EU is sheltering the farm population from rapid adjustments to farm prices by providing compensatory financial payments. For example, farmers who agree either to set aside their land from production (e.g., 10% of their arable area each year) or keep a limited number of livestock per hectare (beef cattle and sheep) receive direct payments (Ilbery, 1992). Another approach has been to compensate farmers in designated areas for reducing their fertilizer applications: this has taken the form of Nitrate Vulnerable Zones (NVZ) in the EU, where the designated NVZs are significant for the collection of water for human consumption (*see* Nitrate Protection Zones in Europe, Volume 3). One useful by-product of these land-use changes has been a modest reduction in GHG emissions from agriculture.

On *sustainable farming*, three dimensions are evident: integrated farming systems (IFS), organic farming and food quality (Ilbery *et al.*, 1997). IFS involves farmers in developing farming practices, such as the management of field margins to increase biodiversity, specified management practices to combat soil erosion (e.g., minimum cultivation, winter cover crops and contour ploughing), biological controls for crop pests and fungal diseases to replace agri-chemicals, green and animal manures to reduce inorganic fertilizers, and crop rotations for land use diversity. They aim to replace non-farm with farm-based inputs. Parallel state assistance for farm diversification aims to reverse the specialization of land use that is characteristic of industrialized agriculture, including the spatial separation of crop and livestock production. For instance, planting trees serves to diversify land use for a variety of objectives, including timber, shelter for livestock, commercial Christmas tree production, amenity and as an energy crop (e.g., willow).

Organic farming (i.e., farming without using inorganic fertilizers and agri-chemicals) is encouraged by financial compensation from the state for loss of income while the transition is made from conventional agriculture to certified organic production. The transition can take from two to four years, depending on regulations in particular countries, and during that time, crop and animal yields are reduced without benefit of the higher price premium that certified organic food attracts. Few countries provide subsidies to organic farming once it is established returns have to be gained through the market. To date, however, organic farming is supported by only a minority of consumers and, as a form of sustainable land use, it is in its infancy (*see* Organic Farming and the Environment, Volume 3).

Speciality food products (SFP e.g., farmhouse cheeses, hams, wines, berries, organic food, fruit and meat pies, meats, yogurts) are emerging as another means of addressing the damaging food health and safety aspects of agricultural intensification. Individual producers are supplying high-income, niche markets at premium

prices, thereby avoiding competition in mass-produced food markets dominated by large corporations. In effect, producers and consumers of SFP are acting together to create alternative food networks (Bowler, 1999). Quality assurance schemes have been developed through self-regulation by producer groups (e.g., cooperatives) and supermarket chains, while the EU has regulations 2092/91 and 2083/92 on biological agriculture (labeling and inspection), regulation 2081/92 on the protection of geographical indications and the protection of designations of origin, and regulation 2082/92 on certi cates of special character. Nevertheless, SFP account for less than 10% of the food market in the EU at their present level of development.

Agri-environmental measures include the direct regulation of pollution within individual countries and, within the member states of the EU, national Agri-Environmental Programs (AEP regulation 2078/92) (Robinson, 1991; Wynne, 1994). AEP represents a shift in the conservation effort away from individual sites (e.g., Nature Reserves, sites of special scienti c interest) towards more extensive tracts of countryside. Each member state has been able to develop and implement its own national schemes, with up to 50% of the cost paid by the European Agricultural Guidance and Guarantee Fund. The agricultural impact of agri-environmental schemes is spatially uneven, however, as reliance is placed on voluntarism and farmer response to nancial incentives.

CONCLUSION

A broad consensus of expert opinion believes that the contemporary level of intensi cation in west European agriculture is not sustainable in terms of economic, social and environmental development. Nevertheless, the level of intensi cation, and its associated problems, vary in magnitude and type between countries and regions. In particular, a broad polarization within western Europe can be detected. On the one hand lie regions favoured by climate, soil, topography or location near the main urban markets, where the continuing development of industrial agriculture and intensi cation can be observed. On the other hand lie regions lacking in those favorable attributes, with farming systems increasingly marginalized by changes in the agri-food system. Most of these regions lie around the periphery of the economic core of the EU. Ironically, the potential for the development of a more sustainable agriculture is greater in these latter regions with their orientation towards more extensive farming systems and practices.

Overarching these considerations, however, are the predicted climate changes associated with human-induced warming of the atmosphere ($+2\,^{\circ}C$ within a range 1.0 3.5 $^{\circ}C$) by 2100, with an associated rise in mean sea-levels ($+0.5$ m within a range 0.15 to 0.95 m) (Adger *et al.*, 1997). Regional estimates within global models (e.g., the Future Agricultural Resources Model of the US Department of Agriculture) predict a reduction in agricultural output within Europe, with wheat production falling by 11.6%, non-grain crops by 10.6% and the number of livestock by 1.5%. Within-Europe estimates are even more uncertain, but in broad terms climatic belts, and their associated crop and livestock production systems, are expected to move northwards within Europe, but at a scale and with consequences that cannot yet be predicted with any accuracy.

REFERENCES

Adger, W N, Pettenella, D, and Whitby, M (1997) *Climate-change Mitigation and European Land-use Policies*, CAB International, Wallingford.

Arden Clarke, C (1992) Agriculture and Environment in the GATT: Integration or Collision? *Ecos.*, **13**, 9 14.

Bowler, I (1985) *Agriculture Under the Common Agricultural Policy*, Manchester University Press, Manchester.

Bowler, I (1999) Endogenous Agricultural Development in Western Europe, *Tijdschrift voor Economische en Sociale Geografie*, **90**, 260 271.

Brouwer, F, Thomas, A, and Chadwick, M, eds (1991) *Land Use Changes in Europe*, Kluwer, London.

Dent, J and McGregor, M, eds (1994) *Rural and Farming Systems Analysis: European Perspectives*, CAB International, Wallingford.

European Commission (1998) *Agricultural Situation in the European Union. 1997 Report*, The Commission, Brussels.

Hampicke, U and Roth, D (2000) Costs of land use for conservation in Central Europe and future agricultural policy. *Int. J. Agricultural Resources, Governance and Ecology*, **1**, 95 108.

Ilbery, B, Chiotti, Q, and Rickard, T, eds (1997) *Agricultural Restructuring and Sustainability*, CAB International, Wallingford.

Ilbery, B (1992) Agricultural Policy and Land Diversion in the European Community, in *Progress in Rural Policy and Planning Volume 2*, ed A Gilg, Belhaven Press, London, 153 166.

Jansen, A and Hetsen, H (1991) Agricultural Development and Spatial Organization in Europe, *J. Rural Stud.*, **7**, 143 151.

Kronert, R, Baudry, J, Bowler, I, and Reenberg, A (1999) *Land-use Changes and their Environmental Impact in Rural Areas in Europe*, UNESCO Man and Biosphere Series 24, Parthenon Publishing, New York.

Laurent, C and Bowler, I, eds (1997) CAP and the Regions: Building a Multidisciplinary Framework for the Analysis of the EU Agricultural Space, INRA, Versailles.

Robinson, G (1991) EC Agricultural Policy and the Environment: Land Use Implications in the UK, *Land Use Policy*, **8**, 301 311.

Robinson, G and Ilbery, B (1993) Reforming the CAP: Beyond MacSharry, in *Progress in Rural Policy and Planning Volume 3*, ed A Gilg, Belhaven Press, London, 197 207.

Wynne, P (1994) Agri-environment Schemes: Recent Events and Forthcoming Attractions, *Ecos.*, **15**, 48 52.

Agricultural Subsidies and Environmental Change

John Lingard
University of Newcastle upon Tyne, Newcastle upon Tyne, UK

With or without subsidies, agriculture as a major land use has a profound effect on the environment; environmental degradation by farmers has been going on for millennia, but many farmers have learnt to look after the natural resources that they use and have responded quickly to economic incentives to do so as they seek ways to sustain their livelihoods. Agricultural activities impact on the environment via soil quality (texture, erodibility, nutrient depletion, moisture balances, salinity and soil conservation, including ood protection and landscape), water systems, including surface and groundwater pollution and irrigation, air quality, including greenhouse gas emissions, biodiversity, wildlife habitats and ecosystems. Environmental change is an inevitable by-product of agricultural activity; policy instruments like subsidies may, by modifying both producer and consumer behavior, increase or decrease the rate of environmental change. Population pressures on the natural resource base are a further important source of environmental change, as an ever-expanding population attempts to satisfy its increasing demands for a wide variety of goods and services.

Conventional, neo-classical production economics theory suggests that input price subsidies/taxes and output price subsidies/taxes will promote intensi cation/extensi cation of production processes. Subsidies will increase the use of various kinds of production inputs, such as fertilizer, irrigation water, pesticides and herbicides; output price subsidies will change the optimal combination or factor proportions with which inputs are used, and they will lead farmers to substitute one crop for another or change between crop production and livestock production processes. Associated with this changing farmer/land user behavior will be different patterns of environmental impacts having both local and wider implications; that is, the environmental impacts will be felt at local, river catchment, regional and global levels. Examples include non-point pollution effects of agricultural activity, water quality and sedimentation, and the global effects due to the carbon balances of agriculture. There is some ambiguity about the role of subsidies and environmental change; subsidies by changing price signals may lead farmers to substitute polluting inputs for non-polluting ones, or to change from production processes which give low emissions (e.g., cereals and sheep) to those giving high emissions (e.g., dairy cows).

AGRICULTURAL SUBSIDIES

The environmental effects of economic support and subsidies to agriculture have recently attracted considerable scientific and political interest. Agricultural subsidies can take many forms, but a common feature is an economic transfer, often in direct cash form, from governments to farmers. These transfers may aim to reduce the costs of production in the form of an input subsidy, e.g., for inorganic fertilizers or pesticides, or to make up the difference between the actual market price for farm output and a higher *guaranteed price*. Subsidies shield sectors or products from international competition. The stated governmental aims of agricultural policy worldwide are many, but essentially involve farm income support and price stabilization. However, by artificially reducing the costs of production or providing taxpayers money for an output that nobody really wants (at least at such high prices), agricultural subsidies encourage wasteful use of materials, energy and natural resources and also encourage over-production.

Analysis of the environmental and economic impacts of agricultural subsidies is exceedingly complex, but many are unquestionably damaging, for example, the practice in forested tropical countries of providing cash incentives for clearing forest land for agriculture and livestock production. Similarly, subsidies to irrigation water, in the form of less than full-cost recovery pricing, encourage over-use of scarce water, and hence, water logging and soil salinization. In contrast, a subsidy to promote and encourage kerosene consumption may be environmentally beneficial if it reduces the demand for fuel wood and deforestation. Deciding which subsidies are, or are not, environmentally benign is extremely hazardous. Boldly stated, agricultural subsidies can encourage the production of environmentally harmful pollution, lead to the excessive use of natural resources and often impose high costs on consumers, taxpayers and government budgets. Their reduction/removal would increase economic efficiency, reduce government spending and, at the same time, improve environmental quality. Farm incomes and profitability will eventually recover following an initial adjustment period.

The exclusion of environmental externalities (e.g., pollution) from the profit and loss accounts of farmers and land users means that environmental damage caused by their economic activities is not paid for by those directly responsible for causing the externality. Private costs differ from *social costs*, and society and the environment must pick up the bill. This is often aggravated by government agricultural support or subsidy programs, which artificially raise the price of agricultural output and further encourage agricultural production and the associated, unpriced environmentally harmful by-products. Support removal, along with complementary policies to internalize social and environmental externalities, will lead to society *getting the prices right* and optimizing the economic system.

There is, however, no precise, easily disentangled, quantitative link between the size and type of agricultural support and the environmental damage caused. The linkages between support and the environment are complex and often indirect and depend upon:

1. agricultural input and output markets, competition within them, and particularly, the price elasticities or responses to price changes;
2. the availability of substitute technologies;
3. taxation regimes and institutional and regulatory frameworks;
4. the biophysical characteristics of the recipient environment, particularly its assimilative capacity. Farming practice effects on the environment will be dependent on site-specific, agri-environmental conditions.

This makes a general environmental assessment of the benefits of agricultural subsidy removal extremely difficult.

MEASURING AGRICULTURAL SUBSIDIES

There are a wide variety of government agricultural support policies and programs worldwide, and the measurement of subsidies is difficult. General Agreement on Tariffs and Trade (GATT) talks and World Trade Organization (WTO) discussions require quantification of agricultural protection by various countries in order to reduce complicated agricultural trade negotiations to measurable dimensions. A benchmark or baseline from which subsidies can be measured is required, and economists often use world prices as this benchmark; that is, the price that could be obtained if a product or resource was sold internationally. This is based upon the concept of opportunity cost to the nation. For example, if wheat sells for \$150 per tonne on the world market and the European Union (EU) pays its farmers \$200 per tonne, EU farmers are effectively being subsidized by \$50 per tonne. Likewise, if fuel oil sells for \$10 per barrel in a country but, if exported, could secure \$15 per barrel, then the domestic market is being subsidized by \$5 per barrel.

Producer subsidy equivalents (PSEs) are a widely accepted measure of the extent of agricultural subsidization (Cahill and Legg, 1989). PSEs measure the overall level of support to producers of a given commodity. Expressed as a percentage, they represent that part of the value of output accounted for by various kinds of market price support, direct income support, indirect income support and other support. If P_W is the unassisted price frequently determined by reference to the world market price, the price at which food imports are available at the border, then PSEs measure the gap between this price and the domestic price that the policy causes. For domestic price, P_D, output, Q, direct income support, D, levies, L, and combined indirect income support and other support, B, the PSE is defined as:

$$\text{Total PSE} = Q(P_D - P_W) + D - L + B \quad (1)$$

$$\text{Per Unit PSE} = \frac{\text{Total PSE}}{Q} \quad (2)$$

The total PSE is thus the net addition to the value of a commodity above the unassisted world price. The level of a country s PSE normally changes whenever the world price, in terms of that country s currency, changes: that is, it is sensitive to foreign exchange rate changes. In most developed countries, PSEs are positive, i.e., farmers are subsidized (and consumers are taxed); the higher the percentage PSE, the higher the per unit subsidy. However, in many developing countries and economies in transition, the opposite is the case and the PSEs are negative; i.e., farmers are taxed and consumers are subsidized a cheap food policy. A negative PSE is interpreted as taxation of farmers rather than subsidization or protection. Within the WTO talks, PSE measurement assumes a critical place. A reduction in PSE-type support is the objective; and reductions in PSEs are used to monitor the progress of trade reform. Table 1 presents some calculated PSEs showing the extent of agricultural subsidization across a variety of countries.

Few developed countries choose not to subsidize agriculture, although, not surprisingly, New Zealand and Australia, as countries relying to a large extent on agricultural exports, come closest to this. New Zealand completely dismantled its agricultural subsidies over the period 1984 1987; following the removal of subsidies, sheep numbers, fertilizer use and pesticide use declined and there was an increase in afforestation. Land prices initially fell by 60% and fertilizer use declined by 50%. However, they soon recovered, and by 1995, land prices were back to 80% of their 1982 values in real terms. Land clearing and overstocking, which had led to widespread soil erosion, were much reduced and the number of full-time farm workers actually increased. Livestock production, formerly encouraged by subsidies, which had

Table 1 PSEs by country (1997)[a]

	PSE (%)
New Zealand	3
Australia	9
US	16
Hungary	16
Canada	20
Poland	22
Russia	26
EU	42
Japan	69
Norway	71

[a] *Source: OECD Secretariat Website, 1998: http://www.oecd.org.*

encroached on to erodible hillsides, is now being intensively practiced on better pasture land, and the hills are being replanted with trees. In 1995, agriculture and related industries accounted for 15% of New Zealand s gross domestic product (GDP) and agricultural products for 50% of New Zealand s exports. The EU and Japan still highly subsidize their farmers by paying them prices well above the comparable international trade prices; rice prices in Japan are some 8 10 times greater than world prices, and EU butter prices are much higher than those in New Zealand which largely determines the world price. The inflated price paid to Japanese rice farmers encourages them to safeguard their crops against any small risk of decreased yield. As a consequence, in 1993, the total rice insecticides market was $1.1 billion, of which 34% was spent in Japan, although Japan produces less than 3% of the world s rice. Subsidizing pesticide prices in Indonesia in the 1980s led to a large and unnecessary use of insecticides; when the subsidy was withdrawn and insecticide use was replaced by integrated pest management (IPM), rice yields continued to increase (Greenland, 1997). Inflated domestic prices can only be maintained by strict controls on international trade import tariffs and levies, quotas and non-tariff barriers; exports can only be sustained by facilitating export subsidies.

The calculated PSEs for other countries are less readily available, but for many African nations and central and eastern European economies in transition, they are probably negative. This means that farmers in these countries are, in effect, being taxed, not subsidized. Ivanova and Lingard (1994) calculated large negative percentage PSEs for a wide range of agricultural commodities in Bulgaria in the mid-1990s. This taxation results in low levels of profitability in Bulgarian agriculture, few incentives to use production inputs other than family labor, unsustainable soil nutrient management practices, obsolete non-functioning irrigation systems and a declining capital base for the production of vineyards, orchards and livestock breeding herds. A profitable agriculture would draw more resources to the sector, thereby ensuring that fewer resources are employed in less productive and more polluting industrial production. Within Africa, marginal land is often used for subsistence food production by squatting, poverty-stricken, landless laborers unable to find paid work. An increase in agricultural product prices to their comparable trade parity prices would increase employment on commercial farms and increase wages and incomes in rural areas. The taxation of agriculture is no real help to environmental improvement.

Agricultural support in the Organisation for Economic Co-operation and Development (OECD) member countries was estimated at $300 billion in 1996, some 1.3% of GDP, although this represented a decrease from 2.2% in 1986 1988 (OECD, 1998). Support to agricultural producers fell from 45% in 1986 1988 to 36% in 1996, as measured by the percentage PSE. The falls have, however, mainly arisen as a result of rapid GDP growth and high agricultural world prices rather than from deliberate policy reform on subsidy removal.

Policy reform driven by international trade pressures from the Uruguay round of the GATT, and subsequently by the WTO, is shifting agricultural support away from market price support and other support measures linked to farm input and output levels towards direct income support. Support which is decoupled from input and production levels does not have the built-in incentives to expand input use and production (and associated environmental externalities), but still aims to maintain farm incomes. Internationally, the policy instruments and levels of support vary widely by commodity and by country. However, market price support and support to inputs probably still account for over 60% of the total agricultural support worldwide. Agricultural subsidies thus stimulate high levels of input and natural resource use, wasteful production processes, and consequently, pollution. Reducing such subsidies is likely to reduce the costs of implementing environmental policy.

An *agri-environmental scheme* is a type of recent, new agricultural subsidy which is environmentally benign and helpful. These encourage land users to undertake agricultural practices with desired environmental results, rather than production enhancing outputs, such as public good provision (landscapes, flora and fauna habitats, biodiversity) or externality reduction. Support is coupled to the undertaking of particular environmentally beneficial farming practices to reward the farmer for the positive externalities that farming practices generate. Agri-environmental measures are being gradually introduced worldwide in developed countries in response to environmental pressures and political lobbying. Such schemes are, however, difficult to design, administer and monitor; frequently, they are merely a mechanism to compensate farmers for foregone market price subsidies. It is too early to determine what their long-term impacts on the environment will be and, in any case, their importance is dwarfed in comparison to market price support. For example, the UK spends approximately $300 million per annum on agri-environmental schemes, whilst the Common Agricultural Policy payments to UK farmers are in the order of $5 billion per annum.

Attempts to use subsidies to rectify market failures should follow the *polluter pays* principle, but rarely do. In terms of property rights, farmers are often assumed to have the right to pollute, and thus, have to be compensated for showing restraint. Likewise, governments have to compensate farmers for providing environmental goods in environmentally sensitive areas in the EU. This can be interpreted as the *polluter gets* rather than the *polluter pays*.

SUBSIDY REFORM AND THE ENVIRONMENT

Agricultural subsidies in rich countries are maintained by import barriers and export subsidies, whilst the agricultural policies of poorer countries tend to discourage farm production. The effects of these agricultural policies on the natural environment are poorly understood and under-researched despite numerous recent empirical trade liberalization studies. Based on the results from his commodity simulation model of world agricultural markets, Anderson (1992) argues that agricultural trade liberalization and the associated subsidy reductions could reduce global environmental damage from farming.

Liberalization would lead to falling agricultural prices in rich countries, resulting in less use being made of farm chemical inputs that pollute the air, soil and water. Chemical fertilizer applications and the use of farm pesticides are strongly correlated with producer price incentives. Lower use of irrigation water would reduce soil salinity problems; and reduced grain feeding of animals would reduce effluent disposal problems. Less intensive agricultural land use could result in more land being devoted to forestry, thus increasing the absorption of CO_2 from the earth s atmosphere.

Liberalization would, in turn, raise world agricultural prices and those in less developed countries (LDCs). This would make agricultural production in LDCs more profitable, increase the demand for farm labor and increase rural wages. Marginal workers could be attracted into the commercial sector, leaving behind subsistence practices on hillsides and resulting in less deforestation and soil degradation, particularly on sloping lands. The price of fuel wood might also rise due to increases in collecting costs causing the substitution of cleaner fuels, such as kerosene. Forests would thus be less depleted (80% of logs felled in LDCs are used as fuel). Higher royalty charges on logging and stronger enforcement of forestry property rights would help too. Optimal environmental policy instruments should be used in conjunction with the removal of distortions to agricultural output and input prices.

The WTO-related expansion of rice cultivation in non-Asian countries could lead to environmental problems in terms of water resources. Intensive cultivation of rice for exports would require expansion in irrigation and drainage infrastructures in Latin America and Africa. In designing such systems, particular attention will need to be given to the problems of water-induced paddy land degradation which have occurred in Asia, such as salinization, soil toxicity build-up and water logging.

Caution should, however, be exercised with respect to these conclusions. They are based on weak links between agricultural prices, production levels, input use, production technologies and environments. Differential population pressures at local, regional and national levels can substantially change, modify and even invalidate the findings. Further research is required.

CONCLUSIONS

Subsidy reform is now a central plank of the environmental policy and international trade agenda. International cooperation in reducing agricultural subsidies will facilitate support removal/reduction and reduce the negative side-effects that may accrue to any one country. Negotiating multilateral reduction of agricultural support which is common in many countries could lead to increased environmental benefits and reduced government expenditures, with little loss of competitive advantage in the agricultural sector of any one country. It is unclear whether the continued subsidy reduction and opening up of markets presents opportunities or threats to farmers in economically marginal, but ecologically valuable land, and what the impact on natural resource management might be. However, a simple message still remains:

> if you want to start saving the environment, stop financing its destruction by agricultural subsidization.

REFERENCES

Anderson, K (1992) Agricultural Trade Liberalization and the Environment: a Global Perspective, *The World Economy*, **15**(1), 153 171.

Cahill, C and Legg, W (1989) Estimation of Agricultural Assistance Using Producer and Consumer Subsidy Equivalents: Theory and Practice, *OECD Economic Studies*, **13**, 13 43.

Greenland, D J (1997) *The Sustainability of Rice Farming*, IRRI and CABI, Wallingford, UK.

Ivanova, N and Lingard, J (1994) Measuring the Effects of Government Transfers from Agriculture in Bulgaria: Calculation of Producer Subsidy Equivalents, *Oxford Agrarian Studies*, **22**(2) 123 137.

OECD (1998) Improving the Environment Through Reducing Subsidies, OECD, Paris.

FURTHER READING

Hampicke, U and Roth, D (2000) Costs of Land Use for Conservation in Central Europe and Future Agricultural Policy, *Int. J. Agric. Resources, Governance Ecol.*, **1**, 95 108.

Agriculture and Salinity

see Salinity and Agriculture (**Volume 3**)

Agriculture, Extensive: Cereal Cultivation

see Cereal Cultivation (Agriculture, Extensive) (**Volume 3**)

Agriculture, Intensive: Animal Production, Feedlots, and Manure Problems in the US

see Animal Production, Feedlots, and Manure Problems in the US (Agriculture, Intensive) (**Volume 3**)

Agriculture, Intensive: Greenhouse Food Production

see Greenhouse Food Production (Agriculture Intensive) (**Volume 3**)

Agriculture, Slash and Burn

see Swidden (**Volume 3**)

Agroecology

see Green Revolution (**Volume 3**)

Agroforestry

Annette Hladik

Eco-Anthropologie, CNRS and Muséum National d'Histoire Naturelle, Brunoy, France

The agroforestry approach combines the study of woody perennials, herbaceous plants, livestock and people, and their interactions with one another in farming and forest systems. It embraces an ecosystem focus considering the sustainability and equitability of land-use systems, in addition to their productivity; consideration of social as well as ecological and economic aspects is implied (Leakey and Newton, 1994).

Agroforestry is the name applied to a large diversity of traditional resource-use techniques and cultural patterns concerned with the management of trees since time immemorial mainly by people living within tropical countries. Agroforestry systems are dynamically linked to the long-term survival of the societies that practice them. Such agroforestry systems might be centered upon forests whose composition is deliberately manipulated. Alternatively they may comprise agricultural systems based on crop production of either annual or pluriannual cultivated plants, mixed in various patterns and cycles on the same land area, together with trees for multipurpose uses (shade, fruit crops, leaf fodder, timber, latex, resin or medicinal products . . . ;) and domestic animals. Around villages, and in small towns of many tropical countries, there exists an ecologically complex landscape made up of species-rich natural forests, and agroforests or community managed forests, alongside pasture lands, intensively cultivated permanent fields, and extensive rotational fields with fallows in different stages of succession. Within this mix of approaches, agroforestry systems show high biological resilience; they control soil erosion, promote nutrient cycling and maintain organic matter. This is a potentially flexible and reproductively viable long-term land-use mosaic provided that the biodiversity is maintained and even increased. It is socially and politically promising for the future development of sustainable agroecosystems.

Agroforestry is an emerging science dating from the 1970s, based on concepts adopted from various tropical regions where flexible land-use mosaics predominate. It links scientists from agronomy and forestry, two fields with research agendas that have always been conducted separately in the developed countries. Agroforestry grew out of new understandings coming from researchers working in tropical forest ecology and anthropology. They stressed the importance of analyzing the dynamic processes that underlie the functioning of forest ecosystems, based on the complex interrelationships between their numerous living species (mutualism instead of competition). They also considered the optimal production of such systems rather than just their maximal production, and paid attention to random events from many sources: climatic, biotic, economic, cultural, and familial.

MANAGED FORESTS AND COMPLEX AGROFORESTRY SYSTEMS

It is generally recognized today that there is no such thing as an untouched primeval forest. People have been present in all forests for many thousands of years. Forest biodiversity has coevolved with humans as well as with other biological species. Everywhere, the forest space has been managed; even before indigenous people practiced agriculture, they

gathered forest products from plants that were tended and protected; and they still do so.

At present, three million indigenous people live in the tropical forests of the central African zone (including about 120 000 Pygmies) and 8.45 million live in the Asian tropics (including about 180 000 hunter-gatherers). In Amazonia, where many more indigenous people lived in the past, today only about 700 000 Caboclos and Amerindians occupy this vast area (estimates are from Bahuchet, 1993). Population densities affect the way forests are managed. In the Brazilian Amazon, where the concept known as *extractivism* prevailed, the procurement of renewable resources used for sale was possible on a more or less sustainable basis because large areas of forests existed, allowing for the creation of extractivist reserves. Yet even here, renewable resources may be seriously over-used and extractivist techniques may threaten biological diversity. Cultivating the forest is the only promising alternative. This is precisely what happened in Java and Sumatra, which have had high population densities (per unit area) for a long time. Here, traditional communities have created complex agroforestry systems based on numerous cultivated tree species planted close to their irrigated paddy fields and dry rice fields in swiddens opened up in forests (*see* Agricultural Intensi cation in Java, Volume 3). Farmers, most often women, need species-rich forests where they could gather a multitude of essential products.

The term *community forest* is generally used by official organizations (for example, within the Food and Agriculture Organization (FAO)), to designate the traditional aspect of agroforestry systems that are managed by local people. But, the term community implies the idea of local ownership, whereas government officials insist that forests belong to the State. The interrelationships that exist between particular people and their forests are constantly being shaped and re-shaped by changing social and political circumstances. In countries where the state claims ownership, officials tolerate local people s use of their forest as usufruct (the right to use another person s property provided that the user does not damage or waste it). When farmers cultivate trees within agrosystems, they assert their ownership over the land. Generally, tree species are planted along the edges of fields, as a traditional way of delineating property boundaries. In some countries, in fact, efforts have been made to place the responsibility of delimiting community forests in the hands of the villagers themselves. This is not an easy process to decree by laws.

In recognizing that diversity (be it biological or cultural) constitutes a fundamental value and a key concept for insuring future sustainability, it is important that it be extended to include land use and land rights. There are various types of ownership even within such categories as private versus public land. There are also diverse forms of common ownership: land may be the property of an entire village, or only of certain descent groups, designated households or family groups. Moreover, within traditional communities, land rights might not be coterminous with rights over the products obtained from the land. Access to forest products is often subject to seasonal rules, even for the owner. In addition, some part of the forest may be set aside for special (ritual, social) use and thus be out of bounds for some community members.

Multi-purpose, traditional agroforests yield products not only for subsistence but also for commercial purposes. Households need additional income to fulfil obligations linked with the modern world (schooling, taxes, clothing and implements). But the problem with commercial production is the absence of infrastructures, mainly transportation facilities to take the products to the market. Moreover, market forces require the flexibility to be able to balance supply and demand particularly because profit margins may be low. The diversity of agroforestry outputs may be either augmented or reduced by consumer choices. Preferences for particular food products are subject to cultural preferences that are often highly arbitrary. There is an urgent need to create and develop new markets for new products. Clearly, the potential productivity and income-generating capacity of agroforestry systems are immense, but the constraints surrounding them are often insurmountable in developing countries.

AGROFORESTRY: TRADITIONAL KNOWLEDGE

The value of a holistic approach, including traditional knowledge about forest ecosystems, is now amply recognized. The term *tradition* does not necessarily imply past and obsolete; it may also suggest modern and promising ways of incorporating change in the modes of thought and behaviors followed by a people throughout the generations. In fact, continuity is a dynamic idea; it accommodates new adaptations and embraces new ways of doing things. As such, it is the opposite of a conservative view. The ability of traditional peoples to invent and to incorporate new methods, new tools, and new species into their resources practices leads to an incredible diversity of traditional resource-use patterns. The survival of small-scale, indigenous communities over generations in their forest habitats is evidence of their success. Indigenous knowledge about forests is thus an essential component of traditional ecological knowledge (TEK).

AGROFORESTRY RESEARCH

History and Literature

One of the first scientific papers clarifying the concept of agroforestry appeared in 1977 (Bene *et al.*, 1977) at

the International Development Research Center (IDRC), Canada. It is a summary of discussions, which took place among experts concerning the relationship and balance between sustainable forest management and agricultural food production. Shortly after, in 1978, the International Council for Research in Agroforestry (ICRAF) was created in Nairobi. At the same time, a new scientific journal appeared entitled *Agroforestry Systems*. Organized at different scales, either national, regional or international, several symposia took place in various countries, including the regional workshop convened in Makokou, Gabon, in the context of *Man and the Biosphere* (MAB)/United Nations Educational, Scientific and Cultural Organizations (UNESCO) on the potential contribution of agroforestry systems to conservation and use (Maldague *et al.*, 1986). Several newsletters are now available, among them *Forests, Trees and People Newsletter* (FTPP/FAO) and *Agroforestry Today* (ICRAF). Bibliographical records containing useful references to agroforestry approaches and practices have also been compiled. Moreover several important books on the subject of agroforestry have been recently published; they are listed at the end of this article.

Universities and research centers nowadays regularly offer programs in agroforestry concepts and management. The subject has become an essential component of an academic programs in the agricultural, sylvicultural and related sciences.

System Analysis

An extensive survey and analysis of traditional agroforestry diversity and functioning has confirmed that such systems are generally complex, reproducing the multi-structure organization of a natural forest (Figure 1). An analysis of this complexity indicates that it is not only created by the manipulation of the plants themselves but also by the modification of the environment where they grow through irrigation, exposure to light, composting and seed selection. One of the most threatening factors requiring careful management is natural pest control through the use of a diversity of plant species and varieties. An integrated approach to the *in situ* amelioration of plant genetic material and the conditions under which different varieties do best, creates a veritable gardened forest. Sometimes, agroforests are entirely reconstructed after a field has been harvested by planting selected species (producing rubber, damar resin, cinnamon, nutmeg, clover and durians) in association with spontaneously regenerating native species (De Foresta and Michon, 1993; see also Figure 2). The complementarity in space and time, and the production of numerous selected, as well as spontaneous, species are the primary advantages of these gardened forests.

In the last few years, foresters from the FAO have placed great importance on trees out of the forest. For FAO scientists there are only natural forests, plus forest plantations; they assume that swidden cultivation leads to deforestation. In fact, a field that has been cleared and cultivated for two or three years will revert to various stages of regenerated forest. One can recognize first regrowth, middle regrowth, young forest and so on, until a mature forest evolves. All these stages and more are specifically named in the local languages. Unfortunately, with increasing population pressure around villages, long forest fallows are often shortened and the forest has a reduced chance to recover enough of its diversity (*see* Shifting Cultivation and Land Degradation, Volume 3). Satellite imagery and geographical information systems (GIS) linking spatial geographical data to events occurring at ground level can be used to obtain an approximate determination of the extent to which an area has been deforested. These essential tools for any type of natural forest management, including conservation, are not without their problems. There may be selective logging and other destructive activities occurring under the forest canopy that cannot be detected by satellite imagery. This is why ground truth is an essential component of all GIS analysis.

Experimenting with New Agroforestry Systems

The new agroforestry research is based on an interdisciplinary approach integrating studies concerning soils, plants (mixed cultivated plant species: annual and perennial crop, herbs and ligneous plants, trees and vines or lianas) with that of animals (cattle and poultry). Research into how human societies manage these plant animal interactions requires not only technical knowledge but also economic and cultural information.

When it comes to testing associated mixed plant species, experimental agroforestry systems are highly simplified when compared with most traditional systems (Figure 3). For example, the alley cropping system promoted by various governmental agencies and non-government organizations (NGOs) involves one crop species (even one variety) plus one shade-tree species, planted in lines. Moreover, alley cropping and hedgerow intercropping do not eliminate the need for chemical fertilizer. Thus, in an effort to test and control for variables, technicians have greatly reduced the complexity of experimental systems and made them more dependent upon outside inputs. An especially problematic case is that of the Taunguya system where peasants are encouraged by foresters to plant trees in lines and grow their crops in the spaces between. This may be a good reforestation procedure but it is not adapted to peoples long-term use of the land. There is always a risk of uncertainty in the creation of artificial systems whose species composition and interactions have not been tested through time. By leaving out the human factor, one decreases the ability of individuals themselves to learn and cope with production constraints.

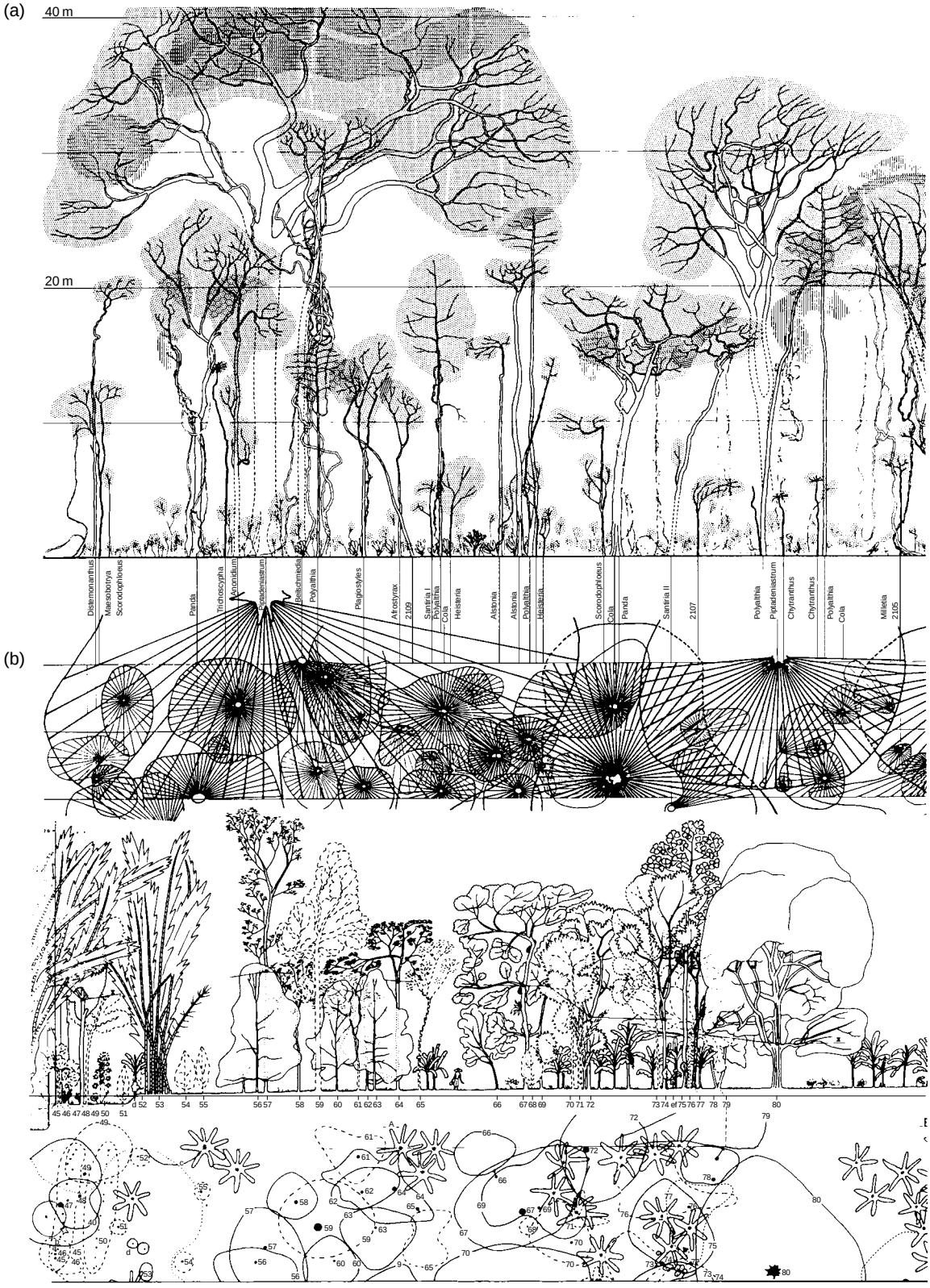


Figure 1 The multi-structure organization of a complex man-made agrosystem, as shown in (b), in Java where banana plants are mixed with forest trees (plot of 10 × 60 m) compared to a species-rich rain forest (as shown in (a)) in Africa (same plot 10 × 60 m with the profile only 5 m wide). (Reproduced from Hladik and Hladik, 1984)

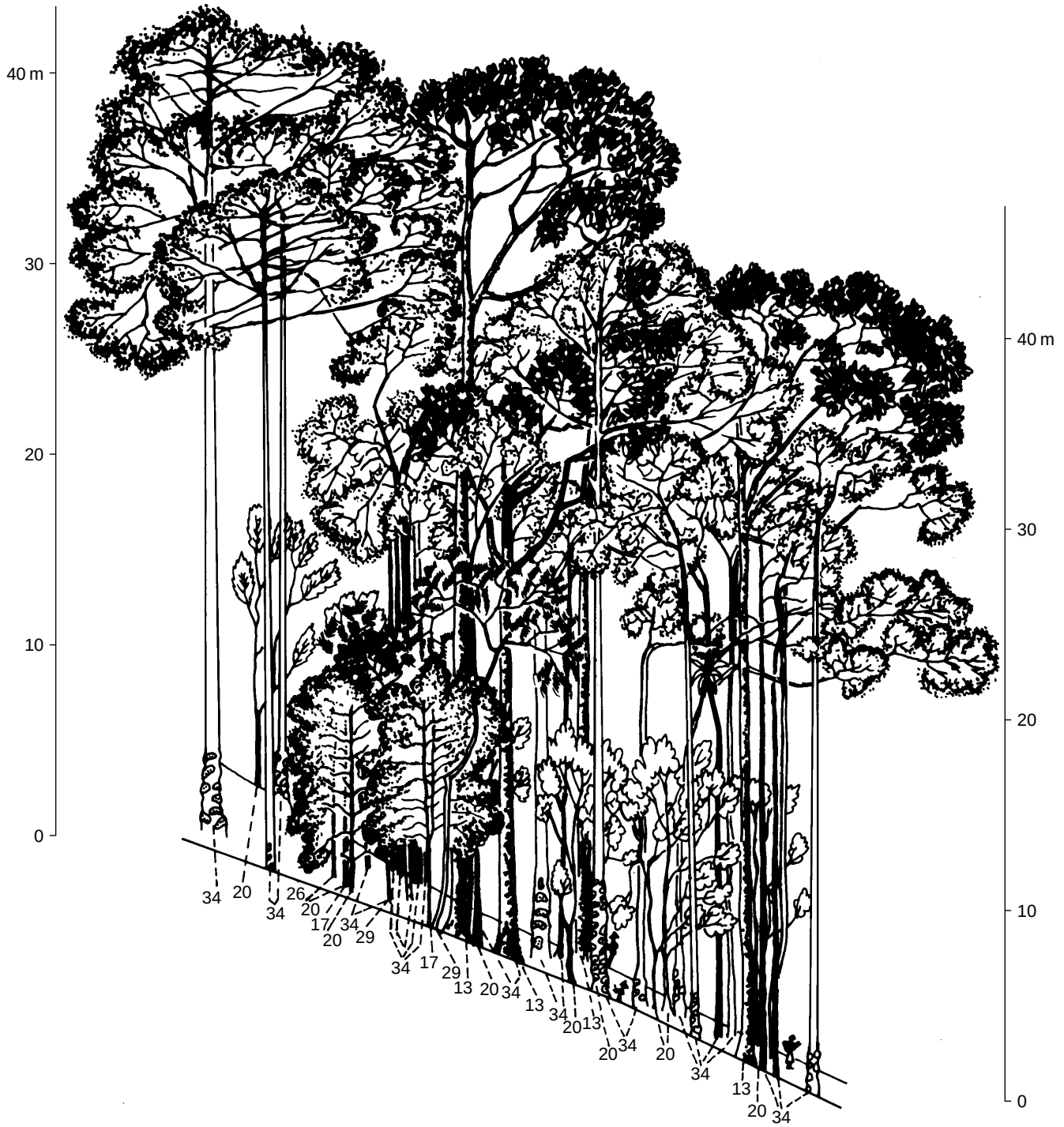


Figure 2 Architectural profile of a damar agroforest (20 × 30 m) in Sumatra where the damar trees are associated with fruit trees (durian, rambutan, mangustan, etc.). (Reproduced from De Foresta and Michon, 1993)

ICRAF has been promoting the use of powerful nitrogen legume species that also help in the recovery of degraded soils. Where there has been a surge of renewed experimental systems, effort is concentrated on the problem of improving fallow periods. Across the tropical world, most agroforestry experiments use the same limited number of exotic legume species. This involves the planting of soil-enriching legume species such as the well-known *Leucaena leucoceplala* during the early stages of fallow periods. But such exotic single species like *Leucaena leucoceplala* can

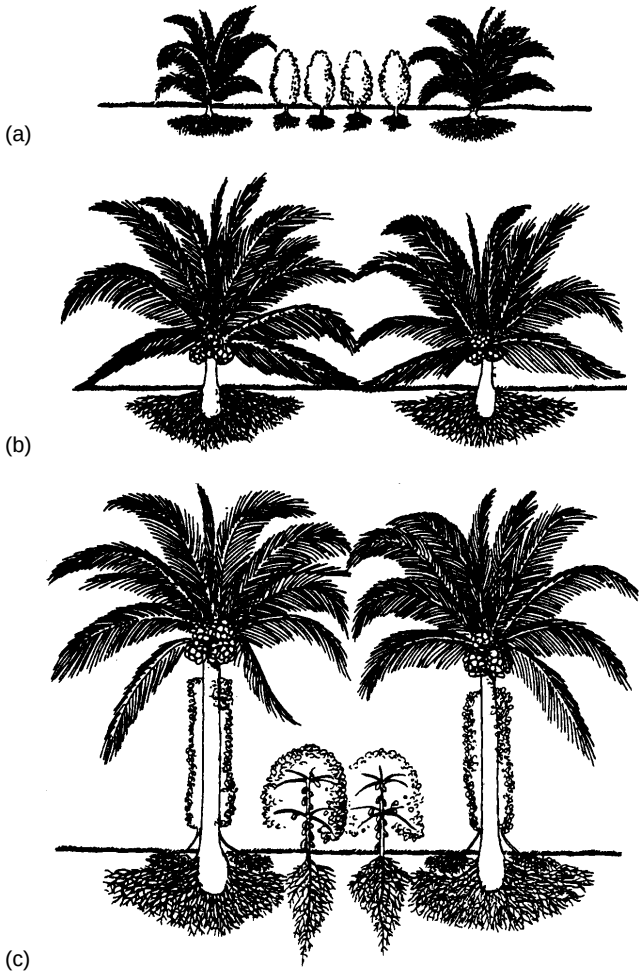


Figure 3 Schematic presentation of the growth phases of coconut palm indicating possibilities for crop combinations. (Reproduced from Nair, 1983). (a) Early phase, up to about 8 years; (b) middle phase, about 8–25 years, creating less than ideal conditions for underplanting; (c) later phase in which sufficient light enters the under-story to make conditions again suitable for a multi-storied combination of coconut, black pepper and cacao plantations

easily become invasive if not carefully controlled (its seeds from dehiscent (self-opening) pods have to be harvested each time). This is why it is important to use a variety of legume species, preferentially native ones that have the added value of contributing other useful products.

Additionally, experimental scientists in agroforestry have been concerned with the genetic improvement and vegetative propagation of an increasing number of indigenous forest tree species, well adapted to the local environment and not yet domesticated. A specific problem in breeding a potentially multi-purpose tree is that a single trait such as crop fruit production is usually selected for improvement whereas other qualities (shade, leaves for fodder, timber) may be ignored.

TEMPERATE AGROFORESTRY SYSTEMS

Finally, it should be pointed out that agroforestry systems are not confined to tropical regions but may also exist in temperate zones where long-term uses of trees have been incorporated into successful agrosystems. Examples are the sugar maple-based system in Southern Canada and Northeastern United States, and the modern Paulownia-wheat intercropping systems of China. Farmers have practiced agroforestry for a long time, together with sylvopastoral systems, forestry grazing, windbreaks or orchard intercropping. In temperate industrial countries, modern agriculture has experienced crises during which traditional cultures were homogenized and landscape was severely modified. The Chinese agricultural landscape also has numerous areas of monoculture, but always interspersed with mixed cropping areas, forming a highly complex mosaic. Additionally, this system is strongly linked to the cultural traditions that have supported China for millennia and that remain strong today among rural populations.

AGROFORESTRY AND GLOBAL CLIMATE CHANGE

Agroforesters have to cope with climate variations (seasons and year-to-year changes); they use numerous and multi-purposed plant species (and varieties). If 21st century climate is going to warm, advantages of the agroforestry systems, including biodiversity and the traditional practices that the local people utilize, will help them cope. However, scientists must be careful to keep enough genetic diversity within cultivars (especially in long-lived trees) when introducing improved selected plants in modern experimental agroforestry.

CONCLUSION

Agroforestry is a relatively new science and research is advancing strongly in developing countries, along with organic farming. Future research and development of agroforestry systems must include agronomic, environmental and economic aspects. It is no longer adequate to demonstrate biophysical increases in productivity without addressing environmental parameters within the socio-economic context of the people concerned. It should be admitted, however, that agroforestry is not a panacea for solving all problems. Agroforestry should be carried out in conjunction with other land-use systems in a mosaic of diverse spatial and temporal practices. It is the total economic value of the forest, natural or managed, that has to be kept in mind. Agroforestry is not a question of production in the abstract, but enriching the potential welfare of human beings around the world.

REFERENCES

Bahuchet, S (1994) *Situation des Populations Indigènes des Forêts Denses Humides*, Office des Publications Officielles des Communautés Européennes (CECA-CE-CEEA), Luxembourg.

Bene, J G, Beall, H W, and Coté, A (1977) *Tree, Food and People: Land Management in the Tropics*, IDRC, 084f, Ottawa, Canada.

De Foresta, H and Michon, G (1993) Creation and Management of Rural Agroforests in Indonesia: Potential Applications in Africa, in *Tropical Forests, People and Food: Biocultural Interactions and Applications to Development*, eds C M Hladik, A Hladik, O F Linares, H Pagezy, A Semple, and M Hadley, MAB Series, 13, Unesco and Parthenon, Paris/Carnforth, 709 724.

Hladik, C M and Hladik, A (1984) L Agroforesterie: Science et Technique d avenir en Afrique Noire, *Le Courrier du CNRS*, 58, 40 43.

Leakey, R R B and Newton, A C (1994) *Domestication of Tropical Trees for Timber and Non Timber Products*, MAB Digest Series, 17, UNESCO, Paris.

Maldague, M, Hladik, A, and Posso, P (1986) Agroforesterie en Zones Forestières Humides d Afrique, Rapport du Séminaire sous-régional, Juillet 1985, UNESCO, Paris.

Nair, P K R (1983) Agroforestry with Coconuts and other Tropical Plantation Crops, in *Plant Research and Agroforestry*, ed P A Huxley, ICRAF, Nairobi, 79 102.

FURTHER READING

Buck, L E, Lassoie, J P, and Fernandes, E C (1998) *Agroforestry in Sustainable Agricultural Systems*, Lewis Publishers, Boca-Raton, FL.

Gordon, A M and Newman, S M (1997) *Temperate Agroforestry Systems*, CAB International, Oxford.

Hladik, A (1986) Perspectives de Développement par l Agroforesterie, in *Bien Manger et Bien Vivre. Anthropologie Alimentaire et Développement en Afrique Intertropicale: du Biologique au Social*, eds A Froment, I Garine de, Binam-bikoi Ch, and J F Loung, L Harmattan-ORSTOM, Paris, 477 487.

Huxley, P A (1999) *Tropical Agroforestry*, Blackwell Science, Oxford.

Young, A (1998) *Agroforestry for Soil Management*, 2nd edition, CAB International, Oxford in association with ICRAF, Nairobi.

Aircraft Emissions

Ulrich Schumann

Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

Aircraft emit gases and particles directly into the upper troposphere and lower stratosphere, mainly in Northern Hemisphere mid-latitudes, where they have an impact on atmospheric composition. These gases and particles alter the concentration of atmospheric greenhouse gases, including carbon dioxide (CO_2), water vapor (H_2O), ozone (O_3), and methane (CH_4). They may trigger the formation of condensation trails (contrails), increase cirrus cover and change other cloud properties, all of which affect the energy and water budgets of the atmosphere and hence may contribute to climate change at the regional and global scale.

The disturbances induced by global aviation cause an additional radiative forcing (heating) of the Earth- atmosphere system by aircraft of about 0.05 $W m^{-2}$ or about 3.5% of the total radiative forcing by all anthropogenic activities in 1992. The values are presently increasing both in absolute and relative terms. Nitrogen oxides (NO_x) emissions from current aircraft are calculated to have increased O_3 by about 6% in the region 30- 60 °N latitude and 9- 13 km altitude. Calculated changes in the total column of O_3 in this latitude range are approximately 0.4%. Calculated effects are substantially smaller outside this region. Emissions of NO_x into the stratosphere above about 20 km might cause a reduction of O_3. An order 1% reduction of O_3 in the stratosphere and a corresponding increase of ultraviolet (UV)-B radiation at the surface may occur if a large (1000 aircraft) eet of supersonic aircraft were to become operational.

Aircraft emissions near airports contribute to local air pollution. The emitted nitrogen oxides (NO_x) reduce the ozone (O_3) concentration in the immediate neighbourhood of the airport and, together with the emitted hydrocarbons, may induce additional O_3 several 10 km downstream from the airport by photochemical smog reactions. Outside the immediate neighbourhood of airports, in regions with high ground traffic and high population density, the emissions and the resultant smog are dominated by other forms of traffic such as motor vehicles, or by industry and domestic emissions. Noise induced by aircraft engines and aircraft structures during take-off and landing is often considered as a significant environmental problem.

TRAFFIC

Aviation is an integral part of the infrastructure of today s society. It plays a vital role for global commerce and private travel. Air traffic has grown strongly in recent decades, see Figure 1, faster than the economy as a whole. In 1997, 13 489 jet, 3213 turbo-prop, and 291 piston-engine commercial aircraft carried 1457×10^6 revenue passengers worldwide, on average over 1766 km distance per flight. The number of jet aircraft increased by 4.8% $year^{-1}$ from 1981 to 1997, turbo-props by 2.2% $year^{-1}$, while the number of aircraft with piston-engines is decreasing. The number of

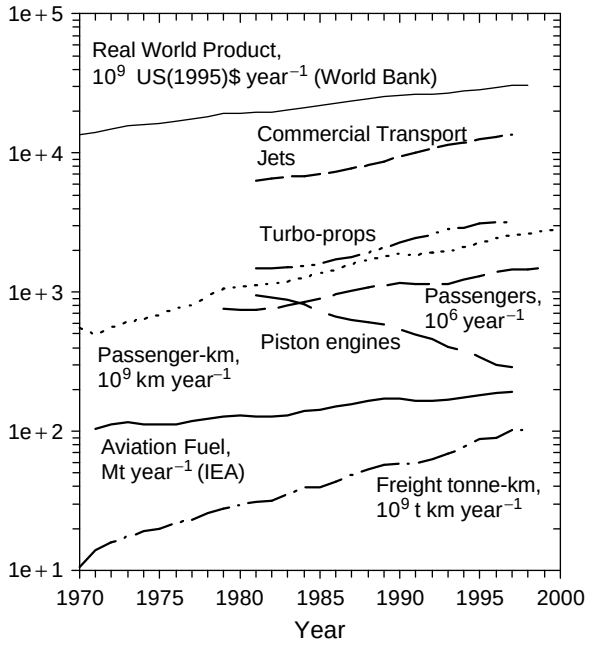


Figure 1 World aviation in the years 1970–1999: civil commercial transport aircraft (three types) registered in International Civil Aviation Organization, ICAO states, and traffic of commercial air carriers as reported by the ICAO (Montreal), aviation fuel production as reported by the International Energy Agency (IEA, Paris), together with the real world economic product (Worldbank, Washington, DC)

revenue passengers increased by 4.2% year^{-1} from 1981 to 1997. Passenger traf c increased by 5.3% year^{-1} to 2573 $\times$ 10^9 year^{-1} revenue passenger-kilometres, and freight traf- c by 7.8% year^{-1} to 103 $\times$ 10^9 freight tonne-kilometres per year in the same period. For comparison, the world economic output (gross national product, GNP, in market prices of 1995) grew by 2.4% annually on average in the years 1991 to 1998 (World Bank, Washington, DC, March 2000).

FUEL CONSUMPTION

Global air traf c consumed aviation fuels at a rate of 130 to 170 Mt year^{-1} during the year 1992 (i.e., about 5 6% of all petrol products), including a military fraction of about 18% (Intergovernmental Panel on Climate Change (IPCC), 1999). The upper bound fuel consumption value is the aviation fuel production reported by the IEA (International Energy Agency), the lower bound results from analysis of air traf c, and aircraft/engine speci c fuel consumption estimates. About 65% of the fuel is consumed at cruise altitudes between 10 and 13 km, see Figure 2. The largest fraction is consumed by wide-body aircraft on long-distance ights. Most of the emissions occur between 30° and 55 °N over the USA, Europe, and the North Atlantic. Globally the fraction of fuel burnt above the tropopause has been estimated to be about 30%. Over the North Atlantic the stratospheric fraction of fuel consumption is about 50% of the annual mean, with larger values during winter. The noon/midnight ratio of air traf c (in terms of fuel consumption) amounts to about 3 globally. Aviation fuel production grew by about 2.6% annually from 1981 to 1997, and was estimated to be about 200 Mt year^{-1} in the year 2000. The fuel consumption has grown at a slower rate than the traf c because of improved aircraft (engine and frame) technology, increased load factors of the aircraft, and a reduced fraction of fuel consumption for military purposes.

For the future, global passenger air travel, as measured in revenue passenger-km, is projected to grow by about 5 6% year^{-1} between 1990 and 2015, whereas total aviation fuel use, including passenger, freight, and military, is projected to increase by 3 4% year^{-1}, over the same period. The military fraction is expected to decrease to 7% in 2015.

EMISSIONS

In burning kerosene (a hydrocarbon mixture with about 13.8% hydrogen mass fraction) with air, engines emit mainly the greenhouse gases carbon dioxide (CO_2) and water vapor (H_2O) (Table 1). Minor emissions formed during combustion in the engine include nitric oxide (NO) and nitrogen dioxide (NO_2) (which together are termed NO_x, the emissions being measured in terms of mass units of NO_2), hydrocarbons (H_xC_y), carbon monoxide (CO), and soot. Soot includes around 10^{15} particles kg^{-1} of burnt fuel with typical diameters of 10 30 nm (Karcher, 1999). Kerosene contains about 0.8 (0.001 3) g sulphur kg^{-1} of fuel. During the combustion in the engines, this sulphur is converted to sulphur oxides (SO_x), mainly sulphur dioxide (SO_2), but partly into sulphur-trioxide (SO_3) and after some cooling and with H_2O, into sulphuric acid (H_2SO_4). The conversion fraction of fuel sulphur to H_2SO_4 in the young plume is in the range of 0.4% to about 10%.

Aviation contributes about 1.6 to 2.2% to the global anthropogenic CO_2 emissions of about 7000 Mt C year^{-1}, 10 to 13% of traf c-originating CO_2, and 2% of all NO_x sources. Other NO_x sources (in units of Mt NO_2 year^{-1}) include biomass burning (17.5), industry and surface traf c (72), microbial activities in soil (11.7), lightning (16.4), and stratospheric sources (2.1; Lee *et al.*, 1997). Aircraft emission amounts of CO and hydrocarbons are much smaller than other anthropogenic emissions and of little importance for air chemistry outside airports. The methane (CH_4) concentration at engine exit may be smaller than in the ambient air. Total aviation NO_x emissions increased faster than fuel consumption over recent decades because of higher

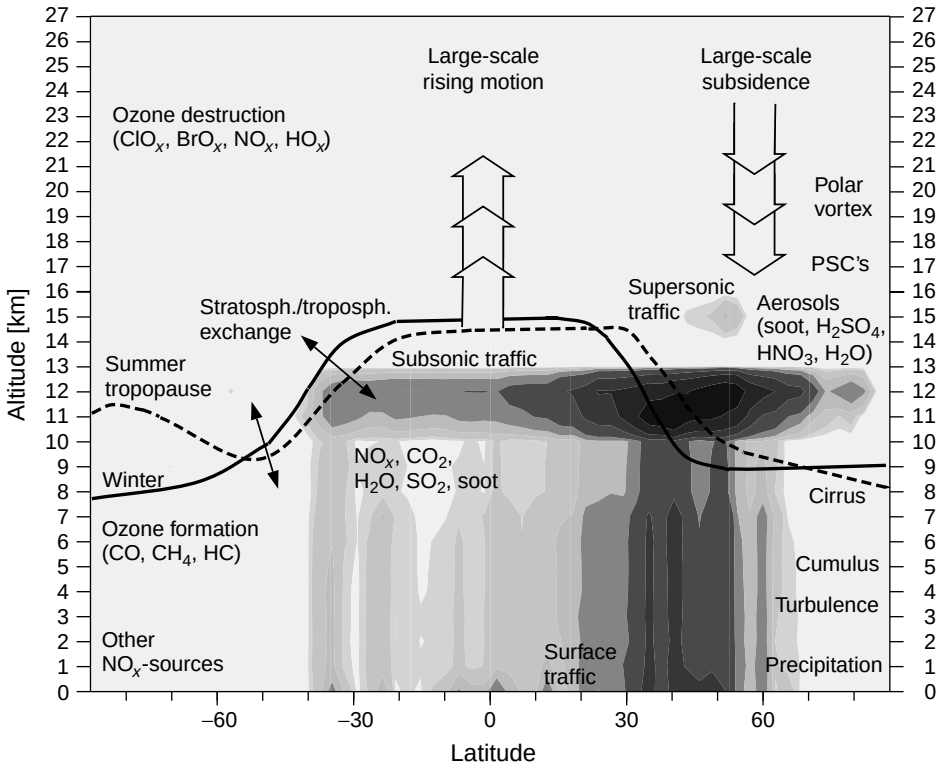


Figure 2 Mean altitude of the tropopause in June and December, and distribution of NO_x emission source rates from aircraft versus altitude and latitude (the emission rate is the larger the darker the shaded area), and indications of mean circulation and some relevant processes

Table 1 Consumed or emitted species, mean emission indices, i.e., mass of emissions per unit mass of burned fuel, for the fleet of aircraft in 1992, total emission rates due to aviation, and comparable emission rates. (Reproduced by permission of Intergovernmental Panel on Climate Change (IPCC), 1999)

Species	Emission index, g kg^{-1} (ranges)	Emission rate (1992) in Mt year^{-1}	Comparable emission rate, Mt year^{-1}	Comparable emission source
Kerosene		140 (139–170)	3140	Total petrol production
CO_2	3160	440 (430–540)	26 000–30 000	Total anthropogenic CO_2 emissions
H_2O	1230	176	45	Methane oxidation in the stratosphere
			525 000	Evaporation of H_2O from Earth's surface
NO_x	14 (12–16)	2	2.9 ± 1.4	Stratospheric sources
			90 ± 35	Total anthropogenic source
Soot	0.04 (0.01–0.1)	0.006	12	Combustion of fossil fuels and biomass
SO_2	0.8 (0.6–1)	0.12	130	Total source from burning fossil fuels
			20–100	Natural sources
			5.4, 8.0	Non-eruptive, eruptive volcanoes
CO	4 (2–6)	0.56	1500	Total anthropogenic sources
H_xC_y	0.6 (0.2–3.0)	0.2	90	Total anthropogenic sources

combustion temperature and pressure in more fuel-efficient modern engines. Other types of emissions decreased per unit of fuel consumption.

CHANGES IN AIR COMPOSITION

The gases and particles emitted by aircraft accumulate in the atmosphere near the flight routes, mainly in the northern mid-latitudes, depending on their residence time. Emissions into the lowermost stratosphere (see Figure 2), just above the tropopause get mixed mainly poleward and downward by the mean circulation, and leave the stratosphere by stratosphere-troposphere exchange processes (see Figure 2). Lower-stratospheric inert emissions reside there for an order of weeks to months. Some of the emissions have shorter residence times because of chemical conversions. For example, NO_x near the tropopause gets converted to nitric acid (HNO_3) within a few days or weeks. The NO_x concentration in the upper troposphere is about 1000 times lower than in urban regions, and the residence time for NO_x emissions near the ground is about ten times smaller than near the tropopause. Therefore, the relatively small amounts of aircraft emissions have notable effects on the NO_x concentration near the tropopause. Accumulations of aircraft emissions of NO_x are measurable regionally near main traffic corridors at least under low wind conditions (Schumann *et al.*, 2000). Increases of the concentration of small particles emitted from aircraft with similar residence times have also been measured near dense flight routes. CO_2 on the other hand, has a lifetime of the order of 100 years and gets distributed essentially over the whole atmosphere. Therefore, the effects of CO_2 emissions from aircraft are indistinguishable from the same quantity of CO_2 emitted at the same time by any other source.

The gases and particles emitted or formed as a result of aviation have an impact on climate both directly and indirectly. The direct effect is due to absorption and scattering of radiation. Indirect effects are many. They result from chemical or physical effects of the aircraft emissions on the gases and clouds, which act as greenhouse gases or radiative scatterers or influence clouds and precipitation, and influence the energy budget and the hydrological cycle on Earth.

CO_2

Concentrations of and radiative forcing from CO_2 today are those resulting from all anthropogenic emissions during the last 150 years. The atmospheric CO_2 concentration increased by about $80\,\mu mol\,mol^{-1}$ since 1850, and is responsible for a radiative forcing of $1.6\,W\,m^{-2}$. For an order of magnitude estimate, one assumes that $1\,W\,m^{-2}$ radiative forcing causes a global change in surface temperature of the order 0.3 1.0 K.

Aviation caused less than $1.4\,\mu mol\,mol^{-1}$, i.e., 1.2 1.7% of the total increase (Sausen and Schumann, 2000). This percentage is lower than the percentage for emissions (<2.4%) because aviation emissions have occurred mainly in only the last 40 years. The related CO_2 radiative forcing due to aviation fuel consumption in the past is $0.02\,W\,m^{-2}$. The corresponding increase in surface temperature is too small to be observable but is computed to be about 0.004 K. For a scenario used in Intergovernmental Panel on Climate Change (IPCC) (1999) with 3.2-fold increase in aviation fuel consumption from 1992 to 2050, one computes a CO_2-concentration increase of about $6\,\mu mol\,mol^{-1}$, causing a temperature and sea level increase of 0.025 K and 3 mm (Sausen and Schumann, 2000). Note that these are scenario calculations covering CO_2 effects only.

O_3

O_3 is a greenhouse gas, which is particularly effective within the tropopause region because of low temperature in this region and low thermal radiation emission at low temperatures. O_3 also shields the surface of the Earth from harmful UV radiation. This shielding increases to first approximation with the column of O_3 (most O_3 resides in the stratosphere). O_3 also acts as a biologically effective air pollutant at the Earth s surface.

O_3 in the stratosphere might get reduced when NO_x concentrations increase there. The photocatalytic cycles for NO_x in the stratosphere are similar to those by which halogen compounds (ClO_x and BrO_x) and hydroxyls (HO_x) cause catalytic O_3 destruction. However, NO_x also binds other O_3-destroying radicals and hence, additional NO_x reduces O_3 only above altitudes of about 20 km. In the troposphere and lower stratosphere, other photocatalytic cycles dominate, which cause oxidation of CO and CH_4 and other hydrocarbons, and cause O_3 formation at a rate which increases with the NO_x concentration at least up to a certain maximum of NO_x concentration (Brasseur *et al.*, 1998). The O_3 formation in the upper troposphere is controlled by the concentration of NO_x, but depends also on the availability of CO, hydrocarbons, and gases that provide sources for HO_x radicals such as O_3, H_2O and acetone. Hence, NO_x emissions from the present fleet of subsonic aircraft, flying essentially below 13 km altitude, cause O_3 increases, while the same emissions from a potential future fleet of supersonic aircraft, flying above 16 km, could cause O_3 destruction. Because of the much larger time scales of photochemistry of O_3 (on the order of weeks), the O_3 concentration change is spread more smoothly over the zonal belt downwind of and adjacent to the main flight corridors. The O_3 change is also notable at far distances from the corridors. Aircraft emissions of NO_x are more effective in producing O_3 in the upper troposphere than an equivalent amount of emission at the

surface. Also increases in O_3 in the upper troposphere are more effective in increasing radiative forcing than increases at lower altitudes.

The NO_x emissions from subsonic aircraft in 1992 are estimated to have increased O_3 concentrations at cruise altitudes in northern mid-latitudes by up to 6%, compared to an atmosphere without aircraft emissions (Intergovernmental Panel on Climate Change (IPCC), 1999). This change in O_3 concentration is of similar magnitude as the natural year-to-year variability. The impact on O_3 concentrations in other regions of the world is substantially less. These increases, on average, tend to warm the surface of the Earth. Computational studies with a global climate model show that aircraft-induced O_3 changes cause temperature changes, which are large enough regionally (order 0.2 K at some latitudes) to be distinguishable from climatic noise at the time scale of a few decades. It seems that aviation induced O_3 is particularly effective in heating the troposphere (Ponater *et al.*, 1999).

CH_4

CH_4 is also a greenhouse gas. Direct contributions from aircraft emissions to methane are unimportant, but NO_x increases in the upper troposphere lead indirectly to reductions of CH_4. The additional NO_x from aircraft enhances the oxidation cycle of CO, which causes local increases in OH radicals (*see* **OH Radical: is the Cleansing Capacity of the Atmosphere Changing?**, Volume 2). Such radicals are short-lived and react with CO both in warm and cold air masses; they react also with CH_4 but essentially only in warm air. The induced OH radicals in the cold upper troposphere reduce the concentration of CO at ight levels. CO has a lifetime of about two months, longer than typical vertical transport times in the atmosphere. Therefore, the CO reduction is felt also near the surface. Reduced CO abundance near the surface leaves more OH radicals from other sources to react there, in warm air masses, with CH_4. Hence, aircraft NO_x emissions indirectly cause a reduction of CH_4. The CH_4 concentration in 1992 is estimated to be about 2% less than that in an atmosphere without aircraft (Intergovernmental Panel on Climate Change (IPCC), 1999). This aircraft-induced reduction of CH_4 concentration is much smaller than the observed overall 2.5-fold increase since pre-industrial times. The reductions in CH_4 tend to cool the surface of the Earth.

H_2O

Water vapor is a greenhouse gas and of chemical importance. The amount of H_2O emitted by aviation is much smaller than the amount of water evaporating at the Earth s surface but comparable in magnitude to the amount of water resulting from CH_4 oxidation in the stratosphere (Table 1). Most subsonic aircraft H_2O emissions are released in the troposphere where they are removed by precipitation within 1 2 weeks. Some fraction of H_2O emissions is released in the lower stratosphere where it can build up to larger concentrations. Because H_2O is a greenhouse gas, these increases tend to warm the Earth s surface. Additional H_2O in the stratosphere may also induce hydroxyl radicals and reduce O_3 either directly or indirectly by forming polar stratospheric clouds, which cause halogen activation. However, for subsonic aircraft, these effects are small compared to the other climatic or chemical effects of aircraft emissions (Intergovernmental Panel on Climate Change (IPCC), 1999).

SULPHATE AND SOOT AEROSOLS

Besides soot, large numbers (about $2 \times 10^{17}\,kg^{-1}$ fuel) of small (radius 1 5 nm) liquid particles are formed in the exhaust plumes of cruising aircraft. These new particles initially form from H_2SO_4, chemi-ions, condensable hydrocarbons, and H_2O; they grow in size by coagulation with other particles and by uptake of H_2O and other condensable gases. Chemi-ions are charged molecule clusters formed during combustion. Charged particles enhance the coagulation ef ciency (Karcher, 1999). The fuel sulphur content has little in uence on the conditions for contrail formation but has strong impact on the size of newly formed particles. The particles from aviation add to the atmospheric aerosol and impact cloud formation. The number of ice particles formed in the young contrail varies by a factor of about 3 for a factor 500 change in fuel sulphur content.

Subsonic aircraft emissions near the tropopause at northern mid-latitudes are a signi cant source of soot and sulphate aerosol concentrations, at least by number. Aircraft generate far less aerosol mass than that emitted and produced at the Earth s surface or by strong volcanic eruptions (Table 1). Aircraft emissions injected directly at 9 12 km altitudes are more important than similar surface emissions because of longer atmospheric residence times and lower background concentrations of pollutants in the upper troposphere. The direct radiative forcing by the accumulated aerosol is small compared to those of aircraft induced gases. The chemical effects of soot and other particles from aviation on O_3 by means of heterogeneous chemistry, and the physical effects of aerosol on cloud properties are not well known, but may be important.

CONTRAILS

Contrails are visible line clouds, see Figure 3, which form from H_2O emitted by aircraft ying in suf ciently cold air. Contrail formation, according to the Schmidt-Appleman

Figure 3 Contrails of various ages at high ambient humidity. (Photo: U. Schumann)

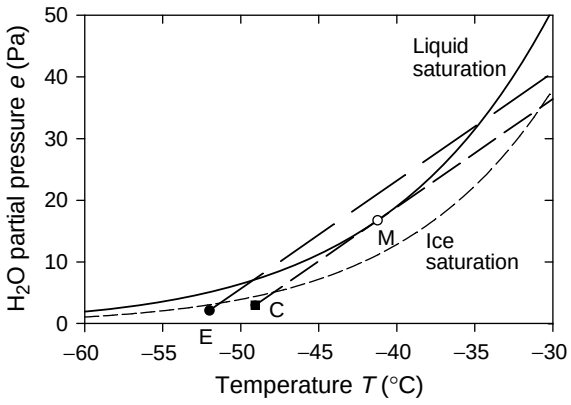


Figure 4 Mixing lines (dashed lines) and saturation curve over liquid water (full) and ice (dashed) in a diagram of partial water vapor pressure e versus temperature T. The mixing lines are plotted for environmental conditions with temperature T in the environment below (point E) and at (point C) the threshold temperature. At point M liquid saturation is reached during mixing under threshold conditions. Contrails remain short if E is below the ice-saturation curve and contrails may grow and persist if E is above ice saturation

criterion (Schumann, 1996), is due to the increase in relative humidity that occurs in the engine plume during mixing of the warm and moist exhaust gases with the colder ambient air, see Figure 4. When the humidity reaches liquid saturation in the young plume behind the aircraft, liquid droplets form, mainly by condensation of H_2O on exhaust particles, which then soon freeze to form ice particles. In the absence of aircraft-induced particles, water in the moist aircraft plume would condense on ambient aerosol entrained into the plume, causing less but larger ice particles, with smaller optical thickness and less radiative impact. Contrails typically form at temperatures below about -35 to $-60\,^{\circ}\mathrm{C}$, see Figure 5, but the actual value

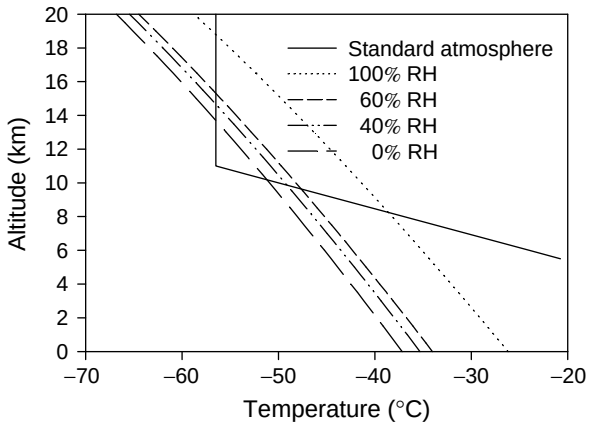


Figure 5 Threshold temperature for 0, 40, 60, and 100% relative humidity over liquid saturation (RH) for overall propulsion efficiency $\eta = 0.3$, kerosene fuel with water vapor emission index $EI_{H_2O} = 1.223$, and combustion heat $Q = 43.2\,\mathrm{MJ\,kg^{-1}}$, and temperature profile of the mid-latitude standard atmosphere versus altitude, z

increases with ambient relative humidity and with the parameter:

$$G = \frac{EI_{H_2O} p c_p}{[\varepsilon Q(1-\eta)]}$$

which is the slope of the mixing line in the diagram of water vapor partial pressure versus temperature, Figure 4. Hence the threshold temperature for contrail formation depends also on fuel and combustion properties in terms of the emission index EI_{H_2O} of water vapor and the combustion heat Q, on ambient pressure p at flight level, on the specific heat capacity of air c_p, the ratio $\varepsilon = 0.622$ of molar masses of water vapor and air, and on the overall efficiency η of the aircraft/engine system at cruise.

Contrails evaporate quickly when forming in dry ambient air. The fractional global coverage by these contrails is less than 0.001% and hence of no importance to the climate. Contrails persist, sometimes for hours, and grow when the ambient air is very humid with humidity above ice saturation. Ice particles in such persistent contrails grow by uptake of H_2O from the surrounding air. About 10 20% of all jet aircraft flights occur in air masses that are humid enough to cause persistent contrails. Ice-supersaturation in these regions is often too small to allow cirrus to form naturally, so aircraft act as a trigger to form cirrus clouds.

Contrail clouds can be identified and discriminated from natural cirrus clouds in satellite data based on their linear shape, see Figure 6, at least up to some age depending on the ambient conditions (Mannstein *et al.*, 1999). In 1996 and 1997, satellite data reveal that linear persistent contrails cover about 0.5 0.7% of the sky at noon over Europe in the annual average. This represents a lower bound for the actual

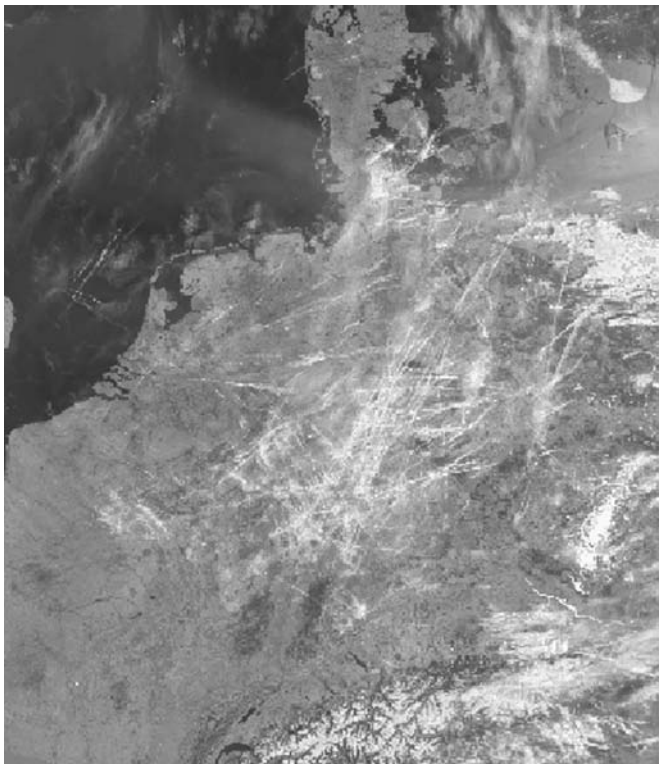


Figure 6 Contrails over mid-Europe on May 5, 1995, 7 : 43 UT, processed from NOAA-12 AVHRR data (as explained in Mannstein *et al.*, 1999). Photo: DLR/DFD

contrail cover. The fraction of cirrus clouds with non-linear shape which originated from old contrails or from aircraft induced particles is unknown. The satellite data indicate a night-time contrail cover which is about 1/3 of that during daytime.

The mean global cover by linear contrails has been estimated to be 0.1% (possibly 0.02 0.2%) for 1992. The contrail cover grows faster with traffic than aviation fuel consumption because air traffic increases mainly in the upper troposphere where contrails form preferentially. Future improvements in the overall propulsion efficiency η will increase the threshold temperature of contrail formation and, hence, the contrail cover.

Contrails cause a positive mean radiative forcing at the top of the atmosphere, in particular during night and over warm and bright surfaces. They reduce both the solar radiation reaching the surface and the amount of longwave radiation leaving the Earth to space. Contrails reduce the daily temperature range at the surface and cause heating of the troposphere. The radiative effects of contrails depend mainly on their coverage and optical depth, both of which are still not well known.

For 0.5% additional cirrus cloud cover regionally, a regional surface temperature increase of the order 0.05 K is expected. Such an increase can hardly be discriminated from natural variability. However, an increase of mean cirrus cloudiness in regions with heavy air traffic from typically 20% to 25% (i.e., by 5% cloud cover) has been found in models to cause a significant regional temperature change of order 1 K. The daily temperature range at the ground decreases with increasing cirrus cover but only little because of stronger sensitivity to lower atmospheric humidity.

CIRRUS CLOUDS

Some correlation between soot and ice particle concentrations have been measured in cirrus clouds which indicates an aviation impact on ice particle formation. Contrails trigger cirrus clouds in air masses which are sufficiently humid to let contrails spread but not humid enough to let cirrus form naturally. On average an increase in cirrus cloud cover tends to warm the surface of the Earth. More than 20% of the Earth is covered with cirrus clouds. Long-term observations of cloud frequencies at some meteorological stations and some satellite data indicate strong increases of the frequency or cover of cirrus clouds (Boucher, 1999; Intergovernmental Panel on Climate Change (IPCC), 1999), for unknown reasons. The increases are particularly large in regions with high air traffic density. The global trends suggest a cirrus cover increase of up to 0.2% since the beginning of jet aviation, in addition to the 0.1% cover by line-shaped contrails. The cirrus cover changes cannot be conclusively attributed to aircraft emissions or any other causes.

CLIMATE EFFECTS

The radiative effects of the emitted greenhouse gases and particles, of changes in O_3, and of changes in cloudiness, can be compared to each other and to climate effects from other causes by using the concept of radiative forcing. *Radiative forcing* is de ned as the net radiative ux change at the top of the atmosphere calculated in response to a perturbation, such as a change in gas concentration or cloud cover (with the stratosphere thermally adjusted to the change in radiative uxes) (*see* **Radiative Forcing**, Volume 1). A positive net ux change represents an energy gain and, hence, a net heating of the Earth System. The global mean climate change of surface temperature at sea level is approximately proportional to the global average radiative forcing. However, the strong concentration of air traf c at the northern mid-latitudes may cause larger regional climate changes. Up to now, such regional changes are not detected.

The radiative effects resulting from aircraft engine emissions are summarised in Figure 7 (Intergovernmental Panel on Climate Change (IPCC), 1999). The uncertainty associated with several of these effects is large. The best estimate of the radiative forcing in 1992 by aircraft is

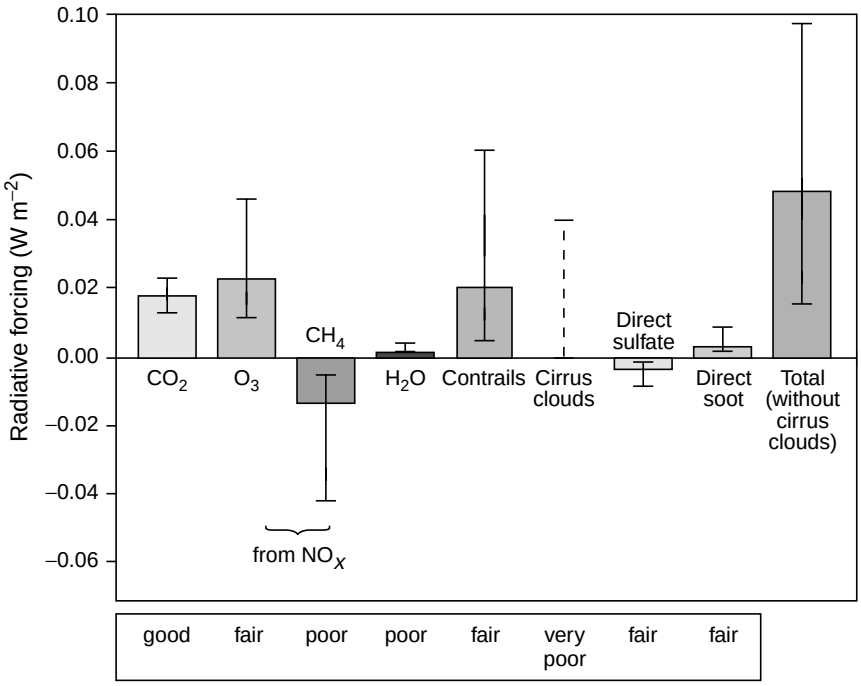


Figure 7 Radiative forcing (W m^{-2}) due to aircraft emissions in 1992 (Intergovernmental Panel on Climate Change (IPCC), 1999). The columns indicate the best estimate of forcing. The error bars denote the range within which the best-estimate value is expected with a 2/3 probability. The available information on cirrus clouds is insufficient to determine either a best estimate or an uncertainty range; the dashed line indicates a range of possible best estimates. The estimate for total forcing does not include the effect of changes in cirrus cloudiness. The evaluations below the graph (good, fair, poor, very poor) are a relative appraisal associated with each component and indicates the level of scientific understanding. (Reproduced by permission of Intergovernmental Panel on Climate Change (IPCC), 1999)

0.05 W m^{-2} or about 3.5% of the total radiative forcing by all anthropogenic activities. The overall radiative forcing by aircraft (excluding that from changes in cirrus clouds) is a factor of 2 4 larger than the forcing by aircraft CO_2 alone.

In the future, for a reference scenario with a 6.7-fold increase in traffic and about a 3-fold increase in fuel consumption from 1992 to 2050, the Intergovernmental Panel on Climate Change (IPCC) (1999) reported a computed radiative forcing by aircraft in 2050 of 0.19 W m^{-2} (3.8 times the value in 1992). For the range of scenarios considered in that assessment, the forcing is projected to grow to 0.13 to 0.56 W m^{-2} in 2050, i.e. 2.6 to 11 times the value in 1992. These estimates of forcing combine the effects from changes in concentrations of CO_2, O_3, CH_4, H_2O, line-shaped contrails, and aerosols, but do not include possible changes in cirrus clouds.

The emissions of NO_x cause changes in CH_4 and O_3, and their influences on the radiative forcing are of similar magnitude, but of opposite sign. Changes in tropospheric O_3 are mainly in the Northern Hemisphere, while those of CH_4 are global in extent so that, even though the global average radiative forcings are of similar magnitude and opposite in sign, the latitudinal structure of the forcing is different so that the net regional radiative effects do not cancel.

The climatic effects of aircraft are superimposed on those caused by anthropogenic emissions of other greenhouse gases and particles, and on the background natural variability. It has not yet been possible to separate the aviation effects in observed climate parameters, such as surface temperature, from those of other natural or anthropogenic disturbances.

ULTRA-VIOLET RADIATION EFFECTS

Besides being a greenhouse gas, O_3 provides a shield against solar UV radiation. The calculated total O_3 column in northern mid-latitudes due to aviation is computed to be about 0.4% in 1992 (Intergovernmental Panel on Climate Change (IPCC), 1999). The increase in O_3 induced by subsonic aircraft decreases the UV radiation. The *erythemal* dose rate, defined as UV irradiance weighted according to how effectively it causes sunburn, is a maximum at northern mid-latitudes and is estimated to have been decreased by aircraft in 1992 by about 0.5% at 45 °N in July. The aircraft effect on UV is small compared to that caused by other O_3 changes.

RADIATIVE FORCING AND UV RADIATION EFFECTS OF SUPERSONIC AVIATION

About 13 supersonic aircraft of type Concorde were in operation until the accident in Paris on July 25, 2000. Their effect on the global atmosphere can be neglected compared to the effects of the far larger fleet of subsonic aircraft. Plans for a new fleet of supersonic aircraft have been discussed but no definite decision has been made to develop such a fleet (Intergovernmental Panel on Climate Change (IPCC), 1999). Supersonic aircraft would cruise at an altitude of about 16 24 km, would emit CO_2, H_2O, NO_x, SO_x, and soot directly into the stratosphere, and contribute to changes in stratospheric O_3, other trace gases and aerosols. Supersonic aircraft consume more than twice the fuel per passenger-km compared to subsonic aircraft. The radiative forcing of civil supersonic aircraft is estimated to be about a factor of 5 larger than that of an equivalent fleet of subsonic aircraft. The calculated radiative forcing of supersonic aircraft depends on the treatment of H_2O and O_3 in models which are difficult to simulate accurately. The effect of introducing a fleet of 500 civil supersonic (Mach 2.4, $EI_{NO_x} = 5\,g\,kg^{-1}$) aircraft is to reduce the stratospheric O_3 column in the Northern hemisphere by -0.4% (range -2.5 to $+0.5\%$, Kawa *et al.*, 1999), and to increase erythemal dose rate by a similar amount.

UNCERTAINTIES

Considerable progress has been made in recent years in quantifying the impact of aircraft emissions on the atmosphere. However, the scientific uncertainties are generally large. The state of scientific understanding of the radiative forcing of the various emissions has been evaluated by the Intergovernmental Panel on Climate Change (IPCC) (1999) as indicated in Figure 7. The evaluation is based on the evidence available, the degree of consensus in the scientific literature, and the scope of the analysis. The error bars and the scientific uncertainty are largest for the complex impact of NO_x on CH_4 reduction, for contrail cover and contrail optical depth, and, in particular, for the indirect effects of aviation on cirrus cloud cover and changes in cirrus properties. The figure accounts for global effects only. More work needs to be done to assess the impact of aviation on the climate in the northern mid-latitudes, where most of the air traffic takes place.

REFERENCES

Boucher, O (1999) Influence of Air Traffic on Cirrus Occurrence, *Nature*, 397, 30 31.

Brasseur, G P, Cox, R A, Hauglustaine, D, Isaksen, I, Lelieveld, J, Lister, D H, Sausen, R, Schumann, U, Wahner, A, and Wiesen, P (1998) European Scientific Assessment of the Atmospheric Effects of Aircraft Emissions. *Atmos. Environ.*, 32, 2327 2422.

Intergovernmental Panel on Climate Change (IPCC) (1999) *Aviation and the Global Atmosphere*, eds J E Penner, D H Lister, D J Griggs, D J Dokken, and M McFarland, Cambridge University Press, Cambridge, 1 373.

Karcher, B (1999) Aviation-produced Aerosols and Contrails, *Surveys Geophys.*, 20, 113 167.

Kawa, S R, Anderson, J G, Baughcum, S L *et al.* (1999) *Assessment of the Effects of High-Speed Aircraft in the Stratosphere: 1998*, Report NASA/TP-1999-209 237, NASA Goddard Space Flight Center, Greenbelt, MD, 1 150.

Lee, D S, Kohler, I, Grobler, E, Rohrer, F, Sausen, R, Gallardo-Klenner, L, Olivier, J G, Dentener, F J, and Bouwman, A F (1997) Estimates of Global NO_x Emissions and their Uncertainties, *Atmos. Environ.*, 31, 1735 1749.

Mannstein, H, Meyer, R, and Wendling, P (1999) Operational Detection of Contrails from NOAA-AVHRR-data, *Int. J. Remote Sensing*, 20, 1641 1660.

Ponater, M, Sausen, R, Feneberg, B, and Roeckner, E (1999) Climate Effect of Ozone Changes Caused by Present and Future Air Traffic, *Clim. Dyn.*, 15, 631 642.

Sausen, R and Schumann, U (2000) Estimates of the Climate Response to Aircraft CO_2 and NO_x Emissions Scenarios, *Climatic Change*, 44, 27 58.

Schumann, U (1996) On Conditions for Contrail Formation from Aircraft Exhausts, *Meteorol. Z.*, 5, 4 23.

Schumann, U, Schlager, H, Arnold, F, Ovarlez, J, Kelder, H, Hov, , Hayman, G, Isaksen, I S A, Staehelin, J, and Whitefield, P D (2000) Pollution from Aircraft Emissions in the North Atlantic Flight Corridor: Overview on the POLINAT Projects, *J. Geophys. Res.*, 105, 3605 3631.

Animal Production, Feedlots, and Manure Problems in the US (Agriculture, Intensive)

Claudia Copeland
US Library of Congress, Washington, DC, USA

Managing the environmental effects of intensive animal rearing and feeding operations is a problem confronting the livestock industry. These facilities, which include con ned feeding operations and feedlots, are a specialized part of the livestock production process, largely separate from cropland agriculture. Manure and waste-handling and disposal problems from intensive animal production are receiving attention as these facilities increase in size and the effects of these problems reach beyond the industry.

The US population of animals in livestock production being raised to feed Americans and other consumers worldwide is very large. Waste produced by these animals is a valuable soil amendment and source of crop nutrients, when applied to land in proper amounts (the traditional waste management approach). However, if not properly used or disposed, or if applied in amounts that exceed plant needs, animal waste or its residuals can leach through soil to contaminate ground water or can be transported by runoff to pollute surface waters. Thus, as animal production has intensi ed concentrating more animals on individual farms, a growing challenge for agriculture is nding suf cient land to dispose of manure, or nding economic alternatives, especially if the supply of land for disposal is insuf cient. The parallel challenge for policymakers is determining if the environmental impacts of animal waste management are signi cant enough to require new remedies and, if so, what strategies are appropriate.

These issues have become prominent in US debates about agricultural and environmental policy. Beyond local and national borders, there are broader potential impacts, as well, re ected in concerns that the trends in livestock production practices and impacts on water and air quality observed in the US are being repeated elsewhere.

OVERVIEW OF ANIMAL AGRICULTURE

Livestock includes cattle (beef, dairy and veal), swine (hogs and pigs), poultry (chicken and turkeys), and sheep and lambs. Livestock is a large component of the farm economy in the US, contributing nearly half of the cash receipts of all US agriculture in recent years. The populations of animals are very large. In 1992, the US inventory of livestock was over 77 million cattle and calves, about 60 million swine, and almost one billion broilers. The inventory gradually shifts in response to changing market conditions and consumer preferences. Changes in geographic location of animals and how they are raised re ect economic considerations, business relationships, and changing technology. These changes have contributed to low prices of food protein globally and led to increasing concerns about several topics, including environmental effects.

Concentration

Livestock production in the US continues to have fewer producers operating at fewer sites. Such concentration offers economies of scale and depends increasingly on modern technologies and better information. This concentration started rst in the poultry industry about 40 years ago, and it is now the most concentrated segment. Broiler production in the US nearly tripled between 1969 and 1992, while the number of farms with broiler houses dropped by 35%. The concentration process has been similar for cattle feed operations and for the dairy industry: overall, the number of dairy farms dropped from 1988 to 1998 by 20%, while the average herd size increased by more than 50%. The US swine inventory has climbed by about 18% since the mid-1980s, while the number of swine farms dropped by 72%. The largest farms have grown larger, so that now, farms with an inventory of at least 1000 head make up less than 3% of the farms, but produce 60% of the swine.

Integration

Growing concentration is an element of broader changes often referred to as industrialization. Trends in farm organization and operation include not only fewer farms of a larger average size, but also vertical integration as farmers enter contracts with processors, or integrators. Under these contracts, producers raise the livestock while integrators actually own the animals, assume marketing risks, and may provide medicine, feed and technical expertise as well. Generally, vertical integration has increased the volume and certainty of supply and improved the market characteristics of the livestock. Producers forgo the risks and uncertainties of the marketplace by becoming contract growers, and as more production comes within these types of relationships, marketing opportunities may decline for those who do not choose to or are unable to participate. In economic terms, ef ciencies are gained for both producers and integrators at the expense of non-participants.

The overall livestock sector has grown since the 1980s, in part to serve expanding demand for protein in a more af uent world. For example, the US has become the largest beef-exporting nation in the world, with between 16 and 20% of world trade in recent years. Swine trade is similar, growing by an annual average of 4% since 1989; the US now accounts for almost 20% of the world s pork exports.

WASTE FROM ANIMAL AGRICULTURE

Animal wastes are predominately solid and liquid manures, but also include used bedding, spilled feed, and dead animals. Manure production in the US is estimated to be almost 112 million tons (dry matter) annually (113.8 million metric tons), or, 13 times more (on a dry-weight basis) than human sanitary waste production. Production varies not only with the animal type, but also with such factors as feed ration, health, animal age, and climate. Trends in the livestock sector, such as increased animal con nement (rather than pasturing) and improved feeds, have increased the amount of manure produced per animal and changed the composition. The larger volume per animal combined with concentration of more animals at a site compounds storage and disposal dif culties for farmers.

Animal waste contains nutrients, including nitrogen, phosphorus, and potassium. Nutrients can be valuable for crops, but their oxygen-demanding characteristics can cause water quality problems. Waste can also contain organic solids, trace heavy metals, salts, bacteria, viruses, other microorganisms, and sediments.

Farmers manage manure and related wastes in many different ways. Manure may be collected at temporary storage facilities (including stacks, ponds and tanks) until it can be treated or utilized. Waste may be treated to convert it to a more useful resource, usually by concentrating the beneficial constituents and decreasing the total volume.

Lagoons are the most common treatment facility. In a lagoon, the manure undergoes continuous anaerobic decomposition. Failure of lagoons and the resulting waste spills have brought much of the recent critical attention to animal agriculture, and some have called for phasing out lagoons. Other types of treatment facilities include composters, solid separators, and settling basins. Holding capacities of storage and treatment facilities are recommended based on the estimated time period the volume of waste must be retained. For example, waste should not be spread while the ground is frozen, or it will be washed into surface waters, so farms in locations where winters are long need a greater storage capacity. The management process ends when the waste is transferred and used. The most common use, by far, is to spread it across the farm fields as a soil amendment and a nutrient supplement. The manner in which it is spread can have environmental implications; for example, spray irrigation has been associated with environmental problems in some situations, while cultivation into the soil generally minimizes that potential.

Manure as an Agricultural Asset

Manure can be a valuable asset for agriculture, generally as a supplement to, or partial substitute for commercial fertilizers. Manure nutrient values, however, vary considerably and any supply of manure must be assessed to determine application rates. Commercial chemical fertilizer, by contrast, has consistent content, with that information supplied by the manufacturer. Also, nutrients from manure are not all immediately available.

A better way to view manure is as a soil amendment that improves many of the physical and chemical properties, as well as the nutrient values, of soil by adding organic material and improving soil structure and the ability to hold water and retain nutrients. These benefits are considerable, in terms of nitrogen and phosphorus needed by crops, plus the added benefits of improved soil quality, decreased runoff and soil erosion potential, and improved soil moisture, or offsetting costs associated with processing, transportation, and management.

Disposal Problems

Disposal has become more difficult for US producers because operations are larger; thus, there is more waste at a single site, and less land is under the control of these operators. Many of the largest US farms lack the capacity to manage manure on the land under their direct control. These difficulties have been compounded by increased concerns about water quality and other problems that may originate with intense and concentrated livestock farms.

In the US, animal waste accumulates unequally because of geographic concentration of livestock, and in many watersheds, the amounts of manure nitrogen and phosphorus generated exceed amounts needed to satisfy crop uptake/removal requirements. Data on manure production in 1997 indicate a significant potential for excessive nutrients, if all manure were spread on land. Manure nitrogen available for application to cropland and pastureland exceeded the potential for crop uptake by 57%. Likewise, manure phosphorus exceeded the potential for crop uptake and pastureland application by 65%. Estimates are that the number of US counties with excess manure production that have the potential to cause water quality problems more than doubled from 1982 to 1997.

Even when farmers control adequate land, proper disposal can still be a problem. The manure must be assessed for its nutrient value, then transported to the site, and spread in the proper amounts; each step is an expense for farmers. Producers often determine application rates based on crop needs for nitrogen. However, if application rates supply the needed nitrogen, there will be occasions when the amount of phosphorus or potassium applied will be excessive, especially after several successive years of application. Soils in areas with concentrated animal operations may accumulate excessive amounts of phosphorus, which normally binds to soil particles, but when particles become saturated, excess phosphorus enters ground water or nearby surface waters.

Manure management problems appear less substantial in the US than in some parts of Europe, especially the low countries, and some European countries are exploring options for addressing this problem. For example, in the Netherlands, where phosphorus saturation is generally believed to be the most serious problem, programs to stabilize manure production and application started in 1987, and since 1991, manure application rates have been declining. The Dutch policy goal is to reach an equilibrium fertilization rate by 2010, when the supply of nutrients from manure plus fertilizer is to be in balance with crop utilization and other losses.

Waste Disposal Options

Alternatives for disposing of manure, other than by land application, are receiving increased attention. The constituents and moisture content of the manure are important

in determining which disposal techniques to use. Shipping costs constrain many options because manure is of low economic value on a volume or weight basis, so it is uneconomic to ship it for long distances unless the waste can be concentrated to decrease the volume or increase the value.

Options under study include composting, burning, and biotech changes to feed supplies that alter the characteristics of the waste. Composting uses microorganisms to turn wastes into relatively stable and odorless material that can be used or sold as a soil amendment. Burning can be used to generate energy, as is done at a power station in East Anglia, UK, but there are limitations. For example, one ton of poultry litter contains about the same amount of energy as 80 US gallons (302.8 liters) of home heating oil, but has the same volume as about 500 US gallons (1892.7 liters). The residue of burned manure is about 10% of the original volume, and contains little of the original nitrogen and sulfur but most of the phosphorus that was initially present. A biochemical change that has received attention is adding phytase (an enzyme) to feed. It has been shown to remove 25% to 40% of the phosphorus and 10% of the nitrogen excreted by causing the animals to use nutrients more efficiently. Phytase does have a cost and must be added to feed. Still other options include converting the waste to fuel and energy, using industrial processes to convert the waste to other useful products, and using it as a feed supplement.

Such approaches are hard to apply widely because of various combinations of physical limitations, high costs, lack of knowledge or management skills by the producer, unfamiliar or untested technologies, and other factors. The costs and difficulties of storing and handling may be the most common constraints.

ENVIRONMENTAL AND PUBLIC HEALTH CONCERNS

In the US, agriculture is believed to be the leading source of water quality impairment in rivers and lakes, affecting 70% of impaired river lengths (waters that fail to meet applicable standards of water quality) and 48% of impaired lake areas. In estuaries, agriculture affects nearly 30% of impaired acres/hectares. In a 1996 assessment, in 22 states that specifically evaluated agriculture s impacts on rivers and streams, animal feedlots and animal holding areas were estimated to be the principal pollutant source in 20% of waters impaired by agricultural practices.

Animal feeding operations have been shown to cause significant environmental and public health problems, including nutrient enrichment of surface and ground waters, contamination of drinking water supplies (with e-coli, for example), fish kills, and odors. Animal waste, if not properly managed, can be transported by water over the surface of agricultural land to nearby lakes and streams. There, the nutrients in animal waste can reduce the oxygen content of the water, leading to algal blooms, fish kills, and threats to other wildlife. Catastrophic events, such as spills from livestock waste lagoons, have occurred in nearly every US state. Solids deposited in water bodies can accelerate eutrophication by releasing nutrients over extended periods. Leaching from manure storage lagoons and percolation through the soil of fields when animal waste is applied has resulted in nutrient contamination of groundwater resources.

Although animal waste is not the only source of pathogens in surface waters, it has been responsible for shellfish contamination in some coastal waters. Closure of shellfish beds and recreational beaches can be necessitated by high fecal coliform counts, both from animal waste runoff and discharge of improperly treated sewage. Animal diseases also can be transmitted to humans through contact with animal feces.

Atmospheric deposition of nitrogen from animal operations is also an environmental concern. This occurs when nitrogen in liquid waste is volatilized as ammonia (NH_3) from anaerobic (oxygen-deprived) lagoons, causing ammonia to evaporate. Volatilization also occurs after land application. Once in the atmosphere, it is converted to forms that are redeposited within 80 160 km on land or in surface waters. These forms of nitrogen are water-soluble, meaning that the nitrogen can adversely affect water quality much like nitrogen fertilizer if it enters a stream as direct surface runoff. Atmospheric transport broadens the environmental dimensions of animal waste management from local to national and even global.

Odor is the most controversial nuisance problem associated with feedlots, and is generally related to manure handling and spreading on fields. Other potential odor sources include wet feed and the decomposition of dead animals. The odors consist of gases, such as ammonia, hydrogen sulfide, methane, and organic compounds produced during decomposition of manure. Although some of the gases are known to be harmful or toxic in large amounts, the principal effect upon humans is annoyance or nuisance. Solutions to odor problems generally involve setbacks, buffers, and other land use planning tools.

POLICY QUESTIONS

Social and political pressure to address the environmental impacts of livestock production has grown to the point that many US policymakers today are asking what to do, not whether to do something. The setting is one where agricultural policy and environmental policy, which traditionally have separate agendas and priorities, come together on some points. A major direction of inquiry is what strategies will be effective and attainable, who should be responsible for developing and implementing strategies to address known and future problems, and who should bear the burdens of new strategies.

One set of policy questions concerns the role of government in addressing animal waste issues. In the US, federal, state, and local governments currently have numerous programs and policies in place, ranging from regulatory programs for implementing water quality requirements, to programs providing technical assistance and financial incentives to farmers for improved waste management practices. Policymakers debate the appropriate balance of regulatory and incentive-based approaches.

State and local governments implement programs on regulation, siting and zoning, technical assistance, and cost sharing. A key question is what degree of federal leadership and national consistency is appropriate and how federal policies will blend with those of states, regions, and localities.

Further, there are questions about additional aspects of government s role generally in addressing animal waste issues. For example, is there a role for government in facilitating exchange or transfer of manure to lessen imbalances between areas with intense animal production and areas with fewer livestock farms where the nutritive value of manure can be used? Is there need for government to encourage certain manure management practices with potential for increasing the nutrient value while reducing off-site damages?

Experts have identified a number of areas of research needed to inform current and new animal waste management policies in the US and elsewhere. Many believe basic research on manure use and disposal is needed. Other research requirements include improved soil tests to determine nutrient application rates and methods, and improved systems to identify water resources at risk from manure and fertilizers (i.e., those where geology, soils, and climate create potential for runoff and erosion). Two related questions are: who should take the lead on research questions (government or industry, for example), and whether some research must precede implementation (e.g., studies on crop uptake of nutrients and application of phytase or other nutrient-reducing enzymes in feed). Efforts to educate and assist farmers in understanding the value of manure as a soil amendment and the use of alternate disposal techniques are also a priority.

FURTHER READING

Aneja, V P, Murray, G C, and Southerland, J (1998) Atmospheric Nitrogen Compounds: Emissions, Transport, Transformation, Deposition, and Assessment, *Environ. Manage.*, 48(4), 22 25.

Caldwell, W J (1998) Land-use Planning, the Environment, and Siting Intensive Livestock Facilities in the 21st Century, *J. Soil Water Conserv.*, 53(2), 102 106.

Mallin, M A (2000) Impacts of Industrial Animal Production on Rivers and Estuaries, *Am. Sci.*, 88(1), 26 37.

National Research Council, Board on Agriculture (1993) *Soil and Water Quality: an Agenda for Agriculture*, National Academy Press, Washington, DC.

US Council for Agricultural Science and Technology (1996) *Integrated Animal Waste Management*, Task Force Report No. 128, USA.

Van der Molen, D T, Breeuwsma, A, and Boers, P C (1998) Agricultural Nutrient Losses to Surface Waters in the Netherlands: Impact, Strategies, and Perspectives, *J. Environ. Qual.*, 27, 4 11.

Anthropocene

see Anthropocene (**Volume 1**)

Anthropogenic

Anthropogenic: of human origin, in contrast to a process or substance occurring naturally. The word is frequently used in the context of biogeochemical cycling; for example, sulfur from volcanoes and sea spray is from natural sources, whereas sulfur from a smelter is anthropogenic.

R E MUNN *Canada*

Anthropogenic Metabolism and Environmental Legacies

see Anthropogenic Metabolism and Environmental Legacies (**Volume 3**)

Aquaculture and Environment: Global View from the Tropics to High Latitudes

Malcolm C M Beveridge
University of Stirling, Stirling, Scotland

Aquaculture, the farming of aquatic organisms, is the fastest growing food sector. Over the past 10 years, production by

weight and value has doubled, as has the contribution of farmed aquatic products to world sh and shell sh supplies. One sh in four on our plates is now of farmed origin. Aquaculture has also become an important source of employment, especially in remote, rural areas.

With the rapid expansion of the industry, however, has come growing concern about the impacts of certain types of aquaculture. Much of the recent losses of mangroves in the coastal zone of Asia and Latin America have been attributed to shrimp farming. The more intensive types of aquaculture use considerable quantities of water and are heavily reliant on shmeal- and sh oil-based diets. Copious quantities of dilute, dif cult to treat wastes, sometimes with residues of chemicals used in disease treatment, may be discharged directly into the environment. The movement and escape of farmed aquatic animals is also causing concern in some parts of the world.

Aquaculture is the aquatic counterpart of agriculture, with aspirations to farm all species that might prove of use, including microorganisms, plants, invertebrates, fish, amphibians and reptiles. Farming in the aquatic context may be differentiated from hunting (fishing) because, as in agriculture, some measure of care or cultivation is involved. A further factor distinguishing it from hunting is ownership of stock. The definition currently accepted by the Food and Agriculture Organization (FAO) of the United Nations accepts that both husbandry and ownership are intrinsic, stating that aquaculture is:

> . . . the farming of aquatic organisms including crocodiles, amphibians, finfish, mollusks, crustaceans and plants, where farming refers to their rearing to their juvenile and/or adult phase under captive conditions. Aquaculture also encompasses individual, corporate or state ownership of the organism being reared and harvested . . ..
>
> (Rana *et al.*, 1999)

Note that aquaculture includes organisms owned and reared other than for food for production of food for other farmed aquatic species (microalgae, such as *Isochrysis galbana* and the brine shrimp *Artemia* spp.), leather for clothing and fashion accessories (crocodiles), jewelry (pearls), fine chemicals (the alga *Dunaliella*, for carotenoids), even pets (aquarium species) are considered as products of aquaculture. Despite the attempts at clarification, gray definitional areas remain. For example, the term ranching, generally applied to the release of hatchery-reared organisms into the environment where they mingle with wild stocks of the same species and are in due course harvested together with the wild stocks, using conventional fishing methods, needs further clarification.

The origins of fish and shellfish farming are lost in history, there being few artifacts to guide the historian. Nevertheless, we know that fish and shellfish farming began in a number of parts of the world and at various stages along the aquatic food supply line, between water and plate, several millennia ago. Aquaculture is, almost by definition, an activity of settled societies and originated among both fishing and wetland farming cultures as well as at points of trade. We can surmise that conditions similar to those initiating the development of agriculture would usually have been necessary, that foraging and hunting (i.e., fishing) were insufficient to satisfy demands for fish and shellfish, resulting in some form of stock management being introduced to increase harvestable supplies. It is believed that a range of proto-aquaculture activities, a term coined to describe the various types and degrees of intervention aimed at increasing the biomass of aquatic organisms that could be exploited by humans, paralleling what is now believed to have happened in agriculture (Harris, 1996), were probably practiced. Precursors to aquaculture can still be found, for example, in the early forms of *vallicoltura* (managed lagoons) initiated by the Romans some 2000 years ago. Although perhaps pre-dated by events in Africa, Asia (China in particular) is widely regarded as the cradle of aquaculture. Aquaculture is thought to have evolved here at a comparatively early stage, perhaps as far back as 3000 years ago, among agrarian societies that farmed wetland areas (Bray, 1984). It is believed that Chinese emigrants took various species of carp and the techniques of fish farming with them first to Korea and then to Japan from where the technology gradually spread to other parts of Southeast Asia. The early Christian church, with its emphasis on meatless days and the development of monastic cultures with strong agrarian skills, was a major factor in the development and spread of aquaculture in Western Europe.

Despite its ancient origins, aquaculture is a largely post-second world war phenomenon. In 1950, global farmed fish and shellfish production was around two million tons (Figure 1) and largely confined to areas of Asia. Why? Most obviously, there was little need to farm fish and

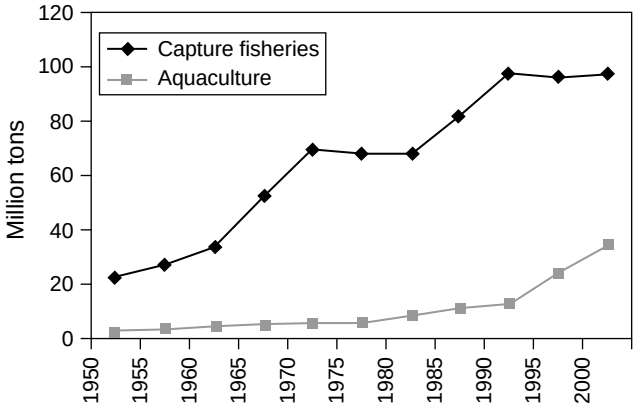


Figure 1 Growth in global fisheries and aquaculture production, 1950–1997. (Based on FAO data)

shellfish for food. Supplies were plentiful, even in the middle years of the present century. The Second World War had curtailed fishing activities and allowed stocks to recover. Growing demand and the development of increasingly effective means of finding and exploiting fish, however, quickly undid the good that the hiatus in fishing had produced, leading to increasing numbers of stocks becoming over exploited. Degradation of aquatic environments through pollution and impoundment of rivers further exacerbated the situation, while the burgeoning global population, combined with urbanization and rapid economic growth stimulated demand.

Over the past half century, aquaculture production has grown nearly 20 fold to 36 million tons in 1997, comprising 16.7 million tons of finfish (49% of the total), 8.5 million tons of mollusks (25%) and 7.7 million tons of aquatic plants (23%) (FAO, 1999). The increasing rate of growth of aquaculture production has resulted in a more than doubling of the contribution to total world fisheries production (capture and culture) from 11% in 1984 to 26% in 1997. Much of the growth has been in farmed freshwater fish production in China where it proved possible to address the demands for fish from the rapidly growing population through universal collectivization of land and other resources, the development of a rigorously planned agricultural system and wide-scale impoundment of rivers. The significance of farmed aquatic organisms is even greater than production statistics indicate. At US$50.4 billion, farmed aquatic products account for half the value of global capture and culture fisheries production. From a human food perspective, more than 25% of fish eaten (one fish in four) comes from aquaculture. There are, however, tremendous regional differences in production (Figure 2). More than 90% occurs in Asia, China alone accounting for more than two thirds of the total. By contrast, Europe and America account for less than 5% and 2% of aquaculture production, respectively. While nearly 90% of aquaculture production is from developing countries and 82% is from low income food deficit countries (LIFDCs) (Tacon, 1998), production in Africa (120 000 tons; 0.3%) remains trivial, especially in relation to the problems that the continent faces in feeding its population.

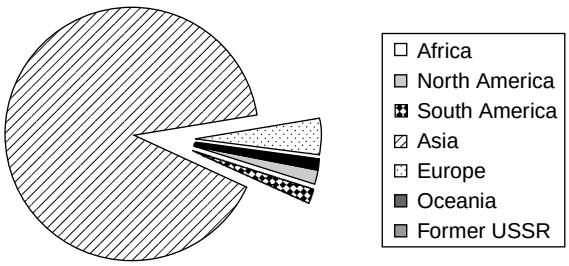


Figure 2 Regional differences (%) in global aquaculture production in 1997. (*Source:* FAO (1999))

Aquaculture remains primarily about rearing fish and shellfish for food; and growth has been largely market-driven. Production of low value species, especially that of carps, has increased tremendously in Asia. However, production methods are intensifying, largely because of increasing pressure on land and water resources (Beveridge *et al.*, 1997a). Growth in incomes in many Asian newly industrialized countries has also fuelled demand for luxury foods, such as live farmed grouper (*Epinephilus* spp.) and seabass (*Lates calcarifer*). Export markets in Japan, Europe and North America for high value farmed aquatic products such as shrimp are strong. The spectacular returns on investment in shrimp farming attracted many investors both in Latin America and in Asia while the implications for foreign exchange arguably influenced the lax environmental regulatory policies of governments. Only in the 1990s have disease and the deleterious effect on mangrove and other coastal resources constrained growth. In Europe, fastest growth has been seen in the production of high market value fish species such as Atlantic salmon (*Salmo salar*) and gilthead seabream (*Sparus aurata*). Throughout the 1990s falling consumer prices have driven the expansion of the market and alternative species/products are now urgently needed. Technical constraints, however, especially with regard to the spawning and larval rearing of marine fish and crustaceans, and production economics have somewhat tempered development.

Aquaculture usually involves maintaining a species in a secure facility under conditions in which it can thrive. Interventions in the life cycle, or husbandry practices, range from exclusion of predators and control of competitors (extensive) through enhancement of food supply (semi-intensive) or even, in intensive aquaculture, the provision of all nutritional requirements. While control over all aspects of an organism s life cycle is desirable, enabling selective breeding for desirable traits such as improved growth, disease resistance or flesh quality, lack of knowledge and economics has constrained reproduction and larval rearing in captivity of many (especially marine) species. Two types of aquaculture system prevail: water-based and land-based. The former includes cages and pens for fish and trestles and long-lines for mollusks while the latter comprise ponds, tanks and raceways used in fish rearing. Mollusks, including species reared for pearl cultivation, are generally farmed in inter-tidal and sheltered inshore coastal areas, where wild caught or hatchery-reared seed (larvae/post-larvae) are typically grown in bags set on trestles or lantern nets or ropes suspended from the water surface. The animals are entirely reliant on prevailing supplies of plankton for food. Crustacean farming is dominated by that of penaeid shrimp in countries such as Thailand, Ecuador, China and Indonesia. Shrimp post-larvae, derived either from wild fisheries or from wild-caught egg-bearing females, are stocked in coastal ponds and

grown for 3 4 months until they reach marketable size. The more intensively managed shrimp farms are heavily reliant on pelleted, nutritionally complete feeds and use pumped seawater to replenish dissolved oxygen and flush away potentially harmful waste metabolites. A range of systems, ponds, tanks or cages is used to farm fish. Ponds predominate in the tropics and sub-tropics to semi-intensively rear carps and other freshwater species, usually in an agricultural setting. By contrast, much culture of marine fish in coastal areas is in floating net cages where intensive rearing methods, involving the use of nutritionally complete fishmeal and fish oil-based diets, are employed (Figure 3).

Aquaculture is reliant on a wide range of natural resources or environmental goods, including:

- land (or space in the sea) in which to site an aquaculture facility;
- materials (timber, steel, etc.) for construction of the system and ancillary facilities (roads, offices, etc.);
- water to support the animals;
- seed (eggs, larvae, past-larvae, fry) for stocking;
- feed and/or fertilizers to enhance production.

Figure 3 Plastic seabass and seabream cages, Mediterranean Spain. This technology is widely used for marine fish culture in northern Europe and is rapidly being adopted in Southeast Asia

Because production is carried out in ecologically open systems, in which materials are imported to the production site and transformed into marketable products through inputs of energy, wastes are inevitably produced. Wastes include:

- uneaten food, fecal and urinary products;
- drugs and chemicals used to combat disease (chemotherapeutants);
- feral (escaped) animals and pathogens that find their way into the environment.

Aquaculture is also dependent upon the environment for essential services: the replenishment of oxygen and the dispersal and assimilation of wastes that otherwise might accumulate in the production system, exerting negative feedback on farmed animal growth and production (see Figure 4). Thus, factors that determine the supply of environmental goods and services, such as changes in land use and pollution, may adversely affect aquaculture. Impacts of aquaculture on the environment stem from the consumption of environmental goods, from the transformation (i.e., farming) process itself and from the production and release of wastes. The overall relationship ranges from the positive (i.e., enhancement of environmental quality) through the relatively neutral, to that which results in environmental degradation. In general, the higher the demand for environmental goods and services, the greater the impacts on the environment.

In the public s mind, much of the controversy surrounding aquaculture revolves around the issue of the impact of aquaculture development on the landscape. While few

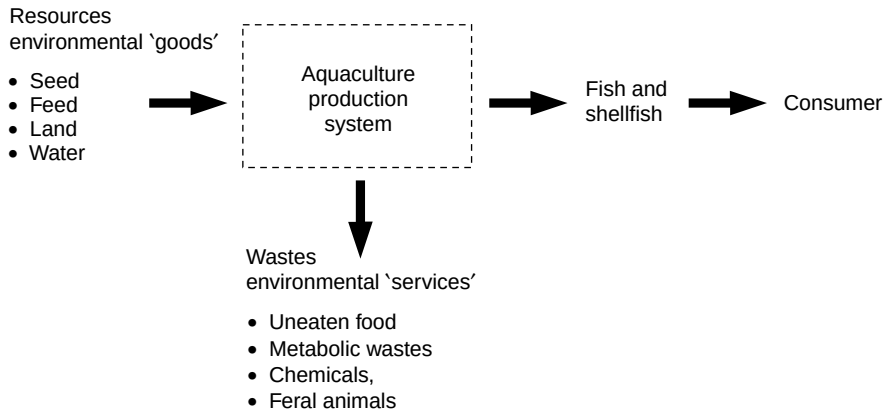


Figure 4 Interactions between aquaculture and the environment, illustrating its dependence on environmental goods and services (modified from Beveridge *et al.*, 1997b)

would claim that aquaculture developments enhance the landscape, there are marked differences among countries and social groups in response. Such issues have led to zoning of developments in areas such as Hong Kong (Beveridge, 1996).

Inland aquaculture is usually located in agricultural land and here the development of ponds can increase habitat diversity (Beveridge and Phillips, 1993). By contrast, tropical marine shrimp and fish culture usually utilizes previously unexploited areas for ponds. Demands for land for shrimp pond construction in Latin America (Ecuador, Colombia) and Southeast Asia (China, Thailand, Indonesia, and the Philippines) have caused the destruction of thousands of hectares of mangrove. Some 65 000 ha of mangrove in Thailand and 100 000 ha in Vietnam, for example, are believed to have been converted into shrimp ponds in recent decades (Primavera, 1998). As a result of seepage, areas of land adjacent to seawater ponds are also often subject to salinization, adversely affecting agricultural activities (Phillips *et al.*, 1993; Primavera, 1998). The removal of mangroves not only represents the loss of a unique habitat type, but it also reduces the availability of food, fuel and building materials for the poorer sectors of coastal communities. Through its role as a nursery habitat for commercially important offshore species and as an important intermediary governing the flow of nutrients and soil from the land to the sea, the loss of mangroves adversely affects coral reefs and offshore fisheries.

Aquaculture uses large quantities of water, use being determined by species and the intensity of the operation. Estimates of water used in semi-intensive and intensive shrimp culture, for example, are around 11 000 21 000 $m^3\,t^{-1}$ and 29 000 43 000 $m^3\,t^{-1}$, respectively (Phillips *et al.*, 1993); requirements for salmonids are typically even higher. Losses from ponds through seepage and evaporation can amount to several hundred cubic meters of water per hectare pond area per day. However, although it often makes greater economic sense to use water for purposes other than aquaculture where it is in short supply, a fish pond on a family farm can also improve water conservation and use.

Most farmed freshwater species are reared in captivity throughout their life cycle. This is desirable from the farmer s point of view since it facilitates selection for economically important traits and is a prerequisite in the manipulation of sex and in the production of year-round supplies of seed. By contrast, marine fish and shrimp culture is still largely dependent upon the supply of seed (eggs, larvae, fry, juveniles or, in the case of shrimp, gravid females) from the wild. Mollusk farming uses both wild-caught and hatchery-reared seed, depending on species and circumstances, but this is unlikely to have much impact on the environment. While undoubtedly there remain many technical barriers to breeding, the economics of production also often do not justify the development of commercial marine fish and shrimp hatcheries. Although the wastage associated with collecting shrimp post-larvae and milkfish fry is widely recognized as tremendous, the vast proportion of material caught being discarded as by-catch (Beveridge *et al.*, 1997a,b; Primavera, 1998), impacts on stocks and on biodiversity are difficult to evaluate given that most stocks exploited in this way are also subject to over-fishing.

Much fish and shrimp culture is dependent upon the provision of trash fish or fishmeal and fish oil-based feeds. It takes four tons of wild fish (mainly anchovies and sardines) to produce one ton of fishmeal. In 1997, some 2.3 million tons of fishmeal and 670 thousand tons of fish oil, about one third of world supplies, were used in aquaculture feeds. Although farming of carnivorous fish such as salmon results in a net loss of fish, Naylor *et al.* (2000) estimate that the production of the 10 most important farmed aquatic species is approximately 1.8 times greater than the quantities of marine fish used in their production. It is also argued that industrially harvested species are fit for no other purpose than reduction to fishmeal.

Irrespective of production method, the establishment of an aquaculture operation can impact on both scenic value and biodiversity. Cage or mussel farming structures act as feed attractant devices (FADs) or substrates for other species while the stocks of fish, shellfish and feed may attract predatory or scavenging fish, birds and mammals (Beveridge *et al.*, 1997b). Immigrants may impose increased predation pressure on wild fish populations in the vicinity of the farm or compete with and displace resident breeding populations, as has been claimed with regard to bird populations in areas in Scotland. Losses of farmed stock to predators can be economically significant and the measures used as deterrents often lead to death, either deliberate (shooting) or accidental (entanglement). Increased levels of activity around farms in isolated areas may disrupt species sensitive to the presence of humans and noise.

Feed losses vary with species and system. Due to currents, losses from marine cages, typically 5 10% of food fed, are generally higher than from other systems. Faecal production from fish and crustaceans is around 230 400 $g\,kg^{-1}$ food ingested. In ponds, much of the solid material remains trapped until harvesting. However, in cages, uneaten food and faecal material continuously falls through the nets towards the sea bottom. As it does so, it tends to break up, increasing the leaching of nutrients into the water column and the area of seabed over which the solids disperse. The major source of dissolved nitrogenous waste, however, is excretory products. Depending on species, dietary protein content and environmental conditions, urinary nitrogen can account for up to 60% of nitrogen released into the environment by intensive fish production (Beveridge *et al.*, 1991).

Fish and shrimp farm effluents and their impacts have been well studied. While hypernutrification, eutrophication

and oxygen depletion of the water column around cage farms is sometimes apparent, given the degree of water movement at most coastal sites, effects are generally only apparent during periods of slack tide. Impacts on the seabed, especially around cage sites, are more readily observable. Loss in sediment habitat heterogeneity is apparent, the area under or immediately downstream of the cages being most affected. Sediments here are largely devoid of oxygen, are highly reducing, and may be devoid of macrobenthos. The anaerobe *Beggiotoa* is often apparent on sediment surfaces. Some 10 20 m distant from the cages, pollution tolerant macrobenthic species such as *Capitella* are found in high numbers. There is usually a return to undisturbed community structure and function within 100 150 m (*see* Aquaculture: Salmon Farming, Volume 3).

Although farmed mussels and other filter feeding mollusks are net removers of nutrients, rafts and long-lines increase rates of biodeposition in the immediate vicinity, reducing biodiversity within the benthic macroinvertebrate community.

A wide range of chemicals is used by the mariculture industry, including compounds found in construction materials (stabilizers, pigments, plasticisers, ultraviolet (UV) absorbents, antifoulants), pigments incorporated into feeds, and chemotherapeutants. There is particular concern about chemotherapeutants because of the amounts used, the ease with which they enter the environment and the lack of knowledge of their effects. Approximately 30 antibacterial compounds are used in aquaculture (Weston, 1996), use being determined by farmed species, pathogens, farm location and intensity of production methods. Greatest use is in marine finfish farms in Europe, Japan and North America and Asian and Latin American shrimp ponds. The quantities used can be enormous. Norway has kept records of antibacterial use by fish farmers for more than a decade (Beveridge *et al.*, 1997b). Quantities used peaked at 49 tons in 1987, but have since declined to only a few hundred kilos per year through the development of vaccines and better disease management.

The bulk of aquaculture production is from a few dozen species, many of which have been translocated. The Pacific cupped oyster *Crassostrea gigas*, for example, is farmed throughout much of the world, far outside its natural range. Aquaculture, however, remains a comparatively minor factor in relation to other causes of species introduction, both deliberate (fisheries development) and accidental (construction of canals, ballast water transport, etc.) (Beveridge *et al.*, 1997b).

The transfer of exotic aquatic organisms poses two threats to biodiversity, from the organism itself and from associated pathogens. Viral pathogens of shrimp and salmon and protozoan parasites of mollusks have been translocated in this way. Farmed organisms will always escape, cages being particularly vulnerable. Escapes occur on a small but almost continuous scale during day to day management of stock (grading, stock movements). Nevertheless, numbers lost in this way may greatly exceed those of neighboring wild populations. Moreover, storm damage or vandalism can result in the periodic releases of many thousands of animals. The fear is that feral organisms become established and reduce biodiversity through habitat modification, competition and predation and by inter-breeding with native stocks (Gross, 1998). Many feral species have become established in freshwaters (Welcomme, 1988). While there is little evidence that exotic feral farmed marine organisms have become established (Beveridge *et al.*, 1997b) or that serious impacts on the environment or on biodiversity have resulted, there has also been little study of the fate of such organisms. In Chile, for example, there is no information on the impacts of the many hundreds of thousands of farmed salmonids that undoubtedly escape each year.

The farming of native species reduces risks of habitat modification and interspecific interactions but may introduce non-adaptive traits (gene combinations) to indigenous stocks, reducing fitness and endangering populations. Farming of local strains may not always be a solution. Native animals brought into the hatchery pass through various selection bottlenecks, either accidental (disease) or deliberate (selection for growth, shape, colour), and within a few generations differ from their wild ancestor (*see* Aquaculture: Salmon Farming, Volume 3).

DISCUSSION

Aquaculture is an increasingly important means of producing aquatic food. It is also apparent that the increases in production and the trend towards intensification are not without cost. There has been insufficient analysis to determine the real contribution of aquaculture to global food supplies or to assess whether it is a more or a less efficient means of producing food than agriculture. However, the more intensive types of aquaculture undoubtedly result in a net loss of fish or protein and despite efforts to find substitutes, demand for fishmeal and fish oil by the aquaculture industry is likely to remain largely unchanged for some time. Aquaculture wastes are characteristically dilute in nature and are difficult and costly to treat and there remain intractable disease problems, treatable only with noxious chemicals. Waste discharges into the environment can cause localized impacts, although their significance at the regional scale is contentious.

Many of the environmental problems have undoubtedly been due to ignorance; after all, modern aquaculture is only a few decades old. Lessons have been learned. Recent European Directives, for example, have led to empowerment of regulatory authorities to demand that environmental impact assessments (EIA) of proposals to site particular types of

aquaculture development in sensitive areas in Ireland, Scotland, Spain and Italy be carried out. Greater control of fish farm effluents, and thus expansion of production, has resulted from reclassification of Scottish salmon marine farms as industrial sites and the necessity to obtain consents to discharge. The rate of destruction of mangroves for aquaculture also seems to be waning. Legislative measures to protect forests are being implemented at last. Areas in Vietnam and Thailand are being replanted and more integrated aquaculture mangrove management projects encouraged. Wastes per unit production have been decreasing in many sectors through a combination of production economics, increased public scrutiny of farming methods and the influence of producer organizations, such as the Scottish Quality Salmon. Initiatives, such as the Global Biodiversity Convention, have led to the development of the Code of Conduct for Responsible Fisheries that includes guidelines pertaining to aquaculture (FAO, 1997). These are increasingly being considered in the regulation of the industry. The consumer too has recently entered the debate, and that more than anything is likely to further improve farming methods.

REFERENCES

Beveridge, M C M (1996) *Cage Aquaculture*, 2nd Edition, Fishing News Books, Oxford.

Beveridge, M C M and Phillips, M J (1993) Environmental Impact of Tropical Inland Aquaculture, in *Environment and Aquaculture in Developing Countries*, eds R S V Pullin, H Rosenthal, and J L Maclean, ICLARM, Manila, 213 236.

Beveridge, M C M, Phillips, M J, and Clarke, R M (1991) A Quantitative and Qualitative Assessment of Wastes from Aquatic Animal Production, in *Advances in World Aquaculture*, Vol. 3, eds D E Brune and J R Tomasso, World Aquaculture Society, Baton Rouge, LA, 506 533.

Beveridge, M C M, Phillips, M J, and Macintosh, D C (1997a) Aquaculture and the Environment: the Supply and Demand for Environmental Goods and Services by Asian Aquaculture and the Implications for Sustainability, *Aquac. Res.*, 28, 101 111.

Beveridge, M C M, Ross, L G, and Stewart, J A (1997b) The Development of Mariculture and its Implications for Biodiversity, in *Marine Biodiversity: Patterns and Processes*, eds R F G Ormond, J Gage, and M Angel, Cambridge University Press, Cambridge, 372 393.

Bray, F (1984) Biology and Biological Technology. Part II: Agriculture, in *Science and Civilization in China*, Vol. VI, Cambridge University Press, Cambridge.

FAO (1997) *Aquaculture Development*, FAO Technical Guidelines for Responsible Fisheries, No. 5, FAO, Rome.

FAO (1999) *Aquaculture Production Statistics (1988- 1997)*, FAO Fisheries Circular, 815 (Rev. 11), FAO, Rome.

Gross, M R (1998) One Species with Two Biologies: Atlantic Salmon (*Salmo salar*) in the Wild and in Aquaculture, *Can. J. Fish. Aquat. Sci.*, 55 (Suppl. 1), 1 14.

Harris, D R (1996) *The Origins and Spread of Agriculture and Pastoralism in Eurasia*, Smithsonian Institution Press, Washington, DC.

Naylor, R A, Goldburg, R J, Pimavera, J H, Kautsky, N, Beveridge, M C M, Clay, J, Folke, C, Lubchenco, J, Mooney, H, and Toell, M (2000) Effect of aquaculture on world fish supplies, *Nature*, 405, 1017 1024.

Phillips, M J, Lin, K, and Beveridge, M C M (1993) Shrimp Culture and the Environment Lessons from the World s Most Rapidly Expanding Warm Water Aquaculture Sector, in *Environment and Aquaculture in Developing Countries*, eds R S V Pullin, H Rosenthal, and J L Maclean, ICLARM, Manila, 171 197.

Primavera, J H (1998) Tropical Shrimp Farming and its Sustainability, in *Tropical Mariculture*, ed S S De Silva, Academic Press, London, 257 290.

Rana, K J, Beveridge, M C M, and Grainger, R (1999) Review of the Definition of Aquaculture, in *Proceedings of FAO Regional Workshop on Aquaculture Statistics Reporting, Bangkok, Thailand*, FAO, Rome.

Tacon, A (1998) FAO Aquaculture Production Update, *Int. Aquafeeds*, 2(1), 13 16.

Welcomme, R L (1988) *International Introductions of Inland Aquatic Species*, FAO Fisheries Technical Paper, No. 294, FAO, Rome.

Weston, D P (1996) Environmental Consideration in the Use of Antibacterial Drugs in Aquaculture, in *Aquaculture and Water Resource Management*, eds D J Baird, M C M Beveridge, L A Kelly, and J F Muir, Blackwell, Oxford, 140 165.

Aquaculture in Asia

Fumio Takashima

Tokyo University Fisheries, Tokyo, Japan

The annual global production by traditional sheries was around 93 million metric tonnes in 1997, whereas aquaculture production in the same year was 36 million tonnes. In many parts of the world, and especially in Asia, aquaculture is becoming a major industry. Asian aquaculture production has been increasing year by year and attained 32.5 million tonnes (about 90% of the world's production) in 1997. About half of it was produced in freshwater and the other half in coastal areas. The main freshwater species produced by aquaculture in Asia are the carps (common carp, Chinese carps and Indian carp), tilapia, eel and cat sh in warm water, and rainbow trout and coho salmon in cold water. Marine species such as yellowtail, seabream, ounder, grouper, milk sh and mullet are important species in many countries. Certain species of shell sh such as tiger shrimp, oyster, clam and abalone are also widely cultured. Among the sea algae, laver, Japanese kelp and wakame are the main species in some countries (Japan, Korea and

China) whereas agar- agar predominates in others (especially Southeast Asian countries).

HISTORY

The common carp originated in Central Asia and was transported into Europe, China and West Java more than 3000 years ago. The technical textbook on carp culture written by Fun Li in China in BC 500 indicates that aquacultural techniques were already developed in ancient times. Milkfish culture in Indonesia also has a long history, fry for culture having been collected along the beaches in Java for more than several hundred years. Rainbow trout, first introduced from North America nearly 100 years ago, is now widely distributed wherever cold water is available. Tilapia was introduced from the Nile River via the Middle East and is now common in Asian countries where warm water is abundant. Culture techniques of eel and tiger shrimp were developed after the Second World War in Asian countries. Several native species can now be cultivated in some countries thanks to the efforts of local researchers and technicians.

According to the Food and Agriculture Organization s 1997 statistics, China is the leading country in aquaculture production with about 24 million tonnes, and Japan was second with about 1.3 million tonnes. In addition, Korea (1 million tonnes), Philippines (958 000 tonnes), Indonesia (912 000 tonnes), Thailand (576 000 tonnes) and Taiwan (270 000 tonnes) produced abundant fish, shellfish and sea-algae.

COMMON CULTURING TECHNIQUES

The aquaculture systems in Asia may be classified in terms of their efficiency of cultivation into three groups: extensive, semi-intensive and intensive aquaculture. The systems may be further classified by means of the type of installation (pond, cage, partition or raft) and of water movement (stagnant, running or circulating water). In some cases, the rearing systems are divided into feeding or non-feeding conditions.

AQUACULTURE IN ASIAN COUNTRIES

China

China not only has one of the largest freshwater and coastal areas in the world, but also the longest history in raising fish. China has developed intensive and integrated polyculture systems that are cost-effective, highly productive and less dependent than the fishing industry on energy sources and modern technology. The achievements of Chinese aquaculture have made China the largest producer of cultured fish in the world. The major approaches for the development of aquaculture in China are: polyculture systems, integrated fish farming systems, artificial propagation of cultivated fish and large-scale production of fish seeds, expansion of culture areas and enhancement of productivity, and strengthening of the prophylaxis and therapeutics of diseases. China produced 24 million tonnes of fish, shellfish and algae in 1997. In this country, extensive culture in warm fresh water is predominant in stagnant ponds, reservoirs and paddy fields. Common carp (*Cyprinus carpio*) and Chinese carps such as grass carp (*Ctenopharyngodon idellus*), silver carp (*Hypophthalmichthys molitrix*), bighead carp (*Aristichtys nobilis*) and black carp (*Mylopharyngodon piceus*) are the dominant species, comprising a production of about 9 million tonnes (74% of all freshwater species). Polyculture is the main method of freshwater aquaculture and has a long history in China. Mixed culture of the four major Chinese carps has been practiced since as early as the end of the Tang Dynasty, about 1000 years ago. The major advantages of polyculture are the efficient use of the space and food available in the pond, the increase in the utilization rate of artificial feeds, and the maximization of the beneficial interactions between compatible species with different habits and ecology. The integrated fish farm is another system of non-feeding aquaculture for omnivorous and herbivorous fish such as common carp and some of the Chinese carps. This system raises livestock such as pigs, ducks or poultry and uses their manure for increasing the abundance of natural food organisms (plankton). The same system is also employed in Southeast Asian countries such as India and Indonesia.

Indonesia

Aquaculture in Indonesia is mainly practiced in fresh and brackish water. The total freshwater fish production in 1997 was 357 000 tonnes, with paddy fields and net cages contributing almost 90% of the production. Freshwater cultivation is also done in stagnant and running water ponds and in bamboo cages in rivers. The rice paddy and net cage cultures are related, in that the farmers rear fish fry up to young (100 g) in the former, and from young to marketable size in the latter. Net cage culture is well developed in some reservoirs in West Java but it is not so common in other areas. Common carp is the dominant species cultured in reservoirs and paddy fields. Tilapias (*Oreochromis niloticus* and *Oreochromis mossambicus*), giant gouramy (*Osphronemus goramy*), Java barb (*Puntius javanicus*) and catfish (*Clarias* spp.) are also important cultured species. The numerous net cages in some reservoirs cause great accumulation of organic waste at the bottom. The suspension of this waste during the rainy season, due to massive water inflow from rivers and convection currents created by temperature changes, causes a sudden decrease of dissolved oxygen and provokes the suffocation of fish. Brackish water aquaculture produced 398 000 tonnes in 1997. The cultures of the

main species, milkfish (*Chanos chanos*, 168 000 tonnes) and tiger shrimp (*Penaeus monodon*, 100 000 tonnes), have a long history and are widespread in Indonesia. Shrimp and milkfish farming is developed in ponds (tambak) built in coastal areas after taking off the mangrove trees. A part of the mangrove tree is sometimes left in the pond to help preserve the environment and this system is known as *tumpangsari*. Tumpangsari use is gradually spreading to other countries. In addition, the biocrete system of shrimp farming ponds has been developed along the southern coast of West Java, in order to avoid the spreading of serious diseases such as that evoked by Monodon baculovirus. In this system, the bottom of the pond is covered with a plastic sheet and the dyke is also covered with a mixture of bamboo, palm fiber, sand and cement. Coupled with technologies such as formulated feed, water quality management, selective breeding, disease control and hatchery techniques, shrimp culture in Indonesia (and also in other countries) has become more efficient, producing more shrimp in less culture area. Farmers gradually have come to realize that intensive shrimp culture must be sustainable and that mangrove destruction for construction of shrimp ponds should be gradually reduced. In the coastal areas of Indonesia, without muddy outflow from rivers, sea-algae for agar agar (Rhodophyceae) is propagated and harvested (157 000 tonnes).

Japan

Most Japanese aquaculture is highly intensive and uses advanced equipment such as net-cages, partitions, raceway ponds and recirculation ponds. Net-cages are square or round in shape and usually consist of a steel frame surrounded by a net. Net-cages are used for yellowtail (*Seriola quinqueradiata*), red seabream (*Pagrus major*), coho salmon (*Oncorhynchus kisutch*) and common carp whereas partitions by net are used for yellowtail and kuruma shrimp (*Penaeus japonicus*). Raceway ponds are common for salmonids. Water recirculation systems are used for eel (*Anguilla japonica*), ayu smelt (*Plecoglossus altivelis*) (freshwater) and olive flounder (*Paralichthys olivaceus*). The stocking density in each system is different depending on the species, the volume of water contained and the exchange rate of the water, which affects the supply of dissolved oxygen. Releasing the juvenile or fry into open waters is called sea farming, and this technique is now quite developed for salmon (*Oncorhynchus* spp.), red seabream, kuruma shrimp, abalone (*Haliotis* spp.), swimming crab (*Portunus trituberculatus*) and ayu smelt. Suspension of shells with spats of oysters (*Crassostrea gigas*) in the sea is a unique method originated in Japan about 100 years ago. This technique was adopted for scallop (*Patinopecten yessoensis*), pearl oyster (*Pinctada* spp.) and sea squirt (*Halocynthia roretzi*) with the modification that juveniles or young are suspended under water inside net bags. Floating nets with spores of laver (*Porphyra ezoensis*) or wakame (*Undaria* spp.) have been recently developed and have started to yield considerable amounts of these species of edible algae. In most cases of finfish and kuruma shrimp farming, artificial food made of fish and vegetable meal and supplements including minerals and vitamins are used. Disease is the most serious problem in the case of highly intensive culture systems and this problem has been tackled with drugs and vaccination programs sanctioned by the government. Genetic control of disease is also being pursued through the experimental introduction of resistant genes into the animals. Studies are now in progress to adapt the *hazard analysis critical control point* (HACCP) concept to aquaculture, and in the near future fish farm products will be required to be HACCP-compliant. This system is a systematic approach to assure the safety of foods that focuses on prevention (early identification of potential biological, physical or chemical hazards to human health, control, and monitoring) rather than remediation of problems. Some of the main species in 1997 were yellowtail (138 000 tonnes), red seabream, (81 000 tonnes), laver (393 000 tonnes), wakame (70 000 tonnes), kelp (60 000 tonnes), scallop (254 000 tonnes), oyster (218 000 tonnes), common carp (12 000 tonnes) and rainbow trout (*Oncorhynchus mykiss*, 13 000 tonnes), among others.

Korea

The Korean aquaculture industry grew rapidly from 1960 until the late 1980s. During this initial stage of development, non-feeding aquaculture species such as common carp and crucian carp (*Carassius auratus*) represented the bulk of the production. Recently, marine finfish, such as olive flounder and red seabream, are farmed under the feeding system and represent nearly 90% of the total production of farmed marine fish. In addition, the production of shellfish such as oyster (201 000 tonnes) and sea-algae (wakame, laver and kelp, 613 000 tonnes) has recently increased. Rainbow trout, Japanese eel, ayu smelt, tilapia and catfish (*Ictalurus punctatus*) have been introduced from abroad, and reared in various parts of the country. The continued growth of the aquaculture industry in Korea is gradually being constrained by deterioration in the quality of water resources, caused by urban and industrial activities as well as aquaculture itself, and market competition within the country.

Philippines

In 1997, the production from aquaculture reached a total of 958 000 tonnes, of which 65% was from seaweeds, 27% from fin fish, 5% from crustaceans and 3% from shellfish.

Milkfish is the most important finfish for aquaculture in the Philippines. Farming of milkfish has gradually shifted from the extensive method to semi-intensive and intensive methods. Cage culture of milkfish, tilapia and other commercially important species is found both in the sea and in fresh water (lakes and rivers). High stocking-density farming in pens and cages is now common. Tiger shrimp farming increased sharply from 1992 1994 but dropped thereafter. Due to the commercial importance of mud-crab (*Scylla* spp.), the aquaculture of this species is now increasing and production reached about 4000 tonnes in 1997. The production of tilapia developed greatly in the last 10 years and is currently about 92 000 tonnes, making the Philippines one of the world s leading producers of tilapia. Genetically developed hybrids of tilapia such as monosex individuals or YY-male are fast gaining consumer acceptance and are being rapidly distributed among the producers. The major commodities produced in mariculture are the seaweeds *Eucheuma* spp. and *Caulerpa* spp., oysters (*Crassostrea* spp.), and mussels (*Mytilus* spp.). In 1997, seaweed production totaled 627 000 tonnes, that of oysters 14 000 tonnes, and that of mussels 12 000 tonnes.

Thailand

Aquaculture techniques in Thailand are fairly advanced, particularly concerning seed production, where the application of techniques for induced spawning has become popular. In 1997, the total production from aquaculture was 576 000 tonnes including the products from fresh water (265 000 tonnes), brackish water (222 000 tonnes), and marine water (89 000 tonnes). Freshwater culture is practiced in four main types of installations, namely paddy fields, ditches, and cages in ponds or lakes. The major species in pond culture are the tilapias, gouramy (*Trichogaster pectoralis*), catfishes (*Clarias gariepinus* and *C. macrocephalus* hybrids, *Pangasius pangasius*), and silver barb (*Puntius gonionotus*). Tilapia has a particularly good potential for future development. Brackish water culture is practiced in the coastal area and concentrates on tiger shrimp culture.

Thailand has the highest shrimp production of the Asian countries (Table 1). Shrimp culture systems in Thailand have three levels of intensity: extensive, semi-intensive, and intensive. Extensive culture systems are characterized by low stocking densities and, therefore, lesser productivity. This system is characteristic of coastal communities with little income. Semi-intensive culture ponds may be created through the intensification of existing extensive ponds. This may include improvements such as clearing, leveling and/or digging canals in the pond bottom, converting from the mere use of tides to the intensive use of pumps for water exchange; controlling stocking density, greater use of fish toxicants to control nuisance species, and provision of feeding. The intensive system for shrimp farming is very similar to the Taiwanese system. The adoption of new technologies allowed a reduction in the production time to four months per crop and raised the production rate to 6000 10 000 kg ha^{-1} $year^{-1}$. Intensive farms do not harvest wild juveniles but rely on the large numbers of fry produced in hatcheries. Due to the high stocking density and the use of supplement feeding, there is a large organic accumulation at the pond bottom after a period of culture. This requires complete drying between successive cultures to facilitate removal of waste and sanitation. The brackish water culture, especially that of shrimp, apparently has reached the maximum sustainable yield. The main species for mariculture in Thailand are shellfish such as blood cockle (*Anadara granosa*), mussel (*Mytilus smaragdinus*) and oysters (*Crassostrea* spp.). There is a great demand for these shellfish in the domestic market.

Table 1 Tiger shrimp production in Asia in 1997

Countries	Production (metric tonne)
Thailand	211 100
Indonesia	99 680
Vietnam	60 000
India	54 483
Philippines	40 102
Malaysia	9380
Taiwan	5161

Taiwan

Aquaculture in Taiwan has a history of more than 300 years and has expanded rapidly in recent times due to factors such as the vast and ever-increasing demand for seafood, highly skilled and motivated aqua-farmers, excellent support from research and extension services and thoroughly integrated peripheral industries and side-business. The high diversity and availability of cultured species (around 200 species are presently cultured) also attest to the significance of the Taiwanese aquaculture industry. The total production in 1997 was 270 000 tonnes. The major species cultured in Taiwan are finfish such as Japanese eel (22 000 tonnes), milkfish (31 000 tonnes) and tilapia (42 000 tonnes). Eel farming is typically an intensive industry with great importance in Taiwan, ranking top in aquaculture in terms of value and quantity of production. Eel ranks top also among exports of agricultural products and most of it is exported to Japan. Milkfish culture was developed in Taiwan about 300 years ago. In general, the management methods of milkfish culture can be divided into the traditional shallow-water culture and the modern deep-water pond culture systems. The former method is based on enhancing the production of benthic algae as food for milkfish by administration of organic fertilizer to shallow (30 40 cm) brackish or saline water

ponds. It is practiced extensively: no feeding, no aeration, with low stocking density and as a result of this, the maximum yield is around 2–2.5 tonnes ha^{-1}. The latter method is based on farming the fish in deep (2–3 m) brackish or fresh water ponds. It is practiced intensively, with the use of automatic feeding with pellet feeds, aeration, high stocking density, and usually produces maximum yields around 8–12 tonnes ha^{-1}.

Tilapia has now become not only the most common food fish in Taiwan but also an export commodity, with the main destination being Japan. Prior to shipment, fish are acclimatized in seawater to eliminate any parasites and the odor of mud that may impregnate the fish in fresh water. Fish are then filleted, vacuum-packed, and frozen. The species of tilapia most intensively cultured are *Oreochromis niloticus* and its red variety. They are mainly polycultured as the primary crop along with carp or eel, or integrated-cultured as a side-product of the production of hog or duck. Tiger shrimp culture developed rapidly in Taiwan up to 1987. However, the production has been decreasing since 1988 because of the outbreak of diseases such as Monodon baculovirus and those caused by bacteria or protozoa. Complicating factors are the absence of disease control techniques, deterioration of pond environment, overstocking, and the use of poor quality fry. As a result of these unfavorable conditions, the production of shrimp was only 5000 tonnes in 1997, while at its maximum in 1987 it was 79 000 tonnes. Taiwan experienced many of the typical problems posed by a rapidly growing agricultural activity such as the improper utilization of limited land and water resources, environment deterioration, disease outbreaks, and poor coordination between production and consumption.

IMPORTANCE OF AQUACULTURE IN THE 21ST CENTURY AND ENVIRONMENTAL ISSUES

Aquaculture has expanded rapidly since 1960 in Asian countries and now provides as much fish and shellfish as freshwater fisheries in the region. It accounts for about one-fifth of the total fish and shellfish production. Aquaculture is expected to make up the difference between the demand for fishery products and the output from conventional fisheries, which already show signs of stagnation or even decrease. Its importance as a source of animal protein in Asia cannot be underestimated in view of the increase in the region's population predicted for the first half of 21st century.

In order to meet these expectations and expand any further, however, Asian aquaculture will have to address the concerns of the consumers regarding the safety of the product and of society as a whole regarding environmental protection. The introduction of exotic species and the use of genetically modified organisms, although justifiable from a production efficiency viewpoint, deserve careful examination. The adoption of the HACCP concept, as explained previously, and the use of biologically reasonable culture practices (e.g., adequate size of facilities, rearing density, supplemental food, etc.) that help prevent the outbreak and spreading of diseases should help allay fears of contamination of the food with any hazards, biological or chemical. These culture practices should also help mitigate environmental problems caused by aquaculture. For example, aquatic farming in coastal areas, especially of shrimp, has driven much of the transformation of mangrove forests in the region and in some areas, such as Thailand, has nearly reached the point above which further expansion would be unsustainable (*see* **Aquaculture and Environment: Global View from the Tropics to High Latitudes**, Volume 3). Culture systems that minimize environmental impacts in mangrove areas, such as the tumpangsari described previously, should be encouraged and the planning of new farms should take in consideration the need for intervening, buffering natural areas.

The expansion of industrial and urban activity, including reclaiming of coastal areas and resort development, is progressively taking away potential sites for aquaculture. Aquaculture also competes with those activities for water, with conflicts arising due to shortage or deterioration in water quality. However, aquaculture is not only a victim, but also a cause of water pollution. In recent years, pollution through excessive feeding has become a serious problem in the effluents of aquaculture farms, which are carried to open waters, or around culture pens. The use of chemicals for the control of algae in ponds is also a potential source of contamination of surface and ground waters. Thus, aquaculture development must deal properly with the potential impacts of organic and inorganic wastes. Many of the above problems might be mitigated by future developments in the field of closed-circulation culture systems that minimize water intake and permit complete control of liquid and solid wastes.

See also: **Aquaculture: Salmon Farming**, Volume 3.

Aquaculture: Salmon Farming

Trevor C Telfer and Malcolm C M Beveridge
Institute of Aquaculture, Stirling, UK

Although the farming of salmon in sea cages began little more than 30 years ago, it has since grown into a multi-billion dollar industry, employing many thousands of people

in rural areas of western Europe, North America and Chile. Farmed salmon is highly nutritious and is now one of the cheapest shes available to western consumers. Despite this, salmon farming has attracted criticism on account of its reliance on shmeal and sh oil-based diets and the impact of the wastes released into the marine environment.

INTRODUCTION

The seven species of salmon (Family Salmonidae) are all indigenous to the Northern Hemisphere (Table 1). For centuries, salmon have been an integral component of the economic, cultural and religious life of the peoples inhabiting the Northwest Pacific coast of America. In Europe, too, the salmon was of considerable economic importance, exploitation often being regulated by legislation. Many stocks of both Atlantic and Pacific salmon, however, are now sadly depleted.

While it is believed that native North Americans may have attempted to manage stocks by transplanting fertilized eggs, salmon farming has its origins in the early 19th century and was precipitated by factors other than farming. In Europe, the rise of leisure among the upper middle classes, combined with the Victorian drive to improve upon nature, led to salmon being successfully spawned in Scotland in the 1830s and subsequently, salmon hatcheries being built on a number of rivers to enhance recreational fisheries (Williamson and Beveridge, 1994). By the beginning of the present century, however, many of the hatcheries had been abandoned due to expense and questionable returns of stocked fish. In Pacific North America, salmon aquaculture was promulgated in the late 1870s in response to concerns about over-fishing, although they were often unable to remedy the deleterious effects of logging and mill and dam construction (NRC, 1996). Salmon farming to produce food for the table, however, only began in the 1960s in Norway but was swiftly followed by developments in Scotland and Ireland. Salmon farming has since spread to more than a dozen countries, including Canada and the United States and even as far as Chile and Australia in the Southern Hemisphere (see Figure 1). Farmed salmon production now approximates that from *capture* (conventional) fisheries, although production of farmed Atlantic salmon, which accounts for nearly 90% of farmed salmon production, has long outstripped capture fisheries production. Salmon farming, however, has done little to alleviate exploitation of wild salmon stocks, as capture fisheries landings remain little changed.

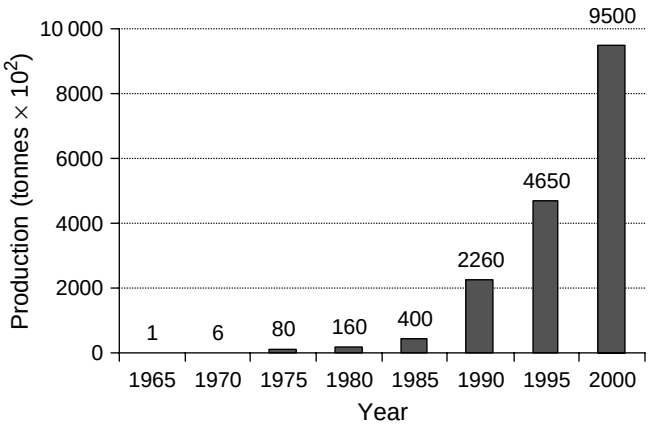


Figure 1 Farmed Atlantic salmon production, 1965–2000. FAO figures (FAO, 1999b) are used for the period 1985–1997; earlier data are from the literature

SALMON BIOLOGY AND FARMING METHODS

Although they begin life in streams and rivers, salmon are anadromous and migrate to the sea, where most of their growth occurs, after a variable period (months years) in freshwater. At the onset of sexual maturity, the fish migrate to their home rivers to spawn. The Pacific salmon die after

Table 1 Salmon and their distribution

Species	Common name	Distribution
Pacific salmon		
Oncorhynchus gorbuscha	Pink, humpy salmon	Rivers flowing into the rivers of the Pacific coast of the Northwest America and the Kamchatka peninsula, Russia
O. nerka	Sockeye	
O. kisutch	Coho, silver	
O. keta	Chum, dog	
O. tschawytscha	Spring, chinook, king	
O. masou	Masu, cherry salmon	Rivers of northern Hokkaido and the Manchuria peninsula
Atlantic salmon		
Salmo salar	Atlantic salmon	Rivers of countries bordering both sides of the northern Atlantic

spawning. Atlantic salmon, however, may return several times to breed. The farming cycle is species and location specific, but is described for Atlantic salmon. Captive, sexually mature broodstock are stripped of eggs and milt in late autumn, the gametes mixed and the fertilized eggs transferred to a hatchery. The eggs hatch in spring and, once the yolk sac has been absorbed, the fish are transferred to tanks and fed on artificial pelleted diets. By the autumn, some 6 months after hatching, a bimodal size distribution is apparent, the larger fish indicating their readiness to smolt and transfer to seawater the following spring. Thus, by the time they are 1 year old, and weighing 35 80 g, the majority of fish will be ready to be transferred to seawater cages or tanks.

More than 90% of farmed salmon production is derived from floating net cages anchored in sheltered inshore areas, such as the fjords of Norway and Chile, the sea lochs of Scotland and sheltered bays and loughs in Ireland. Growth in the sea is rapid. Fish are fed on a largely fishmeal and fish oil-based pelleted diet and can attain a weight of up to 2 kg after one sea winter. Disease and predators (birds, seals and sea lions) take their toll. The tremendous increases in production have led to increasingly saturated markets, driving down prices. Shrinking profit margins have led to wide-scale vertical integration of the industry (i.e., from hatcheries, through production, processing and marketing) and dominance by multinational companies. Production economics have led to year-round production of smolts through photoperiod manipulation, improved diets and feeding methods.

SALMON FARMING AND THE ENVIRONMENT

The relationship between aquaculture and the environment is described elsewhere (*see* Aquaculture and Environment: Global View from the Tropics to High Latitudes, Volume 3; Aquaculture in Asia, Volume 3). Critics point to salmon farming s reliance on industrial fisheries products for food, and the wastes (uneaten food, fecal and urinary products, chemotherapeutants, pathogens and feral (escaped) fish) that are released into the marine environment.

Salmon farming remains almost entirely reliant on fishmeal and fish oil-based diets. Approximately five tonnes of fish are required to make one tonne of fishmeal. Commercial diets are composed of approximately 45% fishmeal and 25% fish oil and as a result it takes something like 2.8 tonnes of fish to produce 1 tonne of salmon. Although salmon farming is a net consumer of fish (Naylor *et al.*, 1998), it may be argued that the fish reduced to fishmeal has little other commercial value than as a source of animal food. It is also a moot point whether farming of salmon may be more energy efficient than the exploitation of wild salmon stocks by fisheries. However, within the aquaculture sector, salmon farming is one of the more important consumers of fishmeal and fish oil, arguably adding to pressures on already fully and over-exploited wild fish stocks (see FAO, 1999a for review of capture fisheries stocks).

Salmon farming wastes are derived from uneaten food, the products of digestion (feces) and metabolic activity (urinary products), and the chemicals used in the treatment of disease and prevention of fouling on the cage nets. Cages rely on maximizing the free exchange of water with the external aquatic environment in order to maintain good water quality to sustain fish growth. However, as a result, the wastes discharged from the cages freely enter the environment.

Waste emissions may be estimated by two methods: direct measurement and mass balance modeling. The former is difficult and time consuming, requiring deployment of sediment traps and analysis of recovered material. Provided that the food conversion ratio (FRC) (amount of food fed/net production of fish) and nutrient content of food and fish flesh are known, nutrient wastes from salmon farms can be estimated by mass balance (Equation 1):

$$X = (\mathrm{FCR} \times \mathrm{A}) - B \quad (1)$$

where X is the nutrients lost to the environment (kg tonnes^{-1}), A is the nutrient content of fish food (kg tonnes^{-1}) and B is the nutrient content of fish flesh (kg tonnes^{-1}).

With improvements in diet formulation, feed technology and feeding methods, nutrient losses per unit production have declined over time. Today, something like 50 60 kg total-nitrogen and 7 9 kg total-phosphorus per tonne farmed salmon production are released into the environment by European salmon farms, less than one-third of the nutrient inputs associated with the industry 25 years ago (see Black, 1998, for review). Such calculations are limited in their usefulness as they give little information on the form of wastes or their bioavailability. However, by including information on uneaten food and by modeling growth and production over time, more detailed information on waste discharges can be derived.

Wastes are principally in one of two forms: particulate (uneaten food and feces) and dissolved (urinary products, leachates from feed and feces). The denser food particles tend to settle quickly and impact on the seabed close to the cages while the less dense feces settle farther from the cages (Chen *et al.*, 1999a, b). Impacts on the environment, especially the benthos, are readily discernible. Sediments under and immediately adjacent to salmon cages typically show signs of organic enrichment, with increased carbon and nitrogen levels (Black, 1998; Midlen and Redding, 1998). Considerable effort has been put into the development of models that can predict the fate of particulate wastes from salmon farms (Silvert, 1994), a pre-requisite to

the appropriate siting and sizing of cage farms. The effects of organic enrichment are to change benthic community composition and metabolism. Typically, macroinvertebrate abundance changes with degree of enrichment, abundance being generally highest at moderate levels of enrichment. There is a decrease in macroinvertebrate community diversity close to the cages and a gradual recovery to background levels as one moves away from the source of the pollution. Certain species, indicative of high levels of organic material, such as *Capitella capitata*, dominate community composition near the cages. Changes in benthic community diversity with distance from cages at a Scottish Atlantic salmon farm site are shown in Figure 2. The zone of impact is dependent upon size of farm, waste inputs, prevailing currents and depths, but typically extends for 25 120 m along the principal current axis (*see* **Aquaculture and Environment: Global View from the Tropics to High Latitudes**, Volume 3).

Copper-based antifoulants are still used on many salmon farms. Copper gradually leaches into the seawater environment and particles of copper-based paints can also be detected in sediments, although sediment concentrations remain low. Farmed salmon have proven susceptible to a range of bacterial diseases such as furunculosis and vibriosis. Gradually, however, reliance on antimicrobial compounds has been superseded by use of vaccines. These have seen quantities in Norway in 1999, for example, fall to 1% of their 1987 value, despite a 10-fold increase in salmon production over the intervening years (see Weston, 1996, for review of situation to 1993).

Some salmon-producing countries have serious problems associated with ectoparasitic copepod (*Caligus elongatus*, *Lepeophtheirus salmonis*) infections and are reliant on treatment with specially formulated pesticides. These

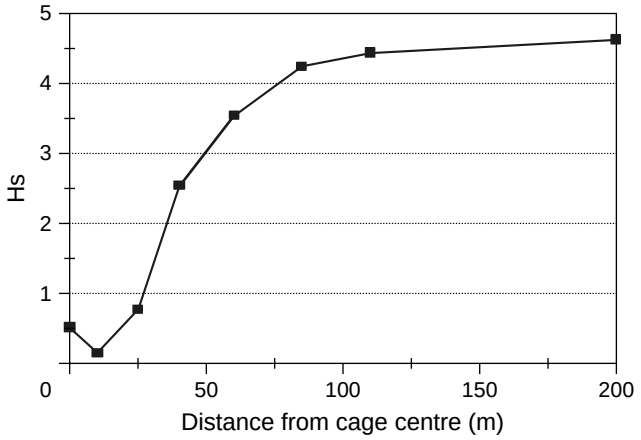


Figure 2 Diversity (Shannon–Weiner Index, Hs) for sediment macrofaunal communities versus distance at a salmon cage site, Scotland. The lower the diversity index value, the more dominated the community is by a few species

are used in two forms: bath and in-feed treatments. The bath treatment *dichlorvos*, for example, is introduced into an enclosed area through which the infected sh swim. Following treatment, the drug is released into the environment. Concerns about its environmental impacts (Lebris *et al.*, 1995; McHenery *et al.*, 1996) have led to the development of in-feed alternatives that are incorporated into the pelleted salmon diets. Release to the environment is from uneaten food or in fecal or excretory material and takes place over a longer time period. However, little is yet known about the ecotoxicology of these compounds. More environmentally benign alternatives, such as hydrogen peroxide and the use of the cleaner *wrasse*, have not proven suf ciently practicable to be widely adopted.

Fish escape from cages in small numbers as a result of day-to-day management of stock. Periodic mass releases, involving tens of thousands of salmon, can also happen as a result of storm damage or vandalism (Beveridge, 1996). In Chile, salmon are exotic and the fate and impact of feral farmed salmon on wild shes are unknown (Buschmann *et al.*, 1996). Atlantic salmon, were introduced to British Columbia for farming purposes in the late 1980s and it is now apparent that feral sh are appearing in sheries catches (McKinnell *et al.*, 1997) and are breeding in the rivers. Even within the home range of the Atlantic salmon, there are concerns about the impacts of feral farmed sh. In Norway, in 1997, more than 50% of Atlantic salmon recorded as caught by shermen were of farmed origin, while in the Faroe islands, as many as 60% of recorded salmon catches in recent years are of farmed origin (Hansen *et al.*, 1999). Genetically distinct, non-interbreeding populations of salmon occur in each river system. The fear is that feral farmed salmon will inter-breed with local populations, introducing genes that reduce the tness of individuals and putting the population at risk. A number of studies have shown escaped farmed salmon can cause long-term genetic changes in natural populations that affect both single-locus and highly heritable quantitative traits, such as growth and sea age of maturity (McGinnity *et al.*, 1997). However, while under some circumstances such changes may lead to reduced tness and productivity, declines in wild stocks are almost certainly due to a complex interaction of factors, including over- shing, habitat modi cation and climate change.

REGULATION OF THE INDUSTRY

Regulation of salmon farming varies from country to country. In Scotland, for example, regulation is governed by implementation of Directives from the European Commission in Brussels. This requires that an Environmental Impact Assessment is conducted by a competent body for each new salmon farm site >100 tonnes biomass or increase in size of an established farm. In addition, as ef uent from

salmon cages is considered an industrial discharge, it is modeled to estimate environmental impact and the impacts monitored by government regulators, such as the Scottish Environment Protection Agency, in the same way as effluents from other industries. Use of pesticides in salmon farming is carefully controlled in Scotland, with each farm site requiring a discharge consent for each chemical that it intends to use. Environmental regulation is usually enforced by granting a maximum production biomass of fish for each site, but may also take account of environmental management procedures, such as the introduction of a fallowing regime. This inevitably controls the amount of effluent entering the environment and allows site recovery between fish production cycles.

Careful site selection is an important part of the regulatory process, allowing new fish farms to be located in areas where environmental capacity will not be exceeded. Selection criteria include projected size of the fish farm (in terms of both production tonnage and area), water depth, water movement, other uses of the water body, proximity to other developments and whether the site is environmentally sensitive (in conservation areas or important wild salmon and trout streams).

REFERENCES

Beveridge, M C M (1996) *Cage Aquaculture*, 2nd edition, Fishing News Books, Oxford.

Black, K D (1998) The Environmental Interactions Associated with Fish Culture, in *Fish Farm Biology*, eds K Black and A Pickering, Sheffield Academic Press, Sheffield, 389 412.

Buschmann, A H, Lopez, D A, and Medina, A (1996) A Review of the Environmental Effects and Alternative Production Strategies of Marine Aquaculture in Chile, *Aquacult. Eng.*, 15, 397 421.

Chen, Y-S, Beveridge, M C M, and Telfer, T C (1999b) Physical, Characteristics of Commercial Pelleted Atlantic Salmon Feeds and Consideration of Implications for Modeling of Waste Dispersion Through Sedimentation, *Aquacult. Int.*, 7, 89 100.

FAO (1999a) *The State of World Fisheries and Aquaculture*, FAO, Rome.

Hansen, L P, Jacobsen, J A, and Lund, R A (1999) The Incidence of Escaped Farmed Atlantic Salmon *Salmo Salar* L., in the Faroese Fishery and Estimates of Catches of Wild Salmon, *ICES J. Mar. Sci.*, 56, 200 206.

Lebris, H, Maffart, P, Bocquene, G, Buchet, V, Galgani, F, and Blanc, G (1995) Laboratory Study of the Effect of Dichlorvos on two Commercial Bivalves, *Aquaculture*, 138, 139 144.

McGinnity, P, Stone, C, Taggart, J B, Cooke, D, Cotter, D, Hynes, R, McCamley, C, Cross, T, and Ferguson, A (1997) Genetic Impact of Escaped Farmed Atlantic Salmon (*Salmo Salar* L.) on Native Populations: Use of DNA Profiling to Assess Freshwater Performance of Wild, Farmed and Hybrid Progeny in a Natural River Environment, *ICES J. Mar. Sci.*, 54, 998 1008.

McHenery, J G, Francis, C, and Davies, I M (1996) Threshold Toxicity and repeated Exposure Studies of Dichlorvos to the Larvae of the Common Lobster (*Homarus gammarus* L.), *Aquat. Toxicol.*, 34, 237 251.

McKinnell, S, Thomson, A J, Black, E A, Wing, B L, Guthrie, C M, Koerner, J F, and Helle, J H (1997) Atlantic Salmon in the North Pacific, *Aquacult. Res.*, 28, 145 157.

Midlen, A and Redding, T A (1998) *Environmental Management for Aquaculture*, Chapman and Hall, London.

Naylor, R L, Goldburg, R J, Mooney, H, Beveridge, M C M, Clay, J, Folke, C, Kautsky, N, Lubchenko, J, Primavera, J, and Williams, M (1998) Nature s Subsidies to Shrimp and Salmon Farming, *Science*, 282, 883 884.

NRC (1996) *Upstream: Salmon and Society in the Paci c Northwest*, National Academy Press, Washington, DC.

Silvert, W (1994) Modelling Benthic Deposition and Impacts of Organic Matter Loading, in Modelling Benthic Impacts of Organic Enrichment from Marine Aquaculture, ed B T Hargrave, Canadian Technical Report on Fisheries and Aquatic Sciences, 1949, 1 18.

Weston, D P (1996) Environmental Considerations in the use of Antibacterial Drugs in Aquaculture, in *Aquaculture and Water Resource Management*, eds D J Baird, M C M Beveridge, L A Kelly, and J F Muir, Blackwell, Oxford, 140 165.

Williamson, R B and Beveridge, M C M (1994) Fisheries and Aquaculture, in *The Freshwater Resources of Scotland: a National Resource of International Signi cance*, eds P S Maitland, P J Boon, and D S McLusky, John Wiley & Sons, Chichester, 317 332.

FURTHER READING

Chen, Y-S, Beveridge, M C M, and Telfer, T C (1999a) Settling Rate Characteristics and Nutrient Content of the faces of Atlantic Salmon *Salmo Salar* L. Feeds and the Implications for Modeling of Solid Waste Dispersion, *Aquacult. Res.*, 33, 395 398.

FAO (1999b) *Aquaculture Production Statistics (1988- 1997)*, FAO Fisheries Circular 815 (Rev. 11), FAO, Rome.

Hargrave, B T (1994) Modeling Benthic Impacts of Organic Enrichment from Marine Aquaculture, Canadian Technical Report on Fisheries and Aquatic Sciences, 1949.

Aquaculture: the FAO Definition

see Aquaculture and Environment: Global View from the Tropics to High Latitudes (**Volume 3**)

Arctic Air Quality

see Arctic Air Quality (**Volume 1**)

Arctic Environment, Oil

see Oil and the Arctic Environment (**Volume 3**)

B

Baltic

see Water Resources: Baltic (**Volume 3**); Baltic Sea (*Case study*, **Volume 4**)

Biodiversity in Freshwaters

see Biodiversity in Freshwaters (**Volume 2**)

Biological Invasions

see Biological Invasions (*Opening essay*, **Volume 2**)

Biomass Burning

Joel S Levine
NASA Langley Research Center, Hampton, VA, USA

In the preface to the book, One Earth, One Future: Our Changing Global Environment, *published by the National Academy of Sciences (1990), Dr Frank Press, the President of the Academy writes:*

> *Human activities are transforming the global environment, and these global changes have many faces: ozone depletion, tropical deforestation, acid deposition, and increased atmospheric concentrations of gases that trap heat and may warm the global climate.*

It is interesting to note that all four global change faces identi ed by Dr Press have one common thread - they are all caused by or related to biomass burning.

Biomass burning also includes vegetation burning is the burning of living and dead vegetation for land-clearing and land-use change, or natural, lightning-induced burning. The bulk of the world's biomass burning occurs in the tropics - in the tropical forests of South America and Southeast Asia and in the savannas of Africa and South America. However, a signi cant amount of biomass burning occurs in the boreal forests of Russia, Canada and Alaska. The majority of the biomass burning, primarily in the tropics (perhaps as much as 90%) is believed to be human-initiated, with natural res triggered by atmospheric lightning only accounting for on the order of about 10% of all res (Andreae, 1991). Biomass burning has been identi ed as a signi cant source of gases and particulates to the regional and global atmosphere.

Biomass burning is an important process of global change for several reasons, including:

1. *Biomass burning is a signi cant source of gases and particulates released to the atmosphere.*
2. *Carbon dioxide (CO_2) and methane (CH_4) produced by biomass burning are greenhouse gases (GHG) that trap heat and may warm the global climate.*
3. *Methyl bromide (CH_3Br) produced by biomass burning leads to the photochemical destruction of ozone in the stratosphere.*
4. *Nitric oxide (NO) produced by biomass burning leads to the photochemical production of nitric acid (HNO_3), the fastest growing component of acid precipitation.*
5. *Atmospheric aerosols or particulates produced by biomass burning absorb and scatter incoming solar radiation, and, hence, impact climate.*
6. *Biomass burning impacts the biogeochemical cycling of nitrogen (N) and carbon (C) compounds from the biosphere to the atmosphere.*
7. *Biomass burning impacts the hydrological cycle by perturbing precipitation run-off and evaporation.*
8. *Biomass burning alters the re ectivity and emissivity of the land cover and, hence, impacts the radiation budget of the land/atmosphere system.*
9. *Biomass burning destroys ecosystems, primarily in the tropics, and, hence, leads to loss of biodiversity and extinction of both ora and fauna.*

10. Particulates produced during biomass burning can lead to a signi cant decrease in atmospheric visibility.
11. Particulates produced during biomass burning and released into the atmosphere can result in signi cant respiratory health problems.

BIOMASS BURNING AS A SOURCE OF GASES AND PARTICULATES TO THE ATMOSPHERE

Biomass burning is a significant global source of gases and particulates to the atmosphere (Crutzen *et al.*, 1979, Seiler and Crutzen, 1980, Crutzen and Andreae, 1990, Levine *et al.*, 1995). Laboratory biomass burning experiments conducted by Lobert *et al.* (1991) have identi ed the various carbon (C) (Table 1) and nitrogen (N) (Table 2) compounds released to the atmosphere by burning. The exact nature and quantity of the gaseous and particulate emissions resulting from biomass burning depend on the particular ecosystem undergoing burning. Lobert *et al.* (1991) used biomass material from several different ecosystems in their laboratory experiments. To quantify the gaseous and particulate emissions from a particular ecosystem, eld measurements are required. Some of the eld experiments to quantify biomass burning in diverse ecosystems are outlined in the following section. The major gases produced during the biomass burning process include carbon dioxide (CO_2), carbon monoxide (CO), methane (CH_4), oxides of nitrogen (NO_x = nitric oxide (NO) + nitrogen dioxide, NO_2), and ammonia (NH_3). Carbon dioxide and methane are GHG, which trap Earth-emitted infra-red radiation and lead to global warming. Carbon monoxide, methane, and the oxides

of nitrogen lead to the photochemical production of ozone (O_3) in the troposphere. In the troposphere, ozone is an irritant and harmful pollutant, and in some cases, is toxic to living systems. Nitric oxide leads to the chemical production of nitric acid (HNO_3) in the troposphere. Nitric acid is the fastest growing component of acidic precipitation. Ammonia is the only basic gaseous species that neutralizes the acidic nature of the troposphere. Particulates, small (usually about 10 μm or smaller) solid particles, such as smoke or soot particles, are also produced during the burning process and released into the atmosphere. These solid particulates absorb and scatter incoming sunlight and hence impact the local, regional, and global climate. In addition, these particulates (specifically particulates 2.5 μm or smaller) can lead to various human respiratory and general health problems when inhaled. The gases and particulates produced during biomass burning lead to the formation of smog. The word smog was coined as a combination of smoke and fog and is now used to describe any smoky or hazy pollution in the atmosphere.

The study of biomass burning is truly a multidisciplined subject, which includes the following areas: fire ecology, fire measurements and modeling, fire combustion, remote sensing of fires, gaseous and particulate emissions from fires, the atmospheric transport of these emissions and the chemical and climatic impacts of these emissions. Over the last two decades, biomass burning has been the subject of many hundreds of papers in scientific journals and nine dedicated volumes (Goldammer, 1990; Levine, 1991; Crutzen and Goldammer, 1993; Goldammer and Furyaev, 1996; Levine, 1996a,b; van Wilgen *et al.*, 1997; Innes *et al.*, 2000; and Kasischke and Stocks, 2000).

BIOMASS BURNING FIELD EXPERIMENTS

Over the last decade, a series of biomass burning field experiments have been conducted to measure and quantify the gaseous and particulate emissions produced by biomass burning in diverse ecosystems. These field experiments also quantified biomass loading, before and after burning, the fire efficiency, the combustion efficiency, the fire temperature and energy released, the fire spread rate, the gaseous and particulate emissions produced during the fire and released to the atmosphere, and the impact of burning on the local ecosystem. These measurable parameters are all ecosystem-dependent. Biomass burning field experiments have been conducted in various ecosystems in Africa, South America, the Pacific Northwest US and Russia. Most of these experiments were planned and implemented under the auspices of the Biomass Burning Experiment (BIBEX), a research activity of the International Global Atmospheric Chemistry (IGAC) Project, part of the International Geosphere-Biosphere Program (IGBP). Some of these experiments are:

1. Africa: South African Fire-Atmosphere Research Initiative (SAFARI) 1992 and 2000, Transport and Atmospheric Chemistry near the Equator-Atlantic (TRACE-A), Experiment for Regional Sources and Sinks of Oxidants (EXPRESSO);
2. South America: Transport and Atmospheric Chemistry near the Equator-Atlantic (TRACE-A);
3. Pacific Northwest US: Smoke, Clouds, and Radiation-C (SCAR-C); and
4. The Bor Forest Island Experiment in Krasnoynarsk, Russia.

ENHANCED BIOGENIC SOIL EMISSIONS OF NITROGEN AND CARBON GASES: A POST-FIRE EFFECT

Measurements have shown that in addition to the instantaneous production of trace gases and particulates resulting from the combustion of biomass matter, burning also enhances the biogenic emissions of NO, nitrous oxide (N_2O) and CO from soil. It is believed that enhanced biogenic soil emissions of NO and N_2O are related to increased concentrations of ammonium (NH_4^+) found in soil following burning. Ammonium, a component of the burn ash is the substrate in nitrification, which is the microbial process believed responsible for the production of nitric and nitrous oxide. The post-fire enhanced biogenic soil emissions of NO and N_2O may be comparable to or even surpass the instantaneous production of these gases during biomass burning.

BIOMASS BURNING AND ATMOSPHERIC NITROGEN AND OXYGEN

Biomass burning is both an instantaneous source (the real time combustion of biomass) and a long-term source (enhanced biogenic soil emissions) of gases to the atmosphere. These gases impact both the chemistry of the troposphere and stratosphere and the biogeochemical cycling of nitrogen (N_2) and oxygen (O_2), the two major constituents of the atmosphere. Through the process of nitrogen fixation, molecular nitrogen is transported to the surface in the form of fixed nitrogen, i.e., ammonium (NH_4^+) and nitrate (NO_3^-). Nitrogen fixation results from both natural processes (biological fixation in root modules in certain agricultural crops and atmospheric lightning) and human processes (the production of nitrogen fertilizer and high temperature combustion). The world s use of industrially fixed nitrogen fertilizer has increased from about 3 teragrams (Tg) of N $year^{-1}$ in 1940 to about 75 Tg N $year^{-1}$ in 1990. The fixed nitrogen in the forms of NH_4^+ and NO_3^- is returned to the atmosphere mainly in the form of N_2, with smaller amounts of N_2O, and still smaller amounts of NO by denitrification and in the form of NO by nitrification.

Burning or pyro-denitrification may also be an important source of nitrogen, mostly in the form of N_2, from the biosphere to the atmosphere. The problem is that it is difficult to quantify the amount of N_2 released during burning (Lobert *et al.*, 1991), however, biomass burning or pyro-denitrification may prove to be an important process in the recycling of nitrogen compounds from the biosphere to the atmosphere.

Burning impacts the concentration of atmospheric oxygen in two ways. Carbon released during the burning of biomass combines with atmospheric oxygen to form CO_2. Hence, burning is a sink for atmospheric oxygen. In addition, biomass burning destroys the very source of atmospheric oxygen its production in the biosphere via the process of photosynthesis in the world s forests.

GEOGRAPHICAL DISTRIBUTION OF BIOMASS BURNING

Areas of biomass burning are varied and include tropical savannas, tropical, temperate and boreal forests and agricultural lands after the harvest. The burning of fuelwood for domestic use is another source of biomass burning. Much of our information on the geographical distribution of biomass burning is based on burning in the tropics, particularly South America, Africa and Southeast Asia. Considerably less information is known about biomass burning in the boreal forest.

BURNING IN THE WORLD'S BOREAL FORESTS

The world s boreal forests cover a little less than 17% of the Earth s land area, yet contain more than 30% of all the carbon present in the terrestrial biome. Although the combined tropical and temperate forests cover almost twice the land area as the boreal forest, the boreal forest contains 20% more carbon than the other two forests combined (Kasischke and Stocks, 2000). There are several important reasons that burning in the world s boreal forests is very important:

1. The world s boreal forests are very susceptible to global warming. Small changes in the surface temperature can significantly influence the ice/snow/albedo feedback. Thus, infrared absorption processes by fire-produced GHGs, as well as fire-induced changes in surface albedo and infrared emissivity are more environmentally significant than in the tropics.
2. In the world s boreal forests, global warming may result in warmer and drier conditions. This in turn may result in enhanced frequency of fire and the accompanying enhanced production of GHGs that will amplify the greenhouse effect.
3. Fires in the boreal forests are perhaps the most energetic in nature. The average fuel consumption per unit area in the boreal forest is on the order of 25 000 kg hectare^{-1} (1 ha = 2.47 acres), which is about an order of magnitude greater than in the tropics. Large boreal forest fires typically spread very quickly, most often as crown fires. Large boreal forest fires release enough energy to generate convective smoke columns that routinely reach well into the troposphere, and on occasion may directly penetrate across the tropopause. The tropopause is at a minimum height over the world s boreal forests. As an example, a 1986 forest fire in Northwestern Ontario (Red Lake) at times generated a convective smoke column 12 13 km, penetrating the tropopause.
4. The cold temperature of the troposphere over the world s boreal forests results in very low levels of tropospheric water vapor. The deficiency of tropospheric water vapor and the scarcity of incoming solar radiation over most of the year results in very low photochemical production of the hydroxyl radical (OH) over the boreal forests. The OH radical is the overwhelming chemical scavenger in the troposphere and controls the atmospheric lifetime of most tropospheric gases. The very low concentrations of the OH radical over the boreal forests will result in enhanced atmospheric lifetimes for most tropospheric gases, including the gases produced by fire combustion. Hence, gases produced by burning, such as CO, CH_4, and the oxides of nitrogen will have enhanced lifetimes over the boreal forest (*see* OH–Radical: is the Cleansing Capacity of the Atmosphere Changing?, Volume 2).

It has generally been believed that biomass burning is primarily a tropical phenomenon. This is because most of the information that we have on the geographical and temporal distribution of biomass burning is largely based on tropical burning. Due to poor satellite coverage, among other things, there is very little information available on the geographical and temporal distribution on biomass burning in the boreal forests, which cover about 25% of the world s forests. To illustrate how our knowledge of the geographical extent of burning in the world s boreal forests has increased in recent years, consider the following: Early estimates based on surface fire records and statistics suggested that as much as 1.5 million ha of boreal forests burn annually (Seiler and Crutzen, 1980). One of the largest fires ever measured occurred in the boreal forests of the Heilongjiang Province of Northeastern China in May 1987. In less than four weeks, more than 1.3 million ha of boreal forest were burned (Cahoon *et al.*, 1994). At the same time, extensive fire activity occurred across the border in Russia, particularly in the area east of Lake Baikal between the Amur and Lena rivers. Estimates based on imagery from the National Oceanic and Atmospheric Administration/Advanced Very

High Resolution Radiometer (NOAA/AVHRR) indicate that 14.446 million ha (35.697 million acres) in China and Siberia were burned in 1987 (Cahoon *et al.*, 1994).

While 1987 was an extreme fire year in Eastern Asia, the sparse database may suggest a trend. Is burning in the boreal forests increasing with time, or are satellite measurements providing more accurate data? Satellite measurements are certainly providing a more accurate assessment of the extent and frequency of burning in the world s boreal forests. However, it is important to point out that if global warming becomes a reality, predicted warmer and drier conditions in the world s boreal forests will result in more frequent and larger fires, and increased burning may have an amplifying effect on global warming!

A recent study (Kasischke *et al.*, 1999) has provided new information of burning in the world s boreal forests:

1. Fires in the boreal forest covering at least 100 000 ha are not uncommon.
2. In the boreal forests of North America, most fires (>90%) are crown fires. The remainder are surface fires. Crown fires consume much more fuel (30 40 tons of biomass material per ha burned) than surface fires (8 12 tons of biomass material per ha burned).
3. The fire record for North America over the past three decades clearly shows the episodic nature of fire in the boreal forests. Large fire years occur during extended periods of drought, which allow naturally ignited fires (i.e., lightning ignited fires) to burn large areas. Since 1970, the area burned during six episodic fire years in the North American boreal forest was 6.2 million ha year^{-1}, while 1.5 million ha year^{-1} burned in the remaining years. There is evidence that a similarly episodic pattern of fire may also exist in the Russian boreal forest.
4. The fire data in the North American boreal forest show a significant increase in the annual area burned over the past three decades, with an average 1.5 million ha year^{-1} burning during the 1970s and 3.2 million ha year^{-1} burning during the 1990s. This increase in burning corresponds to rises of 1.0 1.6 °C over the same period. The projected 2 4 °C increase in temperature due to projected increases in GHGs should result in high levels of fire activity throughout the world s boreal forests in the future.
5. During typical years in the boreal forests, the amounts of biomass consumed during fire ranges between 10 and 20 tons ha^{-1}. During the drought years with episodic fires, the amounts of biomass consumed during biomass burning may be as high as 50 60 tons ha^{-1}. Assuming that biomass is about 50% carbon by mass, such amounts would release 450 600 Tg C globally. These amounts are considerably higher than the often-quoted value for total carbon released by biomass burning in the world s boreal and temperate forests of 130 Tg C globally (Andreae, 1991, see Table 3).

ESTIMATES OF GLOBAL BURNING AND GLOBAL GASEOUS AND PARTICULATE EMISSIONS

Global estimates of the annual amounts of biomass burning from these sources are estimated in Table 3 (Andreae, 1991). Biomass matter is about 45% by weight composed of carbon. Table 3 also gives estimates of the carbon released (Tg C year^{-1}) by the burning of this biomass. Combining estimates of the total amount of biomass matter burned per year (Table 3) with measurements of the gaseous and particulate emissions from biomass burning (Tables 1 and 2) permits estimates of the global production and release into the atmosphere of gases and particulates from burning. Estimates of the global contribution of biomass burning are summarized in Table 4 (Andreae, 1991). The data in Tables 3 and 4 clearly indicate that biomass burning is a truly global process of major importance in the global budgets of atmospheric gases and particulates.

A CASE STUDY: THE CALCULATION OF GASEOUS AND PARTICULATE EMISSIONS FROM THE FOREST AND PEAT FIRES IN KALIMANTAN AND SUMATRA, INDONESIA, IN 1997

The years 1997 1998 saw a series of extensive and widespread El-Nino-related fires in Southeast Asia, South America, Africa, Mexico, Russia, and Florida. The fires in Southeast Asia were particularly extensive and widespread. Throughout Southeast Asia, including Indonesia, the major cause of forest fires is controlled burning for land clearing and land-use change activities. This burning has been an

annual practice for over 100 years. Over the last 15 years, burning for land clearing has increased, as large rubber and oil palm plantations use burning as a fast and inexpensive method of large-scale land clearing. However, forest fires in Indonesia are unique in that they often ignite fires in the underlying peat or coal layers. About 50%, or approximately 269 730 km^2 of Kalimantan and about 35%, or 183 435 km^2 of Sumatra, are underlain with peat and coal. Significant, but smaller deposits of peat and coal are also found on many of the other Indonesian islands, including Java, Southern Sulawesi, Madura and Irian Jaya.

During the last 15 years of burning in Indonesia, it became apparent that large-scale regional episodes of smoke, locally known as haze, were produced. Fires for land clearing produced regional haze conditions in 1982, 1991, and 1994. In 1997, the El Nino/Southern Oscillation once again brought severe drought conditions to Southeast Asia and contributed to uncontrolled fires which produced severe regional haze and air pollution conditions. The 1997 1998 forest and peat/coal fires in Indonesia were described as one of the most broad-ranging environmental disasters of the century.

Some of the consequences of the fires in Indonesia fires included: (1) more than 200 million people exposed to high levels of air pollution and particulates produced during the fires; (2) more than 20 million smoke-related health problems; (3) fire-related damage in excess of $4 billion; (4) the crash of a commercial airliner (Garuda Airlines Airbus 300-B4 on September 26th 1997) in Sumatra due to very poor visibility on landing with 234 passengers killed; (5) the collision of two ships at sea due to poor visibility in the Strait of Malacca, off the coast of Malaysia, on September 27th, 1997 with 29 crew members killed.

International concern about the environmental and health impacts of these fires resulted in a series of United Nations (UN) workshops and reports, including: The World Meteorological Organization (WMO) Workshop on Regional Transboundary Smoke and Haze in Southeast Asia, Singapore, June 2nd 5th, 1998 (Carmichael, 1999), The World Health Organization (WHO) Health Guidelines for Forest Fires Episodic Events, Lima, Peru, October 6th 9th, 1998, (Schwela *et al.*, 1999) and the United Nations Environmental Programme (UNEP) Report on Wildland Fires and the Environment: A Global Synthesis (Levine, 1999).

Calculation of Gaseous and Particulate Emissions from Vegetation and Peat Fires

To assess both the environmental and health impacts of these fires, the gaseous and particulate emissions produced during the fire and released into the atmosphere must be modeled. The calculation of gaseous and particulate emissions from biomass burning or vegetation fires can be

calculated using a form of an expression from Seiler and Crutzen (1980) for each burning ecosystem/terrain type i (where i is the rain forests, dry forests, plantations, chaparral, grasslands, agricultural lands, wetlands, peatlands, etc.):

$$M_i = A_i \times B_i \times E_i \quad (1)$$

where M_i is the total mass of biomass or vegetation consumed by burning (tons) in ecosystem i, A_i is the area burned in ecosystem i (km^2), B_i is the above ground biomass loading in ecosystem i (tons km^{-2}), and E_i is the burning ef ciency of vegetation in ecosystem i (dimensionless). For each ecosystem i, the total mass of carbon $(M(C))_i$ released to the atmosphere during burning is related to M_i by the following expression:

$$M(C)_i = C \times M_i \text{ (tons of carbon)} \quad (2)$$

C is the mass percentage of carbon in the biomass. For most vegetation, $C = 0.45$ (Andreae, 1991). The mass of CO_2 ($M(CO_2)_i$) released during the re is related to $M(C)_i$ by the following expression:

$$M(CO_2)_i = CE_i \times M(C)_i \quad (3)$$

The combustion ef ciency (CE_i) is the fraction of carbon emitted as CO_2 relative to the total carbon compounds released during the re in ecosystem i. For most vegetation res, $CE = 0.90$ (Andreae, 1991).

Once the mass of CO_2 produced by burning is known, the mass of any other species, j, in ecosystem i, X_{ij} ($M(X_{ij})$), produced by burning and released to the atmosphere can be calculated with knowledge of the CO_2-normalized *species emission ratio* ($ER(X_{ij})$). The emission ratio is the ratio of the production of species X_{ij} to the production of CO_2 in the re in ecosystem i. The mass of species, X_{ij} is related to the mass of CO_2 by the following expression:

$$M(X_{ij}) = ER(X_{ij}) \times M(CO_2)_i \text{ (tons of element } X_i) \quad (4)$$

where $X_j = CO$, CH_4, non-methane hydrocarbons, NO_x, NH_3, etc. Typical values for emission ratios for vegetation and peat res are given in Table 5.

An alternate approach involves the use of the *species emission factor* ($EF(X_{ij})$), instead of the species emission ratio ($ER(X_{ij})$). The emission factor gives the mass of gas or particulate produced per mass of biomass consumed by burning in units of g kg^{-1}. To calculate the total particulate matter (TPM) released from vegetation res, we use the following expression:

$$TPM = M \times P \text{ (tons of carbon)} \quad (5)$$

where P is the conversion of biomass matter to particulate matter during burning. For the burning of most vegetation, $P = 20$ tons of TPM $kton^{-1}$ of biomass consumed by re.

Table 5 Some typical emission ratios for tropical forest fires and peat fires

Species	Tropical forest fires (%)	Peat fires (%)
CO_2	90.00	77.05
CO	8.5	18.15
CH_4	0.32	1.04
NO_x	0.21	0.46
NH_3	0.09	1.28
O_3	0.48	1.04
TPM[a]	20 tons tons $kton^{-1}$	35 tons $kton^{-1}$

[a] TPM, TPM emission ratios are in units of tons $kton^{-1}$ (tons of TPM/kiloton of biomass or peat material consumed by fire).

Total Area Burned and Biomass Consumed

Perhaps the major uncertainties in the calculation of gaseous and particulate emissions resulting from res involve poor or incomplete information about four re parameters: (1) the area burned (A); (2) the ecosystem or terrain that burned, i.e., forests, grasslands, agricultural lands, peat lands, etc.; (3) the biomass loading (B), i.e., the amount of biomass per unit area of the ecosystem prior to burning; and (4) the re ef ciency (C), i.e., the amount of biomass in the burned ecosystem that was actually consumed by burning. The area burned can be determined through the use of satellite measurements. Some operational satellite systems to estimate area burned, as well as monitor active res, are listed in Table 6.

Liew *et al.* (1998) analyzed 766 SPOT (Système Pour l Observation de la Terre) quicklook images with almost complete coverage of Kalimantan and Sumatra from August December, 1997. Liew *et al.* (1998) estimated the burned area in Kalimantan to be 30 600 km^2 and the burned area in Sumatra to be 15 000 km^2, for a total burned area of 45 600 km^2. For the calculations illustrated in this article, we have used the Liew *et al.* (1998) estimate for total burned area in Kalimantan and Sumatra of 45 600 km^2 (This is greater than the total area of Denmark or Switzerland). The estimate of Liew *et al.* (1998) represents only a lower limit estimate of the area burned in Southeast Asia in 1997, since the SPOT data only covered Kalimantan and Sumatra and did not include res on the other Indonesian islands of Irian Jaya, Sulawesi, Java, Sumbawa, Komodo, Flores, Sumba, Timor, and Wetar or the res in the neighboring countries of Malaysia and Brunei.

What is the nature of the ecosystem/terrain that burned in Kalimantan and Sumatra? In October 1997, NOAA satellite monitoring produced the following distribution of re hot spots in Indonesia (UNDAC, 1998): agricultural and plantation areas: 45.95%; bush and peat soil areas: 24.27%; productive forests: 15.49%; timber estate areas: 8.51%; protected areas: 4.58%; and transmigration sites: 1.20% (the three forest/timber areas add up to a total of 28.58% of the area burned). While the distribution of re hot spots is not

Table 6 Some current operational satellite fire monitoring systems

1. NOAA/AVHRR Global 1 km imaging systems. Monitors active fires and burned area
2. DMSP (Defense Meteorological Satellite Program)/OLS (Operational Linescan System): Global night-time low light sensor. Monitors active fires
3. GOES (Geostationary Operational Environmental Satellite) Imager: Continental high temporal frequency, coarse spatial resolution geostationary imaging. Monitors active fires and smoke
4. ERS (European Remote Sensing Satellite)/ATSR (Along-Track Scanning Radiometer) (European Space Agency): Global 1 km imaging. Monitors active fires and burned areas
5. ERS/JERS (Japanese Earth Resources Satellite) SAR (Synthetic Aperture Radar) (European Space Agency/NASDA (National Space Development Agency of Japan): Global microwave high-resolution system. Monitors burned area
6. LANDSAT (Land Satellite) TM (Thematic Mapper)/MSS (Multispectral Scanner System): Local, high spatial frequency, low temporal frequency. Monitors burned area
7. SPOT (Système Pour l'Observation de la Terre) CNES (Centre National d'Etudes Spatiales): Local, high-spatial frequency, low-temporal frequency. Monitors burned area.
8. Terra Satellite (EOS-AM1)/Moderate Resolution Imaging Spectrometer (MODIS): Launched December 1999: Measures active fires and burned areas

an actual index for area burned, the NOAA satellite-derived hot spot distribution is quite similar to the ecosystem/terrain distribution of burned area deduced by Liew *et al.* (1998) based on SPOT images of the actual burned areas: agricultural and plantation areas: 50%; forests: 30%; and peat swamp forests: 20%. Since the estimates of burned ecosystem/terrain of Liew *et al.* (1998) are based on actual SPOT images of the burned area, their estimates were adopted in our calculations.

What is the biomass loading for the three terrain classifications identified by Liew *et al.* (1998)? Values for biomass loading or fuel load for various tropical ecosystems are summarized in Table 7. The biomass loading for tropical forests in Southeast Asia ranges from 5000 55 000 tons km^{-2}, with a mean value of 23 000 tons km^{-2}. However, in our calculations we have used a value of 10 000 tons km^{-2} to be conservative. The biomass loading for agricultural and plantation areas (mainly rubber trees and oil palms) of 5000 tons km^{-2} is also a conservative value (Liew *et al.*, 1998). A biomass loading value of 97 500 tons km^{-2} for the dry peat deposits 1.5 m thick was used (Supardi *et al.*, 1993). The combustion efficiency for forests is estimated at 0.20 and for peat is estimated at 0.50 (Levine and Cofer, 2000). Based on the discussions presented in this section, the values for burned area, biomass loading, and combustion efficiency, used in the calculations are summarized in Table 8.

Table 7 Biomass load range and burning efficiency in tropical ecosystems (Scholes *et al.*, 1996 and Supardi *et al.*, 1993)

Vegetation type	Biomass load range (tons km^{-2})	Burning efficiency
Peat	97 500	0.50
Tropical rainforests	5000–55 000	0.20
Evergreen forests	5000–10 000	0.30
Plantations	500–10 000	0.40
Dry forests	3000–7000	0.40
Fynbos	2000–4500	0.50
Wetlands	340–1000	0.70
Fertile grasslands	150–550	0.96
Forest/savanna mosaic	150–500	0.45
Infertile savannas	150–500	0.95
Fertile savannas	150–500	0.95
Infertile grasslands	150–350	0.96
Shrublands	50–200	0.95

Table 8 Parameters used in calculations of emissions from Southeast Asia fires[a]

1. Total area burned in Kalimantan and Sumatra, Indonesia in 1997: 45 600 km^2
2. Distribution of burned areas, biomass loading and combustion efficiency

A. Agricultural and plantation areas – 50%, 5000 tons km^{-2}, 0.20
B. Forests and bushes – 30%, 10 000 tons km^{-2}, 0.20
C. Peat swamp forests – 20%, 97 500 tons km^{-2}, 0.50

[a] See text for discussion of remission estimate range and uncertainty calculations.

Results of Calculations: Gaseous and Particulate Emissions

The calculated gaseous and particulate emissions are summarized in Table 9. For each of the seven species listed, the emissions due to agricultural/plantation burning (A), forest burning (F), and peat burning (P) are given. The total (T) of all three components (A + F + P) is also given. The best estimates of total emissions are: CO_2: 191.485 million metric tons of C (Mt C); CO: 32.794 Mt C; CH_4: 1.845 Mt C; NO_x: 0.971 Mt N; NH_3: 2.585 Mt N; O_3: 7.100 Mt O_3; and TPM: 16.568 Mt C.

Scholes *et al.* (1996) calculated the biomass consumed by burning in 11 different ecosystems in another tropical ecosystem, southern Africa. Scholes *et al.* (1996) performed a detailed statistical analysis of the errors associated with the calculated values of biomass consumed by fire using a statistical procedure which assumes that all error terms (e) are independent. In the error analysis of Scholes *et al.* (1996), the total error (e_{total}), which corresponds to

Table 9 Gaseous and particulate emissions from the fires in Kalimantan and Sumatra in 1997 (for total burned area = 45 600 km^2) (Levine, 1999)[a]

	Agricultural/plantation fire emissions	Forest fire emissions	Peat fire emissions	Total fire emissions
CO_2	9.234 (4.617–13.851)	11.080 (5.54–16.62)	171.170 (85.585–256.755)	191.485 (95.742–287.226)
CO	0.785 (0.392–1.177)	0.942 (0.471–1.413)	31.067 (15.533–46.600)	32.794 (16.397–49.191)
CH_4	0.030 (0.015–0.045)	0.035 (0.017–0.052)	1.780 (0.89–2.67)	1.845 (0.922–2.767)
NO_x	0.023 (0.011–0.034)	0.027 (0.013–0.040)	0.921 (0.460–1.381)	0.971 (0.485–1.456)
NH_3	0.010 (0.005–0.015)	0.012 (0.006–0.018)	2.563 (1.281–3.844)	2.585 (1.292–3.877)
O_3	0.177 (0.088–0.265)	0.213 (0.106–0.319)	6.710 (3.35–10.06)	7.100 (3.55–10.65)
TPM	0.460 (0.23–0.69)	0.547 (0.273–0.820)	15.561 (7.780–23.341)	16.568 (8.284–24.852)

[a] Units of emissions: Mt of C for CO_2, CO, and CH_4; Mt of N for NO_x and NH_3; Mt of O_3 for O_3; Mt C of particulates; 1 Mt = 10^{12} g = 1 Teragram, Tg.

the 3-sigma (99%) confidence level, is estimated using the following expression:

$$e_{\text{total}} = (e^2_{\text{burned area}} + e^2_{\text{fuel load}} + e^2_{\text{combustion completeness}})^{1/2} \quad (6)$$

Scholes *et al.* (1996) assumed the following uncertainties for each calculation parameter: $e_{\text{burned area}} = 30\%$, $e_{\text{fuel load}} = 30\%$ and $e_{\text{combustion completeness}} = 25\%$. In the error analysis of the calculations presented in this article, the error associated with uncertainties in the emission ratio ($e_{\text{emission ratio}}$) (30%) has also been included. The uncertainties in the calculation parameters in the detailed error of analysis of Scholes *et al.* (1996) were adopted in this study, with the exception of the uncertainty in the burned area of 30%. The burned area determination used in Scholes *et al.* (1996) was based on satellite measurements of active fires that were converted to burned area, which is not a simple one-to-one transformation and introduces errors. The burned area determination used in this article was more straightforward since it was based on direct satellite photography of burned areas using SPOT images (Liew *et al.*, 1998). Hence, there is less uncertainty in this burned area determination and an uncertainty of 10% was assumed. (Liew *et al.* (1998) did not give a burned area uncertainty in their paper.) The calculated uncertainty in the emission calculations is 50.2%. The uncertainty range for each species emission is shown in parentheses under the best estimate value in Table 9.

However, it is important to re-emphasize that these emission calculations represent lower limit values since the calculations are only based on burning in Kalimantan and Sumatra in 1997. The calculations do not include burning in Java, Sulawesi, Irian Jaya, Sumbawa, Komodo, Flores, Sumba, Timor, and Wetar in Indonesia or in neighboring Malaysia and Brunei.

It is interesting to compare the gaseous and particulate emissions from the 1997 Kalimantan and Sumatra fires with those from the Kuwait oil fires of 1991, described as a major environmental catastrophe. Laursen *et al.* (1992) calculated the emissions of CO_2, CO, CH_4, NO_x, and particulates from the Kuwait oil fires in units of metric tons per day. The Laursen *et al.* (1992) calculations are summarized in Table 10. To compare these calculations with the calculations presented in this article for Kalimantan and Sumatra (Table 9), we have normalized our calculations by the total number of days of burning. The SPOT images (Liew *et al.*, 1998) covered a period of five months (August December, 1997) or about 150 days. For comparison with the Kuwait fire emissions, we divided our calculated emissions by 150 days. These values are summarized in Table 10. The gaseous and particulate emissions from the fires in Kalimantan and Sumatra significantly exceeded the emissions from the Kuwait oil fires. The 1997 fires in Kalimantan and Sumatra were a significant source of gaseous and particulate emissions to the local, regional, and global atmosphere.

Table 10 Comparison of gaseous and particulate emissions: the Indonesian fires (Levine, 1999) and the Kuwait oil fires (Laursen *et al.*, 1992)[a]

Species	Indonesian fires	Kuwait oil fires
CO_2	1.28×10^6	5.0×10^5
CO	2.19×10^5	4.4×10^3
CH_4	1.23×10^4	1.5×10^3
NO_x	6.19×10^3	2.0×10^2
Particulates	1.08×10^5	1.2×10^4

[a] Units of emissions: metric tons per day of carbon for CO_2, CO, and CH_4; metric tons per day of nitrogen for NO_x; metric tons per day for particulates; one million metric tons = 10^{12} (g = 1 Teragram, Tg).

See also: Remote Sensing, Volume 2; Remote Sensing, Terrestrial Systems, Volume 2; Oil Fires: Kuwait, Volume 3.

REFERENCES

Andreae, M O (1991) Biomass Burning: its History, use, and Distribution and its Impact on Environmental Quality and Global Climate, in *Global Biomass Burning: Atmospheric,*

Climatic, and Biospheric Implications, ed J S Levine, The MIT Press, Cambridge, MA, 3 21.

Cahoon, D R, Stocks, B J, Levine, J S, Cofer, W R, and Pierson, J M (1994) Satellite Analysis of the Severe 1987 Forest Fires in Northern China and Southeastern Siberia, *J. Geophys. Res.*, 99, 18 627 18 638.

Carmichael, G R (Compiler) (1999) *World Meteorological Organization Workshop on Regional Transboundary Smoke and Haze in Southeast Asia*, Vols. 1 and 2, World Meteorological Organization Report 131, Geneva, Switzerland.

Crutzen, P J and Andreae, M O (1990) Biomass Burning in the Tropics: Impact on Atmospheric Chemistry and Biogeochemical Cycles, *Science*, 250, 1679 1678.

Crutzen, P J and Goldammer, J G, eds (1993) *Fire in the Environment: the Ecological, Atmospheric, and Climatic Importance of Vegetation Fires*, John Wiley & Sons, Chicester, 1 400.

Crutzen, P J, Heidt, L E, Krasnec, J P, Pollock, W H, and Seiler, W (1979) Biomass Burning as a Source of Atmospheric Gases, CO, H_2, N_2O, NO, CH_3Cl, COS, *Nature*, 282, 253 256.

Goldammer, J G, ed (1990) *Fire in the Tropical Biota: Ecosystem Processes and Global Challenges*, Springer-Verlag, Berlin, 1 497.

Goldammer, J G and Furyaev, V V, eds (1996) *Fire in Ecosystems of Boreal Eurasia*, Kluwer Academic Publishers, Dordrecht, 1 528.

Innes, J L, Beniston, M, and Verstraet, M M, eds (2000) *Biomass Burning and its Inter-Relationships with the Climate System*, Kluwer Academic Publishers, Dordrecht, 1 358.

Kasischke, E S, Bergen, K, Fennimore, R, Sotelo, F, Stephens, G, Janetos, A, and Shugart, H H (1999) Satellite Imagery gives Clear Picture of Russia s Boreal Forest Fires, *EOS, Trans. Am. Geophys. Union*, 80, 141 147.

Kasischke, E S and Stocks, B J, eds (2000) Fire, Climate Change, and Carbon Cycling in the Boreal Forest, in *Ecological Studies*, Vol. 138, Springer-Verlag, NY, 1 461.

Laursen, K K, Ferek, R J, and Hobbs, P V (1992) Emission Factors for Particulates, Elemental Carbon and Trace Gases from the Kuwait Oil Fires, *J. Geophys. Res.*, 97, 14 491 14 497.

Levine, J S (1999) The 1997 Fires in Kalimantan and Sumatra, Indonesia: Gaseous and Particulate Emissions, *Geophys. Res. Lett.*, 26, 815 818.

Levine, J S, ed (1991) *Global Biomass Burning: Atmospheric, Climatic, and Biospheric Implications*, The MIT Press, Cambridge, MA, 1 569.

Levine, J S, ed (1996a) *Biomass Burning and Global Change: Remote Sensing, Modeling and Inventory Development, and Biomass Burning in Africa*, The MIT Press, Cambridge, MA, 1 581.

Levine, J S, ed (1996b) *Biomass Burning and Global Change: Biomass Burning in South America, Southeast Asia, and Temperate and Boreal Ecosystems, and the Oil Fires of Kuwait*, The MIT Press, Cambridge, MA, 1 377.

Levine, J S, Bobbe, T, Ray, N, Witt, R G, and Singh, A (1999) *Wildland Fires and the Environment: a Global Synthesis. Environment Information and Assessment Technical Report 99- 1*, The United Nations Environmental Program, Nairobi, Kenya, 1 46.

Levine, J S and Cofer, W R (2000) *Boreal Forest Fire Emissions and the Chemistry of the Atmosphere. Fire, Climate Change and Carbon Cycling in the North American Boreal Forests*, eds E S Kasischke and B J Stocks, Ecological Studies Series, Springer-Verlag, NY, 31 48.

Levine, J S, Cofer, W R, Cahoon, D R, and Winstead, E L (1995) Biomass Burning: a Driver for Global Change, in *Environ. Sci. Technol.*, 29, 120A 125A.

Liew, S C, Lim, O K, Kwoh, L K, and Lim, H (1998) *A Study of the 1997 Fires in South East Asia using SPOT Quicklook Mosaics*, Paper Presented at the 1998 International Geoscience and Remote Sensing Symposium, July 6 10, Seattle, Washington, DC, 1 3.

Lobert, J M, Scharffe, D H, Hao, W-M, Kuhlbusch, T A, Seuwen, R, Warneck, P, and Crutzen, P J (1991) Experimental Evaluation of Biomass Burning Emissions: Nitrogen and Carbon Containing Compounds, in *Global Biomass Burning: Atmospheric, Climatic, and Biospheric Implications*, ed J S Levine, The MIT Press, Cambridge, MA, 289 304.

Scholes, R G, Kendall, J, and Justice, C O (1996) The Quantity of Biomass Burned in Southern Africa, *J. Geophys. Res.*, 101, 23 667 23 676.

Schwela, D H, Goldammer, J G, Morawska, L H, and Simpson, O, eds (1999) *Health Guidelines for Episodic Vegetation Fires Events*, World Health Organization, Geneva, Switzerland.

Seiler, W and Crutzen, P J (1980) Estimates of Gross and Net Fluxes of Carbon Between the Biosphere and the Atmosphere from Biomass Burning, *Clim. Change*, 2, 207 247.

Supardi, A, Subekty, D, and Neuzil, S G (1993) General Geology and Peat Resources of the Siak Kanan and Bengkalis Island Peat Deposits, Sumatra, Indonesia. Modern and Ancient Coal Forming Environments, eds J C Cobb and C Cecil, *Geol. Soc. Am. (Special Paper)*, 86, 45 61.

UNDAC (1998) *Mission on Forest Fires*, Indonesia, September November, 1997 International Forest Fire News, United Nations Economic Commission for Europe and the Food and Agriculture Organization of the United Nations, Geneva, Switzerland, No. 18, 13 26.

van Wilgen, B W, Andreae, M O, Goldammer, J G, and Lindesay, J A, eds (1997) *Fire in Southern African Savannas: Ecological and Atmospheric Perspectives*, Witwatersrand University Press, Johannesburg, South Africa, 1 256.

Biomass Burning in Rural Homes in Tropical Areas

Phil O'Keefe

University of Northumbria at Newcastle, Newcastle, UK

Countries with low income, countries with unequal income distributions and countries with low levels of urbanization rely on signi cantly different energy carriers than richer countries. For the cooking, water heating and space

heating needs of some two billion people worldwide the dominant household energy consumption activities wood, crop residues and dung, collectively called biomass, are the dominant fuels. Per capita energy use is approximately one cubic meter per person per year.

Biomass accounts for about 80% of all household energy consumption in developing countries although consumption in urban areas is lower. Women and children are the primary collectors of fuelwood and other biomass energy sources. Air pollution indoors is a major by-product of the traditional use of biomass, which diminishes the quality of life, especially for women and young children.

Despite attempts to directly link biomass eld use with deforestation, the relationship is weak. The primary cause for deforestation is land clearance for agricultural purposes rather than cutting forests for biomass fuels. There is evidence of the negative impact of energy use on wood resources around cities.

The production, transport and distribution of biomass is generally a private sector business. The structure varies from country to country but it is mostly an informal sector activity. As investment needs have been small, biomass investment has been neglected by governments and international nancial institutions. Production costs of traditional fuels have been low because the opportunity cost of labor spent on collection is low and the full cost of reproducing biomass is not included in the price. This is not a small matter because biomass commerce in some cities exceeds electricity sales. Yet because it is a common property resource, there are some exaggerated fears that biomass could disappear, another example of The Tragedy of the Commons.

Directly addressing the biomass problem is dif cult. Fuelwood, usually gathered as a common property resource, is essentially rubbish. Wood needs to be cleared if agriculture is to take place. Dead wood also occurs naturally. Rarely do people cut whole trees, preferring to lop branches not least because they only have hand tools for felling. Rarely do they directly plant trees primarily for rewood. This suggests that, despite the increasing time required to gather biomass, people still consider biomass a cheap fuel that is readily available. It suggests, too, that people view trees as multipurpose, e.g., providing fruit, fodder, thatch and shade; trees only become fuel at the end of useful life, after serving these multipurpose ends.

The broad strategy to address the rural energy problem is to create an energy transition by moving people up the energy ladder (*see* **Biomass Use for Urban Fuels in Developing Countries**, Volume 3). Most conventional fuels, notably electricity and gas, are more expensive than biomass. As a consequence, the energy transition is slow and restricted to high value low intensity enduses such as domestic lighting. The transition will require the eradication of rural poverty.

In the biomass sector itself, there are two strategies available to address energy problems, supply enhancement and increased ef ciency of use. Supply enhancement has focused on four interventions, namely:

1. increased off-take from national parks rather than encouraging vegetation climax as a management goal;
2. plantations;
3. woodlots;
4. *agroforestry*, sometimes called *social forestry*.

The rst three interventions have been largely unsuccessful, not least because wood production is far from point of consumption. With plantations and woodlots, as well, the emphasis on the single exotic species does not t well to local consumption patterns of multipurpose trees. Agroforestry seems to hold the key, locating production in and around family farms. This, however, requires re-educating foresters and agricultural extensionists that trees enhance local livelihood systems.

The second strategy, ef ciency, has focused mainly on improved cooking stoves. These stoves, for which there are many competing designs, replace the traditional open hearths. Most designs have one thing in common an enclosed combustion chamber. That enclosure creates its own problems because, since most cooking is done in the late evening, the light from the open re is lost. Light is a major simultaneous enduse of open res, which is lost by enclosing combustion chambers. Traditional open res have other simultaneous enduses, such as boiling, cooking and space heating, that require designers to consider social, as well as physical, ef ciency. The most successful design has occurred in Africa, where a semi-enclosed stove, modeled on Asian stoves, sells well at urban bus depots for transfer to rural markets.

Developing an energy policy can start from its many practical and strategic advantages of biomass. Firstly, biomass is local and ts local livelihood systems. Secondly, unlike other hydrocarbon fuels, it is carbon (C) neutral and does not contribute to greenhouse gases. Most importantly, for developing countries, it is a saving on foreign exchange.

Biomass should be managed sustainably rather than being mined. The sustainability of the biomass resource can be enforced by:

- pricing biomass at its full reproduction cost rather than its short run exploitation costs;
- devolving management, and political control, to local communities;
- establishing legal instruments for controlled cutting.

Increased prices for biomass fuels will encourage the development and acceptance of more ef cient and less polluting combustion technologies. *See* **Agroforestry**, Volume 3.

FURTHER READING

Hill, R, O Keefe, P, and Snape, C (1995) *The Future of Energy Use*, Earthscan, London.

Leach, G (1992) The Energy Transition, *Energy Policy*, 20(2), 116 123.

Leach, G (1995) *Global Land and Food in the 21st Century: Trends and Issues for Sustainability*, Polestar Series Report No. 5. Stockholm Environment Institute, Stockholm, Sweden.

Munslow, B, Katerere, Y, Ferf, A, and O Keefe, P (1988) *The Fuelwood Trap*, Earthscan, London.

UNDP (1997) *Energy after Rio: Prospects and Challenges*, United Nations, New York.

Van Gelder, B and O Keefe, P (1995) *The New Forester*, Intermediate Technology Publications, London.

Biomass Fuel Power Development

Rik Leemans

National Institute of Public Health and the Environment, Bilthoven, The Netherlands

Plants capture sunlight and use its energy to transform carbon dioxide (CO_2) and water (H_2O) into organic matter. The resulting biomass is the basis of all life on Earth. Traditionally, humans have strongly relied on biomass as a source for food, fuels, building materials and many other uses. During recent centuries, however, the role of biomass as the primary energy source has declined in favor of the more effective and seemingly unlimited fossil fuels.

Since the oil crisis in the 1970s, many countries have tried to develop an independent energy source. Consequently, the use of biomass as an energy carrier has emerged again. For example, Brazil set up an advanced program to produce ethanol from sugarcane, which is used as transportation fuel. The cane residues provide the energy needed to drive the production process. Other biomass programs have been developed in several countries but these have been less commercially viable (Johansson et al., *1993). Liquid fuels from cornstarch and rapeseed in temperate zones have been produced but this has not been very ef cient.*

Many different plant species act as biomass crops. The actual selection is dependent on environmental factors and the energy conversion technology. Traditionally, wood chips, manure and crop residues and other waste materials were used. Nowadays, dedicated energy crops consist of fast growing species, optimized to be cultivated under local conditions. These crops include short-rotation woody shrubs and trees, herbaceous plants, halophytes (in severe environments) and some annual crops, such as oilseeds. The woody and herbaceous biomass is generally transformed into heat or electricity in local plants. The ash is returned to the plantations as fertilizer. Sweden, for example, has developed an extensive system for city heating based on short-rotation willow plantations. Currently, the efficiency of converting biomass into energy is relatively low but this could increase by improved technologies, such as combined heat and electricity generation. A promising technology is to convert biomass into syngas (a mixture of carbon monoxide (CO) and hydrogen (H)) and subsequently use this to generate electricity in a fuel cell.

The increase in atmospheric carbon dioxide (CO_2) concentrations, due to the burning of fossil fuels and deforestation, has generated additional opportunities for modern biomass. The production of biomass withdraws CO_2 from the atmosphere, while its use releases this again. As such, biomass is carbon (C) neutral and contributes to reducing CO_2 emissions or stabilizing CO_2 concentrations. Many assessments of the future energy-carrier mix depict a larger share of modern biomass. The Intergovernmental Panel on Climate Change (IPCC), for example, has already projected a large increase in modern-biomass use in its 1992 scenarios (Pepper *et al.*, 1992). IPCC has further concluded (Watson *et al.*, 1996) that large-scale use of modern biomass and other renewables in combination with a doubled energy efficiency could cut CO_2 emissions by almost 90%. This conclusion was based on the application of known technologies. One of the underlying assumptions, however, is that the conversion of biomass in fuel cells will soon become possible. This means a rapid maturation and implementation of this new technology, which is unlikely.

The use of biomass has more advantages than only reduced CO_2 emissions (Hall and House, 1995). First, the sulfur (S) content of biomass is relatively low, so that burning provides low air-pollution levels, especially when this is done in controlled and efficient furnaces. Second, biomass can be cultivated in many different climates and on different soil and terrain types. The selection of biomass crops can be optimized for local production systems. Third, modern biomass cultivation can contribute to restoring degraded lands. Fourth, biomass crops can diversify farm systems, which result in more robust economic entities. Especially in developing countries, the cultivation of biomass could assist in generating additional income for local farmers. In the industrialized world, biomass could well become a major crop on agricultural land that is taken out of production. Modern biomass crops can easily be produced at small and intermediate scales. Fifth, modern biomass

crops could fulfill additional functions, such as erosion protection, shading, shelterbelts, windbreaks and providing natural habitats for biodiversity. Finally, large-scale use of modern biomass decreases the dependence of imported fossil fuels.

Despite the obvious advantages, disadvantages of modern biomass crops should not be neglected. First, biomass crops are bulky and their transportation requires energy, which reduces the overall energy efficiency. Modern biomass is therefore only transported over short distances and is most suitable for local applications. Modern biomass is thus not as versatile as fossil fuels. The likely future efficient conversion to gas or electricity reduces this disadvantage but requires large investments in research and developing the appropriate infrastructure, such as an electricity grid. Second, the large-scale cultivation of modern biomass requires land. In many regions land availability is a lesser issue because of current overproduction in the agricultural sector. Leftover land could well be used for modern biomass plantations. In developing regions, however, with an increasing demand for food due to larger populations and changing consumption patterns, land could become limited (*see* Biomass Burning in Rural Homes in Tropical Areas, Volume 3). Biomass plantations could well compete with food crops and/or pastures. The competitive abilities are strongly dependent on market prices of biomass and food and many other local environmental and cultural conditions. Competition for land has rarely been comprehensively assessed in studies on the feasibility of biomass crops. Finally, biomass crops generally require longer rotation times. This makes them more prone for disturbances, such as pest and fires. But proper management can probably overcome this.

To increase the share of modern biomass in the mix of energy carriers, research should focus on improving the conversion efficiency of modern biomass for producing heat, electricity, and liquid fuels. Additionally, growing modern biomass crops on a very large scale has consequences for competition with other land uses and possibly impacts on land degradation, deforestation and biodiversity. These impacts, unfortunately, are not very well studied. Overall, one should conclude that modern biomass could become a valuable and environmentally friendly energy carrier, which is truly renewable and can easily be adapted to local and regional conditions and circumstances. Most signs indicate an increased market share of modern biomass in many regions.

REFERENCES

Hall, D O and House, J (1995) *Biomass: an Environmentally Acceptable Fuel for the Future. Proceedings of the Institution of Mechanical Engineers, Part A, J. Power Energy*, 209, 203 213.

Johansson, T B, Kelly, H, Reddy, A K N, and Williams, R H, eds (1993) *Renewable Energy: Sources for Fuels and Electricity*, Island Press, Washington, DC, 1160.

Pepper, W, Leggett, J, Swart, R, Watson, J, Edmonds, J, and Mintzer, I (1992) *Emission Scenarios for the IPCC: an Update. Assumptions, Methodology and Results*, Report, Geneva, 115.

Watson, R T, Zinyowera, M C, and Moss, R H, eds (1996) *Climate Change 1995. Impacts, Adaptations and Mitigation of Climate Change*, Cambridge University Press, Cambridge, 878.

Biomass Use for Urban Fuels in Developing Countries

Phil O'Keefe
University of Northumbria, Newcastle, UK

Most of the world's urban population is in developing countries. Even Africa, largely unurbanized, has a larger urban population, some 250 million people, than that of North America. Over 600 million urban dwellers live in life threatening *or* health threatening *circumstances because of overcrowded housing, dangerous land sites and a lack of basic services. Energy is one such service. In developing countries, urban energy problems are dominated by the household and transport sectors.*

*In the household sector, the single largest enduse of energy is cooking. Biomass resources (*see ***Biomass Fuel Power Development****, Volume 3) predominate, principally charcoal and wood. Charcoal, a pyrolytic conversion of wood, is used because per unit weight it has twice the calori c value of wood and, because of its lower moisture content, is signi cantly cheaper to transport from rural to urban areas. Charcoal making usually requires whole tree destruction. Using traditional charcoal making methods, it takes 10 sacks of wood to make one bag of charcoal, suggesting a signi cant energy subsidy from rural to urban areas. Biomass use creates both indoor pollution and contributes to outdoor air pollution.*

Wood and charcoal, as energy carriers, are at the bottom of what is called the *energy ladder*. Each ring of the ladder corresponds to a dominant fuel used by different income groups. The energy ladder, from bottom rung with the poorest income group, is wood, charcoal, kerosene, liquid petroleum gas (LPG) and electricity.

Movement up the ladder corresponds with greater energy efficiency (i.e., the fraction of energy reloaded from the

carrier that is actually turned into an energy service by the enduse device) and cleanliness. For example, cookstove efficiencies vary from 15% for woodfuel to 65% for gas. Moving up the energy ladder results in lower levels of carbon (C), sulfur (S) and particulate release. If significant numbers of people move up the ladder they will create an energy transition.

In theory, households choose energy carriers on the basis of income. Accessibility, convenience, controllability, cleanliness, efficiency and cost are the basis of choice. In the developing world, however, the transition by such choice is not straightforward. The energy carrier itself, its storage technology and the enduse devices themselves are subject to non-availability and breakdown. Kerosene can be limited because of an absence of foreign exchange to finance imports; LPG bottles may be in short supply; electrically operated machinery is frequently non-operational. Energy service security is important rather than income elasticity in determining energy choice. As a consequence, richer households tend to retain access to all energy-technology combinations rather than moving up the energy ladder.

The lack of movement up the energy ladder is compounded by the hidden subsidies on commercial energy, particularly electricity to urban households. These hidden subsidies mean that insufficient cash is raised by tariffs to allow equipment replacement in the power sector leading, in turn, to frequent system collapse. Energy insecurity blocks energy transition.

Parallel to the urban household energy problem is that of transport. Energy use for transportation grows rapidly with urbanization. As transport is predominantly oil-based, developing countries without oil resources face significant foreign exchange expenditures as transport accounts for more than 25% of all commercial energy in these countries.

Roads provide most transport. Essentially this is a private rather than a public solution leading to significant traffic congestion, and air pollution, as car ownership increases. Technical efficiency tends to be low in developing countries, because of older motor vehicles, although loading factors tend to be higher. Urbanization, economic growth and industrialization also require large increases in freight movement for access to markets.

Improvement in the transport system can be made by:

- moving to more efficient technologies;
- shifting to mass transit systems;
- improving the quality and mix of transport infrastructure.

Policies that encourage large shifts to public transit systems have been quite successful in Singapore and Manila, and reduce overall energy demand. Such shifts do, however, usually require changes in urban land-use planning.

FURTHER READING

Birk, M and Bleviss, D (1991) *Driving New Directions: Transportation Experiences*, International Institute for Energy Conservation, Washington, DC.

Hosier, R H and Dowd, J (1987) Household Fuel Choice in Zimbabwe: an Empirical Test of the Energy Ladder Hypothesis, *Resour. Energy*, 13(9), 347 361.

Salterthwaite, D (1997) Urban Poverty: Reconsidering its Scale and Nature, *IDS Bull.*, 28(2).

UNDP (1997) *Energy after Rio: Prospects and Challenges*, UN, New York.

US Congress Office of Technology Assessment (1991) *Energy in Developing Countries*, OTA-E-486, Washington, DC.

Bioremediation

Bioremediation is the use of biological agents to control or remove pollutants and similar challenges to environmental health. Non-biological cleanup of pollution usually requires an intense physical and chemical attack on the pollutants (e.g., physically gathering oil spills or making them easier to handle by the use of detergents). The ecological hazards of the clean-up agents may be very serious, and clean-up costs are high. Bioremediation seeks to control pollution by using microbes and other biological agents to break down the pollutants to ecologically harmless substances.

The discovery of bacteria that can digest oil led to intensive research to select more active species or strains, with considerable success. A similar approach has been successful in improving the quality of wash water from mining operations, particularly those in which acidity from the rock results in the solution of toxic minerals that must then be removed. The detoxification of organic pollutants by fungi and bacteria has also been tried, and a variety of specific organisms have been discovered or developed by selection or genetic modification that attack specific chemicals (e.g., specific fungi have been found, such as the mushroom *Marasmiellus troyanus*, which can digest benzopyrene, a potent carcinogen from coal tar).

Not all bioremediation uses microorganisms. Mining or quarrying effluent may contain substantial dissolved toxic chemicals as well as suspended solids and silt. Environmental pollution from such sources has been effectively controlled by constructing an artificial swamp on the effluent watercourse, selecting plants (e.g., sphagnum and *Typha*) that have the capacity to sequester the toxic minerals or convert them to non-toxic derivatives, as well as filtering out the sediments. Some freshwater aquatic plants have been

investigated that detoxify mercury, and a selected alfalfa cultivar has been used to clean up soil nitrogen contamination from overfertilization, spills and livestock wastes.

Bioremediation has been applied successfully to sewage effluent, mining wastes, chlorinated solvents, pesticides, agricultural chemicals, oil, gasoline, creosote and a variety of specific organic chemicals including dry-cleaning wastes. However, the use of added microorganisms is not without some hazards. It has been necessary to monitor the environmental consequences of loading the ecosystem with foreign species, which may not simply disappear when the pollution has been cleared up. Similar problems have arisen in other types of biocontrol, such as the importation of foreign predators to control pests, when the predator does not always confine its attention only to the pest in question, but may attack other (sometimes desirable) species when the target organism has been reduced or eliminated. The more recent approach of taking advantage of naturally occurring organisms tends to avoid this problem and reduces environmental stress. However, some ecosystem disturbance usually occurs because it is often necessary to add nutrients to the polluted area to encourage the bioremedial agents. In balance, bioremediation has been a valuable and versatile approach to pollution control, and is usually much cheaper and less ecologically traumatic than other methods of clean-up.

R G S BIDWELL *Canada*

C

Cadmium (Environmental) in the Food Chain

see Environmental Cadmium in the Food Chain (**Volume 3**)

Cadmium in the Rhine River Basin

see Materials Flow Accounting (**Volume 4**)

CAP (Common Agricultural Policy)

see Common Agricultural Policy (CAP) (**Volume 3**)

Capacity, Assimilative

The empirically estimated capability of a natural ecosystem in a particular locale to accept and process wastes into an acceptable form has been called its *assimilative capacity* for those wastes. A century ago that concept was used mostly for the case of untreated human sewage that was led into owing waters. Immediately downstream of the sewage outlet, the natural decomposition processes resulted in offensive and harmful gases like ammonia and methane, and people were expected to stay away from such sites for their own good. The actual numerical estimate of the assimilative capacity, i.e., of how much sewage could safely be released into a particular spot in a particular river, required a prior decision on how far down the river such offensive products of decomposition would be tolerated by humans with other interests in those downstream parts of the river. If an important downstream use was threatened by pollution, then the assimilative capacity for pollution upstream would be set at a low level. The amount of sewage wastes that could be processed ecologically within a particular reach of a river was itself dependent on such factors as volume and rate of river ow, water temperature and whether the river had rapids or was owing only placidly. So the operative estimate of assimilative capacity would have to take such factors into account.

The notion of safety in such circumstances has interesting features. That the waters immediately downstream from an outfall of raw wastes were unsafe was clear. What rendered this regulatory approach safe was that the unsafe locales were zoned, formally or informally, to keep humans at a safe distance. Also, the numerical value of assimilative capacity that was formalized legally was not based on the worst possible scenario, say of some episodic event of excessive loading due to sewerage system malfunction or of unusually low river ow. At such times, downstream users were subjected to higher risks and were expected to ensure their own safety by compensating reactively, say by quickly stopping their own uses of the river until the risks abated. So the safety of downstream users depended on their prior understanding of those risks and then acting appropriately.

It is clear from the above that the concept of assimilative capacity provided a way to place some politically negotiated limit on the externalization of implicit costs by upstream sewage producers at the expense of downstream users of the river. As a river came to be used for more and different purposes, an early measure of assimilative capacity was generally found to be too generous, and then a new, lower assimilative capacity came to be formalized. Or a growing settlement might lead to increases in amount of sewage that would exceed the agreed assimilative capacity. Authorities responsible for sewage disposal then invented cost effective means to remove the more harmful components of sewage so as not to exceed the agreed assimilative capacity of the river. In developed countries such ratcheting down of the concentration of harmful substances released at the sewage outfall occurred by numerous steps. It is now feasible to process sewage technologically so that the ef uent has a quality appropriate for drinking water. If this is deemed to be cost effective, then it may be politically acceptable to assign the river in such a locale an assimilative capacity of zero with respect to sewage wastes. This may occur where politically powerful interests downstream

practice sensitive uses such as angling for sport fish that are particularly intolerant of ecological degradation of the river.

Setting assimilative capacity at zero is consistent with a rigorous interpretation of the *precautionary principle* in which a potential polluter pays the full costs of internalizing complete treatment of all the sewage.

We have used human sewage as an example for which a concept of assimilative capacity has long been used. But that concept was also used for industrial wastes such as gases, acids, bases, heavy metals, oils and complex chemicals that may be released into the waters, soils or atmosphere of a locale. To the extent that the relevant estimate of assimilative capacity was based on scientific understanding, as was usually expected, the amount of scientific information needed quickly became excessive. Part of the complication was due to the occurrence of adverse interactions amongst the ecological consequences of the assimilative processes of the different wastes that acted to exacerbate some of the separate adverse effects. It seldom happened that the adverse effects from one kind of waste resulted in the inactivation of the adverse effects of another kind of waste.

In developed countries this all came to be recognized in the 1960s, say by Sweden when that country triggered the first of the United Nations conferences on the environment in Stockholm in 1972. A complex systems approach, that eventually came to be called an ecosystem approach (*see* **Monitoring in Support of Policy: an Adaptive Ecosystem Approach**, Volume 4) in North America, then began to take the place of the excessively complicated linear approach of legal specification of different assimilative capacities for a large number of different wastes in an industrial city, say. The ecosystem approach led to an inference that a legal requirement for precaution was of critical importance.

As a transitional convention, the concept of assimilative capacity has come to be transformed into a concept of a temporary benchmark of *maximum permissible loading*. With an understanding that such a benchmark will likely be lowered by steps in the future toward a goal of zero discharge from current activities and virtual elimination of such wastes that remain in the environment from previous emissions. For example, a goal of zero discharge and virtual elimination was formalized for the Great Lakes by Canada and the US in 1978 with respect to particular hazardous contaminants; since then temporal benchmarks have been specified for some of these contaminants.

A goal of zero discharge into a nearby river, say, and immediate virtual elimination of all human and pet wastes from a city would likely be perceived as very unrealistic. Sewage infrastructure does malfunction occasionally, storms wash pet wastes from parks into streams, derelict parts of a city do not have sanitary conveniences, etc.

But a goal of zero discharge and virtual elimination, i.e., a legally specified assimilative capacity of zero, may be achievable for some excessively harmful kinds of wastes. These may include particularly hazardous forms of physical substances such as radionuclides, of chemical substances such as organohalides and of biological substances such as viruses and other pathogens. Infinitesimally small quantities of such substances can cause permanent harm to organisms, with initiation of disabilities at the time of fertilization or conception. Such effects include mutagenesis that involves harmful genetic mutations, carcinogenesis that triggers cancer, teratogenesis that causes permanent physical and physiological impairments, etc.

Clarification of what substances have properties for which a zero assimilative capacity is justified became the focus of international legal negotiations late in the 20th century. These negotiations have led to bans on the manufacture of some substances, such as persistent pesticides, that could only be used in ways that resulted in losses to the natural environment which, in effect, had already been assigned zero assimilative capacities for these substances.

HENRY REGIER *Canada*

Carbon, Sulfur, and Nitrogen: Trends in Global Emissions

see Trends in Global Emissions: Carbon, Sulfur, and Nitrogen (*Opening essay*, **Volume 3**)

Cattle Grazing: Impacts on Land Cover and Methane Emissions

Tapas Kumar Bandyopadhyay
Ministry of Environment and Forests, New Delhi, India

Methane (CH_4) is a naturally occurring greenhouse gas. It is generated due to bacterial action on organic material in anaerobic conditions. In nature, it is produced by the decay of organic materials in swamps.

Human activities are also a cause of the release of greenhouse gases into the atmosphere, accounting for about 70% of the total CH_4 emissions. Approximately 20% of total

CH_4 generation in the world is contributed by paddy fields. Livestock contributes about 16% of the total.

CH_4 is generated during the enteric fermentation process and from the anaerobic decomposition of animal wastes. It is estimated that domestic animals account for 94% of the total global CH_4 emissions by animals and human beings. Of that, cattle contribute 74% of animal emissions and sheep and goats contribute 13%.

PATHWAY OF METHANE EMISSIONS FROM ANIMAL SOURCES:

There are two sources of methane (CH_4) emissions from livestock (see Table 1):

1. From the digestive process,
2. from animal wastes.

CH_4 Emissions From the Digestive Process of Ruminants

The generation of CH_4 from the digestive processes of an animal depends upon the type of food intake, quality of food, and quantity of available food. Climatic conditions and seasonality affect CH_4 emissions. In a tropical country like India, higher CH_4 emissions from animals occur during the winter and rainy seasons. The age of an animal is also one of the most important factors for the production of CH_4.

Different types of microorganisms present in the digestive tracts of herbivorous animals are responsible for the production of CH_4 from digestive process of ruminants. Heterotrophic microorganisms decompose organic materials found in the feedlot.

There are three stages for decomposition of organic material in the digestive process of the ruminants. In the first stage the cellulosic materials are hydrolysed by cellulase into simple sugars (e.g., glucose and maltose). The soluble sugars are converted to simple organic acids by anaerobic and facultative bacteria (Figure 1). The organisms responsible for these transformations, called acid formers, function best in a range of pH 4.0 to 6.5. The major product of this step is acetic acid but propionic acid and butyric acid are also produced. Acetic acid is the most important substrate for the final reaction of the sequence, since 70% of the CH_4 produced has been shown to derive from that component. CH_4-producing bacteria (Methanogenes) convert acids to CH_4 and carbon dioxide under strict anaerobic conditions.

CH_4 Emissions From Animal Wastes

Animal wastes are generally composed of organic material and water. Under favorable environmental conditions, organic material present in animal wastes is decomposed by a group of organisms to become CH_4, carbon dioxide and other products. Primary microorganisms responsible for the decomposition of animal waste are heterotrophic bacteria. Heterotrophic bacteria can be classified as aerobic, anaerobic or facultative, depending on their need for oxygen.

Either an aerobic or an anaerobic process can break down the animal wastes. Under aerobic conditions, aerobic and facultative bacteria, using molecular oxygen, decompose the organic material. The end products of this process are carbon dioxide and stabilized organic material. Under anaerobic conditions, the animal wastes are decomposed to CH_4, carbon dioxide and stabilized organic material.

As already mentioned, organic materials in the animal waste are converted to CH_4 and carbon dioxide in three steps.

Step I: Hydrolysis by Enzymes

The first step is the enzymatic hydrolysis of animal waste. The enzyme cellulase is responsible for breaking down cellulose and starch into sugars such as glucose and maltose. Enzyme lipase breaks down lipids into smaller chain fatty acids and enzyme protease breaks down proteins into amino acids.

The extent of enzymatic reactions is controlled by the quantum of enzyme present in the system and the characteristics of the wastes and reaction conditions such as pH and temperature.

Step II: Formation of Small-chain Acids

The next step is the formation of small chain acids such as acetic acid (the main product at this stage), propionic acid and butyric acid.

$$2C_6H_{12}O_6 \text{ (glucose)} + 2H_2O \longrightarrow 2CH_3COOH \text{ (acetic acid (I))} + 2CO_2 + 4H_2 \text{ (hydrogen (II))}$$

Step III: Generation of CH_4

In this final stage, methanogenic bacteria convert acetic acid into CH_4 and carbon dioxide in strictly anaerobic conditions. This reaction process is very sensitive to temperature, pH and the composition of the substrate. The reaction can be expressed as:

$$\text{(I) } 2CH_3COOH \text{ (acetic acid)} \longrightarrow 2CH_4 \text{ (methane)} + 2CO_2 \text{ (carbon dioxide)}$$

$$\text{(II) } 4H_2 \text{ (hydrogen)} + CO_2 \text{ (carbon dioxide)} \longrightarrow CH_4 \text{ (methane)} + 2H_2O \text{ (water)}$$

Table 1 Current CH_4 sources (Tg $year^{-1}$)

Source	Khalil and Rasmussen (1987)	Bingemer and Crutzen (1987)	Cicerone and Oremland (1988)	IPCC (1990)	Fung *et al.* (1990)
Ruminant	120	70–80	65–100	65–100	80
Paddy fields	95	18–91	60–170	–	100
Biomass burning	25	30–100	50–100	20–80	45
Land fills	–	30–70	30–70	20–70	40
Coal mining	–	35	25–45	–	35
Gas leaks and vents	40	0–45	25–50	40–100	40
Other anthropogenic	40	–	–	–	–
Swamps and marches	150	26–137	100–200	100–200	80
Lakes and oceans	23	–	Jun–45	Jun–45	10
Tundra	12	–	–	–	35
Other natural	48	0–30	10–100	–	25
Total anthropogenic	320	183–411	255–535	145–350	350
Total natural	233	26–167	116–345	116–445	150
Total	553	209–578	371–880	261–795	500

IPCC, Intergovernmental Panel on Climate Change.

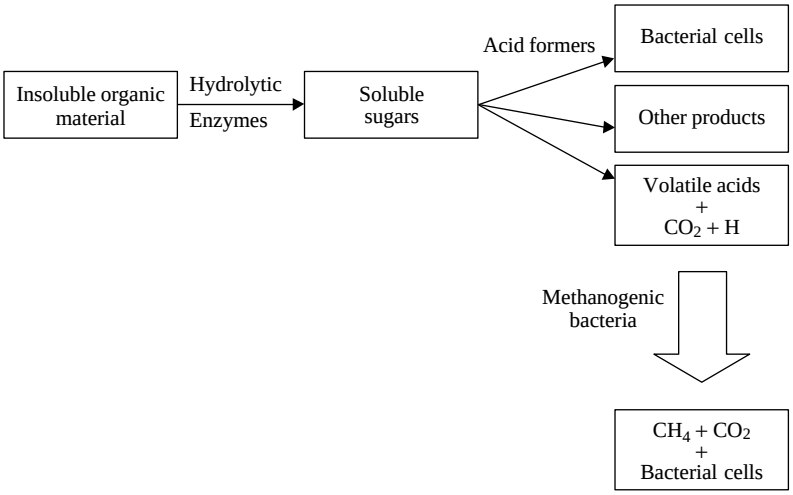


Figure 1 A simplified schematic of the overall mechanism of methane production in the digestive tract of ruminants

CATTLE GRAZING AND PASTORAL CHANGE

Cattle grazing refers to consumption of vegetation on a land area, especially grass, by cows, buffaloes, sheep and goats. The benefits of cattle raising include: the production of manure for fertilizer; the manufacture of products from leather, skin, bone, hair gut and meat; and the production of dairy products. Associated with cattle raising, there are a number of environmental evils such as the destruction of rain forests, competition with wild life for habitat, soil compaction and erosion from trampling, and pseudo problems such as CH_4 emissions from ruminant burping.

Pastures and CH_4 Emissions

There is no effective waste handling system for animal pastures. The manure from animals in these pastures is allowed to lie where it is. The animal waste dries out, reducing anaerobic decomposition. Chen *et al.* (1988) indicate that about 11% of the CH_4 producing potential of the manure is lost during the drying process. Hence, the extent of the CH_4-producing capacity in pastures is very low compared to liquid/slurry or anaerobic lagoons.

The United States Environmental Protection Agency (USEPA) estimates that CH_4 emissions from animal wastes in pasture total about 10.2 Tg year^{-1} out of which the highest contribution, 2.2 Tg year^{-1}, is from Latin American countries (USEPA, 1992) These estimates are only as good as the possibly inaccurate data on animal numbers, size, feed type, and CH_4 production potential on which they are based.

Tropical pastures able to produce year round even without irrigation are potentially more productive than those of sub-tropical or temperate regions. But due to soil moisture, tropical pastures become a better source of CH_4 emissions than temperate pastures.

Causes of Pastoral Change

A livestock management system plays an important role for the movement of animals to pastures. In most developed countries, livestock graze in confined fields or are reared indoors except in Australia and New Zealand where there is little confinement of livestock. In the developing countries in Latin America, Africa and Asia, most livestock graze on communal pastures.

The main causes of pastoral changes are:

1. clearing of tropical rain forests;
2. agricultural expansion into pastoral lands;
3. increased cropping;
4. overgrazing;
5. population pressure;
6. poor pasture management systems.

Present Global Status of Pastures

In developing countries and tropical regions traditionally unirrigated land is generally used for cattle grazing, forest development and dry-land crops.

In many parts of India, the area available for grazing has been reduced due to population pressure and conversion of grazing land into arable land. The pasture and other grazing areas need to be improved by controlled grazing through rotation and deferred, or intermittent, grazing systems.

The pastoral regions in China are located mainly in the Inner Mongolia autonomous region (IMAR), Xinjiang Uygur autonomous region and Qinghai province. At present, about 87 million ha, about one third of the useable grassland in the pastoral regions, are degraded. Every

year, another about 0.66 million ha of grassland become degraded (Liu, 1993). The causes of this trend can largely be attributed to the growth of the animal population, inadequate pasture development, and human population pressure.

In the northern Caspian Sea and Aral Sea region (Gael, 1988) the grazing areas in semi-desert areas are adversely affected due to excessive livestock grazing which has reduced the fodder available from 600 700 kg ha^{-1} to only 50 300 kg ha^{-1}. It has been suggested that the disappearing dominant species of these degraded pastures should be re-sown as soon as possible.

The farmers in Vietnam have traditionally exploited free natural grazing resources for the rearing of buffaloes and cattle (Froberg and Olsen, 1989). In the area in and around Vimu Phu province, the grazing lands are under stress due to the conversion of grazing land to agricultural cultivation on forest plantations or overgrazing.

In many parts of the Middle East and Africa, pastoral lands have become degraded through agricultural expansion into pastoral zones, the loss of critical dry season pasture, nationalization of pastoral resources and the collapse of the traditional common property resource management system (Johnson, 1993). In the Sahara-Sahel transition subzone rapid deserti cation is taking place as a result of excessive biotic pressure, the human population having increased by 226% and the animal population by 237% between 1970 and 1980 (Le Houerous, 1989). Recurrent drought and human population pressures have drastically reduced the density of pasture species in the Nigerian savannas (Igboanugo and Omijeh, 1990). The establishment of agro-pastoral farmers with plantings of eucalyptus has been suggested as a means of improving conditions in this zone.

In certain areas of Mexico intensive cattle grazing has resulted in deforestation. The increased moisture in pasture soils has lead them into becoming a source of CH_4 emission to the atmosphere.

In the State of Rondonia (Brazil), forest soils are a net sink of about 470 mg CH_4 m^{-1} while pastures are a net annual source of CH_4 from the soil of about 1 g CH_4 m^{-2}. The total annual pasture-related CH_4 release for the entire Brazilian Amazon increased from 0.8 Tg in 1970 to about 2.5 Tg in 1990 with a maximum release of 3.1 Tg in 1988. The average rate of increase in CH_4 emission from pasture was about 0.2 Tg CH_4 year^{-1} between 1975 and 1988. This represents between 12 and 14% of the global average rate of change in tropospheric CH_4 content for this time period (Flessa, 1996).

Increased cropping is one of the widely accepted causes of pasture degradation. In Australia, the primary production system used is not sustainable. About half of Victoria s crop and pasture lands are degraded or are at risk due to overgrazing and the failure to adopt an appropriate grazing and crop management system (Bird and Bicknell, 1992).

Strategies for Conservation of Pasture

To avoid irreparable damage to, and erosion of communal grazing lands, a systematic land use management system is needed. An integrated management scheme, ensuring an appropriate interrelation between arable pasture use, irrigation and reduction in grazing pressure is essential.

In China the following counter measures have been recommended to redress the imbalance: human population control and development of alternative income sources for the pastoral regions, pasture development through contracting of grazing land to individual households, joint household or production groups instead of the common grazing system, and a long-term pasture leasing system in a small number of counties with good natural, economic and social conditions in order to encourage investment in pasture development; and controlled stocking.

In order to restore damaged pastures, remedial measures like expansion of cropland, the exclusion of mobile herders from natural resource management schemes and subsidized livestock imports from developed countries must be changed or stopped. Sustainable practices are required for the conservation of pastures by proper pasture management.

LIVESTOCK POPULATIONS AND CURRENT TRENDS OF CH_4 EMISSION FROM LIVESTOCK

World-wide Cattle Population

In developed regions such as the United States, Canada and Europe, suf cient data are available to allow the populations of animals to be analyzed (Table 2). The regional percent of cattle population to total cattle population is given in Table 3.

In the developing regions of Africa and Asia livestock are not con ned. Most cattle there are multi-purpose and there are few specialized dairies. Most livestock graze on communal pastures and with little management of their wastes.

Animals produce signi cant quantities of CH_4 as part of their digestive processes. CH_4 emissions from all animals have been estimated by the IPCC to be between 60 and 100 Tg year^{-1}; accounting for about 16% of total global CH_4 emission from all source.

While the developing countries have a greater percentage of the cattle population than the developed countries, the yield of CH_4 from the digestive processes of cattle depends on the quality and amount of feed that each animal eats. The biggest eaters are dairy cows that receive three times their maintenance level feed. Hence, the average CH_4 production in the cattle of developed countries is also comparable with that in developing countries. This is due to the fact that a large portion in the cattle of developing countries is kept for draught purposes rather than for meat or milk. Such

Table 2 Worldwide cattle population[a] (×1000)

Name of region (Total)	Non-dairy cattle	Dairy cattle	Swine	Sheep	Goat	Buffalo	Others
North America (198 378)	99 199	16 521	66 146	11 336	2422	–	2754
Western Europe (309 131)	56 513	31 099	114 959	93 856	9847	105	2752
Eastern Europe (517 620)	100 866	56 800	152 757	188 159	9018	581	9439
Oceania (269 715)	27 610	4441	5003	228 982	2022	1100	557
Latin America (580 008)	271 771	37 560	78 150	117 312	34 649	3516	37 050
East and Sub-Saharan Africa (283 906)	68 506	12 965	2218	110 395	80 044	–	9778
West and Southern Africa (211 323)	64 692	5769	10 227	68 776	58 904	–	2955
Near East and Mediterranean (33 125 011)	41 046	17 174	174	187 502	64 717	3516	17 145
Asia and Far East (1 385 011)	308 794	45 240	403 231	202 442	253 990	131 604	39 710

[a] Cattle include non-dairy cattle, dairy cattle, swine, sheep, sheep goat, buffalo, horse, mule, donkey and camel.

Table 3 Percent contribution of cattle (cows and bulls) population to the total cattle population

Name of region	Cattle population (×1000)	% contribution cattle population to the total cattle global population
North America	198 378	4.85
Western Europe	309 131	7.56
Eastern Europe	517 620	12.66
Oceania	269 715	6.60
Latin America	580 008	14.20
East and Sub-Saharan	283 906	6.95
Africa West and Southern Africa	211 323	5.18
Near East and Mediterranean	331 274	8.11
Asia and Far East	1 385 011	33.89

developing country cattle are not fattened like dairy cows, and so produce less CH_4 per animal.

Trends of CH_4 Emission From Animal Wastes

A USEPA (1992) report estimates that global CH_4 emissions from animal wastes are about 28 Tg year^{-1}; with a range of about 20 to 35 Tg year^{-1}; about 6 to 10% of total annual anthropogenic emissions. Out of 28 Tg year^{-1}; three regions account for 78% of the total: Europe (Eastern and Western) with 11.4 Tg year^{-1}; (40%); Asia with 7.1 Tg year^{-1}; (25%) and North America with 4.2 Tg^{-1}; (15%) (Table 4).

Table 4 Global livestock waste CH_4 emission by region and system (Tg year^{-1})

Waste management system	North America	Europe	Oceania	Latin America	Africa	Asia	Total
Pasture/range	1.3	2.0	1.2	2.2	1.3	2.3	10.2
Liquid/slurry	0.4	5.8	0.0	0.0	0.0	0.9	7.2
Solid storage	0.1	2.0	0.0	0.0	0.0	0.0	2.1
Anaerobic lagoon	1.5	0.5	0.1	0.0	0.0	0.7	2.8
Dyrlot	0.3	0.1	0.0	0.1	0.1	0.9	1.5
Burned for fuel	0.0	0.0	0.0	0.0	0.0	1.0	1.0
Daily spread	0.1	0.1	0.0	0.1	0.0	0.2	0.5
Other systems	0.6	0.9	0.0	0.4	0.1	1.0	3.0
Total	4.2	11.4	1.3	2.9	1.5	7.1	28.3

Source: Reproduced by permission of the US EPA.

METHODS FOR REDUCING GLOBAL CH_4 EMISSIONS

Technologies and practices exist that can reduce CH_4 emissions by 50 80% from animal waste management systems that are used for large numbers of confined animals.

In most developing countries like India, Bangladesh and Sri Lanka, about 95% of cattle are confined in small animal holdings and only 5% of the total cattle are managed by organized agribusiness. So, in order to reduce CH_4 emissions from animal wastes, animal wastes need to be collected and treated for recovery of CH_4. The recovered CH_4 (bio-gas) can be used as an alternative energy source to heat and light houses.

Reduction of CH_4 emissions can be achieved by modifying the composition of the diet. Most of the cattle in developing countries survive on grass and rice/wheat straw. Hence, it is necessary to find out the exact feed combination that will produce the least CH_4.

Other approaches for reducing CH_4 emission from digestive process of ruminants include: the elimination of protozoa in the ruminants will results in lower CH_4 emissions and may enhance animal performance, improved ruminant performance through the use of locally produced feed supplements and the improvement of fiber digestion efficiency and inhibition of enzyme activity of methanogens through a biotechnological approach.

REFERENCES

Bingemer, H G and Crutzen, P Z (1987) The Production of Methane From Solid Wastes, *J. Geophys. Res.*, 92, 2181 2187.

Bird, P R and Bicknell, D (1992) The Role of Shelter in Australia for Protecting Soils, Plants and Livestock, *Agrofor. Syst.*, 20, 59 86.

Chen, T H, Day, D L, and Steinberg, M P (1988) Methane Production from Fresh Versus Dry Dairy Manure, *Biol. Wastes*, 24, 297 306.

Cicerone, R J and Oremland, R S (1988) Biogeochemical Aspects of Atmospheric Methane, *Global Biogeochem. Cycles*, 2, 299 327.

Flessa, H and Dorsch, P (1996) Influence of Cattle Wastes on Nitrous Oxide and Methane Fluxes in Pasture Land, *J. Environ. Quality*, 25, 1366 1370.

Froberg, A and Olsson, A C (1989) *Grazing and Fodder Production: a Minor Field Study Report for the Forest, Trees and People Project* (ETP) Vietnam, Working Paper, International Rural Development Centre, Swedish University of Agricultural Sciences, No. 122, 79.

Fung, I, John, J, Lerner, J, Matthews, E, Prather, M, Steele, L P, and Fraser, P J (1990) Three Dimensional Model Synthesis of the Global Methane Cycle, *J. Geophys. Res.*, 96(D7), 13 033 13 066.

Gael, A G (1988) Shelterbelts Protecting the Pastures in the Northern Caspian and Aral Regions, *Lesnoe Khozyaistvo*, 8, 33 35.

Igboanugo, A B I and Omijeh, J E (1990) Pasture Floristic Composition in Different Eucalyptus Specifics Plantation in Some Parts of Northern Guinea Savanna Zone of Nigeria, *Agrofor. Syst.*, 12, 257 268.

Khalil, M A K and Rasmussen, R A (1987) Atmospheric Methane: Trends Over the Last 10 000 Years, *Atmos. Environ.*, 21, 2445 2452.

Le Houerous, H N and Le Houerous, H N (1989) The Grazingland Ecosystems of the African Sahel, *Ecol. Stud.*, 75, 27.

Liu, Y M and Longworth, J W (1993) Range Land Degradation in Pastoral Region of China: Causes and Countermeasures, *ACIAR Technical Report Series*, 25, 20 26.

USEPA (1992) Global Methane Emission from Livestock and Poultry Manure, *USEPA Report Series*, Washington.

FURTHER READING

Ahmed, N (1990) The Agricultural Environment in Latin America and the Caribbean and the Greenhouse effect, *Dev. Soil Sci.*, 20, 249 326.

Bandyopadhyay, T K, Goyal, P, and Singh, M P (1996) Generation of Methane from Paddy Fields and Cattle in India, and its Reduction at Source, *Atmos. Environ.*, 14, 2569 2574.

Bayer, W (1990) Behavioral Compensation for Limited Grazing Time by Herded Cattle in Central Nigeria, *Appl. Ani. Behav. Sci.*, 27, 9 19.

Bayer, W and Grell, H (1994) Pastoralism and Fighting Desertification, *Entwicklng Landticher Raum*, 28, 6 9.

Foley, G (1991) *Global Warming*, Panos Publications, London.

Hawke, M F and Jarvis, P G (1991) Pasture Production and Animal Performance Under Pine Agroforestry in New Zealand, *For. Ecol. Manage.*, 45, 109 118.

Hodgkinson, K C and Terpstra, J W (1989) Grazing Pressure by Sheep; Consequences for Pasture Stability in an Australian Mugla Woodland, *Proceedings of the XVI International Grassland Congress*, Nice France, 4 11 October 1989: 1069 1070.

Johnson, D L (1993) Nomadism and Desertification in Africa and the Middle East, *Geojournal*, 31, 51 66.

Lassey, K R and Ulyatt, M J (1997) Methane Emission Measured Directly From Grazing Livestock in New Zealand, *Atmos. Environ.*, 18, 2905 2914.

Ling, H and Chen, Z (1997) Changes in Soil Carbon Storage Due to Over Grazing in Leymus chinensis Steppe in Xilm River Basin of Inner Mongolia, *J. Env. Sci.*, 4, 486 490.

Loker, W M (1993) Feed Resources and Land use on Farms in the Puca upa Region, Peru, *Pasture Tropicales*, 15, 34 38.

Moss, A (1992) Methane From Ruminates in Relation to Global Warming, *Chem. Indust., Lond.*, 9, 334 336.

Stendler, P A and Melillo, J M (1996) Consequence of Forest-to-Pasture Conversion on Methane Fluxes in the Brazilian Amazon Basin, *J. Geophys. Res.*, 101(D13), 18547 18554.

Thakur, D R (1993) Balancing the Ecology and Development of Forests in Himachal Pradesh, *Indian J. Reg. Sci.*, 25, 61 71.

Vavilim, V A and Lokshima, L Y (1997) Modeling Methanogenesis During Anaerobic Conversion of Complex Organic Matter at Low Temperature, *Water Sci. Technol.*, 6–7, 531 538.

Cereal Cultivation (Agriculture, Extensive)

William J Carlyle
University of Winnipeg, Winnipeg, Canada

Extensive cereal cultivation, the most recently developed of the world's main agricultural systems, is found largely in areas of former temperate grasslands. The emphasis from the beginning of cultivation has been on wheat production. The system was initially made possible by the use of machinery, and adaptations of crops and cropping practices to suit sub-humid and semi-arid environments. The adoption of ever larger-scale machinery has led to increased farm sizes, especially since World War II.

The conversion of tens of millions of ha of perennial grasslands to cropland is the clear environmental impact of cereal cultivation. Considerable controversy surrounds other impacts. Soil organic matter has decreased and soil erosion and soil salinity have increased since original cultivation. However, the reduction of summerfallow since 1970 and more recent reductions in tillage have slowed rates of organic matter loss, erosion, and salinity.

Tree planting in some areas to ameliorate the local climate has brought only minor bene ts. Shelterbelts around farmsteads or farm villages have, however, provided beauty, protection from the wind and sun, and a relief from uninterrupted horizons. Field shelterbelts are not extensive overall, but where they have been established, they also provide relief from horizontal horizons, and they help reduce soil erosion by wind.

Climate changes mainly due to urban and industrial activities elsewhere have already begun and are continuing in the extensive cereal districts. It is generally expected that average temperatures will increase in these districts, but there is uncertainty about changes in precipitation. A rise in temperature, without commensurate increases in precipitation and other mitigating measures, would reduce soil organic matter, add carbon dioxide to the atmosphere, and reduce crop yields. If, however, there is an increase of atmospheric carbon dioxide together with increases in temperature, precipitation, and fertilizer use, as well as a reduction in tillage and fallow, signi cant increases in yield could occur.

INTRODUCTION

Extensive cereal cultivation today occupies areas in the mid-latitudes, which were, until recently, mainly temperate grasslands (Figure 1). Cereal cultivation did not begin to replace nomadic herding and ranching in these areas until the late 1800s. By about the 1930s, the perennial grasslands had been largely ploughed up and sown to annual grasses, especially wheat, in the largest short-term expansion of cropland in history.

There are several reasons why crop farming was so long delayed in the temperate grasslands. The climate is sub-humid to semi-arid, with an annual precipitation of about 250 500 mm, but higher amounts occur in such areas as the humid parts of the Argentine Pampas (Figure 1). Moreover, the precipitation is variable from year to year, and periodic droughts occur. In North America, at least, the lack of trees was taken to be an indication that the grasslands were too dry and the soil not fertile enough for widespread crop agriculture (Grigg, 1974). In the Pampas, the estanicia (large estate) owners favored cattle and sheep, and they were generally unwilling to become crop farmers. Perhaps most important, however, was the difficulty of clearing the tough prairie sod by human and animal power with traditional tools (James, 1951). Other important factors were the lack of railways to export the crops and bring in necessary equipment and supplies, including lumber, fuel, and fencing, and the lack of demand for cereal crops on the scale that the grasslands might produce them. An exception is that the steppes north of the Black Sea were partially cleared between the late 1700s and late 1800s by domestic animals and human muscle power in conditions of serfdom and virtual slavery (James, 1951).

CLEARING OF THE GRASSLANDS

The conditions that delayed the development of the grasslands into cropping areas began to change in the late 1800s. Industrialization, urbanization, population growth, and increasing incomes in Northwestern Europe and Eastern North America led to an increasing demand for wheat.

Implements such as the self-scouring steel plough, the binder and reaper, the steam powered thresher, barbed wire for fencing, and steel windmills to draw up water became available in the American Great Plains and Canadian Prairies. Help in establishing cereal cultivation in Australia came from the stripper for harvesting, the stump-jump plough for cultivating among eucalyptus roots in newly cleared land, and a harvester suitable for local conditions. In a similar manner, steel ploughs, reapers, and steam threshers were introduced to the Pampas (Grigg, 1974). Immigrants from Spain and Italy, who became crop-share tenants on the estancias, planted most of the wheat in the Pampas. They would sow wheat for several years, after which the land was sown to alfalfa for cattle feed, and the tenants then moved to other parts of the same estancia or to other estancias to repeat the process (Grigg, 1974). In the Russian steppes east of the Black Sea, plentiful and cheap labor allowed clearing without the degree of

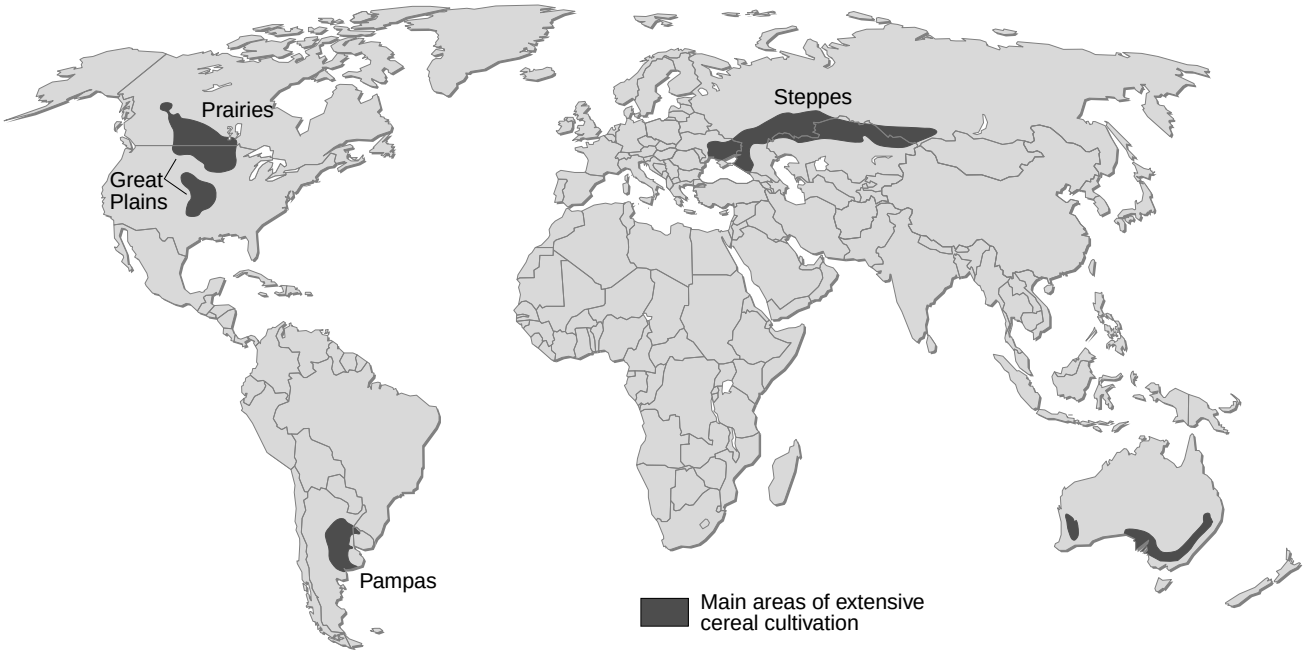


Figure 1 Main areas of extensive cereal cultivation. Excluded here are several smaller areas of extensive cereal cultivation, such as the Paris Basin and the Palouse region of Northwest continental United States of America

mechanization used elsewhere in the temperate grasslands, although mechanization was later adopted by the Soviets.

Strains of wheat were developed which were better suited to the climates of the grasslands, and summer fallowing, i.e., leaving the land without a crop for an entire growing season, was developed to accumulate moisture in semi-arid areas.

The clearing of the grasslands would probably not have occurred as quickly and expansively as it did, however, were it not for the fact that the late 1800s were the railway age. Railways were crucial both to bring in settlers and supplies and to transport cereals to distant markets inexpensively. The replacement of wooden sailing ships by iron steamships also expedited grain exports.

These developments led to the rapid expansion of cereal cultivation into the grasslands, leaving the ranchers and herders in possession of only the drier margins by the end of the main clearing period, which lasted to about 1930 (Figure 2).

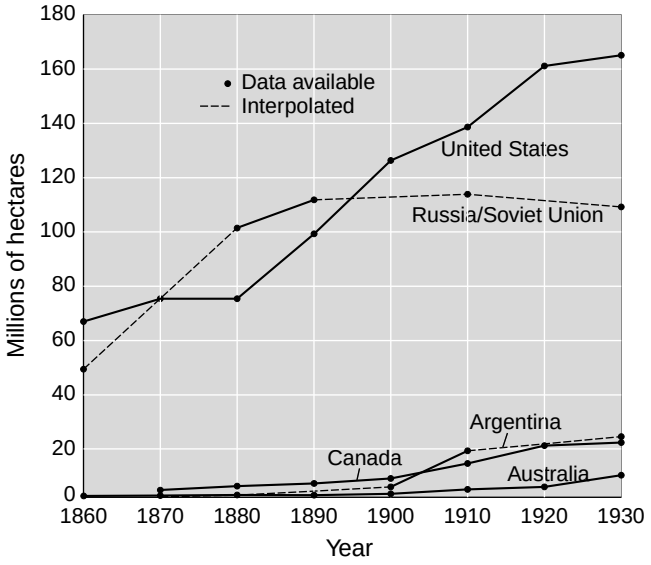


Figure 2 Trends in cropland, 1870 to 1930. Much of the increase in cropland occurred in areas of extensive cereal cultivation

RECENT DEVELOPMENTS

Since the 1930s, and especially since World War II, farming in the former grasslands has become increasingly mechanized and extensive. Combine harvesters, larger and more powerful tractors, and a wide range of other labor saving equipment, together with at land suited to large-scale machinery (James, 1951), have allowed farms to rapidly expand in size.

In the Great Plains and Prairies, where original homesteads were an already sizeable 65 ha, grain farms now are hundreds or thousands of ha in extent. In the Australian grain districts, original farm sizes of 120 160 ha have increased to 500 1500 ha, and even larger. The Pampas began with very large estancias and they have remained large (Rudolph, 1985), although some

estates have been subdivided into smaller owner-occupied holdings. Individual family farms have never been common in the steppelands of Russia, Ukraine, or Kazakhstan. The farm village or *mir* was the traditional farm unit in Old Russia, and these were replaced in the Soviet Union by collectives (*kholkhozes*) and state farms (*sovkhozes*), beginning in the early years of the Stalin era. Sizes of collectives varied, with about 2500 4000 ha being a representative figure by the 1960s. State farms were generally considerably larger.

Beginning in the 1950s, the Virgin and Idle Lands Program was launched by the Soviets. It involved the ploughing up of some 30 40 million ha of land in the drier southern margin of the steppes, largely in Kazakh Soviet Socialist Republic, where annual rainfall decreases to 250 mm (Grigg, 1974). This expansion was accomplished mainly on highly mechanized and very large state farms, which averaged more than 60 000 ha each (Hooson, 1966). Since the collapse of the Soviet Union, the whole system has fallen into disarray, but to date there appear to be very few individual family farms.

In Australia, the establishment of wheat farms at first pushed sheep onto the dry margins. Since the 1930s, however, sheep have been integrated with wheat production. Clover and other forage crops, grown in rotation with wheat, feed the lambs and fix nitrogen in the soil, while the sheep add manure fertilizer to the fields. In addition, trace elements such as copper and molybdenum, originally generally deficient in the grain belt soils, have been added to the soil along with nitrogen and superphosphates, leading to increased yields (Bambrick, 1994).

Wheat is still the dominant crop in all the main areas of extensive cereal cultivation. In the more moist areas, however, cropping has become more diversified, with extensive sowing of barley, canola (rapeseed), sorghum, alfalfa, and other less significant crops. The increase in livestock numbers means that some areas, notably Central and Eastern Ukraine, are now more aptly described as mixed farming regions than as extensive cereal cropping districts.

In all areas of extensive cereal cultivation, crop yields, here exemplified by wheat, have risen substantially in recent years, except in the former Soviet Union (Table 1). This increase has occurred for various reasons, but the increasing use of manufactured fertilizers is a main cause.

EXTENSIVE CEREAL CULTIVATION AND THE ENVIRONMENT

Although extensive cereal cultivation is of recent origin, it has affected the environment in many ways. Chief among these, besides the conversion of tens of millions of ha of natural grassland to crops, are loss of soil organic matter, soil erosion by wind and water, spread of soil salinity, and the planting of trees and shrubs around farmsteads and on fields. The possible effects of anthropogenic alterations of the climate on cereal cultivation will also be examined.

Soil Organic Matter

Under natural conditions, soil organic matter in the temperate grasslands was high, varying from about 3 5% in the drier districts to as high as 10% in the moister areas. However, in Australia, where eucalyptus, shrubs, and grass were the natural vegetation in the areas of extensive cereal cultivation, organic matter levels were in places as low as 1 2%. Converting the grassland to cereal crops led to a decline in organic matter because part of the biomass was removed with the harvest or burned, cultivation accelerated the rate of destruction of organic matter, and there was loss of organic matter through erosion. The practice of summer fallowing further contributed to the loss of organic matter because no crop was sown in the fallow year, hence reducing primary biomass production. Frequent tillage to control weeds during the fallow increased the rates of decomposition of organic matter and soil erosion, and some of the

Table 1 Average yields of wheat, 1974–1998

	1974–1978 (kg ha^{-1})	1979–1983 (kg ha^{-1})	1984–1988 (kg ha^{-1})	1989–1993 (kg ha^{-1})	1994–1998 (kg ha^{-1})
Canada	1869	1917	1755	2136	2239
US	2026	2384	2452	2478	2592
Argentina	1549	1670	1860	2032	2169
Australia	1366	1240	1484	1675	1764
Soviet Union	1502	1528	1705[a]		
Russian Federation	Main countries of the former Soviet Union where there is extensive cereal cultivation			1789[b]	1363[c]
Ukraine				3403[b]	2756[c]
Kazakhstan				925[b]	738[c]

[a] Average for the years 1984–1989.
[b] Average for the years 1990–1994.
[c] Average for the years 1995–1998.
Source: FAO production yearbooks.

Table 2 Tillage practices on cropland in the Canadian Prairies, 1991 and 1996

	Conventional tillage: tillage incorporating most crop residue into the soil (% of total cropland)	Conservation tillage: tillage retaining most crop residue on the soil surface (% of total cropland)	No tillage: no tillage prior to seeding (% of total cropland)
1991	67	26	7
1996	52	32	16

Sources: Statistics Canada, 1991 and Agricultural Censuses 1996.

moisture accumulated during the fallow tended to leach soluble nitrate nitrogen below the rooting zone (Rennie, 1989).

The loss of soil organic matter is considered to be potentially, if not already, serious for several reasons. Organic matter contains many of the nutrients needed for cereal crop growth, notably nitrogen. It improves soil aggregation and tilth, provides a rooting zone for crops, and is important for moisture-holding capacity.

How much soil organic matter has declined is, however, debatable. It is generally agreed that there were rapid losses for about 20 years after initial clearing, then a leveling off to a slower decline. In the Canadian Prairies, there was much concern in the 1970s and 1980s when estimates of soil organic loss since clearing ranged from 50 70% (Gregorich *et al.*, 1995; Rennie and Ellis, undated). These estimates appear to be based on a cropping sequence in which half the arable land was left in summer fallow each year i.e., a grain crop one year followed by fallow the next. A more recent study, in contrast, places the losses of soil organic matter in the order of 15 30% (Gregorich *et al.*, 1995). This study bases its estimate on losses of organic matter since initial cultivation on uneroded soils which, in the Prairies, means mainly soils which have been continuously cropped since clearing, that is, they have not been summer fallowed (E Gregorich, personal communication, February 7, 2000). This estimate appears to be low because until about 1970, summer fallow was practiced throughout the Prairies (Carlyle, 1997).

Whatever the actual losses are, it is apparent that several measures are now reducing the rate of loss of soil organic matter in the Prairies. Perhaps most importantly, between 1971 and 1996, summer fallow decreased in the region from 10.6 6.2 million ha (Carlyle, 1997). Also, even where summer fallow is still practiced, tillage operations to kill weeds during the fallow period have been reduced or eliminated altogether over large areas by using herbicides. Thus, in 1996, herbicides together with less tillage than on conventional fallow were used to control weeds on 37% of fallowed land, and, on 9% of the fallow, tillage was eliminated altogether by using only herbicides to control weeds (Statistics Canada, 1998). Reduced tillage on fallow reduces the rate of loss of soil organic matter by slowing down microbial action and erosion (Janzen *et al.*, 1998). In addition, crop yields have been increasing, partly through the use of manufactured fertilizers, while the decay of a greater biomass increases total organic matter (Table 1). Furthermore, farmers in the region are increasingly adopting reduced or no tillage on the cropped land, practices which leave more plant debris on the soil, thereby reducing erosion and adding to the pool of organic matter compared to traditional tillage methods (Table 2, Figures 3 and 4).

Higher yields, reduced tillage, and a decline in summerfallow have also taken place on the US Great Plains, which should stabilize or increase soil organic matter there (Table 1). A Mediterranean climate with cool, wet winters and hot, dry summers characterizes the extensive cereal zones of Western Australia and the southern part of Southeastern Australia. In these zones, cereals are sown in the autumn (May) to take advantage of the winter rainfall, and are harvested in late spring (November), making the practice of fallowing for additional moisture unnecessary or ineffectual because of very high evaporation rates during the hot summers. Fallow is, however, practiced in the northern districts of Southeastern Australia, where most rainfall

occurs during the summer. Clovers, sheep, and manufactured fertilizers have helped raise cereal crop yields and soil organic matter. In addition, farmers have been rapidly adopting reduced tillage and no tillage, which not only reduce loss of soil organic matter, but also conserve moisture. In Western Australia, for example, it is estimated that each cultivation may cause 2 cm of extra evaporation. In the past five years, up to 400 t ha^{-1} of clay have been added to sandy, water repellent, soils in the extensive cereal districts, which increases yields and hence soil organic matter because of increased soil moisture retention. These measures have resulted, in places, in an increase of soil organic matter since initial cultivation (Crabtree, personal communication, January, 2000).

The loss of soil organic matter in the Pampas appears to be less than in other areas of extensive cereal cultivation. Wheat is generally sown for several years, followed by alfalfa (lucerne) for five or more years, which protects and builds up organic matter in the soil. The author could find no current information on soil organic matter levels in extensive cereal districts of the former Soviet Union. It is, though, apparent that yields of wheat have risen little in recent years, which suggests that methods of tillage, fallowing, and fertilizer use have changed little, at least in the main wheat belt east of the Black Sea in the Russian Federation and Kazakhstan (Table 1). It would therefore seem that there have been few corrective measures to curtail losses of soil organic matter.

Soil Erosion

Estimates of the causes, amounts, and costs of soil erosion in areas of extensive cereal cultivation are even more divergent than those for losses of soil organic matter. For example, soil drifting and dust storms in the Great Plains and Prairies during the 1930s (Figure 5) have commonly

Figure 5 Soil erosion and drifting near Cadillac, Saskatchewan, in the 1930s. (Reproduced with permission of the Prairie Farm Rehabilitation Administration, Regina, Saskatchewan, file photograph 32227)

been ascribed to inappropriate farming techniques exacerbated by drought (Herrington *et al.*, 1997). Malin (1956), however, points out that severe dust storms occurred in the Plains in 1830, long before the land was cultivated, and he states that No more brazen falsehood was ever perpetrated upon a gullible public than the allegation that the dust storms of the 1930s were *caused* by the plow that broke the Plains (Italics Malin s).

More recently, several studies estimated the costs of soil erosion in the Canadian Prairies. These estimates varied from $369 million (Cdn.) per year (Sparrow, 1984) to $430 million (Cdn.) annually (Smith, 1986). Smith (1986) goes further in giving a figure of $1.011 billion annually for losses on the Prairies from all types of soil degradation, with organic matter losses and salinity costing $314 million and $212 million, respectively. These figures refer only to the on-farm costs of degradation; they do not include off-farm damages caused by soil from farms such as silting of reservoirs, damages to peoples health, and infilling of ditches.

In sharp contrast, in studies of soil erosion in the US Great Plains, where conditions are similar to the Prairies, soil erosion was not a major factor.

Alt and Putnam (1989) estimated the amount of wind and water erosion in the Great Plains by focusing on areas where soil erosion exceeded soil formation. They estimated that after 100 years of erosion at 1982 rates, there would be, in the 100th year, a loss of only 1.4% in crop yields in the Northern Plains and 2.6% in the Southern Plains. Moreover, they estimated that about 75% of the cropland in the Great Plains would, in the 100th year, lose less than 2% of its yield from erosion, and 90 95% would lose less than 8%. Crosson and Stout (1983) estimated that losses in mean yields of wheat due to water erosion in the Great Plains for the years 1950 1980 were only 1%. More generally, Crosson and Rosenberg (1989) claimed that The US is the only country in the world with reasonably accurate and comprehensive estimates of soil erosion and its effects .

Recent studies of soil erosion, and soil degradation generally, on the Canadian Prairies suggest that earlier figures were overestimates (Acton and Gregorich, 1995). Wall *et al.* (1995) claim that, in 1991, less than 5% of the cultivated land was at a high to severe risk of wind erosion, and that some 95% of the cultivated land was at a tolerable risk of water erosion. Furthermore, they estimate that the risk of wind and water erosion declined on the Prairies between 1981 and 1991, mainly because of increasing use of reduced tillage and no tillage, and also because of a reduction in summerfallow, a cropping practice (Figure 6). They also state that it is to be expected that the risk of wind erosion had been further reduced by 5 10% between 1991 and 1995. Lerohl and van Kooten (1995) conclude that most of the 1980s soil erosion figures in the Prairies were overestimates, often grossly so, and that soil erosion of farmland did not cause much damage in the region, either on or off the farm.

The disagreements about the severity of erosion in the Prairies and Great Plains, and probably in areas of extensive cereal cultivation elsewhere, appear to arise for several reasons. The high estimates are generally based on the assumption, usually unstated, that the eroded soil is lost forever to agriculture (Lerohl, 1991). They seem, furthermore, to extrapolate dry conditions, such as those of the 1930s and 1980s, into the future. This is of particular importance because wind erosion is the main type of soil erosion in the Prairies, and it appears to be more damaging in dry years (Wheaton and Wittrock, 1991).

The lower estimates consider such factors as rates of soil formation, and the re-deposition of a considerable proportion of the soil moved by wind in particular, but also by water, on farmland elsewhere in the region (Crosson and Stout, 1983; Lerohl and van Kooten, 1995).

Although the estimates of soil erosion vary considerably, most writers urge continuing assessment of soil erosion and remedial measures to reduce it. Although the net effects of soil erosion over large areas may not be severe, individual farmers have much to lose from local erosion. Government agencies are therefore urged to switch from blanket regional measures to actions targeting specific sites and soil types where erosion is above acceptable levels.

Soil Salinity

Soil salinity is a problem in the main areas of extensive cereal cultivation because of the climate and farming practices. In situations where there is water surplus to crop needs, water soluble salts, mainly sulfates, chlorides, carbonates, and bicarbonates of calcium, magnesium, and sodium, move downwards to the water table. These salts can move downslope to discharge areas, and if evaporation is in excess of precipitation (a condition that is common in areas of extensive cereal cultivation), the salts are brought to the surface by capillary action. High levels of salinity have the same effect as drought by making water less available for uptake by plant roots and, if the soil is extremely saline, it will cause water to be drawn out of the plant. As a result, crop yields are lowered or the land is rendered unsuitable for cropping altogether.

Salinity occurs naturally in areas of extensive cereal cultivation (*primary salinity*), but it has become more widespread and severe since the clearing of the grasslands (*secondary salinity*). One reason is that annual cereal crops take up less water than perennial grasslands, and hence more water percolates below the rooting zone than it did before clearing. This raises the water table. The higher water table results in more water being brought to the surface in discharge areas, where it evaporates leaving the salts in the soil (Sparrow, 1984). In areas where it is frequently used, summer fallow appears to be the main cause of the spread of soil salinity. Only about 25% of the moisture accumulated during the fallow is stored in the soil; the remainder either runs off, evaporates, or enters the water table. The latter contributes to an increase in soil salinity. Even without fallowing, sandy soils, especially sandy water repellent soils, cause salinity to spread because of their low water-holding capacity.

As for losses of soil organic matter and soil erosion, estimates of the severity and costs of human-induced (secondary) soil salinity vary. One reason is that most or all of soil salinity has sometimes been attributed to human actions, that is, the role of primary salinity is not given much, if any, consideration. Another is that projections of soil degradation by secondary salinity have been based on a few plot-scale experiments rather than on widespread studies. Examination of soil salinity in the Canadian Prairies illustrates these comments.

One study estimated that 2.3 million ha in the Prairies were affected by soil salinity in the early 1980s, although there was no breakdown for secondary as opposed to primary salinity (Sparrow, 1984). This study claimed that summer fallow was the main cause of secondary salinity, and that secondary salinity was being extended at a rate of 10% per year. The annual costs due to decreases in crop production because of soil salinity were estimated to be at least $260 million (Cdn.). Another study, also published in 1984, focused on secondary salinity (Anderson and Knapik, 1984). It also pointed to summer fallow as the main cause of secondary salinity, but claimed that the often used 10% per year increase in such salinity was an overestimate, and used instead of a 1% per year figure. The cost of secondary salinity because of decreased crop production on the estimated one million ha of land affected in 1984 was estimated to be $105 million (Cdn.), with the annual cost rising to $128 million by 2008. A third study done in the 1980s claimed that secondary salinity affected 2.2 million ha of land in the Prairies. It estimated that secondary salinity was spreading at a rate of 10% per year, and cost an increased loss of $25.8 million (Cdn.) annually from lost crop production, which would amount to $4.42 billion (Cdn.) over the period 1983 2000. As with the other studies, summer fallow was deemed to be the main cause of secondary salinity (PFRA, 1982). Yet another study placed the annual cost of soil salinity, apparently both primary and secondary, at $212 million (Cdn.), with fallow again being the main cause of secondary salinity (Smith, 1986).

A recent study of soil salinity in the Canadian Prairies claims to have made the first attempt to describe salinity in the region in a standard way and to provide a baseline for further measurements (Eilers *et al.*, 1995). This study provides maps and tables showing the amount of farmland in the Prairies affected by various degrees of salinity; low (less than 1% of the land affected), moderate (1 15% of the land affected), and high (more than 15% of the land affected). For the Prairies as a whole, 62% of the total farmland was found to have a low degree of salinity, 36% had a moderate degree, and only 2% had a high degree. Moreover, most farmland in the region was assessed to be at little to no risk of becoming more saline under cropping practices used in 1991, and it was stated that the decrease of summer fallow in recent years had decreased the risk of secondary salinity from this practice.

Experimental plots have shown that leaving strips of tall stubble (grain stalks) at harvest traps more snow than short stubble, and leads to a more even distribution of snow throughout fields. This reduces ponding of water when the snow melts, and reduces the risk of salinization. Furthermore, the additional snowmelt has been shown to increase crop yields. Widespread adoption of tall stubble strips would therefore be beneficial in areas of extensive cereal cultivation in the Northern Hemisphere where snow is a normal winter occurrence.

Soil salinity appears not to be a major problem in the Argentine Pampas, mainly because the alfalfa sown in rotation with wheat allows very little water to seep downwards into the water table (Eilers *et al.*, 1995). It does affect cropland in the US Great Plains, but should be less than it was mainly because of a decline in summer fallow. Salinity is a major problem in areas of sandy water repellent soils in the extensive cereal districts of Australia, and it is expected to

become worse (Crabtree, personal communication, January, 2000). Mitigating measures include adding clay to the soil to reduce seepage into the water table. No current data could be found regarding soil salinity in the former Soviet Union, but it has been a problem in the past, especially in the area east of the Black Sea (*see* Salinity and Agriculture, Volume 3).

Planting of Trees

There were trees mixed with grass in parts of the temperate grasslands before cultivation, for example in the northern part of the Canadian Prairies (parkland), the Eurasian steppes (wooded steppe), and in Australia (eucalyptus, shrubs, and grass). Over most of the pre-cultivation landscape, however, trees were found only along the water courses, leaving vast expanses of treeless grassland. As cropland farmers entered the grasslands in the late 1800s, tree plantings began and have continued to the present day. These plantings have altered the landscape and microclimate of the areas of extensive cereal cultivation, and hence are part of the process of environmental change. However, they are not, strictly speaking, related to extensive cereal

Box 1 Shelterbelts

SHELTERBELTS AND CLIMATE CHANGE

Proposals for tree plantings in the US Great Plains and Canadian Prairies began during the pioneer settlement period of the late 1800s. In the Great Plains, the eminent geologist, F V Hadyn, suggested that large-scale tree plantings should be undertaken because such plantings would lead to a moister climate and improve soil fertility (Rees, 1988). Dean Bessey of the University of Nebraska made similar claims, stating that the Plains were dry because they were treeless (Rees, 1988). In Canada, Professor William Brown suggested that trees be planted in the Prairies because ''A peopled agricultural country without trees is an impossibility'' (Brown, 1884, cited by Rees, 1988). In the late 1800s, it was believed that trees would ameliorate the climate of the Ukrainian steppe grasslands (Rees, 1988).

This concept was revived in the US and in the Soviet Union during the 1900s. During the 1930s, a period of extreme drought in the Great Plains, some 30 000 km of tree belts were planted along the 100th meridian, but with little long-term success because of low survival rates (Burke, 1956). It is, however, in the Soviet Union that the most ambitious plan to modify climate by planting trees was undertaken as part of a wider project known as Stalin's Plan for Reforming Nature (Alisov *et al.*, 1952, cited by Thornthwaite, 1956).

The afforestation part of Stalin's plan, announced in 1948, involved the planting of six million ha of forest strips to protect 120 million ha of farmland on the steppes north of the Black and Caspian Seas (Burke, 1956). These strips, it was claimed, would reduce the loss of snow from farmland by curtailing Arctic winds from Siberia (burani) in winter. In summer, evaporation would be reduced by protecting the cropland from the full force of hot, dry winds (suxovei) from the arid lands of South Central Russia (Burke, 1956). Mitigating the effects of both these winds would, it was asserted, result in increased soil moisture and higher grain yields. Moreover, the Soviets thought that not just the microclimate but the macroclimate of the region could be modified by afforestation. A progress report in 1951 stated that two million ha had been planted, although no mention was made of survival rates (Burke, 1956).

Rees (1988), Thornthwaite (1956), and Burke (1956) all were very skeptical of the efficacy of these plantings in changing the climate for the better. As Rees (1988) put it ''Trees do reduce rates of evaporation, and so conserve moisture, but the plains were dry not because they were treeless but treeless because they were dry.'' Thornthwaite (1956) was confident that the Soviet plan would effect only trivial changes in climate. Burke (1956) claimed that the consensus was ''that such speculation (changing the climate by planting trees) is largely nonsense.''

SHELTERBELTS FOR BEAUTIFICATION AND EROSION CONTROL

Trees were, however, planted in the areas of extensive cereal cultivation for more practical reasons. Shelterbelts around farmsteads, especially if they contained flowering species, beautified surroundings in summer and, if they included coniferous species, did so in winter by adding green to the landscape (Rees, 1988). Such belts also shielded the farmstead from the sun and wind, shut out the endless space of the farmland beyond, and, when mature, could be used for firewood.

The planting of farm shelterbelts took on an almost religious fervor in the United States after the first Arbor Day, a day set aside and dedicated to tree plantings, was declared in Nebraska in 1872. Arbor Day subsequently spread throughout the US and into Canada.

The planting of field shelterbelts to reduce soil erosion by wind began in both the Great Plains and Prairies during the dry 1930s. In Canada, a federal government agency, the Prairie Farm Rehabilitation Administration (PFRA), undertook the distribution of seedlings to farmers. Several areas of farmland highly susceptible to wind erosion were targeted for intensive plantings (Wall *et al.*, 1995). One such project was near Conquest, Saskatchewan (Figure 7). Similar programs have been undertaken in the areas of extensive cereal cultivation in Australia.

Although not known at the time when most farm and field shelterbelts were planted, it is now recognized that they store significant amounts of carbon, and hence reduce emissions of carbon dioxide (CO_2) into the atmosphere (Janzen *et al.*, 1998; Johnston *et al.*, 2000). For example, the PFRA estimates that typical shelterbelt trees contain 162–544 kg C, and that shelterbelt shrubs contain up to 52 tones C km^{-1}.

Box 1 (*continued*)

Figure 7 Field and farm shelterbelts near Conquest, Saskatchewan. (Reproduced by permission of the Prairie Farm Rehabilitation Administration Shelterbelt Centre, Indian Head, Saskatchewan, CAI–76–210A)

cultivation, and therefore they will be separately examined (see Box 1 on shelterbelts).

Climate Change

It is generally agreed that the global atmospheric concentration of greenhouse gases has increased because of human activities. The viewpoint taken here is that it is probable that human activities have contributed and will continue to contribute to climate change, although the precise contribution of humans is not now known.

Whatever the impact of human activities on climate change, it is certain that activities associated with extensive cereal cultivation have not had and will not have a signi cant impact on climate change. Climate change would, however, have impacts on this type of agriculture.

It is generally agreed that global warming has occurred in the main areas of extensive cereal cultivation, and that this trend will continue, and that the greatest warming will occur in the Northern Hemisphere. For example, a recent study estimated that the average annual mean temperature in the Canadian Prairies had increased by 0•9 °C between 1895 and 1996, and that all three global circulation models predicted further increases in mean temperatures in all seasons (Herrington *et al.*, 1997).

If this warming trend continues, crop types and yields could change, depending on the amounts of precipitation and available soil moisture. A decrease in precipitation, which some models predict, would lead to decreased crop yields and a decrease in soil organic matter because of less primary biomass and because soil organic matter varies inversely with temperature (Breymeyer *et al.*, 1996). On the other hand, conservation tillage methods and reduction in summer fallow, trends apparent in many areas of extensive cereal cultivation, could counteract these losses of soil organic matter (Janzen *et al.*, 1998; Breymeyer *et al.*, 1996). Similar results could occur if precipitation increased insuf ciently to offset the higher evaporation rates caused by higher temperatures.

If, however, precipitation increased to the degree that available soil moisture also increased, crop yields would go up, and it might be possible to grow more heat- and moisture-demanding crops such as maize. It should be noted, too, that an increase of carbon dioxide in the atmosphere, providing there is suf cient water, would itself increase yields because carbon dioxide promotes plant growth.

It could well happen that different climatic changes would occur in different parts of even a single extensive cereal zone. This has, in fact, been predicted for different parts of the Canadian Prairies.

McGinn *et al.* (1999) used two climate change scenarios developed from the Canadian Climate Change Global Circulation Model to predict changes in crop yields in Alberta. One scenario involved warming and increased soil moisture, and it resulted in substantial increases in yield; 34% for wheat and 47% for canola. The other scenario involved warming only, which would mean a decrease in overall soil moisture. But, even so, yields were predicted to increase probably because earlier spring seeding would allow more ef cient use of spring precipitation, and more rapid maturity because of higher temperatures would decrease the total amount of water required for plant growth.

In contrast, Raddatz and Shaykewich (1997) predicted a decrease in crop yields on unirrigated land in the eastern part of the Canadian Prairies because of decreased overall evapotranspiration, which would occur because warmer summer temperatures would hasten crop maturity.

CONCLUSIONS

Extensive cereal cultivation has contributed to global environmental change and will continue to do so in the future. It has caused losses in soil organic matter, soil erosion, and the spread of soil salinity. Changes in cropping and cultivation practices, notably a decline in summer fallow and reduced tillage, have reduced soil degradation from these sources. Climate change has taken place in these areas and is expected to continue. Much uncertainty surrounds the effects of climate change on the yields and types of crops. Shelterbelts have been planted in areas of extensive cereal cultivation, and they locally reduce wind erosion and alter the climate, as well as beautify surroundings.

REFERENCES

Acton, D F and Gregorich, L J (1995) *The Health of our Soils*, Centre for Land and Biological Resources Branch, Agriculture and Agri-Food Canada, Publication 1906/E, Ottawa.

Alisov, B P, Drosdov, O A, and Rubenstein, E S (1952) *Kurs klimatologii* (Course in Climatology), Publishing Institute of Hydrometry: Leningrad, Cited in Thornthwaite, C W (1956) Modification of Rural Microclimates, in *Man's Role in Changing the Face of the Earth*, ed W L Thomas, University of Chicago Press, Chicago, IL, 567 583.

Alt, K and Putnam, J (1987) Soil Erosion: Dramatic in places, but not a Serious Threat to Productivity, *Agric. Outlook*, **129**(April), 28 30.

Anderson, M and Knapik, L (1984) *Agricultural Land Degradation in Western Canada: a Physical and Economic Overview*, Regional Development Branch, Agriculture Canada.

Bambrick, S (1994) *The Cambridge Encyclopedia of Australia*, Cambridge University Press, Cambridge.

Breymeyer, A I, Hall, D O, Melillo, J M, and Agren, G I (1996) *Global Change: Effects on Coniferous Forests and Grasslands*, John Wiley & Sons, Chichester, Scope 56.

Brown, W (1884) The Application of Scientific and Practical Arboriculture to Agriculture, *Canadian Economics*, British Association for the Advancement of Science, Montreal Meeting, Cited by Rees, R (1988) *New and naked land*, Western Producer Prairie Books, Saskatoon.

Burke, A E (1956) Influence of Man Upon Nature; The Russian View: a Case Study, in *Man's Role in Changing the Face of the Earth*, ed W L Thomas, University of Chicago Press, Chicago, IL, 1035 1051.

Carlyle, W J (1997) The Decline of Summerfallow in the Canadian Prairies, *Can. Geogr.*, **XLI**(3), 267 280.

Crosson, P R and Rosenberg, N J (1989) Strategies for Agriculture, *Sci. Am.*, **261**(3), 128 135.

Crosson, P R and Stout, A (1983) *Productivity Effects of Cropland Erosion in the United States*, Resources for the Future, Washington, DC.

Eilers, R G, Eilers, W D, Pettapiece, W W, and Lelyk, G (1995) Salinization of Soil, in *The Health of Our Soils*, eds D F Acton and L J Gregorich, Centre for Land and Biological Resources Branch, Agriculture and Agri-Food Canada, Publication 1906/E, Ottawa, 77 86.

Gregorich, E G, Angers, D A, Campbell, C A, Carter, M R, Drury, C F, Ellert, B H, Groenvelt, P H, Holstrom, D A, Monreal, C M, Rees, H W, Voroney, R P, and Vyn, T J (1995) Changes in Soil Organic Matter, in *The Health of Our Soils*, eds D F Acton and L J Gregorich, Centre for Land and Biological Resources Branch, Agriculture and Agri-Food Canada, Publication 1906/E, Ottawa, 41 50.

Grigg, D B (1974) *The Agricultural Systems of the World: an Evolutionary Approach*, Cambridge University Press, Cambridge.

Herrington, R, Johnson, B, and Hunter, F (1997) *Responding to Global Climate Change in the Prairies, Volume III of the Canada Country Study*, Climate Impacts and Adaptation, Adaptation and Impacts Section, Atmospheric Environment Branch, Environment Canada, Prairie and Northern Region.

Hooson, J D (1966) *The Soviet Union*, A Systematic Regional Geography, University of London Press, London, Vol. 7.

James, P (1951) *A Geography of Man*, Ginn and Company, Boston, MA.

Janzen, H H, Desjardins, R L, Asselin, J M R, and Grace, B (1998) *The Health of Our Air: Towards Sustainable Agriculture in Canada*, Minister of Public Works and Government Services Canada, Publication 1981/E, Ottawa.

Johnston, M, Kulshreshtha, S, and Baumgartner, A (2000) Agroforestry in the Prairie Landscape: Opportunities for Climate Change Mitigation through Carbon Sequestration, *Prairie Forum*, **25**(2), 195 213.

Lerohl, M (1991) The Sustainability of Selected Prairie Crop Rotations, *Can. J. Agric. Econ.*, **39**(4I), 667 676.

Lerohl, M L and van Kooten, G C (1995) Is Soil Erosion a Problem on the Canadian Prairies?, *Prairie Forum*, **20**(1), 107 121.

Malin, J C (1956) The Grassland of North America: Its Occupance and the Challenge of Continuous Reappraisals, in *Man's Role in Changing the Face of the Earth*, ed W L Thomas, University of Chicago Press, Chicago, IL, 350 366.

McGinn, S M, Toure, A, Akenrimi, O O, Major, D J, and Barr, A G (1999) Agroclimate and Crop Response to Climate Change in Alberta, Canada, *Outlook Agric.*, **28**(1), 19 28.

Prairie Farm Rehabilitation Administration (1982) *Land Degradation and Soil Conservation Issues on the Canadian Prairies: An Overview*, PFRA, Regina.

Raddatz, R L and Shaykewich, C F (1997) *Impact of a Warming Climate on Evapotranspiration on the Eastern Prairies*, Canadian Meteorological and Oceanographic Society Congress, June 1 5, Saskatoon.

Rees, R (1988) *New and Naked Land*, Western Producer Prairie Books, Saskatoon, Saskatchewan.

Rennie, D A (1989) Managing Soil Conservation, *Agriscience*, December, 8 9.

Rennie, D A and Ellis, J G (undated) *The Shape of Saskatchewan*, Saskatchewan Institute of Pedology, Publication M 41, University of Saskatchewan, Saskatoon.

Rudolph, J D (1985) *Argentina: a Country Study*, US Government Printing Office, Washington, DC.

Smith, S L (1986) *A Growing Concern: Soil Degradation in Canada*, Science Council of Canada, Ottawa.

Sparrow, H (1984) *Soil at Risk: Canada's Eroding Future*, Report of the Standing Committee on Agriculture, Fisheries and Forestry, Senate of Canada, Ottawa.

Statistics Canada (1998) *1996 Census of Agriculture: Selected Data for Saskatchewan Rural Municipalities*, Census of Agriculture, Ottawa.

Thornthwaite, C W (1956) Modification of Rural Microclimates, in *Man's Role in Changing the Face of the Earth*, ed W L Thomas, University of Chicago Press, Chicago, IL, 567 583.

Wall, G J, Pringle, E A, Padbury, G A, Rees, H W, Tajek, J, van Vliet, L J P, Stushnoff, C T, Eilers, R G, and Cossette, J-M (1995) Erosion, in *The Health of Our Soils*, eds D F Acton and L J Gregorich, Centre for Land and Biological Resources Branch, Agriculture and Agri-Food Canada, Publication 1906/E, Ottawa, 61 76.

Wheaton, E E and Wittrock, V (1991) Climatic Change and Wind Erosion by Dust Storms, in *Symposium on the Impacts of Climate Change and Variability on the Great Plains*, ed G Wall, Department of Geography Publication Series, Occasional Paper No. 12, University of Waterloo, Waterloo, 333 335.

Changes in World Marine Fish Stocks

Ian Douglas
University of Manchester, Manchester, UK

An analysis by the Food and Agriculture Organization (FAO) of the United Nations, based on sh harvest records from 1950 to 1994, found that 35% of the most important commercial sh stocks showed a pattern of declining yields. Another 25% showed steady yields but were being shed at their biological limit and are vulnerable to declines if sh- ing levels increase (United Nations Food and Agriculture Organization (FAO), 1997a). The harvest of overexploited sh stocks fell 40%, from 14 million metric tons in 1985 to eight million metric tons in 1994. These numbers masked more precipitous drops in certain sh stocks such as Atlantic cod, haddock, and red sh, which had all but disappeared in some areas of the North Atlantic (Figure 1).

Some 80 million metric tons of sh are presently available for direct human consumption annually. The FAO expects demand to increase to 110 to 120 million metric tons by 2010 as the world population increases. Such a demand could be satis ed only if aquaculture production doubled and over- shing was brought under control so that ocean sh stocks could recover. However, aquaculture growth is likely to be moderate and the ocean catch will probably remain steady at current levels or even decline, leaving a marked gap between supply and demand, and also raising sh prices.

The process of development of a fishery as described by changes in landings with time is schematically represented in Figure 2 as comprising four phases: (I) undeveloped, (II) developing, (III) mature, and (IV) senescent. The phases are related to the relative rate of increase in yield of fish as the fishery grows. The relative rate of increase varies significantly as the maximum long-term yield is approached, reached and overshot. This idea can be used to

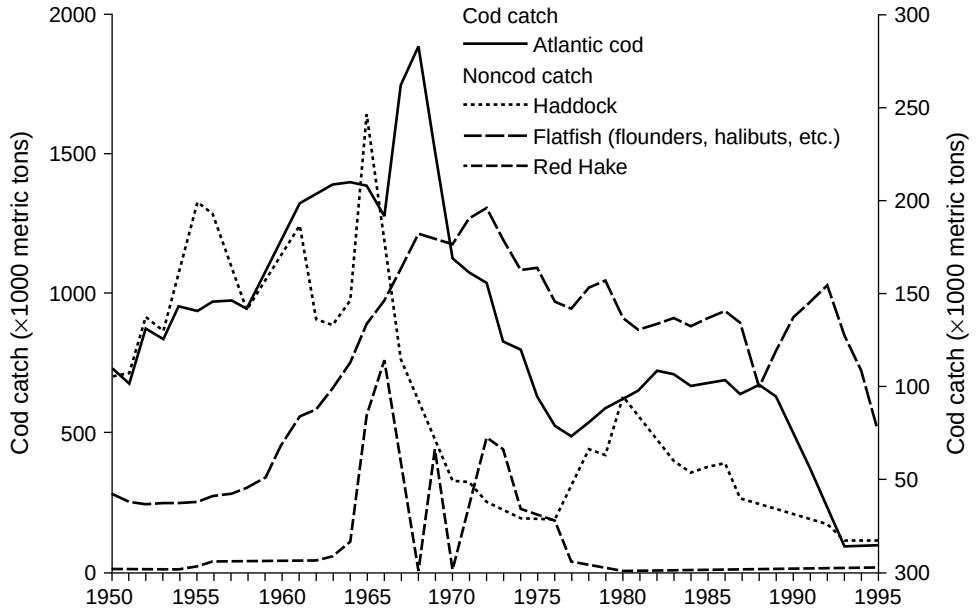


Figure 1 Commercial harvests in the Northwest Atlantic of some important fish stocks, 1950–1995 (*Source:* FAO, 1997c)

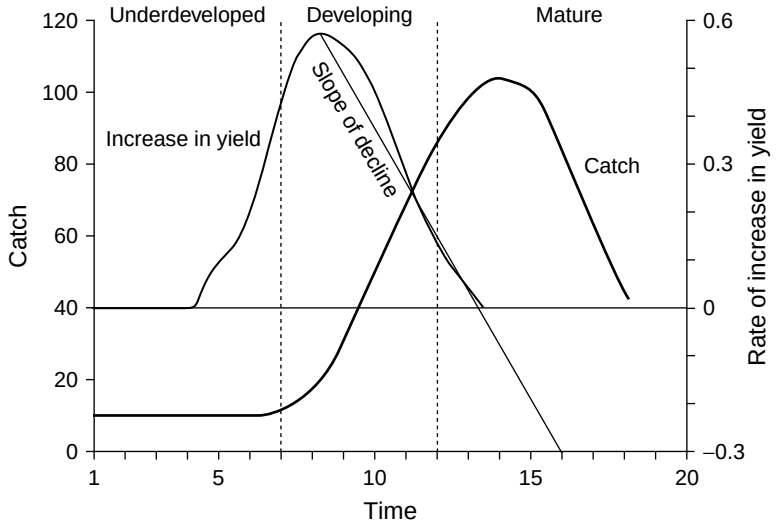


Figure 2 Theoretical development and decline of fisheries showing the derivation of the slope of decline used to predict future trends in world fish stocks (*After:* FAO, 1997c)

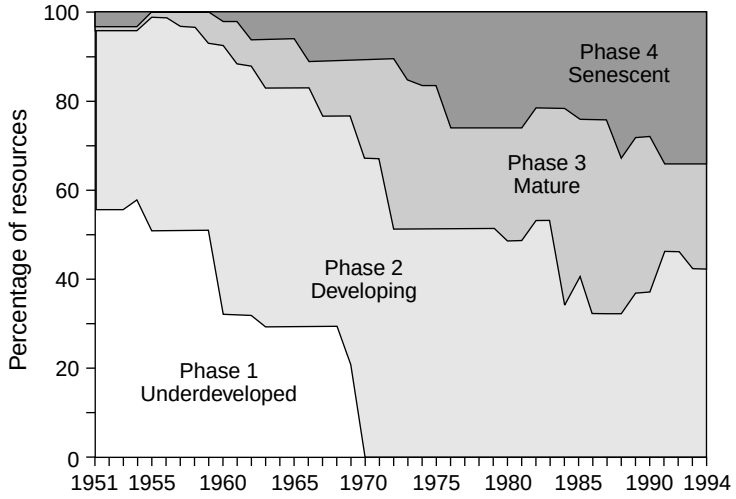


Figure 3 Percentage of the world's fisheries in differing stages of development 1951–1994 (*Source:* FAO, 1997c)

provide a rough assessment of the state of marine resources, globally and by ocean. The relative rate increase is nil in a stable non-developing shery (phase I), but increases rapidly (phases I II) as the shery starts to develop. During the phase of steady growth of the shery, it decreases (phase II) and drops to zero again when the shery reaches its maximum production (phase III). This theoretical pattern is masked by natural uctuations as well as by variability in the accuracy of sheries data. However, the way in which the relative rate of increase declines during phases II and III has been used by the FAO to estimate when full potential has been reached and what the corresponding potential multi-species yield could be.

By 1994 available data indicate that about 35% of the 200 major shery resources had become senescent (i.e., were showing declining yields), about 25% were mature (i.e., were leveling off at a high exploitation level), 40% were still developing, and none remained at a low exploitation (undeveloped) level (Figure 3). Earlier the FAO had concluded that 44% of the stocks for which formal assessments were available, were intensively to fully exploited, 16% were over- shed, 6% depleted, and 3% slowly recovering. These investigations indicated that 69% of the known stocks urgently needed strict management. A global production model used in this investigation indicated that the deepwater high value species were over- shed and that the global shing effort would have to be cut to rebuild the resources.

The decline in shery yields spread from the Atlantic to the Paci c and Indian Oceans as international shing eets went farther and farther from their home ports in the

Table 1 Progression of over-fishing from ocean to ocean, peak harvest year of high value (deepwater) fish by region

Fishing area	Year of maximum harvest	Maximum harvest (000 metric tons)	Recent harvest (000 metric tons)
Atlantic, northwest	1967	2588	1007
Antarctic	1971	189	28
Atlantic, southeast	1972	962	312
Atlantic, western central	1974	181	162
Atlantic, eastern central	1974	481	320
Pacific, eastern central	1975	93	76
Atlantic, northeast	1976	5745	4575
Pacific, northwest	1987	6940	5661
Pacific, northeast	1988	2556	2337
Atlantic, southwest	1989	1000	967
Pacific, southwest	1990	498	498
Pacific, southeast	1990	508	459
Mediterranean	1991	284	284
Indian Ocean, western	1991	822	822
Indian Ocean, eastern	1991	379	379
Pacific, western central	1991	833	833

search for good harvests (Table 1). In the western Atlantic, the breeding population of bluefin tuna, the largest tuna species, fell from around 250 000 adults in 1975 to around 40 000 in 1995. According to the US National Fish and Wildlife Foundation, 14 major commercial fish species in US national waters (accounting for one-fifth of the world s stocks and half of all US stocks) are so depleted that even if all fishing stopped immediately, it would take up to 20 years for stocks to recover.

Since 1994, efforts have continued to obtain international agreements to regulate exploitation of the global fisheries, especially in the case of fish stocks that straddle jurisdictions by migrating out of territorial waters into the high seas. Straddling fish stocks and highly migratory fish stocks have been vulnerable to depletion on the high seas in many regions of the globe:

1. pollock in the Donut Hole of the Bering Sea and the Peanut Hole of the Sea of Okhotsk;
2. orange roughy on the Challenger Plateau in the high seas off the coast of New Zealand;
3. hake, southern blue whiting and squid off Argentina s Patagonian Shelf;
4. jack mackerel off the coast of Chile and Peru; cod in the Barents Sea Loop Hole off the coast of Norway;
5. highly migratory stocks of tuna, dolphin, and shark in the South Pacific Ocean, beyond the limits of national jurisdiction.

A new international agreement was reached in 2000 to conserve the valuable tuna populations of the Western Pacific Ocean from which approximately half the supply of the world s tuna comes. However, the agreement by European ministers in 2000 to reduce fishing quotas for several commercial species in European waters will affect neither the long-distance fishing fleets from EU countries nor the fishing subsidies that fuel over-fishing. According to the World Wide Fund for Nature (WWF), the EU is subsidising its fishing industry by at least 14 000 Euro per fishing boat per year.

Despite the efforts to use aquaculture as a substitute for the harvest of wild fish (*see* Aquaculture and Environment: Global View from the Tropics to High Latitudes, Volume 3; Aquaculture in Asia, Volume 3; Aquaculture: Salmon Farming, Volume 3), farmed fish sometimes compete with native species and spread exotic diseases. Farmed Atlantic salmon often escape from net pens and invade wild populations. Perhaps as many as 40% of all salmon caught in the North Atlantic originate in farms. In addition, more than 255 000 Atlantic salmon also have escaped into the Pacific Ocean since the early 1980s and are routinely caught by fishing vessels from Washington to Alaska. It is possible that farm escapees may hybridize with and alter the genetic makeup of wild populations of Atlantic salmon. Such genetic alterations could exacerbate the decline in many locally endangered populations of wild Atlantic salmon.

CONCLUSION

World fish stocks are in serious decline. The collapse of the Grand Banks cod fishery and the virtual collapse of the cod fishery in the North Sea indicate the magnitude of the problem. Eleven of the world s 15 major oceanic fishing areas have been fished at or beyond their maximum sustainable yield for commercially valuable species and are in a state of decline. Sixty-nine percent of the world s commercial fish stocks are falling and urgent international action to restore them is required.

The fear is that because the cost of rehabilitation of natural fish stocks is escalating exponentially, the damage may become irreversible in some cases. Fish is the major source of protein for many of the world s poorest people, who will be the hardest hit by rising prices and scarcity of sh. The result may be even more exploitation and degradation of the land surface and its soils, as people attempt to substitute vegetables for sh in their diet. Degradation of marine sh stocks could thus eventually lead to further degradation of terrestrial ecosystems and soils.

See also: **Fisheries: Effects of Climate Change on the Life Cycles of Salmon**, Volume 3; **Fisheries: Paci c Coast Salmon**, Volume 3; **Policies for Sustainable and Responsible Fisheries**, Volume 4.

REFERENCES

United Nations Food and Agriculture Organization (FAO) (1997a) *The State of World Fisheries and Aquaculture 1996*, FAO, Rome.

United Nations Food and Agriculture Organization (FAO) (1997b) Fishstat-PC, FAO, Rome.

United Nations Food and Agriculture Organization (FAO) (1997c) FAO Fisheries Circular No. 920, FIRM/C920(En), Rome.

Chernobyl

Burton Bennett
New York, NY, USA

The accident at the Chernobyl nuclear power station located in Ukraine near the border of Belarus occurred on the 26th April 1986. At the station were four reactors, each of 1000 MW capacity, of the graphite-moderated, light-water-cooled Soviet design known as RBMK-1000. Two of the reactors were constructed between 1970 and 1977 (Units 1 and 2) and two were completed in 1983 (Units 3 and 4). The accident happened during a test of the electrical control system as the Unit 4 reactor was being shut down for routine maintenance. The operators, in violation of safety regulations, had switched off important control systems and allowed the reactor to reach unstable, low power conditions. A sudden power surge caused a steam explosion that ruptured the reactor vessel and allowed further violent fuel steam interactions that destroyed the reactor and the reactor building. An intense graphite fire burned for 10 days before it could be brought under control, during which large amounts of radioactive materials were released to the environment.

The Chernobyl accident was the most serious ever to have occurred in the nuclear power industry. The accident caused the early death of 30 power plant employees and re ghters and resulted in widespread radioactive contamination in areas of Belarus, the Russian Federation, and Ukraine inhabited by several million people. In addition, enhanced radiation levels occurred throughout Europe, and trace levels were measurable everywhere in the northern hemisphere.

The radionuclides released from the reactor that caused exposure of individuals were mainly iodine-131, cesium-134 and cesium-137. Iodine-131 has a short radioactive half-life (8 days), but it can be transferred relatively rapidly through milk and leafy vegetables to humans. Iodine becomes localized in the thyroid gland. For reasons of intake of these foods, size of thyroid gland and metabolism, the thyroid doses are usually greater to infants and children than to adults. The isotopes of cesium have relatively long half-lives (cesium-134, 2 years; cesium-137, 30 years). These radionuclides cause long-term exposures through the ingestion pathway and from external exposure to these radionuclides deposited on the ground.

In addition to radiation exposure, the accident caused long-term changes in the lives of people living closest to the reactor, since measures intended to limit radiation doses included resettlements, changes in food supplies, and restrictions in activities of individuals and families. These changes were accompanied by major economic, social and political changes in the affected countries, resulting from the disintegration of the former Soviet Union.

Due to the seriousness of the accident, considerable measurements and assessments have been carried out by national and international organizations. The United Nations Scienti c Committee on the Effects of Atomic Radiation (UNSCEAR) conducted an evaluation of the average doses in separate regions of countries and for the population of the northern hemisphere as a whole that was published in the UNSCEAR report (United Nations, 1988) (*see* **UNSCEAR (United Nations Scienti c Committee on the Effects of Atomic Radiation)**, Volume 4). The experience gained in treating the immediate radiation injuries of workers and re ghters involved in controlling the accident was also reviewed in the UNSCEAR report.

The many measurements made following the Chernobyl accident have been useful to study the characteristics particularly of cesium behavior in the environment, the seasonal features of radionuclide transfers to foods and the enhanced accumulations that were observed in, for example, mushrooms, berries, lake sh, lichens, reindeer, and game animals. The methodology for estimating doses from such accidents have been improved. Due to the uneven deposition of radionuclides caused mainly by rainfall and the importance of particular foods leading to intake of radionuclides into the body, and because inadequate measurements

could be made at the time, especially of the short-lived iodine-131, the estimation of radiation doses to particular individuals requires time-consuming reconstruction efforts.

The average radiation doses to larger population groups in the most contaminated areas of Europe were of the order of 1 mSv or less in the first year following the accident. About twice the first year dose was received during the following few years. This may be compared with the average dose rate from natural background radiation of 2.2 mSv year^{-1}. In more contaminated areas surrounding the reactor, the average radiation doses were on average about 10 mSv during 10 years following the accident. The doses to the thyroid gland, were in some cases, much higher.

Large groups of people were evacuated from contaminated areas within days, weeks, or months after the accident, so the specific radiation doses depend on the particular circumstances of location, diet, age, etc., of each individual.

During the last several years, considerable attention has been devoted to investigating possible associations between health effects in the populations and the exposure to radionuclides released and dispersed following the accident. A review of these results is included in the UNSCEAR report (United Nations, 2000). Of particular note has been the occurrence of numerous thyroid cancers in children who lived in the severely contaminated areas of Belarus, the Russian Federation, and Ukraine. The number of cases, nearly 1500 between 1990 and 1997, is considerably greater than expected based on previous knowledge. The high incidence and the short induction period have not been experienced in other populations, and other factors are most certainly influencing the risk. If the current trend continues, further thyroid cancers can be expected to occur, especially in those exposed at young ages. The most recent findings indicate that the thyroid cancer risk for those older than 10 years of age at the time of the accident is leveling off, while the increase continues for those younger than 4 5 years in 1986.

Apart from the significant increase in thyroid cancer after childhood exposure, there is no evidence of a major public health impact 14 years after the Chernobyl accident. No increases in overall cancer incidence or mortality have been observed that could be attributed to ionizing radiation. The risk of leukemia, one of the major concerns after radiation exposure, does not appear to be elevated even among the workers who participated in the recovery operations after the accident. Neither is there any scientific proof of other non-malignant disorders, somatic or mental, that are related to ionizing radiation.

A majority of the epidemiological studies completed to date are of the geographical type, in which average population exposures are correlated with the average rates of cancer incidence in specific areas and time periods. As long as individual doses with reasonably low uncertainties are not available, the extent to which health effects might be radiation-related remains unclear. The reconstruction of individual doses is a key element in future research on radiation-associated cancers related to the Chernobyl accident. Although the Chernobyl accident could shed some light on our knowledge of the late effects of protracted radiation exposures, it must be recognized that because of the relatively low doses received by the majority of exposed individuals, any increase in cancer incidence or mortality will be difficult to detect in epidemiological studies.

Although the Chernobyl accident caused widespread contamination of local and regional areas of Europe and alarming short-term increases in radiation exposures of the population, the radiation exposures were, in fact, relatively small compared with those caused each year by natural background sources. The accident caused the disruption of the lives of many thousands of people in the immediate areas surrounding the reactor, and many health problems are being attributed to the accident. It appears, however, that the main radiological consequence is the occurrence of numerous thyroid cancers in exposed children. Continued studies will be needed to assess further the exposures and long-term effects of this most serious nuclear reactor accident.

REFERENCES

United Nations (1988) Sources, Effects and Risks of Ionizing Radiation. United Nations Scientific Committee on the Effects of Atomic Radiation, 1988 Report to the General Assembly, with annexes, United Nations sales publication E.88.IX.7, United Nations, New York.

United Nations (2000) Sources, Effects and Risks of Ionizing Radiation. United Nations Scientific Committee on the Effects of Atomic Radiation, 2000 Report to the General Assembly, with annexes, United Nations sales publication, United Nations, New York.

Climate Change: Tourism Influences

see Tourism: Climate Change and Tourist Resorts (**Volume 3**)

Climate, Urban

see Urban Climate and Respiratory Disease (**Volume 3**)

Climatic Extremes

Kelly Sponberg
NOAA, Silver Spring, MD, USA

Increasing concern for global warming and wider, more rapid, media coverage of natural disasters have given climatic extremes greater prominence in human affairs since 1990. Perhaps fueled by the widespread impacts of events such as El Niño, or perhaps because of the activities of global organizations concerned with the environment and natural disasters, more and more people have become intrigued by the way climate creates both severe hardship and great bene ts.

Modern communications and environmental monitoring technology, especially weather satellites, have made us much more aware of extreme climatic events and associated disasters. Analyses of interlinked extremes, especially those associated with El Niño and La Niña occurrences have heightened our awareness of the impact of climate upon society and the important roles that climatologists with good regional knowledge can play in analyzing climatic variability/trends and using those analyses to reduce exposure to disasters.

The following discussion aims to provide a broad context within which to understand climatic extremes and the effect of these events upon society. Every extreme weather or climate event can teach us more about the interactions between humans and the environment. Rather than providing an exhaustive list of extremes, the following paragraphs present an overview of some key concepts and issues.

WHAT ARE CLIMATE EXTREMES?

Climate is simply the average of weather over a period of time. The climatic time scale may cover periods from months to centuries or millennia. Likewise, climate can be local, regional or global. It simply depends on what is being averaged, or more specifically the area and period of interest. In contrast, weather is always very local and measured in days if not hours.

Many weather extremes are a result of, and are fostered by, shifts in the background climate. Nevertheless, damaging weather events do not always reflect abnormal climatic conditions. Some climates always have events such as hurricanes and tornadoes embedded in them.

Most people are familiar with the threat that global warming poses and the warm Northern Hemisphere summers of the 1990s. However, the climate varies from year to year, from decade to decade and on even longer time scales. For instance, El Nino, a phenomenon characterized by warming sea surface temperatures in the Eastern Equatorial Pacific, can disrupt and alter global weather patterns. El Nino itself is a climate scale event, occurring every 2 7 years and requiring months to develop. The disruption caused by El Nino often persists for longer than a year: this global weather pattern has a profound effect on humans and the surrounding biosphere. Estimates of damages associated with the 1997 1998 El Nino place global structural losses at up to $60 billion and deaths reported at 21 000 (*see* El Nino and La Nina: Causes and Global Consequences, Volume 1).

Meteorologists and climatologists discuss extremes technically in terms of such physical phenomena as duration of drought, intensity and volume of rainfall and force of the wind, but most interest in climatic extremes results from the way in which society is affected by these events. For this reason, the popular media and general public think of climatic extremes as events resulting in significant human and property loss. In the past, meteorological records have often been inadequate to build up a picture of the physical characteristics and impact of climatic extremes. However, modern monitoring and information communications technology have permitted more consistent recording and description of these events.

Reconstructing the historical record of the interactions between humans and climate has been a painstaking but worthwhile task. For example, during the well documented winter El Nino events since 1950 (rainfall seasons of 1951 1952, 1957 1958, 1965 1966, 1968 1969, 1972 1973, 1977 1978, 1982 1983 and 1991 1992), rainfall averaged about 37% greater than normal, with a mean anomaly of about 190 mm at San Francisco (Montiverdi and Null, 1998).

THE IMPORTANCE OF CLIMATE EXTREMES

Not only does climate in some way determine the food humans eat, but it influences our fiber and building materials, how we construct our shelters, the clothing we wear, and our energy needs. It is not surprising then that shifts in our climate can greatly disrupt how society functions.

In many industrialized countries, climate extremes cause hardship and misfortune, but government support, emergency response and rescue services, as well as insurance mechanisms, allow for relatively quick recovery while minimizing major effects upon the local and national economy. Not all nations and people are so fortunate. However, events such as Hurricane Mitch, can wipe away decades of development in just a few days, if not moments. For many people it is only through international support and humanitarian aid that previously realized standards of living can be met once again. Even with such support, however, the development of many areas chronically suffers from repeated or sequenced climate extremes. For

example a series of droughts, floods, and storms brought severe hardship upon North Korea in the later half of the 1990s.

While most people will never witness firsthand, or be directly affected by, a climate- or weather-related disaster, increased globalization and international economic interdependence now mean that an event occurring continents away will in some way, even if very minutely, play a role in all our lives.

TRENDS VERSES TRUISMS

While we probably still do not know enough about changes in physical and human systems (and much less the interaction of these two) to be able to confidently say whether climate extremes, and their impacts, are either worsening or decreasing in severity, many current databases report an increase in the damage caused by climatic extremes. When looking at such databases, however, it must be remembered that our ability to observe and record such events has dramatically improved over the past century. We may not know the full impacts of earlier changes in climate and climatic extremes because of a lack of historical data.

The available careful analyses of trends within natural disasters and climate extremes all show that the costs associated with climatic extremes have increased, almost exponentially, over the past few decades. These studies, both global and regional, reflect both population growth and increasing investment in infrastructure and property. On the other hand, fatalities caused by extreme events appear to have decreased considerably. Even with the recent extremes such as hurricane Mitch (*see* Natural Hazards, Volume 3), flooding in Bangladesh, or the Venezuelan mudslides, fatalities due to climate extremes have been reduced considerably. Prior to World War II, several events caused deaths of 500 000 or more people; some events in Asia even resulted in several million fatalities. Again, modern technology and communications, including satellite telephony and helicopter rescues, assist in providing warnings and making evacuations from threatened areas.

Given these trends, on a global scale it remains difficult to determine if the number of climate extremes has increased or if the increasing losses are a result of demographic and societal changes. Studies of events in the US demonstrate that the frequency of most climatic extremes (droughts, floods, hurricanes, and tornadoes) has indeed remained relatively constant. There are some noted decadal fluctuations, such as with hurricanes, but again these do not yet point to a long-term change in event frequency. International experience is likely to be similar to that of the US. However this should not be permitted to obscure the possibility of long-term trends in the present-day climate extremes. Indeed, one lesson is that if natural climatic variability can and does cause significant damage, we should not be surprised that changes in our climate will affect us in more severe ways.

Despite the use in recent years of the term *climate surprise*, to describe events resulting in severe damage, which were not entirely predicted, it should be readily clear that on a longer time scale, surprises do not exist. Fast or slow, harsh or soft, climate variability is a constant, which will continue to significantly affect us.

EXAMPLES OF CLIMATIC EXTREMES

Drought

Drought occurs when an area receives less than normal rainfall over an extended period of time. The exact definition varies among different climatic zones, nonetheless drought can affect any part of the globe.

The human perception of drought is dependent upon the traditional use, and access to, water. People only become aware of drought when human consumption of water and activities using water are hindered by a reduction in precipitation. Not surprisingly, marginal populations dependent upon water for agriculture and transportation are particularly vulnerable to climatic fluctuations that bring about drought in their region.

The most striking feature of drought is that it develops slowly and usually takes a long time to dissipate. As such, the effect on property and populations is a creeping phenomenon that is often hard to notice in its early stages. It is not unusual for a drought to last several years, and as a result the damage to agriculture and food supplies has often led to famine and displaced populations. In severe cases disease often spreads as food supplies dwindle and water supplies become contaminated. Even after a drought subsides, there are often losses that continue to accumulate as a result of land degradation or human migration.

In the early 20th century, China suffered severe droughts and related famines. From 1907 into the early 1940s, more than 35 million died from starvation. The worst impact occurred in the drought of 1907 when upwards of 24 million people died.

Of course other parts of Asia have also suffered from drought. In India the drought of 1900 claimed up to 3.25 million lives. In the Bengal drought of 1943, 1.5 million lives were reportedly lost, and another 1.5 million perished in the 1965 1967 drought famine.

Even in more recent years, drought has continued to affect the Asian landscape. During the 1997 1998 El Nino, dry conditions over much of Indonesia resulted in some

$3 billion worth of damage to crops and property. In addition many associated fires destroyed a reported 3.3 million hectares of land.

Many severe droughts affected Africa throughout the 1970s, particularly in the Sahel of Western Africa. Drought and famine between 1972 1975 claimed over half a million lives. A similar drought occurred between 1984 and 1985, but during this episode some 20 African nations were affected. Among these was Ethiopia, which suffered from prolonged drought from 1981 1985. Some estimate that during the worst period of the event, as many as 20 000 children died a month in Ethiopia.

Drought, however, does not always result is famine or a loss of life. The impacts are dependent upon government, economic, and social support structures. Again, marginalized people are ultimately at a greater risk from drought and other climate extremes.

A severe drought in the US from 1932 1940, also known as the American Dust Bowl, did not result in loss of life or famine but caused severe and long lasting loss of topsoil and associated land degradation. Nonetheless, it severely affected human lives. Some 350 000 farmers were displaced. The drought that affected the US between 1987 and 1989 produced total losses across many sectors of the economy of $39 billion, making that event the most costly recorded climate extreme in US history.

Floods

Floods vary greatly in extent and duration, depending on the size of the river basin in which they occur and the space, time, duration characteristics of the precipitation event, and/or snow melt, producing them. The damage they cause may be exacerbated by the severity of the event, occupation of the flood plain, population density, relative wealth of affected communities, as well as human structures such dams and levee banks. Although flash floods cause a significant amount of damage, the most severe events are often associated with widespread regional flooding in large river basins.

Some of the most extreme flood events have occurred along the Yangtze River in China, where some flooding is expected every year sometime between June and September. However, climatic fluctuations, compounded by changes in land use, can and have caused floods that exceed the capacities of levees and other water control measures. The worst flooding along the Yangtze occurred prior to the end of World War II in which three floods in 1928, 1931, and 1939 each caused over half a million deaths. The event of 1931 was particularly destructive as it resulted in 3.7 million fatalities while affecting another 28 million people. Damage from the event has been estimated at $1.4 billion. The flood of 1931 came at particularly bad time in that it followed the 1928 1930 drought, which itself followed/coincided with the 1928 flood along the Yangtze.

Flooding along the Yangtze continues to cause severe damage today, however the loss of life is much less. Instead, losses are mainly in terms of infrastructure. In 1991 flooding in China caused an estimated $9 billion dollars worth of damage. In 1996 the costs were reported at $20 billion, and as a result of the severe flooding in 1998, $30 billion dollars worth of damage occurred.

Flooding can occur anywhere that rain falls. In central Europe, floods occur along the Danube and its tributaries frequently as a result of snowmelt or heavy rain. In 1838 flooding associated with melting ice in the Danube Valley reportedly destroyed half of Budapest. Although other floods caused significant damage throughout the 20th century, the floods of 1997 were particular destructive. Two weeks of heavy rainfall in July of 1997 caused flooding along the banks of the Danube in Hungary, Austria, and Slovakia. Flooding north of the Danube basin, along the Oder River extended the damage into Germany and Poland. 0.61 million hectares in Poland were flooded with 1360 towns and villages being inundated. In Poland alone, estimated economic losses were placed at $3 billion.

Another historic flood occurred along the Mississippi River in 1993 after record rainfall occurred in the months of June and July. In some places, rainfall was reportedly as much as three or four times above normal. As the rivers connecting to the Mississippi swelled throughout the summer, flooding began to occur downstream of St. Louis as levees broke. In all, some 8.1 million hectares of farmland were flooded and 46 fatalities occurred. Direct economic losses were estimated at $12 15 billion; however, the long term costs to productivity and losses associated with displacement of families were far greater (*see* **Natural Hazards**, Volume 3).

Tropical Cyclones

Tropical cyclones, known as typhoons, cyclones, or hurricanes depending upon the region of the globe, are massive storms created over the oceans. The areas most severely affected by these events are coastal regions, slightly inland areas, and islands located at latitudes greater than five degrees. Although each tropical cyclone is really an individual weather event, these storms are worth mentioning in the context of climate extremes, not only for the severe damage they cause, but also because variations in climate can increase the frequency and intensity of these events. The every-few-years occurrences of the El Niño Southern Oscillation, a phenomenon in the Equatorial Pacific Ocean, is known to affect the Atlantic Ocean hurricane season.

Posing a triple threat of strong winds, storm surges, and heavy rainfall, tropical cyclones can cause very localized and severe damage. On August 24, 1992, category four, Hurricane Andrew caused severe damage along a 100 kilometer path across Southern Florida, US. Winds and floods destroyed 25 524 homes while damaging another 101 241. After striking Florida, Hurricane Andrew made landfall once more in Louisiana as a category 3 hurricane. Sixty-one lives were lost in this one hurricane, and the total cost of the damage in the US was estimated at $26 billion. Well before Andrew, however, hurricanes had caused severe damage in the Atlantic and Gulf of Mexico. On September 8, 1900, Galveston, Texas was stuck by high winds and storm surges from a hurricane. Over half the city was demolished and 10 000 people were left homeless. Prior to the storm Galveston was an up and coming port city, which took advantage of its location and natural harbor. Unfortunately, the city was only some two meters above sea level, thereby leaving it particularly vulnerable to storm surges and flooding. Following the Galveston hurricane, the city built a sea wall to protect itself from future storm surges.

Of all the other low lying areas affected by repetitive flooding associated with tropical cyclones, Bangladesh faces the most severe problems. Most of this densely populated nation is just above sea level, occupying the joint delta of the Brahmaputra and Ganges Rivers. In 1970 a cyclone affected Bangladesh resulted in the worst tropical cyclone damages recorded in the 20th century. As many as 500 000 people were believed to have perished as a result of storm surges and high winds. Construction of storm shelters and other mitigation strategies have greatly reduced the loss of life due to cyclones, however these storms continue to menace the area. In 1991 cyclone 02B struck Bangladesh causing 150 000 deaths and some $1.5 billion worth of structural damage.

Conjunctive Events

Conjunctive climate extremes are events where the impacts are compounded by a concurrent event or when the initial climatological event triggers a man-made disaster. In truth any climate extreme that impacts humans is in someway a conjunctive event, as it likely disrupts agricultural production, transportation, utilities, etc. The famine caused by drought is often the failure of agricultural and food supply systems. Epidemics that result from climatological events are often caused when sanitation and nutrition suffers.

However, a clearly conjunctive event occurred in Egypt in November of 1994 when heavy rains triggered flooding in the southern provinces of the country. On November 2nd, flooding in the province of Asyut toppled a fuel transport which then ignited. Burning fuel on top of floodwaters destroyed several villages. What was not destroyed by water was burned. In total over $140 million worth of damage occurred, 11 000 houses were destroyed, 11 000 houses were damaged, and 110 000 people were affected.

CONCLUSION

The severe impacts of climatic extremes described above emphasize the direct impacts on human beings and economies. The indirect impacts on human society through the effects of climatic extremes on ecosystems may be even greater. The feedbacks from droughts, which dry up stream beds and lower groundwater tables, are felt rapidly by urban societies if water rationing is introduced. Their impacts on rangeland ecosystems grazing stock and irrigated soils may be slower to reach crisis proportions, but can undermine rural economies and eventually affect urban food supplies. Such processes can produce environmental refugees (*see* Environmental Refugees, Volume 4) and even be a factor in civil conflict and war, which in turn can lead to environmental damage (*see* Environmental Changes Driven by Civil Conflict and War, Volume 3). Climatic extremes can readily induce premature deaths, as when elderly people die during heat waves, or people with bronchial ailments die during urban smogs (*see* Urban Sulfurous and Photochemical Smog, Volume 3).

However, people are not the only organisms affected by climatic extremes. Damage, particularly to small island habitats, can greatly reduce or eliminate the habitats of key animal and bird species. Droughts may desiccate small ponds and wetlands, killing species which are unable to adapt to drought conditions. On the other hand, widespread floods in inland Australia usually result in the germination of many dormant seeds, creating a beautiful carpet of wildflowers extending for dozens of kilometers. There are clear thresholds of environmental stress which individual species populations and plant and animal communities can withstand. Yet there is uncertainty over the frequency and duration of the extremes which produce those thresholds. Many suspect that ecosystems have a certain resilience to climatic extremes, which means that their response lags behind climatic change, but detailed palaeoclimatic investigations are only beginning to demonstrate whether or not that is the case.

REFERENCE

Monteverdi, J and Null, J (1998) El Nino and California Precipitation, Western Region Technical Attachment, 97 37, November 21, 1997 (updated June 24, 1998) http://twister.sfsu.edu/elnino/elnino.html.

FURTHER READING

Burton, I, Kates, R, and White, G (1993) *The Environment as Hazard*, 2nd edition, Guilford Press, New York.

Cutter, S (1994) *Environmental Risks and Hazards*, Prentice Hall, Englewood Cliffs, NJ.

EM-DAT: The OFDA/CRED International Disaster Database, www.md.ucl.ac.be/cred.

Karl, T, Nicholls, N, and Ghazi, A (1999) *Weather and Climate Extremes: Changes, Variations and a Perspective from the Insurance Industry*, Kluwer, London.

Kunkel, K, Pielke, Jr, R, and Changnon, S (1999) Temporal Fluctuations in Weather and Climate Extremes that Cause Economic and Human Health Impacts: A Review, *Bull. Am. Meteorol. Soc.*, 80(6), 1077 1098.

National Oceanic and Atmospheric Administration (1999) *Century's Top Weather, Water, and Climate Events*, NOAA.

Office of Coordinated Humanitarian Affairs, various situation reports.

Sponberg, K (1999) *Compendium of Climatological Impacts*, Vol. 1, US National Oceanic and Atmospheric Administration, Office of Global Programs.

Van der Vink, G, Allen, R M, Chapin, J, Crooks, M, Fraley, W, Krantz, J, Lavigne, A M, LeCuyer, A, MacColl, E K, Morgan, W J, Ries, B, Robinson, E, Rodriquez, K, Smith, M, and Sponberg, K (1998) Why is the United States is Becoming more Vulnerable to Natural Disasters, *EOS*, 79(44), Nov. 3, 1998.

Common Agricultural Policy (CAP)

Ian Bowler
University of Leicester, Leicester, UK

The common agricultural policy (CAP) of the European Union (EU) comprises the world's most advanced form of supra-national law and institutional regulation. Despite several attempts at reform, most of the objectives and measures initially negotiated by the member states in the 1950s remain, albeit modi ed to some extent. The pace of reform has been increasing, however, with the reduction in the CAP's protectionist measures under trade liberalization by the World Trade Organization (WTO) and the integration of the CAP into a broader rural development policy for the EU.

The CAP has had a range of positive and negative consequences, the latter now outweighing the former. The main positive bene ts include obtaining security of food supply for consumers in the EU, raising average farm incomes and funding a modernizing and increasingly ef cient farm sector. The principal negative consequences of the CAP have been the production of agricultural surpluses, the disruption of international trade, the restructuring of rural economies and societies, the high nancial cost to EU consumers and taxpayers, and degradation of the environment, especially as regards resources of water, soil and natural habitat.

INTRODUCTION

The CAP of the EU comprises the world s most advanced form of supra-national law and institutional regulation. The objectives and basic principles of the CAP were negotiated by the original six member states of the (now) EU between 1957 (Treaty of Rome) and 1962, with the first measures for a common market in cereals being introduced after 1967 (Bowler, 1985). There have been a number of subsequent reform periods, for example following the Mansholt Plan of 1968 and the MacSharry Reforms of 1992. However, the objectives and measures of the original CAP tend to have been modified and added to, rather than displaced.

OBJECTIVES

The objectives of the CAP are contained in Articles 38, 39 and 110 of the Treaty of Rome, namely: to incorporate agriculture within the common market; to increase agricultural productivity by promoting technical progress; ensuring the national development of agricultural production and the optimum utilization of all factors of production, particularly labor; to ensure a fair standard of living for the agricultural population, particularly by increasing the individual earnings of persons engaged in agriculture; to stabilize agricultural markets; to guarantee regular food supplies; to ensure reasonable prices in food supplies to consumers; and to support the harmonious development of world trade (Fennell, 1997). Many of these objectives are contradictory (e.g., forming a common market and creating harmony in world trade), others are open to political interpretation (e.g., a fair standard of living), while an environmental objective was not introduced until the early 1970s (Brouwer and Lowe, 1999).

PRINCIPAL MEASURES

Seven principal measures have underpinned the CAP: (1) internal commodity price guarantees, but at levels above world market prices, through target prices and intervention agency buying for each of the main farm products; (2) variable import levy protection of the internal market against imports, based on threshold prices; (3) export subsidies for the main farm products in surplus; (4) direct

income supplements for farms in defined areas (e.g., the less favored areas, 1975); (5) financial incentives for farm modernization, farmer retirement and farmer training; (6) limits on production (co-responsibility levies, budgetary stabilizers, production quotas, set-aside of arable land, 1988); and (7) a centrally financed farm budget (i.e., the European Agricultural Guidance and Guarantee Fund (EAGGF)).

The allocation of funds under the EAGGF has favored the guarantee (over 60% of expenditure of the CAP until the late 1990s) over the guidance section. Within the guarantee section, a few farm products have received the majority of price support, either through intervention buying and storage or export subsidies (e.g., milk, cereals, beef until the early 1990s and subsequently crops). The guidance section provides direct income payments and financial assistance for the modernization of farms, food processing and marketing, including the early retirement of farmers.

REFORMS OF THE CAP

Attempts have been made annually, since the early 1980s, to reduce the real value of farm price supports, the degree of protection afforded to internal markets and the level of export subsidies. Progress has been slow, however, and has tended to be overtaken by the rate of increase in agricultural productivity, as reflected in the volume of farm production. Eventually farm output has had to be limited by various types of production quotas, beginning in 1984 with milk, although limitations on sugar beet had been in operation much earlier. The pace and scale of reduction in price supports have been intensified during the 1990s, initially under trade liberalization by the General Agreement on Tariffs and Trade and latterly by the WTO.

Table 1 Selected MacSharry Reforms agreed by the Council of Ministers, 1992

Twenty-nine percent reduction in cereal support prices over three years
Compensation for rotational set-aside (a percentage of a base area)
Cross-compliance between set-aside and area-based income payments on arable production
Removal of co-responsibility levies
New non-rotational set-aside
Accompanying measures providing aid for environmental protection, forestry and early farmer retirement
Quotas in the beef, sheep and milk sectors

Source: adapted from Kay (1998).

Following the 1992 MacSharry Reforms (Table 1), arable area-based direct payments have been added to the principal measures of the CAP as a new form of income supplement, linked in cross-compliance to farm-based set-aside (e.g., 10% of a farm s 1989/1991 base area). Livestock headage payments (beef cattle and sheep) have been limited to fixed stocking densities, as a form of quota. Agri-environmental measures have been introduced through national agri-environmental programs (AEP) as accompanying measures, including payments for the development of organic farming; and financial assistance has been made available for the afforestation of agricultural land. In broad terms, the attempt is being made to decouple farm incomes from the volume of agricultural production.

COSTS AND BENEFITS OF THE CAP

The CAP is associated with a range of positive and negative consequences, the latter now outweighing the former. The main positive benefits include obtaining security of food supply for consumers in the EU, raising average farm incomes and funding a modernized and increasingly efficient farm sector. These benefits are often overlooked in assessments of the CAP, but the improvement in nutritional standards amongst the population of the EU, the assurance of food supplies at reasonable prices to consumers, and the marked increase in quality of life for most of the farm population since the immediate post second world war years cannot be dismissed lightly.

The principal negative consequences of the CAP include the production of agricultural surpluses, the restructuring of rural economies and societies, the high financial cost to EU consumers and taxpayers (e.g., 45 600 m ECU in 1998), and degradation of the environment, especially as regards resources of water, soil and natural habitat. Production surplus to domestic demand initially attracted most political attention, for it was achieved at costs and prices in excess of those on the world market. Surpluses necessitated purchase and storage by intervention agencies (e.g., cereal mountains and wine lakes), and then destruction, export under subsidy, or disposal as food aid. The latter two strategies then served to destabilize national markets in developing countries and international markets for agricultural exporting countries, such as the US, New Zealand, Brazil and Argentina.

Nor have the benefits of the CAP been evenly distributed within the farm sector (Laurent and Bowler, 1997). Most benefit has been to the producers of a relatively few farm products, located in the northern member states, and occupying larger farms. Nor did the CAP protect the farm sector from the economic forces of restructuring, resulting in the migration of a majority of farm workers to the cities, followed in later years by the occupiers

of small and increasingly marginalized farms. In addition, modernized, intensi ed and increasingly large farms have implemented farming practices damaging to the environment, including the pollution of water resources from the excessive application of inorganic fertilizers and the disposal of animal manures on both crop and grass land, the loss of natural habitats, and threats to food health and safety through pesticide residues, infection of food and livestock diseases.

CONCLUSION

The CAP continues to evolve, with new regulations and nancial incentives being provided, to guarantee food quality and safety, deepen and extend the MacSharry Reforms under Agenda 2000 (e.g., further movement from price supports to direct income payments, extending milk quotas, and transferring more funds to the Guidance section of the EAGGF). This last dimension re ects the integration of the CAP into a broader rural development policy for the EU, including the redesignation of Objective 1 and 2 regions for receiving funding and through which agriculture becomes integrated into, rather than separated from, the development of regional economies.

REFERENCES

Bowler, I (1985) *Agriculture under the Common Agricultural Policy: a Geography*, Manchester University Press, Manchester.

Brouwer, F and Lowe, P (1999) *CAP Regimes and the European Countryside*, CAB International, Wallingford.

Fennell, R (1997) *The Common Agricultural Policy: Continuity and Change*, Clarendon Press, Oxford.

Kay, A (1998) *The Reform of the Common Agricultural Policy: the Case of the MacSharry Reforms*, CAB International, Wallingford.

Laurent, C and Bowler, I (1997) *CAP and the Regions: Building a Multidisciplinary Framework for the Analysis of the EU Agricultural Space*, INRA, Versailles.

Community Forest – The FAO Definition

see Agroforestry (**Volume 3**)

Composting

see Soil Mineralization (**Volume 2**)

Consultative Group on International Agricultural Research (CGIAR)

see Green Revolution (**Volume 3**)

Consumption Patterns: Economic and Demographic Change

Jyoti Parikh
Indira Gandhi Institute of Development Research, Mumbai, India

Consumption patterns vary with demographic, economic and cultural differences. Total consumption, of course, goes up as population and income increase, but the structure of that consumption and growth rates of different segments of that structure are revealed by demographic indicators such as rural and urban, age and gender composition of the population among other indicators. Secondly, there is a reverse linkage in that consumption levels in uence demographic variables such as fertility life expectancy and mortality. However, it is difficult to consider the impact of demographic changes separately, without discussing economic prosperity and income levels. Economic prosperity is one of the major forces that drives consumption. In their contribution to Earth Summit at Rio, Parikh et al. *(1991) showed for the first time that 25% of the global population with high income levels in developed countries consumes 75% of the world s resources. Subsequently, Agenda 21 drew attention to consumption and production patterns. Since then, consumption patterns have become a widely recognized issue addressed by researchers, consumer associations in various countries and international organizations as well as in seminars and in print-media. The human development report by the United Nations Development Programme (UNDP) (UNDP, 1998) also devoted a special issue to this topic.*

A breakdown of several commodities by world regions is given in Table 1. What is consumed at various income levels? At a low income, i.e., at subsistence level, one usually consumes primary goods such as cereals, milk, meat and fuelwood. With further rise in income, secondary goods such as petroleum products, cement and fertilizers enter the consumption basket. Finally, at a high income, tertiary goods such as transport vehicles, consumer

Table 1 Long-term trends in private consumption of selected items, by region

Item	Year	World	Industrialized countries	Developing countries	Saharan Africa	Sub-Arab States	East Asia	Southeast Asia and the Pacific	South Asia	Latin America and the Caribbean
Meat (millions of tonnes)	1970	87	57	29	3	2	8	3	3	10
	1995	199	95	103	6	5	53	8	8	23
Cereals (millions of tonnes)	1970	473	91	382	27	20	142	41	112	33
	1995	866	160	706	56	49	236	82	212	57
Total energy (millions of tonnes of oil equivalent)	1975	5575	4338	1237	139	67	407	102	180	306
	1994	8504	5611	2893	241	287	1019	296	457	531
Electricity (billions of kW-hours)	1980	6286	5026	1260	147	98	390	73	161	364
	1995	12 875	9300	3575	255	327	1284	278	576	772
Gasoline (millions of tonnes)	1980	551	455	96	10	12	11	8	6	48
	1995	771	582	188	15	27	38	19	13	72
Cars (millions)	1975	249	228	21	3	2	0.5	2	2	12
	1993	456	390	65	5	10	7	7	6	27
Bicycles produced (millions)	1970	36								
	1995	109	–	–	–	–	–	–	–	–
			–	–	–	–	–	–	–	–
Cigarettes (per capita)	1971		2860	860	Africa[a]	Europe[a]	East Asia[a]	SE Asia/ Pacific[a]	Eastern Mediter-ranean[a]	Americas[a]
	1981		2980	1220						
	1991	1325	2590	1410	480	2080	195	415	890	1530

[a] 1995. *Source:* United Nations Environment Program, 1998.

goods and appliances and material-intensive services are demanded. Not all consumption at a high income is material intensive perfumes, jewelry, and restaurant services, e.g., do not require much material. Thus, consumption levels depend on the kind of population that is growing rich or poor.

Vastly different consumption patterns can be seen in rural and urban areas, especially in developing countries. Some of the differences may be due to low availability of infrastructure and availability of merchandise in the rural areas. For example, in India 70% of the population still lives in rural areas. The ratio of *per capita* expenditure by the rural and urban populations may indicate the disparity. The ratio of the *per capita* expenditure of the average urban person to that of the average rural person is the *rural to urban disparity ratio* (RUDR). These ratios are presented in Table 2 for different commodity groups. On the average, top 10% in urban areas have *per capita* expenditure 9.2 times larger than the rural bottom 50%. While these are large differences within a country, the urban top 10% in India have much lower consumption levels than the average European or North American.

When rural people migrate to urban areas or developing country residents migrate to a developed country, the consumption increases and the patterns also change. Apart from income differences, the change is primarily due to differences in infrastructure such as roads, highways, public transport, airports, power plants, water supply, hospitals and many other services not available often in the rural areas. Despite their ability to pay, rich people in developing countries end up sharing poor infrastructure with the rest of the population. This explains why in developing countries, even very rich persons do not consume as much as an average person in developed countries (World Resources Institute, 1994). Infrastructure differences arise from different investment levels. Thus, investment patterns public and private also determine consumption patterns.

Another critical demographic variable is age structure, e.g., young people require schools, books, goods for sports and travels, whereas seniors require facilities such as hospitals, diagnostic centers and medical care centers. In developed countries, the relatively affluent senior citizens often contribute to a rise in leisure industries. In the developing countries, they depend on their children to take care of them. Their consumption usually falls with age.

Gender also affects consumption patterns i.e., females have different consumption patterns from males. Not many data are available but the differences are particularly evident in the poor households where women do not often get to voice their preferences. Moreover, on their own, a woman puts the interests of her home, children and male members of the family ahead of her own desires.

Cultural differences and investment patterns determine consumption patterns. With globalization, the cultural differences gradually vanish or get blurred.

Consumption *per capita* is highest in the wealthy countries that have both high average personal wealth and high investment in infrastructure and social support systems such as transportation, utilities, hospitals, educational facilities and so on. Such wealthy countries also often have the highest per capita expenditure on military hardware and weapons, adding to the proportion of global resources they consume. All this suggests that consumption by well-off people of wealthy countries is a disproportionately large driver of global environmental change.

Why should consumption levels matter? Sovereign nations do not like to discuss their consumption patterns. Similarly, within nations also people prefer to have freedom to choose their own levels of consumption. However, in addition to inequity in resource utilization, there are some environmental issues which link rich and poor within and across nations. Consumption patterns manifest themselves in environmental degradation such as contributions to ozone depletion and greenhouse gases (GHG) accumulation that pose a threat of global climate change (*see* National Responsibilities for Greenhouse Gas Emissions, Volume 3), pollution of the sea, accumulation of hazardous waste and nuclear wastes causing various types of global

Table 2 Disparity in per capita annual private consumption expenditures, 1989–1990, India[a,b]

	Rural	Urban					
Commodity group	Bot 50%	Mid 40%	Top 10%	Bot 50%	Mid 40%	Top 10%	Urban/rural RUDR
Food	1.0	1.7	2.9	1.3	2.3	4.2	1.4
Clothing	1.0	1.5	1.9	1.6	3.6	8.3	2.4
Fuel and power	1.0	1.4	2.9	2.3	4.7	10.8	3.1
Manufactures	1.0	2.3	7.4	1.8	5.6	16.2	2.2
Transport service	1.0	3.1	10.1	2.1	8.6	33.2	2.9
Services	1.0	2.1	6.1	2.0	5.5	19.3	2.6
Total expenditure	1.0	1.8	3.8	1.6	3.5	9.2	1.9

[a] *Source:* computed by The Indira Gandhi Institute of Development Research (IGIDR) authors Murthy *et al.* (1997).
[b] The commodity groups are formed by aggregating the sectors of input-output matrix for India brought out by the Planning Commission. Each element describes how much inputs of various other commodities are required to produce one unit (value in rupees, say thousand rupees) of that commodity.

environmental stresses apart from resource shortages that follow excessive use. If the consequences of consumption have to be suffered by everybody, it is not a sovereignty issue or freedom of choice. If zero pollution is assured or if environmental concerns are fully internalized for all economic activities, people and countries have the freedom of choice. When this is not so, debate arises.

What can be done by the people?:

- Analyze consumption trends and driving forces along with their economic, environmental and social impacts.
- Involve local communities in addressing this issue in their own context to collectively come up with plans and measures.
- Develop an ethos that rewards the wise use of resources and sets standards to which all societies can aspire. Here media and manner of advertising have a role to play.

Thus, income levels dictate consumption patterns to a large extent, but the demographic variables such as population, rural and urban population shares, age structure and gender also play an important role.

REFERENCES

Murthy, N S, Panda, M, and Parikh, J (1997) Economic Growth, Energy Demand and Carbon Dioxide Emissions in India: 1990 2020, *Environ. Dev. Econ.*, 2, 173 193.

Parikh, J, Parikh, K, Gokarn, S, Painuly, J, Saha, B, and Shukla, V (1991) IGIDR Project Report for Earth Summit, Rio.

United Nations Development Program (1998) *Human Development Report with the then Sustainable Consumption and Production*, United Nations, New York.

United Nations Environment Program (1999) *Industry and Environment; Consumption: Facts and Figures*, Vol. 22 w.4, United Nations, New York.

World Resources Institute (1994) *World Resource Report*, Chapter 1.

FURTHER READING

UNCED (1992) *Agenda 21*, United Nations General Assembly, New York.

Contaminated Lands and Sediments: Chemical Time Bombs?

see Contaminated Lands and Sediments: Chemical Time Bombs? (*Opening essay*, **Volume 3**)

Coral Reefs: an Ecosystem Subject to Multiple Environmental Threats

see Coral Reefs: an Ecosystem Subject to Multiple Environmental Threats (**Volume 2**)

Critical Load

The term *critical load* was first used in the 1980s to define atmospheric deposition limits for sulfur (S) and nitrogen (N) pollutants for preventing acidification. An early definition (Nilsson, 1986) describes a critical load as the highest load that will not cause chemical changes leading to long-term harmful effects on the most sensitive ecological systems. By the late 1980s, the critical loads approach was being discussed as a means for setting pollution emission limits through modeling deposition patterns from emission sources and comparing them with critical loads. While various attempts were made to refine the definition for specific purposes (Bull, 1995), a general definition dating from 1988 (Nilsson and Grennfelt, 1988), was adopted in work under the United Nations Economic Commission for Europe (UN/ECE) Convention on Long-range Transboundary Air Pollution (CLRTAP) (UN/ECE, 1996) for a manual on mapping critical loads (UBA (Umweltbundesamt), 1998): a quantitative estimate of exposure to one or more pollutants below which significant harmful effects on specified sensitive elements of the environment do not occur, according to present knowledge. The manual summarizes agreed methods for calculating critical loads for acidity and eutrophication. It also describes ways of estimating critical levels, thresholds for the direct effects of pollutant gases. Critical loads or critical limits have also been used more recently for other pollutants, e.g., some heavy metals, where there are clearly defined thresholds. However, the term *critical load* is different from traditional ecotoxicological limits, being best applied to ecosystem effects for elements or compounds that are essential to life (Brydges, 1991). They are (usually of necessity) set close to damage levels whilst margins of safety are applied in ecotoxicology (Bull, 1994). In recent years, the European Union, in its Acidification Strategy, and the CLRTAP, in its protocols, have identified critical loads as their key environmental objectives to be achieved through pollution emission decreases (Bull and Fenech, 1999).

See also: Convention on Long-range Transboundary Air Pollution (LRTAP), Volume 4.

REFERENCES

Brydges, T G (1991) Critical loads, Reversibility and Irreversibility of Damage to Ecosystems, in *Electricity and the Environment*, Proceedings of a Senior Expert Symposium, International Atomic Energy Agency, Vienna.

Bull, K R (1994) The Critical Loads/Levels Approach to Gaseous Pollutant Emission Controls, *Atmos. Pollut.*, **69**, 105 123.

Bull, K R (1995) Critical Loads Possibilities and Constraints, *Water, Air, Soil Pollut.*, **85**, 201 212.

Bull, K R and Fenech, G (1999) International Activities to Reduce Pollution Impacts at the Regional Scale, in *Forest Dynamics in Heavily Polluted Regions*, eds J L Innes and J Oleskyn, CAB International, Wallingford.

Nilsson, J, ed (1986) *Critical Loads for Sulfur and Nitrogen*, Nordic Council of Ministers, Copenhagen.

Nilsson, J and Grennfelt, P, eds (1988) *Critical Loads for Sulfur and Nitrogen*, UNECE/Nordic Council Workshop Report, Skokloster, Sweden, Nordic Council of Ministers, Copenhagen.

UBA (Umweltbundesamt) (1998) *Manual on Methodologies and Criteria for Mapping Critical Loads/Levels and Geographical Areas where they are Exceeded*, Federal Environment Agency, Berlin.

UN/ECE (1996) *1979 Convention on Long-range Transboundary Air Pollution*, United Nations, New York and Geneva.

KEITH BULL *UK*

Crown-of-thorns Starfish

see Eutrophication (**Volume 2**); Coral Reefs: an Ecosystem Subject to Multiple Environmental Threats (**Volume 2**)

Curi, Kriton

(1942 1996)

Kriton Curi, Turkish scientist, pioneer, environmentalist, Professor of Civil Engineering at Bogazici University (Turkey), was the founder and director of the Turkish National Committee on Solid Wastes, and the Institute of Environmental Sciences.

He was the director of the Institute of Environmental Sciences, acting chairman of the Civil Engineering Department and a Senator at Bogazici University Senate at various times. Professor Curi was an active member of nine national and international professional associations, he was on the editorial board of three scienti c journals and he was the editor and author of 17 books and more than 100 articles.

Professor Curi had a very strong personality, valuing honor, friendship and loyalty. He was a highly motivated, principled, creative, open minded and self con dent scientist who had clear ideas; he would defend his beliefs and work towards his objectives courageously and with all his means, as if he were on a divine mission. He was the rst scientist in Turkey to create public awareness about environmental protection. He also introduced the concept of appropriate technologies for developing countries.

Professor Curi was principled, uncompromising and persistent regarding:

- national and international laws and by-laws regarding the environment;
- the elimination of environmental damage;
- laws and by-laws on higher education;
- improvement of the quality of university education.

His ideals always had a much higher priority than his personal gain. For example, even though he gave frequent talks and seminars, he would not accept an honorarium or fee; he would rather have the host institution give support to environmental protection, at least plant a tree in memory of the day.

Professor Curi was an industrious academician who loved his job and his students:

- even when he had demanding administrative duties, he would still teach 3 4 courses per semester;
- whenever he visited a town for the rst time, he would visit its solid waste removal site before its downtown;
- he frequently would go to high school students gatherings to speak about environmental problems at a level that students would understand.

Professor Curi was a tireless and effective organizer; he was constantly arranging national and international scienti c meetings, symposiums, seminars and fairs. But above all, Professor Curi was a devoted environmentalist and humanist, promoting harmony between humans and nature and peaceful coexistence between communities and nations. In the last decade, there has been almost no environmentally related national event or action and few international ones that Professor Curi has not been af liated with. These projects include:

- promoting cooperation between Turkey and Mediterranean Countries against forest fires;
- organizing an international conference of Mayors of Mediterranean cities to discuss common environmental issues;
- enhancing cooperation between Turkish and Greek industrialists to reduce water and air pollution in the Aegean Region;
- accomplishing the Vision of International Charter on the Environment by Students (VOICES) Conference, the first and the only environmental conference of students participating from all over the world.

That he was actively pursuing the projects at the time of his unfortunate and unexpected death, is clear evidence of his strong desire to enhance peace and cooperation in the volatile Middle-East Region.

Professor Dr Kriton Curi received many national and international awards and 90 plaquettes given by different institutions and non-governmental organizations for his contributions to the environment. A Kriton Curi Environmental Foundation has been founded in his memory.

GUNAY KOCASOY *Turkey*

D

Deforestation and Habitat Fragmentation in the Amazon Basin

Compton J Tucker, David Skole, Marc K Steininger, John R G Townshend, and Walter Chomentowski

NASA/Goddard Space Flight Center, Greenbelt, MD, USA

Tropical forest extent and deforestation in ~8 000 000 km^2 of the Amazon and Basin of South America were mapped using Landsat satellite images from the mid-1980s and early 1990s. Forest cover extent, including evergreen and deciduous forest, totaled 6 344 000 km^2, while the area of natural non-forest formations totaled 2 000 000 km^2. The area deforested totaled 58 000 km^2 in the middle 1980s; an additional 46 000 km^2 was deforested by the early 1990s. These data document a spatially concentrated deforestation zone in the Eastern and Southern Amazon Basin where deforestation is occurring at an accelerating rate. Contrary to popular perception, tropical dry forests are being cleared at faster rates than any other forest type in the Amazon Basin.

INTRODUCTION

Deforestation has occurred in the tropics throughout history (Tucker and Richards, 1983; Richards, 1984; Hecht and Cockburn, 1989; Williams, 1989, 1990) and has accelerated recently, particularly in areas of seasonally deciduous tropical forests (Schmink and Wood, 1984; Janzen, 1986; Fearnside, 1986, 1993; Houghton *et al.*, 1991; Myers, 1991; Skole and Tucker, 1993; Maas, 1995; Steininger *et al.*, 2000a). Accurate information on the extent of tropical forests and deforestation is essential for estimation of changes in surface energy balance and atmospheric greenhouse gas emissions (Cook *et al.*, 1990; Gash and Shuttleworth, 1991; Houghton, 1991; Keller *et al.*, 1991; Dixon *et al.*, 1994; Fearnside, 1996) and in terms of the global water balance (*see* Deforestation, Tropical: Global Impacts, Volume 3). Precise information about the spatial distribution of deforestation is also necessary to estimate the impacts of habitat destruction and fragmentation on biological diversity (Harris, 1984; Skole and Tucker, 1993; Pimm *et al.*, 1995; Laurence and Bierregaard, 1997; Laurence *et al.*, 1997, 1998; Chiarello, 1999).

Remote-sensing analyses of the Amazon, the largest area of tropical forest on the planet, have demonstrated dynamic deforestation frontiers, particularly in areas near highways or where industrial-scale agriculture is occurring (Fearnside, 1986; Skole and Tucker, 1993; Steininger *et al.*, 2000b). Deforestation in these areas causes high levels of fragmentation of the remaining, uncut forests. This is significant because forests near clearance edges are susceptible to an array of human and bio-climatological impacts (Malcolm, 1994; Laurence *et al.*, 1997, 1998; Cochrane and Schulze, 1999; Nepstad *et al.*, 1999) and the isolation of forest fragments affects local composition and biodiversity (Miller and Harris, 1977; Wilcox, 1980; Karieva, 1987; Laurence *et al.*, 1999; Gascon *et al.*, 2000). It is thus necessary to consider the isolation of forest fragments and the contact zone between deforested areas and intact tropical forest when quantifying the effects of deforestation upon biological diversity. The cumulative affected area of tropical forest where biological diversity is negatively impacted has been reported to be two to three times the actual deforested area (Skole and Tucker, 1993; Steininger *et al.*, 2000b).

Much of the controversy surrounding estimates of tropical deforestation arises from the lack of wall-to-wall measurements. Significant errors result from attempts to estimate tropical deforestation using a sampling approach, because tropical deforestation is highly concentrated spatially (Fearnside, 1986; Skole and Tucker, 1993). Any sampling attempt that does not take this into account will generate highly inaccurate estimates of deforestation (Tucker and Townshend, 2000; Sanchez *et al.*, 1997). Until recently, coarse resolution (• 1 km) satellite imagery has been the only source of data with wall-to-wall acquisitions, largely because fine-resolution imagery has been commercialized and thus expensive (Goward and Williams, 1997). Tropical forest extent has been mapped

with coarse-resolution satellite data (Tucker *et al.*, 1984; Malingreau and Tucker, 1988; Hansen *et al.*, 2000). However, most changes in forest cover occurs in small (• 10 km^2) patches which can not be reliably detected with these data.

AMAZONIAN DEFORESTATION

A systematic study of forest extent and change throughout large areas, such as the Amazon Basin, is only possible through the use of high-resolution satellite data. The data set we report are results from digital analysis of Landsat Multispectral Scanner (80-m resolution) and Thematic Mapper (30 m-resolution) satellite data, coupled with a geographic information system (GIS) to manage the analyzed data.

We classified each Landsat image into seven unique classes (forest; deforestation; regrowth; cerrado or savanna; water; cloud and shadow; and other) and merged these classes into a seamless spatial database. A total of 390 Landsat images have been analyzed for up to three time periods. Each Landsat scene was classified individually since atmospheric and viewing-geometry conditions varied among scenes. The only other alternative would be to attempt normalization of all the data to a constant sun target sensor geometry under specified atmospheric conditions and assume the land surface was invariant. This approach is impossible to achieve with Landsat data alone and was not attempted. However, it may be possible through combined analyses of new satellite data from the Landsat-7 and Terra satellites.

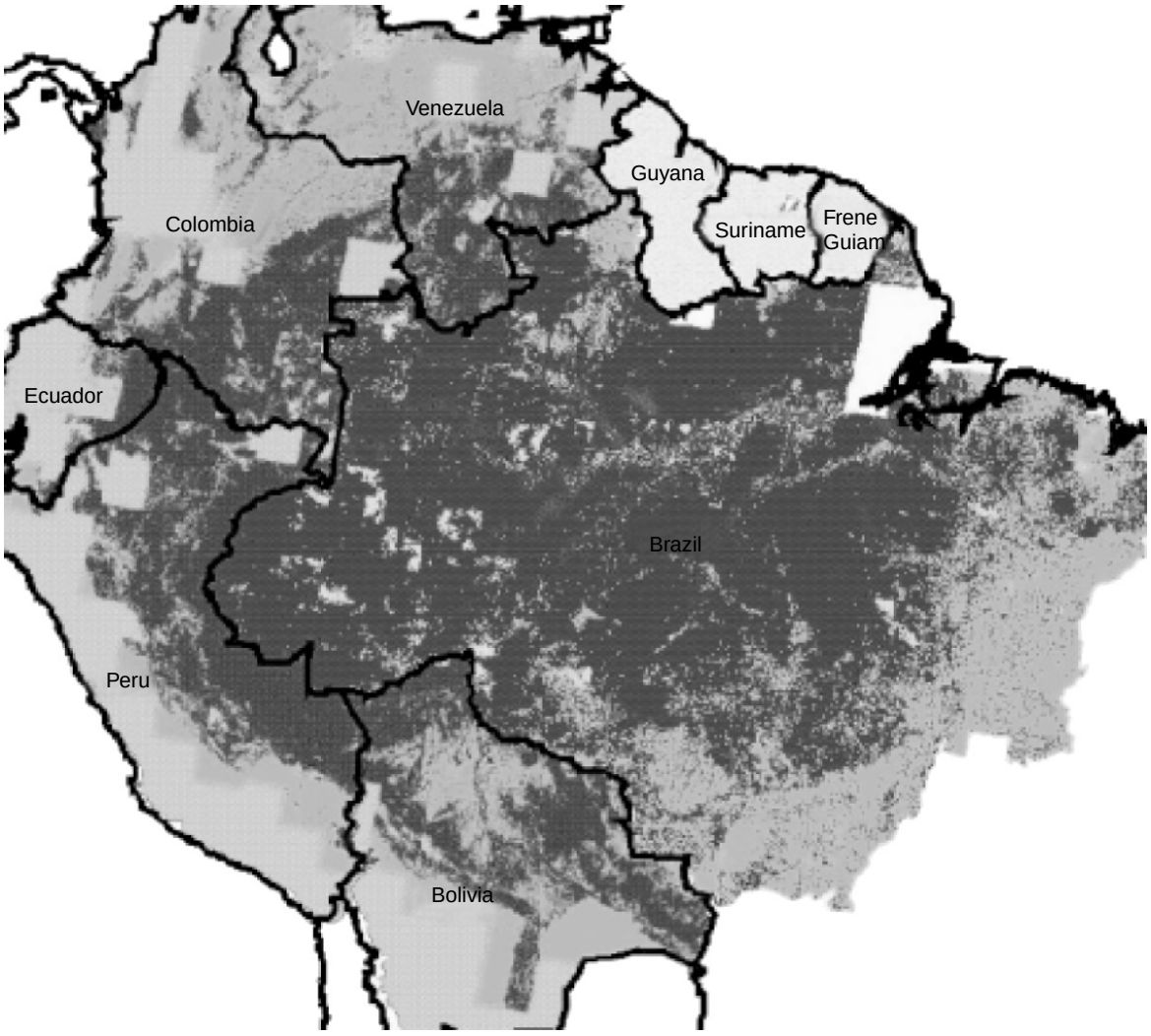


Figure 1 Distribution of Amazonian tropical forest in the early 1990s mapped with Landsat images. This product was produced by the NASA Landsat Humid Tropical Deforestation Mapping Project at the University of Maryland and the University of New Hampshire. Note the degree of fragmentation in the Eastern and Southern part of the basin. Persistent cloudiness has prevented direct observation of forest cover in some areas. Landsat data are still being acquired to fill in gaps caused by cloud interference

Table 1 Estimates of tropical forest area and deforestation for the Amazonian countries. (Data from UMD/MSU are for the Brazilian legal Amazon, and the Amazon basin only in non-Brazilian countries. Data from the FAO are for tropical forest and woodland in areas of • 1000 mm precipitation (FAO, 1993, 1996)

	Tropical forest area (km^2) (FAO)	Deforestation rate (km^2/year) (FAO)	Tropical forest area (km^2) (UMD/MSU)	Deforestation rate (km^2/year) (UMD/MSU)
Country	1990	1980–1990	1992	1986–1992
Bolivia	459 000	5320	426 000	1937
Brazil	4 093 000	36 710	4 093 000	15 200
Colombia	541 000	3670	415 000	883
Ecuador	120 000	2380	51 000	–
Peru	674 000	2710	495 000	1353
Venezuela	457 000	5990	285 400	387

Areas of mature forest, deforestation, young forest regrowth, savanna, water and cloud were mapped and stored in a GIS. Young forest regrowth was treated as deforestation and may constitute 20% to 40% of the total deforestation (Steininger, 1996; Houghton and Skole, 2000). Areas of large-scale deforestation were mapped, as were small clearings associated with indigenous communities, rubber tappers, coca farmers, mining operations, logging camps, airfields, blow-downs, illegal drug operations, and other small disturbances. All visible roads, power line right of ways, pipelines, and similar human-made features were also mapped as deforestation. These data were then merged and joined into a seamless data set for Brazil and the non-Brazilian Amazon Basin (Figure 1). The classification results are supported by field surveys carried out by our team and other research groups in Bolivia, Brazil, Ecuador, and Venezuela.

The largest areas of deforestation, edge-affected and fragmented forest were concentrated in an arc along the southern and eastern fringe of the Amazon Basin, extending from Pará, Brazil southward and westward to Santa Cruz, Bolivia. Deforestation in the interior of the Amazon Basin was closely associated with transportation corridors. Our estimates of deforestation in the Amazon, up to 1992 1993, are lower than other estimates, such as those in the widely cited reports of the Food and Agriculture Organization (FAO) (Table 1; FAO, 1981, 1993, 1996). We are very concerned that Amazonian deforestation is accelerating with the economic integration of the economies of South America. The Santa Cruz, Bolivia area is an example of accelerating deforestation throughout the 1990s, primarily for industrial soybean farming (Steininger *et al.*, 2000b).

The impact of tropical deforestation upon biological diversity is far greater than the deforestation or destruction of habitat *per se*. Our studies show that when edge and isolation effects are combined with actual deforestation, the total area affected negatively for biological diversity increases to two to three times the deforested area (Skole and Tucker, 1993; Steininger *et al.*, 2000b). The preponderance of affected habitat results mainly from proximity to areas of deforestation and not from the isolation of forest fragments. Skole and Tucker (1993) reported a 1988 Brazilian Amazon Basin deforestation of 230 000 km^2, a total 1-km buffer area of the deforestation-forest contact edge of 341 000 km^2, and only 16 000 km^2 of isolated forest fragments. Conversely, in areas of very high deforestation or older settlement histories, the remaining forests are mostly affected by the isolation of fragments (Gascon *et al.*, 2000). It is imperative to minimize both of these kinds of human activities that further fragment large areas of intact tropical forest.

See also: Tropical Forests, Volume 2; Deforestation, Tropical: Global Impacts, Volume 3.

REFERENCES

Chiarello, A G (1999) Effects of Fragmentation of the Atlantic Forest on Mammal Communities in South-eastern Brazil, *Biol. Conserv.*, 89, 71 82.

Cochrane, M A and Schulze, M D (1999) Fire as a Recurrent Event in Tropical Forests of the Eastern Amazon: Effects on Forest Structure, Biomass and Species Composition, *Biotropica*, 31, 2 16.

Cook, A G, Janetos, A C, and Hinds, W T (1990) Global Effects of Tropical Deforestation: Towards an Integrated Perspective, *Environ. Conserv.*, 17, 201 212.

Dixon, R K, Brown, S, Houghton, R A, Solomon, S M, Trexler, M C, and Wisniewski, J (1994) Carbon Pools and Flux of Global Carbon Forest Ecosystems, *Science*, 263, 185 190.

FAO (1981) *Tropical Forest Resources Assessment Project, Forest Resources of Tropical America*, Food and Agriculture Organization of the United Nations, Rome.

FAO (1993) *Tropical Resources Assessment 1990*, FAO Forestry Paper 112, Food and Agriculture Organization of the United Nations, Rome.

FAO (1996) *Forest Resources 1990: Survey of Tropical Forest Cover and Study of Change Processes*, FAO Forestry Paper 130, Food and Agriculture Organization of the United Nations, Rome.

Fearnside, P M (1986) Spatial Concentration of Deforestation in the Brazilian Amazon, *Ambio*, 15, 74 81.

Fearnside, P M (1993) Deforestation in Brazilian Amazonia: The Effect of Population and Land Tenure, *Ambio*, 22, 537 545.

Fearnside, P M (1996) Amazonian Deforestation and Global Warming: Carbon Stocks in Vegetation Replacing Brazil s Amazon Forest, *For. Ecol. Manage.*, 80, 21 34.

Gascon, C, Williamson, G B, and da Fonseca, G A B (2000) Receding Forest Edges and Vanishing Reserves, *Science*, 288, 1356 1358.

Gash, J H C and Shuttleworth, W J (1991) Tropical Deforestation: Albedo and the Surface Energy Balance, *Clim. Change*, 19, 123 133.

Goward, S N and Williams, D L (1997) Landsat and Earth Systems Science: Development of Terrestrial Monitoring, *Photogramm. Eng. Remote Sensing*, 63(7), 887 900.

Hansen, M C, DeFries, R S, Townshend, J R G, and Sohlberg, R (2000) Global Land Cover Classification at 1 Km Spatial Resolution Using a Classification Tree Approach, *Int. J. Remote Sensing*, 21, 1331 1364.

Harris, L D (1984) *The Fragmented Forest*, University of Chicago Press, Chicago, IL, 1 211.

Hecht, S B and Cockburn, A (1989) *The Fate of the Forest*, Verso, London, 1 240.

Houghton, R A (1991) Tropical Deforestation and Atmospheric Carbon Cycle, *Clim. Change*, 19, 99 118.

Houghton, R A, Skole, D L, and Lefkowitz, D S (1991) Changes in the Landscape of Latin America Between 1850 and 1985, II: a Net Release of CO_2 into the Atmosphere, *J. For. Ecol. Manage.*, 38, 133 199.

Houghton, R A, Skole, D L, Nobre, C A, Hackler, J L, Lawrence, K T, and Chowmentowski, W H (2000) Annual Fluxes of Carbon Dioxide from Deforestation and Regrowth in the Brazilian Amazon, *Nature*, 403, 301 304.

Janzen, D H (1986) Tropical Dry Forests: the most Endangered Major Tropical Ecosystem, in *Biodiversity*, ed E O Wilson, National Academy Press, Washington, DC, 130 137.

Karieva, P (1987) Habitat Fragmentation and the Stability of Predator-prey Interactions, *Nature*, 326, 388 390.

Keller, M, Jacobs, D J, Wofsy, S C, and Harris, R C (1991) Effects of Tropical Deforestation on Global and Regional Atmospheric Chemistry, *Clim. Change*, 19, 139 158.

Laurance, W F and Bierregaard, R O (1997) *Tropical Forest Remnants: Ecology, Management and Conservation of Fragmented Communities*, University of Chicago Press, Chicago, IL, 1 409.

Laurance, W F, Laurance, S G, Ferreira, L V, Rankin-de Merona, J M, Gascon, C, and Lovejoy, T E (1997) Biomass Collapse in Amazonian Forest Fragments, *Science*, 278, 1117 1118.

Laurance, W F, Ferreira, L V, Rankin-de Merona, J M, and Laurance, S G (1998) Rainforest Fragmentation and the Dynamics of Amazonian Tree Communities, *Ecology*, 79, 2032 2040.

Ledec, G (1989) *Bolivia Eastern Lowlands Development Project: Appraisal of Natural Resource Planning and Management Component*, World Bank, Washington, DC.

Maas, J M (1995) Conversion of Tropical Dry Forest to Pasture and Agriculture, in *Seasonally Dry Tropical Forests*, eds S H Bullock, H A Mooney, and E Medina, Cambridge University Press, Cambridge, 399 422.

Malcolm, J R (1994) Edge Effects of Amazonian Forest Fragments, *Ecology*, 75, 2438 2445.

Malingreau, J P and Tucker, C J (1988) Large-scale Deforestation in Southeastern Amazon Basin of Brazil, *Ambio*, 17, 49 55.

Miller, R I and Harris, L D (1977) Isolation and Extirpations in Wildlife Reserves, *Biol. Conserv.*, 12, 311 315.

Myers, N (1991) Tropical Forests: Present Status and Future Outlook, *Clim. Change*, 19, 3 32.

Nepstad, D C, Verissimo, A, Alencar, A, Nobre, C, Lima, E, Lefebvre, P, Schlesinger, P, Potter, C, Moutinho, P, Mendoza, E, Cochrane, M, and Brooks, V (1999) Large-scale Impoverishment of Amazonian Forests by Logging and Fire, *Nature*, 398, 505 508.

Pimm, S L, Russell, G J, Gittleman, J L, and Brooks, T M (1995) The Future of Biodiversity, *Science*, 269, 347 350.

Richards, J F (1984) Global Patterns of Land Conversion, *Environment*, 26(9), 6 38.

Sanchez, G A, Skole, D L, and Chowmentowski, W H (1997) Sampling Global Deforestation Data Bases: the Role of Persistence, *Mitigation and Adap. Strat. Global Change*, 2, 177 189.

Schmink, M and Wood, C H, eds (1984) *Frontier Expansion in Amazonia*, University of Florida Press, Gainesville, FL, 1 502.

Skole, D L and Tucker, C J (1993) Tropical Deforestation and Habitat Fragmentation in the Amazon: Satellite Data from 1978 to 1988, *Science*, 260, 1905 190.

Steininger, M K, Tucker, C J, Killeen, T, Ersts, P, Desch, A, and Hecht, S B (2000a) Forest Clearance and Fragmentation in the Tierras Bajas, Santa Cruz, Bolivia: Satellite Data from 1975 to 1998, *Conserv. Biol.*, (in press).

Steininger, M K, Tucker, C J, Townshend, J R G, Killeen, T, Desch, A, Bell, V, and Ersts, P (2000b) Tropical Deforestation in the Bolivian Amazon, *Environ. Conserv.*, 28, 127 134.

Steininger, M K (1996) Tropical Secondary Forest Regrowth in the Amazon: Age, Area and Change Detection with Thematic Mapper Data, *Int. J. Remote Sens.*, 17, 9 27.

Tucker, C J, Holben, B N, and Goff, T E (1984) Intensive Clearing in Rondonia, Brazil, as Detected by Satellite Remote Sensing, *Remote Sens. Environ.*, 15, 255 261.

Tucker, C J and Townshend, J R G (2000) Strategies for Monitoring Tropical Deforestation Using Satellite Data, *Int. J. Remote Sens.*, 21, 1461 1472.

Tucker, R P and Richards, J F (1983) *Global Deforestation and the Nineteenth Century World Economy*, Duke University Press, Durham.

UNEP (2001) An Assessment of the Status of the World s Remaining Closed Forests, UNEP, Nairobi, Kenya.

Wilcox, B A (1980) Insular Ecology and Conservation, in *Conservation Biology, an Evolutionary- Ecological Perspective*, Sinauer, Sunderland, MA, 95 117.

Williams, M (1989) Deforestation: Past and Present, *Prog. Hum. Geogr.*, 13, 176 208.

Williams, M (1990) Forests, in *The Earth as Transformed by Human Action*, eds B L Turner, R W Clark, R W Kates, J F Richards, J T Mathews, and W B Meyer, Cambridge University Press, Cambridge, 179 201.

Deforestation in Historic Times

Michael Williams
University of Oxford, Oxford, UK

Deforestation is as old as human occupation of the Earth, controlled fire probably being coterminous with the emergence of Homo erectus some 0.5 million years ago. Taking the long view, tropical forest destruction since circa 1950 is just a blip on an ever-upwardly sloping curve, probably being no more than between 6–7.5% of the total amount of forest ever cleared. Admittedly, compared to previous deforestation episodes, it is more rapid, more detrimental to global biodiversity, and is occurring in more sensitive and irreversibly damaged environments. But it is nothing new; it is part of the age-old quest of humanity for shelter, food and warmth.

The attraction of forested environments is not difficult to understand. Trees provide wood for construction, for shelter, and for making a multitude of implements. Wood provides fuel to keep warm, to cook food and even to smelt metals. The nuts and fruits of the trees are useful for food, medicines, and dyes, and the roots, young shoots and branches provide food for animals. Cleared forest provides land for growing crops – land, moreover, often covered by a deep humus accumulation and initially rich in nutrients. Clearing requires no sophisticated technology. Humans with stone or flint axes need only boundless energy to fell trees; in contrast, fire and browsing animals can wreak havoc in forested areas with little effort. The substitution of metal for stone axes circa 3500 years ago, and then for saws in the medieval period, eased the backbreaking task of clearing, and accelerated the rate of change, but they did not alter the basic process of destruction and land-use transformation. Power-saws during the last 50 years have made a major impact.

*Three basic questions bedevil the history of deforestation, as they do current concerns (Williams, 1994). What exactly is a forest, what constitutes deforestation, and what was the extent and density of trees at any past given time? Although there are many inconsistencies in various official definitions, for present purposes forest is defined loosely as ranging from a closed-canopy tree cover to a more open woodland (*see *Forest: the FAO Definition, Volume 2). Deforestation is taken to mean any process which modifies the original vegetation, from clear-felling, to thinning, to occasional fire. However, it should not be forgotten that forests regrow, often vigorously. The answer to the final question is partially dependent on answers to the preceding two questions. If human history is taken to mean after the last Ice Age (e.g., 10 000 before present (BP)), then the subsequent climatic fluctuations and ice sheet movements during the late Quaternary caused massive shifts in vegetation belts, especially in the middle latitudes. Palynologists have reconstructed the late Quaternary pollen record of the forest at both local and continental scales, but it is still difficult to capture the kaleidoscopic patterns of change in the distribution of trees. Past vegetation is of interest to climatic modelers who are aware of the role it plays in the radiation balance of the earth and in various bio-geochemical cycles related to climatic change. According to Matthews (1983) the pre-agricultural closed forest probably once covered* $46{\cdot}28 \times 10^6$ *km*^{2}, *and more open woodland* $15{\cdot}23 \times 10^6$ *km*^{2}; *these have been reduced by* $7{\cdot}01 \times 10^6$ *km*^{2} *and* $2{\cdot}13 \times 10^6$ *km*^{2}, *respectively. Historical reconstructions of known clearing episodes and potential rates of clearing based on population numbers support the general magnitude of change as being between approximately* $8{\cdot}05 \times 10^6$ *km*^{2}, *and* $7{\cdot}44 \times 10^6$ *km*^{2} *(Williams, 1991). Thus, the total area of forest in the world has possibly decreased by 15.15% and the woodlands by 13.8%, a massive amount to be sure, but not, perhaps the worldwide devastation that is commonly supposed. Further refinement of these figures is unlikely because of the paucity of reliable records. In the past, clearing was regarded as the most natural thing in the world; it was the first step to improvement and agricultural expansion, and almost no one wrote about it or recorded statistics about it.*

PRE HISTORIC CLEARING

Pre-literate societies throughout the world had a far more severe impact on the forests than is commonly supposed because the increase and spread of people took place in largely forested environments. This was true in Europe from the late Mesolithic and certainly during the Neolithic, and also in the far more chronologically and spatially diverse societies of the Americas, from north to south (Williams, 2000). In Europe, Mesolithic cultures did not avoid forests but actively engaged with them. There is evidence of cultivation, clearing and use of fire for game hunting in England in the upland fringes of the Pennines, North York Moors and Dartmoor, and successive clearings are accompanied by the presence of pollens of light-demanding plants, such as sorrel and ribwort plantain, which could only flourish as a result of clearing. The subsequent 3000 years of Neolithic agriculture/settlement was far more sedentary and stable than previously thought. The significance of the widespread incidence of timbered long houses has been ignored, yet many had been occupied for hundreds of years, which makes the previous dominant paradigm of a slash-and-burn economy unlikely. Rather, the Neolithics sought out gently shelving sites on the floodplain-lower interfluve slope zone, favoring loessic soils because of their fertility and not for their supposed treelessness. Flint and stone

axes made short shrift of the forest, and the flood plains were used for intensive garden cultivation and meadows. Cereal crop yields were sustained for surprisingly long periods on these soils, and shortfalls in diet were made-up by a reliance on cattle, which supplied meat, blood, milk and cheese, and sheep and pigs were also present. Consequently, large numbers of livestock were needed to make it economically feasible to extract milk and meat produce. Herds of between 10 and 20 are too small; 30 50 head are more realistic. A typical six-household, 30-person, locationally fixed settlement would have needed to plant 13.2 hectare (ha) of wheat, and run a 40-head herd of cattle with 40 sheep/goats. If the area used for housing, garden plots, a woodlot, constructional timber, pasture land, meadows and rotational forest is totaled up, then each group of 30 persons needed a little over 6 km^2 of woodland to survive, a staggering 20 ha per person (Bogucki, 1988; Gregg, 1988).

Modern experiments show that flint and stone axes were quite capable of being used for forest felling. Burning and animal grazing, if intensive enough, would have thinned and ultimately eliminated forest in other areas. The process continued unabated during the late Neolithic to early Bronze ages (*circa* 3000 1000 BC). Charcoal layers and successive decreases in forest pollens, followed by increases in cereal and weed pollens in peat deposits, together with interbedded farming and clearing implements, leave one in no doubt about the sequence of events. The evidence for similar processes of early forest disturbance with clearing are unfolding for the Americas so that any concept of virgin forests before European contact is a myth (Denevan, 1992). For example, burnings, swiddens (rotational burning and clearing for cropping), and manipulation of trees in the rainforest of equatorial upland areas may date from at least the early Holocene, and most soils are studded with charcoal. That could be due to natural lightning strikes but as these are usually accompanied by abundant rain in tropical areas, then humans must be responsible for most fires in prehistory. Ethnobotanists think that much of the Amazonian forest is a cultural artefact as native peoples have developed successive resource management strategies to cope with fluctuations in population dynamics. Similar arguments can be made for the Maya lowlands and other parts of tropical Central America.

In temperate North America, the vivid descriptions of indigenous clearing and agriculture in 16th and 17th century travel accounts is substantiated by archaeological and palaeobotanical evidence. From at least 12 000 BP the aboriginal population occupied the rich bottom-lands of the many river systems of the continent though they never abandoned hunting. Progressive clearing of the forests on the flood plains and lower terraces, and the intensification of cropping gradually converted the landscape into a mosaic consisting of permanent settlements and cultivated fields, early successional forests invading abandoned old fields, and remains of the original deciduous forest in the uplands (Delcourt, 1987). By 1000 AD, the Indians of the Woodland Culture were colonizing the fire-prone eastern woodlands.

With the exception of the long-settled savanna-woodland and adjacent belts in West Africa, knowledge about deforestation in Africa is sparse, and may well not have happened on a scale sufficient to be recorded. Whether in Europe, Africa, Asia or the Americas, the record is clear. The axe, together with dibble-and-hoe cultivation and later the light plough, often integrated with pastoral activity in Old World situations, reduced the extent of the forest and altered its composition. Fire was particularly destructive. Delcourt (1987) has suggested that early humans produced four major changes:

1. The increased frequency and magnitude of disturbance resulted in the expansion of non-forested patches or clearing.
2. The increasingly sedentary life style, the development of territorial control, and the high energy investment in the cultivation of crops resulted in a new sort of disturbance in which large areas were kept in the early stages of succession, which allowed the invasion of subsequent weed populations.
3. The selective utilization of plants by humans and animals resulted in long-term changes in the dominant tree structures within forest communities.
4. There were substantial changes in the distributional limits of certain species.

The impact of early humans on the forests was greater than suspected, and greater than many would care to admit. It may well have been one of the greatest deforestation episodes in history.

THE CLASSICAL ERA

Although the pre-historical period overlapped with the conventionally labeled classical one there were important differences. From *circa* 3000 BP to the end of the Dark Ages, increasing population, burgeoning urbanization, mineral extraction, and trade by different cultures and civilizations mainly on the northern rim of the Mediterranean basin brought enormous changes in forest vegetation. More importantly, archaeological and palaeoecological evidence is supplemented by a contemporary literature, e.g., the works of Strabo, Theophrastus, Cicero, Varro, and Columella. For the first time, people recorded what they saw, did to, and thought of, their external world and were conscious of their power to control and even create nature. The primary cause of clearing was either to grow food or facilitate grazing, followed in uncertain order by domestic fuel procurement, shipbuilding and

metal smelting. Tantalizingly, the detail of each process comes in roughly reverse order to its importance. Thus, as ever, clearing for growing food gets little mention as it is subsumed into the larger practice of agriculture. On the other hand, metal smelting looms disproportionately large because of the intense local impact, as at Rio Tinto, Populonia, Lauarion, or Cyprus, although it is doubtful if it was as devastating as it is commonly made out to be. The image of the industrious ploughman who subdued his woodland with flames and plough and who carted off the timber he had felled was common enough. And there were many references to extensive forests, e.g., at Avernian in Campania or the impassable and appalling Ciminian Forest of Southern Tuscany, that had either been vastly diminished or eliminated. But evidence is more abundant about felling to smelt metal or to produce fuelwood for domestic use and baths, to service the general timber trade to Imperial Rome and other cities, as a result of warfare, and particularly through shipbuilding by Venetian and Arab traders during the 7th to 11th centuries AD (Meiggs, 1982; Thirgood, 1981). The marked seasonality of the climate, the prevalence of fire and overgrazing by stock, particularly goats, resulted in the succession to an inferior woodland garigue or marquis which in turn could be degraded to poor pasture. Ultimately, massive erosion with associated deposition in almost tideless coastal locations, led to expansions of wetlands and the widespread onset of malaria by the fourth century BC. Centuries of overgrazing and clearing during the medieval period continued the process.

A general picture emerges of considerable change, but caution is needed not to over-state the extent of forest degradation and erosion or their contribution to economic decline. With a few exceptions, forests furnished the timber needs of the time and in the early Renaissance period, as evidenced by the great fleets launched in Venice, Genoa and Catalonia built from abundant, though diminishing, forests. It is likely that the ultimate denudation of the Mediterranean world came much later and was the product of population pressure during the 19th and early 20th centuries.

THE MIDDLE AGES

The Middle Ages, particularly in Europe, but also in China, witnessed a surge of economic activity in which the forest and its multiple riches played a central part. The extent of the forest clearing in the European continent is not known but it has been suggested that the forests of France were reduced from 30×10^6 ha to 13×10^6 ha between *circa* 800 and 1300 AD, (Bechmann, 1990) and still a quarter of the country was covered, while in Germany and Central Europe, perhaps 70% of the land was forest covered in 900 AD and about 25% remained by 1900 AD (Schluter, 1959), the high point of clearing occurring between 1100 and 1350 AD. The cultural climate of the age was crucial to understanding events. There was an intense religious belief in creating a divine, designed earth, coupled with a need to understand and use nature for practical ends. Lay and ecclesiastical owners cleared vast areas both for the glory of God and personal territorial gain; piety was accompanied by zeal (Glacken, 1967). The driving change was a six-fold increase of population between 650 and 1350 AD. Under-utilized and marginal lands were colonized in the west European heartland, and a massive expansion of settlement occurred in the forests of central and Eastern Europe.

In order to avert famine amongst the burgeoning population, food production needed to increase. Three technical innovations aided this. First, the dominant two-fields with one fallow was replaced by a three-field system, thus shortening the fallow period. This was possible because new crops like oats and legumes helped to fertilize the soil and added to animal and human nutritional requirements. Secondly, the development of the wheeled plough with coulter and mouldboard allowed cultivation to move from light soils onto heavy moist soils that were usually forested. Thirdly, ploughing efficiency was improved by the invention of the rigid horse collar and nailed horseshoes, both favoring the horse over the ox by increasing the speed and pulling power. Whereas the old manorial system was solidified, custom-bound and socially immobile and stratified, the new clearing and expansion produced a more fluid society. Aggrandizing landlords encouraged new settlers onto their lands, and offered generous terms of ownership, disposal and personal freedom, so that clearing contributed in a general way to the emancipation of the common man. This spectacular expansion is the subject of a vast literature. Place names indicating clearing and new settlements in forested areas abound, and rent rolls, charters, and leases show how expansion occurred. Pollen evidence is also useful.

These data show that cultivation expanded from about 5% of land use in the 6th century AD to 30 40% by the late Middle Ages, and the vast tracts of forest were fragmented, thinned or eliminated. It was also an age of organized colonization of the frontier by Germanic settlers into central and Eastern Europe, a thrust, which changed the ethnic and settlement map of that part of the continent. By the end of the 12th century the reduction of forest cover and the rise of personal power by the nobility led to their exerting territorial control on the remaining forests as hunting grounds. This was opposed by the peasants who gathered fuel, grazed stock and made good use of the forest products. The vast body of custom and rights that grew up subsequently to govern the use of the forest was a measure of its increasing scarcity. The various elements interlocked to produce what White

(1962) has called the agricultural revolution of the Middle Ages which asserted the dominance of humans over nature and shifted the focus of Europe from south to north, from the restricted lowlands around the Mediterranean to the great forested plains drained by the Loire, Seine, Rhine, Elbe, Danube and Thames. Here the distinctive features of the medieval world developed a build-up of technological competence, self-confidence, and accelerated change which, after 1500 AD, enabled Europe to invade and colonize the rest of the world. In that long process of global expansion, the forest and the wealth released from it played a central part.

Little is known about clearing in China where massive change to the forests must have happened also. There is a paucity of information on peasant life and livelihood, but an abundance of data on the working of bureaucratic government (Murphy, 1983). The detail of agricultural clearing is murky, but the demands of industry are clearer. A flourishing iron and steel industry in the Shantung region in Northeast China during the Northern Sung (910 and 1126 AD), and the early substitution of coal for charcoal suggest not only precocious technological development but widespread devastation and shortages of fuel. By roughly 1078 AD, production was about 125 000 150 000 tons, only a little less than total European production at the beginning of the 18th century (Hartwell, 1966). Then production declined by a half, whether through exhaustion of fuel, the Mongol invasions, or some other factor, is not known.

EARLY MODERN TIMES

The 450 years from 1492 *circa* 1950 AD are characterized by dramatic and well-documented terrestrial transformations. Europe burst through the confines of the continent with devastating consequences for the forests of the rest of the world. The previous largely land-based civilizations were now replaced by an inter-continental one and the creation of a global economy. The ecological imperialism (Crosby, 1986) of weeds, animals, diseases and ultimately people, that very largely began with Columbus in 1492 AD, ravaged the neo-European world. New economic and social systems were imposed or grafted on to existing societies. The European capitalistic economy commoditized nearly all it found, creating wealth out of nature, whether it was land, trees, animals, plants or people. This radical change occurred against a background of a steadily increasing world population (up more than six-fold from just over 400 million in 1500 AD to nearly 2.5 billion in 1950 AD: and before 1900 AD, most of this increase occurred in Europe). Additionally, rising rates of consumption for raw materials and food with urbanization, agricultural change and industrialization, first in Europe and, after the mid-19th century, in the US and Canada, put even greater strains on the forest resource. In the sparsely settled and mainly temperate areas, neo-European societies were planted and created. After an initial period of near elimination of the indigenes by virulent Old World pathogens, particularly smallpox after 1518 AD, followed by the triumph of Old World crops and stock, permanent settlement began in earnest by the mid-17th century. Although accompanied by much environmentally destructive exploitation the virtues of agriculture, freehold tenure, dispersed settlement, improvement and personal and political freedom were extolled, which all led to a rapid and successful expansion of settlement. Tree growth was considered a good indicator of soil fertility in all pioneer societies, and consequently the bigger the trees, the quicker they were felled to make way for farms. The US was the classic example of this neo-European clearing for agriculture with about 60 000 km^2 being felled by about 1850 AD and a further 660 000 km^2 by 1910 AD (Williams, 1989). It was one of the biggest deforestation episodes ever. Clearing was a combination of sweat, skill and strength and the pioneer farmer on the frontier was seen as the heroic subduer of a sullen and untamed wilderness. Clearing was widespread, universal and an integral part of rural life; and continued well into the early years of the 20th century. The French traveler, the Marquis de Chastellux marveled that;

> Such are the means by which North America, which 100 years ago was nothing but a vast forest, is peopled with three million of inhabitants. Four years ago, one might have traveled 10 miles in the woods I traversed, without seeing a single habitation.
>
> (Chastellux, 1789)

But it was not only the US where the archetypal image (and reality) prevailed of the single pioneer hacking out a life for himself and family in the forest. There were other parts of the world such as Canada, New Zealand, South Africa and Australia where similar pioneer clearing occurred. In Australia, for example, perhaps a total of another 400 000 km^2 of the southeastern forests and sparse woodland were cleared by the early 20th century (Williams, 1988). In the subtropical and tropical forests, European systems of exploitation led to the harvesting of indigenous tree crops (e.g., rubber, hardwoods), but also to the replacement of the original forest by crops grown for maximum returns in relation to the labor and capital inputs, often in a plantation system and often, too, either with slave or indentured labor. Classic examples of this were the highly profitable crops of sugar in the West Indies; coffee (green gold), and sugar in the subtropical coastal forests of Brazil, cotton and tobacco in the southern US, tea in Sri Lanka and India, and later rubber in Malaysia and Indonesia. All flourished at the expense of the forest. In Eastern Brazil, perhaps over half of the original 780 000 km^2 of the vast subtropical forest that ran down the eastern portions of the country had disappeared by 1950 through agricultural exploitation

and mining. In the state of Sao Paulo alone, the original 204 500 km^2 of forest was reduced to 45 500 km^2 by 1952 (Dean, 1995).

A variation of this agricultural clearing occurred in Southern Asia where peasant proprietors were drawn into the global commercial market. Outstanding was the deliberate expansion of rice cultivation in lower Burma by British administrators which resulted in the destruction of about 90 000 km^2 of rain forest between 1850 and 1950. Throughout the Indian subcontinent, the early network of railways meant an expansion of all types of crops, often for cash, that led to massive forest clearing in all parts of the country (Richards and McAlpin, 1983). The European impacts were not the only ones. Traditional societies often exploited their forests vigorously in a manner no more egalitarian or caring than those affected later by European commercial influences. There is plenty of evidence from, e.g., Southwest India and Hunan province in China (Perdue, 1987) from the 16th century onwards, that the commercialization of the forest was not a European invention. In pre-British Southwest India, permanent indigenous agricultural settlement existed side by side with shifting cultivation, and village councils regulated how much forest exploitation could be undertaken by agriculturalists. The forest was not regarded as a community resource; larger landowners dominated forest use in their local areas. Scarce commodities such as sandalwood, ebony, cinnamon, and pepper were under state and/or royal control. In Hunan, in south Central China, a highly centralized administration encouraged land clearance in order to enhance local state revenues so as to increase the tax base and support a bigger bureaucracy and militia. Later migrations into the forested hill country of South China were also encouraged by the state.

Simply, forests everywhere were being exploited and were diminishing in size as a response to increasing population numbers and increasing complexity of society. In the subtropical world, the changes came later than those unleashed by the Europeans with their new aims, technologies, and intercontinental trade links, but no less severe. In all, as much as 216 000 km^2 of forest and 62 000 km^2 of interrupted or open forest were destroyed in South and Southeast Asia for cropland between 1860 and 1950 alone. However, in focusing on distant lands and on the overseas expansion of Europe, one should not forget that Europe itself was being colonized internally during these centuries (Darby, 1956). This was particularly true in the mixed forest zone of central European Russia where over 67 000 km^2 were cleared between the end of the 17th century and the beginning of the 20th century (French, 1983). The insatiable demand for new land to grow crops and settle new societies in all ages has been matched by a rising demand for the products of the forest themselves. For example, the quest for strategic naval stores (masts, pitch, tar, turpentine) and ships timbers made major inroads into the forests of the Baltic littoral from the 14th century onwards and from the southern states of the US after about 1700 AD (Albion, 1926; Lower, 1973). Alternative construction timbers like teak and mahogany were discovered in the tropical hardwood forests after circa 1800 AD.

THE 20TH CENTURY

The pace of transformation increased during the first half of the 20th century. In the Western world, demands for timber accelerated. New uses (pulp, paper, packaging, plywood, chipboard) and relatively little substitution of other materials made timber a crucial resource in energy production, construction and industry generally. It had a strategic value akin to that of petroleum today. Agricultural clearing experienced a downturn with the abandonment of difficult, hard-to-farm lands and a concentration on easier-to-farm open grasslands. Consequently, farmland actually reverted to forest in the eastern US and Northern Europe, leading to a declining net loss of forest. In the tropical world the massive expansion of population by more than half a billion from a base of 1.1 billion resulted in extensive clearing for subsistence, accompanied by an expansion of commercial plantation agriculture. In all, perhaps 235×10^6 ha of tropical forest were lost from 1920 1949.

But the big changes occurred after 1950. While the temperate coniferous softwood forests have just about kept up with the demands of industrial societies for abundant supplies of timber and pulp, the focus of deforestation has shifted firmly to the tropical world. Better health and nutrition has resulted in a population explosion. These often landless-people have moved deeper into the remaining forests and farther up steep forested slopes, often with serious erosional consequences. They have little security or stake in the land and therefore little commitment to sustainable land management. Since 1950 about 550×10^6 ha of tropical forests have disappeared. In addition, the tropical hardwood forests are also being logged-out for constructional timber at an alarming rate, while wood is cut for domestic fuel in prodigious quantities in Africa, India and Latin America. In total cubic terms globally, fuelwood now roughly equals saw-timber extraction about $1{\cdot}8 \times 10^9$ m^3 compared to $1{\cdot}9 \times 10^9$ m^3, and it is forecast to rise rapidly in line with the increasing world population (Eckholm, 1975; Williams, 1991). It is probable that 2 432 000 km^2 of forest was cleared between 1860 and 1978, together with 1 502 000 km^2 of more open woodland (Richards, 1986, 1991).

CONCLUSION

The topic of deforestation is vast and complex, and is no less than a sizeable portion of world history. Many

more regional examples of deforestation from all ages could be cited. Nevertheless, the continuity of the process is striking. Forest clearing is one of the main causes of terrestrial transformation whereby humankind has modified the world s surface, a process, which is now reaching critical proportions (Williams, 1990). Whatever the detail, one thing is certain; with an ever-increasing world population, the process of deforestation will not end in the future. For those living in or near the tropical forest, it will continue to provide land, fuel and shelter for as long as the forests last, and be the mantle of the poor. Many will want to exploit its abundant timber resources while others will continue to restrict its use and preserve it. The tensions between exploitation and preservation will continue.

See also: Forest: the FAO De nition, Volume 2; Policies for Sustainable Forests: Examples from Canada, Volume 4.

REFERENCES

Albion, R G (1926) *Forests and Sea-Power*, Harvard University Press, Cambridge, MA.

Bechmann, R (1990) *Trees and Man: The Forest in the Middle Ages*, Translated by K Dunham, Paragon House, New York.

Bogucki, P I (1988) *Forest Farmers and Stockholders: Early Agriculture and its Consequences in North-Central Europe*, Cambridge University Press, Cambridge.

Chastellux, F J (1789) *Travels in North America in the Years 1780, 1781, and 1782*, Vol. 1, White, Gallaher and White, New York, 47.

Crosby, A W (1986) *Ecological Imperialism: the Biological Expansion of Europe, 900- 1900*, Cambridge University Press, New York.

Darby, H C (1956) The Clearing of the Woodland in Europe, in *Man's Role in Changing the Face of the Earth*, ed W L Thomas, Chicago University Press, Chicago, IL, 183 216.

Dean, W (1995) *With Broadaxe and Firebrand: The Destruction of the Brazilian Atlantic Forest*, University of California Press, Berkeley, CA.

Delcourt, H R (1987) The Impact Of Pre-historic Agriculture and Land Occupation on Natural Vegetation, *Ecology*, 34, 341 346.

Denevan, W M (1992) The Pristine Myth: the Landscape of the Americas in 1492, *Ann. Assoc. Am. Geogr.*, 82, 369 385.

Eckholm, E (1975) *The Other Energy Crisis: Firewood*, Paper No. 1, Worldwatch Institute, Washington, DC.

French, R A (1983) Russians and the Forest, in *Studies in Russian Historical Geography*, eds J H Bater and R A French, Academic Press, London, 1, 27 38.

Glacken, J (1967) *Traces on the Rhodian Shore: Nature and Culture in Western Thought from Ancient Times to the End of the Eighteenth Century*, University of California Press, Berkeley, CA.

Gregg, S A (1988) *Foragers and Farmers: Population Interaction and Agricultural Expansion in Pre-historic Europe*, Chicago University Press, Chicago, IL.

Hartwell, R (1966) A Revolution in the Chinese Iron and Steel Industry during the Northern Sung, 960 1126 AD, *J. Asian Stud.*, 21, 153 162.

Lower, A G M (1973) *Great Britain's Woodyard: British America and the Timber Trade, 1763- 1867*, McGill-Queen s University Press, Montreal, Canada.

Matthews, E (1983) Global Vegetation and Land Use: New High-Resolution Data Bases for Climatic Studies, *J. Clim. Appl. Meterol.*, 22, 474 487.

Meiggs, R (1982) *Trees and Timber in the Ancient Mediterranean World*, Clarendon Press, Oxford.

Murphy, R (1983) Deforestation in Modern China, in *Global Deforestation and the 19th Century World Economy*, eds R P Tucker and J F Richards, Duke University Press, Durham, NC, 111 128.

Perdue, P C (1987) *Exhausting the Earth: State and Peasant in Hunan, 1500- 1850*, Harvard University Press, Cambridge, MA.

Richards, J F (1986) World Environmental History and Economic Development, in *Sustainable Development of the Biosphere*, eds W C Clark and R E Munn, Cambridge University Press, Cambridge, MA, 53 78.

Richards, J F (1991) Land Transformation, in *The Earth as Transformed by Human Action: Global and Regional Changes in the Biosphere over the Past 300 Years*, eds B L Turner, W C Clark, R W Kates, J Mathews, W B Meyer, and J F Richards, Cambridge University Press, New York, 163 179.

Richards, J F and McAlpin, M B (1983) Cotton Cultivating and Land Clearing in the Bombay Deccan and Karnatak, 1818 1920, in *Global Deforestation and the 19th Century World Economy*, eds R P Tucker and J F Richards, Duke University Press, Durham, NC, 68 94.

Schluter, O (1959) *Atlas Ostliches Mitteleuropa* (plate 10), Velhagen and Klasing, Bielefeld.

Thirgood, J V (1981) *Man and the Mediterranean Forest: a History of Resource Depletion*, Academic Press, London.

White, L (1962) *Medieval Technology and Social Change*, The Clarendon Press, Oxford.

Williams, M (1988) The Clearing of the Woods, in *The Australian Experience*, ed R L Heathcote, Longmans-Cheshire, Sydney, 115 126.

Williams, M (1989) *The Americans and their Forests*, Cambridge University Press, New York.

Williams, M (1990) Deforestation: Past and Present, *Prog. Hum. Geogr.*, 13, 176 208.

Williams, M (1991) Forests, in *The Earth as Transformed by Human Action: Global and Regional Changes in the Biosphere over the Past 300 Years*, eds B L Turner, W C Clark, R W Kates, J Mathews, W B Meyer, and J F Richards, Cambridge University Press, New York, 179 201.

Williams, M (1994) Forests and Tree Cover, in *Changes in Land Use and Land Cover: a Global Perspective*, eds W B Meyer and B L Turner, Cambridge University Press, New York, 97 124.

Williams, M (2000) Dark Ages and Dark Areas: Global Deforestation in the Deep Past, *J. Hist. Geogr.*, 26, 28 46.

Deforestation, Tropical: Global Impacts

John H C Gash
Centre for Ecology and Hydrology, Wallingford, UK

Tropical rainforest is typi ed by a dense, continuous canopy stretching up to above 30 m in height, with a great variety of species. It is found in the humid tropics where the dry season is no more than about four months and at other times the rainfall is plentiful. Although rainforest is the world's most luxuriant vegetation, it grows on some of the poorest soils. Most of the nutrients are in the biomass and there is rapid recycling of nutrients from the soil into the vegetation, through the dense mat of roots just below the leaf litter layer. Rainforest also has deep roots, down to below 10 m in some cases, which allow the forest to continue transpiring through the dry season. Where the soil is not deep enough to maintain transpiration through the dry season, rainforest gives way to dry season deciduous, savanna forest.

When rainforest is replaced by agriculture, there are changes to the energy and water balances: less solar radiation is absorbed and more infrared radiation is emitted. The way in which that radiative energy is divided between evaporating water, and heating the air, also changes. Deforestation creates a more extreme local climate with higher daytime and lower nighttime temperatures. If there is large-scale deforestation, this may result in changed weather patterns and climate. Predictions are that large-scale deforestation will reduce rainfall.

Forest soils are very permeable, but deforestation using mechanized extraction techniques and heavy machinery, or continued trampling by cattle after deforestation will compact the soil, increasing the surface run-off and resulting in soil erosion. Deforestation will almost always make river ows increase more rapidly in response to storms, with a greater likelihood of ooding.

After fossil fuel combustion, destruction of rainforest by burning is the second largest contributor to the build-up of atmospheric carbon dioxide. However, the remaining rainforest is estimated to be a net sink for carbon dioxide. The forest takes up carbon dioxide during photosynthesis, but gives it out during respiration, as when vegetation dies and decomposes. The difference between these two uxes is small, and in dry years the forest may be a net source of atmospheric carbon dioxide.

Recognition of the economic value of rainforest both as a sink of carbon and a source of biodiversity may lead to changes in governments' policies on deforestation, but these will only be effective if it becomes more economically attractive for the local population to maintain the forest than to remove it.

INTRODUCTION

Tropical Rainforest

Tropical rainforest is a term commonly used to cover a range of forest types occurring in the humid, Equatorial belt across South and Central America, Africa and South-east Asia. The rainforest is typified by a dense continuous canopy, stretching to above 30 m in height, with a great diversity of plant species, the majority of which are evergreen. Tropical-moist-forest is a more precise definition of this range of biomes, since true rainforest occurs only in areas where there is no dry season and the rainfall is always greater than about 100 mm per month. The forest can extend to areas where there are dry seasons of up to four months, but where the dry season is longer, the rainforest gives way to dry savanna-type vegetation. Here we follow common usage, so that the word *rainforest* applies to the whole range of tropical moist forest (*see* Tropical Forests, Volume 2).

Rainforest Soils and Roots

The paradox of tropical rainforest is that it is the world s most luxuriant vegetation, but that it is growing on some of the world s poorest soils. The soils of the Equatorial tropics are generally low in nutrients and organic matter but the forest has evolved to optimize the use of the available nutrients. This is done so efficiently that almost all nutrients are held in the biomass itself. Recycling of nutrients in the forest is rapid: the high temperatures and humidity, and large insect population promote fast decomposition of fallen leaves or other dead vegetation, and the soil surface is covered by a dense mat of roots which intercepts the nutrients as they are released. The lack of nutrients in rainforest soils often results in deforestation leading to unsustainable agriculture, which is abandoned after a few years. There are also very deep roots, which have been observed at depths below 10 m (Nepstad *et al.*, 1994). These deep roots give the forest access to a store of soil moisture that enables it to survive the most severe dry seasons without significant water stress (Hodnett *et al.*, 1996a,b).

Evaporation from Rainforest

Annual evaporation from tropical rainforest is between 1000 and 1500 mm. Rain falling on to the forest canopy is intercepted and some 10 15% of it evaporates rapidly back into the atmosphere (Lloyd *et al.*, 1988). The rain which is not intercepted reaches the soil and some is later taken up by the roots to evaporate through the leaves in dry conditions (transpiration). Tropical rainforest is typified by a lack of seasonality in evaporation, with little evidence of any reduction in transpiration due to water

stress during dry season drought: rather the forest relies on a strategy of deep roots to access a large volume of stored soil water, which it uses to maintain transpiration. Where the soil is not deep enough to maintain transpiration through the dry season, the rainforest gives way to dry season deciduous savanna forest. The rapid evaporation of intercepted rainfall and the continuous transpiration results in a significant recycling of water vapor over continental rainforest: it is estimated that some 50% of the rain falling in central Amazonia originates from water evaporated by the forest. It is predicted (see below) that deforestation can disrupt this recycling process and result in a reduced rainfall.

DEFORESTATION

Deforestation occurs at several levels: complete clearance for agriculture, multiple clearance of small plots leading to forest fragmentation, or selective logging of valuable species. Large cleared areas can be detected by remote sensing. Detecting small clearance or selective logging is more difficult, but official statistics and socio-economic data can also be used to derive estimates of deforestation. Deforestation rates estimated by the Food and Agriculture Organization (FAO) are shown in Table 1. Current deforestation is not uniformly spread about the tropics, and there are still vast, remote areas where non-existent communications and lack of access to markets have protected the forest from settlement or exploitation. Deforestation is generally concentrated in hot spots, which usually follow a sequence of road building, migration, and clearance. Rondonia, in the southwest of Brazilian Amazonia, and Kalimantan, in Indonesia, are examples of hot spots with current high deforestation rates. However, in the central African countries bordering, and within, the Congo basin, there is relatively little migration and so far road building has resulted in selective logging rather than clear felling.

The increased human activity in logged or fragmented forest increases the incidence of fire. This risk is exacerbated by the open canopy of this type of forest: greater penetration of the wind into the lower canopy causes the vegetation to dry out during dry weather, making the forest more vulnerable to fire. Deforestation is also caused by flooding of reservoirs behind hydroelectric dams, by mineral extraction, not simply the extraction itself, but also the removal of forest to provide charcoal for smelters and by small-scale removal of wood for fuel to dry crops and for cooking.

CLIMATE

Replacing rainforest by agriculture changes the surface radiation and water balance. These changes affect the local climate, but they also have the potential to affect the thermodynamics and the dynamics of the atmosphere at larger scales and therefore change the weather and the climate.

The Radiation Balance

Albedo is the proportion of the solar radiation falling on the Earth s surface, which is reflected. Because forest is efficient at trapping radiation in its deep canopy, forest albedo is low when compared to agricultural land (Gash and Shuttleworth, 1991; Culf *et al.*, 1996). Typically rainforest has an albedo of about 13%, whereas pasture on deforested land has an albedo of about 18%. In addition to solar radiation, the surface receives long-wave, infrared radiation from the atmosphere and also emits it as black body, thermal radiation. The amount of long-wave radiation emitted depends on the temperature of the forest, but, because of the greater turbulence of the air over the forest, its surface temperature is close to that of the air. The surface of short agricultural crops is not so efficiently cooled and will be hotter than the forest s; this higher temperature results in greater long-wave emission. Agricultural land thus reflects more solar and emits more long-wave radiation than forest, resulting in about 10% less net radiation being available at the surface (see Figure 1).

Local Climate

The climate on the forest floor is largely decoupled from the atmosphere above: temperatures are lower than in the

Table 1 The area of remaining tropical forest in 1995, and its annual rate of change (from: State of the World's Forests. 1997, FAO, Rome)[a]

Region	Area of forest in 1995 Thousand km^2	Average annual change, 1990–1995	
		Thousand km^2 year^{-1}	% year^{-1}
Tropical Asia	2797	−31	−1.1
Tropical Africa	5049	−37	−0.7
Central America and Caribbean	241	−6	−2.2
South America	8279	−47	−0.6

[a] For comparison the area of Belgium is 31 000 square kilometers.

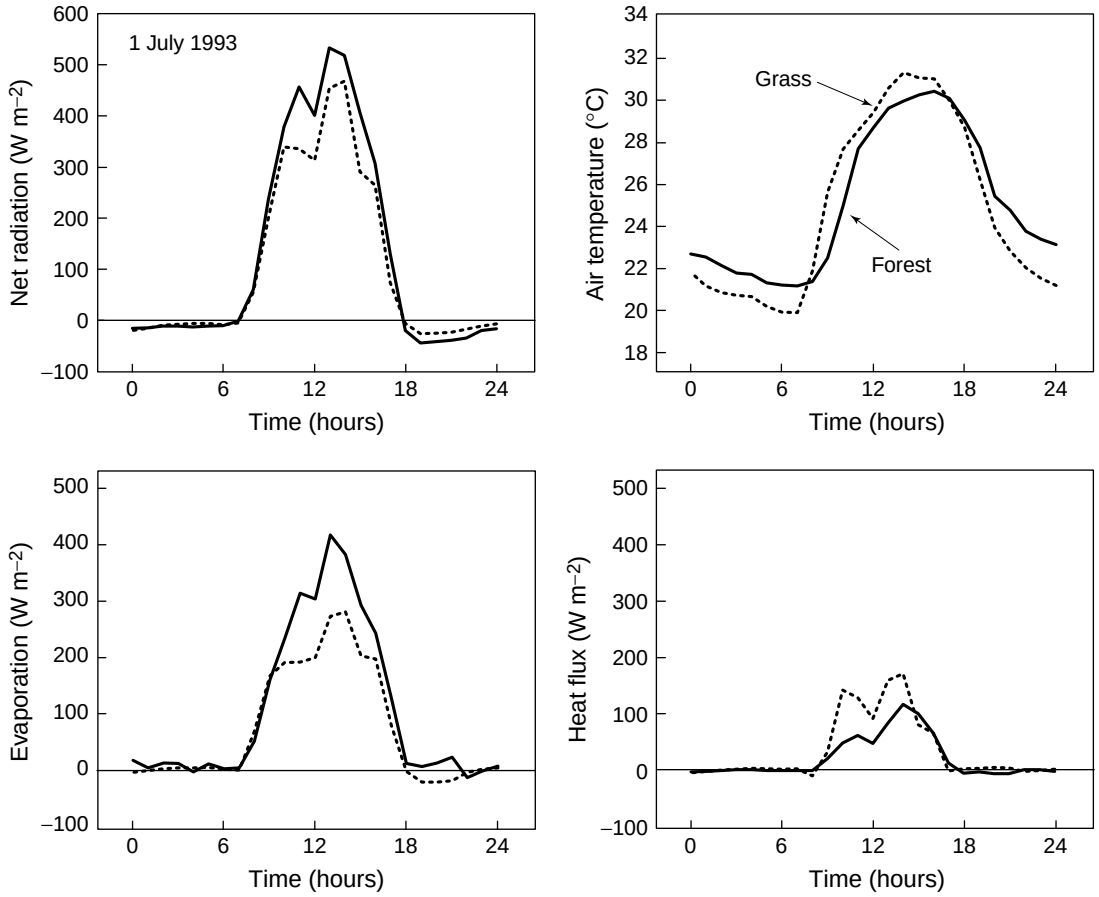


Figure 1 Hourly net radiation, air temperature and the fluxes of evaporation and sensible heat at forest and pasture sites in Rondônia, in Southwest Amazonia, for an example day in the dry season. (Reproduced by permission of the American Meteorological Society from Gash and Nobre, 1997)

air above the forest canopy, but the absolute humidity is slightly higher, giving a much higher relative humidity (Shuttleworth *et al.*, 1985). It is this high relative humidity, coupled with very low wind speeds, that gives the characteristically uncomfortable rainforest conditions. The large biomass of the forest acts to smooth out the daily temperature cycle (see Figure 1). Energy is stored in the biomass during the day and given out at night, so that daytime temperatures are lower over the forest than over deforested land, but at night the forest is warmer.

Evaporation and Sensible Heat Flux

The net radiation at the surface is used for both evaporation and sensible (convective) heating of the atmosphere. Agricultural crops and pasture have shallower rooting depths than forest and can rapidly deplete the store of soil moisture available to them during drought (Wright *et al.*, 1992). Evaporation is then reduced and more energy goes into sensible heat flux. Despite the greater net radiational energy available to the forest, low evaporation from water-stressed pasture can lead to the sensible heat flux from the pasture being greater than that from the forest (see Figure 1).

Regional Climate

At night the land surface is cooler than the atmosphere above and there is a stable boundary layer; but after dawn the surface heats up and convective turbulence develops resulting in a well-mixed boundary layer growing from the surface upwards. The greater the sensible heat flux from the surface, the higher is the top of the convective boundary layer, which may grow up to several kilometers in height. When the height is greater than the condensation level, water vapor in air rising from the surface condenses to form cumulus clouds. Observations of dry season boundary-layer growth taken in Rondonia, in Southwest Amazonia (Nobre *et al.*, 1996) have shown that the differences in sensible heat flux between a large area of forest and a large area of deforestation (more than 50 km across) can result in differences of boundary-layer height of between 700 to 1000 m. Satellite observations of the same area

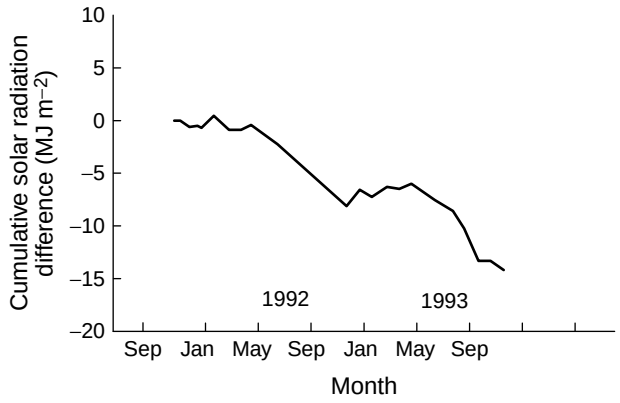


Figure 2 The observed difference in incoming solar radiation between pasture and forest plotted cumulatively for a pair of sites in Rondônia, in Southwest Amazonia. During the dry season (July, August and September) systematically less radiation is received at the pasture site than at the forest. (Reproduced by permission of the Centre for Ecology and Hydrology, redrawn from Culf *et al.*, 1996)

showed differences in cloudiness, with greater cloudiness being observed over the deeper boundary layers associated with the deforestation. This difference in cloudiness was confirmed by observations of systematically lower values of solar radiation reaching the surface shown in Figure 2 (Culf *et al.*, 1996). It is rare to find such convincing, direct evidence that changes in the land surface can affect climate at this regional scale.

Continental and Global Climate

Energy exchange between the land surface and the atmosphere is much greater in the tropics than in temperate latitudes, and it is largely this zonal difference which drives the global atmospheric circulation. Deforestation changes the amount of radiative energy absorbed, and changes the partitioning of that energy between evaporation and sensible heat flux. Just as large-scale changes in sea surface temperature (such as El Nino) can affect global weather patterns, so it might be expected that large-scale deforestation in the tropics could also change the global atmospheric circulation. However, no observational evidence of the effects of deforestation on climate at a continental scale has yet been found. Although there is a wealth of circumstantial evidence pointing to decreased rainfall, rigorous statistical analysis of rainfall records from 48 sites in Amazonia showed almost equal numbers of increasing and decreasing trends (Dias de Paiva and Clarke, 1995). An alternative to observations is provided by modeling studies performed with global circulation models (GCMs). These can be used to simulate the change in climate, which might result from deforestation on scales which have not yet occurred.

GCMs are essentially weather forecasting models, but instead of being used to forecast the weather a few days in advance, they are run for years ahead: first as a control with current conditions and then again with, for example, the land surface in the model having been changed from forest to pasture. Deforestation experiments have been one of the most common uses of GCMs and experiments have been carried out over the past two decades using increasingly realistic models. Most of these have modeled the complete deforestation of Amazonia (e.g., Lean *et al.*, 1996) and almost all have concluded that the deforestation would result in less rain falling over the region. Perhaps more seriously, the models also predict less rain over the adjacent savanna belt and the semi-arid region of Northeast Brazil, which are already prone to serious droughts. Only a few studies have modeled the complete deforestation of all the world s tropical rainforest, but the predicted effect of deforestation is almost always to reduce rainfall over the area that was previously forested (McGuffie *et al.*, 1995). The predicted changes over Amazonia are more than over Southeast Asia and Africa, because the climate in Amazonia is less subject to oceanic influence. Sea surface temperatures and weather patterns in middle and high latitudes have also been demonstrated to be sensitive to global tropical deforestation, but at present it is not possible to come to firm conclusions about the likely size or extent of the changes in global climate which might occur.

HYDROLOGY

One of the most immediate and often most damaging environmental impacts of deforestation is that which can occur to the local hydrology (see Bruijnzeel, 1996). However, the effects are variable, depending on the methods of deforestation used and on the use to which the land is subsequently put. Well-managed removal of forest on good soil, with suitable measures to protect the soil during and after the removal of the timber, can result in successful and sustainable reuse of the land (Critchley and Bruijnzeel, 1996). Unfortunately it is often the case that forest on inappropriate, fragile soils is cleared using techniques that result in long-term damage to the soil structure.

Soil Properties and Erosion

Removing the tree cover does not, in itself, result in greater soil erosion. The impact of raindrops falling on the ground is not made more damaging by removal of the tree canopy. Drops of intercepted rain dripping from the forest canopy reach their terminal velocity in the distance between the top of the canopy and the forest floor and arrive with the same potential to erode the soil as they have in the open. The soil erosion occurs when the protection of the litter, and dense root mat beneath it, is removed and the soil is then exposed to the impact of raindrops (*see* Soil Deterioration and Loss of Topsoil, Volume 3).

Forest soils are generally highly conductive to water flow; even clay soils which normally have low conductivity are full of large pores created by roots and animal activity. This structure may persist after deforestation, below the surface layer of soil, but compression of the soil by machines during extraction of timber, or later by grazing animals, leads to permanent compaction, which reduces the ability of the surface soil layer to conduct water (Tomasella and Hodnett, 1996). This decreased conductivity prevents water infiltrating into the soil and results in water running off the surface, with consequent erosion. Bulldozers and other heavy machinery are particularly damaging to the forest soils and, because there is never frost to break up the soil, this damage persists. Soil erosion is not only a loss of a natural resource for agriculture and forests, it also results in sedimentation of rivers and reservoirs downstream.

River Flows

The greater infiltration of water into forest soils introduces a time lag into the water pathway, so that deforestation will almost always increase the speed of hydrological response, i.e., river flows will increase more rapidly in response to storms, with a greater likelihood of flooding. The reduced evaporation from deforested land (see above) will result in more water being available for river flow. However, this may not always result in increased base flow during dry periods, as the higher runoff in response to storms can reduce the recharge to the groundwater to such an extent that base flow is reduced (see Bruijnzeel, 1996). In temperate latitudes, catchment monitoring has shown that, generally, river flows slowly return to their pre-deforestation patterns if the forest is allowed to re-grow; in the tropics this depends on the care taken to prevent soil compaction during extraction of the timber (*see* Forest Logging Systems in Tropical Countries: Differential Impacts, Volume 3). In general, the humid tropics are not short of water, but decreased base flow can impact water resources used for generating hydroelectricity and, on large rivers, can affect navigation.

At present, the effects of total deforestation on large river basins, such as the Amazon or Congo, cannot be predicted with any confidence. With no change in rainfall, deforestation is expected to result in increased river flow, but negative feedback from the surface to the climate (see above) is expected to reduce the rainfall, thereby reducing runoff. The sensitivity of the system to evaporation is demonstrated by a modeling analysis, which predicted that removing the forest from Amazonia would increase the river flow at Obidos (the last gauging station before the sea on the main channel of the River Amazon) by 21% (Costa and Foley, 1997). The Amazon and the Congo are the major sources of freshwater into the Atlantic Ocean and changes of this magnitude could result in significant impacts on the ocean circulation and therefore climate. The experiments to simulate what these effects might be have not yet been done.

THE CARBON BALANCE

Deforestation as a Source of Atmospheric Carbon Dioxide

It is estimated that some 50% of the world s biomass is stored in tropical forest. The above-ground biomass is about $150\,t\,C\,ha^{-1}$ (tones of carbon per hectare); there are few measurements of below-ground biomass, but it is probably in the range $35\ \ 50\,t\,C\,ha^{-1}$ (see Grace *et al.*, 1999). In comparison, the biomass of pasture or any agricultural crop will be only a few percent of this amount. When forest is burnt, much of the carbon previously stored in the biomass is converted to atmospheric carbon dioxide. Deforestation is second only to the combustion of fossil fuel in contributing to the increase of atmospheric carbon dioxide and the consequent global warming. Fossil fuel combustion and cement production, combined, are estimated to produce 5.5 Gt per year of carbon dioxide and the net emissions from changes in tropical land-use 1.6 Gt per year (Houghton *et al.*, 1996).

Tropical Forest as a Sink for Atmospheric Carbon Dioxide

During photosynthesis, vegetation takes in carbon dioxide and gives out oxygen: this has led to the tropical forests being described as the lungs of the world. However the description is simplistic, because the process of respiration, largely the decomposition of dead organic matter, gives out carbon dioxide, balancing most, sometimes perhaps more than, that taken in by photosynthesis. The net carbon dioxide budget (or net ecosystem production) is the balance between photosynthesis and respiration, and is a small difference between two large numbers.

The net carbon dioxide flux can be measured in the turbulent lower atmospheric boundary layer above the forest canopy, using the eddy correlation method. In this method, measurements of vertical wind speed are correlated with measurements of carbon dioxide concentration: when upward moving air is correlated with an increase in carbon dioxide concentration, there is an upward flux of the gas and *vice versa*. Measurements must be taken rapidly (at a frequency of greater than 20 Hz is best) to measure the flux transported by high frequency turbulence. The method has been used at two sites in Amazonia (Grace *et al.*, 1995; Mahli *et al.*, 1998) and both sites were found to be sinks for carbon dioxide. In central Amazonia, annual rates were estimated to be $30\,t\,C\,ha^{-1}\,year^{-1}$ of photosynthesis and $28.1\,t\,C\,ha^{-1}\,year^{-1}$ of respiration, making the forest a net

Table 2 The estimated net annual carbon fluxes between the atmosphere and undisturbed Amazonia, in response to interannual climate variability and increasing CO_2 concentration (SD is the standard deviation)[a]

Year	Net carbon uptake Pg C year^{-1}	Year	Net carbon uptake Pg C year^{-1}
1980	0.3	1988	0.1
1981	0.7	1989	0.3
1982	0.1	1990	0.3
1983	−0.1	1991	−0.1
1984	0.5	1992	−0.2
1985	0.2	1993	0.7
1986	0.4	1994	0.3
1987	−0.2	SD =	0.3

[a] Negative fluxes indicate Amazonia was a net source of CO_2. (Data from Tian *et al.*, 1998).

sink of 2.3 t C ha^{-1} year^{-1} (Grace *et al.*, 1999). However, the measurement of nighttime fluxes with this method is problematic and the result is sensitive to the treatment of nighttime data (Culf *et al.*, 1999). These figures may thus be overestimates, although analysis of tree growth data (Phillips *et al.*, 1998) also shows accumulation of carbon by tropical forest.

Net carbon dioxide flux is the small difference between photosynthesis and respiration, and because these fluxes must be estimated with quite different models, it is difficult to model the net carbon flux to any worthwhile absolute accuracy. However, models can be used to extrapolate measurements in time (as was done to derive the annual estimates above) or to look at the sensitivity of the carbon balance to change. Photosynthesis varies with climate whereas respiration varies primarily with the temperature of the biomass and soil. Photosynthesis is also expected to increase with ambient carbon dioxide concentration. Table 2 shows how this sensitivity translates into variability in estimated annual net carbon uptake. Using a terrestrial ecosystem model, the net carbon flux of the whole of the Amazonian rainforest is estimated to vary between being a sink of 0.7 Pg C year^{-1}, to being a source of 0.2 Pg C year^{-1} (Tian *et al.*, 1998). The latter years were associated with dry conditions following El Nino events.

ECONOMICS

Generally the market value of rainforest as a whole is low, and its benefits non-marketed and little understood. Conventional investment appraisal techniques are, therefore, more likely to result in development options being interpreted more favorably than a non-development alternative. For example, economic calculations may show that it is beneficial to flood vast areas of forest to build hydroelectric power stations, even though these power stations are very inefficient in terms of the flooded area needed to generate a given amount of electricity. Similarly if the land is producing nothing of commercial value as forest, an economic cost benefit analysis will show that it is beneficial to convert the land to agriculture, however inefficient that agriculture may be. The principles of Ecological Economics challenge the established notion that the value of the environment can be equated with market prices in this way and seek to quantify the value of environmental entities. An example would be ascertaining how much above market prices for timber people are prepared to pay from sustainably managed forest. If in future these values were to be included into cost benefit analyses, very different development alternatives may be produced.

Rainforest may also have increasing commercial value beyond that of the conventional products which it produces. The principle of paying for carbon emission established at the Kyoto conference on climate change and the possibility that rainforest may have economic value as a sink for carbon is causing governments to look again at their plans for deforestation and development (*see* Boreal Forest Carbon Flux and it s Role in the Implementation of the Kyoto Protocol Under a Warming Climate, Volume 4). Similarly, it is now recognized that the biodiversity of rainforest is of value commercially, as a bank of genetic plant and animal material, but also because it is now recognized that all species have an intrinsic existence value. In some countries strict laws are being passed to protect rainforest nations from so-called bio-piracy, the prospecting and export of plant material for commercial exploitation. New awareness by governments of the value of rainforest may change development plans, but it is also essential that the local population be consulted. The pressure to deforest will only be removed when it is more economically attractive for the local population to maintain the forest than to remove it.

DEFORESTATION AND DEVELOPMENT

The populations of tropical countries are growing rapidly and there is a continual need for new land: this creates the pressure to deforest (Jepma, 1995); often this pressure results in official government colonization schemes. Against this is the pressure from conservationists to protect the rainforest in its undisturbed state. Although at first sight it would appear that these objectives are incompatible, increasingly the conservation movement is recognizing that while it is essential to maintain large areas of forest reserves, it is neither practical nor desirable for all of the world s rainforest to be kept as a living museum. Equally, governments and development organizations are now recognizing that simply replacing forest with intensive agriculture is often not a sustainable option. The poor forest soils and the humid tropical climate result either in only short-lived settlements, or in sparsely populated

large ranches, which do little to eliminate poverty. The new thinking is that, rather than replacing forest, agriculture should be incorporated into the forest system and the land managed in an integrated way that protects the tropical environment rather than confronts it. The successful formula is not yet known, and it will no doubt vary from place to place. The challenge for science is to produce the new ideas and solutions. These solutions will need to cross the normal narrow disciplinary boundaries and will certainly need to take account not only of the variability in the physical environment of the tropical rainforest zone, but also differences in culture and human aspirations.

REFERENCES

Bruijnzeel, L A (1996) Predicting the Hydrological Impacts of Land Cover Transformation in the Humid Tropics: the Need for Integrated Research, in *Amazonian Deforestation and Climate*, eds J H C Gash, C A Nobre, J M Roberts, and R L Victoria, John Wiley & Sons, Chichester, 15 55.

Costa, H M and Foley, A (1997) Water Balance of the Amazon Basin: Dependence on Vegetation Cover and Canopy Conductance, *J. Geophys. Res.*, 102(D20), 23 973 23 989.

Critchley, W R S and Bruijnzeel, L A (1996) *Environmental Impacts of Converting Moist Tropical Forest to Agriculture and Plantations*, UNESCO, Paris.

Culf, A D, Esteves, J L, Marques-Filho, A O, and da Rocha, H R (1996) Radiation, Temperature and Humidity over Forest and Pasture in Amazonia, in *Amazonian Deforestation and Climate*, eds J H C Gash, C A Nobre, J M Roberts, and R L Victoria, John Wiley & Sons, Chichester.

Culf, A D, Fisch, G, Mahli, Y, Costa, R C, Nobre, A D, Marques-Filho, A O, Gash, J H C, and Grace, J (1999) Carbon Dioxide Measurements in the Nocturnal Boundary Layer over Amazonian Forest, *Hydrology and Earth System Science*, 3, 39 53.

Dias de Paiva, E M C and Clarke, R T (1995) Time Trends in Rainfall Records in Amazonia, *Bull. Am. Meteorol. Soc.*, 76, 2203 2209.

Gash, J H C and Shuttleworth, W J (1991) Tropical Deforestation: Albedo and the Surface Energy Balance, *Clim. Change*, 19, 123 133.

Gash, J H C and Nobre, C A (1997) Climate Efects of Amazonian Reforestation: Some Results from ABRACOS, *Bull. Am. Meteorol. Soc.*, 78, 823 830.

Grace, J, Lloyd, J, McIntyre, J, Miranda, A, Meir, P, Miranda, H, Moncrieff, J, Massheder, J, Wright, I, and Gash, J (1995) Fluxes of Carbon Dioxide and Water Vapour over an Undisturbed Tropical Forest in South-west Amazonia, *Global Change Biol.*, 1, 1 12.

Grace, J, Mahli, Y, Higuchi, N, and Meir, P (1999) Productivity and Carbon Fluxes of Tropical Rain Forests, in *Global Terrestrial Productivity*, eds H A Mooney and J Roy, Academic Press, San Diego, CA.

Hodnett, M G, Oyama, M D, Tomasella, J, and Marques-Filho, A O (1996a) Comparisons of Long-term Soil Water Storage Behavior Under Pasture and Forest in Three Areas of Amazonia, in *Amazonian Deforestation and Climate*, eds J H C Gash, C A Nobre, J M Roberts, and R L Victoria, John Wiley & Sons, Chichester.

Hodnett, M G, Tomasella, J, Marques-Filho, A O, and Oyama, M D (1996b) Deep Soil Water Uptake by Forest and Pasture in Central Amazonia: Predictions from Long-term Daily Rainfall Data using a Simple Water Balance Model, in *Amazonian Deforestation and Climate*, eds J H C Gash, C A Nobre, J M Roberts, and R L Victoria, John Wiley & Sons, Chichester.

Houghton, J T, Meira-Filho, L G, Callendar, B A, Harris, N, Kattenberg, A, and Maskell, K (1996) *Climate Change 1995, The Science of Climate Change*, Cambridge University Press, Cambridge.

Jepma, C J (1995) *Tropical Deforestation, A Socio-economic Approach*, Earthscan Publications, London.

Lean, J, Bunton, C B, Nobre, C A, and Rowntree, P R (1996) The Simulated Impact of Amazonian Deforestation on Climate using Measured ABRACOS Vegetation Characteristics, in *Amazonian Deforestation and Climate*, eds J H C Gash, C A Nobre, J M Roberts, and R L Victoria, John Wiley & Sons, Chichester.

Lloyd, C R, Gash, J H C, and Shuttleworth, W J (1988) The Measurement and Modeling of Rainfall Interception by Amazonian Rainforest, *Agric. For. Meteorol.*, 43, 277 294.

Mahli, Y, Nobre, A D, Grace, J, Kruijt, B, Pereira, M, Culf, A, and Scott, S (1998) Carbon Dioxide Transfer over a Central Amazonian Rain Forest, *J. Geophys. Res.*, 103(D24), 31 593 31 612.

McGuffie, K, Henderson-Sellers, A, Zhang, H, Durbridge, T B, and Pitman, A J (1995) Global Climate Sensitivity to Tropical Deforestation, *Global Planetary Change*, 10, 97 128.

Nepstad, D C, de Carvalho, C R, Davidson, E A, Jipp, P H, Lefebvre, P A, Negreiros, G H, da Silva, E D, Stone, T A, Trumbore, S E, and Vieira, S (1994) The Role of Deep Roots in the Hydrological and Carbon Cycles of Amazonian Forests and Pastures, *Nature*, 372, 666 669.

Nobre, C A, Fisch, G, da Rocha, H R, Lyra, R F daF, da Rocha, E P, da Costa, A C L, and Ubarana, V N (1996) Observations of the Atmospheric Boundary Layer in Rondonia, in *Amazonian Defor. Clim.*, eds J H C Gash, C A Nobre, J M Roberts, and R L Victoria, John Wiley & Sons, Chichester.

Phillips, O L, Malhi, Y, Higuchi, N, Laurance, W F, Nunez, P V, Vsquez, R M, Laurance, S G, Ferreira, L V, Stern, M, Brown, S, and Grace, J (1998) Changes in the Carbon Balance of Tropical Forests: Evidence from long-term Plots, *Science*, 282, 439 445.

Tian, H, Melillo, J M, Kicklighter, D W, McGuire, D, Helfrich, III, J V K, Moore, III, B, and Vorosmarty, C J (1998) Effect of Interannual Climate Variability on Carbon Storage on Amazonian Ecosystems, *Nature*, 396, 664 667.

Tomasella, J and Hodnett, M G (1996) Soil Hydraulic Properties and van Genuchten Parameters for an Oxisol under Pasture in Central Amazonia, in *Amazonian Deforestation and Climate*, eds J H C Gash, C A Nobre, J M Roberts, and R L Victoria, John Wiley & Sons, Chichester, 101 124.

Shuttleworth, W J, Gash, J H C, Lloyd, C R, Moore, C J, Roberts, J M, Marques-Filho, A deO, Fisch, G, Silva-Filho, V deP, Ribeiro-Mde, N G, Molion, L C B, de Sa, L D A, Nobre, J C,

Cabral, O M R, Patel, S R, and Carvalho de Moraes, J (1985) Daily Variations of Temperature and Humidity Within and above Amazonian Forest, *Weather*, 40, 101 108.

Wright, I R, Gash, J H C, da Rocha, H R, Shuttleworth, W J, Nobre, C A, Maitelli, G T, Zamparoni, C A G P, and Carvalho, P R A (1992) Dry Season Micrometeorology of Central Amazonian Ranchland, *Q. J. R. Meteorol. Soc.*, 118, 1083 1099.

Degraded Ecosystems, Restoration of

see Restoration, Ecosystem (**Volume 2**)

Demographic Change: Indonesian Transmigration

Joan Hardjono
Padjadjaran State University, Bandung, Indonesia

The term transmigration refers to government planned and subsidized resettlement of people from Java's more overcrowded rural areas into less densely populated parts of Indonesia. The program, under which some 3.5 million people were resettled between 1950 and 1990, has been responsible for signi cant expansion of the cultivated area in the major islands of the country other than Java.

Transmigration is essentially an exercise in the utilization of marginal land to reduce population pressure in Java (Table 1). The extent of the environmental impacts of the program is difficult to assess, however, for forest conversion by local small-holders and spontaneous settlers and, particularly, rapid expansion in commercial logging account for much of the environmental transformation that has occurred. For many years the planning of settlements was handicapped by a serious misunderstanding of the physical nature of land resources outside Java and by the assumption that luxuriant natural vegetation was a reflection of fertile soils. Many transmigrant families have been unable to earn an adequate livelihood from agriculture because most of the arable farming systems introduced to provide the economic foundation of settlements have been unsuited to local conditions.

Table 1 Population growth in Java, 1930–2000 (Central Bureau of Statistics, Jakarta)

Year	$\times 10^3$
1930	41 718
1961	62 993
1971	76 102
1980	91 270
1985	99 853
1990	107 574
1995	114 733
2000	123 625[a]

[a] Projection.

The first resettlement schemes, established in southern Sumatra prior to 1940, were based on attempts to introduce irrigated rice growing using the methods developed in Java over many centuries (Pelzer, 1945). While this agricultural system in itself had no negative effects on the environment, the agricultural involution characteristic of intensive rice-growing areas in Java (Geertz, 1963) gradually appeared. At the same time the clearing of land upstream of irrigation infrastructure by loggers, local small-holders and the descendants of the original occupiers of the new irrigation areas, who had been forced by the fragmentation of inherited holdings to seek farm land elsewhere, resulted in the silting up of rivers and canals. This in turn reduced the dry season effectiveness of irrigation networks and led to the economic deterioration of these early settlements.

As conventional irrigation systems are expensive to construct, transmigration policy shifted toward the establishment of tidally irrigated settlements in the coastal swamps of southern and western Kalimantan and eastern Sumatra (Hardjono, 1977). The concept was adopted from the agricultural practices of local cultivators, who had long ago devised a way of using the tidal influence in coastal rivers to lift water on to rice fields and to provide drainage. Very few of these settlements have proved economically viable because planners did not understand that the tidal irrigation system is really a form of shifting cultivation. Local farmers plant coconuts in conjunction with rice and then, as soil fertility declines, make new clearings but return regularly to harvest the coconuts. The traditional system has no adverse effects on the environment since the original vegetation is replaced by perennials. Transmigrants, however, are expected to become sedentary farmers on 2-ha holdings. Given acidic soils derived from peat, they can grow no more than one rice crop a year, nor can they cultivate the secondary food crops like corn that they need for subsistence. The consequence is that much of the land cleared for tidal farming has been abandoned to the growth of weeds and grasses.

By far the greater proportion of government transmigration projects have been based on non-irrigated farming (Hardjono, 1986). The land made available for settlements has for the most part been under poor-grade secondary forest or scrub regrowth and in many instances was already degraded by earlier human activity. Settlements have sometimes been established on logged-over forestry concessions or on land where a shortening in shifting cultivation cycles had led to soil deterioration. In some areas transmigrants have been placed on heavily dissected terrain where erosion is well advanced and in others on land covered in alang-alang grass (*Imperata cylindrica*). Soils have mostly been inherently infertile red yellow podsols, whose shallow upper layers have tended to be destroyed by the use of heavy machinery to prepare settlement sites. Inadequate water resources in many areas have exacerbated these problems. In some places settlers have succeeded in bringing this land under cultivation, thus preventing further degradation. But this has been possible only where subsidized inputs like fertilizers and lime were provided or where project planning included finance for the planting of cash crops like rubber.

The general response of government sponsored transmigrants to the low returns obtained from farming has been to seek seasonal off-farm wage labor or to abandon their holdings completely. The environmental consequences have been two-fold: land degradation has accelerated as neglected fields become covered in alang-alang grass, and encroachment on forested land outside the project area has occurred. In the latter case transmigrants have turned to the felling of trees to produce charcoal commercially, to obtain timber for sale to local sawmills or to make clearings in which to plant cash crops like coffee and pepper in imitation of local small-holders.

Since 1990 government transmigration targets have been considerably reduced because of budgetary constraints and also because of recognition of the fact that arable farming is not an appropriate type of land use (Table 2). Other agricultural patterns have been introduced in the form of privately owned oil-palm plantations and timber estates, which are connected with transmigration only in so far as they depend on transmigrant families from rural Java for their labor force.

In its Indonesian interpretation, transmigration does not include spontaneous migrants who move without government assistance from one part of the country to seek a living in another. Some come from elsewhere in the same province or island while others migrate from Java and Bali. They purchase usage rights to land from local people but, although they adopt the slash and burn techniques of shifting cultivators, they do not plant perennials when fertility declines. As they move on and make new clearings, abandoned fields soon become covered in secondary growth. In the absence of appropriate agrarian legislation, it is difficult to restrict the movement of these settlers, who are responsible for much of the uncontrolled land clearing that has damaged the environment outside logging concessions in recent years.

Table 2 Fully and partly sponsored transmigrant families, 1950–1997 (Department of Transmigration, Jakarta)[a]

Period	No. of families[b]
1950–1954	22 317
1955–1959	32 242
1960–1964	26 456
1965–March 1969	21 633
1969–March 1974	39 511
1974–March 1979	52 084
1979–March 1984	365 977
1984–March 1989	228 422
1989–March 1994	77 485
1994–March 1997	74 565[c]

[a] Figures include local migrants moving within the same province.
[b] Each family has an average of five members.
[c] Preliminary figure.

REFERENCES

Central Bureau of Statistics, Republic of Indonesia, Jakarta.
Department of Transmigration, Republic of Indonesia, Jakarta.
Geertz, C (1963) *Agricultural Involution: The Process of Ecological Change in Indonesia*, University of California Press, Berkeley, CA.
Hardjono, J (1977) *Transmigration in Indonesia*, Oxford University Press, Kuala Lumpur.
Hardjono, J (1986) Transmigration: Looking to the Future, *Bull. Indones. Econ. Stud.*, XXII(2), 28 53.
Pelzer, K J (1945) *Pioneer Settlement in the Asiatic Tropics*, Institute of Pacific Relations, New York.

Demographic Change: Peopling of the Pacific Islands

Patrick D Nunn
The University of the South Pacific, Suva, Fiji

The initial peopling of the Paci c Islands 3500- 800 years ago was a remarkable achievement, one which began tens of generations before any other group of humans developed the

ability to cross such large expanses of open ocean. The linkages between the rst Paci c Islanders and the environments they settled continues to be vigorously debated. Many authors have highlighted the effects of the rst humans on the pristine, vulnerable environments of Paci c Islands, and virtually overlooked the ways in which these environments (and the ways in which they changed) affected human lifestyles. It is now becoming clear that the lifestyles of the rst Paci c Islanders were in uenced considerably by the new environments they encountered, and that subsequent environmental change was the major cause of cultural transformation among many Paci c Island peoples.

OUT OF THE WEST

The first people to colonize the Pacific Islands came from the western Pacific Rim (see Figure 1). The clearest evidence is linguistic; the root of the Austronesian languages, which are the native tongues of most Pacific Islanders, lies in southern China and Taiwan. It is reasonable to suppose that this region was occupied around 10 000 7000 years ago by people practicing fledgling agriculture. During this period, this region was also more extensive because the sea level was lower, the island of Taiwan being connected to the Asian mainland, and the mouths of the Huanghe and Yangtze rivers being far seaward of their present positions (Nunn, 1999).

It may have been sea-level rise and the consequent inundation of many thousands of square kilometers of low-lying arable land that led to the initial exodus of people from this area around 7000 6000 years ago. They took to the oceans and, making their way south through the Philippines archipelago, finally settled the outer islands of Papua New Guinea around 3500 years ago (Kirch, 1997).

There was already a long-established population who spoke the Papuan language (which also include Australian aboriginals) in Papua New Guinea at this time, and it may be that the Austronesian-speaking emigrants could find space to dwell only in the comparatively sparsely-peopled outliers of this vast island group. Although some of the new arrivals remained in Papua New Guinea, others moved on eastward into the Pacific Ocean.

THE LAPITA COLONIZATION

The journeys of the first Pacific Islanders have been tracked precisely by the remains of the distinctively-decorated pots they manufactured. This style of pottery, named *Lapita* after the site in New Caledonia where it was first discovered, occurs from the outer islands of Papua New Guinea through the Solomon Islands, Vanuatu, Fiji, Tonga and Samoa (Kirch, 1997).

Apparently, within a few hundred years around 3000 years ago, Lapita people colonized, for the first time the eastern outer Solomon Islands, Vanuatu, New Caledonia, Fiji, Tonga and Samoa, routinely crossing ocean gaps of several hundred kilometers on vessels loaded with people, pigs, dogs, corms of taro, coconuts and other items intended to recreate their ideal environments on the islands they encountered. This idea of transported landscapes has proved a useful key to understanding many of the similarities in early human lifestyles across tropical Pacific Islands (Kirch, 1997).

It is a remarkable coincidence that the first permanent human settlement of most Pacific Islands occurred only after sea level had begun falling from its postglacial maximum, a time when the cliffed, reefless coasts of these islands would have become replaced by coasts bordered by broad, fertile plains fringed with coral reefs. Indeed it may be that early humans in the voyaging nursery of western Melanesia had paid many visits to island groups a thousand kilometers or more away to the east long before this time but, finding the coasts of those islands unsuitable for permanent colonization, had returned to their home bases in the west to await a time when the coastal environments of those distant islands became more attractive (Nunn, 1994).

INTO THE EAST

The Lapita cultural complex extends no farther east into the Pacific than Tonga and Samoa, perhaps because enough land had been settled to satisfy all the potential colonizers associated with this wave of migration. There seems to have been a pause of perhaps 1000 years in the eastward colonization of the tropical Pacific at this time. Although there is some debate as to when and where eastward colonization resumed, a plausible scenario is that it happened between Samoa and the Marquesas Islands in northeast French Polynesia about 2000 years ago. A conspicuous feature of this phase of Pacific Island colonization is the absence of ceramic use or manufacture, suggesting it may have been accidental rather than deliberate. It is possible that a fishing party out from Samoa was caught in the Equatorial Countercurrent, which is often both stronger and farther south than normal during El Nino events (*see* El Nino and La Nina: Causes and Global Consequences, Volume 1), and swept eastward to a landfall in the Marquesas. From here, the colonization of the rest of French Polynesia and adjacent island groups appears to have been accomplished fairly rapidly. It has been suggested that the clear skies and reliable trade winds associated with the Little Climatic Optimum (c. 1100 700 years ago in the tropical Pacific) facilitated the long-distance voyages from this Polynesian heartland, which saw the first human settlement of Easter Island (Rapanui), the Hawaiian Islands, New Zealand and possibly even Panama in Central America, several generations before European sailors discovered this vast ocean.

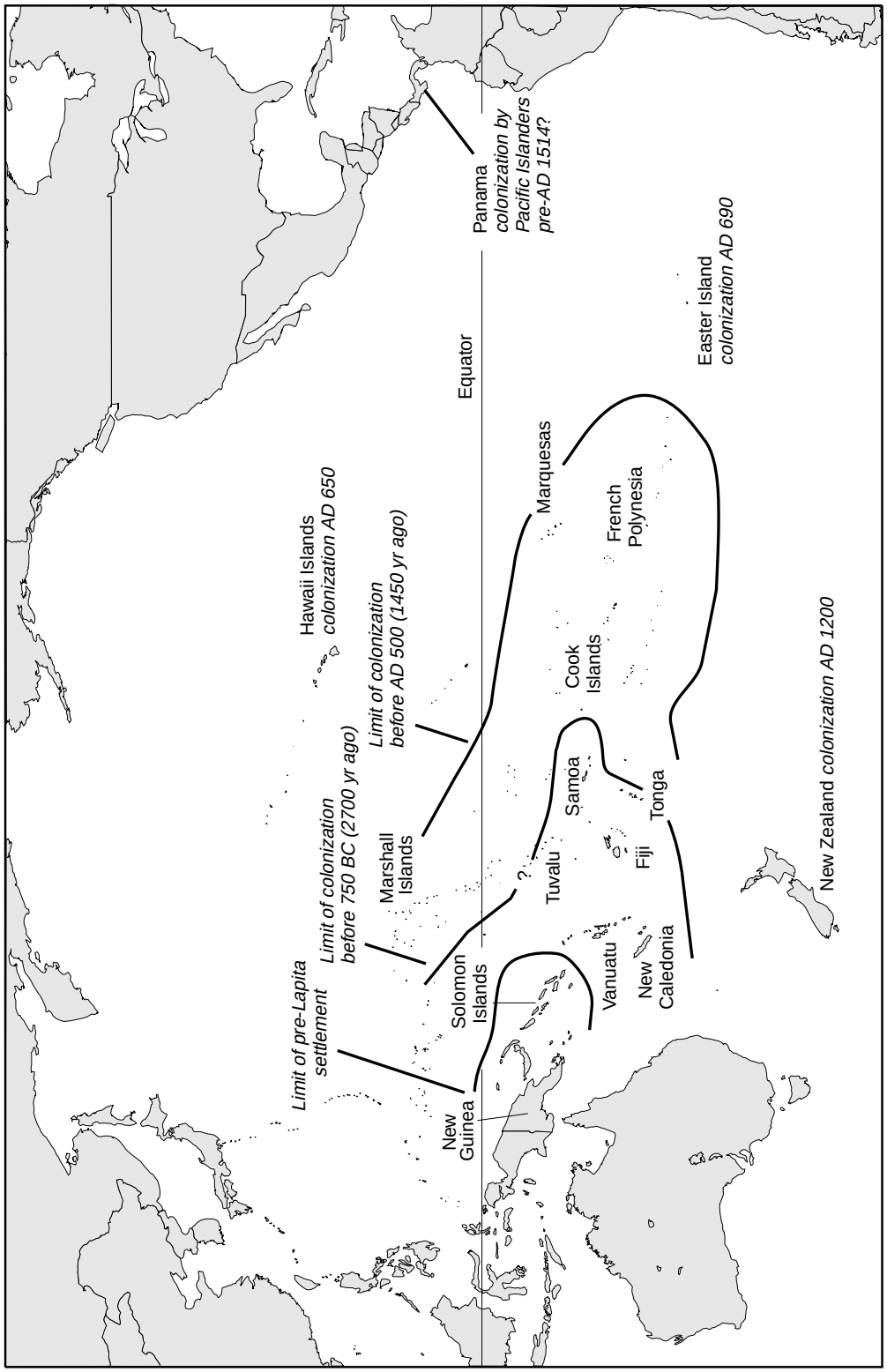


Figure 1 The main Pacific Island groups, showing the chronology of colonization outwards from western Melanesia

EARLY HUMAN–ENVIRONMENT RELATIONS

There is considerable debate about early human environment interactions on Pacific Islands, some authorities claiming that island ecosystems experienced an abrupt and catastrophic change shortly after humans arrived, others inferring that human impact was slower and its effects are obscured by those of non-human environmental changes, particularly those resulting from climate change (Kirch and Hunt, 1997; Nunn 1997, 1999).

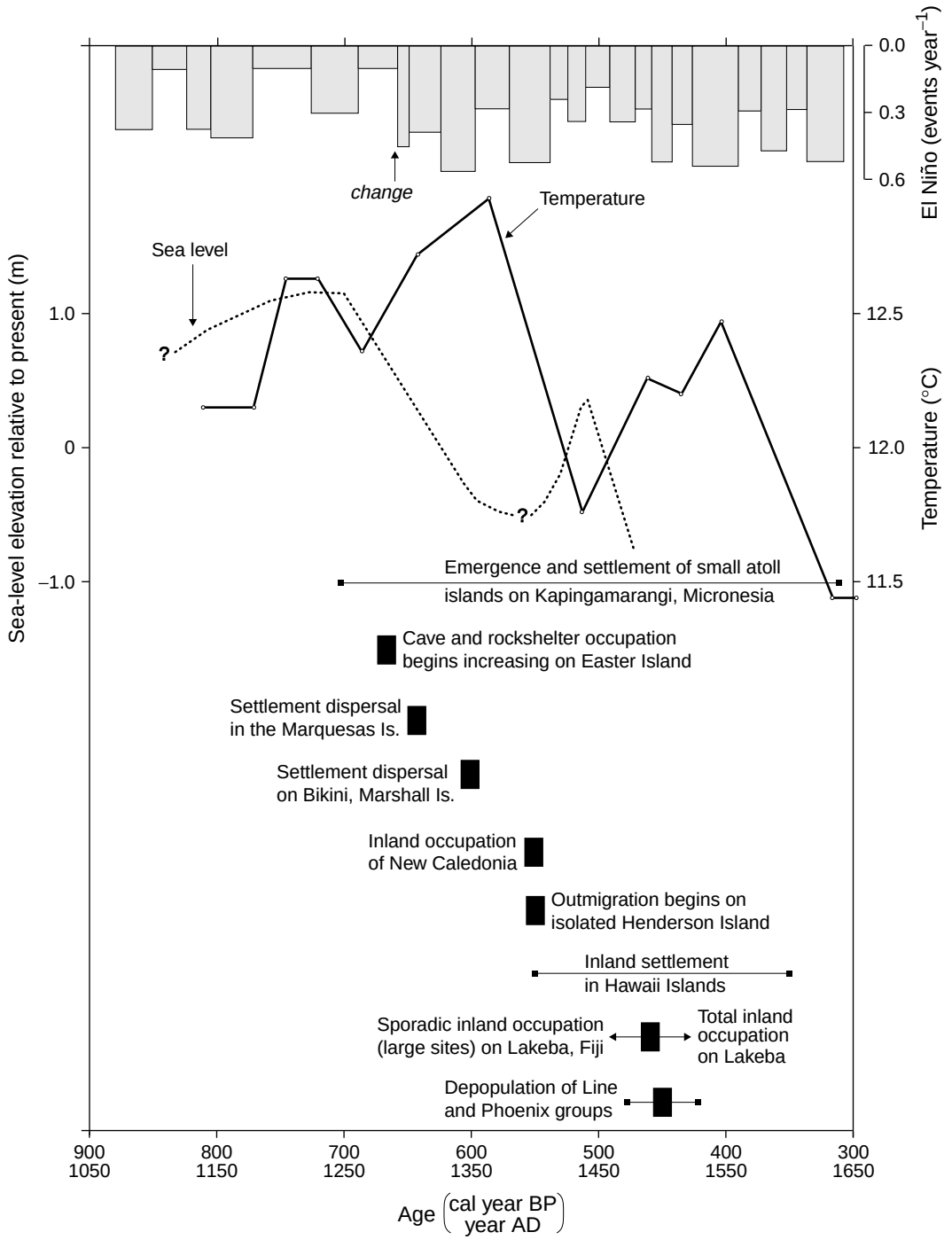


Figure 2 Environmental change and human response manifested by settlement-pattern changes on selected Pacific Islands during the last millennium. All these changes can be interpreted as a consequence of the substantial depletion of the food resource base on many islands associated with the rapid cooling and sea-level fall (and increased precipitation) coincident with the increase in El Niño frequency around AD 1300

It is clear that at the time around 3000 2500 years ago when the Lapita colonization in the southwest Pacific Islands was underway, the climate in the tropical Pacific was becoming cooler and drier, effects that would be expected to have stressed many ecosystems, increasing their vulnerability to rapid change as a result of human arrival. It is possible that much of what has been regarded as a consequence of early human impact is actually (largely) a result of non-human changes, the corollary to which is that, as studies of traditional horticulture in the Pacific Islands suggest, their early inhabitants husbanded rather than recklessly overused the resources available on a newly-discovered island.

The same may not have been true of the central and eastern tropical Pacific Islands on many of which there is evidence that the earliest inhabitants simply moved on to the next uninhabited island once all the birds and bird eggs were gone, as a result of human predation.

Despite the evidence of transported landscapes in different island groups, it is possible to overstress the dominance of human impact at the expense of non-human impacts. This point shows up less well with the low densities of people that settled the Pacific Islands initially but becomes more apparent later.

During the Little Climatic Optimum (1100 700 years ago), for instance, the climate became drier than previously, resulting in the development of water-conservation strategies, primarily agricultural terracing, on many islands. Such strategies required cooperation and, thus, at this time, small isolated settlements were replaced by large nucleated settlements, and the subsequent development of socio-political complexity (Nunn, 2000).

At the end of the Little Climatic Optimum, there was a rapid cooling and sea-level fall, associated with a short-lived rise in precipitation attributable to increased El Nino frequency (Figure 2). This AD 1300 event brought about widespread destruction of the food resource base on many Pacific Islands, resulting in social breakdown that was manifested by the abandonment of large coastal settlements in favor of small, fortified hilltop settlements, and the outbreak of inter-tribal warfare and the associated cannibalism and headhunting, which many writers have supposed to have been endemic to Pacific Island societies (Figure 2). During the subsequent Little Ice Age (600 150 years ago in the tropical Pacific), conditions remained cool but environmental stability resumed and, although people on many islands remained in their hilltop settlements, inter-tribal conflict effectively ended by the time of European settlement around AD 1850 (Nunn, 2000).

Since this time, increased demands on Pacific Island environments have led to many becoming notably less productive, although it remains difficult in many cases to separate human from non-human impacts. The recent warming and increase in tropical cyclone (hurricane) frequency in the South Pacific has caused problems of a magnitude that outweigh deleterious human impacts on many islands (Nunn 1994, 1997).

REFERENCES

Kirch, P V (1997) *The Lapita Peoples*, Blackwell, Oxford.

Kirch, P V and Hunt, T L, eds (1997) *Historical Ecology in the Paci c Islands*, Yale University Press, New Haven, CT.

Nunn, P D (1994) *Oceanic Islands*, Blackwell, Oxford.

Nunn, P D (1997) Keimami sa Vakila na Liga ni Kalou *(Feeling the Hand of God): Human and Nonhuman Impacts on Paci c Island Environments*, 3rd revised edition, Suva, Fiji, School of Social and Economic Development, The University of the South Pacific.

Nunn, P D (1999) *Environmental Change in the Paci c Basin: Chronologies, Causes, Consequences*, John Wiley & Sons, London.

Nunn, P D (2000) Environmental Catastrophe in the Pacific Islands Around AD 1300, *Geoarchaeology*, 15, 715 740.

Demographic Change: the Aging Population

Donald T Rowland

The Australian National University, Canberra, Australia

Population aging is one of the major processes of change in contemporary societies. In the more developed countries, 8% of the population were aged 65 years and over in 1950, compared with 14% in 2000 and a possible gure of 24% in 2050. Less developed countries have lower proportions in the older ages, yet they have the majority of the world's older people. China and India together have greater numbers aged 65 years and over than the whole of Europe, including the former Soviet Union. Textbooks de ne population aging as a rise in the percentages in older ages, but the increase in the numbers of the aged is closely related and important. Even where there are low percentages of older people, families, communities and governments may not have the means to provide adequately for them.

Growth in the numbers of the aged is an unavoidable consequence of falling death rates, while growth in the percentages in older ages is an unavoidable consequence of falling birth rates (fertility). Low mortality and fertility are necessary in a nite world where human welfare is paramount and human numbers cannot expand inde nitely.

Yet world population growth will remain unstoppable until population aging has run its course. Even with low birth rates established, national populations can grow by 50% or more through the expansion of older age groups. Such growth brings with it economic, social and environmental strains requiring major adjustments to an unprecedented situation.

THE PROCESS OF POPULATION AGING

Population aging necessarily accompanies progress towards achieving sustainable development through reducing rates of population growth. All national populations will eventually be old if low birth and death rates become universal. The main phase of growth in the numbers and percentages of older people began in Europe and North America in the 19th century and in Asia, Africa and Latin America from around 1950. Such changes are part of the *demographic transition*, which describes population trends through time and provides a framework for comparing the experience of different countries. The demographic transition identifies progress from a pre-transition stage, with high birth and death rates, to a post-transition stage with low birth and death rates (Table 1). Between them is the demographic transition, during which population growth rises then declines and population aging occurs (Myers, 1990).

Over the demographic transition, national and world age structures evolve from *young*, triangular profiles to *old* rectangular ones. The former indicates high birth rates, the latter low. Figure 1 illustrates these changes, showing the global population in 2000 (young) and the projected situation in 2100 (old). It takes many decades for the declines in birth and death rates to have their full effects on the numbers and percentages in each age group. Thus, growth in the numbers in older ages is typically a delayed result of developments that occurred sixty or more years earlier: the transition in the age structure continues long after completion of the major declines in birth and death rates.

The numbers in older age groups expand throughout the demographic transition because of the long-run effects of lower death rates among the young. The greatest improvements in survival over time occur among infants and children. In the pre-transition stage, up to 50% of infants die before their fifth birthday. Early in the transition falling mortality, due especially to control of infectious diseases and famines, results in more infants and children surviving to maturity and ultimately to the older ages. By the middle of the demographic transition about 80% of children born reach age five and 40% reach age 65 (Table 1). If high birth rates persist, populations remain in the transition

Table 1 Characteristics of populations during and after the demographic transition

	Pre-transition	Mid-transition	Post-transition	Future declining[a]
Crude birth rate[b]	50.0	45.7	12.9	9.8
Crude death rate[c]	50.0	15.7	12.9	14.8
Annual growth rate %	0.0	3.0	0.0	−0.5
Age structure %				
0–14	36.2	45.4	19.2	15.6
15–64	60.9	52.0	62.3	52.7
65+	2.9	2.6	18.5	31.7
Total	100.0	100.0	100.0	100.0
Dependency ratios				
Child[d]	59.0	87.0	31.0	29.6
Aged[e]	5.0	5.0	30.0	60.0
Total	64.0	92.0	61.0	89.6
Percentage surviving (females)				
To age 5	46.8	81.7	98.2	99.6
To age 65	7.8	43.3	83.1	94.2
Life expectancy (females)				
At birth	20.0	50.0	75.0	85.0
At age 5	36.6	55.9	71.4	80.3
At age 65	7.5	11.9	15.7	22.2

[a] Data for Italy, rates refer to 2025–2050, other data to 2050.
[b] Crude birth rate: births/population × 1000.
[c] Crude death rate: deaths/population × 1000.
[d] Child dependency ratio: 0–14/15–64 × 100.
[e] Aged dependency ratio: 65+/15–64 × 100.
Source: Hauser (1976: 66); World Bank (1994: 281); Coale and Demeny (1983); Coale and Guo (1990: 33).

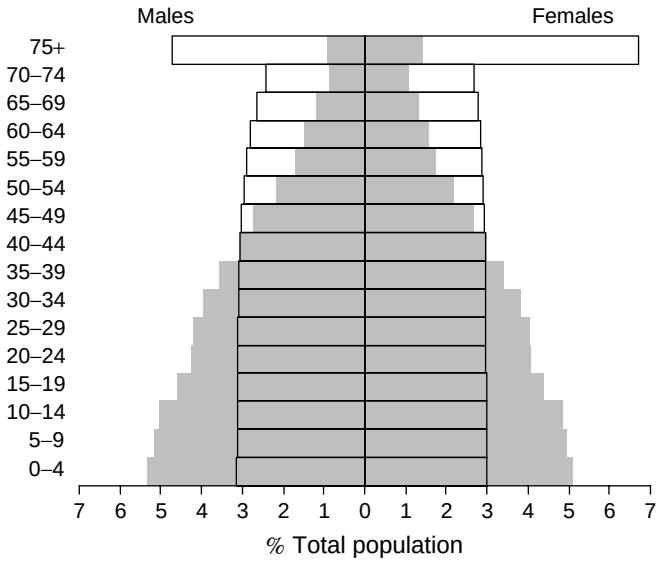


Figure 1 Age structure of the world's population, 2000 (shaded) and 2100. (*Source:* World Bank (1994))

stage and stay demographically young; each newborn group or cohort of children will be larger than the one before it. This situation creates substantial population momentum, or growth potential, arising from larger birth cohorts reaching older ages.

As a result of population momentum, the demographic transition brings spectacular growth in the numbers of the aged. For example, in France, the ultimate outcome of the demographic transition will be a doubling of the total population and a 10-fold increase in the numbers aged 65 and over. In India, a five-fold increase in the total will accompany a 40-fold increase in the aged population. In Mexico and Kenya, the numbers of the aged could increase 100-fold and 200-fold, respectively, by the end of the transition (Chesnais, 1990). The highest figures reflect rapid mortality decline in conjunction with delayed fertility decline.

Whereas improvements in survival sustain cohort numbers for longer and increase population momentum, falling birth rates have the opposite effect: they decrease cohort size and reduce momentum. The percentages in older age groups remain low during the demographic transition until birth rates start to decline. Falling birth rates reduce differences between the size of parent and child generations, as families averaging six or more children give way to families averaging two children. Lower fertility causes young triangular age structures to become narrower at the base and more rectangular (Figure 1). The decrease in the representation of children allows the percentages in older ages to increase. The fall in the birth rate is the principal cause of the expansion of the percentages in older ages from less than 3% (pre-transition) to around 20% (post-transition).

REJUVENATION AND EXCESS AGING

Although the demographic transition envisages unidirectional change towards lower birth rates, reversals in the process of aging can happen. One destined to have a substantial impact in the future is the baby boom that occurred after the Second World War in more developed countries, and was particularly protracted in some such as the United States and Australia. The baby boom was a temporary increase in birth rates arising especially from more universal marriage and childbearing. For a time, the baby boom rejuvenated populations, reversing national trends towards population aging. It also created unusually large cohorts, which will reach the older ages in the second and third decades of the 21st century. In the countries most affected, much forward planning of provision for the aged is to prepare for the aging of the baby boom cohorts.

Other exceptions arise because the post-transition stage of the demographic transition is merely hypothetical: it is not the endpoint of population change. The experience of European countries has been to move to a further stage of below-replacement fertility (van de Kaa, 1997). This produces natural decrease (an excess of deaths over births) and hyper-aging with excess proportions in older ages: the age structures become narrower at the base than in the middle. Examples in the early 21st century could include Italy, Spain, Germany and Japan. Since successive cohorts are smaller than those that went before, such populations develop negative momentum, whereby numbers decrease as smaller cohorts reach older ages (Rowland, 1996). Maintenance of such a situation would ultimately bring extinction. Only a rise in the birth rate could reverse this extreme form of aging, illustrated in Table 1 under the heading future declining population.

International migration is sometimes seen as a means of averting decline and rejuvenating populations, but it is seldom an appropriate response to population aging *per se*. The most rapidly aging national populations are so large that huge and socially disruptive numbers of immigrants would be required to make an appreciable difference to the percentages 65 and over. Also, in the long run, the aging of the immigrants themselves would augment the total numbers in the older ages.

REGIONAL VARIATIONS

Future numbers and percentages in the older ages are necessarily uncertain, but World Bank (1994) projections illustrate possible developments. In the year 2000, about 7% of the world s population of 6.1 billion were aged 65 or more (Table 2). This percentage might double by 2050 (world population 9.6 billion) and treble by 2100 (world population 11.0 billion). During the 21st century,

Table 2 Population aged 65 and over, World Regions, 1990–2100

	1990	2000	2050	2100
Numbers 65+ (millions)				
World	330	419	1,435	2406
More developed regions	147	176	324	335
Less developed regions	182	244	1111	2074
Africa	19	25	149	515
Asia	160	217	877	1408
Northern America	35	39	86	95
Latin America	21	27	137	205
Europe	92	108	178	172
Oceania	2	3	8	10
Percentages 65+				
World	6.3	6.9	15.0	22.0
More developed regions	12.1	13.8	23.7	24.3
Less developed regions	4.5	5.0	13.5	21.6
Africa	3.0	3.1	7.5	19.5
Asia	5.1	5.9	15.6	22.4
Northern America	12.4	12.5	22.9	24.8
Latin America	4.8	5.2	17.0	23.2
Europe	12.7	14.7	24.7	24.1
Oceania	8.9	9.3	19.1	23.3

Source: World Bank (1994).

the age structures of different regions of the world will probably converge toward the demographically old profile expected at the end of the demographic transition. The more developed regions of the world would still have the oldest age structures, but the contrast with the less developed regions would be smaller than in 2000. Due to sustained high birth rates, Africa may be the last major region to complete the transition, with still only 7% of its total 65 plus in 2050. Europe s figure of nearly 25% 65 and over in 2050 reflects the outcome of below-replacement fertility in many countries, as assumed in the projections, leading to negative momentum and population decline. Italy and Japan, with projected figures of around 30% aged 65 and over in 2050 exemplify hyper-aging. Sweden, had the world s oldest national population in the early 1990s; 18% 65 and over (Kinsella and Taeuber, 1993) but it may reach only 22% aged in 2050, provided that its birth rate remains near replacement level (World Bank, 1994).

The World Bank s projections of the total numbers aged 65 and over show the world s total aged population more than trebling between 2000 and 2050, reaching 1.4 billion in the latter year and 2.4 billion in 2100. In 1990, the aged in the demographically younger less developed regions already outnumbered those in the more developed regions. By 2050, 77% of the world s aged may be in the currently less developed regions, rising to 86% in 2100 (Table 2). The demographic giants of China and India, each with total populations of more than a billion, dominate the global picture. By 2050, China and India could *each* have far more old people than Europe and the United States combined. In India, the population aged 65 and over numbered 11 million in 1950 (3.3% of the total population), around 52 million in 2000 (5.2%) and could reach 237 million in 2050 (14.6%).

CONSEQUENCES

Population aging has implications for resource use and environmental change mainly because of its role in overall population growth and the transformation of national age structures. The effects of these demographic developments are mediated through other factors, such as technology, material living standards and international trade. Thus, similar demographic changes can have different environmental consequences, depending on the nature of the society.

Aging of the Labor Force

Shifts in the age structure can affect the productivity of populations through changes in dependency levels and the aging of the labor force, which everywhere are concomitants of overall population aging. A dilemma for less developed countries is whether to maintain the labor force participation of older workers in the modern sectors of the economy to the disadvantage of the young, better educated and potentially more productive; or whether to foster early retirement and prolong the time lived as a potential tax burden. The productivity of workers, however, is far less a function of age than of skills and health status. The principal distinction is between the aged with high functional capacity and those with disabilities. Experience and

lower rates of absenteeism can reinforce productivity in an older labor force. The aged who are not in the mainstream labor force further constitute a valuable resource for society in helping relatives and engaging in voluntary work. The Chinese government has been particularly active in capitalizing on the aged s status, expertise, and free time, including using *granny police* to monitor use of family planning and adherence to the country s one-child policy (Treas and Logue, 1986).

In more developed countries, voluntary and involuntary early retirement are restraining the aging of the labor force. During the 20th century, retirement became established as a stage of later life, as well as one commencing at younger ages. In the United States in 1890, 89% of males aged 55 64 were in the labor force, compared with 68% at ages 65 and over (Davis and van den Oever, 1981). By 1990, these figures had fallen to 68 and 16%, respectively, but changed relatively little in subsequent years to 1997 (US Bureau of the Census, 1998). Early retirement entails not only a shortening of labor force participation, but usually also a prolonging of reliance on publicly funded pensions. In the future, males may spend more than half their lives outside the labor force. The higher participation of women in the labor force has offset much of the lower participation of men (Davis and van den Oever, 1981).

Dependency

The costs associated with providing for the health and welfare needs of larger numbers of the aged raise pressures on governments to increase economic growth, which may heighten conflicts with environmental goals. At the same time, a rising proportion of voters will be above middle age, reinforcing imperatives for policies and programs supportive of the needs of older people. Thus, population aging has implications for the environment through the need to give priority to supporting greater numbers at extra *per capita* costs to the State, in a time requiring critical decisions concerning sustainable use of resources and the environment.

Yet currently, in less developed countries with scarce resources, aging is seldom regarded as a high priority amid the many other demands of social and economic development (Rowland, 1994). There is a strong expectation that the family will provide for older relatives, but this is becoming less tenable in the face of the rising labor force participation of women and the separation of relatives through inter-regional migration (Myers, 1995). In the past, only a small proportion of families had surviving grandparents, which meant that traditional family support for the frail aged was called upon less frequently. Today, the pace of aging in less developed countries is well ahead of institutional preparedness to meet health and social welfare needs (Treas and Logue, 1986).

Demographic dependency ratios compare the numbers of children and older people with the numbers in the working ages. Such ratios are only crude indicators of economic demands, since they take no account that some in the working ages are not economically active while others in the so called dependent ages may be independent in economic terms as well as in other ways. Statistics in Table 1 show that child dependency and total dependency peak in the middle of the demographic transition, at which time there may be around 90 dependants, mainly children, per hundred people in the working ages, 15 64 years. By the end of the transition, total dependency actually falls to around 60 workers per hundred dependants, of whom half are under 15 and half are 65 or more. Thus old, post-transition populations have relatively low levels of dependency. Table 1 shows that a future declining population, exemplified by the World Bank (World Bank, 1994) projection for Italy in 2050, could again have an overall dependency level of 90, two-thirds of whom would be 65 plus. A key question therefore concerns the relative cost of supporting children compared with older people.

Per capita public expenditure on older dependants in more developed countries is said to be two or three times higher than on children. Nevertheless, other estimates suggest that a child absorbs more resources than an old person, with the result that the first 20 years of life cost more than the total years lived after 60. Thus, combining public and private expenditures, children appear more expensive *per capita*. This implies that the decreased cost of supporting children is offsetting the increased cost of supporting the aged. Much of the decreased cost of children, however, will be a saving in private expenditure, which the taxation system cannot easily redirect to the aged (Easterlin, 1996).

Technology and rising standards of education may facilitate adjustments to population aging in more developed countries, as they have to other demographic changes in the past (Easterlin, 1996). Yet continuing economic growth is also likely to be a prerequisite for meeting the needs of greater numbers of older people without compromising living standards and standards of care. Additional responses will be targeting government expenditure on the most needy and encouraging greater self-provision in later life. Such responses are evident in the strategy for aging populations that the Organisation for Economic Co-operation and Development (OECD) (OECD, 1999) has formulated, elements of which are:

- removing financial incentives for early retirement;
- improving job opportunities for older workers, including part time employment;
- reducing public pension benefits and developing advance funded pensions systems;
- achieving a greater focus on cost effectiveness in health and long-term care.

FUTURE ALTERNATIVES

Aging populations have three possible futures: (1) rejuvenation through higher fertility, which would start a new phase of population growth; (2) stabilization of the age structure and population numbers, with growth rates averaging zero; (3) hyper-aging, or super-aging (Myers, 1998) through maintenance of below replacement fertility, leading to excess proportions in the older ages, negative momentum and population decline. In the 21st century, there is likely to be continuing diversity in national population trends, with all three scenarios represented in different places at different times. Yet the second scenario is the only viable long-term global future. Just as the world s population has almost certainly had a young age structure throughout its history (Coale, 1972), the global population will have an old age structure in the future: there is no other age structure consistent with low fertility and low mortality. Nevertheless, a demographically old global population is not a possibility until late in the 21st century, since many less developed countries are still in the midst of the phase of high growth and age structure transformation.

See also: Demographic transition, Volume 5 and Schneider (1999).

REFERENCES

Chesnais, J C (1990) Demographic Transition Patterns and their Impact of the Age Structure, *Popul. Dev. Rev.*, 16, 327 336.

Coale, A J (1972) How a Population Ages or Grows Younger, in *Population: the Vital Revolution*, ed R Freedman, Doubleday-Anchor, New York, 47 58.

Coale, A J and Demeny, P (1983) *Regional Model Life Tables and Stable Populations*, 2nd edition, Academic Press, New York.

Coale, A J and Guo, G (1990) New Regional Model Life Tables at High Expectation of Life, *Popul. Index*, 56, 26 41.

Davis, K and van den Oever, P (1981) Age Relations and Public Policy in Industrial Societies, *Popul. Dev. Rev.*, 7, 1 18.

Easterlin, R A (1996) Economic and Social Implications of Demographic Patterns, in *Handbook of Aging and the Social Sciences*, 4th edition, eds R H Binstock and L K George, Academic Press, San Diego, CA, 84 93.

Hauser, P M (1976) Aging and World-wide Population Change, in *Handbook of Aging and the Social Sciences*, 1st edition, eds R H Binstock and E Shanas, Van Nostrand Reinhold, New York, 58 86.

Kinsella, K and Taeuber, C M (1993) *An Aging World II*, US Bureau of the Census, International Population Reports P95/92-3, US Government Printing Office, Washington, DC.

Myers, G C (1990) Demography of Aging, in *Handbook of Aging and the Social Sciences*, 3rd edition, eds R H Binstock and L K George, Academic Press, San Diego, CA, 19 44.

Myers, G C (1995) Demography, in *The Encyclopedia of Aging*, 2nd edition, ed G L Maddox, Springer, New York, 260 264.

Myers, G C (1998) Emerging Demographic Changes in an Aging World: an Overview, *Australas. J. Aging*, 17(Suppl.), 66 68.

OECD (1999) *OECD Economic Surveys 1998- 1999: Australia*, Organisation for Economic Co-operation and Development, Paris, 1 100.

Rowland, D T (1994) Population Policies and Aging in Asia, in *The Aging of Asian Populations*, ed United Nations, ST/ESA/SER.R/125, United Nations, New York.

Rowland, D T (1996) Population Momentum as a Measure of Aging, *Eur. J. Popul.*, 12, 41 61.

Schneider, E L (1999) Aging in the Third Millenium, *Science*, 283, 796 797.

Treas, J and Logue, B (1986) Economic Development and the Older Population, *Popul. Dev. Rev.*, 12, 645 673.

US Bureau of the Census (1998) *Statistical Abstract of the United States: 1998*, US Bureau of the Census, Washington, DC.

Van de Kaa, D J (1997) Options and Sequences: Europe s Demographic Patterns, *J. Aust. Popul. Assoc.*, 14, 1 29.

World Bank (1994) *World Population Projections*, 1994 95 edition, Johns Hopkins University Press, Baltimore, MD, 1 416.

Desertification, Definition of

Desertification has been defined by the United Nations (UN) as land degradation in arid, semi-arid and dry sub-humid areas resulting from various factors, including climatic variations and human activities (UNCED, 1992).

REFERENCE

UNCED (1992) *Earth Summit Agenda 21, Program of Action for Sustainable Development*, UNEP, New York.

MARTIN WILLIAMS *Australia*

Desertification

Martin Williams

Adelaide University, Adelaide, South Australia

Deserti cation is currently de ned as land degradation in arid, semi-arid and dry sub-humid areas resulting from various factors, including climatic variations and human activities. This de nition was put forward, debated and accepted at the United Nations Conference on Environment and Development (UNCED) held in Rio de Janeiro

in June 1992, colloquially known as the Earth Summit (UNCED, 1992). The UNCED de nition superseded an earlier one used by the United Nations Environment Programme (UNEP) in which deserti cation was de ned as land degradation in arid, semi-arid and dry sub-humid areas resulting mainly from adverse human impact (UNEP, 1992a). The signi cance of the 1992 UNCED de nition is that it allows a speci c role for climatic variations. In contrast, earlier de nitions were almost exclusively focused on human mismanagement and ignored the pervasive in uence of climate.

WHAT IS DESERTIFICATION?

The 1968 1974 drought in the Sahel region of West Africa rekindled international interest in land degradation in Africa and demonstrated very starkly some of the ways in which climatic desiccation and prolonged droughts can contribute to dryland degradation. French hydrologists and meteorologists drew attention to the severity of this drought at the International Geophysical Conference in Grenoble in August 1975. International concern over land degradation and water shortage was the catalyst for the United Nations (UN) to host a Water Conference in Argentina in March 1977, and for UNEP (United Nations Environment Programme) to host the UN Conference to Combat Desertification, Nairobi, August 1977.

At this conference a number of definitions were debated. One proposed that the term desertization be used in a very restricted sense to denote the extension of typical desert landscapes into semi-arid regions with 50 300 mm of rain a year. This definition specifically excluded wetter areas. An opposing view, ultimately accepted and endorsed by UNEP, was that areas of higher rainfall should be included. One outcome of the 1977 Nairobi conference was the definition of desertization as the spread of desert-like conditions in arid or semi-arid areas up to 600 mm, due to human influence or climatic change (Glantz, 1977). Later definitions tended to ignore the climatic factor and even to suggest that drought was a human construct arising from destructive forms of land use. As time progressed, the UNEP definition stressing adverse human impact was widely adopted and was used in the first (1992) but not the second (1997) edition of the *UNEP World Atlas of Deserti cation.*

Political and scientific awareness of desertification was enhanced by the publication of the two editions of the *UNEP World Atlas of Deserti cation* (UNEP, 1992a, 1997). In the comparatively short time between the publication of the first and second editions of this very useful atlas, there was already a much clearer scientific appreciation of the nature and causes of climatic variations, including global floods and droughts. This is very clearly reflected in the editorial discussion of climatic fluctuations in the second edition (UNEP, 1997).

The original roots of the term *deserti cation* connote abandoned or forsaken, and come from *desertum*, the neuter past participle of the Latin verb *deserere*. The present definition refers to land degradation and its possible causes. One consequence of such land degradation is that the land can no longer sustain much in the way of human settlement and so becomes abandoned. In that sense, the present definition is not too far removed from the original Latin meaning.

EXTENT AND SEVERITY OF DESERTIFICATION

It is not easy to measure the extent and severity of desertification. The earlier estimates of desertification put forward by UNEP in 1977, 1984 and 1987 (UNEP, 1992b) have been criticized as exaggerated and unreliable. Some of the estimates were based on questionnaires sent out to government officials in countries often already in the throes of severe drought and widespread social distress and famine. This naturally colored their perceptions of the magnitude of the problem (Thomas and Middleton, 1994). Granting that earlier reports were often overestimates, later estimates did suggest that desertification was a major global scourge. The revised estimates of Dregne *et al.* (1991) indicated that roughly three-quarters of all rangelands on each of the six continents were to some degree degraded. From 15 30% of the land in irrigated areas was degraded. Over half of the rain-fed croplands in Africa, Asia and Europe were considered degraded, over one-third in Australia and South America, and almost one-fifth in North America.

Criticism of the poor quality of the data and its lack of credibility persisted, and prompted UNEP to seek improved and less subjective ways of monitoring desertification.

UNEP therefore approached the International Soil Reference Center (ISRIC) in Holland and in 1987 UNEP and ISRIC began the Global Assessment of Soil Degradation (GLASOD) project. The GLASOD data were used very effectively in both editions of the *World Atlas of Deserti cation* (UNEP, 1992a, 1997). Table 1 is based on the latest published estimates for the global extent and severity of desertification (UNEP, 1997). The lack of adequate ground control remains a fundamental weakness even now, so that these estimates should be considered as provisional, order-of-magnitude values. Irrespective of the less than ideal quality of the data, one conclusion is all too clearly evident. Desertification is a major problem on every inhabited continent. This applies both to the relative proportions of land affected and to the total areas of land degraded.

Table 1 Extent of soil degradation in susceptible drylands, grouped by continent, in millions of hectares[a]

	Type of soil degradation				
Region	Water erosion	Wind erosion	Chemical deterioration	Physical deterioration	Total
Africa	119.1	159.9	26.5	13.9	319.4
Asia	157.5	153.2	50.2	9.6	370.5
Australasia	69.6	16.0	0.6	1.2	87.4
Europe	48.1	38.6	4.1	8.6	99.4
N America	38.4	37.8	2.2	1.0	79.4
S America	34.7	26.9	17.0	0.4	79.0
Total	467.4	432.4	100.7	34.7	1035.2

[a] *Source:* UNEP (1997).

CAUSES AND CONSEQUENCES OF DESERTIFICATION

Part of the reason for the difficulty in defining desertification stems from the very real problem of how to distinguish human impact from the long- and short-term impact of natural climatic variability in arid and semi-arid areas (Hare and Ogallo, 1993). Detection of cause and effect may require careful and sustained interdisciplinary research and long-term monitoring. For example, 17 years of satellite monitoring of vegetation cover along the southern margins of the Sahara revealed considerable variation from year to year in response to yearly variations in rainfall (Tucker and Nicholson, 1999). One conclusion from an earlier study by Tucker *et al.* (1991) extending from 1980 1990 was that at least 10 years of observations were needed to detect any possible trends in the vegetation cover in this region. A similar conclusion almost certainly applies to other dry regions of the world where the annual rainfall is highly variable and the response of annual and ephemeral plants equally so. It is easy to reach wrong conclusions based on too few observations, particularly if the comparison is between the state of the landscape towards the end of a series of wet years and the situation a decade or two later during the height of a prolonged drought.

A good example of this danger comes from the observations of Lamprey (1975) in Northern Darfur and Kordofan Provinces of Western Sudan. Lamprey s team used vehicles and aerial observations to map the boundary between full desert and desert scrub or grassland. They then compared the position of their mapped desert boundary with that defined by Harrison and Jackson after their 1958 survey of the vegetation of the Sudan (Harrison and Jackson, 1958a,b). Comparison of the two boundaries showed that the southern margin of the desert in 1975 was 90 100 km farther south than it was in 1958. This apparent southward shift of the desert margin by 100 km in 17 years, when averaged, gives a mean rate of advance of 5.5 km per year. Such a conclusion is highly misleading and seems to imply that the Sahara is advancing relentlessly at a steady southward rate of 5 6 km year^{-1}. Overgrazing may have caused the change in vegetation mapped by Lamprey in 1975, but it could equally have been a result of the severe drought that began in 1968 and persisted intermittently over the next 15 years. Alternatively, and more plausibly, the loss of plant cover may have been caused by a combination of drought and human impact. This example highlights the difficulty in distinguishing clearly between land degradation triggered by anthropogenic acceleration of natural processes and changes in plant cover arising from sustained rainfall deficit.

Although drought is neither a sufficient nor a necessary cause of desertification, the combination of drought and inappropriate forms of land use such as overgrazing, overcropping, and irrigation with inadequate drainage will all precipitate land degradation. Mabbutt (1978) recognized this when he defined desertification as a change to a more desertic condition, and pointed out that it involved impoverishment of ecosystems, accelerated soil degradation, reduced plant and animal productivity, and impoverishment of dependent human livelihood systems. When there was a combination of climatic stress and land-use pressure, the almost inevitable result was land degradation (Mabbutt, 1978).

Williams and Balling (1996) have recently reviewed the worldwide evidence for desertification. Some of the more obvious signs of dryland degradation include accelerated soil erosion by wind and water; salt accumulation in the surface horizons of dryland soils; a decline in soil structural stability with an attendant increase in surface crusting and surface runoff and a concomitant reduction in soil infiltration capacity and soil moisture storage; replacement of forest or woodland by secondary savanna grassland or scrub; an increase in the flow variability of dryland rivers and streams; an increase in the salt content of previously freshwater lakes, wetlands and rivers; and a reduction in species diversity and plant biomass in dryland ecosystems (Williams and Balling, 1996). Table 2 adds further detail

Table 2 Some causes and consequences of desertification[a]

Trigger factor	Consequences
Direct land use	*Physical processes affected*
Over-cultivation (decreased fallows, mechanized farming)	Decline in soil structure, soil permeability, depletion of soil nutrients and organic matter, increased susceptibility to erosion, soil compaction, dune mobilization
Overgrazing	Loss of biodiversity and biomass, increased soil erosion by wind and water, soil compaction from trampling, increased runoff, dune mobilization
Mismanagement of irrigated lands	Causes waterlogging and salinization of soils, hence lower crop yields, possible sedimentation of water reservoirs
Deforestation (burning, fuel and fodder collection)	Promotes artificial establishment of savanna vegetation, loss of soil-stabilizing vegetation, soil exposed and eroded, soil desiccation, increased frequency of dust storms, dune mobilization
Exclusion of fire	Promotes growth of unpalatable woody shrubs at the expense of herbage
Indirect government policies	Drives over-exploitative land-use practices
Failed population planning policies	Increases need for food cultivation, hence over-exploitation
Irrigation subsidies	Exacerbates flooding and salinization
Settlement policies/land tenure	Forces settlement of nomads, promotes intensive use of land which often exceeds carrying capacity
Improved infrastructure (roads, large dams, canals, boreholes)	Although beneficial, can exacerbate the problem by attracting increased livestock and human populations or increasing risk from salinization, possible lowering of ground-water table below dams, silting up of reservoirs, waterlogging; promotes large-scale commercial activity with little local benefit, flooding may displace people and perpetuate cycles of poverty
Promotion of cash crops and a push towards national/international markets	Displaces subsistence cropping, pushes local people into marginal areas to survive, promotes less resilient monocultures, and promotes expansion and intensification of land use
Price increases on agricultural produce	Incentive to crop on marginal land
War	Valuable resources, both human and financial, are expended on war at the expense of environmental management and the needs of the people; large-scale migration with resultant increased pressure on receiving areas
High interest rates	Forcing grazing or cultivation to levels beyond capacity
Natural	
Extreme drought	Decreased vegetation cover and increased land vulnerability to soil erosion. Creates an environment which exacerbates over-exploitation
Ecological fragility	Impact of land-use practices – impact also depends on resilience of environment

[a] *Source:* Williams *et al.* (1995).

and lists some of the more important causes and associated consequences of desertification.

OVERGRAZING, DROUGHT AND DESERTIFICATION

The Sahel drought that began in West Africa in 1968 and extended along the southern margins of the Sahara to Ethiopia and Somalia drew world attention to the impact of severe drought on desertification processes (Glantz, 1987; Grainger, 1990; Mainguet, 1994). (*Sahel* is Arabic for border or shore and broadly refers to the southern margin of the Sahara in the region between Mauretania and Chad). The Sahel drought afflicted over 500 million people in over 15 African nations, and resulted in widespread crop failure, livestock deaths, famine, and large-scale migration of environmental refugees out of their traditional grazing

lands and villages to urban fringe dwellings and refugee camps.

The severity of the Sahel drought prompted some scientists to ask whether a reduction in plant cover caused by overgrazing may have led to a decline in regional rainfall. During the 1970s, several workers analyzed the possible links between overgrazing and albedo change as a possible cause of prolonged drought in the Sinai desert and the Sahel. Charney *et al.* (1975) suggested that overgrazing in the Sahel reduces the surface plant cover and hence increases the albedo. (Albedo is the proportion of incoming solar radiation reflected from the earth s surface). The increase in albedo would result in lower land surface temperatures during the day and hence in a reduction in local convection, resulting in less rainfall (Figure 1). An obvious way to test these ideas is to measure albedo and temperature over areas of varying plant cover. Albedo measurements from satellites have yielded contradictory data. Surface measurements of albedo in Niger show significant seasonal variation in albedo and a far more limited range between vegetated and non-vegetated surfaces than proposed by Charney and his colleagues.

Perhaps the greatest single weakness of the albedo drought model lies in its failure to explain the globally synchronous pattern of historic floods and droughts over the last few centuries (Ropelewski and Halpert, 1987; Diaz and Markgraf, 1992). The 1980s and 1990s witnessed a resurgence in studies of the links between ocean surface temperature, atmospheric pressure and the global distribution of severe floods and droughts. The pioneering work of Sir Gilbert Walker in the 1920s was rediscovered and extended. In particular, a growing number of studies considered the role of El Nino/Southern Oscillation (ENSO) events and regional anomalies in sea surface temperature in controlling the incidence of major floods and droughts in widely scattered parts of the world (*see* El Nino/Southern Oscillation (ENSO), Volume 1; Southern Oscillation, Volume 1). For example, the Southern Oscillation Index (SOI), defined by Walker in 1924, is a useful tool in predicting year-to-year variations in the flow of the Blue and Main Nile. (The SOI is a measure of the atmospheric pressure difference between Darwin and Tahiti.) Time series analysis shows that there is a statistically significant correlation between years of low Nile flow, drought in Indonesia (shown also by years of narrow tree rings in the teak tree *Tectona grandis*), and years of summer monsoon failure in India and Eastern China (Williams and Balling, 1996). The opposite is also true, with years of extreme flooding occurring synchronously in all of these regions. The example of Indonesia shows that the influence of ENSO events extends well beyond the present drylands. Severe fires devastated parts of the tropical rainforests of Indonesia and Brazil during the 1982 1983 and 1997 1998 ENSO-related droughts, highlighting the importance of fire as an agent of desertification.

Although overgrazing is not responsible for the prolonged nature of the Sahel drought, drought itself can lead to local overgrazing, which can in turn accelerate soil erosion by wind (Figures 2 4) and water, and can precipitate the mobilization of previously vegetated and stable dunes. The eastern borders of the Great Indian Desert of Rajasthan, the southern margins of the Sahara from the Atlantic to the Nile, and the eastern and southern marches of the Gobi desert are all covered in sand dunes that have remained vegetated and relatively stable for the last 10 000 or so years. The shrubs and grasses that have long colonized the surface of these sand dunes act as highly efficient dust traps. The clay- and silt-size desert dust is washed down through the

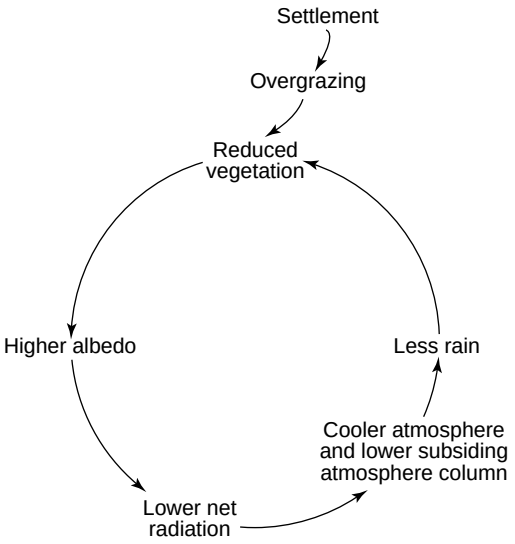


Figure 1 Hypothetical impact of overgrazing and reduced plant cover on rainfall in drylands. (Reproduced from Williams and Balling, 1996)

Figure 2 Dust storm triggered by cattle trampling and overgrazing in the Ethiopian Rift near Metahara. (Photo Martin Williams, February 1974)

Figure 3 Savanna grassland, Awash National Park, Ethiopian Rift, after six years of drought. The healthy status of the vegetation reflects over a decade of cattle exclusion from the Park. (Photo Martin Williams, February 1974)

Figure 4 Impact of drought and overgrazing by cattle on savanna grassland, Ethiopian Rift near Metahara. This photograph was taken immediately west of the Awash National Park shown in Figure 3. (Photo Martin Williams, February 1974)

surface of the dune to form bands of silty clay that parallel the original aeolian bedding planes of the dunes. These clay bands, together with plant roots, give extra mechanical strength to the dune and contribute to its immobilization. The fixed dunes of Africa and Asia are attractive to cultivators and pastoralists alike. The soils are easy to till and crops grow rapidly during the summer rains when photosynthesis is at a maximum. The trees, shrubs and grasses growing on the dunes provide reliable grazing for sheep, goats, cattle and camels. This happy state of affairs can become severely disrupted during times of severe drought combined with over-stocking.

Northern China offers some salutary lessons in this regard. For over 50 years, desertification has been a matter of great concern in Northern China (Ci, 1998; Williams, 2000). The Alashan area in the north-west of Inner Mongolia is the focus of much current attention because of an accelerating influx of sand into the Yellow River from remobilized desert dunes. For example, the dunes along an 80 km reach of the Yellow River left bank opposite the industrial city of Wuhai are currently advancing from the north-west at rates of up to $10\,m\,a^{-1}$. An estimated 80 million m^3 of sand is being blown into the river each year in this sector.

Alashan is one of the driest regions in China and covers an area of about 270 000 km^2. Rainfall declines from about 300 mm in the east to less than 50 mm in the west. Fixed and semi-active dune-fields cover about 90 000 km^2 and are the areas most vulnerable to desertification. Until the 1950s many of these low dunefields and sandsheets were covered in a relatively dense cover of shrubs, trees and grasses. Since that time the human population has doubled and livestock numbers have tripled. Greatly increased stock numbers, the large influx of immigrants from the south, and the occurrence of sporadic but severe droughts over large tracts of Alashan have resulted in widespread and locally severe desertification and associated ecological deterioration. Official local estimates suggest that some 30 000 km^2 of land are now severely degraded and that the rate of desertification is increasing by about 1000 km^2 each year. Monitored rates of dune advance range from more than $10\,m\,a^{-1}$ near the Yellow River to less than $1\,m\,a^{-1}$ farther inland (Williams, 2000). Control of desertification along the Yellow River is now an important element of China s national plans to combat desertification (China s Agenda 21, 1994). Control measures to arrest dune movement include tree-planting and extensive use of the straw mulch checkerboard technique, especially along the main railway line linking Lanzhou to Baotou.

DEFORESTATION, BIOMASS BURNING AND DESERTIFICATION

The importance of fire as an agent of desertification has been noted earlier. Humans have long been makers and users of fire. Archaeological evidence suggests that our ancestors began to use fire well over one million years ago. The Neolithic inception of early plant and animal domestication in widely scattered parts of the globe some 5 10 000 years ago led to clearing of forest and woodland for cultivation. The process accelerated as urban centers developed, with trading empires needing navies of wooden ships, culminating in the enormous demands for timber triggered by the industrial revolution. The tropical forests and savanna woodlands did not escape the impact of axe and fire. The scholarly French forester Aubréville (1949) has documented in great detail the replacement of tropical

African rainforest by secondary savanna and scrub as a result of tree clearing and burning associated with shifting cultivation and used the term *deserti cation* to encapsulate this process. Decreasing fallow cycles prevented tree regeneration and induced a vicious spiral of ever more xeric vegetation. One of his main conclusions was that human-induced deserts are forming today in Africa in areas receiving 750 1500 mm of rain a year.

Another aspect of the regular burning of savanna woodlands (Andreae, 1993) and of the accelerating burning of tropical forests that needs to be emphasized, is the question of feedback processes (Schlesinger *et al.*, 1990). Just as droughts and strong winds may increase the risk of fire, so too can more frequent and more widespread fires have an influence on local and regional climate. The amount of soot emissions (5.7 Tg C year^{-1}) (1 Teragram (Tg) is 1×10^{12} g) and fine particulate emissions (25.0 79.0 Tg year^{-1}) from biomass burning are almost identical to those from fossil fuel combustion (Ghan and Penner, 1992). Carbon released each year from shifting cultivation amounts to 500 1000 Tg C year^{-1}, while permanent deforestation, savanna fires, firewood burning, and burning agricultural wastes release a further 200 700, 300 600, 300 600 and 500 800 Tg C year^{-1}, respectively (Crutzen and Andreae, 1990). About 60% of the humid savannas are burned each year with an estimated fire efficiency of 83%, compared to only 0.6% of tropical forests at a fire efficiency level of 30 40% (Cachier, 1992). These estimates would need to be revised upwards for tropical forests given the greatly increased rates of burning in Brazil, Colombia and Indonesia since 1992.

Extensive tree clearing and burning is not of course restricted to the lowland forests and savannas of Africa, Asia and South and Central America; when conducted in mountainous areas, the downstream effects can also be severe. Removal of forest in tropical uplands can alter the local hydrological balance through increased runoff and reduced infiltration. Accelerated loss of soil from highland catchments can lead to siltation in reservoirs far downstream and not always within the same country. For example, accelerated soil loss from the Ethiopian uplands has led to siltation in several of the major tributaries of the Nile within the Sudan. By 1996, the capacity of the Roseires reservoir on the Blue Nile had been reduced by almost 60% through silt accumulation and that at Khashm el Girba on the Atbara, by 40%.

Hurni (1999) has painted a graphic picture of accelerated soil erosion in some parts of the Ethiopian uplands. In one region in Gojjam Province, the area cultivated rose from 40% in 1957 to 77% in 1995, while natural forestland decreased from 27% to 0.3%. Traditional farming methods recognized that soil losses during cultivation were high and so allowed long years of fallow for the soils to recuperate. The increasing demand for land has meant a reduction in fallow to virtually zero and an expansion of the area under cultivation. According to Hurni, annual rates of soil loss amount to about 40 tonnes ha^{-1} (2 mm a^{-1}) on mountain slopes, but attain rates of over 300 tonnes ha^{-1} (15 mm a^{-1}) during cultivation years, or some 5 10 times more than in non-mountainous areas (Hurni, 1999). These rates are far in excess of the long-term geological rates of erosion documented for this region, which amount to 0.01 mm a^{-1} in the Ethiopian headwaters of the Blue Nile. Of great concern is the progressive increase in soil erosion in the highlands of Ethiopia over the past 100 years. The rate of erosion in the Blue Nile headwaters during the 70 years before completion of the Aswan High Dam was 0.12 0.24 mm a^{-1}. In the 1970s annual soil loss from parts of the Ethiopian plateau was roughly 0.4 1.0 mm a^{-1}, and in the 1990s it was nearly double that rate.

IRRIGATION, DRYLAND SALINITY AND DESERTIFICATION

Salinity has always posed a problem to farmers in semi-arid areas. Rain contains a small amount of marine salt (cyclic salt). In humid areas salt is leached out of the soil and returns to the ocean in rivers. In drier areas leaching is less effective and so various salts tend to build up in the soil profile or at greater depths in the underlying sediments and groundwater. The more soluble sodium chlorides (NaCl) are more readily dissolved and re-precipitated than the less soluble carbonates and sulphates such as calcium carbonate ($CaCO_3$) and gypsum ($CaSO_4 2H_2O$). Since these processes have been operating intermittently over millions of years, the more arid regions of the world where evaporation greatly exceeds rainfall have an abundance of carbonate, gypsum and rock salt or halite. These areas are naturally saline and are not a result of human-induced desertification. Our concern here is with salinity resulting from human actions, especially through irrigation and clearing of native vegetation.

Large-scale irrigation projects in many parts of the arid world have failed to achieve their full potential through inadequate attention to soil drainage, canal maintenance, and the belated discovery of concentrated pockets of highly saline subsoil and groundwater within the irrigated areas. Seepage from irrigation canals can cause waterlogging and accumulation of progressively more salt in shallow closed depressions, which operate as natural evaporating basins (*see* Waterlogging, Volume 3). If the soils beneath these depressions are even slightly permeable, surface salts may be transferred in solution to shallow groundwater aquifers, and hitherto fresh water may become brackish or saline.

Salt accumulation in the irrigated Indus basin in the 1960s prompted a massive remedial effort to lower rising

water tables using tube wells to pump out excess water. In the 1980s efforts to halt dune encroachment in Northern China through irrigated shelter belts of trees poorly adapted to aridity led to salinization through over-irrigation and evaporative concentration of salt in the irrigation water. The Gezira irrigation scheme in Central Sudan provides up to two-thirds of the export revenue of the Sudan from long-staple cotton grown on less than 1% of the total land area. This project is one of the more successful big irrigation projects in the world and there has been minimal salt build-up in the soils except in a few local areas near the White Nile. At least five factors have contributed to this success: skilful local farmers, efficient management of canals, appropriate irrigation schedules, adequate drainage outlets, and very low salt concentrations in Blue Nile irrigation water.

We saw earlier that in dry areas, salt accumulation in soils stems from inadequate leaching by percolating rainwater. The reverse is true of the southern third of Australia, where European settlers cleared extensive areas of native vegetation for pasture and crops. Half the original woodland and forest that grew 200 years ago has now gone from Australia. The settlers were unaware of the role played by deep-rooted eucalypt trees as natural groundwater pumps. (Groundwater recharge under native vegetation is 1 2 mm $year^{-1}$; under wheat cultivation it is 40 120 mm tr^{-1}). The result was a slow but inexorable rise of local and regional groundwater levels, bringing dissolved salts to the surface and killing all plants not adapted to salinity (*see* Salinity and Agriculture, Volume 3).

This process of dryland salinity has resulted in huge loss of productive agricultural land in Southern Australia. In Western Australia, the area of saline land was 1.6 million ha in 1994 or about one-tenth of the cleared agricultural land in that state. In 1992 over 200 000 ha of the Murray Darling basin were in the grip of dryland salinity. Corresponding figures for South Australia and Victoria were, respectively, over 400 000 and 150 000 ha (Commonwealth of Australia, 1996). The losses in terms of agricultural production amount to over Aus.$500 million $year^{-1}$, so that considerable effort is now being devoted to arresting the process and preventing further damage.

In October 1999, the Murray Darling Basin Ministerial Committee published its salinity audit of the Murray Darling Basin which revealed that the economic costs of salinity had been greatly underestimated. Over 2.5 million ha of former agricultural land in Australia are now unusable because of dryland salinity. The cost to the Australian economy is nearly Aus.$1 billion a year, or triple previous estimates. At least 80 towns and cities in regional areas are currently threatened by salt. 20% of South Australia s surface water is too salty for human consumption. Western Australia is losing the use of an area the size of a football field every hour due to salinity and its biodiversity is under threat.

Coping with salinity will be a long slow process. Potential remedial measures include widespread adoption of salt tolerant plants, drainage and diversion of saline water, and planting of deep-rooted perennials, especially along aquifer recharge sites identified through remote sensing and ground control. Australia s dryland salinity problem is a classic example of a creeping environmental problem akin to the progressive desiccation of the Aral Sea (Glantz, 1999).

LESSONS LEARNED FROM DESERTIFICATION

The social, economic and political background to the examples discussed above may differ greatly, but they have four common attributes. First, indiscriminate destruction of the native vegetation has resulted in severe and widespread land degradation by water, wind or salt. Second, the repercussions of this degradation extend well beyond the immediate area prone to desertification. Third, the initial causes of the desertification go back in time for decades or scores of years. Fourth, the onset of degradation was gradual and the resulting loss of productive land was at first almost imperceptible except to a few prescient and acute observers (Glantz, 1999).

See also: Deserts, Volume 1; Land Degradation in the Mediterranean, Volume 3; Deserti cation Convention, Volume 4.

REFERENCES

Andreae, M O (1993) Global Distribution of Fires Seen from Space, *Eos*, 74, 129 135.

Aubréville, A (1949) *Climats, Forêts et Déserti cation de l'Afrique Tropicale*, Société d Editions Géographiques, Maritimes et Coloniales, Paris.

Cachier, H (1992) Biomass Burning Sources, *Encyclopedia of Earth Science Systems*, ed W A Nierenberg, Academic Press, San Diego, CA, 377 385.

Charney, J G, Stone, P H, and Quirk, W J (1975) Drought in the Sahara: a Biogeophysical Feedback Mechanism, *Science*, 187, 434 435.

China s Agenda 21 (1994) *White Paper on China's Population, Environment, and Development in the 21st Century*, China Environmental Science Press, Beijing.

Ci, L (1998) The Impacts of Global Change on Desertification in China, in *Sustainable Development in Arid Zones: Assessment and Monitoring of Desert Ecosystems*, Vol. 1, eds S A S Omar, R Misak, and D Al-Ajmi, Balkema, Rotterdam, 45 60.

Commonwealth of Australia (1996) *State of the Environment Australia, 1996*, CSIRO, Collingwood, Victoria.

Crutzen, P J and Andreae, M O (1990) Biomass Burning in the Tropics: Impact on Atmospheric Chemistry and Biogeochemical Cycles, *Science*, 250, 1669 1677.

Diaz, H F and Markgraf, V, eds (1992) *El Niño: Historical and Paleoclimatic Aspects of the Southern Oscillation*, Cambridge University Press, Cambridge.

Dregne, H E, Kassas, M, and Rozanov, B (1991) A New Assessment of the World Status of Desertification, *Deserti cation Control Bull.*, 19, 6 18.

Ghan, S J and Penner, J E (1992) Smoke, Effect on Climate, in *Encyclopedia of Earth System Sciences*, ed W A Nierenberg, Academic Press, San Diego, CA, 191 198.

Glantz, M H, ed (1977) *Deserti cation: Environmental Degradation in and around Arid Lands*, Westview Press, Boulder, CO.

Glantz, M H, ed (1987) *Drought and Hunger in Africa: Denying Famine a Future*, Cambridge University Press, Cambridge.

Glantz, M H, ed (1999) *Creeping Environmental Problems and Sustainable Development in the Aral Sea Basin*, Cambridge University Press, Cambridge.

Grainger, A (1990) *The Threatening Desert: Controlling Desertication*, Earthscan, London.

Hare, F K and Ogallo, L A J (1993) *Climate Variations, Drought and Deserti cation*, World Meteorological Organisation, Geneva.

Harrison, M N and Jackson, J K (1958a) *Ecological Classi cation of the Vegetation of the Sudan*, Forests Bulletin 2, Khartoum.

Harrison, M N and Jackson, J K (1958b) *Vegetation Map of the Sudan*, Sudan Survey Department, Khartoum.

Hurni, H (1999) Sustainable Management of Natural Resources in African and Asian Mountains, *Ambio*, 28, 382 389.

Lamprey, H F (1975) *Report on the Desert Encroachment Reconnaissance in Northern Sudan, October 21 - November 10, 1975*, National Council for Research, Ministry of Agriculture, Food and Resources: Khartoum, Reprinted in *Deserti cation Control Bull.* (1988), 17, 1 7.

Mabbutt, J A (1978) Desertification of Australia in its Global Context, *Search*, 9, 252 256.

Mainguet, M (1994) *Deserti cation: Natural Background and Human Mismanagement*, 2nd edition, Springer-Verlag, Berlin.

Murray Darling Basin Ministerial Committee (1999) *Salinity Audit of the Murray- Darling Basin*, Murray Darling Basin, website: http://www.mdbc.gov.au.

Ropelewski, C F and Halpert, M S (1989) Precipitation Patterns Associated with the High Index Phase of the Southern Oscillation, *J. Clim.*, 2, 268 284.

Schlesinger, W H, Reynolds, J F, Cunningham, G L, Huenneke, L F, Jarrell, W M, Virginia, R A, and Whitford, W G (1990) Biological Feedbacks in Global Desertification, *Science*, 247, 1043 1048.

Thomas, D S G and Middleton, N J (1994) *Deserti cation: Exploding the Myth*, John Wiley & Sons, Chichester.

Tucker, C J, Dregne, H E, and Newcomb, W W (1991) Expansion and Contraction of the Sahara Desert from 1980 1990, *Science*, 253, 299 301.

Tucker, C J and Nicholson, S E (1999) Variations in the Size of the Sahara Desert from 1980 1997, *Ambio*, 28, 587 591.

UNCED (1992) *Earth Summit Agenda 21: Program of Action for Sustainable Development*, UNEP, New York.

UNEP (1992a) *World Atlas of Deserti cation*, 1st edition, eds N Middleton and D S G Thomas, Edward Arnold, London.

UNEP (1992b) *Status of Deserti cation and Implementation of the United Nations Plan of Action to Combat Deserti cation. Report of the Executive Director*, UNEP, Nairobi.

UNEP (1997) *World Atlas of Deserti cation*, 2nd edition, eds N Middleton and D S G Thomas, Arnold, London.

Walker, G T (1924) Correlation in Seasonal Variations of Weather. IX: a Further Study Of World Weather, *Mem. Indian Meteorol. Dep.*, 24, 275 332.

Williams, M A J (2000) Desertification: General Debates Explored through Local Studies, *Prog. Environ. Sci.*, 2, 229 251.

Williams, M A J and Balling, Jr, R C (1996) *Interactions of Deserti cation and Climate*, WMO and UNEP with Arnold, London.

Williams, M A J, McCarthy, M, and Pickup, G (1995) Desertification, Drought and Landcare: Australia s Role in an International Convention to Combat Desertification, *Aust. Geogr.*, 26, 23 32.

Desertification: The United Nations Definition

see United Nations Convention to Combat Desertification (UNCCD) (**Volume 4**)

Disease, Respiratory

see Urban Climate and Respiratory Disease (**Volume 3**)

Disturbance

see Disturbance (**Volume 2**)

Disturbances, Generated by Humans

see Disturbance (**Volume 2**)

DVI (Dust Veil Index)

The DVI is a measure of climatic effectiveness that was formulated in 1963 by H H Lamb of the British

Meteorological Office. DVI is used to compare the magnitude of volcanic eruptions by measuring the volume of dust and aerosols lofted into the atmosphere by a volcanic eruption and determining its influence on climate. Lamb s principle was that the rays of the sun are inhibited by the volume of dust emitted in an eruption. The climatic effectiveness of any single volcanic eruption may therefore be assessed by calculating its dust or tephra output. Lamb further suggested that stratospheric circulation would move dust from low latitude eruptions around the world while dust from high latitude eruptions would remain in the hemisphere of origin. Different climatic impacts were therefore modeled for eruptions which occurred at different latitudes and longitudes. Besides the latitude of eruption, the following additional parameters were incorporated; the residence time of the dust in the atmosphere, the dust cloud size, the maximum deviation from average temperature and the greatest depletion of incoming solar radiation. By concentrating on the dust volume emitted by different eruptions and by assigning a DVI of 1000 to the 1883 eruption of Krakatau, Lamb constructed a chronology of eruptions. DVIs were calculated for eruptions occurring from 1500 through to 1983. The volume of dust emitted in an eruption was correlated with available temperature records and Lamb s research suggested a correlation between low temperatures and a DVI 100. Importantly, there is a correlation between high DVI values and climatic cooling. His work showed the importance of documenting both volcanic eruptions and climatic anomalies.

The DVI can be calculated in the following three ways based on:

1. the depletion of radiation following the eruption;
2. temperature variations following the eruption; and
3. the amount of solid material dispersed as dust after an eruption.

It is calculated to be:

1. $\mathrm{DVI} = 0{\bullet}97RD_{\max}E_{\max}t_{\mathrm{mo}}$;
2. $\mathrm{DVI} = 52{\bullet}5Td_{\max}E_{\max}t_{\mathrm{mo}}$; and
3. $\mathrm{DVI} = 4{\bullet}4qE_{\max}t_{\mathrm{mo}}$.

$RD_{\max}$ is the largest percentage depletion of direct solar radiation indicated by the average monthly temperatures for the hemisphere concerned; $E_{\max}$ is the largest proportion of the earth affected by the dust veil; $E_{\max}$ is 1 for eruptions between 20° latitude north and south, 0.7 for eruptions between 20 and 35° north and south, 0.5 for eruptions between 35 and 42° north and south, and 0.3 for eruptions occurring above 42° latitude; $Td_{\max}$ is the estimated average lowering of temperature for the most affected year over the middle latitudes of the hemisphere affected; q is the estimated volume in cubic kilometers of solid matter injected into the atmosphere and t_{mo} is the atmospheric residence time for the dust.

Lamb considered that eruptions having a DVI = 100+ were capable of significant impact on climate. Since DVI is a form of prediction of climate change due to eruptions, the higher the DVI measurement of a volcano, the greater the effects of the eruption are on climate. For example, the Tambora eruption in Indonesia possessed a DVI of 1500, which indicates extremely high effects on climate, conversely Mount St Helen s in the US possessed a DVI of only one.

PROBLEMS WITH DVI

As a potential eruption estimate for climate disruption, the DVI is inadequate. It over-emphasizes the significance of the dust output and uses a circular argument to establish the relationship between climatic flux and eruption magnitude. Importantly, the problem is that accurate and precise temperature records do not exist before the 17th century and reliance on observations of climatic severity may be misleading. Moreover, DVI is dependent upon estimates of variable quality and involves the circularity of argument. This means that any temperature deviation from the norm is directly related to the size of an eruption and not to any other external forcing factors such as variation in the output of energy from the sun or fluctuations in the earth s orbit (circular argument).

Problems also lie in the use of Lamb s indices. Values of R, T, and t are calculated from mid-latitude observations, not necessarily where the maximum effect of a low or high latitude eruption would be experienced. Calculating the DVI of an eruption based on recorded values of T (temperature lowering) may be an inconclusive argument. Of Lamb s 250 DVI estimates only 10% were based on measured observations of a decrease in solar radiation or temperature decrease while 48% were based on non-quantitative descriptions of eruptions such as personal accounts and observations. Consequently, it is simply not safe to assume that all temperature departures in eruption years are the result of volcanic activity. Problems also lie in the use of an inadequate eruption chronology, and in the assignment of values for $RD_{\max}$ (greatest depletion in solar radiation) and $E_{\max}$ (latitudinal dispersal of the dust cloud) from proxy records. Inadequate data exist from which to assign a DVI to eruptions prior to the 19th century, or to small eruptions or where several eruptions were closely spaced. As a result there is a risk of entering into a circular argument where temperature declines are used to support estimates of the value of $RD_{\max}$ ($Rd_{\max}$ reflects the greatest depletion in solar radiation) equated as the degree of eruption magnitude. Temperature deviations in eruption years should not be assumed to be the result of volcanic activity. Importantly, temperatures from observations made in a continental interior may not be relevant for the continental fringe and *vice versa*. Estimates of $E_{\max}$ fail to take

into account the possibility of other phenomena such as dust storms being interpreted as being of volcanic origin and problems with D_{max} that is the measure of the degree of eruption magnitude.

APPLICATIONS

An early attempt to refine the use of DVI as a major climatic forcing mechanism was made by Cronin in 1971 by investigation of the periodicity with which volcanoes erupted in the mid-20th century. Unfortunately, several significant eruptions were ignored, importantly the Alaska eruptions of Trident in 1952 and Spurr in 1953. Yet, it was observed that volcanic eruption in lower latitudes where the stratosphere is higher had a greater climatic impact than the eruption of volcanoes at high latitudes where the stratosphere is lower.

In 1974, Bray used DVI in an attempt to correlate volcanic activity and glacier response in the Northern and Southern Hemispheres over the past 40 millennia. It was discovered that the DVI was inadequate as a measurement tool to account for glaciers and climate change. The resultant inconsistent glacier response time indicated further problems with the use of DVI. For further clarification, Bradley and England in 1978 investigated glacier response and significant DVI events. They concluded that DVI was an inadequate measure of climatic effectiveness of a volcanic eruption. Other estimate methodologies of magnitude, the Volcanic Explosively Index (VEI) and the Ice-core Volcanic Index (IVI), were subsequently developed because of several failings of DVI. The DVI was later replaced by the VEI (*see* VEI (Volcanic Explosivity Index), Volume 3).

FURTHER READING

http://www.sunysuffolk.edu/∼mandias/honors/student/volcano/global_warming.html.
http://www.esci.keele.ac.uk/courses/geol317/volcano.html.
http://www.btinternet.com/∼mike.ferris/volcano.htl.
http:www.aber.ac.uk/∼jpg/volcano/lecture2.html.
http://www.aber.ac.uk/∼jpg/volcano/lecture2.html.
http://cdiac.esd.ornl.gov.
http://www.aber.ac.uk/∼jpg/volcano/lecture3.html.
http://bigmac.civil.mtu.edu/home/classes/ce459/public/p14/climate.html.
http://stommel.tamu.edu/∼baum/paleo/paleogloss/node12.html.
http://stommel.tamu.edu/∼baum/cdiac_11-20.html.
http://cdiac/esd/ornl/gov/ndps/ndp013.html.

MARY KEARNEY *Canada*

E

Earthquake Intensity and Magnitude Scales

Earthquakes are rated on two different scales: an intensity scale and a magnitude scale. An intensity scale is useful in classifying and describing damage to terrain and buildings following earthquakes. The arbitrary *Mercalli scale*, expressed in Roman numerals from I to XII, indicating conditions from barely felt to catastrophic total destruction, is used to represent intensity. Earthquake magnitude is estimated according to a system developed by Charles Richter in 1935. Richter established an open-ended, logarithmic scale in which each whole number represented a 31.5-fold increase in the amount of energy released by the earthquake. Thus, 4.0 on the Richter scale represent 31.5 times more energy than 3.0 and 992 times more energy

Table 1 Effects of earthquakes, intensity and magnitude scales and their approximate frequency

Effects in populated areas	Approximate intensity (modified Mercalli scale)	Approximate magnitude (moment-magnitude scale)	World total number year^{-1}
Nearly total damage	XII	>8.0	1 every few years
Great damage	X–XI	7–7.9	18
Considerable to serious damage to buildings; railroad tracks bent	VIII–IX	6–6.9	120
Felt by all, with slight damage to buildings	V–VII	5–5.9	800
Felt by some to felt by many	III–IV	4–4.9	6200
Not felt, but recorded	I–II	2–3.9	500 000

Source: US Geological Survey Earthquake Information Center.

Table 2 Intensities and magnitudes of some major earthquakes since 1970[a]

Year	Date	Location	Number of deaths	Mercalli intensity	Moment Magnitude
1971	Feb 9th	San Fernando, USA	65	VII–IX	6.7
1972	Dec 23rd	Managua, Nicaragua	5000	X–XII	6.2
1976	Jul 28th	Tangshan, China	250 000	XI–XII	7.4
1978	Sep 16th	Iran	25 000	X–XII	7.8
1985	Sep 19th	Mexico City, Mexico	7000	IX–XII	8.1
1988	Dec 7th	Armenia-Turkey border	30 000	XII	6.8
1991	Apr 2nd	Mt Pinatubo, Philippines	847	X–XII	7.8
1991	Oct 20th	Uttar Pradesh, India	1700	IX–XI	6.2
1995	Jan 17th	Kobe, Japan	5500	XII	6.9
1999	Sep 20th	Chi-Chi, Taiwan	2400	X–XII	7.6
2001	Jan 25th	Gujarat, India	2000+	X–XII	7.9

[a] For the impacts of the Mount Pinatubo earthquake and the associated volcanic eruption, *see* **Volcanic Eruption, Mt. Pinatubo**, Volume 1.

than 2.0. A *moment-magnitude scale* has now replaced the *Richter scale* because it is regarded as more accurate than Richter s amplitude-magnitude scale for large earthquakes. Moment magnitude considers the amount of fault slippage produced by the earthquake, the area of the surface (or subsurface) that ruptured, and the nature of the materials that faulted, including how resistant they were to failure. The basis of the intensity and the moment-magnitude scales, as well as values for some major natural earthquakes, are set out in Tables 1 and 2.

FURTHER READING

Christopherson, R W (1998) *Elemental Geosystems*, 2nd edition, Prentice Hall, Upper Saddle River, NJ.

Newhall, C, Hendley, II, J W, and Stauffer, P H (1997) *The Cataclysmic 1991 Eruption of Mount Pinatubo*, Philippines US Geological Survey Fact Sheet 113-197, Online Version 1.1 http://geology.wr.usgs.gov/fact-sheet/fs113-197/.

US Geological Survey 2001 Earthquake Bulletin 01/01/26 03:16:41 UTC 23.40N 70.32E Depth: 23.6 km 7.7Ms INDIA 65 miles (110 km) NNE of Jamnagar, India, http://neic.usgs.gov/neis/bulletin/010126031641.HTML.

IAN DOUGLAS *UK*

Earthquakes Triggered by Human Activities

Ian Douglas
University of Manchester, Manchester, UK

Varying amounts of evidence indicate that several types of engineering and extractive industry activities are suspected of triggering limited localized earthquakes known as induced seismicity. Examples include increased or accelerated activity in seismically active areas or the onset of seismicity in aseismic areas. Earthquakes occur whenever the stresses acting on a point in the lithosphere exceed the strength of rock at that point. Most natural earthquakes occur on pre-existing cracks or fractures where the rock is weakest.

The scienti c study of induced-seismicity began when earthquakes started to be felt in Johannesburg in 1894. By 1908 it was recognized that the gold mining initiated in 1886 was causing these events. The earth movements in this naturally highly stable area increased as the mines became deeper and disturbed the stress eld in the Archean rocks.

Such environmental changes are signi cant because:

1. *much of the earth's lithosphere is stressed near to its failure strength so that even small perturbations in the stress state may lead to induced seismicity;*
2. *many of the induced events occur close to the structures or enterprises that cause them. Thus quite small movements may cause important, economic, social and environmental changes;*
3. *some events can be large causing widespread damage.*

The main causes of induced seismicity are:

1. *reservoir storage: the impounding of large amounts of water behind dams;*
2. *extraction of minerals: quarrying at the surface, shallow mine workings, deep underground workings;*
3. *detonation of underground explosions;*
4. *uid removal: the extraction of oil or the pumping of water from an aquifer;*
5. *uid injection: the injection of high pressure uids into the ground, as in solution mining, waste disposal, geothermal power exploitation, and secondary oil recovery.*

RESERVOIR STORAGE

Reservoir-induced seismicity (RIS) may be divided into three classes:

1. No change in seismic activity;
2. Micro earthquakes, the frequency of which increases with initial inundation and thereafter varies according to water level variation. These tremors eventually subside;
3. Micro earthquakes, the frequency of which increases with the initial inundation and thereafter varies according to water level variation above a threshold value. Subsequent to this, medium size earthquakes may occur, followed by micro earthquakes and eventually stabilization.

Reservoir-induced earthquakes tend to occur close to the reservoir and to have small focal depths. The frequency of small earthquakes is higher than that of large ones and there is a high chance of occurrence in an area of normal or lateral faults. Medium sized earthquakes tend to occur in situations where the height of the dam exceeds 90 m, the stored volume of water is large, typically over 1 km^3, and locally the rocks have many fissures and faults in the crust. The increase in local seismic activity may lead to consequences ranging from numerous small stress-relieving earthquakes to catastrophic earthquakes and slides similar to the events in 1967 that destroyed India s Koyna Dam and resulted in a significant number of deaths.

Most reservoir seismicity is small and as yet there have been only 4 cases which have exceeded moment magnitude scale level (M) = 6:

1. Koyna, India, 1967, M = 6.3
2. Hsinfengkiang, China, 1962, M = 6.1
3. Kariba, Zambia/Zimbabwe, 1963, M = 6.1
4. Kremasta, Greece, 1965, M = 6.2

The Koyna earthquake in Maharashtra state, on December 10, 1967 (M 6.3) is the largest and most damaging reservoir-induced earthquake so far, claiming over 200 lives, injuring over 1500, rendering thousands homeless, and destroying much of the infrastructure in the Koyna Nagar town. The region south of the Shivaji Sagar reservoir has the highest reservoir-induced seismicity (RIS), with more than 150 events of magnitudes M_L 4.0 occurring from 1962 to 1996, of all open water storage reservoirs in the world. The activity is still continuing. Remarkably, most of the earthquakes are concentrated in a 25 km × 40 km area (Gupta, 1992; Gupta *et al.*, 1997). In the 1990s, a second reservoir, the Warna, one-third the size of the Shivaji Sagar reservoir, was completed some 30 km south of the Koyna Dam. It also appears to have created considerable RIS (Gupta *et al.*, 1997).

Seismicity started only a month after water began to be stored behind the 105 m high Hsinfengkiang dam in China in 1959. During the first few months seismicity increased as the water level rose. The main shock of 1962 occurred in an area where north-northwest faults intersected an east-northeast fault system. The local rocks have well developed joints. The main shock created cracks in the dam itself, the riverbank and local buildings. All events occurred within 5 km of the reservoir.

In 1958 Lake Kariba became the largest sustained artificial increase in loading on the earth s crust as it filled with water slowly from 1958 1963. In 1963 several large earthquakes occurred near the deepest portion of the reservoir (Meade, 1991). The earthquake mechanism was dip-slip on normal faults. Later earthquakes in the area, in 1966 1968, appear to have been unconnected with the reservoir.

The level of the Kremasta reservoir in Greece was repeatedly lowered and raised between 1967 and 1972. Careful observations by the project engineer showed that these fluctuations corresponded with levels of increased seismic activity (Therianos, 1974).

EXTRACTION OF MINERALS

Seismicity is primarily caused by changes in the elastic stress resulting from the removal of large masses of rock in mining and surface quarrying. Mining-induced seismic events arise from the redistribution of stress caused by large-scale mining. The total stress state around a mine excavation is the sum of the ambient stress state in the rock mass and the stresses induced by mining. The ambient stress is lithostatic corresponding to the weight of the overburden. Hence there is an increase in stress with depth. This is dangerous not only for those underground, but also causes subsidence. Mining-induced earthquakes of M = 5.7 have been associated with potash workings, while subsidence is particularly frequent in areas of underground salt extraction.

At a quarry in Wappingers Falls, New York, a shallow, 0.5 to 1.5 km deep earthquake sequence occurred in 1974 with a main shock of magnitude M = 3.3. These thrust earthquakes were located directly below the floor of the quarry which had been worked for 75 years down to a depth of 50 m, providing a stress change, albeit over many years, of 1.5 megapascals (MPa). As this region is an area of high horizontal compression, removal of surface material equated to reducing the normal stress. This led to the appearance of pop-up structures in the quarry floor as is often found in the eastern US.

DETONATION OF UNDERGROUND EXPLOSIONS

Underground explosions, as in deep mines also cause local changes in stresses in the rocks, which affect the stability of mine workings, and the overlying rocks.

FLUID REMOVAL

Since 1970 several small earthquakes have occurred in the Western Canada Sedimentary Basin. Many of the epicenters are shallow and are situated near areas of oil and gas production in the western side of the basin. The earthquake of October 19th, 1996, in the Strachan Field, in the Alberta foothills southwest of Rocky Mountain House was most probably triggered by gas extraction.

FLUID INJECTION

The major influence of fluid injection on the stress field is to increase pore fluid pressure. When fluid enters a fracture, it supports part of the normal stress equivalent to the pressure of the fluid. As the fluid has no shear strength, the effective normal stress and the frictional resistance to sliding are reduced. Any earthquakes so caused tend to be small and cause little damage.

The US Army injected chemical waste under various pressures into a waste disposal well drilled into crystalline basement rocks at their Rocky Mountain arsenal from March 1961 to February 1966. A sudden increase in seismic activity in this not usually earthquake prone area suggested that the fluid injection was causing earth tremors. Eventually a linear trend of seismic activity for 5 km from the well became apparent. Events of M = 4.8 5.2 in 1967 caused damage to windows, water pipes, roads and cracking

in concrete pillars. Those events caused $0.5 million of damage, with a total of $8 million damage over the period of pumping.

IMPLICATIONS FOR GLOBAL ENVIRONMENTAL CHANGE

While no one of these induced seismic events will have a global impact, the great 20th century upsurge in all the types of activity producing induced seismicity means that ground disturbance due to engineering works and minerals extraction is becoming an increasing local environmental threat. Now the possibility of induced seismicity is being considered as part of the environmental risks associated with major mining and reservoir projects. For example, although only three unequivocal cases of RIS have been established in Australia, the Talbingo reservoir in New South Wales, the Gordon dam in Tasmania and the Thomson reservoir in Victoria, seismic observations are being made at such reservoirs as the Wivenhoe, Burdekin Falls, Boondooma and Awoonga dams in Queensland (Cuthbertson).

Events at the induced seismicity scale can disrupt local utility infrastructure, especially pipelines and drains and may weaken such structures and river embankments and flood defenses, thus potentially aggravating the impacts of extreme events whose frequency may be heightened by global warming. As with other people-driven impacts on the environment, these cases of induced seismicity have to be seen as part of a cumulative series of small changes which together affect ecosystems and human living conditions.

REFERENCES

Cuthbertson, R, *Reservoir Induced Seismicity* http://quakes.earth.uq.edu.au/gen_info/networks/ris.html.

Gupta, H K (1992) *Reservoir-induced Earthquakes, (Developments in Geotechnical Engineering; Vol. 64)*, Elsevier, New York.

Gupta, H K, Rastogi, B K, Chadha, R K, Mandal, P, and Sarma, C S P (1997) Enhanced Reservoir-induced Earthquakes in Koyna Region, India, during 1993 1995, *J. Seismol.*, 1, 47 53.

Meade, R H (1991) Reservoirs and Earthquakes, *Eng. Geol.*, 30, 245 262.

Therianos, A D (1974) The Seismic Activity of the Kremasta Area, Greece, between 1967 and 1972, *Eng. Geol.*, 8, 49 56.

EC 1991 Directive on Nitrate

see Nitrogen: Agricultural Uses and their Impacts (**Volume 3**)

Ecodevelopment

see Land Transformation in China: 3000 Years of Human Ecological History (**Volume 3**)

Ecological Capacity

William E Rees
University of British Columbia, Vancouver, Canada

Super cially, the concept of human carrying capacity seems fairly straightforward - but consider the following. In 1986, ecologist Garrett Hardin wrote: " The na¨ve question, 'What is the human carrying capacity of the earth?' evokes a reply that is of no human use" (Hardin, 1986). Yet, in 1991, the same author declared that " carrying capacity is the fundamental basis for demographic accounting" (Hardin, 1991). What new evidence had come to light in the intervening ve years? Why had Hardin seemingly changed his mind? Well, there was no new evidence and Hardin's position actually remained rm. The fact is there is no real contradiction between the two statements. Hardin was right in both cases.

*How can we explain this apparent paradox? It seems that the human carrying capacity conundrum resides in certain qualities unique to our species. For other organisms, carrying capacity (*see *Carrying Capacity, Volume 4) is usually de ned as the maximum population that can be supported inde nitely in a de ned habitat without permanently impairing the productivity of that habitat (with allowances for climatic vagaries and other random events) (Rees, 1996; Kirchner* et al., *1985; Meadows* et al., *1992). This de nition is conceptually adequate for game and range management which is concerned with deer, antelope, or cattle. The demands of these animals on their ecosystems are xed and limited to food and shelter, the simplest of material requirements. But humans are hardly simple and present numerous novel challenges to demographers and human ecologists.*

To begin, humans are large omnivorous mammals who live in social groups, so even our basic demands are challenge enough for most natural ecosystems. Moreover, the situation is uniquely complicated because human demands have never been limited to the basics and are constantly increasing. This is because humans are creatures of culture, language, cumulative knowledge and social adaptation. Particularly since the scienti c revolution in the 16th and 17th centuries, human social organization and technology have enabled us gradually to restructure the environment

in the service of ever increasing numbers of people. By the full owering of the Industrial Revolution in the 19th century, we seemed to have gained the power to expand human carrying capacity virtually at will. The global population almost twice doubled to six billion people in the 20th century and resource consumption continues to balloon vastly faster. As a result, humans have become the equivalent of a geological force in changing the face of the Earth. Most signi cantly, the scale and quality of our impact has long been determined more by prevailing values, social behavior, and technology than by mere biophysical needs. This is why Hardin seemed to dismiss " what is the human carrying capacity of the Earth?" as a na¨ve question in 1986.

It is also why the notion of human carrying capacity - of meaningful limits to growth - has no credibility in the important corridors of global power. International organizations and national governments everywhere are dedicated to the proposition that human wellbeing in both the rich and poor countries is tied to maintaining the spiral of continuous economic expansion. Nevertheless, with the acceleration of human-induced (if not human-controlled) global change, it does seem fair to ask whether this vision is sustained as much by the weight of evidence as it is by mere techno-hubris. Are humanity's unique attributes really suf cient to free us forever from the constraints of carrying capacity that seem so rmly to bind all other species? Hardin's answer in 1991 was unequivocal - living within carrying capacity remains fundamental to long-term human welfare.

Continuous advances in technology and the human predisposition to trade are usually cited as reasons for rejecting the notion of human carrying capacity out of hand. Any human population (e.g., a region or country) that can trade surpluses of resource A for needed supplies of resource B need not be restricted in population or economic growth by the theoretical carrying capacity of its own territory. Technology further enables modern humans to reduce or eliminate many of the natural negative feedback mechanisms (such as disease and famine) that limited population growth in earlier times. In general, because humans are toolmakers, we are capable of squeezing more out of [our] environment than undisturbed nature would provide . . . so that the capacity of a given environment to support human populations can expand relatively rapidly (Demeny, 1988). According to some authors: Because of increases in knowledge, the earth s carrying capacity has been increasing throughout the decades and centuries and millennia to such an extent that the term carrying capacity has by now no useful meaning (Simon and Kahn, 1984).

Well, perhaps. Certainly human technological prowess has enabled our species to spread over virtually the entire planet. We have increasingly replaced the natural landscapes of Earth with cultural landscapes dedicated to the human enterprise. By the mid-1980s, Vitousek *et al.* (1986) estimated that humankind was already appropriating, directly and indirectly, almost 40% of the products of net terrestrial primary production. Similarly, the global fishing fleet appropriates seafood for humans that represents 25 35% of net marine primary production in the shallow coastal shelves and estuaries that produce 96% of the catchable fish in the sea (Pauly and Christensen, 1995). Through technology, humans have apparently become the dominant consumer organism in most of the major ecosystem types on the planet.

Of course, no one argues that all this progress comes without a cost. Even in the Neolithic, the invasion of new habitats by humans invariably resulted in local extinctions of easily hunted species; and in technologically augmented modern times we are witness to widespread fish stock collapses, deforestation, accelerating biodiversity loss, soil degradation, falling water tables, atmospheric change, and generalized pollution of air, water, and soil, all on a global scale. All the same, according to conventional wisdom, there is no cause for alarm. Mainstream economists argue that pollution can be managed by outright bans where necessary (e.g., ozone depleting chemicals) or by correcting for market failures through pollution charges and tradable pollution rights that internalize the externalities. As for resource depletion, this poses even less of a problem to technological optimists. They argue that human ingenuity has found functional substitutes for many declining resources over time (petroleum for coal, optical fiber for copper wire, etc.) and there is no reason why this shouldn t continue indefinitely. As Nobel Laureate economist Robert Solow wrote in 1974: If it is very easy to substitute other factors for natural resources, then . . . the world can, in effect, get along without natural resources, so exhaustion is just an event, not a catastrophe (Solow, 1974).

Indeed, belief in near perfect substitution has become a central tenet of neoclassical economics in recent decades. Some adherents even proclaim that resources are more the product of the human mind than they are of nature. In 1990, Walter Block, then Senior Economist at the Fraser Institute (a conservative think tank in Vancouver), argued the benefits of continuous population growth on grounds that additional people can create more resources than they use up, thanks to technological improvements . . . (Block, 1990). More recently, the University of Maryland s late famed growth advocate, Julian Simon, wrote: Technology exists now to produce in virtually inexhaustible quantities just about all the products made by nature . . . , and: We have in our hands now . . . the technology to feed, clothe, and supply energy to an ever-growing population for the next seven billion years . . . (Simon, cited in Bartlett, 1996; see also Simon and Khan, 1984). If this were true and human numbers the only consideration, carrying capacity

would indeed be a non-issue and Malthus ghost could surely take its final rest. But Professor Simon s statement could hardly be further from truth. Under pressure, Professor Simon later corrected his statement to read ... for seven million years . However, as physicist Albert Bartlett (1996) has shown, just 1% annual growth over this much reduced time-frame would result in a human population 30 000 orders of magnitude larger than the estimated number of atoms in the known universe! Simon had obviously not done the math.

CARRYING CAPACITY AS MAXIMUM HUMAN LOAD

> An environment s carrying capacity is its maximum persistently supportable load.
>
> (Catton, 1980)

There is another way of looking at human carrying capacity that enables us to interpret the techno-optimists view of the effects of trade and technology in quite a different light. Because of the variability in human demand, Catton (1980) argued that it makes more sense to define human carrying capacity not as the maximum sustainable population, but rather as the maximum load that can safely be imposed on the ecosphere by people. Total human load is clearly a function not only of population but also of *per capita* consumption, and the latter is increasing even more rapidly than the former due (ironically) to expanding trade and technology. In effect then, Catton is arguing that the world is being required to accommodate not just more people, but effectively larger people. To illustrate, in 1995 the extra-somatic energy being used *per capita* in rich countries like the United States and Canada, was equivalent to the work that might be done by up to 200 ghost slaves (Price, 1995). As a result of such trends, we can argue that load pressure relative to load-bearing capacity is rising much faster than is implied by mere population increases.

One way to estimate the human load on the ecosphere is through *ecological footprint analysis*. This method converts the material demands and impacts of a study population into its land/water area equivalent (Rees and Wackernagel, 1994; Rees, 1996; Wackernagel and Rees, 1996). Thus:

> The ecological footprint of a specified human population is the total area of land and water (ecosystems) required on a continuous basis to produce the resources that the population consumes, and to assimilate the wastes that the population produces, wherever on Earth the relevant land/water is located.

In practice, only a limited range of waste discharges can be readily converted to an areal equivalent, so calculated eco-footprints are underestimates of total load. Nevertheless, the most recent studies show that typical eco-footprints range from less than a hectare per person in poorer countries like India to seven or even ten hectares per person in high income countries like Canada and the United States (Wackernagel *et al.*, 1999). (Note that there are only about two hectares per person of ecologically productive land and water on Earth, so wealthy consumers are unwittingly appropriating more than their fair share of global ecological output.) Such data reveal not only the contribution of resource consumption to human load but also the prevailing inequity across income levels.

Significantly in the present context, the eco-footprint of the present global population of six billion is already as much as 30% larger than the planet the human load has already overshot the long-term carrying capacity of Earth (Figure 1). Indeed, at least two additional Earth-like planets would be required to support just the present world population sustainably if everyone enjoyed Western European material standards. Widespread ecological degradation indicates that even at current average consumption levels, we are living by drawing down cumulative stocks of *natural capital* (*see* Natural Capital, Volume 2) that cannot be replaced in any meaningful human time frame. According to some authors, the apparent expansion of human carrying capacity therefore cannot be permanent (Catton, 1980).

Other studies based on material flows have reached similar conclusions. Using readily available data on waste discharges and on the *assimilative capacity* (*see* Maximum Sustainable Yield (MSY), Volume 4) of biophysical

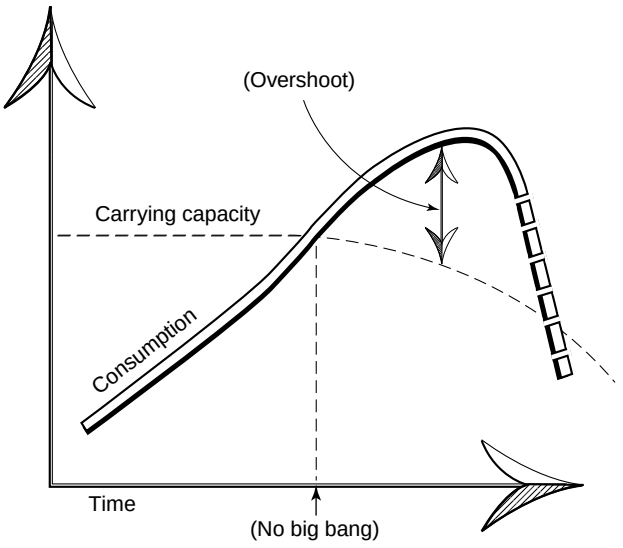


Figure 1 Overshoot is population and economic growth beyond long-term carrying capacity. One effect is accelerating degradation of habitat and supportive ecosystems, i.e., the reduction of future load-bearing capacity. Eventually, the population must come down. The relevant question is, will humanity manage a soft landing or must we suffer an uncontrolled crash. (Drawing by Phil Testamale; *Source:* Wackernagel and Rees, 1996)

systems, the Sustainable Europe Campaign has estimated that sustainability requires an approximately 50% reduction in the material throughput of the global economy in coming decades (Carley and Spapens, 1998). However, in order to create the necessary environmental space for developing countries to claim their fair share of global carrying capacity, developed country reductions in the use of non-renewable resources range between 88 and 94%. Even the World Business Council for Sustainable Development (WBCSD) seems convinced that: Industrialized world reductions in material throughput, energy use, and environmental degradation of over 90% will be required by 2040 to meet the needs of a growing world population fairly within the planet s ecological means (WBCSD, 1993) (*see* WBCSD (World Business Council for Sustainable Development), Volume 4). The need for a tenfold reduction in resource consumption by high income countries is confirmed by the United Nations Environment Programme s (UNEP) Global Environment Outlook 2000 report (UNEP, 1999).

Total human load also exceeds global carrying capacity by other criteria. So far the discussion has been totally anthropocentric. What about other organisms? We have noted that humans appropriate almost half of the products of terrestrial photosynthesis and 35% of marine primary production for themselves. Obviously, production that humans take for their own use is irreversibly unavailable to other species (this is a variation on the competitive exclusion principle). Every additional person, house, automobile, etc., therefore reduces in some measure the capacity of the planet to support other life-forms. In short, as people take over more and more forest, grassland, and other natural habitats for their own purposes, they invariably contribute to a loss of biodiversity including the irreversible loss of other species. The seeming increase in human carrying capacity (as measured by human numbers) is thus accompanied by a decrease in ecosystems capacity. Indeed, habitat takeover combined with overharvesting and pollution are producing what some analysts are calling the sixth extinction (the first 5 extinctions are discussed in Extinctions and Biodiversity in the Fossil Record, Volume 2). The current extirpation rate is 100 1000 times pre-human levels inferred from the fossil and paleontological record (Pimm *et al.*, 1995).

To some, this provides moral or ethical grounds for claiming that people have long since exceeded global carrying capacity. So called deep ecologists, many conservationists, and ordinary people who simply love nature, believe that other species have intrinsic value quite apart from their utility to humans. From this perspective, the millions of other life forms with whom we share the planet have a right to exist and people have no right to take more from nature than they need for material sufficiency. Growth in human numbers and consumption to the point where it destroys the integrity of ecosystems and eliminates species is seen as inherently immoral. By this criterion, we breach human carrying capacity when population or economic expansion compromises the survival of other life forms.

Where does all this leave expansionist economists and other techno-optimists? Again, there is no doubt that advances in technology have enabled the continued expansion of human populations. However, this apparent increase in short-term carrying capacity has actually been accompanied by a real reduction in long-term load-bearing capacity. Technology creates seeming abundance by increasing both the intensity and spatial scale of humanity s exploitation of the ecosphere. Sometimes humans gain by appropriating the habitats and ecological niches of other species at great cost to systems integrity. Sometimes technology is used to accelerate the energy and material flux

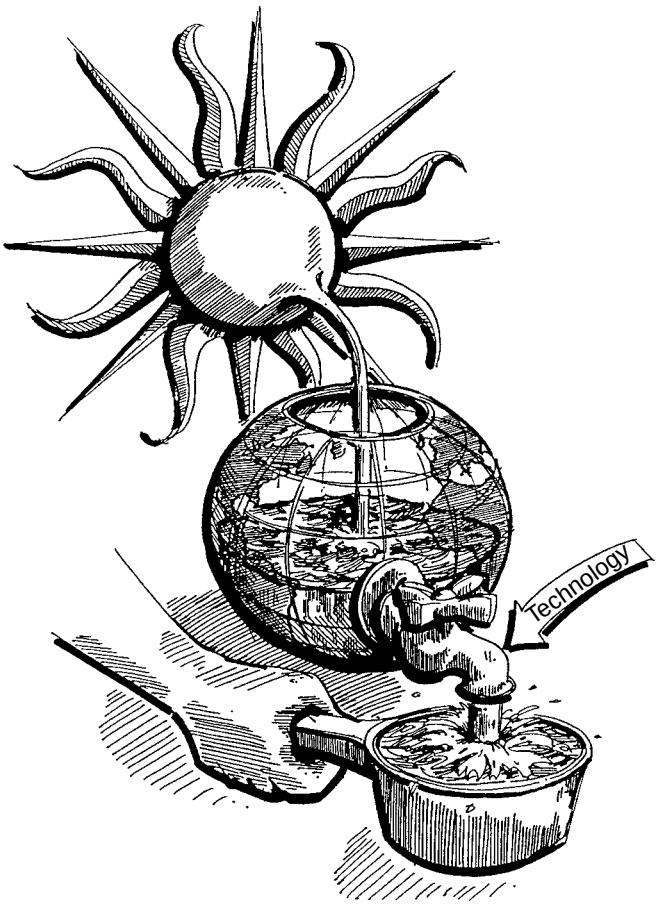


Figure 2 To date, technology has been used largely to accelerate the intensity and rate at which human beings appropriate both renewable and non-renewable resources. Sustainability demands that essential stocks of natural capital not be exploited beyond rates of renewal/replacement and that waste production not exceed the assimilative capacity of ecosystems. (Drawing by Phil Testamale; *Source:* Wackernagel and Rees, 1996)

through exploited ecosystems (Figure 2). In either case, short-term gains in systems productivity permanently erode the resource base and with it the life-support functions of the ecosphere. Energy-subsidized intensive agriculture may be more productive than low input practices in the short-term, but it also increases the rate of soil degradation and groundwater depletion (*see* Agricultural Intensi cation in Western Europe, Volume 3). Similarly, the effectiveness of electronic fish finding devices and high-tech fishing gear overwhelms the reproductive capacity of commercial fish stocks (*see* Changes in World Marine Fish Stocks, Volume 3). The net effect is that more and more people are increasingly dependent on technologically enhanced but potentially unsustainable material flows. Why unsustainable? The technologies themselves are usually based on non-renewable resources (e.g., fossil fuel), and the ecosystems, already over exploited, are now further stressed by human-induced global (e.g., climate) change.

At its best, technological innovation can improve the efficiency of resource use, but again this does not increase load-bearing capacity *per se*. In theory, shifting to more energy and materially productive technologies should enable a defined environment to support a given population at a higher material standard, or a higher population at the same material standard, thereby seeming to increase carrying capacity. However, in either case, we are simply maintaining total human load constant in the vicinity of load-bearing capacity the latter is still ultimately limiting. In practice, we have not done even this well. The steady gains in material efficiency in the post World War II period have been accompanied by increasing *per capita* and aggregate consumption (WRI, 2000). It seems that efficiency gains may actually work against conservation by raising incomes and lowering prices. This gives people more money to spend on cheaper goods and services so total throughput human load has steadily expanded.

The carrying capacity gains from trade are also often illusory. Wealthy nations can exceed domestic limits to growth only by appropriating carrying capacity from the global commons or from other countries all over the world. While commodity trade releases local populations from local resource constraints, it merely displaces some fraction of that population s biophysical load to distant export regions. The resultant increase in population and consumption in importing regions increases the total load of people on the ecosphere but there is no net gain in carrying capacity since trade often reduces the load-bearing capacity of the export regions. Indeed, like technology, trade may even result in lower global load-bearing capacity. Access to cheap imports (e.g., food) may reduce the incentive for importing regions to conserve local natural capital stocks (e.g., agricultural land). Meanwhile, trade accelerates the depletion of natural capital soil, water, forests, biodiversity in the exporting regions, as producers struggle to compete in the global marketplace.

CONCLUSIONS: LIGHTENING THE LOAD TO SOFTEN THE CRASH

Technology and trade have unquestionably enabled the vast expansion of the human enterprise. But it is a confusion of logic to equate the increase in numbers and economic scale with an increase in load-bearing capacity of the earth. The main effect of technology and trade to date has been to enable humans to extend the area and depth of their ecological footprints all over the planet at the expense of other species and the overall integrity of the ecosphere. Far from having been freed from biophysical constraints, people in high income countries today are arguably more indebted than ever to nature and increasingly vulnerable to rapid global change.

So perhaps human carrying capacity ought to be the fundamental basis for demographic accounting after all. There may be any number of different sustainable human populations depending on prevailing values, material standards and technological sophistication. However, for sustainability, any population must live within the *load-bearing capacity* of the planet.

Where do we stand today? The empirical evidence suggests that with present beliefs and prevailing technologies, the existing human population has already significantly overshot global carrying capacity. Unfortunately, because populations in overshoot live partly by depleting even their renewable resource system, they also reduce the future life support capacity of their habitat. Some researchers therefore conclude that the life expectancy of our western techno-industrial society is less than 100 years (Duncan, 1993) and that it will necessarily self-destruct, producing massive ecological damage, social chaos, and megadeath (Smith and Sauer-Thompson, 1998; see also Catton, 1980; Price, 1995; Lyons *et al.*, 1995; Smith *et al.*, 1997).

If we are even to soften what so many see as the inevitable crash, technology must assume a more constructive role. For people to live sustainability within the biophysical carrying capacity of Earth in the 21st century, there must be dramatic reductions in energy and material consumption. Researchers in Europe and North America are calling for a *factor-four* (double the output with half the resources) (Ayers, 1998; Weizsacker *et al.*, 1997) and even a factor-ten revolution in resource productivity (WBCSD, 1993; UNEP, 1999), and are lobbying for the fiscal and economic policies needed to make it happen.

But more efficient technology is not the only way humans can increase their numbers without increasing their load on the Earth we can also adapt behaviorally. Suppose

the increasing stress and spiritual emptiness of consumer society results in the voluntary simplicity movement really catching on in the developed world (increasing numbers of people are, in fact, trying to create meaning in their lives beyond the accumulation of stuff by taking time to rediscover their communities, their families, and themselves). To the extent that materially simpler but spiritually richer living reduces energy and material consumption, the eco-footprints of the world s wealthy would shrink. This in turn would create more ecological space for the poor particularly in the developing world to increase their consumption to an acceptable standard.

Certainly simply to maintain our present course would be a dangerous strategy. Complex systems theory (and an increasing number of real world disasters) suggests that biophysical systems under stress become increasingly brittle. As we push the ecosphere, the oceans, and the atmosphere to their limits, they are likely to exhibit erratically unpredictable or even catastrophic behavior (*see* Monitoring in Support of Policy: an Adaptive Ecosystem Approach, Volume 4). The sudden collapse of fisheries and accelerating climate change may be signals that we are close to the edge of such systemic chaos on a global scale. Thus, for all the apparent success of technology in freeing us from the whims of nature, if the human population and the material economy continue to expand, there is a finite possibility that we could actually be on the verge of extinction, blissfully unaware that a mathematical fiction in the space of the possible is about to become reality. And the really nasty feature is that that may take only the tiniest of changes to trigger the switch (Cohen and Stewart, 1994, 212).

REFERENCES

Ayres, R U (1998) *Turning Point: an End to the Growth Paradigm*, Earthscan Publishers Ltd., London.

Bartlett, A (1996) The Exponential Function XI: the New Flat Earth Society, *Phys. Teach.*, 34, 342 343.

Block, W (1990) Environmental Problems, Private Property Rights Solutions, in *Economics and the Environment: a Reconciliation*, ed W Block, The Fraser Institute, Vancouver, BC, 281 332.

Carley, M and Spapens, P (1998) *Sharing the World: Sustainable Living and Global Equity in the 21st Century*, Earthscan Publications, London.

Catton, W R (1980) *Overshoot: the Ecological Basis of Revolutionary Change*, University of Illinois Press, Urban.

Cohen, J and Stewart, I (1994) *The Collapse of Chaos*, Penguin Books, New York.

Demeny, P (1988) Demography and the Limits to Growth, in *Population and Resources in Western Intellectual Traditions*, eds M Teitelbaum and J Winter, Cambridge University Press, New York, 213 244.

Duncan, R C (1993) The Life-expectancy of Industrial Civilization: the Decline to Global Equilibrium, *Popul. Environ.*, 14, 325 357.

Hardin, G (1986) Cultural Carrying Capacity: a Biological Approach to Human Problems, *BioScience*, 36, 599 606.

Hardin, G (1991) Paramount Positions in Ecological Economics, in *Ecological Economics: the Science and Management of Sustainability*, ed R Costanza, Columbia University Press, New York, 47 57.

Kirchner, J, Leduc, G, Goodland, R, and Drake, J (1985) Carrying Capacity, Population Growth, and Sustainable Development, in *Rapid Population Growth and Human Carrying Capacity: Two Perspectives*, ed D Mahar, World Bank Staff Working Papers # 690, Population and Development Series No. 15, The World Bank, Washington, DC.

Lyons, G, Moore, E, and Smith, J W (1995) *Is the End Nigh? Internationalism, Global Chaos and the Destruction of the Earth*, Avebury, Aldershot.

Meadows, D H, Meadows, D L, and Randers, J (1992) *Beyond the Limits: Global Collapse or a Sustainable Future*, Earthscan Publications, London.

Pauly, D and Christensen, V (1995) Primary Production Required to Sustain Global Fisheries, *Nature*, 374, 255 257.

Pimm, S L, Russell, G J, Gittleman, J L, and Brooks, T M (1995) The Future of Biodiversity, *Science*, 296, 347 350.

Price, D (1995) Energy and Human Evolution, *Popul. Environ.*, 16, 301 319.

Rees, W E (1996) Revisiting Carrying Capacity: Area-based Indicators of Sustainability, *Popul. Environ.*, 17, 195 215.

Rees, W E and Wackernagel, M (1994) Ecological Footprints and Appropriated Carrying Capacity: Measuring the Natural Capital Requirements of the Human Economy, in *Investing in Natural Capital: the Ecological Economics Approach to Sustainability*, eds A-M Jansson, M Hammer, C Folkeand, and R Costanza, Island Press, Washington, DC.

Simon, J and Kahn, H, eds (1984) *The Resourceful Earth: a Response to Global 2000*, Basil Blackwell, Oxford.

Smith, J W and Sauer-Thompson, G (1998) Civilization s Wake: Ecology, Economics and the Roots of Environmental Destruction and Neglect, *Popul. Environ.*, 19, 541 575.

Smith, J W, Lyons, G, and Sauer-Thompson, G (1997) *Healing a Wounded World: Economics, Ecology and Health for a Sustainable Life*, Praeger Publishers, Westport, CT.

Solow, R (1974) The Economics of Resources or the Resources of Economics? *Am. Econ. Rev.*, 64(2), 1 14.

UNEP (1999) *Global Environmental Outlook 2000*, (GEO-2000), United Nations Environment Program, Nairobi.

Vitousek, P, Ehrlich, P, Ehrlich, A, and Matson, P (1986) Human Appropriation of the Products of Photosysthesis, *Bioscience*, 36, 368 373.

Wackernagel, M and Rees, W E (1996) *Our Ecological Footprint: Reducing Human Impact on the Earth*, New Society Publishers, Gabriola Island, BC.

Wackernagel, M, Onisto, L, Bello, P, Linares, A C, Falfán, I S L, Garcia, J M, Guerrero, A I S, and Guerrero, M G S (1999) National Natural Capital Accounting with the Ecological Footprint Concept, *Ecol. Econo.*, 29, 375 390.

WBCSD (1993) *Getting Eco-Ef cient*, Report of the First Antwerp Eco-Efficiency Workshop, November 1993, World Business Council for Sustainable Development, Geneva.

Weizsacker, E U, von Lovins, A B, and Lovins, L H (1997) *Factor Four: Doubling Wealth, Halving Resource Use* (the New Report to the Club of Rome), Earthscan Publications, London.

WRI (2000) *Weight of Nations: Material Out ows from Industrial Economics*, World Resources Institute, Washington, DC.

Ecological Footprint

Ecological footprint analysis is usually applied to discrete human populations. The *ecological footprint* of a specified population is the total area of land and water ecosystems required on a continuous basis to produce the resources that the population consumes, and to assimilate the wastes that the population produces, wherever on Earth the relevant land/water is located. Thus, the eco-footprint calculation is usually based on a specified population s current final demand for goods and services and represents the total load that the study population imposes on the ecosphere. Sometimes, however, it is instructive to eco-footprint particular industries or economic activities. For example, we can use eco-footprint analysis in comparative technology assessment to contrast the relative ecosystem requirements of alternative production technologies or processes.

WILLIAM E REES *Canada*

Ecosystem Integrity

see Ecosystem Integrity (**Volume 4**)

Ecosystems, Urban

see Urban Ecosystems (**Volume 3**)

Ecotourism

see Tourism and Ecosystems (**Volume 3**)

Ecotoxicology

Ecotoxicology, combining ecological and toxicological knowledge, was defined in the early 1970s as the branch of science concerned with the toxic effects of chemical and physical agents on living organisms, especially on populations and communities within defined ecosystems, and including the transfer pathways of those agents and their interactions with the environment. It does not cover, as a rule, the study of toxic effects on humans or domesticated animals, which is the object of toxicology, but has in its purview the environmental pathways of toxicants to humans.

Ecotoxicological knowledge is the basis for the ecological risk assessment of chemicals and physical agents. *Risk* is defined as the probability of occurrence of damage or undesirable effect resulting from exposure to a hazardous substance or agent. *Exposure* is the amount of the chemical or physical agent that reaches the target organism or population.

Consequently ecotoxicological research addresses quantitatively, in a logical sequence, the source term, i.e., the amounts, forms and sites of release of polluting agents; their transfer and transformations as they move among environmental compartments; the amounts reaching the identified biological targets (exposure and dose); and the response (dose effect relationship) of these organisms, populations, communities and ecosystems. This methodology was derived from previous studies on radioactive contamination caused by testing of nuclear weapons as well as by peaceful applications of nuclear energy, which is the subject of *radioecology*.

Ecotoxicology developed rapidly in the 1970s, with the increasing releases of pollutants from fossil fuels combustion and the use of synthetic chemicals. It was spurred by concern originating from the acidification of precipitation, lakes and soils (the acid rain problem), the effects of dichlorodiphenyltrichloroethane (DDT), pesticides and other synthetic chemicals on animal populations. Ecotoxicological research has produced detailed knowledge on the fate and effects of toxicants in the environment, which has been broadly used in defining policy measures such as ecological quality standards or pre-marketing ecotoxicity tests for new chemicals. It has also improved pollution monitoring, for instance through the use of bioindicators, which are organisms that concentrate pollutants, and biomarkers, or biochemical changes resulting from even very low-level exposure to specific chemicals. An ecotoxicological problem of great current interest is that of endocrine disrupters, which interfere with sexual expression, reproductive behavior, etc., of animals in aquatic ecosystems.

For an example of an ecotoxicological assessment, *see* Environmental Cadmium in the Food Chain, Volume 3. For an example of ecotoxicological method, *see* Use of

Epidemiological Data to Develop Air Quality Standards for Short-term and Long-term Environmental Health Bene ts, Volume 4.

PHILIPPE BOURDEAU *Belgium*

Emergy

The leading American ecologist H T Odum developed the concept of *emergy* (Odum, 1988, 1991). He defines emergy as the energy of one type required in transformations to generate a flow or a storage (Huang and Odum, 1991) Emergy is thus the available energy of one kind previously used up directly and indirectly to make a product. Its unit is the emjoule. Since work is required to concentrate materials, higher concentrations of material require more emergy per mass. In other words, emergy is required to concentrate materials and keep them concentrated. The ratio of emergy to mass of materials is a useful measure of work that has been applied to materials. Emergy can be used to represent both the quality and the quantity of energy that any input or product of the economy represents. Furthermore, the emergy concept offers an alternative and complementary viewpoint of value on the biophysical nature of human activity (*see also* Emergy, Volume 5).

If the mining, processing and use of lead in an industrialized society is considered, processes within the earth system add emergy as they converge and concentrate heavy metals in making ore bodies during high temperature and pressure mountain building tectonic events. Human society adds emergy from fuels, machinery and human effort when it mines minerals, refines them and produces such products as lead acid batteries. The greater the work involved in concentrating lead, the higher the emergy per mass. Spatially, concentrations of lead are greatest in cities and along transportation corridors. On the other hand, as lead is dispersed in waste streams, the emergy/mass ratio decreases. In wetlands, the dispersed heavy metals get taken up by plants and then by microorganisms (Odum *et al.*, 2000).

REFERENCES

Huang, S-L and Odum, H T (1991) Ecology and Economy: Emergy Synthesis and Public Policy in Taiwan, *J. Environ. Manage.*, 32, 313 333.

Odum, H T (1988) Self-organization, Transformity, and Information, *Science*, 242, 1132 1139.

Odum, H T (1991) Emergy and Biogeochemical Cycles, in *Ecological Physical Chemistry: Proceedings of an International Workshop, November 8- 12, 1990*, eds C Rossi and E Tiezzi, Elsevier Science, Amsterdam, 25 56.

Odum, H T, Wòjick, W, Pritchard, Jr, L, Ton, S, Delfino, J J, Wòjick, M, Leszczynski, S, Patel, J D, Doherty, S J, and Stasik, J (2000) *Heavy Metals in the Environment: Using Wetlands for their Removal*, Lewis, Boca Raton, FL.

IAN DOUGLAS *UK*

Environmental Cadmium in the Food Chain

J Keith Syers[1] and Michael Gochfield[2]
[1]*Naresuan University, Phitsanulok, Thailand*
[2]*Rutgers University, NJ, USA*

Cadmium (Cd) is important because it is a toxic and carcinogenic element, which occurs widely in the environment, both from natural and anthropogenic sources. Cadmium has been linked to bone disease in Japan and more widely to renal functional abnormalities. The extent to which these abnormalities are clinically signi cant is an important issue, which has engendered substantial controversy and no consensus. Given these human health implications, it is essential that our understanding of cadmium in the environment be more precise.

An expanding world population places pressure on finite natural resources, particularly with regard to establishing world food security. Worldwide, food production growth has outstripped population growth largely due to technology, which has increased yields per hectare (ha). Progress has, however, been uneven and in many regions, significant numbers of people, particularly the poor, women and children, remain seriously malnourished. Nutrient inputs, from several sources, are an important contribution to increased food supply but can have some unintended, adverse effects on ecosystems and humans. Cadmium (Cd) is an impurity in all of these nutrient inputs and poses risks to ecosystems and the human food chain. These are often moderated and regulated by government intervention and policies which should not, however, jeopardize food security, particularly in the poorer parts of the world where population growth, hence the demand for food production, is highest.

Management of food safety to protect human health must be carried out in the wider context of two worlds, a smaller one with food security and current food surpluses, and a larger one with food shortages that are aggravated by rapid population growth and urbanization. The latter, primarily tropical, world is also one where environmental, agricultural, ecological and biomedical data relevant to cadmium and

other toxic substances are most meager and where concomitant health and nutritional stresses are greatest.

SOURCES

- Atmospheric cadmium is mainly derived from metal smelting, the uncontrolled burning of fossil fuels for energy, and incineration of solid wastes and sewage.
- Reducing emissions of cadmium to the atmosphere by improvement in pollution control devices, energy conservation, and the selection of low-polluting fuels must be pursued vigorously.
- Biosolids (sewage sludge) contain relatively small concentrations (mean $<10\,mg\,kg^{-1}$ dry matter (DM) in developed countries) but can add large amounts of cadmium to soil. The disposal of biosolids is therefore a major issue. Restrictions on dumping at sea will force a more extensive disposal of biosolids on land or increased incineration with concomitant release of cadmium to the atmosphere.
- Biosolids have beneficial attributes in improving soil structure and supplying essential plant nutrients. However, concerns for the effects of increased loading of heavy metals, particularly cadmium, will require improved treatment technologies, e.g., source reduction, removal from waste streams, and separation of high cadmium from low cadmium sludge with the application of the former to non-food producing sites.
- Animals exposed to cadmium produce manure which can recycle to soil, adding significant amounts of cadmium as well as valuable nutrients. Where such manures are applied repeatedly in high quantities to localized areas, high cadmium levels may result.
- Phosphatic fertilizers are produced from phosphate rocks which vary in cadmium content from less than one to over $100\,mg\,kg^{-1}$. Igneous deposits have lower cadmium contents but account for approximately 5% of world reserves.
- The development of economically feasible cadmium reduction technologies should be pursued.
- There are regional differences in the sources of phosphate rocks used in fertilizer production, in addition to varying rates of application, so that generalizations regarding cadmium input are fraught with danger.
- Phosphatic fertilizers are essential for agricultural production and in many countries, particularly in Africa, where major increases in food production are required, soils are deficient in phosphate.
- The need and demand for phosphatic fertilizers will continue to increase, and risk management for cadmium must balance the costs and benefits and the risks associated with phosphatic fertilizer use.
- The contribution of each source: atmospheric deposition, animal feeds, biosolids and phosphatic fertilizers, to total cadmium in soil and to bioavailable cadmium requires evaluation at different spatial scales.

FOOD CHAIN

The following points are important in the context of cadmium toxicity:

- Bioavailability of soil cadmium to plants (phytoavailability) is a function of both soil supply parameters and plant demand parameters, and exerts great influence on both ecological and human food chains.
- Soil supply is primarily a function of the soil cadmium concentration, and partitioning of cadmium between solid and solution phases. Partitioning is primarily governed by soil acidity, but other competing (e.g., calcium, Ca) or complexing (e.g., chlorine, Cl) ions in solution, and the type of solid phases present (e.g., clay type and content, amount of organic matter) also influence cadmium partitioning.
- The soil chemistry of cadmium is complex and affects phytoavailability and leaching.
- Movement of cadmium through soil, to plant roots or to groundwater, is an area of increased research activity. Cadmium movement is affected by the distribution between solid and aqueous phases. More accurate measurements of cadmium leaching in a variety of agricultural systems are urgently needed to determine long-term accumulation potential, both in soil and groundwater.
- Plant demand for cadmium and translocation of cadmium to edible parts are moderated by the presence of competing ions in solution (e.g., Ca, hydrogen, H, zinc, Zn). Deficiencies of zinc, in particular, may lead to high cadmium accumulation.
- Growth rates and root proliferation in soil are affected by plant nutrient status (e.g., phosphate availability), so that soil fertility also plays a role in cadmium phytoavailability.
- All plant species appear to access the same pool of soil cadmium, with no evidence of solubilization of less-available pools evident by any species to date. However, due to different root uptake and translocation, very large differences in cadmium accumulation are evident between species and between cultivars within a species. This can be exploited to develop low cadmium cultivars or those suitable for phytoremediation. Improved agricultural practices can reduce cadmium phytoavailability.
- Bioavailability of cadmium to soil microorganisms is much less well defined, but appears to be related to the dissolved cadmium concentration, although soil invertebrates may have exposure to cadmium through the soil solution, dermal contact, and ingestion of soil and food. Invertebrates vary greatly in sensitivity to cadmium. Some soil invertebrates appear tolerant;

many aquatic invertebrates are highly sensitive. The relative importance of these various pathways and the ecologic consequences have not been clearly identified.

- Soil ingestion by grazing livestock or wildlife appears to be a less important exposure pathway than ingestion of herbage. Animals act as a barrier to cadmium transfer through the food chain to humans because of their low intestinal absorption and the sequestration of cadmium in livers and kidneys, leaving muscle tissue (meat) relatively low in cadmium. As with humans, cadmium accumulates with age in animals. Unless significantly large quantities of organs from older animals are consumed, cadmium transfer through this pathway is limited.
- Although there are abundant data on total cadmium concentrations in foods, bioavailability of plant cadmium to animals or humans is not a simple function of food cadmium concentration. It is related to the form of cadmium in the food and other nutritional factors, e.g., Iron (Fe), Zn, Ca and phytate ($C_6H_{18}O_{24}P_6$). Low concentrations of iron, zinc and calcium appear to increase food cadmium bioavailability, especially in the case of rice. Identifying and predicting food cadmium bioavailability is an urgent priority.
- Direct ingestion of soil by humans may also be a significant exposure pathway in highly polluted soils, e.g., eating of soil by children. This is unlikely in agricultural food chains and is usually related to urban or industrial sources of contamination in residential areas. However, conversion of farmland to residential land (suburbanization) is proceeding rapidly, particularly near large cities in developed nations. Therefore incidental or deliberate geophagy may sometimes become important for cadmium.
- Ingestion of cadmium in drinking water was a significant pathway leading to Itai Itai disease in some areas of Japan. Industrial effluents and atmospheric deposition add cadmium to surface waters used for drinking in many parts of the world. Leaching of cadmium, while reducing plant uptake, contributes to ground water contamination. The implications vary with hydrology, aquifer features, and demand. Understanding cadmium in water(s) is of particular importance for ecological risk assessments.

CONSEQUENCES

Human Health

Cadmium is a toxic and carcinogenic element. The International Agency for Research on Cancer has identified cadmium as a known human carcinogen. Epidemiologic studies of cadmium-exposed workers show excess lung cancer. However, cancer has not yet been shown to be a risk from ingestion of cadmium at levels encountered in the environment. The main non-cancer endpoint of concern is kidney damage, particularly to the proximal tubule, but also to the glomerulus. Bone and hematologic disorders have also occurred at high level exposure. A wider range of organ toxicity has been demonstrated in animals.

Risk Assessment

Assessment of the human health consequences of environmental exposure to cadmium and the resulting policy implications should be based on the results of quantitative health risk assessment. The traditional risk assessment paradigm has four steps each of which is subject to major uncertainties.

In the hazard identification step, high level cadmium exposure has clearly caused significant kidney disease and bone weakening, represented by the Itai Itai disease in parts of Japan. Lower level exposures have caused kidney damage in other environmentally exposed populations (e.g., Cadmibel study). Proteinuria is the critical endpoint, but the critical level of proteinuria which represents a significant and unacceptable change is controversial.

The dose response relationship between cadmium exposures and proteinuria remains incompletely characterized. The usual measure of cadmium body burden, urinary cadmium (U-Cd), is itself dependent on kidney damage. The relationship is further confounded by the effects of age on both cadmium level (U-Cd usually declines after the 7th decade) and proteinuria (which normally increases in the elderly). There are associations between occupational exposures (intensity and duration), cadmium levels in urine and kidney, and biomarkers of kidney damage. Both the glomerulus and the tubule can be damaged, resulting in both high and low molecular weight proteins in the urine. In general it is the tubular effect which is considered the most sensitive, although it is likely to be the glomerular damage which is of greatest clinical significance. In the general population exposed to cadmium, proteinuria has been reported at a U-Cd of $5\,\mu g\,Cd\,g^{-1}$ creatinine, with some studies showing elevations at U-Cd levels as low as $2.5\,\mu g\,Cd\,g^{-1}$ creatinine, however criteria for defining elevation have varied among researchers.

Improving the exposure assessment requires quantifying how the relationship among intake, uptake and distribution are mediated by the interaction of several factors under a wide variety of conditions. These include zinc, iron, and calcium status (as well as current levels of these cations in the diet). Because both cadmium concentrations and bioavailability in soil and foods and nutritional status of the studied populations vary, care must be taken in generalizing from the few studies that have been conducted, mainly in Europe and Japan. Intake of cadmium from home gardens varies greatly, and may be a significant source of cadmium for some people in developed nations as it is for subsistence farmers.

The effective dose (dose to the kidney) cannot be measured directly, making risk characterization difficult. Biomarkers of exposure include cadmium in blood (relatively short half-life), U-Cd (correlated with kidney cortex cadmium, but disproportionately excreted when the kidney is damaged), as well as *in vivo* estimations of organ cadmium by neutron activation and X-ray fluorescence, which are not yet widely available or sufficiently sensitive for low level exposures. The low molecular weight proteins, called biomarkers of effect, are often quite well correlated with exposure, but the lack of a completely independent biomarker of exposure is a source of uncertainty in risk assessment.

The World Health Organization has developed a Provisional Tolerable Weekly Intake (PTWI), basing its risk assessment mainly on data from studies of the Japanese Itai Itai disease. Many assumptions and some uncertainty factors are involved which may no longer be appropriate, particularly for other populations, and which may over- or under-estimate risk. The relationship between an average level of $10\,\mu g\,Cd\,g^{-1}$ creatinine in urine and the frequently cited critical concentration of approximately $200\,\mu g\,Cd\,g^{-1}$ tissue in the kidney cortex seems to be reasonably consistent, but the relationship is not linear. The estimation of $7\,\mu g\,kg^{-1}\,week^{-1}$ (corresponding to an allowable daily intake of $1\,\mu g\,kg^{-1}\,day^{-1}$) may benefit from re-examination. Depending on the outcome, the dietary contribution from various sources (some published, some additional ones needed), can be compared to identify populations that may be very close (within an order of magnitude) of the PTWI.

Ecological Effects

Many ecotoxicologic studies have been conducted on selected species. There is wide variability in uptake, storage and effects of cadmium among most groups studied, but data, particularly on effects, are sparse, and interactions among metals need to be examined. Studies need to focus on organisms relevant to particular areas of cadmium contamination. Very few standardized indicators useful for ecotoxicology assessments exist, and they do not represent all trophic levels.

Toxic impacts on one species may be biomagnified through ecosystem food chains, and effects may include changes in the health, longevity, or reproduction of certain species, as well as alterations in the flow of energy and the cycling of matter. Therefore ecological risk assessment becomes much more complex than human health risk assessment. Many studies are carried out on the microcosm (aquarium scale), and these need to be expanded to the mesocosm (pond-scale) level. In addition, studies of terrestrial ecosystems and factors modifying the bioavailability and transfer of cadmium at every trophic level are required.

POLICY

Risk Management

A defining feature of sound risk management is stakeholders, who must live with the options and should have input into the risk process. This requires that the selection of endpoints be relevant to the population, and therefore risk management strategies to prevent a particular hazard, may differ regionally and over time. Risk management must take into account competing risks and must not introduce greater risks.

The precautionary principle (*see* Precautionary Principle, Volume 4 and 5) dictates that, in the absence of data or in the presence of uncertainty, risks should be managed at low levels until data are available indicating that less protection is warranted or acceptable.

Given that cadmium may continue to build up in the environment, and future exposures may be substantially higher than at present, and that Cd is hazardous to human health and ecosystems, it is prudent to keep cadmium levels and bioavailability at safe levels.

Economic Issues

Measures and policies to reduce the potential risks to human health from food produced on soils with different levels of accumulated cadmium may be costly. Internationally-accepted food standards and requirements will change the patterns of world trade in both food products and fertilizers. Imposing food quality criteria will require fertilizer manufacturers, farmers, marketers and consumers to modify their behavior. The costs and benefits to each stakeholder need careful consideration. Controls at the fertilizer input level may be costly to the fertilizer production industry and ultimately to farmers and food consumers, particularly in the developing world where rapidly growing populations create increased energy demand. The removal of cadmium from phosphoric acid (H_3PO_4) (decadmiation) is not cost-effective at present prices. The technical and economic conditions under which decadmiation becomes an economic option need to be established. The costs and benefits of alternative policy measures must be established in order to set socially optimal food safety levels. Their consequences for food security and international trade of less developed countries are particularly important.

CONCLUSIONS

1. Cadmium is a toxic and carcinogenic metal, nearly ubiquitous in the environment. It is hazardous to humans and ecosystems in excessive amounts.
2. Inputs are mainly from atmospheric deposition, application of biosolids, use of phosphate fertilizer, and

from effluents from cadmium-using and recycling industries.

3. These inputs to the environment can be subject to control methods which vary in their economic and social costs.
4. There are significant data gaps related to bioavailability and food chain transfers, particularly under tropical conditions.
5. Caution is needed in generalizing exposure estimates from the existing epidemiologic studies, because of substantial differences between occupational exposure (primarily inhalation) and community exposure (primarily ingestion), as well as between cultures with different diets and perhaps susceptibilities.
6. The most sensitive human health endpoint of concern for human food chain exposure is damage to the proximal tubule of the kidney with resulting proteinuria.
7. The significance of low-level proteinuria and its relationship to exposure duration and levels is controversial and cannot be resolved with the existing epidemiologic data. This limitation precludes recommending specific cadmium levels for the environment.
8. Identifying the susceptible subsets of the population is important, and many factors influencing susceptibility remain to be identified and quantified.
9. There remain significant data gaps in the area of ecological risk assessment, particularly at the ecosystem (rather than single species) level.
10. There is a need to prioritize risk reduction strategies for the major cadmium sources.
11. In addition to the uncertainties concerning food chain and biomedical issues, there remain uncertainties regarding the benefits, costs and potential risks associated with various cadmium management scenarios.
12. The growing populations imposing increasing energy demands and the limited arable land (much with low natural fertility) of developing countries in tropical regions requires special attention with regard to risk balancing.
13. There is an urgent need to support further research on the topics identified above as having high priority.

BACKGROUND INFORMATION

This contribution is based on the results of a Scientific Committee on Problems of the Environment (SCOPE) workshop *Environmental Cadmium in the Food Chain: Sources, Pathways and Risks*, held at the Belgian Academy of Sciences, Brussels on 13 16th September, 2000, which in turn built on prior conferences, including a meeting of the International Agency for Research on Cancer (Gargnano, Italy, 1991), a Workshop on cadmium of the Organization for Economic Cooperation and Development (Saltsjobaden, Sweden, 1995), and a symposium on cadmium (Berkeley, USA, 1997). The first meeting emphasized health effects and the last two soil plant relationships. Each stressed the importance of reducing risks to human health and the environment from cadmium.

The workshop included 32 technical papers in five plenary sessions, two summary reports of prior meetings, eight working group reports, and a follow-up meeting to develop this consensus summary. Further details and recommendations are provided in the Workshop Working Group Reports, which are posted on the SCOPE website (2001).

REFERENCE

SCOPE website (2001) http://www.icsu-scope.org.

FURTHER READING

ATSDR (1999) *Toxicological Pro le for Cadmium*, Agency for Toxic Substances and Disease Research, Atlanta, GA.

Bernard, A M and Lauwerys, R R (1997) Dose-response Relations Between Urinary Cadmium and Tubular Proteinuria in Adult Workers, *Am. J. Ind. Med.*, 31, 116 118.

Cherian, M G (1997) Metallothionein and Intracellular Sequestration of Metals, in *Comprehensive Toxicology: Vol. 3 Biotransformation*, eds I G Sipes, C A McQueen, and A J Gandolfi, Pergamon, New York, 489 503.

Friberg, L, Piscator, M, Nordberg, G F, and Kjellstrom, T, eds (1974) *Cadmium in the Environment*, 2nd edition, CRC Press, Cleveland, OH, 1 248.

Friberg, L, Elinder, C-G, Kjellstrom, T, and Nordberg, G F, eds (1985, 1986) *Cadmium and Health: A Toxicological and Epidemiological Appraisal, 2 Vols*, CRC Press, Boca Raton, FL.

Friberg, L (1950) Health Hazards in the Manufacture of Alkaline Accumulators with Special Reference to Chronic Cadmium Poisoning, *Acta Med. Scand.*, 138(suppl. 240), 1 124.

Hutchinson, T C and Meema, K M, eds (1987) *Lead, Mercury, Cadmium and Arsenic in the Environment*, SCOPE 31, John Wiley & Sons, Chichester.

McLaughlin, M J and Singh, B R, eds (1999) *Cadmium in Soils and Plants: A Global Perspective*, Kluwer, Dordrecht.

McLaughlin, M J, Tiller, K G, Naidu, R, and Stevens, D P (1996) Review: The Behaviour and Environmental Impact of Contaminants in Fertilizers, *Aust. J. Soil Res.*, 34, 1 54.

Mislin, H and Ravera, O (1986) *Cadmium in the Environment*, Birkhauser, Boston, MA, 1 144.

Nordberg, G F, Herber, R F M, and Alessio, L (1992) *Cadmium in the Human Environment: Toxicity and Carcinogenicity, Scienti c Publication No. 118*, International Agency for Research On Cancer, Lyon.

Nriagu, J O (1981) *Cadmium in the Environment*, John Wiley & Sons, New York, 2 Vols.

Nriagu, J O and Sprague, J B (1987) *Cadmium in the Aquatic Environment, Advances in Environmental Science and Technology*, John Wiley & Sons, New York, 1 272, Vol. 19.

OECD (1996) *Sources of Cadmium in the Environment*, Organisation for European Co-operation and Development, Paris, 1 482.

OECD (1996) *Fertilizers as a Source of Cadmium*, Organisation for European Co-operation and Development, Paris, 1 252.

World Health Organization (1995) *Cadmium - Environmental Aspects. Environmental Health Criteria 135*, International Programme on Chemical Safety, World Health Organization, Geneva.

Environmental Change and Human Health: Extending the Sustainability Agenda

see Environmental Change and Human Health: Extending the Sustainability Agenda (*Opening essay*, **Volume 3**)

Environmental Health and Urban Poverty

see Urban Poverty and Environmental Health (**Volume 3**)

Environmental Health – the WHO Defintion

see Policy Responses to Public Health Issues Relating to Global Environmental Change (*Opening essay*, **Volume 4**)

Eutrophication

see Eutrophication (**Volume 2**)

Exergy (High Quality Energy)

see Monitoring in Support of Policy: an Adaptive Ecosystem Approach (*Opening essay*, **Volume 4**)

309

Fisheries: Effects of Climate Change on the Life Cycles of Salmon

Ian A Fleming[1,2] and

Arne J Jensen[1]

[1]*Norwegian Institute for Nature Research, Trondheim, Norway*

[2]*Coastal Oregon Marine Experimental Station, Newport, OR, USA*

Changes in climate are likely to impact water temperature, quantity and quality and thus alter the biological composition, production and function of aquatic environments. Understanding the responses of aquatic life, such as fish, to these changes represents a fundamental ecological challenge. Temperature affects all vital processes of cold-blooded aquatic organisms, including activity, feeding, growth and reproduction. Thus, it is widely believed that climate warming will affect the distributions of marine and freshwater species, resulting in a general shift polewards to reflect their thermal preferences. Moreover, the composition and relative abundance of species in local communities will shift towards those of warmer temperatures. It is also predicted that these effects may be more pronounced towards the poles where warming is expected to be greater than in the tropics. Evidence is growing that such changes are occurring in both marine and freshwater ecosystems in response to the warming of aquatic environments in recent decades. Climate change will not only affect aquatic environments through the direct effects of temperature, but also through other environmental changes coincident with, or driven by, the changes in temperature. Ocean currents are predicted to change and with them so will circulation patterns and regions of nutrient upwelling that affect water quality and biological productivity. For example, the Eastern Pacific Ocean, which has undergone warming in recent decades, has experienced a concurrent decline in productivity, particularly in the near-surface zone. The cause, however, remains poorly understood, but the implications are far reaching for the ecological community. Climate change will also alter annual and seasonal precipitation patterns and increase variability within and among years, which are expected to have profound effects on freshwater environments, affecting lake levels and stream flows. Moreover, if changes in water quantity occur with changes in water temperature, then changes in water quality will likely follow, owing to changes in concentrations of ions, dissolved gases, and organic materials. Thus, the effects of climate change on aquatic environments, both marine and freshwater, are expected to be multifaceted.

SALMON AS AN INDICATOR

Fishes, such as salmon, which spend their lives both in freshwater and marine environments (i.e., diadromous; Figure 1) are ideal for the study of responses to climate change and may prove to be good biological indicators of it. Their life cycle, which takes them from freshwater through estuarine and out into open ocean habitats and then back again, exposes salmon to a diversity of potential effects resulting from climate change. The typical life cycle of anadromous salmon (*Oncorhynchus* and *Salmo* spp.) begins with the spawning of eggs into nests constructed in the gravel of freshwater rivers or streams during the fall (Figure 1). The eggs incubate through the winter and hatch in the spring, with the juveniles emerging from the gravel shortly thereafter. The juveniles remain in freshwater streams, rivers and/or lakes for as little as a couple of months or up to several years (five) before migrating through estuaries and out into the ocean. It is in the ocean where the vast majority of the salmon's growth occurs over a period ranging from a few months up to several years (four). During this oceanic period, salmon appear to be almost continuous travelers and may migrate distances in excess of 5000 km away from the oceanic outlet of their natal river. They then return to their natal river to breed and may migrate up to 4000 km in fresh water to reach their spawning grounds. Accurate homing to the natal river to breed facilitates the evolution of population-specific traits and the intra-specific variation that so characterizes salmon species.

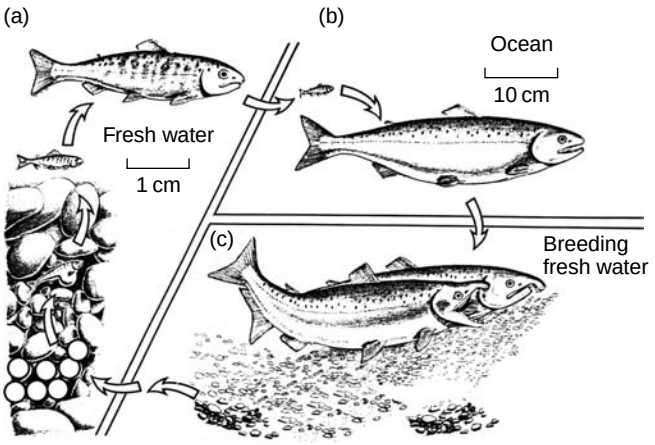


Figure 1 A generalized life cycle of anadromous salmon, involving (a) egg incubation and juvenile rearing in fresh water followed by (b) migration to the ocean for sub-adult growth and then (c) the return to fresh water for breeding

Global warming is predicted to affect freshwater before it affects marine habitats because of the relatively rapid effects of changes in runoff and precipitation on freshwater systems, and the buffering capacity of the oceans. Moreover, rivers and streams are likely to be among the most sensitive aquatic environments to warming because of their small volumes and sensitivity to changes in flow discharge. Thus salmon, with their early and adult life history stages occurring in these environments, will be exposed early on to such direct effects of climate change. In addition, it is generally thought that temperate species, such as salmon, will be more susceptible because warming is predicted to be greater at higher than lower latitudes. Salmon will also be subject to the indirect effects of climate change, such as changes in ocean currents, wind patterns and timing of river flows, that affect the productivity and availability of food, the costs of migration and mortality.

Generally, it is difficult to predict the responses of aquatic animals to climate change, partly because of lack of information about their environmental requirements and the dynamics of their life histories. Salmon, however, have been well studied and quantitative models exist for such things as growth, feeding, egg incubation, thermal tolerance and life history. In addition, a number of long time-series on salmon populations exist by which the responses to climate change may be identified. Finally, the large amount of among-population variation within salmon makes them well suited to study, because the first response of organisms to changing climate will involve changes at this level.

RESPONSES OF SALMON TO CLIMATE CHANGE

Climatic warming is likely to affect salmon species negatively, particularly those populations that rely extensively on freshwater habitats for juvenile rearing. They are likely to face altered water temperatures and precipitation related changes in flow regimes. It is predicted that there would be major alternations in river hydrology resulting from earlier snow melt and increases in the elevations at which freezing occurs. As the flows of many salmon rivers are dominated by snowmelt, summer runoff may decrease and winter runoff may increase.

Climate change may impact salmon at a variety of life stages, beginning at egg incubation (Figure 2). Changes in water temperatures and river flows during this life stage may cause sub-optimal juvenile emergence from the gravel nests and result in increased egg and early juvenile mortality. Egg development is largely temperature dependent, as is the period following hatching when juveniles remain in the gravel for several weeks surviving on the nutrients in their yolk sacs before beginning exogenous feeding and emerging. The timing of emergence from the gravel has likely evolved in response to selection for a presumably optimal time for survival (e.g., temperature and flow), food availability and competitive advantages. If winter temperatures increase following climate change, egg incubation will be shortened and result in earlier juvenile emergence

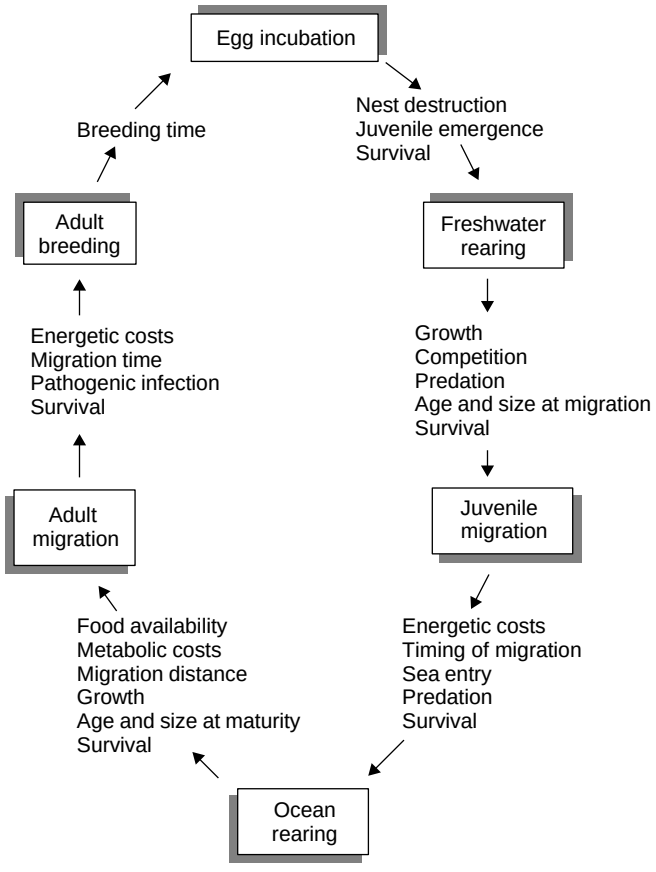


Figure 2 Potential impacts of climate change at various life stages of salmon

from the gravel, in the future. Such a shift away from the optimal timing for juvenile emergence is likely to result in increased mortality. Moreover, if climate change results in increased winter flows, as many models predict, gravel shifts caused by flooding may become more frequent and larger, increasing the susceptibility of incubating eggs to destruction. In addition, increased flows may occur during emergence, when the juveniles are highly susceptible to displacement.

The subsequent growth and survival of the juveniles in fresh water will be affected directly by temperature and flow regimes, and indirectly through changes in the aquatic community in response to the altered hydrology. Each salmon species has characteristic upper (usually below 30 °C) and lower (freezing) temperature limits for survival. Within these limits, there is an optimum temperature range over which feeding and growth occurs (e.g., 6 22.5 °C for Atlantic salmon), but this range declines as the availability of food declines. The effects of climate warming on growth of juvenile salmonids can thus be expected to differ with latitude.

At the southern edge of their distribution area, salmon growth may be retarded and mortality higher because of expanded periods with unfavorably high temperatures. This may be further compounded by the expected reductions in flow and thus habitat during summer months. Generally, stream flow is positively related to habitat space for, and thus abundance of stream fishes. Increased water temperatures may also alter the aquatic community structure, both in terms of the forage base for juvenile salmon and the competitor and predator communities. Expansion of the ranges of more southerly, warmer water communities is expected. Such potentially deleterious effects on growth may, in turn, influence over-winter survival and subsequent marine survival, both of which are size related.

In northern areas, on the other hand, growth rate will increase because of more favorable temperatures during summer and a longer growing season. However, droughts may be more frequent during summer than today, resulting in reduced habitat and food availability, and thus retarded growth. On the whole, however, freshwater growth is expected to be better and the age of migration to the ocean (smolting), which is growth related, lower than today. Moreover, because production of salmon smolts is inversely related to age of smolting due to annual freshwater mortality rates of 40 60%, the smolt production is expected to increase in northern rivers. Thus, losses in production of juvenile salmon in the south may well be overridden, at least initially, by gains in the north.

The spring photoperiod is considered the major environmental cue initiating the transformation from freshwater juvenile to seaward migrating smolt, whereas various river-specific stimuli (e.g., water temperature, water flow, and lunar cycle) regulate the timing of the seaward migration. Such river-specific stimuli likely reflect local adaptations that have evolved to insure an optimal timing of smolt migration in terms of survival and subsequent growth. For example, descending smolts are subject to heavy mortality from fish and birds in the river and estuary. To avoid such predation, smolts migrate down river in schools. If climate change results in reduced spring flows, as is predicted for many temperate rivers, the smolt run may be more dispersed over time and schools smaller in size, resulting in increased exposure to predation. Moreover, the energetic costs of migration may be higher and time taken to descend longer. Changes in river temperatures and flow regimes may also lead to sub-optimal timing of entry into the sea. Timing of sea entry is critical for the early sea survival of smolts because of physiological stress, metabolic costs, food availability and predation, all of which are influenced by sea temperatures.

Once in the ocean, salmon are likely to face further effects of climate change including altered sea surface temperatures, currents and areas of nutrient upwelling, all of which directly or indirectly affect their lives. It has been predicted that modest rises in sea surface temperatures, in line with predictions of global warming over the next half-century, could cause salmon to disappear from much of the North Pacific Ocean. During November March, for instance, sockeye salmon are only found in regions where the sea surface temperature is below 7 °C and thus show remarkably low thermal limits which excludes vast areas of the North Pacific that are otherwise potentially habitable. It is speculated that this is driven by interactions between water temperature and metabolic rate when food supplies are low. The salmon may be avoiding warmer water because food supplies are insufficient to maintain high metabolic rates, which increase exponentially with temperature in cold-blooded animals such as fish. In addition to a reduction in their habitat, a northern shift in its distribution will force salmon to travel farther to reach their breeding rivers, resulting in smaller fish with fewer eggs.

In the Atlantic, where salmon survival, particularly that of out-migrating juveniles, appears related to the areal extent of 8 10 °C water during spring, the situation may be more complicated. Sea surface temperatures may interact with changing currents to either decrease or increase area of the 8 10 °C water during spring and thus, salmon survival. Warmer sea surface temperatures in the Atlantic may also lead to a northward expansion of the range of salmon, as apparently occurred during 1919 1938 when a strong inflow of warm Atlantic water into the Kara Sea, Northern Russia, led to the eastward expansion of salmon.

Warmer sea surface temperatures are also likely to interact with currents to alter the marine communities exploited by salmon, reducing food availability (zooplankton biomass). It has been speculated that this, coupled with higher metabolic costs in warmer water, will lead to smaller

salmon, a prediction that is consistent with evidence that climate variation over the North Pacific Ocean is responsible for long-term (1925 1993) trends in the body size of sockeye salmon.

Upon return to fresh water during their breeding migration, adult salmon are likely to encounter increased water temperatures and decreased river flows associated with climate warming that may cause additional prebreeding mortality and reduced egg deposition. Increased water temperatures influence temperature-related metabolic processes, with energy expenditure during migration increasing exponentially with temperature. Moreover, because of a positive relation between temperature and pathogenic infection rate in salmonids for several bacterial pathogens, increased temperatures will likely increase the susceptibility of breeders to pathogens and parasites. Reduced water flows, in addition to affecting water temperatures, will also influence the ease of migration because it is a primary determinant of upstream fish passage. In general, lower flows create lower velocity conditions that enhance upstream fish passage. Many rivers, however, have obstacles (e.g., falls, barriers) that can only be passed during higher flow conditions. Thus delays in river ascent may occur, increasing the energetic expense for the adults, which do not feed during the breeding migration. Such low flows are also believed to select against large adult body size, which could result in reduced salmon size in many rivers.

Finally, breeding time in salmonid fishes appears adapted to present day thermal and flow regimes in such a way that conditions are favorable for breeding, egg incubation and juvenile emergence and subsequent feeding. Thus, climatic induced changes in thermal and flow regimes are likely to have deleterious consequences for one or more of these life history events that are intimately linked to breeding time.

Climate change is thus likely to impact the life cycle of salmon at a variety of stages (Figure 2). Salmon do have the ability to respond immediately to warming events, such as an El Nino, by altering migration patterns. However, this response is often either inappropriate, or insufficient to compensate for strong negative effects on growth and survival. Thus, while the response of salmon to climate change will, at least initially, involve environmentally mediated changes (i.e., phenotypic plasticity) in life history, it will likely require longer-term genetic change as well. The speed and ability of the life history to respond (through plasticity and genetic change) will be important in determining the extent to which salmon populations become compromised.

See also: Life Cycles, Volume 2; Fisheries: Paci c Coast Salmon, Volume 3.

FURTHER READING

Cox, S P and Hinch, S G (1997) Changes in Size at Maturity of Fraser River Sockeye Salmon (*Oncorhynchus nerka*) (1952 1993) and Associations with Temperature, *Can. J. Fish. Aquat. Sci.*, 54, 1159 1165.

Friedland, K D, Hansen, L P, and Dunkley, D A (1998) Marine Temperatures Experienced by Postsmolts and the Survival of Atlantic Salmon, *Salmo salar* L., in the North Sea area, *Fish. Oceanogr.*, 7, 22 34.

Holbrook, S J, Schmitt, R J, and Stephens, Jr, J S (1997) Changes in an Assemblage of Temperate Reef Fishes Associated with a Climate Shift, *Ecol. Appl.*, 7, 1299 1310.

Jensen, A J (1991) Possible Effects of Climatic Changes on the Ecology of Norwegian Atlantic Salmon (*Salmo salar* L.), *ICES C.M.*, 1991/M:34.

Levy, D A (1992) Potential Impacts of Global Warming on Salmon Production in the Fraser River Watershed, *Can. Tech. Rep. Fish. Aquat. Sci.*, 1889, 1 108.

Mangel, M (1994) Climate Change and Salmonid Life History Variation, *Deep-Sea Res.*, 41, 75 106.

Welch, D W, Ishida, Y, and Nagasawa, K (1998) Thermal Limits and Ocean Migrations of Sockeye Salmon (*Oncorhynchus nerka*) Long-term Consequences of Global Warming, *Can. J. Fish. Aquat. Sci.*, 55, 937 948.

Fisheries: Minamata Disease

Akira Kudo
Kyoto University, Japan

In the 1950s, an environmental disaster brought human misery to a sleepy shing village facing a small inlet called Minamata Bay, in Southwestern Japan. More than 100 people living around the bay died and many thousands more were permanently paralyzed after eating sh contaminated with mercury (methyl mercury, $CH_3\ Hg^+$), an effect known today as Minamata disease.

The Minamata disaster was the first known case of multiple deaths as a result of the natural bioaccumulation of a toxic material in industrial wastewater. It was different from other causes in which deaths followed direct consumption of contaminated water or food.

Minamata Bay s fatal mercury (most of which was in inorganic form, methyl mercury/$CH_3\ Hg^+$) was discharged from a nearby chemical factory in the 1930s 1960s, settling and accumulating in bottom sediments, at concentrations of up to 600 ppm. In 1976, the local Kumamoto prefecture estimated that 150 tonnes had been released into the 3 km^2 bay, where it remained and was chemically transformed naturally into more toxic (more than 1000 times for humans) organic mercury.

In the aquatic environment, mercury maintains an equilibrium between inorganic and organic chemical forms (Kudo *et al.*, 1977). In most cases, the ratio of organic to total mercury ranges from 0.5 to 5% and the rest (95 99.5%) is inorganic.

The ratio of mercury from the chemical factory in Minamata was no exception and measured as 0.5 1.5% methyl mercury (Irukayama, 1966). In other words, methylated mercury was constantly produced (transformed) at the bottom of Minamata Bay from large amount of inorganic mercury discharged from the factory. The production rate reached an equilibrium due to the dilution by tidal sea water and the natural destruction of methyl mercury into the inorganic form. This fact contradicts the still prevalent assumption by many Japanese, including the court decisions, that all methyl mercury was discharged directly from the chemical factory.

The reason why so many human casualties occurred around the Minamata Bay while many aquatic systems in the Western world contaminated by similar level of mercury have not resulted in any human loss lies in the unique aspects of Japanese culture. In Japan, a) fish is heavily consumed as the major protein source and b) harmony in the traditionally social structure is more important than the discovery of new scientific facts. Bureaucrats in Japan will usually take no action until a slowly-evolving consensus has been reached. Furthermore, there is a built-in-system, which makes each bureaucrat s responsibilities extremely vague and diluted.

The professors in the major Japanese universities are part of the bureaucratic system. Few resources were allocated for scientific investigation of the cause of Minamata disease, which proceeded slowly and was limited. During this fact-finding period, most of the casualties occurred. Surprisingly, the cause of Minamata disease was not found by analyzing mercury in the bottom sediments of the bay, but by a unique symptom of mercury poisoning recognized in clinical patients. In other words, the most experienced analytical scientists with modern equipment in Japan were not mobilized to detect Minamata disease. Furthermore, when the chemical company ceased to use mercury in 1968, most research on Minamata disease and on the mercury in the surrounding environment in Japan virtually ceased (see Figures 1 and 2).

The large amount of mercury discharged into a natural system, Minamata Bay, provides a rare opportunity to trace the mercury behavior in the aquatic system. At the same time both the local and central governments have held to their positions since 1960s that none of the mercury from the bottom of the bay has ever moved, escaped or been dispersed to the outside.

A group of scientists measured the dispersal of the mercury from the bay into the surrounding water body of the Yatsushiro Sea after 1975. What really happened to the mercury deposited in the bay prior to 1968? Did it escape into the Yatsushiro Sea? If so, what is the increase of mercury in the surface sediments of the sea? Over 25 years of investigations have produced answers on mercury behavior in the natural environment, which throw doubts on official statements.

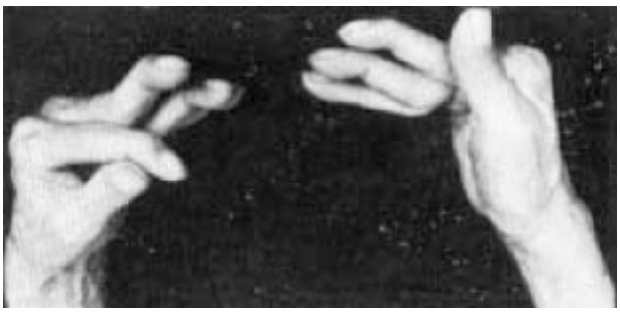


Figure 1 Fingers deformed by Minamata disease. (Photo from Dr M Harada, Kumamoto University)

Figure 2 Decontamination project removing contaminated bottom sediment at Minamata Bay. (Photo from Kumamoto Prefecture)

During the last 25 years, surface sediments have been collected from 24 fixed sampling stations to a depth of 4.0 cm in the Yatsusiro Sea. Researchers have found that the mercury content of the samples was not a cause for concern and that the concentrations outside the bay were below the acutely toxic level. At the same time the investigators spotted a trend in mercury behavior in the 633 km^2 Sea (Kudo and Miyahara, 2000).

Between 1975 and 1984 mercury concentrations gradually increased, with sampling stations close to Minamata Bay showing the greatest rise, from 0.62 ppm in 1975, to 10.3 ppm in 1984. At another, more remote, station, concentrations rose slightly. The pattern was interrupted in 1982, an exceptional year when 400 mm of rain fell in three hours, a 200-year event. Millions of tonnes of clean soils were washed from surrounding mountain slopes to cover the sediments in the Yatsushiro Sea.

Three years after this storm, the government initiated a decontamination project. Mercury concentrations at one sampling station near the bay dropped from 10.5 to 1.56 ppm from 1985 to 1990 and decreased further to a background level of 0.3 ppm in 1997.

The Japanese government eventually took responsible action against mercury pollution in the bay, but it was too slow to act. The decontamination project costing US$500 million and the first such effort in the World, will clean up the bay and protect the people around Yatsushiro Sea for many years to come.

REFERENCES

Irukayama, K (1966) The Pollution of Minamata Bay and Minamata Disease, *Water Sci. Technol.*, 3, 153 165.

Kudo, A *et al.* (1977) Equilibrium Concentrations of Methyl Mercury in Ottawa River Sediments, *Nature*, 270, 419 420.

Kudo, A and Miyahara, S (2000) History of Mercury Migration from Minamata Bay to the Yatsushiro Sea, *Water, Sci. Technol.*, 42, 177 184.

Fisheries: Pacific Coast Salmon

David Welch

Pacific Biological Station, Nanaimo, Canada

The eight species of Paci c salmon (genus Oncorhynchus) form one of the most valuable sheries resources in the world, and have high value to the peoples of the Paci c Rim as an indicator of the health of the natural ecosystem. Paci c salmon are particularly vulnerable to climate change because their complex life histories involve extended periods of life in both the freshwater and marine ecosystem, exposing them to climate disruptions in both habitats.

Climatic changes that disrupt the life cycle of Paci c salmon and reduce the chance of successful breeding are especially serious, because in most species, adults die after breeding. This makes the persistence of populations dependent on successful completion of the life cycle to a much greater degree than in animals where adults have the opportunity to reproduce more than once. As a result, once sexual maturation starts and salmon begin their long migration back through the sea and up the rivers, events that disrupt breeding success leave no chance for the animals to leave and return to breed the next year. In addition, substantial differences in behavior between populations, such as in the timing of egg development, hatching, and spawning, as well as strong delity to the spawning grounds, are believed to be inherited and the result of strong selection to their local environment. Such characteristics may be particularly vulnerable to disruption by climate change.

All species of Pacific salmon begin life as fertilized eggs laid in nests dug out of the gravel or cobble bottoms of lakes and rivers. The embryos then develop over the winter before hatching. Following hatching, the fry emerge from the gravel in the spring. In pink and chum salmon, the fry quickly leave freshwater and enter the ocean. In the remainder of the species, the young salmon make much more extensive use of the freshwater habitat (rivers and lakes) for one or more years before migrating to the sea.

After entering the ocean, a period of time may be spent in estuaries or near-shore habitats where adaptation to the marine environment is completed and rapid growth begins. Most salmon then begin a rapid and highly directed northwards migration along the narrow continental shelf, where most eventually leave the coastal zone and remain in the open ocean for a period of several years. As sexual maturation begins, salmon begin directed homing migrations over long distances in both the ocean and freshwater. Much less is known of the marine phase of the life history than the freshwater phase, but evidence accumulated over the past decade suggests that changes in ocean survival have been at least as serious as changes in freshwater survival.

Most growth and approximately half of the mortality occurs in the ocean. As a result, climatic disruptions that happen at any point in the life history can result in severe impacts on survival rates. Freshwater migrations are affected by changes in water temperature, and rate and timing of river flow. Climate change projections generally indicate that winters will be warmer and wetter in future, while summers will be warmer and perhaps drier. With more precipitation falling as rain rather than snow, the spring run-off is expected to occur earlier and be more intense in future, while river flows in summer will probably be warmer and less rapid, providing less moderation of summer temperature extremes. Young fish migrating to sea may be affected by changed timing of river flows, or mismatch with peaks in ocean production. Adult salmon returning to breed may experience strong river flows and high water temperatures that drain their energy reserves to the point that they are unable to reach the spawning grounds. The unusual warming experienced in the 1990s saw much higher mortality of adult salmon returning to many parts of British Columbia s Fraser River. This was apparently a result of high temperature and stream flows increasing the energy demands on migrating adults, while reduced growth at sea resulted in them returning at smaller sizes and with reduced energy reserves than in the past.

In the North Pacific, elevated temperatures may ultimately reduce salmon populations by limiting their ocean distribution. All species of Pacific salmon sharply avoid warmer temperatures at sea, effectively limiting them to a substantially smaller area of the open Pacific Ocean that they could otherwise occupy. Global warming models project temperature increases that could exclude several species of salmon from the Pacific Ocean within 50 years at expected rates of greenhouse gas increase (see Figure 1). Increasing temperature may also change the structure of the ocean through warming of the surface layer. Warming will increase the temperature contrast with the deep ocean, and may reduce mixing and restrict the input of essential nutrients necessary to fuel the food chain. Such an effect has already been observed in the 1990s, with the thinning of the surface layer resulting in a switch to a nitrate-depleted ecosystem in the Gulf of Alaska for the first time on record.

Several lines of evidence suggest that salmon are strongly adapted to the climate of the relatively recent past. As concentrations of greenhouse gases increase to levels not seen for hundreds of thousands of years, current adaptations may not serve salmon as well in the future. Many critical aspects of the life cycle are completed during brief time periods (e.g., timing of egg hatch, ocean entry for young salmon, breeding of adults). It is thought that evolutionary forces were strongly selective for individuals with specific characteristics appropriate for each population, and that animals deviating from these characteristics did not survive to pass on different behaviors.

Climate change is unlikely to be favorable for most salmon populations, particularly at the southern end of their range, where human populations are high and the value placed on persistence of salmon is greatest. Climate disruption is likely to exacerbate conflicts with other resource users. Water draw-down to support agriculture, and more restrictive land-use regulations on forestry or urban development to protect salmon, brings these competing resource sectors into sharp conflict as Pacific salmon populations become less productive. Hydroelectric dams one of the few energy sources that do not involve generation of greenhouse gases have also been blamed for affecting salmon populations by changing the natural flow of rivers and affecting the survival of both young and mature salmon migrating past the dams. In addition, many fisheries have either explicit legal entitlements (such as treaty rights

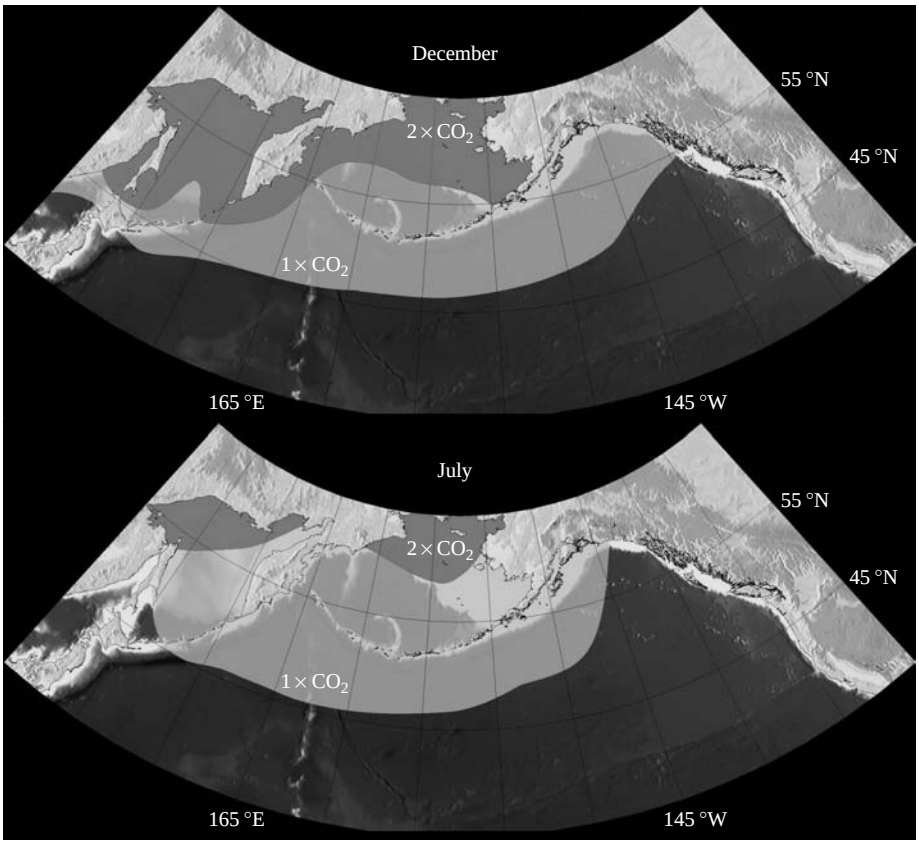


Figure 1 Current ocean distribution of sockeye salmon ($1 \times CO_2$) and projected future distribution if greenhouse gas concentrations double ($2 \times CO_2$). The lines enclose the area of the North Pacific that research shows salmon feed in, and the region of the ocean that may have similar temperatures in future. (Reproduced from Welch *et al.*, 1998)

granted to native fishermen) or implicit claims to the salmon resource based on past access.

Despite these concerns, changes in climate will not be uniformly bad for Pacific salmon. Continued warming, for example, will moderate the harsher climates in northern regions, likely improving conditions for salmon. A few sexually mature Pacific salmon have been reported from the Canadian high arctic in recent years, suggesting that the range of salmon is expanding to the north as climate has warmed. The productivity of many Alaskan salmon populations has also increased over the last few decades as salmon populations in southern regions have fallen. However, it is unlikely that the increased economic gain from the salmon catches in northern regions has outweighed the economic costs of trying to maintain salmon populations in southern regions. At least in North America, recent climate warming has had decidedly mixed effects, and the increasing disparity has resulted in bitter argument over salmon conservation.

See also: Life Cycles, Volume 2; Fisheries: Effects of Climate Change on the Life Cycles of Salmon, Volume 3.

REFERENCE

Welch, D W, Ishida, Y, and Nagasawa, K (1998) Thermal Limits and Ocean Migrations of Sockeye Salmon (*Oncorhynchus nerka*): Long-term Consequences of Global Warming, *Can. J. Fish. Aquat. Sci.*, 55, 937 948.

FURTHER READING

JISAO/SMA Climate Impacts Group (1999) Impacts of Climate Variability and Change in the Pacific Northwest, University of Washington, 1 109.

Levy, D A (1994) *Potential Impacts of Global Warming on Salmon Production in the Fraser River Watershed*, Climate Change Digest No. 94-04, Environment Canada, 1 15.

Fisheries: Pollution and Habitat Degradation in Tropical Asian Rivers

David Dudgeon
The University of Hong Kong, Hong Kong

Anthropogenic in uences threaten the biodiversity of rivers and their associated wetlands at global, regional and local scales, and will impair or signi cantly reduce ecosystem services. Threats to these habitats can be categorised as follows:

1. *Flow modi cation or regulation (impoundment by dams, water extraction for irrigation, etc.);*
2. *Pollution;*
3. *Destruction of instream and riparian habitat (by deforestation and drainage-basin alteration);*
4. *Over-harvesting (mainly of shes);*
5. *Global climate change.*

Freshwater shes, especially those in large tropical rivers, are relatively vulnerable to extinction, and around 20% of the global total is endangered or extinct. Asian river shes are affected by all categories of threat listed above. Deforestation has occurred or is ongoing throughout much of tropical Asia; pollution is ubiquitous and subject to little effective control; and, dam building changes patterns of ow and inundation, blocks migration routes, and fragments populations. Native sh communities will be in uenced also by the spread of exotic species. We lack unequivocal data on the magnitude of their effects, but such invasions are unlikely to be positive. Projected changes in global climate can be expected to interact with anthropogenic threats, especially in rivers where the magnitude and timing of peak and minimum ows may be altered. The effects of such synergy are dif cult to predict, in part because the relative intensity of each threat varies among rivers. This constrains our ability to protect riverine environments. A particular concern is the link between rivers and their oodplains, mediated by the annual cycle of ooding caused by monsoonal rains. Most sheries production along tropical Asian rivers depends on oodplain inundation. Where this is not restricted, yields are high. The lower Mekong Basin, for example, generates an annual catch of 500 000 t, providing around half the protein intake of human inhabitants in that region. In Bangladesh, where oodplains and the Ganges- Brahmaputra Delta occupy about 80% of the land area, ve million people depend on shing for their livelihood. The annual harvest can reach 725 000 t. Fisheries like this are threatened by engineering works intended to mitigate ood damage and prevent river erosion of agricultural land. As seasonally ooded areas diminish, so too does the sh production, which depends upon these areas. Pollution and over shing are signi cant sources of concern also. This article describes the ecological features that make Asian river shes particularly vulnerable to environmental degradation, delineates the present trajectories of change (plus the possible effects of climate alteration on river ows), and sets out the consequences of impact. The prognosis is grim. If current trends continue, sh biodiversity will decline throughout the region.

ECOLOGICAL CONTEXT

Approximately 300 genera of freshwater fishes occur in tropical Asia (for details, see Kottelat, 1989). The family composition of this rich fauna is substantially different from the Neotropics and Africa. Species richness within rivers is high: over 500 species occur in the Mekong River (ranking third in the world for fish diversity after the Amazon and Zaıre); the Kapuas River (Borneo) supports 290 species; the Zhujiang (Pearl River; southern China) has 262; the Chao Phraya (Thailand), 222; the Salween River (Thailand), 150; and, the Mahakam River (Borneo), 147. Nevertheless, inventories of the Asian fish fauna are far from complete. Few countries have made systematic assessments of biodiversity, and published inventories are sometimes based on dated literature or old museum collections instead of recent surveys. One consequence is a critical shortage of information about the conservation status of Asian freshwater fishes, even in countries where species lists are fairly complete. Thus extinctions may not be evident until long after they have occurred.

What is the ecological backdrop for fish biodiversity in tropical Asian rivers? Some of the greatest rivers in the world (ranked by order of discharge) are located within the region: the Chang Jiang (Yangtze River), Brahmaputra, Ganges, Irrawaddy, Zhujiang, Mekong and Indus. Natural lakes, apart from those associated with river floodplains, are of relatively minor importance, and are far exceeded by the extent of freshwater swamp. The ecology of tropical Asian rivers has been reviewed by Dudgeon (1999, 2000 and references therein) who emphasizes the importance of discharge seasonality on the biota and land water interactions. The predominant influence of monsoons generates a characteristic pattern (typified by the Mekong, Figure 1) whereby predictable periods of drought and water scarcity during the dry season alternate with intervals of increased

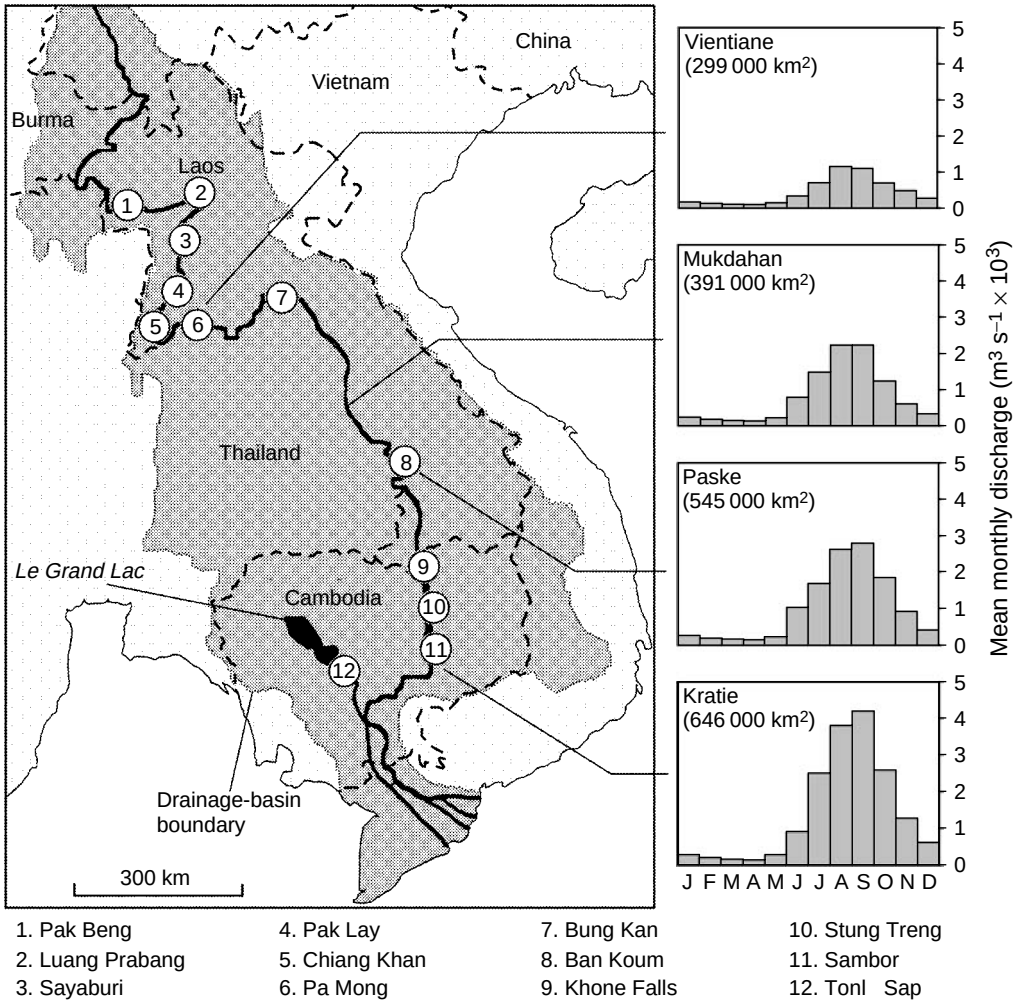


Figure 1 Locations of proposed hydropower projects along the Mekong mainstream in Laos, Thailand, and Cambodia. River discharge at four stations and the drainage basin area of each station are shown also. (Data from van der Leeden, 1975 and Roberts, 1995)

discharge, spates and flood-plain inundation during the wet season. One result is alternating periods of resource scarcity (during the dry season) and resource glut (during the wet season) for most fishes. The monsoon drives flood pulses (i.e., the predictable advance and retraction of water over the floodplain), and fish life histories (which often involve migrations) are timed to exploit land water interactions and energy inputs that occur during the inundation phase. For this reason, much of the fisheries yield from large rivers derives from floodplain habitats (Welcomme, 1976).

The Mekong is a good example of the interaction between fish ecology and discharge. That river s major floodplain lake Le Grand Lac of Cambodia (Figure 1) receives water from the Mekong (via the Tonlé Sap River) during the wet season, but water drains from it in the drier months. The lake fluctuates greatly in extent and depth. From a minimum of around 2500 km^2, the lake may swell to cover 15 000 km^2, inundating tracts of forest. Fish enter the flooded area to feed and breed, and this seasonal inundation sustains the lake fishery. Mekong fishes undertake upstream breeding migrations during the wet season as water levels rise; downstream migrations occur when levels fall during the dry season. The timing of fish migrations differs somewhat according to species and among different parts of the Mekong Basin, but this complex situation is amenable to generalization (Welcomme, 1979). Planktivorous or piscivorous pelagic fishes (whitefishes) cover a greater distance when migrating than bottom-dwelling species (blackfishes). The latter feed on benthic organisms, and make shorter, lateral movements between the main channel and floodplain or inundated forest. The two groups differ with respect to breeding habits: many blackfishes are multibrooded and practice parental care (mouthbrooding, bubble-nest building); they may have accessory breathing organs and can tolerate deoxygenation of stagnant floodplain waters. Whitefishes avoid severe conditions by long-distance migration, and are egg scatterers. Typically, they breed only once per season. An intermediate, relatively species-poor assemblage (greyfishes) have both migratory and static/territorial behavioural components, enabling them to respond facultatively to changing flow conditions. They are generalists, and may thrive where the natural flood regime has been modified.

The characteristic situation in Asian rivers involves fishes using various habitats at different times. Some of these habitats are inundated for part of the year only. Fish must therefore range throughout the river system in response to food availability. Whatever the distance covered, most species depend upon the floodplain for food and breeding sites and are subject to vagaries in the intensity of the monsoon. Thus in some years there will be extensive flooding and relatively high fish production. In years when a smaller area is inundated, production will be lower. While there is some noise in the system introduced by inter-year variation in flood peaks, it is evident that the flood-pulse is not a disturbance in the ecosystem (although human inhabitants of low-lying lands may view it as such). Instead, significant departures from the typical pattern of seasonal flow fluctuation can be regarded as a disturbance. This sets the scene for conflict between the needs of humans and fishes. The engineering response to flow seasonality is to capture and store water during flood times for use during the dry period, the benefits being amelioration of peak discharge (i.e., flood prevention) and a predictable water supply during the rest of the year. In effect, natural flow variability to which fishes are adapted and upon which they depend is smoothed out. This averaging of fluctuating flows is significant because ecological responses to environmental change are usually non-linear. Fish breeding migrations, for example, may not begin until flows have passed a critical threshold; thus a 25% increase in discharge may fail to initiate any population response.

DEFORESTATION AND DRAINAGE-BASIN DEGRADATION

While fishes in tropical Asian rivers are adapted to seasonal flow fluctuations, natural discharge regimes do not always pertain in unregulated rivers. The consequences of drainage-basin misuse and deforestation are evident throughout the region, including increased runoff, sedimentation and flash floods (Dudgeon, 1999 and references therein). In China, for example, devastating floods along the Chang Jiang in 1998 (and preceding years) focused attention on the link between catchment conditions and runoff. Government blamed the floods on years of uncontrolled logging in upland catchments. Similar consequences of drainage-basin degradation led to a logging ban in Thailand after devastating floods in 1988 claimed several hundred lives.

The loss of forest cover in tropical Asia (estimates range from 0.9 2.1% pa, but local variation is substantial) has important implications for river conservation since a significant proportion of the rich fish fauna of the region exploits inundated forest where fruit and terrestrial insects may be significant food sources. For example, forest clearance around Le Grand Lac has led to declining fish catches through a combination of reduced food, erosion and siltation. Silting causes excessive turbidity as particles are resuspended by waves in the increasingly shallow water. This limits primary production, and results in lower fish populations. Forest clearance also threatens an assortment of specialised Southeast Asian fishes confined to soft, acidic, blackwater streams which drain peat swamps. These habitats are favoured sites for timber extraction. Many blackwater fishes are severely threatened, and will become extinct (as has occurred already in parts of Malaysia and Indonesia) if logging and degradation of their habitats continues.

Deforestation in Asia is a major regional issue which, when coupled with seasonally variable river flows, has direct and profound impacts on floodplain residents. Given the extent of human misery caused by annual floods along major Asian rivers (most of which drain degraded catchments), the desire to limit peak flows by trapping floodwaters behind huge dams is understandable. In essence, greater frequency and intensity of flood flows are an inevitable result of deforestation. In turn, floods encourage the construction of protective dams that further alter the natural hydrological regime. The trend in Asia during recent years has been towards more and bigger dams, culminating in the world s largest dam, the Three Gorges Scheme on the Chang Jiang. Approximately 65% of the global total of large dams ($\geq$15 m tall) is situated in Asia; China has almost half of these, plus 10 major dams ($>$150 m tall). India (second in Asia and fifth in world ranking) has over 1100 large dams, and seven major dams. Dam building is not solely a response to floods. Other benefits include storage of irrigation water, and generation of hydroelectric power. The Mekong, for instance, has vast potential generating capacity. In 1994, the Mekong Commission (an autonomous coordinating committee set up by the riparian states) identified 12 sites for dams (most with planned outputs $>$10 000 MW) along the river mainstream in Laos, Thailand and Cambodia (Figure 1).

RIVER REGULATION AND DAMS

The planned construction of a dam array along the Mekong mainstream provides a good example of the potential effects of river regulation on fishes. Changed floodplain inundation patterns, combined with altered flow and temperature regimes, may suppress fish migration and breeding. Mainstream dams will obstruct the passage of long-distance whitefish migrants, including species (such as the pangasiid catfishes) that form the basis of Mekong wild-capture fisheries (for details, see Roberts, 1995). The greatest impacts of the dams will be felt by species confined to the Mekong. Among them are *Pangasius gigas* and *P. kremp* (Pangasiidae), the carps *Probarbus labeamajor* and *Aaptosyax grypus* (a 1.3 m long predator!), and the endangered freshwater herring *Tenualosa thibaudeaui*. They could be driven to extinction by a single mainstream dam. The life history of *Pangasius gigas* the giant 3 m Mekong catfish is obscure, but journeys upstream to Yunnan Province (China) have been mooted. This catfish was once the basis of an important fishery in Cambodia because oil could be extracted from its fatty flesh. Populations began to decline decades ago, and the fishery has collapsed in recent years.

One major dam proposed for the Mekong Basin would block the Tonlé Sap River, connecting Le Grand Lac to the Mekong (Figure 1). A dam on the Tonlé Sap River would impact fish production through changed floodplain inundation patterns, and exacerbate ongoing degradation and sedimentation in Le Grand Lac (see above). Elsewhere in the Mekong Basin, evidence of damage to fisheries is evident at the Pak Mun Dam, Thailand. Fish populations declined from 1991 (when dam construction began) until 1994 (when it was finished). Construction of Nam Theun Hinboun on the Theun River (a Mekong tributary in Laos) began in 1994, notwithstanding a prediction that the river below the dam would be reduced to a series of pools during the dry season with devastating consequences for many of the 140 fish species of the Theun Basin. Other impacts of dams result from the effects of thermal stratification, which occurs in impoundments behind large dams in tropical Asia. Water released from the hypolimnion is cool and oxygen-poor, or may be anoxic containing toxic hydrogen sulfide, with predictable consequences for the downstream biota.

The detrimental effects of dams on river fish are often ignored because of the view that impacts are more than compensated by fisheries in newly created impoundments. Such optimism is based upon a lack of understanding of fish ecology. Species adapted to the fluctuating discharge of a river will not thrive in a stagnant impoundment. Where high fisheries yields have been reported from Asian reservoirs, they depend on cage culture, stocking (usually of major carp), or introduction of exotic species. Such practices are invariably accompanied by a loss of native biodiversity.

Elsewhere in Asia, the conspicuous effects of mainstream dams on migratory river fishes are seen among the Acipenseriformes (sturgeons and paddlefish) of the Chang Jiang the largest river in Asia. Spawning migrations of the Chinese sturgeon (*Acipenser sinensis*) were blocked by the Gezhouba Dam (or Chang Jiang Low Dam) in 1981; fish passages were not provided. This species is anadromous, i.e., it spends part of its life in the sea but swims into rivers to breed. Flow alteration and changed sediment characteristics have reduced spawning success of Chinese sturgeons stranded below the dam. The Gezhouba Dam fragmented populations of the endemic Yangtze sturgeon (*Acipenser dabryanus*) also. It is now virtually extinct in downstream reaches, because it is a potamodromous species (i.e., it undertakes breeding migrations along but within the river) and populations below the dam are unable to complete their upstream breeding migration. Sedimentation has impacted stocks of remaining populations of the Yangtze sturgeon upstream of the dam. The anadromous Chinese paddlefish (*Psephurus gladius*: Polydontidae) has declined drastically in the Chang Jiang because the Gezhouba Dam prevents access to upstream spawning sites, and this species (one of two polydontids on the planet) is liable to become extinct. Furthermore, *Myxocyprinus asiaticus* the only catostomatid (sucker) species in the Old World became rare after completion of the Gezhouba Dam. The Three Gorges Dam may further jeopardize this, and other, potamodromous fish.

More than 3000 dams have been built along the Zhujiang (China s second largest river) during the last 40 years. They have obstructed migrations by *Clupanodon thrissa* (Clupeidae), and stocks of major carp (including *Cirrhinus molitorella*) have fallen to levels where they no longer sustain a viable fishery (Liao *et al.*, 1989). The abundance of the Chinese sturgeon has been greatly reduced, which has conservation implications as this fish is confined to the Zhujiang and Chang Jiang. In the Dongjiang, a major tributary of the Zhujiang, catches of the anadromous clupeid *Tenualosa reevesii* during the 1980s were only 20% of those 20 years earlier; it has been eliminated from parts of the river. Landings of the related *Tenualosa (= Hilsa) ilisha* dwindled to virtually nothing after completion of the Farakka Barrage on the Ganges in 1975. Similar consequences followed obstruction of migration routes by the Chenderoh Dam on Sungai Perak, Peninsular Malaysia. Between 1971 and 1980, catches diminished (Figure 2) and the availability of commercially important fishes fell from nine species to three. Among them, *Probarbus jullieni* (Cyprinidae) occurs in only one other river in Malaysia; it has also been eliminated from the Chao Phraya River (Thailand), and is threatened by dams along the Mekong. The impacts of river regulation involve not only barriers to longitudinal migration but also limitations on lateral movements. Fisheries in the Jamuna River (Bangladesh) have declined by more than 50% since 1983 because flood-control embankments reduce fish access to floodplains. Similarly, the major carp fishery of the Ganges practically disappeared after seasonal inundation of the floodplain was blocked. Along the Indus too, irrigation schemes have led to diminished fish stocks because of inadequate consideration for their need to pass dams, weirs and barrages.

Fish ladders which can mitigate the impact of dams on migrants are not widely used in Asia, and development of passage facilities has scarcely been addressed.

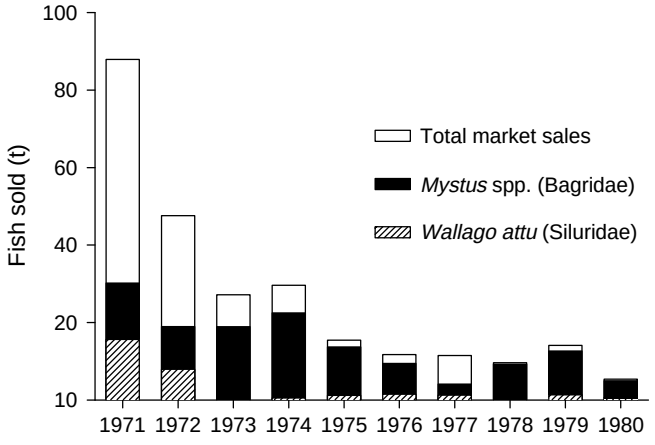


Figure 2 Market sales of river fishes in Perak State, Peninsular Malaysia, 1971–1980. (Data from Khoo *et al.*, 1987)

Fish ladders seem likely to have limited success because most Asian fishes do not jump. Moreover, ladders designed for species (such as north-temperate salmon), which swim upstream, do not allow for the fact that potamodromous Asian fish make downstream migrations as adults. Fishes on return migrations must pass through the dam turbines. Nevertheless, fish passes have been effective in facilitating migration past some dams on rivers in China (Jiangsu Province). Other conservation initiatives include opportunities arising from the Bangladesh National Flood Action Plan to manage capture fisheries through controlled flooding. Potential strategies include the design of flow regulators operated to give adults of major carp access to spawning grounds. Appropriate application of controlled flooding or drainage could permit the passage of water containing carp spawn and fry in a fish friendly manner.

POLLUTION

Water quality varies considerably throughout tropical Asia, but the region s major rivers are among the most degraded in the world. Anti-pollution legislation is in place in some countries, but is weak and/or poorly enforced (Dudgeon *et al.*, 2000). Pollution from agriculture and non-point sources is largely uncontrolled and, as cities and industries expand without adequate waste-treatment facilities, rivers receive increasing quantities of effluent. Many rivers are oxygen-poor for much of the year, but habitat deterioration is especially apparent during the dry season when water flows are insufficient to dilute pollutants. In some instances e.g., Peninsular Malaysia some rivers are so severely degraded that fisheries have collapsed. The situation is especially bad in China where, by 1995, some 80% of the 50 000 km of major rivers were too polluted to sustain fisheries (FAO, 1999). Fish kills (due to industrial wastewater, pesticides, etc.) occur frequently during the dry season, and fish have been entirely eliminated from more than 5% of total river length. Lesser tributaries tend to be more heavily polluted than the main channels (where dilution of contaminants is greater), but even in the mighty Chang Jiang (which once produced 70% of China s freshwater catch), landings fell by half (from 458 000 to 226 000 t) between 1954 and the late 1970s (before the Gezhouba Dam was completed) and have declined ever since. Things are no better in India. The Ganges ranks third in the region by flow volume, yet is probably the most polluted large Asian river. Population densities in the catchment are high, and 25% (600 km) of the river course is grossly polluted by human and animal wastes, as well as industrial effluent. Fish kills are common. Most industries discharging into the river lack waste treatment plants in working order, and some cities along the river (e.g., Varanasi) lack sewage treatment facilities entirely.

In short, most of the major rivers of Asia, as well as a host of minor streams, are grossly polluted. The impacts on the aquatic biota are manifest, and do not require emphasis here. Unfortunately, the political will to control pollution in Asia appears weak, and is likely to remain so as long as economic growth and the attitude of develop now, clean up later remain priorities for Asian governments.

OVERFISHING

Asia is responsible for around 64% of the global total of inland fisheries landings, although it only constitutes 20% of the world continental area (FAO, 1999). China, India, Bangladesh and Indonesia are the top four most important inland fishery countries, with Thailand ranking seventh. What does this indicate about the state or significance of river fisheries? Available figures do not separate yields of aquaculture-based fisheries and enhanced fisheries (where stock are added to water bodies) from natural capture fisheries, nor do they distinguish reservoir from river fisheries. Moreover, the dispersed and informal nature of many fisheries leads to under-reporting of landings. Subsistence fisheries are rarely included, and rivers may yield twice as many capture fishes as are actually reported. Nevertheless, one thing is clear. High catches in Asia reflect heavy exploitation of virtually all freshwaters, plus a substantial long-term increase in the contribution of aquaculture-based and enhanced fisheries to the total. This supplementation has permitted increased yields from certain Southeast Asian countries (Indonesia, Thailand, Laos: Figure 3), although this upward trend has not occurred everywhere (e.g., Vietnam, the Philippines).

There are many anecdotal accounts of declining stocks, but detailed fisheries statistics for Southeast Asian rivers are surprisingly scarce. Some monitoring of stocks would seem prudent. Laos, for example, is land-locked, and must depend entirely upon freshwater fisheries. Despite considerable inter-year variation, capture fisheries based on wild stocks appear to be at or exceeding the limits of sustainable yield. Declines in the average size of fish landed, attainment of maturity at a smaller body size, and reductions in catch per unit effort are among the signs that these fisheries are overexploited. Further consequences include altered community structure, diminished production, and damaged stocks.

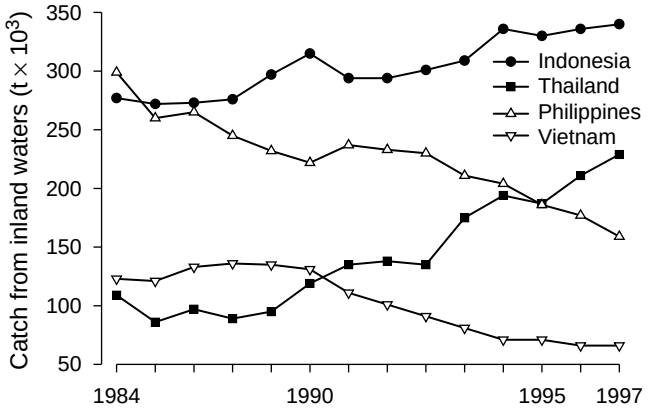


Figure 3 Annual inland fisheries landings, 1984–1997, in selected Asian countries, including yields from aquaculture/enhanced fisheries as well as natural capture fisheries. (Data from FAO, 1999)

The effects of overfishing first become manifest among large species. Populations of the 1 m long cyprinid, *Probarbus jullieni*, in Peninsular Malaysia were devastated by capture of egg-bearing individuals during upstream migrations (before passage was blocked by the Chenderoh Dam, see above). Reductions in *Probarbus jullieni*, *P. labeamajor* (which reaches 80 kg), *Catlocarpio siamensis* (120 kg, formerly the basis of an important fishery), and *Pangasius gigas* (300 kg), apparently due to overfishing, have occurred in Laos, Cambodia and Thailand. Similarly, overfishing in east Sumatran rivers threatens freshwater shark (*Wallago attu*: Siluridae) which has been exterminated over part of its range (see also Figure 2). Some of the reduction in Chinese paddlefish populations in the Chang Jiang (see above) is attributable to overfishing also.

Dramatic declines in capture fisheries such as a 40% reduction in landings by Laotian artisanal fishers between 1984 and 1992 may reflect the use of explosives. Increased availability and use of synthetic chemicals, rather than natural poisons (such as rotenone which must be extracted from plants), may be a contributing factor. Motorboats and refrigeration technology also allow the exploitation and marketing of previously inaccessible stocks. Significant declines in the major carp recruitment in the Zhujiang reflects the use of ever-finer meshed nets and overfishing of spawning grounds, combined with the application of poisons, explosives and electric shocking. Overfishing cannot, however, be separated from the damaging effects of dams and pollution in the Zhujiang, where toxins cause periodic fish kills. Similar caveats apply to rivers such as the Chang Jiang (see above).

CLIMATE CHANGE

The effects of global climate change on flow regimes in rivers is of considerable significance because the influence of water quantity on the life history of tropical fishes seems to parallel the importance of temperature to fishes in temperate latitudes (Meisner and Shuter, 1992). Climate change scenarios derived from general circulation models (GCMs) of the atmosphere suggest that increased carbon dioxide concentrations (and aerosols) would alter rainfall across the tropics. Different GCMs predict different directions and magnitudes of change. Wet seasons may get wetter

or drier; likewise, dry seasons may get wetter or drier. One recent study projects changes in wet-season rainfall over tropical Asia in the range of −5% to +18% (McLean *et al.*, 1998); anticipated dry-season changes are smaller and less consistent. A greater frequency of heavy rainfall or extreme precipitation events can be expected also (e.g., Bhaskaran *et al.*, 1995; Bhaskaran and Mitchell, 1998). Specific predictions for the Mekong include greater seasonal differences in flows and delays in the flood peak (Riebsame *et al.*, 1995); a later onset of the monsoon is seen in a GCM developed for India also (Bhaskaran *et al.*, 1995). Higher temperatures will increase evapotranspiration and may reduce dry-season flows; the magnitude of these projected changes in discharge vary linearly with rising temperatures. Note that the magnitude of annual river discharge may not change under some GCM scenarios because wet-season increases could offset lower flows during the dry season. Nevertheless, because extreme flow events will become more frequent, climate change may stimulate additional dam construction for flood control and irrigation. Substantial uncertainties are added to predictions of climate change and altered river discharge derived from GCMs because of likely interactions with the effects of widespread deforestation in Asia (Fu *et al.*, 1998; Vorosmarty *et al.*, 1998). If wet-season runoff increases (as predicted by most GCMs), sediment loads in rivers will rise, and deforestation will exaggerate this effect.

Regardless of the details of GCM projections, changes in discharge magnitude and seasonality of tropical Asian rivers will alter the extent of floodplain inundation and degrade the quality of fish habitat during non-flood periods. The implications for biodiversity are significant because fish catches (and hence secondary production) from tropical rivers are positively correlated with the extent of inundated floodplain, and fluctuations over time occur in response to inter-annual variations in flood peaks (Welcomme, 1979). The projected increase in frequency of extreme flow events will therefore tend to increase the variability of fisheries yields from rivers. If additional dams and flood-control projects are built in response to climate change, their impacts on fishes (and fisheries) will be detrimental. However, the magnitude of that effect is hard to estimate because, in most instances, the influence of existing structures has not been assessed in detail. Dams appear to have had deleterious effects on ecosystems, so we can assume that further discharge alterations caused by climate change, or the construction of additional dams, will reduce biodiversity further.

CONCLUDING REMARKS

Threats to fishes, generated by a combination of anthropogenic factors, have important implications for riverine biodiversity throughout tropical Asia. The limited data on fisheries landings indicate declines in natural capture fisheries. While figures collected from markets underestimate total yields, long-term population trends are unlikely to be affected by this bias. Data on fisheries in individual rivers or drainage basins, reported at the level of species or species groups, are needed for ecological interpretation of ongoing environmental change. Their lack limits our ability to take remedial action or mitigate anthropogenic impacts. Overfishing is clearly a contributory factor in reduced yields and is, in theory, amenable to regulation or control. However, environmental degradation and loss of fish habitats is a critical problem throughout much of Asia, and is the overriding threat to riverine biota. Reversal of prevailing trends is a major, urgent challenge for the 21st century, if we want to ensure a future for Asian river fishes.

REFERENCES

Bhaskaran, D and Mitchell, J F B (1998) Simulated Changes in Southeast Asian Monsoon Precipitation Resulting from Anthropogenic Emissions, *Int. J. Climatol.*, 18, 1455 1462.

Bhaskaran, D, Mitchell, J F B, Lavery, J R, and Lal, M (1995) Climatic Response of the Indian Subcontinent to Doubled CO_2 Concentrations, *Int. J. Climatol.*, 15, 873 892.

Dudgeon, D (1999) *Tropical Asian Streams: Zoobenthos, Ecology and Conservation*, Hong Kong University Press, Hong Kong.

Dudgeon, D (2000) Large-scale Hydrological Changes in Tropical Asia: Prospects for Riverine Biodiversity, *BioScience*, 50, 793 806.

Dudgeon, D, Choowaew, S, and Ho, S C (2000) River Conservation in Southeast Asia: Conflicts, Constraints and Compromise, in *Global Perspectives on River Conservation: Science, Policy and Practice*, eds P J Boon, G E Petts, and B R Davies, John Wiley & Sons, Chichester, 281 310.

FAO (1999) *Review of the State of World Fishery Resources: Inland Fisheries*, FAO Fisheries Circular No. 942, Food and Agriculture Organization of the United Nations, Rome.

Fu, C B, Kim, J W, and Zhao, Z C (1998) Preliminary Assessment of Impacts on Global Change in Asia, in *Asian Change in the Context of Global Climate Change*, eds J N Galloway and J M Melillo, Cambridge University Press, Cambridge, 308 341.

Khoo, K H, Leong, T S, Soon, F L, Tan, S P, and Wong, S Y (1987) Riverine Fisheries in Malaysia, *Archiv für Hydrobiologie Beiheft, Ergebnisse Limnologie*, 28, 261 268.

Liao, G Z, Lu, K X, and Xiao, X Z (1989) Fisheries Resources of the Pearl River and their Exploitation, *Can. Spec. Publ. Fish. Aquat. Sci.*, 106, 561 568.

Kottelat, M (1989) Zoogeography of the Fishes from Indochinese Inland Waters with an Annotated Checklist, *Bull. Zool. Mus., Univ. Amsterdam*, 12, 1 56.

McLean, R F, Sinha, S K, Mirza, M Q, and Lal, M (1998) Tropical Asia, in *The Regional Impacts of Climate Change. An Assessment of Vulnerability*, eds R T Watson, M C Zinyowera, R H Moss, and D J Dokken, Cambridge University Press, Cambridge, 381 407.

Meisner, J D and Shuter, B J (1992) Assessing Potential Effects of Global Climate Change on Tropical Freshwater Fishes, *GeoJournal*, 28, 21 27.

Riebsame, W E, Strzepek, K M, Westcoat, J L, Perritt, R, Gaile, G L, Jacobs, J, Leichenko, R, Magadza, C, Phien, H, Urbiztondo, B J, Restrepo, P, Rose, W R, Saleh, M, Ti, L H, Tucci, C, and Yates, D (1995) Complex River Basins, in *As Climate Changes: International Impacts and Implications*, eds K M Strzepek and J B Smith, Cambridge University Press, Cambridge, 57 91.

Roberts, T R (1995) Mekong Mainstream Hydropower Dams: Run-of-the-river or Ruin-of-the-river? *Nat. Hist. Bull. Siam Soc.*, 43, 9 19.

van der Leeden, F (1975) *Water Resources of the World*, Water Information Centre, New York.

Vorosmarty, C J, Li, C, Sun, J, and Dai, Z (1998) Drainage Basins, River Systems, and Anthropogenic Change: The Chinese Example, in *Asian Change in the Context of Global Climate Change*, eds J N Galloway and J M Melillo, Cambridge University Press, Cambridge, 210 244.

Welcomme, R L (1979) *Fisheries Ecology of Floodplain Rivers*, Longman, London.

Food Consumption Patterns and their Influence on Greenhouse Gas Emissions

Annika Carlsson-Kanyama

Environmental Strategies Research Group, Stockholm, Sweden

Emissions of greenhouse gases and energy use related to food consumption and production are explored and related to food products, consumption patterns and food trends. The relative emission contributions over the life cycle of food products vary greatly depending on the food product. For some products, processing dominates while for others, farming is the main contributor. Total emission levels over the life cycle vary greatly between food products with similar nutritional qualities. Substitutions leading to lower emission levels include more vegetables grown in the open and less in greenhouses, less meat and more legumes and more fresh and less tinned food. Trends in food consumption patterns, on the other hand, include more meat consumption, more fast and frozen foods, i.e., food almost ready for nal consumption, and more fresh food. In northern Europe, possibilities for production of fresh vegetables may increase as a result of climate change. In other parts of the world, the consequences of this change for food production and consumption may be far less bene cial.

HOW FOOD CONSUMPTION PATTERNS INFLUENCE CLIMATE CHANGE

Emissions of the major greenhouse gases carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O) and hydrofluorcarbons (HCFCs) are closely associated to food production and consumption. During the life cycle of food, which includes crop cultivation, animal husbandry, processing, transportation, storage, preparation and waste disposal, numerous human-induced activities cause emissions of these gases.

It is possible to calculate the greenhouse gas emissions from various foods over their life cycle so that emissions levels for various diets or food products can be compared. Such information can, and is, being used in discussions about how changing food patterns may, or may not, result in decreasing emission levels (Brower and Leon, 1999; VROM, 1999; Dahlin and Lindeskog, 1999).

When such estimates are being made, the following emissions are commonly considered:

- CO_2 fuel cycle emissions from fuels necessary to run the lorries, planes, ships, and tractors that are necessary for producing and transporting food. CO_2 fuel cycle emissions from food processing plants and fodder factories and from the fuels used for generating the necessary electricity for storage, heating, stables and electrical irrigation pumps are included as well. Net emissions of CO_2 that result from clearing forests for the purpose of cultivating food and emissions of CO_2 when organic soils are used for food production should also be considered.
- Emissions of refrigerants (including production emissions of refrigerants) from leaking transport refrigeration equipment in lorries, planes and ships and from refrigeration equipment in stationary storage facilities kept in warehouses, at the retailers and in people s homes.
- CH_4 emissions from hydropower dams used for producing the electricity needed for food production. CH_4 released from domestic livestock enteric fermentation, manure management and flooded rice fields are also part of the analysis.
- Process emissions of N_2O during the manufacturing of nitrogen (N)-fertilizers and N_2O-N yields from farmland as a result of application of nitrogen-fertilizers. N_2O emissions during storage of manure are considered as well.

Thus, food consumption directly and indirectly causes emissions of both energy-related greenhouse gases and greenhouse gas emissions from other sources during all steps in its life cycle. Some estimates of greenhouse gas emissions from various foods and diets have been made with the conclusions that i) relative emissions and energy

use during the life cycle vary depending on product, ii) different properties of foods result in large differences in greenhouse gases during their life cycles. Therefore diet is important for the total emissions from consumption patterns.

Relative Emissions and Energy Use During the Life Cycle of Products

Greenhouse gas emissions over the life cycle of foods can be distributed in very different ways depending on the choices of technologies for farming, transportation, storage etc., and because of the degree of packaging, processing and storage. This is illustrated in Figure 1 where the relative energy inputs over the life cycle of hamburger, bread, onions, pickled cucumber and lettuce are compared. (Energy use is often a good indicator of CO_2 emissions, the major greenhouse gas.) The result is that primary production, including crop farming and animal husbandry, dominates for hamburger, lettuce and cheese. This is because of the high use of fossil energy for heating greenhouses where the lettuce is produced and because of the high-energy use associated with feed production for cattle. For freeze-dried onions, bread and pickled cucumbers, the most energy-demanding stage in the life cycle is processing. This is because the relatively high energy consumption associated with baking, freeze-drying and pickling. These processes are applied to products with low energy demand during primary production and storage.

Mattson (1999) showed that the life cycle greenhouse gas emission profile of carrot purée was dominated by the processing stage due to the high-energy requirements for fabricating the purée from raw carrots. In another study of a cereal-based baby food product, the same author showed that the processing stage dominated the life cycle greenhouse gas emission profile together with the agricultural phase. The agricultural phase included dairy farming for obtaining the milk powder necessary for producing the baby food (Mattson and Stadig, 1999). Dairy cows emit CH_4 and Kramer (1999) suggested an emission factor of 27 grams of CH_4 kg^{-1} raw milk. Andersson (1998) analysed ketchup during its life cycle and showed that food processing and packaging made large contributions to global warming because of their high consumption of fossil fuel. The tomatoes were grown in the open in her study.

Emission Levels and Energy Use During the Life Cycle of Various Foods

There are some fundamental differences between the magnitude of greenhouse gases that can be emitted during the life cycles of foods. Foods that have similar nutritional contents can show very different life cycle emission profiles and this means that there are considerable opportunities for lowering greenhouse gas emissions related to food consumption by changing diet. This has been shown in a few studies that are discussed below.

Using vegetables grown in the open instead of vegetables grown in greenhouses would lower life cycle

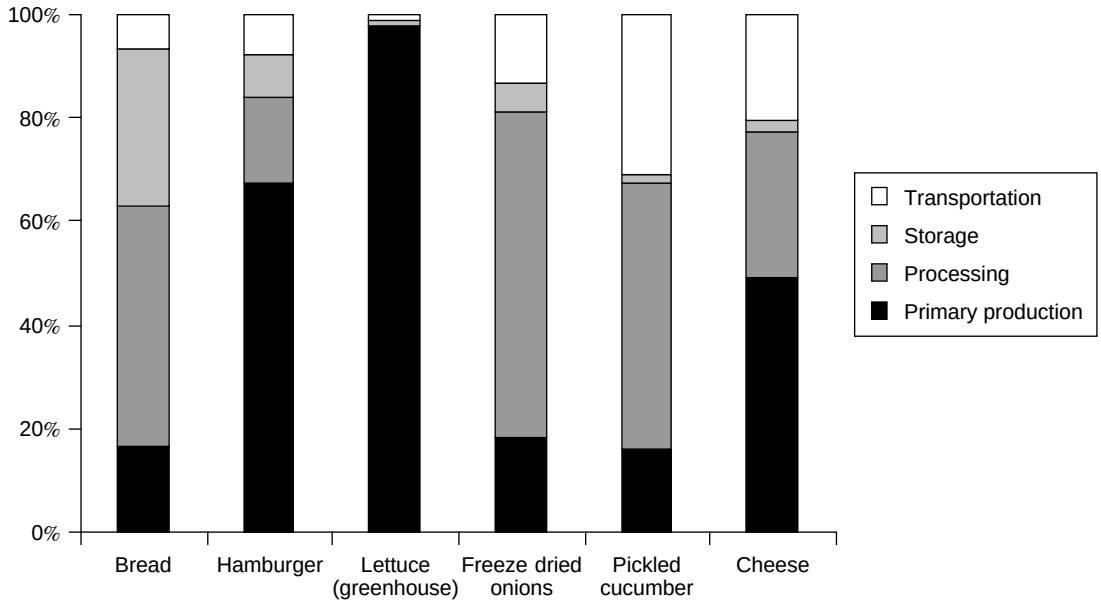


Figure 1 Distribution of energy use during the life cycle of hamburger bread, hamburger, lettuce grown in greenhouse, freeze dried onions, pickled cucumber and cheese (calculated from Carlsson-Kanyama and Faist, 2000)

emissions of greenhouse gases considerably (Carlsson-Kanyama, 1997). When compared by unit weight, open-grown carrots emit only 10% of the greenhouse gases emitted by tomatoes grown in greenhouses (*see* Greenhouse Food Production (Agriculture Intensive), Volume 3).

Coley *et al.* (1998) analysed the life cycle energy used for supplying the UK population with food. They calculated the energy used in the agricultural, transportation and retail sectors for various foods and coupled that information with food consumption data from over 3000 individuals. Coley called energy used in agricultural, transportation and retail sectors for the *embodied* energy of food. The large range in the embodied energy for diets was surprising, according to the authors. Some individuals consumed food with five times more embodied energy than other people did. The authors concluded that there were potentials for significant reductions in fossil fuel related greenhouse gas emissions by simple changes in diets but did not explore these possibilities.

Carlsson-Kanyama (1998) analysed the life cycle greenhouse gas emissions from four different meals consumed in Sweden that had the same nutritional content although they were composed of six different ingredients, carrots, pork, rice, dry peas, tomatoes and potatoes. Table 1 shows the total greenhouse gas emissions for these foods. Both life cycle emissions of greenhouse gases and life cycle energy use were lowest for the vegetable products: carrots, potatoes and dry peas. These foods were grown in open fields and mainly in Sweden. Life cycle emissions of greenhouse gases were highest for rice, pork and tomatoes. Most tomatoes were assumed to have been grown in greenhouses in Sweden or nearby countries, while rice was produced in countries distant from Sweden on irrigated fields. Figure 2 shows the emissions of greenhouse gases in grams of CO_2 equivalents for each of the different meals in Carlsson-Kanyama s study. The meals a d in Figure 2 can be characterised in the following manner. Meal a is purely vegetarian with ingredients mainly from the domestically produced foods carrots, potatoes and dry peas. Meal b is also vegetarian, but with the exotic ingredients rice and tomatoes in addition to dry peas. Tomatoes are considered exotic because either they have to be imported from southern Europe or else grown in greenhouses. Meal c is solely based on animal and exotic foods: rice, tomatoes and pork while meal d contains only domestically produced

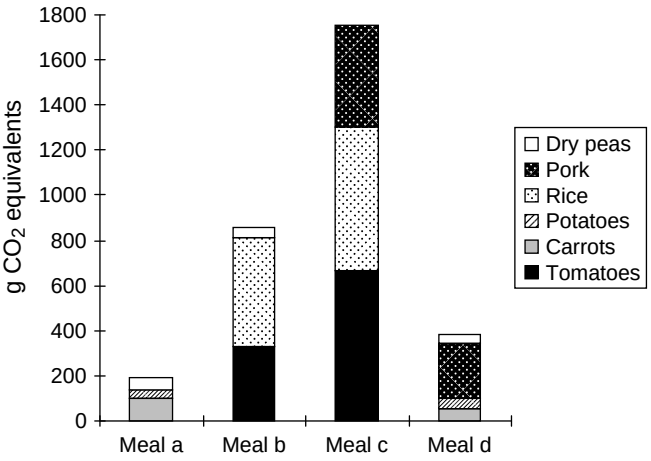


Figure 2 Life cycle emissions of greenhouse gases in CO_2 equivalents from four different meals with the same energy and protein contents (Carlsson-Kanyama, 1998)

food of both animal and vegetable origin. Figure 2 shows that emissions of greenhouse gases during the life cycle of the four meals differ substantially. The animal exotic meal (meal c) emits 1800 g of CO_2 equivalents while the vegetarian domestic meal (meal a) only emit 190 g of CO_2 equivalents. This makes the animal exotic meal nine times more polluting than the vegetarian domestic meal.

Kramer *et al.* (1999) calculated the life cycle emissions of greenhouse gases related to Dutch household food consumption and found that food products from cattle farming determine the CH_4 emissions, while food products from agriculture determine the N_2O emissions. They showed that differences in greenhouse gas intensities within food product categories and between food product categories are significant and that these differences offer possibilities for reduction options. For example, the CO_2 intensity, measured as kg Dfl^{-1}, was 0.59 for minced meat but only 0.17 for carrots and tubers. Examples of reduction options given by the authors are shifting from glasshouse vegetables to vegetables grown on the open ground, substituting meat for eggs, pulses and nuts and substituting tinned, bottled, or frozen vegetables for fresh vegetables. The authors also showed that for food products from cattle farming determines the total life cycle emissions of CH_4 resulting from food consumption patterns while products from agriculture determine the life cycle emissions of N_2O.

Table 1 Life cycle emissions of greenhouse gases in CO_2 equivalents for carrots, tomatoes, potatoes, pork, rice and dry peas consumed in Sweden (Carlsson-Kanyama, 1998)

	Carrots	Tomatoes	Potatoes	Pork	Rice	Dry peas
g CO_2 equivalents kg^{-1}	500	3300	170	6100	6400	680

TRENDS IN FOOD PATTERNS AND HOW THEY CAN INFLUENCE OR BE INFLUENCED BY CLIMATE CHANGE

Historically patterns of food consumption have changed from a diet based mainly on locally produced foods to a diet where exotic fruits, vegetables, spices, meat and processed food are becoming more commonplace (Figure 3).

The global change impacts resulting from the increased meat consumption have been much discussed (e.g., Wirzenius, 2000; Goodland, 1997). They include increased emissions of methane from the digestive tracts of animals, clearing of virgin forests for obtaining grazing land and cultivation of lands not suitable for crop farming with resulting losses of topsoil. Wirzenuis (2000) estimated that animal food systems accounted for roughly two-thirds of the total appropriation of global biomass while their contribution to human diet was merely one-tenth of the total. Such results fuel discussions about possibilities for shifting diets in order to safeguard both the environment and mankind. However, there are other food consumption trends in developed countries that have the potential for changing the amount of greenhouse gases currently released during the life cycle of diets. Those who can afford it in the developing world often emulate consumer culture in the developed countries (Gallagher, 1997), so trends in developed countries are of utmost importance for the whole of the global community. The ongoing nutrition transition among urban populations in the third world includes, among other changes, an increased consumption of meat (Popkin, 1999).

Food patterns depend on a number of factors. The state of the ecosystems is certainly very important for the food supply but wealthy urban populations, a category to which most people in the developed world belong, are not the first to see their food supply affected by environmental change. In the short run, other factors have a far greater effect on their food patterns and that includes

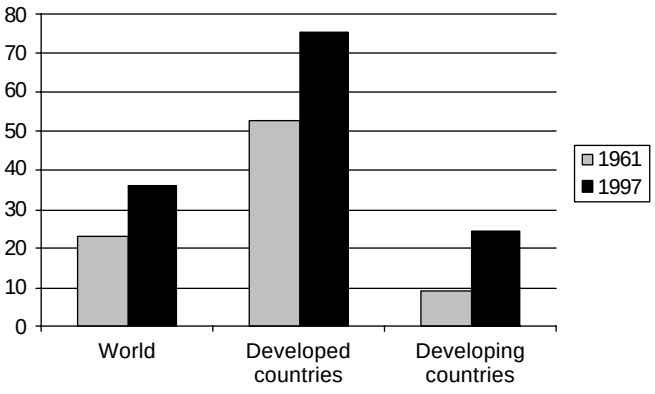


Figure 3 Meat supply in 1961 and 1997 in kg meat person^{-1} year^{-1}. For the whole world and for developing and developed countries (FAO, 2000)

the available food technology, expectations from food suppliers expressed through advertising and marketing and the time and knowledge people have of food preparation. A consistent trend in many developed countries is that knowledge about food preparation and food handling is on the decline, resulting in more cases of food poisoning (Carlsson-Kanyama, 1999).

Food suppliers in Sweden are strongly influenced by trends in the rest of Western Europe and the US (Carlsson-Kanyama, 1999). They believe in some major developments that could impact levels of greenhouse gases emitted during the life cycle of diets in the future.

The major trend is that the supply of fast foods is going to increase in the future food market. Increasingly hurried lifestyles of consumers is considered to be an important factor in this trend, resulting in an increased supply of food that is quick and easy to prepare. Along with an increased supply of more uncomplicated foods, there will be more foods perceived as healthy, according to the same food suppliers. This means more chilled but fewer canned and frozen products. The food suppliers, who do not expect time-starved future consumers to wait for frozen food to thaw, regard the fact that chilled products take less time to prepare as an additional advantage. Partly as a result of the trend towards healthier food, many interviewees also believed in an increased and diversified supply of fruit and vegetables, while many expected meat consumption to decline (Carlsson-Kanyama, 1999).

There are many examples of substitutes for food that require skill and time to prepare, and that emit more fossil fuel related CO_2 during their life cycle than their precursors. Examples are rice versus potatoes or tomatoes versus carrots (Table 1). Rice requires less skill to prepare than potatoes because cooking time can be predicted with higher accuracy and tomatoes require less skill than carrots because carrots have to be ground or boiled. But food is sometimes prepared with equal or less fossil fuel inputs in industries than in homes. A study of energy use during the life cycle of pea soup showed that whether the soup was cooked at home or industrially was insignificant. What really mattered was how much meat the pea soup contained, how the soup or ingredients had been transported to the consumer s home and how much was discarded in the household. A growing numbers of unskilled consumers could also make inefficient use of food resources leading to more organic waste. Food preparation by industry could then be an advantage provided that waste was re-used or re-cycled efficiently. But to what extent will food that is faster and easier to prepare entail longer transportation distances, more elaborate processing and longer storage times? The impact of such activities may be significant in evaluating the impacts of products during their life cycle. There are many

ways of imagining a future supply of faster, easier food: from the local food industry using organically produced food to large-scale industries using food of global origin. Surely, the environmental consequences of these alternatives will differ, as will the consequences for consumer protection.

Ecosystem alteration caused by climate change will, in the long run, significantly alter regional possibilities for food production. For example, crop zones in Western and Northern Europe will expand enabling, for example, successful production of wheat in the northern coastal zone of Sweden where such production is today severely constrained due to the cold climate and short cropping season (Raisanen *et al*., 2000). The possibility of expanding the production of more exotic crops in Northern Europe also seems likely. Examples are maize and grapes, while other examples may be a northern extension of vegetables and fruits cultivation, such as cucumbers, cauliflower and cherries.

These new possibilities certainly seem to fit well with anticipation for an increased consumer demand for fresh and healthy food. In this scenario, local production of fruits and vegetables in Western Europe could increase with beneficial consequences for fossil fuel related emissions from vehicles. Unfortunately, climate change is predicted to have far more serious consequences in parts of the world where many consumers do not have the purchasing power to demand other than locally produced food. With a collapse of local fisheries depending on coral reefs (World Resources Institute, 2000), for example, food patterns common among coastal communities in developing countries could change to diets with less high quality protein. Whether or not this will be of concern for affluent people in Western Europe, who could benefit from the anticipated changes, is an open question. Food patterns now and in the future can both influence and be influenced by climate change. For some people, climate change will offer new possibilities for obtaining a diversified supply of locally produced foods. For others, consequences of this change can threaten their very existence by changing food patterns to less nutritious ones.

REFERENCES

Andersson, K (1998) *Life Cycle Assessment (LCA) of Food Products and Production Systems*, Ph.D. Thesis, School of Environmental Sciences, Department of Food Science, Chalmers University of Technology, Gothenburg, Sweden.

Brower, M and Leon, W (1999) *The Consumer's Guide to Effective Environmental Choices: Practical Advice from the Union of Concerned Scientists*, Three Rivers Press, NY.

Carlsson-Kanyama, A (1997) Food Consumption Patterns and their Influence on Climate Change: Greenhouse Gas Emissions in the Life Cycle of Tomatoes and Carrots Consumed in Sweden, *Ambio*, 27(7), 528 534.

Carlsson-Kanyama, A (1998) Climate Change and Dietary Choices: How can Emissions of Greenhouse Gases be Reduced? *Food Policy*, 23(3/4), 277 293.

Carlsson-Kanyama, A (1999) Changing Food Consumption Patterns: Implications for the Environment with Regards to Past and Future Trends in Sweden, *Consumption Patterns and Climate Change: Consequences of Eating and Travelling in Sweden*, Ph.D. Thesis, Department of Systems Ecology, Stockholm University.

Carlsson-Kanyama, A and Faist, M (2000) *Energy use in the Food Sector: a Data Survey*, Report no. 291, Swedish Waste Research Council, Swedish Environmental Protection Agency, Stockholm.

Coley, A D, Goodcliffe, E, and Macdiarmid, J (1998) The Embodied Energy of Food: the Role of Diet, *Energy Policy*, 26(6), 455 459.

Dahlin, I and Lindeskog, P (1999) *Ett första steg mot mer hallbara matvanor (A First Step Towards More Sustainable Consumption Patterns)*, Stockholm Center of Public Health, (In Swedish).

FAO, Food and Agricultural Organisation (2000) FAO statistical Databases on Food Supply, Accessed at http://apps.fao.org/page/collections?subset=nutrition June 2000.

Gallagher, K (1997) Globalisation and Consumer Culture, Overview Essay, in *The Consumer Society*, eds N R Goodwin, F Ackerman, and D Kiron, Island Press, Washington.

Goodland, R (1997) Environmental Sustainability in Agriculture: Diet Matters, *Ecol. Econ.*, 23, 189 200.

Kramer, K J, Moll, H C, Nonhebel, S, and Wilting, H C (1999) Greenhouse Gas Emissions Related to Dutch Food Consumption, *Energy Policy*, 27, 203 216.

Matsson, B (1999) Environmental Life Cycle Assessment (LCA) of Organic and Integrated Production of Carrot Purée, *Environmental Life Cycle Assessment (LCA) of Agricultural Food Production*, Ph.D. Thesis, Swedish University of Agricultural Sciences, Sweden.

Matsson, B and Stadig, M (1999) Screening Life Cycle Assessment (LCA) of Organic and Conventional Production of Cereal-Based Baby Food, *Environmental Life Cycle Assessment (LCA) of Agricultural Food Production*, Ph.D. Thesis, Swedish University of Agricultural Sciences, Sweden.

Popkin, B M (1999) Urbanization, Lifestyle Changes and the Nutrition Transition, *World Dev.*, 27(11), 1905 1916.

Raisanen, M, Rummukainen, A, Ullerstig, B, Bringfelt, Hansson, H, and Willén, U (1999) *The First Rossby Centre Regional Climate Scenario - Dynamical Downscaling of CO_2 Induced Climate Change*, HadCM2 GCM. Report no. 85, SMHI, Swedish Meteorological and Hydrological Institute, Norrkoping.

VROM (Ministry of Spatial Planning and the Environment) (1999) *Minder energiegebruik door een andere leefstijl? December 1995- juni 1998 eindrapportage*, VROM, Nieuwegein, The Netherlands.

Wirzenuis, S (2000) *Human use of Land and Organic Materials. Modelling the Turnover of Biomass in the Global Food System*, Ph.D. Thesis, Department of Physical Resource Theory, Chalmers University of technology, Gothenburg, Sweden.

World Resources Institute (2000) http://www.igc.org/wri/indictrs/rrbleach.htm, accessed in September.

Forest Logging Systems in Tropical Countries: Differential Impacts

Abdul Rahim Nik
Forest Research Institute Malaysia, Kuala Lumpur, Malaysia

Sustainable production of tropical forests has attracted much debate and become an international issue. This is because tropical rain forests are not only unique in many respects, but also the demands placed on them are intense and increasing in complexity. The growing population in many countries of the humid tropics continues to exert tremendous pressure on the forests for various uses, increasing fuel-wood requirements, expansion of agricultural land, clearing for settlement and commercial logging. By the same token, forests are also susceptible to natural disturbances such as res, landslides, insect attacks and diseases. In addition to tangible bene ts, forests of various types provide other environmental services such as watershed protection, soil conservation, a supply of fresh water and carbon sequestration.

While forest cover in many developing countries has been steadily decreasing, there has been a continued increase in demand for wood products. The global consumption of wood increased by 36% between 1970 and 1994 (FAO, 1997). Similarly, demand for fuelwood, which is the main source of energy for two- fths of the world's population, continues to grow by 1.2% per year.

Realizing the important role of tropical forest, heightened environmental awareness and public pressure have continued to exert in uences upon all aspects of the forestry sector: forest management, harvesting and post-felling operation and including trade. Recognizing that forests should be managed not only to ensure their productive functions, but also environmental services and social bene ts are sustained, has led to efforts to develop criteria and indicators for sustainable management (FAO, 1997). There is a concerted trend towards managing forests as ecological systems with multiple economic bene ts and environmental values.

SELECTIVE LOGGING SYSTEM

Currently, selective logging is practiced for the extraction of valuable timbers from tropical forests. This is so because of the high species diversity of tropical forests; only a small proportion of the trees are suitable for extraction during each logging phase, or rotation. Further, this system has the flexibility to manage highly variable forest conditions and changes in the socio-economic environment (Thang, 1997). While its implementation may vary from country to country, this system entails selection of felling regimes based on inventory data. Depending on the intensity of logging and the rotation time between logging operations, two selective logging systems essentially governed by the silvicultural systems, called the *monocyclic* and *polycyclic* systems, are commonly used. The former involves removal of up to 100% of commercial trees from the forest. The time between successive logging phases or logging cycles depends on the maturity of the main species felled. As the volume of timber removed can be as high as $120\,m^3\,ha^{-1}$, the potential damage to soil and residual trees is also substantial (Bruijnzeel and Critchley, 1994; Bruenig, 1996). Under the polycyclic system, highly selective removal of large trees of desired species is prescribed. Thus, fewer trees and a lower volume of timber are removed, which amounts on average to 8 10 trees ha^{-1} or 20 $30\,m^3\,ha^{-1}$, respectively. Examples of such systems are the Selective Management System in Malaysia, Tebang Pilih Tanam Indonesia in Indonesia and CELOS in Surinam.

LEVEL OF FOREST DISTURBANCE

Since forest logging takes place in several forms and to ease the assessment process, Bruijnzeel (1990, 1992) classified impacts on forests into three levels of disturbance, namely low, intermediate and high. Low intensity disturbance is usually associated with a natural cause such as mortality or tree fall. The other levels of disturbances are associated with human-induced activities: selective logging shifting cultivation, forest plantation and permanent agricultural crops. Forestry activities should be described precisely before evaluating the environmental impacts of each activity. In general, forest logging may cause detectable changes in the environmental variables, depending on the intensity of disturbances and the extent of cover removed. By the same token, forest clearance and forest conversion to other land uses are expected to cause greater impacts on hydrology and soil erosion processes. With increasing emphasis on sustainable forest management, an improved harvesting technique known as *reduced impact logging* (RIL) is being implemented and promoted in various regions. The aim of this technique is to reduce damage on residual trees, soil disturbance, and impacts on wildlife (Sist *et al.*, 1998).

DAMAGE TO RESIDUAL TREES

The extent of damage to residual trees following forest logging or harvesting is critical, for it will determine the time of the next logging cycle. Previous studies showed that damage in conventional forest logging operations has been quite excessive, especially in areas with high logging intensity in which the percentage of damage may reach 30 70% (Nicholson, 1958; Fox, 1968; Dykstra and Heinrich, 1996;

Canonizado, 1978). In areas with lower logging intensity, damage to residual trees has ranged from 10 20%. Logging damage could be reduced by instituting pre-felling of climber cutting. However, with the adoption of RIL techniques, a remarkable reduction in residual tree damage can be attained from 50 200% (Johns *et al.*, 1996; Moura-Costa, 1997). In this instance, achieving a threshold of logging intensity should be promoted as suggested by Watanabe (1992), viz, stands at 30% of stand basal area.

EFFECTS OF FOREST CONVERSION AND FOREST LOGGING ON WATER YIELD

Most of the quantitative information regarding hydrological changes resulting from forest logging or forest conversion is based on catchment experiments usually conducted on relatively small watersheds of less than 250 ha in size (Hewlett, 1982). While there are many variations to conducting experimental catchment studies, a paired-catchment method is most commonly adopted. Nevertheless, one of the problems associated with catchment studies is a leakage problem especially in the case of small headwater catchments (Bruijnzeel, 1990).

The question of water changes after forest removal or land use change has been extensively assessed and reviewed in many parts of the world, including tropical countries (Bosch and Hewlett, 1982; Oyebande, 1988; Bruijnzeel, 1990, 1992, 1997; Bonell and Balek, 1993). The main objective of those investigations was to determine the effects of vegetation and natural plant cover removal, or modification on hydrological variables including water yields and flow regime. This issue has assumed greater importance as forestry and land use changes take place in upstream catchments that served as an important source of water supply for downstream users.

Results obtained from studies in temperate zones as well as in the tropics indicate that almost every well-designed experiment has shown increases in water yield as a response to forest cutting and, in general, the increase is proportional to the amount of canopy removed (Bruijnzeel, 1990; Oyebande, 1988, Table 1). Removing up to 40% of the standing trees causes increases in water yield ranging from 113 188 mm $year^{-1}$ or 59 85% $year^{-1}$ (Abdul Rahim and Harding, 1992). In other words, removal of forest cover generally leads to increased streamflow, while afforestation or reforestation with fast growing species on cleared lands leads to a decline in water yield (Oyebande, 1988). This increase may be permanent when converting tall forest to grassland or shallow rooted agricultural crops, or may be temporary in the case of conversion to tree plantations. With much effort given to reforestation programs in many developing countries in view of diminishing forest areas, more attention should be focused on this issue.

However, the above responses are generally applicable to forests at elevations lower than 1000 m above sea level which, in the humid tropics of southeast Asia comprise lowland and hill dipterocarp forests. Floristically, beyond that altitude, different forest types can be found, namely upper dipterocarp, oak laurel and montane ericaceous forests (Burgess, 1969), which are also known as tropical montane cloud forest (Studtmuller, 1987).

During the regeneration, or the recovery period, the time taken to recover to the pre-logging condition depends on the level of disturbance and the year-to-year variability of rainfall. Nevertheless, detailed information on this process is still limited and sometimes contradictory. The Berembun catchment study in Malaysia, showed that the water yield increment began to decline seven years after selective forest harvesting had extracted about 40% of the commercial stock (Abdul Rahim and Zulkifli, 1994). Similarly, Malmer (1993) found no trend in annual evapotranspiration for five years immediately after forest clearing in Sabah, Malaysia. On the other hand, Kuraji and Paul (1994) claimed that 3 5 years after logging activities in Sabah, the pre-logging water yield was attained. Under temperate conditions, it may take much longer to return to the pre-clearing level of streamflow.

EFFECTS OF FOREST LOGGING ON FLOW REGIMES AND FLOODING

In many places, dry-season flow is an important issue for water supply managers, particularly when the demands are high due to a protracted period without rain. During this period, baseflow and reservoir storage become the only sources of water to meet consumptive requirements of irrigation, domestic use and industry. In addition, a low-flow period is also critical from the water quality point of view, as pollutant concentrations generally become higher. Hence, the impact of forest activity on dry-flow conditions becomes a major concern, not only to forest managers, but more so to water supply engineers.

Evidence with respect to the effects of forest clearing on dry season flow in the tropics seems, at first, contradictory and ambiguous. On the one hand, there are reports of greatly diminished flows whilst on the other hand, significant increases have been observed. This is so because forest conversion has two opposite effects on water. First, it tends to increase runoff and decrease infiltration of water into the ground, mainly due to ground compaction and disturbance, which would lower the water table (Bruijnzeel, 1990). On the other hand, replacement of forest with shallow rooting vegetation with a lower transpiration rate tends to reduce groundwater loss and hence raise the water table. Nonetheless, in interpreting the above contrasting results, one has to take into account the prevailing climatic, pedological and hydrological setting of the area as well as

Table 1 Changes in water yield and stormflow response following logging (adapted from Bruijnzeel, 1990, 1992)

Location	Type of logging	Changes in water yield (mm/year) 1	2	3	4	5	6	7	Changes in stormflow volume	Changes in peakflow discharge	References
Sg. Tekam (A), Malaysia	Dipterocarp forest to cocoa plant	+110 (117)	+706[a] (157)	+353 (94)	+263 (158)%					280% increase	Abdul Rahim (1988)
Sg. Tekam, Malaysia	60% conversion of forest to oil palm	145 (85)	155 (142)	137 (97)%					19 to 37% increase in 1st three years	38% increase after logging/burning	DID (1989); Abdul Rahim (1988)
Mendolong (1) Sabah	Clearfelling rain forest to A. mangium[b]	+397	+522	+89							Malmer (1993)
Mendolong (4) Sabah	Clearfelling rain forest to A. mangium[c]	+197	+170	+80							Malmer (1993)
Mendolong (5) Sabah	Clearfelling rain forest to A. mangium[d]	+460	+262	+468							Malmer (1993)
Berembun (C1) Pen. Malaysia	Selective logging removing 40% of the commercial stocking	165 (70)	142 (55)	175 (72)	188 (85)	161 (73)	168 (67)	113 (59)%	Sign ($P < 0.10$)	ns[e]	Abdul Rahim (1992)
Berembun (C3) Pen. Malaysia	Selective logging removing 33% of the commercial stocking	87 (37)	70 (28)	106 (44)	102 (46)	53 (24)%			ns	ns	Abdul Rahim (1992)
Lien-Hua-Chi, Taiwan	Clearcutting of mixed evergreen hill forest; Regeneration	+448 (58%)	+204[f] (51%)						ns	48% increase	Hsia and Koh (1983)
Babinda, Queensland	Light selective logging	+264 (7.0%)	+323[g] (13.4%)						ns	ns	Gilmour (1977)

[a] 100% clearance.
[b] Non-mechanical, burning, no wood extraction.
[c] Manual log extraction, no burning.
[d] Tractor log extraction, burning.
[e] ns, not significant at <0.05.
[f] Dry year.
[g] Wet year.

the way in which conversion and subsequent land use were carried out (Bruijnzeel, 1989).

Hydrological studies on selective logging in Peninsular Malaysia have clearly shown that increases in water yield following forest removal are largely associated with baseflow augmentation. This is so mainly because of the minimal disturbance to the ground and stream channels, except along forest roads and skid tracks. This, in turn, leaves surface infiltration relatively unaffected. Eventually, a greater recharge to groundwater may occur, resulting from reduced evapotranspiration and, ultimately, a larger baseflow.

There has been a protracted debate and popular belief regarding the role of forests, as to whether or not forest cover upstream areas at catchments will prevent floods downstream. Quite often, forest logging has been blamed for disastrous flooding downstream. For instance, monsoonal floods of the Ganges and Indus have been attributed to uplands tree cutting (World Water, 1981). Controversies arise mainly due to a misunderstanding on the biophysical relationships involved. Among pertinent questions are, for example, whether the perceived increase in flooding that follows forest clearing is due to the removal of tree cover itself or is due to abusive land use. Further, are both the upstream and downstream areas affected?

Scientific evidence supports a link between deforestation and flooding at the local level within a drainage basin of less than 50 000 ha (Bruijnzeel and Bremmel, 1989). In small watersheds, increases in water yield resulting from forest conversion translate directly into increases in stormflow. Individual sub-watersheds will tend to flood in sequence, as storms pass segmentally rather than simultaneously. It has been shown that after forest clearance, peak discharge and stormflow volume significantly increase due to the effect of an increase in soil wetness. However, in the case of selective logging, the effects on peak discharges related to soil moisture would be less significant.

EFFECTS OF FOREST LOGGING ON SOIL EROSION AND SEDIMENTATION

In discussing the impacts on soil erosion and sedimentation, it is instructive to make a distinction between on-site erosion and off-site or downstream effects. The former involves a smaller spatial scale such as a field or hillslope, which produces a much faster runoff while the latter affects a larger area farther downstream. This downstream effect may take a long time to be reflected in a large watershed (Pearce, 1986). The eventual effect, when eroded soil particles are washed downstream, is to reduce the life of reservoir and irrigation systems, diminishing their efficiency, and thus increasing the costs of maintenance.

Interesting results have been synthesized by Wiersum (1984) on surface erosion under various tropical forests and tree crops. Ground cover, rather than canopy extent, is the main cause of surface erosion. Removal of understory vegetation may lead to slightly greater erosion rates, but soil erosion increases markedly upon removal or destruction of the forest litter layer. When trees are harvested, their removal requires trails, skid tracks, and roads, which if not planned and constructed properly, will cause serious erosion and sedimentation. In many instances, the major cause of soil erosion is usually associated with road construction rather than the land use change itself. In Malaysia, the amounts of soil erosion arising from forest roads and skid tracks in the first year after logging are 13 tons ha^{-1} and 10 tons ha^{-1}, respectively. However, the rate of erosion is drastically reduced to 3 and 2 tons ha^{-1} $year^{-1}$ in the second year (Baharuddin, 1995), mainly due to the rapid recovery of undergrowth.

Selective forest logging in the Berembun catchments, Malaysia, led to minimal increases in sediment yield concentrations, and thus sediment yield under both the usual commercial and the supervised commercial types of logging (Baharuddin, 1988). This was because of the low logging intensity as prescribed in the selective logging system coupled with the institution of conservation measures. On the other hand, forest logging in Ulu Segama, Sabah and Hulu Langat, Selangor produced much higher sediment yields (Douglas *et al.*, 1993 and Lai, 1993, Table 2).

To fully assess the impact of land use change from forest to other types, it is necessary to evaluate the link between induced erosion and sedimentation. This depends on two factors the *sediment delivery ratio* and the background level of sedimentation (Chomitz and Kumari, 1996). Essentially, only some of the eroded soil ends up in stream channels and rivers; the reminder is trapped or redeposited farther downslope. Sediment delivery ratio varies inversely with the catchment size and a ratio 0.3 is assumed for basins on the scale of hundreds of square kilometers. Background sedimentation varies remarkably, depending on local geology and the state of on-site land use. Factors to be considered include land management practices, human settlement, extent of road area and topographical features.

EFFECTS ON WATER QUALITY AND THE NUTRIENT BUDGETS

Water Quality

Forest logging affects stream water quality through physical, geochemical and biological processes. Removal of forest cover, especially the undergrowth and litter layer, will not only induce erosion but also enhance leaching of chemical elements. Additional sources of organic matter in the form of logging debris, coupled with greater exposure to sunlight, will enhance microbiological activity, leading to higher rates of decomposition, mineralization and nitrification.

Table 2 Impacts of forest conversion and forest logging activities on sediment yield[a]

	Location	Type of land use change	Area (ha)	Annual rainfall (mm)	Sediment yield (mt ha^{-1} $year^{-1}$) (before)	Sediment yield (mt ha^{-1} $year^{-1}$) (after)
1	Sungai Tekam, Pahang					
	Catchment A	Secondary Forest to cocoa	47	1916	0.17	0.50–1.05
	Catchment B	Secondary Forest to oil palm	96		0.42	1.05–4.14
2	Berembun, N. Sembilan					
	C1	Selective logging, 40% extraction	12.9	2032	0.12	0.27
	C3	Selective logging, 33% extraction (supervised)	29.7	2070	0.12	0.19
3	Ulu Segama, Sabah					
	Sg. Stesen Baru	Selective logging	490	2642	0.6	6.6
4	Hulu Langat, Selangor					
	Sg. Batangsi	Selective logging (on-going)	1980		0.54	28.26
	Sg. Chongkak	Recently logged	1270	2305	0.54	24.7
	Sg. Lui	Logged-over	6880		290	0.90
5	Mendolong, Segama, Sabah					
		Forest to acacia mangium (manual felling/tractor)	18		0.24	2.53

[a] *Sources:* DID (1989); Baharuddin (1988); Douglas *et al.*, (1993); Lai (1993) and Malmer (1993).

The impact of selective forest logging, which removed about 40% of the standing trees in the Berembun, Malaysia, resulted in increases of pH, electrical conductivity (EC) and water hardness (Zulkifli *et al.*, 1993). The largest increases were observed for suspended solids and turbidity by 12- and 9-fold, respectively, in the first year after logging. Similar increases in chemical concentrations were also observed particularly for alkalinity, silicate, iron (Fe), potassium (K) and sodium (Na). However, recovery to pre-logging conditions was attained in between three and five years but increases Fe and K tended to be more prolonged. The impacts were most pronounced during and immediately after storms.

Converting about 60% of a logged-over forest in the Tekam catchment in Pahang, Malaysia resulted in two-fold increases of EC, which was sustained for three years (DID, 1989). The higher EC reflects an overall increase in ionic concentrations. The other noticeable increases were for Fe and K during the land preparation and planting stages. Bare ground resulting from road construction was the major source of Fe. On the other hand, K might be released from soil minerals and decomposing biomass. Following establishment of a forest plantation at Mendolong, Sabah, Malaysia, considerable increases in EC and K were also observed (Malmer, 1993).

Although water quality degradation due to forest operations is inevitable, it is possible to reduce the impact to a manageable level under well-planned logging and by strict adherence to conservation measures. Proper road alignment and buffer strips are the main considerations for reducing logging impacts. In the case of forest conversion/clear-felling, avoiding biomass burning significantly reduces negative impacts on water resources.

Impact on Nutrient Loss

Tropical rainforests are particularly sensitive to site disturbance while their delicate structures, important for conserving nutrients, may be partially or completely destroyed during forestry operations, causing accelerated losses of nutrient. The magnitude of nutrient losses varies from undetectably small amounts in the case of a low intensity disturbance, to serious and permanent reduction in site nutrient capital as a result of a high intensity of disturbance

such as converting rainforest to pasture (Bruijnzeel, 1990). Nutrient removal occurs via biomass harvest, surface erosion and enhanced leaching to groundwater.

Losses of major nutrients following selective logging at Berembun catchment, increased by between 70% to more than 200%. Conversely, the supervised selective logging method showed moderate increases in solute losses, ranging from 13% for Na to 66% for nitrate (NO_3). Besides trapping sediment, buffer strips are also effective in filtering and reabsorbing nutrients before entering a water course (Brouwer, 1995).

An interesting study examining the effects of different methods of establishing the forest plantation on forests previously struck by fire, and forests that had been selectively logged, was conducted in Sabah, Malaysia (Malmer, 1993; Nykvist *et al.*, 1994). After two years of calibration, four of the six catchments were treated involving: (i) clearing of secondary vegetation, burning and planting (W1 and W2); (ii) clear-felling, crawler tractor extraction, burning and planting (W5), and (iii) clear-felling, manual extraction, no burning and planting (W4). The losses of major solutes were consistently higher from the normal practice treatment (W5) compared to manual site preparation with residuals unburned (W4). It is also interesting to note that, within the short time span (2.8 years) after treatment, the hydrologic losses of K and Fe had exceeded losses via stem wood even for the manually extracted and unburned catchment, W4.

Loss of nutrients on an annual basis may be small, but over a long term, the total loss can be substantial. Depending on the chemistry of the rainfall, the time span needed to compensate for nutrient losses following forest harvesting may range from 25 90 years for calcium (Ca), 20 35 years for magnesium (Mg) and 10 50 years for K (Zulkifli *et al.*, 1998). The slower recovery rate for Ca could be attributed to a much larger quantity of Ca for the formation of woody biomass compared to Mg and K. The implication of this study is that nutrient sustainability could be attained if forest logging is carried out over a 50-year rotation.

SUMMARY

The effects of forest logging on hydrological characteristics and water yield not only depend on several bio-physical factors; soil, geology and rainfall regime, but more importantly on the type of logging method used and the extent of forest cover removed. Although the effects of forest removal on hydrological regime and land cover are context-specific, several general inferences can be elicited from the examples reviewed, particularly on the cumulative effects and long-term changes. Cumulative effects of selective logging will be only manageable provided the logging cycle or rotation period is adequately long, and an effective enforcement of management procedures and conservation measures are fully enforced during logging operations. These may include an adherence to the RIL *reduced impact logging* approach, retention of buffer zone, proper planning of logging road and delineation of sensitive and protective zone for biodiversity conservation. Unless these are prescribed, there could be long-term changes in the water balances, geochemical cycles, sedimentation, local microclimate and carbon sequestration functions of forest cover.

REFERENCES

Abdul Rahim, N (1988) Water Yield Changes after Forest Conversion to Agricultural Landuse in Peninsular Malaysia, *J. Trop. For. Sci.*, 1(1), 67 84.

Abdul Rahim, N and Harding, D (1992) Effects of Selective Logging Methods on Water Yield and Streamflow Parameters in Peninsular Malaysia, *J. Trop. For. Sci.*, 5(2), 130 154.

Abdul Rahim, N and Zulkifli, Y (1994) Hydrological Response to Selective Logging in Peninsular Malaysia and its Implications on Watershed Management, in *Proceedings of the International Symposium on Forest Hydrology*, ed T Ohta, University of Tokyo, Tokyo, Japan, 263 274.

Baharuddin, K (1988) Effects of Logging on Sediment Yield in a Hill Dipterocarp Forest in Peninsular Malaysia, *J. Trop. For. Sci.*, 1(1), 56 66.

Baharuddin, K (1995) Effects of Logging Operations on Soil Physical Properties and Soil Erosion in a Hill Dipterocarp Forest of Peninsular Malaysia, Research Pamphlet, No. 119, FRIM, Kepong, Malaysia, 1 22.

Bonell, M and Balek, J (1993) Recent Scientific Developments and Research Needs in Hydrological Processes of Humid Tropics, in *Hydrology and Water Management in the Humid Tropics - Hydrological Research Issues and Strategies for Water Management*, eds M Bonell, M M Hufschmidt, and J S Gladwell, UNESCO Cambridge University Press, Cambridge, 167 206.

Bosch, J M and Hewlett, J D (1982) A Review of Catchment Experiments to Determine the Effect of Vegetation Changes on Water Yield and Evapotranspiration, *J. Hydrol.*, 55, 3 23.

Brouwer, L C (1995) *Nutrient Cycling in Tropical Rain Forest in Guyana*, Tropenbos Program Report, No. 95/2, Guyana.

Bruenig, E F (1996) *Conservation and Management of Tropical Rainforests: an Integrated Approach to Sustainability*, CAB International, Cambridge University Press, Cambridge, 1 339.

Bruijnzeel, L A (1989) (De)forestation and Dry Season Flow in the Tropics: a Closer Look, *J. Trop. For. Sci.*, 1, 145 161.

Bruijnzeel, L A (1990) *Hydrology of Moist Tropical Forests and Effects of Conversion: a State of Knowledge Review*, IHP/UNESCO Publication, Paris and Free University, Amsterdam.

Bruijnzeel, L A (1992) Managing Tropical Forest Watershed for Production: Where Contradictory Theory and Practice Coexist, in *Wise Management of Tropical Forest*, eds F R Miller and K L Adam, Oxford Forest Institute, Oxford, 37 75.

Bruijnzeel, L A (1997) Hydrology of Forest Plantations in the Tropics, in *Management of Soil, Nutrients and water in Tropical Plantation Forests*, eds E K Sadanandan Nambiar and A G Brown, ACIAR Monograph, No. 45, 125 157.

Bruijnzeel, L A and Bremmer, C N (1989) *Highland- Lowland Interactions in the Ganges Brahmaputra River Basin: a Review of Published Literature*, ICIMOD Occasional Paper, No. 11, International Center for Integrated Mountain Development, Kathmandu, Nepal.

Bruijnzeel, L A and Critchley, W R S (1994) *Environmental Impacts of Logging Moist Tropical Forests*, UNESCO/IHP/Free University, Paris, 1 50.

Burgess, P F (1969) Ecological Factors in Hill and Mountain Forests of States of Malaya, *Malayan Nat. J.*, 22, 119 128.

Canonizado, J A (1978) Simulation of Selective Forest Management Regime, *Malay. For.*, 41, 171 175.

Chomitz, K M and Kumari, K (1996) *The Domestic Bene ts of Tropical Forests. A Critical Review Emphasizing Hydrological Functions*, Policy Research Working Paper 1601, The World Bank, Policy Research Department, Washington, DC.

DID (1989) Sungai Tekam Experimental Basin Final Report, July 1977 to June 1986, *Water Resources Publication No. 20*, Drainage and Irrigation Dept., Ministry of Agriculture, Kuala Lumpur, Malaysia.

Douglas, I, Greer, A, Kawi Bidin, J, and Sinun, W (1993) Impacts of Road and Compacted Ground on Post-Logging Sediment Yield in a Small Drainage Basin, Sabah, Malaysia, in *Hydrology of Warm Humid Tropical Regions*, ed J S Gladwell, *IAHS Publ.*, 216, 213 218.

Dykstra, D P and Heinrich, R (1996) *FAO Model Code of Forest Harvesting Practice*, Food and Agriculture Organization of the United Nations, Rome, 1 85.

FAO (1997) *State of the World's Forest 1997*, Food and Agriculture Organization, Rome, 1 200.

Fox, J E D (1968) Logging Damage and the Influence of Climber Cutting Prior to Logging in the Lowland Dipterocarp Forest of Sabah, *Malay. For.*, 31(4), 326 347.

Gilmour, D A (1977) Effect of Rainforest Logging and Clearing on Water Yield and Quality in a High Rainfall Zone of Northeast Queensland, Australia, in *First National Symposium for Hydrology*, eds E M O Loughlin and B J Bren, Institution of Engineers, Melbourne, 156 160.

Hewlett, J D (1982) *Principles of Forest Hydrology*, University of Georgia Press, Athens, Georgia, 1 192.

Hsia, Y J and Koh, C C (1983) Water Yield Resulting from Clearcutting a Small Hardwood Basin in Central Taiwan, *Int. Assoc. Hydrol. Sci. Publ.*, 140, 215 220.

Johns, J S, Barreto, P, and Uhl, C (1996) Logging Damage During Planned and Unplanned Logging Operations in the Eastern Amazon, *For. Ecol. Manage.*, 89, 59 77.

Kuraji, K and Paul, L L (1994) Effects of Rainfall Interception on Water Balance in Two Tropical Rainforest Catchments, Sabah, Malaysia, in *Proceedings of the International Symposium for Hydrology*, ed T Ohta, University Tokyo, Tokyo, 291 298.

Lai, F S (1993) Sediment Yield from Logged, Steep Upland Catchments in Peninsular Malaysia, *IAHS Publ.*, 216, 219 229.

Malmer, A (1993) Dynamics of Hydrology and Nutrient Losses as Response to Establishment of Forest Plantation. A Case Study on Tropical Rainforest Land in Sabah, Malaysia, Published PhD Thesis, Swedish University of Agricultural Sciences, Sweden.

Moura-Costa, P (1997) Reduced Impact Logging Techniques as a Means for Carbon Offsets, in *Proceedings of the FAO/IUFRO Satellite Meeting* held in conjunction with the IUFRO XX World Congress, Tampere, Finland, 4 5 August, 1995, Research on Environmentally Sound Forest Practices to Sustain Tropical Forests, Food and Agriculture Organization of the United Nations, 37 45.

Nicholson, D I (1958) An Analysis of Logging Damage in the Tropical Rain Forest, North Borneo, *Malay. For.*, 21(4), 235 245.

Nykvist, N, Grip, H, Sim, B L, Malmer, A, and Wong, F K (1994) Nutrient Losses in Forest Plantations in Sabah, Malaysia, *Ambio*, 23(3), 210 215.

Oyebande, L (1988) Effects of Tropical Forest on Water Yield, in *Forest, Climate and Hydrology: Regional Impacts*, eds E R C Reynolds and F B Thompson, The United Nations University, Japan, 1 217.

Pearce, A J (1986) *Erosion and Sedimentation*, Working Paper, Environment and Policy Institute, Honolulu, HI, 1 18.

Sist, P, Dykstra, D, and Fimbel, R (1998) *Reduced-impact Logging Guidelines for Lowland and Hill Dipterocarp Forests in Indonesia*, Occasional Paper No. 15, Center for International Forest Research (CIFOR), Bogor, Indonesia, September 1998, 1 50.

Studtmuller, T (1987) *Cloud Forests in the Humid Tropics: a Bibliographic Review*, The United Nations University, Tokyo and Centro Agronomica Tropical de Investigacion y Ensenanza, Turrialba, Costa Rica.

Thang, H C (1997) Concept and Basis of Selective Management System in Peninsular Malaysia, in *Proceedings of the Workshop on Selective Management System and Enrichment Planting*, eds W A Wan Yusoff, A R Abdul Rahman, and M N Mohd. Nizum, Forestry Department Headquarters, Kuala Lumpur, 1 14.

Watanabe, S (1992) Percentage of Felling in the Natural Forest and Damage caused by Felling Operations a Case of Natural Forest in Lokkaido, Japan, in *Beyond the Guidelines - an Action Program for Sustainable Management of Tropical Forests*, International Tropical Timber Organization Technical Series No. 7, 1 146.

Wiersum, K F (1984) Surface Erosion Under Various Tropical Agroforestry Systems, in *Proceedings Symposium on Effects of Forest Land Use on Erosion and Slope Stability*, eds C L O Loughlin and A J Pearce, IUFRO, Vienna and East West Center, 231 239.

World Water (1981) How can Tree Combat Droughts and Floods, *World Water*, 4(10), 18, October.

Zulkifli, Y, Abdul Rahim, N, and Baharuddin, K (1993) *Selective Logging in Malaysia: Impacts on Streamwater Chemistry*, Paper Presented at International Symposium on Forest Hydrology, Canberra, November, 1993.

Zulkifli, Y, Baharuddin, K, and Abdul Rahim, N (1998) Estimating Rates of Nutrient Recovery Following Timber Harvesting in a Second Growth Forest of Peninsular Malaysia, *Forestry Science*, 54, Kluwer Academic Publisher, Dordrecht, 419 429.

FURTHER READING

Hsia, Y J (1987) Changes in Storm Hydrographs After Clear Cutting at a Small Hardwood Forested Watershed in Central Taiwan, *For. Ecol. Manage.*, 20, 117 133.

Moura-Costa, P (1996) Tropical Forestry Practices for Carbon Sequestration: a Review and a Case Study from Southeast Asia, *Ambio*, 25(4), 279 283.

Freshwater Fisheries

see Freshwater Fisheries (**Volume 2**)

Fugacity

Fugacity is a thermodynamic criterion of equilibrium for diffusion of mass within and between phases, and was developed by G N Lewis in 1901 as a surrogate for chemical potential. It can be regarded as the escaping or fleeing tendency of a chemical. It is usually represented by the symbol f and has units of pressure, namely pascals (Pa). For most environmental purposes, fugacity is equal to partial pressure. In many respects, fugacity is analogous to temperature as a criterion of heat equilibrium. It is proportional to concentration, and a proportionality constant (or Z value) is used to convert one to the other. These Z values depend on the properties of the chemical and the medium in which it is dissolved or sorbed. The use of fugacity is advantageous when assessing the fate of chemicals in multimedia environments because it standardizes the escaping pressure or chemical potential in diverse media such as air, water and biota so a uniform measure of chemical potential is applied to all media. Fugacity can be used to reveal the equilibrium status of media in contact, e.g., water and fish, and it can be used (in conjunction with appropriate transport rate parameters) to quantify the direction and rate of movement of a chemical, since a chemical will migrate or diffuse from media of high fugacity to those of low fugacity. The rate of chemical movement is also dependent on the fugacity gradient or difference between the two media, with larger fugacity gradients resulting in greater fluxes. It is thus analogous to Fourier s Law for heat conduction. Fugacity calculations are therefore invaluable when seeking to simulate and predict the rate of a chemical s movement within and between environmental media. Modeling results obtained using the fugacity system can be numerically identical to results from traditional concentration-based modeling systems that employ concentrations and partition coefficients, but the mathematics are usually simpler and more transparent in the fugacity system, especially when the system becomes more complex in terms of the number of compartments and processes. A wide variety of mass balance models based on fugacity principles is available.

For details of applications of fugacity to environmental simulations, see Mackay (2001) and Cowan *et al.* (1994).

REFERENCES

Cowan, C E, Mackay, D, Feijtel, T C J, van de Meent, D, Di Guardo, A, Davies, J, and Mackay, N (1994) *The Multi-Media Fate Model: a Vital Tool for Predicting the Fate of Chemicals*, Proceedings from a Workshop Held at Leuven, Belgium, April 14 16, 1994 and in Denver, CO, November 3 4, 1994, SETAC Press, Pensacola, FL.

Mackay, D (2001) *Multimedia Environmental Models: the Fugacity Approach*, Lewis Publishers/CRC Press, Boca Raton, FL.

DONALD MACKAY *Canada*

G

Geoengineering: a Way to Stabilize Global Climate?

Gregg Marland

Oak Ridge National Laboratory, Oak Ridge, TN, USA

Many human activities are now carried out at such a large scale that they change the environment at regional to global scales. For example, burning fossil fuels releases greenhouse gases to the atmosphere, producing food and ber modi es the earth's surface, and building dams and irrigation systems alters the hydrologic cycle. Such activities have the goal of serving human needs or desires, but they are also, unintentionally, changing the global environment. Can we have large-scale activities deliberately designed to prevent the global environment from changing? Geoengineering suggests that that we might indeed do this on a very large scale.

The term *geoengineering* was first used in 1975 when Cesare Marchetti proposed a plan to collect the carbon dioxide discharged from coal-fired power plants in Europe and to inject it into the stream of subsurface water flowing out of the Mediterranean Sea into the Atlantic Ocean. This geoengineering scheme would put a large quantity of carbon dioxide into the deep waters of the Atlantic Ocean rather than into the atmosphere where it could cause global warming. That very large-scale engineering project would be designed to protect the climate from changes caused by excessive burning of fossil fuels. It is this very large scale and purposefulness that characterize and define geoengineering.

The US National Academy of Sciences used the term geoengineering in 1992 to characterize options to combat or counteract the effects of changes in atmospheric chemistry . They included proposals to plant large areas of trees to remove carbon dioxide from the atmosphere, approaches for increasing the ocean s absorption of carbon dioxide, and options for screening out some of the sunlight reaching the Earth. Counteracting the effect of carbon dioxide from fossil fuel burning could involve activities on the order of millions of km^2 of forests or screening out 2% of incoming solar radiation.

It has been suggested that fertilizing the ocean could increase the biological productivity of ocean waters and thus increase the rate at which carbon dioxide is drawn from the atmosphere. The carbon would then be carried into deep ocean waters where it would be stored for a very long time. Experiments to test the idea that some areas of the ocean are limited in the micronutrient iron have demonstrated that over small areas and short times, adding iron to the surface waters can cause an increase in biological productivity and an increase in carbon dioxide uptake. The long-term consequences of this sort of fertilization are not well understood (*see* Iron Cycle, Volume 2).

Several engineering ways of stabilizing the global climate have been proposed. Mathematical models of the climate system suggest that doubling the concentration of the greenhouse gas carbon dioxide would be as effective in warming the Earth, on average, as would a 2% increase in radiation input from the sun. Rather than decreasing the rate at which we burn fossil fuels, and release carbon dioxide, perhaps it would be possible to geoengineer a decrease in the radiation input from the sun and thus to counteract the effect of increasing carbon dioxide. Possibilities that have been suggested for reflecting away part of the incoming solar radiation include space mirrors, stratospheric dust, large numbers of reflective balloons, and sulfate aerosols in the atmosphere. A recent study suggests that a mesh of very fine aluminum wires in orbit around the earth could be designed to scatter solar radiation and thus to reduce solar input to the Earth by any desired amount. Sulfate aerosols from large volcanic eruptions have indeed resulted in large scale, temporary cooling of the Earth s surface.

Geoengineering might also be used to alter internal processes in the Earth s energy or hydrologic balances. But because of the close coupling of many processes on the Earth s surface is not well understood, it might be difficult to predict consequences and side-effects. Nevertheless,

there are a number of modifications that might be considered to manage the global environment. Modifications in the magnitude or distribution of cloudiness, the circulation of the oceans, the paths of hurricanes, or the area or brightness of ice have all been considered.

A major challenge of geoengineering generally is that we do not understand the Earth System well enough to know the full range of consequences of these activities. Might geoengineering result in positive outcomes in some places and at some times and negative outcomes in other places and at other times? Given this uncertainty, some have suggested that any geoengineering should be limited to activities that can be easily reversed or to activities that are within the range of natural variability.

The attraction of geoengineering is that some of the possibilities appear to be very inexpensive and easy to implement. Counteracting the effect of increasing concentrations of greenhouse gases might be much less expensive than trying to reduce emissions, and it would not depend on the behavior of large populations. It could be simply a matter of deciding what to do and how to share the costs. The issues of risk, responsibility, law, equity, and ethics are not as easy to assess and to deal with.

It has been suggested that geoengineering might provide the last-resort alternative if we are not able to otherwise prevent significant changes in the Earth s climate and we face major impacts on human welfare. Others worry that the mere possibility of geoengineering solutions could weaken our resolve to find other ways to avoid potential problems. Before any geoengineering scheme could be responsibly considered we would need to better understand the complex Earth System than we do now.

FURTHER READING

Flannery, B P, Kheshgi, H, Marland, G, and MacCracken, M C (1997) Geoengineering Climate, in *Engineering Response to Global Climate Change*, ed R G Watts, Lewis, Boca Raton, FL, 379 427.

Marchetti, C (1975) *On Geoengineering and the CO_2 Problem*, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Marland, G, ed (1996) *Climatic Change, a Special Issue on Geoengineering*, 33(3).

National Academy of Sciences (1992) *Policy Implications of Greenhouse Warming: Mitigation, Adaptation, and the Science Base*, National Academy Press, Washington, DC, 433 464.

Teller, E, Wood, L, and Hyde, R (1997) *Global Warming and Ice Ages: I. Prospects for Physics-based Modulation of Global Change*, UCRL-JC-128715, Lawrence Livermore National Laboratory, Livermore, CA, 1 20.

Geographical Information Systems (GIS)

see Geographic(al) Information Systems (GIS) (**Volume 4**)

Geography

Geography is a discipline concerned with understanding the spatial dimensions of environmental and social processes. As defined by the editors of the recent review of advances among American geographers, it is the study and science of environmental and societal dynamics and society-environment interactions as they occur in and are conditioned by the real world (Gaile and Willmott, 2001).

While some geographic study focuses primarily on the distribution of human populations and economic and political activities, a wide variety of investigations have a major environmental emphasis. The latter include: studies of how human activities have altered the global environment; examination of changes in sectors of various natural systems; studies of people nature interactions and their consequences in major regions; and detailed examinations of smaller areas, encompassing either one sector or all sectors of the environment.

Geographers, beginning in 1864 and most recently in 1990, have made a series of reviews of the quantity and quality of human modifications in global land, water, and organic resources (Thomas, 1956; Turner *et al.*, 1990). These have been accompanied by more detailed analyses of changes in vegetation, land use, water management in selected regions, such as for arid and semi-arid lands.

Geographers contributed to appraisals of the possible natural and human consequences of changes in the global carbon cycle (Mather and Sdasyok, 1991; Bolin *et al.*, 1986), and they assessed the social implications of the hazards caused by extreme natural events in high-risk regions (Kasperson *et al.*, 1995; Burton *et al.*, 1993).

On a more local scale, they have carried out an immense number of detailed studies of how human actions have caused ecosystem change, and of how such change is, in turn, reflected in the pattern and quality of social activity.

With their emphasis upon integrated spatial analysis and upon human nature interactions, many geographic investigations are distinguished by their emphasis on regional scales, and by bringing together natural and social sciences.

REFERENCES

Bolin, B, Doos, B R, Jager, J, and Warrick, R A, eds (1986) *The Greenhouse Effect: Climate Change and Ecosystems*, SCOPE 29, John Wiley & Sons, Chichester.

Burton, I, Kates, R W, and White, G F (1993) *The Environment as Hazard*, Guilford Press, New York.

Gaile, G L and Willmott, C J (2001) *Geography in America*, Merrill, Columbus, OH, in press.

Kasperson, J X, Kasperson, R E, and Turner, B L, eds (1995) *Regions at Risk: Comparisons of Threatened Environments*, UN University Press, Tokyo.

Mather, J R and Sdasyak, G V, eds (1991) *Social Change: Geographical Approaches- a Joint USSR- USA Project*, University of Arizona Press, Tucson, AZ.

Thomas, Jr, W L, ed (1956) *Man's Role in Changing the Face of the Earth*, University of Chicago Press, Chicago, IL.

Turner, B L, Richards, J F, and Meyer, W B, eds (1990) *The Earth as Transformed by Human Action*, Cambridge University Press, New York.

GILBERT F WHITE *USA*

Geomorphological Change for Urbanization and Industry

Ian Douglas
University of Manchester, Manchester, UK

Mining and urban development both involve removal of the natural vegetation cover and excavation of the topsoil and often much of the underlying weathered rock and bedrock layers. In new urban developments, small streams are often diverted into culverts or urban drains and minor depressions and valleys are lled in. Steep hillsides may be terraced into a series of home sites by cut and ll operations. Major rivers may be embanked and arti cially straightened. In extreme cases, as in Palma de Mallorca, Spain and Winnipeg, Canada, large new ood channels may be built around the urban center to divert ows away from the city. The new features, replacing the original landforms, are often designed to direct water away from the new developments more effectively, so producing off-site, downstream consequences.

Mining changes the land surface by creating pits and quarries, removing underground material and building up waste tips and tailings deposits. The largest excavations take up many square kilometers of the land surface, while waste tips formerly dominated the skylines of the old deep coal mining areas of Europe.

In extreme cases, the legacies of past mining activity still affect the morphology of major rivers as in the Sacramento River in California, US. Lakes in areas of former gravel working are prominent in the valleys of many rivers, for example, the Thames, Rhine and Seine Rivers in Europe. Removal of mineral resources and water from beneath the ground leads to subsidence creating new surface topographies and often new water bodies (Carbognin, 1985).

These changes in landform may often be individually small, but collectively they are the result of two of the main human drivers of global environmental change: increasing urban development and the mining and quarrying which supplies minerals for industry, and the construction materials required to build all the infrastructure, homes, of ces and factories of cities. Gradually, over the centuries, cities renew their built fabric, often using the demolished remains of earlier buildings as ll and foundations for the latest developments. Thus, the urban surface is gradually raised, again changing the form of the land surface in another way. An exception occurs in cities like Mexico City, where land subsidence is taking place.

In creating these new landforms and altering the workings of geomorphic processes, urbanization and mining have impacts far beyond their municipal or concession boundaries. Impacts of river systems, lakes, wetlands, estuaries, deltas and coastal zones may be particularly marked and persistent. Water quality in some rivers of the Pennines in northern England is still affected by drainage from Roman Pb workings. The impact of the transformation of some major rivers on New Guinea and Bougainville by the release of mine tailings and associated chemicals has been felt in offshore sheries in adjacent seas. Everywhere that rock or clay is extracted, that groundwater is pumped to the surface, coal mined or oil extracted, there are environmental consequences and landform changes. While modern mining technology and regulations can ensure successful restoration of changed landscapes and treatment of mining wastes, not all mines, quarries and wells in all localities in all countries adopt appropriate reclamation practices or have local regulations that are strictly enforced. Thus, the extraction of liquids and materials by mining and quarrying, and the modi cation of the land surface for urban development, will continue to contribute to environmental change long into the future.

LANDFORM CHANGE IN STEEPLANDS

Excavations in mountainous areas are usually highly visible and potentially dangerous. The problems experienced with the collapse of a tailings dam at the Freeport Copper Mine high in the Star Mountains of New Guinea illustrate how earthworks failures can affect hundreds of

kilometers of river channel downstream. Many important mineral resources can only be obtained by working in areas of high instability risk, and restoration after operations have ceased poses particularly difficult problems.

Mountain top removal and valley fill mining as practiced in West Virginia, US involves removing 100 400 m depth of rock from the crest of Appalachian mountain ridges to expose successive seams of coal. The overburden is dumped into adjacent valleys, covering streams and drainage lines, forming a valley fill. The largest individual mines occupy up to 40 km^2. Although now separated by green, forested, intact mountains, many will soon merge together as reclamation creates a new series of landforms of rolling topography, instead of steep ridge and valley relief (Kite and Skousen, 1999). Already, over 2% of the surface of West Virginia has been strip-mined since the 1977 Surface Mining Control and Reclamation Act. At the mine sites, huge draglines work continuously, leveling ridge crests in a matter of months. Instead of forests, many of the valley fills are converted to grassland for cattle grazing. Others are used for recreation as golf courses or football pitches, or for occasional large new shopping malls.

Similar topographic changes in mountain areas are caused by the development of new urban and resort areas in tropical steeplands. On the Main Range, east of Kuala Lumpur, Malaysia, at approximately 2000 m elevation, the Genting Highlands and Bukit Tinggi resort and residential area developments have led to the excavation of large areas of deeply weathered granite and the filling of many minor valleys. In an originally steep, rugged terrain, new golf courses and playing fields have been created, with the surrounding slopes modified into cut and fill terraces for the construction of houses and apartment buildings.

These huge earth-moving operations not only cause major environmental changes in themselves, but they have marked impacts on adjacent ecosystems, local stream flow hydrology and sediment loads, and slope stability. Water quality impacts can generally be controlled during active mining, but many coal seams cannot be mined without some residual acid mine drainage requiring treatment to protect nearby rivers. Valley fills and reclaimed strip mine and mountain top areas operate as artificial, porous areas through which rainwater penetrates to recharge groundwater bodies. Some of the infiltrated water emerges from the fill as highly acid seepages that often flow throughout dry periods when neighboring streams are extremely low. Sometimes the adjacent watercourses have alkalinities below 10 $mg\,l^{-1}$ and are unable to immediately neutralize inputs from reclaimed mine area springs with 1000 2000 $mg\,l^{-1}$ of acid. While reclamation activity offers real opportunities for habitat restoration and improvement, including the amelioration of fish stocks in rivers, some mining operations are unregulated and some companies fail to follow approved mining plans. Much mining in the 19th century had no regulation at all. The legacies of these past activities, and those of present poor practice, continue to affect fish and aquatic resources through erosion and sedimentation, dewatering of wetlands, diverting and channelizing of streams and contamination of surface water and aquifers with toxic chemicals. Some of these mining impacts are strictly geomorphological, altering processes and landforms in river channels and floodplains. Others permeate through food chains and change ecosystems, eventually having impacts on lakes and coastal seas.

SUBSIDENCE DUE TO MINING AND WITHDRAWAL OF FLUIDS

Land subsidence due to coal mining may be caused by the removal of the coal underground. Gravity and the weight of the overlying rock causes the layers of rock to shift and sink down into the void left by the removal of coal. Ultimately, this process can affect the surface, causing the ground to sag and crack and holes to form, which may severely damage or destroy buildings.

Occasionally, the impact of mining subsidence on the environment is catastrophic, destroying property and even leading to the loss of life. Most frequently however, such subsidence causes slight to severe structural damage. The geological situation, actual mining technique and type of mineral deposit determine the nature of the subsidence in any particular locality. Some mining methods result in contemporaneous subsidence, whereas, with others, subsidence may occur long after the mine workings have been abandoned. In the latter instance, it is more or less impossible to predict the effects or timing of subsidence. In the cases of gold mining in the Johannesburg area; room and pillar mining of coal in the Witbank Coalfield, South Africa; longwall mining of coal in the Ruhr district; mining of chalk in Suffolk, England and limestone in the English West Midlands, and solution mining of salt in Cheshire, England, the deposits concerned have often been worked for more than 100 years. Abandoned mines with unrecorded old workings not far below the surface may pose major problems when old mining and industrial areas are being redeveloped (Bell *et al.*, 2000).

Land use after deep mine-induced subsidence may create attractive opportunities. At Pennington Flash near Leigh, west of Manchester, England, subsidence following deep coal mining during the 19th century led to a water-filled depression, which developed around 1890. The depression quickly became a shallow lake that soon began to be used for recreation, particularly water sports such as sailing. After 1970, the area was developed into a country park, with restoration of adjacent derelict land and creation of wildlife refuges. The subsidence-affected area became fully integrated into a new phase of urban regeneration.

Like so many other water bodies, Pennington Flash has a high wildlife and biodiversity potential. While recognized at the local management level, the full potential of so many water bodies related to past extractive industries at the regional level, is often neglected. The full biodiversity and sustainability contribution of these legacies of extractive industries only occurs when they are seen as parts of systems of urban greenspace corridors and nodes within complex metropolitan areas. When thinking of the sustainability of cities and managing materials flows, the whole sequence, from natural land surface, through mining or quarrying, to possible landfilling, restoration and ultimate urban or greenspace uses, has to be considered. Material flows are thus intimately related to land uses and landform changes.

Removal of liquids has had particularly marked impacts on many cities. In the Houston, Texas, US areas, land subsidence first occurred in the early 1990s in areas of groundwater, oil and gas extraction. The subsidence continued throughout the 20th century, mainly because of further groundwater abstraction. The localities and extent of subsidence changed during that time. Before the early 1940s, removal of oil, gas and associated brine caused localized subsidence close to Houston. Nearer the coast, groundwater pumping for public supplies caused more than 40 cm of subsidence between 1906 and 1943 at Texas City. Further exploitation of the groundwater as the only source of supply led to subsidence affecting a large part of Houston, but mainly by less than 30 cm before 1943. The expansion of the petrochemical industry in the 30 years until 1973 put much greater pressure on groundwater resources with nearly 2 m of subsidence occurring along the ship canal between Bayport and Houston. By 1979, the most severely affected areas had been lowered by over 3 m, increasing their vulnerability to hurricanes and storm surges. The Brownwood subdivision of Baytown that had subsided by nearly 2.5 m was so inundated by a 3.5 m storm surge during hurricane Alicia in 1983 that all the homes in the area were abandoned. Now most of the subdivision is a swampy waterfowl habitat (Galloway *et al.*, 2000).

Since the late 1970s, subsidence has largely been arrested along the ship canal and in the Baytown-LaPorte and Pasadena areas, as the development of surface water supplies has allowed groundwater pumping to be reduced. However, as subsidence in the coastal area was stabilizing, subsidence farther inland, to the north and west of Houston, increased with the abstraction of freshwater from the Evangeline aquifer. More than 75 cm of subsidence occurred at Addicks, about 30 km west of Houston, between 1973 and 1996.

Groundwater exploitation has produced subsidence in many cities, among the most well known being Mexico City, Bangkok, Venice and Bologna. In Mexico City, groundwater withdrawal has produced 9 m of land subsidence in places. Most of this compaction has occurred in the uppermost 50 m below the land surface in two highly compressible silt clay beds 25 30 m and 5 10 m thick. From 1877 to 1952, the rate of surface lowering increased steadily, reaching a peak of 30 cm year^{-1} in 1948 1952. After 1952, subsidence slowed to about 5 cm year^{-1} in 1973. Since then, pumping of underground water in the central part of the city has been banned and further subsidence is not a serious problem (Carbognin, 1985).

In Bangkok, the rapid growth of the city outstretched the ability of the municipal authorities to provide the additional water needed, and many private boreholes were sunk. Local settlement of as much as 14 cm year^{-1} occurred, but because the complex sediments of the Chao Phya River deltaic plain had been disturbed and faulted by local earth tremors, differential lowering took place. As one sector of a street or one length of canal was lowered relative to its neighbors, drainage was disrupted and sewers were fractured. Despite the installation of new deep drains well below the zone of faulting-disrupted subsidence, parts of the city can still be inundated for many days during the heavy monsoon rains because of poor drainage.

Groundwater pumping has put the historic world heritage buildings of Venice at risk. Built at sea level on the lagoon, Venice has subsided some 22 cm since 1900. Most of that surface lowering occurred between 1950 and 1970. High water (aqua alta) has occurred more frequently since 1970. Whilst the people-induced subsidence is part of the problem, higher extreme sea levels related to global climate change may possibly be another factor.

In Bologna, groundwater exploitation has affected the historic core of the city, with lowering through differential sinking of foundations of up to 6 8 cm year^{-1}. These irregular changes pose major problems for building maintenance because of the stresses they place on structures. Many other historic Italian cities, including Ravenna, have similar concerns. Few cities where there is exploitation of groundwater, oil or natural brines beneath their buildings can avoid some degree of subsidence. The problem is perhaps most insidious in coastal areas, where the combination of subsidence with rising sea levels greatly increases the vulnerability to the effects of climate change.

QUARRYING AND LANDFORM CHANGE

Around Buxton in the Peak District of England, limestone quarries like Tunstead and Dove Holes have doubled the area of landforms changes by minerals extraction since 1920. Limestone remains the most important non-fuel mineral extracted in Britain, but its removal changes the landscape far more quickly than natural processes of weathering and erosion (Gunn and Gagen, 1989). The natural rate of erosion, or rock removal, by solutional processes

in the Peak District is approximately $70\,m^3\,km^{-2}\,year^{-1}$, equivalent to an annual loss of some 83 500 tons of limestone per year. The amount of limestone removed annually by quarrying rose from 2 million tons (Mt) per year in the period 1900 1920, to 5 Mt by 1949, to a peak of 22.5 Mt in 1973; since then it has fallen back to a more or less steady output of about $18\,Mt\,year^{-1}$. Extrapolating from these figures, the tonnage of limestone removed by human activity in just 100 years would equal the amount lost by natural solution processes over the whole 10 000 years of the Holocene (Gunn and Gagen, 1989).

Quarries can create two situations; voids in the ground or the removal of a topographic feature, such as a mountain top or a ridge of glacial gravels. In the latter case, a topographic feature has been removed by minerals extraction to create new surface landforms. Restoration usually involves grading and consolidating the surface before restoring the topsoil, which is usually removed and stockpiled prior to quarrying. Reshaping of landform remnants, perhaps to recreate something like the original form, may be part of the restoration.

With voids in the ground and conventional quarries, the options for restoration range from retaining the void and converting it to another use, perhaps as a lake or as a site for recreation or industry, to filling the hole and converting it to a landform akin to the original. Use of quarries as landfill sites is common enough and in some cases aggregate removal and landfilling take place one after the other as part of a single operation on one site. Containment systems, with liners of clay and strong plastics, gas and water abstraction pipes and regular daily covering of the waste, are needed to prevent impacts on surrounding areas. When filling is complete, the original land surface contours can be restored and the new surface returned to agriculture or some other land use. Where infilled quarries are later used for new buildings, problems of differential consolidation may be encountered. Several instances of settlement in former chalk pits filled decades ago occur every time there is a severe drought in southeastern England. Filling of old alluvial tin mine workings around Kuala Lumpur, Malaysia has occasionally produced similar problems. The alluvial tin is in a deposit overlying a buried limestone karstic area with many caverns in the bedrock. In a few places, the fill in the tin mine pits has been washed into the underlying karst cavities and buildings on the ground surface have suffered severe subsidence.

Restoration of the English Peak District quarries in an area of high scenic value has used blasting techniques to create more natural quarry faces that plants and animals can colonize more effectively (Cullen *et al.*, 1998; Gagen and Gunn, 1988). Careful landscape design on the floor of the quarry can create water bodies and a variety of habitats. Typical uses for quarry floors include construction, such as shopping malls, or industrial units, with protection of the untreated quarry faces for their geological conservation value.

Sometimes pumping of water from an operating quarry leads to landform change elsewhere, especially in karstic terrain. A large quarry near Rome, Georgia, US, was sunk about a kilometer from the campus of Berry College. Eventually, the quarry workings intersected the water table and water had to be pumped out of the quarry to enable working to continue. This pumping drew down the potentiometric surface in the surrounding rocks. In the limestone below the water table, water moved preferentially towards the quarry through passages enlarged by solution. The water entrained soil and weathered rock particles. This removal of material eventually caused a lowering of the ground surface and the appearance of a depression (sinkhole) at the surface. Many sinkholes appeared on the Berry College campus, some threatening buildings. The water of an artificial lake disappeared into one of the sinkholes (Kastning, 1998). The quarrying had introduced a phase of extraordinarily rapid karst landform development. Sinkhole development is a major problem in many karst areas undergoing urban development and water resource exploitation.

LANDSCAPE MODIFICATION BY AGGREGATE AND BRICK CLAY EXTRACTION

The excavation of building materials, such as brick- and tile-making clays, sand and gravel, which are low in value in relation to their bulk, has traditionally taken place close to the settlement where they are to be used. The Tuileries Gardens in Paris occupy the site of pits where clay was dug to make tiles (tuiles) for the city s roofs. The present Parisian park at Buttes-Chaumont was a belt of quarries and rubbish dumps until the mid-19th century. The 348 ha Central Park in New York was created in 1858 from an area of quarries, swamps, dumps and squatter housing. Valuable agricultural land is being lost around most Indian cities as farmers dig clay to sell to make bricks. Similar local clay extraction is seen around towns in Vietnam.

Both modern river floodplains and former river channels near cities are favored sources of sand and gravel. London has gravel workings along the present Thames floodplain and in the Vale of St Albans where the Thames flowed previously. Glasgow, on the other hand, has exploited gravels deposited under past glaciers. Floodplain gravel workings are operated as wet pits, and are usually left as water-filled areas when operations cease. Such sites are easily converted to lakes for water sports or wildlife refuges. However, the opportunities for successful conversion to recreational and environmental uses may be approaching their limit in some places. For example, the rapid increase in the area of former gravel pits near the head of the River Thames northwest of

Swindon, England has begun to outstretch the demand for water-based recreation and for managed wildlife refuges. Already the Cotswold Water Park provides large areas for water sports, while other pits are used for waterside caravan and weekend home parks or have been made into wildlife reserves. There is unlikely to be enough waste that can be hauled economically and be deposited safely, in gravels overlying limestone aquifers, to justify using these sites for landfill, which in any case is being discouraged by the UK landfill tax. In the Rhine Rift Valley south of Mannheim, Germany, a similar situation exists. The number of water-filled gravel pits has increased greatly since 1960. Here is an example of landform change that begins to sterilize land for future uses, just as the spread of cities over areas of economic clay or gravel resources, sterilizes that resource.

Demands for construction materials in many countries are so great that gravels are extracted directly from active river channels, for example around Beijing, China. In Java, Indonesia, the debris brought down channels by lahars from active volcanoes is exploited as construction material. In Kuwait, sand and gravel extraction often causes severe deflation and dust storms. As urban areas expand, the environmental changes due to construction materials extraction will become more intense and increasingly detrimental to adjacent ecosystems and watercourses, eventually affecting biogeochemical cycling.

IMPACTS OF AGGREGATE EXTRACTION ON ESTUARIES

In the Georges River in southern Sydney, England, three different local changes have affected the estuary since 1940: (a) a shift to a flood-dominated regime from a drought-dominated one in the late 1940s, which increased both water and sediment discharge (like urbanization) in a natural way; (b) sand dredging from the bed of the river in several upstream locations; and (c) dredging and quarrying of floodplain and terrace sediments at Liverpool and Chipping Norton. The shift to a flood-dominated regime:

(a) introduced more water and sediment naturally to the estuary from the 930 km^2 catchment after about 1948.

Sand dredging from the bed of the river created deep holes:

(b) that greatly increased the sub-tidal capacity, thereby increasing the flushing time to evacuate pollutants. The deep holes broke the continuity of the bed, which helped to promote erosion both upstream and downstream of the dredging sites. Additionally, the deep holes helped to undermine the banks in the dredged reaches. Thus, bank erosion became a problem.

Floodplain and terrace quarrying and dredging

(c) should have caused no problems to tidal dynamics, because they were essentially pits away from the river with a 20 m zone left along the banks. However, these banks failed with the passage of floods, and the ponds became connected with the tidal river, which created more problems. The tidal prism, which diminishes towards the upper part of the estuary, was greatly increased by having to fill the ponds on incoming tides and empty them on outgoing tides. This greater volume of water had to pass through the downstream channel more quickly, and thus also helped to promote bank erosion in the middle reaches of the tidal river, which had also been destabilized by dredging. The depth of the ponds dredged on the floodplains greatly increased the sub-tidal capacity at the upper end of the tidal channel. This helped to increase the flushing times by more than 100% (Bridgman *et al.*, 1995).

So the dynamics of the estuary were greatly modified by urbanization and sand dredging for aggregate. In the lower parts, both sub-tidal capacity and tidal prism volume were decreased, and in the upper parts, both were increased. The consequences of this were higher tidal current velocities that helped to promote bank instability and poorer flushing potential in an estuary where major sewerage works were built to discharge treated sewage into the river and the estuary.

COASTAL SUPERQUARRIES IN BRITAIN

The underestimation of demand for aggregates in the 1980s led to the exploitation of rock reserves in areas of high scenic value such as the Mendips and the Peak District National Park. Coastal superquarries provide a viable economic alternative to the further disturbance of such sensitive areas, capable of supplying much of the growth in demand in southeast England and possibly other regions. The first British coastal superquarry occupies a 250 ha estate at Glensanda near Fort William, producing 5 Mt of granite a year, with plans to export 15 Mt per year. The site has sufficient reserves for a hundred years or more. Most of Glensanda s output is, at present, exported to northern Europe and the US. A second site, the proposed Rodel quarry at the southern tip of Harris, has so far been refused planning permission and is the subject of an ongoing public enquiry. This is partly because of concerns over the impact of creating the largest man-made hole in Europe (2•4 • 1•6 km and 500 m deep), with a 1 km long deep sea loch. The effects of water discharges, noise and dust might affect the local environment and fisheries. A further geomorphic impact of the growth of superquarries

would be the effects on estuarine morphodynamics of the construction of up to 12 wharf and port facilities in the Thames and Medway estuaries to transfer the aggregates from ships to trucks. Each wharf site would need at least 5 ha of land, and access roads to accommodate at least 600 vehicle movements (including return trips) per day. The roads and wharves themselves would need more aggregates and other materials and so would add to pressure on resources. The scale of future geomorphic change by superquarry development probably depends on acceptance or rejection of the argument that six superquarries on the Scottish coast are preferable to 45 large inland quarries in the Mendips, Leicestershire or the Peak District, or even 450 sand and gravel pits in south-east England. Promotion of the use of secondary, recycled material would lessen the geomorphic impact of increasing demand for aggregates.

These examples of the impacts of aggregate extraction show that the impacts of simple needs for materials for housing and shelter collectively build up into major environmental modifications.

LAND RECLAMATION BY DUMPING AND INFILLING

Many cities around the world have had to seek to create landforms to house their activities. New airports have been built on artificial islands or peninsulas at Chek Lap Kok Airport, Hong Kong and Kansai International Airport in Osaka Bay, Japan. Runways have been extended into the sea at Sydney and Beirut. The quantities of material involved are large. 102.1 Mm^3 of material were moved to build Chek Lap Kok Airport. Construction of the Airport Island Seawall for the new Central Japan International Airport offshore of Tokoname City, Aichi Prefecture will involve moving some 13 Mm^3 of material.

Other cities have infilled coastal wetlands and reclaimed sectors of their foreshore. At Kota Kinabalu, the capital of Sabah, Malaysian Borneo, filling of the mangrove swamps behind coastal beach ridges for urban development removed the areas where floodwaters used to spread, causing parts of the town to be flooded during storms. In addition, large areas of the foreshore have been reclaimed, both for expansion of the city center and for tourist developments, including a marina and golf course. The disturbance during filling created high suspended sediment concentrations in the sea between the city and the offshore coral reef-fringed islands of the Tuanku Abdul Rahman National Park. The potential for damage to the reefs, themselves a major tourist attraction, is high. Such deliberate geomorphic changes always have some unwanted side effects.

A particularly fine example of an artificial coastal landform is provided by the Leslie Street Spit, which was created by depositing hundreds of thousands of tons of construction waste, large rubble, and other waste materials into Lake Ontario. The Spit just off Toronto, Canada has three cells each designed as a confined disposal facility for the dumping of dredged material. Cell 1 covers 8.2 ha of water area and was filled to capacity with 365 441 m^3 of dredged material by the end of 1987. At that point, disposal operations began in Cell 2, which covers 9.3 ha of water area and has an estimated capacity of 530 000 m^3. Covering an area of 32.1 ha, Cell 3 is the largest of the three cells and has an estimated capacity of 2.2 Mm^3. The total of material moved to make the Spit thus exceeds 3 Mm^3.

The Leslie Street Spit has little relief, and is covered with thin layers of vegetation. Some areas at the outer edge of the spit are in the stage of developing primary beaches. Small plants are invading the rubble and construction debris. However, around Cell 3 and the southern end of the spit, where dumping remains active, no natural beach development is occurring.

The Leslie Street Spit consists primarily of silt, fine sand and clay, the silt portion of the sediment accounts for 47 59% of the sediments, sand 18 34% and clay from 14 23%. The sediments are generally well consolidated with relatively few voids through which contaminated water could penetrate. The water around the Spit contains few contaminants, but some of the dumped material does not meet Canadian open water disposal criteria, although it is within the sediment guidelines for confined disposal. Concentrations of polychlorinated biphenyls and metals in the adjacent lake waters meet the requirements under the Canadian decommissioning guidelines for parkland, residential and agricultural soils. Concentrations of pesticides are also comparable to other waterfront locations in Toronto. These results demonstrate that the disposal cells are effective in their containment of the dredged material and that contaminants from the cells are not leaching into Lake Ontario. This Canadian example illustrates that, with adequate care and attention, surplus material can be dumped in a creative way to create new geomorphic features that create new land use or wildlife conservation opportunities. They are always a contribution to environmental change.

It is to be noted that an active citizens group FRIENDS OF THE SPIT has been active over the years to ensure that the Spit is developed as a conservation area, particularly as a natural site for studying ecological succession. See http://www.interlog.com/• fos/.

RE-USE OF DERELICT LAND (BROWNFIELD SITES)

In the old industrial areas of the world, the regeneration of derelict land is constantly taking place and much brownfield land has been re-used. Half of the new urban development in England has been taking place on land that had

Table 1 Land Justifying Reclamation in Greater Manchester

Year	Hectares of derelict land
1974	3186
1982	4025
1988	2734
1993	3227
1999	3217

previous urban, but not necessarily industrial uses. Much derelict land is highly concentrated in old coal-mining and heavy industry districts. For example, in 1990, the Nord-Pas de Calais region of France, with only 2.3% of the total area of France, had around 12 000 ha of derelict land, more than half the national total. In England, there were 39 600 ha of derelict land in 1995, of which some 27 000 were considered to be contaminated (Kivell, 1999). In 1999, Greater Manchester alone still had some 3217 ha waiting to be tackled (Table 1). Derelict land is being created almost as quickly as it is restored. Stricter regulations are requiring more effective removal of toxic chemicals (*see* Contaminated Lands and Sediments: Chemical Time Bombs?, Volume 3). Many sites are reclaimed for further industrial use or for new retail parks and shopping malls, but major political commitments are being made to endeavor to see that the majority of the need for new homes in countries like the UK is met by building on brownfield sites. Each restoration involves considerable demolition, filling and landscaping, so that the landform of the restored site may differ considerably from that of the original.

Yet, innovative initiatives are now starting to tackle the regeneration of derelict land, particularly closed landfill sites where risks of contamination may remain. Greater Manchester alone has over 900 closed landfill sites The Red Rose Forest, one of 12 semi-government community forest organizations in England and Wales is examining the potential of these sites for re-development as community woodland. Monitoring of 11 woodland initiatives on closed landfill sites in the region has demonstrated mean survival rate of over 80% for all tree species in the first year. The potential for using soil-making material derived from municipal waste for the beneficial regeneration of 23 closed landfill sites was being explored in 2000 2001. Thus, eventually derelict brownfield sites could be restored to new landforms supporting forest ecosystems capable of sequestering some of the carbon dioxide generated by activities on other restored sites.

CONCLUSION

The total area of actively operated, re-used or abandoned mineral workings is relatively small. For example, active mineral workings or former workings not yet reclaimed, and including land used for associated buildings and machinery, occupy 60 043 ha, 0.46% of all land in England. However, the associated landform changes are marked and have impacts well beyond the areas they occupy.

Good examples of landform restoration and control of mine effluents can be found. However, not all countries and local authorities can afford to restore areas of mining and quarrying or to regulate the impacts of the chemicals used in mineral ore processing or added to the environment by mine drainage. The geomorphic changes and associated environmental impacts caused by mining and urbanization will be of increasing environmental significance as more and more of the world s people live in towns and cities.

REFERENCES

Bell, F G, Stacey, T R, and Genske, D D (2000) Mining Subsidence and its Effect on the Environment: Some Differing Examples, *Environ. Geol.*, 40, 135 152.

Bridgman, H, Warner, R F, and Dodson, J (1995) *Urban Biophysical Environments*, Oxford University Press, Melbourne.

Carbognin, L (1985) Land Subsidence: a Worldwide Environmental Hazard, *Nat. Resour.*, 21, 2 12 (Reprinted in *The Human Impact Reader: Readings and Case Studies*, ed A Goudie (1997), Blackwell Press, Oxford, 20 32).

Cullen, W R, Wheater, C P, and Dunleavy, P D (1998) Establishment of Species-rich Vegetation on Reclaimed Limestone Quarry Faces in Derbyshire, UK, *Biol. Conserv.*, 84, 25 33.

Gagen, P and Gunn, J (1988) A Geomorphological Approach to Limestone Quary Restoration in *Geomorphology in Environmental Planning*, ed J M Hooke, John Wiley & Sons, Chichester, 121 142.

Galloway, D, Jones, D R, and Ingebritsen, S E (2000) Land Subsidence in the United States, *US Geological Survey Circular*, 1182, http://water.usgs.gov/pubs/circ/circ1182/.

Gunn, J and Gagen, P (1989) Limestone Quarrying as an Agency of Landform Change, in *Resource Management in Limestone Landscapes: International Perspectives*, eds D Gilleson and D Ingle Smith, Special Publication No. 2, Department of Geography and Oceanography, University College, Australian Defence Forces Academy, Canberra, 173 181.

Kastning, E H (1998) *Potential hydrogeologic and Environmental Impacts of the Proposed Florida Industries Quarry at Six Mile, Vanns Valley, Floyd County Georgia*, http://water.roman.net/kastning1.html.

Kite, J S and Skousen, J G (1999) *Scale and Morphology of Landforms Related to Coal Mining: Trends in the Coal elds of the Appalachian Plateaus*, (modified from a presentation originally given at the Annual Geological Society of America Meeting, 26 October, 1999, Denver, CO), http://www.geo.wvu.edu/• kite/99GSAReclLandform/sld001.htm.

Kivell, P T (1999) Derelict and Vacant Land, in *Applied Geography: Principles and Practice*, ed M Pacione, Routledge, London, 309 320.

Geomorphological Change: Landscape Modification for Recreation

Gerda K Priestley
Universitat Autònoma de Barcelona, Barcelona, Spain

Recreational activities form an integral part of everyday life, especially in advanced, predominantly urban societies. The majority of these activities may be classi ed as sports. Most are urban-based; many occupy small extensions of land; some even take place indoors. The resulting impacts of these activities thus form an integral part of the urban matrix. There are, however, certain outdoor activities, which require the use of large extensions of land, which are normally located in rural areas or on the rural-urban fringe. Some of these outdoor activities generate considerable impacts. Figure 1 provides an overview of the impacts of a number of popular sports, of which golf and skiing are clearly those which generate the most signi cant impacts, although the nature of the corresponding transformations varies considerably.

GOLF

Golf can, in fact, be played on any large open space or field, the only essential requirement being a small hole into which a ball must be dropped. Many 18-hole regulation golf courses require a minimum surface area of 30 ha, and they have been specifically designed for the purpose. There are almost 30 000 regulation golf courses in the world and over 50 million people playing this sport. Golf courses may constitute highly positive elements in the landscape, especially where they fulfill the function of green, open spaces in the suburban expansion of large cities, as is the case in London; or where the presence of water surfaces leads to the enrichment of the flora and fauna.

Nevertheless, attention usually focuses on the negative impacts associated with the construction of golf courses, the most important of which are:

- destruction or modification of the top layers of soil as a result of disturbance during course construction. This can, in some cases, also cause soil erosion;
- excessive demand for water for irrigation when located in regions with a rainfall and humidity deficit;
- loss of water through evaporation from artificial lakes;
- application of chemical fertilizers and other products for the treatment of grass, which can lead to the contamination of aquifers;
- disturbance and modification of adjacent ecosystems, plant communities and fauna.

Obviously the degree of impact varies from one course to another. Two key factors are fundamental in determining such impacts: natural conditions (climate and the physical characteristics of the location of the course) and course design. Golf was originally played in humid regions where

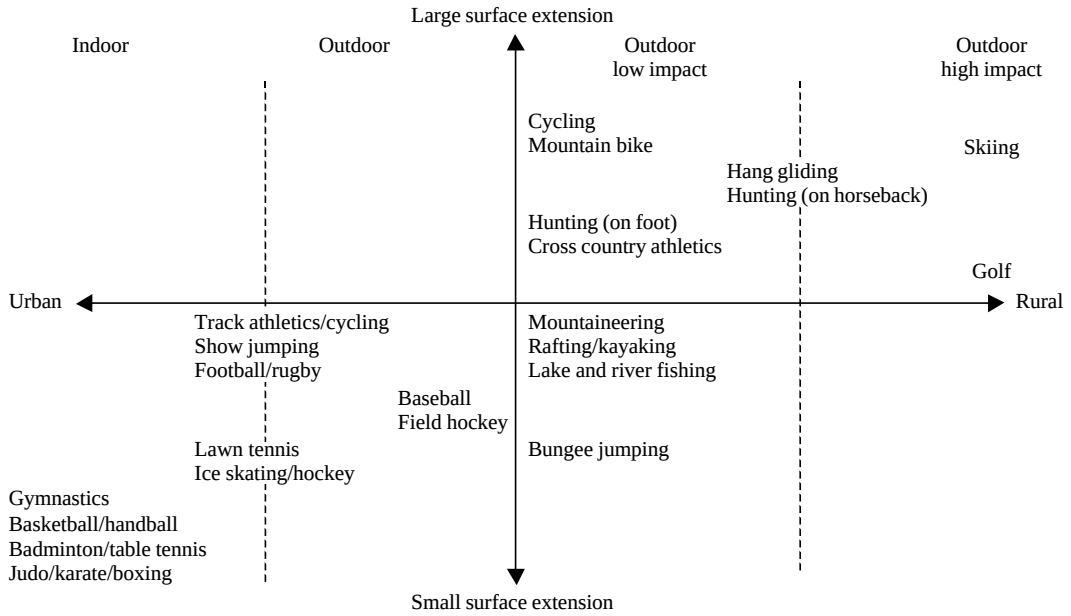


Figure 1 Sports activities: location, extension and impact. Some spectator sports like football produce major secondary impacts, e.g., the provision of huge parking facilities

grass grew naturally. Modern irrigation techniques have permitted its expansion to areas with drier climates, which have gained in popularity mainly as tourism destinations. It is in these regions where the negative impacts of golf courses are potentially most important, as large quantities of chemical products are often required to stimulate the growth of grass, or to counteract the reaction to recycled waste water, which is used for irrigation purposes in the absence of abundant rainfall or fresh water in the aquifers.

The physical characteristics of the location of golf courses (the site chosen) also influence the resulting degree of impact. Measures taken during the construction of a new golf course, such as the remodeling of the landscape, the removal of natural vegetation, or the construction of artificial lakes, can disturb the ecological balance of the environment. Course design is an important factor, not only in determining the impact of these initial measures, but also the recurring impact of subsequent course maintenance. The international trend in course design is towards the emulation of famous North American courses, of which the perfectly landscaped and manicured, world famous Augusta National course in Georgia is without equal. The traditional Scottish natural links (i.e., coastal dunes) courses, such as The Old Course at St. Andrew s, are seldom used as a model for new designs. As a result, most of the courses constructed during the golf boom of the last 20 years include elements that do not exist in the surrounding natural environment. Apart from the basic requirement of a relatively large extension of well cared for grass, the elements most commonly found are artificial lakes and non-native tree and shrub species.

With the prospect of the occurrence of global warming during the 21st century, increasing attention will have to be paid to environmental impacts in the design and management of golf courses. It will be fundamental to adapt course design to the natural environment, and to implement management measures that permit the more efficient use of water, such as intelligent irrigation systems, and the mandatory use of recycled waste water.

SKIING

The impacts of winter sports, basically downhill and cross-country skiing, but also more recent forms, such as snowboarding, can to some extent be compared with those of coastal tourism. Downhill, and to a lesser extent, cross-country, skiing can be considered the counterpart of mass sun/sand/sea tourism on account of its spatial concentration, accompanying construction of hotel and apartment accommodation, and mass but seasonal demand. There are, however, some notable differences between the two forms of tourism: there are fewer areas suitable for winter sports and, as most are located in mountain chains, access is usually more difficult and airports far removed from the resorts. As a result, traffic congestion is common.

The principal permanent impacts of winter sports, which also correspond basically to three different altitude zones, are:

- the presence of access roads, often several lanes wide, totally out of place with respect to the surrounding landscape;
- the accompanying residential and hotel development (in the Alps alone, there are over three million beds available);
- infrastructures associated with the practice of the sport: cable ways, ski lifts and ski jumps.

Other impacts exist that are visible only when snow melts in the warm season and when the tourism demand in many ski resorts remains high as a result of the scenic beauty of such mountainous areas. It is then that the devastating effects of the careless exploitation of natural resources becomes apparent. These include, principally, the clearing of woodland and protective forests; the opening-up of glaciers for winter sports; and land scouring and damage to soils, especially on the lower, beginner ski slopes and access areas to ski lifts. The negative impacts are not limited to such relatively localized elements. Land speculation, as a result of the demand for accommodation at, and even on, ski slopes, has led to the expansion of settlement in ecologically sensitive zones. The accompanying abandonment of agriculture and, consequently, the lack of attention to preserving the landscape, have impaired the natural beauty of the landscape. Furthermore, traditional architectural styles have frequently been replaced by imported ones, requiring damaging construction modes. Impacts are not, however, limited to visual aspects, for the consequences of elemental forces have also intensified. Drainage networks have been disrupted, with the consequent increase in the risk of soil erosion, landslides, floods and avalanches.

As global warming progresses, the future of skiing is somewhat unpredictable, but, certainly, the existing resorts will attempt to maintain demand by applying modern technology to generate snow artificially. New resorts may eventually develop in locations nearer the Poles or at higher elevations, hopefully applying measures to reduce environmetal impacts.

FURTHER READING

Bale, J (1994) *Landscapes of Modern Sport*, Leicester University Press, Leicester.

Hahn, P (1991) Skilauf und Umwelt: Pro und Kontra (Skiing and environment: pro and contra), *Schule Sportstätte*, 4, 7 8.

Standeven, J and De Knop, P (1999) *Sport Tourism*, Human Kinetics, Champaign, IL.

Ward, J, Higson, P, and Campbell, W (1994) *Advanced Leisure and Tourism*, Stanley Thornes, Cheltenham, UK.

Geomorphology

see Geomorphology (**Volume 1**)

GIS (Geographical Information Systems)

see Geographic(al) Information Systems (GIS) (**Volume 4**)

Green Electricity Distribution

see Wind Power Development (**Volume 3**)

Green Revolution

Kenneth A Dahlberg
Western Michigan University, Kalamazoo, MI, USA

The Green Revolution is the name given to a package of fertilizers, pesticides, irrigation, and high-yielding varieties (HYVs) of crop seeds developed in the 1940s speci cally for use outside the temperate zones. The new seed breeding techniques developed in the 1930s for temperate-zone agriculture were used to modify local disease-resistant seed varieties in the developing countries to make them more productive, less sensitive to variations in day length in order to facilitate their wide dissemination throughout different climatic zones, and in the case of rice, to make them earlier maturing to facilitate double cropping. The hope of its promoters, initially led by the Rockefeller Foundation, was that by increasing food production in the developing countries and thereby - it was assumed - alleviating hunger, the Green Revolution would prevent red (i.e., communist) revolutions. By the 1960s, this package was aggressively being spread throughout the developing world by a new network of international research centers and aid programs, with wheat, rice, and maize being the leading crops involved.

A second phase of the Green Revolution emerged in the mid-1980s when new genetic engineering techniques, combined with expanded intellectual property rights (patents) and free trade agreements, gave transnational corporations a much greater ability to shape not only agricultural policies worldwide, but the whole structure of the food system. This power has raised fundamental questions about the ability of democratic societies to set standards for food safety to protect their citizens or to enable them to choose foods produced in environmentally sustainable ways and/or socially just ways which provide farmers and farm workers reasonable incomes and working conditions.

The signi cance of the Green Revolution in terms of global environmental change is that its rst phase extended industrial agriculture in most parts of the world down to the village level, thus dramatically increasing global fossil fuel demands for this highly energy inef cient and environmentally and socially destructive mode of agricultural production. Other global environmental impacts from industrial agriculture include soil erosion, salinization and water logging from irrigation, surface and groundwater contamination from pesticides and fertilizers, and the loss of cultural and biological diversity. The second phase is accelerating these unsustainable trends and is creating a global industrial food system that is highly fossil fuel intensive not only in the production sector, but in processing and distribution - where the average food item is more highly processed, more heavily packaged, and transported increasing distances. By contributing to increasing economic and social inequities at both national and international levels, this system also increases the risks of social instability. The need for more sustainable and socially just food and agricultural systems is clear.

The development of modern industrial agriculture can be traced to an increasing interaction between changes in agriculture and the development of modern science and technology. While the Arabs were the first to practice a science-based agriculture, their knowledge and technologies were ignored or lost as they were driven out of Europe. The values emerging from the Reformation and Renaissance helped lay the foundations for industrial institutions and technologies. The discovery of the New World generated intense interest in new plants and animals (as well as peoples) and the botanic gardens then established reflected an awareness of their strategic economic importance. While there was a very important agricultural revolution which preceded and made possible the industrial revolution, the latter, with its strong urban, industrial, and technological foci, soon led to a general neglect of rural, agricultural, natural, and informal systems. The new and formalized industrial organizations soon became a male preserve, while informal and natural systems became more associated with women and female role models.

It was, however, these natural, agricultural, and informal systems which provided the resources upon which industrialism was built. Over time they have been progressively incorporated into the formal systems of industrialism through the enclosure of land, forests, fisheries, and other common property resources. Typically this was done by placing lower level groups and environments under the

jurisdiction and control of higher level organizations state or corporate.

One of the greatest areas of difference between agriculture and the industrial model lies in the great contextual variability of agriculture where climate, weather, soil, topographic, resource, cultural, social, and institutional variations all profoundly affect its viability as contrasted to industry, where the universalism and rationalism of modern science and technology promote blueprint approaches whereby local variation is fitted to the design, often in a procrustean manner. It is only in the 20th century that there has been the capacity to start systematically applying the industrial model to agriculture.

Thus, while the early practice-based agricultural revolution and the social and political changes associated with the enclosure of the commons had profound societal impacts, they were not science and technology driven. Using modern modes of analysis, such as input/output and rates of growth in productivity, to determine when the most radical changes occurred historically, Grigg (1984) concludes that:

> It seems that European agricultural change was evolutionary rather than revolutionary until this century. It is the period since the 1930s that merits the title [of agricultural revolution].

This modern farming revolution in Europe had parallels in both the Union of Soviet Socialist Republics (USSR) and the United States. In the USSR, the destruction of peasant farming and its replacement with state-owned and managed factory farms offered a new variant on the industrial model of agriculture, one where the standardization of organizational and management practices predominated. In the United States, a standardized production package gradually emerged, based on synthetic fertilizers, pesticides, irrigation, and hybrid seeds. The success of this production package in increasing yields laid the groundwork for the Green Revolution where the underlying idea was that it should be transferred to non-temperate zone countries and colonies.

The Green Revolution has been the subject of much controversy from its beginnings (Karim, 1986). In what can be seen as the first phase of its development roughly from the 1940s to the mid-1980s there were dramatic increases in *per capita* food production in Asia, modest increases in Latin America, but serious declines in Africa. Proponents attributed the gains largely to the Green Revolution, while critics argued that the longer-term environmental, social, and equity costs of these gains were not included in the balance sheet. Some of these concerns were addressed as a network of international agricultural research centers was created. The second phase, which is still ongoing, has involved the increasing incorporation of agriculture into a global food production, processing, and trading system based upon genetic engineering, intellectual property rights, free trade, and the increasing concentration of large corporate power at both national and international levels. This larger global food system is supported by governments and dominated by large transnational corporations.

The first phase of the Green Revolution (1940s mid-1980s) can be understood as the increasingly systematic export of temperate zone industrial agriculture to the developing world which is largely non-temperate zone. Precursors from the 16th century onwards included tropical crops such as sugar cane, indigo, coffee, tea, groundnuts, and rubber, which were developed in the colonies along industrial lines. That is, large-scale production facilities were created called plantations rather than factories often displacing small-scale local subsistence farmers through land expropriation. Labor was also organized in an industrial manner, first with slaves, and later with highly exploited workers. The scale, structure, and management of these plantations all reflected Western industrial approaches and notions of economic rationality. Their direct linkage with the factories and markets of metropolitan countries was a major factor in creating the dual economy of most colonies.

Precursors to the Green Revolution in the white settler colonies involved the immigrants importing their familiar temperate zone crops, animals, and cultivation techniques (Crosby, 1986), plus Western agricultural research approaches (Dahlberg, 1979). Settlers and colonial administrators believed in the superiority of their crops and practices, just as they believed more generally that Western culture, religion, and economic and political institutions were superior, thus justifying the use of the West s genuinely superior military technologies to impose them.

Important as these precursors were as both precedents and indicators of deep-seated Western assumptions about agriculture, their overall impact on the vast majority of rural subsistence peasants, nomads, and villagers was very small. The foundations for the later rapid expansion of the Green Revolution lay more in the creation of centralized colonial infrastructures and administrative control systems which enabled relatively small Westernized enclaves symbolized by Western-style cities to penetrate the colonies and extract their resources.

In the early 1940s, the Rockefeller Foundation, motivated in part by a fear of rising global hunger seen to be caused by rising populations sent a team of plant breeders to Mexico to explore how to increase agricultural production. The Foundation officials and plant breeders involved worked from the widely held assumptions that increased agricultural production would reduce hunger and that technologies are neutral and can be transferred with only minor local adaptations. Because these assumptions serve a number of powerful vested interests, both remain strong even today, in spite of convincing evidence to the contrary (Sen, 1981; Dahlberg, 1989). Critics at the time, notably geographer Carl Sauer, questioned the Foundation s approach, arguing presciently that the production package of HYVs,

synthetic fertilizers, pesticides, and irrigation would favor those already better-off farmers and regions with access to irrigation more than the large majority of dryland peasant farmers located in the more arid regions of Mexico.

Had the Rockefeller Foundation asked what was required to improve the rural welfare of most Mexican peasants, the logical answer would have led them to an interdisciplinary effort to deal with the challenges of dryland farming something the earlier Cardenas government in Mexico (1935 1941) had worked on. Eventually responding in the mid-1960s to criticisms that its hybrid maize varieties were benefiting only better off irrigated areas, the Rockefeller Foundation did create a regional project, the Plan Puebla, in a dryland region. Reviews of this project showed that traditional improved open-pollinated varieties of maize did almost as well under dryland conditions as high-yielding hybrids and had the additional advantage that the farmer could save seed for next year s crop and not have to buy hybrid seed. These studies also showed that in increasing yields, what was most important was to take careful account of a farm s elevation, its soil morphology and moisture, planting dates, and to provide effective credit for small farmers. The clear implication was that the transferable production package was maladaptive and that a farming systems approach to rural development adapted to specific regions was required (Redclift, 1983).

The gradual institutionalization of the Green Revolution was based on the impressive increases in yields the US plant breeders and their Mexican apprentices had achieved. While the original mandate was to explore all major food crops, increasing production proved easiest with wheat, then maize, and more difficult with beans with the former two crops becoming the focus of research. Because a strategy of developing hybrids for wheat was not used, the HYVs developed in the 1950s could be disseminated rapidly throughout the Asian Subcontinent in the mid-1960s when there was a fear of looming malnutrition and starvation (Borgstrom, 1967). The pioneering leader in this wheat research, Norman Borlaug, was awarded the 1970 Nobel Peace Prize for his work (Borlaug, 1997).

With maize, the strategy of focusing on hybrids meant increased production only in those few, already fairly well-off regions where most of the trappings of industrial agriculture were already present, i.e., irrigation, mechanization, distribution networks for seeds, fertilizers, and pesticides, credit, extension services, and markets. As noted above, in dryland conditions, hybrid maize was only slightly more productive and generally more costly. Work on beans again showed the dominance of the production paradigm, where it was not until 1949 that tests were finally made to determine the relative protein and tryptophan content of the improved varieties of this basic dietary staple. While more productive, the improved varieties had less protein and tryptophan. Since then, there have been attempts to breed more usable protein into HYVs but they have not been very successful.

By the 1950s, the Rockefeller program had been expanded to include other Latin American countries and other food crops, especially the potato. Work was also begun in India on maize, sorghum, and millets and in the Philippines on rice. In 1959, the International Rice Research Institute (IRRI) was chartered, with funding coming from both the Rockefeller and Ford Foundations. By 1960, concerns over population growth and possible famines led to a massive Ford Foundation program in India to increase production not only though a package of inputs (HYVs, fertilizers, and pesticides), but accompanying supports (tools, credit, price incentives, and markets). With a 1966 pledge of $1 million, the Rockefeller Foundation backed the transformation of the research institute it had long-supported in Mexico into CIMMYT the Spanish acronym for the International Maize and Wheat Improvement Center. IRRI and CIMMYT became the first two of what has now become an international network of sixteen agricultural research centers known as the *Consultative Group on International Agricultural Research* (CGIAR). The CGIAR is funded by the World Bank and a consortium of other United Nations (UN) agencies, national aid programs, and private foundations. Its budget in 2000 was US$331 million.

The CGIAR has research centers devoted not only to work on various crops (wheat, maize, rice, potatoes), but to livestock, cropping problems in specific climatic zones (tropical, semi-arid, and dry), forestry, agroforestry, living aquatic resources, water management, food policy, and support of national agricultural research programs. The gradual growth of this network, both in scope and diversity, represents an acknowledgment by the international community of the importance of agriculture. It also represents partial acknowledgment by the agricultural establishment of criticisms of the system and the Green Revolution over the years.

These included social, economic, environmental, and gender critiques. A major debate about the nature and direction of rural development was sparked by the development by IRRI of hybrid varieties of rice. The UN Research Institute on Social Development issued a series of reports documenting the Green Revolution s unequal social and economic impacts and their risks. The 1970s energy crisis led to a series of academic studies that highlighted the great energy inefficiencies of industrial agriculture, with the US food system requiring roughly 10 energy calories to deliver 1 food calorie on people s plates (Steinhart and Steinhart, 1974). Environmentalists criticized the high environmental costs of industrial agriculture at home and abroad. Particular emphasis was placed upon the global loss of cultural and biological diversity as well as such local and regional impacts as soil erosion, salinization and water logging from irrigation, surface and groundwater contamination from pesticides and fertilizers, and the loss

of small towns. Beginning in the 1980s, women s groups around the world began an ongoing critique of the huge gender biases found in agricultural research priorities and in the staffing of agricultural ministries, universities, and extension programs worldwide.

In addition to these issues, there were systemic critiques and searches for alternatives. Systemic critiques stressed the ways in which industrial models whether capitalist or socialist focused on production and generally ignored its many social, political, and environmental externalities. The inadequacies of neo-classical economics were shown through assessments of the real-world costs of *industrial agriculture*. These include the health costs to farm workers, farmers, and consumers; the social costs involved in the loss of rural towns and communities, the national and international economic disparities generated through industrial agriculture, the depletion of non-renewable resources, the degradation of renewable resources, and especially the losses of cultural and biological diversity (Dahlberg, 1986; Shiva, 1991).

Others sought out alternatives to industrial agriculture, some existing and some new. Organic farming and biodynamic agriculture, inspired by traditional Asian practices, were brought and adapted to the temperate zones in the 1920s. Newer approaches like *agroecology*, and *regenerative agriculture* and *regenerative food systems* emerged in the environmental and post-oil embargo era, drawing heavily upon ecological and systems theories. Agroecology includes two approaches. One analyzes traditional and indigenous Third World agricultural systems in ecological terms to understand how the demonstrated sustainability of many has been achieved through social, economic, and technological adaptations to the local environment and culture. It focuses primarily on farm and village level systems, but includes a strong concern for land tenure and social justice issues (Altieri, 1995). The other, *permaculture*, is an ecological design approach which applies ecological principles to specific contexts.

Regenerative agriculture and regenerative food systems use an approach based on ecological hierarchy theory. They seek to understand how to regenerate over multiple generations not only local cropping systems and farm families, but rural communities, landscapes, and regions. This approach includes not only production-related matters (e.g., agriculture), but the total food system from production, food processing, preservation, and distribution to preparation, use, recycling, and disposal. Both alternatives also examine the larger setting of agriculture, including traditional issues of social justice, land reform, and rural development as well as newer issues of free trade, aid, technology transfer, corporate power and the increasing disparities between rich and poor countries. Policies needed nationally to spread the more sustainable farming systems developed have also been analyzed (Conway and Barbier, 1990).

Critiques of these and alternative approaches reviewed all agree on one point: that industrial agriculture is unsustainable over the longer term. As the US farm crises of the 1980s persuaded many groups and finally the US agricultural establishment of the need for changes in agriculture, sustainability became the most widely accepted label acknowledging this need. As a not very descriptive term, it, of course, raises its own basic questions: What or who is to be sustained? How? For how long? And at what cost? As indicated, answers from the various alternative approaches include everything from sustaining farm fields, soils, aquifers, and habitats to peasant and farm families and their rural villages and communities to crop germplasm and cultural and biological diversity from the local to the global level. It is also recognized that all this will have to take place within larger frameworks of sustainable societies and a sustainable international system. Throughout, one sees a clear difference between the great emphasis of alternative approaches on context as compared to the standardized blueprint approaches of industrial agriculture.

The second phase of the Green Revolution started emerging in the mid-1980s. International debates over sustainable development, biodiversity protection, and global warming raised questions about the sustainability of industrial society in general, not just industrial agriculture. New technologies and legal and trade doctrines have also shaped this second phase. The development and increasing importance of genetic engineering combined with the ability of corporations to use new intellectual property rights to patent or otherwise own genetically-modified plants, livestock, and other agricultural or food-related organisms has led to a new political economy of food and agriculture. This has been extended to the international arena through new liberalized rules on free trade and their institutionalization in regional trade agreements and the World Trade Organization.

Many key questions have emerged in this dynamically changing setting.

- Do states have sovereignty over the genetic resources found in their country?
- Can they restrict access to or the export of genetic resources?
- Are they entitled to royalties on their genetic resources when corporations genetically modify and then patent them?
- What about the rights of indigenous peoples?
- Can the labeling of genetically modified foods or products be required?
- Can national standards and/or labeling systems be negated by regional and/or international trade agreements?
- Can labeling go beyond the health and safety of a food product to include its place of origin and/or the manner in which it was produced and/or processed (through organic methods; as kosher food; or without the use of child labor).

What appears to be occurring is a transformation of industrial agriculture into a global industrial food system. The main players in the various traditional sectors agriculture, processing, distribution, retailing are beginning to realize this, but are having difficulty in making organizational adjustments.

The response of the agricultural establishment to this emerging global food system has been mixed. It has acknowledged and sought to address many of the specific environmental criticisms emerging from the sustainable development, biodiversity, and global warming debates. As noted, the CGIAR network has broadened its scope. The new President of the Rockefeller Foundation, Gordon Conway, is promoting a doubly green revolution which seeks rural development of the world s poorest regions though sustainable farming systems developed with full farmer participation (Conway, 1998). Another long-term leader argues for an evergreening revolution (Swaminathan, 1996). These responses, however, still focus on agriculture as a largely rural sector and neglect the significant potential of urban agriculture and local food systems (United Nations Development Program, 1996). Beyond that, they don t address the profound organizational and political changes required to achieve their proposed reforms. For example, the development of agricultural biotechnologies still assumed to be neutral has reinforced the functional specialization and production orientation of most existing organizations, making reform even more difficult, especially given the increasing power of private actors with vested interests in current systems and structures.

Particularly crucial is the question of how governments, at whatever level, can regulate and control such developments for the long-term public good. Governments too are structured along functionally specialized lines, which make it difficult for them to address systemic issues. They are hampered in getting advice by similarly fragmented academic traditions. Even in research areas like the environment and global warming, where there has been interdisciplinary work within the natural sciences, there is very little involvement of the social sciences or humanities. Thus, the crucial role of institutions and organizations as well as values and culture in shaping changes in human behavior is neglected (Bennett and Dahlberg, 1990).

Alternative efforts to bridge and integrate these issues include work on ecological economics and sustainable food systems. Ecological economics argues that conventional economics needs to be reconceptualized in terms of the basic principles of thermodynamics and ecology. Sustainable and regenerative food system analysts argue that the fundamental shift in evaluative criteria between industrial and regenerative systems is from production/productivity to health. Food systems thus must be evaluated in terms of their health over multiple generations. The importance of healthy local food systems is stressed since they can not only greatly reduce fossil fuel use, but can be a significant part of sustainable local and/or regional development.

Since agriculture and food systems represent one, if not the major interface between natural and social systems, they are key areas for seeking to reduce the risks of global warming and other types of global change. Major land-use issues, the quality and availability of fresh water, public health, and issues of basic equity and democracy are involved. While there are major conceptual, technological, and organizational barriers to developing sustainable food systems, emerging paradigms and social movements are challenging the traditional industrial paradigm. These need to be translated into political, professional, and personal strategies and actions that will help to overcome and transform the power and inertia of industrial structures and interests. Citizens, professionals, corporations, and governments have to be persuaded that their longer-term interest lies in developing more sustainable systems. They will have to be encouraged to make difficult, but necessary changes in organizational structures and in their ways of life. They must come to realize that just as healthy food is required to nourish the body, healthy food systems are an essential part of sustaining the biosphere.

REFERENCES

Altieri, M A (1995) *Agroecology: The Science of Sustainable Agriculture*, IT Publications, London.

Bennett, J W and Dahlberg, K A (1990) Institutions, Social Organization, and Cultural Values, in *The Earth as Transformed by Human Action: Global and Regional Changes over the Past 300 Years*, ed B L Turner, II, Cambridge University Press, Cambridge, 69 86.

Borgstrom, G (1967) *The Hungry Planet*, Macmillion, New York.

Borlaug, N E (1997) *Norman Borlaug on World Hunger*, Bookservice International, San Diego, CA.

Conway, G R (1998) *The Doubly Green Revolution: Food for All in the 21st Century*, Cornell University Press, Ithaca, New York.

Conway, G R and Barbier, E B (1990) *After the Green Revolution: Sustainable Agriculture for Development*, Earthscan, London.

Crosby, A W (1986) *Ecological Imperialism: The Biological Expansion of Europe: 900- 1900*, Cambridge University Press, New York.

Dahlberg, K A (1979) *Beyond the Green Revolution: The Ecology and Politics of Global Agricultural Development*, Plenum, New York.

Dahlberg, K A, ed (1986) *New Directions for Agriculture and Agricultural Research: Neglected Dimensions and Emerging Alternatives*, Roman and Allanheld, Totowa, NJ.

Dahlberg, K A (1989) The Value Content of Agricultural Technologies and Their Effect on Rural Regions and Farmers, *J. Agric. Ethics*, 2, 87 96.

Grigg, D B (1984) The Agricultural Revolution in Western Europe, in *Understanding Green Revolutions*, eds T P Bayliss-Smith and S Wanmali, Cambridge University Press, Cambridge, 1 17.

Karim, M B (1986) *The Green Revolution: an International Bibliography*, Greenwood, Westport, CT.
Manning, R (2000) *Food's Frontier: The Next Green Revolution*, North Point Press, 225.
Redclift, M L (1983) Production Programs for Small Farmers: Plan Puebla as Myth and Reality, *Econ. Dev. Cult. Change*, 31, 551 570.
Sen, A (1981) *Poverty and Famines: an Essay on Entitlement and Deprivation*, Oxford University Press, New York.
Shiva, V (1991) *The Violence of the Green Revolution: Third World Agriculture, Ecology, and Politics*, Zed, London.
Steinhart, J and Steinhart, C (1974) Energy Use in the United States Food System, *Science*, 184, 307 316.
Swaminathan, M S (1996) *Sustainable Agriculture: Towards an Evergreening Revolution*, Konark, Delhi.
United Nations Development Program (1996) *Urban Agriculture: Food, Jobs and Sustainable Cities*, UNDP, New York.

Greenhouse Food Production (Agriculture Intensive)

Kenneth E Cockshull
Horticulture Research International, Warwick, UK

The best growers of crops in heated greenhouses in northern Europe can produce annual marketable yields of over 600 tons hectare (ha)$^{-1}$ of fresh tomatoes. These extraordinarily high levels of productivity make greenhouse food production one of the most intensive systems in the world. However, these crops are not eaten for either their energy or their protein contents, as water comprises more than 90% of their edible portions. In addition, there is an environmental penalty associated with their production for some 25 TJ (7 million kW h) are required to heat 1 ha of tomato-producing greenhouses for one year in England (Anon, 1999a) and that energy is usually generated from a non-renewable energy source. Furthermore, many of these salad crops could be produced in unheated glasshouses or in the eld without consuming any fuel, but then their season of production in northern Europe would be considerably shorter because of low air temperatures. Producing them in warmer regions could extend their period of availability, as at present, although there is an energy cost associated with transporting the produce to the north.

At present, there is a strong and continuing demand for greenhouse-grown edible products and there are many reasons for this. Most of the products are highly perishable, and freshness is generally perceived as a very desirable quality of such products. In addition, their texture, avor, aroma, taste, color and appearance may contribute to their appeal and some are promoted as possessing health-giving properties because they contain particular vitamins or other compounds, e.g., lycopene, a powerful antioxidant found in tomato (Giovannucci, 1999). A further positive feature for many consumers is that growers in northern Europe and North America can usually give assurances that greenhouse food crops have been produced without the application of any pesticides. Crops grown in modern greenhouses are not now raised in contact with the soil but in water enriched with the correct balance of mineral nutrients (hydroponics) and this too proves to be a good selling feature in North America, though not yet in Europe. Finally, many customers note the absence of surface blemishes on greenhouse-grown products.

The impact of global environmental change on greenhouse food production is unlikely to be very great, at least initially. The principal components of the anticipated environmental change are an increase in the carbon dioxide (CO_2) concentration of the atmosphere accompanied by an increase in average air temperature. Moderate increases in both components are likely to be bene cial for greenhouse production in northern regions. Less fuel will have to be burnt to maintain a given air temperature in heated greenhouses and a higher CO_2 concentration will enable crops to produce higher yields. Provided that the increase in temperature occurs gradually, the greenhouse growers of northern Europe should have time to adapt their patterns of cultivation. Most likely, they will simply adopt the practices already in use in southern Europe and other regions with higher average air temperatures. One indirect effect of environmental change might be that temperatures become too high to sustain crop production for many months of the year and so producers in southern regions would become less competitive. Any reduction in rainfall, however, would have a detrimental effect on all greenhouse growers as it would make irrigation water less available while increasing its cost and even, perhaps, its salt content. Present cultivation systems in greenhouses require a constant supply of water, as they are largely hydroponic, and the salinity of the irrigation water must be kept low otherwise it reduces the yield and quality of the marketable product (Cockshull, 1998).

The primary purpose of a greenhouse is to protect plants from wind and from low air and soil temperatures. However, as green plants need light to survive and to grow, at least part of the greenhouse surface must also be transparent to solar radiation. All the objects within the greenhouse absorb the radiant energy that enters and so their temperature rises but, as the greenhouse shields them from air movement, the resulting increase is much greater than in the field. Energy is slowly lost through the cladding of a

greenhouse by conduction and radiation to the surrounding air. One of the best cladding materials for greenhouses is glass, not only because it is transparent to light and impervious to rain, hail, and snow, but also because it is opaque to long-wave, infra-red radiation. Such radiation, with wavelengths longer than 2.5 μm, is radiated from the warm objects within a greenhouse. The glass cladding then absorbs this radiant energy and, because it cannot escape, the temperature of the air and of the plants within the greenhouse is sustained (Bot and van de Braak, 1995).

Most plants grow and develop to maturity faster when either the soil or the air temperature is raised within the range from 0 25 °C, and so crops can be grown to maturity earlier in a greenhouse than in the field. However, the advances in maturity are not substantial if only solar energy is used to heat the greenhouse because energy is then provided only by day. Furthermore, there is a risk that the crops will die if air temperatures are allowed to fall below 0 °C. These limitations can be overcome by burning a fuel to provide additional energy when required. The most commonly used fuels are non-renewable energy sources, such as kerosene, oil, coal, and natural gas and they are burnt either in hot-air heaters or in a boiler to produce either hot water or steam (van de Braak, 1995).

Plants grow by fixing carbon dioxide (CO_2) from the atmosphere into carbohydrates using radiant energy in the 400 700 nm waveband. The process is called photosynthesis and the waveband is called photosynthetically-active radiation (PAR) (*see* PAR (Photosynthetically Active Radiation), Volume 2). This waveband accounts for about one-half of the radiant energy reaching the earth s surface from the sun. The rate of photosynthesis and, thus, the rate of crop growth is largely determined by the amount of PAR that is absorbed by green leaves (Monteith, 1977). As CO_2 is also required for photosynthesis, the gas is added to the greenhouse atmosphere, either from tanks of the pure liquefied gas or from the combustion of a hydrocarbon fuel. For maximum productivity, the CO_2 concentration should be controlled at about 1000 parts per million volume by (vpm), air temperatures should be kept near the optimum for growth, the greenhouse should transmit as much PAR as possible and as much as possible of the transmitted PAR should be absorbed by the crop s leaves. With tomatoes, fruit yield is directly related to the amount of solar radiation transmitted by the greenhouse (Cockshull *et al.*, 1992) and so it is very important that the external surface of a greenhouse be kept clean. In the quest for maximal light transmission, glass panes as large as 2•0 • 1•25 m are supported on thin extruded aluminum glazing bars that combine strength with lightness and so require relatively little additional structural support. The outcome is that greenhouses with a single glass roof covering up to 10 hectare (ha) (Figure 1), and transmitting 70% of the solar radiation incident upon them, can be constructed (Waaijenberg, 1995).

Figure 1 The greenhouse at Runcton, England, that was built by Dalsem Horticultural Projects B.V. for Van Heyningen Bros. The new greenhouse covers more than 9 ha. (Reproduced by permission of Van Heyningen Bros)

In the south of Europe, film plastic is often the preferred cladding material and in the USA, roofs of inflated, double polythene film are preferred for their simplicity, low cost, high light transmission and high thermal resistance (Ball, 1991; Khosla, 1998).

Within the greenhouse, the transmitted PAR should be absorbed by those leaves nearing full expansion, as they are the most photosynthetically efficient. To maximize absorption, it is best to orient the crop rows east to west as then the incoming solar radiation normally has to pass through foliage and it is only in the early morning and late evening that some reaches the ground without doing so. Yield of tomatoes also increases if the crop density is increased during the first half of the year, in line with the seasonal increases in the amount of solar radiation incident upon the crop (Cockshull and Ho, 1995), presumably because the absorption of solar radiation is increased.

The temperature lift produced by solar radiation can be so great in summer that the resulting air and plant temperatures become supra optimal for crop growth. Furthermore, the water vapor pressure deficit in the greenhouse atmosphere may become so large in summer that plants close their stomata in order not to lose water too rapidly but nevertheless they may wilt. These problems can be lessened in various ways. In Europe, motorized ventilators are installed along the ridge of the roof and are opened when the temperature within the greenhouse becomes too high. In North America, cool air is drawn into greenhouses by electric fans mounted in the greenhouse walls, and drawing the air over a wet pad can further enhance the cooling effect. Alternatively, much of the radiant energy can be stopped from entering the greenhouse by shading the outside of the southern side of greenhouse roofs in summer. Growers also frequently draw a shade over the crop. This is less effective but has the advantage that the shade can readily be

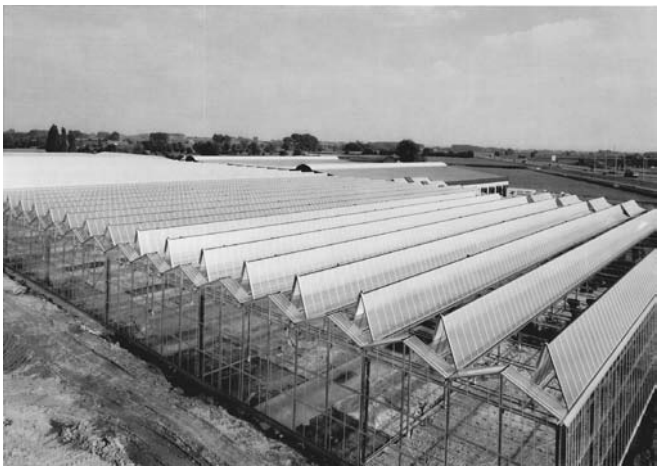


Figure 2 The Cabrio open-roof greenhouse designed by Deforche N.V

withdrawn again. Both of the last two solutions also reduce the amount of PAR that reaches the crop and so will probably reduce yield. One very practical solution is to stop growing crops in greenhouses in the hottest months and restrict their growing season to the cooler months. Other solutions are to design and construct greenhouses in which most of the roof can either be opened or be removed in summer (Figure 2). Film plastics might also be suitably modified by incorporating compounds that transmit less radiant energy in the infra-red waveband without greatly influencing the amount of PAR entering the greenhouse (e.g., Pearson and Wheldon, 1998).

The concept of forcing plants into early maturity appears to have been known since Roman times (Woods and Warren, 1988), but permanent, heated structures for protecting plants were not used in Europe until the 17th century. At that time, it was too expensive to glaze other than small areas of the walls but nevertheless, oranges could be grown in cool climates. By the middle of the 19th century, technology had advanced sufficiently to enable greenhouses to be built with glass roofs and it was then possible to produce pineapple, banana, grapes and other fruits in them. These fruits were produced only for the owner s table and it was not until the end of the 19th century that greenhouses were used to produce food for sale.

Today, the main food crops grown in greenhouses are salad crops, which are of high value (Table 1). Marketable tomato fruit are produced in heated greenhouses from the end of February until the end of October in northern Europe. The crop is usually grown at a daily average temperature in the region of 18 °C in England. This means that about 25 TJ (7 million kW h) of energy must be provided from non-renewable energy sources per hectare of modern greenhouses per year. This represents the combustion of 505 000 liters of heavy fuel oil per hectare. In addition, about 35 TJ of solar radiation are incident per hectare

Table 1 Data from Anon (1999b), © Crown copyright. (Reproduced by permission of the Controller of HMSO)

Greenhouse food crops in the UK in 1998	Planted area (ha)	Yield (tons ha$^{\bullet 1}$)	Value (£ ton$^{\bullet 1}$)
Tomato	263	407	607
Cucumber	183	459	459
Lettuce	596	35	1150
Celery	71	73	865
Sweet pepper	39	181	1006

in an average growing season, i.e., more than the energy input from fuel. With this combined energy input, the best growers can produce over 600 tons of marketable fruit per year (Anon, 1999a) although the average productivity in heated greenhouses in the UK is about 410 tons ha$^{\bullet 1}$ (Anon, 1999b). As the dry matter content of ripe tomatoes is normally less than 6% and its energy content is no more than 75 kJ 100 kg$^{\bullet 1}$, tomato is evidently not a staple foodstuff.

At present, with the exception of lettuce, it is uneconomic to grow these salad crops year-round in greenhouses in northern Europe. However, there is a year-round demand, and the demand in winter is met largely by growers in southern Europe who use both heated and unheated glass and film plastic greenhouses to produce salad crops for export. The heating requirement in the south is much less because of the higher ambient air temperatures. In southern Europe, for example, annual yields from tomatoes grown in heated greenhouses can be as great as those in northern Europe but the energy required is less than 15 TJ ha$^{\bullet 1}$. Energy is then required to transport the produce to the north but the energy cost of transporting 600 tonnes of tomatoes from southern Europe to England by lorry is estimated to be no more than 1.6 TJ. Consequently, producing tomatoes in the south and marketing them in the north requires only 60% of the non-renewable energy cost of producing them in England.

In North America and northern Europe, greenhouse-grown food crops are no longer generally grown in soil but in hydroponic systems using water enriched with appropriate nutrients (Jensen, 1998; Khosla, 1998). This nutrient-enriched water may either be continually recirculated down shallow gulleys containing the plants or it may be applied to inert, rockwool slabs in which the young plants have been established. Computers can now control the composition of the nutrient solution, the frequency of its application and its temperature. Indeed, sophisticated computer programs can now control most aspects of the root and aerial environments in which greenhouse crops are produced, including the CO_2 concentration of the atmosphere. The main advantage is that the grower then has greater control over the growth and development of the crop.

Growers in northern Europe and North America also have a considerable measure of control over the pests and diseases that could infest or infect the crop. Pests are controlled by the introduction of their natural parasites and predators, bred for this purpose by specialist companies. In North America, where fan ventilation is widely used, small insects can also be stopped from entering the greenhouse by fitting filters to the air inlets. As a result, most northern growers can give assurances that no synthetic chemicals have been used to control either pests or diseases during the course of crop production.

In respect of future global environmental change, any increase in the average concentration of CO_2 in the atmosphere will benefit the producers of greenhouse crops, as they will have to provide less CO_2 from other sources. Also, provided that increases in average air temperature continue to occur gradually and the average temperature rise does not exceed 2 °C by AD 2050 in England (Hulme and Jenkins, 1998), northern growers will require less fuel for heating greenhouses. If it does become necessary to change their present patterns of cultivation, they will presumably simply adopt those that are currently employed in warmer regions. A temperature rise will be less beneficial to growers in the warmer regions, as temperatures in their greenhouses may become too high for the production of conventional greenhouse crops. The most serious impact of potential global environmental change would be if it caused any reduction in the availability of water or any increase in either the cost or the salt content of irrigation water. Present cultivation systems for the production of greenhouse foods rely upon a constant supply of water and any increase in the salinity of that water will lead to a reduction in the yield and quality of greenhouse produce (Cockshull, 1998).

Growers are concerned about using non-renewable energy sources to heat greenhouses and many are already adopting measures that use energy more efficiently. One such measure is the cogeneration of electricity and heat in combined heat and power units that use energy more efficiently than in conventional boilers. Additional advantages are that growers can export most of the electricity that is generated, and the CO_2 that is produced as a by-product of combustion of the fuel can be introduced to the greenhouse atmosphere so that the crop synthesizes more carbohydrates per unit of radiant energy. The CO_2 produced in this way costs very little, although the flue gases must first be scrubbed to eliminate pollutants and then cooled before they can be added to the greenhouse.

Governments now propose taxes to penalize the use of non-renewable fuels and the emission of CO_2 to the atmosphere and such measures will have a negative impact on the greenhouse food industry. Considerable work on energy saving in greenhouses has already been done as a result of the sharp rise in energy prices that occurred in 1973. Few of the techniques that were tested then are still in use but the more successful ones could always be revived. The most successful technique was probably the deployment of a sheet of insulating material (a thermal screen or blanket) between the crop and the greenhouse cladding at night. The sheet was rolled up by day but it still reduced light transmission and so produced a corresponding loss of yield. It is not used today because the value of the lost yield now far exceeds the value of the energy that is saved. Double-glazing causes an even greater loss of light (Bailey and Hunter, 1988) and so the cost of fuel has to be very high to obtain any economic gain. Sheard (1979) suggested that greenhouses should use the reject heat from industrial processes and it is also feasible to heat greenhouses using geothermal energy or even renewable energy sources, such as wood or biomass. As yet, however, although there are examples of installations using these techniques, none has been widely adopted.

For the future, it seems likely that the present salad crops will remain popular for many years and that they will continue to be grown in heated greenhouses in northern Europe. However, the food industry is continually seeking novel crops to add to the present range. In the more distant future, functional foods may either be bred or be genetically engineered to contain higher concentrations of vitamins, natural anti-cancer compounds, and even antibiotics, as well as traditional herbal remedies. Provided the products are of high value, are safe, and gain the approval of the general public, they may become the greenhouse crops of the future.

REFERENCES

Anon (1999a) Breaking the 60 kg Target, *Grower*, 132(22), 13.

Anon (1999b) *Basic Horticultural Statistics for the United Kingdom, 1988- 1998*, Ministry of Agriculture, Fisheries and Food, London.

Bailey, B J and Hunter, A (1988) Plant Response and Energy use in Five High Thermal Resistance Greenhouses, *Acta Hortic.*, 229, 165 171.

Ball, V (1991) Greenhouses for the 1990s, in *The Ball Red Book*, ed V Ball, Geo. J. Ball Publishing, West Chicago, IL, 15th Edition, 3 18.

Bot, G P A and Van de Braak, N J (1995) Energy Balance, in *Greenhouse Climate Control*, eds J C Bakker, G P A Bot, H Challa, and N J Van de Braak, Wageningen pers, Wageningen, 135 141.

Cockshull, K E (1998) Plant Responses and Adaptation to Water Issues in the Greenhouse Environment, *Acta Hortic.*, 458, 201 206.

Cockshull, K E, Graves, C J, and Cave, C R J (1992) The Influence of Shading on Yield of Glasshouse Tomatoes, *J. Hort. Sci.*, 67, 11 24.

Cockshull, K E and Ho, L C (1995) Regulation of Tomato Fruit Size by Plant Density and Truss Thinning, *J. Hort. Sci.*, 70, 395 407.

Giovannucci, E (1999) Tomatoes, Tomato-based Products, Lycopene, and Cancer: Review of the Epidemiologic Literature, *J. Natl. Cancer Inst.*, 91(4), 317 331.

Hulme, M and Jenkins, G J (1998) *Climate Change Scenarios for the UK: Scienti c Report*, UKCIP Technical Report No 1, Climate Research Unit, Norwich, 1 80.

Jensen, M H (1998) Hydroponics Worldwide, *Acta Hortic.*, 481, Vol 2, 719 729.

Khosla, S (1998) Status of the Greenhouse Vegetable Industry and Hydroponics in Ontario, Canada, *Acta Hortic.*, 481, Vol 2, 739 742.

Monteith, J L (1977) Climate and the Efficiency of Crop Production in Britain, *Philos. Trans. R. Soc. Lond. B.*, 281, 277 284.

Pearson, S and Wheldon, A (1998) Filtering Out P&D, *Grower*, 7th January 1999, 22 24.

Sheard, G F (1979) Energy and the Future of the Glasshouse Industry, *Rep. Glasshouse Crops Res. Inst.*, 1978, 154 168.

Van de Braak, N J (1995) Heating Equipment, in *Greenhouse Climate Control*, eds J C Bakker, G P A Bot, H Challa, and N J Van de Braak, Wageningen pers, Wageningen, 171 178.

Waaijenberg, D (1995) Construction, in *Greenhouse Climate Control*, eds J C Bakker, G P A Bot, H Challa, and N J Van de Braak, Wageningen pers, Wageningen, 162 170.

Woods, M and Warren, A S (1988) *Glass Houses*, Aurum Press, London.

Greening of Cities

Brad Bass
University of Toronto, Toronto, Canada

The greening of cities has been receiving increased attention in recent years, primarily due to its bene cial effect on modifying microclimates and the magnitude and spatial patterning of the urban heat island as well as reducing energy consumption, emissions and waste gases. Properly sited shade trees can directly reduce residential cooling loads by up to 30% (McPherson et al., *1989). An increased forest canopy in a city also means that some incoming solar radiation is used for evapotranspiration, while non-vegetated surfaces absorb the incoming solar radiation and reradiate it as heat. This heat arti cially raises urban temperatures, creating the urban heat island phenomenon (*see *Urban Heat Island, Volume 3). Reducing the urban heat island by 1 °C can result in decreases of 5% in the total electric load during the summer (Sailor, 1995). Even without reductions in the urban heat island, evaporative cooling has been estimated to reduce the cooling load of some buildings by 20- 30%. During the winter months, a row of trees (a shelter belt) along the wall of a building can reduce the heating load of a building (Minke and Witter, 1982).*

OTHER BENEFITS OF THE URBAN FOREST

However, in addition to reducing energy consumption, trees confer other benefits on a city. These include carbon sequestration, improved air quality, stormwater attenuation, noise reduction, provision of wildlife habitat, increased property value and aesthetics, food production and psychological well-being. For example, on average, over a 40-year period, each tree could sequester 3 tons of carbon (C). Additional tree cover could reduce urban stormwater runoff by 4 18% (Sanders, 1986), and can improve water quality as the pollutants end up being bound in the substrate of vegetation instead of being discharged into the environment.

Higher summer temperatures accelerate the formation of ozone in urban areas. Thus reducing the urban heat island through urban forestry will reduce the severity of smog episodes. In Southern California, simulation models have suggested that reducing the urban heat island by 2 4 °C would be equivalent to converting half of the motor vehicles to zero-emission electric engines (Taha *et al.*, 1997). Reductions in summer energy consumption directly due to shading or evaporative cooling, or indirectly through lowering the urban heat island intensity, would also reduce emissions of greenhouse gases and other compounds associated with the combustion of fossil fuels.

Trees are also effective in filtering gaseous pollution and particulates. Sublethal concentrations of some gases such as sulfur dioxide (SO_2), nitrogen oxide (N_2O), hydrogen fluoride (HF) and ozone (O_3) are neutralized through plant metabolism (Bernatzky, 1978, 1984) and through oxygenation (Grey and Deneke, 1978). Filtering of particulates occurs through sedimentation and active absorption by leaves, branches and stems. Diverting airflow prevents sediment transport, which facilitates the deposition of airborne particles. Particles tend to get trapped in the fine hairs on vegetation and are deposited into the soil or substrate when it rains. In Frankfurt, Germany, it was found that streets with trees have only 10 15% of the number of dust particles found on similar streets without trees (Minke and Witter, 1982).

These benefits are directly or indirectly related to the leaf area density of the urban forest, i.e., the amount of leaf area exposed per unit area of city. The leaf area increases exponentially with the volume of the tree crown, as do the benefits. In order for the crown to reach its full potential, sufficient space must be available for tree growth. In many commercial and industrial areas of the city, this space is not available due to the amount of land that is occupied by hard surfaces and buildings (Figure 1). Even areas that are characterized by low-density residential housing, where large amounts of land may only be occupied by turf grass, tree growth is restricted due to buffers needed to protect sidewalks and other infrastructure.

Figure 1 Restrictions on urban forestry in downtown Vancover, Canada

ROOFTOP GARDENS

One way to increase vegetation cover in urban areas is to use the roofs and walls of buildings (Figures 2 and 3). Rooftop gardens, or green roofs, provide many of the same benefits, and they take advantage of the unused roof space available in most urban areas. In addition, bare rooftops exacerbate the problems associated with removing vegetation from urban areas. For example, the heat from bare rooftops and walls can increase the urban heat island (Terjung and O Rourke, 1981a,b) (*see* Urban Heat Island, Volume 3). In the summer, a typical insulated, gravel-covered rooftop temperature can reach 60 °C. This represents a significant contribution to the urban heat island as well as the cooling load of a building. A rooftop garden can reduce these temperatures through evapotranspiration and insulation.

Based on comparisons of other vegetated and non-vegetated surfaces, evapotranspiration could maintain a summer rooftop temperature between 25 and 35 °C (Luvall and Holbo, 1989). On a large industrial building, evaporative cooling can reduce the cooling load by 20 25% (Abernathy, 1988). Insulation would reduce the variation around the mean temperature. The combined effect could be a reduction of indoor temperatures by 4 °C compared to a case without a green roof (Liesecke *et al.*, 1989). Depending on the actual vegetation, there may be additional reductions in temperature due to shading from the leaves.

During the winter, the rooftop garden would provide additional insulation on the roof. One typical type of rooftop garden, consisting of 20 cm of substrate and 20 40 cm of grass would provide the equivalent of 15 cm of rockwool. This is equivalent to an insulation factor of R20 (Minke and Witter, 1982).

A rooftop garden contributes to improved air quality in several ways. Reducing the urban heat island will decrease the severity of smog. Rooftop vegetation would also reduce the transport of particulate matter from roofs to other parts of the city. As with trees, rooftop vegetation can filter and absorb pollutants. A 2000 m^2 rooftop garden, consisting of unmowed grass, could remove 4000 kg of airborne dirt per year if the lower portion of the canopy was in contact with the moving air (Minke and Witter, 1982). Usually this is not the case, but even if this figure were reduced by 90%, each square meter of grass would still remove 0.2 kg of dirt per year. Broadmeadow and Freer-Smith (1996) summarized the results of several studies comparing the capacity of different plants to absorb nitrogen dioxide (NO_2), ammonia (NH_4) and ozone. Except for maple, birch and pine trees, all other plants may be grown in either a vertical or rooftop garden, including beans, sunflowers, petunias, tomatoes, oats and rye grass.

A large component of urban water resource management is devoted to managing stormwater runoff. Since large parts of the city are now impermeable, water has to be diverted through artificial systems, or it runs off over the surface. Stormwater runoff has contributed to water quantity and quality problems and during extreme precipitation events, it can lead to flooding and erosion. Stormwater can also be managed through storage, infiltration and retention. Alternatives to hard infrastructure include downspout disconnection, rain barrels, cisterns, swales adjacent to parking lots, retention ponds, artificial wetlands and living machines, and the use of porous materials for roads, driveways and parking

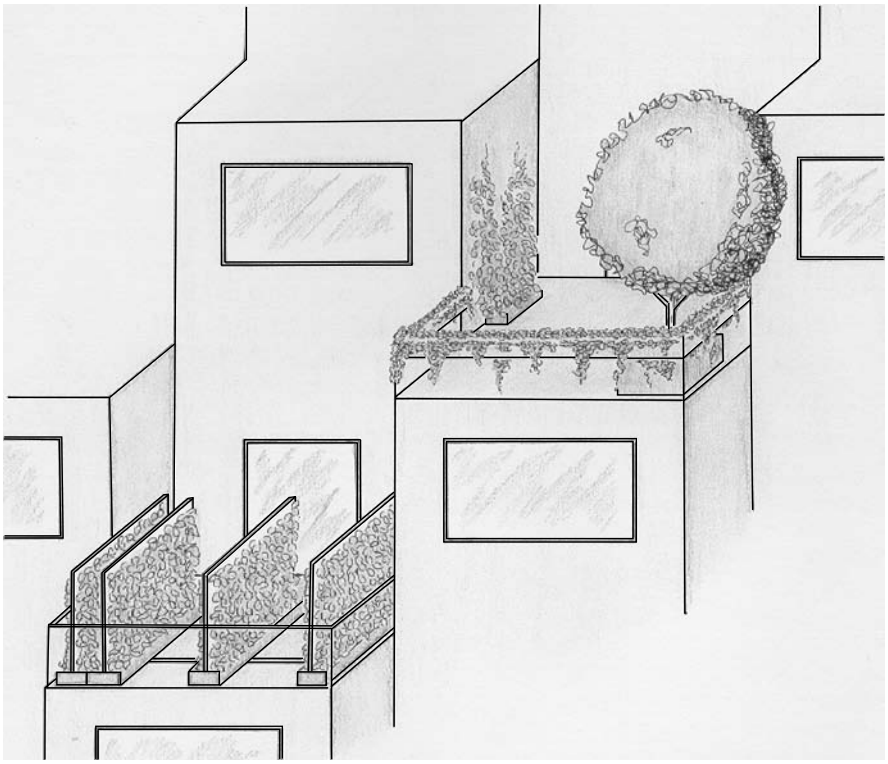


Figure 2 Rooftop garden design for high-density residential areas (J Flemming, Division of the Environment, University of Toronto)

Figure 3 Using balconies to create vertical gardens, Frieburg, Germany (A Kenney, Faculty of Forestry, University of Toronto)

lots as well as urban forestry, rooftop gardens and vertical gardens.

On a rooftop garden, nearly all rainwater from small rain showers is stored in the substrate and returned to the atmosphere by evapotranspiration. Rainwater that is not retained is released at a much slower rate and reaches storm drains after the peak flow of runoff from paved areas. In Portland, Oregon a rooftop garden with a 7 cm vegetation layer produced no runoff during a three month summer period although the retention was not as high during a continuous, heavy rainfall (Thompson, 1998). Rooftop gardens can also be used to improve water quality. Heavy metals can be bound in the substrate, as much as 95% of cadmium (Cd), copper (Cu), lead (Pb) and 16% of zinc (Zn) (Johnston and Newton, 1996). In Europe rooftop gardens (Thompson, 1998) are being designed to treat wastewater that would normally be discharged as raw sewage.

Thus even if sufficient space for trees is available, rooftop gardens would still be an important component in reducing stormwater runoff, the urban heat island and air pollution. Rooftop gardens also provide a means for recreating habitats to restore biodiversity in urban areas. Rooftop environments are similar to arid or semi-arid environments such as might be found in a prairie landscape. In the US, rooftop gardens have been used to recreate prairie habitats (North American Wetland Engineering, 1998) and in Canada, the rocky alvars of Manitoulin Island (Reid, 1998).

In Europe, there are two examples where rooftop gardens have been used as part of a wildlife corridor through the city. The first was designed for birds, insects and airborne seeds, and the second design created habitats that are analogous to islands to isolate plants whose seeds are not spread by air (Peck *et al.*, 1999). Butterflies, bees and birds have all been found in gardens above 19 stories (Johnston and Newton, 1996).

There are two basic types of rooftop gardens or green roof systems extensive and intensive. Extensive green roofs are characterized by their low weight, low capital cost and low maintenance. Typical extensive gardens consist of various grasses and wildflowers, a lightweight growing medium, a membrane to repel roots and moisture from the roof and a drainage layer to move moisture off the roof. Intensive green roofs are usually designed to resemble gardens, and may consist of various structures and a variety of plants including trees. Whereas extensive rooftop gardens are often inaccessible to the public, intensive rooftop gardens are designed to be accessible and are characterized by their increased weight, capital cost and maintenance requirements. One false preconception is that the space for rooftop gardening is limited to only flat roofs. Sloping roofs can also accommodate vegetation, and some rooftop gardens, such as hydroponic systems, may even be better suited to such roofs as gravity can be used to help circulate the water.

VERTICAL GARDENS

Walls provide an even greater amount of space for vegetation in the form of vertical gardens. Such gardens, or green walls, can comprise vines growing adjacent to a wall, window shades that utilize plants, or vertical hydroponic systems. Even in areas that are suitable for trees, they can be used to cover windows that cannot be shaded by trees due to the height or specific architectural features of a building.

Observations indicate that vertical gardens also reduce the heat flow into the building, and their surface temperature is lower than that of a bare wall, thus reducing the urban heat island. A series of experiments in Japan suggested that vines could reduce the temperature of a veranda with a southwestern exposure (Hoyano, 1988). In Germany, the vertical garden surface temperature was 10 °C cooler than a bare wall when observed at 1:30 pm in September (Wilmers, 1989). For a building in South Africa, consisting of two 10 mm fiber-cement sheets with 38 mm of fiberglass insulation, a computer simulation estimated that a vertical garden would reduce summer daytime temperatures on the surface by 5 °C (Holm, 1989).

In some cases, vertical gardens might be the preferred alternative. For example, on a multi-story building with a lot of exposure to afternoon sun during the summer, shading the windows with vertical gardens will reduce the overall cooling load significantly more than will a rooftop garden (Figure 4). However, the widespread adoption of green window shades will require additional research to devise structures to support the weight and to cope with the higher wind speeds that occur at higher elevations on multistory buildings.

CONCLUDING REMARKS

Rooftop and vertical gardens confer other benefits such as extending the life of the building envelope, spurring new economic development, and providing space for horticultural therapy and for urban agriculture. Nevertheless, neither of these technologies has been widely adopted in many countries, despite the historical precedents for doing so. Rooftop and vertical gardens can be traced back to the hanging gardens of Babylon and were known to exist in the Roman Empire, both as a response to urban population densities and to beautify institutional buildings, such as mausoleums. In Pompeii, shopkeepers grew vines on their balconies, and the ancient historian Pliny wrote about trees being imported for rooftops (Farrar, 1996). The Vikings layered the walls and roofs of their homes with turf to protect against wind and rain, and sometimes used seaweed to insulate roofs (Donnelly, 1992). During the Renaissance, steeply terraced gardens and rooftop gardens were common in the city of Genoa (Gorse, 1983).

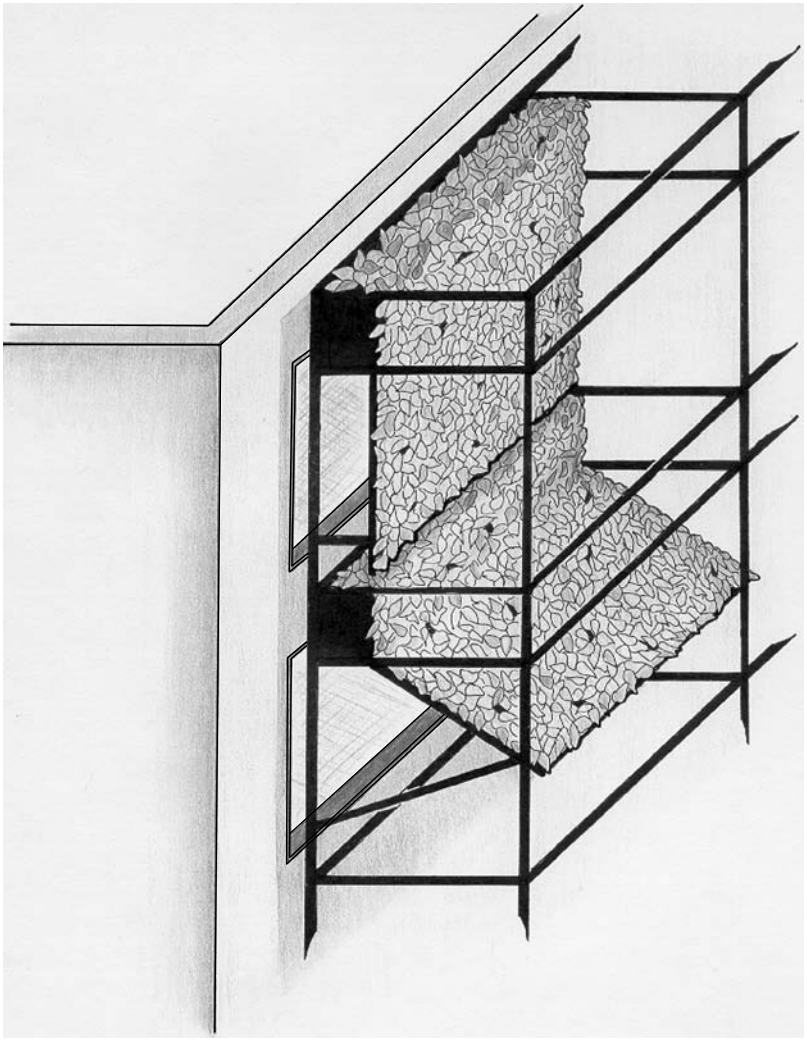


Figure 4 Vertical garden as a window shade (J Flemming, Division of the Environment, University of Toronto)

Vertical gardens in the form of hanging gardens also existed in pre-Columbian Mexico (Goode, 1986), India, and in some of the Spanish homes of 16th 17th century Mexico (Flower, 1937). In Russia, hanging gardens were favored in the 17th century Kremlin (Titova, 1990). In 18th century France, vertical gardens were constructed for the sake of aesthetics (De Lorme, 1996), and Le Corbusier and Frank Lloyd Wright, made extensive use of rooftop gardens in their designs (Curtis, 1986; Hoffman, 1995).

Recent advances in rooftop gardens have led to a technology that consists of far more than plants and soils. Proponents of rooftop gardens now advocate the use of the term *green roof infrastructure* to refer to the combination of plants, substrate, membranes and drainage layers, the latter two often being what separates one provider from another (Figures 5 and 6). Vertical gardens are still most commonly viewed as vines growing on a wall, perhaps because a clearly identified industrial sector has not yet emerged to develop and market this technology as a green

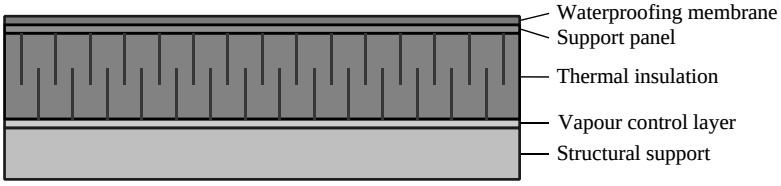


Figure 5 Conventional rooftop infrastructure (K Liu, National Research Council, Canada)

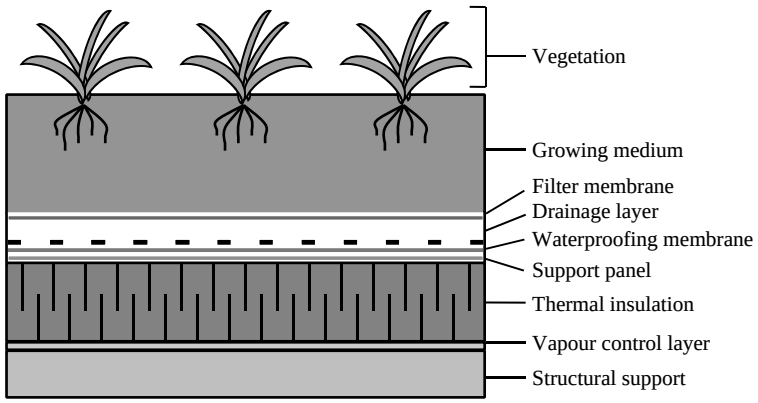


Figure 6 Green roof infrastructure (K Liu, National Research Council, Canada)

wall infrastructure. Nevertheless, green roof infrastructure and vertical gardens provide the capacity to extend the benefits of the urban forest into areas of the city where leaf area density is restricted by land use or zoning restrictions. Even where urban forestry can be expanded, it is still important to be able to mitigate the impacts of bare roofs and walls on the urban environment. By the year 2005, over 50% of the world s population will live in cities (Bindé, 1998) (*see* Urban Population Change, Volume 3). Thus, these new technologies will provide the basis to use vegetation as part of a strategy to manage the increased environmental stresses that will accompany the growing urban populations.

REFERENCES

Abernathy, D (1988) Roof Spray Cooling Systems, *Plant Eng.*, 42, 178 180.

Bernatzky, A (1978) *Tree Ecology and Preservation*, ed M Treshow, Elsevier Scientific, Amsterdam.

Bernatzky, A (1984) *Air Pollution and Plant Life*, John Wiley & Sons, NY.

Bindé, J (1998) Cities and Environment in the 21st Century, *Futures*, Vol. 30, 499 518.

Broadmeadow, M S J and Freer-Smith, P H (1996) *Urban Woodland and the Bene ts for Local Air Quality*, The Copyright Unit, Her Majesty s Stationary Office, Norwich.

Curtis, W J R (1986) *Le Corbusier: Ideas and Forms*, Phaidon Press, Oxford.

De Lorme, E (1996) *Garden Pavilions and the 18th Century French Court*, Antique Collector s Club, Suffolk.

Donnelly, M (1992) *Architecture in the Scandinavian Countries*, The MIT Press, Cambridge, MA.

Farrar, L (1996) Gardens of Italy and the Western Provinces of the Roman Empire from the 4th Century BC to the 4th Century AD, BAR International Series No. 650, Oxford.

Flower, W P (1937) Mexican Motif: Casa Alvarado, *Bull. Pan Am. Union*, 71, 316 324.

Goode, P (1986) Mexico, in *The Oxford Companion to Gardens*, eds G Jellicoe and S Jellicoe, Oxford University Press, Oxford.

Gorse, G L (1983) Genoese Renaissance Villas: a Typological Introduction, *J. Garden Hist.*, 3, 255 289.

Grey, G W and Deneke, F E J (1978) *Urban Forestry*, John Wiley & Sons, NY.

Hoffman, D (1995) *Understanding Frank Lloyd Wright's Architecture*, Dover Publications, NY.

Holm, D (1989) Thermal Improvement by Means of Leaf Cover on External Walls a Simulation Model, *Energy Build.*, 14, 19 30.

Hoyano, A (1988) Climatological uses of Plants for Solar Control on the Effects on the Thermal Environment of a Building, *Energy Build.*, 11, 181 189.

Johnston, J and Newton, J (1996) *Building Green, a Guide for using Plants on Roofs, Walls and Pavement*, The London Ecology Unit, London.

Liesecke, H J, Krupka, B, and Brueggemann, H (1989) *Grundlagen der Dachbegruenung, Zur Planung, Ausfuhrung und Unterhaltung von Extensivbegruenungen und Einfachen Intensivbegruenungen, Patzer Verlag*, Berlin-Hannover.

Luvall, J C and Holbo, H L (1989) Measurements of Short Term Thermal Responses of Coniferous Forest Canopies Using Thermal Scanner Data, *Remote Sens. Environ.*, 27, 1 10.

McPherson, E G, Simpson, J R, and Livingston, M (1989) Effects of Three Landscape Treatments on Residential Energy and Water Use in Tuscon, Arizona, *Energy Build.*, 13, 127 138.

Minke, G and Witter, G (1982) *Haeuser Mit Gruenem Pelz, Ein Handbuch Zur Hausbegruenung*, Verlag Deiter Fricke GmbH, Frankfurt.

North American Wetland Engineering (1998) Ecology, Construction and Benefits, in *ReNature/NAWE Green Roof Product Literature*, MN, May 18.

Peck, S, Callahan, C, Bass, B, and Kuhn, M (1999) *Greenbacks from Green Roofs: Forging a New Industry in Canada. Status Report on Bene ts, Barriers and Opportunities to Green Roof and Vertical Garden Technology Diffusion*, Canada Mortgage and Housing Corporation, Ottawa, Canada.

Reid, R (1998) Habitat for the Hardy, *Seasons*, 4, 14 22.

Sanders, R A (1986) Urban Vegetation Impacts on the Hydrology of Dayton, Ohio, *Urban Ecol.*, 9, 361 376.

Sailor, D J (1995) Simulated Urban Climate Response to Modifications in Surface Albedo and Vegetative Cover, *J. Appl. Meteorol.*, 34, 1694 1704.

Taha, H, Douglas, S, and Haney, J (1997) Urban Climates and Heat Islands: Albedo, Evapotranspiration and Anthropogenic Heat, *Energy Build.*, 25, 99 103.

Terjung, W H and O Rourke, P A (1981a) Energy Input and Resultant Surface Temperatures for Individual Urban Interfaces, Selected Latitudes and Seasons, *Arch. Met. Geoph. Biokl., Ser. B*, 29, 1 22.

Terjung, W H and O Rourke, P A (1981b) Relative Influence of Vegetation on Urban Energy Budgets and Surface Temperatures, *Bound-Layer Meteorol.*, 21, 255 263.

Thompson, W (1998) Grass-Roofs Movement, *Landscape Architecture, The Magazine of the American Society of Landscape Architects*, 88, 47 51.

Titova, N (1990) Rooftop Gardens, *Science in the USSR*, 5, 20 25.

Wilmers, F (1988) Green Melioration of Urban Climate, *Energy Build.*, 11, 289 299.

Groundwater Withdrawal and the Development of the Great Man-made River Project, Libya

Hedley Moorwood
Brown and Root NA Limited, Surrey, UK

In many arid and semi-arid areas, the use of groundwater has transformed the environment, either by establishing agricultural projects in desert regions or by expanding agriculture on the margins of the desert. Libya provides a good example of such transformations.

Libya has traditionally had an agricultural based economy with most of its population centered on the farming areas on and around the mountains of the Jebel Nefusa and the Jebel Akhdar at the coast. The water supply for these areas was taken from rainfall and shallow coastal aquifers. Other populated areas were centered on oases, most of which are inland.

The increasing demand for water caused by the growth of the Libyan economy could not be sustained by the minor aquifers near the coast and over-pumping led to a consequent reduction of water quality. However, the water need for the expanded coastal farming had to be met.

Exploration following the discovery of oil identi ed huge reserves of groundwater in four major interlinked basins, inland. The water slowly drains from the southern highlands through the underground rock formations of these basins to the Mediterranean, a passage that possibly takes hundreds of thousands of years. The water is replenished intermittently during periods of high rainfall. The last of these occurred in the region approximately 6500, 8000 and 9000 years ago and there is evidence that the Tibesti and Tassili mountains enjoyed wet conditions from 19 000 to 13 000 years ago when the rest of the Sahara was suffering arid conditions.

This resource was rst used to establish large agricultural schemes in the desert but these are socio-economically unattractive as they are far from population centers. Water is now being brought instead by the Great Man-made River to where it is needed along the fertile coastal strip with its existing infrastructure. This massive project is designed to bring water from large well elds in the desert to supply towns and agricultural projects along the whole coastal strip of Libya. An eventual 6.5 million cubic meters of water per day will be supplied through thousands of kilometers of pipe of up to 4.0 m in diameter. The major coastal cities of Libya are already bene ting from potable water supplied from the desert by the Project which has been under construction since the mid 1980s.

WATER USE AND RESOURCES

Libya has had a thriving coastal agriculture for thousands of years centered on the areas of the Jebel Nefusa and the Jeffara Plain in the west and the Jebel Akhdar in the east. (The Jebels are upland areas or mountains.) The main centers of population (see Figure 1) are found in these two areas with a third being situated some 700 kilometers (km) south of the Jebel Nefusa in the numerous oasis towns grouped around Wadi Shati and Wadi Ajal. To the west and south of this are the isolated oases of Ghadames and Ghat.

In the east, the coastal plain around Benghazi to the west of Jebel Akhdar is not as large as the Jeffara Plain and generally has a limestone cap-rock surface which has to be removed before agriculture can be effective or economic. The coastal side of the Jebel Akhdar itself is extremely fertile and is the main agricultural zone. South of the Jebel, the individual oases of Ajdabiya, Jalu, Awjelah, Tazerbo, Rebianah, Kufra and Oweinat are strung out to the southern border with Sudan.

In the central zone, the coastal town of Sirt is the main center with the Al Jufrah oases of Hun, Waddan and Suknah about 300 km to the south. Beyond that there is virtually no habitation, the country being dominated by the basalt highlands of the Haruj Al Aswad and farther south by the Tibesti Mountains.

Except for the areas on and near Jebel Nefusa and Jebel Akhdar that receive annual rainfalls of up to about 300 and 450 mm, respectively, evaporation exceeds precipitation over the remainder of the country and desert conditions prevail.

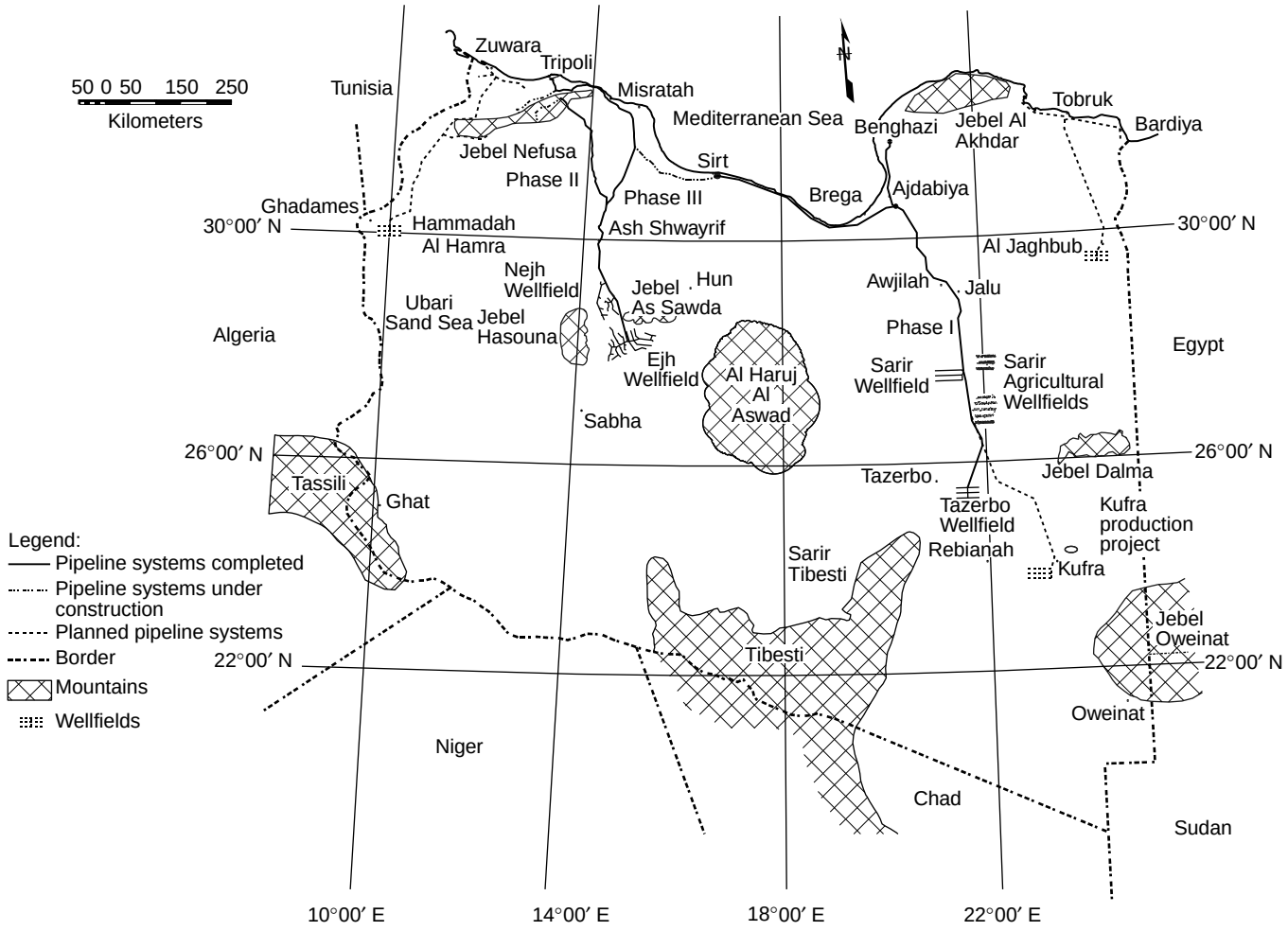


Figure 1 Location of GMRP well fields

HISTORICAL WATER USE

Traditional farming used shallow hand-dug wells, cisterns and springs for irrigation. In historical times as far back as over 2000 years ago, the agricultural area was extended inland into the system of wadis or dry river valleys cutting down through the Jebels to the coast. After heavy rain, water would flow along the wadi beds that were farmed by using trapped storm run-off caught in cisterns or dams built across the wadis (*see* Wadi, Volume 3). Extensive terracing was already developed on the hillsides of the Jebel Nefusa. In the present day, the coastal wadi systems and terracing are not farmed to the same extent, the main farming being done in large farms with better access on the plateaus of the Jebels or the plains.

Rain falls on the coastal Jebels from October to April but it tends to be in localized intense events. This results in a lot of run-off that does not provide the soil with as much recharge as it could and does not necessarily fall when or where the crops most need the rain. The use of dams and cisterns in wadis to trap the run-off sustained the level of traditional agriculture; and some indirect recharge from wadi flow into the limestone coastal aquifer helped to maintain water levels.

THE DEVELOPMENT OF AGRICULTURE AND WATER RESOURCES

In the 1960s when the oil industry was being expanded, a booming economy developed and with the continuing exploration for oil came the discovery that extensive water reserves were held in four main basins underlying Libya. The introduction of new technology and the changes in farming practice that followed resulted in the need to meet new water demands. To provide the greater quantities of irrigation and domestic water needed, pumped tube wells were used. The newly available water and Mediterranean climate at the coast enabled a full range of fruits, vegetables and cereals to be produced in ever increasing quantities for an expanding population. Dramatic environmental change

thus began in the early Seventies with major agricultural projects being established in both coastal areas and at oases to meet the aim of self-sufficiency in agricultural produce.

Oil exploration continued to prove the vast reserves of groundwater that were held inland beneath the desert. A systematic program of exploratory water well drilling began to provide details of the aquifer systems, the quality and quantity of water held in each aquifer, its nearness to the surface and rate at which it could be abstracted from the formations.

Meanwhile, in the coastal regions, the higher levels of abstraction by tubules could not be sustained by the coastal aquifers. Overpumping of the wells led to a large drawdown of the water table. Water quality deteriorated owing to saline intrusion and upcoming of lesser quality water. At the oases, expansion was limited by the unwanted effect of the increased abstraction lowering the shallow water table, which started killing off the natural, often shallow-rooted vegetation such as palm trees, and drying up lakes and springs and shallow wells. Perversely, in some oases where artesian wells were constructed, leakage caused localized flooding or oversaturation of the soil. Both of these types of situations caused marked changes to the environment of the oases concerned.

CENTER PIVOT IRRIGATION IN THE DESERT

Groundwater abstraction for the development of agriculture close to centers of population therefore had to be controlled and the Water and Soil Department within the Secretariat of Land Reclamation and Agricultural Reform, now the General Water Authority of Libya, was formed to manage the water resources. To prevent more overpumping of the coastal aquifers, the expansion of agriculture was focused on desert areas with good soil chemistry where large quantities of good quality groundwater were present. At two such locations, Kufra and Sarir (see Figure 1) center pivot irrigation schemes were started.

The first such project to pioneer large-scale desert agriculture outside the USA was built near Kufra in the late sixties, about 1000 km from the coast. The drive to use the desert for agricultural expansion led to two center pivot irrigation schemes being constructed in the mid Seventies at Sarir about 500 km south of the Jebel Akhdar. All were built in the eastern part of the country following the extensive oil and water exploration there. Kufra was an early choice because its abundance of near-surface groundwater was known. Sarir was chosen in line with the philosophy of using the desert for agriculture. The exact location was determined from mathematical modeling of the aquifer system and from water quality studies. It has a relatively high water table and flat sandy terrain and its development in its desert location far from any oases, prevented any adverse effects on existing centers.

Irrigation is provided from groundwater. Each center pivot irrigation system is centered on a well and supplied with electric power. The power operates the well pump motor and a wellhead motor that rotates a wheeled irrigation gantry from a center pivot by the wellhead. The gantry has several wheeled sections joined together to make up to the required length and the water is sprayed from it down on to the crops. The gantry slowly rotates to form an irrigated crop circle of approximately one km in diameter. The crop types grown are generally alfalfa or cereals that are suited to the irrigation system and the climate. Soil drainage and build up of salts owing to evaporation have to be monitored (*see* Salinity and Agriculture, Volume 3).

At the Kufra Production Project, the wells are spaced approximately 1.1 km apart in a matrix arranged to suit the topography of the hills to the east of Kufra. In the case of the Sarir agricultural projects, the wells are spaced every 2.5 km in double rows, 10 km apart to reduce the drawdown and spread the cone of depression of the water table. Monitoring of the water table at Kufra and Sarir has shown that the aquifers exhibit an unconfined response with the lower layers mirroring the upper layers after an initial drawdown where clay layers are present. Kufra shows a rate of drawdown at the center of the well field of 0.6 m year^{-1} over 20 years and drawdowns at the centers of the Sarir North and South agricultural well fields were recorded as less than 0.5 m year^{-1} over 7 and 11 years operation, respectively.

There are offices, accommodation buildings, workshops, storage facilities and plant and equipment at each project to run and maintain the facility and to plant, harvest and transport the crops. With the water from these wells, the desert blooms with large circles of green and the projects continue to operate although all suffer the ravages of the harsh desert environment on the equipment. There are 53 wells at the Kufra Production Project and 82 wells at the Sarir North and 157 wells at the Sarir South Agricultural Project.

The Kufra project is the most effective as it provides produce for the Kufra oases and therefore has little transportation costs and much less staffing and logistics costs as many of the staff live in Kufra. Sarir on the other hand, is situated almost 200 km from the nearest oasis and has to be manned and supported continuously. The cost of supplying the food, fuel and spares and transporting the produce hundreds of kilometers to market plus keeping a workforce operating in the desert away from their families in an inhospitable, isolated environment, makes the project much less socio-economically attractive and more difficult to run.

Although these schemes show that, with water, the desert can be used to produce crops, the area used in agricultural terms is small and the limitations of scale and their economics demanded another solution. If the coastal aquifer was already being overused and it was impractical to take agriculture to the desert where groundwater was available,

then the solution had to be that the good quality water should be piped to the fertile coastal areas. This would take advantage of existing centers of population and develop new agricultural areas away from the disadvantage of the desert environment, at the same time relieving the demand on the coastal aquifer. This solution gave rise to the Great Man-made River Project in 1982.

THE GREAT MAN-MADE RIVER PROJECT

The Great Man-made River Project is one of the largest civil engineering/water supply projects in the world today. The outline of the Project is shown on Figure 1. It carries groundwater in buried pre-cast concrete cylinder pipes, up to 4.0 m in diameter, from well fields in the desert to reservoirs and tanks and turnouts at various locations along the coast. These supply municipalities and large agricultural reservoirs. The agricultural reservoirs then provide water to new farms at each location through networks of distribution pipes. The system of agricultural reservoirs and farms is currently being expanded under the control of the Great Man-made River Water Utilization Authority.

Preliminary conceptual studies of the scheme were followed by investigations for agricultural zones, well fields and pipeline routes. To establish the water demand, potential agricultural areas were identified. Soil surveys were carried out to check soil chemistry and permeability to prevent excessive salt build-up that would harm plants and to select crops that would be tolerant of the conditions and of the water quality to be provided. These surveys and the positions of existing communities determined the end points for the system. Water demand was reviewed for each location, based on factors such as planned crop type, area to be irrigated, proposed irrigation systems and soil permeability. This planning process required the identification of water sources able to provide sufficient quantity of acceptable quality for agricultural use. Groundwater quality was also checked against municipal guidelines for potability (drinkability) because although the project was conceived as an agricultural project, the increased salinity of the coastal aquifers meant that alternative municipal supplies were also needed. The levels of chloride in the coastal aquifers were affecting taste and had accelerated corrosion in much of the water supply networks of the coastal cities, causing increased water loss and a greater maintenance burden.

The outcome of these investigations was a design to supply from desert well fields, an initial 2.0 million m^3 (i.e., metric tons) of water per day through its eastern system (Phase I) to the coastal demand areas from Sirt to Benghazi. The western system (Phase II) is to provide up to 2.5 million m^3 of water per day to the coast from Misratah to the west of Tripoli and to the Jebel Nefusa, south of Tripoli. Both Phase I and Phase II are already supplying water to the main cities and population centers along the routes of the pipelines. Phase III, which has begun, will interconnect Phases I and II and studies are underway to increase the flexibility of the system interface. A later part of Phase III will provide an additional 1.68 million m^3 of water by branching off southwards from the pipeline at Sarir and extending it into the Kufra basin where another well field will be constructed. Future Phases are in the planning stage or under design. A well field situated in the region of Ghadames in the west will pipe 0.25 million cubic meters of water per day to the coast at Zuwara. A supply to Tobruk of 0.13 million m^3 of water per day from a future well field near Al Jaghbub, approximately 300 km south of Tobruk is being designed. In the south, investigation has been underway to assess the Sarir Tibesti as a possible source from which to pipe water to Sirt.

The total cost of Phase I and Phase II is of the order of 12 billion US dollars. However, when compared to providing desalinated water in similar quantities, the unit cost of water is only 5 25% of that of desalinization, the range of cost varying with the difference in cost of the infrastructure of the relative schemes.

This vast Project already incorporates about 900 wells and well pumps, over 3000 km of 4.0 m diameter pre-cast concrete cylinder pipe, 1000 km of collector pipeline, 4000 km of haul road, a 90 mW power station, a multi-product fuel pipeline, over 4000 km of transmission and distribution power lines, 5 major pump stations, water treatment plants, flow control stations, regulating tanks and reservoirs with 35 000 000 m^3 of storage, 100 supply points, agricultural reservoirs and water distribution systems, 9 telemetry control stations interlinked with microwave radio and optical fiber networks, telecommunications and 9 operation and maintenance complexes with accommodation and office buildings for the operating body, the Great Man-made River Authority.

ENVIRONMENTAL IMPACT

The environmental impact of the project itself is limited in that the pipe and the majority of the structures are buried. The pipe has one to two meters cover and only the tops of most structures are visible because the floor level of the tanks and pump stations are built to tie in with the bottom level of the pipe, often seven meters below ground. The pipeline and structures do not therefore affect migration or movement of wild animals or of livestock and there are few settlements or crops to be disrupted as the pipeline runs mainly through desert areas.

The predicted change in piezometric head in the aquifers at the nearest towns has been checked and agreed with the General Water Authority as acceptable over the 50 year design life of the project and there are no local areas of natural vegetation where an increase in water table depth would cause problems. The monitoring of the water

table and the total abstraction from the well fields of all the projects and Municipalities in Libya allows the predictions to be refined and long-term future abstractions to be controlled and planned.

The well fields were placed well away from oases and population centers to prevent or minimize any significant interference drawdown affecting them or other existing projects. The pipeline route was studied throughout its length and any sites of known historical interest were avoided in the planning stage. Detours were made around any such sites encountered during the excavation for the pipeline. In the fertile coastal farming areas, the right of way was drastically reduced and the route planned to minimize disruption and loss of land use. Farming over the pipeline after construction and reinstatement of the ground is permitted.

Erosion protection has been introduced at cuts and wadi crossings to minimize damage from heavy rain or storm events and careful provisions made at every road and utility crossing in conjunction with all the relevant authorities and utility companies. The line is maintained by the Operating Authority which checks for erosion damage and reinstates protection where necessary. No damage or exposure of the pipe by wind erosion has been recorded.

Phase II of the Great Man-made River Project was the subject of an environmental impact study that found minimal harmful effects or potential resulting from the construction of the Project.

The results of the operation of the Project are beneficial to the population and to agriculture in the quality and quantity of water supplied. There are many other benefits following from its construction including development of the infrastructure, parallel upgrading of power and telecommunications, training and expertise and growth of Libyan companies and businesses through use of local produce and commodities.

GROUNDWATER SYSTEMS AND THE GREAT MAN-MADE RIVER PROJECT

The exploration work that followed the discovery of the huge resources of groundwater revealed the presence of

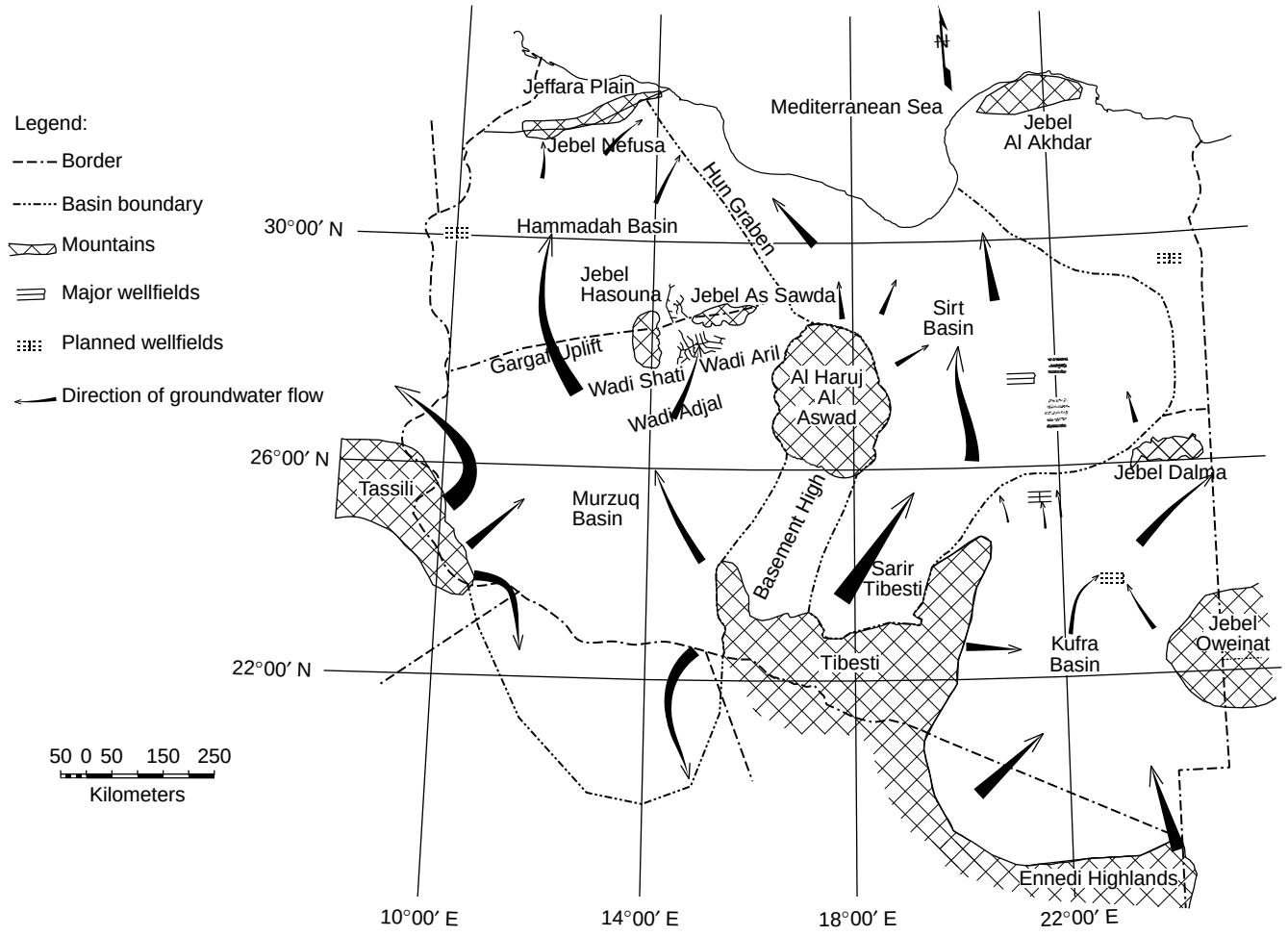


Figure 2 Groundwater flow, Libya

four main basins in Libya: the Kufra basin in the southeast, the Sirt basin in the northeast, the Murzuq basin in the southwest and the Hammadah basin in the northwest (see Figure 2). Each basin was found to contain a series of water bearing formations (aquifers) with water of different qualities depending on the origin and chemical composition of the host rocks. The aquifers are interconnected to varying extents between the basins and the whole system is considered to be draining in a quasi-steady state condition, from the southern highlands northwards to the coast. Drainage also occurs southwards from the highlands into Chad and Niger, northeastwards into Egypt from the Kufra basin and westward into Algeria from the Murzuq basin. The middle aquifer system in the Hammadah basin (the Triassic/Jurassic/Lower Cretaceous aquifer) extends as far as Morocco and is called the Continental Intercalaire.

The southern basins are continental in formation with predominantly sandstone aquifers while the northern basins suffered marine intrusions and the upper aquifers are predominantly limestone. The water quality, in terms of salinity, generally deteriorates northwards.

The ability of the aquifer system to maintain the head gradients while supporting the outflow to springs and oases, evaporation from sabkhas and other free water bodies, the water demand of the population as well as traditional agriculture and evapotranspiration losses, required that there should be a water balance for the system. With virtually no rainfall of any duration or regularity south of the coastal mountains, Jebel Nefusa and Jebel Akhdar, there is effectively zero recharge to the system. The inflow into the system is therefore considered to be provided by the depletion of the groundwater mounds in the highland areas of the Tibesti and Tassili mountains and of the southern Jebels. For the western system, this was modeled at 1.5 2.0 mm year^{-1}.

For the Great Man-made River Project, the major criteria for location of the well fields on this system were that the well fields should produce good quality water that could be economically abstracted as near to the coast as possible. Hydrogeological considerations were paramount with the cost of the pipeline length being balanced against the cost of the wells and pumping costs. The volume of water to be abstracted and the expected drawdowns precluded the use of areas with high water tables, as these were usually associated with oases where shallow wells and vegetation would be affected.

Thus the depth to the water table was not the main factor in the selection of well field locations but one of many cost factors. More important was the yield and thickness of the aquifer to sustain the abstraction and the unconfined condition to limit the spread of interference drawdowns.

The presence of old subterranean river channels was known at both Sarir and Kufra and those at Kufra were dramatically shown by the Space Shuttle radar imagery which mapped their course below the sand. This assists greatly with the understanding of the hydrogeology, especially of an upper aquifer, and the geomorphology of the area but was not available at the time of the Phase I well field design. Satellite imagery was used in Phase II to study the course of wadi drainage into the well field and was a factor used in zoning of transmissivity within the model.

PHASE I WELL FIELDS

Geological and topographical mapping together with exploratory drilling for oil and water allowed the structure of the basins to be determined and the flow system to be conceptualized. Mathematical computer models of the aquifer systems were developed in the seventies starting with the Sirt basin (Wright and Edmonds and then M. Ahmad) where the most data were available. The development of the exploratory data resulted first in the center pivot agricultural project well fields located at Sarir and then the Great Man-made River Project Phase I well fields being located at Sarir and Tazerbo. The Agricultural Projects and the Great Man-made River Project well fields at Sarir are all placed in the most northerly extent of a lobe of good quality water in the unconsolidated sediments of the upper aquifer, the Middle to Upper Miocene of the Sirt basin. The Tazerbo well field is located about 70 km from Tazerbo oasis near the northern boundary of the Kufra basin with the Sirt basin and was designed to draw water from the aquifers of the Kufra basin.

Both well fields had been modeled in the early eighties and a drawdown of the water table of about 70 90 meters in the centers of the Sarir and Tazerbo well fields was predicted after 50 years of abstraction of 1.0 million m^3 of water per day from each well field. At Sarir, 126 wells with a nominal flow rate of 92 liters s^{-1} are arranged in three rows, 10 km apart with a spacing of 1.3 km between wells. At Tazerbo, there are 108 wells with a nominal flow rate of 120 liters s^{-1} also arranged in three rows with the same spacing as at Sarir.

Owing to the closeness of the Tazerbo well field to the boundary between the basins and because of the complexity encountered there of the interconnected aquifers and the impermeable formations (*aquitards*) between them, additional modeling of the well field area was subsequently carried out by the Consultants, Brown and Root.

Exploratory wells were drilled to identify the geology and layering of the aquifers and core samples analyzed to obtain the age and permeability of the rock. The aquifers were found to be older Devonian formations rather than the Nubian deposits previously assumed. However, the geological complexity of the system and the lack of water abstraction data in the area for use in the non-steady state calibration of the model meant that it could not be used as a predictive tool as no unique solution for the calibration

could be achieved. Nevertheless, a considerable increase in the understanding of the aquifer system at the well field was achieved from the model, and the depth of some wells amended in consequence.

At Tazerbo oasis, deep wells are being drilled to provide a source of water that will not be affected in case the cone of depression from the Project s well field extends as far as the village. This removes dependency on shallow wells and provides compensatory irrigation water for shallow rooted natural vegetation that would be affected if the water table in the upper aquifer was lowered as a result. At Sarir, there are no oases or centers of population and the agricultural projects are far enough away not to be affected by the drawdown.

At both well fields, water quality analyses were carried out for the whole range of chemical and microbiological constituents to check for suitability for agricultural and potable use for corrosion potential. As the main pipeline is made of concrete, the concentration of carbon dioxide present in the waters was specifically measured because it was necessary to establish if it would result in the water dissolving the cement in the concrete pipe and weakening it. Excess free carbon dioxide can be removed from the water but removal of too much will cause deposition of calcium carbonate (scaling) on the inner surface of the pipe which will reduce the flow. Neither of these is desirable so to prevent either from occurring, the level of free carbon dioxide was established at which the waters remain neutral to the pipe. This showed that water treatment plants and upstream internal coating of the collector pipes were needed for Tazerbo well field.

PHASE II WELL FIELDS

In the west, the situation is complex. Several towns and oases exist on the northern edge of the Murzuq basin along the Wadi Shati and Wadi Ajal where the groundwater flows from the Tassili Mountains emerge. To the south of the oases, the desert surface of the Murzuq basin extends to the Tassili Mountains. To the north of the oases, the terrain is dominated by the escarpment of the Gargaf uplift and the mountains of Jebel Hasouna and Jebel As Sawda. The stony plateau of the Hammadah Al Hamra, cut by major wadis, then extends northward across the Hammadah basin to the foothills of the Jebel Nefusa with the oases of Al Jufrah to the east. To the west are the Ubari sand seas.

Several areas had previously been considered as possible locations for the well fields but none had been chosen. This and the size of the planned abstraction and the need to prevent any unacceptable interference with the numerous oases, towns or existing agricultural projects led to the decision to model the whole western aquifer system to its aquifer boundaries as had previously been done for the Al Jufrah project. By this means, the lateral continuity of the aquifers, the physical condition at those boundaries and the hydrogeological features of the original system were taken into account.

The GEASS quasi-3D finite element program developed by Pizzi and Sartori of Geomath was chosen to model the multi-layered aquifer system envisioned which was based on that postulated by Pallas with later modifications by the General Water Authority. This considers each basin to have three main aquifer layers with the lowest underlying both basins and covering almost the whole of western Libya. The middle and upper layers in each basin interconnect with each other but the main interconnection is with the underlying aquifer. The modeling, done by Geomath, was controlled by a joint committee set up under the Great Man-made River Authority with the General Water Authority, Brown and Root and their respective consultants for hydrogeological interpretation.

The water levels prior to 1970 were taken as the effective steady state condition as any traditional abstraction before that date had not caused any change in drawdown. The model was calibrated by examination and adjustment of the hydrogeological parameters to achieve close correspondence of the computed water table heads with measured values. Post 1970 abstraction and water level data were collated by the General Water Authority and used to calibrate the model in the non-steady state.

Once the model was shown to compute correctly the water table resulting from the abstractions up to 1989, it was used to test the system response to future bulk abstractions in various locations. The final well field location was chosen to tap the Cambro-Ordovician Devonian sandstone aquifer, the lowest aquifer layer that underlies both the Murzuq and the Hammadah basins. The well fields were located northeast and east of Jebel Hasouna where the Cambro-Ordovician rises up at the Gargaf uplift and outcrops and where the aquifer is therefore unconfined with only the northern edge of the well field becoming confined as the aquifer dips northwards under the Upper Cretaceous limestone. This allows the well field to utilize the specific yield of the unconfined area. The model is sensitive to regional specific yield so from the non steady state calibration, the minimum value that correctly matched the measured zero drawdown of the unconfined area up to 1989 was used.

Taking abstraction rates derived from exploratory well pump tests, it was determined that 440 production wells would be required to provide 2.0 million m^3 day^{-1} of water with piezometer wells to monitor the aquifer response. These wells are distributed over an area of 11 800 km^2 using similar spacing to those used in Phase I well fields: 1.5 km between wells in rows 10 km apart. A further seven exploratory wells were drilled to give a better spread of hydrogeological data in the area, to confirm the extent of the aquifer system in the well field area, to give more

information on drilling conditions and to confirm the water quality. Rock cores were tested for permeability and old pump test data were reworked.

The age of the water was checked and a mixed flow identified with quality varying across the area. Water quality and corrosivity studies were carried out using depth samplers, corrosivity meters, corrosion coupons and microbiological arrays. From the hydrochemistry and these data, the blend quality of the water was assessed and the water was found to be generally satisfactory for agricultural and potable use with minor treatment. As a result, carbon dioxide degassing has been introduced to remove its aggressivity to the pipeline and corrosion resistant high-grade stainless steels and fiber reinforced plastic used in the wells.

Owing to schedule constraints, large-scale mapping of this area was done by using existing aerial photography to the best scale possible. The layout of the well field was adjusted to the topography using the large scale mapping and rationalized by modeling future simulations to obtain the best response which was a maximum of about 100 meters drawdown at the center of the northeast well field after 50 years. The number of production wells was increased by 10% for standby purposes and these were added to the well field layout. The design of the well field was done by linking the results of this modeling to an hydraulics program and a costing program which defined the pipework design and the location of the structures and the connections. From this, the total pumping head at each well was calculated.

The model was also used to simulate future drawdowns that would occur on other projects and centers and to assess the future expansion scenarios to obtain 2.5 million m^3 of water day^{-1}.

The modeling exercise allowed the Phase II well field to be placed as far north as possible and away from population centers, which resulted in large cost savings and acceptable minimal interference effects and provided design criteria for much of the project. Being a regional model, it is also available for other studies and it has already been used by the General Water Authority to model the well field near Ghadames in the Triassic Jurassic Lower Cretaceous aquifer for the next Phase of the Project.

PLANNED ABSTRACTION FOR FUTURE PHASES

The Great Man-made River Project is an ongoing project that is being expanded and interlinked. The pipeline between Phase I and Phase II has been started and the studies and exploration for a model of the future well field in the Kufra basin are underway. It is intended to use a similar approach to modeling the basin and well field as that used in Phase II but the remoteness of the area means that data sources for calibration are very limited. Data collection is therefore extremely important and preliminary field trips to map the geology and gather data from existing old exploratory wells and areas of natural vegetation have been made. Aerial photography has been assembled and the existing near-surface seismic data from oil exploration reworked. A remote sensing study has already been carried out and the exploratory program for the modeling has been agreed. The model that is to be built from this will be interrogated to minimize the environmental impact of the chosen well field location.

The well field at Ghadames for the planned abstraction to supply Zuwara in the west has already been modeled and the project is being started. The results of the exploratory drilling at Al Jaghbub in the east to supply Tobruk are very encouraging. Approval for the scheme has been given and design studies are underway.

OPERATIONAL AND ENVIRONMENTAL ASPECTS

Sarir well field has been supplying water to the coastal demand areas of the Phase I system from Benghazi to Sirt since 1991 and Tazerbo well field is also completed. The northeast Jebel Hasouna well field has been producing water for Tripoli and the other Phase II locations since 1996.

Data are collected from production wells by the telemetry systems and from the piezometer wells by data recorders to monitor the performance of the well fields, the flows and drawdowns of the wells and the cones of depression against the predicted values. These are used for continual upgrading of the database of measured values and will be used for the later recalibration of the models. The telemetry is also used to zone production within the well fields to meet the demand taking into account the performance of each well. The 10% standby capacity allows for maintenance of the wells.

The environmental impact of the groundwater abstraction of the Great Man-made River Project itself is minimal owing to the remoteness of the well fields and the precautions taken. The amount of water abstracted, although large, is tiny compared to the volume available and at the planned rate of abstraction, the water mounds in the southern highlands driving the system will be recharged in geological time at the next pluvial before any significant depletion of them occurs with or without the Great Man-made River Project. The buried nature of the pipelines means that the system has little effect on the environment and the supply of the water to the coastal regions will alleviate the harmful effects of overpumping of the coastal aquifers that have already taken place and provide improved potable water quality. Although the quantity of water is large, the increase in new agricultural areas is small

in comparison to existing agriculture areas. As a result, no major environmental change is expected.

ACKNOWLEDGMENTS

This article is based on the groundwater development, modeling and hydrogeology carried out under Mr A S Kuwairi of The Great Man-made River Authority by its engineers and its Consultant Dr A Binsariti; by Mr O Salem, Mr A ElSonni *et al.* of The General Water Authority of Libya and by Messrs. H Moorwood, R Pim and P Harrup *et al.* of Brown and Root and its Consultant Dr J W Lloyd.

REFERENCES

Ahmad, M U (1983) A Quantative Model to Predict Safe Yield for Well field in the Kufra and Sarir Basins, Libya, *Groundwater J.*, 21(1), 58 66.

Lloyd, J W, Binsariti, A, Salem, O, El Sunni, A, Kwairi, A S, Pizzi, G, and Moorwood, H (1997) The Groundwater Assessment for the Western Jamahiriya System Well field, Libya, eds Fei Jin and N C Krothe, *Proc. 30th Int. Geol. Congr.*, 22, 258 269.

Pallas, P (1980) Water Resources of the Socialist Libyan People s Arab Jamahiriya, in *Geology of Libya*, eds M J Salem and M T Busrewil, Academic Press, London.

Pim, R H and Binsariti, A (1994) The Libyan Great Man-made River Project, Paper 2: The Water Resource, *Proc. Instn. Civ. Engrs., Wat. Marit. Energy*, June, 106, 123 145.

Pizzi, G and Sartori, L (1984 85) Interconnected Groundwater Systems Simulation (IGROSS) a Description of the System and a Case History Application, *J. Hydrol.*, 75, 255 285.

Townley, L R (1990) *Aquifem-N: a Multi-layered Finite Element Aquifer Flow Model User's Manual and Description*, CSIRO, Division of Water Resources, Perth, Western Australia.

Wright, E P, Benfield, A C, Edmund, W M, and Kitching, R (1982) Hydrogeology of the Kufra and Sirte Basins, Eastern Libya, *Q. J. Eng. Geol.*, 15, 83 103.

H

Harmful Algal Blooms (HABs)

I J Hodgkiss
The University of Hong Kong, Pokfulam Road, Hong Kong

Occasionally, many coastal areas of the world experience the rapid formation of dense, visible patches of microscopic marine plants near the surface of the water offshore. When these plants contain reddish pigments, they give a red appearance to the sea, leading to the popular description of red tides.

Harmful algae are microscopic, single-celled plants that live in the sea. The world s oceans are home to a dazzling array of small plant and animal life, which together we call *plankton*. The plants, the algae or phytoplankton, are the base of the entire marine food web, leading eventually to the largest marine mammals the whales, the largest marine fish (the whale sharks) and, of course, to humans who feed on the fruit of the seas.

When these algae grow in large numbers we call them blooms and these blooms may reveal the colors of the algal pigments red, brown, yellow, etc. The most obvious are those we call the red tides, caused by minute algae called dinoflagellates.

Sometimes these blooms can have adverse impacts on the environment, its inhabitants, and on people when those inhabitants are used as food. The impacts of these harmful algae blooms (HABs) may be due to the sheer numbers of cells clogging fish gills; or they may be due to the deoxygenation resulting from such large numbers of breathing cells; or, in a small number of cases (not more than a few dozen), they may actually produce poisonous or toxic substances.

These toxins may be accumulated in shellfish, as they filter the water for food and oxygen, until they reach levels that can harm humans or other consumers in the marine food web. Or, they may kill fish, particularly those held in pens or cages in mariculture areas. HABs are global phenomena and research studies all over the world suggest that the nature and extent of this HAB problem has been expanding over the last few decades. There are now reports from almost every coastal country. Explanations for this increase range from natural species dispersal by ocean currents, to an increase in human-related and global activities such as transport of species in ship ballast water, increased inshore activities and pollution, to more vigilance. This last reason has little validity where increases have been recorded in the past 20 years by monitoring programs that have changed little over that time.

Phytoplankton bloom dynamics are related to nutrient supply, and this might result from riverine inflows of water, containing nutrients derived from natural sources, being drained into the rivers from their surrounding catchments during rainfall, or sewage effluent inflows generated by Man, or the upwelling of nutrient rich waters from the oceans depths. This process of nutrient enrichment (*eutrophication*) can thus be natural or man-made (*see* Eutrophication, Volume 2). However, evidence is accumulating from all over the world that accelerated eutrophication, i.e., the speeding up of this natural process by people-made inputs of domestic, agricultural and industrial wastes, is stimulating algal growth in coastal waters and leading to undesirable changes. Thus, a persistent and global pattern has emerged of a long-term increased frequency of HABs associated with coastal marine pollution.

The nature of the HAB problem has changed considerably since 1975 in the US. Previously problems arose in scattered localities in a few states, but now virtually every coastal state is threatened, in many cases over large geographic areas and by more than one harmful or toxic algal species. The number of toxic blooms, the economic losses from them, the types of resources affected, and the number of toxins and toxic species have all increased markedly, not only in the US, but throughout the world. Few US coastal waters are unaffected by HABs. In the 1990s, fish kills in the middle Atlantic states were linked to a newly discovered organism called *P esteria*. Florida and other Gulf Coast states are subject to fish kills from *Gymnodinium breve* red tides. The tropical areas of the Florida Keys, Hawaii, and

Puerto Rico have experienced ciguatera fish poisoning for some time.

In southeast Asia, red tides have been observed in many coastal waters such as bays in the Philippines, Brunei Bay and off Sabah, Malaysia. The casualties due to the consumption of contaminated fisheries products have also been significant. A red tide reported in Hong Kong in late March 1998 affected about 1000 of Hong Kong s 1500 fish farms and killed two-thirds of the marine fish farm population at an estimated loss of HK$120 million. It forced the closure of the popular Repulse Bay beach. Local pollution by phosphorous in industrial sewage, probably from heavy industries like dye factories and shipbuilding plants in Guandgong province, was blamed for the algal bloom.

Climate change could alter local habitats and change the areas where ciguatera toxin is produced, but such changes might be undetected by local fishermen or regulators. Warming would expand the range for ciguatera and other biotoxins and could lead to increased biotoxin poisoning.

To most scientists, the future appears to be clear a concerted effort is necessary to protect our marine environment before the effects of waste disposal become overwhelming. In many countries this effort has begun, others have yet to recognize the problems.

FURTHER READING

Anderson, D M (1994) Red Tides, *Sci. Am.*, 271(2), 52 58.

Anderson, D M (1995) Toxic Red Tides and Harmful Algal Blooms: a Practical Challenge in Oceanography, *Rev. Geophys.*, 33(Suppl.), 1189 1200.

Anderson, D M (1997) Turning Back the Harmful Red Tide, *Nature*, 388(2), 513 514.

Hodgkiss, I J and Ho, K C (1997) Are Changes in N : P Ratios in Coastal Waters the Key to Increased Red Tide Blooms? *Hydrobiologia*, 352, 141 147.

Okaichi, T, Anderson, D M, and Nemoto, T, eds (1989) *Red Tides: Biology Environmental Science and Toxicology*, Elsevier, New York.

Health: the WHO Definition

see WHO (World Health Organization) (**Volume 4**)

Heat Island, Urban

see Urban Heat Island (**Volume 3**)

Holocene: Climate Changes and Society

Burkhard Frenzel
Universität Hohenheim, Stuttgart, Germany

People are affected by climate, but in turn they also contribute to changes in climate. The key ideas of these two-directional interactions since the last Ice Age include:

1. *it is repeatedly stated that climate changes triggered migrations of huge numbers of people, such as the Migrating German Tribes in Europe and the raiding and migrating Mongols and Huns moving into Eastern and Central Asia and into Eastern, Central and Western Europe;*
2. *climate change may have triggered profound changes in, or a reorganization of, socio-economic systems, even when migration did not occur;*
3. *robust, stable human societies may have been able to buffer climate change, while other societies affected by internal political, demographic or economic dif culties may have been less able to adapt to climate change;*
4. *human economic activities may have strongly in uenced climate directly or indirectly, e.g., by forest clearances, excessively intensive land use, or commercial and industrial processes.*

Since the interactions between climate and society are highly complex and because these connections began in the dim and distant past, events early in history need attention as well as those of the recent industrial past.

The basis for considering the influence of changing climate on human societies is good knowledge of past climates, in time and space, and of the amplitude of past climatic oscillations. Originally it was held that during postglacial times (the Holocene, i.e., the last 11 550 years) several major changes of climate had taken place. They were thought to have produced important changes in the vegetation cover of the continents. It was held that these major shifts in climate had happened at the transition between periods characterized by more or less constant climatic conditions, the character of which was derived from plant-fossil assemblages in Norwegian peat bogs (Blytt, 1882). This was the basis for the well-known names of preboreal, boreal, atlantic, subboreal and subatlantic, the ages of which were estimated by assumptions about the velocity of sedimentation and by archaeological findings. Later on, many more indicators of past climates were used and better methods of dating palaeoecological samples were developed, especially the radiocarbon method (^{14}C (carbon); e.g.,

Godwin, 1975). It soon became evident that comparable steps in the vegetation history of various regions, such as Europe, had happened asynchronously. Thus, although names like preboreal, boreal and atlantic were still used in many countries, it became necessary to divide the Holocene chronostratigraphically, i.e., by the acceptance of certain well-defined ^{14}C-ages to differentiate between these biozones. This meant that during the Holocene, vegetation changes, even when they were highly marked, were much more governed by biological processes than by climate. On the other hand, analysis of a wealth of indicators of past climates, e.g., stable isotopes in the cellulose of tree-rings or in annually layered lake sediments, showed that the Holocene climate changed repeatedly and that the phases of change could be traced over vast regions of the globe, if not everywhere (e.g., Lamb, 1972; Frenzel *et al.*, 1992). The variability of climate appears to have increased considerably by about 6000 BP (before present; Frenzel *et al.*, 1997) with amplitudes of mean temperature in the month of August of about 2 4 °C. Using tree-rings, annually layered lake sediments or annual ice layers in inland-ice masses, it was demonstrated that the early ^{14}C dating method was not precise enough for reliable comparisons of radiocarbon data with historical data. In turn, this facilitated calibration of ^{14}C dates against tree-ring data (Stuiver *et al.*, 1998) so that physical age data of the past can be directly compared with historical data. This is important, if the relationship between climate change and events in early human societies and their economies are to be investigated.

Many claims are made that climate change triggered important changes in human society. No doubt, this triggering happened near the climatic boundary of permanent land use, e.g., at Lake Titicaca (Binford *et al.*, 1997) and during the Iron Age in Norway (Jankuhn, 1980). Yet the question remains whether this also holds for areas in more or less optimal climatic conditions. In this respect, the spread of Neolithic cultures from the Near East to Europe raises some questions. At that time, agriculture and livestock-raising spread from typically Mediterranean climates to typical warm- and cool-temperate continental to subatlantic and atlantic climates. Even though, at that time, minor changes of climate occurred, it is unlikely that these peoples could totally absorb the climatic contrasts within such a large area. Either migrating tribes or the diffusion of cultural traditions, or both, were at that time strong enough to spread into other climatic provinces. Here, climate cannot have been decisive, though a typical Mediterranean climate differs markedly from those of Central and Western Europe. Evidently new survival and adaptation techniques allowed this migration to occur.

As previously stated, the climate changed repeatedly during the Holocene, especially during the last 6000 years. At that time, particularly after about 3600 BP, several tribal migrations and movements of groups of people occurred. A comparison with the known history of climate change (Figure 1) reveals that sometimes periods of climatic change coincide with the start or the end of migrations, but often they do not. One interesting example is the raiding by the Mongols at the end of the 12th and beginning of the 13th centuries from the Altai Mountains and Mongolia into Eastern and Central Europe. This was the time of the so-called Medieval Climatic Optimum, which, in Central and Western Europe at least, in contrast to the concept of an uninterrupted climatic optimum, consisted of periods of cooler and moister and of warmer and drier climates (*see* Medieval Climatic Optimum, Volume 1). Yet these well-planned raids were carried out from 1190 AD onwards, with reconnaissance raids in various directions, mainly to the west and southwest. Only after these raids, and examination of the political and military situation in countries to the west, did the main attack occur from 1236 to 1241 AD. Of course, this example must not be generalized uncritically. However, it cannot be concluded that climate change in the past caused many tribes or groups of people to migrate and to occupy the territory of other peoples, had the migrating groups not been strong enough to do so. Thus, the interrelationships between changes of climate and the behavior of early human societies have to be analyzed as individual cases without premature generalization.

In this respect, yet another factor should be considered overpopulation, which cannot be sustained indefinitely if the climate becomes harsher or if groundwater resources decline. In these cases, society may collapse due to inappropriate decisions about resource management or unwise use of the environment. The other general problem in the relationship between climate and human societies is the impact of people upon climate, one of the causes of atmospheric changes. People began to influence global and regional climates several thousand years ago. Initially this was through changes in surface albedo, which was different in summer and winter in temperate and higher latitudes, and by forest clearance, which caused changes in surface heat and moisture fluxes.

Secondly, people caused considerable changes in tropospheric chemistry by burning wood, by agriculture, by overgrazing, and by ore smelting over several millennia. To estimate the influences and consequences of forest clearance on climate, computer simulations have been run, studying either global climate (Chase *et al.*, 2000) or regional climates (e.g., Dumenil and Lie , 1999). It turns out that, up to the present, global climate seems to have been influenced only weakly. However, strong regional differences exist, which, if they are to be believed, could affect human activities considerably.

Prediction of the consequences of man-made changes in tropospheric aerosols is even more difficult. Hong *et al.* (1998) reported on the deposition of copper aerosols

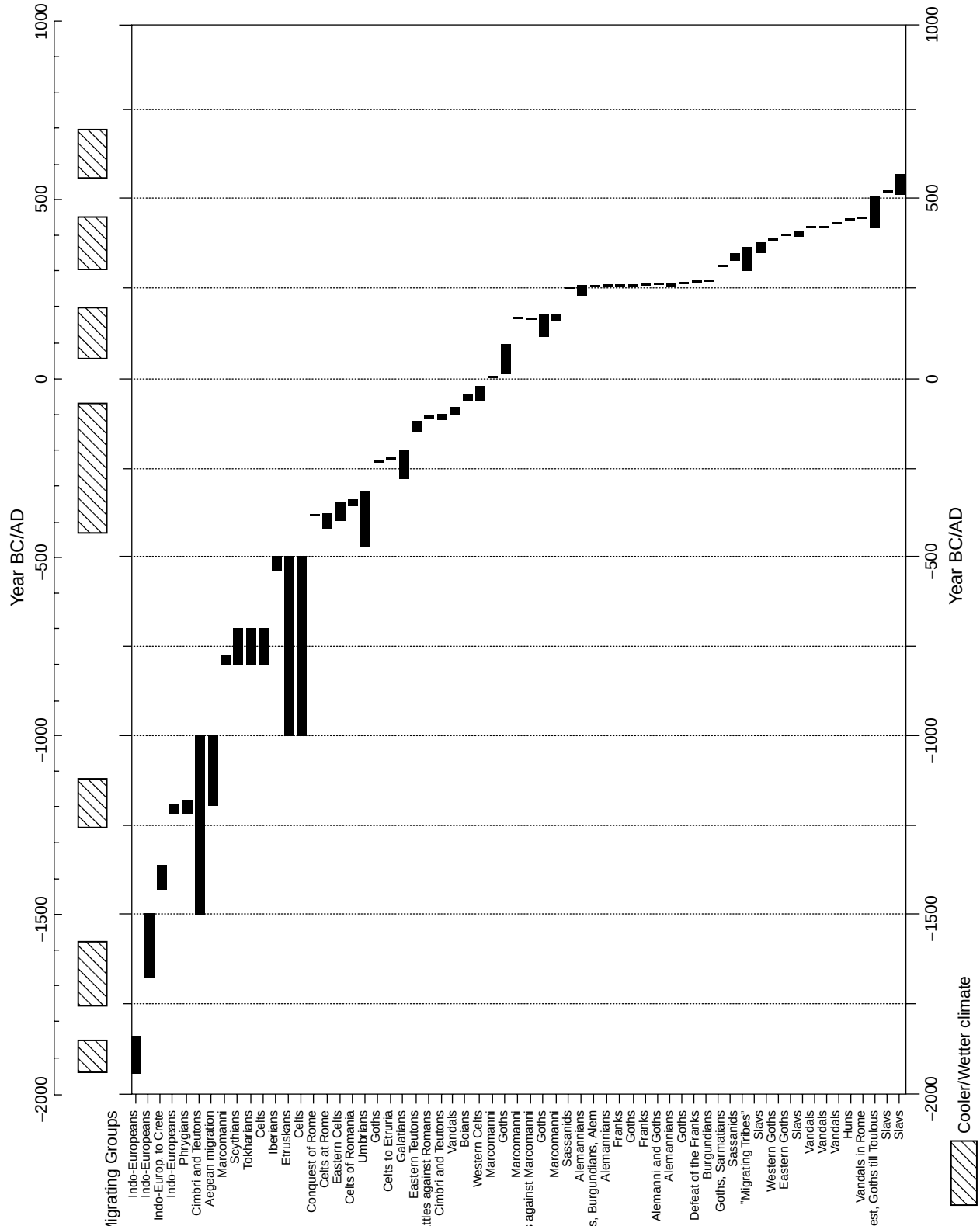


Figure 1 Times of migrations, of pressures of various tribes on other communities and periods of more pronounced climatic change in Europe. Diagonal hatching: relatively cold periods, generally moist summers. Black horizontal bands: phases of cold climate. No bands: periods of warm and generally dry summers. (Climate data from stable isotopes ($\delta^{13}C$ and $\delta^{2}H$) in the cellulose of accurately dated tree-rings)

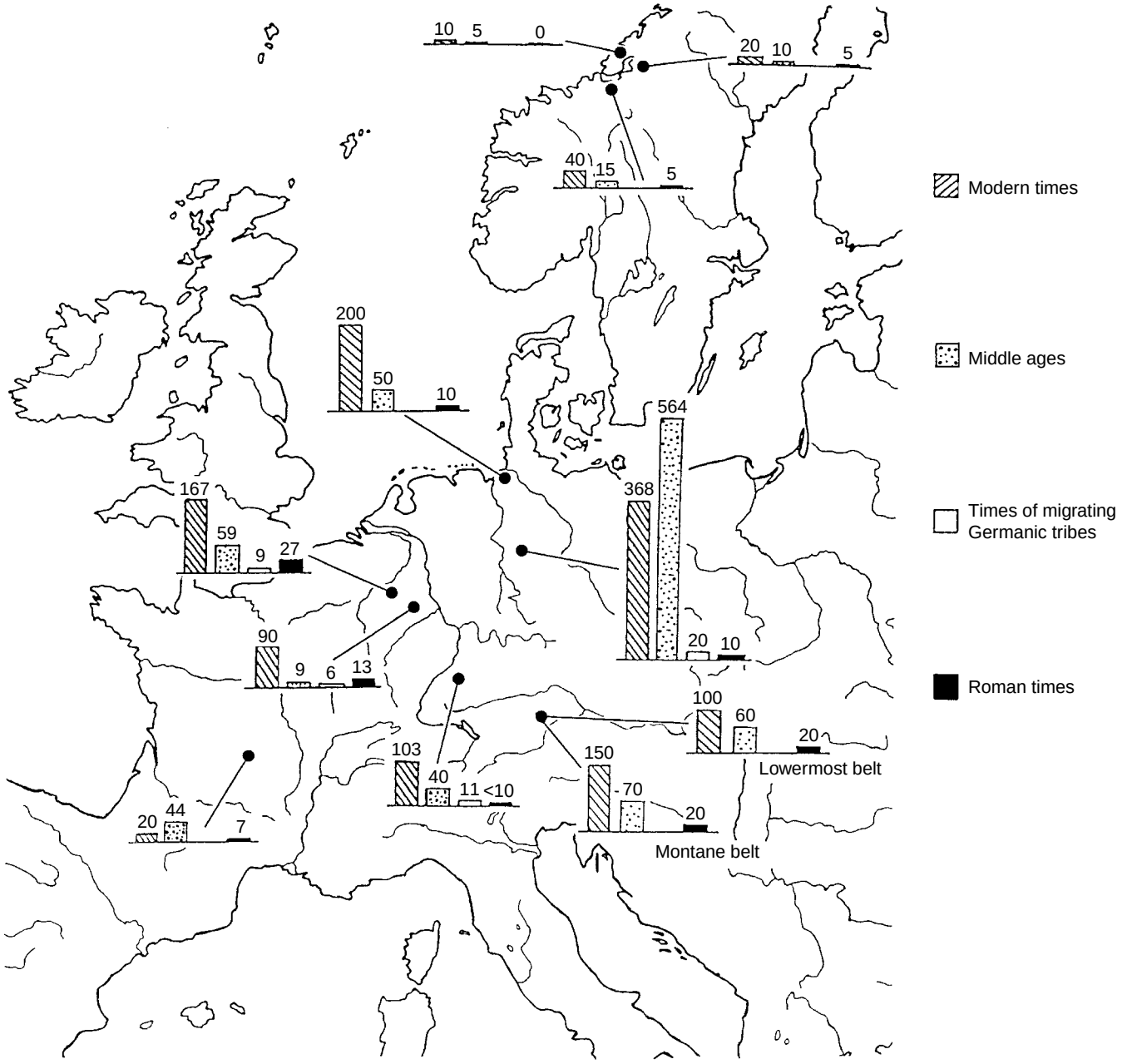


Figure 2 Lead (Pb) content as an indicator of commercial or industrial activities of man in ombrogenous peat-layers (in parts per million)

in Greenland during Roman times. Comparable investigations also took place in Southern Sweden and Central Europe (e.g., Shotyk *et al.*, 1998; Gorres, 1991). It became evident that the deposition of people-made aerosols has a long history. Yet calculation of source area emissions from the deposition rates has only just begun. Remarkable regional differences in the deposition rates of certain chemical elements in Europe (Figure 2) were the result of contrasts in agricultural practices, particularly plowing, during historical times, and of ore smelting and mineral processing. This is paralleled by observations of an increase in the deposition intensity of minerogenic dust in deep-sea cores off the West African coast. On the other hand, at the boundary of semiarid regions, as in Northwestern China, steppe soils and some forest soils are frequently covered by eolian dust. This may be interpreted as indicating the influence of the people-driven destruction of the natural vegetation cover, which loosened the soil and allowed

winds to blow away dust. (Some Chinese Scientists believe that the Silk Road contributed to desertification in the surrounding areas.) In turn, these tropospheric aerosols should have enhanced the formation of haze, fog and clouds, thereby influencing moisture and heat balances. Yet we are only at the start of pertinent but difficult investigations. These studies are difficult because many indirectly interrelated processes and factors have to be considered.

It must be concluded that much more research is needed on the interrelations between human societies and climate change, relying not only on modeling, but also on precise field investigations.

See also: Migrations: The Environmental Challenge of Population Movements, Volume 3.

REFERENCES

Binford, M W, Kolata, A L, Brenner, M, Janusek, J W, Seddon, M T, Abbott, M, and Curtis, J H (1997) Climate Variation and the Rise and Fall of an Andean Civilization, *Quatern. Res.*, 47, 235 248.

Blytt, A (1882) Die Theorie der Wechselnden Kontinentalen und Insularen Klimate, *Botan. Jahrb.*, 2, 1 50, 177 184.

Chase, T N, Pielke, R A, Kittel, T G F, Nemani, R R, and Running, S W (2000) Simulated Impacts of Historical Land Cover Changes on Global Climate in Northern Winter, *Clim. Dyn.*, 16, 93 105.

Dumenil, L and Lie , S (1999) Impacts of Deforestation and Afforestation in the Mediterranean Region as Simulated by the MPI Atmospheric GCM, *Max Planck Institut für Meteorologie, Report 301*, MPI Hamburg.

Frenzel, B, Pécsi, M, and Velichko, A A, eds (1992) *Atlas of Palaeoclimates and Palaeoenvironements of the Northern Hemisphere: Late Pleistocene - Holocene*, Fischer, Stuttgart.

Frenzel, B, Boulton, G S, Glaser, B, and Huckriede, U, eds (1997) Glacier Fluctuations During the Holocene, *Paläoklimaforschung/Palaeoclimate Res.*, 24, 182.

Godwin, H (1975) *History of the British Flora. A Factual Basis for Phytogeography*, Cambridge University Press, Cambridge.

Gorres, M (1991) Uber den Eintrag anorganischer Nahrstoffe in ombrogene Moore als Indikator der Ehemaligen Aerosolbelastung, *Dissertationes Botanicae*, 181, Cramer, Berlin.

Hong, S M, Candelone, J P, Petterson, C C, and Boutron, C F (1996) History of Ancient Copper Smelting Pollution During Roman and Medieval Times in Greenland Ice, *Science*, 272, 246 249.

Jankuhn, H (1980) Klima-Anderungen seit dem Neolithikum und ihre Bedeutung fur Siedlung und Wirtschaft bis zur Wikingerzeit, *Veröffentlichungen der Joachim Jungius Gesellschaft der Wissenschaften*, Hamburg, 44, 115 139.

Lamb, H H (1972) *Climate: Present, Past and Future, 1: Fundamentals and Climate Now*, Methuen, London.

Shotyk, W, Weiss, D, and Appleby, P G (1998) History of Atmospheric Lead Deposition Since 12 370 C-14 yr BP from a Peat Bog, Jura Mountains, Switzerland, *Science*, 281, 1635 1640.

Stuiver, M, Reimer, P J, Bard, E, Beck, J W, Burr, G S, Hughen, K A, Kromer, B, McCormac, G, van der Plicht, J, and Spurk, M (1998) Intcal 98 Radiocarbon Age Calibration, 24 000 0 cal BP, *Radiocarbon*, 40, 1041 1083.

Human Disturbance of the Earth System: Dynamics and Complexities

see Human Disturbance of the Earth System: Dynamics and Complexities (*Introductory essay*, **Volume 3**)

Human Impacts on Marshes

see Marshes, Anthropogenic Changes (**Volume 3**)

Hurricane George, Impacts of

see Sustainable Cities Policies (**Volume 4**)

Ice Storm in Montreal, January 1998

see Sustainable Cities Policies (**Volume 4**)

Indus Basin: a Case Study in Water Management

Ian Douglas
University of Manchester, Manchester, UK

The Indus rises in the Himalayas in the Tibet region of China at 4900 m above sea level in some of the steepest country on earth. Great glaciers and melting snows feed the stream that descends precipitously through steep gorges for the first 800 km or so of its 2900-km course. These snow and ice melt waters from the Karakoram in the Upper Indus Basin (UIB) supply much of Pakistan's hydroelectricity, irrigation water and domestic supplies. Their effective management requires a deep understanding of the physical processes that govern meltwater production in the UIB.

Through the mountains, the Indus flows northwest across Jammu and Kashmir, then bends to the southwest and descends to the hot, dry plain of the western Punjab. Punjab means land of the five rivers. They are the five rivers that empty into the Indus there: the Jhelum, Chenab, Ravi, Beas, and Sutlej. Next, the Indus crosses the Sind plain of southern Pakistan and reaches the Arabian Sea of the Indian Ocean through a delta some 210 km wide. Below the confluences of the Punjab rivers, the Indus widens and its gradient decreases. As its mean velocity falls with the gentler slope, much silt is deposited on its bed, raising its level above the sandy plain. Embankments are constructed to contain floodwaters, but occasionally they give way and floodwaters spread rich alluvial soils over vast areas of the plain.

In Pakistan and the Indian sector of the basin, cultivation of the rich alluvial soils has been the main economic activity for centuries. The Indus irrigation system is still the world's largest. In 1993, irrigation occupied 80% of Pakistan's 21.25 million hectare (ha) of cropland and employed about 55% of the labor force. The sustainability of the irrigation system is thus essential for the great majority of the people of Pakistan and for their neighbors in the Indian part of the Punjab.

Modern irrigation engineering work commenced around 1850, with the construction of large canal systems by the British administration. In many cases, old canals and inundation channels in Sindh and Punjab were revived, modernized and developed into the greatest canal irrigation system in the world. At partition in 1947, the international boundary between India and Pakistan split the integrated irrigation scheme of the Bari Doab and the Sutlej Valley Project between the two newly independent countries. Conflict arose because the headworks fell to India while the canals ran through Pakistan. This led to a disruption in the water supply in some parts of Pakistan. The dispute continued for some years until it was resolved through the mediation of the World Bank by a treaty between Pakistan and India (1960) known as the Indus Waters Treaty. According to this agreement, the flow of the three western rivers of the Indus basin: the Indus, Jhelum, and Chenab (save for a small quantity used in Kashmir) is assigned to Pakistan, whereas the entire flow of the three eastern rivers, the Ravi, Beas, and Sutlej, is reserved exclusively for India (see Figure 1).

INDUS BASIN PROJECT

To replace the water supplies lost to India through the diversion of waters of the eastern rivers, as provided under the 1960 treaty, the Indus Basin Project, involving the construction of two large dams, the Mangla and Tarbela, five barrages, one siphon and eight inter-river link canals, was undertaken by the Pakistan Water and Power Development Authority between 1960 and 1977. The Tarbela Dam was, at

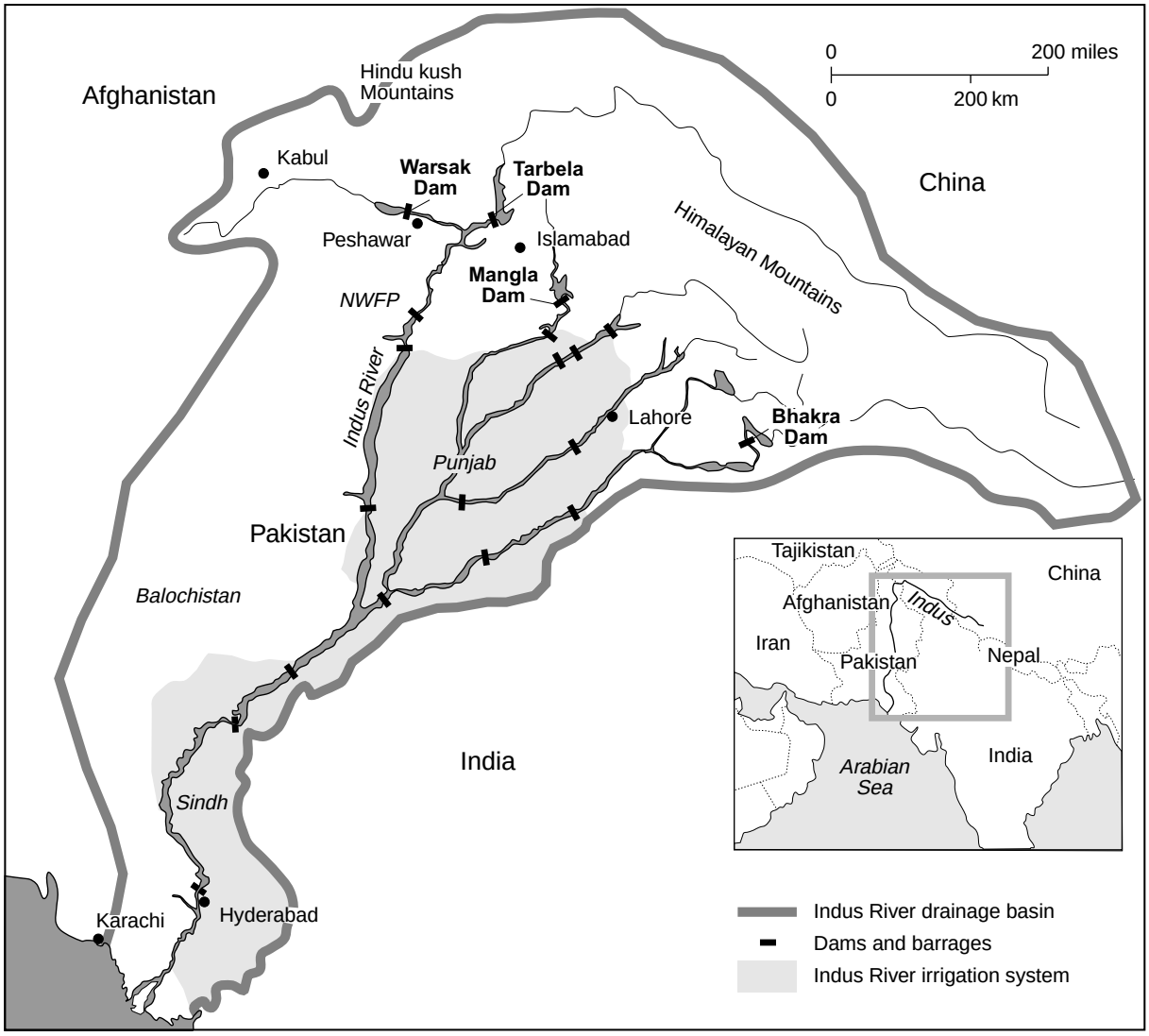


Figure 1 Map to show the major dams, barrages and irrigation areas of the Indus basin

the time, the world s largest earth/rock filled dam. The first large earthfill dam, the Mangla Dam on the River Jhelum was commissioned in 1967. The dam is 120 m high and has a live storage capacity of 5940 million m^3 with a power generating installed capacity of 1000 MW. The Tarbela Dam on the River Indus was completed in 1975 and put into operation in 1976. It is 160 m high and 3000 m long, with an original live storage capacity of 10 740 million m^3 and installed capacity of 3500 MW of hydropower generation (see Figure 2).

The eight inter-river link canals, some the size of large rivers, have a total length of over 622 km and over 400 principal structures with discharge capacities varying between 116 615 $m^3 s^{-1}$. At full capacity, a total of almost 3000 $m^3 s^{-1}$ could be diverted though these canals to transfer water of three western rivers, namely, Chenab, Jhelum and Indus, to canals formerly dependent on the three eastern rivers, the Sutlej, Beas and Ravi. The regularity of the water supply improved significantly after the commissioning of Mangla Dam and the improvement of the various barrages. The original design irrigation water supply rate was 0.21 $l s^{-1} ha^{-1}$, implying that about 80% of the available agricultural land could be used to grow crops in any year (40% in each season). The improved supplies, augmented by supplementary irrigation from shallow ground water sources, enabled the cropping intensity to be increased to over 130%, with about one-third of the land being double-cropped. However, this resulted in significant salinity and water logging problems.

After partition, a further three barrages were also constructed on the River Indus, the Guddu, Taunsa and Kotri Barrages. The downstream province of Sindh was greatly affected by all the changes upstream. Not only was the

Figure 2 The Indus in the foothills of the mountains before reaching the area of the major dams

Table 1 The annual distribution recommended in 1970 in million m^3

Punjab	64 715
Sindh	54 050
North West Frontier Province	6567
Baluchistan	3333
Total	128 665

quantity and quality of water reaching the irrigation channels of Sindh affected by all the diversions and uses upstream, but also water allocations within the Pakistan sector sometimes appeared to allow less than seemed equitable to be diverted to the Sindh. The Sindh Province complained about such inequity. Table 1 shows the annual distribution that was recommended in 1970.

This did not end arguments about distributions. When suddenly, in the Water Accord of 1991, the quantity of usable water was shown to be 144 415 million m^3 instead of the customarily accepted 128 000 to 130 000 million m^3, accusations were made that the new figure was false and fabricated just to keep the people of Sindh believing, on paper, that they were getting 60 000 million m^3 of water.

WATER SHORTAGES AND PRESSURES ON AGRICULTURAL LAND

Successions of wet and dry years have meant great irregularities in the annual water transfers, especially to Sindh. In some dry years, the total flow in the Pakistan Indus system is as low as 111 000 million m^3 with resulting water shortages. Between 1971 and 1990, only two years are said to have adequate flows to meet all irrigation demands. However, the period from 1991 1997 was wetter, with adequate flows in five of the years. In 1999, conditions were much drier. On November 14, 1999, all non-perennial canals of the Sakro Irrigation Division were closed until further notice due to non-availability of water in the Indus River. No water was available for any of the off-take canals originating from Kinjhar Lake. The water shortage would be met by re-adjustment of flow in the canal network to ensure that water supply schemes and standing crops obtained the quantities of water due to them.

The Sindh Province often suffers losses because of acute water shortages during the sowing months of April, May and June. Further problems arise in the lower reaches of the Indus when, sometimes for up to 10 months of the year, there is hardly any flow of water below the Kotri Barrage. Lack of freshwater outflow leads to a risk of intrusion of brackish water that threatens to damage crops in the Thatta District. Any further reduction in discharges below the Kotri Barrage would lead to ecological problems in the Thatta District, including damage to mangrove forests, shrimp cultivation and the palla fishery which earns much foreign exchange for the country.

Pakistan s population has continued to grow rapidly, rising from around 40 million in 1950 to 140 million in 1995, with a doubling to 280 million expected by 2025. The rate of increase has slowed from 3.7 2.8% per year, but nevertheless there are 4 million more mouths to feed every year. Despite movement of people into the cities, with Karachi and Lahore growing by over 4% every year, the pressures on irrigated agricultural land have grown. In the period 1820 1850 Sindh cultivated about 400 000 ha of land and Punjab about 300 000 ha. By 1995, Sindh was cultivating 3.6 million ha and Punjab almost 6.5 million ha. The cultivated land *per capita*, however, has decreased drastically from 0.3 ha per head in 1950 to 0.13 in 1995. There is thus enormous pressure to increase production by whatever means possible, including increased applications of water, fertilizers and pesticides.

Because the early canals were unlined and the drainage system was insufficient to flush salt off irrigated areas, about half the irrigated land has been increasingly affected by salinity; white sterile areas are visible throughout the plains. New solutions to the salinity problems are being sought and new canals are being dug.

SILTATION AND SALINITY

The glacier meltwaters carry huge quantities of fine sediment, some of which contributes to the gradual siltation of the great reservoirs as at Tarbela. Water released from the reservoirs has relatively low sediment concentrations but picks up sediments from the channel walls and beds, quickly becoming heavily laden with sediment. Silt gets deposited in the distributary canals, reducing their capacity. Sometimes contractors have been able to dredge the

canals, and maintain the original trapezoidal cross-section. But other times, they have just built up the canal banks, creating perched waterways above the land. Failures and floods overtop the banks. One of the major construction activities in maintaining the canal system is to rehabilitate these banks.

Waterlogging and salinity are the principal threats to the sustainability of irrigated agriculture in Pakistan. The environmental costs of the salinity problems in the Upper Indus alone are estimated at US$300 million. However, these salinity problems are more severe downstream along the lower Indus in Sindh Province. Drainage is necessary to combat them. The area requiring drainage, inclusive of areas where projects have been completed or are in progress, is about 7.75 million ha, of which 2.36 million ha are covered under on-going projects, 0.9 million ha are in areas with completed projects which require further rehabilitation, and 5.39 million ha are in areas with no drainage facilities.

The methods of disposal of saline effluent currently used in Pakistan are not environmentally acceptable on a large scale or on a permanent long-term basis. Twenty-four of the 33 million tons of salt brought into the Indus basin annually are retained and recycled within the basin: (i) 2.2 million tons are stored in leaky evaporation ponds in the Cholistan Desert which threaten the Pajnand Abbasia Canal on the lower edge of the Upper Indus; (ii) 2.6 million tons are stored in Pakistan s largest inland freshwater lakes (Manchar and Hammal) on the right bank of the Indus, which are now threatened with extinction; (iii) 0.7 million tons are trapped in the Peshawar Valley; (iv) 18.5 million tons are inadvertently stored in the aquifer through direct seepage from canals, watercourses, and drains; and (v) the rest is disposed into the Arabian Sea via the Lower Indus River itself. In the longer term, the only viable option is disposal of the effluent outside the system by developing a national surface drainage system by extending the left bank outfall drain northwards and constructing trans-Punjab link drains.

However, people in the Sindh fear that such a project might have severe impacts on their environment and livelihoods. Potential risks are seen as a rise in the groundwater table, effects on the ecology of the lower Indus, especially below the outfall of the right bank drain below Sehwan, impacts on the drinking water supplies of the cities of Hyderabad and Karachi and possible slow destruction of the Manchar and Keenjhar lakes by eutrophication.

POWER GENERATION AND COMPETITION FOR WATER

Further environmental concerns have arisen over the multi-purpose Kalabagh Dam Project designed to:

- generate 3600 MW of low cost hydro-electric power;
- replace the live storage capacities of Mangla and Tarbela reservoirs lost due to progressive siltation;
- provide additional storage on the Indus for better irrigation supplies to crops; and
- regulate and control flood peaks in the Indus.

Although this project is actively being considered for implementation with the consensus of all provinces, some argue that the Kalabagh dam will not only render large areas close to the dam infertile but will also accelerate soil degradation throughout the newly irrigated areas. If Kalabagh gets built, even more resources for drainage works than the huge amounts already envisaged under the 25-year national drainage plan will be required. The dam will deprive the lower Indus delta of the fertile silt and freshwater required to nourish its land and forests. Already, the silt volume has declined from 200 million t year^{-1} in 1947 to 36 million t year^{-1} in 1995.

IMPACTS OF CLIMATE CHANGE

Climatic change is most likely to affect the Indus irrigation system through increased variation of the monsoons. Doubling of carbon dioxide in the atmosphere, according to some models, could increase average summer monsoon rains by up to 60% (Second Asia Pacific Seminar, 1993). The resulting floods would hit the densely populated areas that produce most of the food, fiber and fodder in the country; they would destroy irrigation systems and crops, especially cotton, which is the main cash crop of Pakistan. Erosion and landslides from the aggravated rains in the mountains would accelerate reservoir sedimentation. However, another frequently used model (Watson *et al.*, 1997), by contrast, foresees summer monsoon rains decreasing, placing heavy stress on winter wheat, the main food staple. In either case, changes in agriculture and redesign of irrigation will be essential.

Although Pakistan is less vulnerable to sea level rise than some of its South Asia neighbors, the Indus delta, south of Karachi, already retreating because of a sharply reduced silt load, could lose up to 25% of its area if seas rise. Saline water would intrude and mangroves would be destroyed, harming fisheries, which account for 30% of exports. Coastal resources will need more protection.

Pakistan s forests now cover less than 5% of the nation s land. Small shifts in average temperature would lead to large shifts in latitude of optimal growing zones for many tree species. Forest resources will need to be preserved and enhanced for their intrinsic timber and biodiversity values, as well as their hydrologic importance. Integration of forest management with other activities, especially agriculture and irrigation, will be important for a sustainable future.

Climate change may shift the balance of cost effectiveness toward investment in the revival of riverine forests with consequential implications for irrigation.

The Indus River thus illustrates the complexities and conflicts in a basin where the great majority of people depend on irrigation. Competition for water, not only at the international level, but also between provinces and between uses (irrigation, power, and urban supplies) aggravates the management problems. Environmental responses to changes in water flows, such as waterlogging through over-irrigation, and soil degradation through salinity, require remedies that lead to new or additional environmental impacts. Saline drainage waters carrying agricultural chemical residues lead to the eutrophication of lakes. Raised embankments along partly silted irrigation channels produce greater seepage into groundwater aquifers, which produce higher water tables, and thus more waterlogging. The challenge in managing such a system is to try to reduce the problems at source and not to merely transfer the problems, silt or salt, farther down the basin and merely substitute environmental impact in one area for environmental impact in another. In the context of the rising Pakistani and Punjabi populations, the challenge is difficult, but the combination of good understanding of the dynamics of the hydraulic and ecological systems and engineering skill has the potential to create a much more sustainable irrigation system.

See also: Salinization, Volume 2; Salinity and Agriculture, Volume 3; Waterlogging, Volume 3.

REFERENCES

Second Asia Pacific Seminar (1993) *Report of the Second Asia Paci c Seminar on Climate Change*, 29 31 March 1993, Bangkok, Thailand, http://www.eic.or.jp/eanet/en/pol/apscc/2nd/.

Watson, R T, Zinyowera, M C, and Moss, R H (1997) *Summary for Policymakers, the Regional Impacts of Climate Change: an Assessment of Vulnerability*, a Special Report of IPCC Working Group II Intergovernmental Panel on Climate Change, http://www.ipcc.ch/pub/sr97.htm#tropasia.

FURTHER READING

http://sangat.org/creed, Citizens Alliance in Reforms for Efficient and Equitable Development (CREED).

http://www.ilri.nl/icid/journal.html.

Industrial Agriculture

see Green Revolution (**Volume 3**)

Insect Pests and Global Environmental Change

Richard Harrington
IACR, Rothamsted, UK

Insect pests are detrimental to mankind through direct competition for food resources and through transmission of diseases to crops, cattle and humans. As such, they cause huge losses of life, especially in poorer countries, and huge economic losses everywhere. Pest status is affected to some degree by virtually every aspect of the environment and thus the effects of environmental change on the impact of pests are of great relevance worldwide.

Whilst certain vertebrates, especially rodents, are important pests (e.g., in Indonesia the rice eld rat causes 17% pre-harvest yield losses, costing US$ one billion per year), invertebrates, and insects in particular, are predominant in terms of their impact worldwide and are the sole focus of this contribution. The label 'pest' cannot necessarily be applied to an insect throughout its range. Whether it is a pest will depend on whether it coincides in time and space with a host, which is of value to humans, and whether, in the absence of arti cial control, it is able to reach population levels that cause economic damage.

WHAT ASPECTS OF ENVIRONMENTAL CHANGE ARE LIKELY TO AFFECT PEST STATUS?

The potential rate of increase of all invertebrates is strongly dependent on temperature, and their survival is impaired at low and high temperatures. Changes in both mean temperature and the value and frequency of extremes can hence have a major impact on the status of pests. Rainfall is also especially critical to survival. Some pests are unable to tolerate extremes of drought, whilst others are disadvantaged by extremes of wetness. Other meteorological variables such as wind speed and direction may influence pest status, but effects are likely to be minor compared to temperature and rainfall. Increased concentrations of carbon dioxide (CO_2) increase carbon nitrogen ratios of plants. This leads to the need for crop pests to eat more in order to obtain adequate dietary nitrogen. This may be compensated by increases in plant biomass or levels of carbon-based defenses (Bezemer and Jones, 1998; Cannon, 1998). It is important to bear in mind the likelihood that responses to changes in any given variable will interact with responses to changes in other variables, not only meteorological (Figure 1). This may produce outcomes that are counterintuitive or synergistic to those expected when considering one variable in isolation. For example, economic

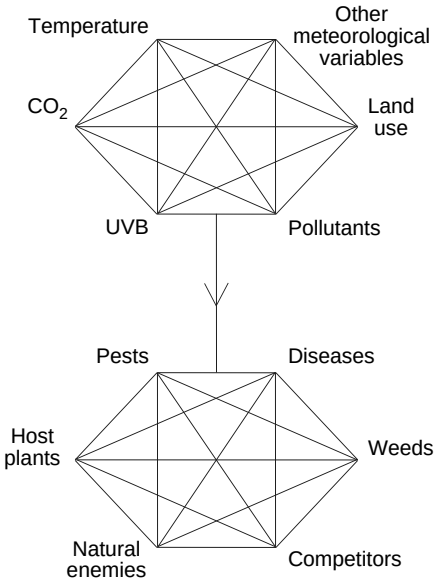


Figure 1 Interactions within and between abiotic and biotic components of climate change

and other considerations may lead to changes in land use and agronomic practice, which override any effect of climate on pest status. Unfortunately, little experimental work has been done to examine the interacting effects of multiple climatic or other environmental variables (Lawton, 1995). For this reason, it is useful to examine long-term datasets on pests, as any trends or relationships associated with environmental variables will occur despite all the interactions taking place in the system. There are few such long-term datasets, but a good example (The Rothamsted Insect Survey) is given later.

POTENTIAL EFFECTS OF CLIMATE WARMING

For insects currently limited by low temperature, some intuitive potential effects of a warmer climate are listed below. The realization, or otherwise, of this potential, will depend on the many complex interactions within and between the biotic and abiotic components of change alluded to above, and on adaptations in the species concerned. For species currently limited by high temperature, the intuitive effects of warmer conditions will often be the reverse of those outlined below.

Increase in Rates of Development and Reproduction

Rates of development and reproduction increase between the developmental threshold and an optimum temperature. In temperate regions, where insects generally experience conditions below that optimum for most of the year, increases in temperature generally increase developmental and reproductive potential. Generation times become shorter, and this often results in more generations in a year.

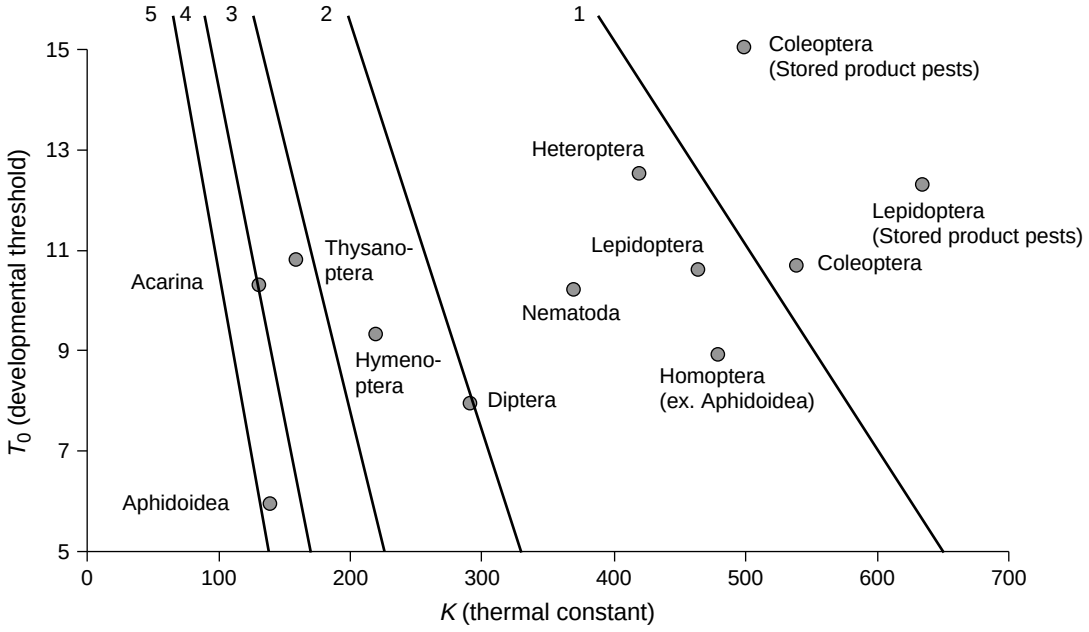


Figure 2 Number of extra generations likely for various insect taxa under a warming of 2 °C. The development threshold is the temperature (deg C) below which no development occurs. The thermal constant is the accumulated number of degree days above that threshold required for the insect to complete a generation (e.g., egg hatch to egg hatch). The lines 1 to 5 show the number of additional generations that each insect group shown is expected to get through in a year with 2 deg warming. (Reproduced with permission from Yamamura and Kiritani, 1998)

Figure 2 shows the expected increase in annual generation number for a range of insect taxa. However, some species are constrained, through responses to other stimuli, to a set number of generations each year, and changes in generation time must be compensated by increases in the duration of a dormancy period.

Extension of the Active Season

Warmer conditions may extend the period during which temperatures are above the threshold for development, reproduction and activity. This may advance the timing of phenological events such as the start of a migration or the progress to the next developmental stage, leading to an earlier completion of a particular activity or developmental cycle. Alternatively it may facilitate increased development or activity at both ends of their normal period of occurrence. The responses of pests will again depend on interactions with other stimuli such as day length. For example, insects often enter diapause in response to photoperiodic cues. Unless the duration of food availability is also extended, pests will be unable to benefit from the increased potential given by changes in temperature alone.

Overwintering Success

Winter mortality as a result of low temperature is a key factor in the dynamics of many temperate pests, especially those that do not enter diapause but continue activity throughout the winter as temperature permits. Warmer winters, or reduction in the frequency of extremely cold periods, may thus improve the prospects for survival of such species. However, species which are quiescent in winter may be disadvantaged by periods of unusual warmth, as activity increases, possibly in the absence of food, and depletes vital energy reserves required for survival until adequate food supplies are available.

Extension of Range

Warmer conditions may be expected to promote the poleward extension of the range of species currently limited by low temperature, or the altitude at which they can survive. In temperate regions, a 1 °C increase in temperature is equivalent to a shift of approximately 200 km in latitude or 140 m in elevation (Parry *et al.*, 1989). At predicted rates of temperature increase, this could mean a potential range extension of 4 km per year in latitude or 3 m in elevation. Whether this occurs will depend partly on the availability of food sources in the new location. Generally, insects have the ability to extend their range far more quickly than do plants, and many pest insects easily have the potential to track climate change. Many plants are unlikely to be able to track climate change, although in the case of crops, if growing conditions are suitable they can, of course, be moved by humans. International trade also offers the opportunity for species to spread. It is not necessarily the case that if a species, which is a pest in one area, arrives in a new location, it will be a problem there as well. The reverse is also true. A species that is not a pest in one location may become one if it establishes, e.g., in an area relatively free of natural enemies or where hosts have not evolved adequate defensive strategies.

In the case of vector-borne diseases, the survival of the pathogen as well as the vector may be limited by temperature. For example, malaria can generally occur where the winter isotherm is around 16 °C as, although the vector can survive lower temperatures, the parasite does not develop below this temperature (Patz and Martens, 1996). Higher temperatures also increase mosquito biting rates and shorten the incubation period of the pathogen. The altitude at which this disease becomes endemic, and other diseases of humans (such as dengue and yellow fever) or of cattle (such as east coast fever) is likely to increase as temperatures increase. Malaria has been observed in unusually warm conditions at high elevations in Africa where it is not endemic, and dengue reached an altitude of 1700 m in Mexico during an unusually warm summer in 1998.

POTENTIAL EFFECTS OF CHANGES IN RAINFALL

In the case of the diseases mentioned above, the larvae of the vectors generally require warm, standing water, and particular problems develop when conditions are both abnormally warm and wet. In East Africa, such conditions may occur in strong El Nino years, and thus if these can be predicted, there is the possibility of taking appropriate steps to control the diseases.

Under conditions of drought stress or ozone episodes, there is a tendency for increased susceptibility of hosts, whether plant, animal or human, to pests and the diseases they transmit. However, this is not always the case and the impact will be determined by the balance of advantages and disadvantages conferred on the pest. In the case of plant hosts, phloem sap becomes more concentrated, leading to it acting as a richer source of amino acids for sap feeders. However, the concentration of a range of secondary plant compounds tends to increase under drought stress, leading to changes in the attraction of plants to a pest. The physical nature of the host plant may also change. Leaves may become thicker or waxier, or may change color, which could make them less palatable or attractive to pests.

Although drought stress often leads to increased susceptibility of hosts to pests, it may also alter relationships between the pests and their pathogens. For example, the spread of fungal pathogens of insects is often impaired by dry conditions, again favoring the pest. Heavy rain may also be detrimental to pests through direct effects on dislodging, but if the pest is able to recolonize its host, dislodging may result in increased disease spread.

CHANGE IN INTERACTIONS WITHIN OR BETWEEN TROPHIC LEVELS

Changes in climate may affect the rates of development of pests and their natural enemies differently, leading to variations in control potential. For example, it has been shown that ladybirds tend to control aphids better at higher temperatures as the consumption rate of aphids by ladybirds increases more rapidly with temperature than does the reproductive rate of aphids.

Changes in climate may affect interactions between species within or between trophic levels by disrupting synchrony. For example, many pests are dependent on synchrony between time of bud burst (or flowering) and the emergence of feeding stages. Surprisingly, there is little evidence of feeding stages being cued directly by bud burst. Instead, similar external cues often affect both bud burst and the appearance of feeding stages. However, the requirements are often complex combinations of day length and specific combinations of periods and extents of chilling and warming. Requirements are rarely identical for the plant and pest, and have developed over evolutionary time. It is quite conceivable that under the current rapid changes in climate being experienced, synchrony between trophic levels could become decoupled as a result of subtle differences in the impacts of changes on the timing of cues determining the phenology of constituent species (Harrington *et al.*, 1999).

Spatial synchrony may also be compromised by climate change. It is known from studying the fossil record that spatial associations between species are generally transient. This is likely to be due to the improbability of any two species sharing an identical set of environmental requirements and responses to given combinations of biotic and abiotic factors. Thus, as these factors become dissociated, so do species assemblages. The potential for such dissociation has been shown in laboratory studies on fruit flies. Flies were reared in a series of linked incubators in order to provide artificial temperature clines. The outcome of competition between three species of fruit fly depended on the temperature range and whether or not a parasitoid species was present. Temperature determined the outcome of the interaction between the fruit flies and their parasitoids, and different species of fruit fly dominated or became extinct accordingly (Davis *et al.*, 1998).

ADAPTATION

The extent to which pests are able to adapt to environmental change will depend on the genetic diversity of the population, the distribution of tolerant phenotypes, migratory tendency and competition. Genetically diverse populations are more likely than homogenous ones to include individuals better adapted to new conditions than the bulk of the population. In such cases, it may be expected that the better adapted genotypes will gradually displace those which are less tolerant of change. Adaptation will also be aided by mobility, as the more mobile a species, the greater the probability of suitably adapted founder individuals arriving from elsewhere.

OBSERVATIONS FROM A LONG-TERM DATASET

Because of the complex interactions that must be considered when assessing the potential impacts of environmental change, combined with interspecific, intraspecific and even intraclonal variability in responses, attempts at generalization will be largely unproductive. Process-based modeling can help to demonstrate the combined effects of some of these interactions, but it is unrealistic to expect to be able to parameterize all the relevant interactions in any system. Long-term datasets can be useful in demonstrating impacts of environmental change, and in validating predictions. However, predictions based on long-term datasets suffer from the inevitability that the range of values of relevant variables at the time for which a prediction is being made, will fall outside the multidimensional cluster of values experienced during the construction of the relationships used for that prediction. Such extrapolation may lead to erroneous predictions.

Some trends that may be due to environmental changes are apparent in a comprehensive set of data on aphids (Woiwod and Harrington, 1994). This group of insects may be expected to be amongst the most sensitive indicators of temperature change because of their low developmental thresholds and short generation times (Figure 2). They are predominantly pests of temperate crops and, as such, tend to be limited more by low than by high temperature. Expected increases in temperature are, therefore, intuitively likely to benefit these insects.

The Rothamsted Insect Survey has been operating a network of suction traps in the UK since 1964 in order to monitor flying aphids. The traps are emptied daily and all aphids identified. Similar traps operate throughout much of Europe

Figure 3 Location of suction traps for aphid monitoring in Europe. Seventy-five sites in 19 countries; shaded countries are EU members

(Figure 3), providing what is probably the most extensive, standardized dataset for any terrestrial invertebrate group. For many species at many sites, there is a remarkably strong relationship between the timing of the start of the spring migration and winter temperature (Figure 4). The milder the winter, the earlier the migration. This is probably due to a combination of more rapid development and reproduction and reduced mortality due to low temperatures. This conclusion is backed up by the fact that aphids which pass the winter as an egg tend to show very much weaker relationships between migration time and winter temperature than those that pass the winter in the active stages. Eggs are very much more hardy than the active stages and are hence less affected by low temperature. However, they are in diapause and, unlike the active stages, cannot take advantage of warm periods with a burst of development and reproduction. Thus, biological expectation shows through in these purely statistical relationships. The clear implication from this is that milder winters will result in earlier aphid migrations and hence in earlier threat to crops. The extent of change in economic damage will depend partly on the extent of change in the planting and emergence times of crops. As an approximation, the aphid species examined show an advance in the timing of the spring migration of about fourteen days for every degree centigrade rise in temperature. A rise in numbers trapped during the critical spring and early summer period is also found. At this time, crops tend to be particularly sensitive to damage caused by aphids through removal of plant sap and through transmission of plant viruses.

Plant viruses add another layer of complexity to the potential impact of climate change on problems caused by aphids. The extent of the complication can be understood by considering the example of barley yellow dwarf virus

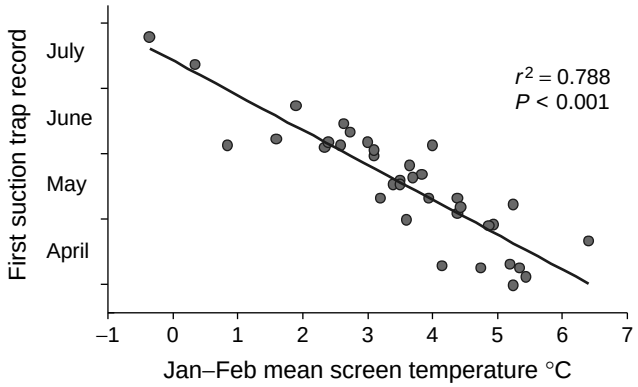


Figure 4 Relationship between winter temperature and the time of the start of the migration of the peach-potato aphid at Rothamsted, UK

(BYDV) in the UK. The virus is important worldwide and has a range of grass and cereal hosts and a range of strains, which are transmitted with different efficiencies by different vector aphid species. One strain is particularly common in maize and is transmitted efficiently by the corn leaf aphid. Maize is a minor crop in the UK but with climate warming, may become more viable and hence more widespread. The corn leaf aphid is also rare in most parts of the UK in most years, but the expansion of maize growing will encourage spread of the aphid. The aphid has no hardy egg stage and an increased frequency of warm winters will further aid its survival. At higher temperatures, other aphids become capable of transmitting the maize strain of BYDV to cereals such as wheat and barley. Therefore, warmer conditions, through their effect on interactions between host plants, aphids and viruses may render the maize strain of BYDV important in wheat and barley for the first time in the UK.

The long-term nature of the aphid dataset makes it possible to look for trends with time that may be attributable to climate change, as well as trends with meteorological variables themselves. In an analysis involving five species in the UK, there is evidence for an advance in phenology of 3 6 days over a 25-year period during which the temperature rose by 0.4 °C (Fleming and Tatchell, 1995).

RESPONSE OPTIONS

Whilst there is good reason to expect pest problems to increase, there is a danger that over-emphasizing the risks will lead to a false perception of current problems and an increased prophylactic use of pesticides. Control programs which include monitoring and take account of ecological and economic factors in determining the need for using chemicals are crucial to prevention of a vicious circle of increased reliance on pesticides and increasing resistance to those pesticides. Pest management decisions also need to be integrated into the broader context of total resource management. For instance, pesticides may reduce food sources for birds and other animals of conservation interest.

It is also important to realize that the efficacy of control methods may change as environmental conditions change. It is well known that the persistence of many conventional pesticides is very reliant on temperature, rainfall and soil moisture content. Newer technologies are not immune from such considerations. For example, cotton that has been genetically modified to include a bacterial gene that produces an insect toxin is more efficacious when grown at higher temperatures. No doubt examples could be found where the reverse is true, and it cannot be assumed that any given control method will be equally effective under the environmental conditions of the future.

REFERENCES

Bezemer, T M and Jones, T H (1998) Plant-insect Herbivore Interactions in Elevated Atmospheric CO_2: Quantitative Analyses and Guild Effects, *Oikos*, 82, 212 222.

Cannon, R J C (1998) The Implications of Predicted Climate Change for Insect Pests in the UK, with Emphasis on Non-Indigenous Species, *Glob. Change Biol.*, 4, 785 796.

Davis, A J, Lawton, J H, Shorrocks, B, and Jenkinson, L S (1998) Individualistic Species Responses Invalidate Simple Physiological Models of Community Dynamics under Global Environmental Change, *J. Anim. Ecol.*, 67, 600 612.

Fleming, R A and Tatchell, G M (1995) Shifts in the Flight Responses of British Aphids: a Response to Climate Warming? in *Insects in a Changing Environment*, eds R Harrington and N E Stork, Academic Press, London, 505 508.

Harrington, R, Woiwod, I, and Sparks, T (1999) Climate Change and Trophic Interactions, *Trends Ecol. Evol.*, 14, 146 150.

Lawton, J H (1995) The Response of Insects to Environmental Change, in *Insects in a Changing Environment*, eds R Harrington and N E Stork, Academic Press, London, 3 26.

Parry, M L, Carter, T R, and Porter, J H (1989) The Greenhouse Effect and the Future of UK Agriculture, *J. R. Agric. Soc.*, 150, 120 121.

Patz, J A and Martens, W J M (1996) Climate Impacts on Vector-borne Disease Transmission: Global and Site-Specific Analyses, *J. Epidemiol.*, 6, S145 S148.

Woiwod, I P and Harrington, R (1994) Flying in the Face of Change The Rothamsted Insect Survey, in *Long Term Research in Agricultural and Ecological Sciences*, eds R A Leigh and A E Johnston, CABI, Wallingford, 321 342.

Yamamura, K and Kiritani, K (1998) A Simple Method to Estimate the Potential Increase in the Number of Generations Under Global Warming in Temperate Zones, *Appl. Entomol. Zool.*, 33, 289 298.

Inter-basin Transfer (IBT) for Water Supplies

Geoff Petts
The University of Birmingham, Edgbaston, UK

Historically, water availability constrained the location and size of settlements. Across the globe, population growth and urban development in drylands is placing considerable strains on water resources requiring water transfers from relatively humid areas, often supported by water-storage reservoirs. Even within relatively humid regions there are demands for water transfers to meet the demands of urban populations. Today, desert cities in the USA such as Phoenix, AZ, and Las Vegas, NV, and Adelaide in Australia depend upon the importation of water from distant regions, while in many areas of the world, modern industrialized agriculture is heavily reliant upon IBT schemes. IBTs have become a feature of regional water supply infrastructure and are a key component of river regulation schemes discussed elsewhere in this volume.

IBTs have been de ned as (Davies et al.*, 1992): the transfer of water from one geographically distinct river catchment or basin to another or from one river reach to another. Impacts include arti cial ow levels in the donor (reduced ows) and recipient (augmented ows) streams, a change of water quality in the recipient stream, and the breakdown of biogeographical barriers to the dispersal of animals, plants and their associated diseases. The manipulation of river ecosystems through IBTs is occurring on a global scale and the conservation of riverine biodiversity is being severely compromised (Davies* et al.*, 2000).*

A HISTORY OF IBT

Major river diversions have been attempted from at least the 17th century. Many were built for navigation, requiring reservoirs to maintain water levels in summit sections, or for flood protection. One of the most ambitious attempts was the 36 km-long navigation canal to link Lake Neuenberg and Lake Geneva in Europe (Vischer, 1989). This was notable in that it would have crossed the great European water divide between the catchments of the Rhine and Rhone; 46 locks were required from the summit to Lake Geneva, 80 m below. Begun in 1638, the project ran out of funds and was never completed.

TRANSFERS IN THE UK

Early transfers of water utilized the simple observation that water could be transferred easily under gravity along relatively low gradient channels cut across the contours of hillslopes, over a watershed to a town in an adjacent river basin. In the UK, the first water-supply reservoir was constructed at Whinall in Scotland to supply the rapidly urbanizing center, Greenock. Over the next 40 years, earth dams over 15 m high were built in England as part of the increasingly sophisticated water systems to power the industrial revolution and to provide water supplies to such industrialized areas within the southern Pennines (Figure 1). Transfers from the catchwater basins left their streams dry during periods of low flow. In most cases, the transferred waters were used within an urban area before being discharged as untreated wastewaters into the receiving streams. Thus, between 1848 1884 a chain of five separate reservoirs were constructed to exploit the resources of the river Etherow. This Longdendale Scheme supplied water by gravity to Manchester. The dams command a catchment area of 78 km^2 and have a combined capacity of about 15.5 million m^3. However, even at this early time, storage in two of the reservoirs was allocated exclusively to maintain a compensation flow in the Etherow to meet the needs of mill owners downstream.

Larger schemes of water transfer were advanced towards the end of the 19th century. Parliamentary Bills enabled the construction of large reservoirs in the Vyrnwy and upper Wye valleys of Wales to supply Liverpool (1880) and Birmingham (1892), respectively. Both schemes involved dams over 40 m high and water transfers of more than 100 km. Vyrnwy, with a capacity of nearly 60 million m^3, was the largest reservoir in Europe at the time. Later, in 1919, proposals were advanced to transfer water from Haweswater in the English Lake District to Manchester. Today, of the 1310 streamflow gauging stations in the UK, 40% are significantly affected by IBTs for water supply, either by abstraction from a reservoir or river intake (26%) or by effluent return via sewage works or industrial works (27%).

At the regional scale, supply and demand are inversely correlated. West east contrasts in the UK are dramatic: rainfall is above 750 mm in the west, rising to over 1500 mm in the headwater catchments, and below 750 mm in the east, falling to about 500 mm in some areas; population densities are about 250 persons km^{-1} in the west rising to nearly 500 in the east; and forecasts of the public water supply deficit for 2021 range from a continuing surplus in the west to a deficit of over 250 megaliters per day (Mld) in the east under a high-use scenario. West-to-east water transfers offer a solution to any water deficit in eastern England (Figure 1). The River Severn is a typical donor river being heavily regulated, by reservoirs in mid-Wales and by the Shropshire Groundwater Scheme. The river supports abstractions of 2300 Mld approximately 500 Mld for the city of Birmingham. The River Trent receives over

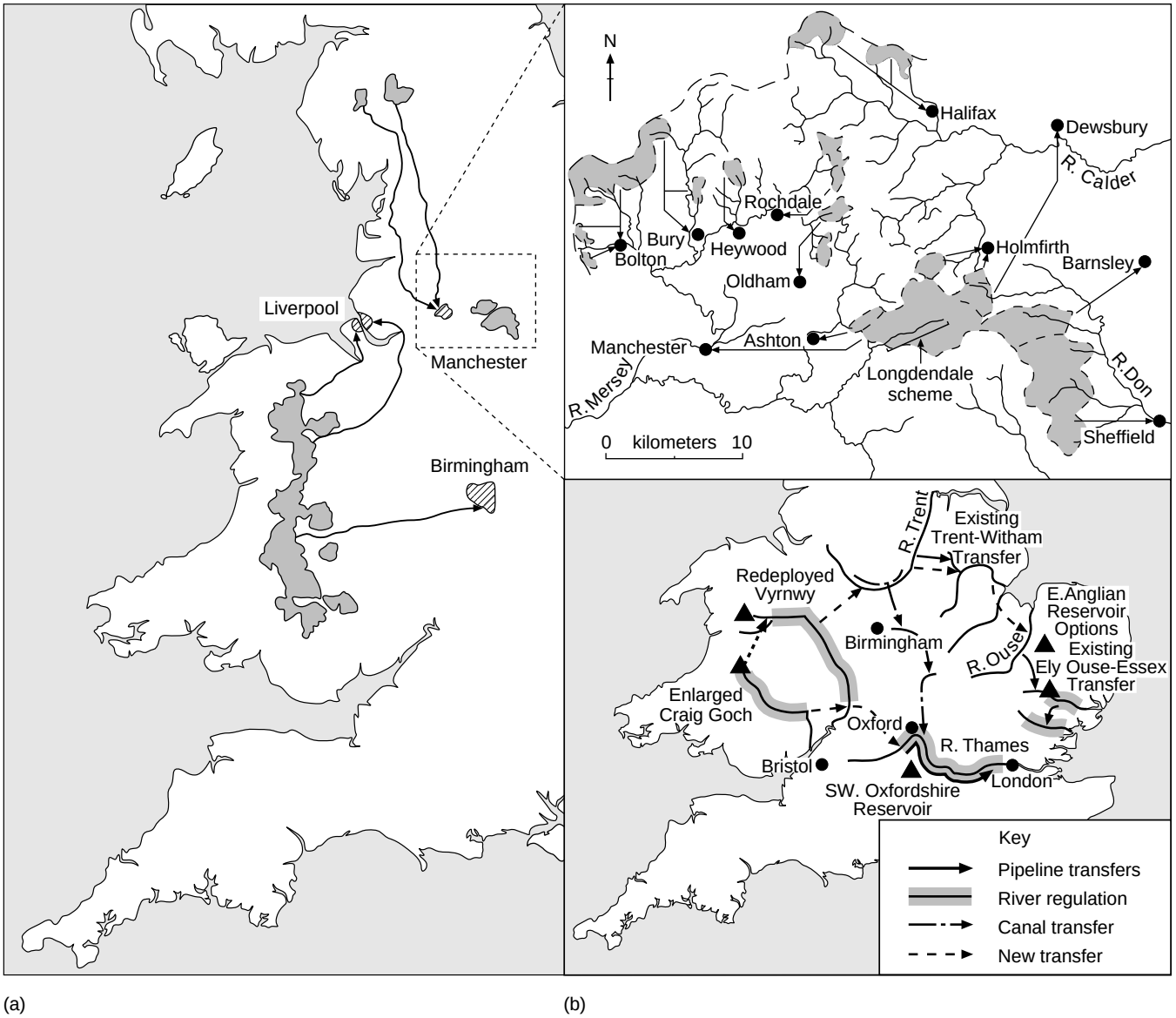


Figure 1 (a) IBTs to supply Central and Eastern England: 19th century catchwater schemes and associated transfers in the southern Pennines and longer-distance schemes to supply Manchester and Birmingham; and (b) possible future scenarios to meet the needs of the growing population in South-east England

500 Mld of effluent from Birmingham s two major water recovery works, and this has markedly altered the pattern of flows in the river, especially during dry summers when artificial inputs represent up to 130% of the natural flow! Following the improvements in water quality, one solution to the deficits projected for Eastern England is to further exploit the enhanced flows in the River Trent through new transfers.

GLOBAL EXAMPLES

The relatively small-scale IBTs that have been developed to counteract regional inequalities of water resources are mirrored by much larger schemes, both existing and proposed, worldwide (Golubev and Biswas, 1985; Petts, 1994). The exploitation of catchwaters to increase the water yield to a storage reservoir is widespread practice, especially for hydro-electric power generation. In Switzerland, the Grande Dixence Scheme, completed in 1965, links intakes on mountain and proglacial streams by 90 km of tunnels to the Lac des Dix, thereby increasing the reservoirs catchment area to 357 km^2 and its water yield to about 420 million m^3. La Grande Riviére hydroelectric complex in Quebec provides a second example: its catchment has been increased by 43% by IBTs from adjacent basins.

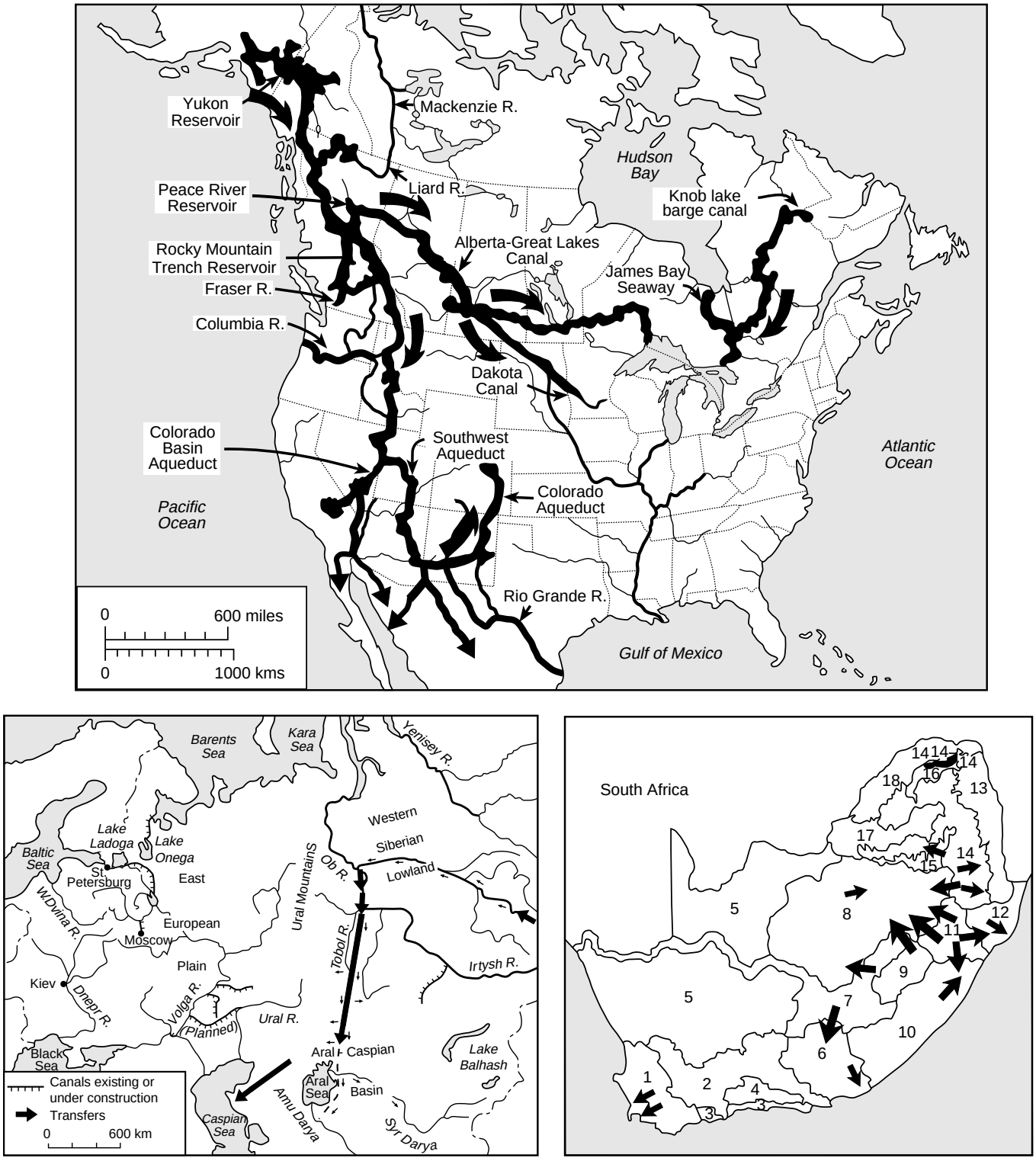


Figure 2 Examples of IBTs plans for North America and Siberia-Central Asia, and inter-eco-region transfers in South Africa (eco-regions are labeled 1–18). (Reproduced from Davies *et al.*, 2000)

Schemes to meet regional water deficits include California s State water project built between 1963 and 1973, which diverts flow from better-watered northern areas to the drier central and southern parts of the state. Again the scheme involves water storage reservoirs and river regulation in the north, and a 715 km aqueduct and pumping stations to redistribute the water. Plans for an even larger scale, long distance IBTs have been proposed. One of them was the North American Water and Power Alliance (NAWAPA) scheme (Figure 2), put forward in the 1960s as a comprehensive solution to North America s water supply problems. The project was designed to transfer 136 km^3 of water each year and would have required 240 reservoirs the largest impounding 693 km^3 of water and the highest dam would have been 543 m high! However, it is now highly criticized, especially in Canada and is unlikely to proceed.

A similar gigantic solution to regional water shortages has been proposed for Central Asia (Figure 2). The expansion of irrigation agriculture and population growth in the Aral Sea basin has led to a severe water deficit, manifest by a 20 m fall in sea level since 1960 causing the shoreline to recede by over 100 km in places (*see* Aral Sea, Volume 4). Large-scale, long distance water transfers have been considered since the 1930s as a solution to water deficiency in the region. The Siberian Rivers Diversion Project of the 1970s proposed a north-south, 2500 km-long, transfer of up to 200 km^3 of water each year. At the present time, such large-scale proposals have not been developed.

ECOLOGICAL IMPACTS

River basins are often seen as the fundamental units of landscape, the watersheds that divide basins effectively separate one river system from another. Within a catchment, water flow from headwaters to mouth with declining altitude creates a succession of in-stream and riverine habitats associated with a longitudinal sequence of animals and plants. The transfer of water between basins can have a range of impacts upon both the donor and receiving rivers, not least by disrupting the natural pattern of ecological change (Davies *et al.*, 1992; 2000).

Abstractions from donor streams can cause these rivers to become dry, especially during summer periods. Thus, the La Grande Scheme in Quebec caused about 1000 km of donor river to experience reduced flows. It was estimated that the proposed Siberian Diversion Project would have reduced levels along the River Ob by between 1.5 and 2.5 m, reducing the area of floodplain and severely impacting fish production which is dependent upon floodplain habitats for spawning and feeding for juveniles, and extending the duration of the winter freeze in the lower river and Ob Gulf. On other donor rivers, flows are augmented by reservoir releases, often eliminating important marginal habitats, such as water-water shallows for fish fry or mud banks for wading birds. Similar impacts can arise in the recipient streams, but flow releases can be pulsed and seasonal, causing additional stresses on biota. A suite of water quality and physical changes are coupled to the releases, including changes in nutrient concentrations and altered thermal characteristics.

Physical processes drive ecological processes across a range of spatial scales from the region to the basin, the reach and the habitat patch. IBTs break down the barriers between regions, basins and reaches, and water transfers can cause the passive transport of animals and plants across natural divides.

Large-scale transfers can cross divides between eco-regions. Climate, geology and landscape history control the physical processes that drive ecological processes and the patterns of biological productivity and diversity. Consequently, river basins can be grouped geographically into eco-regions, each with distinctive ecological characteristics (Omernik and Bailey, 1991). Thus, in South Africa 18 aquatic eco-regions have been defined using biotic distributions. IBTs have crossed divides between eco-regions especially in the east (Figure 2) and particularly involving four donor eco-regions (Orange, Montane, Tugela and Northern Uplands) and four receiving eco-regions dominated by the Vaal, but also including the Drought Corridor, Eastern Seaboard and St Lucia Complex. In Australia, there are 16 IBTs that cross eco-region boundaries and in the USA, 111 IBTs cross water-resource sub-regions (Davies *et al.*, 2000).

The loss of species through altered flows in the donor and recipient streams, the mixing of gene pools, the risks associated with the transfer of non-indigenous species, disease vectors, parasites and invasive species, are significant but as yet there is little information available. In one example, from South Africa, fish species passed through an 82 km-long tunnel between the Orange River and the Great Fish River where they successfully established themselves (Cambray and Jubb, 1977).

Since the 1970s, the rising concern for the environment and the realization of the huge cost of large-scale IBTs has dampened enthusiasm for them. Nevertheless, the looming water crisis, with many areas of the globe expected to be suffering water stress by the middle of this century, may lead to renewed enthusiasm for IBTs. Carefully planned operational rules governing the timing and volumes of abstractions and releases, and rate of change of flows can mitigate many ecological impacts but, as yet, there appear to be no technologies available to prevent biological transfers. It has been argued (Davies *et al.*, 2000) that biological transfers through IBTs will not only severely impact the speciation process

but also evolution itself; an impact, they suggest, that may be as serious as the release of genetically modified organisms!

Analyses of the possible effects of climate change on water supply systems suggest that most systems could be highly vulnerable to changes in climate inputs and that a small change in inputs could lead to large changes in system performance. The magnitudes of such changes vary greatly according to the scenarios used. Isolated single reservoir systems, including those dependent on small-scale IBTs, such as the 19th century works in the English Pennines, could be the most vulnerable. Larger, integrated schemes, using a variety of sources, including major IBTs, would be more resilient, especially if water demand management became more effective.

REFERENCES

Cambray, J A and Jubb, R A (1977) Dispersal of Fish Via the Orange-Fish tunnel, South Africa, *J. Limnol. Soc. SA*, 3, 33 35.

Davies, B R, Thoms, M C, and Meador, M (1992) The Ecological Impacts of Inter-basin Water Transfers, and their Threats to River Basin Integrity and Conservation, *Aquat. Conserv.: Mar. Freshwat. Ecosyst.*, 2, 325 349.

Davies, B R, Snaddon, C D, Wishart, M, Thoms, M, and Meador, M (2000) A Biogeographical Approach to Inter-basin Water Transfers: Implications for River Conservation, in *Global Perspectives on River Conservation: Science, Policy and Practice*, eds P J Boon, B R Davies, and G E Petts, John Wiley & Sons, Chichester, 427 440.

Golubev, G N and Biswas, A K, eds (1985) *Large-scale Water Transfers: Emerging Environmental and Social Experiences*, Tycooly, Oxford.

Omernik, J M and Bailey, R G (1991) Distinguishing between Watersheds and Eco-regions, *J. Am. Wat. Resour. Assoc*., 33, 935 949.

Petts, G E (1994) Large-scale River Regulation, *The Changing Global Environment*, ed N Roberts, Blackwell, Oxford, 262 284.

Vischer, D (1989) Impact of 18th and 19th Century River Training Works: Three Case Studies From Switzerland, in *Historical Change of Large Alluvial Rivers: Western Europe*, eds G E Petts, H Moller, and A L Roux, John Wiley & Sons, Chichester, 19 40.

IPCC (Intergovernmental Panel on Climate Change)

see IPCC (Intergovernmental Panel on Climate Change) (**Volume 1**)

International Rice Research Institute (IRRI)

R D Hill
The University of Hong Kong, Hong Kong

Founded in 1960, the Institute is located at Los Baños, south of Manila, Philippines. It continues to play a major role in researching technical aspects of wet rice production and management.

IRRI has played a notable role in raising rice crop yields at periods when Asian populations, especially, were growing rapidly and where the prospects for increasing production to meet basic food needs by expanding the cultivated area were limited by lack of suitable land and by high costs of developing the little that there was.

The Institute was founded in 1960 at the initiative of the Ford and Rockefeller Foundations whose directors were well aware of the economic wisdom that foresaw burgeoning populations and a potentially serious food supply crunch with rice yields stagnating at around 1.5 tons ha^{-1}, except in Japan and other temperate-country producers. They were equally aware of the possibility of national strife/revolutions and their appeal to hungry people; and they saw green revolutions as an alternative. Six years after the Rockefeller Foundation s George Harrar and Warren Weaver proposed a rice research institute, it became a reality, occupying its present site at Los Banos, near Manila. Its initial success in rice breeding and management led to support from governments internationally, whose representatives sit on the corporate Board of Directors responsible for overall policy and control.

The need for rapid expansion of production at first led to decisions to concentrate upon the relatively easy-to-solve problems of raising yields by improved water control, controlling pests, and enhancing fertilizer response by designing a series of high-yielding varieties that were to change the very architecture of the plant. Later work aimed to enhance grain quality without sacrificing yields, still with the aim of matching environment to the plants as much as the plant to the environment (some would say more). Consequently some of the more intractable problems have received less attention. Improving drought resistance is inherently difficult in rice, for it is by origin a swamp plant. Varieties grown on flat land fed only by rainfall and trapped water have been difficult to improve, even though globally, most rice-farmers occupy this environment. IRRI has not been notably successful in developing cold-tolerant varieties, in which field Japanese breeders excel, or in

breeding types that give good yields where grown in brackish water, an area in which Chinese breeders have succeeded. One persisting criticism of IRRI s approach is that some of its improved varieties require rather precise conditions to achieve good results, with very poor yields should those requirements not be met, whereas traditional varieties, though low-yielding, can cope better with the vagaries of the weather.

Further criticisms are that improved rices mainly benefit relatively wealthy farmers, who alone can meet the greater cost of inputs, and that the benefits of greater output flow mainly to those supplying inputs such as insecticides, herbicides and fertilizers. It is also argued that the rice HYV (High Yielding Variety) package has led to environmental damage, seen for example in increased eutrophication, reduction in the quality and quantity of rice-field fish and in increased vulnerability to new kinds of pests as a result of the reduction of biodiversity consequent upon the use of single varieties over very large areas.

In response to these criticisms, it may be remarked that research, like much else, is driven, at least in part by basic economic considerations. The pay-off, in the form of keeping food production in Asia substantially ahead of population growth for four decades, has been a substantial one. It is also a moot question whether any major technological intervention in agriculture can be socially neutral. Eutrophication and soil toxicity are partly matters of over-use of fertilizers, herbicides and pesticides, in much the same way as improper use of agricultural chemicals poses health risks to users. Where double cropping of rice replaces single cropping, conditions certainly become adverse for rice-field fish and they may become positively dangerous to eat where chemicals are overused. Rice-field ecology is unquestionably modified by IRRI varieties and new techniques. Once relatively minor pests have become major the brown planthopper is an example. But further work has shown that it is possible to breed a degree of resistance into rice plants in the same way as it is feasible to improve fertilizer response genetically or to create varieties more tolerant of low or high temperatures at flowering. The number of rice HYVs has increased greatly since the days of IR6, the Institute s first major release some 40 years ago, so that biodiversity in HYV rice is steadily improving. This is much assisted by the Institute s collection of rice genotypes, many no longer grown.

The Institute s tasks are thus far from completion. While human population growth rates have fallen in Asia, the number of new stomachs to be filled is still many millions a year. Competition for the best rice land from urban growth, from other, more remunerative forms of rural production is increasing, so that the Institute s role in environmental change continues and is as pressing as ever.

FURTHER READING

Chandler, Jr, R T (1982) *An Adventure in Applied Science: a History of the International Rice Research Institute*, IRRI, Los Banos.

IRRI (1985) *International Rice Research: 25 Years of Partnership*, IRRI, Los Banos.

Irrigation: Environmental Impacts

Peter Crabb
Australian National University, Canberra, Australia

Irrigation is the arti cial application of water to soil, to supplement natural precipitation and so overcome any soil moisture de ciencies that would inhibit plant growth. It can permit crop production where none would be possible under natural conditions. It can extend the natural growing season, ensure that moisture is available at critical periods in plants' growth cycles, and increase both the quantity and quality of crop production. Because irrigation alters the length, characteristics and effectiveness of plants' growing periods, it creates an arti cial environment. In turn, new agricultural landscapes are created, with new socio- economic communities, their particular nature depending largely on the crops that are grown. The new landscapes are in marked contrast to adjoining areas of non-irrigated agriculture, or no agriculture at all. Irrigation has been described as a catalyst for rural change. Few human activities have brought about more environmental change than irrigation, locally, regionally and globally. It is a form of environmental change with a long history, making possible population growth and urban development, but in turn resulting in considerable environmental damage.

THE GROWTH AND GLOBAL EXTENT OF IRRIGATION

Irrigation is an ancient activity, dating back perhaps 8000 years at Jericho in the Jordan Valley (Hillel, 1994). It was

the basis of many of the earliest hydraulic civilizations in the Tigris Euphrates, Orontes, Nile and Indus Valleys. For the Sumerians on the Mesopotamian Plain, irrigation made possible increased food production and the development of the world s earliest urban society. Irrigation was critical to the pre-Hispanic civilizations on the coastal plain of Peru, in Mexico, and in what is now the USA states of New Mexico and Arizona. Whilst the irrigation systems were simple, their development, maintenance and water allocation arrangements required social stability and a high degree of communal organization. This was perhaps even truer for the terraced irrigation areas found in many Asian countries. This remains a valid observation, both in long-established societies, such as those of Japan and many Southeast Asian countries, and in the more recently established irrigation areas.

By the year 1800, an estimated 8 million ha of land were being irrigated. Over the next 100 years, the figure increased to over 40 million ha, much of the growth being due to the construction of large projects in the Indian sub-continent. The real expansion in the world s irrigated area, the *Blue Revolution*, was from the 1950s to the early 1970s, with the annual growth rate of between 2 and 4% being higher than that of the world s population. On a *per capita* basis, the area of irrigated land reached its peak of 48 ha per 1000 people in 1979. Since then, the situation has been reversed, with the irrigated area growing at no more than 1% a year and, with continued population growth, a continuing decline in the *per capita* figure (Table 1). Estimates of the United Nations Food and Agriculture Organization (FAO) indicate that in 1995, there were 258.9 million ha of irrigated land in the world. This was 17.3% of the total area of land classed as arable and under permanent crops, some 1499.9 million ha.

Table 1 Global expansion of irrigated land and population since 1800[a]

Year	Irrigated area, in millions ha	Population, in billions	Irrigated area in ha per 1000 people
1800	8	1.00	8.0
1900	40	1.50	26.7
1930	80	2.00	40.0
1950	94	2.50	37.6
1961	139	3.07	45.3
1965	151	3.35	45.1
1970	169	3.71	45.5
1975	190	4.08	46.6
1979	209	4.37	47.8
1980	211	4.45	47.4
1985	226	4.86	46.5
1990	239	5.30	45.1
1995	255	5.69	44.9

[a] Ghassemi *et al.* (1995); Rhoades (1997); FAO (1987).

Among the individual countries with one million ha or more of irrigated land in 1995 are some very high figures, especially for a number of developing countries (Table 2). However, lower figures in both developed and developing countries must not be taken to suggest that irrigation is of less importance. There are very few countries where irrigation is not undertaken.

Asia has 179 million ha of irrigated land, some 69.1% of the world s total; India, China and Pakistan alone account for 121.1 million ha, 46.8% of the world total (*see* Water Use: Future Trends, and Environmental and Social Impacts, Volume 3). In North and Central America, the largest areas are in the United States and Mexico. In Africa, much of the irrigated land is in the Nile basin, notably Egypt and Sudan. Irrigation is widespread in Europe, but particularly in countries in the south east of the continent and those bordering the Mediterranean. Brazil, Peru, Argentina, and Chile are the South America countries with the largest irrigated areas. Australia is one of 24 countries with over 2 million ha under irrigation. However, there is considerable variation from one country to another, especially in terms of irrigated land as a percentage of all arable and permanently cultivated land.

The area of irrigated land is an indirect indicator of the extent and growth of water use by agriculture. Globally, agriculture accounts for some 65% of withdrawals and 87% of consumption, with irrigation accounting for all but a small portion. Unfortunately, data on actual water use are unreliable in many ways and have to be treated with extreme caution (Engelman and LeRoy, 1993).

THE BENEFITS DERIVED FROM IRRIGATION

Globally and nationally, the greatest benefit of irrigation is its contribution to the production and security of food supplies. It is estimated that some 36% of the world s crop output comes from the 17% of the total arable area that is irrigated (Upton, 1994), including almost 30% of all food and 46.5% of all grains. As Postel (1993) noted, irrigation is now a cornerstone of global food security. The higher yields farmers ••• get with a controllable water supply [are] vital to feeding millions added to our numbers each year . For many countries, such as China, India, Indonesia, Israel, Japan, North and South Korea, and Peru, irrigation supports over half of domestic food production; in Pakistan and Egypt, 80 and 100%, respectively. The importance of irrigation for domestic food production is reflected in the fact that, based on FAO data, close to 60% of the world s total irrigated area is devoted to paddy rice (FAO, 1987; Jones, 1995, 112 113). It is also seen in the development of irrigation for domestic food supplies and export commodities in water-scarce countries in the Middle East and North Africa. In some developing countries, such

Table 2 Irrigation in countries with more than one million ha of irrigated land in 1995[a]

Country and continent	Irrigated area, in '000 ha, 1970	Irrigated land as a percentage of total arable land, 1970	Persons per irrigated hectare, 1970	Irrigated area, in '000 ha, 1995	Irrigated land as a percentage of total arable land, 1995	Persons per irrigated hectare, 1995
Egypt	2843	100.0	11.6	3283	100.0	18.9
Madagascar	330	13.9	19.4	1087	35.0	13.7
Morocco	340	4.5	45.0	1258	13.6	21.1
South Africa	1000	8.2	22.8	1270	8.0	32.6
Sudan	1750	23.1	7.9	1,946	15.1	13.7
Africa	7620	4.5	47.3	12 282	6.5	58.6
Mexico	3583	15.5	14.3	6100	22.3	14.9
USA	16 000	8.4	12.8	21 400	11.4	12.5
North and Central America	20 955	7.9	15.2	30 142	10.9	15.1
Argentina	1280	3.9	18.7	1700	6.3	20.5
Brazil	796	1.5	120.4	3169	5.6	50.2
Chile	1180	24.1	8.0	1265	29.5	11.2
Colombia	250	4.9	83.2	1037	20.7	34.5
Peru	1106	39.3	11.9	1753	46.1	13.4
South America	5681	5.1	33.5	9826	8.9	32.3
Afghanistan	2340	29.2	5.7	2800	34.8	7.0
Azerbaijan	[b]	[b]	[b]	1000	11.5	7.5
Bangladesh	1058	11.6	63.0	3305	33.8	35.8
China	38 123	37.2	21.8	49 857	37.9	24.5
India	30 440	18.4	18.2	54 000	31.9	17.2
Indonesia	4370	24.1	27.5	4580	14.3	43.1
Iran	5200	33.1	5.5	7264	42.0	9.4
Iraq	1480	29.6	6.3	3525	64.0	5.7
Japan	3415	61.9	30.6	2745	65.4	45.6
Kazakstan	[b]	[b]	[b]	2380	7.4	7.1
Korea DPR	500	25.0	27.8	1460	73.0	15.1
Korean Republic	1000	43.5	31.9	1206	57.2	37.2
Kyrgyzstan	[b]	[b]	[b]	1077	75.6	4.1
Myanmar	839	8.0	32.3	1555	15.4	29.0
Pakistan	12 950	67.0	5.1	17 200	82.1	7.9
Philippines	826	10.5	45.4	1580	17.3	42.9
Saudi Arabia	365	42.0	15.7	1478	42.5	12.4
Syria	451	7.6	13.9	1089	19.4	13.0
Thailand	1960	14.2	18.6	4642	22.5	12.5
Turkey	1800	6.6	19.6	4186	15.1	14.5
Turkmenistan	[b]	[b]	[b]	1300	87.8	3.1
Uzbekistan	[b]	[b]	[b]	4000	81.6	5.7
Vietnam	980	16.2	43.6	2000	31.3	36.9
Asia	109 727	24.9	19.2	179 013	35.3	19.2
France	750	3.9	67.7	1630	8.5	35.6
Greece	730	18.7	2.0	1328	34.0	7.9
Italy	2561	17.2	21.0	2710	22.6	21.1
Romania	731	6.9	27.7	3110	31.0	7.3
Russian Fed.	[b]	[b]	[b]	5360	4.0	27.7
Spain	2379	11.6	14.2	3527	17.5	11.2
Ukraine	[b]	[b]	[b]	2585	7.6	20.0
Europe	10 728	7.4	42.8	25 069	7.9	29.0
Former USSR	11 100	4.8	21.9			
Australia	1476	3.5	8.5	2317	4.8	7.7
Oceania	1588	3.7	12.2	2605	4.9	10.9
World	167 399	11.9	22.1	258 937	17.3	21.0

[a] *Source:* FAO (1987); FAO (1998).
[b] Included in Former USSR.

as Bangladesh and Pakistan, irrigated commodities provide the bulk of their export earnings (Ghassemi *et al.*, 1995).

Whilst the contribution of irrigation to food supplies and the alleviation of poverty is particularly evident in the developing world, irrigation s wider benefits are perhaps more clearly evident in the developed world. On farms, there is greater diversity and stability of production. The increased quantities and often higher value agricultural production support processing and farm service industries, such as the manufacture, installation and maintenance of irrigation equipment. High-value agricultural commodities frequently require other high-value farm inputs, such as planting and harvesting equipment, fertilizers, and pesticides. Their provision creates additional employment and larger rural communities. In the Canadian province of Alberta, only 4% of its arable land is irrigated, but this accounts for 20% of its agricultural production. Of the overall economic benefits derived from irrigation, Canadian farmers receive 15%, with the remaining 85% shared by local communities, the province, and Canada as a whole. Twenty years ago, an Australian study indicated that in New South Wales, each $1.00 (AUD (Australian Dollar)) of irrigation farm output and income generated $5.00 worth of output and income beyond the farm.

Irrigation development often brings associated benefits, such as hydro electricity generation, flood mitigation, and opportunities for recreation. Whilst it is difficult to put accurate values on many of these benefits, some estimates can be made: for example, sport fishing in Alberta s irrigation waters is worth some $17 million (CAD (Canadian Dollar)) a year.

Many governments have seen such rural development and the associated multiplier effects as the key benefit of irrigation and the major reason for their direct involvement. Especially in the past, government influence in irrigation was all-pervasive in many countries, and not simply in terms of the large subsidies provided for the establishment and operation of irrigation schemes, as in the western United States and Australia s Murray Valley. More recently, in Egypt, there have been moves to establish irrigation on the shores of Lake Nasser in Aswan Province, one of the least developed parts of the country and which has benefited little from the huge reservoir, and in the Western Desert (Pearce, 1997).

THE COSTS OF IRRIGATION

The benefits of irrigation are substantial; over 1/3 of the global harvest cannot be ignored. However, the benefits have not been achieved without wide-ranging and substantial costs, at farm, regional, national and even international levels.

The capital costs of irrigation schemes, covering water storage and reticulation (water distribution networks), land preparation, and the on-farm irrigation systems, are substantial and increasing (Yudelman, 1994). On top of these are the operating costs. Especially at the farm level, costs will vary depending on the nature of the schemes and the commodities produced. In the developing world, labor inputs are high and costs low; in the developed world, sophisticated equipment replaces high-cost labor. Energy is a significant component of both capital and operating costs. The construction of dams, storages and reticulation schemes consumes large quantities of energy. So, too, does the operation of irrigation schemes, either through fossil fuels or human or animal labor, especially for pumping. In the United States, about 10% of the total energy expended each year in crop and livestock production is used to pump irrigation water (Pimentel *et al.*, 1997). With the intensification of agricultural production, the increased inputs, such as fertilizers and other chemicals, all involve higher energy costs.

The establishment of irrigation brings major changes in terms of population, new communities and social conditions. With the intensification of agriculture comes the intensification of communities, demands for a new range of support services, and even different lifestyles, not least because of the inflexible nature of irrigation farming (crops have to be watered just as dairy cows have to be milked). The creation of large reservoirs and the establishment of irrigation schemes can result in the costly disruption and relocation of existing communities, as has happened in many parts of Asia and Africa.

A critical cost issue is the price irrigators pay for their water. In both developed and developing countries, it is evident that the economics of water use for irrigation have been distorted by enormous subsidies. Data from different countries indicate subsidies of 50 90% of costs (Falkenmark and Lundqvist, 1998). In hardly any situation are the full costs covered and, in many cases, water charges do not even cover the operating costs of the supply systems (see for example, Postel, 1999, 230 232). When water is essentially free or low priced, its waste is encouraged, as there is no incentive to improve the efficiency of its use. In many cases, the situation is further complicated by the fact that the value of the commodities that are produced is less than the real cost of the irrigation water that is used to grow them. On top of this, the agricultural commodities may also attract government subsidies. In South India, electricity for water pumps is free. Such situations reflect government attitudes to agriculture in general and irrigation in particular. Their attitudes are extremely difficult to change. Farmers in California pay about two per cent of what urban users are charged per unit of water (Pimentel *et al.*, 1997). In spite of their general unwillingness to do so, there is clearly a need for irrigators to pay a realistic price for the water they consume.

The impacts of the introduction of irrigation on a region s land and water resources are equivalent to a change of climate. Most of the impacts and their costs stem from excessive and inefficient water use. Irrigation requires a reliable water supply (though it can never be absolutely drought proof), made possible by engineering works of all kinds, from simple pedal pumps that laboriously lift water from a river to the most substantial and sophisticated storage and reticulation systems. The result is substantial alteration of the natural hydrological cycle, with significant environmental consequences for both surface water and groundwater. The demands of irrigation often result in major reductions in overall river flows and significant changes to river flow patterns, including the elimination of many flood flows. These and other changes have devastating impacts on all aspects of aquatic ecosystems (Postel *et al.*, 1996). Examples include the Murray, Colorado, Tigris Euphrates, Nile, Ganges, Huang He and Chao Phraya rivers, though among the most dramatic are the diversions from the Syr Darya and Amu Darya to expand cotton production and the consequent virtual destruction of the Aral Sea (Precoda, 1991; Postel, 1999) (*see* Aral Sea, Volume 4). Northern Nigeria provides more recent examples of environmental costs exceeding the benefits of irrigation (Barbier and Thompson, 1998).

The excessive extraction of water for irrigation is not confined to surface sources. Many groundwater resources are being used beyond their recharge rates, a practice that is not sustainable. At huge cost and with few benefits, Libya s Great Manmade River Project is pumping water from aquifers under the Sahara Desert to irrigate its coastal areas (McLoughlin, 1991; Pearce, 1991) (*see* Groundwater Withdrawal and the Development of the Great Manmade River Project, Libya, Volume 3). In Saudi Arabia, fossil water is being used to grow wheat, which is exported at a loss, though cuts in subsidies reduced production by more than 60% over 1994 1996 (Postel, 1998). Quite apart from the waste of a non-renewable asset, questions are being raised about the geological impact of using so much fossil water. On the US High Plains, depletion of the Ogallala Aquifer has contributed to the termination of irrigation on about one million ha, the significant reduction of surface stream flows, and threatened this semi-arid area s entire economy. However, with institutional and technical innovation, educational campaigns, and land and water stewardship of most farmers, long term irrigation appears likely in the High Plains (Kromm and White, 1997).

Paralleling the costs of water extraction are the consequences of water application. Because of inefficiencies of the reticulation systems (especially seepage and evaporation from open channels), as much as 60% of water is lost between its source and the crop root zone; in developing countries, the figure reaches 80 90% (Barrow, 1987, 202). Further losses occur on the farms, as excess water passes through the root zones and, where drainage is non-existent or inadequate, contributes to rising water tables, inevitably resulting in water logging and salinisation. Salinisation of land and streams is the scourge of irrigation (and increasingly of many dry land farming areas, such as Australia s Murray-Darling Basin (MDBMC, 1999)). Salinisation and associated soil degradation resulting from irrigation have a long history of causing the demise of irrigation schemes, from those of the Middle East in ancient times to India, China, and the United States in the 20th century (Hillel, 1994; Ghassemi *et al.*, 1995; Postel, 1999). Between 20 and 50% of the world s irrigated land is affected. In California s San Joaquin Valley, land is being permanently taken out of production because of salinisation. This is tragic, yet more cases are being added. In the West African Sahel, some 150,000 ha were under rice by the mid 1990s. It has become the chief staple as a result of the establishment of irrigation along the Senegal and Niger rivers. However, the soil and water conditions are not suitable, land being rendered useless in some areas after only one season by a concrete-like salt crust, essentially due to the absence of drainage (Madeley, 1993). This is but one illustration of the way that irrigation alters a region s society, economy and ecology.

Downstream, other water users and the aquatic ecosystems have to cope with the polluting and damaging effects of saline drainage water from the irrigation schemes, as well as the higher salinity levels that come with reduced flows. In addition, there are often other water-born pollutants from intensive irrigation production, some of which can be toxic. The downstream environmental impacts of such exported pollution are rarely, if ever, taken into account in the costing of irrigation agriculture.

Health problems are also often associated with irrigation, especially in tropical and sub-tropical areas, including bilharzia (as along the Nile and Volta) and malaria (as in Indus Valley).

IS IRRIGATION SUSTAINABLE?

Given the enormous economic, social and especially environmental costs of irrigation, there is much evidence to support the view that irrigation is not sustainable. Yet it has survived in some regions. For example, the rice terraces of Benaue, northern Luzon, were built at least 2000 years ago to maximize production in this very hilly region. Despite concerns being expressed about deforestation and soil erosion, the continued existence of the terraces depends on a sophisticated social organization for their labor-intensive operation and maintenance.

Whilst much of the world s irrigation continues to be undertaken in ways essentially similar to those practiced for hundreds of years hence creating the unsustainable situation dramatic changes are taking place that bring into question an unqualified negative response to the critical

question. Irrigation sustainability is technically possible, but only by making the changes that are needed to sustain and rejuvenate it (Rhoades, 1997).

As has been indicated, present patterns of water use, in terms of both consumption and contamination, are unsustainable. In particular, excessive water withdrawals for irrigation are contributing to a growing number of major environmental problems, quite apart from such disasters as the Aral Sea (Postel *et al.*, 1996). Issues of water allocation and priority of use will have to be addressed by governments and societies. Even the movement of a relatively small quantity of water out of irrigation would greatly benefit other users, including the environment, with little or no cost to irrigation.

Water will have to be used more efficiently, especially by irrigators. The performance of many irrigation schemes is sub-optimal, in economic, social and environmental terms, in both developed and developing countries. Not only do poorly managed schemes create serious problems, the reduction or elimination of management deficiencies would mean improved resource use, especially of water, and a reduction in the perceived *need* for further irrigation developments. For example, Chilean farmers now irrigate 22% more land with the same amount of water as before.

Significant improvements can be made to water distribution. The replacement of open channels by pipes virtually eliminates seepage and evaporation losses. With a pressurized system, farmers can take delivery of water when they require it, rather than on the basis of a pre-determined schedule that may bear no relation to crop needs. On the farm, various methods are available to improve water application efficiency, such as micro-jet sprinklers, drip hoses and underground systems, and the use of tensiometers. As important as the irrigation technology are the management skills of the farmers, as studies in the Riverland and Sunraysia districts of Australia s Murray Valley have shown. Investigations of differences in rice yields between research stations and farms showed that water control was responsible for 23% of the difference between maximum and actual yields (Yudelman, 1994). Excessive watering is not only wasteful but also unnecessary. Provided it does not occur at critical stages in a plant s development, water stress can lead to higher yields, lower costs and spectacular increases in water use efficiency (Jerie *et al.*, 1992; Upton, 1994). Benefits can also be gained from partial irrigation, the practice of scientifically under-irrigating crops to reduce yield in a controlled way. The last increment of water applied to a crop results only in a small yield increase. With limited water, the water saved through partial irrigation can be applied to other land where the incremental yield increase is large (Trimmer, 1990).

Paralleling improved efficiency in the delivery and application of water is the need for drainage to remove excess water from the plants root zone and prevent its addition to the groundwater (Letey, 1994). Many irrigation schemes have no drainage facilities, while those that do exist are often inadequate and inefficient. The absence of efficient drainage is the major cause of water logging and salinisation.

There are other issues. In terms of crop management, greater emphasis will have to be given to high value crops and to those that use less water (Pimentel *et al.*, 1997). For example, should rice be grown in semi-arid areas? As water is an economic good, pricing structures will have to be changed to reflect accurately the real cost of irrigation water. Realistic prices would be a major incentive for greatly increased water use efficiencies. With the development of water markets, almost all other users are able to pay much higher prices for water than irrigators, except for those farmers producing high-value agricultural commodities. Given the decreasing availability of fresh water, recycling irrigation drainage and other water provides an additional source (Shuval, 1987), though the use of marginal quality water requires much more careful management, while the long-term cumulative effects are largely unknown (Tanji, 1997). The same observations apply to the use of saline water for irrigation. In many locations, marginal quality water is blended with fresh water, though this reduces the available quantity of good quality water (Rhoades, 1997).

Critical to the future of irrigation is the rehabilitation of existing schemes. Even though irrigated agriculture has led to increased output, most irrigation systems are operating well below their capacity and are giving much lower increases in agricultural output than planned (Yudelman, 1994). Further, Because fewer and fewer new sources of water are available, the emphasis of irrigation investment has shifted away from new facilities towards rehabilitating and upgrading existing ones (Jones, 1995: xi). Because of the increasing problems with many existing schemes, as much land is being lost through water logging and salinisation as is being added by new schemes. More than 20% of the world s irrigated area is affected by salinisation, which is spreading by up to 2 million ha a year (Postel, 1999, 92 93). In both developed and developing countries, there are tremendous benefits to be gained from the rejuvenation of under-performing irrigation schemes of all kinds, including those that may seem to meet the best standards (Yudelman, 1994).

CONCLUSION

Irrigation is a major dilemma. On the one hand, in order to feed the world s increasing population, it needs to expand by some 2.25% a year, while appropriate intensification and expansion of irrigated areas may account for more than 50% of food requirements by the year 2025 (Pereira *et al.*, 1996). To these views must be added the likely impacts

of global warming, which could increase world irrigation needs by 26% to maintain current production (Pimentel *et al.*, 1997). On the other hand, global patterns of water use are unsustainable (Postel *et al.*, 1996). Climate change is already making its impact on irrigation. For example, irrigation water supplies are drying up in Mediterranean Europe. Water availability will be a serious constraint to achieving the food requirements projected for 2025. The need for irrigation water is likely to be greater than anticipated, and the available supply of it less than anticipated (Postel, 1998). Already, there are serious regional water problems, in many cases largely due to the huge demands of irrigation, as in California and parts of China.

Clearly, There is a point in water resources development when water can no longer be matched to the economy, but the economy must be matched to the water available (Teclaff, 1985, 568). This is true for all natural resources, in terms of both their use and the environmental consequences of their use. At current rates of expansion, by 2025, all the water resources now considered to be available for irrigation will have been exploited. In the view of Postel (1999: 259), irrigation agriculture is unprepared for the era of water scarcity and competition . Further, as Goodland *et al.* (1991: 2) state, sustainable development has to be development without growth in throughput beyond environmental carrying capacity , with development meaning getting qualitatively better not quantitatively bigger .

This is the critical point for the future sustainability of irrigation. It requires change in both the biophysical, cultural and political environments of irrigation. Irrigation and irrigators will have to change. Ultimately, the need to balance the conflicting water interests of urban residents, farmers, industry and natural ecosystems now evident in California will have to be faced worldwide (Engelman and LeRoy, 1993; Reisner, 1986). It is inevitable that water will have to be re-allocated to other uses from irrigation (including garden watering), in California and in many other parts of the world. In particular, the absolute necessity of maintaining environmental integrity has to be accepted. New water supplies will have to come from conservation, recycling, reuse, and improved water use efficiencies. In turn, such measures will reduce the long-term environmental damage that is so often associated with irrigation.

Change in the cultural and political environments of irrigation is equally critical. Irrigation is not only a form of agricultural production and land use but a human activity. In particular, it involves human decisions about how water is allocated, managed and valued (NRC, 1996). This must change. The continuation of current practices will not even maintain current levels of food production, let alone increase them. A major United States study, *A New Era for Irrigation*, concluded: It is critical that the same resourcefulness that has made irrigation such an important economic and cultural activity over the past 100 years be brought to bear in the future (NRC, 1996, 7).

See also: Indus Basin: a Case Study in Water Management, Volume 3.

REFERENCES

Barbier, E B and Thompson, J R (1998) The Value of Water: Floodplain Versus Large-Scale Irrigation Benefits in Northern Nigeria, *Ambio*, 27, 434 440.

Barrow, C (1987) *Water Resources and Agricultural Development in the Tropics*, Longman Scientific and Technical, London.

Engelman, R and LeRoy, P (1993) *Sustaining Water: Population and The Future of Renewable Water Supplies*, Population and Environment Program, Population Action International, Washington, DC.

Falkenmark, M and Lundqvist, J (1998) Towards Water Security: Political Determination and Human Adaptation Crucial, *Nat. Resour. Forum*, 22, 37 51.

FAO (1987) *FAO Production Yearbook 1986*, UN Food and Agriculture Organisation, Rome.

FAO (1998) *FAO Production Yearbook 1997*, UN Food and Agriculture Organisation, Rome.

Ghassemi, F, Jakeman, A J, and Nix, H A (1995) *Salinisation of Land and Water Resources: Human Causes, Extent, Management and Case Studies*, University of New South Wales Press, Sydney, and CAB International, Wallingford.

Goodland, R, Daly, D, and El Sarafy, S (1991) *Environmentally Sustainable Economic Development: Building on Brundtland*, Environment Working Paper No. 46, World Bank, Washington, DC.

Hillel, D (1994) *Rivers of Eden: The Struggle for Water and the Quest for Peace in the Middle East*, Oxford University Press, New York.

Jerie, P, Goodwin, I, and Mitchell, P (1992) Regulated Deficit Irrigation Strategies in *National Irrigation Convention Proceedings: Technology for Improved Irrigation Ef ciency and Productivity*, Australian National Committee on Irrigation and Drainage, Melbourne, 73 83.

Jones, W I (1995) *The World Bank and Irrigation*, The World Bank, Washington, DC.

Kromm, D E and White, S E (1997) *A Half Century of Change in American High Plains Irrigation*, Canadian Water Resources Association 50th Annual Conference, Lethbridge.

Letey, J (1994) Is Irrigated Agriculture Sustainable? in *Soil and Water Science: Key to Understanding Our Global Environment*, eds R S Baker, G W Gee, and C Rosenzweig, Soil Science Society of America, Madison, 23 37, Special Publication 41.

McLoughlin, P F M (1991) Libya s Great Manmade River Project: Prospects and Problems, *Nat. Resour. Forum*, 15, 220 227.

Madeley, J (1993) Will Rice Turn the Sahel to Salt? *New Sci.*, 140(1894), 35 37.

MDBMC (1999) *The Salinity Audit of the Murray-Darling Basin: a 100-Year Perspective*, Murray-Darling Basin Ministerial Council, Canberra.

NRC (1996) *A New Era for Irrigation*, National Academy Press, Washington.

Pearce, F (1991) Will Gaddafi s Great River Run Dry? *New Sci.*, September 7.
Pearce, F (1997) Deluge of criticism greets Nile irrigation plan, *New Sci.*, 153, 2066, 5.
Pereira, L S, Gilley, J R, and Jensen, M E (1996) Research Agenda on Sustainability of Irrigated Agriculture, *J. Irrig. Drain. Eng.*, 122, 172 177.
Pimentel, D, Houser, J, Preiss, E, White, O, Fang, H, Mesnick, L, Barsky, T, Tariche, S, Schreck, J, and Alpert, S (1997) Water Resources: Agriculture, the Environment, and Society, *BioScience*, 47, 97 106.
Postel, S (1993) Water and agriculture, in *Water in Crisis: A Guide to the World's Fresh Water Resources*, ed P H Gleick, Oxford University Press, New York, 56 66.
Postel, S L (1998) Water for Food Production: Will There be Enough in 2025? *BioScience*, 48, 629 637.
Postel, S (1999) *Pillar of Sand. Can the Irrigation Miracle Last?*, W W Norton & Company, New York.
Postel, S L, Daily, G C, and Ehrlich, P R (1996) Human Appropriation of Renewable Fresh Water, *Science*, 271, 785 788.
Precoda, N (1991) Requiem for the Aral Sea, *Ambio*, 20, 109 114.
Reisner, M (1986) *Cadillac Desert: the American West and its Disappearing Water*, Penguin Books, New York.
Rhoades, J D (1997) *Sustainability of Irrigation: an Overview of Salinity Problems and Control Strategies*, Canadian Water Resources Association 50th Annual Conference, Lethbridge.
Shuval, H I (1987) The Development of Water Reuse in Israel, *Ambio*, 16, 186 190.
Tanji, K K (1997) Irrigation with Marginal Quality Waters, *J. Irrig. Drain. Eng.*, 123, 165 169.
Teclaff, L A (1985) *Water Law in Historical Perspective*, W S Hein, Buffalo.
Trimmer, W L (1990) Applying Partial Irrigation in Pakistan, *J. Irrig. Drain. Eng.*, 116, 342 252.
Upton, M (1994) Spreading the Benefits of Irrigation, *ICID Bull.*, 43 1 12.
Yudelman, M (1994) The Future Role of Irrigation in Meeting the World s Food Supply, in *Soil and Water Science: Key to Understanding Our Global Environment*, eds R S Baker, G W Gee, and C Rosenzweig, Soil Science Society of America, Madison, WI, 15 21, Special Publication No. 41.

Irrigation: Induced Demise of Wetlands

A Dennis Lemly
Virginia Tech University, Blacksburg, VA, USA

Irrigation is essential to support agricultural production in many arid and semi-arid regions of the world. However, there is often an important trade-off. The bene ts from irrigation can be offset by environmental costs, particularly to wetlands, and the many functions they provide to society; for example, water, food, and aesthetic uses. The intensive irrigation practices typically used in modern agriculture place high demands on water supplies. Native wetlands and their associated wildlife populations can be eliminated in a short period of time because water removal for use in irrigation robs wetlands of their source of water and they shrink in size or dry up. For example, in the western US, wetland losses due to agricultural irrigation exceed 90% in many locations, and have been occurring largely unimpeded for the past 100 years (Reisner, 1986).

In addition to the direct loss of wetland areas, wetlands can lose their capacity to function normally if agricultural pesticides and herbicides in the surface run-off from irrigated elds or subsurface irrigation drainage contaminate the water sources. Subsurface drainage was identi ed in the United States in the early 1980s as a major cause of wetland change. Irrigation water trapped by shallow clay lenses must be removed or drained away, otherwise it saturates the root zone and kills crops. The resultant drainage contains elevated concentrations of soil trace elements, salts, and other constituents, and ultimately reaches major wetlands, streams, and rivers. Fish and wildlife populations have been poisoned by drainage water contaminants at several locations and human health has been jeopardized (Lemly et al., *1993; Lemly, 1994). Thus, although reductions in water supplies have caused the demise of wetlands for centuries, there are more recent threats that are just now being identi ed and investigated.*

Wetlands continue to play a direct role in supporting many human populations through their varied functions, products, and attributes (Hruby et al., *1995). There are also numerous indirect roles. For example, wetlands are important in removing greenhouse gases, such as carbon dioxide, from the atmosphere. The global assimilation of these gases may be detrimentally affected by wetland losses associated with irrigation. Wetlands are a sensitive public issue because they provide a habitat and staging grounds for migratory birds, which have important aesthetic, recreational, and conservation values.*

Major irrigation-related impacts to wetlands have continued to take place throughout the world during the past 25 years despite international conservation efforts, such as the Ramsar Convention and the Rio Conference (Lemly et al., *2000). One thing is certain - nding an environmentally acceptable balance for irrigated agriculture is more complicated, yet more necessary, than ever before.*

GLOBAL EXAMPLES

Africa (Table 1; Figure 1)

The most extensive wetlands within the African continent are the large seasonally inundated floodplains. These

Table 1 Some examples of the impact of irrigation on Africa's floodplain wetlands

Wetland	Status	Cause of impact	Period of impact	Effects on wetland and wildlife	References
Logone Floodplain, Cameroon	Includes the Waza National Park	Upstream barrage for irrigation and riverside embankment to protect irrigation scheme	1970 to present	Reduced wet season inundation and loss of grazing for wildlife, including ungulates and elephants	Drijver and Marchand (1985); Wesseling *et al.* (1996)
Floodplain of the Benue River, Cameroon	Close proximity to three NPs	Large dam and reservoir for 29 000 ha of irrigation	1982 to present	Loss of important wildlife habitat beneath the reservoir, reduction in downstream inundation, disruption of migration routes of large mammals, including buffalo, antelope and elephant	Drijver and Marchand (1985)
Hadejia–Nguru Wetlands, north-east Nigeria	Partly NP and nature reserves	Two large dams, one barrage and two large-scale irrigation schemes upstream. Expansion of small-scale irrigation within the wetlands	1971 to present	Reduced wet season inundation leading to declines in wildlife habitat. Removal of natural vegetation cover. Reports of river pollution with agricultural chemicals	Hollis *et al.* (1993); Thompson and Hollis (1995); Thompson (1995)
Phongolo Floodplain, South Africa	Includes the Ndmu Game Reserve	Dam constructed upstream for 40 000 ha of irrigated land	1970 to early 1990s	Changes to hydrological regime of the floodplain, including reduced flood flows and inundation extent. Recently, an artificial flood regime has been established	Acreman (1994); Bruwer *et al.* (1996)
Senegal Delta, Senegal	Includes the Djoudj National Park, a Ramsar site, World Heritage site and NP	Upstream dam for irrigation and hydropower, downstream barrage to prevent saline intrusion, river embankments to prevent flooding, expansion of irrigated agriculture	1985 to present	Loss of wetland habitat to numerous rice schemes. Reduced water supplies to some wetland areas, reduced fish populations, encroachment of freshwater vegetation into areas of open water	Drijver and Marchand (1985); N'diaye (1997); Vincke (1996)

NP, national park.

wetlands, including the Inner Delta of the Niger in Mali, the Sudd of the Sudanian Upper Nile and Zambia s Kafue Flats, provide valuable habitats for wildlife. They support a huge variety of bird life and are vital for many migratory species that winter within them. The presence of water within many floodplains during the dry season provides important grazing areas for wild ungulates during periods when the surrounding drylands are desiccated.

In addition to their importance for wildlife, Africa s floodplains play a vital life-support role for a significant

Figure 1 Major African floodplain wetlands that have been detrimentally affected by water management schemes implemented for agricultural irrigation

proportion of the continent s population (Drijver and Marchand, 1985). Human populations have, for centuries, utilized the agricultural, fisheries, hunting, grazing and water resources provided by floodplains (Adams, 1996). These wetlands have rightly been termed *the heart of Sahelian life systems*.

Historically, the different uses of floodplain resources were integrated economically and ecologically. Floodplains have been able to support wildlife and people. However, over the last 40 years, many dams have been built as part of large-scale schemes of water management for irrigated agriculture (Thompson, 1996). In 1987, Van Ketel and others identified 114 dams which were detrimentally affecting wetlands within West Africa alone. Drijver and Van Wetten (1992) stated that by the year 2020, all Sahelian wetlands would be subject to the impacts associated with upstream dams. These impacts will have profound consequences for the wildlife and human populations which depend on wetlands.

At a local level, the expansion of small-scale irrigation has led to the replacement of natural vegetation or rainfed agricultural land with small plots (typically 1 4 ha), in which relatively high value crops, such as wheat, peppers, onions, sweet potato and tomatoes, are grown. The expansion of this form of agriculture was aided by a sharp rise in wheat prices in the 1970s and 1980s resulting from a ban on wheat imports into Nigeria (Kimmage, 1991). This coincided with the dissemination of approximately 70 000 portable irrigation pumps by the Kano State Agricultural and Rural Development Program. This change in agricultural practices enabled areas previously considered as marginal for cultivation to be cropped. The quality of wetlands and their suitability for wildlife have been altered due to shifts in vegetation and increased salinity. Muslim and Umar (1995) indicated that agricultural intensification has also resulted in increasing concentrations of chemical pollutants, largely from fertilizers, within the wetlands.

Table 2 A sample of 20 key wetlands in Australia that have been changed by irrigation

Wetland	Status	Cause of impact	Period of impact	Effects on wetland and wildlife	References
Barmah/Millewa Forest	Ramsar	Diversion and regulation of water for irrigation. Timber harvesting. Flows in the wetland regulated	1936 to present	About 50% of wetland lost. Decreased flood frequency and changes in season of flooding. Wetland vegetation less reliant on flooding; replaced by more aquatic vegetation	Bren (1988, 1992)
Wetlands of the Border Rivers	No reserve status	Diversion and regulation of water for irrigation	1980 to present	Decreased wetland flooding	Kingsford (1995)
Bowling Green Bay	NP, Ramsar	Water augmentation to inflowing rivers from irrigation			Blackman *et al.* (1996)
Carrum Swamplands		Swamps drained for agriculture	1870–1950	Wetland area declined from 4050 ha to 81 ha	Norman and Corrick (1988)
East coast Wetlands of New South Wales	Includes Kooragang Island Nature Reserve	Drainage schemes to limit flooding damage to agricultural and urban land	1900–1980s	Drainage affected 96% of wetland area on Macleay river floodplain	Goodrick (1970); Pressey (1989)
Gwydir Wetlands		Diversion and regulation of water for irrigation. Seventy percent reduction in floods of 100 Gl per month. Flooding frequency declined halved, from 1 in 2 to 1 in 4 years. Drains and channelization. Elevated levels of herbicides and insecticides (endosulfan) detected		Reduction in areas of aquatic vegetation, reed beds from 4000 ha to 250 ha. Abundance and diversity of waterbirds declined from 58 to 45 species	Debus (1989)
Hattah–Kulkyne Lakes	NP, Ramsar, Biosphere reserve	River regulation has reduced inflows to the lakes from the Murray River	1900 to present	Increased blooms of cyanobacteria. Reduced flood frequency	Hull (1996)
Herbert River Floodplain	NP (part), wetland reserves, fish habitat reserve	Decline in water quality caused by agriculture. Sugar cane production destroyed most of wooded swamps and many other wetlands. Drains replaced natural watercourses			Blackman *et al.* (1996)
Kerang Wetlands	Ramsar site wildlife reserves	Water levels held permanently high in wetlands. Increased salinity. Levees to control flooding, channeling of water courses. Saline water disposed in some wetlands	1936 to present	Prolonged inundation caused a loss of aquatic vegetation adapted to wet and dry cycles. Increased salinity affected fauna and flora	Hull (1996)
Koo-wee-rup Swamp		Drainage for agriculture		No wetland remains, declined from 40 000 ha. No aquatic plants or animals	Norman and Corrick (1988)
Lake Albacutya	NP, Ramsar site, national heritage	Historical records indicate reduced flooding because of upstream diversion of water for irrigation. Rising saline groundwater	1920s to present	Degradation of vegetation communities	Hull (1996)

Macquarie Marshes	Nature reserve, Register of National Estate, Ramsar	Diversion of water upstream mainly for irrigation	1969 to present	Reduced wetland size by at least 40–50%. Number of waterbirds and species in decline. Wetland vegetation in decline	Kingsford and Thomas (1995)
Menindee Lakes	Included in Kinchega National Park	Water levels kept artificially high for irrigation and public water supply. Tandou Lake no longer a wetland	1950s to present	Aquatic vegetation killed by too much flooding. Reduced numbers of waterbirds. Waterbird community dominated by piscivores. Erosion of lakeshores	Kingsford (1995)
Murray River Wetlands		Water levels kept artificially high for irrigation on 35% of wetlands	1900 to present	Death and deterioration of floodplain vegetation	Bren (1990); Pressey (1990); Smith and Smith (1990)
Murrumbidgee River Wetlands		Water levels kept artificially high; 62% of wetland area with water levels controlled locally		570 ha of floodplain eucalypt trees killed; ducks did not breed where water levels were highly controlled. Breeding of herons, egrets, and cormorants dependent on flooding of red gum. Loss of some breeding habitat for egrets	Briggs *et al.* (1994); Thornton and Briggs (1994); Briggs *et al.* (1997)
Narran Lake	Nature reserve	Water diversion for irrigation	1985 to present	Reductions of flow by 31%	Kingsford (1995)
Ord and Parry River Floodplains	Nature reserve and Ramsar site	Building of reservoir (Lakes Argyle and Kununurra). Reduced flooding. Ord irrigation scheme diverts water.	1972 to present	Prevented migration of fish; impact on downstream wetlands. Submerged seasonal wetlands (e.g., Packsaddle Swamp)	Lane and McComb (1988)
Swan Coastal Plain		Drainage of wetlands for agriculture, roads, urban development. Increased water levels	Up to 1964	Seventy percent of wetlands lost or significantly altered by clearing of vegetation	Halse (1989)
South-west Western Australia	Affected nature reserves include Lakes Toolibin (Ramsar site); Dumbleyung, Coyrecup, Pinjarrega, Nonalling, Beverley Lakes	Clearing of dryland vegetation for agriculture	1900–1960s	Salinization of wetlands is common	Lane and McComb (1988); Froend *et al.* (1987)
Lower South-east Wetlands of South Australia (include Bool Lagoon)	Bool and Hacks Lagoon (Ramsar site; conservation park)	Drainage to increase availability of lands for grazing livestock	1863–1981	Loss of wetlands; affected 53% of flood prone areas; more than 93% loss of permanent lakes and swamps	Lane and McComb (1988)

At a regional and continental level, a far more significant impact on the acreage of wetlands has resulted from the construction of large-scale irrigation schemes and associated dams. For example, in the Hadejia Nguru River Basin in Nigeria alone, there are currently over 20 dams, all upstream of the wetlands. The largest are associated with two major irrigation schemes located on the Hadejia River or its tributaries. The Kano River Irrigation Project (KRIP) is the biggest scheme at present. Tiga Dam, the largest dam in the basin, which was constructed during the period 1971 1974, provides irrigation water. The plan for KRIP envisaged its development in two distinct phases. A total irrigated area of 27 000 ha was planned for Phase I (KRIP-I) while Phase II (KRIP-II) would add a further 40 000 ha. The construction of Phase I began in 1977, and to date, around 14 000 ha have been completed.

The second major irrigation scheme, known as the Hadejia Valley Project (HVP), is under construction. The scheme involves a barrage across the Hadejia River just upstream of the Hadejia Nguru Wetlands and a planned 12 500 ha of formal irrigation. At present, around 7000 8000 ha have been completed. Challawa Gorge Dam on the Challawa River, upstream of Kano, is linked to the HVP. The dam, which was completed in 1992, is designed to store water during the wet season and subsequently release it downstream for use on the HVP.

The development of these large-scale schemes has had a tremendous impact on the hydrology of the river basin. These impacts have been concentrated in downstream areas. By retaining and utilizing water upstream, the schemes have reduced the magnitudes of downstream flood discharges and thereby reduced the area inundated within the floodplain. Any further expansion in the area of large-scale irrigation within the basin will clearly have tremendous implications for wildlife and human communities within the wetlands. In particular, plans for 84 000 ha of irrigation and a massive dam at Kafin Zaki on the, as yet, undammed Jama are River will reduce flood extents and associated wetlands by an average of over 50% (Thompson, 1995).

Australia (Table 2; Figure 2)

Agricultural irrigation has caused major changes in the wetlands of Australia. Hydropower generation and urban development have destroyed wetlands, but agriculture has caused most wetland loss.

The 1960s marked the beginning of the great development of water resources within Australia. In New South Wales, for example, a 25-year plan was initiated in 1971 to spend A$723 million (about US$60 million, in 1997 dollars), primarily on building dams (WRIC, 1971). Australia now boasts the largest per capita water storage in the world

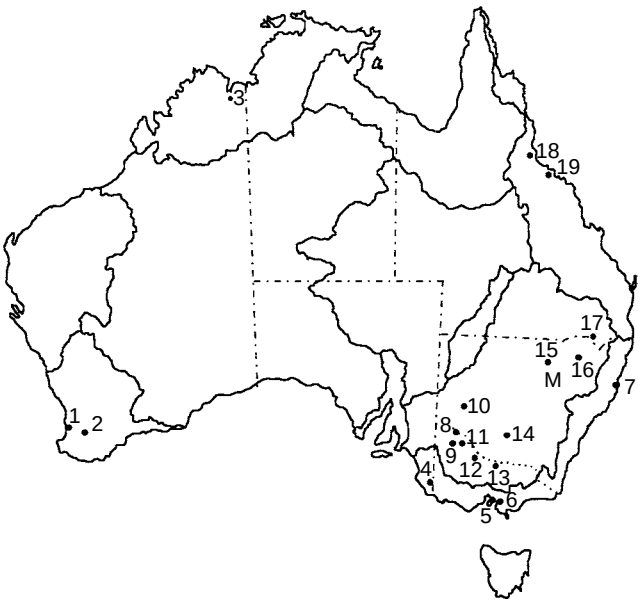


Figure 2 A sample of significant wetlands in Australia that have been detrimentally affected by irrigated agriculture. Lines mark the major drainage divisions in Australia. M marks the location of the Macquarie Marshes and the numbers refer to the following wetlands: (1) Swan Coastal Plain; (2) Lake Toolibin; (3) Ord and Parry River Floodplains; (4) Bool Lagoon; (5) Carrum Wetlands; (6) Koo-wee-rup Swamp; (7) Macleay River Floodplain; (8) Murray River Wetlands; (9) Lake Albacutya; (10) Menindee Lakes; (11) Hattah–Kulkyne Lakes; (12) Kerang Wetlands; (13) Barmah–Millewa Forest; (14) Murrumbidgee Wetlands; (15) Narran Lake; (16) Gwydir wetlands; (17) Wetlands of the Border Rivers; (18) Herbert River Floodplain; (19) Bowling Green Bay Wetlands

(Wasson *et al.*, 1996). In New South Wales, there are 144 large storage reservoirs, each with a capacity of greater than 1000 Ml (10^6 l) (Kingsford, 1995). These have had a significant impact on wetlands and water resources in the country. The management of water for irrigation, mainly with large dams, has severely affected river flows (Maheshwari *et al.*, 1995) and wetlands (Pressey and Middleton, 1982). With the changed water regime, many wetlands now store water permanently, which has killed aquatic vegetation adapted to wet and dry cycles.

A more insidious problem is the impact of the diversion of water upstream on the extent and duration of flooding. A notable example is the Macquarie Marshes. In 1896, the first weir was built on the Macquarie River to manage water. Nine large dams, five major weirs and several minor weirs, a water transfer scheme, and numerous other regulatory structures (Kingsford and Thomas, 1995) store water for later release to a large irrigation industry. About 89% of the water diverted from the Macquarie River is used for irrigation. Irrigation increased after Burrendong Dam was built in 1967. Burrendong Dam has a storage capacity 70%

larger than any other regulatory storage in the Macquarie Valley. Irrigated cotton was first grown in the valley in 1967 (McHugh, 1996), and now accounts for more than half of the water use for irrigation. It peaked in 1993/1994 when 543 000 Ml were diverted for irrigation. Originally flooding up to 1 280 000 ha, the Macquarie Marshes now inundate only about 130 000 ha during large floods (Kingsford and Thomas, 1995).

The impact of irrigated agriculture on the wetland and its water birds is profound. The total amount of water reaching the Macquarie Marshes is a highly significant predictor of wetland size (Kingsford and Thomas, 1995). Due to reduction in flows as a result of the diversion of water upstream, mainly for irrigated agriculture, the wetland is now conservatively 40 50% smaller than it was before water was diverted (Kingsford and Thomas, 1995). Before the Burrendong Dam was built, about 50% of the water, which passed the upstream gauge at Dubbo, reached the Macquarie Marshes. Now only about 21% of this water reaches the Marshes, despite no changes in the catchment rainfall in the period 1944 1993 (Kingsford and Thomas, 1995). Wetland trees, such as river red gums, have declined by about 50% during the period 1934 1998; reed beds experienced similar declines between 1963 and 1972 (Brander, 1987). Water couch (*Paspalum paspaloides*) has declined by 40% in some areas during the period 1949 1991, with exotic dryland vegetation replacing it. The numbers of water birds and the numbers of species declined in the period 1983 1995 (Kingsford and Thomas, 1995). Reduced flows have also impacted the size of breeding colonies of Ciconiiformes. Conservatively, this has probably meant that there are now about 100 000 fewer nests over a period of 11 years as result of water that was diverted from the Macquarie River. Fish surveys also indicate a substantial decline in diversity and biomass.

Irrigation upstream of the Macquarie Marshes is also raising water tables, which has caused salinity problems that are contributing to long-term increases in the salinity of water reaching the Macquarie Marshes (*see* Irrigation: Environmental Impacts, Volume 3; Salinity and Agriculture, Volume 3). Four pesticide products occur in the river: endosulfan, atrazine, prometryn and parathion. Sampling during the summer of 1994 1995 at two sites at the beginning of the Macquarie Marshes detected levels of endosulfan greater than $0.01\,\mu g\,l^{\bullet\,1}$, which is above the threshold at which toxic impacts to aquatic ecosystems begin to occur.

North America (Table 3; Figure 3)

Limited availability of water was one of the major obstacles to early settlement in many arid and semi-arid regions in western North America (Reisner, 1986). In the US, the shortage of water and perceived need to homestead on desert lands led to the establishment of the US Bureau of Reclamation in 1902 (then named the Reclamation Service). The primary mission of this agency was to *reclaim* unproductive lands by bringing water to arid regions and turning desert into farmland. This reclamation began in earnest with the completion of the Newlands Water Project in Nevada in 1915 and reached a peak with the building of such massive water storage/diversion projects as the Hoover Dam on the Colorado River, the California Aquaduct across the Mojave Desert, and the Central Valley Irrigation Project in California (Reisner, 1986). However, the price for this reclamation and associated increase in agricultural production was a sharp decline in the amount of water available to native wetlands.

A combination of reduced freshwater inflows and contaminated subsurface irrigation drainage has seriously degraded wetlands throughout western North America. The impacts to wildlife have been severe, and the conservation benefits provided by these wetlands have dropped precipitously (Lemly *et al.*, 1993; Lemly, 1994). Battles for water rights between settlers, agriculture, urban developers, and water authorities have raged for well over a century (Reisner, 1986), but only in the last few years have the water needs of remnant native wetlands gained attention and become the focus of water management policies. A notable example is the wetlands located in the Lahontan Valley of northwestern Nevada.

An extensive wetland ecosystem, comprising some 70 000 ha, existed in low areas of the Carson Desert of northwestern Nevada prior to the completion of the Newlands Irrigation Project in 1915. Of this amount, 14 000 ha in Stillwater Marsh, 11 000 ha in Carson Lake Marsh, and about 11 000 ha at the mouth of the Carson River in the Carson Sink (Fallon NWR) were terminal drainage areas for the Carson River and were directly impacted by the Newlands Project. Since the initiation of the project in 1905, the quantity and quality of the water reaching the terminal wetlands has declined precipitously. The water quantity has been reduced due to the diversion of water from the Carson River for use in irrigating crop land, primarily to grow alfalfa that is harvested and sold as livestock feed in Nevada and neighboring states. The water quality has declined because of saline, contaminated subsurface drainage formed as a by-product of irrigation. The net effect has been the elimination of 59 000 ha of wetlands in exchange for the water used to irrigate 22 000 ha of land. A total of 84% of the native wetlands are gone and the remaining marshland receives water that contains up to 100 times the historic concentrations of dissolved solids, including toxic trace elements such as mercury, arsenic, and selenium (Hoffman *et al.*, 1990; Hallock and Hallock, 1993). The changes in water quality have led to extensive declines in marsh vegetation in the remaining wetlands (Hoffman *et al.*, 1990; Lemly, 1994).

Table 3 A sample of wetlands in North America that have changed due to irrigated agriculture

Wetland	Status[a]	Cause of impact	Period of impact	Effects on wetland and wildlife	References
Stillwater Marsh	NWR, Ramsar, hemispheric reserve for shorebirds	Water diversions for crop irrigation resulted in • 90% reduction of inflows and produced contaminated subsurface drainage	1915 to present	Loss of 71% of wetlands; production of shorebirds cut by • 60%; death and deformity of birds and fish due to contaminants in irrigation drainage; mortality of endangered species	Hoffman *et al.* (1990); Hallock and Hallock (1993); Lemly (1994)
Kesterson Marsh	NWR	Contaminated subsurface irrigation drainage	1978–1985	Death and deformity of thousands of waterfowl and shorebirds; all marshland (2388 ha) drained and sediments excavated. No wetlands remain	Zahm (1986); Hoffman *et al.* (1986); Marshall (1985)
Grasslands Marshes	NWR, SWMA, 70% private ownership	Freshwater inflows replaced by contaminated irrigation drainage	1950 to present	Accumulation of drainwater contaminants in shorebirds to toxic levels for reproduction	Ohlendorf *et al.* (1987); Hothem and Welsh (1994)
Rasmus Lee Lake, Goose Lake Wetlands	SWMA	Contaminated subsurface irrigation drainage	1986 to present	Bio-accumulation of selenium to toxic levels; death and deformity of waterfowl and shorebirds	Peterson *et al.* (1988); See *et al.* (1992a,b)
Ouray Wetlands	NWR, SWMA	Contaminated subsurface irrigation drainage	1960 to present	Bio-accumulation of selenium to toxic levels; death and deformity of waterfowl, shorebirds, and fish; mortality of endangered species	Stephens *et al.* (1988, 1992); Waddell and Stanger (1992); Hamilton (1998)
Benton Lake Wetlands	NWR	Freshwater inflows replaced by contaminated irrigation drainage	1970 to present	Bio-accumulation of trace elements to toxic levels in waterfowl, shorebirds, and fish; reproductive impairment in birds	Knapton *et al.* (1987); Lambing *et al.* (1987)
Mexican Riverine Wetlands	WP, SWMA	Diversion of freshwater inflows for irrigation; salinization of surface and ground water supplies	1963 to present	Loss of • 50% of wetlands and shallow water habitat; local loss of up to 68% of native fish species; total extinction of 15 rare or endangered fish species	Conteras-Balderas and Lozano-Vilano (1994)
South Saskatchewan Basin Wetlands	NP, PP, PWMA	Diversion of freshwater inflows for irrigation; salinization of surface and ground water supplies	1940 to present	Loss of • 40% of wetlands and shallow water habitat; declines in bird usage and production	Livingstone and Campbell (1992); Gilbert and Ramey (1995)

[a] NWR, national wildlife refuge; SWMA, state wildlife management area; WP, wildlife preserve; NP, national park; PP, provincial park; PWMA, provincial wildlife management area.

Figure 3 A sample of significant wetlands in North America that have been detrimentally affected by irrigated agriculture. L marks the location of the Lahontan Valley Wetlands, which include Stillwater Marsh, and numbers identify the following wetlands: (1) Kesterson Marsh; (2) Grasslands Marshes; (3) Rasmus Lee Lake and Goose Lake Wetlands; (4) Ouray Wetlands; (5) Benton Lake Wetlands; (6) Mexican Riverine Wetlands; (7) South Saskatchewan River Basin Wetlands

The irrigation-induced effects on wildlife have been severe. Fish populations are greatly reduced, and the species composition has changed since historical times. For example, the Lahontan cutthroat trout (*Oncorhynchus clarki henshawi*), a Federally listed endangered species, has been totally eliminated. The largemouth bass (*Micropterus salmoides*), an introduced species that once supported an extensive recreational fishery, has also entirely disappeared. Most of the native species are now absent from the wetlands and those that remain are pollution-tolerant fishes, such as common carp (*Cyprinus carpio*) and mosquitofish (*Gambusia af nis*). The once diverse fish forage base supporting American white pelicans has been greatly reduced. The river otter (*Lutra canadensis*), mink (*Mustela vison*), frogs, and turtles characteristic marsh species that were once numerous are all gone. Other species have experienced precipitous declines in numbers, to the point that they are now uncommon e.g., the American curlew (*Numenius americanus*). Muskrats (*Ondatra zibethicus*), freshwater clams and aquatic snails were once abundant throughout the wetlands, but only small, remnant populations remain. Freshwater clams, for example, are only present in Stillwater Point Reservoir, which is the area with the lowest dissolved solids concentration (Hallock and Hallock, 1993).

Waterfowl production is now much reduced compared with historical conditions. The numbers of nesting birds, as well as the percentage of successful waterfowl and shorebird nests, have steadily declined due to a loss of wetland acreage and emergent vegetation, coupled with exposure to contaminants in irrigation drainage water. Only about 25% of nests succeed in hatching even one chick, and the birds that do fledge contain elevated concentrations of mercury and selenium. Moreover, the water in many marsh areas is toxic to aquatic life because of the accumulated salts and trace elements that are leached out of crop land during irrigation and carried to the wetlands in drainwater (e.g., selenium, boron, molybdenum, lithium, arsenic). There are also public health concerns. Selenium, for example, accumulates in biota to concentrations that are 10 000 times those present in water, resulting in tissue residues in waterfowl that are four times the safe levels for human consumption (Hallock and Hallock, 1993).

OUTLOOK FOR THE FUTURE

The degree of wetland change caused by irrigation over the past century has reached a critical point on a global scale. Many key wetlands have lost the ability to sustain the critical hydrological and biological functions responsible for multiple benefits to the environment and society. Particularly troubling is the possibility that future changes in global climate will affect precipitation patterns and increase the extent of arid and semi-arid regions. This would increase the need for irrigation to support agriculture and accelerate declines in remaining native wetlands. The repercussions are significant. Loss of wetlands may contribute to alterations in climate by reducing the amount of carbon taken up and sequestered in wetland plants, thereby enhancing atmospheric CO_2 and the greenhouse effect. Climate change means alteration of agricultural practices and crops. Thus, there is an intricate cycle in operation; wetlands are detrimentally affected by water withdrawals for irrigation, but at the same time, destroying wetlands may contribute to climate changes that detrimentally affect agriculture. It is important to recognize that the agricultural systems we now depend on are directly affected by what society does to wetlands.

Loss of wetlands also translates to reduced biodiversity on a global scale. This fact has not gone unnoticed. Recent developments, such as the United Nations Biodiversity Convention (Denny, 1994), show a willingness to resolve environmental issues related to wetlands. Agenda 21, a global partnership for sustainable development in the 21st century, was established at the UN Conference on Environment and Development in Rio de Janiero in 1992. Although the goals are noble, conferences such as this may merely establish *political correctness* rather than institute meaningful changes in environmental management policy. Care must be taken that the euphoria born at the

Rio Conference is not allowed to degenerate into a series of platitudes in which words like *biodiversity* and *sustainable development* become political words of convenience rather than words of true meaning. (Denny, 1994) This quote highlights the key failing of past efforts much is said and written, but little is done. Moreover, the track record established over the past 40 years suggests that *sustainable development* is often a political code name for *perpetual growth*, used strategically to make development more palatable to those who would oppose it (Willers, 1994).

Irrigation threats to wetlands are often disguised within a sustainable development scenario i.e., that continued growth and development are compatible with environmental constraints. As noted by Pearce *et al.* (1989), sustainable development has come to mean whatever suits the advocacy of the individual or group concerned. Advocacy for sustainable agriculture under a false pretense of environmental compatibility has been a principal factor guiding water policy decisions for decades. Environmental management has been intense and largely in one direction toward the expansion of irrigation. Yet, the recurring result has been that so-called sustainable development has guaranteed the deterioration of wetland ecosystems and loss of biodiversity.

Recent surveys reveal that a steady growth of irrigated agriculture is continuing to occur (e.g., Anonymous, 1997). Water demands for domestic and industrial use will no doubt expand substantially over the next 30 years (Postel *et al.*, 1996), making the issue of wetland values more important than ever in resolving problems. Detailed economic valuation of wetland resources and application of the resultant ecosystem *capital* in water negotiations is being used to develop management plans that protect wetlands. For example, the New South Wales government in Australia reached decisions for managing the Macquarie Marshes based in a large part on ecological grounds (values associated with wildlife preservation) despite considerable lobbying by the irrigation industry with arguments that about A$35 million (about US$3 million) and about 500 jobs would be lost (Morrison and Kingsford, 1997). With cooperation and strategic planning, local citizens and wetland managers are having a significant effect on how water policies are developed and implemented. This is a positive sign that historic trends in wetland change due to irrigation are beginning to reverse.

REFERENCES

Acreman, M C (1994) The Role of Artificial Flooding in the Integrated Development of River Basins in Africa, in *Integrated River Basin Development*, eds C Kirby and W R White, John Wiley & Sons, Chichester, 35 44.

Adams, W M (1996) Economics and Hydrological Management of African Floodplains, in *Water Management and Wetlands in Sub-Saharan Africa*, eds M C Acreman and G E Hollis, IUCN, Gland, Switzerland, 21 33.

Anonymous (1997) 1996 Irrigation Survey Reflects Steady Growth, *Irrig. J.*, Jan/Feb.

Blackman, J G, Perry, T W, Ford, G I, Craven, S A, Gardiner, S J, and De Lai, R J (1996) *Queensland. A Directory of Important Wetlands in Australia*, 2nd edition, Australian Nature Conservation Agency, Canberra.

Brander, D (1987) Environmental Changes in the Southern Macquarie Marshes: 1934 to 1987. Honours Thesis, University of New South Wales, New South Wales, Australia.

Bren, L (1988) Effects of River Regulation on Flooding of a Riparian Red Gum Forest on the River Murray, Australia, *Regul. Rivers*, 2, 65 77.

Bren, L (1990) Red Gum Forests, in *The Murray*, eds N Mackay and D Eastburn, Murray Darling Basin Commission, Canberra, 231 242.

Bren, L (1992) Tree Invasion of an Intermittent Wetland in Relation to Changes in the Flooding Frequency of the River Murray, Australia, *Aust. J. Ecol.*, 17, 395 408.

Briggs, S V, Hodgson, P F, and Ewin, P (1994) Changes in Populations of Waterbirds on a Wetland Following Water Storage, *Wetlands (Australia)*, 13, 36 48.

Briggs, S V, Thornton, S A, and Lawler, W G (1997) Relationship between Hydrological Control of River Red Gum Wetlands and Waterbird Breeding, *Emu*, 97, 31 42.

Bruwer, C, Poultney, C, and Nyathi, Z (1996) Community Based Hydrological Management of the Phongolo Floodplain, in *Water Management and Wetlands in Sub-Saharan Africa*, eds M C Acreman and G E Hollis, IUCN, Gland, Switzerland, 199 211.

Contreras-Balderas, S and Lozano-Vilano, M L (1994) Water, Endangered Fishes, and Development Perspectives in Arid Lands of Mexico, *Conserv. Biol.*, 8, 379 387.

Debus, S J (1989) Moree Watercourse, *RAOU Newslett.*, 81, 5.

Denny, P (1994) Biodiversity and Wetlands, *Wetlands Ecol. Manage.*, 3, 55 61.

Drijver, C A and Marchand, M (1985) Taming the Floods: Environmental Aspects of Floodplain Development in Africa, Center for Environmental Studies, University of Leiden, Leiden, The Netherlands.

Drijver, C A and Van Wetten, C J (1992) Sahel Wetlands 2020: Changing Development Policies or Losing Sahel s Best Resources, Center of Environmental Science, University of Leiden, Leiden, The Netherlands.

Froend, R H, Heddle, E M, Bell, D T, and McComb, A J (1987) Effects of Salinity and Waterlogging on the Vegetation of Lake Toolibin, Western Australia, *Aust. J. Ecol.*, 12, 281 298.

Gilbert, C L and Ramey, J B (1995) Riverine Wetlands of the Alberta Prairie, Alberta Parks and Wildlife, Edmonton, Alberta.

Goodrick, G (1970) *A Survey of Wetlands of Coastal New South Wales*, CSIRO Division of Wildlife Research Technical Memorandum Number 5, CSIRO, Canberra.

Hallock, R J and Hallock, L L (1993) Detailed Study of Irrigation Drainage in and Near Wildlife Management Areas, West-central Nevada, 1987 1990. Part B. Effect on Biota in Stillwater and Fernley Wildlife Management Areas and Other Nearby Wetlands, Water-Resources Investigations Report 92-4024B, US Geological Survey, Carson City, NV.

Halse, S A (1989) Wetlands of the Swan Coastal Plain Past and Present, in *Swan Coastal Groundwater Management Conference - Proceedings, Publication No. 1/89*, ed G Lowe, Western Australian Water Resources Council, Perth, 105 108.

Hamilton, S J (1998) Selenium Effects on Endangered Fish in the Colorado River Basin, in *Environmental Chemistry of Selenium*, eds W T Frankenberger and R A Engberg, Marcel Dekker, New York, 297 313.

Hoffman, R J, Hallock, R J, Rowe, T G, Lico, M S, Burge, H L, and Thompson, S P (1990) Reconnaissance Investigation of Water Quality, Bottom Sediment, and Biota associated with Irrigation Drainage in and Near Stillwater Wildlife Management Area, Churchill County, Nevada, 1986 1987, *Water-Resources Investigations Report 89-4105*, US Geological Survey, Carson City, NV.

Hollis, G E, Adams, W M, and Aminu-Kano, M, eds (1993) *The Hadejia- Nguru Wetlands: Environment, Economy and Sustainable Development of a Sahelian Floodplain Wetland*, IUCN, Gland, Switzerland.

Hothem, R L and Welsh, D (1994) Contaminants in Eggs of Aquatic Birds from the Grasslands of Central California, *Arch. Environ. Contam. Toxicol.*, 27, 180 185.

Hruby, T, Cesanek, W E, and Miller, K E (1995) Estimating Relative Wetland Values for Regional Planning, *Wetlands*, 15, 93 107.

Hull, G (1996) Victoria, in *A Directory of Important Wetlands in Australia*, 2nd edition, eds R Blackley, S Usback, and K Langford, Australian Nature Conservation Agency, Canberra, 605 757.

Kimmage, K (1991) The Evolution of the *Wheat Trap*: The Nigerian Wheat Boom, *Africa*, 60, 471 500.

Kingsford, R T (1995) Ecological Effects of River Management in New South Wales, in *Conserving Biodiversity: Threats and Solutions*, eds R Bradstock, T D Auld, D A Keith, R T Kingsford, D Lunney, and D Sivertsen, Surrey Beatty & Sons, Sydney, 144 161.

Kingsford, R T and Thomas, R F (1995) The Macquarie Marshes in Arid Australia and their Waterbirds: A 50-Year History of Decline, *Environ. Manage.*, 19, 867 878.

Knapton, J R, Jones, W E, and Sutphin, J W (1987) Reconnaissance Investigation of the Quality of Water, Bottom Sediment, and Biota in the Sun River Project, West Central Montana, *Water-Resources Investigations Report 87-4244*, US Geological Survey, Missoula, MT.

Lambing, J H, Jones, W E, and Sutphin, J W (1987) Reconnaissance Investigation of the Quality of Water, Bottom Sediment, and Biota in the Bowdoin National Wildlife Refuge and Adjacent Areas of the Milk River Irrigation Project, Northeastern Montata, *Water-Resources Investigations Report 878-4243*, US Geological Survey, Missoula, MT.

Lane, J A K and McComb, A J (1988) Western Australian Wetlands, in *Conservation of Australian Wetlands*, eds A J McComb and P S Lake, Surrey Beatty & Sons, Sydney.

Lemly, A D (1994) Irrigated Agriculture and Freshwater Wetlands: A Struggle for Coexistence in the Western United States, *Wetlands Ecol. Manage.*, 3, 3 15.

Lemly, A D, Finger, S E, and Nelson, M K (1993) Sources and Impacts of Irrigation Drainwater Contaminants in Arid Wetlands, *Environ. Toxicol. Chem.*, 12, 2265 2279.

Lemly, A D, Kingsford, R T, and Thompson, J R (2000) Irrigated Agriculture and Wildlife Conservation: Conflict on a Global Scale, *Environ. Manage.*, 25, 485 512.

Livingstone, A J and Campbell, I A (1992) Water Supply and Urban Growth in Southern Alberta: Constraint or Catalyst? *J. Arid Environ.*, 23, 335 349.

Maheshwari, B L, Walker, K F, and McMahon, T A (1995) Effects of Regulation on the Flow Regime of the River Murray, Australia, *Regul. Rivers Res. Manage.*, 10, 15 38.

Marshall, E (1985) Selenium Poisons Refuge, California Politics, *Science*, 229, 144 146.

McHugh, S (1996) *Cottoning On. Stories of Australian Cottongrowing*, Hale & Iremonger, Sydney.

Morrison, M and Kingsford, R T (1997) The Management of Inland Wetlands and River Flows and the Importance of Economic Valuation in New South Wales, *Wetlands (Australia)*, 16, 83 98.

Muslim, I and Umar, A I (1995) Water Resources and Agricultural Development in Jigawa State, in *The Critical Water Resources of the Komadugu- Yobe Basin*, ed M Aminu-Kano, Proceedings of a Workshop on the Management of the Water Resources of the Komadugu Yobe Basin, National Institute for Policy and Strategic Studies, Kuru, Jos, Nigeria, and Hadejia Nguru Wetlands Conservation Project, Nguru, Nigeria, 49 62.

N diaye, A (1997) Ecological Problems Caused by Dams in West Africa: Case study of Djoudj National Park, Senegal, in *Development Policies, Plans and Wetlands*, eds R C Prentice and R P Jaensch, Proceedings of Workshop 1 of the International Conference on Wetlands and Development, Kuala Lumpur, Malaysia, Wetlands International, Kuala Lumpur, Malaysia, 142 145.

Norman, F I and Corrick, A H (1988) Wetlands in Victoria: A Brief Review, in *The Conservation of Australian Wetlands*, eds A J McComb and P S Lake, Surrey Beatty & Sons, Sydney.

Ohlendorf, H M, Hothem, R L, Aldrich, T W, and Krynitsky, A J (1987) Selenium Contamination of the Grasslands, a Major California Waterfowl Area, *Sci. Total Environ.*, 66, 169 183.

Pearce, D, Markandja, A, and Barbier, E B (1989) *Blueprint for a Green Economy*, Earthscan, London.

Peterson, D A, Jones, W E, and Morton, A G (1988) Reconnaissance Investigation of Water Quality, Bottom Sediment, and Biota Associated with Irrigation Drainage in the Kendrick Reclamation Project Area, WY, 1986 1987, *Water-Resources Investigations Report 87-4255*, US Geological Survey, Cheyenne, WY.

Postel, S L, Daily, G C, and Ehrlich, P R (1996) Human Appropriation of Renewable Fresh Water, *Science*, 271, 785 788.

Pressey, R L (1989) Wetlands of the Lower Clarence Floodplain, Northern Coastal New South Wales, *Proc. Linn. Soc. NSW*, 111, 143 155.

Pressey, R L (1990) Wetlands, in *The Murray*, eds N Mackay and D Eastburn, Murray Darling Basin Commission, Canberra, 167 181.

Pressey, R L and Middleton, M J (1982) Impacts of Flood Mitigation Works on Coastal Wetlands in New South Wales, *Wetlands*, 2, 27 44.

Reisner, M (1986) *Cadillac Desert: The American West and its Disappearing Water*, Viking Penguin Press, New York.

See, R B, Peterson, D A, and Ramirez, Jr, P (1992a) Physical, Chemical, and Biological Data for Detailed Study of Irrigation Drainage in the Kendrick Reclamation Project Area, WY, 1988 1990, *Open-File Report 91-533*, US Geological Survey, Cheyenne, WY.

See, R B, Naftz, D L, Peterson, D A, Crock, J G, Erdman, J A, Severson, R C, Ramirez, Jr, P, and Armstrong, J A (1992b) Detailed Study of Selenium in Soil, Representative Plants, Water, Bottom Sediment, and Biota in the Kendrick Reclamation Project Area, WY, 1988 1990, *Water-Resources Investigations Report 91-4131*, US Geological Survey, Cheyenne, WY.

Smith, P and Smith, J (1990) Floodplain Vegetation, in *The Murray*, eds N Mackay and D Eastburn, Murray Darling Basin Commission, Canberra, 215 228.

Stephens, D W, Waddell, B, and Miller, J B (1988) Reconnaissance Investigation of Water Quality, Bottom Sediment, and Biota associated with Irrigation Drainage in the Middle Green River Basin, UT, 1986 1987, *Water-Resources Investigations Report 88-4011*, US Geological Survey, Salt Lake City, UT.

Stephens, D W, Waddell, B, Peltz, L A, and Miller, J B (1992) Detailed Study of Selenium and Selected Elements in Water, Bottom Sediment, and Biota Associated with Irrigation Drainage in the Middle Green River Basin, UT, 1988 1990, *Water-Resources Investigations Report 92-4084*, US Geological Survey, Salt Lake City, UT.

Thompson, J R (1995) Hydrology, Water Management and Wetlands of the Hadejia Jama are Basin, Northern Nigeria. PhD Thesis, University of London, London, UK.

Thompson, J R (1996) Africa s Floodplains: A Hydrological Overview, in *Water Management and Wetlands in Sub-Saharan Africa*, eds M C Acreman and G E Hollis, IUCN, Gland, Switzerland, 5 20.

Thompson, J R and Hollis, G E (1995) Hydrological Modeling and the Sustainable Development of the Hadejia Nguru Wetlands, Nigeria, *Hydrol. Sci. J.*, 40, 97 116.

Thornton, S A and Briggs, S V (1994) A Survey of Hydrological Changes to Wetlands of the Murrumbidgee River, *Wetlands (Australia)*, 13, 1 13.

Van Ketel, A, Marchand, M, and Rodenburg, W F (1987) West Africa Review: A Comprehensive Listing of Water Management Projects and an Outline of Their Effects on Wetlands, Edwin Report No. 1. Center for Environmental Studies, University of Leiden, Leiden, The Netherlands.

Vinke, P P (1996) The Integrated Development Program for the Left Bank of the Senegal River Valley, in *Water Management and Wetlands in Sub-Saharan Africa*, eds M C Acreman and G E Hollis, IUCN, Gland, Switzerland, 145 153.

Waddell, B H and Stanger, M C (1992) The Influence of Selenium on Incubation Patterns and Nesting Success of Waterbirds at Ouray National Wildlife Refuge, UT, Contaminant Report R6/400S/92, US Fish and Wildlife Service, Salt Lake City, UT.

Wasson, B, Banens, B, Davies, P, Maher, W, Robinson, S, Tait, D, and Watson-Browne, S (1996) Inland Waters, Chapter 4, in *Australia: State of Environment 1996*, eds N Alexander and R Taylor, An Independent Report Presented to the Commonwealth Minister for the Environment by the State of the Environment Advisory Council, CSIRO Publishing, Melbourne, Australia, 1 68.

Wesseling, J W, Naah, C A, Drijver, C A, and Ngantou, D (1996) Rehabilitation of the Logone Floodplain, Cameroon, through Hydrological Management, in *Water Management and Wetlands in Sub-Saharan Africa*, eds M C Acreman and G E Hollis, IUCN, Gland, Switzerland, 185 198.

Willers, B (1994) Sustainable Development: A New World Deception, *Conserv. Biol.*, 8, 1146 1148.

WRIC (Water Resources and Irrigation Commission) (1971) *Water Resources of New South Wales*, Government Printer, Sydney.

Zahm, G R (1986) Kesterson Reservoir and Kesterson National Wildlife Refuge: History, Current Problems, and Management Alternatives, *Trans. N. Am. Wildl. Natl. Res. Conf.*, 51, 324 329.

K

Karst

Karst is landscape or topography dominated by surface features produced by the dissolving of rocks by water passing over and through them and by development of underground passages and cavern systems. Usually the term is restricted in use to carbonate rocks (limestone or dolomite) that are characterized by sinkholes, caves, sinking streams, springs and solution valleys. In this case the dominant dissolving action is by dilute carbonic acid in the water formed by the combination of water with carbon dioxide derived from both the atmosphere and biotic activity in vegetation and soils.

Among the world s most celebrated karst areas are those of Slovenia and Croatia (where the word comes from), the Guilin tower karst along the Li River in southern China and the Mammoth Cave area of the US. However karst underlies many important populated regions, including much of Florida, the Kuala Lumpur area of Malaysia and large areas of Cuba and northern Vietnam.

Karstic landscapes are particularly sensitive to environmental degradation. Problems such as modifications to groundwater supply, contamination of groundwater, and inducement of land instability abound in populated karst regions.

IAN DOUGLAS *UK*

Kovda, V A

(1904- 1991)

Viktor Abramovich Kovda, corresponding member of the USSR Academy of Sciences, Professor, was born in Vladikavkaz into a post office clerk s family. He graduated from Kuban Agricultural Institution (Krasnodar). After post-graduate studies he became a student of Professor Wigner in Switzerland.

Kovda worked in the Dokuchaev Soil Science Institute and in Moscow State University, where he was Chairman of the General Soil Science Department (1953 1981). In 1969 1970 he founded the Agrochemistry and Soil Science Institute in Pushino (Moscow region), where he worked until his death. Throughout his lifetime Kovda worked in many different international organizations, whose focus was centered around the problems of soil science and environment conservation.

He was the Head of the United Nations Educational, Scientific and Cultural Organization (UNESCO) Department of Natural and Exact Sciences (Paris, 1959 1965). During this period he raised and solved some important global scientific problems (soil reclamation, aridization, relationships between the biosphere and human activities, and construction of a soil map of the world). In 1968, Professor Kovda was elected President of the International Soil Science Society. From 1973 up to 1976 he was President of the Scientific Committee on Problems of the Environment (SCOPE).

Due to his encyclopedic knowledge of foreign and Russian literature and his numerous trips abroad, Kovda gained a deep insight into the problems of global soils and biosphere changes as a consequence of anthropogenic influences and global climatic change.

Kovda played a leading part in the world scientific community directed towards understanding the role of soils and soil cover in the planetary biosphere, people s responsibility for maintaining biosphere functioning, and the need to handle nature conservation problems on a global scale.

Kovda is one of the founders of the contemporary theory of desertification, aridization and soil salinity. He developed the scientific principles of the origin of saline soils, he solved both theoretical and practical problems of chernozem irrigation, and he worked out the concept of the prevalence of paleohydromorphism on the Great Plains and its influence on the formation of soil cover. Following V I Vernadski, he stimulated the world community to

study the whole complex of environmental conservation problems.

Kovda demonstrated for the first time that aridization and desertification are global processes. In 1974 at a SCOPE meeting, Kovda stressed that the problem of the biochemical cycles of chemical elements, both natural and anthropogenic, is very important. In his activities Kovda constantly adapted fundamental scientific developments to the needs of practical users.

FURTHER READING

Kovda, V A (1973) *The Principles of Pedology*, 1st and 2nd editions, Nauka, Moscow (in Russian and Chinese).

Kovda, V A (1980) *Land Aridization and Drought Control. Westview Special Studies in Natural Resources and Energy Management*, Westview Press, Boulder, CO.

Vasil yevskaya, V D, Evdokimova, T I, Zimovets, B A, and Kydeyarov, V V (1996) V.A. Kovda Scientist, Teacher, Citizen on the Occasion of the Ninetieth Anniversary of His Birth, *Eurasian Soil Sci.*, 2, 4 7.

VERA VASLL EVSKAYA *Russian Federation*

Kuwait, Oil Fires

see Oil Fires: Kuwait (**Volume 3**)

L

Lahars

Lahar is an Indonesian word for a rapidly flowing mixture of rock debris and water, a volcanic mudflow or debris flow that originates on the flanks or slopes of a volcano particularly in the wet tropics. As a hazardous volcanic event, lahars occur in the presence of large quantities of water in association with any type of volcano. Lahars form in a variety of ways: from crater lakes; the rapid melting of snow and glaciers; intense rainfall on loose volcanic rock deposits; breakout of a lake dammed by volcanic deposits; as a consequence of debris avalanches and predominantly by the rapid melting of snow and ice by pyroclastic flows. Once a pyroclastic flow mixes with a stream or rainwater mixes with a still hot pyroclastic material a hot lahar forms. Lahars consist of fluids of water and volcanic particle mixtures comprised of particles 50% less than sand-sized with the remainder ranging from clay-size to boulder-size. Lahars travel great distances ranging up to 300 km traveled mostly in valleys and have a viscous consistency of wet concrete capable of a range of velocity activity. Their spontaneous but disastrous impact can be deadly. The velocity of a lahar depends on the slope and water content; the lower the viscosity, the faster it flows. They attain speeds of $1\,\mathrm{m\,s^{-1}} = 2{\bullet}55\,\mathrm{m\,h^{-1}}$ and on steep slopes velocity ranges from $1.3\,\mathrm{m\,s^{-1}}$ to $40\,\mathrm{m\,s^{-1}}$. Flows of lower viscosity are therefore more dangerous because of the drowning risk compared to high viscosity flows that are only capable of removing vegetation and causing structural failure. Lahars become increasingly diluted with water after interacting with pyroclastic surges downstream from snow and ice and transform into regular floods. This was experienced during the eruption of Nevado del Ruiz, Armero, Colombia in 1985 that caused the death of 21 000 people. This first occurred at Mount St Helen s, US, where hot pyroclastic surges transformed to lahars, to hyperconcentrated streamflow and then to normal floods. Besides the initial destruction and altered river courses, lahars subsequently leave deposits that impede local river channels. These deposits of volcanic sediment require dredging to allow navigation to resume and are often used as construction materials. Lahars do not necessarily happen at the time of the eruption, they may occur many years later. Unfortunately, lahar prevention is only possible by use of barriers similar to those used for lava flows, this provides only minimal protection. Some Indonesian villages position artificial mounds so that people can quickly access higher ground away from the flow. However, this preventive measure requires an adequate warning system. To determine the threat of mudflows, rain gauges or flow sensors are used to monitor the amount of precipitation.

REFERENCES

http://www.sunysuffolk.edu/~mandias/honors/student/volcano/global_warming. html.
http://www.esci.keele.ac.uk/courses/geol317/volcano.html.
http://www.btinternet.com/~mike.ferris/volcano.htl.
http://volcanoes.usgs.gov/Hazards/Effects/LaharEffects1_caption.html.
http://magic.geol.ucsb.edu/~fisher/lahar.htm.

MARY KEARNEY *Canada*

Lake Baikal

Anson W Mackay
University College London, London, UK

Lake Baikal, located in southeast Siberia, is one of the most unusual freshwater ecosystems in the world. It formed within a series of rift zones over 30 million years ago, making it by far the oldest living lake. Lake Baikal is also the deepest lake in the world, with a maximum depth of 1642 m. This great depth ensures that the lake contains the world's largest volume of surface freshwater, some 20% of global resources, and 80% of resources found within the former Soviet Union (fSU). Lake Baikal has a remarkable freshwater ora and fauna: over 2500 animal and plant species have been identi ed; of which it is estimated over

75% are endemic (that is, they are found nowhere else in the world today). Biodiversity levels in the lake are extremely high, principally because of its great age allowing evolutionary processes to proceed, but also because the entire water column is saturated with oxygen, allowing extensive, deep-water, endemic faunal communities to exist.

Over the last 30 years, Lake Baikal has been one of the fSU's environmental causes celêbres, due to both increasing exploitation of resources within the lake, and perhaps more importantly, increasing contamination of its water by pollution from industrial growth in the catchment. In recognition of Lake Baikal's global ecological signi cance, the United Nations Educational, Scienti c and Cultural Organization declared it a World Heritage Site in 1996.

Lake Baikal consists of three basins of differing geological age within a series of rift zones. The basins are separated by two underwater highs, the Selenga Shallows in the south (separating the south and middle basins), and the Academician Ridge, separating the middle and north basins (Figure 1). The rift zones are associated with intense seismic activity, and are currently expanding at a rate of c. 20 mm year^{-1} (SB-RAS, 1996). Accordingly, Lake Baikal also occupies the deepest continental depression at 1164 m below sea level (Table 1).

In terms of area, Lake Baikal is the world s seventh largest freshwater lake, but it contains the world s largest volume of freshwater (over 23 000 km^3) because of its great depth (Kozhov, 1963). Lake Baikal is, therefore, an important freshwater resource. The majority of water enters the lake through some 365 rivers, although there is only one outflow from the lake, the Angara River in the south basin. The Angara flows northwards through the heart of Russia, before it eventually merges with the Yenisei River and then enters the Arctic Ocean. Two-thirds of the water entering Baikal is contributed by just three rivers: the Selenga, the Upper Angara and the Barguzin Rivers which yield 47, 13 and 6% of water input, respectively. The largest, the Selenga River, has its source in the Mongolian highlands.

Lake Baikal is strongly influenced by a continental climate: mean daily temperatures in January fall to an average of c. $-25\,^{\circ}$C, whilst in July they are in the region of c. $+18\,^{\circ}$C. Winters are long, cold and dry, whilst summers are relatively hot, short and wet. The surface of the lake freezes every winter (late December in the north basin to January in the south basin) until the following May (in the south basin), and June (in the north basin). Ice thickness during this time reaches between 80 120 cm. Ice formation and cover are affected by local and regional climatology, although influences from the North Atlantic Oscillation have also recently been found. The lake itself promotes local temperature differences, with the formation of a local climate within 25 km^3 of its shores, with slightly warmer winters and slightly cooler summers.

Table 1 Major morphological characteristics of Lake Baikal

Length (km)	636
Mean width (km)	49.3
Total area (km^2)	315 000
Maximum depth (m)	1642
Mean depth (m)	738
Volume (km^3)	23 000

The catchment area of Lake Baikal is vast (over 540 000 km^2), spanning south-eastern Siberia and northern Mongolia (Kozhov, 1963). It has recently been split into zones, defined by both physical and economic geographies, which has been important in securing World Heritage Site status (SB-RAS, 1996). It is also a UNESCO Biosphere Reserve (*see* Biosphere Reserves, Volume 2). The zones are defined as: (i) the central zone, which includes the lake and natural parks and reserves around the lake; (ii) the buffer zone, which includes the rest of the physical catchment area both within Russia and Mongolia; and (iii) the zone of atmospheric impact, which occupies an area immediately outside the physical catchment area to the west and north-west of the lake. All three together comprise the Lake Baikal Region, which lies at the junction between several biogeographically distinct regions (e.g., taiga, steppe, tundra and high mountain deserts). Approximately 10% of the 2500 species of vascular plants in this region are endemic, whilst the faunal communities are even more diverse, and in the mountain regions at least, endemism is especially common (SB-RAS, 1996). The biota in Lake Baikal is derived from three principal sources: (i) local endemics; (ii) marine immigrants; and (iii) Asian immigrants (Kozhova and Izmest eva, 1998).

Lake Baikal is home to the world s only freshwater seal, or nerpa (*Phoca sibirica*), which is also Baikal s only mammalian endemic. It is thought that the nerpa are descended from the ringed seal *P. hispida*, (i.e., that they have marine ancestry) and have been inhabitants of Lake Baikal since the last Ice Age (Kozhov, 1963). Population levels are in excess of about 70 000 individuals, of which up to 6000 are culled annually. Adults generally live to an age of 50+ years. In recent decades, the fluctuating populations of seal have been the focus of environmental concern, outlined in more detail below. The majority of species within the fish population are also endemic, although there have been some recent introductions of more common species such as perch and roach. Some fish species are extremely unusual and have very specific adaptations to their environment. For example, the *Abyssocottus* genus only exists in the very deep waters of Baikal, whilst another genus, the *Comephorous*, is specifically adapted to live in the open water, has large pectoral fins, and a scaleless, translucent

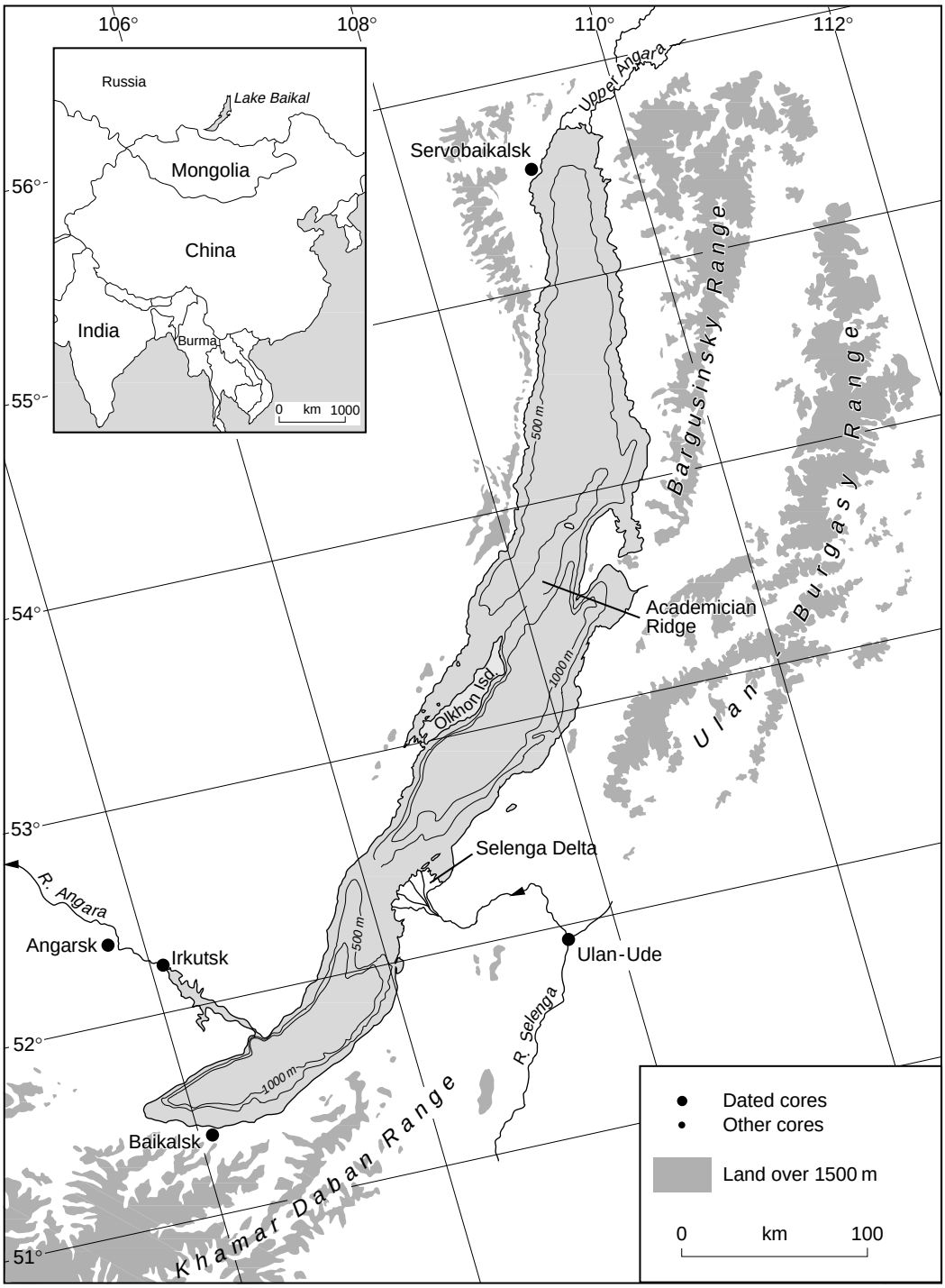


Figure 1 Map of Lake Baikal and its immediate catchment. Major towns are indicated, as are the rivers, Selenga and the Angara, and the two underwater highs, the Selenga Delta and the Academician Ridge

body and reduced ossification. Relatively few species of fish are commercially exploited in Lake Baikal, the most important one being the omul, *Coregonus autumalis migratorius*.

Levels of endemism increase further in the invertebrates, especially the crustaceans (e.g., gammarids), the molluscs (e.g., gastropods), the annelids (e.g., polychaetes) and free-living flat worms (e.g., Platyhelminthes).

Along the coastlines of Baikal, there are few species of higher plants, except for local occurrences of *Myriophyllum*, *Potomogeton*, and *Elodea canadensis*, a recent

introduction. Lower plants, especially the algae, are more diverse, of which perhaps the diatoms (Bacillariophyta) are the most important and interesting group. Diatoms are unicellular plants encased by ornamented silica shells, or valves. The hundreds of endemic diatom species in Baikal ensure that they are important primary producers in the lake, comprising approximately 57% of the lake s planktonic community. Diatoms are by far the dominant producer of biogenic silica, contributing around 98% of all sedimentary silica.

Lake Baikal is an important site for studying past global climate change for a number of reasons: (i) it is sensitive to changes in insolation patterns caused by orbital forcing; (ii) it is close to the boundaries of two globally important weather systems (the Siberian High and the Asiatic Monsoon); and (iii) Lake Baikal contains one of the few very long lacustrine (i.e., lake) records of environmental change that is comparable to marine records. The Baikal rift zone has never been glaciated, and since its formation, over 7500 m of sediment have accumulated. The sediments are composed largely of biogenic silica (from diatoms) and clastic and fluvial input from Baikal s catchment. These sediments contain the remains of pollen grains and fossil diatoms (amongst other environmental proxies) going back millions of years, which can be used by scientists to reconstruct past environmental change in the lake and its catchment. For example, pollen analysis can tell us when the catchment of the lake was covered by forest, e.g., in warmer climates such as the present-day one, or when steppe vegetation was dominant, indicative of cooler climate. The diatoms can provide information on past climate change around Lake Baikal from as little as c. 150 years ago (e.g., Mackay *et al.*, 1998) to as far back as 5 million years ago (e.g., Williams *et al.*, 1997). Past climate in the Baikal Region has responded to both global changes relating to the earth s orbit, and to changes in the Earth s ice-cover, despite the continentality of its location. These changes not only manifest themselves on biological activity in the lake, but have also had direct influences on lake-water levels, which have risen and fallen in the past by hundreds of meters.

Human influence around Lake Baikal has been evident since at least the Neolithic period, as small populations hunted nerpa on a seasonal basis. Asian populations, such as the Buryats and the Mongols have also existed around Baikal for centuries, but the first Russian settlers did not arrive until the mid-17th century. However, only since the 1950s have environmental concerns for the lake developed into national and international issues. These concerns focus mainly on point-source pollution from two paper and pulp mills (built in the 1960s 1970s) in the towns of Baikalsk in the southern basin, and Selenginsk on the Selenga River. During the 1980s, the environmental campaign against these industries gathered pace, and extended to include logging in the catchment, over-fishing, the dumping of effluents from other factory waste products and partly treated sewage, and atmospheric contamination from recent increases in industrial emissions. The largest sources of sulfur dioxide (SO_2) and particulates are from oil-fired power stations at the town of Angarsk. It has recently been established that pollutants indicative of fossil fuel combustion, i.e., spheroidal carbonaceous particles (SCPs) have increased in the Baikal region over the last 50 years (Rose *et al.*, 1998). SCPs are only produced by high temperature combustion, and are, therefore, unambiguous indicators of anthropogenic contamination. Concentrations of these particles in Baikal show spatial differences, with highest concentrations found in the southern basin, where industrialization and urbanization are highest. Small increases are also evident in the north basin, where, in recent decades, new centers of urbanization have been developed, e.g., the town of Severobaikalsk. These increases closely mirror levels of fossil fuel combustion in the catchment. As well as SCPs, heavy metal input into the lake is also increasing, such as quantities of lead, copper, zinc and vanadium. However, attributing heavy metal concentration increases solely to anthropogenic contamination is more difficult than for SCPs, as these metals are also produced naturally in the catchment, and will increase in the lake as the catchment is disturbed. Current levels of metals in Baikal are greatest in the southern basin, but the effects of these pollutants are still debated. For example, it is well known that toxic levels of mercury and cadmium can cause immuno-suppression in animals, and whilst these poisons can be detected in Lake Baikal, their concentrations are still low, and are similar to those of Northern Hemisphere background sites, such as the Arctic region.

Perhaps the greatest threat to the lake s biota, especially the nerpa, lies with increasing levels of persistent organic pollutants (POPs), polychlorinated biphenyls (PCBs) and dichlorodiphenyltrichloroethane (DDT). It has been shown that the Baikalsk mill caused locally enhanced levels of persistent organic chlorine compounds in the southern basin. Trends are complicated however. Concentrations of POPs in the air above Baikal are similar to those in remote regions, such as the Arctic, although levels in fish are comparable to those in the Great Lakes in North America, whilst concentrations in the nerpa were similar to those in animals found in, e.g., more polluted European waters, such as concentrations in the Caspian seal (Kucklick *et al.*, 1994). In the late 1980s, thousands of nerpa died from a morbillivirus, and environmentalists suggested that the seals immune systems had been compromised by increasing pollution levels, including POPs and heavy metals. On the other hand, many other scientists firmly believe that changes in the biota of the lake can be explained by natural variation, and during the 1990s, international research programs were established to collate scientific evidence on

the past and present environmental state of Lake Baikal (under the auspices of the Baikal International Center for Ecological Research).

Bioaccumulation of organic chlorine compounds, including DDT and PCBs, in the nerpa is clearly a problem, even though the physical environment of Baikal (water and sediments) contains relatively low levels of pollutants. A direct link has not yet been made between increasing organochlorine pollution and morbidity and mortality in *P. sibirica*, although mechanisms have been proposed. Much more research needs to take place, and it is encouraging in the meantime that several measures to reduce pollution in the lake have already been taken, such as converting the two pulp and paper mills into closed systems.

Further, pollution from other industrial and domestic wastes, increasing eutrophication (nutrient enrichment), uncontrolled fishing and logging of the catchment were also linked to declines in fish growth rates, the disappearance of endemic animals, and a general influx of more cosmopolitan flora and fauna. For example, *C. autumalis migratorius* is non-migratory and its numbers were severely depleted since the 1950s due to over-fishing, but since the introduction of catch controls, populations have increased significantly. One potential threat from increasing pollution of the Baikal ecosystem is the decline of numbers of endemic plant and animal species as a whole, due to deteriorating water quality. Diatoms, however, provide an excellent record of past environmental conditions, as they are very sensitive to changes in water quality (e.g., pH and nutrient enrichment) and climate. By examining diatoms in surface sediments, Mackay *et al.* (1998) found no evidence that the endemic flora in the lake had been affected by pollution. Therefore, there has been no deterioration in water quality due to pollution across the main body of water in Lake Baikal, although there is evidence for very localized eutrophication in the shallow waters of the Selenga Delta region.

REFERENCES

Kozhov, M (1963) *Lake Baikal and its Life*, W Junk, The Hague.

Kozhova, O M and Izmest eva, L R (1998) *Lake Baikal: Evolution and Biodiversity*, Backhuys, Leiden.

Kucklick, J R, Bidleman, T F, McConnell, L L, Walla, M D, and Ivanov, G P (1994) Organochlorines in the Water and Biota of Lake Baikal, Siberia, *Environ. Sci. Technol.*, 28, 31 37.

Mackay, A W, Flower, R J, Kuzmina, A E, Granina, L Z, Rose, N L, Appleby, P G, Boyle, J F, and Battarbee, R W (1998) Diatom Succession Trends in Recent Sediments from Lake Baikal and their Relationship to Atmospheric Pollution and to Climate Change, *Philos. Trans. R. Soc. Lond.*, 353, 1011 1055.

Rose, N L, Appleby, P G, Boyle, J F, Mackay, A W, and Flower, R J (1998) The Spatial and Temporal Distribution of Fossil-fuel Derived Pollutants in the Sediment Record of Lake Baikal, Eastern Siberia, *J. Paleolimnol.*, 20, 151 162.

SB-RAS (1996) *The Present and Future State of the Lake Baikal Region, Part 1*, eds V E Ermikov, N Meshkova, N Prittwitz, and O Podoinitsyna, Studio Design Folio, Novosibirsk.

Williams, D F, Peck, J, Karabanov, E B, Prokopenko, A A, Kravchinsky, V, King, J, and Kuzmin, M I (1997) Lake Baikal Record of Continental Climate Response to Orbital Insolation During the Past 5 million Years, *Science*, 278, 1114 1117.

Lake Victoria, Water Quality Issues

see Lake Victoria (*Case study*, **Volume 4**)

Land Degradation in the Mediterranean

John B Thornes
King's College, London, UK

Land degradation is the reduction of the land's capacity to grow agricultural products. It mainly takes the form of soil erosion. Globally it is an important process that has accelerated over the centuries as pressure on the land for production has increased and as agricultural technology has changed. It is intimately linked to off-site effects such as ooding, reservoir siltation and water shortage and is often viewed as an integral part of the wider problem of deserti cation.

In Europe, lands around the Mediterranean Sea are most affected, especially the areas with low rainfall totals, high summer temperatures and seasonally concentrated rainfall. These climatic factors, which vary greatly in space and time, lead to a poor vegetation cover, particularly when they are linked to land husbandry practices unsuitable for the environment. Poor soils and steep mountain slopes exacerbate the problem. The changes in land management over the last century, related to North European markets, have also accelerated it.

The major approach to mitigating land degradation is to reconstruct the land cover. Forestry, the generally preferred option, may not be an optimal solution. Aside from the technical xes, such as terracing, check dams and reservoirs, it is the human dimensions that have to be addressed by policy makers.

LAND DEGRADATION AND DESERTIFICATION

Land degradation is the progressive reduction of the land s capacity to grow agricultural products. It comprises two main processes, the loss of the soil itself that acts as a framework for plant growth, and the loss of fertility. The latter normally accompanies the former because most of the nutrients are stored near the soil surface. This means that a modest amount of erosion may lead to a substantial loss of fertility. Soil erosion has other impacts. The lost material is transported by surface water and rivers to off-site locations where, by its accumulation, it may cause other problems. For example, it may cause siltation of reservoirs, burial of crops or of irrigation equipment. Accumulation of deposits in channels increases the risk of flooding and may cause direct damage to residential property. Soil degradation is inseparably linked to hydrology. Soil acts as a sponge for rain and, if its storage capacity is reduced by soil erosion, more of the rain is lost to runoff and therefore more erosion occurs. Therefore, there is a vicious circle of feedback. The more erosion that occurs, the more that is expected to occur. The end of this cycle is only reached when all the soil has been stripped and the surface largely compromises bare rock, from which rainwater runs off more or less instantaneously. Not only does more water run off into rivers, but it also gets there more quickly. Water on the surface can reach speeds of $1\,\mathrm{m\,s^{-1}}$. Water traveling through the soil may reach $1\,\mathrm{m\,day^{-1}}$. Water traveling through rock may take one day to travel 1 cm, depending on the type of rock.

Deserti cation is land degradation due to climate and human causes, but this definition, from the United Nations (UN), includes water and vegetation in the term land, so it refers not only to soil erosion, though this is generally thought of as the main indicator of desertification. Countries and even international agencies have often tended to exaggerate the extent of land degradation for political purposes, and there has been a lively debate about this in recent years. Thomas and Middleton (1994) claimed that desertification was a myth perpetuated by United Nations Environment Programme (UNEP), to draw attention to its International Convention, though they themselves were contributors to UNEP s *Atlas of World Deserti cation*. It is, however, extremely difficult to obtain accurate estimates of erosion because of the huge variety of conditions and the crude methods of its measurement. Often rates of erosion are estimated from small plots inadequate in size. Sometimes they are very few in number so that it is impossible to obtain a statistically representative sample. Sometimes the rates are generated by simulated artificial rainfall that fails to replicate the characteristics of real rainfall. The plots have strong biases arising from their construction, with water and sediments leaking down the cracks where the metal edge has been installed. Another approach is to estimate the total load of sediments carried by a river from a basin. Potentially this is the more accurate approach. It involves multiplying the discharge (flow rate of the river) by the concentration of sediment in the flow. There are many other techniques, for example measuring the depth of exposure of tree roots that appear to mark former ground surface; or fence posts, whose original placement and age is known. Not all systems are so crude and recently the concentration and profile of radioactive fall-out in the soils has been used. Generally, however, rates of erosion are extremely difficult to obtain with much confidence. What is clear is that they vary greatly over very small distances. As a consequence, estimates are made with models ranging from the empirical universal soil loss equation (USLE) to sophisticated computer models such as WEPP (Water Erosion Point Predictor) and EPIC (Erosion Point Impact Calculator). These models have been developed and tested mainly by the United States Department of Agriculture, but despite the vast resources involved in their production they can still only give rough estimates that are little better than those obtained from field experiments. So we really do not know how much or where or when land degradation has occurred.

Computer models (Kirkby and Neale, 1987) and field data converge on the view that Mediterranean regions have some of the highest erosion rates in the world. The general plot of erosion against mean annual rainfall adjusted for evapotranspiration is shown in Figure 1. Although this is for North American data, the overall relationship represents causal factors and should hold in the Mediterranean as well (see discussion). The curve is highest where effective rainfall is about 300 400 $\mathrm{mm\,year^{-1}}$, typical of large areas of the drier parts of the dry lands surrounding the Mediterranean Sea. The argument runs as follows. In drier areas (to the left of the peak) there is insufficient rain to create runoff. At the extreme, with no rainfall, there can be no runoff. As the rainfall increases, erosion increases relatively quickly, peaking on bare soils. To the right of the peak the continuing increase in mean annual rainfall means that vegetation (which suppresses erosion) grows faster. The increase in erosion produced by more rainfall cannot keep up with the decrease caused by more vegetation.

One important omission from this general approach is the effect of seasonality. One of the strongest features of Mediterranean environments is the strong contrast between summer and winter. Summers are hot and dry and winters are cool and wet. The stronger the seasonality, the stronger is the erosion. So much so that, as Kirkby and Neale (1987) show with model studies, where most of the rainfall falls in one month, the curve continues to rise as rainfall increases beyond the Mediterranean peak, so that with strong seasonality, the wetter areas of the Mediterranean can still have high erosion rates (see Figure 2).

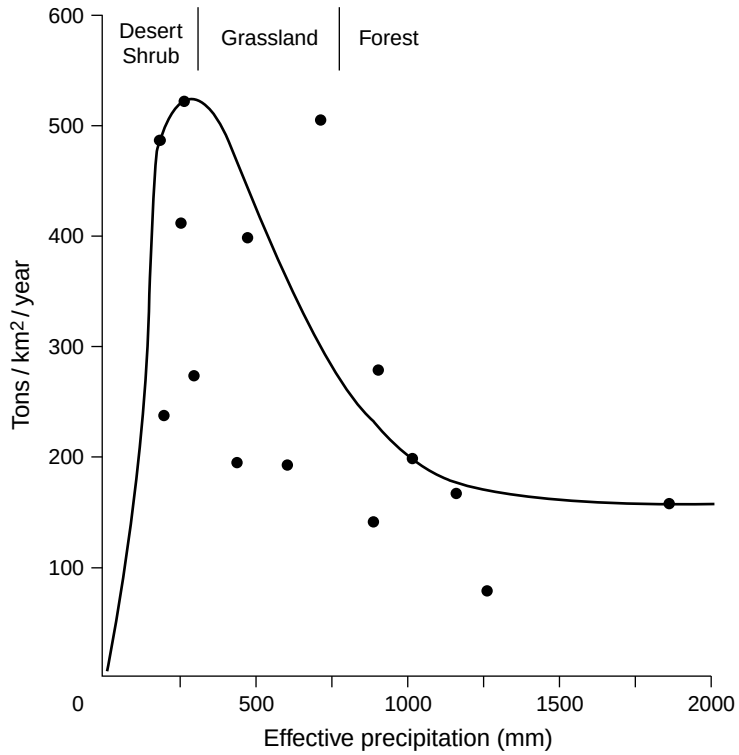


Figure 1 A general plot of erosion against mean annual rainfall, adjusted for evapotranspiration, based on data from the United States according to Langbein and Schumm (1958)

Rainfall intensity is the amount of rain that falls in a given period of time, measured in mm hour^{-1}. The more intense the rainfall, the more quickly the soil fills up and water starts to flow over the surface, because the soil is saturated (*saturated overland ow*). Alternatively, if the rain is coming down faster than it can infiltrate into the soil, it again is forced to flow over the surface (*Hortonian overland ow*). Consequently storms with higher intensity produce more erosion. It is generally true that the kinetic energy of rainfall is related directly to intensity (higher intensity, higher kinetic energy). High kinetic-energy rain detaches the soil particles by rainfall impact, a process called *splash erosion*. The splashed particles are thrown into the overland flow and carried away. Winter storms often have low intensities (20 30 mm hour^{-1}). By contrast summer storms have highest intensities (say 60 mm hour^{-1} in about 1 year in 5). Some of the most extreme intensities are found in the autumn storms when depressions, formed over the Mediterranean Sea, move inland and the air is forced to climb high mountains near the coast (orographic lifting). The intensities may reach 140 160 mm hour^{-1} and the storms may be sustained for many hours.

The strong seasonality in rainfall leads to sharp seasonal variation in soil moisture as shown in Figure 3. Throughout the winter, soil moistures are typically high, though they still respond to individual rain events, as indicated by the vertical black bars. After the spring rains, there is a dramatic fall in soil moisture as the summer heat begins to develop. Soil moisture remains low throughout the hot dry summer months, rising sharply again with the autumn rains, to winter high values.

LAND DEGRADATION AND VEGETATION

A direct consequence of this regime is that plants (agricultural as well as wild) have difficulty in producing enough growth to survive the summer drought. Roughly speaking, the amount of visible plant (above ground biomass) is directly proportional to rainfall in dry areas. So vegetation in Mediterranean drylands, that typically have 300 500 mm year^{-1} of rainfall, is quite sparse. Moreover, the vegetation comprises mainly bushes which can tap water at great depths and are ecologically suited to the overall environmental conditions, including human activity. As a result, when land is abandoned, the recovery of vegetation is often in the form of bushes, low trees and herbaceous shrubs. This is called *matorral* or *monte* in Spain, *maquis* in France, *macchia* in Italy and *phrygana* in Greece. This vegetation has a variable impact. Bushes may concentrate overland flows to cause higher rates of erosion and tiny stream channels called *rills*. They also provide a good general cover, sometimes almost as good as small pine

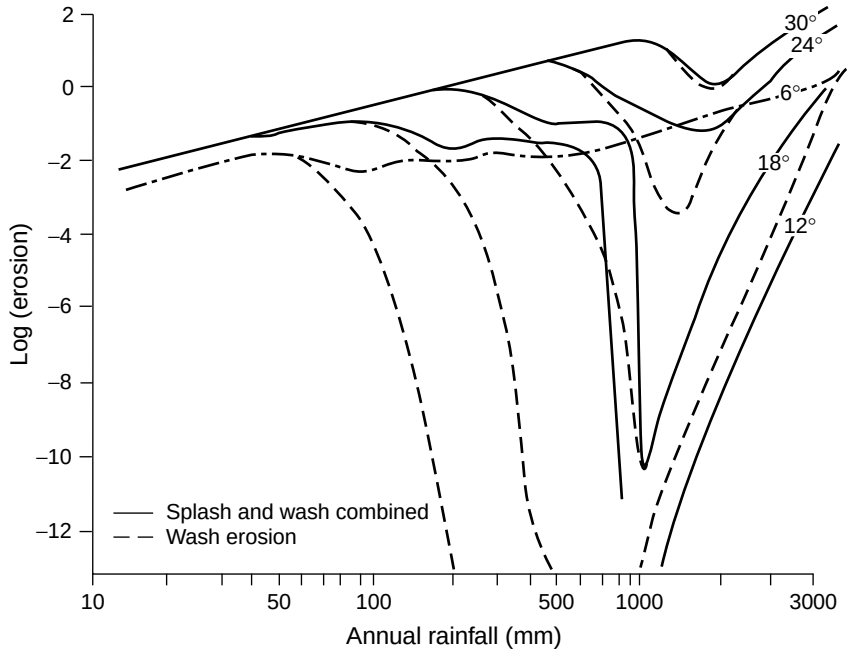


Figure 2 Kirkby and Neale's model output on the relationship of erosion and rainfall that emphasizes the effect of the concentration of rainfall into certain months of the year and the direct effects of temperature (T °C). *Wash* refers to erosion caused by water running over the soil surface. *Splash* refers to the process whereby soil is detached by the impact of falling raindrops. *Erosion* is the combination of the detachment of soil by splash and wash and its transportation away from the point of detachment. The curves are for different slope angles

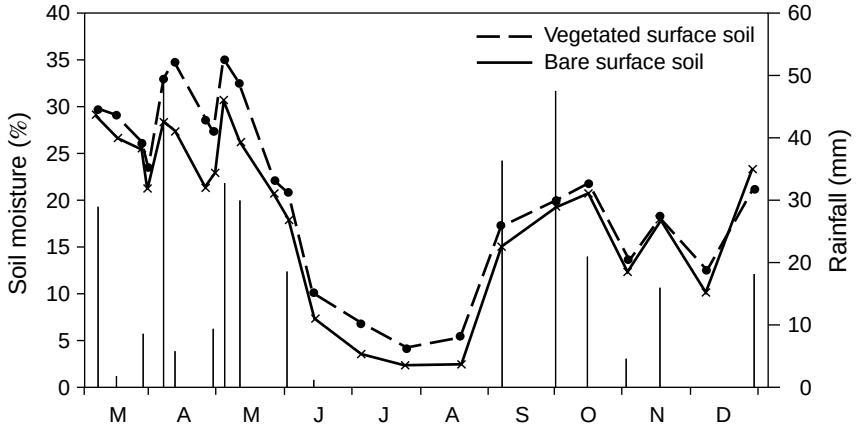


Figure 3 Soil moisture variations through time in the dry Mediterranean (Murcia Spain, 350 mm rain year^{-1}) based on data collected by Professor F Lopez-Bermudez, University of Murcia, Spain

trees, and so inhibit other forms of erosion. The vegetation breaks up the raindrops and reduces their kinetic energy. The soil organic matter creates tubes and holes in the soil. This encourages infiltration and reduces runoff.

The relation of erosion and runoff to vegetation has been established and re-affirmed for over 30 years and is shown in Figure 4. The lower curve shows that, as plant cover increases, soil erosion falls very dramatically so that, as a rough rule of thumb, by the time that plant cover has reached 30%, soil erosion has fallen to 30% of the bare soil value. This shows that soil erosion is very sensitive to plant cover between 0–30%. This relationship is quite sensitive to rainfall intensity and the curve is most dramatic for LI and less dramatic for HI. As a result, quite moderate amounts of biomass can be as effective in reducing erosion as large amounts of biomass. Matorral or grass may be as good as trees. Conversely, anything that reduces cover in the 0–30% danger zone for vegetation can have a quick and dramatic impact on erosion rates.

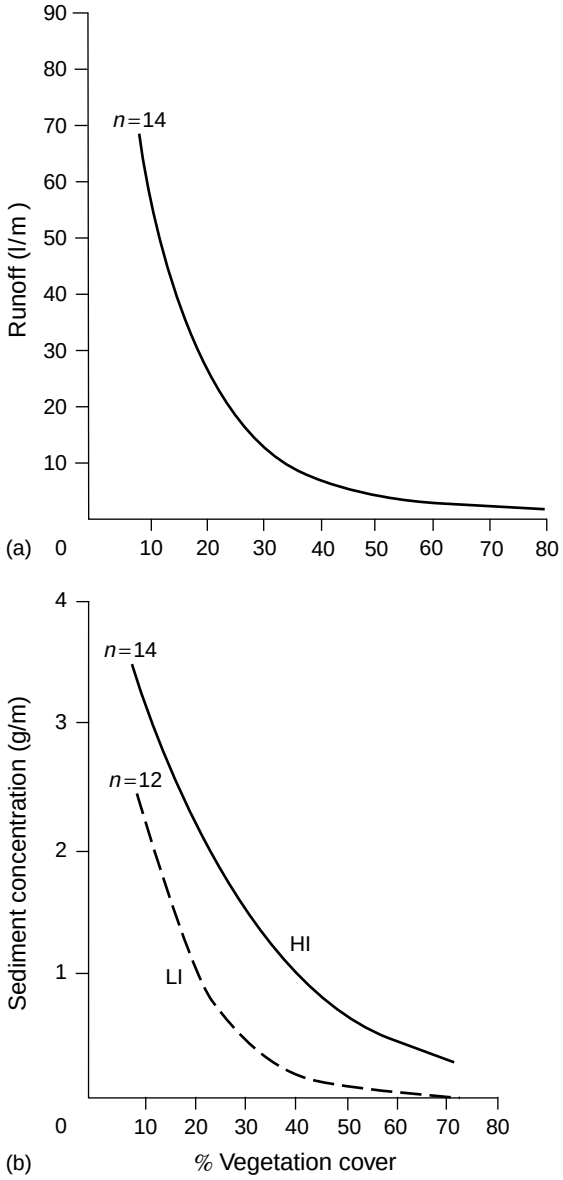


Figure 4 (a) the relationship of vegetation (x, cover percent) to erosion (y, proportion of bare soil erosion rate) according to Elwell and Stocking (1976). (b) The effects of high intensity (HI) and low intensity (LI) rainfall, as well as vegetation cover, according to Francis and Thornes (1990). The horizontal axis is the percentage of the ground covered by plants when looked at in plan view (i.e., from above). The vertical axis is the erosion rate as a fraction of that on bare soil

LAND DEGRADATION AND CLIMATE CHANGE

The impact of climate change on degradation is very complex. Given the estimated 3 °C warming, the carbon dioxide (CO_2) doubling and the expected reduction in winter rainfall that coupled models anticipate for the Mediterranean, Diamond and Woodward (1998) have modeled the expected impact on semi-natural vegetation in the Iberian Peninsula and have estimated the likely consequences for soil degradation. They indicate a marked reduction in vegetation biomass for bush and grass by 2050 AD, especially in areas where the vegetation is already quite marginal (those that have less than 30 mm of mean annual rainfall and more than seven months year^{-1} dry). These are notably in the provinces of Almeria, Jaen, Murcia and Granada. The effects on both bush and grass are much less dramatic in the more humid areas, such as the Autonomous Region of Catalonia and the Basque country of the northwest. In the dry areas, small changes in the already marginal vegetation cover must lead, according to the arguments made earlier, to substantial increase in runoff and erosion as the cover falls to less than the 30% critical amount. On the other hand, the response is likely to be complex, with a boom and burst type oscillation as the rainfall varies enormously from year to year and erosion and plant growth compete for dominance.

THE HUMAN IMPACT ON LAND DEGRADATION THROUGH THE CENTURIES: THE IBERIAN PENINSULA

Early man seems to have had little impact on erosion in the Western Mediterranean. Research suggests that the main features of erosion that we see today (the gully systems) were already in existence in Bronze Age Spain. Therefore, the widely held idea that erosion is mainly due to forest clearance has to be questioned. However, six millennia of traditional Mediterranean farming have resulted in significant amounts of soil erosion since prehistoric times. The traditional combination of bush and tree crops (vines and olives), animal grazing (goats and sheep) and cereal crops (wheat and barley) have taken their toll. The tree cropping is usually accompanied by the removal of weeds and grass between the trees to conserve water and nutrients and this reduces the cover. Grazing directly takes off the protective leaves and compacts the soil under the hooves of animals that may be in large flocks. In the mesetas of Central Spain, the ancient system of *mesta*, herding sheep and cattle along public trails, could lead locally to complete devastation of soils.

A major shock to the west Mediterranean system came with the Moorish invasions at the end of the first millennium, which brought citrus fruit cultivation and traditional irrigation practices. Irrigation had already been introduced in the form of Roman aqueducts, but Moorish irrigation meant that an economy already based on scarce water resources was moved further towards the margin and that the extensive agricultural economy was replaced by a more intensive one. The Moorish concept of the landscape as

a garden, however, meant that there was an emphasis on better land husbandry. The terrace as a device for water and soil conservation was already widespread, extending the possibilities for agriculture. As the Moors were expelled in the late 15th century, new immigrants from other parts of the Mediterranean took over the land and the struggles for production and ownership commenced and continued for half a millennium.

By the end of the 18th century in Spain, major land holding changes were in progress. The Royal Decree of 1793, applicable in Extremadura, attempted to increase cereal output by allowing farmers to plow up their pasturelands and saw the beginning of the end of the mesta system. Between 1798 and 1808, the first sales of church properties occurred. These changes were accelerated by the invasion of the French army that brought about major changes in land husbandry in the Spanish landscape. The population rise (12.5% between 1780 and 1820) forced the pace of agricultural production. By 1820 this surge had ended, as demonstrated by the increased exportation of agricultural products. Many more changes liberated land in the 19th century. Some of these changes had profound consequences for land degradation, particularly the redistribution of church lands, the best of which went to capable wealthy large landowners, whilst the poorest went into the public domain as common land that continued to be degraded by over-grazing well into the 20th century.

The next huge changes occurred in the middle of the 20th century. Mountain areas under vines, such as Eastern Andalucia, were impoverished by pestilence and emigration, so that land care became almost non-existent and terraces began to fail from lack of maintenance. The initiation of serious degradation in the Sierra Nevada can be traced back to the early years of the 20th century.

In the better agricultural areas, the pressure to increase production with the expansion of regional and global markets, coupled with the rise of tourism with its insatiable demand for water and market garden products, led to strong intensification of agriculture in the 1950s. Other technological changes mechanization, improved pumping systems and fertilization all added to the pace of modernization, adding pressures to land that had suffered the ravages of centuries. By 1965, the stamp of contemporary agriculture, and its attendant consequences for land degradation, had already been imposed. Entry to the European Community further added to these pressures through the last 30 years of the 20th century, as the emphasis on competitive productionist agriculture prevailed until the end of the millennium.

Spain and the other Mediterranean nations were already aware of their plight. But water received more attention than land, even though the two are inextricably related. By 1879, the Water Law allowed tax concessions on newly irrigated land and the consequent demand for water led to a boom of reservoir construction, exacerbated by the Spanish law of 1883 that allowed a 50% state subsidy for all major construction work. Other Mediterranean countries were quick to follow, though the early emphasis was on canals rather than on reservoirs. A slower uptake of irrigation resulted from the need for farmers, as well as engineers, to learn the techniques and skills of irrigation agriculture. Not least of all, they had to learn about the dangers of salinisation, a serious cause of land degradation in susceptible Mediterranean drylands and in coastal areas.

Although the Common Agricultural Policy (CAP) of the European Union (EU) strives to maintain the sustainability of rural populations, decrease the disparities between the gross national productivity of member states and incorporate measures that protect the environment, over the years there have been measures that have exacerbated the land degradation problem. The most general and most important of these policies have encouraged agricultural intensification by subsidies that have led to increased production in marginal areas, most notably for the support of mechanization and livestock farming (Papadimitriou and Mairota, 1998). A specific example is the generous economic support given to seed oil (sunflower oil) producers by the EU. It is claimed that it has led to the collapse of olive oil prices relative to sunflower oil and, in turn, to the abandonment of olive groves in southern France and the Aegean Islands. Land abandonment does not invariably lead to land degradation, but can lead to an increase in animal husbandry, more intensive grazing and selective management of the vegetation cover, all of which increase the risk of degradation.

MITIGATING LAND DEGRADATION

The land degradation problem has now been recognized and highlighted in the UN s International Convention on Combating Desertification and Drought (ICCD), which has a special annex referring to the drylands around the Mediterranean. The ICCD has called upon affected countries (including Portugal, Spain, Italy and Greece) to set up National Focal Points and National Committees to establish the nature and causes of the problem and to develop National Action Plans to mitigate the problem.

Mitigation is a broad activity that can range from restoration to reclamation and rehabilitation. The US National Academy of Sciences defined these terms 25 years ago, according to the effects of these actions in relation to the initial conditions in the context of mining. At the simplest level, *rehabilitation* means that the land will be made useful, but with a different land-use and different species. *Reclamation* means that the site will be similar in ecological functioning after disturbance and will be habitable for similar, but not necessarily the same, species. *Restoration* means that the exact conditions before disturbance due to agriculture will be replicated after the remedial action.

Allen (1988) prefers the word reconstruction, a compromise term meaning to build an image of what something was in its original form. Although these terms were originally introduced for land disturbed by mining, they are useful in reminding us that mitigation involves more than simply covering the ground with vegetation to make it look visually more appealing.

Because most Mediterranean states have a land husbandry policy almost invariably dominated by a forestry ethic, it is especially important to be clear about what is understood, and intended by, mitigation. Forestry seems the obvious resort to providing cover and is usually the first option to be considered. A recent investigation of the action taken by Mediterranean governments indicates that they have almost invariably turned to forestry as the solution (Burke and Thornes, 1998). Because the original objectives of the action were never clearly specified, it is impossible to judge their success. The difficulties of evaluating the impacts of afforestation on runoff have been well documented (Rodda, 1976) on the basis of experiments carried out in the USA in the 1920s and 1930s. Moreover, in the face of public criticism, afforestation has been justified *post-hoc* on the basis that it sustains rural activity and is economically viable, both of which are debatable.

If we take the view that the primary objective is to reduce the loss of nutrients and to minimize off-site effects, then the major task is the reduction of soil erosion to a natural rate. Because erosion is a natural geological process that has varied through geological history, it is inconceivable that it can, and should be, reduced to zero. Because the rate of soil production is slow, it is essential to conserve what remains for future generations. At the same time, we have to reckon with achieving the sustainability of rural communities, especially agricultural production. This is a complex activity especially in the mixed farming economies that are widespread in the Mediterranean lands. Soil conservation also implies catchment water retention and hence reduced flooding and reservoir siltation and is certainly more than ecological reconstruction.

The restoration of plant cover is the highest priority, and this is naturally taking place as a result of a very widespread problem land abandonment throughout the Mediterranean, as in the rest of Europe. There have been strong objections, on aesthetic and practical grounds, to the use of exotics such as *Eucalyptus* spp., especially in Portugal. The problem is not only which species to use, but also where and when to plant. Computer simulations have shown that different patterns of planting in space and time have different success rates, but that these also depend on the climate history in the first decade or so after planting. If there is a drought punctuated by heavy storms, then regeneration can be very slow, the cover taking 13 20 years to reach the desired 30%. If regeneration is concentrated in the lowest third of the catchment, reductions in runoff and sediment yield are much more significant than if the upper third of the catchment is abandoned. A ramped (steadily increasing) abandonment is more effective than a steadily declining abandonment and so on. Of course, these are model results, though we are confident that the model fairly represents the plant growth and erosional processes.

These factors are beyond the control of foresters who do not usually have a choice in the matter, certainly not over the weather. Usually the least desirable land is abandoned first, often rocky land at the head of the catchment, difficult to access and farm, and so this is the first to receive attention from foresters on a lowest cost, least argument basis. Political and economic constraints will usually determine the rate of abandonment. In Italy, the reduction in the male population in World War II led to a rapid increase in land abandonment.

Other technical fixes are used as well as forestry. Where land is abandoned, terraces that were built hundreds of years ago are prone to collapse and serious soil erosion may ensue. Maintenance of these structures is part of the required husbandry package.

In water and drought management, hard solutions have generally been preferred, in the form of reservoirs in good (winter) times and check dams to slow down or halt the loss of sediment to downstream areas. The great emphasis on large dams to resolve water resource issues over the last 200 years continues today, even though world-wide, the emphasis has shifted to land husbandry and integrated catchment management as the preferred strategy.

Above all, given that land husbandry has been promoted and practiced since the times of Plato, we feel obliged to ask why land degradation is still a serious problem. Part of the answer at least must rest in the failure of politics and politicians to recognize the essential need to involve those directly engaged in land husbandry in the formulation and implementation of policies based on sound planning. Above all, an integrated, holistic view of the catchment must be taken to ensure that changes made in one place do not create fresh problems in another. European-wide directives and spatially uniform agricultural subsidies are not appropriate instruments for the temporally, spatially and socially extremely variable mosaic of Mediterranean landscapes.

See also: Deserti cation, De nition of, Volume 3; Blue Plan, Volume 4; Deserti cation Convention, Volume 4; Mediterranean Sea, Volume 4.

REFERENCES

Allen, E B (1988) Introduction, in *The Reconstruction of disturbed Arid Lands*, ed E B Allen, AAAS Selected Symposium 109, Westview Press, Colorado, 1 4.

Burke, S and Thornes, J B, eds (1998) *Actions taken by Local Governmental and Non-governmental Organisations to Mitigate Deserti cation in the Mediterranean*, Contract No ENVA-CT95-1083, Concerted Action Report 1, European Commission, Brussels.

Diamond, S and Woodward, I (1998) Vegetation Modelling, in *Atlas of Mediterranean Environments: The Deserti cation Context*, eds P Mairota, N Geeson, and J B Thornes, John Wiley & Sons, Chichester.

Elwell, H A and Stocking, M F (1976) Vegetal Cover to Estimate Soil Erosion Hazard in Rhodesia, *Geoderma*, 15, 61 70.

Francis, C F and Thornes, J B (1990) Runoff Hydrographs from three Mediterranean Vegetation Cover Types, in *Vegetation and Erosion*, ed J B Thornes, John Wiley & Sons, Chichester, 363 385.

Kirkby, M J and Neale, R H (1987) A Soil Erosion Model Incorporating Seasonal Factors, in *International Geomorphology, Part II*, ed V Gardner, John Wiley & Sons, Chichester, 189 210.

Langbein, W B and Schumm, S A (1958) Yield of Sediment in Relation to Mean Annual Precipitation, *Trans. Am. Geophys. Union*, 39, 1076 1084.

Papadimitriou, F and Mairota, P (1998) Agricutural Land Use in the Mediterranean EU Countries, in *Atlas of Mediterranean Environments: The Deserti cation Context*, eds P Mairota, N Geeson, and J B Thornes, John Wiley & Sons, Chichester.

Rodda, J C (1976) Basin Studies, in *Facets of Hydrology*, ed J C Rodda, John Wiley & Sons, London, 257 298.

Thomas, D and Middleton, F (1994) *Deserti cation: Exploding the Myth*, John Wiley & Sons, Chichester.

FURTHER READING

Brandt, C J and Thornes, J B, eds (1996) *Mediterranean Deserti cation and Land Use*, John Wiley & Sons, Chichester.

Hudson, N (1999) *Land Husbandry*, Batsford, London.

Mairota, P, Geeson, N, and Thornes, J B, eds (1998) *Atlas of Mediterranean Environments: The Deserti cation Context*, John Wiley & Sons, Chichester.

Margaris, N S, Koutsidou, E J, and Giourga, C E (1998) Agricultural Transformations, in *Atlas of Mediterranean Environments: The Deserti cation Context*, eds P Mairota, N Geeson, and J B Thornes, John Wiley & Sons, Chichester.

Land Reclamation from Seas

Iris Möller
University of Cambridge, Cambridge, UK

Land reclamation from seas refers to the arti cial drainage of coastal areas and their subsequent enclosure and conversion to freshwater land. The newly created land surface is used as a sea defence and/or for agricultural, industrial, or residential purposes and has to be continuously drained to prevent the reversion to brackish or saltwater wetland. Land reclamation for agricultural purposes dates back to 2000 BC and has occurred in many parts of the world, leading to fundamental changes in the con guration of the world's sedimentary coastlines and contributing signi cantly to global environmental change over the past century.

Methods of coastal land reclamation differ with respect to the intended use of the reclaimed land, the nancial and technological means available, and the socio-political framework within which the reclamation scheme is carried out. Schemes range from the enclosure of existing wetlands by dykes and the construction of small brushwood groynes on intertidal mud ats to the arti cial planting of coastal wetland vegetation, the construction of docks or arti cial islands. Land reclamation from seas can have many undesired, local, regional, and global ecological and socio-economic side effects. Increases in coastal population combined with global sea-level rise and a fundamental uncertainty about the long-term sustainability of such large-scale human intervention in the natural evolution of the coast are likely to provide an unprecedented global challenge.

THE GLOBAL CONTEXT

Low-lying estuarine and coastal areas with their rich fishing grounds, fertile alluvial soils, and potential for transport and trade route development have attracted human settlement since the early days of civilization. Globally, more than 50% of people currently live within 60 km of the sea. As a result, human society has had to develop a variety of ways of dealing with the challenging nature of the sea while harnessing its many resources. On the one hand, the development of sea defense systems has been necessary to protect human lives, buildings and agricultural land from the destructive effects of waves and tides. On the other hand, the need for food and thus agriculturally productive land and space for settlements and industry has driven the development of methods of land reclamation from the seas. This has been particularly important in areas such as the southern North Sea, where the presence of shallow-sloping intertidal flats favors natural accretion of sediment. More than 20% of the land surface of the Netherlands consists of land that has been reclaimed from the sea (Stuyfzand, 1995) (Figure 1). The increasing industrial development of the world s coastlines over the past few decades has introduced a new dimension to land reclamation schemes. In addition, new scientific insights into the complexity of natural coastal processes have increased awareness of the potential environmental and economic impacts of reclamation schemes.

A BRIEF HISTORY OF LAND RECLAMATION FROM SEAS

It is not clear when and where the first attempts to claim new land from the sea began. There is no doubt, however, that early civilizations such as the Sumerians of lower Mesopotamia succeeded in draining the low-lying alluvium deposited by the rivers at the coast as early as 2000 BC for agricultural purposes. These societies also used the salt-tolerant vegetation of the intertidal areas seaward of their enclosure dykes as food and for medicinal purposes (Wagret, 1968). It is likely that the early hydraulic knowledge developed by the Sumerians in Mesopotamia made its way to Italy where, 1500 years later, the Etruscans transformed the swampy marshland of, e.g., the Po valley into productive market gardens. The Etruscans passed much of this knowledge on to the Romans but significant land reclamation in Northern Europe did not start until the seventh and eighth centuries when it became common practice in the region

of Flanders and Zeeland (now the southern part of the Netherlands).

Early land reclamation in Northern Europe was initially linked to the initiatives of individual farmers. By the 13th century, however, much land was reclaimed by monks who were encouraged by the local nobles and thus, land reclamation became closely linked to the national importance of the church and local and national politics. Land reclamation schemes continued to be predominantly driven by the need for agricultural land until the growth of regional and world trade in the Middle Ages increased the need to expand ports and related industrial development at the coast. It could be argued that land reclamation for industrial and residential purposes began as early as the 12th Century in the Netherlands. At this time, the practice of land reclamation was actively encouraged by wealthy merchants from the flourishing Flemish market towns of, e.g., Ghent and Bruges. Most of the reclaimed land on which the first settlements developed in Flanders and Zeeland, however, were initially reclaimed for agricultural purposes.

Land reclamation for the sole purpose of industrial use is a more recent development associated with the expansion of a large number of ports and port cities around the world. In the Netherlands, for example, large areas of intertidal wetland were reclaimed near Rotterdam to expand its internationally important port and related industries. Between 1950 and 1970, the port was extended by reclaiming an area of 11 km^2 from the sea (Pinder and Witherick, 1990). Other large-scale reclamations for the purpose of port development have occurred around the world, e.g., in New York, Marseilles and Le Havre in France, Southampton in the UK (*see* Shipping: Harbors and Ports, Volume 3), Zeebrugge in Belgium, Hong Kong, and Singapore. Nowhere has there been a more urgent need for land reclamations for urban use than in Japan. This is exemplified by the development of Tokyo metropolitan area with a population of over 20 million (Figure 2) (Pinder and Witherick, 1990).

Most recently, the increasing industrial need for reclamation of land from the sea has ceased to be limited to the simple seaward displacement of the shoreline and has involved new schemes that focus on the building of offshore islands in deeper water. One example of such a scheme is Kansai International airport in Osaka Bay, Japan, built entirely on such an artificial island. The increasing economic and industrial development at a global scale has led to large-scale wetland reclamation in previously less developed countries. In China, for example, some 1.19 million ha of tidal lands were reclaimed for agriculture, salt production, shrimp ponds, ship berths and other uses during 1950 1985 (Yu, 1994).

In addition to the potential agricultural and industrial use, land reclamation has been driven by the need for flood protection and to prevent the spread of disease. This was the case in Lagos lagoon, Nigeria, for example, where almost 400 ha of swamp were reclaimed between 1900 and 1946 (Mabogunje, 1968).

METHODS OF LAND RECLAMATION

Methods of land reclamation from the sea have varied through time and depend to a large extent on the natural characteristics of the coastal environments, the purpose of the reclamation, and the technical, financial and organizational means available. Initially, coastal land reclamation did not involve the artificial planting of vegetation or the placement of material in the intertidal zone but focused either on draining existing coastal wetlands (such as salt marshes or mangrove swamps) or on encouraging natural sedimentation and the formation of coastal wetlands before draining.

The earliest reclamation schemes focused on draining existing vegetated coastal wetlands, making them suitable for agriculture and protecting their inhabitants from flooding. Thus the Sumerians created ditch and canal systems to drain the low-lying alluvium deposited by the Mesopotamian rivers (Wagret, 1968) and the early land reclamation in the Netherlands began with small individual projects by local landowners who would construct mounds (terpen) to protect themselves and their cattle from coastal flooding (Knights, 1979). Causeways were constructed between these mounds and formed the first dykes (embankments) and *polders* (enclosed low lying

flat areas with controlled water levels and protection from surrounding higher water levels).

In addition to draining existing wetlands, methods were later developed to encourage the formation of wetlands from non-vegetated mud flats. Such methods rely on the successful accumulation and retention of sediment on the tidal flat. This can only be achieved in conditions of ample sediment supply and low wave and tidal current velocities. Furthermore, vegetation growth is only possible if surface elevations are raised sufficiently to allow enough exposure of plant seedlings between individual inundations and if sufficient drainage is available.

Simple breakwater/ditch constructions started to be used as early as the 13th century in the Netherlands and Northern Germany, to take advantage of the favorable natural conditions (i.e., ample sediment supply and hydrodynamic conditions favoring the settlement/retention of sediment and vegetation establishment). Brushwood breakwaters and small banks were used seaward of dykes to decrease tidal currents and waves in the intertidal zone and thus encourage sediment deposition and the colonization of the mud surface by vegetation. These methods are still being used and have led to the creation of approximately 160 000 ha of reclaimed land in the northern part of the Netherlands and Germany (the Waddenzee and Schleswig-Holstein) (Knights, 1979).

In addition to schemes that encourage the natural sedimentation and vegetation colonization of mud flats for the purpose of land reclamation, the artificial planting of coastal wetland vegetation in the intertidal zone has been attempted in many parts of the world. In 1996, e.g., the government of Bangladesh launched a large-scale program of planting mangrove trees outside coastal embankments. Although the aim was to provide added protection from tidal surges and create an economic resource and wildlife habitat, circa. 60 000 ha of the newly created wetland areas have subsequently been drained and used for cattle, goat, and buffalo grazing (Saenger and Siddiqi, 1993). In Europe, the planting of salt marsh vegetation, such as *Spartina townsendii*, to encourage the progression from mud flat to salt marsh became common practice in the early 20th century (Wagret, 1968).

Throughout history, technological development has had a large effect on methods of reclamation. This is well illustrated in the history of the Netherlands. While early reclamations were carried out on a small scale and by using simple ditches to drain the land, technical advances in the 16th and 17th century led to improved pump drainage driven initially by more efficient windmills, then by steam engines, and most recently by diesel and electric pumps. Such new technologies allowed a much larger scale approach to land reclamation. This type of drainage and enclosure method spread to a large number of areas within Europe. Over 29% of the total current land area (3.4 million ha) of the Netherlands is now pump-drained (compared to circa. 400 000 ha that are drained by gravity) (Knights, 1979). Further advances in technology have led to more advanced ways of land reclamation. The artificial placement of large amounts of sediment near the shore, for example, has been used to raise surface elevations before or after enclosure. This has been successfully used in the construction of artificial islands (e.g., in Osaka Bay, Japan) and industrial port developments (e.g., in Southampton, UK, where 1090 ha of the estuary were claimed in this way (Doody, 1992)).

REGIONAL AND GLOBAL ENVIRONMENTAL IMPACT

The success and long-term sustainability of coastal land reclamation can only be achieved if the methods and designs of such schemes are based on a thorough understanding of the natural (local and regional) conditions of erosion, accretion, wave and tidal action. Ignorance of the natural conditions and processes can not only lead to the failure of the land reclamation scheme itself but can also contribute to coastal environmental problems on a regional and global scale (Viles and Spencer, 1995) (Figure 3).

Undesired ecological and economical consequences have resulted from many land reclamation schemes. In the case of the mangrove plantation scheme of Bangladesh, for example, accretion has been so rapid that much of the land no longer supports mangroves and the objective of creating intertidal habitats and optimal timber production has had to be compromised (Saenger and Siddiqi, 1993). Other studies have shown that the reduction in intertidal area caused by reclamation in estuaries or partially enclosed tidal inlets may significantly alter the tidal prism (the difference between maximum and minimum amount of water stored in the tidal section of the estuary) and the balance between tidal flood and ebb currents. Such changes can lead to the silting up of tidal channels and rivers and this in turn can have implications for the navigability of these waterways and the dredging of channels may become necessary. An example of this can be found in the Westerschelde in the Netherlands, where embanking and poldering has led to a reduction in the tidal prism, increased sedimentation and the need for continuous maintenance dredging of the access channel to the port of Antwerp (Smit *et al.*, 1997). In addition to these physical effects of land reclamation, the conversion of saline intertidal environments to agricultural land can alter the chemical composition of adjacent ground and coastal waters. In natural coastal wetlands, for example, dead organic material gets broken down into iron sulfide or iron pyrites by bacteria that thrive in the waterlogged, anaerobic, conditions. When

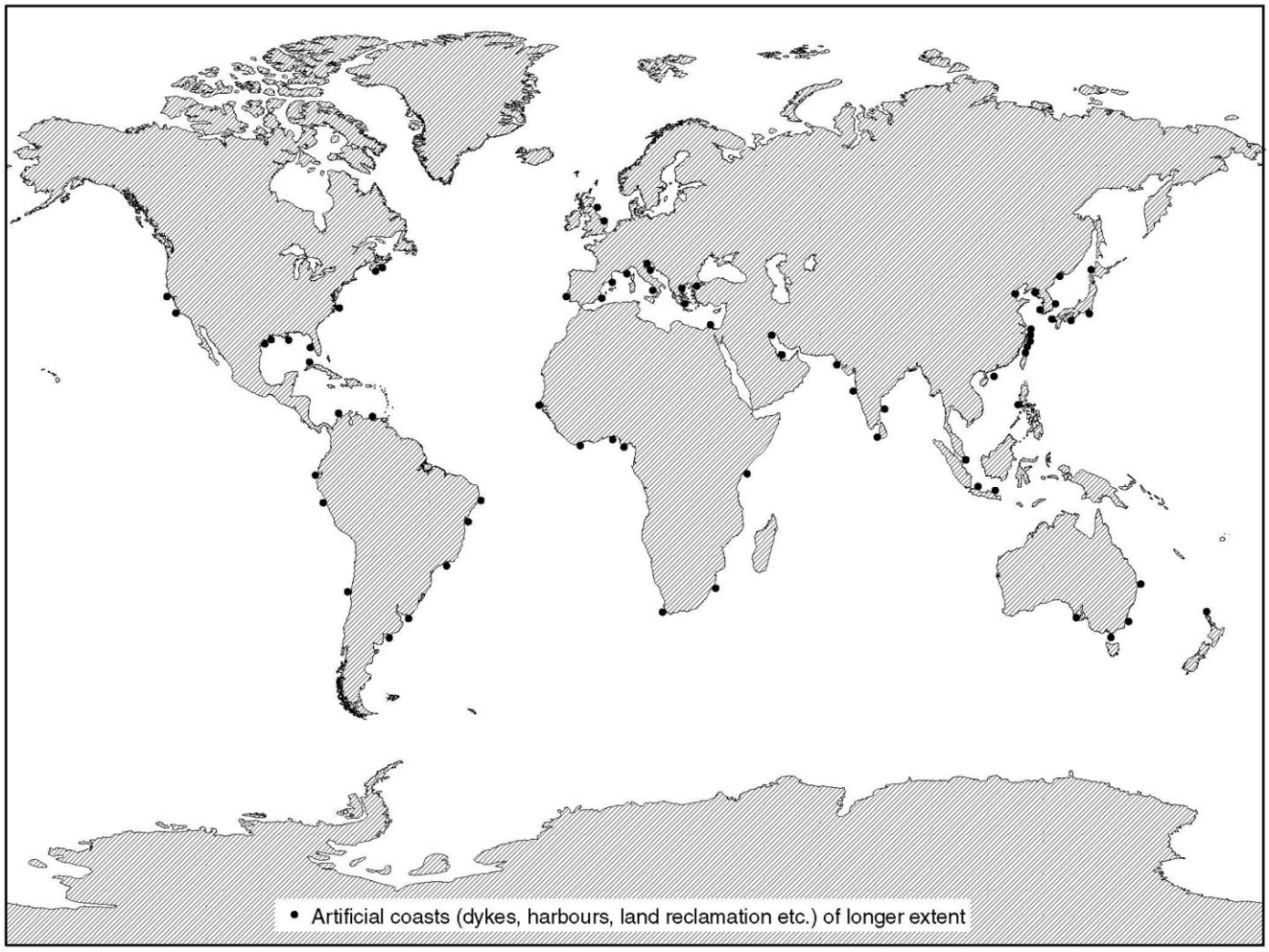


Figure 3 World distribution of major coastal problems related to artificial coasts (land reclamation, dykes, harbors, etc.,) of large extent. (Reproduced by permission of Arnold Publishers in Viles and Spencer, 1995)

these soils are drained as part of reclamation schemes, the iron pyrites come into contact with air and oxidize to form sulfuric acid (*see* **Acid Sulfate Soils**, Volume 3). When this acid reacts with clay in the soil, aluminum hydroxide ($Al(OH)_3$) may form and asphyxiate fish. This effect has been observed in relation to the mangrove areas of Southeast Asia that have been converted to rice fields or other agricultural uses (Anderson, 1991). Other chemical and hydrological problems may result from the intrusion of polluted water from adjacent rivers or saline water from the adjacent coast into the groundwater in continually drained reclaimed areas that lie below sea level (Stuyfzand, 1995).

The loss of intertidal, vegetated or non-vegetated, habitat, as a result of land reclamation itself can have important environmental impacts. Significant losses of intertidal wetlands due to reclamation have been reported from many parts of the world. In the UK, for example, over 56 000 ha of intertidal salt marsh have been reclaimed for agriculture (Doody, 1992) and in China the total area of mangroves declined from 50 000 ha to less than 20 000 ha between 1991 and 1994 (Yu, 1994), largely due to reclamation for agricultural use. In Korea, 1070 km^2 of intertidal land has been reclaimed mainly for agricultural purposes (Hong, 1991). As wetlands have a range of important functions at the local, regional and global level, such large-scale wetland losses are a matter of some concern. At the local level, wetlands act as important wave energy buffers (Knutson, 1988; Möller *et al.*, 1999), sediment stores and traps/filters for pollutants and toxic substances (Williams, 1991). At a more regional level, coastal wetlands provide habitats for a large range of plant, bird and invertebrate species. More than two-thirds of the commercially used fish and shellfish species of the Gulf and Atlantic coasts of the US, for example, depend on coastal or estuarine marshes as spawning or feeding grounds (Williams, 1991). Globally, tidal wetlands and estuaries are thought to constitute 3.18% of the total aquatic primary production

(De Vooys, in Bolin *et al.*, 1979) and there is growing evidence for the importance of coastal wetlands as part of the main biogeochemical cycles. In terms of the global carbon (C) cycle, it is thought that the uptake of carbon by the natural organic soils of temperate wetlands lies between 57 and 83×10^6 tons per year (Armentano and Menges, 1986). The large-scale loss of these environments due to reclamation is likely to have significant effects on global climate through influencing atmospheric carbon concentrations.

LAND RECLAMATION IN THE CONTEXT OF SEA LEVEL CHANGE

In most cases the reclaimed land (whether it is used for agricultural or industrial purposes) needs to be actively prevented from reverting back to land that is liable to flooding by the sea and/or periodic inundation by the tides. If land reclamation has taken place in areas where sea levels are rising relative to the land surface, such protection can require costly engineering structures such as, e.g., the extensive dyke systems of the Netherlands.

While many coastal areas experience a regional isostatic sea level decrease (due to a regional increase in the elevation of the land), others experience sea level rise (due to a regional decrease in the elevation of the land). In addition, global climate change may cause future global eustatic sea level rise (due to a change in the volume of water in the oceans). In 1995, the IPCC (Intergovernmental Panel on Climate Change) estimated that the latter effect could result in an increase in global sea level of 15 50 cm by the year 2100.

In the context of such potentially high rates of sea level rise, the protection of reclaimed land from inundation by the sea may become economically and environmentally unsustainable on some coastlines. In such circumstances, it is not unknown for reclaimed land to revert back to intertidal sand/mud or marshland. In some countries, the option to abandon reclaimed land and to recreate mud flats and/or saltmarshes is strongly supported by conservationists and is becoming an integral part of coastal management schemes. If such abandonment is approved by coastal management authorities (and involves the deliberate breaching of sea defenses or the failure to repair storm damage), however, a multitude of political, economic and social issues (such as, e.g., compensation for the loss of agricultural land), need to be addressed.

Increasingly refined estimates of future sea level change and an increased knowledge of coastal morphological evolution will be required if the sustainability of particular past and present coastal land reclamation schemes is to be assessed in the context of sea level change.

CONCLUSIONS

The reclamation of land from the sea has remained an important means by which to harness coastal resources and satisfy human demands for space in the coastal zone. On a regional scale, however, the socio-economic driving forces behind this have changed through time. While the need for agricultural land dominated early land reclamation schemes, the need for industrial developments and residential space has become more urgent in many industrialized countries over the past few decades. Similarly, the geographic focus of land reclamation from seas is shifting. Throughout history, Northern Europe (and especially the Netherlands) has been the region with the most significant gain in land due to reclamation from the sea. In recent years, however, the increasing economic development of low-lying coastal regions of previously less developed countries is likely to increase the need for agricultural, industrial, and residential coastal land in other parts of the world. The regional and global rise in coastal population and future predictions of global sea-level rise are likely to exacerbate the problems that are associated with the large-scale losses of intertidal habitats that inevitably result from land reclamation from the seas.

REFERENCES

Anderson, I (1991) Land Reclamation Poisons Coastal Waters, *New Sci.*, 30, 11.

Armentano, T V and Menges, E S (1986) Patterns of Change in the Carbon Balance of Organic Soil Wetlands of the Temperate Zone, *J. Ecol.*, 74, 755 774.

Bolin, B, Degens, E T, Kemp, S, and Ketner, P, eds (1979) *The Global Carbon Cycle*, John Wiley & Sons, Chichester.

Doody, J P (1992) The Conservation of British Saltmarshes, in *Saltmarshes. Morphodynamics, Conservation and Engineering Signi cance*, eds J R L Allen and K Pye, Cambridge University Press, Cambridge, 80 114.

Hong, S Y (1991) Assessment of Coastal Zones Issues in the Republic of Korea, *Coast. Manage.*, 19(4), 391 415.

Kay, R and Alder, J (1999) *Coastal Planning and Management*, Routledge, New York.

Knights, B (1979) Reclamation in The Netherlands, in *Estuarine and Coastal Land Reclamation and Water Storage*, eds B Knights and A J Phillips, Saxon House, Farnborough, 209 233.

Knutson, P L (1988) Role of Coastal Marshes in Energy Dissipation and Shore Protection, in *The Ecology and Management of Wetlands. Vol. 1 Ecology of Wetlands*, ed D D Hook, Timber Press, Portland, OR.

Mabogunje, A L (1968) *Urbanization in Nigeria*, University of London Press, London.

Moller, I, Spencer, T, French, J R, Leggett, D J, and Dixon, M (1999) Wave Transformation over Salt Marshes: a Field and

Numerical Modeling Study from North Norfolk, England, *Estuar. Coast. Shelf Sci.*, 49, 411 426.
Pinder, D A and Witherick, M E (1990) Port Industrialization, Urbanization and Wetland Loss, in *Wetlands: a Threatened Landscape*. Institute of British Geographers Special Publication Series. 25, Wetlands, ed E Williams, Blackwell, Oxford, Chapter 8, 234 266.
Saenger, P and Siddiqi, N A (1993) Land From the Sea: the Mangrove Afforestation Program of Bangladesh, *Ocean Coast. Manage.*, 20, 23 39.
Smit, H, Kop, R, and Westmacott, S (1997) Strategies to Combine the Functions of Ecology and Navigation in the Scheldt Estuary, *J. Chartered Instit. Water Environ. Manage.*, 11, 251 256.
Viles, H and Spencer, T (1995) *Coastal Problems. Geomorphology, Ecology and Society at the Coast*, Edward Arnold, London.
Stuyfzand, P J (1995) The Impact of Land Reclamation on Groundwater Quality and Future Drinking Water Supply in The Netherlands, *Water Sci. Technol.*, 31(8), 47 57.
Wagret, P (1968) *Polderlands*, Methuen, London.
Williams, M (1991) Understanding Wetlands, in *Wetlands: a Threatened Landscape*, ed M Williams, Blackwell, Oxford, 1 41.
Yu, H (1994) China s Coastal Ocean uses: Conflicts and Impacts, *Ocean Coast. Manage.*, 25, 161 178.

Land Subsidence

Land subsidence is the lowering of the land surface elevation as a result of changes that take place underground. Common causes of land subsidence from human activity are: pumping water, oil and gas from underground reservoirs; dissolution of limestone aquifers (sinkhole formation); collapse of underground mines; drainage of organic soils; and initial wetting of dry soils (hydrocompaction). Land subsidence causes many problems including: (1) changes in elevation and slope of streams, canals and drains; (2) damage to bridges, roads, railways, storm drains, sanitary sewers, canals and levees; (3) damage to private and public buildings; and (4) failure of well-casings from forces generated by compaction of fine-grained materials in aquifer systems.

Traditional surveying techniques to monitor subsidence, such as global positioning systems campaigns, provide accurate point measurements over limited areas. Synthetic aperture radar interferometry is now used to monitor small-scale surface displacements over large areas, such as major sectors of cities, providing a denser coverage and more complete information of surface height changes.

IAN DOUGLAS *UK*

Land Transformation in China: 3000 Years of Human Ecological History

Rusong Wang
Beijing, China

Land transformation from natural ecosystems to croplands in China started 7000 years ago and has been ourished for the last 3000 years in the middle reaches of the Yellow River. Based on the well-developed human ecological tenet that " Man and nature be as one" , China has enabled nearly a quarter of the world's population to survive for thousands of years on about one- fteenth of the world's farmland.

Dramatic land use and land cover changes have taken place in China in the last two decades since the country was opened to the world and began the transition from a planned to a market economy. Although bringing prosperity and hope to the people, the pace, depth, and magnitude of this transition have exerted severe ecological stresses on both local human living conditions and regional ecosystems.

Faced with the sharp con icts between old traditions and the new transition, an eco-development campaign towards sustainable development has been initiated in some regions of the country aimed at physical and human ecological restoration.

THE OLD TRADITION: HUMAN ECOLOGICAL THOUGHTS ON SUSTAINABILITY IN ANCIENT CHINA

China is a huge country with a population of 1.2 billion, and an area of 9.6 million km^2, two-thirds of which are in highland, hillside or marginal land areas. Its one million km^2 of farmland is mainly located in the eastern and southeastern parts of China. Based on the geographical and demographic conditions, the country can be divided into three regions: the coastal region with 48% population, 13.5% land, 62% gross domestic product (GDP); the interior region with 41% population, 25% land, 32% of GDP; and the remote region with 11% population, 61.5% land, and 6% of GDP. The country s *per capita* farmland amounts to only 0.1 ha. China s climate is characterized by the summer monsoon, which brings high rainfall in July and August, but dry winter conditions. Water and land are the two limiting factors for the life-supporting system. The ecological disasters of flooding, famine and pest outbreak have beset the Chinese for at least 3000 years. In part they are caused by overloading of the carrying capacity of the land in densely populated areas, such as the middle and lower reaches of the Yellow and Yangtze River Basins. As a traditional

agricultural society, ancient China paid more attention to land sustainability. To survive from the marginal water and land, people had to plan, design and manage their environment in accordance with nature by resorting to making efficient use of resources from their native ecosystem. There were no wastes in ancient China, either in towns or rural areas: all wastes were degradable and used as fuel, forage or fertilizer for local ecosystems.

An example is Chinese agriculture. With 7000 years history, Chinese agriculture has maintained sustainable production and soil fertility with the holistic human ecological view has played an important role. The traditional, widely accepted practice of agro-ecological engineering plays a key role in maintaining this densely populated land. Its purpose is to design and maintain a sustainable agricultural ecosystem by enhancing the mechanism of material regeneration and recycling, maintaining ecological integrity and self-reliance. Measures include intercropping, interplanting, total metabolism farming (combination of cropping, husbandry, fishery, forestation and agro-products processing), integrated pest management, soil cultivation through organic and green manure, and biogas systems. In hundreds upon hundreds of villages in China agriculture was sustainable for centuries and the people enjoyed a harmonious country life with nature, with cities and with each other.

Contrasted with the mechanical thoughts dominated in western industrial societies, ancient Chinese human ecological thought emphasized system nexus and dialectical thinking, advocating that man and nature were as one. The ideas of Tian (heaven or nature), Di (earth or resources) and Ren (people or society) and their relationships have been accepted for thousands of years and formed the concept of Chinese human sustainability. The most fruitful period was from Spring and Autumn Time (770 476 BC) to Warring States Time (475 221 BC), when various philosophical schools, including Confucianism, Taoism, Legalism, Yin Yang, Logicianism, etc., flourished. The result is an organized set of principles for managing the relationship between people and the environment. These principles include: Dao-Li (physical relationship with the universe, geography, climate, etc.); Shi-Li (planning and management of human activities, such as agriculture, warfare, politics, medicine.); and Qing-Li (ecological ethics, psychological feelings, attitudes, motives and values towards the environment). Other important principles are: Yin Yang (negative and positive forces playing upon each other provide the basis for all ecological relationships); Wuxing (five fundamental elements and movements within any ecosystem promoted and restrained by each other); Zhong Yong (things should not go to their extremes but take a moderate way) and Feng Shui theory (Wind Water in Chinese expressing the geographical and ecological relationship between human settlements and their natural environment).

The theory of Yin Yang and the Five Elements is the main theory about the structure and origin of the universe, including the relationships of Man and nature. Yin Yang is a theory on the relationship and rules of things and phenomena. Yang originally referred to the sun or the heaven, while Yin refers to the moon or earth. Yang means male, positive, hot, bright, dry, hard, etc.; Yin represents female, negative, cold, dark, damp, soft, etc. The interaction between Yin and Yang produces all things and phenomena and maintains a specific balance. The interdependence and transformation between Yin and Yang result in the dynamics of nature and human society (Wang and Qi, 1991).

Wuxing theory (the Five Elements) divides nature into five agents, i.e., wood (living organism), fire (energy), soil, metal (chemicals) and water. They are promoted and restrained by each other. Wuxing theory was used to explain the network relationships within nature, society and the human body. All natural phenomena and human activities are in this network and connected with each other. Any component either too strong or too weak will put negative impacts on other components as well as on itself through feedbacks. For example, inappropriate behaviour towards nature will be punished by nature. Social change and dynasty succession are also related to the Five Elements. The holistic view implied in the theory of Yin Yang and the Five Elements claims that Man is nourished by nature, grows up as the seasons change, exchanges energy and materials with the environment and thereby is always affected by nature. Meanwhile, Man is also able to adapt and reform the environment.

The principle of Zhong Yong, a common philosophy of Confucianism and Taoism, claims that a contrary result would occur if something developed to an extreme. Therefore, it is recommended that one be Chongho (appropriate, harmonious, neither too hurried nor too slow), and Yongchang (taking things easy and moderate, adapting to nature), Mingcheng (showing good sense in natural laws and social activities). The great tolerance of the Chinese people to difficult times was affected by the thought of Zhong Yong.

Feng Shui theory, despite its superstitious components, is a kind of ecological theory for landscape planning and design. It deals with the ecological integrity between Man and the land as follows (Wang and Hu, 1999):

- totality: geographical continuity, hydrological cycling, ecological integrity, and cultural consistency;
- harmony between structure and function, internal and external environments, implicit and explicit layout, nature and Man, objective being and subjective value, material and spiritual goals;
- mobility: constant wind and flowing water, vertical and horizontal flow, meandering streams, undulating and far reaching;

- vitality: luxuriant, flourishing and productive fauna, flora, soil and aquatic biomes;
- purity: clean and transparently clear water and air, quiet and secluded surroundings, carrying capacity never overloaded;
- safety: spacious, open, easy to disperse and defend, disaster resistant;
- diversity and heterogeneity of landscape, ecosystem, species, society and culture;
- sustainability: negative and positive interlocking feedbacks, self-reliance, self-maintaining, sufficiency and efficiency, appropriate exploitation and development.

The first part of the Chinese word *sheng* means Man on the land . The modern landscape is a kind of human-dominated ecosystem driven by technological and social behaviour, sustained by natural life support systems, vitalised by ecological processes. Shijun Ma termed it a social-economic-natural complex ecosystem (Ma and Wang, 1984) that is driven by energy, money, power and the human spirit. The ten eco-cybernetic principles in ancient Chinese human ecological philosophy fall into three groups: competition for resources and available living space; symbiosis among: (1) Man and nature; (2) different groups of human beings; (3) any human ecological unit and its upper level ecosystem; and (4) self-reliance to sustain the structural and functional stability of an ecosystem through self-organization and recycling (Wang *et al.*, 1990; Wang and Hu, 1994).

THE NEW TRANSITION: RAPID LAND TRANSFORMATION IN RECENT DECADES

China is experiencing rapid growth through regional urbanization and rural industrialization. During the past 20 years, the GNP and the number of cities and towns have increased threefold. Nearly all types of environmental pollution experienced in industrializing countries and of ecological deterioration characteristic of poor developing countries have emerged in China. During the past 50 years, for example, the total farmland was reduced by a quarter.

The Yellow River, the second largest river in China, first began periodic cessations of flow in 1972 due to the excessive human impacts on the upper and middle reaches resulting from urbanization and industrialization (*see* Yellow River, China: a Case Study in Water Resources, Volume 3). Thousands of reservoirs, waterworks, canals and huge engineering schemes in the upper and middle reaches have weakened the river s vitality, mobility and eco-service functioning. As a result, the frequency, duration, and length of cessation have increased year by year, and the date when flow ceases has become earlier and earlier every year since 1972. The average annual duration of no flow was 21 days in the 1970s, 36 days in the 1980s, 122 days in 1995, 133 days in 1996, and 226 days in 1997. The cessation started 85 and 102 days earlier in 1996 and 1997, respectively, than in 1995. Urban and industrial demands for water have been met at the expense of water use to maintain natural ecosystems and agriculture. This in turn has led to severe regional ecosystem degradation.

The long duration, widespread flooding in the middle reaches of the Yangtze River basin in 1998 is another significant example of Feng Shui deterioration. Though the total rainfall in the watershed in 1998 was close to the median level over the last 100 years, and less than that in 25 other years, the river level at most stations exceeded or approached the highest ever recorded. The damage was much heavier than before with a direct economic loss of more than 166.7 billion yuan Ren Min Bi (RMB). The direct causes of this disaster are eco-capital loss, eco-service function degradation, blocking of eco-throughput and the ecological fragmentation of Feng Shui integrity (deforestation, inappropriate hydrological engineering construction, wetland reclamation, biodiversity loss and sedimentation of rivers and lakes). For example, 14.4 million m^3 of forest timber were cut every year in Sichuan and Yunnan provinces in the upper reach of the Yangtze. The total area of soil erosion increased from 0.36 million km^2 in the 1950s to 0.57 million km^2 in the 1990s, a 57% increase in 30 years. The area of Dongting Lake has decreased from 6000 km^2 in the 19th century to 2600 km^2 now. Its capacity decreased from 29.3 billion m^3 in 1949 to 17.8 billion m^3 in 1998. One hundred and twenty thousand km^2 of lakes were transformed into cropland. Of 1066 lakes in the 1950s, only 182 remained in the 1990s. A total of 56.7 billion m^3 of the capacity of 22 main lakes were lost in the Yangtze river basin. As a result, the flood disaster frequency in the basin increased dramatically from once every 83 years during 276 1524, 20 years during 1525 1851, five years during 1852 1949 and four years during 1949 1980s to once a year in the 1990s (Wang *et al.*, 2000).

Though drought and flooding are contrasting disasters, they have both resulted from the deterioration of Feng Shui in both the Yangtze and the Yellow River Basins. Inappropriate resource exploitation, wetland reclamation, deforestation, short-term and localized engineering project and inappropriate settlement construction, have destroyed or disrupted the integrity of the landscape ecosystem. The ecosystem s ability to retain water, conserve the soil and purify wastes has been reduced. The vitality of flora, fauna and soil communities has been weakened. The diversity of the landscape, ecosystem and culture has been greatly reduced. The harmony between people and nature has been destroyed by the misleading attitude that Man has to conquer nature. As a result, the Qi of the basins either were exhausted or stagnated, and the natural service functions declined dramatically. The desire for short-term economic profits has overwhelmed the need to preserve long-term

ecological benefits. The disasters that have happened in the two watersheds are warning signs of more severe disasters in the future.

The Yangtze River Delta, with a population of 73.71 million and an area of 99 530 km^2, has a crucial role in the national economy. Its area, population, GDP and revenue are 1.04, 6.09, 15.42 and 21.8% of those of all China, respectively. For more than 3000 years, it has been the richest economic and cultural area and most beautiful landscape in China. The area used to be the land of rice and fish , the hometown of gifted scholars and beauties , and the paradise on earth due to its marvelous natural and artificial landscape or Feng Shui. The industrialization of China was initiated from this region at the beginning of the 20th century, resulting in the emergence of the megacity of Shanghai, the Suzhou Wuxi Changzhou urban complex and many small towns in the area.

In the last 10 years, dramatic social, economic and environmental changes have taken place in the region especially along the Suzhou Wuxi Changzhou corridor. Flexible development policies, viable decentralized institutions, the long tradition of industrialization, strong technical support from large cities such as Shanghai, and rich human and natural resources all stimulated the extraordinary growth of the local economy. The area has become one of the most rapidly urbanizing and rural industrializing areas in China. Its annual growth rate of GDP has been 50% higher than that of the country as a whole since 1992. The urban population in the region has increased by 60% and population in small towns has increased by 2.7 times; around eight million workers have migrated into the region after 1980.

Compared with that of the mid-1980s, the urbanized area in the Yangtze Delta in 1993 increased by 17.57%, much higher than that of the state average level of 2.5%, and road traffic has increased by 46.67%. In Shanghai, the built-up area was 86 km^2 in 1949, 276.35 km^2 in 1984, 324 km^2 in the early 1990s, and presently 365 km^2; more will be added after the construction of the new Pudong development area is finished. In the same time, agricultural production in the region has been shrinking due to the great shift of arable land into urban and industrial use. Accordingly, the percentage of agriculture in the total production is declining year by year, and food export has been replaced by food import in most areas of the region.

Rural industrialization in the Yangtze delta is proceeding much faster than that of the cities and the average income *per capita* of the region is nearly twice that of the whole country. The economic development in this region is based on township and village enterprises (TVE), which have become a significant source of both income and pollution. According to a statistical survey in 48 small towns around the Taihu Lake watershed, the total annual amount of sewage discharge is $3{\bullet}7 \times 10^8$ t $year^{-1}$, of which industrial wastewater discharge occupies 90%. The quality of Taihu Lake has been reduced by one grade every 10 years since the 1970s. In 1995, 70% of the Taihu water failed to reach the national grade III standard. In 1992, more than 50% of the water in the Dianshan, Yangcheng and Gehu lakes failed to meet the national grade III standard. Only 1.8% of treated wastewater meets the state environmental standard. Due to over-exploitation of underground water and deforestation, the land has visibly been degraded. The over-exploitation of underground water has caused a severe fall in the water table. Subsidence caused by over-exploitation of underground water in the region affects about 8000 km^2. As a result, human health has been significantly affected. The outbreak of hepatitis in Shanghai in 1988, where 230 000 citizens were infected due to food pollution is just one example of the consequences of degraded environmental conditions.

A main driving force for this transition is the replacement of Chinese human ecological traditions by high consumption and high impact activities. Ecological farming is being replaced by petroleum agriculture, with organic manure being displaced by chemical fertilizers, and pesticides replacing the traditional integrated pest control methods. Household excrement and urine are being carried away through pipes mixed with domestic chemicals. A deluge of plastic bags, packaging materials and advertisement papers is taking the place of traditional organic garbage. Frozen or preserved food from supermarkets is consumed instead of fresh meat and vegetables from the free markets. Some enterprises are following the development strategy of other industrializing countries: pollution first, treatment later; speed first, efficiency later; and chaos first, sustainability later . In one word, reductionism is replacing holism, diversified human ecosystems are being replaced by mono-production and mono-culture, the self-reliant life style is being displaced by a modern life style involving high energy consumption and high environmental impacts (Wang *et al.*, 1996).

ECO-DEVELOPMENT: A NEW TRANSITION BASED ON AN OLD TRADITION

Fortunately, based on the long tradition of human ecological thought in China, an eco-development movement led by ecologists and supported by local government has been operating in several villages, counties and towns in China since the 1990s. The essential idea of eco-development is to plan, design, manage and construct the human ecosystem s function of production, living, and sustenance according to ecological cybernetics. The ultimate goals of eco-development are wealth, health and faith. *Wealth* measures the structural state of the monetary assets, natural assets (minerals, water, forest, soil, air and biodiversity), human resources and social resources (institution, arts, etc.). *Health* measures the functional state: individual and social health,

economic health and ecosystem health. *Faith* measures the behavioral mode: attitudes, traditions, customs, perceptions, ideas, concepts, values, beliefs, ethics and life styles. Systematically responsible planning, totally functioning engineering and ecologically sound management are the three determinants of sustainable development (Wang and Ouyang, 1996).

Several pilot projects have been initiated, including more than 50 provincial and 29 national Comprehensive Experimental Communities for Sustainable Development (CECSD), 51 national Eco-Agricultural Counties (EAC) and 100 Ecological Demonstration Districts (EDD). Some fruitful theoretical and applied results have been obtained. Local engineers, planners, and managers have adopted sustainability measures and recommendations, which will be the model for the development of other regions. The project focuses on the comprehensive use of natural resources; comprehensive management of the environment; and comprehensive development of people. It operates by introducing effective appropriate technology, developing systematically responsible institutions, and cultivating an ecologically sensitive culture (Wang *et al.*, 1998).

As one of the new sustainable development strategies, ecological agriculture has been implemented in many of these pilot studies since the early 1980s. There are more than 2000 eco-agricultural experimental sites in China ranging from village to county level which cover 123 000 km^2 and 23.92 million people. By 1995, the annual increase of crop yield was over 15%, the yield per unit of arable land increased by 10% since 1980, respectively, 4.5 and 9.2 times higher than those of counties without eco-agricultural development. *Per capita* income at the experimental sites was 12% higher than the average of neighboring regions. For example, the *per capita* income in 105 eco-agricultural experimental sites in Hubei province in 1995 was 32% higher than the average level in the province.

Baiquan county in Heilongjiang province which has suffered severe soil erosion due to intensive land reclamation, is one of the EACs. The reclamation ratio of soil in the county before 1990 reached 77.1%. Sixty percent of the total territory and 71% of the arable land were subject to erosion before 1990. Water eroded 6 mm of topsoil per year and wind eroded 4 mm $year^{-1}$. The organic matter in the soil was reduced from 8% at the beginning of reclamation to nowadays 2 3%, the lowest one being even below 1%.

Since 1990, an eco-development aimed at harnessing soil erosion has been initiated throughout the county. Approximately 150 eco-restoration zones and 32 comprehensive resource-based zones were established. More than 7000 gullies have been regulated, 36 000 ha terraced fields have been restored, 40 000 ha of sloping land has been cultivated along contour lines and 13 000 ha of ridges have been planted in 10 years. Four thousand, seven hundred and thirty-three ha of former cultivated steep slope cropland has been restored to grassland and grass forestry ecosystems. This has stimulated local livestock husbandry. Approximately 15 000 tons of wood were saved through widespread use of firewood-efficient stoves and solar energy houses. One thousand, four hundred and thirty energy-efficient tobacco curing houses were built, which saved 1430 tons of coal annually. Additionally, there are 84 solar energy houses, totaling about 7621 m^2, and 5150 solar energy hog cotes, which increased pig weights by 250 g. Finally, the area of fuel wood has reached 26 000 ha, and the recycling of crop residues and straw to fields resulted in an increase of organic content in the soil by 26.9% in 10 years (Wang and Yan, 1998).

Through 10 years implementation of eco-agriculture, agricultural gross output, gross income and *per capita* income have increased by 190%, 360% and 540%, respectively. Forest cover increased by 19.7%, 5% higher than in neighboring areas. Soil erosion control areas covered 67.7% of the total eroded area, resulting in a 37% reduction of surface runoff, a 50% decrease in soil loss, and a decrease in wind speed by 10.14%.

Before 1980, Xishan village in Shandong province was a poor village where food had to be imported. The village was subsidized by the government. The *per capita* income was less than 90 yuan RMB and the total fixed assets of the village were less than 20 000 yuan RMB in 1978. The eco-village development initiated in 1985 started from courtyard planning, with a 160 m^2 lotus pond in front of each house, and a 64 m^2 grape garden in front of the lotus pond. Around the village, a 2 km long and 30 m wide river was dug, outside of which was planted an orchard of 66.7 ha. There is also a 1 ha park at the back of the village, which makes the environment of this village more beautiful and viable. Pigs and chicken are raised in a piggery where pigs are kept in the basement, while chickens are on the upper floor. In the basement there is an 8 m^2 methane pond. The excrement of people and livestock are placed in this pond to produce biogas. Fermented waste is used as fertilizer in the yard. In this way, the *per capita* household income exceeded 2000 yuan RMB, some households have even reached 6000 yuan RMB, which is 20 70 times higher than the income in 1978 (Lu *et al.*, 1995). In 1990 a series of methane fermentation ponds, covering 2200 m^2, were built to treat the excrement of 5000 pigs, 6000 ducks, and 300 cows. The ponds provide about 400 m^3 biogas per day for residents use and about 7500 kg of high quality organic fertilizer per day for agriculture.

The eco-village development has changed this village dramatically in 10 years. Grain production increased from 12 900 kg ha^{-1} in 1987 to 19 770 kg ha^{-1} in 1995, with a *per capita* output of 1305 kg. The *per capita* income increased from 870 yuan RMB in 1987 to 2511 yuan RMB in 1995. The land productivity has multiplied from 14 505 yuan ha^{-1} in 1987 to 54 520.5 yuan ha^{-1} in 1995. The organic matter

in soil increased from 1.1% in 1987 to 1.3% in 1995. Meanwhile, the chemical fertilizer input was reduced from 1290 kg ha^{-1} in 1987 to 765 kg ha^{-1} in 1996.

Another example of the CECSD is the eco-county development in Dafeng, a middle-size county located in the northern part of Jiangsu Province with 740 000 people and 2367 km^2 of land. It is also one of the EACs and the EDDs. The county s economy was less developed in comparison with counties in the southern part of the province, where rapid industrialization started in 1978. The severe environmental problems in South Jiangsu province warned the people of Dafeng that they must search for an alternative way of realizing industrialization while maintaining a harmonious relationship with nature (Wang *et al.*, 1990).

The main goal of the Dafeng eco-county development is the promotion of a synthesis of the biophysical, socio-economic and psycho-cultural factors of production, consumption and sustainability in a long term, large scale and holistic policy-making framework. In striving towards this target, technological, institutional and cultural instruments are used together with cultivation of the mechanisms of competition, symbiosis and self-organization. A compromise between decision-makers, entrepreneurs, experts and the public is as important as the exploitation of the external and internal potentials. The evaluation indicators are: production efficiency; quality of life; institutional harmony; people s capability; ecological order. Twenty-two demonstration sites for ecological engineering have been organized in the county including ecological agriculture, ecological industry, ecosystem conservation, and community development. One hundred technological societies and 10 technological service networks have been set up. Considerable progress has been achieved in social, economic and environmental development since Dafeng eco-county planning was adopted, by the Yancheng municipal government in 1989. The county s gross domestic production in 1996 was nearly eight times greater than that of 1986 before the plan was initiated. Environmental quality has improved or has been maintained at the same level as that of 1980 when there were few industries, and the soil organic content has increased from 1.19% in 1986 to 1.23% in 1996. At the same time, the forest cover has increased by 15.7%. The county has won more than 10 national awards for its eco-development by relevant ministries and international agencies since 1989.

REFERENCES

Lu, B Y, Li, X C, and Ying, D (1995) Study on Rural Eco-yard, *Agric. Environ. Protection*, 5, 11 14.

Ma, S and Wang, R (1984) Social-Economic-Natural Complex Ecosystem, *Acta Ecol. Sin.*, 4(1), 1 9.

Wang, R, Fang, J Y, Fong, Z W, and Gao, L (1996) *Wealth, Health and Faith - Sustainability Studies in China*, China s Science and Technology Press, Beijing.

Wang, R and Hu, D (1994) From Economic Prosperity to Ecological Sustainability a Theoretical and Practical Concern of Sustainable Development in China, *J. Environ. Sci. (China)*, 6(4), 389 401.

Wang, R and Hu, D (1999) Totality, Mobility and Vitality: Feng Shui Principles and Their Application to the Blue Network Development of Yangtze Delta, in *Land and Water: Integrated Planning for a Sustainable Future*, ed H C M Heyning, ISoCARP, The Netherlands.

Wang, R and Ouyang, Z (1996) Ecological Integration Probing the Scientific Methodology of Human Sustainable Development, *Chin. Sci. Bull.*, 41(special issue), 47 67.

Wang, R and Qi, Y (1991) Human Ecology in China: Its Past, Present and Prospect, in *Human Ecology - Coming of Age: an International Overview*, ed S Suzuki, Free University Brussels Press, Brussels, 183 200.

Wang, R, Ouyang, Z Y, and Zhao, Q T (1990) On Ecological Construction, *J. Environ. Sci. (China)*, 12(3), 1 12.

Wang, R, Ren, H, and Ouyang, Z (2000) *China Water Vision: The Eco-sphere of Water, Life, Environment and Development*, China Meteorological Press, Beijing.

Wang, R and Yan, J (1998) Integrating Hardware, Software and Mindware for Sustainable Ecosystem Development: Principles and Methods of Ecological Engineering in China, *Ecol. Eng.*, 11, 277 290.

Wang, R, Yan, J, and Mitch, W J (1998) A Promising Approach Towards Sustainable Development in Developing Countries, *Ecol. Eng.*, 11, 1 16.

Wang, R, Zhao, Z L, and Dai, X L (1990) *Human Ecology in China*, China Science and Technology Press, Beijing.

Life-cycle Assessment

see Environmental (Eco) Labeling (**Volume 4**)

Liming of Acidifying Forests

see Nitrogen Deposition on Forests (**Volume 2**)

Low Concentration Methane Bearing Gases

Edan Prabhu
FlexEnergy Inc., Mission Viejo, CA, USA

Methane, a major constituent of natural gas, is generated from many sources related to human activities. Enormous

quantities of such methane are emitted each year, most commonly as part of the carbon cycle, through the anaerobic decomposition of vegetable and animal matter. Over time, methane in the atmosphere converts into carbon dioxide, and the carbon dioxide is then photosynthesized back into vegetable matter. Examination of its prevalence and sources suggests potential uses for the low concentration methane now emitted into the atmosphere. If low concentration methane can be captured and converted into carbon dioxide and water, useful energy could be extracted, providing an important step towards sustainable energy supply.

METHANE IN THE ATMOSPHERE: A MAJOR GREENHOUSE GAS

Methane (CH_4) is the principal constituent of natural gas, but is also generated in large quantities as a by-product of various natural and human created systems. Methane is created by anaerobic decomposition of biological matter, such as in rice fields, landfills, farms, wetlands, termites and other sources. Methane is also released during coal, oil and natural gas production, the result of decomposition of biological matter in geological time, possibly under far more aggressive conditions.

In some cases, methane is available in relatively high concentrations, between 40 100% of gas flows, such as in natural gas production systems, large landfills and large water treatment facilities. Such sources of methane are being successfully used on a commercial basis today. In most circumstances, however, the methane is available only in very low concentrations (10% or less), levels at which energy recovery is thought to be economically unfeasible (*see* Methane: Industrial Sources, Volume 3).

The US Environmental Protection Agency (EPA) states: As a contributor to climate change, methane is second only to carbon dioxide . The EPA also states Over the last two centuries, methane concentrations in the atmosphere have more than doubled from about 700 parts per billion by volume (ppbv) in pre-industrial times to 1730 ppbv in 1997 largely due to human-related activities (Figure 1). The doubling of methane concentrations corresponds to a four-fold increase in world population. The EPA rates methane s global warming potential (GWP) as 21 on a scale where carbon dioxide has a GWP of 1. Thus, methane is 21 times more potent as a greenhouse gas than carbon dioxide. The EPA also states that a reduction of about 10% in methane emissions from sources due to human activities alone would halt the annual rise in methane concentrations, thereby significantly contributing to mitigation of climate change (*see* Methane, Volume 1).

Total global annual methane emissions estimated by the EPA for 1990 were about 350 teragrams, or 350 million tons, of which about 70% came from anthropogenic sources and 30% from natural sources.

Anthropogenic activities such as fossil fuel production, animal husbandry (digestive processes of ruminant livestock and manure), rice cultivation, and landfills release significant quantities of methane (Figure 2), fundamentally altering the composition of the earth s atmosphere. In industrial nations such as the US, methane is primarily generated by decomposition of wastes in landfills, manure and enteric (intestinal) fermentation associated with domestic livestock, natural gas systems, and coal mining. In developing nations, the methane is produced primarily by rice cultivation and enteric and manure decomposition in ruminant livestock (*see* Cattle Grazing: Impacts on Land Cover and Methane Emissions, Volume 3).

Natural sources of methane include wetlands, gas hydrates and permafrost, termites, oceans, freshwater bodies, non-wetland soils, and other sources such as wildfires.

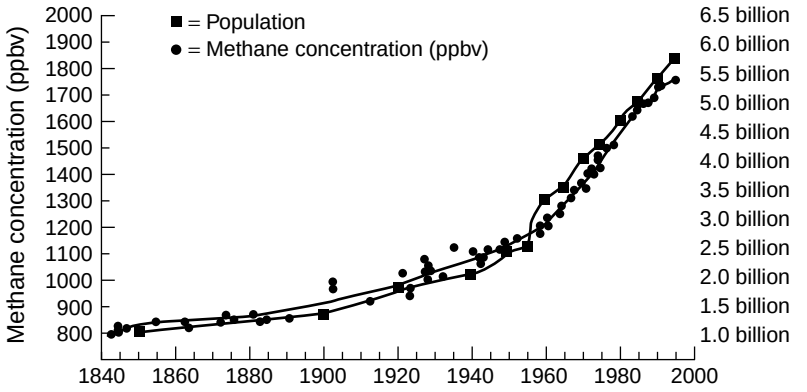
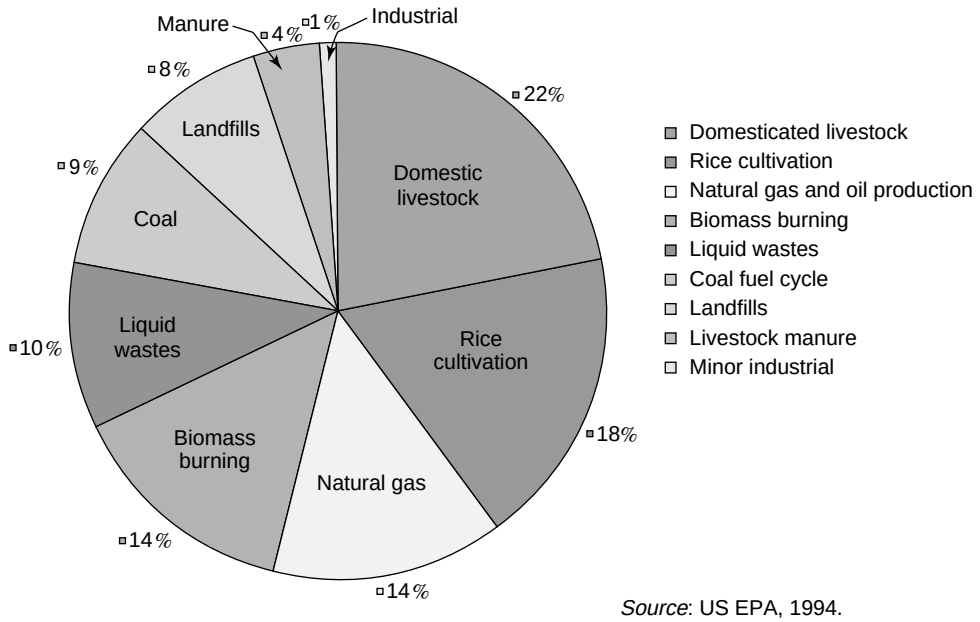


Figure 1 Increases in methane concentrations compared to world population growth (1841–1996)



Source: US EPA, 1994.

Figure 2 Sources of total global methane emissions from anthropogenic sources

In an ironic twist, in some localities, global warming may actually increase the natural production of methane by creating larger wetlands and swamps, exacerbating the emissions problem. The emissions from natural sources are roughly divided 50% from wetlands and 50% from other sources. The two largest anthropogenic contributors are domestic livestock with 22% and rice cultivation with 18%, together a staggering 40% of anthropogenic emissions.

The US EPA and other government agencies around the world have rigorous programs to reduce methane emissions. The EPA focuses on the main anthropogenic sources in the US, on high concentration methane emitters such as landfills and animal manure, with the goal of capturing and recovering the gas. There has also been some effort to prevent formation of low energy methane emission, for example, by altering cattle feed to reduce their methane production.

TECHNOLOGIES FOR USING METHANE TO PRODUCE USEFUL ENERGY

How much of this methane could be recovered as fuel, and how significant a fuel would it be? The 350 million tons of methane emitted each year are the equivalent of roughly three billion barrels of oil a year, or three percent of total world energy consumption. Obviously there would be a large payoff if this energy could be recovered. Even if only a tenth of it could be used for power generation, it would greatly dwarf all solar and wind energy combined. How possible and feasible would this be?

The existing programs by the EPA and others focus on those streams of methane that are easily recoverable; these are generally those where the methane is produced and recovered in relatively high concentrations. For example, the best landfills generate gas with about 40 50% methane, and anaerobic digesters for cattle manure generate gas with about 50 80% methane. The fuel is sometimes captured for use in reciprocating engines, gas turbines and in some cases used to produce heat and steam for electricity. In a few instances, the methane is piped to other locations for industrial or commercial uses. However, most of the methane emissions are simply vented into the atmosphere. If a significant portion of this methane is to be captured, it will most likely be recoverable only in much lower concentrations. In order for a fuel to be converted into useful energy, its chemical energy must be converted into heat and then into useful mechanical or electrical energy. Traditional combustion engines developed for using petroleum based fuels, require a methane concentration of at least 10%, but prefer no less than 100% methane.

While traditional engines cannot utilize the low concentration methane, there are fresh approaches that can make a difference. Technology developed during air pollution research has made catalytic converters and combustors available that remove very low concentrations of hydrocarbons from tailpipe and other gases, down to the parts per million (ppm) threshold. Theoretically, a catalytic combustor could actually reduce the 1730 ppbv of methane in ordinary air, but doing so is impractical and the process would not be self-sustainable. However, it may well be feasible to usefully recover the energy from methane concentrations as low as 1 2%. Modern microturbines routinely

operate in these fuel ranges. They generally begin with a highly concentrated fuel that is diluted after combustion. By converting the turbines to catalytic combustion, it will be feasible to run them directly on low concentration methane fuel gases. Such catalytic combustion is already in use for certain large turbines, and one company is now developing the Flex-Microturbine™ to use catalytic combustion for very low methane fuels. Should that effort be successful, the path to recovery of low energy gases could be radically altered.

COLLECTION AND LOCAL CONSUMPTION OF LOW CONCENTRATION METHANE GAS

How feasible is it to collect gases from methane emitting sources such that the resulting concentrations are at least 1 2% of methane? One percent methane is 10 000 ppm, which is over 2000 times the concentration of methane in ambient air. However, methane rises from landfills, rice fields, animal pastures, animals themselves, coal mines, oil and gas drilling, production and transport operations often in concentrations of 10 000 ppm or greater. Can simple collection systems be devised? Perhaps landfills could be covered with large sheets to collect the emanating gases; methane-bearing gases from coal mines could be diverted rather than diluted; oil and gas operations could be adapted to collect these waste fuels; large feedlots could be enclosed and the methane collected.

Size becomes a critical issue. It would be prohibitively expensive to cover very large areas, or to transport low energy gases through long distances to central power plants. It would also be prohibitively expensive to develop means to boost the methane concentrations by treating the gas stream. However, microturbines are small and easy to move. It may be feasible to install small portable power plants, so-called distributed generation, that locally converts the methane into electricity and useful heat. Distributed generation is one of the fastest growing new means of power generation and could very well become a major part of the generation mix in the next ten years. Microturbines, in particular, are rapidly becoming commercial with a marked increase in production in the last few years.

THE BENEFITS OF LOW CONCENTRATION METHANE RECOVERY

In addition to the immense value derived by consuming methane for greenhouse abatement, several other benefits will accrue.

The most immediate benefit would be the generation of electricity. Electricity is inexpensive to transport and easy to use. It is by far the most desired fuel on the planet, the true engine for growth. In addition to the electricity generated, in some instances the waste heat from individual small power plants may be useful for space or water heating or pasteurizing milk. In a world of rapidly fluctuating fossil fuel prices, the fuel, for the present, at least, is free for the taking.

The catalytic combustion needed for recovery of low concentration methane assures virtually no NO_x. NO_x, formed in most other engines, is the main ingredient of urban smog and a contributor to acid rain.

Many rural areas lack electricity. It may become feasible to supply those areas directly using local fuels. For example, many developing nations that grow rice, have large herds of cattle, or contain swamplands may be able to capture and convert the methane. Methane is considered non-toxic; a rice field could simply be covered with a transparent sheet to collect the gas. The methane-laden air inside may actually be breathable, at least for short periods.

The environmental impact of such power plants would be almost entirely beneficial. No new major structures would be needed, no wastewater, no solid disposal problems. The new energy would displace fossil fuels that cause damage during their production, use and disposal. Small power plants could be sited with no need for complex licensing review. The footprint of these plants on their local environment would be less than that of fossil power plants, solar or wind energy systems. They would tap nature s energy in the least obtrusive way, making them the most environmentally friendly source of renewable energy. Unlike wind and solar power, electricity could be generated any time of day or night to deliver dependable power to small, remote villages without need for a large central grid. The recovery of this waste fuel would have the potential to provide at least one percent of the world s energy needs, perhaps much more.

THE FUTURE

In the eternal carbon cycle, all carbon goes from carbohydrate in plants to carbon dioxide in the atmosphere and back to carbohydrate in plants. Along the way, some of it is first converted into methane. This methane may be a valuable source of energy for sustainable living. It is recommended that public policy programs place heavy emphasis on its recovery and use. There may be value in developing means to increase such methane production by mimicking and enhancing natural digestion processes. Even more methane could then be generated and used, unobtrusively tapping into the eternal cycle for comfort and growth.

See also: Volatile Organic Compounds, Biogenic (VOCs), Volume 2; Methane: Industrial Sources, Volume 3.

FURTHER READING

EPA website http://www.epa.gov/ghginfo/questions.htm, Methane and Other Greenhouse Gases.

EPA Report, *US Methane Emissions, 1990- 2020, Inventories, Projections and Opportunities for Reductions*, available on website http://www.epa.gov/ghginfo/01-ES-Intro.pdf.

EPA website http://www.epa.gov/ghginfo/topic1.htm#measure1, Methane in the Atmosphere, Measurement of Historical Concentrations, Measurement of Current Concentrations.

EPA website http://www.epa.gov/ghginfo/table1-1.htm, 100 Year Global Warming Potentials for Selected Greenhouse Gases.

EPA website http://www.epa.gov/ghginfo/topic4.htm, Current Global Methane Emissions.

EPA website http://www.epa.gov/ghginfo/topic4.htm#global, Current Global Methane Emissions.

M

Malnutrition, Infectious Diseases and Global Environmental Change

David Pimentel
Cornell University, Ithaca, NY, USA

Global climate change is likely to have major effects on public health. Higher than present temperatures may encourage the spread of mosquito vectors and other public health pests. In addition, heavy rainfall and ooding may encourage increases in mosquito populations, leading to outbreaks of diseases such as malaria.

Droughts may reduce crop production, intensifying the malnutrition problem that currently exists (WHO, 1996a,b; Conway, 1997). Malnutrition itself is a serious disease but equally important is the fact that malnourished people are also more susceptible to infectious diseases, such as malaria and cholera, and the effects of environmental stresses, such as smoke and chemical pollution (Pimentel et al.*, 1998).*

Climatologists project that greenhouse gas emissions will probably increase average global temperatures by approximately 3 °C in temperate regions within the next 50 years or so, increase rainfall in some regions and cause droughts in others, and intensify hurricanes and other tropical storms (Rosenzweig and Hillel, 1998). In addition, sea levels may rise about half a meter.

In this article, the potential effects of global climate change on malnutrition and other diseases are assessed.

GLOBAL CHANGE

Assessing the impact of carbon dioxide and other greenhouses gases on the world s climate is difficult because at this juncture, we have not fully defined the size of atmospheric, biotic, and oceanic reservoirs as sources and sinks of carbon dioxide and other greenhouse gases (Hominick, 1993; Rosensweig and Hillel, 1998). Nonetheless, it is well documented that burning fossil fuels and removing and burning forests does increase the levels of atmospheric carbon dioxide and other greenhouse gases. Levels of atmospheric carbon dioxide have risen from 280 ppm in the 1850s to more than 350 ppm today, and are expected to reach 400 600 ppm by the year 2035 (Hominick, 1993; Rosensweig and Hillel, 1998). Without significant societal changes, concentrations of atmospheric carbon dioxide and other gases will continue to increase.

Atmospheric and other scientists conclude that if present trends continue, the combined effect of atmospheric carbon dioxide and other greenhouse gases will increase the earth s average temperature by 1.5 4.5 °C before the middle of the century, with warming most pronounced in the Arctic. Rosensweig and Hillel (1998) have reported that average global temperatures have already risen approximately 0.5 °C in the past century. Although this temperature increase seems indicative of global climate change, others question whether a significant global temperature increase has occurred, and instead attribute any warming to the natural climate variability on earth.

Exactly when and how much the climate will change depends on the complex interactions of the atmosphere, oceans, and biosphere. Climate changes will not be uniform over the different regions of the earth. Winter temperatures in the middle and high latitudes can be expected to rise to more than twice the world average rise (Rosensweig and Hillel, 1998). Summer temperatures will also rise, but less than during other seasons. Furthermore, rainfall is expected to increase significantly in some regions, such as the eastern portion of Africa, and decrease in others, such as the central portion of the United States (Hominick, 1993; Rosensweig and Hillel, 1998).

MALNUTRITION WORLDWIDE

Undernutrition

Malnutrition, which includes inadequate intake of calories, protein, and numerous essential vitamins and minerals, is a major disease and is in part related to environmental degradation and rapid population growth. Malnutrition prevails in regions where the overall food supply is inadequate, where populations lack economic resources to purchase

food, and/or where political unrest and instability interrupt food supplies. In addition, rapidly expanding human populations intensify food supply problems by diminishing the *per capita* availability of cropland, water, and biological resources (Pimentel and Pimentel, 1996).

In 1950, 500 million people (about 20% of the world population) were considered malnourished (Grigg, 1993). Today more than three billion people (about 50% of the world population) suffer from malnutrition (WHO, 1996a). This is the largest number and highest rate of malnourished people in history. Each year, between 6 and 14 million people die both directly from malnutrition and from malnutrition related diseases (Murray and Lopez, 1996). Malnutrition problems are also on the increase in the United States, especially among the poor.

In many parts of the world, especially in developing countries, severe shortages of vitamin A cause prevalent blindness, illness, and death. For example, in the Sahelian region, as well as in West and East Africa, *per capita* consumption of vitamin A has been declining during the last 10 20 years, while associated serious eye problems have been increasing (ACC/SCN, 1992). Worldwide, approximately 250 million children are vitamin A deficient (WHO, 1996b). Each year, vitamin A deficiency causes approximately two million deaths and three million serious eye problems, including blindness (Murray and Lopez, 1996).

Similarly, *per capita* iron intake has been declining during the last 10 30 years, especially in sub-Saharan Africa, South Asia, China, and South America, primarily because of the overall shortages of food that are resulting in widespread inadequate nutrition (ACC/SCN, 1992). Globally, more than two billion people are iron deficient, severely enough in most cases to suffer from anemia (WHO, 1996b). An estimated 20% of global malnutrition deaths are attributed to anemia (Murray and Lopez, 1996). In general, people with anemia have difficulty working, as well as limited resistance to other diseases. In addition, about 1.6 billion people live in iodine deficient environments and suffer from iodine deficiency disease (WHO, 1996b).

Malnutrition is frequently complicated by parasitic infections, and is typically found in poverty-stricken areas with inadequate sanitation. Malnourished individuals, especially children, are seriously affected by parasitic infections, because these infections can further reduce the amount of nutrients a person can obtain from his or her food. Intestinal parasites frequently diminish both appetite and food intake. Their presence increases loss of nutrients by causing diarrhea and dysentery. Hookworms, for instance, can suck as much as 30 ml of blood from an infected person each day, gradually weakening individuals and lowering their resistance to other diseases (Hotez and Pritchard, 1995). Because an estimated 5 20% of an infected individual s daily food intake is used to offset the effects of these parasitic illnesses, the overall nutritional status of a parasite-infected person is greatly diminished over time (Pimentel and Pimentel, 1996).

Decreasing Food Production

Based on current rates of increase, the world population is projected to double to more than 12 billion in less than 50 years (PRB, 1998). As the world population continues to expand at a rate of 1.4% per year, with more than a quarter million people added each day, providing adequate food becomes an increasingly difficult problem. Conceivably, the numbers of malnourished will reach 4 5 billion in future decades.

Reports from the Food and Agriculture Organization (FAO) of the United Nations and the US Department of Agriculture, as well as numerous other international organizations, further confirm the serious nature of the global food supply problem (RS and NAS, 1992; NAS, 1994; WHO, 1996a). For example, the *per capita* availability of world cereal grains, which make up 80 90% of the world s food supply, has been declining for the past 17 years (Kendall and Pimentel, 1994).

Thus, as the world population continues to expand, more pressure than ever before is being placed on the finite basic resources that are essential for food production. Unfortunately, while the human population grows exponentially, food production can only increase linearly. Furthermore, degradation of land, water, energy, and biological resources, all vital to agriculture, continues unabated (Pimentel *et al.*, 1997a; Pimentel and Pimentel, 1996).

More than 99% of the world s food supply comes from the land; less than 1% is from oceans and other aquatic habitats (FAO, 1991). The continued production of an adequate food supply is directly dependent on ample quantities of fertile land, fresh water, energy, and natural biodiversity. Obviously, as the human population grows, the requirements for these resources will escalate. Even if these resources are never completely depleted, their supply will decline significantly on a *per capita* basis because they must be divided among more people.

Throughout the world, fertile cropland is being lost from production at an alarming rate. While true for all cropland, it is best illustrated by the diminishing amount of land now devoted to cereal grains. Soil erosion by wind and water, in addition to overuse, has been responsible for the loss of about 30% of the world s cropland during the past 40 years (WRI, 1994; Pimentel *et al.*, 1995). Once fertile soil is lost, it takes 500 years or more to form a mere 25 mm (1 inch) of fertile soil. For crop production, at least 150 mm of soil is required (*see* Soil Deterioration and Loss of Topsoil, Volume 3).

Most of the replacement for eroded and unproductive agricultural land is coming from cleared forests and

marginal land. The need for more cropland accounts for more than 60% of the world s deforestation (Myers, 1994). Despite such land replacement strategies, available world cropland *per capita* is declining and now stands at only 0.27 ha *per capita*. This is only about 50% of the 0.5 ha *per capita* considered the minimum for the production of a diverse diet, similar to that of the US and Europe (Lal and Stewart, 1990). China now has only 0.08 ha *per capita*, or about 15% of the accepted minimum.

Rainfall and its collection in rivers, lakes, and vast underground aquifers provides the water needed by humans for their survival and diverse activities. Ample supplies of fresh water are critical for all vegetation, including crops. All plants use and transpire massive amounts of water during the growing season. For example, a hectare of corn, producing about 8000 kg ha^{-1}, will transpire more than five million l of water during one growing season (Pimentel *et al.*, 1997b). This means that more than 8 to 10 million l of water must reach each hectare of corn during the season. In total, agriculture consumes about 70% of the world s fresh water supply, far more than any other human activity, thus leaving relatively little available for other uses (Postel, 1996).

Water resources are currently under great stress as densely populated cities, states, and countries increase their withdrawal of water from rivers, lakes, and aquifers every year. For example, by the time the Colorado River reaches Mexico, it is diminished to a trickle into the Gulf of California (Sheridan, 1983). Also, the great Ogalla aquifer in the central US is suffering an overdraft rate that is about 140% above recharge rate (Gleick, 1993). Water shortages in the US and elsewhere in the world are already reflected in the *per capita* decline in crop irrigation during the past 20 years.

To compound the water problem, about 40% of the world s population live in regions that directly compete for shared water resources (Gleick, 1993). In China, for example, more than 300 cities already are short of water, and these shortages are intensifying as Chinese urban areas expand (WRI, 1994). Competition for water resources among individuals, industries, and regions both within and between countries is growing throughout the world community (Gleick, 1993).

In addition to the quantity of water available, its purity is also important. Diseases associated with impure water and unsanitary systems rob people of their health, nutrients, and livelihood. Again, these problems are most serious in developing countries, where about 90% of diseases can be traced to a lack of pure water (WHO, 1992). Worldwide, about four billion cases of disease are contracted from impure water and approximately six million deaths are caused by waterborne disease each year (Pimentel *et al.*, 1997b).

Energy from many sources (especially fossil energy) is also a prime resource used in food production. Nearly 75% of the fossil energy used each year throughout the world is consumed by the people living in developed countries. Of this, about 17% is expended in the production, processing, and packaging of food products (Pimentel and Pimentel, 1996). In particular, the intensive farming technologies characteristic of developed countries rely on massive amounts of fossil energy for fertilizers, pesticides, irrigation, and for machines that substitute for human labor. In contrast, developing countries use fossil energy primarily for fertilizers and irrigation to help maintain yields, rather than to reduce human labor inputs (Giampietro and Pimentel, 1993).

Because fossil energy is a finite resource, its depletion accelerates as populations expand and their food requirements increase. The US Department of Energy warns that the United States will exhaust most or all of its oil reserves within the next 15 20 years (BP, 1994).

Worldwide, *per capita* supplies of oil resources show a significant decline since 1979 and this trend can be expected to continue (Duncan, 1999). Fossil energy, especially natural gas, for fertilizer production is having a negative impact on available amounts of fertilizers. For example, *per capita* fertilizer use has declined about 23% during the past decade (IFDC, 1998).

In addition to all these problems forcing the decline in per person food production, future global climate change will further depress food production. The projected higher temperatures will increase some insect pest, weed, and plant pathogen problems that are already causing the loss of more than 40% of all potential food production despite all the pesticide being applied worldwide (Pimentel, 1997; Rosensweig and Hillel, 1998; Chakraborty *et al.*, 1998). In addition, water shortages will also intensify (Postel, 1999). Higher temperatures will increase evapotranspiration from plants and land, and rainfall in some regions will decrease (Postel, 1999). These complications will result in some decline in global crop yields.

INFECTIOUS DISEASES

Waterborne infections account for about 80% of all infectious diseases worldwide, and 90% of all infectious diseases in developing countries (Epstein *et al.*, 1994). With global warming, projections suggest that many of these waterborne diseases will increase. For example, cholera, a well known waterborne disease, continues to be a serious global problem. *Vibrio cholerae* outbreaks are closely associated with climatic cycles (e.g., El Nino) and ocean plankton blooms (Colwell, 1996). These associations suggest that a potential influence of global warming will increase the spread and incidence of certain diseases, such as cholera (Colwell, 1996). In 1988, there were about 50 000 cases of cholera,

but by 1991 the number of cases had risen to 600 000; deaths from the disease rose from 2000 to 18 000 over the same period (Gleick, 1993). However, from 1991 to 1995, the number of cholera cases declined worldwide to about 210 000 per year, due to effective public health efforts.

Schistosomiasis, which is associated with contaminated fresh water, is also expanding worldwide and currently causes an estimated one million deaths annually (Gleick, 1993). This expansion is due to an increase in suitable habitats for the snail intermediate host population due to various human activities, including the construction of dams and irrigation channels (Shiklomanov, 1993). Global warming will probably contribute to the increase in schistosomiasis because the availability of suitable habitats, including warm water ecosystems, for the snail intermediate host will also increase (Martens *et al.*, 1997).

Mosquito-borne malaria is also associated with aquatic ecosystems. This disease infects more than 500 million people each year, killing, approximately 2.7 million (Marshall, 1997; Travis, 1997). Approximately 90% of all malaria cases occur in Africa, as do 90 95% of the world s malaria-related deaths. Between 1987 and 1994 in Rwanda, the increase in malaria incidence was greater than 500% and it continues to rise (Loevinsohn, 1994).

In some regions of Asia and South America, malaria is also on the rise. For example, in 1991, Peru recorded about 34 000 cases, but by 1992 that number had increased to about 55 000 cases per year (WHO, 1996c). Similarly, in Bangladesh, malaria prevalence increased from about 33 000 cases in 1988 to 64 000 cases in 1991; and to 125 000 cases in 1993 (WHO, 1996d).

Environmental changes, including more polluted water, have fostered the high incidence rates of malaria. Future global warming will also favor the spread and breeding of malaria-transmitting mosquitoes (Loevinsohn, 1994; Lindsay and Birley, 1996; Martens *et al.*, 1997; Woube, 1997; Lindsay and Martens, 1998; Kindson and Indaratna, 1998). Also, with some African populations doubling every 20 years, more people are living in close proximity to the water ecosystems most suitable for mosquito breeding. Concurrently, mosquito vectors are evolving resistance to insecticides (insecticides that also are polluting aquatic ecosystems), and the protozoan pathogens are evolving resistance to anti-malarial drugs, making control efforts less effective than before (Georghiou, 1990; Olliaro and Cattani, 1996).

In addition to its effects on global climate change, deforestation can cause other global changes. For example, deforestation in parts of Africa has exposed land to sunlight and promoted the development of temporary pools of water, thus favoring the breeding of malaria-transmitting mosquitoes, *Anopheles gambiae* (Coluozzi, 1994). Changes in human activities associated with deforestation can also change the incidence of exposure of humans to vectors (Walsh *et al.*, 1993). For instance, deforestation can expose humans to vectors that were previously zoophilic, leading to the transmission of various zoonotic agents, and can change the ecology of the land, providing new and different larval breeding sites.

Other infectious diseases transmitted by insect vectors that may increase with global climatic change include: onchoceriasis, filiarasis, dengue, yellow fever, Japanese encephalitis, and equine encephalitis (McMichael, 1993; McMichael and Beers, 1994; Murray and Lopez, 1997; Martens *et al.*, 1997; Lindgren, 1998; Patz *et al.*, 1998) (*see* Infectious Diseases, Volume 2; Urban Poverty and Environmental Health, Volume 3; Policy Responses to Public Health Issues Relating to Global Environmental Change, Volume 4).

LAND DEGRADATION AND DISEASE

Humans can acquire chemical pollutants and pathogens directly from the soil, by direct contact with it and/or indirectly through consumption of contaminated food and water. For example, sometimes soil particles themselves can be pollutants, entering eyes, nose, and mouth and acting as an irritant and/or allergen. Increases in temperature and an increase in wind intensity will intensify human exposure to these types of pollutants.

Cleared and exposed soil is especially susceptible to wind and water erosion. Wind erosion can cause serious health problems by blowing soil particles and microbes into the air. These wind-borne particles irritate respiratory tracts and eyes and aggravate allergies and asthma. Erosion can also disperse toxic chemicals, such as heavy metals and pesticides, leading to increases in contaminated food and water. Furthermore, erosion strips soil of its nutrients, lowering food crop productivity and ultimately reducing human nutrition. Professor R. Lal of Ohio State University reports that the addition of carbon dioxide of soil organic matter exposed because of soil erosion each year is 10 times greater than that contributed by all fossil combustion world wide.

As people invade more and more natural ecosystems and increasing amounts of land are cleared exposing more soil, the chances of humans becoming infected by helminths (like hookworms), and microbes (like pathogenic *Escherichia coli*), increase (WHO, 1992). Such increases were observed in Nepal, a mountainous country that experiences serious soil erosion and severe disease problems; 87% of the population was infected with helminths in 1984 (Suguri *et al.*, 1985; Metz, 1991). Children suffer greater morbidity from helminthic infections than adults because children need more protein than adults per kilogram of body weight, and under severe parasitic infections may be unable to utilize protein efficiently enough to remain healthy.

In addition, many species of helminths that infect humans are found in soil contaminated by human feces, thereby increasing the cycle of exposure. Worldwide, approximately two billion people are estimated to be infected with one or more species of helminths, either by direct penetration or by consumption of contaminated food or water (Hotez *et al.*, 1996). The most prevalent helminths are hookworms (*Necator americanus* and *Anclostoma duodenale*), *Strongyloides stercoralis*, and *Ascaris lumbricoides*. In locations where sanitation is poor and people are overcrowded, as in parts of urban Africa, up to 90% of the population may be infected with one or more helminths (Stephenson, 1994).

CONCLUSION

Global climate change will probably increase the prevalence of human diseases and the number of deaths caused by these diseases (*see* Environmental Change and Human Health: Extending the Sustainability Agenda, Volume 3). The ecology of increasing disease is exceedingly complex because of the diversity of infectious organisms and the impacts of environmental degradation on the prevalence of diseases. Rapid expansion of human populations has become a major factor in the rise of human diseases. Humans living in crowded urban areas provide an ideal ecosystem for the resurgence and rapid spread of old diseases, as well as the development and spread of new diseases (Pimentel *et al.*, 1998). Unprecedented increases in air, water, and soil pollutants, including organic and chemical wastes, further stress humans and increases disease prevalence. In particular, malnutrition enhances the susceptibility of humans to infectious pathogens and other diseases. Global climate change will probably improve the environment for growth of many vectors and intermediate hosts of diseases, thereby also increasing the spread and prevalence of diseases.

Global climate changes are also improving the environment for the development of some other diseases, such as malnutrition. Climate changes can be expected to further increase the susceptibility of food crops to certain pests, which could intensify food shortages and malnutrition. A concurrent problem is the rapid expansion of the number of environmental refugees and refugees from wars, floods, and other problems (Myers, 1993). These people, living in poverty and desperate for food, flee their homes in their search for survival. Their prevalent malnutrition and dislocation-related stress provide ideal conditions for further development of various diseases in these dense populations under environmental stress (*see* Refugees and Human Conflict, Volume 3; Environmental Refugees, Volume 4).

Many factors influence the ecology of the increasing number of human diseases now occurring worldwide. Currently, 40% of disease-related deaths result from diverse environment factors, including chemical pollutants, tobacco, and malnutrition. This growth in diseases is expected to continue as the global climate changes. Infectious diseases, which cause 37% of all deaths throughout the world, also are expected to increase.

To prevent diseases, poverty and malnutrition from worsening, the growing imbalance between an escalating human population and the earth s limited amount of natural resources must be dealt with. The crowding of people into urban areas; movement of populations into new environments; increased use of chemicals that pollute soils, water, and air; misuse of antibiotics that foster resistance in disease microbes; and growing malnutrition are all contributing factors to the ecological increase of human diseases worldwide.

See also: Urban Climate and Respiratory Disease, Volume 3; Viral Diseases and the Influence of Climate Change, Volume 3.

REFERENCES

ACC/SCN (1992) *Second Report on the World Nutrition Situation*, Vol. 1, Administrative Committee on Coordination, Subcommittee on Nutrition, United Nations, Washington, DC.

BP (1994) British Petroleum Statistical Review of World Energy, British Petroleum Corporate Communications Services, London.

Chakraborty, S, Murray, G M, Magarey, P A, Yonow, T, O Brien, R G, Croft, B J, Barbetti, M J, Sivasithamparam, M, Old, M J, Duszinski, M J, Sutherst, R W, Penrose, L J, Archer, C, and Emmett, R W (1998) Potential Impact of Climate Change on Plant Diseases of Economic Significance to Australia, *Australas. Plant Pathol.*, 27(1), 15 35.

Coluzzi, M (1994) Malaria and the Afrotropical Ecosystems: Impact of Man-made Environmental Changes, *Parasitologia*, 36, 223 227.

Colwell, R R (1996) Global Climate Change and Infectious Disease, *Science*, 274, 2025 2031.

Conway, G (1997) *The Doubly Green Revolution: Food for all in the 21st Century*, Penguin Press, London.

Duncan, R (1999) World Population and Energy-Use: the End of Exponential Growth, c. 1973, Manuscript.

Epstein, P, Ford, T, Puccia, C, and Possas, C D A (1994) Marine Ecosystem Health. Implications for Public Health. Diseases in Evolution, Global Changes and Emergence of Infectious Diseases, *Ann. N. Y. Acad. Sci.*, 740, 13 23.

FAO (1991) *Food Balance Sheets*, Food and Agriculture Organization of the United Nations, Rome.

Georghiou, P (1990) Overview of insecticide resistance, *Am. Chem. Soc. Symp. Ser.*, 421, 18 41.

Giampietro, M and Pimentel, D (1993) The Tightening Conflict: Population, Energy Use, and the Ecology of Agriculture, The NPG Forum, October, Negative Population Growth Inc., Teaneck, NJ.

Gleick, P H (1993) *Water in Crisis*, Oxford University Press, New York.

Grigg, D B (1993) *The World Food Problem*, Blackwell, Oxford.

Hominick, W (1993) *The Impact of Global Change of Disease*, Cambridge University Press, Cambridge.

Hotez, P J, Hawdon, J M, Cappello, M, Jones, B F, Ghosh, K, Volvovitz, F, and Xiao, S (1996) Molecular Approaches to Vaccinating Against Hookworm Disease, *Pediatr. Res.*, 40(4), 515 521.

Hotez, P J and Pritchard, D T (1995) Hookworm Infection, *Sci. Am.*, 272(6), 68 75.

IFDC (1998) Global and Regional Data on Fertilizer Production and Consumption 1961/62 1995/96, Muscle Shoals, International Fertilizer Development Center, AL.

Kendall, H W and Pimentel, D (1994) Constraints on the Expansion of the Global Food Supply, *AMBIO*, 23(3), 198 205.

Kidson, C and Indaratna, K (1998) Ecology, Economics, and Political Will: the Vicissitudes of Malaria Strategies in Asia, *Parassitologia*, 40(1 2), 3946.

Lal, R and Stewart, B A (1990) *Soil Degradation*, Springer-Verlag, New York.

Lindgren, E (1998) Climate Change, Tick-borne Encephalitis and Vaccination Needs in Sweden: a Prediction Model, *Ecol. Model.*, 110(1), 55 63.

Lindsay, S W and Birley, M H (1996) Climate Change and Malaria Transmission, *Ann. Trop. Med. Parasitol.*, 90(6), 573 588.

Lindsay, S W and Martens, W J M (1998) Malaria in the African Highlands: Past, Present, and Future, *Bull. WHO*, 76(1), 33 34.

Loevinshon, M E (1994) Climatic Warning and Increased Malaria Incidence in Rwanda, *Lancet* (North American Edition), 343(8899), 714 718.

Marshall, E (1997) African Malaria Studies Draw Attention, *Science*, 276, 275 299.

Martens, W J M, Jetten, T H, and Focks, D A (1997) Sensitivity of Malaria, Schistosomiasis, and Dengue to Global Warming, *Clim. Change*, 35(2), 145 146.

McMichael, A J (1993) *Planetary Overload: Global Environmental Change and the Health of the Human Species*, Cambridge University Press, Cambridge.

McMichael, A J and Beers, M Y (1994) Climate Changes and Human Population Health: Global and South Australian Perspectives, *Trans. R. S. S. Aust.*, 118(1 2), 91 98.

Metz, J J (1991) A Reassessment of the Causes and Severity of Nepal s Environmental Crisis, *World Dev.*, 19(7), 805 820.

Murray, C J and Lopez, A D (1997) Alternative Projections of Mortality and Disability by Cause 1990 2020: Global Burden of Disease Study, *Lancet* (North American Edition), 349(9064), 1498 1504.

Murray, C J L and Lopez, A D (1996) *The Global Burden of Disease*, World Health Organization, Geneva.

Myers, N (1993) *Gaia: an Atlas of Planet Management*, Anchor Press/Doubleday, Garden City, NY.

Myers, N (1994) Tropical Deforestation: Rates and Patterns, in *The Causes of Tropical Deforestation*, eds K Brown and D W Pearce, UBC Press, Vancouver, 27 41.

NAS (1994) *Population Summit of The World's Scienti c Academies*, National Academy of Sciences Press, Washington, DC.

Olliaro, P and Cattani, J (1996) Malaria, the Submerged Disease, *J. Am. Med. Assoc.*, 275(3), 230 233.

Patz, J A, Martens, W J M, Focks, D A, and Jetten, T H (1998) Dengue Fever Epidemic Potential as Projected by General Circulation Models of Global Climate Change, *Environ. Health Perspect.*, 106(3), 147 153.

Pimentel, D (1997) Pest Management in Agriculture, in *Techniques for Reducing Pesticide Use: Environmental and Economic Bene ts*, ed D Pimentel, John Wiley & Sons, Chichester, 1 11.

Pimentel, D, Harvey, C, Resosudarmo, P, Sinclair, K, Kurtz, D, McNair, M, Crist, S, Spritz, L, Fitton, L, Saffouri, R, and Blair, R (1995) Environmental and Economic Costs of Soil Erosion and Conservation Benefits, *Science*, 267, 1117 1123.

Pimentel, D, Harvey, C, Resosudarmo, P, Sinclair, K, Kurtz, D, McNair, M, Crist, S, Spritz, L, Fitton, L, Saffouri, R, and Blair, R (1997a) Land Use, Erosion, and Water Resources, in *Water Resources: Environmental Planning, Management, and Development*, ed A K Biswas, McGraw-Hill, New York, 37 71.

Pimentel, D, Houser, J, Preiss, E, White, O, Fang, H, Mesnick, L, Barsky, T, Tariche, S, Schreck, J, and Alpert, S (1997b) Water Resources: Agriculture, the Environment, and Society, *BioScience*, 47(2), 97 106.

Pimentel, D and Pimentel, M (1996) *Food, energy and society*, University Press of Colorado, Niwot CO.

Pimentel, D, Tort, M, D Anna, L, Krawic, A, Berger, J, Rossman, J, Mugo, F, Doon, N, Schriberg, M, Howard, E, Lee, S, and Talbot, J (1998) Ecology of Increasing Disease: Population Growth and Environmental Degradation, *BioScience*, 48(10), 817 826.

Postel, S (1996) *Dividing the Waters: Food Security, Ecosystem Health, and the New Politics of Scarcity*, Vol. 132, Worldwatch Institute, Washington, DC.

Postel, S (1999) *Pillar of Sand: Can the Irrigation Miracle Last?* W.W. Norton, New York.

PRB (1998) World Population Data Sheet, Population Reference Bureau, Washington, DC.

Rosenzweig, C and Hillel, D (1998) *Climate change and the Global Harvest*, Oxford University Press, New York.

RS and NAS (1992) The Royal Society and the National Academy of Sciences on population growth and sustainability, *Popul. Dev. Rev.*, 18(2), 375 378.

Sheridan, D (1983) The Colorado an Engineering Wonder Without Enough Water, *Smithsonian*, February, 45 54.

Shiklomanov, I A (1993) World Fresh Water Resources, in *Water in Crisis: a Guide to the World's Fresh Water Resources*, ed P Gleick, Oxford University, Oxford, 13 24.

Stephenson, L S (1994) Helminth Parasites: a Major Factor in Malnutrition, *World Health Forum*, 15(2), 169 172.

Suguri, S, Tongu, Y, Inatomi, S, and Pradhan, H D (1985) A Survey on Human Parasite Infections in Nepal, *Jap. J. Parasitol.*, 34(4), 285 292.

Travis, J (1997) Hitting Malaria Parasites Early and Hard, *Sci. News*, 151(2), 23.

Walsh, J F, Molyneux, D H, and Birley, M H (1993) Deforestation: Effects on Vector-borne Disease, in *The Impact of Global Climate Change on Disease*, ed W Hominick, Cambridge University Press, Cambridge, 55 75.

WHO (1992) *Our Planet, Our Health: Report of the WHO Commission on Health and Environment*, World Health Organization, Geneva.
WHO (1996a) *Micronutrient Malnutrition: Half the World's Population Affected*, World Health Organization, 13 November 1996, 78, 1 4.
WHO (1996b) *The World Health Report 1996*, World Health Organization, Geneva.
WHO (1996c) *World Malaria Situation in 1993*, Weekly Epidemiological Record, World Health Organization, 71(January), 18 20.
WHO (1996d) *World Malaria Situation in 1993*, Weekly Epidemiological Record, World Health Organization, 71(9 February), 41 44.
Woube, M (1997) Geographical Distribution and Dramatic Increases in Incidences of Malaria: Consequences of the Resettlement Scheme in Gambela, SW Ethiopia, *Indian J. Malariol.*, 34(3), 140 163.
WRI (1994) *World Resources 1994- 95*, World Resources Institute, Washington, DC.

FURTHER READING

Bouma, M J, Poveda, G, Rojas, W, Chavesse, D, Quinones, M, Cox, J, and Patz, J (1997) Predicting High-risk Years for Malaria in Colombia Using Parameters of El Nino Southern Oscillation, *Trop. Med. Int. Health*, 2(12), 1122 1127.
McClanahan, T R and Muthiga, N A (1998) An Ecological Shift in a Remote Coral Atoll of Belize over 25 years, *Environ. Conserv.*, 25(2), 122 130.

Marine Mammal Exploitation: Whales and Whaling

C Scott Baker[1] and Phillip J Clapham[2]
[1]*Auckland University, Auckland, New Zealand*
[2]*Northeast Fisheries Science Center, Woods Hole, MA, USA*

The utilization of marine mammals for subsistence or pro t has a long tradition in human history. Seals, sea lions, walruses and sea otters have been hunted for their fur, meat and oil since Neolithic times. Some of these species were later the target of organized commercial hunting, primarily for their pelts. Other coastal species, such as dolphins, manatees and dugongs have been hunted on an ongoing, if limited, basis by many societies around the world. In general, this exploitation did not differ markedly in scale or method from the hunting or trapping of terrestrial mammals, although in the case of the now extinct Stellar sea cow, the result was nonetheless devastating.

The hunting of whales stands out in the history of human resource exploitation because of its global expanse, its dependence on progressive technical innovation and the sheer scale of the slaughter. From its origins along the shores of the Basque province during the 11th century through to the factory whaling eets patrolling the Southern Ocean during the 20th century, commercial whalers have exploited an international biological resource for short-term gain, even when this was clearly at the expense of a long-term, sustainable utilization. In acting as though the oceans were an inexhaustible resource, the whaling industry brought about a tragedy of the commons on a grand scale. Time and again, agreement on limits to this exploitation was reached only when it was no longer pro table to hunt depleted species or populations. In the case of illegal whaling by the Soviet Union, even these belated agreements were ignored. With the measures of protection now in effect, some populations of whales are increasing in numbers and a few are approaching their pre-exploitation abundance. Other populations, however, remain precariously small and may be drifting towards extinction.

THE WHALERS' PRINCIPAL PREY

Although directed hunting of cetaceans has often involved smaller species such as dolphins and porpoises, the focus of commercial whaling was the great whales. This includes only one odontocete (toothed cetacean), the sperm whale, a pelagic species that underpinned the expansion of the New England whale fleet in the 18th and 19th centuries. Beaked whales, including the northern bottlenose whale and Baird s beaked whales, have been the target of more localized hunting at times during the 20th century. All of the other hunted large whales were mysticetes (baleen whales), a group that includes the largest of all animals, the blue and fin whales. There are 11 recognized species of baleen whales, divided into four families. These include the rorquals (blue, fin, sei, Bryde s, minke and humpback whales), the balaenids (the bowhead whale, as well as both northern and southern right whales), the gray whale, and the pygmy right whale. Several unresolved taxonomic issues exist concerning species or subspecies designations in blue, Bryde s, minke, bowhead and right whales.

Most of the baleen whales have a cosmopolitan distribution, but there are four exceptions. The bowhead is a Northern Hemisphere species that spends much of its life in the Arctic and is usually closely associated with sea ice. The Bryde s whale is found in all major oceans, but is the only mysticete whose distribution is concentrated in tropical or subtropical waters year-round. The gray whale is found today only in the North Pacific, although a population

existed in the North Atlantic as recently as the 1700s. The cause of its extinction is unknown, and may have been related to either whaling or natural factors, or a combination of the two. Finally, the pygmy right whale, a poorly understood Southern Hemisphere species, is the only mysticete to have never been a regular target of exploitation.

Most baleen whales undertake seasonal migrations from productive high-latitude summering areas, where they feed, to mating and calving grounds in warm waters, where fasting generally occurs. Some migrations, such as those of the humpback and gray whale, are the longest recorded for any mammal. Bowheads never leave high latitudes and, as noted above, Bryde s whales remain in warm water year-round. In other species, migratory patterns vary. Because the oil content of whales was greater during the summer feeding season or during migration from the feeding grounds, whalers tended to focus their hunting at these times. Whales on the return migration from the winter breeding grounds were dry, having lost between one-third and one-half of their renderable oils from fasting.

OPPORTUNISTIC ENCOUNTERS AND SUBSISTENCE HUNTING

Historically, the utilization of large whales took a variety of forms. Many primitive societies took advantage of occasional drift (stranded) whales as a supplement to other sources of nutrition and a source of material (bone and baleen) for tools, utensils and ceremonial items. The Wampanoag of the northeastern coast of the United States (US) and the Maori of New Zealand, for example, frequently availed themselves of beached animals, but did not actively engage in directed hunting of whales until colonial times. A relatively small number of societies took to sea in pursuit of whales. For these people, particularly the Inuit of the Arctic, whaling was, and continues to be, a central part of both their diet and culture. The origins of this aboriginal whaling probably date back thousands of years. A carving on a sandstone wall in southeastern Korea made about 6000 BC clearly shows a boatload of men harpooning a whale.

Methods of capture varied in directed aboriginal whaling. A variety of cultures in the Pacific region used darts tipped with the poison aconite to kill whales. This must have been a wasteful hunt: given the delayed effect of this poison, the number of animals eventually caught probably represented a small percentage of those actually struck and killed. Early harpoon-based fisheries used shafts tipped with stone points; in some Arctic hunts, sealskin floats were attached to the harpoon line to make the submerged whale easier to follow, and to impede its ability to move and dive, thus hastening the lethal strike with a lance or other weapon. Directed aboriginal whaling relied mainly on slow coastal species such as humpback, right, gray and bowhead whales. A number of cultures exploited the herding instinct of pilot whales by driving them into shallow bays where they could be killed easily; in addition, Arctic peoples often made use of ice entrapments (sassats) to kill entire pods of toothed whales, notably narwhals and belugas.

Examples of aboriginal whale hunts survive today in places as far apart as Indonesia, Greenland and North America. Many native societies that came into contact with the expanding European commercial whale fishery adopted its more efficient methods. By the late 1800s, this included the replacement of hand-thrown implements with guns or explosive harpoons. In some cases, modern subsistence hunting was initiated after contact with commercial whaling (as occurred during the 19th century in the Caribbean island of Bequia and the South Pacific Kingdom of Tonga), or was resumed after a long hiatus (such as with the Makah Indians of the northwestern coast of the US). Japan, which has a long history of opportunistic and directed whaling, developed systematic commercial hunting in the 17th century based on entangling whales in nets. This allowed villages to take species that had previously been difficult to capture by harpooning alone, including gray, right, humpback and fin whales.

EARLY EUROPEAN AND YANKEE WHALING

The earliest sustained commercial whaling, for which reliable records exist, was the Basque right whale fishery in the Bay of Biscay. Market records and tax documents indicate that this enterprise was operating as early as the 11th century in southern France and northern Spain. The Basques were instrumental in the expansion of European whaling into the New World: their whalers were operating by 1530 in Labrador, where they killed right whales as well as bowheads. Declining catches combined with general European exploration resulted in the opening of new whaling grounds in the Arctic in the early 17th century. This expansion was led by the Dutch and Germans, closely followed by the British and French, with the Basques often employed as harpooners. Bowheads were the principal targets of the Arctic fishery, and exploitation was so intense that the species was effectively extirpated from some areas, notably Spitsbergen. During the same period, colonial settlers in New England began hunting right whales. This pursuit began in earnest in the late 1600s and crashed by 1750, when few whales remained in the coastal waters of the region.

Early European whalers, although using large sailing vessels to find whales, had relied on towing the carcasses to shore for processing, a practice that limited them to coastal or ice-edge species. Despite this limitation, the slow moving right and bowhead whales were depleted (and the gray whale was extinct) in the North Atlantic by the late 18th century and whalers were forced to look to other

oceans and species for profit. This expansion was led by the Yankee whalers of New England and aided by technological improvements, including larger vessels fitted with a central brick furnace and large iron pots, referred to as the try-works, used for rendering the oil from the blubber and scraps of the whale. In 1791, Thomas Jefferson, in his role of Secretary of State, wrote of the American whalers, They followed the whales farther and farther into the ocean, enlarging their vessels to 60, 100 and 200 tons. Having extended their pursuit . . . , they fell in with the spermacaeti whale.

At the beginning of the 19th century, New England whalers were venturing into the Southern Hemisphere in pursuit of right and sperm whales. In 1835, the first North Pacific right whales were killed in the Gulf of Alaska, and soon after, the fishery expanded to the western side of the ocean and the Okhotsk Sea. By 1849, right whales were sufficiently scarce in areas of the North Pacific to cause many whalers to switch their focus to the Bering and Beaufort Seas, where newly discovered stocks of bowheads were being plundered. Later in the 19th century, and continuing into the 20th century, shore stations on the west coast of North America sprang up, taking abundant humpback and gray whales for a few years before local depletion caused the stations closure.

MODERN WHALING AND THE SOUTHERN OCEAN

During the 1860s, two resourceful whalers (one Norwegian, the other American) independently introduced an innovation that was to change forever the methods of whaling. The harpoon gun (fired from the shoulder or mounted in more powerful form on the bows of catcher boats) permitted whalers to dispatch whales more swiftly than had ever been possible with hand-thrown harpoons and lances. This invention, together with the equally revolutionary replacement of sail with steam power, rendered even the fastest rorquals fair game, and ushered in the modern era of industrial whaling. Additional improvements in efficiency resulted from use of the compressor to inflate carcasses that would otherwise sink and be lost. Finally, in 1925 the introduction of the factory ship freed whalers from any remaining dependence on land stations for processing and thus permitted independent, fully mechanized whaling on the high seas.

The technical advances of modern whaling coincided with the discovery of the richest whaling grounds on Earth. In 1904, a Norwegian whaler ventured into the Southern Ocean near the island of South Georgia and noted with astonishment, I see them in hundreds and thousands. Every species of baleen whale except the bowhead, gray and Bryde s existed in vast numbers in the extraordinarily productive feeding grounds of the Antarctic. With this discovery, the stage was set for a slaughter that has few parallels in the history of wildlife exploitation. Between 1904 and 1980, approximately two million whales were killed in the Southern Hemisphere.

Although catch reports to international agencies for this period showed rapid declines in most species, the full extent of this exploitation was not revealed until 1994 when a group of Russian scientists presented the true catch records of several Soviet factory ships. From shortly after World War II until the late 1970s, the Soviet vessels systematically falsified their records, vastly under-reporting the true catches of some species. For humpback whales, the reported catch was 2710 but the true catch was 48 477. For the southern right whale, which had been depleted in the 19th century and protected by international agreement since the 1930s, the reported catch was four but the true catch was 3354. Accounting for these newly available records shows that the true extent of modern whaling in the Southern Hemisphere included over 200 000 humpbacks, 350 000 blue whales, almost 400 000 sperm whales and a staggering 725 000 fin whales. Although we will never know the true abundance prior to whaling, it is probable that between 80 and 95% of the pristine populations of these species were killed during this period.

THE MORATORIUM AND STATUS OF STOCKS

The International Whaling Commission (IWC) was established 1947 in recognition of the interest of the nations of the world in safeguarding for future generations the great natural resources represented by the whales stocks. Unfortunately, this admirable objective was not quickly or easily achieved. Although whale stocks were already in decline, and a few species had been protected by previous international agreements, the IWC made little effort and had little power to imposed catch limits. By the 1960s even the seemingly boundless stocks of rorquals in the Southern Hemisphere, particularly the humpback and blue whales, had all but disappeared. In 1972, the United Nations recognized the need for urgent action and called for a moratorium on whaling to allow the recovery of stocks. Not until 10 years later, however, did the IWC admit failure of its past management and agree to the need for a moratorium. This moratorium took effect in 1986, banning all commercial whaling and providing time to formulate an adequate management plan for any future hunting. Following several years of development and debate, the IWC agreed to a Revised Management Procedure based on a computer model to determine levels of allowable catch for any future hunting.

Today, some whales are recovering from overexploitation, although few have begun to approach their pre-exploitation abundance. The California gray whale is an exception: after nearly a century of protection, it is now

thought to number 26 000 individuals, equaling or exceeding pre-exploitation abundance. Some humpback and southern right whale populations appear to be experiencing strong growth, despite the massive damage inflicted upon them in the 20th century. Nonetheless, a number of exceptions exist involving populations that remain dangerously small. Among the most critically endangered are the North Pacific and North Atlantic northern right whales, bowhead whales of the eastern Arctic, western Pacific gray whales and most blue whale populations. In addition, some local populations appear to have been almost entirely extirpated; examples include the humpbacks and blue whales that once fed in huge numbers around South Georgia, New Zealand humpback whales, Spitsbergen bowheads, Gibraltar fin whales, North Atlantic right whales, and blue whales off Japan.

The reasons for the variability in recovery of stocks are unknown but could include a loss of cultural tradition or fidelity to specific migratory destinations. Recent behavioral and genetic studies indicate the importance of a calf s early experience in determining migratory traditions. If these traditions are lost by local extinction, the pattern of recovery will be patchy. Another possibility is the as yet poorly understood consequences of past illegal or pirate whaling. With the exception of the former Soviet scientists data, which are still being evaluated, most illegal whalers kept poor catch records. Finally, even low levels of ongoing illegal hunting could be depressing recovery in some stocks. Molecular genetic surveys of whale meat markets of Japan and Korea show that protected species and stocks are still available more than a decade after the moratorium.

ENVIRONMENTAL CHANGE AND THE FUTURE

The ecological impacts of whaling have been much debated, especially for the Antarctic ecosystem where whales were the dominant predator. The principal prey for virtually all mysticetes in the Southern Ocean is krill (euphausiid crustaceans). Even a moderate size whale, such as a humpback, probably requires more than a ton (900 kg) of krill a day for much of the 90 120 day feeding season. By some estimates, the depletion of whales has liberated more than 150 million tons of krill a year into the Antarctic marine ecosystem. This is several times more biomass than taken each year in all the fisheries of the world (estimated to be 60 70 million tons in 1980). Although there is evidence that populations of other predators (such as penguins and some pinnipeds) have expanded as a result of increased krill availability, it is unlikely this has replaced the role of whales. In particular, these other predators do not migrate and the incidental transport of nutrients to tropical waters by formerly abundant whale populations is now vastly reduced. Data on possible density-dependent effects in surviving whale populations are more equivocal, and there is little evidence to support the hypothesis that the recovery of some whale populations is inhibited by interspecific competition from others.

Humans do not as yet colonize the oceans; consequently, unlike many terrestrial mammals, competition for habitat space is not strictly an issue for whales. However, cetaceans are increasingly threatened with human competition for food resources, as well as with habitat degradation from pollution, noise, ship traffic and proliferation of fishing gear. The North Atlantic right whale population, depleted by whaling prior to the 20th century, is now struggling due to mortality from collisions with ships and entanglement with fishing nets in the busy coastal waters of the northeastern US and Canada. The impacts of global climate change on whales (and thus on whaling) are more difficult to assess but could constitute a serious threat in the Southern Hemisphere. Studies have shown that increased ultra violet radiation from ozone depletion can inhibit growth of phytoplankton on which krill feed. Independently, global warming may be influencing the ecological dynamics of invertebrates in the Antarctic marine system. In some areas, salps (tunicates), which are not a dietary component of whales, seem to be displacing krill. Interestingly, data on whales and whaling also provide an indication of global climate change. Whaling records (collected since 1937) appear to show an abrupt southerly movement of the sea-ice edge between the mid 1950s and early 1970s. This suggests a decline of some 25% in the area covered by sea ice, and could imply changes in Antarctic deep-water formation and, consequently, biological productivity. The actual impact of these changes on the recovery of whales remains to be seen. Like all other life on Earth, however, whales are now unwitting participants in a great and uncontrolled experiment in global ecology and climate change which began with the onset of the industrial revolution.

FURTHER READING

Aguilar, A (1986) A Review of Old Basque Whaling and its Effect on the Right Whales (*Eubalaena glacialis*) of the North Atlantic, *Rep. Int. Whaling Commission*, Special Issue 10, 191 199.

Baker, C S and Palumbi, S R (1994) Which Whales are Hunted? A Molecular Genetic Approach to Monitoring Whaling, *Science*, 265, 1538 1539.

Baker, C S and Palumbi, S R (1995) Population Structure, Molecular Systematics and Forensic Identification of Whales and Dolphins, in *Conservation Genetics: Case Histories from Nature*, eds J Avise and J L Hamrick, Chapman and Hall, New York, 10 49.

Bengtson, J L and Laws, R M (1985) Trends in Crabeater Seal Age at Maturity: an Insight into Antarctic Marine Interactions, in *Antarctic Nutrient Cycles and Food Webs*, eds

W R Siegfried, P R Condy, and R M Laws, Springer-Verlag, Berlin, 669 675.
Brown, S G and Lockyer, C H (1984) Whales, in *Antarctic Ecology*, ed R M Laws, Academic Press, London, 717 781.
Bryant, P J (1995) Dating remains of Gray Whales from the Eastern North Atlantic, *J. Mammal.*, 76, 857 861.
Clapham, P (1997) *Whales of the World*, Voyageur Press, Stillwater, MN, 1 132.
Clapham, P J and Brownell, Jr, R L (1996) Potential for Interspecific Competition in Baleen Whales, *Rep. Int. Whaling Commission*, 46, 361 367.
Clapham, P J and Brownell, Jr, R L (2000) Vulnerability of Migratory Baleen Whales to Ecosystem Degradation, *Convention on Migratory Species Spec. Pub.*, (in press).
Clapham, P J, Young, S B, and Brownell, Jr, R L (1999) Baleen Whales: Conservation Issues and the Status of the Most Endangered Populations, *Mamm. Rev.*, 29, 35 60.
Cumbaa, S L (1986) Archaeological Evidence of the 16th Century Basque Right Whale Fishery in Labrador, *Rep. Int. Whaling Commission*, Special Issue 10, 187 190.
Delamare, W K (1997) Abrupt Mid-twentieth Century Decline in Antarctic Sea-ice Extent from Whaling Records, *Nature*, 57 60.
Ellis, R (1991) *Men and Whales*, Alfred A Knopf, New York, 1 542.
Evans, P G H (1987) *The Natural History of Whales and Dolphins*, Facts on File Publications, New York.
IWC (1997) Report of the IWC Workshop on Climate Change and Cetaceans, *Rep. Int. Whaling Commission*, 47, 293 313.
Kellogg, R (1929) What is Known of the Migrations of some of the Whalebone Whales, *Annu. Rep. Smithsonian Inst.*, 1928, 469 494.
Kock, K H and Shimadzu, Y (1994) Trophic Relationships and Trends in Population Size and Reproductive Parameters in Antarctic High Level Predators, in *Southern Ocean Ecology: the BIOMASS Perspective*, ed S Z El-Sayed, Cambridge University Press, Cambridge, 287 312.
Laws, R M (1977) Seals and Whales of the Southern Ocean, *Philos. Trans. R. Soc. Lond.*, 279, 81 96.
Little, E D and Andrews, J C (1980) Drift Whales at Nantucket: the Kindness of Moshup, *Man in the Northeast*, 23.
Loeb, V, Siegel, V, Holmhansen, O, Hewitt, R, Frase, W, Trivelpiece, W, and Trivelpiece, S (1997) Effects of Sea-ice Extent and Krill or Salp Dominance on the Antarctic Food Web, *Nature*, 387, 897 900.
Mead, J G and Brownell, Jr, R L (1993) Order Cetacea, in *Mammal Species of the World*, eds D E Wilson and D M Reeder, Smithsonian Institution Press, Washington, DC, 349 364.
Mead, J G and Mitchell, E D (1984) Atlantic Gray Whales, in *The Gray Whale*, eds M L Jones, S L Swartz, and S Leatherwood, Academic Press, New York, 33 53.
Melville, H (1851) *Moby Dick*, Bantam Classics edition, 1967, 1 594.
Mizroch, S A and York, A E (1984) Have Pregnancy Rates of Southern Hemisphere Fin Whales, *Balaenoptera physalus*, Increased? *Rep. Int. Whaling Commission*, Special Issue 6, 401 410.
Neale, P J, Davis, R F, and Cullen, J J (1998) Interactive Effects of Ozone Depletion and Vertical Mixing on Photosynthesis of Antarctic Phytoplankton, *Nature*, 585 589.
Omura, H (1986) History of Right Whale Catches in the Waters around Japan, *Rep. Int. Whaling Commission*, Special Issue 10, 35 41.
Reeves, R R (2001) The Origins and Character of Aboriginal Subsistence Whaling: A Global Review, *Mamm. Rev.*, 32, in press.
Sanderson, I T (1956) *Follow the Whale*, Little, Brown and Company, Boston, MA, 1 423.
T nnessen, J N and Johnsen, A O (1982) *The History of Modern Whaling*, C Hurst, London, 1 788.
Tuck, J A and Grenier, R (1987) A 16th Century Basque Whaling Station in Labrador, *Sci. Am.*, 126 136.
Yablokov, A V (1994) Validity of Whaling Data, *Nature*, 367, 108.

Marine Pollution: the GESAMP Definition

see GESAMP (Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection) (**Volume 4**)

Marsh, George Perkins

(1801 - 1882)

George Perkins Marsh, US diplomat and scholar, provided one of the first extensive examinations of environmental degradation, in *Man and Nature; or, Physical Geography as Modi ed by Human Action* (1864). In this book he discussed the extinction of species, the consequences of forest clearing for streamflow, and other topics. Based on extensive study and travel, his portrayal was of a harmonious world torn apart by greed and ignorance. To check the abuse of nature, it was necessary, Marsh argued, not to reject all exploitation, but to exercise foresight, guided by a scientific understanding of the world.

Marsh s essential message was not entirely novel. Writers in France, Germany, Italy and elsewhere in Europe (especially the geographer Alexander von Humboldt), as well as scientists in India and other colonies, had expressed concerns during preceding decades about human impacts on the environment, particularly the effects of forest clearing on species, streams, and local climate. Marsh drew on many of these writers (he could read in 20 languages), and also observed conditions while traveling in the Mediterranean region and Europe. Thus, *Man and Nature* was especially significant because it brought together an array of existing knowledge and experience, synthesizing in an accessible format the perceptions of scientists in various parts of the world, and explaining their relevance to North America and other temperate regions.

Marsh s views were also shaped by his personal experience of land clearing and settlement in his home state of Vermont. By 1847 he had noted the consequences of clearing: erosion, streams in which droughts alternated with floods, impoverished fish communities (Marsh, 1848). Ten years later he advised the Vermont Legislature on the declining fisheries of the state. In his report he anticipated *Man and Nature*, describing a nature in the shorn and crippled condition to which human progress has reduced her (Marsh, 1857).

Although quoted elsewhere, Marsh was especially influential in North America. *Man and Nature* sold well, and within a decade was described as a classic. His critique of the once widespread view of deforestation as improvement proved persuasive: many Americans became convinced that clearing led to unregulated streamflow, deterioration in climate, and poorer conditions for farming. In response, Congress established a national forestry commission and forest reserves in 1873. His influence was also eventually felt by 1900 through the progressive conservation movement, when Gifford Pinchot of the US Forest Service, with others, advocated efficient use of resources through professional management. Closely linked to the advocacy of scientific management in business and government, the conservation movement accelerated the application of expertise to the management of forests, water and other resources. More recently, with the widespread emergence of environmental concerns since the 1960s, *Man and Nature* has been recognized as a prescient warning of the hazards of environmental degradation.

REFERENCES

Marsh, G P (1848) *Address Delivered before the Agricultural Society of Rutland County*, September 30, 1847, Rutland.

Marsh, G P (1857) *Report, Made Under Authority of the Legislature of Vermont, on the Arti cial Propagation of Fish*, Free Press Print, Burlington, VT.

Marsh, G P (1864) *Man and Nature, or, Physical Geography as Modi ed by Human Action*, reprint edition, ed D Lowenthal, Harvard University Press: Cambridge, MA, 1965.

FURTHER READING

Lowenthal, D (2000) *George Perkins Marsh: Prophet of Conservation*, University of Washington Press, 1 650.

STEPHEN BOCKING *Canada*

Marshes, Anthropogenic Changes

Pat Hutchings
The Australian Museum, Sydney, Australia

Tidal marshes occur on sheltered shores between approximately mean high water level of neap tides and the highest high waters of spring tides, and are periodically inundated with water ranging from fully marine to brackish waters. They are often referred to as salt marshes. Tidal marshes occur worldwide, from subpolar regions to the tropics, and depending upon latitude they may be dormant for varying periods of time in the winter, or may continue growing all year although often at varying rates.

No detailed worldwide inventory of tidal marshes exists although the largest area occurs in North America where extensive *Spartina* marshes are found along the east coast (occupying in excess of two million ha; Frey and Basan, 1985) and marshes dominated by *Distichlis* are present on the west coast (Dawe and White, 1986). Australia has more than one million ha of salt marsh (IPIECA, 1994). Tropical marshes have fewer plant species than temperate ones (Adam, 1990), although whether this relationship is also valid for the associated fauna and microflora is not known. Attempts to classify the vegetation of marshes based on floristic and vegetational criteria have been made by Chapman (1977). Within a tidal marsh, plant species are commonly zoned according to tidal height. The zonation may be dynamic with the marsh plants facilitating deposition of fine sediments, leading to increasing height relative to tidal datum and changes in species composition. Similarly, the distribution, abundance and species composition of the fauna are related to plant species distribution and structure as well as latitude, although local factors such as leaf litter, surface topography, salinity, sediment

characteristics and the distribution of surface microflora all play a major role (Hutchings, 1994). However, there are relatively few quantitative studies on the distribution and abundance of the macrofauna of tidal marshes, and many species exhibit major fluctuations in populations over time.

The terrestrial and seaward boundaries of tidal marshes fluctuate as sediment patterns and tidal regimes change. Along the New England coast, accretion rates in salt marshes over the last few thousand years have kept pace with increases in relative sea level. This has allowed the salt marshes to extend both landward over previously terrestrial or freshwater wetlands and seaward onto mudflats (Niering and Warren, 1980). However, landward encroachments of marshes are increasingly prevented by the presence of seawalls and other barriers. Encroachment of adjacent mangroves into salt marshes may lead to loss of such habitats unless active measures are taken to maintain the marshes (Burchett *et al.*, 1998). The reasons for this mangrove encroachment in parts of Sydney Harbor are not fully understood, although increasing urbanization of the catchment and with it increasing sedimentation may be a contributing factor. In coastal areas experiencing tectonic subsidence such as SE Australia and Irian Jaya, fringing mangroves are encroaching into the salt marshes as land levels drop (cited by Bird, 1993).

So not only do we not have a detailed inventory of the areas currently occupied worldwide by this important ecosystem (Adam, 1990, and references therein), we have limited knowledge of how these areas have fluctuated globally since the last major sea level rise, 6000 years ago (Adam, 1990; Bird, 1985; Viles and Spencer, 1995 for specific examples), although we have some detailed studies of individual areas during the last 100 years or so. For example, in the Fal estuary in SE England, the salt marsh advanced seawards by up to 900 m between 1878 and 1973, while its landward margins were colonized by scrub and woodland vegetation (Bird, 1985). This clearly illustrates the potential dynamic nature of salt marshes both in terms of distribution and species composition. So it is unclear as to whether the total area occupied by tidal marshes has increased or decreased globally since the last major sea level rise, although changes have certainly occurred within regions. A variety of processes control the distribution of tidal marshes, including sea level and associated tidal regimes and sedimentation and accretion processes. Knowledge of the distribution of tidal marshes prior to 6000 years ago is limited, as marshes provide few fossilized deposits. Kennish (1986) has suggested that the available area for the development of tidal marshes was small before this time, and the period of rapidly fluctuating sea levels between 15 000 and 6000 years ago would also have prevented their development.

The hydrological and topographical conditions which enable tidal marshes to develop are often identical to those needed for ports and associated settlements. For example, in Australia all but one of the capital cities and most other cities and large towns are located on an estuary or a sheltered bay and have associated ports. This means that tidal marshes are often in close proximity to centers of human population, and are therefore subjected to anthropogenic impacts. Such impacts may have occurred over many centuries. In addition, in many countries coastal populations are generally increasing, e.g., in California and the east coast of Australia.

The following is a brief summary of the impacts of humankind on tidal marshes:

1. A long established use of salt marshes has been the clearing of vegetation for the creation of ponds in which sea water is trapped and then allowed to evaporate to produce salt, a practice still prevalent in many countries.
2. Tidal marshes are often reclaimed by dredging and or drainage for developments such as marinas and port developments (Pinder and Witherick, 1990). Such major development may require the redirection of wave energy, changing tidal and wave patterns in the area. Wave-deflecting barriers incorporated into the development of Port Botany, Botany Bay, redirected waves across the Bay leading to massive erosion of tidal marshes.
3. The proximity of tidal marshes to ports makes them vulnerable to oil spills, and they are efficient at trapping oil. The ability of the marshes to recover varies and in some cases it is suggested that it is preferable to allow the marsh to recover naturally (IPIECA, 1994) than to try to remove the oil either physically or chemically. Hopefully, an increasing use of booms will reduce the numbers of oil spills or at least the extent of cover.
4. Tidal marshes are affected by changes in freshwater input and storm water discharges, both in terms of quantity and quality of input. Redirection of storm water in the Swan River, Perth has led to a loss of tidal marshes and increases in the area occupied by the sedge *Typha* (Thurlow *et al.*, 1986). Such changes will also presumably affect the microalgal and meiofaunal communities and this may impact on the macrofaunal communities. Storm water runoff with high levels of nutrients will also impact on these communities.
5. Increasing damming of rivers for fresh water may lead to the salinity wedge encroaching further upstream, changing salinity patterns and so modifying marsh communities, both animals and plants.
6. Deliberate plantings of *Spartina* in southern England in the late 1800s and the accidental formation of a hybrid *Spartina anglica* which is now the dominant saltmarsh species throughout Northern Europe, has drastically altered the ecology of this area (see Adams, 1990 for a discussion). As this species is tolerant of some tidal submergence, it has colonized mud flats and may have

reduced the value of these sites for wading birds (Davis and Moss, 1984; Millard and Evans, 1984). The hybrid has been planted in New Zealand, southern Australia and China, and forms extensive colonies replacing endemic communities. At least in Australia such a practice has stopped, but the species is still colonizing new areas.

7. While tidal marshes are relatively weed free because of the salinity conditions which prevail, some introduced species have become pests. In Australia these include the sedge *Juncus acutus* and the American species *Baccharis* ssp. which has been declared a noxious weed in Queensland and has been implicated in causing asthma.
8. Pumping offshore for oil and gas and extraction of ground water has led to subsidence along some coastal areas which has then led to sea level rises and flooding of coastal areas (see Bird, 1993 for details). In other areas, offshore mining for sand has modified rates of accretion in coastal areas.
9. Tidal marshes have been used for livestock grazing for centuries, which is important in northern Europe (Dijkema, 1984), eastern Canada (Roberts and Robertson, 1986) and Japan (Ishizuka, 1974), primarily cattle and horses, although the practice is slowly declining in these areas. The grazing regime is a major factor controlling diversity of plant species (see Adam, 1990 for discussion of impact of grazing).
10. As tidal marshes often occur near populated areas, they may be subjected to frequent spraying of insecticides to control mosquitoes and biting midges, which also impacts on other invertebrates as well; in addition, the physical damage occurs as a result of trampling through the marsh. In the past, swamps were often drained to control these insects.

While the above impacts could potentially be quantified for particular regions, this has not been done. It is far more difficult to evaluate the impact of global warming and associated rising sea levels and increased frequency of storms, and to separate the impact of these factors from the naturally dynamic nature of tidal marshes, which fluctuate in area and species composition over time. In addition, tidal marshes are impacted by a variety of localized anthropogenic impacts as discussed above. So several factors, natural and anthropogenic, are responsible for changes in area or type of tidal marshes. Separating these factors is not a feasible option at this stage as there is no detailed global inventory of tidal marshes or changes over time. Certainly in some areas, loss of tidal marshes has occurred but whether losses have been cancelled out by gains elsewhere is not known. However, tidal marshes are diverse systems and a system in SE Asia cannot be equated with an Arctic marsh.

While in the past, tidal marshes could migrate landwards with rising sea level, increasingly this will be prevented by the construction of seawalls. Although some data from the Gulf of Mexico and the Atlantic coast of the USA suggest that rates of accretion in most cases exceed rates of sea level rise, where land is available for boundary movement. With increasing temperatures, the species composition of tidal marshes will change, not just in terms of species, but also in terms of structure and presumably associated changes in the microflora and both macro- and microfauna. Limited data suggest that this will affect bird populations that either roost in the tidal marsh after feeding on nearby mud flats or those species which feed in the marshes. Increasing storm activity may lead to changes in sediment distribution and erosion, and the ability of them to recover will in part be dependent on the frequency of such events.

Certainly global warming and changes in sea level, together with other anthropogenic impacts discussed above, will change the distribution and structure of tidal marshes. Whether these changes will lead to a reduction in the areas occupied by this habitat is unclear. However, certainly in tidal urbanized areas physical barriers will curtail any expansion of marshes inland and this will lead to losses of this habitat on a regional scale. But will these losses be compensated by expansion in non-developed areas? Also unclear is the impact of either physical loss of tidal marshes, or changes in species composition and how this will affect nearby tidal waters. Expanding tidal marshes tend to function as net carbon importers but as the marsh matures, it may become a carbon exporter as well as being important in the transformation of compounds (Adam, 1990 and references therein), with considerable variation occurring between marshes. So changes in species composition of tidal marshes induced by rising sea levels and global warming will certainly change the productivity of the marshes. However, few attempts have been made to construct budgets for individual tidal marshes, let alone regional or global ones, so the impact of changes in the distribution and composition of these habitats is difficult to assess or how these will impact adjacent ecosystems. One should therefore practice the precautionary approach and attempt to prevent any further loss of tidal marshes, and to undertake further studies to complete our understanding of the role of various types of tidal marshes.

See also: Mangrove Ecosystems, Volume 2.

ACKNOWLEDGMENTS

I should like to thank Paul Adam for useful discussions and for making his library available to me.

REFERENCES

Adam, P (1990) *Saltmarsh Ecology*, Cambridge Studies in Ecology, University Press, Cambridge.

Bird, E C F (1985) *Coastline Changes*, Wiley Interscience, Chichester.
Bird, E C F (1993) Submerging Coasts, The Effects of a Rising Sea Level on Coastal Environments, John Wiley & Sons, Chichester.
Burchett, M D, Pulkownik, A, Grant, C, and Macfarlane, G (1998) Rehabilitation of Saline Wetlands, Olympic 2000 Site, Sydney (Australia) 1: Management Strategies Based on Ecological Needs Assessment, *Mar. Pollut. Bull.*, 37(8 12), 515 525.
Chapman, V J (1977) *Wet Coastal Ecosystems*, Elsevier, Amsterdam.
Davis, L V and Moss, D (1984) *Spartina* and Waders the Dyfi Estuary, in *Spartina anglica in Great Britain*, ed P Doody, Nature Conservation Council, Attingham Park, 37 40.
Dawe, N K and White, E R (1986) Some Aspects of the Vegetation Ecology of the Nanoose Bonell Estuary, Vancouver Island, British Columbia, *Can. J. Bot.*, 64, 27 34.
Dijkema, K S (1984) Development and Classification of Main Salt Marsh Biotopes in Europe, in *Salt Marshes in Europe* ed K S Dijkema, European Committee for the Conservation of Nature and Natural Resources, Strasbourg, 8 15.
Frey, R W and Basan, P B (1985) Coastal Salt Marshes, in *Coastal Sedimentary Environments*, 2nd edition, ed R A Davis, Springer-Verlag, New York, 225 301.
Hutchings, P (1994) *Epifauna of Mangroves and Salt Marshes, Jervis Bay Baseline Studies Final Report, May 1994*, Commonwealth of Australia, CSIRO Division of Fisheries, 617 654.
IPIECA (1994) *Biological Impacts of Oil Pollution: Salt Marshes*, Petroleum Industry Environmental Conservation Association, London.
Ishizuka, K (1974) Maritime Vegetation, in *The Flora and Vegetation of Japan*, ed M Numata, Elsevier, Amsterdam, 151 172.
Kennish, M J (1986) *Ecology of Estuaries, Physical and Chemical Aspects*, Vol. 1, CRC Press, Boca Raton, FL.
Millard, A V and Evans, P R (1984) Colonization of Mudflats by *Spartina anglica*: Some Effects on Invertebrate and Shore Bird Populations at Lindisfarne, in *Spartina anglica in Great Britain*, ed P Doody, Nature Conservation Council National Research Council, Attingham Park, 41 48.
Niering, W A and Warren, R S (1980) Vegetation Patterns and Processes in New England Salt Marshes, *Bioscience*, 30, 301 307.
Pinder, D A and Witherick, M E (1990) Port Industrialization, Urbanizations and Wetland Loss, in *Wetlands: a Threatened Landscape*, ed M Williams, Institute of British Geographers Special Publication 25, Blackwells, Oxford, 234 266.
Roberts, B A and Robertson, A (1986) Salt Marshes of Atlantic Canada: Their Ecology and Distribution, *Can. J. Bot.*, 64, 455 67.
Thurlow, B H, Chambers, J, and Klemm, V V (1986) *Swan-Canning Estuarine System. Environment, Use and the Future*, Waterways Commission Perth West Australia, Report No. 6.
Viles, H and Spencer, T (1995) *Coastal Problems. Geomorphology Ecology and Society at the Coast*, Edward Arnold Member of the Hodder Headline Group, London.

Materials Flows for Mining and Quarrying

Ian Douglas and Nigel Lawson
University of Manchester, Manchester, UK

The materials ows due to mining and quarrying directly link industrial activity to the sustainable use of resources and are a major component of land surface change. To the minerals or ore extracted must be added all the overlying rocks and soils that are removed and dumped elsewhere, even temporarily. The total mass of material moved in these extractive industries exceeds the mass of sediment carried annually by all the world's rivers to the sea and is approximately three times the amount of mineral matter actually entering industrial processing streams. The hidden ows of overburden removal and ore processing waste have to be taken into account when considering the contribution that mining and quarrying make to global environmental change.

INTRODUCTION

Mining and quarrying involve some the greatest transformations of the landscape through human activity; often creating huge pits and waste heaps. Mineral extraction is broadly divided into three basic methods: open-pit or surface, underground, and solution mining. Open workings are the dominant form of extraction of the main commodities mined or quarried: coal and aggregates. Surface, or open-pit mining requires rock, soil and vegetation removal to reach mineral deposits. The waste rock, or overburden, is piled near the mine. The workings have large energy requirements and produce emissions to the atmosphere and discharges to nearby water bodies. For any particular mine, these hidden flows are often greater in magnitude than the mass of mineral or ore extracted for processing. This extraction of materials from the natural environment and their transfer to urban and industrial areas has to be considered as a geological and geomorphological agent contributing to global environmental change.

Despite having relatively static total populations, the industrialized economies are increasing their use of materials, particularly construction materials, as households become smaller but more numerous and individuals acquire more possessions. In the rapidly industrializing countries, urban and industrial building and the construction of roads and other infrastructure are proceeding apace. For example in China, the production of aggregates and associated hidden flows more than doubled in seven years, rising from 2313 million metric tones (Mt) in 1989 to 5403 Mt in 1996

(Chen and Qiao, 2000). However per capita flows remain highest in the wealthier countries, the flows associated with aggregates being approximately 10 metric tonnes (t) per person per year in the US and approximately 4.2 t per person per year in China.

The key quantities to be addressed in establishing the materials flows due to mining and urbanization are the annual masses of rock and earth surface materials extracted, including overburden and mineral processing residues. In addition, account should be taken of pollutant releases to the atmosphere, water and the soil, as well as of the chemicals, fuels and materials in machinery used in the mining and mineral processing activities.

CALCULATING THE OVERBURDEN AND HIDDEN FLOWS

National mining and quarrying statistics record the net amounts of material produced for each commodity extracted, the run of mine production. However, these figures do not account for all the overburden removed before extraction can start and the waste created during mineral processing, the hidden flows associated with mining and quarrying (Adriaanse *et al.*, 1997; Douglas and Lawson, 1997). Whilst it can be argued that during open pit mining, especially coal mining, much of the overburden and locally stored waste is eventually replaced in the hole from which it came, even that local temporary shift of materials involves a large energy consumption and a change in the nature of the ground surface. In addition, there is always a risk that some of the stockpiled overburden will be eroded or lost through drains to nearby streams and that the final land restoration will result in a somewhat different landscape from that which existed before mining. The quantities of material shifted by mining and quarrying will vary with the mineral being extracted, the technology used to extract it, the geological situation of the deposit, the age and life of the mine, and the precise management policies of the mine operator. Nevertheless, this ecological rucksack of mining and extractive industries has to be estimated.

Researchers, including Sherlock (1922) and Hooke (1994), have tried to gain an idea of these *hidden* flows using estimated multipliers of the mineral production to obtain a figure for the total amount of materials shifted by mining and quarrying. For each mineral product, a given additional quantity of earth surface materials is removed as overburden or as waste from ore treatment. The ratio of this additional quantity to the amount of mineral produced can be used as a multiplier to obtain the total volume of material moved. To obtain the multipliers used in this article, case study examples of mine overburden quantities cited in the literature were examined for more general applicability; and approximate multipliers were established for as many mineral commodities as possible. These approximations (Table 1) were then extensively revised through consultation with a wide range of experts involved with the mineral industries (Douglas and Lawson, 1997). This provides a list of multipliers based on commodities that can be applied to world figures. Their application to individual countries may be less safe, particularly if a single mine, or group of mines, in an unusual geologic formation is responsible for virtually all the production of a particular mineral in that country.

The multiplier for coal plays a major role in the magnitude of the global figure of total material extracted. Many argue that because the excavated material in open pit mining is usually replaced in the same hole, it should not be included in the calculations of hidden flows. However, we assume no overburden replacement in this assessment of earth surface change even though good site restoration programs can, in time and in certain locations, successfully return the land to up to 80% of original productivity. Energy and machines have to be used to dig out and later replace the material. The nature of the replaced material is not exactly the same and a local change has occurred which contributes to the overall global change caused by extractive industries.

INFORMAL AND UNRECORDED MINING ACTIVITY

National mineral production statistics vary in accuracy, and there is probably a certain amount of unrecorded, and often illegal, removal of gravel and rock in every country. Some countries have a large amount of small-scale artisanal mining, some of which may be unregulated and unrecorded. Gold Fields Mineral Services of London s estimate of 175 t of gold being produced in 1975 by informal miners worldwide may be an underestimate. Three million carats of diamonds may be mined informally each year in Africa. Because of low recovery rates when informal artisanal gold and gem miners open up sites with very low yields, these operators probably shift more than twice the mass of material per net tone of pure mineral extracted than do their formal mining counterparts. The overburden: mineral ratio may be as high as 3.7 M to 1 for artisanal gold mining in the Amazon, as compared with a ratio of about 1 M to 1 for large mechanized gold mining. These operations alone could easily result in the movement of over 400 Mt of material per annum, equivalent to nearly 1% of the global mass moved during mineral extraction.

In many parts of the developing world, building materials are worked as small enterprises. People seek the clay, sand, gravel and stone that they need in the nearest most

convenient place and it is inconceivable that these activities are accurately recorded. In Vietnam, for example, at headlands along the coast, individual entrepreneurs quarry granite, while around towns and cities, small brick clay pits leave a series of derelict hollows and degraded soils. In inland China, villagers operate small crushers making aggregates from rock quarried from tiny rock outcrops near their fields. Around Beijing, many farmers have removed a depth of 1 2 m of clay from their fields to feed local brickworks and have then resumed cereal or vegetable cultivation on the lowered surface. Many hundreds of thousands of tones of material are worked in this way and in inefficient

Table 1 An example of the development of multipliers to calculate the hidden flows in mining activity

Commodity	Multipliers from the literature	Expert opinions	Recommended multiplier	Sources
Copper bearing ores: Cu content of copper ores and concentrates and all other copper bearing ores and concentrates intended for treatment of copper recovery	Copper ore = 1000% waste (Sherlock, 1922)	1% copper in ore and need to move 5 tons per 1 ton ore, i.e., ×500 (Highley)	×450	Highley = ×500
	1.23% = average of six mines (Mining Annual Review, 1996)	Bingham Canyon 0.7% copper in ore (Gilluly, *et al.*)		Accept Tilton and Highley average
	Bougainville quoted 1 ton waste to 1 ton ore (Winslow)	×337–×550 (Tilton in Eggert quoting US Bureau of Mines)		
	Freeport ore = 2.5% copper (Mitton)	0.49% average for US production in 1990–1994 (US Bureau of mines)		
	Copper content of ores hardly 1%, 165 material to one refined metal (Masini and Ayres)	0.77% concentration (Blackburn and Dennen)		
		×603 in US (Ayres and Ayres)		
Nickel bearing ores: (Ni content of nickel ores and concentrates intended for treatment for nickel recovery)	1.3% concentration (Blackburn and Dennen)	3% Ni content, and predominantly opencast and 1 : 15 ore/overburden (Manning)	×560	Basis 2% Ni content in ores and 75% opencast
	1.89% = average Ni content of eight mines (Mining Annual Review, 1996)			
Gold bearing ores: (Au content of gold ores and concentrates and all other gold bearing ores and concentrates intended for gold recovery)	8 grams per ton of ore (Blackburn and Dennen)	0.9–4.0 grams Gold per ton of ore, but 1 : 9.5 ore to waste, in the Andes (Fox)		
	Less than 1/3 oz per ton (Gilluly *et al.*)	5.4 grams per ton in 4 mines in Canada		

Table 1 (*continued*)

Commodity	Multipliers from the literature	Expert opinions	Recommended multiplier	Sources
	Bougainville 0.56 grams per ton of ore and ore = half of mined material (Gilles in Winslow)	0.99 grams per ton in 3 mines in Australia (Mining Annual Review, 1996)	×950 000	Average Australia, Canada and South America linked to production = 2.61 grams per ton and assumes ×2.5 overburden/ore
		×2 680 730 in US (Ayres and Ayres)		

Source: Douglas and Lawson (2000).

unauthorized mines. Road construction in remote areas, such as logging roads in Borneo, uses up to 370 t of gravel per kilometer (km), and may involve the unrecorded quarrying of over 4000 t of rock per year in each major Borneo logging concession (assuming some 10 km of road construction or repair each year).

The global amount of material moved by these informal, and largely unrecorded, mining activities is difficult to assess. Assuming that such mining involves 250 t per person per year and noting that in India alone about 200 000 informal miners may be shifting about 50 Mt of material each year, about 2.5% of all mining related materials displacement in India (Lawson and Douglas, 1997). Whilst these activities result in substantial amounts of earth movement, they probably add no more than 1 5% to the global mass of materials moved during the extraction of minerals.

THE RELATIVE MAGNITUDES OF NATURAL AND EXTRACTIVE INDUSTRY MATERIALS FLOWS

The masses involved in these deliberate shifts of material at the earth s surface are globally more than double the amount of material transfer to the oceans by rivers (Table 2). In North and Central America, Europe and Australasia, the mining of hard coal alone shifts more material than is carried by river erosion processes to the sea, including the products of soil erosion induced by human activity. Only in Asia, where the steeplands and intense human occupancy of the land cause extremely high rates of river erosion, does natural erosion exceed the movement of materials by the mining of just three commodities: coal, iron ore and copper ore.

The comparison suggests that deliberate human activity, in terms of excavation and filling of sectors of the earth s surface, is the most efficient geological agent on the earth at present. Natural processes of erosion vary greatly in space and time, with enormous local transformations when earthquakes or volcanic eruptions occur. Even so, most of the world s river basins are greatly affected by human activities, their sediment discharge including eroded soil from cultivated fields and waste discharges from homes and factories. Few river basins have no human interference. Natural rates of erosion range from less than $10\,t\,km^{-2}\,year^{-1}$ in lowland rivers on resistant rocks, to over $1000\,t\,km^{-2}\,year^{-1}$ on the slopes of volcanoes or in the steep, earthquake prone uplands of New Guinea. Erosion rates in areas affected by human activities may be one or two orders of magnitude greater.

Even if we take those accelerated rates of sediment discharge of major rivers, we find that individual rivers such as the Mekong, Ganges and Yangtze discharge a smaller mass of material annually than the global production of an individual mineral commodity in a single year and that the fluvial sediment discharges to the oceans from the continents are either similar to, or less than, the total movement of materials for minerals production on those continents. Thus, when we compare river loads with mineral production and overburden movement, we are examining the similarity of people accelerated natural process activity to deliberate shifting of material. The comparison brings out the magnitude of the materials fluxes involved.

River basin based analyses of natural and people driven materials flows offer an alternative to regional accounts. Investigations in the moist temperate environment of Northern Spain and in the humid tropical environment of Puerto Rico indicate the scale of materials shifted by extractive industries against the work of natural fluvial processes. The Spanish Besaya Basin has 13.9 million Mt per year of material excavated for mining, but the river only removes 0.07 Mt of sediment per year to the sea. For the Rio Loiza in Puerto Rico, 1.24 Mt of sand and gravel are dug out of the river channel for urban construction each year, while the

Table 2 Comparisons between denudation rates and mining activities, *Source:* Douglas and Lawson (2000)

Continent	Area (10^3 km^2)	Total load (solid and dissolved) (t km^{-2} year^{-1})	Natural erosion Mt year^{-1}	Coal, hard including 4.87 : 1 hidden flows 1995 (Mt year^{-1})	Coal, brown and Lignite including 9.9 : 1 hidden flows 1995 (Mt year^{-1})	Iron ores, including 5.2 : 1 hidden flows 1995 (Mt year^{-1})	Copper including 450 : 1 hidden flows (1995) (Mt year^{-1})
North and Central America	25 349	117	2966	4413	1139	348	1.314
South America	17 611	125	2201	180	1	712	1337
Europe[a]	10 948	92	967	3467	6771	587	529
Asia[a]	43 608	412	17 966	8990	1287	1097	679
Africa	30 335	59	1789	993	<1	156	286
Australasia	8923	30	267	944	505	457	297
Total	136 324		26 156	18 987	9703	3357	3357

[a] The United States Bureau of Mines published figures are for the whole of the Russian Federation. We have therefore split them 50% to Europe, 50% to Asia.

Table 3 Global mineral production and associated earth materials movement 1995

Commodity 1995	Production: net weight (million tons)[a]	% World	Multiplier	Production: gross weight (million tons)	% World	Remarks
Coal, Hard	3787	19.19	4.87	18 444	32.05	
Building Stones: granite, porphyry, sandstone including marble and travertines	10 430	52.85	1.36	14 186	24.65	Net production figures: by correlation with energy consumption
Aggregates: limestone flux and calcareous stone, gravel and crushed stone, sand, silica and quartz[c]						
Coal, Brown + lignite	930	4.71	9.9	9204	15.99	
Copper ores Cu content	9.3	0.05	450	4190	7.28	
Petroleum, crude	3065	15.53	1.02	3489	6.06	Gross weight includes 375 000 thousand tons oil shale and oil sand production wastes[b]
Iron ores Fe content	604	3.06	5.2	3138	5.45	
Gold ores Au content	0.002	0.00001	950 000	2138	3.71	Net production figures: Nötstaller, 1997
Phosphates, natural	119	0.60	4	477	0.83	
Nickel ores Ni content	0.72	0.003	560	403	0.70	
Bauxite, crude ore	101	0.51	3	302	0.52	
Clay	154	0.78	1.5	231	0.40	
Zinc ores Zn content	6.9	0.035	32	222	0.38	
Salt	17	0.84	1	166	0.29	
Gypsum, crude	99	0.50	1.2	119	0.21	
Kaolin	28	0.14	4	114	0.20	
Diamonds, Industrial and gem (221 738 thousand carats)	0.00005	0.00	2 380 000	109	0.19	
Lead ores Pb content	2.7	0.014	32	88	0.15	
Ilmenite concentrates	3.4	0.017	25	85	0.15	
Manganese ores Mn content	11	0.056	6	67	0.12	
Peat, for agriculture and fuel	42	0.21	1.25	52	0.09	
Iron pyrites, unroasted	7	0.04	5	36	0.06	
Gasoline, natural	36	0.18	1	35	0.06	
Uranium ores U content	32.97	0.0002	900	33	0.06	

(*continued overleaf*)

Table 3 (*continued*)

Commodity 1995	Production: net weight (million tons)[a]	% World	Multiplier	Production: gross weight (million tons)	% World	Remarks
Fullers Earth	8.0	0.04	4	32	0.06	
Bentonite	7.32	0.04	4	29	0.05	
Chalk	23	0.11	1.2	27	0.05	
Potash Salts K_{20} content	22	0.11	1	22	0.04	
Tin ores Sn content	0.18	0.001	100	18	0.03	
Andalusite	1.24	0.006	9	11	0.02	
Magnesite	9.5	0.05	1.2	11	0.02	
Abrasives, natural	8.7	0.04	1.2	10	0.02	
Fluorspar	4.0	0.02	2	8.0	0.01	
Barytes	4.2	0.02	2	8.5	0.01	
Slate	5.4	0.03	1.5	8.1	0.01	
Talc	6.5	0.03	1.2	7.8	0.01	
Chromium ores Cr content	3.3	0.017	2	6.8	0.01	
Tantalum and Niobium concentrates	0.05	0.0002	100	4.6	0.008	
Asbestos	2.5	0.01	1.5	3.7	0.007	
Borate minerals	3.4	0.02	1	3.4	0.006	
Natural gas	2.6	0.01	1	2.6	0.005	British Geological Survey World Mineral Statistics, conversion rate: 0.022 m^3 = 20 grams
Tungsten ores W content	0.024	0.0001	100	2.4	0.004	
Antimony ores Sb content	0.15	0.0007	9	1.3	0.002	
Graphite, natural	0.65	0.003	2	1.3	0.002	
Arsenic trioxide	0.045	0.0002	20	0.91	0.002	
Cobalt ores Co content	15.20	0.00008	20	0.30	0.0005	
Vanadium ores V content	0.032	0.0002		0.00	0.00	By product
Zirconium concentrates	0.94	0.005	100	0.00	0.00	By product
Silver ores Ag content	13.26	0.00007		0.00	0.00	By product
Sulfur	18	0.09		0.00	0.00	By product
Total	19 735	100		57 549	100	

[a] Source of production figures, net weight (unless otherwise stated): Department of Economic and Social Affairs, Statistical Division of the United Nations.

[b] Oil won from oil bearing shales and sands is included in the net figure for production of Petroleum, crude. However, oil from these sands and shale deposits requires additional materials movement of 22 times the net quantity of oil produced. Oil products produced from oil bearing shales and sands by the major producing countries amounted to 17 857 thousand tons in 1994 (United States Bureau of Mines, 1994).

[c] Aggregates and building stone production accounts for over 52% of the net weight of all mined or quarried material. Published UN data only cover 22 countries that produce about 50% the aggregates and building stones used. There are few figures covering the developing world. The only available global estimates of aggregates and building materials production have been made by correlating US production figures with gross domestic product (GDP) (Evans, 1993; Hooke, 1994). AS GDP includes exports, it is a variable predictor. Alternatively, an extrapolation could be based on population or energy consumption. As development involves building roads, housing and industrial facilities that use both aggregates and energy, data on energy consumption may be a good predictor of aggregate use. Absence of efficiency considerations in energy figures may be compensated for by the additional labor use. Available production figures for national, United Nations aggregate and building stone production correlate better with energy consumption than with GDP or population. Aggregates – energy: 80% of results range 9% $\pm$ the median. Aggregates – GDP: 80% of results range 43% $\pm$ the median. Aggregates – population: 80% of results range $>$100% $\pm$ the median.

sediment yield to the river is 2.80 Mt (Douglas and Lawson, 1997).

CONCLUSION

Globally, the total mineral production recorded in published data for 1995 was 19.7 billion tons (see Table 3), but the actual total amount of material mined or quarried from the earth s surface was of the order of 57.5 billion tons, i.e., some 37.8 billion tons of waste, overburden or spoil were also moved indicating that the hidden flows are about three times greater than the recorded ones (Douglas and Lawson, 1997).

REFERENCES

Adriaanse, A, Bringezu, S, Hammond, A, Moriguchi, Y, Rodenburg, E, and Schutz, H (1997) *Resource Flows: The Material Basis of Industrial Economies*, World Resources Institute, Washington, DC.

Chen, X and Qiao, L (2000) Material Flow Analysis of Chinese Economic-environmental System, *J. Nat. Resour.*, 15(1), 17 23.

Douglas, I and Lawson, N (1997) An Earth Science Approach to Material Flows Generated by Urbanization and Mining, in *Regional and National Material Flow Accounting: From Paradigm to Practice of Sustainability (Wuppertal Special 4)*, eds S Bringezu, M Fischer-Kowalski, R Kleijn, and Y Palme, Wuppertal Institute for Climate, Environment and Energy, Wuppertal.

Douglas, I and Lawson, N (2000) The Human Dimensions of Geomorphological Work in Britain, *J. Ind. Ecol.*, 4(2), 3 28.

Evans, A M (1993) *Ore Geology and Industrial Minerals: an Introduction*, Blackwell, Oxford.

Hooke, R Le B (1994) On the Efficacy of Humans as Geomorphic Agents, *GSA Today*, 4(9), 223 225.

Notstaller, R (1997) Socio-Economic Potential of Artisanal and Small-Scale Gold Mining. Paper presented at the *United Nations Industrial Development Organization Expert Group Meeting on UNIDO High Impact Programme " Introducing New Technologies for the Abatement of Gold Mercury Pollution Deriving from Artisanal Gold Mining"* , Vienna, 1 3 July.

Sherlock, R L (1922) *Man as a Geological Agent*, Witherby, London.

US Bureau of Mines (1994) *Minerals Yearbook 1994 - Volume 1 Metals and Minerals*, US Government Printing Office, Washington, DC.

FURTHER READING

Lawson, N and Douglas, I (2000) Urban Metabolism, Materials Flows and Sustainable Development: a Geoenvironmental Perspective, *Proceedings of the 5th International Symposium on Geoenvironmental Technology and Sustainable Development, August 9- 133, 1998*, CEEST University of Massachusetts, Lowell, 3 12.

Mediterranean: Land Degredation

see Land Degradation in the Mediterranean (**Volume 3**)

Methane: Industrial Sources

Lee Beck

United States Environmental Protection Agency, Research Triangle Park, NC, USA

Methane concentrations are increasing in the Earth's atmosphere. The exact mechanisms responsible are not known well enough to quantify the individual types of sources. Part of the increase is due to natural processes such as decay of organic material, and some of the change is caused by chemical reactions in the atmosphere. However, the greatest part of the increase in atmospheric concentrations is due to anthropogenic (human) activity. Part of this increase is the result of direct activity (e.g., fossil fuel combustion) and part is indirect activity (e.g., partial thawing of permafrost due to anthropogenic global warming). The industrial sources described in this article are part of the direct component.

One of the dif culties in quantifying methane emissions from industry is the inconsistent treatment of source characterization within the industrial sector. Industrial wastewater treatment lagoons and the petroleum industry are the most signi cant sources of methane emissions presented in this article. Some sources that could have been included (e.g., from the natural gas industry) are not treated here.

SCOPE OF MATERIAL PRESENTED IN THIS ARTICLE

Treatment of methane emissions from the industrial sector is not consistent across publications. Several authors include coal mining and/or natural gas production in the industrial sector (e.g., Sapoundjiev *et al.*, 1999: 805). In fact, many sources refer to natural gas production, transmission, and distribution systems as the natural gas industry. Electricity production, agricultural sources, and municipal solid waste (MSW) landfills are sometimes also included in the industrial sector. Sometimes sources are mixed even within the same report. For example, the *industry* module of the US Environmental Protection Agency s (EPA s) global model called the Atmospheric Stabilization Network includes only non-combustion industrial sources. Industrial combustion is accounted for in the *energy* sector (Lashof

and Tirpak, 1990: VI-15 VI-18). This is to allow for emissions accounting in the model by balancing energy supply with energy demand. When quantifying methane emissions elsewhere in the report however, the industrial sector includes combustion sources (boilers, heaters, lighting, etc.) (Lashof and Tirpak, 1990: IV-15). Methane from landfills is included in the industrial sector (Lashof and Tirpak, 1990: IV-19 and VI-18) as is agriculture (Lashof and Tirpak, 1990: V-48). Landfills and agriculture are treated separately in most other studies (e.g., Adler, 1994; Hogan, 1993).

Some studies separate stationary fossil fuel combustion from other industrial sources (Berdowski *et al.*, 1993: 133). The Intergovernmental Panel on Climate Change (IPCC) Working Group II publication considers energy production to be part of the industrial sector when characterizing the sector in the part of the report presenting mitigation assessment options (IPCC, 1996: 655), yet treats them separately in the section which assesses impacts and adaptation (IPCC, 1996: 367). As with Lashof and Tirpak (1990), this is probably due to the different objectives of the different sections of the report. However, these inconsistencies illustrate the difficulty in defining industrial sources.

There are also different methods for reporting individual sources within manufacturing operations, even when they clearly are part of the industrial sector. Some reports on greenhouse gas emissions combine all fossil fuel combustion into one category, which includes electricity generation as well as industrial boilers and other industrial fossil fueled processes. Other works group demand by common sources such as lighting and motors into their own separate categories that include both residential and industrial sources.

A primary rationale for deciding what to include in this article is to minimize redundancy. For this reason, coal mines, the natural gas industry, and MSW landfills are not covered here. (For methane from landfill, *see* Urban Wastes, Volume 3). Peat mining is included here since it is a methane source characterized in the referenced literature but is not included elsewhere in this volume.

Sources covered in this article are categorized as:

1. industrial combustion sources;
2. the petroleum industry (a significant methane emitter in the industrial sector);
3. coke production (also individually cited in the literature as a methane emission source);
4. peat mining;
5. wastewater treatment lagoons and industrial landfills;
6. miscellaneous industrial processes.

Some of the data presented are *emission factors* (amount of emission per source or activity). Some data are presented as total emissions, taking into account the number of sources. Further, some emission estimates are global totals and some are for the US only. This is because there frequently is insufficient information to extrapolate the US estimates to global totals. Because of the overlap or exclusion of sources presented in this article, compared to other presentations of methane and/or the industrial sector, it is imperative that the reader confirms the scope of material herein with the scope from other sources prior to any attempt to use the data to quantify anthropogenic methane emissions. Failure to do so may result in double counting of some sources and/or not counting of others.

INDUSTRIAL COMBUSTION SOURCES

This segment of the industry includes industrial boilers, process heaters, and other devices that burn fuel for heat. There are a number of ways to estimate methane emissions from these sources. A method used by the EPA is to apply an emission factor of 1.4 kg methane per $2.8 \times 10^4\,m^3$ of gas consumed to the annual consumption rate of about $2 \times 10^8\,m^3$ (Hogan, 1993: 7-7). This yields about 0.01 Tg annually for this source. That same study estimated methane emissions from wood-fired equipment used in the industrial sector as 0.2 kg of methane per kg of non-methane volatile organic compounds (NMVOCs). For the reported annual rate of 0.485 Tg of NMVOCs (Hogan, 1993: 7-7), this yields an annual methane emission rate of about 0.1 Tg methane from this source.

Energy usage is another way to estimate the greenhouse gas emission contribution of the sector. One major study estimates the following breakdown of energy usage for the US industrial sector: process heat applications, 40%; boilers, 30%; electrical requirements, 15%; and feedstocks, 15%. In that study, boilers are defined as steam producers, and process heat applications are defined as industrial non-steam producers such as ovens, furnaces, dryers, and kilns (Lashof and Tirpak, 1990: IV-15).

THE PETROLEUM INDUSTRY

The petroleum industry can be divided into five segments:

1. Production: this includes emissions at the oil well and at the pumps, tanks, and compressors associated with the well. Some equipment such as compressors used to transport natural gas to sales are not included here. They are considered part of the natural gas industry.
2. Transportation of crude oil: this segment includes all movement of crude oil from the production segment to the oil refinery. This includes truck, marine, rail, and pipeline transportation.
3. Refining: this segment includes all equipment that transforms the crude oil into fuels such as gasoline.

Crude storage, refinery operations and equipment, and product storage tanks are included here. It does not include chemical plant operations such as ethylene manufacture, plastic/rubber operations, and specialty products, even though these operations are sometimes integrated with the refinery operations. These latter operations are included under the chemical industry.

4. Product transport: this segment includes the truck, marine, rail, and pipeline transport of products and/or intermediate products from the refinery to the end-user.
5. End use: this segment includes the final distribution (e.g., gasoline filling stations) and use (e.g., in cars and trucks).

Information on methane emissions for this sector is available from only the first three categories listed above (Harrison *et al.*, 1999: 8). In other words, emissions are estimated for sources from the wellhead to the exit from the refinery.

Methane emissions from the petroleum industry can be characterized as fugitive, vented, and combusted. *Fugitive emissions* are unintentional leaks from equipment or storage vessels. These include pump packings, seals, pipelines, and leaks. *Vented emissions* are those typically released to the atmosphere as normal operating practice. Examples are dehydrator vents, tank vents, gas operated pneumatic devices, and maintenance practices such as blowdowns. *Combusted emissions* are unburned methane fuel from sources such as engines, burners, and flares.

Emissions from the US petroleum industry were estimated by determining how many emitting items there are, called the *activity factor*, and how much methane is emitted from each item, called an *emission factor*. The activity factor and emission factor are multiplied to determine the emissions.

Harrison *et al.* (1999) used 1993 as the base year for which the most recent complete data were available. For comparison purposes, the study also estimated emissions for the years 1986 1992. This was done by applying the 1993 activity data to the activity profiles for each of the prior years. Emission factors were assumed not to have changed over the study period. Only the 1993 data are presented here.

The production segment of the US petroleum industry emitted 1.7 Tg of methane in 1993. This represents about 90% of the emissions characterized by the study. Within the production segment, about 35% of the emissions were from oil tank venting, 17% from pneumatic devices, and 13% from compressor fugitives (Harrison *et al.*, 1999: 46).

Refineries emitted about 9% of emissions (0.18 Tg) and transportation the remainder (0.027 Tg). Of the total industry emissions, about 60% were vented, 30% were fugitive, and 10% were combustion related (Harrison *et al.*, 1999: 48 52).

MANUFACTURE OF COKE

Metallurgical coke is produced by heating coal to drive off the volatile material. The remaining residue, coke, is almost pure carbon. It is used primarily in the steel making process as a reducing agent and for fuel. The process takes place in coke oven batteries, which consist of a series of refractory lined ovens. In a typical battery, 7370 kg of coke and 1500 kg of coke oven gas are produced from 10 000 kg of coal (EPA, 1984: 3 8). The coke oven gas is cleaned at a by-product plant where the hot gas is quenched with water and the resulting tar separated. Usually this tar goes through additional fractionation and separation to produce products such as naphthalene and other heavy organics. About 40% of the cleaned gas from the by-product plant is returned to the underside of the ovens to provide the fuel for the coking process. The cleaned coke oven gas is approximately 50% hydrogen and 32% methane by volume, with carbon monoxide, carbon dioxide, nitrogen, and other trace gases making up the remainder.

Though most of the process takes place under negative pressure, there is some leakage of methane to the atmosphere. These leaks amount to about 3.2 tonnes of methane per million tonnes of coke produced (Adler, 1994, 7). Another author reports methane emissions of 0.319 Tg globally (Piccot *et al.*, 1996: 22 764).

PEAT MINING

World peat production was reported by Lacroix (1993) to be 23 million tonnes per year. About half of the production reportedly occurs in the former Soviet Union, with practically all of the additional production taking place in Finland and Ireland (Adler, 1994: 7 9). The mining operations produced emissions of about 2 ± 1 Tg of methane per year. However, the methane emission estimate is highly uncertain. Lacroix (1993) reports that it may be low, given that it does not include emissions from the total peat production cycle and does not include dewatering and briquetting, which likely produce additional methane (Lacroix, 1993, 530). One author reports concern over the accuracy of Lacroix s estimate (Piccot *et al.*, 1996: 22 762).

WASTEWATER TREATMENT LAGOONS AND INDUSTRIAL LANDFILLS

Emission from industrial wastewater treatment lagoons is highly uncertain. The source is generally associated with facilities which have a large organic content such as food processing industries. The methane is produced as a result of organic decomposition. The EPA reports that by extrapolating emissions from individual lagoons, global methane emissions from this source may be on the order of 20 25 Tg per year, and are expected to increase over the next decades

(EPA, 1990: A-8). Another EPA report estimates 25 Tg per year from this source (Hogan, 1993: 1 4). Most of the increase, as with most other sources of greenhouse gases, is expected to take place in developing countries. By contrast, emissions from domestic (municipal) wastewater treatment facilities are much lower than from industrial sources. Emissions from wastewater treatment facilities in the US are reported as 0.2 Tg each year from 1990 1997 (EPA, 1999: 7-4 and 7-5).

One study assumes that the methane emission from wastewater treatment lagoons is a function of the organic content of the wastewater, expressed as biochemical oxygen demand (BOD). This study provides global estimates of methane by multiplying the global wastewater outflow by wastewater BOD concentrations, using industry-specific concentration factors. This number was then multiplied by the fraction of anaerobically digested wastewater and a methane emission factor of 0.22 kg methane per kg BOD for the anaerobic fraction. References are given for the BOD estimates for various industrial waste waters. The global estimate of methane emissions for industrial wastewater treatment obtained using this method was 26 40 Tg per year (Adler, 1994: 10-3 10-11).

Much information has been published on methane emissions from MSW landfills; however, industrial landfills are not as well characterized. The EPA greenhouse gas inventory cites emissions from industrial landfills annually for the US from 1990 (0.7 Tg) to 1997 (0.9 Tg). The EPA study assumed that emissions from industrial sites are 7% of total emissions from MSW landfills. Methane emissions from MSW landfills totaled 10.6 Tg in 1990 and 12.6 Tg in 1997. 1.5 Tg of methane from all landfills was recovered and combusted in 1990 and 1.8 Tg was recovered/combusted in 1997. No information is given on how much of this recovery took place at the industrial sites (EPA, 1999: 7-2 and 7-3).

MISCELLANEOUS INDUSTRIAL PROCESSES

The scientific literature reports several other industrial sources as having an as yet unquantified potential for methane emissions. In some cases, it is because methane (usually from natural gas) is used as feedstock. Such is the case for the manufacture of hydrogen, ammonia, methanol, and various organic chemicals. In most of these processes, hydrogen is stripped from the methane molecule. The remaining carbon is oxidized to carbon monoxide or carbon dioxide. Production of methanol may increase as cleaner transportation options are pursued. Methanol can be used both as a direct fuel and as a source of hydrogen in fuel cells. In the US, nearly $340 \times 10^9\ m^3$ of hydrogen is used in ammonia production, $900 \times 10^9\ m^3$ by refineries, and $25 \times 10^9\ m^3$ for methanol production (IPCC, 1996: 660). Though emissions were not estimated, one publication reports methane emission factors, in kilograms per tonne, for the following industrial processes: 11 for carbon black, 4 for styrene, 2 for methanol, 1 for ethylene, and 0.4 for ethylene dichloride (EPA, 1999: 3 15).

Kraft paper manufacturing processes, printing and publishing operations, organic chemical manufacturing, and miscellaneous unnamed manufacturing sources are estimated to emit about 0.5 Tg of methane annually (Piccot *et al.*, 1996: 22 763).

The manufacture of silicon carbide, an abrasive, also has the potential to emit methane. In this case, quartz reacts with carbon from petroleum coke, and methane is produced. Emissions from this process for the US are estimated to be 0.001 Tg annually (EPA, 1999: 3 16).

There is also a potential for significant emissions from the non-fuel mining industry. Methane concentrations ranging over several orders of magnitude (0.0012 to $9.3\ cm^3\ kg^{-1}$) are present in various rock types and in mineral and metal ores (Lacroix, 1993: 530 531).

Table 1 summarizes emissions estimates presented in this article.

FUTURE PROSPECTS FOR INDUSTRIAL SECTOR METHANE EMISSIONS

IPCC (1996), 655 cites four driving forces that will affect future emissions of greenhouse gases from the industrial sector:

Table 1 Summary of Information on methane emissions from industry

Source	Scope	Annual emissions (Tg)	Reference
Industrial combustion	US	0.1	Hogan, 1993
Petroleum industry	US	1.7	Harrison *et al.*, 1999
Coke manufacturing	Global	0.3	Piccot *et al.*, 1996
Peat mining	Global	23	Lacroix, 1993[a]
Wastewater treatment	Global	25	Hogan, 1993
Wastewater treatment	US	0.2	EPA, 1999
Paper, chemicals and miscellaneous	Global	0.5	Piccot *et al.*, 1996
Silicon carbide manufacturing	US	0.001	EPA, 1999

[a] Unconfirmed, see text.

1. growth in industrial production (quantitative changes);
2. changes in the sectoral pattern of industrial growth (qualitative changes);
3. reduction in the energy used to produce specific products (energy efficiency improvements);
4. changes in energy carrier composition, which relates primarily to fuel switching (e.g., coal to natural gas, fossil to renewable sources).

The IPCC (1996) conclusions cited above address all greenhouse gases; however, the same drivers apply to methane emissions from the industrial sector. Certainly changes in production and/or the kinds of materials produced (items 1 and 2) will affect the quantity of methane emitted. Also changes in fuels used (item 4) will affect methane emissions (e.g., switching to natural gas will increase leakage to the atmosphere). Reduction in energy usage (item 3) will reduce methane emissions to a lesser extent than carbon dioxide emissions but will have the added benefit of fewer opportunities for natural gas leakage.

Most of the *per capita* quantitative increase in the industrial sector will occur in the developing countries. However because of the potential for qualitative changes in industrial development, the quantitative increases may not be proportionally as great as would otherwise be expected (IPCC, 1996: 657 658). This contention is sometimes referred to as *leapfrogging* (*see* Leapfrogging Technology, Volume 4) to indicate that new technology can learn from the past and not be subject to the same evolutionary trends (Lashof and Tirpak, 1990: V-50).

See also: Methane, Volume 1.

REFERENCES

Adler, M J (1994) *International Anthropogenic Methane Emissions: Estimates for 1990*, EPA-230/R-93-010, EPA (US Environmental Protection Agency) Office of Policy, Planning, and Evaluation, Washington, DC.

Berdowski, J, Olivier, J, and Velt, C (1993) Methane Emissions from Fuel Combustion and Industrial Processes, in *International IPCC Workshop: Methane and Nitrous Oxide - Methods in National Emissions Inventories and Options for Control*, ed A R van Amstel, RIVM Report No. 481507003, National Institute of Public Health and Environmental Protection, The Netherlands.

EPA (US Environmental Protection Agency) (1984) *Benzene Emissions from Coke By-product Recovery Plants - Background Information for Proposed Standards*, EPA-450/3-83-016a (NTIS PB84-209477), Office of Air Quality Planning and Standards, Research Triangle Park, NC.

EPA (US Environmental Protection Agency) (1990) *Methane Emissions and Opportunities for Control - Workshop Results of Intergovernmental Panel on Climate Change, Response Strategies Working Group*, EPA-400/9-90/007, Office of Air and Radiation, Washington, DC.

EPA (US Environmental Protection Agency) (1999) *Inventory of US Greenhouse Gas Emissions and Sinks: 1990- 1997*, EPA-236/R-99-003, Office of Policy, Planning, and Evaluation, Washington, DC.

Harrison, M R, Shires, T M, Baker, R A, and Loughran, C J (1999) *Methane Emissions from the US Petroleum Industry*, EPA (US Environmental Protection Agency) EPA-600/R-99-010 (NTIS PB99-128159), Air Pollution Prevention and Control Division, Research Triangle Park, NC.

Hogan, K B (1993) *Anthropogenic Methane Emissions in the United States: Estimates for 1990*, EPA-430/R-93-003, EPA (US Environmental Protection Agency) Office of Air and Radiation, Washington, DC.

IPCC (Intergovernmental Panel on Climate Change) (1996) *Climate Change 1995, Impacts, Adaptations and Mitigation of Climate Change: Scienti c- Technical Analyses*, Cambridge University Press, Cambridge.

Lacroix, A V (1993) Unaccounted for Sources of Fossil and Isotopically-enriched Methane and their Contribution to the Emissions Inventory: a Review and Synthesis, *Chemosphere*, 26(1 4), 507 557.

Lashof, D A and Tirpak, D A (1990) *Policy Options for Stabilizing Global Climate*, EPA 21P-2003.1, EPA (US Environmental Protection Agency) Office of Policy, Planning, and Evaluation, Washington, DC.

Piccot, S D, Beck, L, Srinivasan, S, and Kersteter, S L (1996) Global Methane Emissions from Minor Anthropogenic Sources and Biofuel Combustion in Residential Stoves, *J. Geophy. Res.*, 101(D17), 22 757 22 766.

Sapoundjiev, H, Trottier, R, and Aube, F (1999) Heat Recovery from Lean Industrial Methane Emissions Environmental and Economic Benefits of CFRR Technology, in *Greenhouse Gas Control Technologies*, eds B Eliasson, P W F Riemer, and A Wokaun, Pergamon.

Migrations: The Environmental Challenge of Population Movements

Ronald Skeldon

University of Sussex, Falmer, UK

The relationship between human migration and the environment is complex. Often, environmental change is seen as a root cause of population movement. Periods of desiccation in Central Asia have long been considered a factor in stimulating the great movements of nomad hordes that wrought so much devastation to the civilizations of Europe, the Middle East and China from the 12th to the 15th centuries, for example. Soil erosion, or the destruction of the

tropical forest, have been seen as factors creating more local movements of peoples in our own day. In very different contexts, the image of environments over owing with milk and honey, or land and gold in the Americas, to generations of Europeans and Chinese, attracted mass migrations to these open lands. However, the very movement of peoples, both pushed and pulled, causes environmental change through the spread of new technologies and new ideas, or simply through the introduction of increased numbers of people into environments unsuited for such an in ux. Yet, it is not a simple equation of population numbers and resources, but a complex matrix of institutional and political factors, in addition to the measures of population and environment.

It must also be made clear, however, that it is not only the great migrations that can represent a signi cant challenge to human and physical environments; small numbers of movers, too, can bring profound change. There can perhaps be no more dramatic example than the introduction of the human immunode ciency virus (HIV) and AIDS (acquired immune de ciency syndrome) into the United States, possibly by a single short-term visitor. The resulting rapid diffusion of the disease, the climate of fear among large sections of the population, and the eventual massive policy response changed people's behavior irrevocably in the West. In those areas in developing countries worst af icted, the disease is diffused rapidly along trucking routes, again by a minority of migrants, and the resulting high mortality has led to the virtual abandonment of lands and villages in parts of sub-Saharan Africa.

Throughout history, trading diasporas have linked small numbers of circular migrants and the exchange of goods, ideas, genes, and microbes that have generated profound change, both positive and negative, in the global environment. Four great human migrations of the past and present are identi ed for discussion: the migrations out of the Eurasian continent; movements from rural to urban areas; refugee and forced movements; and the new migrations of the late 20th century. This simple typology should not be seen as establishing four mutually exclusive forms of migration, but they form convenient divisions through which to examine the challenges for the environment posed by the movement of peoples.

THE MIGRATIONS OUT OF EURASIA

The Great Migration refers to the movement out of Europe, across the Atlantic, from the second half of the 19th to the early 20th century. Between 1860 and 1914, some 52 million people left Europe, mainly for the Americas but also for Australasia (Baines, 1991; Nugent, 1992). Many returned to their home countries, but this wave of migration was critical to the peopling of the North American continent, the transformation of once subsistence economies and the emergence of the United States as a major world power. That great migration, however, was but part of the much longer, and continuing, interchanges of population between Europe and Asia and the Americas. Beginning in a persistent form from the late 15th century, that Columbian exchange (Crosby, 1972) transformed the global environment through the introduction of new diseases, staple food crops, types of human settlement, and ways of doing things (*see* Biological Invasions, Volume 2).

The great human migrations in and to the Americas need to be placed within the overall context of the expansion of Europe that saw the ultimate creation of global empires. Smaller groups of European migrants traveled to literally every part of the world, not just as settlers but also as administrators and captains of commerce and industry. The resultant network of communities provided the human framework for the so-called global system in which large parts of the world are interrelated in a functional economic system. Small groups of migrants within colonial systems transformed rural landscapes through the expansion of commercial cropping. Perhaps the most dramatic example of this environmental transformation was the introduction of the plantation system, which brought a visual unity of coconut palm, sugar cane, or rubber trees to vast stretches of once heterogeneous forest. One account of the significance of the movement of plants, not just plantation crops, is given in Hobhouse (1985).

The conditions created by European expansion allowed for the vast movement of other peoples, some forced but others more voluntary. Movements of slaves out of Africa, later of indentured servants out of India and China, and of independent laborers seeking their fortunes, also from China, contributed to a transformation of the world as it was then known. Perhaps as many as 10 million slaves were landed alive in the Americas by the late 19th century, with an equal number having died between their place of capture and their destination (Lovejoy, 1989). Large parts of Africa suffered economic and demographic decline as a consequence, even if other parts of the continent undoubtedly profited as a result of the trade. By the early 20th century, some 1.5 million Indians had been moved outside the subcontinent as indentured laborers, and some three million Chinese were shipped out of China between 1801 and 1925 (Clarke *et al.*, 1990; Pan, 1998).

The global networks of European, Indian and Chinese communities that were established through these great migrations by the end of the period of European expansion laid the basis for the subsequent transformation of our environment today by human movements. An integral part of the demographic expansion out of Europe, however, was the establishment of urban centers for administration, trade, and the exploitation of resources. These laid the foundation for much of the second great human migration: the movement from rural to urban areas.

RURAL TO URBAN MIGRATIONS

Great cities long predate the colonial age in Africa, Asia, and South and Central America. The size, wealth and splendor of cities in China from the 12th century, and in India in the 16th and 17th centuries, overshadowed those of Europe before the industrial revolution. From the middle of the 18th century, however, Europe led the world in the growth and economic and political power of its cities. The overall level of urbanization in the developing world in 1800 was probably no more than about 8.3% compared with almost 12% in Europe (Bairoch, 1988). The change in the proportion of population that is urban by region from 1800 is shown in Figure 1. Of a total world population of around 978 million in 1800, about 50 million lived in towns. Since then, the world s population has risen more than six-fold to exceed six billion, while the numbers living in urban areas in the year 2000 have increased some 60-fold to almost 2.9 billion (United Nations, 1980, 1998b).

The explosive growth in the size of cities and in the level of urbanization occurred in Europe in the 19th century, but in developing countries that growth started essentially after the end of the Second World War. Currently, about 65 million people are being added to the world s urban population every year, compared with about 15 million being added to the rural sector. Over 60 million of the annual increase in urban population is being added to cities in countries of the developing world. Thus, the environmental challenge of population growth as we move into the 21st century is essentially one of the growth of the urban population in the developing world.

While migration is critical to the emergence of rapidly growing urban centers, mortality in early European cities was so high that the growth of these cities could only be sustained through the constant supply of migrants from the countryside, but migration is only one of three components of urban growth. The natural increase of the urban population itself is often the major component and currently accounts for about 58% of global urban growth (McGee and Griffiths, 1998). The reclassification of formerly rural areas as urban is also an important contributor to the total increase in urban population in any area, but one notoriously difficult to measure with any accuracy. The movement of people from village to town was, however, fundamental to the shift in the relative balance from a rural to an urban society. The rural to urban migration brought about the concentration of population at specific locations and the eventual depopulation of other regions.

The migration from village to town should not be thought of in terms of a simple, unilinear movement of people abandoning their rural homes for a life in the city. Rural to urban migration is a complex system that involves much return movement back to the villages and human circulation between town and country that binds the two into an integrated economic system (Skeldon, 1990). Wealth flowing in cash and kind back to the villages has been

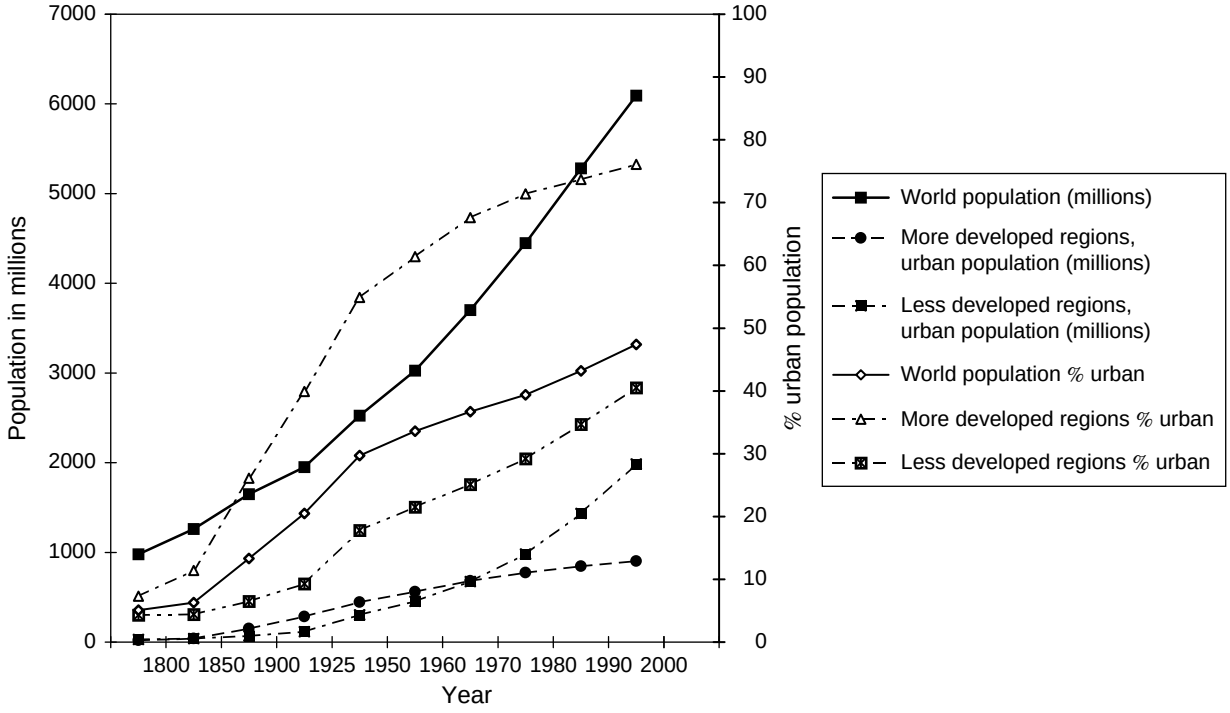


Figure 1 Growth of the urban population of the world by major region, 1800–2000. (Data from United Nations, 1980 and 1998b)

Table 1 The 15 largest urban agglomerations in the year 2000, ranked by projected population size in millions

Tokyo	28.0
Mexico City	18.1
Bombay	18.0
Sao Paulo	17.7
New York	16.6
Shanghai	14.2
Lagos	13.5
Los Angeles	13.1
Calcutta	12.9
Buenos Aires	12.4
Seoul	12.2
Beijing	12.0
Karachi	11.8
Delhi	11.7
Dhaka	11.0

Source: United Nations (1998b).

shown to be a support to rural life, initially at least, in both developed and developing worlds during the transition to more urban societies.

An integral part of this process has been the emergence of very large urban areas, or mega-cities, which are often the capital and single largest city in a country or large region. While most of these used to be in the countries of the present developed world, the majority now are to be found in poorer parts of the world (Table 1). These cities have posed, and continue to pose, some of the major challenges facing our environment. The problems of supplying adequate housing, transportation, health services and sewage disposal, as well as providing the means to earn a basic living for densely packed and rapidly growing populations, have taxed the minds of city administrators in all these huge metropolises. Between 70 and 95% of all new housing in the cities of the developing world is estimated to be constructed illegally, and the average number of persons per room ranges between three and five times that of the developed world; up to 66% of urban populations in the developing world have no adequate systems of sanitation and several hundred million have no access to clean drinking water (Satterthwaite, 1998) (*see* Urban Population Change, Volume 3).

Pollution of air and water and the rapid spread of disease among the squalid conditions of the urban poor are thus as characteristic of the Calcutta, Lagos or Rio de Janeiro of today as they were of Cobbett s early 19th-century London the great wen ... The monster called the metropolis of empire . The teeming masses of the underemployed and unemployed have provided crowds to threaten and overthrow governments and the conditions of the urban poor have generated such a multitude of informal activities for survival that the criminal activities among them are seen as a threat to social order and decency. Yet, despite all the problems, the great cities have survived and prospered as one of the most enduring symbols of our time.

It would be unfair to attribute all these ills either to migration or to migrants. In the developing world of today, the city-born are usually found to have higher levels of unemployment than migrants from the rural sector, and the city-born, too, are often shown to have greater involvement in criminal activities. Yet, migration brings different groups together, facilitating the emergence of new identities while at the same time creating tensions over the perceived fear of the outsider. The resulting ferment has seen many of the environmental challenges met through the concentration of minds and wealth in the great centers of learning, commerce, and industry that are the mega-cities. While many challenges remain, particularly in the cities in the developing world, it is no exaggeration to say that the migration to the cities has provided the human foundation for the transformation of our environment to an urban world.

FORCED AND REFUGEE MIGRATIONS

Warfare, the destruction of crops and property, and the capture of slaves have been integral parts of the history of civilizations everywhere, so too has flight from these depredations been a universal response. More recently, however, the emergence of the nation state as the basic political unit of the modern world institutionalized flight on an unprecedented scale. With the nation state s emphasis on common identity, any group deemed to be outsiders is at risk of persecution and expulsion. State conflicts, the emergence of new nations, and the ethnic or ideological cleansing of the 20th century have generated another of the great human migrations: the flows of refugees. Some 30 million people were displaced in Europe alone by the Second World War (Moch, 1992), between 15 and 17 million people moved upon the partition of India and Pakistan in 1947, and some 10 million people fled to India in 1971 from what is now Bangladesh to escape the civil war in that area which had been aggravated by cyclone damage and famine conditions.

Although the world has ostensibly been at peace since the early 1950s and has subsequently experienced a period of continuous and rapid economic growth, the number of refugees rose substantially from under two million in 1965 to 13.2 million in 1996. The number of people considered to be of concern to the United Nations High Commissioner for Refugees in 1996 was double this number, or 26 million, which included the 13.2 million refugees, plus 3.3 million prior refugees who had since returned home, 4.7 million internally displaced, and another 4.9 million in danger of being displaced by conflict (United Nations, 1998a). Most of these people are to be found in the poorest countries of the world in sub-Saharan Africa and western South Asia,

which often incorporate marginal environments, with environmental concerns looming large as both cause and consequence of the displacements (*see* Refugees and Human Conflict, Volume 3).

It is perhaps in the area of forced migrations that the link between the environment and population movement is most explicit in the literature (*see* Environmental Refugees, Volume 4). Natural disasters caused by typhoons and hurricanes, floods, volcanic eruptions, and earthquakes illustrate the challenges posed to populations by the environment, although the effects are more likely to be seen in increased mortality than in migration, with population displacement, while intense, being localized and short-term. Famine, the disaster most capable of displacing large numbers of people, is likely to be more the result of human mismanagement and political ambition resulting in conflict than of natural factors. Famine and its accomplices, disease and war, make up a triple affliction that sees migration as a basic survival strategy resulting in the movement of hundreds of thousands of people and, as indicated above, was as common in the history of Europe as it is in much of sub-Saharan Africa today. The environmental refugee, or the person apparently forced to move for environmental reasons, is likely to have moved for a combination of reasons (Black, 1998).

The forced migration of such large numbers moving in so short a period of time is also seen to create intense pressure on the environments to which the people move. Those fleeing toward towns for protection are faced with finding adequate food in non-agricultural economies, making worse the already over-crowded and unsanitary conditions in the urban centers. Those displaced to rural areas are seen to exceed the carrying capacity of the often marginal environments to which they are moved, leading to the destruction of local resources, specifically forests, land, and groundwater. As in the case of those who fled to urban areas, the closely packed populations create severe problems for the adequate disposal of human waste and provide ideal conditions for the spread of disease. While such situations do exist and severe environmental destruction can result from the forced movement of populations for whatever reason, dispassionate analysis suggests that these conditions are neither characteristic of all refugee settlement areas nor necessarily inevitable (Black, 1998).

Displaced migrants have the capacity to help themselves and it is incumbent upon relief agencies to provide the conditions that can best allow the displaced to achieve these objectives. Such a strategy is similar to that proposed to provide the urban poor, many of whom are rural-to-urban migrants, with another path to development (De Soto, 1989), the means to help themselves. Given the large numbers of forced migrants and the fact that they are not an exceptional type of population movement in our modern world, it is perhaps best to consider them as an integral part of the total pattern of migration. As such, their challenges for the environment can be seen within a more general development framework. The issues then blend with those that have come to dominate concerns over international population movements since about the mid-1960s, or the new global migrations.

More long-term environmental changes leading to migration, such as desertification and future rises in sea level as a result of global warming, have the potential to displace very large numbers indeed. For example, in 1995, the Intergovernmental Panel on Climate Change projected a rise in sea level of 20 cm by 2020 and 30 100 cm by 2099. A rise of this magnitude would inundate almost one fifth of the present land area of Bangladesh within a century, directly affecting some 70 million people. Some small island countries would virtually vanish. The disappearance of densely settled, low-lying cultivable land, however, would merely be one of the more obvious direct effects of such a rise in sea level. The encroaching sea would also contaminate lenses of fresh water and gradually intensify the effects of tropical storms in populated areas.

The impact of global warming on a rise in sea level is still a matter of considerable debate and it is difficult to convince governments to plan for long-term shifts that may appear to be hypothetical when more immediate issues are at hand. However, the mere possibility that such an event might occur can undermine confidence, discourage investment and lead to long-term movements away from small island countries. For example, the prospects are dire indeed for large, densely populated areas such as those around the Bay of Bengal or the Gulf of Siam, where there is neither the tradition of building defenses against the sea (as in the Netherlands, for example) nor the capital available to do so.

THE NEW MIGRATIONS

Our present age has been called the age of migration (Castles and Miller, 1993). Although the absolute number of international migrants, as measured by migrant stock in each country, increased from 75 million persons in 1965 to 120 million in 1990, the annual rate of increase was only slightly higher than for the population as a whole and the proportion of people remaining outside their countries of origin is small, fluctuating between 2.1 and 2.3% over this entire period (Zlotnik, 1998). The reasons for designating our era as the age of migration lie elsewhere. The direction and composition of international population flows changed markedly over the period, with the movement to the settler societies of North America and Australasia being dominated by migrants from non-European sources, and Europe itself emerging as a significant destination for migration from developing countries.

These changes in the direction and composition of migration were associated with the transition to persistent low

fertility among European peoples that brought about a slowing in the growth of the labor force and an ageing of populations. With the post-war economic boom in progress, immigrants, initially from southern Europe and later from further afield, met the demand for labor. The societies of the developed worlds of Europe, North America and Australasia thus experienced a major transformation through ethnic mixing on a scale unprecedented in recent history. Such mixing had characterized societies at various stages prior to the emergence of the ideal of the nation state, and the late 20th century has seen a return to a polyethnic norm (McNeill, 1986).

The creation of multicultural states presents one of the greatest challenges to our modern world and one in which the environment looms large. The bringing together of peoples with very different attitudes to the environment and to its use can provide a basis for conflict. Different consumption patterns, attitudes towards trees, the interpretation of local building and zonation codes, or the disposal of garbage can create tensions between long-standing residents and new immigrants. Migrants from highly urbanized and affluent Hong Kong, for example, have difficulty understanding local Canadian sensitivities towards their approach to house styles and neighborhood trees that in Hong Kong is normal but in Vancouver is considered crass and uncaring. That many of the landscapes being altered by migrants are more romanticized than real, and have themselves been imposed by previous generations of European migrants on native populations, simply illustrates the complex relationships between the environment, the relative wealth of migrants, and attitudes that can be seen as racist (Ley, 1998).

Scapegoating and resentment toward wealthy immigrants on the part of the local population at the local level can find broader national expression, as they have in Australia where environmental concerns have been used as a justification to limit immigration altogether. Increased consumption of scarce resources raises Malthusian fears of the limits to growth and a limited carrying capacity (*see* Carrying Capacity, Volume 4) that, if exceeded by a future generation of migrants, will threaten the lifestyles of existing citizens (Hugo, 1996). In Australia, the fear of the culturally different outsider bringing environmental destruction has been a vocal, if as yet largely ineffectual, dimension of the immigration debate. While such local and national attitudes ignore the very real positive contribution that migrants make to their host economies and societies, they clearly reflect the difficulties of creating a single nation out of many peoples living in one environment.

A significant component of the modern migration flows is the short-term movements of tourists. The annual number of tourists increased from 69 million in 1960 to 454 million in 1990. The rising affluence of the developed countries and the advent of rapid and cheap air transportation facilitate movement for recreation to all parts of the world, including Antarctica. While almost three-quarters of the number of tourists visit developed countries, increasing numbers are going to remote and often fragile environments in search of the exotic. High mountain and island ecosystems are particularly at risk from any unregulated influx of visitors and an enduring challenge remains to ensure adequate plans for the conservation of heavily used sites and the preservation of other wilderness areas (*see* Tourism and Ecosystems, Volume 3).

Human migration is, along with fertility and mortality, one of the basic components of demographic change and, as such, must be included as a critical factor in the equation relating population size and resource use. With the decline of fertility, migration has emerged as *the* variable in causing populations to increase or to decline in some areas. For example, international migration accounted for almost 88% of the population growth of Europe between 1990 and 1995 but only 33% for North America (United Nations, 1998a). More important than its impact on growth, however, is its effect on the distribution of populations, concentrating people in urban areas and leading to the depopulation of rural areas, and its effect on the composition of populations, spreading ideas and patterns of consumption around the world and bringing very different peoples together. Human migration has been, and remains, one of the major forces behind environmental change and is an integral part of overall development (Skeldon, 1997).

REFERENCES

Baines, D (1991) *Emigration from Europe 1815- 1930*, Macmillan, London.

Bairoch, P (1988) *Cities and Economic Development: From the Dawn of History to the Present*, University of Chicago Press, Chicago, IL.

Black, R (1998) *Refugees, Environment and Development*, Longman, Harlow.

Castles, S and Miller, M J (1993) *The Age of Migration: International Population Movements in the Modern World*, Macmillan, London.

Clarke, C, Peach, C, and Vertovec, S (1990) *South Asians Overseas: Migration and Ethnicity*, Cambridge University Press, Cambridge.

Crosby, A W (1972) *The Columbian Exchange: Biological and Cultural Consequences of 1492*, Greenwood, Westport, CT.

De Soto, H (1989) *The Other Path: The Invisible Revolution in the Third World*, Harper and Row, New York.

Hobhouse, H (1985) *Seeds of Change: Five Plants that Changed the World*, Sidgwick and Jackson, London.

Hugo, G (1996) Environmental Concerns and International Migration, *Int. Migr. Rev.*, 30, 105 131.

Ley, D (1998) The Rhetoric of Racism and the Politics of Explanation in the Vancouver Housing Market, in *The Silent Debate: Asian Immigration and Racism in Canada*, eds E Laquian,

A Laquian, and T McGee, Institute of Asian Research, University of British Columbia, Vancouver, 331 348.

Lovejoy, P E (1989) The Impact of the African Slave Trade on Africa: A Review of the Literature, *J. Afr. Hist.*, 30, 365 394.

McGee, T G and Griffiths, C J (1998) Global Urbanization: Towards the Twenty-First Century, in *Population Distribution and Migration*, United Nations, New York, ST/ESA/SER.R/133, 49 65.

McNeill, W H (1986) *Polyethnicity and National Unity in World History*, University of Toronto Press, Toronto.

Moch, L P (1992) *Moving Europeans: Migration in Western Europe since 1650*, Indiana University Press, Bloomington, IN.

Nugent, W (1992) *Crossings: The Great Transatlantic Migrations, 1870- 1914*, Indiana University Press, Bloomington, IN.

Pan, L (1998) *The Encyclopedia of the Chinese Overseas*, Archipelago Press, Singapore.

Satterthwaite, D (1998) Health and Environmental Problems in the Cities of the Developing Countries, in *Population Distribution and Migration*, United Nations, New York, ST/ESA/SER.R/133, 150 177.

Skeldon, R (1990) *Population Mobility in Developing Countries: a Reinterpretation*, Belhaven, London.

Skeldon, R (1997) *Migration and Development: a Global Perspective*, Longman, Harlow.

United Nations (1980) *Patterns of Urban and Rural Population Growth*, United Nations, New York, ST/ESA/SER.A/68.

United Nations (1998a) *World Population Monitoring 1997: International Migration and Development*, United Nations, New York, ST/ESA/SER.A/169.

United Nations (1998b) *World Urbanization Prospects: The 1996 Revision*, United Nations, New York, ST/ESA/SER.A/170.

Zlotnik, H (1998) International Migration 1965 96: An Overview, *Popul. Dev. Rev.*, 24, 429 468.

Minamata Mercury Disaster

see Mercury in the Environment (**Volume 2**)

N

National Responsibilities for Greenhouse Gas Emissions

Ian Douglas
University of Manchester, Manchester, UK

When considering the human drivers of global environmental change and discussing what can be done to reduce the pace of such change, responsibility for greenhouse gas emission reduction always comes to the fore. One measure is to simply look at national tonnages of gas emitted and then to call for a percentage reduction in that tonnage. However, more equitable measures look in each country at the amount of gas emitted per person, per unit of gross domestic product (GDP) and per unit land area and the deviation of those individual country gures from the global averages per person, per unit GDP and per km^2. Reliable data are not available on a standard basis for all gases, but those for fossil fuel derived carbon dioxide are considered reliable. For some countries, data for the three main greenhouse gases (carbon dioxide, methane and nitrous oxide) are now available.

Many factors affect differences in the emissions per person in a country, for example, the types of fuel used for heating or cooling, cooking or to generate electricity, population density, degree of industrialization, availability of private cars and vehicle kilometers traveled (people tend to drive longer distances in sparsely populated areas), and whether fossil fuels are extracted or processed within the state.

Differences in emissions per unit GDP (here expressed as per $1000 GDP) re ect contrasts in economic conditions between countries, the type of economic activity, and the ef ciency with which energy is used throughout the country. Differences in emissions per km^2 re ect population density, land use and the intensity and type of industrial activity. Some countries have vast areas of thinly populated, inhospitable land, while others are highly urbanized and have intensive agriculture.

ADEQUACY OF DATA FOR EVALUATING FAIRNESS AND EQUITY OF RESPONSIBILITY FOR EMISSIONS

Notions of fairness and justice underpin international negotiations to reduce greenhouse gas emissions. The United Nations Framework Convention on Climate Change (UNFCCC) (*see* United Nations Framework Convention on Climate Change and Kyoto Protocol, Volume 4) contains concepts of burden sharing and common but differentiated responsibilities (*see* Common but Differentiated Responsibility Principle (Stockholm/Rio), Volume 4) that are based on the idea that nations that have contributed most to the problem of climate change should do most to solve it. This is inherently the polluter pays principle (*see* PPP (Polluter Pay Principle), Volume 4). Application of this notion suggests that national targets for the reduction of greenhouse gas emissions should be based on the historical contribution of each nation to global emissions. The most important factor in determining this contribution is the level of emissions *per capita*.

Although investigations of burden sharing or differentiation usually make emissions *per capita* as the foremost criterion on which emission reduction targets should be based. However, other criteria to be considered include: the ability to pay (usually measured by GDP *per capita*), emissions per unit GDP, dependence on fossil fuels and emissions per unit land area. Clearly, emissions may be high in particular countries because they are burning fossil fuels to manufacture commodities widely used elsewhere in the world, such as the petrochemical industries of the Netherlands, Venezuela and the US.

Due to measurement difficulties and lack of data, the figures for emissions *per capita* usually used refer to carbon dioxide emissions from fossil fuel burning, cement production and gas flaring only. Such figures suggest that the US has the world s highest emissions *per capita*. However, the provisions of the UNFCCC require the industrialized countries that have signed up to emissions reduction targets to compile and submit to the United Nations systematic and comprehensive inventories of emissions from all sources and sinks. This additional information

includes the impact of agriculture, grazing, wetland use and forest management, some countries removing greenhouse gases from the atmosphere through these activities, others releasing them. Some of these data now permit more comprehensive evaluations of greenhouse gas emissions for some countries. Nevertheless, the only analysis for all countries deals solely with carbon dioxide emission from fossil fuel burning, cement production, and gas flaring.

FOSSIL FUEL BURNING, CEMENT PRODUCTION, AND GAS FLARING CARBON DIOXIDE EMISSIONS

The Largest National Gas Emitters by Total Tonnage

The US continues to be the largest single source of fossil fuel related carbon dioxide emissions reaching an all time high of 1447 million metric tons of carbon in 1996. In fact, US emissions are approximately 58% higher than those of the world s second largest emitter, the People s Republic of China. Over the past 20 years, China s fossil fuel carbon dioxide emissions grew at an annual rate of 5.3% reaching almost 918 million metric tons of carbon in 1996. From 1990 1996 alone, emissions from fossil fuel consumption and cement production rose 40%. Growth has occurred largely in the use of coal, not surprising given that China is the world s largest coal producer. In 1950 coal use contributed 99% of the total Chinese energy-driven carbon dioxide emissions, but only 77% in 1996. Liquid fuels now contribute 14% of emissions and have grown appreciably over the past decade. Since 1992, total carbon dioxide emissions from Russia have dropped 19% to 431 million metric tons of carbon, still the third largest emitting country in the world and the largest emitter of the republics comprising the former USSR. After 1987, Japanese carbon dioxide emissions grew by 30% to almost 319 million metric tons of carbon in 1996. Recent growth reflects a return to late 1970s consumption levels for liquid petroleum products and continued contributions from coal and natural gas. Japan is the world s largest importer of coal (126 000 metric tons) and liquefied petroleum gas (15 000 metric tons), and second largest importer of crude oil. India has experienced dramatic growth in fossil fuel carbon dioxide emissions, which have increased 6.1% per year since 1950, as India has climbed to the world s fifth largest fossil fuel carbon dioxide emitting country. Since 1990 alone, total emissions have risen 47.7% to 272 million metric tons of carbon in 1996. Fossil fuel carbon dioxide emissions from the UK in 1996, 152 million metric tons of carbon, are comparable to emission levels of the 1950s. Since 1950, the UK has dropped from third to seventh in rank among fossil fuel carbon dioxide emitting nations.

The Highest Per Capita Emitters

Several small states with major refineries and petrochemical operations, such as the US Virgin Islands with 33.22 tonnes *per capita* ($t\,cap^{-1}$), Qatar 18.19 $t\,cap^{-1}$, United Arab Emirates 9.40 $t\,cap^{-1}$, and Netherlands Antilles 8.75 $t\,cap^{-1}$, head the ranked list of countries by *per capita* emissions. Of the countries with the 30 largest total emissions (Table 1) the US heads the list at 5.48 $t\,cap^{-1}$. *Per capita* values for the US are the highest of the industrialized world but remain well below the 1973 high, roughly equivalent to late 1960s levels, and only 22% higher than the 1950 level. In 1996, 40% of US fossil fuel emissions came from the consumption of petroleum products.

Russia has a 1996 *per capita* emission rate of nearly three metric tons of carbon per person that is almost three times the global average and represents the third highest rate of the region behind the Czech Republic (3.37) and Estonia (3.04). Japan s *per capita* fossil fuel carbon dioxide emission rate has risen in 1996 to an all time high of 2.54 metric tons of carbon. China s population has doubled over the past forty years and now exceeds 1.25 billion people. *Per capita* emissions increased by a factor of almost eight over this period but their 1996 rate of 0.7 t carbon *per capita* is well below the global average. With the world s second largest population of nearly one billion people, India s *per capita* emission rate for 1996, of 0.29 t carbon cap^{-1} is well below the global average (1.13). It is the smallest *per capita* rate of any country with fossil fuel carbon dioxide emissions exceeding 25 million metric tons of carbon.

This contrast in per capita emissions between the world s two largest countries in population terms, China and India (0.76 and 0.29 $t\,cap^{-1}$, respectively) and the two biggest emitters, the US and Russia (5.48 and 2.65 $t\,cap^{-1}$, respectively) illustrates the great gulf in responsibility for release of carbon dioxide to the atmosphere between the people of relatively poor countries and those that are wealthy or have a highly developed industrial base (Figure 1).

The Lowest Per Capita Emitters

Forty out of the 210 countries in the database emit less than 0.10 $t\,cap^{-1}$. The majority of them are in Africa, but they include countries with large cities and substantial populations such as Kenya, Ethiopia and Mozambique. Pakistan which emits 0.18 $t\,cap^{-1}$ and ranks 158th in *per capita* terms, is ranked 37th in terms of gross national emissions. Indonesia is 22nd in national tonnage of emissions but 138th in *per capita* terms while India is 5th and 142nd, respectively (Table 1). The two sets of figures tell different stories.

Table 1 Carbon dioxide emissions from fossil fuel burning, cement production, and gas flaring in 25 of the top 30 total emitting countries in 1997 expressed as tonnes of carbon

Country	Tonnes per capita	Tonnes per 10 km^2	Tonnes per $1000 GDP	Total (tonnes)
US	5.48	1.62641	0.23	1 489 648
China (PRC)	0.75	0.983498	2.15	913 768
Russian Fed.	2.65	0.231291	1.19	390 616
Japan	2.51	8.397004	0.08	316 164
India	0.29	0.94141	1.12	279 899
Germany	2.77	6.509692	0.12	227 364
UK	2.41	5.881457	0.15	142 096
Canada	4.42	0.145202	0.24	133 890
South Korea	2.55	11.82022	0.35	116 701
Italy	1.94	3.785724	0.11	111 323
Ukraine	1.97	1.733443	0.92	100 427
Mexico	1.06	0.523731	0.29	99 964
Poland	2.47	3.134255	1.11	95 413
France	1.59	1.688384	0.074	92 878
South Africa	2.23	0.708675	0.74	86 532
Australia	4.71	0.108159	0.3	86 336
Spain	1.68	1.333173	0.14	66 584
Indonesia	0.32	0.359373	0.45	65 103
Thailand	0.95	1.115543	0.46	56 992
Turkey	0.85	0.702182	0.31	54 042
Venezuela	2.25	0.579831	0.85	51 144
Netherlands	2.83	13.04717	0.14	44 256
Argentina	1.05	0.137498	0.15	37 629
Malaysia	1.7	1.086897	0.55	35 710
Czech Republic	3.25	4.334239	1.06	33 495

Sources: emissions total and per capita from Marland *et al.*, 2000; land area and GDP from http://www.wri.org.

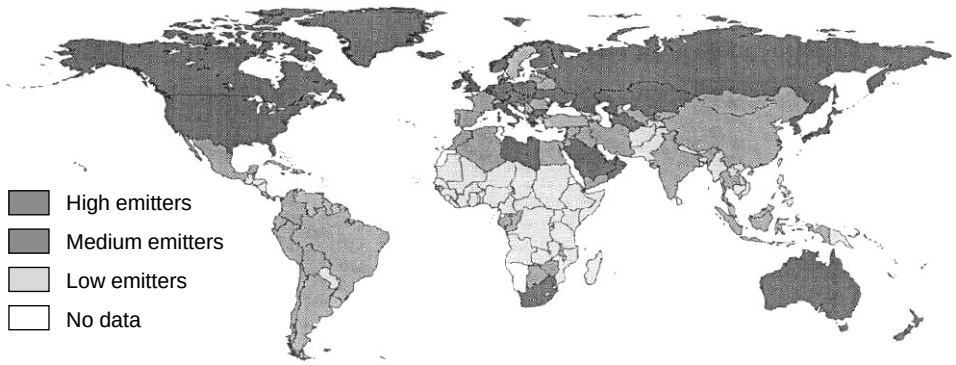


Figure 1 Per capita emissions of carbon dioxide by country 1996. The map shows the countries in which the 20% highest emitters live (high emitters) and the 20% lowest emitters (low emitters)

The Highest Per Unit GDP Emitters

Among the 30 largest emitters, China is by far the biggest emitter of carbon dioxide per unit GDP at 2.15 t per $1000 (Table 1), reflecting a great reliance on coal for transportation, industry and domestic cooking and heating. Other countries with heavy industrial structures, such as Poland, the Czech Republic and the Russian Federation follow China. India also has heavy industry, but a relatively small GDP in relation to its total population and thus has a similar emissions per unit GDP to the Central and Eastern European countries at 1.12 t per $1000 GDP.

The Lowest Per Unit GDP Emitters

Japan and France (0.08 and 0.074 t per $1000, respectively) have low carbon dioxide emissions per unit GDP

because they have considerable reliance on hydropower and nuclear power for energy. Many Western European nations, such as the UK, Italy and Germany have emissions of 0.11 0.14 t per $1000, while Canada, Australia and the US are in the range 0.23 to 0.30 per $1000.

The Highest Emitters Per km^2 Land Area

With emissions of 13.0 t km^{-2} the Netherlands has the highest rate among the 30 countries with the largest emissions, followed by South Korea with 11.8 t km^{-2}. These figures may reflect the high population densities and the concentrations of petrochemical and steel making industries in those countries. Japan, Germany and the UK are the next three highest countries with emissions of 8.38, 6.51 and 5.88 t km^{-2}, respectively.

The Lowest Emitters Per km^2 Land Area

Large countries with arid, cold or mountainous areas and relatively small populations are the lowest emitters per unit area, Australia emitting 0.11 t carbon dioxide km^{-2}, Argentina, 0.13 and Canada 0.14. While the Russian Federation emits 0.23 t km^{-2}, China emits 0.98 and the US 1.62.

CALCULATIONS OF TOTAL GREENHOUSE GAS EMISSIONS PER CAPITA

New calculations of greenhouse gas emissions per person for all 35 Annex B parties to the Kyoto Protocol (the industrialised nations that have signed up to emissions reduction targets) cover emissions of the three main greenhouse gases (carbon dioxide, methane and nitrous oxide) in 1995 from all sources and all sinks measured in carbon dioxide equivalents (CO_2-e) (Turton and Hamilton, 1999). On this basis, Australia has the world s highest greenhouse gas emissions per person at 26.7 tonnes; this is twice the average level for all other wealthy countries (13.4 t cap^{-1}) and 25% higher than emissions per person in the US (21.2 t cap^{-1}). While the US has higher emissions from energy (20.6 t cap^{-1}

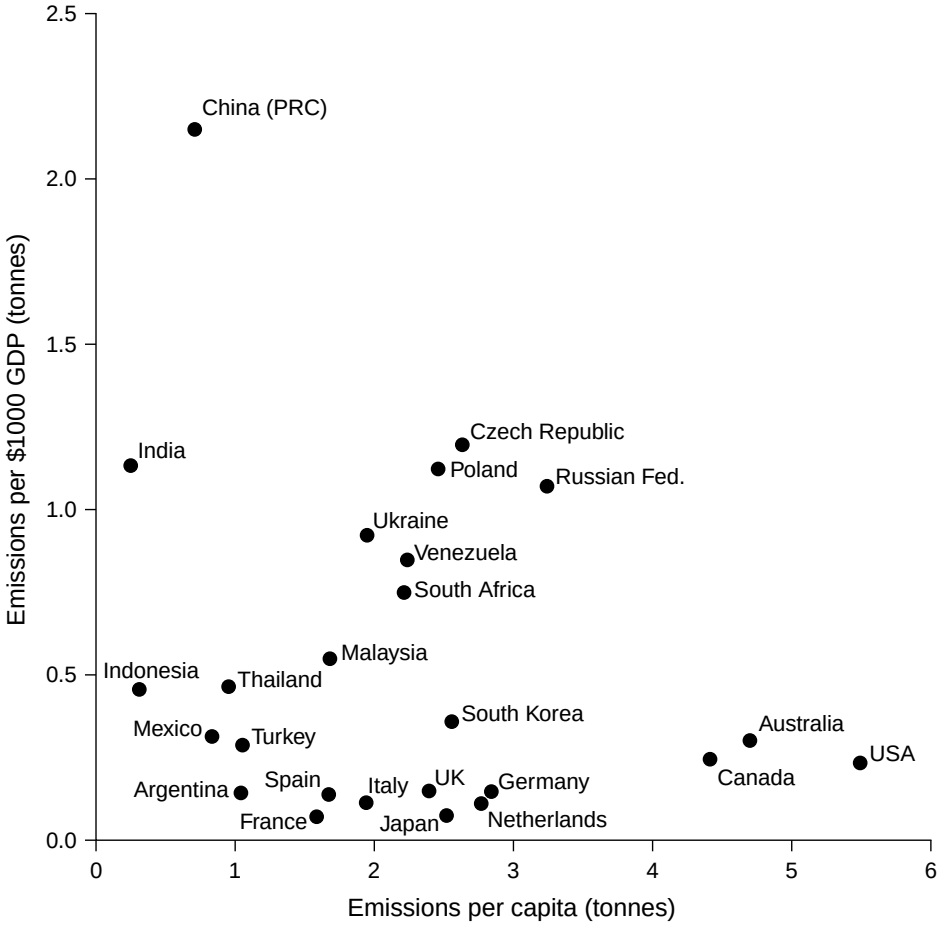


Figure 2 Plot of tonnes carbon dioxide emissions per $1000 GDP against tonnes carbon dioxide emitted per capita for the countries in Table 1

compared to Australia s 17.6 t cap^{-1}), Australia has much higher levels of emissions from agriculture and land-use change.

In descending order, the six nations with the highest per capita emissions are: Australia (26.7), Luxembourg (24.2), US (21.2), Canada (20.6), New Zealand (17.3) and Ukraine (16.7). The next five countries have emissions per capita of 14 15 tonnes. Luxembourg s very high emissions are due to the presence of a large steel plant. New Zealand has low energy emissions (due to the predominance of hydroelectricity) but very high emissions from agriculture (due to the large number of sheep). These are offset to some extent by the net sink provided by forests in that country.

Among larger countries at the other end of the scale, France (7.8 t cap^{-1}), Germany (12.6 t cap^{-1}), Spain (7.1 t cap^{-1}), Italy (9.0 t cap^{-1}) and Japan (9.5 t cap^{-1}) are notable. Their low emissions are due to a combination of energy efficiency, industrial structure and the use of nuclear power.

CONCLUSION

Plots of emissions per $1000 GDP against emission *per capita* for 25 high total emission countries (Figure 2) show the contrasting positions of the three countries with the highest total emissions. China is high per unit GDP and fairly low *per capita*. The Russian Federation is moderately high in both per unit GDP and *per capita* terms, while the USA is low in per unit GDP terms, but highest in *per capita* terms. Countries like Thailand, Mexico and Argentina are low on both criteria.

When emissions per $1000 GDP are plotted against emissions per km^2 (Figure 3), the majority of the 25 countries crowd into the lower left-hand corner of the figure, but the densely populated industrial countries, like South Korea and

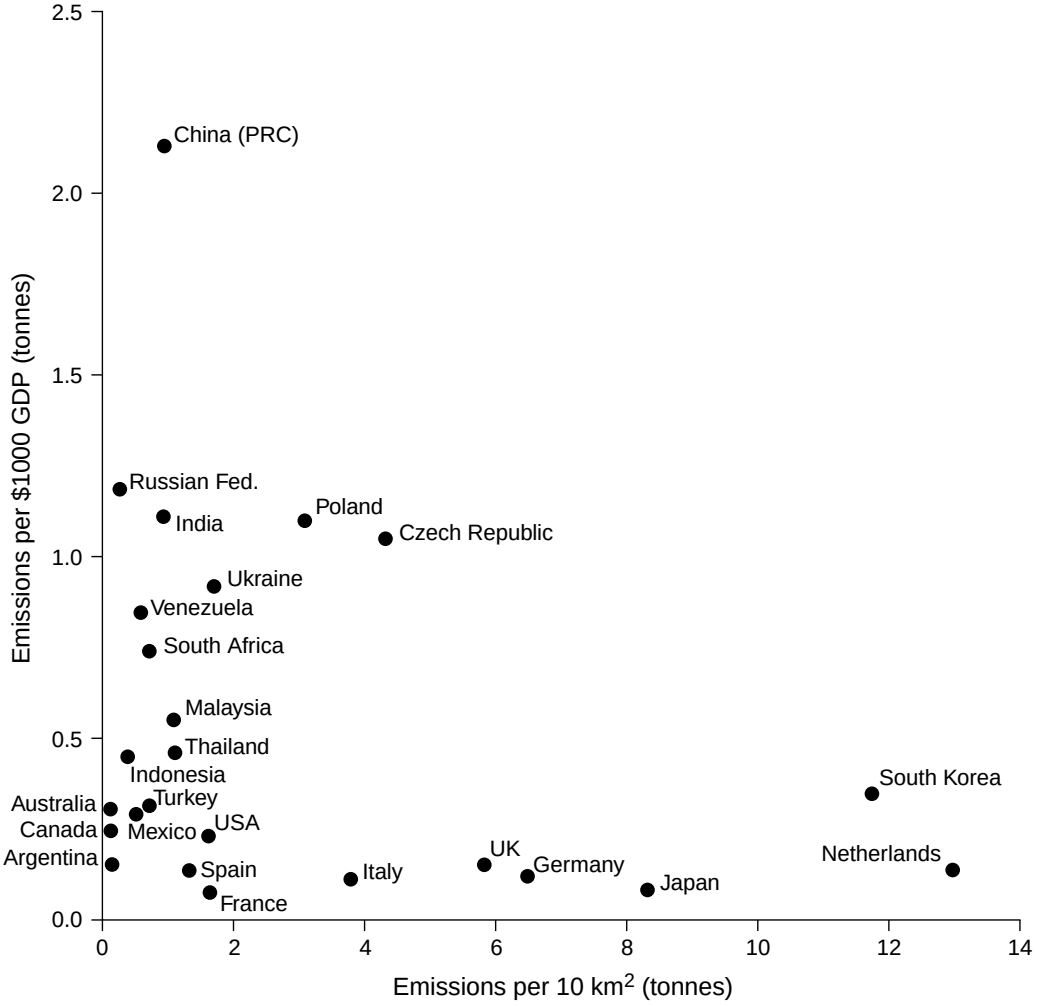


Figure 3 Plot of tonnes carbon dioxide emissions per $1000 GDP against tonnes carbon dioxide emitted per 10 km^2 for the countries in Table 1

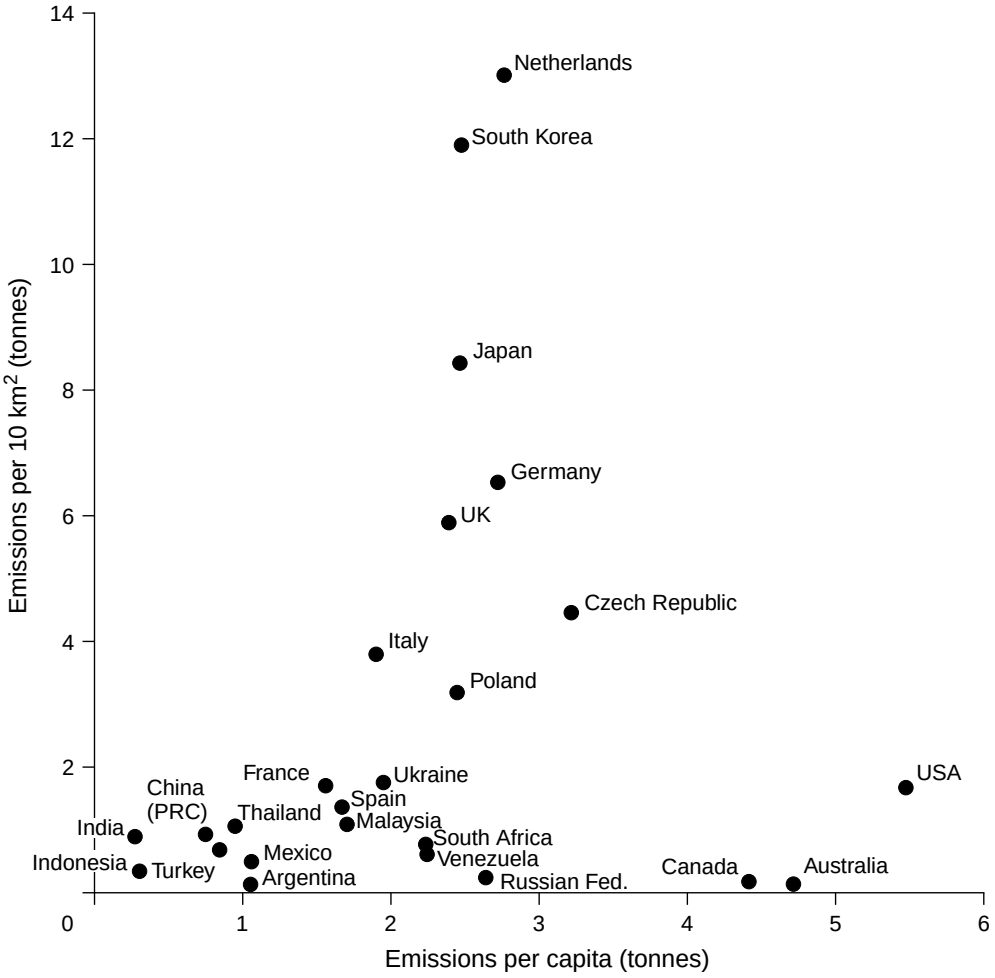


Figure 4 Plot of tonnes carbon dioxide emissions per 10 km^2 against tonnes carbon dioxide emitted per capita for those countries in Table 1

the Netherlands plot towards the lower right. The countries with enormous reliance on coal and relatively old and less efficient industrial facilities plot towards the top left-hand corner.

The plot of carbon dioxide emissions per $1000 GDP (Figure 4) separates the 25 countries more clearly, with the high *per capita* countries in the lower right corner and the high per GDP countries in the lower left. Interestingly, the newly industrialising countries, such as Malaysia, Thailand and Venezuela, plot towards the centre of the graph.

These graphs and tables demonstrate that responsibility for greenhouse gas emissions is extremely unequally distributed amongst countries. One fifth of the world s population accounted for almost two thirds (62%) of all emissions in 1996 while another, much poorer, fifth accounted for less than 2%. Industrialized countries, with relatively slowly growing populations, have the largest *per capita* emissions, as Tables 1 and 2 show. Growth in *per capita* emissions, by contrast, tends to be most rapid in less wealthy countries, many of whose populations are growing rapidly. The gap in *per capita* emissions by nation is narrowing, but so slowly that there is no reason to expect relatively equal *per capita* emissions at any point in the future.

The fifth of the global population collectively contributing less than 2% of all greenhouse gas emissions (over a billion people) could rightfully argue that they are not contributing to climate change at all. Since their modest *per capita* greenhouse gas emissions, if emulated by everyone else in the world, would result in stable or even falling greenhouse gas concentrations in the atmosphere. This poor population is growing rapidly, however, and in escaping poverty, their descendants greenhouse gas emissions are likely to come closer to the global average and be less sustaining of the global climate. The more people there are in the world, the lower the average *per capita* emission will have to be if climate change is not to accelerate.

Table 2 Total emissions of the three main greenhouse gases (carbon dioxide, methane and nitrous oxide), breakdown by source and per capita emissions for Annex B countries, 1995 (Mt CO_2-e)

	Energy					Land-use change and forestry			Population 1995 (million)	Per capita emissions (t CO_2-e/ capita)
	Fuel combustion	Fugitive	Industry	Agriculture	Waste		Other	Total		
Australia	291.77	25.58	7.45	87.36	16.36	51.87	1.55	481.94	18.07	26.67
Luxembourg	9.16	0.04	0.41	0.51	0.08	−0.3	0.01	9.92	0.41	24.19
US	5206.4	202.49	96.43	268.23	236.44	−428	0	5581.99	263.17	21.21
Canada	478.96	48.2	36.34	25.04	19.47	0	3.31	611.32	29.62	20.64
New Zealand	24.95	1.19	2.74	44.33	2.77	−13.49	0.2	62.69	3.66	17.13
Ukraine	671.17	130.81	33.7	50.5	19.68	−51.98	7.25	861.12	51.55	16.7
Denmark	58.91	0.71	1.31	16.17	1.55	−0.96	0.46	78.14	5.23	14.94
Ireland	33.27	0.23	2.58	19.28	2.95	−6.23	0.75	52.83	3.6	14.68
Netherlands	183.66	3.57	7.61	18.31	9.13	−1.7	1.48	222.06	15.46	14.36
Czech Rep.	130.37	8.51	5.22	3.45	3.02	−5.45	0.3	145.42	10.33	14.08
Belgium	112.83	0.95	14.27	11.52	4.99	−2.06	0.06	142.56	10.14	14.06
Germany	885.13	24.57	50.31	61.52	39.9	−30	0	1031.43	81.66	12.63
Lithuania	37.75	0.55	2.64	7.15	3.49	−8.85	4.09	46.81	3.72	12.58
UK	533.77	23.94	28.93	26.19	38.44	9.95	1.53	662.75	58.61	11.31
Finland	57.33	0.08	1.77	4.64	2.79	−10.5	0.08	56.19	5.11	11
Russian Fed.	1607.27	297.2	24.37	114.53	41.04	−568	9.95	1526.37	148.2	10.3
Poland	365.18	18.9	13.76	22.87	17.96	−41.95	0.23	396.94	38.59	10.29
Greece	84.79	1.03	8.33	8.37	2.77	0	0	105.29	10.45	10.08
Iceland	1.77	0.08	0.46	0.29	0.04	0	0.01	2.65	0.27	9.81
Slovakia	45.99	2.25	3.43	4.24	1.45	−5.12	0.19	52.42	5.37	9.76
Bulgaria	59.34	5.57	8.18	3.4	10.96	−7.52	0.06	79.98	8.41	9.51
Japan	1162.1	3.55	68.65	20.65	28.54	−94.62	1.51	1190.38	125.57	9.48
Slovenia	13.6	1.07	0.64	2.35	1.6	−2.29	1.79	18.75	1.99	9.42
Italy	425.2	10.07	29.31	41.84	21.65	−24.51	12.39	515.95	57.3	9
Norway	29.89	2.35	8.52	3.88	6.78	−13.64	0.34	38.12	4.36	8.74
Austria	50.05	2.46	11.49	5.41	4.63	−13.58	4.13	64.59	8.06	8.01
France	365.79	14.33	40.79	48.88	19.15	−46.8	9.93	452.06	58.14	7.78
Portugal	47.92	0.26	4.01	6.33	13.77	−1.15	0.27	71.41	9.92	7.2
Spain	221.62	13.41	18.85	37.64	15.3	−28.97	0.04	277.88	39.21	7.09
Hungary	58.97	6.62	2.28	3.06	6.11	−4.8	0	72.23	10.23	7.06
Switzerland	40.95	0.34	2.71	5.84	2.85	−5.1	0.12	47.72	7.08	6.74
Estonia	20.93	0	0.22	0.84	0.67	−13.27	0	9.39	1.49	6.3
Monaco	0.08	0	0	0	0.05	0	0	0.13	0.03	4.3
Sweden	56.29	0.02	5.17	4.2	1.28	−30	0.25	37.21	8.83	4.21
Latvia	12.16	0.46	0.13	5.81	0.64	−10.48	0.04	8.76	2.52	3.48
Total	13 385.31	851.36	542.98	984.59	598.3	−1409.5	62.33	15 015.37	1106.36	13.57

Source: Turton and Hamilton, 1999.

These contrasts between nations pose severe problems in negotiations that affect the causes of climate change. They need to be understood in order to appreciate the difficulties which negotiators face. Some argue that emissions rights should become proportional to people rather than remaining proportional to accumulated income and environmental impact (Global Commons Institute, 1999). Others make the case that there have always been centers of high consumption of energy *per capita* and per unit GDP and that these centers change over time. According to them, setting a common *per capita* standard would stultify the industrial activity which has over the last 200 years made it possible for the world to support a six-fold expansion of its human population.

REFERENCES

Global Commons Institute (1999) *Climate change - A global problem: Contraction and Convergence - A global solution*, http://www.gci.org.uk/contconv/cchist.html.

Marland, G, Boden, T A, and Andres, R J (2000) Global, Regional, and National CO_2 Emissions, in *Trends: A Compendium of Data on Global Change*, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, TN.

Turton, H and Hamilton, C (1999) *Greenhouse Gas Emissions Per Capita of Annex B Parties to the Kyoto Protocol*, Submission to the Senate Environment References Committee Inquiry into Australia s Response to Global Warming, November 1999, The Australia Institute, Canberra.

Natural Hazards

Kenneth Hewitt
Wilfrid Laurier University, Waterloo, Canada

Natural hazards refer to conditions or agents in the planetary environment that have caused damage to life and property, and are likely to do so again. They include extremes of heat and cold, hurricanes and droughts, earthquakes and landslides, forest fires and locust swarms. They are natural in the sense of occurring spontaneously, as processes of the planetary environment. Objective physical or biological relationships and laws govern their behavior. They may occur whether humans are present or not. Of course, human activity can, and increasingly does, intervene to modify the incidence and severity of these dangers, or the damaging events associated with them.

Used strictly, the term hazard has an interactive and an evaluative meaning. Natural conditions become hazards where a potential exists for harm to humans - a function of their exposure to destructive forces and their vulnerability or lack of defenses against such forces. However, most studies treat the hazard as, literally, a natural process such as volcanism, a virus or a flood wave. The field is organized in terms of natural agents, with specializations in geophysical, biological and environmental hazards. Moreover, the work has mainly addressed the more extreme or calamitous dangers.

The importance of natural disasters is reflected in continuing large death tolls and property losses. In the late 20th century, agencies like the Red Cross and the United Nations Disaster Relief Organization recognized some 210 natural disasters in an average year, worldwide. These involved annual death totals of about 128 600 and 90 000 people severely injured. Some 4.8 million people were made homeless and, in all, about 137 million directly affected. The evidence suggests that the global losses have been growing over recent decades. Although notoriously hard to define, economic costs in the 1990s were put at about US$88 billion annually (IFRCRCS, 1997). This takes no account of thousands of other less destructive episodes not reported in major disaster inventories.

Global environmental change may increase the intensities and/or change the temporal and geographical incidence of natural hazards such as droughts, storms and floods. Trends, such as sea-level rise, or degraded ecosystems, may be dangerous in themselves. They can also magnify other hazards such as storm wave impacts, or wildfires and insect infestations, respectively. There are dangerous interactions with novel human technologies and activities, as demonstrated in photochemical smogs, acid precipitation, Arctic haze, drug-resistant diseases and biocide-resistant pests.

In this context, the need for detailed and reliable knowledge of natural processes and environments, has never been more urgent. However, risks from nature, or the dimensions of natural conditions that we need to address, are altered and complicated by societal change. Of singular importance are the redistribution of populations and intensification of land uses, rapid urbanization, megaprojects, ecological abuses and disruptions occasioned by warfare (UN/DHA, 1994). Hazards research increasingly recognizes the complexities resulting from changing technologies and activities, and changing human exposure or vulnerability to hazardous agents.

While trying to retain a sense of the latter here, we will focus mainly on the characteristics and scope of dangerous geophysical agents and related disasters. Problems arising from global environmental changes, examined in more detail elsewhere in these volumes, will be indicated throughout.

NATURAL HAZARDS AND DISASTERS

Natural conditions can cause harm over a broad range of scales and intensities. There are everyday hazards and chronic forms of damage. They include most weather-related traffic accidents, common childhood infections and the almost imperceptible, repetitive deterioration of buildings by weathering and molds. These tend to be tolerated or dealt with routinely by citizens, communities and well-established agencies.

There are also extreme and unusual conditions that cause sudden, extensive bouts of damage and may lead to natural disasters. Usually, they become threatening and destructive because they exceed manageable magnitudes, intensities, or thresholds. It might be extremes of heat and cold, the height of floodwaters, winds above a certain speed, or the length of time and/or area covered by drought conditions.

Other distinguishing features and problems of natural disasters relate to casualties or property damage highly concentrated in space or time (Quarantelli, 1998). The harm may overwhelm local or regional, sometimes national, capacities to respond, or to sustain acceptable living conditions. Extraordinary measures are required. As a distinctive field of enquiry, the study of natural hazards has mainly addressed these more extreme or disastrous conditions.

Damage and disaster can also occur with unusual or untimely combinations of conditions, not individually threatening (see the section on compound hazards). There are geographic or out-of-place dangers: conditions common in some places but so rare, or formerly absent, in others, that people lack the capacities and experience to withstand them. Conversely, activities, equipment, or domesticated plants and animals that are well suited or adapted to one environment, may not be in another.

In none of these cases is a hazard a separate reality. On the one hand, every component of the natural environment may be harmful for some persons, or at certain times, including what are, otherwise, treated as natural resources. On the other hand, even phenomena we may think of as only and always destructive, like earthquakes, tornadoes, termites and forest fires, have been shown to contribute to the long-term health or renewal of the biosphere. They are not just unfortunate natural phenomena. For example, the torrential rains associated with hurricanes contribute significantly to groundwater reserves in the US southeastern states.

For these reasons, natural hazards are as much about human adaptation and preparedness against dangers in given habitats, as the intrinsic properties of natural agents. Global environmental change presents both the prospect of changes in hazardous agents, and transformed human relations to them.

Etiology of Hazards

Nevertheless, we need to have in mind just how they become hazardous, or contribute to disaster. Any damaging event will involve a series of factors, in which natural processes can be identified as:

1. Generative conditions: those that give rise to dangerous forces, agents or events. Examples include the tectonic forces in the Earth s crust, and fault mechanisms that give rise to earthquakes; conditions over the tropical mid-ocean regions that can generate hurricanes, or dry weather and the state of forest vegetation in promoting wild fires.
2. Primary damaging agents: the conditions or events that directly initiate and cause damage, such as strong ground motion with the arrival of earthquake waves, volcanic explosion, hurricane winds, or a conflagration. These are what the term natural hazards is most often used to refer to.
3. Secondary and tertiary hazards: damaging agents triggered by other hazards, such as landslides set off by earthquakes, or mass fires by electrical storms. Tertiary hazards are follow-on dangers such as disease outbreaks after floods. Often, these secondary or tertiary hazards, or their impacts, are simply lumped together with the primary agent. However, they have their own distinctive properties as hazards, and reflect the setting in which a primary hazard occurs. They show the importance of intervening habitat conditions.
4. Intervening habitat conditions: these moderate, magnify or redirect damaging agents. Examples would be the influence of topography on snow drifting or on cold air drainage in frost events; coastal zone morphology in wave action; or the age and composition of forest stands in the fire hazard. Secondary hazards and their importance, mostly reflect such intervening conditions.

The factors listed above involve the natural habitat. Equally, intervening conditions created by human activity may modify natural hazards or give rise to secondary artificial hazards. The urban mass fires that followed the San Francisco, 1906 and Tokyo, 1923 earthquakes, illustrate how such factors transform the type and severity of damage triggered by natural forces. Moreover, losses in natural disasters always depend upon whom and what are exposed to given natural agents, their state and available protections. How human exposure and vulnerability, and adjustments to and defenses against, dangers enter into natural hazards will be only briefly indicated here.

DOMAINS AND CLASSES OF NATURAL HAZARD

Hazards investigations mainly follow the division of the planetary environment into atmosphere, hydrosphere, cryosphere, lithosphere and biosphere. It will be appropriate, therefore, to review our physical understanding of hazards in such terms. Remember, however, that few destructive events are confined to just one process or natural sphere. A disaster initiated in one domain may involve more destructive processes in another: for example, floods or landslides triggered by rainstorms, forest fires set by lightning, and epidemics following drought or flood. Moreover, human activities mostly occur at interfaces of these spheres, such that harm and response in any disasters are constrained by aspects of all four.

The full range of dangers is enormous, and can only be outlined. We will then concentrate on those known to be especially damaging, or likely to be sensitive to global change, and disasters that illustrate these points.

Atmospheric Hazards

The most widespread and frequent environmental disasters are those initiated by atmospheric processes (Table 1). In part, this relates to the diversity and changeability of weather conditions, and recurring extremes. In part it reflects the continuous exposure of people and property to them.

Atmospheric hazards can be treated as single conditions such as high winds, severe cold, evaporation losses. However, in storm systems and any disaster situation, several hazardous forces usually occur together. In some, such as tornadoes, the forces involved are so great that few affordable structures or activities can be protected from destruction. In the modern world, because more or less elaborate arrangements have been put in place to

Table 1 Classes of atmospheric hazards

Single condition/process	Multiple process or event hazards
Temperature (high/low)	(Meteorological) drought
'cold spell', heat wave	Temperature inversion
Fog	
(high) Winds	Wind storm
Hydrometeors:	
Rain, rainfall	Rainstorm, mid-latitude depression
Snow, snowfall	Snowstorms, 'Whiteouts' Blizzards
Freezing rain ('glaze')	Glaze or ice storms
Lightning	Thunderstorms/hailstorms
	Tornadoes
	Tropical cyclones: Hurricanes (Atlantic Ocean) Typhoons (Pacific) Cyclones (Indian Ocean)

cushion people and property from the weather, serious harm usually results from their breakdown or destruction, or from failures of common protective measures and preparedness. Vulnerable people are suddenly exposed to harsh conditions in cities with failure of fuel and power lines, for example, on highways or in recreational areas.

The incidence and severity of dangerous atmospheric agents vary greatly in different climatic regions. Some are largely or wholly confined to particular regions. Alternatively, it is the relative frequency, or combinations, of conditions that vary. There are hazards identified mainly with certain broad zonal divisions organized in latitudinal belts, from equatorial to polar regions. There are azonal environments, such as coastal zones, mountain belts or continental interiors, which generate their own distinctive combinations of hazards.

Global Change and Atmospheric Hazards

Climate modification has the potential to alter the magnitude and frequency of hazardous atmospheric agents, their geographical scale and incidence. It is thought that a relatively greater incidence of damaging events in recent decades is, at least partly, due to more severe weather or changing weather patterns.

Recent human impacts on the atmospheric environment have undermined some of its singular benefits for life on Earth. Perhaps the most critical relate to the ways in which the atmosphere transmits and absorbs radiant energy. As an absorber of, or barrier to, incoming high-energy solar and cosmic radiation, the atmosphere shields living tissue from harmful irradiation. The critical role of the stratosphere in this regard is threatened by ozone-stripping industrial chemicals, causing unprecedented holes in the stratospheric ozone layer. Rapid increases in greenhouse gas emissions affect the atmosphere s capacity to absorb and retain thermal energy, and can induce global warming. Associated changes in weather conditions, perhaps climatic zones, threaten existing ecosystems, human economies and health.

Increases in atmospheric particulate matter from fossil fuels and other industrial activity, from desertified land surfaces or increased deliberate or inadvertent burning of vegetation, may have a range of dangerous consequences. They may block sunlight, lower local temperatures in agricultural zones, and suppress cloud formation. Meanwhile, these dangers become coupled with the beneficial roles of atmospheric circulation, in transfers of heat and moisture around the globe. The same conveyors also transport toxic materials from industrial, urbanized, war-torn, or modernized agricultural regions, to distant places.

Hydrological Hazards

Hydrology includes all aspects of the movement and storage of moisture, including water in the atmosphere (hydrometeors like rain and snow), lithosphere (ground water) and biosphere (soil, wetlands and transpiration). Moisture-related risks can arise where human society interacts with each of these. However, the focus of hydrological hazards has been on surface waters, especially streams and rivers, freshwater lakes, glaciers, and the oceans, and on acute shortages of fresh water (Table 2). Drought, the absence or deficiency in fresh water supplies, although it arises most often from atmospheric conditions, is generally examined as a hydrological hazard.

Table 2 Classes of hydrological hazards

Single condition/process	Multiple process or event hazards
Runoff (overland, stream)	Floods:
Snow-on-the-ground	Riverine
Groundwater	Lakeshore
Freeze-thaw	Coastal (marine)
Sea ice	Storm surge
Icebergs	Ice-infested waters
Wave action	Storm waves
Glaciers	Waterlogging
	Glacier 'surges'
Evaporation and transpiration	Natural dams: Glacier Landslide Vegetation
Salinization (of freshwater resources)	Droughts: Meteorological Runoff Aquifer

Danger tends to arise with extremes of occurrence or behavior; including various powerful hydrodynamic processes, excessive or deficient moisture, and high or low water conditions. These hazards may also be defined through single processes such as stream flow and evaporation, or in events such as floods, droughts, and glacier surges, that combine several hazardous conditions.

Attention to flood and drought hazards has outstripped all others. This is hardly surprising. The former comprise the single most widely reported, lethal, and costly disasters. They occur in nearly all habitable land areas. Major droughts, though less frequent than floods, when they do occur tend to affect the greatest numbers of people (UN/DHA, 1994).

The importance of floods and droughts reflects the close association between water supply or water bodies, and human populations, including most towns and cities. Risks arise mainly through the forms of human dependence on, and use of, water as a resource, including large demands for domestic, agricultural, municipal and industrial consumption. People and property may also be exposed to floods in tapping biological resources of seas, lakes and rivers; through water-borne transportation; mercantile and naval strategy; and water-related amenities and recreation. Water-borne disease, in terms of numbers of victims, is probably the greatest threat associated with these relationships and adverse effects on the aquatic environment. This is illustrated by many recent outbreaks or epidemics of cholera, bilharzias, and *Escherichia coli*, among others, and infestations by water-dependent insect vectors like malaria-carrying mosquitoes.

Global Change and Hydrological Hazards

Recent climatic trends suggest that both floods and droughts are likely to become more severe in future. Climate warming is likely to intensify the thermal and evaporative conditions in drought. However, flood risks are also likely to increase, even in the same areas. Greater warming of ocean and freshwater surfaces, and higher temperatures generally, should drive higher rates of evapotranspiration. This increases the amount of moisture entering the atmosphere and, ultimately, its turnover through condensation and precipitation. Storminess is also likely to intensify, increasing the amount of precipitation in more severe events and, hence, flood risks. Some regions, especially coastal and high mountains, are likely to see higher, if more variable, precipitation. Moreover, water quality, already tending to deteriorate world wide, is likely to be adversely affected by increased drought conditions, and larger fluctuations. Meanwhile, availability of fresh water in most regions is declining, while there is ever-greater exploitation of, and conflict over, water resources.

Table 3 Classes of geological and geomorphological (lithospheric) hazards

Single condition/ process	Multiple process or event hazards
Seismicity	Earthquake
Earth/rock materials:	
Quick clay	Liquification
Quicksand	
Radionuclides	
Radioactivity	Domestic radon gas
Geothermal heat	Volcanic eruptions
	Ash (tephra) fall
	Lava flows
	Pyroclastic flows
	Gas or vapour eruption
	Explosive eruption
Mass movement	Landslides:
	Rock slide, rock avalanche
	Debris- and mudflow
	Earthflow
	Dirty avalanche
	Density currents (lake)
	Submarine slides
Subsidence	

Geological Hazards

These arise from processes of the lithosphere or solid earth. They fall into three broad classes, seismo-tectonic, volcanic, and geomorphic or earth surface processes (Table 3). Earthquake disasters are the most frequent and destructive, followed by volcanic eruptions. Landslides are the most widespread source of damages, but less often the primary cause of disasters. Danger, usually of a more long-term and dispersed kind, may also arise where earth materials contain and release toxic chemicals, or ionizing radiation (*see* Contaminated Lands and Sediments: Chemical Time Bombs?, Volume 3).

Ultimately, the hazards arise from the large scale and continuing evolution of the Earth s crust. They depend upon isostatic, tidal and geothermal conditions, and crustal processes now interpreted in terms of plate tectonics. Of special relevance are the great stresses, volcanism and intensified erosional processes that occur at plate margins. They include the extending crust at the mid-ocean ridge systems, but especially collision zones, whether of continents as exemplified by the Alpine Mediterranean and Himalayan mountain zones, or continent sea floor margins, such as surround the Pacific Ocean.

Geomorphic hazards derive from subaerial or landscape forming processes. Most are actually triggered by atmospheric, geological or hydrological extremes. They are highlighted as geomorphic hazards when the main impacts come from erosion, transport and deposition of earth materials. The distribution of geomorphic hazards is strongly

dependent on the relief and ruggedness of land and submarine surfaces the composition, structures and weathering of rock and soil masses. Most are triggered by climatic and hydrological conditions.

The dangers and damages from landslides tend to be greatly underestimated. This is partly because few individual events affect large areas or numbers of people. Mainly, it is because the worst landslide damages occur as secondary hazards in events identified with other processes. In particular, they occur with earthquake, volcanic, flood and severe storm disasters. In more than half of all earthquake disasters, slope failure and landslide impacts are greater causes of death and destruction than primary seismic motion (Hewitt, 1997, 226).

With respect to human safety, geological hazards mainly involve processes that suddenly alter or destroy the stability of the land surface. The solid earth affects material security mainly through the sites of human settlement and, especially, the support of built structures. The destruction in earthquakes, landslides, and the more lethal volcanic eruptions, is largely to buildings. Most fatalities occur when buildings collapse on people inside or nearby, usually peoples homes. Under modern conditions, geological hazards are also identified with lifeline disasters. They commonly sever, block or destroy the built infrastructure of roads and railways, water, power and fuel lines. Geomorphic and volcanic events may also ruin more or less extensive areas of cropland and natural vegetation, burying them in ash fall, lava flows, and in mass movements, of water- or wind-borne sediments.

Global Change and Geological Hazards

Tectonics and volcanism reflect long-term geosphere developments, rather than global change over decades or centuries. Nevertheless, dramatic changes in risks from these processes are likely as a result of changes in human exposure and vulnerability. Meanwhile, global environmental change involves significant concerns in the scope and scale of geomorphic hazards.

Earth surface processes are directly affected by climatic hydrological controls. Geomorphic activity is also modified in landscape changes by human agency (Hewitt, 1989). Increases in sediment yields with intensified cultivation or urbanization, and increased land sliding after deforestation, are well documented. The artificial fragmentation of rivers is already associated with more or less severe geomorphic, as well as biotic, effects, including downstream erosion and upstream siltation following dam building (Dynesius and Nilsson, 1994). In fact, all of the more important landform-shaping processes are being drastically modified by human agency or climate change streams and rivers, glaciers, freeze thaw in soils, and the active layer over permafrost, coastal and aeolian action, and biological activity over and in the near-surface zone.

Dangerous contaminants introduced into geomorphic systems are less visible but, perhaps, more critical for human health and harder to reverse. In addition to well-known problems with contaminated ground waters, earth surface processes of all kinds may selectively pick up, carry and deposit contaminated sediments in distinctive ways and patterns (Stone and Droppo, 1994; Mallin, 2000). The sources of pollutants are industries, fossil fuels, municipalities, mechanized agriculture, factory raised farm animals, and mineral and biotic resource extraction. However, natural geomorphic processes redistribute and concentrate these dangers. Only recently have their roles in the toxic waste stream started to be fully recognized and addressed (Callow and Petts, 1994). Strictly speaking, it is part of the burgeoning problems of complex hazards at the interface of modern technologies, industries, land uses and natural processes (see following sections).

Biotic Hazards

These involve dangers to humans from other organisms or conditions in the living world. They arise both from human interdependence with the rest of the biosphere, and efforts to control or segregate parts of it for human use. The latter leads to competition with other organisms that are then called pests, weeds, infestations, etc. However, not only are damaging impacts of modern societies on the biosphere a great cause for concern, they are a large and integral part of biological hazards that come to plague society (Table 4).

Compound and Complex Hazards

Each domain of the natural environment gives rise to a variety of conditions that may threaten human societies. Each has been the focus of particular hazards studies. However, as noted, actual damaging events involve at least three, and usually all four, natural spheres. In addition to hazards in one natural domain that trigger secondary ones in another, disasters often involve compound hazards adverse combinations of atmospheric, hydrological, biotic and geological conditions (Table 5). The case studies of natural disasters provide ample illustrations.

In recent years, the idea of complex emergencies has been proposed, and the growing significance of such situations underlined. From a hazards perspective, these are events in which natural hazards combine with harmful technologies and/or social processes. Experience with severe famines in Africa give particular support for the notion. Drought and crop loss do not, alone, decide who starves, which people become refugees, and what their needs are. Usually, economic and social change are equally involved. In most cases, that includes armed conflict, misrule and developments that reduce food security. Failures of national

Table 4 Classes of biological hazards

Single condition/organism	Multiple process or event hazards
	Disease outbreak/epidemic
Bacteria (Pneumonia)	Bubonic plague
Protozoa (Giardia, Malaria)	Yellow fever
Fungal (Pneumcystas)	Influenza pandemic
Viruses (Measles, HIV, Dengue)	
Algae	Harmful algal bloom 'Red tide' (toxic algal blooms)
Sexually transmitted diseases Chlamydia, Syphilis, HIV, etc.	
Domestic animal disease: Hoof and mouth disease 'Mad cow' disease	Outbreaks
Domestic crop diseases Blights, rusts	Outbreaks, crop failures
Plant hazards or 'weeds' Water hyacinth, Prickly Pear	Plant infestations, 'Invasions'
Insects 'pests' Locust/grasshopper Lice, cockroaches	Insect plague/infestation
Animal 'pests' Rats, mice	Infestation/population explosion Rat 'plague' Rabbit infestation
Wild animal disease Rabies Shellfish poisonings Neurotoxic poisoning (e.g., whales)	

Table 5 Compound and complex hazards

Compound natural hazards
Permafrost and 'active layer' processes
Volcanic mud flows (lahars)
Volcanic floods ('jökulhlaupts')
Landslide-, and vegetation-dam outburst floods
Tsunami (seismic sea wave) coastal flood
Complex (natural + artificial agents)
Smog
Arctic haze
Acid precipitation
Artificial dambreak flood
Urban firestorms
Famines
Environmental warfare agents
Irrigation-induced waterlogging
Combustion generated greenhouse gases and global warming

and international humanitarian assistance occur for similar reasons (IFRCRCS, 1997).

Recognizing such interrelationships among hazardous processes is of singular importance for anticipating how they may change with global environmental change. Even where one sphere is mainly affected, it can magnify, redirect or alter the significance of hazards in another. Sea level rise is likely to increase the dangers from tsunamis, storm surges, and coastal erosion. Changes in weather patterns and vegetation cover will be reflected, especially, in dangers from hydrological and geomorphic processes, in earthquakes and wild fires. Each of these risks will change with changing human technologies, activity and vulnerability.

We can anticipate that modern disasters, whatever the specific agent that triggers them, become complex emergencies. Earthquakes and severe storms in urban centers are now complicated by the presence of stores of flammable and volatile chemicals. Some of the most serious and enduring dangers and losses from floods are the toxic wastes flushed from agricultural operations, chemical industries, or overloaded sewage treatment plants. Again, one can see how global change raises the prospect of more and more severe, interactions between natural and artificial systems, increasing the incidence of such complex disasters.

THE MAJOR FORMS OF NATURAL DISASTER

Tropical Cyclones

These are the greatest storms on Earth. They are known as hurricanes in the Atlantic and eastern Pacific, typhoons in the western Pacific and cyclones in the Indian Ocean. They

develop over open areas of the warm, tropical oceans, but do their greatest harm where they pass over oceanic islands, and/or strike mainland coasts. Most disasters occur on sub-tropical eastward- and equatorward-facing shores of the three oceans (*see* Hurricanes, Typhoons and other Tropical Storms – Descriptive Overview, Volume 1; Hurricanes, Typhoons and other Tropical Storms – Dynamics and Intensity, Volume 1).

The storms involve a number of destructive forces, including high winds, torrential rains, and intense electrical or lightning activity. Over the ocean, the storms may also generate waterspouts and, over land, tornadoes (Pielke and Pielke, 1997).

The high winds can cause significant property damage directly, and from the impact of debris carried by them. Power and communication lines are disrupted and fallen debris closes access routes. Usually, however, the greatest loss of life and devastation are due to the storm surge. This derives from a piling up of water in the low-pressure eye of the cyclone. On reaching land, the surge can be higher than the highest tides. It may be further elevated by wind-driven wave action.

Notorious in this respect are Bay of Bengal cyclones that reach the low-lying, deltaic coasts of the Ganga Brahmaputra delta. The surge in the November, 1970 cyclone, drowned between 400 000 and 500 000 people in Bangladesh (then East Pakistan), the most lethal of these storms in the past century.

In other cases, torrential rainfall has caused the greatest damage; usually through floods and/or landslides triggered by the rain. Hurricane Mitch, where it affected the Central American republics in November 1998, is an example.

Case Study 1

Hurricane Mitch, October 26 November 4, 1998, exemplifying the dangers of intensified tropical storms, and of increasingly vulnerable land uses and populations.

This was the most lethal Atlantic hurricane in 200 years, and the fourth strongest on record. Especially critical was the path followed by the storm and the way it stalled over the Central American republics of Honduras, Nicaragua, Guatemala, and El Salvador. Here, where the greatest damage occurred, the storm s impact was greatly aggravated by recent changes in land use, and redistribution of vulnerable populations. There were also losses in Costa Rica, Belize, Panama, Mexico and the US.

The storm attained Category 5 intensity highest on the Simpson Saffir Hurricane Intensity Scale for over 33 h, with sustained winds of over 290 km h^{-1} and gusts of over 320 km h^{-1}. Its destructive potential was maintained or recovered over 10 consecutive days. An unusually warm year, probably related to El Nino global warming, may have increased the intensity, and perhaps the likelihood, of this hurricane.

However, the heaviest rainfall was over volcanic mountain ranges, when a downgraded Mitch drifted slowly over Central America, stalled by a blocking frontal system. As much as 125 190 cm of rain fell in two to three days. Secondary hazards induced by the rainfall, especially major mud slides and flooding, caused the greatest loss of life and destruction.

Preliminary estimates of the death toll were 10 000+, with 12 000 missing. Some 50 000 people were injured. Honduras was the hardest hit with 6500 deaths, 11 760 reported missing, and 1 million made homeless. There were major agricultural losses, Guatemala lost 95% of its banana crop and 30% of its cattle. There was huge and widespread damage to the sugar cane, coffee, corn and bean crops.

Forest removal and land use changes, notably the spread of large commercial plantations in the previous two to three decades, were blamed for increased flood potential and erodibility of soil. Enforced redistribution of peasant populations into more mountainous areas also encouraged further forest clearance and destabilized slopes. Meanwhile, other poor and displaced persons were concentrated in coastal zones, and along river courses. This increased their exposure to, and vulnerability from, mud slides and floods.

Tornadoes

Unlike hurricanes, tornadoes are associated with frontal storm conditions, mainly in sub-tropical and mid-latitude continental areas. They occur with intense, convective storms, and usually in the summer half of the year. About 70% of all reported tornadoes occur in the US Midwest.

The tornado is a funnel cloud that descends to earth from intense convective storms. It comprises a vortex of whirlwinds around a central, low-pressure area. The highest known wind speeds and enormous numbers of lightning strikes to earth occur with tornadoes. All combine to magnify storm destructiveness and the problems of emergency response (Church *et al.*, 1993).

The width and length of the path where the funnel touches down is as critical as the dangerous forces in it. Rarely is the swath more than a few hundred meters wide or a few kilometers in length. Occasionally, however, a tornado may remain active and in contact with the ground for tens or even hundreds of kilometers, or repeatedly lift and descend again. The damage swath may reach 2 3 km wide. An example was the US Tri-state tornado of March 1925, which traveled more than 350 km across Missouri, Illinois and Indiana, killing 689 persons and devastating more than 20 towns and cities. Storm systems may also generate multiple funnels. In the so-called super outbreak of April, 1974 in the US, 147 tornadoes were recorded, touching down in 13 states, and killing 335 persons.

If increased heating of the atmosphere continues, it may lead to more, or stronger tornado activity, and a geographical expansion of the zones in which tornadoes are more frequent. It is difficult to judge what is happening on the basis of disaster trends, because of enormous variations in damage from tornadoes of the same size and intensity, and comparable damages from those of very different scale. More important are variations in land use and densities of human occupancy.

Winter Storms

The snow and ice storms of higher latitudes and elevations have not been as lethal or destructive as tropical cyclones. However, they are a major, and growing, source of disruption and economic loss in some of the wealthiest countries. Winter storms have become, above all, a threat to modern urban areas and infrastructure, especially to power, light and transportation services for cities.

Of particular concern are large area snowstorms, blizzards, and ice storms. *Blizzard* refers to a winter storm with winds in excess of about $56\,km\,h^{-1}$, temperatures less than $-10\,^{\circ}C$ and very low or zero visibility. Wind and temperature combine to create severe wind chill for humans and animals. There must be blowing snow involved, but actual snowfall may or may not be a large. Snow whipped up from the ground helps to cause severe snow drifting and whiteouts an extreme visibility hazard.

About half the world s land area is subject to snowfall. However, snow-related disasters and costs are overwhelmingly concentrated in major urbanized regions and, to a lesser but growing extent, in winter recreational areas. Closing of transportation routes in and between cities, typify what happens in most winter storm disasters. This is reflected in the exceptional number and share occurring in North America, where the media invariably report them as paralyzing cities.

Ice storms, also referred to as freezing rain or glaze storms, involve rain falling from relatively warm air aloft, into sub-zero air below. The raindrops may be super cooled, but remain liquid. When they fall on vegetation and exposed structures, they freeze instantly. Heavy coatings of ice on overhead wires can cause the wires to break, especially in windy conditions. Branches of trees are bent over or broken off by the weight of ice. Overhead wires, structures or vehicles on which they fall are damaged, and roads blocked. Coatings of ice on roads lead to traffic accidents or close highways.

It is likely that climatic changes will change the incidence, geography and, possibly, intensity of freezing rain events. The case study below indicates the potential of intensified or shifting patterns of winter storms, and the vulnerability to them of urban-industrial areas.

Case Study 2

The January, 1998 Ice Storm in Eastern Canada and North-eastern US (Soulard *et al.*, 1998; Phillips, 1998).

A succession of winter storm fronts moving across central and eastern Canada created exceptional conditions for the formation of freezing precipitation. In the worst-hit areas of South-eastern Ontario and southern Quebec, between 75 and 100 mm of freezing rain fell. Temperatures rarely rose above freezing at ground level throughout the episode. They dropped below $-10\,^{\circ}C$ for several days following the storms, greatly aggravating the difficulties of survivors and recovery operations. Damage also occurred in northern New York state, parts of New England and the Canadian Maritimes.

The area affected contained 18% of Canada s population and 56% of Quebec s, 19% of Canada s urban space, and 57% of Quebec s. Damages included thousands of kilometers of downed power lines, including collapse of hundreds of steel towers supporting high-tension lines. Over 4.5 million persons were without power for some time; more than 300 000 for over a week. Most highways, commuter trains, airports, telephone systems, newspaper deliveries and water filtration plants were shut down, and 2.8 million workers were unable to get to work in the worst period.

Some 25% of Canada s dairy cattle were endangered. Farms could not milk them without electricity. Most milk that was obtained could not be sent to market and was dumped. Over 5.5 million maple trees, supplying the maple syrup industry, were damaged or destroyed. Millions of other trees along city streets, in parks, wood lots and forested areas were damaged or destroyed.

About 100 000 people took refuge in emergency shelters, and 16 000 troops were mobilized for rescue work. Total economic losses were estimated at about 1.5 billion. The 35 fatalities (six in the USA) seem small, but the forms of death due to secondary hazards such as house fires, carbon monoxide poisoning from improper use of heating equipment, and hypothermia are typical dangers.

It was suggested that synoptic conditions over the North Atlantic, possibly relating to global warming or El Nino, helped intensify the unseasonal interaction of relatively warm, moist oceanic air, with cold, Arctic air. Freezing rain is common in the areas affected, which have a history of such disasters. However, most of the region has not recorded such quantities before, and the problem usually occurs in early spring. In the 1990s, however, Southern Canada experienced an unusual series of mid-winter warm spells and freezing rain episodes.

Floods

Floods are mainly a problem of riverside or coastal zones, the latter including lake and marine shores. A wide range of

generative conditions is involved. For rivers, heavy rainfall is the most common source and factors that govern rainfall-runoff relations, and flood routing in streams, have been a major focus of investigations. In higher latitude and altitude regions snowmelt or ablation of glacier ice can be an important source, tending to create flood-prone rivers.

In some cases, disasters involve natural impoundments or dam-burst floods. Natural obstructions include ice-jams of river ice in spring break-up; landslides, avalanches, lava flows, and vegetation, especially log jams. Inundation of land above the impoundments may be damaging, but the sudden failure of dams and resulting flood waves, can be highly destructive events (Schuster, 1986). The collapse of man-made dams is often brought about by natural extremes, and the behavior of resulting flood waves resembles those from natural dams.

Lakeshore floods can derive from contributing streams and may coincide with riverine floods. They also involve sieches oscillations of lake levels caused by storms or landslides. Marine coastal floods derive from exceptionally high tides, storm surges, and seismic sea waves or tsunamis (see herein). They may be magnified by wave action. These tend to be considered as part of the atmospheric or geologic hazards initiating them, yet their impacts are more like those of flood hazards.

Case Study 3

The Mississippi flooding of 1993, illustrating the dangers of intensified or changed patterns of rainfall-generated floods, interacting with more vulnerable land uses.

In 1993, floods occurred on the Mississippi and most of its major tributaries, forming the largest flood in 133 years. The main flood lasted 79 consecutive days, longer than the damaging 1971 flood, the previous record. The Mississippi at St Louis, Missouri was above flood stage from April 1 September 30, or 144 days. Some settlements were inundated for more than 160 days. Monthly mean discharges were above normal for the first nine months of 1993.

The generative conditions involved an unusual and persistent weather pattern related to jet stream position and high pressure in the northern mid-west United States. Blocked by the latter, humid air from the Gulf of Mexico produced sustained heavy precipitation. Precipitation was also above average for the first nine months of 1993, and twice the normal amounts for North Dakota, Kansas and Iowa. Moreover, an unusually cool 1992 had left water tables high and the ground in a saturated condition.

Damages affected a cluster of nine midwestern states, from North Dakota and Nebraska, to Missouri and Illinois. There were 48 deaths; 56 000 homes were destroyed or damaged and 26 000 people were evacuated. Direct losses were estimated at US$11 billion, while total losses, including indirect costs, such as lost income and production, were put at $20 billion. Some 3.2 million ha of cropland and 75 towns were completely inundated. More than 400 counties were declared disaster areas. River waters carrying, especially, sewage and agricultural chemicals contaminated water supplies. A two-month halt in Mississippi barge traffic affected a large fraction of national bulk transport, including coal, oil and grain. Inundation, erosion and sedimentation damage to roads was estimated at $500 million. There were serious damages to river control works, which led, in turn, to much of the other harm. Some 9300 km of levees were damaged, and 1083 out of 1576 levee systems broken or overtopped.

The 1993 flood is generally regarded as the most destructive flood on record for the region. However, while initiated by exceptional weather and hydrological conditions, the impacts were also attributed to dependency upon river engineering works (that failed) and land use changes. Over 80% of wetlands within the flood areas are said to have been drained since the 1940s, greatly reducing natural restraints on flooding. Meanwhile, as much as 70% of the homes and subdivisions inundated were part of urban expansion or other flood plain occupancy since the 1971 flood! This was despite past recognition of the importance of flood plain management to reduce exposure and vulnerability to floods.

The event illustrates how the combination of atmospheric extremes, land use change, failed control works, or otherwise inadequate adjustments to natural hazards, lead to disaster. This happened in a very wealthy country with, otherwise, outstanding achievements in water science and hazards research!

Droughts

Drought is, above all, a killer for agricultural crops and animals. It is increasingly also an economic hazard for urban industrial communities. Drought tends to affect human societies in substantially different ways from floods and in different phases of activity. Their link as hydrological extremes is largely incidental (Hewitt, 1997).

Droughts arise mainly from atmospheric conditions affecting wide territories. However, their effects on humans or other organisms can be felt in or through all forms of fresh water supply. Three main classes of drought hazard are recognized:

1. Meteorological drought, identified, especially, with lack of precipitation. It can also depend upon weather conditions that induce excessive evapotranspiration. The main atmospheric conditions tending to bring meteorological droughts are persistent high pressure systems with subsiding air; descending winds in the lee of mountain ranges; and air masses which, even if they contain abundant moisture, lack instabilities or cloud-, and precipitation-generating mechanisms. The world s desert regions involve one or more of these

as permanent conditions. They recur annually where there is seasonal drought. Most disasters arise where such conditions are abnormal, temporary, or of unusual duration.

2. Runoff drought, involving reduced stream flow, or drying up of surface waters including ponds and lakes. Though usually due to meteorological drought in the drainage basin or its headwaters, the impact is felt through dependence on surface waters. Low stream flow may be accompanied by declining water quality or contamination of the water that is available.
3. Aquifer drought involves loss of ground water sources due to declining water tables and drying up of springs and wells. Again, it tends to reflect meteorological drought, but may be exaggerated by human depletion of ground waters in regions or periods of drought.

Meanwhile, landforms, soils and vegetation cover are critical intervening variables that influence the moisture-conserving properties of sites. They can moderate or hasten the impact of dry spells. For this reason, the severity or the development of droughts involve fairly complex spatial patterns. The more severe consequences for agriculture also depend upon the drought-resistance of crops, cultivation practices, seasonal moisture demand and the possibilities for supplemental irrigation.

The special and severest association between drought and disaster concerns food security and famines. Most of the worst famines have been accompanied by drought and exacerbated, if not triggered, by drought-related crop and animal losses. However, not even the great, modern famines of the tropical world, or any since the early years of the 19th century, have been due to absolute food shortage in the food-producing or economic system, to which the victims belonged. Human actions, or inaction, can be decisive in whether food is provided to the needy, from areas of surplus.

Although both are treated as hydrological hazards, there are important differences between the development and impacts of drought and floods. Droughts develop slowly but may last weeks or months, even years (Smith, 1992, 246). Drought is often not recognized or declared until a dry spell has gone on longer than the time in which most floods arrive, peak and end. However, this drawn-out aspect can also provide opportunities for varied and geographically extensive human responses.

Floods are mostly linear or patchy in spatial extent, following water courses or coastal zones, and reflecting their topography. Severe droughts tend to embrace extensive regions. The areas most prone to moisture stress, or feeling its effects earlier, tend to be normally well-drained and drier areas. Those most prone to floods are poorly drained and flood plain areas usually the last to succumb to drought.

Earthquakes

Some 1.5 million people were killed in earthquakes in the 20th century, more than half after 1950. Perhaps 10 times as many suffered serious injury, and more than 100 million had their homes destroyed. The disasters follow directly from the occurrence of large earthquake events, mainly, but not exclusively, in regions with relatively high levels of seismicity.

The impacts of earthquakes on society depend on ground motion, especially so-called strong motion. This is influenced by near surface conditions, as well as the strength of, and distance from, the original seismic event. The local severity of earthquakes is modified by the bedrock and soil; by landforms and surface or groundwater conditions; and by vegetation cover. Seismic waves can be amplified by certain slope and landform conditions, at the interfaces of bedrock and unconsolidated materials, and in areas of deep alluvium. These are referred to as micro seismic conditions. Foundations may prove highly unstable in certain clay and sandy deposits, especially if wet. These are particular dangers of coastal and flood plain areas. Destructive landslides often occur in earthquakes (see herein).

In general, the more severe damages within disaster zones are associated with locally unfavorable choice of sites, and built structures made vulnerable by bad design. Most reflect a pattern of ill-advised occupation of sites that are hazardous because of microseismic conditions. This has become the critical factor of changing earthquake risk, notably through rapid urbanization. The importance of these relations between microseismic conditions and occupancy in earthquakes, are detailed in reports for the California, 1906, Kanto Plain, Japan, 1923, Alaska, 1964, Skopje, 1963, Niigata, 1964, Lomo Prieta, 1989 and Kobe 1995 disasters among others (Hewitt, 1997, 208 217).

A related concern, given the proliferation of megaprojects, is the evidence that large dams and fluids artificially injected deep into the ground can trigger earthquakes. A major disaster occurred around the Koyna Dam in India in 1967, directly related to the dam. A series of tremors around Denver, Colorado in the 1960s not very damaging, but disturbing for residents were traced to liquid wastes being pumped into a deep well (*see* Earthquakes Triggered by Human Activities, Volume 3).

Volcanic Hazards

Volcanic eruptions give rise to a variety of highly destructive forces and conditions (Fisher *et al.*, 1997). The type, scale, sequence and duration of eruptive episodes determine which forces occur, or can do the greatest harm. The main distinctions for disaster potential involve whether and how explosive the eruption is; composition of the ejected material and how this affects its behavior, and development of associated volcanic or constructional landforms.

Vulcanologists recognize 11 different types of eruption according to different combinations of these conditions (Blong, 1984; Chester, 1993).

The most lethal and destructive events involve explosive eruptions and pyroclastic (hot ash) flows. However, secondary hazards triggered by eruptions have together caused even more fatalities. There are volcano-induced tsunami and volcanic mudflows or lahars. The latter may be initiated by rapid melting of snow and ice on the summit area of the volcanic cone; where volcanic vents erupt under ice caps; or with a sudden outburst from a crater lake. Sometimes, lethal gas eruptions occur. The cloud of carbon dioxide emitted from the crater lake, Nyos, in Cameroon, August, 1986, killed over 1700 people by asphyxiation (McCall *et al.*, 1992).

Most volcanic eruptions have their main or only destructive impacts within a few kilometers or tens of kilometers of the volcano or eruptive vent. However, volcanic ash from the larger, explosive eruptions may be carried for thousands of kilometers, endangering transportation and reducing sunlight far away. Eruptions of exceptional explosive force and/or heat can loft large quantities of fine ash and smoke into the stratosphere where it can remain long enough to be transported around the globe. This can affect hemispheric or global climates over months, years, and, perhaps decades (Rampino and Self, 1982). The Tambora, Indonesia eruption of 1815, ejected some $35\,km^3$ of pyroclastic material three orders of magnitude larger than the Mt St Helens, 1980 eruption. Unseasonable cold, weak sunlight and summer frosts were reported in Europe and North America for the following two years. There were severe shortages of food and collapse of grape and other harvests (Post, 1977). This is by no means the largest known eruption in the geological record. A recurrence of such a relatively rare event, could drastically affect the pace, and possibly the direction, of climate change.

The more evident forms of increased risk from volcanoes relate to changing human exposure and vulnerability (Pflaker and Ericson, 1978). Many regions with active volcanoes show increasing concentrations of population, intensified land use or activities bringing more persons into areas threatened by volcanism. Often the volcano itself is an important resource and attraction, having exceptionally good soils and aquifers, or religious and aesthetic value. They are major tourist destinations, from the Caribbean to Japan, where volcanoes are one of the attractions. Included are active volcanic cones like Mt Etna in Sicily, whose winter snow cover supports a ski industry. Ever-growing investments in facilities and infrastructure help to attract growing numbers of visitors. Given that little or nothing can be done about eruptions when they occur, much improved monitoring and effective forecasts can save lives, but disastrous property losses are likely to continue increasing (Chester, 1993) (*see* Volcanoes and Cities, Volume 3; Volcanoes and the Environment, Volume 3).

Landslide Hazards

Landslides, resulting mainly from slope failure, are distinctive and widespread causes of disasters (Brabb and Harrod, 1989). Some of the more destructive ones involve catastrophic rockslides, debris flows, slump earth flows and mud flows. Avalanches are a concern in snow-covered mountains.

As noted above, the dangers and damages from landslides are underestimated because most of the larger examples or multiple failures tend to occur with other hazards such as floods and severe storms. The most destructive examples have occurred in conjunction with earthquakes. The larger part of the devastation, and some 180 000 deaths in the Kansu, China catastrophe of December, 1920, came from slope failures, and huge land slips in weak loess and alluvial deposits.

The May 1970 Peruvian earthquake triggered an avalanche of ice and entrained debris some 50 100 million m^3 in volume. Descending from Mount Nevados Huascaran (6665 m) in the Cordillera Blanca, it traveled 14.5 km and attained speeds in excess of $300\,km\,h^{-1}$. A tongue of the debris buried and totally destroyed the town of Yungay. Some 18 000 people were killed. Thousands of other landslides occurred and, especially coastal settlements, destructive ground failure occurred by subsidence, fissuring and liquefying. Together, these secondary effects were responsible for most losses over an area of some $103\,000\,km^2$ and the most lethal earthquake in the Western Hemisphere.

Deforestation and other forms of land use change or ecological abuse, increase the scale and likelihood of slope failures. Artificially excavated slopes along roads and other construction projects can, and generally do, increase the occurrence of destructive slope failures. Human exposure to, and damages from, landslides increase with the spread of buildings, facilities and activities over steep slope areas, especially in mountains and cliffed coastal areas. As indicated, mass movement dangers are likely to increase with intensified, changing geographies of atmospheric and hydrological hazards.

Tsunamis

Tsunami is a Japanese word meaning (great) harbor wave. These are the largest and most destructive water waves on Earth. They are usually triggered by undersea earthquakes, sometimes volcanic eruptions, occasionally great landslides or volcanic materials plunging into bodies of water. They become hazards where the waves strike and inundate coastal areas, thus combining aspects of hydrological and geological dangers (*see* Tsunamis, Causes and Consequences, Volume 1).

Most tsunami disasters occur along coasts within a few tens of kilometers of the undersea earthquake producing them. However, some result from waves that travel thousands of kilometers before striking mid ocean islands or distant coastlines. There have been some notable examples around the Pacific Ocean.

Tsunamis travel at speeds one or two orders of magnitude faster than wind-generated waves. In the deep ocean this may be from $500\,\mathrm{km\,h^{-1}}$ to perhaps $900\,\mathrm{km\,h^{-1}}$ over the deepest waters (Abbott, 1999). However, here they cause little or no danger, having wavelengths measured in tens or hundreds of kilometers. Their small rise and fall can go unnoticed by shipping. When they reach a shelving coast, however, or enter bays and inlets, they can translate into huge roll waves many meters high.

The arrival of a tsunami may be preceded by a sudden retreat of the sea. Sometimes people rush down to the shore to see what has happened, only to be engulfed by the returning great wave. Moreover, several waves are commonly triggered by the same event, arriving separately on a distant coast. In this case too, thinking that the danger is over after the first one, people are caught in a later one, which may also be the largest.

Some of the most lethal incidents of these kinds followed the Chilean earthquake of May 1960. With one of the largest magnitudes ever recorded, the undersea earthquake generated a series of tsunamis that swept back and forth across the Pacific for some days. In Chile, they killed about 1000 persons, mainly in the third and largest wave, when people had run down to the shore to try to rescue items swept up in the first two waves. In Hawaii, 61 people were drowned and 282 injured, and 165 persons died in Japan caught out, even though warnings were issued and relatively accurate predictions were made of arrival times.

The tsunami hazard itself is likely to become more severe with global environmental change, but not because there are any indications of increased, or stronger seismic and other triggers. Sea level rise can change the scale and scope of these events where they come ashore. Recall that a few centimeters of rise can actually mean huge volumes of water closer to shore, and available for incorporation in the wave.

This is also an example of how change in one sphere can alter dangers from another. Global warming leads to thermal expansion of ocean waters, and reduction of the Earth s glacier and permafrost, which adds water to the oceans. As a result, a hazard whose own geophysical generative conditions may not change is magnified. Meanwhile, ever-greater human occupancy of coastal zones is increasing human exposure. In many regions, the range of more vulnerable persons, activities and items near shorelines, is also growing.

Biomass or Wild Fires

Wild fires are out-of-control burnings, fed by forest, grassland and peat land whether started by natural or human agency. These fires have been a recurring natural phenomenon of regions that combine wet and dry seasons, or occasional drought in an otherwise cool and/or humid region. Dry conditions are generally needed to set the stage for large and extensive mass fires. However, humid conditions are required to produce sufficient biomass to feed the fires.

Especially susceptible have been forest and brush lands in mediterranean and sub-mediterranean climates winter rains, summer drought savanna and parkland in sub-humid tropical climates; steppic grasslands, and monsoonal areas with a marked dry season. Continental interiors with cool, snowy winters and droughty summers are also subject to recurring, large fires. The changeable weather and windy conditions of sub-humid and even humid mountain environments can also lead to serious fire hazards.

In addition to broad climatic variations in plant growth, and types of vegetation, the length of time needed to accumulate substantial biomass in particular, dead, dry, inflammable material can differ markedly between grassland, brush and forest ecosystems. Forests usually take much longer to regenerate from past, large fires than grassland, while brush will tend to lie somewhere in between.

Fire weather is more than just a matter of dry conditions. Lightning starts most wildfires. Windiness is an important, aggravating factor in the more serious burns and conflagrations. Strong winds aid combustion and the movement of fire. They make fire-fighting much more difficult and hazardous. The worst fires generate their own winds, the worst being the firestorm, a meteorological phenomena approaching tornado-like winds and temperatures exceeding $1000\,^{\circ}\mathrm{C}$.

Meanwhile, burnt-over areas become more susceptible to soil loss by water and wind erosion. In seasonally humid climates, when tree and grass covers are lost, the returning rains are likely to generate more, or worse floods, landslides or mud and debris flows.

It is widely thought that global warming will increase the forest fire hazard in mid-latitude and boreal forests. The problem is thought to be particularly severe for the great forestlands of Russia, Canada, Alaska, Scandinavia and parts of Southern Argentina and Chile. Competitive extractive industry practices, and large-scale insect infestations, may continue to aggravate the fire hazard and effects of climate change. In the tropical lowlands and highlands even areas that are humid for much of the year environmental abuse has rendered surviving forest areas much more susceptible to fire. There are very few forest areas left that have not been opened up to greater fire risks from human activity, including deliberate burning.

In many contexts, fire is a traditional technique of land clearance or pest control, and for slash and burn agriculture. This has often seemed sustainable with careful use in experienced hands. However, excessive, careless or deliberately destructive use of fire can be harmful. Combined with drought, it can lead to immense, out-of-control burnings, as in Central Africa and Amazonia in recent decades, and the regional fires in Indonesia, 1997 and 1998 (see herein).

Conversely, artificial fire suppression, to protect forested areas, may also be ill advised. In the end, it can increase the hazard or ultimate scale of fires by aiding the build-up of flammable biomass. This was seen as an important factor in the devastating forest fires in Yellowstone National Park, USA, in 1988 (Romme and Despain, 1989).

Thus, given the importance of weather and biomass conditions in the wild fire hazard generally, global climatic and land use changes seem likely to make the problem worse. Recent increases in numbers, or more extensive fires, have accelerated the destruction of forest ecosystems and forest resources. The fires have also become major regional health and economic dangers due to smoke, haze, and smog. Meanwhile, there is a positive feedback danger as large fires contribute to greenhouse gases. The conditions responsible suggest these hazards will continue to grow without a, hitherto unprecedented, international commitment to reversing the human practices that aggravate them.

Case Study 4 – The Indonesia Wild Fires, 1997

A total of some 5 million hectares burned in Indonesia, mainly in East Kalimantang (part of the island of Borneo), and Sumatra. Unusually dry weather was a major factor in the timing, scale and uncontrolled spread of the fires, and related to a strong El Nino event. The setting on fire of extensive peat lands was a grave concern, since they are likely to burn underground for years or decades, and flare up whenever dry conditions return. Meanwhile, the smoke and haze from the fires spread over much of South East Asia. Reduced visibility and yellow smog affected many areas for two to three months.

Damages included large losses to the forest, plantation, agricultural, and tourism sectors; enormous health threats from smoke and haze; evacuations to escape fire and smoke, and magnification of already severe ecological damage and loss of biodiversity. Direct losses in Indonesia were around US$2.2 billion, mainly in health care and loss to the timber, agricultural and tourism sectors (EEPSEA/WWF, 2000). Malaysia and Singapore lost more than US$0.6 billion, mainly in health costs and lost tourism revenues. Some argue that the real losses were much larger in Indonesia through long-term health costs, loss of indirect benefits of forests such as control of erosion and floods, lost biodiversity, loss of business confidence and withdrawal of foreign investment. This also came at a time of acute national economic and political crisis, and expanding social conflicts.

Fire is commonly used to clear land for agriculture, plantations and other developments. However, letting fires burn out of control has become much more common, as a way to cover up illegal logging; for plantation developers to clear land and force out small peasant farmers and indigenous peoples and for the latter to retaliate. The scale of the fires, including those in the peat formations of Kalimantang, also reflects two decades of massive forest clearance, vast plantation and resettlement schemes, illegal logging, and unchecked ecological abuse (Saharjo, 2000). In fact, about half the burnt area was actually in plantation and agricultural areas, and 20% or about 1 million hectares, forested land. However, much of the plantation and agricultural land, and some 30% classed as unproductive, represent recent clearance of forests.

Other reports suggest that 1997 was one of the worst years for fires elsewhere in tropical lowland forests. Large numbers of major, uncontrollable fires occurred in Amazonia, Central and West Africa as well as SE Asia. Droughts or higher temperatures were factors. However, in all areas, human activities were implicated in the scale and scope of the fires especially fire use itself, past forest clearance, land use changes and conflicts over land. Perhaps such wild fires need to be identified as a complex hazard, rather than a natural, let alone merely a biological, one.

CONCLUDING REMARKS

In the context of global environmental change, natural hazards or risks involving them, may increase due to one or more of the following:

1. intensification, or higher magnitudes of given natural forces and events;
2. higher frequencies or recurrence of given dangerous conditions;
3. combinations of adverse conditions, that become more frequent, prolonged or intense;
4. changing geographies of one or all the above.

Moreover, with or without any of these, life could become more hazardous, due to:

1. human modifications in the development or intensity of natural forces, whether intended or not, e.g., the roles of deforestation or urbanization on the frequency and scale of floods;
2. interaction of artificial processes with natural ones to create complex, perhaps novel, dangers, e.g., photochemical smogs combining weather conditions and

industrial pollution; or failures of engineered river control works in floods;
3. changing human exposure and/or vulnerability to some or all natural hazards.

Also, we need to pay particular attention to the fact that life could become less hazardous with:

1. (effective) hazard and vulnerability mitigation, including improved forecasting, defenses against hazards, risk-averting land use planning and design of the built environment, adaptive strategies, and assistance, especially, to those people and activities that are at greatest risk;
2. improvements in emergency preparedness, humanitarian assistance in, and recovery from, the disasters that will continue to occur.

On-going changes in human land uses and other activities not only modify natural processes, but also alter the ways in which humans are put at risk from age-old hazards (Blaikie *et al.*, 1993; Brookfield, 1999). We noted how snowfall hazards in North America, drought and famines in Africa, lethal storm surges around the Bay of Bengal, and wild fires, have been transformed by land use and societal changes. Broad, adverse trends such as desertification, deforestation, reduction of perennial snow and ice covers or sea level rise alter the danger from a range of other natural hazards. Often, it is difficult to decide which of the above, or which combinations, are more critical.

Techno-scientific knowledge and practice, now confronts the sense in which dangers from natural processes are as much manufactured by human activity in the language of Ulrich Beck (1999) and others as givens of the planetary setting. This is partly about environment-changing or high-risk technologies, and increased or changing endangerment of human populations. However, there are also the expectations of, and claims made for, scientific understanding, estimation and management of risks.

The need for detailed and reliable knowledge of environmental conditions has never been more urgent. But geophysical understanding has, increasingly, to be employed in creative, and interdisciplinary contexts. Not least among the complex demands now being made, are those requiring geophysical and techno-scientific work, as it bears upon human safety, to be developed in ways that are more accountable and transparent to society at large.

Forecasting how global environmental change will affect disaster incidence, or which hazards and damages should be of greatest concern, is complicated by uncertainties about the future of natural processes. But even more uncertainty surrounds human developments, whether from novel technologies and their adoption, or conflicts including those that turn violent and undermine response capabilities. Most obviously, recent struggles concerning whether, how, and how far, governments, the corporate sector, and particular regions of the world, will address climate change, have large implications for research into, and responses to, related natural hazards.

REFERENCES

Abbot, P L (1999) *Natural Disasters*, 2nd edition, WCB/McGraw-Hill, Boston, MA.

Alexander, D (1993) *Natural Disasters*, Chapman and Hall, New York.

Beck, U (1999) *World Risk Society*, Polity Press, Blackwells, Oxford.

Blaikie, P, Cannon, T, Davis, I, and Wisner, B (1994) *At Risk: Natural Hazards, People Vulnerability, and Disasters*, Routledge, London.

Blong, R J (1984) *Volcanic Hazards: a Source Book on the Effects of Eruptions*, Australia Academic Press, Sydney.

Brabb, E E and Harrod, B L, eds (1989) *Landslides: Extent and Economic Signi cance*, A A Balkema, Rotterdam.

Brookfield, H (1999) Environmental Damage: Distinguishing Human from Geophysical Causes, *Environ. Hazards: Human Policy Dimensions*, 1/1, 3 12.

Chester, D (1993) *Volcanoes and Society*, Edward Arnold, London.

Church, C, Burgess, D, Doswell, C, and Davies-Jones, R, eds (1993) *The Tornado: its Structure, Dynamics, Prediction and Hazards*, Geophysical Monograph 79, American Geophysical Union, Washington, DC.

Dynesius, M and Nilsson, C (1994) Fragmentation and Flow Regulation of River Systems in the Northern Third of the World, *Science*, 266, November, 753 761.

EEPSEA/WWF (2000) *The Indonesia Fires and Haze of 1997: the Economic Toll*, Economy and Environment Program for SE Asia and World Wildlife Fund for Nature, May 29, 1998, http//www.geocities.com/rainforest/2701/eepsea3.html.

Hewitt, K (1989) Human Society as a Geological Agent, in *Quaternary Geology of Canada and Greenland*, ed R J Fulton, Geological Survey of Canada/Geological Society of America (The Geology of North America, Vol. K-7), 619 634.

Hewitt, K (1997) *Regions of Risk: a Geographical Introduction to Disasters*, Longmans, London.

IFRCRCS (1997) *World Disasters Report, 1997*, International Federation of Red Cross and Red Crescent Societies, Oxford University Press, London.

Mallin, M A (2000) Impacts of Industrial Animal Production on Rivers and Estuaries, *Am. Sci.*, 88/1, 26 37.

McCall, G J H, Laming, D J C, and Scott, S C, eds (1992) *Geohazards: Natural and Man-made*, Chapman and Hall, New York.

Pflaker, G and Ericson, G E (1978) Nevados Huascaran Avalanches, Peru, in *Rockslides and Avalanches 1 Natural Phenomena*, ed B Voight, Elsevier, NY, Chapter 8, 277 314.

Phillips, D (1998) The Ice Storm, *Can. Geog.*, March/April, 36 37.
Pielke, Jr, R A and Pielke, Sr, R A (1997) *Hurricanes: Their Nature, and Impacts on Society*, John Wiley & Sons, New York.
Post, J D (1977) *The Last Great Subsistence Crisis in the West*, Johns Hopkins University Press, Baltimore, MD.
Quarantelli, E L, ed (1998) *What is a Disaster? Perspectives on the Question*, Routledge, London.
Romme, W H and Despain, D G (1989) The Yellowstone Fires, *Sci. Am.*, 261 (November), 37 46.
Saharjo, B H (2000) *The Role of Human Activities in Indonesian Forest Fire Problems, Laboratory of Forest Protection, West Java, Indonesia*, http//www.wwf.com/forests.
Schuster, R L, ed (1986) *Landslide Dams: Processes, Risk, and Mitigation*, Geotechnical Special Publication No 3. American Society of Civil Engineers, New York.
Soulard, F, Trant, D, Filoso, J, and Van Wesenbeeck, P (1998) Ice Storm! *Can. Social Trends*, 11-008 Winter, 21 23.
Stone, M and Droppo, I G (1994) In-channel Surficial Fine-grained Sediment Laminae. Part ii: Chemical; Characteristics and Implications for Contaminant Transport in Fluvial Systems, *Hydrol. Process.*, 8, 113 124.
UN/DHA (1994) Situation Reports 1 4.

Nigeria Oil Industry – Case Study

see Oil Industry: Cultural Impacts and Environmental Risk Tolerance (**Volume 3**)

Nitrate Leaching

Leaching is the slow downward percolation of water through the soil, which results in the removal of water-soluble minerals such as nitrate, from the upper soil horizons, and their accumulation in surface or groundwater. Nitrate, the oxidized form of nitrogen, is the end-point of the two-stage oxidation (nitrification) process mediated by chemoautotrophic bacteria. Nitrate is highly oxidized, soluble and biologically available. Thus nitrate leaching leads to a deterioration of drinking water quality and undesirable effects on ecosystems.

A L HEATHWAITE *UK*

Nitrate Protection Zones in Europe

Hadrian F Cook
Imperial College at Wye, Ashford, UK

High levels of nitrate contamination occur in parts of Europe in both surface water bodies and groundwater. There is concern where affected waters are destined for human consumption, and nitrate also contributes to eutrophication of lowland rivers, lakes and reservoirs. This article describes the de nition and operation of nitrate protection zones, designed to reduce nitrate loading of natural waters at source. This is through land use prescriptions and the management of inorganic and organic fertilizers where the origin lies in agriculture. In addition, nitrate protection policies increasingly consider nitrates from non-agricultural sources. Also discussed are uncertainties associated with scenarios of climate change. The European Union (EU) is the main focus with some discussion of the issues for Central and Eastern Europe.

A *nitrate protection zone* may be defined as a land area coterminous with the whole or part of: a surface water catchment, a groundwater capture zone around a specific source, or a vulnerable aquifer. Within this area, land management or other activities are constrained with the stated aim of controlling nitrate loading of waters. Generically, nitrate protection zones are examples of contaminant-specific water protection zones. Within the definition are included specific environmental policy schemes, such as *nitrate sensitive areas* (NSAs) and *nitrate vulnerable zones* (NVZs) (both described in this article), which are concerned with the implementation of detailed policy prescriptions for the reduction of nitrate contamination of waters from agricultural sources (*see* Nitrogen: Agricultural Uses and their Impacts, Volume 3).

The nitrate anion (NO_3^-) arises from the dissociation in water of nitrates, including those of calcium, magnesium and potassium. Nitrate in soils is essential for supplying nitrogen (N) for the synthesis of protein in plants; in waters, nitrate can present problems for both human health and the environment.

Nitrate ingestion in humans has variously been linked to stomach cancer, methaemoglobinaemia (or blue baby syndrome) and childhood diabetes. Medical scientists remain uncertain about the true risk to health, and epidemiological studies are inconclusive. There is indeed some evidence for a role for nitrate in preventing heart disease, controlling skin infections and killing certain gastrointestinal pathogens.

However, loading of the environment with nitrate, especially as the problem affects lowland aquatic ecosystems, remains a more tangible problem. Combined with phosphorus, nitrate in waters is frequently associated with eutrophication of aquatic ecosystems causing blooms of green and blue-green algae in reservoirs (the latter sometimes toxic) during the summer.

Figure 1 shows a simplified *nitrogen cycle* (*see* Nitrogen Cycle, Volume 2). In agricultural systems, nitrogen is lost through nitrate leaching from soils to water, ammonia volatilization and nitrous oxide (N_2O) and dinitrogen (N_2) losses by denitrification in soils and groundwater. N_2O is estimated to contribute around 5% of greenhouse gas warming potential in the atmosphere. Soil nitrogen (of many species) is boosted through regular inorganic and organic fertilizer application. Microbial activity is responsible for the release, in the first instance, of ammonium (NH_4^+). This arises from the ammonification or mineralization of organic matter, followed by nitrification, i.e., ammonium oxidation to nitrate. Nitrate ions are highly soluble, negatively charged (as is the bulk of charges in the soil matrix) and therefore prone to leaching; i.e., removal from soils by the passage of water. However, the proportion of soil nitrate leached directly from fertilizer applied is normally small, and considered to be controllable by careful application (*see* Nitrogen: Agricultural Uses and their Impacts, Volume 3).

Other nitrogen sources supplying nitrogen to soils and the environment include atmospheric fixation by leguminous plants, farmyard washings, slurries and silage deposition from the atmosphere (dry and wet) and sewerage, including leaking sewers and cesspools. Non-agricultural sources of nitrogen in waters are variously assessed as between 8 40%. In Southern England, nitrogen loading from the atmosphere, as ammonium, nitric acid and other forms occurs in the range 25 50 kg N ha^{-1} $year^{-1}$; comparable levels have been detected in the Netherlands.

Raised nitrate concentrations in the chalk ground water of Thanet (East Kent, England) have been detected since the middle decades of the 19th century, and were originally attributed to sewage arising from urbanization, although intensive horticulture and agriculture have long been the favored land uses. The chalk aquifer here is largely unconfined, meaning that no geological protection is afforded from overlying strata, and any leachate from the soil, or discharges of contaminated water, percolates vertically to the water table. Rising concentrations in the 1950s and 1960s in chalk sources across Southern and Eastern England were detected at a sufficient rate to cause concern; in the 1970s the link was established with the application of inorganic fertilizers in arable agriculture. However, nitrates in both groundwater and surface waters present problems across Europe.

For example, Figure 2 shows potential groundwater hotspots for nitrate contamination that were identified from model calculations by the European Environment Agency (EEA). Large areas of central and western Europe are potentially affected and, due to excess groundwater recharge from irrigation, areas of Southern Spain and the Southern Russian Federation may have a worse problem than Figure 2 predicts. Assessment of the real nitrate problem for groundwater across Europe is more difficult because data coverage, standards and sampling regimes all vary. However, EEA identifies actual hotspots within the territories of: Austria, Denmark, France, the Netherlands, Slovenia, Belgium, Finland, Switzerland, Spain, Greece, Portugal, Romania, Republic of Moldova, Hungary, Lithuania, Bulgaria, and Poland. To this list, the UK must be added. Overall, the problem is worst in the intensively farmed countries of North and Western Europe and has become an area of policy development by the EU, formally the European Community (EC).

Rising concentrations in natural waters caused a maximum admissible limit to be set at 50 mg nitrate/liter (11.3 mg nitrate-as N/liter) in the EC Drinking Water Directive 80/778/EEC. The guide level for human consumption is 25 mg nitrate $liter^{-1}$. Other relevant EC directives include the EC Urban Waste Water Treatment Directive (91/271) covering sewage effluent, and the EC Groundwater Directive (80/68) that requires protection against hazardous substances in groundwater, but does not cover diffuse pollution. The EC Nitrate Directive 91/676/EEC set the 50 mg nitrate $liter^{-1}$ standard for all waters vulnerable from agricultural pollution and requires member states to designate NVZs within their territories. Eventually, the quality of all waters may be protected by an EU-wide water framework directive.

Nitrate can be stripped from water by anion exchange systems that pass water through resin. This is increasingly common, but expensive. Alternatively, reactor systems are used; these utilize the natural process of microbial denitrification (the reduction of NO_3^- to the gases N_2O and N_2) through the use of a carbon source in anaerobic situations. Blending high and low-level water sources for public supply is another option, but only as a temporary measure. On Thanet, waters from an ion exchange denitrification plant are blended with high nitrate sources to reduce concentrations. However, stripping and blending relate to water put into public supply and cannot improve the status of natural waters *in situ*.

Although neither the sources nor the behavior of nitrate mobile in waters is open to simple description, many agree that control of the problem at source is the best sustainable option for reducing environmental loading. Point sources of nitrogen contamination are readily controllable by codes of good agricultural practice set by governments; however, diffuse nitrogen loading at source has

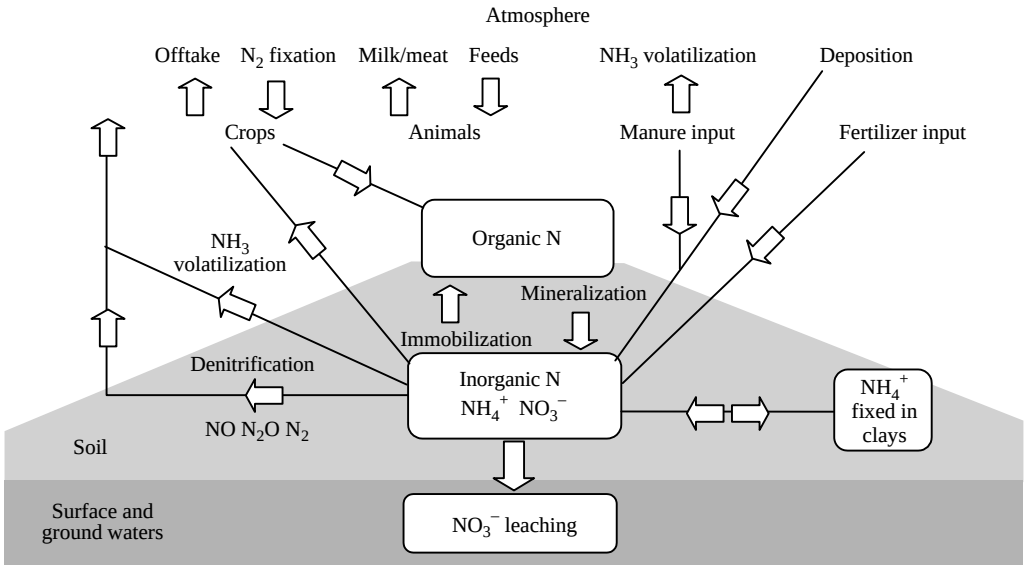


Figure 1 A simplified nitrogen cycle

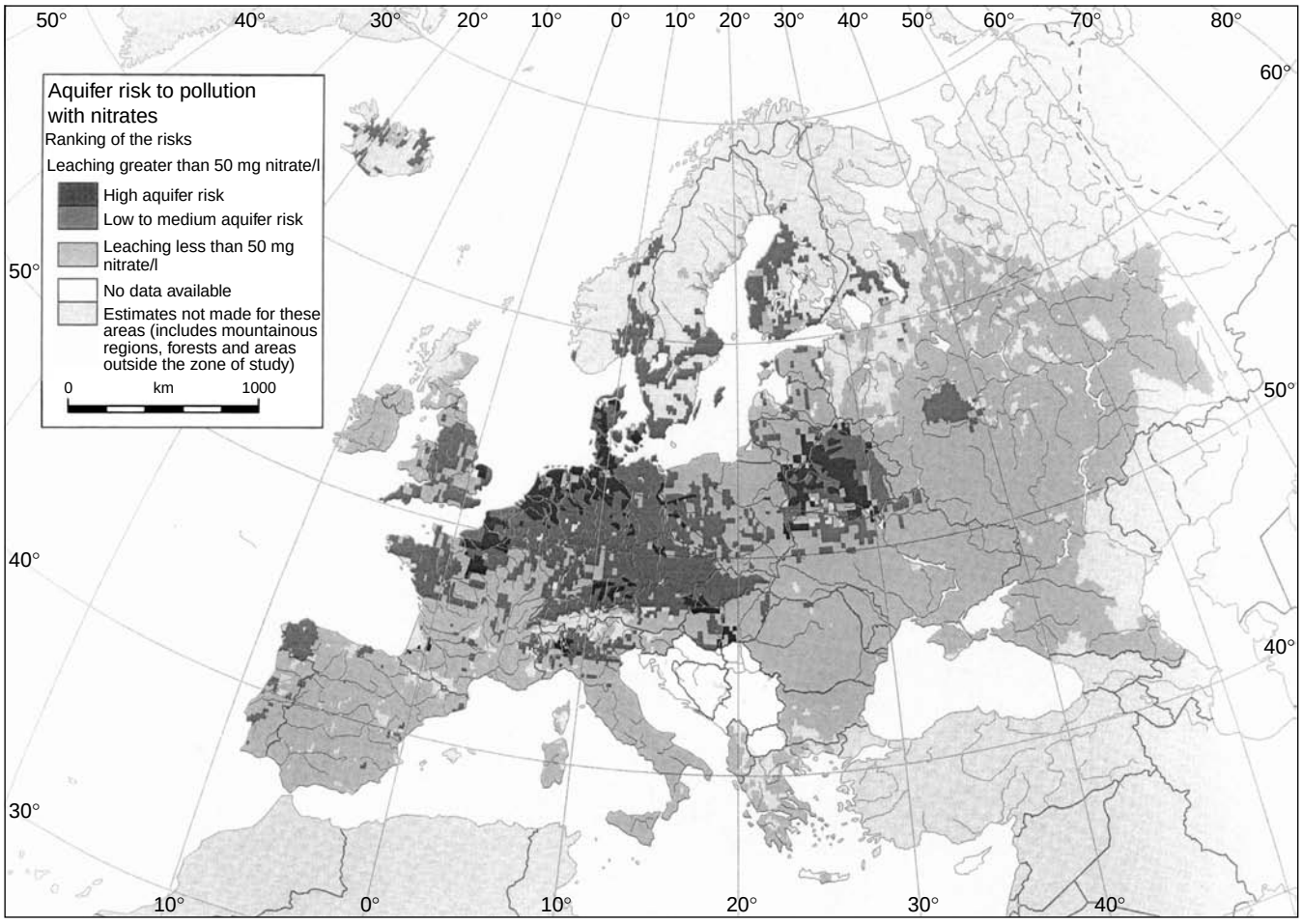


Figure 2 Potential nitrate ''hotspots'' for groundwater contamination (from Stanners and Bowdeau, 1995)

to be via land use measures. While agricultural yield is compromised, the reduction in cost of nitrate fertilizer is reduced.

EU member states have responded differently in implementing protection zone policies at the national level, depending largely on the severity of the nitrate problem within their territories. The very notion of nitrate protection zones, however, requires elaboration through the following points.

First, a nitrate protection zone suggests an uneven environmental response to what may be adverse and unsustainable agricultural practices within the catchment area for a water source. If these areas are in some way protected from the worst excesses, there remains the possibility of agricultural intensification elsewhere resulting in an expansion of any attendant problems. Nonetheless, there is a compelling argument in terms of economic efficiency for the targeting of land areas for environmental protection where farmer compensation and policing by environmental institutions is concerned.

Second, therefore, there is a tendency to protect only water sources. Ignoring the wider environment results in calls for widespread protection measures across whole aquifers and surface catchment areas.

Third, those concerned with whole-farming systems would favor the widespread adoption of measures deemed less damaging than conventional agriculture involving high agrochemical input systems and regular plowing, as in short-cycle crop rotations. Alternative systems have to be closely scrutinized, but as yet no clear benefits have been found; for example, poor management of organic farming systems produces little or no reduction in nitrogen loss.

Fourth, maximum use of available environmental information should be employed. For example, the absence of detailed soil mapping information in targeting for nitrate protection zones is an omission in Britain, yet it is known that leaching from shallow and lighter textured topsoil is great at certain times of the year.

What emerges is a convenient differentiation between source protection (aimed to protect individual boreholes, wells, springs and surface water abstraction points), and resource protection whereby larger-scale, hydrographic features may be protected, such as the entire unconfined portion of an aquifer or a surface water catchment. Some EU countries, especially the Netherlands, Germany, Denmark and parts of Belgium, have a serious nitrogen management problem due to a high density of livestock. In effect, Germany, The Netherlands and Denmark have designated their entire territories NVZs in an effort to achieve resource protection for all waters, while the UK, Belgium and France have preferred the targeted approach. In the case of France, however, the problem in certain regions is so great that the policy response has been for large areas, including the whole of Brittany, to be designated as vulnerable.

Land use policy responses to the drinking water and nitrate directives employ measures aimed at reducing nitrate losses below the soil (or horizontally into surface waters) through land use prescriptions. This is to be achieved by restricting the amounts and timing of organic and inorganic fertilizer applications, planting catch (crops that absorb nitrogen) crops, establishing low input grass on arable land and avoiding the plowing of grassland to avoid diffuse nitrate pollution. In addition, there is to be careful management of slurries and manures, according to codes of good practice, so as to avoid point source pollution.

Other effective alternative land use measures for the protection of surface waters are buffer strips and wetland areas. Nitrate that is horizontally mobile in shallow groundwater in alluvial subsoil is attenuated by the emplacement of fertilizer-free grass buffer strips parallel to watercourses, provided there is no under-drainage. The strips may be ≥ 5 m wide, and are an option promoted by several European agricultural and conservation agencies. Another approach is the use of existing, creation of new, or re-instatement of former wetland areas where relatively immobile surface waters can reduce nitrate loading in the environment.

The UK response has been to target geographical areas, and the first were NSAs under the Drinking Water Directive (*see* Nitrogen: Agricultural Uses and their Impacts, Volume 3). Since 1994, the number of NSAs was expanded to 32, incorporating 10 pilot areas (around 10 000 ha of agricultural land), initiated under the Water Act 1989. These were chosen to typify the geology and land management in areas where nitrate concentrations in abstracted groundwater was high.

In NSAs, the farmer is compensated for a supposed loss of earnings caused by a reduction in fertilizer inputs, switch of crops and even a switch to unfertilized, low-density grazing grassland under the arable premium scheme. There was an initial uptake of 86% of the agricultural land areas designated in 1991; longer-term enrollment was around 70%.

Designed to reduce the residual nitrogen in soils at the end of the growing season in autumn and winter, and to maximize nitrogen uptake during this period, the NSA scheme is a long-term experiment in land targeting, expected to run until 2003. The results are proving successful, especially where cover crops have been established before spring sowing. Under the arable premium scheme, concentrations leaving the soil at 5 mg liter^{-1} nitrate, or even less have been recorded. The basic scheme, on the other hand, provided for arable farming to continue with compensation for lower outputs in terms of yield. It has led to a variable response but around a 30% reduction in

leached nitrogen. This is attributed to cover crops and catch crops that absorb the nitrogen during autumn. Livestock farmers were asked to produce a manure plan; thereby providing some control on the storage and disposal of manure to land. There is a closed period for application depending upon soil type and crop.

Concentration of nitrate in groundwater is slower to respond than in soil water, mainly because of the longer residence time of water in the unsaturated zone of aquifers. Although certain early results are encouraging, it remains highly probable that wider-scale restrictions on land management, including fertilizer use and plowing, will be required if there is to be a widespread reduction in nitrate concentrations below the EU limit for water leaving the subsoil in agricultural areas.

A wider-ranging policy instrument in UK has involved the designation of NVZs, covering some 600 000 ha across England and Wales, with one example in Scotland in an area of intensive horticultural production (*see* Nitrogen: Agricultural Uses and their Impacts, Volume 3). NVZs derive from the nitrate directive and are concerned specifically with environmental protection from nitrates, rather than only the quality of drinking water destined for human consumption. Member states are required to identify water bodies where the EU limit for nitrates is actually or potentially exceeded, and compulsory measures must be taken to reduce pollution upstream of the sampling points.

The largest areas of designation protect surface waters, examples being the River Waveney (East Anglia), the Rivers Blackwater and Chelmer (Essex) and Rivers Leam, Cherwell and Great Ouse (a large area between Coventry and Dunstable) although groundwater is also protected, including Thanet in Kent. All the NSA areas are nested within NVZs. Compulsory NVZ measures are not always so stringent as in voluntary, compensated NSA policies. The restriction on fertilizer use is, in reality, not to exceed the economic recommendation. However, for manures good practice potentially involves additional storage and spreading over crops for some farmers.

To summarize the key NVZ rules:

- Nitrogen fertilizer applied must be according to recommended crop requirements on arable land and grass. Manures are limited to 210 kg ha^{-1} arable (to become 170 kg ha^{-1} in 2002) total nitrogen and 250 kg ha^{-1} total nitrogen to grassland.
- Closed periods for the application of nitrogen fertilizers are set at 1 September 1 February for arable land and 15 September 1 February for grassland. For organic manures on sandy and shallow soils, the closed period is 1 August 1 November for arable land, and 1 September 1 November for grassland.
- There are controls on manure-spreading during periods of waterlogged, flooded, snow covered or frozen soil and on steep slopes.
- Spreading of all fertilizers should be even and accurate.
- There are restrictions on fertilizer applications near watercourses and wells.
- There has to be adequate slurry storage.
- Good record-keeping of fertilizer applications is required.

Protection zones for surface catchments are relatively easily defined and the stretches of channel involved are easily identified. However, groundwater protection zones are more complicated. In simple terms, catchment areas from which boreholes draw water are derived from modeling subsurface flow at the water table using variable parameters. The main variables used are transmissivity, hydraulic conductivity, effective porosity and recharge. Calibration of the model is important and this is achieved by comparing output of the model runs with real water level data. By varying parameter values and their combinations, it is possible to derive various realizations of the catchments, and the NVZ boundary is based upon the line that represents the most realistic local condition.

Figure 3 shows the NVZ area designated at Moulton in East Anglia. Modeling boundaries are translated into hard boundaries in the landscape, typically field boundaries. Potentially, the final area included in a groundwater

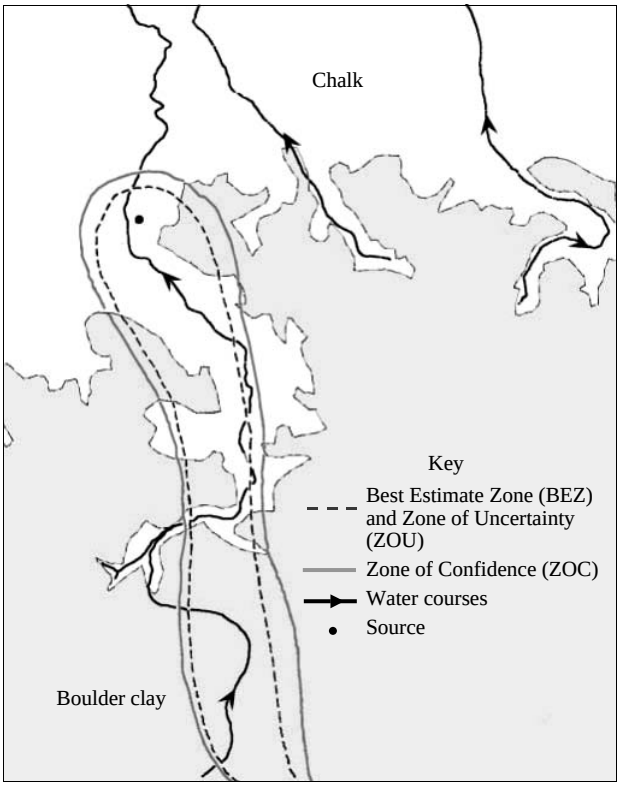


Figure 3 The Nitrate Vulnerable Zone designated at Moulton in East Anglia, UK. (Reproduced by permission of the Environment Agency for England and Wales)

NVZ may be a combination of surface water and groundwater protection zones, this being dependent upon local geology.

While the UK provides some interesting examples of policy development and implementation, it is by no means the most radical in approach. For example, Danish agriculture faces severe problems and is arguably the most regulated in Europe. Blanket restrictions under the nitrate directive have been brought about due to pollution problems affecting groundwater (the major sources of drinking water) and marine coastal waters. Around 62% of the land area is farmed, and nitrate pollution is largely attributed to the livestock sector, a problem apparent since the 1980s. Regulation of nutrients is by action plans aimed to reduce not only nitrogen pollution, but also that from phosphorus.

Disposal of manure is therefore key to the operation of successful nitrogen management, and spreading is restricted to periods of crop take-up and when risk of run-off is low. For applications between harvest and 1 February, the quantity applied must not exceed 50 kg N ha^{-1}. Danish farmers are required to prepare crop rotation and fertilizer accounts. Fertilizer must be reduced by an amount equal to the readily available manure nitrogen, i.e., assuming no losses to volatilization or leaching. For example, in the fertilization year 2001/2002 a requirement of 60% of nitrogen supplied from pig slurry or 55% cattle slurry must be used. A percentage adjustment is made in this way for each year in order to account for the residual effect of the particular organic fertilizer applied. While the policy is mostly concerned with livestock manure management, maximum available nitrogen (all sources) is set for arable crops, taking account of soil texture.

In Hungary, outside the jurisdiction of the EU, water protection zone policy is developing apace. An ambitious program was implemented for groundwater protection in 1996. Of some 1600 well fields, 643 (around two-thirds of the country s drinking water supply) with a further 78 prospective sources, are considered vulnerable to a range of pollutants. These are mostly over unconfined aquifers, hence the initial implementation of non-pollutant-specific groundwater protection zones. Around each source there are up to five concentric zones, defined at 20 days, six months, five years and 50 years travel time for subsurface flow at the water table, with the total recharge area designated if required. The policy requires assessment of current problems, their remediation and prevention of future contamination; agriculture and sewerage are included as potentially polluting surface activities.

The Hungarian Government seeks harmonization with the EC Nitrate Directive (91/676), the EC Urban Waste Water Treatment Directive (91/271) and the EC Groundwater Directive (80/68). Protection zones are designated in accordance with these directives, including not only NVZs against pollution from nitrates of agricultural origin, but also nitrates from sewage. Codes of good agricultural practice are to be implemented and there has been a thorough evaluation of the nitrate content of surface and groundwaters. In due course, there will be harmonization with the new water framework directive covering all waters.

In the states of Eastern Europe, the trend of nitrogen fertilizer use in the 1990s was downward, reflecting the poor economic conditions. Hungary also has a downward trend. In Western Europe, nitrogen fertilizer use for the same period peaked in the Netherlands, Belgium and Denmark, while it is expected that low growth will level off in the UK, France and Germany.

Yet another dimension to the nitrate issue relates to the impact of climate change in Europe. For instance, in the UK, temperatures are expected to rise from between 0.1 and 0.3 °C per decade in the first half of this century. The pattern of crop production across Europe will likely be perturbed by a northward shift in the agro-ecological zones resulting in changes in nitrogen and organic fertilizer application.

Soil organic matter loss in cultivated land, another probable outcome of climatic warming, is central to considerations of soil carbon, water and nitrogen balances and hence affects the potential for nitrate leaching. However, patterns of change in the loading of nitrate to waters are uncertain. Because raised nitrate losses may be observed following drought and storm events, it is possible that short-term increases in organic matter mineralization will result from raised temperatures. This will make areas of shallow and sandy soils over unconfined aquifers particularly vulnerable. Furthermore, changes in precipitation would increase opportunities for nutrient leaching while temperate clay soils prone to shrinkage and swelling, when exposed to a warmer climate, may also experience increased cracking, further increasing drainage rates and agrochemical losses.

There remain great unknowns for the longer term. Alone, raised atmospheric carbon dioxide has been shown to cause assimilation of soil carbon through increased primary productivity, and has the potential for nitrogen immobilization. Changes in the hydrological balance during (for example) drier summers would also reduce contaminant dilution in rivers and aquifers across Europe and wetter winters would have the opposite effect. There would also be reduced opportunities for denitrification in a drier climatic regime, while both wetter and drier soils in temperate Western Europe may also slow the turnover of soil nitrogen. Should desertification become a reality throughout Southern Europe, an accelerated loss of organic matter would be limited by its own depletion, arguably bringing an end to accelerated nitrate loss and, potentially, to cultivation itself.

The direction, severity and uncertainty of such changes will have to be reflected in policy formulation for nitrate protection zones.

The EU, and indeed certain accession states (including Hungary), are moving towards tight regulation on agriculture in order to control nitrate pollution, and many member states are working hard in order to comply with the appropriate directives. There remains an element of experiment, especially because of long residence times in aquifers and because of the need to establish workable policy instruments at the farmer/land-use interface: a situation aggravated by such variables as changing economic circumstances and scenarios of climate change. Greater success may be achieved by wider-scale measures in certain EU member states (including the UK) with better use of environmental information and further research and monitoring. Control of nitrogen loading from sewage and atmospheric deposition would also greatly assist in the control of nitrate pollution. The EC Urban Waste Water Treatment Directive (91/271) is showing the way to widening controls on nitrate pollution from non-agricultural sources; that from air pollution is more difficult.

ACKNOWLEDGMENT

Dr Grant Edwards of the Imperial College at Wye, Dr Paul A Hart of the Environment Agency of England and Wales and Dr Eunice Lord of ADAS, Ministry of Agriculture, Fisheries and Food for England and Wales, provided critical comments on the draft manuscript.

FURTHER READING

Addiscott, T M, Powlson, D S, and Whitmore, A P (1991) *Farming, Fertilizers and the Nitrate Problem*, CAB International, Wallingford.

Archer, J R and Marks, M J (1997) *Control of Nutrient Losses to Water from Agriculture in Europe*, The Fertilizer Society, York.

Cook, H F (1998) *The Protection and Conservation of Water Resources: a British Perspective*, John Wiley & Sons, New York, 1 336.

Ministry of Agriculture, Fisheries and Food (1999) *Tackling Nitrate from Agriculture*, PB 4401, MAFF London, 1 58.

Nixon, S, ed (1999) *Groundwater Quality and Quantity in Europe*, European Environment Agency, Copenhagen, 1 123.

Stanners, D and Bourdeau, P, eds (1995) *Europe's Environment: the Dobris Assessment*, European Environment Agency, Copenhagen.

Wilson, W S, Ball, A S, and Hinton, R H, eds (1999) *Managing Risks of Nitrates to Humans and the Environment*, Royal Society of Chemistry, Special Publication No. 237, 1 348.

Nitrate Sensitive Areas (NSAs)

see Nitrogen: Agricultural Uses and their Impacts (**Volume 3**)

Nitrate Vulnerable Zones

see Nitrogen: Agricultural Uses and their Impacts (**Volume 3**)

Nitrogen: Agricultural Uses and their Impacts

A L Heathwaite
University of Sheffield, Sheffield, UK

To date, environmental change related to the agricultural nitrogen (N) cycle is essentially a developed world issue but is likely to become a developing world issue unless action is taken to control elevated nitrogen losses to water and the atmosphere. Nitrogen imbalance leads to a range of environmental problems including contribution to greenhouse gases, direct damage to vegetation from nitrogen oxides, eutrophication and acidi cation. Not all problems have an agricultural source but agriculture is implicit in widespread environmental damage.

Plant production responds to increases in the supply of available nitrogen. This relationship is exploited by the agricultural system. In many developed countries, the use of nitrogen fertilizers has increased 100-fold in the post-war years but fertilizer application is generally inef cient. Thirty to 50% of fertilizer nitrogen applied may be lost to the atmosphere through denitri cation and volatilization; a further 30% may be lost through the leaching of nitrogen compounds. Such loss constitutes a major environmental problem and a source of global environmental change. There is good evidence that nitrogen pollution shifted from local to regional scale (Burt et al., *1993) when it became clear that reduction in point source inputs (e.g., industrial discharges, sewage outfalls) were not giving the expected decrease in nitrogen loading to surface waters, groundwaters and coastal seas. Today it is recognized that non-point, primarily agricultural, sources of contamination play a central role in the degradation of both terrestrial and aquatic ecosystems. Environmental management, in the form of legislation and voluntary codes of good agricultural practice has been introduced to combat the decline in the quality of*

terrestrial, aquatic and atmospheric systems. Time lags in both the agricultural system and ecosystem response mean that it may be some time before the effects on the environment of agricultural management are felt.

THE AGRICULTURAL NITROGEN CYCLE

The global nitrogen (N) cycle is described in volume 2 of this encyclopaedia (*see* Nitrogen Cycle, Volume 2). The key processes operating in the nitrogen cycle are:

- fixation from the atmosphere by bacteria;
- mineralization of organic nitrogen in soil organic matter to ammonium by soil bacteria: a key process governing the availability of mineral nitrogen for plants and leaching loss;
- immobilization through plant uptake;
- nitrification through bacterial oxidation of ammonium to nitrite then nitrate (*see* Nitrate Leaching, Volume 3);
- denitrification through reduction of nitrate to nitrite with further reduction to nitrous oxide (N_2O, a greenhouse gas) and gaseous nitrogen;
- volatilization of nitrogen compounds to gaseous ammonia;
- deposition and leaching.

Understanding the agricultural nitrogen cycle in the context of environmental change is largely a question of resolving nitrogen imbalances within the cycle and nitrogen losses from it. Climate, especially rainfall and temperature, is the key control on nitrogen loss from agricultural land owing to its effect on soil moisture, drainage, crop growth and nitrogen utilization. However, the agricultural nitrogen cycle is also driven by human action and as a consequence, land use and land management factors become important; these are discussed in detail later. Figure 1 demonstrates the changing scale of water quality problems.

Figure 2 shows a simplified nitrogen cycle for the agricultural system. Inputs (fertilizers, manures, feed) rarely balance offtake from crops and livestock produce. Furthermore, natural nitrogen inputs from atmospheric deposition have been enhanced by agricultural activity (see later). The resulting nitrogen surplus is the main source of nitrogen-related environmental change in ecosystems. Nitrogen fertilizer may be applied in urea, ammonium or nitrate form. For temperate climates, warm, moist soils mean that the various nitrogen forms are rapidly converted to nitrate that is easily lost by leaching owing to its high solubility. Nitrogen present in manures and some inorganic fertilizers may also be lost through volatilization as gaseous ammonia, with return via atmospheric deposition. If the soil is waterlogged, some nitrogen may be lost to the atmosphere through denitrification. Rivers, lakes and coastal seas act as nitrogen sinks removing nitrogen through denitrification, sedimentation and biotic uptake. The rate of denitrification varies; lakes may remove up to 50% imported nitrogen but the rate of denitrification in rivers is much lower.

ENVIRONMENTAL CHANGE: IDENTIFYING IMBALANCES IN THE AGRICULTURAL NITROGEN SYSTEM

Overview

Surplus nitrogen compounds are emitted and introduced into ecosystems as a result of three main types of human activity:

- combustion of fossil fuels (especially vehicle emissions);

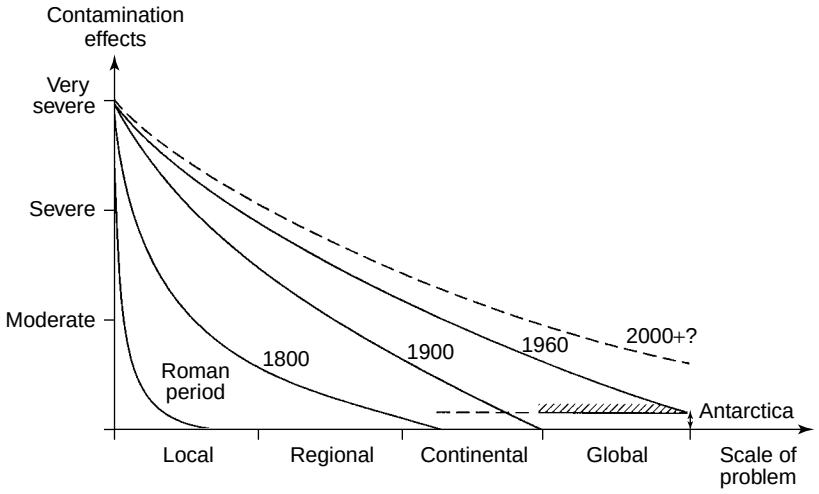


Figure 1 The changing scale of water quality problems over time. (Reproduced with permission from Burt *et al.* (1993))

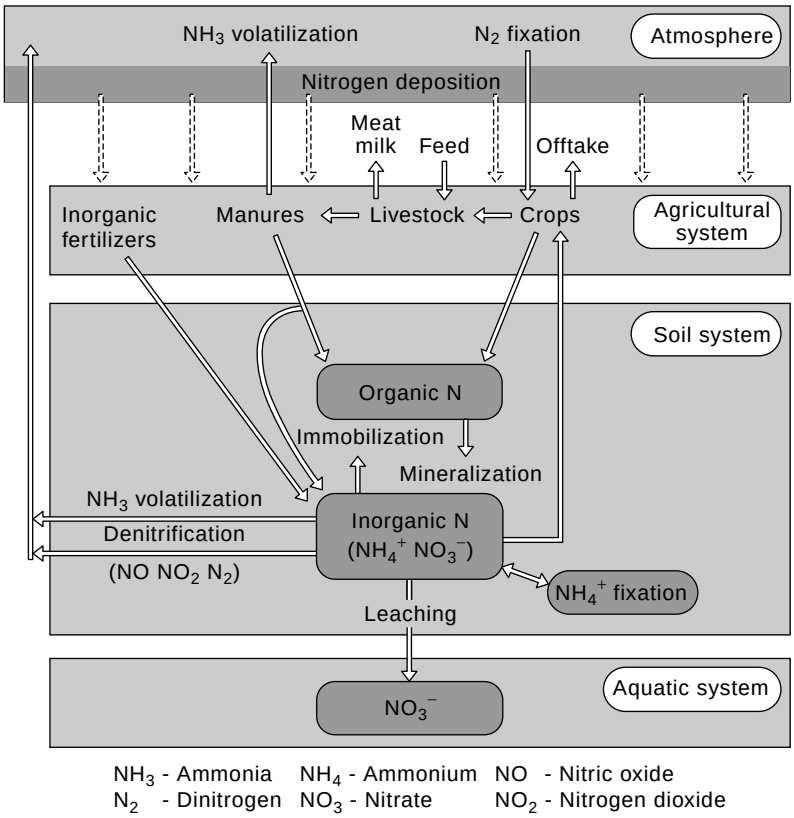


Figure 2 A simplified agricultural nitrogen cycle. (Reproduced with permission from the Ministry of Agriculture, Fisheries and Food (1999))

- direct discharges of sewage and industrial effluents;
- intensive agriculture (the focus of this article).

The introduction of higher yielding crop varieties supported by adequate supplies of fertilizer nitrogen has provided large increases in the yield of most major agricultural crops. Increases in nitrogen use similarly underpin higher livestock stocking densities. As a consequence of this intensification, agricultural land is now the main source of nitrate in most rivers and groundwaters (Burt *et al.*, 1993). Indeed, increased nitrate leaching is common to all intensive farming systems throughout the developed world. Among the other sources of excess nitrogen are increased nitrogen loading from human (sewage) sources and increased nitrogen flux from the atmosphere, although the latter has, in part, an agricultural source.

Changes in Land Use

A key problem in the agricultural use of nitrogen might be termed untimely nitrate, i.e., nitrogen present in the soil that is excess to crop requirements. Whilst climate and soil residual nitrogen are the ultimate controls on nitrate leaching, land management is becoming increasingly important. Untimely nitrogen can be directly linked to land use and land management as discussed later.

Arable Cultivation

Arable cultivation modifies soil structure and the rate of microbial breakdown of organic matter, releasing potentially leachable nitrogen. Arable land has frequently been cited as the key source of nitrogen enrichment in surface waters, lakes, and the marine ecosystem (Burt *et al.*, 1993). Critical land management practices are cultivation and drilling date and the amount and timing of nitrogen fertilizer applications. The risk of nitrate leaching from arable land increases immediately post-harvest when plant uptake is zero but nitrogen release through mineralization is high. Just as different arable crops leave different nitrogen residues in soil, research suggests that some soil types (sands and chalks) and some crops (potatoes, beans, peas, oilseed rape) are more leaky with regard to nitrogen than others.

Whilst it is possible, with careful management, for arable crops to take up the equivalent of 90 100% of the nitrogen applied from fertilizers, more commonly up to 60% nitrogen applied is lost by leaching and denitrification, with small losses though volatilization. Research has shown that the most effective way of minimizing nitrogen leaching

loss is appropriate use of cover crops coupled with careful budgeting of nitrogen fertilizer use (see later).

Grassland

Nitrogen applications to encourage high grass dry matter production are an essential part of grassland livestock management. Relative to arable crops, grassland can absorb a high fertilizer nitrogen supply. Although average nitrogen application to fertilized grass in the UK is 145 kg N ha^{-1} it may be as much as 700 kg N ha^{-1} for dairy farms. The sensitivity of grassland systems to nitrogen loss depends primarily on climate (particularly rainfall following dry spells), and mineralization of nitrogen from soil organic matter. The use to which the grass is put is also important. For example, up to 400 kg fertilizer N ha^{-1} may be applied to cut-grass before substantial leakage occurs. This is possible owing to the continuous vegetation cover, nitrogen removal in the crop and an extensive root system that directs nitrogen to the soil organic matter store. However, ploughing grassland and excessive cultivation, for example, reseeding grass leys, make a significant contribution to nitrate leaching. Nitrogen loss up to 4 ton ha^{-1} has been reported from ploughed permanent grassland although these effects are short-lived. If grassland is used for livestock production, up to 80% of all nitrogen consumed through grazing may be returned as dung and urine. Here nitrogen surpluses arise because offtake (milk, meat) relative to input is small and applications of manures and slurries coupled with direct excretion, increase the potential for nitrogen loss.

Changes in Inputs to the Agricultural System

Nitrogen fertilization is viewed as a vital part of productive grassland and arable crop management. Soils with a long history of nitrogen fertilizer application (both inorganic fertilizers and organic manures) commonly have a large mineralizable nitrogen fraction, forming a potentially large leachable nitrogen source. However, the relationship between nitrogen fertilizer application and nitrate leaching is not straightforward with factors such as climate, soil type and drainage and fertilizer timing and application rate influencing the magnitude of nitrogen losses from agricultural land. Recent research has clearly shown that two agricultural practices lead to excessive nitrogen loss: (i) use of fertilizer nitrogen in excess of crop requirement, and (ii) application of organic manures at excessive rates or at inappropriate times.

The replacement of manures by commercial inorganic fertilizers over the past 50 years is of particular significance for the agricultural nitrogen cycle. Animal manures and slurries spread on arable and grassland are commonly regarded as a waste product, with commercial fertilizers applied regardless of the nutrient content of manures. In the UK, for example, recent estimates of the potential value of animal manures are around £150 M a^{-1} but many farmers perceive the nutrient value of manures to be low relative to inorganic fertilizers, which are more convenient to use and relatively inexpensive. The manure problem is acute in northeast USA, eastern England and The Netherlands, where specialization in intensive livestock units such as pigs and poultry has produced specific problems relating to the available land area to recycle large volumes of animal waste. Current UK data show around 67 M ton animal manures are produced annually (c. 340 000 ton nitrogen) from housed livestock. A further 45 M ton a^{-1} excreta are deposited directly by grazing cattle, sheep and pigs. Manures are applied to c. 16% arable and 48% grassland.

Manure-nitrogen is mainly present as ammonium or organically bound nitrogen. The risk of nitrogen loss from agricultural land varies with manure type. Slurries and poultry manure, for example, have a high available nitrogen content and thus high potential for nitrate leaching. Two problems have recently emerged that continue to frustrate attempts to minimized nitrogen loss from livestock waste: (i) the nitrogen content is highly variable making nitrogen-balancing difficult, and (ii) ammonium-nitrogen in manures is unstable and readily volatilized as ammonia gas. Although a natural process, volatilization is greatly enhanced in intensive livestock production systems.

Changes in Atmospheric Nitrogen Loading and Emissions

Atmospheric nitrogen concentrations are increasing in developed countries as a result of emissions from vehicles and power stations, ammonia volatilization, and denitrification. Both ammonia volatilization and denitrification are enhanced under intensive agricultural production. Ammonia (NH_3) has only been regarded as an atmospheric pollutant in the last 10 years but is likely to become a major contributor to atmospheric change unless controls are implemented. Our understanding of NH_3 emissions, transport, and deposition effects remain limited. Gaseous ammonia may be returned to the soil surface through dissolution in rainwater to form ammonium. Ammonium generates acidity through release of hydrogen ions as it is oxidized to nitrate by soil bacteria. In The Netherlands, the most livestock intensive country in Europe, ammonia deposition is estimated to account for c. 20 30% of total acidification with at least 80% of these emissions coming from livestock production. Around 55% of emissions are derived from cattle, 17% from fertilizers, 11% from sheep, and 8 and 7% from pigs and poultry, respectively. Annual emissions of ammonia in Europe are c. 7 M ton $year^{-1}$ (as nitrogen) which is roughly equivalent to NO_x emissions from industry and vehicles. But whereas NO_x emissions contribute to long-range pollutant transport, most ammonia emissions are localized.

Table 1 On-farm ammonia emissions (%) by source

Livestock buildings	Liquid manure (slurry)	3–13
	Solid manure	3–12
Manure storage	Liquid manure (slurry)	2–10
	Solid manure	17–23
Field application	Liquid manure (slurry)	3–34
	Solid manure	2–18

Ammonia emissions are controlled by soil properties, pH, dry matter content and meteorological controls (especially wind speed, temperature, humidity). Ammonia emissions are greatest from slurry application to land, whilst solid manures emit more NH_3 during storage (Table 1).

Recent developments in emissions control from livestock housing include: separating urine from solid manure, improving the design of liquid manure stores, and constructing housing linked to biofilters to remove nitrogen. In the field, ammonia emissions may be reduced by up to 85% by slurry injection or incorporation rather than surface application.

AN EVALUATION OF THE CONSEQUENCES OF NITROGEN IMBALANCE

The consequences of agricultural-driven change encompass soil nitrogen saturation and soil acidification as a result of nitrogen deposition, water quality deterioration as a result of nitrate leaching, and eutrophication of coastal seas. Knock-on effects include changes in species composition in many aquatic ecosystems, mobilization of heavy metals in acidified catchments, and potential damage to human health through nitrate-rich drinking waters.

Impacts on Water Quality

Nutrient enrichment of receiving waters is undesirable owing to the resulting changes in the freshwater or marine ecosystem and the implications for drinking water affected by toxic algal growth. Nitrate reaching fresh and marine waters essentially acts as a fertilizer causing rapid aquatic macrophyte and algal growth and leading to undesirable ecological changes including anoxia, changes in species composition and loss of sensitive species.

Surface Water Quality

On a global scale, less than 10% of rivers may be classified as pristine (Heathwaite *et al.*, 1997). The average nitrate concentration in European rivers is 4.5 mg NO_3^- N per liter^{-1} compared with 0.25 mg NO_3^- N per liter^{-1} outside Europe. In the eastern USA, similar trends of increasing nitrate concentrations in rivers occur, although absolute concentrations at c. 1.0 mg N per liter^{-1} remained low until the 1990s. Overloading the capacity of forested land to retain nitrogen and elevated atmospheric nitrogen deposition is blamed for the recent increases in nitrogen concentrations in USA surface waters.

Groundwater Quality

The average concentration of nitrate is groundwater is increasing by 1 2 mg per liter^{-1} year^{-1} in intensively cultivated areas (Burt *et al.*, 1993). Shallow groundwater systems are particularly susceptible and may have nitrate concentrations in excess of 25 mg NO_3^- N per liter^{-1}. In the USA, highest nitrate concentrations are recorded in wells within 30 m of the land surface. The increased nitrogen loading to groundwater is thought to be mainly associated with fertilizer application, and is strongly correlated with land use. Owing to the time lags involved in nitrogen percolation to the groundwater table, the nitrate concentration in water extracted from bore holes for drinking water purposes is expected to continue to increase, even if fertilizer applications cease. Elevated nitrate concentrations in bore holes recorded in developing countries appear at present to be primarily related to human sewage disposal, rather than to agriculture.

Eutrophication of Marine and Coastal Waters

Elevated nitrate concentrations make a significant contribution to eutrophication in some marine and coastal waters because nitrogen is a key limiting nutrient for both phytoplankton and benthic algae in marine waters. In the 1970s, marine eutrophication was largely a phenomenon occurring in the coastal seas of northern industrialized regions. Marine eutrophication today may be considered a global environmental problem ranging from the Gulf of Mexico to Chesapeake Bay, the North Sea and the Straits of Malacca. Coastal seas such as the North Sea receive c. 1.5 M ton N a^{-1} from European Union (EU) states; nearly 70% of this input is derived from rivers carrying agricultural runoff. The symptoms of marine eutrophication include algal blooms, anoxia, formation of H_2S and fish kills.

Impacts on Human Health

Although opinion remains divided over the link between human health and nitrate in water, within the EU since 1980, the concentration of nitrate in drinking water has been regulated by the EU Drinking Water Directive. There are two main health concerns: methaemoglobinaemia and stomach cancer. Methaemoglobinaemia or *blue baby syndrome* is a well-documented medical condition thought to arise where ingested nitrate impairs the oxygen carrying capacity of the blood. Ingested nitrate is converted to nitrite in the gut. High concentrations of nitrite result in the oxidation of hemoglobin, ultimately restricting its ability to carry oxygen. The oxidation process occurs more readily in infants. Here it is compounded by a poorly developed enzyme reduction system that restricts the retransformation of methaemoglobin to hemoglobin, potentially resulting in

oxygen starvation and a characteristic bluish tinge. The condition does not affect adults because they are able to convert any nitrite formed back to nitrate, which is then excreted. The risk of blue baby syndrome is low. The World Health Organization (WHO) reported 2000 cases worldwide between 1945 and 1986. Risk is greatest in rural areas where high nitrogen fertilizer use may be enhancing nitrate concentrations in drinking waters.

The link between nitrogen and stomach cancer is poorly understood. Nitrosamine compounds possibly formed from food amines and nitrate in saliva are known to be carcinogenic but the link between nitrate ingestion and nitrosamine formation is less clear. There is little evidence to suggest a link between high nitrate concentrations in surface waters and the incidence of stomach cancer but this may be a function of the time lag for cancer to develop.

THE NEED FOR MANAGEMENT OF AGRICULTURAL NITROGEN

Much of the debate surrounding the need for control of agricultural nitrogen losses centers on who should bear the cost of any nitrogen pollution control strategy. Should the polluter pays principle be supported and largely preventative measures be adopted at a cost to the farmer? Conversely, should curative measures be imposed where the cost is borne by water supply companies and ultimately the taxpayer? A cynic might argue that the taxpayer pays in effect three times for nitrogen pollution: by subsidizing farming, by paying to clean up the resulting pollution, and through possible effects on human health.

In Europe, the critical legislation relating to agricultural nitrogen is the EU Drinking Water Directive. The Directive describes a precautionary approach to nitrogen control and provides a framework for standardizing legislation on the quality of drinking water for member states. The maximum admissible concentration (MAC) for nitrate in drinking water is 11.3 mg NO_3^-N per liter^{-1} or 50 mg NO_3^- per liter^{-1}. The MAC was reviewed by the WHO in 1997. On the basis of the latest scientific evidence, they concluded that the MAC should remain at the maximum guideline level of 50 mg NO_3^- per liter^{-1}. To achieve this level of nitrogen control, strategies preventing nitrate inputs to watercourses or treating nitrogen inputs once they have reached watercourses must be devised. These are reviewed below.

CURATIVE STRATEGIES FOR CONTROL OF NITROGEN POLLUTION DERIVED FROM AGRICULTURE

Curative strategies of nitrogen control are primarily aimed at drinking water treatment. The various approaches include water substitution or blending with low nitrogen water, and nitrogen removal through biological/chemical treatment or storage. The latter is primarily used in drinking water supply reservoirs and operates on the principle that bacteria in the bottom sediments will reduce nitrate concentrations. Biological or chemical treatment includes stimulation of denitrification whereby nitrate is reduced to N_2 or N_2O, by addition of a carbon source (e.g., methanol) to a fluidized sand bed to encourage the growth of nitrogen-removing bacteria. This approach is not without its own problems because although dinitrogen (N_2) is an acceptable end product, nitrous oxide (N_2O) is a serious atmospheric pollutant. The balance between $N_2 : N_2O$ production is largely controlled by soil factors (nitrate concentration, pH, oxygen concentration and organic matter content), and for treatment, the characteristics of the fluidized bed.

Preventative Strategies for the Control of Agricultural Nitrogen Losses

Curative strategies are essentially short-term nitrogen control measures; in the long-term, preventative measures aimed at land use control are viewed as the cheapest strategy. However, most preventative strategies will achieve only slow reductions in nitrogen loss from agricultural land. Rapid reductions may be achieved only with radical shifts in land use from intensive arable and horticultural crops to extensive, unfertilized grassland. In most developed countries such a shift is unacceptable, thus preventative controls are essentially a question of damage limitation: accepting that some level of nitrogen pollution will continue in order to maintain the productive capacity of agricultural land. Thus, in the UK, for example, whilst agricultural policy is aimed at limiting losses of nitrate and other forms of nitrogen pollution from agriculture by promoting practices that reduce such losses, restricting fertilizer use is not seen as economically viable for farmers operating on tight crop yield margins. It would also be some time before any reduction in nitrate concentrations in receiving waters was recorded owing to time lags in the system.

Essentially, three main options exist for preventative management of agricultural nitrogen to reduce nitrate concentrations in receiving waters:

- changes in land use (e.g., conversion of arable to low intensity grass as implemented in the set-aside scheme in the UK);
- changes in inputs (e.g., extensification);
- changes in management (e.g., timing of fertilizer applications).

The key proposed changes in farming practice contained in the various schemes described below are:

- accurate determination of crop nitrogen requirements;
- nitrogen fertilizer application in late spring in order to maximize crop uptake;

- split fertilizer applications rather than one or two major dressings;
- use of slow-release fertilizers;
- use of autumn cover crops to reduce the area of bare ground;
- provision of buffer zones between arable land and receiving waters.

The discussion here relates primarily to schemes promoted in the EU. Similar approaches exist in the USA, where more recently there has been a shift towards whole-farm management strategies.

Nitrate Sensitive Areas

The nitrate sensitive areas (NSA) scheme introduced by the UK Ministry of Agriculture, Fisheries and Food (MAFF) in 1990 has voluntary restrictions on nitrogen use in selected groundwater catchments and nitrate advisory areas, where good farming practice is promoted (Ministry of Agriculture, Fisheries and Food, 1998). The scheme was extended under the EU Agri-Environment Regulation to a total of 32 NSAs in 1994. Within NSAs, farmers are offered payments in return for adopting specified nitrate-reduction practices, including reductions in nitrogen fertilizer inputs below economic optimum rates, use of over-winter cover crops on arable land and conversion of arable to low input grassland. Results from the various NSA schemes show that reductions in mean nitrate concentration in drainage waters might be achieved (Burt *et al.*, 1993).

The EU Nitrate Directive

The EC Directive on Nitrate (91/676/EEC) adopted in 1991, is designed to protect waters against pollution caused by nitrate derived from agricultural sources. Member states must reduce nitrate pollution by introducing controls on agriculture in catchments where the nitrate concentration either exceeds $50\,mg\,l^{-1}$ or is at risk of doing so. The Directive provides scope for integrated controls on land use through designation of nitrate vulnerable zones (NVZs) where farming practices are restricted and the timing and application of fertilizers is controlled. To date, the UK Government has designated 68 NVZs covering 600 000 ha in England and Wales. A mandatory program of control on the amount and timing of nitrogen fertilizers and manures was imposed in the 68 NVZs in December 1998. Designated NVZs include buffer zones in riparian areas designed to reduce nitrate concentrations entering receiving waters by creating an environment in which denitrification is high. Recent research suggests that buffer zones between agricultural land and watercourses or bore holes reduce nitrogen inputs to receiving waters only if the hydrological pathways of nitrogen transport are intercepted (Haycock *et al.*, 1997). Buffer zones should perhaps be viewed as short-term measures of nitrogen pollution control. Furthermore, encouraging nitrogen removal through denitrification passes the problem from the land to the atmosphere because denitrification increases nitrous oxide emissions.

MAFF Code of Good Agricultural Practice for the Protection of Water

The Water Code (Ministry of Agriculture, Fisheries and Food, 1998) instigates voluntary controls on the application of animal manures and inorganic fertilizers to land. The Code is primarily concerned with guiding fertilizer timing and application rates and the use and disposal of animal manures. An application limit of $250\,kg\,N\,ha^{-1}\,a^{-1}$ is proposed to minimize nitrogen pollution risk; this limit does not include manure voided directly to land by livestock. Land use controls are included in the Code by limiting nitrogen fertilizer application within 10 m of a watercourse and 50 m of a bore hole or well.

THE FUTURE: AGRICULTURAL POLICY AND MANAGEMENT

The available research on agricultural nitrogen indicates two critical areas requiring further development for successful control on nitrogen loss from the agricultural system. The first is an improvement in predictive capacity to test various proposals for agricultural land use and fertilizer management. To this end, several models have been developed to predict changes in nitrogen loss as a result of specific changes in land use and management. These include NCYCLE (grassland), SUNDIAL (arable), MANNER (manures), INCA (hydrochemistry, Whitehead, 1990) and LEACHM (soils, Hutson and Wagenet, 1992). The INCA and LEACHN models are less agriculturally biased but include realistic in-river processes and groundwater/surface water interactions, respectively. In nitrogen modeling research, there is a move from sophisticated, data-hungry models towards simple semi-distributed models that estimate nitrogen loss on the basis of the limited data available. Such models commonly operate with a spatial resolution around 1 km and are driven by data availability, which in most developed countries includes land cover, livestock numbers, crops grown, climate, and soil physical properties.

The second critical area of agricultural nitrogen research is development of an integrated catchment management approach. A holistic approach is required because environmental problems unfortunately interact. Thus, buffer zones decrease nitrate leaching but they also enhance nitrous oxide emissions. Similarly, cover crops reduce nitrogen loss but increase the risk of pest and disease problems. Recent developments in farm systems studies pioneered in the USA may provide a means of developing more environmentally balanced approaches to the agricultural nitrogen cycle. UK research on, for example, minimal impact dairy systems (MIDAS) and nitrogen flows within complete grass/arable

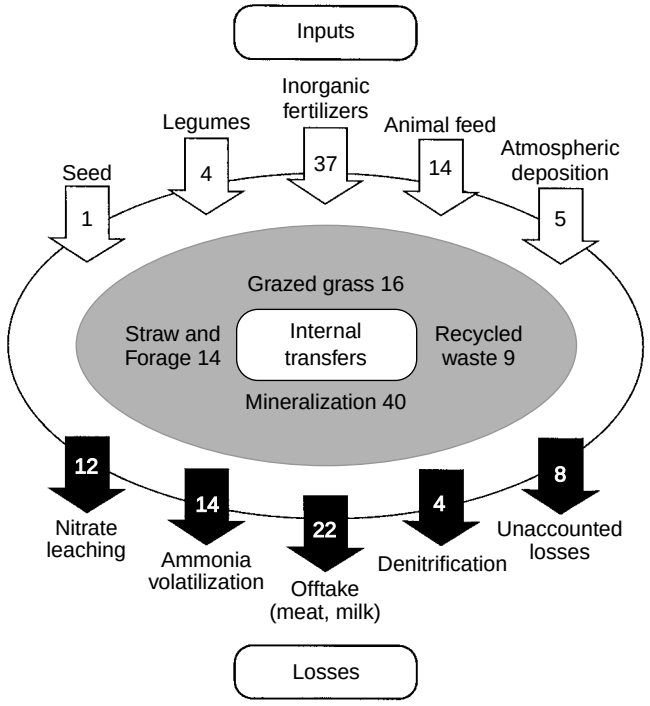


Figure 3 Measured nitrogen flows for a mixed grass/arable farm, southwest England. (Reproduced with permission from Jarvis (1999))

enterprises (Figure 3) has shown that improved management of nitrogen by, for example, adjusting fertilizer inputs together with manure management plus incorporation of maize silage, can reduce nitrate leaching whilst maintaining production. Such integrated studies demonstrate the way forward in achieving a balanced approach to nitrogen in agricultural systems.

See also: Nitrogen Fixers, Volume 2; Animal Production, Feedlots, and Manure Problems in the US (Agriculture, Intensive), Volume 3; Nitrate Leaching, Volume 3; Nitrate Protection Zones in Europe, Volume 3; Trends in Global Emissions: Carbon, Sulfur, and Nitrogen, Volume 3.

REFERENCES

Burt, T P, Heathwaite, A L, and Trudgill, S T (1993) *Nitrate: Processes, Patterns and Management*, John Wiley & Sons, Chichester, 444.

Haycock, N E, Burt, T P, Goulding, K W T, and Pinay, G (1997) *Buffer Zones: their Processes and Potential in Water Protection*, Foundation for Water Research, Marlow, 336.

Heathwaite, A L, Johnes, P J, and Peters, N E (1997) Trends in Nutrients, in *Water Quality Trends and Geochemical Mass Balance*, eds N E Peters, O P Bricker, and M M Kennedy, John Wiley & Sons, Chichester, 139 169.

Hutson, J L and Wagenet, R J (1992) *LEACHM: Leaching Estimation And Chemistry Model: a Process Based Model of Water and Solute Movement Transformations, Plant Uptake and Chemical Reactions in the Unsaturated Zone*, Continuum Vol. 2, Water Resources Institute, Cornell University, Ithaca, NY, Version 3.

Jarvis, S (1999) *Tackling Nitrate from Agriculture*, MAFF Publications, PB4401, London, 18 26.

Ministry of Agriculture, Fisheries and Food (1998) *Code of Good Agricultural Practice for the Protection of Water*, MAFF Publications, PB0585, London, 97.

Ministry of Agriculture, Fisheries and Food (1999) *Tackling Nitrate from Agriculture*, MAFF Publications, PB4401, London.

Whitehead, P G (1990) Modelling Nitrate from Agriculture to Public Water Supplies, *Philos. Trans. R. Soc. London: Series B*, 329, 403 410.

Nuclear Waste: Geological Issues

Charles Curtis

University of Manchester Environment Center, Manchester, UK

For the last 40 years nuclear ssion has made a major contribution to global electric power generation. Currently 31 countries generate nuclear power and 17 rely on this source for more than 25% of their total demand. Looked at another way, had this power been generated by conventional combustion, current atmospheric carbon dioxide (CO_2) levels would have been substantially higher than they are and, presumably, we should be further down the uncertain path of global warming and climate change. Nuclear power generation probably represents the only current technology capable of substantial expansion that does not produce greenhouse gases.

The problem with nuclear power is that it generates radioactive wastes, which are hazardous to all life. Although the radiation, which generates the hazard, naturally declines in intensity with time, certain components within the waste remain active for thousands or even millions of years. The waste must therefore be isolated from the natural environment for time periods much longer than civilization as we know it has existed, therefore, for very much longer than our record of engineered structures. Most scientists and engineers who have worked with the waste problem argue for disposal of the waste in engineered structures deep underground. Earth scientists have two inputs to the debate: they are familiar with rock formations and also with geological and geochemical processes which take place over

time intervals of the same order as required for safe waste containment structures.

This article explores some of the technical issues which must be addressed by the nuclear industry if it is to demonstrate that radioactive waste can be managed safely and that, in consequence, nuclear power generation is a sensible choice for the future. It seems likely that any decision, whether it is to close down or to expand nuclear power generation, will have far-reaching social and economic consequences. On the one hand, there are those who see nuclear power as a threat to the long-term survival of the environment whereas others nd it dif cult to see how the legitimate aspirations of the developing world can be met without a huge increase in electric power generation. There are dif cult choices ahead.

THE NATURE OF THE PROBLEM AND WHY A GEOLOGICAL PERSPECTIVE IS NEEDED

Introduction

Rather more than 450 nuclear power plants (NPPs) are currently operational worldwide (see the section on nuclear power generation: different plants and products). Radioactive waste, mostly in the form of highly radioactive spent-fuel assemblies, continues to accumulate. To date, however, an insignificant fraction of the accumulated waste, both from power generation and military programs, has been disposed of (*disposal* is generally accepted to mean consignment to a facility that, once sealed, would require no further management: a final or permanent management solution): it is held in numerous temporary surface stores around the world. Most technologists working with national and international radioactive waste management programs have argued that deep geological disposal (hundreds of meters burial within stable rock formations) is a safe and the correct solution. There is much public concern, however, about the safety of such disposal. Special interest groups opposed to the nuclear industry have systematically reinforced this concern. The problem really is to decide which is the best management option for the planet: accepting that we have accumulated a very substantial inventory of hazardous radioactive waste which must be dealt with even if all nuclear power generation were to cease tomorrow. There is also the certainty that major global environmental changes will occur within the lengthy period during which the waste will pose a threat to health and the environment.

The Health Hazard, Real and Perceived

The threat to human health and the environment posed by radioactivity is straightforward: radioactive waste is one of many highly toxic materials generated by our society. The objective of radioactive waste management is simply to protect the public from this health threat (regulation and regulators are there to ensure the required level of protection).

Radioactivity is perceived widely to be uniquely dangerous. Perhaps this is because radiation is unseen, hence mysterious. Radioactivity and threat to life are also inevitably associated with the horrors of Hiroshima and Nagasaki.

Surprising to many, there are grounds for regarding radioactive wastes as far less of a threat than many other toxic wastes:

- Radioactive substances are easily detected and identified in minute concentrations (much more easily than most organic chemical toxins).
- The toxicity (dose risk relationships (see below)) of radioactive substances is very much better known than that of the vast majority of chemical compounds.
- Radioactive substances have always been regulated very much more closely than other toxic materials; their amounts and locations are much better known. The volumes of radioactive wastes are tiny compared with non-radioactive toxic wastes.
- Their amount, therefore threat, declines naturally with time, whereas elemental poisons like lead, cadmium, mercury and arsenic last for ever.

A geological perspective is needed because the half-lives of some radionuclides are very long: a small fraction of the radioactive threat lasts for hundreds of thousands or even millions of years. Only earth scientists work daily with processes that operate on such time scales. Long-term management solutions have thus tended to rely on designing repositories in geological settings where the mobility of radionuclides will be naturally constrained: in other words the radioactive elements and the hazard they represent will be contained. The existence of metalliferous ore deposits (including uranium ores) demonstrates that metals remain immobile over geological time periods in the natural subsurface environment. We can determine the age of such deposits and demonstrate them to be millions (or even hundreds of millions) of years old. Under the right geological conditions therefore, some radionuclides should be completely immobilized. Such natural occurrences, when studied in the context of waste disposal, are referred to as *natural analogs*.

Health-dose-risk and the Safety Case (over Geological Time!)

Radioactivity; Sources, Types and Damage Caused

Ionizing radiation is potentially harmful to all living organisms. Complex molecular structures, such as are found in all cells, can be damaged. Chemical bonds are ruptured or caused to rearrange such that molecular shape and

function are modified. Some of these modifications change the way cells grow or replicate, perhaps causing cancer or genetically transmittable mutations. Waste management aims to protect man and the biosphere from these harmful effects primarily by shielding (not allowing the radiation to come into contact with them).

Radiation is emitted when an unstable atomic nucleus disintegrates. This is now measured in units which correspond to single atom disintegrations, bequerels (Bq). Obviously these units are tiny since atoms are so numerous (1 g of hydrogen contains 6×10^{23} atoms). Every radioactive isotope (radionuclide) emits characteristic radiation as it disintegrates. It is useful to recognize a number of broad radiation categories. The most damaging, in health terms, is α-radiation. Table 1 summarizes the properties of the most common categories.

It should be remembered that since the Earth is naturally radioactive, there is an irreducible background radiation level and we are all subjected to significant radiation doses. To this natural background must be added man-made radiation, overwhelmingly medical in origin (especially X-rays). The source and dose of the principal contributors to background radiation are listed in Table 2.

Radioactivity, Dose and Risk

A dose is the amount of radioactivity absorbed by the body (or a part of the body). The different kinds of radiation are more or less harmful to living cells. Some parts of the body are more susceptible than others. An effective dose takes account of these variables and relates the radiation received to its harmful effect (units are sieverts rather than bequerels). These measurements are obviously very much more meaningful in health terms than simply documenting disintegrations of any kind (becquerels). The harm or hazard can be expressed in risk terms which allows the hazardous nature of radiation to be compared quantitatively with hazards from different sources: in turn permitting reasonable analysis of the costs and benefits of the nuclear industry and any remedial measures that may be contemplated.

Table 1 Types of radiation

alpha (α)	Helium nucleus (charged He^{2+}) massive	Ionizing
beta (β)	Electron: high speed, low mass, low momentum	Ionizing
gamma (γ)	Zero mass, charge, electromagnetic wavelength $<10^{-11}$ m^{-1}	Ionizing
Neutron	(Reactor product): creates radioactive isotopes in reactor body	

(unit = becquerel (Bq) = 1 disintegration per second)

Table 2 Typical background radiation levels as doses (units of sieverts – related directly to bequerels but taking account of the more damaging nature of some radiation)

Annual UK average: 2.50 mSv year^{-1}, 88% background, 12% man-made		
Air	1.30 mSv year^{-1}	Radon
Ground	0.35 mSv year^{-1}	^{40}K, Th, U
Food/drink	0.30 mSv year^{-1}	^{40}K
Medical	0.30 mSv year^{-1}	X-rays, irradiation
Cosmic rays	0.25 mSv year^{-1}	Air travel

Fundamental Protection Strategies

There are two basic approaches to managing the health risk posed by additional (over background) radiation from nuclear fission waste products. The first is to isolate (also immobilize or fix) the waste radionuclides within containers and store these within a shielded environment (the shielding to absorb all emitted radiation). The danger lies in some part of the radionuclide charge escaping the confines of the store and dispersing into the accessible environment hence, potentially, to living organisms via air, water or the food chain.

The shielding required is technically straightforward. Fortunately, α-radiation has very little penetration and a sheet of paper is sufficient shielding! The great danger from α-emitting radionuclides is only if they enter the body (as does naturally occurring radon gas). Containment is thus the problem, not shielding. β- and γ-radiation are much more penetrating, therefore, more comprehensive shielding is required. Relatively thin barriers of heavy metals (e.g., lead) or thicker barriers of steel or concrete can be constructed to provide completely effective shielding. The rocks of the earth also provide excellent shielding hence burial underground is wholly effective.

The alternative strategy is to allow discharges to the atmosphere (gases) and oceans (radionuclides dissolved in water) in such a way that the huge volumes of atmosphere and oceans dilute the discharged radionuclides to final concentrations which are radioactive at levels far lower than natural background. The additional risk is effectively zero, unless such huge discharges are allowed that background levels rise significantly. The calculations to evaluate these are simple and reliable. Current international conventions, however, are moving away from the discharge (dilute and disperse) strategy in favor of containment. One reason for this, perhaps, is the realization that dilution cannot be safely assumed for all radionuclides under all circumstances; mechanisms as yet undiscovered may afford reconcentration pathways to hazardous concentrations. Natural processes can and do concentrate elements to high

levels; without such mechanisms, there would be no ore bodies!

Regulation

Regulation aims to ensure that risks associated with radioactivity are quantified and then monitored against agreed limits. Internationally, imposed risks (of cancer or genetically transmittable mutation) of less than 1 in 10^6 are agreed as broadly acceptable. As can be seen from Table 3 this risk is extremely small when compared with risks we all take every day. It is also very small compared with that due to natural background radiation. By any standard, these regulatory risk standards, which apply not to the general public at large but to some small hypothetical group (the critical group) with maximum conceivable exposure, are conservative. Furthermore, many regulators have stipulated that the performance requirement for a waste repository is for an unlimited time. This literally means forever, with a fully quantitative safety case for 10^4 years, a semi-quantitative case for 10^6 years and reassurances that there will be no pulse at any time thereafter. It should be remembered that the hazard continuously decays: the radioactivity of spent fuel declines exponentially with time from the termination of the fission process.

The Fission Process and the Wastes Requiring Management

Nuclear Fission and the Products of the Fission Process

The release of thermal energy from nuclear fission arises because the nuclei of fissionable isotopes (notably ^{235}U and ^{239}Pu) are so destabilized by neutron capture that they split into fragments with release of radiation (α, β, γ, neutrons) and enormous amounts of energy. Nuclear power plants (NPPs) convert most of the energy into heat thence to electricity. One of the most unfortunate aspects of the fission process is that it is almost random in the fragments produced. Each single disintegration generates two or three specific isotopes of particular chemical elements. Within a fuel rod, numerous isotopes of very many different chemical elements are formed and, unlike the vast majority of isotopes in the natural environment, most derived from the fission process are unstable, themselves further decaying, sometimes in several steps with radiation emitted at each.

It is this very large array of different chemical elements and isotopes that poses the major management challenge. Each chemical element has unique properties that determine its solubility and migration properties within the geological subsurface, in the natural environment and food chains and also in any containment structure. For example, ^{127}I is a troublesome waste radionuclide in that it is very soluble in water as the anion I^- (iodide) and this interacts hardly at all with natural materials (see the later section on the ease with which different radionuclides move around in the natural environment). It is dangerous because the thyroid concentrates iodine: and radiation from ^{127}I may contribute to enhanced cancer risks. This mobility also means that iodine is difficult to remove from contaminated waters by a purification plant.

Predicting the performance of a deep geological repository (or any other management solution) thus must depend upon a knowledge of the geochemical cycles and environmental chemistry of a large proportion of the elements within the Periodic Table.

Table 3 Doses, risks illustrated by reference to background radiation levels[a]

Maximum individual regulatory risk to man (of developing terminal cancer or genetic defect)	$<10^{-6}$ year^{-1} or 1 in 1 000 000 year^{-1}
This compares with average risk of death per year	
Smoking 10 cigarettes a day	1 in 200 year^{-1}
Natural causes, age 40	1 in 850 year^{-1}
Road accident	1 in 9500 year^{-1}
Accident at home	1 in 26 000 year^{-1}
Accident at work	1 in 43 500 year^{-1}
Radiation exposure at 1 mSv year^{-1}	1 in 80 000 year^{-1}
Background radiation level, Table 1, is 2.5 mSv year^{-1}	1 in 32 000 year^{-1}

[a] Note, however, that there are different kinds of risks. Some are voluntarily accepted (smoking, car travel) whereas others are involuntary (fallout from a nuclear accident). The outcomes are also perceived in very different ways: some with dread, others as tolerable.

Nuclear Power Generation: Different Plants and Products

Rather more than 450 NPPs are currently operational worldwide. Most modern plants are pressurized water reactors (PWRs) or boiling water reactors (BWRs) which use ceramic uranium oxide fuel pellets encased within inert metal tubes (zirconium alloys, for example). Some older reactor designs, such as the UK Magnox reactors, use metallic uranium within magnesium cladding.

Once the energy has been extracted, the used or spent fuel assemblies contain the fission products trapped within the residual fuel. The assemblies also contain heavy radionuclides formed by neutron capture by uranium, including fissionable isotopes of plutonium (Pu). The oxide fuels, being not unlike naturally occurring uranium oxides, are chemically relatively stable. They are potentially less likely to react/dissolve and release trapped radionuclides into the natural environment than chemically much less stable metallic fuels.

Radionuclides are also generated by neutron bombardment (capture) of components of the fuel assembly and the reactor itself: ^{60}Co is one such.

Waste Forms and Categories

Obviously the bulk of the radioactivity generated by power generation lies within the spent fuel, but this is not all of the waste generated by NPPs. Throughout their working lifetime, relatively small amounts of waste are generated by maintenance. For example, ion exchange resins are continuously used to extract traces of radionuclides from primary and secondary cooling waters. The operations associated with refueling also create small volumes of waste. Both intermediate level wastes (ILW) and low level wastes (LLW) are generated and both pose health hazards hence storage and disposal is carefully regulated. A typical upper limit for LLW is 4 Gbq tonne^{-1} for α-emitters and 12 Gbq tonne^{-1} for β- and γ-emitters (1 GBq $= 10^9$ Bq). Higher radiation levels than these define ILW. High level waste (HLW) is so radioactive that significant heat is generated and cooling is required during storage. HLW includes both spent fuel and the bulk of reprocessing products (see below).

Another defining characteristic of waste is the half-life of the isotopes present. Short-lived wastes can be stored until activity has dropped to low levels making them much less hazardous. Many of the isotopes used in medicine and as tracers in experimental science, for example, have half-lives measured in hours or days. After ten or so half-lives, the residual activity is so low that they can be disposed of without special precautions. Many radioactive isotopes, however, remain dangerously active for thousands of years and some for many millions. In this context, of course, it is again worth remembering that non-radioactive toxic wastes such as cadmium (Cd), arsenic (As) and mercury (Hg) never decay!

A problem for long-term management is that many ILW and LLW waste forms contain organic matter (paper, cloth, ion-exchange resins). These materials are themselves unstable and will modify the chemical environment in which the radionuclides are stored or disposed of (see below). The nuclear industry understands these problems and separates all waste into streams with carefully documented compositions.

This was not necessarily true in the past, particularly in the early days of the industry. Then the principal and very urgent objective was Pu production for weapons. A consequence of the arms race is that all the major nuclear powers have major waste problems inherited from these early days. Furthermore, the amounts of waste generated in the past were much higher (per unit of energy generated) than arise from modern NPPs. We thus have two problems: management of waste arising from current activities and a much larger and more complex problem inherited from the past.

Conditioning, Reprocessing and the Once-Through Cycle

Some waste streams are more difficult to manage than others for reasons other than radiation levels. Liquids and gases are obviously more mobile than solids, therefore less easily contained. Many ILW and LLW waste streams contain organic matter (paper, cloth) and this is chemically unstable, breaking down to a range of more reactive and mobile substances. This is precisely similar to the degradation of organic matter to carbon dioxide, methane and organic acids (leachates) within landfill systems.

The first step in radioactive waste management (after good housekeeping in inventory description) is conditioning. This is the term used for treating and packaging waste to generate a more stable waste form, both physically and chemically. Much LLW and ILW is compacted. It is then placed in steel or concrete containers and grouted with cement-based filler.

There are two fuel cycles. The first is the once-through cycle which is designed to use fissionable fuel once only: extremely radioactive spent fuel assemblies are planned to be sealed within inert copper or stainless steel containers and these flasks can then be disposed of to a deep geological repository.

The alternative fuel cycle strategy involves reprocessing. Spent fuel is dissolved in acid, and fissionable uranium (U) and Pu are recovered for further use as fuel. All three products are at least potentially hazardous. The residual acid solutions are extremely radioactive, chemically reactive and potentially very mobile. Conditioning requires evaporative concentration and conversion to silicate glass (vitrification) which is then solidified within stainless steel flasks. These would be disposed of to a deep repository, much like spent fuel. Although valuable as fuel, U and Pu are potentially dangerous in the wrong hands and therefore demand strict security as an element of management. There is considerable debate about Pu management options; some argue for making it safe by mixing with HLW (the spent fuel standard), effectively reversing the reprocessing stage. It is technically very much more difficult to separate fissionable Pu from spent fuel than it is to manufacture a nuclear device from stolen Pu.

The relative merits of the two fuel cycles are the subject of intense debate. Ultimately, reprocessing can deliver much more energy from a given amount of U fuel than the once-through cycle. Whether or not reprocessing and the use of Pu for power generation become important as a global electrical power source will depend on developing energy demand worldwide, availability of raw materials and the environmental costs associated with different energy conversion technologies.

The Ease with which Different Radionuclides Move Around in the Natural Environment

The mobility of radionuclides obviously determines the likelihood (and rapidity) of their entering the food chain and significantly harming living organisms. Mobility equally dictates their behavior within engineered systems: how difficult they are to contain. This mobility depends not at all on their radioactivity nor on the health hazard they represent but entirely on their chemical properties. The design of containment systems depends primarily, therefore, on the chemical properties of the many different radionuclides generated by nuclear fission and the chemical form they have when produced as wastes.

The principal variables are solubility and degree of interaction with the solid constituents of the engineered barrier system or rocks in the subsurface. The most soluble radionuclides are those which form simple ionic species such as Na^+ and Cl^-. Negatively charged species tend to react least with rocks and minerals because most minerals have negatively charged surfaces. Radionuclides such as ^{36}Cl and ^{127}I are thus relatively mobile, being hardly retarded by these interactions.

The oxides/hydroxides of many metals are much less soluble in alkaline environments; hence concrete (highly alkaline) greatly reduces the mobility of such species in addition to physical containment when used as part of a barrier system. This is a further benefit conferred by the cementitious grout utilized in waste packaging. Another influence is the presence or absence of oxygen, which is generally present in soils and shallow groundwater but absent at depth. The solubility of metals such as uranium or technetium is much greater in oxidizing environments because different ionic species are formed (U^{4+} extremely insoluble, U^{6+} (as UO_2^{2+}) very soluble).

PRINCIPAL THREATS TO THE LONG-TERM INTEGRITY OF MANAGEMENT SOLUTIONS

Introduction

In looking to the longer term, beyond, say 10^3 years, it is important to assess the natural processes which could (and would) impair the containment properties of any store or repository. A good starting place is the performance of natural materials: the mineral constituents of soils, unconsolidated sediments and rocks. These should provide a good yardstick against which to judge the performance of synthetic materials such as cement and composites.

Chemical reactions, whether at the earth s surface (weathering) or deep in the subsurface (diagenesis) are relatively slow and continuous. Most other natural processes which would threaten a containment system, however, are physical, essentially discontinuous with most of the action taking place over very short time periods.

Chemical Degradation/Modification: Continuous Reactions

Chemistry of Rock Degradation Reactions

At the earth s surface, all rocks are broken down to unconsolidated soil products by chemical weathering. The chemistry is very simple with just two core reactions; dissolution and hydrolysis/hydration. For example, quartz dissolves readily (in contrast to some received wisdom) to produce dilute silicic acid:

$$SiO_2 + 2H_2O = Si(OH)_4 \quad (1)$$

All substances are soluble to some extent in pure water; hence rain (very low ionic strength) is a particularly aggressive reactant. The second important reaction type requires both hydrogen ions and water in order to generate charged solute species. Hydrogen ions are abundantly available in soils from organic matter breakdown products, including carbon dioxide (CO_2).

$$CO_2 + H_2O = H^+ + HCO_3^- \quad (2)$$

These reactions normally produce both solid and dissolved products: potassium feldspar under intense weathering breaks down to give soluble silicic acid and potassium with a residue of extremely insoluble aluminum hydroxide (a small proportion of that also inevitably dissolving too).

$$KAlSi_3O_8 + H^+ + 7H_2O = Al(OH)_3 + 3Si(OH)_4 + K^+ \quad (3)$$

Provided there is lots of rainwater available and the temperature is reasonably high (chemical reactions all speed up with increasing temperature), all rocks disintegrate to leave residues rich in hydrated oxides of aluminum and iron. In the humid tropics where both rocks and unconsolidated materials are commonly converted to duricrusts of which bauxites are aluminum rich and ferricretes the iron-rich equivalents.

Environmental Controls on Weathering Reactions

The rate (and completeness) of rock breakdown depends obviously on the availability of pure water capable of taking solute species into solution. Waters with dissolved salts already present are less aggressive because the rate of dissolution is a function of the degree of undersaturation. For both these reasons stone buildings survive longer in arid than humid climates, other things being equal. This applies also to extreme cold where no water is available for reaction; but then other forces may come into play! The basic observations of soil science thus tell us that high temperatures, high rainfall, topography (rapid movement of water away from the weathering site not allowing

the build up of solutes) and high organic productivity (high hydrogen ion input; itself promoted by rainfall and temperature) promote mineral dissolution and rock disintegration. Acid precipitation further accelerate surface weathering. It should also be noted that the salt content of pore waters increases with depth in soil profiles, accordingly they become chemically less aggressive. This pattern extends to depth within groundwater. Equally, rates of pore water movement decline sharply from the unsaturated to the saturated zone of soils and again this trend continues within groundwater. Chemical reactions which lead to mineral dissolution and rock disintegration are thus at their fastest at the Earth s surface and rates decline markedly into the subsurface.

Weathering rates have been quantified in numerous studies of building stones (and gravestones) dating back many hundreds of years. Equally, denudation studies offer quantitative estimates of the rate at which the land surface is being lowered in different climatic and topographic settings. Such studies generate data relevant to assessments of waste management strategies in different places, with various purposes and for the kinds of time interval needed for radioactive waste isolation.

Evaporative concentration of weathering solutions in arid environments can cause rock disintegration by a different mechanism; salt precipitation in pore space. Within the deeper subsurface, however, mineral precipitation leads to cementation rather than disintegration. In fact the norm when sediments are buried is diagenetic cementation. Were this not the case, of course, we would not have physically strong sedimentary rocks, notably sandstones and limestones; perhaps the most common of all building materials.

One other important environmental control on mineral stability and solute mobility is the presence or absence of excess oxidants (especially molecular oxygen) or reducing agents (notably organic matter). In water-saturated environments such as the geological subsurface, free oxygen is almost invariably absent whereas in drained soil systems, oxygen is present. Rocks with minerals stable only in reducing environments will be stable in the subsurface but not at the surface, and *vice versa*. Very energetic redox reactions ensure rapid degradation reactions when the environmental setting changes. The oxidation of sulfide minerals when brought to the surface as a result of mining operations is the most familiar example (with disastrous environmental consequences) (*see* Redox Potential, Volume 2; Contaminated Lands and Sediments: Chemical Time Bombs?, Volume 3).

Differential Stability of Solids

It takes only one mineral from an assemblage to break down and the rock will fall apart. The analogy of a chain and its weakest link is apt. From both observation and theory, the relative chemical stability of all common rock-forming minerals is known; hence it is possible to assess with confidence the geochemical stability of different rock types. A further obvious control is porosity and permeability since chemical reaction demands water access. Other things being equal, extremely tight rocks are most resistant to degradation.

Potential Host Rock Types For Deep Repositories

Three different rock types (lithologies) have been widely advocated by different agencies as suitable for deep radioactive waste disposal sites. These are crystalline rocks, shales and salt. Each has its advantages and disadvantages. Crystalline rocks have high mechanical strength and thermal stability (good engineering properties) but are often fractured. The resulting permeability may be high and is difficult to model (therefore to demonstrate safety). Shales and salt, by contrast, are low permeability rocks, which behave plastically: they are naturally self-sealing and also easy to excavate. Shale is a rather poorly defined term covering clay rich rocks, which vary from poorly consolidated, highly plastic muds and clays to less porous, more indurated and sometimes fissile rocks: mudstones or shales proper. This progression from mud through to shale records age and burial history; the older the formation, the more likely it is to be indurated. Salt is thermally stable and highly conducting, both positive attributes not shared by shales. On the other hand, salt is potentially very mobile were there to be a change in groundwater regime. All three lithologies presently host underground stores or laboratories: Sweden, Canada and Finland in crystalline basement, Belgium and Switzerland in shales and the USA in salt (Waste Isolation Pilot Plant (WIPP) site).

Theory and Demonstration: Building Confidence

This very brief review of a substantial and mature science leads us to rather surprisingly precise statements about settings and about materials that are likely to survive for long time periods: a blueprint both for containment (packaging, engineered containment) and for good geological settings. Natural analog studies allow these general observations and conclusions to be focused on specific issues in nuclear waste management. Studies of ancient uranium ore bodies, for example, prove that, in certain situations, solution and migration have not occurred over millions of years: natural containment has been complete. In this particular case, of course, it is the natural radioactivity that gives us precise dates for the original mineral formation. In a similar way, the stable isotope equivalents of key radionuclides can be studied in a range of geological environments and ages. Such studies should build confidence in management strategies for the necessary very long management terms required by regulation.

Climate and Climate Change (Arid/Humid Changes, Sea level change, Warm/Cool Transformation ± Glaciation)

From the previous discussion it follows that climate change (time scale 10 to 10^3 years) will influence the stability of rocks and synthetic materials at the surface and in the subsurface. Chemical degradation rates at the surface will decline markedly going from a humid to an arid climate, and *vice versa*. In desert environments the effects of physical disturbance last for decades whereas all traces of such disruption are quickly lost in the humid tropics where chemical and biological activity are both intense. As yet, the science of global climate modeling is insufficiently advanced for any kind of confident long-term (centuries) prediction. Historical rates of change are well documented in the geological and oceanographic literature and give some indications of stability and uncertainty.

Deep subsurface environments are very much less sensitive to climate and, obviously, to climate change. Within an arid basin of low topographic variation, groundwater will be saline and, almost certainly, stratified. The density difference between meteoric water and saline brines is sufficiently great to demand high pressure-heads for displacement. Consequently, a change to a more humid setting would simply establish strong salinity/density gradients in groundwater and effective isolation of deep from shallow water systems. So geological (and other) materials at depth are very much less susceptible to chemical modification than those at the surface and also less sensitive to climate change.

There are examples of deep penetration of fresh, aggressive water which apparently has caused significant dissolution (porosity and permeability enhancement: important in oil and gas reservoirs) but these seem invariably linked to high differential topography.

There is abundant evidence of sea level change throughout geological history clearly preserved in virtually all sedimentary rock successions. Such changes will modify groundwater regimes in the coastal zone, delineated by the extent of rise/fall. Seawater will easily displace freshwater but not the reverse so the overall effect is probably to promote the development of salinity stratified waters in sedimentary basins (the observed norm).

The glaciation end of the cycle lowers sea level to a minimum. Glaciers themselves, however, are in hydraulic continuity with groundwater and can generate high hydraulic heads which could lead to anomalous displacement of groundwater with glacier derived meteoric water. Zones of actual or potential glaciation therefore pose substantial questions about the nature of chemical reactions in deeper rock sequences: uncertainties not present in non-glacial settings.

Sea level change and, especially, glaciation lead also to physical erosion. In both cases the effects are more or less continuous with transgression of shorelines, down-cutting of river systems and deep glacial excavation.

Discontinuous Physical Erosion Processes

Unlike glaciation or sea level change, the essentially statistical nature of weather means that most erosion by storm or tidal surge (or other events) means that almost all the action takes place once every 10^2 or 10^3 years with really not much happening in between. Risks of this kind are easily identified and evaluated: it is surely rather surprising, even allowing for the benefits of hindsight, to note just how many NPPs and manufacturing facilities have been constructed so very close to the ocean in Europe and in North America. Like chemical degradation, the physical threat to materials and structures at the Earth s surface is very obviously much greater than that to similar materials and structures located in deep geological formations.

Earthquake, Volcanicity Tsunamis

There are also disruptive events generated by crustal processes. Stress releases cause fault displacements. Individual fractures are confined to very small, two-dimensional zones and the probability of one intercepting (and destroying) a structure is very low. Moreover, the distribution of fault zones on the Earth s surface is well characterized and those with high activities can be avoided without difficulty. A much bigger problem is the energy released by displacement events (earthquakes). This can destroy structures by shaking or secondary land deformation (foundation failure, landslips, etc.) over 10 s to 10^4 km^2. Submarine energy releases cause waves which can travel great distances and which have devastating consequences for coastal zones they intercept (tsunamis).

MANAGEMENT PRACTICE

Waste Management Principles and Practice

In spite of 50 years of radioactive waste generation, no nation has yet put into practice any long-term solution for spent fuel and high liquid wastes (HLW). The International Commission for Radiological Protection and the Earth Summit (1992 UN Conference on Sustainable Development) cite the principle of intra-generational equity: that management burdens should not be passed on to future generations. This argues clearly for disposal: there will be no future management requirement. For a variety of reasons the opponents of nuclear power argue against disposal in favor of temporary surface storage. One case made is that available scientific and technological knowledge on waste disposal is not good enough; that we will not hand on a worse problem (a failing repository: although no

systematic analysis of such failure and its consequences has, to my knowledge, been undertaken). It is also true, however, that a solution to the waste management problem would give the green light to the nuclear industry. These opposing views and values are not for debate here but the alternatives are clear. We either surface store (for some indeterminate period) or we attempt some more permanent solution, deep disposal being the apparently obvious choice. Where does the geological perspective lead us in this debate? As outlined above, shielding is easy but containment is not. The principal threat to health is contamination of surface and groundwater used for drinking and agriculture.

Surface Store

The least expensive form of management in the short term is surface storage. Dry stores take conditioned waste forms (see the section on monitored, retrievable deep store) and store them in a facility which:

- provides security (against terrorist acts and thefts of fissionable material);
- provides shielding (against worker/public irradiation);
- provides cooling (for HLW and spent fuel);
- allows continuous monitoring (for containment failure);
- is designed to contain leakages (since very accessible to the biosphere);
- provides a facility for repackaging when failure occurs.

The key element here is that continuous management is required (with finite doses for management staff). The design of the facility will be such as to minimize the effects of weathering (protection from the elements) and to minimize the effects of catastrophic natural events. The design life of store buildings is likely to be tens of years, certainly not hundreds or thousands.

Provided management is assumed, no long-term safety case is required. Delivery of such stores should be relatively straightforward and modern NPPs may install sufficient spent fuel storage capacity for their entire working lifetimes.

It is inconceivable that the regulatory requirements for disposal could be made for surface disposal. Exposure to both natural degradation processes (weathering, erosion, flooding) and man-induced failure (accidental damage, terrorism, economic breakdown with consequent failure of management regime) would all seem to be almost infinitely greater at the surface than deep beneath the ground in a reasonably well chosen location.

Deep Disposal

Although a simple comparison of surface and deep disposal suggests the latter to be very much safer in the long-term, possible failure mechanisms have been studied in great detail as part of the preparation of post-closure performance assessment: as required by regulators. The geosphere (i.e., the rocks above the repository) provide perfect shielding but must be evaluated for possible radionuclide escape pathways. In general, repository design concepts are based on the so-called multiple barrier approach. Conditioned waste is placed within sealed canisters which, in turn, are surrounded by engineered barriers (absorptive clay, concrete, etc.) which provide both physical and chemical containment. The geosphere itself acts as a barrier to radionuclide migration and provides physical protection for the engineered barriers. The properties of the geosphere (rock permeability, hydrodynamic regime, geochemical environment, tectonic stability, presence or absence of natural resources likely to be sought, sensitivity to climate change) are evaluated at the site selection stage, final choice obviously being guided by some form of multi-attribute analysis.

Monitored, Retrievable Deep Store

In recent years, consultation exercises have revealed that the general public doubt the safety case made for deep repositories (in one sense, of course, proof is unattainable: it is fundamentally impossible to prove the negative). Special interest groups aiming to end all nuclear power generation encourage and sustain this doubt. They see that a good, safe solution to the waste problem would offer a green light to expansion of the nuclear industry. This being the case, it follows that it is not in their interest to find good solutions to waste management problems.

The case for proving the very long-term safety of a repository is very difficult: harsh regulatory targets have to be met for times much longer that the history of the human race! The only way to demonstrate such safety cases is via risk analysis and mathematical modeling: neither is easy to explain nor is likely to give reassurance to the non-specialist. For this reason thought is now being given to designing deep repositories with some of the apparent benefits of surface stores. The integrity of the conditioned waste can be continuously monitored up until the time of closure (and for some time thereafter) and, up to that time, waste can also be retrieved easily.

It is also apparent that a fail-safe element can be designed into deep repositories. The failure of institutional control would remove management and immediately expose surface stores to terrorist threats or accidental damage in the short-term whereas deeply stored/disposed material would be much less vulnerable.

Very Deep Disposal

Various suggestions have been made for very deep disposal of HLW or Pu at great depths (several kilometers) by

drilling. In the most elegant designs, sufficient heat is generated by the waste just to melt surrounding crystalline rock. On subsequent dissipation of some of the heat so generated, the melt would fuse, encapsulating the waste in a large impermeable container. Retrieval would not be possible but once so entombed, it appears very unlikely that significant waste would ever escape within the decay period of almost all radionuclides.

THE GEOLOGICAL PERSPECTIVE IN SUMMARY

Geologists, and earth scientists more generally, work with natural materials and processes which take place on time scales comparable with those required for nuclear waste to be contained and stored (until its radioactivity has decayed to such low levels that they no longer pose an hazard to Man and the environment). These time intervals are many times longer than the totality of civilization such that conventional technology has absolutely no relevant record of performance.

The regulators, nevertheless, have stipulated that any repository designed for permanent disposal (i.e., a sustainable waste management solution) must meet safety standards extending far beyond the likely survival of *Homo sapiens*. The only way that future performance can be assessed is by mathematical modeling. The only reasonable validation is by analysis of processes spanning equivalent time periods historically: geological studies of natural analogs.

Geological studies demonstrate that natural circumstances have prevailed which, over millions of years, have isolated materials analogous to conditioned radioactive wastes. It therefore seems likely that repository designs, which maximize the benefits of natural geological containment are likely to meet the most stringent (and barely sensible?) regulatory requirements. Rock properties such as low permeability and high chemical retardation are not rare. Sites with very low groundwater flow tendency, which are tectonically stable, which are insensitive to the effects of weather and of climate change should not be too difficult to locate.

The geological perspective thus suggests that deep geological disposal is a safe option and certainly much safer than long-term surface storage where conditioned wastes are hugely more exposed to degradation (natural and man-induced) and eventual leakage directly to the accessible environment and biosphere. Deep geological disposal similarly reduces vulnerability to global environmental change, particularly to the suggested increasingly common extreme meteorological events.

The information upon which this perspective is based must be made available to the general public such that informed debate can take place and informed policy developed. The geological perspective is only one of several which must be brought to bear on this complex and pressing problem. As time passes, the problem gets more urgent as most of the world s economies are becoming less robust. How many will be able to meet the costs of management and, indeed, how many of the stronger economies have actually made real provision to meet the costs? Within this depressing scenario, those groups who regard delay as success may be celebrating pyrrhic victories.

ACKNOWLEDGMENTS

This review has benefited greatly from discussions with numerous colleagues, especially those present and past members of the UK Government Radioactive Waste Management Advisory Committee (RWMAC). The views expressed here, however, are purely personal and do not necessarily reflect those of either colleagues or RWMAC.

FURTHER READING

House of Lords (1999) *Select Committee on Science and Technology: Management of Nuclear Waste*, HL Paper 412, London, HMSO, 1 90.

OECD (1991) *Review of Safety Assessment Methods*, Nuclear Energy Agency, OECD, Paris.

POST (1997) *Radioactive Waste - Where Next?* Parliamentary Office of Science and Technology, London, 1 99.

Royal Society (1998) *The Management of Separated Plutonium*, The Royal Society, London, 1 14.

The Royal Society and The Royal Academy of Engineering (1999) *Nuclear Energy, The Future Climate*, The Royal Society, London, 1 76.

Warner, F and Kirchmann, R J C, eds (2000) *Nuclear Test Explosions: Envionmental and Human Impacts*, SCOPE 59, John Wiley & Sons, Chichester.

Witherspoon, P A (1996) *Geological Problems in Radioactive Waste Disposal*, US Department of Energy contract DE-ACO3 76SF00098, 1 270, Ernest Orlando Lawrence, Berkeley National Laboratory, CA.

Nuclear Winter

The enormous direct impact of nuclear explosions in warfare was dramatically demonstrated in the Japanese cities of Hiroshima and Nagasaki on August 6th and 9th, 1945, causing the death of 120 000 people from the heat and the shock waves of the explosions. The explosive power of the two weapons was equivalent to 15 and 21 ktons of trinitrotoluene (TNT), corresponding to energies of 6.3

drilling. In the most elegant designs, sufficient heat is generated by the waste just to melt surrounding crystalline rock. On subsequent dissipation of some of the heat so generated, the melt would fuse, encapsulating the waste in a large impermeable container. Retrieval would not be possible but once so entombed, it appears very unlikely that significant waste would ever escape within the decay period of almost all radionuclides.

THE GEOLOGICAL PERSPECTIVE IN SUMMARY

Geologists, and earth scientists more generally, work with natural materials and processes which take place on time scales comparable with those required for nuclear waste to be contained and stored (until its radioactivity has decayed to such low levels that they no longer pose an hazard to Man and the environment). These time intervals are many times longer than the totality of civilization such that conventional technology has absolutely no relevant record of performance.

The regulators, nevertheless, have stipulated that any repository designed for permanent disposal (i.e., a sustainable waste management solution) must meet safety standards extending far beyond the likely survival of *Homo sapiens*. The only way that future performance can be assessed is by mathematical modeling. The only reasonable validation is by analysis of processes spanning equivalent time periods historically: geological studies of natural analogs.

Geological studies demonstrate that natural circumstances have prevailed which, over millions of years, have isolated materials analogous to conditioned radioactive wastes. It therefore seems likely that repository designs, which maximize the benefits of natural geological containment are likely to meet the most stringent (and barely sensible?) regulatory requirements. Rock properties such as low permeability and high chemical retardation are not rare. Sites with very low groundwater flow tendency, which are tectonically stable, which are insensitive to the effects of weather and of climate change should not be too difficult to locate.

The geological perspective thus suggests that deep geological disposal is a safe option and certainly much safer than long-term surface storage where conditioned wastes are hugely more exposed to degradation (natural and man-induced) and eventual leakage directly to the accessible environment and biosphere. Deep geological disposal similarly reduces vulnerability to global environmental change, particularly to the suggested increasingly common extreme meteorological events.

The information upon which this perspective is based must be made available to the general public such that informed debate can take place and informed policy developed. The geological perspective is only one of several which must be brought to bear on this complex and pressing problem. As time passes, the problem gets more urgent as most of the world s economies are becoming less robust. How many will be able to meet the costs of management and, indeed, how many of the stronger economies have actually made real provision to meet the costs? Within this depressing scenario, those groups who regard delay as success may be celebrating pyrrhic victories.

ACKNOWLEDGMENTS

This review has benefited greatly from discussions with numerous colleagues, especially those present and past members of the UK Government Radioactive Waste Management Advisory Committee (RWMAC). The views expressed here, however, are purely personal and do not necessarily reflect those of either colleagues or RWMAC.

FURTHER READING

House of Lords (1999) *Select Committee on Science and Technology: Management of Nuclear Waste*, HL Paper 412, London, HMSO, 1 90.

OECD (1991) *Review of Safety Assessment Methods*, Nuclear Energy Agency, OECD, Paris.

POST (1997) *Radioactive Waste - Where Next?* Parliamentary Office of Science and Technology, London, 1 99.

Royal Society (1998) *The Management of Separated Plutonium*, The Royal Society, London, 1 14.

The Royal Society and The Royal Academy of Engineering (1999) *Nuclear Energy, The Future Climate*, The Royal Society, London, 1 76.

Warner, F and Kirchmann, R J C, eds (2000) *Nuclear Test Explosions: Envionmental and Human Impacts*, SCOPE 59, John Wiley & Sons, Chichester.

Witherspoon, P A (1996) *Geological Problems in Radioactive Waste Disposal*, US Department of Energy contract DE-ACO3 76SF00098, 1 270, Ernest Orlando Lawrence, Berkeley National Laboratory, CA.

Nuclear Winter

The enormous direct impact of nuclear explosions in warfare was dramatically demonstrated in the Japanese cities of Hiroshima and Nagasaki on August 6th and 9th, 1945, causing the death of 120 000 people from the heat and the shock waves of the explosions. The explosive power of the two weapons was equivalent to 15 and 21 ktons of trinitrotoluene (TNT), corresponding to energies of 6.3

and 8.8×10^{12} J, respectively. In the meanwhile, despite some reductions in nuclear arsenals in the US and the former Soviet Union, the total worldwide number of nuclear weapons is still more than 10 000, each with explosive powers averaging 20 30 times those of the Hiroshima and Nagasaki weapons. These weapons are mainly in the possession of Russia and the US. Although the impact of a nuclear war with the current arsenals can not be scaled up linearly from the experiences in Hiroshima and Nagasaki, the devastating direct impact of a nuclear war on the nations involved in such a nuclear conflict would still be enormous.

However, the potential impact on humanity resulting from the effects of a nuclear war was actually, for a long time, much underestimated. Following an article by Crutzen and Birks in the beginning of the 1980s, a large number of scientists drew attention to the large climatic and environmental consequences, which would result from the large amounts of highly sunlight-absorbing black smoke, which would be injected into the atmosphere by the many fires that would start burning after nuclear explosions in cities and in oil depots (Crutzen and Birks, 1982). As a consequence of the enormous loading of the atmosphere with the black smoke, and its transport over much of the Northern, and part of the Southern Hemisphere, sunlight at the Earth s surface would be much reduced over large tracts of the Earth. Also, as a consequence of the major reductions in sunlight, temperatures at the Earth s surface could become greatly reduced, possibly to below freezing even during the summer. Together, the nuclear war could upset crop production over much of the Earth, putting a large fraction of the Earth s population at risk of starvation and death, including people living in the attacking nation. The combination of darkness and freezing temperatures is known as *nuclear winter*, a name first coined by Richard Turco, now at the University of California, Los Angeles.

REFERENCE

Crutzen, P J and Birks, J W (1982) The Atmosphere After a Nuclear War: Twilight at Noon, *Ambio*, XI, 114 125.

FURTHER READING

Harwell, M A and Hutchinson, T C (1985) *Environmental Consequences of Nuclear War*, SCOPE 28, John Wiley & Sons, Chichester, Vol. 2.

Pittock, A B, Ackerman, T P, Crutzen, P J, MacCracken, M C, Shapiro, C S, and Turco, R P (1986) *Environmental Consequences of Nuclear War*, SCOPE 28, John Wiley & Sons, Chichester, Vol. 1.

PAUL CRUTZEN *Germany*

O

OH–Radical: is the Cleansing Capacity of the Atmosphere Changing?

see OH–Radical: is the Cleansing Capacity of the Atmosphere Changing? (**Volume 2**)

Oil and the Arctic Environment

David J Thomas
Axys Analytical Services Ltd., Sidney, Canada

Petroleum hydrocarbons have been detected in all abiotic and biotic compartments of the circumpolar arctic environment and have both local and regional sources. One of the greatest concerns for ecological impact in the arctic environment is related to oil spills because the severity and persistence of the impacts will be greater than those in temperate and tropical environments where biological recovery is faster due to higher growth rates, longer growing seasons, faster generation turnover times and lower ages at maturity. Moreover, physical characteristics of the Arctic such as permafrost, sea ice regime, severe temperature and remoteness could make clean up of oil spills exceedingly slow and in some cases impossible leading to long-term exposures and impacts.

The traditional de nition of the Arctic, namely that area of the earth north of the Arctic Circle, is generally inadequate for environmental applications because it fails to take into account physical, geographical and ecological characteristics. Instead, the Arctic Environment can be thought of as the area north of the tree line and the 10°C July isotherm; speci cally, the terrestrial and marine areas north of the Arctic Circle and north of 62°N in Asia, and 60°N in North America modi ed to include the marine areas north of the Aleutian chain, Hudson Bay and parts of the North Atlantic Ocean including the Labrador Sea.

Oil is a general term used to describe a wide variety of natural substances of plant, animal or mineral origin and a group of synthetic compounds. In this paper, oil is taken to mean crude oil or petroleum, a substance which consists primarily of a series of complex hydrocarbon compounds such as alkanes, naphthenes, aromatics and alkenes. Crude petroleum is a complex and variable mixture of hydrocarbons of different molecular weight and structure, and contains as many as 300 different petroleum hydrocarbon compounds.

SOURCES

Petroleum hydrocarbons in the arctic environment have both natural and anthropogenic sources. The primary natural source is seeps from geological formations. Smaller quantities of specific petroleum hydrocarbon compounds are produced naturally by bacteria, fungi, and higher plants and during the incomplete combustion of organic materials such as wood, coal and oil. Anthropogenic sources include chronic discharges from oil and gas exploration, development and production activities, accidental releases from oil spills, discharges from shipping and atmospheric deposition. A summary of possible sources is shown in Table 1.

Natural Oil Seeps

The presence of petroleum seeps has been documented in many areas of the Arctic including Canada (Mackenzie River Delta), the US (several locations along the Beaufort Sea Coast), the Barents Sea and near Spitzbergen (Figure 1). The largest seepages occur in north-flowing rivers (e.g., Mackenzie River and Ob River) which eventually discharge the petroleum hydrocarbons into the Arctic Ocean. The rate of natural seepage of oil into the arctic marine environment relative to total input is thought to be greater than the 15% estimated for the global marine environment (GESAMP, 1993).

Table 1 Sources of petroleum hydrocarbons to the Arctic environment

Natural	Oil seeps Biogenic synthesis Transport by ocean currents Transport by northward-flowing rivers Transport by air movements from subarctic and temperate regions
Anthropogenic	Transport by ocean currents Transport by northward-flowing rivers Transport by air movements from subarctic and temperate regions Oil and gas exploration, development, production (operational discharges, blow-outs) Transport of crude and refined oil (accidents from tankers at sea, tanker ballast washing, leakage/spillage from pipelines and tanker trucks on land) Transportation (discharges and spills from shipping, emissions and leakage from vehicles) Land-based discharges and run-off (municipal waste water, refineries, industrial waste water, combustion of wood and fossil fuels)

Oil and Gas Exploration, Development and Production

The Arctic may contain some of the world s largest oil and gas reserves. Current proven reserves are approximately 17 billion m^3 of oil and it is estimated that undiscovered, conventionally recoverable oil may exceed another 20 billion m^3.

Significant hydrocarbon exploration has taken place both offshore and on land in the Canadian Arctic during the past 30 years, but no permanent production facilities exist except for those at Norman Wells on the Mackenzie River. Production is approximately 2 million m^3 per year.

In the US Arctic, large-scale production of oil has occurred since 1977 when the Prudhoe Bay fields began shipping oil south via the Trans-Alaska Pipeline System. It is estimated that about 7 billion m^3 of oil are still recoverable from the Prudhoe Bay and adjacent fields (AMAP, 1998). In addition, expected production from undeveloped offshore areas will be about 0•08 0•43 $\times$ 10^9 m^3 and 0•05 0•24 $\times$ 10^9 m^3 for the Beaufort Sea and the combined Chukchi Sea Hope Basin region, respectively (MMS, 1996).

Oil and gas exploration and production activities in the Russian Arctic occur primarily (1) in the Nenets Antonymous District of the Arkhangelsk region, (2) in the Komi Republic within the Pechora River drainage basin, and (3) in the Yamal-Nenets Antonymous District of Western Siberia. Total annual production exceeds 100 million m^3 of crude oil, 740 trillion m^3 of natural gas and 2.2 million m^3 of gasoline (figures in tonnes have been converted to m^3 assuming a specific gravity of 0.9). Large oil and gas fields have been discovered along the continental shelf areas of the Barents, Kara and Pechora Seas.

The only remaining significant oil and gas activity in the Arctic occurs in the Norwegian portions of the Barents Sea and in the Norwegian Sea. Total proven reserves are approximately 400 million m^3 of oil and natural gas liquids, and 350 million m^3 of gas (NPD, 1995). Undiscovered reserves are estimated to be similar in size to proven reserves.

Chronic point source losses of petroleum hydrocarbons to the environment are associated with most oil and gas exploration, development and production activities. Primary sources of relatively minor petroleum hydrocarbons include discharge of oil-base drilling muds, oiled drill cuttings, rig wash, produced water and periodic accidental losses of refined petroleum products.

Blow-outs, tanker spills and leakage, however, pose the largest threat from oil pollution to terrestrial and aquatic ecosystems in the Arctic. To date two large-scale spills have occurred: (1) 1989: the EXXON Valdez tanker accident where 35 000 tonnes of crude oil entered the marine environment (*see* Exxon Valdez Oil Spill, Volume 5), and (2) 1994 1995: the pipeline rupture at Usinsk, Siberia where more than 100 000 tonnes of crude oil flooded rivers, lakes and tundra. In both cases, significant damage was done to the flora and fauna of the arctic environment, damage that will take decades to repair.

Shipping

Currently, inputs of petroleum hydrocarbons to the arctic environment related to routine marine transportation are limited by the small amount of ship traffic. The most common arctic shipping is associated with sealifts to isolated communities and industrial facilities, ice breaker support for sealift traffic, fishing, seismic and oceanographic research and supply of offshore drilling operations. The potential exists for large inputs, however, as a result of the likely expansion of traffic on the Northern Sea Route (NSR). The NSR is a system of sea lanes north of Asia joining the Kara and Barents Seas in the west with the Bering Sea in the east. It was officially opened by the former Soviet Union in 1987; annual inputs of petroleum hydrocarbons, largely from illegal dumping of bilge waters and accidental fuel spills, vary between

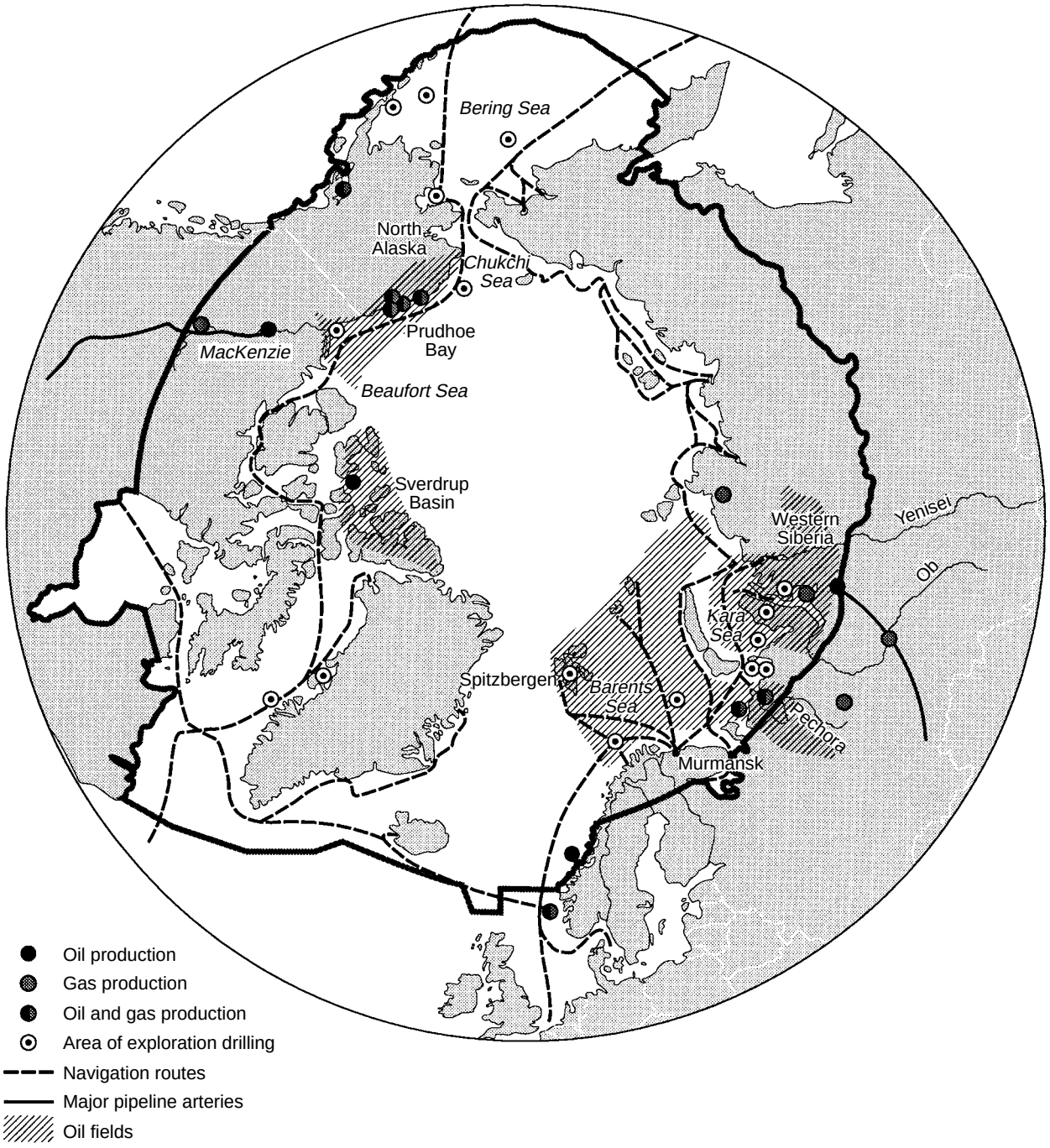


Figure 1 Main oilfields, production and exploration wells of the Arctic showing pipelines and navigation routes

15 and 1500 tonnes, depending on year (Melnikov *et al.*, 1996).

Local Municipal Sources

Most human settlements in the Arctic are low in population and generally constitute a limited source of petroleum hydrocarbons. Exceptions occur in the Russian Arctic where cities can be large, industrial activity significant, and only poor infrastructure development exists to deal with municipal and industrial wastes. For example, Murmansk (pop. 500 000) on the Kola Peninsula has no municipal waste treatment facilities at all. Annually, it discharges about 65 million m^3 of untreated waste water containing 56 tonnes of petroleum products directly into Kola Fjord. Total annual discharge of wastewater to the Siberian Shelf

coastal zone from coastal settlements is approximately $36 \times 10^6\,m^3$ and $21 \times 10^6\,m^3$, untreated and treated, respectively (AMAP, 1998).

CONCENTRATIONS OF PETROLEUM HYDROCARBONS IN THE ARCTIC ENVIRONMENT

The largest quantity of data for petroleum hydrocarbons in the Arctic pertains to Russia. The main reason for this is not that petroleum hydrocarbons are necessarily more problematic in Russia, but rather that Russia is the only circumpolar nation that has conducted systematic, long-term monitoring of petroleum hydrocarbons in the Arctic.

The geographical coverage of samples for petroleum hydrocarbon levels across the Arctic is generally poor because most sampling has been limited to relatively accessible locations close to sites of human activity. Samples are more commonly available from the aquatic environment (marine and freshwater) than from the terrestrial environment. There are especially few measurements of petroleum residues in biota as such measurements have not been emphasized in most sampling programs. This is largely because petroleum hydrocarbons are not generally persistent in the environment. Additionally, although many of the petroleum compounds, particularly aromatic hydrocarbons, are rapidly bioaccumulated, they are also readily metabolized by many organisms, particularly fish and other vertebrates.

Direct comparison of the values obtained by the various programs which have measured petroleum hydrocarbons in the Arctic are often impossible because of typically wide differences in the analytical methods, instrumentation, quantification, and reporting used to produce the results. Some values, particularly those produced by non-specific analytical techniques, tend to overstate the amount of petroleum hydrocarbons present in a sample. Other values, especially those which are based on the sum of individual compounds, tend to understate the total amount of petroleum hydrocarbons present. The lack of consistency among the data sets severely limits their usefulness for analyses of spatial and temporal trends.

It is also important to note that few of the petroleum hydrocarbon compounds, which comprise crude oil, are unique to crude oil. Many have terrestrial biogenic, marine biogenic, pyrolytic or diagenetic sources. Consequently, values reported as petroleum (or oil) residues often include components from non-petroleum sources.

Concentrations in the Arctic Marine Environment

In general, the concentrations of dissolved petroleum hydrocarbons (detected by infrared) for samples from the Russian Arctic are higher in river estuaries than at offshore locations. The concentrations are also generally higher in surface than in subsurface waters. Both trends strongly suggest anthropogenic sources.

In the shelf seas, dissolved petroleum hydrocarbon concentrations are usually highest just off of the river mouths, and tend to decrease with increasing distance from these areas. Reported concentrations of dissolved petroleum hydrocarbons in marine waters of the Russian Arctic are generally much higher than those found in North American areas.

The highest concentrations of petroleum hydrocarbons are observed in the vicinity of seaports, such as in Kola Fjord near the port of Murmansk. During winter, when the waters of the Kola Fjord are relatively stagnant, concentrations of total hydrocarbons in surface water and bottom water exceed the maximum permissible concentration (MPC) of $50\,mg\,l^{-1}$ set by the Russian Federation by factors of 150 and 36 126, respectively. The average concentrations of paraffin oil in surface water and bottom water are five times the MPC and 35 times the MPC, respectively. In contrast, during summer, when water movements throughout the fjord are greatly increased, concentrations of total hydrocarbons seldom exceed two times the MPC (ACOPS, 1995).

The concentrations of total petroleum hydrocarbons in marine sediments from the Russian Arctic are highly variable. The highest concentrations have been reported from the Ob and Yenisei estuaries, and the Baydaratskaya Gulf.

The few measurements of dissolved, low molecular weight hydrocarbons in the Canadian Beaufort Sea indicate that methane is relatively high for marine waters and highest in concentration near the seabed, suggesting that the sediments are a methane source. Concentrations of methane, ethane and ethylene in some samples from the Chukchi Sea, Bering Sea and Norton Sound also show a near-bottom enrichment, also indicating a sedimentary source for these compounds. As high ethane, propane and n-butane concentrations were not found in conjunction with the high methane concentrations in the American and Canadian Beaufort Sea samples, petrogenic activities can be ruled out as a source. Very few measurements have been made of alkanes in seawater because background concentrations are generally well below historical detection limits. The few data that do exist, however, indicate individual alkane concentrations are in the low part per billion range and are derived from terrestrial plants, not crude oil.

Taken as a whole, the reported concentrations for alkanes, aliphatics and petroleum hydrocarbons in arctic marine sediments are highly variable in part because few have been normalized against particle size. Interpretation of sedimentary hydrocarbon assemblages generally strengthens the conclusion that the hydrocarbon

content off most coastal and shelf sediments usually reflects multiple sources: specifically, there are marker compounds present which are characteristic of both crude oil and terrestrial plant materials (Thomas, 1988; Steinhauer and Boehm, 1992; Yunker *et al.*, 1993).

Low levels of petroleum hydrocarbon contamination have been detected in various biota samples (including flounder, sculpins and zooplankton) throughout the arctic marine environment.

Concentrations in the Arctic Freshwater Environment

It is well documented that the Russian Arctic rivers, especially in areas of oil and gas exploration and production such as the lower reaches of the Ob River, are heavily contaminated with petroleum hydrocarbons. Data from the Russian Federal monitoring network show that mean concentrations of hydrocarbons in water samples obtained from this river, in the areas of the most extensive oil extraction activities, and determined with IR-spectrophotometry, exceeded the established standards for freshwater (MPC $= 50\ \bullet \mathrm{mg\,l^{-1}}$) by 12 16 times, although it is generally accepted that for analytical reasons these values are over estimated (AMAP, 1998). By contrast, dissolved petroleum hydrocarbons in the Mackenzie River occur at sub-nanogram to low nanogram per liter concentrations, which are considered to represent background (Erickson and Fowler, 1987).

Mean concentrations of petroleum hydrocarbons in bottom sediments at the mouths of Russian Arctic rivers during the period 1991 1993 were in the range of about 30 60 $\mathrm{mg\,l^{-1}}$, and thus similar to data from the Mackenzie River, and rivers along the US Beaufort Sea Coast. The highest concentrations were found in the Pechora and Ob Rivers.

A survey of petroleum hydrocarbon contamination of the bottom sediments in the Pechora River Basin was made in 1994 1995, in connection with the Usinsk oil spill. In the immediate vicinity of the oil spill site, the concentration of oil was approximately 1.4%. In the mouth of the Kolva River, the hydrocarbon concentration decreased to 0.05%. Farther downstream, however, where the Usa River flows into the Pechora, the bottom sediments were strongly contaminated with hydrocarbons (1.6%), a result of particular hydrological conditions of the river at this site which are favorable for sediment accumulation. Farther downstream, the concentrations did not exceed 100 $\mathrm{mg\,l^{-1}}$. These observations emphasize the importance of factors such as particle composition and hydrological conditions in influencing the distribution of petroleum hydrocarbons in rivers and also proximity of sources; even relatively large volume oil inputs are primarily a local-scale environmental phenomenon.

Concentrations in the Arctic Terrestrial Environment

Data on the concentrations of petroleum hydrocarbons in soils are only available for Russia. Background levels appear to be approximately 10 40 parts per million. A great potential for pollution of the Russian Arctic terrestrial environment is associated with spills and leakage from oil pipelines. A recent example is the 1994 Komi spill in the drainage area of the Pechora River. A survey of the spill site made in the autumn of 1994 showed that the total area affected by accidental and operational releases was 1 320 000 $\mathrm{m^2}$, with a mean level of soil contamination of 0.3%. The concentration of oil in the soil 5 m from the main leakage site was 15%.

Further large volume losses of oil from pipelines in the Russian Arctic seem likely. Almost all of the oil and gas is transported from the extraction fields to processing plants by pipelines, most of which have exceeded their design life and have not been well maintained over the years.

BIOLOGICAL EFFECTS FROM OIL SPILLS

The potential effects of petroleum pollution on the various arctic ecosystems (terrestrial, freshwater and marine) depend on a number of factors including type of oil, size of spill, time of year (temperature, photoperiod, life cycle stage of biota), location (proximity to sensitive and vulnerable biota), etc. The duration and severity of the impacts will generally depend on how all the factors combine to influence oil spill countermeasures and to influence the many weathering processes, which begin to modify the oil immediately after it is spilled. These factors include evaporation, photochemical degradation, dissolution, bacterial mineralization, emulsification, adsorption and mixing. In general, biological effects of oil spills are similar in tropical, temperate and arctic ecosystems for related or similar types of biota. The important environmental differences are those, which affect the distribution, composition, and physical state of the petroleum hydrocarbons in space and time, i.e., their bioavailability. Typical biological effects can be lethal and sub-lethal. They are expressed in the ecosystem by loss of certain species; changes in species biomass, diversity, distribution and abundance; effects on reproduction and growth; and short-term and long-term toxic physiological effects. The severity of any effects is usually thought to be greater in arctic ecosystems than in more temperate ones because it will generally take arctic ones longer to recover from oil damage. This is because of typically slower growth rates, shorter productive season lengths, longer generation turnover times and longer ages to maturity.

An oil spill on land generally will be more confined than a spill in the aquatic ecosystem. However, the duration

of exposure and effects could be longer in the terrestrial setting if the oil migrates downward through the soil where degradation is inhibited.

A similar protracted spectrum of effects is also possible in the marine ecosystem if oil is spilled under multi-year ice. In this case, the oil remains practically unchanged until the ice thaws creating the possibility of extensive spreading and longer exposure periods to toxic components of the crude oil.

It is also important to note that due to the fragility of the arctic ecosystems (particularly the terrestrial and freshwater ecosystems) a quicker recovery could be achieved from a spill by not further damaging sensitive environmental components as a result of the countermeasures effort.

Significant oil spills can also have potentially serious social implications in the Arctic. Many of the indigenous peoples in the circumpolar countries have diets and social structures that rely heavily on the use of local flora and fauna. This traditional reliance on the land together with the related hunting and gathering activities provides a rich cultural heritage and folklore, strong social interactions and a strong sense of identity among indigenous peoples of the Arctic. Consequently, long- or even relatively short-term displacement or alteration of the access to traditional food items caused by oil spills or the resulting cleanup operations could have far reaching consequences.

POTENTIAL EFFECTS OF GLOBAL CLIMATE CHANGE

It is commonly accepted that the effects of global climate change will be considerably greater in Arctic regions than on a global average. Measurements over the past 40 years, for example, confirm that the rate of global warming has been greatest between 40 °N and 70 °N, thus overlapping the area between approximately 60 °N and 70 °N where most of the oil and gas activity occurs.

Climate warming will have a profound effect on the distribution and characteristics of sea ice, snow cover, tundra, surface water drainage, permafrost and vegetation, among others. This will lead to new spatial and temporal overlaps between Arctic biota and present and future oil and gas activities, thereby having a significant influence over how and when exploration, construction and production of oil and gas reserves occurs. It will probably also increase the impact of oil spills and the subsequent cleanup activities. Sound science-based decision making and the use of proven, appropriate engineering practices will be essential to ensure that development does not occur in a way that exceeds the ability of engineering and management to operate safely and in an environmentally friendly way in the Arctic.

Two other concerns and research priorities should be mentioned:

1. The arctic permafrost may weaken or even collapse in a warmer climate, releasing methane and creating problems for pipeline construction and overland transport (*see* Permafrost, Volume 1).
2. Methane clathrate deposits will need to be investigated (*see* Methane Clathrates, Volume 1).

REFERENCES

ACOPS (1995) *Identi cation and Assessment of Land-based Activities in the Russian Federation that Contribute to the Degradation of the Arctic Marine Environment*, A report prepared by the Arctic Working Group of the Advisory Committee on Protection of the Sea (ACOPS), Moscow.

AMAP (1998a) *AMAP Assessment Report: Arctic Monitoring and Assessment Programme (AMAP)*, Oslo, Norway.

AMAP (1998b) *Petroleum Hydrocarbons* http://www.amap.no/assess/soaer/Q.htm.

GESAMP (Joint Group of Experts on the Scientific Aspects of Marine Pollution) (1993) *Impact of Oil and Related Chemicals and Wastes on the Marine Environment*, IMCO/FAW/UNESCO/WMO/IAEA/UN/UNEP, Joint Group of Experts on the Scientific Aspects of Marine Pollution, Reports and Studies No. 50.

Melnikov, S A, Vlasov, S V, Ryzhov, O V, and Semushin, D S (1996) *An Overview of the Results of Monitoring of the Levels of Petroleum Hydrocarbons and Polycyclic Aromatic Hydrocarbons in the Environmental Compartments of the Seas and the Coast of the Russian Arctic in 1989- 1995*, Federal Service of Russia for Hydrometeorology and Environmental Control, St. Petersburg.

MMS (Minerals Management Service) (1996) *Outer Continental Shelf Oil and Gas Leasing Program: 1997- 2002 (draft)*, Minerals Management Service (Document 95 0061), US Department of the Interior, Herndon, Virginia.

NPD (Norwegian Petroleum Directorate) (1995) *Petroleum Resources on the Norwegian Continental Shelf*, Norwegian Petroleum Directorate Report, February 1995.

Steinhauer, M S and Boehm, P B (1992) The Composition and Distribution of Saturated and Aromatic Hydrocarbons in Nearshore Sediments, River Sediments and Coastal Peat of the Alaskan Beaufort Sea: Implications for Detecting Anthropogenic Hydrocarbon Inputs, *Mar. Environ. Res.*, 33, 223 253.

Thomas, D J (1988) The Tuktoyaktuk Harbor Benthic Biological Monitoring Programme, A report prepared by Seakem Oceanography Ltd. for Environment Protection (DOE), Yellowknife, NWT.

Yunker, M B, MacDonald, R W, Cretney, W J, Fowler, B R, and McLaughlin, F A (1993) Alkane, Terpene, and Polycyclic Aromatic Hydrocarbon Geochemistry of the Mackenzie Shelf, Riverine Contributions to Beaufort Sea Coastal Sediment, *Geochim. Cosmochim. Acta.*, 57, 3041 3061.

Oil Fires: Kuwait

Tahir Husain
Memorial University of Newfoundland, St. John's, Canada

The Gulf War aftermath witnessed two historical environmental episodes, i.e., the largest oil spill on record and torching of more than 700 oil wells in the Kuwaiti oil fields. Other impacts were those associated with the damage to the terrestrial environment caused by massive troop movements, the release of ammunition during ground and air battles, potential release of toxic materials into the environment during the war, and the deployment of land and sea mines. The estimates of the volume of oil spilled into the marine environment as a result of the Gulf War ranged from 7 to 11 million barrels. Over 640 km of shoreline extending from Kuwait to Abu Ali Island in Saudi Arabia was impacted by the oil slick.

The problems arising from the torching of more than 700 oil wells in the Kuwaiti oil fields in February 1991 are discussed here. According to the estimates by the Kuwaiti Oil Company (1991), 605 oil wells were on fire, 108 wells were damaged and 46 wells were gushing oil on a continuous basis. Approximately 22 000 tons of sulfur dioxide (SO_2), 18 000 tons of soot, and thousands of tons of carbon monoxide and oxides of nitrogen (NO_x) were emitted by these wells as a result of 2.5 million barrels of oil lost initially on a daily basis (Husain, 1995). The unburned and partially oxidized oil was separated into: (a) heavy drops which fell to the ground within the vicinity of the fire and contributed to the formation of oil lakes, and (b) fine droplets of oil which behaved like a spray which traveled with the plume for some distance and then fell to the ground.

EFFECTS ON THE TERRESTRIAL ENVIRONMENT

The most obvious environmental problem during the Gulf War was due to the impact caused by the largest oil spill on record and torching of more than 700 oil wells in the Kuwaiti oil fields by the Iraqi troops before the conclusion of War in February 1991. The plume from the burning oil field as shown in Figure 1 was towards the south and southeast covering the eastern part of the Kingdom of Saudi Arabia. The environmental consequences of the Persian Gulf War on the terrestrial environment in the region can be summarized as follows (Husain, 1995; Husain and Amin, 1994; WMO, 1992):

1. The smoke plume was usually below 2 km and was never detected above 6 km from the base of the oil fires. The particulate counts within a few kilometers from the source were very high ($10^5\,cm^{-3}$) but decreased with distance. At about 300 km from the fire the particle counts were in the range of 5000 to 15 000 cm^{-3} while beyond 1000 km from the source, the counts decreased to the range of 300 500 cm^{-3}.
2. The plume was usually not photochemically active until it had traveled over 200 km. Within the plume at about 1000 km downwind, the ozone concentration was found to be relatively high but there was no indication of a regional oxidant episode.
3. The meteorological conditions in the area during March to October 1991 were such that materials in the plumes were largely confined to the horizontal cloud without much mixing down to the surface. Hence, pollution levels at population centers in the region, although higher than normal, were less than initially predicted.
4. The deposition of soot on the ground surface within 100 km from the source was more than 500 mg m^{-2}. Between 100 and 300 km downwind from the source, the deposition was estimated to be in the range of 100 to 500 mg m^{-2}. The soot deposition level between 300 and 500 km radius downwind was found to be in the range of 50 to 100 mg m^{-2}. The effect of soot from oil fires was observed even beyond 500 km from the source.

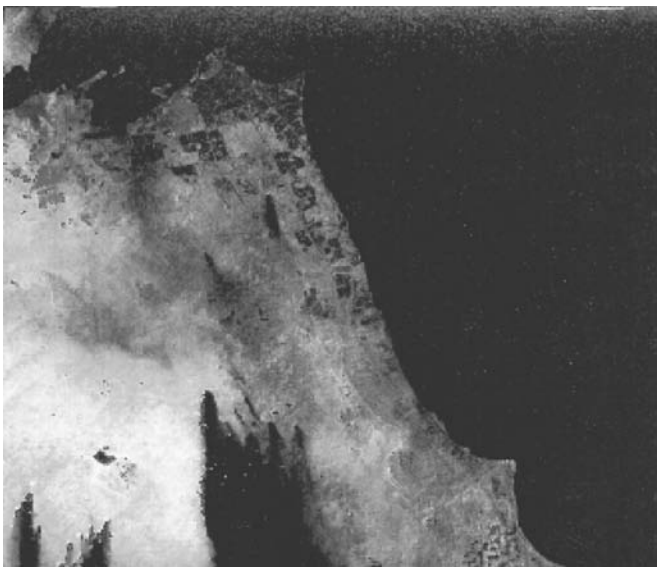


Figure 1 Plumes from Kuwaiti oil fields during oil fires in 1991

Soil samples were collected from different locations near the source of the oil fires and from other locations in Kuwait. The analysis of these samples showed that the concentration of organic constituents on the surface of the soil was much higher than in the deeper soil (Husain, 1995).

Table 1 Comparison of gaseous pollutants in post-war (1991) and pre-war conditions

Stations	SO_2 ($\mu g\,m^{-3}$)	NO ($\mu g\,m^{-3}$)	NO_x ($\mu g\,m^{-3}$)	Non-methane hydrocarbons ($mg\,m^{-3}$)	Methane ($mg\,m^{-3}$)
Mansoria (30 km north of Burgan oil field)					
Pre-war 1990	4.69	30.00	77.60	0.29	1.15
1991	5.99	16.40	43.20	1.33	1.74
Increase (as a %)	+28	−45	−44	+352	+50
Rega (about 15 km northeast from Burgan oil field)					
Pre-war 1990	3.65	35.20	83.60	0.27	1.13
1991	7.82	120.40	161.60	0.41	1.09
Increase (as a %)	+114	+242	+93	+48	−4
Average increase (as a %)	+71	+98	+25	+200	+23

The concentration of inhalable particulates was high and in many cases it was above the permissible limit of $340\,\mu gm^{-3}$. On the other hand, the level of gaseous pollutants was usually lower than the permissible limit but comparing the value in 1991 during oil fires with the prewar situation, it was found that the pollutants were elevated by 200% in some cases. Table 1 lists the concentration levels during the Gulf War and the prewar situation at two locations in Kuwait.

EFFECT ON WEATHER AND THE MARINE ENVIRONMENT

About 10% of the sun s radiation was transmitted through the thickest part of the super composite plume, 75 80% of the sun s radiation was absorbed, and the remainder was scattered by the smoke in the super composite plume (Hobbs and Radke, 1992). A temperature drop of up to 10 °C in Kuwait and in northern Saudi Arabia was recorded. Seawater temperature in the Gulf was considerably lower during 1991 as compared to previous years. The measurements at about 250 km from Dhahran towards Kuwait showed a decline of 2.5 °C in mean sea temperature in 1991 mainly because of the smoke plume covering the area. Total solar radiation in the Arabian Gulf coast near Dammam in 1991 was only 79% of the 1990 recorded value. Due to the reduction in mean temperature and solar radiation, fisheries in the Arabian Gulf were severely affected (McCain *et al.*, 1993). The shrimp catch by the Saudi Fisheries Company in 1992 was significantly reduced compared to the level before the Gulf War, with estimated loss to the prawn fisheries of US$55 millions. In 1992, the catch of adults and juveniles was only $1.5\,tons\,boat^{-1}\,year^{-1}$ compared to 195 205 $tons\,boat^{-1}\,year^{-1}$ in 1989 1990.

The spawning biomass index also fell to about 1.8% of its pre-war year level. The prawn landing fell to 2.5 tons in the first half of 1992 compared to 4000 tons in 1989 (Mathews *et al.*, 1993). A plankton survey showed that eggs/larval abundance declined significantly in 1992. One of the causes of such a decline in prawn population was the decrease in solar energy on the prawn spawning grounds, the deposition of soot on the water surface and the oiling of tidal and sub-tidal areas in the region.

CONCLUSIONS

The air pollutants from the Kuwaiti oil fires were complex mixtures of many organic and inorganic compounds. Some of these compounds were toxic to human health and their concentrations were elevated significantly compared to pre-war and post-war conditions. A detailed investigation should therefore be conducted on the long-term effects of these pollutants by a reliable quantitative risk assessment methodology incorporating emission estimate, assessment of exposure levels, dose-response relationships, and risk characterization.

REFERENCES

Husain, T and Amin, M B (1994) Kuwaiti Oil Fires Particulate Monitoring, *Atmos. Environ.*, 28(13), 2235 2248.

Husain, T (1995) *Kuwaiti Oil Fires: Regional Environmental Perspectives*, Pergamon Press, Oxford.

Hobbs, P V and Radke, L F (1992) Airborne Studies of the Smoke from the Kuwaiti Oil Fires, *Science*, 256(15), 987 990.

Mathews, C P, Kedidi, S, Fita, N I, Al-Yahya, A, and Al-Rashhed, K (1993) Preliminary Assessment of the Effects of 1991 Gulf War on Saudi Arabian Prawn Stocks, *Mar. Pollut. Bull.*, 7, 251 271.

McCain, J, Beard, D W, and Fadlallah, Y H (1993) Influence of the Kuwait Oil Well Fires on Seawater Temperature in the Western Gulf, *Mar. Pollut. Bull.*, 27, 79 83.

WMO (1992) *Report of the Second WMO Meeting of Experts to Assess the Response to and Atmospheric Effects of Kuwaiti Oil Fires*, Geneva, Switzerland.

Oil Industry: Cultural Impacts and Environmental Risk Tolerance

William R Stanley
University of South Carolina, Columbia, SC, USA

Petroleum is extracted in company-controlled concessions, obtained through marketing arrangements with national oil companies, especially in the important producing states, or purchased on the open or spot market from other producers. Pro t margins, however, generally are highest when the petroleum is corporate controlled, from producing eld to transport to re nery and eventually, to sales at the retail gas station or equivalent. Diversity in supply is standard for industrial operations and the oil industry is no exception. While it has long been an accepted facet of the petroleum industry that exploration for oil is worldwide in scope, some producing regions are more important than others. Provided there would be suf cient economic incentives, coupled with the absence of political constraints in the oil importing countries, the proven gigantic reserves and relatively low production costs would make it possible for the Gulf oil producing countries alone to provide for the needs of the entire world into the next several decades, possibly longer. Were this the case, one might legitimately ask why petroleum is sought and extracted in such physically hostile and monetarily expensive environments as the Arctic north of Alaska, the depths of the North Sea and the Grand Banks of Newfoundland or semi-impenetrable tropical forests of three continents? The answer is alarmingly simple. The day is past when oil is found seeping from easily tapped, above ground ssures. Petroleum companies have but one basic product to sell, and corporate nancial health is contingent upon a continuous ow of the raw material. Since production elds on the scale of the ones in eastern Saudi Arabia or in the Caspian Sea region are the exception and new ones of this magnitude are increasingly unlikely to be discovered, the industry seeks economic and political risk diversity in supply. In some cases, this means exploring for oil in physically remote or hostile environments. It is a given fact that the cost of nding oil in such locations and bringing it to market is certain to be high.

OPERATIONS IN THE ARCTIC AND TROPICS

Environmental degradation, possibly resulting from the extraction of oil in diverse climatic, topographic, vegetative, cultural and maritime zones, varies by site and sometimes by nationality or size of the oil company. Weak governments in search of economic development at almost any cost clearly are attractive situations for petroleum companies not overly concerned for pollution. That this once was the case cannot be disputed. Today, however, oil companies share a common bonding in that, if one is guilty of some infraction or has been careless to the extreme, then all are painted with the same ugly brush. There have been well-publicized occasions when the so-called ugly brush carried severe financial penalty. Therefore, it is in all of their interests that adequate concern be given to the environmental impacts of their operations and to the level of technical response on hand in case of emergencies. This is not to suggest that sloppiness has been vanquished or that accidents never will occur. Rather, it reinforces the notion that all petroleum operators have a financial and moral obligation to insure environmentally friendly operations to the maximum extent possible.

Frequently more severe in the magnitude of dislocation, and usually longer lasting, are the negative cultural impacts to the local community resulting from the exploration and production of hydrocarbons. These negative impacts also are found associated with the routes used to transport the petroleum or gas to the next stage in the production marketing continuum. This culturally disruptive face of the industry is relatively late in surfacing in the public s consciousness. Nevertheless, it is now being taken seriously, often more so than the better publicized environmental issues. One problem is to be able to identify the potential negative issues prior to the commencement of a project. This is not always easy. Indeed, some projects have been extolled in the first instance by the leadership at all levels of government as offering the local inhabitants an opportunity to share in the benefits of economic change. Only later did these same proponents discover that the financial costs associated with cultural disruption outweighed any possible combination of economic benefits! Some Canadian First Nations people have recently had land claims settled, and are sending out some encouraging signals about pipeline construction.

At least one major hydrocarbon transport scheme, the Mackenzie Valley pipeline project for evacuating natural gas from the Beaufort Sea in Canada s Northwest Territories for sale in the United States, was canceled because of this potentially negative cultural impact dimension. Proponents of the scheme, however, were critical of the reasoning for cancellation. This particular project has not been resurrected and other industry projects have been delayed.

To assess the environmental and social impacts of the oil industry, two widely separated projects in contrasting physical settings are examined. One is in Nigeria (West Africa) where oil has been extracted since the 1950s and where its presence (and who would control revenues) was an important consideration in events leading to Nigeria s tragic civil war in the 1960s. Today, oil revenue is not only the mainstay of this country s foreign exchange, but it has

also been responsible for substantial change and possibly even some economic development. Nevertheless, Nigeria is in a state of almost constant turmoil and the quest for oil and oil revenue has negatively impacted those living near production facilities. The second case study is Alaska s North Slope, astride the coastline facing the Arctic Sea. Here, the study focuses on the important issues behind exploration and production in a pristine but climatically harsh natural environment. Nigeria is densely populated and many people reside in close proximity to petroleum operations. In contrast, the indigenous population in this northernmost portion of Alaska is numerically small and, until recently, generally lacking in political influence. Fewer still reside in close juxtaposition to petroleum operations.

Case Study – Nigeria

Nigeria is a developing country where the great mass of the approximately 100 million population lives close to poverty. The oil industry has operations both in the rainforest and adjacent offshore areas. This country has suffered economic decay, political collapse, military dictatorship, religious antagonism resulting in serious loss of life, decline in agricultural production, urban crises and, ironically, persistent shortages of gasoline at the retail pump due to the large volumes of refined products shipped illegally to neighboring countries or otherwise entering the black market. Several major on-shore and off-shore oil spills have occurred and illegal tapping of a gas pipeline connecting a refinery near the southern city of Warri to northern markets has resulted in two recent fires that claimed the lives of scores, perhaps hundreds of people. Attempts to obtain precise casualty figures floundered on the unwillingness of survivors to report to hospitals for fear of arrest.

Oil revenue has fashioned a remarkable economic landscape. Four-lane highways crisscross the country and schools and hospitals are readily observable in most urban and in many rural places. Particular to this analysis is the impact of petroleum operations at the local level. A new administration building or road may mean less than the prospect of steady employment and attractive wages with an oil firm. Consider a villager living near a constantly burning oil flare. Environmental stress is all the more intense if that person or family has no direct link with the oil firm, such as employment. He assumes that the oil company is taking care of itself financially, for why is it here. He realizes that government may not support his village in practice, irrespective of what is said in the state capital or in Lagos. His children likely have less opportunity for obtaining any of the coveted school or university scholarships than do children in the city.

Whether government cannot deliver the desired level of services or whether the money for providing such services goes astray somewhere along the way is not the question. The villager knows what is being done, the means used to accomplish the task(s), and what is not being done. He is just as apt to listen to BBC s World Service radio broadcast as is someone in the city. To him, the manner in which the oil company is operating is nothing less than a marvel of efficiency. Wages are paid in full, and aside from mandatory and well-publicized deductions, the worker receives full accounting. For better or worse, life has been substantially changed. At the same time, he realizes that the village clinic is in the Dark Ages when compared to the nearby company clinic, and while he and his family may have access to the company health service, there are others here and elsewhere not so fortunate.

Set in motion is at first a trickle, then a stream of requests to the oil firm s management for improving the local setting: books and furniture for the school, an access road, drinking water from wells and more. Government is responsible for these but the villager knows only too well that government is slow to respond. He may even suspect that the oil company is not the responsible party for providing such basic services. He probably appreciates that some of his requests and those of others are an imposition upon the company. By the same token, he assumes that he is simply using the company as a proxy of government. It is better to get the task completed, and why not, for the company surely will charge government for the cost of any socio-economic or infra structural assistance. Government will be paying to have done what it is unable to do itself or has not accomplished because the necessary funds never reached their destination. The oil company is a means often the only means to gain the attention of Government!

The oil company manager emphasizes the fact that his company can help Nigeria best by doing an efficient job in the locating, extracting and marketing of oil. The better the performance, the more taxes and royalties are paid to government, and, as a result of these taxes, the more services can be provided by government. Nevertheless, there has been a tidal wave of raised expectations because of oil and, because of the weak existing infra structural base students lacking classrooms or books, unsanitary drinking water, unreliable utilities, abdominal communications and inadequate medical care development is seen to be lacking.

> We are operating in areas where need is everywhere; we have both a professional and humanitarian wish to be of assistance. Furthermore, being of assistance is good business practice and each of the companies in Nigeria budgets funds for good neighbor services. This we accept as the price of doing business, and no less important, because indeed we are good neighbors. Let s face it, if the village outside our compound contracts the plague, then our employees also are at risk. It is better to help out in the health and sanitation sectors from the outset. I employ the same logic for helping to furnish the local primary school or in assisting in providing electricity or in making available

> clean potable water. Frankly, I don t feel the same sense of urgency for a village 3 km distant.
>
> (Stanley, 1990)

Close observation of petroleum operations in Nigeria suggests that residents frequently tolerate relatively severe environmental degradation if economic opportunity is present. Unfortunately, the latter is available to only a relatively small component of the impacted population. Even so, oil spills, fish kills, destruction of vegetation or air pollution does not seem to generate the sometimes vitriolic responses experienced in more affluent parts of the world. Pollution generally is tolerated, if not accepted, due to lack of empowerment and questionable local concern for the environment. In 1985, an analysis of 1100 questionnaires offered to Nigerians residing in the oil impact area showed that nearly 77% of respondents answered in the negative when asked if they or their families had benefited personally from oil operations (Stanley, 1990). The tropical forest regions of the world too often are used to cover or hide man s littering of the landscape. Pollution can be significant but seldom so severe that payments to the most affected cannot resolve the problem locally. Littering and misuse of the landscape is no less severe in arctic regions, but here the effects are more difficult to hide and, in Alaska, the most affected populations often have vocal and influential spokespersons not to mention friends and supporters in and out of government.

Case Study – North Slope of Alaska

Alaskan oil (or its problems) is embellished in the minds of many as representing a special dimension of what is wrong with this industry. The reader is reminded that the Exxon Valdez disaster was maritime in origin. The knowledge that it could have occurred almost anywhere in the three-quarters of the earth s surface covered by water does not lessen its connection to Alaska, nor to the company and port whose names the oil tanker carried. That said, let us go back in time to 1969 when there was no pipeline connecting the Prudhoe Bay field to Alaska s ice free and year-round maritime accessible south coast and when the only hydrocarbon activity at Valdez was the occasional arrival of a small products tanker for local gas stations and the regular departure of liquefied natural gas vessels carrying Alaska s natural gas to Tokyo for electricity generation.

By the mid-1960s, it was clear in industry circles that the Alaska North Slope oil reservoir in the vicinity of Prudhoe Bay was substantial if not world class and, as such, well worth pursuing in economic terms. A major handicap was climate. The northern areas of Alaska are bitterly cold in winter and the offshore Beaufort Sea is frozen for much of the year. The issues, then, were two-fold. First would be the development of a townsite, extraction field, pumps and pipeline and storage capacity for the oil removed through drilling. Second was the problem of how to evacuate the oil once it was extracted. The first issue was daunting but something the petroleum industry could resolve once sufficient resources were directed to establishing a permanent facility at Prudhoe Bay. The second issue was more perplexing and, in retrospect, proved to be a genesis of the Exxon Valdez disaster.

In the summer of 1969, Humble (Exxon subsidiary) Oil s largest US Flag Tanker, the 119 000 dwt SS Manhattan had completed an unusual refit. The tanker s bow was reinforced with so much steel that the ship acquired icebreaker qualities. All other steel faces of the vessel were reinforced in order to withstand the punishing effects of ice. These decisions were made at least three years earlier and were predicated on the knowledge that proper development and steady oil extraction in Alaska s North Slope required a means of shipping the product on a scheduled basis. At the time, the Pacific Coast of the US, unlike its eastern seaboard, was not experiencing shortages of oil (gasoline). There was and still is not a trans-continental petroleum pipeline, one that would permit the shipping of oil from one coast to another, from surplus regions to areas lacking sufficient petroleum. Oil from Alaska was not meant to serve markets in the west. Thus, a maritime route through the Bering Strait and into the Beaufort Sea might have been technically feasible but impractical since the petroleum still would be on the Pacific Coast. There is, however, a maritime route to northern Alaska by way of the North Atlantic, from the east. This is by way of Newfoundland and the Labrador Sea, the Davis Strait separating Baffin Island and Greenland, Lancaster Sound to the south of Devon Island and, once through, the use of two possible routes into the Beaufort Sea. The difficulty is that these waters, painted blue on maps, are frozen solid for all but several weeks of the year. It was in preparation for such unusually severe physical conditions that the SS Manhattan was fitted-out and deployed for its epic journey.

The voyage was timed for late summer departure from Houston, which would bring the Manhattan into the deep Arctic at a time of maximum melting and ice movement away from shore (northward). Two icebreakers accompanied the Manhattan, one Canadian and one American. The voyage was a press extravaganza. The Manhattan carried scores of scientists to test practically everything, biologists to report on the fauna, oceanographers, climatologists and, not least, a constant presence of Canadian officials. The political consequences of this voyage have been long lasting, not least of which is the introduction of a permanent Canadian military force on Baffin Island. Not only were some officials in Ottawa concerned for their southern neighbor and how she might respond to a long term use of Canadian Arctic waters for a critical resource, there also was some concern for such a massive territory assigned to Canada on atlases but almost totally lacking in Canadians.

This large hunting and gathering culture realm, to which much of Alaska can be joined, contains few residents and they live in widely dispersed villages, government centers or mining camps.

The voyage northward was not uneventful and on several occasions the tanker was ice bound, dependent upon icebreaker assistance to break free. In spite of the reinforced steel plates, punctures from the ice were recorded. Eventually, but only after much travail, the SS Manhattan anchored offshore from Prudhoe Bay. This voyage was a technical success but served to convince oil company executives that an alternative route for evacuating oil would be necessary if ever production were to commence in the North Slope.

There were several reasons behind this decision and all revolved around the almost impossible task of guaranteeing ship arrivals at Prudhoe Bay. Scheduled tanker visits are necessary for the orderly removal of oil already extracted and stored in on-shore holding tanks. Drilling and storage in the Arctic requires heating, without which the oil would freeze once brought to the surface. Imagine the consequences if ice were to prevent tankers access to the region, and the holding tanks at Prudhoe bay no longer had storage capacity. Drilling operations would have to be halted with enormous costs associated with restarting the process. Pack ice deeper than usual or winter storms not only would hinter tanker operations; such conditions would enhance the probability of punctures or other types of maritime disaster. Envision a fully loaded tanker caught up in extreme winter weather conditions and suffering even a modest oil spill. A comparable spill in the Persian Gulf is bad, but at least the physical environment in that part of the world offers the prospects for recovery vessels and personnel to attempt a clean-up at sea. How would this be managed in a region with almost 24 hour winter darkness and in temperatures lower than minus 50 °C? Clean-up would be impossible. Imagine further the ecological consequences of an oil spill in waters with constantly low temperatures? These were the issues that caused the industry to explore alternative evacuation routes and eventually to construct a pipeline connecting the North Slope to the ice free port of Valdez.

At the time and prior to the Exxon Valdez, primary concern was for the multi-billion dollar pipeline and for the twin issues of interference with traditional animal migration routes and what heating the pipeline might mean for permafrost stability. In retrospect, the evacuation of oil by pipe has been relatively incident free, certainly nothing on the scale of the tanker grounding. Clearly, however, there have been cultural dislocation problems amongst the indigenous population and a powerful environmental lobby at this writing has succeeded in preventing further oil and natural gas developments in northern Alaska.

FUTURE PROSPECTS

Political turmoil is evitable and national energy security issues are bound to resurface. Automobile manufacturers undoubtedly will allocate increasing resources to developing marketable electric or electric hybrid vehicles. Electric utilities will endeavor to reduce their dependence on high sulfur coal and petroleum (which essentially guarantees greater use of nuclear generated electric power). Wind generated power will gain more adherents and increased spending; tidal, geothermal and solar power will have a greater role in future energy use. Nevertheless, fossil fuel consumption can only increase as greater numbers of people in developing countries seek and achieve the good life. Those residing in countries with high current levels of energy use give no indication of a collective willingness to tighten the belt and reduce energy consumption. An ever-increasing demand for petroleum and all forms of energy can only lead to a future scenario of supply disruptions, environmental degradation and outright military struggles for still existing oil reserves. The world is constantly stretching safe economic and environmental parameters while economic growth remains dependent upon increased petroleum use. New technologies could reduce the rate of future economic environmental collapse but this is problematic. Will future observers look back to the beginning this Millennium with envy or disdain?

POSTSCRIPT

Alaskan oil is shipped to markets in the eastern US, but by way of the Panama Canal. Tankers collecting oil in Valdez are too large to traverse the Canal so the oil is off loaded or reloaded into smaller tankers at both entrances.

REFERENCE

Stanley, W (1990) Socioeconomic Impact of Oil in Nigeria, *GeoJournal*, 22, 67 79, 75 76.

Oil Shales and Tar Sands

Joe Macquaker
University of Manchester, Manchester, UK

Oil shales and tar sands are alternative sources of hydrocarbons to conventional oil, gas and coal deposits. Oil shales are any shallow buried, organic matter-rich rocks

that yield oil upon pyrolysis (i.e., heating in an inert atmosphere to approximately 500 °C). Oil shales are potential viable reserves if they generate more oil and gas than is used in the extraction process. This requirement, in practice, means that oil shales must contain at least 5% total organic carbon. Tar sands are sands and sandstones (sedimentary rocks with a grain size in the range 0.0625 - 2.0 mm) that contain abundant heavy viscous oil in their pore space. The oil in tar sands at near-surface temperatures and pressures is not mobile and, therefore, can only be extracted using specialist techniques that typically involve heating. In the past, oil shales have been extensively mined and processed as a source of oil. The first plant to extract oil from oil shale was established in Bathgate, Scotland in 1851 by James Young. This was followed by widespread shale oil industries developing in countries such as Australia (1865), Brazil (1881), New Zealand (1900), Switzerland (1915), Sweden (1921), Spain (1922) China (1929) and South Africa (1935). Despite much commercial interest, tar sands have not been commercially exploited on a large scale, although pilot plants have been developed in Canada to extract the Athabasca tar sands, and in the Orinoco region of Venezuela. In the early years of the 21st century, there are few commercial oil shale and tar sand reserves being exploited due to the high cost of production from these deposits. Worldwide, however, hydrocarbon reserves within these regions are significant, and governments and industry continuously review their potential as a resource. Plans are currently being developed to utilize oil shale resources in Queensland (Australia), Estonia and Jordan and also to expand tar sand extraction operations in Canada.

Inevitably as conventional sources of crude oil and gas run short in the mid to late 21st century, and the real cost of crude oil reaches more realistic levels, the economics of extracting hydrocarbons from oil shales and tar sands will become more favorable. It is likely, therefore, that increasing attention will be placed on these resources, in addition to coal, assuming that hydrocarbons continue to underpin the world-wide economy. The successful separation of hydrocarbons from these alternative resources without sustaining enormous ecological damage represents a real challenge.

OIL SHALES

Composition

Oil shales are most commonly siliciclastic-rich, organic matter-bearing mudstones (sedimentary rocks with a grain size • 0.0625 m, e.g., Dinantian oil shales in the Midland Valley of Scotland). A few, however, are carbonate-rich organic matter and clay-bearing mudstones (e.g., Cretaceous, Green River Formation in UT, CO and WY, US). Oil shales are characterized by having very low densities (0.95 to 1.05 g cm^{-3}) in comparison to other rock types. The organic matter component of oil shale is predominantly *kerogen* (defined as the organic fraction that is not soluble in common solvents) with minor *bitumen* (defined as the organic fraction that is soluble in common solvents). Kerogen is very variable, and comprises both material that is easily recognizable: for instance, cysts derived from green algae (e.g., *Botryococus* spp. and *Tasmanites* spp., which are particularly common in torbanite from the Dinantian of the Midland Valley of Scotland, boghead coal from the Permian of Autun, France and the Stuart Shale from the Tertiary of Queensland) and material that is amorphous, which is probably derived from bacteria. In addition to these organic components, oil shales also contain: (i) detritus derived from subaerial weathering (e.g., clay minerals such as illite, illite/smectite and quartz), (ii) productivity-derived components (e.g., shell debris), and (iii) diagenetic (cement) components (most commonly pyrite, and a variety of carbonate cements e.g., calcite, dolomite, ankerite and siderite).

The primary depositional fabrics in oil shales are rarely disrupted by burrowing organisms; thus in both field and hand specimens, they commonly appear thinly bedded and laminated (10^{-2} to 10^{-3} m scales). At an individual bed-scale, the grain-size within depositional units commonly either becomes finer upward from silt to clay, or exhibits thin alternating laminae composed of either couplets of clay and organic matter, or couplets of carbonate cement and organic matter. Compressed faecal pellets, derived from zooplankton, are commonly observed within some units.

Origin

All oil shales are deposited subaqueously in either lakes or shallow seas. The preservation of organic matter within these rocks requires a number of key prerequisites to be met. Firstly, it is important that there was high organic productivity in the water column to supply the primary organic material to the sediment water interface. Secondly, it is crucial that this organic material was not diluted by other components during deposition. Thirdly, a significant proportion of the primary organic matter must survive prior to being incorporated into the sediment. Finally, it is vital that following deposition, heating induced by burial was insufficient to drive off a large proportion of the key volatile hydrocarbons.

Primary productivity within the water column mainly occurs in the photic zone, and is controlled by the availability of nutrients (mainly phosphate and nitrate) and light. On continental shelves and in lakes, the key nutrients are supplied by rivers and, to a lesser extent, upwelling. This inevitably means that zones of highest productivity occur close to river mouths and beneath regions of local upwelling. In these settings, the proportion of organic matter that actually reaches the sediment is controlled by a

complex trade-off between absolute water depth, the depth of the photic zone, the concentration of nutrients in the surface layers, and the effects of dilution by non-organic components. These lead to most oil shale precursor sediments being deposited in regions bypassed by the main clastic supply, yet beneath zones that are nutrient enriched relative to surrounding areas.

Producing abundant organic matter is only part of the equation that leads to the formation of an oil shale. It is also vital that this primary organic material is not completely destroyed prior to being incorporated into the sediment. Organic matter is most commonly destroyed by organisms using it as a reducing agent, either in the water column or during the first few meters of burial. This degradation of organic matter is driven by a combination of organisms using aerobic (macro and microorganisms that use oxygen as an electron acceptor) and anaerobic (microbial organisms that use alternative oxidizing agents to oxygen, e.g., sulfate, nitrate, manganese) respiratory pathways. The simplest mechanism to reduce the proportion of organic matter destroyed prior to its incorporation into the sediment is to limit its exposure to prolonged microbial degradation. This is most simply accomplished by rapidly transporting it through the water column (i.e., depositing it in relatively shallow water) and then burying it quickly to a depth where it can no longer be extensively oxidized by bacteria (depths below which sulfate-reducing bacteria thrive, typically 0.2 0.5 m). In most depositional settings, rapid burial is accompanied by significant clastic dilution of the organic component, inevitably reducing the organic content of the sediment and so preventing the formation of oil shales. In oil shales, rapid burial without the organic matter being diluted by clay and silt components is most commonly caused by episodic sedimentation events (e.g., storms, floods or density currents). Organic matter preservation is also enhanced by the bottom waters containing no dissolved oxygen (i.e., being anoxic). The absence of oxygen in these settings restricts the time available for aerobic organisms to metabolize the organic matter, thus allowing a greater proportion of organic matter to reach the sediment and be buried. Deposition in depositional basins where the bottom waters are anoxic causes the sediment fabric to be undisrupted by burrowing organisms and primary lamination to be preserved.

Restricting the ability of aerobic organisms to colonize the bottom waters can also be accomplished in settings where either all or part of the water column is hypersaline. Here, faunal diversity is dramatically reduced, and the preservation potential of organic matter is commonly much greater. The positive effect of reduced benthic colonization of the water column is partially offset, however, by more sulfate being available than is present in either marine or lacustrine environments, acting as an oxidizing agent. This causes increased anaerobic microbial degradation in these hypersaline settings. Where oil shale deposition has occurred in hypersaline settings, organic matter is commonly intimately associated with evaporitic minerals. For instance, in the Green River Formation (a Cretaceous-aged oil shale from WY, CO and UT, US), organic matter is interspersed with trona deposits, suggesting that oil shale deposition here occurred at least partially in hypersaline environments.

Lacustrine (lake-dwelling) settings are particularly favorable places for the preservation of organic matter, as the concentration of dissolved sulfate is commonly lower than in normally marine settings. Anaerobic decay via sulfate reduction is, therefore, reduced and the potential for organic matter preservation is higher than in marine systems. Preservation in these locations is further enhanced because lakes are more prone to being stratified than continental shelves. This stratification causes the bottom water layers to be anoxic and to enhance the preservation potential of organic matter still further.

It is crucial that if an oil shale deposit is to be a viable resource, it must not have been heated much above 80 °C during its burial history. Above this temperature, the volatile hydrocarbon components are driven off, and the residual organic component becomes enriched in carbon relative to hydrogen, leaving it with a much lower calorific value.

Extraction

Prior to extracting oil from an oil shale, the rock has to be mined (using either opencast or sub-surface methods) and transported to the pyrolysis plant. Once at this plant, oil is typically extracted from the shale by placing crushed fragments in a retort and heating them in an inert atmosphere to approximately 500 °C. This process breaks down the kerogen into short chain liquid hydrocarbons and also liberates various gases e.g., methane, oxygen, nitrogen and sulfur compounds. In most early plants (19th century in Scotland, France) heat was supplied to the retort externally by burning coal, and thus the process was energy expensive. In later (20th century) plants, however, heat was supplied internally by either burning gas (France, Brazil), using spent hot shale (Russia) or by the insertion of hot ceramic spheres (US) into the unprocessed rock. These methods resulted in an increase in efficiency relative to the early externally heated retorts. Recently, recycling techniques have been employed that use both the surplus carbon produced after pyrolysis (coke) as a fuel, and the hot shale itself to provide the necessary heat to drive the pyrolysis reactions. Together these dramatically improve the efficiency of these most modern retorts. In spite of these advances, however, extracting oil from oil shales is still energy intensive, generates significant volumes of mineral waste (oil shale waste occupies a larger volume than the original oil shale, generates significant volumes of dust and

is depleted in many key plant-nutrients making it difficult to bioremediate), and conventional mining techniques cause significant ecological damage.

In Estonia, oil shale reserves, rather than being retorted, are first powdered and then burnt directly as fuel to provide energy for electricity generation. As oil shale is a dirty fuel, the combustion process generates large volumes of dust, and can cause atmospheric pollution because of the emission of sulfur compounds.

Reserves

Significant oil shale reserves exist in many parts of the globe. Estimates of the value of these reserves are highly variable and depend on geopolitical variables. Among the most important are those in Estonia and northwestern Russia (four billion tonnes, which are likely to be used for electricity generation), Brazil (800 billion barrels of oil), Australia (28.9 billion barrels), US (560 billion barrels), and Jordan (27 billion barrels).

TAR SANDS

Composition

Tar sands contain both hydrocarbon (15 20%) and mineral grains (80 85%). Hydrocarbons infill intergranular and grain dissolution porosity within tar sands. The hydrocarbons comprise heavy oils with a density range 0.93 g m^{-3} to 1.06 g m^{-3}, and are predominantly composed of asphaltenes and some resins, in addition to minor alkanes and alkenes. The proportion of alkanes and alkenes in the oil fraction of these rocks is significantly less than in conventional oil reserves. The mineral grains in tar sands most commonly comprise quartz, rock fragments and feldspar.

Origin

Tar sands were deposited in several environments, e.g., shallow marine, fluvial and lacustrine settings. Their original depositional setting, however, is not so important when it comes to their relative potential value as a tar sand resource. Their crucial attribute in this context is that they contain more than 15% porosity (either intergranular or grain dissolution). This prerequisite, in practice, means that they have not been subjected to extensive cementation, which usually implies that they have not been significantly buried. Once in place, the porous sandstones have the potential to become tar sands if they are charged with oil migrating from source rocks to the surface of the Earth. In the sub-surface, oil migration pathways are commonly horizontal, which causes tar sand deposits to mainly occur on the margins of petroleum-forming basins. Overall, tar sands may be considered as near-surface petroleum reservoirs.

Once established, bacteria, yeast and fungi degrade the hydrocarbons within the sands, as they are close to the surface of the Earth, in the zone affected by groundwater circulation patterns. In these environments, microorganisms are able to oxidize the volatile oil fraction, particularly the short chain (• C 20) unbranched molecules, to alcohols, ketones and a variety of acids. This biodegradation process leaves the longer chain, more complex, heavier-oil fractions concentrated in the pore space where these form the characteristic heavy oils that fill the pore spaces. This fractionation process is commonly facilitated by evaporation.

Extraction

The hydrocarbons in tar sands are extracted either by mining followed by transport to a plant and then steam extraction (Canada), or by *in situ* steam extraction (Canada and Venezuela). In the former process the tar sands are mined in large open cast pits, the raw material is then transported to a plant where the hydrocarbons are separated from the mineral component of the rock. This process utilizes a combination of treatments that involve steam and hot water injection, addition of caustic soda, followed by diluting the resulting heavy oil with naphtha prior to separating the light oil fractions with a centrifuge. In contrast, the *in situ* techniques involve drilling a pattern of wells, injecting steam at the periphery of the pattern and driving the volatile hydrocarbons towards a central drainage well. The latter process produces an emulsion composed of oil and water termed *Orimulsion*, which can be used as a low-grade fuel. Future use of horizontal drilling technology will inevitably lead to more efficient extraction operations.

Reserves

The potential oil reserves in Athabasca, Canada and the Orinoco region of Venezuela are enormous, with estimates being as high as 1.7 and 1.5 trillion barrels of oil, respectively. Additional reserves are also present in Russia. Unfortunately, the economics associated with extracting oil and the environmental damage that results from mining operations are such that in the early years of the 21st century, tar sands are not yet economic reserves of hydrocarbons.

FURTHER READING

Macquaker, J H S, Gawthorpe, R L, Taylor, K G, and Oates, M J (1998) Heterogeneity, Stacking Patterns and Sequence Stratigraphic Interpretation in Distal Mudstone Successions: Examples from the Kimmeridge Clay Formation, UK, in *Recent Progress in Shale Research*, eds J Schieber, W Zimmerle, and P Sethi, Schweizerbart sche Verlagsbuchhandlung, Stuttgart, 163 186.

Tissot, B P and Welte, D H (1984) *Petroleum Formation and Occurrence* 2nd edition, Springer Verlag, Berlin, 1 699.

Organic Farming and the Environment

Carol Morris and Michael Winter
Cheltenham and Gloucester College of Higher Education, Cheltenham, UK

As a major land user across the globe, the impact of agriculture on the environment is of unquestionable signi cance. Agriculture necessarily involves a modi cation of natural ecosystems in order to enhance their potential for food production. However, the profound extent to which the natural environment has been adversely affected by agricultural activity has recently become a matter of scienti c, public and political concern. Examples of these extensively documented, but spatially variable, effects include biodiversity loss through the damage, removal and neglect of habitat and landscape features, the pollution of surface and ground waters by livestock ef uents and agrochemicals, and soil erosion. The environmental problems within agriculture are widely attributed to a conventional or productivist farming system. This is based upon a high input- high output logic, particularly with regard to the use of chemical inputs, and has provided the model for agricultural development throughout most of the 20th century. In response to these problems, there is a recognized need to develop more environmentally sustainable systems of agriculture, of which organic farming represents one variant. Others include integrated farming systems (Morris and Winter, 1999), and environmentally friendly farm management practices which emphasize the production of environmental goods such as hedgerows, woodlands and ponds. Within many of the developed market economies, organic farming is currently generating considerable interest and support. A range of environmental bene ts is believed to accrue from organic farming, but only some are presented here: the impact of organic farming on landscape quality; biodiversity; and soils and natural resources. The disbene ts of the approach are also given brief consideration. Stolze et al. *(2000) provide a more comprehensive review of the environmental impacts of organic farming in Europe.*

BACKGROUND

Driven by mounting evidence of the environmental impacts of conventional production systems, and consumer concerns about the safety of food produced using chemical inputs, organic farming was given official recognition in Europe and North America in the 1980s. However, the origins of present day organic farming lie in a number of alternative agriculture movements during the early 20th century, inspired by a variety of philosophical currents. This dimension remains important within the contemporary organic farming movement. For many organic practitioners, organic farming is an agricultural philosophy as well as a farm management system. The ideological aspect extends beyond production methods. The organic movement also challenges conventional food processing and marketing arrangements and the distance these have created between producers and consumers. In their place, the organic movement advocates a decentralized system of marketing and a more intimate relationship between consumers and producers. However, the recent growth in interest in organic food, exemplified through its presence in the major supermarket chains, has led some commentators to argue that organic farming has become commoditized and divorced from its philosophical foundations (Tovey, 1997).

Organic farming has been defined as an approach to agriculture where the aim is to create integrated, humane, environmentally and economically sustainable agricultural production systems, which maximize reliance on farm-derived renewable resources and the management of ecological and biological processes and interactions, so as to provide acceptable levels of crop, livestock and human nutrition, protection from pests and diseases, and an appropriate return to the human and other resources employed (Lampkin, 1994: 4 5). Rather than thinking of organic as the type of inputs used, Lampkin argues that it should be thought of as referring to the concept of the farm as an organism, in which all component parts the soil minerals, organic matter, microorganisms, insects, plants, animals and humans interact to create a coherent whole (Lampkin, 1994: 5). Lampkin principal features of organic farming are:

- protection of the long-term fertility of soils by, e.g., maintenance of organic matter levels;
- provision of crop nutrients indirectly by the use of relatively insoluble nutrient sources made available by soil microorganisms;
- nitrogen (N) self-sufficiency, e.g., through the use of legumes and recycling of livestock wastes;
- weed, disease and pest control through a variety of physical and cultural methods, e.g., crop rotations and use of resistant varieties, and limited (preferably minimal) chemical intervention;
- extensive, welfare-friendly management of livestock;
- consideration to the wider impact of farming on the environment (e.g., avoidance of pollution) and the positive management of wildlife and natural habitats.

Organic farming is readily distinguished from other types of farming systems by the application of legislated and voluntary standards and certification procedures.

The area of land farmed organically has witnessed a significant expansion in recent years. In Europe, e.g., the area under organic production increased from 100 000 to 600 000 hectare (ha) between 1985 and 1993 (Lampkin, 1994), and to over one million ha in 1997 (UK Agriculture Departments, 1998). However, in the European Union as a whole, less than 5% of the agricultural area is in organic production, which is just over 1% of holdings. Moreover, the rate and extent of conversion varies greatly between nation states. The highest concentrations are in Sweden (12% of all farms), Austria (9%) and Finland (4%), but conversion is also now taking place rapidly in southern countries, particularly in Italy and Spain. A recent European-wide study revealed that organic farms have the following characteristics (Offermann and Nieberg, 2000):

- larger on average than comparable conventional farms (although average organic farm size varies considerably in relation to national averages) and like their conventional counterparts are increasing in size over time;
- natural conditions are worse and soil quality is lower than the national average and relatively more organic farms are located in less favored areas;
- labor use per ha of utilizable agricultural area (UAA) is on average 10 20% higher than conventional equivalents, although labor use is very much dependent on farm type. As in conventional systems, labor use in organic farming is falling.
- a greater diversity in land use, largely because of the importance of integration between enterprises in these systems;
- stocking rates are consistently lower than on conventional farms, although significant variation occurs with farm type.

ORGANIC FARMING AND LANDSCAPE QUALITY

A European-wide research project recently set out to evaluate the relative contributions of organic and conventional farms to landscape quality, encompassing the concept of cultural environment (Kuiper, 2000). Findings varied between the case study countries, with the differences between compared organic and conventional holdings sometimes negligible, particularly where the organic holdings were small and isolated. However, the aggregated observations suggest that organic farms are more likely to make a superior contribution to landscape quality arising from, e.g., their more positive approach to the management of landscape features, greater biodiversity and less soil erosion.

ORGANIC FARMING AND BIODIVERSITY

Organic farm management practices seek to encourage biological diversity, e.g., through the use of crop rotations, maintenance of ground cover, protection and management of natural habitats and avoidance of practices such as use of artificial pesticides likely to disturb species and habitats. Organic standards typically specify a minimum area on a farm (e.g., 5%) that should be devoted to natural habitats. This reflects the recognition of the relationship between habitat features and biological control of pests as well as a commitment to the positive management of these features for nature conservation objectives. Evidence from a variety of European research projects has demonstrated higher levels of abundance of butterflies and spiders in organic than conventional farming systems (Feber *et al.*, 1997; Cobb *et al.*, 1999). The cropping patterns within organic farming (notably the use of grass leys, complex rotations and the structural complexity of the organic cropped habitat), rather than management practices, are believed to account for this pattern. Greater floral diversity, in field margins, arable fields and grassland has also been observed on organic farms within Europe (van Elsen, 2000). In this case, organic farm management practices are important, particularly the absence of agrochemicals which prevents defoliation, allows more flowers to set seed and allows a greater variety of nectar sources for insects (Cobb *et al.*, 1999).

ORGANIC FARMING, SOILS AND NATURAL RESOURCES

Soil is a key agricultural resource and soil structure one of the most important soil properties influencing crop production. A well-developed soil structure enables effective drainage and nutrient supply to the crop and assists the soil s ability to resist erosion. The impact of conventional and organic systems on soil structure in two farms in the US corn-belt have been compared (Gerhardt, 1997). The management methods associated with the organic system, i.e., organic inputs and crop rotations, contributed to an improved soil structure evidenced by the increase in organic matter content, soil biota presence and activity. Conventional management methods were found to degrade soil structure as a result of a lack of organic inputs, conventional tillage, monoculture and use of chemical inputs. While conceding that good management practices in both organic and conventional systems can reduce soil erosion problems, Conacher and Conacher (1998) suggest that the build-up of soil organic matter, maintenance of a protective surface cover and minimum tillage in organic farming favor a reduction in soil loss and its associated problems.

Loss of nitrogen from a farming system can lead to water pollution and human health problems. Organic farming potentially reduces these losses due to lower nitrogen

inputs because mineral fertilizers are not permitted, and livestock densities tend to be lower than in conventional operations. Stolze *et al.* (2000) show that the soil nitrate content on organic farms is lower than on conventional farms during the critical autumn period. They also point to improvements in organic husbandry designed to deal with the high nitrate run-off that can occur when grass and clover leys are ploughed. Likewise, comparing nitrogen surpluses on a limited number of Danish conventional and organic pig and dairy farms, Dalgaard *et al.* (1998) demonstrate that a reduction in nitrogen surplus is possible by converting from conventional to organic dairy farming. However, nitrogen surpluses are greater in organic than conventional pig farming and are attributed to differences in production methods. Scenarios for widespread conversion to organic dairy farming were predicted to require a significant increase in the total dairy farm area, undermining production levels from other farm types.

Using a slightly different approach based on secondary information from a collection of representative data, Cobb *et al.* (1999) attempted to compare the relative environmental burdens of organic and conventional farming. A range of significant impacts on the resource base were examined and calculated in £/ha: nitrous oxide (HNO_2), methane (CH_4) and ammonia (NH_4) emissions to the atmosphere, nitrogen leaching to ground water, carbon dioxide (CO_2) emissions and energy use on the farm and from the manufacture of nitrogen fertilizer, and total pesticide footprint. The total range for organic farming was calculated at £10 15, compared to £25 40 for conventional. These researchers argue that their calculations provide robust evidence of a higher natural resource cost arising from conventional than organic systems.

THE DISBENEFITS OF ORGANIC FARMING

Organic farming undoubtedly is a more environmentally friendly approach, particularly when compared with conventional, high input specialist arable or livestock operations. However, the environmental advantages are less clear when compared to sensitively managed, conventional mixed farms, extensive livestock systems and arable systems employing an integrated management system (Clark *et al.*, 1998). Furthermore, organic farming may be associated with some environmental disbenefits. For example, in the UK, nature conservationists have expressed concern where species rich, extensively farmed grassland has been ploughed out for incorporation into an organic rotation. In Australia, some natural (e.g., farm yard manure) or alternative (e.g., sewage sludge) soil amendments used to promote fertility in organic systems can present a pollution threat, and particular types of natural/alternative pesticides carry a health risk (Conacher and Conacher, 1998). Similar concerns have been raised elsewhere. A recent illustration was the temporary removal from sale of organic mushrooms by one of the UK s largest supermarket chains, prompted by food poisoning fears arising from contamination of the organic growing medium. Thus, it cannot be unequivocally assumed that the organic option is always environmentally benign and can be guaranteed to produce the safest and healthiest food.

In addition, conventional producers are discouraged from converting to organic farming due to some economic disbenefits, notably lower yields. For example, in Europe, yields of cereals are 60 70% and pulses 80% of those from comparable conventional farms. Data on livestock performance is more limited, but studies demonstrate that the yield difference is smaller in organic animal production than it is for crops. Lower yields can be largely attributed to the prohibition of input use and restrictions on stocking rates in organic systems. Although this may discourage some farmers, attuned to a high-input high output mentality, reduced yields may be offset by reduced costs and higher prices received for organic products. Again, European-wide research on the economics of organic farming suggests that the profits of organic farms are quite similar to, and in many cases higher than, those of comparable conventional farms (Offermann and Nieberg, 2000).

CONCLUSIONS

Further scientific research is required to provide precise quantification of many of the environmental, health and economic benefits and disbenefits of organic farming. Available data on the environmental impacts display a number of problems. For example, research tends to focus upon one element such as soil structure, butterfly populations or landscape quality, within an organic system and compares this with its conventional equivalent, rather than comparing the systems as a whole. Even when whole system analyses are undertaken, there is potential for considerable variation in management practices within both organic and conventional systems, making the collection of comparable data between systems undergoing organic transition very difficult. Additionally, research is frequently conducted at the plot or field scale, although whole farm analyses are becoming more common.

These caveats aside, the potential of organic farming to make a positive contribution to a more environmentally friendly agriculture is now widely accepted (Stolze *et al.*, 2000). However, it will only be realized if sufficient land area is converted to this system of production. Achieving higher levels of organic conversion is contingent upon public, and policy support, research and development programs, a well-developed advisory system, adequate processing, distribution and marketing structures, willing and able farmers, and sufficient economic returns. Although predictions of the future extent of organic farming, both in

terms of the UAA and market share, are constantly being revised upwards, in the final analysis it seems unlikely that this approach will be universally adopted. This reinforces the point that organic farming is just one possible option in the development of more environmentally sustainable agriculture.

REFERENCES

Clark, M, Ferris, H, Klonsky, K, Lanini, W, van Bruggen, A, and Zalom, F (1998) Agronomic, Economic, and Environmental Comparison of Pest Management in Conventional and Alternative Tomato and Corn Systems in Northern California, *Agric., Ecosyst. Environ.*, 68, 51 71.

Cobb, R, Feber, R, Hopkins, A, Stockdale, L, O Riordan, T, Clements, B, Firbank, L, Goulding, K, Jarvis, S, and Macdonald, D (1999) Integrating the Environmental and Economic Consequences of Converting to Organic Agriculture: Evidence from a Case Study, *Land Use Policy*, 16, 207 221.

Conacher, J and Conacher, A (1998) Organic Farming and the Environment, with Particular Reference to Australia: a Review, *Biol. Agric. Hortic.*, 16, 145 171.

Dalgaard, T, Halberg, N, and Kristensen, I (1998) Can Organic Farming Help to Reduce N-losses?, *Nutr. Cycling Agroecosyst.*, 52, 277 287.

Feber, R, Firbank, L, Johnson, P, and Macdonald, D (1997) The Effects of Organic Farming on Pest and Non-Pest Butterfly Abundance, *Agric., Ecosyst. Environ.*, 64, 133 139.

Gerhardt, R (1997) A Comparative Analysis of the Effects of Organic and Conventional Farming Systems on Soil Structure, *Biol. Agric. Hortic.*, 14, 139 157.

Kuiper, J (2000) A Checklist Approach to Evaluate the Contribution of Organic Farms to Landscape Quality, *Agric., Ecosyst. Environ.*, 77, 143 156.

Lampkin, N (1994) Organic Farming: Sustainable Agriculture in Practice, in *The Economics of Organic Farming*, eds N Lampkin and S Padel, CAB International, Wallingford, 3 9.

Morris, C and Winter, M (1999) Integrated Farming Systems: the Third Way for European Agriculture? *Land Use Policy*, 16, 193 205.

Offerman, F and Nieberg, H (2000) *Economic Performance of Organic Farms in Europe*, in *Organic Farming in Europe: Economics and Policy*, Vol. 5 in series, Hohenheim, Stuttgart.

Stolze, M, Piorr, A, Haring, A, and Dabbert, S (2000) *Environmental Impacts of Organic Farming in Europe*, in *Organic Farming in Europe: Economics and Policy*, Vol. 6 in series, Hohenheim, Stuttgart.

Tovey, H (1997) Food, Environmentalism and Rural Sociology: on the Organic Farming Movement in Ireland, *Soc. Ruralis*, 37, 21 38.

United Kingdom Agriculture Departments (1998) *Review of the Organic Aid Scheme*, MAFF, London.

van Elsen, T (2000) Species Diversity as a Task for Organic Agriculture in Europe, *Agric., Ecosyst. Environ.*, 77, 101 109.

FURTHER READING

Rewanold, J P, Glover, J D, Andrews, P K, and Hinman, H R (2001) Sustainability of three apple production systems, *Nature*, 410, 926 930.

Trewayas, A (2001) Urban Myths of Organic Farming, *Nature*, 410, 409 410.

536

Phosphorus: Global Transfers

Vaclav Smil

Winnipeg, Canada

Although relatively rare in the biosphere, phosphorus (P) plays several key roles in the chemistry of life, above all due to its presence in nucleic acids (deoxyribonucleic acid, DNA and ribonucleic acid, RNA) and in adenosine triphosphate, the life's carrier of energy. Phosphorus – unlike carbon, C, nitrogen, N and sulfur, S – does not form any long-lived atmospheric compounds and hence its global cycle is just a part of the grand, and slow, process of denudation and geotectonic uplift. But on a small scale the element is rapidly recycled between organic and inorganic forms in soils and water bodies. Human activities now annually move more than four times as much phosphorus as did the natural processes during the preagricultural era.

Together with nitrogen and potassium, phosphorus is one of the three macronutrients needed by all plants, and its applications in inorganic fertilizers have been essential for raising crop yields during the 20th century. The nutrient is also frequently the limiting factor of aquatic productivity, and even its relatively low concentrations present in the runoff from fertilized fields can cause eutrophication (excessive nutritional enrichment) of both fresh and ocean waters, an environmental change with numerous undesirable consequences. Fortunately, appropriate agronomic practices and wastewater treatment can moderate the environmental impact of anthropogenic phosphorus.

IMPORTANCE OF THE ELEMENT

Phosphorus is one of the three macronutrients required by plants. It is particularly important for the growth of young tissues, flowering and seed formation (cereal grains contain 0.25–0.45% phosphorus (P), more than is contained in any other plant tissues). The element is also an essential micronutrient in humans: its daily intakes are usually around 1.5 mg capita^{-1}, well above the recommended daily minima of about 1.2 mg (National Research Council, 1989).

P is absent in cellulose, hemicellulose and lignin, the three polymers that make up the bulk of the plant mass, as well as in proteins that compose the metabolic tissues of heterotrophs. However, the element is abundant in bones and teeth: hydroxyapatite, a calcium phosphate, makes up 60–70% of their mass. The element is indispensable in phosphodiester bonds linking mononucleotide units of RNA and DNA, and all life processes are energized by transformations of adenosine triphosphate.

NATURAL STORES AND FLOWS OF PHOSPHORUS

Comprehensive assessments of the global P cycle have been much rarer than the studies of carbon or nitrogen flows (Smil, 1990; Jahnke, 1992; Mackenzie *et al.*, 1993; Tiessen, 1995; Smil, 2000; Figure 1; Table 1). The Earth's crust is the largest reservoir of the element, nearly all of it bound in apatites, calcium phosphates containing also iron, chlorine or the OH group. Soluble phosphates are released by weathering, but they are rapidly transformed to insoluble

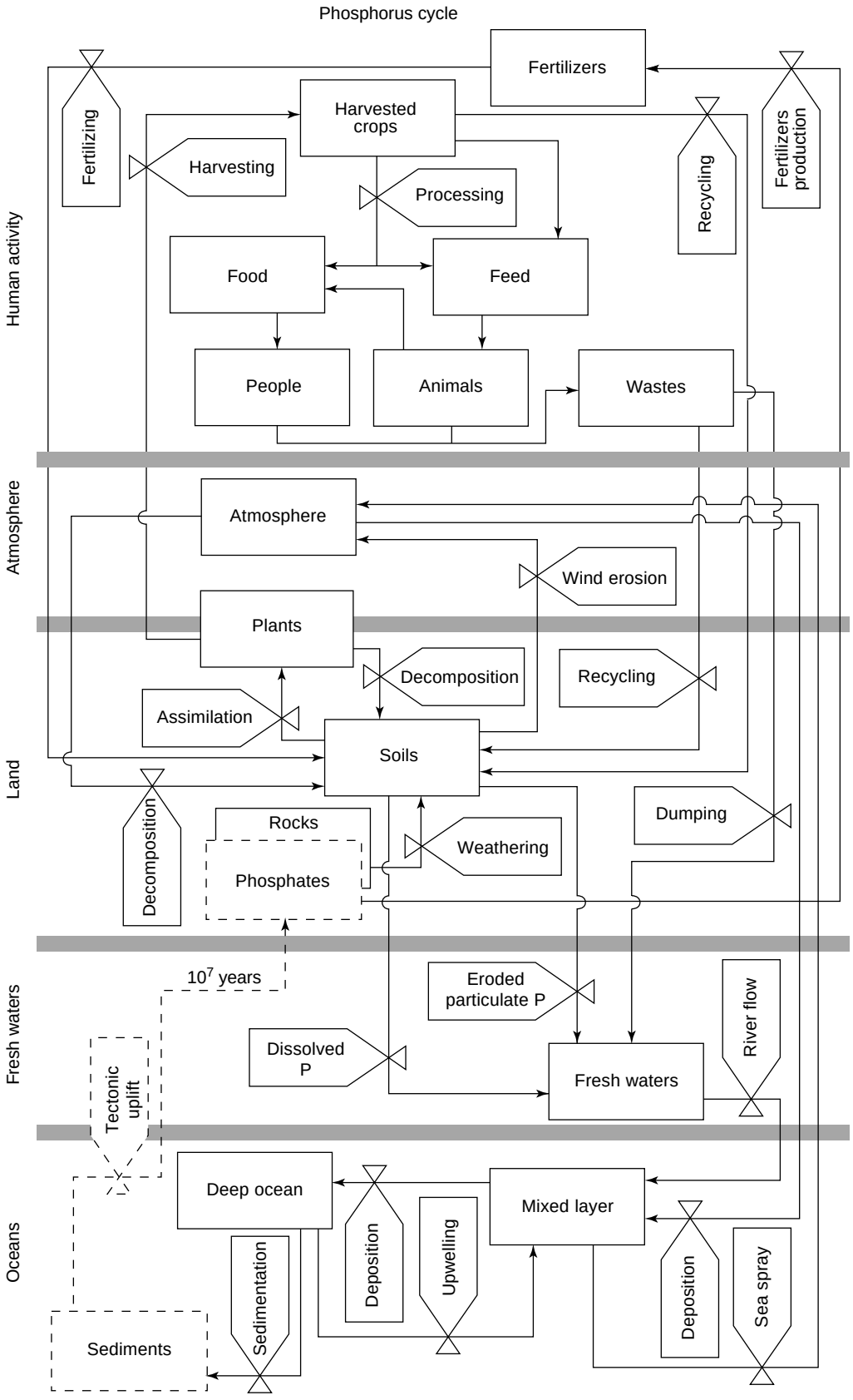


Figure 1 Global phosphorus cycle

Table 1 Major biospheric reservoirs and fluxes of phosphorus

Phosphorus reservoirs	Total storage (Mt P)
Ocean	93 000
Surface	8000
Deep	85 000
Soils	40–50
Inorganic phosphorus	35–40
Organic phosphorus	5–10
Phytomass	570–625
Terrestrial	500–550
Marine	70–75
Zoomass	30–50
Anthropomass	3
Phosphorus fluxes	**Annual rate (Mt P year^{-1})**
Atmospheric deposition	3–4
Erosion and runoff	25–30
Particulate phosphorus	18–22
Dissolved phosphorus	2–3
Plant uptake	970–1300
Terrestrial	70–100
Marine	900–1200
Burial in marine sediments	20–35
Tectonic uplift	15–25

compounds in soils. Consequently, plants must be able to absorb P from very dilute solutions and concentrate it 100 1000 fold in order to meet their needs, and P released by decomposition of biomass is rapidly reused.

The element is even less available in fresh waters and in the oceans where its particulate forms sink into sediments and where soluble phosphates are rapidly recycled in order to support phytoplanktonic photosynthesis in surface waters. Only in the areas of upwelling are surface layers enriched by the influx from deeper waters. Phosphorus in marine sediments can become available to terrestrial biota only after the tectonic uplift re-exposes the minerals to denudation. The element s global cycle thus closes only after tens to hundreds of millions of years. However, the temporary interruptions of the element s oceanward journey are indispensable for the Earth s life.

The terrestrial phytomass stores about 500 Mt of the element, plant growth assimilating up to 100 Mt P year^{-1}. Phosphates dissolved in rain and dry-deposited in particles amount to only about 3 Mt a year. Soils store about 40 Gt P, with no more than 15% of this total bound in organic matter. Natural denudation and precipitation transfer annually about 10 Mt of particulate and dissolved phosphorus from land to the ocean. The marine phytomass stores only some 75 Mt phosphorus but because of its rapid turnover, it absorbs annually about 1 Gt P from surface water. Mixing between sediments and the surface layer is effective only in very shallow waters, and high near-surface concentrations of the nutrient only occur in coastal areas receiving phosphorus-rich runoff. Coastal seas contain only some 0.2% phosphorus in the ocean but they support a disproportionately large share of marine productivity, and the top 300 m of the open ocean contain less than one tenth of all marine phosphorus.

ANTHROPOGENIC MOBILIZATION OF PHOSPHORUS

Human activity intensifies biospheric phosphorus flows in five main ways:

1. accelerated erosion and runoff caused by large-scale conversion of natural ecosystems (mainly forests and grasslands) to arable land;
2. settlements and transportation links;
3. recycling of organic wastes crop residues, animal manures and human urine and faeces to fields;
4. releases of untreated, or insufficiently treated, urban and industrial wastes to streams and water bodies;
5. applications of inorganic fertilizers; and combustion of biomass and fossil fuels (Smil, 2000; Table 2).

Although the last process is the least important source of phosphorus, it is the largest anthropogenic source of carbon and a large source of fixed nitrogen (*see* Phosphorus Water Quality Objectives, Volume 3) Trace amounts of phosphorus present in wood and coal remain overwhelmingly in the ash and less than 0.3 Mt P are released annually in airborne particulate matter.

Recycling of organic wastes has always been a mainstay of traditional farming and it remains an important

Table 2 Human intensification of the global phosphorus cycle (all values are in Mt P year^{-1})

Fluxes	Natural	Preindustrial (1800)	Recent (2000)
Natural fluxes intensified by human actions			
Erosion	• 10	• 15	• 30
Wind	• 2	• 3	• 3
Water	• 8	• 12	• 27
River transport	• 7	• 9	• 22
Particulate P	• 6	• 8	• 20
Dissolved P	• 1	• 2	• 2
Biomass combustion	• 0.1	• 0.2	• 0.3
Anthropogenic fluxes			
Crop uptake	–	1	12
Animal wastes	–	• 1	• 15
Human wastes	–	0.5	3
Organic recycling	–	• 0.5	• 6
Inorganic Fertilizers	–	–	15

ingredient of well-balanced modern agriculture. Progressing urbanization and diffusion of sewers have gradually reduced the often intensive recycling of human wastes, once characteristic of many regions of Asia and Europe. Nevertheless, the recycling of crop residues remains highly desirable, not only because of the return of plant nutrients but also because of its contributions to better tilth, higher water absorption capacity and reduced soil erosion. During the 1990s, global crop harvests produced about 3.75 Gt (dry matter) of residues (mostly cereal straws). About half this phytomass was recycled, returning (with average content of 0.05 0.1% phosphorus) 1 2 Mt P to arable soils.

Animal wastes are potentially a very large source of recyclable phosphorus. Every year, over four billion large mammals and some 10 billion domestic birds produce together about 2 Gt of dry waste containing up to 20 Mt P. Only the wastes produced in confinement are recyclable but, because of their bulk, uneven distribution and excessive cost of application beyond a limited radius from the source, not all of this manure is actually applied to fields. The best estimates indicate annual return of up to 8 Mt P, with the highest applications (exceeding 50 kg P ha^{-1}) in such regions of high animal density as the US Corn Belt, Northwestern Europe and East Asia.

The global population of six billion people now discharges every year about 3 Mt P in its wastes. Nationwide production densities of human waste are as high as 9 kg P ha^{-1} of cultivated and settled land in densely populated countries (Egypt, Japan), and more than 100 kg P ha^{-1} in large cities. A large share of human waste produced in rural areas of the low-income countries is deposited on land, but progressing urbanization puts a growing share of human waste into sewers and then, treated or untreated, into streams or water bodies. Since the 1940s phosphorus-containing detergents have added another major source of waterborne phosphorus. Neither the primary sedimentation of urban sewage (which removes only 5 10% of the initial phosphorus in the sludge), nor the use of trickling filters during the secondary treatment (removing 10 20% phosphorus) withhold enough of the nutrient to prevent eutrophication, undesirable enrichment of waters with the nutrient that most often limits the growth of phytoplankton.

Conversion of forests and grasslands removes the canopies and litter layer that protect the soil and opens the way for accelerated soil erosion which increases typically by one or two orders of magnitude, from 0.1 2 t ha^{-1} to 10 20 t ha^{-1} a year; overgrazing of permanent pastures has a similar effect. The additional phosphorus lost due to these changes is difficult to quantify with any degree of confidence. I have estimated that over 20 Mt more phosphorus are lost annually from farmland and pastures than erosion would have removed from the same area of land if it were covered by undisturbed natural ecosystems. Enhanced anthropogenic loss of dissolved phosphorus adds another 1 Mt P $year^{-1}$.

Production of inorganic fertilizers began during the 1840s with the treatment of P-containing rocks with dilute sulphuric acid. The resulting ordinary superphosphate (OSP) contains 7 10% phosphorus, ten times as much as recycled phosphorus-rich manures. Discovery of huge phosphate deposits in Florida (1870s), Morocco (1910s) and Russia (1930s) laid the foundation for the rapid post-World War II expansion of the fertilizer industry. Consumption of phosphatic fertilizers reached 5 Mt P in 1960 and it peaked at more than 16 Mt P in 1988. After a 25% decline by 1993 the global use began rising once again (International Fertilizer Industry Association, 1999; Figure 2). The top three producers (USA, China and Morocco) now account for about 2/3 of the global output. Major fertilizers besides OSP include triple superphosphate (TSP, containing about 20% P) produced by treating phosphate rock with phosphoric acid, and combinations with nitrogen (mono- and diammonium phosphate), calcium and potassium (UNIDO and IFDC, 1998).

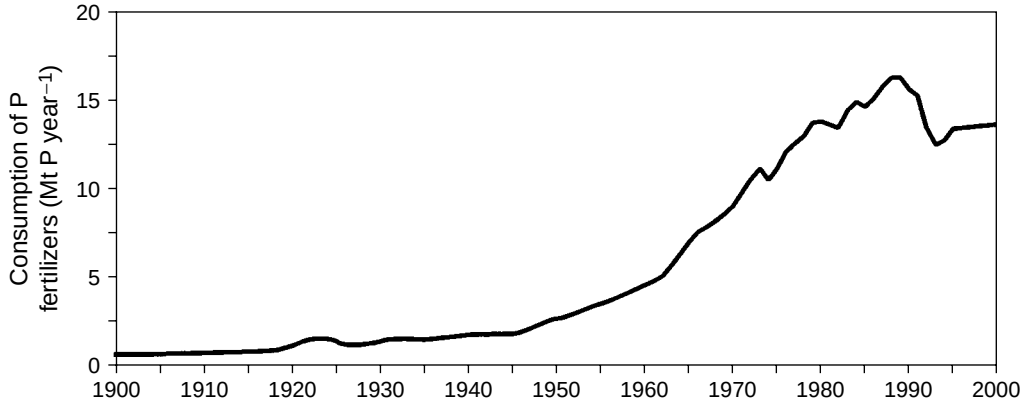


Figure 2 Consumption of inorganic phosphatic fertilizers, 1900–2000

DEPENDENCE ON INORGANIC FERTILIZERS

Modern cultivars whose harvests now dominate the production of food and feed remove annually between 15 35 kg P ha^{-1} of cereals, 15 25 kg P ha^{-1} of leguminous and root crops and 5 15 kg P ha^{-1} of vegetables and fruit (Pierzynski and Loan, 1993). Global food harvests now assimilate about 12 Mt P in crops and in their residues while no more than 4 Mt P are supplied annually by weathering of P-bearing rocks and by atmospheric deposition (Smil, 2000). Applications of phosphatic fertilizers have thus become indispensable for producing today s harvests and this dependence will only increase.

Combination of an early stabilization of human population at less than 10 billion people, less meaty diets in affluent countries (where about 60% of all cereals are now used as animal feed), and relatively modest increases in animal food consumption in low income countries would significantly lower the growth rate of global food demand during the first half of the 21st century but they would still result in a substantial rise of overall phosphorus requirements.

During the 1990s global applications of inorganic fertilizers averaged just over 10 kg P ha^{-1} of arable land, with continental means ranging from a mere 3 kg P ha^{-1} in Africa to over 25 kg P ha^{-1} in Europe (International Fertilizer Industry Association, 1999). These means hide enormous inter- and intra-national differences, as well as uneven use on different crops. For example, in Africa applications are adequate on some plantation crops while tens of millions of the continent s subsistence peasants have never applied any phosphorus at all. International surveys indicate that nearly 2/3 of all applied phosphorus is used on cereals, 10% on oilseeds and about 5% on vegetables (Harris, 1998). Crops receiving the highest applications are corn in the US Corn Belt (around 60 kg P ha^{-1}), Japanese rice (over 40 kg P ha^{-1}) and Chinese and West European winter wheat (about 35 kg P ha^{-1}).

PHOSPHORUS IN AGRICULTURAL SOILS

Phosphorus applied in both organic and inorganic fertilizers is involved in complex reactions that transform a large part of soluble phosphates into much less soluble compounds (Figure 3). Particularly important in this regard are reactions between superphosphates and Al, Fe, K and Mg. The resulting compounds are relatively insoluble and hence immobile in soil. This process of P fixation has been known since 1850 and it has received an enormous amount of research attention (Khasawneh *et al.*, 1980; Morgan, 1997). The traditional argument is that the fixation is rapid, dominant and irreversible. Accordingly, phosphate applications were seen as overwhelmingly futile albeit laborious and costly exercises of moving the element from extracted rocks to insoluble soil compounds (Tinker, 1977).

This misconception resulted in decades of excessive use of phosphorus in agriculture: global N : P ratio of applied fertilizers was just around 1.5 by the late 1940s. Mounting evidence of relatively high efficiency with which crops use the nutrient has finally changed that wasteful practice (Karlovsky, 1981; Isermann, 1990). By the late 1990s the global fertilizer N : P ratio stood around 6 : 1. Recovery rates of 50 60% of applied phosphorus are not uncommon, and maxima are above 70%. Even the global mean appears to be close to 50%, virtually the same as the rate for inorganic nitrogen. Decades of relatively high P applications and the gradual release of a part of the initially fixed nutrient have resulted in many fertilized agro ecosystems have considerable surpluses of available phosphorus (Edwards *et al.*, 1997). Present annual accumulations commonly amount to several kg P ha^{-1}, surpassing 30 kg P ha^{-1} in the most intensively cultivated regions.

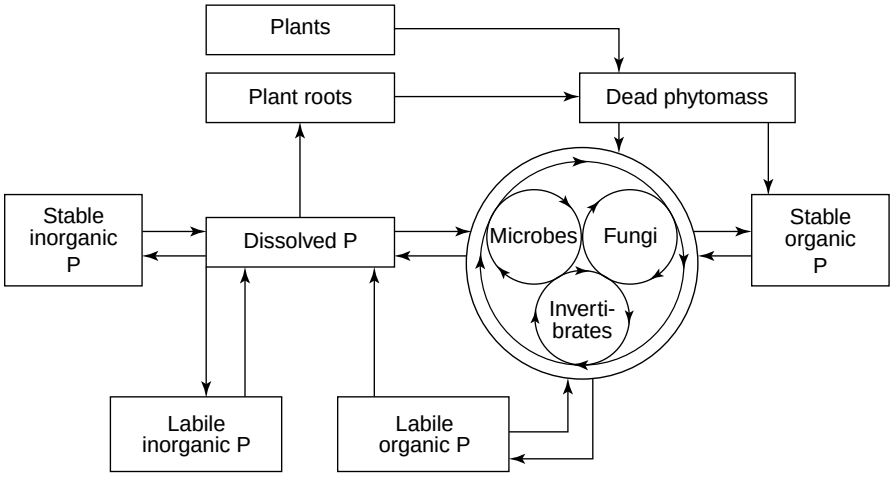


Figure 3 Phosphorus cycle in soil

PHOSPHORUS IN WATERS

The low solubility and high sorption rates of phosphorus compounds limit their leaching into ground waters, leaving diffuse runoff from fertilizers and manures and direct, point discharges of urban and industrial wastes as the two principal ways of enriching surface waters with phosphorus. Because a single atom of phosphorus supports the production of as much phytomass as 16 atoms of nitrogen and 106 atoms of carbon, even relatively low additions of the nutrient can cause eutrophication (*see* Eutrophication, Volume 2), excessive nutritional enrichment of waters (Harper, 1992; McComb, 1995; Tunney *et al.*, 1997). Mere 10 μg P l^{-1} will support algal growth that will greatly reduce the clarity of water. As total phosphorus in water increases, the standing phytomass goes up linearly. It may include unsightly scum-forming algal blooms and filamentous cyanobacteria, as well as thick coats of algae on any submerged surfaces.

Concentrations above 50 μg P l^{-1} will result in deoxygenation of bottom waters during the decomposition of the accumulated biomass; this can cause extensive summer fish kills and reduce or eliminate many aquatic species: pike may disappear while bream and roach, species adapted to turbid waters, become dominant. Offensive taste and odor of water filled with algae requires expensive treatment before drinking, and neuro- and hepatotoxins released by decomposition of some of these blooms may be a serious health hazard to livestock and people. Studies of polluted lakes and estuaries have shown that excessive eutrophication can be generally prevented if annual loadings are lower than 1 g P/m^2 (10 kg P ha^{-1}) of water surface but much higher loadings may not cause serious problems as long as the average retention time of imported nutrients is short.

Eutrophication also seriously disrupts coastal ecosystems in regions receiving high P inputs. For example, it has affected the growth of corals in parts of the Australian Great Barrier Reef (Bell and Elmetri, 1995). In the North Sea it repeatedly creates anoxic conditions and toxic algal blooms while a third of the sea bottom of the Baltic Sea proper (its southern part) is now intermittently deprived of oxygen, the condition which also results in formation of toxic H_2S by S-reducing bacteria and precludes the survival of previously very common mussels and bottom fish (Jansson and Dahlberg, 1999) (*see* Water Resources: Baltic, Volume 3; Baltic Sea, Volume 4).

MODERATING THE HUMAN IMPACT

There is no shortage of ways to lower the anthropogenic loss of phosphorus. By far the most effective measure would be to reduce the intake of animal foods in affluent countries where *per capita* consumption is now commonly above 60 kg $year^{-1}$ of meat and that of dairy products above 100 kg $year^{-1}$. Lowering these intakes by 15 35% would still allow the enjoyment of diets with plenty of animal foods. In addition, since 2/3 of all phosphatic fertilizers are used on cereals and 3/5 of all grains in rich countries are used as animal feed the need for phosphatic fertilizers would decline by 5 15% without any investment.

An approach that is much more likely to be successful relates to best management practices aimed at reducing phosphorus applications or limiting the post-application losses (Sharpley and Withers, 1994). Indirect measures include reduced tillage, crop rotations and recycling of crop residues; direct ways include applications of phosphatic fertilizers derived from repeated soil testing, balanced amounts of the three macronutrients, maintenance of appropriate soil pH, and incorporation, rather than broadcasting, of phosphorus compounds.

Phosphorus in sewage can be effectively controlled either by the use of coagulating agents or by microbial processes (Imperial College of Science, 1997). Addition of flocculating reagents (most commonly salts of iron or aluminum) can remove 70 95% of all phosphorus initially present but the increased bulk of the resulting sludge and high cost of reagents favor microbial treatment of waste water. This treatment relies first on anaerobic conditions for phosphorus release and then during the aerobic stage on a rapid uptake of soluble P by phosphate-accumulating microorganisms. While conventional sludge has 1.5 3% phosphorus, activated sludge contains up to 7% phosphorus and the effluent has less than 1 mg P l^{-1}. Phosphorus-rich sludge can be either recycled or it can be dried and incinerated and the nutrient recovered from ash.

The best way to deal with waste phosphorus is to precipitate it in discrete crystalline products of reasonable purity that can be used either directly as fertilizers or as raw materials. Nearly a dozen of these processes are now commercially available or in an advanced research stage; they produce magnesium ammonium phosphate or calcium phosphate (which is ideal for further processing as it is indistinguishable from phosphate rock), and they release effluent with less than 0.5 mg P l^{-1}. Naturally, the eventual mass deployment of these efficient controls will depend on their cost and on the standards set by national environmental regulations.

In many cases, perhaps most notably in the Great Lakes, eutrophication has been greatly reduced due to the elimination or restricted use of phosphorus-based detergents and increased phosphorus removal from sewage. In other cases it remains a serious problem: although total Dutch phosphorus inputs were cut by about 50% by 1995 compared to the year 1985, phosphorus-saturated soils remain a large source of excessive runoff and many surface waters have phosphorus concentrations an order of magnitude above the natural levels (Uunk, 1991). And while phosphorus loadings may have stabilized, or even substantially declined in

many lakes in North America, Europe and Japan, they are rising in China and India, and eventually they will increase also in Africa.

Yet another long-term environmental problem associated with applications of phosphatic fertilizers and recycling of manures and sewage sludges is the presence of cadmium (Van Kauwenbergh, 1997). Cadmium is the most enriched element in phosphate rocks, occurring in concentrations almost 70 times higher than in average shale. Maximum daily intake should not exceed 50 70 μg Cd day^{-1} for most adults while the recent European intakes are already as high as 40 μg Cd day^{-1}. Consequently, about a dozen countries (including Australia, Denmark, Germany, Japan and the Netherlands) have already set Cd limits for fertilizers.

REFERENCES

Bell, P R F and Elmetri, I (1995) Ecological Indicators of Large-scale Eutrophication in the Great Barrier Reef lagoon, *Ambio*, 24, 208 215.

Edwards, A C, Withers, P J A, and Sims, T J (1997) *Are Current Fertilizer Recommendation Systems for Phosphorus Adequate*? Fertilizer Society, York.

Harper, D (1992) *Eutrophication of Freshwaters,* Chapman and Hall, London.

Harris, G (1998) An Analysis of Global Fertilizer Application Rates for Major Crops, International Fertilizer Industry Association, Paris.

Imperial College of Science (1997) *Phosphorus Removal and Reuse Technologies*, Selper Publications, London.

International Fertilizer Industry Association (1999) *World Phosphorus Fertilizer Consumption*, International Fertilizer Industry Association, (http://www.fertilizer.org).

Isermann, K (1990) Share of Agriculture in Nitrogen and Phosphorus Emissions into the Surface Waters of Western Europe against the Background of their Eutrophication, *Fertilizer Res.*, 26, 253 269.

Jahnke, R A (1992) The Phosphorus Cycle, in *Global Biogeochemical Cycles*, eds S J Butcher, R J Charlson, G H Orians, and G V Wolfe, Academic Press, London, 1 379.

Jansson, B-O and Dahlberg, K (1999) The Environmental Status of the Baltic Sea in the 1940s, Today, and in the Future, *Ambio*, 28, 312 319.

Karlovsky, J (1981) Cycling of Nutrients and their Utilization by Plants in Agricultural Ecosystems, *Agro-Ecosystems*, 2, 127 144.

Khasawneh, F E, Sample, E C, and Kamprath, E J (1980) *The Role of Phosphorus in Agriculture*, American Society of Agronomy, Madison, WI.

Mackenzie, F T, Ver, L M, Sabine, C, Lane, M, and Lerman, A (1993) C, N, P, S Biogeochemical Cycles and Modeling of Global Change, in, *Interactions of C, N, P and S Biogeochemical Cycles and Global Change*, eds R Wollast, F T Mackenzie, and L Chou, NATO ASI Series, Springer Verlag.

McComb, A J, ed (1995) *Eutrophic Shallow Estuaries and Lagoons*, CRC Press, Boca Raton, FL.

Morgan, M A (1997) The Behaviour of Soil and Fertilizer Phosphorus, in *Phosphorus Loss from Soil to Water*, eds H Tunney, O T Carton, P C Brookes, and A E H Johnston, CAB International, Wallingford.

National Research Council (1989) *Recommended Daily Allowances*, National Research Council, Subcommittee on the Tenth Edition of the RDAs, National Academy Press, Washington, DC.

Pierzynski, G M and Loan, T J (1993) Crop, Soil, and Management Effects of Phosphorus Test Levels, *J. Prod. Agric.*, 6, 513 520.

Sharpley, A N and Withers, P J A (1994) The Environmentally-sound Management of Agricultural Phosphorus, *Fertilizer Res.*, 39, 133 146.

Smil, V (1990) Nitrogen and Phosphorus, in *The Earth as Transformed by Human Action*, eds B L Turner, R W Kates, J Richards, J T Mathews, and W Meyer, Cambridge University Press, New York.

Smil, V (2000) Phosphorus in the Environment, *Annu. Rev. Energy Environ.*, 25, 53 88.

Tiessen, H, ed (1995) *Phosphorus in the Global Environment: Transfers, Cycles and Management*, John Wiley & Sons, Chichester.

Tinker, P B (1977) Economy and Chemistry of Phosphorus, *Nature*, 270, 103 104.

Tunney, Carton, O T, Brookes, P C, and Johnston, A E H, eds (1997) *Phosphorus Loss from Soil to Water*, CAB International, Wallingford.

UNIDO and IFDC (1998) *Fertilizer Manual*, Kluwer, Dordrecht, 1 615.

Uunk, E J B (1991) *Eutrophication of Surface Waters and the Contribution of Agriculture*, The Fertilizer Society, Peterborough.

Van Kauwenbergh, S J (1997) *Cadmium and Other Minor Elements in World Resources of Phosphate Rock*, The Fertilizer Society, York.

Phosphorus Water Quality Objectives

Ilija Vukadin

Institute for Oceanography and Fisheries, Split, Croatia

During the last three decades, there has been a several-fold growth in use of nitrogen (N) and phosphorus (P), particularly in agricultural and chemical production, such as in detergents and in food additives. Many factors are important in controlling levels of plant biomass in aquatic ecosystems, but phosphorus is generally considered the key element that controls or limits productivity. The enrichment by nutrients and suspended matter causes accelerated growth of algae

and higher forms of plant life and produces an undesirable disturbance to the balance of organisms present in the waters. Excess phosphorus inputs into surface waters are a major determining factor in eutrophication phenomena associated with the impairment of commercial and recreational fishing, local water quality problems and aesthetic damage.

Viewed in terms of water quantity, it is agriculture that uses the most water, followed by industry and cities, leaving barely enough for the natural environment to exist. From the standpoint of water quality, the greatest polluter of water is industry, followed by municipalities and agriculture.

The aim of this paper is to briefly discuss the main feature of the present flux of nutrients through agro-urban and aquatic ecosystems, and the measures available to reduce or stop the great losses of nitrogen and phosphorus. Special attention will be paid to phosphorus, because this element is looked upon as a limited natural resource.

During the last three decades there has been a several-fold growth in use of nitrogen and phosphorus, for example, agriculture production, in chemical production such as in detergents and as food additives. Cola drinks, for example, contain 180 mg P dm^{-3}, a concentration about 350 times higher than that of the effluent (0.5 mg P dm^{-3}) from modern sewage treatment plants operating with dephosphatation. Feeding a growing population means increased human nutrient excretion (Forsberg, 1994).

The mobility of nitrogen differs from that of phosphorus. Compared to the easily mobile nitrate ion, phosphorus is fixed in different systems, at least temporarily. The modern agro-ecosystem accumulates about 60% of the annual input of phosphorus.

Since the amount of phosphates from municipal sewage contributes about 40% of total P-loading in European countries, of which about 40-50% is due to detergent phosphates, the reduction of detergent phosphates could have a beneficial effect in reducing the total load.

PHOSPHORUS WATER QUALITY OBJECTIVES

Many factors are important in controlling the levels of plant biomass in aquatic ecosystems, but phosphorus is generally considered to be the key element that controls or limits productivity (*see* Soil Amelioration, Volume 3). Nitrogen plays a secondary role, although it can sometimes become important (*see* Nitrate Leaching, Volume 3). Another reason why emphasis is given to phosphorus in eutrophication studies is that phosphorus is often derived primarily from sources related to human activities. The main sources of phosphorus loading are well recognized and could be sufficiently quantified, although there are some uncertainties particularly in evaluating loading from non-point sources of agricultural origin. The contribution of agricultural diffuse sources results in a mean value of 17% of the total P load, although values as high as 30% are attained, and even higher (see Figure 1).

Our estimates of point and non-point source nutrient loads to the Bay indicate that they contribute near to equal amounts of phosphorus to the loading of the Bay.

Continuous monitoring of seawater along the eastern Adriatic coast has been carried out for the past decade. The goal of this research is to observe the influence of urban and industrial effluents on the quality of the receiving seawater. The nutrient balance has been calculated (Vukadin, 1992) for the whole Adriatic Sea (see Table 1). River runoff and domestic wastes are very probably the

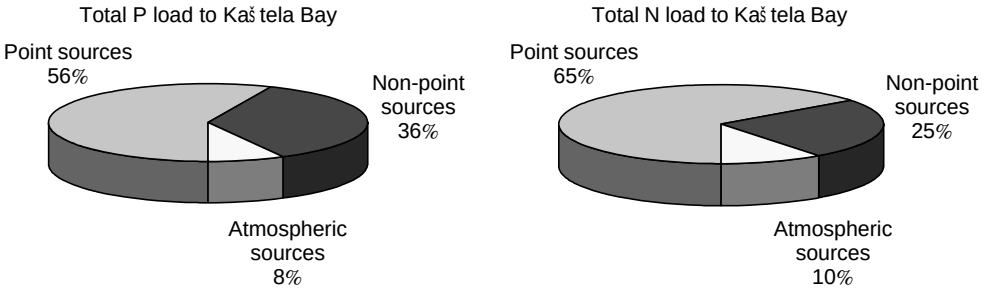


Figure 1 Source of nutrients in Kaštela Bay, Croatia

Table 1 Annual pollution input into the Adriatic Sea (• 10^3 t year^{-1})

Pollutants	Rivers	Domestic effluents	Industrial effluents	Agricultural	Total
Phosphorus (P)	79	2.5	0.5	3	85
AD	14 000	2200	–	–	16 200

AD, anionic detergent.

causes of eutrophication and phytoplankton blooms in the northern and middle Adriatic (Vollenweider *et al.*, 1992).

The values of anionic detergents (AD) and phosphorus found in the Adriatic coastal waters can be classified as sufficiently safe (values AD lower than 1 10 mg dm^{-3}, i.e., the acute values of toxicity of linear alkylbenzene sulfonates for aquatic life). These values are still far beyond these levels and are in compliance with proposal quality standards, (Vukadin *et al.*, 1997).

Two different approaches for obtaining water quality criteria have been followed in different countries:

- without considering the type and use of the water body receiving the contaminant;
- adapting the quality of the pollutant to the natural characteristics of the receptors (rivers, lakes, coastal waters) and taking into account the characteristics of the pollutants (toxicity, persistence, bioaccumulation).

The well-known Organization for Economic Cooperation and Development (OECD) models provide a method of interpreting and using nutrient load estimates in water management effort (Table 2). The model attempts to relate the classical limnologic trophic terminology of oligotrophic, mesotrophic, eutrophic and hyper trophic to specific in-lake levels of several water quality parameters.

Table 2 OECD boundary values for trophic categories

Category	Total P concentration (mg dm^{-3})
Oligotrophic	• 10
Oligo–mesotrophic	10–20
Mesotrophic	20–50
Eutrophic	50–100
Hypertrophic	• 100

Recognition of the probabilistic nature of the responses of lakes to given P-loadings is an important aspect in assessing the trophic response of a lake. Some probability distributions, based on the OECD database, are presented in Figure 2 (Chiaudani *et al.*, 1988). A lake with an average phosphorous concentration of 10 mg m^{-3} would have a 10% probability of being ultraoligotrophic, 26% of being mesotrophic, only 1% of being eutrophic and no probability of being hypertrophic.

As already mentioned, phosphorus inputs into surface waters are a major factor in eutrophication phenomena associated with the impairment of commercial and recreational fishing, local water quality problems and aesthetic damage. Phosphates from a number of different sources enrich surface waters, but restriction or a ban on their use in detergent alone would significantly reduce (30% 50% of the total phosphorous content comes from raw sewage) the nutrient loading, and would significantly reduce the rate of eutrophication.

The effective and long-term control of eutrophication can only be achieved by reducing nutrient loads by adequate preventive measures, supported by appropriate administration and legal powers. Basic options are:

- chemical phosphorous removal in treatment plants and the development and improvement of enhanced biological phosphorous removal;
- ban or reduction of phosphates in detergents;
- pollution control measures relating to agricultural practices; and
- regional planning including land use control.

THE ROLE OF PHOSPHATES IN DETERGENTS

About 50 years ago, it was discovered that a particular phosphate salt, sodium triphosphate (STP)-$Na_5P_3O_{10}$, has outstanding properties as a detergent. This compound

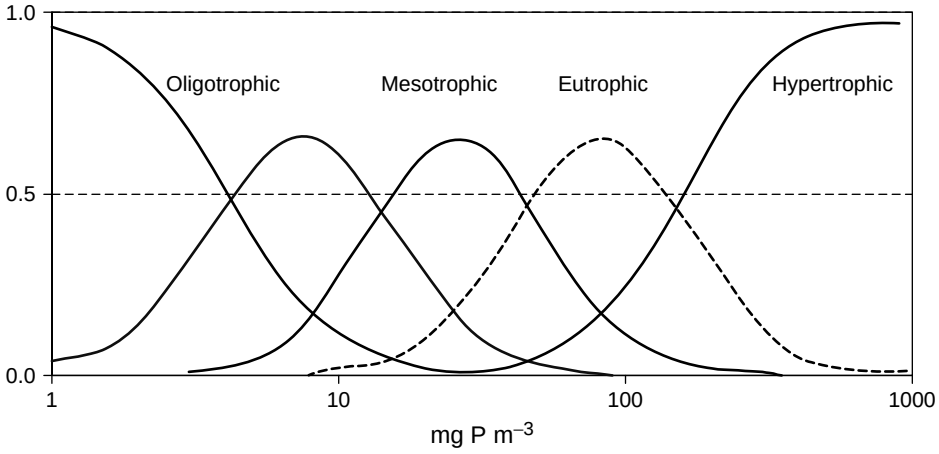


Figure 2 Probability distribution of trophic categories, based on total phosphorus

reduces water hardness (by sequestering magnesium and calcium ions) and detaches certain stains from fabrics.

STP is an excellent compatible with other ingredients used in detergent manufacture; in particular, it is stable to oxidizing agents (i.e., sodium perborate) during storage and washing. Technically, STP is easy to manufacture as a convenient powder, and does not produce any significant health hazards during its manufacture and use. Of course, the higher the level of STP in detergents, the better its performance. This is why, in the late 1960s and early 1970s, North American and European laundry detergents contained up to 60% STP. According to the nature of the material to be washed, several categories of detergents are available:

- light-duty detergents for delicate materials (30 40 °C);
- detergent for colored and synthetic fabrics (60 °C);
- detergents for pre-washing;
- heavy-duty detergents for high temperature washing (up to 90 95 °C).

In liquid detergents, STP is usually replaced by sodium (or potassium) pyrophosphates which play a role similar to that of STP (See Table 3).

As a result of recent political and environmental pressure, STP levels in laundry detergents have decreased to the 20 30% range. Detergent manufacturers have reformulated their products, trying to achieve a balance between consumer demands for clean laundry and the public demand for less phosphate in detergents. Below approximately 25% STP, the precipitation of Ca_3 $(PO_4)_2$ starts to occur in detergents. This damages fabric fibers and washing machine heaters. The search continues for an alternative to the triphosphates in detergents, hoping to find a more effective or cheaper substance. From the late 1960s, environmental concerns about the use of phosphate in detergents strongly intensified this research. Of the many organic and inorganic materials proposed as a substitute, there are only two options: the sodium salt of nitrilotriacetic acid and synthetic sodium aluminum silicates (zeolites).

Table 3 Composition of typical European detergents (%)

Constituent and function	Light-duty powders	Heavy-duty powders
STP, builder	2–15	20–25
Surfactants	25–32	10–15
Sodium perborate, bleaching agent	–	15–25
Sodium Sulphate, filler	10–40	0–30
Sodium silicate	–	5–10
Minor ingredients and water	Up to 100	Up to 100

REFERENCES

Chiaudani, G and Premazzi, G (1988) *Appraisal of the Possible Methods of Combating the Threat of Eutrophication in Community Waters*, Commission of the European Communities, (J R C) Ispra, Italy.

Forsberg, C (1994) The Large-scale Flux of Nutrients from Land to Water and the Eutrophication of Lakes and Marine Waters, *Mar. Pollut. Bull.*, 29, 409 413.

Vollenweider, R A, Rinaldi, A, and Montanari, G (1992) Eutrophication, Structure and Dynamics of a Marine Coastal System: Results of Ten-year Monitoring Along the Emilia Romagna Coast (Northwest Adriatic Sea), Marine Coastal Eutrophication, *in Science of the Total Environment*, Supplement 1992, Elsevier, Amsterdam.

Vukadin, I, Dujmov, J, Sucevic, P, and Stojanoski, L (1997) Enrichment and Distribution of Anionic Detergents and Phosphate in the Coastal Waters of The Eastern Adriatic, *Fresenius Environ. Bull.*, 6, 84 94.

Vukadin, I (1992) Impact of Nutrient Enrichment and its Relationship to the Algal Bloom in the Adriatic Sea, *Mar. Poll. Bull.*, 23, 145 149.

FURTHER READING

OECD (1982) *Eutrophication of Water Monitoring, Assessment and Control*, OECD, Paris.

Polders

see Land Reclamation from Seas (**Volume 3**)

POPs: Persistent Organic Pollutants

see International Environmental Standards (**Volume 4**)

Population Change, Urban

see Urban Population Change (**Volume 3**)

Population Trends, Global

see Global Population Trends (*Opening essay*, **Volume 3**)

Prehistoric People and Forest Soils

Ian Simmons
University of Durham, Durham, UK

During the 19th century, a great deal was learned about the nature of soils in pre-industrial agriculture, since the 'improvement' of cropping in the European nations' empires was actively sought. That knowledge has often been the basis of the interpretation of relationships between prehistoric people and soils. One nding is the idea of a tension between the forest as a land bank, convertible to farming, and its use as a resource of many kinds: wild food and wood products most notably. During the 20th century it was realised that the forest-farming connection was also the site of traditional ecological knowledge and this has informed a view of prehistory that the farming systems were usually not 'primitive' but sensitively tuned to the capacities of the environment and the cultural desires of the people.

Forest soils have never been globally homogeneous but they do share some ecological characteristics. In their natural state they generally exhibit tight nutrient cycles, for instance. The mineral nutrients in the ecosystem circulate between the tissues of the tree and the soil so that while they are in the living vegetation they are bound into complex molecules. They then fall as litter and are decomposed into simpler forms and enter the soil. The root system of the tree, possibly aided by mycorrhizal fungi, then takes up the mineral ions in solution and incorporates them once more into living tissues. The times and states when the ions might be lost to runoff or downwash are few. The soils under such vegetation tend to be deeply penetrated by humic material whose concentration attenuates gradually down the profile so that visible separation of the horizons in the soil profile is absent and the organisms which perform the task of decomposition are adapted to shady conditions. The water relations of the system are also dominated by the trees: a canopy in leaf intercepts a proportion of the incident precipitation and allows it to evaporate back into the atmosphere. The water which reaches the soil is partitioned: some runs through the soil and into runoff, but another proportion (the relationship between the two components is very variable according to climate, forest type and topography) is taken up by the tree roots and transpired back to the atmosphere via the stomata in the tree s leaves. In the wet climates of Northwestern Europe, acid brown earth soils under deciduous forest were transformed into podsols during the mid-Holocene period (Dimbleby, 1962).

If by prehistoric, we mean pre-literate or before European contact (rather than simply pre-industrial), then the human impact upon these types of flows and cycles may be seen as either obviously temporary (on a time-scale of a decade or less) or apparently permanent on a time-scale of centuries. In the first case, *shifting cultivation* has been common in many biomes (Nye and Greenland, 1960) (*see* Shifting Cultivation and Land Degradation, Volume 3). While it is most commonly associated with the tropics (sometimes called *slash-and-burn agriculture*), shifting cultivation was probably also used by prehistoric populations in Northern Europe when they first replaced hunter-gatherer groups. It was practised in the coniferous forests of Finland until the 19th century. The second instance records the quasi-permanent removal of forests for grassland management (for the production of domesticated or wild animals) or for tillage. In some places, permanence was probably not intended but became necessary to support growing human populations: in others, the forest reclaimed the land if the population decreased after, e.g., famine or epidemic disease.

In ecological terms, clearance of forest carries certain quite fundamental consequences for the dynamics summarised above. The removal of trees, for example, takes away the dominant plant of the ecosystem. Thus, those elements of flora and fauna which depend upon the tree s products for their life cycles and nutrition (and especially the soil flora and fauna) are deprived of their niches. The absence of trees usually allows daytime temperatures at ground level to rise and thus exceed the tolerances of some of the litter-layer organisms, so that the rapid processing of plant litter no longer occurs. But human activity often involves burning the accumulated litter and the debris from felling, so that the cleared area to some extent mimics the pre-existing conditions and indeed pH may rise initially (Sillitoe, 1995). Fire may however bake the top of the soil and kill the remaining flora, fauna and components of the seed bank. Nitrogen is lost in smoke and from rapidly decomposing organic matter at the soil surface (Lal *et al.*, 1986). In any case, the replacement vegetation under agriculture is likely to produce less litter (since most of the plant material will be harvested by humans) and so the humus content of the soil becomes lower. This in turn makes for a less tightly bound crumb structure and so the soil is more prone to erosion, especially on steep slopes. Since the trees have gone, the soils are more open to the impact of the intensive rainfall often found in, e.g., the tropics. Vulnerability to erosion and downwash is therefore increased: bare soil may yield 100 times as much silt as forest. The absence of tree rooting systems also means that the water-pump effect of the trees is lost. Neither canopy evaporation nor root-trunk-leaf transpiration systems is present to remove water and so the soil may either shed the water quickly (which it may do if it is baked by fire, for instance) thus enhancing its erosive power, or it may stay in the upper

horizons of the soil and waterlog them. Hence removing only some trees to form a clearing may result in the death or retarded growth of others. In very wet environments, such as that of upland Britain in the mid-Holocene, the plants able to tolerate the wet and acid conditions fail to break down completely when dead and accumulate to form considerable depths of peat, even on water-shedding slopes. Just as such peat may act as a sponge and hold water, as did the root layers of the original forest, so downwashed soil may accumulate at the foot of a slope and be of considerable agricultural value in its new setting. In a similar fashion, upland forests in the Rocky Mountains of North America hold the snowpack until well into the spring, making for a steady release of water.

So the magnitude of environmental impact upon forest soils of clearance by prehistoric people has been quite high and not always covered over by later changes: the peats of upland Britain are still there, for instance. But later periods of human occupation, with advanced technology, have often produced even greater changes. It is worth noting that one of humankind s major alterations of the face of the earth is the agricultural terrace, which acts as a check on the downward movement of soil and water (padi rice is an extreme case) which have often been mobilised or released by forest removal. Also on the positive side, shifting agriculture seems to have been a reasonably productive system provided that the exhaustion of the cleared plot could be followed by a period in which the nutrient and water cycles were reconstructed by colonising woody vegetation which ended in the re-establishment of closed-canopy forest. Traditional shifting agriculture, for example, loses very little soil down slope. Colonial administrators in the tropics often thought of shifting cultivation as primitive and wasteful without recognising that fallow under forest was an exact analogy of fallow periods in temperate agricultural systems. As in the management of most biological resources, however, too heavy a cropping regime (i.e., too short a fallowing cycle) produces poorer crops and less resilient agro-ecosystems.

REFERENCES

Dimbleby, G W (1962) *The Development of British Heathlands and their Soils*, *Oxford For. Mem.*, 26, Oxford.

Lal, R, Sanchez, P A, and Cummings, R W (1986) *Land Clearing and Development in the Tropics*, A A Balkema, Rotterdam.

Nye, P H and Greenland, D J (1960) *The Soil under Shifting Cultivation*, Commonwealth Bureau of Soils Technical Bulletin 51, Harpenden.

Sillitoe, P (1995) Fallow and Fertility Under Subsistence Cultivation. II. Soil Fertility, *Singapore J. Tropical Geogr*., 16, 101 115.

Producer Subsidy Equivalent (PSE)

see Agricultural Subsidies and Environmental Change (**Volume 3**)

Projection of Future Changes in Climate

see Projection of Future Changes in Climate (*Opening essay*, **Volume 1**)

Public Health – the WHO Definition

see Policy Responses to Public Health Issues Relating to Global Environmental Change (*Opening essay*, **Volume 4**)

R

Recreation, Landscape Modification for

see Geomorphological Change: Landscape Modification for Recreation (**Volume 3**)

Red Tide

Red tide is a common name for an algal bloom accumulating as a dense visible patch with reddish tinges in the surface of the water in seas in coastal areas. Under such conditions, the water appears to be red. The term red tide is thus a misnomer because they are not associated with tides and they are not usually harmful. On the other hand, algal species that are harmful may never reach the densities required to discolor the water. Thus, the term *harmful algal blooms* is now preferred and used by scientists to refer to marine algal bloom phenomena containing toxins that cause negative impacts (*see* Harmful Algal Blooms (HABs), Volume 3).

IAN DOUGLAS *UK*

Redox Potential

see Redox Potential (**Volume 2**)

Refugees and Human Conflict

Phil O'Keefe and John Kirkby
University of Northumbria at Newcastle, Newcastle, UK

Throughout the world, 50 million people are displaced from their homes - a population almost as great as that of the UK. Put in stark terms, one person in every 120 has been driven from home. Some 13 million of them are legally recognized as refugees: they have left their country to escape persecution, violence or armed con ict. The remainder, a mixture of stateless people, asylum seekers, the internally displaced and returnees, are sometimes recognized. They all need humanitarian assistance. They urgently need help with food, shelter, water and medical care. Just over half of that 50 million receive help, largely through the United Nations (UN) and international non-governmental organizations (NGOs). The remainder must survive as best they can.

The 1990s were dominated not only by complex, protracted emergencies, but also, in the former Soviet Union, former Yugoslavia and Africa, by protracted local wars that frequently gave birth to new states or challenged national sovereignty. Early on in the decade, the Gulf War (1991), a politico-military intervention, colored all subsequent humanitarian assistance.

The Gulf War heralded a new era of cooperative measures in defense of international security. It also posed a humanitarian problem because the Western Alliance was unwilling further to protect the Iraqi population, especially the Kurds. Subsequent episodes, in Bosnia and Somalia, showed an increasing willingness to intervene in the internal affairs of a sovereign nation state but at a cost. In

Bosnia, the cost was the realization that peacekeeping to allow the delivery of humanitarian aid was dangerous when there was no peace to keep; humanitarian agencies were also conscious that negotiating access with the combatants frequently involved rewarding them with humanitarian supplies as the price for access to the needy. The cost in Somalia was a flawed intervention that rendered an international community increasingly reluctant to engage in other complex emergencies, particularly when it meant a military involvement. The cost of this reluctance was the withdrawal of the UN Assistance Mission from Rwanda and the subsequent genocide. The position was only reversed in Kosovo.

The Gulf War, the North Atlantic Treaty Organization s action in former Yugoslavia and the UN intervention in East Timor raise, in their different ways, difficult questions for humanitarian assistance. Can military intervention legitimately be used without a UN mandate to protect refugees? Can systematic violations of human rights go unchecked? Could active peacekeeping in support of humanitarian assistance take place without military intervention? Can the UN respond to its charter and rise to the defense of a shared humanity? Will the international community respond to peace with the commitment to consistency that it sometimes gives to war?

A complicating factor is that not only are refugees victims of war, but also of natural disasters (earthquakes, floods, droughts, etc.). In this latter case, climate change and sea level rise over the next 100 years must also be considered (not discussed in this paper).

HUMAN IMPACT OF EMERGENCIES

One of the most challenging humanitarian tasks is the relocation of people usually called *repatriation*. According to the US Committee for Refugees, the number of people in need of emergency assistance, in January 1998, was around 36 million. The single largest group was internally displaced persons estimated at over 18 million. Repatriation is an inappropriate term for relocating internally displaced persons.

The United Nations High Commission for Refugees (UNHCR) presents other data. On the 1st of January, 1999, the total number of people of concern was 21 459 620, of whom 11 491 710 were refugees. The agency estimates that, worldwide, there are a total of 30 million internally displaced persons, of whom five million receive UNHCR assistance. It suggests that the total number of people forced to flee their homes is about 50 million.

Although there are common characteristics, relocation in each emergency is always unique, not least because of different histories of architecture and settlement patterns. Relocation now often takes place in conflict or at least in unstable or insecure peace processes. Refugee and internally displaced person camps may themselves become targets of war. Most relocation is voluntary repatriation, that is spontaneous relocation without external supervision. Voluntary relocation raises a dilemma for humanitarian agencies. Their duty to protect does not allow them to encourage return to unsafe areas, but if refugees and internally displaced persons exercise their right to return, what responsibility for assistance and protection can the humanitarian agencies offer? Even with successful relocation, is there is a gap between resources available for humanitarian assistance and longer-term development needs? If relocation is not linked to the rehabilitation of productive capacity, a cycle of displacement is likely to emerge.

The increase in complex disasters has produced an expansion in humanitarian and agency aid and in the amounts of food allocated for emergency relief. Bilateral donor expenditure (excluding food aid) rose to US$3.5 billion in 1994. If emergency food aid and contributions to multilateral institutions from the member countries of the Organization for Economic Cooperation and Development/Development Assistance Committee are included, the total for 1994 would exceed US$6 billion, some 12% of official development assistance expenditure. From the late 1980s to the early 1990s, the volume of bilateral emergency assistance and food aid doubled. In the same period, aid monies and monies channeled through multilateral organizations declined.

Until recently, emergency aid seemed to be relatively well protected, despite declining official development assistance. It was a popular cause, but was always attached to specific events, frequently those in which politicians and pop-stars could present their humanitarian credentials on television (other catastrophes often went unreported and under aided). Nevertheless, the most recent available figures show that distress relief expenditure fell marginally as a proportion of total official development assistance from 5.86% in 1994 to 5.19% in 1995 (excluding food aid). For some agencies, the fall has been even more pronounced, forcing them to retrench (IFRC, 1997).

The European Union (EU), largely through the European Community Humanitarian Office, has become the single largest humanitarian donor; together with the US it provides almost half of the global humanitarian funds. Their distribution is markedly skewed however; for example, three disasters in the Northern Hemisphere (Iraq, Yugoslavia and Haiti) took almost 60% of total bilateral emergency spending in 1995, which resulted in a *per capita* expenditure twice that for African emergencies in the same year. Much of that disparity arises, of course, from differences in local costs.

CHANGING CONTEXT OF HUMANITARIAN ISSUES

In humanitarian interventions, less attention is now paid to the sovereignty of individual nation states and more to human rights. Peacekeeping, conflict resolution and preventive diplomacy are increasingly deployed. Expenditure is difficult to classify since relief, rehabilitation, reconstruction, development or prevention cannot easily be separated in protracted emergencies. The scope of expenditure is widening to include local capacity building, restoring democracy and civic institutions and promoting good governance.

Gender and environment are crosscutting issues in humanitarian assistance. Much of the diplomacy underlying this changing political landscape is out of sight in regional bodies such as the EU, or in dialogue between donor governments and the UN. Humanitarian assistance is marked not so much by donor fatigue as by donor frustration. Political activity to prevent emergencies is becoming more commonplace, including early warning systems for emerging conflicts, and a drive to ensure, in the aftermath of Rwanda, that humanitarian assistance should not be used as a substitute for political intervention.

Debate about the civil and political rights that must underpin the delivery of humanitarian assistance is increasing. The distinctive feature of a rights-based approach is its legal foundation. If it is to be effective, a legal apparatus must be created to monitor the respect, promotion, protection and fulfillment of human rights.

Military interventions, partly based on experience in Bosnia and Somalia, have moved from small vehicles and lightly armed peacekeeping forces, to heavy artillery and battle troops to ensure access to beneficiaries for humanitarian agencies. Armed forces are willing, where necessary, to work alongside humanitarian agencies, although some agencies find this a difficult relationship. Security for humanitarian personnel is increasingly a problem, as they and the assistance they bring become targets for protagonists in complex emergencies.

INSTITUTIONAL TRENDS

Institutional trends in the delivery of humanitarian assistance are clear. Donor pressure, and the internal drive for change, have made responses to complex and protracted emergencies a major issue for the UN. Control is increasingly moving from individual agencies to the secretary general s office and there is a drive to set the complex emergency agenda before both the UN General Assembly and the UN Security Council. Coordination is being strengthened by concentrating overall responsibility for it with the UN Office for the Coordination of Humanitarian Affairs. These moves have been paralleled by efforts to unify fund raising into consolidated appeals.

Individual UN agencies have been undergoing similar change, for example, UNHCR s responsibilities have widened; it has an increasing caseload of refugees and internally displaced persons to support, to repatriate and to resettle. It now has direct responsibility for beneficiaries and so frequently takes the operational lead in the field.

The International Committee of the Red Cross (ICRC) and the International Federation of the Red Cross and Red Crescent (IFRC) have also experienced change over the last decade. Both have expanded their roles, particularly in addressing the needs of internally displaced persons where a UN response is sometimes inappropriate. Donor agencies respect the leverage that ICRC and IFRC can generate because of their neutrality, although concern has been expressed about mandate drift between the two.

Institutional change has also influenced the NGOs. Their involvement in humanitarian assistance has grown very quickly over the last decade, new international and local organizations have emerged and global NGO families have been strengthened. UN agencies are increasingly selecting leading NGO families, rather than individual organizations, as their partners for particular disasters. NGOs compete amongst themselves, and amongst themselves and UN agencies for humanitarian funds and, more generally, for access to development funds. Their mandates have also drifted; partly as a consequence of increased competition, relief organizations have moved towards development, and *vice versa*. Donor agencies, including the UN, are increasingly attempting to evaluate their comparative capacity and performance. As donor pressure increasingly demands low cost, high quality, integrated assistance, many NGOs seek to reclaim a measure of independence, especially in matters of neutrality and impartiality, by reaffirming that their prime responsibility is to pursue justice for the beneficiaries.

These institutional changes have produced calls for legal and technical criteria by which institutions can be judged. This includes the establishment of a code of conduct for the *International Red Cross and Red Crescent Movement and NGOs in Disaster Response*, the *SPHERE Project* that establishes objective standards for the delivery of all humanitarian assistance and the *Active Learning Network on Accountability and Performance in Humanitarian Aid*. Parallel changes have taken place in the academic discussion of protracted emergencies. Key concepts are vulnerability, entitlement and livelihood systems and a demand-side approach, even if such concepts are difficult to capture in the improved early warning systems.

CHANGING APPROACHES

The number and quality of early warning systems for protracted emergencies, e.g., sea level rise, drought, have improved. However, predictions about future political events and social change are inherently difficult and

uncertain and there are limits to formalized models as well as to political risk analysis.

Conflict prevention is a problematic venture in which humanitarian agencies, by themselves, have little or no influence. Thus, they are expected to respond effectively and efficiently to developments that are difficult to predict, and over which they have little control.

Humanitarian actors do not sufficiently recognize that a pervasive humanitarian presence will distort local coping mechanisms and livelihood systems and may additionally fuel, rather than prevent, the conflict in a protracted emergency.

Obtaining access for the provision of food and medical assistance is crucial for saving lives, and humanitarian agencies are increasingly concerned to resolve access problems caused by reluctant governments, rebels and by anarchic violence. However, sometimes the affected population needs security more urgently than it needs material supplies.

Security is also an issue for humanitarian workers and, in many places, has been violated and goods looted and misused in violation of the principles of humanitarian aid. These situations tend to occur when there is a serious imbalance between assistance and protection and where the humanitarian agencies have been unable to protect civil victims of the conflict. Nevertheless, the right of humanitarian access is strongly affirmed, even where it contradicts the principle of national sovereignty.

There has been much discussion on creating effective links between relief and development. However, dealing with protracted emergencies and countries in-between conflict and peace is difficult, and thought will be required about the role of government, especially about its responsibility for reintegrating returning refugees and internally displaced persons. As the volume of assistance for victims of conflict cannot be sustained by development budgets, humanitarian agencies will have to improve the quality and sustainability of their rehabilitation activities. This will require the increased use of local resources, human and material, to strengthen the local economy and to build capacity for local governance.

See also: Environmental Refugees, Volume 4.

FURTHER READING

Danida (1999) *Danish Humanitarian Assistance Evaluation*, 8 volumes, Danida, Copenhagen (also available at www.evaluation.dk).

IFRC (International Federation of the Red Cross and Red Crescent) (1999) *World Disasters Report*, Edigroup, Chere-Bourg.

Kirkby, J, Howorth, C, O Keefe, P, and Collins, A E (1999) A Survey of Evaluation Experience in Complex Emergencies, in *Evaluating Humanitarian Aid*, Wageningen Disaster Studies, Rural Development Sociology Group, Wageningen University.

MacRae, J and Zwi, A (1994) *War and Hunger*, Zed Press, London.

UNHCR (1997) *The State of the World's Refugees: a Humanitarian Agenda*, Oxford University Press, Oxford.

Regenerative Agriculture

see Green Revolution (**Volume 3**)

Restoration of Degraded Ecosystems

see Restoration, Ecosystem (**Volume 2**)

Río de la Plata

see Water Resources: Río de la Plata (**Volume 3**)

River Regulation

Geoff Petts
The University of Birmingham, Edgbaston, UK

Human societies depend upon water for household use, for industrial processes and for irrigation. Viewed from space, the Earth is a green and wet planet, but water scarcity is a major global problem and in 1990, 40% of the world's population of 5.3 billion suffered water shortage. Agriculture is the largest water user, accounting for nearly 80% of consumption. In many areas, dry climate, intermittent drought, and population growth imply an increasing need for water security in agriculture as well as increasing water needs for domestic and industrial supply.

River regulation is the control of river ow for human use. Traditionally, these uses include water supply, ood control, wastewater disposal, and navigation. Large dams and water storage reservoirs are key components of river regulation schemes as are networks of embankments, weirs and locks.

Throughout history, water-dependent civilizations have experienced drought and drought-related pollution, poverty

*and war. Civilizations dependent upon rain-fed agriculture have been less powerful than those that achieved control of the variations of river ows to sustain irrigation agriculture. Over the past century, and increasingly, river regulation has been closely linked to major transfers of water from one basin to another (*see *Inter-basin Transfer (IBT) for Water Supplies, Volume 3).*

Dams and their associated river regulation schemes are often the single largest infrastructure investment undertaken by a country. Until the 1970s, they were seen as engineering challenges offering enormous opportunities for harnessing a seemingly under utilized and renewable resource. During the 1970s and 1980s, the damaging impacts of these schemes upon the environment and social systems were forcefully demonstrated.

Strains on water supply systems in this century will relate not only to continuing population growth but also to rapid urban development in many arid and semi-arid regions, requiring even more intensive development of river regulation and inter basin transfers. River regulation, especially for agricultural and urban water supplies, and for hydroelectric power production, is still a key component of development programs. Scienti c research over the past 20 years has now established the basis for evaluating the sustainability of new river regulation proposals and for planning for the sustainable development of river corridors.

A HISTORY OF RIVER REGULATION

The control and appropriation of water has been fundamental to the building of human civilizations for more than 2500 years. Most of the early civilizations were dependent upon the management of floodwaters to irrigate agricultural lands. The rise of civilizations in the Middle East, Northern India, Southeast Asia, China and Meso-America reflected the sophistication of technology and the power of the centralized state authority, which emerged out of water control. In some cases, the fall of these civilizations related to natural drought or soils becoming infertile because of salt accumulation resulting from mismanagement of the irrigation systems. The combination of hydraulic agriculture, a hydraulic government, and a single-centered society constituted what Wittfogel (1956) referred to as a hydraulic civilization.

The earliest dams for the purpose of river regulation were constructed over 5000 years ago and they were widespread throughout the Mediterranean area by Roman times. By 2000 BC, irrigation agriculture, the development and distribution of potable water supplies to urban centers, and elementary flood control were widespread. However, the first 3000 years of river regulation were dominated by the desire to control river flows in order to increase food production and to generate food surpluses. River regulation was also to provide the foundation for industrial growth using waterpower and urban centers based upon water-borne trade. In England about 3000 BP, early towns, such as Egham on the River Thames, developed as focal points for trade along inland waterways and the Domesday records of 1086 indicate that there were already more than 5000 water mills. Pound locks, the catalyst for a rapid expansion of inland waterway trade, a product of the Italian Renaissance, were widespread across Europe by the mid-16th century. The Industrial Revolution saw water assuming an ever-widening role, for power, navigation, water supply, and waste disposal. With urban and economic growth came wetland drainage, canalization of natural channels for boat traffic, embanking and dredging of rivers for flood control, and the construction of ever-larger water storage reservoirs.

The basis for the modern era of river regulation was established between the 16th and 19th centuries through advances made by surveyors, cartographers, architects, and military engineers. Progressively, the art of river regulation was replaced by a scientific method based upon mathematical calculations and physical laws. The control of major rivers was seen not only as a practical necessity, but also as a way to capture the public imagination, enhance the status of the engineer, and become symbols of national prowess (Cosgrove and Petts, 1990). Gigantic works of water engineering came to signify, especially to the French and British, the natural dominion of imperialist powers over the environments and peoples of other lands. By the end of the 19th century all the major rivers of Western Europe had been regulated. In 1817, the training of the braided Alsatian section of the Rhine had been initiated to establish a fixed, single channel with benefits for navigation, flood control and land drainage. The latter was not only required to create more agricultural land but also to combat malaria by eliminating mosquito habitat. One of the greatest works of the 19th century was the regulation of the River Tisza, which drains the southern and western Carpathians; the works shortened the river by 340 km!

Faith in the power of technology to control the natural environment spread with colonial expansion to the New World and then to the Third World. Led by the popular pioneering vision of Man s struggle to tame nature and entrepreneurs motivated by the desire for economic growth, river regulation was to form the core of plans to develop entire river basins.

Grand water projects were politically and socially acceptable, and they had a major role in generating employment, especially during the inter-war years, including the Great Depression of the 1930 s when, in the US:

> In the interests of the national welfare there must be national control of all running waters of the United States, from the desert trickle that might make an acre or two productive to the rushing flood waters of the Mississippi (US National Resources Planning Board 1934, cited in National Research Council 1992, 42).

The 221 m high Hoover Dam on the Colorado River, USA and its 37 • 10^9 m^3 reservoir Lake Mead, constructed in 1935, provided one model for water development: the creation of massive impoundments. A second model was provided by the Tennessee Valley Authority (TVA) in the US, formed in 1933. Total control of the river was sought by using many, large multi-purpose dams to exploit the full potential of the catchment s water resources (water supply and hydroelectricity) and of the river for navigation, whilst at the same time achieving control over flooding. Today, the TVA has 32 large dams and there are 20 other dams in the basin.

Throughout the US there followed 30 years of construction of integrated systems within large river basins (the Columbia, Colorado, and Mississippi Missouri rivers), consisting of new large, multi-purpose storage reservoirs and expansion and improvement of the inland navigation systems.

By 1965 virtually every river in Europe and North America had been regulated and even small streams had been incorporated into water resource plans. On other continents progress towards river regulation moved rapidly, such that few rivers remain that are not affected by regulation in some way.

Globally, the construction of dams over 15 m high (large dams, as defined by the International Commission on Large Dams) accelerated from less than 100 per year in the 1950s to a peak of global dam building activity, during the 1970s of more than 500 per year (World Commission on Dams, 2000). Today there are more than 45 000 dams over 15 m high. However, one estimate suggests that including small dams, approximately 800 000 dams have been constructed worldwide.

Today, a mega-project can be defined as one that meets one or more of the following criteria: the dam is higher than 150 m, the dam volume is more than 15 000 • 10^3 m^3 or the reservoir storage capacity is greater than 25 000 • 10^6 m^3. Currently, there are 410 mega-projects constructed, under construction or planned. The largest dams are higher than 275 m or impound reservoirs with capacities of over 150 • 10^9 m^3 (Table 1). The largest schemes, such as that proposed for the Caroni tributary of the Orinoco river in Venezuela (Figures 1 and 2), would convert entire river systems into chains of lakes impounded by large dams.

Dams account for 18% of the electricity generated worldwide, provide reliable water supplies in many of the world s expanding urban centers, vital flood control in some areas, and played a key role in expanding irrigated land from 95 220 million hectares between 1950 1975.

ENVIRONMENTAL IMPACTS

In the past, the need to conserve flora and fauna, let alone entire ecosystems, was not simply viewed as a constraint on progress, as it is in many developing economies today; it was irrelevant. The legacy of progress in river regulation over the past 200 years, has been pollution, dried-up rivers, widespread destruction of wetland and aquatic habitat and

Table 1 The world's highest dams and reservoirs

Rank	River	Country	Dam	Year completed	Height (m)
1	Vakhsh	Siberia	Rogun	1990	300
2	Vakhsh	Siberia	Nurek	1980	300
3	Dixence	Switzerland	Grande Dixence	1961	285
4	Inguri		Inguri	1980	272
5	Vajont	Italy	Vajont	1961	262
6	Grijalva	Mexico	Chicoasen	1980	261
7	Bhagirathi	India	Tehri	1995	261
8	Tons	India	Kishau	1995	253
9	Guavio	Colombia	Guavio	1989	246
10	Columbia	Canada	Mica	1973	242
Rank	River	Country	Reservoir	Year completed	Volume (m^3 • 10^3)
1	Dneiper	Ukraine	Kakhovskaya	1955	182 000
2	Zambezi Zambia	Zimbabwe/ Kariba		1959	180 600
3	Angara	Russian Federation	Bratsk	1964	169 270
4	Nile	Egypt	Aswan High	1970	168 900
5	Volta	Ghana	Akosombo	1965	148 000
6	Manicouagan	Canada	Daniel Johnson	1968	141 852
7	Caroni	Venezuela	Guri	1986	138 000
8	Yenesei	Russian Federation	Krasnovarsk	1967	73 300
9	Peace	Canada	Bennett WAC	1967	70 309
10	Zeya	Russian Federation	Zeya	1978	68 400

Figure 1 Guri Dam and hydroelectric power plant on the Caroni River, Venezuela (see Table 1 and Figure 1)

species extinctions. Three indicators of unsustainable water use at the end of the 20th century were:

- the mighty Huang He (Yellow River) (China) and Colorado (US) rivers no longer flowed to the sea in the dry season (*see* Yellow River, China: a Case Study in Water Resources, Volume 3);
- the surface area of the Aral Sea (Kazakhstan) shrunk by 25 000 km^2 because of water abstractions from the Syr Darya and Amu Darya rivers (*see* Aral Sea, Volume 4);
- salinization caused severe loss of fertility in 20% of irrigated lands: (*see* Salinity and Agriculture, Volume 3).

Studies of the environmental impacts of dams (e.g., Petts, 1984) are important for the development of predictive models for the dual purposes of sustaining natural ecosystems and meeting human needs for water resource development and hazard control. Particular attention has been given to the fate of fisheries. Atlantic Salmon (*Salmo salar*) disappeared from many rivers of Western Europe following river impoundment; in Eastern Europe the Sturgeon (*Acipenseridae*) have been similarly affected, whilst in Northwest US the numbers of Pacific Salmon (*Oncorhynchus* spp.) declined dramatically subsequent to dam construction. Along tropical rivers, fisheries are a particularly important component of indigenous cultures; flood regulation by dams can cause a dramatic decline in river fish stocks.

The development of new fisheries in reservoirs is often seen as a benefit of dams. In Africa, for example, reservoir fisheries are important both by virtue of the area they occupy (40 000 km^2 for reservoirs exceeding 10 km^2) and because they contribute 150 000 t of fish; 10% of the overall inland fishery yield (Kapetsky and Petr, 1984). However, in many cases the fish production in the reservoir only replaces that lost from the river downstream, whilst in many other cases the amount of fish lost downstream of the dam surpasses the potential of the dam itself. Thus, below the Kainji dam in Africa, 6000 t of fish have been lost and the reservoir behind the dam yields only 5400 t. Reservoirs that support fish populations and fisheries where yields are higher than the rivers they submerge can be found only where the productivity of the natural river is low (Welcomme, 1985).

Most ecological changes resulting from river regulation can be related to one of four situations:

1. The introduction of dams creates barriers especially to the migration of anadromous fish. Fish passage facilities have been introduced on many dams, including ladders, locks, lifts, and collection and trucking, but with varying degrees of success. The primary problem appears to be the ability of the fish collection system to intercept or to attract migrants by providing entrances at proper locations and with suitable hydraulic conditions. The design and efficiency of fishways is dependent upon a detailed knowledge of the swimming capabilities and behavior of migrating fish; for many species this knowledge is not yet available. Irrespective of the height of the dam, delayed migrations (upstream and downstream) related to fish movement through a reservoir can adversely affect the survival and reproduction of migratory species.
2. All dams change the distribution of flows downstream, they trap sediments being transported along the river, and the retention of water in reservoirs can alter the water quality of the river downstream. In most cases, the flow variability is reduced, the average flow is reduced and sediment loads are dramatically reduced. Nutrient levels and temperatures of water released from the reservoir depend on the depth of withdrawal; deep-water releases from the hypolimnion of the reservoir can cause particular problems, discharging cold anoxic water.

   Such limnological impacts affect the river ecosystem for a long distance downstream of the dam. For example, for 400 km below Glen Canyon Dam on the Colorado, the water is now too cold for most native fish (Holden and Stalnaker, 1975). In other cases, production of algae and plants within regulated rivers can be enhanced by reduced turbidity, flood regulation, and increased water temperatures. The cumulative influence of runoff from non-regulated tributaries and inputs from riparian sources causes the effects of a reservoir to decrease downstream, such that the linear effect of regulation is partly dependent upon the structure of the drainage network. Nevertheless, a down river variation in organic life, with the most dramatic change occurring within the first 2 km below the impoundment, often

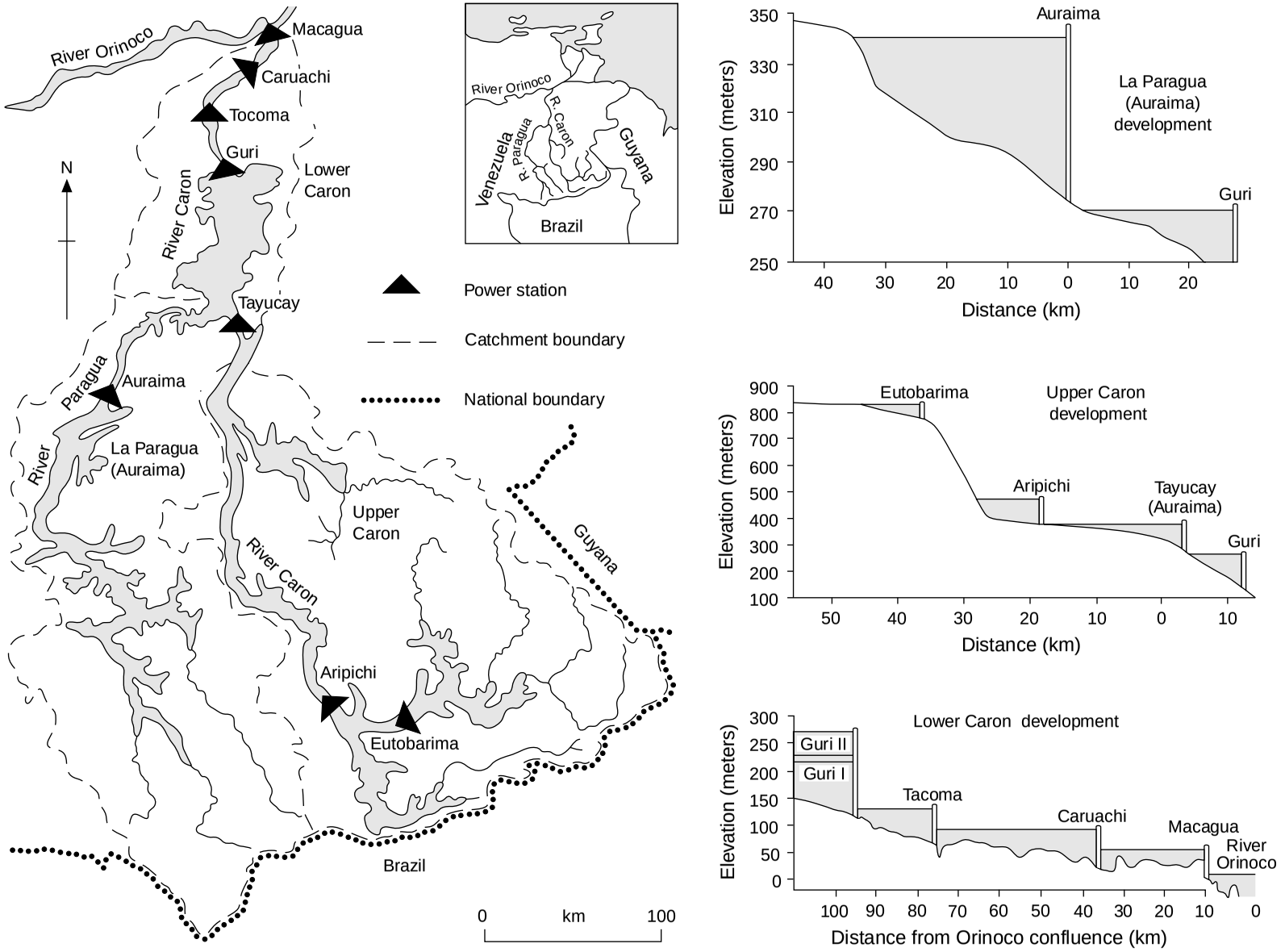


Figure 2 A proposal to regulate the Caroni River Venezuela. The scheme would convert a fast flowing river into a series of slow flowing lakes impounded by large dams

occurs below dams. Such changes are characterised by a lower diversity and higher density of invertebrates. The cumulative effects of river regulation may be even more marked producing major changes in the chemistry of coastal seas and having marked effects on the chemistry of nearly enclosed seas. For example, long term data on the discharge of water and nutrients by the Danube into the Black Sea show a reduction of more than 60% in the dissolved silicate concentration of water in the central Black Sea. The consequent changes in the ratio of silicon to nitrogen in the sea appear to be larger than that caused by eutrophication alone and is thought to be responsible for a dramatic change in phytoplankton species composition from diatoms (siliceous) to coccolithophores and flagellates (non-siliceous). The upstream damming of the Danube at the Iron Gates appears to have been a key factor in causing these changes (Humborg *et al.*, 1997). River regulation has contributed to changes in the biogeochemistry of the entire Black Sea and probably to alterations to food web structures and biogeochemical cycling in coastal seas in many parts of the world.

3. Geomorphological changes are induced by changes of the hydrological and sediment transport regimes. Changes of channel morphology are associated with changes of the instream hydraulics (depth, velocity and shear stress distribution) and substrate characteristics, as well as space and shelter availability. The dimensions of river channels adjust to the hydrological and sediment transport regimes, and a range of channel changes have occurred below dams (Petts, 1984; Williams and Wolman, 1984). Changes in flow and sediment transport have had dramatic impacts on habitats for wildlife (Figure 3).
4. Biodiversity is reduced by the isolation of the river channel from its alluvial plain and by the loss of primary habitats. Separation of the major bed of the river (its floodplain) from its minor bed (the main channel) has had a considerable effect on the ecological diversity of the highly productive alluvial corridor and riverine fish populations. Natural river corridors are biologically diverse and highly productive (Naiman and Decamps, 1990). Floodplain forests, sustained by flooding and periodic rejuvenation by channel erosion, contain a diversity of habitat patches of different types and ages. In turn, the range of habitat types and ages sustains a rich diversity of biota. The elimination of access to backwaters, floodplain lakes, and marshes has severely impacted fish stocks (Welcomme, 1979) by eliminating spawning habitats and refuges during periods of lethal and/or critical water quality.

A fundamental problem in the development of a general model of ecosystem response to river regulation is the failure of many studies to consider observed changes within an appropriate time scale. Most studies of the impacts of dams have focussed on impacts at a single point in time, often within five years of dam closure (Petts, 1987). A wide range of physical, chemical, and biological factors interact to determine the character of a river and to maintain the river system in a state of quasi-equilibrium. The ecosystem recovery process cannot proceed in advance of the physical recovery process. Whilst new regulated hydrological and water quality regimes will stabilize relatively rapidly, over 5 10 years, the physical adjustment of channel morphology can require a time scale measured in tens or even hundreds of years.

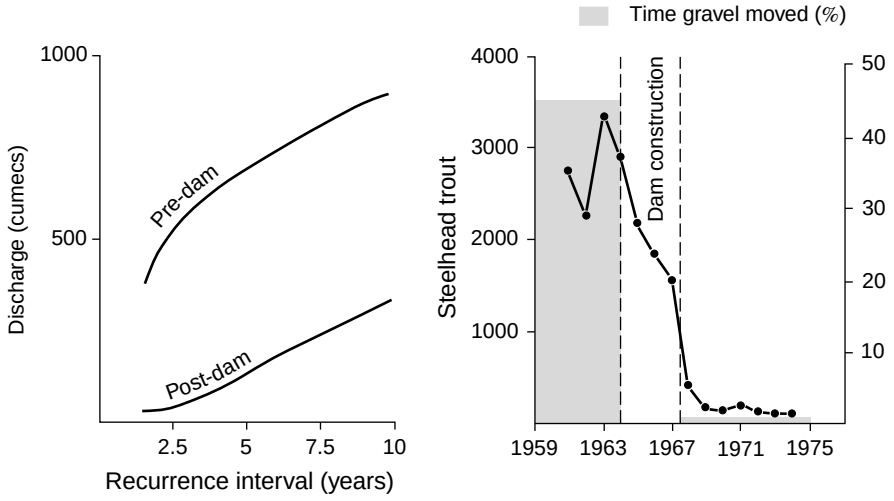


Figure 3 Impact of the Lewiston dam and reservoir on the Trinity River in Northern California, US. Reduced flood magnitudes have dramatically reduced the frequency of sediment movement, causing channel change, and a marked decline in fish numbers

LESSONS FOR THE FUTURE

Despite the growth in environmental awareness, for many, water running to the sea is perceived as a lost resource; floods are viewed as major hazards; and wetlands are seen as wastelands. Yet the nature conservation value of river corridors and the ecological services they provide are dependent on the natural hydrological regime, and its related biochemical and geomorphological dynamics.

The real question is not whether rivers should be regulated but what steps should be taken to reduce the adverse environmental impacts to a minimum. There are two important lessons for the future. First, all dams will cause environmental changes. Any change in the hydrological regime will cause ecological change. However, research on ecological dynamics within regulated systems is beginning to define the sensitivity of species, communities and environmental processes to hydrological change; approaches are being developed to enable selection of sites for development that are least sensitive to change; and options are being developed to mitigate unacceptable ecological changes.

Secondly, in many cases the most dramatic impacts have been shown to result from poor planning, ineffective administration, and bad management of the water resources scheme. A dam and reservoir may have been the catalyst for change but often the major impacts resulted from the lack of appropriate planning, administration, and operational management. The failure to consider the secondary and tertiary impacts of water development schemes on the environment and on people during the early stages of project assessment. One of the most serious of such secondary impacts of dam construction is degradation of soil fertility by waterlogging, salinization and alkalization caused by irrigation without adequate drainage facilities or controls on rates of water application. Many of the health problems associated with water development projects are also linked to secondary impacts including irrigation and the immigration of people with either new strains of endemic diseases or new diseases not normally found in the project area. For example, schistosomiasis and river fluke infections are often high in areas of perennial irrigation.

Sustainable socioeconomic development needs dams; it also requires water engineering projects that are environmentally sound. A strategy for sustainable river regulation to conserve biological diversity and important ecosystem functions along river corridors requires:

1. the allocation of water and land to meet ecosystem needs;
2. comprehensive catchment and national planning and coordination and in some cases international plans that incorporate both development and conservation objectives;
3. community participation in the decision-making process and project management;
4. long-term monitoring.

However, despite dialogue to formulate environmental guidelines for water projects, such guidelines are often difficult to implement, have seldom been taken seriously by governments and are commonly ignored by project officers of development agencies.

River regulation works are built with a prospective lifetime of around 100 years. Reservoirs are usually managed with a defined flood storage capacity. Such planning usually assumes no long-term change in climatic or river regime. Many existing projects have seen severe consequences of upstream land use and land cover change, particularly through increased runoff and siltation. Global climate change will add to these impacts and exacerbate problems of decision making in relation to storage of water for downstream users and flood detention capacity. As changes in storm intensity and frequency may aggravate both upstream soil erosion and downstream channel stability, the designed pattern of river regulation may become inadequate to cope with future events. Stresses on aquatic and riparian ecosystems may increase. Reservoir managers and river regulators need to consider all these possible impacts of climate change when formulating their sustainable catchment management and river regulation plans. For information on the science and management of regulated rivers there is a dedicated international journal: *Regulated Rivers: Research and Management*, published by John Wiley. For an interesting example on water management in a river basin.

See also: Indus Basin: a Case Study in Water Management, Volume 3.

REFERENCES

Cosgrove, D E and Petts, G E (1990) *Water, Engineering and Landscape*, Belhaven, London.

Holden, P B and Stalnaker, C B (1975) Distribution and Abundance of Mainstream Fishes of the Middle and Upper Colorado River Basins, 1967 1973, *Trans. Am. Fish. Soc.*, 104, 217 231.

Humborg, C, Ittekkot, V, Cociasu, A, and Bodungen, B V (1997) Effect of Danube River Dam on Black Sea Biogeochemistry and Ecosystem Structure, *Nature*, 386, 385 388.

Kapetsky, J M and Petr, T (1984) *Status of African Reservoir Fisheries*, Food and Agricultural Organisation of the United Nations, Rome.

Naiman, R J and Decamps, H (1990) *The Ecology and Management of Aquatic-terrestrial Ecotones*, Carnforth, Parthenon.

National Research Council (1992) *Restoration of Aquatic Ecosystems: Science, Technology and Public Policy*, National Research Council, National Academy Press, Washington, DC.

Petts, G E (1984) *Impounded Rivers*, John Wiley & Sons, Chichester.

Petts, G E (1987) Timescales for Ecological Change in Regulated Rivers, in *Regulated Streams: Advances in Ecology*, eds J F Craig and J B Kemper, Parthenon, NY.

Welcomme, R L (1979) *Fisheries Ecology of Floodplain Rivers*, Longman, London.

Welcomme, R L (1985) *River Fisheries*, Technical Paper 262, FAO, Rome.

Williams, G P and Wolman, M G (1984) *Downstream Effects of Dams on Alluvial Rivers*, US Geological Survey Professional Paper 1286.

Wittfogel, K A (1956) The Hydraulic Civilisations, in *Man's Role in Changing the Face of the Earth*, ed W L Thomas, University of Chicago Press, Chicago, IL, 152 64.

World Commission on Dams (2000) Dams and Water: Global Issues, Local Experiences, www.dams.org/global/.

FURTHER READING

Petts, J I (1999) *Handbook of Environmental Impact Assessment, Blackwell Science*, Oxford.

S

Salinity and Agriculture

Arthur Conacher
University of Western Australia, Crawley, Australia

Extensive coastal and inland areas in arid and semi-arid environments, worldwide, are characterized by the natural accumulation of soluble salts in soils, groundwater and ephemeral streams and lakes. Such areas are referred to as areas of primary salinity. In contrast, regions where salt concentrations in soils and water have increased (salinization) as a result of human actions are referred to as being affected by secondary salinity. These are the areas of increasing concern which are the subject of this article.

Agriculture has long been known to cause salinity problems; whether by irrigation, cultivation under natural, rainfed regimes (dryland agriculture), intensive or even extensive pastoralism. The problem, very simply, is related to excessive concentrations of soluble salts in water or soils. Discussions of the salinity problem are usually couched in terms of the adverse effects of high salt concentrations on agriculture and downstream water users - human consumption, industry and infrastructure - with relatively less emphasis on natural ecosystems. Even fewer writers comment on the collapse of civilizations because of salinization. Direct causes of the problem involve changes to landscape hydrology and are reasonably well understood. Solutions, usually involving political, social and economic responses, human attitudes and perceptions, as well as technical measures, are much more complex.

SALINITY CAUSED BY IRRIGATION AGRICULTURE

Irrigation in arid and semi-arid environments inevitably, it seems, results in *waterlogging* (*see* Waterlogging, Volume 3) and *salinization* of the land and water. The best known early example is Mesopotamia, where the development of early civilizations and settlements was dependent on irrigated farming in the valleys of the Tigris and Euphrates. Gradually, changes in environmental conditions were reflected in the agricultural crops. The increase of salinity in ancient Southern Mesopotamia (now Iraq) affected the growth of wheat, which is salt-sensitive, which was gradually replaced by the more salt-tolerant barley. From grain impressions in pottery dated 5500 years BP (before the present), the proportions of wheat and barley were equal. One thousand years later wheat comprised only 20% of the crop and 1500 years later only 2%, being abandoned by 3700 BP (Goudie, 2000). Eventually, in the kingdoms of Sumer and Akkad, even the substitute barley could not grow and people were forced to abandon the area (Ghassemi *et al.*, 1995).

Historic evidence from other parts of the world, including China, the Indus River basin, South America and Arizona, indicates that the problem of human-induced salinization permeates history. Now-desertified parts of Roman North Africa, the bread basket of ancient Rome, are another example. Remnants of historic irrigation channels can be identified from aerial photographs, and salinization of this previously productive region was yet another factor contributing to the fall of the Roman Empire. The persistence of some springs and remnants of forests in the Tell and the Atlas indicate that the North African climate has been relatively stable for the last 4000 years. Land degradation, therefore, was primarily a consequence of human actions (Conacher and Sala, 1998). Here as elsewhere, people were forced to move on, often causing salinization in the newly settled areas.

More recently, in the later part of the 19th century, British engineers constructed large-scale irrigation projects in India and Pakistan to increase food production and reduce the risk of famine. Of some 17 million hectares (Mha) of irrigated agricultural land in Pakistan, in the Indus Plain, at least 25% of the land had become salt-affected by the late 1970s, with the electrical conductivity (EC) of soils exceeding 4000 micro Siemens per centimeter (and 8% with $EC > 15\,000\,\mu S\,cm^{-1}$). A decade later, surveys showed that groundwater tables below 8.8 Mha of the irrigated lands were within 3 m of the soil surface at the end of the monsoon season (Ghassemi *et al.*, 1995). As discussed below, soil salinity problems are closely associated with the proximity of groundwater to the soil surface. Goudie

(2000) reported estimates that salt-affected and waterlogged soils comprised 50% of the irrigated area in Iraq, 23% of all Pakistan, 50% in the Euphrates valley of Syria, 30% in Egypt and over 15% in Iran. The irrigated San Joaquin Valley in Southern California is another recent example. Of more than 2.26 million irrigated hectares, 900 000 ha experience high salinity and at least 600 000 ha are characterized by high water tables (Conacher and Sala, 1998).

Salinity problems in irrigated agricultural areas affect not only the irrigated crops but also the quality of the water beneath the soil (the groundwater) and in the streams or rivers, which are usually associated with these areas. Downstream uses of water for irrigation, to water stock, to maintain wetland ecosystems, or for industry or even human consumption may be affected adversely. The Murray-Darling Basin, which drains one-seventh the area of the Australian continent, is an example of all the above aspects. River water salinity increases progressively down the Darling and Murray rivers as a result of return flows of irrigation water and concentration by evaporation. River discharges are being manipulated to provide environmental flows to preserve important wetland ecosystems under the international Ramsar convention. The one million inhabitants of the city of Adelaide near the mouth of the Murray River rely on the river for about 40% of their water supply. Yet the State of South Australia, of which Adelaide is the capital, has no control over most of the Murray-Darling Basin, which is largely under the jurisdiction of other states Victoria, New South Wales and, to a lesser extent, Queensland. This provides a nice case study of realpolitik, with the Murray-Darling Commission being given the responsibility of dealing with the problem (Murray-Darling Basin Commission, 1993; Conacher and Conacher, 2000).

The most common salts which affect water users include sodium, calcium and magnesium combined with chloride and sulfate. But in some areas other toxic elements are also present. Concentrations of selenium, e.g., exceed 4000 parts per billion (ppb) in parts of the San Joaquin Valley in California, in contrast to naturally occurring amounts of 0.2 and 0.1 ppb in fresh and marine water, respectively. Selenium can be transmitted through the food chain, with toxic effects on humans and animals. At one location in the Valley, some 20% of waterfowl nests in wetlands were found to contain deformed embryos and 40% had at least one dead embryo (Tanji *et al.*, 1986; reported in Conacher and Sala, 1998).

Causes of the Problem

The fundamental problem concerns the manipulation of water and all water contains some salts. In irrigation agriculture, the problem occurs when more water is added to the soil (by irrigation practices or from leaking irrigation channels) than is drained from the subsoil and used by plants. As a result, water accumulates in the soil and evaporates from the surface recalling that irrigation is generally practiced in hot and relatively dry environments leaving the salts behind. Over time, the concentrations of the salts in both soils and water increase, eventually reaching toxic levels (with the concentration of concern varying according to the land use). Waterlogging and problems relating to poor soil structure are commonly associated with salinization. Theoretically, many crops can be irrigated with water containing relatively high concentrations of soluble salts provided that the soils are thoroughly drained or flushed (Hollaender *et al.*, 1979). This requires larger quantities of water than are usually available to irrigators and well-drained, coarse-textured soils and groundwater tables many meters below the soil surface. Since irrigation areas are often located in low-lying, generally flat, alluvial flood plains of river systems in relatively arid areas where the rivers have ephemeral or seasonal flow regimes, these requirements are rarely, if ever, met. Moreover, irrigation can raise water tables from quite considerable depths to within 1 2 m of the surface within only a few years, especially if drainage is poor.

Depths to groundwater tables are a key variable. If the soil surface is within the capillary fringe of the groundwater table, then water can be drawn to the surface and evaporated, leaving the soluble salts behind. Theoretically, capillary heights are almost infinite. They are greatest with silt-textured materials (particle sizes between 0.002 and 0.02 mm in diameter) and least in coarse sand. Most soils or sedimentary (alluvial) materials are heterogeneous, with capillary tubes being interrupted by the presence of coarse layers. A practical rule of thumb of about 2 m is often used; but the real possibility of capillarity (and hence salinity) developing over significantly greater depths to groundwater must be recognized (Peck, 1978). Problems of waterlogging and salinization can also arise from the development of perched water tables, if subsoil layers are relatively impermeable. Water infiltrating into the soil can be held up (perched) close to the surface, above these relatively impermeable layers, and prevented from reaching the deeper groundwater aquifers. Such layers often develop as a consequence of repeated cultivation and in response to waterlogging and deterioration of soil structure. Thus it is not always necessary for the water table of the permanently saturated zone to rise close to the surface for salinity problems to develop.

Salts are present naturally in varying concentrations in underlying rocks and sediments, and in rain and dustfall. Salts released from rocks by chemical weathering eventually reach the sea (an interesting question is whether the salinity of our seas is increasing, and if not, why not), with some released into the atmosphere as salt spray, returned to the land in rain and eventually to the sea (so-called cyclic salts). Coastal areas with onshore winds receive

considerable quantities of salt in this way from 10 to as much as 50 kg ha^{-1} for each 100 mm of rainfall each year. In more arid areas, even at some distance from the sea, or where underlying lithologies or groundwaters contain soluble salts in appreciable quantities, salts tend to accumulate over time as there is insufficient rain to flush them from the system. Increased salt concentrations in soil water in the root zone of plants affect osmotic pressure potentials across the root/soil interface. Normally, plant roots take up water (and nutrients) by a positive pressure differential across the root/soil interface. But where salt concentrations are sufficiently high, the osmotic pressure differential is reversed. Water movement then is (potentially) from the plant to the soil. Even though the plant may be standing in water, it dies of thirst.

SALINITY CAUSED BY DRYLAND AGRICULTURE

The development of secondary salinity problems in non-irrigated agricultural areas seems to be a much more recent problem or at least, one which has been recognized only recently. Australia provides a good case study. The first accounts of the development of salinity problems were recorded from Western Australia and Victoria late in the 19th century, soon after the natural woodland vegetation was cleared for agriculture by the European settlers. By 1996, more than 2.4 Mha of agricultural land in Australia had been rendered unproductive due to increased concentrations of soluble salts in the soils, up from only 0.4 Mha as recently as 1982 (Conacher and Conacher, 2000). All States have areas of secondary, dryland soil salinity. Western Australia is the worst affected, with more than 1.8 Mha in 1996. This is predicted to increase to over 6 Mha by equilibrium (i.e., when groundwater tables have stopped rising): the equivalent national prediction exceeds 12 Mha. Increased salinity of water in streams, rivers, lakes and dams is an associated and possibly more serious problem. In the southwest of Western Australia, all rivers which have their headwaters in the agricultural areas are unfit for human consumption, even though they were fresh before the development of agriculture. This has had serious ecological as well as economic consequences, particularly as numerous stream and other wetland habitats for aquatic fauna have been destroyed.

Similar problems of secondary soil salinization occur in the North American Great Plains, from Alberta, Manitoba and Saskatchewan in Canada down through Montana and the Dakotas in the US. More than 600 000 ha have been rendered unproductive in that region. Dryland salinity also occurs in parts of South Africa, Turkey, Thailand, India and Argentina.

Ghassemi *et al.* (1995) cite a 1991 source in providing data for the global extent of human-induced salinization (Table 1). It is not clear whether the data refer to salinity in dryland as well as irrigated agricultural areas, although the Australian data imply that they do, or what the proportions of the two forms of salinity are. Moreover, as a result of improved survey techniques, the situation in Australia, at least, is now known to be much more serious than the 0.9 Mha indicated in their Table.

Table 1 Global extent of human-induced salinization

Continent	Millions of hectares
Asia	52.7
Africa	14.8
Europe	3.8
North and Central America	2.3
South America	2.1
Australia	2.4
Total	78.1

Source: Ghassemi *et al.* (1995: 19); (Australian data from Conacher and Conacher, 2000, 57).

Causes of the Problem

As in irrigated areas, secondary salinity in dryland agricultural areas is caused by too much water in the wrong places. And again there is the striking irony of this situation in semi-arid environments, where the single most important limiting factor for crop growth is insufficient water. However, the mechanisms leading to the accumulation of excess water are quite different.

Fundamentally, the problem in dryland agricultural areas is caused by the replacement of indigenous vegetation with introduced crops and pastures. In Australia, the native vegetation is usually a woodland association. The plants have deep root systems, which enable them to draw on subsoil moisture for the entire year, sometimes from considerable depths ($>$20 m). In contrast, the introduced agricultural species have shallow root systems and they grow only during spring and early summer. Thus, less water is transpired by the agricultural plants than by the native vegetation, resulting in the accumulation of water in the soil.

There have been various studies into the extent of this imbalance, particularly in the south west of Western Australia. In general terms, in areas with mean annual rainfalls of around 300 400 mm, the amount of excess water in the landscape is only of the order of 6 15 mm yr^{-1}, which infiltrates to groundwater and also forms perched soil-water systems. Because valley floor gradients in the ancient landscape of the semi-arid, Western Australian wheat-belt are very gentle, and hydraulic conductivities of the deep, intensely weathered (kaolinized) lateritic soil materials are very low (as little as 0.009 m day^{-1}), the rate at which groundwater drains from catchments is as little as 0.05 0.3 mm yr^{-1}. Thus, since clearing of the native vegetation, more water is entering the groundwater systems

than is able to drain from them. As a result, groundwater tables in the Western Australian wheatbelt are rising at rates of between 0.02 and 0.3 m yr^{-1} in areas receiving less than 350 mm of rain a year, to 0.15 1.5 m yr^{-1} in areas with more than 500 mm of rainfall annually. Mixing of these deeper, saline groundwaters with near-surface, relatively fresh aquifers is common (Conacher and Conacher, 2000).

As in irrigated areas, saline water accumulates to within the capillary fringe of the soil surface, water is evaporated and soluble salts are left behind, to concentrate in the soil and at the surface. In many situations salty groundwater also emerges at the surface as saline seeps, which then contribute salty water to adjacent streams.

The consequences are similar to those in irrigated areas loss of soil structure and fertility, increased erosion by water and wind on the bare, salt-affected surfaces, waterlogging, reduced agricultural productivity, salinization of streams, rivers and lakes, and severe degradation of habitats with the loss of several species. Additionally, many farm houses and towns in Australia were built adjacent to (then) fresh streams and rivers. The water was used for human consumption, for steam locomotive boilers, to irrigate vegetables and fruit trees, for recreation, and for a range of other purposes. Unfortunately, these low-lying areas have been amongst the first to be affected by waterlogging and increased salinity. As a result, many farm houses and their adjacent building complexes have had to be abandoned, and some 80 rural towns in Australia are being affected by rising, saline groundwater tables or the development of perched water tables. Mortar and bricks fret and crumble, playing fields and home gardens become dust(mud)bowls, and the basements of hotels and other large buildings are filling with salty water.

In North America, the original tall prairie grasses were also replaced by agricultural crops and pastures. Here the hydrological disruption is more finely balanced than in Australia, as the differences in transpiration of water between introduced and indigenous plants, and therefore the amount of excess water infiltrating into the soil, are much less. The nature of the soils and underlying materials, and the composition of the salts, also differ from those in Australia. But whilst the consequential saline seeps are very similar, it has been easier for land managers in North America to rectify the hydrological imbalance.

TWO SPECIFIC SITUATIONS

There are instances where neither of the above situations accounts for the development of secondary salinization. One occurs worldwide, and not only in semi-arid or arid environments. In coastal (including estuarine) areas, overpumping of fresh groundwater (sometimes for irrigation, often for urban water supply) results in seawater incursions into the groundwater aquifer (Goudie, 2000). The problem has been reported from Israel, California, Bahrain, the United Arab Emirates, the Nile Delta (due to the Aswan Dam rather than direct pumping of groundwater) and in more humid environments, from London and Liverpool in England, Germany, the Netherlands, Japan, the eastern seaboard of the US, and Perth (Western Australia).

A second situation concerns the Aral Sea in the Commonwealth of Independent States (CIS), previously the USSR (Ghassemi *et al.*, 1995) (*see* Aral Sea, Volume 4). In the 1950s, the Amu Darya and Syr Darya (rivers) were diverted from the Aral Sea to develop irrigated lands. Those irrigation areas produced 90% of the former USSR s cotton output and 40% of its rice. But the result was to cause severe shrinking of the Aral Sea. Between 1960 and 1989 its level dropped 12.9 m and its area decreased by 40%, and the problem has continued since then. The effects on the ecology of the Sea and on the fishing communities which depended on it have been disastrous. In terms of salinity, the exposed lakebed has been stripped by strong winds and the highly saline sediments deposited over an extensive area. The chemistry of the salts (and probably other toxic ingredients) appears to be such as to have caused a range of serious illnesses in people living in the affected area, a problem which is still the subject of research.

FAILURE TO IMPLEMENT SOLUTIONS

The situation of the Aral Sea in particular, but also the many other extensive problems of secondary salinity around the world, raise the obvious question: why do they persist? The problems have been known for a long time especially in the case of irrigation agriculture and the causes are well understood. Why, then, have apparently obvious preventative measures, let alone effective solutions, not been implemented, and why do the problems not only continue but become more severe?

There is an increasing feeling in certain quarters that there may not be solutions. In some instances, such as the Aral Sea, the areas irrigated by the Amu Darya and Syr Darya could be closed and the river flows redirected to the Sea. Social, economic and political factors work against this solution. In other instances, clearing of natural vegetation in areas potentially subject to secondary, dryland salinity could be prevented. Legislation to control clearing in parts of Australia has been introduced, but its enforcement leaves much to be desired again for social, economic and political reasons.

In other cases there may well not be a practicable, long-term solution. Irrigation needs relatively large amounts of fresh water to flush salts from the system. But water is in short supply in arid and semi-arid environments, the soils progressively lose their structure and become

increasingly impermeable, and the problems of waterlogging and salinization inevitably increase until the land has to be abandoned. Lining irrigation channels with impermeable materials and using drip rather than flood irrigation systems can slow the rate of salinization but will not prevent it.

Research in the Western Australian wheatbelt indicates that, in the catchments of saline areas, it is very difficult to plant crops or trees which will both transpire sufficient water and earn an income for the farmer. Nevertheless, extensive tree planting schemes are being undertaken which are the thrust of the state s salinity strategy (Government of WA, 1996) but the schemes are unlikely to solve the salinity problem and they will not return the land to its previous agricultural productivity. Relatively small saline seeps can and do respond to management, but the extensive, salinized valley floors which characterize much of the wheatbelt are a much more difficult proposition. It is being suggested that up to 30% of the wheatbelt may have to be abandoned. Agronomic and environmental efforts can be concentrated on the remaining 70% which may eventually provide greater total yields than are currently being obtained from the entire wheatbelt, and alternative commercial (or other) uses can be found for the (predicted) saline 30%. Unfortunately this still begs the question of the salinized water systems and the socio-economic consequences of the loss of infrastructure.

Deep drainage is another alternative which has been touted from time to time. But this is technically not a feasible proposition in most cases (due to the low gradients of the valleys and low hydraulic conductivities of the soil materials), it is very costly, and it would only exacerbate the problem of water salinity downstream.

A major problem now and in the past has been the lack of political will to take effective preventative or remedial actions. In Western Australia, reputable scientific, government researchers had demonstrated the nature and causes of the secondary salinity problem, and warned against further clearing for agriculture, in the 1920s and again in the 1930s. The warnings were not heeded, just as they are being paid insufficient attention in 2000 in Queensland especially, where clearing continues at breakneck speed. Decades of boards, committees and agreements in the Murray-Darling Basin (mainly involving New South Wales and Victoria), clearing bans in South Australia, and the more recent state salinity strategy in Western Australia, have failed to halt the inexorable spread of soil and water salinity problems across the nation. Unfortunately the farming electorate to which the politicians pay heed either do not understand or do not wish to understand the scientific messages.

Pannell (2000) considers that lack of awareness is not the main factor responsible for the slow or non-adoption of remedial measures. Rather, he argues, the major factors are related to the economic costs and benefits of current treatment options, the difficulties of testing the options, long time scales, externalities and social issues. A specific study by Kington and Smettem (2000) in one Western Australian wheatbelt catchment takes the argument further. They found that tree planting of 37% of lower slopes would need to be comprehensively adopted by all farmers; or that 60% or more of the upper catchment would need to be revegetated. However, existing catchment management policies (regulatory, cooperative and market-driven) were found to have inherent implementation problems and are inadequate for the purpose. The authors further suggest that perhaps no policy approach, even if fully implemented, would abate the salinity problem. They ask, therefore, whether it would not be more appropriate to develop policy strategies aimed at minimizing the inevitable social and environmental impacts and therefore extend current policy beyond its almost exclusive focus on dryland salinity management.

In contrast, in parts of North America the more tractable problem of secondary dryland salinity (in comparison to the Australian situation) has been treated successfully by judicious planting and management of *lucerne* (alfalfa). These plants are relatively deep rooted and high transpiring, and are therefore able to return the hydrological system back to its previous equilibrium. The main difficulties were to involve the farmers in the saline seep catchments in a cohesive and coordinated approach to the problem (Black *et al.*, 1981).

IMPLICATIONS OF CLIMATE CHANGE

There are two aspects of climate change which are relevant to the problem of secondary salinization; global warming, and changes in precipitation.

While global mean surface temperatures have increased in recent decades, the changes have not occurred uniformly around the world. In some parts of the globe the increase was as much as 4 °C whereas in a few areas temperatures actually declined (see map reproduced in Kerr, 2000). Continued warming of currently cold regions may make cropping possible in areas now unsuitable or marginal for agriculture, notably in the northern parts of the North American and the Eurasian continents, leading to the hydrological changes and resulting salinization which have been discussed here. On the other hand continued temperature increases in hotter areas, in parts of South America, Africa, Eastern Europe, South and Southeast Asia and Australia, may encourage efforts to increase irrigation, again with adverse consequences with regard to salinity.

Predicted changes in precipitation are even more uncertain and spatially and seasonally variable. With reference to Australia, e.g., winters in the southwest of the continent

are likely to be drier and summers wetter, as synoptic systems shift to the south. Increased summer rainfall in this Mediterranean climate type region will not benefit agriculture, but drier winters may conceivably result in less water accumulating in groundwaters with potentially beneficial consequences as far as the secondary salinity problem is concerned. On the other hand, increased summer rains in the agricultural areas of eastern Australia may well lead to expanded areas being cropped, with adverse implications for salinization.

A FINAL COMMENT

This article has discussed largely intractable problem of secondary salinity. It occurs throughout the world, usually in association with irrigation or dryland agriculture. But it needs to be recognized that salinity (and waterlogging) is only one, highly visible, manifestation of severely disrupted ecosystems. Other manifestations include loss of soil fertility, reduced soil structure, increasing soil acidity, increased erosion by water and wind, accumulation of chemical residues in soils and water, and loss of biodiversity and wildlife habitat. These problems result in decreased agricultural production and offsite environmental problems such as eutrophication of water bodies, the accumulation of chemical residues in food and species extinctions.

The ecosystems have been disrupted by agricultural practices. In a world with a burgeoning population and therefore increasing demands on increasingly scarce good land to produce more food and fiber, this is not good news. Yet in many places (e.g., around the Mediterranean Basin), agricultural and pastoral land has been abandoned. We have to learn to farm much more cleverly and more intensively on the small areas of good land, and maintain the world s biodiversity on the remaining, extensive areas with low actual and potential agricultural productivity. And we have to learn how to distribute food much more efficiently and equitably from those regions with surpluses (Europe, North America, Australasia) to those experiencing widespread famines and starvation, notably in Africa.

See also: Irrigation: Environmental Impacts, Volume 3.

REFERENCES

Black, A L, Brown, P L, Halvorsen, A D, and Siddoway, F H (1981) Dryland Cropping Strategies for Efficient Water Use to Control Saline Seeps in the Northern Great Plains, USA, *Agric. Water Manage.*, 4, 295 312.

Conacher, A J and Conacher, J L (2000) *Environmental Planning and Management in Australia*, Oxford University Press, Melbourne.

Conacher, A J and Sala, M, eds (1998) *Land Degradation in Mediterranean Environments of the World: Nature and Extent, Causes and Solutions*, John Wiley & Sons, Chichester.

Ghassemi, F, Jakeman, A J, and Nix, H A (1995) *Salinization of Land and Water Resources: Human Causes, Extent, Management and Case Studies*. University of New South Wales Press, Sydney.

Goudie, A (2000) *The Human Impact on the Natural Environment*, 5th edition, Blackwell, Oxford.

Government of WA (1996) *Western Australian Salinity Action Plan*, Agriculture WA, Department of Conservation and Land Management, Department of Environmental Protection, Water and Rivers Commission: a Report to the Minister for Lands and the Minister for the Environment, Government of Western Australia, Perth.

Hollaender, A, Aller, J C, Epstein, E, San Pietro, A, and Zaborsky, O R, eds (1979) *The Biosaline Concept: an Approach to the Utilization of Under-Exploited Resources*, Plenum Press, NY.

Kerr, R A (2000) Can the Kyoto Climate Treaty be Saved from Itself?, *Science*, 290, 920 921.

Kington, E A and Smettem, K R J (2000) Evaluation of Policy Approaches to Dryland Salinity Management in the Kent River Catchment, in *Land Degradation*, ed A J Conacher, Kluwer Academic, Dordrecht, 347 361.

Murray-Darling Basin Commission (1993) *The Murray-Darling Basin Commission - Managing Australia's Heartland*, Murray-Darling Basin Commission, Canberra.

Pannell, D J (2000) Explaining Non-Adoption of Practices to Prevent Dryland Salinity in Western Australia: Implications for Policy, in *Land degradation*, ed A J Conacher, Kluwer Academic, Dordrecht, 335 346.

Peck, A J (1978) Note on the Role of a Shallow Aquifer in Dryland Salinity, *Aust. J. Soil Res.*, 16, 237 240.

Tanji, K, Lauchli, A, and Meyer, J (1986) Selenium in the San Joaquin Valley: a Challenge to Western Irrigation, *Environment*, 28, 8 16.

Salinity, Primary

see Cereal Cultivation (Agriculture, Extensive) (**Volume 3**)

Salinity, Secondary

see Cereal Cultivation (Agriculture, Extensive) (**Volume 3**)

Salmon, Pacific Coast

see Fisheries: Pacific Coast Salmon (**Volume 3**)

Shelterbelts

see Cereal Cultivation (Agriculture, Extensive) (**Volume 3**)

Shifting Agriculture

see Swidden (**Volume 3**)

Shifting Cultivation and Land Degradation

Bryant Allen
The Australian National University, Canberra, Australia

Shifting cultivation is a form of agriculture where fields are rotated from place to place instead of crops being rotated around fields. During the rotation the majority of fields are not cultivated, but are left in fallow. During the fallow, natural plant successions develop on the field. In the archetypal shifting cultivation system, the fallow successions reach tall forest stages, before the trees are again felled and burned. However, shifting cultivation takes many forms. Although it is a very old form of agriculture, it is neither primitive, backward nor wasteful of resources. But because increasing pressure on shifting cultivation systems, by population growth and economic development, inevitably leads to the destruction of forests, shifting cultivators have been blamed for deforestation since Roman times. It is still routine to blame so-called slash and burn agriculture for contemporary deforestation, when the real causes lie deep in the competitive political and economic behavior that typifies all human societies.

SHIFTING CULTIVATION AS A SYSTEM OF AGRICULTURE

Shifting cultivation is a form of agriculture in which the cultivated or cropped area is shifted regularly to allow soil properties to recover under conditions of natural successional re-growth. In a shifting cultivation system, at any particular point in time, a minority of fields will be in cultivation and a majority will be in various stages of natural re-growth. Over time, fields are cultivated for a relatively short time, and allowed to recover, or are fallowed, for a relatively long time. Eventually, a previously cultivated field will be cleared of the natural vegetation and planted in crops again. Thus, fields in established and stable shifting cultivation systems are cultivated and fallowed cyclically.

Although fallow fields are not cropped, they are not unproductive. During the fallow period, shifting cultivators use the successional vegetation species for fencing and construction, firewood, thatching, ropes, clothing, tools, carrying devices and medicines. It is common for fruit and nut trees to be planted in fallow fields to such an extent that many are in fact orchards. Commonly, fallow re-growth species attract birds and animals and so indirectly increase the amount of food provided through hunting. Fallow tree species also draw nutrients to the surface from deeper in the soil profile. Particular species that have been shown to improve soil conditions are deliberately planted or protected from slashing or burning in fallows. Fallow shrubs and tree species protect the soil surface from direct sunshine and from the impacts of heavy tropical rainfall, either directly, or through the deep layers of leaf litter that build up over time. This litter and the fallow shrubs and trees themselves provide ecological niches for a myriad of parasitic plants, worms, insects, arachnids and small animals. Stable shifting cultivation systems significantly enhance and sustain biodiversity. This is in direct contradiction of popular views of shifting cultivation as destructive of forest resources.

Shifting cultivation becomes destructive only when systems become unstable. A critical relationship that governs the stability of shifting cultivation systems exists between the time for which the land is cultivated and the time for which it is fallowed. These parameters determine whether or not the shifting cultivation system as a whole suffers a net loss of nutrients over time. A system in which there is a net loss of nutrients with each cycle will eventually become unstable. Resources will be degraded unless actions are taken to arrest the losses.

The longer a field is cropped, the greater is the loss of soil organic matter, the reduction in the cation-exchange capacity and in nitrogen and phosphorus, and the greater the increase in acidity. Continued cropping leads to reductions in soil porosity and infiltration capacity and the loss of seeds of naturally occurring plant species from soil seed banks. In a stable shifting cultivation system, the fallow is long enough for the natural vegetation to recover to the state that it was in before it was cleared, and for the soil to recover to its pre-cropping condition. During fallow periods soil temperatures are lower, wind and water erosion is much reduced, nutrient cycling becomes closed again, nutrients are extracted from the subsoil, soil fauna increases, acidity is reduced, soil structure, texture and moisture characteristics improve and seed banks are replenished.

No optimum relationship exists between the length of the cropping period and the length of the fallow period. In

favorable agricultural environments, cropping periods can be longer and fallow periods shorter, than in less favorable agricultural environments. In favorable environments, soil conditions at the beginning of a cropping cycle will be better and fallow successional stages will proceed faster, than in less favorable environments. Nevertheless, even in the most favorable environments, it is likely that if the cropping period is extended beyond a certain point, the fallow conditions required for an adequate recovery of soils and vegetation will be jeopardized.

If the fallow period is shortened, there will be less time in which the soil recovery processes and vegetation successions can take place. The length of fallow period required to prevent net loss of nutrients will again depend on the quality of the environment, which will in turn, determine the rate at which recovery occurs. But sooner or later, if the fallow period continues to be reduced, an observable change will occur in the fallow vegetation. Secondary forest may be reduced to shorter, thinner stemmed, fewer, woody bush or jungle species, bush may be reduced to scrub and tall grasses which may be in turn reduced to short grasses. Less directly observable, but nevertheless critical, changes will also occur in the soil.

Changes in environmental conditions that happen subsequent to either a lengthening of the cropping period or a shortening of the fallow period often result in a fall in crop yields. It is not difficult to perceive how a shifting cultivation system, once destabilized, can proceed into a vicious circle of declining yields and in order to maintain overall production, a further shortening of fallows or lengthening of cropping periods, which in turn leads to further degradation of environmental conditions. This process, its causes and possible solutions are discussed further below.

Shifting cultivation is also known as *swidden agriculture* (*see* Swidden, Volume 3), an old European term, and, in some cases, is wrongly termed bush fallowing, or slash and burn. Stable shifting cultivation systems can exist where the fallow successions do not proceed beyond grasses. There are shifting cultivation systems where the slashed vegetation is never burned. In some systems, planting occurs beneath the standing trees and as the crops mature, the trees are felled on top of the crops to let in more sunlight. Shifting cultivation systems can be irrigated. Shifting cultivators may mulch the soil surface of fields or create mounds in which green manure is placed as compost. On slopes it is common for barriers to be constructed along the contour to reduce soil erosion. Crops may be mixed together or planted separately in different parts of the same field or in completely separate fields. A complex cycle of shorter and longer fallows may follow the cultivation of different crops. It is not uncommon for farmers to employ shifting cultivation systems in conjunction with sedentary (or non-shifting) systems, often using one system in one environment and the other in another.

The secondary forests created by shifting cultivation are commonly much richer in plant and animal resources useful to humans than primary forests, even though they may be less biodiverse. Shifting cultivators view the forest as an agricultural landscape of fields at various stages in a regular cycle. People not accustomed to living in forests cannot see the fields for the trees. Rather they perceive an apparently chaotic landscape in which trees are cut and burned randomly and so they characterize shifting cultivation as ephemeral, pre-agricultural, or as primitive and as a stage to be progressed beyond. Shifting agriculture is none of these things.

Stable shifting cultivation systems are highly variable, closely adapted to micro-environments and are carefully managed by farmers during both the cropping and fallow stages. Shifting cultivators possess a highly developed knowledge and understanding of their local environments and of the crops and wild plant species that they exploit. Complex and highly adaptive land tenure systems exist under shifting cultivation. Furthermore, introduced crops for food and for sale have been skillfully integrated into shifting cultivation systems.

PRIMITIVE, BACKWARD, WASTEFUL, UNPRODUCTIVE?

Shifting cultivation is negatively stereotyped. Shifting cultivation systems are perceived both by the supposedly well informed as well as the general public, as primitive, backward, wasteful, unproductive, exploitative and the cause of widespread environmental degradation. Shifting cultivators are blamed for the destruction of much of the world s tropical forests, and thus for land degradation, atmospheric pollution and global climatic change.

Contemporary manifestations of these attitudes towards shifting cultivators are often political and reflect competition between land-occupying farmers, migrant settlers, loggers and international capital for unfettered access to tropical forests (see next section). However, the attitudes can be traced back to the late 19th and early 20th century, to European colonial administrations in tropical and subtropical South Asia, Southeast Asia and South America. Colonizers characterized the indigenous occupants of the colonies and their cultures as primitive and hence viewed their agricultural systems, which were commonly a form of shifting cultivation, also as primitive. In addition, there developed a fundamental misunderstanding that shifting cultivators were nomadic wanderers who selected sites for cropping at random in previously unused forest and that they destroyed the forest by slashing and burning.

These attitudes prevented otherwise astute observers of tropical environments from understanding the social and economic conditions that explain why shifting cultivation systems become unstable and hence destructive of

environments. It is ironic that a major pioneering work published in the 1950s and renowned for its demonstration of the extent of human impacts upon the Earth contains a chapter on agriculture in the tropics that states No matter where we go we find primitive agriculture was carried on at the expense of forest and makes a distinction between primitive tropical horticulture and agriculture as practiced by higher cultures (Bartlett, 1956). This chapter implies all shifting cultivation systems are destructive of resources. Yet in the same book another chapter describes the massive loss of the forests in Europe prior to the beginning of the 20th century and details the social and economic conditions, other than shifting cultivation, which brought about their destruction (Darby, 1956). Many of the conditions responsible for the deforestation of Europe are similar to those causing deforestation today. It is worthwhile therefore commenting upon what happened in Europe.

SHIFTING CULTIVATION IN EUROPE

Shifting cultivation was still practiced as a viable and stable form of agriculture in many parts of Europe and west into Siberia at the end of the 19th century and in some places well into the 20th century. In the Ruhr in the late 1860s a forest-field rotation system, known as *Reutbergwirtschaft*, was using a 16-year cycle of clearing, cropping and fallowing with trees to produce bark for tanneries, wood for charcoal and rye for flour (Darby, 1956, 200). Swidden farming was practiced in Siberia at least until the 1930s, using specially selected varieties of swidden-rye (Steensberg, 1993, 98). In Eastern Europe and Northern Russia the main swidden crops were turnips, barley, flax, rye, wheat, oats, radishes and millet. Cropping periods were usually one year, but were extended to two or three years on very favorable soils. Fallow periods were between 20 and 40 years. In Finland in 1949, Steensberg (1993, 111) observed the clearing and burning of a 6 ha swidden 400 km north of Helsinki. Birch and pine trees had been cleared over a period of a year and the logs sold for cash. A fallow of alder (*Alnus*) was encouraged to improve soil conditions. After the burn, turnip was sown for sale and for cattle feed. Shifting cultivation was disappearing in this part of Finland because of a loss of agricultural labor to the industries of the towns. Steensberg (1993, 110 152) provides eye-witness descriptions of shifting cultivation being practiced in Sweden in the 1900s, and in Estonia, Poland, the Caucasus, Serbia, Bosnia, Hungary, Switzerland, Austria and Germany in the 1930s to the 1950s.

That these agricultural practices survived from the Neolithic into the middle of the 20th century amidst the sweeping changes that occurred in Europe over that period, suggests they were adaptive and in themselves, were not massively destructive of the environments in which they were practiced. This raises the question: if it was not shifting cultivation that led to the disappearance of European forests, what was it?

The earliest written accounts of forest destruction in Europe begin around 1000 BC in the histories of Homer, Thucydides and Plato and in Strabo s *Geography*. Forests in Southern Europe were exploited for ship building, and urban development, the manufacture of casks, pitch and charcoal, as well as being cleared for agriculture. An increased demand for ships, which were manufactured completely from forest products, followed upon an intensification of trade and warfare. Although goat herding is singled out as an important cause of environmental degradation, a more important cause of forest destruction was the practice in some places of granting ownership rights to those who clear-felled forests and brought the land into permanent cultivation. Evidence that circumstances other than agriculture were the major causes for forest destruction was the recovery of tree cover in many parts of the Roman empire from 400 BC to around 500 AD following the collapse of the Roman economy and industry. Darby observes that by 400 AD land that had once been tilled became derelict and overgrown and quotes Lactantius who wrote that in many places cultivated land became forest (Darby, 1956, 186). The other major causes of forest destruction in the hot dry summer Mediterranean environment were wild fires that became more common following human disturbance of the forests.

In Central and Northern Europe the use of stone tools and fire in agriculture is well established in the palynological and archaeological record from the Neolithic. Here, just as in Southern Europe, the destruction of the forest was not through shifting cultivation but the result of more intensive agriculture, following the invention of the plough and of the growth of trade, mining, smelting, tanning, building and urban construction. Nearly constant warfare, including the demands of naval shipbuilding, also contributed more to the destruction of the forests than shifting cultivation.

By the Middle Ages in Europe, large areas of forest were being cleared and converted into arable land in association with the development of feudal tenurial practices. From the 1500s to the 1700s the demands of iron smelters for charcoal, increasing industrial developments and the discovery and expansion of colonial empires as well as incessant warfare that increased the demand for shipping to levels never previously reached, all combined to deforest Europe. With the loss of the forest, shifting cultivation became restricted to the periphery of Europe, where permanent agriculture was uneconomic, transport costs constrained logging or terrain prevented the use of draught animals or tractors. It has disappeared from even these refuges since 1945, as agriculture has become increasingly capital intensive, rural areas have become depopulated and the remnant European forests themselves

have been revalued ecologically, economically and socially as watershed protection and places of recreation and tourism.

SIMPLE SOCIETIES, SHIFTING CULTIVATION AND ENVIRONMENTAL CHANGE

The forests of Europe were destroyed by the seemingly inexorable advances of civilization, industrialization and warfare, the sort of advances that many of those who criticized shifting cultivators in the 19th century thought were desirable and indicative of higher cultures. The same kinds of processes have continued to lead to the destruction of tropical forests in the last decade of the 20th century. So advances in civilization, now known as *development*, have not resolved these problems. The problems are located not in the practice of a particular form of agriculture, but within the fundamental relationships that human societies have with their environments. In complex developed economies, these relationships become very elaborate and are difficult to comprehend. However, in simple economies, where agriculture is the major source of wealth creation, they can be easier to understand.

A growing body of archaeological and palynological evidence indicates that simple human societies also brought about extensive changes to their environments before the establishment of any sort of state, feudal or capitalist, and before the development of large-scale mining, smelting or shipbuilding industries. In these societies, agriculture was the driving force in the economy and shifting cultivation was the most common type of agriculture practiced. By examining the relationships between social and economic change and the agricultural change in these societies, insights can be gained on contemporary social and economic change and global environmental change, and the place of shifting cultivation in these relationships.

Questions about relationships between the rise and fall of the Mayan civilization of the Yucatan Peninsula in equatorial America and shifting cultivation were raised as early as 1930 and continue to be debated today. Archaeological evidence suggests the development of Mayan society and economy began around 250 AD. A mere 700 years later it reached its apogee, by which time the population may have reached two million people. There followed a precipitous decline that left the great cities and ceremonial centers vacant and overgrown with jungle vegetation. The causes of this decline are uncertain; but warfare and the exhaustion of agricultural land are commonly cited (Meggers, 1954; Dumond, 1961; Turner, 1974).

Polynesian settlers appear to have followed similar paths in New Zealand and the Pacific Islands. Within 500 years of their arrival around 1100 AD, the Maori turned substantial areas of coastal New Zealand from forest into scrub and fern, in the process causing the elimination of many bird and animal species (Kirch and Hunt, 1997). In the restricted environments of the Pacific islands, including Fiji and Hawaii, clear evidence of early extensive erosion and change of vegetation is presumed to have been caused by shifting cultivation on slopes. Soils washed from slopes were deposited in valley bottoms as a rich, swampy alluvium. The newly created environments were then exploited to develop intensive, irrigated fields for taro (*Colocasia*). The change from shifting cultivation to intensive irrigated fields occurred in association with a rapid growth in population and the development of elaborate and highly socially stratified chiefdoms (Kirch, 1984). In the larger, temperate latitude islands of New Zealand, the presumed course of events took a different path. There the stimulus for population growth was the hunting of large birds to extinction, during which time forests in drier areas were destroyed by burning. This was followed by the development of intensive agriculture in favorable environments, based mainly on sweet potato (*Ipomoea batatas*) and a reliance on the gathering of two main wild plant species (bracken fern and *Cordyline*) in less favorable environments. These changes, as in the smaller islands, were accompanied by population growth, the competition for the occupation of the best environments, increasing complexity in social organization and endemic warfare (Anderson, 1997).

The record of human-induced changes in environments is longer in New Guinea than in most places, agricultural activities probably beginning 5000 years ago, and possibly as long ago as 9000 years. However, the most spectacular changes, in both societies and environments, are believed to have occurred in the central highlands of the island within the last 1000 years, in association with the introduction of a crop new to New Guinea, the sweet potato (Golson, 1982a,b). One of the most striking signals of the relatively recent intensification of agriculture is the sudden increase in sedimentation rates in small lakes.

The question posed by these and the numerous other examples that could be cited of simple societies that have intensified their agricultural systems in association with increases in population, social complexity and warfare is not whether or how shifting cultivation was responsible for the extensive changes to landscapes and environments. Rather it is why simple societies of shifting cultivators in the tropical forest of Yucatan, or the highlands of New Guinea, began to grow in numbers, to develop stratified and sometimes complex social hierarchies and to fight each other with increased frequency and ferocity.

At first sight, the greatest stimulus to the intensification of a shifting cultivation system is a growth in population. If no other changes occur within the system, for each extra person to be fed from agriculture, a small extra amount of land must be cultivated. The total amount of land available is the land being presently cropped and all of the land in fallow. If the area occupied by the system

is not expanded into previously unused land, or cannot be expanded because it is already occupied by neighbors, then either the cropping period must be extended or the fallow period shortened. The environmental consequences of changing these relationships have been explored above.

At least two problems exist with the population growth hypothesis. First, population growth rate in most preindustrial shifting cultivator societies has been shown to be very low over the long term. Second, no human societies are known where people work only to eat. People engage in social relations with each other; and agricultural produce, together with other commodities, is used in the conduct of these relationships. Studies of these relationships are the focus of two attempts to understand the nexus between human societies and their environments, one an explanation of a particular situation and the other a general exploration of the problem.

In a study of the Duna in the Southern Highlands of Papua New Guinea, a group in the process of moving from shifting cultivation into permanent field agriculture postsweet potato, Modjeska (1982) argued for the development of two self-amplifying feedback loops of ecological and social causation. The trigger was very slow population growth and the slow expansion of agriculture to meet the demands of this growth. This set in motion the first feedback loop, the use-value loop. As more forest was cleared, there was a decline in wild food resources and protein produced from hunting, which was substituted for by an increase in domestic pig raising. An increase in domestic pigs required a further expansion in agriculture. The greater protein available from the larger number of pigs increased human fertility and survival rates and resulted in faster population growth.

Increasing numbers of people also set in motion the second or exchange-value loop of social causalities. More people meant greater numbers of human interactions, including increased opportunities for conflict. There arose the need for a means to mediate these relationships. The Duna (and other highlanders in New Guinea), says Modjeska, substituted pigs for humans and began to exchange pigs to compensate for losses of humans killed in warfare, or in marriage or from deaths from natural causes. The demand for pigs thus increased and the production of pigs increased to meet the demand, with all of the consequences observed in the use-value loop. This is what Brookfield (1972) calls social production in his distinction of the importance of production for social reasons over production for human nutritional requirements alone. Increased pig production and the exchange of pigs to mediate relations between individuals and groups created opportunities for leadership and management of resources and some men gained authority over other men and all women. As groups became more complex, competition between men and between groups increased, increasing the opportunities for conflict.

The outcome of the operation of the two loops, one bringing about ecological change and the other social and economic change, is an expanding and intensifying agricultural system. Forest is converted to grassland, the population grows at an increasing rate and expands geographically and the society increases in complexity and stratification.

Another explanation of the relationships between simple agricultural societies and their environments is that of Ellen (1982, 252 270). Ellen does not attempt to separate use-values from social production. He argues that almost all of the materials required by humans to live (with perhaps the exception of air) are obtained through social relations of production and that these relations proliferate and are modified in numerous ways. The values that humans attribute to items produced from the environment arise out of cultural arrangements and not from the objects themselves, a restatement of Karl Sauer s dictum that resources are cultural appraisals . Humans frequently translate actual objects into culturally conceived forms, an example being the translation by the Duna of the pig into an item of compensation and redemption. As a result, two fundamental processes underlie the ecology of human social systems. The first is that of obtaining material items from the environment with subsequent alteration and circulation through social relations, and the second is the giving of material values to the items which will affect how important it is to obtain, circulate or alter them. Environmental pressures are thus mediated through social relations.

Transitions in ecological systems and in social systems do not proceed at the same rate. The rate of evolution of new varieties of plants and animals proceeds slowly and is determined mainly by natural selection but also partly by human interference and adaptation, as in the domestication of a wild species. Social evolution however, proceeds rapidly because humans have the ability to learn and to communicate their knowledge to each other and to future generations. Most social systems have the tendency to increase in complexity and sooner or later they come into conflict with, or into contradiction (Friedman, 1979, 1982) with their environments. What happens around the point of contradiction will determine the extent of the environmental degradation that will occur. Of particular importance is the ability of a society to change, to invent or to innovate technologically and sociologically, in order to overcome the contradiction without incurring continuing environmental degradation, or social disintegration.

An economic study of what occurs at the points of conflict with specific reference to shifting cultivation is that of Ester Boserup (1965). Boserup argues that low intensity farming, extensive shifting cultivation for example, has lower labor costs than more intensive farming systems. She argues that given a choice, a human group will always choose the technique which has the lowest absolute labor cost rather than the highest yield. This assertion is

questionable, but Boserup s model nevertheless retains its interest. At the point of conflict, yields will have become unsatisfactory. Boserup argues, contra Malthus, that rather than population always overwhelming resources, humans will invent a new agricultural technique or adopt an existing innovation that will boost yields. The new or adopted technology is adapted to the new environmental conditions created by the degradation that has occurred already, even though the society will pay for the increases in the form of higher labor costs. Examples of such changes are the adoption of new higher yielding crops, the exchanging of a digging stick for a hoe, or a hoe for a plough, or the development of irrigation systems. The controversy over Boserup s proposal is in part over whether intensive systems are more costly in labor terms, and whether humans bring about change in their agricultural systems before environmental degradation forces them to do so (*see* Technological Society and its Relation to Global Environmental Change, Volume 5).

A number of very important things happen in the passage from simple to more complex societies and agricultural systems (Ellen, 1982, 272 273). External demands for cash crops on a local production system may destabilize the local ability to regulate human environment relations. Parts of the agricultural system may become more specialized, species diversity may be reduced or lost, wild plant and animal resources reduced or lost and ecosystems become more fragile. Improved communications result in a higher rate of innovation and hence a greater rate of change. Higher rates of change and increased differentiation in the society give rise to increased conflict. Increased differentiation also leads to larger numbers of individuals not producing anything, and more being produced by fewer, such that the system as a whole becomes less efficient. Greater organization and specialization results in greater complexity, technical division of labor and a greater codification of cultural responses with more extensive social control. All of these things increasingly isolate large parts of society from the environment on which their consumptive demands are impacting. Hence responses to changes in environments that have a direct and immediate impact on humans in simple societies, become indirect, heavily mediated and muted in complex societies.

SHIFTING CULTIVATION IN THE CONTEMPORARY WORLD AND GLOBAL ENVIRONMENTAL CHANGE

No human societies remain isolated from the world economy and the forces of globalization. All humans are to a lesser or greater extent, part of one enormous and highly complex society. It is to be expected, that shifting cultivation systems will be influenced by global economic and social forces and that the responses of shifting cultivators to changes in their agricultural systems will be similarly influenced. Yet shifting cultivators continue to be blamed for environmental outcomes as if they were isolated in time and space from the rest of the world.

The estimated rate of deforestation in Southeast Asia in 1990 was 3.4 million ha year^{-1} (FAO, 1990, quoted in Potter, 1993). In Indonesia alone it was estimated that 1.31 million ha year^{-1} were being lost, 368 thousand ha year^{-1} from Sumatra and 377 thousand from Kalimantan, of which 144 thousand ha were due to the fires of 1982 1983. Since those estimates were made, huge fires ravaged Indonesian forests during the 1997 98 El Nino-associated drought.

Shifting cultivation was assessed by the Food and Agriculture Organization (FAO) to be one of the causes of this deforestation while logging was not. The apparent discrimination against shifting cultivators caused a confrontation between FAO and environmental groups, who saw the FAO supporting commercial logging interests against the rights of indigenous people (Potter, 1993, 108). Other independent studies of the problem note that despite lack of government control over forests and the dominance of a political elite in the logging industry, the causes of deforestation are highly complex. It is true that the logging industry has driven roads into the forests, built logging camps and caused severe environmental damage with heavy machinery. The industry is replete with instances of corruption at the highest level that results in an almost complete lack of control over where logging is practiced and how it is practiced. The loggers have also provided paid employment to former subsistence farmers. One of the outcomes of cash incomes and better access to the outside world has been rapid population growth among indigenous groups of former shifting cultivators placing pressure on their traditional long fallow farming systems.

Many farmers have taken advantage of the improved road access to urban markets by planting cash crops, such as rubber or pepper. Increased cash incomes have been spent on chain saws, which have enabled larger areas to be cleared for cultivation faster. Fallow periods have been radically reduced and cropping periods extended. Serious poverty elsewhere in the country has brought thousands of land-hungry settlers into the cut-over forests along the logging roads. The settlers practice what appears to be shifting cultivation but which is in fact a one-cycle slash, burn, continuous cropping, with no intention to fallow for a long time. Clearing of trees and the permanent cultivation of fragile soils in a tropical environment with little attempt to replace lost nutrients causes rapid degradation of the environment on a large scale. The loss of forest in Indonesia, Thailand and the Philippines during the 1990s was preceded by major ecosystem disruptions in Vietnam, Laos and Cambodia in the 1970s and 1980s as a result of warfare. Forests were sprayed with defoliants, thousands of

rural forest-dwelling people uprooted from their homes and roads driven into previously isolated areas.

The loss of the tropical forests of Southeast Asia is the particular outcome of the general possible outcomes described by Ellen (see previous) when small local ecological and social systems become part of a larger system. When the previous relatively stable ecological relationships are destabilized, degradation can occur rapidly. Similar forest losses and fragile ecosystem destruction are occurring in the Amazon Basin, by large-scale state-sponsored colonization of forest land (Becker, 1995, 61) and in Central Africa where endemic armed conflict is destabilizing rural settlement and farming communities on a massive scale.

CONCLUSION

In the tropical developing world, shifting cultivation in its many diverse forms remains an important practice. Shifting cultivation was one of the very first forms of agriculture practiced by humans and its survival into the modern world suggests that it is a flexible and highly adaptive means of production. However, it is also a seriously misunderstood practice. Many casual observers cannot see past the clearing and burning of standing forest and do not perceive often ecologically stable cycles of cropping and fallowing. Nevertheless, shifting cultivation systems are particularly susceptible to rapid increases in population and to economic and social change in the larger world around them. The blame for the destruction of forest resources is often laid on shifting cultivators. However, the forces causing the rapid tropical forest loss at the end of the 20th century are those that led to the destruction of the forests of Europe at the beginning of the millenium. They are essentially related to urbanization and industrialization but are all too often driven by the application of the latest technology to extract ever more resources from the environment in pursuit of political and military power by competing groups. Shifting cultivators then, as now, are swept along in changes that are largely not of their own making.

Studies of small, isolated and pre-capitalist groups and their relationships with their environments suggest that the roots of the contemporary problem lie deep in human behavior. However, it has been shown that simple societies can adapt to the changes that come from within, population increase, competition and conflict, and which impact on the environment. But they are less able to adapt to external economic, political and social forces beyond their control. Globalization is pushing shifting cultivators all over the world into direct contact with world economic, social and political forces. The outcome is often severe environmental degradation. Global climatic changes may further increase stresses upon shifting cultivation systems.

Shifting cultivation systems are extremely diverse. Wide variations are found in land use, animal husbandry, crop germplasm and successional re-growth species, as well as in the societies themselves. Continued exploitation of shifting cultivators and their forest resources will result in continued rapid degradation of the environment, loss of species and loss of human knowledge. If the biodiversity, agrodiversity and the knowledge of local environments that is held within shifting cultivation societies is to be conserved, the people who practice shifting cultivation must be the subject of political and economic equity at national and international levels.

REFERENCES

Anderson, A (1997) Prehistoric Polynesian Impact on the New Zealand Environment: Te Whenua Hou, in *Historical Ecology in the Paci c Islands: Prehistoric Environmental and Landscape Change*, eds P V Kirch and T L Hunt, Yale University Press, New Haven, 271 283.

Bartlett, H H (1956) Fire, Primitive Agriculture, and Grazing in the Tropics, in *Man's Role in Changing the Face of the Earth*, ed W L Thomas, University of Chicago Press, Chicago, 692 720.

Becker, B K (1995) Undoing Myth: the Amazon, an Urbanized Forest, in *Brazilian Perspectives on Sustainable Development of the Amazon Region*, Vol. 15, eds M Clusener-Godt and I Sachs, Man and the Biosphere Series, UNESCO, Paris, 53 89.

Boserup, E (1965) *The Conditions of Agricultural Growth: the Economics of Agrarian Change Under Population Pressure*, George Allen and Unwin, London.

Brookfield, H C (1972) Intensification and Disintensification in Pacific Agriculture: a Theoretical Approach, *Pac. Viewpoint*, 13(1), 30 48.

Darby, H C (1956) The Clearing of the Woodland of Europe, in *Man's Role in Changing the Face of the Earth*, ed W L Thomas, University of Chicago Press, Chicago, 183 216.

Dumond, D E (1961) Swidden Agriculture and the Rise of Maya Civilization, *Southwest. J. Anthropol.*, 17, 301 316.

Ellen, R F (1982) *Environment, Subsistence and System: the Ecology of Small-Scale Social Formations*, Cambridge University Press, Cambridge.

Friedman, J (1979) Hegelian Ecology: Between Rousseau and the World Spirit, in *Social and Ecological Systems*, eds P C Burnham and R F Ellen, Academic Press, London, 253 269.

Friedman, J (1982) Catastrophe and Continuity in Social Evolution, in *Theory and Explanation in Archaeology: the Southampton Conference*, eds C Renfrew, M J Rowlands, and B A Segraves, Academic Press, New York, 175 176.

Golson, J (1982a) The Ipomoean Revolution Revisited: Society and the Sweet Potato in the Upper Wahgi Valley, in *Inequality in New Guinea Highlands Societies*, ed A Strathern, Cambridge University Press, Cambridge, 109 136.

Golson, J (1982b) Kuk and the History of Agriculture in the New Guinea Highlands, in *Melanesia: Beyond Diversity*, eds R J May and H Nelson, Australian National University, Canberra, 297 307.

Kirch, P V (1984) *The Evolution of the Polynesian Chiefdoms*, Cambridge University Press, Cambridge.

Kirch, P V and Hunt, T L (1997) *Historical Ecology in the Paci c Islands: Prehistoric Environmental Change and Landscape Change*, Yale University Press, New Haven.

Meggers, B J (1954) Environmental Limitations on the Development of Culture, *Am. Anthropol.*, 56(5), 801 824.

Modjeska, C N (1977) *Production Among the Duna: Aspects of Agricultural Intensi cation in Central New Guinea*, PhD thesis, Australian National University, Canberra.

Potter, L (1993) The Onslaught on the Forests in South-East Asia, in *South-East Asia's Environmental Future: the Search for Sustainability*, eds H C Brookfield and Y Byron, United Nations University Press and Oxford University Press, Kuala Lumpur.

Steensberg, A (1993) *Fire-clearance Husbandry: Traditional Techniques Throughout the World*, Poul Kristensen, Herning.

Turner, B L (1974) Prehistoric Intensive Agriculture in the Mayan Lowlands, *Science*, 185(4146), 118 124.

FURTHER READING

Brookfield, H C and Byron, Y (1993) *South-East Asia's Environmental Future: the Search for Sustainability*, United Nations University Press and Oxford University Press, Kuala Lumpur.

FAO (1990) *Interim Report on Forest Resources Assessment 1990 Project, with Reference to Asia and the Paci c Region Forest resources Assessment 1990*, Food and Agriculture Organization, Rome.

Humphries, S (1993) The Intensification of Traditional Agriculture Among Yucatec Maya Farmers: Facing up to the Dilemma of Livelihood Sustainability, *Hum. Ecol.*, 21(1), 82 102.

Rao, Y (1989) Forest resources of tropical Asia, in *Environment and Agriculture - Environmental Problems Affecting Agriculture in the Asia-Paci c Region*, World Food Day Symposium, Food and Agriculture Organization, Bangkok, 1 20.

Thomas, W L (1956) *Man's Role in Changing the Face of the Earth*, University of Chicago Press, Chicago.

Shipping: Harbors and Ports

Brian Hoyle
University of Southampton, Southampton, UK

The global maritime transport system and the world cityport system that it serves both re ect and in uence the physical environments within which they are set. Transport is a central dimension of the national and global production systems that are reshaping the world. Global economic integration relies upon ef cient maritime transport, and interlinkages between global environmental change, ports and shipping re ect policy and technology as well as socio-economic conditions.

Harbors, ports and the cities with which they are associated provide essential mechanisms and contexts for economic and social development and change. The evolving world cityport system depends upon many factors, including the biophysical environment. Relationships between port systems and environmental circumstances are never static, but change as a consequence of multidirectional in uences and effects. In both advanced and developing countries, technological skills increasingly facilitate ef - cient port operation but the development of port-industrial systems is often accompanied by adverse effects on the environment. Global environmental factors affect port systems everywhere. Speci c events such as oil spillages can have serious consequences on the local environment of affected ports and harbors.

Ports and shipping perform critical roles in the development process. While expanding port systems generally bene t regional and national economies, they can have negative effects on local or regional environments. Urban redevelopment in older port areas no longer used by shipping presents challenges and opportunities that can yield substantial environmental improvements. Any global sea-level rise induced by climatic change will adversely affect low-lying coastal zones worldwide and many port-city systems will face substantial problems.

TRANSPORT AND THE ENVIRONMENT

Relationships between transport and the environment are extremely complex, whether considered from scientific perspectives or in political terms. Concern about the environmental consequences of transport development and operation is not new, and transport takes a central position in environmental debates. The development of land-based transport modes roads, railways, canals and airports involves many potentially damaging impacts on land and landscapes as well as many forms of noise, water and air pollution (Hunter *et al.*, 1998; Hayashi *et al.*, 1999). Throughout the world, maritime transport and port development produce environmental problems, effects and controversies. The changing technology of maritime transport, the scale and character of modern port operation and development, the increasing globalization of world trade, and their associated patterns of economic development and urban growth all produce planning debates in which environmental issues loom large. In both advanced and less-developed countries, debates linking policy, technology, ecosystem resilience and rates of environmental and economic change influence decision-making processes.

Maritime and coastal conditions currents, tides, winds, water depths, natural harbors provide permissive or restrictive environmental frameworks which facilitate or constrain the operation of the global maritime transport system. Widespread concern for the possible impacts of the system upon the environment, in terms of land and water use (involving consumption, congestion, pollution and competition) affects the functioning and development of the ships, ports and trade relations which make up this transport system (Goudie, 1993). Maritime disasters involving oil spillage, the fouling of beaches and coasts, and the loss of wildlife and habitats, often highlighted in the media, sometimes reflect inappropriate juxtaposition of oil terminals and conservation areas (as at Milford Haven, UK). Port development, promoted in the context of changing socio-economic structures, invariably affects delicate coastal ecosystems; the instability of both ecosystems and local human society is influenced by changes in maritime technology. While modern technology means that environmental constraints on maritime transport development are less than in past times, concern for the environment has increasingly influenced the operation and expansion of maritime transport systems. Nevertheless, contrary to what is often supposed, the development of modern ports can have significant beneficial effects on local environments.

PORTS, HARBORS AND THE CITY/PORT INTERFACE

Every port needs a good harbor a sheltered area of water within which the port can develop. Natural or artificial, the harbor is the environmental and engineering context within which the economic, trade and transport roles of the port occur. Relationships between the harbor, in its natural or improved state, and the development of port facilities, are very variable. Within its harbor, a port is fundamentally a central place of economic and cultural interchange, and specifically a place where the mode of transportation changes from land to water-borne systems. Ports function as transfer points in the flow of ideas, goods, services and biological elements of change. Ports are also points through which exotic species may spread: the role of shipping in the ecology of the black rat, for example, is significant.

Ports originate, grow or decline in response to demands from their customers; they serve as nodes in multimodal transport systems (Hayuth, 1985; Slack, 1998), and may become important industrial centers, sources of employment and major urban complexes. The term *cityport* is sometimes used to describe multifunctional settlements of this kind. Not all ports are located on the shore of seas and

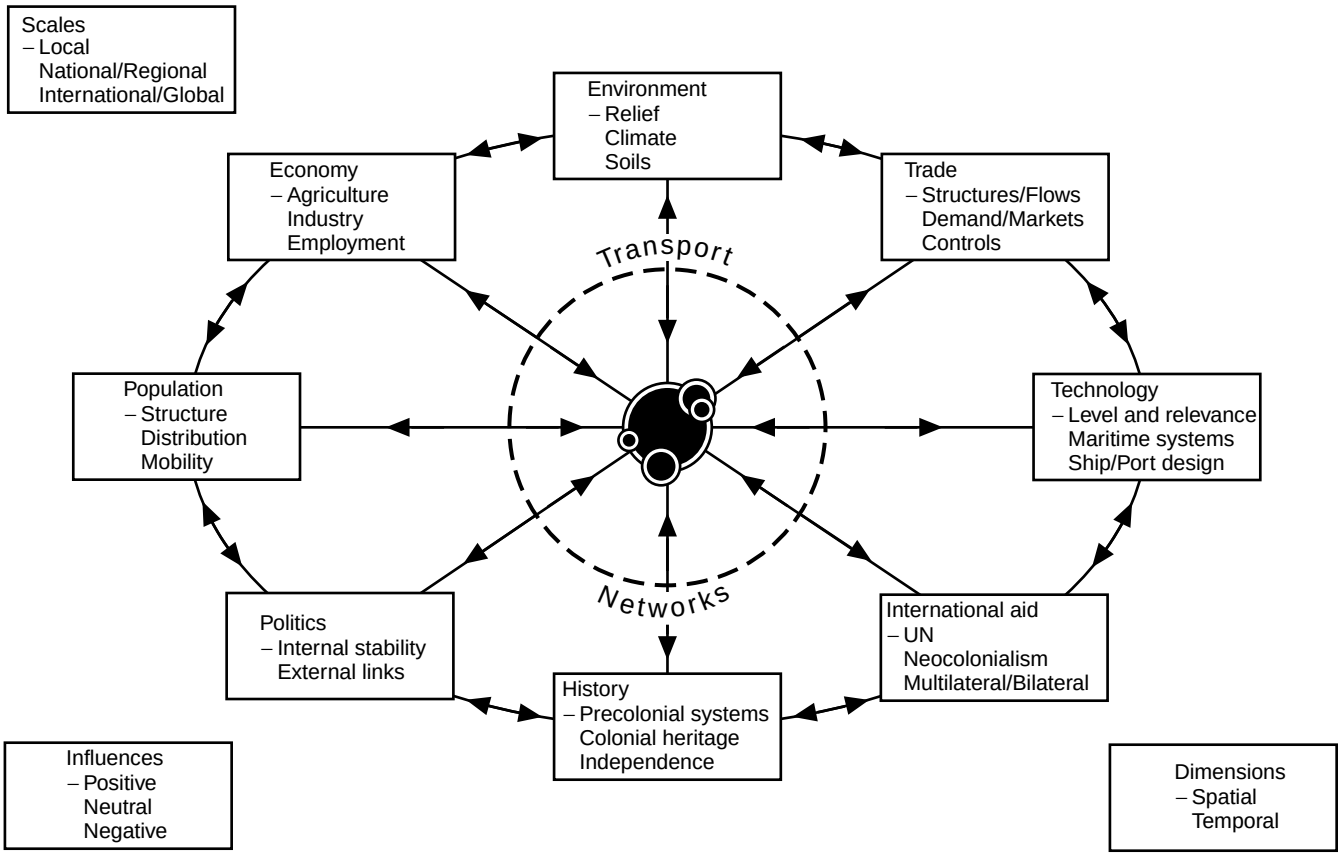


Figure 1 The world cityport system. (The first version of this diagram was published in Hoyle and Pinder, 1981)

oceans: inland ports on rivers (Montreal, Canada; Manaus, Brazil) and lakes (Chicago, USA; Kisumu, Kenya), sometimes a very long way from the open sea, are major elements in national and global maritime transport networks.

Throughout human history the essential problems of port development have been: how to adapt port activity and growth to the physical environment; how to respond to technological innovation, especially to changes in ship design; how to survive inter-port competition; and how to serve local needs, on the waterfront and in the city, as well as to reflect national and global conditions. At various scales and in both spatial and temporal terms, a wide range of interdependent factors is involved in the growth of transport networks and cityport systems (Figure 1). Whether the viewpoint is local or global, each factor (including the environment) may influence others, or the system as a whole, in positive, neutral or negative ways.

Shipping and maritime trade normally underpin the origins and early growth of coastal cities, although links between port functions and urban activities are often less intimate today than in the past. The growth of ports is basically affected by four factors: the land situation, the water situation, the land site and the water site. The water site (i.e., the physical conditions of the harbor) often provides the initial stimulus to development, but unless conditions in the other three categories are favorable, the settlement is unlikely to prosper. The level of port development, measured in terms of cargo-handling capacity or the volume and value of commodity flow, is chiefly a reflection of the land situation the extent and pattern of economic development including transport infrastructures in the hinterland and the demand for port services thus produced. The land site provides a more direct control over urban growth patterns and the operation of landside port functions; and the water situation affects relationships with major trans-oceanic shipping routes. The prosperity of a port ultimately depends, however, upon the efficiency with which it is able to respond to demand from its customers and on the degree to which it is able to maintain and enhance its competitive position. Adaptability to environmental, technological, socio-political and economic change is clearly a critical dimension of port policy.

Within these four inter-related contexts, environmental influences are critical directly and indirectly, and port development ultimately depends on the assessment of situation and site factors in an environmental context. In changing port-city interrelationships, environmental controls may be seen as a filter (Figure 2), intended to reduce pollution and harmonize development. In this port-city interface model, which expresses the inability of cityport sites to absorb not only rapidly-changing and expanding port development but also successive phases of urban

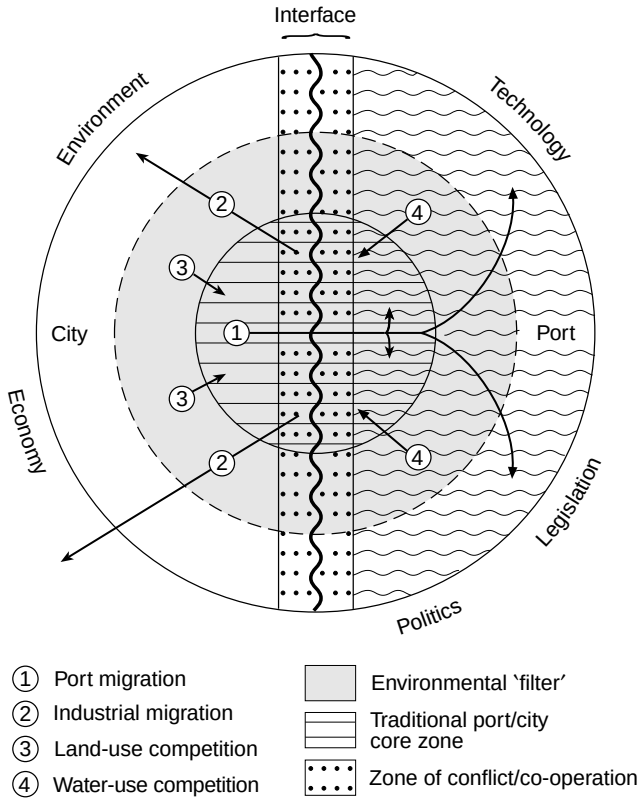


Figure 2 The port-city interface model. (Reproduced from Hoyle *et al.*, 1988)

growth, urban land uses are separated from maritime activities by a zone of conflict or cooperation characterized by decay or by regeneration. Port development migrates downstream in search of bluecoast sites with deeper water and more land/water space, while port-based industries migrate to greenfield sites. Competition for the use of the most attractive available waterfront locations comes from developments such as housing, offices and marinas. These processes involve negative and positive changes to nearby ecosystems and the built environment (AIVP, 1994). They are affected by and also influence the way local people and non-governmental organizations (NGOs) perceive environmental conditions and changes (Cau, 1996).

PORT DEVELOPMENT IN ADVANCED COUNTRIES

Faced with finding the most efficient and cost-effective response to customers demands, port authorities see themselves as driven by inter-port competition and constrained by acknowledged environmental issues and responsibilities. In advanced countries, port-city industrialization has developed in a transport-dependent context that has involved complex relationships with, and

impacts on, the environment. Cityport industrialization, with all its risks and opportunities, has been traditionally based on the juxtaposition of maritime transport the cheapest form of transport for bulk goods and the opportunity to develop processing or manufacturing activities at the break-of-bulk point where the mode of transportation changes. Such transformation activities based on flows of import or export goods have long provided port-city employment, encouraged urbanization and created or exacerbated environmental problems including land, water and air pollution.

This system changed radically after 1960 as growth in demand, production and transport led to a revolution in maritime technology. The widespread use of standard-sized containers for rapidly increasing transport of goods by sea and land led to multimodal and intermodal transport systems involving origin-to-destination movements from an interior location point in one continent via two or more ports and a sea voyage to an inland destination in another continent that reduce or obliterate the traditional port s break-bulk processing and transformation function. However, the evolution of the global economy has also involved vastly increased flows of oil, ores and other bulk commodities, and many of these shipments originate in developing countries and terminate in advanced industrialized countries (e.g., crude oil from the Middle East and West Africa to Europe, and iron ores from West Africa to Japan and the USA). Links between suppliers and customers in such cases are made technically possible by the development of very large crude carriers (VLCCs) for oil and very large ore carriers, and by the simultaneous development of very large port facilities where such carriers can be received and where it often makes sense to process goods before onward movement.

As ships get larger, the differentiation between large hub ports and smaller feeder ports increases (*see* Transport: Global Freight and Passenger Flows, Volume 3). Demand for oil and steel in the 1960s and 1970s meant that previously unexploited crude oil reserves (e.g., in Indonesia, Nigeria and the North Sea) and iron ore resources (e.g., in Quebec, Western Australia and West Africa) were developed. The invention of ULCCs (ultra-large crude carriers) for oil in the 1970s and 1980s facilitated flows but created problems for ports, thus accelerating port concentration. The global pattern of trade and transport in oil and oil products produces a pattern of oil slicks within which there are long-term abuses which can be reduced by good practice of the diluting capacity of the oceans. Accidents involving vessels of these types, pollution emergencies, can have devastating environmental impacts on coasts, beaches and wildlife. The loss of the *Torrey Canyon* off Cornwall in 1967 (*see* Torrey Canyon, Volume 5) led to agreement by international oil companies to pay compensation for massive coastal pollution. The true long-term impacts of maritime and coastal oil spills are unknown. Some oiled birds may recover, but oiled mangroves and coral may never do so (Newson, 1995). The restructuring of global oil production, refining and transport industries from the 1980s has, however, yielded some beneficial effects in certain coastal zones (Harcombe and Pinder, 1996).

From the 1960s, the response of some port authorities, and of many regional and national governments, to the increasing scale of maritime goods transport and associated port facilities was to develop maritime industrial development areas (MlDAs) which capitalized on economies of scale arising from bulk handling and from the geographical juxtaposition of heavy industry (oil refining, steelworks) and the maritime transport system used to import the raw materials for these industries. Traditional ports cannot cope with this scale of activity, because MIDAs require vast amounts of land and water space for storage and transforming industries as well as for the reception of vessels and the loading and unloading of their cargoes.

Today the basic economic and logistic arguments for MIDA-style developments in coastal zones are no longer accepted at face value (they never were in the UK). There is a much stronger awareness of coastal environments and the threats posed to them by industrialization. Planning controls have been greatly strengthened in European countries. In Japan it is now increasingly difficult to develop new port-industrial areas (except on artificial land, as at Kobe); and Japan invests in heavy industry in the countries producing required raw materials. MIDAs now incorporate the importation of semi-finished goods, the conservation of coastal zones, the diversification of industrial activity, and residential and other non-industrial urban functions. The MIDA idea has led over several decades to less purely economic, more socially acceptable and more environmentally aware views and policies concerning how port cities should grow, how coastal zones should be conserved, and how the relationships between ports, cities and industries should be managed.

Venice, Italy

Two European examples, Venice (Italy) and Southampton (UK), illustrate the environmental issues associated with port-industrial growth and the substantial increases in scale involved. Venice (Figure 3) exemplifies Italy s concentration of industrial development in coastal zones, with the expansion of port facilities from the Marittima zone opened in the western part of the city in the 1880s and the creation of the industrial zone of Porto Marghera on the mainland during the early decades of the 20th century (Soriani, 1996).

Port-industrial development in the Venice area, especially during the post-World War II period, has involved environmental impacts of great significance. Industrial waste

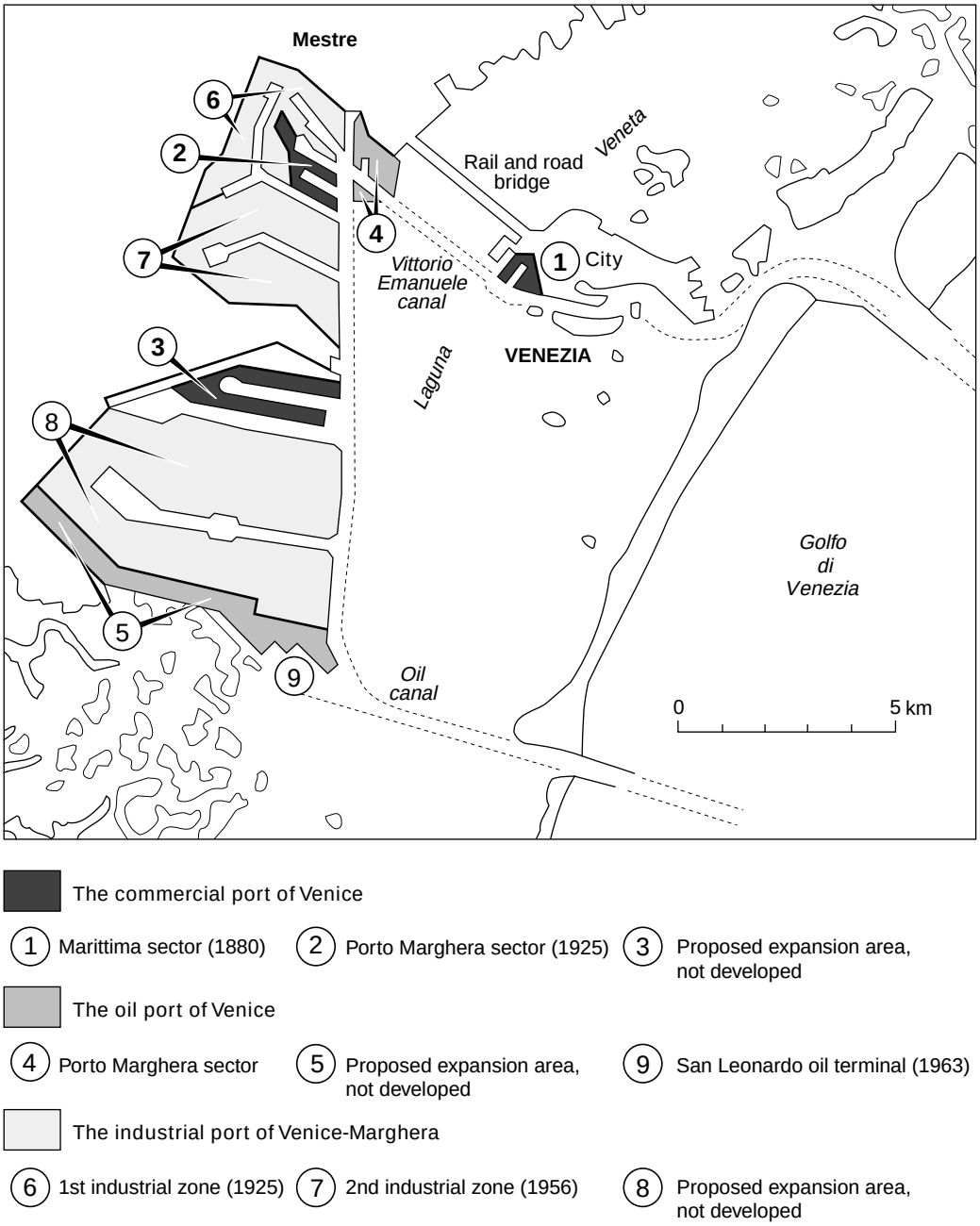


Figure 3 Port and industrial development in the Venice lagoon, Italy. (Reproduced from Hoyle, 1996)

has strongly contributed to the modification of the lagoon s ecosystem. Rapid industrial growth at Porto Marghera after 1950 involved chemical, steel and oil-refining plants, which discharged wastes directly into the lagoon, inducing deterioration of water and sediment quality. Although strictly regulated from the 1980s, serious damage is still caused by the release of pollutants of industrial origin. Dredging operations, especially those designed in the 1960s to facilitate the access of increasingly large vessels, seriously affected the morphology of the lagoon and its hydrological regimes. The introduction of coastal barriers to protect the entrances to the lagoon, and the excavation in the 1960s of an oil canal designed to allow large tankers to enter the port without passing the city s historic core, also increased the influence of marine forces on the dynamics of the lagoon. In addition, the abstraction of ground water for industrial purposes appears to have contributed to the slight but significant gradual reduction in the height of Venice above sea level.

On 4th November 1966 Venice was flooded as a consequence of exceptionally high tides, a strong southeasterly gale and inadequate coastal protection. This dramatic

event may have marked the end of port-industrial expansion within the Venice commune. Public concern led to widespread action aimed at the protection and conservation of the traditional historic core of the city. A United Nations Educational, Scientific and Cultural Organization (UNESCO) report published in 1969 stressed the demographic characteristics of the city, inhabited by an increasingly ageing population and by the accelerating migration to the mainland, and questioned the continuing integrity of Venice in the context of the Porto Marghera developments. New research centres were established during the early 1970s to study the environmental problems of the area. Private, mostly foreign, organisations took steps to finance the restoration of fabric of the historic city. The Italian Government promulgated a 6th Special Law for Venice in 1973 outlining policies for the lagoon area as a whole. The idea of the third industrial zone was abandoned, and part of the already reclaimed land was flooded again.

The question of the protection of Venice has not yet been solved. Many investments have been made to reduce polluting emissions and waste in recent times. Nevertheless, the relationships between the environment and port-industrial development is still a thorny problem, both in the context of navigation and in relation to industrial accidents. Venice continues to be flooded by high tides far more often than previously. Issues related to the protection of Venice, in socio-political and environmental terms, now fundamentally condition the future of the Porto Marghera port-industrial complex. The 1973 special law for Venice provided, *inter alia*, for the diversion of oil traffic away from the lagoon. By the late-1990s this law had not been put into effect. Oil traffic represents about 45% of total port traffic; revenues from oil movements underpin the financial viability of the port and provide the raw material for the refining and petrochemical industries in the port-industrial area. Projects aimed at connecting Porto Marghera industries by pipelines with other oil terminals on the northern Adriatic coast have been proposed. However, a lack of clear signals from the national government and the expectation of further job losses have combined to postpone a solution to this problem. The result is that the underlying controversy arising from conflicting factors involved in coastal zone management in this area is often rekindled, thus preventing continuity in the planning process.

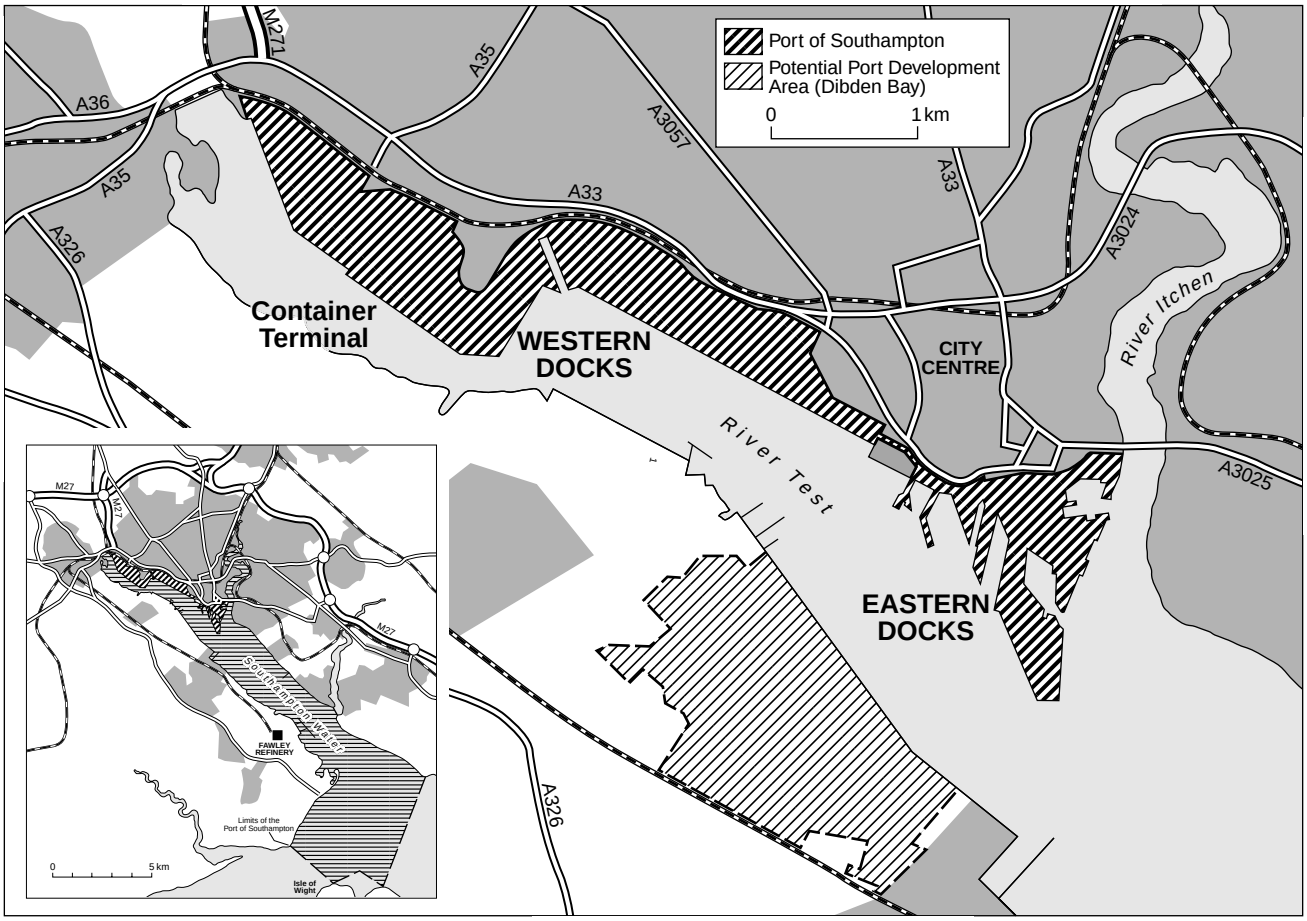


Figure 4 The port of Southampton, UK. (This map is an updated version of one published in Hoyle, 1986)

Southampton, UK

Located on a peninsula at the confluence of two small rivers, the port of Southampton (Figure 4) was in the early 20th century a major transatlantic passenger port and gateway to the British overseas empire. Today it is an important commercial gateway orientated towards the European continent and the Far East, with worldwide container trades. The environmental advantages of the site include extensive, sheltered deep water and, in earlier times, a reasonably secure defensive location. As a result of complex tidal conditions in the English Channel, there is high water for long periods, important for ship maneuvers.

Disadvantages include the distance from the open waters of the Channel (in comparison with Portsmouth), restrictions on modern urban transport circulation imposed by the shape of the peninsula, and a lack of natural land sites suitable for modern port facilities. All Southampton s 20th-century port facilities have been developed on artificial reclaimed land (*see* Land Reclamation from Seas, Volume 3). The Eastern Docks, which extended the natural peninsula, have been partly redeveloped for recreational, commercial and residential functions, and for the University of Southampton s research center for deep-sea oceanography and marine geology. Further development upstream, beyond the modern container port, is very unlikely; but reclaimed land nearby was set aside in 1967 at Dibden Bay, a brownfield site created from port dredging between the 1940s and the 1970s. Pressure on existing facilities means that the development of new container berths on this site is being promoted. Preliminary consultations and studies have addressed, in particular, the environmental and ecological implications of creating new berths (with associated road/rail links and river channel deepening) in a development zone close to residential areas and to the New Forest. The needs of wildlife and plants as well as people are central to the port authority s proposals, submitted to the UK government as a basis for a planning enquiry and for eventual landscaping and construction. The port authority and the government are anxious to balance the needs of the port, the local communities and the environment. Despite some local opposition, port expansion is widely recognized as essential if Southampton is to maintain and enhance its global competitive position as a maritime transport hub.

PORTS AND PORT CITIES IN LESS-DEVELOPED COUNTRIES

Ports and port cities have played a large part in the development of the less-developed countries of the modern world, often as gateway settlements from the standpoint of a colonizing power and as windows on a wider world for the societies and economies of coastal and interior zones (Hilling, 1996). Imperial port cities, in Australia and North America as well as, for example, in India and Africa, were critical nodes in the establishment of both transport networks and urban systems, helping to develop dependent territories within a globalizing economy. These key elements normally provided foundation stones of continuing importance in the context of present-day urban and transport systems.

Environmental conditions in less-developed countries are often inimical to transport network development as a whole and to port growth in particular. Good natural harbors are infrequent and often inconveniently located whereas modern ports are often needed where site conditions are unpropitious. While this is generally the case, it is especially so in tropical Africa. Dangerous surf on the Atlantic coasts of West Africa, and coral growth along the Indian Ocean shores of East Africa, are not only problems for navigators and travelers but also for port engineers, and they increase the costs of port development. Good natural harbors such as the drowned river valleys at Mombasa (Kenya), the lagoons at Lagos (Nigeria) and the sheltered bay at Dakar (Sénégal) have encouraged port and urban development on a large scale; but the land sites of some good harbors such as Freetown (Sierra Leone) are unsuited to large-scale city growth and the land situation of others such as Mtwara (Tanzania) are badly located in relation to the economic and transport geography of their hinterlands. Numerous artificial harbors have been created, as at Tema (Ghana) and Cotonou (Benin), to provide sheltered deep-water accommodation for shipping in association with economic development where the coastal environment does not provide adequate shelter for shipping.

Ports provide one set of dynamic instruments through which the external relations of developing countries are conducted and through which their environments are continually threatened, modified or enhanced. Examples from West and East Africa illustrate these principles.

Abidjan, Côte d'Ivoire

The Atlantic shores of the West African republic of Cote d Ivoire are generally low and sandy, completely lacking in shelter, but backed by a series of deep-water lagoons. The capital port city of Abidjan has been shaped both by its lagoon-based physical environment and by liberal economic policies that have led towards active participation in international maritime trade. Created in 1950 by the opening of the artificial Vridi Canal linking the Gulf of Guinea with the Ebrié Lagoon, and aligned to avoid siltation from the sea the port experienced rapid growth in traffic and facilities until the early 1980s, and the city has added numerous industrial and service functions to its commercial role. Modern Abidjan is a remarkable example of adaptation to a very difficult physical site and of African port-city growth within an economy responding to local, national and regional needs.

The environmental cost of these economic and social achievements has been the severe pollution of the waters of the lagoons on which cityport development has occurred. To the east of the cityport, the Ebrié lagoon was formerly linked directly to the estuary of the Comoé River and the open sea, and tides contributed to the regular renewal of the lagoon waters. This link has been closed as the urban road network has intensified, and the Vridi Canal is inadequate to ensure a normal circulation of the lagoon waters. Pollution has resulted from natural causes and from urban waste. The proliferation of parasitic water plants (notably the water hyacinth) reached serious, and seemingly uncontrollable, levels in many parts of the lagoon by the mid-1990s. The treatment of urban waste is limited, bacteriological pollution is high, and the port-industrial zones of the cityport complex contribute substantial quantities of hydrocarbons to the water bodies. Solutions to these problems require the reopening of the Comoé estuary, stricter control on port-industrial effluents and more efficient urban waste management. Rapid cityport development has produced increasing environmental pollution, and a more socially responsible policy of waste management and environmental control is required as urban populations and port-industrial activities continue to increase.

Mombasa, Kenya

A different kind of adaptation to problematical physical site conditions is illustrated by the Indian Ocean port of Mombasa (Kenya) (Figure 5). A series of drowned river valleys or *rias* resulting from Pleistocene changes in relative sea level provide the essential geomorphological

and geographical context for port development at Mombasa, as elsewhere along the East African coast for example at Dar es Salaam (Tanzania). Mombasa is unusual in that the port has two harbors: the island on which the town center and many of the port facilities are located lies between two arms of a complex double ria, and each of these two harbors has been developed for commercial purposes in quite different ways. Mombasa Old Harbor, rather narrow and relatively shallow, was well suited to medieval navigation and now provides shelter for smaller craft, while Kilindini Harbor (the place of deep water) provides modern facilities. The significance of Mombasa as a seaport in both medieval and modern times is largely due to the geographical juxtaposition on this site of these two harbors contrasted in area, depth and capacity.

These conditions have enabled Mombasa, unlike many other ports, to adapt successfully to functional and navigational changes over a long period of time, and this cityport is now overwhelmingly the dominant port, urban and industrial complex on the Kenya coast. Environmental conditions have generally favored the growth of the port and town and have rarely proved a constraint. The development of Mombasa as a port over the centuries has been conditioned not by environmental change but by the perceptions of successive generations concerning the opportunities provided by the specific environmental characteristics of the island and its harbors for the development of trade and settlement.

Today, although urban pollution (especially on the coastal fringes of the Old Town) is a growing problem, water pollution is not a serious issue as the harbors are open to tidal fluctuations and the movements of vessels (especially oil tankers) are carefully controlled. Improvements in operational efficiency, management restructuring and the revitalization of the port in the competitive context of the East African port system as a whole are critically important current policy objectives. Environmental conditions in terms of land and water sites do not present any insuperable obstacles to physical port expansion in the foreseeable future. Additional deep-water berths and container terminals could be developed at Kipevu or on the southern shore of Port Reitz where a maritime industrial development zone has been proposed (Hoyle, 1999).

WATERFRONT REDEVELOPMENT: THE TRANSFORMATION OF THE PORT–CITY INTERFACE

An aspect of port-city change that has received much attention in recent years around the world is the phenomenon of urban waterfront redevelopment. While this is not, directly, a result of global environmental change in a physical sense, it nevertheless expresses a worldwide trend in the reshaping of cities on water and in the refashioning of the human environments that ultimately depend on the conservation and use of water bodies. Although waterfront redevelopment has spread to innumerable cities of many kinds, it is nevertheless particularly associated with port cities (Hoyle *et al.*, 1988).

The main reason for recent changes and problems in the port/city interface zone, and ultimately for the global waterfront redevelopment movement, is that traditional cityport sites cannot absorb rapidly changing and expanding port development and successive phases of urban growth. Cities and ports which once were intimately connected and finely balanced have grown apart, and both ports and cities have found ways of adjusting to their new relationship. In many instances this process stems essentially from the fact that many cityport land and water sites were environmentally ideally suited to the relatively limited forms of urban development and port activity which characterized, for example, ancient or medieval times the examples of Venice and Southampton illustrate this point but such sites are totally unsuitable for modern ports and cities which require far more extensive sites for their development where factors affecting ports and those affecting cities are less closely associated.

In Europe, waterfront redevelopment is regarded as essentially maritime in character, linked to port cities, while in North America it is seen more as a process of urban renewal, as many settlements with a waterfront introduce redevelopment in a context that has little or nothing to do with port activity. Often a major issue is how to get rid of the railways and highways that skirt the shoreline, replacing them with accessible green spaces. The spatial diffusion of waterfront redevelopment started in North America in the 1960s (in Boston, Baltimore and San Francisco) and spread to European port cities such as London in the 1970s and 1980s and then to Australia and Japan (Figure 6). In the 1990s it became important in newly industrializing and developing countries too for example in SE Asia, in the Caribbean, and in South and East Africa. Diversity is obviously a keynote of this diffusion, but common elements include the maintenance or provision of public access to the water s edge; opposition to (and prevention of) insensitive and inappropriate development in waterfront zones; enhanced rapprochement rather than continuing or increased separation between waterfront zones and urban cores; the conservation and sensitive development of open green space; the cultivation of difference and of a sense of community; and the maintenance of links with the past.

An emphasis on greening the port-city waterfront environment is often a key to successful integrated redevelopment in this context. As the movement spreads to less-developed countries in a context of modernization, urbanization and globalization, clear links are perceived between port-city urban renewal and other socio-economic sectors: water supply, housing, employment, tourism. Costs

are high, progress is slow and socio-economic progress is often associated with environmental deterioration.

IMPACTS OF RISING SEA LEVEL

An aspect of global environmental change that directly affects harbors and ports in the longer run is the prospect of substantial sea-level rises as a result of continued global warming. If the global warming now being detected were eventually to involve substantial sea-level rise, all the world s harbors and ports would be affected. Some ports might find deeper water alongside existing facilities an advantage; others would find their operation restricted and their functions curtailed. It is more likely, however, that low-lying coastal zones, including some urban areas in port cities, would experience serious flooding. If, however, some lowering of sea levels were eventually to occur (unlikely in the next million years), the consequences for ports would be far more serious, as many would find themselves with insufficient deep water and, *in extremis*, at some distance from the sea. The examples of Rye (UK) and Bruges (Belgium) show how coastal evolution can effectively reduce a port s traffic over a relatively short time period, although in such cases soil erosion and local sedimentation may explain more than will global sea-level change.

CONCLUSION

Links between global environmental change and global maritime transport systems are expressed at one level very specifically in terms of water quality, sea-level change, and the impact of port-industrial and urban development on air, land and water pollution. In a much wider sense, however, the changing global environment is clearly a major factor in the dynamic evolution and operation of international maritime transport systems and port-city development. Environmental changes, and world shipping, are two interdependent dynamic systems that influence one another over longer or shorter time periods, at many different levels of socio-economic development, and in association with a wide variety of other factors and systems.

The redevelopment of port-city waterfronts is a feature of late 20th-century urban environments. As the new millennium unfolds, and as new centers of global innovation and change emerge and strengthen in the Pacific rim, the forces of globalization, intermodal transport systems and the communications revolution mean that ports and port cities may lose some of their traditional significance as the need to adapt and transform intensifies. The relative success of port-city change and development will depend in part on integration between communities and localities, so that appropriate development based on environmental conservation and social objectives can at least gain ground against (if not ultimately supersede) commercial development driven only by the profit motive. This demands politico-economic restructuring, a sense of scale, an appreciation of interdependence, and a reassessment of environmental priorities.

See also: Transport: Global Freight and Passenger Flows, Volume 3; Transport Infrastructure, Volume 3.

REFERENCES

Association Internationale Villes et Ports (1994) *Villes Portuaires, Acteurs de l'Environnement*, 4th International Conference, Montreal, 1993, AIVP, Le Havre.

Cau, L (1996) Environmental Perception and Planning: the Case of Plymouth s Waterfront, in *City Ports, Coastal Zones and Regional Change: International Perspectives on Planning and Management*, ed B S Hoyle, John Wiley & Sons, Chichester, 61 82.

Goudie, A S (1993) *The Human Impact*, 4th edition, Blackwell, Oxford.

Harcombe, S and Pinder, D A (1996) Oil Industry Restructuring and its Environmental Consequences in the Coastal Zone, in *Cityports, Coastal Zones and Regional Change: International Perspectives on Planning and Management*, ed B S Hoyle, John Wiley & Sons, Chichester, 83 102.

Hayashi, Y, Button, K, and Nijkamp, P, eds (1999) *The Environment and Transport*, Edward Elgar, Cheltenham.

Hayuth, Y (1985) Containerization, Inter-Port Competition and Port Selection, *Marit. Policy Manage.*, 12, 293 303.

Hilling, D (1996) *Transport and Developing Countries*, Routledge, London.

Hoyle, B S (1986) The Transformation of the Port of Southampton, in *Ports et Mers: Mélanges Maritimistes offerts à André Vigarié*, ed J J Charlier, Caen, Paradigme.

Hoyle, B S, ed (1996), *Cityports, Coastal Zones and Regional Change: International Perspectives on Planning and Management*, John Wiley & Sons, Chichester.

Hoyle, B S (1999) Port Concentration, Inter-Port Competition and Revitalization: the Case of Mombasa, Kenya, *Marit. Policy Manage.*, 26, 161 174.

Hoyle, B S (2000) Global and Local Change on the Port City Waterfront, *Geog. Rev.*, 90(3), 395 417.

Hoyle, B S and Pinder, D A, eds (1981) *Cityport Industrialization and Regional Development: Spatial Analysis and Planning Strategies*, Pergamon Press, Oxford.

Hoyle, B S, Pinder, D A, and Husain, M S, eds (1988) *Revitalising the Waterfront: International Dimensions of Dockland Redevelopment*, Belhaven, London.

Hunter, C, Farrington, J, and Walton, W (1998) Transport and the Environment, in *Modern Transport Geography*, eds B S Hoyle and R D Knowles, John Wiley & Sons, Chichester, 97 114.

Newson, M (1995) The Earth as Output: Pollution, in *Geographies of Global Change: Remapping the World in the Late Twentieth Century*, eds R J Johnston, P J Taylor, and M J Watts, Blackwell, Oxford, 333 353.

Slack, B (1998) Intermodal Transportation, in *Modern Transport Geography*, eds B S Hoyle and R D Knowles, John Wiley & Sons, Chichester, 263 289.

Soriani, S (1996) The Venice Port and Industrial Area in a Context of Regional Change, in *Cityports, Coastal Zones and Regional Change: International Perspectives on Planning and Management*, ed B S Hoyle, John Wiley & Sons, Chichester, 235 248.

FURTHER READING

Goldrick, M and Merrens, R (1996) Toronto: Searching for a New Environmental Paradigm, in *City, Capital and Water*, ed P Malone, Routledge, London, 219 239.

Slash and Burn Agriculture

see Swidden (**Volume 3**)

Soft Energy Path

Soft energy path is an energy future that combines significant increases in energy efficiency with the development of relatively environmentally benign, and renewable, energy sources, including solar power, wind power, geothermal energy, biomass and small-scale hydroelectric generation (*see* Biomass Burning, Volume 3; Wave Power Development, Volume 3; Wind Power Development, Volume 3; Environmental Politics, Volume 5).

ROBERT PAEHLKE *Canada*

Soil Amelioration

J Keith Syers
Naresuan University, Phitsanulok, Thailand

To ameliorate a soil means to enhance its productive capacity by improving soil properties, which may be inherently poor or have been degraded by Man's activities.

Improving the productive capacity of soils by soil amelioration has a long history. For example, practices such as drainage (to remove excess surface and subsurface water), marling (to improve the mechanical composition and nutrient status of sandy soils), and the addition of animal manure and human excreta (to provide soil organic matter and nutrients) have been used widely.

Soil degradation, accelerated by Man's activities, has increased the need for ameliorative practices. These include nutrient input from a range of sources, but particularly from chemical fertilizers, to combat nutrient depletion; the addition of lime to correct soil acidity, which has increased in many countries because of industrial emissions and the use of legumes, which acidify the soil; and drainage to combat salinity in conjunction with the use of by-product gypsum ($CaSO_4$) (when available and affordable) to reverse sodication. Our fundamental understanding of nutrient depletion, acidi cation, and salinization (salinity and sodication) is good, but in the case of sodication (process whereby exchangeable sodium is increased) and to some extent nutrient depletion, there is a need to develop effective ameliorative management practices which can be used at the farm level.

Bioremediation of soils contaminated with toxic organic pollutants (e.g., hydrocarbons and halocarbons) involves biodegradation brought about by enzymes. Bioavailability, including the physical accessibility of the organic molecules to the microorganism producing the enzyme, is critical. Molecular techniques are being used with a view to accelerating the biodegradation of more resistant organic molecules.

HISTORICAL CONTEXT

Soil amelioration has a long history, with a wide range of practices being used over the centuries to increase crop productivity.

Probably one of the oldest and the simplest practices of soil amelioration is that of land drainage whereby the water table is lowered, thus improving the oxygen status of the soil. This is vitally important for the growth and development of most plant species. Drainage systems range in complexity from simple ditches, to ridges and furrows that remove surface water, to clay tiles placed at depth to transport excess sub-surface water away, to the currently-used perforated plastic pipe, again placed at depth, often about 75 cm below the surface. Wet, heavy-textured soils are cold and difficult to cultivate. Only by installing field drainage systems can such soils be made productive.

Soil ameliorants, traditionally added to soils in historic time, include clay, lime and chalk, seaweed, animal manure, and human excreta. Initial use of these materials predated the development and use of chemical fertilizers. These materials are not fertilizers because they contain insufficient quantities of the three major plant nutrients nitrogen (N), phosphorus (P), and potassium (K). However, some of them contain useful quantities of plant nutrients and they were essentially the only source of nutrients up until the beginning of the 20th century.

Until around the middle of the 19th century in the UK, the amelioration of low-fertility, coarse-textured (sandy) soils was achieved by the practice of marling (for an account of the early practices of marling and chalking, see Hall, 1945). Marl contains calcium carbonate in association with clay, and when added to sandy soils, it substantially increases soil fertility and productive capacity. For example, in Norfolk (UK), the usual practice was to excavate pits down to the marl or clay underlying the surface sands, remove the marl, and spread it at the rate of 100 to 300 loads per hectare on the surface of sandy soils. Not only did the marl improve the texture but it added calcium carbonate (a liming material) and some plant nutrients (e.g., P and K). Chalking involved the addition of very large quantities of chalk (almost pure calcium carbonate). At Rothamsted Experimental Station (UK) up to 250 tons of chalk per hectare were added to experimental sites in the 18th and early 19th century (Johnston, 1991). The practices of marling and chalking were at their height during the early 19th century, but the rise in labor costs and the development of chemical fertilizers in the early 20th century resulted in their demise.

Before the advent of chemical fertilizers, the large quantities of animal manure returned to the soil were the major source of plant nutrients, in Europe and Asia in particular. Soil physical properties, such as soil structure and water-holding capacity, were improved through the addition of large quantities of decomposable organic material, which produced humus in the soil.

The addition of human excreta to the land is often regarded as being a Chinese technology, and it is sometimes forgotten that the transport of night soil from urban areas to rural areas for disposal on land was an important way of reversing the flow of nutrients and reclaiming areas of low-land peats, as in northern England. Public health regulations in many countries now restrict the use that can be made of human wastes in agriculture to sewage sludge, which has been digested to remove pathogens (Evans, 1998). Restrictions on the dumping of human wastes at sea and of uncontrolled incineration have increased the need for land disposal. This has led to legislation in several countries, particularly in Europe, to limit any adverse effects of the heavy metals present in sewage sludge. However, it seems likely that there will be further restrictions on the disposal of sewage sludge on land and controlled incineration may become a much more important outlet for disposal.

PRESENT SITUATION

At the present time, there is increasing recognition of the importance and impact of soil degradation on crop productivity and environmental quality, and thus the need for soil amelioration. This is particularly the case in the ecologically-fragile semiarid tropics, where the need for increased food production is great (Syers *et al.*, 1996). Soil degradation may be defined as a temporary or permanent lowering of a soil s productive capacity. It is an insidious process which results in a reduction in soil fertility and crop productivity. Although widely recognized as a global problem, the geographical distribution of soil degradation is not well understood.

The major types of soil degradation include wind and water erosion, which displaces soil material, and several *in situ* chemical degradation processes (including nutrient depletion and loss of organic matter, acidification, salinization, and pollution) and physical degradation processes (including compaction, crusting and sealing, and waterlogging) (*see* Waterlogging, Volume 3). It is in the context of soil degradation that soil amelioration has its greatest impact, with the technologies ranging from low-technology liming to high-technology bioremediation. Examples of some of these follow.

NUTRIENT INPUT

Nutrient depletion is a characteristic feature of agriculture in most developing countries. This occurs primarily through crop removal in harvested products, and crop residues that are not returned, and by leaching and soil erosion. Nutrients, in inorganic or organic forms, are added to soils to improve the nutrient-supplying power of low-fertility soils or to balance losses from productive farming systems, which are essentially restricted to the developed world. Resource-poor farmers are usually unable to provide the required nutrient inputs in the form of fertilizers and further productivity declines are inevitable. Recycling of low nutrient-containing crop residues in subsistence agriculture can only promote continuing poverty. Organic agriculture, without the input of essential nutrients from external sources, will not solve the problems of world hunger and poverty.

Chemical fertilizers have had an enormous impact on crop production in developed countries, to the point that surpluses of several agricultural commodities now exist, although several commentators consider that such agricultural surpluses are insufficient to protect the world in years of bad harvests. In lesser-developed countries, where the need for external nutrients is invariably greater, the supply of chemical fertilizers is sometimes erratic. Even where chemical fertilizers are in adequate supply, their price frequently restricts use to rates that are less than those required for optimum crop production. This situation is particularly serious in Sub-Saharan Africa where mining of soil nutrients is occurring because removals and losses are exceeding inputs. It is for this reason that the World Bank and its partners are interested in implementing a program of soil fertility recapitalization.

Phosphorus is now a limiting nutrient in many acid, weathered soils of the sub-humid and humid tropic (*see* Phosphorus Cycle, Volume 2; Phosphorus Water Quality Objectives, Volume 3). Amelioration of phosphorous deficiency can be achieved gradually with moderate, seasonal applications of phosphorous at rates which are sufficient to increase the available soil phosphorous levels. Alternatively, this can be achieved almost immediately with a large one-time application of P. An excellent example of amelioration of degraded soils is provided by the work in Southeast Asia of the Potash and Phosphate Institute (located in Singapore). In Southeast Asia, approximately 40 60 million hectares of natural forest have been removed by recent human activity, and following short-term cropping have reverted to shrub and grass savannah. The soils are acid and their productivity is limited by P deficiency. In Indonesia, a large ameliorative application of reactive phosphate rock of about 1000 kg ha^{-1} is sufficient to supply enough available P for annual and plantation crops for up to eight cropping seasons. The economics of this practice are very favorable (von Uexkull and Mutert, 1995), but the most critical issue is the financial outlay for the initial heavy application of phosphate rock. Because reactive phosphate rock materials have liming value (equivalent to about 30% by weight of the phosphate rock material), an added benefit from this strategy is the partial amelioration of soil acidity.

CORRECTION OF SOIL ACIDITY

Acidification is a natural process, which is enhanced by Man s activities. The continuous movement of weakly acidic rainwater through the soil leaches soluble basic cations, particularly calcium (Ca) and magnesium (Mg). Natural acidification has been accelerated by the use of legumes, which fix nitrogen and which generate ammonium-nitrogen; the heavy use of ammonium-containing fertilizers (and manure and slurry), the nitrogen from which can cause eutrophication of surface waters; and emissions of acidic sulfur dioxide and nitrogen oxides into the atmosphere and their return to the soil. It is only when the ammonia and ammonium salts are converted to nitrate (nitrified) by soil microorganisms and the nitrate leached that acidity (H^+) is developed. This may be shown as follows:

$$NH_4^+ + 2O_2 = NO_3^- + 2H^+ + H_2O \qquad (1)$$

The H^+ thus generated further accelerates the loss of basic cations.

Because Ca is the dominant cation in many agricultural soils, where soil pH is maintained in the range of 5.5 7.0, calcium or lime loss is usually given emphasis in the acidification process. Thus, the lime requirement to maintain soil pH, or return it to a desired level, has considerable practical significance. The value depends on soil pH and the capacity of the soil to buffer changes in soil pH, in addition to the amount and acidity of the rain. Work at Rothamsted Experimental Station (UK) indicates that basing the lime requirement on the need to neutralize soil acidity, rather than balancing Ca loss, appears to be a more correct approach (Goulding and Annis, 1998). Recent work in the UK indicates that lime requirements for agricultural land have doubled between the late 1960s and early 1970s, and the late 1980s and early 1990s, a fact attributed to increased fertilizer nitrogen use and atmospheric deposition.

As mentioned herein, in the section on the historical context, the use of lime and chalk has a long history in much of Western Europe. The benefits of lime are well established and include positive effects on the availability of certain plant nutrients (for example, phosphorus, potassium, sulfur and molybdenum). However, increasing soil pH by liming will decrease the availability of manganese and boron. Lime addition can also improve the physical properties of soils. Acid, degraded soils often have a poor structure, which restricts air and water movement, and frequently develop surface crusts (discussed herein), which reduce water infiltration and restrict germination. Liming can assist in preventing crusting in coarser-textured soils and can improve the friability of heavy-textured soils.

Whereas liming materials are usually accessible to farmers in the developed world, resource-poor farmers in less-developed countries invariably find the cost of lime prohibitive, even when it is available locally. The result is that very large areas of acid soils remain unproductive, for example in Southeast Asia (*see* Acid Sulfate Soils, Volume 3).

REVERSING SALINIZATION

The changes to soils involving the increase in salts is usually referred to as salinization (*see* Salinity and Agriculture, Volume 3). Salt-affected soils are often separated into saline soils, which contain excess salts (usually the chlorides (Cl) and sulfates (SO_4) of sodium (Na), Mg, and Ca in the soil solution, and sodic (or alkali) soils, which have elevated levels of Na on the exchange complex. Saline and sodic soils occur naturally and also as a result of irrigation, which redistributes salinity within the soil, and transport salts and or Na to new locations. Unlike the more extensive saline soils, sodic soils have been studied less intensively (Sumner, 1993). Salinization through the intrusion of sea water often occurs in coastal areas.

Swelling and dispersion are associated with the highly hydrated Na ion in sodic soils and are the primary reasons for the degradation of soil physical properties, including

low infiltration and poor conductivity of water through the soil, and crusting of the surface (discussed herein). Also, the high salt concentration in the soil solution of a saline soil makes it more difficult for plants to abstract water through their roots, resulting in moisture stress which can adversely affect crop growth and development.

Definitions of *salinity* and *sodicity* are comparatively easy but in reality it is impossible to separate the effects of salts (salinity) from those of sodicity (Na), as they often go together. Furthermore, the level of salt in a sodic soil is important in determining physical properties. This has implications for the development of management practices to reverse salinization.

In theory, removing salinity should be easy and should only require supplying sufficient water, over crop requirements, to leach out the soluble salts. This requires that the soil is permeable and that the water table, into which the salts are leached, is below the rooting zone of the crop or can be brought below the rooting zone by installing a suitable drainage system. Because a high water table is often also a cause of salinity, drainage is a prerequisite for ameliorating saline soils. Where groundwater is naturally high and the soil profile permeable, it may be necessary to install a pumping system, thus adding to the cost of reversing salinization. However, the recovered groundwater can be pumped into irrigation ditches and diluted with irrigation water, if the quality (i.e., salinity) permits.

In arid and semi-arid areas, rarely does the luxury of surplus water exist for the flushing out of salts and frequently farmers have to contend with salinity or even abandon the land.

The amelioration and management of sodic soils is less well understood. Salt-transport models provide valuable information on the chemistry and physics of sodicity amelioration. However, this has not been translated into validated management practices for application at the farm level. As with salinity, ensuring adequate drainage is the primary consideration during the reclamation of sodic soils. Replacement by calcium of some of the Na on the exchange complex is the secondary consideration. For soils containing gypsum, flooding a well-drained soil is usually sufficient but when the amount of gypsum is low or absent, soils containing appreciable amounts of exchangeable Na, or free sodium carbonate, will disperse and become impermeable if wetted by high-quality irrigation water or by rain. Maintaining a sufficiently high salt concentration during the removal of exchangeable Na is essential. The slow release of calcium from gypsum is sufficient to progressively remove exchangeable Na down the profile, provided drainage is adequate. By-product gypsum, as phosphogypsum from the phosphate fertilizer industry or flue gas desulphurization gypsum, produced in coal-fired power stations, is the most suitable because mined gypsum is too slowly soluble. Thus, gypsum has found fairly extensive use for the amelioration of sodic soil in several countries, including Australia, Israel, and the US.

PREVENTING CRUST FORMATION

Although dispersive soils usually contain high levels of Na, it is known that highly-weathered, acid soils disperse under conditions of low electrolyte (salt) concentration in the soil solution and with mechanical energy inputs, such as raindrop impact. Disintegration of aggregates at the surface, under raindrop impact and dispersion of the clay at low electrolyte concentrations is followed by movement of clay particles into the pore spaces, which become blocked, thus lowering the infiltration of water and increasing runoff and erosion. This is referred to as surface crusting (Sumner and Stewart, 1992). Movement of clay sometimes leaves loose sand on the surface. The surface crust usually consists of a compacted layer, approximately 1 mm in thickness, underlain by a clay-enriched layer.

The formation of a surface crust can be alleviated by reducing raindrop impact, by using a mulch; adding manure to stimulate biological activity; and/or by adding by-product gypsum to supply electrolyte at the soil surface.

BIOREMEDIATION OF CONTAMINATED SOILS

Soil pollution is a form of soil chemical degradation resulting from the addition of contaminants (or pollutants) to the soil. It is usually a feature of specific industrial sites and industrial regions. Examples include contamination from hydrocarbons (e.g., oil), halocarbons (e.g., chlorobiphenols), and heavy metals (e.g., lead) and radionuclides.

Bioremediation uses biological systems to accelerate the breakdown or transformation of chemical pollutants to less harmful forms. Although plants are increasingly being used to extract heavy metals and radionuclides from contaminated soils, by the process of phytoremediation, it is with microorganisms and organic pollutants that bioremediation finds its greatest potential application (Head, 1998). Here, the ability of microorganisms to biodegrade organic pollutants to non-toxic end-products determines the applicability of bioremediation. In some cases the end products are water and carbon dioxide.

The natural process of biodegradation is the underlying basis of bioremediation. This involves the action of specific enzymes, produced by particular groups of microorganisms, on specific pollutants or groups of pollutants. Not all compounds can be degraded in this way. For example, whereas low- to mid-molecular weight alcohols and hydrocarbons are readily biodegradable, many of the halocarbons tend to be resistant.

Several microorganisms produce dehalogenase enzymes which are effective in the process of dehalogenation, for

example, in the removal of chlorine. Similarly, certain microorganisms can degrade chlorophenols by the action of monooxygenase, which produces phenol and eliminates chlorine. The physical access of microorganisms to pollutants in situ is also important in influencing biodegradation. Bioavailability to microorganisms can be reduced by interaction of the pollutant with organic and inorganic components in the soil, sometimes referred to as pollutant aging.

In the application of soil bioremediation it is essential to identify the target chemical to be degraded. Bioremediation, involving the natural microflora and natural environmental conditions, is often the first choice. However, biostimulation through nutrient addition (particularly with nitrogen and phosphorus) and bioaugmentation or seeding with specific biodegrading organisms to increase the rate and extent of biodegradation, is thought to be worthwhile.

There are exciting, new developments in molecular microbiology that are starting to impact on bioremediation. For example, new biochemical pathways are being developed using plasmid-mediated genetic information exchange between strains of microorganisms. This is being done with a view to accelerating the biodegradation of more resistant molecules.

Bioremediation is a very interesting and potentially cost-effective way of cleaning-up environmental contamination by toxic organic materials. However, much more work is needed before it is accepted as a routine treatment technology.

REFERENCES

Evans, T E (1998) The Agricultural use of Biosolids (Sewage Sludge), *Int. Fertilizer Soc. Proc.*, 409, 1 27.

Goulding, K W and Annis, B (1998) Lime, Liming, and the Management of Soil Acidity, *Int. Fertilizer Soc. Proc.*, 410, 1 36.

Hall, A D (1945) *The Soil: an Introduction to the Scienti c Study of the Growth of Crops*, fifth edition, revised by G W Robinson, John Murray, London, 1 322.

Head, I M (1998) Bioremediation: Towards a Credible Technology, *Microbiology*, 144, 599 608.

Johnston, A E (1991) in *Long-Term Ecological Research: an International Perspective*, ed P Risser, SCOPE 47, John Wiley & Sons, Chichester.

Sumner, M E (1993) Sodic Soils: New Perspectives, *Aus. J. Soil Res.*, 31, 638 750.

Sumner, M E and Stewart, B A, eds (1992) *Soil Crusting: Chemical and Physical Processes. Advances in Soil Science*, CRC Press, Boca Raton, FL, 1 372.

Syers, J K, Lingard, J, Pieri, C, Ezcurra, E, and Faure, G (1996) Sustainable Land Management for the Semiarid and Subhumid Tropics, *Ambio*, 25, 484 491.

von Uexkull, H R and Mutert, E (1995) Global Extent, Development, and Economic Impact of Acid Soils, *Plant Soil*, 171, 1 15.

Soil Deterioration and Loss of Topsoil

Robert Evans
Anglia Polytechnic University, Cambridge, UK

Soil deterioration arises in a number of ways, one of which is the loss of topsoil. The impacts caused by soil deterioration can be seen over the short and long-term, onsite and, in the case of erosion of the topsoil, offsite. Indeed the offsite impacts of soil erosion may, over the short term, be much more important than the impacts on the land, and its users, which are directly affected by erosion.

Onsite soil deterioration can reduce the quality of the soil through a decline of fertility, acidi cation, salinization, or the loss of soil depth, all of which have corresponding impacts on crop and animal yields and soil and landscape biodiversity. In some instances there can be a total loss of the soil resource in terms of food, ber and timber production, for instance, when land is built on. Offsite impacts are related to the transport of soil particles from the land into water courses, reservoirs and ultimately the oceans.

The deterioration of soil is often closely bound up with social and economic factors, and can be further exacerbated by population pressure. Without a thorough understanding of the causes of soil deterioration and loss of topsoil and how these are often related to the use of the land, which in turn is frequently linked to social, economic and political factors, it is unlikely that the problem of soil deterioration will be resolved and the land used in a more sustainable way.

PROCESSES OF SOIL DETERIORATION

Most soil deterioration processes are related to how the land is used, but two, acidification and pollution, are also related to humankind s impacts on the atmosphere. Acidification can be induced both by application of fertilizers to the land or by airborne pollutants from industrial processes.

Loss of Topsoil by Erosion

Soil erosion, especially by water and wind, is the prime and most obvious cause of soil deterioration. Its manifestation can often be seen, as a dust storm or large gullies dissecting the landscape. For example, erosion came to the fore as a problem in the USA in the 1930s (the Dirty Thirties), as a result of the dust storms over the Great Plains and widespread water erosion of arable land (Bennett, 1939)

due to poor land management by farmers striving to make a living under dire economic circumstances. It is noteworthy, however, two decades before that Iceland had established a soil conservation department to combat wind erosion initiated by overgrazing and exposure of vulnerable soils on volcanic rocks or sandy glacial outwash. Concern about soil conservation was carried around the world from the USA, especially to countries which were then colonies of European states. Conservation techniques were often imposed in these countries, as they still are, with little understanding of the social, economic and political problems associated with erosion.

Much research in the USA focussed on intense rainfall as the major driver of erosion, and it was considered by many researchers that countries with generally low intensity rainfalls, such as those in Western Europe, did not have an erosion problem. That was a fallacy (Evans, 1996). Much erosion research in the USA and elsewhere has been based on plot experiments, often the plots were little more than 20 m in length and a few meters wide. Such experiments may have allowed replication of treatments and may have been good science, but they did not lead to a full understanding of the processes of water erosion. Only over the last two decades has the importance of landform been fully appreciated, for it is in depressions and valley floors that much erosion occurs. Plots cannot simulate such situations. Since field, rather than plot, assessments of erosion became more widespread, especially in Europe, there has been a better understanding of water erosion, its processes and the social and economic factors which drive it.

Agriculturists and many soil conservationists have written that soil erosion is a natural phenomenon, with the implication that because of this, erosion is less of a problem than it really is. Indeed, erosion of topsoil (and subsoil) is a natural phenomenon, but without the activities of humankind it would be much less widespread than it is now. High rates of erosion can occur in little altered landscapes not settled by large numbers of people, for they are usually inhospitable places. Glaciated mountains, especially those receiving high precipitation and still undergoing uplift and subject to earthquakes, such as the New Zealand Alps, the Himalayas and Andes, all suffer high natural rates of erosion. For example, the mean annual rate of erosion of the New Zealand Alps is estimated to be $100\,m^3\,ha^{-1}$, a surface lowering of $10\,mm\,year^{-1}$. Semiarid regions with sparse vegetation cover can erode severely when rare intense storms occur, *badlands* are the most severe example of this. Wind erosion can also be effective in deserts. However, elsewhere, and that is most of the world s land surface, where vegetation protects some 30% or more of the soil surface, erosion is rare.

Most landscapes were wooded, even many present day savannas, until humankind settled them in appreciable numbers and cleared the trees. In steeply sloping localities, tree clearance can destabilize slopes so that rapid mass movements such as landslides and landslips occur much more frequently and extensively, as in New Zealand. Triggering of mass movements is widespread in the mountainous tropics as slopes are deforested or, more commonly, settled for agriculture as population pressure increases. Even in temperate Britain there is much evidence that in the past, in periods of tree clearance, mass movements and gullying occurred.

Clearing trees from land changes its hydrology. More water becomes available for runoff, but erosion need not occur if the vegetation cover remains intact. However, it is likely that gullying will take place in valley floors and that stream heads will retreat upvalley, especially if animals track valley floors. The initiation of gullying and valley floor incision was widespread, for example, in the USA and South Eastern Australia when range lands were opened up for grazing in the late 19th century, and is widely seen now in much of tropical Africa. The increase in grazing pressures could sometimes be accompanied by climate change, as in the USA, which exacerbated erosion. Grazing pressures in the past could have initiated erosion which still continues today, for example, the Loess Plateau in China. However, some Chinese scientists consider that the erosion and desertification was triggered by the opening of the Silk Road trade route across the prairie grassland by creating a linear source of dust and sand which spread downwind over the centuries.

When land is cleared for agriculture, wind and, especially, water erosion can become widespread and sometimes severe. Compared to natural rates of erosion, rates can increase by orders of magnitude and in the case of water erosion will be associated with increased runoff and flooding. The impacts of erosion induced by agriculture (infilled valleys, thinned valleyside soils) can be seen not just in countries where intense rainfalls can cause severe present day erosion, but also in countries characterized by gentle rains. There is much evidence for widespread erosion since the Europeans settled North America. Similar evidence is found on the chalk, limestone and sandstone outcrops of Britain, much of it dating from clearance of woodlands up to 4000 years ago, but also due to cultivation in times of population and economic pressures in Romano-British and Medieval times and the present day. What happened in European and European-settled countries in the past is now happening widely in developing countries as their populations expand.

Peat Wastage

Peat soils when drained shrink and oxidize. Not only do topsoils disappear, so do subsoils. In Britain, for example, the East Anglian peat Fenlands have largely disappeared

since they were first drained in the mid-17th century and up to 4 m of peat has wasted away. The subsoil exposed is mostly of inferior quality to the peat and the lowered land surface has to be protected from flooding. In many parts of the world, especially South East Asia, new areas of peat land are presently being drained to add to the large coastal and deltaic areas already converted to agriculture.

LOSS OF NUTRIENTS FROM THE SOIL

Cereal and timber crops and animals remove from the soil, via harvest or slaughter, the nutrients they have taken from the soil which are stored in the seed, wood, meat, bone or hide. In a sustainable farming system the nutrients removed from the soil have to be replaced. Soils on grazed ranges will slowly deteriorate as rarely is any effort made to replace nutrients. In arable systems some effort must be made to replace the nitrogen, phosphorus, potassium and micro-nutrients needed by the crops otherwise yields will decline to a point which makes it not worthwhile farming the land. In the tropics where *slash and burn* agriculture is practiced, (*see* Prehistoric People and Forest Soils, Volume 3; Shifting Cultivation and Land Degradation, Volume 3) the arable land is left for grasses, shrubs and trees to recolonize so that when cleared and burnt in the future their ashes will return nutrients to the soil. A similar system was used by Mesolithic settlers in Britain.

As population increases, agriculture can become more intensive. To replace nutrients the land can be left fallow for a season to allow inputs from the atmosphere, especially nitrogen, to accumulate in the topsoil. Weeds can be cultivated into the topsoil to return organic matter and nutrients. If animal manure is available this too can be added to the land. Animals can be grazed on common land and then herded onto the arable land, or they can graze the vegetation that is left after crop harvest which comprises the crop stubble, unharvested plants or seedheads and weeds, and their dung returned to the soil to fertilize it. Such systems are widespread in many developing countries and were extensive in Britain until the 18th century.

In Europe, and earliest in the Netherlands, fallow land was gradually replaced by an arable rotation. In Britain this formed the basis for a four course rotation which became widespread in the late 18th century. In such a system, root crops or grass replaced the fallow and these were fed to cattle and sheep and their dung returned to the land. Yields increased greatly. However, for this system to work well, land had to be enclosed and farmers had to have secure tenure and markets for their products. This system offered a sustainable route for European farmers to take as populations grew. Such a route does not seem easily available in the present day, for climatic and social reasons, in many tropical countries of the world.

In the middle and late 19th century farmyard manure was in turn replaced in Europe and North America by manufactured fertilizers. Modern agriculture in the developed world and in those parts of the tropics where the *green revolution* came into being in the 1960s and 1970s, now depends on high inputs of artificial fertilizers (*see* Green Revolution, Volume 3). Ironically, in Britain, for example, because gases emitted to the atmosphere from industry and power stations are now much cleaner, insufficient sulfur is being returned to the soil and crop yields suffer unless fertilizers containing sulfur are added to the land.

Loss of Organic Matter from the Topsoil

Organic matter is a crucial component of the topsoil, for it not only holds and releases nutrients and water to the crop in disproportionate amounts to those released from mineral soil, but it also makes the soil better structured and resistant to erosion. The exposure of organic matter to the atmosphere by cultivation leads to its oxidation. After many decades an equilibrium is reached, but this may be below the level needed to maintain soil structural stability. Organic carbon is a surrogate for organic matter and a level of 2% is considered by some authorities to be needed to give a topsoil structural stability and so be resistant to erosion. The amount of organic carbon in a topsoil is positively related to its clay content. However, even in Britain with its temperate climate, many soils do not contain sufficient clay to retain 2% organic carbon. In hotter and drier climates less organic matter is held in soils in proportion to their clay content.

Salinization

Salinization of the soil comes about in two main ways. The first occurs on land irrigated by water containing salts, and such land is widespread in the semi-arid parts of the world. If the salts, especially sodium chloride, are not flushed through the soil they are brought to the surface by evaporation and are deposited on the surface. Such salinity inhibits growth and can be toxic to plants. The second way is due to the removal of trees which changes the hydrology and causes the groundwater table to rise to the surface, where salts contained in the water are deposited. As a result, in many localities in Australia, for example, plants in valley floors die, exposing bare soil which is then susceptible to water erosion. Soil salinity can also occur when adjacent seawater or fossil saline water intrudes into groundwater because of overabstraction, as occurs in many deltaic regions such as the Nile and Mekong deltas (*see* Salinization, Volume 2; Irrigation: Environmental Impacts, Volume 3; Salinity and Agriculture, Volume 3; Waterlogging, Volume 3).

Acidification of the Topsoil

When organic matter in topsoils breaks down, it releases weak acids. Over time, most topsoils will acidify unless they remain under (native) woodland where leaf fall recycles nutrients including calcium to maintain the soil s nutrient and acidity profile. However, if the air becomes polluted with acidifying gases such as sulfur dioxide, from burning fossil fuels in homes and industry, this may be *scavenged* from the air by plants. The acid water can then seep from the vegetation into the soil. An increase in acidity (decline in pH) will have to be rectified because many crops will not grow and yield well on acid soils (pH < 6.0). For natural or semi-natural vegetation, more acid tolerant plants will replace those which prefer higher pHs. When soils become very acid (pH < 4.0) aluminum may be released and in solution this is toxic to both plants and animals.

Acidification can also take place when nitrogenous fertilizers are applied to land year after year if the resulting weak acid released, and leached into the soil, is not neutralized by liming. The draining of soils, often developed in former estuarine deposits which contain iron pyrites, as in East Anglia, England or in South East Asia will lead to oxidization of the pyrites to form sulfuric acid and so result in extremely acid soils.

Other Soil Deterioration Mechanisms

Pollution

Pollution of soils and vegetation by emissions of heavy metals from smelting works can lead to the deterioration of the vegetation cover and ultimately to the exposure of soil to erosion.

Compaction

Compacted soils hold and release less water to the plant and dense and capped soils inhibit seedling germination. Topsoil structure is partly related to organic matter content (see above) but soils are also liable to be compacted by machinery or by animals. Land cleared of tropical forest by bulldozers is compacted more than that cleared by hand or by using light equipment drawn by animals. Using larger agricultural machinery increases the risk of soil compaction, especially if the land is cultivated or harvested when it is too wet. Animals can compact the soil if they congregate in large numbers, for instance, around a watering hole.

Waterlogging and Construction

Waterlogging of the soil can happen when surface drainage is obstructed by humankind s actions and land *built over* for homes, industry, commerce, and transport is removed from agriculture. In many parts of the world it is often the better agricultural land which is being swallowed up in cities.

EVIDENCE FOR SOIL DETERIORATION

There is a massive literature on the deterioration or degradation of the world s soils and the loss of topsoil by erosion. Unfortunately, much of the literature is of low quality, and there is little useful quantitative information. For example, much of the world s information on rates of soil erosion is based on plot experiments which have little relevance to what is actually happening in farmers fields or common grazing lands and which may overstate the rates of erosion by a factor of 10 (Evans, 1995).

Oldeman *et al.* (1991) have produced a world map of the status of human induced soil degradation. This is based on expert opinion. Although there have to be many reservations about the quality of this information, it is at present the best there is. Four degrees of soil degradation are classified, *light*, somewhat reduced productivity of the terrain; *moderate*, greatly reduced productivity; *strong*, soils are not reclaimable at farm level, but may be restored by major effort; and *extreme*, unreclaimable and beyond restoration.

North America is least affected by degradation (Table 1), Europe most, but it is Africa and Central America which contain the highest proportions of most degraded land. The areas of land affected by degradation seem small until they are compared with the areas of land suitable for agriculture (Table 2). Except for North America, between one fifth

Table 1 Impacts of human-induced soil degradation

	Light (%)[a]	Moderate (%)[a]	Strong (%)[a]	Extreme (%)[a]	Total (%)
World	5.8	7.0	2.3	0.1	15.2
Asia	6.9	8.1	2.5	<0.1	17.6
Africa	5.8	6.5	4.2	0.2	16.7
South America	5.9	6.4	1.4	–	13.7
Central America	0.6	11.6	8.3	–	20.6
North America	0.9	4.1	<0.1	–	5.0
Europe	6.4	15.2	1.1	0.3	23.0
Australasia	10.9	0.4	0.2	<0.1	11.7

[a] See text for definitions.
Source: (Reproduced from Oldeman *et al.*, 1991).

Table 2 Land suitable for agriculture and the proportion of it which is degraded

	Suitable land (% of total land)	Degraded land (% of suitable land)
World	60.5	24.9
Asia	55.1	31.9
Africa	60.4	27.5
South America	77.6	17.7
Central America	73.9	27.9
North America	40.7	12.4
Europe	87.7	26.3
Australasia	60.9	19.2

Source: (Reproduced from Oldeman *et al.*, 1991).

and one third of the land of the continents suitable for agriculture has been affected by soil degradation.

Erosion by water and wind account for by far the largest proportion of the world s soil degradation (Table 3). Loss of nutrients is most important in South America, salinization in Asia, compaction in Europe and water logging in Central America. The types of degradation defined by (Oldeman *et al.*, 1991) relate well in the main to those described above. However, they attribute the loss of topsoil by water erosion just to *sheet erosion*, and their category *terrain deformation* includes rills, gullies and landslides. They claim the loss of topsoil by sheet erosion is the most important type of soil degradation but that appears unlikely if, as (Evans, 1995) argues, sheet erosion is generally an ineffective process in comparison to riling and gullying. It seems more likely that, as must be the case in North America, sheet erosion must include (small) rills. Estimates of degradation attributed to loss of topsoil by sheet erosion must be treated with circumspection therefore. In many localities, but probably not in North America, the loss of soil and crop productivity attributed to sheet erosion is probably related to loss of soil nutrients (see below).

Deforestation has triggered most degradation in South America, Asia and Europe (Table 4), overgrazing in Africa

Table 3 Type of soil degradation

	Percent of total degraded land							
	World	Asia	Africa	South America	Central America	North America	Europe	Australia
Loss of topsoil	46.8	48.9	41.5	39.1	33.5	62.6	42.4	79.4
Terrain deformation	8.8	10.0	4.5	11.5	40.1	–	10.0	1.1
Wind	27.9	29.7	37.7	17.2	7.3	36.2	19.3	15.9
Peat wastage	0.2	0.2	–	–	–	–	1.2	–
Loss of nutrients[a]	6.9	1.9	9.1	28.0	6.7	–	1.5	0.4
Salinization	3.9	7.0	3.0	0.9	3.7	–	1.7	0.9
Acidification	0.3	0.5	0.3	–	–	0.1	0.1	–
Pollution	1.1	0.2	<0.1	–	0.6	–	8.5	–
Compaction	3.5	1.3	3.7	1.6	0.2	0.9	15.1	2.2
Waterlogging	0.5	0.1	0.1	1.6	7.8	–	0.4	–

[a] Includes loss of organic matter.
Source: (Reproduced from Oldeman *et al.*, 1991).

Table 4 Factors causing soil degradation

	Percent of total degraded land				
	Deforestation	Overgrazing	Cultivation	Domestic[a]	Industry
Asia	39.9	26.4	27.3	6.2	0.1
Africa	13.6	49.2	24.5	12.7	–
South America	41.0	27.9	26.2	4.9	–
North and Central America	11.4	24.0	57.6	7.0	–
Europe	38.2	22.7	29.1	0.4	9.5
Australasia	11.6	80.6	7.8	–	–
World	29.4	34.5	28.1	6.8	1.2

[a] Overexploitation of the vegetative cover for domestic use.
Source: (Reproduced from Oldeman *et al.*, 1991).

and Australasia, and cultivation in North and Central America. Industrial pollution as a cause of degradation is most widespread in Europe.

In most parts of the world water erosion, comprising interill (rainsplash and sheetwash) and especially channel (rill and gully) erosion (ISSS, 1996), is the major process degrading topsoils. The tentative map of the global pattern of suspended sediment yield shows where erosion rates are highest (Figure 1). However, the yield includes sediment derived from erosion both of the land surface by water and mass movements and from river channels, and identifies areas of high natural erosion rates as well as areas of high accelerated erosion rates. In some regions, for example, South East Asia, both high rates of natural and accelerated erosion occur, whereas elsewhere, mountainous regions especially if glaciated will also have high rates of natural erosion. High rates of erosion attributable to agriculture and grazing are found in the USA, the Mediterranean and Eastern Africa. The amount of sediment which reaches stream and river channels and the oceans is only a fraction of that eroded from the land, for much of the eroded soil is deposited before it reaches channels (Walling, 1983). However, much of the silt, clay and organic fractions, which are eroded from the land, will reach river channels. Thus, Figure 1 gives an indication of where accelerated erosion is occurring but is not helpful with regard to rates of erosion of the topsoil.

Rates of water erosion measured on experimental plots are frequently reported (Pimental, 1993), but unfortunately little of this information can be translated into *reality* (Evans, 1993). Rates of topsoil loss occurring in farmers fields or eroding rangelands are not widely reported but often the data implies that actual rates of erosion are much less than those recorded in plot experiments (Evans, 1995). Estimates of amounts of soil eroded often can not explain the losses in crop productivity of the soil, for losses are often less than a few millimeters in depth a year. In many parts of the world the decline in productivity is probably related not to the mass of the soil lost but to the loss of nutrients attached to silt, clay and organic particles which are transported from the soil by erosion and, especially, by the removal from the soil by the crop of nutrients which are not then replaced by the farmer. The loss of organic matter by oxidation when exposed to the atmosphere and the diminution of the role of organic matter in recycling nutrients, especially phosphorus, may be crucial, particularly in the tropics (Webster and Wilson, 1998).

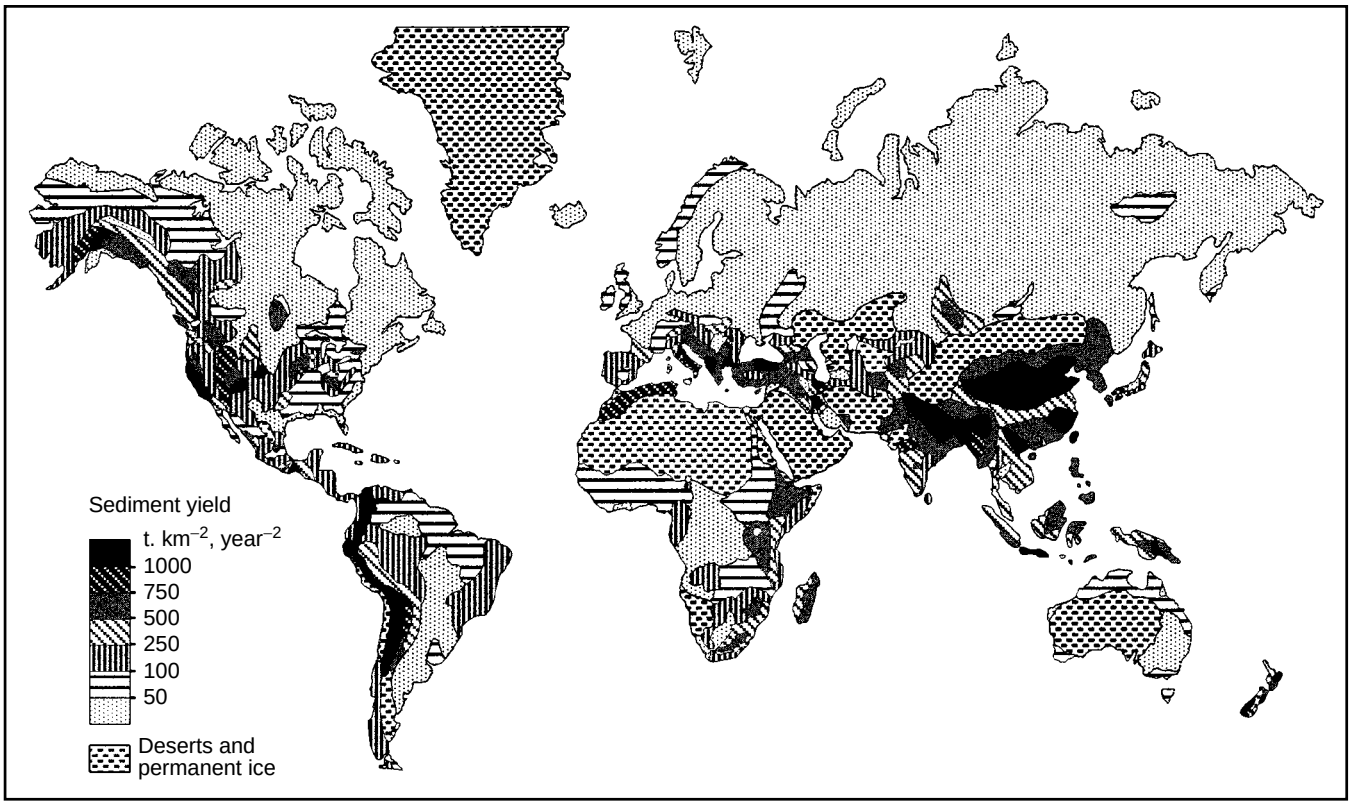


Figure 1 A tentative map of the global pattern of suspended sediment yield (Reproduced with permission from the Soil and Water Conservation Society in Walling, 1988) – includes all sediment from both disturbed and undisturbed areas of the Earth

SOIL DETERIORATION IN THE FUTURE

The outlook for soils in the next few decades is at best unsure, and is most likely to be bad. Soils will continue to deteriorate if land use continues as now. Also, as more forests are cut down and converted to pasture or arable land, wider areas will be affected by mass movements, runoff and erosion, and more peatlands will be at risk of wastage. Erosion of topsoil by water and wind will continue if large scale intensive agriculture continues and especially if these techniques expand into areas of less intensive agriculture. Cultivation techniques are available to alleviate erosion, such as conservation tillage which leaves a good protective cover of vegetation and stubble on the ground into which the crop is drilled. However, for large scale agriculture this technique relies on the use of a herbicide to kill weeds which compete with the crop. The herbicide could have an adverse impact on the environment s biodiversity and also pollute water courses. Grazing induced erosion will continue unless less intensive grazing regimes are introduced in many parts of the world.

Unless nutrients removed by grazing animals and crops are replaced by fertilizers soils will continue to deteriorate. However, before they can be used, for many in the world fertilizers will have to be made easily available and affordable. Moreover, over-application of fertilizers may lead to pollution and enrichment of drinking waters. Soils under continuous cultivation will loose organic matter and soil structure will deteriorate unless ways of recycling organic matter can be devised, such as the application to the land of compost made from urban household waste (*see* Soil Amelioration, Volume 3).

The use of nitrogenous fertilizers and emissions from industry could continue to cause acidification and pollution of topsoils. However, in developed countries laws to curb industrial pollution, such as the European Union s Integrated Pollution and Prevention Control regulations, may alleviate the problems of acidification and pollution. Judicious liming of the soil could keep acidification at bay, but that is only feasible where lime is easily available and cheap to buy. Salinization will continue unless more care is taken in the future when land is irrigated and unless techniques for alleviating salinity are brought into use more widely. The use of low pressure tyres on farm machinery, and varying of the depth of cultivation, could help in the future to alleviate soil compaction. Presently, it seems most unlikely that the building over of topsoils will cease in the future and hence good quality agricultural land will continue to be lost.

What appears a bleak picture could be made worse in those localities most affected by global warming. However, predicting where those localities will be is not easy at the present time, because of the coarseness of the resolution of the general climate circulation models. Increases in precipitation and in the occurrence of large storms could trigger mass movements on slopes and increase runoff and erosion. Increasing dryness could lower erosional thresholds on grazed land because vegetation could become sparser and recover from grazing more slowly. It will be where land use changes with climate change that impacts on the soil are likely to be most severe. Thus, forest clearance coinciding with increased rainfall or an increase in the area of land under cultivation as temperatures warm up at higher altitudes and latitudes will increase the chances of erosion happening.

CONCLUSIONS

There is much evidence worldwide, from experiments and observation, for the deterioration of soil and for the loss of topsoil. But there is little good quantitative evidence for the extent and size of the deterioration, even for the most studied aspect, accelerated soil erosion.

There is little apparent acceptance by farmers that soil erosion is a problem. This is because in developing countries subsistence farmers cannot afford to conserve their soils. And, like their colleagues in the developed world, generally they are not affected enough by erosion (and deposition) for them to want to do anything about it. Interrill erosion does not remove plants, nor do rills which run between rows of crops, and deposits rarely bury much crop. Only gullies and their deposits seriously interfere with the task of growing crops, and often these are of infrequent occurrence in a soil landscape.

Declines in crop yields can probably most often be attributed to declines in soil fertility. In those localities in the developed world where declines could be attributed to erosion thinning the soil they will be hidden by the application of fertilizers. In many localities it will likely take two or more generations for high rates of rill erosion to strip sufficient depths of soil to affect yields.

The short term impacts of accelerated erosion are most apparent off the farm; flooding, damage to property, siltation of water courses and damage to fisheries, pollution of water bodies and the oceans. These damages are costly to remedy and are often borne by that part of society which does not farm the land.

This is not to say the long-term impacts of erosion are unimportant; they are. In England and Wales for instance, it is estimated (Evans, 1996) that since the land was cleared of forest for agriculture the loss of crop production resulting from the thinning of topsoils and the exposure of impoverished subsoils is of the order of 10%. At present day prices this can be valued at £700 million a year.

Few countries have good quality information on the extent of erosion and the risk of erosion. There is a need to monitor erosion and other processes which lead to the deterioration of soils, but the monitoring needs to be done in farmers fields and on rangeland. We need to know what is actually happening, not what is considered to be happening based on anecdotal, local, experimental plot or modeled evidence. This information is especially needed so that good predictions of risk of deterioration based on reality can help us alleviate the probable adverse impacts of future global warming.

The rapid concreting over of land for houses, industry and roads may well at the present time be the world s most severe problem of soil deterioration. Although there is awareness of the problem in some countries, notably the UK and USA, there is a dearth of evidence on its impacts.

In the future, unless humankind makes a determined effort to use its soils wisely, soils will continue to deteriorate and disappear. The implications for long-term food, fiber and timber production could be severe. Over the short term, in many less developed parts of the world soil fertility needs to be maintained or, even better, improved, whilst in many localities accelerated soil erosion needs to be stopped to curtail such off-farm impacts as flooding, pollution and sedimentation. Planning regulations are needed to protect soils from being built on.

REFERENCES

Bennett, H H (1939) *Soil Conservation*, McGraw-Hill, New York.

Evans, R (1993) On Assessing Accelerated Erosion of Arable Land by Water, *Soils Fert.*, 56, 1285 1293.

Evans, R (1995) Some Methods of Directly Assessing Water Erosion of Cultivated Land a Comparison of Measurements Made on Plots and in Fields, *Prog. Phys. Geogr.*, 19, 115 129.

Evans, R (1996) *Soil Erosion and its Impacts in England and Wales*, Friends of the Earth, London.

ISSS (1996) Terminology for Soil Erosion and Conservation.

Oldeman, L R, Hakkeling, R T A, and Sombroek, W G (1991) *World Map of the Status of Human-induced Soil Degradation: an Explanatory Note*, 2nd edition, International Soil Reference and Information Centre, Wageningen, United Nations Environment Programme: Nairobi GLASOD.

Pimental, D, ed (1993) *World Soil Erosion and Conservation*, Cambridge University Press, Cambridge.

Walling, D E (1983) The Sediment Delivery Problem, *J. Hydrol.*, 65, 209 237.

Walling, D E (1988) Measuring Sediment Yield from River Basins, in *Soil Erosion Research Methods*, ed R Lal, Soil and Water Conservation Society, Ankeny, 39 73.

Webster, C C and Wilson, P N, eds (1998) *Agriculture in the Tropics*, Blackwell Science, Oxford.

Soils, Human Impacts on

see Sedimentary Records of Long-term Ecological Change (*Opening essay*, **Volume 2**)

Sulfurous and Photochemical Smog

see Urban Sulfurous and Photochemical Smog (**Volume 3**)

Sustainable Irrigation?

see Irrigation: Environmental Impacts (**Volume 3**)

Swidden

Swidden, in modern usage, means shifting cultivation. Swidden is frequently said to be an old English term but it is not found in dictionaries of old English, or in a thesaurus of old English with extensive references to agricultural practices. The *Oxford English Dictionary (OED)* describes swidden as a northern dialect word to describe an area of land that has been cleared by slashing and burning. Swidden, also spelt as swizzen (to burn or singe), was first used in published English in 1788 by W H Marshall in his *Rural Economy of East Yorkshire* to describe the burning off of heath.

There is also a long published history of the use of a single word to describe shifting cultivation in Scandinavia. The Swedish word *svdjebruk* and the Finnish word *kaskeaminen*, are translated as swidden by Raumolin (1987). Raumolin finds the first papers on swidden agriculture were published in Swedish in 1750 and in Finnish in 1753.

The OED implies that the modern use of the term began in 1951, when K G Izikowitz in *Etnologiska Studier* wrote In English it is sometimes called shifting cultivation or slash and burn. There is no single word in English which covers the meaning .... In searching for an English word I have taken ... a dialect word, *swidden* . A number of similar terms for shifting cultivation are encountered in the literature. They include *ladang* (Malaysia and Java), *juma* (Sumatra), *kaingin* (Philippines), *jhum* (Assam), *taungya* (Burma), *milpa* (Guatemala), and *chitimene* (Uganda).

The association amongst swidden, burning and primitiveness is both unfortunate and wrong. Shifting cultivation is not a primitive form of agriculture, nor does it always involve burning. It was a very early form of agriculture. That it survives today suggests that it is highly adaptive.

REFERENCE

Raumolin, J, ed (1987) *Special Issue on Swidden Cultivation*, Suomen Antropologinen, Helsinki, 278.

FURTHER READING

Izikowitz, K G (1951) *Lamet, Hill Peasants in French Indochina, Etnologiska Studier 17*, Etnografiska museet, Gotenburg.

Roberts, J, Kay, C J, and Grundy, C (1995) *A Thesaurus of Old English*, King s College, London.

Simpson, J A and Weiner, E S C, eds (1989) *Oxford English Dictionary Online*, 2nd edition, www.lib.ucdavis.edu/hss/English/.

BRYANT ALLEN *Australia*

Tidal Power Development

David Elliott

The Open University, Milton Keynes, UK

Tidal flows in rivers have been used since medieval times in tidal mills for corn grinding. However, it is only recently that tidal power has been seen as a potentially large source of energy.

Some very ambitious schemes have been considered for electricity production from barrages across major estuaries in the UK, Canada and elsewhere, but so far the only significant scheme is the 240 megawatt (MW) Tidal Barrage on the Rance estuary in Brittany in France, which was commissioned in 1968. There are some smaller barrages elsewhere, in Canada and China, for example, but although the global tidal resource is put at around 100 000 MWs (capable of generating perhaps 300 TWh per annum – which is equivalent to the UK's overall electricity requirements) this major renewable source of pollution free-power has not yet been taken up significantly.

Looking at the basic physics and engineering of tidal power can identify the main reasons for this. The rise and fall of the tides is caused primarily by the gravitational pull of the moon on the seas, which, modified by the smaller influence of the sun, produces a complex but highly predictable pattern of tides around the world. At any one point, there are two tides (approximately) every 24-h period, as the earth turns on its axis, but at most sites the tidal rise is small, whereas in others, like the Severn estuary in the UK, it can be enhanced by complex geographical funneling and resonance effects.

The construction of barrages or dams across estuaries allows water to be trapped at high tide, and this head of water can be used to power turbines, as with hydro electric power plants. The main difference is that the head of water is much lower than with hydro dams, as is therefore the power output – which is proportional to the square of the head. In addition, the continually shifting twice daily tidal pattern means that power can only be generated at certain times, and these may not coincide with periods of consumer demand. The end result is that expensive plant is not used fully, and although two-way operation and energy storage options exist, they are expensive.

In the longer term, when they have paid off their large initial capital costs, barrages, like hydro plants, can generate power relatively cheaply: after all, their fuel is free. They can also pay back their energy costs quite quickly – the energy used in construction and materials is only around 7% of the energy likely to be generated in a plant's lifetime, and relatively minor upgrades to turbines can extend the lifetime to a hundred years or more. However, given the short-term perspective adopted by most investors at present, barrages are not seen as economically attractive.

This economic problem can be exemplified by looking at one of the most developed proposals. The Severn estuary offers one of the best sites in the world given its large 11-m tidal range. The scheme proposed by the Severn Tidal Power Group, a consortium of leading UK construction firms, would involve the construction of an 18-km long barrage from Weston super Mare in Somerset across the Bristol Channel to the Welsh coast. It would have some 8600 MWs of generating capacity in the form of turbines embedded in the barrage structure and could generate around 6% of the UK's annual electricity requirements. However, it would cost around £10 000 million to build. So far, no private investor has emerged willing to take on this project; and large-scale publicly funded projects are currently not politically popular.

While the economic problem is probably the central issue, there are also potential environmental problems. Put simply, if you block up an estuary you are likely to have impacts on the local eco-system. Of course, some of the effects may be positive and the precise balance will be site specific. For example, given the lowered tidal range upstream, the decrease in the area of regularly exposed mud flats on the Severn might reduce the attractiveness of the site to some species. However, given the reduction in tidal flow, suspended silt would drop out, so that light would penetrate deeper, and the biological productivity of the water would increase, making it more attractive to

other species. Nevertheless, several environmental groups, fearing net adverse effects on migratory birds in particular, have come out against barrages.

With barrages in effect stalled as an option, attention has turned in the UK and elsewhere to another approach the use of tidal streams. Whereas barrages make use of the vertical tidal rise, use can also be made of the horizontal flow of tides the so-called tidal currents. A number of tidal current turbines have been developed: they are effectively like undersea wind turbines.

Although, as with barrages, the technology can only be deployed in specific sites, the technical potentials look good one study suggested that the UK could obtain nearly 20% of its electricity from such sources. The economic potential is less clear. In principle, tidal current turbines should be much cheaper than barrages since there is no need to build expensive (and environmentally invasive) dams across estuaries: tidal turbines can be free standing, mounted on the sea bed, or suspended from pontoons, out to sea. However, the power has to be brought back to land by expensive marine power cables and, as with offshore wave energy devices, tidal turbine systems would have to be designed to survive storms. As a result, at present, the technology is seen as only likely to be marginally competitive.

Tidal power offers a significant source of pollution-free energy, but so far only a few small to medium scale projects exist or are planned, perhaps the most promising being the 30 MW tidal current turbine system planned for installation on a causeway in the Philippines. With concerns about climate change growing, it could be that, if projects like this are successful, then the tidal option will be taken up seriously.

See also: Tides, Atmospheric, Volume 1.

Tourism and Ecosystems

Joanna Burger
Rutgers University, Piscataway, NJ, USA

One aspect of global change that is often overlooked is the increase in movement of people on a temporary basis; people y within and among continents for business and pleasure. Travelling for pleasure, or tourism, is increasing worldwide as western economies boom and developing economies allow their citizens to attract tourists from abroad, and to travel themselves.

Tourism includes visiting friends and relatives, visiting archaeological or cultural features, basking in unique or pleasurable ecosystems (beaches, mountains, deserts), seeking exciting or challenging hunting or shing opportunities, enhancing other recreational activities, and visiting natural and ecological attractions. Ecotourism, aimed at providing the visitor with opportunities to view plants and animals within their natural ecosystems, is growing at an incredible pace. Ecotourists travel to the world's parks and nature reserves, and are attuned to the ideal of sustainable development. Ecotourism is a partnership among the travel industry, tourists and the conservation community to enhance environmental sensitivity (Ehrenfeld, 1992). People travelling on their own or on tours are visiting a wide array of habitats, and merely mentioning East Africa conjures up images of vast herds of zebra, wildebeest and gazelles, stalked by cheetahs during the day, and lions and leopards by night. Other world-renowned attractions include the Galapagos Islands made famous by Charles Darwin, the Antarctic with its colonies of penguins and nesting seals, Madagascar with its unusual primates, and the Great Barrier Reef with its spectacular sh diversity, colonial seabirds, and nesting sea turtles. Tourism, however, can also involve exploring towns, cities, and natural areas around home, with short trips of a few days.

Whenever people move about they leave their footprint on plant and animal communities, as well as on the ecosystem itself. In some cases, the imprint is short-lived: plants are trampled for only a brief period and then recover, or animals move temporarily, but return once the intrusion has passed. In others, the effects of tourism are longer-lasting; soil is compacted, plants die out, sea turtles abandon their nesting sites for less optimal places, iguanas are forced into other habitats or exposed to predators, birds abandon colony sites or foraging areas, and mammals leave suitable habitats to move farther into cover. Sometimes the effect of tourism is to reduce the time that animals can forage effectively in optimal habitat, increase the time they devote to vigilance for people and predators, increase their risk of predation, or expose themselves or their young to thermal stress.

While plants and animals, and ecosystems themselves, are resilient and have mechanisms to adapt to pressures from people, the problems faced by many species, communities and ecosystems are a direct result of increasing numbers of people, in different habitats, in more remote areas, for a longer period of time. While the effects of global changes in habitat, and habitat loss itself (Reaka-Kudla et al.*, 1997) are well known, the effects of tourism on naturally functioning ecosystems are less well studied. Such effects are often subtle, involving minor shifts in habitat, slight changes in species composition, or small declines in reproductive success, growth rates and survival. Small effects are cumulative, changing the overall health of plant and animal communities, and ultimately the ecosystems themselves. For*

species that are threatened or endangered, the effects of tourism can be catastrophic. As tourism increases globally, the effects on ecosystems will increase disproportionately, requiring careful environmental planning and management to sustain natural ecosystems.

In this article global changes in tourism, particularly ecotourism, are explored as they relate to Man's effect on individuals, populations, species, communities, ecosystems, and the landscape. Ecotourism is both a cause of concern for conservation, and a possible avenue of support for conservation. For example, the vast numbers of tourists that visit the Galapagos Islands contribute directly to their preservation through monetary rewards to the Ecuadorian government and through support of local guides, hotels, restaurants, boats, and other facilities (DeGroot, 1983). Similarly, many islands in Australia's Great Barrier Reef depend on ecotourism, yet are partially degraded by decades of resource use (Walker, 1991). Too many tourists strain fragile ecosystems, leading to declines of sensitive species, and reductions in the quality of the experience for the tourists themselves. Environmental degradation due to tourism is in many ways a common issue (Ostrom et al., *1999): each tourist who visits a place derives the full pleasure, while incurring only a small cost of the damage caused by that visit.*

TOURISM IS ON THE RISE

Tourism of all kinds is on the rise worldwide, as well as in countries such as the US, and is the largest business in the world (Miller, 1993). Tourism itself is difficult to define, since people can travel for the day, the weekend, or for extended periods of time, and can travel far from home, or tour around nearby. In this article I refer to tourism as travel by people, away from home, to see and observe natural, cultural, and political sights. By any measure, global tourism is not evenly divided among regions. Europe receives the greatest number of tourists, followed by the Americas and East Asia (Table 1, Figure 1, after World Tourism Organization, 1998, 1999).

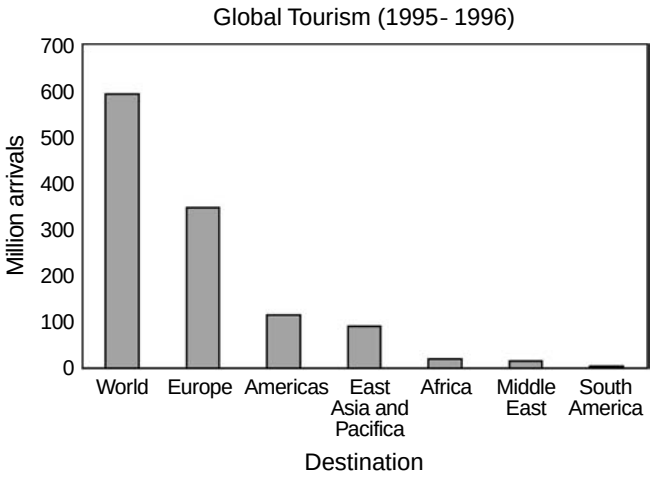


Figure 1 Tourism arrivals by region (after World Tourism Organization, 1999)

Ecotourism, as well as recreational activity, is also on the rise. For the US as a whole, based on a national survey, walking is the most popular recreational activity, followed by sightseeing, picnicking, swimming, fishing, bicycling, and bird watching (Cordell *et al.*, 1999). Some types of tourism and recreation are increasing more than others (Figure 2). Bird watching, hiking, backpacking, downhill skiing, and primitive camping are the five fastest growing activities in the US. For example, 21.2 million Americans (12%) were at least casual birdwatchers in 1982, while in 1995 the number had grown to 54.1 million (27% of Americans, Cordell *et al.*, 1999). Although some of these activities can be accomplished very near to home, most involve tourism.

Another measure of ecotourism is the number of people visiting National Parks or other reserves. In the US, the number of people visiting all sites under the National Park Service (including historic sites, battle fields, military parks, national monuments, national parks, preserves, seashores, and scenic rivers) has been increasing for many years, as has the number of people visiting the National Parks (Figure 3, after National Park Service, 1999). Despite the

Table 1 Ten most visited sites of the US National Park System (after Wiesenfeld, 1999)

Site	State	Hectares	Visitors in 1997
Blue Ridge Parkway	North Carolina, Virginia	35 802	18 373 279
Golden Gate National Recreation Area	California	30 146	13 803 382
Great Smoky Mountains National Park	Tennessee, North Carolina	211 280	9 965 075
Lake Mead National Recreation Area	Arizona, Nevada	605 328	8 528 422
Gateway National Recreation Area	New York, New Jersey	10 770	6 807 945
Natchez Trace National Parkway	Alabama, Tennessee	20 952	5 995 978
George Washington Memorial Parkway	Virginia, Maryland, D.C.	2934	5 844 204
Grand Canyon National Park	Arizona	492 876	4 791 668
Delaware Water Gap National Recreation Area	New Jersey, Pennsylvania	27 210	4 752 100
Statue of Liberty National Monument	New York, New Jersey	23	4 738 388

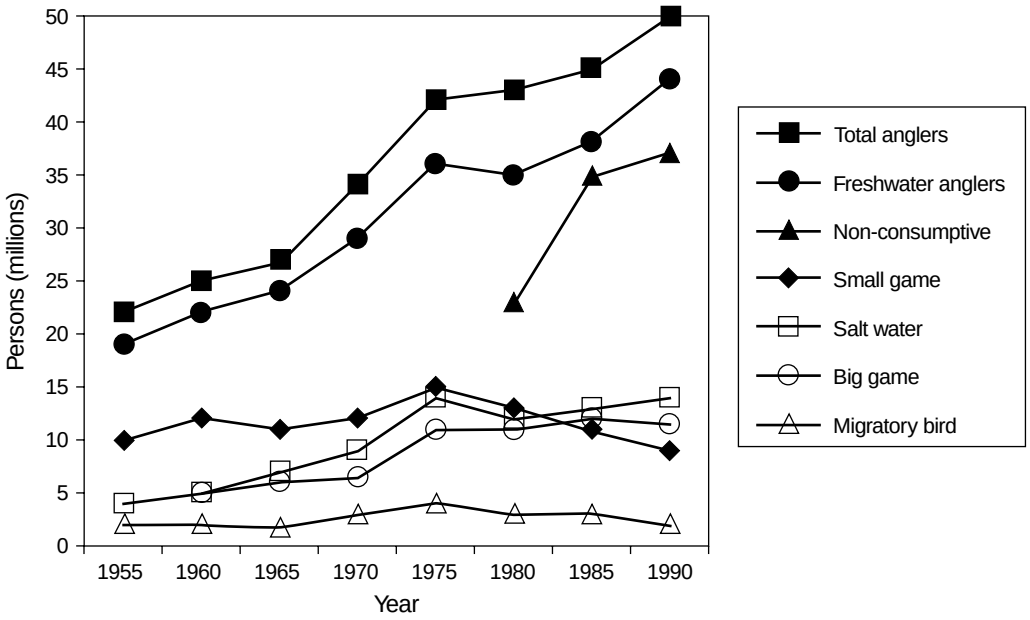


Figure 2 Growth in some forms of tourism and recreation in the US (after Knight and Gutzwiller, 1995)

wide variety of sites under the National Park Service, nine of the ten most frequently visited sites are parks and recreational areas (including parkways, Table 1); ecotourism is on the rise.

Marine and coastal tourism deserves special attention because of the fragility of these systems, because of a rapid increase in interest in the coastal landscapes, and because of the concentrations of people along coasts (Miller, 1993; Agardy, 1993). As human populations increase, these trends will continue, placing additional pressures on the plants and animals in these ecosystems. The pressures are two-fold: (1) an increase in human disturbance because increasing numbers of people will visit the parks, reserves, preserves, and other wild ecosystems; (2) increasing human populations and development will reduce available habitat elsewhere, leaving the protected ecosystems as islands.

TOURISM AND ECOSYSTEMS

There are many effects caused by tourists, whether they be ecotourists or otherwise, ranging from compaction of soil and increasing erosion, to destroying vegetation, altering the kinds and number of plants, displacing plants and animals, disturbing animals so that they abandon breeding or foraging habitats, increasing predation or piracy rates, and lowering survival and reproductive success. The kinds of disturbances caused by tourists vary by proximity, by species and the complexity of the ecosystem (Liddle and Scorgie, 1980; Knight and Gutzwiller, 1995; Liddle, 1997), as well as the type of tourists, and their locations (Mathieson and Wall, 1982; McNeely *et al.*, 1992; Ceballos-Lascurain, 1996).

The effects of tourists on ecosystems are multiple and complex (see summary in Table 2). Many of these effects are compounded by the great increases in the number of tourists visiting natural environments (McNeely *et al.*, 1992). There are effects for all major groups of species, and at all levels of organization. Whether people are walking over woodland trails, abandoned fields, sand dunes, or alpine tundra, soil is compacted, dramatically changing the soil microfauna and the ability of plants to germinate and flourish. Even walking over rocky outcroppings disturbs rock structure and lichens. Changes in soil compaction affect erosion, which in turn affects water dynamics, and ultimately stream flow, nutrient loss, and pollutant movement.

When soil is compacted, plants are affected as well, both by changes in their growing environment and by direct trampling by people, who also disperse a variety of native and exotic seeds among ecosystems. Changes in abundance and species composition of plants in turn alters all the animal communities dependent upon them. Thus, for all groups of organisms there are two types of tourist effects: direct affects that affect the organism itself (i.e., killing it, retarding its growth, and exposing it to predation, competition, or piracy), and indirect effects that include changing its microclimate, the plants or animals it depends upon, and its predators and competitors.

In turn, populations, community structure and function, and eventually ecosystems, are affected because of the cumulative effect of small-scale changes on soil and individual plants and animals (Figure 4). Some particularly

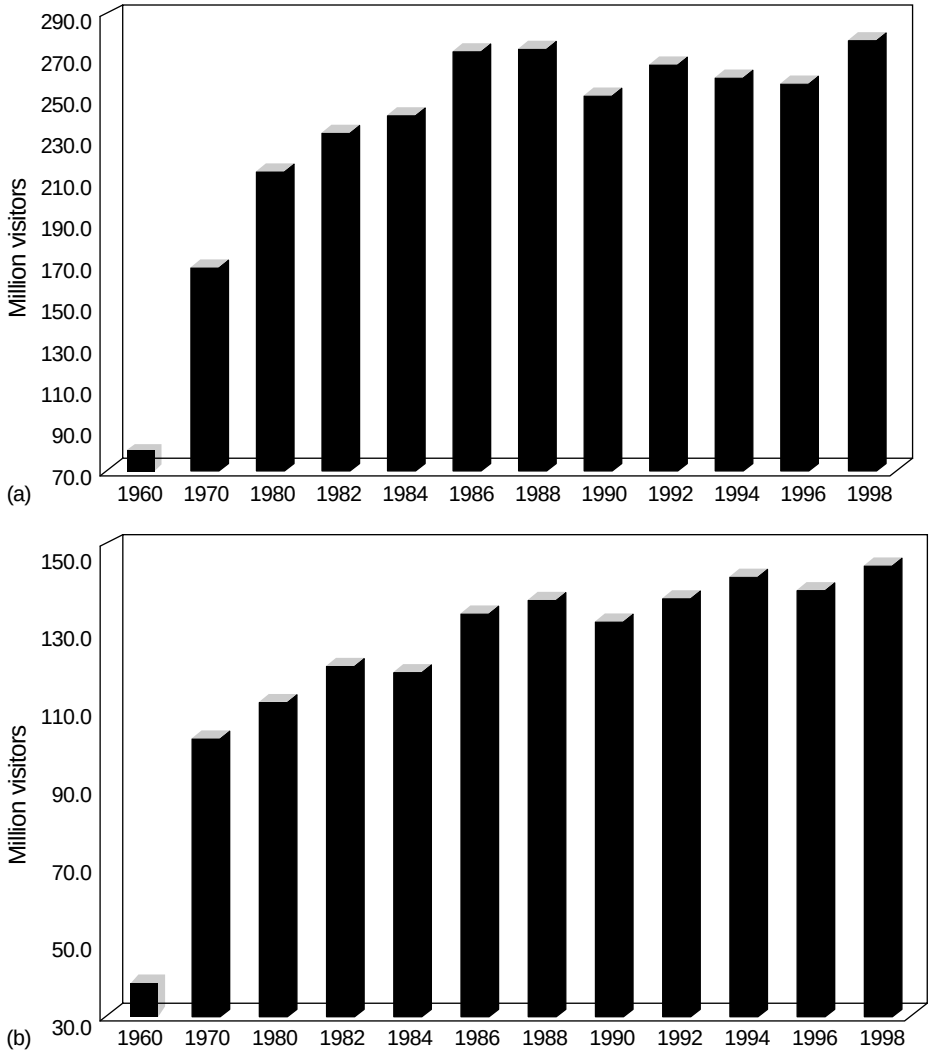


Figure 3 Number of visitors to all parks under the US National Park Service, including battlefields, historic parks, monuments, reserves and national parks (a), and number visiting the National Park sites (b)

sensitive species are affected merely by the presence of people, even though they are trying to be inconspicuous. For example, when predators, such as the big cats (lions and cheetahs), are surrounded by people, they hunt less frequently and have lower hunting success (McNeely *et al.*, 1992). Small or isolated ecosystems are particularly at risk, such as coral reefs, small islands and other delicate ecosystems (Mathieson and Wall, 1982). The feeding, breeding, and resting behavior of a wide diversity of animals are affected by increasing numbers of tourists (Mathieson and Wall, 1982).

There are also global effects, such as decreases in air quality because of air pollution from increased car exhausts, the destruction of coastlines, and the pollution of rivers (Mathieson and Wall, 1982). Waste materials generated by tourists can adversely affect water quality, and thereby ecosystems, in many different aquatic systems, including lakes, rivers and coasts; pathogens may also be introduced. Further, local geological features are often adversely affected by tourists who cause disruptions by merely walking there, or who remove materials (Mathieson and Wall, 1982). All of these changes affect both local and regional ecosystems.

For managers, the question then becomes how to reduce the effect of tourists, while still allowing ecotourists to have exciting and meaningful experiences. Reducing effects, however, requires detailed knowledge of how tourists affect a wide range of species, communities, and ecosystems.

ECOTOURISM, RECREATION AND BIRDS: A CASE STUDY

There are many types of ecotourists that affect and are affected by ecosystems and their component parts, but

Table 2 Possible effects on ecosystems by tourists moving through ecosystems

Area affected	Affects
Physical environment	Soil compaction Shifts in drainage erosion Increased addition of nutrients (from human waste and garbage) Shifts in depth of organic matter Changes in hydration
Litter, organic matter and soil invertebrates	Disruption of soil microfauna Changes in particle size Changes in earthworm population because of soil compaction Reduced soil anthropod numbers
Vegetation	Trampling, pruning, abrasion Shifts in species, biomass Damage to plant root Reduced root depth due to compaction
Reptiles	Direct killing of snakes to protect tourists Disturbances during egg-laying that lead to predation or sub-optimal selection of nest sites Lowered reproductive success due to compaction of soil over nests or hatchlings Inadvertent running over hatchlings Lights disorienting hatchlings/sea turtles and preventing them from reaching the sea Vehicles inadvertently running over reptiles
Birds	Decreases in time available for foraging Shifts in habitat Decreases in nesting Changes in nesting locations Shifts in time at the nest Leaving eggs exposed to thermal stress and predators Early abandonment of nest site by brood Changes in heart rate Frequent flushing from preferred resting sites Too rapid nest departure leaving eggs exposed to thermal stress or predators Increases in human predation of eggs and chicks Increased time away from colony site Lowered hatchling or fledgling rates Difficulties with hunters, prey
Mammals	Decreased foraging time Increased defense of young Disturbances during hibernation forcing relocation Withdrawing into sub-optimal habitat Running to exhaustion (from off-road vehicles or snow-mobiles) Changes in migration patterns Changes in escape routes Changes in heart rate Accidents while fleeing from people over rocky terrain
Community	Population declines Changes in predator population Lowered reproductive success and shifts among species
Ecosystem	Increased erosion of topsoil and destruction of soil and litter organisms Shifts in species composition Declines in species diversity Increases in invasive species

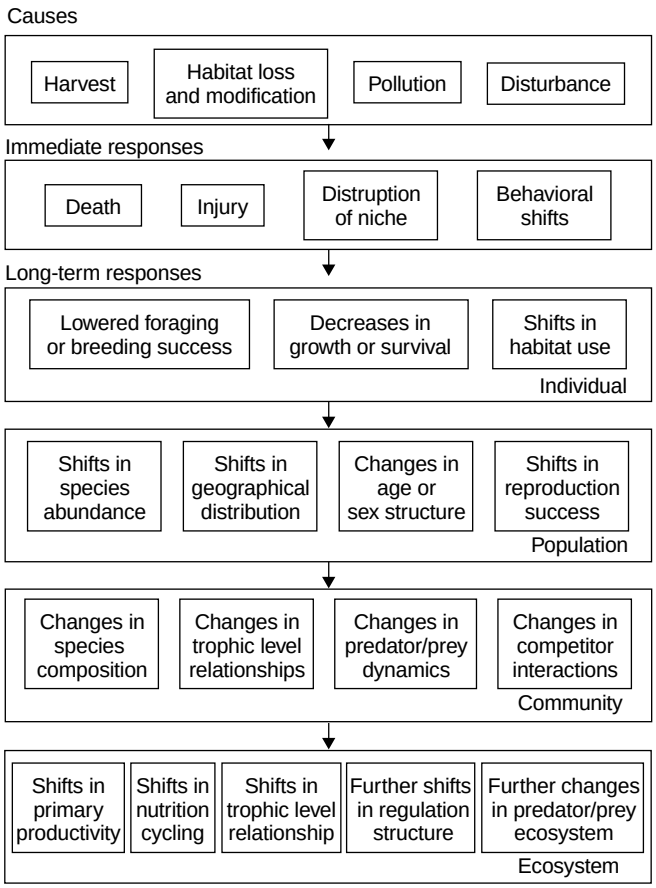


Figure 4 Model for the effects of tourism on organisms and ecosystems

basically they fall into two categories: people who specifically travel to observe or experience animals and plants in their natural habitats, and people who enter natural environments as part of some other pursuit, but who can affect ecosystems. The latter include a wide range of recreationists, such as walkers, joggers, hikers, campers, off-road vehicle users, and boaters. People who travel to see or experience natural habitats, and their plants and animals, include people who come to watch, photograph or study birds, butterflies, mammals or other natural spectacles. For ecotourists to receive maximum pleasure from their experience, the plants and animals must be in excellent condition. This includes high species diversity, unwary and visible animals, untrampled plants, and high numbers (massive herds of ungulates on the African plains, large numbers of nesting seabirds and basking iguanas on the Galapagos).

While the intent of people may differ between the two major classes of ecotourists, the overall effect on ecosystems does not. Twenty people walking through a fragile sand dune, whether to search for rare birds or to reach the beach, disrupt delicate dune vegetation and sand placement. Twenty people walking over the rocks of the Galapagos disrupt nesting seabirds or basking iguanas, regardless of whether they are photographing them, watching them, or merely walking to a fishing or swimming spot. Twenty people swimming over a coral reef, whether they are photographers, scientists, or merely divers watching fish, disrupt the normal rhythm of the sea creatures they are watching.

Because birds are generally diurnal, medium to large (that is, not microscopic), and live mostly in the same terrestrial space as people, they are enticing to ecotourists, and excellent subjects for scientific study. There is considerably more research on the effects of a wide range of human activities on birds than on all other groups of plants and animals. In this section the effects of people on birds are examined as a case study.

There are three major types of studies: (1) investigator effects; (2) experiments aimed at simulating ecotourists; and (3) observations of ecotourists around birds. Each will be discussed briefly because they illustrate how tourists and others affect birds.

Investigator Effects

Investigator disturbances were first noted for birds, such as seabirds and wading birds, that nest in large colonies, where a large proportion of the birds leave their nests when an investigator enters the colony. While scientists were interested in understanding how they affect birds so that they could reduce these effects, such information provided a basis for management programs.

The effects of investigators have been examined both experimentally and retrospectively, where the number of observers in a colony over a number of years is compared to some measure of effect (reproductive effect, colony size). For example, 21 years of data on Northern Fulmars (*Fulmarus glacialis*) indicated that hatching and breeding success was lower in years when 12 or more observers were present, compared to years with fewer investigators (Ollason and Dunnet, 1980). Other scientists have noted that birds will not establish territories and breeding colonies if people are present, some species desert nesting colonies if too many scientists walk through a colony (particularly during the early phases of nesting), and in some species, nesting density is lower in sections of the colony that are heavily visited (see Burger and Gochfeld, 1994).

During incubating and chick care, birds often take flight when there is an intrusion, exposing eggs or chicks to thermal stress and predation (Kury and Gochfeld, 1975; Roberts and Ralph, 1975; DesGranges and Reed, 1981), resulting in lowered hatching or fledging success (Hunt, 1972; Nettleship, 1975). Chicks can get lost and perish (Veen, 1977). Neighbors or marauders sometimes kill and eat chicks that wander from their territories as a result of human intrusion (Burger and Gochfeld, 1990, 1994). Some

parents and their brood abandon the nest site early when exposed to human disturbance, and move to other places farther away or where there is cover.

Many species of birds habituate (become accustomed) to the presence of investigators, who then cause little effect (Burger and Gochfeld, 1994; Brown and Morris, 1994), and disturbances can be reduced by providing cover for chicks (Quinn, 1984). Investigators also protect nesting colonies by their very presence, reducing vandalism, egging, and hunting (Haynes-Sutton, 1987).

This work demonstrated that investigators can have a negative effect, but that experience and knowledge can minimize the disturbances so that reproductive success is not reduced. Further, the factors affecting human disturbance for nesting birds were determined from this work, species (or reproductive stages) that were particularly vulnerable were identified, and methods of reducing human disturbance were described (reduced time in the colony, use of the same pathways where possible, walking slowly through any colony or nesting area, wearing dull clothing, and not looking directly at birds).

Experimental Studies

It soon became apparent that investigators could simulate the behavior of tourists and others approaching birds while they foraged or bred, which provided a paradigm for understanding how people affect birds, and a methodology for risk reduction and management. Experimental studies usually take the form of approaching birds to determine the distances at which they change behavior or fly; this has the advantage that the technique can be used on foraging and resting birds, as well as on nesting ones. Birds that are habituated learn to quickly distinguish whether a person is walking directly toward them or will pass a few meters away (Burger and Gochfeld, 1981).

For wading birds (herons, egrets) and seabirds that nest along coasts, distances that birds flush from their nests varies by species, but data suggest that nesting wading birds require people to stay at least 100 m away, and some seabirds require people to stay 200 m away to avoid undue disturbance to established colonies (Erwin, 1989). Greater distances are required to allow birds to establish new colonies. In a similar study in Florida, the distances that birds flushed from their nests when approached by people were used to derive management guidelines for pedestrian tourists and motor boats (Rodgers and Smith, 1994). Remarkably, the disturbance distances for wading birds were all about the same and for tern and skimmer colonies the distances (100 m) clustered at about 180 m. These data suggest that even with increasing human use of coastal ecosystems, preserving nesting colonies of birds will require maintaining these 180 m distances even in the face of increasing human density.

Even seabirds that nest on remote islands and apparently seem undisturbed by the presence of tourists respond physiologically by increasing their heart rate, by moving slightly away from advancing tourists, and by nesting farther from established paths. Boobies nesting on the Galapagos Islands, move away from tourists that pass as close as a meter from their nests, while birds that are over 6 m away seldom respond; albatross chicks on Midway Atoll in the northern Pacific Ocean likewise move back from approaching tourists (Burger and Gochfeld, 1999).

For songbirds nesting solitarily in shrubs and trees in forests, intrusions reduce singing (Gutzwiller *et al.*, 1994), which could affect their ability to acquire mates and defend territories. Response time even depends upon how many birds of the same species are nearby (Gutzwiller *et al.*, 1998). Such intrusions can result in nest abandonment and decreased nest attentiveness, leading to decreased reproductive success.

Birds that forage close to the ground are particularly vulnerable to predators, and can be easily disturbed by tourists or other people. When disturbed, birds either lose valuable foraging time, or fly away, where they are often forced to other habitats where food may be less plentiful and predators more abundant. Birds that forage alone or in flocks are equally vulnerable to human intrusions from walkers, joggers, birdwatchers or others moving toward them.

Many factors influence how close foraging or resting birds allow approach before altering their behavior or flying away. Birds in suburban areas where they have become habituated to people are less wary than those in rural areas (Cooke, 1980). Small birds often allow a closer approach by humans than do larger birds. Other factors that affect how birds respond include the number of people in a group, number of birds in the flock, time of day, and whether birds are migratory or not. In India, for example, where birds are not persecuted for religious reasons, resident birds allow far closer approach than do migrants from farther north where they are persecuted and hunted (Burger and Gochfeld, 1991).

Observational Studies of Tourists and Birds

Another way to assess the effects of tourists on birds is to design studies to specifically record how birds respond whenever tourists come by. The methodology is to quantify the amount of human disturbance in terms of type, frequency, proximity, and duration, and to document the response of birds to each type of disturbance. Such studies include understanding the effects on nesting and foraging birds. Tourist and recreational intrusions into forests can alter the abundance and distribution of songbirds, as well as changing species composition.

Intrusions into bird colonies in heavily populated coastal areas are difficult to document because most states protect

colonial species; many states post such colonies, and some fence them or employ wardens. However, observational studies have demonstrated that when picnickers disturb nesting herring gulls (*Larus argentatus*), hatching success declines by half (Hunt, 1972). Similarly, intrusions by tourists and others into Heerman s gull (*Larus heermanni*) colonies reduced fledging success by up to 75% (Anderson and Keith, 1980). Similar declines due to human intrusions have been documented for several other species. Colonial species disperse outward from the colony to forage, usually in coastal bays and estuaries or in the nearby ocean, rendering them less vulnerable to people at this time.

As more and more tourists visit beaches and other coastal regions throughout the world, the space available for nesting and foraging birds decreases and their exposure to human activities and noise increases. Shorebirds nesting solitarily on beaches are highly vulnerable to tourists because it is not easy to protect one nest here, another one 200 m away, and still another 500 m away. While fences can be placed around nests, it is far more difficult to protect these species while they and their chicks forage on beaches (Burger, 1994). For example, piping plovers (*Charadrius melodus*), a US endangered species that nest along the Atlantic Coast and along the Mississippi River in the US, nest on sandy beaches and forage at the water s edge, just where tourists want to walk, jog, swim, and sunbathe. Part of the plover s recent decline seems directly due to increases in the number of tourists along beaches, which restricts foraging time (Burger, 1994). The number of seconds that plovers are alert during foraging increases directly with the number of people on beaches (Figure 5). While adults can partially compensate for the lost foraging time by feeding at night, this is more difficult for chicks that have not mastered difficult tactile foraging skills. Further, continued disturbance while foraging scatters broods, and parents lose chicks that have run too far (Burger, 1991). Since the number of people is continuing to increase on these beaches, plovers will continue to lose foraging time, overall survival will decrease, and populations will continue to decline, unless managers exclude visitors during critical nesting periods.

With increasing numbers of people concentrating along coastal margins, birds nesting on estuarine islands are more exposed to tourists as well. Many types of recreational boats are being produced worldwide, especially personal watercraft (PWC) (Figure 6). PWC can drive into areas that are too shallow for other craft, destroying delicate submerged vegetation, disrupting bottom sediments, harming fish and crayfish nurseries, and disturbing nesting and foraging birds (Burger, 1998). Partly it is the noise that disrupts nesting birds, forcing them from their nests again and again, until they finally desert the colony site. As more and more people use PWCs, the pressure on ecosystems will increase, unless legislation restricts their use.

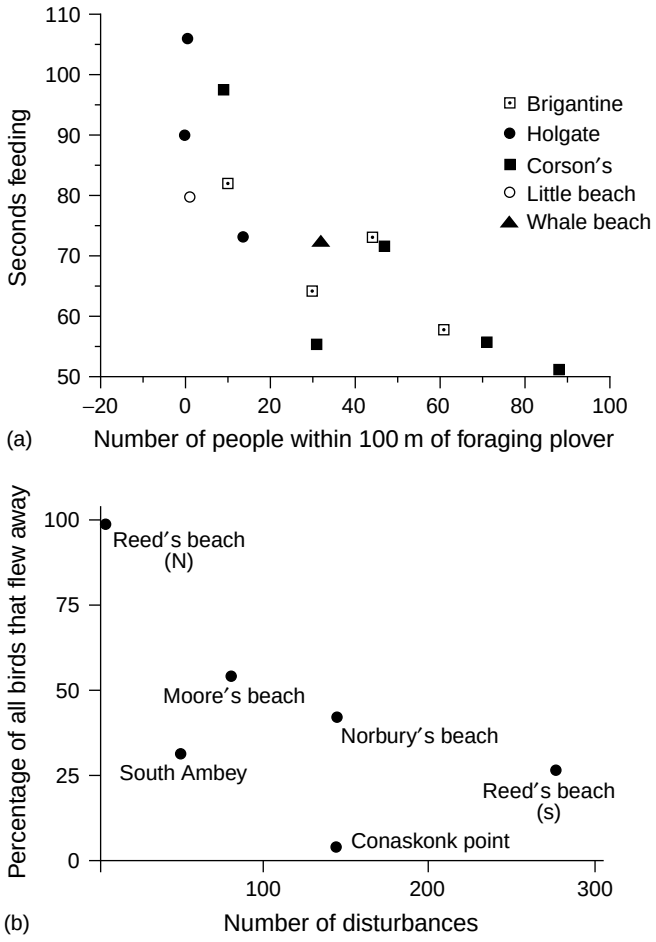


Figure 5 Dose–response relationship of the number of tourists and foraging behavior of (a) breeding piping plover (after Burger, 1994) and (b) migrant shorebirds (after Burger, 1986) (Cape May, NJ)

Migrant shorebirds are particularly vulnerable because they have very little time to lay down sufficient fat reserves to fly long distances, northward to breeding habitats or southward to overwintering grounds. Flocks of birds foraging along the shore are affected by the presence of tourists on the beaches by avoiding beaches with people, by moving from people (and their associated dogs) whenever they pass, of by ceasing to feed or guard their young when people pass by. The number of foraging shorebirds that fly away is directly related to the number of disturbances at Cape May, New Jersey (Figure 5, Burger, 1986; Burger *et al.*, 1995), a world famous birding spot where the number of tourists increases yearly. Further, when some members of the foraging flock take flight, other do as well, even though they are much farther away (Roberts and Evans, 1993). Tourists and others on migration staging areas can impact shorebird movements by moving shorebirds from preferred resting sites, and by displacing them from optimal foraging areas (Pfister *et al.*, 1992).

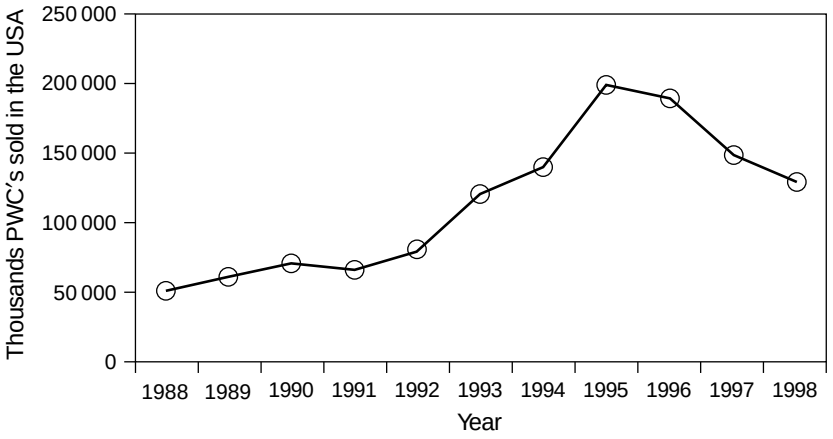


Figure 6 Sales of PWCs (personal water craft) in the US from 1987 to 1997 (after Burger, 1998)

Shorebirds and other species feeding on coastal beaches are particularly vulnerable to tourists because they are forced to move when people come by. However, even birds foraging in the water respond to tourists moving along dikes to watch them (Burger and Gochfeld, 1998). Gallinules (*Gallinula chloropus*), rails (*Porzana carolina*), ibises (*Plegadis falcinellus*) and herons (*Egretta caerulea*, *Egretta tricolor*) all spend less time foraging when people are near, and obtain less food than when there are no intrusions. Variations in the time they devote to foraging are a function of the number of people present, and the amount of noise made by people. Birds that are close to dikes when people pass usually swim away or fly a short distance before returning to foraging, while those that are farther from the tourists simply paddle away while continuing to feed. Such information is critical to managing tourists and birds because it means that where possible there should be a buffer between foraging birds and tourists, and decreasing noise levels is central to reducing the effects on foraging birds. It also suggests, however, that as the number of tourists increases, the problem will be exacerbated by the increasing noise that additional people, even those who are relatively quiet, will make.

Other birds that are vulnerable to disturbances by tourists are exotic, rare, or charismatic birds, such as birds of prey. Tourists frighten eagles off their nests, exposing developing eggs to cold stress, and human disturbance accounts for nearly all nest losses in some areas (Bocker and Ray, 1971). Ospreys (*Pandion haliaetus*) have abandoned nests after frequent disturbances by boaters, bird watchers, and photographers (Dunstan, 1968).

OTHER SPECIES

While the effects of tourism on birds have been extensively studied, there are fewer studies with other animals. Conservation of fish may largely reside in the hands of fisheries management, but many marine resources are affected by ecotourists. Marine sanctuaries are being developed in many places in the world, and direct concern for the effect of ecotourists on fish, sea turtles, seabirds, and marine mammals is being expressed. Yet there are few quantitative studies of effects.

Currently there is widespread concern about the effect of ecotourists on marine mammals, such as gray whales (*Eschrichiius glaucus*) in Southern California, humpbacks (*Megaptera* spp.) in Alaska and Hawaii, and Belugas (*Delphinapterus leucas*) in Alaska and the St. Lawrence River in Canada (Blane and Jackson, 1994). In many cases tourism boats get as close as possible, for as long as possible, and there is little information on the short- or long-term effects of these boats. Increasing pressures to visit species such as the northern elephant seal (*Mirounga angustirostris*) are resulting in management strategies such as limiting the number of visits, limiting the number of trails, limiting the time on trails, and enforcing non-disturbance policies (Magoon *et al.*, 1998).

There is almost no work on the effect of tourism on reptiles, except for the interest in sea turtles and Galapagos iguanas (Johnson *et al.*, 1996). Although watching sea turtles nest is an exciting experience, female sea turtles accompanied by a group of 15 people spend less time camouflaging their nests and return to the sea over a less direct route than unwatched turtles (Johnson *et al.*, 1996). Although reproductive success does not differ, ecotourism must be accompanied by protection of nests. The number of tourists interested in watching sea turtles nest is likely to increase. In the past decade there has been a 24-fold increase in visitations to Costa Rica s Tortuguero National Park. The park has responded with extensive tour guide training programs to ensure the protection of endangered sea turtles (Jacobson and Robles, 1992), but few data are available on the effects of ecotourists. Turtle watching in SE Asia has often been unregulated with adverse effects on turtle populations through overexploitation.

Experimental work in Costa Rica indicated that basking black iguanas (*Ctenosaura similis*) evaluate the risk from approaching people, discriminate whether they are walking directly toward them or will pass by, and respond accordingly (Burger *et al.*, 1991, 1992). A tourist approaching directly and looking at an iguana is the most threatening; partly the iguana response is due to eye size.

Population declines in wood turtles (*Clemmys insculpta*) in Connecticut over a 20 year period were directly correlated with increases in people in the surrounding communities, and in the number of recreational/tourist permits issued for turtle habitat (Garber and Burger, 1995). Remarkably, the recreational/tourist pressures were not great, but the increases in human presence nonetheless led to declines and shifts in the age distribution; there were no new recruits into the population. Such studies indicate that, for some species, even small numbers of recreationists and tourists can have devastating effects.

As with birds, land mammals suffer some of the same effects because of tourists (Liddle, 1997). Bears can be disturbed during their winter hibernation by skiers or others (Goodrich and Berger, 1994). Many mammals respond to increasing interactions with people by becoming wary and withdrawing. For example, the number of white-tailed deer (*Odocoilus virginianus*) observed from snowmobiles was lower on Saturdays and Sundays and was negatively correlated with the number of snowmobiles in St. Croix State Park, although the deer merely hid (Dorrance *et al.*, 1975). In contrast, caribou (*Rangifer tarandus*) change their migration route, avoiding areas where they were hunted by snowmobiles (Calef, 1976).

Like birds, some mammals habituate to the presence of tourists, and then can cause problems: when bears habituate they become dangerous, whereas when deer habituate they become nuisances, moving among habitation to consume garden and foundation plantings and spread Lyme disease. Some animals, such as foxes (*Vulpes vulpes*), habituate sufficiently to tourists and other people that they begin to live near human habitation although they still run from groups of noisy people. Habituation of some mammals, such as the large cats, can be deadly, as has happened in Africa when careless staff of lodges put out bait for large cats so that people can see them at close range.

While there are far fewer studies of the effects of ecotourists on groups of vertebrates other than birds, they indicate a full range of effects that are similar to those demonstrated for birds. They illustrate that there is much still to learn about the nature, magnitude, and permanence of tourist effects. These studies indicate that environmental management and planning are required to reduce the effects of tourists on the well-being, survival and reproduction of wild populations.

TOURISM AND ENVIRONMENTAL PLANNING

In 1980 the World Tourism Organization (WTO) recognized that there should be a close relationship between environmental planning and tourism, and that both natural and cultural resources must be conserved and protected (Inskeep, 1987). Such planning requires comprehensive environmental analysis of carrying capacity, development and application of local and regional strategies, planning for tourist facilities, maintaining the quality of these facilities and the ecosystems that attract the tourists, and monitoring the health of these ecosystems, as well as the tourist facilities. Planning and development of ecotourism should also include economic analyses of both direct and indirect effects of that tourism. What are the appropriate multipliers for the economic costs and benefits of tourism, on both the local culture and the local ecosystems?

Tourism may represent a significant threat to the environmental health of the very places that tourists want to visit. While many planners believe that ecotourism may provide an economic incentive to preserve habitats, more attention is devoted to planning for ecotourists than assessing the effects of those tourists on the ecosystems themselves. Many regions fail to take into account the fundamental issues surrounding tourism, particularly ecotourism (Table 4, Butler, 1991). While scientists may be determining the effects of ecotourists on plants and animals, this information may not be accessible to managers and planners.

Table 3 Arrivals (thousands) of tourists from abroad by region (after World Tourism Organization, 1998)

	1980	1993	1994	1995	1996
Africa	7329	18 605	19 009	20 117	21 553
Asia/Pacific	21 480	71 094	76 788	81 356	89 186
Europe	185 442	309 758	330 630	333 855	348 999
Middle East	7467	71 348	12 747	13 465	14 084
Central America/Caribbean	8238	15 184	16 092	16 641	16 885
North America	47 321	77 324	77 906	80 491	85 223
South America	5809	11 250	12 573	13 480	14 565
South Asia	2242	3511	3866	4200	4332

Table 4 Issues surrounding environmental planning, tourism, and ecosystem protection

Knowledge gaps	Dimension of tourism
	Potential growth of ecotourism
	Economic value of ecotourism
	Methods for containing ecotourists
	Data on effects of ecotourism[a]
	Data on carrying capacity for ecotourists
	Dose–response data[b]
	Role of education in reducing effects
	Methods of involving local peoples
	Methods of working in interdisciplinary teams
Technology gaps	Tools to determine sustainability
	Tools for maintaining restricted populations
	Methods for determining effects of tourism
	Methods for valuating ecotourism and biota
Philosophy gaps (agreement gaps)	Over desired level of ecotourism
	Over which types of tourism to encourage
	Over amount of wilderness to protect
	Over desired level of tourist development
	Over type of ecotourism[c]
	Over balancing industry/tourism[d]

[a] Effects of ecotourists on most groups of animals and plants have not been examined in detail.
[b] The relationship between the number of tourists and degree of effect has not been determined for many types of ecotourists or groups of organisms.
[c] For example, woods might be preserved for ecotourists interested in migrating birds, but old fields might be preserved for butterfly watchers.
[d] Is it better to cut down the Asian forests for timber, or preserve them for the tourist value of high species diversity.

Ecotourism has largely advanced on a piecemeal basis, but this will not be possible in the future because of increasing numbers of tourists, and increasing pressure on natural resources. Ecosystems and their component plant and animal communities must be protected or the stimulus for ecotourism will disappear. Studies of the type described above for birds are essential for a wide range of animals to assess the factors that allow tourists and animal populations to flourish side by side, even with increasing pressure from increasing numbers of tourists. Thus, both monitoring and research will play a more critical role as human populations and tourism increases. To plan for tourism within a healthy environment, interdisciplinary teams will be required (Grenier *et al.*, 1993, Table 4).

INCREASING GLOBAL TOURISM AND FUNCTIONING ECOSYSTEMS

The data presented above suggest that birds, as well as other animals, are vulnerable to increasing numbers of tourists during their breeding, foraging, and migrating phases. They respond both to features of the disturbance, as well as to their own constraints. Disturbance features that are important include the number of people, their activities, their noise levels, their location, and past behavior of humans (i.e., degree of hunting or persecution). Animal life history parameters that are important are flock, herd or colony size, activity (foraging, stage of breeding), location (on land, in trees, or in the water), location in the flock or herd, available cover, predators that are present, and physical stresses (e.g., thermal). While animals can habituate, there are some human activities that are not amenable to habituation (such as the noise of PWCs), and there are minimal distances that cannot be violated or the animals will cease their activities, losing valuable time when they could be feeding, engaged in breeding, or other equally important behaviors.

If the past is any indication of the future, tourism will continue to increase. Further, ecotourism, the form that has a direct effect on ecosystems, will also increase. Another complexity is the probability of significant climate change and rising sea level before the end of this century. This would modify, or even destroy some habitats, particularly along low-lying sea coasts. See for example, various articles in volume 1, and the overview paper: Intergovernmental Panel on Climate Change (IPCC): an Historical Review, Volume 4.

While direct habitat loss is clearly the greatest global threat to ecosystems, we also need to be concerned for the preservation of the plants and animals within the ecosystems that we choose to protect. There are a number of different types of protection available, ranging from national parks and reserves, where recreation and other human activities are often encouraged, to preserves and wilderness areas where human uses are severely restricted.

The feedback mechanisms between tourism and ecosystems are also extremely important. Tourists have the potential to affect the plants and animals within ecosystems in profound ways, as illustrated above. The tourists can affect the landscape itself. However, as tourists affect these animals and plants, they also can render the ecosystem less attractive to other tourists. This will in turn decrease the number of tourists that come to see that ecosystem, which may allow the system to recover. Thus, there is a very delicate balance between tourism and healthy ecosystems.

As more and more private land is developed, the pressures on public lands will increase, making the preservation of fragile and exciting ecosystems paramount. Understanding the relationships between the behavior of tourists and the survival and reproduction of plants and animals,

their communities and ecosystems, is critical to preserving functioning ecosystems for the future.

ACKNOWLEDGMENTS

I thank M Gochfeld for valuable comments on the manuscript, M McMahan for literature and web searches, R Ramos for graphics, and the Department of Energy for support while writing this manuscript (consortium for risk evaluation with stakeholder participation, AI#DE-FC01-95EW55084 DE-FG-26-OONT 40938).

REFERENCES

Agardy, M T (1993) Accommodating Ecotourism in Multiple Use Planning of Coastal and Marine Protected Areas, *Ocean Coastal Manage.*, 20, 219 239.

Anderson, D W and Keith, S O (1980) The human Influence on Seabird Nesting Success: Conservation Implications, *Biol. Conserv.*, 18, 65 80.

Blane, J M and Jackson, R (1994) The Impact of Ecotourism on the St. Lawrence Beluga Whales, *Environ. Conserv.*, 21, 267 269.

Bocker, E I and Ray, T D (1971) Golden Eagle Population Studies in the Southwest, *Condor*, 73, 463 467.

Brown, K M and Morris, R D (1994) The Influence of Investigator Disturbance on the Breeding Success of Ring-billed Gulls *(Larus delawarensis)*, *Colonial Waterbirds*, 17, 7 17.

Burger, J (1986) The Effect of Human Activity on Shorebirds in Two Coastal Bays in the Northeastern United States, *Environ. Conserv.*, 13, 123 129.

Burger, J (1991) Foraging Behavior and the Effect of Human Disturbance on the Piping Plover *(Charadrius melodus)*, *J. Coast. Res.*, 7, 39 52.

Burger, J (1994) The Effect of Human Disturbance on Foraging Behavior and Habitat Use in Piping Plover *(Charadrius melodus)*, *Estuaries*, 17, 695 701.

Burger, J (1998) Effects of Motorboats and Personal Watercraft on Flight Behavior Over a Colony of Common Terns, *Condor*, 100, 528 534.

Burger, J and Gochfeld, M (1981) Discrimination of the Threat of Direct Versus Tangential Approach to the Nest by Incubating Herring and Great Black-backed Gulls, *J. Comp. Physiol. Psychol.*, 95, 676 684.

Burger, J and Gochfeld, M (1990) *The Black Skimmer: Social Dynamics of a Colonial Species*, Columbia University Press, New York.

Burger, J and Gochfeld, M (1991) Human Distance and Birds: Tolerance and Response Distances of Resident and Migrant Species in India, *Environ. Conserv.*, 18, 158 165.

Burger, J and Gochfeld, M (1993) Tourism and Short-term Behavioral Responses of Nesting Masked, Red-footed, and Blue-footed Boobies in the Galapagos, *Environ. Conserv.*, 20, 255 259.

Burger, J and Gochfeld, M (1994) Predation and Effects of Humans on Island-nesting seabirds, in *Seabirds on Islands: Threats, Case Studies, and Action Plans*, eds D N Nettleship, J Burger, and M Gochfeld, Bird Life International, Cambridge, 39 67.

Burger, J and Gochfeld, M (1998) Effects of Ecotourists on Bird Behavior at Loxahatchee National Wildlife Refuge, Florida, *Environ. Conserv.*, 25, 13 21.

Burger, J and Gochfeld, M (1999) Role of Human Disturbance in Response Behavior of Laysan Albatross, *Diomedea immutabilis*, *Bird Behav.*, 13, 23 30.

Burger, J, Gochfeld, M, and Murray, B G (1991) Role of a Predator s Eye in Risk Perception by Basking Black Iguana *(Ctenosaura similis)*, *Anim. Behav.*, 42, 471 476.

Burger, J, Gochfeld, M, and Murray, B G (1992) Risk Discrimination of Eye Contact and Directness of Approach in Black Iguanas *(Ctenosaura similis)*, *J. Comp. Psychol.*, 106, 97 101.

Burger, J, Gochfeld, M, and Niles, L J (1995) Ecotourism and Birds in Coastal New Jersey: Contrasting Responses of Birds, Tourists, and Managers, *Environ. Conserv.*, 22, 56 65.

Butler, R W (1991) Tourism, Environment, and Sustainable Development, *Environ. Conserv.*, 18, 201 209.

Calef, G W (1976) Numbers Beyond Counting, Miles Beyond Measure, *Audubon*, 78, 42 61.

Ceballos-Lascurain, H (1996) *Tourism, Ecotourism and Protected Areas*, IUCN, Gland, Spain.

Cooke, A S (1980) Observations on How Close Certain Passerine Species will Tolerate an Approaching Human in Rural and Suburban Areas, *Biol. Conserv.*, 18, 85 88.

Cordell, H K, Herbert, N G, and Pandolfi, F (1999) The Growing Popularity of Birding in the United States, *Birding*, 31, 168 176.

DeGroot, R S (1983) Tourism and Conservation in the Galapagos Islands, *Biol. Conserv.*, 26, 291 300.

DesGranges, J L and Reed, A (1981) Disturbances and Control of Double-Crested Cormorants in Quebec, *Colonial Waterbirds*, 4, 12 19.

Dorrance, M J, Savage, P J, and Huff, D E (1975) Effects of Snowmobiles on White-tailed Deer, *J. Wildl. Manage.*, 39, 563 569.

Dunstan, T C (1968) The Biology of Osprey in Minnesota from 1963 1968, *Loon*, 40, 109 112.

Ehrenfeld, D (1992) The Business of Conservation, *Conserv. Biol.*, 6, 1 3.

Erwin, R M (1989) Responses to Human Intruders by Birds Nesting in Colonies: Experimental Results and Management Guidelines, *Colonial Waterbirds*, 12, 104 108.

Garber, S D and Burger, J (1995) A 20-year Study Documenting the Relationship Between Turtle Declines and Human Recreation, *Ecol. Appl.*, 5, 1151 1162.

Goodrich, J M and Berger, J (1994) Winter Recreation and Hibernating Black Bears Ursus Americanus, *Biol. Conserv.*, 67, 105 110.

Grenier, D, Kaae, B C, Miller, M L, and Mobley, R W (1993) Ecotourism, Landscape Architecture and Urban Planning, *Landscape Urban Plann.*, 25, 1 16.

Gutzwiller, K J, Wiedenmann, R T, Clements, D L, and Anderson, S H (1994) Effect of Human Intrusion on Song Occurrence and Singing Consistency in Subalpine Birds, *Auk*, 111, 28 37.

Gutzwiller, K J, Marcum, H A, Harvey, H B, Roth, D J, and Anderson, S H (1998) Bird Tolerance to Human Intrusion in Wyoming Montaine-forests, *Condor*, 100, 519 527.
Haynes-Sutton, A H (1987) Human Exploitation of Seabirds in Jamaica, *Biol. Conserv.*, 41, 92 124.
Hunt, G L (1972) Influence of Food Distribution and Human Disturbance on the Reproductive Success of Herring Gulls, *Ecology*, 53, 1051 1061.
Inskeep, E (1987) Environmental Planning and Tourism, *Ann. Tourism Res.*, 14, 118 135.
Jacobson, S K and Robles, R (1992) Ecotourism, Sustainable Development, and Conservation Education: Development of a Tour Guide Training Program in Tortuguero, Costa Rica, *Environ. Manage.*, 16, 701 713.
Johnson, S A, Bjorndal, K A, and Bolten, A B (1996) Effects of Organized Turtle Watch on Loggerhead (*Caretta caretta*) Nesting Behavior and Hatchling Production in Florida, *Conserv. Biol.*, 10, 570 577.
Knight, R L and Gutzwiller, K J (1995) *Wildlife and Recreationists: Coexistence Through Management and Research*, Island Press, Washington, DC.
Kury, C and Gochfeld, M (1975) Human Interference and Gull Predation in Cormorant Colonies, *Biol. Conserv.*, 8, 23 32.
Liddle, M (1997) *Recreation Ecology: the Ecological Impact of Outdoor Recreation and Ecotourism*, Chapman and Hall, New York.
Liddle, M J and Scorgie, H R A (1980) The Effects of Recreation on Freshwater Plants and Animals: a Review, *Biol. Conserv.*, 17, 183 206.
Magoon, O R, Converse, H, Baird, B, and Miller-Henson, M (1998) *Taking a Look at California's Ocean Resources: an Agenda for the Future*, Vol. 1, ASCE, Reston, VA.
Mathieson, A and Wall, G (1982) *Tourism: Economic, Physical and Social Impacts*, Longman, London.
McNeely, J A, Thorsell, J W, and Ceballos-Lascuratin, H (1992) *Guidelines: Development of National Parks and Protected Areas for Tourism*, IUCN, Madrid.
Miller, M L (1993) The Rise of Coastal and Marine Tourism, *Ocean Coastal Manage.*, 20, 181 199.
National Park Service (1999) Annual Visitors to United States National Parks, http//www2nature.nps.gov/stat/summary master.htm.
Nettleship, D N (1975) Breeding Success of the Common Puffin (*Fratercula arctica* (L)) on Different Habitats at Great Island, Newfoundland, *Ecol. Monogr.*, 42, 239 268.
Ollason, J C and Dunnet, G M (1980) Nest Failures in the Fulmar: the Effect of Observers, *J. Field Ornithol.*, 51, 39 54.
Ostrom, E, Burger, J, Field, C B, Norgaard, R B, and Policansky, D (1999) Revisiting the Commons: Local Lessons, Global Challenges, *Science*, 284, 278 282.
Pfister, C, Harrington, B A, and Lavine, M (1992) The Impact of Human Disturbance on Shorebirds at a Migration Staging Area, *Biol. Conserv.*, 60, 115 126.
Quinn, J S (1984) Egg Predation Reduced by Nest Covers During Researcher Activities in a Caspian Tern Colony, *Colonial Waterbirds*, 7, 149 151.
Reaka-Kudla, M L, Wilson, D E, and Wilson, E O (1997) *Biodiversity II*, Joseph Henry Press, Washington, DC.
Roberts, H C and Ralph, C J (1975) Effects of Human Disturbance on the Breeding Success of Gulls, *Condor*, 77, 495 499.
Roberts, G and Evans, P R (1993) Responses of Foraging Sanderlings to Human Approaches, *Behavior*, 126, 29 43.
Rogers, Jr, J A and Smith, H T (1994) Set-back Distances to Protect Nesting Bird Colonies from Human Disturbance in Florida, *Conserv. Biol.*, 9, 89 99.
Veen, J (1977) Functional and Causal Aspects of Nest Distribution in Colonies of the Sandwich Tern (*Sterna s. sandvicensis* Lath.), *Behavior*, 20 (Suppl.), 1 123.
Walker, T A (1991) Tourism Development and Environmental Limitations at Heron Island, Great Barrier Reef, *J. Environ. Manage.*, 33, 117 122.
Wiesenfeld, L P (1999) *World Almanac and Book of Facts 1999*, Primedia, Mahwah.
World Tourism Organization (1998) Yearbook of Tourism Statistics, World Tourism Organization, Madrid.
World Tourism Organization (1999) Global Tourism Growth, http//hubcap.clemson.edu/-prtm/wtotourinfo.html.

Tourism as a Global Driving Force for Environmental Change

Poh Poh Wong
National University of Singapore, Singapore

In the last 50 years, the movement of people within and between nations in pursuit of leisure has been on an unprecedented scale. Restricted in the past to only the wealthy, tourism is now enjoyed by the masses, made possible by a decrease in the real cost of travel, a higher standard of living and available leisure time. The business of providing and marketing services and facilities for pleasure travelers is now the world's largest and most diverse industry. Tourism, itself, consists of diverse groups of industries covering the scale from multinational corporations to individually run enterprises. A wide range of tourist activities can be identi ed - visiting friends, beach holidays, coach tours, country walks, museum visits. Despite its growing complexity and scale, it represents an activity certainly well recognized, and has distinct impacts on the environment and society.

Tourism has become one of the most significant forces in environmental change in the latter half of the 20th century. It causes environmental changes affecting air, land and

water, particularly through its tendency to concentrate on certain areas such as coasts, mountains, wilderness areas, urban areas and small islands. The environmental impacts can be positive, e.g., the enhancement of local beautiful vistas, but often the impacts are negative, e.g., the inability of beach resorts to cope with the high demand by tourists during the summer months.

Tourism s response to an increasing awareness of environment and the need to reduce the environmental degradation has taken several directions. One response is *ecotourism* or nature-based tourism, which has become an important instrument for nature conservation. Another approach is to ensure that tourism development adopts proper environmental measures, for example, carrying out environmental impact assessment before implementation of a project. All the measures taken are towards making tourism a less negative force on environmental change and achieving sustainable tourism development.

However, measures to protect the environment have not become widespread, and the future scenario of environmental change by tourism is far from positive. Travel is expected to increase rapidly, leading to more greenhouse gas emissions by cars and aircraft. As more land is used for tourism development, additional degradation is to be expected. Plants and animals continue to be under threat. Fragile ecosystems such as sandy beaches, coral reefs and mangroves will come under increasing pressure. Tourism s constant search for new areas, for example, the Antarctic, and the rise of *adventure tourism* will put the environment on the defensive.

GLOBAL TOURISM

As defined by the World Tourism Organization (WTO), a tourist is a temporary visitor staying at least 24 hours in another country for pleasure or business. Tourism is now the largest industry in the world, constituting the largest item in world trade and the largest number employed. It is also one of the fastest growing industries in the world, averaging nearly 10% per annum from 1950 1970, although it has slowed down since then (Figure 1). For 1997, the

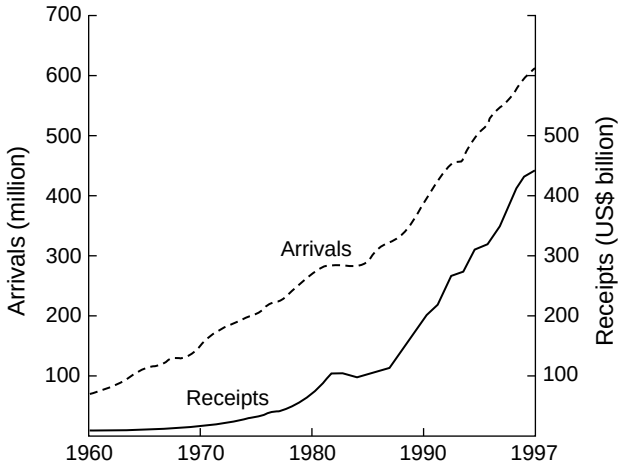


Figure 1 World tourist arrivals and receipts, 1960–1997. (Data source: WTO, 1998)

preliminary figures were 612 million tourists generating an estimated US$444 billion (excluding international transport). By 2010, tourism is projected to reach more than one billion tourists generating US$1500 billion (WTO, 1998). Its economic importance has grown more rapidly than the number of tourists. It had registered double-digit growth from 1950s except in the 1990s. As a group of economic activities, tourism is the world s largest industry in terms of employment. According to the World Travel and Tourism Council, 230 million or about 10% of the world s workforce are employed in the tourism sector. This is projected to increase to 328 million in 2010.

Tourism is unevenly distributed and highly concentrated in certain regions and countries. WTO, the most important intergovernmental body for the tourist industry, divides the world into six tourist regions: Africa, the Americas, East Asia/Pacific, Europe, the Middle East and South Asia. Europe and the Americas account for the lion s share of the world s tourism but are decreasing in favor of East Asia/Pacific and Africa. East Asia/Pacific has more than doubled its share since 1980 (Table 1). The remaining regions, each averaging a few per cent, have increased marginally.

Table 1 Regional share of world tourism, 1960, 1980, 1996

	Arrivals (per cent)			Receipts (per cent)		
Region	1960	1980	1996	1960	1980	1996
World	100.0	100.0	100.0	100.0	100.0	100.0
Africa	1.08	2.57	3.62	3.54	2.57	1.92
Americas	24.10	21.51	19.61	35.71	24.11	26.01
E Asia/Pacific	1.02	7.53	14.99	2.84	8.28	18.75
Europe	72.64	64.99	58.67	56.56	60.28	50.63
Middle East	0.91	2.62	2.37	0.84	3.29	1.78
South Asia	0.26	0.79	0.73	0.51	1.46	0.90

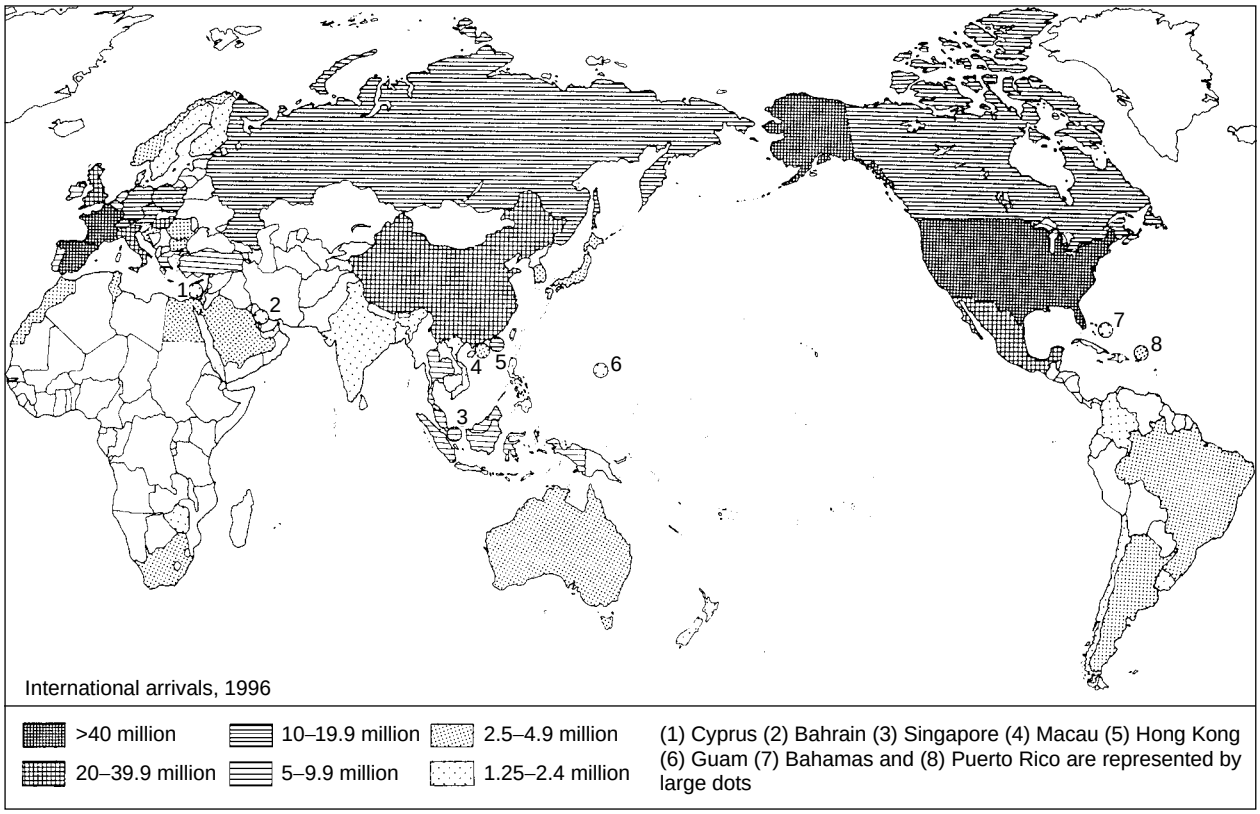


Figure 2 World tourism, 1996. (Data source: WTO, 1998)

Table 2 World's top tourism destinations, 1996

Rank 1996 (1980)	Country	Arrivals, 1996 (millions)	% of World's arrivals, 1996
1 (1)	France	62.41	10.49
2 (2)	United States	46.49	7.82
3 (3)	Spain	40.54	6.82
4 (4)	Italy	32.85	5.52
5 (7)	United Kingdom	25.29	4.25
6 (18)	China	22.77	3.83
7 (8)	Mexico	21.41	3.60
8 (10)	Hungary	20.67	3.48
9 (13)	Poland	19.41	3.26
10 (6)	Canada	17.33	2.91
11 (5)	Austria	17.09	2.87
12 (•)	Czech Republic	17.00	2.86
13 (9)	Germany	15.21	2.56
14 (•)	Russian Federation	14.59	2.45
15 (28)	Hong Kong SAR	11.70	1.97

A more quantitative idea of the volume of global tourism can be obtained by looking at individual countries. Tourism remains largely an activity between developed countries with the developing countries accounting for a 30% share of global arrivals. There is a high concentration of international tourism within certain developed and developing countries. Outside Europe and North America, the countries receiving more than 5 million arrivals in 1996 are limited to China, Hong Kong, Mexico and some countries in Southeast Asia (Figure 2). This picture is further supported by the top tourism destinations (Table 2), the top tourism spenders (Table 3) and the largest flows between generating and receiving countries. The most significant change has been China which has developed rapidly as a tourist attraction since its opening to the outside world in the late 1970s. The biggest spenders are

Table 3 World's top tourism spenders, 1996

Rank 1996 (1980)	Country	Tourism expenditure, 1996 (US$ billions)	Tourism expenditure, 1996, % of World
1 (1)	Germany	50.82	13.40
2 (2)	United States	48.74	12.86
3 (6)	Japan	37.04	9.77
4 (3)	United Kingdom	25.45	6.71
5 (4)	France	17.75	4.68
6 (13)	Italy	15.52	4.09
7 (10)	Austria	11.81	3.12
8 (5)	Netherlands	11.37	3.00
9 (9)	Canada	11.09	2.93
10 (•)	Russian Federation	10.72	2.83
11 (8)	Belgium	9.90	2.61
12 (12)	Switzerland	7.48	1.97
13 (37)	South Korea	6.96	1.84
14 (24)	Taiwan	6.49	1.71
15 (21)	Sweden	6.44	1.70

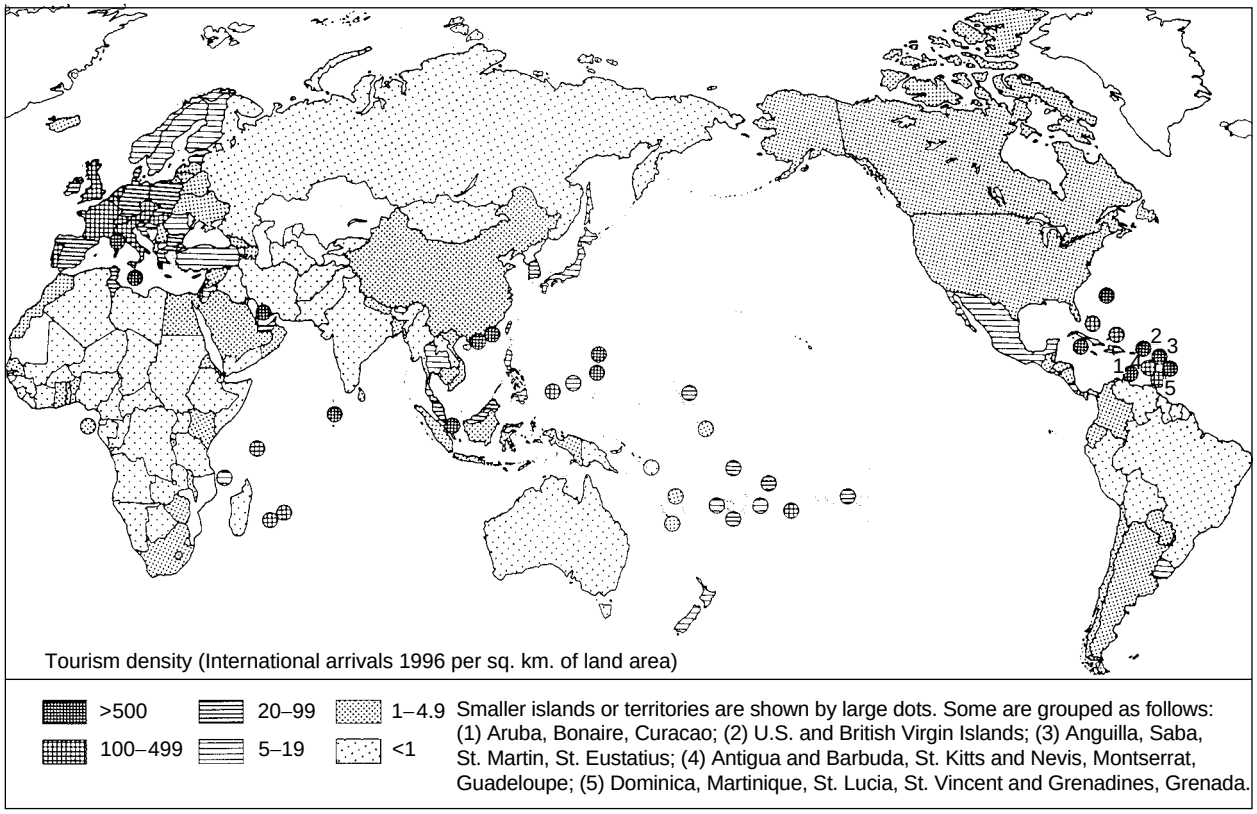


Figure 3 World tourism intensity, 1996. (Data source: WTO, 1998)

from the developed countries, joined by South Korea and Taiwan.

Available WTO data for 1996 and 1995 also show well-developed global patterns of flows between generating and receiving countries. The largest flows are across the Canada United States border. Except for the Mexico flow to the United States, the other largest flows are between neighboring countries of Europe, with Italy, France and Spain being the major recipients. More than 3/4 of international travel is short-haul and often between developed countries in Europe, North America and Asia. Most arrivals are intra-regional rather than inter-regional flows.

The picture of global tourism is markedly differently in some cases, when tourism intensity is compared to the land area of the various countries (Figure 3). Many island states of the Caribbean, Micronesia and the Indian Ocean, and

city-states are heavily frequented by tourists. Those in the medium-sized range include the European countries. The large countries of more than 1/2 million km^2 have lower densities.

Domestic tourism or travel entirely within a single country, may be more important than international tourism, as it involves about 10 times more people than the international trade. Up to 80% of the population of industrialized countries is involved in domestic tourism. Domestic tourists outnumber those traveling abroad in countries of large size (e.g., US) or geographical diversity (e.g., France). The potential for further growth of domestic tourism is in the developing regions, e.g., Southeast Asia.

Several factors and processes make tourism a global force. The most significant has been the technological innovation in global transport. Before 1945, travel was mainly by trains, boats or coaches. The airplane has made distant destinations tourist attractions. The Mexican resort of Acapulco was probably the first international resort to depend primarily on airborne tourists. After 1945, the airplane made possible the merging of the pleasure peripheries of different continents to global tourism (Turner and Ash, 1976). The Europeans have the Mediterranean, the Americans have the Caribbean and the Japanese are in the Far East. Mass tourism by air transport produced several distinct tourist flows: the annual flood of North Americans to Europe and the Caribbean, the Europeans to the Mediterranean resorts particularly in Spain, and the Japanese to other parts of East and Southeast Asia.

The introduction of wide-bodied jets in the 1970s reduced the price of air travel and increased affordability of tour packages. Overall, travel costs have declined 40% in real terms (after adjusting for inflation) in the past 40 years. The decrease in real cost of travel has been accompanied by rising disposable incomes and longer holidays for workers. Modern tourism has become particularly attractive, as there is a high level of professionalism in making travel safe, comfortable and enjoyable. A wide range of facilities and services is available, new travel products are being created and new destinations developed to increase the volume of tourist traffic.

ENVIRONMENTAL CHANGES

The realization that tourism is a portent force in environmental change is fairly recent (Table 4). In the 1950s, there was little concern about the environment when tourism was just beginning to develop rapidly. Conservation that started earlier in the century was limited to specific environmentally outstanding areas which themselves were attractions for tourists. Tourism grew rapidly in the 1960s and concern about its negative environmental impacts started in the 1970s. In the early 1980s, the WTO published reports on carrying capacity, saturation and management while tourism growth extended to more exotic areas. By the 1990s the response of tourism to environmental concerns has been in two major ways, ecotourism and sustainable tourism development (Hudman, 1991).

The environmental change by tourism is underpinned by the fact that the environment is and remains the foundation of the tourism industry. Mathieson and Wall (1982, 97) wrote:

> In the absence of an attractive environment, there would be little tourism. Ranging from the basic attractions of sun, sea and sand to the undoubted appeal of historic sites and structures, the environment is the foundation of the tourist industry.

The environment can mean both natural and built environments. Whether natural or man-made, it is inevitably changed or modified by the tourism process, as tourism is an environmentally dependent industry. Many impacts are related to the construction of general infrastructure and tourism infrastructure itself.

Environmental change by tourism is influenced by tourism s spatial concentration in specific types of environment, such as coastal areas, mountain and wilderness areas, rural areas, urban areas and small islands. Each ecosystem has its inherent environmental characteristics (Thuens, 1993). Environmental change is dependent on several

Table 4 Environmental concerns and tourism's response

Decade	Environment	Tourism
1950s	Enjoy and use	Age of exploration and involvement. Beginning of mass tourism
1960s	Awareness, public involvement and protests	Development, great growth. Elements of environment as attractions
1970s	Institutionalization. Creation of Environmental Protection Agency in USA. Concern over visual pollution, air and water	Growth and achievement. Marketing. Academic studies on impacts
1980s	Concern for toxic substances in the environment. Acid rain, global warming, ozone layer	Expanding world markets and technological advancements
1990s	Deforestation, desertification, climate change, global impacts	Ecotourism. Sustainable development

factors, such as the resilience of the ecosystem, the intensity of tourist use, the time factor and the transformational character of the tourist development (Cohen, 1978). Generally, environmental stresses increase with increasing volume and concentrations of tourists. One can expect little impact from an individual tourist but when the numbers are large and the tourist site is overused, negative impacts will occur.

To date, the verdict on the environment impact of tourism is that it is detrimental. The consensus is that the net economic impact is usually beneficial, but the evidence on environmental and social impacts shows more conflicts than benefits. The environmental impacts of tourism are known to have been overwhelmingly negative, based on well-documented examples from developed and developing countries. In some cases, the causal links to tourism may be debated, but harmful environmental impacts have occurred and the concern seems justified. Some environmental benefits are brought by tourism, such as in the conservation of some areas or environmental enhancement, especially in the re-development of derelict and redundant waterside areas.

As the impacts of tourism on the environment are varied and complex, there are several ways of classifying them. One simple approach differentiates direct impacts from indirect impacts; the former arise from the presence of tourists and the latter from the infrastructure created for tourism. Tourist impacts can be graded in terms of degree or intensity of impacts, such as the pressure on physical resources, the damage on the natural and the built environment, and the destruction or loss of land, ecosystems and built environment (Cazes, 1992). It is useful to distinguish impacts on the natural environment from impacts on the built environment. This has been incorporated into the classification of adverse environmental impacts of tourism (Table 5).

Table 5 Adverse impacts of tourism on the environment

Natural environment

(A) Pressure on natural resources
  1 Energy depletion
  2 Water supply
  3 Land use
  4 Soil erosion

(B) Harm to wildlife/habitat and biodiversity loss
  1 Trampling and clearance of vegetation
  2 Loss of forest cover
  3 Disturbance to wildlife
  4 Damage to coral reefs
  5 Damage to species

(C) Pollution
  1 Air pollution
  2 Untreated wastewater
  3 Solid waste and litter
  4 Noise pollution

Built environment

(A) Over-urbanization
(B) Destruction of natural landscapes

Given the scale and continued growth of global tourism, its impacts are wide ranging and are of increasing concern. The following sub-sections look at tourism as a force in global environmental change using examples from Asia, Africa, Europe and Australasia.

Air Pollution and Global Warming

Tourism contributes to air pollution, as cars, buses and air travel are the most important modes of transport of tourists. Within resorts and tourist sites, pollution from tourist vehicles is likely to be significant. In cities, tour buses, taxis and recreational driving are the main causes.

Air travel contributes to air pollution in the stratosphere. It is estimated that 60% or more of all air travel is for tourism, which is thus responsible for a significant share of air pollution by aviation (Genot, 1997). Airplanes are major sources of greenhouse gases particularly carbon dioxide, water vapor and nitrogen oxides, which are known to be harmful to the ozone layer. Although air travel accounts for 3% of nitrogen oxide emissions arising from human activities, as much as 25 30% of the total nitrogen oxide in the upper troposphere could be due to aircraft emissions (*see* Aircraft Emissions, Volume 3).

Land Resources Consumption and Land Degradation

In modern times, tourism contributes to land resources consumption and land degradation in various ways. To begin, tourism has a voracious need for land for airports, hotels, leisure complexes, golf courses and infrastructure. In extreme cases, this has led to the displacement of local communities and the loss of access to public areas for the local population. The problem is most acute in developing countries as the local population loses land resources that could sustain them, for example, the Kenyan fishing communities lost their livelihood as fishing area was designated for marine parks. The Indonesian farmer is displaced as coastal land in Bali and Lombok is taken up for resort development.

In Southeast Asia, the development of golf courses for tourists, especially the Japanese tourists, is an issue. Golf is increasingly a popular game and a status symbol in Japan. Due to high membership costs and lack of land, most of the estimated 15 million Japanese golfers never have a chance to play except on driving ranges. While detailed statistics are not available, it is estimated that 1.5 10% of Japanese travelers play golf while on overseas trips (Cameron, 1996). Many golf courses have been developed in Southeast Asia to cater to Japanese tourists, who numbered 2.49 million

in 1996 (*see also* Geomorphological Change: Landscape Modi cation for Recreation, Volume 3).

Golf courses create a number of environmental problems. The clearing of land sometimes leads to the removal of forests and forest reserves. Large amounts of water, chemicals and often non-native greenery create artificial ecosystems. In Malaysia, over 160 courses were created mainly by clearing rainforest areas. Each course uses enough water to meet the needs of 100 Malaysian farming families. In Indonesia, farming land and nature reserves were replaced by some 90 golf courses which annually consume more water, pesticides and fungicides than all the country s farms combined. Thailand s 200 courses are a drain on water resources (Cameron, 1996). Fertilizers and pesticides are used on golf courses and posed a threat to water resources. If water quality is reduced through pesticides and fertilizer run-off, it is almost impossible to restore the land back to its natural state.

Tourism is known to be a major consumer of water and competes with other land uses. For example, water is required for the tourists, swimming pools, the washing of bed linen and floors, watering gardens and golf courses. Tourists consume many more times *per capita* consumption of water than do the local populations. This situation becomes critical when a destination suffers from water shortage (seasonal or chronic) and water has to be diverted from other users, such as the farmers and inhabitants. In the Mediterranean, water is a major issue as tourism peaks during the dry season. On one Mediterranean island, each tourist uses enough water for eight local inhabitants (Genot, 1997). In Barbados, the tourist per capita consumption of water is 6 10 fold of that of the local population (Gajraj, 1989).

Tourism can contribute to soil erosion, which is usually associated with the removal of protective vegetation cover. Tourists moving constantly over vegetated areas create several impacts. Careless scrambling can cause the rate of movement downslope of unconsolidated chalk-rubble-soil, in a temperate woodland, to rise from its usual 5 cm year^{-1} to 30 cm year^{-1}. Normally, leaf litter layers help to prevent soil erosion and increase moisture. However, a newly opened nature trail in a temperate woodland, used by 8000 people, can lose 50% of its litter in a week (Ceballos-Lascuráin, 1996). Litter is broken up and easily dispersed by the wind.

Soil compaction occurs from trampling by people and the hooves of ponies and pack animals. This occurs along trails and is very common at campsites. It leads to impeded drainage in the soil, increased runoff and finally soil erosion. The problem is accelerated in areas where wind and water can remove soil easily. For example, in tropical areas with heavy rainfall, rills develop rapidly and form gullies. Sand-dune systems are particularly vulnerable to wind erosion when the vegetation cover is broken.

Mountain areas have been traditionally used for agriculture, forestry and reservoir development. In recent years, tourism development in mountainous areas competes with such uses and has an impact on the environment. Skiing is particularly popular in the mountain areas of developed countries. In Europe, the Alps have approximately 40 000 ski runs and 14 000 ski lifts (Ward *et al.*, 1994). Visually, the resulting steel pylons, overhead cables, lifts and tows are unattractive. More seriously, the construction of ski resorts means trees are removed, slopes reshaped and there may be an increase in avalanches. During seasons of insufficient snow, ski resorts depend on snow machines using compressed air and water sprayed at high pressure when temperatures are below freezing. This results in a serious drain on the local water supply. Through the use of snow machines, the skiing season is extended which has an impact on wildlife habitat and flora, increased damage to slopes and accelerated soil erosion. It is estimated that a cubic meter of alpine pasture can hold 90 liters of water. If dug up and prepared for skiing, it retains only 40 liters of water (Ward *et al.*, 1994). The higher proportion of unabsorbed water increases the likelihood of soil erosion and flooding.

In recent years, trekking tourism has been increasingly popular in the mountain areas of some developing countries and its impacts have, to date, been detrimental (Box 1). Trekking in more gentle or tropical highland environments shows similar environmental results. In the hills of Northern Thailand, trekking started in the early 1960s and has become an established industry with more than 200 outlets handling this activity in Chiang Mai, the major city in northern Thailand. It involves about 100 000 trekkers per annum. With an average of 3 nights per trek, they generate a minimum of 300 000 trekker nights per year (Dearden, 1993). Overall, the direct economic benefits are small but significant in an area of poverty. However, there are serious environmental impacts, particularly deforestation of the bamboo, which is used to make rafts for trekkers. An estimated 400 000 stems of bamboo are required annually (Dearden, 1993). Trekking has an impact on water quality, arising from sanitation and from solid waste disposal, as few sanitary facilities are available in the villages.

Biodiversity Loss and Habitat Destruction

According to various estimates, the global extinction rate of all species is 3 130 species daily or 2 8% of all species from 1990 2015. Plants are disappearing at the rate of 2000 species per year (German Federal Agency for Nature Conservation, 1997). Although tourism contributes marginally to overall biodiversity loss, it assumes a more significant force when it is located in or near fragile habitats or habitats of high biodiversity, such as coastal ecosystems, coral reefs, ocean islands, rainforests and mountain ecosystems (*see* Tourism and Ecosystems, Volume 3).

Box 1 Impacts of tourism in the Himalayas

Nepal is a landlocked country of 141 415 km^2 in the Himalayan mountains. Four-fifths of its 33 million people are in agricultural activities. Although opened to mass tourism in 1951, the country did not develop rapidly as a destination for adventure travel until the 1970s. The number of tourists increased from 4000 in 1960 to more than 100 000 in 1977, 200 000 in 1986, and about 365 000 in 1995 (Weaver, 1998). The tourists are attracted to the country's natural resources and typically 20% go for mountain trekking and another 60% for a combination of trekking, jungle safaris, river rafting and ethnic tourism. Although tourism has become the largest foreign-exchange source, it has also caused severe environmental degradation.

Environmental problems caused by tourism include deforestation, solid waste management, sanitation and water supply, overgrazing (by pack animals) and erosion of paths and trails. Large amounts of litter and garbage are left along mountain trails and campsites. The Everest trail is often described as the Mount Everest garbage trail, and this is true of many other trails. Non-biodegradable materials are the biggest concern but even bio-degradable materials take years to decompose due to the low temperatures. An average group of 15 people trekking for 10 days can generate 15 kg of non-biodegradable and non-burnable garbage (Rai and Sundriyal, 1997).

The Nepalese rely virtually on wood as the only source of energy for cooking and heating. This has resulted in deforestation, which has been aggravated by tourism. It is estimated that each trekker in the 1970s used 106 kg of firewood for a 15-day trek (Weaver, 1998). Fuelwood consumption is itself influenced by several critical factors, such as the type of tourist activity, length of trip, number of support staff and the use of fuelwood substitutes. The deforestation problem became extremely serious by the early 1980s and the government instituted regulations for trekkers to carry fuelwood, so as to eliminate gathering of fuelwood along the trekking routes. However, this regulation did not apply to the tourist lodges, and the problem is thus transferred to them. The result can be devastating. Depending on the concentration of tourists, a tourist lodge in one area was using 220 kg of fuelwood per day or ten times that of a local household (Gurung, 1991). Deforestation leads to soil erosion and landslides. One hectare of cleared forest is estimated to lose 30–75 tons of soil annually. The lodges are also poorly equipped for the disposal of human waste and litter, leading to water pollution and odor problems.

As tourism in Nepal remains a much-needed economic alternative, measures have been taken to lessen the environmental impact. The Annapura Conservation Area Project was set up in 1986 to address problems of nature conservation, human development and tourism management. It has a program of integrated conservation and development implemented in stages. The deforestation is tackled by many measures, including alternative energy sources, energy-saving technologies, reforestation and community forestry management. The tourist carrying capacity is not necessarily fixed but is a function of technology and resources available to tackle the impact. If tourism is managed properly, it is possible for an area to recover (Gurung and De Coursey, 1994).

When tourism development has taken place in or near to some areas of high levels of biodiversity, its impacts lead to irreversible damage, including species extinction as in the case of the destruction of tropical rainforests and coral reefs. The impacts from tourism can set off a chain reaction with results seemingly out of proportion to the initial impact. For example, the collection of the triton mollusc (Giant Triton) by recreational divers has been linked to the spread of crown-of-thorns starfish, which has destroyed large areas of coral reefs in the Pacific.

Tourism can change the percentage cover, species diversity, growth rate, age structure and habitat diversity of vegetation. The impacts range from the complete removal of natural vegetation to other impacts not immediately evident. The greatest damage to vegetation cover is in the initial use of the area where fragile species disappear and recovery is dominated by other species. The reproduction rate is reduced in trampled areas. Innocent acts by tourists, such as flower picking and plant collecting, can lead to a loss of individual species.

Camping and hiking contribute to further negative impacts on the vegetation cover. Vegetation is removed for walking sticks, tent poles and firewood. One study found that the quantity of small timber declined significantly up to 100 metres from the side of the road and this is attributed to campers collecting wood (Ceballos-Lascuráin, 1996). An additional danger is fire. In general, the damage from trampling is greater than that caused by camping. The creation of rubbish dumps favors soil alteration and the introduction of exotic species. For example, Australian country roads are lined with apple and pear trees originating from fruit thrown from cars. Clothing or horse riding, particularly on certain trails or camping sites in national parks, also carries seeds.

Tourism impacts wildlife in several ways, such as loss or changes to habitats, impacts on natural patterns of feeding, breeding and predator-prey relationships (Mathieson and Wall, 1982). In particular, the consumption of wildlife by tourists has a harmful effect. For example, the demand for seafood has a severe impact on local fisheries and threatens wildlife population in marine parks. In the Caribbean, the spiny lobster and conch shell have been much reduced by the demand from hotels (Jackson, 1986).

The market for souvenirs has led to an over-exploitation of a number of now endangered species. The demand for tortoiseshell is a threat to the loggerhead turtle in Sri Lanka. Black coral for jewelry is widely in demand in the Caribbean and the species is now confined to areas where

collection is difficult. Tourism is largely responsible for growing demand for other wildlife souvenirs such as gorilla hands for ashtrays, fur from endangered cats, reptile skins for boots, belts and bags, ivory and horn, many of which are being poached from protected areas. While the Convention on the International Trade in Endangered Species puts a ban on international trade of listed species of animals and plants threatened with extinction, it is ineffective on such trade within countries.

An idea of the impact of tourist vehicles on large animals can be seen in the Maasai Mara National Reserve of Kenya. The tourist vehicles usually approach close to the cheetah, leopard and the lion and encircle them for photography. The angle (arc) of encirclement can vary from a quadrant (90°) to a full circle (360°). An increase in encirclement results in increased levels of walking by cheetahs and lions, particularly between 90° and 270°. Presumably, at less than 270° animals can still escape, but at 360°, escape is extremely difficult. The animals walk away when tourists approach them too closely. With distances of under 5 meters, they are greatly disturbed. The more sensitive cheetah moves off at distances of more than 20 m. Overall, the vehicles interfere with animal activities, particularly their mobility, which restrict their search for prey, mating and seeking cover (Ceballos-Lascuráin, 1996).

Coral Reef Degradation

Coral reefs are one of most severely endangered habitats. They are rapidly being destroyed by activities such as dredging, mining and over-fishing. This occurs in the areas near to, and of, high population density in coastal southeast Asia, around the Indian Ocean and tropical islands. In many of these countries, coral mining is a traditional activity as the material is used for construction and production of lime. Local fishers use destructive methods of fishing, such as the use of dynamite and poison. Compared to barrier reefs and atolls, fringing reefs have suffered the greatest damage from overexploitation. The most serious cause comes from sedimentation resulting from human activities including tourism. Other causes include unsound agricultural and forestry practices, poor management of watersheds, exploitation of mangroves, construction and dumping of mining tailings and effluents (*see* Deforestation in Historic Times, Volume 3). In Sri Lanka, the major portions of reefs along the west and south coast have been destroyed. In the Philippines, almost 70% of the reefs are seriously damaged and about 80% or more in certain areas of eastern Indonesia (Lundin and Lindén, 1993). In Indonesia, coral reef destruction has accelerated dramatically in recent years and is reaching epidemic proportions as the economic and social crisis deepens in the country.

Coral reefs are an attraction for tourists and often provide the major impetus for resort development in the tropics. At the same time, tourism has been responsible for their degradation, arising from the direct effects of diving and snorkeling, to infilling and sedimentation caused by tourism construction (Table 6). Structural damage occurs when reef flats and adjacent coral reefs are used for the construction of landing strips on small islands. Construction of hotels and condominiums may seriously disturb the sandy beaches and lead to their erosion (Maragos, 1993). Tourism activities have increased the rate of destruction not only through construction, but also through removal of corals for sale as souvenirs, damage from diving, boating, snorkeling, and other recreational activities. Also, the quality of water, which is critical to coral growth, is affected by pesticide or fertilizer runoff from golf courses and untreated wastewater from hotels and supporting facilities (Dubinsky and Stambler, 1996).

Tourism and recreation have already damaged the coral reefs in Kenya, Tanzania, Mozambique, Madagascar, Reunion, Mauritius, Comores, Seychelles and Sri Lanka in the Indian Ocean and in Malaysia, Thailand, Philippines and Indonesia in Southeast Asia. Tourism is considered by the World Conservation Union to be a problem for coral reefs in Thailand, Philippines, Indonesia, Kenya and some of the oceanic islands, particularly the Maldives and Seychelles. Often, the main impact on coral reefs is not from the activities of tourists. The significant impacts are from tourist-generated pollution and coastal development (Box 2). Tourist facilities are often concentrated in narrow coastal strips and semi-enclosed bays. Coastal structures such as breakwaters and jetties lead to an alteration in water circulation, beach erosion and sedimentation. Coastal pollution results from the discharge of poorly treated wastewater into the sea.

Table 6 Impacts of tourism on coral reefs

Construction
Building continuous to reefs
Breakwater construction
Resort construction
Outfall construction
Extraction of coral sand and limestone
Recreational activities
Reef walking
Boulder moving
Anchor damage
Diver damage
Small boat damage
Fishing
Collection of corals and shells
Pollution
Herbicides and pesticides
Antifouling paints and agents
Spilled petroleum
Sewage and detergents
Sediments

Box 2 Effects of tourism on coral reefs in the Red Sea

The Red Sea has unusually warm water. Helped by favorable physical conditions, the reefs at the Gulf of Aqaba typify the northern Red Sea coral reef, consisting of a narrow reef flat and a reef slope plunging almost vertically into deep water. The calm sea, clear water and rich coral reefs form the basis of the Red Sea's rapid development as a tourist destination shared by three countries, Egypt, Jordan and Israel.

Approximately 19, 50 and 100% of the Red Sea coast of Egypt, Jordan and Israel, respectively, are developed for coastal tourism that already has had several effects on the coral reefs (Hawkins and Roberts, 1994). At Hurghada, the inshore reefs are seriously damaged and only the offshore reefs 4–16 km from the coast support the diving industry. Both fringing reefs and those farther offshore are still healthy at Sharm-el-Sheikh. Several impacts arising from construction and infrastructure have occurred: coastal infilling; sedimentation from construction (large amounts due to dust blown by wind and not through rain wash, which occurs in other tropical regions); sewage disposal, irrigation, desalination and solid waste (through inadequately processed wastewater or seepage from hotel gardens which use treated wastewater); effects on local fisheries (demand for fresh seafood, marine-curio trade); direct impacts of diving and snorkeling.

The intensity of use of reefs is expected to increase as Israel plans for a further 43% increase in coastal tourism, Jordan 100% and Egypt a massive 11-fold expansion. More than half of the planned expansion in Egypt is concentrated in the established resorts of Hurghada and Sharm-el-Sheikh. The reef ecosystem does not have the resilience of a sandy ecosystem and cannot support the same density of tourists as sandy coasts. Attempts have been made to work out the carrying capacity of reefs for diving. Hawkins and Roberts (1994) suggested that Sharm-el-Sheikh can support 10 000–15 000 dives per site per year without serious degradation compared to 4000–6000 dives per site per year in Bonaire, Netherlands Antilles. However, with the planned increase in tourism development, Sharm-el-Sheikh will have 40 000 dives per site per year despite doubling the number of sites and spreading out the dives. In Israel, increased pressure is on the existing reefs and development has proceeded south to encroach on the Eilat Coral Reserve.

If tourism development in the Red Sea proceeds as projected, widespread reef degradation is expected. To some extent, this can be ameliorated by other measures, such as limiting pollution to reefs, preventing sewage discharge and nutrient-rich effluents, and reducing the proportion of visitors who dive. On a more positive note, industrial development has been restricted by tourism as reefs have an economic value to be preserved. In the long run, tourist development would still determine the fate of the coral reefs: the option of a long-term controlled development or short-term financial gains.

The damage on corals is dependent on the fragility of the corals and percentage cover by living corals. Branching or leafy corals are easier to damage than rounded corals and visible damage is more conspicuous on reefs with a high percentage coral cover. For example, a ship s anchor and chain can damage up to 20 square meters of corals on a good reef. Consumptive pursuits, collecting and spear fishing, have selectively depleted certain species, such as spider conch, murex, cone and cowrie shell, pearl oyster, seafan, black coral, grouper and lobster. The crown-of-thorns starfish usually devastates large areas, and the reefs take up to 10 40 years to recover, provided no fresh infestation occurs.

Coast and Small Island Degradation

On a global scale, tourism has the most impact on coasts and small islands as they are highly favored by tourists (Figure 4). Both coasts and small islands are particularly fragile areas, partly because they are comprised of closely linked ecosystems such as the beach, coral reefs and mangroves (Knight *et al.*, 1997). Also, other land uses and activities are attracted to these ecosystems and altogether they combine to have an impact greater than tourism. Tourism itself has been responsible for many distinctive urbanized coasts.

The Mediterranean Sea was the world s first major coastal tourism area, developed for Europeans to escape the northern winter. It has experienced severe pressure from urban growth, much of it associated with tourism. About 4400 km^2 of land around the Mediterranean Sea is for tourist accommodation and supporting structures. Water pollution is of increasing concern particularly for the older tourist resorts. Data from the Mediterranean Action Plan showed that the average water consumption for the international and the domestic tourist was 250 and 150 $l\,person^{-1}\,day^{-1}$, respectively, and 60% of this was for sewage disposal. Each tourist generated 0.9 $kg\,day^{-1}$ of garbage (German Federal Agency for Nature Conservation, 1997). Tourism development along the coasts of Spain, France, Italy, Greece, Yugoslavia and Tunisia contribute to large quantities of waste materials. Pathogens are introduced into the sea through raw or inadequately treated sewage. The addition of nutrients has hastened the process of eutrophication and results in algal blooms and red tides. Diseases such as cholera, typhoid, viral hepatitis and dysentery can be transmitted through contaminated seafood. There is a growing concern in the European Community at the environmental pressure generated by tourism expansion (*see* Land Degradation in the Mediterranean, Volume 3; Blue Plan, Volume 4; Mediterranean Sea, Volume 4).

The fragility of the coastal systems and its impact by tourism is being repeated in the developing countries, especially those in the tropics. Blessed with beaches, year-round sunshine and the added attraction of coral reefs,

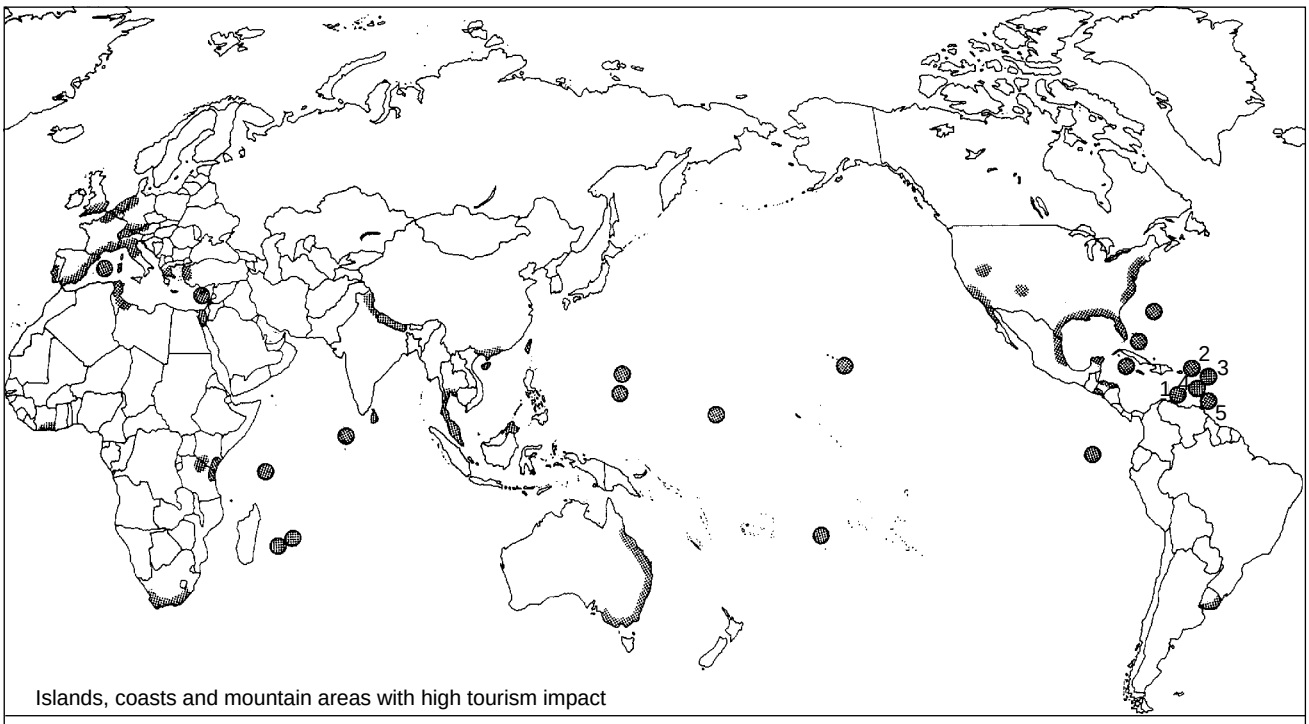


Figure 4 Areas of high environmental change by tourism

many tropical developing countries and small islands have become coastal honey pots. They attract tourists throughout the year from America, Europe, Australasia and Japan, each group replacing the other, depending on the season. Tourism development has largely proceeded spontaneously in these countries and has given rise to many problems. The tropical marine-coastal ecosystem with its coral reefs and coastal forests is one of most severely affected by tourism development (Figure 5). Coastal water pollution is a common problem where tourism development has been unplanned and wastewater facilities are inadequate (Box 3).

Tourism is a significant economic activity on many islands in the Caribbean, the Mediterranean, the Pacific Ocean, the Indian Ocean and Southeast Asia (Figure 4). Some have been developed exclusively for tourism. Extreme cases are the island resorts or ıles-hotels of the Maldives in the Indian Ocean. They are basically coral islands of atolls and are actively sought for tourism development because of their insularity, an easily recognizable and marketable image and a unique combination of land and sea. They have become very accessible with modern transportation. In most cases, the small population rules out domestic tourism and they are dependent on international tourists. But they are very vulnerable because of limited size and less diversified resources. Solid waste production *per capita* is higher than the Maldivian average, up to 16.5 kg of solid waste per visitor per week (Brown *et al.*, 1997). Traditionally, the water is obtained from groundwater wells and rain-fed cisterns. Water availability and quality decline as tourist demand increases. The freshwater lens is reduced and the mean salinity of the depleted water source has more than doubled. The resorts are dependent on desalinated water and imported bottled water. Also, rising sea level is a long-term threat to the island resorts. Erosion has occurred and in some cases, inappropriately designed coastal structures have been deployed.

It should be noted that the impacts of tourism on island ecosystems could be less than from other industries, such as agriculture, fishing or mining. Such islands also suffer from pressures of overpopulation, inappropriate urban development and land clearance. Tourism is sometimes an easy target for blame, as it is often owned by foreigners.

ECOTOURISM AND SUSTAINABLE TOURISM DEVELOPMENT

The global concern of the negative environmental impacts of tourism became evident in the 1980s. Tourism could not be considered purely in terms of economic gains without attention to the environment. Further environmental change

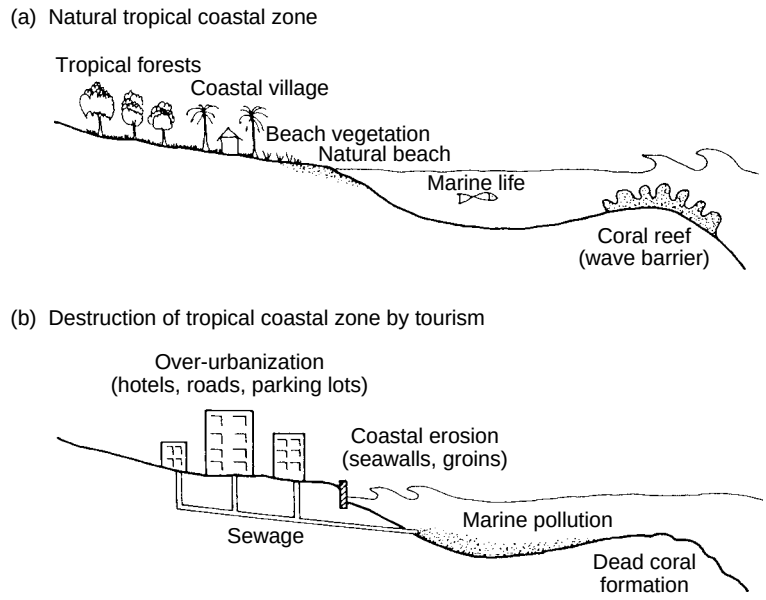


Figure 5 Environmental change to a tropical coastal zone by tourism

Box 3 Water pollution at Pattaya

Pattaya is located 140 km southeast of Bangkok, Thailand. It grew rapidly from a fishing village to a major tourist resort in less than three decades. Its first international hotel was constructed in 1964 and significant tourism development started in the late 1960s. The number of visitors increased from 259 000 in 1972 to 600 000 in 1982, and 2.45 million in 1990 (Smith, 1992).

As coastal land was developed for tourism, land farther inland came under development. The traditional coastal land use was replaced by a more typical strip or ribbon development in which the physical form of the resort was linear and parallel to the coast. The resort developed into a concrete jungle of buildings crammed tightly together, alternating with roads parallel to the coast. Traffic congestion, noise and overcrowding were common. One 300-m stretch with bars, nightclubs and massage parlors known as 'The Strip' gives Pattaya its unsavory image.

As Pattaya developed its tourism, there was a demand for infrastructure, particularly for water supply and wastewater treatment. Water supply was a problem. Water was trucked to the resort until the supply mains reached Pattaya in 1981. However, the increasing demand for water as the resort expanded has resulted in water supply deficiencies since 1989. Water shortages also occur during the dry season.

By the mid-1970s, the seawater was polluted by resort-generated wastewater as well as effluent from the tapioca factories in the Pattaya hinterland. Two wastewater treatment plants were established in 1984 and 1986. The total wastewater of Pattaya more than doubled from 36 500 m^3 in 1986 to 76 000 m^3 in 1990, but the plants could handle only 13% of the total before discharging the rest into the sea. In 1989, sea pollution became critical and the water was declared unsafe for swimming. In 1991 and 1992, measurements of total coliform count were 10–24 times above the recreational water quality standard set at 1000 MPN 100 ml^{-1} (Charoenca, 1993). By 1992, marine pollution was cited as the major cause of Pattaya's declining number of tourists.

Clearly in Pattaya, the major cause of marine pollution is the overload of treatment plants and the insufficient collection system. More than 200 hotels and many condominiums do not comply with official environmental standards. Some major hotels with water-treatment facilities keep them switched off to reduce costs. Also businesses and restaurants contribute to the wastewater problem.

The urban growth of Pattaya has produced other negative environmental impacts, such as the removal of tree cover, replaced by a mix of buildings, roads and overhead wires and a maze of signboards. The wetlands and rivers are filled or converted into open sewers. The beach is littered by solid waste, and wildlife habitats have been destroyed. The coral reefs have been removed or destroyed by recreational activities and wastewater. Beach erosion occurs because of the unnecessary use or poor design of coastal structures. Garbage collection is inadequate to deal with the solid waste generated within the resort.

In light of declining tourist arrivals to Pattaya, the government has set up a program to reduce water pollution and improve the environment. Attention is given to improve the capacity of the existing wastewater treatment plants and to provide two new plants that would service North Pattaya and South Pattaya, respectively. Other measures are to improve traffic conditions and the beach (Wong, 1998).

by tourism will be influenced by two major developments: development of ecotourism as an alternative to mass tourism and sustainable tourism development (*see* Agenda 21, Volume 4; Sustainable Development, Volume 4).

Ecotourism had its origins in the late 1970s, growing from the concern about environmental impacts and the search for an alternative to mass tourism. It is defined as environmentally responsible travel to relatively undisturbed natural areas to enjoy nature and local culture; it has a low visitor impact and provides for socio-economic involvement of the local population (Ceballos-Lascuráin, 1996). Ecotourism is supposed to promote conservation. But it has quickly become the buzz word to sell a variety of products, such as ecotour, ecotravel, ecologically sensitive adventures, eco(ad)ventures, ecocruise, ecosafari, which may not be committed to conservation at all (Goodwin, 1996).

Due to a lack of reliable data, various estimates put the ecotourism industry at US$10 billion in 1989 to a liberal US$200 billion in 1990 and as high as 20 25% of leisure travel. Anecdotal evidence from companies offering ecotours indicates a growth rate of about 20 25% per year (Weaver, 1998). Ecotourism is expected to grow rapidly in the next 20 years.

The growth of ecotourism and related forms has serious implications on rates of environmental change and this depends on the scale of ecotourism and a number of other factors. Weaver (1998) sees ecotourism operating at several scales: comprehensive, where the entire country is oriented to deliberate ecotourism; diversionary, where isolated ecotourism sites are near and augment mass-tourism products; and regional, which involves the adoption of deliberate ecotourism at a subnational scale, such as in the mountains or wilderness. The pace of environmental change depends on the relative development of ecotourism at these three scales.

The future development of ecotourism also depends on a number of factors.

One significant factor is the extent of protected areas. Ecotourism can become a threat to protected areas and ecosystems that are already under threat from other pressures (Goodwin, 1996). There is not enough formal protection for natural habitats that are being destroyed by lumbering, agriculture and other activities incompatible with the maintenance of a natural state. Only 4.6% of Africa, 4.4% of Asia and 6.4% of South America are formally protected. Often the nature parks are not well protected. Ameliorating measures, such as buffer zones (Costa Rica) and the maintenance of adjacent natural habitats along a common frontier (Tanzania, Kenya) are required. The measures taken to protect the environment must include better co-ordination of land use, the involvement of local population and integration with other economic sectors.

Another factor is local community involvement. Generally, it is vital to include the local population in decision-making and implementation for moral, conservational, economic and practical reasons. Tourist organizations, including government and international agencies and NGOs are important in regulatory and facilitating roles. Lastly, political and social stability are essential, as ecotourism cannot be divorced from the broader socio-political context. For example, in Rwanda, successful local ecotourism based on the mountain gorilla collapsed suddenly in the wake of social and political upheaval.

Sustainable tourism development has become a desirable goal with the spread of the concept of sustainable development. It can be defined as tourism meeting the needs of present tourists and hosts while protecting and enhancing the opportunity for the future . The concept is clear but is difficult to put into practice and there may be an inherent contradiction. It is not possible to create culturally and environmentally benign forms of tourism, when tourism is essentially structured upon exploitation of these very resources.

Sustainable tourism and ecotourism are not synonyms as many forms of ecotourism may not be sustainable, and if ecotourism is to contribute to sustainable development, then careful planning and management will be required. Griffen and Boele (1997) concluded that there is no universally applicable model for sustainable tourism development at the national level. Much depends on the expectations of the tourism sector, the destination s own political and administrative structures, differing standards of economic development and socio-cultural characteristics. Tourism generates changes, many of which are detrimental. One must be realistic in acknowledging that the problem is complex, and any solution has to be based on the necessity of accommodating further growth.

One approach is to look at the innovations related to environmental issues that could reduce the negative environmental impacts of tourism. According to Buckley (1996), much of it can be found in the research of environmental science, engineering, information systems and management. The innovations should cover not only product innovations, but also process innovations (e.g., energy saving, water saving, better waste management) and management innovations (e.g., procedures to motivate staff, local inhabitants and tourists into acting responsibly on environmental issues) (Hjalager, 1996) (*see* EA (Environmental Assessment), Volume 4). For fragile ecosystems such as a coast, a more interdisciplinary approach is required.

> The work of the architects, engineers and hotel accountants needs to be complemented by social scientists studies of tourist demand, use and satisfaction, by more detailed environmental examination from coastal geomorphologists, biogeographers, ecologists and other natural scientists and by research from environmental and resource economists.
>
> (Pearce and Kirk, 1986)

Integrated Coastal Management can be used for coastal ecosystems and for reefs facing multiple threats (White *et al.*, 1997; Wong, 1998).

SCENARIOS FOR ENVIRONMENT CHANGE

Ecotourism and sustainable tourism development have yet to make a substantial impact to redress the environmental change wrought by mass tourism. Many areas of the earth and several ecosystems, especially mountain areas (*see* Tourism: Climate Change and Tourist Resorts, Volume 3), coasts and small islands are still facing serious environmental change by tourism at present or in future (Figure 4). Based on these areas and the projected growth of tourism, some possible scenarios for environmental change caused by tourism can be suggested.

1. The coastal systems will continue to be assaulted by tourism as coastal tourism remains the most important type. This is especially true in many developing countries, which have the natural resources of sea and sand. Coastal tourism is likely to proliferate without adequate measures to protect the environment. By 2025, 95% of the Mediterranean coast will be urbanized, supporting 500 million people and 200 million tourists with 150 million cars (Gajraj, 1988). While there is a plan to alleviate the problems in the Mediterranean, which has experienced some of the worst of coastal tourism development, other coastal areas may be severely affected before similar plans are implemented.
2. Many natural ecosystems will continue to bear the impact of increasing numbers of tourists. The extent of change will depend very much on the carrying capacity, the resilience of terrestrial ecosystems and the implementation of management strategies to lessen the negative impacts. The adoption of environment-friendly technologies can also help to reduce the negative impacts.
3. Many studies show that tourist attractions follow a product-life cycle, going through stages of discovery/exploration, involvement, development, consolidation, stagnation and finally, decline (Butler, 1980). All new tourist areas, including forested and mountainous areas face this prospect. Of international concern is the continent of Antarctic, which had attracted more than 40 cruises with 5600 tourists in the 1991 1992 season (Gartner, 1996). The increasing number of cruise ships and tourists poses a threat to the fragile polar environment.

CONCLUSION

In the latter half of the 20th century, mass tourism, or people moving within and between countries in pursuit of leisure, was on an unprecedented scale. This has wrought changes to the environment, many of which have been negative. This was not fully realized until the 1980s. Some measures have subsequently been taken to prevent and minimize the environmental degradation.

The concern and response of tourism to environmental change has crystallized around ecotourism and sustainable tourism. New ways in technology and management can help to achieve sustainable tourism development. But both have yet to make a substantial impact.

At the same time, there are still many areas and fragile ecosystems under threat from tourism. Environmental change will spread to new tourist areas especially in the developing countries of Asia and Eastern Europe.

REFERENCES

Brown, K, Turner, R K, Hameed, H, and Bateman, I (1997) Environmental Carrying Capacity and Tourism Development in the Maldives and Nepal, *Environ. Conserv.*, 24, 316 352.

Buckley, R (1996) Sustainable Tourism: Technical Issues and Information Needs, *Ann. Tourism Res.*, 23, 925 928.

Butler, R (1980) The Concept of a Tourist Area Cycle of Evolution: Implications for Management of Resources, *Can. Geogr.*, 24, 5 12.

Cameron, O (1996) Japan and South-East Asia s Environment, in *Environmental Change in South-East Asia*, eds M J G Parnwell and R L Bryant, Routledge, London, 67 93.

Cazes, G (1992) *Fondements Pour une géographie du Tourisme et des Loisirs*, Bréal, Paris.

Ceballos-Lascuráin, H (1996) *Tourism, Ecotourism and Protected Areas*, IUCN, Gland.

Charoenca, N (1993) Tropical Coastal Pollution: a Case Study of Pattaya, in *Contributions to Training in Coastal Zone Management in the Asia-Paci c Region and Report of the First NETTLAP Resources Development Workshop for Education and Training at Tertiary Level in Coastal Zone Management*, eds J E Hay and L M Chou, UNEP/ROAP Network for Environmental Training at Tertiary Level in Asia and the Pacific, Publication No. 7, UNEP, 123 142.

Cohen, E (1978) Impact of Tourism on the Physical Environment, *Ann. Tourism Res.*, 5, 215 237.

Dearden, P (1993) Cultural Aspects of Tourism and Sustainable Development: Tourism and the Hilltribes of Northern Thailand, in *Tourism and Sustainable Development: Monitoring, Planning, Managing*, eds J G Nelson, R Butler, and G Wall, Department of Geography Publications Series No. 37, University of Waterloo, 165 178.

Dubinsky, Z and Stambler, N (1996) Marine Pollution and Coral Reefs, *Global Change Biol.*, 2, 511 526.

Gajraj, M (1989) Warning Signs, *Tourism Manage.*, 10, 202 203.

Gartner, W C (1996) *Tourism Development*, Van Nostrand, New York.

Genot, H (1997) Tourism, in *The Global Environment: Science, Technology and Management*, eds D Brune *et al.*, 1, 264 275. VCH Verlagsgesellschaft mbH, Weinheim.

German Federal Agency for Nature Conservation (1997) *Biodiversity and Tourism*, Springer, Berlin.

Goodwin, H (1996) In Pursuit of Ecotourism, *Biodivers. Conserv.*, 5, 277 291.
Griffen, T and Boele, N (1997) Alternative Paths to Sustainable Tourism: Problems, Prospects, Panaceas and Pipe-dreams, in *Tourism and Economic Development in Asia and Australasia*, eds F M Go and C L Jenkins, Pinter, London, 321 337.
Gurung, H (1991) Environmental Management of Mountain Tourism in Nepal, *ESCAP Tourism Rev.*, 7, 204 235.
Gurung, P G and De Coursey, M (1994) The Annapurna Conservation Area Project: a Pioneering Example of Sustainable Tourism? in *Ecotourism: A Sustainable Option?* eds E Cater and G Lowan, John Wiley & Sons, London, 177 194.
Hawkins, J P and Roberts, C M (1994) The Growth of Coastal Tourism in the Red Sea: Present and Future Effects on Coral Reefs, *Ambio*, 23, 503 508.
Hjalager, A-M (1996) Tourism and the Environment: the Innovation Connection, *J. Sustainable Tourism*, 4, 201 218.
Hudman, L E (1991) Tourism s Role and Response to Environmental Issues and Potential Future Effects, *Rev. de Tourisme*, 4/1991, 17 21.
Jackson, I (1986) Carrying Capacity for Tourism in Small Tropical Caribbean Islands, *UNEP Industry Environ.*, 9(1), 7 10.
Knight, D, Mitchell, B, and Wall, G (1997) Bali: Sustainable Development, Tourism and Coastal Management, *Ambio*, 26, 90 96.
Lundin, C G and Lindén, O (1993) Coastal Ecosystems: Attempts to Manage a Threatened Resource, *Ambio*, 22, 468 473.
Mathieson, A and Wall, G (1982) *Tourism: Economic, Physical and Social Impacts*, Longman, London.
Maragos, J E (1993) Impact of Coastal Construction on Coral Reefs in the US-affiliated Pacific Islands, *Coast. Manage.*, 21, 235 269.
Pearce, D G and Kirk, R M (1986) Carrying Capacities for Coastal Tourism, *UNEP Industry Environ.*, 9(1), 3 7.
Rai, S C and Sundriyal, R C (1997) Tourism and Biodiversity Conservation: the Sikkim Himalaya, *Ambio*, 26, 235 242.
Smith, R A (1992) Coastal Urbanization: Tourism Development in the Asia Pacific, *Built Environ.*, 18, 27 40.
Thuens, H L (1993) Possibilities for Sustainable Tourism Development in Selected Least Developed Countries and Viet Nam, *ESCAP Tourism Rev.*, 11, 1 18.
Turner, L and Ash, J (1976) *The Golden Hordes: International Tourism and the Pleasure Periphery*, Constable, London.
Ward, J, Higson, P, and Campbell, W (1994) *GNVQ Advanced Leisure and Tourism*, Stanley Thornes, Cheltenham.
Weaver, D B (1998) *Ecotourism in the Less Developed World*, CAB International, Wallingford.
White, A T, Barker, V, and Tantrigama, G (1997) Using Integrated Coastal Management and Economics to Conserve Coastal Tourism Resources in Sri Lanka, *Ambio*, 26, 335 344.
Wong, P P (1998) Coastal Tourism Development in Southeast Asia: Relevance and Lessons for Coastal Zone Management, *Ocean Coastal Manage.*, 38, 89 109.
WTO (1998) *Yearbook of Tourism Statistics*, 50th edition, WTO, Madrid.

Tourism: Climate Change and Tourist Resorts

Jörg Stadelbauer
Albert-Ludwigs-Universität, Universität Freiburg im Breisgau, Germany

The increasing economic and social importance of tourism is challenged by global changes, above all climate change. Rising rainfall instead of snow in mountain regions of medium altitude, earlier snow melting in spring, shortening of the annual snow cover period in high mountain skiing areas, rise of the sea level and withdrawal of the coastline are the main consequences in the temperate zone. They mainly affect coastal and mountain tourism. In semi-arid and arid parts of the sub-equatorial zone, water de cits will cause a fall in lake levels so that present-day resort tourism will be threatened directly. Indirect consequences are caused by de cits in the water supply for inhabitants and resorts. In the tropics some small islands whose population is currently living from tourism are threatened by increasing sea level and may disappear totally. The article examines the general features of probable climate change and its consequences for tourism and gives regional examples taken from different continents and regions.

INTRODUCTION

Economic Significance of Tourism

Today, tourism is one of the most important global economic activities. In 1998, 625.5 million international tourists were estimated to have spent a total of US$445 billion. These figures are based on the statistical definition of the World Tourism Organization, which classifies tourists as foreigners crossing an international boundary and staying at least 24 hours in the other country, with the exception of transit passengers and foreign workers. In addition, the expenditive data are incomplete because only the official money change is registered and additional expenses are estimated. A larger number of local tourists, estimated at 5 to 5.2 billion, do not cross international boundaries. Thus, tourism is the most important element of overall population mobility. At the same time, it is still a growing economic factor. It is believed that about 12 13% of worldwide gross domestic product (GDP) is gained from tourism, and that almost one in eight jobs depends on it. In addition, indirect economic effects result from influences on other economic sectors (including transportation, construction, agriculture, local manufacturing, information and communication services). The annual rates of growth differ regionally from each other; even on the economically

weak African continent they reach up to 15%. The highest number of tourist arrivals was recorded in Europe, with 372 million or 59.5% of the total number of international tourists. The so-called developing countries are not among the leading participants in international tourism, but add 27% to the global income gained from tourism. A significant decline has occurred in the Commonwealth of Independent States (CIS) where international tourism collapsed after the break-up of the Soviet Union. Some countries depend to an extremely high degree on incomes from international tourism. Island states like the Bahamas, Antigua and Barbuda, Barbados and the Maldives have more than 55% of their total of trade and services involved in tourism (Gormsen, 1996). In Europe, Spain, France and Italy are the leading recipients of arriving guests. Among the smaller countries, tourism adds nearly 8% to the Swiss GDP and contributes 9% of jobs. Tourism holds third place among Swiss exports after the chemical industry and machine building. The highest gross incomes gained from international tourism are obtained by countries that are strong in other economic sectors as well (US, France, Italy, Spain). Social phenomena like migration directly depend on tourism.

Global Change and Tourism

Global change comprises not only climatic or other environmental changes, but also includes a large number of phenomena such as the consequences of population growth (development of total population, increase of the concentration in urban settlements, especially in megacities, change of the age structure with overproportional increase of aged persons in highly developed countries, socio-demographic differences between highly and less developed countries (cf. Rayner and Malone, 1998)); pressure on the use of natural resources, mainly energy, soil and water, technological changes, especially in telecommunication and the internet and their economic consequences, economic globalization, political change, social change and ethnic and cultural diversity. The following discussion emphasizes the likely consequences of climate change and only partially considers the other processes of change.

Tourism and Climate

Tourism is not only an important economic factor and social structure; it also depends highly on climatic conditions and is sensitive to climate change. Therefore, global climate change will exercise a significant influence on tourism. This can be assumed not only for different types of landscapes (mountains, coastal areas and small islands), but also for different climatic zones. While in regions near to the poles, warming will occur and the temperate zones may see improved climatic conditions, the subtropical continental regions will undergo higher aridity and experience significant water supply deficits. In the mountains of the temperate zones, global warming will cause decreasing snow cover and increasing rainfall and, thus, shorten the skiing season and generally alter seasonal patterns of tourism. It is necessary, indeed, to take into consideration the inter-annual variability of weather conditions, which normally exerts a significantly higher influence on leisure and tourism activities than simple climatic means suggest. Particularly threatened are low-lying islands, where already a small rise in sea level threatens the economic basis of tourism. In this respect, peripheral centers of tourism remote from major urban areas are more affected by climate change because they have fewer possibilities of adaptation and cannot compensate for the change by attracting visitors from nearby regions.

Due to the large number of influencing factors, scholars dealing with the consequences of climate change deny any climatic determinism, and emphasize the whole range of factors influencing the spatial relationships of tourism (Elsasser *et al.*, 1998). Research on the effects of global climate change is rather recent, but refers to earlier works in which the relationship between climate and cultural development was constructed deterministically on consequences of climatically induced hazards. These were analyzed in simple causal chains (impact approach) because climate seemed to be the determining factor of touristic development. A main reason for this line of argument was that global change was merely identified as climate change without taking notice of indirect consequences. Gloomy future predictions of the consequences of climate change for tourism do not take into account how tourism is constantly subject to changes in fashion and to the development of new alternative opportunities and attractions. Thus, in reality tourism has to adapt and change, and climate change is but one factor in the dynamism of the industry.

General Indirect Consequences of Global Change Influencing Tourism

Global climate change will influence tourism at first, and above all, in coastal and in mountain regions. If sea level rises by a few meters as expected, present coastal settlements will be threatened, including important tourist destinations. For yacht harbors and modern marinas, this development causes high additional costs so that some investments will be unprofitable. For the tourist centers in mountain areas, quantity and the annual distribution of snowfall will be modified and cause changes in seasonality. Another consequence will be the changing runoff from the mountains, not only altering the natural water balance, but also affecting the availability of water for agricultural, industrial and human use including water-based recreation. In high mountain regions the length of the winter season

will decrease, thus forcing the tourist industries to either use artificial snowmaking or to move to higher altitudes (if possible) or to diversify by offering additional attractions in order to extend the tourist season.

A significant part of the temperate zone will experience a rise in temperature. This may be positive in respect of demand for recreation, but higher temperature also causes higher evaporation so that the availability of fresh water diminishes. Furthermore, such activities as hunting and fishing are influenced immediately if the animals alter their habits or migrate into other zones and regions.

Increasing aridity may cause human migrations in areas like the near and Middle East with urban population increases (Watson *et al.*, 1998, 246). Such developments could reduce the political stability and affect the tourist resources of those regions.

Climate change also affects human health care (*see* Environmental Change and Human Health: Extending the Sustainability Agenda, Volume 3). If social services of highly developed countries are to be maintained, the threats to health may increase the demand for climatic health resorts and spas. However, higher temperatures will extend the areas in which tropical diseases originate, particularly those of dengue and yellow fever (Loevinsohn, 1994). The Aedes mosquito, which spreads dengue fever, reacts sensitively to changes of temperature and precipitation. It has already spread over larger areas probably as an effect of changed climatic conditions (Hales *et al.*, 1966). The dangers of the transfer of pathogenic agents increase with tourism.

REGIONAL ASPECTS

Polar and Subpolar Zones

In the Polar Regions, forecasts indicate that indigenous peoples may have to change their lifestyles. Economic gains from tourism could increase with global warming. In the Northern Hemisphere, this effect will occur earlier than in the Southern Hemisphere, where it will be slowed down by the ice cover of Antarctica. Adventure tourism and cruises to Antarctica could be threatened by global warming because of its impacts on the fauna. Although the polar and subpolar zones will experience the greatest climatic change, they have considerably less significance for tourism than the warm temperate zone and the subtropics.

North America

In North America, the mountain regions will show a rise of the regional snowline and a decrease of natural runoff during the snowmelt period. Consequently, the winter ski season will be shortened. For the downhill ski season in Ontario, Wall (1988) (cf. McBoyle and Wall, 1987) calculated an increase of temperature by 3.5 to 5.7 °K and a total increase of annual precipitation by 9%. An analogous calculation for Quebec (Lamothe and Peáriard Consultants, 1988) quotes a shortening of the ski season by 50 70%. Some lost days may be made up by snowmaking devices. Further consequences are seen in the transport sector if less has to be spent, and fewer jobs are involved in snow clearing and winter road maintenance. On the other hand, the summer recreation season may become longer. Districts that are not connected to large regional water supply systems may suffer a seasonal water deficit, which would influence tourism.

Europe

For Europe, all modeling of climate change must take into consideration not only global warming but also several anthropogenic factors, impact of changing land-use cover and even the influences of environmental politics. The most probable forecast indicates an increasing winter temperature. At the same time, it is assumed that winter precipitation will increase by up to 20%. In Northern Europe, summer precipitation will also increase (Watson *et al.*, 1998, 158). Severe storms may occur in NW Europe and threaten the coastal regions. Great Britain may experience summer warming, by increasing summer tourism and the income so derived (Abegg *et al.*, 1997, 107). In the more continental regions of Central Europe, water deficits and a decline in touristic traffic could occur.

In lakes, the changes to runoff regimes will have an impact on tourism. Lake levels will probably be almost the same as at present, but a possible decrease in water input and outflow will affect the water quality and its suitability for swimming, because nutrient loadings will increase and lead to eutrophication.

In the European Alps, this climatic change will cause a deficit of snow during winter months. If the regional warming in the Swiss Alps reaches only 2 °K, the snowline for adequate skiing will be elevated by 300 m. At the moment, ski resorts above 1200 m enjoy an abundance of snow. However, after significant warming, only stations above 1500 m, and on slopes facing the sun, above 2000 m, will have sufficient snow (Abegg and Froesch, 1994, 330). This might cause a decrease of secure skiing areas from 195 (85% of all 230) to 144 skiing areas (or 63%) (Burki, 1998, 155; Elsasser *et al.*, 1998, 152). For winter resorts, the consequence would be either artificial snowmaking or an institutional change to new touristic attractions, with greater regulation of visitor movements if the burden of the remaining snow-reliable areas grows too quickly.

In the Mediterranean region, the main problem will be the growing water deficits. The declining runoff from large rivers will probably cause an intrusion of saline water into

coastal aquifers and make freshwater supply more difficult. Sea level rise will exert additional stress on coastal zones and resort and harbor infrastructure.

Asia

As for Asia, it may be supposed that in the forthcoming years and decades, the intensified melting of high mountain glaciers will cause an increase in runoff, which may threaten touristic infrastructure. Within the mountainous areas, the altitudinal belts will be elevated so that touristic infrastructure also must move. The consequence is an increased need for development of hitherto less accessible areas. Tropical Asian countries have only recently become involved in tourism but are gaining more and more income from such activity. In areas threatened by higher flood risks from increased runoff or more intense typhoons, threats to rural areas will lead to higher migration to urban centers where better flood control may exist. Changes in the timing and intensity of the monsoon could have consequences for mountain tourism in the South Asian Himalaya (Price and Barry, 1997, 433). Tourist infrastructure near coastal zones is menaced by the rise of the sea level, which will create enormous difficulties for island tourist destinations such as the Maldives. But such a rise in sea level would also affect the coastal zone of South East Asia and the delta regions of some large rivers like Indus, Brahmaputra, Menam Chao Phraya and Mekong.

Africa

Tourism in Africa is expanding rapidly despite the continent s economic, social and ecological difficulties in particular rivers. Wildlife tourism in the national parks and resort tourism in West and East Africa have grown. Water availability will be the restricting factor. If, within the next 50 years, mean annual temperatures increases by 1.6 °K in the Sahara and by 1.4 °K in the inner tropical zone (Hernes *et al.*, 1995), decreasing water supplies will affect both the population and short-term visitors. Already, two-thirds of the rural, and one-quarter of the urban population lack secure access to good drinking water. By 2025, as many as 600 million Africans may not have an adequate water supply (Watson *et al.*, 1998, 49). Increasing aridity will lead to a decline in the wildlife population and diminish the attractiveness of some nature parks in the Sahel zone, and in East and South Africa. This effect will have negative consequences for tourism. In the coastal regions, small rises of sea level may threaten the existence of tourist resorts and the livelihoods that depend on them. Natural attractions like Victoria Falls can lose their appeal if the water volume decreases as it did in 1992, and thus visitor numbers may decline. Water sports on lakes (sailing, water skiing, fishing, etc.) may decline, and attractive wetlands like the Okavango Delta may be threatened. Economic investments in the tourism sector will have difficulty repaying shareholders. Hotels near inland lakes and rivers will see a reduced number of visitors if the water deficit increases. Finally, increasing interannual climatic variability, decreases in water supply and water quality and resulting migration will increase the problems of disease control and the fight against epidemics, making Africa a more risky area to visit. Africa is indeed extremely vulnerable to the impacts of climate change.

Australia and New Zealand

In Australia and New Zealand tourism is already the most important earner of foreign exchange, accounting for 13% of all export income. Its importance is increasing (Watson *et al.*, 1998, 111, 135). Global warming will probably cause a sea level rise of 15 100 cm by the year 2100, which will threaten the most frequented shore regions, some coastal resorts, and some marinas. Primarily, it is expected that severe storms will cause large floods. On the other side, sea level rise may favor the growth of corals within the large reefs, and thereby strengthen the attractiveness of the Great Barrier Reef. In the mountainous regions, winter tourism will decline due to the retreat of snow cover and snow security. Thus, the ski season will shorten and become increasingly marginal (Konig, 1999, 147). This will mainly influence ski centers of some national parks (Kosciusko National Park, New South Wales; Alpine National Park, Victoria). The consequence is the potential loss of nearly 12 000 seasonal jobs (Konig, 1999, 148). It is estimated that at least 60 70 days of winter season are necessary in order to maintain the skiing industry financially. This will be possible in most winter resorts if the increase of temperature during the next three decades does not exceed 0.3 °K (favorable scenario); a warming of 1.3 °K (negative scenario) would put the future of winter tourism in doubt (Konig, 1999, 250).

Oceans and Islands

In some small island states in the Caribbean Sea, the Pacific and the Indian Oceans, tourism has become the most important economic sector. On some of those islands, the number of visitors greatly exceeds the number of inhabitants. In Antigua, tourism accounted for 69% of the GDP in 1995, in the Bahamas 53%, and in the Maldives 50%. Sea level rise will increase the loss of beaches by erosion, the danger of flooding and the salinization of freshwater aquifers. Thus, tourism would become unsustainable. On the atoll islands in the Pacific, which only rise 1 2 m above sea level, sea level rise threatens the very existence of the islands. In a scenario from the early 1990s, calculating

a sea level rise of 2 9 mm year^{-1} menaces all touristic activities concentrated near the shoreline. On the islands of the Caribbean Sea, 50% of the population, and an even higher percentage of tourists, are concentrated in a 2 km broad coastal strip that is extremely vulnerable. It might be possible to erect protective barriers against the sea but the high costs of such an undertaking and a loss of scenery would be involved.

Conclusion

Abegg *et al.* (1997, 107) stress the fact that in the temperate zone, seasonal and spatial shifts in regional and local climate change will affect the tourist industry. However, it is not possible to identify local winners and losers among the regions because the changing structure of tourist opportunities will alter the status of localities, and cannot be predicted. The response and adaptation of the tourist industry to changed natural conditions is not well understood. Added to this is the fact that models of climatic change cannot provide small-scale detail. Today one supposes that changes of the structure of tourism will not be caused by first-order, but by secondary consequences. Therefore, how far climate change reduces the economic significance of tourism will depend on the adaptive capability of the tourist industry itself.

REFERENCES

Abegg, B (1996) Klimaanderung und Tourismus: Klimafolgenforschung am Beispiel des Wintertourismus in den Schweizer Alpen, Schlu bericht NFP 31, Zurich.

Abegg, B and Froesch, R (1994) Climate Change and Winter Tourism. Impact on Transport Companies in the Swiss Canton of Graubunden, in *Mountain Environments in Changing Climates*, ed M Beniston, Routledge, London, 328 340.

Abegg, B, Konig, U, Burki, R, and Elsasser, H (1997) Climate Impact Assessment im Tourismus, *Die Erde*, 128, 105 116.

Burki, R (1998) Klimaanderung und Skitourismus. Wie Wirkt Sich Eine Klimaanderung Auf Eine Skitouristische Nutzung Aus? *Geogr. Helvetica*, 53(4), 155 161.

Elsasser, H, Abegg, B, Burki, R, and Konig, U (1998) Beitrage der Tourismusgeographie zur Klimanderungsfolgenforschung, *Geogr. Helvetica*, 53(1), 150 154.

Gormsen, E (1996) Tourismus in der Dritten Welt. Ein Uberblickuber Drei Jahrzehnte Kontroverser Diskussion, in *Tourismus in der Dritten Welt*, eds G Meyer and A Thimm, Mainz Universitat, 11 46.

Hales, S, Weinstein, P, and Woodward, A (1996) Dengue Fever Epidemics in the South Pacific: Driven by El Nino Southern Oscillation? *Lancet*, 348, 1664 1665.

Hernes, H, Dalfelt, A, Berntsen, T, Holtsmark, B, N ss, L O, Selrod, R, and Aaheim, H A (1995) *Climate Strategy for Africa*, Cicero Report No. 1995(3), Oslo.

Konig, U (1999) Climate Change and Snow Tourism in Australia, *Geogr. Helvetica*, 53(3), 147 157.

Lamothe and Peáriard Consultants (1988) Implications of Climate Change for Downhill Skiing in Quebec, CCD 88 03, *Environment Canada*, Downsview, Ontario, Canada, 1 12.

Loevinsohn, M (1994) Climatic Warming and Increased Malaria Incidence in Rwanda, *Lancet*, 343, 714 718.

McBoyle, G R and Wall, G (1987) The Impact of CO_2-induced Warming on Downhill Skiing in the Laurentians, *Cah. Géogr. Québec*, 31(82), 39 50.

Price, M and Barry, R G (1997) Climate Change, in *Mountains of the World. A Global Priority*, eds B Messerli and J D Ives, Parthenon, New York, 409 445.

Rayner, S and Malone, E, eds (1998) *Human Choice and Climate Change*, Vol. 1, Battelle Press, Columbus, OH.

Wall, G (1988) Implications of Climate Change for Downhill Skiing in Ontario, CCD 88 05, *Environment Canada*, Downsview, Ontario, Canada, 1 16.

Watson, T, Zinyowera, M C, and Moss, R H, eds (1998) *The Regional Impacts of Climate Change. An Assessment of Vulnerability*, Cambridge University Press, Cambridge.

FURTHER READING

Abegg, B, Konig, U, and Maisch, M (1994) Klimaanderung und Gletscherskitourismus, *Geogr. Helvetica*, 49(3), 103 114.

Beniston, M, ed (1994) *Mountain Environments in Changing Climates*, Routledge, London.

Besancenot, J P (1990) *Climat et Tourisme*, Masson, Paris.

Bruce, J P, Lee, H, and Haites, E F, eds (1996) *Climate Change 1995: Economic and Social Dimensions of Climate Change*, Cambridge University Press, Cambridge.

Konig, U (1998) *Tourism in a Warmer World: Implications of Climate Change due to Enhanced Greenhouse Effect for the Ski Industry in the Australian Alps*, Wirtschaftsgeographie und Raumplanung, 28, Zurich.

Krupp, C (1995) *Klimaänderung und die Folgen. Eine Exemplarische Fallstudie über die Möglichkeiten und Grenzen Einer Interdisziplinären Klimafolgenforschung*, ed Sigma, Berlin.

Parry, M L (1993) Geographers and the Impact of Climate Change, in *The Challenge of Geography: a Changing World - a Changing Discipline*, ed R J surJohnston, Oxford University Press, Oxford, 138 147.

Parry, M L, Blantan de Rozari, M, Chong, A L, and Panich, S (1992) *The Potential Socio-Economic Effects of Climate Change in South- East Asia*, United Nations Environment Program, Nairobi.

Smith, K (1990) Tourism and Climate Change, *Land Use Policy*, 7, 176 180.

Sterr, H (1993) Klimawandel, Klimawirkungsforschung und die Rolle der Geographie, *Geographische Rundschau*, 45(3), 196 198.

Wall, G (1992) Tourism Alternatives in an Era of Global Climatic Change in *Potential and Problems in the Development of Tourism*, eds W Smith and W Eadington, Philadelphia, PA, 194 215.

Wall, G and Badke, C (1994) Tourism and Climate Change: an International Perspective, *Sustainable Tourism*, 2, 193 203.

Whetton, P H, Haylock, M R, and Galloway, R (1996) Climate Change and Snow Cover Duration in the Australian Alps, *Clim. Change*, 32, 447 479.

Toxic Wastes: Generation and Disposal: a Case Study of UK Practice and Legislation

Stephen L Willetts
The Willetts Partnership, Atherstone, UK

This essay deals with the past practices for the disposal of toxic wastes generated by industry in the UK. It then explores how and why legislation evolved in order to control the potential effects of such disposal practices.

Many practices for the disposal of toxic wastes were gradually phased out because of the implementation of the legislation, both domestic and, more recently, international. Disposal by way of land ll became the preferred route - if not the only remaining option. It might be said, therefore, that the UK has a legacy of land lled toxic wastes and that this may well erupt because of the predicted forthcoming global environmental changes.

The size and scope of the potential legacy is quanti ed and some of the predicted effects discussed. The conclusion of the study is that, whatever the predicted effects might be, they are, at least, manageable and should not be disastrous!

So-called toxic wastes have always been disposed down the drains, up the chimney or dumped on land: the options are straightforward sea fill, air fill or landfill! The former two methods have the added advantage that the wastes are swept away from the place even the country where they were produced. This is the classic *dilute and disperse* philosophy and some would argue that there is nothing at all wrong with this approach to waste management so long as we do not overload the self-purification capacity of the receiving environment.

The landfill approach perhaps unfortunately, keeps the problem closer to home and, therefore, warrants a more diligent management philosophy. In the UK, toxic wastes have generally been co-disposed with household and other readily biodegradable rubbish. Proper site management should avoid the obvious problems of violent and potentially destructive chemical reactions taking place as a result of mixing incompatible types of wastes.

There is a more insidious problem with landfills however, and that is the inevitable generation of highly polluting liquors called leachate: this is the liquid formed as a result of rain flushing through the wastes and dissolving chemicals from the wastes themselves and their products of biodegradation. The chemical constituents of the toxic wastes landfilled can, therefore, be leached out: this is especially so in the aggressive, acidic conditions produced initially during the, usually anaerobic, biodegradation process.

This leachate, if not contained within the site by natural or artificial barriers, can seep through the base and sides of the site and contaminate groundwater or can overtop the site and pollute surface streams: both of these water bodies can be sources of potable water, of course.

Toxic wastes were not recognized as being problematic in terms of their disposal in the UK at least until a celebrated incident near Nuneaton in 1972. Certainly, prior to that date, controls on their discharge to sewers and other waters were strictly controlled by virtue of a system of formal consents, as were discharges to the atmosphere from their incineration: disposal to land, however, was not expressly controlled by legislation. The seminal incident that forced a rethink on their disposal to land involved the discovery of several abandoned drums of cyanide wastes on derelict land on which children regularly played.

That is not to say, however, that toxic and other wastes had not already been recognized singularly as being in need of proper management as regards their pollution potential. Indeed, as early as 1953, the then Minister of Housing and Local Government commissioned a technical committee to report on The Pollution of Water by Tipped Refuse : it concluded that controlled tipping was generally a satisfactory method of refuse disposal.

In 1964, another technical committee was commissioned to report on The Disposal of Solid Toxic Waste : it concluded that there was a definite need for a formal authorization procedure administered regionally and established by legislation.

Events rather overtook themselves with the Nuneaton incident, however, and the Deposit of Poisonous Waste Act was hurriedly implemented on 30th March 1972. Producers of poisonous, noxious or polluting wastes were required to give at least three days notice of any intended deposit to the statutory local authorities. Such wastes were not, in fact, defined in the Act, but the terms had been given legal recognition as a result of earlier court cases. The local authorities were thus in a position to build up a picture and database of such poisonous waste disposal transactions throughout the UK.

During this period, the environment was being firmly pushed onto the political agenda: Rachel Carson s *Silent Spring* and the Club of Rome s *Limits to Growth* were raising serious concerns about the consequences of environmental pollution. These publications helped to encourage the promulgation of several new pieces of waste management legislation.

For example, the attractively descriptive Dumping at Sea Act of 1974 came about as a response to the Oslo and

London Conventions, which sought to prohibit the dumping in international waters of certain categories of wastes and to provide the requirement for formal permits to dump those wastes that were to be allowed with precautions. At its peak, sea dumping accounted for around one-and-a-half million tonnes of industrial waste and 10 million tonnes of contaminated sewage sludge each year. The practice has now been phased out as a result of international agreements.

Incineration of wastes at sea, usually in the North Sea, accounted for the disposal of over 100 000 tonnes per year of halogenated residues during the 1970s. The advantage over land-based incineration was simply that no expensive gas cleaning systems were employed. The practice was phased out in the early 1980s.

During this time, the UK government seized the opportunity of taking stock of its existing piecemeal environmental legislation, replacing it with the encyclopedic Control of Pollution Act in 1974. This was the first integrated approach in the world to protecting the environment. The Act dealt with four major areas in pollution: waste on land; water pollution, both of the sea and inland water; pollution of the atmosphere; and pollution by noise.

The greatest impact on the waste disposal industry came from the introduction of a formal licensing regime for all waste disposal sites. The authorities could impose any condition they saw fit in a license. These conditions became ever more stringent as environmental concerns grew, particularly over groundwater protection, and from the requirements of European Directives.

The old-style *dilute and disperse* landfills or as the then Secretary of State for the Environment euphemistically termed them, *dilute and attenuate* were gradually to be phased out. The new *concentrate and contain* philosophy of landfill management demanded impermeable liners on the base and sides of sites and the managed removal of leachate for proper treatment and disposal elsewhere in order to prevent any migration of polluting leachate either downwards or laterally from the site. Upon their closure, sites would have to be sealed by impermeable caps to keep rainfall out thereby reducing the amount of leachate produced.

It also became common to set loading rates for the types and maximum amounts of toxic wastes to be deposited as a proportion of biodegradable wastes. The theory behind this was that organic species in the toxic wastes would eventually biodegrade so long as their concentrations were kept modest and the practice of co-disposal, as in the UK, would work as long as the remaining disposed refuse provided sufficient biological buffering that the microorganisms could cope with the otherwise toxic loading.

Even materials that might be thought to be extremely refractory to biodegradation for example, polychlorinated biphenyls (PCBs) and cyanides will biodegrade in the right conditions.

The 1974 Act also made special provision as to the disposal of waste of a particularly dangerous kind or of a kind that was difficult to dispose of: the term special waste was introduced. This predated the European Hazardous Waste Directive by some eleven years. The implementing regulations came into force in 1981 at which time the Deposit of Poisonous Waste Act was repealed.

Essentially, the term *special waste* was defined as any controlled waste which was dangerous to life ••• or has a flash point of 21 degrees Celsius or less and waste was to be regarded as dangerous to life if a single dose of not more than 5 cm^3 would be likely to cause death or serious damage to tissue if ingested by a child of 20 kg body weight or exposure to it for 15 minutes or less would be likely to cause serious damage to human tissue by inhalation, skin contact or eye contact .

It is difficult to imagine how this definition could be tested practically. The idea of rows and rows of 4-year old children queuing up to have waste samples rammed down their throats, rasped into their skin and douched into their eyes is both unacceptable and unrealistic.

The regulations introduced a comprehensive system of consignment notes to document the movement of the waste from cradle to grave thus providing proper records of waste sources and destinations.

These regulations continued until repealed and replaced in 1996 by new ones that fulfill certain European Community (EC) Directive obligations. The current definition of *special waste* is now far less prosaic referring, as it does, to the European Hazardous Waste List.

The 1999 Landfill Directive ensures that toxic wastes are only deposited into effectively sealed landfill sites independently from other wastes. The apparent driving philosophy is that future generations will be able to recover them if technologies are developed for the reuse or recycling of the wastes. If not, at least they will, henceforth and in perpetuity, have been entombed, bottled-up, isolated, for all intents and purposes, from the environment: the Directive will see to that!

Essentially, then, the UK has a legacy of toxic wastes in landfill sites with little control and even less knowledge up until the mid-1970s. Thereafter, landfill sites were subject to rudimentary regulatory controls that have evolved into our present-day management practices. For about two decades, toxic waste landfill sites were not specially lined and leachate has gradually or sometimes suddenly seeped, leaked or spilled away into the aquatic environment. Concentrate and contain is now the dominant philosophy as is the maintenance of accurate and rigorous records for posterity: if previous practices were suspect, they have now effectively been, or very soon will be curtailed.

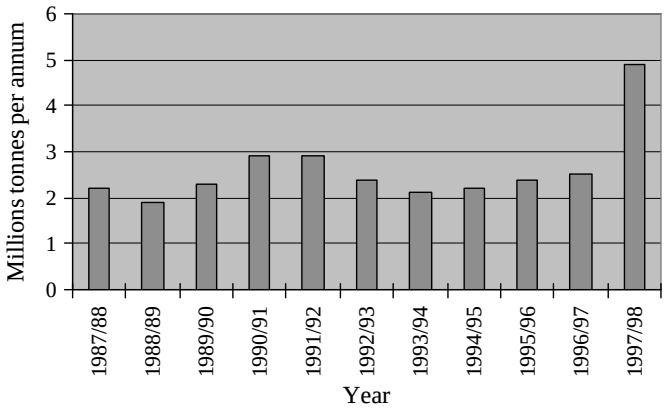


Figure 1 UK special waste arisings 1987/1988–1997/1998

The legacy can be roughly quantified from official data first published from about 1984/5, although the government always heavily qualifies the accuracy of the figures. The big jump to the 1997/98 levels is said to be a consequence of the change in definition of special wastes in late 1996. The picture is illustrated in Figure 1.

As can be seen in Figure 1, between two and three million tonnes of special wastes have been produced each year and with some consistency until the change of definition. Although a large mass, this tonnage is only some 0.005% of the total waste produced each year.

The disposal routes used for special wastes have remained relatively constant and are illustrated in Figure 2.

It has to be noted, however, that incineration will typically produce anything up to 30 40% by weight of ash and gas cleaning residues: chemical treatment will produce anything up to 60 70% by weight of residues. These will ultimately need disposal to landfill.

From 1970 to the present, possibly up to 5 6000 licensed landfill sites operated in the UK at one time or another. Those already closed will remain capped and forgotten unless and until something untoward happens and leachate starts to emerge. Only a select few of these, perhaps no more than a hundred, will have received significant amounts of toxic waste.

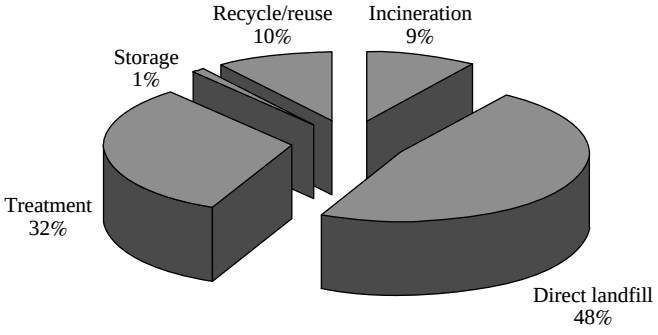


Figure 2 Special waste disposal routes

The useful Friends of the Earth Landfill Campaign Guide of June 1997 estimated that there were then about 3700 operational landfill sites in the UK. Of these, about 100 sites accounted for almost one half of all waste that is landfilled: these are the ones most probably taking special wastes. The picture is one of a few, large, strategically located sites and hundreds of others apparently serving purely local markets.

Most of the landfill sites used for special wastes will have been located in naturally impermeable clay strata for water pollution prevention reasons and logic dictates that they will also have been located as close as possible to the manufacturing industries that generated the wastes.

How will the predicted global environmental changes affect these sites? The main aspects to consider here are coastal erosion, coupled with, or as a consequence of, a rise in sea levels, increasingly heavy rainfall and rising groundwater levels. The issue of most significance will be the release, sometimes sudden, sometimes insidious, of toxic, heavy metals into the environment.

It is reasonable to discount any significant effects from landfill gas since the proportion of biodegradable organic species in toxic wastes is insignificant compared with that provided by ordinary domestic and commercial wastes: the relative contribution of methane derived from toxic wastes and the consequent contribution to global warming can, therefore, be discounted. In any event, in the long term, biodegradation of these organic species will result in the wastes becoming environmentally benign, leaving only the toxic heavy metals to consider.

Of course, landfill sites located in the coastal zone will be prone to inundation and scouring out by rising sea levels and coastal erosion. Such wholesale removal of the wastes into the sea would, almost certainly, produce

immediate toxic effects on marine organisms. Longer term bioaccumulation in these organisms is, also, almost certain with the scope and extent of the resultant damage being dependent upon the dispersion caused by tides and currents. Shellfish will be the most susceptible to poisoning by heavy metals and especially their bioaccumulation but short-term damage to commercial fisheries could be expected, too.

As regards increasingly heavy rainfall, this will inevitably lead to increased volumes of leachate being produced from active and closed landfill sites: even sites capped with clay and/or synthetic materials do allow infiltration which contributes to leachate production. The consequence will be a need for vigilant leachate monitoring, removal and treatment or an otherwise accumulation within the sites with the real prospect of escape into both ground and surface waters. If this does happen, there will be increased costs of treatment of any abstracted water prior to its supply for use: either that or the need to abandon specific polluted sources as a consequence of contamination by leachate.

It is also possible that increased river flows will inundate and scour out landfills situated on or close to the river banks or in the river flood plains. This might result in potentially catastrophic release of toxic substances causing fish kills and other damage. Fortunately such incidents tend to be transient in their effects simply owing to the volume and flow rate of the water, but not always so: there are cases of the plug flow of pollutants wiping out huge stretches of rivers.

Rising groundwater levels will also affect landfill sites. There is the potential for sealed sites to be floated out of the ground by upward hydraulic pressure: certainly, there have already been instances of liners having been lifted out because of rises in local groundwater levels owing to pump failure. If this were to happen, then the result would be land contamination in the immediate vicinity with the need to collect and remove the wastes elsewhere for disposal followed by remediation with clean materials of the former landfill site.

There is also the prospect of rising groundwater entering into landfill sites, including entry through lining systems by virtue of increased hydraulic pressure, contributing to increased volumes of leachate and its potential escape.

All of these impacts of global environmental change on our legacy of tipped, toxic wastes are credible, the probability depending upon the location and circumstances of individual sites. None of the impacts needs be disastrous: all are manageable.

What is needed to avoid real problems is vigilance and an accountable responsibility for this. Fortunately, policy on this is emerging: someone out there is listening.

FURTHER READING

Cope, C B, Fuller, W H, and Willetts, S L (1983) *The Scienti c Management of Hazardous Wastes*, Cambridge University Press, Cambridge.

Institute of Wastes Management (1995) *The Role and Operation of the Flushing Bioreactor*, Report of the IWM Sustainable Landfill Working Group, IWM, Northhampton.

Porteous, A (1985) *Hazardous Waste Management Handbook*, Butterworths, London.

Transboundary Movements of Hazardous Wastes

Iwona Rummel-Bulska

United Nations Environment Programme, Geneva, Switzerland

The transboundary movement of hazardous wastes became of global concern in the 1980s. This is because of the growing export (both legal and illegal) of hazardous wastes from industrialized to developing countries. The practice was a response to rising costs for hazardous waste disposal in rich countries at a time when costs for such disposal in developing countries were minimal. The export of hazardous wastes from the highly industrialized Organization for Economic Cooperation and Development (OECD) countries to less developed non-OECD nations grew dramatically from 1970 to 1995. Greenpeace estimated that in the 20 years before 1989 approximately 3.6 million tons of hazardous wastes were exported, but as much as 6662.6 million tons were shipped in only 5 years between 1989 and 1993. Exact gures on the quantity of hazardous wastes involved are not available but the United Nations Environment Programme (UNEP) estimates the total annual transboundary movement of such wastes amounts to be between 300 and 500 million metric tons.

Among the highly publicized cases contributing to media attention and consequent public concern about transboundary movements were the dumping of polychlorinated biphenyls (PCBs) from Italy into a farmer s field in Koko, Nigeria, and the case of the toxic waste ship Khian Sea in 1987 1988 (Kreuger, 1999). As a result, UNEP sponsored negotiations on drafting a global treaty (the so-called Basel Convention), which was adopted by 116 governments in 1989, and which came into force in May 1992, when it received the necessary 20 ratifications. In March

1994, the second Conference of the Parties of the Basel Convention adopted a full ban on all hazardous waste trade to developing countries (*see* Basel Convention, Volume 4).

Eager to avoid negative publicity as well as to circumvent laws against the dumping of toxic wastes, many companies are said to disguise their deadly exports with benign labels. Greenpeace estimates that 86 90% of all hazardous waste shipments destined for developing countries are purported to be materials for recycling, reuse, recovery, or humanitarian uses. Many importing countries have discovered that toxic wastes by another name still pose major problems. Indonesia was importing foreign plastics for recycling until 1992 when it found that 40% of the imported material could not be recycled and that approximately 10% of it was toxic and imposed a ban on all importation of plastic waste.

The environmental and human impacts of the transboundary movement of hazardous waste may be severe. Reports from the Philippines reveal that even legitimate hazardous waste recycling operations approved by the government are causing toxic emissions, and in many cases creating residual hazardous waste more toxic than the original waste. During 1993 1996, Singapore, Saudi Arabia and Australia were the three main exporters of drained or undrained scrap batteries to the Philippines. People living near a lead re-smelting facility at the recycling plant complained of nausea, burning eyes, sore throats and various respiratory ailments. Tests on the blood lead levels of children in the area revealed blood lead levels ranging from 8 micrograms per deciliter ($\mu g\,dl^{-1}$) to as high as $29\,\mu g\,dl^{-1}$. Blood lead levels as low as $10\,\mu g\,dl^{-1}$ cause problems such as decreased intelligence, impaired hearing and stunted growth. Soil, river sediment and vegetation samples collected near the recycling plant had lead levels greatly exceeding the limits of lead in soil and plant matter for crop cultivation and human consumption. The problem of transboundary waste disposal has not yet gone away, even between OECD countries. For example, after considerable public outcry and political debate, the City Council of Toronto (Canada) recently decided (October 2000) to send garbage to a land-fill site near Detroit (USA) rather than to an abandoned mine (Adams Mine) 570 km to the north in a relatively unsettled part of the Province of Ontario. Although admittedly, the wastes will not technically be hazardous, the example shows that the temptation to dispose of waste materials in any way possible is still an option that is considered by political bodies.

REFERENCE

Kreuger, J (1999) What s to Become of Trade in Hazardous Wastes? *Environment*, 41, 11 21.

Transport and Species Dispersal

Ian F Spellerberg
Lincoln University, Aotearoa, New Zealand

With increasing volumes and speed of transport of people and cargo around the world, it is no surprise that transportation has aided the distribution of species and consequently has contributed much to environmental change. Roads and vehicles, trains, barges, ships and aircraft have all contributed to species dispersal. Road vehicles transport seeds, trains carry dust mites, ships have transported African mosquitoes to Brazil. Ocean liners have been vectors for land birds, butter ies and other insects have been have been found in cars transported across the world, and bats have been transported from France to the UK amongst a load of oak planks. Containers have been found to be vectors of many organisms including a raccoon transported from the US to France.

For thousands of years humans have contributed to changes in the dispersal of plants, animals and other organisms. We have mixed and stirred the world s biogeography either deliberately or by accident. Some people talk of homogenization of the world s biota. Many species have been deliberately introduced to new areas mainly as a new source of food (or some other kind of resource) and to control pests. One of the best known examples of programs of intentional introductions was in New Zealand. Plants and animals were transported from Europe, North America and other countries under a series of legislative acts commencing in 1861. Some of these introduced species are now extremely abundant and invasive pests. In Australia, one of the most spectacular examples of invasive species was the spread of the European rabbit. This originated from a shipment of rabbits imported from England in 1859 and within three years they had become a potential pest.

The infrastructure for transport has also contributed to the dispersal of species. Roadsides, edges of railway tracks and canals have all in some way acted as linear habitats and helped to facilitate the dispersal of many species. The story of the Oxford Ragwort (*Senecio squalidus* L.) is a classic example. This South European species, now abundant in many parts of Britain was once growing in the Oxford University botanical garden in 1770s. By the 1830s, it had spread to the Oxford railway station and from there it spread along the Great Western Railway.

Perhaps the most well researched example of species being dispersed by transport (Laird, 1984) is that of the

role of road vehicles in the dispersal of plants (particularly weeds) and seeds (Wace, 1977). As early as 1930 it was suggested that the spread of weeds along roadsides had occurred as a result of seeds being transported in mud attached to vehicles. A study of seeds found in the mud on a car that had traveled over 15 000 km around Germany revealed 124 species with a total of 3926 seedlings (Schmidt, 1989). In Australia, Wace found seeds from 259 species collected over 2.5 years from cars in Canberra. Some concern has been expressed about the danger of alien and invasive weed species being transported by cars into protected areas. Weeds are indeed transported into national parks by cars but it is believed that rather than attempting to prevent this movement, it is better to detect and eradicate existing weed infestations.

Ships and other forms of aquatic transport (together with fishing nets and other equipment) are vectors for dispersal of species. For thousands of years, marine life such as barnacles, molluscs, seasquirts and algae must have been dispersed around the world on the bottoms of the many forms of marine transport. Beetles native to Europe (especially ground dwelling and from dry conditions) have been transported in ships ballast to eastern North America. Shipping, and even recreational boating, have been important in the relatively recent invasion of North American fresh waters by the Zebra Mussel (*Dreissena polymorpha*). This species is now the most important fouling organism in the US and Canada causing much ecological damage and causing considerable expense.

Cargo ballast water has been known to be a vector for dispersal of marine species since the start of the 19th century and the transportation of pathogenic species is of increasing concern. Ships from Latin America have been found to have cholera in their ballast on arrival in Australia and the US. The ballast water and sediments in the ballast tanks are known to have contributed to the spread of many planktonic species and benthic species (Williams *et al.*, 1988) and ballast water is now considered to be the largest single vector of exotic species. More and more invasions of exotic marine species are being linked to ballast water and there are considerable threats to aquaculture.

Aircraft are contributing to accidental species dispersal. The dispersal of mosquitoes, other insects and species of fungi via international aircraft has become a well-known phenomenon; particularly worrying has been the transport of disease vectors. Aircraft have unintentionally transported barn owls, bats, snakes and rodents.

The trade liberalization stemming from treaties such as General Agreement on Tariffs and Trade (GATT), together with the growing volume and speed of transport have increased the risk of unintentional spread of species via transportation (Laird, 1984). Many introduced species have become invasive and are contributing to unwanted global environmental change. Article 8(h) of the 1992 Convention on Biological Diversity calls for states to prevent the introduction of, control or eradicate those alien species which threaten ecosystems, habitats or species .

Dispersal and introduction of invasive and alien species by ballast water has been addressed and is being managed by several countries. The United Nations International Marine Organization (IMO) is also facilitating management to reduce the risks of change to indigenous communities caused by invasive species transported by ballast water. The biosecurity risks posed by air cargo are being researched and some authorities are ensuring that containers have been cleaned before loading at airports.

The huge volumes of road transport pose a considerable risk for the spread of weeds. Traditional mechanized commercial car washing is known to be merely a cosmetic operation leaving an abundance of weed seeds in the vehicle borne biota (Wace, 1977). More effective cleaning methods will need to be developed especially for vehicles being imported. A very recent issue has been the potential spread of foot and mouth disease from Europe to North America.

See also: Biological Invasions, Volume 2; Plant Dispersal and Migration, Volume 2.

REFERENCES

Laird, M (1984) *Commerce and the Spread of Pests and Disease Vectors: Paci c Science Congress, 1983, Dunedin, New Zealand*, Praeger Publishers, New York.

Schmidt, W (1989) Plant Dispersal by Motor Cars, *Vegetatio*, 80, 147 152.

Wace, N (1977) Assessment of Dispersal of Plant Species, the Car-borne Flora in Canberra, *Proceedings of the Ecological Society of Australia*, 10, 167 186.

Williams, R J, Griffiths, F B, Van der Wal, E J, and Kelly, J (1988) Cargo Vessel Ballast Water as a Vector for the Transport of Non-indigenous Marine Species, *Estuar. Coast. Shelf Sci.*, 26, 409 420.

Transport: Global Freight and Passenger Flows

Peter Rimmer
The Australian National University, Canberra, Australia

The goal of global transport is to provide links for movements of freight and passengers between and within six continental areas: Africa and the Middle East, Asia, Australasia,

Europe, North America and Central and South America. Over time, the relative importance of transport modes performing this task - water, road, rail, pipeline and air - has changed with innovations in transport (and communications) technologies, and shifts in institutional and organizational formats. Canals, railways, private motorized transport and commercial aviation represent past technological jumps; and shifts from anti-monopoly through central control and nationalization to competitive policies, re ect historical movements in the nature of regulatory regimes. All transport modes are environmentally intrusive but only since the early 1970s has the nature and scale of their global impact become a key issue for policy makers. The industrialization of Asia has led to fundamental changes in the structure of the world's transport system. Asia is now linked, through an east- west main street in the Northern Hemisphere, with Europe and North America. Conversely, Africa, Australasia and Central and South America in the Southern Hemisphere have been relegated to the status of cul-de-sacs. They are connected to main street by north- south feeder links through major hubs in the Northern Hemisphere, which are already served by extensive continental feeder systems.

The stark representation of the present global transport infrastructure shown in Figure 1 reflects an ongoing search for full efficiency from integration. The development of the global hub and spoke system aimed at reducing costs to the user and stimulating technological and organizational innovation, has been accompanied by greater public concern about environmental degradation. Regional emissions of nitrous oxides, and global emissions of carbon dioxide, compound local problems of noise, atmospheric lead pollution, vibration, and visual intrusion and community severance. Thus, two issues are raised: how has this pattern of global transport emerged since the early 1970s, and what are the environmental challenges and policy implications of transport development?

In pursuing these issues, a distinction is made between freight and passenger movements between and within the six continents. This distinction reflects the fact that the restructuring of the world s transport system since the early 1970s was to meet the two-fold needs of transnational corporations for: (a) door-to-door freight systems to knit together their dispersed production sites, and (b) seamless passenger services for personnel and their diffused clients. The extent to which these global local needs have been met, and their resultant environmental repercussions, are discussed in separate studies of freight and passenger systems.

A major drawback to this analysis is the quality of transport statistics. There are useful data on global bulk and liner shipping and air transport movements (FR, 1999; CIY, 1998; IATA, 1999; UNCTAD, 1998). Data on local (regional) movements are inadequate except for the European Union (EU) and North America (Gerondeau, 1997; DETR, 1998; ECMT, 1995; ECMT, 1999). The UN (1997) covers all modes road, rail, sea, air and inland waterways but recourse has to be made to commercial sources to compensate for many omissions (JUTS, 1998; IUR, 1999; TMC, 1997).

Even then, information on road transport is restricted to numbers of vehicles in use worldwide and their uneven distribution (Table 1). There are 494 million passenger cars worldwide, or one car for every 11 people (TMC, 1997). Countries with more than 400 cars per 1000 people include Australia, New Zealand, United States and Canada; and those experiencing marked congestion include Germany,

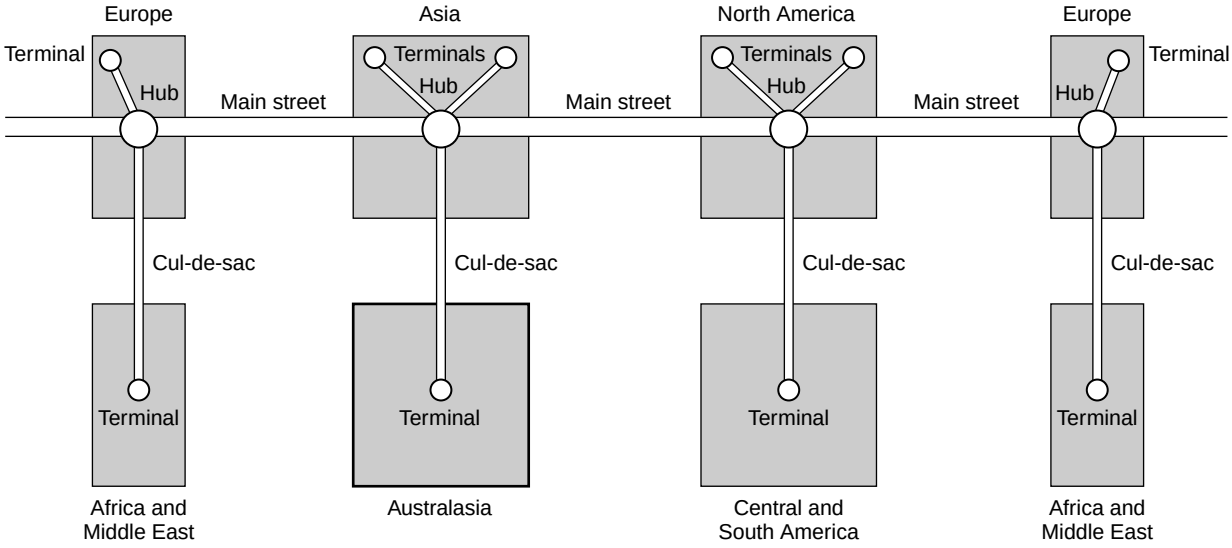


Figure 1 The global hub and spoke system

Table 1 Regional distribution of motor cars and trucks and buses, 1995[a]

	Passenger cars		Trucks and buses		Total	
	Mill	%	Mill	%	Mill	%
Asia	74.1	15.0	40.0	24.7	114.2	17.4
Europe	203.6	41.2	26.4	16.3	230.0	35.1
North America	149.1	30.2	70.6	43.6	219.7	33.5
Africa	12.2	2.5	5.5	3.4	17.7	2.7
Middle East	11.4	2.3	4.3	2.7	15.7	2.4
Central and South America	33.7	6.8	11.8	7.3	45.5	6.9
Australia	10.3	2.1	3.1	1.9	13.4	2.0
	494.4	100.0	161.7	100.0	656.2	100.0

[a] European figures are distorted by the inclusion of all of the Russian Federation.
Source: reproduced using data from EB (1999).

Israel, Italy, Japan, Kuwait, Mexico, Romania, South Korea, Switzerland, Taiwan, Thailand, United Arab Emirates and the United Kingdom (IRF, 1996; IRTU, 1996). This information is inadequate for any in-depth discussion of world motorization and its environmental repercussions. Indeed, there is little information covering the expected growth in car ownership and traffic in low-income countries and their impact on trans-boundary pollution problems. Given these data problems, attention is concentrated on freight and passenger movements between continents, with only scant reference to those originating and terminating within continents.

FREIGHT

The demand for freight transport is derived from economic growth and the volume of international trade. An increase in industrial production, and an expansion of trade, result in an increasing demand for freight transport. Although the rate of economic growth has been declining since the early 1970s, it has been uneven, as witnessed by strong economic growth in East Asia and low-income countries. While the rate of production has been declining, world trade has been reviving buoyed by the global integration of production, reflecting fewer impediments to trade and innovations in transport and communications.

Intercontinental Systems

Land connections are possible between Asia, Europe and Africa, and North, Central and South America. Much effort has been invested in intercontinental rail freight systems between Asia and Europe, but the Trans-Siberian Railway is of diminishing importance and the Trans-China Railway has yet to prove itself. Consequently, deep-sea shipping dominates mainline intercontinental freight systems ahead of air cargo, with road, rail, water, air and pipeline providing the necessary urban and regional feeder systems within individual continents.

Deep-sea Shipping

A distinction is made in shipping between carriers that handle commodities in bulk (e.g., oil, iron ore, coal and grain) and liners that handle general cargo on a regular schedule. Bulk cargoes, as has been the case since the 19th century, still depend on either shipper or manufacturer (e.g., car carrier or pulp and paper vessels). Deep-sea liner shipping has undergone marked organizational changes since the early 1970s to meet the increase in container cargoes. There is an active market for the sale of bulk ships but only a small resale market for container ships. Bulk operators buy assets, but container operators purchase transport and communications networks. Bulk carriers are still involved in the shipping business, but increasingly, liner shipping is part of the wider maritime global hub and spoke transport system, which seeks to meet the demands of manufacturers for a truly global reach in door-to-door services (logistics).

Bulk Shipping

Bulk ships transport high volume low unit value substances such as crude oil, iron ore, coal and grain in shipload lots between producing areas and major industrial processing centers. This freight business is transacted in the ship charter market, which is characterized by marked daily fluctuations in freight rates. As many variables are involved, the ship charter market is risky. Freight rates on some bulk commodities can account for over one-third of the delivered price. Given these slim margins, a few cents difference can determine trading profitability or loss. This explains the close attention paid by maritime research institutes to forecasting likely short- and long-term trends in the bulk shipping market for major commodities, notably crude oil, iron ore and coal (no figures are available for the transport of hazardous materials).

Movements of crude oil totaled 1415 million tonnes in 1995 16% smaller than the peak year of 1979 due to conservation efforts and a switch to other energy types. The Middle East played a pivotal role in supplying major markets in East Asia that were supplemented by Southeast

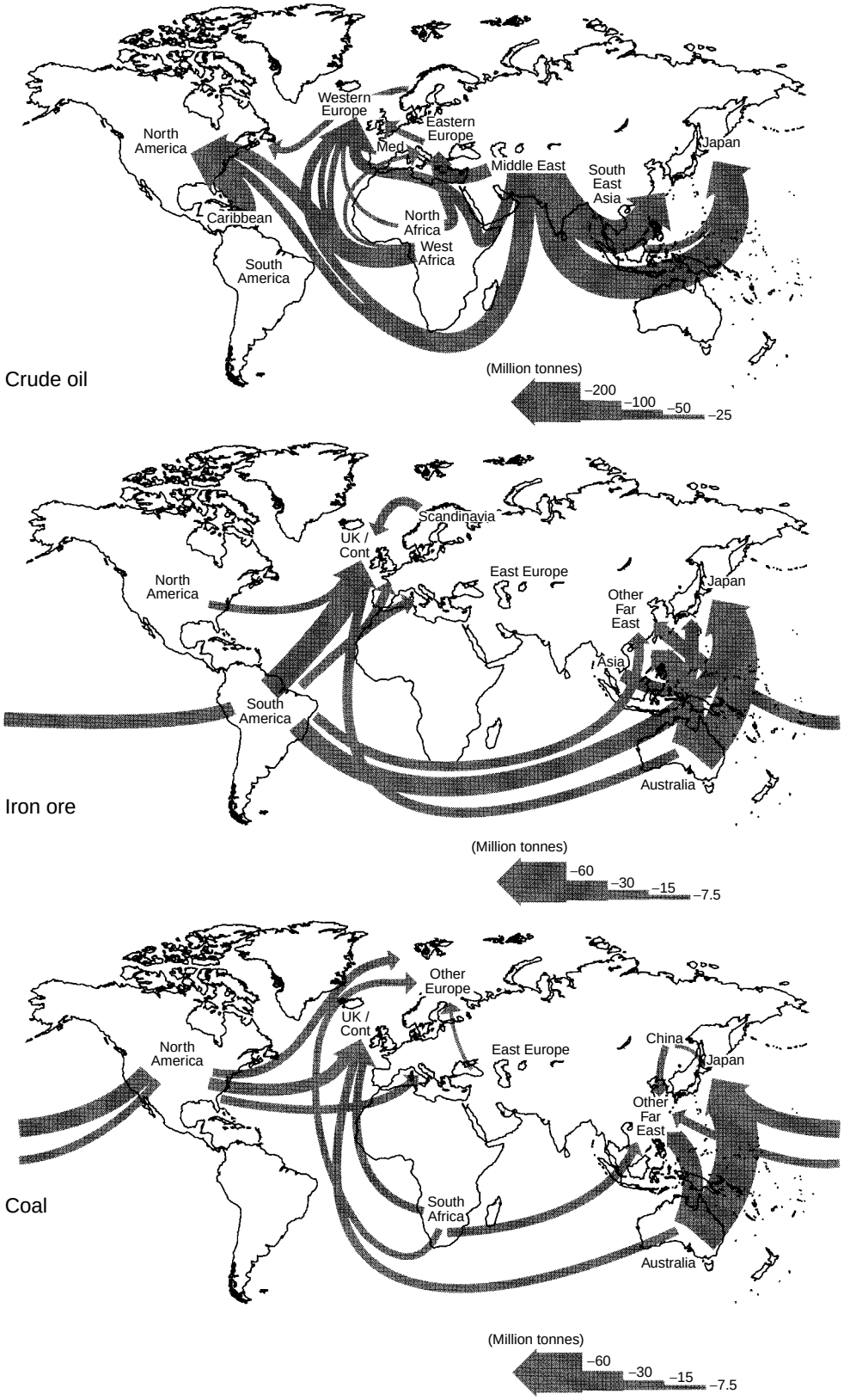


Figure 2 Major crude oil, iron ore and coal movements, 1995. (Reproduced using data from FR, 1999)

Asia (Figure 2). Western Europe augmented its Middle Eastern supplies with those from the Near East, North and West Africa and North America from the Caribbean, West Africa and the North Sea.

In 1995, sea-borne trade in iron ore, totaling 402 million tonnes, was closely patterned on crude steel production. South America, notably Brazil, was the largest producer that supplied Europe, the largest market, supplemented by Australia. However, Australia s main markets were in Japan and other Asian countries (China, South Korea and Taiwan), each of which also received smaller amounts from South America.

International sea-borne trade in coal of 423 million tonnes in 1995 made it the largest dry bulk commodity trade. Australia was the major supplier to Japan and the Far East. It also supplemented flows to Europe that were largely drawn from North America and South Africa.

The emphasis on forecasting demands for these major commodities stemmed from the bulk shipping market adhering closely to the principles of classical economics. Trends in commodities are matched against the mid-year supply of available shipping tonnage, to identify the supply-demand gap. A comparison of the supply-demand gap with freight rates provides the basis of forecasting longer-term trends in the shipping markets. When the tonnage is tight, freight rates rise. Conversely, when it is slack, freight rates decline.

Since the late 1960s this simple world of the classical economist has been mired by environmental problems reflecting an apparent decline in the standards of bulk shipping. In 1967, the Torrey Canyon ran aground off the coast of southwest England, spilling 120 000 tons of oil into the sea (*see* Torrey Canyon, Volume 5). Photographs of birds covered with oil had a profound impact on public perception of environmental degradation. This disaster led the International Maritime Organization to adopt a series of measures aimed at preventing operational oil pollution and accidental pollution, reducing the impact of accidents, providing compensation and helping implementation.

Shipping safety and pollution from ships improved, but a series of disasters from the late 1980s highlighted the need for fresh actions. In 1989 the tanker, Exxon Valdez, spilled oil off the coast of Alaska (*see* Oil Industry: Cultural Impacts and Environmental Risk Tolerance, Volume 3; Exxon Valdez Oil Spill, Volume 5); between 1989 and 1991 several bulk ore carriers sank off the Australian coast; in 1992 oil was spilt by the tanker Aegean Sea off Spain; in 1993 the tanker Maersk Navigator collided with an empty Japanese tanker in the Straits of Malacca and the tanker Braer was stranded off the Shetland Isles.

This spate of shipping disasters led to investigations by ship classification societies that attributed the sinkings to physical causes of failure corrosion, damage to the side structure and local overloading. The Australian government s Ships of Shame Inquiry pinpointed low freight rates as the underlying cause of a reduction in maintenance and a loss in structural integrity (Australia, 1992). The high-risk ships that foundered off Western Australia were found to be over 15 years old, carried iron ore, registered with flag states with high casualty rates and lacked seaworthiness due to corrosion (BTCE, 1994). An identical picture was echoed by the British government s Donaldson Inquiry (1994) into the Braer stranding that also emphasized shipping and transport mismanagement as key factors (e.g., crew proficiency, the chain of command, and better training and avoidance of fatigue).

Following these inquiries, a package of regulations designed for safer ship operation and environmental protection has been developed (Figure 3). This package includes international conventions and codes agreed to by governments and implemented by flag states. Individual states have also implemented measures to compensate for flag states and ship owners unwilling to accept their responsibilities for ship maintenance, management and operation. Private organizations likely to suffer financial damage from sub-standard ships have also taken further steps to eliminate rogue ships and reduce the risk of accidents (e.g., oil companies have proposed tanker traffic safety zones). These measures stemming from the hostile, political and regulatory environment generated by the ships of shame has not succeeded in eliminating the real culprits. A more targeted approach focusing on attacking the defective vessels, which account for less than 25% of the world s fleet may be more effective.

Liner Shipping

The present pattern of long-distance, deep-sea liner shipping is reflected in the large volume of container movements along main street, with its distinct east west alignment (Figure 4). The next largest set of flows is the north south movements that link a number of hub ports or load centers on the east west route. Gateway hubs include Singapore between Asia and Australia; Rotterdam between Europe and Africa; and Los Angeles Long Beach and Miami between North and South America.

This pattern of deep-sea liner shipping was brought about by technological change, undermining the previous historic system in which shipping companies were organized into rigid cartels or conferences. They operated regular multi-port routes for general cargo using a complicated rate schedule. The high cost of handling conventional cargo under this system led to the progressive introduction of containerization. Although the costs of operating container vessels and shore-based handling equipment exceeded those of conventional shipping, the companies merging their conference rights and establishing consortia offset the high fixed costs. Newcomers, mostly based in East Asia, challenged these consortia, by introducing cheaper container ships, lower

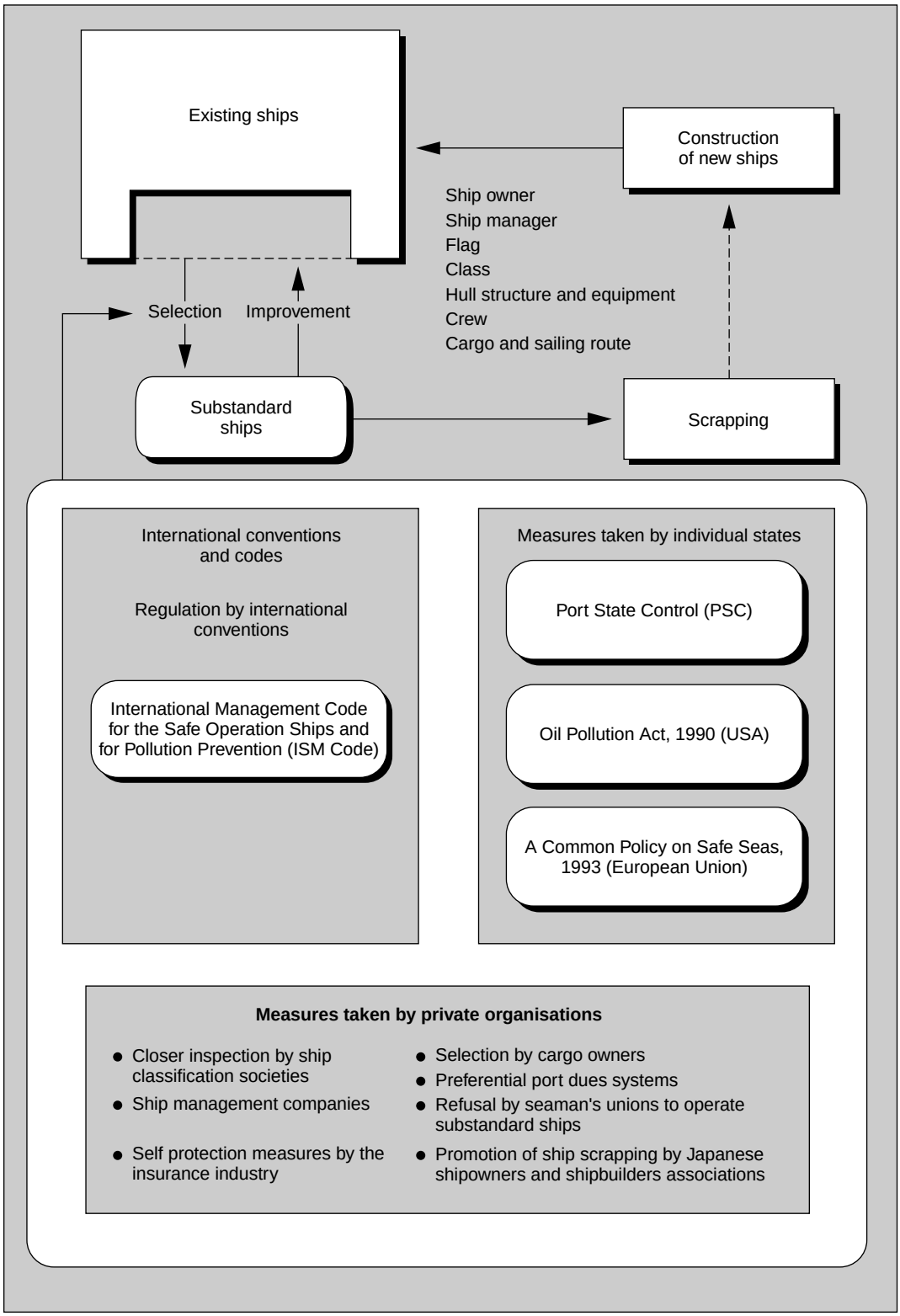


Figure 3 The process of identifying substandard ships. (Reproduced from MOL, 1995)

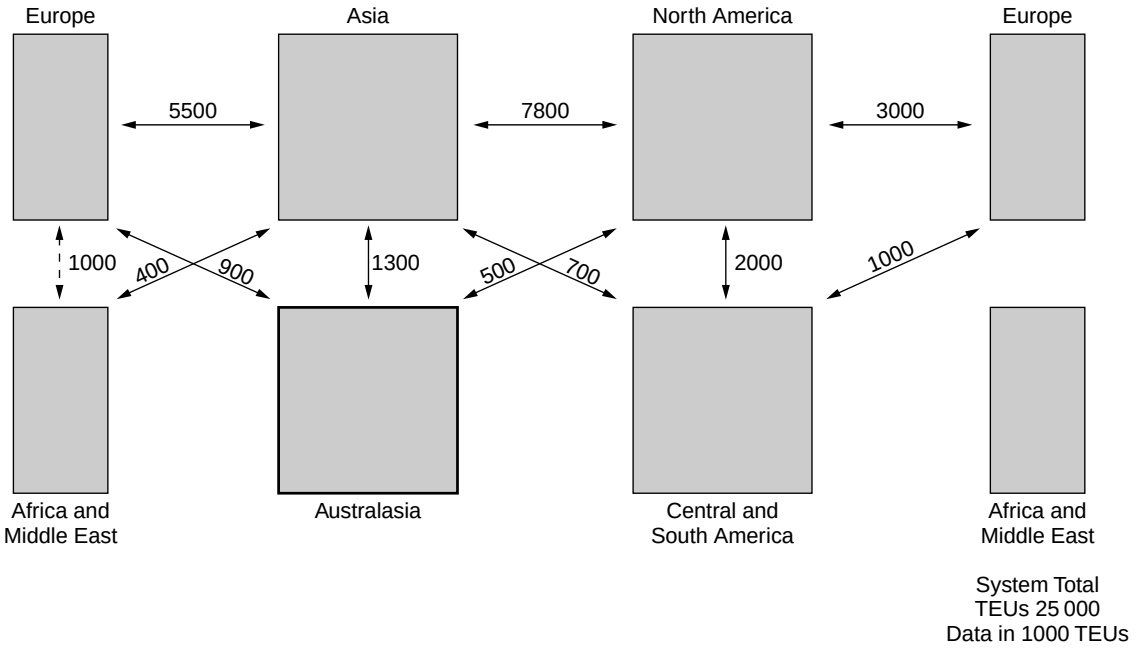


Figure 4 World container movements, 1996. TEU stands for twenty-foot equivalent units. (Reproduced using data from Iten, 1998)

freight rates and more flexible route configurations (e.g., round-the-world and pendulum services). The intensified competition led the established consortia to improve productivity and reduce costs. As these strategies did not provide a decisive competitive edge, many new entrants from Asia were allowed admittance into consortia.

These developments resulted in East Asia being recognized as a market in its own right, and direct Europe Far East and Far East North America services ensued. This strategy was supported by the introduction of ever-larger vessels, fixed day-of-the-week services, speedier turnaround, and faster and more reliable land-based transport connections. The shipping companies also used double stack container trains to extend their reach across North America to link their trans-Pacific and trans-Atlantic services. In the process, shipping companies were transformed into multimodal transport operators (MTOs) with interests in both sea and land transport.

During the 1990s, the MTOs ordered larger ships and formed transnational alliances that allowed the diversification of risks through the joint use of containers, container ships, terminals, feeder services, equipment and inland transport. Although individual alliance members maintained separate marketing identities, these groupings were designed to meet the demands of producers, such as IKEA, Nestlé, Philips, and Wilson Sporting Goods, for a global, door-to-door logistics package. Without the base load provided by these contracts, operators were at a distinct disadvantage and only about 20 of them could operate at this global scale.

Some operators have also taken shares in intercontinental cargo airlines to complement their shipping activities. These activities encompass both the activities of: (a) general cargo carriers handling high-value merchandise and operating on an airport-to-airport basis in collaboration with local freight forwarders, and (b) integrated express carriers, typified by DGS, Federal Express, TNT, and UPS, handling time-sensitive documents, parcels and packages.

The environmental impacts of changes in deep-sea shipping and air cargo driven by global forms of production are less marked than those associated with bulk shipping. Essentially, they are local, being concentrated around international gateways. Plans are afoot to move these gateways to more remote sites. Already, conventional container ports on the east west route are being replaced by a necklace of deep draft, offshore, mega transshipment hubs to accommodate the larger vessels (e.g., Gioia Tauro in Italy, Mina Raysut in Oman, Freeport in the Bahamas, and Manzanillo in Panama). Cargo will be transshipped from these hubs to smaller ports by feeder vessels.

Feeder Systems

Changes in the role of transport within the production process has resulted in the increased integration of intercontinental transport activities with national transport systems, through the optimal development of intermodal transport operations capable of handling intermediate and final product movements. Mega-carriers have linked their intercontinental sea and air container operations, through gateway

hubs, with domestic road, rail, water and air transport systems to seek full efficiency gains as part of just-in-time systems and reduced stockholdings.

This process of integrating national economies into the global market place has been facilitated by the deregulation and privatization of national transport sectors, which hitherto had been used as instruments of regional industrial and social policies. This change has allowed increased foreign participation. Consequently, longer-haul, container freight systems cutting across national borders are being developed through the EU, and the North American Free Trade Agreement. These pan-continental systems are still embryonic in Asia (though progress has been made in China). The full range of cost savings from these multimodal activities between firms operating at a global scale has yet to be attained. Investments in key bottlenecks are still required. Gaps in international transport networks have still to be overcome, particularly in Africa and the Middle East, Australasia and Central and South America. All have utopian plans for integrated transport networks long-distance highways, railways, land bridges and inland waterways to open up their continents.

Over time, there has been a shift towards movements by road and air, at the expense of water and rail transport (trends which the EU is trying to reverse, given the expected saturation of roads and airports). Most growth has been focused on short-distance, frequent delivery, road-based local systems handling final product movements through new logistical systems. Heavier road vehicles have also had an impact in urban areas though their intrusion has been perceived largely in terms of social impacts (noise and safety) rather than as noxious gas emissions. Efforts to mitigate the effects of urban goods movement have included the provision of loading spaces, peak hour truck restrictions, transport corridors to port areas, restrictions on truck size, hazardous material truck routes and urban distribution centers (Ogden, 1992). The proper pricing of freight and passenger road traffic may offer the best way of overcoming conflicts between economic and environmental objectives.

PASSENGERS

Intercontinental passenger services by rail and road are possible between Asia and Europe and Africa, and North and Central and South America. However, air transport has dominated intercontinental passenger movements since the introduction of jet aircraft in the 1960s, which reduced sea passenger services to leisure travel, except for short-distance trips. Air, rail and, on occasions, water transport provide regional connections with intercontinental airport gateways but road transport is important and dominates urban connections.

Air

Intercontinental air passenger traffic patterns, like those of deep-sea liner shipping, exhibited a strong east west pattern of flows along main street in passenger kilometers during 1998 (Figure 5). Again, north south flows are closely linked with hub airports on main street. These gateway hubs include Los Angeles and Miami between North

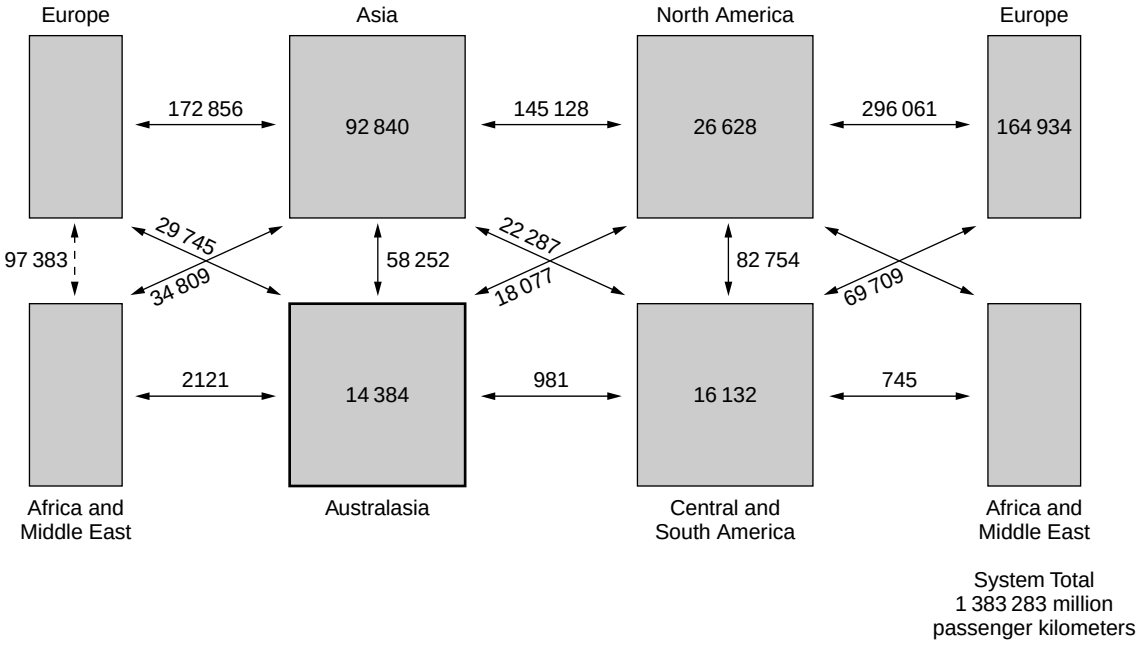


Figure 5 International air traffic on scheduled flights, 1998. (Reproduced using data from IATA, 1998)

and South America; Singapore and, to a lesser extent, Hong Kong and Tokyo between Asia and Australasia; and London, Paris, Rome and Athens between Europe and Africa. The hub function is used by resident airlines to take advantage of economies of traffic density.

There is, however, no freedom of the skies in air transport to match the freedom of the seas in deep-sea liner shipping. Although passenger airlines operate commercially, they do so under a franchise from national governments. Traffic rights are allocated by national governments through bilateral air service agreements established under the 1944 Chicago Convention, and have the status of treaties. While airlines can choose the type of aircraft and schedules, their route structures, seat numbers, number and frequency of flights and seat prices are dictated by these bilateral agreements.

Since the late 1970s, the airline industry has been restructuring to accommodate the demands of multinational corporations for seamless services and greater international social mobility involving migrants and seasonal tourist traffic. As traffic rights are not easily renegotiated, there has been pressure since the 1978 US Airline Deregulation Act from the US government to progressively liberalize the right to fly on specific international routes. There has also been an associated drive to privatize or commercialize national flag carriers still in public ownership. This strategy allows the entry of low-cost carriers and reduces state subsidies. Some airports have also been privatized.

The introduction of progressively larger and faster aircraft since the 1980s, especially the Boeing 747-series and 777-series and the 555 seat super jumbo the Airbus A3XX has promoted organizational changes. Managements have sought to transform their airlines into very large companies or mega-carriers that offer long, non-stop intercontinental flights between metropolitan regions. The corporations have expanded their route networks and dispersed operations to different parts of the globe to meet the greater competition between airlines unleashed by deregulation and the need to increase returns to shareholders.

Globalization, deregulation and privatization have profoundly affected the air industry s workforce as they provide the basis for exploiting an international division of labor, because airlines recruit cabin staff from countries with lower labor costs. Some workers have experienced pressures aimed at cutting labor costs through the threat of outsourcing to meet global industry benchmarks (e.g., catering, maintenance, cleaning, ground handling and information technology). Others have been urged to be flexible and intensify their labor effort through quality programs so that higher productivity and improved service quality can be provided at key points of customer contact.

Further restructuring has occurred as mega-carriers have experimented with strategic alliances to grow globally by evading the myriad of government restraints on access to markets and restrictions on the majority ownership of national carriers. Since the 1980s, alliances have grown ten-fold to number 500 in May 1998. The growth of intercontinental alliances among this number has been slow because of limited progress with deregulation on a global scale, and the difficulties of securing partners in dispersed markets. However, four major global alliances comprising most of the world s top 25 carriers have been formed, which collectively account for half of the world s passenger market. They are Star (12 members including United Airlines and Lufthansa); Oneworld (nine members including British Airways); Atlantic Excellence (three members including Air France and Delta); Qualiflyer (eight members including Swissair); and Wings (four members including Northwest and KLM Royal Dutch Airlines). Although further realignments among alliances are possible, real mergers are still precluded by bilateral regulations.

Global alliances have allowed large carriers, who dominate key airports, to generate hub-and-spoke operations in a more open market, in a way not dissimilar to those in deep-sea liner shipping. Mega-carriers have gained from economies of scale and scope through the expansion of their networks and the use of computer reservation systems to consolidate their market position. Certain airports have been advantaged when they have become most-favored hubs for an alliance. Some air passengers have enjoyed greater convenience and service quality and reduced fares from the restructuring of the airline industry, though gains to other passengers and some workers are less clear. The objective of a seamless service for transferring passengers from one carrier to another still remains a distant goal. It depends on the international air transport network being integrated with urban and regional planning. Much attention is being paid to moving airports from urban areas to mitigate the effects of noise and overcome the problem of curfews.

Feeder Systems

The car dominates interurban trips, with the road to and from the airport being often the most heavily trafficked. Use of rail and air transport increases with the distance and duration of the journey. Car usage hinges on the size of the city-region. The modal split between car and bus/rail depends on the efficiency of the mass transit system and the general policy for land use, parking, traffic control and general infrastructure capacity. This trend towards increased car use in urban and regional areas affects the mix of work and leisure travel. Despite energy and pollution problems, car use has continued to expand, even during periods of high fuel prices. The major task facing planners is safety problems as there are no medium-term problems with energy.

Exceptions are the major metropolitan city-regions where congestion and pollution from traffic are serious issues,

particularly in the suburbs. Traffic calming measures and noise limits have been employed. Urban mass transit and suburban rail transport offer alternatives to some car use for individual transport. If the railways, underground and light transit are to attract passengers from cars, these services not only have to offer low tariffs but high quality services (frequency, speed and comfort). Attractive interurban services are provided by high-speed trains (Japan s Shinkansen and the French TGV). Planners in the EU want to develop regional inter-metropolitan services, typified by Eurostar (Paris Brussels London) and Thalys (Paris Brussels Amsterdam). Separating railway undertakings into track and rail operators would assist the efficiency of this process. Efforts to link integrated railways with regional air transport networks have also been made. Regional air travel has also been assisted by open skies accords within the EU, which match those adopted in North America. These have led to the introduction of low cost airlines in Europe such as buzz, easyJet, Go, Ryanair and Virgin Express.

CONCLUSION

Should transport follow the demands of business and the public? Contradictory pressures are placed on policy makers as we all wish to be protected from transport s environmental nuisances where we live, but want quality infrastructure to save time when we travel. There is no sign of any slackening of long and short-distance transport demand due to economic globalization and leisure travel.

Salvation is still being sought in technological advances. Teleworking and the virtualization of our economies may mean less travel. However, communications technologies may merely complement or encourage further personal transport because there is no immediate substitute for face-to-face contact in transferring complex knowledge. Thus, there is no guarantee that information technology will diminish our use of transport until a saturation level is reached in face-to-face contact (Ciccantell and Bunker, 1998; Nijkamp *et al.*, 1998).

We will still have to reconcile our conflicting aspirations between respecting the environment and meeting economic needs, to achieve sustainable development. There are contrasting opinions as to whether our current global transport system is sustainable. Some think technological developments and building more infrastructure will be sufficient while others, conscious of current traffic trends in lower income countries, do not. Even where the imperative of becoming more environmentally conscious is accepted, many motor vehicle restraint measures to achieve this end are unpopular (e.g., road pricing adopted in Singapore). The challenge will be to make these measures palatable at a political level.

See also: Aircraft Emissions, Volume 3; Shipping: Harbors and Ports, Volume 3; Transport Infrastructure, Volume 3; Urban Sulfurous and Photochemical Smog, Volume 3.

REFERENCES

Australia (1992) *Inquiry into Ship Safety, Report from the House of Representatives Standing Committee on Transport Communications and Infrastructure*, Canberra, Australian Government Publishing Service.

BTCE (1994) *Structural Failure of Large Bulk Ships*, Bureau of Transport and Communications Economics Report 85, Canberra, Australian Government Publishing Service.

Ciccantell, P and Bunker, S G, eds (1998) *Space and Transport in the World-system*, Greenwood Press, Westport, CT.

CIY (1998) *Containerisation International Yearbook 1998*, Emap Business Communications, London.

DETR (1998) *Transport Statistics Great Britain 1998 Edition*, Department of the Environment, Transport and the Regions, The Stationery Office, London.

Donaldson Inquiry (1994) *Safer Ships, Cleaner Seas*, Report of Lord Donaldson s Inquiry into the Prevention of Pollution from Merchant Shipping, CMD 2560, HMSO, London.

EB (1999) *99 Britannica Book of the Year*, Encyclopaedia Britannica, Chicago, IL.

ECMT (1995) *European Transport Trends and Infrastructural Needs*, European Conference of Ministers of Transport, Paris.

ECMT (1999) *Which Changes for Transport in the Next Century*? Introductory Reports and Summary of Discussions, 14th International Symposium on Theory and Practice in Transport Economics, Innsbruck, 21 23 October 1997, European Conference of Ministers of Transport, Paris.

FR (1999) *Fearnley's World Bulk Trades, 1998: an Analysis of 1997*, Fearnresearch, Oslo.

Gerondeau, C (1997) *Transport in Europe*, Artech House, Norwood, MA.

IATA (1998) *World Air Transport Statistics, No. 42: WATS 6/98*, International Air Transport Association, Montreal.

IATA (1999) *World Air Transport Statistics, No. 43: WATS 6/99*, International Air Transport Association, Montreal.

IRF (1996) *World Road Statistics*, International Road Federation, Geneva.

IRTU (1996) *Handbook of International Road Transport*, International Road Transport Union, Geneva.

Iten, J (1998) Prospects of Success in Liner Shipping, *Int. Transp. J.*, 42/98, 46 47.

IUR (1999) *International Railway Statistics*, International Union of Railways, Paris.

JUTS (1998) *Jane's Urban Transport Systems, 1998- 99*, 17th edition, Jane s Information Group, Coulsdon.

MOL (1995) Safety Management: Safe Operation and Environmental Protection, Mitsui O.S.K. Lines, Tokyo.

Nijkamp, P, Rienstra, S A, and Vleugel, J M (1998) *Transportation Planning and the Future*, John Wiley & Sons, Chichester.

Ogden, K W (1992) *Urban Goods Movement: a Guide to Policy and Planning*, Aldershot, Ashgate.

TMC (1997) The Automobile Industry: Toyota and the World 1997, Toyota Motor Corporation, Tokyo.

UN (1997) *Statistical Yearbook*, 42nd Issue, Department of Economic and Social Affairs, Statistics Division, UN, New York.
UNCTAD (1998) *Review of Maritime Transport 1998*, UN, New York and Geneva.

Transport Infrastructure

William Young
Monash University, Victoria, Australia

Transport infrastructure provides the life-blood of our society. Ensuring it is sustained at a high level in the new millennium is essential for our quality of life to continue. Transport provides the ability for people to move (mobility) and to partake in activities (accessibility). It is a derived demand. That is, it provides an opportunity to satisfy another demand. Transport is rarely undertaken for its own sake. It is undertaken to carry out some other task. Within the context of a derived demand, transport is also a service, which can be purchased and used. Notwithstanding the importance of transport in providing quality of life, there are costs associated with its provision. These include: construction and maintenance of transport infrastructure; the opportunity cost associated with the land used to provide transport; the growth of cities and the reduction of agricultural land; road trauma; the use of energy resources; emissions and their impact on human health, the biosphere and the build-up of greenhouse gases; and stratosphere ozone depletion. The trade-off between the bene ts of transport (mobility, accessibility) and it costs (safety, pollution, energy use) cannot be overlooked and must be incorporated into transport decision-making. This paper addresses the global impacts of transport on the environment. It provides a history of transport infrastructure development, emphasizing the progressive improvement in mobility and access. It explores the impact of transport on the distribution of urban activity, energy, greenhouse gas emissions and ozone depletion. The paper concludes with discussion on transport infrastructure policies and how they relate to the environment.

THE BENEFITS OF TRANSPORT INFRASTRUCTURE

Improving mobility and accessibility have driven the development of transport infrastructure throughout human existence. Transport infrastructure has moved through a number of revolutions as people have created new ways of increasing mobility and accessibility. The main improvements have been in speed, comfort, reliability, flexibility, and cost of travel. The transport system most desired by the community dominates at any one point in time. Previous transport modes do not become extinct; their use may persist long after they have been replaced by new technology.

Land Travel

About four million years ago early humans were distinguished from other great apes by their erect walking style. Human creativity enabled them to built crude toboggans in 7000 BC. Draught animals revolutionized transport by dragging loads in 6000 BC. The development of the wheel dates to 3550 BC and two-wheel chariots were used in 2000 BC (McWhirter, 1999).

Roads and Cars

For many years wheeled vehicles moved over open land. Tracks were developed when vehicle movement occurred regularly. The development of roads substantially increased the speed to travel. The development (52 AD) of a road network of over 85 000 km enabled the Romans to move their military vehicles rapidly from one location to another. The Silk Road was forged in 130 AD, running 6400 km between Shanghai and Xian. The Incas (1534) developed an impressive early road system in South America. The development in travel by land allowed the exploration, settlement and development of continents.

In the 18th century, new methods of road building were pioneered in the UK and Europe. A series of layers of stone was used to even out the road surface and increase the speed of movement. Asphalt, a mixture of stones and bitumen, was first used in 1835 in France (Lay, 1992). Traffic lights were first installed in 1868 with freeways, motorways and parkways introduced in the 1900s (McWhirter, 1999). Improved civil engineering construction techniques, materials and design have allowed bridges to increase from spans of 4.3 m (14 AD) to 1780 m in 1998 (Bennett, 1998), and tunnels to increase from 1.0 km in 550 BC, to the tunnel under the English Channel (37.98 km) built in 1994 (McWhirter, 1999). These improvements allow people to move more quickly and to have an increasing set of options available to them, hence improving both mobility and accessibility.

The automobile s history began with the steam-powered gun carriage built by the French military engineer Nicholas Cugnot in 1769. Gottlieb Daimler and Karl Benz built the first petrol-driven internal combustion engine in 1885. By 1895, there were 300 cars manufactured in the US. Mass production of cars was started in 1901. The number of vehicles in the world has grown from 35 million in 1930 to 666 million in 1997. The availability of cars varies across the world. The US had 760 vehicles per 1000 people in 1997. Italy, Canada, Germany and the UK had 575,

557, 536 and 478, respectively (McWhirter, 1999). The flexibility, comfort, convenience, reliability and speed of the automobile have ensured its adoption by many people as a preferred method of travel. People have shown their desire to purchase mobility and accessibility. Moderate long-term economic growth of a few percent a year contributes to a growth in car ownership and mobility (Nijkamp and Perrels, 1991). Clearly there is still considerable potential for growth in car use in countries like China and India, which have only eight and six vehicles per 1000 population, respectively.

Rail

Wheels running on track can be traced back to the Babylons in 2250 BC. The railway was first used in 1681. Its rapid growth in the period 1840 1900 allowed cheap mass transport. This rapid growth occurred because of, and supported a number of developments. The industrial revolution provided the mechanism for building rail infrastructure. Rail allowed the movement of large volumes of workers from suburban home sites to industrial areas. Rail infrastructure began to decline as cheap reliable methods of road travel were introduced.

Pipelines

Aquaducts and pipelines provide an alternative to road and rail for transporting water, gas and oil.

Water Travel

Even in early days, people could ford rivers and reach some islands at low tide. The use of logs, row-boats, papyrus reeds lashed together (3500 BC), sails (3500 BC), improved navigation techniques (first compass 1150; satellites, global positioning systems in the 1990s), steam engines (1800), iron hulls (1850), turbines (1880), hydrofoils (1910) and hovercrafts (1955) have provided substantial increases in speed and facilitated the movement of large numbers of people (McWhirter, 1999). The development of travel by sea allowed the exploration, settlement and development of remote regions of the earth.

Only natural landforms and nature (sea currents, ice, wind), at sea limit the movement of boats. The access to land is constrained by the size, depth, and efficiency of ports. Freight movement has been made more efficient with the introduction of containers and more recently improved work practices and technological innovation (crane design, etc.). Passenger transport by sea has decreased over the last few decades but was the major form of movement between countries separated by sea up until the first half of the 20th century.

Harbors, lakes and rivers facilitated movement within continents. The major transport infrastructure, other than ports, developed to facilitate water travel was canals and lochs. Canals can be traced back to 5 BC, where they were constructed in Iraq. The world s first great canal was China s Great Canal, started in 6 AD and completed in 1283. It was 1700 km long. Canals were introduced in approximately 1760 in the US. Their use grew and more canals were constructed until land based rail systems replaced many of them.

This history of sea travel presents a move towards increased speed (below one knot to in excess of 60 knots (111 km hour$^{\bullet 1}$) presently), increased carrying capacity (below 1 m^3 to over 250 000 m^3 presently), at a decreasing cost per unit moved.

Air

Initially people movement in the air involved balloons (1793). Kites were used in China in 1000 BC, but the first free flight took place in the summer of 1853 in the UK using a glider. The first heavier-than-air flight was by the Wright brothers in 1903. Jet engines were introduced in 1937 and now there are Boeing 747 jumbo jets with a seating capacity of 340 persons, and the possibility of 1000-passenger double deck airbuses being introduced in 2003 (McWhirter, 1999). These improvements are reducing the cost of travel. For instance, in 1945 an air trip between Sydney and London represented approximately 130 weeks work for the average person, in 1998 it took 2.5 weeks to earn the fare. Clearly air travel is increasing mobility and accessibility through increased speed (below 1 km hour$^{\bullet 1}$ to over 700 km hour$^{\bullet 1}$), passenger carrying capacity (one person to over 340 passengers) and decreasing cost per trip.

Recent developments in air travel have moved humans into space. Space travel moved from the first manned space mission in 1961 to regular space shuttles in 1981. Travel in space is allowing the exploration, and ultimately, the settlement and development of other parts of the universe.

These developments provide people with greater choice, and increasing use of air travel indicates that people want this opportunity. In contrast with automobile travel, the movement of airplanes is not constrained by roads, although air traffic is coordinated by air traffic control. The airports provide a major infrastructure constraint and cost. Considerable expense is directed at planning, building and maintaining these facilities.

Communications

Transport enables the movement of people and goods. An important dimension of transport is the facilitation of communication. Substitutes for communication-related travel involve semaphore (140 BC) and telegraph (1844), with teleconferencing, e-mail, internet, world wide web

and facsimile machines revolutionizing communications in the 1990s. Considerable debate is directed at whether communications will decrease the need for travel. Evidence is present on both sides (Salomon, 1998) with the clear conclusion that communications will decrease the need for some travel (e.g., work) but will increase the need for other forms of travel (e.g., freight).

Remarks

The history of transport and communications is categorized by improving mobility and accessibility through increased speed, comfort, reliability and affordability. The desire for increased mobility and accessibility is unlikely to decrease. However, the costs associated with the provision of transport infrastructure, along with the social and environmental costs, may result in mobility and hence accessibility being constrained in some manner. The trade-off between the benefits of transport (mobility, accessibility) and the costs (safety, pollution, and energy use) cannot be overlooked and must be incorporated into transport decision-making.

THE COSTS OF TRANSPORT INFRASTRUCTURE

There is a clear interaction between the world s social, economic and physical environment. A healthy economy facilitates social and physical growth. A healthy social environment encourages economic and physical development. A healthy physical environment provides a basis for sustaining economic and social activities. Within this broad interaction, transport has a direct impact on the distribution of land use and the environment. The interaction between human activity, transport and the environment is shown in Figure 1.

Transport's Interaction With Land Use

Transport infrastructure provides the platform on which travel decisions are made. Activity between land uses generates the trips made on the transport system. For instance, a university may create one trip per day per student, while a shopping center may generate one trip per day for each 100 m^2 of shopping space. The provision of transport infrastructure influences this demand, in that better provision of transport will encourage more people to travel. This influence has been formalized in the context of a law of the road, which states that building more and better quality roads will increase traffic.

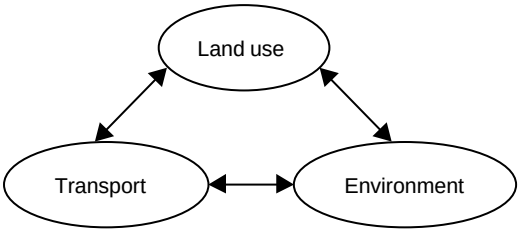


Figure 1 Land use/transport/environment interaction

Equally important is the impact of transport on land use. This comes in two forms. First, the provision of transport takes up land and hence will reduce the amount of land that can be used for other purposes. Some types of transport infrastructure are more land intensive. Secondly, transport influences the distribution of land use. A high speed, flexible transport system allows people to choose from a variety of homes and workplaces. The low-density nature of cities in Australia and North America are consequent on the rapid growth of these cities at a time when residents had ready access to an automobile and a considerable road network.

Hence, the provision of a high quality transport system may result in a decrease in the land available for other activities. Since many cities are often built on prime agricultural land, the progressive spread of cities and the use of land for transport means less high quality land is available for the production of food and flora. The reduction in flora will have a direct impact on the quality of life in cities.

The Impacts of Transport on the Environment

The impacts of transport on the environment are many. These impacts may occur locally or globally. Local impacts may include automobile accidents (including property damage, injuries and deaths), noise, vibration, road trauma, photochemical air pollution (smog), lead pollution, air-borne particulates, visual intrusion, roadside pollution, heavy vehicle movement, and construction. Global impacts are related to the use of arable land, energy, and emissions (as related to climate warming, stratospheric ozone depletion and acid rain). This section will focus primarily of the global energy and emissions questions.

Energy

Transport is one of the many activities that consume energy. Hence when considering energy use, the broad range of consumption activities should be considered. The interactions between the long term creation of energy sources, increases in human consumption of energy, changing population preference, and economic prosperity mean that existing forms of non-renewable energy may eventually disappear or become prohibitively expensive as they become more scarce. For instance, transport accounts for approximately 27% of the total energy consumed in Australia (BTCE, 1991). Road transport accounts for 23.6%, rail

1.3%, air 1.8%, and sea 0.7%. The present increase in fuel use is approximately 2% per annum. If energy usage in transport continues to grow, as predicted, at 2% per annum, fuel consumption will be approximately 50% higher in only 20 years.

The spread of fuel usage is not consistent across the world. Newman and Kenworthy (1989) compared the transport fuel usage in 32 cities. They found that the average resident in a US city uses approximately twice as much fuel as Australian city residents, 4 times that of a European resident and 10 times that of residents in Hong Kong, Tokyo and Singapore. This discrepancy is influenced by the fuel efficiency of vehicles, the private/public transport split, cultural norms, social expectations and the personal mobility of the residents.

Improving the fuel efficiency of cars is a key factor in reducing fuel consumption. Consideration of this can be divided into two parts: reducing the fuel consumption of new vehicles and removing old vehicles from the fleet or improving their fuel consumption. The introduction of vehicles with lower fuel consumption $km^{\bullet 1}$ may reduce overall fuel consumption. General decreases in fuel consumption in European and North American countries are of the order of 1.5% per annum with average levels of approximately 8.0 litres per 100 km travelled (BTCE, 1991, 1996). These improvements in fuel economy are not reflected directly in the consumption of the total car fleet since not all cars are new. Reductions in the fuel consumption of the vehicle fleet of 0.8% have been measured in the same regions as above with the average consumption being of the order of 10 litres per 100 km travelled (BTCE, 1991, 1996).

Even within the context of moving people and goods, transport fuel consumption is only part of the total fuel cycle which also includes: the amount of fuel used to deliver energy to the end use; the generation and transmission losses for electricity; and the energy used in fuel extraction, refining and distribution. In the case of liquid fuels used in transport, adjustment to the full fuel cycle basis is estimated by adding around 10% to the end-use energy consumption.

The desire for travel by motor vehicles is unlikely to diminish because of the quality of the mobility and accessibility it offers, even though the frequency of traffic gridlocks may increase. Hence the method of propulsion needs exploration. Alternate fuels, electrical power or solar power may provide opportunities to reduce energy consumption by cars.

Emissions

Vehicle emissions include carbon dioxide (CO_2), carbon monoxide (CO), hydrocarbons (HC, including methane (CH_4)) and other volatile organic compounds (e.g., aldehydes), nitrous oxides (N_2O), nitrogen oxides (NO_x), lead (Pb), particulate matter of varying composition, and chlorofluorocarbons (CFCs). These can have detrimental effects on humans and the biosphere. These impacts can be local (smog, increasing toxicity, respiratory infections), regional (formation of reactive substances, acid rain, suspended particles reducing visibility) and global (depletion of stratospheric ozone layer, greenhouse effect). Weather conditions (such as wind), sunlight levels (photochemical pollutants) and local geography (such as hills and valleys) complicate the occurrence and reactions involving emissions and their impact on a region. The sources, sinks, health and welfare effects, greenhouse gas impacts, and basic chemical interactions of transport emissions, illustrate the need to consider all pollutants when looking at the impacts of particular changes, not just those relating to the major greenhouse gas (CO_2). The impacts of each greenhouse gas are:

- Equilibrium in the global carbon cycle is necessary for continuous human existence. Urban transport contributes carbon dioxide to the atmosphere. The other contributors are agriculture, commercial firms, manufacturing, mining, and residential. The automobile is the main source of transport carbon dioxide emissions in many countries. Carbon dioxide is non-toxic, but is a major contributor to the greenhouse effect. It can also dissolve to produce acid.
- Carbon monoxide primarily arises from the incomplete combustion of hydrocarbons. Although a rich air fuel mixture produces high levels of carbon monoxide emissions, they are also produced with lean and stoichiometric air fuel mixtures. Carbon monoxide interferes with the absorption of oxygen by red blood cells, impairing perception and muscle response. It has been associated with reduced work productivity through its impact on respiratory functions and circulation.
- Nitrogen oxides (NO, N_2O_2 and NO_2) are created by fossil fuel combustion. Natural sources of such gases are negligible, and motor vehicles create most urban concentrations. Nitrogen oxides are absorbed by plants and are significantly reduced by the use of catalytic converters on motor vehicles. Like CO they are considered an indirect greenhouse gas since they contribute to ozone formation. They cause lung damage and increase susceptibility to respiratory infections. They can damage plants and contribute to acid rain and brown pollutant haze.
- The main sources of CH_4 are fossil fuel combustion, release from coal mining and natural gas distribution, fermentation in livestock, wood fires, release from rice paddies and landfills. The main absorber for CH_4 is the hydroxyl radical, forming carbon dioxide and water. Reduction in concentrations of this radical has

increased the average time CH_4 emissions remain in the atmosphere to 11 years. This may account for current CH_4 levels being double those of pre-automobile days. Methane is a direct greenhouse gas and also acts as a minor ozone precursor.

- Non-methane hydrocarbon emissions have a variety of sources. They are created through incomplete combustion of fossil fuels in motor vehicles, evaporation of fuels in refineries, and in petrochemical manufacturing. Significant levels of non-methane hydrocarbons also come from the evaporation from fuel in motor vehicle tanks, carberators and crankcases. About half of these emissions come from motor vehicles.
- Motor vehicles are minor sources for nitrous oxide (N_2O).
- Sulfur oxides generally occur because of impurities in fossil fuels. Motor vehicle emissions account for about 15% of urban SO_x. These emissions generally stay in the air for about a week before being removed by rain, absorbed by vegetation or consumed by soil bacteria. They cause coughing and aggravate asthma, bronchitis and lung disease. Sulfur oxides are the main precursors to acid rain, which destroys vegetation, acidifies lakes and damages buildings.
- Stratospheric ozone depletion is facilitated by CFCs. These are man-made compounds that are used as solvents, refrigerator fluids, spray-can propellants and foam agents, due to their stability and safety. A major emission source has been motor vehicle air-conditioners. Since they are stable compounds, CFCs have a long atmospheric lifetime, and no natural sinks. High radiation levels are required to decompose the CFCs and release chlorine. In turn, the chlorine reacts strongly with stratospheric ozone, creating the Antarctic ozone hole (*see* several articles in Volume 1, e.g., Stratosphere, Ozone Trends, Volume 1). The Montreal Protocol called for a 50% reduction from the 1986 levels in the use of CFCs, by 1996. CFCs are being phased out in many countries, since current policy is to control losses of CFCs in their use in and servicing of air conditioners. However, the greenhouse contribution of their replacements may still be considerable (*see* several articles in Volume 4, e.g., Environmental Responses: an Overview, Volume 4).

Given the unlikely move towards lower levels of mobility and accessibility, there is a need to reduce emission levels through technological change or changes in fuel use.

BALANCING THE BENEFITS AND COSTS OF TRANSPORT INFRASTRUCTURE

Transport provides mobility and accessibility. These are desirable attributes, and the human race has spent many years trying to improve them. Unfortunately they come at a cost. For instance, increased speed increases the probability of being killed in an accident, traffic noise, emission levels, energy consumption, etc. To reduce these costs while maintaining the desirability aspects of transport is a challenge. The challenge should be seen within the overall constraint of sustainability. Sustainable development means using, conserving and enhancing the community s resources to increase and improve the total quality of life to guarantee the maintenance of ecological processes. In calling for a sustainable transport solution, the scale of the task can be reduced, transport can be made more efficient and/or alternate methods of propulsion can be used. These options can be facilitated by:

A Vision

An overriding vision of the future urban area is required (Black, 1992). A key issue is the development of assessment and evaluation procedures that take into account economic, social, and environmental factors to ensure that the best choices and solutions for the future are made. Some urban development objectives are: cater for population preference, increase efficiency, encourage economic growth, provide equitable access to urban services, reduce energy consumption, and improve the urban environment. It is obvious that there are conflicts amongst these objectives. Decreasing environmental pollution may result in increased energy consumption. Catering for population preference may result in increased urban inefficiency. Trading off these objectives may require putting sustainability into urban development decisions. A clear focus on the main areas of growth in travel (road and air) may also provide a clear direction for such evaluation and policies (BTCE, 1996).

Pricing

A key link between transport infrastructure provision and the demand for its use is the price paid for access to transport. Hensher (1993) points to the need to correctly price the social and environmental costs associated with transport use, in order that transport users may receive signals about the impact of their travel decisions. The BTCE (1996) points to road user charges which recognize the external costs associated with the environment as an effective measure for reducing greenhouse emissions and energy use. Competitive neutral pricing, where each mode bears the full cost of its operation, could be facilitated by road tolls, congestion pricing and road pricing (Taylor and Taplin, 1998).

Integration

The integration of transport and land-use planning, development and use, will ensure more efficient use of transport infrastructure. In countries with strong economies

and flexible social regimes, people often wish mobility and access to desired activities. The appropriate use of the transport system may require an integrated multi-modal view of transport (road, rail, tram, bike, bus, and pedestrian). The development of multi-centered cities with strategically located transport interchanges may optimize transport use. Locating workplaces, activity centers, service and other facilities closer to where people live may increase the sustainability of a city. Urban consolidation may allow more efficient use of public transport use for labor-intensive regional centers. Urban villages could reduce travel through the location of service industries closer to the users.

Technology

Technology can influence the driver, vehicle and road system, resulting in reductions in emissions and energy use. The reduction in fuel consumption has been discussed earlier. Vehicle control mechanisms can also reduce fuel consumption. ITS information systems can reduce travel time and provide information on driving behavior and route choice. Technology change will play a large part in developing a sustainable transport system.

Overlaid on these approaches to developing a sustainable future are people making decisions in a constrained environment. The general community, business, politicians, governments, planners and developers make transport decisions. Further, the ownership, operation and maintenance of infrastructure by the private sector is well underway. The responsibility for developing a more sustainable and equitable transport system lies with all members of the community. They clearly have a role in determining the vision since it is they and their descendants who must live with the consequences of their actions.

REFERENCES

Bennett, D (1998) *The Creation of Bridges*, Lothain Books, Tasmania.

Black, J (1992) *Sustainable Development and the Urban Transport Planning and Policy Process*, Biennial Congress of Royal Australian Planning Institute, Canberra, Australia.

BTCE (1991) *Reducing Greenhouse Gas Emissions from Transport: Some Tax Policy Options*, Submission to the Industry Commission Inquiry Into Costs and Benefits of Reducing Greenhouse Gas emissions, Bureau of Transport and Communications Economics, Canberra, Australia.

BTCE (1996) *Transport and Greenhouse: Cost and Options for Reduced Emissions*, Bureau of Transport Economics Report 94, Canberra, Australia.

Hensher, D A (1993) *Socially and Environmentally Appropriate Urban Futures for the Motor Car*, Transportation, 20(1), 1 19.

Lay, M G (1992) *Ways of the World: a History of World's Roads and the Vehicles that used them*, Rutgers University Press, New Brunswick, NJ.

McWhirter, N (1999) *Book of Millennium Records*, Virgin Publishing, London.

Newman, P and Kenworthy, J (1989) *Cities and Automobile Dependence*, Gower Allen and Unwin, London.

Nijkamp, P and Perrels, A (1991) New Transport Systems in Europe: a Strategic Exploration, in *Cities of the 21st Century: New Technologies and Spatial Systems*, eds J Brotchie, M Batty, P Hall, and P Newton, Halsted Press, New York.

OOE (1993) *Victorian Externalities Study*, Office of Environment, Melbourne, Australia.

Salomon, I (1998) Technology Change and Social Forecasting: the Case of Telecommunication as a Travel Substitute, *Transp. Res. C*, 6(1/2), 17 45.

Taylor, M A P and Taplin, J H E (1998) Travel Demand, Energy and Pollution Impacts of Alternate Congestion Pricing Regimes, *Proceedings of the International Conference on Transportation into the next Millennium, Singapore*, 1 12.

U

Urban Climate and Respiratory Disease

Jonathan A Patz
Johns Hopkins School of Public Health, Baltimore, MD, USA

As the 21st century begins, the concern about many local and regional air pollutants has become less acute in the United States, Canada, and other developed nations. The implementation of regulatory measures and the development of cleaner technologies have resulted in the improvement of some indicators of air quality. Many developing countries have a different experience because they have not yet developed or adopted many of the regulatory measures of the industrialized world to reduce the levels of local and regional air pollution (United Nations Environment Programme (1991); World Bank (1997)). However, for both developed and developing nations, evidence is growing that global climate changes, caused by the emission of greenhouse gases may exacerbate the impacts of air pollution on human health and the environment.

HISTORICAL PERSPECTIVE ON AIR POLLUTION

From Local to Regional to Global

Historically, the developed nations have gone through a series of transitions with respect to air pollution problems. By the mid-1800s, the combustion of fossil fuels had fouled the urban environment in industrial cities. At that time, smoke, soot (particles), sulfur dioxide (SO_2), and nitrogen dioxide (NO_2) were the predominant substances of concern. These pollutants caused respiratory irritation, hindered visibility, and blackened buildings. Later, the advent of the automobile introduced the problem of increasing concentrations of nitrogen oxides (NO_x) and volatile organic compounds (VOCs) in the urban environment. As chemical precursors to ground-level ozone, these pollutants resulted in a new type of photochemical smog, consisting primarily of ozone whose formation was catalyzed by sunlight, first studied in Los Angeles.

Following the 1952 disaster in London, England, where a four-day period of smog caused an estimated 4000 deaths (Logan, 1953), it became strikingly clear that local accumulations of air pollution were hazardous. The adage, the solution to pollution is dilution , was fervently followed, leading to the construction of higher smokestacks and the relocation of polluting industries farther away from urban centers. While local air quality benefited from the greater dispersion of pollution afforded by higher smokestacks, problems of regional air pollution, such as acid rain, emerged or were exacerbated.

Acid rain demonstrates the shortcomings of addressing pollution from only a local perspective. The regional problem of acid rain was actually first described in the mid-1800s in reference to the effect of industrial emissions on precipitation in the British midlands. It has long been recognized that acid rain is widespread in northern Europe and eastern North America. By the mid-1980s, roughly half of the sulfuric acid rain falling in eastern Canada originated in the US, just as much of the Scandinavian acid precipitation has been traced to industrialized areas of central Europe and the United Kingdom (Schindler, 1988). Recent studies have led to discoveries of areas affected by acid rain in the western US, Japan, China, the former Soviet Union, and South America.

FOSSIL FUELS: THE COMMON SOURCE OF AIR POLLUTANTS AND GREENHOUSE GASES

Fossil fuel combustion is a common source of emissions of carbon dioxide (CO_2) and the six criteria air pollutants: carbon monoxide, NO_2, particulates, SO_2, ozone, and lead. The use of fossil fuels, particularly gasoline, also releases air toxins, such as benzene, toluene, and VOCs. Fossil fuel combustion produces nearly all of the CO_2 released into the atmosphere from human activities.

Therefore, co-benefits can be achieved by reducing the emissions of fossil fuels that are a common source of criteria air pollutants, air toxins, and greenhouse gases. For

example, if nations abided by the Kyoto Protocol to reduce their emissions of greenhouse gases, what might be the benefits in terms of the effects on human mortality? An interdisciplinary working group examined this question by combining models of energy consumption, carbon emissions, and associated emissions of fine particulate matter (PM). PM has been linked to increased mortality (Dockery and Pope, 1994; Pope *et al.*, 1995). Even at levels below national ambient air quality standards (NAAQS), the presence of PM increases daily rates of cardiorespiratory mortality and total mortality in the US and Europe (Katsouyanni *et al.*, 1993). In children, PM levels also correlate with increased hospital admissions, school absences, and medication use, as well as a reduction in lung function (measured as peak respiratory flow rates) (Bascom *et al.*, 1996).

Under a climate policy scenario approximating that of the Kyoto Protocol (e.g., the US emitting 7% below 1990 levels of greenhouse gas emissions), approximately 700 000 premature deaths due to exposure to PM might be averted annually by the year 2020 as compared with the business-as-usual forecast (Working Group on Public Health and Fossil-Fuel Combustion, 1997) (*see* Business-as-usual Scenarios, Volume 5). This analysis showed the potential near-term co-benefits of long-term policies to mitigate global climate change.

CLIMATE CHANGE AND LEVELS OF AIR POLLUTANTS

Secondary Air Pollutants

Fluctuations in weather have the most influence on pollutants that arise from chemical reactions in the atmosphere. These *secondary pollutants*, such as ozone and acid rain, are derived from a mixture of pollutants directly emitted, which are termed *primary pollutants*. Even with the improving air quality in the US and Canada, urban ozone pollution and regional acid rain persist as problems, and have the potential to be exacerbated if air temperatures increase. Ozone forms secondarily in the lower atmosphere by a temperature- and sunlight-dependent reaction between NO_x and VOCs, precursors of the formation of ozone.

The major source of NO_x in North America is fossil fuel combustion used in transportation and electric utilities. Since 1970, the number of vehicle miles traveled (VMT) in the US has increased at a faster rate than that of the overall population (Figure 1). From 1970 to 1994, car ownership in Canada rose from 310 to 484 vehicles per 1000 people (Last *et al.*, 1998). Transportation is also a major source of anthropogenic VOCs. Other sources of VOCs include incinerators, gasoline vapors, paints and solvents, and some trees, wetlands and other vegetative sources.

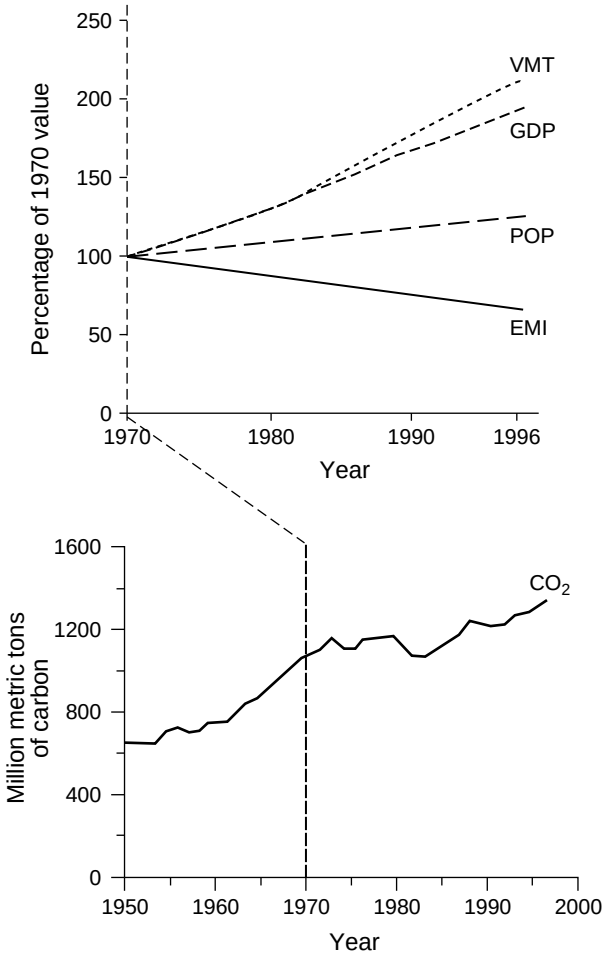


Figure 1 Trends in emissions of criteria air pollutants (EMI), emissions of carbon dioxide, and related factors. VMT • vehicle miles traveled

INFLUENCE OF WEATHER ON OZONE FORMATION

Tropospheric ozone, as a photochemical oxidant and the main ingredient of urban smog, is pervasive and difficult to control. The reaction between NO_x and VOCs to form ozone is catalyzed by ultraviolet (UV) radiation and requires relatively high ambient air temperatures. Meteorological factors that could theoretically influence the levels of tropospheric ozone include UV radiation, air temperature, wind speed, and atmospheric mixing and transport.

In general, there is a direct correlation between temperature and levels of ozone (Kamens *et al.*, 1982; Grey *et al.*, 1987; Samson, 1988). Increases in atmospheric temperature accelerate photochemical reaction rates in the atmosphere and tend to increase the rates at which tropospheric ozone and other oxidants are produced (Morris *et al.*, 1989, 1995). Studies of ambient temperature have reported that successive episodes of high temperatures characterize years

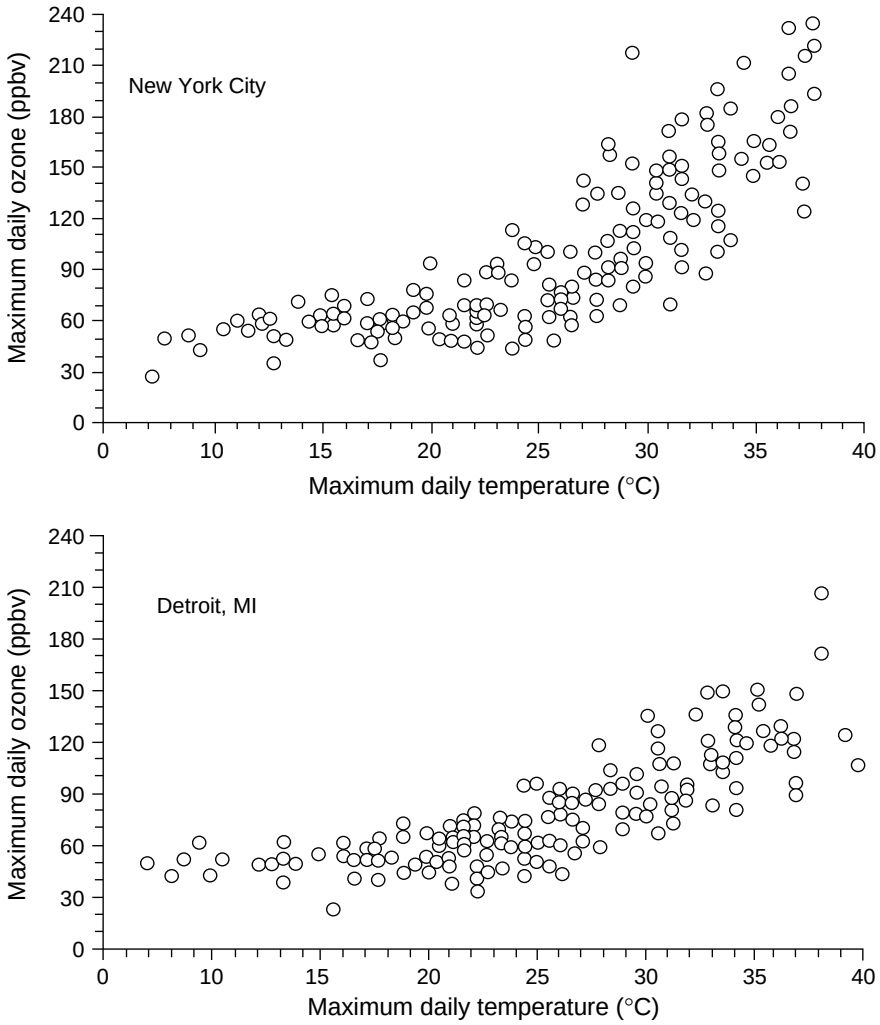


Figure 2 Relationship between maximum daily ozone concentration and maximum daily temperature in two US cities

with seasonally high levels of ozone. The relationship is nonlinear, and above a temperature of 32 °C (90 °F), there is a strong correlation between temperature and levels of ozone (Figure 2). However, levels of ozone may not always increase with an increase in temperature. For example, the production of ozone is reduced when the ratio of VOCs to NO_x is low.

In the eastern US and in Europe, the majority of days when the levels of ozone exceed air quality standards occur in conjunction with slow-moving high-pressure systems around the time of the summer solstice (National Research Council, 1991). This is the period of greatest sunlight when solar radiation is most intense and air temperatures are high. The relatively high levels of ozone in the US during 1988 and 1995 were likely due in part to hot, dry, and stagnant conditions. During 1995, one of the hottest years on record, 32% of Americans, or 71 million people, resided in counties in which ozone levels exceeded the NAAQS (Environmental Protection Agency, 1996). Episodes of high levels of ozone generally last from three to four days on average and extend over a large area, i.e., greater than 600 000 km^2.

Several factors are associated with high-pressure systems that are conducive to the photochemical production of ozone. Large high-pressure systems often create an elevated subsidence inversion of the normal temperature profile, trapping pollutants in the shallow boundary layer at the Earth s surface. Winds associated with major high-pressure systems are generally light, allowing greater accumulation of pollutants. Clear, warm conditions generally associated with large high-pressure systems afford ample sunlight to catalyze the photochemical production of ozone (National Research Council, 1991).

Conditions in Los Angeles illustrate how the combined characteristics of meteorology and topography can adversely affect the air quality. This area is dominated by a persistent high-pressure system off the Pacific Ocean, which forms an elevated inversion layer trapping pollutants over

the city. Topography (river valleys or regions surrounded by mountains) also provides pockets to trap primary pollutants and ozone, often capped by an inversion layer. Due to the region s low latitude and relatively clear skies, plenty of UV radiation is available for photochemical production of tropospheric ozone.

In recent decades, summer climatic conditions in the US may have become more suitable for the formation of ground-level ozone. From 1949 to 1995, the number of summer heat waves (defined as more than three consecutive days with average apparent temperatures exceeding the 85th percentile) increased by 20% (Gaffen and Ross, 1998). Climatic conditions could become more suitable for ozone episodes under projections of global warming, but the extent of ozone formation also depends on air mass and wind conditions.

INFLUENCE OF WEATHER ON OZONE TRANSPORT

Urban smog has long been viewed as a local issue. However, both ozone and its precursors can travel relatively long distances in the atmosphere and can be transported from region to region and across international boundaries. Such processes of transboundary pollution transport have implications for distant downwind populations. Analyses by such groups as the Ozone Transport Assessment Group (OTAG) has confirmed that, in the eastern US, some areas experience a considerable influx of ozone and cannot meet air pollution standards by local measures alone.

The distance that ozone can be transported ranges from 150 500 miles (roughly 240 800 km). For example, ozone precursors from New York City are transported by prevailing winds on a roughly 300-km northeast trajectory through Connecticut and as far as northeastern Massachusetts (Cleveland *et al.*, 1976). Consequently, southwestern Connecticut has the highest concentrations of ozone in this region. The mid-Atlantic and northeastern US are affected by even broader regional ozone transport. While the highest emissions of the ozone precursors, VOCs and NO_x, come from large urban metropolitan areas, the largest elevated point source of NO_x emissions comes from the industrial Ohio River valley (Guinnup and Collom, 1997). Based on spatial patterns and transport considerations, general control measures should clearly target urban nonattainment areas (areas that do not meet NAAQS), as they contribute significantly to their own ozone problems. However, with the Ohio River valley s high density of NO_x point sources, NO_x controls targeted for this region may be an important measure for effectively reducing ozone pollution for the entire eastern US and Canada.

HEALTH EFFECTS OF OZONE

Ozone causes irritation of respiratory and mucous membrane (eyes, nose, and throat). As a strong oxidant, ozone can cause both cellular and structural damage in the lung. Three types of respiratory responses to ozone exposure are: (1) irritative cough; (2) reduced lung capacity, as well as decreased expiratory flow rates; and (3) inflammation of the airway lining or submucosa (Bascom *et al.*, 1996). An inflammatory response usually develops within an hour after exposure and persists for up to 24 hours. Exercise, of course, can increase the dose inhaled due to higher ventilation rates, and the most vulnerable populations include asthmatic children in the northeastern US, where daily ozone levels may remain elevated for six to eight hours (Bascom *et al.*, 1996).

Epidemiological studies of emergency room and hospital admissions indicate ozone s significant health effect. During the summertime (when many viral respiratory infections have subsided), exposure to ozone contributes an estimated 10 20% of all admissions to hospital due to respiratory conditions. On the days with the highest levels of ozone, exposure to ozone can account for nearly half of all respiratory admissions (Environmental Protection Agency, 1996). Even levels below the NAAQS can cause reversible decrements in respiratory function, including airway constriction, reduced lung volume, irritating cough, and chest pain (Bascom *et al.*, 1996).

The effects of chronic exposure to ozone are less well understood. However, an analysis of pulmonary function data from the second National Health and Nutrition Examination Survey (NHANES II) revealed some association between ambient concentrations of ozone and the loss of lung function over time (Schwartz, 1989). Some evidence also suggests that increased incidence and severity of asthma are related to exposure to ozone, although it is difficult to separate the effect of exposure to ozone from concomitant exposure to PM (Environmental Protection Agency, 1996). In Mexico City, ozone has been linked to increased admissions to hospitals for respiratory infections and asthma in children (Romieu *et al.*, 1996). Ozone has been implicated in the exacerbation of asthmatic reactions by reducing the amount of allergen required to induce symptoms (Koren and Bromberg, 1995; Koren and Utell, 1997).

POTENTIAL IMPACTS OF CLIMATE CHANGE ON AIR POLLUTION IN THE US

As population growth and increases in energy consumption lead to increased fossil fuel emissions, the accumulation of CO_2, once thought to be a clean and harmless by-product of combustion, now poses indirect public health and ecological ramifications via global climate change. The change that is

likely to have the most effect on conventional air pollutants is an increase in temperature.

Higher temperatures due to climate change, all other things held constant, may increase concentrations of ozone and acid aerosols, as well as emissions of particulates and allergens (see the section on climate variability, biomass, air quality, and health: forest fires and allergens). A rise in temperature could affect concentrations of ozone by modifying the factors that affect the production of ozone (National Research Council, 1991; Patz *et al.*, 2000), including:

1. Elevated temperature Ozone formation is highly dependent on temperature. Urban areas tend to absorb and hold more heat, a process termed the *heat island effect* (*see* Urban Heat Island, Volume 3). The urban heat island effect can drive up air temperatures by 4 5 °C or more (Landsberg, 1981). The urban heat island may significantly contribute to the chemical formation of ground-level ozone (Quattrochi *et al.*, 2001). Also, NO_x and VOC precursors of ozone are concentrated in congested cities. Urban air quality could therefore be particularly altered by higher global temperatures.
2. Increased frequency and intensity of stagnation periods Climate change could result in more persistent or intense high-pressure systems that provide favorable conditions for the reactions that produce ozone in the atmosphere.
3. Increased water vapor concentration In addition to enhancing the formation of acid aerosols, increases in water vapor can potentiate the formation of ozone (Penner *et al.*, 1989).
4. Increased emissions of precursor pollutants Forests, shrubs, grasslands, wetlands, and other sources of natural hydrocarbons (VOCs) emit greater quantities of these compounds at higher temperatures (*see* Volatile Organic Compounds, Biogenic (VOCs), Volume 2). Soil microbial activity may also increase with warmer temperatures, leading to an increase in NO_x emissions.

On the other hand, factors that could lead to reduced ozone concentrations include the following:

1. Increased cloud cover A more vigorous hydrologic cycle due to global warming could lead to an increase in the frequency of cloudy days. More cloud cover, especially in the morning hours, could diminish reaction rates, thus lowering ozone formation.
2. Increased thickness of the boundary layer of air Higher temperatures might be associated with greater convection, resulting in reduced atmospheric stability and higher wind speeds. If precursor pollutants (VOC or NO_x) are mixed in a greater volume of air, they will be less concentrated and less ozone will be formed (Smith and Tirpak, 1989).

The net effect of these factors is unclear; however, modeling studies combining these variables are being performed to assess the effects of climate change on air quality.

CLIMATE VARIABILITY, BIOMASS, AIR QUALITY, AND HEALTH: FOREST FIRES AND ALLERGENS

Extremes in climate, such as heat waves or severe droughts, are difficult to predict, but are anticipated to become more prevalent under climate change scenarios. Severe droughts with accompanying forest fires have the potential to affect air quality. For example, in 1997 1998, El Nino-driven droughts contributed to the development of expansive forest fires in Indonesia, Mexico, Central and South America, Florida, Canada and several other parts of the world. The extended duration of biomass burning affected air quality over regions far greater than the areas where burning actually occurred. The exposure of humans to fire smoke has been associated with irritation of the throat, lungs, and eyes. In addition, fire smoke carries large amounts of fine particles, which exacerbate respiratory problems, such as asthma and chronic obstructive pulmonary disease (Duclos *et al.*, 1990).

In June and July of 1998, 2277 fires burned approximately 200 000 ha in both rural and urban areas in Florida (Karels, 1998). An analysis of records from several hospitals in Volusia and Flagler Counties showed that visits to the emergency department increased for asthma (by 91%), bronchitis (132%), and chest pain (37%) (Centers for Disease Control and Prevention, 1999).

Another interaction between climate variability and biomass may be found in the release and dispersal of spores and pollen, which can act as allergens. Allergens are not criteria air pollutants, but are included due to their temperature-dependent effect on air quality for allergic people. Pollen counts from birch trees (the main cause of seasonal allergies in northern Europe) appear to increase with increasing temperature (Ahlholm *et al.*, 1998). Pollen counts for Japanese cedar significantly increase in years when summertime temperatures are unusually high (Takahashi *et al.*, 1996; Tamura, 1997). Finally, in the aftermath of flooding, molds and fungi grow on interior surfaces, which can aggravate or trigger airway reactivity among susceptible asthmatics.

See also: Use of Epidemiological Data to Develop Air Quality Standards for Short-term and Long-term Environmental Health Bene ts, Volume 4.

ACKNOWLEDGMENTS

Special thanks to Anne Grambsch, US Environmental Protection Agency (EPA), for her information on climate

change and air pollution modeling studies. This contribution is adapted with permission from Aron and Patz (2001).

REFERENCES

Ahlholm, J U, Helander, M L, and Savolainen, J (1998) Genetic and Environmental Factors Affecting the Allergenicity of Birch (Betula pubescens ssp czerepanovii [Orl] Hamet-Ahti) Pollen, *Clin. Exp. Allergy*, 28, 1384 1388.

Aron, J L and Patz, J A, eds (2001) *Ecosystem Change and Public Health: A Global Perspective*, John Hopkins University Press, Baltimore, MD, 526.

Bascom, R, Bromberg, P A, Costa, D A, Devlin, R, Dockery, D W, Frampton, M W, Lambert, W, Samet, J M, Speizer, F E, and Utell, M (1996) State of the Art: Health Effects of Outdoor Air Pollution, *Am. J. Respir. Crit. Care Med.*, 153, 3 50.

Centers for Disease Control and Prevention (1999) Surveillance of Morbidity During Wildfires Central Florida, 1998, *MMWR (Morbidity and Mortality Weekly Report)*, February 5.

Cleveland, W, Kleiner, B, McRae, J, and Warner, J (1976) Photochemical Air Pollution: Transport From the New York City Area into Connecticut and Massachusetts, *Science*, 191, 179 181.

Dockery, D and Pope, C (1994) Acute Respiratory Effects of Particulate Air Pollution, *Ann. Rev. Public. Health*, 15, 107 132.

Duclos, P, Sanderson, L M, and Lipsett, M (1990) The 1987 Forest Fire Disaster in California: Assessment of Emergency Room Visits, *Arch. Environ. Health*, 45(1), 53 58.

Environmental Protection Agency (1996) *National Air Quality and Emissions Trends Report, 1995*, Environmental Protection Agency, Office of Air Quality Planning and Standards, Washington, DC.

Gaffen, D J and Ross, R J (1998) Increased Summertime Heat Stress in the US, *Nature*, 396(10), 529 530.

Grey, M, Edmond, R, and Whitten, G (1987) *Tropospheric Ultraviolet Radiation: Assessment of Existing Data and Effect on Ozone Formation*, Environmental Protection Agency, Research Triangle Park, NC.

Guinnup, D and Collom, B (1997) *Telling the OTAG Ozone Story with Data*, Final Report, Vol. I, Executive Summary, OTAG Air Quality Analysis Workgroup, June 2, 1997.

Kamens, R, Jeffries, H, Sexton, K, and Gerhardt, A (1982) *Smog Chamber Experiments to Test Oxidant-related Control Strategy Issues*, Environmental Protection Agency, Research Triangle Park, NC.

Karels, J (1998) Wildland Fire Season in Review, *Florida Fire Service Today*, 6, 8 19.

Katsouyanni, K, Pantazopoulou, A, Touloumi, G, Tselepidaki, I, Moustris, K, Asimakopoulos, D, Poulopoulou, G, and Trichopoulos, D (1993) Evidence for Interaction Between Air Pollution and High Temperature in the Causation of Excess Mortality, *Arch. Environ. Health*, 48(4), 235 242.

Koren, H S and Bromberg, P A (1995) Respiratory Responses of Asthmatics to Ozone, *Int. Arch. Allergy Immunol.*, 107(1 3), 236 238.

Koren, H S and Utell, M J (1997) Asthma and the Environment, *Environ. Health Perspect.*, 105(5), 534 537.

Landsberg, H E (1981) *The Urban Cimate*, Academic Press, New York.

Last, J, Trouton, K, and Pengelly, D (1998) *Taking Our Breath Away*, David Suzuki Foundation, Vancouver.

Logan, W P D (1953) Mortality in the London Fog Incident, 1952, *Lancet*, 336 338.

Morris, R, Guthrie, P, and Knopes, C (1995) *Photochemical Modeling Analysis Under Global Warming Conditions*, AWMA Annual Meeting, AWMA, San Antonio, TX.

Morris, R E, Gery, M W, Liu, M K, Moore, G E, Daly, C, and Greenfield, S M (1989) *Sensitivity of a Regional Oxidant Model to Variations in Climate Parameters*, Environmental Protection Agency, Washington, DC.

National Research Council (1991) *Rethinking the Ozone Problem in Urban and Regional Air Pollution*, National Academy Press, Washington, DC.

Patz, J A, McGeehin, M A, Bernard, S M, Ebi, K L, Epstein, P R, Grambsch, A, Gubler, D J, Reiter, P, Romieu, I, Rose, J B, Samet, J M, and Trtanj, J (2000) The Potential Health Impacts of Climate Variability and Change for the US: Executive Summary of the Report of the Health Sector of the US National Assessment, *Environ. Health Perspect.*, 108, 367 376.

Penner, J E, Connell, P S, Wuebbles, D J, and Covey, C C (1989) *Climate Change and its Interactions With Air Chemistry: Perspectives and Research Needs*, Environmental Protection Agency, Washington, DC.

Pope, C, Bates, D, and Razienne, M (1995) Health Effects of Particulate Air Pollution: Time for Reassessment? *Environ. Health Perspect.*, 103, 472 480.

Quattrochi, D A, Luvall, J C, Rickman, D L, Estes, M G, Laymon, C A, and Howell, B F (2001) A Decision Support Information System for Urban Landscape Management Using Thermal Infrared Data, *Photogramm. Eng. Remote Sens.*, 66(10), 1195 1207.

Romieu, I, Meneses, F, Ruiz, S, Sienra, J J, Huerta, J, White, M C, and Etzel, R A (1996) Effects of Air Pollution on the Respiratory Health of Asthmatic Children Living in Mexico City, *Am. J. Respir. Crit. Care Med.*, 154(2 Pt 1), 300 307.

Samson, P (1988) *Linkages Between Global Climate Warming and Ambient Air Quality*, Global Climate Linkages Conference, Washington, DC.

Schindler, D W (1988) Effects of Acid Rain on Freshwater Ecosystem, *Science*, 239, 149 157.

Schwartz, J (1989) Lung Function and Chronic Exposure to Air Pollution: a Cross-sectional Analysis of NHANES II, *Environ. Res.*, 50, 309 321.

Smith, J B and Tirpak, D A, eds (1989) *The Potential Effects of Global Climate Change on the United States*, Report to Congress, US Environmental Protection Agency, Washington, DC, EPA-230-05-89-050.

Takahashi, Y, Kawashima, S, and Aikawa, S (1996) Effects of Global Climate Change on Japanese Cedar Pollen Concentration in Air Estimated Results Obtained from Yamagata City and its Surrounding Area, *Arerugi*, 45(12), 1270 1276.

Tamura, Y, Kobayashi, Y, Watanabe, S, and Endou, K (1997) Relationship of Pollen Counts of Japanese Cedar to Weather Factors in Isehara City, Kanagawa, *Nippon Jibiinkoka Gakkai Kaiho*, 100(3), 326 331.

United Nations Environment Programme (1991) *Urban Air Pollution*, United Nations Environment Programme, Nairobi, Kenya.
World Bank (1997) *Clear Water, Blue Skies*, World Bank, Washington, DC.
Working Group on Public Health and Fossil-Fuel Combustion (1997) Short-term Improvements in Public Health From Global-Climate Policies on Fossil-fuel Combustion: an Interim Report, *Lancet*, 350(9088), 1341 1349.

FURTHER READING

http://capita.wustl.edu/otag/reports/aqafinvol. I/animations/v1. exsumanimb.html (Date Last Revised 6/2/1997).
Patz, J A, Engelberg, D, and Last, J (2000) The effects of changing weather on public health, *Annu. Rev. Public Health*, 21, 271 307.

Urban Ecosystems

Ian Douglas
University of Manchester, Manchester, UK

Cities are often seen as places where people's lives are most removed from nature, but in many ways, as the consequences of severe geophysical events such as earthquakes and oods demonstrate, they are places where the vulnerability of people to the impacts of extreme climatic and tectonic events is heightened (Douglas, 1999). People and all other living beings in cities indeed are as dependent on nature and the global environment as in any other habitat. Urban settlements are globally important in the environmental context because they house over half the world's people, because they concentrate the burning of fossil fuels and the release of greenhouse gases, and because they are dependent on food and other resources from land and aquatic ecosystems throughout the world. Urban settlements occupy between 1 and 2% of the total land area of the world, and only 11% of the land in crowded England. However, the world urban area is growing rapidly and is often made up of good quality agricultural land that previously supported intensive food production.

There are the same basic kinds of interactions in cities as in other ecosystems. However, those interactions are greatly affected not only by the structures that people have built and the energy they import, but also by people's cultures, behaviors, social organization and economy. In addition, cities cannot provide all the energy, water, food and other materials their people require from the land occupied by the city alone. Cities have to import such things from outside their territory and necessarily export manufactured, used and waste items and substances back out to the rural environment. Thus great cities are biologically parasites in their use of the vital resources, air, water and food. The bigger the cities, the more the demands on the surrounding countryside and the greater the danger of damaging the natural environment host (Douglas, 1983). Thus the importance of urban ecosystems and other human settlements in global environmental change is related less to the area covered with buildings and other structures than to the role of settlement-related activities in promoting land-cover change generally, through land-use change and land-cover conversion to meet the requirements of urban metabolism (Douglas, 1994). Being centers of intense human activity and industry, cities are also nodes of concentrated atmospheric emissions. Per capita emissions of carbon dioxide and many other greenhouse gases tend to increase with urban development, as recent trends in China and India suggest. Thus the functioning of urban ecosystems is a key element in global environmental change.

INTRODUCTION

In many ways, urban areas are just a complicated kind of ecosystem, requiring new interactions among researchers to understand cities as ecosystems. Ecologists, social scientists, economists, and engineers are all involved, along with city, county, state and national agencies. Inevitably, interactions with individual citizens, families and community groups form a major part of the success of any applied urban ecosystem study. All attempts to improve living conditions and the environment of deprived urban areas indicate that no lasting progress is possible without community commitment and involvement.

Three broad views of urban ecology are current. One view sees urban ecology as the study of natural assemblages of plants and animals in the urban environment. Another approach views the city as a whole and examines the broad relationship between the human population, ecologically significant consumption, and the sustainability of essential energy and materials flows (Rees, 1992) (*see* Anthropogenic Metabolism and Environmental Legacies, Volume 3; Ecological Footprint, Volume 3). The third perspective sees urban areas as a series of varied habitats for human beings and other organisms. In urban social geography and sociology this standpoint is sometimes termed *human ecology* (Knox and Pinch, 2000). Cities are sometimes seen as the places where people are the most removed from nature, but in reality they represent a set of habitats to which organisms adjust. Every part of a city has some form of life, and some segments of cities teem with wildlife.

URBAN ECOSYSTEMS HABITATS FOR PLANTS AND WILDLIFE

In terms of the analysis of natural assemblages of plants and animals, urban ecology investigates the biosphere in towns and cities using ecological methods. Although urban growth appears to destroy wildlife habitats and reduce the diversity of flora and fauna (Bridgman *et al.*, 1995), open spaces and food sources in cities provide a great variety of ecological situations that are exploited by particular types of plants and

Table 1 Types of urban land cover and biogeographic conditions

Nature of land cover	Type of location	Nature of biotic zone	Type of urban ecosystem	Trend over time
Paved, roofed, densely urban complexes devoid of open areas of vegetation and water bodies	City central business districts. Large shopping centers and car parks. Some industrial zones	Cliff/organic detritus (Walls of tall buildings are cliff-like. Organic waste may litter streets and other spaces between buildings)	Anthropogenic	Birds and insects develop an increasing number of niches. Micro-organisms colonize
Suburban mosaic of houses, roads, gardens and mature trees	Older, spacious inner suburbs	Old urban savanna	Urbanophile	Already diverse and mature, future changes depend on management
Corridor zones of wild plants	Railway, canal, power line and some arterial road reservations	Grassland/weed complex	Peripheral	Increasing biotic diversity unless interference occurs
Landscaped parks and open spaces	City parks, recreation grounds, golf courses	Mowed grassland	Anthropogenic and modified natural	Careful management keeps biotic diversity low
Derelict land construction sites	Abandoned industrial sites, old waste tips	Abiotic/weed complex	Urbanophile	Temporary opportunities for plant colonization removed when development is completed
New suburbs, devoid of mature trees and high grassland: cultivated/garden ratio	Outer suburban areas, new housing estates, some modern industrial estates and office building complexes	New urban savanna	Urbanophile	Biotic diversity increases as gardens become established. Rate of increase depends on occupants' attitudes
Grassland on reclaimed soil with streets, car parks and buildings but few or no mature trees	Inner city redevelopment areas with large open spaces, modern flats or row houses, some new industrial estates	New urban savanna/mowed grassland	Anthropogenic/ urbanophile	Management tends to preserve low biotic diversity
Small woodland and rural areas within the city	Patches of woodland left as quasi-natural areas, urban commons or heaths	Remnant ecosystem/natural islands	Relict	Old woodland diversity maintained if human interference is not increased
Water bodies	Lakes, rivers and reservoirs	Lake/stream aquatic complexes	Modified natural	Status dependent on management
Wetlands and much modified water bodies	Marshes, sewage farms and gravel pits	Derelict/weedy grassland and aquatic complex	Urbanophile	If neglected, biotic diversity increases. Likely to be much disturbed

Source: Douglas, 1983.

animals. Within each of these urban habitats, equilibrium develops between the colonizing plants and their physical and biological environment.

The nature of the land cover affects the climate, nutrient status and diversity of intra-urban ecosystems, as city habitats and biotic communities may be called. At one extreme is the completely covered, built up central business district, at the other, the relics of the rural landscape in the form of remnant woodlands, bush land, copses, commons and urban heaths (Table 1). Between these two extremes is a gradation from near natural ecosystems to completely artificial ones (Table 2). Only a small part of any urban area is totally paved and roofed; most of it is a mosaic of such impermeable surfaces and vegetated areas. For example, roads and parking lots cover 10% of Los Angeles. The vegetated areas range from carefully tended gardens to patches of derelict industrial land (*brown elds*) and areas of mature trees and shrubs. This floristic diversity is related in part to the nature of the urban soil, much of which is fill or rubble, often highly compacted, on which many native plants have difficulty in becoming established. The plant diversity also reflects the deliberate or inadvertent introduction of exotic species and modification of the plant nutrient status of soils by fertilizers, compost, dumping of wastes. The nature of intra-urban ecosystems is thus partly a function of land management, varying in time and space according to changes in land ownership and land use. The changing, impermanent, patchwork nature of the factors affecting these plant and animal habitats means that no accurate, comprehensive classification of intra-urban ecosystems can be made. Table 1 merely indicates some broad types of urban biotic zones particularly relevant to European, North American and Australian cities.

Trees form a significant part of any urban landscape. Their major benefits vary with situation and climate. In hot climates they provide welcome shade, in windy areas, shelter and in polluted places they help to remove some of the atmospheric contaminants. In urban settings, deciduous trees can reduce dustfall by 27% and intercept 9% of suspended particles in the air. Coniferous trees have even greater capture rates, reducing dustfall by 38% and intercepting 13% of suspended particles (Wheater, 1999). Trees remove other pollutants, such as nitrogen oxides, sulfur dioxide, carbon monoxide, ozone and halogens, although some, especially sulfur dioxide, fluoride and ozone, damage receptor trees. Possibly 5% of total air pollution is eliminated by urban trees in suburban Chicago. Trees also reduce noise, double rows of evergreen trees being highly effective.

Trees use large quantities of water. While this may help to alleviate some urban drainage problems, large trees of certain species can so lower local water tables that settlement of building foundations occurs or roots penetrate drain pipes. Nevertheless, urban forestry is now a major part of city management and for many organizations urban ecology is primarily concerned with trees in cities.

In the US, 80% of the population lives in metropolitan areas and on average, tree canopies from about 75 billion trees cover 33% of this land area. The impact of these urban forests on runoff is of growing interest as efforts to protect water quality focus on non-point sources within urban watersheds. Tree planting is one of several *best management practices* examined in an experimental residential

Table 2 Schemes for characterizing urban ecosystems[a]

Areas used by Simms (1975) to discuss urban birds	Types of urban ecosystem	Urban biocenoses as described by Duvigneaud (1974)	City land use types described by urban geographers
Central business district	Cliff/organic detritus Old urban savanna	Anthropogenic biocenoses	Central business district
Nearer suburbs	New urban savanna	Urbanophile biocenoses	Inner suburbs
Outer ring	Landscaped parkland	Anthropogenic and modified natural biocenoses	Outer suburbs
(1) Estates and factories			Industrial estates
(2) The open spaces			
Edge of the countryside	Remnant ecosystem Natural	Relict biocenose Peripheral biocenoses	Rural–urban fringe
Rivers, lakes and reservoirs	Lake/stream aquatic complexes	Modified natural biocenoses	Water based recreation areas
Marshes, sewage farms and gravel pits	Derelict/weedy grassland and aquatic complexes	Urbanophile biocenoses	Industrial wasteland and wetlands

[a] Duvigneaud uses biocenosis (plural: biocenoses) as a term for a plant and animal habitat that is not necessarily exactly parallel to what most ecologists would call an ecosystem. The emphasis is more on the life support conditions than on specific assemblages of life forms.

retrofit in Los Angeles. The area has been converted into a mini-watershed that retains runoff on site and stores roof runoff for landscape irrigation. Such regulation of urban drainage at source, around the roofs and gardens, on which the rain falls, has major implications for groundwater levels and river flows in and out of urban areas. It would exert some influence on the global flux of water to the oceans if applied on a citywide scale over most major urban areas.

THE METABOLISM OF URBAN ECOSYSTEMS

Cities exist by bringing things together from different places and transforming them into other things. This supply and conversion of food and raw materials involves complicated chains and both good and bad side effects. These materials flows are assisted by monetary changes and by flows of energy and water.

Materials flows through urban areas can be envisaged as part of the global biogeochemical cycling of materials between the lithosphere, hydrosphere, atmosphere and the biosphere. People exploit biogeochemical cycles when they exploit substance stocks and create stocks in the economy, through the materials flows involved in industrial, agricultural and urban activities. This biogeochemical cycling can be conceptually extended to incorporate an urban, or industrial system, in which the urban fabric is added as an additional repository of substances, alongside the lithosphere, biosphere, hydrosphere and atmosphere (Odum, 1971; Douglas, 1983). Indeed, the urban fabric is constantly being renewed, demolition creating a recycling opportunity for crushed concrete to be used as aggregate. The urban fabric can thus be interpreted as the urban deposit (Wilburn and Goonan, 1998), a form of renewable resource, constantly being restocked with new structures. In this way, people s activities are seen as part of the processes changing the earth s surface at geological scales.

The acquisition, turnover, stocking and ejection of materials by human settlements was termed *urban metabolism* by Wolman (1965) (*see* Anthropogenic Metabolism and Environmental Legacies, Volume 3). His classic study concentrated on local problems of air pollution, waste and water supply, but since 1980 the metabolism approach has been extended to a broader concern about the impacts of the materials demands of human settlements on other ecosystems at the global scale (White, 1994). At least since the early 16th century, the European expansion of trade to a global enterprise has witnessed the richer people in the richer countries appropriating global ecosystems to meet the needs of their increasingly urban citizens (Jardine, 1996). Gradually the great cities have become more and more dependent on a wider and wider range of sources of food, raw materials and manufactured goods. No longer are vegetables and fruit seasonal in appearance, the places from which they come simply vary through the year. Thus it can be argued that urban ecosystems need far more land for their support than that which they actually occupy. To that extra area of land must be added land used for waste disposal, water supply and recreation. This land requirement, calculated on a per capita basis for any city, is termed the urban ecological footprint (Rees, 1992 and Ecological Footprint, Volume 3).

Nevertheless, many studies continue to examine urban energy, water and materials fluxes at the whole-city and intra-urban ecosystem scales. The Human Settlements as Ecosystems: Metropolitan Baltimore, Maryland, from 1797 2100 study (Pickett *et al.*, 1999), for example, focuses on surface/atmosphere energy exchange, hydrologic and nutrient flux, atmospheric deposition, the import/export of raw and processed materials and waste products, and fluxes of human populations and capital. It is based on the input/output mass balance concept that has been applied widely in ecosystem ecology. It considers the whole-city scale (sub-regional), and the small watershed scale in order to foster comparison with past, present and future studies of natural and human dominated ecosystems. The urban mass balance equation inevitably includes many fluxes normally not found in ecosystem studies (e.g., inputs of imported food, sewage outputs, fluxes of commodities and money).

URBAN ECOSYSTEMS AS HABITATS FOR HUMAN BEINGS

The city may be examined as a human habitat just as it is a bird or mammal habitat. Human beings have spent virtually all their evolutionary history in rural environments engaged in manual labor. Now more and more of them are engaged in sedentary occupations in cities. While there appears to be little evidence that mankind is ill suited to the life style of cities, there is evidence that some urban environments are quite unsuitable habitats. In the mid-20th century, French country children tended to be shorter, have slower growth rates and reach puberty later than city children. However, although such contrasts could be interpreted in terms of better living conditions and improved diet in cities, the physical superiority of city dwellers mainly arises from the inter-marriage of migrants coming from different parts of France to the cities. Parisian children whose parents were born in Paris were shorter than the children of in-migrant parents.

Human physique, the development of brain cells and mental processes are influenced by both socioeconomic status and environmental conditions. Children in poor housing in Western countries as well as those of squatter settlements in Third World cities may have their physical and intellectual development handicapped by their habitat even though the habitat itself is a result of the social economy of the city. The human environment is sometimes the result of

choice not solely of socioeconomic forces. The clustering of the Pakistani population in cheap but poor housing in certain British cities in the 1970s, for example, was seen as a desire to have surplus funds to repatriate to Pakistan and to gain the advantages of proximity to ethnic organizations and shops.

Little escape from poverty can exist in social systems that confine part of the population to poor housing that itself directly affects human wellbeing through health risks. Lack of basic sanitation in the housing of the bulk of the world s poor creates public health hazards through contaminated water supplies or other sources of pathogen transmission resulting from inadequately treated sewage. Depending on soil and water table conditions, human wastes in some cities may be handled by dug latrines (as in Addis Ababa) or by open ditches (as in much of Jakarta). Such conditions, amplified by crowding, increase the likelihood of many infectious diseases, of which typhoid fever and cholera are only the two best known. In spite of advances in epidemiological techniques, the interaction of these problems of health and housing is so complex that the evidence from empirical studies is by no means conclusive.

Many studies in Western countries indicate that relationships between urban physical environmental quality and many social problems are generally weak. The most important associations between social problems, disease and socioeconomic conditions in the US are links between poverty, sudden death, broken homes, lack of schooling, juvenile delinquency and drug offences. Nevertheless, before these conclusions are extended to other areas, it has to be remembered that living conditions in many deprived areas of American cities are far better than those in the poorest parts of cities in Latin America, Africa and most of Asia. When examining cities as ecosystems, the analyst has to follow a narrow path between environmental determinism and completely social explanations (Douglas, 1983).

Perhaps it is in the field of public health, whose significance for improvements in urban living conditions goes back to at least the early 19th century, that the view of the city as an ecosystem is most practically relevant. A specialist in preventive medicine may identify a death due to chronic bronchitis in a badly heated, poorly ventilated, one room aged person s dwelling as the result of excessive cold. A social worker on the other hand may see the excessive cold as the product of social inequality and a political system unable to cope with the need to redistribute wealth and to ensure adequate health care provision and geriatric services. In this example, the death could have been prevented by a better welfare system, but in the absence of the appropriate support, the victim was vulnerable to extreme cold weather. Analysis of human ecology in urban areas thus requires identification of the interplay of biophysical and social variables and definition of the levels of ecological analysis being pursued in any given enquiry.

One specific usage of the terms urban ecology and ecological analysis has been in the differentiation of residential areas within cities (Knox and Pinch, 2000). Sometimes known as *social area analysis*, this type of analysis generated a debate on people poverty versus place poverty in which attention is drawn to the so-called ecological fallacy of assuming that individuals in a particular area share the overall attributes of that area. The temptation in this analysis is to use readily accessible census data to characterize wards, or even census enumeration or collectors districts within wards, on the basis of such deprivation indicators as lack of a hot water supply, no bath or shower, unemployment, and low car ownership. While clear parallels emerge between patterns of ill health and living conditions in many such analyses, there are substantial differences in the kinds of poor people who live inside and outside areas of major concentrations of poverty. The complications of these delicate analyses and their interpretation should not be allowed to discourage their pursuit. If urban areas are to be made more habitable, and thus if the ideals of the United Nations Habitat Agenda are to be achieved, integrated biophysical and social analyses of urban ecosystems will be required.

CONCLUSION

Cities of all sizes, collectively, occupy pivotal positions in the patterns of global ecology. Cities consciously exclude natural processes, substituting mechanical devices made possible by intensive use of fossil fuels. Rather than using the solar energy continuously falling on their streets and buildings, they dissipate it as excess heat. At the same time, cities import immense quantities of concentrated energy in various forms, most of it derived from petroleum coaxed from the ground in distant landscapes. Urban centers evacuate the water falling upon their roofs and streets as runoff through concrete pipes and channels into the nearest bay or river and, at the same time, bring water in from distant landscapes through similar concrete channels. From other ecosystems throughout the world, cities import nutrients in the form of food, use it once, then send the residues out through pipes as sewage waste and packaging and solid remains are trucked to a convenient landfill or incinerator. Many cities however, have long had traditional rubbish collections for recycling, such as the rag and bone men that once toured London streets, or the rag pickers found in cities like Nairobi and Cairo today, with their modern equivalents of kerb side bottle, paper and can collections or recycling centers in supermarket car parks. Such resource re-use could be greatly increased to slow down the expanding impacts of cities on other ecosystems.

While the ecological footprints and global impacts of this urban metabolism are key elements in global environmental change, for the majority of the world s people who now live in urban areas, the day-to-day characteristics of intra-urban

ecosystems are perhaps more immediately important. Their housing and health conditions are intimately related to local vulnerability to geophysical and technological hazards and to the consequences of poor diet, smoking, lifestyle, educational opportunities and employment. With social differentiation becoming more marked, from China to America, the gulf between the urban dwellers who drive global change through their purchases and their waste generation and the city people who suffer the consequences of increased vulnerability and poor living conditions will inevitably widen. The analysis of global and within-city implications of the functioning of urban ecosystems thus makes a major contribution to the understanding of the factors driving global change and their local consequences.

See also: Geomorphological Change for Urbanization and Industry, Volume 3; Sustainable Cities Policies, Volume 4; Love Canal, Volume 5.

REFERENCES

Bridgman, H, Warner, R F, and Dodson, J (1995) *Urban Biophysical Environments*, Oxford University Press, Melbourne.

Douglas, I (1983) *The Urban Environment*, Arnold, London.

Douglas, I (1994) Human Settlements, in *Global Land Use/Cover Change*, eds W B Meyer and B L Turner, Cambridge University Press, Cambridge, 149 169.

Douglas, I (1999) Physical Problems of the Urban Environment, in *Applied Geography: Principles and Practice*, ed M Pacione, Routledge, London, 124 134.

Duvigneaud, P (1974) L ecosystème Urbs , *Mémoires de la Société de Botanique de Belgique*, 4, 5 35.

Jardine, L (1996) *Worldly Goods: a New History of the Renaissance*, Macmillan, London.

Knox, P and Pinch, S (2000) *Urban Social Geography*, 4th edition, Prentice Hall, Harlow.

Odum, H T (1971) *Environment Power and Society*, Wiley Interscience, New York.

Pickett, S T A, Burch, Jr, W R, and Grove, J M (1999) Interdisciplinary Research: Maintaining the Constructive Impulse in a culture of Criticism, *Ecosystems*, 2, 302 307.

Rees, W E (1992) Ecological Footprints and Appropriated Carrying Capacity: What Urban Economic Leaves Out, *Environ. Urban.*, 4, 121 130.

Roots, B L, Chant D A, and Heidenreich, C E (1999) *Special Places: The Changing Ecosystems of the Toronto Region*, UBC Press, Vancouver, 1 342.

Simms, E (1975) *Birds of Town and Suburb*, Collins, London.

Wheater, C P (1999) *Urban Habitats*, Routledge, London.

White, R R (1994) *Urban Environmental Management - Environmental Change and Urban Design*, John Wiley & Sons, Chichester.

Wilburn, D R and Goonan, T G (1998) Aggregates from Natural and Recycled Sources: Economic Assessments for Construction Applications a Materials Flow Analysis, *US Geological Survey Circular*, 1176.

Wolman, A (1965) The Metabolism of Cities, *Sci. Am.*, 213(3), 179 190.

Urban Heat Island

James A Voogt

University of Western Ontario, London, Canada

The urban heat island (UHI) is the characteristic warmth of urban areas compared to their (non-urbanized) surroundings. Generally it refers to the increase in air temperatures, but it can also refer to the relative warmth of surface or subsurface materials. The UHI is an inadvertent climate change that arises because of atmospheric and surface modi cations in urban areas. Changes to the surface radiation and energy balances lead primarily to reduced cooling rates in urban areas, compared to those of its surroundings. The UHI has important implications for human comfort and health, urban air pollution, energy management, and urban planning. UHIs of cities in hot climates are generally non-bene cial as they lead to larger cooling loads and energy use, increased human discomfort and higher air pollution concentrations. The high rate of urbanization of the world's population, particularly in developing countries, means that increasing numbers of people will be affected by heat islands in these cities. In temperate and cold climates heat islands may provide some bene ts, especially in the cold season because of reduced heating loads, thereby reducing energy use. In these cases impacts of the UHI may be mostly bene cial in the winter, and undesirable in the summer.

UHIs are not responsible for global warming because cities occupy only a tiny fraction of the Earth's surface. Urban climates are a local climate effect, but they do complicate our efforts to detect global climate change, because many climate recording stations have been located in or near cities, and their records may be in uenced by the UHI. Global climate change will add an additional thermal burden to urban areas, accentuating UHI impacts. Whether the relative warmth of a heat island will be enhanced under global climate change is less clear, because of unknowns regarding the speci cs of frequencies of future weather conditions, which exert a strong control on heat island magnitude. Efforts to mitigate UHIs in warm climates have shown promise. These provide not only local environmental bene ts but also result in feedbacks that reduce emissions of greenhouse gases and therefore contribute to mitigation efforts of global climate change. The UHI may serve as a current analog for future global environmental change.

TYPES OF URBAN HEAT ISLANDS

The UHI most often refers to the increase of air temperature, in the near-surface layer of the atmosphere within cities relative to their surrounding countryside. Observations

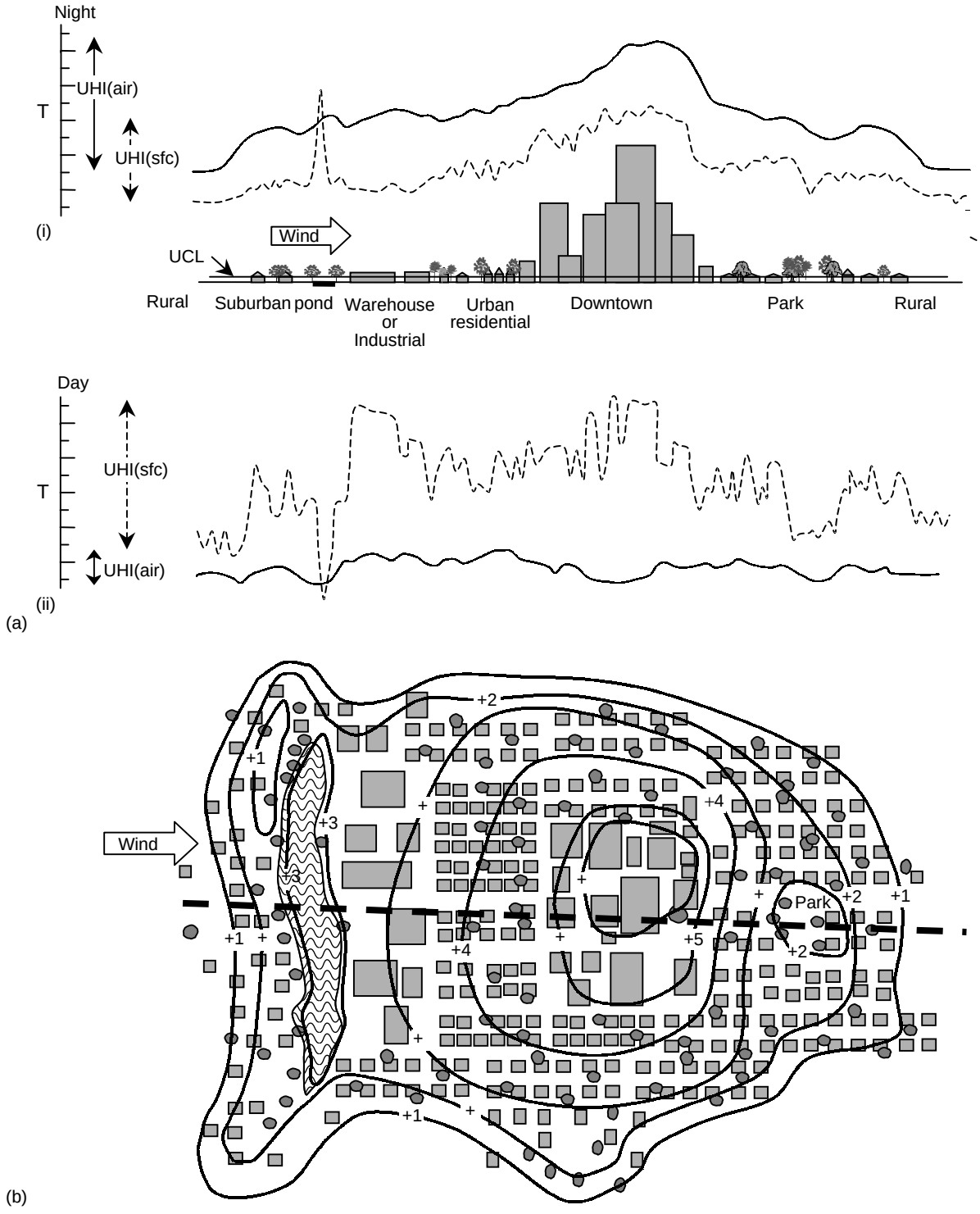


Figure 1 UHI characteristics. (a) Cross-sections of air temperatures measured within the UCL (urban canopy layer) and surface temperatures (e.g., as observed by a remote sensor) under optimum heat island conditions during (i) nighttime and (ii) daytime. (b) Plan view of spatial patterns of air temperature, which make up the nighttime UHI

of most UHIs are derived from air temperature measurements made within the layer of the urban atmosphere below the tops of buildings and trees, known as the urban canopy layer (UCL; see Figure 1a). These observations may be made using networks of fixed stations or from mobile traverses, for example of vehicle-mounted thermometers.

More recently, remote sensors operating in the thermal infrared wavelength region and mounted on aircraft or satellites have been used to observe the surface UHI, with high spatial resolution. These sensors detect radiation emitted and reflected by the surface, rather than temperature directly, so their output is often termed the apparent surface temperature. The temperature may be substantially different from the true surface temperature because of the preferential view of horizontal surfaces by the sensor, reduced transmission of radiation to the sensor by the atmosphere, and the reflectivity of the surface in the waveband of the sensor. The surface temperature is very sensitive to changes in surface conditions, and therefore shows much greater spatial variability and temporal variation between day and night, than does air temperature (Figure 1). Thus, although surface and air temperature heat islands are related, they are not the same, and care should be taken to distinguish between them.

URBAN HEAT ISLAND CHARACTERISTICS

Spatial Characteristics

The UHI is so named because the pattern of isotherms (contours of equal temperature) forms an island shaped pattern. The actual pattern for a given city depends on the configuration of the urbanized area, but typically a large gradient of air temperature forms near the urban rural boundary followed by a more gradual rise to the city core. Within the built-up area, the pattern is influenced locally by surface features such as parks, water bodies and more or less densely built-up areas of the city (Figure 1b). The spatial pattern of isotherms loosely follows the urbanized area, but may show evidence of being transported (advected) downwind. The topographic setting of the city (e.g., a coastline or valley location) may add further complexity to the spatial characteristics of the UHI, as topographic effects (e.g., sea breezes) interact with urban effects.

An overall measure of the magnitude of the heat island is the heat island intensity; the difference between the background rural and highest sustained urban air temperature. This definition avoids uncharacteristically large (or small) heat islands that could result from localized anomalously cool (warm) rural and warm (cool) urban measurements.

HEAT ISLAND FORMATION

Temporal Development

UHIs are primarily nocturnal features that arise as a result of differences in the cooling rates between urban and rural areas. Differences in the cooling rates are greatest when skies are clear and winds are light. Heat island intensities generally increase with time from sunset although exact timing of the peak value may vary depending on weather conditions and season (Figure 2). In some cases, negative heat islands or cool islands, may characterize some inner-city urban areas in the middle of the day due to the slow uptake of heat by the city and large shaded areas from tall buildings.

The temperature differences created by a strong UHI can influence local circulation patterns when regional-scale winds are light. A heat island circulation, similar to a sea breeze, may be initiated, with a low-level inflow towards the city center from outlying rural regions, rising air over the city, and a return flow aloft towards the rural areas. Advective (horizontal) transport of heat, humidity and pollutants by these circulations can have important implications on the dispersion of pollutants in and around urban areas and on the development of the UHI itself.

Formative Factors

The differences in cooling rates that underlie the formation of the UHI, are influenced by surface characteristics and atmospheric conditions. Direct addition of heat to the atmosphere through human activities, known as anthropogenic heat, can also play an important role in heat island formation. Typically, anthropogenic heating tends to be greatest

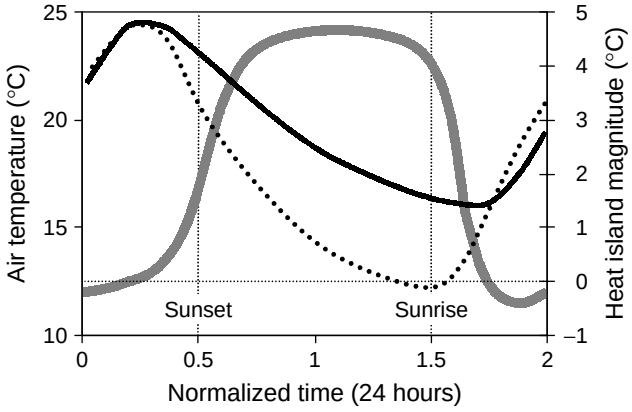


Figure 2 Typical diurnal development of urban and rural air temperatures (thick lines), and heat island magnitude (thin line) under clear and calm conditions. Results will vary depending on the climate region and season. (Adapted from Oke, 1982)

in the winter season in cold climates because of the need for space heating, but the magnitude of anthropogenic heating is affected by intensity of energy use, power generation and type of transportation system in the urban area (Taha, 1997).

Surface Controls

The primary surface controls for most UHIs are surface geometry and surface thermal properties (Oke *et al.*, 1991). Surface geometry refers to the obstruction of the sky by buildings and other objects on the urban surface expressed through a term known as the *sky view factor* (SVF). The reduced SVF of many urban surfaces, particularly those on the ground between buildings, restricts their ability to lose heat by radiative transfer, because the cold radiative sky is replaced by relatively warm surfaces of buildings, trees etc. Surface geometry may also provide a sheltering effect that reduces convective heat losses. Decreases in SVF almost always lead to increased UHI intensity: although if buildings are very closely spaced, shading may reduce absorption of solar radiation during the day sufficiently to produce a cool island.

Surface thermal properties, especially thermal admittance (μ) control the uptake and release of heat by the surface and subsurface. Dense or wet materials such as metal, stone, saturated clay soils, are easily able to absorb and release heat (high μ). Dry and less dense materials with good insulating properties such as wood, sand or aerated concrete have low μ values. Conditions favorable for large UHIs exist where and when the urban area is characterized by relatively dense building materials (and therefore larger μ) and the rural area by smaller μ (dry surfaces). Seasonal changes in the value of μ often affect rural areas, and highlight the importance of remembering that the UHI is a difference measure. Model simulations and observations suggest that the heat islands of humid tropical cities may be expected to be small because of large rural μ values, whereas tropical dry cities may have larger heat islands (see Table 1).

Table 1 Suggested causes of the UHI[a]

Cause	Description
Surface geometry	Increased surface area and trapping of solar radiation by multiple reflections leads to warming due to greater absorption of solar radiation Closely spaced buildings reduce the sky view factor and reduce radiative heat loss, particularly at night Sheltering of buildings reduces convective heat loss from the surface and near-surface air
Surface thermal properties	Urban building materials are better stores of heat; they have higher heat capacity, and larger surface thermal admittance
Surface conditions (dryness, snow cover)	Waterproofing of the urban surface by buildings and pavement reduces evaporation; more energy is channeled into sensible heat that can heat the air rather than into latent heat (heat taken up by evaporation of water) During periods of snow cover, plowing of city surfaces and increased soiling of urban snow cover can yield increased absorption of radiation in the city
Anthropogenic heat	Heat released by urban energy use in buildings and vehicles and from humans
Urban greenhouse effect	The polluted and warmer urban atmosphere emits more thermal radiation downwards towards the city surface. Enhanced urban humidity may also contribute to this effect

[a] Adapted from Oke (1982); Oke *et al.* (1991).

Atmospheric Controls

The primary atmospheric controls on UHI development are wind and cloud (Table 2). The strongest UHIs are observed when skies are clear and winds are calm. A clear sky enhances rural cooling by allowing radiative heat losses to the relatively cold night sky. As wind speeds increase, turbulent mixing exponentially reduces temperature differences in the near surface air. Atmospheric humidity acts similarly to clouds by reducing the net radiative cooling potential of the surface. High atmospheric humidity will reduce heat island intensity; lower humidity allows greater surface radiative cooling.

Advection is also an important control on UHI intensity. For example, with onshore winds, coastal cities may experience advective cooling of urban temperatures in the summer season when sea surface temperatures are cooler than those of the land, and advective warming during the winter season when the sea surface temperature is relatively warmer. Under such conditions, it is important to assess the relative impact of advection on both the rural and urban stations, because the UHI is a difference measure. Local-scale advection induced by heat island circulations can also modify local temperatures and humidity and alter cooling rates.

Table 2 Atmospheric controls on UHI magnitude

Control	Effect
Wind speed	Increased winds increase turbulent mixing, suppress UHI magnitude
Wind (advection)	Horizontal transport of heat: warm advection can enhance UHI, cold advection may suppress UHI[a]
Cloud	Cloud reduces radiative cooling by the surface Low, thick clouds have the largest effect (warmer), high thin clouds the least
Humidity	Variable effects

[a] Effects depend on the spatial extent of advection and the rural and urban measurement locations.

Heat Island Intensity

Combining the most favorable atmospheric and surface conditions, the maximum observed heat island magnitudes are approximately 12 °C (Figure 3). The maximum heat island intensity can be related to city population, but is more fundamentally related to surface characteristics such as the SVF. More typically, heat island intensities are in the range 1 3 °C, even for large cities, indicating the importance of wind and cloud effects. Seasonally, heat islands in temperate climates are a maximum during summer and fall when weather conditions are more favorable.

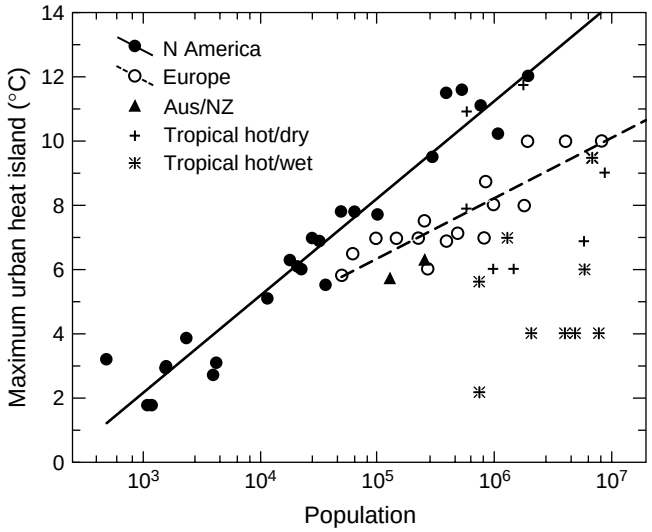


Figure 3 Maximum UHI magnitude plotted versus population. Slope of the relationships differ for cities in different parts of the world depending in part on the urban characteristics, the characteristics of the rural surroundings, and the prevailing climate. (Reproduced from Oke (personal communication))

IMPACTS OF UHI

Local Impacts

UHIs can provide both positive and negative effects for cities (see Table 3). For cities in warm climates, or temperate climate cities in the summertime, the UHI increases the use of energy for air conditioning (Figure 4a). The increase in utility demand can feed back to increase emissions of greenhouse gases used to generate electricity, which can further enhance the need for cooling, and also degrade local air quality. Higher urban temperatures can increase the formation of urban smog because both emissions of precursor pollutants and the atmospheric photochemical reaction rates, which produce smog, are enhanced (Figure 4a) (*see* Urban Sulfurous and Photochemical Smog, Volume 3). During heat waves, especially in temperate areas, the UHI effect exacerbates heat stress, particularly at night (Figure 4b). Further health implications arise from the fact that urban warmth provides a preferential site for spread of vector-borne disease.

In colder climates, UHIs may yield positive effects such as reduced energy use in the cold season, reduced length of snow cover, and a longer growing season.

Global Impacts

UHIs themselves are not responsible for global warming; they are a local climate modification, and altogether urban areas comprise less than 0.25% of the Earth s surface area (Sagan *et al.*, 1979). However, UHIs do have global impacts in other ways. Many of the long term observing stations used to construct the globally averaged surface temperature record have been subject to urbanization over their period of record. The effect of the UHI can lead to a contamination of the long-term temperature record, where the ability to detect global climate change is made

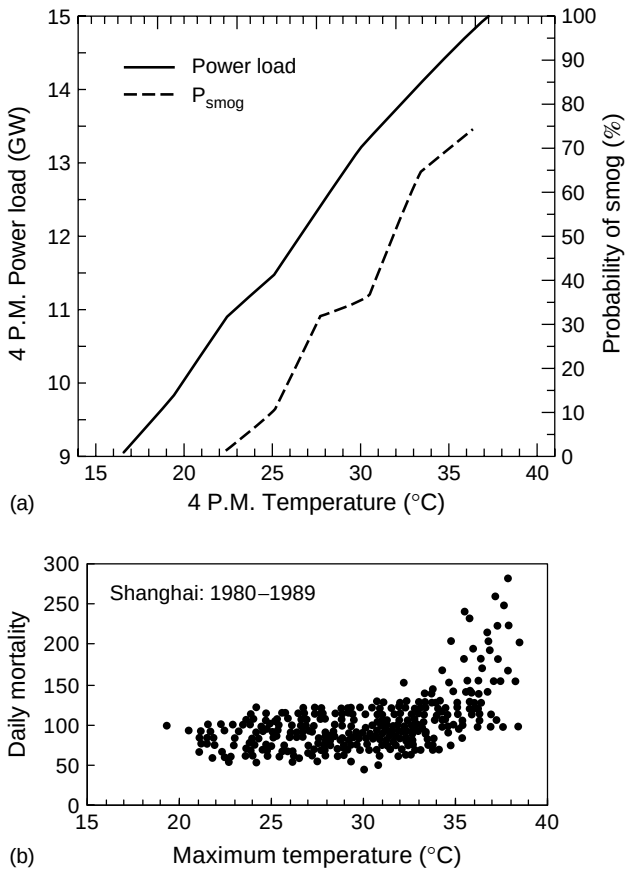


Figure 4 Impacts of warmer urban temperatures on (a) peak power load and probability of smog (reproduced by permission of Elsevier Science from Rosenfeld *et al.*, 1995) and (b) mortality (reproduced by permission of Birkhäuser-Verlag from Kalkstein and Smoyer, 1993)

more difficult because of the superposition of local climate effects. Current techniques (e.g., Karl *et al.*, 1988) used to remove urban effects may be inadequate (Oke, 1997; Changnon, 1999).

Urban areas are also the sites of most of the greenhouse gas emissions that contribute to global climate change. Emissions originating from urban areas contribute to both local and global climate modification. About half of the world s population currently lives in cities and this value is expected to increase to 61% by 2030 (UNPF, 1999). The most rapid rates of increases are occurring in the developing world (UNEP, 2000) where UHIs yield primarily negative impacts. Further urbanization will increase the area of altered local climates, the fraction of emissions originating from cities, and the number of people affected by UHI.

UHI as an Analog for Global Warming

The climate modifications that occur within urban areas may be used as an analog for examining changes anticipated from global climate change (Oke, 1997); many large cities have undergone substantial warming over the past 100 years due to local climate modifications arising from urbanization. These modified urban climates show similarities in terms of the rates and magnitude of climate change expected with projected future climate changes, and therefore cities may provide important laboratories for assessing impacts of, and adaptation strategies to climate change on both local and global scales (Oke, 1997).

MITIGATION OF URBAN HEAT ISLANDS

Reducing the magnitude of the UHI through various mitigation strategies is possible. Strategies that have received the most attention include the use of trees and vegetation to shade the surface and provide evaporation cooling, and increasing the reflectivity of surfaces such as roofs to reduce the amount of solar radiation absorbed (Rosenfeld *et al.*, 1995) (*see* Greening of Cities, Volume 3). Mitigation of the UHI can provide benefits in terms of improved air quality and human health through reduced energy use, and generally improved human comfort, particularly in cities with warm climates.

EFFECT OF GLOBAL ENVIRONMENTAL CHANGE ON URBAN HEAT ISLANDS

Finally, one might reasonably ask how UHIs may themselves be altered under anticipated global climate change. The answer to this question is not clear; the formative factors underlying the UHI will continue to operate, albeit from a warmer ambient temperature base. So we may reasonably expect UHI to continue, and to expect that cities will be warmer than they currently are. It is unlikely that the intensity of the heat island will increase with global warming because the temperature gradient between city and country is expected to remain similar. Further assessment of how the frequency of strong versus weak UHIs may change under future climates requires detailed knowledge about the type and frequency of future weather patterns expected for a given location, because of the strong controls exerted on the UHI intensity by weather factors. This information is poorly known at present. At best it seems reasonable to expect that, under future climates, cities will be warmer than they are now with the combined local impacts of the UHI and global environmental change, contributing negatively to the environment of future urban areas in warm climates, and the summer conditions in cities in cold climates, but with a possible overall benefit in temperate and cold climate regions. The high rate of urbanization, particularly in tropical cities, means that increasing numbers of people will be exposed to impacts resulting from UHI in the future.

REFERENCES

Changnon, S A (1999) A Rare Long Record of Deep Soil Temperatures Defines Temporal Temperature Changes and an Urban Heat Island, *Clim. Change*, 42, 531 538.

Karl, T R, Diaz, H F, and Kukla, G (1988) Urbanization: its Detection and effect in the United States Climate Record, *J. Clim.*, 1, 1099 1123.

Kalkstein, L S and Smoyer, K E (1993) The Impact of Climate Change on Human Health: some International Implications, *Experientia*, 49, 969 979.

Oke, T R (1982) The Energetic basis of the Urban Heat Island, *Q. J. R. Meteorol. Soc.*, 108, 1 23.

Oke, T R (1997) Urban Climates and Global Change, in *Applied Climatology: Principles and Practices*, eds A Perry and R Thompson, Routledge, London, 273 287.

Oke, T R, Johnson, G T, Steyn, D G, and Watson, I D (1991) Simulation of Surface Urban Heat Islands Under Ideal Conditions at Night. Part 2: Diagnosis of Causation, *Bound.-Layer Meteorol.*, 56, 339 358.

Rosenfeld, A H, Akbari, H, Bretz, S, Fishman, B L, Kurn, D M, Sailor, D, and Taha, H (1995) Mitigation of Urban Heat Islands: Materials, Utility Programs, Updates, *Energy Build.*, 22, 255 265.

Sagan, C, Toon, O B, and Pollack, J B (1979) Anthropogenic Albedo Changes and the Earth s climate, *Science*, 206, 1363 1368.

Taha, H (1997) Urban Climates and Heat Islands: Albedo, Evapotranspiration, and Anthropogenic Heat, *Energy Build.*, 25, 99 103.

UN (1998) *World Urbanization Prospects: the 1996 Revision*, United Nations, New York.

UNEP (2000) *Global Environment Outlook 2000*, United Nations Environment Program, Nairobi.

UNPF (1999) *The State of World Population 1999*, United Nations Population Fund, New York.

FURTHER READING

Brown, R D and Gillespie, T J (1995) *Microclimatic Landscape Design: Creating Thermal Comfort and Energy Ef ciency*, John Wiley & Sons, New York.

Landsberg, H E (1981) *The Urban Climate*, Academic Press, New York.

Roth, M, Oke, T R, and Emery, W J (1989) Satellite-derived Urban Heat Islands from three Coastal Cities and the Utilization of such Data in Urban Climatology, *Int. J. Remote Sens.*, 10, 1699 1720.

Urban Metabolism

see Anthropogenic Metabolism and Environmental Legacies (*Opening essay*, **Volume 3**)

Urban Population Change

David Satterthwaite
International Institute for Environment and Development, London, UK

By 2000, the world's urban population totaled some 2.8 billion people - a near four-fold increase in size since 1950 (United Nations, 2000) and a more than 10-fold increase since 1900 (Graumann, 1977). The proportion of the world's population now living in urban centers is close to 50%, having risen from less than 30% in 1950 and less than 15% in 1900. It is likely to exceed 50% within the rst decade of the new millennium (United Nations, 2000). Many aspects of urban change are unprecedented, including the size of the world's urban population, the number of countries becoming more urbanized and the size and number of very large cities.

It is important to understand these urban trends in terms of the very large economic, social, political and demographic changes that have underpinned or in uenced them. Foremost among these is the much increased scale in the world's economic activities and the shifts this has brought in employment patterns, from agricultural and pastoral activities (highly concentrated in rural areas) to industrial, service and information activities (highly concentrated in urban areas). The globalization of world production and trade (including the very rapid expansion in the value of international trade) has been an important part of this and has in uenced urban trends in most nations. Many cities owe their prosperity to key roles within this increasingly internationalized production system, including the world cities (see section on world cities herein). There are also the demographic changes that so in uence urban change, including rapid population growth rates in much of Latin America, Asia and Africa after World War II (although in most cases these have declined signi cantly) and changes in the size and composition of households and age structures (UNCHS (Habitat), 1996).

There are also many political changes that have had profound in uences on urban change in the last 50 years in many nations, especially the dissolving of the European powers' colonial empires that helped underpin rapid urban change in most of Africa and Asia between the 1950s and the 1980s, and, more recently, the dissolving of the Soviet Union and its economic bloc and the breaking up or reshaping of many nations in Europe. Wars and civil con icts in Europe and Africa have brought major shifts in populations; the number of international refugees has risen to unprecedented levels and a considerable proportion have come to live in cities, either for refuge or to seek new bases for their livelihoods (Castles and Miller, 1993). Changed political structures and economic policies have also brought major

urban changes in many other areas - notably in China with its very rapid economic growth and urban change since the late 1970s. In much of Latin America, urban systems and trends have been reshaped during the 1980s and 1990s by the introduction of or return to democratic rule, the shift in economic policies from import substitution to export promotion, serious economic problems and, in many nations, decentralization and stronger democracy at the level of local government. Thus, while we can discern broad trends towards increasingly urbanized societies in much of the world, the scale and nature of such trends and their underlying causes differ very much from country to country and even within each country, regionally and over time. In looking at global change, it is important not to forget how unique social, economic, political and demographic structures are in uencing urban change within each location.

INTRODUCTION

Despite the long-term trend towards more urban societies, more than half the world s population still lives in rural areas and a significant proportion of the urban population lives in small administrative centers and market towns in which much of the workforce depends on agriculture or on enterprises serving rural populations. Only a very small proportion of the world s population lives in very large cities (see section on the world s largest cities herein). Despite the speed at which urban populations have grown, there is also a perhaps surprising continuity among the world s most important cities. As the section herein on the regional distribution of the world s urban population will describe in more detail, most of the world s largest cities have long histories as important cities, and most such cities have also experienced significant slow-downs in their rate of population growth. Many had more people moving out than moving in during the last recorded inter-census period. And for many nations, the speed with which they urbanized slowed down during the 1980s, in part because of serious economic problems. The world in 2000 is less urbanized and far less dominated by very large cities than most people had predicted. For instance, Mexico City likely has around 18 million people by 2000 not the 31 million people predicted 20 years ago. Calcutta likely has around 13 million, not the 40 50 million people that was predicted during the 1970s.

We also know little about the scale and nature of urban change in recent years. For most nations, the most recent figures available for urban populations come from censuses held in 1990 or 1991, and we have little idea of the scale of urban change within nations and the growth (or decline) of city populations since then. This is why data on city populations in the tables in this article are for 1990. The scale and nature of urban trends during the 1990s will only become apparent when new data become available for censuses held in 2000 and 2001. All regional and global figures for 1995 or 2000 are therefore estimates made by extrapolating trends from the 1980s. For many African nations, there is no census data for the last 15 20 years so even regional data for 1990 may be inaccurate. There is also the fact that each nation uses different criteria regarding how it defines its urban population (or distinguishes urban centers from other settlements), which means that the scale of the world s urban population is strongly influenced by the urban criteria used within the largest population nations. If the Indian or Chinese government chose to change the criteria used in their censuses to define urban centers, this could increase or decrease the proportion of the world s urban population by several percentage points and there are good reasons for thinking that the current criteria used in China considerably understate the size of its urban population (Kirkby, 1994). Revisions by, for instance, the Nigerian or Brazilian census authorities could significantly alter their region s level of urbanization. Thus, the world s level of urbanization is best understood not as a precise figure (47.4% in 2000) but as being between 40 55%, depending on what criteria are used to define an urban center.

Another source of uncertainty regarding urban change arises from the different ways in which large city boundaries can be defined for instance, is a city s population that which is within the city proper, the built up area, the metropolitan area or a wider planning region? The current population of most of the world s largest urban areas, including London, Los Angeles, Shanghai, Beijing, Jakarta, Dhaka and Bombay/Mumbai can go up or down by many million inhabitants, depending on which boundaries are used to define the area within which their population is counted (Satterthwaite, 1996; UNCHS, 1996). Different boundaries also mean different population growth rates so London, Los Angeles, Tokyo, Buenos Aires or Mexico City can be correctly stated as having populations that have declined or expanded in recent decades, depending on which boundaries are used. This is why any international list of the world s largest cities and their population growth rates risks considerable inaccuracy, as some city population figures are for large urbanized regions with thousands of square kilometers while others are for older city boundaries with a few hundred square kilometers.

THE REGIONAL DISTRIBUTION IN THE WORLD'S URBAN POPULATION

Most of the world s urban population is now outside Europe and North America (Table 1). Asia alone contains close to half the world s urban population, even if more than three-fifths of its population is still in rural areas. Even Africa, which is generally perceived as overwhelmingly rural, now

Table 1 Some basic urban statistics

	Urbanization level		Proportion of the world's urban population		Number of the world's 100 largest cities			
	1950	1995	1950	1995	1800	1900	1950	1990
Africa	**14.6**	**34.9**	**4.4**	**9.8**	**4**	**2**	**3**	**6**
Eastern Africa	5.3	22.4	0.5	1.9	0	0	0	0
Middle Africa	14.2	32.7	0.5	1.1	0	0	0	1
Northern Africa	24.7	48.0	1.8	3.0	3	2	2	4
Southern Africa	38.2	47.8	0.8	0.9	0	0	1	0
Western Africa	10.2	36.3	0.9	3.0	1	0	0	1
Americas	**52.8**	**74.5**	**23.8**	**22.4**	**3**	**21**	**26**	**29**
Caribbean	35.4	61.7	0.8	0.9	1	0	1	1
Central America	39.8	66.3	2.0	3.2	1	1	1	3
Northern America	63.9	76.2	14.6	8.8	0	16	18	13
South America	42.9	77.4	6.4	9.6	1	4	6	12
Asia	**17.4**	**34.7**	**32.6**	**46.3**	**64**	**23**	**32**	**44**
Eastern Asia	18.0	36.9	16.1	20.4	29	16	18	22
South-central Asia	16.6	28.7	11.1	15.3	24	5	8	12
South-eastern Asia	14.8	33.5	3.6	6.3	5	1	5	7
Western Asia	26.7	67.6	1.8	4.4	6	1	1	3
Europe	**52.4**	**73.5**	**38.2**	**20.8**	**29**	**51**	**37**	**19**
Eastern Europe	39.3	70.2	11.5	8.5	2	9	7	4
Northern Europe	72.7	83.5	7.6	3.0	6	16	7	1
Southern Europe	44.2	64.1	6.4	3.6	12	8	8	6
Western Europe	67.9	81.5	12.8	5.7	9	18	15	8
Oceania	61.6	70.1	1.0	0.8	0	2	2	2
World	29.7	45.3						

The statistics for 1950 and 1995 were taken or derived from data in United Nations (1998). The data on the distribution of the world's 100 largest cities is an updated version of a table taken from Satterthwaite (1996).

has a larger urban population than either North America or South America.

Levels of urbanization in certain regions increased dramatically between 1950 and 1995 (Table 1) for instance, in Africa from 15 31%, in West Asia from 27 68% and in Eastern Europe from 39 70%. However, the growth rates of urban populations and the rates of increase in levels of urbanization are not unprecedented (Preston, 1979; Satterthwaite, 1996) Many countries in West Europe or North America had periods when their level of urbanization increased as rapidly. In addition, many cities in North America figure among the world s fastest growing cities during the 20th century (see section on world s most rapidly growing cities herein).

It is also important not to assume that the factors underlying urban change are comparable across time or across nations. For instance, one of the main reasons why many cities in Africa grew so rapidly during the 1960s and 1970s was that they began from such a small base, because the European colonial powers had kept down their populations by imposing restrictions on the rights of their national populations to live and work in urban centers. Thus, one of the reasons why urban populations grew so rapidly just before or after the end of colonial rule was the removal or weakening of the colonial apartheid-like controls on population movements (see Potts, 1995). Another reason why the urban population of many sub-Saharan African cities grew so rapidly during the 1960s and 1970s was the achievement of political independence and the development in their urban centers of the institutions associated with an independent state and the development of a higher education system that had been so undeveloped under colonial rule. A third reason was the division of what, under colonial rule, had been one country into two or more, with each new country needing a capital (Bairoch, 1988).

Data on how the regional distribution of the world s 100 largest cities has changed from 1800 through to 1990 (Table 1) show how, historically, Asia has always had a high proportion of the world s largest cities, and that South and Central America and North Africa also have long had large cities (Chandler and Fox, 1974; Bairoch, 1988). By 1990, Asia had 44 of the world s 100 largest cities and this proportion is likely to increase. In most of the world s regions, there is a perhaps surprising continuity in the list of the largest cities over time. An analysis of population trends among what are currently the world s largest cities found that more than two-thirds of the world s million-cities in 1990 were already important cities 200 years ago while around one-quarter have been important for at least 500 years (Satterthwaite, 1996). This analysis also showed

a perhaps surprising comparison in that it is North America and sub-Saharan Africa that stand out as having most new cities among their largest cities, i.e., cities that now have more than a million inhabitants that had not been founded or did not exist as urban centers by 1800 (ibid). Table 1 also highlights how few of the world s largest cities Europe had in 1990, unlike in 1900 when it had more than half the world s 100 largest cities. In part, this reflects the growing economic importance of other regions. But in part, it reflects the way that urban form has changed with much of the working population of European cities not living within city boundaries.

THE WORLD'S LARGEST CITIES

Most of the world s largest cities had relatively slow population growth rates during the 1980s (Table 2); more than half had growth rates of less than 2% a year. Many of these cities had more people moving out than moving in (so their population growth rate was lower than their rate of natural increase) including Mexico City, Sao Paulo, Calcutta, Buenos Aires and Rio de Janeiro. Among these 30 cities, only two had a population growth rate that exceeded 5% a year during the 1980s Dhaka and Lagos and the estimate for Lagos is probably too high (the 1991 census suggested a much lower population growth rate). However, as Table 2 shows, very large cities with relatively modest population growth rates can still have large annual average increments in their populations.

One notable change in recent decades is the scale of the world s largest cities and the size to which they have grown is unprecedented. While there are examples of cities over the last two millennia with populations of one million or more inhabitants, the city or metropolitan area with several million inhabitants is a relatively new phenomenon, London being the first to reach this size in the second half of the 19th century (Chandler and Fox, 1974). The average size of the world s largest cities has increased considerably. In 1990, the average size of the world s 100 largest cities was around 5.1 million inhabitants. This compares to 2.1 million inhabitants in 1950, around 700 000 in 1900 and just under 200 000 in 1800 (Satterthwaite, 1996).

Table 2 The world's largest urban agglomerations in 1990

Urban agglomeration	Population (thousands – 1990)	Annual average growth rate 1980–1990 (%)	Annual average increment in population 1980–1990 (thousands)
Tokyo	25 069	1.4	322
New York	16 056	0.3	46
Mexico City	15 130	0.9	124
Sao Paulo	15 082	1.5	214
Shanghai	13 342	1.3	160
Bombay/Mumbai	12 246	4.3	418
Los Angeles	11 456	1.9	193
Buenos Aires	11 144	1.2	122
Calcutta	10 890	1.9	186
Beijing (Peking)	10 820	1.8	179
Seoul	10 544	2.4	226
Osaka	10 486	0.5	50
Rio de Janeiro	9682	1.0	94
Paris	9334	0.4	40
Moscow	9048	1.1	91
Tianjin	8786	1.9	152
Cairo	8633	2.3	178
Delhi	8207	4.0	265
Manila	7968	3.0	201
Karachi	7945	4.7	292
Lagos	7742	5.8	336
London	7653	(0.1)	(9)
Jakarta	7650	2.5	167
Chicago	6792	0.0	1
Istanbul	6544	4.1	215
Teheran	6360	2.3	129
Essen	6353	0.0	2
Dhaka	6219	6.8	301
Bangkok	5901	2.3	118
Lima	5826	2.8	143

Although there is a large and growing literature on what are termed mega-cities (Rakodi, 1997; Gilbert, 1996; Lo and Yeung, 1996), which is well justified by the problems faced in managing such large population concentrations and by their key economic and political roles, two points should be noted: the very small proportion of the world s population that actually lives in them and the extent to which they are concentrated within the world s largest economies.

There is no agreement as to the definition of a mega-city but if a city becomes a mega-city only when it has 10 million or more inhabitants, there were only 12 mega-cities by 1990 and these included less than 3% of the world s population. The association between the *per capita* gross national product (GNP) of a country and its level of urbanization is well known i.e., that in general, the higher the *per capita* GNP, the higher the level of urbanization. Most of the nations with the most rapid increases in their levels of urbanization 1960 to 1990 were also the nations with the most rapid economic growth. What is perhaps less well known is the high concentration of the world s largest cities in the world s largest economies. In 1990, the five largest economies (US, China, Japan, Germany and the Russian Federation) had half of the world s mega-cities and 37% of the world s million-cities (Satterthwaite, 1996). All but one of the world s other mega-cities were in the next 10 largest economies which also concentrated close to one-quarter of the world s million-cities.

THE WORLD'S MOST RAPIDLY GROWING LARGE CITIES

Although the historical continuity in the world s largest cities was noted above, there are many important cities today that did not exist in 1900 or were only very small settlements. In some nations, the emergence of new large cities is fundamentally changing the spatial distribution of urban populations. For instance, in the US, what was originally considered as de-urbanization in the 1970s and 1980s was in part the emergence of new very large cities in the South and West which came to challenge the size and economic importance of the large Eastern cities (Bourne, 1995).

Very large cities rarely figure in lists of the world s most rapidly growing cities, although they inevitably did so when they were smaller. The larger a city s population at the beginning of any period for which growth rates are being calculated, the larger the denominator used to divide the increment in the city s population to calculate its growth rate. Although cities such as Lagos, Karachi and Dhaka are often said to be among the world s fastest growing cities, there are hundreds of smaller cities with higher population growth rates. In most nations, an analysis of urban centers growth rates in any inter-census period during the 20th century will show some relatively small urban centers with population growth rates of • 10% per year far above even the fastest growing city in Table 1.

Although it is often assumed that the world s most rapidly growing cities are concentrated in Latin America, Asia and Africa, this is not entirely true. Table 3 shows how several cities in the US rank among the most rapidly growing cities in the world during the 20th century. For instance, Nairobi is often held up as one of the world s most rapidly growing cities, but both Miami and Phoenix in the US had larger populations than Nairobi in 1990, although all were small settlements in 1900. The population of Los Angeles was around one-tenth that of Calcutta in 1900, yet in 1990, it had about the same number of people in its metropolitan area. The growth of Tampa (US) from a small town in 1900 to a metropolitan area of over 2 million by 1990 or Las Vegas from a small town in 1950 to a metropolis of more than a million inhabitants by 1995 also place them among the world s most rapidly growing cities during the 20th century. This should not be taken to imply that the social, economic, demographic or political underpinnings are comparable. In addition, these and other examples of rapidly growing cities in high-income countries do not alter the fact that most of the large cities in the world with the fastest population growth rates in recent decades are in low and middle-income countries.

THE WORLD CITIES

Structural changes within the world economy have helped to reorder the relative importance of cities around the world. Certain regions and cities have been particularly successful in attracting new investment and concentrating innovation for instance, Silicon Valley and Orange County in California, Bavaria in Germany, and certain cities in southern Europe and the Asian tigers (Castells and Hall, 1994). Meanwhile, many of the most successful regions had urban forms that were less dominated by a large central city and that had new enterprises developing in a network of smaller cities and greenfield sites. Most major port cities and traditional industrial centers in Europe and North America (and some elsewhere) that grew to become among the world s major cities during the second half of the 19th and first half of the 20th century lost importance and most had population declines during the 1980s.

The large and increasing share of the world s economy controlled by multinationals has led to certain cities becoming what are often termed global or world cities, as they are key command and control points of the world economy (Friedmann and Wolff, 1983; Sassen, 1994). Some are unambiguously world cities as they are the commanding nodes of the global system London, New York and Tokyo (Sassen, 1994). It is interesting to note that Tokyo

also remains one of the world s major concentrations of industrial production whereas New York and London do not. It is a reminder of how advanced telecommunications systems and logistical management of production can separate almost completely, in spatial terms, the production process from those who manage and finance it and the producer services that the managers and financiers need (ibid).

Another category of world cities are those that articulate large national economies into the system (Paris, Madrid, Sao Paulo) or subnational (regional) economies (Chicago) or simply have a commanding multinational role (Miami, Singapore) (Friedmann, 1993). Global cities are not necessarily the same as large cities as many cities that are not among the world s largest have major international roles for instance, Singapore and Zurich. Several of the world s largest cities do not owe their size and economic base to their role within global production but to being national capitals in more populous nations with a high concentration of political power there; for instance, Delhi and Cairo.

A PREDOMINANTLY URBAN FUTURE?

Most publications about the future predict or assume that the world will continue to urbanize far into the future, even if most now admit that the speed of the change and the size to which the largest cities are likely to grow have been overstated. In its most recent projections, which considerably reduce the scale of urban growth predicted up to 2025 compared to previous projections, the United Nations (UN) suggests that the world will have 59% of its population living in urban areas by the year 2025 (UN, 1998). Africa is predicted to have some 52% of its population living in urban areas by then. But such projections should be viewed with caution. Given the historic association between economic growth and urbanization, a steady increase in the level of urbanization in less urbanized nations is only likely to take place if they also have steadily growing economies. While stronger and more buoyant economies for the world s lower-income nations should be a key development goal, the current prospects for most such nations are hardly encouraging, within the current world economic system. Many of the lowest income nations have serious problems with political instability or civil war and most have no obvious comparative advantage on which to build an economy that prospers and thus urbanizes.

There are also grounds for doubting whether a large proportion of the world s urban population will come to live in very large cities. As noted above, many of the largest cities have very slow population growth rates. For some, further expansion is hindered by resource shortages; for instance, water shortages will inhibit the development of many large (and smaller) cities in semi-arid regions. Many large coastal cities are particularly at risk from the sea level rises and more frequent and severe extreme weather events that will come with global warming. Meanwhile, much new investment is going to particular medium size cities well located in relation to the largest cities, and to transport and communications systems, and these are helping to create less primate-city dominated urban systems. In addition, in regions with advanced transport and communications systems, rural inhabitants and enterprises can enjoy standards of infrastructure and services and access to information that historically have only been available in urban areas. Thus, both low- and high-income nations may urbanize less than anticipated, although for very different reasons.

REFERENCES

Bairoch, P (1988) *Cities and Economic Development: From the Dawn of History to the Present*, Mansell, London, 1 574.

Bourne, L S (1995) *Urban Growth and Population Redistribution in North America: a Diverse and Unequal Landscape*, Major Report 32, Center for Urban and Community Studies, University of Toronto, Toronto, 1 41.

Castells, M and Hall, P (1994) *Technopoles of the World: the making of 21st century Industrial Complexes*, Routledge, London, 1 275.

Castles, S and Miller, M J (1993) *The Age of Migration: International Population Movements in the Modern World*, MacMillan, London, Basingstoke, 1 306.

Chandler, T and Fox, G (1974) *3000 Years of Urban Growth*, Academic Press, New York.

Friedmann, J (1993) *Where we Stand: a Decade of World City Research*, Paper prepared for the Conference of World Cities in a World System, Center for Innovative Technology, April, 1 37.

Friedmann, J and Wolff, G (1983) World City Formation, *Int. J. Urban Regional Res.*, 6(3), 309 343.

Gilbert, A, ed (1996) *The Mega-city in Latin America*, United Nations University Press, Tokyo, 1 282.

Graumann, J V (1977) Orders of Magnitude of the World s Urban and Rural Population in History, *UN Popul. Bull.*, 8, United Nations, New York, 16 33.

Kirkby, R (1994) Dilemmas of Urbanization: Review and Prospects, in *China: The Next Decades*, ed D Dwyer, Longman Scientific and Technical, Harlow, 128 155.

Lo, F C and Yeung, Y M, eds (1996) *Emerging World Cities in Paci c Asia*, United Nations University Press, Tokyo, 1 528.

Potts, D (1995) Shall we go Home? Increasing Urban Poverty in African Cities and Migration Processes, *Geogr. J.*, 161(3), November, 245 264.

Preston, S H (1979) Urban Growth in Developing Countries: a Demographic Reappraisal, *Popul. Dev. Rev.*, 5(2), 195 215.

Rakodi, C, ed (1997) *The Urban Challenge in Africa - Growth and Management of its Large Cities*, United Nations University Press, Tokyo, 1 628.

Sassen, S (1994) *Cities in a World Economy*, Pine Forge Press, Thousand Oaks, London, New Delhi, 1 157.

Satterthwaite, D (1996) *The Scale and Nature of Urban Change in the South*, IIED Working Paper, IIED, London, 1 29.
UNCHS (Habitat) (1996) *An Urbanizing World: Global Report on Human Settlements, 1996*, Oxford University Press, Oxford.
United Nations (1998) *World Urbanization Prospects: the 1996 Revision*, Population Division, Department of Economic and Social Affairs, United Nations ST/ESA/SER.A/170, 1 190.
United Nations (2000) *World Urbanization Prospects: the 1999 Revision*, Population Division, Department of Economic and Social Affairs, ESA/P/WP.161, 1 128.

Urban Poverty and Environmental Health

David Satterthwaite
International Institute for Environment and Development, London, UK

This contribution concentrates on the environmental health problems associated with urban poverty and overcrowding in countries where such problems are particularly acute. As such, it concentrates on urban areas in Africa, Asia and Latin America. This does not mean that there is no association between urban poverty and environmental risk in other regions such as Europe and North America. All over the world, low income groups generally face more serious environmental health problems than wealthier groups (as they cannot afford housing with the best environmental quality and as they take on more dangerous work) but the scale of the environmental health burden suffered by low income groups in high income countries is generally much smaller. So too are the differentials between low and high income groups. For instance, in many urban areas of Africa, Asia and Latin America, the risk of a child dying before reaching the age of ve is often 20 to 30 times higher in the poorest quality low income areas compared to areas with a predominance of middle or upper income groups and environmental health problems contribute much to this differential.

This concentration on environmental health problems associated with urban poverty should not be taken to imply that environmental health problems are less serious for the rural poor. But there are important differences in context: for instance as urban areas have different occupational structures (and associated health risks) and as they concentrate people, production and their wastes which contributes to very serious environmental health problems, in the absence of effective environmental and public health policies.

WHAT ENVIRONMENTAL HEALTH INCLUDES

The World Health Organization s (WHO) definition of environmental health includes not only the health aspects of the human environment but also the technical and administrative measures for improving the human environment from a health perspective (WHO, 1979, 1989). This means it must include a concern for the different environments in which women, men and children live and work; including homes and residential neighborhoods, workplaces, schools and other public facilities and the roads, walkways and transport systems that connect them. It also implies a concern for the environmental aspects of disease and injury prevention which include prevention oriented health care and disaster preparedness policies.

To help understand and manage them, environmental health problems can be grouped into various categories. The different ways of doing so reflect different professional disciplines or priorities and the ways in which different problems have been assigned to different public agencies. The basis for categorizing them can be:

- the nature of the problem (or the hazard it presents to people) e.g., infectious or parasitic disease-causing agents, chemical pollutants and physical hazards, lack of access to resources (for instance freshwater) and the generation of hazardous wastes;
- the context within which the environmental problems occur e.g., housing, workplace, school, city-wide problems;
- the sector into which the problem falls and/or the division of institutional responsibilities for its resolution (e.g., solid waste management, air pollution control);
- (for infectious and parasitic diseases) the biology of the disease causing agent or based on the media through which infection takes place (air, food, water, soil, animals, insects or other organisms that are disease vectors or carriers);
- (for chemical pollutants) the kind of pollutant, the concentration and whether it arises from point or non-point sources.

Here we consider the environmental health context within the residential areas in which low income groups are concentrated, and include some discussion of other serious environmental health risks and who is most vulnerable to them. But prior to this, there is a brief discussion of urban poverty and of its multiple links with environmental health.

URBAN POVERTY

Although historically, the definition and measurement of urban poverty have generally been based only on income

level or consumption level, there is now widespread agreement that this narrow focus excludes other important aspects of deprivation (Wratten, 1995; UNCHS, 1996; IDS, 1997). These include:

- *inadequate, unstable or risky asset bases.* Assets are particularly important for allowing low income groups to avoid deprivation because they provide the means by which economic stresses or shocks can be managed, including being able to afford the cost of treatment when sick or injured and to cope with the loss of earning when an income earner cannot work;
- *limited or no right to make demands within the political system.* The possibilities for poorer groups of obtaining appropriate public responses are generally much less if there is no legislative and political framework that guarantees civil and political rights: for instance the right to have representative government, the right to organize and to make demands and to get a fair response, and the right to procedural justice. Inadequacies in the provision for water, sanitation, drainage, garbage collection and health care for the urban poor and the lack of protection against forced eviction are often linked to undemocratic and unaccountable political structures;
- *poor quality/insecure housing and lack of basic services.* From an environmental health perspective, the most serious problem is housing that has inadequate provision for the infrastructure and services that are essential for environmental health. This inadequate provision imposes a large health burden and also high economic and other costs.

Another important aspect of deprivation which also contributes to poverty is *discrimination.* For instance for women, there is the discrimination they face in labor markets and in obtaining housing, credit and services or the discrimination faced by certain groups based on their ethnic origin or caste.

If urban poverty is understood in terms not only of inadequate income but also of deprivation associated with lack of assets, political influence and access to basic services, then urban poverty is itself the most significant predictor of environmental health problems (see WHO, 1995a; Bradley *et al.*, 1991).

HOUSING AND BASIC SERVICES

Housing and the basic services it should contain (including adequate provision for water, sanitation, drainage and garbage removal) provide the primary line of defence against environmental health hazards in the locations where most people spend most of their time. Estimates for 1990 suggested that at least 600 million urban dwellers in Africa, Asia and Latin America live in life and health threatening shelters and neighbourhoods because of poor quality and generally overcrowded housing with inadequate infrastructure and service provision (Cairncross *et al.*, 1990; WHO, 1992, 1999). Most of these live in tenements, boarding houses or precarious housing often built on illegally occupied or subdivided land. They also included tens of millions of people who have no shelter at all, including those who sleep outside (on pavements, in shop doorways, in parks or under bridges) or in public buildings such as railway, bus or metro stations (UNCHS, 1996). Although considerable progress has been made since 1990 in improving housing conditions, for low income groups, in certain cities and more effective approaches are now being more widely applied, the number of people suffering serious environmental health burdens in urban areas probably increased significantly during the 1990s. This is partly because urban populations continue to grow rapidly in most of Africa, Asia and Latin America, partly because of weak and ineffective urban governance, and partly because of continued increases in urban poverty in many nations (see for instance Kanji, 1995; Minujin, 1995).

The association between urban poverty and serious environmental health problems is obvious. Virtually all of those living in life and health threatening homes and neighbourhoods subsist on below poverty-line incomes, and it is their low income which is the main reason why they do not move to less dangerous kinds of accommodation. The health burdens they suffer and their economic consequences (in terms of work days lost and costs of treatment) are a major cause of their poverty (see for instance Pryer, 1993). In cities where most of those with low incomes live in shelters with adequate provision for water, sanitation, electricity and drainage and have access to health care, the environmental health burden associated with poverty is much reduced. The contribution that improved housing and service provision can make to poverty reduction is often forgotten, especially where urban poverty is conceived only in terms of income-levels.

Most of the environmental health burden for those living in poor quality housing that lacks basic services arises from infectious and parasitic diseases. These are environmental because they are transmitted through environmental media (air, water, soil, food) or via insect or animal vectors. Many diseases thrive when provision for water, sanitation, drainage, garbage collection and health care is inadequate or where it breaks down. As a result, cities can become among the most health threatening of all human environments as disease causing agents and disease vectors multiply, as the large concentration of people living in close proximity to each other increases the risk of disease transmission, and as health care systems become unable to respond rapidly and effectively (WHO, 1999).

At any one time, close to half of the urban population in Africa, Asia and Latin America is suffering from one or more of the main diseases associated with inadequate provision for water and sanitation (WHO, 1999). Every year, more than five million people die from illnesses linked to unsafe drinking water, improper excreta disposal and unclean domestic environments (WHO, 1995b). Improved water and sanitation can bring great benefits in terms of, improved health, reduction of expenditures (on water vendors and on treatment from diseases) and much reduced physical effort (especially for those who have to collect and carry water from standpipes or other sources far from their shelters).

The proportion of urban dwellers with provision for piped water and sanitation has increased over the last 20 years but with the rapid increase in urban populations, the number of people not served by water supplies and sanitation has increased, not declined. Recent estimates suggest that close to 300 million urban dwellers are still not served by water supplies and close to 600 million are without sanitation. But these figures understate the problem. For water supply to be adequate, it must be of good quality, close-by, and readily available. Many governments include in their official statistics of people adequately served all households with public standpipes, even when hundreds of people have to share each standpipe and fetching and carrying water from distant standpipes is a time consuming and arduous task. In addition, many public standpipes are poorly maintained or connected to water supply systems that function intermittently (Hardoy *et al.*, 2001).

For low income groups, provision for the removal and safe disposal of household and human wastes is generally even worse than provision for water. Most government statistics on provision for sanitation exaggerate the proportion of the urban population adequately served, as they include as adequately served those who have to share poorly maintained latrines or rely on communal latrines that are poorly maintained, difficult to access and lack facilities for washing. Tens of millions of urban dwellers have no access to a toilet and so have to defecate outside or into plastic bags (wrap-and-throw) (ibid). It is also common for much of the solid wastes generated not to be collected. It is usually the predominantly low income districts that have the least adequate provision which means that wastes accumulate on open spaces and streets, clogging drains and attracting disease vectors and pests (rats, mosquitoes, flies).

Overcrowded homes, inadequate water supplies and inadequate facilities for preparing and storing food greatly exacerbate the risk of food contamination and microbially contaminated food contributes to a high incidence of acute diarrhoea and to many other food borne diseases (McGranahan, 1991). Many environment related diseases are significant contributors to under-nutrition: including intestinal worm infections (it is common for a high proportion of the urban poor to have large worm burdens) and diarrhoeal diseases.

OVERCROWDING

Overcrowded dwellings are considered a major health risk but it is difficult to single out the precise impact of overcrowding from the other environmental health problems that are common in overcrowded housing (such as the inadequacies in provision for water, sanitation, garbage collection and facilities to safely prepare, cook and store food) and from other problems facing their inhabitants (for instance inadequate incomes leading to under-nutrition). It is also difficult to specify what aspect of overcrowding has the greatest health impact; for instance too many persons per room or per bed or inadequate indoor area per person.

Some of the world s leading causes of premature death are from airborne infections; for many, their transmission is aided by the overcrowding and inadequate ventilation that is so common in the housing used by lower income groups; from tenements and cheap boarding houses to one or two-room shacks. They include acute respiratory infections (the single largest cause of infant and child death worldwide) and tuberculosis (the single largest cause of adult death worldwide). They also include measles, also one of the largest causes of infant death and a major cause of infant and child morbidity and mortality in poor urban areas (Foster, 1990) although the health impact of measles has been considerably reduced through vaccination programmes. Overcrowding increases the risk of transmission for these and for other infectious diseases; it also increases the risk of multiple infections, the risk of infection early in life (when immune systems are still immature) and the risk of severe infection (UNCHS, 1995). However, overcrowding will have less significance as a risk factor where the population (especially children) are protected by immunization programmes and by health care services that provide rapid and appropriate treatment for acute respiratory infections.

Few studies have sought to isolate the health impacts of overcrowding. A study in the city of Bissau (Guinea Bissau) found that many persons per bed and many children under five within a family were among a range of risk factors for childhood mortality. In a low income settlement in Jakarta, persons per room and number of children under five per room were significant factors for diarrhoea and respiratory diseases (UNCHS, 1995). Accidents in the home are often among the most serious causes of injury and premature death among low income urban populations and overcrowding is likely to be a major risk factor, along with homes made of flammable materials and the use of open fires or mobile stoves and candles or kerosene lights.

THE RANGE OF ENVIRONMENTAL HEALTH PROBLEMS

Table 1 shows the range of environmental health problems that are common in urban areas in Africa and much of Asia and Latin America. The problems include not only those emphasized above but also those of the workplace and the wider city, including ambient air pollution and physical hazards from traffic accidents and disasters. These also take a heavy toll on human health. For instance, worldwide, more than 1.5 billion urban dwellers are exposed to levels of ambient air pollution that are above the recommended maximum levels and an estimated 400 000 additional deaths each year are attributable to ambient air pollution (WHO, 1999). Worldwide, around 900 000 people are killed each year by traffic accidents and many more are seriously injured (WHO, 1995a). In wealthier cities where provision for water, sanitation, drainage and garbage collection has improved, traffic accidents have often become among the most serious causes of injury and premature death. Environmental factors are also important contributors to psycho-social health problems and the contribution of these to total health burdens has been under estimated (Harpham and Blue, 1995). Low income groups in urban areas have greater exposure to the environmental stressors that can cause mental health problems e.g., overcrowding, high noise levels and poor quality infrastructure and services (ibid, Ekblad, 1993).

Looking to the future, there are also a considerable range of environmental health problems associated with global environmental change (*see* Environmental Change and Human Health: Extending the Sustainability Agenda, Volume 3) that could have very serious consequences for urban populations. For instance, global warming will bring rising sea levels and an increased frequency and severity of extreme weather conditions and this will greatly increase the risk of flooding in thousands of urban areas, including many of the world s largest urban areas that are located on coasts. There are also the health problems associated with more frequent and severe heat waves; and increased temperatures in cities can aggravate air pollution problems. Warmer average temperatures will permit an expansion in the area where many tropical diseases can occur. As in the case of most other environmental hazards, low income groups are generally least able to take measures to reduce these risks in their homes (and many of them live in high risk areas e.g., flood plains), while cities in low income countries have the least resources to build up defences against these risks.

VULNERABILITY AND SUSCEPTIBILITY

Certain people or households are more at risk from environmental hazards because they are:

1. Less able to avoid them;
2. More affected by them; or
3. Less able to cope with the illness, injury or premature death they cause.

Such individuals or households are generally termed as vulnerable groups. But to ensure a more precise understanding (from which more appropriate responses can be developed), it is worth distinguishing between susceptibility (where the increased risk is related to endogenous factors such as a person s nutritional status, the state of their immune system or genetic factors) and vulnerability (where it is social, economic or cultural conditions that increase the risk e.g., through an increased likelihood of exposure to environmental hazards or less capacity to cope with or adapt to an illness or injury).

The distinction between vulnerability and susceptibility is perhaps clearest when considering infants and young children since they are all particularly susceptible to a wide range of environmental hazards compared to most other age groups but those who live in poor quality housing are far more vulnerable. The nature of young persons susceptibility changes as they get older. They are particularly susceptible to certain chemical pollutants from the time of their conception through their development in the womb and infancy. They are particularly susceptible to many diseases from their birth onwards, until their immune system has developed (and, for many serious childhood diseases, strengthened through vaccines). Young children are particularly susceptible to physical hazards, from the time they begin to crawl: for among other reasons, no knowledge of danger and a desire to explore and experiment. However, their vulnerability to the environmental hazards to which they are particularly susceptible is much increased by poor quality housing that lacks adequate provision for water, sanitation, drainage and garbage removal and as such also lacks safe indoor and outdoor living and play environments (Bartlett *et al.*, 1999). People with limited mobility, strength and balance can be considered as more susceptible to physical hazards because they are at greater risk of injury. Young children, older groups and people with physical disabilities are more at risk from homes built on unsafe sites such as slopes or flood plains. Asthmatics are more susceptible to certain common air pollutants. Pregnant women as well as their foetuses are particularly susceptible to many external environmental agents (WHO, 1999).

Much of a person s vulnerability is determined by their income (and thus whether they can afford adequate quality housing with basic services) and by the quality and extent of provision for basic services, such as water, sanitation, drainage, garbage collection, health care and emergency services.

Some of these, such as the quality of housing and provision for water and sanitation, have a major influence

Table 1 A summary of the main city-related environmental problems

Context	Nature of hazard or problem	Some examples
Within house and its plot	Biological pathogens	Water-borne, water-washed (or water-scarce), airborne, food borne, vector-borne (including some water-related vectors). Insufficient quantity of water may be as serious in terms of health impact as poor water quality. Quality of provision for sanitation is also very important. Overcrowding/poor ventilation aiding transmission of infectious diseases
	Chemical pollutants	Indoor air pollution from fires, stoves or heaters. Accidental poisoning from household chemicals. Occupational exposure for home workers
	Physical hazards	Household accidents – burns and scalds, cuts, falls. Physical hazards from home-based economic activities. In poor quality housing, inadequate protection from rain, extreme temperatures
Neighbourhood	Biological pathogens	Pathogens in waste water, solid waste (if not removed from the site), and local water bodies. Disease vectors often breeding in standing water, drains or garbage. If sanitation is inadequate, many people will defecate on open sites, bringing faecal contamination. If a settlement is served by communal standpipes, latrines and/or solid-waste collection points, these need intensive maintenance to keep them clean and functioning well
	Chemical pollutants	Ambient air pollution from fires or stoves; also often from burning garbage if there is no regular garbage collection service. Air and water pollution and wastes from cottage industries and from motor vehicles
	Physical hazards	Site-related hazards, e.g., housing on slopes with risks of landslides; sites regularly flooded, sites at risk from earthquakes. Traffic hazards. Noise
Workplace	Biological pathogens	Overcrowding/poor ventilation aiding transmission of infectious diseases
	Chemical pollutants	Toxic chemicals, dust
	Physical hazards	Dangerous machinery, noise
City (or municipality within larger city)	Biological pathogens	Quality and extent of provision for piped water, sanitation, drainage, solid waste collection, disease control and health care at city or municipal level a critical influence on extent of the problems
	Chemical pollutants	Ambient air pollution (mostly from industry and motor vehicles; motor vehicles' role generally growing); water pollution; hazardous wastes
	Physical hazards	Traffic hazards. Natural disasters and their unnaturally large impact because of inadequate attention to prevention and mitigation
	Citizen's access to land for housing	Important influence on housing quality directly and indirectly (e.g., through insecure tenure discouraging households investing in improved housing and discouraging water, electricity and other utilities from serving them)
	Heat island effect and thermal inversions	Raised temperatures a health risk, especially for susceptible groups (e.g., elderly, very young). Air pollutants may become trapped, increasing their concentration and the length of people's exposure to them
City-region (or city periphery)	Resource degradation	Soil erosion from poor watershed management or land development or clearance; deforestation; water pollution; ecological damage from acid precipitation and ozone plumes
	Land or water pollution from waste dumping	Pollution of land from dumping of conventional household, industrial and commercial solid wastes and toxic/hazardous wastes. Leaching of toxic chemicals from waste dumps into water. Contaminated industrial sites. Pollution of surface water and possibly groundwater from sewage and storm/surface runoff
	Pre-emption or loss of resources	Freshwater for city pre-empting its use for agriculture; expansion of paved area over good quality agricultural land
Links between city and global issues	Non-renewable resource use	Use of fossil fuels and other mineral resources; loss of biodiversity; loss of non-renewable resources in urban waste streams
	Non-renewable sink use	Persistent chemicals in urban waste streams; greenhouse gas emissions, stratospheric ozone depleting chemicals
	Overuse of finite renewable resources	Scale of consumption that is incompatible with global limits for soil, forests, freshwater

on the extent to which environmental hazards can be avoided; others, such as health care services and emergency services, are a major influence on the extent to which a person can cope with any illness or injury. In general, low income groups are more vulnerable because they are the least able to afford the homes that protect against environmental hazards, treatment when sick or injured and time off work to recover.

Those groups within the population which face discrimination in obtaining access to environmental services (for instance particular ethnic groups or immigrant groups) are also vulnerable. Women are generally more vulnerable than men to many environmental hazards because of gender roles (i.e., as a result of the particular social and economic roles that women have). Women are generally more severely affected than men by poor and over crowded housing and by inadequate health care because they take most responsibility for looking after children, caring for sick family members and managing the household. Those who are responsible for water collection and its use for laundry, cooking and domestic hygiene suffer most if supplies are contaminated and difficult to obtain; these people are generally women or girls. In households with high levels of indoor air pollution, it is generally women (or girls) who inhale larger concentrations of pollutants over longer periods (WHO, 1992).

The greater exposure that women face to many environmental hazards is often linked to the difficulties they face in finding better quality housing. This in turn is linked to the discrimination they face in obtaining employment, purchasing or renting housing, and obtaining credit.

Certain low income groups are particularly vulnerable to environmental hazards because they live in particularly hazardous areas or particularly dangerous housing because these fulfill other more pressing needs, such as being cheap and well located for income earning opportunities. Examples include centrally located tenements, cheap boarding and rooming houses and illegal or informal settlements located on dangerous sites (for instance slopes at risk from landslides or flood plains).

Certain groups within the labour force are more vulnerable to environmental hazards because of high levels of risk within their occupational environment.

REFERENCES

Bartlett, S, Hart, R, Satterthwaite, D, de la Barra, X, and Missair, A (1999) *Cities for Children: Children's Rights, Poverty and Urban Management*, Earthscan, London.

Bradley, D, Stephens, C, Cairncross, S, and Harpham, T (1991) *A Review of Environmental Health Impacts in Developing Country Cities*, Urban Management Program Discussion Paper No. 6, The World Bank, UNDP and UNCHS (Habitat), Washington, DC.

Cairncross, S, Hardoy, J E, and Satterthwaite, D (1990) The Urban Context, in *The Poor Die Young: Housing and Health in Third World Cities*, eds J E Hardoy, S Cairncross, and D Satterthwaite, Earthscan Publications, London, 1 24.

Ekblad, S (1993) Stressful Environments and their Effects on Quality of Life in Third World Cities, *Environ. and Urban.*, 5(2), October, 125 134.

Foster, S O (1990) Measles, the Ultimate Challenge in Urban Immunization, in *Universal Child Immunization - Reaching the Urban Poor*, Urban Examples 16, UNICEF, New York.

Hardoy, J E, Mitlin, D, and Satterthwaite, D (2001) *Environmental Problems in an Urbanizing World*, Earthscan Publications, London.

Harpham, T and Blue, I (1995) *Urbanization and Mental Health in Developing Countries*, Avebury, Aldershot, UK.

IDS (1997) Special Issue on Urban Poverty, *IDS Bull.*, 28(2), April.

Kanji, N (1995) Gender, Poverty and Structural Adjustment in Harare, Zimbabwe, *Environ. and Urban.*, 7(1), 37 55.

McGranahan, G (1991) *Environmental Problems and the Urban Household in Third World Countries*, The Stockholm Environment Institute, Stockholm.

Minujin, A (1995) Squeezed: the Middle Class in Latin America, *Environ. and Urban.*, 7(2), 153 165.

Pryer, J (1993) The Impact of Adult Ill-health on Household Income and Nutrition in Khulna, Bangladesh, *Environ. Urban.*, 5(2), 35 49.

UNCHS (1995) *Human Settlement Interventions Addressing Crowding and Health Issues*, United Nations Centre for Human Settlements (Habitat), Nairobi.

UNCHS (Habitat) (1996) *An Urbanizing World: Global Report on Human Settlements, 1996*, Oxford University Press, Oxford.

WHO (1979) *Agreed Terms on Health Effects Evaluation and Risk and Hazard Assessment of Environmental Agents*, Internal Report of a Working Group, EHE/EHC/79.19, World Health Organization, Geneva.

WHO (1989) *Glossary of Terms on Chemical Safety for Use in IPCS Publications*, WHO/ICS/89.27, World Health Organization, Geneva.

WHO (1992) *Our Planet, Our Health*, Report of the WHO Commission on Health and Environment, World Health Organization, Geneva.

WHO (1995a) *The World Health Report 1995: Bridging the Gaps*, World Health Organization, Geneva.

WHO (1995b) *Community Water Supply and Sanitation: Needs, Challenges and Health Objectives*, A48/INF.DOC./2, 28 April 1995.

WHO (1999) Creating Healthy Cities in the 21st Century, in *The Earthscan Reader on Sustainable Cities*, Chapter 6 ed D Satterthwaite, Earthscan Publications, London.

Wratten, E (1995) Conceptualizing Urban Poverty, *Environ. and Urban.*, 7(1), 11 36.

FURTHER READING

Hardoy, J E, Mitlin, D, and Satterthwaite, D (2001) *Environmental Problems in an Urbanizing World*, Earthscan Publications, London.

Urban Sulfurous and Photochemical Smog

Derek M Elsom
Oxford Brookes University, Oxford, UK

In trying to satisfy the demands of a growing population for energy, food, consumer products, water, waste disposal, housing, services and transport, vast amounts of air pollutants are being added to the atmosphere. Air pollution has become a severe problem in many cities in recent decades, particularly those experiencing rapid industrialization such as Bangkok, Beijing, Cairo, Calcutta, Delhi, Hong Kong, Mexico City, Manila, Santiago, São Paulo, Seoul, Singapore and Taipei (UNEP/WHO, 1992, 1994).

Serious air pollution problems also continue in industrialized cities such as Athens, London, Los Angeles, New York City, Paris and Tokyo. Although pollutant emissions from industrial and residential sources contribute signi cantly to the pollution problems, it is emissions from the world's 800 million motor vehicles which are giving rise to the worst urban air pollution problems.

Air pollutants can adversely affect the health of people. They can produce immediate (acute) symptoms in people, which may be reversible, as well as longer-term (chronic) effects, which may be moderated but not readily cured. In some cases pollution can result in premature death. The health of as many as 1.6 billion people living in urban areas throughout the world may be at risk from poor air quality (Elsom, 1996). Pollutants can acidify soils and freshwater, harm plants and animals, erode buildings and historic artifacts, reduce visibility, and affect local and global climate.

Actions are being taken to reduce pollutant emissions and improve air quality in urban areas but, in some cities, new sources of pollution are being added at a rate which is straining the ability of authorities to control them. Attempts to reduce emissions include tting pollution control equipment to industrial stacks and motor vehicles, changing to cleaner fuels, increasing the combustion ef ciency of furnaces, boilers and vehicles, implementing land-use plans which reduce the distances traveled by motor vehicles, improving public transport systems and placing restrictions on the use of cars. These measures may take many years to become fully effective and so, in the meantime, some cities have introduced emergency measures to reduce emissions from motor vehicles, industry and even homes during periods of very poor air quality (smog).

TYPES OF POLLUTANTS AND THEIR SOURCES

The type of pollutants emitted into the atmosphere depends upon the nature of the polluting activity, especially the type of fuel being burned and, for some activities, the raw materials being processed. Typically boilers and furnaces of manufacturing industries and electricity generating stations burn coal, lignite (brown coal) or heavy fuel oil and so emit sulfur dioxide, nitrogen oxides and particulates. If natural gas is used, then the emissions are predominantly nitrogen oxides. Industrial plants and waste incinerators may release toxic chemicals (chlorine, ammonia, fluoride) and metals (e.g., lead, cadmium).

Petrol engine motor vehicles produce a range of pollutants including carbon dioxide, carbon monoxide (indicative of incomplete combustion), nitrogen oxides (about 90% as nitric oxide and 10% as nitrogen dioxide) and volatile organic compounds (VOCs) such as benzene, 1,3 butadiene and other hydrocarbons. Diesel engine vehicles are more fuel efficient than petrol engine vehicles and emit lower amounts of benzene, carbon dioxide and carbon monoxide. However they emit sulfur dioxide and fine particulates. The latter are less than 10 μm (PM_{10}) or even smaller ($PM_{2\cdot5}$) and so are tiny enough to penetrate deep into the lungs. Research suggests that an increase of 10 $\mu g\,m^{\cdot 3}$ in PM_{10} concentrations in urban areas is associated with a daily rise in the death rate by 1% such that PM_{10} particles may be responsible for 60 000 premature deaths a year in the US and as many as 10 000 in the UK, mainly among the elderly and those with pneumonia and chronic lung disease (*see* Use of Epidemiological Data to Develop Air Quality Standards for Short-term and Long-term Environmental Health Bene ts, Volume 4).

Nitrogen oxides and VOCs, emitted mainly from motor vehicles, electricity generating stations and industrial plants, react with sunlight to create photochemical oxidants, notably ozone. Ozone levels show a strong diurnal variation. During the early morning, emissions of nitric oxide and VOCs increase as traffic flows increase. As sunlight increases, nitric oxide rapidly converts to nitrogen dioxide. The decrease of nitric oxide and the increase of nitrogen dioxide is crucial to ozone formation since nitric oxide destroys ozone. Around 09:00 h, VOCs peak and the VOCs, nitrogen dioxide and sunlight begin to work together to create ozone. The chemistry is aided by carbon monoxide, which acts as a catalyst. As ozone levels increase, VOCs and nitrogen dioxide decrease. Ozone formation ceases with the exhaustion of nitrogen dioxide. Ozone peaks around midday and early afternoon. Depending on the city and the atmospheric situation, ozone levels may remain high during the afternoon, drifting away from the city, and only beginning to diminish after dark (Bridgman, 1997). Ozone concentrations tend to be higher in the suburbs and over rural areas downwind of large

conurbations as it takes time for this secondary pollutant to form and because emissions of nitric oxide from traffic deplete ozone in traffic-congested city centers. Ozone and other photochemical oxidants such as peroxyacetyl nitrate (PAN) cause eye and nose irritation, respiratory illness and damage to materials (e.g., rubber, paint) and some plants.

MOTOR VEHICLE SOURCES

In urban areas, road transport is becoming the dominant source of many of the key pollutants affecting the health of urban residents. In recent decades there has been a phenomenal surge in the number and usage of motor vehicles, which reached a world total of 800 million in 1998. For example, the number of vehicles in Sao Paulo, Brazil is increasing at a rate of nearly 300 000 vehicles a year, having already reached a total of 5.7 million.

Emissions from individual vehicles depend on the constituents of the fuel, the design and size of the engine, the external conditions in which it is being driven (road surface, air temperature), the speed (and whether the car is accelerating/decelerating or under stress due to load or road gradient), its age and level of maintenance. Emissions from motor vehicles can be reduced by improvements in fuel and vehicle technologies and by transport and traffic management measures. Economic instruments (e.g., fines, charges, taxes and subsidies) can be used to increase the effectiveness, shorten the time taken to become effective, or help fund all or part of the cost of such measures.

Reformulated fuels can reduce emissions of certain pollutants by changing the fuel constituents. The introduction of unleaded petrol and the reduction of the lead content of leaded petrol has resulted in a marked lowering of airborne lead concentrations in many countries. In 1997, around 80% of all petrol sold worldwide was unleaded. Low-sulfur diesel fuel has helped reduce emissions of sulfur dioxide and fine particulates. Alternative vehicle fuels such as compressed natural gas (CNG) offer significant emission reductions in many pollutants. Electric battery-powered vehicles offer the potential for zero emissions in cities (but they produce pollution at the power stations where the electricity is generated to charge the batteries) as do vehicles powered by hydrogen fuel cells. Vehicle-targeted measures include governments setting emission limits for new and used vehicles, which are then tightened periodically. This can lead to marked improvements in engine design and a requirement for the fitting of pollution control equipment (e.g., catalytic converters and diesel particulate traps). Regular vehicle inspection and maintenance programs help encourage motorists to comply with the standards. Random roadside inspections are useful in identifying the worst polluting vehicles (e.g., poorly maintained) which make a disproportionate contribution to total vehicle emissions.

Transport and traffic management measures aim to reduce the number of private vehicles, their frequency of use, and the distances they are driven. In particular they attempt to encourage a modal switch from private cars to public transport, walking and cycling. Some traffic management measures are designed to create situations where vehicle engines operate efficiently, consuming less fuel and keeping pollutant emissions as low as possible. Measures include vehicle access restrictions, parking controls, traffic calming and smoothing, urban traffic control, park-and-ride facilities, road pricing, parking controls and priority lanes for public transport and high occupancy vehicles.

The problem of vehicle emissions may be made worse in developing countries because of the use of low quality fuels, a dominance of old cars characterized by poor fuel efficiency, poor vehicle maintenance and overloading of vehicles, leading to frequent breakdowns and traffic congestion, limited use of vehicle emission controls (e.g., catalytic converters), weak emission standards and lax enforcement, limited off-street parking facilities leading to on-street parking and congestion, and an inadequate public transport system (Onursal and Gautam, 1997; Wijetilleke and Karunaratne, 1995). The situation is often made worse by narrow canyon-like streets in the city center not designed for high traffic flows. Growth of cities may be largely uncontrolled as squatter settlements and shanty towns proliferate and the urban area expands. In many cities it is difficult to implement land use policies that locate new residential and commercial developments close to existing public transport facilities. Consequently there is a large increase in the number and length of journeys made by private cars, worsening congestion and increasing vehicle emissions. Curitiba (2.5 million population) in southeast Brazil has avoided the worst of these transport problems by ensuring urban growth takes place along radial axes and providing a dense network of prioritized express bus services along the axes and the concentric roads linking them. Three-quarters of all commuters use this service (Rabinovitch and Leitman, 1996).

INDUSTRIAL SOURCES

Emissions from electricity-generating stations and industrial plants can cause serious problems locally, as well as contribute to acid deposition farther afield. Measures to reduce emissions from these sources include increasing efficiency in energy production and consumption, using cleaner fuels and ensuring more widespread installation of effective pollution control equipment such as electrostatic precipitators (which use electrically charged wires to attract dust particles) to remove particulates, and flue gas desulfurization systems (scrubbers) to reduce sulfur dioxide emissions. Most industrialized countries have implemented legislation which specifies emission and technological standards. This

is enforced by requiring businesses to obtain a license before beginning operation, which has to be renewed annually subject to confirmation that the legislative requirements are being fulfilled. The stringency of pollution control in Europe and the United States contrasts markedly with that of many developing countries. In cities in some rapidly industrializing countries, industries processing and manufacturing chemicals, petroleum, metal and paper as well as electricity-generating stations lack even basic pollution controls. Extensive construction activities and unpaved roads give rise to a high particulate loading of the urban atmosphere. The authorities may well be aware of these problems and wish to improve air quality but simply lack the resources to do so effectively. Financial support from international bodies (e.g., the World Bank) will be needed.

RESIDENTIAL EMISSIONS

Much of the world s population relies on poor quality fuels such as biomass (*see* Biomass Use for Urban Fuels in Developing Countries, Volume 3), kerosene, low-grade coals and lignite for cooking and heating purposes. These can create high levels of ambient (outdoor) particulate and sulfur dioxide concentrations as well as poor indoor air quality adding to the pollution caused by other sources such as tobacco smoking. The shanty and squatter settlements of expanding cities in developing countries pose particular problems. People living in such poverty have no choice but to use cheap and polluting fuels. Air quality improvements can be achieved by increasing the efficiency of cooking stoves and heating appliances and shifting to fuels such as anthracite, oil, electricity and natural gas but this requires significant financial investment. Measures to reduce emissions from new residential areas, notably in parts of Europe, include installation of district heating systems using waste heat from nearby power stations. Ultimately, improved building insulation, energy conservation measures and more energy-efficient household appliances can reduce energy consumption and consequently reduce pollutant emissions.

AIR QUALITY STANDARDS

Governments of most countries as well as the World Health Organization have set health-based environmental air quality standards; threshold pollutant concentrations above which adverse effects are believed to occur (Table 1). Health-based standards (criteria) are derived from epidemiological studies of community groups and from laboratory toxicological experiments using human volunteers or animals. They are set to protect the health of the population, especially sensitive groups such as those with preexisting heart and lung diseases, the elderly, young children and pregnant women and their unborn babies. Such groups may form one-fifth of a national population. To such groups at risk may be added people taking vigorous physical exercise (e.g., participants in outdoor sports, cyclists and joggers).

Table 1 World Health Organization health-based guideline values for selected pollutants

Pollutant	Guideline value	Averaging time
Carbon monoxide	100 mg m^{-3} (86 ppm)	15 min
	60 mg m^{-3} (52 ppm)	30 min
	30 mg m^{-3} (26 ppm)	1 hour
	10 mg m^{-3} (9 ppm)	8 hour
Ozone	120 μg m^{-3} (60 ppb)	8 hour
Nitrogen dioxide	200 μg m^{-3} (105 ppb)	1 hour
	40 μg m^{-3} (21 ppb)	Annual
Sulfur dioxide	500 μg m^{-3} (188 ppb)	10 min
	125 μg m^{-3} (47 ppb)	24 hour
	50 μg m^{-3} (19 ppb)	Annual

SUMMER AND WINTER SMOG

A period of a few hours or a few days during which the concentrations of one or more pollutants greatly exceed health-based standards is called a *pollution episode* or *smog* (derived from the words SMoke and fOG). It occurs during stagnant meteorological conditions when atmospheric processes provide only limited dispersal and dilution of the pollutant emissions. A short-term increase in emissions can add to the problem such as from traffic emissions during rush hours or heating of homes during a cold spell. An anticyclone (high air pressure synoptic system) is usually present, characterized by light winds and the capping of the lower atmosphere by an intense temperature inversion which limits vertical mixing. Some cities such as Athens, Los Angeles and Santiago experience a subtropical (Mediterranean) type of climate, characterized by stagnant weather conditions for an entire season which creates the potential for almost continuous high pollutant concentrations lasting several months (Elsom, 1996; Turco, 1997).

Typically, smog occurs either during warm and sunny stable weather or during cold and foggy stable conditions. These are sometimes described as summer and winter smog types, respectively. However, such descriptions are more appropriate to mid- and high-latitude climates than to the tropics, where it is the dry season that triggers the smog. It should be stressed that although stable weather conditions result in pollution concentrations rising above health-based air quality standards, smog would not occur if pollutant emissions were not high to begin with. In other words, human activities are the fundamental cause of smog rather than the weather conditions. Topography and altitude can worsen smog. When a city lies in a basin or valley, the

slopes act like the sides of a box and a temperature inversion may form over the lid of the box, so trapping the pollutants. Oxygen levels are reduced at high altitude so combustion is less efficient in cities such as Mexico City (2250 m) while photochemical activity may be enhanced. Land and sea breezes can play a part in aggravating coastal city pollution problems such as in Athens, Los Angeles, Melbourne and Sydney. During the night an offshore land breeze sweeps the pollutants out to sea, but then an onshore sea breeze brings them back the next day.

URBAN AIR QUALITY MANAGEMENT

The ultimate aim of urban air quality management is to keep air pollution concentrations below levels (*air quality standards*) that pose a threat to people or the environment and, in areas where such levels are exceeded already, to implement a plan to return pollution concentrations to below air quality standards as soon as is practicable (Elsom, 1996; Longhurst *et al.*, 2000).

Air quality management first requires an assessment of whether pollutant concentrations exceed air quality standards, or may do so in the future, as well as an understanding of the nature and relative importance of the various pollutant sources. This involves monitoring current levels of air pollution, identifying and compiling an inventory of sources of the pollutants, using air quality models to calculate pollution levels where monitoring results are not available, and using the models to predict changes in concentrations over time, both short-term (hours, days) and long-term (years, decades). It is then necessary to develop, test and implement a plan to ensure that a healthy and environmentally acceptable air quality is achieved and then sustained in the future. That plan will need continually updating and modifying as the nature and sources of air pollution problems change (Box 1).

SMOG ALERT SYSTEMS

Reducing pollutant emissions to a minimal level will eliminate smog eventually. However, smog poses a serious risk to public health and because it may take many years before improved technologies and pollution control legislation take effect, short-term emergency emission reduction measures

Box 1 London: from sulfur smog to photochemical smog

From the mid-19th century through to the early 1960s London and other UK cities such as Glasgow, Manchester and Sheffield experienced frequent winter smog, commonly called peasoupers, during which levels of particulates (black smoke) and sulfur dioxide were so high that pollution caused hundreds and sometimes thousands of premature deaths amongst the elderly, infants and the sick (e.g., nearly 4000 excess deaths were attributed to the December 1952 smog in London) and left a legacy of chronic bronchitis sufferers amongst residents. The main sources contributing to these winter smogs were emissions from household open fires, industrial plants, power plants and railway steam engines. Inefficient combustion technology, the reliance on coal as the dominant fuel and relatively weak legislation all helped to ensure Britain suffered many decades of winter smog.

Following the notorious 1952 smog, the situation began to change and the frequency and severity of smog were reduced as emissions of pollutants were reduced dramatically. This was achieved through improvements in technology (especially with respect to improving fuel combustion efficiency), the availability of cleaner fuels (oil, natural gas and electricity replaced coal increasingly), urban redevelopment (much terraced housing with inefficient heating systems was replaced by high-rise dwellings with efficient systems), power station replacement (electricity became generated nationally by a smaller number of larger, efficient power stations located outside urban areas), the replacement of steam locomotives by diesel and electric powered engines, social changes (growing affluence resulting in a desire for more efficient and convenient home heating systems to replace open coal fires), public and media pressure for a cleaner environment, and landmark legislation in the form of the Clean Air Acts of 1956 and 1968. The Clean Air Act of 1956 enabled local authorities to introduce smoke control areas, requiring heating and cooking appliances in homes to run on smokeless fuels (some of which, such as natural gas, were sulfur free too). Minimum heights for industrial stacks were set and improved industrial pollution control was specified.

The marked reduction of suspended particulates, which had contributed to fog formation and persistence, resulted in local climate change. The frequency of dense fogs decreased and the number of hours of bright sunshine in November in central London nearly doubled from an average of 38 hours during 1921–1950 to reach 70 hours on average during 1961–1990. For comparison sunshine hours in the surrounding countryside over the same period increased by just eight hours.

Although London no longer experiences sulphurous smog, these episodes have now been replaced by smog caused by motor vehicle pollutants. This was evident during the London episode of 12–15 December 1991 when a period of cold, stable and foggy weather resulted in one-hour nitrogen dioxide levels peaking at 423 ppb (809 $\mu g\, m^{-3}$) while levels of fine particulates (PM_{10}) and benzene increased markedly too. Smog in and around London is no longer restricted to the winter and has become a year-round problem. As the cocktail of urban generated pollutants drift away from the cities, photochemical reactions on warm sunny days create ozone so spreading the smog problem across the country. European pollutants ignore national boundaries so Dutch, French and German emissions add to the ozone and PM_{10} pollution problems of Southeast England.

Table 2 Air pollution episode alert levels in Ile de France (Paris)[a]

Level	Sulfur dioxide	Nitrogen dioxide	Ozone
1 Information only	200(75)	200(105)	130(65)
2 Information and measures	350(130)	300(155)	180(90)
3 Alert and stringent measures	600(225)	400(210)	360(180)

[a] Pollution values are measured as hourly means in $\mu g\,m^{-3}$ (ppb) and at least two monitoring stations must register exceedances for declaration of a level.

may be implemented in the meantime. Most episodic management or smog alert schemes implement traffic restraint measures and place curbs on industry but also make some attempt to tackle residential emission sources. The measures are intended to reduce the severity and duration of smog. The ultimate objective would be to implement measures in advance of, and during, a period of stagnant meteorological conditions in order to prevent pollution concentrations ever reaching unacceptable levels.

Two or sometimes three pollutant threshold concentrations (smog alert levels) are commonly employed to alert the public to periods when air quality may pose an increasing risk to their health (Table 2). The first stage usually involves the issuing of a health advisory warning, especially to sensitive groups within the community. Advice is given on what actions they can take to minimize risks to their health, such as to stay indoors and avoid vigorous exercise. Requests are made for the public to take voluntary action to reduce emissions such as asking them not to undertake inessential car journeys. Subsequent smog alert stages introduce mandatory pollution control measures.

MEASURES TO TACKLE SMOG

Several major cities (Athens, Mexico City, Milan, Paris, Rome, Tehran) have introduced a scheme banning the use of half the cars in the city center during smog, the restriction being determined by the license-plate number, cars with odd-numbered plates being banned on odd-numbered dates (Box 2). Banning vehicles according to their level of emissions may be more effective but would require vehicles to display exemption permits. An experiment at Heilbronn, Germany, was conducted in June 1994 when it was closed for four days to all vehicles, except those fitted with three-way catalytic converters and emission diesel engines meeting the most stringent limits, as an experiment to measure the air quality impacts. The result was a 40% reduction in road traffic and a 50% increase in public transport use resulting in nitrogen dioxide concentrations falling by 40% and benzene concentrations being halved. Moving traffic is less polluting than congested stationary traffic (at least below about $88\,km\,h^{-1}$), so during gridlock conditions Madrid introduces short-term bans on parking along the main thoroughfares. In Mexico City, road repairs are halted to minimize traffic congestion, the use of government vehicles are halved, and schools are closed so there is no need to use vehicles to take children to school.

Some cities are planning to link their traffic light coordination system with pollution monitoring networks. This would enable an area within the city experiencing, or forecast to experience, poor air quality to be protected from further traffic emissions by redirecting vehicles along alternative routes and/or discouraging motorists from entering that area by charging them for entry. Effective

Box 2 Paris smog restrictions

Traffic restrictions were introduced in Paris on 1 October 1997 after the previous day's pollution levels reached smog alert level 3 (NO_2 one-hour threshold of $400\,\mu g\,m^{-3}$) for the first time since legislation empowering the authorities to restrict vehicle movement came into force. Around three million vehicles enter or leave the city each day and 1.1 million use the ring road around Paris. On this date only vehicles with an odd-numbered license plate were allowed access into the city. Public transport was free (being subsidized by the government), speed limits were reduced (from 80 to $60\,km\,h^{-1}$) on the Paris ring road, residential parking was free to try to persuade commuters to leave their cars behind, only priority vehicles were used by Parisian authorities, sports were stopped in schools, tourist buses were banned and industrial activities had to decrease emissions (the largest of three power stations near Paris ceased operation). Vehicles with three or more passengers, delivery vans and emergency vehicles were allowed to enter the city as well as electric and gas powered vehicles. In the future, exemptions (green stickers) will be given to less polluting vehicles, such as those fitted with a three-way catalytic converter. Around 1000 police officers were deployed to enforce the restrictions but, as this was the first time the alert procedure had been implemented, drivers breaking the ban were not fined Fr 900 ($150) but were asked to park their cars and walk or use public transport. The number of passengers on buses increased by 15% and on the Metro and trains by 10%. Traffic levels in central Paris fell by 17%, on the ring road by 6%, and around Paris by 14%. Emissions of nitric oxide and carbon monoxide were reduced and nitrogen dioxide levels fell and remained below alert level 1 (one-hour of $200\,\mu g\,m^{-3}$).

ways will be required to communicate to car drivers what actions they should take and to inform them of any charge that will be made. Communication may be via roadside variable-message signs and in-vehicle units (containing a smart card for any charge to be deducted). Schemes exploring this approach are being developed in Athens, Bangkok, Bristol, Helsinki, Gothenburg, Kanagawa and Leicester, while Singapore already has a sophisticated road pricing system. Its electronic scheme began in late 1998. Cars are charged $2 per pass during the congested 08:00 09:00 h period but this drops to $1 during the half an hour before and after this period. This is intended to smooth the peak morning traffic flow and reduce peak hourly pollutant concentrations but the charges could be raised and extended to other times of the day during smog to discourage even more motorists.

Emissions from residential areas, businesses and public buildings can be decreased during smog by urging people to reduce the level of heating during winter and air conditioning in summer as well as to reduce energy consumption in general. Use of garden incinerators and the lighting of bonfires can be curtailed. During summer smog episodes, the use of barbecues may be discouraged as well as that of certain household and decorating products (e.g., solvent-based paints) which release VOCs.

Industry may be required to reduce emissions by 30%, by half or even cease production completely. Exceptions may apply to industrial activities providing essential community services such as electricity generating stations unless a national grid system of electricity generation allows local stations to close. Reductions in industrial emissions can be achieved by switching to, say, low-sulfur fuels, which have been stocked in readiness for such an occasion. During a photochemical smog, closure of businesses emitting large amounts of VOCs (e.g., petrol stations not fitted with vapor recovery systems on the pumps, paint-spraying shops and dry cleaning stores) may be considered because of their contribution to ozone formation. Pollutants emitted from a very tall stack may penetrate a capping temperature inversion. In such a situation, the emissions may not contribute to the local smog problem if the emissions drift away from the area. On the other hand, some industries with low stacks may be required to cease production altogether or to switch their emissions from shorter stacks to taller ones, if such an option is available. However, if upper winds are light, pollutants emitted by tall stacks above a nocturnal capping inversion may remain there overnight and be brought back down rapidly to the surface the next morning when the sun warms the ground and convection develops (a process called *fumigation*).

The effectiveness of various episodic management measures depends upon many factors. Current smog alert schemes indicate that traffic restraint measures may have a significant effect on carbon monoxide levels within the area in which the restriction is imposed, and nitrogen dioxide levels to a lesser extent. High concentrations of PM_{10} are associated with regional transfer into the city as well as local sources so reductions in traffic and industrial emissions locally can influence only the latter component. Levels of ozone are more likely to be reduced if the restrictions on vehicles and industry apply over a wide regional area, not simply the city center. Reduction of sulfur dioxide levels may be achieved by reducing emissions from coal-fired electricity generating stations and industrial plants with low stacks in the vicinity as well as diesel engine vehicles but, as in many developing countries, where the major cause of high levels of sulfur dioxide is public transport diesel engine buses, then the problem may be harder to resolve.

URBAN AIR POLLUTION AND GLOBAL CLIMATE CHANGE

Ultimately, the solution to the smog problem and poor air quality in general is to reduce pollutant emissions within an urban area to a very low level. As the baseline or average pollution level in an urban area declines, so the severity and frequency of smog ought to decrease (Tables 3 and 4).

Table 3 Number of days exceeding the one-hour federal ozone standard (exceeding 0.12 ppm or 240 $\mu g\, m^{-3}$) and Stage 1 ozone episodes (exceeding 0.20 ppm or 400 $\mu g\, m^{-3}$) in the South Coast Air Basin (includes Los Angeles), 1976–1999[a]

Year	Federal standard	Stage 1 episode
1976	194	102
1977	208	121
1978	187	116
1979	191	120
1980	167	101
1981	180	99
1982	149	63
1983	152	84
1984	173	97
1985	158	83
1986	164	79
1987	160	66
1988	178	77
1989	157	54
1990	130	41
1991	130	47
1992	143	41
1993	124	24
1994	118	23
1995	98	14
1996	90	7
1997	68	1
1998	62	12
1999	41	0

[a] Several decades of pioneering and increasingly stringent control legislation have begun to show the benefits.

Table 4 Number of days exceeding specified levels of the air pollution index (IMECA) for ozone in Mexico City, 1988–1997. The one-hour national health-based standard for ozone is 0.11 ppm (216 μg m^{-3}) which is represented by an IMECA index of 100[a]

Year	• 100	• 150	• 200	• 250	• 300
1988	330	232	69	11	1
1989	331	136	15	3	0
1990	338	235	94	28	4
1991	350	284	162	56	7
1992	329	264	125	39	12
1993	324	240	85	14	1
1994	346	263	100	4	0
1995	325	263	94	7	0
1996	327	235	71	5	0
1997	322	219	55	3	0

[a] Emission-reduction measures have reduced the ozone peaks but not the length of time of exceedances of the standard (typically 11 months each year).

In situations where pollution is imported from outside a city s boundaries, emissions will need to be reduced on a regional, national and even international scale too.

Urban air pollution problems will be affected by global environmental changes. Stratospheric ozone depletion has permitted increased penetration of ultraviolet radiation and this may enhance photochemical activity, resulting in increased ground level ozone concentrations. Global warming, due to the greenhouse effect, may lead to a worsening of pollution problems. For example a region may experience more frequent and intense anticyclones and rising temperatures may increase evaporative emissions of VOCs. Clearly there is a need to monitor air quality changes and to model the likely effects of changing environmental factors that may affect pollutant emissions and ambient concentrations.

REFERENCES

Bridgman, H (1997) Air pollution, in *Applied Climatology: Principles and Practice*, eds R D Thompson and A Perry, Routledge, London, 288 303.

Brimblecombe, P and Bentham, G (1997) The Air That We Breathe: Smogs, Smoke and Health, in *Climates of the British Isles*, eds M Hulme and E Barrow, Routledge, London, 243 261.

Elsom, D M (1996) *Smog Alert: Managing Urban Air Quality*, Earthscan, London.

Longhurst, J W S, Elsom, D M, and Power, H (2000) *Air Quality Management*, WIT Press, Southampton, Boston, MA.

Onursal, B and Gautam, S P (1997) *Vehicular Air Pollution: Experiences from Seven Latin American Urban Centers*, Technical Paper, 373, The World Bank, Washington, DC.

Rabinovitch, J and Leitman, J (1996) Urban Planning in Curitiba, *Sci. Am.*, March, 26 33.

Turco, R P (1997) *Earth Under Siege: From Air Pollution to Global Change*, Oxford University Press, Oxford.

UNEP (United Nations Environment Programme)/WHO (World Health Organization) (1992) *Urban Air Pollution in Megacities of the World*, Blackwell, Oxford.

UNEP (United Nations Environment Programme)/WHO (World Health Organization) (1994) Air Pollution in the World s Megacities, *Environment*, 36, 4 13, 28 37.

Wijetilleke, L and Karunaratne, A R (1995) *Air Quality Management: Considerations for Developing Countries*, World Bank Technical Paper, 278, The World Bank, Washington, DC.

Urban Wastes

Stephen T Washburn and Cheryl A Doyle
ENVIRON Corporation, Princeton, NJ, USA

It has become increasingly dif cult to safely handle the enormous amounts of waste produced by urban societies, as cities throughout the world continue to grow. Over the past 50 years, when the average population of the world's 100 largest cities more than doubled, the demand for waste disposal in urban areas has strained (and sometime overwhelmed) municipal resources. For example, Tokyo has grown so quickly, quadrupling its population since 1950, that municipal authorities have devised ways to build islands of waste in Tokyo Bay (United Nations Publications, 1997). In Bombay, at least two million residents have no sanitary facilities at all, and most sewage that is collected is discharged into local waterways (United Nations Publications, 1997).

The waste disposal challenges facing urban areas vary widely from one region to another, re ecting important

differences between cities in developing and industrialized countries. In poorer, less industrialized cities, concerns regarding waste disposal frequently focus on very local, immediate effects, including disease spread by untreated human waste and uncollected garbage, as well as pollution of surface and groundwater resources.

Richer, industrially developed cities, on the other hand, generally have the resources to process sewage and collect garbage, but higher rates of consumption and lower rates of recycling can create more widespread, even global, environmental effects. For example, according to the World Resources Institute, the resources consumed and greenhouse gases emitted to support cities in developed countries are much greater than those needed to support poorer cities in developing regions, on a per capita basis (EPA, 1992).

COMPONENTS OF THE URBAN WASTE STREAM

By volume, municipal solid waste (MSW) is by far the largest component of the overall urban waste stream. MSW generally consists of household garbage, commercial waste (i.e., waste from businesses, including stores and restaurants), and non-hazardous industrial waste.

The volume of MSW generated per person is generally much higher in wealthy, industrially developed urban areas than in poorer, rural areas. For example, high-income countries generate 300 1000 kg per person each year, while low-income countries generate between 100 and 200 kg per person each year (United Nations Publications, 1997). Wealthy cities also produce a higher percentage of wastes requiring final disposal, largely because the higher-income societies do not have as great a motivation to recover and recycle, and often no tradition of doing so.

In industrially developed countries, MSW consists primarily of food wastes, garden trimmings, paper and paperboard, glass, metals, and plastics; in the US, paper and paperboard represents the largest fraction (about 40% by weight) (Corbitt, 1999). By contrast, MSW in developing countries tends to have a much greater percentage of organic wastes, and much less plastic, paper, and other manufactured goods. For example, the organic content of MSW in Africa may exceed 70%; and plastics, glass and metals constituted less than 10% of MSW in four African cities studied (UNEP, 1999).

Although MSW is often considered non-hazardous, it may also include a variety of potentially toxic substances, such as oil-based paint, paint thinners, cleaning chemicals, used oil, batteries, and pesticides. Estimates of the percentage of such household hazardous waste in MSW range from 0.5 to 5% in industrialized countries. Because of their potential to cause environmental contamination, the impact of disposing of such wastes may be considerably greater than these low percentages may suggest.

Although MSW is produced in the greatest volumes, each of the following types of waste can also pose challenges to urban waste disposal systems:

- *Sewage and sewage sludge*: In the wealthiest countries, effective sewage systems have been developed to safely dispose of human biological waste. In developing countries, however, especially those undergoing rapid urban growth, sewage collection, disposal, and treatment present major problems. Inadequate sewage disposal can increase risk of disease resulting from exposure to pathogens in the waste, and cause ecological damage, particularly to rivers, lakes, and streams. In Shanghai, China, for example, serious water pollution and supply problems have been reportedly caused by the release of about four million $m^3\ day^{\bullet\ 1}$ of untreated human waste into the Huangpu River (United Nations Publications, 1997). Disposal of sludge from sewage treatment plants is also becoming a greater issue, particularly in more industrialized cities. This sludge can contain high concentrations of metals and other potentially toxic substances, and as a result some jurisdictions are placing increasingly stringent restrictions on traditional disposal options such as application to agricultural lands and ocean dumping.
- *Industrial waste (liquid and solid)*: Developing and transition countries, as well as some industrialized countries, have only recently begun to address the environmental effects of industrial waste disposal. Liquid and solid wastes from industrial processes are most frequently disposed of in dumps or simple landfills, although some types may also be burned. In economic terms, the environmental legacy of improper industrial waste disposal is perhaps clearest in the United States, where it has been estimated that approximately $1 trillion will ultimately be spent on the cleanup of former waste disposal sites (EPA, 1992) (*see* Love Canal, Volume 5).
- *Hospital/medical waste*: Hospital and medical wastes typically consist of common waste (office and kitchen waste), pathogenic or infectious waste (including sharps), and hazardous wastes (mainly from laboratories). While many types of common hospital/medical waste can be handled with regular MSW if proper separation can be achieved, infectious and hazardous wastes should receive special handling to prevent the spread of disease and other environmental problems.
- *Construction and demolition (C&D) debris*: C&D waste consists of cement, bricks, asphalt, wood, and other typically inert materials. In industrial countries, C&D waste can represent a significant fraction of the overall waste stream, and may contain asbestos, polychlorinated biphenyls (PCBs) and other potentially toxic substances (*see* Geomorphological Change for Urbanization and Industry, Volume 3).

METHODS OF WASTE DISPOSAL

Land Disposal

Land disposal ranges from open dumping, which may pose a serious risk of pollution and disease (UNEP, 1999), to sanitary landfills, which incorporate engineering features (such as liners and methane gas management) to reduce the potential for releases to the environment. Sanitary landfills are generally more common in developed countries, especially near urban areas where population densities are highest. For example, in Africa and Central America, most of the MSW that is not recycled is sent to open dumps; there are few, if any, sanitary landfills. In Latin America, where land disposal is used for about 60% of the MSW, open dumping is most common, although some large cities have state-of-the-art landfills. In North America, land disposal is used for 65 to 70% of MSW (down from more than 80% in the 1980s), and sanitary landfills are required by legislation in most areas (UNEP, 1999) (*see* Wastes: Land ll and Land Raising, Volume 3).

Local environmental effects of land disposal can include ground and surface water pollution, as well as the spread of disease if waste is not properly covered. At some landfill sites, methane gas can also accumulate and create an explosion hazard. Regional and global environmental effects of land disposal are less obvious, perhaps, but may represent a long-term threat to the global climate. For example, carbon dioxide and methane, which have been identified as greenhouse gases capable of contributing to global warming, are both generated by landfilling. It is likely that carbon dioxide emissions from landfills represent only a minor contributor to global loadings to the atmosphere; for example, fossil fuel combustion accounts for 99% of total carbon dioxide emissions in the US (US Department of State, 1997; US Environmental Protection Agency, 1998). For methane, however, landfills may represent a more significant percentage of total emissions, and landfill liners and covers can increase methane production by producing anaerobic conditions. In the US, which has many covered sanitary landfills fostering anaerobic conditions, approximately 35% of all methane emissions have been attributed to land disposal. In Venezuela, on the other hand, where open dumping under aerobic conditions is common, land disposal reportedly contributes less than 10% of all methane emissions (US Department of State, 1997; US Environmental Protection Agency, 1998).

Composting

Composting is the practice of promoting biodegradation of the organic constituents in solid wastes and sludges, thus reducing the volume and potential hazard of the waste. When compost is used as a soil conditioner, fertilizer, or growth medium, it benefits the environment by reducing the need for artificial fertilizers and by reducing the amount of waste to be landfilled. Although composting results in emissions of carbon dioxide, a greenhouse gas, aeration of compost piles prevents the generation of methane, a considerably more potent greenhouse gas that is produced in large quantities by landfilling, as discussed above (US Environmental Protection Agency, 1998).

In order for composting to be a sustainable component of any MSW management system, it is necessary to have a market for the compost. Many Asian countries have a long history of compost production and use. In northern Europe, centralized composting is a major component of MSW management, but most attempts at centralized composting have been unsuccessful in Latin America and the Caribbean. In Africa, despite the high organic content of the waste stream, composting is a little used waste management practice. In North America, bans on landfilling of garden trimmings in some areas have resulted in an increase in backyard composting, although it has generally been found that compost can be given away but not sold (UNEP, 1999).

Source Reduction, Materials Recovery, Reuse and Recycling

It is often possible to reduce the amount of materials requiring final disposal by simply reducing the amount discarded, or retrieving reusable or recyclable items from the waste stream. In addition to reducing the potential for adverse local effects, studies have shown that source reduction and recycling also represent good ways of reducing greenhouse gas emissions (US Environmental Protection Agency, 1998). Agenda 21, an agreement reached at the United Nations Conference on Environment and Development in Rio de Janeiro in 1992, noted that waste reduction and environmentally sound reuse and recycling should be the primary target in the waste management hierarchy (UNEP, 1999).

While some poorer societies have strong traditions of waste recovery and reuse, it is only relatively recently that most industrialized countries have begun to focus on reducing waste at its source, or recycling used materials instead of discarding them. The primary reason for this difference is economic: poorer societies often reuse and recycle in order to save the cost of new items and conserve resources. In richer societies, on the other hand, there may even be economic disincentives to reuse and recycling; for example, in Europe, almost 15% of the workforce are engaged in the production of packaging (European Commission, 1996). This potential loss of jobs could represent a hurdle to accomplishing source reduction through a switch to reusable packaging. Further, in industrialized countries, packaging is an important component of marketing for many products.

In some developing countries, more than half of the urban waste may be recycled or reused. For example, Bangalore,

a large Indian city, produces 3000 tonnes of waste a day; 2700 tonnes of which is recycled or reused, leaving only 300 tonnes for disposal (United Nations Publications, 1997). Countries such as India, Vietnam and China reuse most of their municipal wastes; in South and West Asia, food wastes are sold to farmers for use as compost, with some large hotels even auctioning them off (UNEP, 1999).

In the industrialized countries, source reduction and recycling efforts typically increase only when the cost of landfilling or other forms of disposal becomes excessive (for example, as a result of more stringent environmental controls), or if natural resources are significantly depleted. In the past several decades, however, some industrialized countries have begun to reduce the amount of urban waste requiring final disposal. For example, in North America, approximately 20 25% of the solid waste stream is currently recycled, compared to 10% recycled in 1975 (UNEP, 1999).

Incineration/Open Burning

Open burning or incineration of MSW can reduce the weight of the waste requiring land disposal by up to 75%, and the volume by up to 90%. While the solid residues of incineration still require disposal, and may be considered hazardous, they can sometimes be used in construction, reducing the amount to be landfilled.

In North America, the use of incineration varies from region to region, due in part to differences in local regulations and in the availability of landfill space. In the more densely populated northeastern US, for example, 40% of solid waste is incinerated, while a total of only about 2% of the MSW is incinerated over the rest of the country. There are fewer incinerators in Canada than in the US, and the province of Ontario has banned construction of any new facilities.

In Europe, incineration is relatively common, in part because of the shortage of land in many areas. Northern European countries rely greatly on incineration: 35% to 80% of residential waste stream in Western Europe is incinerated, with state-of-the-art pollution control and energy recovery. Incinerators in Eastern Europe are often older, with less sophisticated environmental controls.

Incineration is not common in developing countries because the cost is high relative to land disposal, and because wastes in such areas are difficult to incinerate due to high moisture and low energy content. Some former developing countries now incinerate waste, however; for example, Singapore incinerates 90% of its MSW (UNEP, 1999).

Potentially hazardous air emissions are usually a key issue in evaluating waste disposal through incineration. For example, there are concerns regarding the potential human health and ecological effects of polychlorinated dibenzo-p-dioxins and polychlorinated dibenzofurans (dioxins) emitted from MSW and hospital waste incinerators, even with increased use of air pollution controls. The effects of dioxins may not be limited to the immediate vicinity of the source; studies have shown that environmentally persistent chemicals such as dioxins can be transported globally. Waste incineration in industrialized countries has been identified as a major source of dioxins; according to one study, almost 40% of the global total dioxin emissions from identified sources have been attributed to waste incinerators in Japan (UNEP, 1999).

Incineration produces emissions of carbon dioxide and nitrogen oxides; the latter is a greenhouse gas with particularly strong global-warming potential (*see* Global Warming Potential (GWP), Volume 1). However, properly controlled combustion with energy recovery may actually reduce net greenhouse gas emissions by cutting emissions from burning fossil fuels for energy production, as well as emissions from landfills (US Environmental Protection Agency, 1998). Energy produced by incineration is generally converted to either electricity or steam; in the US, 90% of the MSW incinerators with energy recovery produce electricity. Steam generation is more common in Europe, due to a ready market for steam created by a widespread reliance on district heating (UNEP, 1999).

Several innovative incineration technologies have been proposed to reduce toxic emissions. For example, plasma systems operating at extremely high temperatures (e.g., 2000 °C or more) have been developed to provide more complete destruction of waste and minimize emissions of dioxin and other organic compounds. Such systems are more typically used for hazardous waste than for MSW, and have had limited commercial applicability to date.

Surface Water Disposal

Untreated or partially treated sewage is commonly discharged to waterways, leading to the potential spread of disease, particularly if the water is also used as a source of drinking water. Dumping of other types of solid and liquid wastes into rivers, lakes and oceans is also relatively common, since most urban areas are situated near major bodies of surface water. In rapidly developing countries that have limited environmental controls, bacterial contamination of surface water is widespread; in industrialized countries, concerns often focus on discharges of potentially toxic substances in wastewater.

LOOKING FORWARD

The population growth experienced in our cities over the past several decades is not expected to abate in the near future. In fact, according to data published by the United Nations and the World Resources Institute, it is projected that by 2025, the global urban population will be 5 billion

(or about 60% of the world population), compared to 2.5 billion in 1995 (United Nations Publications, 1997; World Resources Institute, 1997). This continuing population growth in the cities of the world will create increasingly difficult waste management and disposal challenges. These challenges will require a combination of local, regional, and global initiatives to reduce the amount of waste generated, and better manage the waste that cannot be eliminated.

A number of international bodies, including the United Nations Environment Programme (UNEP) are actively involved in supporting worldwide waste management efforts. Many cities and countries, whether they have a long history of industrialization or are experiencing recent and rapid development, are also recognizing the need for integrated planning and policy making. While some issues, such as reducing greenhouse gases, may be more global in nature, many decisions must be made regionally or locally.

In order for waste management efforts to be successful, they must be sustainable, both culturally and economically. As cities grow, cultural and economic conditions can change dramatically, affecting the sustainability of different waste management strategies. However, it appears clear that virtually all urban areas will need to develop a waste management hierarchy consisting of waste minimization; resource recovery; recycling; treatment (including incineration); and final disposal.

REFERENCES

Corbitt, R A (1999) *Standard Handbook of Environmental Engineering*, McGraw-Hill, New York.

EPA (1992) Study Finds Potential Hazardous Waste Cleanup Costs May Top $1.5-Trillion, *Inside EPA*, January 3, 1992.

European Commission (1996) *European Sustainable Cities: Report of the Expert Group on the Urban Environment*, March, Brussels, accessed at http://europa.eu.int/comm/dg11/environment/urban/home_en.htm#report.

UNEP United Nations Environment Programme Chemicals (1998) *Dioxin and Furan Inventories: National and Regional Emissions of PCDD/PCDF*, Geneva, Switzerland.

UNEP United Nations Environment Programme International Environmental Technology Center (1999) *International Source Book on Environmentally Sound Technologies (ESTs) for Municipal Solid Waste Management (MSWM)*, accessed at http://www.unep.or.jp/ietc/ESTdir/pub/MSW.

United Nations Publications (1997) *Cities of Today, Cities of Tomorrow*, accessed at http://www.un.org/Pubs/CyberSchoolBus/special/habitat/units/un01pg01.htm and subsequent pages.

United Nations Publications (1997) *Cities of Today, Cities of Tomorrow, City Pro les*, accessed at http://www.un.org/Pubs/CyberSchoolBus/special/habitat/profiles/index.html and subsequent pages.

US Department of State (1997) Inventory of US Greenhouse Gas Emissions, in *The 1997 US Climate Action Report*, Chapter 3, Submitted by the United States of America under the United Nations Framework Convention on Climate Change, accessed at http://www.epa.gov/oppeoee1/globalwarming/publications/emissions/US1997/index97.html.

US Environmental Protection Agency, Office of Solid Waste and Emergency Response (1998) *Greenhouse Gas Emissions from Management of Selected Materials in Municipal Solid Waste*, accessed at http://www.epa.gov/globalwarming/actions/waste/reports.html.

World Resources Institute (1997) *A Guide to the Global Environment: The Urban Environment*, accessed at http://www.wri.org/wri/wr-96 97/.

FURTHER READING

Atmospheric Research and Information Center, *ACE Information Programme, Global Climate Change*, 27. Climate Change and Waste Management, accessed at http://www.doc.mmu.ac.uk/aric/ace/factsheets/gcc27old-pdf.pdf.

IFAT 99, Refuse Disposal a Central Theme, accessed at http://argus.cs.tu-berlin.de/intern/ifat-e.html.

US Country Studies Program (1995) *Interim Report on Climate Change Country Studies*, accessed at http://www.gcrio.org/CSP/IR/title.html.

V

VEI (Volcanic Explosivity Index)

VEI is a logarithmic measure of the explosiveness of volcanic eruptions, which is used for the classification of types of eruption and method of comparison of past and present eruptions. As well as the direct effects of an erupting volcano, there is also a climate impact caused by stratospheric aerosol clouds, which may be produced by powerful eruptions. These clouds contain contaminant sulfuric acid aerosol layers that block out the rays of the sun and produce global cooling. Every year there are about 60 volcanic eruptions that may be quantified. In 1982, Newhall and Self proposed that it was the volume of dust injected into the stratosphere which determined the climatic effectiveness of an eruption and they developed a measure to determine a value of energy release for volcanoes using the volume of material erupted, height of the ash plume and duration of the eruptive phase as well as additional observations. VEI is dependent on the following five parameters.

1. Magnitude: determined from the volcanic ejecta volume.
2. Intensity: determined by the volume of ejecta per unit time calculated from column height and velocity.
3. Dispersive power: determined from column height.
4. Eruption violence: comparable to intensity for spontaneous eruptions.
5. Destructive potential: determined by the devastation extent caused by an eruption.

VEI is based on volcanologically derived data that removes the circularity of argument built into the application of Rdmax and Emax in estimating the value of the dust veil index (DVI) (*see* DVI (Dust Veil Index), Volume 3). In calculating VEI, an open-ended index was assigned to known eruptions on a scale of zero to eight ranging from zero for non-explosive eruptions to eight for the largest explosive eruption in history to characterize and evaluate volcanic eruptions follows (see Table 1).

An eruption with a VEI of zero such as Mauna Loa in Hawaii will have less than 10 000 m^3 of ejecta including lava and ash that will not reach a height above 0.1 km. Durations of eruptions from VEI zero to one will have up to one hour of continuous blasting. With a VEI of four at least 108 m^3 of ejecta is produced from a column height of 10 25 km but will not have any significant climatic effects. A VEI of eight has not been recorded in the last 10 000 years; however, in the past two decades there have been many volcanoes like Mount Pinatubo and El Chichon with high VEIs that have had a significant global cooling effect. A volcano with a VEI of six such as Mount Pinatubo is a highly explosive volcano in comparison to most other volcanoes. Mount Pinatubo located in the Philippines erupted in June of 1991 with a VEI of six. It ejected

Table 1 VEI criteria[a]

	Descriptive	Volume of ejecta (m^3)	Eruption column height (km)
0	Non-explosive	• 104	• 0.1
1	Small	104–106	0.1–1
2	Moderate	106–107	1–5
3	Moderate–large	107–108	3–15
4	Large	108–109	10–25
5	Very large	109–1010	• 25
6	Very large	1010–1011	• 25
7	Very large	1011–1012	• 25
8	Very large	• 1012	• 25

[a] *Source:* http://users.aber.ac.uk/jpg/volcano/lecture2.html.

Table 2 VEI index[a]

VEI	Description	Plume height	Volume	Classification	How often	Example
0	Non-explosive	• 10^2 m	10^3 m^3	Hawaiian	Daily	Kilauea
1	Gentle	10^2–10^3 m	10^4 m^3	Haw/Stromboli	Daily	Stromboli
2	Explosive	1–5 km	1 E106 s m^3	Strom/Vulcan	Weekly	Galeras, 1992
3	Severe	3–15 km	1 E107 s m^3	Vulcanian	Yearly	Ruiz, 1985
4	Cataclysmic	10–25 km	1 E108 s m^3	Vulc/Plinian	Tens of years	Galunggung, 1982
5	Paroxysmal	• 25 km	1 km^3	Plinian	10^2 of years	St Helens, 1981
6	Colossal	25 km	10 km^3	Plin/Ultra-Plin	10^2 of years	Krakatau, 1883
7	Super-colossal	• 25 km	10^2 km^3	Ultra-Plinian	10^3 of years	Tambora, 1815
8	Mega-colossal	• 25 km	10^3 km^3	Ultra-Plinian	10^4 of years	Yellowstone, 2 ma

[a] *Source:* http://volcano.und.nodak.edu/vwdocs/eruption. scale.html.

20 Mt of sulfur dioxide (SO_2) into the stratosphere that produced a measured cooling effect in the tropics of 2 °C and a reported global decrease of 0.5 °C. The global cooling effects reached their peak in 1993, more then two years after the eruption. Similarly, El Chichon in Mexico erupted in April, 1982, with a VEI of five. It ejected 74 Mt of SO_2 into the stratosphere that had a comparable effect that caused a global cooling of 0.2 °C. Studies showed that estimates of magnitude are climatically independent and different latitude locations have no effect.

Table 2 identifies various VEIs defined in terms of plume height and consistency of occurrence.

As documented, the most powerful volcanic eruption in recent history occurred in Indonesia in 1815. The Tambora eruption with a VEI of seven ejected more than 150 km^3 of material into the atmosphere to heights of about 30 km. The expelled debris cut sunlight by approximately 30% and the average global temperature was reduced by 1.5 °C. The year of 1816 is often referred to as the year without a summer.

The VEI can also be used to classify the type of eruption. The five different types of eruptions recognized are: Hawaiian (the smallest), Strombolian, Volcanian, Plinian and Ultra-Plinian (the largest). An example of a Hawaiian eruption, a non-explosive volcanic eruption with a VEI of zero, is Mauna Loa compared to Yellowstone, a mega-colossal eruption with a VEI of eight, that is an Ultra-Plinian.

Additional scales for measuring the size of a volcano are the DVI and the ice-core volcanic index (IVI), which quantifies widespread ice acidity.

REFERENCES

http://users.aber.ac.uk/jpg/volcano/lecture2.html.
http://volcano.und.nodak.edu/vwdocs/vwlessons/kinds/act5.html.

FURTHER READING

http://hpccsun.unl.edu/nebraska/VEI.html.
http://volcano.und.nodak.edu/vwdocs/eruption. scale.html.
http://www.esci.keele.ac.uk/courses/geol317/volcano.html.
http://www.oxford-brookes.ac.uk/geology/8307/volchaz/tsld021.htm.
http://www.sunysuffolk.edu/• mandias/honors/student/volcano/global. warming.html.

MARY KEARNEY *Canada*

Vienna, Urban Metabolism

see Anthropogenic Metabolism and Environmental Legacies (*Opening essay*, **Volume 3**)

Viral Diseases and the Influence of Climate Change

Charles Guest
Australian National University, Canberra, Australia

This article outlines the range of in uences on the occurrence of viral diseases, a group of major and dynamic problems in public health. The nature of selected conditions most probably associated with climate variation, and therefore presumably with long-term climate change, is described. Human viral diseases include dengue fever, encephalitis syndromes borne by mosquitoes and ticks, Rift Valley fever, Ross River virus disease and yellow fever. In animals, increased arboviral disease activity of akabane, bluetongue and bovine ephemeral fever is expected. Prevention and mitigation of these problems, with epidemiological and other research prospects, are considered.

INTRODUCTION

Environmental factors are critical to the occurrence of disease. The idea is not new: links between climate and human health were proposed in Hippocrates treatise *Airs, waters and places*, in the fifth century BC. For the study of medicine, Hippocrates first recommended an understanding of the seasons, with investigation of how the various winds, waters, site, soil and other local environmental characteristics might influence the occurrence of disease. Environmental influences on health continued to receive attention up to the 17th and 18th centuries.

This focus changed in the late 19th century, with the germ theory of disease. The concept that a specific microbe reaches a susceptible host to cause distinctive patterns of morbidity and mortality came to dominate Western medicine. More recently, it has become clear that susceptibility and resistance to disease involve a wide range of factors, including physical environmental phenomena such as climate.

Global environmental change may have profound effects on many forms of life, including the cellular microorganisms and viruses (McMichael *et al.*, 1996). Viruses are not cells, but are completely dependent on cellular hosts for the production of energy. Consequently, with very few exceptions, viral diseases lack sensitivity to antibiotics, and are therefore of great public health concern.

The microbiological agents of vector-borne communicable disease are carried by animals. These diseases include arthropod-borne viral (arboviral) diseases, the primary focus here. First, to illustrate the ecological principles that relate global environmental change and human health, consider climatic influences on viruses in general.

CLIMATIC VARIABILITY AND VIRAL DISEASES

The epidemiological indicators of links between long-term variation in climate and the occurrence of viral disease have a range of time-scales. They come from more immediate (and sometimes venerable) observations on annual changes, through to more speculative, theoretical studies which depend on computer models of the long-term climate change that is now underway (Patz *et al.*, 2000).

Annual Cycles

In temperate and subtropical countries, death rates are highest in winter (Guest *et al.*, 1999). Viral diseases contribute to these rates, but even this relationship is partly obscured by social and behavioural adaptations to cold. Some adaptations prevent an unknown proportion of winter deaths, especially in countries at higher latitudes and altitudes, but more indoor living during colder periods increases the opportunities for transmission of communicable disease.

Interannual Variability

The El Nino/southern oscillation (ENSO) determines, in part, global climate anomalies with altered rainfall and other climatic patterns that are particularly relevant to vector-borne diseases. For example, the incidence of mosquito-borne Rift Valley fever (see in the following sections) has been associated with El Nino-driven flooding in Kenya (Linthicum *et al.*, 1999).

Long-term Climate Change

The frequency of extreme weather events is expected to increase in the coming century. This is more likely to be important to variations in disease patterns than the small changes in the averages of temperature and other climatic variables. It is difficult to predict changing incidence of viral diseases in relation to changing rainfall patterns, as higher temperatures may also increase evaporation. However, one effect of increased flooding and stagnation of water may be greater prevalence of vector-borne diseases. An increase in droughts could increase the incidence of gastrointestinal infections.

VECTOR-BORNE DISEASES

Vector-borne diseases prevail in ecological cycles that include the arthropod vector (mosquitoes, ticks, etc.), with human and other animal hosts. The survival of vectors is sensitive to climatic and other environmental influence, so that the distribution of some of these diseases may expand, while others may contract in future. The geographical range of the vectors may change, but it does not necessarily follow that the incidence of disease will reflect these changes.

For aeons, shifts in insect populations across latitudes may have taken place in response to fluctuating minimum temperatures or other phenomena. At present, many additional factors, such as land clearing, the movement of peoples into bush and jungle areas, and increasing irrigation are influencing the breeding habits of disease vectors and bringing humans into closer proximity with them.

With their short generation time, insects are sensitive biological indicators of terrestrial ecosystems. For example, temperature, humidity and wind may determine whether or not vectors are able to transmit disease in a given location. Rodent populations are another group that both may transmit disease to humans and respond rapidly to environmental change.

Two human diseases that are caused by microorganisms that are more complex than viruses should be mentioned. Malaria, of pre-eminent world importance as a

communicable disease, results from a protozoan parasitic infection of blood cells. Malaria is clearly sensitive to changes in climate, so changes in the territory of this disease, according to altitude and latitude, are expected in future (*see* Malnutrition, Infectious Diseases and Global Environmental Change, Volume 3). Lyme disease (vector-borne, but bacterial) has assumed major public health importance, particularly in the United States in recent years. It is possible that milder winters may increase the range of this disease. Third, consider red tide (discolouration of seawater, caused by toxic dinoflagellates (Halstead, 1982; Humpage *et al.*, 1993). This example of non-human disease suggests that the ecological niche of shellfish is highly sensitive to changing climatic conditions.

The remainder of this article is devoted to the possible climatic influences on arboviral diseases. It is also necessary to appreciate the possible contribution of factors other than climate to the emergence or reemergence of disease (Shope, 1994):

1. The pathogen may have been introduced from afar into an area where transmission could be amplified if vectors and reservoir hosts are present;
2. The vector may have been transported artificially or by natural means into the area;
3. Susceptible hosts may have migrated into the area, e.g., during conflict;
4. Environmental change in an enzootic site (where a disease is usually prevalent in animals) may increase the force of transmission of a pathogen; or
5. The pathogen may have evolved.

ARBOVIRAL DISEASES

Generally, arboviral disease is seasonal, depending on the period during which the arthropod takes its blood meals. Many factors influence the prevalence and season of disease, including climate (rainfall, temperature), vector competence, density of vectors, transport of virus to distant places, level of immunity in the vertebrate populations, the length of *extrinsic incubation period* (the interval between acquisition of the infectious blood meal and the time when the arthropod becomes capable of transmitting the agent), and the length of *intrinsic incubation period* (the interval between invasion of the body by the virus and the appearance of clinical disease in the human host), density of vectors, transport of virus to distant places, and level of immunity in the vertebrate populations.

Dengue

Dengue fever, a mosquito-borne viral disease, has become a widespread tropical urban health problem (Gubler, 1998). Of vector-borne diseases, only malaria affects more people world wide. WHO estimates that 2500 3000 million people in over 100 countries are at risk of both dengue and dengue haemorrhagic fever a serious and often fatal variant of the disease resulting from a second infection by another of the four strains of the virus. Millions of clinical cases occur each year, including up to 500 000 cases of the more severe form of the disease.

Dengue s highest incidence occurs during periods of high rainfall and humidity. Increased temperature may increase dengue transmission in coming decades, but many factors under more direct human control (such as patterns of urbanisation or the effectiveness of mosquito-control programs) may be more critical.

Increases in the latitudinal and altitudinal range of dengue have been projected with climate change. As the survival of *Aedes aegypti* mosquitoes is reduced when the temperature increases to above 42 °C, however, the incidence of dengue may eventually decline in the hottest parts of its current range (Chan *et al.*, 1999).

Increased transmission intensity could lead to a lowering of the average ages of primary and secondary infections (Jetten and Focks, 1997). As infants and adolescents appear most vulnerable to the more dangerous form of dengue, it is possible that with climate change, the morbidity and mortality associated with this disease could increase significantly.

With detailed information on local weather patterns, breeding sites for mosquitoes and the immunity status of the population, it may become possible to predict dengue outbreaks.

Encephalitis

Several viral infections of the brain, frequently fatal, may be influenced by global environmental change. In the Americas, these include Eastern Equine, Western Equine and Venezuelan Equine encephalitis. While their range and prevalence may change, again the importance of environmental changes other than climate should also be considered. Take, for example, the emergence of Eastern Equine encephalitis in Massachusetts. Centuries of farming, with the draining of swamps and deforestation, could have reduced the prevalence of the mosquito vector by the early twentieth century but also lowered immunity in the human host, leading to outbreaks in 1938 and subsequently.

Warm wet winters followed by dry summers have favoured the occurrence of St Louis Encephalitis, the most common form of arboviral encephalitis in the United States. With climate change, a northward extension of its current range is possible.

In Australia, since 1917, there have been seven major epidemics of Murray Valley Encephalitis. Epidemics have followed summers with high rainfall in the major catchment

areas of central and eastern Australia. The breeding cycle of the main mosquito vector, *Culex annulirostris*, is critical to the seasonal occurrence of this disease, which has peaked in the late summer-early autumn. Why epidemics have not occurred in southeastern Australia since 1974 remains unclear. With climate change, however, an increase in this disease is predicted, as variations of summer rainfall may promote the breeding of waterbirds and mosquitoes.

European tick-borne encephalitis is seen in temperate zones in the Northern Hemisphere. Its distribution may expand with global warming. In Sweden, for example, increased incidence of tick-borne encephalitis has been reported in association with milder winters (Lindgren, 1998).

Related syndromes in animals include the North American encephalitides, viral diseases occurring episodically in horses (and less frequently in humans). Weather variables have been used to develop predictive models for equine encephalomyelitis in the United States (Francy *et al.*, 1992).

Hantaviruses

Hantavirus pulmonary syndrome and haemorrhagic fever with renal syndrome have been recognised to be associated with changes in rodent populations, in turn related to variation of season and land use.

Rift Valley Fever

First described in 1930, epizootics of Rift Valley fever, sometimes accompanied by epidemic human disease, have occurred throughout sub-Saharan Africa. The acute viral infection may produce subclinical or mild disease in adult sheep and cattle, but is often fatal to young animals. High fever, severe headache, myalgia, and exhaustion comprise the acute syndrome in humans, while some develop fatal encephalitis and ocular disease.

Wilson (1994) traced the roles of environmental change (especially unusual rainfall), migration of ungulates, economic development, including the presence of dams and irrigation systems, human demography, behaviour and travel on the dynamics of the *Aedes* and *Culex* mosquito vectors, resulting in changes in endemic prevalence, and the frequency and intensity of epidemics.

Ross River Virus Disease

This syndrome of epidemic polyarthritis normally lasts a few days to several weeks, causing high morbidity. Widespread in Australia, the attack rates for Ross River virus disease are higher, for example, than rates associated with the vector-borne disease of greatest profile in North America (Lyme disease).

The greatest incidence of Ross River virus disease occurs in the wet season in northern Australia, and in the summer to mid-autumn period in southern Australia, a seasonal pattern that depends on the life cycle and habits of the mosquito vectors. Temperature, humidity and rainfall affect the incidence of this disease, while human activity (such as the development of irrigation) is also clearly important (Curson, 1996).

Climate change is expected to alter the incidence and distribution of Ross River virus disease. In coastal areas, sea level rise may enhance the penetration of tidal saline waters into estuarine and tidal river habitats, by consequence extending the breeding grounds of the mosquito vectors (e.g., two saltmarsh-breeding species, *Aedes camptorhynchus* and *vigilax*). Moreover, greater numbers of humans, particularly along the eastern seaboard, are expected to become exposed.

Yellow Fever

Yellow fever is a mosquito-borne disease occurring in tropical Africa and America. With the potential for urban outbreaks, the range of yellow fever may also be changing.

FUTURE CHALLENGES

Controlling these viral and other diseases will require improvements in both surveillance and risk communication. Future disease surveillance activity should examine how changing climate could influence both the environmental sources of pathogens and human behaviour (Patz *et al.*, 2000). Public health experts should become more familiar with the work of urban planners and *vice versa*.

Climate change will be important for these diseases and probably many others (such as Japanese Encephalitis) for which the climate-disease relationship has not yet been researched adequately. So far, research has principally been undertaken from centres in developed countries, but more detailed analyses from other regions are now underway (Githeko *et al.*, 2000). In animals, other possible examples of increasing arboviral disease activity include akabane, bluetongue and bovine ephemeral fever.

As this work goes to press, the outbreak of foot and mouth disease in the UK in the year 2001 has illustrated graphically some of the global forces additional to climate that must be considered in the changing environment. These forces included the de-regulation of the economy, with pressure on farmers. The decline of state veterinary services may also have contributed. Spread of this virus was promoted by the long-distance (and sometimes illegal) movement of animals, humans and vehicles, at vast economic and social cost. The effectiveness of intervention by vaccination as opposed to the policy of mass animal slaughter

was debated intensely. The panic moved to the centre-stage of politics. (See also Woolhouse and Donaldson, 2001).

Climate and disease should not be investigated separately from the behaviour of human populations or the practice of public health. Research priorities include the analysis of disease patterns in relation to climatic variables, but not restricted to them. The development of predictive models is needed to inform public policy in health and other sectors (Woodruff *et al.*, 2000). Research on preventive responses (including but not limited to the development of vaccines) is required. Adaptation to changing patterns of disease also needs research. To wait and see, an alternative preferred by some governments, could add an avoidable future burden of morbidity and mortality.

See also: Environmental Change and Human Health: Extending the Sustainability Agenda, Volume 3; Malnutrition, Infectious Diseases and Global Environmental Change, Volume 3.

REFERENCES

Chan, N Y, Ebi, K L, Smith, F, Wilson, T F, and Smith, A E (1999) An Integrated Assessment Framework for Climate Change and Infectious Diseases, *Environ. Health Perspect.*, 107, 329 337.

Curson, P (1996) Human Health, Climate and Climate Change: an Australian Perspective, in *Climate Change: Developing Southern Hemipshere Perspectives*, eds T W Giambelluca and A Henderson-Sellers, John Wiley & Sons, Chichester, 319 348.

Francy, D B and Wagner, B A (1992) Equine Encephalomyelitis in the United States, 1980 1990, *Anim. Health Insight*, Spring, 1 8.

Githeko, A K, Lindsay, S W, Confalonieri, U E, and Patz, J A (2000) Climate Change and Vector-borne Diseases: a Regional Analysis, *Bull. WHO*, 78, 1136 1147.

Gubler, D J (1998) Dengue and Dengue Hemorrhagic Fever, *Clin. Microbiol. Rev.*, 11, 480 496.

Guest, C S, Willson, K, Woodward, A J, Hennessy, K, Kalkstein, L S, Skinner, C, and McMichael, A J (1999) Climate and Mortality in Australia: Retrospective Study, 1979 1990, and Predicted Impacts in Five Major Cities in 2030, *Clim. Res.*, 13, 1 15.

Halstead, B W (1982) *Paralytic Shell sh Poisoning Guide*, World Health Organization, Geneva, 66.

Humpage, A R, Rositano, J, Baker, P D, Nicholson, B C, and Steffensen, D A (1993) Paralytic Shellfish Poisons from Freshwater Blue-green Algae, *Med. J. Aust.*, 159, 423 7.

Jetten, T H and Focks, D A (1997) Potential Changes in the Distribution of Dengue Transmission Under Climate Warming, *Am. J. Trop. Med. Hyg.*, 57, 285 287.

Lindgren, E (1998) Climate and Tick-borne Encephalitis in Sweden, *Conserv. Ecol.*, 2, 5 7.

Linthicum, K J, Anyamba, A, Tucker, C J, Kelley, P W, Myers, M F, and Peters, C J (1999) Climate and Satellite Indicators to Forecast Rift Valley Fever Epidemics in Kenya, *Science*, 285, 397 400.

McMichael, A J, Haines, A, Slooff, R, and Kovats, S, eds (1996) *Climate Change and Human Health*, World Health Organization, Geneva.

Patz, J A, Engelberg, D, and Last, J M (2000) The Effects of Changing Weather on Public Health, *Annu. Rev. Pub. Health*, 21, 271 308.

Shope, R E (1994) The Discovery of Arbovirus Diseases, *Ann. N. Y. Acad. Sci.*, 740, 138 145.

Wilson, M L (1994) Rift Valley Fever Virus Ecology and Epidemiology of Disease Emergence, *Ann. N. Y. Acad. Sci.*, 740, 169 180.

Woodruff, R E and Guest, C S (2000) Teleconnections of the El Nino Phenomenon: Public Health and Epidemiological Prospects, in *El Niño: History and Crisis*, eds R Grove and J Chappell, White Horse Press, Cambridge, 89 108.

Woolhouse, M and Donaldson, A (2001) Managing Foot-and-Mouth, *Nature*, 410, 515 516.

Vitality Fertilization of Forests

see Nitrogen Deposition on Forests (**Volume 2**)

Volcanic Eruptions

see Volcanic Eruptions (**Volume 1**)

Volcanic Eruptions: Mt Merapi, Indonesia

Sutikno

Gadjah Mada University, Yogyakarta, Indonesia

Merapi volcano is one of the 26 most active volcanoes in Indonesia (Kusumadinata, 1979), and is one of about a dozen volcanoes in the world which are in a state of persistent eruptive activity (Tazieff, 1983). The Merapi volcano is located in Central Java, Yogyakarta Special Region Province, Republic of Indonesia; the 2911 m high peak is located at 7° 32′ 30″ S, 110° 26′ 30″ E, or about 25 km north of the Yogyakarta urban area.

The earliest record of Merapi activity was in 1006. Since then historical records show that big eruptions occur every 9- 16 years, while smaller eruptions occur at about 2- 3 year intervals. This activity is associated with two intersecting faults under the volcano. The consequences are:

eruption, lava ow, glowing cloud/nuée ardente (local term for the nuée ardente is wedus gembel - big white sheep), ash fall and lahar (mud) ow. Merapi eruptions are not explosive. The magma rises slowly and piles up above the active vent of the volcano to build a lava dome. The lava dome is unstable and partly slides down along the slope together with pyroclastic materials. If associated with heavy rain, a rain lahar occurs.

Living near an active volcano, such as the Merapi Volcano is full of danger and risk. Eruptions in 1930 and 1954 killed 1369 and 64 people, respectively (Kusumadinata, 1979). The Merapi activity with nuée ardente on November 22nd, 1994 killed 85 people, 22 people were seriously wounded, 17 of whom had to be hospitalised. Flora and fauna were destroyed, mostly caused by high temperature of the glowing cloud of 300 600 °C. From May to July 1998 Merapi was active again with nuée ardente and ash fall but without any loss of life. Archaeological remains show that the danger of the Merapi Volcano has been known for a very long time. Some Hindu temples (8 9th century) in the Merapi volcanic area were buried by volcanic material more than 2 m deep, and subsequently partly restored, such as the Sambisari Temple. In 1991, archaeological artefacts made from gold, supposedly from the Hindu period, were found in Wonoboyo village, in the eastern part of Yogyakarta town. Buried natural features, such as ancient rivers, palaeosols and charcoal were found everywhere in the sub-surface layers in the Merapi volcanic area.

In order to prevent deaths due to Merapi eruptions, the Volcanological Bureau has delimited four danger zones based on the magnitude of the danger:

1. *the prohibited zone*: an area located near the source of the danger, on the volcanic cone and the upper slope of the Merapi; this zone is not inhabited and naturally is not arable due to the pressure of very recent volcanic materials and lack of water;
2. *the level one danger zone*: an area where at any time a big eruption could cause danger, either by nuée ardente, bomb, lapilli, ash fall or their combination. This zone is located just below the prohibited zone, on the middle volcanic slope; this zone is inhabited and the land is utilised for settlement, dry fields and agroforestry;
3. *the level two danger zone*: an area located along the river banks flowing down from the volcanic slope of the Merapi Volcano; the main danger of this zone is rain lahar; there are eight main rivers flowing from the Merapi volcanic slope, i.e., Woro, Gendol, Kuning, Boyong, Krasak, Batang, Putih, Blongkeng and the Pabelan River; all of the rivers flowing through highly populated areas, and fertile and productive agricultural lands;
4. *the safe zone*: an area free from the threat of nuée ardente, lapilli, volcanic bombs, ash fall and rain lahar. Most of the urban areas of Yogyakarta are located in this zone (Kusumadinata, 1979).

People living in danger zones one and two should be prepared mentally and physically for volcanic activities, especially in an active period. Merapi volcano eruptions cause environmental changes physically, biologically and culturally. The physical environmental change is caused by volcanic deposits, such as lava flow, pyroclastic, nuée ardente, volcanic ash and lahar deposits. In an area covered by nuée ardente, the environmental changes are tremendous; dense and very hot masses of material that move rapidly are able to kill humans and animals, destroying and burning wooden material and vegetation, as shown by the tragedy of November 22nd, 1994 described earlier. All the trees facing in the direction of the nuée ardente flow were burnt, while the trees on the opposite side were still fresh. The ash fall deposit in a Merapi eruption varies in thickness while its spatial distribution depends on the position of the vent hole and the wind direction. In the eruption of May to July 1998, the thickness of the ash fall in the highest settlement was generally between 2 5 cm. The ash covered the agricultural land and most of the cash crops were damaged. The trees with wide and dense leaves fell down due to the weight of ash. Farmers had to wash grass in order to feed their cattle. Roofs of houses were also covered by volcanic ash, and the owners had extra work to do. The volumes of the volcanic deposits of the eruptions in 1969 and 1984 were estimated to be 5.5 million m^3 and 6.5 million m^3, respectively. The huge mass of volcanic materials covered the steep volcanic slopes, where precipitation is high (more than 3500 mm annually). In this saturated condition the material flowed down the slopes as rain lahar along the main stream, in some cases overflowing the banks and destroying settlement areas or agricultural land.

Even though living in an active volcanic area is dangerous, the area is still densely populated. Fertile soil, and plenty of water are magnets that attract people, who therefore refuse to migrate permanently to other locations. There are reasons why they do not like to move, such as:

1. Their forefathers have lived there for a very long time and historically they feel safe from the hazard; they believe that Merapi is their friend, and thus most of them are not afraid of Merapi activity and even believe that Merapi activity offers some advantages.
2. They believe that an accident can occur anywhere and anytime; and they believe the Merapi activity is only temporary, so they are willing to move temporarily if Merapi is active.
3. Most of them feel happy living in their settlement environment, because they have a piece of land with

fertile soil, water is available and they have enough foodstuff for their family.
4. Usually in the rural areas, they have a close relationship with their family and their neighbors, they do not want to lose this; they have a slogan in Javanese, mangan ora mangan waton kumpul which means whether we eat or not, the important thing is to live together ; they do not know whether a new place could bring certainty into their life.

Government policy with respect to people who do not want to leave the danger zone can be divided into three measures:

1. an early warning system; using detailed information from the Bureau of Volcanology and coordinated by the Governor and Regent; in some parts of the danger zone sirens were installed; in areas where the siren s sound is not heard, the local people use traditional tools, made from bamboo or wood (local term is kentongan) to indicate the danger of an eruption;
2. construction of physical structures to protect settlements from the lahar (mud) flow, such as the Sabo dam and embankment;
3. preparing evacuation and relocation of settlements.

Based on the accident on November 22nd, 1994 when the nuée ardente from Merapi rushed into Turgo village, the government prepared a relocation settlement area in the safe zone. The relocation settlement consists of single houses, public facilities and barns, such facilities are normal or even better than in other rural areas in mountainous areas in Java. Only a few people have stayed in the relocation settlements; the government regulation states that the people can work in their former fields, but do not allow them to stay permanently in the first danger zone; temporarily they obeyed the government regulation but after a few months they went back home again. The people who live along the rivers descending from the Merapi Volcano collect sand, gravel and stone to use as construction material. Formerly they collected the material manually, but when demand for materials increased rapidly due to development projects, private companies came with machines and trucks to collect and transport the materials to the urban area. After some time a conflict arose between the local people and the private companies. It meant that in a short time a lot of material was dug out from the land surface or from the river beds. Every day hundreds of trucks carry material out from the river beds mainly to the urban area. There are some impacts due to the withdrawal of the material from the river beds, i.e.:

1. most of the roads for transportation of surficial material are damaged by heavy trucks; in the dry season the air along the road becomes dusty;
2. along the river courses there are some intakes on the riverbank for irrigation purposes. Due to the lowering of the riverbed caused by withdrawal of the material, the irrigation channel offtakes cannot obtain enough water. This lack of water reduces the production of rice in the field fed by the channel. In the dry season most of the rice fields cannot be supplied by irrigation water;
3. as their income from the rice field decreases, farmers seek other sources of income to meet their daily needs. Some of them work collecting surficial material, which means that the area of sand and gravel exploration increases rapidly. In the rainy season the river course sometimes changes and vertical erosion becomes more active.

The Merapi volcano thus brings both benefits and problems. Constant vigilance and preparedness is required by everyone seeking to exploit the benefits by taking the risk of living in the areas likely to be affected by eruptions and the associated nuée ardentes and lahars.

REFERENCES

Kusumadinata, K, ed (1979) *Data Dasar Gunungapi Indonesia*, Volcanological Survey of Indonesia, 1 820.

Tazieff, H (1983) Monitoring and Interpretation of Activity on Mt. Merapi, Indonesia, 1977 80: a Practical Example, Civil Defense, in *Forecasting Volcanic Events*, eds H Tazieff and J C Sabroux, Elsevier, Amsterdam, 485 494.

Volcanic Explosivity Index (VEI)

see VEI (Volcanic Explosivity Index) (**Volume 3**)

Volcanoes and Cities

Grant Heiken

Los Alamos National Laboratory, Los Alamos, NM, USA

A volcanic eruption near a metropolitan area is capable of major destruction of property and loss of life. Globally, there are 60 large cities located near or on historically active volcanoes; including two megacities (Manila and Mexico City) and the population at risk is over 40 million people (1990 census estimates).

THE HAZARDS (SEE FIGURE 1)

- The greatest risks to life and property are posed by the fast-moving volcanic density currents called *pyroclastic ows* or surges. These rapidly moving, hot mixtures of volcanic ash, pumice, and gas possess energies capable of destroying stone buildings. During the 1902 eruption of Mont Pelée, Martinique, 29 000 lives were lost when volcanic density currents destroyed the town of St. Pierre. Infrequent but large volcanic density currents from the Philippines Taal Volcano once swept across the area now occupied by Manila.
- Although less dangerous than volcanic density currents, ash and pumice fallout can cover large areas, and damage the regional electrical and water infrastructure. Fallout also causes respiratory ailments, smothers plants and insects, and creates enormous clean-up problems (like a snowfall that fails to melt). In 1997, a light ashfall from Popocatépetl slowed down life in Mexico City, with respiratory irritation, slick roadways, and closure of the International Airport. Kagoshima City, Japan is regularly showered with ash from the nearby Sakurajima volcano, and ash cleanup is a part of the public works department s regular duties.
- Volcanic mudflows are generated after water and volcanic ash mix on the volcano, or ashfall deposits on hillsides are saturated by heavy rains. Prehistoric mudflows swept across the valleys of Mexico and Puebla with severe effects on Preclassic cultures. In 1985, volcanic mudflows from Nevado del Ruiz, Colombia, flowed through the town of Armero, located on the Río Lagunillas 70 km from the crater, killing 25 000 people. Mudflows from Mount Rainier, Washington are the most worrisome potential hazard for the lowlands near Seattle and Tacoma.
- Lava flows are a part of history for Catania, Sicily, where lava flows from an eruption of Mt Etna in 1669 traveled 17 km and destroyed the western part of the city. Lava flows have erupted intermittently from Mount Vesuvius during the last 370 years, damaging towns on the volcano s slopes.
- Gases are the least obvious but the most insidious of volcanic hazards. Volcanic gases, including carbon dioxide (CO_2), sulfur dioxide (SO_2), and hydrogen sulfate (H_2S), are heavier than air and can flow along valley bottoms and accumulate in closed areas such as craters, suffocating any breathing creatures. Effects near cities are mostly localized, but an eruption of CO_2 from a crater lake in the Cameroon highlands in 1986 flowed down two river valleys, suffocating 1700 people in five villages. It has been suggested that many of the deaths in Pompeii during the AD 79 eruption of Vesuvius were caused by CO_2 suffocation.
- Tsunamis generated by submarine caldera collapse, or by structural failure and collapse of a composite cone near the sea have damaged or destroyed coastal cities during historical time. The most famous eruption is that of Krakatau, Indonesia in 1883, when most of the 36 000 deaths in coastal towns along the Sunda Straits were caused by volcanogenic tsunamis. Most large cities with this potential hazard are located around the Pacific Rim, and especially around the coasts of Japan, the Philippines, Indonesia, and Papua-New Guinea.

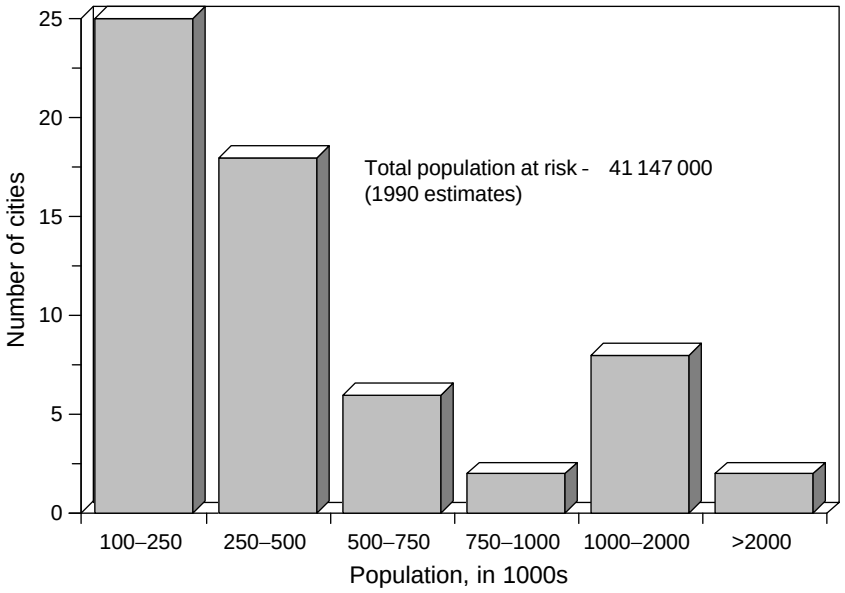


Figure 1 Cities of • 100 000 population near potentially active volcanoes

Table 1 List of cities of • 100 000 population near potentially active volcanoes

Yerevan, Armenia
Vancouver, Canada
Ibague, Colombia
Manizales, Colombia
Pasto, Colombia
Popoyan, Colombia
San José, Costa Rica
Quito, Ecuador
Cuenca, Ecuador
San Salvador, El Salvador
Santa Ana, El Salvador
San Miguel, El Salvador
Bandung, Indonesia
Bogor, Indonesia
Cianjur, Indonesia
Tasikmalaya, Indonesia
Magelang, Indonesia
Yogyakarta, Indonesia
Surakata, Indonesia
Malang, Indonesia
Medan, Indonesia
Jakarta, Indonesia
Semarang, Indonesia
Catania, Italy
Napoli, Italy
Torre del Greco, Italy
Kagoshima City, Japan
Kumamoto, Japan
Oita, Japan
Nagano, Japan
Marbashi, Japan
Morioka, Japan
Murovan, Japan
Asahikawa, Japan
Miyazaki, Japan
Odawara, Japan
Matsumoto, Japan
Sendai, Japan
Akita, Japan
Cuernavaca, Mexico
Guadalajara, Mexico
Ciudad de Mexico, Mexico
Orizaba, Mexico
Toluca, Mexico
Ciudad de Guatemala, Guatemala
Auckland, New Zealand
Managua, Nicaragua
Arequipa, Peru
Angeles City, Philippines
Olangapo, Philippines
Manila, Philippines
Quezon City, Philippines
Batangas, Philippines
Bacolod, Philippines
Petropavlovsk-Kamchatsky, Russia
Las Palmas, Canary Islands, Spain
Santa Cruz, Canary Islands, Spain
Portland, USA
Seattle, USA
Tacoma, USA

THE GLOBAL ENVIRONMENT

The infrequent, large volcanic eruptions that have had global environmental effects, such as the 1815 eruption of Tambora, Indonesia, have also affected the world s cities (see Table 1). The 1815 eruption, responsible for the year without a summer, caused crop failures and famine in North America, eastern Europe, and China. In addition to stress related to food shortages, the gloomy, hazy days caused by the substantial sulfuric acid aerosol in the stratosphere that led to anxiety and civil unrest (*see* Volcanic Eruptions, Volume 1).

Because of the proximity of volcanoes such as Popocatépetl, Mexico (Mexico City and Puebla), Guagua Pichincha, Ecuador (Quito) and Taal, Philippines (Manila) to densely populated metropolitan areas, volcanologists are becoming involved in urban disaster mitigation planning, moving into the city. In a few cities, volcanic hazard studies are being added to integrated models of urban infrastructure (e.g., transportation, energy, civil defense, telecommunications, water distribution, and public health), and the social and political framework. Prevention of large-scale disasters in cities on or near volcanoes requires more than the traditional hazard mapping and volcano monitoring. The community has to be involved through extensive education and public awareness campaigns. Risk evaluation must rely heavily on modeling, and requires visualization of the physical processes involved, and their effects, in a manner that is easily grasped by emergency planners, the insurance industry, policy makers, and the public.

Volcanoes can catastrophically change environments in nearby towns and cities by simply destroying them. Widespread ash fallout and gases can also cause environmental catastrophes, with effects ranging from respiratory problems to seasons without bees for plant pollination. More indirect effects of large eruptions include global temperature decreases and possibly the enhancement of El Nino phenomena. In any case, with ever-increasing urban populations, volcanoes constitute a major environmental threat (*see* Volcanic Eruptions: Mt Merapi, Indonesia, Volume 3).

FURTHER READING

Blong, R J (1984) *Volcanic Hazards - a Sourcebook on the Effects of Eruptions*, Academic Press Australia, North Ryde, 1 424.

Francis, P (1993) *Volcanoes - a Planetary Perspective*, Clarendon Press, Oxford, 1 443.

Sigurdsson, H, ed (1999) *Encyclopedia of Volcanoes*, Academic Press, New York, 1 1417.

Simkin, T and Siebert, L (1994) *Volcanoes of the World*, 2nd edition, Geoscience Press, Tucson, AZ, 1 349.

Volcanoes and the Environment

Susan L Donoghue
University of Hong Kong, Hong Kong

The surface of our Earth is studded with restless volcanoes, living reminders of the forces that have helped shape our world since its inception. The distribution of most of these volcanoes is governed by complex interactions between large mobile crustal plates that t together like pieces of a jigsaw to form the surface of the Earth. Some 50- 70 volcanoes erupt every year, but few of these have signi cant or lasting impacts. The greatest immediate dangers and possible changes to the global climate arise from volcanoes that erupt on time scales of hundreds to thousands of years. Several cataclysmic eruptions of the past few centuries have resulted in a volcanic forcing of climate and a temporary marked cooling of Earth's surface temperatures, with wide-ranging consequences.

It is no accident that many of the world's people live in the shadows of active and dormant volcanoes. This perilous existence is governed by our need for resources and because of this, people are prepared to risk their lives in return for the many bene ts that volcanism can bring. Although volcanoes are typically seen as agents of death and destruction, their ery breath also helps to sustain life by providing us with an essential resource - extensive regions of the most fertile soils in the world, capable of producing successive rich harvests to feed an ever increasing population.

THE FIRE MOUNTAINS

The theory of Plate Tectonics, a revolution in the earth sciences, holds that the rigid outer shell of the Earth is broken into a mosaic of plates that are in constant relative motion. These plates interact at their boundaries giving rise not only to the world s earthquakes but to the many diverse physical features of Earth s landscape including the myriad of volcanoes that lace Earth s surface (Figure 1).

Large concentrations of volcanoes occur above subduction zones, areas where lithospheric plates moving towards each other collide. The process of subduction of the underlying oceanic crust generates magma within the upper mantle, which rises beneath and intrudes into the overlying crustal layers. This magma eventually erupts at Earth s surface as lava, forming volcanoes.

Volcanoes formed above subduction zones occur as arcuate chains within the oceans where they are termed island arcs and along the margins of continents that lie above subduction zones, forming active continental margins. Subduction along the margins of the Pacific Plate

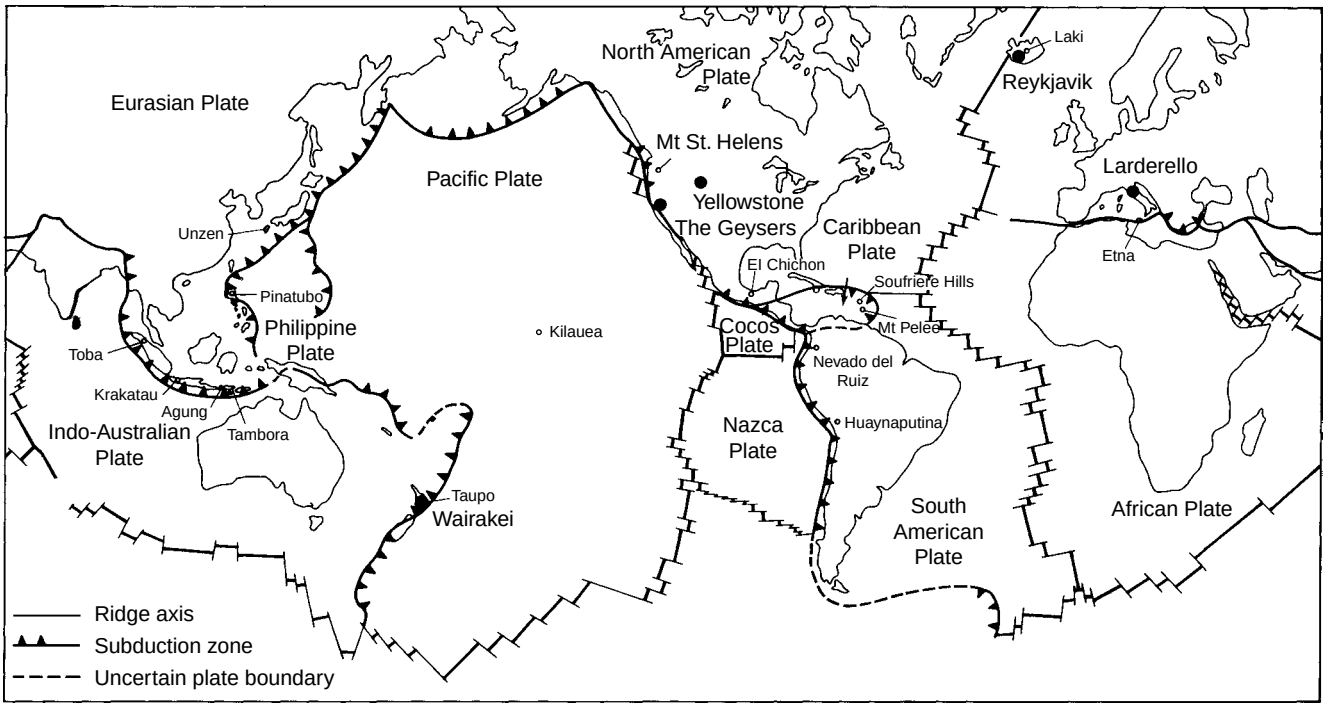


Figure 1 The location of some of the world's active volcanoes and developed geothermal fields in relation to the major tectonic plates of the Earth's surface. Subduction along the margins of the Pacific Plate has created a Ring of Fire where two-thirds of the world's active volcanoes are found

has created a spectacular belt of volcanism termed the Pacific Ring of Fire which is home to no less than 500 of the world s most active and explosive volcanoes.

Eruptions at subduction-related volcanoes are typically violent. Several to tens of cubic kilometers of tephra and gases may be explosively ejected into the atmosphere. Much of this tephra later falls back to ground, blanketing the landscape downwind of the vent. These eruptions are also known for their propensity to generate lahars (*see* Lahars, Volume 3) and pyroclastic flows, the latter being the most terrifying and certainly one of the deadliest of all volcanic phenomena. Lahars generated during the 1985 eruption of Nevado del Ruiz in Colombia overran the town of Armero, 50 km from the volcano, burying 23 000 citizens. Pyroclastic flows generated during the 1902 eruption of Mt Pelèe on the Lesser Antilles Island of Martinique incinerated the town of St Pierre, killing 27 000.

Not all of Earth s active volcanoes, however, occur at subduction zones. Thousands of active volcanoes dot the ocean floors in regions along mid-oceanic ridges, enormous fractures running the length of the ocean basins. These are areas where plates are being pulled away from one another at rates of 6 15 cm year^{-1}. Eruption of lava along the crests of these ridges generates new oceanic crust compensating for the destruction of oceanic crust at subduction zones. In Iceland, where a segment of the Mid-Atlantic ridge is exposed on land, birth of new crust is witnessed with great regularity. Icelandic eruptions are characterized by outpourings of enormous volumes of fluid basaltic lava and spectacular fire-fountaining episodes along the rift.

Active and extinct volcanoes are also found in seemingly anomalous locations away from plate boundaries. They occur where localized and relatively stationary upwellings of mantle material, termed hot spots or mantle plumes, sit beneath the crust, scorching the underside of the tectonic plate and eventually burning holes through it. Magma erupts to form volcanic islands in oceanic settings and isolated volcanic centers in continental settings. The Hawaiian islands, a chain of volcanoes formed as the overriding Pacific plate migrates over the hot spot, and Yellowstone volcanic center, are two such examples. Eruptions at hot spot volcanoes are more effusive than explosive in character, building broad shield volcanoes and lava plateaux.

Every year an estimated 50 70 volcanoes erupt somewhere on Earth. A few volcanoes, such as Kilauea (Hawaii) and Stromboli (Italy) are in more or less continuous eruption, but for most volcanoes, short-lived and often violent outbursts punctuate much longer periods of relative inactivity or dormancy. Dormant volcanoes, liable to awaken with very little warning, present the greatest threat to populations living in volcanic regions. Take, for example, the sudden awakenings of Washington State s Mount St. Helens in 1980 after a 600 year slumber, Mexico s El Chichón volcano in 1982 (which was thought to be extinct), and the recent 1995 1997 eruptions at Soufriére Hills Volcano on the Caribbean Island of Montserrat after a 400 year respite.

The potential loss of life from future eruptions is enormous. At risk are some 350 500 million people living on or near potentially active volcanoes, many of the people inhabiting towns that have been repeatedly rebuilt following previous disasters. Every century an estimated 100 000 people die as a result of volcanic activity. Twenty-seven volcanoes including Nevado del Ruiz (Colombia), Mt Unzen (Japan), El Chichón (Mexico), Tambora (Indonesia) and Laki (Iceland) have earned special distinction, each having killed more than 1000 people in the last 300 years.

In an effort to reduce the global loss from natural disasters, the United Nations declared the 1990s an International Decade for Natural Disaster Reduction (*see* IDNDR (The International Decade for Natural Disaster Reduction), Volume 4). The emphasis has been on cooperative international effort to mitigate disasters, especially in developing countries. In response to this, teams of scientists the world over have been busy establishing volcano surveillance networks and undertaking risk assessment studies at 16 of the world s most pernicious volcanoes, dubbed the decade volcanoes. Their aim has been to provide adequate warning of imminent activity to allow safe evacuation of populations living near volcanoes. By the close of this Decade program, their tremendous efforts will have meant the saving of hundreds of thousands of lives.

AGENTS OF CHAOS

Although immediate death and destruction are the most frequently described effects of volcanic eruptions, there may also be severe post-disaster social and economic effects. Historically, large magnitude eruptions and events of particularly long duration have resulted in widespread famine, disease and massive displacements of human populations. In recent years, the effects of famine and disease have been greatly reduced through international relief programs, but the problem of how to cope with mass human migration in the aftermath of disasters remains. This is a particularly pertinent issue for less developed Ring of Fire countries presently home to nearly half of the world s population and many of the largest growing cities. In the last two decades alone, eruptions in the Americas, Philippines and Indonesia have displaced in excess of one million people, with consequent hardship and economic distress. Such displacements of population occur when large areas of land engulfed by volcanic sediments are rendered temporarily or permanently uninhabitable. Recent eruptions at Mt Pinatubo (Philippines) and Soufriére Hills Volcano (Montserrat) are pointed illustrations of this problem.

Mt Pinatubo, one of a chain of volcanoes of the Luzon volcanic arc, had not erupted in 600 years. However, on

June 15, 1991, after several months of volcanic unrest, activity culminated in a climatic eruption described as the eruption of the century. Within one hour of the commencement of the eruption an enormous and vigorous eruption cloud developed which then spread laterally, at stratospheric levels, for 400 km. As the cloud expanded, tephra rained across 7500 km^2 of central and Western Luzon. At the same time, numerous pyroclastic flows roared down the slopes of the volcano in-filling and blocking major drainage systems with tens of meters of pyroclastic material. Rivers forced to flow unconfined across broad pyroclastic fans, mixed with loose debris generating voluminous debris flows that traveled down all major tributaries draining the volcano. Numerous lahars were also triggered by heavy rains associated with the untimely landfall of typhoon Yunya the same day (Figure 2).

The immediate effects of the eruption on local communities were disastrous. Several hundred lives were lost, millions of head of livestock killed, and an estimated 200 000 km^2 of farmland, the heart of Luzon s rice growing region, were lost beneath thick ash and lahar deposits. A staggering half a million people had been made homeless. In the longer term scientists projected that devastating lahars would likely occur for another 10 years as monsoon rains remobilize the 5 7 km^3 of loose pyroclastic material left resting on the slopes of the volcano. From June 1991 onwards literally hundreds of rain-generated lahars flowed down the main drainage pathways impacting communities up to 50 km away from the volcano and killing at least 850 people. In 1995, lahars breached hastily constructed dykes built to protect the heavily populated communities of Angeles City, San Fernando, Porac and Bacolor. Torrents of mud flooded prime agricultural lands and entombed Bacolor killing 100 people and leaving a further 100 000 homeless (Figure 3).

Much of the farmland around Pinatubo is now uninhabitable wastelands. This, together with the continued threat of lahars, has effectively prevented permanent resettlement of

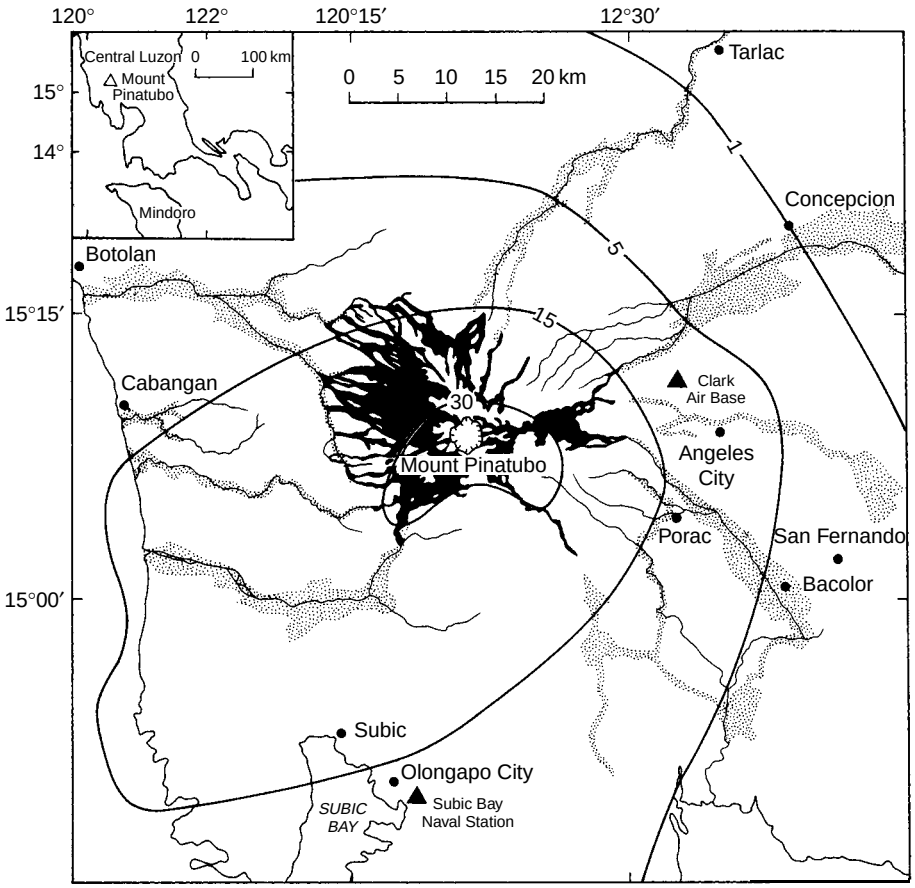


Figure 2 A composite map showing the distribution of tephra-fall deposits (isopachs are in centimeters) and pyroclastic flow deposits (in black) from the climatic June 15 1991 eruption at Mt Pinatubo. A thick blanket of tephra covered much of central and Western Luzon burying the heart of the rice bowl region. Also shown is the extent of deposition from hundreds of lahars (in stipple) that flowed down all major drainage systems between 1991 and 1993. Adapted from Janda *et al.* (1995); Paladio-Melosantos *et al.* (1995); and Umbal and Rodolfo (1995). (Reproduced with permission from the United States Geological Survey and the Philippine Institute of Volcanology)

Figure 3 In 1995 rain-generated lahars breached dykes built to protect communities, burying Bacolor town and surrounding agricultural lands in torrents of mud. This 1997 photograph of Bacolor town shows old housing buried beneath recent lahar deposits. The new housing alongside has been elevated in an attempt to mitigate damage from recurring lahars. (Reproduced with permission of M Watanabe, Tokyo Institute of Technology)

the region and the resumption of vital economic activities. In fact, the physical landscape has been so transformed that today many of those who once lived on the slopes of Pinatubo are unable to locate their homelands, let alone return to them. The indigenous Aeta people, who prior to the eruption had farmed the slopes of the volcano, have lost not only their homelands but also their lifestyle, and presently face an uncertain future as they attempt to adjust to an urban existence.

Soufriére Hills Volcano located on the tiny island of Montserrat, in the Lesser Antilles volcanic chain, began erupting on July 18, 1995. The eruptions marked the beginning of an extended three year period of sporadic and at times violent activity, and the first eruption in recorded history for this peaceful Caribbean hideaway. Most affected were regions to the South, located downwind of the volcano. As eruptive activity escalated through 1995 1997 and towns were repeatedly showered with ash and threatened by pyroclastic flows, citizens in the South were progressively evacuated to designated safe zones in the North of the island. By August 1997, the volcano had claimed back two-thirds of the island, destroying in the process the capital of Plymouth. With continuing displacement of residents from the South, growing and intolerable pressure was exerted on the limited resources available in the North of the island, a situation made worse by ongoing eruptions which forced authorities to zone even more land as unsafe. The area occupied by safe zones was shrinking and conditions within them deteriorating. Hundreds of evacuees were living in cramped sub-standard shelters with little immediate prospect for improvement in living conditions. Eventually, with the threat of further activity and the growing realization that they may never be able to return to their homes in the South, two-thirds of the population left Montserrat in September 1997, accepting offers to be resettled on other Caribbean islands or in Britain.

FIRE AND ICE

The detrimental impacts of volcanism may also extend beyond these regional effects to influence the planet as a whole and specifically its climate. Our present understanding how volcanic eruptions affect climate derives from a much improved understanding of eruptive processes, revised interpretations of written accounts of famous historic eruptions, and new geologic and climate-proxy evidence of the effects of volcanism in prehistoric and historic times. These all lend support to the concept of volcano forcing as a mechanism of climate change.

Any sustained explosive eruption that generates a high-altitude (stratospheric) eruption plume has the potential to influence global climate. Such eruptions are termed Plinian and are characteristic of those at subduction-related volcanoes. Tephra, and hot gases released from the magma, are forcefully ejected into the atmosphere, forming a low-level eruption cloud above the vent. Initially, the momentum of this cloud allows it to propagate upward to heights of many hundreds of meters. However, in order for the cloud to reach higher stratospheric altitudes, its density must be reduced below that of the surrounding atmosphere. This will occur if sufficient atmospheric air is entrained into the cloud and heated. Buoyancy forces will then dominate and strong vertical convection will allow the cloud to continue to rise tens of kilometers until it reaches an altitude where its density becomes equal to that of the surrounding atmosphere. Here, the plume will begin to spread laterally forming a characteristic stratospheric umbrella cloud (Figure 4).

Volcanic ash carried to high altitudes in the convecting plume rains out from the base of the umbrella cloud over a period of several days to months, blanketing the landscape below it. Due to the short residence time of ash in the stratosphere, the effect of the particulate loading on Earth s global radiation budget is negligible. Rather, it is the gas content of the plume, specifically the sulfur dioxide loading, that is of importance. Once injected into the stratosphere, sulfur dioxide gradually reacts to form sulfuric acid aerosols, clouds of fine liquid droplets dispersed in air. These aerosols may persist for several years during which time they reduce the receipt of solar radiation at the Earth s surface by scattering and absorbing solar radiation. Their effect is to reduce tropospheric temperatures. The degree of cooling is a function of both the optical density (the extent to which aerosols scatter and absorb solar radiation) of the aerosol and its latitudinal distribution as it spreads with the atmospheric circulation. Eruptions at equatorial latitudes are

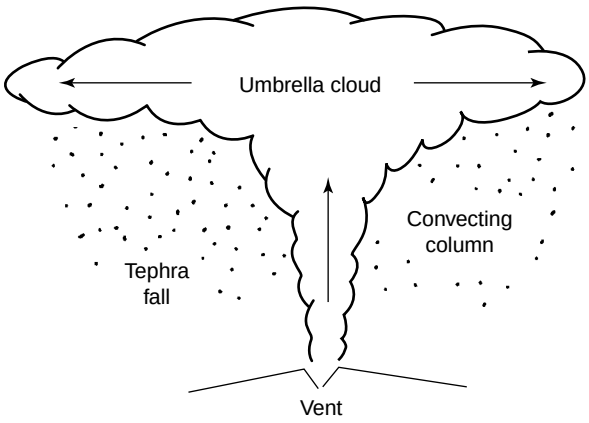


Figure 4 Schematic diagram showing a buoyant convective eruption column and stratospheric umbrella cloud. Tephra falls from the base of the umbrella cloud blanketing the landscape below. (Reproduced from Sparks *et al.*, 1997 by permission of John Wiley & Sons)

nicely positioned to disperse aerosols into both hemispheres and so have far greater potential to affect global climate.

In recent years, eruptions such as the 1982 El Chichón and 1991 Pinatubo eruptions have had small but measurable effects on global climate. Sulfuric acid aerosols from these eruptions encircled the Earth in less than three months, reducing mid-latitude Northern hemisphere temperatures by 0.3 0.5 °C (Figure 5).

The overall tendency of upper-atmospheric winds is to spiral towards the poles where downward movement of air masses aids sedimentation into the ice caps. As a result, sulfate concentrations in polar ice cores provide a direct measure of the deposition of volcanically produced sulfuric acid. Ice layers showing high sulfate concentrations correlate with well documented recent eruptions, and such acid spikes can be used to identify probable climate forcing volcanic events in early historic and prehistoric times. Identified are the eruptions of Taupo (New Zealand) in 181, Huaynaputina (Peru) in 1600, Tambora (Indonesia) in 1815, Krakatau (Indonesia) in 1883 and Gunung Agung (Indonesia) in 1963, amongst many others.

Historical accounts of some of these events describe volcanic winters, characterized by deteriorating climatic conditions and markedly cooler summer temperatures for up

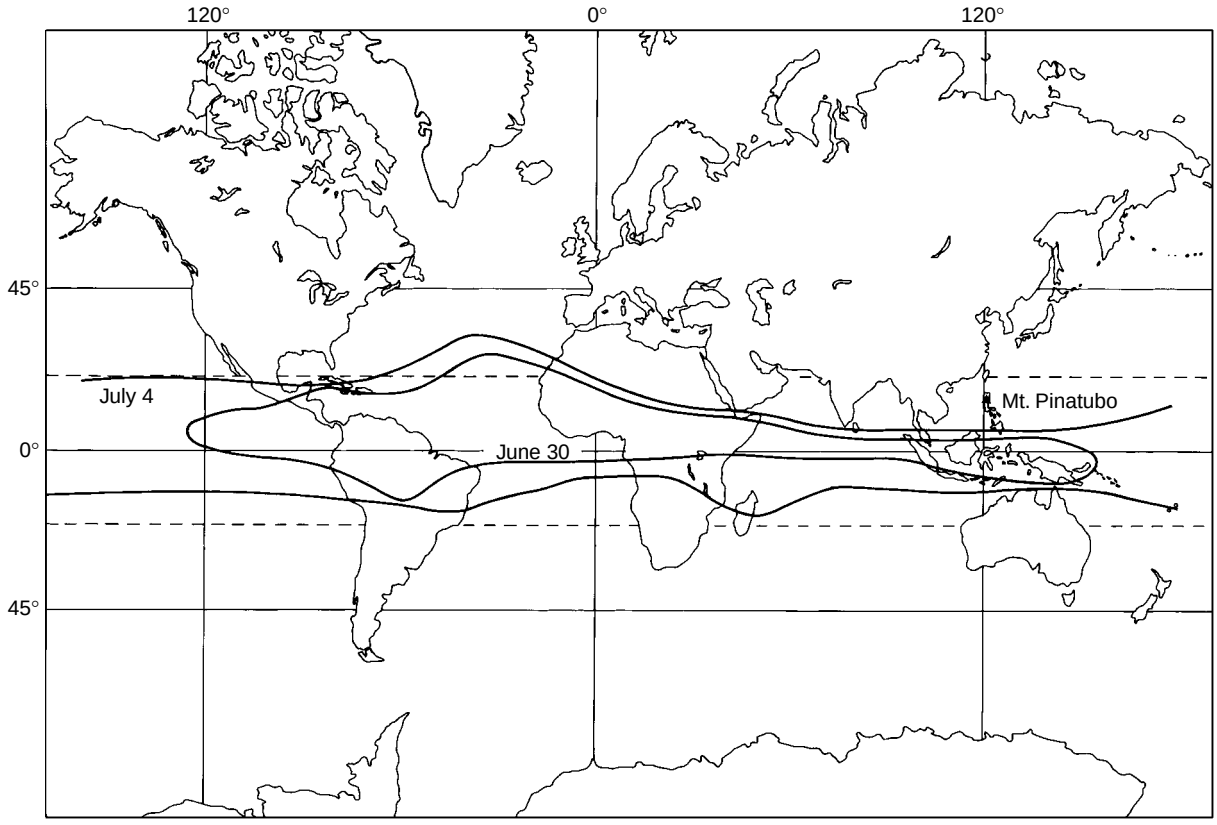


Figure 5 Stratospheric distribution of the aerosol cloud formed following the climatic June 15, 1991 eruption of Mt Pinatubo. The aerosol circled the Earth in 21 days and spread between 30 °N and 10 °S, causing measurable cooling (0.3–0.5 °C) of the troposphere. The distribution of the cloud was mapped by a TOMS (total ozone mapping spectrometer) carried aboard a satellite. (Reproduced with permission from Self *et al.*, 1995. Data courtesy of Gregg Bluth and Arlin Krueger, NASA Goddard Space Flight Center)

to 3 years following the eruption. The effects of a volcanic winter can be devastating, resulting in widespread crop and livestock losses, famine and deaths of thousands of people.

Some eruptions produce acid spikes at both poles, indicating widespread circulation of sulfur-rich aerosols. These eruptions are assumed to be of global importance and may be intimately linked with major climatic global perturbations. One such eruption occurred 74 000 years ago at Mt Toba in Indonesia, during a period of climate change associated with the last glacial cycle. The eruption produced a stratospheric sulfuric acid loading in excess of one billion tons resulting in an estimated 3 5 °C decrease in global surface temperatures for up to three years, but more importantly, a lowering of summer temperatures of • 10 °C in areas adjacent to those of established ice-cover. This temperature change would have been sufficient to tip the balance toward annual snow accumulation and the establishment of a snow albedo feedback mechanism. Dramatic increases in the planetary surface albedo would thus have triggered rapid glaciation and an acceleration of the global cooling already in progress.

The Toba eruption may itself have been triggered by a lowering of sea levels associated with the onset of glaciation. Throughout the Quaternary (the last 2 million years), periods of increased global volcanic activity, especially at volcanoes along coastlines, appear coincident with periods of rapidly fluctuating sea levels. During low sea levels, hydrostatic pressures acting to confine magma at depth are reduced, so allowing magma to ascend through new and reactivated crustal fractures caused by increased crustal tension. During periods of high sea level, the undermining of a volcanic edifice and periodic flank collapses may trigger decompression-related explosive eruptions. Indeed, studies at Mt Etna, on the coast of Sicily, suggest that past explosive eruptions are linked to rapidly changing sea levels in the Mediterranean area. Given that more than 90% of all active volcanoes are located within 20 km of the sea, the predicted rise in global sea levels attributable to global warming may have important implications for future eruptive activity, both in terms of eruption periodicity and impact on global climate.

CRUCIBLES OF CREATION

The scale of devastation wrought by volcanic eruptions in historic times has led us to view volcanoes primarily as agents of destruction and to overlook the immense tangible benefits that volcanism has presented to mankind, from extremely fertile agricultural soils to a source of renewable energy. These benefits derive from geologic processes that have acted over often very long periods of time measurable on geologic timescales and rarely within an individual s lifetime. Subtle as they are, they drive the economies of many developing nations, helping to sustain their typically dense populations. They may also be a source of additional wealth for many industrialized nations.

The most important of these benefits is the development and rejuvenation of soils. Of all the world s soils, those found in volcanic regions are probably the most fertile and have the highest potential for agricultural production, owing to the special nature of the parent material from which they develop. Termed *andisols*, these soils develop in tephra or pyroclastics, that is, from ejecta formed by the explosive disruption of magma. Pyroclasts are composed of minerals that crystallized within the magma and volcanic glass formed from the residual non-crystalline melt. They are classified according to their grain size, with material finer than 2 mm called *ash*. The constituent minerals and glass contain an abundance of elements such as potassium, magnesium and calcium which are essential plant nutrients, allowing plants to thrive. These nutrients become available to plants after many tens to hundreds of years of weathering in typically warm and wet tropical climates. In addition to their natural fertility, andisols also have excellent textural and structural characteristics providing for good natural moisture retention, drainage, and ease of tilling. Light dustings of ash from intermittent eruptions provide a means of rejuvenating heavily utilized soils or soils leached by monsoon rains by acting as a potent slow-release fertilizer. Thus, it is no accident that the densest populations of the world are found living on and around the slopes of active and dormant volcanoes. These are the rice bowl areas of developing nations such as Indonesia and the Philippines, which also support the world s best tea, coffee and sugar cane plantations.

Additional and economically important benefits include geothermal energy and ore deposits. Geothermal systems are found in areas of high regional heat flow where cooling intrusive magma bodies, at depths anywhere between several hundreds of meters to several kilometers, transfer heat and magmatic gases to subsurface groundwater reservoirs, hence their intimate association with young volcanic regions and the boundaries of Earth s tectonic plates.

The oil crisis in the 1970s and the subsequent concern over the finite nature of fossil fuel reserves saw great interest in the potential of geothermal systems as energy sources. Historically, those countries that successfully harnessed geothermal energy, principally for electrical energy generation, were wealthy and industrially developed. Italy s Larderello field began producing geothermal power in 1904, with plants in New Zealand, the USA, Japan and Iceland established shortly after. Developments in the 1980s and 1990s, however, have seen rapidly expanding capacity in the Ring of Fire countries and particularly within the Philippines and Indonesia. The Philippines is now the world s second largest producer next to the USA. Geothermal production is expected to continue in the Philippines to help meet industrial needs, part of an overall strategy to achieve

recognition as an industrialized nation. Today around 20 countries have significant contributions to their electrical supply from geothermal energy, and total global geothermal capacity is around 8000 MWe.

Iceland, situated above one of the world s hot spots and astride the mid-Atlantic oceanic ridge, is in an enviable situation of having potential geothermal reserves well in excess of the population s requirements. The very existence of this island as a developed nation is irrevocably entwined with the diverse and innovative development of its geothermal resources in overcoming the restrictions imposed by the harsh sub-arctic climate. Hot igneous intrusions beneath the 200 or more Icelandic volcanoes support numerous large high-temperature geothermal systems. Harnessing of geothermal energy began at Reykjavik in the 1930s, utilizing hot-water reservoirs for central heating of homes. Subsequent developments utilize geothermal waters and steam for electricity generation in concert with hydroelectric plants, for a range of commercial and industrial applications such as the heating of greenhouses and the drying of diatomaceous lake sediments used to manufacture high-quality industrial filters.

However, as with all energy sources, development of geothermal energy offers both advantages and disadvantages. Obvious advantages over the use of fossil fuels are the renewable nature of the reserves, and the lack of greenhouse gas emissions and consequent contribution to global warming. Disadvantages include resource management problems, potentially serious environmental pollution, and technological limitations to the exploitation of many geothermal fields.

Where geothermal reserves are extracted beyond their replenishment rates there is water depletion and pressure reduction. Excessive removal of geothermal water from the reserve can also result in surface subsidence. Water under the high temperatures and pressures associated with geothermal systems is capable of dissolving large amounts of solutes such as calcium carbonate and silica from the country rock[•]. On release of pressure and cooling, much of this solute can precipitate from solution. This effect is seen in areas where superheated waters vent naturally at the surface as hot springs or geysers and where spectacular formations such as the travertine terraces (composed of finely crystalline calcium carbonate) at Mammoth Hot Springs in Yellowstone may form. For the geothermal engineer, the formation of choking calcite and silica precipitates in pipes is a constant problem. Further, the residual brine left after the use of geothermal water and steam in electricity production is often a rich cocktail containing compounds of arsenic, mercury and boron which when discharged into surface waters can be a serious pollutant.

[•] *Country rock* is a general term for older rock (either sedimentary, metamorphic, or igneous) which has been intruded by magma or penetrated by mineral veins (Lapidus, 1987)

The future of geothermal energy development may not rely solely on superheated steam and pressurized hot water reservoirs presently harnessed, but on geopressurized zones and hot dry rocks. Geopressurized zones are areas where heat, water and methane are trapped under abnormally high pressures beneath thick layers of sedimentary rock, at depths of 3 km or more. These great depths prevent exploitation with current technologies. Hot dry rocks are areas where heated rocks exist near the surface, but where there are no associated reservoirs of water to carry away this heat. Experiments in harnessing this heat through the injection of cold surface waters and the recovery of the heated return water, though technically difficult, look promising. The potential energy resources of these reservoirs is enormous, but for the moment it is the dry steam reservoirs such as occur in the USA at the Geysers in California that are the most economically viable.

High-energy geothermal systems generated by volcanoes are among the most efficient natural concentrators of valuable ore minerals because factors critical for the creation of high temperature geothermal systems also favor formation of mineral deposits. Active volcanic systems are almost always too hot to mine and so ores are largely extracted from ancient or extinct volcanic complexes. Geothermal systems have the potential to deposit commercial concentrations of many elements including gold, silver, copper, lead, zinc, and sulfur. To illustrate the extraction of volcanic ores, we can consider the mining of sulfur from a chain of volcanoes along the crest of the Altiplano in the central Andes of South America. Here, economic deposits of particulate sulfur are found disseminated within the country rock and deposited around fumaroles. Sulfur extracted from no less than one hundred of the highest worked mines in the world (at around 6000 m above sea level) contributes to the global production of some 60 million tons per year which is destined for use in the manufacture of sulfuric acid, plastics, detergents, fertilizer, vulcanized rubber, and a host of other products.

See also: Volcanic Eruption, El Chichon, Volume 1; Volcanic Eruption, Mt. Pinatubo, Volume 1; Volcanic Eruptions, Volume 1; Volcanic Eruptions: Mt Merapi, Indonesia, Volume 3.

DEFINITIONS

Pyroclasts are particles of volcanic origin expelled through volcanic vents. They are formed by fragmentation of magma and country rock surrounding the vent. They are classified according to their grain size as either blocks or bombs, lapilli or ash (Sparks *et al.*, 1997).

Tephra is synonymous with pyroclast and the terms tephra-fall and tephra-flow are used to indicate their mode of transport and deposition. Tephra-fall is used for all pyroclasts transported within and deposited from an eruption

column or cloud and tephra-flow is used for pyroclasts transported and deposited by pyroclastic flows (Sparks *et al.*, 1997).

Pyroclastic flows are hot, high-concentration, gas pyroclast dispersions that flow down the slopes of volcanoes. Their origins are from collapse of an eruption column that is unable to rise buoyantly (Sparks *et al.*, 1997).

A *lahar* is a rapidly flowing mixture of rock debris and water (other than stream flow) from a volcano. The term is restricted to sediment water flows of volcanic origin. In non-volcanic environments the term debris-flow is generally used (Smith and Fritz, 1989) (*see* Lahars, Volume 3).

REFERENCES

Janda, R J, Daag, A S, Delos Reyes, P J, Newhall, C G, Pierson, T C, Punongbayan, R S, Rodolfo, K S, Solidum, R U, and Umbal, J V (1995) Assessment and Response to Lahar Hazard Around Mount Pinatubo, in *Fire and Mud*, eds G Newhall and R Punongbayan, University of Washington Press, Washington, 107 140.

Lapidus, D F (1987) *Collins Dictionary of Geology*, Harper Collins, UK.

Paladio-Melosantos, L O, Solidum, R U, Scott, W E, Quiambao, R B, Umbal, J V, Rosolfo, K S, Tubianosa, B S, Delos Reyes, P J, Alonso, R A, and Ruelo, H B (1995) Tephra Falls of the 1991 Eruptions of Mount Pinatubo, in *Fire and Mud*, eds G Newhall and R Punongbayan, University of Washington Press, Washington, 513 536.

Self, S, Zhao, J X, Holasek, R E, Torres, R C, and King, A J (1995) The Atmospheric Impact of the 1991 Mount Pinatubo Eruption, in *Fire and Mud*, eds G Newhall and R Punongbayan, University of Washington Press, Washington, 1089 1115.

Smith, G A and Fritz, W J (1989) Volcanic Influences on Terrestrial Sedimentation, *Geology*, 17(4), 375 376.

Sparks, R S J, Bursik, M I, Carey, S N, Gilbert, J S, Glaze, L S, Sigurdsson, H, and Woods, A W (1997) *Volcanic Plumes*, John Wiley & Sons, Chichester.

Umbal, J V and Rodolfo, K S (1995) The 1991 Lahars of Southwestern Mount Pinatubo and Evolution of the Lahar-Dammed Mapanuepe Lake, in *Fire and Mud*, eds G Newhall and R Punongbayan, University of Washington Press, Washington, 951 970.

FURTHER READING

Fisher, R V, Heiken, G, and Hulen, J B (1997) *Volcanoes*, Princeton University Press, New Jersey.

Fishpool, I (1997) Living with a Volcano, *Geogr. Rev.*, 11, 2 12.

Hammer, C U, Clausen, H B, and Langway, C C (1997) 50 000 Years of Recorded Global Volcanism, *Clim. Change*, 35, 1 15.

Horrocks, J, (1998) Death and Destruction on Montserrat, *Geogr. Rev.*, March, 18 22.

http://www.buss.co.uk/geosci/igahome.html.

Koening, J B (1973) Worldwide Status of Geothermal Resources Development, in *Geothermal Energy*, eds P Kruger and C Otte, Stanford University Press, CA, 15 58.

McGuire, W J (1992) Changing Sea Levels and Erupting Volcanoes: Cause and Effect? *Geol. Today*, July–August, 141 144.

Oppenheimer, C (1993) Mines in the Sky, *Geol. Today*, March–April, 66 68.

Pierson, T C, Janda, R J, Umbal, J V, and Adage, A S (1992) *Immediate and Long-term Hazards from Lahars and Excess Sedimentation in Rivers Draining Mount Pinatubo*, Philippines, US Geological Survey Water-Resources Investigations Report, 92 4039.

Rampino, M R and Self, S (1993) Climate-volcanism Feedback and the Toba Eruption of • 74 000 years Ago, *Quatern. Res.*, 40, 269 280.

Shoji, S, Nanzyo, M, and Dahlgren, R A (1993) Volcanic Ash Soils: Genesis, Properties and Utilization, *Dev. Soil Sci.*, 21.

Tiglao, R (1996) March of the Sludge, *Far Eastern Econ. Rev.*, February 1, 26 27.

Weaver, K F (1977) Geothermal Energy: the Power of Letting Off Steam, *Natl. Geogr.*, October, 566 579.

Weintraub, B (1982) Fire and Ash, Darkness at Noon, *Natl. Geogr.*, November, 660 684.

Zielinski, G A, Mayewski, P A, Meeker, L D, Whitlow, S, Twickler, M S, Morrison, M, Meese, D A, Gow, A J, and Alley, R B (1994) Record of Volcanism since 7000 BC from the GISP2 Greenland Ice Core and Implications for the Volcano-climate System, *Science*, 264, 948 952.

W

Wadi

Wadi: an Arabic word for a watercourse that is dry for much of the time but which is subject to flash floods, sometimes catastrophic. Within the United Nations Educational, Scientific and Cultural Organization International Hydrological Program, there is a wadi hydrology network (Salih, 1999). The objectives of the network are:

- to enhance the knowledge and understanding of wadi systems;
- to develop the human and institutional capacities needed for the sustainable management of wadi systems;
- to develop the human and institutional capacities needed for the sustainable management of wadis and their water resources in arid and semi-arid areas.

REFERENCE

Salih, A (1999) Wadi Hydrology Regional Wadi Hydrology Network of the International Hydrological Program (IHP), *Waterway*, 17, 5 7.

R E MUNN *Canada*

Waldsterben

see Nitrogen Deposition on Forests (**Volume 2**); Waldsterben (**Volume 5**)

Waste Dumps in Megacities: Case Study of Istanbul

Günay Kocasoy
Boğaziçi University, Istanbul, Turkey

Istanbul, with a population of 14 million, is located on two continents, Europe and Asia, acting as a bridge between them. Every day 6000 tons of solid waste is generated on the European side and 3000 tons on the Asian side. Wastes of Istanbul were dumped into the sea until 1953. Then the disposal of the solid wastes into open dumps became a common practice until the publication of the Solid Waste Regulation at 1992. Since then, the Municipality of the Metropolitan City of Istanbul has started construction of transfer stations and sanitary land lls. Currently there are six transfer stations and two sanitary land lls in Istanbul. Solid wastes are collected and transported to transfer stations and emptied into silos by the local district municipalities. After their volume has been reduced by 20%, they are transported in silos to the nal disposal areas and buried by the Municipality of the Metropolitan City of Istanbul.

Industrial solid waste is collected and transported to the nal disposal areas by industry itself. Medical wastes are handled in a special way and collected and transported separately to the medical waste incinerator of the Municipality. Construction of solid waste recovery and compost plants is included in the programme of the Municipality for the coming years.

One of the most significant problems of a city is the management of the solid waste generated. Handling the solid waste properly is very significant especially for cities like Istanbul, whose population is continuously increasing because of the migration from rural areas and other cities. Today, the population of Istanbul has reached 14 million and it is expected to reach 17 million by the year 2020. As a result of the increase in the population, 9000 tons of municipal solid waste is generated per day.

Istanbul, situated on two continents, is a bridge between Europe and Asia. Six thousand and 3000 tons of solid waste are generated on the European and Asian sides of Istanbul, respectively. The residential areas, hotels, offices, commercial areas, street cleanings, parks and recreational areas are the sources of solid wastes generated. The annual waste generation estimates for the future are given in Table 1.

According to Table 1, the solid waste production *per capita* is $0.66\,\mathrm{kg\,day^{-1}}$ on the European side while it is $0.56\,\mathrm{kg\,day^{-1}}$ on the Asian side; the average for the whole city is $0.63\,\mathrm{kg\,day^{-1}}$.

The physical and chemical properties of Istanbul s solid waste change from region to region and also from season to season. Therefore, the composition of the solid waste generated in Istanbul is estimated by dividing the city into five regions according to the socio-economic level of the inhabitants and the type of fuel used for heating in the district.

1. Commercial districts.
2. Low socio-economic level, shanty houses, using coal or wood for heating. Thirty percent of the population live in this region.
3. High socio-economical level, using solid fuel. Twenty percent of the population live in this region.
4. High socio-economical level, using liquid fuel. Twenty percent of the population are living in this region.
5. Normal average socio-economical level, using solid fuel. Thirty percent of the population live in this region.

Table 1 Annual solid waste production in Istanbul (millions of tonnes) (Curi *et al.*, 1998)

	Years			
Location	1990	2000	2010	2020
European	1.11	1.71	2.34	3.01
Asian	0.55	0.97	1.35	1.74
Total	1.66	2.68	3.69	4.75

Composition of the solid waste by year, in percentage, is given in Table 2. These percentages will change because of the increase in the use of natural gas and low-ash coal. As a result, the ash content is expected to decrease. Also, although the amount of organic materials is increasing, the percentage of it decreases because of the increase in the total amount of solid waste.

The most common method of disposal of solid waste used to be open dumps. Istanbul had five open dumping areas, named Kemerburgaz, Halkalı, Umraniye, Aydınlı and Yakacık. All of them, except Kemerburgaz and Halkalı, were legally closed but rehabilitation activities are still going on. These were the places where solid waste was simply dumped and left as it was. Wastes were neither compacted nor covered. Also, the dumping areas did not have any leachate or gas collection systems. Therefore, they were potential sources of very serious environmental problems.

With the publication of the Solid Waste Control Regulation in 1992, the Metropolitan Municipality started to construct transfer stations and sanitary landfills according to international standards. Six transfer stations and two sanitary landfills have been constructed since 1995.

The transfer stations at Halkalı, Baruthane and Yeni Bosna with 53, 65 and 50 silos, respectively, and a total capacity of $7500\,\mathrm{tons\,day^{-1}}$ are located on the Asian side. The other three transfer stations, Aydınlı, Hekimbası and Kucuk Bakkalkoy, having 55, 69 and 45 silos and a total capacity of $7500\,\mathrm{tons\,day^{-1}}$ are on the European side. The volumes of the silos are $32\,\mathrm{m^3}$. The distances from the transfer stations to the sanitary landfills are 38 51 km. Collected solid wastes are brought by truck to the transfer stations, emptied into the vertical silos and pressed, reducing the volume by 20%. Then the silos are carried to the final removal areas.

There have been two sanitary landfills in operation in Istanbul since 1995. One, called Odayeri, is located on the European side and the other one, named Komurcuada is on the Asian side. Odayeri sanitary landfill has an available area of 14 hectare (ha), but it has a projected area of 2000 ha for future developments. It has the capacity

Table 2 Change in solid waste composition according to year (%) (Curi, 1990; Kocasoy, 1996)

Content	1980	1981	1983	1984	1985	1986	1987	1992	1996
Organic material	60.8	60.60	59.00	51.8	43.25	42.70	44.62	45.00	48.00
Paper	10.15	18.80	13.12	10.73	11.72	10.20	11.92	14.50	8.40
Textile	3.22	31.00	1.40	3.21	4.87	3.10	4.28	5.60	2.90
Plastic	3.05	3.10	4.04	4.83	7.13	6.22	11.87	9.50	11.00
Glass	0.65	0.70	0.55	1.36	1.71	3.61	3.42	3.80	4.60
Metal	1.43	1.50	6.49	4.82	4.34	4.56	2.28	2.20	2.30
Ash	16.16	14.60	10.04	18.45	16.84	20.15	14.12	15.00	13.20
Miscellaneous	4.54	6.90	5.00	4.79	10.14	9.46	7.43	4.40	6.30

to store 4500 tons of solid waste per day. The area of the other sanitary landfill, Komurcuada, is 6 ha, but an additional 50 ha is reserved for the future. The target is 198 ha for the next 25 years. The base of the landfill areas was made impermeable (10^{-8}) by a layer of clay 60 cm thick and a high density polyethylene (HDPE) membrane 2 mm thick. Leachate collection pipes were installed above the impermeable layer and the collected leachate is transferred to treatment plants. Also gas collection chimneys were placed every 50 m. It is planned to convert the collected gas into electric energy in the future.

The local district municipalities are responsible for the collection and transportation of all municipality solid wastes to the transfer stations while the Metropolitan Municipality is responsible only for the collection of solid wastes from the main streets, public parks and then the management of all solid waste after it has been brought to the transfer stations. The Metropolitan Municipality is also responsible for the construction and operation of the final disposal areas.

The Municipality does not collect industrial solid waste. It is the responsibility of the industry itself to collect and transport domestic solid waste to the disposal sites and hazardous waste to the incineration plant located at Kocaeli. Health-care waste is collected and transferred to the incineration plant separately by the Municipality.

Silos of compacted wastes are transported from the transfer stations to the sanitary landfills by trucks and emptied onto specially built platforms; from there they are placed in prepared cells by bulldozers and pressed by compressors. At the end of the daily storage operation, the cells are covered with soil. Every day 4500 tons of waste is brought to Odayeri and 3000 tons of solid waste is brought to the Komurcuada sanitary landfill. The rest of the solid waste is dumped in open dumps.

There are no effective recycling and recovery programs in Instanbul although one recovery plant operates as a pilot plant. Construction of compost plants is also within the program of the Municipality.

REFERENCES

Curi, K (1990) Katı Atıkların Toplanması ve Uzaklastırılması, *Belediyeler ve Cevre Sorunlar* , Katı Atık Turk Milli Komitesi, Istanbul.

Curi, K, Ekinci, E, and Kocasoy, G (1998) Ulusal cevre Eylem Planı, Planning State of Turkey, Ankara.

Kocasoy, G (1996) Solid Waste Management Policy for Turkey, *Proceedings of the Twelfth International Conference on Solid Waste Management*, University of Pennsylvania and Widener University, Philadelphia, PA.

Waste: the Global Mass and its Management

W S Fyfe
University of Western Ontario, London, Canada

There is no doubt that one of the major causes of recent global change is the increase in human population and the impact of the present life-support technologies for this population on the global environment. It is interesting to look at dictionary de nitions, of the word waste (vastos): to desolate, ravage, devastate, damage, ruinous, neglect, useless, spend uselessly, foolishly, squander, decay, exhaust, consume the livelihood.

We would not have six billion humans on the planet (soon eight to ten billion) without certain technological systems. These all involve waste production and waste management or mismanagement. The key components include:

1. *energy production;*
2. *water quality and quantity;*
3. *food related to soil quality and quantity, water and climate, biodiversity;*
4. *arti cial materials (steel, glass, concrete, etc.) all derived from the outer 5 km of the crust of the Earth by mining technologies;*
5. *the chemical industries, from plastics to pesticides, etc.;*
6. *the management of urban wastes; today most of the people of the world live in cities.*

ENERGY

There is no doubt that a major component determining the quality of life is the availability of energy. The present world is divided into those who waste energy, those who are trying to conserve and those who desperately need more energy. Thus in the US, energy consumption *per capita* is 317 gigajoules; Germany, 170, Japan, 141; China, 25; India, 10; Nigeria, 7 and Afghanistan, 1 (Fyfe, 1999).

Today, most energy is produced from fossil fuels (coal, oil, gas) which are not renewable on human time scales. Smaller sources of energy include nuclear energy, whose use has greatly increased over the past 30 years, hydroelectricity and solar energy (wind, waves, light), whose use is increasing (Table 1). All energy technologies are associated with the production of wastes.

First, with fossil carbon fuels the dominant waste is carbon dioxide. As the famous Swedish chemist Arrhenius stressed in 1896, if we continue to increase the carbon dioxide concentration of the atmosphere, the planet will

Table 1 World Energy Consumption (*Source:* Fyfe, 1999)

	Petajoules (1993)	Percent increase since 1973
Solid carbon fuels	92 000	36
Liquid carbon fuels	134 000	11
Gas carbon fuels	78 146	72
Geothermal and wind	1463	641
Hydro	8554	86
Nuclear	23 646	1365

become warmer, the greenhouse effect! Today, world data are clear. In the year 1800, atmospheric carbon dioxide was 280 parts per million (ppm) by volume, while today is over 350 ppm. Our energy industries have changed the air we breath and slowly, the impacts on human health are appearing (Haines and McMichael, 1999) (*see* Environmental Change and Human Health: Extending the Sustainability Agenda, Volume 3). *Per capita* emissions of carbon dioxide are 20.8 tonnes in the US, 10.2 Germany, 5.8 France, but if all people in the world produced carbon dioxide *per capita* as with North America, it would not be a greenhouse, it would be Venus!

But when one examines the chemistry of fossil carbon fuels, there are many other forms of atmospheric and environmental pollution. Thus atmospheric methane concentrations are increasing rapidly (*see* Methane: Industrial Sources, Volume 3). With high sulfur coals, which are common with up to 10% sulfur, acid rain is related to combustion. And some coals contain halogens at moderate concentrations, particularly, chlorine and fluorine, and during combustion acid gases and even organo-halogen components can be produced.

The chemical composition of coal, the major fuel of China and India, is highly variable. Coal can contain a host of trace elements like lead and arsenic. It is amazing that when a very large coal powered electric plant was recently built in Northern Thailand, the detailed chemistry of the coal had not been determined. Normally, engineers are interested in moisture, sulfur and ash only. We discovered that tones of the coal being used was very rich in arsenic, 1000 ppm. Arsenic pollution of soil and moisture is a huge world problem. A recent report by Pearce (2000) states: unless urgent action is taken, one in ten deaths in Southern Bangladesh could be due to arsenic poisoning . There are many sources of arsenic pollution but coal is one that must be carefully watched. It is interesting to note that India and China mine over 1.5 billion tonnes of coal per year. This implies moving something like 3 km^3 of rock per year!

The ash and waste produced from the coal industry is of the order of 1 km^3 per year. It is only recently that there have been detailed studies of the chemical composition of coal ash. Some coals are rich in uranium, and even gold. But more commonly, because of the bio-organic origin of coal, there may be significant concentrations of bio-essential elements (potassium, calcium, phosphorus, •••). Some coal ash can even be useful for soil remediation and as a fertilizer.

COMBUSTION GAS DISPOSAL

There is increasing interest in the possibility of controlling and using combustion gases (see Fyfe, 1999). And if combustion gases are pumped into appropriate rock formations, the acid rain problem can be reduced. The reactions also generate energy and there can be a bonus of warm water.

After fossil fuels, nuclear energy is the major producer of electrical energy. During the period 1985 1996 nuclear energy production increased 56%. In the 1940s and 1950s, most scientists saw a future for clean nuclear energy. Today, that view has changed. First, particularly since the Chernobyl disaster, one must be sure there are no accidents with nuclear power plants, an impossible guarantee.

Only recently have we become aware of the problems of nuclear waste management, isolation from the biosphere. And recent work has shown (see Pedersen, 1993) that complex communities of microorganisms are active to depths of several kilometers. The problem is urgent as stressed by Pease and Roberts (1996) who show that there is enough plutonium on the Earth s surface to make 200 000 nuclear bombs. Nuclear waste isolation is an urgent international responsibility (*see* Nuclear Waste: Geological Issues, Volume 3).

Is hazardous, toxic waste isolation from the biosphere possible? With nuclear waste, we are concerned with times of a few million years for decay of the most dangerous species. But there are wastes, as with arsenic, mercury and many other chemical elements and compounds, where there is no half-life. Fortunately geological and geochemical data show that long-term isolation is possible. We have many examples of materials formed more than a billion years ago, materials sensitive to corrosion and chemical transport, which are perfectly preserved. Such minerals include species like elemental copper, many types of metal sulfides, even massive deposits of oil, gas, and salt! But a problem with nuclear and other toxic wastes is that nations which have the nuclear reactors may not have the best host rocks for disposal. Also, recent work in Sweden has shown that events like an ice age can lead to disturbance of an underground site.

IS CLEAN ENERGY POSSIBLE?

Fortunately the answer to this critical question is positive. We have two vast energy resources on planet Earth. If we could collect and use about 12 hours of solar energy, it would provide all that is needed. Solar energy exists

in many forms, such as radiation, wind, tides and ocean currents (*see* Wave Power Development, Volume 3; Wind Power Development, Volume 3). The costs of wind, photovoltaic, energy are rapidly diminishing and the use expanding. As reported in the *New Scientist*, 16, October, 1999, Germany is planning the two largest wind energy systems in the world. These sites, farms, will produce power on the scale of 1000 MW (cf. nuclear plants), with minimal pollution.

The other great source of clean energy is from the thermal gradient in the rocks of our planet. In most situations, the temperature increases 30 °C per km depth. Near volcanic sites, rocks near the surface can be at temperatures over 1000 °C and are produced on and in the continents and on a massive scale in submarine environments. The rate of production of heat is sufficient to supply the energy needs in many nations. And as stressed by many, the land disturbance, and of wastes associated with solar and geothermal energy is minimal. It is interesting to note that as reported by the Worldwatch Institute (2000) the annual growth rate of wind power is 22%, solar electric, 16%, geothermal 4%, oil gas 2%, nuclear 1%, coal 0% in the period 1990 1998. There is hope for clean energy.

MATERIALS

Advanced societies use a very complex array of inorganic and organic materials. These include stone, concrete, brick, ceramics, glass, steel, copper, zinc, aluminum, plastics, paper and a huge array of chemicals. In advanced societies something like 20 tonnes of rock are used annually per person. The great demands are for construction purposes, buildings, roads, railways, etc. If all people in the world of 2050 were to use rock on this scale, it would approach a volume of 100 km^3 per year, more than all volcanism on land and ocean floors. Yes; Vernadsky (1986) was correct!

In our populated world the land surface becomes more precious and must not be wasted. Giant shallow, open cast, mining operations destroy the land surface. Almost all our inorganic resources come from the outer 5 km of the Earth s crust. Some very deep mines, mainly for gold, approach 5 km depth. It is interesting to look at world mining figures (World Resources Institute, 1999). For example, iron ore 554 million tonnes; aluminum ore 27 million tonnes; copper ore 10 million tonnes. An interesting example of mining and waste is provided by gold with total annual production over 2 million kg. In the rocks mined for gold, the normal concentration is about 3 g per tonne (3 ppm). Thus for the total production about 1 billion tonnes of rock is mined, ground to a powder and the gold extracted by mechanical methods or with chemicals like cyanide. Recently in one of our university classes, we weighed the gold rings of the young students. The average weight was 10 g so the student was wearing the product of processing about 3 tonnes of rock! (*see* Materials Flows for Mining and Quarrying, Volume 3).

Gold illustrates the general problem of quality control of all extraction processes. Gold ores are normally associated with iron sulfide minerals. And very frequently there is a moderate concentration of arsenic or lead. The wastes from gold mining and gold extraction can be highly toxic and environmental disasters occur frequently. For example in a recent report in *Harper's Magazine*, October, 1999, it is stated:

> Constrained by environmental and labor regulations at home, American, Canadian, Australian and European mining companies have turned increasingly to remote, developing countries, digging mines of unprecedented size and destructiveness in relative obscurity. Many of these projects have caused significant deforestation, forced migration, and the release of industrial waste (cyanide, arsenic, lead and copper) into nearby waterways, killing all life in several rivers

A recent study by Nriagu (1996) on the history of global metal pollution is alarming. Thus as metal production has increased so has atmospheric pollution which in turn leads to soil and water pollution. Thus for lead, annual production increased from about 5 million tonnes in 1850 to 30 million tonnes in 1990. Lead emission to the atmosphere changed from about 0.5 million tonnes to 4 million tonnes over the same period. In many regions today, children have high levels of lead in blood with serious health effects (World Resources Institute, 1999). However, in countries where only lead-free gasoline is sold, children s blood no longer shows elevated levels of lead.

The use of metals illustrates well the need for quality control at all stages in their production and the need for recycling. Birkett (2000) describes the new European Union proposal for massive reuse and recovery of all materials in automobiles. And these approaches are needed for all materials extracted from the Earth.

An elegant example of waste prevention is given by highway construction in Sweden. A new multilane highway was proposed to link major cities in Sweden. Normally, one would aim for the shortest distance. But when the geology, the nature of the surface (rocks, soil, water) was carefully studied, it was found that by using a slightly longer route, the costs were reduced by 50% and the environmental impact vastly reduced.

WASTE AND URBAN SYSTEMS

In 1950, the urban population of the world was about 1 billion, today about 3 billion and is projected to rise to 5 billion by 2030 (*see* Urban Population Change, Volume 3). In the world today 47% of people live in urban regions, in Europe 75%. North America, 77%, South America 80%, Africa 38%, Asia, 38% and all cities are projected to grow.

Urban wastes are massive and complex in their variety (*see* Urban Wastes, Volume 3). A typical example is provided by paper and paper products. Today, *per capita* use of paper is near 0.05 tonnes worldwide. The greatest waste of paper is in North America at a figure of 0.25 tonnes per person, compared with Europe 0.09 tonnes per person. Given quality control, paper can be recycled. In 1993 the lead nations were Austria and Spain with 78% recycling compare to the US at 34%. In many nations, paper products if carefully collected and separated can be recycled or even used as fuel for electricity production.

World data on total urban solid wastes per capita for 1993 (World Resources Institute, 1999) are impressive. For example, Lagos Nigeria, 0.3 kg per day, Paris, France 1.3 kg, Toronto, Canada 1.4 kg, New York, US 1.7 kg, Shanghai, China 0.8 kg, Dubai 2.3 kg. If one examines world data on urban wastes (see the *Economist*, 1999) the annual municipal wastes in the developed nations averages about 500 kg per person. If this was the average for a world with 10 billion humans, given that most wastes have a density near $1\,g\,cm^{-3}$, world production would be of the order of $5\,km^3$ per year, a figure comparable to the output of all annual continental volcanism.

But as well as typical urban domestic wastes, there are hazardous wastes and a variety of industrial wastes, which are produced on an even larger scale in industrial nations. These include wastes from the chemical industries, the metallurgical mining industries and the like. A dramatic example of industrial waste problems was the cyanide spills from Romanian mining.

A new initiative from the European Union will produce regulations that will require that in automobile manufacture, at least 85% of materials from used automobiles, including plastics, must be reused (Birkett, 2000)! This illustrates the general principle that quality control at the input determines the ease and economics of recycling the wastes.

It is interesting to examine the history of urban architecture and waste. Before electricity, before the large use of fossil fuels and before air conditioning, great attention was given to efficient construction. I was very impressed visiting some very old houses in Saudi Arabia during a very hot summer. The materials used were designed to reflect the sun, windows were placed to catch the night breezes. Today, with many of our massive modern buildings, such factors are rarely considered.

WATER

For many nations there is a growing crisis in the supply of clean water. Almost 40% of the average annual precipitation is manipulated by man. And after we use fresh water, its chemistry and biology are changed. The growth of urban populations has increased the problems and in many nations urban waste waters are discharged to the local rivers and oceans untreated. The world situation has been described by the World Resources Institute (1999) with particular attention to the situation in China. In most of Europe over 90% of wastewater is subject to some form of treatment. But for Lagos, Nigeria, only 2% is treated, Bombay 10%, Rio de Janeiro, 23%. We have recently been flooded by reports of the pollution on Copacabana Beach, Rio, with massive death of fish due to improper treatment and disposal. Normally the treatment involved some form of use of aerobic bacteria. With careful control, the sludges can be used for various compost, applications. Often, before the final fluid is discharged, low levels of chlorination are used (Baird, 1999).

SOIL

One of the most critical issues facing the present and future world is providing sufficient food of nutritional quality. As recently reported in *Nature* (Nov. 25, 1999, 341) to turn a blind eye to 40 000 people starving to death every day is a moral outrage . Many would say that 40 000 is a very conservative estimate. Food production in large part depends on soil quality, water and climate. As Brown and Wolf (1984) of the Worldwatch Institute wrote soil erosion: quiet crisis in the world economy (*see* Soil Deterioration and Loss of Topsoil, Volume 3). These authors show that over the past decades topsoil loss has approached 1% per year globally. We have shown in many places, that when soils are eroded by wind or water, the fine particulate clays are removed. These materials with large surface areas are vital to water retention and nutrient cycling between soil microorganisms and plants. Global satellite pictures dramatically show the patterns of soil erosion, an ultimate case of waste of a major component of our life support systems. Quality soil formation from rock materials takes times measured in 100s or 1000s of years. There are many examples of how wastes can be used for soil remediation (*see* Soil Amelioration, Volume 3).

See also: Anthropogenic Metabolism and Environmental Legacies, Volume 3; Contaminated Lands and Sediments: Chemical Time Bombs?, Volume 3; Fisheries: Minamata Disease, Volume 3; Geomorphological Change for Urbanization and Industry, Volume 3; Nuclear Waste: Geological Issues, Volume 3; Toxic Wastes: Generation and Disposal: a Case Study of UK Practice and Legislation, Volume 3; Urban Wastes, Volume 3; Waste Dumps in Megacities: Case Study of Istanbul, Volume 3; Wastes: Land ll and Land Raising, Volume 3.

REFERENCES

Baird, C (1999) *Environmental Chemistry*, W H Freeman, New York.

Birkett, D (2000) The Road to Full Recovery, *Chem. Britain*, Feb, 2000, 46 48.
Brown, L R and Wolf, E C (1984) Soil Erosion: Quiet Crisis in the World Economy, *Worldwatch Institute, Paper*, 60, 1 49.
Fyfe, W S (1999) Clean Energy for 10 Billion Humans in the 21st Century: Is it Possible? *Int. J. Coal Geol.*, 40, 85 90.
Haines, A and McMichael, A J (1999) *Climate Change and Human Health*, The Royal Society of London, London.
Nriagu, J O (1996) A History of Global Metal Pollution, *Science*, 272, 223 224.
Pearce, F (2000) Danger in Every Drop, *New Scientist*, Feb, 12, 16 17.
Pease, S and Roberts, L (1996) Bombs to Kilowatts: the Plutonium Problem, *Science and Public Affairs*, The Royal Society of London, London, Spring, 39 42.
Pedersen, K (1993) The Deep Subterranean Biosphere, *Earth Science Reviews*, 34, 243 260.
The Economist (1999) *Pocket World in Figures*, John Wiley & Sons, New York.
Vernadsky, V I (1986) *The Biosphere (Based on the French Edition of 1929)*, Synergetic Press, London.
Wilson, E O (1992) *The Diversity of Life*, Harvard University Press, MA.
World Resources Institute (1999) World Resources, 1998 1999, Oxford University Press, (this report is published every two years).
Worldwatch Institute (2000) *State of the World 2000*, W W Norton, NY.

FURTHER READING

Ernst, W G (2000) *Earth Systems: Processes and Issues*, Cambridge University Press, Cambridge.
Koshland, D E (1993) Clean Thoughts on Clean Air, *Science*, 261, 1371.

Wastes: Landfill and Land Raising

Peter Jones
Biffa Waste Services Ltd., Buckinghamshire, UK

Land ll is the oldest and most obvious waste activity associated in the human mind with the management of discarded materials. Historically this position was shared with reuse but the growing sophistication of materials and the complexity of products in the waste stream result in steadily increasing global tonnages to this exit route. This article explores aspects of this process in terms of the sustainability challenge. Speci cally, we discuss land ll morphology, land ll economics, in uences on compositional shifts (demand and supply side), and future challenges.

The role and signi cance of land ll management systems in different national frameworks is subject to a complex interplay of factors involving geology, climate, input waste stream policies and engineering expertise. The unpredictability of these constantly changing variables by de nition results in a counsel of caution with regard to the use of land ll as a strategic element in end-life waste management. The key strategic issue is whether, at local, regional or national level, it is easier to manage the unpredictable gaseous and liquid impacts of land lls or to apply equivalent energy and nancial resources to front-end segregation and specialist processing systems. At its heart is the question: are centralized aggregated processes simpler to manage and measure than geographically dispersed and segregated processes?

LANDFILL MORPHOLOGY

There will always be a need for some form of long-term storage facility for waste products whether in combination or segregated (Figure 1). Technology is readily available to reduce the mass of a wide variety of waste streams but governmental and corporate choice is restricted by straight economics. As long as landfill remains considerably cheaper than alternative process routes, the greater the volume that will be entombed in this fashion. Even where via a framework of fiscal and budgetary pressures or market forces those cost differentials are narrow, landfill entombment remains a best practicable option for a wide range of difficult substances, the mass or toxicity of which renders alternative routes unviable. Prominent examples include bottom ashes from waste incineration plants, certain types of foundry sands and contaminated soils (Figure 2).

Engineering practice in landfills has accelerated in response to the complexity of waste stream mixing introduced with sites accepting wet/dry, organic/inorganic and industrial/domestic streams representing the greatest challenge for containment engineers. The basic prerequisite introduced by regulatory pressures is that externalities and pollution arising from the landfill activity proper (in gaseous, liquid or disamenity (dust, odors and windblown pollution)) are reduced to an absolute minimum (Figure 3). In practical terms, in a sector where measurement systems are still represented as emergent technologies, the challenge for landfill operators is to benchmark individual site performance on these factors against industry-wide norms or expectations of zero performance standards. Against this background, national governments have thus fluctuated between outright bans (Denmark) to higher regulatory and monitoring standards (UK) to flexible input strategies coupled to significant liability and legal exposure (USA).

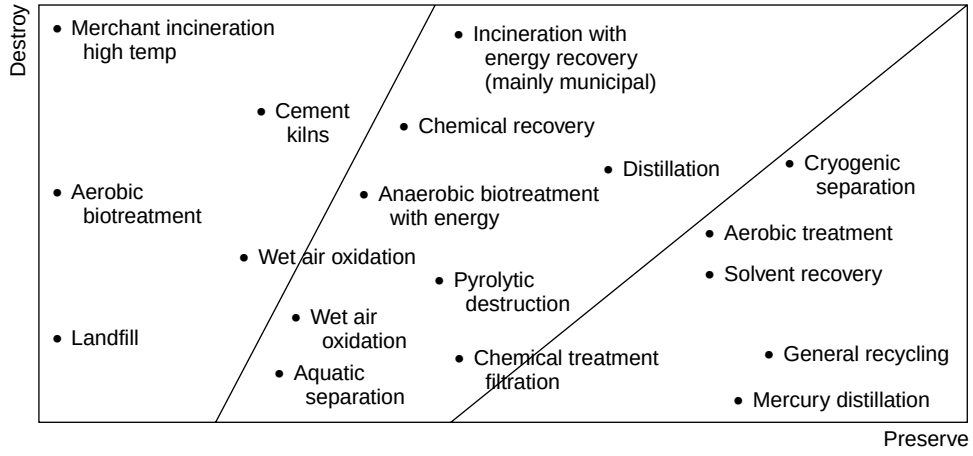


Figure 1 Waste options/processes

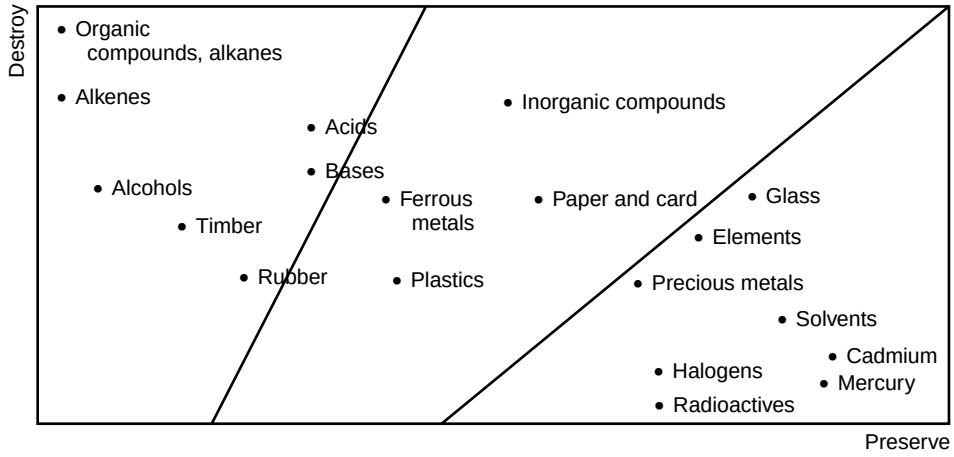
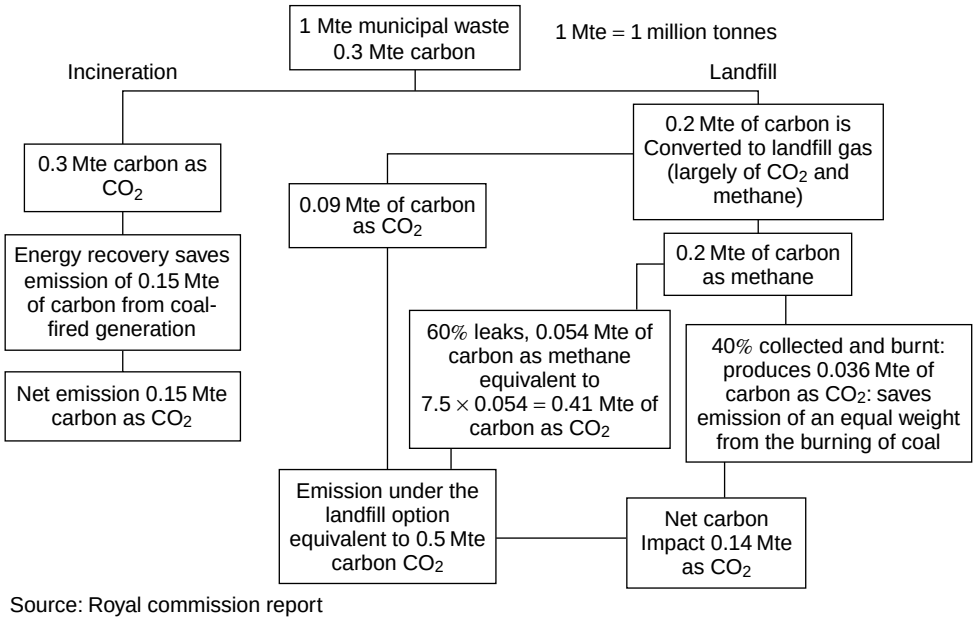


Figure 2 Waste options/products



Source: Royal commission report

Figure 3 Incineration versus landfill emissions of greenhouse gases

Each of these core policy frameworks produces a different emphasis in terms of engineering management and cost profiles.

Engineering Issues

Landfills are complex material processing mechanisms which operate within a fairly simple framework. In essence a chemical reaction is introduced by limited quantities of oxygen, moisture/fluids, organic materials, soluble metals and compounds and non-reactible materials. Over time scales varying from hours to decades, the result is a displacement of mass in this material into gaseous methane and carbon dioxide, liquids containing inorganic materials in solution, precipitates and stable, inert material. The reactions between each of these diverse materials do not occur in a consistent and controllable manner. They are driven by waste inputs related to material supply and demand in complex economic situations. Given the wide diversity of origin of wastes, coupled to the variable geology of their local position, it is feasible to produce a constant output stream only by careful attention to on-site engineering in respect to construction of working cells together with associated pipework and gas collection infrastructure.

Nevertheless, the broad range of combined outputs from landfill sites is often readily identifiable. Work has been undertaken on methane CO_2 compositional analysis and the factors influencing it as well as on similar factors affecting leachate composition and basicity. Mass balance evaluation of this nature for entire landfill sites is the most productive area of research. This is largely because, at the operational level, regulatory frameworks require the landfill site to be managed by phased development in units or cells. Such construction techniques are enforced by the need to concentrate releases of gas and liquids on a day by day basis throughout the working lifetime of the operation. Thus individual cells which often operate as mini-self-encapsulated landfills within the main body of the landfill are themselves subject to regular daily cover or capping by inert materials which further inhibit local disamenity on a day by day basis. The effect of such management practices, however, is to inhibit cross reactivity of materials within a site as well as the transboundary migration of gaseous and leachate liquid by-products in ways which are optimal to the operation of the site (Figure 4).

Much work continues to be done on the trade-offs between the avoidance of localized pollution during the construction phase of the landfill and maximizing the long-term productivity of the site for converting methane and other gases for power generation.

In turn this has triggered further debate on whether best engineering practice on landfills should encourage tight day to day containment (which often results in extensions to the active cycle of the site life) or accelerated biodegradation (which reduces whole-life activity cycles but may compromise localized disamenity during site operation). The choice between such options is often a function of location, geology and input streams; these points are evaluated in the section on future challenges.

LANDFILL ECONOMICS

As is common with most items of capital plant in any engineering application, one has to distinguish between operating and capital overhead costs. The assessment of landfill sites in financial accounting terms is often as complex as the variables involved in their physical operation. Historically, given an abundant supply of new landfills capable of meeting existing and future demands, it was perfectly acceptable in the UK for new landfills to be valued at their direct acquisition cost and depreciated on a straight-turn basis over their anticipated life cycle. In the UK, in common with other countries where the planning system allows public participation, the replacement of void on a like-for-like basis is often inhibited. In such circumstances the cost of replacement void in the UK trebled from £0.7 to £2 m^{-3} between 1990 and 1999.

Cost allocation strategies with regard to remediation and ongoing engineering thus vary between operators. This is a function of their approach in terms of the engineering process selected (see above) and the quality of leachate gas management retrieval systems. The most predictable cost factor relates to routine site operation in the form of compactor operation, *in situ* management and checking facilities. Nevertheless, additional variations creep in, given that different operators adopt different operating management systems in the form of ISO accreditation and other standards. Pricing strategy in such circumstances tends to be opportunist, marginal and designed to operate on the basis of charging what the market will bear (Figures 5 and 6).

For those landfills accepting multi-material inputs, pricing decisions may be even more erratic, often reflecting the degree of local monopoly control enjoyed by the operator. Where two or more operators exist within a 30 km radius, pricing tactics tend to reflect cash flow and profit expectations on a marginal basis. Where single operators have no immediate competition, gate fees tend to reflect the entry price of the next best available processing technology for specific material streams. Inputs of classes of material classified as hazardous or special, offer premium price options which contributes substantially to net margins. Volumes are restricted, however, in the UK by license conditions which prevent carryover of unused quotas from one week to the next (Figures 7 and 8).

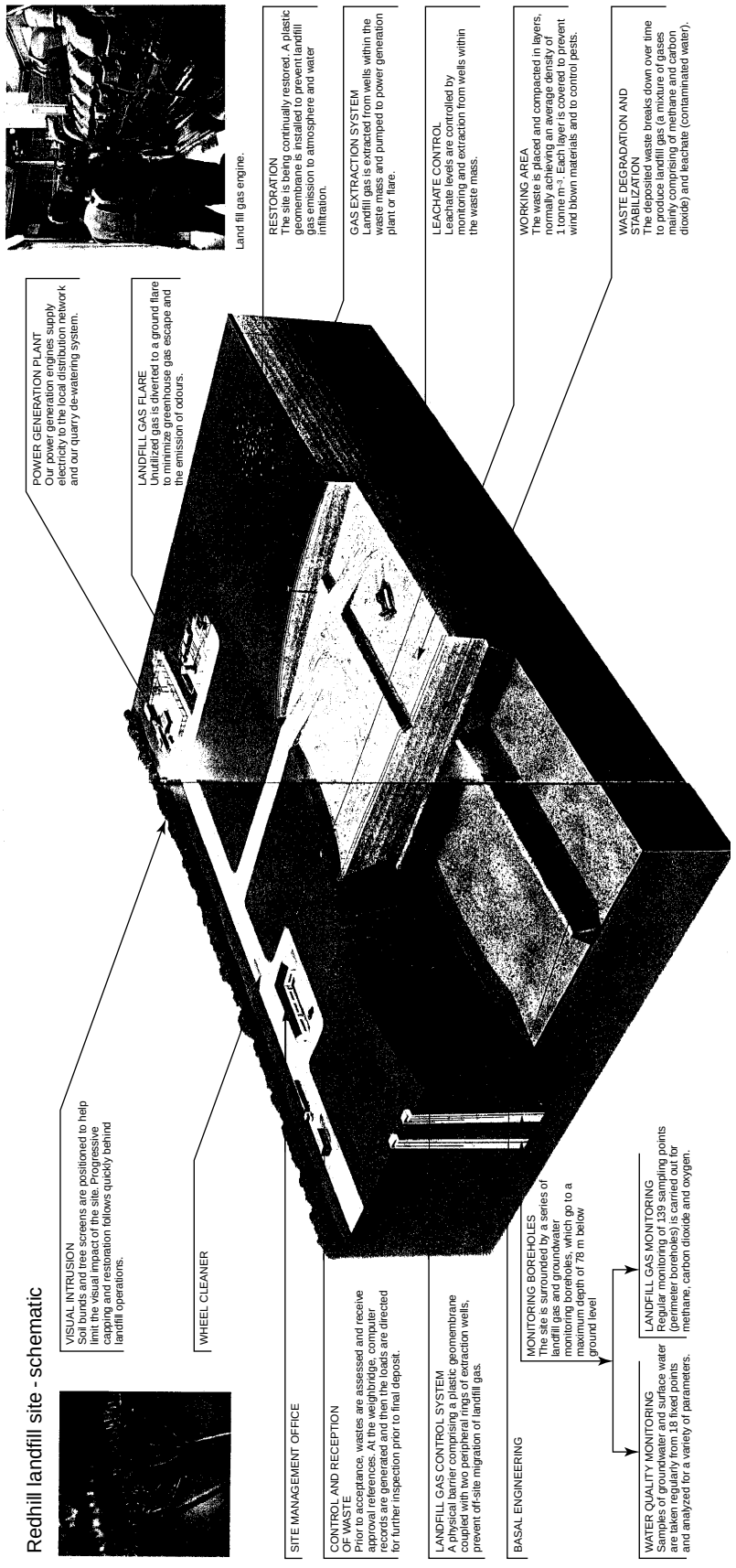


Figure 4 Cross section of a landfill site

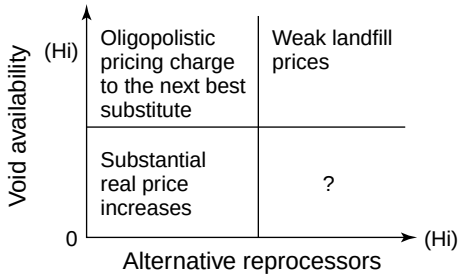
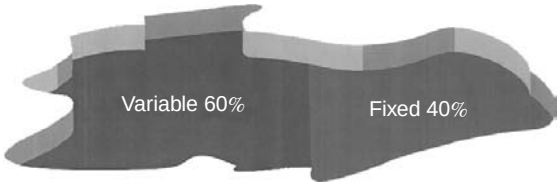


Figure 5 Local landfill pricing strategies

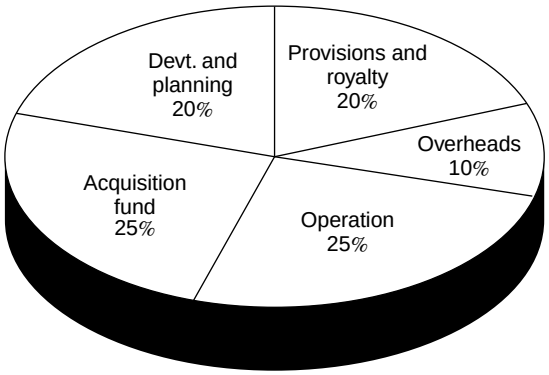
Landfill	8–20	Neg.
Tax	11–2	Neg.
Composting	20	Neg.
EFW	26	Neg.
Glass recycling	3	Neg.
Card/Paper	15	Pos.
Ferrous	0	Neutral
Plastic	0–30	Pos.

Figure 6 Gate fee structures (£ sterling) of the spring of 2001



Total operational = £8 m^3

Figure 7 Profile of containment landfill costs (2 mm^3 + void)



Tax exclusive income £12 per tonne

Figure 8 Landfill operation cost profile of the year 2000

INFLUENCES ON COMPOSITIONAL SHIFTS

Supply Side Factors

Supply side factors are those that influence variability in the composition of the waste coming to the landfill. As public consciousness is raised with regard to the management of end-life materials, such factors tend to be driven by governmental policy or strategy. The most significant of those which operate in Europe in 1999 are:

- Producer responsibility: whereby companies or industry sectors introducing particular products into the environment are responsible for the final disposal of these products at the end of their useful life. Such organizations may see landfill as a desirable objective but their thinking is often influenced by governmental policies involving targets or objectives which divert their material away from landfill. A Traded Pollution Permits framework designed to provide an incentive for choosing landfill diversion strategies may reinforce such tactics. In the UK landfill context, the most significant of these are the Packaging Directives relating to organic and inorganic inputs in the packaging supply chain.
- Landfill taxes are also a significant contribution to the process by which governments can raise the competitiveness of other forms of waste reprocessing. In the late 1990s, British waste stream charges for mixed waste arisings to landfill increased from around £5 $tonne^{-1}$ to £10^{-1}, yet even with landfill taxes of an additional £11^{-1} the incentive was still insufficient to encourage substantial diversion. Regarding customs data the moving annual total (MAT) tonnages to landfill have fallen 25% in 3 years to 70 million tonnes (latest data December 2000) (Figure 9). More important, the municipal contribution to overall total inputs has risen from 25% in January 1998 to 40%.
- Virgin input taxes such as on newly quarried aggregates may also encourage landfill diversion activity. Such mechanisms raise the cost of freshly mined material particularly aggregates and construction products thus encouraging the sorting and cleaning of materials such as construction and demolition waste which would otherwise have passed to landfill.
- Shifts in behavior of significant waste-generating sectors such as municipal and household arisings moving toward higher levels of source segregation in response to government-set targets or other factors can also impact on compositional profiles to landfill.
- Outright bans have an obvious and immediate effect depending on the time cycle over which they are phased. Examples in the UK include threats to the landfilling of liquid hazardous materials, tyres, clinical waste (incontinence pads), biodegradable materials and certain hazardous substances such as insecticides, pesticides and heavy metals. The key focus of EU

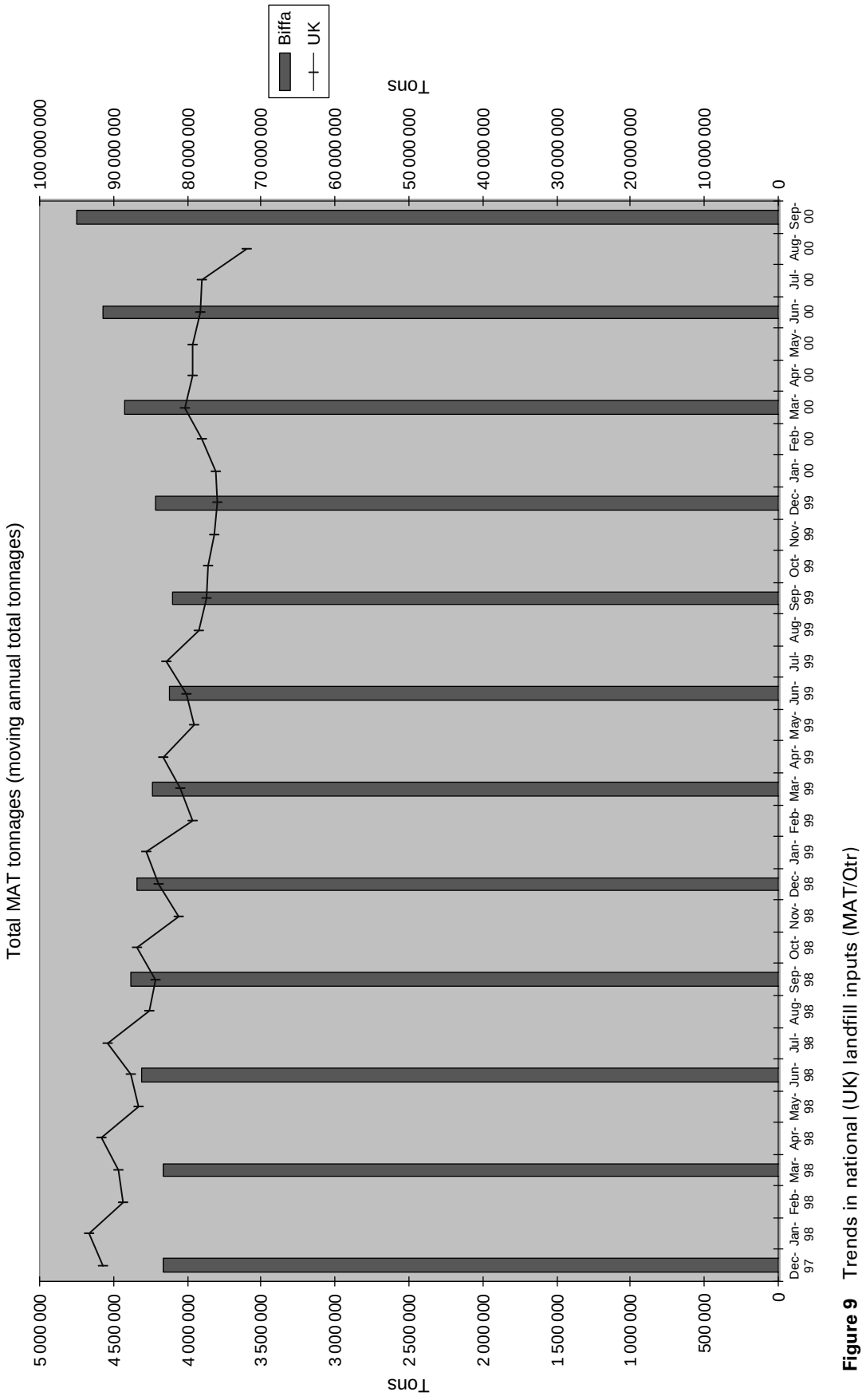


Figure 9 Trends in national (UK) landfill inputs (MAT/Qtr)

- By 2020 at latest, reduce biodegradable municipal waste (BMW) to landfill to 35% of total BMW in 1995. Intermediate targets of 75% by 2011 and 50% by 2013
- Ban on co-disposal
- Ban on tyres (by 2003 for whole tyres; by 2006 for shredded tyres)
- Ban on liquid wastes, infectious clinical waste, certain hazardous wastes by 2001

Figure 10 Landfill directive

strategy lies in reducing organic inputs to landfill with the consequence that the biodegradable activity in the landfill site is slowed and the active phase of the site thus extended. The most significant cumulative impact of this process is seen in the UK Landfill Directive which calls for a reduction of 65% in the tonnage of biodegradable municipal waste inputs by 2016 on 1995 levels (Figure 10).

FUTURE CHALLENGES

To Mono-dispose or Co-mingle

The persistent fundamental question on the true function of landfills remains as to whether they are ongoing chemical bioreactors with the specific purpose of neutralizing their content or whether they are long-term surrogate storage units for segregated, mono-material pollutant streams which can be held in the best practicable environmental medium until such time as innovative technologies permit them to be mined and exploited at some point in the future?

Dry Versus Wet Entombment and Management

As bioreactor units, landfill sites can biodegrade over an extended life cycle, major sites with organic materials often having a life expectancy in excess of 30 years. That life cycle can be reduced by absolute bans on all liquid inputs or pre-preparation of waste streams prior to burial. Conversely it can be accelerated significantly by the active introduction of liquid or aquatic waste streams along the concept of the flushing bioreactor. In this scenario, landfill sites are flushed on an accelerated basis of precipitates, heavy metal compounds and sulfides which may otherwise contribute to longer-term risk and management costs. The effect of the latter process is to foreshorten the active biodegrading phase of the landfill life cycle so that the site can be capped and sealed within a more manageable time frame.

Added Value Options

Different operators and sites follow different strategies in terms of income generation. Landfill gas may be engineered as a product of landfill sites to maximize income streams from power generation. Such initiatives often depend on innovative kick-start funding mechanisms from government to offer subsidies for energy production from renewable sources. One such has been the UK government s non-fossil fuel obligation that offered entry fees of 6.5 pence per kilowatt hour against a mean coal-fired power station baseline of 2.2 pence in the early 1990s. As the scheme expanded, economies of scale were identified and levels of knowledge on installation and exploitation increased significantly to the point that power generation costs from landfill gas now match those from gas-fired combined heat and power plants without the need for support subsidies.

Landfills as Material Management Operations

Given the range of uncertainties and the challenges to input streams which arise from a mix of external market and government-inspired regulations in different countries, many landfills are now becoming final resting places for materials. However, more importantly, many are now beginning to attach more significance to their role as material presorting and quality control centers. As taxes for disposal increase, landfill operators find it attractive to install primary processing capacity to sort specific elements of their captive waste stream. This is clearly profitable where average segregation costs are less than or equal to the prevailing level of tax. Their marketing and logistics infrastructure put them in a preferential position to attract such waste streams. In addition, both rail and road connections can add to their operational and economic flexibility. Such operations may also bring about a re-appraisal in the basic advantages of landfill as a last resort provided it meets scientifically quantifiable targets as a genuinely best practicable environmental option.

FURTHER READING

Biffa (2000) *A Question of Balance*, available from http://www.biffa.co.uk/publications/index.html, and references therein.

Wastes, Urban

see Urban Wastes (**Volume 3**)

Water Resources: Baltic

Erkki Leppäkoski
Åbo Akademi University, Turku, Finland

The Baltic is a brackish water epicontinental sea, a fjord-type super estuary common to about 250 rivers. It is non-tidal, small (415 000 km^2, water volume 21 700 km^3, i.e., covering 0.1% of the area and 0.002% of the volume of the world's seas) and shallow (maximum depth 459 m, mean depth 56 m; 17% of bottom areas • 10 m deep). It is long (1400 km from 54 °N to 66 °N), isolated from the North Sea by both geographical (sill depth 18 m, total opening circa 25 km) and ecological (coldness and less than one- fth of the salinity of the oceans) barriers. The annual in ow of salt water corresponds to about 2% of its total volume. Surplus water passes through the Danish sounds, resulting in a positive water balance. The residence time of water in the Baltic proper is estimated at 25- 40 years.

Seasonal variations are distinct, increasing in proportion to the distance from the entrance area. The northernmost parts are covered by ice 140 days year^{-1}.

The Baltic is a young, unstable sea. After the latest deglaciation, a freshwater lake was formed, followed by a short marine period lasting until about 9000 before present (BP). Gradual crustal uplift cut off the oceanic connection, and the Baltic changed into a fresh-water lake again. Around 7500 BP, it became brackish. Consequently, all animal and plant species living in the Baltic regions (including humans) are immigrants. Few survive in brackish water. Living close to their tolerance limits, even slight salinity and temperature changes may seriously affect distribution and dominance relationships. Increasing rainfall and river run-off have reduced salinity in recent decades. The ecosystem is thus increasingly fueled by waterborne inputs of organic matter produced outside the marine ecosystem. The oxygen periodically decreases to zero in the bottom water; poisonous hydrogen sulfide is formed, killing all bottom-living animals. These lifeless bottoms, Europe s largest deserts , periodically comprise 100 000 km^2 (25% of the total area). In the southern Baltic, postglacial Arctic relict species are threatened: their populations have poor resilience to catastrophic events like oxygen deficiency. Since the 1950s, Arctic and Arctic-boreal fauna have decreased. Atlantic-boreal and cosmopolitan species are now dominant.

Eutrophication is a major environmental concern in the Baltic region. The drainage basin (population 85 million) is shared by 14 industrialized countries. It is also subject to atmospheric fallout of nitrogen and harmful substances such as heavy metals. The initially oligotrophic (or slightly mesotrophic) Baltic ecosystem is becoming eutrophic. Inputs from multiple sources of nitrogen have increased about four times, and those of phosphorous, eight times. Most of this increase has taken place since 1955, particularly during the 1970s and early 1980s. Recently, the phosphorous inputs have decreased, but the concentrations continuously fluctuate at high levels.

During the 20th century, water transparency in the coastal and open Baltic decreased by 2 5 m. The macroalgal community, especially bladder-wrack (the main community-forming macroalga), has suffered from eutrophication (reduced light penetration, substrate changes, overgrowth by epiphytic algae, and increased grazing pressure). Perennial brown and red algae have been replaced by annual filamentous green and brown algae.

Massive phytoplankton blooms, mainly of cyanobacteria (some of which are toxic) are common in the Baltic proper; in late summer, one-sixth of the sea may be affected.

On deeper oxygen-poor bottoms devoid of bioturbating animals, laminated sediments with distinct winter and summer layers are formed. Some 25% of recent sediments in the Baltic proper are laminated. Since the 1960s, they have expanded over 2000 km^2 annually. In shallower areas, the biomass of soft bottom fauna has increased 3 5-fold, which is related to an almost 2-fold increase in the deposition of organic matter in sediments since the 1920s.

Since the early 1700s, the maximum annual extent of ice cover of the Baltic has decreased from an average of 240 190 $\times 10^2$ km with the ice in some rivers breaking up 15 days earlier. The expected sea level rise resulting from global warming will be compensated by the parallel land uplift that has been occuring since the last glaciation (maximum 0.9 m per 100 years) in the northern parts. However, in the south, rising waters will aggravate the inundation threats in low-lying coastal regions.

Three fish species (cod, sprat, and herring) account for 80 90% of the landings. Cod is the most sensitive to eutrophication requiring a salinity of • 11 parts per trillion (ppt), and an oxygen content of • 2 ml l^{-1} to breed. Increasing eutrophication increases the risks of oxygen deficiency in the saline near-bottom waters.

The following factors have caused annual fish catches to rise from 100 000 to 1 000 000 tons since the 1950s: intensified fishing, eutrophication and decreased numbers of seals (down from several hundred thousand in 1900 to 10 000 by the mid 1970s). Consequently, man has become the top consumer in the food webs of the Baltic ecosystem.

River-spawning stocks of Baltic salmon are at risk. They are found in less than 40 rivers instead of almost 100 rivers a century ago. Almost 90% of salmon stocks result

from farmed smolts implying serious genetic depletion risks.

The accidental introduction of non-native species spread with ballast water or via canals, or intentionally introduced into adjacent fresh water bodies, is an indicator of global change. Over 100 non-indigenous aquatic species have been recorded in the Baltic. They become common members of the fouling community in shallow waters, especially at river mouths and coastal inlets. Most of them originate from warmer areas. Global warming is expected to increase future introductions with devastating impact.

See also: Baltic Sea, Volume 4.

FURTHER READING

Ambio (1980) *The Baltic*, 9, 112 193.
Ambio (1988) *Ecotoxicology*, 17, 360 405.
Ambio (1990) *Eutrophication*, 19, 102 175.
Ambio (1992) *Seals and Protection of Seals*, 21, 494 606.
Ambio (1997) *Agriculture and Environment*, 27, 410 478.
Baltic Marine Environment Protection Commission Helsinki Commission (1996) Third Periodic Assessment of the State of the Marine Environment of the Baltic Sea, 1989 1993, Baltic Sea Environment Proceedings 64A (available at http://www.helcom.fi).
Elmgren, R (1989) Man s Impact on the Ecosystem of the Baltic Sea: Energy Flows Today and at the Turn of the Century, *Ambio*, 18, 326 332.
Leppakoski, E and Bonsdorff, E (1989) Ecosystem Variability and Gradients. Examples from the Baltic Sea as a Background for Hazard Assessment, in *Chemicals in the Aquatic Environment*, ed L Landner, Springer-Verlag, Berlin, 6 58.
Leppakoski, E and Mihnea, P E (1996) Enclosed Seas under Man-induced Change: a Comparison between the Baltic and Black Seas, *Ambio*, 25, 380 389.
Rapport, D J (1989) Symptoms of Pathology in the Gulf of Bothnia (Baltic Sea): Ecosystem Response to Stress from Human Activity, *Biol. J. Linn. Soc.*, 37, 33 49.
Voipio, A (1981) *The Baltic Sea*, Elsevier Oceanography Series 30, Amsterdam.

Water Resources: Great Lakes Case Study

Henry Regier
University of Toronto, Toronto, Canada

North America's chain of ve Great Lakes are headwaters within the St Lawrence River Basin. The Basin is shared by Canada, the United States and many tiny nations of Aboriginals of two major language groups. The two lakes lowest in the watershed, Ontario and Erie, have the highest densities of humans, and were severely degraded ecologically by 1960. Shallow southerly parts of the two middle lakes, Huron and Michigan, were equally degraded. Superior, the most northerly lake, was less degraded. All this degradation was due to careless industrialization, strong urbanization and poor land use practices. Many of the environmental abuses that were generated within the Basin have been partially corrected or remediated since 1960, at a direct cost of many billions of dollars. While this has been happening, researchers have shown that this Basin is also part of a global pattern of biospheric degradation. It is caught in: hemispheric processes of atmospheric transport of hazardous contaminants northward and eastward; worldwide shipping practices that involve transport in ballast water of exotic species, which then are released in Great Lakes ports; growing demands for export of water to drier areas of the continent with growing political pressure by entrepreneurs that the water be declared a commodity for purposes of international free trade; and migration of people into the Basin. Of course, the Basin has been contributing vast amounts of greenhouse gases and ozone-destroying chemicals to the biosphere, but its people are committed to reducing these atmospheric loadings gradually.

Fortunately, there has never been a major war in the Great Lakes Basin and not even minor wars for nearly two centuries. The Aboriginal Anishinabek and Haudenaushaunee peoples were generally considerate of European invaders, who usually did not reciprocate such consideration.

Throughout the 19th and 20th centuries, Americans and Canadians have cooperated implicitly or explicitly in the processes that led eventually to the intense ecosystemic debasement of the more southerly waters that are downstream in the Basin. Since about 1950 these two countries have collaborated in processes of remediation that have led to partial rehabilitation of the Basin s lands and waters. The programs under the numerous inter-governmental agreements undertaken to correct the environmental abuses are being celebrated for innovation with respect to participatory governance in the context of an ecosystem/bioregion approach (*see* Ecosystem Stability, Volume 2). The many enclaves of Aboriginal people each has a treaty with one of the two larger nations; gradually they have been empowering themselves to participate in correcting the ecological ills of the Basin for which they themselves can seldom be blamed.

Some kinds of regional abuses are still largely uncontrolled, e.g., intense smog events due, in part, to transboundary atmospheric flows in a northerly and easterly direction, e.g., from Chicago to Detroit to Hamilton/Toronto

to Montreal to Quebec City. Each conurbation is in both a downwind and an upwind position with respect to such massive fumigations. Not surprisingly, each is more agitated about being the downwind victim of smog fumigations rather than the upwind abuser that is contributing smog substances to be wafted downwind.

The current set of inter-jurisdictional Basin-wide programs and agreements does not suffice with respect to a need to correct the processes of biospheric degradation, as became evident in the 1990s.

The waters of the Great Lakes, as any other waters, act as temporary storage places for hazardous contaminants such as organohalides and polycyclic aromatic hydrocarbons. In water, these substances have strong tendencies to become complexed with weak chemical linkages to suspended clay or organic substances, and especially to living things. While temporarily in free form, such chemicals may escape into the air and then be transported downwind on particulates, to be washed out again by atmospheric precipitation.

In the 1990s it was found that pesticides that were used recently to the south and west of the Basin, notably Toxaphene and dichlorodiphenyltrichloroethane, were being transported into the lakes. Some of this new material then cycles harmfully among the biota through food webs, causing cancer and endocrine-related developmental handicaps, before being destroyed naturally, being buried naturally in sediments, or escaping again to the atmosphere to be transported farther east and north. Under some reasonable scenarios of reductions in the loadings of such contaminants to the regional and global environment, quantitative models predict that contamination of waters of the Great Lakes may fall to non-detectable levels sometime within the 21st century, depending on the particular chemical contaminant. Meanwhile generations of humans and other creatures will suffer harm.

Some of the threats posed by exotic creatures, e.g., the zebra mussel, transported into Great Lakes with ballast water of ocean-going ships, were perceived by researchers in 1980. Efforts in the mid-1980s by two binational commissions to stop such introductions to the Great Lakes had little effect on the shipping industry, in part because the designs of ballast systems in ships make retrofitting for purposes of preventing release of live exotic organisms almost impossible. Also, the leaders of the shipping industry appeared to have little if any concern about ecological phenomena, and could hide behind foreign registries of their vessels. The abundance of exotics released by ocean-going vessels was reduced somewhat by the late 1990s, but still to an inadequate extent. Meanwhile, the ecological associations of these lakes have been forever changed because of these exotics, most of which are found to be undesirable. Thus it is now unknown what the ecological outcome will be of ecosystemic rehabilitation programs which have cost many billions of dollars, in part because the exotic species are participating in unpredictable ecological reorganization.

The Great Lakes drain northeastward into the Gulf of St Lawrence and the northwest Atlantic Ocean. The mixing of the fresh and saline waters generates ecosystemic phenomena such as plankton blooms that are important for the reproduction of some valued migratory fish and other marine species. Thus the water that flows down the St Lawrence River is not wasted as is presupposed by some proponents of diversion of Great Lakes southward and westward to drier lands outside the Basin.

Economic studies of the various notions to divert Great Lakes waters have all shown that governmental subsidies would be necessary, i.e., that free market processes would not justify such expenditures over an alternative, say, of more efficient and less costly conservation of available water in those drier lands. Nevertheless, attempts to have any and all water of the Great Lakes Basin declared a commodity under current conventions of international free trade persist. Presumably the entrepreneurs expect to receive governmental subsidies for any ventures to divert or ship the water to drier lands. Aboriginals and naturalists, among others, object to such commodification, since they take the water to be part of the habitat of countless species of living things. The abundance, and even survival, of some of these aquatic creatures would likely be threatened by major withdrawals of water.

A fourth global process important to the Great Lakes Basin is the migration of people. In the 1990s about 40 million people lived in the Basin. Current assessments of climate warming in coming decades do not indicate catastrophic adverse effects, and might even imply that the Basin could generally become more attractive climatologically to mobile people. However, the current rate of urbanization is cause for concern. Further, the extant infrastructure of public utilities related to water, sewage, heating and cooling of buildings, etc., is largely obsolescent in the older parts of the Basin s cities. Hence, rapid population growth due to immigration into the Basin combined with postponement of replacement of obsolete infrastructure will likely exacerbate urban environmental problems.

The Great Lakes Basin was richly endowed with many kinds of resources that were developed and abused during the 19th and 20th centuries. As corrective measures were undertaken regionally late in the 20th century, toward a better but unpredictable ecosystemic state, it became apparent that the Basin was contributing and receiving massive abuses to and from other parts of the biosphere. No doubt, that will be a major challenge for all of the 21st century.

See also: Great Lakes Region of North America, Volume 4; Transboundary Water Resource Management: The Example of the Canada/US IJC (International Joint Commission), Volume 4.

Water Resources: Río de la Plata

Gustavo J Nagy[1], M Gómez-Erache[1], and A C Perdomo[2]

[1]*Facultad de Ciencias, Montevideo, Uruguay*
[2]*SOHMA, Montevideo, Uruguay*

The R´o de la Plata in South America is one of the world's largest and oceanographically most signi cant river and estuarine systems. Its economic value for transportation, sheries, and tourism is enormous, and its sustainable development an imperative (Wells and Daborn, 1997). Its basin, the most important in Latin America in terms of population and economic development, is changing as a result of both El Niño/Southern Oscillation (ENSO) related variability and such human driven activities as soil erosion and fertilizer application. As a consequence of continuing demographic and economic growth in the basin, environmental deterioration will be dif cult to arrest.

STUDY SITE

The Río de la Plata is a large, river-dominated, microtidal, wind driven, nutrient rich, stratified, and plankton-based coastal system (Figure 1) fed by the rivers Paraná and Uruguay (freshwater inflow into the estuary $\sim 7\ \ 12 \times 10^{11}\,m^3\,y^{-1}$). River flow typically varies on both seasonal and interannual timescales between 22 000 and 28 000 $m^3\,s^{-1}$, with extreme values, i.e., more than 30 000 $m^3\,s^{-1}$ and less than 20 000 $m^3\,s^{-1}$ during El Nino and La Nina, respectively (Nagy *et al.*, 2001, in press).

Wind and water discharge rate control the seasonal variability of the salinity field in the upper layer of the river. The lowest wind influence and a maximum in river runoff result in a lower salinity regime, whereas onshore winds and a minimum of riverine runoff induce a high salinity condition (Guerrero *et al.*, 1997).

ENVIRONMENTAL VARIABILITY AND CHANGE

The main global environmental change processes in the Río de la Plata are:

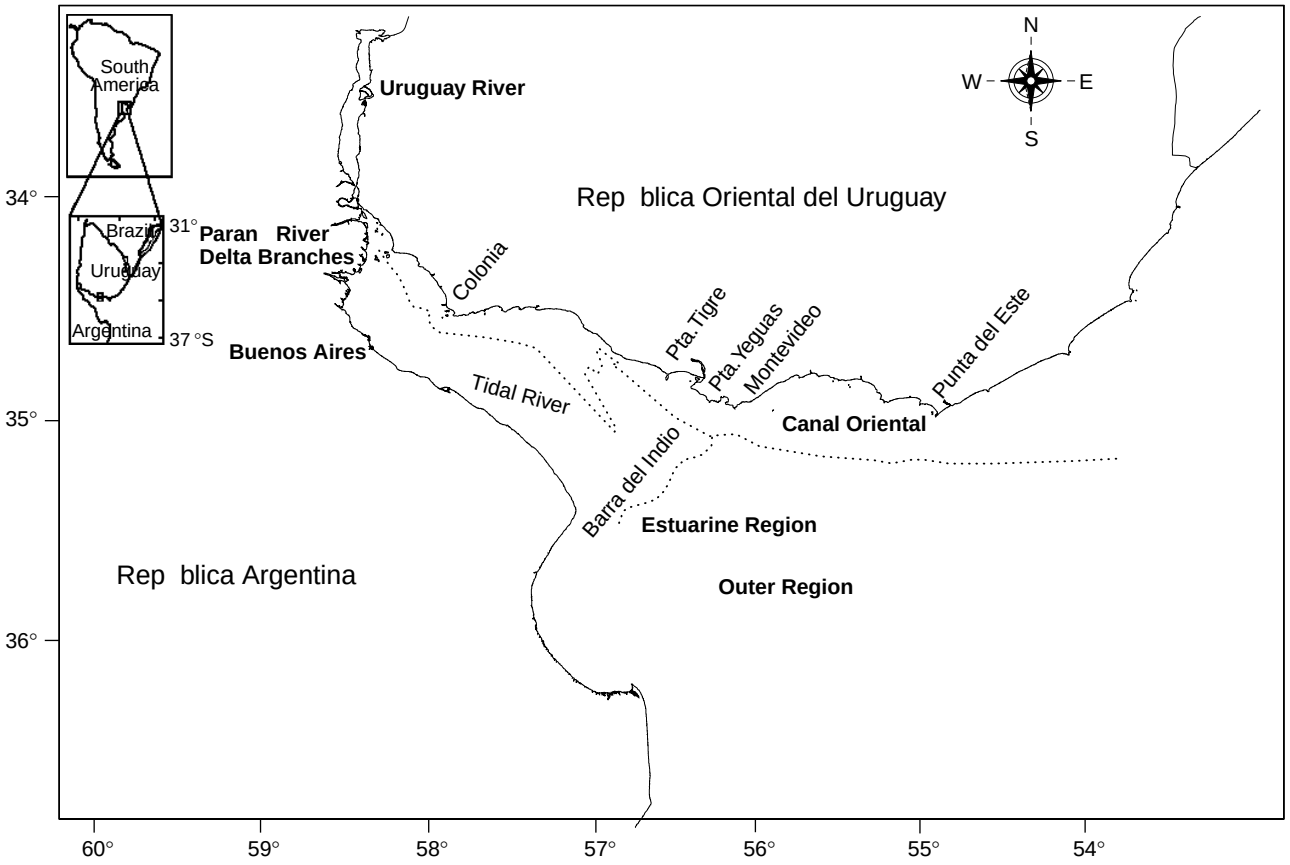


Figure 1 Río de la Plata system location and regions. The Uruguayan coast (Canal Oriental) is deeper at each cross-section and consequently is more stratified (greater bottom-surface salinity difference)

1. *Hydroclimatic variability and change*: a shift occurred in the Río de la Plata basin freshwater inflow in the early 1970s (García and Vargas, 1998). The variability and total volume of river discharge have both increased by about 25% in the last 30 years. Most of these fluctuations and part of this trend are associated with ENSO-related variability (Nagy *et al.*, 1996, 1997). Since 1972, four strong El Nino events (1972, 1983, 1992, 1997) and two strong La Nina events (1988, 1998) strongly affected the Río de la Plata basin and coastal system. Over the long term, river discharge fluctuations control yearly salinity variability (Figure 2) and induce sea level changes (± 10 cm) at Montevideo (Nagy *et al.*, 1996, 1997).
2. *Eutrophication*: The Río de la Plata receives primary treated urban wastewater and fertilizers from the basin. Point sources from the Buenos Aires metropolitan area account for about 25% of the total nutrient load to the river (Pizarro and Orlando, 1985), estimated to be about $3{\bullet}5 \times 10^6$ mol dissolved inorganic phase (DIP) d^{-1} and 64×10^6 mol dissolved inorganic nitrogen (DIN) d^{-1} (Nagy, 2000). Because of *in situ* sink processes like organic matter production and denitrification (Nagy, 2000), nutrient values in the estuarine region are within the range considered by Nixon (1995) as typical for most subtropical coastal zones with low human-induced eutrophication (Gómez-Erache *et al.*, 2000). Direct atmospheric deposition of nitrogen (N) has been estimated to be to the order of 1 2% of the total nitrogen load (Nagy, 2000). This value is consistent with the estuarine area/watershed area ratio $\simeq$1.2%, suggesting low deposition (Ocean Studies Board, 2000).

ENVIRONMENTAL CHANGES RELATED TO HYDROCLIMATIC CHANGE AND NUTRIFICATION

Such changes include:

- Decrease of mean yearly salinity (Figure 2) at Montevideo (Nagy *et al.*, 1997).
- Increase of vertical saline stratification (Nagy *et al.*, 2000).
- Increase of diffuse nitrates load. After the El Nino flood in 1983, when the water discharge increased by about 50%, the nitrate concentrations rose by 25%, whereas phosphate concentrations were 50% lower; therefore, the nitrate load and the N/P ratio increased by about 75%.
- Low dissolved oxygen values (• 50% saturation) in the bottom layer (Nagy *et al.*, 1997, 2000).
- Downward displacement of the estuarine front, affecting the fish spawning zone and the catchment.
- High nitrogen flux per unit land area (Nixon, 1995); the livestock wastes reported by the author seem to be too high.
- High diffuse and point source phosphorus inputs from the Uruguayan coast to the Río de la Plata via fertilizers and detergents (Sommaruga *et al.*, 1995).

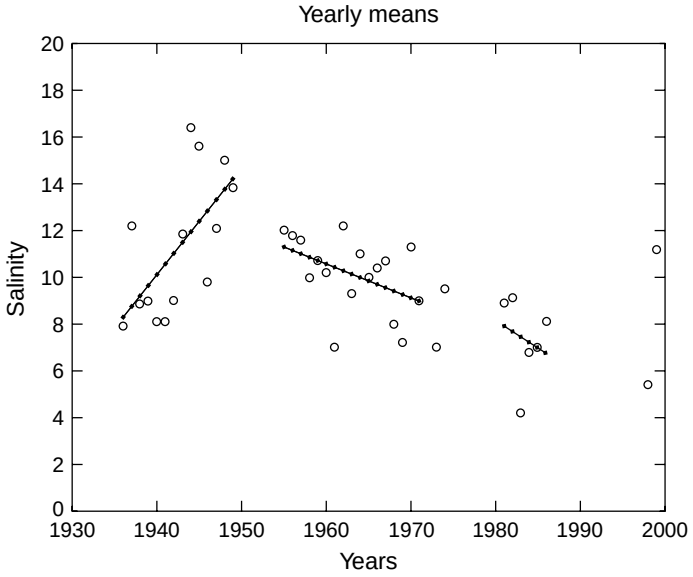


Figure 2 Yearly mean salinities at Montevideo. The series is discontinuous because of the lack of data for several periods. The means are highly correlated with total yearly river flow into the estuary, especially with the river Uruguay discharge. Note the great variation between 1998 (strong El Niño) and 1999 (strong La Niña). The long-term trend and extreme values of salinity match the river inflow trend and fluctuations

Figure 3 Turbidity front image (November 17th, 1999) during the peak of La Niña signal in the Río de la Plata. The front was displaced three to five tens km up-river of its average location because of very low river flow. Consequently, both residence time within the system and biological activity increased, and algal pigment concentrations reached eutrophic values (up to 38 μg chlorophyll-a l^{-1})

- Freshwater cianobacteriae blooms (coming from the river Uruguay) and nitrogen fixation (De León and Yunes, 2001).
- Harmful and toxic algal blooms (Méndez *et al.*, 1996).
- A slight increase of water level (+3 cm) at Montevideo; most yearly fluctuations are correlated with freshwater inflow (Nagy *et al.*, 1997), and therefore, with the ENSO signal.
- Eutrophic values of algal biomass (• 25 μg chlorophyll-a l^{-1}) have been found:

    1. close to the emissions of urban waste waters, i.e., in Montevideo bay (Gómez-Erache *et al.*, 2000); and
    2. seaward of the turbidity front during La Nina low river flows in November 1999 (Figure 3).

GLOBAL CHANGE PERSPECTIVES

Stratified systems, like the Río de la Plata along the Uruguayan coast, are highly susceptible to oxygen stress in the bottom layer as a response to nutrient over-enrichment and the consequent increase of productivity and organic matter respiration (Rabalais *et al.*, 1999). Excess nitrogen (a limiting nutrient), stimulates primary production. However, as anoxia does not occur in the Río de la Plata because of mixing by the wind, denitrification is also stimulated by excess nitrogen, acting as a feedback on both carbon dioxide (CO_2) and nitrogen (N_2) nitrous oxide (N_2O). If river flow increases, as in 1998, vertical stratification, allochtonous organic matter, and oxygen stress would also increase. If anoxia develops, it could act as a positive feedback: both available nitrogen for organic matter production and oxygen deficit increase. If this is coupled with increased human-driven nutrification, in the long-term the system could become heterotrophic (net carbon dioxide emission).

ACKNOWLEDGMENTS

The authors are grateful to the Eco-Plata Project, financed by the International Development Research Center-Canada (IDRC-Canada), and to the Interamerican Institute for Global Change (IAI).

REFERENCES

De León, L and Yunes, J S (2001) First Report of a Mycrocystin-Containing Bloom of the Cyanobacterium Microcystis Aeruginosa in the La Plata River, South America, *Environ. Toxicol.*, 16(2), in press.

García, N O and Vargas, W M (1998) The Temporal Climatic Variability in the Río de la Plata Basin Displayed by the River Discharges, *Clim. Change*, 38(3), 359 379.

Gómez-Erache, M, Vizziano, D, Muniz, P, and Nagy, G J (2000) *The Health of the R´o de la Plata System: Northern Coast, Uruguay. Coastal Zone Canada 2000 Case Study/Workshop: Nutri cation of Coastal Waters*, BoFEP Program, New Brunswick, Saint John, Canada, September.

Guerrero, R E, Acha, E, Framinán, M, and Lasta, C (1997) Physical Oceanography of the Río de la Plata Estuary-Argentina, *Cont. Shelf Res.*, 17(7), 727 742.

Méndez, S, Severov, D, Ferrari, G, and Mesones, C (1996) Early Spring Alexandrium Tamarense Toxic Blooms in Uruguayan Waters, in *Harmful and Toxic Blooms*, eds Y T Oshima and Y Fukuyo, Intergovernmental Oceanographic Commission of UNESCO, 113 116.

Nagy, G J, Martínez, C M, and Forbes, E A (1996) Physical Environment Variability at the Montevideo Coastal Zone. I: Historical Evolution, in *Eco-Plata 1996 International Conference - Towards a Sustainable Development of the Uruguayan Coastal Zone of the R´o de la Plata*, Abstract of Scientific Papers, 4.

Nagy, G J, Martínez, C M, Caffera, M R, Pedrosa, G, Forbes, E A, Perdomo, A C, and López Laborde, J (1997) The Hydrological and Climatic Setting of the Río de la Plata, Chapter 2, in *The R´o de la Plata. An Environmental Overview*, eds P G Wells and G R Daborn, An Eco-Plata Project Background Report, Dalhousie University, Halifax, Nova Scotia, 248.

Nagy, G J (2000) *Dissolved Inorganic NP Budget for the Frontal Zone of the R´o de la Plata system. Estuarine Systems of the South American Region: C, N, P Fluxes 2000*, LOICZ-UNEP Workshop Reports, eds V Dupra, S V Smith, J I Marshall Cross, and C J Crossland, Loicz Reports and Studies, Vol. 15, LOICZ, Texel, The Netherlands.

Nagy G J, Gómez-Erache, M, Martínez, C M, and Perdomo, A C (2000) *Nutrient Over-Enrichment Surveillance: Quintennial Time-Scale Modeling and Monitoring in the R´o de la Plata Coastal System (Uruguay-Argentina)*, Nutrient Over-Enrichment Symposium, Poster Session, National Academy of Sciences, Washington, DC, October.

Nagy, G J, Gómez-Erache, M, López, C H, and Perdomo, A C (2001) *Nutrient, Behavior and Eutrophication Symptoms in the R´o de la Plata River Estuary System*, ed E Orive, *Proc. 31st Annu. Symp. Estuarine and Coastal Sciences Association Hydrobiologia*, in press.

Nixon, S W (1995) Coastal Marine Eutrophication: a Definition, Social Causes and Future Concerns, *Ophelia*, 41, 199 219.

Pizarro, M J and Orlando, A M (1985) Distribución de Fósforo, Nitrógeno y Silicio disuelto en el Río de la Plata, *Serv. Hidr. Naval. Secr. Marina*, H-625, 1 57.

Ocean Studies Board (2000) *Clean Coastal Waters. Understanding and Reducing the Effects of Nutrient Pollution*, Ocean Studies Board and Water Science and Technology Board, Commission on Geosciences, Environment, and Resources, National Research Council, National Academy of Sciences Press, Washington, DC.

Rabalais, N N, Turner, R E, Justic , D, Dortch, Q, and Wiseman, Jr, W J (1999) *Characterization of Hypoxia. Topic 1 Report for the Integrated Assessment of Hypoxia in the Gulf of Mexico*, NOAA Coastal Ocean Program Analysis Series, NOAA Coastal Ocean Program, Silver Spring, MD, 1 167.

Sommaruga, R D, Conde, D, and Casal, J A (1995) The Role of Fertilizers and Detergent for Eutrophication in Uruguay, *Fresenius Eviron. Bull.*, 4, 111 1116.

Wells, P G and Daborn, G R (1997) *The R´o de la Plata: an Environmental Overview. An Eco-Plata Project Background Report*, Dalhousie University, Halifax, Nova Scotia, 256.

Waterlogging

Waterlogging occurs when water is near, at, or above the surface of the land, resulting in the total saturation of the soil. It can be permanent or intermittent, the latter due to seasonal or other conditions. It has a number of causes, all of which are associated with inadequate or impeded drainage, including over-irrigation, rising groundwater levels, and seepage, as well as periods of high rainfall and flooding. The effects of waterlogging depend in large part on how long it lasts. While a soil is waterlogged, there is no energy for plant growth, nutrient uptake stops, plant roots are drowned and die (except for those of hydrophytes, plants that grow in water, either submerged, emergent or floating), soil microbes and other organisms die, and there is increased likelihood of plant and animal diseases. As a result, plants are damaged, their productivity reduced, and they may eventually die. Waterlogging affects a significant proportion of the world s irrigated lands, the result of inadequate or non-existent drainage; a 1977 United Nations study estimating the area affected at some 22 million ha. The costs are substantial. For example, in the late 1980s, it was estimated that farm production from irrigated dairy pastures in northern Victoria (Australia) could be increased by over $AUD50 million per year if waterlogging problems could be eliminated. Where waterlogging is more or less permanent, the anaerobic conditions reduce the ability of bacteria to break down the dead plant materials. The accumulation and only partial decomposition of plant materials can result in the formation of peat, especially in temperate humid conditions. Waterlogging also destroys soil structure and leads to the formation of gley soils (characterized by the reduction of iron and other elements producing their blue-gray mottling). Waterlogging is often a precursor to salinization.

PETER CRABB *Australia*

Wave Power Development

Les Duckers
Coventry University, Coventry, UK

Wave power technologies are being developed to generate electricity, heat or mechanical power from ocean waves.

Wind, passing over the surface of water, gradually transfers some of its energy into the water to create waves. If the wind has a reasonable velocity and persists for a significant time across a long stretch of water, then the resulting waves will be large and powerful. This is why certain locations around the World enjoy favorable wave climates for electricity generation: for them the predominant winds have traveled across several thousand kilometers of ocean and in so doing have generated annual average wave power levels of around 50 kW m^{-1} of wave front.

The exploitable World wave power resource is estimated at 2000 GW, largely to be found in offshore locations where the water is deeper than 40 m. The shoreline resource, although easier to exploit, has lower power density (typically around 20 kW m^{-1}) since some of the energy content of the waves is dissipated as the waves approach the shore.

Concern for the security of energy supply during the oil crisis of the 1970s prompted modern research and development in wave power. Shore-mounted oscillating water column (OWC) and the offshore Edinburgh Duck were among the rst concepts to provoke interest and both continue to be developed.

Now, though, the motivation comes from the low environmental impact of electricity generated from wave power, which emits about 11 g of carbon dioxide (CO_2) kW h^{-1} compared to 730 g CO_2 kW h^{-1} for average UK electricity generation.

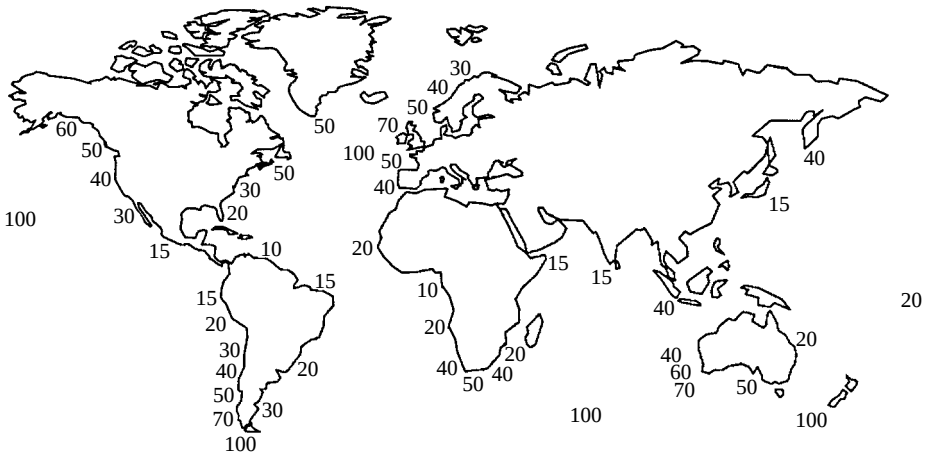


Figure 1 Estimated world potential of wave technologies. The shoreline unit (1–50 GW): fixed to shore in wave power densities of about 20 kW m^{-1}. The offshore unit (200–5000 GW): probably floating in deep water, with wave power densities of about 50 kW m^{-1}. The figures are the author's estimates of the technically achievable powers. Approximate wave powers are given as kW m^{-1} of wave front

SHORE-MOUNTED TECHNOLOGIES

Many of the prototype schemes are shore mounted and incorporate an OWC (oscillating water column) driving a self-rectifying air turbine. These schemes have few moving parts, and with judicious choice of structural size and geometry, overall performance can be impressive. Micro-OWCs are deployed in huge numbers to power navigation buoys with a few kilowatt, a 600 kW shore mounted scheme was built in Norway in 1985 and 1 MW scheme is installed in Scotland. India, Japan, China, Portugal are some other countries which are piloting OWCs.

Tapchan (Tapered channel) is also a shore-mounted device but has the advantage of some inherent energy storage. As waves approach the mouth of the Tapchan they become compressed in width and so build up in height. It is the design of this Tapchan, which enables the scheme to harvest energy from the ocean waves arriving at the mouth of the channel. The 350 kW prototype, built in 1985 in Norway, had a 40 m wide horn shaped collector capable of accepting energy from waves with a range of frequencies and directions. A 1.1 MW plant is under construction in Indonesia.

OFFSHORE TECHNOLOGIES

Large-scale offshore devices like the Duck and Clam will be deployed in deep water and so can exploit the higher wave power densities to found there. The numbers plotted on the World map in Figure 1 are the average deep water wave power density in kW m^{-1} of wave front. Since these offshore systems will have to float, they are designed to be large enough to span several wave crests. In this way the wave force on the structure is averaged at an acceptably low value permitting the structure to survive and moreover to provide a reasonably stable frame of reference. Arrays of hundreds of such devices could form wave farms and could transmit substantial amounts of electricity to the Irish, UK and Norwegian grids. The collective output from an array will be smoother than from individual devices. Offshore devices would transmit electricity via subsea cables. Modular construction methods should permit minimization of capital costs and ensure rapid deployment and early operation of each unit.

The Edinburgh Duck was originally envisaged as many cam-shaped bodies linked together on a long flexible floating spine, which was to span several kilometers of the sea. Individual ducks respond to incident waves by rotating with respect to the common spine.

The SEA-Coventry Clam is an 80 m diameter rigid toroid, rated at 3 MW, delivering an average of 1 MW, and floating in deep water. Twelve air cells are arranged around the circumference of the toroid and these cells are coupled together by an air ducting, which contains 12 air turbines.

POINT ABSORBER TECHNOLOGIES

Float systems, with a body floating on the surface but moored to the seabed via a pump are attracting some attention. These can act as point absorbers, which draw in energy from a greater width of water than their own physical diameter (i.e., they have capture width ratios greater than one).

The hose pump wave energy converter has been developed in Sweden and is intended to pump sea water from an array of devices fixed to the seabed to an elevated reservoir on land. A Pelton wheel extracts energy from the water as it is released from an upper reservoir back to the

sea. A hose-pump lightbuoy is undergoing pre-production tests, and evaluation of wave power plants in Ireland, Spain, Sweden and the USA have been carried out.

Variations on the systems mentioned above or new ideas are being developed in Denmark, the Netherlands, Norway, USA, Korea and Australia.

PREDICTED WAVE ENERGY COSTS

The predicted cost of electricity generated from wave power schemes is dropping as designs improve and as construction techniques are refined. To illustrate this, one of the OWC designs had a cost of US\$0.3 $kW\,h^{-1}$ in 1982. By 1993, OWC designs were costed at US\$0.11 $kW\,h^{-1}$ and projections suggest that costs will fall to US\$0.06 $kW\,h^{-1}$ by 2005 and lower still by 2010.

FURTHER READING

Duckers, L J (2000) Wave Power, *Eng. Sci. Educ. J.*, 9(3), 113 122.

Proceedings of the Second European Wave Power Conference, Lisbon (1995) 8 10 November, European Commission ECSC-EC-EAEC, Brussels, ISBN 92-827-7492-9.

Proceedings of the Third European Wave Power Conference (1998) Patras, Greece, 30 September 2 October.

The European Wave Atlas, www.ineti.pt/weratlas/index.

Thorpe, T W (1992) *A Review of Wave Energy*, Report for the DTI, ETSU Report, R-72.

Thorpe, T W (1998) *Overview of Wave Energy Technologies*, Report for the DTI, AEAT-3615.

Wave Power (1999) Moving Towards Commercial Viability International One-Day Seminar ImechE, London 30th November.

Weather Extremes and Climate Impacts: a Case Study for the United States

Roger A Pielke, Jr

National Center for Atmospheric Research, Boulder, CO, USA

Weather events are among the most pervasive in uences on human affairs. Good weather can enhance society's efforts, bad weather often reveals vulnerabilities and sometimes leads to human and environmental disasters. The degree to which society exploits good weather and reduces its vulnerabilities to bad weather is a function of how society organizes and informs itself and its decision processes in the face of various typical and extreme weather events. This essay provides a perspective on the manner in which researchers acquire information about extreme weather impacts and how societies cope (and do not cope) with extreme weather. Such information and coping strategies are increasingly important as societies around the world see their economic and human vulnerabilities change and grow with increasing population and wealth.

What are weather impacts?

In the 1970s, climate became of increasing interest to many societies around the world. Events included the failed Peruvian anchovy harvests in 1971 and 1973; the 1972 1974 drought in the African Sahel; a severe 1972 winter freeze in the Soviet Union; and in 1974, floods, drought, and early frost in the US Midwest. The 1977 winter in the eastern US was the coldest ever recorded and summer was one of the three hottest in a century. As a consequence of these extreme events and their impacts, decision makers in influential national and international settings began paying significant attention to the relation of weather, climate, and human affairs (*see* Climatic Extremes, Volume 3).

The concept of weather impacts and related societal problems includes:

- extreme events, including droughts, hurricanes, floods, blizzards, tornadoes, thunderstorms (including hail), etc.;
- the benefits of good weather, meaning favorable conditions for a particular activity;
- routinely disruptive weather, defined as not extreme, but significant enough to warrant behavioral adjustments.

Most public and political attention focuses on disasters and extreme events, but the other types of weather are likely to have significant impacts on society as well. This essay focuses on extreme events.

THE CHALLENGE OF ASSESSING DAMAGES OF EXTREME WEATHER AND STORMS

In the aftermath of any extreme event such as a hurricane, there are many considerations in measuring its costs (of impacts). Any assessment of impacts resulting in an estimate of total damages associated with a disaster must pay explicit attention to assumptions guiding the analysis in order to facilitate interpretation of the estimate. The analyst needs to pay attention to five factors that can complicate damage assessment: contingency, quantification, attribution, aggregation, and comparison.

Contingency: The Problem of Multiple-order Impacts

When an extreme event strikes a community, it leaves an obvious path of destruction. For example, following a hurricane, as a result of high winds and water from a storm surge, homes, businesses, and crops may be destroyed or damaged, public infrastructure may also be compromised, and people may suffer injuries or loss of life. Such obvious impacts can be called direct impacts because of the close connection between event and damages. The costs associated with direct impacts are generally easiest to assess because they come in discrete quantities. In the US, federal insurance payouts are one measure of direct impacts, as are federal aid, public infrastructure reconstruction, and debris removal.

Secondary impacts are related to direct impacts. Generally, secondary impacts occur in the days and weeks following an event. For example, a flood may destroy a water treatment plant. The direct impact is the cost associated with rebuilding the plant; secondary impacts include the costs associated with providing fresh water for local citizens. In general, such secondary impacts are more difficult to assess because they require estimation and are part of an existing social process; e.g., estimating the costs of providing fresh water in lieu of the costs of normal operation of the plant without the disaster.

Other impacts on time scales of months and years occur and can easily be imagined but can be hard to quantify. For example, a flood may destroy a number of businesses in a community resulting in a decrease in tourist visits, which in turn leads to a shortfall in sales tax collection. As a result, community services that had been funded from sales tax revenues may suffer, leading to further social disruption and thus additional impacts. Estimation of the costs associated with such impacts is difficult because of numerous confounding factors. In short, an extreme event causes various impacts that reverberate through the social system for short and long periods. Separating the signal from the noise of ongoing social processes becomes increasingly difficult as the impact becomes further removed in time and in causation from the event s direct impacts.

Attribution: The Problem of Causation

Related to contingency is attribution. In the aftermath of a natural disaster, people are quick to place blame on nature, e.g., the flood caused billions of dollars in damages. However, it is often the case that natural disasters are a consequence of human failures. Damage is often a result of poor decisions of the past and inadequate preparation, rather than simply the overwhelming forces of nature. It is often at the intersection of extreme events and poor preparation that a disaster occurs. An important aspect of learning from a disaster is to understand what damages and casualties might have been preventable and which were not. Gross tabulations of damages neglect the question of why damage occurred, and often implicitly blame nature rather than society.

Quantification: The Problem of Measurement

How much is a life worth? Or put in practical terms, how much public money are people willing to pay to save one more life in the face of an environmental hazard? According to one review, the public assigns between $2.0 million and $10.9 million as the value of a human life. The difficulty associated with assigning an economic value to a human life is representative of the more general problem of assessing many of the costs associated with weather impacts. Similar questions might include: what is the value of a lost ecosystem, park, or unrecoverable time in school, etc.? What are the costs associated with psychological trauma?

Often, an extreme event impacts many aspects of society that are not explicitly associated with an economic measure (e.g., well-being). As a consequence, any comprehensive economic measurement of an extreme event s impact necessitates the quantification of costs associated with subjective losses. Therefore, the assumptions that one brings to assessment of value can affect the bottom line. Care must be taken to make such assumptions explicit in the analysis.

Aggregation: The Problem of Benefits and Spatial Scale

Estimates of impacts from disasters rarely consider benefits. Consider the following example: following a hurricane that severely damages agricultural productivity in a region, commodity prices rise nationwide. Thus, while farmers in the affected region may lose money, farmers outside of the region may actually incur significant benefits. At a national level the hurricane may thus lead to net economic benefits. Even in the immediate vicinity of the storm track, the rainfall may be beneficial in replenishing depleted reservoirs and groundwater.

The above example points to two sorts of issues: benefits and spatial scale. Arguably, following every disaster some individuals and groups realize benefits in some way from the event. Should such benefits be subtracted from a hurricane s total impact? Further, the picture of damages depends upon the scale of the analysis. For the same event a county may experience complete devastation, the state only moderate impacts, and the nation positive benefits. Transfers of wealth through disaster aid further complicate the picture. Since there are multiple valid spatial scales from which to view a hurricane s impacts, careful attention must be paid to the purposes of loss estimates. Furthermore, it is important to remember that impacts go beyond those things

that can be expressed in dollars suffering and hardship are losses independent of scale.

Comparison: The Problem of Demographic Change

As a consequence of the challenges facing meaningful impact assessment, comparing impacts across time and space is problematic. In many settings, past extreme events would certainly leave a greater legacy had they occurred in more recent years. Yet, damage statistics often go into the historical record noting only the event and economic damage (usually adjusted only for inflation). Such statistics can lead to mistaken conclusions about the significance of damage trends. Since population and property at risk to extremes have changed dramatically in recent decades, such statistics may grossly underestimate vulnerability. Therefore, care must be taken in the use of bottom-line damage estimates to reach policy conclusions.

The Bottom Line: Apples with Apples, Oranges with Oranges

There are many ways to measure the costs associated with a weather or climate impact. There is no one right way. The method chosen for measurement of the costs of damages depends upon the purposes for which the measurement is made, and therefore must be determined on a case-by-case basis. No matter what method is employed when assessing or using the costs associated with an impact, the analyst needs to ensure at least two things. First, the analyst needs to explicitly describe the assumptions that guide the assessment: what is being measured, how, and why. Second, apples should be compared with apples and oranges with oranges. If the purpose is to compare the impacts of a recent event with one in the past or a weather disaster with an earthquake, the methods employed ought to result in conclusions that are meaningful in a comparative setting.

IMPACTS TREND DATA

Following the discussion above, a three-tiered sequence is useful for assessing weather and climate impacts. Direct impacts are those most closely related to the event, such as property losses associated with wind damage. Secondary impacts are those resulting later from the direct impacts. For example, an increase in medical problems or disease following a hurricane would be a secondary impact. Tertiary impacts are those that follow long after the storm has passed. A change in property tax revenues collected in the years following a storm is an example of a tertiary impact. The impacts discussed below are direct impacts.

Hurricanes

In the US, after adjusting only for inflation, hurricanes were responsible for an annual average of $1.6 billion for the period 1950 1989, $2.2 billion over 1950 1995, and $6.2 billion over 1989 1995. For comparison, China suffered an average of $1.3 billion (unadjusted) in damage related to typhoons over the period 1986 1994. Significant tropical cyclone damage is also experienced by other countries including those in south-east Asia, along the Indian Ocean (including Australia), islands of the Caribbean and Pacific, and in Central America (including Mexico). While a full accounting of these damages has yet to be documented and made accessible, it is surely in the billions of dollars, with a reasonable estimate of about $10 20 billion annually (1995 $).

In the US, 196 people lost their lives as a result of hurricanes over the period 1986 1995. Experts have estimated that world wide, tropical cyclones result in approximately 12 000 23 000 deaths. Tropical cyclones have been responsible for the largest losses of life due to a natural disaster. For instance, in April 1991, a cyclone made landfall in Bangladesh resulting in the loss of more than 140 000 lives and disrupting more than 10 million people (and leading to $2 billion in damages). A similar storm resulted in the loss of more than 250 000 lives in November 1970. Similarly, thousands of lives have been lost in China, India, Thailand, Honduras, and the Philippines in recent years.

Society has become more vulnerable to hurricane impacts. The trend of increasing losses during a relatively quiet period of hurricane frequencies should be taken as an important warning. When hurricane frequencies and intensities return to levels observed earlier in the 20th century, then losses are likely to increase to record levels unless actions are taken to reduce vulnerability.

Inhabitants along the US Atlantic and Gulf Coasts are fortunate in that hurricane watches and warnings are readily available, as are shelters and well-conceived evacuation routes. However, this should not give reason for complacency the hurricane problem cannot be said to be solved. Disaster planners have developed a number of scenarios that might result in a large loss of life in the US. For instance, imagine a situation of gridlock as evacuees seek to flee the Florida Keys on the only available road. Or imagine New Orleans, with much of the city below sea level, suffering the brunt of a powerful storm, resulting in tremendous flooding. Scenarios such as these highlight the need for constant attention to saving lives. Since the nature of the hurricane problem is constantly changing as society changes, the hurricane problem can never be said to be solved.

Floods

By any measure, floods have a significant impact on society. The Red Cross estimates that, worldwide, more than

1.5 billion people over the 25-year period ending in 1995 have felt the impact of floods. Of that total, they estimate that more than 318 000 people were killed and more than 81 million were made homeless. In addition, over the period 1991 1995 flood-related damages totaled more than US$200 billion (not inflation adjusted) worldwide, representing close to 40% of all economic damages attributed to natural disasters in this 5-year period. In the 1990s, significant flood damage occurred in eastern Europe (1997, damages totaling US$4 6 billion), US and Canada (1997, US$1.75 billion), China (1996, US$26.5 billion), central Europe (1993 and 1995, US$4 billion), and the US Midwest (1993, US$16 billion).

Reliable and accurate knowledge of flood impacts is important because decision makers allocate scarce resources to flood-related concerns based on their interpretation of past trends in and expectations for the future of the magnitude and causes of flood impacts. In addition, local actions taken to reduce vulnerability to floods are based on public and decision-makers perceptions of the past and expectations for the future. Regrettably, implementation of policies in response to floods is hampered by a lack of specific knowledge of past trends in flood impacts on society and, more importantly, the factors that underlie those trends. Indeed, the International Federation of Red Cross and Red Crescent Societies has written:

> The lack of systematic and standardized data collection from disasters, man-made or natural, in the past is now revealing itself as a major weakness for any developmental planning. Cost benefit analysis, impact analysis of disasters or rationalization of preventative actions are severely compromised by unavailability and inaccuracy of data or even field methods for such data collection.

Some have speculated that the steady increase in the average annual economic damages related to floods in the US indicates a change in climate; some blame population growth and development; others place the blame on federal policies; and still others suggest that the trend distracts from the larger success of the nation s flood policies. Empirical evidence from a number of cases clearly shows that climate, population growth and development, and flawed policy all play a role in trends in the costs of flooding in the US, but the state of knowledge is such that the relative contribution of each factor is poorly understood. The US case seems typical: policy makers face difficulties in assessing the magnitudes and causes of floods and in evaluating the effectiveness of past responses.

A consequence of the lack of data on the spatial extent of floodplains in the US is the difficulty in assessing trends in flood frequencies and current levels of population at risk to floods. Trends in population at risk to flood events are important in any determination as to whether societal vulnerability to floods is decreasing, increasing, or remaining relatively constant. One can easily hypothesize that increasing population and urbanization in the US has led to a commensurate increase in population at risk. Yet, one can also hypothesize that the various societal responses to floods may have more than compensated for population growth and, in fact, fewer people are today at risk to flood events. Currently, reliable data to assess trends in populations at risk from flood events are lacking.

Accurate determination of property at risk to flooding faces many of the same obstacles. Differences in estimates of people and property at risk to floods are attributable to actual demographic changes, but also to differences in floodplain definitions, and to the lack of reliable data. The data that do exist allow for only gross generalizations.

CONCLUSIONS

The information presented above is consistent with the findings of the Intergovernmental Panel on Climate Change (IPCC). Even on a global scale it is difficult for scientists to discern trends in extreme events. As the IPCC (1996, 173) notes:

> Overall, there is no evidence that extreme weather events, or climate variability, have increased, in a global sense, through the 20th century, although data and analyses are poor and not comprehensive. On regional scales there is clear evidence of changes in some extremes and climate variability indicators. Some of these changes have been toward greater variability; some have been toward lower variability.

For many years, decision-makers have assumed that climate remains constant, at least over the relevant period of record. So it was possible to estimate statistical return periods from historical time series. In recent years, however, scientists and policy makers have come to realize that we live in a climate that is changing in ways that are difficult to assess and predict. In other words, both the distribution of events and the central tendency may be changing. A consequence is that climate change or variability may be responsible for some of the variance in weather-related deaths and damages.

However, it is difficult to attribute the documented recent increases in economic impacts in the US to fluctuations in climate. This is primarily due to the fact that the strongest signal in the impacts record is increased societal vulnerability.

In this context, studies of extreme events have often called for greater emphasis to be placed on the following areas:

Vulnerability Mapping

Communities would benefit from a greater understanding of exposure and risk. Many countries, particularly developing countries, lack the resources to undertake such studies. Even some US communities have never been evaluated for

flood risk. In addition to mapping where natural hazards could occur, community planners could map the neighborhoods where highly vulnerable groups such as the disabled, poor, and elderly reside. This information can provide a tool for targeting education and other mitigation programs, evacuation plans, distribution of relief supplies, and other response services to where they are most needed.

Land Use

Much development around the world occurs in the most vulnerable locations, including low-lying shorelines and inland flood-prone areas. Land use planning is typically in a rudimentary stage in those regions. For example, coastal areas of parts of the southeastern US have experienced explosive growth over recent years. In 1990, Dade and Broward Counties in south Florida had larger populations than all 109 counties from Texas through Virginia on the Gulf and Atlantic coasts in 1930. Some policies for managing land in vulnerable areas include zoning, regulation, taxing, creating special property districts in which governments may exercise eminent domain, and public acquisition of land at risk.

Promulgation and Effective Enforcement of Building Codes

Building codes are an important component in strengthening housing stock and reducing vulnerability. In disaster-prone areas, effective building codes should address factors such as the ability of structures to withstand high wind velocities. For example, South Florida s building code requires structures to be able to withstand wind velocities not less than 120 mph (53.2 m s^{-1}) at a height of 10 m above the ground. Many engineers maintain that using steel or composite fasteners to hold on a roof would save homes. However, even the best codes will not reduce vulnerability without adequate enforcement. One study comparing damage from two hurricanes of roughly equal size and intensity found that building code enforcement was a primary factor in reducing structural damage to buildings during hurricanes. Other strategies are necessary in those parts of the developing world where much of the construction of homes occurs informally and does not go through a code review and inspection process.

Improved Forecasts, Warnings, and Evacuations

While major improvements in weather forecast accuracy have been made, thanks to advances in technology, considerable opportunity exists for further improvement in both the technical production of forecasts and their use by decision-makers. In addition, some countries need better national communications systems to disseminate warnings. Dissemination of effective warnings can be a challenge in countries such as India where warnings need to be translated into several languages. It is likely that those who suffered the greatest impacts of Hurricane Mitch in Central America had little or no warning of its approach and little opportunity to evacuate even if warned.

For evacuation to be effective, communities must develop data banks on various factors including: what areas need to be evacuated, how many people need to be evacuated, how will the public respond to an evacuation, what roads will be used, and where will evacuees be sheltered? Improvements to the evacuation process can range from widening roads, to building new causeways, to limiting population growth in areas most at risk.

Reduce Environmental Degradation

Deforestation is occurring at alarming rates in some parts of the world. Loss of natural vegetation reduces the soil s ability to absorb rainwater, and contributes to the instability of hill slopes. These factors can lead to highly dangerous landslides that result in hundreds of deaths. Development of wetlands and beaches interferes with the ability of these natural environmental features to absorb floodwaters and act as seawalls. Poor land use and natural resource management, while a reflection of poverty, is also caused by government policies that subsidize unsustainable activities and practices. These policies often result from conscious decisions and trade-offs made by government officials. If the decision-making processes are opaque and corrupt, there is little that people can do to exert pressure to correct these environmentally harmful practices.

Addressing Poverty and Poor Health

Many factors that increase societal vulnerability to extreme weather, particularly in the developing world, will not be solved without a concurrent focus on alleviating poverty, improving public health and provision of secure land tenure in safe sites. Poor people frequently build homes through the informal housing sector, making building codes and land use restrictions meaningless. They often settle on marginal lands such as steep slopes and flood-prone areas. Poverty is also a factor in environmental degradation practices such as deforestation. Vulnerability reduction has clear links with economic development and public health.

REFERENCE

IPCC (1996) *Climate Change 1995: Impacts, Adaptations and Mitigation of Climate Change: Scienti c- Technical Analyses*, Contribution of Working Group II to the 2nd Assessment Report of the IPCC, Cambridge University Press, New York, 1 879.

WHO (World Health Organisation)

see WHO (World Health Organization) (**Volume 4**)

Wind Power Development

Derek Taylor
Open University, Milton Keynes, UK

Harnessing winds for mechanical power purposes is thought to have originated in China or Persia in ancient times. Windmills became a key form of mechanical power in Europe; the number peaked at over 300 000 two centuries ago. Denmark began the rst tentative steps towards using wind power to generate electricity in the last decade of the 19th century. There were periodic attempts to develop wind power in the early part of the 20th century, but it was not until the 1970s and 1980s that wind power began to be used seriously (Taylor, 1996). By the end of the 1990s, wind turbines had become a successful mature technology capable, in suf ciently windy areas, of generating clean electricity at a cost approaching that of electricity from conventional power stations.

Most of the technical issues have been addressed, and clusters of wind turbines known as wind farms have become established as types of power stations (Figure 1). Most wind farms have used wind turbines rated between 300 and 600 kW but commercial wind turbines are increasing in size and are now available with ratings of 1 MW, 1.5 MW, 2 MW and 2.5 MW. Over 1001 MW class machines have been installed in Europe (Taylor, 1999).

Figure 1 Typical wind farm. (*Source:* photograph by Dr Derek Taylor 1998)

The global potential for electricity production from land based wind power has been estimated to be approximately 20 000 terawatt-hours year^{-1} (Van Wijk *et al.*, 1991) or about twice the 1987 global electricity consumption of 9000 TWh. If this level of generation was achieved and displaced electricity from coal fired power stations, emission of about 17 billion tons of carbon dioxide (CO_2) per year would be avoided. The European Wind Energy Association proposed ambitious new targets for wind power development in Europe in 21st century of 8000 MW by 2000, 40 000 MW by 2010 and 100 000 MW by 2020.

Because wind turbines do not emit any pollutants, the environmental impacts are generally positive and when substituted for fossil fuelled electricity power plants, wind turbines will result in major reductions in air pollution, thus benefiting the local environment. There are also reductions in acid rain and in emissions of CO_2 the major greenhouse gas. Each kWh produced by a wind turbine will avoid 860 1000 g of CO_2 if substituting for a coal fired power station. Wind power is a rapid and cost effective means of achieving cuts in greenhouse gas emissions; it can be installed quickly and economically and is a key technology for combating climate change.

As well as improving the quality of forest and water ecosystems by offsetting the use of power stations producing acid rain, it is possible that wind turbines can actually help to enhance their immediate local ecosystems and environments and may help to increase the biodiversity at their location sites (Percival, 1998) and offshore wind turbine platforms appear to be able to increase fish stocks by acting as reefs. Wind power development may also be able to protect wilderness areas or valued landscapes because landowners of such areas will be able to receive a sustainable income from wind turbines on their land, while leaving 99% of it unused, thus reducing the need to develop or plough up the land for economic benefit. A land use activity capable of generating a sustainable form of revenue is likely to be of growing economic, environmental, social and employment importance.

In operation, inland conventional steam thermal power stations lose substantial amounts of water in their cooling systems by evaporation, so substituting wind turbines will help to reduce this water use and thus be able to assist strategies for water conservation.

It is these environmental benefits, coupled with progressively reducing costs, which is helping to drive the industry which generated sales approaching some 2 billion US dollars in 1998, and is considered to be the fastest growing energy source world wide with a 35% increase in the installation rate achieved in 1998 compared to 1997 (Flavin, 1998). By the end of 1998, the estimated world wide installed capacity of wind power plants had exceeded 9600 MW (capable of generating approximately 21 TWh year^{-1}) and Germany had the largest capacity of

over 2800 MW in place. It was estimated that by the end of 1999, Germany would have a wind power capacity of over 4000 MW (Hinsch, 1999).

There are no major negative environmental impacts resulting from wind energy development but there are subjective public acceptability factors that may limit the extent of their use. The acceptability factors are not strictly environmental but are dependent on how many of the public that feel that wind turbines are attractive and/or prefer the cleaner generation characteristics of wind energy to the impacts of other methods of electricity generation, compared to the number of people that feel them to be unattractive. Unlike the environmental impacts of other technologies, the effects of wind turbines are readily reversible as they can be easily removed at a later date.

Noise from turbines was considered intrusive in some projects in the early days of wind power development, but it is becoming less of problem today as the major manufacturers are taking great care to avoid generating noise. The regulatory authorities of the countries where wind energy is becoming established impose maximum noise levels permitted to be audible at certain distances from a wind turbine.

The next big opportunity for wind energy is offshore development. Two successful wind farms are operating in the seas around Denmark and are providing valuable practical experience in offshore wind energy development, so much so that the Danish government plans to develop 750 MW of offshore wind power by 2008 and 4000 MW by 2030. Offshore projects are also operating in Sweden and the Netherlands and are planned for the UK. The offshore wind energy resource around Europe is substantial, and major energy and marine companies appear to be seriously interested in developing potential projects.

A further development underway in a number of countries (notably the UK and the USA) is that of so-called *green electricity distribution*. In the liberalized electricity supply markets in these countries, a number of companies are offering to supply electricity from assured renewable energy sources. Wind power is expected to play a key role in providing pollution free electricity to these schemes.

REFERENCES

Flavin, C (1998) *Worldwatch Brie ng " Wind Power Sets New Record in 1998: Fastest Growing Energy Source"* , Vital Signs Brief 98 7, Worldwatch Institute, December, USA.

Hinsch, C (1999) More Jobs from German Wind Power Boom, *New Energy*, September, BWE (Bundesverband WindEnergie, i.e., German Wind Energy Association), Germany.

Percival, S (1998) *Birds and Wind Turbines: Managing Potential Planning Issues*, BWEA 20 Conference, British Wind Energy Association, Cardiff, UK, September.

Taylor, D A (1996) Wind Energy, in *Renewable Energy - Power for a Sustainable Future*, ed G Boyle, Oxford University Press, UK, 267 314.

Taylor, D A (1999) Wind Energy, in *T265 Update Supplement*, eds B Everett and G Boyle, T265 UPD, Open University Press, UK, 29 34.

Van Wijk, A A, Coelingh, J P, and Turkenburg, W C (1991) *Global Potential for Wind Energy*, Proceedings of EWECO91 Conference, Amsterdam, Netherlands, 367 371.

World Energy Consumption Trends

Ian Douglas
University of Manchester, Manchester, UK

Energy is necessary for all human activities. The most important energy types are:

1. *biomass;*
2. *solid fuels (coal, lignite and peat);*
3. *petroleum products and natural gas;*
4. *electricity generated from running water and wind;*
5. *nuclear power.*

*Other sources of energy such as the tides, waves, solar energy and geothermal heat are less common, but becoming more signi cant in special situations (*see *Wave Power Development, Volume 3).*

World annual energy consumption rose from 363.74 Billion Megajoules (10^{15} J) in 1990 to 384•83 $\times$ 10^{15} J in 1995 and was predicted to rise by 2015, in a high economic growth scenario to 685•93 $\times$ 10^{15} J and to 508•22 $\times$ 10^{15} J under low growth conditions. Developed countries consume a disproportionate share of the energy supply. The US uses 25% of the world's energy output for only 5% of the world's population. The developing (non-Organization for Economic Cooperation and Development (OECD)) countries with 85% of the world's population use less than 50% of the world's energy output. However as the populations and economies of the non-OECD countries grow, their energy use is beginning to catch up to that of the industrialized (OECD) countries (Figure 1).

Projected future expansion in energy use under a high growth scenario sees coal rich countries such as India and China increasing coal production and use, while everywhere else use of natural gas would grow even faster than coal (Figure 2). High growth would lead to particularly rapid increase in oil use in countries like China (Figure 3). A

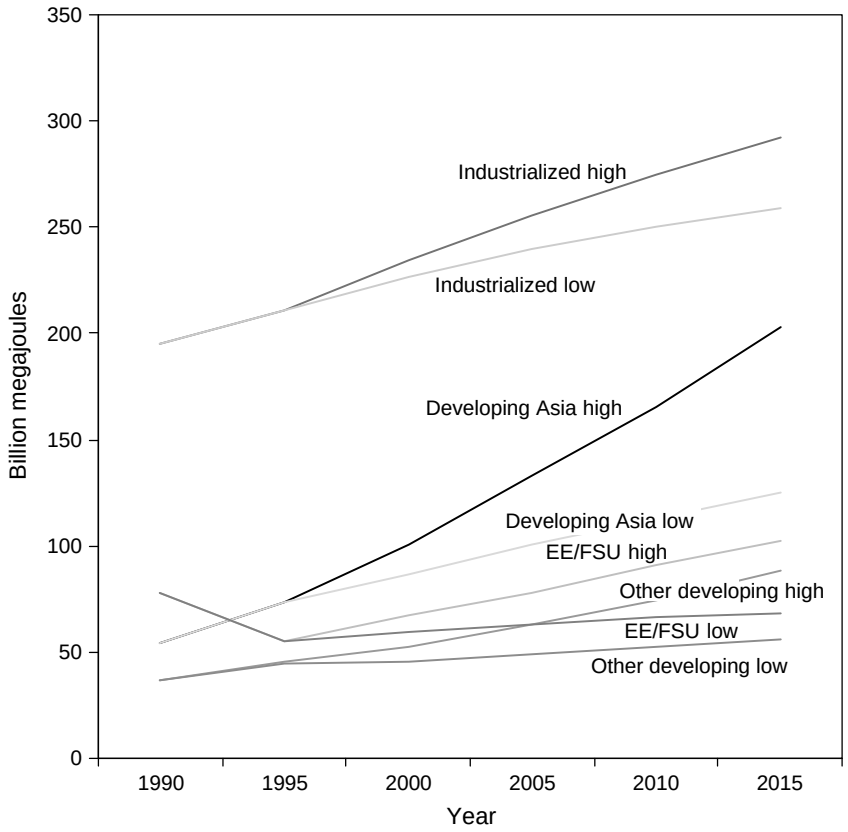


Figure 1 World total energy past and projected future use by region (based on data from International Energy Outlook, 1997) (EE/FSU denotes Eastern Europe and the Former Soviet Union)

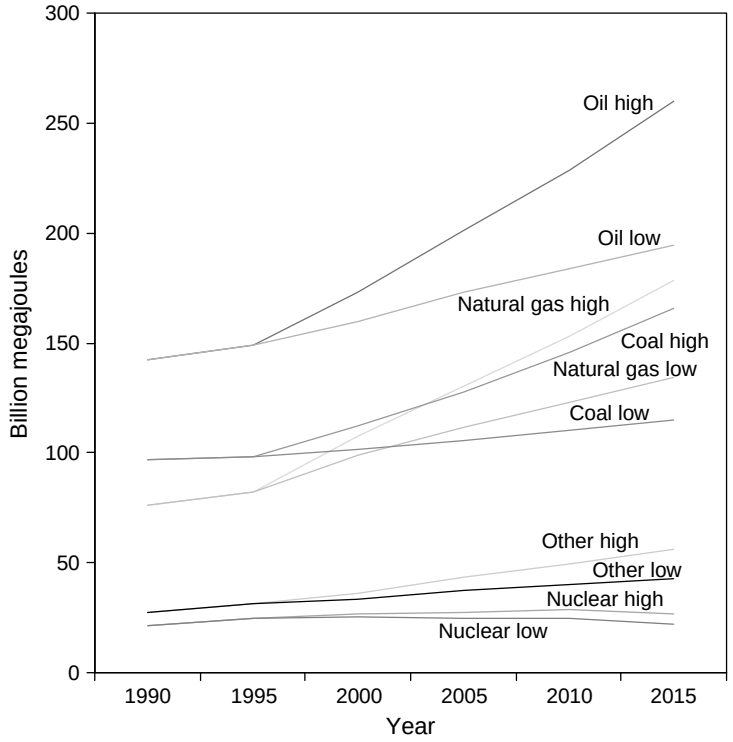


Figure 2 Past and projected future trends in global annual use of different sources of energy (based on data from International Energy Outlook, 1997)

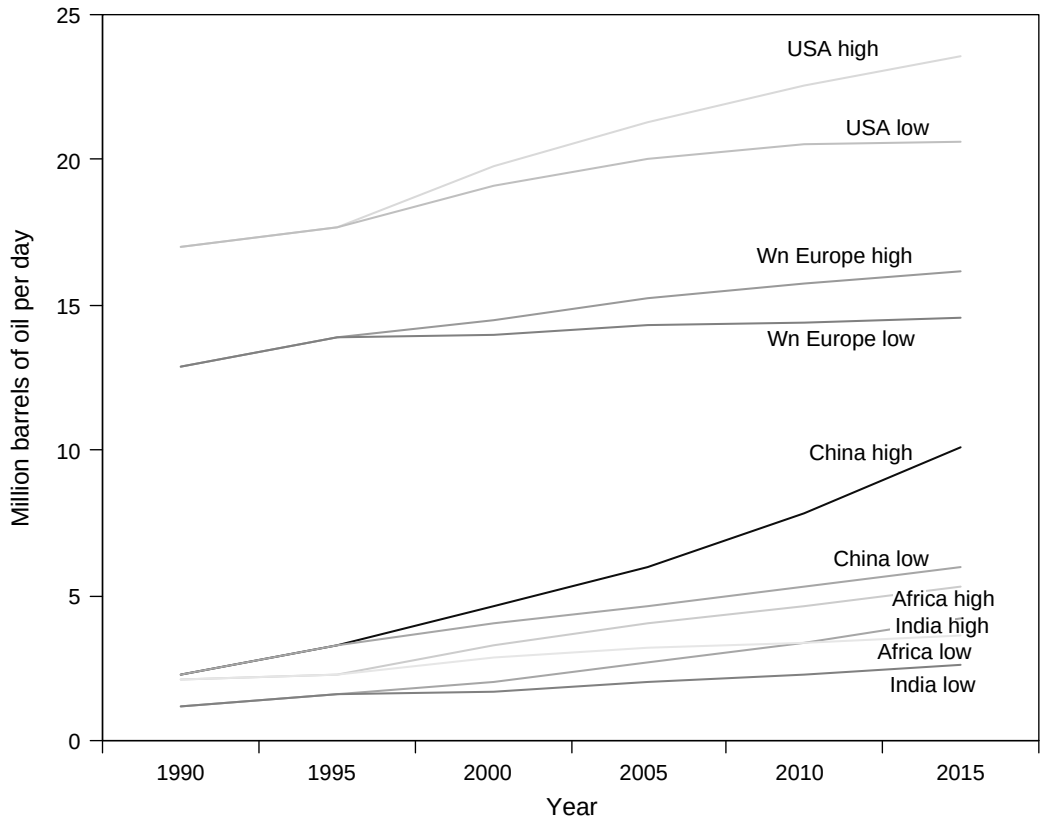


Figure 3 Past trends and future predictions of oil consumption in different part of the world (based on data from International Energy Outlook, 1997)

low growth scenario would still see a 3.4% annual growth in energy use in India and 3.1% in Mexico, compared with 0.4 in the Middle East, 0.7% in the US and 0.9% in Australasia.

PRESENT ENERGY USE

Present patterns of energy use show that the rich, industrial countries generally import energy. One extreme example is Japan, which has no oil and only a little coal and imports virtually all of its energy. Singapore, a much smaller country, is even more completely dependent on imported energy to maintain its high standard of living, yet it has a major petrochemical industry. At the other extreme are the countries of sub-Saharan Africa which have the lowest per capita fuel use rates in the world, little commercially produced energy, and a high dependence on wood fuel in rural areas and charcoal and wood in cities and towns. These countries are highly vulnerable to future shortages of biomass fuel due to the present high rate of deforestation.

Many major energy exporting countries, such as Mexico, Iran, Nigeria, South Africa, China and Indonesia, do not have particularly high per capita rates of fuel use. The wealth gained from fuel exports is not always redistributed equitably within such countries. Industrial countries use a variety of sources of energy, especially from coal, oil, natural gas, hydropower and nuclear energy. Most other countries depend on one source of fuel, China at present getting about 75% of its energy from coal, and accounting for some 25% of the world s coal production. South Africa similarly has high dependence on coal, while most countries in the Middle East and North Africa rely on oil (Rogers and Feiss, 1998).

FUTURE ENERGY USE

Projections of future energy use are difficult and there are variations in the predictions made. For example the World Energy Council s reference case shows demand rising more than 50% from 1995 2020, while those of the International Energy Agency and the US Department of Energy suggest a 34 44% rise in demand by 2010 (World Resources Institute, 1996). All three models suggest that growth in demand will be highest in Asia and Latin America. Demand will increase most markedly in developing Asia, largely as a result of rapid economic growth in India, China and Southeast Asia. The proportion of the

total world energy production used by the industrialized countries will gradually fall, but their per capita energy use will remain well above that of other countries for several decades.

A major factor in future energy demand will be the expansion in motor vehicle transport. The world currently has just over 500 million private cars, but is projected to have over 1 billion by 2020. Most of the growth in motor vehicle numbers will occur in urban areas in developing countries, although per capita vehicle ownership will remain highest in the industrialized countries for many decades. For example, there are already 4.5 million vehicles in Sao Paulo, Brazil, compared to 2.1 million in New York. The number of cars in Sao Paulo is increasing by 1000 per day. China is undergoing an unprecedented increase in motor vehicle numbers.

Petroleum is thus expected to remain the largest energy source, but the rapid expansion in natural gas powered electricity generation means that under high economic growth, world use of natural gas will begin to exceed that of coal before 2005. The exact mix of energy sources depends on the particular scenarios in each predictive model and on whether governments are assumed to have introduced measures to promote energy efficiency and the use of renewable fuels.

Practically all projections envisage that renewable energy sources (included under other in Figure 2) will play a small but growing role in the total energy supply in the next three decades. Already biomass fuels supply over 10% of global energy needs and a much higher proportion in developing countries, but with low efficiency of use. Hydroelectricity could be greatly expanded but large projects, such as the Three Gorge Dam in China, are highly controversial and schemes such a series of hydropower stations along the Mekong River in Southeast Asia, are unlikely to materialize.

Table 1 Energy related environmental degradation (after Gupta and Asher, 1998)

	Environmental degradation related to each stage of production and use			
Type	Accession and processing	Transportation	Utilization	Disposal of residuals
Petroleum	Seepage, burning of gas in oilfields, Subsidence due to withdrawal of fluids	Pipeline leakages, oceanic oil spills, permafrost pipeline problems	Air pollution, particulates from diesel fuel, emission of greenhouse gases	Management of oil wastes, risks of groundwater contamination
Solid fuels (coal, lignite, peat)	Opencast pits, landform destruction (e.g., ridge top excavations) lignite workings, destruction of wetlands through peat extraction	Coal dust	Air pollution, emission of sulfur dioxide and carbon dioxide particulates, local smog, regional acid rain, global warming	Restoration of mined land, subsidence over old coal workings, power station fly ash, mine waste dumps
Nuclear power	Uranium tailings, release of radioactive matter and heat	Risk of accidents during transport of nuclear materials and spent fuel rods	Problems of secure radioactive waste collection	High level nuclear waste disposal, restoration of areas mined for uranium; decommissioning of old nuclear reactors
Hydroelectricity	Dam and reservoir related problems			
Biomass Burning	Deforestation, soil erosion, modification of river regimes, release of air pollutants during charcoal manufacture		Household air pollution: health risks from smoke; air pollution, release of carbon dioxide and particulates; global warming	Disposal of ash, restoration of deforested areas

Projections suggest a faster rate of increase, in energy supply from renewable sources than from fossil fuels in areas like Western Europe. Here increased use of wind, solar energy and farm produced biomass is in part encouraged by government policies, such as the UK s non-fossil fuel obligation imposed on its privatized electricity distribution companies. Nevertheless, it is unlikely that such sources will account for more than 5% of world energy supplies by 2020 unless much stronger encouragement is given to their adoption.

The World Energy Council projections under an ecologically driven scenario assume deliberate fiscal and regulatory measures, as well as strong energy conservation incentives, to promote the use of renewable energy sources. With such actions and the development of new technologies, the percentage of the supply coming from renewable sources could grow to 12% of the global total by 2020. Even so, the 21st century will still be predominantly powered by fossil fuels with severe implications for future emissions of greenhouse gases.

ENERGY AND THE ENVIRONMENT

Serious environmental degradation at the local, regional and global scales is produced by the consumption of all types of non-renewable energy (Table 1). Use of renewable energy, in hydroelectricity production and biomass burning, also produces severe impacts on the environment. Dams and reservoirs may cause deterioration of river channels and of the immediately adjacent physical environment (*see* River Regulation, Volume 3). The barrages and water bodies disrupt freshwater ecosystems and movement of migratory fish. The mass of water in the largest reservoirs may place such pressure on the earth s crust that they cause earth tremors (*see* Earthquakes Triggered by Human Activities, Volume 3). Biomass burning may destroy forests and woodlands and release carbon dioxide, smoke and particulate matter to the atmosphere (*see* Low Concentration Methane Bearing Gases, Volume 3 and Biomass Fuel Power Development, Volume 3). However, the majority of the severe environmental problems caused by energy use, stem from the use of fossil fuels at all stages of the production consumption and waste disposal train. Problems arise during mining, preparation, transportation, use and residue disposal (Table 1) (*see also* Oil and the Arctic Environment, Volume 3). Normally nuclear power should only release emissions to the atmosphere in the expenditure of energy to manufacture components and to construct the reactors and generators. However, in the event of leak at a nuclear reactor, the consequences are extremely serious. Major environmental issues arise when radioactive waste has to be dealt with (*see* Nuclear Waste: Geological Issues, Volume 3).

Global fossil fuel burning was expected to grow by 32% (low growth) to 78% (high growth) between 1995 and 2015 (International Energy Outlook, 1997) and would create increases in global energy related carbon dioxide emissions of the order of 40 80%, depending on the growth scenario adopted. A considerable part of the growth in emissions would come from the accelerating growth in energy use in countries like India and China (Figure 3). However, even with that rapid growth, per capita emissions from developing countries will remain well below those of the industrialized countries (*see* National Responsibilities for Greenhouse Gas Emissions, Volume 3). If the world per capita fossil fuel burning related emissions were at the level of the average industrialized countries, world carbon dioxide emissions would be three times those already projected by 2010. The problem of emissions is extremely serious, for even if global emissions could be held at current levels, the concentrations of carbon dioxide in the atmosphere would be unlikely to stabilize for several hundred years.

REFERENCES

Gupta, A and Asher, M G (1998) *Environment and the Developing World*, John Wiley & Sons, Chichester.

International Energy Outlook (1997) *World Total Energy Consumption High Economic Growth case and Low Economic Growth Case*, http://www.eia.doe.gov/oiaf/ieo97/appa3.html.

Rogers, J J and Feiss, P G (1998) *People and the Earth*, Cambridge University Press, Cambridge.

World Resources Institute (1996) *World Resources 1996- 1997*, Oxford University Press, Oxford.

World Energy Council

The 75-year-old World Energy Council (WEC) is the world s leading multi-energy organization, with member committees in over 90 countries, including most of the largest energy-producing and energy consuming countries. The United Nations-accredited, non-governmental, non-commercial and non-aligned organization covers all types of energy, including coal, oil, natural gas, nuclear, hydro, and renewables, and is a UK-registered charity based in London. WEC s mission is to promote the sustainable supply and use of energy for the greatest benefit of all.

WEC offers a wide variety of services, programs and activities to its members, the energy industry at large and the general public. WEC is well known for its authoritative reports, analyses, research, case studies, medium and long-term energy projections, and policy and strategy

recommendations. The work of the organization spans the entire energy spectrum, focusing on such topics as market restructuring; energy efficiency; energy and the environment; financing energy systems; energy pricing and subsidies; energy poverty; ethics; benchmarking and standards; use of new technologies; and energy issues in developed, transitional, developing countries.

WEC maintains a world energy information data base for its members and has issued four seminal reports: *World Energy: Looking Ahead to 2020* in 1978; *Global Energy Perspectives 2000- 2020* in 1989; *Energy for Tomorrow's World: Realities, Real Options, The Agenda for Achievement* in 1993 and *Energy for Tomorrow's World - Acting Now!* in 2000 (World Energy Council, 2000).

WEC, which counts some of the world s major energy suppliers and power generation equipment manufacturers among its patrons, gives special attention to the need to bring reliable power supplies to the world s two billion people who still lack them. The executive assembly chairman, Jima Adam, has urged WEC to have a global conscience: we must price energy appropriately to help the poor gain access without subsidizing the wealthy, and we need to foster financing partnerships between developed and developing countries linked to environmental goals (Adam, 2000a).

Commenting on responses to the international agreements to reduce greenhouse gases and alleviate the trends towards global climate change, WEC has said:

1. developed countries, with few exceptions, have not done enough by way of domestic actions or through transfer of technology, finance and know-how to fulfill the commitments they have entered into;
2. the developed countries that achieved the aim of returning their emissions to 1990 levels by the year 2000, by a substantial margin, did so for reasons largely unconnected with steps taken to mitigate climate change;
3. developing countries, while understandably pointing to failures of developed countries to meet their commitments, have sometimes acted in ways which obstruct relevant projects and financing being undertaken jointly with investors from developed countries (World Energy Council, 2000).

In addition, WEC has advocated careful attention to the future expansion of nuclear energy and hydropower.

> Our commitment to reduce greenhouse gas emissions requires that we grow the world s nuclear power capacity as one element of our strategy. Instead of opposing nuclear power at every opportunity, the industrial world should step up development of advanced, high efficiency nuclear power plants ••• Just as with nuclear power, the world s hydro-electric capacity needs to be greatly expanded and this means developing large projects as well as small projects.
>
> (Adam, 2000b)

WEC sees these measures as an essential part of meeting its three goals of energy accessibility to all at reasonable prices, energy availability in terms of reliable uninterrupted supplies and energy acceptability in terms of the social and environmental impact of types of supply and source.

The WEC is located on the 5th Floor, Regency House, 1-4 Warwick Street, London, W1B 5LT, UK. Tel: +44 (20) 7734 5996; Fax: +44 (20) 7734 5926; Website: http://www.worldenergy.org/wec-geis/.

REFERENCES

Adam, J (2000a) *Executive Assembly Chairman's Report, World Energy Council Annual Report 1999 Committee Reports*, http://www.worldenergy.org/wec-geis/wec_info/open.plx?file=annual_report/ar99/Committee_Reports/AR99PJAdam.htm.

Adam, J (2000b) *India Energy Day, New Delhi, India, 24 November 2000*: Keynote Address by Jim Adam, Chairman of the Executive Assembly, World Energy Council, http://www.worldenergy.org/wec-geis/wec_info/open.plx?file=/structure_organisation/ea/EAkeynote1100.htm.

World Energy Council (2000) *Energy for Tomorrow's World - Acting Now*, http://www.worldenergy.org/wec-eis/etwan/open.plx?file=exec_summary/exec_summary.htm&print=1.

IAN DOUGLAS *UK*

Y

Yellow River, China: a Case Study in Water Resources

Changming Liu
Chinese Academy of Sciences, Beijing, China

The Yellow River is the second largest river in China (between 96- 119°E and 32- 42°N). The length of its main course is 5464 km from its headwater in Qinghai province to the Bo Sea in Shandong province with a drainage area of 752 443 km^2. The population is 107 million (1997) with an average population density of about 143 per km^2. The area of farmland totals 12.6 million hectare (ha), of which about 40% is under irrigation using the Yellow River's water resources. Per capita *grain production reached 352 kg in 1997. About 70% of the grain yield was produced on irrigated lands. The urban population of about 25.06 million comprises 23.4% of the total.* Per capita *gross domestic product (GDP) was 4530 Yuan equivalent to about US$550. Water resources in the basin total 71.9 billion m^3 including annual average runoff of 58 billion m^3 and groundwater of 13.9 billion m^3 while the Yangtze River has an annual average runoff of around 1000 billion m^3. The Yellow River is characterized by its hyper-concentrated sediment load. The measured annual sediment runoff during 1919- 1996 averaged 1.69 billion tonnes, which mainly comes from the Loess Plateau of the Yellow River's middle reaches.*

MAJOR PROBLEMS IN WATER EXPLOITATION

The Yellow River faces severe water shortages, as well as an increasing threat of floods and deterioration of the ecology and environment in the basin. The main channels of the lower Yellow River have dried up in 22 of the 28 years of records from 1972 to 1999, causing water shortages. In the 1990s, such events occurred every year as never before. The most severe situation was in 1997 and the duration of the dry riverbed totaled 226 and 330 days without water flowing through the river estuary outlet into the Bo Sea. The maximum length of a dry river bed in 1997 reached 704 km, about 90% of the length of the lower Yellow River (786 km), so the Yellow can be called an ephemeral stream or temporary river nowadays. Moreover, at the same time, the drying of the river has also been occurring in the main tributaries of the mid-upper reaches of the Yellow River (see Figure 1).

MAJOR IMPACTS OF THE DRYING UP CHANNEL UPON THE ENVIRONMENT

Ecologically, frequent drying up of water courses would cause a number of eco-environmental problems. The most critical one would be found in wetlands along the Yellow River s estuary and delta areas where the national wetland protected area lies with an extent of 150 000 hectare (ha), which would be influenced by the depletion of water. Obviously, this damages the biodiversity of the wetlands. Disastrously, under successively drying up courses the lower Yellow River would reduce its carrying capacity in transporting the hyper-concentrated sediment loads resulting in channel aggradation, consequently in a decrease of channeling water flows. This implies an increasing threat of floods in the areas along the lower Yellow River in the case of heavy rainstorms.

THE MAIN REASONS FOR THE DRYING-UP OF THE RIVER'S DOWNSTREAM COURSES

1. Climatically, most of the river basin lies in semiarid and arid regions. Its water balance features high evaporation and low runoff. The annual precipitation of the river basin averages only 475.9 mm while the mean annual evapotranspiration is 388.3 mm. The annual runoff is only 87.6 mm, equivalent to just 18.5% of the precipitation. Moreover, the runoff varies greatly in space and time, being highest in summer but fluctuating greatly from year to year. The lowest recorded annual

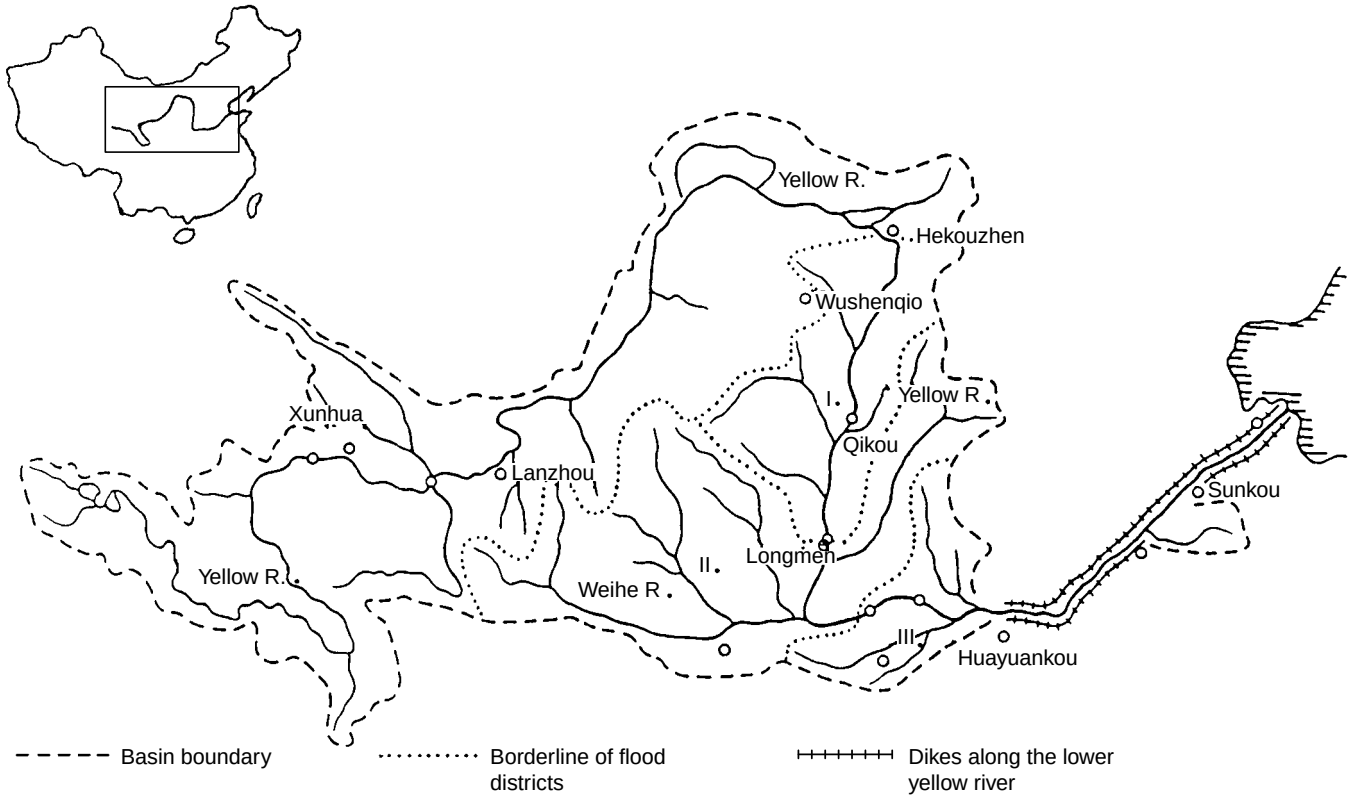


Figure 1 The Yellow River

runoff between 1919 and 1997 was only 27.5 billion m^3 in 1928, while the highest yearly runoff, in 1964, reached 93.3 billion m^3. In particular, there were successive low-water years in 1922 1932, 1969 1974 and 1990 1997. Of course, the dry years in the 1990s were the reason for the successive drying up of the river courses.

2. With intensive socio-economic development, water consumption has increased year by year, from 12.2 billion m^3 in the 1950s to about 30 billion m^3 in the 1990s, an increase of about 2.5 times. Most river water withdrawal goes to the agricultural sector, which takes about 92% of all the water used in the basin. The total installed withdrawal in the whole basin has reached 6000 $m^3 s^{-1}$ of which 4000 $m^3 s^{-1}$ is in the lower reaches, much exceeding the availability of the river s natural runoff (1840 $m^3 s^{-1}$).
3. Institutionally, water resource management is inadequate in terms of poor control of water withdrawals among the provincial authorities along the river courses. Administrative bodies control water rights; and water allocation is not adjusted in accordance with water regimes. For instance, each province takes the same amount of water from the river in both dry and wet years.

PROJECTION OF THE YELLOW RIVER'S WATER RESOURCES DEVELOPMENT IN THE 21ST CENTURY

Average annual runoff and exploitable groundwater in the basin totals 71.9 billion m^3. *Per capita* water volume in 1993 was 543 m^3 and is estimated to be 459 m^3 in 2010 and 402 m^3 in 2050. With rapid socio-economic growth, water consumption will continue to increase. Water demand in the river basin is expected to be 53 billion m^3 in 2010, 58 billion m^3 in 2030, and 60.7 billion m^3 in 2050. Accordingly, a water deficit would last till the middle of the century, when water demand will approach zero growth. The situation will become even worse if climate change brings warmer and drier conditions to the region.

MAJOR STRATEGIC CONSIDERATIONS

Based on the principle of sustainable development, three countermeasures are critical:

1. The establishment of a unified management system and formulation of a unified river basin management commission under the leadership of the state council and the setting up of an executive authority

implementing unified management, thus coordinating local administrations and water users.

2. Development of a water-saving policy incorporating water demand management as a countermeasure against wasting of water (implementation of a marketing policy including water fee regulation to protect water resources against pollution, thus maintaining the ecological equilibrium of the river basin).
3. Formulation of a special law for managing the Yellow River basin which is an extremely large unit covering nine administrative bodies of eight provinces and one autonomous region; and two administrative bodies of Hebei province and Tianjing City outside the basin. The 11 administrations have their own rights to obtain water from the Yellow River. Such water allocation was made by the state council in 1987. In addition, there are several hydro-power stations and watershed administrations with their own operational rights in the river basin. Therefore the formulation of a law to manage the Yellow River is highly advisable.

See also: Land Transformation in China: 3000 Years of Human Ecological History, Volume 3.

ACKNOWLEDGMENT

This study was supported by G1999043601.

Encyclopedia of
Global Environmental Change

egec

The volumes in this Encyclopedia are:

Volume One
The Earth system: physical and chemical dimensions of global environmental change

Volume Two
The Earth system: biological and ecological dimensions of global environmental change

Volume Three
Causes and consequences of global environmental change

Volume Four
Responding to global environmental change

Volume Five
Social and economic dimensions of global environmental change

Encyclopedia of
Global Environmental Change

Editor-in-Chief

Ted Munn

Institute for Environmental Sciences, University of Toronto, Canada

4

egec

Responding to global environmental change

Volume Editor

Mostafa K Tolba

International Center for Environment and Development, Cairo, Egypt

JOHN WILEY & SONS, LTD

Copyright © 2002 John Wiley & Sons, Ltd, Chichester
Baffins Lane, Chichester
West Sussex PO19 1UD, UK

National: 01243 779777
International: (• 44) 1243 779777
e-mail (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on http://www.wiley.co.uk
or http://www.wiley.com

Copyright Acknowledgments
A number of articles in the Encyclopedia of Global Environmental Change have been written by government employees in the United Kingdom, Canada and the United States of America. Please contact the publisher for information on the copyright status of such works, if required. In general, Crown copyright material has been reproduced with the permission of the Controller of Her Majesty's Stationery Office. Works written by US government employees and classified as US Government Works are in the public domain in the United States of America.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except under the terms of the Copyright Designs and Patents Act 1988 or under the terms of a license issued by the Copyright Licensing Agency, 90 Tottenham Court Road, London, W1P 0LP, UK, without the prior permission in writing of the publisher.

Other Wiley Editorial Offices

John Wiley & Sons Inc., 605 Third Avenue,
New York, NY 10158-0012, USA

Wiley-VCH Verlag GmbH, Pappelallee 3,
D-69469 Weinheim, Germany

Jacaranda Wiley Ltd, 33 Park Road, Milton,
Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01,
Jin Xing Distripark, Singapore 129809

John Wiley & Sons (Canada) Ltd, 22 Worcester Road,
Rexdale, Ontario M9W 1L1, Canada

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library.

ISBN: 0 471 97796 9

Typeset in 10pt Times by Laser Words Private Limited, Chennai, India
Printed and bound in Great Britain by Antony Rowe, Chippenham, Wiltshire

Editorial Board

Editor-in-Chief

Ted Munn
Institute for Environmental Studies, University of Toronto, Canada

Volume Editors

Volume One ***The Earth system: physical and chemical dimensions of global environmental change***
Dr Michael C MacCracken, *Lawrence Livermore National Laboratory, California, USA* and
Dr John S Perry, *formerly, National Research Council USA*

Volume Two ***The Earth system: biological and ecological dimensions of global environmental change***
Professor Harold A Mooney, *Stanford University, Stanford, USA* and
Dr Josep G Canadell, *GCTE/IGBP, CSIRO Sustainable Ecosystems, Canberra, Australia*

Volume Three ***Causes and consequences of global environmental change***
Professor Ian Douglas, *University of Manchester, Manchester, UK*

Volume Four ***Responding to global environmental change***
Dr Mostafa K Tolba, *International Center for Environment and Development, Cairo, Egypt*

Volume Five ***Social and economic dimensions of global environmental change***
Mr Peter Timmerman, *IFIAS, Toronto, Canada*

International Advisory Board

Dr Joe T Baker
Commissioner for the Environment, ACT, Australia

Professor Francesco di Castri
CNRS, France

Professor Paul Crutzen
Max Planck Institute for Chemistry, Germany

Professor Eckart Ehlers
IHDP/IGU Germany

Professor José Goldemberg
Universida de São Paulo, Brazil

Dr Robert Goodland
World Bank, USA

Professor Hartmut Grassl
WCRP, Switzerland

Professor Ronald Hill
University of Hong Kong, Hong Kong

Dr Yu A Izrael
Institute of Global Climate & Technology, Russia

Professor Roger E Kasperson
Clark University, USA

Professor Peter Liss
University of East Anglia, UK

Professor Jane Lubchenco
Oregon State University, USA

Mr Jeffrey A McNeely
IUCN, Switzerland

Professor Thomas R Odhiambo
Hon. President, African Academy of Sciences and
Managing Trustee, RANDFORUM, Kenya

Sir Ghillean T Prance
University of Reading, UK

Professor Steve Rayner
Columbia University, USA

Contents

Preface to the Encyclopedia of Global Environmental Change xi

Preface to Volume Four xvii

Environmental Responses: an Overview 1
Trends in Environmental Management in the Last 40 Years 15
Public-driven Response to Global Environmental Change 21
Scientific Responses in an Era of Global Environmental Change 36
Policy Responses to Public Health Issues Relating to Global Environmental Change 47
Industrial Responses to Stratospheric Ozone Depletion and Lessons for Global Climate Change 65
Emerging Environmental Issues 72
Adaptation Strategies 80
Sustainable Energy Policies 86
Dematerialization and Sustainable Development 96
Eco-efficiency 106
Monitoring in Support of Policy: an Adaptive Ecosystem Approach 116

Adaptive Environmental Management 139
Agarwal, Anil 140
Agenda 21 140
Andersen, Stephen O 143
APN (Asia–Pacific Network) 143

Basel Convention 146
Benedick, Richard Elliot 148
Blue Plan 148
Boreal Forest Carbon Flux and it's Role in the Implementation of the Kyoto Protocol Under a Warming Climate 149
Bourdeau, Philippe F J 152
Brinkhorst, Laurens Jan 153
Bruce, James 153
Brundtland, Gro Harlem 154

Capacity Building 156
Carrying Capacity 156
CITES (Convention on International Trade in Endangered Species of Wild Fauna and Flora) 157
Clean Development Mechanism 158
Cleaner Production 158
Climate Change: an Emerging Issue in the Global Agenda 163
Club of Rome 166
Coastal Zone Management 166
Common but Differentiated Responsibility Principle (Stockholm/Rio) 168
Conference of Parties 168
Convention on Biological Diversity 169
Convention on Long-range Transboundary Air Pollution (LRTAP) 173
CSD (Commission on Sustainable Development) 178
Cumulative Environmental Assessment 178

Demand Management 181
Dematerialization of the Economy 183
Desertification Convention 183
Distance Learning and Environment 186
Dooge, James 191
Dowdeswell, Elizabeth 192
DPSIR (Driving Forces–Pressures–State–Impacts–Responses) 193

EA (Environmental Assessment) 194
ECA (Economic Commission for Africa) 194
ECLAC (Economic Commission for Latin America and the Caribbean) 195
Eco Taxes 195
Ecological Dumping 198
Ecosystem Health 199
Ecosystem Integrity 199
EEA (European Environment Agency) 200
EIS (Environmental Impact Statement) 201
Emissions Trading 201
Endangered Species 201
Environmental Assessment of Major Development Projects 203
Environmental Data Report 204
Environmental Education 206
Environmental (Eco) Labeling 210
Environmental Policies for the World Oceans 212
Environmental Refugees 214
ESCAP (Economic and Social Commission for Asia and the Pacific) 218
EU (European Union) 218
FAO (Food and Agriculture Organization) 219

GATT (General Agreement on Tariffs and Trade) 221
GEO (Global Environment Outlook) 221

Geographic(al) Information Systems (GIS) 222
GESAMP (Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection) 226
Global Environmental Change and Natural Disasters 227
Global Forest Watch 231
Goldemberg, Jose 233
Goodman, Gordon 234
Green Investment 235
GRID (Global Resource Information Database, of the United Nations Environment Programme (UNEP) Division of Early Warning and Assessment) 237

Holdgate, Martin 239

IAI (Inter-American Institute for Global Change Research) 241
ICREA (International Commodity-related Environmental Agreements) 242
ICSU (International Council for Science) 242
IDNDR (The International Decade for Natural Disaster Reduction) 242
IDWSSD (International Drinking Water Supply and Sanitation Decade) 244
IFIAS (International Federation of Institutes for Advanced Study) 244
IHDP (International Human Dimensions Programme on Global Change) 245
IIASA (International Institute for Applied Systems Analysis) 246
IMO (International Maritime Organization) 246
Indicators on Environmental Quality and Sustainable Development 247
Industrial Ecology 248
Integrated Assessment, Definition of 249
Integrated Assessment 250
Integrated Assessment Models of Global Change 253
Integrated Pest Management (IPM) in an Era of Global Environmental Change 261
Intergovernmental Panel on Climate Change (IPCC): an Historical Review 265
Internalization 278
International Environmental Prizes 279
International Environmental Standards 281
IOC (Intergovernmental Oceanographic Commission) 287
IPCS (International Program on Chemical Safety) 288
IRPTC (International Register of Potentially Toxic Chemicals) 289
ISO (International Organization for Standardization) 290
IUCN (The World Conservation Union) 290
Izrael, Yuri A 291

Joint Implementation 292
Kassas, Mohamed 292

Lang, Winfred 294
LEAD (Leadership for Environment and Development) 294
Leapfrogging Technology 295
London Dumping Convention 297
LRTAP (Convention on Long Range Trans-boundary Air Pollution) 297

MacNeill, Jim 299
Materials Flow Accounting 300
Maximum Sustainable Yield (MSY) 307
MEPC (Marine Environment Protection Committee) 308
Montreal Protocol, Multilateral Fund for the Implementation of 309
Multi-issue Assessments 316

National and Local Responses to Desertification 319
National Self-interest: a Major Factor in International Environmental Policy Formulation 323
Natural Hazards: Social and Policy Responses 328
No Regrets Principle 334

Obasi, G O P 335
OECD (Organization for Economic Co-operation and Development) 335
Ozone Layer: Vienna Convention and the Montreal Protocol 337

Policies for Sustainable and Responsible Fisheries 343
Policies for Sustainable Forests: Examples from Canada 351
Policies that Promote Sustainable Energy Futures 356
Policies to Achieve Sustainable Agriculture 358
Policies to Achieve Sustainable Tourism 362
Policy Exercises: a Tool for Solving Environmental Policy Dilemmas 366
PPP (Polluter Pay Principle) 368
Precautionary Principle 369

Quotas in International Environmental Negotiations 371

Risk Management in an Era of Global Environmental Change 375

SCEP (Study of Critical Environmental Problems) 382
SCOPE (Scientific Committee on Problems of the Environment) 383
Seismic Risk (and Risk Assessment) 383
SMIC (Study of Man's Impact on Climate) 385
Social Assessment 386
Social Impact Assessment 387
Socially Responsible Investment 393
Soil Erosion 395
Speth, James Gustave 397

START (Global Change System for Analysis Research and Training) 398
Stockholm and Beyond 398
Strategic Environmental Assessment 403
Strong, Maurice 406
Sustainable Cities Policies 406
Sustainable Development 411
Sustainable Development of Energy Resources in Small Island Developing States (SIDS) 412
Sustainable Development Policies in Small Island Developing States 416
Sustainable Development Policy 422
Sustainable Transportation 426

Tickell, Crispin 436
Toepfer, Klaus 436
Tolba, Mostafa 437
Tradable Permits for Greenhouse Gases 438
Tradable Permits for Limiting Stratospheric Ozone Depletion 446
Tradable Permits for Sulfur Emission Reductions 450
Transferable Fisheries Quotas (TFQs) 455

ULYSSES (Urban Lifestyles, Sustainability, and Integrated Environmental Assessment) 460
UNCED (United Nations Conference on Environment and Development) 460
UNCHS (United Nations Centre for Human Settlements) 461
UNCLOS (United Nations Convention on the Law of the Sea) 462
UNCTAD (United Nations Conference on Trade and Development) 463
UNDP (United Nations Development Program) 464
UNEP (United Nations Environment Programme) 464
UNESCO (United Nations Educational Scientific and Cultural Organization) and the Environment 465
UNFPA (United Nations Population Fund) 466
UNIDO (United Nations Industrial Development Organization) 468
United Nations Conferences on the Environment 469
United Nations Convention to Combat Desertification (UNCCD) 474
United Nations Environment Programme (UNEP) Regional Seas Programme (RSP) 476
United Nations Environment Programme (UNEP) Saskawa Environment Prize 477
United Nations Framework Convention on Climate Change and Kyoto Protocol 478
UNSCEAR (United Nations Scientific Committee on the Effects of Atomic Radiation) 485
Use of Epidemiological Data to Develop Air Quality Standards for Short-term and Long-term Environmental Health Benefits 485
Valued Ecosystem Component 493
Vavroušek, Josef 493
Villach Conferences 494
Volvo Environment Prize 495

Ward, Barbara 497
Watson, Robert T 498
WBCSD (World Business Council for Sustainable Development) 498
WCED (World Commission on Environment and Development) 499
WFC (World Food Council) 500
White, Gilbert 500
WHO (World Health Organization) 501
World Commission on Environment and Development (WCED) 503
World Conservation Union (IUCN) Red List of Threatened Species 508
Worldwatch Institute 509
WRI (World Resources Institute) 510
WTO (World Trade Organization) 510
WWF (World Wide Fund for Nature) 511

CASE STUDIES

Regional Responses to Global Environmental Change 515

Seas

Baltic Sea 517
Black Sea 521
Caspian Sea 524
Mediterranean Sea 528

Freshwater

Aral Sea 534
Great Lakes Region of North America 536
Lake Victoria 539
Nile River 542

General

Climate Change Assessment, United States 544
Dobr s European Environment Assessment Process 547
Eco-engineering to Promote Ecological Sustainability in China in an Era of Global Environmental Change 548
MINK (Missouri, Iowa, Nebraska, Kansas) Study 550
RAINS (Regional Air Pollution Information and Simulation) 552
Transboundary Water Resource Management: The Example of the Canada/US IJC (International Joint Commission) 553

Alphabetical List of Articles 557

Preface to the Encyclopedia of Global Environmental Change

THE ENVIRONMENT

The word *environment*, whose dictionary meaning is simply that which surrounds, has in the last few decades become a buzzword , encompassing an exceedingly diverse array of elements and social issues. Taking the original meaning as a point of departure, it is clear that we humans depend totally on the environment provided by planet Earth for the food we eat, the water we drink, and the air we breathe. Thus changes in this environment must be of vital concern. Will Earth continue to sustain humans in a way that also encourages the ourishing of the other living things with whom we share the planet? This question has loomed ever larger as it has become more evident that human activities have been inducing major changes in all of the compartments of the global environment. We have converted forests and savannas to farms and cities; we have exhumed ancient treasures of fuels and minerals; we have used the rivers and winds as convenient sewers; and we have released entirely new chemical compounds and organisms into the environment. In the 1960s, the scientific community began to use the word *environment* in this new non-specialist sense. Soon too, Departments of the Environment were created by many governments, and new scientific journals began publication while others were re-named. For example, the *International Journal of Air Pollution* became *Atmospheric Environment*! In the ensuing decades, the world community has come to see the environment in many different ways, as a life-support system, as a fragile sphere hanging in space, as a problem, a threat and a home.

GLOBAL ENVIRONMENTAL CHANGE

A broader and deeper understanding of the global aspects of environmental concerns emerged in the 1970s and 1980s, and a new phrase *global environmental change* acquired popular currency. Paleoresearch had revealed that environmental change was far from new, and by no means the sole result of the actions of heedless humans. Since the planet's formation, virtually every element of its environment has been undergoing massive changes on all space and time scales. Oxygen waxed and carbon dioxide waned in the atmosphere. Continents moved about the planet's surface like scum on a soup kettle. Great ice sheets grew and shrank. Above all, the force of life (the biosphere) emerged as a dominant driver of planetary change.

However, another vital insight began to emerge about 1980: the inescapably interlinked nature of these many environmental changes. On the very longest time scales, continental drift moved lands into different climates – but also changed the climate of the globe itself. Photosynthetic life changed the atmosphere – but also made possible more advanced life forms that could take advantage of the new environment. On shorter time scales, atmosphere and ocean often interact to produce the massive changes in the Southern Pacific that we term El Nino and La Nina, whose consequences extend across the planet, and profoundly affect even our socioeconomic systems. Indeed, we have come to realize that human-induced perturbations in the environment are becoming increasingly large, and are potentially coming to dominate the natural workings of the complex and interdependent global system that sustains life on Earth. Humans and their global environment are no longer independent; they are ever-increasingly becoming interdependent components of a single global system.

Thus, the term *global environmental change* has come to encompass a full range of globally significant issues relating to both natural and human-induced changes in Earth's environment, as well as their socioeconomic drivers. This implicitly includes concerns for the capacity of the Earth to sustain life that have motivated the development of studies of global change and sustainable development in the last few decades. Analyses of global environmental change therefore demand input from the social sciences as well as the natural sciences (and indeed also from the engineering and health sciences) – necessitating an inescapably interdisciplinary approach.

Scientists from many disciplines have been attracted to this growing field of global environmental change. This is particularly noticeable in the biological sciences through the encouragement of IGBP (International Geosphere–Biosphere Programme), which invites ecologists to expand their field of vision from the plot and landscape scales to the regional, continental and global ones, and to interact with scientists from other disciplines in exploring environmental change at these larger scales. Indeed this trend has encouraged publishers and scientific societies to introduce new journal titles, and these publications from all accounts appear to be ourishing.

At the same time, human social, economic, and cultural systems are rapidly changing under the in uence of growing globalization. In the economic sphere, for example, today's discourse centers on multinational corporations, the Global

Economy, and the Globalization of Trade. Thus, the agendas of most of the physical, life, and social sciences increasingly focus on global-scale changes.

THE EVER-CHANGING ENVIRONMENT

Since the launching of Earth-viewing satellites, we have been able to see a constantly updated moving picture of our planet. Great weather systems sweep around the globe, waves of green and brown ebb and ow over the continents with the seasons, as do white waves of ice and snow. Nevertheless, a remarkable feature of the Earth system has been its relative stability. Since the dawn of life, the planet's environment has remained within a range of conditions that has supported life. Moreover, with some notable and perhaps cautionary exceptions, changes from year to year, decade to decade, millennium to millennium, have been modest, particularly during the last 10 000 years. Many features of human society exhibit similar behavior. Civilizations and cultures evolve slowly for the most part in response to environmental, economic, and social driving forces over their long lives. Great cities like Rome endure through the ages. A citizen of ancient Babylon probably would have little difficulty understanding the politics of modern Toronto.

But today we appear to be entering an era of change unprecedented since Babylon was a cluster of mud huts. There are many reasons to believe that changes greater than humankind has experienced in its history are in progress and are likely to accelerate.

Virtually every measure of human society – numbers of people and automobiles, airplanes, energy consumption, generation of waste – is increasing exponentially. While the values of a few indicators are leveling off, the magnitudes of annual increases of others remain immense.

Major changes are becoming evident in many critical elements of the environment – increasing carbon dioxide concentrations, stratospheric ozone depletion (not to mention the stratospheric Antarctic ozone hole and the possibility of a similar Arctic event), rising sea level, declining productivity of soils, widespread collapses of fisheries, dramatic declines in biodiversity, destruction of tropical forests, etc. Within living memory, the countryside around large cities has been swallowed up by suburban developments and highways. Humans are stepping on the gas pedal of the planet's environment, and perhaps recklessly breaking the survivable speed limit of global change.

Paleoclimatological studies have shown that the stable environment that we take for granted has not always prevailed. Indeed, the relatively brief period in which human civilization has developed is somewhat unusual in its equable stability. Neolithic hunters in Europe some 13 000 years ago saw their climate shift from temperate to glacial in a single short lifetime. The evolution of human society is punctuated by wars, pestilence, and technological revolutions. Major – perhaps even catastrophic – change could occur in the future, because we know it has occurred in the past.

Rapidly advancing understanding of both natural and human systems – and above all the ability to translate that understanding into quantitative models – has enabled us to explore the future of our global society and its global climate with unprecedented realism. Although prediction of the single future path that we will follow is inherently unpredictable, it is possible to map a broad range of future environmental trajectories that we might take, each completely consistent with our understanding of how the system works. Such scenario-building exercises amply confirm our concerns that the changes of the 21st Century could be far greater than experienced in the last several millennia. Business-as-usual for human society appears to imply business-as-highly-unusual for the global environment.

THE INTERLOCKING BIOGEOPHYSICAL – SOCIOECONOMIC SYSTEMS

Recognition came in the 1970s that many of the environmental issues are inter-connected through the biogeochemical cycling of trace substances, especially carbon, sulfur, nitrogen and phosphorus. In fact, in a prescient statement on the main environmental research priority of the 1980s, Mostafa Tolba (then Executive Director of the United Nations Environment Programme) and Gilbert White (then President of the Scientific Committee on Problems of the Environment) drew the attention of both the scientific and the science-policy communities to the need to understand the major global biogeochemical cycles in order to maintain the global life support systems in a healthy state (Tolba and White, 1979). Quoting from that statement,

> We draw attention to the fundamental scientific importance of understanding the biogeochemical cycles which link and unify the major chemical and biological processes of the Earth's surface and atmosphere. The results will have practical significance for all of us who inhabit an Earth with limited resources and who, by our actions, increasingly affect the quality of the human environment.

So it is that many of the global environmental issues – acid rain, stratospheric ozone depletion, climate change, nitrogen over-fertilization – are inter-related through the global biogeochemical cycles.

One of the interesting results of the study of the Earth's history has been the discovery of the global teleconnections of the Earth system, with some major climatic shifts occurring simultaneously in the two hemispheres. In an analogous way, there has been recognition for a long time that human social, economic, and cultural systems are globally inter-related. That these are connected in turn to the

Earth system was implicitly recognized as early as the 1972 Stockholm Conference on the Human Environment.

To manage the human responses to this enormously varied but at least moderately coupled world system in an era of increasing global change through the diverse array of local, national and international organizations is indeed a daunting challenge. An essential first step is to describe past and present states, then to explain the various phenomena observed, and finally to develop predictive models (or at least a range of scenarios) describing the future behavior of this total Socioeconomic–Cultural–Environmental system. In this process, environmental scientists must learn how to assimilate better the new information constantly being received, although uncertainties, often large, will continue to exist. Within this mix, rational and effective policies must be developed that will balance the risks and costs of global environmental change in an adaptive way. Only then will we be able to even formulate, much less start to implement, rational and effective policies that cope effectively with global change. The prospect of change tempts us to think in terms of winners and losers. However, such analyses often do not play out in simple ways. For example, a modest increase in rainfall would cause farmers to rejoice while vacationers would despair. However, greater farm production can lead to lower commodity prices, thereby reducing farm incomes while making vacation food purchases less expensive. Human society is pretty well adapted to the present environment, so change is necessarily a challenge.

In the longer-term, much depends on the ability of societies to respond, to adapt. Societies will differ in the resources – natural, human, and technological – that are available to them. They will also differ in terms of the values and priorities they attach to physical, social, and environmental goals, and in the social and political mechanisms that they employ to seek these goals. These differences in the human world of generations yet unborn may be as great – and as significant – as the changes in the global environment. A major challenge for our times is to develop frameworks for understanding complex interdisciplinary issues of this complexity.

No discussion of change and the future can be complete without consideration of risk. Projections of the future, however imaginative or soundly based, necessarily center on plausible, surprise-free scenarios – population will increase, economies will advance, climate will change – all typically slowly and smoothly. However, such projections made at the beginning of the 20th Century would have missed two World Wars, the automobile, aviation, space travel, television, and McDonald's in Beijing. By definition, genuine surprises cannot be predicted. However, an understanding of the impacts of past surprises may help us to make our society and our world somewhat more robust in the face of the unknown surprises that await us.

THE HUMAN DIMENSIONS OF GLOBAL ENVIRONMENTAL CHANGE

The human dimension of global change has been defined as the ways in which individuals and societies contribute to global environmental change, are in uenced by it, and adapt to it. Many empirical studies have been undertaken that describe human activities in physical terms, based on various kinds of indicators. Rates of deforestation, urbanization, and changing levels of emissions of greenhouse gases are only some measurable contributors to environmental change. But the dynamics of human activities and global change are much more complex, and they re ect the complexities of the human–nature relationship.

While the sheer burden of human activity on the planet is important, we know that the major forces at work are human beings operating together through political systems, corporations, interest groups, and beliefs that sway whole peoples. These raise questions about the nature of choice, about the hopes, dreams, and frustrations that impel people forward, or block them. At the moment, for instance, the dream of the good life – which has been a critical element of every religious tradition in the world – is increasingly being defined in terms of material possessions and powerful images that are shown on global media. Can this particular version of human happiness be sustained on a limited planet? Some people say that it can; others say that it cannot.

The central questions about the role of human beings in global environmental change revolve around social, cultural, economic, ethical, and even religious issues. These are becoming more and more pressing, and more and more foundational as human beings deliberately or inadvertently modify more and more of the planet. It is also obvious that in this generation the modification of organisms and ecosystems may well extend to the modification of human beings themselves. Among the fundamental questions are: What motivates us towards saving or harming the environment? How do we see ourselves with regard to nature? What is our responsibility to this and future generations? Who do we think we are, and who would we like to become? Appropriately enough, Volume Five of this Encyclopedia wrestles, in many different voices, with these ultimate questions that remain intimately linked with the sweep of physical, chemical, biological, geographical, and institutional changes documented and discussed throughout the earlier four volumes.

The Brundtland philosophy urges us not to reduce options for future generations. Implementation of this idea is, however, difficult. For example, it often seems more compelling to alleviate current poverty than to protect the environment and renewable resources for future generations. Many scientists agree that new approaches are needed to meld the social with the natural sciences in the policy arena. Some of the new methodologies that go beyond the physical sciences

are described in Volumes Four and Five of this Encyclopedia, e.g., post-normal science, integrated environmental assessment, and the precautionary principle.

In a commentary in which he described the 21st Century as the century of the environment, Edward O. Wilson began by referring to the growing *consilience* (the interlocking of causal explanations across disciplines) so that the interfaces between disciplines become as important as the disciplines themselves. He then stated that this interlocking amongst the natural sciences will in the 21st Century also touch the borders of the social sciences and humanities. In the environmental context, environmental scientists in diverse specialities, including human ecology, are more precisely defining the arena in which that species arose, and those parts that must be sustained for human survival (Wilson, 1998).

This is a major challenge for environmental scientists. Already through DNA techniques, for example, it is possible to trace back connections amongst prehistoric peoples through the last Ice Age to modern times.

THIS *ENCYCLOPEDIA OF GLOBAL ENVIRONMENTAL CHANGE*

More than two million papers are published every year in science and medicine, a twentyfold increase since 1940 that taxes the resources of concerned citizens, scientists, university departments, research institutes and libraries. At the same time, many policy-motivated organizations find it difficult to draw together the necessary expertise for resolving the newly emerging environmental issues. The scientific literature relating to the environment is burgeoning. However, research syntheses are in general still scattered across a broad spectrum of journals and books, and information on global environmental change is not readily available in an inter-related way. In particular, it is quite uncommon for contributions from the natural sciences to appear in the same journal or workshop proceedings as contributions from the social sciences and the humanities.

This *Encyclopedia of Global Environmental Change* is a comprehensive and integrated reference in the broad area of global environmental change that will be conveniently accessible and productively usable by students, managers, administrators, legislators, and concerned citizens. The Encyclopedia consists of five volumes of interrelated material:

Volume 1 The Earth system: physical and chemical dimensions of global environmental change (Michael C MacCracken and John S Perry, editors)

Volume 2 The Earth system: biological and ecological dimensions of global environmental change (Harold A Mooney and Josep G Canadell, editors)

Volume 3 Causes and consequences of global environmental change (Ian Douglas, editor)

Volume 4 Responding to global environmental change (Mostafa K Tolba, editor)

Volume 5 Social and economic dimensions of global environmental change (Peter Timmerman, editor)

The first four volumes cover the broad issues of the science and politics of global environmental change. Volume Five adds an often understated but extremely important aspect, linking as it does, global environmental change to the socioeconomic, cultural and ethical dimensions of human societies. Here will be found a rich panoply of writings by people who are not natural scientists but who have thought deeply about the environment. It places global environmental change in a refreshing historical, sociological and cultural context. In many contributions, the time horizons of most interest are the last hundred and the next hundred years. However, some contributions dip backward millions of years. Throughout, the emphasis is upon the dynamics of the various processes discussed – how and why did they change? A second recurrent theme is the interconnection of processes and changes – What produces change? What is impacted by change? Finally, we attempt to deal even-handedly with natural and human-induced change, and with impacts on both the natural world and human society. From the numerous diverse articles in the Encyclopedia, we believe that the user can obtain a coherent picture of this complex and dynamic system of which we all are a part.

To assist in promoting this coherence, each volume begins with a group of extended essays on major topics that embrace the field covered in that volume. These are intended to provide an introduction to the topic, a convenient road map through cross-references for exploring the Encyclopedia. Then there follows, in alphabetical order, shorter articles on a variety of scientific topics, descriptions of scientific programs, definitions, acronyms, and biographies of leading contributors to the field – from Charles Darwin, through the Russian ecologist Vernadski, to the three most recent environmental Nobel Laureates – Crutzen, Rowland and Molina. Indeed, these definitions, biographies and acronym definitions are, we believe, a uniquely valuable feature of the Encyclopedia. In the case of acronyms of international organizations and programs, it is no exaggeration to state that a young environmental scientist requires not only a good understanding of science, but also a good knowledge of acronyms, if they are to follow the discussions taking place at many international meetings! Also included in the alphabetic listings are abundant cross-references to related topics in the same or other volumes.

The substantive scientific articles that comprise the meat of the Encyclopedia are original contributions by active scientists from around the world, and thus

represent authoritative and up-to-date summaries of the state of current knowledge, direct from the producers of this knowledge. A number of these articles break new ground in synthesizing and summarizing our understanding from novel viewpoints and in unconventional ways. Thus, readers will find a wide variety of styles and approaches within the articles, re ecting in a unique way the rich diversity of today's world science. The articles have also had the benefit of careful reviews, particularly by our Editors.

The scientific essays and some of the program descriptions begin with a few italicized paragraphs written for non-specialists. These are not intended to be abstracts of the paper to follow, but rather are aimed at providing the reader with an introduction into why the topic is important and where it fits into the broader aspects of global change – a kind of encouragement to read on. Reading on should not be too difficult a task, since most of the scientific essays are written at the level of journals such as *Scienti c American* and *AMBIO* that are intended for the non-specialist. The Encyclopedia will be a valuable source of information for everyone with a general interest or a need-to-know in the various environmental fields (the natural sciences, socio-economics, engineering, the health sciences, and policy analyses) particularly as they relate to global-scale environmental change, its drivers, and its consequences. It is also expected that among the audiences for each volume will be practitioners and researchers in the fields covered by the other four volumes. We believe that this rather unusual – indeed unique – Encyclopedia will be used in a variety of ways.

1. Some people will employ it simply as a convenient source of information on specific topics – What has been happening recently to the sea ice in the Arctic Ocean? Who was Roger Revelle? What is soil mineralization? What is ICSU or IGBP? What is the Kyoto Protocol? What is deep ecology?
2. Serious students in the environmental sciences may begin by reading one or all of the introductory essays, which present highly compressed crash courses in a range of central topics.
3. We believe that the substantive specialized articles that constitute the bodies of the volumes are productively and enjoyably readable in their own right.
4. Finally, a case has been made by one of the volume editors (Peter Timmerman) for encyclopedia browsing as an enjoyable and productive pastime (*see* Encyclopedias: Compendia of Global Knowledge, Volume 5).

We believe that these five volumes are in the tradition of the human aspiration towards the compilation of global knowledge that sparked the preparation of the first encyclopedias in the 18th Century.

Our Editors have had the benefit of a marvellous postgraduate education in the course of the Encyclopedia's evolution! We are immensely grateful to the many, many authors in our virtual university who have contributed their wisdom to this project.

ACKNOWLEDGMENT

Many helpful comments and draft paragraphs were received from the Volume Editors during the preparation of this Preface.

REFERENCES

Tolba, M K and White, Gilbert F (1979) Global Life Support Systems, SCOPE Newsletter, No. 7, October. 2 pp.

Wilson, Edward O (1998) Integrated science and the coming century of the environment, SCIENCE 279, 2048–2049.

Ted Munn
Editor-in-Chief

Preface to Volume Four

Volume Four deals with human responses to global environmental change. The opening essays discuss actions that have been taken, or are contemplated, to deal with global environmental change – separately by the public, by governmental and intergovernmental bodies, by engineers in search of technological solutions, by business and industry, by the scientific community and by various sectoral interests such as the fishing, forestry and tourist industries. There follow about 100 shorter articles that elaborate on these themes. Topics covered include strategies for adapting to global change, tradable emission permits, the Biodiversity Convention, the Montreal Protocol, various tools that assist in the development of better long-term environmental policies (environmental assessment methods, environmental dispute resolution, the design of early warning monitoring systems, indicators of sustainability, etc.), the precautionary principle, and Agenda 21.

Over the past three decades, the pace of global environmental change has been increasing, and there are warning signals that this trend may accelerate. This has triggered much re ection (and action) within the environmental policy community, and there is not enough space in the Encyclopedia, let alone in this volume, to describe all of the important developments that have taken place. So we have had to be selective, while at the same time trying not to miss any of the major responses. Thus the volume covers:

1. Policies that have been developed with a view to achieving sustainability (of energy, transport, fisheries, cities, tourism, development, •••);
2. Intergovernmental and other environmental meetings that have had significant impacts on progress being made in environmental policy formulation (particularly the Stockholm Conference; the UNCED Rio Summit; Kyoto and the various world conferences organized by the UN and its Specialized Agencies);
3. The main principles agreed during the past three decades (polluter pays; common but differentiated responsibilities; no regrets principle; precautionary principle, •••);
4. Concepts (cleaner production; eco-efficiency; dematerialization; adaptive environmental monitoring strategies, •••);
5. Tools for assessing the state of the environment and dealing with environmental problems (environmental and social assessments; material ow accounting; tradable permits; eco-labelling; adaptation strategies; capacity building, •••);
6. The major International Environmental Conventions (Ozone; Climate Change; Hazardous Wastes; Biodiversity; Basel Convention; CITES; Desertification; LRTAP; Vienna Convention and Montreal Protocol, •••);
7. Examples of Actions taken to deal with regional environmental problems (Mediterranean; Black Sea; Baltic; Great Lakes; Caspian Sea; Aral Sea; Lake Victoria, RAINS, MINK). For convenience, these regional cases have been grouped together at the end of the volume, preceded by an introductory article written by Osama El-Kholy);
8. Short descriptions of organizations that have contributed significantly to the development of long-term environmental policy formulation (Club of Rome; IPCC; Commission on Sustainable Development; WBCSD; European Environmental Agency; ICSU; IPCS; Worldwatch Institute; UNEP; World Resource Institute, •••);
9. Short biographies of people who have played significant roles in the environmental response field (Gro Brundtland; James Dooge; Maurice Strong; Mostafa Tolba; Gilbert White; Anil Agarwal; Barbara Ward; José Goldemberg, •••).

We believe that in spite of space limitations, this volume provides ample information on the progress that has been made in the last 30 years in the field of human responses to global environmental change – by institutions, academia, governments, lawyers, economists and the whole array of people around the world involved in this noble task.

Mostafa K Tolba
Editor of Volume Four

Environmental Responses: an Overview

MOSTAFA K TOLBA

International Center for Environment and Development, Cairo, Egypt

Over two million or so papers are published every year in various disciplines of science, including medicine. Assessing such a massive output, is quite impossible. However, there is no doubt that understanding in all branches of science has deepened tremendously in recent decades. This is particularly true in the environmental field (Tolba *et al.*, 1992).

Once consequence of these developments has been the increasing attention given to long term futures (50 or even 100 years ahead). During the 1970s and 1980s, much was learned about the global nature of environmental systems. Particularly impressive are the insights gained into the biogeochemical cycling of elements essential for life, notably carbon, nitrogen, oxygen, phosphorus and sulfur, called the environmental life-support system by Tolba and White (1979), who have emphasized that many large-scale environmental issues – climate warming, stratospheric ozone depletion, acid rain and soil degradation – are inter-related through disruptions in the global cycles of these elements.

The last two decades have seen considerable advances in understanding the behavior of geophysical systems, and of plant and animal populations, and in the development of a capacity to model them mathematically. If changes within a system are sufficiently small for it to be expected to remain close to the range of previous experience, it is possible to develop rational environmental management plans by applying the ideas of carrying capacity of a region, sustainable yield of a renewable resource, and assimilative capacity of a watershed or airshed. These tools have developed considerably in the last 20 years. Holling introduced the idea of resilience of an ecosystem, which means that although most ecosystems can withstand environmental stresses within certain limits, they may not survive a major shock.

A related idea that has been much popularized in recent years is that of *chaos*, which is a measure of the sensitivity of a system to initial conditions. It was first studied in the context of weather forecasts by Lorenz (1963) who demonstrated that the atmosphere quickly loses its predictability.

These and other concepts have been drawn together in the development of new approaches to environmental management. The ecological insights into the dynamics of ecosystems and the critical conditions governing stability or change have led in turn to analyses of how far it is practicable to use certain systems sustainably. The concept of sustainable development embraces economic, social and ethical elements as well as ecological ones, but it is clear that development is sustainable only if it is ecologically sound (IUCN/UNEP/WWF, 1991). The intellectual challenge for ecologists is that sustainability must be achieved not simply under conditions of slow change, but also under the rapid environmental and socio-economic transformations expected in coming decades. Studies that attempt to evaluate these changes and to manage the resulting impacts are, therefore, becoming increasingly important. In this connection, the scientific problem of how to achieve re-development of degraded ecosystems has been given increased attention in the last decade.

There have of course, been major advances in many other branches of both basic and applied science, which have contributed in important ways to the intellectual context within which environmental understanding has evolved. Physics, chemistry, mathematics and geophysics have all contributed much. Within the engineering sciences, the idea of industrial metabolism and material ows has been pursued by many. Major advances in medical science have contributed considerably to human quality of life. Smallpox has been eradicated and there is now a much better understanding of other major infections and about the causes of cancer. On the other hand, the appearance of many new kinds of industrial chemicals, and rapid advances in measurement techniques have led to the appearance of a suite of new health issues relating to toxic substances such as PCBs (Polychlorinated Biphenyls) and pesticides. In fact, a new scientific discipline, environmental toxicology (or ecotoicology) has been born.

Although social scientists are making increasingly important contributions to our understanding of the environment, progress has not been as spectacular as in the natural sciences. This imbalance is gradually being corrected as the roles of human population growth, resource consumption and life-styles are increasingly recognized as major factors transforming the biosphere. It is now appreciated that it is impossible to draw up plans and strategies for environmental management without understanding the sociology and economics of the people involved. Resource planning requires models of demographic trends and resource consumption patterns. The role of special groups – women, children and the aged – are being closely analyzed. The traditional wisdom of indigenous people is also being evaluated for the insights it can provide into sustainable resource use. Developments in social science research have led in turn to the development of new techniques to promote environmental management. These range from evaluations of the environmental impacts of new policies and industrial developments, examined through environmental impact assessment (which increasingly considers economic and social factors) to combinations of regulatory and economic instruments to promote optimal resource use and minimal pollution. Such principles as the Precautionary Principle, the Polluter Pays Principle and the User Pays Principle have originated at least as much from the social sciences as from the world of pollution-control technologies.

Environmental ethics' is a term that has recently come into use to describe the principles that should guide humanity's relationship to the environment. The history of environmental ethics under other names is long, encompassing the whole range of humanity's experience and use of nature. However, the current global environmental situation has given fresh impetus to the exploration of the values underpinning modern society. Some of this is being carried out in the philosophical community, some in religious groups, and some simply among ordinary people who want to know what they ought to be doing to think globally, and act locally.

There are two different kinds of ethical approach to the environment. The less radical kind suggests that we can apply standard ethical understandings derived from philosophy or theology. This is the approach used by the International Union for Conservation of Nature and Natural Resources (IUCN), the United Nations Environmental Programme (UNEP) and the World Wildlife Fund (WWF) in their publication *Caring for the Earth: a Strategy for Sustainable Living* (1991). The second kind is more radical, and argues that a complete rethinking of our relationship to the environment is required, and that by doing that, we shall inevitably by forced to change many things we take for granted, including what we mean by ethics. These two kinds of environmental ethics have been referred to as shallow ecology and deep ecology, although both use the term ecology in a sense somewhat removed from its scientific meaning (*see* Deep Ecology, Volume 5).

RESPONSES TO ENVIRONMENTAL PROBLEMS

At the National Level

In the development of human societies, interactions with the environment were predominantly at local, and especially village, level. This is still the case in most of the world. In the 1970s, 75% of the population of developing countries was rural, although by 1980 this figure had fallen to 71%, and 63% by 1990. Agriculture continues to provide over two-thirds of total employment in many of the lower-income countries, and to provide over three-quarters of their exports (MacNeil *et al.*, 1991). However, the development process has in most countries been accompanied by the progressive expansion of the industrial sector, and urbanization has brought the growth of municipal utilities and aggregated many human impacts on the environment.

Local policies governing human interactions with the environment are as ancient as history. Technical ways of apportioning hunting rights, and controlling how land and water are used, or livestock managed, unquestionably began long before we had written records. National policies for the apportionment of land, protection of forests and control of hunting and fisheries are also ancient. This is not surprising, for there is nothing more fundamental to a nation than its land and water; and the rights to the use of such resources must inevitably have been among the first to be guarded by the central organization of a state.

Environmental policy has become increasingly cross-sectoral and integrated. While it is necessary to have specific environmental policy objectives for each of the many sectors of a modern complex state, it is also necessary to have a broad over-view and system of guidance, designed to optimize the use of the natural capital which the environment represents.

Environmental policy is developed by institutions at many levels, including local authorities, citizens' groups, environment and development non-government organizations (NGOs), religious groups and industry. All have their place and importance alongside national governmental institutions.

Another trend of the past 20 years has been the recognition of the interdependence of nations and the need for this to be expressed in regional policies and institutions that span broad areas of the environment, and for whose management a number of sovereign states need to come together. The most obvious examples are large river basins (40% of the world's population live in the catchment area of a river shared by several states), and areas of sea whose ecosystems are interlinked, and which need managing as units.

The diversity observed in policies and institutions is both inevitable and adaptive. The world's environment is

Box 1 Major international agreements

EARLY INITIATIVES

1906 Convention concerning the Equitable Distribution of the Waters of the Rio Grande for Irrigation (US–Mexico water treaty)
1909 Boundary Waters Treaty concluded between Great Britain (on behalf of Canada) and the United States
1911 Convention for the Protection and Preservation of Fur Seals
1923 Convention for the Preservation of the Halibut fishery of the Northern Pacific Ocean and the Bering Sea
1931 Convention on the Regulation of Whaling
1940 Convention on Nature Protection and Wildlife Preservation in the Western Hemisphere
1946 International Convention for the Regulation of Whaling
1951 International Plant Protection Convention
1954 Convention for the Prevention of Pollution of the Sea by Oil
1958 Convention on Fishing and Conservation of the Living Resources of the High Seas
1959 Antarctic Treaty
1964 Agreed Measures on the Conservation of Antarctic Fauna and Flora
1963 Treaty Banning Nuclear Weapon Tests in the Atmosphere, in Outer Space and Under Water
1967 Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, including the Moon and other Celestial Bodies

LATER INTERNATIONAL AGREEMENTS

Species and Habitat

1971 Convention on Wetlands of International Importance Especially as Waterfowl Habitat (Ramsar Convention) IUCN/International Waterfowl and Wetlands (IWRB)
1972 Convention Concerning the Protection of the World Cultural and Natural Heritage (World Heritage Convention) UNESCO
1973 Convention on International Trade in Endangered Species (CITES) UNEP
1979 Convention on the Conservation of Migratory Species of Wild Animals (Bonn Convention) UNEP
1992 Biodiversity Convention (UNEP)
1994 UN Convention on Control of Desertification
1980 Convention on the Conservation of Antarctic Marine Living Resources (CCAMLR) – CCAMLR Secretariat
1991 Protocol to the Antarctic Treaty on Conservation

Marine Pollution

1972 Convention on the Prevention of Marine Pollution by Dumping of Wastes and Other Matter (London Dumping Convention) (IMO)
1973 International Convention for the Prevention of Pollution from Ships and the Protocol of 1978 Relating Thereto with Annexes (MARPOL 73/78) (IMO)
1982 UN Convention on the Law of the Sea (LOS Convention) – UN Office of Ocean Affairs and Law of the Sea (OALOS)
1990 International Convention on Oil Pollution Preparedness, Response and Cooperation (IMO)

Hazardous Substances

1986 Convention on Early Notification of a Nuclear Accident (International Atomic Energy Agency (IAEA))
1986 Convention on Assistance in the Case of a Nuclear Accident or Radiological Emergency (IAEA)
1989 Convention on the Control of Transboundary Movements of Hazardous Wastes and their Disposal (Basel Convention) (UNEP)

Pollution of the Atmosphere

1979 Convention on Long-Range Transboundary Air Pollution (LRTAP) United Nations Economic Commission for Europe (UN/ECE)
1985 Helsinki Protocol on the Reduction of Sulphur Emissions or Their Transboundary Fluxes by at least 30%
1988 Sofia Protocol Concerning the Control of Emissions of Nitrogen Oxides or Their Transboundary Fluxes
1985 Vienna Convention for the Protection of the Ozone Layer (Vienna Convention) (UNEP)
1987 Montreal Protocol on Substances that Deplete the Ozone Layer (Montreal Protocol) (UNEP)
1992 UN Framework Convention on Climate Change

International organizations depend for their effectiveness on the commitment of governments (and in a few cases a wider membership) to support them. All have governing bodies representative of their members, in which policies are laid down. Many have national committees or institutions with which they have forged close links. The international and national institutional structures have thus evolved together.

diverse: peoples have adapted to it culturally and have adopted environmental policies that are an expression of sustainable living under local conditions.

The depletion of natural resources led belatedly to the adoption of international agreements and regulations in the shape of regional fisheries agreements and the International Whaling Convention. The other parallel trend has been the extension of controls by coastal states outward to sea so that strict national jurisdictional rules could be applied, thus restricting open access to the resource. This has been a classic pattern for fisheries in most parts of the world, with the extension of territorial waters from the original 3 mile (5 km) limit to 12 miles (19 km) and the more recent establishment of exclusive economic zones (EEZs) extending 200 miles (320 km) from land. In parallel, there have been developments of controls in the shape of net sizes and other regulations on fishing gear, together with seasonal regulation and the establishment of sanctuary areas.

On land, resource utilization has largely been regulated by grants of land ownership, either to individuals or to groups. Where there was communal ownership, elaborate policies backed by enforceable rules evolved early in history. Water management in arid regions has been subject to tight codes of practice from very early times, for reasons that are obvious given the limiting role of water in agriculture in such locations. In parts of Northwestern Europe, grazing rights in common pasturage have often been apportioned in terms of a right to particular farms within a village to graze a specified number of animals on the shared pasture, a particular unit of grazing carrying the right (for example) to graze one horse or one cow or three sheep. In Africa, pastoral practices have also been subject to unwritten, but elaborate, traditions.

These traditional resource-use policies were both adaptive and sustainable as long as the number of people seeking to exploit the resource did not rise dramatically. Many of the policies, however, have broken down under the combined pressures of population growth, the development of technologies that have permitted more intensive use of the resource, or the superimposition of alien systems, often in conjunction with colonial expansion.

The evolution of environmental policies up to 1970 was sectoral, reactive and uneven. In many countries, it was intimately linked to the granting of rights to certain sections of the population, often to the exclusion of others. Sometimes the rights of one sector to damage the environment in the pursuit of personal land had pre-eminence over the general interest of the community to be provided with clean air, safe water supplies and a productive environment. In many countries, industrial development and urban growth assumed priority over the protection of air, water and land and the safeguarding of agricultural interests. Many cities were permitted to expand over prime farmland, and the cash value of land approved for urban and industrial development was commonly much greater than that for agricultural use, re ecting the increased potential earnings from the land under the new pattern of use rather than its intrinsic worth to the community. The need for systems of economic evaluation that truly re ected the service to the community provided by wild land and agricultural land had not been recognized, and economic cost-benefit equations and the judgement process behind permission for land conversion tended to favor the urban and industrial sectors and the short-term over the long-term.

Institutions for environmental management also grew sectorally and sporadically. By 1970, many developed countries had departments of government that were concerned with aspects of environmental management notably agriculture, fisheries, forestry, town and country planning, industrial development, transport policy and central government's responsibilities for local and municipal affairs. At local level, there was commonly devolved responsibility for land management and development control within the area of the authority's interests, and for the control of local sources of pollution. In developing countries, a number of Ministers of Health had responsibilities that included environmental matters. However, coherent environmental policies based on monitoring and with guiding strategies generally did not exist in any country.

There was a burst of activity just prior to the 1972 Stockholm United Nations (UN) Conference on the Human Environment. Sweden established a National Environmental Protection Board in 1969, while the United States established its Council on Environmental Quality and its Environmental Protection Agency (EPA) in 1970, under a National Environment Policy Act (NEPA). The United Kingdom established a Royal Commission on Environmental Pollution and a Department of the Environment in 1970, and Canada established its Department of Environment in the same period. In 1971, Japan established its Environmental Agency and France established a Ministry for the Environment and the Protection of Nature. It was not, however, until after the Stockholm Conference that Departments of Environment and cross-sectoral coordinating machinery for environmental affairs were established in many countries. The same pattern also characterized national law. Among the Organization for Economic Cooperation and Development (OECD) nations, only four major national environmental laws were passed in the five years 1956–1960, 10 in 1961–1965 and 18 in 1966–1970, but between 1971 and 1975, 31 such measures were adopted (Holdgate *et al.*, 1982).

The period up to 1970 was also characterized by a widening in uence of what can broadly be termed *ecological thinking* on national policies, laws and institutions and a growth in citizens' groups concerned with environmental issues. These private organizations – the ancestors of

today's worldwide non-governmental environmental movement – were founded to focus on very specific issues such as the protection of birds and the safeguarding of sites of outstanding importance for their natural beauty and wildlife. The vast proliferation of environmental NGOs, many of them with wide cross-sectoral interests, and their expansion to cover international affairs, has been a post 1970 phenomenon.

These trends were undoubtedly stimulated by increasing concern in developed countries over the adequacy of the world's environmental systems to support mounting human populations, and the recognition in the 1960s of the unforeseen environmental impacts of modern industry, and especially pesticides (Carson, 1962). However, at the time of the Stockholm Conference, these concerns were only beginning to reach developing countries, which had their own preoccupations, especially with the pollution of poverty.

In 1972, therefore, the national situation around the world was one of extreme unevenness in policy, law and institution. While there was worldwide recognition of the need for productive agriculture and forestry and the protection of fisheries, in most countries these had not been given an environmental label. Environmental protection and pollution control were largely a concern of northern developed countries. A few of the latter had established Departments of the Environment and national environment protection agencies, backed by special councils to establish national targets across the whole environmental field but these countries were in the minority. The scientific base for environmental policy was fragmented, and there were few national reports on the state of the environment and few national monitoring systems. Indeed, the first real urry of such national reports was provoked by the Stockholm Conference itself, which received 80 such documents (McCormick, 1989).

Trends in the Industrialized Countries

The emergence of environmental policy in the industrialized countries over the past 25 years has been characterized by the following main features.

- *A recognition of the need for sectoral policies to take a greater account of environmental impacts and constraints*. Agricultural policies for example, have been broadened to include much tighter controls over pesticides, fertilizers and farm wastes. Energy, industry and transport policies have all devoted increasing attention to reduction in emissions to air and water, and to the handling of hazardous wastes. Environmental impact assessment has been a feature of development control in most developed countries.
- *An increasing recognition of the need for cross-sectoral policies*. Links between environment, energy, and transport policies have been strengthened. Pollution control policies have become cross-sectoral, seeking the best practicable environmental option.
- *Replacement of a largely reactive approach to pollution control by an anticipatory and preventive one*. There has been an increasing focus on pollution abatement at source. The precautionary principle and the polluter pays principle are more and more widely recognized. Screening of potentially hazardous substances before their introduction to the general market has become more stringent partly as a reaction to the unforeseen impacts of persistent organochlorine pesticides and chloro uorocarbons. The concept of *critical loads* (defined as the maximum amount of pollutant an ecosystem can withstand without stress) has increasingly come to guide policy.
- *Increasing interest in economic instruments as an incentive to energy and pollution control*. This development recognizes that while strict regulation remains necessary, within a market economy such instruments are often the most efficient way of bringing about the necessary environmental improvement.
- *Increasing promotion of energy ef ciency, energy conservation, and environmentally sound processes in industry, transport and domestic environments*.
- *Recognition of the international, and often regional, nature of many environmental problems*.
- *Increased attention to public information and participation*. It has become evident that the most effective way of ensuring the success of environmental policies is to involve all sectors of the community in the discussion of their objectives, and to inform the public about the state of their environment and the success of the actions being taken to protect it.
- *Better environmental science and monitoring*. These are the foundations of sound policy and public understanding. Through them, the loop between policy, implementation and effect can be closed. Due to the international dimension of many environmental problems, monitoring is increasingly subject to international coordination, and linked into global programmes.

During the 1970s and 1980s, as in the preceding decades, unforeseen accidents continued to force sudden adjustments in policy. These were one cause of the action early in the 1970s to prevent the pollution of the sea from the dumping of toxic materials and through accidents involving large oil tankers. In the 1960s and 1970s, there had been a prevailing view of the ocean as an effectively limitless disposal system, but it was clear by the time of the Stockholm Conference that this was not a resource that should be relied on. Since that date a whole series of national and international

Table 1 Investment in environmental protection in OECD member countries, 1970–1985

	As percentage of GNP	As percentage of all investment
Public sector	0.13–1.17	0.24–2.87
Public sector	0.08–0.86	0.28–1.62
Total	0.82–1.52	1.28–3.15

measures curbing marine pollution have been implemented. Many of the latter are regional in scope.

Disasters at Flixborough, Bhopal, Seveso, Basel and Chernobyl led in turn to the demand for more stringent action on the safety of industrial plants, better protection and higher health standards for workers in industry, more certain containment of hazards, close controls over the movement of hazardous waste and improvements in the safety of nuclear reactors, all of which were re ected in both national and international responses. These events were a salutary reminder of the fact that risk is inherent in many complex industrial operations and there was, therefore, a need for commitment to improving risk assessment and to the judgement of new policy on the basis of its results.

As a result of these evolutions in policy, there has been increased investment in the developed countries in three particular spheres: environmental protection itself (Table 1), governmental investment in environmental research (Figure 1), and public information.

The volume of public information has increased enormously. Most developed countries have published government policy statements about the state of the nation's environment, and the intentions of the national authorities in this field.

A number of countries have established independent bodies to act as evaluators of the state of the environment or of particular environmental problems, publishing their reports for the information of the community at large.

A final important development in policy has been the requirement in many OECD countries that there should be public access to information on the environment held by public authorities, especially including the results of monitoring imposed by government. The difficulty has lain in striking an appropriate balance between the public interest and the need for industry to maintain commercial confidentially over technical advances that secure market advantages. The main categories of information released into the public domain have been the quantities and nature of hazardous materials held by particular industrial plants, and the quantities of discharge to the environment of specified substances (Tolba *et al.*, 1992).

Public concern, expressed through the Green Consumer Movement regarding the environmental friendliness of products, has led to a further policy of requiring the labeling

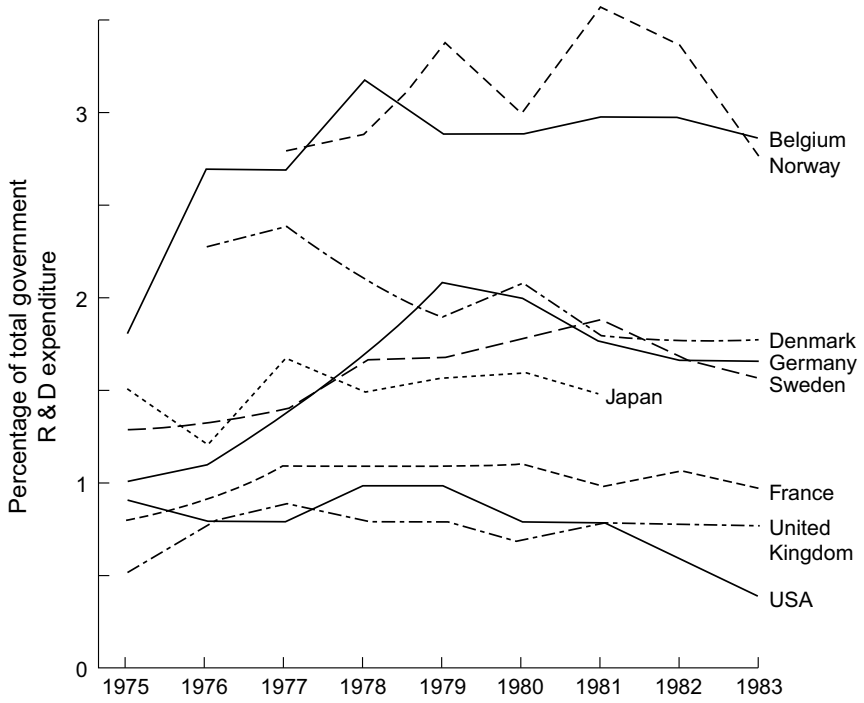


Figure 1 Government expenditure on environmental research and development in selected countries, 1975–1983. These expenditures refer to pollution abatement, but do not include expenditures concerning pollution through changes in production processes themselves. (Data were obtained from OECD, 1985)

of consumer goods to include information that will allow a prospective purchaser to judge whether or not they meet criteria of environmental acceptability. Labeling schemes, pioneered in Germany and extended in Canada and Japan, are now being adopted in the Nordic countries, Australia, New Zealand, the European Community and elsewhere (*see* Environmental (Eco) Labeling, Volume 4).

Laws and Instruments

There has been a major evolution of environmental law in the industrialized countries between 1970 and 1990. Table 2 gives some general statistics. The surge in environmental legislation between 1950 and 1970 appears to have been followed by a more measured process from 1970 onwards. The character of the legislation has changed to bring about a more integrated and cross-sectoral series of policies, and increasingly to apply within the territory of individual countries' international obligations entered into on a regional or global basis.

The emergence of an increasing body of law and regulation has been paralleled by changes in the way in which legal instruments are interpreted and enforced in a number of countries. The public has acquired an increasing right to challenge the way in which government departments and official agencies carry out their functions. This move to make public authorities accountable to the public at large, through the courts, for the achievement of national policy goals in the environmental field is a development that has clear implications for the future.

During the 1960s and 1970s increasing numbers of countries have recognized that economic instruments can, within market economies, be an effective means of improving the environment and then maintaining high environmental quality. By 1988, 153 different economic instruments were reported to be in use in the OECD countries, and 31 were of other kinds. User and administrative charges were commonest, and included fees for waste collection and sewerage services, while administrative charges were usually for licences to release substances to the environment. Not all of these charges covered the full economic cost of the activity in question, some simply recovering the costs of the administrative activities of the licensing authority. Among the other economic instruments in use were charges added to the cost of projects in order to encourage recycling, for example of oil, batteries and containers (OECD, 1991). Differential taxes, though not a direct charge, have been found useful in in uencing consumer behavior at an individual and household level. For example, Denmark, Finland, Germany, The Netherlands, New Zealand, Sweden, Switzerland and the United Kingdom impose different levels of taxation on leaded and unleaded gasoline, while the Netherlands, Sweden, Japan and Germany use taxation as an instrument to promote low pollution vehicles.

The evolution of natural *resource accounting* as a base for economic instruments that will in turn promote sound use of natural resources, was an active phenomenon in many industrialized countries in the late 1980s. This new development has been a re ection of the fact that gross national product and related aggregate economic accounts do not include the degradation of natural resources.

Table 2 Environmental laws adopted by OECD member countries, 1950–1984

Type of legislation	1950–1959	1960–1969	1970–1979	1980–1984
General	0	2	8	7
Water	3	6	16	4
Wastes	0	1	13	2
Air	1	6	8	4
EIA	0	2	8	1
Others	0	3	11	5

Institutions

Institutional responsibility for the environment in industrialized countries is commonly divided between the national and the state/district/municipal/local authority levels. Moreover, there are commonly divisions, at both these levels, between departments with overall responsibility for policy and agencies that have operational responsibility for setting technical standards, granting operational consents, levying charges, monitoring performance and prosecuting infringements.

The governmental sector has no monopoly of institutions concerned with the national environment. During the last three decades there has been an increasing emergence of self-regulation institutions within industry, linked to the development of new and environmentally compatible technologies in which sectors of industry have commonly cooperated. A third institutional development has been the massive growth in NGOs, usually in the form of groups whose business it is to demand high standards of environmental performance, protect particular regions or interests, or draw attention to inadequacies of performance.

Toward the end of the 1980s there was increasing concern within governments over the inadequacy of having sectoral Departments of Environment established alongside others, when it was apparent that the environment was actually the natural capital of nations, and the foundation of quality of life and prosperity (WCED, 1987).

Many governments have, therefore, established new machinery to better coordinate the activity of the sectoral departments.

In the industrial field, Chambers of Commerce, national business councils and specialized industrial organizations are playing an increasing part in developing environmental policy. Some such organizations are concerned with promoting products perceived to be superior in energy use

or environmental impact, while others seek to represent industrial sectors in dialogue with government.

Trends in Developing Countries

The database on national action within the developing countries is far less extensive than for the developed countries. However, as a broad generalization the developing world follows a pattern similar to that of the developed countries, but with a different range of concerns and on a different time-scale.

The primary environmental concern in the developing world is obviously the management of land and fresh waters in order to maintain the production of food and other essential resources, and the exports which are the mainstay of these economies. Development is clearly essential in order to improve the quality of life, eliminate poverty and support the infrastructure needed in order to deliver the health care, education and other institutions essential to the national future. It is, however, increasingly recognized that development processes must be based on care for the environment, and this is well illustrated by the large number of developing countries that have prepared substantial national conservation strategies (in the form of national environmental plans for the sustainable use of their resources). Many such countries have also provided reports on the state of the environment, statements of environmental policy, and in some cases considerable volumes of statistics (Tolba *et al.*, 1992).

In most countries in the Asia–Pacific region, the primary thrust of environmental policy is mostly the maintenance of agricultural and forest productivity, often in the face of pressures from mounting population growth and a con ict between the requirement for maximum revenue in the short-term and the protection of interests of the future. However, a number of countries have enunciated very broad statements of environmental policy and several are becoming increasingly preoccupied with managing industrial growth. For example, the integrated national planning of the economy, including environmental planning, its development through national land-use planning and the associated planning of environmental pollution control, have all been stressed as policy goals in China.

Similarly, Indonesia has stated its commitment to development based on the sustainable use of the country's resources, with improved planning and management of the environment, better use of human resources and natural resources such as land and water, and more attention to the environmental dimension in land and resource-use practices. Similar broad policy statements are evident in, e.g., national reports from the region to the UN Conference on Environment and Development.

Environmental laws in this region are diverse. Some countries have a relatively wide-ranging and sophisticated series of laws (e.g., India, Malaysia, The Philippines and Sri Lanka) while others (e.g., Brunei, Maldives and the Pacific Island nations) have rather less developed environmental legislation. Institutions in the region are similarly diverse.

However, as a general summary, ministerial and departmental authority for environmental interests now exists in most of the states of the region. A considerable number have national environmental protection agencies which are increasingly concerned with the overall planning of the management of natural resources. Few, if any, of these existed in the early 1970s. The cross-sectoral integration of these interests is, however, far from easy, especially given the complexities of the environmental situation in the region. The rural level, and especially the village communities, are often the key entities in environmental management, and yet it is not easy to involve them in overall policy strategies to develop land and water use systems that are built on local resource availability, people's own knowledge and traditions, local skills and needs and ecological circumstances are, however, now being developed.

A corresponding increase can be seen in the number, membership and effectiveness of NGOs concerned with environment in this region. Asia, and South Asia in particular, has the largest number of NGOs in the developing world. It also has a more active NGO involvement in policy making than other regions.

Environmental policy in the African countries has been especially concerned with agricultural development, the management of the publicly-owned forest estate, the management of wildlife resources, especially for tourism and above all the acceleration of the development process.

Africa is an immensely diverse continent, with a wide range of human traditions, climates and vegetation types. A number of African countries have developed national conservation strategies or national environmental plans. However, overall national conservation strategies and environmental plans are relatively few and far between. Those that do exist display a strongly sectoral structure within government, with very little cross-sectoral machinery yet established, despite the call for integration in many conservation strategies and environmental plans. This is a challenge for the future, which a number of governments are beginning to address.

Environmental law in Africa generally follows the same sectoral pattern. In a number of countries, however, separate laws put in place by sectoral ministries have been linked together in over-arching legislation, and some new laws have been introduced.

In many African countries environmental institutions also take the form of sectoral departments, though this too is changing. By the mid 1990s, 37 out of 51 countries on the continent had established Ministries of Environment, Natural Resources or Nature Conservation/Protection. In the remainder, responsibility for environment and natural

resource management remained with other, usually sectorally based, ministries or agencies. Most of these are concerned with agricultural, forestry, lands and natural resources, tourism and wildlife or variants on these themes.

During the last three decades there has also been a rapid growth in NGOs and citizens' groups dealing with environmental issues. An umbrella organization for African NGOs working in the environment, the African NGOs Environment Network (ANEN), was formed in 1982 with a membership of 21 organizations. This number increased more than 10-fold in the first six years of its existence, and more than doubled again in the next two years. In 1990, the membership was 530 organizations in 45 countries. The ANEN database of NGOs working in the field of environment and development in Africa in 1990 listed over 3700 NGOs. Although African NGOs are involved in the same range of activities – from advocacy and policy development to grass-roots conservation and development efforts – a higher proportion is involved in practical environment and development work than among their counterparts in other regions. A 1989 list produced by ANEN names 39 nationally organized indigenous NGOs undertaking conservation and rehabilitation work in respect of one or more of the following areas: deserts and arid land; rivers and lake basins; forests and woodlands.

Many countries in Latin America and the Caribbean have published national development strategies during the past three decades, although few (apart from Costa Rica) have produced national conservation strategies and environmental plans.

Environmental profiles have been produced by five English-speaking countries as a first step in resource assessment, and as a base on which full environmental strategies might be developed. The process of establishing cross-sectoral policies, in countries where environmental damage has occasionally been caused by the pursuit of sectoral goals, remains an urgent need. These reviews and reappraisals have already led to some extremely striking changes in national policy.

There is a considerable diversity of legal instruments bearing on environmental resources in the Latin America and Caribbean region, but this tends to be fragmented and sectoral, and there are few overall national bodies of environmental law which come together to provide an integrated framework for the protection of the environment and the sustainable management of its resources. Specific legislation has, however, dealt with the protection of forests, the safeguarding of National Parks and other protected areas, the regulation of settlements, and the protection of air and water from pollution.

Many countries in Latin America have a decentralized system, each with a Regional Council on Economic and Social Planning (CORPES). Each of these bodies formulates regional plans, prepares investment programs and coordinates and supervises the activity of various sectoral agencies. This structure was developed following an extensive discussion of a national environmental profile, which recognized the need for coordination of national activities, with an appropriate balance between central and regional administrations, and with the creation of a national environmental system in which the Administrative Department of Natural Resources and Environment would have a leading role in the formulation of overall strategy.

The number of NGOs and citizens' groups in the region dealing with the environment has also grown rapidly It is estimated that there are currently more than 6000 NGOs in the Latin American and Caribbean region that have goals related to environmental protection, most of them having been formed within the last 20 years.

INTERNATIONAL RESPONSES

Before World War II, few international environmental agreements existed. There were also few institutions, with the distinguished exception of the World (originally International) Meteorological Organization, founded in 1873 to coordinate world weather observations and promote the understanding of climate, and the International Council of Science which, founded in 1919 as the International Research Council, began to promote and coordinate international science. Most of the environmental agreements concentrated on protecting migratory birds and animals, managing transfrontier river basins and boundary waters, and conserving marine fish and mammals.

However, from 1944 onwards there was a surge of institution building, the Food and Agriculture Organization (FAO), the World Health Organization (WHO) and the United Nations Educational, Scientific and Cultural Organization (UNESCO) were all established between 1945 and 1948. Within a few years new agreements were concluded protecting the oceans and living marine species, governing Antarctica, and banning nuclear tests under water, in the atmosphere and outer space, and in Antarctica. In addition, the Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, including the Moon and Other Celestial Bodies, signed on 27 January 1967, obliged states engaged in studying or exploring outer space to avoid harmful contamination and adverse changes in the environment of the Earth.

World War II highlighted the importance of consultation and dialogue to identify potential sources of friction early on, and so avoid future con icts. The UN Charter not only established the UN itself in 1945, but contemplated that specialized agencies in economic, social, cultural, educational, health and related fields would be linked to it with a view to the conditions of stability and well-being which are necessary for peaceful and friendly relations among nations .

The Charter also called for the establishment of regional agreements to further these ends.

As this network of international institutions evolved, it was inevitable that specialized organizations and programmes would be founded to promote the cooperation that is essential if transboundary resources and the global commons are to be preserved. The IUCN, founded in 1948 with both state and non-governmental members, was charged with protecting natural areas and species. Others, including the International Maritime Organization (IMO), the International Labor Organization (ILO), and the WHO stemmed from efforts to protect the human environment by ensuring human health and safety.

These institutional developments have proceeded in parallel and in close association with the evolution of international policy and philosophy on environmental issues. One of the most striking aspects of this evolution has been the way in which environmental concepts and terms have entered the language of international policies and diplomacy. One possible reason for this growth may have been the perceived political neutrality of the environment, allowing environmental actions to be promoted as a bridge across doctrinal divides.

THE DEVELOPMENT OF INTERNATIONAL ENVIRONMENTAL INSTITUTIONS

The UN Conference on the Human Environment, held at Stockholm in 1972, is commonly thought to be the event where international debate on the environment began, but it was, in reality, the culmination of a considerable process of discussion. It was also a presentation of rather different points of view. The industrialized world came to discuss international solutions to pollution problems, which resulted from the growth of industrial activity. Their emphasis was on regulation and pollution control. They were also concerned with nature conservation and especially with threats to particular species (such as whales) and habitats (such as rain forests). The developing countries expressed little interest in these problems as compared with the pollution of poverty and the inefficiency of resource use caused by underdevelopment.

This divergence became a serious issue in the months leading up to the Stockholm Conference, and at one point seemed to endanger it. An important meeting that began to bridge the gap in perception was held at Founex, Switzerland, in June 1971. The report of this meeting clearly and cogently defined many of the issues that were to confront governments from both developed and developing countries. At Founex it became clear that development and environment were two sides of the same coin.

The breadth of activity within the UN system was familiar to the governmental delegations at Stockholm, many of whom were also members of non-UN institutions such as the OECD and IUCN, who were playing an active part in evaluating and responding to emerging environmental concerns. There were three main thrusts behind the Stockholm agenda:

1. The recognition that the developing nations faced massive environmental problems – especially linked to poverty – that needed to be overcome by development, while the developed countries had problems because their development had followed the wrong course.
2. The growing scientific understanding of the interrelatedness of natural systems.
3. Growing public concern over the cumulative impacts of human activities on global environments. This had been stimulated by disasters such as the 1967 wreck of the Torrey Canyon (the world's first supertanker accident) and reports of damage to wildlife through persistent contamination by DDT, mercury and other substances.

In addition there was the increasing recognition that neither individual nations, nor the North or the South acting alone, could adequately protect the global environment.

The UN family was hardly in a position to respond coherently to the Stockholm agenda. Those agencies with programmes directly relevant to the environment, such as FAO, UNESCO and WHO, presented points of view that were naturally linked closely to their formal mandates. Forest depletion was seen through the eyes of foresters, soil loss by agronomists and soil scientists, and so on. Stockholm came before the period when all applicable natural sciences became grouped as environmental science. The holistic, or systems, approach was still something being studied by a few academics. Moreover, social scientists hardly ever participated in the debate. Economic questions were discussed largely in relation to development and natural resource depletion. Government organizations were similarly sectoral (and there were, as yet, hardly any Departments of Environment). The overall approach was compartmentalized, and the diplomatic community did not regard environment as a central concern at all.

The Stockholm Conference was the culmination of a two-year preparatory process and the beginning of a new and intensified programmme. In the preparatory process, an Intergovernmental Working Group drew up plans, later endorsed by the conference and put into operation by the UNEP, for a global environmental monitoring system (GEMS), and an International Referral System for Sources of Environmental Information (INFOTERRA). Concern over marine pollution led to the negotiation of the Convention on the Prevention of the Pollution of the Sea by the Dumping of Wastes and Other Matter, finalized in London in October 1972. In parallel, conventions on wetland conservation (Ramsar), world heritage sites (Paris), and the control of trade in endangered species (CITES) (Washington) were stimulated and pressed forward.

Given the background, it is remarkable that the Stockholm Conference in fact decided on a system of environmental coordination for the UN that took 15 years for many governments to adopt for themselves. Although the resolutions were not systematically organized, and were truly global or even regional in scope, they nonetheless covered a wide spectrum of issues. The need for global information (as manifest through the establishment of *Earthwatch*) led towards global standard setting. The establishment of the UNEP, with its voluntary fund, was a major advance, and a spur both to the development of the UN system and to the adoption of a more integrated approach by governments, slowly though this proceeded. The limiting factor in the success of the UNEP machinery may well, in fact, have been the slowness in the adoption of similarly inter-sectoral approaches at the national level.

The Stockholm Conference was also a success because it created tremendous public interest in the environment and provided directives for international and national action. It began a process that linked environment inseparably with development, an issue explored further at the UN Conference on Environment and Development (UNCED or as commonly known, the *Earth Summit*) held in Rio de Janeiro, Brazil, 1992.

A comparable approach was adopted but with less success in response to mounting concern over desertification. As the Stockholm Conference closed, the 1968–1973 Sahel drought was at its height. Widespread crop failure and livestock deaths brought great hardship to people living in arid areas in five African countries. A UN Conference on Desertification, held in Nairobi in 1977, adopted a Plan of Action with 28 recommendations. Unfortunately, the subsequent response was slow and limited. Some training programmes were initiated, and a few national action plans were prepared (although even fewer were implemented). The call for contributions to a UN Special Account to support action against desertification was also disappointing. The 1977 Conference calculated that US 2.4 billion should be spent annually. However, because of opposition by a number of developed countries to the principle of special funds, only eight governments actually contributed to the account. The special account was ultimately closed, as decided by UN General Assembly (December 1989) at the recommendation of UNEP's Governing Council.

THE ENVIRONMENT PROGRAMS OF THE UN AGENCIES

Over 70 UN programs and agencies are listed in directories of the UN system. The vast majority carries out functions relevant to national economic and social development and to managing the global commons. These functions may be grouped roughly into data collection and research, policy formulation (including international treaties and soft law codes and standards), and development cooperation activities, such as training and education, technical assistance and project funding. Virtually all of these agencies and programs also contribute to global awareness by making available to the public, information describing the issues and objectives of their programs.

As the interconnections between avoiding environmental degradation, alleviating poverty and developing the human potential are more clearly understood, it has become increasingly apparent that virtually every organization in the UN system has some bearing on the environment and sustainable development. The Stockholm Conference prompted many agencies and programmes to examine their activities and highlight environmental components. Many of the major global conferences following Stockholm which explored in depth issues related to food and fresh water, science and technology, population, women, climate, renewable energy sources, desertification and other topics, led to the establishment of additional inter-governmental and secretariat organs within the UN system. Yet, in a perverse manner, the proliferation of environmental components and narrow, issue-oriented organs has increased the isolation of each, instead of enhancing an integrated, multisectoral approach to problem solving.

The Report of the World Commission on Environment and Development (WCED, 1987) and the UNEP's report: *The Environmental Perspective to the Year 2000 and Beyond* (UNEP, 1988) were a major spur to new thinking and especially to a recognition that an integrated approach was essential. They have been complemented by several significant regional responses including *From Our Own Agenda*, for Latin America and the Caribbean, *Economic Policies for Sustainable Development*, prepared under the auspices of the Asian Development Bank (1990), and a long-term perspective study by the World Bank, *From Crisis to Sustainable Growth*, dealing with sub-Saharan Africa. While these three latter documents are not specifically about the environment, they demonstrate how environmental considerations are now deeply integrated into development planning.

ENVIRONMENTAL POLICY AND LAW

Many of the UN agencies provide guidance on environmentally-sound development, and prepare action plans, strategies, principles, codes and guidelines as well as legally binding international treaties. Some of these state objectives and principles, while others provide operational guidelines and practices. The more general ones are designed for review every five to ten years, while many of the more specific ones provide institutional mechanisms to evaluate and update their specifications (*see* International Environmental Law, Volume 5).

REGIONAL INSTITUTIONS AND ACTIONS

The numbers of regional political and economic organizations that in one way or another are concerned with environment and development questions have also increased greatly since 1972. They include the original regional security arrangements that came into being after World War II as contemplated in Chapter VII of the UN Charter; the five UN regional economic commissions established directly under the economic and social council (ECOSOC): numerous regional offices of global UN agencies and programs, the regional development banks, and several other organizations and conferences that bear no relationship to the UN system at all, such as the OECD and the South Pacific Forum. This list is further swelled by the increasing decentralization of bodies such as IUCN, WWF and the growth of NGOs on the regional scale.

A primary reason for the growth in regional environmental activities has been the recognition that many severe environmental questions – such as control of transboundary air pollution, shared river basin management, and the protection and use of regional seas and their coastal areas – span national frontiers, and cannot be dealt with unless the states that share them act together for mutual benefit. Many species of animal, and especially birds, migrate or move across frontiers. Animal-borne diseases, ranging from malaria to rabies, are carried by their vectors from country to country and demand coordinated preventative measures. Cargoes of potentially hazardous substances are carried through regional seas, necessitating cooperative arrangements to guard against, and deal with, accidents.

At the regional level, policy, law and institutions have commonly evolved together. A problem, once admitted, leads to bilateral or multilateral regional discussion. Agreements then reached define the policies it has been agreed to pursue, and these require translation into national legal instruments and administrative practices, which in turn affect the nature of national institutions.

International institutions are commonly set up as a framework for the onward development of policy, for monitoring its effectiveness, and for sharing expertise and resources.

The United Nations Conference on Environment and Development (1992)

As stated earlier there has been a series of global conferences convened by the UN on various development issues, which dealt in detail with the environmental aspects related to the conference subject.

At the 20th Anniversary of the Stockholm Conference, the UN convened another major conference on environment and development in Rio de Janeiro, Brazil, 1992. The Conference squarely cemented the close relationships between environment and development and addressed clearly the issue of poverty and the role of major groups in society. The Conference was attended by the highest number of Heads of State and Government ever to attend a UN Conference.

The Conference adopted the carefully worded and agonisingly negotiated Rio Declaration. It adopted a plan of Action for the 21st century called Agenda 21, which synthesized and updated the outcome of all previous conferences and meetings. Commitments were made at the conference to work effectively towards the implementation of Agenda 21 and particularly to assist developing countries in this regard.

Five years later the UN General Assembly convened a Special Session to review the implementation of Agenda 21. The Special Session concluded that while there were positive steps towards achieving sustainable development, the situation in 1997 was worse than in 1992. Th Special Session adopted a Programme for the Further Implementation of Agenda 21 hoping to rectify the situation.

THE CENTRAL CHALLENGE

The complexity and magnitude of the environment/development *problematique* was recognized during the preparatory process for the Stockholm Conference. Twenty years later, at Rio, the level of political interest in the environment was very different. Yet, ironically, many of the essential points recognized by the participants in the Founex meeting in 1991 are still almost 30 years later, being raised in debate.

The major problem of underdevelopment is back on the agenda of the developing countries in a major way. The debate over the over-consumptive, wasteful life-styles in the developed countries has, if anything, intensified. Many people are again asking the kinds of questions posed in *Limits to Growth*, e.g., concerning the possibility, let alone the desirability, of unabated growth in energy demand and the private transport sector. In the 1970s, these questions generally related to energy resources running out. Today they focus on the limits of the environment as a sink or provider of services, but the central issue of nature's capacity to tolerate human impact remains.

The massive expansion of national and international institutions that has occurred during the past 30 years or so has produced as much sectoralization and fragmentation as synthesis. Only now is the world community really beginning to come to grip with the underlying economic and demographic forces that impede sustainable development.

It is now clear that development strategies need substantial adjustment, and that this must go far beyond the technology of environmental management to incorporate trade debt, and social infrastructure. It is also evident that environmental costs and benefits must be incorporated into the technologies and processes of development from the

initial planning stages. There is now much talk of a greater reliance on market mechanisms as a mode of environmental control. The use of economic instruments and full environmental cost accounting, largely unexplored until the mid-1980s, is the new fashion in environmental governance. Many governments are insisting that the costs of natural resources depletion must be accounted for and that those who generate pollution and waste must pay the price. Many of the 1972 Stockholm Conference recommendations were about reducing uncertainty through new research and monitoring and waiting for the resulting clarification. Now, when there are far more answers, the precautionary principle is gaining general acceptance.

What has been particularly lacking is an appreciation of the complex linkages between sectoral activities, macroeconomic policies and sustainable development and how to internalize them in policies and development strategies. This next step will require far greater collaboration among all concerned agencies, and a far greater effort to ensure that best use is made of all available resources for the promotion of sustainable development. It will also require better procedures to review and evaluate the effectiveness of on-going programs.

One can safely say that despite manifest failures and inadequacies, there have been important advances in international cooperation during the past decades. The trends toward closer cooperation in program planning, specific activities, funding and regional action are a heartening sign that the world's institutions can function as a community, while retaining the distinctive strengths of each. We now possess institutions and infrastructure for dealing with virtually every conceivable kind of environmental issue.

REFERENCES

Carson, R (1962) *Silent Spring*, Houghton Mif in, Boston, MA.

Holdgate, M W, Kassas, M, and White, G F (1982) *The World Environment, 1972- 1982*, Tycooly International, Dublin.

IUCN/UNEP/WWF (1991) *Caring for the World: a Strategy for Sustainability*, IUCN, Gland, Switzerland, 1–133.

Lorenz, E N (1963) Deterministic Nonperiodic OW, *J. Atmos. Sci.*, 20, 130–141.

MacNeill, J, Winseminus, P, and Yakushiji, T (1991) *Beyond Independence. The Meshing of the World's Economy and the Earth's Ecology*, Oxford University Press, New York and Oxford.

McCormick, J (1989) *The Global Environment Movement*, Belhaven Press, London.

OECD (1991) *The State of the Environment in OECD Member Countries*, Organization for Economic Cooperation and Development, Paris.

Tolba, M K, Kassas, El Kholy, O A, El Hinnawy, E, Holdgate, M W, McMichael, D F, and Munn, R E (1992) *The World Environment 1972- 1992*, Chapman and Hall, London.

Tolba, M K and White, G F (1979) *Global Life Support Systems, Joint Press Release of UNEP and SCOPE*, 1–2.

UNEP (1988) *Environmental Perspective to the Year 2000 and Beyond*, UN Environment Programme, Nairobi.

WCED (1987) *Our Common Future. Report of the World Commission on Environment and Development*, Oxford University Press, Oxford and New York.

Trends in Environmental Management in the Last 40 Years

OSAMA A EL-KHOLY

Arabian Gulf University, Manama, Bahrain

One of the oft-repeated slogans in management textbooks is that what cannot be measured, cannot be managed. This seems to lead – at least at a first glance – to the conclusion that the management of environmental affairs should be a straightforward business. Once we have regulations of one sort or the other, specifying desirable properties of ambient conditions, on the one hand, and discharges to the environment, on the other, this should provide a conceptual framework for environmental management. This is of course true, except that it is much easier said than done!

Here is a preliminary list of reasons why this is the case: (a) How do we define the environmentally desirable? (b) What are the boundaries of the system we are managing? (c) What are the scientific-technical and socio-economic constraints that bound the choice of management tools, and specify, in the final analysis, the optimum mix of such options? And last, and certainly not least (d) What are the feasible objectives of the management exercise?

Looking back over the path we have trodden over the last three decades in environmental management, we can see that we have been grappling, with different degrees of success and wisdom, with one or more of these questions. With hindsight, we can identify today a few significant signposts along the road, each leading to an interesting change of course, as well as to a new destination. We are still doing this and will most probably continue to do so for some time to come.

Each stretch of this tortuous road had its buzz word. Four such stages can be distinguished and these are reviewed brie y.

When Rachel Carson's book *Silent Spring*, was published in 1962, it shocked American society, and eventually the industrialized world at large (Carson, 1962). It argued that pollution, caused by harmful chemicals and radiation, was the major threat to the environment and life on the planet.

In the second chapter of the book entitled *The Obligation to Endure*, she declares that the

> ••• most alarming of all man's assaults upon the environment is the contamination of air, earth, rivers and sea with dangerous and even lethal materials. This pollution is for the most part irreversible••• To adjust to these chemicals would require time on the scale that is nature's••• And even this, were it by some miracle possible, would be futile, for the new chemicals come from our laboratories in an endless stream.

Towards the end of the chapter, she cries out

> ••• why should we tolerate a diet of weak poisons, a home in insipid surroundings, a circle of acquaintances who are not quite our enemies, the noise of motors with just enough relief to prevent insanity? Who would want to live in a world which is just not quite fatal?

A dire warning indeed and a powerful expression of agony and despair.

The chemical industry was up in arms. It went on the war path claiming that the book was based on bad science leading to wrong conclusions. A massive media campaign was organized nationwide extolling the benefits brought about by the chemical industry since World War II. The counter claim was that without these huge benefits mankind would be plunged into a dark era of pestilence and hunger. It was the public pronouncement of support by the president of the United States himself that finally upheld her warning, putting pollution at the top of the environmental agenda, at least in the industrialized countries.

Pollution was, in fact, the initial thrust of the 1972 Stockholm Conference on the Human Environment. Yet in gearing up for the world gathering, two significant developments changed the course of events, both before and during the conference itself. First, the Founex meeting (UNEP, 1981) brought into the argument an entirely new gamut of thorny problems and objectives. The environmental agenda,

i.e., the management objectives, now had to embrace such issues as poverty (dubbed the worst polluter), technology transfer and even world trade patterns. Next, came the Swedish case study blaming the severe pollution of Swedish lakes on sulfur emissions from chimneys hundreds of kilometers away, in Norway, Finland, Poland, Czechoslovakia, Germany, the Benelux countries, France and the UK. Suddenly, the boundaries of the system acquired global dimensions rather than Carson's local one. These made the deliberations of the conference quite difficult. Apart from changes in the system boundaries, the con ict between development objectives and those of environmental stewardship came out in the open. The Chinese delegation led the movement that did not consider pollution as the main problem for developing countries. Turning the argument on its head, it declared that it was prepared to live with some pollution in its fight against disease! We can see traces of this con ict in the Stockholm Declaration itself. In almost every one of the items in the preamble listing the environmental agenda, pollution comes first, followed by issues such as disturbing the ecological balance, or depletion of resources. While distinguishing between the priorities and objectives of developing and developed countries, it acknowledges the existence of a growing class of environmental problems of regional or global dimensions that require extensive cooperation among nations. However, the principle recognizing the sovereign right of states to follow their environmental policies while calling upon them to ensure that they do not cause damage to the environment of other states is number 21 in a list of 26 principles. Even in calling for the development of international law regarding liability and compensation for the victims of environmental damage in areas beyond their jurisdiction, pollution is again singled out and mentioned by name. The objectives of Stockholm were indeed a tall order. It is obvious that such diverse objectives could not have been all addressed at once. There had to be priorities.

COMMAND AND CONTROL MANAGEMENT

It comes as no surprise that pollution abatement was at the top of the agenda. Emphasis was on the treatment of emissions and ef uents to comply with certain permissible qualities. The main environmental management tool was command and control. The state now acted as the guardian of public interest, deciding on standards for permissible releases into the environment, keeping a watchful eye on compliance, and punishing the wayward for their misdeeds.

Limits set on discharges became more stringent and extended in scope to cover more types of discharges. Around the mid-1970s, a new generation of chemicals management legislation emerged calling for pre-marketing or pre-manufacturing assessments of new industrial chemicals. This proved to be practically impossible and many chemicals were in actual use while still awaiting satisfactory assessments. No wonder that relations between the regulating agencies and the regulated community became progressively adversarial and socially destructive.

In most developing countries the command and control approach failed ignominiously for obvious reasons. Ludicrously inappropriate standards copied from more developed communities were adopted. In Egypt, the first clean air legislation specified no less than 93 properties of emissions to be monitored! Government machinery was totally ill-equipped to monitor compliance or enact enforcement. By the end of the 1970s, it became obvious that not only was this approach socially expensive, but also remarkably ineffective and inefficient.

THE NEW SOCIAL DIMENSION IN THE ENVIRONMENTAL AGENDA

Pollution abatement may have been at the top of the environmental agenda at the operational level in the 1970s, but it would be misleading to leave the impression that this was the only item on the agenda. In fact there were many other initiatives, action programs, strategies, declarations and conventions, both at the regional and the international levels. One or two initiatives that went almost unnoticed, only to come back with a vengeance later on are worth noting. United Nations Environment Program (UNEP), which had just taken shape and started operating, called for eco-development and development without destruction. But these were still very much at the conceptual rather than the operational level. In 1974, the Cocoyoc symposium, organized jointly by UNEP and United Nations Conference on Trade and Development (UNCTAD), looked into the economic and social factors inherent in environmental degradation. In examining the increasing scarcity of resources it raised the issue of consumption patterns in the developed nations, relating them to environmental degradation in the developing countries. This came soon after the first Club of Rome report, *Limits to Growth*, and its almost total rejection by nearly everybody: the business community, the modeling community, and the developing countries. The business community emphasized that resource availability responds to scarcity through the price mechanism and technological innovation. The modeling community was alarmed by the crudity of the model. The Third World abhorred the notion that economic growth should be limited while two-thirds of the world lived in poverty, responding with its own, the Bariloche model. However, the model's major impact was to set many people seriously thinking about a conserver society, perhaps for the first time. Towards the end of the decade, the concept

of environmental ethics was again revived. It gave fresh impetus to the exploration of values underpinning modern society. By 1980, the debate now encompassed, amongst other things, the issue of the responsibility for future generations (Partridge, 1980). The discussion finally culminated, in the Brundtland report of 1987, as the concept of *sustainable development*.

However, these and other efforts at understanding the environment-development problematique were at the conceptual level. They had little impact on the operational aspects of environmental management, which had reached a dead end in its concentration on pollution abatement. A change of course was obviously in the offing.

FROM ABATEMENT TO PREVENTION

During the latter half of the 1970s, the concept of low and non-waste technologies (LNWT) emerged, shifting emphasis to pollution prevention rather than pollution abatement. With the benefit of hindsight, we could well wonder how it had taken us so long to remember the old adage that prevention is better than cure . But let us remember that tinkering with the innards of a functioning system is not welcome. Attention now came to be focused more and more on so-called clean technologies. However, a number of complicating considerations soon became apparent. First, there is no clear-cut distinction between the clean and the dirty. Furthermore, the issue should logically extend to the products themselves, their design, the materials used in making them, and their disposal, i.e., the assessment should encompass the whole life cycle of a product from cradle to grave (or cradle to cradle as the Wuppertal school preached later on). Finally, the very premises on which products or services are provided, and the value systems behind them, were questioned. Although this was well articulated in what UNEP labeled *cleaner production* (*see* Cleaner Production, Volume 4), the emphasis in the 1980s at the operational level was on pollution prevention, and mainly on production technologies and processes. It did not extend, at first, to addressing this hierarchy of issues. Almost every day new examples of remarkable combinations of pollution reduction and economic benefit were publicized, both in the industrialized and the industrializing countries. Almost one-third of these examples were mainly good housekeeping measures. Payback periods for their adoption were remarkably short and capital expenditure upfront was quite small and almost negligible in many cases. The European Preventive Environmental Protection Approaches in Europe (PREPARE) added an extra dimension, viz. catalyzing industry/government cooperation and the transfer of information on pollution prevention. In the USA, prevention teams helped small businesses in adopting pollution prevention technologies.

This new trend of the regulator joining hands with the regulated, rather than commanding and controlling, was a healthy development initiated by the Dutch. Covenants were concluded between the two parties, specifying what needs doing and what it takes to do it. This resulted in a welcome change in relations and in confidence-building measures. Transnational corporations (TNCs) now committed themselves publicly to quantitative reductions in polluting discharges over specific time horizons. The Triple P now stood for Pollution Prevention Pays, rather than the 1975 Polluter Pays Principle of the OECD (*see* PPP (Polluter Pay Principle), Volume 4). This came to be regarded as a win-win situation. Win-win came to be an article of faith accepted worldwide, heralding a new phase in environmental management.

Gradually, the law of diminishing returns dampened this euphoria. The idea became more and more difficult to sell; once the easier options resulting in substantial savings were exhausted, win-win became more problematic and its limitations became gradually clearer. DuPont came up with their version of the 80/20 principle – 80% of all possible environmental benefit came from the first 20% of costs. Sophisticated methods for identifying these 20% cost measures were being developed. Other social complications arose, sometimes preventing the implementation of some corporate priorities.

THE SOCIAL RESPONSIBILITY OF BUSINESS

But before we move on to the next phase we need to take note of a particular class of events and the reactions they triggered, namely, industrial accidents causing serious damage to the environment.

Bhopal, in particular, will always remain as a very significant watershed in environmental history. The accident occurred in a plant belonging to a well-known TNC in a developing country. The extent of damage and human suffering was horrendous. It suddenly became clear that response to previous accidents in industrialized countries (Flixborough or Seveso; *see* Seveso, Volume 5) had to move to a much higher level, and to take a global dimension, and also at the social, rather than just at the technical front (SCOPE, 1989).

Industry woke up to the need to improve its public image. The Community Awareness and Emergency Response (CAER) program was initiated by the Chemical Manufacturers Association (CMA) in the USA. The public was now kept informed about business operations, the hazards and risks involved, and plans for dealing effectively with emergencies. Similar programs were also initiated in Europe by the European Chemical Industry Council (CEFIC). Industry was now seriously seeking closer relations with society. The Responsible Care program, started in Canada in the mid-1980s, became adopted worldwide.

This declared the *moral responsibility* of business towards the safety and well-being of people and the environment, throughout the life cycle of its products. The chemical industry has indeed come a long way since it denounced Rachel Carson three decades earlier.

Furthermore, the International Chamber of Commerce (ICC) Second World Industry Conference on Environmental Management (WICEMII) of 1991 confirmed business commitment to the new concept of sustainable development, now receiving the hallmark of world acceptance in the Brundtland report. The ICC enshrined its commitment in a new business charter. A World Business Council for Sustainable Development (WBCSD), and even local councils, were formed.

FROM POLLUTION PREVENTION TO ECO-EFFICIENCY

Enter the third phase, that of *eco-ef ciency* (*see also* Eco-efficiency, Volume 4). This moved upstream from the tail end of pollution control to the very beginning: to better resource utilization. The conceptual forays towards the end of the 1980s, whether Ayres' *industrial metabolism* (Ayres, 1989) or, later on, Frosch's *industrial ecosystems* (Frosch and Gallopoulos, 1991) emphasized the need for dealing with all material ows in interconnected industrial systems that form part of the larger environmental system. It emphasized the importance of minimizing inputs of energy and raw materials, on the one hand, and utilizing wastes as inputs to new processes, on the other.

Business was now moving beyond compliance to a more aggressive commitment to exceeding regulatory requirements. It was now venturing into a vast arena of continuous improvement. The new buzz word was *eco-ef ciency*. Unlike pollution prevention, which focuses on reducing harmful discharges, eco-efficiency is concerned now also with material inputs. Reductions in these inputs by factors of four were being demonstrated as feasible targets through an increasing number of examples in many walks of life. A factor of 10 was now set as the target for sustainable development worldwide. Dematerialization, first mentioned in the late 1980s came out of the cold to become the new moving target to be pursued indefinitely by business (*see* Dematerialization and Sustainable Development, Volume 4). Eco-efficiency now replaced pollution prevention as the route to benefiting both the bottom line and the environment.

While eco-efficiency was clearly desirable environmentally, it was less obviously so economically. In this brave new world, the environment became more and more a cost center. The challenge was to build a business case for its implementation. Public image in an atmosphere of more vocal public opinion and aggressive environmental non-government organizations had to be factored in, as well as the impact on market acceptance and penetration measured in new ways. Environmental reports by enterprises were published, using a variety of approaches and measures of eco-efficiency. In the absence of a standard methodology and performance indicators, there was confusion often leading to skepticism. Analysis of these reports showed that eco-efficiency was not really being measured.

In 1994, an article in the *Harvard Business Review* argued against adopting this approach as the foundation of corporate environmental strategy, pointing out the danger of environmental planning being guided by wishful thinking and comforting buzz words. Greening could no longer be regarded as a win-win strategy for all corporations, all the time. It was time for a change of course.

THE AGE OF SUSTAINABLE DEVELOPMENT

By the time the Rio Earth summit was held in 1992, the tiny stream of ideas and approaches forming the environmental agenda had become a major river fed by a multitude of tributaries of new ideas and experiences, that certainly exceed in their number, and significance, the rather eclectic review given here. No wonder that Agenda 21, in its attempt to acknowledge each one of these, came out to be such a thick volume providing a rather forbidding checklist of what needs doing, together with rather dubious estimates of the cost of doing it.

Be that as it may, UNCED was particularly significant in two respects. First, the signing of two new global environmental conventions, emphasizing the urgency of the global dimension of environmental problems. It is worth noting here, in passing, that the principle of ensuring that activities within the jurisdiction and control of states do not cause damage to the environment of other states has now jumped from number 21 on the list of principles in the Stockholm Declaration to the very top of that in the Rio Declaration (*see* Sustainable Development, Volume 4).

The second was enshrining the concept of *sustainable development* firmly and squarely at the global level. No one would henceforth discuss environmental issues without at least paying lip service to sustainable development.

The question now is: how has this affected environmental management, particularly in the business community that had already committed itself to the concept in a charter and set up its own council, the World Business Council for Sustainable Development (WBCSD), even before Rio (*see* WBCSD (World Business Council for Sustainable Development), Volume 4)? Governments and the public have long been concerned with the environment. Business has now entered the field as a third and responsible player. Here are a few examples:

- Before 1992 was over, a book was published by Stephan Schmidheiny on behalf of the WBCSD

(Schmidheiny, 1992). Schmidheiny is a unique phenomenon – a successful businessman who does not mince words when advocating the responsibilities of business towards society, particularly in environmental stewardship. The book calls for profound changes in the goals and assumptions that drive corporate activities and change in the daily practices and tools used to reach them. This, for the author, means a break with business-as-usual mentalities and conventional wisdom that sidelines environmental and human concerns. He goes on to say: the painful truth is that the present is a relatively comfortable place for those who have reached positions of mainstream political or business leadership. This is the crux of the problem of sustainable development, and perhaps the main reason there has been great acceptance of it in principle, but less concrete actions to put it into practice.

Strong fiery words indeed, but how do they translate into actions?

- At the level of WBCSD, its working groups produced a plethora of reports on many topics, such as eco-efficiency, sustainable production and consumption. It now remains to see this converted into tangible operating practices on which new generations of business personnel will be trained. Corporate social responsibility (CSR) is the new buzz word that strengthens the voluntary approach, and the involvement of all stakeholders, away once and for all, from command and control. Working groups are busy thrashing this out in an attempt to formulate a working model combining everyday business activities with the well-being of society, and accepting accountability to different stakeholders. In a sense, business is taking over some of the traditional responsibilities of governments. This is a dangerous and slippery path. The WBCSD is particularly, and quite rightly, concerned about the role of each party in a new business–government–civil society partnership.

Let us first admit in all honesty, that there is as yet no worldwide consensus as to what sustainable development means. All the more reason that we do not quite know yet what it means operationally. The jury is still out and we must not be too severe in judging current efforts to interpret the term and be guided by it.

- The trouble with sustainable development is that it is abstract, complex and subversive. Business is uncomfortable with such qualities. Pollution prevention, eco-efficiency and other strategies are useful; but they in fact dodge the issue for a simple reason. They are local, straightforward issues amenable to relatively simple and fairly clear analysis and quantification of some sort or the other. Sustainable development, on the other hand, is a global and fiendishly complex issue calling for much deeper and broader involvement – something business shuns for obvious reasons. It had to bring sustainable development to a clearer and more manageable size and form. The way business has done this so far seems to have been essentially to equate sustainable development to eco-efficiency.

Stated bluntly this means that we will be less worse off tomorrow then we are today. In doing this we are in a sense doing our best within the constraints of the *status quo*. This is a far cry from what Schmidheiny is calling for.

TAKING STOCK, LOOKING AHEAD

To re ect on our experience so far will help us to clarify our understanding of how and why we arrived where we are today, to take stock of our past experience, and to base our future plans on the lessons we have learnt:

1. One striking feature is that the ideas that happen to be the fashion at any time are seldom new. We have heard them before but failed to grasp their significance. Reading today the Founex report, or even the first Meadows report, or the Cocoyoc declaration, or revising the recurrent warnings over decades of possible dramatic climate change brought about by mankind, or reminding ourselves that responsibility for future generations is much older than the Brundtland report, to cite but a few examples, recalls, as Victor Hugo once remarked, that their time had not come when they were first enunciated.
2. One reason is – and will continue to be – our incomplete knowledge of the complex natural and social issues we are grappling with. This is hardly a confidence-inspiring foundation for policy-making. We only act – or rather react – when the environment springs a surprise on us. Such extraordinary events end up improving our knowledge; but the next emergency when it comes will, more often than not, be somewhere else. No profound global change can be predicted on the basis of recognizing one specific hazard, or through one scientific discipline. We are still unable to encompass all imaginable effects or to predict all possible interactions in time and space, in order to plan meaningful interference. We now realize that environmental effects are time bombs – they are very slow and could go on practically unnoticed for decades before they explode into irreversible change.
3. Scientific discovery is not in itself a social force. We need a credible consensus amongst scientists before

the politicians start acting. Science has not always managed to establish, beyond reasonable doubt, a clear cause–effect relation. While scientific uncertainty is the driving force of scientific progress, it also paralyses decision-making. Unfortunately, some scientists become advocates, confusing the decision-making process even more.

4. Faced with this daunting challenge we have not been standing idle. If we still do not quite know the dynamics of environmental change, we have at least come up with some new ideas, e.g., the precautionary principle, environmental economics, the need to correct market deficiencies, the necessity for all stakeholders to cooperate sincerely and consistently. We will no doubt continue to improvise new approaches.

What we must never do is to forget that this is a long and hard struggle that calls for sustainability of effort as we seek to achieve the ultimate goal of economic, social and environmental sustainability.

REFERENCES

Ayres, R V (1989) Industrial Metabolism, in *Technology and the Environment*, eds J H Ausubel and H E Sladovich, National Academy of Engineering Washington, DC.

Carson, R (1962) *Silent Spring*, Houghton Mif in, Boston, MA.

Frosch, R A and Gallopoulos, N E (1991) Towards an Industrial Ecology, in *The Treatment and Handling of Wastes*, eds A D Bradshaw, R Southwood, and F Warner, Chapman & Hall, London.

Partridge, E, ed (1980) *Responsibilities for Future Generations*, Prometheus Books, Buffalo, NY.

SCOPE 40 (1989) *Methods of Assessing and Reducing Injury from Chemical Accidents*, John Wiley & Sons, Chichester.

Schmidheiny, S and the Business Council for Sustainable Development (1992) *Changing Course: a Global Business Perspective on Development and Environment*, MIT Press, Cambridge, MA.

UNEP (1981) In Defence of the Earth: The Basic Texts on Environment – Founex, Stockholm, Cocoyoc, United Nations Environment Programme, Nairobi.

Public-driven Response to Global Environmental Change

B KASEMIR[1,2], F GASSMANN[3], S F LIENIN[3], C C JAEGER[4], AND A WOKAUN[3]

[1]*Swiss Federal Institute for Environmental Science and Technology (EAWAG), Duebendorf, Switzerland*
[2]*Harvard University, Cambridge, MA, USA*
[3]*Paul Scherrer Institute (PSI), Villigen, Switzerland*
[4]*Potsdam-Institute for Climate Impacts Research (PIK), Potsdam, Germany*

Global environmental change (GEC) is directly related to the general public. On the one hand, GEC is ultimately shaped by the activities of people, be it energy- and material-intensive lifestyles in industrialized countries or poverty-driven deforestation and other land use changes in developing countries. On the other hand, it is the public which is directly affected by GEC and which suffers from living conditions changing for the worse.

Conventional wisdom suggests that two elements lead to public support of a development towards sustainability: setting the right prices for energy and other resources, and increasing public awareness of the dangers of GEC by information and education. In this opening essay, we argue that this approach by itself is not suf cient. Not only have newspaper articles like " Double the Price Without Pain: What Happened to Oil Shocks?" (International Herald Tribune, 2000) publicly put a question mark on how much the mechanisms of today's economies are in uenced by energy prices. But, more importantly, recent research on public responses to GEC discussed in this essay shows that citizens are not supportive of strongly increasing energy prices by scal measures. This is not caused by any lack of environmental awareness. The same research nds that citizens around the globe are very concerned with prospects of GEC, and strongly support preventive action. But they expect solutions to be found without hard-hitting price increases. Technological change in combination with related lifestyle changes will be essential in nding solutions to GEC problems that are supported by the public. Opportunities for such processes can only evolve in a complex network of actors in science, business, and politics together with the general public. Figure 1 illustrates this system of different actors, which is complex due to many interactions and feedback-loops (not shown in detail). The nature of complex systems suggests that new solutions, e.g., for less polluting transportation, can emerge but cannot be planned by any unique agent. Our essay focuses on the dynamics of the response to GEC evolving from the complex interplay between the public, politics, business, and science. Studying this complex system in more depth will require a combination of natural and social sciences, of quantitative and qualitative methods, and of modeling and empirical research.

In our essay, we will discuss this complex system as follows. In the section on studying public reactions to global environmental change, we highlight the importance of understanding public reactions for the formulation of sustainable development policies. We will discuss recent research on public response to GEC mainly on the basis of three major studies. These are, first the global Health of the Planet survey, second a study of US residents' views on global climate change, and third the European research project, Urban Lifestyles, Sustainability and Integrated Environmental Assessment (ULYSSES) focusing on citizen views on climate change and policy options. These three projects and their different methodologies will be discussed in order to explore the responses to GEC of citizens with different cultural backgrounds. Results of these studies indicate that most citizens, both in industrialized and developing countries, are concerned about GEC and do not accept the lack of absolute scientific proof as a reason to delay preventive action. On the other hand, they are not willing to support a major increase of energy prices to induce a reduction of greenhouse gas emissions. This implies an essential role for technological change fostered

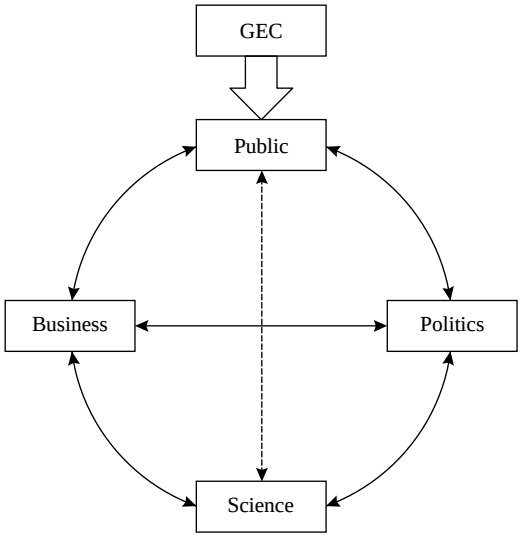


Figure 1 Response to GEC results from a complex interplay between the public and actors in business, politics, and science

by science and business, which are often perceived to have responsibility here. Their interaction with citizens and policy makers in complex social systems will be crucial for shaping technological change combined with lifestyle changes as an effective response to global environmental change.

In the section on perceptions of technological opportunities, we follow the pathway paved of the results of public opinion studies and discuss possibilities and problems of technological change and related lifestyle changes in response to GEC. We focus on questions related to development towards a sustainable energy system, which requires the application of technological innovations in both energy production and energy consumption. The emerging picture for the global energy requirements in the next hundred years clearly shows that drastic action is needed which consists not only in the development of new technologies and their implementation but also of behavioral attitudes of consumers. This can be illustrated best in the context of transportation, which has been the fastest growing energy sector during the last two decades. Urbanization, rationalization and globalization have changed living conditions dramatically leading to a fast growing demand for transportation of people and goods. Technological innovations towards more energy-efficient transport are of great importance but must be accepted by the public. This underlines the importance of investigating the dynamics of the whole system constituting the public in its double role of citizens and consumers and the other main players? politics, business and science.

In the section on complex dynamics: innovation and lifestyle changes, we discuss new approaches essential for analyzing such systems. We argue from the viewpoint of the interdisciplinary science of complex systems. This novel perception of natural systems, developed during the last three decades, has revealed basic misconceptions and misbeliefs of traditional natural scientists and bears the potential to reconcile natural with human sciences. The combination of non-linearity and feedback-loops in complex systems leads to a variety of possible dynamics which often differ from common expectations based on linear extrapolation from the present state. We will make clear that seemingly impossible but important changes in socio-economic systems, if they occur, often happen abruptly and without warning. These changes are unpredictable surprises that can be in uenced by different factors only within narrow time intervals or *windows of opportunity*. This view leaves room for hope for innovation and lifestyle changes inducing development of a sustainable future, but at the same time, sets clear limits to our possibilities to create it according to our own design. Appropriate action can only be developed in a common effort involving all actors in Figure 1. Finally, we discuss the application of some principles of the science of non-linear dynamics to an innovative process, describing the substitution of conventional, by innovative cars.

In the final section we re ect on implications of recent research on public responses to GEC on the interplay between the public, politics, business and science and we discuss how *social learning* (*see* Social Learning in the Management of Global Atmospheric Risks, Volume 5) in this complex system could be both studied and supported by future research, as an interdisciplinary effort including close dialogue between the public and other stakeholders.

STUDYING PUBLIC REACTIONS TO GLOBAL ENVIRONMENTAL CHANGE

The Importance of Public Reactions

Effective responses to GEC will have to include major changes in our socio-economic system. While some simpler environmental problems could be tackled by reducing some toxic byproducts of certain production processes, responses to climate change require large shifts right at the heart of our industrialized cultures. Referring to reduction scenarios discussed in the Intergovernmental Panel on Climate Change (IPCC) context, (Kempton, 1991) has stressed that if world leaders decide to reduce greenhouse gas emissions by two-thirds, such a large reduction will require consumer and worker cooperation as well as citizen consent that major societal changes are worth the effort . Public support is an essential requirement for any effective response to global environmental change.

This suggests that studying reactions of citizens concerning climate change and related policy options is a central element in developing effective response strategies.

Such studies can build upon a large body of research on environmental consciousness and behavior in general (see, e.g., the comprehensive overview by Brand, 1997). In the last decades, various studies have focused specifically on perceptions of climate change issues (see, e.g., the overview by Thompson and Rayner, 1998). Bearing in mind the scale of changes needed in the longer run for effective climate protection, the crucial research goal here is to understand the role of the public in supporting changes of a complex socio-economic system. This includes collective action by many players, and not just isolated individual reactions in an otherwise static social environment.

> Reducing carbon dioxide and taking further steps to curtail global warming will require collective action and institutional change (both intra- and internationally) and are unlikely to result primarily from people changing their consumption patterns on an individual basis. Indeed, one could argue that public support for environmental regulation and incentives, and for politicians who will implement them, is far more essential than voluntary change in individual consumer behaviour when it comes to achieving a global-level public good' like a (reasonably) stable climate.
>
> (Dunlap, 1998)

In this section we will thus discuss research on public responses to GEC from the perspective of what its findings imply for the complex interplay between the different actors in society. What does the public expect from science, politics and business, and what can be expected from the public? We will illustrate these questions on the basis of discussing three major studies which have recently targeted responses of the public to environmental issues, and to GEC in particular.

1. Global Responses: Health of the Planet Survey

The international Health of the Planet' survey on environmental attitudes aimed to give voice to individual citizens around the world concerning the environment and economic development issues affecting their lives (Dunlap *et al.*, 1993a,b). This large questionnaire study was carried out in 24 nations both in developed and in developing regions around the world, with sample sizes between 770 (for Finland) and nearly 5000 (for India). Overall, nearly 30 000 questionnaire responses were collected and analyzed. The questionnaires were completed in face-to-face meetings with survey researchers, thus minimizing problems, e.g., of illiteracy.

A highly interesting result of the Health of the Planet survey concerns the comparison of views on environmental issues between developed and developing countries. The researchers were surprised that the views expressed were much less polarized than has often been assumed. Citizens of both developed and developing nations were found to accept a share of responsibility for environmental problems. The most frequent response to the question who is more responsible for today's environmental problems in the world? was that industrialized and developing countries are both equally responsible (Dunlap *et al.*, 1993a). Respondents from developing nations were nearly as likely as their developed country counterparts to mention overpopulation as the major cause of environmental problems within developing countries themselves. And not much difference was found between residents of developed and developing countries concerning the degree of blame placed on consumption of the world's resources by industrialized countries (Dunlap *et al.*, 1993b). Major differences in international climate diplomacy between the two sets of countries can be expected to persist related to the question whether long-term emission rights should be based on grandfathering (favoring developed countries with high current emissions) or on *per capita* emissions (favoring developing countries with low current *per capita* emissions). However, the results of the Health of the Planet survey suggest that citizens in developing and developed countries are basically not pointing the finger of blame at each other. This should encourage policy makers in both sets of countries to look for a consensus on dealing with GEC problems.

Furthermore, a striking finding of the Health of the Planet survey was that careful analysis of the data did not support the post-materialist hypothesis that higher economic security in developed countries would generate post-materialist values leading to higher environmental concern compared to developing countries. Indeed, national wealth and environmental concern seemed to be negatively rather than positively correlated. Even if the level of environmental concern is comparable in developed and developing regions, one might have assumed that citizens in developed countries worry more about GEC than their counterparts in developing countries, which face severe local environmental degradation. However, the survey found that this is not the case. While citizens in poor nations were indeed much more likely to rate local community problems as serious than were residents of wealthy nations, they rated global environmental problems nearly as serious as did those in the wealthy nations (Dunlap *et al.*, 1993a). Concerning actions that developed countries could take to support developing countries in protecting the environment, technological assistance was often mentioned, next to providing educational information. However, the role of technology was perceived ambiguously. Respondents often saw technology both as a cause of environmental problems, and as a part of the solution. Also, the survey overall found willingness to pay somewhat higher prices if this supports environmental protection.

The Health of the Planet survey also found issues where the views of developing and developed country respondents differed from each other. This was the case, for example, concerning the questions: Which group within

a given country was mainly responsible for environmental protection? and; How much can citizens effectively contribute to solving environmental problems? While most often government was felt to be primarily responsible for environmental protection, also individual citizens and citizen groups were often seen in this role. The response that citizens should have primary responsibility for environmental protection

> is somewhat more likely in the developing than in the industrialized nations and is especially pronounced among Latin American nations. This pattern would be more distinct were it not for the abnormally low level of responsibility assigned to citizens in Russia and Poland [which Dunlap *et al.* had included in the group of developing nations], which could re ect the low levels of citizen efficacy in these formerly totalitarian nations . (Dunlap *et al.*, 1993b)

Also with regard to the question, How much effect do citizens (both individually and as citizens groups) have on solving environmental problems? respondents from developing countries were more likely than those of developed countries to think citizens can play an effective role in solving environmental problems. Again, the Eastern European nations included in the survey (Russia, Poland and Hungary) were striking exceptions in that here, citizen action was seldom seen as very effective. Overall, these findings mean that in many nations (especially in developing countries) citizens do not simply delegate responsibility to policy makers but also assume a share of responsibility for environmental protection themselves.

In a subset of the Health of the Planet survey, citizens in six nations (Canada, the US, Mexico, Brazil, Portugal, and the Republic of Russia) were asked specifically about global warming (Dunlap, 1998). While other aspects of GEC like ozone depletion, loss of rain forests and jungles, and large-scale water pollution were perceived as even more worrying, global warming was in general seen as very serious. Challenging again the post-materialist hypothesis, respondents from Brazil, Mexico and Portugal were overall more likely to see global warming as having very harmful effects than their counterparts from Canada, the US or Russia. Overall, the respondents saw global warming effects on natural systems (either by creating various ecological problems or by changing weather patterns) as more worrying than effects on social systems. Choices that people have of where to live and economic well-being are the least likely realms to be seen as being harmed – suggesting that economists' emphasis on economic impacts [•••] ignores the public's crucial concerns ••• (Dunlap, 1998).

2. US Residents' Views

In contrast to the Health of the Planet survey, a series of studies of US residents' views on global climate change has combined questionnaire survey methods with semi-structured interviews of a small but diverse sample of citizens. A first study has been based on the analysis of just fourteen ethnographic interviews (Kempton, 1991). In a further step, 43 interviews were conducted together with the collection and analysis of 142 questionnaire responses from US residents with a wide range of different backgrounds (Kempton, 1997). The semi-structured interviews were based on open-ended questions, follow-up probes for unexpected topics raised by the informants, and paraphrases for verification. Kempton suggests that such interview techniques, originally developed by anthropologists to study foreign cultures, are recommended when the interviewer initially has little understanding of the native concepts. Such methods contrast with survey research, in which the questions and range of possible answers are already known and precise answer frequencies are measured (Kempton, 1991). Such interviews are often taped and selectively or fully transcribed, so that exact quotations can illustrate the *mental models* of the interviewees as directly as possible. Thus, interview studies usually provide more in-depth insights into the perceptions of a smaller number of respondents than surveys. In the case of this study, the small number of interviews was backed up by a questionnaire study with at least a medium-sized sample.

One set of findings of this study concerns lay conceptualizations of the greenhouse effect. These were found to be rather different from scientists' descriptions. The research team reports, e.g., that lay people often confused the greenhouse effect with ozone depletion or tropospheric air pollution. In general, Kempton concludes that citizens fit new information on global warming into prior concepts or cultural models easily accessible to them, instead of receiving the information as it was meant by the scientific authors. Different conclusions have been drawn from this mismatch between citizens' and analysts' conceptualizations of global climate change. The leading researcher of this study sees the current mismatch as evidence for the difficulties in putting a new concept [i.e., climate change] on the lay person's cognitive map (Kempton, 1991). He suggests that it is vital to change this situation by continuing media coverage and education, because citizen knowledge is needed to keep leaders on the right track .

In contrast to that, Dunlap remains skeptical that public support for environmental protection will hinge upon laypersons becoming highly informed about technical aspects of global warming (Dunlap, 1998). He sees it as no surprise that lay people try to make sense of global climate change by associating it with existing and more accessible mental models like air pollution, ozone depletion, or ecological problems in general. Instead of worrying about the mismatches with expert concepts, he thinks that public concern over global climate change is unlikely to disappear in the foreseeable future precisely because it is woven into the fabric of the ecological problematique' that seems

firmly (if not deeply) embedded in public consciousness . A similar interpretation is given by Thompson and Rayner, who argue that

> For many lay people, the climate change issue is perceived as part of a wider problem concerning humankind's disturbed relationship with nature. In this sense, climate change and ozone depletion are the same thing: they are members of the category of environmental insults deriving from industrial society. If industrial society is the problem, and if climate change and ozone depletion are interchangeable symbols or prototypes of the problem, then the specific etiology of each simply may not be relevant to the respondents of these surveys.
>
> (Thompson and Rayner, 1998)

It may well be, this debate suggests to re ect on the question: To what extent are technical details of GEC relevant for citizens, or To what extent does Dunlap's analysis apply that in an era of re exive modernization, it may well be more crucial for the public to know which experts to trust and distrust than to understand the details of their respective arguments (Dunlap, 1998).

A second set of findings from Kempton's studies concerns environmental values. In agreement with the overall results of the Health of the Planet survey, Kempton and his colleagues report that US residents express strong support for protecting the environment in general and preventing global climate change in particular (Kempton, 1997). This seems based on ethical values which include the perceived intrinsic value of nature, in addition to religious reasons or the human use aspects of ecological systems. The strongest environmental value to emerge was a desire to preserve the environment for one's descendants, sometimes including concern for more than one coming generation. Also, the weather seems to be a natural system with a special cultural meaning, and people seem to have a tendency even in our technology-based society, to feel that it is wrong for humans to tamper with it (Kempton, 1991). Kempton suggests that an important part of citizen support for environmental protection is local, e.g., the desire to preserve local recreational and scenic places and to protect one's own health .

Concerning specific climate change impacts, Kempton's findings suggest that species preservation is not highly valued in the abstract by the majority of the US public. Rather, e.g., rising food costs were perceived as more worrying than species extinction, and most informants justified species preservation in anthropocentric terms. While weather patterns were seen as becoming more unpredictable, with storms and violent weather more common, most respondents connected the greenhouse effect primarily to higher temperatures rather than to the more complex climate effects expected by scientists.

Regarding policy options, Kempton found that there often was a lack of connection between the greenhouse effect and energy efficiency. He finds that people do not really understand what energy efficiency is, and when analysts talk about reducing energy use, lay people [in the US] tend to interpret it as decreasing energy services . US residents were found often to see energy taxes as negative, although not very strongly so. However, a policy of adaptation without prevention was not supported by the interview partners. He interprets that in the children/descendants reference frame, informants did not acknowledge discounting of future costs . Average Americans are aware that measurable environmental disturbance is now global in scale, ••• and they want decisions to be based not only on cost and benefits but also on our responsibility to leave a healthy planet for our descendants (Kempton, 1991).

3. European Perspectives: the ULYSSES Project

Instead of using individual interview methods, the European ULYSSES project has developed a specific form of *focus group* procedures (called *integrated assessment* (IA) (*see* Integrated Assessment, Definition of, Volume 4; Integrated Assessment Models of Global Change, Volume 4) focus groups) to study views of citizens and other stakeholders on global change issues as part of IA processes (Kasemir *et al.*, 2000). IA strives to provide more useful information to decision-makers than can be achieved with traditional disciplinary research (for more on IA concepts, see Weyant *et al.*, 1996). The IA focus group method allows exploration of spontaneous associations by the participants (e.g., in collage work) as well as the interaction with current research findings (usually by the use of computer models in the focus groups), before the participants themselves summarize their views on the focal topic. Within ULYSSES, IA focus groups were conducted in seven metropolitan areas across Europe in order to study views of ordinary European citizens on climate and energy issues. Approximately 600 European citizens participated in these groups.

The IA focus group method as developed by ULYSSES tries to avoid both extremes mentioned above concerning the role of expert knowledge in citizen debates on global environmental change. Scientific findings on GEC issues are seen as an important reference point, in that the groups of citizens are confronted in a series of meetings with current expert assessments of global change issues. But it is not the goal of the exercise to study to what extent the citizens in the IA focus groups assimilate these expert views. Rather, the aim is to understand what views on climate change and climate policy options emerge in the confrontation with scientific findings and with other citizens' points of view.

IA focus groups are not designed to elicit individual consumer preferences (which are better studied by observing actual consumption behavior). Rather, in line with the contention by Dunlap (quoted at the start of this section) that collective action is essential for any effective response

to global change, these focus groups are intended to show how ordinary citizens develop preferences about public choices and how these preferences evolve and are expressed in a dynamic social setting. Because the principal purpose of ULYSSES was to study views emerging in discussions between different types of participants (thought to be more indicative of public opinion dynamics than a study of the views of isolated segments of the population), it was important that the focus groups be diverse with respect to age, gender, income, educational level, and attitudes toward the environment. To that end, the potential participants were carefully screened, with quotas established for manual workers and college graduates; those who felt that environmental problems were important and those who did not; and those who favored environmental regulation and those who did not.

In agreement with findings of Dunlap's global and Kempton's US study, European citizens who participated in ULYSSES focus groups often saw the prospect of climate change as very frightening. In general, the ethical aspects of global change, such as the duty to preserve nature or responsibility toward future generations, were important to participants. Worries about the frightening aspects of climate change were not restricted to the impacts in the regions where the participants lived. Impacts in other parts of the world, especially in developing countries, mattered to the European focus group members. Somewhat in contrast to Kempton's findings that higher temperatures rather than more complex climate effects were of major concern to US citizens, the European ULYSSES study found that, while rising temperatures were often mentioned, other consequences (notably a rise in sea-level and the increased frequency of extreme weather events) were seen as more worrying. Significantly, the impacts of climate change were usually not perceived in economic terms but rather in terms of the deterioration of basic living conditions. This supports Dunlap's finding that economic impacts are not at the center of citizens' worries about climate change.

One of the aspects stressed in the ULYSSES study was the role of uncertainties surrounding climate change and its impacts. The ULYSSES researchers took care to confront the participants with the uncertainties in global change science, and not just with some *best guess* scenario. As a result, the focus group participants were often disappointed that science cannot be more definitive about these issues. However, in most cases uncertainty about future climate change was not seen as a reason for not taking preventive measures. Expectations of the deterioration of basic living conditions around the world, ecological destruction, and danger to future generations made a wait-and-see policy unacceptable to the participants.

While participants usually did not frame climate impacts in economic terms, economic arguments were readily offered in discussing response options. In line with Dunlap's finding that globally there is some willingness to pay higher prices for environmental protection, a certain increase in energy prices was often seen as a useful measure by the European citizens participating in the ULYSSES study. However, in the focus group discussions it became clear that not many participants supported hard-hitting increases in energy prices. The reasons for rejecting very-high energy prices included fairness. For example, it was generally not seen as fair if only the rich could afford to continue using their cars. However, some participants thought that slow and steady increases in energy prices would induce technological change towards greater energy efficiency, and most believed that more investment in research and development was necessary.

Overall, and in contrast to the findings of Kempton for US residents discussed above, the European participants in the focus groups often saw a society with much lower energy use than today as being very desirable. Many of the people participating in the focus groups wanted to reach this goal without excessive increases in energy prices. They tended to feel that the development of innovative, energy efficient products is crucial for making this goal feasible and also thought that researchers and firms have the responsibility for it. For example, it was felt that much more energy efficient cars could already be produced, but that industry was delaying the release of such innovations. From the point of view of many European citizens, what seems required in the climate policy debate is an ethical discussion concerning which climate impacts are unacceptable, together with a discussion of how to attain large reductions in energy use at low cost. Fostering related technological change will be essential for this.

PERCEPTIONS OF TECHNOLOGICAL OPPORTUNITIES

The research on public responses discussed above implies considerable responsibility for politics, business, and science to share in the exploration of low-cost options for global change mitigation by supporting major processes of technological change. According to traditional views of technological change, science would come up with different technological options for a well-defined problem, politics would support the technology option being selected based on independent scientific information, and business would develop, produce, and sell the corresponding technology to consumers or industry if profit results. In reality, technological change is often a very time-consuming process which is dominated by a series of coincidences rather than planning, and, finally, leads sometimes to results very different from the initial expectations (Grubler, 1998, and references therein). It is the complex network between business, politics and science, which makes the difference. Because some links between business, politics and science

dominate the system and other possibly important links are missing, information and financial resources, most often, are not accessible to some key actors in innovation processes. Technological innovation might not lead to expected results, if promising technological solutions cannot be communicated by science, politics relies on particular interest groups, and business is focused on short-term profit only. It seems reasonable that well organized and continuous information exchange and support between politics, science and business can improve the chances that even visionary long-term goals can be achieved (still success is not guaranteed, see section on complex dynamics: innovation and lifestyle changes). In addition, public acceptance is often the key element: numerous examples exist of technological options with great potential which have not been accepted by the public for various reasons. Thus, the traditional approach (especially promoted by scientists), that technology just has to be developed and transferred, is hardly a successful way to induce technological change towards sustainable development.

In this article, technological change refers to a concept of technology which is not only defined by both hardware and knowledge for production and use but includes technology diffusion, resulting from the spread of technological innovation via different communication channels in time and space based on acceptance among members of a social system (Grubler, 1998). The various actors in science, business, politics, and the public shape the different stages of the technological life cycle consisting of invention, innovation, and, finally, diffusion. Concerning technological change as a response to GEC, various processes like introduction, substitution, and cut-off processes of complete technological systems or even technology clusters as well as efficiency and end-of-pipe measures on already existing technological systems are of fundamental importance. Developing countries with strong economic growth could profit from the introduction of new technological systems (such as power plants or transport systems) to increase their productivity and close the gap to with industrialized countries without making the same mistakes in wasting resources. In the industrialized countries, substitution (e.g., innovative fuel cell cars for conventional cars) and increasing efficiency (e.g., more efficient fuel consumption by conventional cars) of technological systems are needed to maintain (and increase) the productivity but decrease the consumption of resources and the production of waste. Cutback processes are necessary to restrict production of dangerous products (e.g., chloro u-orocarbons (CFC)) whereas end-of-pipe processes reduce the environmental impact by filtering out (e.g., sulfur-based components) or transforming (e.g., by catalysors in car engines) wastes or emission components. Technological changes involving large capital-intensive facilities (such as power plants) are often crucial since the environmental impact is most significant. However, the long time scales for introduction or substitution processes have to be taken into account, huge investments in (basic) research and development, as well as public acceptance have to be guaranteed. Recently, the concept of an efficiency revolution involving decentralized and smaller as well as highly adaptable technological systems has been renewed (von Weizsacker *et al.*, 1997). It is argued that investment in technological change of this kind has several advantages like short time scales for development and implementation, high payback rates for profit-oriented business and improved living quality for people.

In the following section, we will focus our discussion on technological change involving energy systems since the world consumption of primary energy, which increased by a factor of more than four during the last 50 years, is directly related to one major element of GEC, climate change. The primary energy consumption *per capita* measured in kilowatts (kW) is an interesting indicator of the present situation. A kW is a thousand watts, and a watt is a rate of using energy. The present worldwide average of 2 kW *per capita* results from a very broad distribution including developing countries such as Ethiopia with 0.29 kW, Europe with 4.8 kW and the United States with more than 10 kW *per capita* (World Resource Institute, 1997; Imboden and Jaeger, 1999). Two simple scenarios outline potential limit cases for future development. If, on the one hand, an increased world population of 10 billion people head for the specific energy consumption of the US, a global energy consumption level ten times larger than in 1990 would result. If, on the other hand, the global average of 2 kW *per capita* was kept constant, the global energy consumption would only increase due to growth of the population, e.g., double with respect to 1990 in case of 10 billion people. Considering the strongly increasing energy demand of developing countries, it is obvious that the ambitious approach of keeping the average at 2 kW could only result from a less heterogeneous distribution of the energy consumption *per capita*. For example, if the industrialized countries could reduce their energy consumption *per capita* by a factor $2-5$ due to a huge efficiency increase, and the economically growing developing countries could restrict their energy consumption *per capita* to 2 kW by establishing efficient energy use. We refer to this scenario as the *global 2 kW society*.

An energy system satisfying the consumption of an energy efficient global 2 kW society but based on fossil resources would, however, still lead to an increase of the atmospheric carbon dioxide (CO_2) concentration exceeding double its pre-industrial value before the year 2100 which is generally considered as a level inducing significant climate changes. Thus, technological change affecting the energy production side plays a key role in reducing CO_2 emissions as well. The Intergovernmental Panel on Climate Change

(Houghton *et al.*, 1995) suggests that the overall emissions have to be decreased from the present 8 Gt C (Giga-tons carbon) per year (including 6 Gt C due to consumption of fossil energy and 2 Gt C due to deforestation) to a level of 40% (3 Gt C per year) within the next 100 years. To achieve this goal, the burning of forests has to be stopped and the contribution of fossil energy has to be reduced from currently approx. 77% (World Energy Council and International Institute for Applied System Analysis, 1995) to 25–30% in 2100 (taking increasing efficiencies into account). The rest of 70–75% must be contributed by different technologies without producing CO_2 emissions. A global 2 kW society of 10 billion people which satisfies the requirements of the IPCC could have the following contributions of energy resources in the year 2100 (Wokaun, 1999a): wind energy 5–10%, geothermal energy 5–10%, solar energy 10–15%, biomass 25%, hydropower 10%, nuclear energy 10%, fossil energy 25–30%. Massive introduction and substitution processes have to be undertaken to build a sustainable energy production system of this kind. Technological change on the energy production side involves large facilities and the above mentioned problems have to be taken into account.

The global final energy consumption can be divided into the following energy sectors (International Energy Agency and Organization for Economic Cooperation and Development, 2000) industry 35%, transportation 28%, residential/commercial 33%, and non-energy use 4%. More than 50% of the global final energy consumption (major part of the transportation and residential sectors) directly results from activities of the public and, therefore, depends on the behavior and decision-making of the consumers. In many countries, the transportation sector has been the fastest growing energy sector and has overtaken the industrial sector. In the following section, we focus on technological change in the sector referred to as the transportation of people looking at some of the key problems on the way to a 2 kW society. The transportation sector, in particular, is very interesting from the point of view of the science of complexity (see section on complex dynamics: innovation and lifestyle changes) since technological change within this sector is connected with a large variety of processes, including all actors of Figure 1 whose strategies underlie numerous environmental, economical, and social criteria.

Passenger travel demand and mode choice follow regular patterns rooted in regularities in time and money expenditures (Schafer and Victor, 2000). At highly aggregate (world-regional) levels, humans spend a roughly constant fraction of their income and time on transportation. A roughly constant fraction of travel money expenditures implies that travel demand rises approximately in proportion to income, while a roughly constant travel time budget (1–1.5 hours *per capita* and day) requires the rising travel demand to be satisfied with higher-speed modes. Walking is replaced by public transport, then private cars, and finally modes which involve high-speed trains and airplanes. Based on these regularities in time and money expenditures, land-use constraints, and projected growth in world economy and population, the volume of passenger traffic world-wide might increase from 23.4 trillion Pkm (passenger kilometer) in 1990 to about 100 trillion Pkm in 2050, and worldwide shares will change dramatically in the coming decades. High-speed traffic might gain a market share of about one third of the world passenger traffic volume by 2050 compared to 9% in 1990 whereas car traffic might supply about 40% in 2050 compared to 53% in 1990. These results suggest that changes in the transportation sector will have a tremendous in uence on GEC. We discuss some aspects of technological changes in the most important traffic modes for the transportation of people-high-speed trains and cars.

Aviation is the least environmentally sustainable of all transport modes at present and shows the least promise of becoming sustainable. Largely because commercial aircraft are usually better occupied than cars, fuel-use per passenger kilometer is, on average, only slightly higher than for cars. However, in the most recent report of the IPCC (Penner *et al.*, 1999) it is concluded that burning a given amount of fuel in an airplane appears to be equivalent to burning about three times as much fuel in a car because exhaust gases seem to have a much more potent greenhouse effect 10–11 kilometers above ground than at sea level (mainly due to the increased production of the greenhouse gas ozone, which is the transformation product of exhausted NO_x). Thus, one return Trans-Atlantic trip by airplane has an equivalent global warming impact as about two years of automobile travel of an average European or North American. The globalized economy, resulting in growing demand for business trips, and modern lifestyle, with travel-intensive leisure activities, has led to an annual rate of increase of air travel of about 6% in the 80's and 90's with a slightly reduced increase of 4–5% until 2000 (Air Transport Action Group, 2000). This dramatically increasing demand has exceeded the technological improvements needed for energy efficiency, although aircraft are about 70% more fuel efficient per passenger-km than 40 years ago. It is projected (Penner *et al.*, 1999) that future engine and airframe design improvements will result in a 20% improvement of fuel efficiency by 2015 and a 40–50% improvement by 2050 relative to aircraft produced today. However, the realization of such improvements in the form of substitution processes takes a long time since typical life expectancies of aircraft are 25 to 30 years. Alternatives to kerosene-based fuels are promising but long-term options, such as hydrogen fuel, which would reduce CO_2 and NO_x emissions from aircraft, require new approaches to aircraft design and supply infrastructure. Technological changes on a shorter time scale will affect air traffic management with a reduction potential for fuel consumption in

the range between 8–18%. Smart techniques for guidance, separation, coordination, and control of aircraft movements are promising since they have the potential to improve security and energy efficiency at the same time.

The public is a key player for the further development of aviation. The outcome is uncertain since contrary aspects ranging from the lifestyle-driven demand for high-speed travel to public opposition against expanding airport activities characterize the present situation. Complex decision-making processes in uenced by public opinion have a high impact on the realization of the potential technological improvements since, first, safety considerations have priority and, second, there is a considerable trade-off between energy efficiency and noise protection measures (quieter airplanes use more fuel). Regulatory and economic options in the form of increasing fuel prices or aviation taxes (Penner *et al.*, 1999) could successfully reduce the explosive growth of air traffic but public opposition (see section on public reactions to global environmental change) has to be taken into account. Thus, technological alternatives combined with changes in travel behavior that aim for the substitution of aircraft travel are the most interesting options. Most aircraft ights over land are still in a short-distance range and could be substituted by ecologically more efficient high-speed trains, which could eventually offer mobility at comparable speed and/or superior travel comfort. Intercontinental travel by airplane is, of course, much harder to replace since travel by sea is much more time-consuming and needs to be combined with leisure or working activities to be attractive for a certain sector of travelers.

In the railway sector, a relatively large variety of technological options exist. The conventional steel-wheel-on-steel rail technology still dominates and ongoing technological progress ranges from improved electric and diesel power trains, reduction of train weight and improved aerodynamics to new design concepts for conventional trains, like the successful Italian Pendolino tilt-technology, which enables up to 35% increased velocities on curves and increased travel comfort for passengers. The next stage comprises innovative rail concepts including modifications in track infrastructure. High-speed trains like the French TGV and the German ICE are compatible with the existing track system but profit from high-speed tracks with reduced curve radii, which enables top speeds greater than 300 km h^{-1} in regular service. Introduced about one decade ago, they are widely accepted by consumers in many countries. However, their potential for further progress in speed and, thus, competition on a time basis with ights of up to 1000 km is limited. Thus, several concepts for new high-speed railway systems based on magnetic levitation or jet-powered hover trains, like the Japanese MLU and the German Transrapid , with velocities from 350 km h^{-1} up to more than 500 km h^{-1}, are being developed. Another new technology, the Eurometro , would operate below ground in partially evacuated tunnels running on magnetic propulsion, levitating, and guiding systems. It is argued (National Research Program NRP 41, 2000) that such a system would be much more resource- and energy-efficient than other transport modes (about 5–10 times compared to short-distance aircraft, 6 times compared to cars, and about 3 times compared to conventional trains). However, such large-scale technological change would have to rely on high and long-term investment, broad political and public acceptance as well as international cooperation. Critics of such high-speed systems (von Weizsacker *et al.*, 1997) argue that these systems are not compatible with existing infrastructures and the needed high investments, which should instead be used to improve either conventional technology combined with much increased travel comfort or, if new infrastructures are desired, to build more exible, consumer-oriented train systems that could increasingly compete with car traffic. The Cyber-Tran system (National Center for Advanced Transportation Technology, University of Idaho), for example, would be based on large numbers of small, light, and therefore energy-efficient vehicles operating on elevated guide-ways at speeds of up to 240 km h^{-1}. Flexible operation would be based on consumer call options and a high number of small vehicles.

Apart from technological considerations concerning travel speed, railway systems have a huge potential to improve their travel comfort and, thus, compete increasingly with car or air traffic. Simplified booking and ticketing procedures, improved home computer and internet access, and, most importantly, novel on-board-entertainment concepts for families or groups would improve travel comfort significantly. In addition, railway traffic could be much better synchronized with other traffic modes, in particular, public transport and cars. The implementation of car sharing, which operates from railway stations, can be very attractive for consumers in terms of travel comfort as well as costs. In the end, public acceptance of novel train travel forms is as important as technological criteria for selecting appropriate train systems and for replacing air and car traffic by train traffic.

At present, car traffic is the backbone of transportation of people in many countries. The worldwide success story of the car is strongly connected to the fact that it enables a high degree of individual mobility which is an important feature of modern lifestyles. However, the huge increases in car traffic during recent decades has led to a situation where people in most cities all over the world suffer from car emissions, noise and daily traffic gridlocks. Furthermore, the contribution of the car segment to global warming is very significant. Considerable effort is being undertaken by regional, national, and international actors to reduce environmental impacts. The strategy consists, most often, of

measures in three categories. Modern city planning applies concepts seeking to avoid car traffic or enabling a shift towards more efficient public transport. For car traffic that cannot be avoided or shifted, smoothing measures have a large potential. Smoothing measures comprise improvements in car operation and technology. With regard to car operation;

1. occupancy of cars is low, in Switzerland, for example, 1.1 persons per car for work trips, and about 2 persons per car for leisure trips. It is a considerable challenge to integrate efficient forms of car sharing in consumer-accepted travel habits;
2. fuel consumption of cars depends heavily on their technical operation. An efficient driving style, which consists of defensive driving in high gears, results in a reduction of fuel consumption of about 10–25%.

A major shift in the behavior of consumers can only result if concerned stakeholders including manufacturers develop new lifestyle priorities in car driving based on security and efficiency. In some European countries, for example, EcoDrive is an established ecolabel which is supported by authorities in the form of lessons on eco-efficient driving and even efficiency competitions (*see* Environmental (Eco) Labeling, Volume 4).

Long-term improvements in car technology will have a decisive impact to the contribution of car traffic on global warming. Besides the improvement of air and rolling resistance of a car, two aspects of technological development predominate. A reduction of car weight is a necessary prerequisite for any drastic reduction of fuel consumption. At present, we are confronted with the general tendency towards even heavier and more powerful cars, which results from increasing demands for security and comfort as well as lifestyle-driven desires. Several approaches for weight reduction exist which relate to very different design concepts. The automobile in its present form is a vehicle optimized for a broad range of travel patterns, in particular for medium to long-distance travel. A comprehensive lightweight-strategy for the large variety of conventional cars is not realized yet, although some manufacturers refer to first successes in producing cars with an increasing amount of lightweight components and significantly decreased fuel consumption. Manufacturers, in general, argue that production costs increase significantly because of still expensive lightweight components. They leave it to future research and development to make possible the use of alternatives to steel without serious drawbacks in costs, security or comfort. A lightweight-strategy can also be developed in a very different way in the form of a car that would be optimized for city or low-distance use at lower velocities and, thus, could be built with radically reduced weight in the range between 200 and 500 kg. The use of such vehicles could be combined with public transport or railway systems and could therefore substitute most of the city or short-distance car traffic based on heavy vehicles. However, a major shift of consumers' habits would be part of this more radical lightweight-strategy. The second starting point for improved fuel efficiency is the innovation of power train systems. There exist a couple of options apart from conventional combustion engines, including battery-driven electric cars, hybrid cars combining elements of electric and combustion engines as well as fuel cell cars (Adler, 1996). The car industry seems to follow a two-sided strategy: on the one hand, the conventional combustion engine is constantly improved and, on the other hand, considerable effort is undertaken to develop fuel cell cars which are driven by hydrogen fuel. Fuel cell cars are of considerable interest since there exist several short- to long-term options to produce hydrogen fuel in a sustainable way ranging from on-board reformation of methanol gained from biomass gasification to solar production of hydrogen (Wokaun, 1999b). A fuel cell car package based on solar hydrogen could be one of the major challenges in the transportation sector in the coming decades. Large-scale technological change, involving the introduction of a new fuel supply infrastructure as well as changes towards public acceptance of new car concepts, are necessary prerequisites for this transition process.

The technological options for the worldwide fast-growing transportation sector would make an important contribution to possible solutions of pressing environmental problems, but technology is only one element in the process of change. Several other aspects have to be considered. First, transformations in all sectors have to be compatible with each other. Second, technological change has to be accepted by people living in very different societies with different histories, various norm and value systems, divergent ethical-philosophical backgrounds, with different priorities for basic needs and possibilities to cope with them. It is important that political and public acceptance is considered in the early stage of technological research (in the form of stakeholder-dialogues including e.g., IA focus groups) to support decision-making in the context of, for example, the selection of appropriate car concepts or train systems. In general, it has to be taken into account that some technological options, if not supported by lifestyle-changes, might drive the system as a whole into an even less sustainable direction. In spite of the widely held conviction that large-scale technological change, including behavior changes, are impossible and that developing-country imitation of the unsustainable lifestyle of industrialized countries is unavoidable, the evolving science of complex systems adds reasons to believe that a large-scale transition process is more than a mere esoteric dream.

COMPLEX DYNAMICS: INNOVATION AND LIFESTYLE CHANGES

During the last three decades, a major paradigm change induced by novel concepts arising in the natural sciences and associated with the notions of *chaos* and *self-organization* has lead to the development of a new science of complexity (Mainzer, 1997; Strogatz, 1998). This approach tries to circumvent important limitations encountered by traditional reductionistic methods when dealing with complex systems such as ecosystems, the global climate system, living organisms, the immune system, socio-economic systems or the human brain (Gallagher and Appenzeller, 1999). An important contribution of the developing new science is the insight that some central conceptions and beliefs of traditional natural scientists are wrong. So is the seemingly logical idea that a deterministic system can be predicted for an extended period of time. When computers began to transform science in the 1960s, the idea of P S Laplace (1749–1827), often referred to as the *demon of Laplace*, let scientists believe that e.g., weather forecasts could be extended over any desired period, if only enough computing power and observations would be available. Though computing power as well as memory size has grown a billion-fold in the meantime, weather forecast periods could be extended by a modest factor of about 10 only. Another old misconception probably dates back to Aristoteles (384–322 BC) who believed in the continuity of natural development. The idea was assimilated by the classical Romans and put into the well-known expression *natura non facit saltum* (nature does not make leaps) which found its way into the heads of today's scientists. New observations indicate that, even the largest system on our planet, the global climate system, can be subject to abrupt transitions (Greenland Ice-core Project (GRIP) Members, 1993) which could hardly contrast more with the idea of a gradual development.

These two basic characteristics of complex systems, limited predictability and prevalence of abrupt transitions, share common roots: *non-linearity* and *circular causality*. Both are ubiquitous in natural systems, where the same causes can produce different effects depending on the state of the system (non-linearity) and long chains of interlinked cause-effect relationships can lead to closed feedback loops (circular causality). An example for non-linearity is the relation between carbon dioxide concentration in the atmosphere and the resulting warming force, pushing the global climate towards another state. The already raised atmospheric CO_2 concentration increase of 32% above pre-industrial level exerts a larger climate changing force than an equivalent future increase will do. Because of the existence of many feedbacks in the climate system (circular causality), however, this saturation effect does not necessarily mean that the measured global mean temperature increase of 0.4–0.8 °C since the end of the 19th century will be an upper bound for the effect of a further anthropogenic CO_2 concentration increase of another 32%. Feedbacks including e.g., changes of the Earth's vegetation cover could induce abrupt climate change involving a large temperature increase not directly related to CO_2 growth (Houghton *et al.*, 1995). Such abrupt climate change could be induced, as soon as a yet unknown critical CO_2 concentration will be exceeded.

The most important outcome of complexity science, is probably the identification of common traits of spatio-temporal evolution for very different systems: abrupt climate change and an epilepsy episode are, from the complex systems point of view, related phenomena and their basic dynamics can be understood with rather simple mathematical models. Applied to the social phenomenon of lifestyle changes, complexity science leads us to expect the following characteristics:

The Seemingly Impossible Happens Abruptly

Looking at a given socio-economic system with its intricate interactions and interdependencies, its couplings to the existing infrastructure, to the complicated network of trade relations within itself and with other but perhaps different socio-economic systems, its correlation with the value system, beliefs and behavior of people, we are astonished by the near perfect match of the myriad of components which support each other. The classical explanation for this outcome is that the system was intentionally designed and created by the people involved. On the grounds of this explanation, it seems plausible that an intelligent dictator could make an even better design. However, why did the great historical experiments with planned economies fail? Complexity science gives an answer by suggesting a more dynamic and process-oriented view, in which small random events help a system find one of its *multiple* equilibrium states. These stable states are emergent properties of the system naturally evolving by self-organization processes. External planning efforts would act on the system as additional implied forces driving it along another path into a different state not necessarily being characterized by the efficiency and stability typical for the natural equilibria. Such a forced state is unstable in the sense that the system relaxes to one of its natural states as soon as the force is removed. Free market economies, subject to moderate planning forces only, have the possibility to evolve through their internal dynamics into one of their natural equilibria, explaining their superior stability and efficiency (Waldrop, 1992; Arthur, 1999). Because every equilibrium state shows the above mentioned astonishing match of all subsystems, another equilibrium cannot be understood on the basis of the present one and therefore seems to us impossible: today's lifestyle may not be a good reference for anticipating tomorrow's lifestyle (and e.g., its matching transport system).

From this point of view, the prominent trait of complex systems to jump from one to another equilibrium rather than changing gradually and slowly is plausible. As the components cannot be changed one by one (they would not match any more the whole system), a lot of them have to be transformed simultaneously. Due to strong couplings and feedbacks, this transition phase is short and therefore changes are *abrupt*. However, if some components almost instantaneously react to the ever-changing situation during a state transition, a new match cannot easily be established: an irregular *chaotic* phase is the system's answer. As chaos always looms in the backyard of transitions, the intuitive hesitation of politicians against changes is well founded.

Important Changes are Unpredictable, and Irreversible Surprises

Everyday life experience tells us that complex systems can slowly be brought to a critical point, where their reaction is very intense. Due to our incomplete knowledge of the internal dynamics of complex systems, we do not know of the exact location of such critical boundaries and sometimes not even that they exist. A system forced slowly in a certain direction might exceed such an invisible and so far unknown threshold and its sudden movement towards another equilibrium is a *surprise*. An abrupt global climate change at a so far unknown critical CO_2 concentration is in the same way unpredictable as a sudden change in lifestyle at critical levels of civilization, environmental damage, education, wealth, etc. In addition, an abrupt change modifies the system in such a way that a direct return to the original state is not possible. A hypothetical jump of the global climate to a warmer state beyond a critical CO_2 concentration would be *irreversible* in the sense that the system might return to the original cooler state only after a considerable reduction of CO_2 concentrations, perhaps even to below its current value (Gassmann, 1998). Because such a reduction would be impossible or at least take an excessive length of time, an abrupt climate change would have to be considered as a practically irreversible development.

Be At The Right Place In Exactly The Right Moment

Recent stability investigations reveal that e.g., complex hydrodynamic systems are almost always rather insensitive to external disturbances, even if they are in a chaotic state. So, the legendary butter y has to disturb the weather system at the right place at exactly the right moment to be effective. Computer simulations show transitory and localized sensitive hot spots, reminiscent of windows of opportunity in business or of the phenomenon called *kairos* in psychology. This Greek term for weather or time is used there to describe the narrow time window when a psychiatric treatment would be effective. In addition, this insight into the dynamics of complex systems could explain the significance of historical moments and gives us hope that the development of human history could be in uenced. At the same time, it indicates to us that we should adopt a rather modest attitude, realizing that important changes cannot be programmed to occur at the moment of our choice, but can only be induced when the system is ripe to react. This view questions versions of neo-classical economic theories that were inspired by the successful classical physical theories (applicable for rather simple and artificial systems only) and led, e.g., to the development of general equilibrium models with unique equilibrium states and poor dynamics. The science of complexity offers radically different and hopefully more realistic paradigms exhibiting *multistability* and *path-dependency*. But contrary to neo-classical theory, this dynamic rather than static equilibrium is of temporary nature: The possibility of abrupt changes to another state induced by external or internal uctuations is a permanent reality. This everlasting potential for change can be seen as a threat, but also as giving hope that our society can move towards a more sustainable future, with a better match between mankind and nature.

These concepts can be used to study very different complex systems. We concentrate on the example of a system of actors running a car eet and being affected by innovative car developments. The present car eet industry based on conventional fuels, which is run by a network of actors (customers, manufacturers, and the fuel supply industry) acting under economical, environmental, and political constraints, constitutes a complex system with many interlinked circular cause-effect relationships. The established norms and values, especially those developed in the industrialized nations, have led to a context supporting and stabilizing the traditional car types. In a first transition stage, an innovative car, e.g., powered by fuel cells, with low noise and CO_2 emissions would hardly match existing infrastructure, actual taste, and economical expectations of its potential customers. Every innovation process would be doomed to fail, if it does not take into account this initial situation.

In a social system, every individual contribution is integrated into a macroscopic pattern called *lifestyle* (*see* Life Style, Private Choice, and Environmental Governance, Volume 5), giving a context that guarantees stability of the system by circular causality (Kuppers, 1998). This relationship is analogous to the situation of an ecosystem modifying an initially hostile environment into one that tolerates a first variant of a new ecosystem. This pioneer-ecosystem pushes the environment further into a more favorable state, supporting a next step in its development. By this self-sustaining circular relationship, isolated ecosystems can survive and even expand in a general environment that, in the unmodified state, would not permit survival.

In the same sense, an innovative setup within the car industry dominated by conventional cars has to create its own environment supporting the ongoing innovation process. It can be catalyzed best, if some major concerns are part of an *innovation network* enabling strong cooperation. Not only technical aspects are optimized due to these interactions and in the course of the innovation process, but the system of norms and values changes which leads to a modified context with new opportunities for new car types. Reciprocally, the car type has to be modified until it finds a match within emerging new norms and values. As information on adverse aspects of conventional cars and prerequisites for public acceptance of the new car type have to be evaluated at the same time with the technical optimization, strong links between science, manufacturers and the public are absolutely necessary. Technical feasibility, economical success and social acceptability have to reproduce each other in a stable way, inducing an abrupt change in a carefully chosen innovation-friendly environment. For further information on the dynamics of complex systems, *see* Chaos and Predictability, Volume 1; Chaos and Cycles, Volume 2; Non-equilibrium Ecology, Volume 2; Monitoring in Support of Policy: an Adaptive Ecosystem Approach, Volume 4.

A new interaction could result if, for example, independent research institutes were to develop, with the help of IA focus groups and surveys, revolutionary car concepts that found broad acceptance among consumers. If such novel concepts, which might not be found in more interest-driven market research of car manufacturers, could be communicated to business, the at first isolated innovation region would be stabilized and in a late stage, a revolutionary car type could even be made. The links established during the innovation process would help to integrate characteristics of the operation of the new car type into the lifestyle of consumers: disadvantages of innovative cars according to today's norms and values would be turned into advantages by modifications of these norms and values. This cooperation of all involved actors would finally arrive at a point where traditional cars would be considered as old fashioned and would therefore be rejected. Today's ideals of car types with impressive power and size would lose their dominance and would give way to new values better matching the conditions for sustainable development.

CONCLUDING REMARKS

Conclusions on Citizens' Responses to GEC

What does the research on public responses to GEC discussed above imply for the interplay of the public with politics, business and science? First, the Health of the Planet survey indicates that concern about the environment in general and global environmental problems in particular seems to be on roughly the same level in developing and in developed countries. The finding that citizens of both sets of countries are not pointing the finger of blame at each other should encourage policy makers around the globe to look for a consensus on dealing with global environmental change. Often, not only government, but also individual citizens and citizen groups were seen as responsible and effective players in environmental protection. This means that citizens do not simply delegate responsibility to policy makers but are willing to assume responsibility themselves.

All three studies discussed indicate that citizens in different parts of the globe see climate change as a clear threat. Some aspects of this concern were found to be different in the three studies. While the study of Kempton *et al.*, stressed the importance placed by US residents on preserving the local environment, the ULYSSES study found that many European citizens are concerned with climate change impacts around the globe. While Kempton's results indicate, e.g., an anthropocentric justification of species preservation or its secondary importance compared to rising food costs for US residents, both the findings of the Health of the Planet survey and the ULYSSES project imply that impacts on social systems, phrased in economic terms, were not the main concern of the citizens participating in these two studies. However, Kempton's US study and the European ULYSSES project agree in the conclusion that ethical values, including a desire to preserve the environment for further generations, lead to clear citizen support for a preventive climate policy.

Concerning the role of climate change science, the studies came to different conclusions. It seems clear that there is a mismatch between expert and lay conceptualizations of climate change (see Bostrom *et al.*, 1994). However, recommendations range from increasing public education about climate science (favored by Kempton *et al.*) to considering the support by citizens for environmental protection based on a holistic fabric of the perceived ecological problematique' (implied by Dunlap). A mediating suggestion is to confront study respondents with scientific findings, but then to focus on what views on climate change and climate policy options emerge in this confrontation (as done in ULYSSES IA focus groups), rather than on how much of the expert assessments were assimilated by the participants.

Of major importance for the topic of this essay is a set of findings regarding energy prices and technology. Dunlap's survey results imply that globally there is some willingness to pay more for environmental protection. The more in-depth discussions with European citizens in the ULYSSES study support this willingness to pay somewhat more for the case of energy, but also show clearly that hard-hitting increases in energy prices could not be supported. This is in line with the finding by Kempton that energy taxes were often seen as negative by his US respondents. The ULYSSES study found clear positive associations with a

low-energy future by European citizens, who seem much more supportive of reducing overall energy use than their US counterparts participating in Kempton's study. But in both cases, while climate protection was clearly supported, high energy prices seem not to be perceived as the way to get there. This implies an essential role for technological change fostered by science and business. While technology was seen as one cause of environmental problems, Dunlap's findings also indicate as well citizens' hope that the right technological change can help to find a solution. The ULYSSES study found that (at least for the case of European citizens) the development of innovative energy-efficient products like low-energy cars is seen as crucial, and science and industry are perceived to have responsibility here. It will be their interaction with citizens and policy makers in complex social systems that will be crucial for shaping a combination of technological change and lifestyle changes as humankind's response to global environmental change.

Conclusions on Social Learning and Future Research

While technology plays a major role for a development towards a sustainable society, the technological imperative *can* implies *ought* , as widely accepted in the golden age of engineering, has to be replaced by more comprehensive approaches. Complex processes of *social learning* which result from a new quality of interaction between the public and science, business, and politics, will shape the chances for transition processes including combinations of technological change and lifestyle changes and, thus, the feasibility of effective GEC mitigation options in the future. *See* Social Learning in the Management of Global Atmospheric Risks, Volume 5.

Social learning (see, e.g., Social-learning-group, 2000) starts with the awareness of all of the actors in a complex system. Dialogues between all actors will unavoidably create new questions not confined to single areas and, therefore, open up new horizons for innovative solutions of pressing problems. According to the science of complexity, modified interaction patterns between different actors in such a network bear the potential to invoke a self-organization process leading to abrupt transitions or trend changes, which happen as a surprise to single actors in contrast to expectations of linear evolution.

In the following, we focus specifically on the interaction between the public and science, which represents, to our understanding, an especially weak link in the network of Figure 1. In principle, public and science interact in both directions. The more established interaction consists of scientists informing the public. Specifically in the last years, many research institutes all over the world have defined new programs to increase information to citizens and tax payers about their research activities. This important interaction, however, cannot compensate for the fact that there is as yet almost no direct reciprocal impact of the public on science.

The public does not just stand aside in innovation processes, profiting or suffering passively from the results of scientific discoveries and respective technical innovations. It is heavily involved in the process due to decision-making of consumers and voters. Public participation in early stages of research projects with relevance to GEC could not only make the techno-scientific enterprise more transparent but also more successful, as public acceptance would be considered from the very beginning. A dialogue between scientists and people, e.g., in form of IA focus groups, could help experts avoid linearly extrapolating their technical knowledge and neglecting important consequences of innovations. On the other hand, if lay people's creative fantasy exaggerates technological possibilities, integration of people in such a process can help to increase public acceptance for scientists' problems like, for example, the lack of simple solutions to complex problems.

New forms of interactions, e.g., between the general public and scientists, call for new tools to support communication. In this context, computer simulation and modeling could be very useful. We do not think that a future sustainable world can be designed on the basis of mathematical models; this would be reminiscent of the efforts to establish planned economies during the last century. However, we suggest that the communicative potential of models of complex systems should be exploited. Contrary to traditional scientific beliefs, adequate models will hardly be found as a by-product of disciplinary scientific research, but will require a highly interdisciplinary scientific enterprise of its own. And using such models effectively will require carefully prepared processes of model-stakeholder interactions. As a necessary prerequisite for such an endeavor, scientists have to adapt a more holistic view allowing them to see their own research field as a valuable contribution, rather than the only or at least the main part of the scientific adventure. We consider it as an exciting challenge that an adequate research agenda for the decades to come, when questions in the grey area between natural and social sciences arise, can only be addressed in a collaborative effort involving science, business, politics, *and* the public.

REFERENCES

Adler, R (1996) *Automotive Handbook*, Robert Bosch Corp., Stuttgart, Germany.

Air Transport Action Group (ATAG) (2000) *The Economic Benets of Air Transport*.

Arthur, W B (1999) Complexity and the Economy, *Science*, 284, 107–109.

Bostrom, A, Morgan, M G, Fischoff, B, and Read, D (1994) What do People Know about Global Climate Change? Part 1: Mental Models, *Risk Anal.*, 14(6), 959–970.

Brand, K W (1997) Environmental Consciousness and Behaviour: the Greening of Lifestyles, in *The International Handbook of Environmental Sociology*, eds M Redclift and G Woodgate, Edward Elgar, Cheltenham, 204–217.

Dunlap, R E (1998) Lay Perceptions of Global Risk. Public Views of Global Warming in Cross-National Context, *Int. Sociol.*, 13, 473–498.

Dunlap, R E, Gallup, G H, and Gallup, A M (1993a) *Health of the Planet: Results of a 1992 International Environmental Opinion Survey of Citizens in 24 Nations*, The George H. Gallup International Institute, Princeton, NJ.

Dunlap, R E, Gallup, G H, and Gallup, A M (1993b) Of Global Concern, *Environment*, November 1993, 6–15 and 33–39.

Gallagher, R and Appenzeller, T (1999) Beyond Reductionism: Introduction to 10 Following Articles on Complexity in Different Scientific Disciplines, *Science*, 284, 79–109.

Gassmann, F (1998) Greenhouse Effect – Modeling and Reality, in *The Co-Action Between Living Systems and the Planet*, eds H Greppin, R Degli Agosti, and C Penel, University of Geneva, Geneva, 43–66.

Greenland Ice-core Project (GRIP) Members (1993) Climate Instability during the Last Interglacial Period Recorded in the GRIP Ice Core, *Nature*, 364, 203–207.

Grubler, A (1998) *Technology and Global Change*, Cambridge University Press, Cambridge.

Houghton, J T, Meira Filho, L G, Callander, B A, Harris, N, Kattenberg, A, and Maskell, K (1995) *The Science of Global Change: Contribution of Working Group I to the Second Assessment of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge.

Imboden, D and Jaeger, C C (1999) Towards a Sustainable Energy Future, in *Energy: The Next Fifty Years*, ed OECD, OECD, Paris, 63–94.

International Energy Agency (IEA) and Organization for Economic Cooperation and Development (OECD) (2000) *World Energy Outlook*, OECD, Paris.

Kasemir, B, Schibli, D, Stoll, S, and Jaeger, C C (2000) Involving the Public in Climate and Energy Decisions, *Environment*, April 2000, 32–42.

Kempton, W (1991) Lay Perspectives on Global Climate Change, *Global Environ. Change*, 1, 183–208.

Kempton, W (1997) How the Public Views Climate Change, *Environment*, November 1997, 12–21.

Kuppers, G (1998) The Selforganization of Social Systems: a Simulation of the Social Construction of Knowledge, in *Computer Simulations in Science and Technology Studies*, eds P Ahrweiler and N Gilbert, Springer Verlag, Berlin, 145–155.

Mainzer, K (1997) *Thinking in Complexity - The Complex Dynamics of Matter, Mind and Mankind*, Springer, Heidelberg, Peking (Chinese translation) and Tokyo (Japanese translation).

National Research Program NRP 41 (2000) *Ecological Effects of Eurometro*, Bern, Switzerland.

Penner, J E, Lister, D H, Griggs, D J, Dokken, D J, and McFarland, M (1999) *Aviation and the Global Atmosphere: a Special Report of the IPCC Working Groups I and III in Collaboration with the Scienti c Assessment Panel to the Montreal Protocol on Substances that Deplete the Ozone Layer*, Cambridge University Press, Cambridge.

Schafer, A and Victor, D (2000) The Future Mobility of the World Population, *Transpor. Res. A*, 34(3), 171–205.

Social-learning-group (2000) *Learning to Manage Global Environmental Risks: a Comparative History of Social Responses to Climate Change, Ozone Depletion and Acid Rain*, MIT Press, Cambridge.

Strogatz, S H (1998) *Nonlinear Dynamics and Chaos*, Addison-Wesley, Reading, MA.

Thompson, M and Rayner, S (1998) Cultural Discourses, in *Human Choice and Climate Change. The Societal Framework*, Vol. 1, eds S Rayner and E L Malone, Battelle Press, Columbus, 265–343.

von Weizsacker, E, Lovins, A B, and Lovins, H L (1997) *Factor Four: Doubling Wealth - Halving Resource Use: The New Report to the Club of Rome*, Earthscan, London.

Waldrop, M M (1992) *Complexity: The Emerging Science at the Edge of Order and Chaos*, Simon and Schuster, NY.

Weyant, J, Davidson, O, Dowlatabadi, H, Edmonds, J, Grubb, M, Parson, E A, Richels, R, Rotmans, J, Shukla, P R, Tol, R S J, Cline, W, and Frankhauser, S (1996) Integrated Assessment of Climate Change: an Overview and Comparison of Approaches and Results, in *Economic and Social Dimensions of Climate Change: Contribution of Working Group III to the Second Assessment Report of the Intergovernmental Panel on Climate Change*, eds J Bruce, H Lee, and E Haites, Cambridge University Press, Cambridge, 367–396.

Wokaun, A (1999a) *Erneuerbare Energien*, Teubner, Stuttgart.

Wokaun, A (1999b) Towards Sustainable Mobility, in *Greenhouse Gas Control Technologies: Proceedings of the 4th International Conference on Greenhouse Gas Control Technologies in Interlaken, Switzerland*, eds B Eliasson, P W F Riemer, and A Wokaun, Pergamon Press, 899–904.

World Energy Council and International Institute for Applied System Analysis (1995) *Global Energy Perspectives to 2050 and Beyond*, London.

World Resource Institute (1997) *World Resources 1996- 97: a Guide to the Global Environment*, Oxford University Press, Oxford.

Scientific Responses in an Era of Global Environmental Change

PHILIPPE BOURDEAU

Université Libre de Bruxelles, Brussels, Belgium

" Panta rhei, oùden menei" (All is in ux, nothing is stationary), said Heracleitus 2600 years ago. What is new about global environmental change (GEC) is the increased rate and scale of change imposed on the environment by human activities. Science is playing a major role in detecting evidence of this new trend and in predicting its frequently negative consequences, as well as in devising responses to it. Its function remains essential but will also have to evolve. Early on scientists foresaw several of the large-scale environmental issues that subsequently developed, such as acidi cation and the impact on climate of the release of carbon dioxide (CO_2) in the atmosphere from coal combustion, but these warnings went initially unheeded. In the late 1960s the international scienti c community as a whole, represented, inter alia, by the International Council for Science (ICSU) and its Scienti c Committee on Problems of the Environment (SCOPE), or ICSU - SCOPE, became deeply concerned with the new environmental threats. SCOPE identi ed and assessed the main environmental problems of the time and made recommendations, which provided an important input to the 1972 UN (United Nations) Conference on the Human Environment. Thereafter, it kept a continuous watch to detect emerging issues, such as the depletion of the ozone layer or the loss in biodiversity. At about the same time, other international scienti c bodies planned and coordinated large research programs, such as GARP (Global Atmospheric Research Program), WCRP (World Climate Research Program), MAB (Man and Biosphere), IGBP (International Geosphere Biosphere Program), etc., in order to acquire additional knowledge necessary to respond effectively to global environmental threats. These programs provided a framework for de ning and funding research priorities at the national, supranational (e.g., European Union (EU)) and international level. This was made possible through cooperation between the science community and bodies such as WMO (World Meteorological Organization), UNEP (United Nations Environment Programme), WHO (World Health Organization), UNESCO (United Nations Educational, Scienti c and Cultural Organization), OECD (Organization for Economic Cooperation and Development), IUCN, etc. Assessments of research ndings, for example those of IPCC (Intergovernmental Panel on Climate Change), paved the way for the signing of various international conventions (e.g., Stratospheric Ozone Layer (1985), Climate Change and the Protection of Biological Diversity (1992)). It was soon realized that human factors were all important if successful solutions were to be found and implemented, hence the rise of research programs like the IHDP (International Human Dimension Program). Many environmental issues are complex and fraught with uncertainty, if not indeterminacy. This of course is compounded when the human element is factored in, hence the recognition of the precautionary principle and the need for humility on the part of scientists. This is now increasingly recognized and efforts are made to develop integrated environmental assessments taking into account all aspects of an issue and involving all stakeholders. Science also played a most important role in providing the basis for the development of technologies aimed at monitoring the environment and at curing environmental damage, or better yet, at preventing it through better industrial processes and cleaner products. The private sector is increasingly involved in funding R&D in this area. The paradigm of sustainable development, widely adopted since the 1992 UN Conference on Environment and Development, implies the integration of environmental concerns in all sectors of the economy and should guide future research priorities. Some of the main emerging issues for the 21st century are likely to be old issues, which take on new aspects as the result of the rapidly changing socio-economic scene due to globalization of the economy, the development of information technologies leading to the information society, and the growing impact of biotechnologies, as well as new social expectations in general. Surprises

cannot be excluded, however, despite progress in monitoring, modeling, and forecasting. Science must, of course, continue to improve the understanding of the basic life-support systems of the planet and there is still much to do in this area but it must also be very much concerned with policy-related research. This means supporting the goal of environmentally sustainable development through the better use of natural resources (eco-ef ciency) and modi cation of human behavior. " Science should be on tap, not on top," said Churchill. Policy-making must always be based on the best available science but many other, more subjective, factors have to be taken into account. The precautionary principle and scienti c information should be applied complementarily, not in opposition, to each other, as some would now like to see it. Finally, scientists should respect ethical standards with regard to scienti c integrity, quality assurance, recognition of uncertainties and of biases, as well as communication with decision makers and the public at large.

INTRODUCTION

The environment has never been static. All is in ux said a Greek philosopher 26 centuries ago. Astronomical and geological factors have produced considerable variations in climate and sea level over past millennia. Ecological succession has transformed terrestrial and aquatic biomes. However, man-made change has increased at an ever-growing rate and reached global proportions in the last 100 years, altering the composition of the atmosphere, profoundly disturbing biogeochemical cycles and drastically reducing biological diversity. Some of these changes have occurred as surprises while scientists predicted others, sometimes long before their actual appearance.

In the following sections the historical development of the GEC issue is discussed from the perspective of scientists. It is not intended to provide a comprehensive overview of scientific activities pertaining to the issue, but to discuss brie y the role of scientists in identifying and clarifying problems and their contributions in managing them, and to see what the future may have in store. The examples given are mainly at the international level but it should be realized that most major national research projects and programs on GEC have been designed and implemented in an international framework or context.

THE ROLE OF SCIENCE IN ENVIRONMENTAL POLICY

On defining environmental policy, as aimed mainly at managing man-made environmental risks (but also man-enhanced natural risks), a rational approach is first to gain adequate knowledge of the causes and effects of the phenomenon under consideration in order to assess the importance of the risk (*risk* is defined as the probability of occurrence of an undesirable effect or hazard). Thus, scientific information is necessary, if not sufficient, for successful control of the risk. There should also be the technical means to counter it, the political will to do it, and the necessary economic resources.

As recognized by Holdgate and White (1977), environmental issues, to which science can make profound contributions, may be grouped in three categories:

- those solved successfully by the application of existing scientific knowledge;
- those which existing knowledge could solve but where social or economic factors limit action;
- those requiring more scientific knowledge.

Scientists have been involved with man-made environmental change in various ways:

- identifying issues, elucidating their causes and effects and determining effective solutions and counter-measures (e.g., air pollution by sulfur dioxide (SO_2) and smoke);
- pursuing basic research on the structure and functioning of the earth systems (climate, biogeochemical cycles, energy ows, ecosystems, etc.);
- interpreting monitoring data, collected for other purposes, such as the study of fundamental processes, thereby detecting new issues (e.g., monitoring of ultraviolet (UV) radiation as a measure of stratospheric ozone above the Antarctic);
- performing comprehensive assessments of scientific information pertaining to an environmental issue, in order to present the information in a policy-relevant way (e.g., SCOPE at the non-governmental and the IPCC at the inter-governmental levels) and to define gaps in knowledge and research needs;
- carrying out state-of-the-environment reports at various geographical scales to assist decision making and inform the public.

While, initially, natural sciences were almost exclusively involved, the socio-economic disciplines have increasingly been called upon to resolve problems as it became evident that:

- perception of risk plays an important role;
- solutions often imply changes in human behavior;

- some environmental issues are characterized by complexity and uncertainty and bring into play the precautionary principle.

EARLY SCIENTIFIC CONTRIBUTIONS TO THE UNDERSTANDING OR PREDICTION OF MAN-MADE ENVIRONMENTAL PROBLEMS

Arrhenius in 1896 and Chamberlain in 1899 reasoned that the luxuriance of vegetation in the Carboniferous period was associated with the effect of high carbon dioxide (CO_2) concentration in the atmosphere. Arrhenius also foresaw the problem of acid rain.

In 1923, W Vernadski, a member of the Academy of Sciences of Russia, published a book in French (*La Géochimie*) presenting to a Western audience the results of Russian research at the end of the 19th and early 20th century (Vernadski, 1924). He outlined the carbon cycle, giving estimates of the amount sequestered in biomass and of the quantities released in burning coal and their possible impact on climate. He concluded, rightly, that humans had become a new geological force, moving chemical elements, such as metals, in amounts of the same order of magnitude as natural phenomena. He also mentioned the phytoplankton blooms in the Adriatic Sea, which form gelatinous masses and break down the silica balance in seawater. Vernadski introduced the *noosphere* concept as an addition to the geosphere and the biosphere, to designate the collective human thought, which impacts the environment (*see* Noosphere, Volume 5) (could we consider the Internet as a concrete manifestation of the noosphere?).

Continuous measurements of atmospheric CO_2 concentration were initiated in 1957 on Mauna Loa in Hawaii and have provided an invaluable baseline for monitoring changes in the carbon cycle. Similar measures were made in Scandinavia from 1955–1960.

The chemical constituents of precipitation were measured through a special observation network organized by the International Meteorological Institute of Stockholm University as early as 1954, and systematic observations were undertaken in the USSR in 1957, while in the USA measurements were made in 1955 and 1956, then resumed in the 1970s. All showed the rise of sulphates, nitrates and acidity.

Photochemical oxidants were identified as causing smog in Los Angeles in the 1950s. They were found to be secondary pollutants formed from nitrogen oxides and unburned hydrocarbons. This led to the adoption of catalyzers in the exhaust systems of automobiles and the banning of lead additives in gasoline since they poisoned the catalyzers. Elsewhere and later on, lead additives were phased out when epidemiological research showed the negative correlation between children's mental development and blood lead levels at even very low values.

The threat to stratospheric ozone caused by chloro uorocarbons (CFCs) emission was suspected by Rowland and Molina as early as 1972–1973. Stratospheric ozone had been measured with the Dobson spectrophotometer since the 1930s. Observations were started in the Antarctic during the International Geophysical Year (IGY) of 1957 and fortunately continued every year from March to October. This is how Farman discovered the so-called ozone hole in the Antarctic in 1982. It is interesting to note that satellite observations (Nimbus 7) from the same period did not detect the phenomenon because values lower than a certain level (180 Dobson units) were automatically deleted as being due to experimental error (Gribbin, 1988). The confirmation of ozone destruction by CFCs and other chemicals led to the adoption of the international convention for the protection of the ozone layer (Vienna, 1985) and its Montreal protocol of 1987. This was the first global environmental convention.

Inland waters became a concern with regard both to quantity and quality. The International Hydrological Decade (1965–1974), followed by the International Hydrological Program, assessed water stocks. Eutrophication of lakes and rivers became a large problem in many countries in the mid 1950s. The OECD-sponsored research in this area identified phosphorus as the main culprit. This made it possible to implement successful lake rehabilitation projects.

The IUCN, now the World Conservation Union, recognized early on the threat caused by the destruction of natural vegetation, in particular tropical forests, and the importance of the preservation of biological diversity.

The marine environment was the concern of, *inter alia*, the International Council for the Exploration of the Sea (ICES) together with the International Ocean Commission (IOC), and the Monaco Laboratory of the International Atomic Energy Agency. It organized inter-calibration exercises for the quantitative determination of very low concentrations of marine pollutants, following such events as the Minamata Bay mercury contamination and the detection of dichloro-diphenyl-trichloro-ethane (DDT) and polychlorinated biphenyl (PCB) residues in marine life.

Pollution by man-made chemicals was also a concern on land. Research showed that organo–mercuric compounds used for seed dressing killed birds in Sweden, that DDT made spring silent in the USA (Carson, 1962) and that it reduced predatory bird populations through thinning of eggshells and behavior modifications. As a consequence these pesticides were banned in many countries.

Mention should also be made of research on radioactive contamination, which was carried out in several countries, starting with the USA, in the Columbia River near Hanford (Washington) during World War II and culminating during the period of intensive nuclear weapons testing up to the test-ban treaty of 1963. Megaton-size hydrogen bombs detonated in the atmosphere generated global fallout, which

could be traced in all environmental compartments and along food chains, and it was discovered later that DDT residues circulated around the planet in the same way as strontium 90. The tracing of exposure pathways made possible by the high sensitivity of radiation detection, combined with radiobiological research, enabled the development of a methodology of radioprotection (critical radionuclide, critical pathway, critical target group) which was later adapted, *mutatis mutandis*, to chemical pollution and spurred the development of *eco-toxicology* (*see* Ecotoxicology, Volume 3).

INTERNATIONAL SCIENTIFIC ACTIVITIES IN THE ENVIRONMENTAL FIELD

The examples of early research, given above, led to more concerted programs aimed at gaining a better scientific understanding of all aspects of environmental issues. Many programs were carried out and funded at the national level and it would be impossible to review them here. However, most of these activities, especially if dealing with GEC, were related more or less closely to international efforts, which often provided them with a framework or at least a context. Therefore, only international projects and programs are brie y discussed hereafter, without any claim to comprehensiveness.

United Nations Environment Programme (UNEP) (*see also* UNEP (United Nations Environment Programme), Volume 4)

In considering science and environmental policy, special mention must be made of the role of UNEP. Originating from the 1972 UN Conference on the Human Environment in Stockholm, which was based, as already mentioned, on a considerable scientific input, UNEP is often called the environmental conscience of the UN system. It has remained a program, not an agency or organization, providing a mechanism to integrate a large number of intergovernmental, non-governmental and national activities aimed at protecting the environment.

Despite its structural weakness and chronic under-funding, UNEP has many achievements to its credit, in the areas of global monitoring, environmental information and assessment, pollution control, chemical data banks (over 8000 chemicals in the International Register of Potentially Toxic Chemicals (IRPTC)), protection of marine resources, biodiversity (e.g., the Global Biodiversity Assessment of 1995, carried out by over 1500 scientists), climate (UNEP was co-founder of the IPCC, see later), desertification control, cleaner production, technology transfer, environmental education and awareness. One of its great successes is undoubtedly the 1985 Vienna Convention on the protection of the ozone layer, already mentioned, which was solidly grounded in science.

UNEP has always considered that one of its most important functions was the promotion of environmental science and information. It coordinated the production of a large number of state of the environment reports. More recently it started producing global environment outlooks (GEOs) – GEO 1, GEO 2000. These aim at providing an overview of major global environment concerns, their causes and impacts, and they identify, from a regional perspective, the priority issues that the international community should address. They are also meant to generate a continuous global assessment process through a scientific and policy-consultative mechanism with strong regional participation, by means of a network of public and non-governmental collaborating centers.

The global environment monitoring system (GEMS) was set up early by UNEP as a worldwide collective effort to monitor and assess the health of the global environment. Later, the global resource information database (GRID) was added to provide reliable geo-referenced data and information useful for policy purposes and decision-making. These two activities and the GEO process have now been integrated in a new Environmental Observing and Assessment Strategy, which enhances UNEP's role in this area and enjoys strong political support. The functional elements of this strategy are: assessment and reporting, environmental observing, data analysis and integration, and strategic oversight and early warning. This last element implies strengthening ties to the scientific community to ensure that the best scientific insights are available and, conversely, that research priorities to resolve uncertainties or to guide management decisions are made known to scientists.

To respond to the challenge of the goal of environmentally sustainable development, UNEP has moved into the area of environment and economics, concentrating on methodologies. This includes research on valuation of environmental and natural resources to be used in cost–benefit analysis, risk assessment, environmental impact assessment of projects and programs, and internalization of environmental costs. Work is also done on integrated environmental and economic accounting (IEEA) for national accounts in order, *inter alia* to bring out the implications of development plans on the depletion and degradation of natural resources, as well as on economic instruments.

From its inception, UNEP's leadership has consistently supported scientific research and worked closely with the ICSU and, in particular, SCOPE.

ICSU (*see also* ICSU (International Council for Science), Volume 4)

ICSU, now appropriately called the International Council for Science, has played a pioneering and major role,

on the international scene, in applications of science in environmental policy formulation.

As an independent body representing science on a worldwide level, ICSU is ideally suited to serve as a partner to policy makers to provide scientific understanding of problems of the global environment. As such, it helped, through its SCOPE committee, the UN in preparing the 1972 Stockholm Conference on the Human Environment and was the scientific advisor for the 1992 Rio United Nations Conference on Environment and Development (UNCED). ICSU's environmental involvement, however, dates back to 1957 when it organized the IGY, which served as a model for subsequent international environmental research and assessment programs (Greenaway, 1996). IGY was very successful in bringing scientists from many countries to make coordinated measurements of physical parameters at the time of the cold war, and it led eventually to the Antarctic Treaty. It served as a model for other international programs (*see* IGY (International Geophysical Year), Volume 1).

The International Biological Program (IBP), ran from 1964–1974 (biological processes require generally more time to be studied than physical phenomena). Its main thrust was to investigate biological productivity on land, in freshwater and marine environments. It also included a section on human adaptability. It resulted in a very large body of information on the productivity of ecosystems of all kinds and in many places and, as such, provided a valuable basis for later assessments of the effects of human disturbance on the environment (*see* International Biosphere Program (IBP) a Foundation for Global Studies, Volume 2).

SCOPE (*see also* SCOPE (Scientific Committee on Problems of the Environment), Volume 4)

SCOPE, was established by ICSU in 1969, following an initiative of T F Malone. Its mandate is to advance knowledge of the in uence of humans on their environment, as well as the effects of these environmental changes upon people ••• as well as to serve as a non-governmental, interdisciplinary and international council of scientists and as a source of advice for the benefit of governments ••• . Its function is thus to produce authoritative assessments of environmental problems in order to identify and prioritize research needs and to summarize scientific knowledge for use by policy makers.

As an interface between science and decision making, SCOPE has focused on major environmental issues such as environmental monitoring, climate, biogeochemical cycles, environmental chemicals and eco-toxicology, the environmental consequences of nuclear war, the role of biodiversity, invasive species, etc., and more recently to environmental sustainability as well as environment and economics. It has spawned a number of international research programs, such as the WCRP and IGBP mentioned later.

Man and Biosphere (MAB) (United Nations Educational, Scientific and Cultural Organization (UNESCO)) (*see also* UNESCO (United Nations Educational Scientific and Cultural Organization) and the Environment, Volume 4)

IBP, a non-governmental scientific program of ICSU led to the still-running intergovernmental MAB program of UNESCO. Its general objective is to develop the basis within the natural and social sciences for the rational use and conservation of the resources of the biosphere and for the improvement of the global relationship between Man and the environment; to predict the consequences of today's action on tomorrow's world and thereby to increase Man's ability to manage efficiently the natural resources of the biosphere.

MAB launched a large number of projects dealing particularly with the ecological effects of human activities on tropical and temperate forests, grasslands and savannas, arid ecosystems, freshwater and wetland systems, tundras, and interestingly, the ecological aspects of urban systems. After almost three decades, MAB has increased considerably the knowledge of the biosphere and the role of humans in its modifications. One of its major achievements is the network of *Biosphere Reserves* (*see* Biosphere Reserves, Volume 2), of which there are now 352, some of them very large, providing not only a rich basis for the study of biodiversity and its preservation, but also sites for the Global Terrestrial Observing System (GTOS), see the section on the global observing program. Over the years MAB has evolved and is now focusing on new approaches to facilitate sustainable development.

IUCN, Now the World Conservation Union (*see also* IUCN (The World Conservation Union), Volume 4)

While only partially a research organization, IUCN, which celebrated recently its 50th anniversary, has been an important international body whose mission to conserve the integrity and diversity of nature, and ensure that any use of natural resources is equitable and ecologically sustainable is embedded in science.

It is not a pure NGO (non-governmental organization) but a union of governments, governmental agencies (such as the Environmental Protection Agency in the USA) and NGO's working at protecting nature in the field and at the policy level. It has enlisted the help of many scientists and works closely with ICSU and its unions and committees as well as with UN organizations. Without going into detail, suffice it to say that, besides its well-known *red lists of threatened*

species, it was the first to advocate sustainable development in its World Conservation Strategy in the late 1970s, a concept adopted by the Brundtland Commission in 1987 and endorsed by the 1992 UNCED in Rio, and now the official policy of many countries and, among others, the EU. IUCN was also responsible for drafting the convention on the protection of biological diversity, which was eventually signed in Rio.

The UN Food and Agriculture Organization (FAO)

The mandate of the FAO is to raise levels of nutrition and standards of living, to improve agricultural productivity and to better the conditions of rural populations. Its motto is the access of all people at all times to the food they need for an active and healthy life. It is active in plant and animal production, forestry, fisheries, land and water development, nutrition, etc., but also in social and economic policies. A current priority is to encourage sustainable agriculture and rural development by means of a long-term strategy for the conservation and management of natural resources.

Together with the Consultative Group on International Agricultural Research (CGIAR), an informal association of donors which it sponsors jointly with UNEP and the United Nations Development Program (UNDP), FAO addresses the environmental problems generated by the otherwise beneficial green revolution which was initiated by the Rockefeller Foundation in the 1950s.

In pursuing its general objectives, FAO promotes scientific and technological research and maintains data bases of interest to environmental policy, for example in the areas of climate, water and soil resources, terrestrial ecosystem monitoring, etc.

World Meteorological Organization (WMO) (*see also* WMO (World Meteorological Organization), Volume 1)

WMO is concerned with weather and climate. It is among the sponsors of the WCRP and the IPCC. It also publishes periodic reports on the stratospheric ozone layer and other data of interest to environmental policy.

WCRP

This program is jointly supported by the WMO, ICSU and the IOC, itself linked to UNESCO. Initiated in 1980, it succeeded GARP, dating from 1967. It is aimed at developing the fundamental scientific understanding of the climate system and climate processes needed to determine to what extent climate can be predicted and the extent of Man's in uence on climate. The program includes not only observational and diagnostic research leading to a quantitative understanding of climate processes, such as exchanges between atmosphere, ocean and sea-ice, interactions between radiation, land surface and the hydrological cycle, but also the development of global climate models able to simulate the present climate and to predict climate variations at different time and space scales. It has a number of projects known by their acronyms: GEWEX (Global Energy and Water Cycle Experiment), CLIVAR (Climate Variability and Predictability), TOGA (Tropical Ocean Global Atmosphere), WOCE (World Ocean Circulation Experiment), etc., the results of which have provided substantial inputs to the assessments carried out by the IPCC, see later.

IGBP (*see also* IGBP Core Projects, Volume 2; IGBP (International Geosphere–Biosphere Programme), Volume 2)

As WCRP developed, it became evident that there should be a broadening of climate change to the concept of global change. SCOPE was instrumental in promoting this idea. G F White, its president at the time, and M K Tolba, Executive Director of UNEP, issued a declaration on Global Life Support Systems in 1979 in which they said ••• we draw attention to the fundamental scientific importance of understanding the biogeochemical cycles which link and unify the major chemical and biological processes of the earth's surface and the atmosphere. The idea was picked up by ICSU and eventually the IGBP was launched in 1986. Like other international environmental research programs, it provides a framework developed and controlled by a central scientific committee, within which countries and bodies such as the EU can fund and implement their own scientific priorities but in an integrated and coordinated way, thus using their resources more efficiently. Central funding is a persistent problem in these cases, hence the creation of the International Group of Funding Agencies (IGFA), which attempts not only to provide what was called glue money for this purpose but also to see that at the national or regional level relevant research is financed.

The goal of IGBP is to describe and understand the interactive physical, chemical and biological processes that regulate the total Earth system, the unique environment that it provides for life, the changes that are occurring in this system, and the manner in which they are in uenced by human actions.

The science plan was published in 1990 after long and careful preparation. Six key questions are currently addressed by eight so-called core projects (acronyms given below). These are:

- how is the chemistry of the global atmosphere regulated and what is the role of biological processes in producing and consuming trace gases? (International Global Atmospheric Chemistry Project, IGAC);

- how will global changes affect terrestrial ecosystems? (Global Change and Terrestrial Ecosystems, GCTE and Land Use and Land Cover Change Project, LUCC);
- how does vegetation interact with physical processes of the hydrological cycle? (Biospheric Aspects of the Hydrological Cycle, BAHC);
- how will changes in land-use, sea level and climate alter coastal ecosystems, and what are the wider consequences? (Land–Ocean Interactions in the Coastal Zone, LOICZ);
- how do ocean biogeochemical processes in uence and respond to climate change? (Joint Global Ocean Flux Study, JGOFS and Global Ocean Ecosystem Dynamics, GLOBEC);
- what significant climate and environmental changes have occurred in the past and what were their causes? (past global changes, PAGES).

There are also crosscutting activities dealing with data and information systems, global analysis, interpretation and modeling (Global Analysis, Integration, and Modeling, GAIM), analysis, research and training at the regional level (global change system for analysis, research and training, START).

Thus, IGBP is a strategic research effort, which, while dealing with fundamental scientific issues, focuses on those questions that are most likely to lead to a practical and predictive capability, and consequently contribute to policy decision-making. IGBP has already produced a wealth of data. Just to mention one example, much of the scientific basis for the 1995 report (up-dated in 1997) of IPCC Working Group (WG) II on the impacts of climate change comes from IGBP-related research.

International Human Dimension Program (IHDP) (*see also* IHDP (International Human Dimensions Programme on Global Change), Volume 4)

After the initiation of IGBP, the international scientific community came to realize that the response of human societies to GEC was as important to understand as the physical mechanisms involved, in order to devise policies to mitigate and adapt to it. Hence the IHDP which was launched by the International Social Science Council (ISSC) in 1990 and consolidated in 1996 when ICSU joined in.

Human dimensions are defined as the ways in which individuals and societies contribute to GEC, are in uenced by it and mitigate it and adapt to it.

After a slow start, IHDP seems to be running well, with, currently, four science projects:

- LUCC jointly with IGBP;
- GEC and human security (GECHS);
- institutional dimensions of GEC (IDGEC));
- industrial transformation (IT).

DIVERSITAS (*see* DIVERSITAS, Volume 2)

This is a program aiming at coordinating a worldwide research effort on biological diversity. Established in 1991, it is sponsored by UNESCO and various bodies of ICSU, including SCOPE and the International Union of Biological Sciences (IUBS) thus involving NGOs and an intergovernmental bodies. Its goal is to provide accurate scientific information and predictive models of the status of biodiversity and sustainability of the use of biotic resources as well as to build a worldwide capacity for the science of biodiversity. It is expected to provide scientific input to the implementation of the Convention on Biological Diversity. Its achievements have been somewhat limited so far. Noteworthy is the work done by SCOPE on the ecosystem function of biodiversity and on the problem of invasive species, the latter now undertaken with the financial support of GEF (Global Environmental Facility), a joint program of the World Bank UNDP, and UNEP, see later). DIVERSITAS is now organizing an International Biodiversity Observation Year (IBOY), which should begin in 2001.

Global Observing Programs

The purpose of these programs is to provide a continuous check of the environmental status of the planet. They include the Global Climate Observing System (GCOS), which collects and disseminates data, on, e.g., the El Nino phenomenon, the Global Ocean Observing System (GOOS) with modules on pollution, living marine resources, coastal problems, and the Global Terrestrial Observing System (GTOS) looking at changes in land quality, freshwater resources, biodiversity, pollution, etc. Using remote-sensing data from space as well as ground level observations, the overall system aims at bringing science and policymaking closer together in order to address global change issues.

The systems are not yet all operational. If properly funded, they should provide invaluable help in framing and assessing the effectiveness of worldwide environmental policies.

IPCC (*see also* Intergovernmental Panel on Climate Change (IPCC): an Historical Review, Volume 4)

IPCC is a unique body. Placed at the interface between science and policy for the issue of man-induced climate change, it was set up in 1988 by WMO and UNEP with a scientific mandate but within a political institutional framework. Because of its *modus operandi* it is actually more a transitional zone that a sharp interface. Its most scientific component is WG I, which provides assessments of the science of climate change while WG II assesses the impacts of climate change. Initially, WG III was established to formulate response strategies but in 1992 WGs II and III

were merged with the task of assessing response options in addition to impacts; and a new WG III was created to look at the socio-economic dimension of climate change and other cross-cutting issues.

The IPCC decision-making process involves three levels: the scientific core in each WG, the WG plenary and the full plenary. In the plenaries, every line in the draft reports is discussed at length until a consensus is reached. This leads sometimes to situations where statements accepted by the scientists are watered down to satisfy representatives from countries with no notable scientific expertise but strong vested economic interests in, e.g., fossil fuels.

IPCC, however, has succeeded in enlisting the participation of most leading scientists in the relevant fields who review all published literature (including that originating from WCRP and IGBP). It has produced so far two major assessment reports with executive summaries (1990 and 1995) plus many scientific and technical papers, and is preparing its third assessment report. One of its most well known statements (1995) is that the balance of evidence suggests that there is a discernible human in uence on global climate. There are still some *bona de* scientists, however, who dispute this statement, not to speak of various national and international industrial lobbies.

Four years after the establishment of IPCC the United Nations Framework Convention on Climate Change (UNFCCC) was signed in 1992 in Rio and a Conference of the Parties was subsequently set up. The fact that the Kyoto Protocol of 1997 includes obligations to reduce greenhouse gases emissions (5.2% below the 1990 levels by 2012!) is considered a mark of the success of IPCC, even though the reduction is much too small to achieve what IPCC estimates to be necessary to slow down global warming to an acceptable rate.

Be that as it may, IPCC may be considered as a model exemplar in the science-environmental policy dialogue. Its integrated assessments have so far not been replicated for other global environmental issues.

ASCEND 21

ASCEND 21 stands for an Agenda of Science for Environment and Development into the 21st Century. It was prepared by ICSU in cooperation with the Third World Academy of Sciences (TWAS) through a conference, which took place in Vienna in 1991. Aimed at defining a science agenda for the next 10–20 years, it was to provide a consolidated contribution to the 1992 UNCED Conference (where ICSU was the official scientific advising body) as well as a perspective for the future direction of international science in the areas of environment and development.

It proved to be a most valuable exercise in that top scientists from around the world jointly produced a broad review of the issues of highest priority, including the unknowns and uncertainties, and recommendations for mid and long term research. It covered the major driving forces (population, land use, industry and wastes, energy, health), scientific understanding of the Earth system (global cycles, atmosphere, marine and coastal systems, land resources, fresh water resources) and the contribution of science to environment and development strategies (quality of life, public awareness, capacity building, policies for technology, institutional arrangements).

The conference provided a fundamental input for the formulation of Agenda 21 adopted in Rio.

Global Environmental Facility (GEF) and Scientific and Technical Advisory Panel (STAP)

The GEF was established in 1991 as a joint program between UNDP, UNEP and the World Bank, in the context of preparation for UNCED, in order to help solve global environmental problems. Essentially, it provides funding, mostly to developing countries, to meet the costs of measures to fulfill international obligations in four areas related to GEC, namely the protection of biodiversity, of the ozone layer and of international waters, as well as the reduction of greenhouse gases emissions.

It was soon realized that before taking decisions to spend large sums of money on projects, GEF needed sound scientific advice, hence the setting up of STAP. STAP provides objective, strategic and technical advice on GEF policies, operational strategies and programs and also performs reviews of projects, using a roster of experts. One of the main studies it is currently conducting is the global international waters assessment (GIWA).

Commission for Sustainable Development (CSD), (*see also* CSD (Commission on Sustainable Development), Volume 4)

The UN CSD was set up to ensure effective follow-up of UNCED and in particular of the implementation of Agenda 21. It does not have any scientific program as such but some of its activities call for scientific information and research. An example is the definition and testing of indicators of sustainable development, which has spurred much research on this subject throughout the world.

OECD (*see also* OECD (Organization for Economic Co-operation and Development), Volume 4)

The OECD, sometimes called the rich countries' club, although concerned mostly with economic issues, was a pioneer in the assessment of environmental chemicals, i.e., manufactured substances that found their way into the environment and could cause harm to humans and ecosystems. The reason for concern by an economic

organization was that uneven constraints put to the marketing of new chemicals would interfere with free trade. Testing systems were devised for the evaluation of degradability, pathways, exposure patterns, toxicological and eco-toxicological effects, etc., stimulating research in these fields. More recently OECD was instrumental in organizing work-sharing in the assessment of high production volume chemicals, marketed since before systematic testing was imposed.

OECD has also done significant work in environmental data reporting on the basis of the pressure–state–impact (PSI) scheme, on assessment of country environmental performance, on the development of a core set of environmental indicators and it is now much involved in environmental economic issues, sustainable consumption and production, as well as eco-efficiency.

World Health Organization (WHO)–International Program on Chemical Safety (IPCS)

WHO has been concerned with environmental health and, in cooperation with UNEP and ILO (International Labor Organization), launched in 1980 the IPCS which developed a number of valuable activities in two directions: to establish the basis for the scientific health and environmental risk assessment for safe use of chemicals, and to strengthen national capabilities for chemical safety, especially in developing countries. IPCS has promoted methodological research on health and environmental assessment of chemicals and established guideline values for many chemicals. It was the initiator of the International Forum on Chemical Safety in 1994, where government representatives meet to reach consensus in implementing Chapter 19 (Chemicals) of Agenda 21. WHO has recently downgraded IPCS, which is unfortunate considering the persisting concern with chemical risk, to wit the current scare about endocrine disruptors.

United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR)

UNSCEAR reviews and assesses scientific data on the biological effects of ionizing radiation and has published much information on environmental radioactive contamination, critical radionuclides, pathways and population groups. It provides background material to the international commission on radiation protection (ICRP) for setting and revising radioprotection norms.

United Nations Economic Commission for Europe (UNECE)

The UNECE (which covers also the USA and Canada) was an early player in environmental issues and provided a good example of successful regional science-policy cooperation. A major achievement was the 1979 Geneva Convention on long-range transboundary air pollution (LRTAP), originating from concern over acid precipitation and based on a well-coordinated research effort (involving among others the International Institute of Applied System Analysis (IIASA) where scientists from the Soviet block cooperated with those from the West). This involved inventories of SO_2 emissions in Europe and modeling of transport so that countries could be assigned emission reduction targets. The convention established a broad framework for cooperative action on reducing the impact of air pollution and set up a process for negotiating concrete measures to control emissions of air pollutants through legally binding protocols.

To implement this process a scientific program was set up, called EMEP (co-operative program for monitoring and evaluation of the long-range transmission of air pollutants in Europe). Its role is to supply governments with qualified scientific information to support the evaluation and further development of the various protocols on emission reductions (sulfur, nitrogen oxides, volatile organic compounds, heavy metals, persistent organic pollutants). It does this by collecting emission, air and precipitation data and modeling atmospheric transport and deposition of air pollution.

CURRENT CHALLENGES AND PRIORITIES FOR GEC SCIENCE

In attempting to discuss the role of global environmental science in the 21st century, one has first to consider the rapidly changing societal context in which environment policy has to operate. Spurred by geopolitical upheavals (e.g., the collapse of the Soviet empire) and technical progress (in particular information and communication technologies), the phenomenon of globalization has accelerated with the free movement of capital and goods, the general adoption of the Western model, the decreased power of governments in relation to that of multinational corporations and even of NGOs. The World Trade Organization (WTO) is now taking decisions affecting the global environment and that of individual countries, sometimes even ignoring existing multilateral environment agreements, and the environmental constituency in the UN is no match for it. There are also rising social and environmental expectations and changing consumer preferences.

The paradigm of sustainable development, with its social, economic and environmental components, has been formally adopted worldwide. It is generally agreed that, despite the vagueness of the concept, sustainable management and use of natural resources and sustainable patterns of production and consumption are the only ways to meet the needs of present and future generations. A central concept in this respect is *eco-ef ciency* which aims to make increases in production of goods and services generate proportionally

less environmental damage and less use of natural resources (*see* Eco-efficiency, Volume 4).

To strive toward this goal implies the integration of environmental and economic decision-making and the inclusion of an environmental dimension in all sectoral policies (agriculture, industry, energy, transport, tourism, etc.). A logical approach is the driving forces, pressures, state, impacts and responses (DPSIR)-framework (*see* DPSIR (Driving Forces Pressures State Impacts Responses), Volume 4). In the past, responses dealt mainly with correcting the impacts on human health and on the environment. Then came action on pressures, such as emission reductions of air and water pollutants. In future the emphasis should be put on action at the driving force level. This can be done either by modifying the processes and products, which provide a certain service or in managing demand for this same service, or more likely both. Policies should aim at optimizing the responses, helped by science, which should not only identify the issues but also contribute in determining the most effective options. Once a policy is implemented, its effectiveness should be checked, using performance indicators, especially at the economic sector level. Here again is a priority for research.

Another consideration is that of complexity and uncertainty. Major issues relating to natural systems (such as climate) are complex in themselves and fraught with uncertainty and this is compounded by the need for modifications of human behavior in order to implement solutions. Adding to the negative attitude toward science, which has been developing, particularly in the area of risk assessment and management, new ways must be found to achieve effective interaction between scientists, whether they be in the natural or social sciences, and all other stakeholders. Hence the development of *post-normal science*, i.e., science which takes account of the uncertainty of quantitative data and the biases of scientists themselves (*see* Post-normal Science, Volume 5). In this respect, the implementation of the precautionary principle poses problems. Precaution should not be opposed to science-based regulation but the two should be complementary. As Churchill said during World War II, science should be on tap, not on top, in the sense that decisions should always be based on the best available scientific information but that may not suffice. Scientific risk analysis is unavoidably and inextricably intertwined with subjective framing assumptions, values, trade-offs and expectations of surprise (Stirling, 1999).

Policy-related research is under strong pressure to provide magic numbers (as few as possible), to facilitate the tasks of political and administrative decision makers, hence the current high popularity of indicators. While these can undoubtedly be useful tools, much care must be exercised in selecting them and especially in clarifying their limits. Indicators are useless if the data on which they are based are lacking or unreliable. Hence the need for continuing time series of good quality data, and accessibility thereof, on the five components of the DPSIR scheme. These concern not only environmental, but also economic and social parameters. This huge task is currently undertaken by such bodies as the United States Environmental Agency (USEPA) or the European Environment Protection Agency, for their respective territories, while UNEP attempts to coordinate the information worldwide through its new Monitoring and Assessment Strategy.

It is now broadly recognized, as acknowledged again in a 1998 joint report of UNEP and the World Bank, that it is in a holistic, integrated manner that global environmental issues must be tackled, taking account of linkages between these issues (climate change, loss of biodiversity, stratospheric ozone depletion, freshwater degradation, land degradation and desertification, deforestation and unsustainable use of forests, marine environment and resource degradation, persistent organic pollutants) but also between environmental issues and human needs (food, shelter, health and clean water).

Against this background GEC science should of course continue to concentrate on basic research into the functioning of the life support systems of the planet and how these are affected by human activities, as done by WCRP and IGBP (which might avoid some of the surprises) but it should also detect and tackle emerging issues and do more policy-related research. As confirmed by a recent survey of environmental experts carried out by SCOPE for UNEP's GEO 2000, the results of which are summarized by Munn *et al.* (1999), the issues of high priority are not necessarily new. These issues could be either transformations of old issues in response to increasing scientific and technological knowledge and to evolving socio-economic, cultural and environmental conditions, or policy issues, such as the long-term environmental impacts of economic globalization.

Among the emerging science-driven issues those already well-recognized are: sustainable management of natural resources; socio-economic drivers of environmental change (population, globalization, changing production and consumption patterns, transport, tourism, urbanization); assessing the interactions amongst the global mega-issues, such as population, environment, development; intensification of environmental stressors, such as exposure to man-made chemicals or disruption of biogeochemical cycles. Another set of emerging issues consists of those with low or unknown probability but unacceptable consequences, such as natural disasters, biotechnology (genetically modified organisms (GMOs)), the rapid evolutionary nature of microbes and other pests and the collapse of ecosystems.

Policy-driven issues included, *inter alia*, environmental impacts of proposed strategies and policies (such as the Kyoto protocol on greenhouse gases emissions), implementation issues (e.g., how to implement the precautionary

principle) and how to make progress with well-recognized problems such as road transport.

The experts consulted in the SCOPE survey defined a large number of research needs related to these emerging environmental issues. Methodological needs came up preeminently as could be expected, considering the complexity of most issues and the necessity to combine quantitative natural science data with qualitative social science data. More specific priorities included: comprehensive environmental and health research in the tropical regions; ecosystem restructuring and movement in response to global change; more work on biogeochemical cycling, especially the decoupling of the carbon and nutrient cycles and the effects of the global nitrogen overload; the effects of multiple stresses on ecological and social systems; perturbations of the hydrological cycle; the environmental implications of GMOs; health-environment links; mechanisms of carbon dioxide sequestering in nature and through technology. Other environmental engineering R&D needs concerned chemicals (disposal of persistent substances, reduction of residues in food, risk assessment, chemical time bombs), cleaner production in general, improvement of water management systems and efficient energy systems. Economic issues such as natural resource and ecological accounting, the valuation of ecological services, etc., were also mentioned.

Essentially the same priorities for GEC science were identified in more general terms by the UNESCO–ICSU Conference on Science for the 21st century held in Budapest in June 1999. In its Science Agenda-Framework for Action, the Conference said, areas requiring special attention include the freshwater issue and the hydrological cycle, climate variations and change, oceans, coastal areas, polar regions, biodiversity, desertification, deforestation, biogeochemical cycles and natural hazards. It called for the vigorous pursuit of the international research programs such as IGBP, as well as for long-term monitoring of all components of the earth system. It also stressed the need for a solid scientific basis for technological development directed toward safe and clean production, greater efficiency in resource use and more environmentally friendly products.

No less importantly, and as general recommendation for all fields of science, the conference called for high ethical standards, in terms of scientific integrity and quality control, knowledge sharing, communication with the public and education of the younger generation. This applies particularly to GEC science, considering the need to establish relations of trust among all stakeholders to achieve progress toward the goal of sustainable development.

REFERENCES

Carson, R (1962) *Silent Spring*, Houghton-Mif in, Boston, MA, 1–368.

Greenaway, F (1996) *Science International. A History of the International Council of Scienti c Unions*, Cambridge University Press, Cambridge.

Gribbin, J (1988) *The Hole in the Sky, Man's Threat to the Ozone Layer*, Bantam Books, New York.

Holdgate, M W and White, G F (1977) *Environmental Issues*, SCOPE Report 10, John Wiley & Sons, Chichester.

ICSU, Brennan, M (1992) *ASCEND 21. An Agenda of Science for Environment and Development into the 21st Century*, eds J C I Dooge, G T Goodman, J W M la Rivière, J Marton-Lefèvre, T O'Riordan, and F Praderie, Cambridge University Press, Cambridge.

Munn, T, Whyte, A, and Timmerman, P (1999) Emerging Environmental Issues: a Global Perspective of SCOPE, *Ambio*, 28, 464–471.

Stirling, A (1999) *On Science and Precaution in the Management of Technological Risk*, Internal Report to the Forward Studies Group, European Commission, May 1999, University of Sussex, UK.

UNEP (1999) *GEO 2000 (Global Environment Outlook)*, UNEP's Millenium Report on the Environment, Earthscan Publications Ltd., London.

Vernadski, W (1924) *La Géochimie*, Félix Alcan, Paris, 1–403.

Policy Responses to Public Health Issues Relating to Global Environmental Change

ALEYA EL BINDARI HAMMAD

The Robert S Wagner School of Public Health Services, New York, NY, USA

Health security rests on the principle of equity, which allows human beings to live with the security of knowing that they have access to health care which is affordable, relevant and of good quality. Health security is needed throughout the entire life span of human beings, and encompasses all aspects of the right of every individual to the highest attainable standard of physical and mental health, including the right to food in suf cient quantity and of good quality, the right to decent housing, the right to live and work in an environment where known health risks are minimized, and the right of access to education and information on health.

The WHO Task Force on Health in Development (1994) Geneva.

There are three inextricable links in the chain of human development - health security, economic development, and environmental protection. At the end of the 20th century, only one of these has seen great positive change from its condition at the beginning of the century. Economic growth has taken place at an unprecedented rate. When wealth is examined from a safe distance, and as a mathematical mean, the planet is, scally at least, wealthier several-fold. Yet for all the additional comfort and improvements in quality of life that this should herald, we nd that we do not fare so well. The environment teeters dangerously on the brink of complete destruction. Health indicators show severe regression in the human condition; old diseases are back with a vengeance, and new ones are daily created by our poor understanding and use of our resources, both natural and derivative. What is perhaps most worrisome is that public health, the eld which contributed most to the improvements in human health and which spurred all development and innovation, is in decline. The global decline in the profession is mirrored in the decline in the health of each individual in any society, developed or developing. In early 1940, the discovery of therapies, especially the sulfonamide group of drugs and insulin, contributed to an increase in therapeutic interventions and gave tremendous con dence to new technical and scienti c approaches in medical care. With the control of infectious diseases and the acceleration of scienti c discovery in the medical eld, the in uence of public health diminished, giving hospital-based curative services a more prominent place, and gradually the lion's share of resources. Nevertheless, public health in the mid 20th century contributed tremendously to the improvement of the general environment and to major programs for controlling communicable diseases such as cholera. Public health strategies were pivotal in malaria control programs and in the growth of the health profession. This last was accomplished by an investment in institutional capacity-building through the establishment of schools for health workers for the development of public health cadres such as sanitarians, nurses, and health controllers. It is during this period that quarantine and other preventive services were established, through the leadership of WHO (World Health Organization) greatly enhancing the free movement of people, goods and services between countries.

Declarations, conventions, and conferences have made a virtual truism of the statement that human development and environmental protection are mutually dependent. And yet, at a time when we are armed both with the knowledge that human beings are at the center of concerns for sustainable development, that the right to development must be ful lled so as to equitably meet developmental and environmental needs of present and future generations, and that human beings are entitled to a healthy and productive life in harmony with nature, the actions taken so far continue to treat health and environmental concerns as separate disciplines (Rio Declaration, 1992). Conferences on health are held separately from those on the environment. Economic conferences pay at least super cial attention to environmental questions but almost none to health concerns. Health protection and promotion have received little recognition in environmental policies and development plans, and have

rarely been identi ed as an important factor in environmental or development programs. That the most pressing problems of ill health and premature death are caused by biological agents in the human environment - in water, food and soil - seem to be nothing more than curious side effects, that can be solved without direct attention to themselves.

The unacceptable numbers of infants and children dying from diarrheal diseases is largely the result of contaminated food or water. Likewise, hundreds of human beings, very often at the peak of their productivity, suffer unnecessarily from debilitating intestinal infections seeping them from their energies; malaria continues to take its toll unabated. Whilst many of these health risks are found in the developing world, there still are serious environmental health problems shared by both industrialized and developing countries. These include respiratory ailments caused by biological agents in the air, both indoors and outdoors, and chemical and physical hazards in their home, workplace or wider environment.

The third link in the chain - economic growth, based on market models - has widened rather than narrowed the gap between health and the environment. Social conditions are treated in the market as costs and risks, not as prerequisites for and aspirations of any civilization. The emphasis of the market on the preeminence of the individual, the vast monetary fortunes amassed by players in the market, and the global celebration of consumption and ostentatious display of wealth, have each contributed to a growing public perception that there is no need to look out for the needs of other individuals. Others are to be viewed as competitors. This perception has contributed to the decline of the notion of society and as such, to the decline of any institutions and professions that place value and respond to humanity as a society of beings interdependent on each other and on the living environment.

Sustained, unregulated, economic growth has been shown, in a growing body of evidence, to have negative consequences on the environment and biophysical resources. Yet here again, where inroads have been made against the cycles of production and consumption that deplete resources at an alarming rate, they have been in the evidence presented of the risks posed to the individual's health.

It is as if the concept of human development has set about to sabotage itself, by pitting the values of one link - economic growth and individual prosperity - against the values of the other two - protection and enhancement of all those factors which contribute to harmony. In part as a result of this, public health concerns continue to be pushed to the margins. Yet sustained economic growth, particularly if the desired outcome is measured by the levels at which consumption occurs in the West, is unsustainable in its impact on the environment and unpalatable in its consequences for the whole human condition - taking into account the physical, psychic, intellectual, emotional, and mental dimensions.

It is perhaps time for those concerned about human development to make known to the public, government, and private sectors the link between environmental and health concerns, and to use their conjoined strength to put an end to business as usual.

It is our ability to take these facts and devise ways of better managing the interaction between human activities, and the physical and biological environment that will enable us to meet current and future needs sustainably.

THE BEGINNINGS OF PUBLIC HEALTH AWARENESS

Health is worth the world, There is nothing more precious than health! are common sayings in many cultures. Usually, they are evoked with nostalgia by those who are passing through either temporary or permanent episodes of ill health. The human being very often has taken for granted the enjoyment of good health and is reminded of it only when he/she loses it.

From ancient history to the present day, most cultures have continued to exercise rituals and practices and have adhered to various rules to restore health and sometimes, to protect it. Some of these traditional practices work, and extensive research has and continues to be undertaken to discover the therapeutic basis for such healing practices. Many traditional healing practices draw their ingredients from the environment – plants, natural springs, the sand, the sea, the earth, and the atmosphere. It is the living environment that is the cornerstone for good health.

In addition to traditional healing practices, public health is as old as humankind and there is evidence that many public health measures were practiced by ancient civilizations. It has evolved throughout the years to reach its present form, which dates back to 19th century Europe during the Industrial Revolution. At that time, public health concerns were linked predominantly to sanitary action required to address problems caused by environmental and lifestyle factors. More specifically, in the first instance, they dealt with issues emanating from overcrowding, food and potable

water resulting from rural-urban migration. Many of the public health laws that were passed, such as those in England, dealt with environmental issues of sanitation, clean water, and clean cities. These activities proved to be very successful in the emergence of developed countries. With the development of germ theory in the 1870s, the introduction of massive vaccination and immunization practices was added to public health functions. It is this phase that ushered in many of the social services addressed to community groups, such as school health services and community health projects.

In 1923, Winslow provided a definition of public health that has been widely used to define its scope and function, and provide the basis upon which public health programs were designed. In 1952 and 1953, expert committees from the WHO amended slightly Winslow's definition to include both mental and physical health, as follows:

> Public health is the science and art of preventing disease, prolonging life, and promoting mental and physical health and efficiency through the organized community efforts for the sanitation of the environment, the control of communicable infections, the education of the individual in personal hygiene, the organization of medical and nursing services for the early diagnosis and preventive treatment of disease and the development of social machinery to ensure to every individual a standard of living adequate for the maintenance of health, so organizing these benefits as to enable every citizen to realize his birthright of health and longevity.
>
> (WHO, 1952)

Studies of the period spanning 1801–1971, undertaken by scholars such as McKeown, showed that health indicators such as those in England in the mid-1970s, improved primarily as a result of better standards of living and public health initiatives like better sanitation, clean water and better housing, rather than by the provision of medical care *per se*. The dictum that prevention is better than cure proved its worth in countries such as China and Sri Lanka and in the State of Kerala, India, in which there was a strong commitment to the goals of equity and concerted efforts to ameliorate the conditions of disadvantaged and poorer social groups. Many resources for satisfying health needs were health-related resources from other sectors, namely the entitlement to adequate food; basic education and knowledge, which enhanced the capacity to cope with health problems, particularly the capacity of women to deal with maternity and child care; and a well-distributed social and economic infrastructure leading to improved housing, water, sanitation and transportation facilities, including ready access to health services. All these improvements built the capacities for families and communities to lift themselves out of conditions of absolute poverty and the cluster of health risks it created. Even at income levels below US$350, life expectancies at birth in China and Sri Lanka as early as 1983 were estimated at 69 and 67 years, respectively – higher than the average life expectancy of 65 years for the upper-middle income group of countries, with an average *per capita* income of US$2050 (WHO, 1986).

Parallel to this, scientific knowledge and the development of technologies for curing diseases were progressing at a remarkable rate. Gradually, emphasis on public health and the development of the peripheral infrastructure weakened, and more weight was given to curative care, which increasingly absorbed the larger part of resources. As a result, urban-based curative hospitals were given priority over peripheral health units, greatly contributing to inequities in health status both within and between countries. The dominance of urban-based medical care exists to this day, and has contributed to the gap between urban- and rural-based health care, and ultimately between those who can afford and those denied health services. The inequity in health status is a phenomenon that characterizes many countries, both developed and developing.

The dominance of urban-based medical care in developing countries is reinforced by two very important values that predominate. The first is the appeal that medical high technology has on people and the value they place on obtaining this type of service. The second is the gradual decline in the quality of health care and in employment conditions in the public sector; and an attendant lowering in the confidence of individuals in state-owned and state-operated health institutions, occurred as governments began to economize on public expenditures, and the real wages and salaries of public servants began to fall behind those paid by the private sector. Patients were encouraged directly or indirectly to seek more and more services in the private sector if they had the ability to pay. Patients and the wider population began to perceive a difference – real or imagined – between services which further affected the quality of services delivered by the public sector. Public policy has so far failed to resolve these dilemmas. It should be stated, however, that by their very nature, the services provided by the private sector are demand-led and will therefore concentrate in the provision of services in areas where purchasing power is high (WHO, 1997a).

The costs of individual medical interventions have escalated around the world, but the state of health of the people has not improved in relation to expenditure. In fact, the resurgence of diseases such as cholera, schistosomiasis, tuberculosis, and the emergence of new diseases like the acquired immunodeficiency syndrome (AIDS) pandemic, has brought into focus the urgent need to revisit the whole question of how best to promote and ensure healthier nations within development strategies, and has focused attention again on public health measures as central to any debate on health.

Debates on health cannot be separated from debates on overall development. Patterns of development are dominated by economic forces, and recently, the model

of the market economy has been widely accepted. This development has promoted the drive towards systems that are consumer-oriented. The excessive use of market-driven policies has had negative social consequences on the poorest segments of society. This scenario has led to environmental degradation, taking place at an alarming rate, and continues to have serious consequences on health, as will be discussed in the ensuing paragraphs.

ENVIRONMENT AND HEALTH

Human beings live within an environment which constitutes an array of elements – physical, chemical, social, cultural and economic – varying according to local geography, season, infrastructure, time of day, and activity undertaken. Even though there is an intimate relation between the individual and the environment, this relation is highly complex. There is no single best way of viewing this relationship, which reveals all interactions and possible entry points for public health interventions, especially as the situation evolves and is constantly changing.

Global environmental change itself is a complex issue that is both the cause and the consequence of human dynamics. Before the middle of the 20th century, scant attention was paid to the notion that the manner in which humans interacted with their environment, and exploited resources, could permanently and negatively affect these same. If anything, innovations in the taming of the wild and harnessing of resources for human consumption were seen as advancements in the development and civilization of humanity, and perhaps rightly so. Innovations in water distribution and water purification through piping, damming, separation of potable water from sewage, for example, went a long way in the eradication of diseases that had threatened to destroy urban populations, and availed humans of natural resources for the production of energy. The same was true for medical, agricultural and industrial innovation.

Health depends on the ability to understand the interaction between the physical and biological environment, and the human activities undertaken. There are two concerns to be taken into consideration. The first is to respond concretely to people's needs, especially for health; the second is ecological sustainability, ensuring that natural resources are not depleted and natural systems not damaged or degraded by actions undertaken by humans, or for the benefit of humans. Meeting the basic human needs of present and future global populations for food, water and energy without depleting or damaging the global resource base as well as avoiding the adverse health and environmental consequences can be achieved only if individuals have the knowledge and the means to in uence action.

In 1989 at the first European Conference on the Environment and Health, environmental health was defined as:

> Environmental health comprises those aspects of human health and disease, that are determined by factors in the environment. It also refers to the theory and practice of assessing and controlling factors in the environment that can potentially affect health.

The conference agreed that important factors in the environment that affected health included, chemical and microbial agents and those factors, which affected the broad physical, psychological, social and aesthetic environment.

Beginning with the most basic needs, environmental health development stages can be characterized as:

- Physical survival – food, drinking water and shelter.
- Awareness raising – knowledge and understanding of factors affecting environmental health.
- Health and social care, environmental protection – responsive and appropriate health services, establishment of environmental management institutions.
- Preventive measures – controlling environmental pollution, prevention of illness.
- Promotion of environmental health – health supportive community structures, healthy and environmentally sound behavior and lifestyles.
- Well-being.

CHALLENGES FOR PROTECTING HEALTH

The economic circumstances and crisis, which predominated in the 1960s, and persisted into the 1980s and 1990s, resulted in major economic reforms and structural adjustments leading to a drastic curbing of public health expenditures. Old policies that were based on the fundamental values of equity and public responsibility, which characterized the public health strategies of the early part of this century up to the mid-1980s, were gradually replaced by new policies that promoted the market economy and privatization.

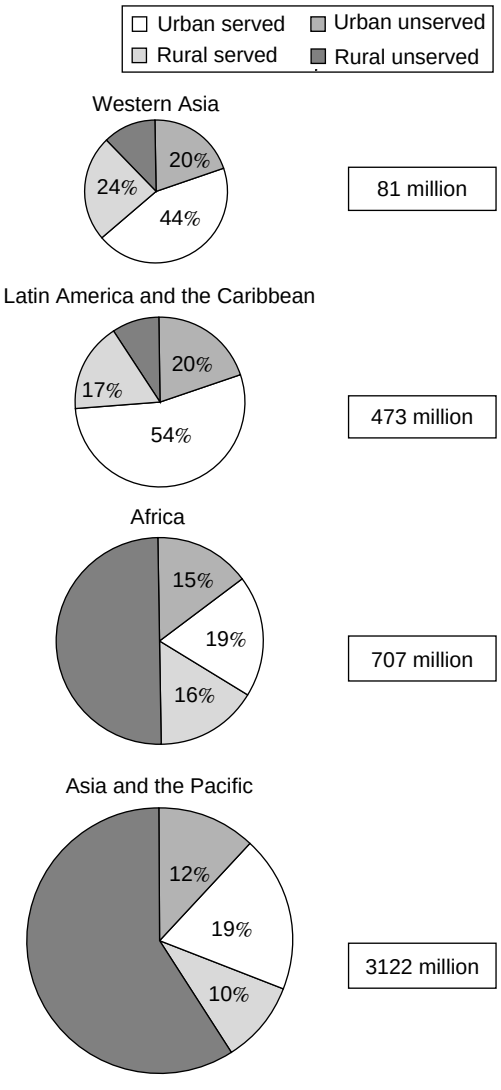
The failure of macroeconomic policies to ensure equitable distribution of wealth has contributed to increased poverty. Not only have poor people remained poor, but so have large numbers of poor nations, producing pockets of deprivation and subhuman living conditions in the urban areas that are sometimes worse than those in the rural areas (WHO, 1997b).

The contamination of water, air and food affects millions of people worldwide, exposing humans, in varying degrees, to pathogens and subsequent adverse health effects. In most situations, exposure to a wide array of health hazards occurs simultaneously, frequently resulting in the aggravation, exacerbation, and complication of health complaints. The problem of air pollution has reached global proportions, and the guidelines for air quality provided by WHO are often exceeded by a substantial margin. The developing countries register the most indoor exposure with negative

consequences on health, affecting more than 1000 million people (WHO, 1997c).

The contamination of water and food, inadequate sanitation, and little or no treatment of sewerage have contributed to much of disease incidence. In the developing countries, only 10% of sewage is believed to receive any type of treatment before being discharged into the environment. Much of the infections, injuries and vector-borne diseases are due to inadequate collection and disposal of solid waste, especially in the urban areas. The number of people without adequate sanitation is on the rise and is estimated to have surpassed 3000 million by the year 2000 (see Figures 1 and 2) (WHO, 1997c).

Water, especially safe water, is essential for life and for health. Up to 800 million people gained access to safe drinking water between 1990 and 1994. However, because of population growth, the net gain of percentage of the population covered was much smaller. It is believed that the scarce water supplies will constitute a major political issue in the 21st century, and will be the cause of wars and con icts between countries and regions. On the other hand, water pollution from the discharge of sewage and industrial waste remains a major problem leading to unacceptable human exposures due to the consumption of contaminated drinking water. Recently, the contamination of coastal water has become a major concern of many countries. Estimates suggest that if the provision of safe water and appropriate sanitation coverage are achieved for populations who are currently at risk, it would lead to the reduction of diarrheal deaths by two million, and the reduction of the prevalence of schistosomiasis and trachoma by 150 million and 75 million cases, respectively (see Figure 3) (WHO, 1997c).



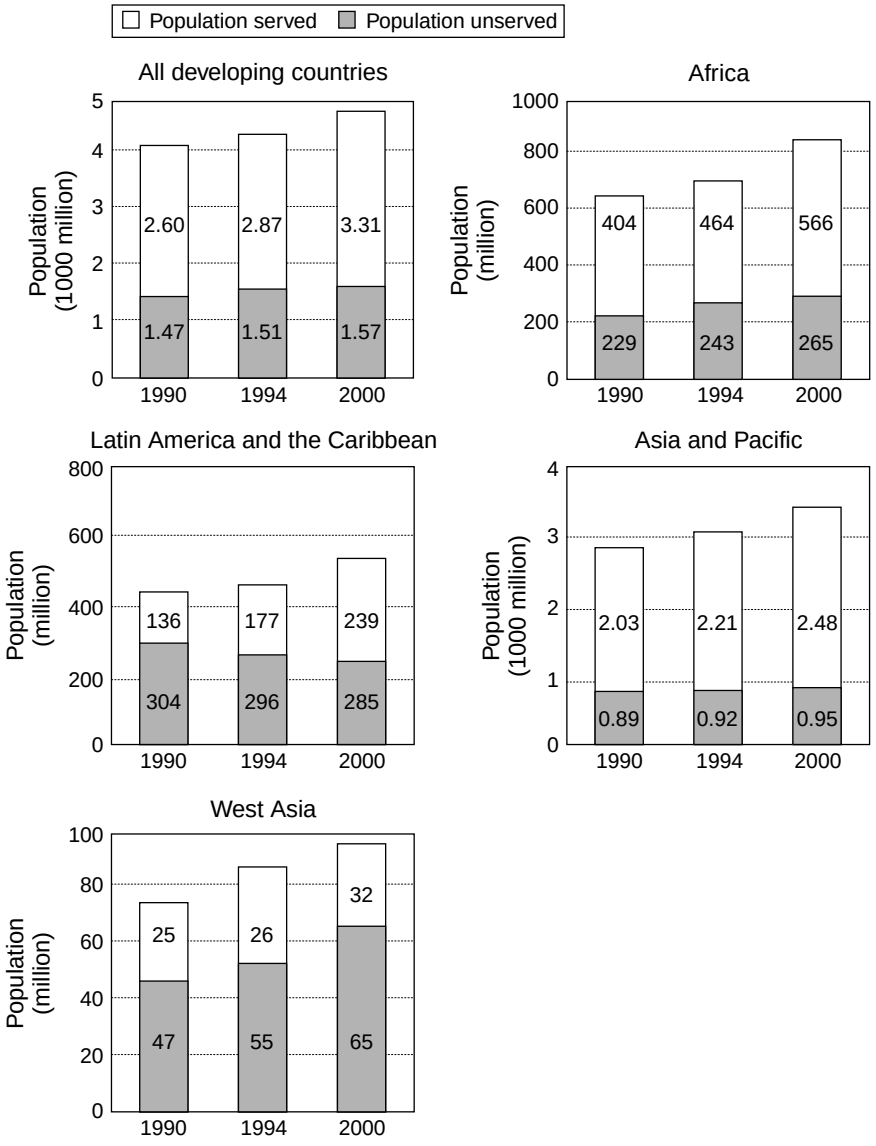
Source: WHO/UNICEF, 1996.

Figure 1 Sanitation coverage by region, 1994

Poor management of water resources and agricultural development has been largely responsible for the substantive increase of vector-borne disease such as malaria, schistosomiasis and dengue fever. There is no doubt that if past mistakes are to be avoided and lessons are to be drawn from activities undertaken, strong and coherent harmonization and planning between the relevant sectors concerned is paramount in order to ensure a more integrated approach in implementing health and environment projects.

Because food is essential to a healthy life, if not appropriately monitored nationally and internationally, it can become the route for many pathogens and toxic chemicals. The race for opening new markets by the food industry has led to the increasing use of chemicals in food production. These include food additives, pesticides, animal drugs, and other agro-chemicals such as fertilizers. The latter may be responsible for different types of health effects such as nausea, vomiting, impaired immune function, birth defects, neotoxicity, and cancer. Food contamination monitoring is becoming an increasingly important feature of food control program in both developed and developing countries. Whereas the monitoring and enforcement of public health regulations seem to be working in developed countries, which constantly uncover the tendency of industry to violate those regulations, standards and norms, developing countries may be especially vulnerable where either public health regulations are lacking, or their enforcement is ineffective.

There are many recent examples where the enforcement of such a regulation has prevented the major spread of food-borne diseases during the period of 1997–1999. The new variant of Creutzfeldt-Jakob disease (nvCJD) – a degenerative brain condition that appeared to be linked to the consumption of beef and other products from cattle suffering from, or infected by the agent-causing bovine spongi-form encephalopathy (mad cow disease) – resulted in the slaughtering of cows in the UK, and a ban on the export of beef from the UK to the European countries. The



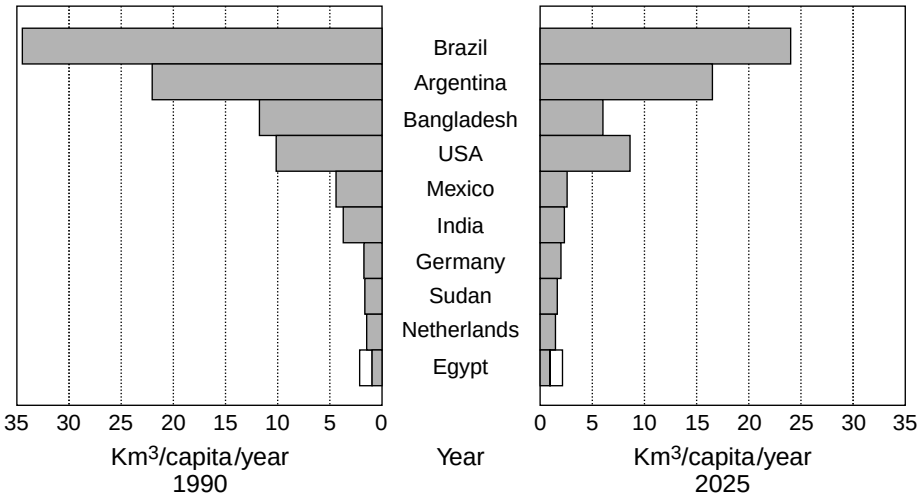
Source: WHO/UNICEF, 1996.

Figure 2 Sanitation services coverage to the year 2000 in developing countries

latter incident has put pressure on the European Union (EU) to tighten controls on animal feed. A recent food scare in Belgium was caused by the apparent contamination of feed with dioxin. The ARD German television network reported that companies in Brittany and Central France were putting residues from septic tanks and waste plants, into the feed of pork and poultry. Because of the threat to health, food-store chains in Germany withdrew the sale of poultry that might have been fed on meal mixed with sewage (The International Herald Tribune, 1999). These incidents did not happen in Europe alone. The contamination of chicken in Hong Kong resulted in the elimination of thousands of chickens by public health authorities to prevent further spread of disease throughout the territory.

The World Trade Organization plans to refer to Food and Agriculture Organization (FAO)/WHO codex alimentarius the arbitration of disputes involving food, health and safety requirements that seem to be on the increase, provided these are used as the basis of non-tariff barriers. Consequently, food contamination monitoring is becoming an increasingly important feature of food control programs in both developed and developing countries.

The contamination of foods by chemicals is also becoming an area of grave concern for the developing countries, and the major cause of diarrheal diseases. Statistics indicate that food-borne diseases of microbiological origin are increasing worldwide. This may be due to a slackening of standards, poor personal hygiene, and poor food control and



*UN medium population projection

Source: Based on data from Gleick, 1993 and UN, 1995.

Figure 3 Per capita annual fresh water available (km^3) in 1990 and 2025 for selected countries

inspection procedures. The health impacts range from mild indisposition to life-threatening illness. The people most affected are the poor, who are also vulnerable to lack of food and under-nutrition.

The contamination of soil by pathogens as well as by chemicals is a major contributor to human exposure to environmental health hazards. Whether contamination is due to poor sanitation or application of agricultural chemicals, improper disposal of hazardous waste or deposition of air-pollutant particles such as lead, it has become of particular concern. Soil contamination leads to ground-water pollution and uptake of contaminants by crops that are later on consumed by people (see Figure 4).

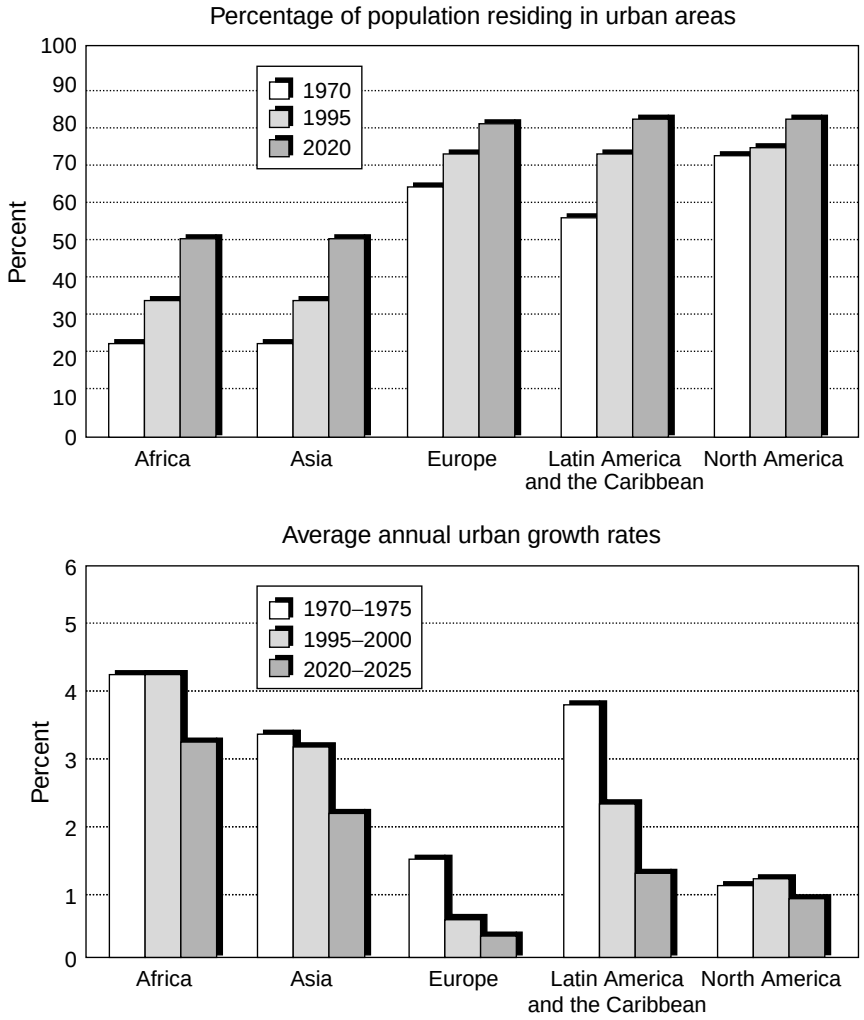
It is not surprising that with the present rate of urbanization, inadequate housing is becoming a prominent contributor to poor health. The absence of a safe and secure structure, lack of sanitation facilities, overcrowding, and lack of tenure and security, have resulted in ill health and the spread of infectious diseases. Around 50% of the poor in developing countries live in conditions of extreme deprivation.

With the emphasis placed on profit-making at any cost, indiscriminate industries have mushroomed in developing countries, with little or no respect for safety and health standards in the workplace. Environmental health hazards in the workplace are very serious, affecting a significant proportion of the workforce. The garment, bidi and carpet industries are notorious examples of hazardous working conditions with harmful effects on health. Many of the laborers are young children who are af icted with physical and mental injuries, in addition to infections and chronic ailments, which make them invalids for the rest of their lives. A great proportion of them do not survive beyond 35 years of age. Many occupational injuries and diseases are not reported, and therefore go undetected. Those that are recognized cause 125 million people to suffer each year from occupational injury (The International Herald Tribune, 1999).

The rather recent discoveries of global climate change and ozone layer depletion, points to further adverse effects on the health of populations. Among the hazards that are forecast are the increasing severity of heat waves, changes in the areas affected by vector-borne diseases, and an increase in ultraviolet exposure resulting in more skin cancers and cataracts.

Different environmental factors can contribute simultaneously to the occurrence of a given disease. This is the case for acute respiratory infections in children, for example, which is linked to several factors, including exposure to indoor air pollution, crowding and poor housing, and malnutrition. For respiratory disease to actually develop, exposure to other factors must occur. The prevention of acute respiratory infection therefore, will depend on a combination of interventions directed at the different contributing factors as illustrated in Table 1 below.

It is necessary to underscore the importance of getting away from the single-bullet interventions, frequently known as vertical programs, that not only absorb many resources, but have been unable to address other important aspects which contribute to the predominance of a disease. Only policies that enhance an integrated approach, ones that address multiple factors, can truly contribute to the health conditions that are potentially related to several environmental exposure situations.



Source: WHO, 1997.

Figure 4 Urbanization levels and urban growth rates by region, 1970–2025

HEALTH OUTCOMES AND VALUE SYSTEMS

Up to a few decades ago, resources on our planet were considered infinite, there to be exploited. It was only in the early 1960s that harmful environmental change began to be systematically recorded. Threats to human health were traced to the same development processes that had affected the environment and development. The first environmental protection or green activist movements focused their efforts on exposing environmental degradation as a result of industrial processes, by highlighting the negative health consequences on individuals. Their strategy, in bringing home the impact of working conditions and the manner of waste disposal, set much of the tone for the ensuing debate about environmental restoration. If industry was responsible for the destruction of the environment, let industry be exposed and brought to account. Tremendous success was seen, governments enacted bills – like the Occupational Safety and Health Act in the US – to protect the lives and health of workers. But when the same approach began to be applied to actions whose direct adverse in uences could be seen, such as deforestation, much more limited headway was made (see Figure 5).

The movement to protect the environment falls into two classes. The first tackles the health outcomes of environmentally unfriendly strategies. If the contamination of water by myself causes cholera or dysentery or typhoid and could in turn kill me, then I have a vested interest in clean water. Immediate problems are identified, and immediate results are seen from changes in behavior. This strategy aims at behavior modification. The more the problem is personalized and is seen as threatening to one's own survival and well-being, the more successful the strategies aiming at behavior modification.

Table 1 Potential relationships between exposure situations and health conditions

	Exposure situations					
Health conditions of concern	Polluted air	Excreta and household wastes	Polluted water of deficiencies in water management	Polluted food	Unhealthy housing	Global environmental change
Acute respiratory infections	•				•	
Diarrheal diseases		•	•	•		•
Other infections		•	•	•	•	
Malaria and other vector-borne diseases		•	•			
Injuries and poisonings	•		•	•	•	•
Mental health conditions					•	
Cardiovascular diseases	•				•	•
Cancer	•			•		•
Chronic respiratory diseases	•					•

Source: WHO, Health and Environment in Sustainable Development: five years after the Earth Summit, 1997.

The second strategy tackles the value systems that humans employ in their use of natural resources. And these are much harder to identify and to change. Value systems are seen in the way in which humans identify their relationship with the rest of life on the planet. They are also seen in the national and global social, economic, cultural, and educational systems. Because these notions are abstract to the individual, and because changes in attitudes do not afford tangible beneficial or even measurable results, there is less incentive for humans to undertake to change them.

Government policies and programs vary according to the prevailing value system. For example, if equitable sharing of available wealth is part of this value system, policies will be designed to promote social equity. Similarly, if health is a priority, resources will be channeled into health protection. Both of these were illustrated by the previously mentioned examples of China, Sri Lanka and Kerala. The concept of sustainable development has introduced an additional component of value systems, namely intergenerational equity, which stipulates meeting the needs of the present without compromising or threatening the ability of future generations to meet their own needs. Obviously, the debate concerning the right balance between the allocation of resources for the protection of the environment and health, and the allocation of resources for lifestyle enhancement still continues, despite warnings that progress is slowed down when populations suffer from ill health, and life-support systems are degraded.

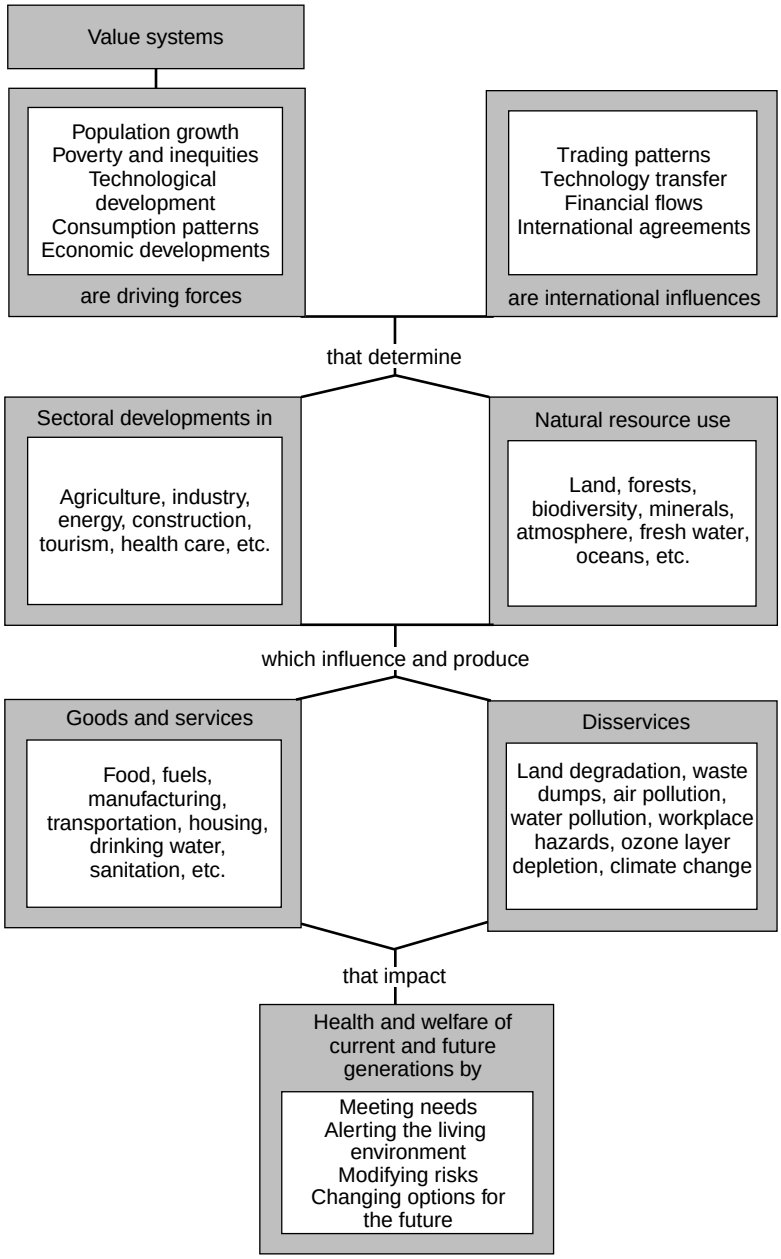
People worldwide must be made aware of what is at stake. They need to be enlightened and empowered. Several steps must be taken to mobilize people's relevant participation. National and international frameworks providing the necessary knowledge as well as the resources, must be made available.

CONTINUING CHALLENGES TO THE ENVIRONMENT AND PUBLIC HEALTH

Population Dynamics

Any discussion on health, the environment and development has to include a discussion on population – its size and its effects on consumption levels. This is because the latter largely determine the impact that the human population will have on the environment. It must also include an analysis of distribution of income and assets in order to allow appreciation of the difficulties faced by those people whose health is compromised by a lack of them. Since macroeconomic policies affect all sectoral policies, it is important to know the macroeconomic framework used by governments to set their economic, social, and environmental policies. This is important in order to assess whether appropriate changes are made to promote health and sustainability.

The world's population grew more than five-fold between 1800 and 1990 to reach 6000 million people in 1999 (UN, 1994). Up to 80% of the global population are in the developing countries. Projections suggest that world population will reach 8000 million by 2020 (UNFPA, 1997). If the world population continues to grow at this



Sustainable development can be described in terms of the elements in this figure, interlinking in a logical manner: the driving forces, intermediate elements and the eventual impact or effect on the health and welfare of current and future generations.

Source: Mr. Nitin Desai, Under-Secretary-General for Policy Coordination and Sustainable Development, UN in "Health and Environment in Sustainable Development: Five years after the Earth Summit", WHO, Geneva, 1997.

Figure 5 Elements of sustainable development

rate, resources such as water will be severely depleted by then. This will bring a host of new health threats to add to those existing threats that will be severely exacerbated.

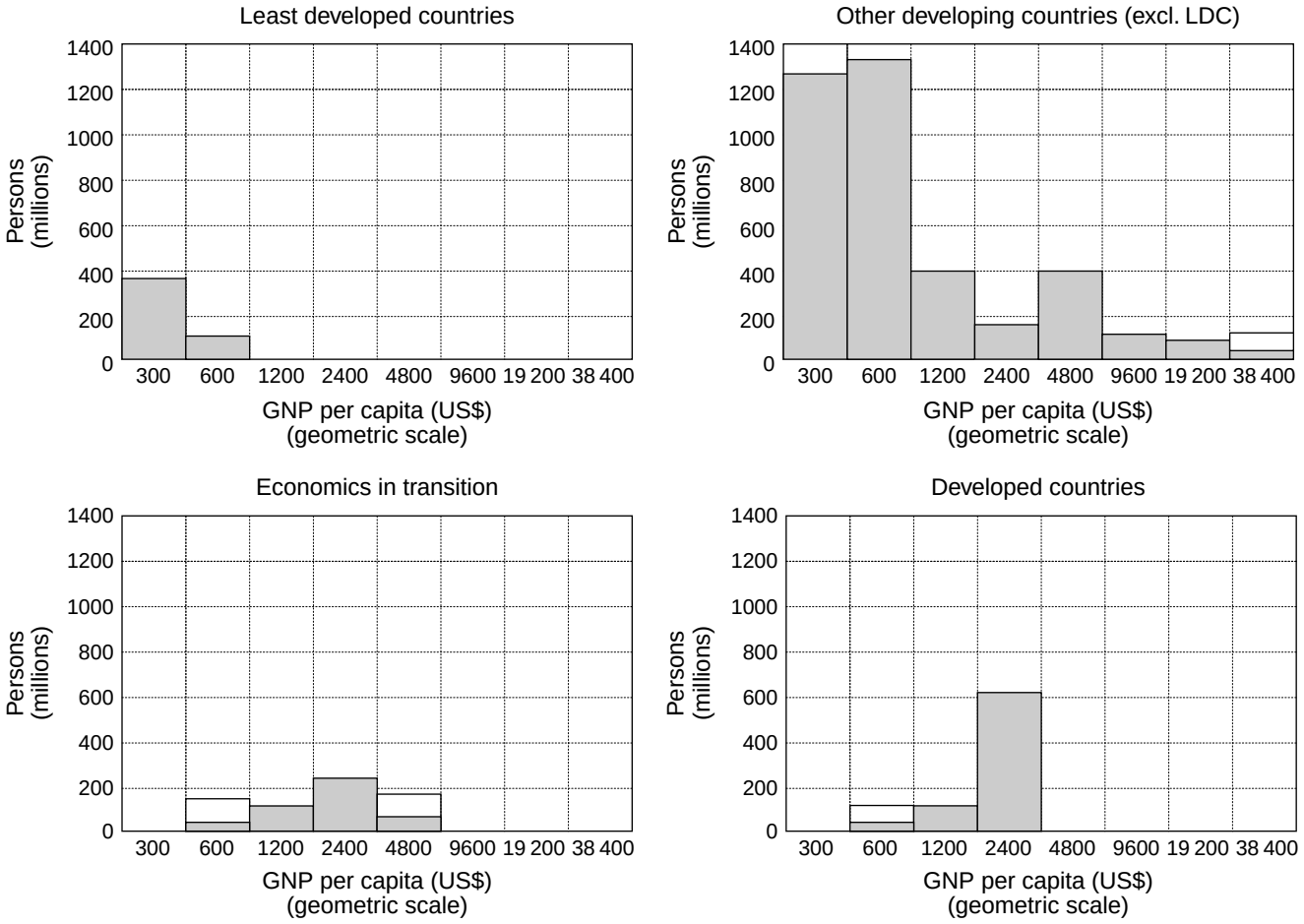
In developing countries, where population growth is still high, pressures on scarce resources make it difficult to improve living conditions. Existing infrastructures are rundown, as city authorities are unable to provide the

extra services required by the rapid population growth. In developed countries, population levels are largely static, enabling a greater level of prosperity and leading to increased consumption and even over-consumption with little regard to their possible planetary impact: the depletion of non-renewable resources, the degradation of soil and water resources, and the emission of gases that threaten the ozone layer and climatic stability. The question of sustainability is undoubtedly being raised as many more countries are aspiring to comparable levels of consumption (see Figure 6).

The United Nation Development Programme (UNDP) 1998 Human Development Report brings out the stark facts on consumption patterns and illustrates the consequences of both over-consumption in the developed world and under-consumption and human deprivation, the lot of the poor in the developing world. The report brings to light the ever-expanding consumption patterns that put strains on the environment. It identifies emissions and wastes that pollute the earth and destroy ecosystems, and highlights growing depletion and degradation of renewable resources that undermine livelihoods. It points out that in the final analysis, it is poor people and poor countries who are unable to protect themselves, and bear many of the costs of unequal consumption. According to the report, a single child born in the industrial world contributes more to consumption and pollution over his/her lifetime than do between 30 and 50 children born in the developing countries (UNDP, 1998).

The report goes on to recommend that poor countries need to accelerate their consumption growth, without following the path taken by the rich and high-growth economies over the past half century. It points out that the developing countries today can leapfrog to patterns that are pro-environment and pro-poor, by ensuring also that consumption of over one billion poor people be raised (UNDP, 1998).

In addition to consumption patterns, there are environmental implications of high population growth, which are far-reaching for many low-income countries.



Source: Based on data from the World Bank in "Health and Environment in Sustainable Development: Five years after the Earth Summit", WHO, 1997.

Figure 6 Share of GNP per capita (US$) by population size and level of development, 1993

Indeed, a strong relationship exists between population growth and poverty. The source of much suffering, ill health and death, poverty can act as an additional pressure on the environment. Poor people having few or no resources, have littlechoice but to exploit rather than protect the environment. Given that experience shows that fertility rates are closely associated with infant mortality rates and child health in general, a sustained global decrease in fertility will depend on general health improvements, including the amelioration of living standards. In fact, steps that reduce infant and child mortality also tend to promote reduction in fertility rates, namely: a secure livelihood, better health and education of mothers and girls, improved water supplies, sanitation and nutrition, together with effective reproductive services integrated into health care systems available to all. It is important to sustain reductions in fertility by implementing programs that remove the economic necessity for poorer groups to have large families, since in many poor countries, children are the principal source of income and provide the only social security available in old age. Programs such as the ones developed by WHO in seven African countries, whereby investing in poor women in raising income levels, coupled with functional literacy and community-based health services, contributed in improving the conditions of these women and their children. This approach provided women with a viable alternative for gainful living other than having more children (WHO, 1992).

Population Structure and Changing Health Concerns

Policies on population dynamics are intended to minimize the pressure on natural resources, primarily by minimizing the number of people who need these resources – although we have seen that 20% of the world's people in the highest income countries account for 86% of total private consumption expenditure, the poorest 20% a minuscule 1.3% (WHO, 1992). The carrying capacity of the planet is calculable, and the growth rate of the world population exceeds the rate at which resources to support human life can be replenished. In the 20th century, much attention was given to reducing fertility rates, particularly in developing countries, where the infrastructure to support citizenry is either weak or non-existent. Much less has been done on investing in human development, in reducing the gap between the rich and the poor, in expanding the knowledge base for all people, and in applying technologies that are environmentally sustainable for poor and af uent consumers.

The manner in which population growth is treated is instructive about the prevalent attitudes towards public health. In developing countries, much attention is given to the distribution of prophylactics – condoms and birth control pills that are mainstays of the pharmaceutical industry. The argument is that the use of condoms prevents the spread of sexually transmitted diseases, limiting the need for clinical intervention, reducing the risk of pregnancy, and preventing the spread of AIDS. Birth control pills, by enabling mothers to either space pregnancies or avoid unwanted births, put power in the hands of women, and limit both child and maternal morbidity and mortality. Programs supported by international donors and lending institutions support the distribution and educational campaigns associated with condom and pill use. But the use of these programs continues to be limited. Several factors contribute to this. First, the high-technology nature of the industry that produces prophylactics means that most developing countries either have to import them, or create industries for their local production. Second, education is a key component of the success of contraceptive use, which in many parts of the world where illiteracy is high means additional investment in the programs.

Poverty on the Increase

There is no doubt that people lacking economic and other resources, those denied opportunities to improve their condition, and those suffering from natural or man-made disasters and/or warfare are affected by the most severe environmental problems. Inequity is a major factor affecting the health-and-environment link. As long as large-scale inequity remains, the environment and health status will be continuously compromised. About 2900 million people live in countries where the *per capita* gross national product (GNP) is less than US$600. In contrast, about 800 million people live in countries where *per capita* GNP is greater than US$9600, the highest being Luxembourg where *per capita* GNP is US$37 320 (UN Statistical Yearbook, 1995).

Poverty as re ected by global inequity is actually increasing. In 1970, the richest 20% of humanity had roughly 30 times more income than the poorest 20%. Today, the richest one-fifth receive more than 60 times as much as the poorest one-fifth does. The wealth of the richest people in the world is staggering. The net worth of the world's 358 richest people is larger than the combined annual income of the poorest 45% of the world's population, that is 2300 million people (UNDP, 1996).

Between one-fifth and one-quarter of the world's population live in absolute poverty. More than 90% of this group, live in developing countries. A combination of increasing numbers of people, land given over to large-scale agro-business, and inequality in land tenure is not only leading to population movement, but also forcing those with least opportunities to settle in environmentally poor areas with the lowest potential for development.

Poverty and environmental degradation are inextricably linked. The world's poorest people are found predominantly in two types of areas: in remote and ecologically fragile

rural areas, and at the margins of expanding urban areas (Leonard, 1989). The precarious conditions of both expose poor populations to health risks that range from inadequate water and fuel supplies, to water-borne diseases, ambient and indoor air pollution, and to natural and technological disasters. The least developed countries have been affected by diminishing overseas development assistance and slow economic growth. Those living in absolute poverty include a high proportion of women, children, refugees and other displaced persons who lack access to education, employment, health care and other basic resources. These groups often interlink and overlap because the fundamental cause of their condition is the same – lack of power, low status, and a lack of opportunities.

Women, Poverty and Diminishing Resources

It has been stated that women provide the best illustration of the combined impacts of poverty, unemployment and social disintegration, on health and quality of life. Because women account for more than 70% of the agricultural workforce in many parts of Africa, and in others they are one-quarter to one-third of the sole breadwinners of the world's households, their poor health status has serious repercussions for economic development.

In spite of the overall improvement of women's health as shown by their increased life expectancy, many of them suffer loneliness, alienation, disability and poverty during old age (WHO, 1997c).

Although there has been a positive relationship between women's literacy, fertility, as well as prevention of disease and health conditions that affect them, women are still denied access to education. Illiteracy prevents them therefore, from understanding how their body functions, how they can prevent disease and protect themselves, and how they can avoid engaging in harmful practices. Insufficient access to existing methods of disease prevention has contributed to the problems they face on sexually transmitted diseases. They have had to contend with the use of technologies that were devised for men when attending to agricultural chores. And they have had to survive with very inadequate nutritional intake whose cumulative effects have permeated the different phases of their lives, sometimes transmitting the effects of deficiencies to their offspring (WHO, 1997c).

In a highly competitive climate where there is emphasis on the production of cheap tradable goods, women have been the ones recruited to work with the lowest possible wages in working conditions that are full of health risks and expanding long hours, at the expense of their health. When there are layoffs to be made, it is the women who are the first to lose their jobs. The high incidence of acute and chronic health problems is testimony to their poor working conditions.

Almost everywhere, women are victims of violence ranging from sexual abuse to social and economic disadvantage. Gender violence is a universal plague, grossly under-reported. In some countries, domestic violence is the leading cause of injury among women of childbearing age; up to 35% of women's visits to emergency treatment centers are on account of violence. Studies show that the prevalence of violence against pregnant women ranges from 7–20%, and in some instances, even more, and that such violence is more common than many other conditions routinely screened for during pregnancy (World Health Report, 1997). Violence against women is an important cause of morbidity and mortality throughout their life span, and more recently, elderly women have been particularly the target of aggression.

Finally, it is women who must cope with the stress and pressures as mothers, wives, daughters, sisters and caregivers, sometimes during famines and wars, often bearing the extra burden of coping with the unemployment of a husband or a child, the drug abuse of a brother, or the rape of a daughter (World Health Report, 1997).

Violence Reaching Epidemic Proportions

As we enter the 21st century, we carry with us the signs and symptoms of a deep social crisis facing almost all societies. This crisis is characterized by persistent high unemployment, poverty, and social disintegration. The most glaring results of social disintegration related to poverty and unemployment is demonstrated by the worldwide increase of violence. Violence is a major public health issue. It is the most complex problem that the public health field has ever addressed, and will be one of the most threatening factors in uencing sustainable development. Since there is no single cause, there is no single solution.

While there is no single direct relationship between poverty and violence, a close association does exist between inequity and violence. It is not that the poor are intrinsically more violent than other members of society, but rather that the inequities they suffer combined with disempowerment, fear, insecurity, oppression, hopelessness, and the depression these cause, are contributing factors to violent behavior.

Violence may erupt as a result of structural factors such as patterns of economic, social and political domination; lack of exemplary leadership that may stir the anguish of those who see no future reward for their efforts, or those for whom the future seems bleaker than the present, or to any situation marked by cruelty, discrimination, rejection, or forcible and arbitrary domination. Relatively easy access to ever-more sophisticated weapons of destruction, has allowed the level of violence in societies to escalate. New structures of violence, borne out of the overwhelming temptation of quick and easy access to huge fortunes,

have emerged such as transnational criminal organizations engaged in illegal trade in narcotics, arms and other goods. The frequent association of violence with drugs and alcohol use has been recognized. This has often led to the corruption of social values, and even of institutions such as the judicial systems. A violent culture transmits the values that sustain it – in the family, at the school, in the arts, or in the mass communication media.

What has violence – be it towards children, associated with gender, or ethnic or race – to do with public health policies and environmental consequences?

The cost of violence is incalculable in terms of mortality, disability, destruction of the infrastructure, and damage to the environment, let alone the financial implications of reconstruction and rehabilitation. The injuries inflicted by modern weapons of war, have had devastating effects on the life of the civilian population, and on the destruction of the environment. Water resources have been polluted, crops and harvests have been destroyed, and fertile agricultural land has become barren.

The use of landmines, for example, has not only inflicted irreparable pain and suffering – maiming men, women, and children – but their effects continue to be felt even when the conflict has ceased. The use of chemical weapons represents an alarming threat to the populations that were exposed and to the environment, and will continue to threaten the health of the population exposed long after they have been deployed. The inhalation of toxic substances seriously enhances the risk of damaged immune system and fatal cancer. The rehabilitation of war victims is an additional social and economic requirement, carrying with it tremendous cost implications. Many war victims may never recover, and others may be permanently disabled.

In some countries, health gains that were obtained with great effort and commitment in the last decade, were wiped out as a consequence of war, tribal, or ethnic conflict. Mozambique, Angola and Sierra Leone are dramatic examples where children paid the price of extraordinary violence. The latest report by United Nations Children's Fund (UNICEF) on the child mortality survey in Iraq in 1999, provides alarming figures. Infant mortality rates have increased from 47 deaths per 1000 live births, for the period 1984–1989, to 108 per 1000 live births for the period 1994–1999. The under-5 mortality rate has increased over the same period from 56 deaths per 1000 live births to 131 deaths per 1000 live births (UNICEF, 1999).

Other consequences include destruction of the health infrastructure, pollution of water, leading to epidemics difficult to contain in situations of conflict. Violence draws off from the already limited resources required for the prevention and treatment of diseases. But the financial costs are but one aspect; the added costs includes those associated with the insecurity and fear arising from violence. Demands for more and better law enforcement police and military forces divert resources from a socially oriented strategy.

Finally, the total cost of all forms of violence – be it at home, in schools, in the community, or emanating from local conflicts that inflict major human rights abuses, uprooting many people, and fueling the refugee crisis – lead to mental and physical traumas that may never be measurable. The health consequences are likely to be long term and passed on to later generations.

A Glimpse of Hope

It is encouraging that some countries in Europe are increasingly denouncing lack of commitment and action by the rich countries, in helping poor countries overcome unbearable poverty. At a meeting held in Utstein Abbey in Western Norway in 1999, ministers for international development discussed how to make more progress in the fight against poverty. Although they concluded that they did not possess the power to reverse the current poverty trends, they shared the belief that they had a responsibility for building stronger coalitions for change. They focused particularly on the need to improve coordination among donors and to support the fight against corruption, and to deliver on the promise of deeper debt relief, strengthening the ability and resolve of the international community to respond to conflict and crises. They agreed on an eleven-point agenda that will be subsequently reviewed for progress.

The Dutch Minister for Development Cooperation, the Norwegian Minister for Development and Human Rights, the British Secretary of State for International Development, and the German Minister for Development Cooperation went on to state that the poorest countries were weighed down with too many donors, and that scarce government capacity was tied up in meetings with the donors and preparing reports to justify the use of aid funds. They stated that donors too often determined priorities and tied aid to the purchasing of goods and services from their countries. They proposed that the donor community should not undermine the efforts of poor countries or try to run their own programs. The donor community must be willing to let go and see their partners in real command of development strategies.

Referring to poverty, the ministers commented on two critical factors: first was the collective ability of bilateral and multilateral debtor countries to secure sufficient financing based on transparent and fair burden sharing; second was the ability to transform debt relief into real poverty reduction and sustainable development. Strengthening the social dimension of the International Monetary Fund's (IMF's) structural adjustment facility for the poorest countries was considered a key contribution. Many more issues were dealt with and concrete proposals made, which were not only creative but were for the first time, sincere in

illustrating the partnerships that have to be forged between countries in need of assistance and the donor community (Herfkens *et al.*, 1999).

Lifestyles, Health and Environmental Implications

Af uence by most definitions means increased *per capita* consumption of food, goods and services. This increased consumption and production depletes natural resources and produces wastes, some of which can have severe effects on the environment and human health. Although population growth contributes to these developments, more important are the lifestyle patterns that drive people to consume. This lifestyle is predominantly found in af uent countries but is rapidly being adopted in all countries. In terms of sustainability, three of the most important consumption issues are diet (particularly the consumption of meat), consumption of certain raw materials and persistent chemicals, and the consumption of fossil fuels and associated carbon dioxide emissions. The consumption of grain, beef and water has tripled, while paper use has risen six times over the past 45 years (UNDP, 1998).

Richer populations not only consume greater variety and amounts of food, increasing the environmental effects of transport and storage of food, but also greater amounts of specific foods, which likewise have environmental implications. Increased demand for meat has many consequences. For example, it intensifies marine fishing, pushing some fisheries towards extinction and shifting the ecological balance of the oceans; it increases the demand for pasture, thus increasing the rate of deforestation and associated loss of biodiversity in some areas; and it diverts more of the world's basic food supply to animals, thereby making agriculture and fishing less efficient in meeting human nutritional needs (FAO, 1996).

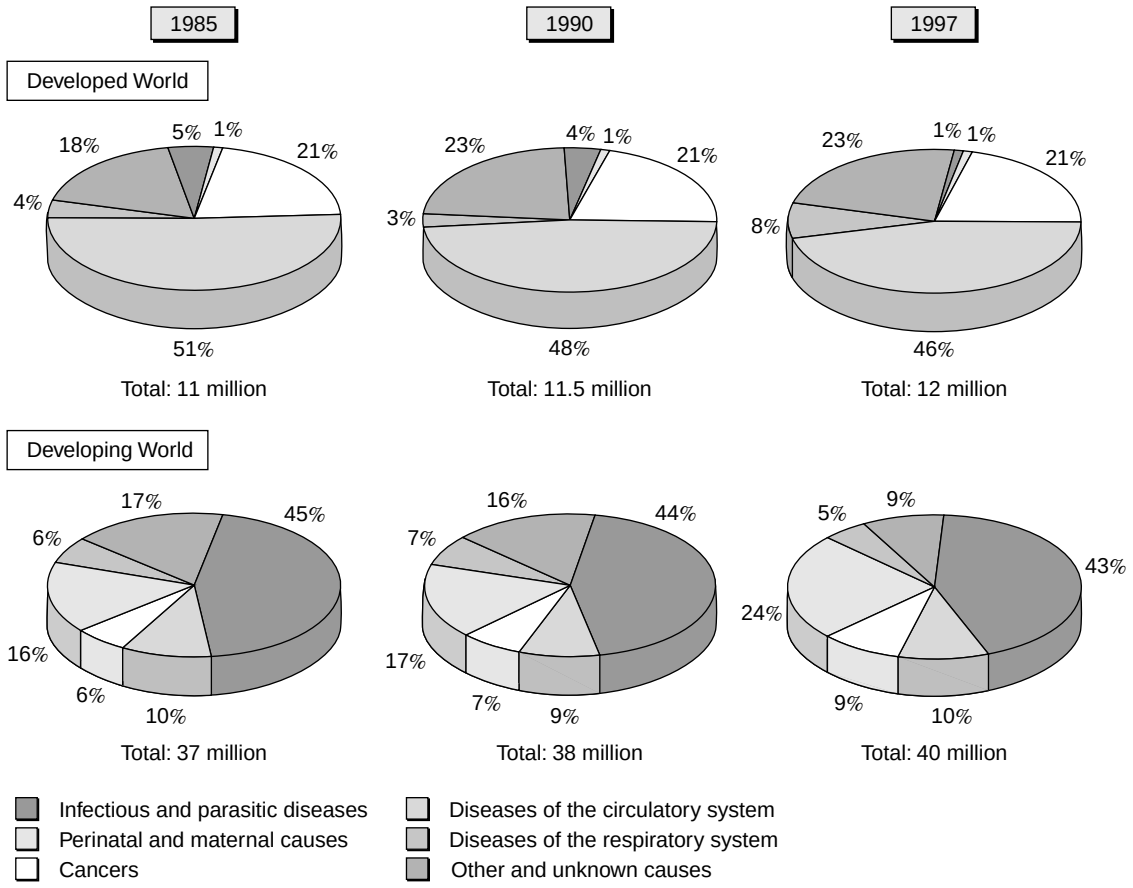
Another lifestyle issue that has grown in importance especially among youth groups in the world, is engaging in health-damaging behavior such as drug and alcohol consumption, smoking, encouraging the cultivation of coca and other crops for purposes other than their therapeutic reasons, and their illicit marketing worldwide. The consumption of the world's raw materials, especially toxic metals such as lead, cadmium (Cd) and mercury (Hg) have great environmental health risks. Obviously depending on the degree to which human populations are exposed, the production figures are such and their emissions, especially of lead (Pb), are the subject of great concern.

Many examples can be cited on the harmful effects of the preoccupation with making a quick profit at the expense of health and sustainability of the environment, characterizing the lifestyles of modern societies. It is sufficient to look at the epidemiological transition and the estimated trends in mortality for major causes of death, to appreciate the effect on human health. In the industrialized world, control of communicable diseases as a major cause of death has been replaced by non-communicable diseases such as cardiovascular, cancers, injuries and accidents and rising incidence of mental health problems. In the developing countries, the major cause of disease still remain infectious diseases with an increasing rise in non-communicable diseases, illustrating the double burden of disease facing the third world. An alarming phenomenon coming into the picture is the quality of inequity that affects health status differentials between as well as within countries. As we arrive into the 21st century, mental health conditions, cardiovascular diseases, cancers, allergies, chronic respiratory diseases, occupational hazards, and finally violence, are becoming a major scourge. All these constitute significant public health problems, whose root causes emanate from the socioeconomic, political and environmental context in which people live. There is no doubt that the more that action is undertaken as an integral part of the environment, the more the benefits will accrue to health (see Figure 7).

Clearly, sustainable development policies should emphasize longer-term broad-spectrum interventions touching upon the driving forces operating in human societies. In many developing countries, this would mean tackling inequity, poverty, and population issues, thereby contributing to the control of land degradation and deforestation, biodiversity loss, soil erosion, and food insecurity. In developed countries, inequity also needs tackling as evidenced by the fact that sizable population groups live in squalor and relative poverty. Emphasis should be placed on reducing unsustainable consumption, curbing the use of non-renewable fuels, and reducing the generation of solid wastes to minimize transboundary pollution, toxic waste problems, and global environmental change. To implement such proactive preventive approaches successfully, development policies and planning need a long-term horizon with health and environment concerns forming an integral part of the associated planning. A supportive environment for health is free from major health hazards, satisfies the basic needs of healthy living, and facilitates equitable social interaction. Supportive environments, on the other hand, encompass where people live, their local community, their home, where they work, where they learn, and where they play. By looking at the physical, the social, the spiritual, the economic, the political, and above all, the involvement of the people themselves and all civil society, there is hope that protecting our health and our environment can become a legacy to the future generations (WHO, 1997).

CONCLUDING REMARKS

Today, in pursuit of development, societies are constantly confronted with tradeoffs; between health and economic development, between the increased generation of income and the protection of the people, and between sound



Source: The World Health Report 1998.

Figure 7 Causes of death: distribution of deaths by main causes, by level of development, 1985, 1990, 1997

environmental practices and making business profitable. The global value system has been strongly biased in favor of economic gain, quick profit, or cutting down on essential expenditures at the cost of other invaluable attributes such as health and a sound environment. The global value system advocating caring for societies in a caring world is fading (WHO, 1997).

The dominant issue facing policy makers is the respective roles of the market and the State. Taking health as an example, policy makers must be made aware of the inherent limitations and imperfections of a health care system that is controlled by market forces. Although imperfections are inherent in any market, the value of health is such that the potentially grave and dramatic human consequences that could occur in an inappropriately managed health market cannot be allowed (WHO, 1997).

Improved public health is an indispensable pre-condition for the enjoyment of other goods and services. The goods and services needed to protect health and cure disease are essential to individuals in a way in which other goods are not. Consequently, health and the goods and services needed to achieve it are not tradable in the same way as other commodities. This attribute makes health unique and not a commodity like others that can be bought and sold (WHO, 1997).

Technological advances must be anchored firmly to a foundation of human values and to individual rights to life and health. If not, they will imperil the basic social institutions, which order society. Abuses are likely to occur in any system, whether private or public. This anchoring will not be done by the market nor will it be done by the state acting alone. It must be achieved through societal choices made collectively by the state, the market and civil society (WHO, 1997).

It is becoming increasingly clear that the environmental factors that most affect health are linked to underlying pressures on the environment. These pressures are mainly due to factors such as population growth, inequitable resource distribution, consumption patterns, technological development and components of economic development. Since these factors and the health hazards they create emanate from activities of various sectors, effective action to protect health will require coordination and collaboration between various sectors.

There will always be a part of the public domain in health which must remain a public responsibility; e.g., the protection of people from communicable diseases, the control and elimination of disease vectors, and the control and regulations of known health risks. This is because the benefits of these activities accrue to people in an indivisible form. They create, by their very nature, an environment where benefits are collectively enjoyed, and are not available to their delivery through market mechanisms to individual beneficiaries (WHO, 1997).

The health sector must assume its share of responsibility by being the guide and a partner in these actions, so that health concerns are represented appropriately at all stages of implementation. It has an essential advocacy role to play in highlighting the links between health, the environment and sustainable development, when future policies are developed and actions planned. A much stronger partnership between the health sector and other sectors, and an alliance with the environment activities is vital for the successful reduction of health threats arising from poor environmental conditions.

See also: Environmental Change and Human Health: Extending the Sustainability Agenda, Volume 3; Theories of Health and Environment, Volume 5.

REFERENCES

FAO (1996) *1996 World Food Summit: Technical Background Document*, 3 Volumes. FAO, Rome.

Herfkens, E, Johnson, H F, Short, C, and Wieczorek-Zeul, H (1999) If We Are Serious, We Do Something About Poverty, *International Herald Tribune*, August 10, 6.

International Herald Tribune (1999) *EU Queries France Over Animal Feed*, August 14–15, 6.

Leonard, H J (1989) *Environment and the Poor: Development Strategy for a Common Agenda*, Transaction Books, New Brunswick.

Rio Declaration (1992) www.un.org/riodeclaration.

UN (1995) *UN Statistical Yearbook 1995*, UN, New York.

UNDP (1996) *Human Development Report 1996*, Oxford University Press, New York.

UNDP (1998) *Human Development Report 1998*, Oxford University Press, New York.

UNFPA (1997) *Task Managers' Report for 1997 Special Session of the General Assembly. Demographic Dynamics and Sustainability*, UNFPA, New York.

UNICEF (1999) *Child and Maternal Mortality Survey 1999: Preliminary Report. Iraq*, (Joint publication by UNICEF and Iraqi Ministry of Health).

United Nations Department for Economic and Social Information and Policy Analysis (1994) *World Population Prospects: The 1994 Revision*, (Population Division, Document ST/ESA/SER. A/145), UN, New York.

WHO (1952) *Expert Committee on Public Health Administration*, WHO, Geneva.

WHO (1986) *Intersectoral Action for Health*, WHO, Geneva.

WHO (1992) *Health Dimensions of Economic Reform*, WHO, Geneva.

WHO (1994) *Health in Development: Prospects for the 21st Century*, Report of the First Meeting of the Task Force on Health in Development, 27–30 June 1994, (Document WHO/DGH/94.5.), WHO, Geneva.

WHO (1997a) *The World Health Report 1997: Conquering Suffering, Enriching Humanity*, WHO, Geneva.

WHO (1997b) *Health and Environment in Sustainable Development: Five Years After the Earth Summit*, WHO, Geneva.

WHO (1997c) *Health: The Courage to Care*, A Critical Analysis of WHO's Leadership Role in International Health by the Task Force on Health in Development, eds G Gunatilleke and A El Bindari Hammad, WHO, Geneva.

FURTHER READING

Annest, J L, Pirkle, J L, Makuc, D, Neese, J W, Bayse, D D, and Kovar, M G (1983) Chronological Trend in Blood Lead Levels Between 1976 and 1980, *N. Engl. J. Med.*, 308, 1373–1377.

Cahill, K M, ed (1995) *Clearing the Fields: Solutions to the Global Landmine Crisis*, (a joint publication of Basic Books and the Council on Foreign Relations), Basic Books, New York.

CSD (1997) *Overall Progress Achieved Since the United Nations Conference on Environment and Development: Protecting and Promoting Human Health. Report of the Secretary General*, Commission on Sustainable Development (E/CN.17/ 1997/2/ Add.5), New York.

CSD (1997) *Overall Progress Achieved Since the United Nations Conference on Environment and Development: Combating Poverty. Report of the Secretary General*, Commission on Sustainable Development (E/CN.17/1997/2/Add.2), New York.

CSD (1997) *Overall Progress Achieved Since the United Nations Conference on Environment and Development: Environmentally Sound Management of Solid Waste and Sewerage-Related Issues. Report of the Secretary General*, Commission on Sustainable Development (E/CN.17/1997/2/Add.20), New York.

CSD (1997) *Comprehensive Assessment of Freshwater Resources of the World: Report of the Secretary General*, Commission on Sustainable Development (E/CN.17/1997/9), New York.

CSD (1997) *Overall Progress Achieved Since the United Nations Conference on Environment and Development: Changing Consumption and Production Patterns. Report of the Secretary General*, Commission on Sustainable Development; (E/CN.17/ 1997/2/Add.3), New York.

CSD (1997) *Overall Progress Achieved Since the United Nations Conference on Environment and Development: Combating Poverty. Report of the Secretary General*, Commission on Sustainable Development (E/CN.17/1997/2/Add.2), New York.

De Macedo, C G (1992) *The Crisis of Public Health: Re ections for the Debate*. (Scientific Publication No. 540), Pan American Health Organization, Washington, DC.

FAO (1994) Water Policies and Agriculture, *The State of Food and Agriculture, 1993*, FAO, Rome, Italy.

FAO (1994) *Compendium of Food Consumption* (Vols. 1–2), FAO, Rome, Italy.

FAO (1995) *FAO Production Yearbook 1995* (Vol. 49), FAO, Rome, Italy.

Gratz, N G and Knudsen, A B (1996) *The Rise and Spread of Dengue, Dengue Hemorrhagic Fever and its Vectors: a Historical View (up to 1995)*. (unpublished document CTD/FIL (DEN) 96.7) WHO, Geneva, Switzerland.

HMSO (1988) *Public Health in England: The Report of the Committee of Inquiry into the Future Development of the Public Health Function*, London.

Kickbusch, I (1989) Approaches to an Ecological Base for Public Health, *Health Promotion*, 4(4), 265–268.

Kickbusch, I (1989) The New Public Health Orientation for The City, *The New Public Health in an Urban Context.*, (WHO Healthy Cities Papers, No. 4). WHO, Copenhagen.

Litsios, S (1996) *The Tomorrow of Malaria*, Pacific Press, Wellington.

McKeown, T (1976) *The Role of Medicine: Dream, Mirage or Nemesis*, Nuffield Provincial Hospitals Trust, London.

OMS (1992) *Pour une Nouvelle Action de Santé Publique*, (WHO/HRH/92.1), OMS, Geneva.

Rockefeller Foundation (1985) *Good Health at Low Cost*, Proceedings of a Conference held in Bellagio, Italy, April 29–May 3, 1985, eds S B Halstead, J A Walsh, and K S Warren, The Rockefeller Foundation, New York.

Sachs, J (1999) Helping the World's Poorest, *The Economist*, August 14th–20th, 17–20.

UN (1995) *Report of the International Conference on Population and Development, Cairo, 5- 13 September 1994* (document A/conf. 171/13/rev. 1) UN, New York.

UN (1996) *The Beijing Declaration and the Platform for Action. Fourth World Conference on Women, Beijing, China, 4- 15 September* 1996, UN Department of Public Information, New York.

UNCHS (1996) *Report of the United Nations Conference on Human Settlements (HABITAT II), Istanbul, Turkey, 3- 14 June 1996*, (A/CONF. 165/14) UN, New York.

UNCHS (1996) *An Urbanizing World: Global Report on Human Settlements, 1996*, Oxford University Press, Oxford.

UNCTAD (1995) *The Least Developed Countries: 1995 Report*, United Nations Conference on Trade and Development, UN, New York.

UNDP (1994) *Human Development Report 1994*, Oxford University Press, New York.

UNDP (1995) *Human Development Report 1995*, Oxford University Press, New York.

UNEP (1988) *Assessment of Chemical Contaminants in Food: Report on the Results of the UNEP/FAO/WHO Program on Health-related Environmental Monitoring*, UNEP, Nairobi.

UNEP (1991) *Freshwater Pollution* (GEMS Environmental Library No. 6), UNEP, Nairobi.

WHO (1952) *Expert Committee on Public Health Administration*, First Report: WHO Technical Report Series No. 55, WHO, Geneva.

WHO (1992) Commission on Health and Environment, *Our Planet, Our Health*, WHO, Geneva.

WHO (1992) Commission on Health and Environment, *Report on the Panel on Urbanization*, WHO, Geneva.

WHO (1992) Commission on Health and Environment, *Report on the Panel on Energy*, WHO, Geneva.

WHO (1992) Commission on Health and Environment, *Report on the Panel on Industry*, WHO, Geneva.

WHO (1992) Commission on Health and Environment. *Report on the Panel on Food and Agriculture*, WHO, Geneva.

WHO (1996) *New Challenges for Public Health*. Report of an Interregional Meeting: Geneva, 27–30 November 1995. WHO, Geneva.

WHO (1997) *Health and Environment in Sustainable Development*, WHO, Geneva.

WHO (1998) The World Health Report 1998. *Life in the 21st Century: a Vision for All*, WHO, Geneva.

WHO (1999) *The World Health Report 1999: Making a Difference*, WHO, Geneva.

WHO (1999) *Promoting Health Through Women's Functional Literacy And Intersectoral Action: Lessons from the Nigerian Project*. (Document WHO/HPD/98.2.) WHO, Geneva.

Wiman, R and Partonen, T, eds (1999) *Putting People at the Center of Sustainable Development*, STAKES, Helsinki.

Women's Health Collection (1998) *Women, Vulnerability and HIV/AIDS: A Human Rights Perspective*, eds A Gómez and D Meacham, Latin American and Caribbean Women's Health Network, Chile.

Women's Health Journal (1999) Latin American and Caribbean Women's Health Network, January–March 1, 99.

World Bank (1994) *Making Development Sustainable*, The World Bank Group and the Environment, The World Bank, Washington, DC.

www.who.org.

www.bbc.co.uk.

Industrial Responses to Stratospheric Ozone Depletion and Lessons for Global Climate Change

STEPHEN O ANDERSEN

US Environmental Protection Agency, Washington, DC, USA

When the Montreal Protocol on Stratospheric Ozone Protection was signed in 1987, few alternatives to ozone-depleting substances (ODSs) were available or even imagined. Industry predicted that technical progress would be slow, that costs would be exorbitant, that valuable equipment would be abandoned, and that safety might be compromised. As it turned out, these predictions were never realized. In fact, new technologies are more energy ef cient, maintenance cost is reduced, and systems are more reliable than those based on chloro uorocarbons (CFCs) and other ODSs.

The rst decade of technical progress under the Montreal Protocol occurred during a period of growing concern for the global environment and increasing corporate environmental leadership. During this same period, other international environmental treaties were developed including the Basel Convention on the Control of Transboundary Movements of Hazardous Wastes and their Disposal, the Convention on Biological Diversity, the Convention to Combat Deserti cation, and the Framework Convention on Climate Change and its Kyoto Protocol. The 1992 United Nations Conference on Environment and Development (the Rio Earth Summit) intensi ed global awareness of the importance of environmental protection to economic development and began integrating governmental and non-governmental activities.

This contribution documents and illustrates some of the extraordinary individual, association, and corporate contributions to the remarkable success of stratospheric ozone protection and draws important lessons for climate protection. After the Montreal Protocol was signed in 1987, industry in both developed and developing countries responded to the threat of stratospheric ozone depletion by rolling up their sleeves to get the job done.

THE IMPORTANCE OF INDUSTRY LEADERSHIP TO THE MONTREAL PROTOCOL

The Montreal Protocol is a remarkable success in a variety of ways. Foremost, it is environmentally successful because the controls placed on production of ODSs are working, atmospheric concentrations of chlorine and bromine are coming down and ozone depletion will soon end. It is diplomatically successful because it is proving that the global commons can be wisely managed with stringent obligations, adequate financing, global cooperation, and exible procedures. It is economically successful because it was affordable to business and virtually unnoticed in consumer prices. It is technically successful because engineers used the necessity of change to thoroughly redesign systems, resulting in better efficiency, better reliability, and better overall value for consumers.

One reason for this success is that industry and government learned to cooperate to accomplish common objectives. Looking back, it is possible to trace many unanticipated ways that organizations with little previous experience working together ultimately helped each other to overcome technical challenges.

UNITED NATIONS ENVIRONMENT PROGRAM WELCOMED INDUSTRY LEADERSHIP

Until the Montreal Protocol, it was a tradition of United Nations organizations to only seek advice from government and academic experts who were recommended through national diplomatic channels. United Nations Environment Program (UNEP) Executive Director Mostafa Tolba

Box 1 Technology and economic assessment panel

- 1999: 225 members from 46 countries (Argentina, Australia, Belgium, Brazil, Canada, Chile, China, Colombia, Denmark, Egypt, Finland, France, Germany, Hungary, India, Indonesia, Israel, Italy, Japan, Jordan, Kenya, Korea, Malaysia, Mauritius, Mexico, Morocco, Netherlands, New Zealand, Norway, Pakistan, Philippines, Poland, Russia, Singapore, South Africa, Spain, Sweden, Switzerland, Tanzania, Thailand, Tunisia, Uganda, UK, USA, Venezuela, and Vietnam)
- Since 1989, over 700 members from 50 countries

broke with this tradition when he formed the Technology Review Panel (later renamed the Technology and Economic Assessment Panel (TEAP)) to advise the Parties to the Montreal Protocol on how fast the controls could be strengthened. Dr Tolba allowed the TEAP organizing chairs to recruit technically respected and globally in uential industry experts from outside the diplomatic chain of command, instructed members to act independently from their sponsoring organizations, and empowered TEAP to report directly to the Parties to the Montreal Protocol. Contrast this with the Kyoto Protocol where experts are predominantly from government, universities, and institutes and where reports are reviewed and edited prior to presentation to Parties. The daring action was premised on Dr Tolba's belief that the world's best technical experts would know the most about emerging technology and therefore would be far more technically knowledgeable than government and academic experts with access only to published research. Another anticipated advantage was that these experts would be able to craft detailed strategies to remove barriers to new technology and to create business opportunities where alternatives to ODSs would ourish. The Technology Review Panel (TEAP) was formed in 1989 with 116 experts. Over 80% of its members were from industry and less than 10% were from government (Box 1).

CORPORATE LEADERSHIP PRECEDED NATIONAL AND INTERNATIONAL REGULATORY ACTION

Corporate environmental leadership on ozone layer protection began in 1975 when the S.C. Johnson Company announced plans to phase out the use of CFC as an aerosol product propellant. Publicity from the announcement increased public concern for the ozone layer and encouraged product boycotts. Rapid introduction of hydrocarbon propellants gave consumers an environmental choice and persuaded regulators that the impact of a CFC aerosol ban would be minimal. Within 3 years, sales of aerosol products had plummeted as S.C. Johnson took over market share. The United States banned aerosol products in 1978 followed by Sweden, Canada, and Norway.

The US Environmental Protection Agency (EPA) and the Japan Ministry of International Trade and Industry (MITI) promoted corporate leadership worldwide.

Until the Montreal Protocol was signed in 1987, there was very little cooperation between industry and governments. Industry was mostly represented by associations or special interest organizations chartered to defend CFCs by any means. Companies that were members of these organizations were warned to not express any concern for ozone depletion or any technical optimism that could aid or encourage CFC controls. Within companies, research engineers were discouraged from investigating CFC alternatives: The ozone layer is healthy, CFCs do not destroy ozone, technical optimism could encourage regulation, regulation will harm the company; harm to the company hurts stockholders, employees, and customers.

Industry began to support stratospheric ozone protection after the Montreal Protocol was signed. This corporate environmental leadership resulted from at least eight separate motivations:

1. Public interest organizations including Friends of the Earth, Greenpeace, and the Natural Resources Defense Council were creating public demand for CFC-free products. School children were starting letter writing campaigns and boycotts against products such as foam food packaging.
2. The EPA stratospheric ozone office – managed by Eileen Claussen, John Hoffman, and Stephen Seidel – was young and open to innovation. They reasoned that cooperation with business could speed regulatory results and they gave their staff experts extraordinary authority to experiment with innovative programs. In Japan, MITI and industry associations quickly organized to encourage rapid action to protect the ozone layer.
3. Environmental leadership companies with strong environmental ethics began to organize to halt their own CFC use. These companies included AT&T, Digital Equipment Corporation, Motorola, Nissan, Nortel, and Seiko Epson. These corporate leaders believed that technical cooperation could speed technical development of alternatives and substitutes.
4. Developing country governments and companies were advocates of ozone layer protection. Developed countries have historically been in the vanguard of environmental protection because they were the first to experience the negative environmental effects of industrialization and because they were financially able to respond. Developing country populations are often most affected by environmental despoilment and least able to respond to its destructive effects. The lost productivity and increased public spending for health

Box 2 Extraordinary Mexican leadership

- Aerosol product phase-out before protocol adopted
- Invented open-air HC Filling Technology
- Diplomatic advocate
- First country to sign Montreal Protocol
- First country to ratify Montreal Protocol
- First semi-aqueous electronics facility
- Second no-clean electronics facility
- First ODS Country Program
- First article 5(1) to accelerate phase-out

Box 3 Technology from East European and developing countries

- Brazil Low Solids Flux w/o Solvent
- Brazil Caster Oil for Spray Foam Blowing
- China and India Eco-Frig
- East Germany Hydrocarbon Refrigerator
- India Neem Methyl Bromide Replacement
- Korea New HFC-134a Pathway
- Mexico Open Air HC Aerosol Filling
- Poland Military Halon Bank Management

programs resulting from environmental degradation slows economic development. Environmentally concerned children in uenced their families. Consumers mounted protests against CFC products. Businesses strived to shift from obsolete ODS technology to modern CFC-free technology that was often more profitable. Box 2 illustrates the leadership of developing countries by listing actions taken by Mexico that had a global in uence.

5. The United States Military determined that weapons systems were dependent on CFCs and halons for production, operation, and maintenance and sought to eliminate the supply vulnerability that would come from regulatory controls. The US military leadership inspired other military organizations to cooperate on global environmental protection. Cooperation hastened acceptance of new technology and reduced transition costs. Military leadership inspired suppliers to develop solutions, stimulating work by some of the best inventors in the world. Global military leadership moved the North Atlantic Treaty Organization (NATO) to take the unprecedented step of writing directly to UNEP Executive Director Mostafa Tolba in support of the proposed accelerated phase-out of ODSs. NATO also endorsed technology cooperation and modification of standards to allow production and maintenance of military equipment without the use of chemicals that deplete the ozone layer.
6. Companies dependent on CFCs had over-trusted CFC manufacturers' claims that the ozone layer was healthy and that CFCs were not a culprit. It was a sobering wake-up call for customers when DuPont admitted ozone depletion and supported an international treaty. Companies using CFCs and other ODSs were angry, fearful, and motivated to seek alternatives.
7. Suppliers of alternatives to ODS technology and inventors viewed the Montreal Protocol as a business opportunity. Controls under the Montreal Protocol caused ODS prices to increase and several countries accelerated that price in ation with taxes. For the first time in a decade, ozone protection became a marketing issue. It was significant that CFC suppliers were admitting that ozone depletion was real and that prices would be increasing. The Treaty was a regulatory mandate; governments were working to eliminate barriers to new products and to encourage substitution. Some of the most important technology to replace ODSs was developed and implemented first in developing countries (see Box 3).
8. Environmental awards from EPA, UNEP, and other organizations motivated, inspired and rewarded global citizenship. EPA presented its first annual Stratospheric Protection awards in 1990 and by 2000 had recognized almost 420 individuals, corporations, and organizations from 32 countries (Australia, Belgium, Brazil, Canada, Chile, Dominican Republic, France, Germany, Hungary, India, Indonesia, Ireland, Italy, Japan, Kenya, Malaysia, Malta, Mexico, Netherlands, Norway, Philippines, Poland, Singapore, Spain, Sweden, Switzerland, Taiwan, Thailand, United Kingdom, United States, Uruguay, and Venezuela). UNEP presented 19 awards in 1985, 23 awards in 1987, and one award in 1998 to individuals, corporations and organizations from 19 countries (Australia, Brazil, Canada, Denmark, Egypt, Finland, Germany, Greece, Japan, Malaysia, Mauritius, Mexico, Netherlands, New Zealand, Poland, Sweden, Switzerland, United Kingdom, and the United States). On the 10th Anniversary Celebration of the Montreal Protocol EPA presented Best-of-the-Best Stratospheric Ozone Protection Awards to 71 previous award winners making the most exceptional contributions in the first decade. The winners were from 12 countries: Australia, Brazil, Canada, Germany, Japan, Malaysia, Mexico, Netherlands, Switzerland, United Kingdom, United States, and Venezuela.

CASE STUDY: GLOBAL TECHNOLOGY COOPERATION – MEXICO CFC SOLVENT PHASE-OUT PARTNERSHIP

Material for this case study was provided by Jorge Corona, Head, Foreign Affairs Section, Mexican Environmental Commission of the National Chamber of Industries of Mexico (CANACINTRA). Mr Corona organized the Program for Technology Transfer in the Electronics Industry with US

EPA, UNEP and Nortel, CANACINTRA and the Mexican Electronics Manufacturers Association.

With the support of environmental non-government organizations (NGOs), the Instituto Mexicano del Aerosol A.C. organized the first voluntary sectoral phase-out of CFCs by any developing country. The association and its members achieved the phase-out of CFCs in cosmetic and pesticide products at least 5 years faster than most developed countries and used new technology that was more profitable than the CFC technology it replaced. This favorable experience caused Mexican industry to propose a rapid CFC phase-out in other sectors to take full advantage of superior new technology.

Industry representatives decided to leapfrog CFC solvents as they expanded and modernized factories. The Mexican Solvent Partnership was organized by the Government of Mexico, Camara Nacional de la Industria de la Transformacion, the Canadian telecommunication company Nortel, the International Cooperative for Ozone Layer Protection (ICOLP) (ICOLP member corporations included AT&T, The Boeing Company, British Aerospace, Compaq, Digital Equipment Corporation (DEC), Ford, Hitachi, Honeywell, Hughes Aircraft Company, IBM, Lockheed-Martin, Matsushita Electric, Mitsubishi Electric, Motorola, Nortel/Northern Telecom, Ontario Hydro, Texas Instruments, and Toshiba), and the US EPA. This may have been the first time governments allowed a private company (Nortel) to be an organizing and managing partner in global environmental protection. The Partnership showcased CFC-free technology at Mexican factories of multinational companies and involved Mexican factory managers and solvent cleaning experts through workshops, technology study tours, and training. Nortel assigned a full-time coordinator to identify barriers and to work with the Mexican government and industry authorities to speed technology change.

The Mexican leadership motivated multinational companies to concentrate on modernizing their Mexican production facilities. Within 3 years, AT&T built their first CFC-free factory in Mexico using aqueous cleaning and Nortel built a new factory using no-clean soldering that eliminates the need to clean with CFCs. Mexico was the location of the newest, most environmentally acceptable technology – later duplicated throughout the world.

With the help of multinational companies and associations and the Multilateral Fund, Mexico virtually phased out CFC solvents by 1996.

CASE STUDY: GLOBAL TECHNOLOGY COOPERATION – PHASE-OUT OF CHLOROFLUOROCARBONS IN THAILAND

Yuichi Fujimoto provided material for this case study. Mr Fujimoto was Thailand project director for the Japan Electrical Manufacturers' Association (JEMA).

At a 1990 seminar conducted for the Association of South-East Asian Nations (ASEAN), it was reported that electronics and refrigerator enterprises from Japan, the United States and Europe were primarily responsible for dramatic rates of increase in CFC consumption in Thailand. Responding to this situation, the government of Thailand and multinational companies decided to motivate a rapid phase-out through voluntary pledges.

At the March 1992 Japan–US–Thailand Technical Seminar on Ozone Layer Protection, 35 Japanese and 11 American companies pledged specific schedules for phasing out CFCs and 1,1,1-trichloroethane from Thai operations and joint ventures (AT&T, Dai-Ichi Densikogyo, Digital Equipment Corporation, DuPont, Ford, Fujikura Cable, Fujitsu, Hitachi, Honda, Honeywell, IBM, INOAC, Matsushita, Minebea, Mitsubishi Electric, Motorola, NEC, NHK Spring, Nippondenso, Nissan Motor Company, Nissin Electric, Nitsuko, Nortel, Philips Semiconductor, Signetics, Sanyo, Seiko Epson, Sharp, Sundstrand, Texas Instruments, 3M, Toshiba, Toyota, Yamaha, and ZEXEL).

The original plan was that each multinational company would separately solve its own technical problems. This plan worked well in the electronics manufacturing sector but was unsuccessful for refrigerator manufacturing. Seven Japanese companies and their joint ventures produced household refrigerators in Thailand. Three of these companies imported all the critical parts necessary to complete the phase-out but the other four companies used compressors supplied by a single Thai company. It was therefore crucial that the Thai compressor manufacturer redesign components to achieve high energy efficiency and reliability using refrigerant HFC-134a. The Thai compressors had to be competitive in price and quality with foreign suppliers.

The government of Thailand announced the end of 1996 as its target date for eliminating CFCs used as refrigerants and as blowing agents for insulating foam in domestic refrigerators. This challenging target date was only one year later than the phase-out date for developed countries and was the world's first phase-out for the refrigerator sector in any developing country.

The local compressor manufacturing company was scheduled to supply the new HFC-134a compressors in time for the new refrigerators to be produced by the end of 1996. Unfortunately, prototype compressors tested by the Japanese companies were unacceptable due to manufacturing residue contamination that clogged refrigerant tubes and poor lubrication that caused rapid wear. Despite significant effort to resolve the problems, the local company found itself unable to produce reliable compressors in time for the phase-out.

No other local compressor suppliers could be found with the capability to supply enough compressors to meet the deadline established by the Thai government. The Japanese

companies considered compressors from manufacturers other than the local company but rejected this option because they had pledged to help Thailand build local manufacturing capability. Another option was to postpone the phase-out and wait for a supply of reliable compressors from the local company but this option was rejected because the Thai government was eager to keep the phase-out on schedule.

The only remaining option required additional technology cooperation with the local compressor company. The refrigerator parent companies that were buying compressors in Thailand – Hitachi, Matsushita, Mitsubishi Electric and Toshiba – and the JEMA launched a Tiger Team (Tiger Teams are motivated experts who work as a team to solve problems – confident, decisive, and successful) to assist the local company to improve reliability and to help with its implementation. The US EPA attended follow-up meetings in Thailand and worked with the Thai compressor manufacturers' parent company in the United States to promote its participation in the technical support project.

The ambitious phase-out goals were ultimately achieved through corporate leadership and unprecedented technical cooperation. The seven Japanese refrigerator companies and their joint ventures completed the CFC phase-out by the end of 1996. The Thai Government complemented this technical achievement by prohibiting the manufacture or import of foreign CFC refrigerators after 1996.

CASE STUDY: SUPPLIERS AND CUSTOMERS WORKED TOGETHER TO CHOOSE CHEMICAL SUBSTITUTES

Fluorocarbon manufacturers pooled their resources to quickly and thoroughly test the environmental, health, and safety effects of promising new refrigerants. Individual chemical companies took unusual risks by investing millions of dollars to build facilities to manufacture the new chemicals, even before toxicity and performance studies were completed. In addition, they developed the thermodynamic tables for the new refrigerants in record time, enabling new equipment to be quickly designed and built. Manufacturers of cooling and refrigeration equipment worked closely with the chemical companies to design and build new systems based on the new refrigerants, and to test and optimize their performance.

CASE STUDY: SECTORS THAT RARELY COMMUNICATED JOINED FORCES TO ADVOCATE REFRIGERANT RECYCLING

In 1987, automobile air conditioners were the largest single source of controllable CFC emissions in the US. Normal CFC leakage through rubber hoses, connectors, and seals was very high, and typical repair procedures intentionally released large amounts of CFCs into the atmosphere. CFCs were inexpensive, and no acceptable machines were available to capture and reclaim the CFCs. When confronted with a poorly functioning air conditioner, a service technician typically diagnosed the problem by recharging the system, and then locating the leak by observing the CFC discharging to the atmosphere. The technician would then vent the remaining refrigerant into the atmosphere, repair, recharge, and leak test the system again. Do-it-yourself car owners saved money by merely recharging leaking systems – often recharging several times a year, without fixing the leak. CFCs also entered the atmosphere when an air conditioning component had to be removed to gain access to other automobile systems.

In January 1988, EPA organized its first meeting of the Ad-Hoc Automobile Recycling Working Group. The proposed standard was based on the assumption that some level of CFC contamination would be acceptable, that it was the obligation of recycling equipment manufacturers to design equipment capable of cleaning to a standard acceptable to automobile manufacturers, and that equipment would have to be tested by an independent laboratory to prove its cleaning performance.

By 1989 the Mobile Air Conditioning Society (MACS), the Society of Automobile Engineers (SAE) and EPA developed a CFC recycling standard that was endorsed by international vehicle manufacturers for repair under new-car warranty. Many automobile manufacturers required franchised dealers to employ the new recycling technology. Underwriters Laboratory, also represented on the working group, simultaneously developed the equipment certification program that was available by the time vehicle manufacturers accepted recycling.

Companies that were fiercely competitive in product markets allowed their engineers to jointly develop ozone-safe technology.

In 1989 UNEP formed the TEAP to conduct the first global assessment of alternatives and substitutes to ODSs. A majority of TEAP members were from industry and research institutes, with a minority from governments and universities. TEAP experts were surprised to learn that some technically sophisticated companies had inadvertently avoided typical dependence on ODSs and these experts quickly realized that TEAP could become a catalyst for technical evolution.

The role of the TEAP Solvent Technical Options Committee (TOC) in commercializing no-clean soldering to replace the use of CFC-113 solvents illustrates this cooperation. When the German company SEHO demonstrated controlled atmosphere soldering, electronics manufacturing experts on the solvents, TOC recognized the theoretical potential but were unimpressed by the soldering results. It was obvious to the TOC experts that the component parts were too oxidized, that the ux was mismatched, and

that the soldering wave was poorly formed. However, the promise of the technology was so compelling that experts at AT&T, Ford, Motorola, and Nortel persuaded their companies to experiment with the technology and decided to work as a team.

Some experts concentrated on the ux composition, while others believed that the ux was chemically suitable and worked out ways to apply it to the board more precisely. Still others concentrated on the mixture of gases in the soldering chamber. Several months into development, one company discovered that gas monitoring and control calibrations were critical. Better calibrations dramatically improved the soldering quality. The team consulted with ux suppliers who grasped the opportunity for developing and commercializing new products and intensified development in cooperation with the electronics manufacturers. EPA encouraged the work by documenting and publishing the global environmental advantages that such a technology could provide. TEAP reports reported the progress and alerted potential customers to the possibility of a technical breakthrough.

Meanwhile, engineers at AT&T Bell Laboratories were developing state-of-the-art spray uxing machines to apply precisely the optimal amount of ux to the locations in the printed circuit boards where necessary. Nortel was developing equipment to verify ux concentrations on production boards, and Motorola was experimenting with soldering ultraminiaturized circuits with hybrid components including optical devices and exible connectors. One by one, companies satisfied internal quality controls and moved from laboratory scale to pilot studies and finally to full implementation. During implementation, experts from the intercompany team continued to cooperate to debug operations and optimize performance.

It is impossible to say just when the engineering team realized that their no-clean technology would revolutionize electronics assembly. Engineers who had cautiously reported performance as good as with CFC solvents began to report improved performance. Line managers cautiously increased the speed of soldering to rates never achieved with conventional soldering and found that in some cases defect rates actually *decreased*.

EPA and its contractors tracked and documented the technical development and operating parameters and, in cooperation with the ICOLP published the first no-clean handbook in order to make the know-how and technology available worldwide.

CASE STUDY: JOINT AND INDEPENDENT TECHNOLOGY EVOLUTION PROVIDED GLOBAL CONFIDENCE

In many ways, the implementation of recycling in the stationary sector paralleled that in the automotive sector.

Another issue that will become important to the industry as we face the challenge to protect climate is the *global warming potential* of HFC-134a (*see* Global Warming Potential (GWP), Volume 1). Controls are considered for HFC-134a mindful of its environmental impact, the quantity of use and emissions throughout the world, the importance of its applications, and the availability of technically feasible alternatives. Industry has spent a great deal of time and money commercializing HFC-134a for important air conditioning and refrigeration applications. Is industry capable of exercising self-control, and avoiding unnecessary and conspicuous uses of HFC-134a? Can chemical suppliers resist the temptation to market HFC-134a into cosmetic aerosol applications or for the proposed self-chilling beverage that exhausts several ounces of HFC-134a to cool each can of drink? Can all countries implement best containment practices and recycling for vehicle HFC-134a air conditioners?

Potentially excessive emissive uses of HFC chemicals make them easy targets of regulation. However, it is within the power of industry to recognize these vulnerabilities, to look beyond this quarter's sales figures, and to take the important actions necessary to protect investments and markets over the longer term.

LESSONS FOR CLIMATE PROTECTION

When the Kyoto Protocol on Climate Protection was signed in 1997, a vocal majority of global industry predicted dire consequences from possible regulations to reduce the emissions of greenhouse gases – an end to economic growth, drastic changes in life-style, mass starvation, and political chaos.

Although climate protection is more daunting than ozone layer protection, corporate and military leadership is already showing the world that technology can reduce greenhouse gas emissions. Climate protection pledges by multinational companies (Box 4) and new pro-environment business organizations (Box 5) are consolidating business support and advocating prompt action.

The 1998–2000 EPA Climate Protection Awards winners illustrate this leadership. British Petroleum is reinvesting in sustainable energy and advertises BP: Beyond Petroleum . Trigen Energy supplies heating, cooling, and electricity with half the fossil fuel pollution of conventional generation and Enron Corporation is drastically reducing the emissions from natural gas pipeline compressor stations. Compaq Computer Corporation is integrating environmental considerations into its products, facilities, and operations, while Intel is making semiconductors that automatically save energy. IBM Corporation is the first multinational electronics company to announce phospho uorocarbon (PFC) emissions reduction goals; and Royal Philips Electronics is showing its customers how to reduce energy use by

Box 4 Corporate climate leadership pledges

2% Energy reduction per year	Dow
40% PFC reduction by 2003	IBM
50% PFC reduction by 2010	Motorola
65% Emission Reduction by 2010	Dupont
25% Reduction by 2007–2010	Mitsubishi Electric, Ontario Hydro, Philips, Sanyo, and United Technology
10% Energy reduction by 2008	British Petroleum
10% Energy reduction by 2002	Shell
20–30% Less energy intensity	Japanese & German Steel
10% Renewable Energy by 2010	Dupont

PFC, perfluorocarbon

Box 5 Industry associations advocating climate protection

- Alliance for Responsible Atmospheric Policy
- Business Council for Sustainable Energy
- International Climate Change Partnership
- International Cooperative for Environmental Leadership
- Pew Center on Global Climate Change
- World Semiconductor Council

one-third or more. Trane air conditioning chillers supply extraordinary energy efficiency and near zero refrigerant emissions and Whirlpool invented the super-efficient refrigerator that surpasses energy standards by 30–40%. Honda and Toyota have commercialized the world's first hybrid passenger vehicles and Nissan is lowering emissions and fuel use with new direct injection technology. The US Naval Surface Warfare Center stern ap reduces the power required to propel a ship by 6–14%. McDonald's Corporation's Earth Effort makes climate protection a high priority and the Walt Disney Company Environmentality encourages Disney cast members and employees to undertake positive environmental practices at home and on the job (Box 5).

CONCLUSIONS AND HOPE FOR THE FUTURE

Technology cooperation under the Montreal Protocol mobilized experts from fiercely competing companies to share information and to speed technology development and application. Military organizations broke with precedents and made ozone layer protection a priority for environmental security. Leadership companies announced ambitious goals that stimulated suppliers to do it better without ODSs. Public gratitude and confidence rewarded innovation.

The history of stratospheric ozone protection is repeating itself in climate protection. Multinational companies are declaring climate protection goals. Industry associations are supporting the Kyoto Protocol. National governments are choosing emission reduction strategies. Perhaps most remarkably, industry is working with government to pursue voluntary measures, complemented with regulatory incentives to protect the climate.

See also: Ozone Layer: Vienna Convention and the Montreal Protocol, Volume 4.

Emerging Environmental Issues

R E MUNN AND PETER TIMMERMAN

University of Toronto, Toronto, Canada

*Emerging environmental issues are often thought of as shocking and surprising events, like the discovery of the ozone hole in the Antarctic or the oil spill of the Exxon Valdez. However, most of the issues that are expected to emerge in the 21st century are actually intensi cations or complications arising from already existing trends or problems. One of the important developments of the modern world has been the increasing attempt to predict or forecast the future, as a basis for planning: this has become even more of an imperative as the potential impacts of human decisions and activities stretch further out in time. H.G. Wells once remarked that the 20th century would be a race between education and catastrophe - the arrival of the 21st century has not invalidated the remark, merely moved the nish line. Futures studies, scenario building, forecasting and computer modelling of issues like climate change are all current attempts to foresee possible emerging environmental issues (*see *Futures Research, Volume 5; Scenarios, Volume 5).*

EMERGING ENVIRONMENTAL ISSUES: DEFINITION

It is important to be clear at the outset that an emerging issue is not necessarily an issue that no one has ever heard of, or that comes as a shocking surprise. In a questionnaire survey conducted by United Nations Environmental Programme (UNEP)/Scientific Committee on Problems of the Environment (SCOPE) by the authors and Anne Whyte in 1999, the following working definition of emerging issues was set out:

Emerging environmental issues are defined as those issues (both positive and negative) which are not generally recognized but which will have significant impact on human and/or ecosystem health in the 21st century. They can have impact globally or regionally and can include local issues that occur in many parts of the globe. Emerging environmental issues are associated with one or more of the following situations:

1. One or more of political, social, economic, financial, institutional or technological developments that are likely to cause changes in current trends of human activity which in turn lead to significant environmental changes.
2. New evidence or theory which suggests potentially large environmental changes, but which are currently either not widely accepted or are considered unproven. This would include new recognition of synergisms in human activity, which produce significant environmental change.
3. Lack of adequate policy, action or leadership on an existing issue, which will become more significant, or more urgent in the future.

Examples of past emerging environmental issues cited in that questionnaire included: the effects of pesticides on birds (1950s); the long range transportation of atmospheric air pollutants (1960s); the death of the Aral Sea through uncontrolled water abstraction (1970s); the elimination of guinea worm through water management (1980s); the collapse of North Atlantic fish stocks (1990s).

From this definition it can be seen that numerous kinds of environmental issues can come under the rubric of *emerging*. What this suggests is that a wider exploration of the context out of which these issues emerge, their characteristics, and their implications for environmental policy would be worthwhile. The rest of this article attempts to do this, first, by a discussion of what can be called the arrival of the future; second, by an analysis of some of the recent results of probes into possible emerging issues; and third by some conclusions concerning how to manage emerging issues.

THE ARRIVAL OF THE FUTURE

The 20th century has seen human activities grow to the point where they are capable of affecting and in uencing many of the physical systems of the planet. The increase in our physical capacities ought to be matched by an increase in our ability to understand and predict the consequences

of our decisions – in some sense, to learn how to manage ourselves. There has in fact, been a slow and painful growth of a new set of institutions, ethics, and techniques to begin to cope with our changing status. Central to this growth has been an increasing concern with the future.

At the beginning of the 18th century, the idea of the future barely existed, except as the predictable end of time in the last judgment. With the rise of the idea of progress and development, the future began to be more interesting; and ironically, this happened at the same time as the past began to stretch out further and further backwards. By the beginning of the 20th century, especially in the works of writers like Wells and Verne, predictions of the future became part of popular culture. More directly, through the rise of bureaucracies and planning as part of social management, nation states began to project and implement policies and structures for 10, 20, 30, and more years into the future, particularly since the 1950s. Modeling, and more recently, computer modeling, has made everyone familiar with projections and forecasts as a basis for decision-making.

Now, at the beginning of the 21st century, the future is a subject of widespread debate, controversy, and premature colonization (for example, as one of many, see Kennedy (1993)). Forecasting and expert projections lead decision-making in many arenas, in spite of the ability of events and short-term responses to crises to derail the best laid plans. The credibility of these forecasts, and the weight placed upon them, are central features of contemporary life. For example, decisions are made according to economic forecasts whose predictive ability is not apparently better than other kinds of forecasting. The great current debates over the ozone hole, climate change, and others are in part debates over whether we can have faith in computer models.

The future is casting a longer and longer shadow back over the present, and of course, *vice versa*. Our decisions today are ramifying further and further into the distant future. Similarly, our lack of decision making, given lag times, means that we are using the future as a storehouse for the buildup of risks that we are unable to respond to adequately now. The more that environmental issues deal with cumulative effects and long-term impacts, the more important socially and politically the notion of the future becomes, and the more troublesome.

Who are we to believe? What trust do we have in prediction? Is the present any guide to the future, or is it totally misleading in an era of rapid change and transformation? How are we to probe ahead?

As already mentioned, individual and collaborative quantitative and quantitative/qualitative future modeling for environmental issues has been developing steadily, at least since *Limits To Growth* (Meadows *et al.*, 1972). More qualitative future searching includes:

- popular or mass surveys of citizen opinions about the future or future concerns sometimes, leading to self-fulfilling prophesies;
- key actor interviews; in depth interviews with experts on possible futures in their area of expertise;
- modified Delphi techniques (A *Delphi technique* is an iterative process of expert consultation, gradually focussing on mutually acceptable outcomes, for example, the continuing Millennium Project of the United Nations University).
- scenario creation, for example, the various activities of the Global Business Network, including *The Futures of Women* (McCorduck and Ramsey, 1996).
- general efforts at envisaging futures through future studies (*see* Scenarios, Volume 5).

There is a burgeoning literature on this subject ranging from books to articles (e.g., Southern Africa Beyond the Millennium: Environmental Trends and Scenarios to 2015, *Environmental Planning Issues*). So extensive has this kind of literature become that there is a volume of 450 abstracts of essential books, reports, and articles just on environmental issues and sustainability (*Environmental Issues and Sustainable Futures*, ed Michael Marien, from the World Futures Society and IUCN, 1997).

A particularly interesting trend is the spread of creative approaches to considering the future from, for instance, science fiction, to a wider professional and occasionally public constituency. One version of this, under the general label of backcasting, involves the generation of different alternative futures for a given place or problem, and then using expert techniques to map how one would approach a preferred outcome. This is a looser form of long-range planning. Similarly, the availability of inexpensive computing power has spawned the rise of a number of expert consultative workshops where possible alternatives are assessed and winnowed in some detail. These approaches have their drawbacks, but their most significant advantage is that they can provide something of a feel for the complexities of future issues as they emerge.

CHARACTERISTICS AND TAXONOMIES OF EMERGING ISSUES

This whole discussion presupposes that issues emerge out of the background of a stable system, which then proceeds to manage them, and then new issues emerge, and so on. To think this way naturally promotes the identification of surprising issues that threaten the *status quo*, single issues that can be picked out of the surrounding noise, or issues whose emergence is not so much emergence into the world as their emergence on the agenda of powerful nations or media.

Reference to reports such as *GEO-2 Outlook* (for which the UNEP/SCOPE Questionnaire already cited was preparatory) suggests that for the expert international scientific community, the main emerging issues are rooted in long-term systemic change, the intensification and ramification of existing issues, and the spread of existing issues from one part of the world to another. Surprises, while not discounted, are not of great concern: rather, one might say it would be surprising if some of the stresses on major ecosystems did not result in some unexpected expressions or specifically unpredictable breakdowns.

More theoretically, Kalhok and Gomes (2000) in a working paper for the UNEP/SCOPE Project suggested the following taxonomy of emerging issues, based on historical examples:

Unforeseen Issues

New issues rarely appear without warning; however, their lack of frequency does not undermine their importance.

Chlorofluorocarbons (CFCs) Induced Ozone Depletion

This was virtually unheard of outside the scientific community at the time of Stockholm. Ozone depletion gained some attention by the public following a 1974 front page story in the *New York Times*, but it suddenly became an alarming issue worldwide for scientists, policy makers and the general public, following the discovery of a large ozone hole over Antarctica in 1985. CFCs, which were reported in 1973 to be chemically inert and harmless to the environment, were later determined to be the primary cause of ozone depletion.

Unlikely Events

Certain events may be known to be possible, but their severity, timing and location is unanticipated. With the expansion of human communities and industrial activity, such events have the potential for increased frequency and impact.

Oil Spills

For example, the Torrey Canyon (1967); the Amoco Cadiz (1978); the Exxon-Valdez (1989); Persian Gulf (1991).

Accidental Poisonings and Toxic Chemical Events

For example, methylmercury poisoning, Minimata, Japan (1959); PCB poisoning (Itai-Itai), Kyushu, Japan (1960s); dioxin leak, Seveso, Italy (1976); methyl isocyanate leakage, Bhopal, India (1984); chemical warehouse fire, Basel, Switzerland (1986).

Severe Smog and Air Pollution Events

For example, London (1952), Indonesia forest fires (1997).

Nuclear Accidents

Urals (1958); Three Mile Island (1979); Chernobyl (1986).

Biological Invasions

Zebra mussels, Great Lakes (1980s); Mesquite trees, intentional introduction with unexpected agricultural effects, Sudan (• 1950).

Emergence of New Perspectives

Existing issues may be brought into new light by new developments or findings or by media involvement.

Pollution

For example, *Silent Spring*, the 1962 publication about effects of pesticides and herbicides; discovery of toxic substances in the Love Canal, NY; discovery of contaminants in the Arctic, far from areas of use.

Acid Rain

The Swedish Case Study (1972) presented at Stockholm stimulated international concern and action.

Climate Change

The media sensationalized the unseasonably hot summer of 1988 as evidence of global warming which stimulated scientific research and increased public awareness.

Tropical Deforestation

Satellite images of tropical deforestation in the 1980s vividly demonstrated the extent of the loss of biodiversity.

Consequences of Resource Management

For example, ecological, social, economic and health impacts of the Aswan Dam and the shrinking of the Aral sea from water diversion for irrigation.

Changing Trends

Discontinuities in an expected course of events can be ecologically and economically devastating.

Climate Change

For example, increased occurrence of severe weather events and El Nino.

Resource Depletion

For example, Oil crisis (1970s); Atlantic cod fishery (1990s).

Shifts in Environmental Perception

The way in which scientists, policymakers and the general public view environmental issues changes, often with remarkable rapidity.

1. Earthrise photo from Apollo II (1969) instilled a perception of Earth as a fragile, unified ecosystem;
2. Focus of issues moved from local to national to international due to increased awareness of the transboundary nature, or spatial extent of many issues, e.g., marine and air pollution, habitat loss;
3. Focus on interdisciplinary and interconnectedness of issues: e.g., population, pollution and resource depletion, which were once viewed as independent issues are now also seen in the context of broader issues such as sustainable development, climate change and biodiversity loss, which themselves are interconnected.

A related taxonomy would refer to the internal structure of emerging issues.

Intensification

An existing issue becomes radically more difficult, with a new intensity of impacts.

Extensification

An existing issue spreads, affecting a new range or space.

Boundary Crossing

An existing issue moves over its existing boundary or category, complicating institutional responses.

New Emergent Properties

An existing issue is transformed by synergy or complex internal properties into a completely new problem.

Emergencies/Surprises

A new problem emerges out of nowhere.

Cascades

Higher level problems that set off new responses at lower levels.

Flip-flops

Systems under stress that flip from one state of the system to a quite different state very quickly.

New Problems from Old Solutions

A solution turns into the next big problem.

Other kinds of taxonomy are available, e.g., based on different kinds of risk assessment, or as functions of a scenario-based approach.

THE EMERGING ENVIRONMENT: OVERVIEW

The results of the SCOPE/UNEP Questionnaire (2000) and further consultations suggest that while there are specific priority emerging environmental issues, it is the overall systemic changes underway that are of greatest concern. These changes are very difficult to manage, because they are insidious, cumulative, and widespread, crossing boundaries and jurisdictions. They include ecosystem degradation, loss of biodiversity, major reductions in available freshwater, and so on.

These issues are emerging out of a situation wherein human beings are affecting and restructuring for their own purposes an increasing fraction of the physical and biological systems of the earth. In spite of a number of efforts, it is not known how many people the earth can support, and how the burden of a substantially increasing population with increasing expectations is to be sustained. Some commentators believe that technological innovation will ease this burden substantially in the coming decades; others believe that at least locally, many ecological systems have reached or are already near the breaking point.

As part of the growth of human influence over the Earth, we are also witnessing the spread of a social and economic development pattern originating in the West, which has been amazingly successful to date. The globalization of this pattern, which includes commitments to resource exploitation, increasing free trade, and the promotion of consumer demand as an engine of growth, is of profound environmental significance. For example, one of the basic assumptions of free trade is the cheapness of international transportation, which allows product components to be made in one part of the world and shipped to other parts for assembly. However, the cheapness of transportation is due in large part to the lack of internalizing of the environmental costs of fossil fuel emissions, a major cause of the expected global climate change.

It is also worth noting that (at least in developed countries) there is a trend away from the point source pollutants which figured so largely in the early days of environmental politics, that is, the belching smokestack and the pipe disgorging chemicals directly into a river. This kind of pollution has not disappeared by any means, but it is being replaced by non-point-source pollutants that come from a myriad of small users, or from the overall lifestyle choices of more affluent peoples.

Overall, we can see that the preferred solution to environmental problems is increasingly considered to be economic growth as a spur to improvement, and yet economic growth, while increasing efficiency of production, seems to promote excessive consumption. We are presumably in a transition period between the environmental problems associated with the former, towards environmental problems associated with the latter.

THE EMERGING ENVIRONMENT: THE ROLE OF SCIENCE

At the same time, thanks to the efforts of the scientific community in the past half century, we are beginning

to understand the complexity of global ecological cycles, and the need to promote biodiversity and the preservation of ecological integrity as basic goals for a sustainable society. The examples of dichlorodiphenyltrichloroethane (DDT) and the results of the atmospheric testing of nuclear weapons in the 1960s, which ironically led to great increases in our understanding of the pathways through which chemicals and ionizing radiation move, remind us that emerging issues are often surprises and challenges to our understanding of basic processes, and prod us towards seeking better methodologies of detection and assessment.

The last 50 years have seen enormous advances in ecosystem science and in the development of policies that may allow us to slow down or cope with environmental change in a sustainable manner. The progression of knowledge started with integration among and between the various physical and life sciences, and has led more recently to dialogue with the health and social sciences and engineering.

The role of science is of course central to the identification of many emerging issues. Most obviously, the increasing use of applied science in technology is a major force in human development; but the increasing ability of science to detect and monitor changes in the physical world, either through better understanding or through technical aids such as satellites, is also important. The past 50 years have seen revolutions in understanding how environmental systems function, how complex their interconnections are, and how difficult it often is to determine the impacts over extended distances of space and periods of time of environmental changes.

One of the most important single changes has been the ways in which science and the scientific community have emerged as significant factors in the identification and shaping of emerging issues on national and international political agendas. The international accords on the protection of the ozone layer, biodiversity, climate change, and others, were not only framed and sustained by scientific input; but in some cases (such as the Climate Change Convention) were initiated by members of the scientific community. Because of this shift in their role, the scientific community has begun to develop new mechanisms for consultation and reasoned sifting of evidence, of which the most prominent is probably the Intergovernmental Panel on Climate Change (IPCC).

While some may see these developments as the politicization of science, others suggest that we are entering an era of post-normal science (*see* Post-normal Science, Volume 5) where the management of complex systems under conditions of future uncertainty for human benefit requires a more open acknowledgment of the biases and agendas of scientists and scientific bodies. This is particularly important if one agrees that the definition of an emerging issue is in part dependent on past histories, present priorities and future hopes and fears.

The way the issues are defined obviously affects the way in which the solutions are defined, including the mandates for responding, the resources and institutions brought to bear, and so on. It is, for instance, possible to ask why certain emerging environmental issues have been defined and categorized in certain ways. It took until the 1990s for women's education to be widely acknowledged as a central component of environment and development. Climate change is rarely seen as partly the result of a need to subsidize cheap transportation by fossil fuel use as the basis for increased global trade.

For science and policy purposes, certain sets of distinctions may be useful. We can distinguish between:

- new issues that emerge because of scientific evidence or theory, and which are currently unknown or unproven;
- issues that are well known and proven, but which are not yet on the policy agenda. That is, they are taken seriously by scientists, but not by policy makers.

There may be many reasons for the existence of the second group, ranging from the inability of the scientific community to articulate clearly the characteristics or consequences of the issue. Inherent uncertainty in the projections (making it easy for policy people to adopt a wait-and-see attitude) and outright refusal of the political community to respond in spite of repeated warnings.

We can also distinguish between:

- issues that are connected to one or more point sources, and are therefore relatively easy to pinpoint and manage;
- issues that are rooted in overall social trends or derive from non-point sources, and are therefore much harder to eliminate.

An example of this distinction is between the debate over the elimination of CFCs from the atmosphere versus the much more intractable issue of decreasing fossil fuel use to slow the onset of climate warming.

A third distinction is between:

- issues that are global in scope, but with regional and local consequences;
- issues that are emerging primarily at the regional and local level.

This is not a hard and fast distinction. Purely regional issues can be driven by global trends. Nevertheless, there are issues whose expression or characteristics are firmly regionally based.

Other scientific characteristics of certain emerging issues can be described. They tend to be complex, with multiple

linkages between causes and effects. They are often synergistic, with widespread consequences, and with multiple causes. These issues are often controversial, and characterized by uncertainty. This is sometimes because the scientific uncertainty is too great, due to lack of evidence, the available statistical data are unreliable, or the response of complex systems is unpredictable. On the other hand, scientific uncertainty can be used as an excuse for delaying appropriate political responses. This in turn may be the reason for an issue to progress from being a twinkle in a scientist's eye to being a fully edged problem that is still increasing because it is not yet being addressed by policy makers.

For some issues, local and regional dynamics are currently better understood than global ones; in others, the local scale is less well understood, while the overall nature of the problem is fairly clear. For example, the introduction of new environmental technologies to developing countries is widely seen as a preferred solution to many problems; but it is already clear that the process of introduction can be disruptive and oppressive. The notion that developing countries might have solutions to export to developed countries for their environmental problems is almost never considered.

As in the case of environmental technology, a widespread problem with emerging issues in general is that the socio-economic variables involved in predicting the sequence from issue identification through to impacts are just too complex. Human responses are difficult to predict and/or are understudied in relation to the problem, or often just assumed. This inability to demonstrate consequences weakens the case for immediate action, and policy makers then tend to adopt a wait-and-see attitude.

Many emerging issues are the expression of unsustainable and unintegrated management, where among other things, the relationships among and between issues are being ignored. Further complicating the situation is the fact that the social and political will to respond to the issue is often not there, due to other priorities, short-term economic goals, and a lack of champions or constituencies for the issue, and/or a lack of readily available knowledge.

More and more, though much work is rooted in studies carried out within a single discipline (e.g., meteorology), the dimensions of each issue (e.g., climate change) are increasingly expanding into more and more fields (ecology, chemistry, engineering, the health sciences, economics, sociology, policy analysis, etc.). As an example, the provision of clean drinking water has traditionally been a problem to be solved by hydrologists and engineers. However, when drinking water quality began to deteriorate in New York City in the early 1990s, an environmental assessment of policy options included an economic analysis. In this case, the economically preferred alternative was found to be also the ecologically preferred one. The cost of a new filtration plant would have amounted to between six and eight billion dollars, plus operating costs of 300 million dollars annually. On the other hand, the cost of protecting the watershed (in the Catskill mountains) by eliminating the sources of pollution (sewage, fertilizers and pesticides) in that area was only about 1–1.5 billion dollars (Chichilnisky and Heal, 1998).

These considerations make the identification and management of emerging issues incredibly complex, and as stated often already, subject to interpretation and myriad obstacles, conceptual as well as physical or economic.

THE MANAGEMENT OF EMERGING ISSUES

While there has been a plethora of futures studies, Delphi methods, and scenario constructions in the past few years, almost none of these approaches have dealt with how best to manage the issues that emerge from these studies. They have contented themselves with identifying the issues, often in a scatter gun approach. A more useful approach would relate the characteristics of emerging issues to the characteristics of the management systems that are required to cope with them.

For instance, in Western democratic systems, the social and political systems, while increasingly dependent on forms of forecasting, continue to be focussed on the very near-term demands of the electoral cycle. There are only marginal resources for coping with future issues: the early warning systems in place (the antennae) are assumed to be part of the mandate of the academic and scientific communities scanning the horizon; and the technological and economic communities seeking out new opportunities. However, even these marginal resources can be significant.

Nevertheless, a central problem for the policy makers is that emerging issues are on contested terrain. Because of the uncertainties involved, the hopes and fears they carry with them projected into the future, and the complex relations they have with existing issues, emerging issues can be very controversial, and therefore hard for policy makers to respond to adequately. They are often ash points, institution redefiners, triggers for a cascade of further issues. They are often seen as best avoided: with all the current problems, why take on new ones?

On the other hand, the appropriate management of emerging issues can:

- help shape internal priorities, and build on immediate-term successes;
- provide for some resilience and safety nets in the face of inevitable surprises;

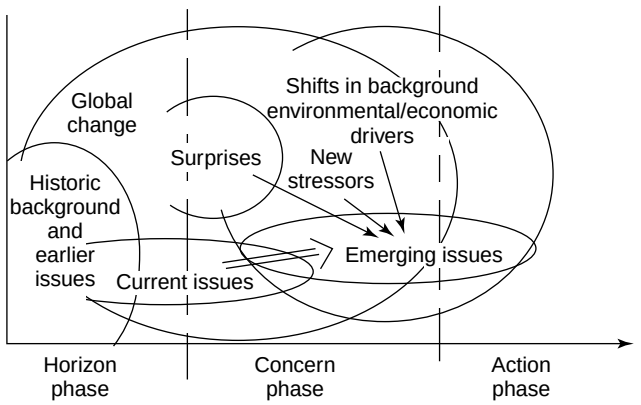


Figure 1 Phases of emerging issues

- open a useful window for wider public input into the policy-making process by connecting emerging issues to larger concerns and hopes.

One useful approach, differing from the earlier taxonomy (by characteristics of emerging issues), is to consider how issues emerge into the public domain (or the scientific domain), always keeping in mind the caveats mentioned earlier about reinforcing the *status quo*. In this approach, the continuity of emerging issues with already existing issues is emphasized: they have a life cycle, which moves them from below the horizon, into the domain of concern, when forces and resources begin to be mobilized to respond, and then into action. This can be followed by reaction, recurrence, or transformation into the next set of emerging issues.

This is sketched out in Figure 1. For policy makers, what is important are the phase shifts of concern, and the triggers that alter the response. There may be different management responses appropriate for each phase.

Horizon Phase/Horizon Issues

Horizon issues are issues that have simply been identified, but have not yet crystallized to the point where their implications and potential impacts can be determined. They may have been identified: their characteristics are still unclear.

In dealing with horizon issues, management needs to take into account:

1. there is a great deal of noise and early debate surrounding horizon issues;
2. these issues are still very uid and in many cases not subject to clear definition;
3. there is a need to gather information from a wide range of sources, and not to lock in on the issue too early;
4. the role of science in this period is to ensure that basic rules of scientific evidence and experiment are being followed in the literature, and if research is being carried on internally. Because of the embryonic status of the issue, in-house science is likely to be most in uential in assessing the credibility of the issue, suggesting appropriate analogies from other related or hitherto unrelated cases, and so on.

Concern Phase/Issues of Concern

In this phase, levels of concern depend upon different criteria, for instance:

1. scientific consensus forming;
2. substantial public concern and input;
3. potential for serious risk;
4. appearing in multiple jurisdictions;
5. availability of appropriate responses;
6. new stressors, real or imaginary, identified;
7. predicted impacts are beginning to appear, etc.

In most bureaucratic systems, this phase is characterized by the initial formulation of policy, creation of task forces, mobilization of research towards policy conclusions, etc.; still falling short of accountable activity.

Action Phase

The Action phase takes place when the issue has entered the full formal policy formulation stage and has become a current issue subject to the usual procedures.

For policy purposes, this kind of sequence of management phases is normal practice, but it is hardly ever articulated very well. For instance, issues that are still inchoate and uncertain don't tend to be immediately enshrined in legislation. The response is rather to wait and see; to scan for the best information; to determine if this issue is worth spending time and resources on; to find out what the leading persons in the field think, and so on. As the article on Social Learning discusses (*see* Social Learning in the Management of Global Atmospheric Risks, Volume 5), study of the ways society learns about new environmental issues is still in its infancy. The movement from the concern phase to the action phase is subject to all kinds of forces, including champions of the issue, strong nongovernmental organizations (NGOs), appropriate available information, media coverage, etc. Only a small part of this whole emerging response to emerging issues can be captured simply by looking at official government action.

Sometimes this system works reasonably well: sometimes, the strategy fails. Wolf is cried, and the wolf in fact appears at the door. The discounted radical scientist turns out to be right; society refuses to acknowledge the problem for social, economic, and even religious reasons; resources are misapplied; monitoring systems are inadequate; obstacles are overcome too late.

EARLY WARNING INDICATORS

An essential element of any strategy for managing emerging issues is to design monitoring systems that can provide early warning of impending change. There are fundamental difficulties in designing such systems, including:

- Current environmental and socio-economic monitoring systems were originally designed for other purposes, often yesterday's emerging issues and not tomorrow's. They underpin and reinforce current institutions and approaches to dealing with issues. These monitoring systems may generate new kinds of warnings only if there is bleeding or indirect in uence from some quite different stress that affects an existing monitored stress.
- Changing existing monitoring systems to handle new issues may affect the continuity of valuable time series, and should be done with care. The ability to interpret or reinterpret (to mine) existing data sets is often a critical alternative.
- Monitoring systems often are not designed within a larger design framework, so that (for example) socio-economic variables and environmental variables operate on quite different temporal and spatial scales. Linking socio economic with environmental variables is fraught with a myriad of technical difficulties.
- Multiple stresses affect ecological systems, and it is often difficult to connect cause and effect. For example, if forest dieback occurs, is it due to air pollution, drought, bark beetles, nitrogen overfertilization, a particularly sever winter or hot summer, or all of the above?
- Similarly, there may be long lag times between stress and response, such as happens in poorly buffered lakes that are subjected to long regimes of acid rain, and only when the buffering capacity is used up, will there be precipitous changes in lake ecosystems. Recent work on the possibility of non-linear ip- ops of ecosystems from one state to another highlight another aspect of potential sudden change in strongly stressed environments that may not be caught beforehand by conventional monitoring.

Historical experience with emerging environmental issues shows that it was not until the fundamental processes involved were reasonably well understood that future trends and scenarios could be predicted with reasonable chances of being correct. In the meantime, early warning systems must be adaptive: more in the nature of probes than a new generation of monitoring systems for today's problems that will become obsolete or rigid (*see* Monitoring in Support of Policy: an Adaptive Ecosystem Approach, Volume 4).

This supports the idea of using Biosphere Reserves and other model ecosystems as early warning systems in their own right, indicating through changes in their resilience and vulnerability the interactive effects of multiple stresses and larger global shifts in environmental dynamics.

Along the same lines, given the difficulties involved, a basic management strategy for emerging environmental issues must involve the precautionary principle (providing for enough working space for margins of error in natural and human processes) so that no emerging issue becomes large enough to endanger the capacity of the entire system to respond creatively to unexpected change (*see* Precautionary Principle, Volume 4).

REFERENCES

Chichilnisky, G and Heal, G (1998) Economic Returns from the Biosphere, *Nature*, 391, 629–630.

Kennedy, P (1993) *Preparing for the 21st Century*, Harper Collins, New York.

McCorduck, P and Ramsey, N (1996) *The Futures of Women*, Time Warner, New York.

Meadows, D H, Meadows, D L, Randers, J, and Behrens, W W (1992) *Limits to Growth*, Universe Books, New York (A Report to the Club of Rome).

FURTHER READING

Linthurst, R A, Bourdeau, P, and Tardiff, R G (1995) *Methods to Assess the Effects of Chemicals on Ecosystems: SCOPE 53*, John Wiley & Sons, Chicester.

Munn, R E (1995) *Atmospheric Change in Canada: Assessing the Whole as well as the Parts*, Institute for Environmental Studies and Environment Canada, Toronto.

Munn, R E, Timmerman, P, and Whyte, A (2000) Emerging environmental issues: the Walter Orr Roberts lecture, *Bull. Amer. Meteorol. Soc.*, 81, 1603–1609.

Munn, R E, Whyte, A, and Timmerman, P (1999) Emerging Environmental Issues: a Global Perspective of SCOPE, *Ambio*, 28(67), 464–471.

OECD (1998) *The OECD Megascience Forum: Workshop on Global-scale Issues*, OECD, Paris.

Risser, P G (1991) *Long-term Ecological Research: an International Perspective: SCOPE 47*, John Wiley & Sons, Chicester.

United Nations Environment Programme (1999) *Global Environment Outlook 2000*, Earthscan Publications for UNEP, London.

Adaptation Strategies

IAN BURTON

Meteorological Service of Canada, Downsview, Canada

The concept of adaptation has come into prominence recently as a result of its importance in human response to climate change. In this connection, adaptation refers to " the adjustments in ecological, social or economic systems in response to actual or expected climatic stimuli, their effects or impacts. It refers to changes in processes, practices, and structures to moderate potential damages or to bene t from opportunities associated with climate change."

The concept of adaptation has a long intellectual history and this is brie y sketched below as a background to its present use in global climate and environmental change. A summary of some ideas about the anatomy of adaptation is provided as well as some ideas about the emergence of a science of adaptation.

Three kinds or levels of adaptation are described and their potential role in sustainable development is discussed.

CONTEXT, DEFINITIONS, HISTORY

The word *adaptation* has been propelled into prominence in the context of global environmental change as a result of its use in the United Nations Framework Convention on Climate Change (UNFCCC), signed at the Earth Summit in 1992. The convention makes a distinction between mitigation and adaptation as responses to climate change. *Mitigation* refers to those actions that might be taken or promoted to prevent or slow climate change, by controlling the level of concentrations of greenhouse gases in the atmosphere. This includes reduction of emissions, such as the reduction of carbon dioxide (CO_2) emissions by increasing energy efficiency, fuel switching, and the use of renewable sources of energy. It also includes various forms of carbon sequestration through which the stock of carbon in forests, and soils, might be increased.

By contrast adaptation refers to the:

> adjustments that are possible in practices, processes, or structures of systems to projected or actual changes of climate. Adaptation can be spontaneous or planned, and can be carried out in response or in anticipation of change in climate conditions.
>
> (Intergovernmental Panel on Climate Change (IPCC), 1996)

Adaptation can be used to moderate potential damages or to benefit from opportunities associated with climate change.

The UNFCCC uses the concept of adaptation to apply to both natural and socio-economic systems. There is, in fact, a long history of the term *adaptation* being used in this way, going back at least to Charles Darwin (1809–1882), and the philosopher and sociologist Herbert Spencer (1820–1903). For Darwin, adaptation was the result of the process of natural selection acting upon natural variability of biological forms and attributes. For Spencer, adaptation is a feature of the fitness of human beings and their prospects for survival. The term *adaptation* in the social sciences thus came, in the 19th century, to be associated with extreme *laissez-faire* policies, and served as a rationalization for racist and other discriminatory beliefs. Although such beliefs are now widely discredited, the old debates about the relative importance of genetic (nature) versus environmental (nurture) in uences have resurfaced in the current ideas of sociobiology (Wilson, 1975, the most authoritative summary of the socio-biological approach to human adaptation). See also Wilson (1998), and its extensions into evolutionary psychology and other social sciences (Wright, 1994), an excellent summary for the general reader of the application of sociobiological approaches into the social sciences and psychology. This intellectual history complicates the use of the word *adaptation* in the current global environmental change context, and it is important therefore to be clear about what sort of adaptation is meant, and what definition is being employed (Dubos, 1965, a classical synthesis of adaptation in human biology and physiology with special reference to human health. See also Dobzhansky, 1962).

The *Oxford English Dictionary* defines adapt' as meaning to fit, to make suitable , or to alter so as to make fit for a new use . Two examples are given of adaptation.

The structure of the outer ear is adapted to collect and concentrate the vibrations , and a comedy adapted from the French . The distinction is important because the processes of biological and social adaptation are so different. It would be less confusing if a specific word were to be employed for each. In fact, the geographer Gilbert White consciously chose to use the phrase *human adjustment* rather than adaptation, when studying the way in which human societies respond to ood hazard (White, 1945, a classic early study that pioneered the approach of human adaptation). The Intergovernmental Panel on Climate Change (IPCC) definition cited above does indeed use adjustment as a synonym for adaptation when referring to socio-economic adaptation.

In addition to the mobility issue, some ecologists are concerned about nitrogen overfertilization and about the harmful effects of global change on species competition, which in both cases might result in biodiversity losses. In the nitrogen overload situation, some species would ourish while others would be seriously stressed, leading to a changed composition of ecosystems (*see* Nitrogen Cycle, Volume 2). In the latter case, some species will adapt more rapidly to global change (particularly to increasing carbon dioxide concentrations and to climate warming) so that species competition will favour some species over others (*see* Plant Competition in an Elevated CO_2 World, Volume 2).

In the early phase of the science and policy debates about climate change (say 1975–1992), interest focused on the adaptation of species and ecosystems to climate change or biological adaptation. Anthropogenic climate change constitutes a threat to natural ecosystems, and there is a risk of mass extinctions (reductions of biodiversity) since the rate of change is projected to be much more rapid than the time required for biological adaptation. Of course plants and animals can, given time enough, migrate (not strictly an adaptation in biological terms) and thus move with the climate. For example, the climatic zone now occupied by the boreal forest that extends in a wide belt across North America and Eurasia is expected to move some 200–300 km farther north. This has led to proposals to create migration corridors, or belts of protected land extending on a south–north axis in order to facilitate migration. This may, in theory, help the more mobile species, such as animals, to survive, but to the extent that they are an integral part of an entire ecosystem and require its support, this form of adaptation depends upon the least mobile elements in the system. More extreme cases of the threat of extinction can occur in small island ecosystems and in mountain regions where there is no possibility of adaptation by migration as the climate changes.

More recently (since 1992), questions of socio-economic adaptation have been added to the concerns about non-human adaptation, and this interest continues to grow as the social and economic implications of climate change become more evident. The scientific consensus is that anthropogenic climate change is now happening, and at the same time it is becoming clear that the stabilization of greenhouse gas concentrations in the atmosphere cannot be easily achieved and will take time. The question is no longer, Is human activity changing the climate?' but What are the risks of the impacts of climate change and how can they best be managed?' (Dotto, 1999, an excellent account for the general reader of the risks of climate change, focusing more on mitigation than adaptation).

WHY ADAPT?

In the context of the climate change science-policy negotiations, there are two main reasons for the assessment of adaptation. First, it is important to know more about the means (including the costs) of reducing the impacts of climate change by adaptation, in order to compare this with the means and costs of mitigation. It is often assumed that there is a trade-off between mitigation and adaptation; that is to say that the more successful the mitigation strategy (as proposed in the Kyoto Protocol, for example), the less the need will be for adaptation, and *vice versa*. While this idea has some theoretical appeal, in practice it will be difficult to get enough of either. The stabilization of greenhouse gas concentrations, and agreement on how this is to be achieved, is an enormously complex task and will require considerable time. Meanwhile the rapid rise in losses from extreme atmospheric events (hazards such as oods, tropical storms and droughts) that has occurred over the past three decades (1970–2000) shows that, in some respects, human societies are often not well adapted to current climate variability. Also, in order for a trade-off to be politically acceptable, there has to be some way whereby the winners (those who benefit from climate change or are least damaged) to compensate the losers. At the international level, no such regime or authority exists and solutions have to be found through negotiations, which are protracted, and in a number of respects still inconclusive.

The second reason for the focus on adaptation has to do with the development of adaptation strategies. It is agreed that adaptation to climate change has the potential to substantially reduce the adverse impacts of climate change, and to enhance the beneficial effects. The new attention to adaptation therefore raises many questions. These include the extent to which socio-economic systems may be expected to adapt autonomously (i.e., by individual, community, or corporate decision-making) and the extent to which planned governmental intervention is required. To the extent that governmental action is required, how are specific adaptations to be identified, assessed and prioritized? And how are adaptations to be paid for? This is also a crucial question at the international level, because the developed country

parties to the UNFCCC have agreed in principle to assist the developing country parties (especially the most vulnerable) in meeting the costs of adaptation (*see* Common but Differentiated Responsibility Principle (Stockholm/Rio), Volume 4).

Despite the fact that the climate (as measured by changes in annual average temperature) is projected to increase more towards the poles (high latitudes) and less in tropical regions (low latitudes), the present judgment is that the people and places most vulnerable to climate change are to be found in the poorer low latitude countries. This is partly because some tropical ecosystems and agricultural systems are often already close to the limits of heat tolerance, but perhaps more importantly because such countries are relatively lacking in adaptive capacity. The capacity to adapt to climate change is a function of wealth (financial resources); availability and access to technology; the level of information and skills or trained human resources; the nature of the existing social infrastructure; the strength or degree of development of social institutions, as well as the distribution of wealth in a society or equity. Thus, the rationale for assistance to the most vulnerable developing countries rests upon the fact that they have least adaptive capacity and that they have historically contributed proportionately less emissions of greenhouse gases that have created the problem in the first place.

In the international political climate of the late 20th and early 21st centuries, a Spencerian *laissez-faire* approach is not acceptable. Therefore, the developed-country parties to the UNFCCC have recognized their own responsibility to provide assistance in meeting the costs of adaptation. In order to facilitate this financial support, the Conference of the Parties (COP) to the Framework Convention have recognized the Global Environment Facility as the financial mechanism for the convention. Parties to the convention are required to make periodic national communications in which they report progress in both mitigation and adaptation. To facilitate these national communications, assistance is available to developing country or non-Annex 1 Parties. For the purposes of addressing adaptation, three stages have been established under the COP. These are in the short term:

- *Stage I*: planning, which includes studies of possible impacts of climate change, to identify particularly vulnerable countries or regions and policy options for adaptation and appropriate capacity-building.

In the medium and long term, the following stages are envisaged for the particularly vulnerable countries or regions identified in Stage I.

- *Stage II*: measures, including further capacity-building, which may be taken to prepare for adaptation, as envisaged by Article 4.1(e) (of the Framework convention).
- *Stage III*: measures to facilitate adequate adaptation, including insurance, and other adaptive measures as envisaged by Article 4.1(b) and 4.4.

At the time of writing (July, 2000) some 25 developing countries have completed their first national communications, and the COP has agreed in principle to move from Stage I to Stage II adaptation.

In some countries and regions, a case can be made for adaptation measures to be taken immediately on the basis of the short-term benefits that would accrue. Six reasons are commonly cited:

- Climate change is already occurring and further changes cannot be prevented, therefore there is no justification for a wait-and-see policy.
- Anticipatory and precautionary adaptation is often more efficient and less costly than forced, last minute, emergency adaptation, or relief, rehabilitation and reconstruction after an extreme climatic event.
- Climate may be more abrupt and more pronounced than current models and estimates suggest. Low probability and unexpected events are possible.
- Immediate benefits can be obtained from better adaptation to current climate variability and extreme atmospheric events, which will also help reduce vulnerability to longer run climate change; the so-called *win-win responses*.
- Immediate benefits can also be obtained by removing or halting current maladaptive policies and practices that may tend to increase present and future vulnerability.
- Climate change brings new opportunities as well as threats. Future benefits can be obtained by adapting now in an anticipatory way.

Recognition of the validity of these arguments has led to an elaboration of the concept of adaptation in the context of climate change. A number of experts have developed ideas about the structure of adaptation, referred to as the anatomy of adaptation. Smit *et al.* (2000) provide a structure of the concept of adaptation to climate change. The rest of that special issue of *Climate Change* contains a further 12 papers on various aspects of societal adaptation to climate change and variability to develop what they refer to as the science of adaptation (Smit *et al.*, 1999). Simultaneously with these developments, scientists in many countries are beginning to carry out research on adaptation with a view to strengthening understanding and to creating adaptation strategies or programs (Bwango *et al.*, 2000).

ADAPTATION TO WHAT?

The first question to be addressed about adaptation is the nature of the stimulus or change that prompts the need to adapt. In the case of climate change, the stimulus is

commonly described as the output of general circulation models (GCMs). Initially, these were equilibrium models based upon a doubling of the CO_2 equivalent concentrations in the atmosphere. More recently transient models have been developed which show changes in climate (mean temperature and other variables) over time, beginning with the present. The output of GCMs is used to study impacts upon human and natural systems, and it is these impacts that are the direct source of stimulation to which an adaptive response may occur.

There are many sources of uncertainty in this process. For example, the level of future concentrations of greenhouse gases in the atmosphere will depend upon the amount of emissions from human activities, and this in turn depends upon the rate of population growth, the mix of types of energy used, the rate of economic growth, and the environmental protection or sustainable development policies adopted and put into practice. There are also uncertainties in the atmospheric models themselves, such as the rate of absorption of CO_2 into the oceans and the land biomass. The scale of the GCMs is such that the outputs are generalized over large regions, and local variations in future climate are not well captured. In addition, the outputs have been largely in terms of mean values and have less information on variability, especially climate extremes. However, it is climate variability and extreme events that have larger significance for impacts than changes in mean values.

The task of identifying and selecting adaptation strategies is greatly complicated by the uncertainties of climate change and climate change impacts. To many policy makers, it seems pointless to worry about what adaptation strategies to adopt in response to the impact, decades into the future, of an uncertain climate upon unknown economic and social conditions.

On the other hand, much can be learned from present adaptation policies and practices. Much attention has therefore been focused on present-day adaptation to climate, as well as trying to learn from past experience with climate variability. Lamb (1982) is a classical reference work on the in uence of climate on society through historical times; adaptation is addressed by implication throughout the text. Adaptation strategies to future climate change can best be seen as a process of incremental improvement in present adaptation.

WHO AND WHAT ADAPTS?

All socio-economic systems and natural systems adapt to climate change, including its variability. It is generally agreed that both human and natural systems are now well adapted to present climatic conditions. Human beings have successfully occupied most of the climatic regions of the planet, and thrive in them to varying degrees. There is virtually no climate projected for the coming decades and centuries that is not now found somewhere on earth and where human society is not successfully adapted. Except on the very long time scales, climate varies more over space than over time.

Individual species adapt over long periods of time by genetic variation and natural selection. Migration of species is also a form of adaptation. Species of plants and animals coexist in complex patterns of ecosystems, but whole ecosystems cannot migrate intact because some species migrate much more rapidly than others. It seems likely, therefore, that the projected climate changes will occur too rapidly for many species to adapt naturally, and that unless human intervention and management can be successfully employed, many species may be lost. Climate change poses a substantial risk to biodiversity. Human beings also adapt to climate on a physiological level, and can become acclimatized to different conditions.

In socio-economic systems, much adaptation takes place at the individual and community levels. Farmers choose crops and the timing and manner of cultivation according to their understanding of climate and climate risks. Houses, roads, bridges, ports, drainage systems, transmission towers and the whole array of human-built infrastructure are designed to withstand the climate extremes of a specific location. Many design features result from centuries of experience and trial and error. For example, local farmers in the Andes have known for centuries the optimum height above sea level to grow maize: at too high an elevation, there is danger of frost while at too low a height, there is danger of drought. How will these people adapt to climate warming? In modern societies, design standards and criteria are carefully worked out by professional experts backed up and implemented by government-approved regulations. So in addition to adaptation at the individual and community level, governments play an important role in facilitating and sometimes requiring adaptation.

Private sector corporations and businesses also adapt to climate, both in the design of their facilities and operations and in responding to variations in market demand prompted by climate variability. In short, climate is such a pervasive factor that everything and everyone adapts. For the development of adaptive strategies to future climate change, there already exists a strong basis of knowledge and experience in adapting to present and past climates.

There are signs however, of less than adequate adaptation in some circumstances. The loss of life and property from extreme climatic events (oods, tropical storms, droughts and the like) has risen rapidly over the past three decades. This increasing vulnerability is not a result of climate change, but arises from population and economic growth and the increasing occupation of hazardous and marginal lands. Present-day climate risks are not adequately taken into consideration in development activities. Burton and Van Aalst (1999), in a World Bank report, describe the

need to take adaptation to climate change and variability into account in Bank lending and other activities; see also Burton *et al.* (1993), which deals mainly with human adjustments to natural hazards, and discusses the value of adjustments in relationship to climate change. For this reason, some specialists refer to *maladaptation* to indicate those practices that result in increased vulnerability to present-day climate, and by inference to projected future climate change. In this connection, Burton (1996) discusses the concept of adaptive capacity.

Decisions and policies directed at adaptation to climate change are not taken in isolation from many other changing circumstances. Thus a farmer is concerned not only with present and future climate but also with changes in market prices for farm products, changing government policies on price supports and crop insurance, new technologies, and many other individual and local factors. Similarly with other sectors in uenced by climate and projected climate change, including forestry and water resources, coastal zones, public health and so forth, decisions about response to climate are made in the context of other rapidly changing circumstances. In this mix of factors, climate is currently often not an important or primary consideration. At all stages of adaptation, therefore, consideration has to be given to the context in which adaptation occurs. At its broadest level this has been interpreted to mean adaptation in the context of sustainable development (Intergovernmental Panel on Climate Change, Third Assessment Report, 2001; see especially Chapter 18, Adaptation to Climate Change in the Context of Sustainable Development and Equity').

CLIMATE CHANGE AND SUSTAINABLE DEVELOPMENT

A broad conceptual approach to climate and sustainable development is shown in Figure 1 (Burton, 1997). The climate cycle is represented by eight ovals, and points of human intervention are represented by seven rectangles. Human impacts on the environment (I), are defined in terms of population (P), level of consumption (or af uence, A) and technology (T) in the well known I • PAT formula (*see* DPSIR (Driving Forces Pressures State Impacts Responses), Volume 4). These factors give rise to greenhouse gas emissions

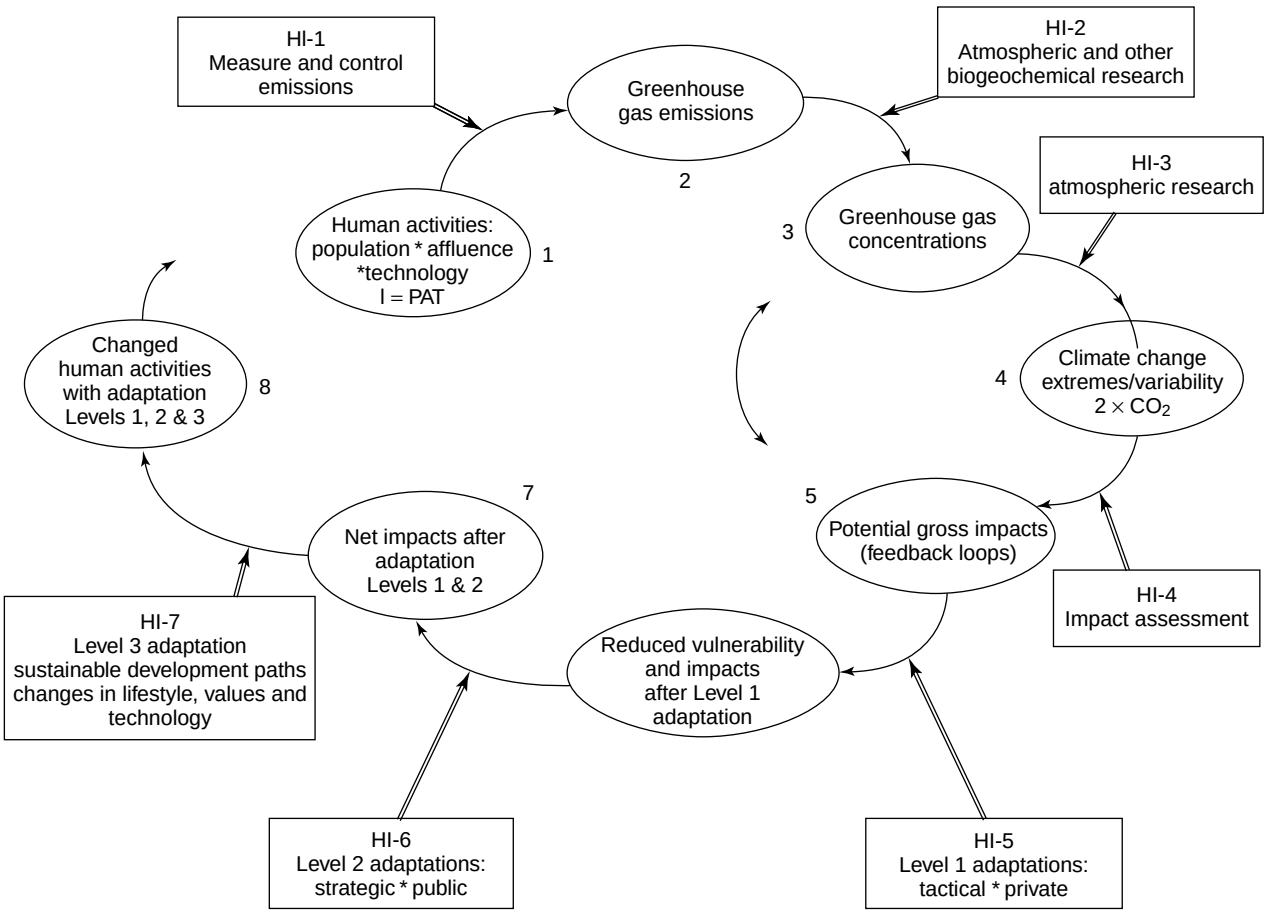


Figure 1 The climate cycle, adaptation and sustainable development; HI, human intervention (reproduced with permission from Burton, 1996)

(oval 2), and in turn to greenhouse gas concentrations (oval 3), and then climate change including variability and extremes (oval 4). Interventions in the cycle include the measurement and control of emissions (box 1), and atmospheric, biochemical, and other research (boxes 2 and 3). Climate changes give rise to impacts (oval 5) and these impacts can have feedback effects, such as the release of methane (a greenhouse gas) in Arctic regions due to permafrost melting. Research on impacts (box 4) is carried out to define the gross or primary consequences of climate change. Three types or levels of adaptation can reduce vulnerability. The first (box 5) is spontaneous or tactical adaptation carried out largely by individuals, communities and the private sector. The second type of adaptation (box 6) refers to those measures and policies that involve specific governmental action, often of a more strategic kind. The collection of such activities can be described as adaptation strategies. Adaptation strategies can fail, or produce contrary or counter-intuitive results if they do not correctly take into account the spontaneous adaptations included in box 5.

Adaptation strategies also contribute to the third type of adaptation, which includes many potential actions and policies directed towards sustainable development such as changes in technology, lifestyle choices, and human values and priorities (box 7). Thus the accumulation of adaptation measures at these three levels changes the basic I • PAT formula giving rise to climate change in the first place. Adaptation to climate change is therefore a potentially important element in the movement towards the sustainability transition (National Academy of Sciences, 2000, an assessment of the present global environmental circumstances of humanity and the steps needed to achieve the sustainability transition).

REFERENCES

Burton, I (1996) The Growth of Adaptation Capacity: Practice and Policy, in *Adapting to Climate Change: an International Perspective*, eds J B Smith, N Bhatti, G V Menzhulin, R Benioff, M Campos, B Jallow, F Rijsberman, M I Budyko, and R K Dixon, Springer, New York, 55–67.

Burton, I (1997) The Evolution of Adaptation in Response to Climate Change, in *Biometeorology, Part 2, Vol. 1, Proceedings of the 14th International Congress of Biometeorology*, September 1–8, 1996, Ljubljana, Slovenia, eds A Hočevar, Z Črepinšek, and L Kajfež-Bogataj, Ljubljana, Slovenian, The International Society of Biometeorology, 44–52.

Burton, I and van Aalst, M (1999) *Come Hell or High Water - Integrating Climate Change Vulnerability and Adaptation into Bank Work*, Washington, DC, The World Bank Environment Department, 1–60.

Burton, I, Kates, R W, and White, G F (1993) *The Environment as Hazard*, 2nd edition, Guilford Press, New York, 1–290.

Bwango, A, Wright, J, Elias, C, and Burton, I (2000) Reconciling National and Global Priorities in Adaptation to Climate Change: an Illustration from Uganda, *Environ. Monit. Assess.*, 61(1), 145–159, this paper outlines the potential con icts between national and global priorities, and illustrates the development of an adaptation strategy for Uganda. This is a special issue of the journal devoted to adaptive management.

Dotto, L (1999) *Storm Warning: Gambling with the Climate of our Planet*, Doubleday Canada Limited, Toronto, 1–332.

Dubos, R (1965) *Man Adapting*, Yale University Press, New Haven.

Dobzhansky, T (1962) *Mankind Evolving*, Yale University Press, New Haven.

Intergovernmental Panel on Climate Change (1996) *Climate Change 1995: Impacts, Adaptations, and Mitigation of Climate Change: Scienti c-Technical Analysis*, Contribution of Working Group II to the Second Assessment Report, Cambridge University Press, Cambridge.

Intergovernmental Panel on Climate Change (2001) *Climate Change: Impacts, Adaptation and Vulnerability*, Third Assessment Report, Working Group II, Cambridge University Press, Cambridge.

Lamb, H H (1982) *Climate, History and the Modern World*, Methuen, New York, 387.

National Academy of Sciences (2000) *Our Common Journey: a Transition Towards Sustainability*, National Academy Press, Washington, DC.

Smit, B, Burton, I, Klein, R J T, and Street, R (1999) The Science of Adaptation: a Framework for Assessment, *Mitigation Adap. Strat. Global Change*, 4(3/4), 199–213.

Smit, B, Burton, I, Klein, R J T, and Wandel, J (2000) An Anatomy of Adaptation to Climate Change and Variability, *Clim. Change*, 45(1), 223–251.

White, G F (1945) *Human Adjustment to Floods: a Geographical Approach to the Flood problem in the United States*, Research Paper No. 29, University of Chicago, Chicago, IL, 1–225.

Wilson, E O (1975) *Sociobiology: the New Synthesis*, Belknap Press of Harvard University Press, Cambridge, MA, 1–697.

Wilson, E O (1998) *Consilience. The Unity of Knowledge*, Alfred Knopf, New York.

Wright, R (1994) *The Moral Animal: Evolutionary Psychology and Everyday Life*, Vintage Books, Random House, New York, 1–467.

Sustainable Energy Policies

THOMAS B JOHANSSON

Bureau for Development Policy, United Nations Development Programme, New York, NY, USA

Sustainable development, a goal that the United Nations and its Member States have committed themselves to, implies meeting the needs of people alive today without compromising the ability of future generations to meet their needs. Since energy use is linked to poverty, opportunities for women, security, demographic trends, environmental degradation and economic growth, it can be a lever for improvements in all of these aspects of sustainable development. Access to affordable commercial energy dramatically increases the choices and opportunities available to individuals, which is one way of de ning human development. Indeed, thanks to the ready availability of affordable energy services, many individuals today enjoy a degree of comfort, mobility and capability unprecedented in human history. Energy services refer to the end bene ts of the energy system, such as illumination, mobility and comfortable temperatures and are also required for virtually all, commercial and industrial activities, from agriculture and industry to telecommunications. Meanwhile, two billion people (a third of humanity) nd their time, health and choices diminished by a reliance on traditional fuels for cooking and heating and no access to electricity. Meeting the world's growing demands for energy in conventional ways, however, would be a human and environmental disaster, leading to greater inequity, jeopardizing economic growth and threatening human and ecosystem health. But there are options for using energy in ways that simultaneously address the growing demand for energy services and the necessity of reducing the negative externalities of energy production and use. These options include: improving energy ef ciencies, increasing the use of renewable resources and accelerating the development and deployment of advanced technologies that allow conventional forms of energy to be used in environmentally benign ways.

The magnitude of the challenge is enormous. Energy policy must consider not only the two billion currently without access to commercial fuels and electricity, but also the new demands that will result from population growth; by 2050, another three billion people are likely to need commercial energy. Energy suppliers will need to decarbonize the energy systems radically to stabilize carbon dioxide (CO_2) concentrations in the atmosphere. Governments will need to work with the private sector to reduce local and regional pollution linked to energy use. Fortunately, as demonstrated by various energy scenarios in the literature, the physical and technical resources are available (or in the development stage) to achieve this. Whether they will be effectively used to meet the urgent challenges confronting us depends largely on the policy initiatives that will be taken over the next few decades. Power plants, re neries and other energy investments have long lifetimes. So if we do not make good decisions now, in another 20 years we will be locked in to those choices, and certain opportunities may be lost. Energy scenarios also show that the changes needed to achieve sustainable development are signi cant, and are unlikely to occur in the current policy environment.

The objectives of energy policies for sustainable development are to supply adequate and affordable energy services to support widespread economic growth and social development, and to do so in ways that curb environmental degradation and health-damaging pollutants. Energy used in ways that support human development in all its social, economic and environmental dimensions is what is meant here by *sustainable energy*. Key strategies to achieve this include:

- widen access to affordable and reliable energy supplies;
- make markets work more effectively to support sustainability, through accurate price signals and regulatory approaches;
- mobilize investments in sustainable energy;
- increase energy efficiency in all energy conversions, especially at the point of end use, where considerable cost-effective potential for improvement still exists;

- encourage the use of renewable energy resources (modern forms of biomass, solar photovoltaics, and wind and geothermal energy seem particularly promising);
- accelerate progress of energy technologies (energy efficiency, renewables and conventional fuels) along the energy innovation chain, from research and development (R&D) through deployment and diffusion;
- support capacity building and technological innovation and leadership in the developing countries, where most of the growth in energy demand is likely to occur.

While these strategies can be applied across the board, the specific policy instruments for implementing them may vary with the vastly different situations found among and within different countries. Different social circumstances, economic conditions, resource endowments and cultural factors need to be considered in designing appropriate energy policies.

The approaches discussed here mainly require action at the national level. However, with the growing integration of world markets, harmonization of policies at the international level (for example, energy efficiency standards or environmental taxes) could increase their effectiveness. Globalization, and the growing power of multinational corporations, is one example of the larger trends that policy makers need to consider in designing energy policies. Others include:

- restructuring and liberalization of energy markets, which would bring more competition and private activity to the energy sector, and would change government's role in this arena;
- the emerging information revolution, with its potentially enormous but difficult-to-predict structural consequences;
- greater public participation in decision-making, which is likely to keep equity issues high on the political agenda.

During the last three decades, several global studies have employed scenarios to try to understand the role that energy systems will have in determining our future. Whereas some of the earlier scenarios were prompted by fears of oil shortages and price volatility, concerns have broadened. Now it is clear that ample fossil fuel resources, at reasonable prices, will be available to meet the world's growing demand for energy services, through this century at least. Thus, there will not be a resource-driven change in the world's energy systems for a long time to come.

Two other critical energy-linked challenges are proving to be more problematic: addressing vast inequities in opportunities for different groups of people (especially as these relate to poverty, health and women's lives) and maintaining ecological balance. Both must be addressed in order to achieve sustainable development.

Scenarios (*see* Scenarios, Volume 5), which are dynamic models that analyze the interactions of systems under certain sets of assumptions, suggest that it is possible to simultaneously expand energy services (and redress energy inequities) while limiting environmental degradation associated with energy use. Two scenarios developed in 1998 by the International Institute of Applied Systems Analysis and the World Energy Council (a business-as-usual scenario, and one that is more ecologically driven) compare

Table 1 Characteristics of sustainability in two energy development scenarios in the second half of the 21st century compared with 1990

Indicator of sustainability	1990	Business as usual[a]	Ecologically driven[b]
Eradicating poverty	Low	Medium	Very high
Reducing relative income gaps	Low	Medium	Very high
Providing universal access to energy	Low	High	Very high
Increasing affordability of energy	Low	Medium	Very high
Reducing adverse health impacts	Medium	High	Very high
Reducing air pollution	Medium	High	Very high
Limiting long-lived radionuclides	Medium	Very low	High
Limiting toxic materials[c]	Medium	Low	High
Limiting GHG emissions	Low	Low	Very high
Raising indigenous energy use	Medium	Low	Very high
Improving supply efficiency	Medium	High	Very high
Increasing end use efficiency	Low	Medium	Very high
Accelerating technology diffusion	Low	Medium	Medium

Source: UNDP *et al.*, 2000.
[a] Corresponds to B scenario in the World Energy Assessment.
[b] Corresponds to C1 scenario in World Energy Assessment.
[c] For this row only, the qualitative indicators are not based on quantitative features of the scenarios, but were specified by the author on the basis of additional assumptions.

key environmental, economic and energy indicators as they unfold over this century given different assumptions about how energy systems will develop. Both hold population projections constant, and achieve the goal of sustained economic growth. But the other effects, as they play out over time, were substantial. The scenario exercise suggested that continuing in the present direction will result in adverse health impacts, an environmental burden of toxic materials, and increasing greenhouse gas emissions, without fully addressing the current inequities in the distribution of benefits from energy use.

The more ecologically driven scenario achieves a substantial reduction in atmospheric carbon concentrations by the end of the century, as well as considerably less acidification (as re ected by reduced sulfur dioxide (SO_2) emission levels). Achieving these long-term consequences would require ambitious policy measures, including carbon and environmental taxes, to curb environmental damage. This path also demands continued improvements in energy efficiency, a greater reliance on renewable energy sources and a high level of technological innovation, which allows for the use of fossil fuels in non-polluting ways. The ecologically driven scenario, listed in Table 1 assumes a nuclear phase-out by the end of this century. However, nuclear power could be part of the energy supply mix, if the key problems associated with nuclear energy (cost, safety, waste disposal and weapons proliferation) could be adequately resolved. Table 1 compares the implications of the two scenarios from the standpoint of sustainability. The remainder of this article discusses how we can get beyond business-as-usual, and head toward a sound future.

WIDENING ACCESS

A pre-requisite to a sustainable future is greater equity in the distribution of energy services (and therefore, more equality in the opportunities afforded to individuals). Yet despite tremendous efforts to improve energy services to rural populations in the last 20 to 30 years, the population without access to electricity has remained approximately the same in absolute numbers, around two billion people. Approximately the same number rely on traditional fuels (dung, crop residues, and firewood) for cooking and heating. An effective strategy to address the energy needs of the rural populations is to promote the climbing of the energy ladder. This implies moving from the solid biomass fuels to the most convenient and efficient form of energy appropriate to the task at hand, usually liquid or gaseous fuels (coupled with more efficient stoves) for cooking and heating, and electricity for most other uses (including commercial and industrial activities). As new, cleaner technologies become available, it is also possible to introduce new rungs on the energy ladder, and gain even greater efficiencies and environmental acceptability.

Over the long term, using energy in ways that produce near zero harmful emissions will be needed. However, in the near term, small amounts of conventional forms of commercial energy could greatly improve living conditions in rural areas. Meeting the cooking needs of two billion people with conventional energy carriers (such as liquefied petroleum gas), would require an amount equivalent to less than 1% of current commercial energy consumption worldwide, but would constitute a much more efficient use of energy (by about a factor of three) compared to traditional fuels. This means that if the world's poorest people were to move to liquid fuels, their primary energy use (and the resulting health and environmental impacts) would initially decrease, with no corresponding decrease in energy services.

In rural areas, small-scale decentralized approaches may be more appropriate and cost effective than extending energy grids. Successful implementation of such small-scale technologies seems to depend on the active participation and empowerment of the people who use them. Cooperatives, non-governmental organizations and local groups can effectively support delivery of energy services and management of natural resources.

Affordability is critical. However, because the rural poor already pay a high price for the limited energy services they receive, they are often willing and able to pay more than one might expect for energy services if appropriate financing options are offered. The economics of providing basic electricity to rural households should be evaluated according to the costs of supplying comparable energy services through less efficient carriers. Supportive financial and credit arrangements (including micro-credit arrangements of the kind now in existence) will be needed to introduce commercial energy to those excluded from markets, especially in rural areas.

Because commercial energy can be a necessary lever for breaking out of a cycle of poverty, temporary subsidies may be effectively employed, especially within a context of integrated development initiatives. Such subsidies could be looked at as part of a new social contract, in which competitive conditions in the energy sector (a key element of ongoing energy sector reforms) are complemented by the establishment of a public benefits fund to protect various public goods that the market would not. Funds could be generated by small, but non-bypassable wires/pipes charges on electricity and gas providers, and could be used to widen access and support technological innovation.

Such a strategy could be employed while still taking advantage of the forces of competition for efficient allocation of resources. Consider, for example, the rural energy concession, an arrangement in which a substantial market is awarded to a supplier, along with an obligation to serve

all customers within a specified market. If such a concession were awarded competitively, market forces would be brought into play to find the least costly mix of energy technologies. While some subsidization might still be necessary, the competitive process would help ensure that the public funds provided the minimum subsidy required to satisfy the concessionaire's obligation to provide modern energy to all.

Policy approaches to address the challenge of supplying modern energy to rural areas include:

- decentralized rural electrification, using wind turbines, solar photovoltaics or diesel generators, where grid extensions are inefficient;
- public participation in planning rural energy systems;
- innovative credit schemes such as locally managed micro-credit arrangements, are a good example, especially to help consumers meet the hurdle of high initial costs associated with moving up the energy ladder;
- temporary subsidies to help promising energy technologies overcome barriers to commercialization;
- special rates or temporary subsidies to provide enough energy to support basic human needs (possibly supported by a wires charge);
- concessions that offer a significant customer base in return for energy investments in underserved areas;
- bilateral and multi-lateral support for innovative approaches to sustainable energy;
- a social contract, that couples highly competitive markets with an obligation to protect public benefits.

MAKING MARKETS WORK BETTER: PRICING AND REGULATORY APPROACHES

Driven by the force of competition, markets are unsurpassed in their ability to allocate resources efficiently and drive costs down. For that reason, many of the approaches suggested here attempt to harness the power of the marketplace, guiding it toward the achievement of sustainable development goals. Markets never work perfectly however. The effectiveness of energy markets is often hampered by price distortions and lack of competition. Market imperfections also result in inadequate investment levels, low levels of energy efficiency, and lack of support for energy innovation, subjects to be addressed in subsequent sections.

Removing Subsidies

The first obstacle to fair competition in energy markets is the significant subsidies for conventional energy in many countries. These publicly supported cost reductions, which may be granted to producers and consumers (directly, through price reductions, or in less visible forms, through tax breaks, market support, or inadequate metering) often fail to help the most needy and discourage least-cost choices and efficient energy use. In addition to working at cross-purposes to the aims of sustainable development, energy subsidies (exclusive of the transport sector) represent public costs on the order of US$100–200 billion per year. Since these support mainly fossil energy, they create a huge hurdle for new technologies, renewables and energy efficiency improvements to overcome.

Policy approaches to level the playing field include:

- transparent accounting procedures to make subsidies visible;
- phasing out of permanent subsidies and cross-subsidies to conventional energy (fossil fuels, large hydropower and nuclear energy).

An important exception to the general rule of removing subsidies is lifeline rates, that is, the provision of small amounts (perhaps 50 kilowatt hours per household per month) of commercial energy supplies, at little or no cost, to help vulnerable groups meet basic needs or to break out of a cycle of poverty. Where such subsidies are well targeted, and well designed (to encourage use, but not inefficiency), the resulting benefits, including access to the outside world through radio and television, electric light for reading and studying, time savings, health benefits and so on, outweigh the costs.

Internalizing Externalities

A second obstacle to a level playing field is that the market, in the absence of government intervention, fails to account for social and environmental costs of energy production and use, which are called *negative externalities*. Prices that do not re ect these costs tilt the playing field to the advantage of fossil fuels at the expense of societal aims (and make it difficult for more sustainable forms of energy to compete). Finding ways to accurately figure externalities into the pricing equation proves to be difficult, however, partially because there is no consensus on how to measure their costs in monetary terms. These external costs are, however, estimated to be quite substantial – in some cases comparable to, or in excess of, the cost of the primary energy itself.

Nevertheless, a number of interesting policy initiatives are attempting to integrate externalities into prices at the national level, through gasoline taxes and other means. In many European countries, taxes on gasoline are well above the base cost of the fuel itself. Differential fuel taxes can be seen as another way to start to account for externalities. They have been adopted in a number of countries to support ethanol fuel or compressed natural gas use, both of which reduce oil import dependence and improve urban air quality. Generally speaking, targeted taxes that penalize the use of environmentally harmful forms of energy are

more effective than a general energy tax. From a theoretical point of view, carbon taxes are the simplest and most consistent method for internalizing the cost of mitigating climate change. A mix of energy and carbon taxes would address a wider range of environmental impacts associated with energy production and use.

From a national perspective, internalizing environmental costs can undermine the competitive position of domestic industries. That's why, in a world where markets are becoming more integrated, international cooperation to harmonize environmental initiatives could be an important step toward a more sustainable future.

Fine tuning taxes on energy use so as to more accurately re ect the environmental impact and other externalities of use can encourage energy efficiency and the use of cleaner forms of energy. An additional dividend is that the chain of cause and effect between energy production and use and resulting externalities becomes clearer and the polluter-pays principle better established (Jefferson, 2000). Such taxes can be used to offset other taxes or to support many of the measures discussed here, particularly technological innovation and wider access.

Policy options to incorporate external costs into prices include:

- differential fuel taxes;
- environmental taxes;
- environmental regulations (which have associated costs);
- carbon taxes;
- emissions trading.

Regulatory Approaches

Throughout the 20th century, grid-based energy systems were looked at as natural monopolies. But new technical developments, such as the smaller scale of new power plants, the drive for greater efficiency, and the wave of liberalization are creating new system models. Increasingly, various functions of these former monopolies have been unbundled, that is, the production, transmission and distribution functions are no longer vertically integrated. Rather, they are operated by independent entities that compete in the arenas of generation and distribution. This has opened up the gas and electricity sectors to increased competition, including that from independent power producers (IPPs). Although in some circumstances this diminished role for government in supplying and distributing energy could be seen as a threat to public welfare, it also presents an opportunity for a different kind of government participation. Indeed, implementing regulatory measures to protect societal objectives and maintain fair competition may be a more appropriate and effective role for government.

In the process of liberalization, corporations are created that act under the provisions of the legal, economic and regulatory systems of the country. Changes in these framework conditions often demand readjustments, with associated costs and changes in competitiveness, and are therefore often resisted. However, once a new framework is introduced to replace a monopoly situation, the process of change has already begun. This transitional period, while the rules are in ux, often presents a timely opening in which to introduce and negotiate provisions that address social and environmental concerns. Because of ongoing market reforms, many countries are now in a position to take advantage of this window of opportunity.

Regulatory approaches to protect public benefits include:

- measures to widen competition, for example, by guaranteeing IPPs access to power grids and giving energy service companies opportunities to bid on supply contracts;
- creating a public benefits fund (see above) to support sustainable energy developments;
- allowing prices to re ect marginal costs of supply;
- obligations to serve specific regions in return for exclusive markets (concessionaire arrangements);
- obligations to buy (such as Renewable Portfolio Standards, that help bring down the cost of clean new technologies);
- energy efficiency and performance standards;
- reporting requirements to ensure transparency.

MOBILIZING INVESTMENTS IN SUSTAINABLE ENERGY

Since capital investment is a prerequisite for energy development, the lack of capital for energy investments can thwart economic development. A major challenge in raising funds for energy investments is the perceived risk to investors and adequate rates of return. Returns in the energy sector are often lower than those in other infrastructure investment alternatives. And within the energy sector, rates-of-return considerations favor large-scale, conventional projects over smaller-scale, clean, and innovative energy alternatives. Likewise, the scale and dispersed nature of investments in energy end-use efficiency do not attract capital under current models.

Scarce public sector funds, especially in developing countries, are sought by many projects besides energy supplies including rural development, education and health care. Because energy supply, more than any other alternative, is often seen as capable of generating revenues early on, the energy sector is increasingly viewed as suitable for private sector investment. But private capital is not owing into many countries at an adequate rate to support energy sector development.

Foreign direct investment (FDI) in energy projects has been estimated at about 5–15% of the total, which amounted to some $400 billion in 1997. While FDI is an important source of financing for energy projects, it is generally commercially motivated. The sponsor of investments not only expects to recover the initial capital but also counts on competitive returns. Where political systems and markets are unstable, or where the rules are changing, risks are high. The 47 least developed countries receive little or no FDI. To a large extent, foreign investors lack confidence in the ability of developing country energy projects to provide stable (and competitive) returns until the investment has been recovered.

Official development assistance (ODA), in contrast, is publicly supported aid delivered through grants. While the level of ODA increased from US$34 billion in 1984 to US$56 billion in 1997 or 0.25% of the Organization for Economic Cooperation and Development (OECD) gross domestic product (GDP), it is still far below the 0.7% target that has been agreed to by the industrialized countries.

Policy options to encourage investments in sustainable energy include:

- encouraging good governance to create conditions more attractive to investors;
- creating multi-laterally backed insurance to protect energy sector investments;
- added incentives to invest in sustainable energy, such as the *Clean Development Mechanism*, whereby industrialized countries may be able to earn credits for investing in carbon abating projects in the developing world (*see* Clean Development Mechanism, Volume 4);
- increasing ODA to agreed-upon levels and tying some of it to capacity building and investments in sustainable energy.

IMPROVING ENERGY END-USE EFFICIENCIES

If energy resources with zero or near-zero emissions were widely available and affordable, the need for energy efficiency and conservation would not be so urgent. But conventional fossil fuels are likely to dominate world energy use for the next half century. In the absence of clean energy technologies, energy efficiency is crucial to meeting the need for increased energy services while reducing environmental and health burdens. In theory, energy efficiency improvements of an order of magnitude are possible (based on analysis using the second law of thermodynamics). Whereas increasing energy efficiency in generation and distribution tends to be driven by the market, improvements in end-use efficiency are more difficult to realize. This is true in part because such improvements represent decentralized activities that require wide participation. Because the substantial potential benefits of energy efficiency measures are incremental and dispersed, it is not an easy issue to mobilize support for.

From a societal perspective, achieving greater end-use energy efficiency generally requires less investment than new generation (freeing up investment for other purposes), and it means fewer energy-related environmental externalities to deal with. From an individual perspective, it can mean significant savings over the long-term. However, many obstacles prevent energy efficiency from reaching its economic potential, including:

- lack of information, technical knowledge and training;
- uncertainties with regard to the performance of investments in new and energy efficient technologies;
- lack of capital or financing possibilities;
- high initial and perceived costs of more efficient technologies;
- high transaction costs (for searching and assessing information and for training);
- lack of incentives for careful maintenance;
- differential benefits to the user compared to the investor (for example, when monthly energy bills are paid by the renter rather than the property owner);
- consumer patterns and habits.

A number of the policy initiatives to address these barriers are relatively simple and effective. Correct pricing signals, as discussed earlier, can make energy efficiency investments more attractive. Additional policy options to overcome barriers include:

- labeling schemes (tied to testing programs) to better inform consumers;
- procurement policies to maximize economies of scale, buy down costs, and reduce transaction costs;
- voluntary or mandatory efficiency standards for appliances, vehicles and buildings;
- technical training in new energy efficiency technologies and their maintenance;
- credit mechanisms to encourage customers to make energy efficiency investments;
- greater reliance on energy companies to provide energy services, which may also include end-use technologies, rather than simply energy supplies.

This second point can be illustrated by a technology procurement policy that was successfully introduced by the Swedish energy agency (NUTEK) in the late 1980s. NUTEK convened a group of public and private companies who represented a large share of the appliance market, and issued a bid for more energy efficient, freon free refrigerators. The winning model had energy requirements that were one-third less than the most efficient model then on the market.

DEVELOPING AND ENCOURAGING DIFFUSION OF SUSTAINABLE ENERGY TECHNOLOGIES

Current technologies are not adequate to deliver the energy services that will be needed in the long-term and simultaneously protect human health and the environment. Thus, a prerequisite for the achievement of a sound energy future is accelerated progress for new technologies along the energy innovation chain, from R&D, to demonstration, deployment and diffusion (Table 2). This process can take decades and be stalled at any point along the way.

Over the long term, radical changes in energy technology will be needed to support sustainable development aims. During this new century, some combination of the following technologies, which offer high levels of inherent cleanliness and safety, will be needed:

Table 2 The energy innovation value chain

Stage of effort	Laboratory	Pilot to market	Achieving lower costs per unit	Widespread diffusion
Activity	R&D	Demonstration projects	Buy down of cost	Increasing investments in sustainable energy technologies
Key barriers	Government austerity makes allocation of funds for energy R&D problematic	Government austerity makes allocation of funds for energy R&D problematic	Financing incremental cost	Transaction costs
	Difficulty for private sector in capturing benefits	Difficulty for private sector in capturing benefits	Cost uncertainty	Prices for competing
		Risks	Technological and other risks	Technologies exclude externalities
		Large capital costs		Lack of retail finance
				Lack of information
Policy options to address specific barriers	Increased public funding (national and international)	National and international support for demonstration projects	Obligations to buy (such as Renewable Portfolio Standards, which guarantee markets for new technologies)	Obligations to buy
	Tax deductions for private R&D	Low cost or guaranteed loans	Obligations to supply (such as the Non Fossil Fuel Obligation in the UK)	Obligations to supply
	Collaborative research partnerships and funding	Project aggregation to reduce costs	Tax incentives	Green labeling and marketing
		Temporary price guarantees for products of demonstration projects	Large-scale procurement policies	Renewable energy concessions
			Temporary subsidies	Capacity building
				Clean development mechanism

Source: UNDP *et al.*, 2000.

- fuel cells for transport, stationary power and combined heat and power generation;
- fossil fuel derived hydrogen (H_2) with sequestration of the separated carbon dioxide (CO_2);
- modernized biomass technologies for production of liquid and gaseous fuels and electricity;
- photovoltaic and wind power technologies coupled to energy storage to provide dispatchable power;
- wide deployment of much more energy-efficient end-use technologies.

Fortunately, a wide range of promising advanced energy technological options can already (or may soon be able to) address energy needs in sustainable ways if adequate attention is given to energy innovation. Long lead times for new technology dictate that intensive energy innovative activity is needed today, to establish a portfolio of new energy technology options to choose from over the longer term.

Direct government support is more likely to be needed for radically new technologies than for incremental advances, which can generally find support from the private sector. Scarce public resources can be used more effectively to support emergent technologies offering major public benefits not fully valued in market prices, that have steep learning curves and good prospects for market penetration after the incentives or subsidies are phased out. Because of the risks involved, and the fact that clean energy does not necessarily command higher prices than conventional energy, targeted policy approaches are needed, both to encourage supply (push) and stimulate demand (pull) (Geller, 1999).

Research and Development (R&D)

While R&D is the first step in the energy innovation chain, both public and private sector expenditures on energy R&D have generally been falling. Energy R&D expenditures are low in relation to both the direct value of energy expenditures in the overall economy and in relation to the sustainable development challenges posed by conventional energy. Moreover, much of this money has been used to support nuclear energy technologies; in recent years, only about 10% goes toward energy efficiency and another 10% toward renewable energy technologies.

Several OECD economies are now moving away from direct state support of R&D toward a complex mix of incentives based on:

- regulatory requirements on private industry to develop technologies with low carbon emissions;
- tax incentives for the development of low carbon technologies that are neutral between technologies, such that the choice of which technologies to develop can be delegated to private initiative;
- marketable permit systems, such as the proposed (under the Kyoto Protocol) programs of Joint Implementation and the Clean Development Mechanism; and
- special financing facilities, such as the Global Environment Facility, which blend their own concessionaire finance with the hard finance of the Multilateral Banks and industry to achieve a softer financial blend for innovative environmental projects.

These are all clearly more market-oriented initiatives. However, the scale of existing incentives is insufficient in light of the emerging environmental problems: and the continuing competition from fossil fuels.

Moreover, the scope of energy R&D should be extended in order to address the particular energy needs of the developing world, which will account for most of the world's incremental energy requirements. New energy technologies should be tailored to the specific resource endowments, capabilities and cultures of individual countries.

Demonstration through Widespread Diffusion

Even after the viability of energy technologies has been established, decades of demonstration projects and market growth may be required before these new technologies can command major market shares. Demonstration projects, however, are generally costly, risky and difficult to finance. Although public funds may be crucial at this stage, one way of limiting risks is to bring in the private sector as well. Some of the most successful demonstration projects have been those where the government role has been to set performance and cost goals, and to provide some capital, while the private sector takes responsibility for finding ways to meet the goals, and also shoulders some of the financial risk.

Buying down Costs for New Technologies

Some public support or incentives may be needed to bring clean energy technologies to the point where they can compete with conventional energy. Economies of scale in production, learning and experience curves, and cost reductions through competition can all help drive the costs of new technologies down. Typically, manufactured goods show cost declines of about 20% for each cumulative doubling of production, and this pattern can be seen for a number of rapidly growing renewable technologies, including windmills, solar photovoltaics and gas turbines. But the market alone will not typically be able to overcome the higher initial costs of new energy technologies.

Temporary subsidies to promising, emergent technologies that offer major public benefits not captured in prices, can take various forms, including tax incentives, subsidized loans or guaranteed prices, government procurement and competitive market transformation initiatives.

The UK's Renewables Non Fossil Fuel Obligation and California's Renewal Portfolio Standards are examples of initiatives that encourage cost buy downs by stipulating that suppliers include a specified percentage of renewable energy in their portfolios. Green pricing, whereby customers can choose to pay higher prices for environmentally friendly energy, is a market-based mechanism to initially boost production, even at higher costs, and ultimately bring down prices.

FOSTERING TECHNOLOGICAL INNOVATION AND LEADERSHIP IN DEVELOPING COUNTRIES

In many respects, the achievement of sustainable development depends on what happens in the developing world, where energy demands are set to grow substantially in any scenario of economic success. However, thanks to technological developments, the developing world need not retrace the wasteful and polluting path taken by the industrialized world. Indeed, from the standpoint of their natural endowments (the higher incidence of solar insolation, for instance), nascent infrastructure and growing markets, developing countries are well positioned to profit, both socially and economically, from deploying clean energy technologies at an early stage of development.

Developing (and transitional) countries need to further develop their own resources (human, natural and technological) so that they can create energy systems appropriate to their own circumstances. But they also need assistance in terms of technology transfers, financing and capacity building. The declining share of ODA relative to other investment capital suggests that much of this support will need to be led by the private sector, and/or by private/public sector partnerships. International industrial collaboration offers one means by which the private sector could gain markets while fostering the migration of new technologies to developing countries. Public sector agencies, private research institutes and regional institutes that provide training in technological management are additional possibilities to further technology sharing and capacity building.

Capacity and Institution Building

If developing countries are to meet their energy needs, they will require more human capital and stronger institutions. Technological and managerial capabilities are critical for successful technology transfer and innovation. Human capital will be needed to form and staff public sector agencies and research institutes and to create and staff the private companies that will produce, market and install energy efficiency, renewable energy, and clean fossil fuel technologies. Without strong private sector actors and markets as well as sustainable energy programs and agencies, sustainable energy initiatives in developing countries are likely to be *ad hoc*, limited in scope, and insufficient to overcome the full set of barriers inhibiting large-scale implementation.

Strong government/private sector collaboration has been a key characteristic of many successful market development programs. This means involving the private sector in program design as well as implementation. Successful national energy efficiency and renewable energy centers and programs are already operating in Brazil, China, India, and Eastern Europe. These centers and programs work with the private sector, utilities, and other entities to introduce and widely disseminate energy efficiency and renewable energy measures and to:

1. provide education and training;
2. conduct R&D and demonstrations;
3. adopt and implement efficiency standards;
4. perform product testing, labeling, and certification;
5. conduct marketing and promotion;
6. encourage local manufacturing of energy efficiency and renewable energy technologies;
7. help establish and support energy service companies;
8. provide financing and/or financial incentives;
9. develop and analyze policy reforms.

Taking the Lead

Although the principle of technological *leapfrogging* (*see* Leapfrogging Technology, Volume 4) is widely accepted, conventional wisdom cautions against developing countries taking the lead in the deployment of emerging energy technologies, and assuming the risks involved in commercializing technologies and systems not widely in use elsewhere. In fact, the development assistance community has not typically supported innovation in the energy sectors. But there are several reasons to question this view.

First, inherently cleaner energy resources would mean that developing countries could effectively address their growing air quality and other environmental challenges with less regulatory infrastructure than would be required if these goals were pursued by mandating and enforcing the use of increasingly more stringent end-of-pipe control technologies.

Also, rapidly growing markets, with high demand for energy-intensive industries, like those found in many developing countries, offer favorable conditions for energy technology innovation. In developing countries, large, rapidly growing markets, plus lower wage rates at all levels of technical, scientific and engineering professions make both manufacturing and product development attractive options for technology owners. When new technologies are introduced in developing countries, some components could

be produced locally. Local manufacture implies not only local job-creation opportunities but also potentially lower costs than for the same technologies imported from the industrialized world, leading to large domestic markets and opportunities for export growth.

Taking advantage of the current window of opportunity for market leadership will be a challenging task. Three prerequisites for successful outcomes are the establishment of an appropriate legal, regulatory, and policy framework; the rational pricing of energy; and the creation of strong domestic capital markets.

Policy options to encourage technological innovation in developing countries (many of which could benefit from the support of bilateral and multilateral financing and development cooperation agencies) include:

- R&D and demonstration projects emphasizing clean energy supply and energy end-use technology innovation;
- encouraging the introduction of new energy technologies through international joint ventures and other mechanisms;
- adopting and enforcing strong energy efficiency and environmental standards in developing countries so that new infrastructure is state of the art (and so that polluting technologies do not get locked in);
- attractive financing, market development assistance, and government procurement for products made by clean energy technology entrepreneurs whether these companies are locally owned, international, or joint venture.

Because of the long lifetimes of power plants, refineries and other energy investments, even vigorous policy measures will not produce major changes in energy systems and emissions before 2020. But the seeds of the post-2020 world will have been sown by then. Thus, choices about the world's future energy systems are relatively wide open now. This window of opportunity is particularly significant for those developing regions where much of the infrastructure has yet to be installed, offering the possibility of a very rapid introduction of new, environmentally sound technologies. Once infrastructures are in place and rates of economic growth decline, a phase of largely replacement investments will begin. In this phase, changes can be made, but they will take much longer to affect the average system performance. If we do not make good decisions now, by around 2020 we will be locked in to those choices, and certain development opportunities might not be achievable later. Thus, sustainable development dictates a global perspective, a very long time horizon and an early initiation of policy measures (Nakičenovič, 2000).

A key challenge for human society now is to move beyond the tangible pressures experienced today and to manage global resources with future generations in mind. Taking such actions, in the face of competing short-term interests, will essentially require a paradigm shift. For this shift to occur, the sustainability debate will need to move to center stage, and be accompanied by greatly raised levels of public awareness, information and commitment. Continued dialogue and consensus building within the international community, between the public and private sectors, and within society at large is needed to advance sustainable energy policies at the national, regional and global levels.

ACKNOWLEDGMENTS

This paper draws heavily on material prepared by convening lead authors (and their respective teams of experts) of the World Energy Assessment, especially Dennis Anderson, José Goldemberg, Michael Jefferson, Eberhard Jochem, Nebojsa Nakićenović, Amulya K N Reddy, Hans-Holger Rogner, Kirk Smith, Wim C Turkenburg and Robert H Williams. Janet Jensen provided excellent editorial assistance.

FURTHER READING

Nakičenovič, N, Grubler, A, and McDonald, A (1998) *Global Energy Perspectives*, Cambridge University Press, Cambridge.

PCAST (1999) *Powerful Partnerships: The Federal Role in International Cooperation on Energy Innovation*, President's Committee of Advisors on Science and Technology, Panel on International Cooperation in Energy Research, Development, Demonstration and Deployment, Office of Science and Technology Policy, Washington, DC.

Reddy, A K N (1999) Rural Energy: Goals, Strategies and Policies, *Economic and Political Weekly*, 34(4), 3435–3445.

Reddy, A K N, Johansson, T B, and Williams, R H (1997) *Energy After Rio: Prospects and Challenges*, United Nations Development Programme, New York.

UNDP (1991) *Human Development Report*, Oxford University Press, New York.

United Nations Development Programme, United Nations Division of Economic and Social Affairs, and World Energy Council (2000) *World Energy Assessment: Energy and the Challenge of Sustainability*, ed J Goldemberg, United Nations Development Programme, New York.

World Bank (1994) *World Development Report*, World Bank, Washington, DC.

World Bank (1996) *Rural Energy and Development: Improving Energy Supplies for Two Billion People*, World Bank, Washington, DC.

Dematerialization and Sustainable Development

MOSTAFA K TOLBA AND JIH C YANG

International Center for Environment and Development, Cairo, Egypt

When the term sustainable development was coined more than 15 years ago, it clearly meant that the patterns of development being pursued were not sustainable. Sooner or later, we had to refocus our attention and have serious plans for achieving sustainable development. Agenda 21, adopted at the 1992 Rio Summit (the United Nations Conference on Environment and Development), charted the general path in this direction, but we need detailed plans. Recognizing that the root cause of our sustainability problems is the persistent imbalance between the growth and the dematerialization rates of our economic activities, it follows that the solution lies in accelerating dematerialization, so that it may catch up with growth and nally put a stop to growth's runaway demands on the Earth's natural resources. Dematerialization denotes acts that reduce the consumption of materials (energy, water, land, forests, minerals, etc.) in each unit of economic output. The issue of dematerialization can be reduced to achieving a balance as given by a simple equation: dematerialization rate = growth rate.

From this simple relationship, we may begin to explore the challenges we face, and comprehend the actions we must take, however dif cult they may be. Extending this simple logic, we may also grasp, quantitatively, the size of one of sustainable development's most complicated issues - the North/South dispute. Simple calculation would show that just one percentage point of extra dematerialization performance by the developed world (the North) would create enough room under the overall sustainability cap for the developing world (the South) to grow essentially unrestrained out of their under-development and unsustainable ways.

Dematerialization is the key to our sustainability struggles, and traditional approaches such as ef ciency improvements and cleaner production can only go so far. Ef ciency (or cleanliness) is neither easy nor, as a resource, limitless. If it has to progress as fast as growth, it will not be able to keep up for long. For dematerialization to have a realistic chance to catch up with growth, growth must itself become a lot less material dependent. The outcome will hinge upon the ways in which society progresses in the information age. Will it make growth ever more material-less? Or will paradigms such as E-commerce accelerate sales and consumption of the world's natural resources? Such are the crucial questions upon which our sustainability fates hang.

REFOCUSING SUSTAINABLE DEVELOPMENT

If sustainable development is going to have a meaningful future, our views on how to achieve that desirable state will have to change. When people talk about the subject nowadays, they speak most of the time about what we shouldn't do, sometimes about what we should do, but almost never about how to do it. Even when the subject of actions is raised, suggestions tend to be disparate, anecdotal and lacking the structure and rigor that characterize the action plans of serious enterprises. Sustainable development has been around too long to be still without a coherent plan or even a compelling scenario. Can anything so important be talked about without a blueprint?

Implementation of sustainable development will be, by all measures, one of the toughest actions we will ever take, and so far we are not exactly overloaded with ideas. This however is no reason to avoid the subject. If it is indeed so critical to the future of mankind, then we have no choice but to face it head-on. What follows herein is an attempt to outline the how-to-do-it aspects of sustainable development.

SUSTAINABLE DEVELOPMENT'S OBJECTIVE IS TO CHANGE DEVELOPMENT

Curiously, few people nowadays seem to remember that sustainable development began as a movement to change development. Just look at the definition of sustainable

development given by the World Commission on Environment and Development (WCED) in 1987: sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their own needs .

The people who wrote this didn't waste time getting to the point. They didn't mention pollutants, wastes, toxics, energy use and its effects on climate, losses of habitats and biodiversity, fresh water shortages or an assortment of other well-known sustainability issues, which are only symptoms. It's the causes that WCED was worried about, and the causes are much more troublesome. They are the most instinctive and normal things we humans do – the pursuit of development.

Too Much Growth is a Threat to Sustainability

Over the last 100 years, economic activities of mankind have grown 80-fold, and this growth has increased our demand and consumption of natural resources many, many fold. If we keep this up for the next 100 years, our economic activities will expand 80 folds more. Today's environment is under pressure, can we imagine what would happen then. That is only for the next 100 years, but we humans expect to be around for a lot longer than that. It's simple arithmetic. We can't go on the way we are going now.

Too Little Growth is a Threat to Sustainability

If too much growth is the primary reason for our sustainability problems, then too little growth wouldn't be far behind. Today, a full one-third of all humans live on the borderline of abject destitution, and their ways of life are not only humanistically unacceptable but also among the most environmentally damaging. Sustainable development for them is to grow, and to do so in an unrestrained manner for as long as it takes to reach an acceptable level of subsistence. That means the developed world must, in addition to taking care of the sustainability consequences of its own growth, go the extra distance to make room for the developing world to grow out of its underdevelopment miseries as fast as they can.

Even the Transition is a Threat to Sustainability

Not only are too much growth and too little growth the main causes of our sustainability problems, even the transition from developing to developed is problematic. When developing countries develop, they will most likely do so with not only the most labor-intensive but also the most material-intensive kinds of economic activities. High value-added, innovation-driven economic activities, low in energy and material intensities are the forbidden games only the most technically advanced countries can play. Can we blame the developing countries for undertaking material-intensive actions until they finally manage to develop? What are their options than to stay underdeveloped? They do not have other options. To stay underdeveloped is not an option.

CAN GROWTH BE STOPPED?

There are, theoretically, two kinds of things we could do to rein in runaway growth. One is to stop growth, and the other is to change the way we grow so the accumulation of environmental damages will stop. Can growth be stopped? The answer is no. There are several reasons.

1. Growth has been an integral part of human development for a very long time. It is probably a little late to try to persuade humans to give up growth now.
2. Sustainability problems are long-term. They are only painful and jolting enough for the majority of humans to want somebody else to mend their ways, not themselves.
3. There simply must be room for growth for the developing world.

Thus, making development sustainable is our only real option. The WCED statement, for example, called for development to be changed rather than for growth to be stopped. This view, however, has become more and more blurred as sustainable development becomes more and more popularized. There are even rumbles within the environmental community itself that the term sustainable development be replaced by sustainability because the word development is so very offensive to the environment. This would be most unwise. How are we going to change development if we keep avoiding it?

THE DYNAMICS OF SUSTAINABLE DEVELOPMENT

What makes development sustainable or not sustainable? The answer is actually quite simple (only the doing part is tough). What is sustainable development? It is as simple as: a form of development that is sustainable.

If we cannot do away with growth, then the only way to make development sustainable is to have growth with no added environmental impact.

How do we do this? There is actually only one way: to dematerialize as fast as we grow.

Here the word dematerialization denotes acts that reduce the consumption of materials (energy, water, land, forests, minerals, etc.) in each unit of economic output.

Or in equation form: dematerialization rate = growth rate.

Here dematerialization rate is measured as the rate of decrease of material intensity, which in turn is defined as

units of a certain material consumed (like kilo–liter oil equivalent in the case of energy) per unit of economic output (like dollars).

This simple equation tells us the root causes of our sustainability problems and answers many of the questions concerning the implementation of sustainable development. It gives the criteria on how sustainable development would have to be, and spells out the size of our tasks. It even lets us monitor how close or far away we are from our targets. It is the governing equation for the implementation of sustainable development.

Herman Daly, a leading economist who originated steady state economics, has argued that the answer to our sustainability problems lies in distinguishing growth from development, where growth is mere accumulation of wealth while development, on the other hand, is the improvement in quality of life. What kind of quality was he referring to? Simply stated, if it doesn't meet the dematerialization = growth requirement, it is not going to be of sufficient quality to achieve sustainability.

UNDERSTANDING DEMATERIALIZATION

Dematerialization is not a new thing. A part of our evolution has always been to become more efficient, to use less and less energy and materials to produce more and more goods and services. Neither is it new in sustainable development. However, as the basis for the actions we must take to make development sustainable, dematerialization needs to be looked at with considerably more care and rigor than we have in the past.

One of the most common mistakes made in studies on dematerialization is the lack of understanding that dematerialization should always be measured against growth. Dematerialization rate is the rate of change of material intensity (units of material like energy, water, etc. consumed per dollar of economic output, like gross national product (GNP) in the case of a nation) just like growth rate is the rate of change of GNP. So if in 1998, the energy intensity of a certain country is 1 kilo–liter oil equivalent per dollar of GNP, and in 1999 the figure went down to 0.97, the dematerialization rate for that country in 1999 would be 3% year^{-1}. *Dematerialization de cit* (i.e., growth rate minus dematerialization rate) rather than dematerialization itself, is the important parameter here. It is the size of the deficit that measures how close or far away we are from a state of sustainability, not the rate of dematerialization itself. We are better off with a 1% dematerialization rate when the growth rate is 3%, for example, than 2% when the growth rate is 5%. If we fail to appreciate this, we would draw very wrong conclusions.

Our current state of understanding on the subject of dematerialization is full of this kind of error. The following is an example. Figure 1 shows a sample of China's growth and energy dematerialization performances during its robust growth years. Dematerialization is not the problem but dematerialization deficits surely are.

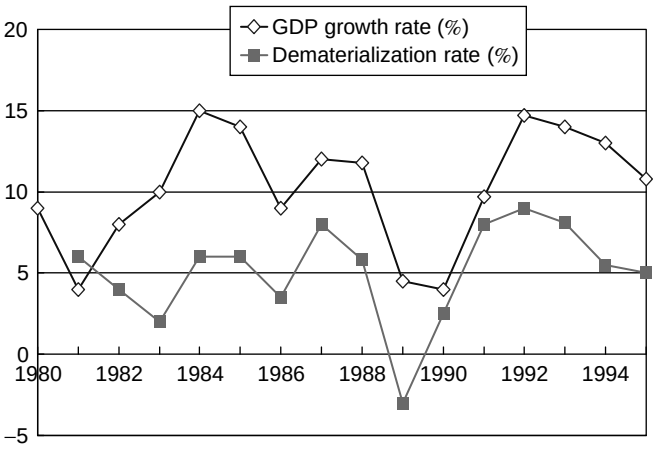


Figure 1 Growth and energy dematerialization rates for China for the years 1980–1995

Technology's Role in Sustainability

Technology almost always dematerializes. Historically, each new generation of technologies has almost always been more efficient and less material intensive. But if technology is so good to the environment, how is it that the results it brought were so bad? This is the typical growth versus dematerialization situation. Technology by itself is not bad. The trouble is that it has always done a lot more for the growth side of the equation than for the dematerialization side. An example is information technologies. They may have very low material intensities, but if, say, E-commerce stimulates growth and increases the demand and consumption of materials more than it displaces the share of heavily material-dependent industries' share in our overall economic activities, E-commerce would be bad for sustainability, not good.

Is a 1% Energy Efficiency Improvement a Positive Sign?

It is almost standard practice now to cite the fact that our energy intensity has been improving by about 1% a year since the 1800s as a sign that we are moving continuously on a positive dematerialization course. This cannot be more incorrect. Not only is this not a sign of improvement, the fact is that efficiency gains have consistently been far outpaced by the growth rates of our economic activities, and therefore the increases in demand and consumption of energy, is the very reason why we have such a major climate change issue now.

Do we have a Surplus of Dematerialization Technologies?

Many people contend that we have plenty of technologies available for our dematerialization purposes. The fact is that these technologies are plentiful only because of the dismal dematerialization performance we have managed to achieve so far. If we begin to dematerialize at a rate close to that of our growth rates, our stock of materials will be depleted in a hurry. Studies by the Electric Power Research Institute of the US have indicated that the existing stock of energy efficiency technologies would be enough to improve energy efficiency by 24–44%. How long will this last if we start to dematerialize energy at the kind of rates dictated by the Kyoto Protocol targets (which are essentially a form of dematerialization = growth requirement)? Less than 10 years.

Two kinds of people on opposite ends of the environmental spectrum – environmental idealists and right-wing contrarians – tend to over-dramatize our past performances and current capabilities in dematerialization. The former want people to believe dematerialization is easy, so everybody should be compelled to do it. The latter want people to believe dematerialization is easy, so nobody would be forced to do it. In reality, dematerialization is not easy at all, but everybody needs to do it.

Are Development and Environment no Longer Antagonistic Under Sustainable Development?

Many people have said that environmental and developmental interests are now no longer antagonistic, but aligned under the banner of sustainable development. This kind of optimism is grossly premature. If our dematerialization deficits are not narrowing, then things are not any more aligned or any less antagonistic than ever before. Antagonism between environment and development did not end when we invented the term sustainable development. It will end only when development finally becomes truly sustainable.

How Much Corporate Environmental Action Would Be Enough?

If a company published in its company environmental report that it has improved its energy or water intensity 5% last year, was that a good result? Yes and no. If the company's sales grew 10%, then 5% efficiency improvement was not sufficient. Now, if that company dematerialized certain natural resources 20% (twice its 10% revenue growth rate), was that a good result? Still yes and no. Very large gains in material intensity are most often the result of certain old technologies being phased out. The aggressive targets usually end when the phasing-out is completed (and the focal point of the next report would be shifted to another topic). Dematerialization has to happen at equivalent or greater than growth rates on a continuous, year-in and year-out basis. It is a forever kind of effort.

Most of what corporate-based environmental actions have achieved has been the reduction of pollutants and wastes. This progress is good, but this is the easiest part of our sustainability struggle. Pollution problems are largely solved in the developed world anyway. What industry needs to do now is to dematerialize the natural resources out of its products and production processes.

Businesses are where our society's capabilities are concentrated. They have the responsibility to show leadership in our toughest sustainability battles. At a minimum, they need to reduce the material intensities of their business activities at least as fast as the growth rates of their revenues; otherwise they would be asking the rest of society to subsidize them in our struggles to reach sustainability.

INDUSTRIAL AGE DEMATERIALIZATION: CLEANER PRODUCTION

When we talk about dematerialization, we usually associate it with things like cleaner production, pollution prevention, efficiency improvements, renewable energy, industrial ecology, design for environment and other such familiar actions. For convenience, let us group them together and call them *cleaner production* (the UN's Environmental Programme definition of cleaner production covers essentially everything of environmental worth).

Cleaner Production: How Far can we Run with It?

Unlike pollution control, which draws upon technologies developed for its specific purposes, cleaner production relies to a large extent on pushing harder the dematerialization attributes of mainstream technologies, such as new processes; to advance the causes of sustainability. But the development of these mainstream technologies is driven mainly by competition and growth purposes. Isn't there a contradiction here? What would we get at the end of the day, more dematerialization or more growth?

This question should be broken down into two parts. The first part is: do we need these mainstream technologies to play a major role in sustainable development? The answer is absolutely yes. Human affairs are driven by these mainstream technologies. They have got to become cleaner and more eco-efficient. All elements of cleaner production must be pushed to their very limits, and all kinds of incentives, including both sticks and carrots, must be brought to bear to make them happen. But will they be pushing dematerialization or growth harder? That's the second part of the question.

We are a long way from being able to control the environmental outcomes of technology. Most, if not all, of

the important technologies come out of the growth side, for growth purposes. They can become cleaner, for sure, but is it reasonable to expect these technologies, created for growth reasons, to do more for dematerialization than for growth? Pushing cleaner production may narrow the gap between growth and dematerialization, but can it bridge it? Let us look at what happened in the 1970s during the energy crisis – a time when the pressure and the incentives for dematerializing energy were greatest.

Probing the Limits of Cleaner Production

Growth and energy dematerialization rates of six developed economies are shown in Figure 2 for the period 1987–1996, which was relatively uneventful and is thus treated as a norm, and in Figure 3 for 1971–1982, which were the energy crisis years.

We note first that even though energy dematerialization (efficiency) rates were generally about one percentage point higher during the crisis years than during the normal period, most of these countries were still a long way from erasing their dematerialization deficits. We note also

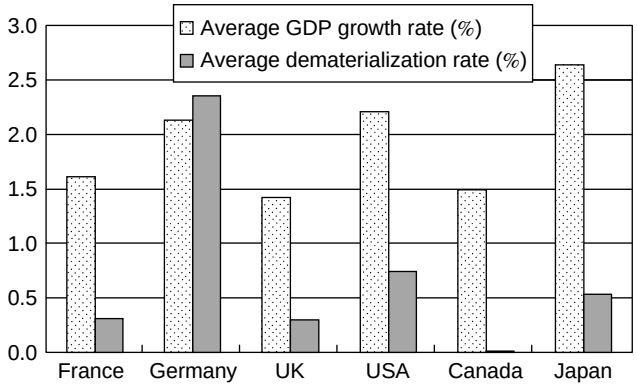


Figure 2 Growth and energy dematerialization during the early Information Age (1987–1996)

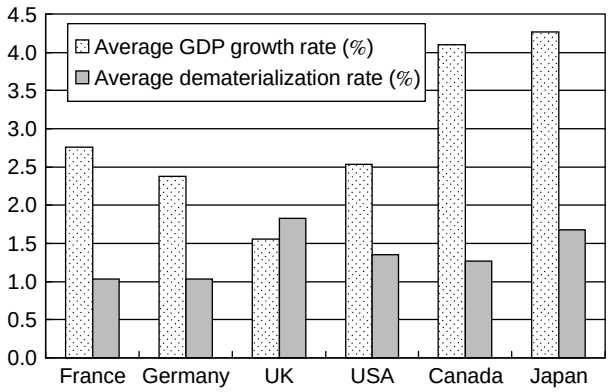


Figure 3 Growth and energy dematerialization during energy crisis years (1971–1982)

that 1987–1996 were the years when we first realized the threat of climate change and the need to reduce energy use to mitigate it. Yet if we have done anything to reduce the threat, we surely cannot find it in the energy dematerialization data shown in Figure 2 (Germany's good showing was mainly the result of reunification).

Businesses are usually vigilant about keeping their energy costs down, for competitive reasons. Businesses that stay in business would have to be efficient. Thus, Figure 2 represents the best that business can do under normal circumstances, with normal investment/return considerations. There is of course room for improvement, but that would take extraordinary actions and inducements.

It is premature to pass judgment on the efficacy of cleaner production, corporate environmentalism, etc. After all, it is less than a decade since Rio, when many of these dematerialization efforts were just getting started. However, sooner or later the effects must register on graphs like Figures 2 and 3. We need to see the effects of environmental actions on dematerialization versus growth charts, or they would show much motion but little movement.

Now back to the question, Would cleaner production bridge the dematerialization versus growth gap? The fact that we didn't bridge the energy gap even during the energy crisis years tells us that the answer is most likely not. To bridge the gap we need something outside the realm of the Industrial Age when cleaner production was spawned. That something is to be found in the Information Age.

INFORMATION AGE DEMATERIALIZATION: BITS-FOR-ATOMS

Bits are the fundamental unit of digitized information, while *atoms* are the fundamental unit of materials. The phrase bits-for-atoms, means the substitution of information for material. It is an obvious fact that our struggle to reach sustainability will take place largely during the Information Age. Given this obvious fact, it is truly bewildering that sustainable development has taken in so many fringe issues scarcely related to its core tasks, and yet has remained so distant and withdrawn from this one thing that is surely going to dominate human development for the foreseeable future. This will have to change.

Shapes of the Information Age

To get a feel for the Information Age, let us first look at its raw power. In the next 15 years, a combination of the expansion in usable bandwidth and advances in digital signal compression will increase the transmission capacity of bits by 5–6 orders of magnitude. When those bits reach us, the electronics at our fingertips will be able to process them about 1000 times faster. That means our power to access and use information will expand by 100 million to a

billion times in the next 15 years, or about 10 times every two years.

That is not all. What will truly transform our lives will actually be the so-called connection power. We have right now about 200 million computers that are connected together through the Internet and other networks at a rapid pace. By 2005 there will be 500 million of them. Mathematicians long ago proved that the intelligence of a network is proportional to the square of the number of its constituent members. That still is not all. Non-computer chips, which are already permeating all kinds of products and equipment, totaling six billion today and growing faster than computer processors, are also being increasingly interconnected through wired or wireless means.

Sustainability Promises of the Information Age

What would the Information Age bring to our dematerialization versus growth race? Could it be the force that finally reverses dematerialization's long-running losing battle with growth? Or could it push growth so much harder that, in spite of all information technology's ample dematerialization attributes, our race towards sustainability would actually get worse at a faster pace?

We will look at the positives first. Indeed, it has been said that the only way to dematerialize is to chip away the material contents of Industrial Age economic activities through cleaner production. Can't growth happen with little or no material dependency? Yes it can, in the Information Age. Information-based economic activities are much less material intensive, and many of their fastest growing elements are virtually material-less. With information power growing the way it is, many things are possible and on time horizons much shorter than we think. Bits-for-atoms is not just the foundation to our emerging network economy, it is also where our brightest hope for a breakthrough to sustainability lies.

To explore the potential of this bits for atoms shift in our growth, let us compare the material intensity of the Industrial Age to that of the Information Age. Table 1 shows the energy consumption share and gross domestic product (GDP) share of Taiwan's heavily material dependent industries. For a total consumption of 34% of Taiwan's energy, these industries contributed less than 7% to its GDP. This should be compared with Taiwan's present information industry. Data from the Hsinchu Science Park, the so-called Silicon Valley of Taiwan, will be used as a case in point. Studies show that its energy intensity was approximately one-eighth of the all-industry average, and its water use intensity was even smaller – in the order of one-fiftieth. Compared to its Industrial Age counterparts, information age growth is essentially material-less.

How important would bits-for-atoms be in our sustainability battles? If the bits-for-atoms transition in growth doesn't happen, then our only weapon would be cleaner production (in the broad UN sense). If future growth is 3% a year, then cleaner production would have to deliver 3% dematerialization a year to make the sustainability balance. That's a 20 times improvement in 100 years, 400 times in 200 years. Can cleaner production keep delivering these kinds of improvements? No. The cleaner you become, the tougher and costlier it is to become even cleaner. Sooner or later, cleaner production plateaus and further improvements would essentially run out. How soon we would come to this kind of dead-end depends on how fast we dematerialize. If we continue on our business-as-usual mode, we would never even know that we have this dead-end problem. Even if we dematerialize at close to the current rate of growth, we will run into this dead-end, perhaps in the next 30–50 years. Bits-for-atoms must build up a head of steam by then because it will be the only solution left.

Sustainability Pitfalls in the Information Age

While full of promise, the Information Age's beneficial effects on moving towards sustainable development are by no means a sure thing. Far from it, as a matter of fact. We are already quite a number of years into the Information Age and the bits-for-atoms shift has not even made a dent on the material content of our growth. If there is one thing that the rapidly advancing Information Age technologies will be sure to bring, it is not dematerialization, but

Table 1 Profiles of heavily material-dependent industries in Taiwan

	Energy consumption		Production	
	Amount (10^6 KLOE)	Share in national energy consumption (%)	Total sales (1991 $\$10^6$)	Share in GDP (%)
1977	0.77	36.9	–	–
1982	1.04	36.2	4618	6.5
1987	1.55	36.4	7885	7.1
1992	2.03	34.6	10 308	6.5
1997	2.72	34.7	14 330	6.7

GDP, gross domestic product.

increases in our productivity, and therefore greater acceleration in growth. Quoting *Wired* (July 1997), the so-called magazine of the networked world:

> Historians will look back on our era as an extraordinary moment. They will chronicle the 40 years from 1980 to 2020 as the key years of a remarkable transformation. In the developed countries of the West, new technologies will lead to big productivity increases that will cause high economic growth. And then the relentless process of globalization, the opening up of national economies and the integration of markets, will drive the growth through much of the rest of the world.

This may be nirvana for business executives but nightmares for implementers of sustainable development. Epochal technological advances are pushing the developed countries of the world onto new S curves, giving them the impetus for faster growth heretofore unrealizable for these mature economies. That is not even the most troublesome part. With explosive communications, technologies driving the globalization of not only knowledge but also ideas, economic democracy and free-market entrepreneurship will ourish in the Third World, unleashing many new Taiwanese and Koreans into the newly industrialized ranks. Two countries on the move are China and India, the first and second most populous states on Earth each with quality technical capabilities, low cost labor supply and large domestic markets, to create and sustain comparative advantages and robust economic performances for years to come. If we think growth is bringing too much pressure to our ecosystems now, we have to think again. All signs are pointing to faster growth, and therefore greater dematerialization challenges.

Indeed, in all likelihood our sustainability aspirations will suffer the bad side of the Information Age, before it will see the good side. Have we seen any signs that growth is less material-dependent or that dematerialization is catching up with growth? Mankind seems to have this very nasty habit – whenever we are given something that is twice as efficient, we want to use it four times as often (some people call this the rebound effect). Esther Dyson talked about this behavior in her book *Release 2.0: a Design for Living in the Digital Age* (Dyson, 1997). For her, advances in communication technologies didn't reduce her travel at all; she is actually traveling a lot more because she can now manage her office from the road!

Would information power be so potent that it will, for the first time, make growth so scarcely material-dependent that it overcomes the rebound effect and begins to promote dematerialization faster than it does growth, until finally it erases our dematerialization deficits? Or would it just push growth so hard that all of our best dematerialization efforts would be futile? That is almost the most crucial question in the sustainable development issue, and the Information Age.

CONSERVATION AND INCENTIVES

Whichever Age we are in, Industrial or Information, conservation will have to be a central and lasting part in our dematerialization struggles. The 1% extra dematerialization rates for energy during the energy crisis years (Figure 3) were largely the result of redoubled conservation efforts. Water conservation in the Los Angeles Basin has held today's consumption at the level of 30 years ago. Many metals are now recycled at very high rates (copper more than 90%, aluminum and lead over 80%, zinc 60%) that have prolonged the use of reserves, kept prices level and netted substantial returns.

Three things are important to understand about conservation.

1. enlightenment and behavior changes only play a small part. Easy options can only do so much. People who want to reduce wastes or improve efficiencies often discover that things are not as easy as they look and there are often substantial costs involved.
2. different conservation options cost different amounts. Large early gains are possible through exploitation of the easy and inexpensive options but the same rates of improvement cannot be kept up. For that to happen, costs typically would have to increase and there needs to be further and continuous technological innovations.
3. economics is what drives conservation. If we don't get enough conservation, the way to get more is not through persuasion and education. It is by changing the economics.

Conservation is Big Money

Conservation is not the esthetic endeavor that most people subconsciously perceive it to be. It involves big money. A 1% increase, for the US, on top of its present energy efficiency improvement rates, for example, would result in savings greater than a trillion dollars over the next 20 years (savings from dematerialization are compounded). This kind of money, however, doesn't grow on trees. Making it happen would entail upfront investment on the order of hundreds of billions of dollars by American corporations and citizens. That money is just for energy, only one percentage point of it. That's the kind of money involved when we talk about dematerialization and sustainable development.

Why are Incentives Needed?

Do people fork over this kind of money for the love of their neighbors? Most certainly not. Somebody would have to make it worth their while.

It is fashionable in sustainable development discussions nowadays to talk about the many technologies we have at our disposal that could not only cut deeply into our

material intensity but also earn us a nice profit doing it. One wishes that things were so simple. The fact is, yes, we will indeed earn a profit with these technologies. However, the majority of energy efficiency technologies actually have lower lifetime costs than their less efficient counterparts, but many of them never make it to the market because, as investment options, they don't measure up. There are two main reasons. One, their rates of return, or payback periods, are not good enough to compete with other investment options. Secondly, besides saving money they don't really do much else, like strengthening their potential investors' competitive positions or enhancing their core competencies.

What would it take to make people invest in efficiency and dematerialization technologies that do not compete favorably with other investment options? Two things. One is *command and control* (*see* International Environmental Law, Volume 5), such as product efficiency standards. The other is *incentives*. The best approaches would be those that combine the best attributes of each.

Incentives: Innovations Versus Brute Force

Command and control looks convenient. We pass a certain regulation or adopt a certain standard and the job is more or less done. In reality though, it takes considerable persuading for industry to accept new more stringent product efficiency standards, for example; and three years later new standards will again have to be negotiated upwards to keep efficiency gains rising. The result is that this kind of program often starts out well but soon runs into its own sustainability problems.

When people think of *incentives*, they typically think about taxes, fees and charges, which are put in place to change people's behavior. Their benefits are two. One, they discourage people from doing certain things because the costs are too high. Secondly, they give additional investment incentives because the materials or energy saved now cost more. Many economists believe that anything can be corrected if we get the prices right.

But, if this is so, why is it that there are still so many inefficiencies uncorrected? The reason is that, more often than not, price fixing is neither the most cost-effective nor the most efficient solution. Furthermore, new taxes, fees or price hikes often have only one-time effects. After a while we will have to raise them further if we want the improvements to continue. That could become very unpopular.

Experience of the demand-side management programs run by the American utilities, the market-based programs of many water utilities, and many cases with permits and trading, has shown that there are many innovative approaches that give considerably better results than straight taxes, fees or charges. They are a lot more popular. The most direct incentives such as rebates, which reimburse cash right into the hands of people who have made efficiency investments, are often far more cost-effective than price fixing. The reason is simple. At any one time there would be only so many people contemplating making efficiency or dematerialization type investments. It is a lot cheaper to provide incentives only to these people rather than raising the prices for everybody. Lawrence Berkeley Laboratory's analysis showed that the differences were as large as seven or eight times.

We should remember that all incentives are not created equal, some are more cost-effective than others, and innovations are always preferred to brute force. Mixtures of command and control and incentives are often the best approach. We could, for example, adopt a standard that stipulates a lower efficiency for a product to be allowed into the market, but sweeten it by giving the best 20% of these products a 30% purchase rebate. Or, the government could offer certain monetary rewards and set minimum efficiency-gain targets to let businesses come up with the best and the most innovative ways to bid for them. The possibilities are endless. In the age of sustainable development, these kinds of programs would have to flourish to the maximum extent possible.

PUTTING A NUMBER ON THE NORTH/SOUTH ISSUE

The North/South issue is one of the most complex and politically charged issues in sustainable development. It boils down to this: how much extra responsibility should the North bear so that the South may be given enough room to grow. We have been for decades mired in international wrangling between North and South bureaucrats, and there is little daylight in sight. The debates have been conducted often without common grounds or even common metrics, which are prerequisites for meaningful negotiations.

Why can't we take a scientific approach to the whole thing? Can we put a number on how much extra should the North do? We can, in fact. We have just never tried. It would be impossible only if we want the answer to be quantitatively precise. But quantitatively meaningful answers are not difficult at all, as demonstrated as follows.

We will divide the world into two blocks – a North block representing economies that have *per capita* incomes above 10 000 US dollars, and a South block below it (Table 2). Both are growing now at approximately 3% a year.

The state of sustainability as defined by the dematerialization = growth requirement can then be expressed with the following equation:

$$\bullet 1 - S_0 \bullet [1 + g_1 - d_1]n + S_0[1 + g_2 - d_2]n = 1 \quad \bullet 1 \bullet$$

Here S stands for the South-block total GDP as a fraction of total global GDP. The North would thus be $(1 - S)$.

Table 2 North and South Characteristics. Data Source: Britannica Book of the Year, 1996

	North	South
Average Per Capita GDP (US$)	23 000	1000
Number of Countries	48	169
Total GDP (US$ billion)	19 326	4983
Population (million)	846	4823

g and d, respectively, stand for the average growth and dematerialization rates over the n year span. Subscript o stands for the base year (year 0). Subscripts 1 and 2 denote the North and the South, respectively. The equation expresses the assumption that world material consumption (expressed as the sum of North and South consumptions) stays constant for the next n years.

Solutions to this equation are shown in Figure 4. It can be seen that if the North were to dematerialize at 1% beyond its minimum requirement (3%), the South would gain approximately 2% of extra growth (or 2% of allowable dematerialization deficit). How much difference would this 2% allowable dematerialization deficit or growth surplus make? A big difference. From Figure 5, it can be seen that the South would be able to expand its 50-year growth

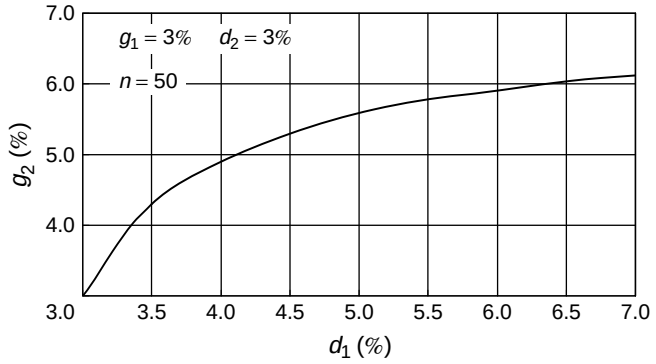


Figure 4 Sustainable South growth rates at various North dematerialization rates

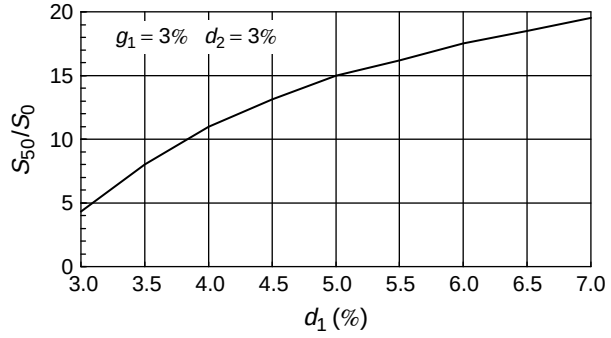


Figure 5 Sustainable 50-years South GDP expansions at various North dematerialization rates

from 4.4 times under the baseline scenario (minimum dematerialization by the North) to 11 times (1% extra dematerialization). This should be pretty close to or even greater than unrestrained growth. At this rate, North/South income disparity at the end of the 50-year period would be reduced from approximately 4 : 1 to 1.5 : 1. Human ills far beyond sustainability problems would be greatly relieved if this would actually occur.

Is it unrealistic for the South to dematerialize as fast as the North? Not necessarily. When the South starts to grow in any substantial way, it will shed its inefficient ways and, if governments do things right, build its growth on the newest and therefore the most dematerializing technologies from the North, while the North must wait out the depreciation of its own older technologies.

The demonstration given above, while simplified, should give us a quantitatively meaningful feel for the size of the North/South issue of sustainable development. Can we put a number on how much the North should give? Yes we can. The answer is 1% extra dematerialization beyond meeting its own minimum requirements.

HOW OUR SUSTAINABILITY BATTLES ARE WON: A SCENARIO

One day we may begin to take the implementation of sustainable development seriously, and start to put forth real efforts to balance growth with dematerialization. We will call it *day one*. Thirty years after day one, dematerialization rates will finally catch up with growth rates, and we will finally reach a sustainability balance. The balance will still be tenuous because neither of the two main driving forces for dematerialization – cleaner production and bits-for-atoms, is at a steady state. Cleaner production began quickly but slowed down, running into increasingly higher costs and diminishing returns. Bits-for-atoms started more slowly but began to accelerate as people's well-being is increasingly derived from virtual things. At the same time, as underdevelopment gradually disappears due to information-age stimulated globalization and development in the developing world, growth rates across the globe will slow down. One hundred years after day one, the contribution by cleaner production will have run out. But we are not worried. Bits-for-atoms has taken over completely. Growth is now slower and no longer material dependent. Dematerialization deficits are automatically erased. Sustainable development has been achieved.

REFERENCES

Dyson, E (1997) *Release 2.0: a Design for Living in the Digital Age*, Broadway Books, New York.

Schwartz, P and Leyden, P (1997) The Long Boom: a History of the Future 1980–2020, *Wired*, July.

FURTHER READING

Ausubel, J H (1998) The Environment for Future Business: Efficiency Will Win, *Pollut. Prevent. Rev.*, 8(1), 39–52.

Ayres, R U *et al.* (1995) *Achieving Eco-ef ciency in Business*, World Business Council for Sustainable Development.

Daly, H E (1998) *Beyond Growth*, Beacon Press, Boston, MA.

Electric Power Research Institute (1991) Proceedings, Fifth National Demand-Side Management Conference, CU-7394, July.

Kelly, K (1997) New Rules for the New Economy, *Wired*, September.

Schmidt-Bleek, F (1995) Increasing Resource Productivity on the Way to Sustainability, *UNEP Industry and Environment*, October–December.

Von Weizsacker, E, Lovins, A, and Lovins, H (1974) *Factor Four: Doubling Wealth, Halving Resource Use*, Earthscan, London.

Yang, J C (1998) *Sustainability Challenges in the Information Age*, OECD Proceedings: Globalization and the Environment: Perspectives from OECD and Dynamic Non-member Economies.

Eco-efficiency

ROBERT U AYRES

Center for the Management of Environmental Resources, Fontainebleau, France

Eco-ef ciency is a business strategy that emerged in the 1990s, in response to general recognition that the end-of-pipe treatment approach, dominant in the 1970s and 1980s, is too limited. In particular, it cannot adequately deal with global threats, such as the accumulation of greenhouse gases in the atmosphere and acidi cation of soils. The eco-ef ciency strategy is, essentially, to radically increase the value-added to material/energy inputs, which is tantamount to reducing the inputs per unit of service provided to customers. Implementation of this strategy at the company level implies a systematic search for win-win opportunities, i.e., opportunities to reduce environmental emissions pro tably.

Undoubtedly there are many such opportunities waiting to be uncovered, and the search for them should be a high priority. However, it is unlikely that the major global environmental threats (such as climate warming and sea level rise) can or will be eliminated by the voluntary adoption of process changes that are pro table in the near term. For instance, the evidence is strong that major new technologies, such as solar power or wind power, could only be competitive with existing established energy technologies after a considerable investment in research and development and capital accumulation to achieve necessary economies of scale and learning. Such investments would not necessarily be pro table in comparison with other more conventional opportunities. The essential role of government, therefore, is to create the necessary incentives for the private sector to make the needed investments.

The article discusses alternative policy approaches, ranging from regulation by government agencies to taxation of emissions or taxes on resource consumption. More innovative approaches are getting attention now. One, which has theoretical economic bene ts, is the creation of tradeable emissions permits (now a part of the Kyoto protocol). Another possibility is the use of tradeable consumption permits. Meanwhile, the eco-ef cient strategy for the private sector can be reformulated as a systematic program aimed at moving away from the sale of products to the sale of services, thereby treating the products themselves as capital assets. This change in perspective has very real consequences in terms of product design, for instance. Some sectors of industry (notably chemicals) have voluntarily moved in the direction of acknowledging downstream responsibility for their products. Government policies encouraging recycling and take-back will encourage this trend. Organizational and engineering implications of these new approaches are discussed.

> Eco-efficiency is reached by the delivery of competitively priced goods and services that satisfy human needs and bring quality to life, while progressively reducing ecological impacts and resource intensity throughout the life cycle, to a level at least in line with the earth's estimated carrying capacity.
>
> WBCSD (1995)

INTRODUCTION: HISTORICAL BACKGROUND

The deteriorating state of the human environment began to achieve headline status in the late 1960s. *Earthday* in 1970 marked the creation of the Environmental Protection Agency (EPA) in the US and comparable agencies in other western countries. This was followed quickly by the publication of *Spaceship Earth* by René Dubos and Barbara Ward, and the United Nations Stockholm Conference in 1972. This resulted in the institutionalization of environmental concerns at the international level through the creation of an Environmental Directorate at the Organisation for Economic Co-operation and Development (OECD) and a new United Nations agency, the United Nations Environmental Program (UNEP). However, the initial approach to environmental protection at the practical level was largely limited to what has been termed end-of-pipe treatment of sewage, large-scale industrial polluters of air and water and motor vehicles. In the following decade the scope of government programs was expanded in various ways, to include solid wastes and especially toxic and hazardous wastes.

In the 1980s the limitations of this end-of-pipe treatment approach began to become clear, especially with regard to stratospheric ozone depletion, acidification and greenhouse warming. These problems are all attributable to emissions that are either excessively costly or impossible to remove from waste streams. The case of acidification is a partial exception, in that there is a realistic possibility of removing sulfur dioxide from the ue gases of large coal-burning power plants through ue-gas desulfurization (FGD) technology. However, this technology is quite expensive and energy intensive and cannot realistically be applied to small industrial boilers or households (which dominate coal consumption in China, for instance). The potential for treatment of NO_x in combustion products is quite limited, and NO_x emissions continue to increase, despite the introduction of catalytic post-combustion treatment technologies in power plants and some motor vehicles. The notion of removing and sequestering carbon dioxide from power plant ue gases has been given serious consideration (as a hypothetical alternative to other approaches to carbon dioxide reduction). However, the same objections apply; even if this were technically feasible for some large electric power plants, it would not be applicable to smaller combustion sources. The case of chloro uorocarbons (CFCs) is the most straightforward: to the extent that recycling was feasible, it was already being done for the most part. Emissions were mostly leaks, which are inevitable and cannot be recaptured.

In the early 1980s, these facts created widespread concern that economic growth in the developing countries seemed to be in con ict with the imperatives of environmental protection, especially to the extent that the latter might involve increased costs. The response at the international level was the creation by the UN General Assembly of the Commission on Environment and Development, chaired by the then Prime Minister of Norway, Gro Harlem Brundtland. The commission prepared a major report, Our Common Future (1987, *see* WCED (World Commission on Environment and Development), Volume 4), which tried to make the case that environmental protection is an essential element of economic development. In so doing, it offered a definition of sustainable development that was necessarily vague enough to achieve political consensus but which, consequently, provided little operational guidance. One obvious prerequisite of long-term sustainability, however, is the reduction of anthropogenic pressures on the environment. These pressures arise from the competition between human needs and desires for food, housing and living space and the needs of other species. Above all, they arise primarily from extraction and processing of natural resources and using the environment as a sink for the disposal of waste ef uents and garbojunk.

More recently, the notion of *sustainable consumption* has emerged, to re ect the recognition that consumption of physical materials and energy derived from non-renewable fuels cannot exceed some finite limits imposed by the carrying capacity of the Earth. The biosphere can only absorb a limited amount of waste products (including, but not limited to carbon dioxide (CO_2)) associated with the extraction, processing and disposal of materials. There are many indications that the limits are not far away if, indeed, they have not already been exceeded.

It is now becoming clear that economic development in developing countries over the next half century at recent growth rates, combined with population growth that is already unavoidable, would require a five- to tenfold increase in the consumption of natural resources (Friends of the Earth, undated). This magnitude of increase may well be physically impossible, due to resource scarcities and capital constraints. It would certainly impose enormous stresses on the environment.

But even if such rapid gross national product (GNP) growth were feasible from an economic perspective, it would be inconsistent with ecological sustainability. For these reasons, a number of experts have proposed that the total worldwide consumption of materials (especially hydrocarbon fuels) should be stabilized. But allowing for inevitable global population growth, which will certainly increase by more than 60% in the next 50 years, and allowing for economic growth in the non-industrialized countries, especially China and South Asia, stabilization at the global level can only be achieved by sharp *reductions* in the consumption of materials in the industrialized West. Von Weizsacker *et al.* (1994, 1998) suggest a reduction factor of four. Others suggest the need for an even more dramatic tenfold reduction (Factor Ten Club, 1994). Whether the appropriate divisor is four or ten does not matter so much as the fact that any reduction at all means a major U-turn in economic development strategy (Ayres, 1996; Ayres and Weaver, 1998).

Bearing in mind these, and other problems, a new bottom-up way of thinking about emissions reduction also began to emerge in the 1980s. It has been variously called source reduction, waste reduction, clean technology, low-waste (or non-waste) technology, design-for-environment (DFE), dematerialization, industrial ecology and eco-efficiency. There are minor differences among these approaches. For instance, waste reduction emphasizes energy conservation and materials utilization efficiency. Clean technology emphasizes technical change that reduces emissions at the source, such as the classic example of substituting water soluble paints and cleaning agents for coatings and cleaning agents requiring volatile hydrocarbons or chlorocarbons. DFE emphasizes the potential for reducing wastes by making products more repairable, more remanufacturable or more recyclable. Industrial ecology emphasizes the potential for recycling wastes from one industry as feedstocks to another, as biological organisms recycle nutrients. (Oxygen is a waste product of photosynthesis,

while carbon dioxide is a waste product of animal and plant respiration.)

THE ECO-EFFICIENCY STRATEGY

The term eco-efficiency was first used in the book *Changing Course*, an input to the Earth Summit conference of the UN Commission for Environment and Development (UNCED) held in Rio de Janeiro, June 1992 (Schmidheiny, 1992). Some of the early works including the underlying ideas of eco-efficiency are given in Smart (1992) and Fussler and James (1996).

All of the bottom-up approaches summarized above share a common focus on process-change to reduce emissions at the source rather than to remove or treat them later on. Eco-efficiency can be regarded as an umbrella that subsumes all of the others. But it differs from the others in that it is being developed and promoted by a business group – the World Business Council for Sustainable Development (WBCSD) (*see* WBCSD (World Business Council for Sustainable Development), Volume 4) – as a strategy for business. Whereas the other approaches are typically defined by an outcome, eco-efficiency is a concept more firmly based on a business perspective. The basic approach, slightly simplified, is to increase the value-added to customers per unit of materials/energy input.

Eco-efficiency suggests the objective of maximizing value added per unit resource input. This idea is essentially equivalent to maximizing resource productivity at the company level (taking into account scarce environmental resources as well as energy and raw materials), rather than simply minimizing wastes or pollution associated with a given product.

From a management standpoint, there seem to be four key elements. They have been identified as follows (WBCSD, 1995):

1. providing real services based on the customer's needs;
2. assuring economic viability for the company;
3. adopting a systems (life-cycle) viewpoint with respect to both processes and products;
4. recognizing, at the company's policy level, that the environment is finite, that the carrying capacity of the Earth is limited, and that the company bears some responsibility regarding the environment.

The first two elements have always been, and remain, firmly in the domain of corporate management. However, with regard to elements 3 and 4, adopting a life-cycle viewpoint and recognizing that the earth has limited carrying capacity may not be enough to ensure that a company's actions are consistent with the imperatives of global sustainability.

It is possible to estimate the effect if all people worldwide were to buy or use a given product. Consider, for example, that nearly everybody wants to own a car. What are the implications for highway construction, loss of valuable farmland, petroleum consumption, air pollution, etc.? Such calculations, and questions, are essential aspects of the life cycle perspective. But they offer little or no guidance as to how an individual company should behave. This is because the carrying capacity is characteristic of the system (i.e., the Earth and the biosphere) as a whole. In short, it would seem that long-term sustainability for the Earth system is not a manageable criterion for an individual product or a company.

Proponents of eco-efficiency argue that there are many win-win opportunities for value creation that are also environmentally friendly. Some firms have introduced slogans expressing this idea, such as PPP for Pollution Prevention Pays (3-M Corp.) and WRAP for Waste Reduction Always Pays (Dow Chemical Co.). Unfortunately, these slogans cannot be taken too literally. The thrust of eco-efficiency is to actively search for, and create, such opportunities. However, most economists are quite skeptical about claims of win-win possibilities, at least at the level of process change. In the view of many experts, pollution prevention and waste reduction are more likely to be costly than profitable. There is evidence from the experience of end-of-pipe treatment to support this skeptical view.

If the global carrying-capacity criterion is beyond the control of firms operating in a competitive market, what then? One theoretical possibility would be regulation, at the global level, of total materials and energy consumption and waste output. Regulation could be accomplished in either of two ways. The first might be termed command and control, implying an authoritarian central bureaucracy to allocate production, akin to the central planning agency of a communist country. This approach would be the ultimate nightmare for most businessmen. It would certainly be politically unacceptable in today's world.

The second approach would be favored by most economists on theoretical grounds, in any case. It would be to internalize the social and environmental costs of excessive materials production and consumption by imposing taxes or ef uent charges. The notion of internalizing can be understood in the context of economic theory as eliminating externalities, or external costs, which are costs to a third party, not directly involved in a buy-sell transaction. Pollution, noise, congestion, and so on, are all externalities, in that they are costs that have to be borne by people who had no part in (and no gain from) the economic activities and transactions that caused the problem. Generally, such costs should be charged to those who caused them. This idea is sometimes expressed as the *polluter pay principle* (*see* PPP (Polluter Pay Principle), Volume 4). In the case of pollution, it is generally optimal for the cost to be charged to the party that has the greatest potential for eliminating the pollution through process or product redesign,

namely the manufacturer. In principle this could be done by imposing a tax equal to the social cost of pollution, and using the revenues to compensate those who suffer damage (Goulder, 1995). The tax would raise the cost of materials and/or energy consumption, discouraging trivial and low-value uses, and encouraging conservation. Tax revenues would then be used (in principle) to pay for pollution abatement or remediation, or to compensate those who suffer damage. The controversial carbon tax is an example of such a proposal.

Unfortunately, while this neat solution might work in some cases, it has drawbacks. One is that the social and environmental costs in question are impossible to calculate accurately. Some would inevitably be penalized too much and others too little. Moreover, experience suggests that a special-purpose tax collected by government is unlikely to be used as the theory requires (i.e., to abate the pollution or compensate those who are harmed by it). It is far more likely that the tax would be lumped in with general revenues and re-allocated to other purposes favored by law makers and their constituents. Also, the administration and enforcement of such a tax would itself be costly.

Another possible solution involving less governmental (i.e., bureaucratic) intervention is the adoption of *tradable emissions permits* (*see* Tradable Permits for Greenhouse Gases, Volume 4; Tradable Permits for Limiting Stratospheric Ozone Depletion, Volume 4; Tradable Permits for Sulfur Emission Reductions, Volume 4). This avoids the need for centralized global administration, but introduces other problems. In the case of carbon emissions from fossil fuels, at least, this approach is being seriously considered at the international level. But, as the 1997 Kyoto Conference revealed, to obtain a binding and enforceable agreement among a hundred or more sovereign nations, each concerned with protecting or advancing its national interests, is a very difficult challenge.

An alternative to taxes or tradeable emissions permits would be *tradeable consumption quotas*. This follows directly from the notion of sustainable consumption mentioned above. At the international level it would be no more difficult, and possibly easier, to negotiate than tradeable permits. One virtue would be that it would force the public to recognize that consumption – in the case of fossil fuels, at least – is essentially equivalent to pollution. To reduce the latter it is necessary to limit the former. Moreover, since the quotas could be allocated to individuals and traded directly through computerized markets, government intervention would be limited to determination of the quotas, and supervision of the allocation and trading system. The sellers of surplus quotas (underconsumers) would receive the income and spend it as they like. Buyers (including all industrial users) would simply pay the market price (Ayres, 1998). For them the effect would be similar to that of a carbon tax, except that the money would go directly to consumers rather than to a government agency.

A consistent business strategy can achieve the desired result without either excessive regulation or excessive use of fiscal mechanisms. Many businesses could voluntarily adopt the strategy of internalizing the product or, more simply, selling service instead of products. Since material wastes and environmental pollution are directly attributable to the production and consumption of material products, the only way to reduce wastes emissions in the long run is to reduce throughputs.

Hence, regardless of the choice of government policy instruments, whether regulatory or fiscal, the ultimate goal of an eco-efficiency strategy for companies must be to sell only the immaterial service content embodied in durable goods, retaining ownership and responsibility for the goods themselves. The effect, in the long run, would be to eliminate most durable goods, from cars to refrigerators, from the consumer market – along with consumer goods normally required for operation and maintenance, such as fuel, lubricants and antifreeze. It would also ensure a much higher degree of re-use, renovation, remanufacturing and recycling. This combined strategy would result in a gradual, but ultimately radical, *dematerialization* of the economy. Consumer goods, such as food products, cosmetics, pharmaceuticals, household cleaners, and so on do not embody enormous amounts of materials, except as packaging. The problem of reducing excessive use of packaging materials can be dealt with in other ways (*see* Dematerialization and Sustainable Development, Volume 4).

ECONOMIC/ORGANIZATIONAL PERSPECTIVE

Can radical dematerialization be accomplished without sacrificing economic growth and prosperity? This is a contentious but essential question. As already noted, mainstream economists are generally skeptical. A macroeconomic perspective can help to clarify the issues here. Until the late 1960s economists were inclined to conceptualize economic activity as a closed loop between production, consumption and investment, all measured in monetary terms. The connection with real material objects and the physical world from which materials are extracted, processed, made into products and finally disposed of as wastes, was never considered to be an essential aspect of economics until 30 years ago (Ayres and Kneese, 1969). Even now, these links are ignored in the great majority of economics textbooks and research papers.

However, the material aspects are vitally important to any consideration of long-term sustainability: human activity now competes with natural processes in terms of materials fluxes, especially with regard to the most toxic metals and the most environmentally harmful substances. Most

environmental problems are attributable either to land-use changes or materials consumption. It is true that emissions of wastes and pollutants are the most obvious problem. But it is also true that inputs become outputs. Every raw material extracted from the environment is a potential waste, and most become actual wastes within a few months. Only a small fraction is incorporated into durable goods, and very few durables survive more than a few decades.

The problem is mechanisms: economic growth in the past two centuries can be attributed largely to two growth engines. The first and most important was and is the dramatic decline in raw material prices, especially fossil fuel prices, beginning with the first industrial uses of coal and steam power in 18th century Britain. The availability of steam power facilitated the mining and transportation of coal, and eventually the generation of electricity and finally the production of all other goods and services. The process continued with the discovery and exploitation of petroleum and natural gas. The effect has been a long-term decline in the cost of mechanical power (i.e., work) delivered to the point of use. This, in turn, has permitted the substitution of machinery for human and animal labor and an enormous increase in output.

The second of the two growth engines has been *returns to scale*, which makes products cheaper to make as demand increases. In both cases the impetus to economic growth arises from the fact that demand for a product tends to increase as (real) prices fall. This phenomenon is called the *price elasticity of demand*. Falling consumer prices, as noted above, result from increasing use of fossil energy and mechanical power and exploiting economies of scale in manufacturing. The two cycles together are illustrated in Figure 1.

But the problem is that both drivers of growth are inherently materials intensive. The first depends on increasing the overall use of input materials (fuels) to replace labor. The second depends on increasing output of manufactured goods to achieve economies of scale. Thus the only way to reduce the materials intensity of the economic system – it would seem – is to lose the benefits of the two fundamental growth mechanisms or break the link to materials consumption. In the case of the first growth engine – declining energy/power costs – dematerialization requires development of other non-polluting sources of power, especially the sun, without reversing the downward cost trend. Perhaps this is technically feasible (despite many difficulties).

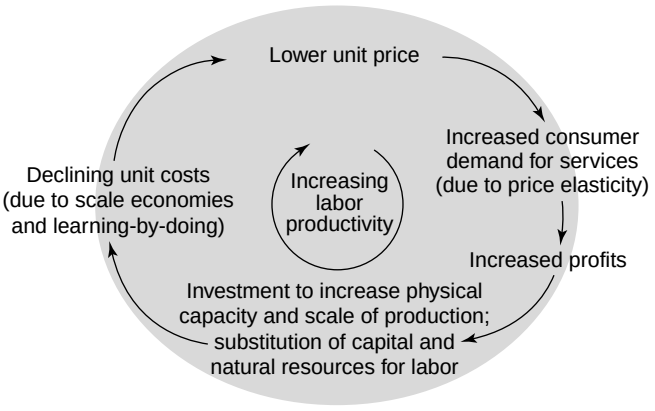


Figure 1 Old growth cycle (Salter cycle)

To enjoy continuing declines in consumer prices without depending on mass production and its associated economies of scale is less straightforward, if only because it is not really a technological problem, although information technology can probably play an important role. As already mentioned, the critical transition requires converting firms from a product orientation to a service orientation. This is fairly easy in some cases; it might simply mean renting or leasing capital goods and taking them back for remanufacturing on a scheduled basis rather than selling them. For consumer durables, however, the problem is much more complex. At bottom, it implies a change in the incentive structure of the competitive free-market system itself.

All durable goods in the service-based economy of the future would ideally be owned and maintained by service providers. In fact, they should be regarded as productive assets, to be used as efficiently as possible and conserved as long as possible. This means a change in the growth mechanism. One of the apparent advantages of an economy based on services is that it is inherently more labor-intensive than the mass-production manufacturing and throw-away economy. Repair, renovation, disassembly and remanufacturing are inherently more labor-intensive than original equipment manufacturing (Stahel and Reday-Mulvey, 1981; Stahel, 1994). These activities are much less able to exploit economies of scale or learning-by-doing precisely because they are less repetitive and standardized. For workers, however, they will be more interesting and will require higher levels of skill.

But what seems at first like an advantage in terms of job creation (especially from a European perspective) is a disadvantage from the perspective of labor productivity. The growth mechanism that depends on economies of scale to reduce prices and increase demand would operate in reverse: more labor per unit output means higher unit costs and that means falling demand. A fundamentally different economic mechanism is needed to reduce service costs as such. The likely mechanism for cost reduction in the service sector is *information technology* (IT). The other, and increasingly important, mechanism for cost-reduction is research and development applied to add value to services. Taken together these can be described as the new growth cycle (Figure 2).

Figure 3 shows schematically the limits of eco-efficiency as applied to the manufacturing economy, *vis-à-vis* the

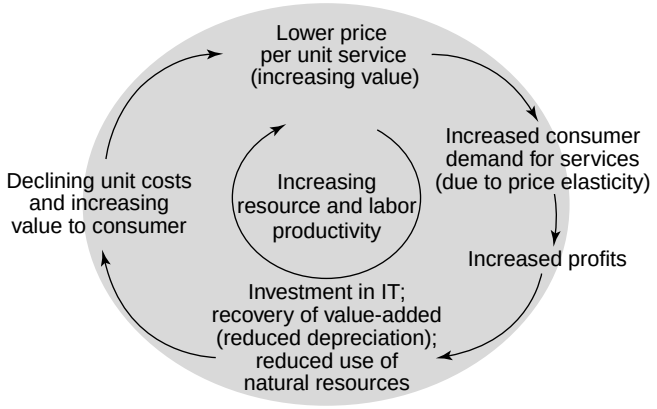


Figure 2 New growth cycle

service economy. It is convenient and evocative to label the latter as the spaceship – or zero-emissions – economy, in as much as it is an economy in which all materials must be recovered, repaired, re-used, remanufactured or as a last resort recycled sooner or later. In the spaceship economy nothing can be wasted without penalty.

ENGINEERING PERSPECTIVE: TO CLOSE THE MATERIALS CYCLE

Having considered the problem from a high-level economics perspective, it is now appropriate to consider the technical possibilities for approaching the goal of zero emissions. From an engineering point of view there are, essentially, three types of technologies to reduce wastes and emissions. These are:

1. energy and materials conservation;
2. product life extension: re-use, repair, renovation, remanufacturing, recycling;
3. waste mining: conversion of wastes into useful products.

All of these strategies are implicitly subsumed under the eco-efficiency umbrella. Note that waste treatment (at the end of the pipe) is not included, for the simple reason that waste treatment does not actually reduce the volume or mass of wastes, except in the rare cases where it creates a potentially useful product. One example is the use of FGD waste, which consists largely of calcium sulfate, as a substitute for primary gypsum in the manufacture of plaster wallboard. This is currently being done in Kalundborg, Denmark, where FGD waste from the ASNAES power plant is used as feedstock for the GYPROC wallboard plant (Ayres and Ayres, 1996). In that case it fits under the third category (waste mining). Otherwise, waste treatment simply shifts the waste from one environmental sink – air or water – to another (e.g., land). In too many cases, waste treatment has removed an emission from the air only to leave it in water, or vice versa. The substitution of aqueous cleaning agents for hydrocarbon or chlorocarbon solvents is an example of the first. The incineration of water treatment or sewage sludges is an example of the second. Incineration converts solid wastes into airborne wastes, usually to save landfill space. This sort of transfer is seldom desirable in the long run.

The first major category of emission reduction technologies generally includes efficiency improvements and process changes (source reduction, clean technology, etc.). Conventional energy-saving technologies, such as

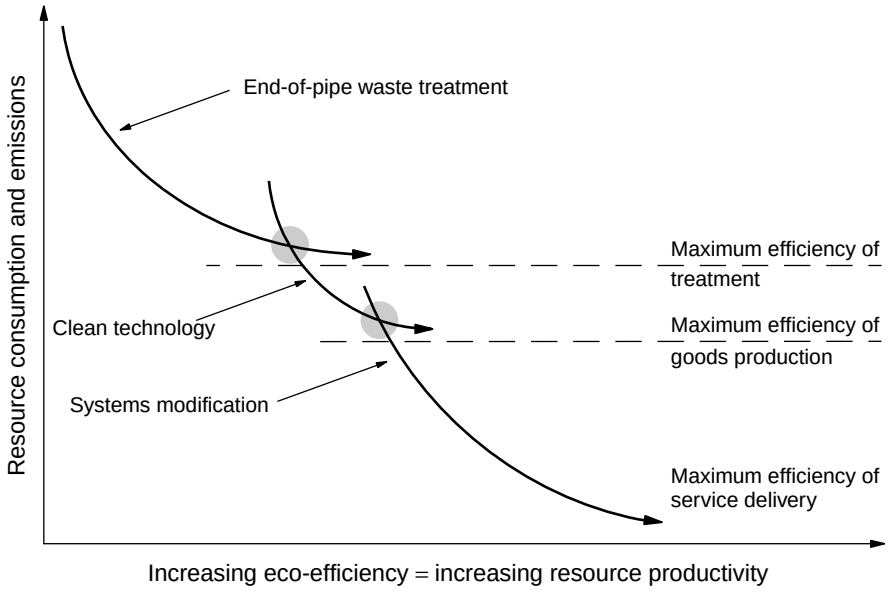


Figure 3 Three stages of eco-efficiency

co-generation, improved insulation, compact fluorescent lights, microwave ovens, smaller lighter vehicles, and so on, fall into this category. The same is true of a variety of industrial process innovations and material substitutions.

The second approach to reducing emissions is to extend product life by some combination of re-use, repair and renovation, remanufacturing and recycling. Re-use is very common among certain categories of consumer goods, such as clothing, shoes, blankets, books and furniture. Much of this is informal (e.g., clothing for refugees collected by churches), but there are fairly well-developed secondary markets for some categories of goods, such as books, bicycles, personal computers and automobiles; these items are frequently sold by departing students to incoming students at universities, for instance. Computerized bulletin boards within major institutions facilitate resale and re-use, and some of these information exchanges are reaching wider audiences through the Internet.

There is a major opportunity for increasing re-use in the area of shipping containers. Transport packaging – mainly consisting of corrugated paperboard boxes and wooden pallets – accounts for 16% of the US municipal waste stream. Re-usable packaging has long been standard for milk producers and soft drink bottlers, as well as some breweries. But in other cases, it is still relatively rare. Reusable packaging is technically feasible in virtually all cases. In fact, it offers a number of quantifiable environmental and economic benefits apart from waste prevention, including resource conservation and lower unit costs (Saphire, 1994).

Renovation and remanufacturing are options applicable mainly to capital goods and consumer durables. Truck and aircraft engines have been routinely remanufactured for a long time. The concept is closely related to a more recent concept of design for environment, which is now subsumed in the broader concept of *industrial ecology* (*see* Industrial and Anthroposystem Metabolism, Volume 3). But the fundamental point is that products that are designed with remanufacturing in mind will be designed differently than products designed for once-through consumption and disposal. They will, for instance, be designed for longer life and easier disassembly (Stahel, 1982, 1994; Navinchandra, 1991; Frosch and Gallopoulos, 1989). Firms that have begun to convert themselves from durable goods manufacturers to service companies include Xerox (copiers), Pitney-Bowes (postage meters), IBM, and a variety of others in businesses as diverse as elevators, furnace linings, truck tires and office carpets (Ayres *et al.*, 1997; Ferrer, 1997; Ayres, 1998).

Recycling of materials is also an important part of an integrated system; the technologies and economic factors are well described elsewhere in this encyclopedia.

The third way to reduce waste emissions is to convert them into useful products. This is a core idea of industrial ecology. Consider the famous Danish town of Kalundborg, where the idea was first implemented in practice (Frosch and Gallopoulos, 1989; Ehrenfeld and Gertler, 1997). The town has three primary producers: a petroleum refinery (Statoil), a coal fired electric power plant (ASNAES), and a large pharmaceutical and enzyme manufacturer (Novo Nordisk). By integrating these activities, it has been possible to reduce wastes dramatically while increasing value added. Thus the power plant supplies waste heat to the pharmaceutical plant, as well as serving a fish farm and a greenhouse. It also produces fly ash that is sold to a cement manufacturer and synthetic gypsum (calcium sulfate) from its FGD facility, to a wallboard manufacturer. The refinery supplies waste hydrocarbons as supplementary fuel to the power plant, and sulfur from its desulfurization process to a chemical company that makes sulfuric acid. Finally, the pharmaceutical manufacturer generates organic wastes that are consumed by the fish or used as fertilizer in the greenhouse.

Other potential opportunities to utilize industrial process wastes can be identified (Ayres and Ayres, 1996). For instance, phosphate rock processing (to produce fertilizer) yields a form of synthetic gypsum as well as by-product fluorides that can replace mined fluorspar for manufacturing synthetic cryolite (for the aluminum industry), or fluorides used in toothpaste. Coal burning power plants generate large amounts of fly ash and bottom ash, that can be utilized as an additive to Portland cement or, as a substitute for bauxite as a source of alumina.

Agriculture and agro-industry also offer opportunities for integration and synergy. Agro-industrial operations that utilize organic wastes on a large scale are both technically and economically feasible in some cases. Such wastes are available on an industrial scale in tropical plantations, sugar refineries, oil mills, breweries, pulp/paper mills, meat processing, dairy processing, and so on. In some cases, the integration process began long ago and is already virtually complete. For instance, large-scale meat processing in the US and Europe generates very little true waste – except some water pollution – because economic uses have been found for virtually every part of the animal, including hides, hair, bones, hooves, blood and even glands. The same is largely true of cane sugar, grain and oilseed milling in industrialized countries. Sugarcane waste (bagasse) and sugar beet processing wastes are underutilized.

However, smaller scale food and fiber processing activities, especially in the tropics, still generate a surprising amount of potentially utilizable organic waste. Paddy husk (from rice) can be used in paper-making, as a fuel for mushroom culture. Its potential value is said to be greater than that of the rice itself. R A Wickramasinghe of the Central Environmental Authority of Sri Lanka has compiled an astonishing list of potential uses of different parts of the coconut palm, ranging from the major commercial use as oil (for margarine, soap and candles) to activated

charcoal, wood, roof thatch, rope, brooms, vinegar, sugar, arrack, fodder, vitamin supplements, wound dressings and even teeth cleaners. However, he points out that certain wastes, such as coir dust (from coconut fiber processing) have virtually unlimited life in storage, but no known uses. Four million tons of this material (which is 90% water) have accumulated in Sri Lanka alone. The items mentioned above all came from materials that were distributed at a workshop on the development of non-waste technologies in developing countries.

Several interesting possibilities for utilizing tropical wastes have been identified in the United Nations University Zero Emissions Research Initiative (UNU ZERI) program, making use of mushroom farming, fish farming or other secondary sources of food or feed (Pauli, 1994; UNU, 1995). For instance, a small beer brewery in Namibia is now successfully recycling its water and organic wastes into a series of potentially profitable secondary activities including mushroom growing and fish-farming, while simultaneously saving fresh water.

Wood pulp and paper production probably still constitute the biggest single source of industrial water pollution in the world, despite great improvements in recent years. The fundamental problem is that, for all types of paper except newsprint, roughly half of the organic substance of the wood, namely lignins and ligno-cellulosics, is essentially wasted. In the most efficient modern pulp and paper mills in Scandinavia and North America, these wastes are burned for energy recovery. But, given the high moisture content, relatively little net energy is actually recovered and all of the potential chemical building blocks in the material is lost.

One of the most interesting long-term possibilities, also identified in the UNU ZERI program, is the development of an innovative steam cracking technology for separating the cellulose fibers from pulpwood and simultaneously cracking the high molecular weight lignins and ligno-cellulosic compounds. Lignocellulose, in particular, can be cracked to yield cellulose (about 30%) hemicellulose and lignin. The latter can be utilized mainly as a binder for particle board. However, hemicellulose is a valuable feedstock. It can be hydrogenated to produce xylitol (a sweetener), hydrolyzed to carbohydrates (sugars) which can be converted to ethanol, or dehydrogenated to produce furfural, from which a number of herbicides and rubber chemicals are subsequently derivable. The Tigney process was originally developed in Canada, but much of the further development has been carried out in Estonia.

ORGANIZATIONAL PERSPECTIVE: FROM PRODUCTS TO SERVICES

While some of the above concepts are technologically feasible and potentially economically attractive, they do not seem to self organize themselves spontaneously. The Kalundborg case in Denmark, now approaching 30 years of age, remains unique (Frosch and Gallopoulos, 1989; Ehrenfeld and Gertler, 1997). The economic system, as now structured, does not seem to encourage the emergence of integrated industrial ecosystems except in very special cases. The explanation is open to dispute. Regulatory barriers are one of the problems. However, the most plausible explanation is that it is not possible to optimize at the system level without a long-term central planning and coordination authority at the same level (Ayres, 1996).

This planning and coordinating function might be accomplished most naturally by centralized ownership, i.e., vertical integration and top-down management of the whole system. But one should not rule out *a priori* the possibility of other modes of organization, such as the Japanese keiretsu, that would be effective. A very powerful marketing organization, a city or county, or even a non-profit foundation, might be able to provide this essential planning and coordination function, for example.

An even more interesting possibility might be for the public authorities to create an integrated local utility for each industrial park to provide electric power, process heat, space heat, water and waste disposal services within the complex. In the first place, such an integrated utility would automatically incorporate several internal synergies, such as the potential for co-producing heat and power, and the potential for reducing solid waste disposal costs and electric power generation costs by incineration. In the second place, such an enterprise would have a strong incentive to create a joint venture with a building materials group that could utilize y ash and FGD wastes to cement, concrete products and wallboard, much of which could be used on site. However, in its absence, the decentralized pure market system, such that each activity (firm) interacts with the others only by competitive buying and selling, is unlikely to achieve the desired degree of integration.

In short, the inability of a group of decentralized manufacturing firms to optimize the use of by-products and secondary resources, and thus to minimize overall waste and pollution, constitutes a kind of market failure. In any event, waste reduction through process change is only the tip of the iceberg. It presupposes no change in the mix of goods and services produced and consumed by the economy. However great the potential for waste reduction (and dematerialization) by process change alone, there must be much larger potentials at the system level.

However, it is necessary to begin at the product level. This implies extending the useful lives and increasing the recoverability of value from durable goods. It also implies reducing the dissipative uses of materials (including fuels). A firm that simply sells its products to others and retains no responsibility for subsequent maintenance or disposal has no incentive to facilitate longer life – quite

the contrary – and little incentive to reduce operating cost if that means significantly higher initial cost. In fact, a manufacturing firm has a clear economic incentive to minimize product life so as to encourage early replacement. (The automobile companies, in particular, have been accused in the past of adopting a deliberate policy of planned obsolescence (Galbraith, 1968)).

It follows that major opportunities to reduce waste by extending life, and reducing operating and maintenance costs require one of two things. Either there must be long-term responsibility on the part of the manufacturer itself, or there must be a service provider with enough economic clout to dictate design specifications. In either case, some version of vertical integration, by ownership or by contract, appears to be needed. This can be termed *product internalization*, to suggest the economic effect without using language that has other connotations. The important economic implication is that in a product-as-service regime, products are capital assets and must be treated as such. Dissipative uses will be discouraged naturally. Owners of such product assets will have every reason to preserve them, upgrade them and recover value from them. Thus, conservation of value added by manufacturing is another way of characterizing the new paradigm.

One of the major implications of internalization of durable goods by service providers would inevitably be a sharp increase in re-use, renovation, reconditioning, remanufacturing and recycling. Here again, one can foresee significant opportunities for double dividends, since a durable product that is designed for long life, efficient reconditioning and eventual remanufacturing can offer a much higher value per kilogram of embodied material over its useful life than a product designed for minimum initial cost.

For instance, if automobile engines were truly designed for remanufacturing, the most complicated mechanical parts, such as the crankshaft, camshaft and the engine block itself would be made of very hard and durable materials, but with easily replaceable wearing surfaces such as roller bearings, piston rings and cylinder linings. The remanufacturing process would consist of disassembly, washing, replacement of all wearing parts, including gaskets and valves, and re-assembly. This could be partly mechanized, depending on batch size, although it would definitely be more labor intensive than the original manufacturing process. However, such engines would last indefinitely.

Leasing of durable goods by the manufacturer is the most obvious, and possibly the most common way to achieve effective internalization. Many kinds of capital equipment, including aircraft, are already leased by specialized leasing firms. However, in most cases this is purely a financial arrangement to free up the buyer's working capital. The aircraft, computer or machinery leasing firms are simply financial intermediaries; they take no responsibility for either the operation or the maintenance of the equipment itself. These responsibilities remain with the end-user, not the manufacturer. Here the aircraft case is the best model, since airlines – service providers – handle their own maintenance, or use specialized subcontractors, and have a very strong in uence on airliner design.

The case of rental car firms is very different. Car rental firms do take responsibility for maintenance, but only for relatively new vehicles. Operating costs (fuel) are simply passed on to consumers. Rental cars are invariably sold through the second-hand market while still relatively new, normally after only one or two years. Car rental firms, in practice, are merely marketing arms for the automobile manufacturers (and are mostly owned by them). They have no in uence whatsoever on vehicle design. Insurers potentially have some in uence in this domain, but they rarely choose to exert it. (For example, insurers could do a great deal to in uence manufacturers to make straightforward design changes that would reduce the costs of car theft; up to now they have neglected to do so; instead, costs have simply been passed on to consumers.) Thus, car rental, as currently practiced, cannot be regarded as a significant step towards internalization. It would become so only if service providers were forced to compete to sell services directly (per trip or per passenger km traveled) and if final disposal responsibility remained with the original equipment manufacturer.

The issue of industry structure in a service-oriented society is also important. From the standpoint of maximizing service output (and revenue) per unit of materials and energy input, the optimum organizational structure would seem to be a vertically integrated company whose operations encompass the entire life cycle of the product, from cradle to grave. The structure of AT&T (prior to its forced breakup) was close to this. One subsidiary, Western Electric Co. designed and manufactured telephones and all the various kinds of switching and transmission equipment for all operating divisions of AT&T. The individual Bell companies handled all maintenance for subscribers, and owned all the phones. When any equipment was replaced, it was recycled (or remanufactured) as a matter of course. The breakup of the system opened the way for independent equipment manufacturers, but cut the links between design, manufacturing, maintenance and recycling. Now old phones and other telephone equipment are not recycled, but are mostly junked and sent to landfills.

The same sort of vertical integration once characterized the computer industry. In its heyday, IBM manufactured all the equipment, wrote software, leased the computers to its customers, did the maintenance under contract, and took back the old equipment when the next generation machines were available. It is only to be expected

that, despite the rapidity of technological obsolescence, some re-usable electromechanical parts and components were recovered and re-used in less demanding machines, such as card punchers, sorters, magnetic tape readers, drums and printers. Why not? The customers were paying for service, and the service was guaranteed by IBM. Xerox Corp. had a similar rental with take-back policy in the early days, which also eroded over the years. Now, a smaller proportion of the Xerox machines are leased, which has reduced the number of returned end-of-life machines to just one-half of those retired. These cases illustrate the trade-off between vertical integration and competition.

REFERENCES

Ayres, R U (1996) Creating Industrial Ecosystems: A Viable Management Strategy? *Int. J. Technol. Manage.*, 12(5/6), 608–624.

Ayres, R U (1998) *Turning Point: An End to the Growth Paradigm*, Earthscan, London.

Ayres, R U and Ayres, L W (1996) *Industrial Ecology: Closing the Materials Cycle*, Edward Elgar, Aldershott, UK.

Ayres, R U, Ferrer, G, and Van Leynseele, T (1997) Eco-efficiency, Asset Recovery and Remanufacturing, *Eur. Manage. J.*, 15(5), 557–574.

Ayres, R U and Kneese, A V (1969) Production, Consumption and Externalities, *Am. Econ. Rev.*, 282–297.

Ayres, R U, ed and Weaver, P M, assistant ed (1998) *Eco-Restructuring: Implications for Sustainable Development*, United Nations University Press, Tokyo.

Brundtland, G H, ed (1987) *Our Common Future*, Oxford University Press, New York.

Ehrenfeld, J and Gertler, N (1997) Industrial Ecology in Practice: the Evolution of Interdependence at Kalundborg, *J. Ind. Ecol.*, 1(1), 67–79.

Factor Ten Club (1994) *Carnoules Declaration*, Factor Ten Club, Wuppertal, Germany.

Ferrer, G (1997) The Economics of Tire Remanufacturing, *Resour., Econ. Recycling*, 19, 221–255.

Friends of the Earth (undated) *Ten Times Better: Fair Shares in Environmental Space*, Friends of the Earth, Amsterdam.

Frosch, R A and Gallopoulos, N E (1989) Strategies for Manufacturing, *Sci. Am.*, 261(3), 94–102.

Fussler, C with James, P (1996) *Driving Eco-Innovation*, Pittman, London.

Galbraith, J K (1968) *The New Industrial State*, 1st edition, Houghton Mif in, Boston, MA.

Goulder, L H (1995) Effects of Carbon Taxes in an Economy with Prior Tax Distortions: an Intertempororal General Equilibrium Analysis, *J. Environ. Econ. Manage.*, 29(3), 271–297.

Navinchandra, D (1991) Design for Environmentability, *Proceedings of 1991 ASME Design Theory and Methodology Conference*, American Society of Mechanical Engineers, Miami, FL.

Pauli, G (1994) *Zero Emissions Research Initiative (ZERI)*, United Nations University Tokyo (draft prospectus).

Saphire, D (1994) *Delivering the Goods: Bene ts of Reusable Shipping Containers*, Special Report, INFORM, New York.

Schmidheiny, S with the Business Council for Sustainable Development (1992) *Changing Course: A Global Perspective on Development and the Environment*, MIT Press, Cambridge, MA.

Smart, B (1992) *Beyond Compliance: A New Industry View of the Environment*, World Resources Institute, Washington, DC.

Stahel, W R (1982) The Product Life Factor, in *An Inquiry into the Nature of Sustainable Societies*, ed S G Orr, Houston Area Research Center, Houston, TX.

Stahel, W R (1994) The Utilization-focussed Service Economy: Resource Efficiency and Product Life Extension, in *The Greening of Industrial Ecosystems*, eds B Allenby and D Richards, National Academy of Engineering, National Academy Press, Washington, DC.

Stahel, W R and Reday-Mulvey, G (1981) *Jobs for Tomorrow: The Potential for Substituting Manpower for Energy*, Vantage Press, New York.

United Nations University (1995) Towards a Virtual Biorefinery Zeri Report, United Nations University, Tokyo.

von Weizsacker, E U, Lovins, A B, and Lovins, L H (1994) *Faktor Vier*, Droemer Knaur, Munich.

von Weizsacker, E U, Lovins, A B, and Lovins, L H (1998) *Factor Four: Doubling Wealth, Halving Resource Use*, Earthscan, London.

Ward, B and Dubos, R (1972) *One Earth: The Care and Maintenance of a Small Planet*, Penguin Books, Harmondsworth, UK.

WBCSD (1995) Report of the World Business Council for Sustainable Development, 2nd Antwerp Eco-Efficiency Workshop, Antwerp.

Monitoring in Support of Policy: an Adaptive Ecosystem Approach

JAMES J KAY[1], MICHELLE BOYLE[1], AND BRUCE POND[2]

[1] *University of Waterloo, Waterloo, Canada*
[2] *Ontario Ministry of Natural Resources, Peterborough, Canada*

*The development of a monitoring program is usually regarded as synonymous with generating indicators. This simplification reflects the belief that there exists a correct set of things to measure. Once these measures are discovered, there remains only the task of collecting the data and making the results available. Such a traditional approach permits monitoring programs to be developed that are disassociated from the context; that is, people and their concerns. Ungrounded initiatives can produce fragmented programs that are not effective and do not use resources efficiently. (For example an inventory of one Biosphere Reserve in which we were involved revealed over 60 separate and uncoordinated monitoring initiatives.) Or, their abstruse utility may cause them to fizzle out before they are even implemented (*see *Biosphere Reserves, Volume 2).*

These shortcomings occur despite the best intentions and conscientious work. The difficulty lies in the fact that normal science and available problem-solving tools are insufficient when dealing with complex systems. Issues of sustainability are persistently untidy and demand an examination of the wider ecological-economic system to resolve. Thus, conventional monitoring methods applied to sustainability tend to fail. New approaches and tools that incorporate complexity into monitoring programs are required.

A reframing of science, propelled by the need to unravel complexity, is emerging under the rubric of post-normal science (Funtowicz and Ravetz, 1993; Kay et al., *1999) (*see *Post-normal Science, Volume 5). In this article, we present a scheme where the activities of monitoring and assessment are treated as an integral component of an adaptive ecosystem approach to sustainability and health.*

In this ecosystem approach, governance, management and monitoring form a triad of activities that are carried out in the context of an issues framework of human concerns and an explicit conceptual model of the ecological-economic system. Taken together, the issues framework and the conceptual model focus the discussion of sustainability: the sustained health and integrity of the ecological-economic system. Governance, management and monitoring activities then inform decisions to resolve the necessary tradeoffs and continually chart a course to sustainability.

In most Western jurisdictions, the issues framework and conceptual model, which animate an ecosystem approach, take the form of policies and/or regulations, which are enacted as part of a legislative process. In this traditional mode, monitoring, as described in this article, is undertaken in support of policy. However, there are forms of virtual governance emerging which have no formal legislative powers and whose governance does not take the form of traditional policy. For example, United Nations Biosphere Reserves are often overseen by an informal and ad hoc assembly of concerned individuals and non-government organizations. Even though these individuals may have no legal power, they do act as stewards of the Reserves and in effect, manage them. Monitoring, as discussed herein, is meant to be in the service of such initiatives as well as the more traditional policy-based approaches to environmental and sustainability issues.

COMPLEX SYSTEMS THINKING, SETTING THE STAGE

A central challenge of successfully resolving the issues of sustainability for ecological-economic systems (see Figure 1) is learning to deal with complexity; that is, understanding situations in which dynamics are dominated by *self-organizing phenomena* (Kay *et al.*, 1999; Kay and Schneider, 1994; Schneider and Kay, 1994b). A theoretical basis for such an understanding is emerging and is referred

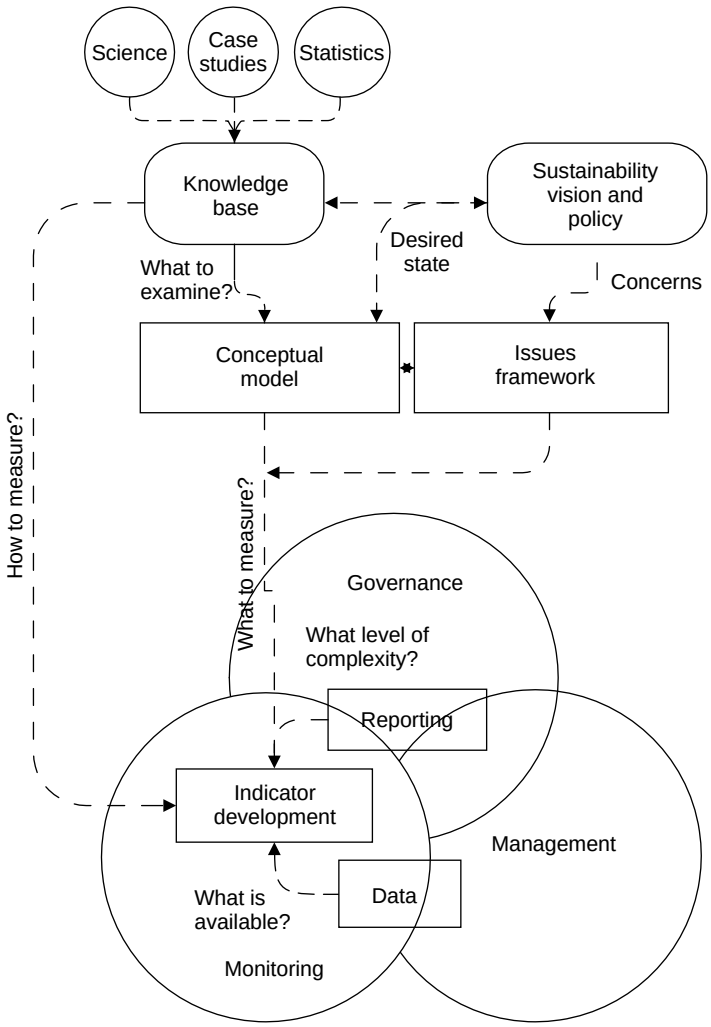


Figure 1 Monitoring in context; the arrows show an influence on indicator development

to as complex systems thinking (see Table 1). Its main elements are hierarchy and self-organization theory.

The hierarchical nature of complex systems requires that they be studied from different types of perspectives and at different scales. There is no one correct perspective. Rather, a diversity of perspectives is required for understanding. By their nature, such systems are self-organizing. This means that their dynamics are largely a function of positive and negative feedback loops. Linear, causal mechanical explanations of their dynamics are precluded. In addition, emergence and surprise are normal phenomena in systems dominated by feedback loops. Inherent uncertainty and limited predictability are inescapable consequences of these system phenomena.

Complex systems organize about *attractors*. Even when the environmental situation changes, the system's feedback loops tend to maintain its current state. However, when system change does occur, it tends to be very rapid and even catastrophic. Precisely when the change will occur, and to what state the system will change, are generally not predictable. In a given situation, there are often several possible system states (attractors) that are equivalent. Which state is currently occupied is a function of its history. There is not a correct state for the system, although there may be a state which is preferred by humans. This new understanding of complexity suggests a different role for science and monitoring in the quest for sustainability.

Hierarchy

There is growing comprehension that sustainability issues must always be examined within their broader context. Every system is a component of another system and is itself, made up of systems. So, a wetland must be understood in the context of the subwatershed it is a part of, and in terms of the processes and species, which make it up. The body of thinking which deals with these issues is called *Hierarchy Theory* (Ahl and Allen, 1996). Its central tenent

Table 1 Properties of *complex systems* to bear in mind when thinking about ecological-economic systems

Non-linear: behave as a whole, *a system*. Cannot be understood by simply decomposing into pieces, which are added or multiplied together
Hierarchical: are *holarchically nested*. The system is nested within a system and is made up of systems. Such nestings cannot be understood by focusing on one hierarchical level (holon) alone. Understanding comes from the multiple perspectives of different *types* and *scale*
Self-organizing (internal causality): are non-Newtonian, not a mechanism, but rather *self-organizing*. Characterized by: goals, positive and negative feedback, autocatalysis (self-reinforcing processes), emergent properties and surprise
Window of vitality: must have enough complexity but not too much. There is a range within which self-organization can occur. Complex systems strive for *optimum*, not minimum or maximum
Dynamically stable?: may not have equilibrium points (that is they may not exist for the system)
Multiple steady states: do not necessarily have a unique preferred system state in a given situation. *Multiple attractors* can be possible in a given situation and the current system state may be as much a function of historical accidents as anything else
Catastrophic behavior: experience sudden unpredictable behavior and state changes
Bifurcations: moments of unpredictable behavior when a system's development may proceed in either of two divergent directions
Flips: sudden discontinuities, rapid change
Holling four-box cycle: the ongoing process of birth, growth, death and renewal, e.g., the shifting steady state mosaic of a wildfire-driven forested landscape (Holling, 1986)
Chaotic behavior: have limited predictability. Our ability to forecast and predict is always bounded regardless of how sophisticated our computers are and how much information we have

is that sustainability issues can only be understood in terms of systems embedded in systems which are also embedded in systems or, in the vernacular of hierarchy theorists, as *nested holons*.

Hierarchy theory requires that a study of complex systems begin by careful consideration of the types of perspectives required and the appropriate scales of investigation. For example, the traditional ecological perspective, which dictates an approach based on studies of population dynamics of individual species, is not helpful in understanding fire dynamics in ecosystems. Rather, understanding of fire phenomena comes from studies of communities and landscapes over local and regional scales.

The researcher must take care to identify the human sustainability issues at hand and the appropriate perspectives and scales of investigation necessary to deal with these issues in an ecological-economic context. This identification process can only occur in the context of human values and requires bringing a diversity of views to bear on the question at hand.

To do otherwise is to court disaster. For example, research on ecosystems and sustainable food production in some areas of the upper Amazon watershed have historically centered on cattle production and its effects. A community-based hierarchical system analysis of the situation revealed the main food source to be fish with cattle playing an incidental role. Thus, community health and development initiatives focusing on beef production were sorely misplaced. Another example comes from an early Canadian environmental impact assessment that examined the effects of development on reproduction of caribou herds. Later, it was noted that the most serious effect of development was on the food source of the caribou, but this was not studied as part of the impact assessment. The study focus was on the reproductive habitat of a single species, and ignored the (much more relevant) broader food chain.

The challenge for science is to abandon the normal approach of searching for the single correct model for dealing with a problem. Instead, the message of hierarchy theory is that we must develop a manner of investigation that uses a diversity of different perspectives and models, which brings different players to the table, and which synthesizes the different perspectives together into understanding.

Self-organization and Attractors

Complex systems have multiple possible operating states or attractors, and may shift or diverge suddenly from any one of them (Holling, 1986; Kay, 1991; Ludwig *et al.*, 1997). For example, a portion of a natural area (A natural area is a section of countryside which has been set aside to be maintained in its natural state e.g., a national park or Biosphere Reserve) in Southern Canada is a closed soft maple swamp in a wetland community. However, the amount and duration of the ows of water can radically alter this operating state. Drying events, such as an extended drought, could change the operating state to an upland forest community or grassland with associated vegetation structure. If extended periods of ooding do occur, high water levels would shift the operating state to a marsh ecosystem. The shift happens because red and silver maple are tolerant to ooded conditions within 30–40% of the growing season. If ooding events are greater than this threshold, the forest trees die and give way to more water-tolerant herbaceous marsh

vegetation. The feedback mechanism, which maintains the swamp state is evapotranspiration (i.e., water pumping) by the trees. Too much water overwhelms the pumping capability of the trees and not enough shuts it down. The point of this example is that the current ecosystem state is a function of its physical environment and the accidents of its history. A single dry or wet season can change, for decades, what is on the landscape.

Each of the ecosystem states, in this example, is as ecologically healthy and appropriate as the others. There is not a right community for this landscape. Each is equally right. Thus, scientists cannot tell decision-makers or policy makers which of these states is ecologically better. Scientists can only provide information about the different tradeoffs each state represents. A decision regarding which of these states to promote is necessarily a value decision. In fact, since water ow is the key in uence which determines the organizational state of this portion of the natural area, decisions about which ecosystem state to promote hinge on what kind of land use and development is allowed on the lands adjacent to the natural area; clearly a political decision.

A similar example comes from the management of the same natural area (Lister and Kay, 2000). Currently, the aquatic ecosystem consists of a fast moving, highly oxygenated cold water stream in which brook trout thrive. However, another state, which is normally incompatible with the survival of brook trout, is emerging: a slow moving, low oxygen, warm water stream interrupted by ponds and small wetlands. In this case, beaver and muskrat are dominant species that shape and maintain the habitat and its constituent communities. The current ecosystem trajectory is tending towards this state due to several factors, including an invading population of beaver and purple loosestrife. The beaver dam the stream, thus providing more habitat for beaver; and the purple loosestrife choke out the other species in the wetland. Both factors slow the water ow, increase water temperature and decrease dissolved oxygen. The emergence of this attractor was a surprise to the local managers of the natural area. However, managers operating at the much larger scale of the drainage basin (of which this natural area is part) were not surprised, as they have followed the beaver migration into the area for years.

The group responsible for the natural area would like to maintain both the trout stream and the beaver and their ponds. However, this may not be feasible since the two ecosystem states are potentially mutually exclusive. A difficult choice may have to be made to favor one over the other. The irony is that ecosystem managers may have to trap beaver (Canada's national animal) and keep them out of the natural area in order to preserve a pristine state which will likely require intensive human management to maintain.

In making this choice, the values of various stakeholders and community members will play an essential role in the decision-making process. The role of the scientists will be to inform decision-makers about the ecological options, the tradeoffs and uncertainties involved, and various strategies for in uencing what happens on this landscape. However, scientists cannot inform us about the correct way to proceed, nor, can they predict with certainty what will happen in the situation. So, the role of science in decision-making for sustainability changes from problem solver (in the sense of providing a solution for the situation) to the role of facilitating understanding about the biophysical realities of the situation. In this manner, both experts and stakeholders contribute to the resolution of the situation.

Narratives

By their nature, ecological-economic system dynamics cannot be captured in a single model. Multiple descriptions are required to deal with complexity, and self-organizing phenomena cannot be explained in terms of a linear mechanistic causal description. Consider the case of shallow lakes, such as Lake Erie in North America. Two different attractors for shallow lakes have been identified (Carpenter and Cottingham, 1997; Kay and Regier, 1999; Regier and Kay, 1996; Scheffer, 1998). In the *benthic* state, the ecosystem is characterized by high water clarity and lake bottom vegetation. As nutrient loading increases the turbidity of the water, the ecosystem crosses a catastrophe threshold and ips into a hypertrophic, turbid, phytoplankton ecosystem, i.e., a *pelagic* state. Lakes exist that regularly ip between the two attractors. At least three quite different descriptions of such a lake will be needed, one for the pelagic state, one for the benthic, and one for the intermediate stage as the system ips between attractors.

The nature of these descriptions is also quite different. The description is in terms of the feedback loops that tend to maintain the ecosystem in its current state. Of particular importance are those feedback loops which buffer the system from changes in external in uences. In the case of Lake Erie, the benthic ecosystem has elaborate feedback schemes, operating at different spatial and temporal scales, to limit the phosphorous in the water column. The pelagic state has elaborate schemes to accomplish just the opposite. Describing the ip from one attractor to the other involves accounting for how environmental in uences (acting at different spatial and temporal scales) disable one feedback system while enabling another. Such a description takes the form of a multi-layer narrative which recounts the ecosystem's operation from different perspectives and scales (Kay and Regier, 1999; Kay *et al.*, 1999). While individual elements may consist of traditional scientific models and descriptions, synthesizing these elements together into a narrative transcends them to invoke a more complete picture of the system. Narratives are a cornerstone of the description of the dynamics of complex systems. They are also the

end product of monitoring programs undertaken as part of an adaptive ecosystem approach to sustainability and health.

Ecological-economic Systems as Self-organizing Holarchic Open (SOHO) Systems

Our premise is that resolving sustainability issues for ecological-economic systems entails understanding these systems as SOHO systems. Such an understanding comes from thinking through the hierarchical nature of these systems by considering issues of type and scale, and the bounding and nesting of the system. This hierarchical systems description of the important processes and structures, and their relationships and context, is essential. In addition, the self-organizing behaviors of the system need to be identified, described, and understood in so far as is possible. This involves identifying the attractors accessible to the system, the feedbacks which maintain the system at the attractors, the external in uences which define the context for a specific attractor, and conditions under which ips between attractors are likely. The overall understanding of system behavior that comes from studying aspects of a SOHO system is summarized in the form of a narrative. (A more detailed discussion of these notions can be found in Kay *et al.*, 1999.)

As mentioned above, a hierarchical system description requires identification of the important aspects of the system. What is deemed important can only be identified by examining the context of human values and human sustainability issues associated with the given situation. Therefore, in order to undertake a SOHO approach, the stakeholders and actors involved in the situation must participate. As well, their roles, issues, concerns and the aspirations that they bring to the situation must be understood. This understanding is summarized as an issues framework for the situation, which includes a vision for the future.

The process of identifying the important ecological and societal aspects of the system proceeds on the basis of the human issues framework and vision. Together, the SOHO system description and the issues framework are the starting point for an adaptive ecosystem approach to sustainability and health (Figure 2).

MONITORING AS PART OF AN ADAPTIVE ECOSYSTEM APPROACH TO A SUSTAINABLE SOCIETY

The complexity of resolving sustainability issues, necessitates a SOHO system approach. SOHO systems have the ability to decouple their behavior from environmental change, to generate novel behavior, and to self-organize in response to change (in a way that we cannot determine). These properties thus limit our capacity to predict how such situations will unfold. In principle, in many situations it will not be possible, or even appropriate, to make accurate quantitative models which forecast the future. It may be more useful to generate qualitative narratives of how the situation might unfold.

The premise of the traditional approach to management (anticipatory management) is that it is possible to predict and anticipate the consequences of decisions. Once all the necessary information is gathered to make a scientific forecast, the right decision can be made. This approach is not valid when dealing with complex systems, however. Given the limitations imposed by complexity, the focus of management and decision-making strategies must be on maintaining the capacity to adapt to changing environmental conditions. Adaptive management involves a very different agenda than anticipatory management (Gunderson *et al.*, 1995; Holling, 1978). In adaptive management, differences between how the future actually unfolds and how it was envisioned are seen as opportunities for learning. This is in sharp contrast to anticipatory management, which sees such deviations as errors to be avoided. Much of the agenda of adaptive management is learning through experimentation rather than focusing on error avoidance. Monitoring plays a crucial role in adaptive management. It provides feedback necessary for adaptive learning, information to generate narratives, and advice to determine if alternative courses of action are required.

At the core of an adaptive ecosystem approach is the premise that a sustainable society maintains itself in the context of the larger ecological system of which it is part. The formulation of a sustainable society involves realizing a vision of how the landscape of human and natural ecosystems should co-evolve as a self-organizing entity. Decision-making comes to be understood for what it has always been, finding our way through partially undiscovered country rather than charting a scientifically determined course to a known end point. A framework for this undertaking is presented in Figure 2.

In this post-normal approach to sustainability, science takes on a different function (*see* Post-normal Science, Volume 5). The scientist's role in decision-making shifts from inferring what will happen, that is, making predictions which are the basis of decisions, to providing decision-makers and the community with an appreciation, through narrative descriptions (scenarios), of how the future might unfold. Through these SOHO descriptions, science informs society about known ecological constraints and possibilities (see the top left hand box in Figure 2). People provide an image of how they would like to see the landscape of human and natural ecosystems co-evolve (see the top right hand box in Figure 2). A dialogue must ensue (the diamond box in Figure 2), which explores the desired and the feasible options and reconciles these in a vision of how to proceed. Scientists inform this dialogue by providing future narratives that will evolve as scientists participate, as

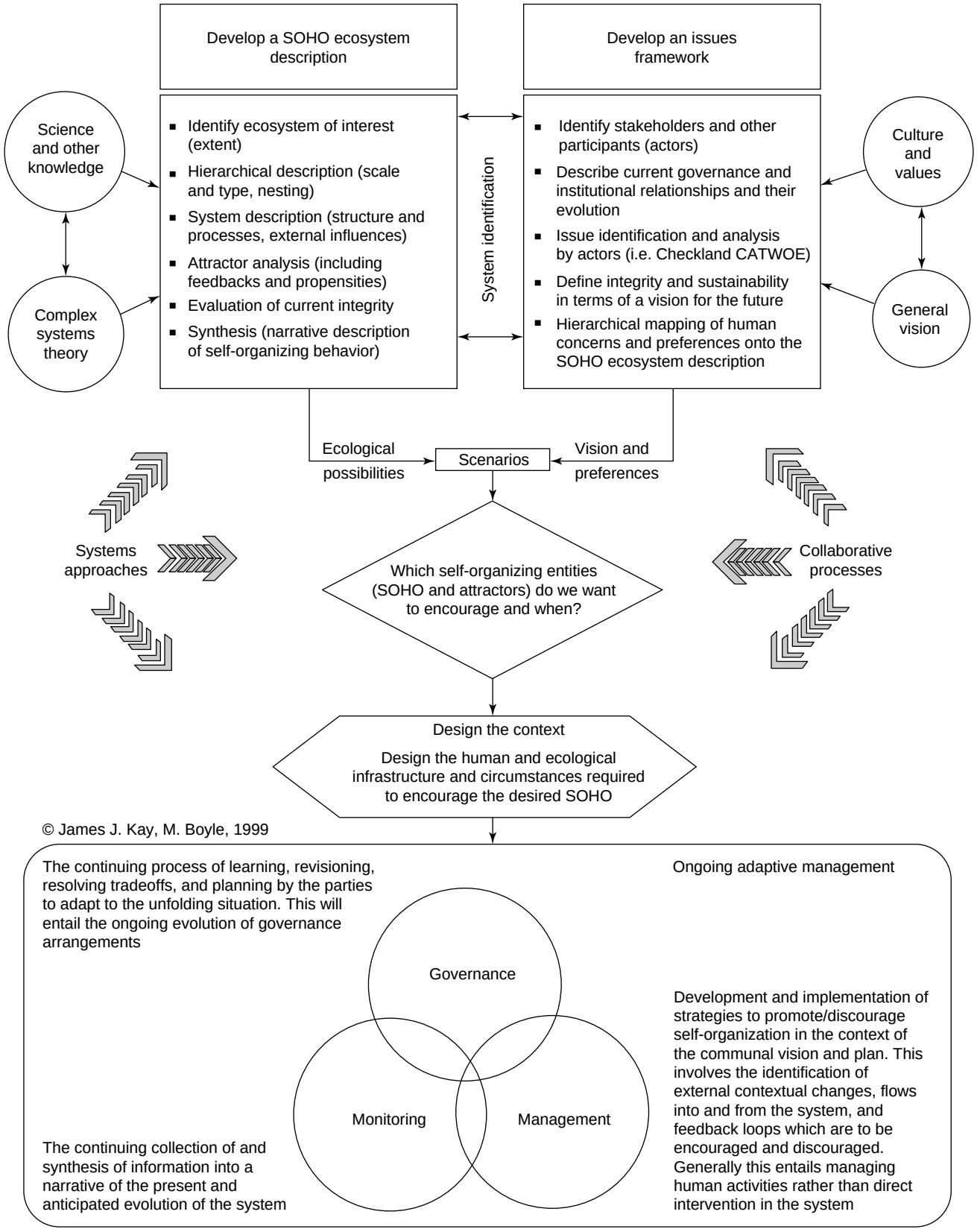


Figure 2 An adaptive ecosystem approach to sustainability and health; SOHO stands for self-organizing holarchic open systems

equals, with others in the process of resolving the vision. Having resolved a community vision for the future, the next phase is to design an adaptive program for the realization of the vision.

This adaptive program consists of a plan and infrastructure for the triad of activities, governance, management and monitoring. *Governance* refers to the continuing process of learning, revisioning, resolving tradeoffs, and planning to adapt to the unfolding situation. The ongoing evolution of governance arrangements and structures is required to accommodate these activities. *Management* is the activity of translating the vision into reality. It involves the development and implementation of strategies to promote or discourage specific forms of self-organization (i.e., trajectories to particular undesirable attractors or system states) in the context of the communal vision and plan. This means maintaining the context for the self-organizing complex SOHO systems, rather than intervening in the system in a mechanical way (Ahl and Allen, 1996). Maintaining the context involves identification of external contextual changes, ows into and from the system, and feedback loops to be encouraged or discouraged. Generally, management concentrates on the relationship between humans and natural ecosystems, and guides the human side of the relationship. Monitoring is the activity of observing the human and natural systems and synthesizing the observations into a narrative of how the situation has actually unfolded and how it might unfold in the future. This narrative is used as the basis for governance and management; that is, for learning, revisioning, and adapting human activities as the human and natural ecosystems co-evolve as a self-organizing entity.

THE ENTERPRISE OF MONITORING

As stated at the beginning of this essay, the traditional view of monitoring tends to focus solely on indicators. By implication, monitoring is restricted to collecting data and reporting. Our position is that it is precisely this narrow interpretation of monitoring that is responsible for its impotence in the service of decision-making and policy. Our thesis is that monitoring, done in isolation from a decision-making process and community needs, is an academic exercise whose utility is left to happenstance. This is why we have gone to some length to set out the context for monitoring. The details of the activity of monitoring follow.

The input to a monitoring program is a statement about: which SOHO ecological-economic system is the focus of attention; what aspects are of importance; how it is hoped that these will change over time (i.e., a narrative of the future); and, probably, a plan or strategy for making these aspirations happen. This statement is an end result of the process depicted in Figure 2 before the box labeled ongoing adaptive management . Such a statement is referred to in this discussion as a *sustainability vision.*

Traditionally, in most Western jurisdictions, the sustainability visions have taken the form of policies and/or regulations, which are enacted as part of a legislative process. In this traditional mode, monitoring, as described herein, is undertaken in the support of policy. However, there are forms of virtual governance emerging which have no formal legislative powers and whose governance does not take the form of traditional policy. For example, United Nations Biosphere Reserves are often overseen by an informal and ad hoc assembly of concerned individuals and non-government organizations. Even though these individuals may have no legal power, they do act as stewards of the Reserves and in effect manage them. While they have no instruments that resemble policy in its traditional sense, they most certainly will have a vision and a plan for the Biosphere Reserve. Similarly, coalitions of concerned citizens can emerge around a sustainability issue (for example, air or water quality issues in a region) and take it upon themselves to make changes in society to alter the current situation. Monitoring, as discussed herein, is meant to be in the service of these initiatives as well as the more traditional policy based approaches to environmental and sustainability issues.

The approach to monitoring that follows is not appropriate for all situations. For example, if one is simply trying to enforce a specific regulation, like water quality standards, then the monitoring exercise is simply one of appropriate surveillance to trigger an enforcement process. Similarly if one is trying to resolve a normal scientific issue, like the effect of a fertilizer on a crop, a different approach to monitoring is appropriate. However by the same token, approaches to monitoring that are appropriate to these situations are not appropriate for dealing with the broader issues of sustainability.

An output of a monitoring program is a narrative of how the organization of the SOHO system has actually changed and how it might unfold in the future. This narrative is compared to the narrative of the sustainability vision. The comparison is used to re ect on all aspects of the sustainability vision and on the appropriateness of the monitoring program itself. This assessment is the basis for learning on the part of those undertaking management and governance. In this way the monitoring program is a re exive activity.

The Monitoring Process

Figure 3 depicts the different activities involved in the process of monitoring. (The activity of developing a monitoring program is discussed below.) It should be expected that as part of an adaptive approach, the monitoring program will need to be revised from time to time. Data

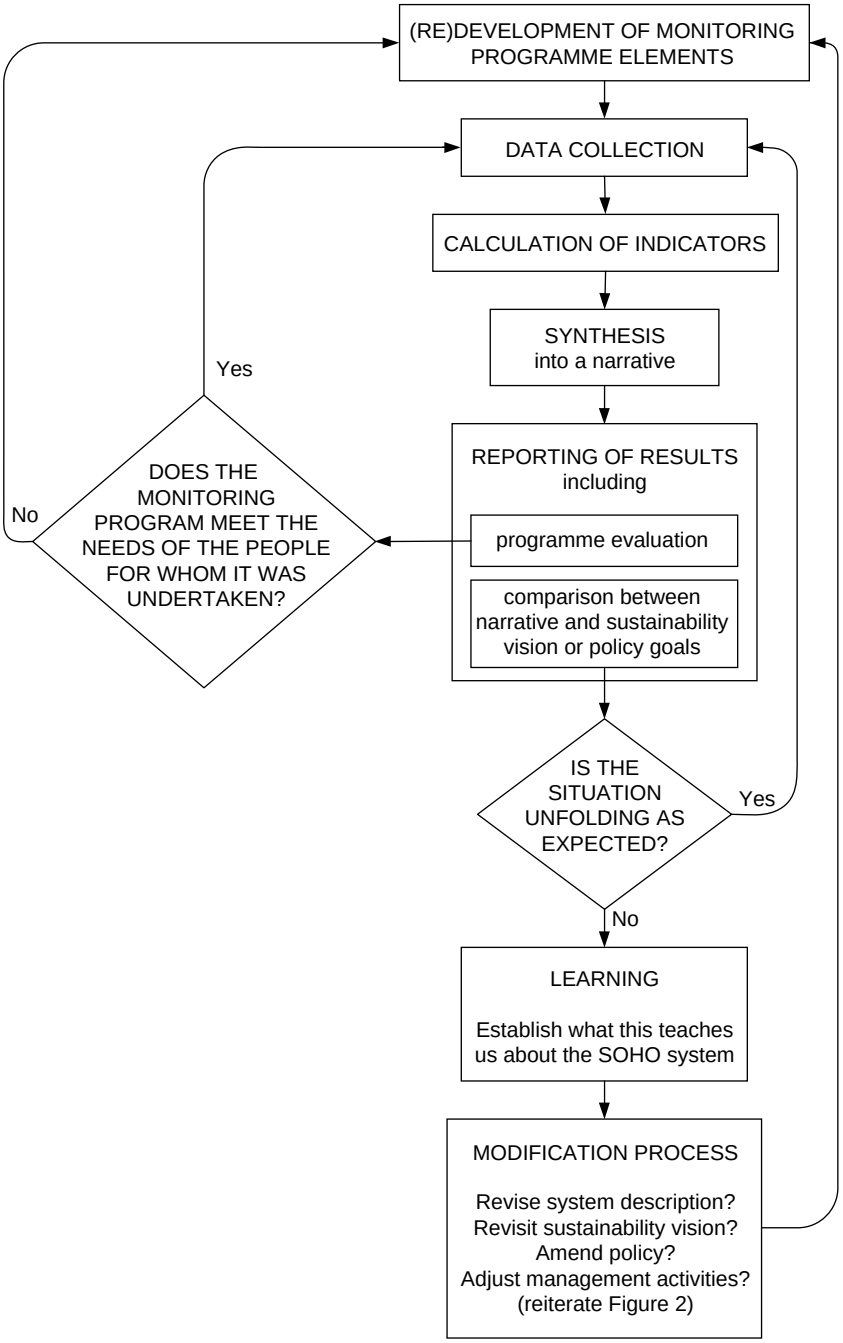


Figure 3 The monitoring process

collection and calculation of indicators are the activities traditionally associated with monitoring and there is an abundant literature on these activities. Synthesis is about generating the narrative about the current and future status of the ecological-economic system and comparing it to the goals and aspirations expressed in the sustainability vision or policy. The next stage is reporting. This is the communication of the findings of the monitoring and assessment to those involved in management and governance. It is a crucial element of the interface between the triad of activities that make up an adaptive approach to sustainability.

The report provides the opportunity to perform two crucial evaluations that are at the heart of the learning process which makes this approach adaptive. One evaluation involves the monitoring program itself: Is it meeting the needs of the people for whom it is undertaken? The other evaluation is an assessment of whether or not the

situation is unfolding as articulated in the sustainability vision or policy.

Given that ecological-economic systems are complex, it is reasonable to expect that the situation is not unfolding in accordance with the sustainability vision or policy. If this is indeed the case, then it should trigger an investigation into what the discrepancies teach us about the behavior of the ecological-economic system as a SOHO system (such discrepancies may herald global environmental change). Discrepancies should also trigger a reevaluation of the policy. Does our description of the system, and what is deemed important, need to be revised? Is the vision still feasible, or is a revisioning exercise in order? Are new management strategies or tactics required? In essence, one should go through another iteration of the process outlined in Figure 2. This reiteration may very likely result in modifications to the monitoring program itself.

Attributes of a Monitoring Program

Based on a review of literature (Boyle, 1998) on monitoring and indicators, and from consideration of the principles of complex systems theory and post-normal science, we have compiled a list of attributes which a comprehensive monitoring program should possess. This list is quite separate from characteristics of the system of interest or from the indicator selection criteria. Our research revealed that there is often confusion between attributes, which apply to a monitoring program or to the set of indicators as a whole, and criteria that serve to evaluate individual potential indicators. (The latter are discussed separately under the section Indicator Development.) The following list of attributes is not only helpful to keep in mind during the design of a monitoring program, but it also becomes an important tool in evaluating whether the program actually provides the required information and meets the needs of the users.

A comprehensive monitoring program should:

- Be relevant and decision supportive. The monitoring program must be relevant to both the issues of concern, and to the users of the information. The information provided to decision makers must be understandable, timely and as forward looking (anticipatory) as possible so that decisions may be made before the system is adversely affected.
- Take into account considerations of scale and type. The monitoring program should provide information at a range of spatial and temporal scales (hierarchically nested) that are appropriate to the vision yet also be sensitive to changes (temporal, spatial and across groups). The program must examine different types or perspectives of the system and consider abiotic, biotic and cultural factors. A separate issue is that data must be collected and reported at appropriate, but not necessarily the same, scales.
- Be based on a conceptual systems model that explicitly recognizes relationships between society and environment. The conceptual model of the SOHO system must be scientifically valid and should re ect the integration between society and the environment in several ways. It must be nested to deal with considerations of scale and type (as above) and be based on ecosystem and institutional boundaries. Human systems must be acknowledged as a subset of the ecological systems that support them and linkages between the economy and the health of ecological systems must be made. It is also important to monitor the feedbacks between society and the environment.
- Allow for an overall integrated evaluation of the system. In order to evaluate the state of the system, the monitoring program should collect a variety of different types of measures. It is important to monitor the sensitivity to magnitude, direction and duration of known stresses, (including socioeconomic and cultural ones) but also to monitor for ecological integrity (which would reveal unanticipated problems). Continuous measurement from stressed to non-stressed conditions tracks the pathways of change between states. An evaluation of the system should be able to assess: its organization (structure and process); states; the quantity and quality of ows of energy and materials; and, whether changes are reversible or controllable. Historical and baseline information can be used to identify natural ranges of variability and trends over time. The monitoring program should thus re ect our knowledge of naturally occurring changes as part of normal system behavior and responses. Cumulative effects should be measured and compared with threshold values for the system. Emergent properties of the system should also be detected and evaluated. Finally, all system assessments should rely on both scientific, objective measures as well as experiential, subjective ones.
- Be adaptive and exible. The program must be adaptive and exible enough to: deal with environmental (including catastrophic) changes; incorporate new information, technology and scientific research; and adjust to changes in the political context or societal values. It is useful to have a monitoring scheme applicable to diverse operational situations at different scales and for different ecological systems, yet the program must also be tailored for specific applications.
- Be practical. A program design is useless if it is not practical and cost effective. Employing existing expertise, data sets and tools (e.g., geographical information systems (GISs) and modeling methodologies) can save time and money and incorporate the work of others.

DESIGN OF THE ELEMENTS OF A MONITORING PROGRAM

Many sources in the literature offer steps or stages to follow in designing a monitoring program. In reality, of course, the process is not linear. Information and knowledge become available piecemeal and many insights are gained along the way. Steps are taken and retaken. We have chosen to separate the design steps and the process of monitoring, which are two quite different activities. A monitoring program, then, is made up of elements (Table 2) that are collectively used in the enterprise of monitoring. Developing the elements of the program should not be confused with the development of the protocol for the activity of monitoring, that is, the monitoring process. The latter is about designing a human activity system as depicted in Figure 3, the former is about designing tools for the analysis of ecological-economic systems.

Table 2 Elements of a monitoring program

An issues framework: the impetus of a monitoring program is always to assess progress towards a set of human goals. Hence, a clear articulation of the goals and users of the information is the foundation of any monitoring program
A conceptual model of the world: the model represents how we look at the world in the context of the goals. It serves to delineate the system that should be monitored. It provides a framework that relates the indicators to each other in the context of the overall system being monitored
A set of indicators: the indicators characterize the system being examined in a meaningful way for the users
A methodology for data collection and storage: carefully laid out procedures to address the practical and technical issues involved in data collection must be established to ensure accuracy, consistency and statistical robustness. Equally important is the storage of the data, so that it is accessible and usable in the future (*see* **Metadata**, Volume 2)
A methodology for calculating indicators: the data collected will have to be manipulated in order to derive values for the indicators. Again, a method to do this accurately, consistently and in a statistically appropriate manner is required
A process for synthesis: synthesizing the information that the indicators provide into an overall narrative of the status of the system is essential to completing the central task of the monitoring program, that is, to assess progress towards the human goals and aspirations which motivated the enterprise
A methodology for reporting: The values of the indicators and the results of the system assessment (the narrative which results from the synthesis) must be reported to the intended audience or users of the information. A methodology for presenting it in a clear, purposeful, and timely manner for decision making is crucial to the utility and success of the monitoring program

The relationship among these elements is shown in Figure 1. Strictly speaking, the issues framework and conceptual model should be developed as part of an overall adaptive ecosystem approach to sustainability and health, rather than as specific elements of a monitoring program. These elements would be produced as part of the analysis of the situation, as described by the right- and left-hand top boxes of Figure 2. They would provide the context for the monitoring program. However, it is usually the case that the development of a monitoring program is not initiated in the context of a larger approach to sustainability. If this occurs, the development of a monitoring program (in particular the development of the first two and last two elements of the program in Table 2) can be used as impetus to stimulate the development of the governance and management activities necessary to complement monitoring.

Issues Framework

The decisions that guide our choices and actions in relation to the environment are inextricably rooted in human concerns that are based upon our worldview and shaped by our culture and values. These considerations determine our general vision of how we want the world to be. In many cases, where a traditional approach is taken, the perspective brought to bear on a problem remains implicit in the problem-solving procedure. The same can be said of decision-making criteria. This fact is often a source of conflict as sustainability is not primarily a technical, scientific, or managerial problem. It is about dealing with people and their diverse cultures, interests, visions, priorities and needs. Pursuing sustainability entails participation of these myriad stakeholders, each having a legitimate perspective, and arriving at a resolution that all can live with. The purpose of the issues framework is to bring the diverse interests, visions, priorities and needs of the stakeholders into focus in an integrated way.

It is difficult to bring stakeholders with diverse beliefs and opinions together to resolve often highly contested and emotional issues. These problems, though significant, are not insurmountable. See, for example, Ramírez (1999) and Burgoyne (1994) for methods in stakeholder analysis and collaboration. Related discussions on public participation in environmental planning may be found in Warriner (1997) and Beckenstein *et al.* (1996). We have found that a variant of the Soft Systems Methodology (SSM) (Bunch, 2000; Checkland, 1981; Checkland and Scholes, 1990; Wilson and Morren, 1990) is a useful way to proceed to develop an issues framework that recognizes and incorporates the perspectives of all stakeholders.

The issues framework is designed to explicitly: (a) identify the actors and stakeholders; (b) describe and define the situation; (c) determine the important issues and concerns; and (d) disclose the values applied in making decisions

and judgments. The issues framework takes the form of a hierarchical mapping of human concerns and preferences onto the SOHO ecosystem description (the conceptual model). Thus, it facilitates comparison of various scenarios that re ect the socioeconomic and ecological possibilities. It defines integrity and sustainability for the situation in terms of a vision for the future. The issues framework lays the foundation for developing a clear statement of vision and articulated preferences in terms of characteristics or features of the desired ecological-economic system. Therefore, the sustainability vision directs the choice of information, which should be collected by the monitoring program.

Conceptual Model

The conceptual model for a specific situation describes how the important considerations (derived from the issues framework) are related in a dynamic system. Such a description requires complex systems thinking. The practical result is a framework for a narrative of the behavior of the ecological-economic system of interest. A comprehensive conceptual model should naturally lead to indicators by effectively telling us where to put meters'.

The explicit development of a conceptual model is crucial for the activities involved in monitoring, management and governance. Our literature review revealed, however, that (a) many monitoring programs do not explicitly define the assumptions underlying the model, and (b) there are many different types of conceptual models in use. (Some examples are outlined below in the subsection entitled Traditional Conceptual Models.) Furthermore, we found that none of the existing conceptual models adequately deals with the complexity of ecological-economic systems. Little attention is paid to the considerations of scale and type, holons, nesting, self-organizing entities, attractors, feedback loops etc. or in short, a systems description to provide the framework for narratives. This means, more often than not, that the monitoring program does not provide sufficient information to generate a system description that will be able to deal with sustainability issues.

Traditional Conceptual Models

A review of traditional conceptual models used in monitoring programs can be found in Boyle (1998). Many current monitoring programs are based on a *stress-response model*. When an ecological system is subjected to a disturbance that it cannot absorb, it will change in some way. The stress-response model (Rapport *et al.*, 1985) attempts to link stresses with responses and therefore provides early warning or predicts how, and when, a system will respond to in uences. There are many examples of indicator frameworks which utilize this model, for example Canada's State of the Environment Reporting Program (Environment Canada, 1995). An improvement in this monitoring approach is the inclusion of *cumulative effects*, that is, compounding effects from more than one source or persistent additions from the same source (Cocklin *et al.*, 1992). Monitoring under the stress-response model is limited, however, to addressing questions regarding known cause and effect relationships in a system. The notions of monitoring for *ecosystem health* and, more recently, *ecosystem integrity* are efforts to monitor and evaluate the status of the system as a whole. *Ecosystem health* monitoring focuses on attributes of the system that give an overall indication of any problems in the system, analogous to taking a person's temperature in the human health model. (See, for example, Breckenridge *et al.*, 1995 for measuring the health of rangelands; and Svanberg, 1996 on ecosystem health indicators for the Baltic Sea.) *See* Ecosystem Health, Volume 4; Ecosystem Integrity, Volume 4.

Evaluations of ecosystem health are based on an assumption of homeostasis (as is human health in a medical context). This assumption is generally not appropriate for ecological systems. A system could completely change from a forest to a wetland because of a river diversion, for example. The system may still be perfectly healthy, only it is now a wetland. To incorporate this range of possible healthy systems, the idea of ecological integrity has emerged to assess the system's crucial characteristic; its ability to reorganize itself (Woodley *et al.*, 1993). Thus, integrity includes responses to a changing environment, or even change from within, as long as its ability to self-organize is maintained (Kay, 1991; Kay and Regier, 2000). Integrity is a concept that has been used most often in monitoring natural and protected area (Noss, 1995; Soulé, 1995; Woodley, 1993).

Straddling the evolution of the ecological conceptual model, the integration of social and economic dimensions has been recognized as important. A shift is also occurring in the human goals that motivate monitoring activities. People are increasingly concerned with living sustainably and wish to assess progress toward this goal. In the last few years, there has been an explosion of sustainability monitoring initiatives (Hodge, 1996; Maclaren, 1995; Nilsson and Bergstrom, 1995; Rees, 1996).

An effort to combine both the social and ecological is the *pressure-state-response model* (Adriaanse, 1993). This model suggests three different, but related and integrated, types of indicators for each sustainability issue. Pressure indicators measure pressures on the environment due to human activities. State indicators measure the quality of the environment, and the quality and quantity of natural resources. Finally, response indicators measure how, and how well, society is doing in response to environmental changes and issues. The advantage of this framework is that it results in a group of indicators for each issue that address the sources of the problem, its current status, and what is being done about it. The Dobriš assessment, used by the

European Environmental Agency, uses the driving forces pressure state impacts response (DPSIR) model of assessment (*see* Dobr s European Environment Assessment Process, Volume 4). This model takes the *pressure-state-response* model one step further by adding the human activities which are the driving forces behind the pressure, and the impacts on human activities and the environment due to the environmental state change (*see* DPSIR (Driving Forces Pressures State Impacts Responses), Volume 4 or the European Environmental Agency WWW site, http://themes.eea.eu.int/showpage.php//?pg• 37183).

The conceptual model we propose integrates these earlier models with the notions of a systems approach to complexity.

The Basis for the Conceptual Model: the Notion of Self-organization

Ilya Prigogine in his Nobel Prize winning work, showed that spontaneous coherent behavior and organization (e.g., tornadoes) can occur and are completely consistent with thermodynamics (Nicolis and Prigogine, 1977; Nicolis and Prigogine, 1989). The key to understanding such phenomena is to realize that one is dealing with open systems with an enduring (not necessarily constant) ow of high quality energy (that is, *exergy*). In these circumstances, coherent behavior appears in systems almost magically. Prigogine showed that this occurs because the system reaches a catastrophe threshold and ips into a new coherent behavioral state. (This is evident, for example, in the vortex, which spontaneously appears when draining water from a bathtub.)

In examining the energetics of open systems, Kay (1984) and Schneider and Kay (1994a) have taken Prigogine's work one step further. An open system with exergy pumped into it is moved away from equilibrium. But nature resists movement away from equilibrium. (This is the second law of thermodynamics restated for non-equilibrium situations.) When the input of exergy and material pushes the system beyond a critical distance from equilibrium, the open system responds with the spontaneous emergence of newly organized behavior that uses the exergy to build, organize and maintain its structure. This dissipates the ability of the exergy to move the system further away from equilibrium. As more exergy is pumped into a system, more organization emerges, in a stepwise way, to dissipate the exergy. Furthermore these systems tend to get better and better at grabbing resources and utilizing them to build more structure, thus enhancing their dissipating capability.

The theory of non-equilibrium thermodynamics suggests that the self-organization process in systems proceeds in a way that: a) captures more resources (exergy and material); b) makes more effective use of the resources; c) builds more structure; d) enhances survivability. These seem to be the kernel of the propensities of self-organization in systems. How these propensities manifest themselves as positive and/or negative feedback loops is a function of the given environment (context) in which the system finds itself embedded, as well as the available materials, exergy and information.

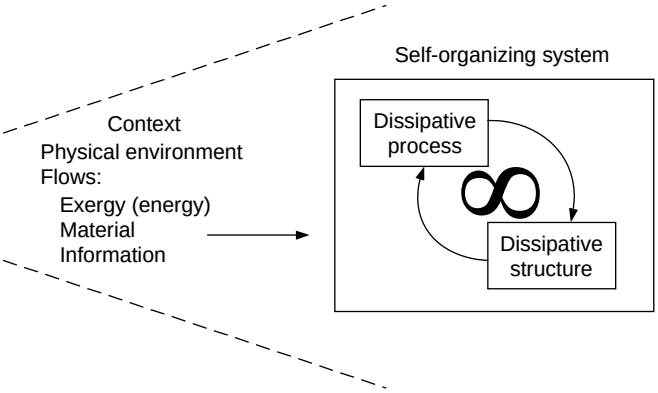


Figure 4 Complex systems as dissipative systems

This conception of self-organization, as a dissipative system, is presented in Figure 4. Self-organizing dissipative processes emerge whenever sufficient exergy is available to support them. The details of the processes depend on the raw materials available to operate them, the information present to catalyze the processes, and the physical environment. The interplay of these factors defines the context for, or the constraints on, the set of processes which may emerge. (Generally speaking, the specific processes that emerge from the available set are uncertain.) Once a dissipative process emerges and becomes established, it manifests itself as a structure.

(In the case of a vortex in the bathtub water, the exergy is the potential energy of the water, the raw material is the water and there is no historical information; the dissipative process is water draining, and the dissipative structure is the vortex. The vortex will not form until enough height of water is in the bathtub and if too much height of water is present, the vortex ips into a different form of ow.)

In an ecological setting, examples of the structures are the individuals of species, breeding populations, forests etc. The processes are reproduction, metabolism, evapotranspiration, etc. The context is the available set of nutrients and energy sources in a physical environment. The information includes the genetic codes and biodiversity at various scales.

The Model

We have integrated the traditional conceptual models with the notion of ecological-economic systems as dissipative complex systems to produce a conceptual model that serves as a monitoring measurement framework (refer to Figures 5–7). In this model, the elements of the landscape (e.g., woodlots, wetlands, farms, neighborhoods) that

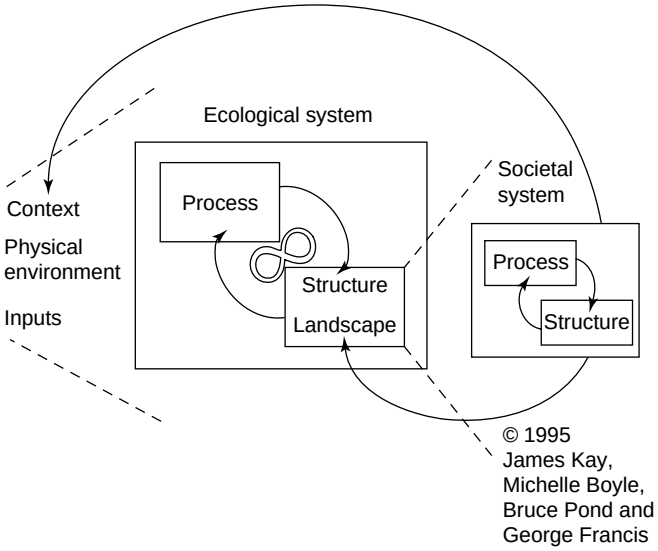


Figure 5 Conceptual model; single horizontal holarchical level

make up the societal and ecological systems, are seen as self-organizing entities set in an environmental context. Self-organizing entities are understood through consideration of their constituent processes and structures, and the relationships between these. (For example, in a woodlot: processes would be evapotranspiration and growth of biomass; structure would be the species that make up the woodlot; and a description of the relationship between these processes and structure would be Holling's four-box model of the ongoing process of birth, growth, death and renewal in ecosystems (Holling, 1986, 1992)). The processes involve the ows of material, energy and information. The structures are the objects (i.e., trees) we see on the landscape. The processes allow for the emergence, and support, of structures, which in turn allow for the emergence of new processes and so on. The recognition of this recursive relationship between process and structure separates this conceptual model from more traditional ones.

Once the relevant processes and structures have been identified, a suite of variables may be specified which can be used to monitor the state of the processes and structures. These variables must be sufficient to define the wellbeing (in the case of ecosystems, the integrity (Kay, 1991; Kay and Regier, 2000)) of the self-organizing entity.

Ecological communities provide exergy, materials and information required for human societies to sustain themselves. This is depicted in Figure 5. The societal system depends on the ow of exergy, materials and information from the ecological system to support its processes and structures. These ows, along with the biophysical environment provided by the ecological systems, are the context for societal systems. The context constrains the possible societal processes and structures in a specific location. While Figure 5 illustrates a single ecological system providing the context for the societal system, the reality is that it is a suite of adjacent ecological systems (for example: woodlots, fields and wetlands adjacent to a farm) that provide this context. Alterations in these adjacent systems will alter the context and thus the possibilities for the system in question.

However, the societal system can also in uence the ecological system in two ways. The first in uence is through changes in the structure of the ecological system (for example: cutting trees down in a woodlot, filling wetlands, and all the human activities which involve removing or dismantling ecological structures on the landscape). Such actions, of course, alter the ows from the ecological systems to the societal systems and thus create a feedback structure on the landscape. This is represented in Figure 5 by the lower arrow back from the societal system to the structure in the ecological system. The feedback to the societal system occurs because changes in the ecological structure change the context for the societal system.

The second in uence occurs when the context of the ecological system is altered by the societal system. For example, the runoff into a wetland or stream may be altered by human activities on adjacent properties. It is depicted in Figure 5 by the upper arrow from the societal system back up to the context of the ecological system. This in uence is qualitatively different than the structural in uence just discussed.

The resulting feedback loop has more steps and accordingly is more indirect. By changing the context of the ecological system, the societal system affects the ecological processes and, in turn, the ecological structure and ultimately the societal system's own context. For example, modifying the runoff into a waterway can dramatically alter the character of the waterway, and hence the type of fish found in it and therefore the sport fishery and associated economic system.

To summarize this discussion, each self-organizing entity resides in an environment that provides: (a) the biophysical surroundings in which the entity exists; and, (b) ows of exergy, material and information that the entity depends upon for the continuation of the self-organizing processes, which maintain its structure. The biophysical surroundings, in conjunction with the ows into the system, constitute the context for the self-organizing entity. It is necessary to monitor this context, since changes in it will effect the processes of the entity.

Referring to Figure 5, the relationship between societal systems and ecological systems is three-fold:

1. Ecological systems provide the context for societal systems. That is, they provide the biophysical surroundings and ows of exergy, material and information that are required by the self-organizing processes of the societal systems.

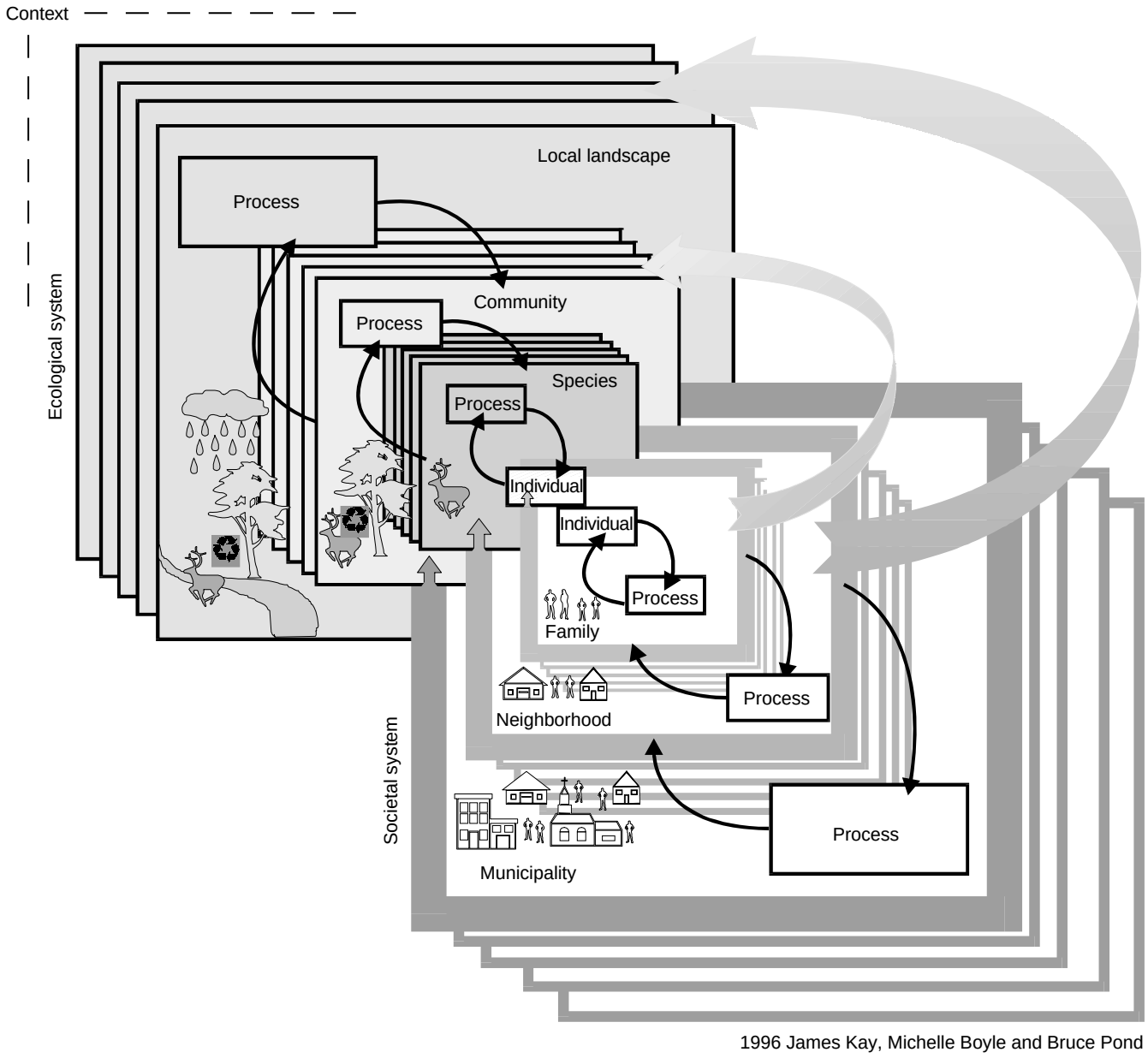


Figure 6 Conceptual model: three levels

2. Societal systems can alter the structures in ecological systems. (For example, cutting down a woodlot, removing beaver from a watershed.) Changes in the ecological structure can then, of course, alter the context for the societal systems themselves.
3. Societal systems can alter the context for the self-organizing processes of ecological systems, e.g., a change in the drainage patterns into a wetland, or a change in the nutrient runoff into a shallow lake from adjacent agricultural fields. Changes in ecological processes can alter ecological structure and consequently the context for societal systems.

Each of these represents a qualitatively different aspect of the relationship between ecological and societal systems that should be considered when monitoring.

The Hierarchical Structure of the Model

Figure 5 applies to one hierarchical level, but as observed earlier, sustainability issues can only be understood in terms of nested holons. Figure 6 illustrates this idea. On the ecological side, *local landscape* can be thought of as a subwatershed, for example. The hydrological cycle is an example of a process in the subwatershed. The structures which make up the subwatershed are the ecological

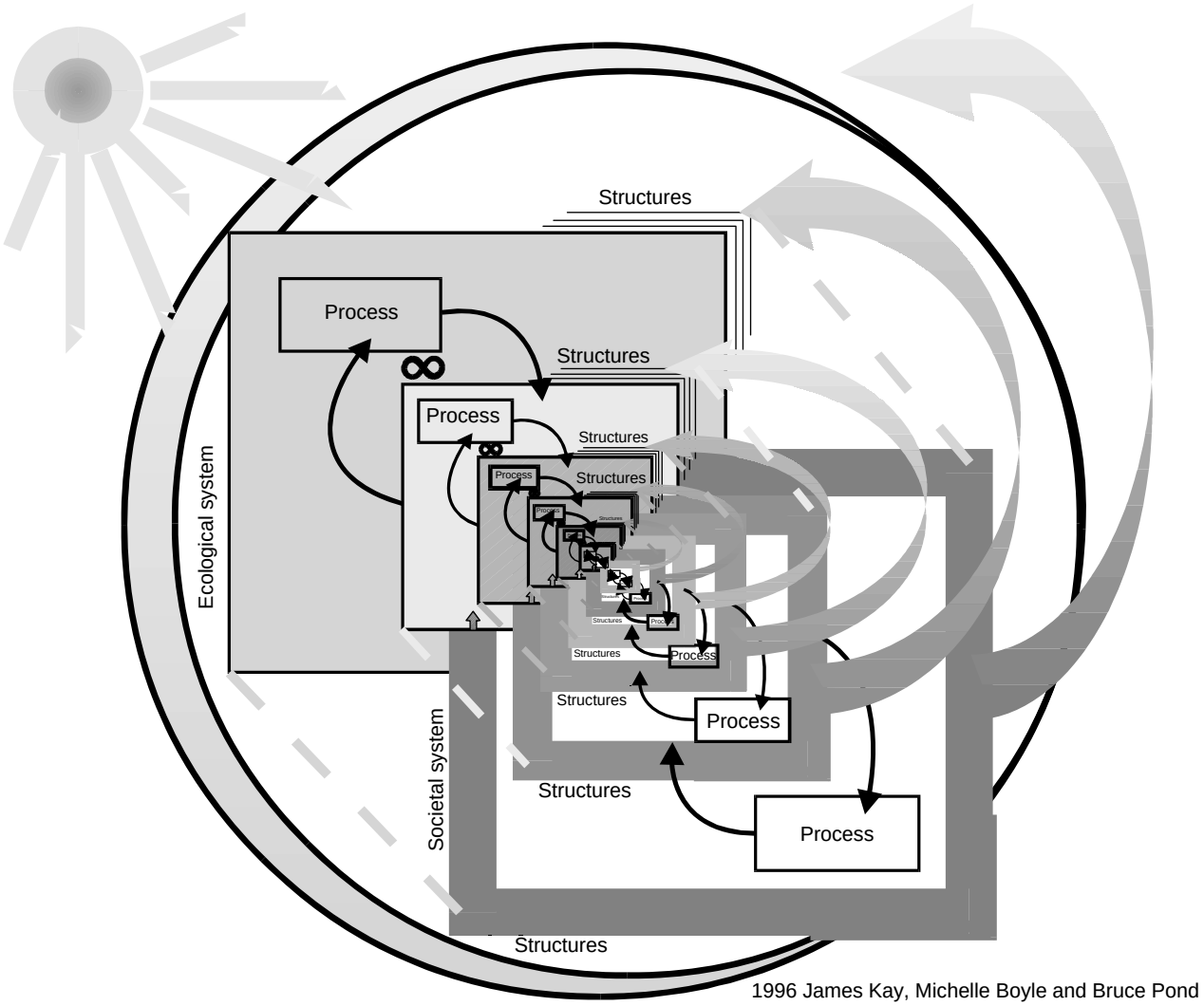


Figure 7 Conceptual model: the big picture

communities (e.g., woodlots, wetlands, open fields, etc.). The communities are in turn made up of species. On the societal side, municipalities are scattered across the local landscape. These in turn are made up of neighborhoods, which are made up of families and businesses.

In many cases, the local subwatershed defines the context for the local municipality. However, the municipality can, and does, directly modify the ecological communities in the subwatershed and thus its own context. Similarly, the context for local neighborhoods is determined by the adjacent ecological communities. Nonetheless, the local neighborhood is quite capable of in uencing ecological communities, through direct structural change (such as harvesting wood from a woodlot), or changing the context of an ecological community (for example, changing drainage patterns into a wetland).

Figure 6 and the examples are meant to be illustrative and not exhaustive, although they do demonstrate that such changes can cascade through the nested holons to ultimately affect individual families and businesses. What is unique about this conceptual model is that it explicitly identifies these connections and relationships and provides a framework for deciding what to monitor.

Figure 7 shows the full conceptual model, which would be used as a template to develop a particular conceptual model in which the important levels, processes, structures, environmental context, and in uence/feedback considerations are specifically identified.

The Conceptual Model and Monitoring

This conceptual model identifies a set of aspects which should be monitored, that is, should have indicators associated with them. These are summarized in Table 3. In effect, Figures 5–7 define different points where a monitoring

Table 3 Aspects of the SOHO system which should be monitored

(a)	State of the wellbeing of the societal and ecological entities:
	measures of self-organizing processes
	– *examples: evapotranspiration, recycling*
	measures of self-organizing structure
	– *examples: biomass, species diversity*
(b)	The context of the societal entities:
	measures of biophysical surroundings
	– *example: average humidity*
	measures of flows into the system
	Exergy: – *example: fuel wood consumed*
	Materials: – *example: aggregate consumption*
	Information: – *example: flow of hormone like chemicals from the environment into food*
(c)	Influences of the societal systems on ecological systems:
	Structural influences
	– *examples: wetlands destroyed, woodlots removed*
	Contextual influences
	measures of biophysical surroundings
	– *example: average temperature*
	measures of flows into the system
	Exergy: – *example: phosphorous loading*
	Materials: – *example: drainage flows*
	Information: – *example: hormone like chemicals into the food chain*

meter should be plugged into the ecological-societal system. A monitoring program should include:

1. Measures of the state of wellbeing of each of the self-organizing entities (ecological and societal). These are represented in the figures by process and structure boxes.
2. Measures of the context for the societal system, that is, the biophysical surroundings and the ows into the societal system from the ecological system. This is depicted in the figures as the societal system in the shadow of the ecological systems.
3. Measures of the contextual and structural in uences of the societal systems on the ecological. These are represented in the figures by the arrows going from the societal system, up one level, and into the ecological structure. These arrows show how societal in uences on the ecological systems, through changes that these in uences cause in the process and structure of the ecological system, can alter the context for the societal system. So the societal in uences on ecological systems are part of a feedback loop which affects the societal systems.

Sustainability is about maintaining the well-being of the combined ecological–societal system. This requires that its self-organizing processes and structures be maintained. This will happen naturally, if we maintain the context for self-organization in ecological systems, which in turn will maintain the context for the continued well-being of the societal systems. This condition for sustainability is incorporated in the conceptual model.

Thus, this conceptual model provides a framework for monitoring for sustainability that is more elaborate than conventional approaches. Traditionally, one would monitor external (usually negative) stresses in uencing the system and its response. In this model, we emphasize the importance of the hierarchical monitoring of:

- both process and structure and their inter-relationship in order to understand the current wellbeing of the system;
- the context of the ecological and societal system in terms of the biophysical surroundings of each, and the ow of exergy, materials and information into the system;
- the contextual and structural in uence of societal systems on ecological systems and hence ultimately on the societal system itself.

Without consideration of each of these aforementioned aspects, we cannot decide where indicators may be placed to most effectively fulfil their purpose. For example, if the intent of a monitoring program is to anticipate future changes in order to make wise management decisions in the present, indicators must be chosen to monitor the factors which in uence the processes and structure of the system. Measures focusing on the current state of the system will only reveal changes that have already occurred and will thus be too late. This is particularly true with regards to monitoring to detect the effects of global change.

Also, monitoring within the boundaries of the defined system of interest is not sufficient to gain an understanding of its dynamics or to anticipate changes, both of which are extremely important in protecting areas or certain land uses. We must broaden our scope to examine the context that constrains or in uences the behavior of the system, and do this from a hierarchical perspective.

When using this model, it is important to bear in mind that self-organization inherently involves internal causality. A self-organizing system has the ability to maintain itself at an attractor despite changes in its environment. So, the environment may change substantially without the system exhibiting major change. Self-organizing systems can respond in a synergistic way to a collection of environmental changes. Thus, changes in the system can rarely be tied unequivocally to a specific environmental, policy or management change. Furthermore, such systems have the capability to generate new behavior that may emerge independently of changes in the environment. The capacity to organize and maintain itself about an attractor is the hallmark of a self-organizing system. It is important to remember that discussions of the behavior of this model cannot

be in terms of one-to-one linkages of contextual elements (i.e., stresses) to the state of processes or structures. Rather, the explanation must be in terms of morphogenetic mutual causality, that is, narratives (Kay and Regier, 1999).

Indicators and their Development

Indicators are often confused with a monitoring program. They are an important part of monitoring, but they must be generated, selected and implemented within the context of all the elements of a monitoring program. Indicators describe the status of different attributes of a system. However, the conceptual model is required to provide a framework that integrates the indicators. As well the data collection and calculation methodologies and a process for synthesizing all the information into a narrative of the system are required if the indicators are to be of any use. Without these elements of a monitoring programme, the indicators cannot be used to describe and assess the overall status of the ecological-economic system in question.

Table 4 Indicator development activities

Generate indicator selection criteria: selection criteria capture the issues and constraints regarding the desired set of indicators are used to screen potential indicators. They must be agreed upon by all those participating in the development of the monitoring program elements
Generate potential indicators: a set of potential indicators consists of possible measures, which emerge from consideration of the issues without regard for constraints. All potential indicators should be recorded and evaluated
Select indicators: Not all potential indicators will meet all the criteria equally well and there will always be trade-offs when choosing some indicators over others. A method must be in place to assess the feasibility of potential indicators and to make a subjective judgement on their evaluation against the criteria

Indicator development consists of three activities: enumerating indicator selection criteria, generating potential indicators and selecting indicators to be implemented for the monitoring program. These activities are summarized in Table 4.

Figure 8 is schematic for organizing information while generating criteria and potential indicators. It delineates a set of considerations that directly or indirectly in uence indicator selection, as well as the relationship between them. For each consideration (box, in Figure 8), criteria for indicator selection and questions to guide the task of generating indicators are formulated. An explanation of the diagram follows, while compiled lists of selection criteria and questions for indicator generation can be found in Tables 5 and 7, respectively.

The development of indicators is motivated by the need to build a narrative of how the ecological-economic system is developing so as to guide human activities. The *sustainability narrative* articulates a vision for the desired development of the system and identifies issues or concerns of import to the users of the indicators. This consideration defines *what we want to know*. As such, it also serves as the entry point for incorporating, into the choice of indicators, changes (over time) in the issues or concerns of users.

The *conceptual model* defines the structural and functional components of the biophysical and societal systems

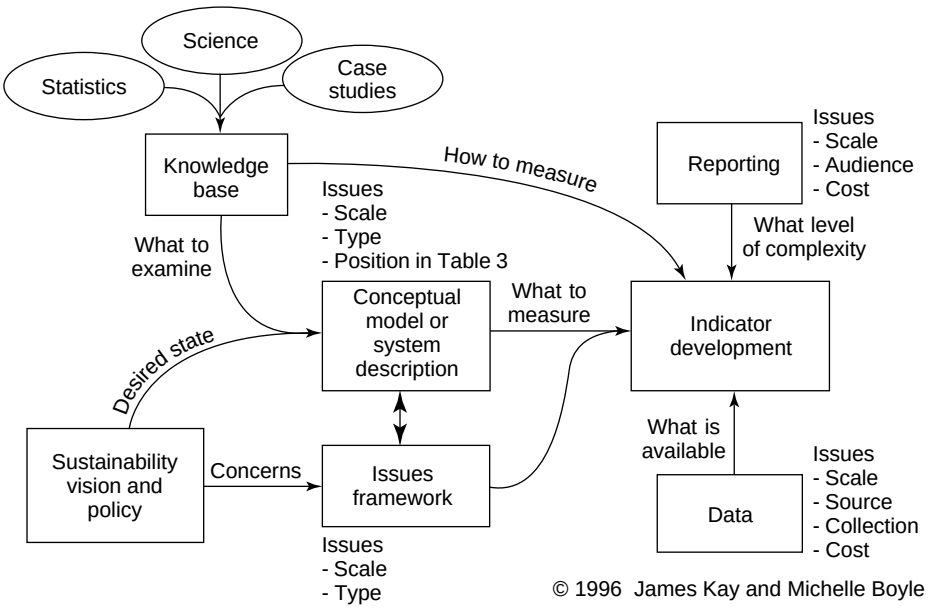


Figure 8 Considerations for indicator development. Legend: • Influence on development of indicators

Table 5 Indicator selection criteria (refer to Figure 8)

Categories of consideration	Indicator selection criteria
Conceptual model of the system	Indicators should: clearly relate to specific societal or ecological elements of the SOHO description
Issues framework	Indicators should be: clearly relevant to specific articulated goals and objectives for sustainability
Knowledge base	Indicators should be: scientifically valid; statistically and analytically sound; demonstrated by case studies to be practical
Data	Indicators should use data that are: available and accessible (easy to obtain and maintain at a reasonable cost); accurate; comparable over time; complete with historical information (to establish baseline conditions and thresholds)
Reporting	Indicators should provide information that is: understandable to potential users; unambiguous; easy to use; at the appropriate scale for decision-making

Table 6 Indicator category

Level I: indicators available for immediate use All essential criteria are met
Level II: indicators with demonstrated potential for use Most criteria are met but, for example, data are not currently being collected, or there is a lack of historical data
Level III: possible indicators There is evidence that these measures would be worthwhile indicators, but further scientific research and/or case studies are required to confirm their utility

Table 7 List of questions for potential indicator generation

(1) What is the sustainability goal or objective?
(2) What are the potential indicators?
(3) Which aspect of the conceptual model does each indicator measure?
(4) At what scale(s) will the indicators be reported?
(5) What is the rationale for each of these indicators (i.e., what does it tell us about the sustainability goals and issues)?
(6) Are there any confounding factors or problems which may limit the use of the indicator?
(7) Data sets: what measurements are required? At what scale will the data be collected? Who is responsible for data analysis and reporting?
(8) Data sources: who has the data set? Who is responsible for its collection?
(9) Which category does the indicator fall into (Level I, II or III)?

and describes the linkages and feedbacks between them. It provides a reference frame, which identifies the appropriate scales at which monitoring should occur and for which indicators are needed. It also provides a framework which shows the relationships between the indicators and some sense of how useful or timely the resulting information will be. The *issues framework* extracts specific management or user concerns and arranges them into a hierarchical framework to associate them with the appropriate scale and type. Thus, the conceptual model and the issues framework are highly interdependent and both facilitate the decision of *what to measure*.

Our *knowledge base* also determines what we elect to monitor. Science, from the perspectives of biology, ecology, geology, hydrology, etc. aids in understanding the processes and structure of the system and the effects of specific in uences or stressors. Complex systems theory adds valuable insight about how we perceive the system and what questions to ask when approaching sustainability issues. Statistics offers rules and standards for sampling and measurement as well as techniques for data analysis and transformation that ensure quality results. Case studies allow us to test chosen indicators and evaluate their utility and feasibility in a practical setting. The amalgamation of these is the body of knowledge that helps us not only to judge *what things are important to examine*, but also *how to make measurements* and *calculate indicators accurately and consistently*.

There are also many considerations attached to the *data* themselves that affect which indicators are chosen. *Data availability* encompasses considerations related to cost, collection, maintenance of databases, accessibility, and who is responsible for each task. The amount (i.e., historical extent) and the quality of data available are also important.

Finally, the development of indicators must be done with an awareness of the *reporting* requirements and the needs of the users. This category provides parameters for the *appropriate level of detail* and *the relative number of indicators*, which affects the aggregation and/or integration of data to compose the indicators.

It should be noted that there are three independent considerations of hierarchy, which are relevant when developing indicators for a monitoring program. We must be concerned with:

1. the scale and type of focus for the situation of interest;
2. the scale of data collection;
3. the scale of reporting the results.

The scale and type of focus may be ascertained from the specific sustainability issue and is re ected in the issues framework. The scale of data collection need not have anything to do with the level of focus for the sustainability issue since information can be aggregated, desegregated, or integrated to be meaningful for decision-making. Likewise, the results may be reported at several different scales.

Indicator Selection Criteria

Using the above considerations as a framework and drawing upon the literature (Boyle, 1998), we generated a list of criteria (Table 5) to be used in evaluating potential indicators. These are properties *each* indicator should possess. Note that these are relatively generic criteria, which apply to many different monitoring initiatives. For a particular project, however, there would be project specific criteria that would be added to this list. Keep in mind that some important characteristics of the information that the monitoring program should provide (for instance anticipatory' or monitoring for specific threats') need not be a quality of every indicator.

There are some criteria that, while desirable, are not necessarily essential. Thus tradeoffs are inevitable. For example, an otherwise excellent indicator may require data which are expensive to obtain, or for which the necessary historical data are not available. Thus, the list of indicator criteria may be used as a checklist to assist in indicator selection. Judgement, and careful consideration of each indicator, is still required. The use of indicator categories is one way to accomplish this.

Indicator Categories

The purpose of grouping possible measures into categories is to distinguish between indicators that are immediately useful (Level I) and those that could potentially be used in the future (Levels II and III). Funding limitations, unavailable data or untested measures are all considerations, which would restrict the immediate implementation of an indicator. Those indicators placed in Levels II and III become a wish list that may become feasible at a later time. This method ensures that these potentially valuable indicators are not completely discarded. For some, the case could be made that the benefit derived from the indicator is worth the extra expense or research required in order for it become a Level I indicator. The final set of indicators is chosen after detailed feasibility investigations for each indicator have been conducted and tradeoffs have been considered. Table 6 defines the three indicator categories.

Note that potential indicators are associated with indicator categories on the basis of an initial assessment. Only those indicators that are known to reasonably meet all the criteria and could be immediately implemented are placed in Level I. If there is some uncertainty about an indicator meeting the criteria, or it is known to be lacking data or scientific validity, it is placed in Level II and III for further investigation. Thus, some potential indicators in Levels II and III may be discovered to be, or developed into, Level I indicators at a future date.

Guiding Questions to Develop Potential Indicators

The considerations described in Figure 8 are synthesized, in Table 7, into a list of questions that will assist in the generation of potential indicators. We have found such a set of questions to be very useful in guiding the process of indicator development and encourage those involved in developing indicators to generate a set of guiding questions appropriate for their own circumstances.

Designing the Remaining Elements

We brie y discuss the design of the remaining elements of the monitoring programme (see Table 2) as they are either very case specific or are the subject of much literature in their own right.

A Methodology for Data Collection and Storage

Carefully laid out procedures that address the practical and technical issues involved in data collection must be established to ensure accuracy, consistency and statistical robustness. This is especially important when different people will be collecting the data over time. An emerging design concern is the technical expertise of those collecting data. Increasingly community groups and volunteers are engaged in data collection. The reliability of the data then, very much depends on the methodology being designed to allow the keen, but technically unsophisticated, to collect the data in ways which are replicable. In any case the design of the methodology must re ect the capabilities of the users of the methodology, if the resulting data is to be useful.

Equally important is the storage of the data, so that they are accessible and usable in the future. Much has been written on this subject, but in essence there must be a way of recording the information that allows someone in the future

(typically a decade) to understand what measurements were taken in the field and to be able to confidently assess their reliability and to repeat them. Failure to do this has often resulted in data being rejected simply because their reliability is unknown, thus effectively wasting the time, effort, and money spent in collecting the data in the first place (Wiersma, 1995). *See also* Data Banks: Archiving Ecological Data and Information, Volume 2; Metadata, Volume 2.

A Methodology for Calculating Indicators

The data collected will have to be manipulated in order to derive values for the indicators. Again, a method to do this accurately, consistently and in a statistically appropriate manner is required. This method must be easily applied by those doing the calculations.

A Process for Synthesis

Synthesizing the information that the indicators provide, into an overall narrative of the status of the system, is essential to completing the central task of the monitoring program, that is, to assess progress towards the human goals and aspirations which motivated the enterprise. The synthesis should in effect bring the conceptual model to life. How one does this is an art form akin to story telling. As such, the form of the narrative is very much dependent on the audience and thus is very case specific.

A Methodology for Reporting

The values of the indicators and the results of the system assessment must be reported to the intended audience or users of the information. While it seems trite to say this, a methodology for presenting it in a clear, purposeful, and timely manner for decision-making is crucial to the utility and success of the monitoring program. Many a monitoring report has collected dust on a shelf because the authors paid no attention to the issue of packaging and marketing it. In fact, generating the report should be treated as a marketing exercise and it should be recognized that a number of different reporting products may be necessary for different audiences. With the advent of the World Wide Web and browsers with capabilities to access GIS and computer-aided drafting software in quite transparent ways, some very exciting possibilities for reporting the results of monitoring programs are emerging (*see* Geographic(al) Information Systems (GIS), Volume 4).

IN CLOSING

We have argued that monitoring needs to be part of an adaptive ecosystem approach if it is to contribute to the sustainability of ecological-economic systems. Monitoring provides feedback to the two other elements of an adaptive ecosystem approach, governance and management. At the core of the approach is a conceptual model (describing the situation in terms of a SOHO systems description) that addresses the issues of complexity, and an issues framework that integrates the hopes, aspirations and concerns of the various actors into a working definition of sustainability. A central tenet of the ecosystem approach is that the path to sustainability is one of tradeoffs. Science can illuminate the tradeoffs but a resolution, that is, the choice of path, is a political decision and in the domain of governance. Thus, an ecosystem approach must bring together science and the socio-political economic realities of the situation in a synthetic and systematic way.

In an adaptive ecosystem approach, governance is seen as the ongoing process of resolving the tradeoffs and charting a course for sustainability. Management is about operationalizing sustainability. Monitoring provides feedback on how the situation is actually unfolding and where it appears to be going.

An adaptive ecosystem approach cannot be undertaken without a monitoring program. The reality of irreducible uncertainty, associated with complex situations, limits our ability to predict how the situation will unfold, and hence where the path to sustainability lies. Thus, any successful approach to sustainability must not only employ traditional anticipatory methods but also methods for adapting to the way in which the situation actually unfolds. The feedback provided by the narrative, which results from a monitoring program, is the source of learning that is required for a successful adaptive approach to sustainability.

In this context the monitoring program must be capable of providing a narrative description of how the ecological-economic system is developing and the implications of this for the path to sustainability. To do this requires collecting information that is clearly tied to the SOHO conceptual model and the issues framework, and reporting this information in a form which can be used for governance and management. Indicator development must be undertaken in this broader system's context, and not as a purely scientific exercise. Otherwise, it is more likely than not to produce yet another set of measures which satisfy disciplinary curiosity, but provide little input to the process of resolving the ecological-economic challenges confronting our society.

For further information on non-equilibrium dynamics of complex systems, *see*: Chaos and Predictability, Volume 1; Chaos and Cycles, Volume 2; Non-equilibrium Ecology, Volume 2; Public-driven Response to Global Environmental Change, Volume 4.

REFERENCES

Adriaanse, A (1993) *Environmental Policy Performance Indicators - a Study of Indicators for Environmental Policy*

in the Netherlands, Sdu Uitgeverij Koninginnegracht, The Hauge.

Ahl, V and Allen, T F H (1996) *Hierarchy Theory: a Vision, Vocabulary and Epistemology*, Columbia University Press, New York.

Beckenstein, A, Long, F, Arnold, M, and Gladwin, T (1996) *Stakeholder Negotiations: Exercises in Sustainable Development*, Mirror Higher Education Group, Chicago, IL.

Boyle, M (1998) *An Adaptive Ecosystem Approach to Monitoring*, University of Waterloo, Ontario, URL ersserver.uwaterloo.ca/jjkay/grad/mboyle/.

Breckenridge, R P, Kepner, W G, and Mouat, D A (1995) A Process for Selecting Indicators for Monitoring Conditions of Rangeland Health, *Environ. Monit. Assess.*, 36, 45–60.

Bunch, M J (2000) *An adaptive ecosystem approach to the rehabilitation and management of the Cooum River in Chennai, India*, University of Waterloo, Ontario, URL, ersserver.uwaterloo.ca/jjkay/grad/mbunch.

Burgoyne, J G (1994) Stakeholder Analysis, in *Qualitative Methods in Organizational Research*, eds C Cassell and G Symon, Sage Publications, London, 187–208.

Carpenter, S R and Cottingham, K L (1997) Resilience and Restoration of Lakes, *Conserv. Ecol.*, 1(2), URL, www.consecol.org/vol1/iss1/art2.

Checkland, P (1981) *Systems Thinking, Systems Practice*, John Wiley & Sons, Toronto.

Checkland, P and Scholes, J (1990) *Soft Systems Methodology in Action*, John Wiley & Sons, Toronto.

Cocklin, C, Parker, S, and Hay, J (1992) Notes on Cumulative Environmental Change 1 : Concepts and Issues, *Environ. Manage.*, 35, 31–49.

Environment Canada (1995) *State of the Environment Report*, Environment Canada, URL, http://www.ec.gc.ca/soer-ree/English/1996Report/Doc/1–3-1.cfm.

Funtowicz, S and Ravetz, J (1993) Science for the Post-normal Age, *Futures*, 25, 739–755.

Gunderson, L H, Holling, C S, and Light, S S (1995) *Barriers and Bridges to the Renewal of Ecosystems and Institutions*, Columbia University Press.

Hodge, R A (1996) Indicators and their Role in Assessing Progress toward Sustainability, in *Geoindicators - Assessing Rapid Environmental*, eds A R Berger, W J Iams, and A A Balkema, Rotterdam, 19–26.

Holling, C S (1978) *Adaptive Environmental Assessment and Management*, John Wiley & Sons, Chichester.

Holling, C S (1986) The Resilience of Terrestrial Ecosystems: Local Surprise and Global Change, in *Sustainable Development in the Biosphere*, eds W M Clark and R E Munn, Cambridge University Press, Cambridge, 292–320.

Holling, C S (1992) Cross-scale Morphology, Geometry, and Dynamics of Ecosystems, *Ecol. Monogr.*, 62, 447–502.

Kay, J J (1991) A Non-equilibrium Thermodynamic Framework for Discussing Ecosystem Integrity, *Environ. Manage.*, 15, 483–495.

Kay, J and Regier, H (1999) An Ecosystem Approach to Erie's Ecology, in *The State of Lake Erie (SOLE) - Past, Present and Future. A tribute to Drs. Joe Leach & Henry Regier*, eds M Munawar, T Edsall, and I F Munawar, Backhuys Academic Publishers, The Netherlands, 511–533.

Kay, J, Regier, H, Boyle, M, and Francis, G (1999) An Ecosystem Approach for Sustainability: Addressing the Challenge of Complexity, *Futures*, 31, 721–742.

Kay, J J and Regier, H (2000) Uncertainty, Complexity, and Ecological Integrity: Insights from an Ecosystem Approach, in *Implementing Ecological Integrity: Restoring Regional and Global Environmental and Human Health*, eds P Crabbé, A Holland, L Ryszkowski, and L Westra, Kluwer, The Netherlands, 121–156.

Kay, J J and Schneider, E D (1994) Embracing Complexity, The Challenge of the Ecosystem Approach, *Alternatives*, 20, 32–38.

Lister, N and Kay, J J (2000) Celebrating Diversity: Adaptive Planning and Biodiversity Conservation, in *Biodiversity in Canada: an Introduction to Environmental Studies*, ed S Bocking, Broadview Press, Toronto, 189–218.

Ludwig, D B, Walker, B, and Holling, C S (1997) Sustainability, Stability, and Resilience, *Conserv. Ecol.*, 1(7), URL, www.consecol.org/vol1/iss1/art7.

Maclaren, V (1995) *Developing Indicators of Urban Sustainability: A Focus on the Canadian Experience*, ICURR Press, Toronto.

Nicolis, G and Prigogine, I (1977) *Self-Organization in Non-equilibrium Systems*: John Wiley & Sons, New York.

Nicolis, G and Prigogine, I (1989) *Exploring Complexity*, W H Freeman, NY.

Nilsson, J and Bergstrom, S (1995) Indicators for the Assessment of Ecological and Economic Consequences of Municipal Policies for Resource use, *Ecol. Econ.*, 14, 175–184.

Noss, R F (1995) *Maintaining Ecological Integrity in Representative Reserve Networks*, World Wildlife Fund, Canada, and World Wildlife Fund, United States.

Ramírez, R (1999) *Stakeholder Analysis and Con ict Management*, ed D Buckles, International Development Research Center, Ontario, Canada/The World Bank, Washington, DC.

Rapport, D J, Regier, H A, and Hutchinson, T C (1985) Ecosystem Behavior under Stress, *Am. Nat.*, 125, 617–640.

Rees, W F (1996) Revisiting Carrying Capacity: Area-based Indicators of Sustainability, *Popul. and Environ.*, 17, 195–215.

Regier, H A and Kay, J J (1996) An Heuristic Model of Transformations of the Aquatic Ecosystems of the Great Lakes, St. Lawrence River Basin, *J. Aquat. Ecosyst. Health*, 5, 3–21.

Scheffer, M (1998) *Ecology of Shallow Lakes*, Chapman and Hall, London.

Schneider, E and Kay, J J (1994a) Life as a Manifestation of the Second Law of Thermodynamics, *Math. Comput. Model.*, 19, 25–48.

Schneider, E D and Kay, J J (1994b) Complexity and Thermodynamics: Towards a New Ecology, *Futures*, 24, 626–647.

Soulé, M E (1995) Biodiversity Indicators in California: Taking Nature's Temperature, *Calif. Agric.*, 49, 40–44.

Svanberg, O (1996) Monitoring of Biological Effects, *Resour. Conserv. Recycling*, 16, 351–360.

Warriner, K (1997) Public Participation and Environmental Planning, in *The Environment and Canadian Society*, ed T Fleming, International Thomson Publishing, Toronto.

Wiersma, B (1995) *Looking for the Forest Among the Trees*, National Academy of Science, Washington, DC.

Wilson, K K and Morren, G (1990) *Systems Approaches for Improvement in Agriculture and Resource Management*, Macmillan, New York.

Woodley, S, Kay, J, and Francis, G (1993) *Ecological Integrity and the Management of Ecosystems*, St. Lucie Press, FL.

Woodley, S J (1993) Monitoring and Measuring Ecosystem Integrity in Canadian National Parks, in *Ecological Integrity and the Management of Ecosystems*, eds S Woodley, J Kay, and G Francis, St. Lucie Press, FL, 155–176.

A

Adaptation Strategies

see Adaptation Strategies (*Opening essay*, **Volume 4**)

Adaptive Environmental Management

The notion of *adaptive environmental management* (AEM), also termed active adaptive management or experimental management, has evolved over time, but its focus has always been on managing resources under uncertainty. Uncertainties are pervasive in resource management: data are almost always limited and both internal and external processes, including management practices, can cause ecosystem behavior to switch dramatically and unexpectedly (Holling, 1973).

The emphasis placed on reducing such uncertainties distinguishes AEM from other approaches to resource management (Holling, 1978). At their best, these other approaches usually strive for a process of informed trial-and-error through which available knowledge is used to produce a best-guess management strategy. Ongoing scientific research, as well as monitoring the (ecological, social and economic) results of implementing this strategy produce new information, which is then used to refine the best guess. In this process of informed trial-and-error, management operations are typically focused solely on meeting specific socio–economic targets and environmental standards. Efforts to specifically reduce uncertainties occur through standard (reductionist) scientific research, which is conducted at small scales on selected components of the ecosystem. Because this small scale, piecemeal approach ignores component interactions, it is relatively ineffective when used alone at reducing uncertainty in the large scale, operational management of complex ecosystems. In contrast to informed trial-and-error, AEM is a process in which a considerable amount of management's operational effort is intentionally directed at reducing uncertainties. AEM uses operational activities as tools not only to manage, but also to conduct experiments in operational settings (Walters, 1986), and thus test different possible management strategies at operational scales. Thus, in contrast to informed trial-and-error, AEM uses management experiments to supplement standard scientific research in reducing uncertainty.

AEM is most useful when there is substantial uncertainty about the effects of management policies. In certain situations, informed trial-and-error may be more appropriate. Conducting management experiments and constructing monitoring systems over large areas for ecosystems with long and erratic response times is expensive and technically challenging. The large scale of management experiments can also pose unacceptable ecological risks.

Implementing AEM often depends on a shift in awareness of institutions and society to the realization that experimentation at operational scales can be a cost-effective approach to long-term resource management. Bringing about this shift may involve a variety of social, political, and institutional changes (Gunderson *et al.*, 1995; Lee, 1993). Because AEM's management experiments involve both people and institutions affected by the management strategy in designing the experiments and in interpreting the results, AEM has been described by some as social learning (Lee, 1993) (*see* Holling, C S, Volume 2; Social Learning in the Management of Global Atmospheric Risks, Volume 5).

REFERENCES

Gunderson, L H, Holling, C S, and Light, S S, eds (1995) *Barriers and Bridges to the Renewal of Regional Ecosystems*, Columbia University Press, New York.

Holling, C S (1973) Resilience and Stability of Ecological Systems, *Annu. Rev. Ecol Syst.*, 4, 1–23.

Holling, C S (1978) *Adaptive Environmental Assessment and Management*, John Wiley & Sons, London.

Lee, K N (1993) *Compass and Gyroscope: Integrating Science and Politics for the Environment*, Island Press, Washington, DC.

Walters, C J (1986) *Adaptive Management of Renewable Resources*, Macmillan, New York.

RICHARD A FLEMING AND JAMES A BAKER *Canada*

Agarwal, Anil

(1948-)

Anil Agarwal was trained as a mechanical engineer at the Indian Institute of Technology, Kanpur. He is an environment advocate, journalist and analyst in India. In 1980, he founded the Center for Science and Environment, one of India's first environmental non-governmental organizations (NGOs) to analyze and study the relationship between environment and development and create public consciousness about the need for sustainable development. In 1982, the Center published a pioneering *Citizens' Report on the State of India's Environment*, which provided an overview of the level of degradation and its impact on the people of India.

In 1985, Agarwal co-edited *The Second Citizens' Report on the State of India's Environment*. In 1986, the government of India presented Agarwal with its prestigious Padma Shri award.

In 1986, Prime Minister Rajiv Gandhi invited Agarwal to address the Union Council of Ministers to speak to them about the importance of the relationship between environment and development. Thereafter the Prime Minister asked him to address Members of Parliament on the same issue. He addressed all the 27 Parliamentary Committees, which oversee the work of different Union Ministries.

During this period, Agarwal also chaired the world's largest network of environmental NGOs based in Nairobi, the Environment Liaison Center. In 1987, the United Nations Environment Programme (UNEP) elected Agarwal to its Global 500 Honor Roll for his work both in the national and international arena.

In 1992, Agarwal founded and continues to edit *Down to Earth*, India's first science and environment fortnightly, in order to take environmental issues beyond the limited constituency of environmentalists.

At the national level Anil, among several other things, is the Chairperson of the Society for Environmental Communications, New Delhi, a member of the National Environment Council, Government of India: the Task Force on Control of Air Pollution, Government of Delhi, and the Standing Committee, National River Conservation Authority, Government of India, New Delhi.

At the international level he has been extremely active. Among other things he is Patron, Schumacher College, Devon, UK; Member of the International Working Group on Sustainable Livelihoods, United Nations Development Program, New York, and of the World Commission on Water for the 21st Century, co-sponsored by the United Nations Educational, Scientific and Cultural Organization (UNESCO), the Food and Agriculture Organization (FAO), the United Nations Development Programme (UNDP), UNEP, the World Health Organization and the World Bank.

Agarwal has received a large number of awards from his country, India, as well as from Hawaii and France and from two United Nations Organizations: FAO and UNEP. Anil Agarwal, over the past 20 years, has published 18 books in London, New Delhi, Washington, Rome and Geneva. Some of his books have been translated into French, German and Danish.

His books cover drugs, land, water, energy, food and agriculture, oods and droughts, forests, global warming, women and the landless Aral Sea and environmental economics in addition to his three citizen's reports on the State of Environment in India.

MOSTAFA K TOLBA *Egypt*

Agenda 21

Mary P Williams Silveira
United Nations Economic Commission for Europe, Geneva, Switzerland

Agenda 21 was adopted by the United Nations Conference on Environment and Development (UNCED) in Rio de Janeiro, in 1992, by consensus. As the name implies, it is intended as an Agenda for the 21st century, in which emphasis is placed on decision-making that integrates social and economic concerns with sustainable management of the environment.

Implementation of Agenda 21 is the responsibility of governments, non-governmental actors and individuals. To monitor implementation, the General Assembly established the United Nations (UN) Commission on Sustainable Development (CSD) in 1993. In 1997, the General Assembly carried out a ve-year review of Agenda 21 and adopted a program for the Further Implementation of Agenda 21.

INTRODUCTION

Although efforts had been made by the international community to address major issues of environmental degradation and conservation since the late 1960s, by December 1989, the United Nations General Assembly, in its Resolution 44/228, expressed deep concern at the serious degradation of the global life-support systems and warned of ecological catastrophe. Political understanding of the issues was not keeping pace with scientific knowledge, and environmental issues were largely approached as the provenance of environmentalists.

The General Assembly therefore, in the same resolution, resolved to convene UNCED, in Rio de Janeiro, Brazil, 3–14 June 1992. The conference was to elaborate strategies and measures to halt and reverse the effects of environmental degradation in the context of strengthened national and international efforts to promote sustainable and environmentally sound development in all countries.

Preparations for the conference spanned more than two years. During this period, negotiations began on the documents that would be adopted by the conference, including Agenda 21.

AGENDA 21

Agenda 21 is formally the Programme of Action for Sustainable Development. It was adopted by consensus by the 178 governments that participated in UNCED. Subsequently welcomed by the General Assembly in its Resolution 47/191 of 29 January 1993, Agenda 21 addresses the pressing problems of today and also aims at preparing the world for the challenges of the next century. It re ects a global consensus and political commitment at the highest level on development and environment cooperation. (United Nations Department of Public Information, 1993, Preamble, paragraph 1.3).

Agenda 21 is comprehensive: it addresses needed actions in 39 areas, spanning four clusters of issues: economic and social; conservation and resource management; participation; and implementation.

It is complex: the common thrust throughout is long-term, strategic, integrated and participatory decision-making.

Agenda 21 is inclusive and universal: it places a major emphasis on broad participation not only from governments, the UN and other intergovernmental organizations, but also from nine major groups in civil society (see below).

Agenda 21 is forward looking. It is a program to take the world into and through the 21st century, implicitly founded on the concept of inter-generational equity (see also Principle 3 of the Rio Declaration on Environment and Development, which states that the right to development must be fulfilled so as to equitably meet developmental and environmental needs of present and future generations (United Nations Department of Public Information, 1993).

Agenda 21 is a consensus program. It does not carry the force of international law, but, having been adopted by 178 governments, it does bring with it a strong moral obligation to ensure full implementation.

CONTENTS

Within its 40 chapters, Agenda 21 specifies 115 programs and approximately 2500 actions to be implemented by governments at the local, national, regional and global levels. Each chapter (issue) also identifies priorities, targets, cost estimates, modalities and means of implementation through new and additional financial resources, technology transfer, and strengthening of institutional capabilities at national and international levels.

The economic and social issues are international cooperation (chapter 2), poverty (3), changing consumption patterns (4), demographic dynamics (5), human health (6), human settlements (7), and integrating environment and development in decision-making (8) (the numbers in parenthesis refer to chapter numbers in Agenda 21. Actual chapter titles are more descriptive and longer than the shorthand names provided here).

Issues of conservation and management of resources for development refer to the atmosphere (9), integrated planning and management of land resources (10), combating both deforestation (11) and desertification and drought (12), mountain development (13), agriculture and rural development (14), biological diversity (15), biotechnology (16), oceans and seas (17), freshwater resources (18), toxic chemicals (19), hazardous wastes (20), and radioactive wastes (22), and solid wastes and sewage (21).

Issues of participation are geared toward strengthening the nine major groups listed in Agenda 21 as women (24), children and youth (25), indigenous people and their communities (26), non-governmental organizations (27), local authorities (28), workers and their trade unions (29), business and industry (30), the scientific and technological community (31) and farmers (32).

Finally, the means of implementation are finance (33), technology transfer (34), science (35), education, public awareness and training (36), national mechanisms and international cooperation for capacity-building (37), international institutional arrangements (38), international legal instruments and mechanisms (39) and information for decision-making (40).

IMPLEMENTATION

To review progress at the international, regional and national levels in the implementation of recommendations and commitments contained in, *inter alia*, Agenda 21, the General Assembly, in its Resolution 47/191 of 22 December 1992, established the UN CSD (*see* CSD (Commission on Sustainable Development), Volume 4) as a functional commission of the Economic and Social Council. The Commission meets in regular session once each year and reviews specific issues of Agenda 21 according to a multi-year program of work. To facilitate its review, the Commission has requested governments of all member states to provide, on a voluntary basis, annual reports of their national implementation of Agenda 21.

Agenda 21 calls for its activities to be implemented at all levels and by all local, national and international actors, including governments and local authorities, intergovernmental organizations and non-governmental organizations, other major groups and individuals.

Particular stress is given in virtually every chapter to implementation through national strategies and action plans. One example of implementation is the adoption of National Agendas 21 by 56 governments (this figure is based on national reports submitted by governments annually, on a voluntary basis, to the United Nations Commission on Sustainable Development (UNCSD). The last count was undertaken in 1998, following the sixth session of the CSD), or 30% of the member states of the UN. Many more governments have adopted similar comprehensive strategies for sustainable development at the national level. Governments have also undertaken a broad range of programs and other activities to implement all of the issues raised in Agenda 21. Examples of these are available on the UN system-wide website for sustainable development (this website is maintained by the UN Division for Sustainable Development and is located at the following Internet address: http://www.un.org/esa/agenda21/natlinfo).

Other examples rest at the local level, where local authorities, non-governmental organizations and other major groups have been important catalysts behind the creation of Local Agendas 21. As of 30th November 1996, more than 1800 local governments in 64 countries were involved in Local Agenda 21 activities. An additional 933 municipalities from 43 countries were in the process of establishing Local Agendas 21, and 879 municipalities were just getting started (the information is taken from the Local Agenda 21 Survey, published in February 1997 by the International Council for Local Environmental Initiatives, in cooperation with the United Nations Department for Policy Coordination and Sustainable Development.

Youth have been active in facilitating implementation of Agenda 21. In 1994, a group of international students wrote, designed and published a children's version of Agenda 21. Young people from approximately 100 countries worked together to produce the book that has subsequently sold over 250 000 copies in 17 languages.

REVIEW AND FURTHER PROGRAM

It was understood from the beginning that the objectives and activities of Agenda 21 would need to be revised continuously. The General Assembly decided therefore to conduct a five-year review of the implementation of Agenda 21. This review took place at the nineteenth Special Session of the United Nations General Assembly (UNGASS), in June 1997. UNGASS, in resolution S-19/2 of 19 September 1997, while acknowledging action on a number of fronts, expressed concern over the lack of real progress in the implementation of Agenda 21. It adopted a program for the Further Implementation of Agenda 21 (United Nations Department of Public Information, 1997), which reaffirms that:

> that Agenda 21 remains the fundamental program of action for achieving sustainable development. ••• We are convinced that the achievement of sustainable development requires the integration of its economic, environmental and social components. We recommit to working together – in the spirit of global partnership – to reinforce our joint efforts to meet equitably the needs of present and future generations.

In addition to reaffirming Agenda 21, the program for the Further Implementation of Agenda 21 established a new target by calling for the formulation and elaboration of national strategies for sustainable development that re ect the contributions and responsibilities of all interested parties in all countries by the year 2002. It raised the visibility of certain issues contained in Agenda 21, including energy, transport, sustainable tourism and industry. And it committed governments to undertake the next comprehensive review of Agenda 21 in the year 2002 (United Nations Department of Public Information, 1997, 1–2).

REFERENCES

United Nations Department of Public Information (1993) *Agenda 21: The United Nations Programme of Action from Rio*, United Nations Sales Publication No. E.93.I.11, New York.

United Nations Department of Public Information (1997) *Earth Summit • 5: Programme for the Further Implementation of Agenda 21*, New York.

Andersen, Stephen O

(1948-)

Stephen O Andersen, a United States Citizen, born in Logan, UT on 17 January, 1948, is Director of Strategic Climate Projects in the US Atmospheric Pollution Prevention Division. Prior to that Dr Andersen was Deputy Director of the Environmental Protection Agency (EPA) Stratospheric Protection Division.

Andersen studied business administration and economics as an undergraduate and received his PhD in Agricultural and Natural Resources Economics from the University of California, Berkeley (1974).

Prior to joining EPA he worked for environmental and consumer non-governmental organizations (NGOs) including Sierra Club Research, the Environmental Law Institute, and the Consumer Energy Council. He was also a Professor of Economics and Environmental Studies at the College of the Atlantic and the University of Hawaii.

Andersen studied stratospheric ozone depletion and climate change in the early 1970s as a member of the Climatic Impact Assessment Project that investigated the consequences of SSTs. In 1986 he was hired by EPA's stratospheric ozone protection office. His EPA responsibilities included analysis of regulatory impact, cost, and trade, technology cooperation, and sector projects in halons, foams, mobile air conditioning, and military uses.

Steve Andersen (as he is known to his friends) was one of the main pillars in the implementation of the Montreal Protocol on Substances that Deplete the Ozone layer. In 1986 Dr Andersen brought Soviet authorities together with the National Aeronautics and Space Agency (NASA) scientists and EPA senior managers, to gain acceptance for a US ozone monitoring satellite to be carried into space on a Soviet rocket, in order to overcome the loss of access to space resulting from the Challenger accident. He helped organize military support for ozone layer protection including projects to change military specifications requiring use of ozone-depleting substances (ODSs) and three international conferences on the importance of the military to ozone layer protection. In 1992 he helped the North Atlantic Treaty Organization (NATO) take the unprecedented step of writing directly to the United Nations Environmental Programme (UNEP) supporting the proposed accelerated phase-out of ozone-depleting substances. NATO also endorsed technology cooperation, exchange of information regarding halon banks and the harmonization of standards to allow production and maintenance of military equipment without the use of chemicals that deplete the ozone layer.

Andersen also helped organize global technology cooperation including the early phase-out of chloro uorocarbons (CFCs) in refrigerator manufacturing in Thailand and the corporate pledge to help Vietnam avoid increased dependence on ODSs. He served on the team that commercialized no-clean soldering and the team that is phasing out ODSs from solid rocket motors. Dr Andersen helped organized the Halon Alternatives Research Corporation and the International Co-operative for Ozone Layer Protection.

Andersen co-chaired quite a number of the Montreal Protocol Panels, Committees and working groups of the Ozone Assessment Panels. He has actively participated in workshops and conferences concerning the protection of the ozone layer. In recognition of his achievements, Dr Andersen has received several medals and awards from UNEP, the United States, Japan and Brazil.

In 1995 Dr Andersen began work on climate protection by providing the United Nations with a blueprint for undertaking evaluation of the technical and economic feasibility of climate protection including the involvement of experts from industry. In 1996 he successfully negotiated cooperation of electronics manufacturers in Japan, Korea, Singapore, and Taiwan to search for alternatives to the use and emissions of per uoro-compounds as etching agents in semiconductor manufacture.

See also: Industrial Responses to Stratospheric Ozone Depletion and Lessons for Global Climate Change, Volume 4.

MOSTAFA K TOLBA *Egypt*

Aral Sea

see Aral Sea (*Case study,* **Volume 4**)

APN (Asia–Pacific Network)

The Asia–Pacific Network for Global Change Research is an inter-governmental network whose mission is to foster

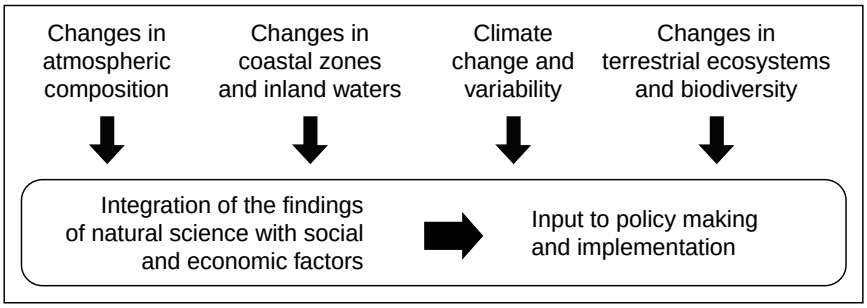


Figure 1 APN research framework

global environmental change research in the Asia–Pacific region, increase developing country participation in that research, and strengthen interactions between the science community and policy makers.

The APN believes that international cooperation among governments and scientists will help increase the understanding of the complex mechanisms and impacts of global environmental change on ecosystems and human society in the Asia–Pacific region. This is necessary to identify and address the problems that may arise from that change.

The vision of the APN is that scientists and institutions within the countries of the Asia–Pacific region will together conduct high quality, cooperative, multidisciplinary research on the region's key global environmental change issues and that the countries will use the results of that research in their policy making processes.

Currently 21 governments participate in the APN (Australia, Bangladesh, Cambodia, China, Fiji, India, Indonesia, Japan, Korea, Laos, Malaysia, Mongolia, Nepal, New Zealand, Pakistan, Philippines, Russia, Sri Lanka, Thailand, Vietnam and the USA) together with scientists throughout the region and the wider global change research community.

The APN was established in response to the 1990 White House Conference on Science and Economics Research Related to Global Change, at which then US President Bush invited the countries of the world to join the United States in creating three regional networks. Those three networks are the APN, ENRICH (European Network for Research in Global Change) and the IAI (Inter American Institute for Global Change Research).

After a series of planning workshops the APN was formally launched in 1996. Its main focus has been the support of global environmental change research projects which cannot be conducted on a national basis and require regional cooperation, and which strengthen research capability at the national and regional levels.

In order to achieve its goals the APN conducts a number of different activities including:

- supporting collaborative research and training activities;
- organizing planning and scoping workshops to develop new research projects;
- disseminating global change information to scientists and policy makers;
- providing opportunities for governments and scientists to discuss regional research priorities and other global change issues.

To date the APN has made direct contributions to 65 global change and capacity building projects, and helped to promote in-kind contributions from a number of countries and organizations for this work. The APN has also initiated and funded scoping workshops, symposiums, an inventory of global change research and a number of other initiatives. It produces a quarterly newsletter and maintains a homepage on the Internet: http://www.apn.gr.jp.

The broad framework of the research interests of the APN is set out in Figure 1. The primary concern of the APN is to provide an input to policy-making through the integration of the natural and social sciences. Accordingly, understanding the human dimensions of global change is at the heart of the APN's concerns. The APN has sponsored research and capacity building activities in areas such as:

- food and water security;
- regional climate modeling;
- land use and cover change;
- urbanization;
- marine and coastal zone studies;
- aerosol data synthesis.

The APN believes that working in partnership with other organizations involved in global change research is essential to maximize the resources available and to deliver the

best possible results. In particular the APN cooperates very closely with START (the System for Analysis Research and Training in Global Change), the International Geosphere Biosphere Program (IGBP), the International Human Dimensions Program (IHDP) and the World Climate Research Program (WCRP). It has also developed links with a number of other regional bodies with an interest in global change research.

HIROAKI TAKAGI *Japan*

Assessment

see Integrated Assessment (**Volume 4**)

Assimilative Capacity

see Maximum Sustainable Yield (MSY) (**Volume 4**)

B

Baltic Sea

see Baltic Sea (*Case study*, **Volume 4**)

Baltic: Water Resources

see Water Resources: Baltic (**Volume 3**)

Basel Convention

Jennifer Clapp
Trent University, Peterborough, Canada

The Basel Convention on the Control of Transboundary Movements of Hazardous Wastes and their Disposal is an international legal treaty governing cross-border shipments of toxic wastes. The impetus for this convention was the growing export in the 1980s of hazardous wastes from industrialized to developing countries. This practice emerged as a response to rising costs for hazardous waste disposal in rich countries at a time when costs for such disposal in developing countries was minimal. As the nature of this trade became apparent, work began on drafting a global treaty to control it. The United Nations Environment Programme (UNEP) sponsored negotiations on this treaty, which was adopted by 116 governments in Basel, Switzerland, in March 1989. The treaty came into force in May 1992 when it received the necessary 20 rati cations (Tolba and Rummel-Bulska, 1998). As of April 2000, there were 135 parties to the convention. The types of wastes covered by the treaty are those which are toxic, poisonous, explosive, corrosive, amable, eco-toxic or infectious.

The principal objectives of the Basel Convention are to:

1. reduce the transboundary movement of hazardous wastes;
2. encourage hazardous wastes to be treated and disposed of as close as possible to their source of generation; and
3. minimize the generation of hazardous wastes.

To facilitate the achievement of these objectives, the provisions of the original 1989 convention provided a regulatory framework for the movement of wastes (Kummer, 1995; O'Neill, 2000).

There are several cases under which the export of hazardous wastes is strictly forbidden by the convention. Parties are prohibited from exporting wastes to states that have banned its import. This upholds states' sovereign right to refuse waste imports if they so choose. Parties are also to refrain from trade in hazardous wastes with non-parties unless there is a bilateral or regional agreement under which wastes are to be disposed of in no less environmentally sound a manner than that outlined in the convention (Article 11). The purpose of this latter trade restriction is to encourage states to become parties to the convention.

For exports of wastes to states that have not prohibited its import, the principle of prior written notification and consent is to be adhered to (Article 6). This requirement was intended to allow countries to accept or turn away such imports on a case-by-case basis at their own discretion. Parties that intend to export hazardous wastes must notify the importing country in writing in advance of a hazardous waste shipment. If the latter agrees in writing to accept the waste, and provided there is assurance that the waste will be disposed of in an environmentally sound manner, then the transaction is allowed. Waste exports that do not follow these provisions are considered by the convention to be illegal. Parties are asked to only export wastes if they themselves lack the means to dispose of those

wastes in an environmentally sound manner, or if the wastes are considered a raw material in the importing country.

The control system established by the 1989 Basel Convention was criticized by a number of developing countries and environmental non-government organizations (NGOs) as being weak (Puckett, 1992). This was mainly because the convention only sought to control the export of wastes and did not ban their trade between rich and poor countries. Moreover, it did not effectively put a halt to the trade in hazardous wastes destined for recycling operations in developing countries. The issue of whether to incorporate a ban on the trade in wastes between rich and poor countries within the Basel Convention was the subject of heated debate in the early 1990s. In 1994, the conference of parties to the Basel Convention adopted a decision to ban outright the export of hazardous waste destined for both disposal and recycling from the highly industrialised Organisation for Economic Cooperation and Development (OECD) countries to non-OECD countries (Decision II/12). This decision was made on the grounds that developing countries were not capable of disposing of such waste in an environmentally sound manner.

The Basel Convention was formally amended to this effect in 1995, with the ban on waste exports for disposal taking place immediately, and the ban on waste exports for recycling taking effect December 31st, 1997 (Decision III/1). The amendment asks parties listed in Annex 7 to refrain from the export of wastes to countries not listed in Annex 7, once the amendment is ratified. Annex 7 lists members of the OECD, European Commission (EC) and Liechtenstein. To come into force, this amendment requires ratification by 62 parties. As of November 2000, 22 ratifications have been deposited with the Secretary General of the United Nations. The Basel Ban has continued to be a source of controversy amongst the parties, as some have argued that such a trade ban is in contravention of international trade rules enforced by the World Trade Organization (Krueger, 1999).

In 1998, progress was made by the technical working group of the convention in establishing a system for classifying wastes and clarifying precisely which wastes are covered by the treaty. This clarification has gone some way to assure those involved in the global trade in recyclable materials that many of the wastes traded by them are not classified as hazardous, and are thus not controlled by the convention.

In 1999, a protocol on liability and compensation was adopted by the parties. The drafters of the original convention had hoped to include provisions on liability and compensation for damages arising from the transboundary movement of wastes in the original convention, but agreement on these matters was not secured at that time. Instead, Article 12 of the Basel Convention calls on parties to develop such a protocol as soon as practicable. The driving force for the development of this protocol was the concern of developing countries that they lacked funds and the technology to deal with illegal imports of hazardous waste and their impact on their local environments. The protocol thus provides rules for the compensation of countries that are victims of environmental damage due to waste imports, and determines liable parties that must provide that compensation.

Also in 1999, the parties adopted the Basel Declaration on Environmentally Sound Management, which outlines the agenda for the next decade of the convention. This declaration emphasizes the importance of prevention and minimization, as well as the environmental sound management of recycling and disposal of hazardous wastes. This marks a new direction for the convention, as it broadens the focus of the convention as it puts more emphasis on the specific objective of waste management (Clapp, 2001).

REFERENCES

Clapp, J (2001) *Toxic Exports: The Transfer of Hazardous Wastes and Technologies from Rich to Poor Countries*, Cornell University Press, Ithaca, NY.

Krueger, J (1999) *International Trade and the Basel Convention*, Earthscan, London.

Kummer, K (1995) *International Management of Hazardous Wastes: The Basel Convention and Related Legal Rules*, Oxford University Press, Oxford.

O'Neill, K (2000) *Waste Trading Among Rich Nations: Building a New Theory of Environmental Regulation*, MIT Press, Cambridge, MA.

Puckett, J (1992) Dumping on Our World Neighbors: The International Trade in Hazardous Wastes, and the Case for an Immediate Ban on All Hazardous Waste Exports from Industrialized to Less-Industrialized Countries, *Green Globe Yearbook*, Oxford University Press, Oxford.

Tolba, M and Rummel-Bulska, I (1998) *Global Environmental Diplomacy: Negotiating Environmental Agreements for the World, 1973- 1992*, MIT Press, Cambridge, MA.

FURTHER READING

http://www.basel.int.

http://www.ban.org.

Krueger, J (1999) What's to Become of Trade in Hazardous Wastes? The Basel Convention One Decade Later, *Environment*, 41(9), 10–19.

Benedick, Richard Elliot

A career diplomat and former Deputy Assistant Secretary of State, Richard Benedick has had a distinguished career in global environmental affairs. Ambassador Benedick was the chief US negotiator and a principal architect of the historic 1987 Montreal Protocol on protecting the ozone layer. His acclaimed book, *Ozone Diplomacy: New Directions in Safeguarding the Planet* (Harvard University Press, 1991, revised 1998), is being translated into Japanese and Spanish and is used in universities around the world.

After serving several years on the International Advisory Board of Battelle, he joined their Washington, DC, office in 1998 as a Deputy Director of the Environmental and Health Sciences Division of Pacific Northwest National Laboratory. He is a Visiting Fellow at the Wissenschaftszentrum Berlin. He has lectured at numerous universities, serves on several boards, and consults for international agencies, governments and industry. From 1990 to 1994, he was Special Advisor to Secretaries-General of both the United Nations Conference on Environment and Development (Rio de Janeiro, 1992) and the International Conference on Population and Development (Cairo, 1994), assisting the negotiations and drafting chapters of Agenda 21 and the Cairo Programme of Action. He has organized or presided over numerous international conferences and negotiations on environment, development, population and science policy.

Benedick served on diplomatic assignments in Iran, Pakistan, Paris, Bonn and Athens. As Deputy Assistant Secretary of State for Environment, Health, and Natural Resources, he supervised policy formation and international negotiations on climate change, biotechnology, ozone, oceans, tropical forests, wildlife conservation, and AIDS. He has also headed policy divisions at the State Department responsible for population affairs and for development assistance, and was selected for the Senior Seminar; the US Government's highest study program. He has led many US delegations and has testified before Congress and foreign parliaments.

Elected a Fellow of the World Academy of Art and Science in 1991, Benedick has also received two of the highest presidential career public service honors (the Distinguished and two Meritorious Service Awards), the State Department's John Jacob Rogers medal, and the 1997 United Nations Tenth Anniversary Ozone Award. Other distinctions include: two State Department Superior Honor awards; visiting fellow, National Center for Atmospheric Research; senior fellow, World Wildlife Fund; awards from the Climate Institute and the Population Reference Bureau; the *Tonisssteiner Kreis*; Phi Beta Kappa; and a citation in *Who's Who in America*.

He received an AB *summa cum laude* from Columbia, an MA in economics from Yale, and a doctorate in international finance from the Harvard Graduate School of Business Administration. He was also an Evans Fellow at Oxford University in metaphysical poetry, and is author of over 90 publications in the US and abroad, including *Industrial Finance in Iran* and *The High Dam and the Transformation of the Nile*.

MOSTAFA K TOLBA *Egypt*

Beneficiary Pay Principle

see Eco Taxes (**Volume 4**)

Biodiversity in Freshwaters

see Biodiversity in Freshwaters (**Volume 2**)

Black Sea

see Black Sea (*Case study*, **Volume 4**)

Blue Plan

In 1975, all the Mediterranean countries, increasingly concerned by the pollution of their common sea, signed the Barcelona Convention and launched their Mediterranean Action Plan (MAP). Since some 80% of marine pollution comes from land-based sources, it was considered necessary to analyze the root causes of this process, to see how it might evolve, and more generally to look for socio-economic development patterns which would avoid environmental degradation in the Mediterranean Basin. This continued systemic and prospective

study of current trends and possible futures is known as the *Blue Plan*.

Using the method of scenarios, the Blue Plan has explored the interactions between development sectors (agriculture, industry, energy, tourism, transport), environmental components (soil, water, vegetation, coastline, sea), demographic dynamics and land occupation changes, in order to suggest avenues towards sustainable development to the Mediterranean countries and the European Commission. Its initial results (Grenon and Batisse, 1989) aroused keen interest in the region. The report revealed how trend scenarios could prove socially and environmentally unacceptable well before 2025, particularly for the countries South and East of the Basin, and outlined the major requirements for a more sustainable future (*see* Mediterranean Sea, Volume 4).

The Blue Plan keeps updating its databases, checking the validity of its scenarios, monitoring changes in the region and publishing thematic studies on a variety of subjects, such as the Mediterranean vision on water, population and the environment presented at the second World Water Forum in The Hague. It now serves as the main support centre for the Mediterranean Commission for Sustainable Development established in 1995, and to this effect prepares reports on such topics as water demand, sustainable tourism, environmental indicators and impacts of free trade.

A major step forward, in line with a main recommendation of the Blue Plan, was the establishment in 1995 of the Euro-Mediterranean partnership, supported by a fund of the European Union called MEDA. The MEDA project on environmental statistics is conducted by the Blue Plan, which also hosts a MEDA project for a regional information system on water.

Blue Plan operations are made possible through a specific non-governmental organization (NGO) under French law and carried out in a regional activity centre located at Sophia Antipolis, near Nice, where a staff of some 15 people works in close contact with experts from all countries in the region. About half of the total annual budget of approximately US$1.5 million is provided by MAP, the rest coming from the European Union (including MEDA) and the French Ministry of Environment.

Future publications will include further issues in the Blue Plan Fascicules series, a comprehensive compendium of indicators, and a new report on environment and development in the Mediterranean to be issued in 2003.

REFERENCE

Grenon, M and Batisse, M (1989) *Futures for the Mediterranean Basin*, Oxford University Press, Oxford.

MICHEL BATISSE *France*

Blue Planet Prize

see International Environmental Prizes (**Volume 4**)

Boreal Forest Carbon Flux and it's Role in the Implementation of the Kyoto Protocol Under a Warming Climate

Kaz Higuchi
Meteorological Service of Canada, Toronto, Canada

Since the signi cant increase in industrial activities in Europe in the mid-1800s, the atmospheric carbon dioxide (CO_2) concentration has exponentially climbed from a value of about 280 parts per million (ppm) to over 360 ppm at present, and is still increasing at a rate of about 1- 1.5 ppm per year. This increase is equivalent to about 2- 3 Pg of carbon (Pg C = 10^{15} grams of carbon) per year in the atmosphere which now contains about 730 Pg C of CO_2. We know relatively well that about 5 Pg C is emitted into the atmosphere every year from various fossil fuel combustion processes. In addition to this industrial source of CO_2, approximately 1.5 Pg C per year is emitted from tropical deforestation. Of these two major sources of CO_2, totaling 6.5 Pg C, it is estimated that about 2 Pg C per year is absorbed by the oceans, while about 3 Pg C per year remains, on average, in the atmosphere. From our present knowledge of the existence of only three major carbon reservoirs in the world (atmosphere, ocean, land biosphere), a carbon sink of about 1.5 Pg C per year is required in the terrestrial biosphere to balance the global carbon budget. A major research activity in the carbon cycle community is to nd out where the sinks are in the land biosphere, and what processes control the net uptake of the atmospheric carbon and how these processes might change under a changing (warming) climate. This scienti c issue is politically important because of its possible in uence on the way the terms of the Kyoto Protocol are implemented by nations such as Canada and the United States. The Kyoto Protocol legally binds countries that signed the agreement to a collective reduction in greenhouse gas emissions below 1990 levels. The Protocol allows countries to use net emissions from land-use change in forestry and agriculture in calculating their compliance with the terms of the Protocol.

To tackle the important but extremely difficult problem of long-term changes in the strengths of the biosphere carbon sinks and sources, scientists usually employ either the *bottom-up* or the *top-down* approach. In the latter case, an atmospheric model, constrained by oceanic carbon dioxide (CO_2) uptake, globally-balanced land biospheric exchange of CO_2, and the observed atmospheric CO_2 concentration measurements from a global CO_2 observing network, is used to estimate spatial and temporal variations in the carbon sources and sinks in the terrestrial biosphere. One of the recent studies using the top-down method has estimated a significant terrestrial sink of about 1.7 Pg C per year in North America during the late 1980s and the early 1990s (Fan *et al.*, 1998). However, climate change impact on net carbon storage in the terrestrial biosphere under an increasing atmospheric CO_2 concentration cannot be made using a top-down approach.

In the bottom-up approach, scientists generally employ the eddy-correlation method using a tower situated in an ecologically homogenous environment to estimate uxes of CO_2 from vegetation. These uxes are then interpreted in terms of daily and seasonal biological and chemical processes taking place in the environment represented by ux measurements. These measurements tell us that the boreal forest in North America has been acting, at least at some locations, as a net sink of about 1–2 metric tons of C hectare $(ha)^{-1}$ $year^{-1}$ (*see* Boreal Ecosystem Atmosphere Study (BOREAS), Volume 2). If one were brave enough to multiply this figure by the boreal forest area in North America of about $2{\bullet}5 \times 10^8$ ha (1 ha = 10^4 m^2), one obtains a North American sink estimate of about 0.25–0.5 Pg C per year. Using a bottom-up approach, it is very difficult however, if not impossible, to make regional and global statements about the spatial and temporal variations in the terrestrial net ux of CO_2 within the context of changing climate. In order to overcome these shortcomings to some extent, the Ameri ux program was initiated in North America to systematically measure CO_2 uxes from as many different ecosystems in North America as possible.

Another approach, called the boundary-layer budget (BLB), has been used by some investigators to estimate average CO_2 uxes over an area of 10^2–10^4 km^2. In this approach, which can be classified as a bottom-up approach, the atmospheric CO_2 concentration in the lower atmosphere is used as an input into a highly simplified atmospheric boundary layer model to calculate the CO_2 ux from the biosphere. This method was used, for example, on a 6-year (1991–1996) record of atmospheric CO_2 concentration values from a mixed boreal forest site (Fraserdale) in Ontario, Canada. Flux estimates were found to be similar in magnitude to those obtained by measurements on ux towers (Shashkov *et al.*, unpublished).

There appears to be some evidence to indicate the existence of a large terrestrial sink in the Northern Hemisphere at the present time, with the North American biosphere possibly playing a significant role. To assess the validity (or reality) of the existence of this terrestrial sink, and to interpret it as a function of various biogeochemical processes within different ecosystems, scientists have started to employ a number of ecosystem models of hierarchical complexity. These models constitute an essential scientific tool to understand how changes in physical environment (such as temperature, precipitation and atmospheric CO_2 concentration) can affect CO_2 exchange between the land biosphere and the atmosphere. Questions we need to ask are: Do these models, which represent mechanisms of carbon cycling in various ecosystems, give net CO_2 ux estimates in agreement with the results obtained from the top-down and/or the bottom-up approaches? What controlling mechanisms governing the carbon ow through the biosphere are important in in uencing the CO_2 ux across the atmosphere-biosphere interface? How are these mechanisms changed as climate becomes warmer with increasing concentration of CO_2 in the atmosphere? A recent study by a group of scientists (Schimel *et al.*, 2000) suggests, based on the results of three different ecosystem models driven by historical meteorological and atmospheric CO_2 concentration data, that for the period 1980–1993 the ecosystems in the United States have been acting as a net sink of about 0.3 Pg C per year, of which about 0.1 Pg C per year is attributed to the CO_2 fertilization of the photosynthesis and warming climate, and about 0.2 Pg C per year to such processes as regrowth on abandoned agricultural and harvested forest lands. (The term *CO_2 fertilization* is used to denote enhanced photosynthesis in plants exposed to a high ambient CO_2 concentration.) These estimates are consistent with carbon inventory models, but are about an order of magnitude lower than those obtained by the top-down approach. Carbon inventory models attempt to keep a running account of carbon ow through various carbon-based products in our society.

The northern high latitude terrestrial biosphere above 50 °N constitutes about 25% of the global vegetated land area, and is composed of about 800–900 Pg C soil carbon and 100–150 Pg C above-ground vegetation carbon. It is dominated by boreal forest ecosystems, of which about 30% are found in Canada. A recent study by a group of British scientists (White *et al.*, 2000) shows that the terrestrial biosphere located north of 50 °N has been accumulating carbon at a rate of about 0.4 Pg C per year, mostly due to the CO_2 fertilization from the increasing CO_2 concentration in the atmosphere. This sink was found to increase, at least in their ecosystem model called *Hybrid* driven by an atmospheric climate model output, to 0.8–1.0 Pg C per year by 2050 under a warming climate. Changes in the physical

climate variables, such as temperature and precipitation, are found to cause the northern high latitude biosphere to become a sink. The above estimates are similar to those obtained in a study using a two-dimensional coupled climate/land carbon cycle model (Chan *et al.*, 1996). In the latter study, the land ecosystem model employed an empirical expression for the soil respiration as a function of soil temperature and carbon content. The study shows that the climate effect alone results in a large soil carbon loss in the northern high latitude regions, causing the biosphere in these regions to become a net source of CO_2. However, inclusion of the CO_2 fertilization effect results in very little change in the soil carbon content in these regions, while producing a relatively large increase in the vegetation carbon content due to enhanced photosynthesis. These estimates are not inconsistent with a net CO_2 uptake estimate for the Canadian forest of about 0.06 Pg C per year over the last few decades. It should be noted that a carbon budget inventory model shows the Canadian forest to be a small source of CO_2. Various values of estimated CO_2 uptake mentioned so far are summarized in Table 1.

Many ecosystem models assume that respiration will increase with an increase in atmospheric temperature. However, some recent findings indicate that soils in high latitude ecosystems may decompose slower than usually assumed in these models, thus reducing the effect of respiration dependence on temperature. In fact, it has been found that forest soils in Finland, for example, increase in carbon content with increase in mean annual temperature.

Given the extreme complexity of the physiological structure and processes of a biological system, it is going to be very difficult to predict how precisely various ecosystems in the terrestrial biosphere will respond in terms of carbon redistribution to a warming climate. It will not be a surprise to find out then that there will be a large variation in the biospheric response to various environmental changes, depending on which ecosystem one is dealing with. There are con icting opinions about the importance of the dependence of respiration on temperature, as well as about the impact of nitrogen deposition. Photosynthetic response to increasing CO_2 concentration and temperature in the atmosphere is not very clear, although there is consensus that it must vary significantly among various vegetative species. Although there appears to be some agreement on the significance of the uptake of the atmospheric CO_2 by the Northern Hemisphere biosphere on a relatively short time scale, with boreal forest ecosystems contributing about 20–30% of the 1.5 Pg C per year missing sink, there is great deal of uncertainty at the present time if the sink will continue on a long time scale under a warming climatic regime. Some scientists believe that soil respiration will begin to dominate over the next 30–50 years, as atmospheric temperature rises more and CO_2 fertilization begins to saturate. If this turns out to be true, then the terrestrial biosphere, now being viewed as a net sink of anthropogenic CO_2, will become a significant source of natural CO_2, adding to the already increasing CO_2 emission from fossil fuels. Reduction of the uncertainty in the net CO_2 ux from the terrestrial biosphere is regarded as one of the top priority activities by the carbon cycle research community.

POLITICAL IMPLICATIONS

Results from such research activities in the global carbon cycle research community will have major global political, economic and social implications. The Kyoto Protocol would legally bind industrialized countries such as the United States, Canada, Japan and the European Union to reduce their overall greenhouse gas emissions by about 5% from 1990 levels during the period from 2008 to 2012. Canada's reduction target is 6%, whereas that of the US is 7%. Article 3.3 of the Kyoto Protocol allows countries to use as credits against their intended emission reduction targets, amounts of CO_2 *saved* by land-use change in forestry via such activities as reforestation (planting of trees on land which was once forested but cleared in the last 50 years) and afforestation (planting of trees on land that has not been covered by forest for over 50 years). However, many unanswered questions related to such basic issues as the definition of a forest, and how to measure changes in the carbon content of various ecosystems have hindered the ratification of the Protocol. In the recent Intergovernmental Panel on Climate Change (IPCC) Special Report: Land Use, Land-Use Change, and Forestry, the IPCC has attempted to address

Table 1 Values indicate net ecosystem uptake of atmospheric CO_2

Model type	Present time (Pg C per year)	Under warming climate (Pg C per year)
Top-down	≈1.7 (N. America, mainly US)	
Bottom-up	≈0.2–0.5 (Boreal forest, Canada)	
Inventory model	Slight source (Canada)	
Ecosystem models	≈0.3 (US)	
	≈0.4 (North of 50 °N, Hemispheric)	≈0.5–1.0 (North of 50 °N, Hemispheric)

some of those unanswered issues (Kaiser, 2000; IPCC, 2000). The IPCC Special Report emphasizes the importance of defining what a forest is, since countries will get carbon credit for trees grown since 1990, or lose credit for trees destroyed since 1990. The IPCC Special Report also notes the importance of distinguishing forests that are harvested commercially and then replanted from forests with increased carbon sequestration from other reforestation activities.

Aside from some obvious moral issues, such processes as the CO_2 fertilization effect will likely induce many countries to start planting fast-growing trees like pine and birch to meet their targets, while possibly altering fossil fuel emission by very little. Indeed, the IPCC Special Report suggests that this is possible. Ecosystem models, which indicate that the terrestrial biosphere north of 50 °N will continue to grow as a net sink for CO_2 over the next 50 years, are supportive of such human-induced land-use change to absorb excess CO_2 in the atmosphere. But what if they are wrong? These models, although complex, are still very primitive. Simply put, the future role of the boreal forest, and indeed the entire global terrestrial biosphere, as a net sink or source of CO_2 seems to depend on the difference between CO_2 fertilization and soil respiration. These processes are very complicated and are governed by many complex interactive biochemical feedback mechanisms that scientists do not understand very well at the present time. Understanding carbon cycling in the boreal forest, as well as in other land ecosystems, has very wide and significant global political and economic implications in the 21st century.

See also: Boreal Forest, Volume 2; Tradable Permits for Greenhouse Gases, Volume 4; United Nations Framework Convention on Climate Change and Kyoto Protocol, Volume 4.

REFERENCES

Chan, Y-H, Chan, D, and Higuchi, K (1996) On a Two-dimensional Energy Balance Climate Model Coupled Interactively to a Land Carbon Cycle Model, *Tellus*, 48B, 329–346.

Fan, S, Gloor, M, Mahlman, J, Pacala, S, Sarmiento, J, Takahashi, T, and Tans, P (1998) A Large Terrestrial Carbon Sink in North America Implied by Atmospheric and Oceanic Carbon Dioxide Data and Models, *Science*, 282, 442–446.

Kaiser, J (2000) Panel Estimates Possible Carbon Sinks', *Science*, 288, 942–943.

IPCC (2000) Land Use, Land-Use Change, and Forestry, IPCC WGII Special Report, Cambridge University Press, Cambridge, 1–377.

Schimel, D, Melillo, J, Tian, H, McGuire, A D, Kicklighter, D, Kittel, T, Rosenbloom, N, Running, S, Thornton, P, Ojima, D, Parton, W, Kelly, R, Sykes, M, Neilson, R, and Rizzo, B (2000) Contribution of Increasing CO_2 and Climate to Carbon Storage by Ecosystems in the United States, *Science*, 287, 2004–2006.

White, A, Cannell, M G R, and Friend, A D (2000) The High-latitude Terrestrial Carbon Sinks: a Model Analysis, *Global Change Biol.*, 6, 227–245.

FURTHER READING

Adam, D (2001) Royal Society Disputes Value of Carbon Sinks, *Nature*, 412, 108.

http://www.royalsoc.ac.uk/policy/csink_ann.pdf

Bourdeau, Philippe F J

Professor Philippe Bourdeau is a former director of environmental science, research and development programs in the Commission of the European Union (CEC), and a past-president of the Scientific Committee on Problems of the Environment (SCOPE) (*see* SCOPE (Scientific Committee on Problems of the Environment), Volume 4), bringing to those organizations the benefit of his very considerable insight into the problems of ecotoxicology. How does the biosphere respond to the stresses imposed on it by the increasing loads of potentially toxic substances dumped into the environment? Do biosphere responses mimic those of humans? Are the effects of multiple toxins additive or multiplicative? These are questions that have interested Philippe for many years.

Philippe Bourdeau obtained his first degree (in agronomic engineering: water and forests) in Gembloux, Belgium in 1949, followed by a PhD in ecology at Duke University in 1954. He then taught at North Carolina State (1954–1956) and Yale University (1956–1962, with a short break from 1958 to 1960 as Dean of Faculty at the University of the Belgian Congo). During his Yale years, Philippe was instrumental with George Woodwell in establishing the Gamma Forest Project at Brookhaven National Laboratory on the ecological effects of nuclear war. In 1962 Philippe became head of the radiobiology service at the Euratom Joint Research Center in Ispra, Italy. Finally in 1971 he joined DG XII of the CEC, where he remained until he retired as Director in 1991; he was in charge of the research programs

of the European Union in environment, marine sciences and technologies, and non-nuclear energy. Since 1971, Professor Bourdeau has been a professor at the Free University of Brussels, where he teaches courses in radioecology, ecotoxicology and agriculture and the environment. He has also chaired an international committee that has sponsored the creation and implementation of a Master Degree Program in Environmental Management across a network of 14 European universities.

Professor Bourdeau has had a distinguished international career. He was President of SCOPE from 1995 to 1998 and earlier he was editor to several of its synthesis volumes in ecotoxicology. Between 1991 and 1994, he was head of a Task Force of the European Community Commission that led to the creation of the European Environmental Agency (EEA) in Copenhagen (*see* EEA (European Environment Agency), Volume 4). He also led the team which produced the EEA Dobris Assessment published in 1995 (*see* Dobr s European Environment Assessment Process, Volume 4) and is author of more than 100 scientific papers. Bourdeau is a Commandeur de l'Ordre de la Couronne (Belgium).

R E MUNN *Canada*

Brinkhorst, Laurens Jan

(1937-)

Laurens Jan Brinkhorst was born in Zwolle, The Netherlands on 18 March 1937. He studied Law at Leyden University (The Netherlands) and received a masters degree in Public Law and Government in 1961 at Columbia University in New York.

He started his career with Shearman, Sterling & Wright, a major law firm in New York, became Lecturer at the Faculty of Law in Leyden and then Professor of European Law at Groningen University (1966). He started his political career as State-Secretary for Foreign Affairs in charge of European Affairs (1973), and later was elected as a Member of the Second Chamber of the Netherlands for Democrats 66 (1977), where he was parliamentary leader in 1981/1982.

Subsequently he pursued a European career: he was Head of the Delegation of the Commission of the European Communities (CEC) in Japan from 1983–1987 and became Director General for Environment and Nuclear Safety (1987–1994). In June 1994, he was elected a member of the European Parliament where he sits in the European Liberal and Democratic group. Brinkhorst is a part-time Professor at the Faculty of Law of Leiden University and is a Visiting Professor at the Law Faculty of the University of Lausanne, Switzerland on International Environmental Law. During his work as Director General for Environment in the European Community (EC), Brinkhorst played a role in the negotiation of the famous Montreal Protocol on the Control of Substances that Deplete the Ozone Layer. He also led the Environment Delegations during the crucial period of negotiations between December 1986 and September 1987, the date of the adoption and signature of the Montreal Protocol.

Laurens Jan Brinkhorst is a board member of His Royal Highness Prince Charles's Business & the Environment Program Cambridge UK, and of the International Institute for Sustainable Development, Winnipeg, Canada.

He has received Decorations from Italy, Denmark, Belgium, The Netherlands and Luxembourg. He has published several books and numerous articles on European Integration and EC Law. He speaks six languages.

MOSTAFA K TOLBA *Egypt*

Bruce, James

(1928-)

James P Bruce, meteorologist and environmental consultant, was born in Toronto, Canada, in 1928. He graduated from the University of Toronto in 1948, was a weather forecaster with the Canadian Meteorological Service and obtained an MA in Physics (meteorology) from Toronto in 1952.

Seconded to the Ontario government in 1955, he established a provincial ood warning system, taught at the University of Toronto and organized a federal hydrometeorological research and applications unit. In 1967 he was appointed founding director of the Canada Centre for Inland Waters and, in 1977, became an Assistant Deputy Minister in Environment Canada with responsibility for the national forestry, wildlife, inland waters and lands directorate programs. In 1980 he transferred back to meteorology as Assistant Deputy Minister responsible for weather forecasts and warnings, national air quality monitoring, atmospheric research and the Canadian Climate Centre.

During the 1970s Bruce helped negotiate Great Lakes water quality agreements with the United States and, for

the International Joint Commission, co-chaired the Science Advisory Board and later the Great Lakes Water Quality Board. He also chaired several intergovernmental boards including the Prairie Provinces Water Board and the Mackenzie River Basin Board. Later he chaired a Canadian scientific steering committee on acid rain and international committees and conferences such as those which negotiated the Helsinki Protocol for sulfur dioxide emission reductions and the 1985 Villach conference on greenhouse gases and climate change.

Leaving government service, Bruce joined the World Meteorological Organization (WMO) Secretariat in Geneva where he served as Director of Technical Co-operation and as Acting Deputy secretary-general from 1986 to 1989. He assisted in establishing the Intergovernmental Panel on Climate Change (IPCC) and later co-chaired an IPCC working group dealing with economic and social aspects of climate change. Returning to Canada, Bruce served as chair of the United Nation's (UN) Scientific and Technical Committee for the International Decade for Natural Disaster Reduction (1990–1994) and of the Canadian Climate Program Board (1992–1997).

In the late 1990s Bruce assisted WMO in advising member countries on the management of their national meteorological and hydrometeorological services and on the feasibility of an international El Nino center in Ecuador. He has provided consultation to the Canadian government regarding assistance to Latin America in disaster preparedness, advised Caribbean countries on dealing with climate change and chaired the UN University's advisory committee on water, environment and health.

Bruce is co-author of the widely used *Introduction to Hydrometeorology* (1967) and has contributed many papers and reports for several books and conference proceedings dealing with climate, water, natural disasters and related issues. He is a Life Member of the Canadian Meteorological and Oceanographic Society and a Fellow and Honorary Member of the American Meteorological Society. In 1986 he was awarded the Patterson Medal for distinguished service to meteorology. In 1994 he was awarded WMO's International Meteorological Organization (IMO) prize for exceptional world-wide contributions in meteorology and hydrology and received an honorary doctorate of environmental studies from the University of Waterloo and, in 1996, the Massey Medal from the Royal Canadian Geographical Society. In 1997 he was made an Officer of the Order of Canada and a Fellow of the Royal Society of Canada.

Bruce currently serves on several Canadian boards, is vice-chair of the Board of the Winnipeg-based International Institute for Sustainable Development and is Senior Associate with Global Change Strategies International, Inc. He lives in Ottawa, Ontario.

MORLEY THOMAS *Canada*

Brundtland, Gro Harlem

(1939-)

Gro Harlem Brundtland, born in Oslo, is recognized for her outstanding contributions towards enhancing international cooperation in environment and development among other remarkable achievements in her continuing career. As a mother of four, married to the renowned political scientist, Arne Olav Brundtland, she is a role model to women in fulfilling their aspirations and convictions.

A medical doctor by profession from the University of Oslo (1963) with a Master of Public Health from Harvard University in 1965, she joined mainstream politics in 1974 when she was appointed to the Norwegian Cabinet as Minister for the Environment. This began a period when she acquainted herself with international issues of increasing concern in environment and development. In 1975, she was appointed Deputy Leader of the Social Democratic Party and in 1981 became the Party Leader and first female Prime Minister of Norway, indeed a post few women had achieved, comparable to Golda Meir, Margaret Thatcher and Indhira Gandhi who Mrs Brundtland held in highest esteem. In her 10 years as Prime Minister she made headlines when she appointed 8 women to her Cabinet out of 18 members. She strongly believed that ••• any cabinet with less than around 40% women will be seen as non-representative of the people . She also persuaded her party to introduce a gender quota in all public appointments.

Gro Harlem Brundtland was appointed Chairman of the World Commission on Environment and Development (WCED) by the Secretary General of the United Nations at a time when environment became a top priority issue in the world's international development agenda. The Commission was established to propose long-term environmental strategies for achieving sustainable development by the year 2000 and beyond, recommend ways whereby concerns for the environment are translated into greater cooperation among nations and define shared perceptions of long-term environmental issues. In her address to the inaugural session of the WCED, Dr Brundtland set the tone by underlining the interdependence of the heretofore discordant themes of sustainability, equity, security and international cooperation. She opined that, ••• the limiting factor is our incapacity to put all together as a guide to future action. At the inaugural meeting of WCED, Geneva, 1984, Gro, as she is known among friends, stated

> We share a world economy; a world environment; and a stake in world development and a decent and dignified human condition of life. We must learn to think globally and in a long-term perspective. The world is shrinking rapidly. No single region or nation can isolate itself from the rest of the world. They share the responsibility of a common future.

The WCED report, *Our Common Future* is recognized for fostering the concept of *sustainable development* and the Commission's recommendations led to the convening of the Earth Summit; United Nations Conference on Environment and Development (UNCED) in Rio de Janeiro in 1992.

As Director General of the World Health Organization (WHO), a post which she took up in 1998 and which brought her closer to her original calling in medicine, she bases her leadership of the WHO on the belief that a healthy and well-educated population is every country's greatest resource. She states that, Good health is a prerequisite for economic growth and social progress. Health issues should be addressed as an integral part of the overall process of economic and social development .

Gro Harlem Brundtland has, over the years, been recognized through many awards among them the prestigious Third World Prize (1988), the Indhira Gandhi Prize (1988), and the Charlemagne Prize in 1994, just to list a few. Her interest in international security, peace, environment and development continues in various other posts that she holds including Vice-President of the Socialist International, Member of the Carnegie Commission on Preventing Deadly Con ict, and Member of the International Task Force On Security Council Enforcement. She previously served as a Member of the Palme Commission on Disarmament and Security.

Her life has touched and changed many in different circles and circumstances and under her leadership, radical opinions on women, politics, development and the environment were rekindled. Her central message has always been ethical and educative:

> ••• to succeed will require a new ethic, a new moral imperative and a new level of international awareness. It will require, more than anything else, a new thinking, a mind-set that has as its basic tenet a commitment to uphold the rights of peoples and nations to participate equitably in the resources of this planet and the development processes that they support.

MANSOUR KHALID *Kenya*

C

Cadmium: Management in the Environment, Including Food Chains

see Environmental Cadmium in the Food Chain (**Volume 3**)

Capacity Building

Capacity building is the strengthening of individuals and institutions, as well as the networks and systems in which they operate, so that they function better, are more innovative, and more resilient to threats to undermine them. Thus, capacity building does not only refer to training for individuals, but also to developing the institutional and policy infrastructure, which is essential to their success. Capacity building is a key element in human development, which was defined by the United Nations Development Program (UNDP) in 1990 as a process of enlarging people's choices so that they can lead long and healthy lives, be educated and enjoy a decent standard of living. Continuous capacity building is seen as an integral part of sustainable development.

The International Council for Science (ICSU) has defined capacity building in science as consisting of activities that lead to the strengthening of a corps of qualified scientists with a supporting infrastructure, including facilities and working conditions, which enables them to conduct research, education, training and advisory work. A number of international efforts have been launched to undertake capacity building for the environment and for global change, including Capacity 21 (UNDP); System for Analysis, Research and Training in Global Change Science (START-ICSU); and Leadership for Environment and Development International (LEAD) (*see* LEAD (Leadership for Environment and Development), Volume 4).

These programs recognize that research is not only conducted in universities, but also in government departments, non-governmental organizations and, at local levels, by community organizations. Strengthening the research system will only lead to sound environmental decision-making if there are also trained people and adequate institutions in policy and management to implement the research findings. Much of the focus in capacity building for environment has been in strengthening the national capacities of developing countries. Less attention has been paid to gender equity in science, and to training leaders who can strengthen research and policy systems from within.

ANNE WHYTE *Canada*

Carrying Capacity

Henry Regier
University of Toronto, Toronto, Canada

An organism of a species has natural needs that include food, shelter and a way to dispose of its wastes. A simplistic perspective takes these three considerations as a start from which to calculate the carrying capacity of an ecosystem or landscape for a local population of a particular species.

Basic needs of an organism can be inferred by observing organisms through their life span. Once those needs are known, and provided that the species has little adaptive capability for substituting abundant for scarce goods, the capability of the larger ecosystem to provide for these needs can be assessed. Dividing the measure of an ecosystem's capability by the measure of the needs of an individual organism provides an estimate of the total number of that population which can be sustained indefinitely in the ecosystem. This crude estimate of the total number is the *carrying capacity* of that ecosystem for that species.

In calculating carrying capacity, some particular supply factor necessary for the population may be inferred to be relatively less available than all the others. Such a limiting factor acts as an immediate constraint on the number or size of the population, if the particular species has no way of substituting elements of some abundant factor for the one in short supply. Predation, disease infestation and competition, as demands on the population, may also act as limits. Ecologists with interest in increasing the productivity of a particular species in an ecosystem, say of a species valued by hunters or fishers, may then seek to relax the constraint of a limiting factor through cultural interventions. Such management may include supplying additional food or shelter appropriate to the species, reducing disease infestation, and controlling competitors and predators.

Organisms in general have some self-organizing capabilities that relate to their basic needs. Thus, the life cycle of a species evolves with respect to seasonal availability of elements of these needs. Often some kind of reciprocity emerges between different species to reduce adverse effects on each other, e.g., from competition for scarce resources or other limiting ecosystemic features. Behavioral and social capabilities emerge which may guide an organism so that it does not induce strongly adverse ecosystemic reactions to its activities while satisfying its needs. Hence, it is seldom easy to estimate a species' carrying capacity realistically from a simple factorial analysis within a particular ecosystem. A more realistic measure may be inferable from comparative factorial analyses of different populations of the same species in somewhat different habitats.

A concept of carrying capacity is sometimes applied to the populations of humans in a region or even in the biosphere. Humans have complex behavioral, learning and social capabilities to intervene in the natural system, constructively and destructively with respect to the total number of humans whose basic needs can be accommodated. Malthusians tend to emphasize the destructive and wasteful capabilities while Cornucopians emphasize the constructive and conservative capabilities of humans.

An expert, with an extremely biocentric value-based criterion that satisfaction of human needs should not lead to the extinction of any other species, might assess the carrying capacity of the biosphere for humans as only a small fraction of the current population, i.e., that carrying capacity was surpassed long ago. Alternatively, an extremely technocentric and homocentric expert may guess that many times the present number of humans could thrive in a biosphere that would be almost completely artificial. For example, Caesar Marchetti, a futurist working at IIASA in Austria, once designed life-support systems for a trillion-person world. In the latter case, the limiting factor could be the technical capability for our planet to radiate waste heat to outer space, so that humans don't cook the whole surface of the globe.

Any realistic concept of the interrelationships of a species population and the larger ecosystem is complex. A simplistic notion of carrying capacity may lead researchers and managers astray, especially if applied to humans. Depending on how highly we value non-human nature, and whether we are willing to live at the subsistence level, or insist on an extravagant life style, estimates of the biosphere's carrying capacity may vary from 1 to 100 billion people.

Caspian Sea

see Caspian Sea (*Case study,* **Volume 4**)

CITES (Convention on International Trade in Endangered Species of Wild Fauna and Flora)

The international wildlife trade, worth billions of dollars annually, has caused massive declines in the numbers of many species of animals and plants. The scale of over-exploitation for trade aroused such concern for the survival of species that the CITES was adopted in Washington DC in 1973 to protect wildlife against such over-exploitation and to prevent international trade from threatening species with extinction. CITES entered into force in July 1975, and had 146 Parties in 1999. Its objectives are to ensure, through international cooperation, that the international trade in species of wild fauna and ora does not threaten survival in the wild of the species concerned, and to protect endangered species from over-exploitation by means of a system of import–export permits issued by a management authority under the control of a scientific authority. The Conference of Parties (COP) meets about every 2 years to consider assigning species to one of three appendices. Appendix I prohibits (with limited exceptions) the commercial international trade in wild-caught specimens of species threatened with extinction. Appendix II assigns the responsibility to exporting States to control, through a permit system, such trade in other species, which could become threatened with extinction if there were no such restriction. And Appendix III requires parties to control trade in specimens of still other species, which have been protected in

certain states and listed by those states. The secretariat of CITES is located in Switzerland, with a staff of about 25 and an annual budget of around US$5 million. The convention is governed between its COP and by the CITES Standing Committee which consists of about 16 members representing the major geographic regions and elected at the COP. The work is supported by several scientific and technical committees, including the Animals Committee, Plants Committee, Identification Manual Committee, and Nomenclature Committee; each of these is composed of persons chosen by the Parties. The website of CITES is http://www.cites.org.

JEFFREY A MCNEELY *Switzerland*

Clean Development Mechanism

The *Clean Development Mechanism* (CDM) is defined in the Kyoto Protocol to the United Nations Framework Convention on Climate Change. The CDM will become operational upon entry into force of the Kyoto Protocol.

The roots of the CDM can be found in Article 4.2 of the Convention which includes a provision that parties to the Climate Change Convention may implement policies and measures jointly with other parties.

In the Kyoto negotiations, the CDM evolved from a proposal by the Brazilian delegation for a Clean Development Fund, which would have received levies from parties found to be in non-compliance with their abatement commitments. Instead, the CDM was defined.

Under the CDM, projects are implemented in developing countries. Such projects both contribute to the sustainable development of these countries and generate emissions lower than an appropriately selected baseline, and thus earn certified emission reductions (CERs). Developed countries (listed in Annex I to the Convention) may use such CERs to comply with part of their commitment to limit and reduce emissions.

The CDM is subject to the authority of the Conference of the Parties (COP) serving as the Meeting of the Parties (MOP) to the Kyoto Protocol (COP/MOP) and is supervised by an executive board under the COP/MOP. Emission reductions are certified by operational entities designated by the COP/MOP. Private and public entities may participate in project implementation and acquire CERs. In this regard, it is also envisaged that the CDM would also provide a channel for investment into CDM projects.

A share of the proceeds from certified project activities would be used to cover administrative expenses and to assist those developing countries that are particularly vulnerable to the adverse effects of climate change meeting the costs of adaptation.

The COP, at its fourth session (COP 4) held in Buenos Aires, Argentina, in November 1998, decided that COP 6 (held in the Hague, the Netherlands in November, 2000) should finalize negotiations on principles, modalities, rules and guidelines for the CDM with a view to their adoption by the first session of COP/MOP following entry into force of the Kyoto Protocol. COP 5, held in Bonn, Germany, in October 1999, considered proposals by Parties on a number of such principles, modalities, rules and guidelines, which will be further developed during the period leading up to COP 6. Current work covers such aspects as approaches to project baselines, modalities for the monitoring and verification of projects, for the certification of emission reductions and for the designation of operational entities. Issues such as the composition and terms of reference of the Executive Board, the determination of the share of proceeds for administrative and adaptation costs and ways and means to arrange the funding of project activities, will need to be addressed.

TAHAR HADJ-SADOK *Germany*

Cleaner Production

Osama A El-Kholy
Arabian Gulf University, Manama, Bahrain

In May 1989, the United Nations Environment Program (UNEP) Governing Council asked UNEP to promote the establishment of a network to "allow the transfer of an environmental protection strategy". This was prompted by a growing realization that the cost of dealing with pollution resulting from human activities, particularly in industry and agriculture, after it has been generated (the so-called end-of-pipe approach) has been increasing dramatically over the years since the Stockholm Conference of 1972. A new approach now emphasized that, rather than wait until pollution is generated, it would be more sensible to attempt to prevent it in the first place.

As thinking moved along these lines, the investigation extended to consider other issues not looked into so far. These covered product design, the materials used to produce it, the production processes, and even the very social environment that created demand for such a product or service.

It also became clear that we must continue along this path, always doing better than before, up to the theoretical ideal of zero pollution.

The strategy of cleaner production (note the 'er' in cleaner, and the use of production rather than technology or product) formulated by UNEP to articulate this new approach was launched in a conference in Canterbury in 1990. The strategy is now adopted worldwide and applied in different ways, having shown a remarkable ability to interface and/or assimilate many of the ideas that developed over the last decade or so in environmental management.

FROM POLLUTION ABATEMENT TO POLLUTION PREVENTION

Concern about pollution from modern industry and agriculture in the industrialized countries was the main force behind organizing a United Nations (UN) Conference on the Human Environment held in Stockholm in 1972. This was one of the series of UN-sponsored conferences that dealt with global problems (desertification, water, population, etc.). The Stockholm Declaration at the end of this conference placed pollution at the top of the environmental agenda at the time.

Emphasis in the 1970s was on pollution abatement. Emissions and ef uents had to be treated to satisfy certain requirements before being released to the environment. The major industrialized countries were reported in the 1970s to be spending billions of dollars on the control of industrial pollution. In the then Federal Republic of Germany the share of investment in industrial pollution control between 1971 and 1977 was more than 5% of total industrial investment. In Japan, the ratio was as high as 10.6% in 1973, rising to 17.1% in 1975 under the in uence of more stringent legislation, before dropping to 5% in the early 1980s. In the USA, the ratio was 5.8% in 1975. The expenditure of private industry on pollution control in some industrialized countries varied between 0.8 and 1.7% of the gross national product (GNP) over the period 1972–1986. On average, countries with the most stringent environmental requirements spent about 1.5% of their GNP on pollution abatement. Of this, the manufacturing sector in the USA and Germany accounted for one quarter of this expenditure, or about 0.4% of GNP. A small number of industrial sectors (chemicals, metal products and machinery) accounted for a large part of this expenditure. As for the split between capital investments and operating costs of pollution treatment equipment, recurrent operating costs were in general, much higher, reaching almost five times the capital expenditures in the USA.

The situation was becoming progressively more serious as new legislation on the use of chemicals was promulgated in the mid-1970s. New industrial chemicals now had to go through stringent, expensive, and time-consuming tests before being marketed. Furthermore, the cost of monitoring compliance to the new requirements was mounting as more personnel, both in industry and in government, became involved in monitoring polluting releases. One estimate is that over US$300 billion was spent worldwide each year on environmental projects, most of which was spent on purchasing and maintaining treatment technologies. With the promotion of the *polluter pays principle* by the Organisation of Economic Co-operation and Development (OECD) in 1975, it was the consumer who foot the bill in the final analysis. This situation could not go on for obvious reasons.

During the latter half of the 1970s the concept of low and non-waste technologies emerged, now shifting the emphasis to pollution *prevention* rather than pollution *treatment*. The 1989 decision of UNEP's Governing Council was perhaps in response to this shift of emphasis. The UNEP Industry and Environment Office in Paris assembled a small core group of experts to work out a program for the implementation of the Governing Council decision.

It is both interesting and instructive to look back at the way thinking of the core group developed as work proceeded. First, they agreed that no justifiable distinction can be made between clean and dirty technology. There is no absolutely clean technology: one technology can only be shown to be cleaner than another. This means that there will always be room for improvement to develop still cleaner and cleaner technologies as time goes on. The ultimate ideal goal is a technology that produces no pollution at all, as if the production process is going on inside a bubble into which raw materials are introduced and out of which come only products (the bubble theory). Further consideration led to realization of the need to take into account the pollution a product produces as it is being used (e.g., the exhaust gases from a car). Expansion of the scope of the analysis finally extended both ways, downstream of manufacture and use of a product to its disposal as waste; and upstream to the supply side and the pollution produced in extracting and processing the inputs of raw materials and semi-finished goods. Consequently, the investigation needs to address products also, rather than technologies only. It has to extend across the whole life cycle of products (the so-called *cradle to grave analysis*). It has to investigate product design, production process, and management practices. Cleaner production was finally defined as in Box 1.

There is more to this compact definition than meets the eye:

1. Cleaner production is defined as a strategy, i.e., it extends well beyond technical fixes.
2. It is an all-embracing integrative view of the issue, that does not isolate and concentrate on one aspect only of the problem.

Box 1 What is Cleaner Production?

Cleaner production is the continuous application of an integrated preventive environmental strategy to processes and products to reduce risks to humans and the environment.

For production processes, cleaner production includes conserving raw materials and energy, eliminating toxic raw materials and reducing the quantity and toxicity of all emissions and wastes before they leave a process.

For products, the strategy focuses on reducing impacts along the entire life cycle of the product, from raw material extraction to the ultimate disposal of the product.

For services, it incorporates environmental concerns in designing and delivering services.

Cleaner production requires changing attitudes, responsible environmental management and evaluating technology options.

3. It emphasizes the continuity of the endeavor. We can always do better, *ad in nitum.*
4. The last paragraph means that we must also look into the social environment that created the demand for this product in the first place, and in uenced how it was to be produced.

The complexity of the concept can be more clearly articulated through a hierarchy of substitutions at different levels, taking as an example the motor car:

1. *the process level*: improving the combustion process (combustion chamber design, lean burning, etc.);
2. *the component level*: adding a new component (a catalytic converter to the exhaust pipe, an example of the end of pipe approach);
3. *the sub-system level*: adopting an entirely new technology (the electric car, fuel cells, hybrid engines);
4. *the system level*: adopting a new system (replacing private cars by mass transit, public transport systems);
5. *the value system*: questioning the very premises that created the need for the product (e.g., alternative landscape planning, value systems, work styles).

This hierarchy of substitutions moves the analysis from the purely technical all the way up to the social level, emphasizing the all-embracing, integrative nature of this definition.

UNEP has continued to hold senior level expert group meetings (Paris, 1990; Canterbury, 1992; Warsaw, 1994; Oxford 1996; Seoul, 1998). The meetings take stock of the status of the cleaner production movement worldwide, and review the work of an increasing number of specialized working groups that provide practical technical advice in application of the cleaner production concept as well as in policy formulation, and design practices.

THE CLEANER PRODUCTION STRATEGY IN PRACTICE

On the practical level, it is claimed that a review of the results of applying the cleaner production approach in 500 case studies of industrial firms has shown 70–100% reductions in some air emissions, water discharges and of the production of hazardous and non-hazardous solid wastes, and that the payback period was quite short, ranging from 3 years to less than 1 year. The late 1980s saw an increasing number of programs in many industrial countries, east and west, based on the cleaner production approach outlined above. One pioneering project in Europe is the Project on Industrial Successes with Waste Prevention (PRISMA) initiated in 1985 by the Netherlands Organization of Technology Assessment. When the preliminary results of field research started in 10 companies in 1988 became available, no less than 164 prevention options were identified. These included minor changes in procedures and installations as well as drastic innovations in products or processes. In 30% of the cases these changes were in the category of good housekeeping; another 30% in material and raw materials; 30% in changes in equipment; and the rest in process modifications (Sybren do Hoo *et al.*, 1991). Some good housekeeping measures effected 25–30% reductions in chemicals used. In a number of companies, technological changes resulted in reductions in waste and emissions of 30–80%. Occasionally, it was possible to eliminate a noxious waste ow completely. The use of alternative raw materials resulted in a 100% reduction in emissions of substances such as cyanide (in zinc plating) and solvents (in decreasing parts).

The cost-benefit analysis showed a heterogeneous picture with a few extremely favorable peaks. In one extreme case an investment of DFl 2500 saved the company DFl 124 000 *per annum*. Another company recouped an investment of DFl 800 000 within a year (saving more than DFl 1 000 000 *per annum*). Of the 45 options actually implemented by then, 20 turned out to be cost saving, and 19 to be neutral. It was found, moreover, that these measures had further indirect benefits, such as improvement in product quality.

Within EUREKA (the co-operative research program in Europe), the EUROENVIRON program began promoting Preventive Environmental Protection Approaches in Europe (PREPARE) in 1990 on the basis of the Dutch PRISMA project. PREPARE sought also to catalyze industry/government cooperation and the transfer of information of pollution prevention and cleaner production.

In the United States, prevention teams have been in operation for some years as part of the Technical Assistance Program , helping small businesses in adopting pollution prevention technologies. Groups of external consultants assist industrial firms in formulating and implementing a pollution prevention program. Evaluation of their work indicates that these programs were not only commercially

profitable and environmentally beneficial, but also useful at the macroeconomic level, as well as helpful in clarifying gaps in current knowledge and identifying areas for research.

Mexico and Brazil have achieved some noteworthy successes in organizing efforts for pollution control at various levels of government. The program in Cubatao, home of a large industrial complex in the state of Sao Paulo in Brazil, is one important example of a unique institutional system. Enforcement powers were delegated by contract to the Environment Management Company and an integrated control approach was started in 1984. Some of the results achieved by the end of 1989 in air pollution abatement are summarized in Table 1. Similar success was also achieved in reducing water pollutants.

In China in the early 1990s, UNEP, with funding from the World Bank, joined hands with the National Environment Protection Agency (NEPA) to promote cleaner production throughout China.

In 1993, when the preparatory phase was completed, substantial economic and environmental benefits were identified, even at this early stage of the program:

- Sixty-seven cleaner production options, ranging from managerial changes to process modifications were implemented. Most of these were of the low–no cost type.
- Over US$350 000 saving was achieved for investment of less than US$16 000.
- An average of more than 50% of the chemical oxygen demand in the waste water was eliminated.

In the latter phases, experts were trained to carry out cleaner production activities throughout the country, particularly in small and medium enterprises (SMEs). General and sector specific manuals were published and policy studies carried out. This program has resulted in a gradual shift from investment in end-of-pipe equipment to integrated cleaner production approaches throughout the country.

Many other initiatives in East Europe, the Mediterranean Basin, India, South East Asia, and in Africa could also be cited. The United Nations Industrial Development Organization (UNIDO) cooperated with UNEP in establishing national cleaner production centers (NCPCs). The first batch of eight NCPCs were started in 1994 with funding from European donors and located in Brazil, China, Czech Republic, India, Mexico, Slovak Republic, United Republic of Tanzania and Zimbabwe. NCPCs are spreading gradually worldwide with self-financing in some countries and funding from donors in most cases (Tunisia, Vietnam, Costa Rica, Nicaragua, Croatia, El Salvador, Guatemala and Uzbekistan). A review carried out in July 1998 states that during the period from 1994 up to July 1998, 405 plants participated in in-plant assessments, 1016 consultants and participants were trained, 16 000 persons participated in seminars and more than 7700 attended workshops. The review summarizes the main lessons learnt as:

- A 5-year period is necessary to develop a NCPC into a service-oriented unit.
- A system of certification would enhance the status of cleaner production experts.
- Cleaner production can easily be incorporated into and promoted by other related bodies, such as environmental management systems (EMS) or ISO 14000. (the UNEP/International Chamber of Commerce (ICC)/Fédération Internationale des Ingénieurs-Conseils (FIDIC) EMS Training Resources Kit explains the relationship between EMS/ISO 14000 and cleaner production).
- While the methodology of cleaner production is applicable to all industrial sectors, sector-specific technical expertise needs to be developed.
- An integrated program, rather than a project approach is necessary for sustainable cleaner production implementation.

After almost a decade since the launch of the cleaner production program by UNEP, it is useful to take stock of past experiences and to look ahead into the future:

- There is no doubt that cleaner production has gained remarkable popularity over a short period of time. This is mainly due to the fact that launching cleaner production coincided with a paradigm shift in pollution management from end-of-pipe treatment to pollution prevention. It may seem to have taken a long time to apply the time-honored principle, *prevention is better than cure* but we had to deal with the natural resistance to changing established practices in a functioning establishment. In good economic times, the costs of end-of-pipe treatment could be easily absorbed with other costs. Furthermore, there were no compelling reasons to internalize environmental costs in the total cost of products. It must also be admitted that regulations in the 1970s and early 1980s focused

Table 1 Reduction in atmospheric pollution in Cubatao, Brazil

Pollutant	Initial emissions in July 1984 (kg day^{-1})	Emissions in December 1989 (kg day^{-1})	Actual reduction (%)
Particulates	316 350	77 949	75
Fluorides	2620	400	85
NH_3	8736	205	97
NO_x	61 085	47 561	22
HC	90 000	11 970	86
SO_2	78 353	49 527	37

on end-of-pipe solutions and offered few incentives to pollution prevention initiatives.

- In newly industrializing countries and developing countries, weak monitoring of compliance with environmental regulations, and enforcement mechanisms, mitigated against adoption of cleaner production practices. One could also cite the reluctance of most banks to fund cleaner production projects, due to lack of experience in financial risk assessment of environmental projects.
- However, cleaner production received a tremendous boost once demonstration projects provided concrete evidence of the economic advantages of a host of simple low–no cost actions as mentioned earlier. Pollution prevention came to be considered as a win-win policy. The 3Ps, which stood for the *polluter pays principle* in 1975, now stood for *pollution prevention pays*. Several transnational corporations mainly in the chemical industries, announced their unilateral commitment to reduce the pollutants they discharge by fixed amounts over specific time horizons. Eventually, the law of diminishing returns showed that there is a limit which is reached when the easy, low–no cost options are exhausted. Du Pont revived Pareto's 80/20 principle pointing out that the challenge facing management is to identify the actions that account for the first 20% of costs that achieve maximum environmental benefits.
- Most developing countries still have a long way to go, reaping environmental benefits together with economic benefits through relatively simple, short pay-back periods, low–no cost options. What they now need is the local expertise, enlightened management and relatively small upfront financing of the more complex options that have proven economic benefits. Demonstration projects in developing countries continue to show that application of cleaner production practices, more often than not, yield immediate unmistakable economic and environmental benefits, particularly in SMEs. Unfortunately, and for obvious reasons mentioned earlier, this is exactly where

Box 2 International Declaration on Cleaner Production (adopted at Fifth International High-Level Seminar on Cleaner Production, Korea, 1998)

We recognize that achieving sustainable development is a collective responsibility. Action to improve the global environment must include the adoption of more sustainable production and consumption practices.

We recognize that cleaner production and other preventive strategies such as eco-efficiency, green productivity and pollution prevention are preferred options and require the development, support and implementation of appropriate policies and practices.

We understand cleaner production to be the continuous application of an integrated, preventive strategy applied to processes, products and services to reduce risk and to pursue economic, health, safety and environmental benefits.

To this end we are committed to:

Transparency sharing our cleaner production experience:

- by fostering and promoting dialogue on the implementation of this strategy;
- through communication of the benefits to external stakeholders.

Integration encouraging the integration of preventive strategies:

- into all levels of our organization;
- through EMS, and tools such as environmental performance evaluation, and cleaner production, environmental impact, and life cycle assessments.

building cleaner production capacity:

- by developing and conducting awareness, education and training programs to facilitate the practice within our organization;
- by encouraging the inclusion of the concepts and principles into educational curricula at all levels.

Creating innovative solutions:

- by promoting a shift of priority from end-of-pipe to preventive strategies in our research and development policies and activities;
- through the development of products and services which are environmentally efficient and meet consumer needs.

Leadership using our influence:

- to encourage the adoption of cleaner production and sustainable consumption practices through our relationships with stakeholders.

Implementation taking action to adopt cleaner production:

- with continual improvement by setting challenging goals and regularly reporting progress through established management systems;
- by encouraging new and additional finance and investment in preventive technology options, and promoting cleaner technology cooperation and transfer between countries;
- through work with UNEP and other partners and stakeholders to support this declaration and review the success of its implementation.

cleaner production has not yet penetrated on any significant scale. Positive technical/financial support on an adequate scale is necessary. Such intervention is cost-effective and arrangements could, and have been, concluded with enterprises to pay back part, if not all, of the resulting savings to cover the cost of implementation.

- Cleaner production has stood the test of time. The original definition has demonstrated its compatibility with, and even ability to take under its wing, new ideas in sound environmental management (e.g., eco-efficiency, resource accounting, sustainable consumption, ISO 14000 standards or life cycle assessment and eco-labelling, etc.).
- Although cleaner production has become something of a buzz word in common use, there are clear signs that it is often used in restricted senses that deprive it from its all-embracing advantage. The most common form of misuse is equating it to pollution prevention particularly in production processes. Perhaps the social aspect of consumption patterns is the one field that has not yet been investigated in depth in operationalizing the definition formulated by UNEP a decade ago.
- With the launch of the International Declaration on Cleaner Production (Box 2), cleaner production has entered a new phase, aimed at creating a multi-stakeholder partnership. It is hoped that this will increase proper understanding and awareness of cleaner production, as well as increasing demand and renewed commitment of cleaner production practitioners. Led by the UNEP Executive Director and the Korean Environment Minister, 64 ministers, CEOs directors of environmental agencies, NGOs and academic institutions signed the Declaration which continues to attract more signatories.

FURTHER READING

Hirschhorn, J H and Oldenburg, K (1991) *Prosperity without Pollution: The Preventive Strategy for Industry and Consumers*, Van Nostrand Reinhold, New York.

Huisingh, D (1989) Cleaner Technologies through Process Modifications, Material Substitutions and Ecologically-based Ethical Values, *Ind. Environ.*, 12(1), 7.

Sybren de Hoo, Bezet, H, Crul, M, and Dieleman, H (1991) *Prisma: Industrial Success with Pollution Prevention*, Nederlandse Organisatie voor Technologisch Aspectenonderzoek (NOTA), Den Haag.

Tolba, M K, El-Kholy, O A, El-Hinnawi, E, Holdgate, M E, McMichael, D F, and Munn, R E, eds (1992) *The World Environment 1972- 1992: Two Decades of Challenges*, Chapman & Hall, London.

UNEP (1993) Cleaner Production Worldwide, UNEP, Paris.

UNEP (1994) Cleaner Production, *Ind. Environ.*, 17(4).

UNEP (1996) Cleaner Production, *Ind. Environ.*, 19(3).

UNIDO/UNEP (1998) National Cleaner Production Centres, Activity Report, June 1998 (mimeograph).

Climate Change: an Emerging Issue in the Global Agenda

Michael Zammit Cutajar
United Nations Framework Convention on Climate Change, Bonn, Germany

This article discusses developments with respect to the UN Framework Convention on Climate Change up to early 2001. Climate change is a quintessentially global issue. The global climate system is a mega-resource that humanity - especially agrarian humanity - has taken for granted for millennia. It is being destabilized by the economic and demographic growth of the last two centuries. Emissions of carbon dioxide and other heat-trapping gases from all sources and all social origins - wasteful emissions and survival emissions alike - mix indiscriminately in our common atmosphere. There they enhance the natural greenhouse effect, making the world warmer than it should naturally be.

The international response to climate change is a negotiating process, from which a strategy is emerging, so far embodied in the 1992 United Nations Framework Convention on Climate Change, and its 1997 Kyoto Protocol. The first step, the Convention, is an agenda-setting framework, based on principles. Its objective is to stabilize greenhouse gases in the atmosphere at safe levels. It provides for: cooperation and exchange of information; leadership by industrialized countries in modifying long-term emission trends; financial and technological support for the efforts of developing countries; and responses to vulnerability. It has been in force since 1994 and, with over 180 parties, is nearly universal. The Protocol would take a significant further step by binding industrialized countries to verifiable emission limitation and reduction targets. Its first commitment period aims at an aggregate reduction of greenhouse gas emissions of some 5% below 1990 levels, in the period 2008- 2012.

THE DYNAMICS OF NEGOTIATION

The negotiating process is driven primarily by the scientific assessments of the independent Intergovernmental Panel on Climate Change (IPCC), the third of which (2001) confirmed evidence that distinguishes the human contribution

to global warming from natural climate variability. The message of the mainstream scientists, in summary, is that the global climate is changing as a result of human activity (*see* Intergovernmental Panel on Climate Change (IPCC): an Historical Review, Volume 4).

There is some dissent from these findings of the IPCC. This is vocal but limited. In some cases, it is discredited by association with vested economic interests that have much to lose from action to use fossil fuels less wastefully.

Perhaps more disturbing is the wait-and-see approach advocated by many business and industry groups, allied to predictions of job losses. This defensive attitude does not give due weight to the costs of doing nothing – the costs of the impacts of climate change on vulnerable ecosystems, communities and economies. The vulnerable segments of global society do not wield economic clout, it would seem.

It is encouraging, however, to note that the *precautionary principle* (*see* Precautionary Principle, Volume 4), which is enshrined in the Convention, is attracting an increasing number of adherents from within the industrial and business sectors. The insurance industry took the lead in this direction, accompanied by companies with an interest in promoting energy conservation, and producers of solar and wind power. Even within the fossil fuel industry, natural gas producers have tended to favor action to limit emissions of carbon dioxide, since their product contains less of this greenhouse gas than coal or oil. And major petroleum companies based in Europe have shifted towards a precautionary approach.

The message of the IPCC has difficulty getting through, beyond those involved in the climate negotiations (mainly in environment ministries), to the economic policymakers, public and corporate. Admittedly, it is sometimes difficult to see through the scientific clothing in which the message is wrapped. More needs to be done to deliver it clearly and loudly. More fundamentally, though, the long-term nature of the problem makes it difficult to engage political and public attention. Few citizens, and even fewer politicians, are concerned with a problem that will face their grandchildren's grandchildren.

Thus, the effort must continue to convince Northern policymakers that they cannot buy their way out of the problem without making the necessary structural adjustments in domestic production and consumption. For their part, Southern policymakers need to be convinced that mitigating and adapting to climate change should figure in their present priorities for sustainable development. The message to both is that action now will cost less than action later. A further message to the South, sadly familiar, is that it is the poor countries who are most vulnerable and who will bear the burden of adverse climatic impacts.

POSITIVE STRATEGIES

In both cases, North and South, it is possible to identify actions that make economic sense for the countries concerned while also addressing climate change by limiting greenhouse gas emissions (*see* No Regrets Principle, Volume 4). This approach, known as no regrets enables climate strategists to address the problem without basing the justification for action on climate science alone. Thus, for example, one can seek emission reductions by promoting policies to deal with more tangible environmental problems such as urban air quality, traffic congestion and the spread of transport infrastructure – problems that voters and taxpayers recognize more readily as their own problems, here and now. Local governments and municipal authorities can be lead players in such a strategy.

Beyond the no-regrets approach, one senses that, behind its defensive walls, industry is exploring possibilities for making money out of cleaner, climate-friendly technologies. Regular articles in the press about experimental low-emission motor vehicle engines are perhaps the most important indicator of this technological exploration, automobile transport being the source of the most rapid growth in carbon dioxide emissions. A climate-friendly technology fix does not contradict economic growth, provides opportunities for profitable investment and bridges North/South economic interests. Consequently, technology – its development, application and transfer – is a regular theme in discussions under the Convention though, so far, little has been done to move beyond debate in this area. Positive measures to promote transfer, through public-private partnerships, are likely to be needed.

CONSUMPTION PATTERNS

Less is heard about consumption patterns and lifestyles. On the one hand, the American way of life is held sacred and it is considered to be political suicide to tamper with it. On the other hand, this same way of life is being globalized through the exponential spread of communications, despite the fact that it is not replicable on a global scale at forecast levels of population and given present technology.

Discussion under the Convention does not neglect the consumption aspect totally. Proposals have been tabled concerning technical standards for consumer goods and fiscal measures that stimulate changes in consumption habits. However sensible such proposals may be, they are caught up in a negotiation where some (such as the European Union) argue that such measures should be adopted by all industrialized countries in order to maintain a level playing field in international trade, while others (such as the United States) press for maximum exibility to

each country to choose how best to achieve reductions in emissions.

MAKING COMMITMENTS EFFECTIVE

While the first phases of negotiations have concentrated on action by industrialized countries to limit their emissions, it is clear that the objective of the Convention – the stabilization of atmospheric concentrations of greenhouse gases at a safe level – cannot be achieved by the presently developed countries alone. They are obliged to take the lead in cutting emissions and modifying long-term emission trends, because of their historical contribution to concentrations and their high emissions per head. However, their efforts could be negated by the growth of emissions from developing countries, needed to meet the needs of these countries for material well-being.

Yet developing countries will only consider limiting their emissions in ways that would not place a new constraint on their economic growth, and even then not until they are convinced that developed countries are serious about cutting theirs. The evidence so far is not convincing; no Organization for Economic Co-operation and Development country achieved the initial target under the Convention (returning emissions to 1990 levels by the year 2000) on account of policies explicitly addressing climate change. Some such as the UK and Germany achieved the target for other reasons. One can argue that it does not matter how they achieve the aim as long as they do so, but a happy accident is not a rational strategy and will not be replicated. Currently (early 2001) negotiations are under way on ways of putting teeth into the Kyoto commitments of developed countries. These, in turn, are experimenting with domestic measures and economic instruments aimed at achieving emission limitation at minimum economic and social cost.

FUTURE NEGOTIATIONS

The forward vision that can be derived from these early steps is of a series of negotiating rounds among the parties, leading to progressive degrees of commitment and institutional development – not unlike the history of trade negotiations since 1948. The commitments will be differentiated according to responsibility and capacity. The strategy will be implemented by the parties themselves; several multilateral and regional agencies and programs will support them.

Efficiency in limiting future emissions of greenhouse gases is vital for success in achieving the ultimate objective of the Convention. Equity in distributing responsibility for these limitations is vital for political engagement. Visions of the economic opportunities of a climate-friendly future would transform the language of negotiation from burden sharing to win-win.

This linkage with the economy relates the climate change scenario to that of globalization and global sustainable development. Climate change is a by-product of two centuries of economic growth, so action to address it must pass through the economy of the 21st century. Greater efficiency and cleaner technologies are needed in the energy, transport and industrial sectors, among others. This can be a driver of technological change, green jobs and green profits. The international response can also generate a new push in favor of the transfer of technology and know-how to developing countries and capacity-building.

A novelty of the emerging strategy is its use of international economic instruments. The mechanisms of the Kyoto Protocol would allow industrialized countries to meet part of their emission targets offshore, through economic transactions. One mechanism – *emissions trading* (*see* Tradable Permits for Greenhouse Gases, Volume 4) – would permit an industrialized country to acquire part of the emission allowance of another. The other two mechanisms *joint implementation* (*see* Joint Implementation, Volume 4) and the *clean development mechanism* (CDM) (*see* Clean Development Mechanism, Volume 4) – allow the acquisition of credits for emissions averted by investments in climate-friendly projects in industrialized or developing countries. Together, the three mechanisms would create a new tradable commodity, namely carbon; a new service industry: the verification and certification of transactions in the carbon market; and a new motivation for foreign direct investment in low-carbon technologies.

One of the two project-oriented mechanisms, the CDM, also has the aim of assisting developing countries in achieving sustainable development and generating finance to help them adapt to the adverse effects of climate change.

All three mechanisms seek to engage the private sector in advancing the international strategy. They are thus in tune with the spirit of the times.

See also: Common but Differentiated Responsibility Principle (Stockholm/Rio), Volume 4; Conference of Parties, Volume 4; Intergovernmental Panel on Climate Change (IPCC): an Historical Review, Volume 4.

Climate Change Assessment, United States

see Climate Change Assessment, United States (*Case study*, **Volume 4**)

Club of Rome

Ricardo Diez-Hochleitner
President, The Club of Rome, Madrid, Spain

The Club of Rome is a global think tank and center of innovation and initiative. As a non-pro t, non-governmental organization (NGO), it brings together scientists, economists, businessmen, international high civil servants, heads of state and former heads of state from all ve continents who are convinced that the future of humankind is not determined once and for all and that each individual can contribute to the improvement of our societies.

Founded in Rome 1968, the approach of the Club of Rome to the solution of the world problematique is the identification of problems, their analysis in a global, interdisciplinary and long term perspective, the research of future alternative solutions and the elaboration of scenarios for the future. The results of this work are communicated to public and private decision-makers and to the general public. The methods for such communications are reports, conferences, and the internet. The Club of Rome's essential mission is thus to act as an independent, global, non-official catalyst of change.

The first report commissioned and published by The Club of Rome was *The Limits to Growth*, a book which produced a world-wide impact (more than 12 million copies have been sold in 37 languages). Its thesis was interpreted in many different ways. It stressed above all, for the first time, the importance of the environment, and the essential links with population and energy. This was a particularly striking illustration of what is meant by the global problematique. It was understood as a break and change of paradigm.

The Club of Rome is governed by three complementary principles:

1. A global perspective in examining issues with the awareness that the increasing interdependence of nations and the globalization of problems pose predicaments beyond the capacity of individual countries.
2. Holistic thinking and the seeking of a deeper understanding of complexity within the contemporary problems – political, social, economic, technological, environmental, psychological and cultural – which the Club of Rome terms the world problematique.
3. An interdisciplinary and long-term perspective focussing on the scenarios, alternative choices and policies determining the destiny of future generations, because this perspective is too often neglected by governments and other decision-makers on account of short term interests.

The report *The Limits to Growth*, a seminal one for the Club, has been followed by 26 other reports, on problems ranging from education, energy, the impact of microelectronics on society, to governance, the role of NGOs in development and the environment. Several others are under preparation.

It is important to emphasize that, in addition to their role as information sources, these reports must also serve two essential functions. First, they represent the starting point of a process and not its culmination. Their publication allows the Club, which has no claim to possessing the truth, to launch wide-ranging debates to discuss the conclusions and so share with a variety of audiences, from governments to universities and the media, its thinking and suggestions for resolving the problems affecting the world. Secondly, they are the starting point for certain initiatives implemented at the highest levels of decision-making, both public and private, with a view to fostering changes to strategic factors.

Every year, the Club of Rome holds its Annual Conference in a different region of the world. In addition to the working sessions reserved for members of the Club, these meetings bring together personalities from the region or the world who are invited to take part in discussions on a particular aspect of the world problematique. These meetings have established fruitful inter-personal relations with leaders and activists in the region, enabling a better understanding of the region's specific problems and its perception of global issues and the role it would like to play in them. Beside these annual meetings, the Club of Rome organizes occasionally regional meetings on regional aspects of the world problematique.

However, the Club itself tends normally to adopt a low profile, and the passionate debate sparked by *The Limits to Growth* has been the only and unexpected exception to this desire to operate discreetly. We believe that we are sometimes more effective when we work behind the scenes.

For further information, see http://www.clubofrome.org.

Coastal Zone Management

Mahmoud Kh El-Sayed
University of Alexandria, Alexandria, Egypt

Coastal zones throughout the world have historically been among the most heavily exploited areas because of their rich resources. If these coastal resources are to be maintained and restored, effective action is urgently needed. To

answer this need, a management system has been designed: integrated coastal zone management (ICZM).

ICZM is a process of governance consisting of the legal and institutional framework necessary to ensure that development and management plans for coastal zones are integrated with environmental (including social) goals and are made with the participation of those affected. The purpose of ICZM is to maximize the bene ts provided by the coastal zone and to minimize the con icts and harmful effects of activities upon each other, on resources and on the environment (World Bank, 1996).

It is worth noting that the terminology of coastal zone management (CZM), ICZM, integrated coastal and marine areas management, and coastal resource management, are all used interchangeably.

The coastal zone is generally understood to be the place where the land meets the ocean, comprising significant marine and terrestrial areas. The limits of the coastal zone are often arbitrarily defined, differing widely among nations, and are often based on jurisdictional limits or bounded by reasons of administrative ease.

Coastal areas contain some of the world's most productive and diverse resources. These resources are used and exploited for economic and social objectives. Population pressure, particularly the growth of urban areas, combined with the rapid expansion of industry and tourism associated with services, utilities and activities has created a worldwide concern about sustainable development of these areas.

These sectoral activities produce combined environmental impacts resulting in marine and fresh water pollution, air pollution, loss of marine resources, loss of natural land resources and land degradation, destruction of historic and architectural heritage, loss of public access to the coast, noise and congestion. Natural risks and hazards, such as earthquakes, forest fires and oods, pose additional hazards to coastal areas (United Nations Environment Program, 1995). In addition, serious environmental problems are likely to arise in association with potential impacts of expected climate change and sea level rise in coastal and deltaic areas worldwide. Policy makers are faced with the challenge of ensuring economic development while limiting the adverse impacts of such development on natural areas and protecting human life and infrastructure. Sectoral development interests, which prevail in coastal areas, make decision-making even more difficult and require mechanisms for cross-sectoral cooperation.

In both developed and developing countries, the coastal zone is likely to undergo the most profound change in the near future. Already more than 60% of the world's population lives within the coastal areas and many of these people may have to cope with significant rises in sea level before the end of the century. Consequently, unless careful environmental management and planning are instituted, severe con icts over coastal space and resource utilization are likely, and the degradation of natural resources will close development options (World Bank, 1996).

A basis for action, objectives, and activities related to Integrated Management and Sustainable Development of Coastal and Marine Areas, including Exclusive Economic Zones were presented in Chapter 17 (para 17.3–17.7) of Agenda 21. Recognizing the intensifying threats and the environmental pressure in coastal areas, the United Nations Conference on Environment and Development (UNCED) in Rio de Janeiro, 1992, emphasized that:

> The marine environment – including the oceans and all seas and adjacent coastal areas – forms an integrated whole that is an essential component of the global life-support system and a positive asset that presents opportunities for sustainable development; and that the coastal area contains diverse and productive habitats and ecosystems important for human settlement, development and local subsistence.

The Conference recommended that a new integrated approach to marine and coastal area management and development, at the national, subregional, regional and global levels be developed to minimize con icts and to provide for optimal sustainable resource use.

The concept of developing CZM plans, guidelines and principles was therefore considered as priority for most of the coastal nations all over the world.

Since the directives, priorities and the decision-making process differ from national, regional and global perspectives regarding sustainable development; there are many definitions of CZM. However, integration, sustainability and the socio-cultural dimension remain as common subjects in most definitions.

The World Coast Conference '93 Statement defines ICZM as a continuous and evolutionary process for achieving sustainable development, that involves the comprehensive assessment, setting of objectives, planning and management of coastal areas and resources, taking into account traditional cultural and historical perspectives and con icting interests and uses.

The common elements of the CZM process are:

- problem identification and assessment;
- program design and preparation;
- program adoption and financing;
- program implementation;
- monitoring and evaluation.

In all of the above stages, strong political will and commitment at all decision-making and stakeholder levels are needed if the CZM process is to be carried out successfully.

The development and enforcement of a CZM plan requires an adequate implementation structure, which needs to be defined in terms of the following dimensions:

- institutional;
- legislative;
- financial;
- human resources.

REFERENCES

United Nations Environment Program (1995) *Guidelines for Integrated Management of Coastal and Marine Areas*, UNEP Regional Seas Reports and Studies, No. 161.

World Bank (1996) *Guidelines for Integrated Coastal Zone Management, Evironmentally Sustainable Development Studies and Monographs Series, No. 9*, eds J C Post and C G Lundin, World Bank, Washington, DC, 1–16.

Command-and-control

see International Environmental Law (**Volume 5**)

Commission on Sustainable Development (CSD)

see CSD (Commission on Sustainable Development) (**Volume 4**)

Common but Differentiated Responsibility Principle (Stockholm/Rio)

The *common but differentiated responsibility principle* (Stockholm/Rio) is a significant product of the United Nations Conference on Environment and Development (UNCED). The principle affirms that developed nations accept that they bear primary responsibility for the present state of environmental degradation, and in the meantime all nations must cooperate to prevent further degradation. Principle 7 of the UNCED Rio Declaration (1992) asserts:

> States shall cooperate in a spirit of global partnership to conserve, protect, and restore the health and integrity of the Earth's ecosystem. In view of the different contributions to global environmental degradation, States have common but differentiated responsibilities. The developed countries acknowledge the responsibility they bear in the international pursuit of sustainable development in view of the pressures their societies place on the global environment and of the technologies and financial resources they command.

The principle is also embedded in the Framework Convention on Climate Change and The Convention on Biological Diversity (both signed during the UNCED). The principle of differentiated responsibility has also been articulated in previous international agreements; for example the 1972 Stockholm declaration.

See also: Convention on Biological Diversity, Volume 4; UNCED (United Nations Conference on Environment and Development), Volume 4; United Nations Framework Convention on Climate Change and Kyoto Protocol, Volume 4.

KEVIN S HANNA *Canada*

Conference of Parties

When a convention/treaty/protocol is adopted by a plenipotentiary conference, i.e., a conference of government representatives who are officially authorized to adopt the convention, it is then open for signature together with what is called the *nal act*. The latter records what happened at the time of adopting the convention, who attended and any reservations made. To the final act, is annexed the treaty itself. Government representatives may sign only the final act or may sign the final act and the treaty/convention. After signing the convention it goes through the process of ratification. Ratification means endorsement by the relevant authority in the country. This is generally the parliament but could be a revolutionary council or a high committee of the party in the single party system countries.

A specific number of ratifications are required for the treaty to enter into force. This varies from one convention to another and the number is stipulated in the body of the convention. Once the required number of instruments of ratification is deposited with the depository, the treaty enters into force. Countries which ratify the treaty constitute the *Conference of the Parties* which meets regularly, normally every one or two years, to follow up on the implementation of the convention, take the necessary steps for its amendments, approve the budget for its secretariat and so on.

MOSTAFA K TOLBA *Egypt*

Control of Biological Invasions

see Biological Invasions (*Opening essay*, **Volume 2**)

Convention on Biological Diversity

Jeffrey A McNeely
IUCN, The World Conservation Union, Gland, Switzerland

The 1992 United Nations Conference on Environment and Development (the Earth Summit) put biological diversity on the international agenda when its participants signed the Convention on Biological Diversity (CBD). The CBD entered into force on the 29th December 1993 and by the end of 1999 had 178 State Parties. Its secretariat is based in Montreal, Canada, with some 34 staff and an annual budget of US$9.7 million (E-mail: secretariat@biodiv.org). The CBD commits the nations of the world to the objectives of conserving biological diversity, using biological resources sustainably, and sharing equitably the bene ts arising from the use of genetic resources. Its objectives thereby bring together elements that address biological, economic, and ethical concerns. It is the rst global agreement to address comprehensively all levels of biodiversity, from genetic resources to species and ecosystems.

The idea of such a convention began at the Third World Congress on National Parks, held in Bali, Indonesia, in October 1982 and was developed further by International Union for the Conservation of Nature's (IUCN's) Environmental Law Centre over the following several years. In August 1988, United Nations Environmental Programme (UNEP) Executive Director Mostafa Tolba convened a high level panel of experts to advise him on whether a global biodiversity convention was timely, and if so, what it should include. UNEP then convened a series of intergovernmental meetings to develop the CBD for signature at Rio de Janeiro. The CBD is a landmark legal instrument in that it takes for the rst time a comprehensive approach to the conservation of biodiversity and sustainable use of biological resources. It recognizes the justi cation for conservation on the basis of ethics, economic bene t and human survival. However, the CBD goes beyond conservation and sustainable use to encompass such issues as ensuring access to genetic resources, sharing of bene ts arising from the use of genetic material, dealing with potential hazards of biotechnology, and promoting access to information and technology.

The CBD also implicitly recognizes that biodiversity is unevenly distributed on our planet. While the economically prosperous industrialized countries are typically relatively poor in biological wealth, the developing countries of the tropics have many more species and much greater genetic diversity. Thus the responsibility for conserving biodiversity lies especially with the actions to be taken in tropical countries, at the very time when the use of these biological resources may be critically important for developing their domestic economies. The CBD recognizes that it is in the enlightened self-interest for the industrialized countries to provide nancial and other kinds of support to the tropical countries to enable them to carry out their obligations under the convention and thereby provide global bene ts in terms of better management of biological resources.

The CBD is also innovative in its construction as an umbrella agreement, leaving it up to individual Parties to determine how its provisions are to be implemented while providing advice on how this might be accomplished. Containing no lists of species or sites to be protected, the articles of the convention express overall goals and policies rather than elaborated obligations and targets.

CONTENTS OF THE CONVENTION

The CBD defines biological diversity as the variability among living organisms from all sources including, *inter alia*, terrestrial, marine, and other aquatic ecosystems and the ecological complexes of which they are a part; this includes diversity within species, between species, and of ecosystems . While biodiversity has many definitions, the fact that nearly 180 governments have accepted this one gives it particular power. The CBD also recognizes the principle in international law, that states have the sovereign right to exploit their own resources according to their own environmental policies, but that they also have the responsibility to ensure that activities within their jurisdiction or control do not cause damage to the biodiversity of other states. It calls for the parties to cooperate with others both directly and through international organizations on issues of biodiversity, and to involve the private sector in doing so.

Article 6 of the CBD builds the framework for national implementation by calling on each party to prepare a national biodiversity strategy and action plan; to include biodiversity concerns into the plans and policies for other sectors such as agriculture, fisheries, forestry, and tourism; and to incorporate biodiversity into cross-sectoral matters such as land-use planning, trade and economic policy. By the year 2000, nearly 150 countries had either published their national biodiversity strategy or were actively engaged in preparing it.

BUILDING THE KNOWLEDGE BASE

Knowledge is the foundation upon which action in the field of biodiversity can be based. Therefore, Article 7 of the convention encourages parties to identify and monitor genes, species, and ecosystems that the party considers particularly important (because of special economic, social, or biological value to the country), and to identify and monitor its activities that may have significant adverse impacts on biodiversity. The information generated will enable countries to determine their own priorities for conserving biodiversity and using biological resources sustainably, and foster international exchange of information. However, consistent with the notion of a precautionary approach, parties also recognize that lack of knowledge is not a valid reason for not taking appropriate action to conserve biodiversity.

IN SITU CONSERVATION

Conserving ecosystems and species in their natural settings is the essence of the convention. Parties are encouraged in Article 8 to emphasize *in situ* conservation through actions such as: establishing and managing a representative system of protected areas; conserving and sustainably using biological resources within and outside protected areas; promoting environmentally sound and sustainable development in areas adjacent to protected areas; rehabilitating degraded ecosystems and promoting the recovery of threatened species; controlling alien species and the risks associated with genetically modified organisms; protecting threatened species and populations, and regulating processes and activities that may adversely affect biodiversity. Article 8 thus seeks an appropriate balance between preservation and sustainable use. It also specifically calls for cooperation between countries in providing financial or other support to developing countries to implement the measures called for by the CBD. By the end of 1999, Parties to the CBD had agreed actions on forests, alien invasive species, and coastal marine systems.

EX SITU CONSERVATION

Since *ex situ* conservation takes place outside natural habitats, for example, in zoos, botanical gardens, and seed banks, it is meaningful primarily for genes and species as a complementary effort to *in situ* conservation. Under Article 9, parties are encouraged to implement appropriate *ex situ* measures in support of conserving ecosystems and natural populations of species, preferably implementing these measures in the country where the genes or species originate.

SUSTAINABLE USE

The CBD defines sustainable use as the use of components of biological diversity in a way and at a rate that does not lead to the long-term decline of biodiversity . It thereby moves beyond the traditional focus on the species being harvested to instead address the system affected by any harvest. Under the CBD, parties are also encouraged to promote the sustainable use of biological resources by integrating conservation and sustainable use into national decision-making. Adopting measures for using biological resources which avoid or minimize adverse impacts on biodiversity, supporting local populations to develop and implement remedial action in degraded areas, and encouraging cooperation between government and the private sector in developing methods for the sustainable use of biological resources.

INCENTIVES FOR CONSERVING BIODIVERSITY

Biodiversity often is lost by accident, as an unintended consequence of other activities, such as building a dam or a road, converting a forest into a plantation, or extending agriculture into habitats that previously had been used for other purposes. Such development activities are often heavily subsidized, while the incentives for conservation tend to be much weaker. Therefore, the Convention encourages the adoption of economically and socially sound measures which act as incentives, such as proper pricing of biological resources to re ect true cost and the use of tradable rights in their management. Such measures can be effective means for encouraging the conservation of biological diversity and the sustainable use of biological resources.

RESEARCH, TRAINING, PUBLIC EDUCATION AND AWARENESS

Recognizing the importance of generating new information on biodiversity, the parties agree to promote active programs for research, scientific training and technical education in fields related to biodiversity. The CBD urges countries to cooperate in applying new scientific information to conserving biodiversity and using biological resources sustainably, with particular attention given to developing countries. Parties recognize that biodiversity will only be conserved if the general public is aware of its value and supports actions being advocated under the convention to promote public understanding of biodiversity issues, and to cooperate internationally in developing appropriate educational and public awareness programs in this field.

ASSESSING IMPACTS

Environmental impact assessments and strategic environmental assessments can be useful tools in determining how development projects are likely to affect or be affected by biological resources and to identify the options for minimizing these effects. The parties agree to introduce appropriate procedures for assessing the environmental impact of projects, programs, and policies that could significantly affect biodiversity, to seek to avoid or minimize any negative impacts on biodiversity, and to facilitate appropriate public participation in these procedures. The CBD also calls for states to exchange information about activities, which are likely to damage biodiversity and to promote arrangements for emergency responses in such cases.

ACCESS TO GENETIC RESOURCES

Genetic materials of plant, animal, microbial or other origin often have economic value. Seeds of agricultural crops, particular animal breeds, and genes from bacteria form important components of biodiversity and provide the foundation for a number of commercial activities, including the development of improved agricultural varieties and pharmaceuticals. Recognizing the sovereignty of states over their natural resources, the CBD encourages parties to facilitate access to genetic resources within their jurisdiction for environmentally sound use, under mutually agreed terms and with prior informed consent. These latter terms are difficult to define and have fostered new approaches to national legislation that ensure fair access and equitable sharing of benefits.

TRANSFERRING TECHNOLOGY

Fulfilling many of the convention's objectives will require that developing countries have access to appropriate technology. Parties are encouraged to provide or facilitate access to these technologies, including genetic engineering techniques, conservation technologies, and methods of habitat rehabilitation. Developing countries are to be extended fair and most favorable financial terms in this regard. Access to and transfer of technology is to be consistent with the protection of *intellectual property rights* (IPRs): no simple matter. Benefits from genetic resource based technologies are to be shared with the parties providing the genetic resources. The convention also includes elements relating to safe trade, handling and use of living modified organisms (LMOs). The first protocol of the CBD was negotiated to address these latter issues. The so-called Cartagena Protocol, named after the Colombian city where detailed negotiations began, was adopted as a formal Biosafety Protocol during a special session of the Conference of Parties (COP) held in Montreal, Canada, in January 2000 and has far reaching implications for trade in biotechnology. This protocol is the first international treaty to comprehensively address genetically modified organisms. The protocol will enter into force when 50 countries have formally ratified it.

Transfer of technologies relevant for the conservation and sustainable use of biological diversity, continues to be a major area of unfulfilled expectations of most developing country parties. Technology is an important tool for *in situ* and *ex situ* management of biodiversity. For example, new agricultural technologies have the potential to increase yields to feed the growing human populations while reducing unintended adverse environmental impacts. Modern biotechnology could provide tools for understanding the living world and thus may greatly aid in assessing and monitoring biodiversity at the national, regional and global levels. Transfer of relevant technologies will be facilitated and enhanced through the involvement of the private sector in the area of bioprospecting and biosafety in joint ventures with national institutions or local private sector entities.

EXCHANGING INFORMATION

The CBD encourages its parties to exchange information about biodiversity. The CBD also recognizes that many indigenous and local communities have a close and traditional dependence on biological resources and that their knowledge and practices can make a significant contribution to the conservation of biological diversity and the sustainable use of biological resources. Parties are therefore encouraged to protect and recognize customary use of biological resources in accordance with sustainable traditional practices. The CBD also encourages the maintenance and wider application of relevant indigenous knowledge, innovations, and practices as well as the equitable sharing of benefits arising from their use. They have established an internet-based clearing house mechanism for helping to do so.

PROVIDING FINANCIAL RESOURCES

Each Party to the CBD undertakes to provide within its capabilities the necessary funds and incentives for national activities undertaken to implement the Convention. It also recognizes that developing countries may require additional financial assistance to enable them to implement the convention and benefit from its provisions. Under the convention, the developed countries are called upon to provide new and additional financial resources to developing countries. One response has been the use of the financial resources of the Global Environment Facility (GEF), which was established and is operated by the World Bank, the

United Nations Development Programme, and the UNEP to provide the incremental costs necessary to assist developing countries in dealing with global environmental problems: the GEF, *inter alia*, the financial mechanism for implementing the CBD. The GEF spends over US$100 million per year in this endeavor, while many developed countries also provide considerable funding directly to developing countries amounting in total to several hundred million dollars per year. The CBD is therefore one of the relatively rare international conventions that also directly results in substantial economic incentives in support of its implementation.

THE WORKINGS OF THE CONVENTION

The CBD established a COP (*see* Conference of Parties, Volume 4) that meets on a regular basis. By mid-2000, it had met five times, as follows:

1. COP1: Nassau, Bahamas (November–December 1994);
2. COP2: Jakarta, Indonesia (November 1995);
3. COP3: Buenos Aires, Argentina (November 1996);
4. COP4: Bratislava, Slovakia (May 1998);
5. COP5: Nairobi, Kenya (May 2000).

The COP is advised by its Subsidiary Body on Scientific, Technical and Technology Advice (SBSTTA) which meets between sessions of the COP, normally in Montreal. The COP establishes *ad hoc* working groups to address particular issues. Noteworthy in this regard has been its work to implement Article 8(j), which addresses the issue of indigenous peoples and traditional knowledge. The CBD offers the only effective international mechanism where these issues can be discussed at a substantive and technical level; this helps integrate emotive human rights issues within the environmental issues affecting biodiversity.

As an umbrella agreement, the CBD is able to address a wide range of important issues. During its third meeting, the COP reasserted its aspiring role as a focal point in relation to other relevant international instruments and processes, especially on issues such as forests, intellectual property rights (IPRs), and agriculture. The degree of these interactions and the leadership level adopted by the COP are variable. Perhaps the issue where interaction is strongest is agriculture: the CBD addresses land resources; water resources; plant, animal and microbial genetic resources; wildlife; air and climate; farm inputs; wild sources of food; traditional knowledge; marketing conditions for agricultural products; land-use pressures; and agroforestry. The COP has staked out a leading position in the field of agricultural biodiversity, taking the opportunity to link concerns regarding biodiversity conservation and sharing of benefits with the mainstream economy, drawing on balanced attention to the three objectives of the CBD. The COP sent a message to the December 1996 session of the Food and Agriculture Organisation Commission on Plant Genetic Resources for Food and Agriculture (PGRFA) to negotiate the revision of the Convention on the International Undertaking on Plant Genetic Resources in harmony with the CBD. It also specified COP's willingness to consider a protocol on PGRFA under the CBD.

Consistent with its mandate under the CBD, the COP has also sought to in uence intergovernmental discussions on forests to ensure that forest biodiversity concerns are addressed covering issues such as the underlying causes of biodiversity loss, analysis and mitigation of human impacts on forest biodiversity, economic valuation of biodiversity components, and development and use of criteria and indicators for biodiversity. The work program for forest biodiversity includes development of technologies necessary for conservation and sustainable use of forest biodiversity, it takes an ecosystem approach, and incorporates traditional systems of conserving biodiversity in forests. The COP has also called for scientific analysis of the ways in which human activities, especially forest management practices, in uence biodiversity and sought ways to minimize or mitigate negative in uences arising from such practices.

Through its COP, the CBD has also stimulated considerable progress on biodiversity in the seas. For example, COP2 adopted the Jakarta Mandate on Marine and Coastal Biodiversity. It provided a checklist covering five areas in which CBD Parties should take actions for conservation and sustainable use in these habitats: integrated marine and coastal area management; marine and coastal protected areas; sustainable use of marine and coastal living resources; mariculture; and alien species. Major international bodies were invited to improve their existing activities and develop new actions, which promote the conservation and sustainable use of marine and coastal biodiversity, taking into account the recommendations contained in the Jakarta Mandate. A three-year process will elaborate the recommendations for action in the above five areas, and possibly others. To assist in the implementation of the process, the CBD Executive Secretary has established a roster of experts, and he draws upon a wide range of inputs from governments, intergovernmental organizations (IGOs), nongovernmental organizations (NGOs) and others.

Many people believe that conserving biodiversity and using biological resources sustainably will benefit all of society. However, conserving biodiversity, using biological resources sustainably, and equitably sharing the benefits of such use, all involve social costs and benefits that are borne unevenly across segments of society. Therefore, the CBD calls for policies to promote the equitable sharing of these costs and benefits in ways that improve the wellbeing of the poor and weaker sections of society. Activities to implement the Convention thus need to be supported by

all sectors of society, including local communities, NGOs, the scientific community and industry. The CBD offers a unique opportunity for this diverse mixture of interests to work toward the same broad objectives. This desired international consensus can be attained if all sectors of society contribute to the agreed goals of the CBD.

The CBD website is found at http://www.biodiv.org/.

Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES)

see CITES (Convention on International Trade in Endangered Species of Wild Fauna and Flora) (**Volume 4**)

Convention on Long-range Transboundary Air Pollution (LRTAP)

Lars Björkbom
Stockholm, Sweden

The LRTAP was signed in Geneva in November 1979 by 34 member states of the United Nations Economic Commission for Europe (UNECE), which also includes Canada and the US. The Convention came into force in 1983 and today (early 2000) has 45 parties out of 55 of the UNECE members. The growth of number of parties over the period is mainly a re ection of the geo-political changes that have taken place, but also re ects the growing awareness of international interdependence when trying to achieve national environmental goals in minimizing negative effects to ecosystems and human health of various airborne pollutants, which are transported long-range over national borders.

The LRTAP is a framework convention focusing on information gathering and exchange of knowledge to support parties in their air pollution abatement policies. Its foremost scienti c tool from the start was the incorporation of the European Monitoring and Evaluation Programme (EMEP), which had been built up as part of the preparatory process. Later, other expert groups have been established on effects of air pollutants and information gathering and assessment of abatement technologies.

*The LRTAP countries would, at the time, not allow any legal provisions for extending international cooperation to mutual abatement efforts in the form of protocols. There is still no such provision in the framework agreement but today there are de facto eight protocols, ve of which are in force. One of the protocols (1984) regulates the nancing of the EMEP. The others deal with reductions of separate pollutants (sulfur dioxide 1985 and 1994, nitrogen oxides 1988 and volatile organic compounds (VOCs) (*see *Volatile Organic Compounds, Biogenic (VOCs), Volume 2) 1991). The later protocols regulate groups of compounds with similar characteristics and effects, e.g., protocols on heavy metals (HMs) and on persistent organic pollutants (POPs) (1998). The latest protocol to abate acidi cation, eutrophication and ground-level ozone, signed in 1999, further regulates reductions of emissions of sulfur, oxidized and reduced nitrogen and VOCs, which all interact in different ways to create environmental and human health problems in the three effect areas listed in the name of the protocol. None of the protocols has all LRTAP Parties as parties or signatories. Also, the cluster of countries is not the same for each country's protocol, which re ects the fact that different countries have different priorities on different ecological and health issues.*

The LRTAP had its roots in the East–West tension during the Cold War, the Principle 21 adopted at the 1972 United Nations (UN) Conference on Man and the Environment in Stockholm, and Scandinavian concerns about low resistance of waters and soils (low buffering capacities) to acidifying pollutants, whose main sources were sulfur emissions in continental Europe and the British Isles. In the Final Act of the Conference on Security and Cooperation adopted in Helsinki in 1975 during a temporary thaw in East–West relations, UNECE countries agreed to cooperate in a number of areas: disarmament, human rights and economic affairs, which included environmental affairs. When the thaw again disappeared in the late 1970s, the only area that was not considered too dangerous for cooperation was environment. A proposal from the Soviet Union in late 1976 to the UNECE to cooperate in the field of the environment (and also in the field of energy) was, however, acceptable to the western countries. When discussing what particular problems should be dealt with, Sweden and Norway saw their chance to bring acidification on to the agenda as being a transnational problem and also a test of the implementation of Principle 21, which states that although nations have sovereign rights over their own environment, they are not allowed to pollute their neighbors or the environmental commons. After more than a year of tough negotiations, a high-level meeting of UNECE members in November 1979 adopted the LRTAP.

Within the convention, parties agree to cooperate, in exchange for information relating to national emissions of air pollutants that might have transboundary effects, and measures that they are taking to abate such emissions. The objective laid down in article 2 is to endeavor to limit and, as far as possible, to gradually reduce and prevent air pollution including long-range transboundary air pollution. Parties shall develop ••• policies and strategies which, shall serve as a means of combating the discharge of air pollutants (art 3) and exchange information on, and review, their policies, scientific activities and technical measures aimed at combating ••• the discharge of air pollutants (art 4) and exchange available information on national, subregional and regional policies and strategies for the control of sulfur compounds and other major air pollutants (art 8).

The LRTAP is served by the UNECE secretariat, but the UNECE has no decision-making power. This rests with the executive body (EB), which is the Conference of the Parties to the LRTAP (*see* Conference of Parties, Volume 4).

The initial central scientific tool related to the convention is the Cooperative Programme for Monitoring and Evaluation of the Long-range Transmission of Air Pollutants in Europe (EMEP), which was in existence already in 1977, but became an integral part of the LRTAP when it came into force. A number of subsidiary bodies to the EB have successively been established, relating to various articles under the LRTAP or to separate Protocols or EB decisions, relating to research on effects, assessment of emission abatement techniques, negotiation and review of protocols and implementation of protocols. All these subsidiary bodies have, through EB decisions, over time and as needs have arisen, established (and, less frequently, concluded) international coordinating programs, centers and Task Forces (TFs) to support the quality of their advisory role to the EB. (For the 1999 organization of work and how it has been developing over time, see Figure 1).

During the negotiations of the LRTAP, a few countries (in particular Norway and Sweden) pressed for a more action-oriented agreement, which would include specific international emission-reduction obligations for parties. The time was, however, not ripe in 1979 for such proposals. It was not even possible to get a provision in the LRTAP that could form a legal base for developing protocols to reduce emissions of air pollutants. It is interesting to re ect on the fact that today (early 2000), four such Protocols are in force and a further three adopted and signed by a large majority of LRTAP parties that regulate quantitatively defined reductions in emissions of the most important air pollutants, which are transported over national boundaries within (and in certain cases beyond) the UNECE region. There is also a Protocol on the financing of EMEP.

From a legal point of view, each Protocol is an independent international agreement. None of the Protocols (already in force or expected to come into force) has all LRTAP parties as parties (or signatories) and the composition of parties (or signatories) to each Protocol differs for different protocols, which re ects different priorities of parties to problems regulated in each Protocol. There has been, however, a growing tendency over time that more and more LRTAP parties also become parties to the Protocols. This might re ect a growing awareness among all LRTAP parties that most of the problems related to transboundary air pollution can not be effectively attacked without common international abatement responsibilities, but also the fact that later Protocols have allowed for differentiations of national abatement obligations, when they are based on scientific assessments of each party's contribution to the problems.

The first three substantive Protocols on emission reductions of sulfur (1985), nitrogen oxides (1988) and VOCs (1991), respectively, are often called protocols of the first generation. Their common characteristic is a at rate approach, meaning that each party shall implement the same percentage of emission reductions of the polluting compound over the same given period of time.

In the 1985 Protocol on the Reduction of Sulfur Emissions or their Transboundary Fluxes, all contracting parties were obliged to reduce their emissions of sulfur dioxide by at least 30%, by 1993, based on their emissions in 1980.

In the 1988 Protocol concerning the Control of Emissions of Nitrogen Oxides or their Transboundary Fluxes, the general rule is that parties should freeze their emissions by 1994 to the emission levels of 1987. There was, however, some exibility introduced in this Protocol. The US was allowed to have 1978 as the base year for calculating their emissions in 1994 because they could demonstrate, relative to other parties, early introduction of three-way catalysts to curb exhausts from cars, which were then and still are a dominant source of emissions of nitrogen oxides.

In the 1991 Protocol concerning the Control of Emissions of VOCs or their Transboundary Fluxes, the general rule was to reduce emissions of VOCs by 30% by the year 1999 based on assessed emission values from a choice of a year between 1984–1990. Some parties from countries in Central and Eastern Europe (CEE) with relatively small total emissions were allowed to freeze their emissions based on the same optional base year. Parties with areas larger than 2 million km^2 could choose a limited part of their territories as management areas. Canada and Norway chose this option when ratifying the Protocol. Norway was allowed to include its vast marine area allotted to them under the Law of the Seas agreement, where Norway has large offshore oil and gas production, which is an important source for VOC emissions.

All three Protocols were considered as a first step to control and reduce emissions of the relevant pollutants. In the 1988 NO_x Protocol, the review article prescribed that a revised or second step protocol should *inter alia* be based on the *critical load* approach (*see* Critical Load, Volume 3). This has also been the case for two of the second-generation protocols, which have dealt with the same pollutants (plus ammonia) as the earlier ones. The critical load is a quantified value of how much deposition of a pollutant a receptor can accommodate without losing its buffering capacity. A similar value for concentrations of pollutants in the atmosphere is applicable for effects of ground-level ozone to *inter alia* crops and human health, and is called the *critical level*. The critical load concept has so far, however, not been applicable to the two Protocols, which regulate the control of POPs and heavy metals (HMs), both adopted in 1998, where alternative methods of scientific risk assessments of the various effects of the various chemicals and metals on ecosystems and human health have been criteria for banning or severely restricting their production and/or use. Common to all protocols of the second generation is the fact that they are guided by

scientific considerations, where up-to-date knowledge of the negative effects of the controlled pollutants is the key characteristic. Another common feature of these protocols is that they allow differing emission reductions or other obligations for individual parties, the goal being to use optimal cost-effective solutions.

The 1994 Protocol on Further Reduction of Sulfur Emissions can today be considered as a first test to apply the critical load's approach. Based on information of emission sources and volumes, atmospheric trajectories and data on depositions (*see* RAINS (Regional Air Pollution Information and Simulation), Volume 4), critical load maps produced on different ecosystems' sensitivity to sulfur depositions in 150 • 150 km grid squares covering all of Europe west of the Urals, and cost curves for emission abatement measures in different countries, optimized integrated assessment scenarios were produced that could give guidance to negotiators on the most cost-effective way, on a regional scale, to reduce sulfur emissions. The long-term objective was to achieve depositions below critical loads for sulfur all over Europe, to abate acidification. It was realized that this goal could only be reached in a series of steps. The Protocol covered a first step, aiming at a 50% closure of the gap of exceedances of critical loads between 1990 depositions and critical loads in each grid square, translated into emission reduction quantities for each country by the year 2000 based on 1980 emission values. Although this ideal gap closure scenario was modified in the final negotiating phase (as expressed in Annex II to the Protocol) owing to political and economic concerns expressed by some delegations, the critical load or effect-related approach proved its value to all participating governments. An option was open to parties with areas of more than 2 million km^2, to meet their requirements within a defined sulfur dioxide management area that contained emissions that had significant impacts beyond their national boundaries. This option was used by Canada only.

The approach was, however, considered a half-measure. After all, acidification is not caused by sulfur deposition only, but also by nitrogen, both oxidized and reduced. Therefore, the EB considered that the objective should be to abate acidification and eutrophication, and given the fact that NO_x, together with VOCs, under exposure to sunlight produce ground-level ozone, this effect area should also be attacked in the new NO_x Protocol (see Figure 2). This multi-effect/multi-pollutant approach was embarked upon in 1994. This required much further research on deposition and concentration effects on ecosystems, materials and human health, as well as laborious refinement of the integrated assessment model used for the 1994 Sulfur Protocol. Much work to establish reasonably accurate emission inventories for the relevant

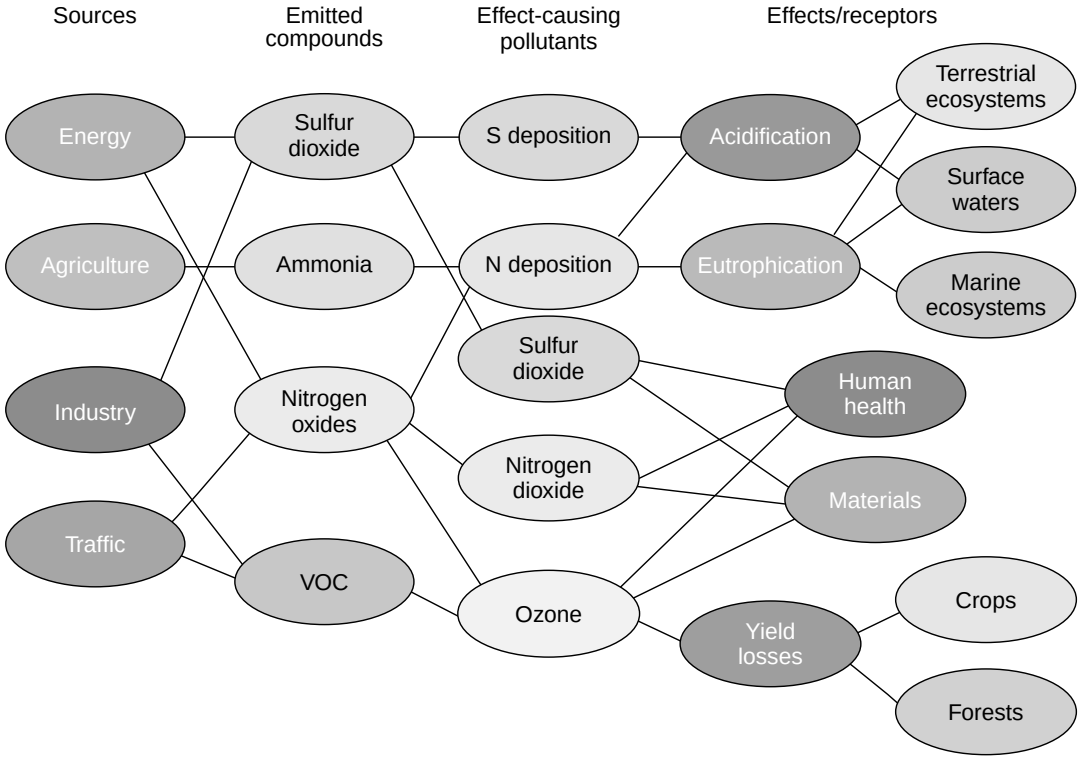


Figure 2 The interrelationship between sources of air pollution, emitted compounds, effect-causing pollutants and effects receptors. (Reproduced by permission of IVL)

pollutants and assessment of abatement costs for the many emission sources involved also had to be done. New insights into the atmospheric chemical processes leading to ozone formation, and the behavior of ammonia when released to the air and when deposited in ecosystems, were developed. With this impressive knowledge base, a Protocol to abate acidification, eutrophication and ground-level ozone was ready to be adopted in late 1999. Parties will be legally obliged, in a first step, to make substantial reductions of all pollutants involved (defined in absolute volumes in Annex II to the Protocol) in relation to their national contributions to the effects area by the year 2010, based on the emission levels of 1990. When implemented, exceedances above critical loads for acidification and eutrophication will be limited to the most sensitive areas in Europe, and the reduction of background ozone concentrations will significantly relieve pressure on human health, crops and semi-natural forests. However, in 2010 there will still be large areas with depositions above critical loads and several million people will live in areas where critical levels for oxidants will be exceeded. Further steps towards the final objective of the Protocol, to reduce deposition of acidifying and eutrophying pollutants below critical loads and concentrations below critical levels in the UNECE region, are to be negotiated within the framework of the Protocol after periodic reviews assessing the relevant scientific development and new techniques and other measures to further abate the pollutants. By using this integrated approach, it is calculated that the overall abatement costs will have been reduced by more than 30%, in comparison with the at-rate pollutant-by-pollutant approach used in the first generation of protocols under the LRTAP.

Two further Protocols on HMs and on POPs were developed in a parallel exercise and both were adopted in 1998. The HMs Protocol regulates emissions of cadmium, lead and mercury. The objective of the Protocol is to minimize effects on human health and adverse effects on the environment. Its central obligation is to curb emissions by applying strict limits to the major source categories emitting metals that can be achieved by using *best available techniques, not entailing excessive costs* (BATNEEC), which are quantified and described in a number of annexes. Efforts to ban metals in products were less successful, owing to con icting international trade rules. Phasing-out of leaded gasoline over a period of years is, however, required and will have its main effects in CEE countries. The Protocol also introduces measures to lower emissions of mercury from products. The Protocol includes provisions for adding other metals to the Protocol if international action is deemed necessary.

The objective of the POPs Protocol is to control, reduce, or eliminate discharges, emissions and losses of POPs to the environment. Three measures are used as obligations. For some substances, production and use are eliminated, and for others, use is severely restricted. For those substances which are formed unintentionally as a result of combustion or processing activities, total annual emissions are reduced from a reference year, applying stiff limits using BATNEEC. The 16 substances listed in the present Protocol are aldrin, chlordane, chlordecon, dichlorodiphenyl, trichloroethane, dieldrin, dioxins and furans, endrin, heptachlor, hektachlorobenzene, hexachlorocyclohexane, hexabromobiphenyl, mirex, polycyclic aromatic hydrocarbons, polychlorinated byphenyls and toxaphene. A process included in the Protocol allows for substances to be added or current obligations to be modified as new information is obtained. This regional POPs Protocol is seen as a major step towards global control of these substances, which are presently under negotiation under the UN Environment Programme umbrella.

With the latter three Protocols, the most important known air pollutants, which are transported over national borders, are now controlled within the framework of the LRTAP. These Protocols are designed to stand the test of time, and have provisions for regular reviews and amendments. Both effective implementation and cost-optimal abatement measures need to be based on the most up-to-date scientific knowledge. Preserving the impressive scientific network that has been built up during the 20 years of development under the LRTAP is a key to the future effectiveness of the Convention. It has been an inspirational and legal focal point for both international and national air pollution abatement in the UNECE region. The implementation record has so far been impressive. Substantive reductions of sulfur emissions have been recorded, but also reductions in absolute terms of nitrogen oxides and VOCs emissions have taken place over the period and with the new Protocols in force, there will be forceful management instruments to achieve the objective laid down in article 2 of the LRTAP to gradually reduce and prevent transboundary air pollution in the UNECE region. Of course, this does not mean that air pollution in the region has reached sustainable levels, but the work under the LRTAP indicates that this goal can be reached through revisions of the Protocols foreseen over the next few decades. Its effectiveness so far should be an inspiration to other regions in the world.

See also: LRTAP (Convention on Long Range Transboundary Air Pollution), Volume 4.

COP (Conference of Parties)

see Conference of Parties (**Volume 4**)

Cradle to Grave Analysis

see Cleaner Production (**Volume 4**)

Critical Load

see Quotas in International Environmental Negotiations (**Volume 4**)

CSD (Commission on Sustainable Development)

The Economic and Social Council of the United Nations (UN) established the CSD as a functional commission of the Council on the 12th February 1993, in response to a resolution of the UN General Assembly. The resolution of the General Assembly on 29th January 1993 endorsed the recommendations of the UN Conference on Environment and Development (UNCED) in particular, the chapter of Agenda 21 concerning Institutional Arrangements which, *inter alia*, recommended the establishment of the Commission.

The Commission is composed of 53 members elected from among state members of the UN and members of the specialized agencies for a term of office of three years.

The Commission meets annually for a period of 3 weeks, now reduced to two.

The Commission was established to ensure effective follow-up to the UNCED Conference, as well as to enhance international cooperation, and rationalize the intergovernmental decision-making capacity for the integration of environment and development issues, and to keep under review the implementation of Agenda 21, recognizing that it is a dynamic program that could evolve over time, and make recommendations, as appropriate, on the need for new cooperative arrangements related to sustainable development.

The Assembly also decided that the Commission, in the fulfillment of its functions, will also monitor progress in promoting, facilitating and financing, as appropriate, access to and transfer of environmentally sound technologies and corresponding know-how, in particular to developing countries, on favorable terms, including concessional and preferential terms, as mutually agreed, taking into account the need to protect intellectual property rights as well as the special needs of developing countries.

The Commission has already met seven times. Part of each meeting (lasting for two to three days) is attended by high-level ministers. The Commission is endeavoring to coordinate activities of the UN system covering areas of sustainable development, which include economic growth, social development, and environmental protection and management.

MOSTAFA K TOLBA *Egypt*

Cultural Eutrophication/Cultural Oligotrophication

see Policies for Sustainable and Responsible Fisheries (**Volume 4**)

Cumulative Environmental Assessment

R E Munn
University of Toronto, Toronto, Canada

A cumulative environmental assessment (CEA) is an assessment of the cumulative environmental effects over decades of a proposed development (e.g., a power station or waste dump), including:

- *the cumulative environmental effects of the development itself (e.g., slow release of heavy metals into ponds or the atmosphere);*
- *the multiplying effects on the local environment of other developments that may subsequently be attracted to the area (e.g., construction of an electricity-generating station may attract industry to the area);*
- *the interactive effects on the local environment of long-term regional and global environmental and socioeconomic changes (e.g., climate change, globalization, free trade agreements). (In many jurisdictions, however, CEAs are not required to include the effects of such large-scale changes.)*

The causes of local change may be environmental (e.g., drainage of wetlands), socioeconomic (e.g., population growth, highway construction) or both. Often, the changes are piecemeal over decades (sometimes referred to as *nibbling*), and thus difficult to predict. Overlapping political

responsibilities may sometimes contribute to the problem (CEARC/NRC, 1985).

Some engineers and planners think that the preparation of a CEA is an impossible task. Nevertheless, more and more jurisdictions require that developers submit CEAs in addition to environmental impact assessments (EIAs), (called environmental assessments, EAs in some jurisdictions). In Canada, for example, Bill C-13 (the 1992 Canadian Environmental Assessment Act) includes the clause that consideration be given to any cumulative effects that are likely to result from the project in combination with other projects or activities that have been or will be carried out. This Act does not include assessment of the effects of long-term large-scale change, although a hydrological development, for example, is usually designed to operate successfully for 100 years, during which time the rainfall regime might change drastically.

Factors to be considered in the selection of methodologies to be used in CEAs include the following:

- The assessor must be concerned with effects that may not become apparent for several decades, and in some cases, more than a century into the future.
- The spatial scales of interest must be much larger than in the case of an EIA. For example, a CEA of a proposed recreation area to be located in a rural setting must include not only the area but also the transportation corridors to and from urban centers.
- Because of the complexities of socioeconomic and ecological systems, an ecosystem approach (*see* Monitoring in Support of Policy: an Adaptive Ecosystem Approach, Volume 4) must be used. As examples, CEA study boundaries, wherever possible should coincide with those of naturally occurring ecological zones such as watersheds. Post-construction monitoring systems should be directed towards detecting changes in the stresses operating on the system as well as on the environmental effects.
- The CEA should include consideration of relevant long-term regional and global change factors such as urbanization and climate change.
- As interest in cumulative effects extends further and further into the future, the uncertainty factor increases more and more, due to:

1. The inherent uncertainties in socioeconomic and political scenarios, for even 10 years ahead, e.g., the unexpected rise in oil prices in the 1970s, and the fall of the Berlin wall in the late 1980s.
2. The failure of simulation models to predict many kinds of cumulative effects due to non-linearities, feedbacks, and incomplete knowledge of initial and boundary conditions.
3. The lack of methodologies to test model performance. Historical analogues are usually difficult if not impossible to find, and even if an environmental audit is performed 10 or 20 years after a development has been launched, there is no way of telling whether a better site could have been selected for the development.

With these concerns in mind, several CEA methodologies are being used, but it is still too early to recommend any single one of them.

1. *Risk assessment methodologies*: the risk assessment field contains a rich panoply of methodologies, some of which might be useful for assessing long-term cumulative effects.
2. *Multiple scenario construction*: because the future is so uncertain, it may be useful to select a range of future scenarios, to study the environmental effects of each, and to bracket the possible range of environmental effects that might ensue. See, for example, Stigliani *et al.* (1989) and Policy Exercises: a Tool for Solving Environmental Policy Dilemmas, Volume 4.
3. *Adaptive management strategies*: planners are well aware of the fact that long-term predictions of future land use are often wrong. They therefore periodically update their plans as new information comes in. Holling (1978) has recommended that this approach be used in EIAs. Of course, once a reservoir, for example, has been built, it is very difficult to make changes to its design. Wherever possible, however, options should be kept open as long as possible in preparation for the unexpected.

See also: EA (Environmental Assessment), Volume 4; EIS (Environmental Impact Statement), Volume 4; Environmental Assessment of Major Development Projects, Volume 4; Integrated Assessment, Volume 4; Intergovernmental Panel on Climate Change (IPCC): an Historical Review, Volume 4; Social Assessment, Volume 4; Social Impact Assessment, Volume 4.

REFERENCES

CEARC/NRC (1985) Cumulative Environmental Effects: a Binational Perspective, CEARC/FEARO, Env. Canada, Ottawa, 1–175.

Holling, C S, ed (1978) *Adaptive Environmental Assessment and Management*, John Wiley & Sons, Chichester.

Stigliani, W M, Brower, F M, Munn, R E, Shaw, R W, and Antonofsky, M (1989) Future environments for Europe: some implications of alternative development paths, *Sci. Total Environ.*, 80, 1–102.

FURTHER READING

Buckley, R (1998) Cumulative Environmental Impacts, in *Environmental Methods Review: Retooling Impact Assessment for*

the New Century, eds A L Porter and J J Fittipaldi, The Press Club, Fargo, ND, 95–106.

Preston, E M and Bedford, B L (1988) Evaluating Cumulative Effects on Wetland Functions: a Conceptual Overview and Generic Framework, *Environ. Manage.*, 12, 565–583.

Spaling, H and Smit, B (1993) Cumulative Environmental Change: Conceptual Frameworks, *Environ. Manage.*, 17, 587–600.

Treewek, J and Hankard, P (1998) Ecological Impact Assessment, in *Environmental Methods Review: Retooling Impact Assessment for the New Century*, eds A L Porter and J J Fittipaldi, The Press Club, Fargo, ND, 263–272.

D

Demand Management

A P Lino Grima
University of Toronto, Toronto, Canada

Careless technology and lack of adequate policy instruments have made it possible for producers and consumers of environmental and natural resources to avoid part of the costs of production and consumption by passing these costs to others. These costs (e.g., water and air pollution) are referred to as technological externalities in environmental economics (Baumol and Oates, 1979). One way of reducing these externalities is to invoke regulation and prohibition. Indeed in the case of highly noxious pollutants, outright prohibition is called for. However, in most other cases there are several economic incentives and legal mechanisms for reducing these externalities (e.g., regulation, transferable emission permits, liability for compensation as in the US Superfund provisions). How can the demand for goods and services that deplete scarce resources or pollute our environment be reduced while respecting consumer choice (and, by implication, an economically ef cient allocation of capital and labour)? Demand management (sometimes referred to as demand-side management, or demand design) is one such exible policy tool to enhance conservation and move toward sustainable development within a market system based on consumer sovereignty. Demand management refers to policies designed to in uence consumer behavior by setting the price of a good or service (e.g., municipal water) to the level of producing and delivering it. In more recent literature the term has been used to include any technical, institutional, or educational effort to encourage resource conservation.

In general, the emphasis in environmental and natural resource management has been on the supply fix, i.e., increasing the supply (for example, of municipal water or electricity or highways or paper products) by expanding the infrastructure to meet the requirements for water, electricity, and so on. This emphasis on the supply fix re ects the view that supplies of environmental and natural resources are plentiful and that the demand for goods and services based on such resources are fixed. Indeed demand is often described as need or requirements. In this mind-set the supply fix becomes the only fix; and prohibition or regulation through improved technology becomes the only forms of environmental management. This mind-set is slowly changing to the view that higher levels of material production and consumption are compatible with sustainable resource use (WCED, 1987).

In an early empirical study of *demand management* through marginal cost pricing (Grima, 1972, 5), the first mention of the terms demand engineering or demand management was traced to a 1960 report by the US Senate Select Committee on National Water Resources, 1960. Demand management aims to in uence consumer choices by making the consumer pay a price that is as close as possible to the incremental cost of production, including any costs borne by others (e.g., health costs caused by air pollution). Like other environmental management options, demand management has its strength and weaknesses and the choice among the wide range of options should not be made on purely ideological grounds.

One limitation should be emphasized right at the outset: demand management and other economic incentives and legal mechanisms for reducing externalities complement rather than replace a regulatory and monitoring framework. To state an obvious example, residential electricity demand management requires metering, an appropriate pricing schedule and regular billing. So in a third-world city that lacks metering and billing facilities, residential demands for water or electricity are not amenable to demand management. This example leads to another limitation of price-dependent demand management: it is more relevant to reducing pressure on resources in af uent countries than in less developed countries.

Demand management works best for resource-based services that exhibit rising incremental costs (i.e., economies of scale have been exhausted or the cost of resources is rising) or services that exhibit daily or seasonal peak demands (e.g., for municipal water, energy for heating or cooling and highway use). Meeting peak demands requires higher levels of investment (e.g., for water storage towers) and at

least some higher operating costs (e.g., energy for pumping the water). It is therefore reasonable that the users of services during peak periods should pay for these increased costs. In addition, consumer choice is allowed and the policing/monitoring of peak pricing is not likely to be expensive. For example, it should be administratively easy to impose higher prices for water use during the summer or to require higher prices for transportation users during the rush hours. These higher costs to the consumers would provide a continuing incentive to reduce peak use and therefore increase the efficiency of the water (or public transport) system.

One of the positive aspects of demand management is that it enhances conservation by giving consumers correct signals and helping them to avoid policy traps. For example, in Canada in the late 1990s, more than half of the new vehicles on the road were relatively speaking gas-guzzling mini-vans and sport utility vehicles. These new car owners were faced with steeply rising gasoline prices in late 1999 and early 2000. However, adjusting to these higher energy costs in the short term is difficult or costly or both because of the long lifetime of vehicles. If an appropriate gasoline tax had been imposed to meet the costs that energy users impose on society as a whole in the form of NO_x and CO_2 new car owners would have received a more appropriate signal. Another example where a smaller pay-off may be expected is the deposit for the returnable beer bottle in Ontario (as opposed to the recyclable glass for wine and the aluminum can for pop). The average beer bottle in Ontario is used 20–30 times with small but not negligible savings in energy, litter and cost. It is particularly important that consumers in developing economies are given appropriate signals in order to avoid costly policy traps. In this sense, demand management is quite relevant to less af uent countries.

Demand management requires that the consumer of energy, water and other resource-based goods and services face costs (i.e., prices) at the correct level. If electricity at peak seasons and time costs more to produce and deliver, then electricity rates should re ect this. If energy use imposes costs on society in the form of climate change, then prices should re ect this too. Consumers would still have the choice to pay the higher price or reduce consumption, presumably through the adoption of more energy-efficient technology.

In most cases, prices are set by a number of producers who compete for market share and a number of consumers who seek to find the lowest price possible. So prices are set at the intersection of the demand and supply curves without too much administrative interference or collusion. Any deviation from this model would be a deviation from perfect competition and also a deviation from the ideal (i.e., efficient or optimal) allocation of the factors of production. However, most water and electricity rates and mass transportation prices are prices set by monopolies that provide municipal water, electricity, mass transport, etc. So how does one compare the prices with respect to the optimal allocation of resources that the market system aims to achieve?

In a market economy, incremental (or marginal) utility that consumers derive from a good or service is equal to the price per unit of a good or service. On the production side the volume of the good or service offered for sale is such that the incremental cost of producing the last (or marginal) unit is equal to the revenue, i.e., the price charged to the consumer. In brief, both the level of consumption and production are adjusted in such a way that the cost of the last unit produced and the utility of the last unit consumed are both equal to the price. Under these conditions, producers would not want to produce more or less of the good or service and the consumers would not want to consume more or less of it. In economic terms, we have an efficient allocation of scarce labor, human-made capital and natural resources. One should put in the usual disclaimer that it is necessary to make assumptions of perfect competition and the absence of externalities for this argument to be tenable. In addition, one has to assume a fair distribution of wealth and income before one accepts economic efficiency to be also ethically acceptable. However, the absence of perfect competition or a fair distribution of income and assets, is not a good reason to throw out the pricing system, which in general works much better than any central planning. The lack of perfect competition and social equity should, however, act as a spur to improve the market system. Demand management meets this criterion because ideally demand management sets consumer price equal to the incremental price of production, taking into account any external costs imposed by consumers on society and any costs of meeting peak demands.

Much of the early work on demand-side management focused on correct pricing and this type of mechanism is still a very useful step. However, later work by Brooks and others (Brooks, 1994; Grima and Paine, 1985) have expanded the concept especially with respect to its application to less developed countries. For example, marginal cost pricing (or even sensible pricing) of irrigation waters would free water for higher value uses such as municipal water. However, in economies where metering is non-existent and where it is politically and socially sensible to keep farmers on the land, emphasis on the conventional equimarginal principle is probably misplaced. Instead the focus needs to be directed towards technical, institutional and educational efforts to conserve water in the current uses.

To summarize, regulation and prohibition are no longer considered to be the only tools for achieving environmental goals. Market-oriented policy options such as demand management (coupled with appropriate monitoring and regulatory regimes) could play a key role in the transition from a consumer society to ecologically sustainable economic

development. The more rational prices that would face consumers could reduce the rate of depletion of non-renewable natural resources and could also reduce the rate of environmental degradation. Recent work on demand-side management has tended to include a broader range of conservation strategies in addition to the original emphasis on marginal-cost and peak-responsibility pricing.

REFERENCES

Baumol, W J and Oates, W E (1979) *Economic, Environmental Policy and the Quality of Life*, Prentice Hall, Englewood Cliffs, NJ.

Brooks, D B (1994) Economic, Ecology and Equity: Lessons from the Energy Crisis in Managing Water Shared by the Israelis and Palestinians, in *Water and Peace in the Middle East*, eds J Isaac and H Shuval, Elsevier, Amsterdam, 441–450.

Grima, A P (1972) *Residential Water Demand: Alternative Choices for Management*, University of Toronto Press, Toronto.

Grima, A P and Paine, R L (1985) Urban Water Conservation, *GeoJournal*, 11(3), 257–263.

US Senate Select Committee on National Water Resources (1960) *Water Resources in the United States*, 86th Congress, 2nd Session, Committee Print No. 31. Washington, DC, GPO, March.

WCED (1987) *Our Common Future*, World Commission on Environment and Development, Chairperson G H Bruntland, Oxford University Press, Oxford.

Dematerialization and Sustainable Development

see Dematerialization and Sustainable Development (*Opening essay*, **Volume 4**)

Dematerialization of the Economy

Dematerialization is a shorthand term for significantly reducing the materials intensivity of the economic system. The need for this arises from a series of related propositions that are increasingly salient. Firstly, the world economy, as currently structured, is extremely materials intensive, in that the rich countries extract, process and discard far greater tonnages of materials than do the poor and developing countries. Secondly, economic growth in the past has also been based on mechanisms (such as the substitution of machines powered by fossil fuels for human and animal muscles) that are inherently materials intensive. Moreover, this mechanism, together with economies of scale, is still important for increasing labor productivity. Thirdly, some materials, like metals, are recyclable but others (especially fossil fuels) are not. Most materials extracted from the environment automatically become wastes and pollutants in a relatively short time; this is a simple consequence of the law of conservation of mass (Ayres and Kneese, 1969; Ayres, 1989). Fourthly, the waste assimilative capacity of the earth is becoming overloaded. In many respects, it is the scarcest resource. Finally, the developing countries must exploit many of the same growth mechanisms that the industrial countries did in reaching their present state of af uence. It follows that, to allow scope for economic growth in the developing world without triggering irreversible environmental degradation, it will be necessary for the advanced industrial countries to adopt fundamentally different policies and approaches. In brief, these changes require far more strict conservation of materials (including redesign to facilitate repair, renovation and remanufacturing), and development and exploitation of renewable sources of energy (especially solar power). Apart from reducing their own demand for raw materials and fossil fuels, this change of direction is important because the new materials-conserving and non-polluting renewable energy technologies that must be developed by the industrialized countries are also the technologies that will also be adopted in the developing countries as they grow.

See also: Dematerialization and Sustainable Development, Volume 4.

REFERENCES

Ayres, R U (1989) Industrial Metabolism, in *Technology and Environment*, eds J Ausubel and H E Sladovich, National Academy Press, Washington, DC.

Ayres, R U and Kneese, A V (1969) Production, Consumption and Externalities, *Am. Econ. Rev.*, June.

ROBERT AYRES *France*

Desertification Convention

Martin Williams
Adelaide University, Adelaide, Australia

At the United Nations (UN) Conference on Environment and Development (Earth Summit) held at Rio de Janeiro in June

1992, deserti cation was de ned as: " land degradation in arid, semi-arid and dry sub-humid areas resulting from various factors, including climatic variations and human activities" (UNCED, 1992). Until then, the de nition used by the United Nations Environment Programme (UNEP) attributed all deserti cation to human activities and excluded climate as a causal factor. Under this scenario, deserti cation was regarded simply as " land degradation in arid, semi-arid and dry sub-humid areas resulting from adverse human impact" (UNEP, 1990). Another important initiative to emerge from the 1992 Earth Summit was a request to the UN General Assembly to establish an intergovernmental committee for negotiating a convention on deserti cation. The General Assembly (47th session, 1992) established such a committee, the negotiations were successful, and the Deserti cation Convention was duly signed by 87 countries in Paris on 14- 15th, October, 1994. The full title of the convention is as follows: United Nations Convention to Combat Deserti cation in those Countries Experiencing Serious Drought and/or Deserti cation, particularly in Africa. To appreciate the need for such a convention, it is useful to understand some of the reasons for the inclusion of drought in the title and for the particular emphasis on Africa.

PRELUDE TO THE DESERTIFICATION CONVENTION

As long as desertification was considered to be a result of adverse human actions in arid and semi-arid areas, there was little incentive to consider an international approach to the problem. The prevailing view until the early 1970s was that desertification re ected inappropriate forms of land use (Aubréville, 1949), and that its impact was highly localized (Mainguet, 1994; Thomas and Middleton, 1994). The responsibility for managing desertification must therefore reside within the countries affected, so that desertification was seen to be a matter for national, but not for international, concern. All this changed with the 1968–1974 Sahel drought. This drought, which began in the Sahel region of Africa in 1968, brought below-average rainfall to every country immediately south of the Sahara for the ensuing fifteen years (Hare and Ogallo, 1993). More than any other single factor, the Sahel drought was the catalyst in focussing world attention on the reality of severe and widespread drought as a cause of famine and extreme social distress (Kassas, 1975a). Droughts had af icted other parts of the world besides Africa, but nowhere else in the world at that time had the scale and severity of the drought been so great (Tucker *et al.*, 1991; Middleton and Thomas, 1992, 1997; Williams and Balling, 1996; Williams, 2000).

The UN responded very belatedly to the famine in Africa and held a conference on desertification in Nairobi in 1977. This conference produced a plan of action to combat desertification that was endorsed by the UN General Assembly later that year and entrusted to the UNEP for implementation. In the years that followed, it soon became apparent that the plan of action was not achieving its aims and that desertification was becoming an increasingly severe problem, particularly in Africa (Darkoh, 1989; Dregne *et al.*, 1991; Glantz, 1977, 1987; Grainger, 1990). Reasons for the failure include a lack of political commitment by individual countries, inadequate infrastructure and financial support, and failure to involve local communities.

IMPEDIMENTS TO SUCCESS IN THE COMBAT AGAINST DESERTIFICATION

Table 1 lists a number of the more important reasons for past and present failures in dealing with desertification. Some relate to a lack of adequate monitoring procedures, some to inappropriate forms of participation, some to inadequate resources and poor use of existing resources, and some to awed planning.

Frustrated by the lack of substantive progress in dealing with desertification in Africa, the African environment ministers met in November 1991 in Abidjan in order to prepare a common position paper for Africa for presentation to the 1992 Earth Summit. In particular, they were unanimous in requesting that a new convention to combat desertification should be a specific recommendation to be included in Agenda 21. They were successful. They also enjoyed the support of many other countries facing similar problems of desertification, including China (China's Agenda 21, 1994; Ci Longjun, 1998). Chapter 12 of *Agenda 21* is entitled: Managing Fragile Ecosystems: Combating Desertification and Drought. It contains a very useful analysis of the causes and consequences of desertification, and a set of highly practical action plans for use at a variety of national and international scales (UNCED, 1992). Paragraph 12.40 (p. 104) provided the impetus for what followed and deserves to be cited in full:

> The General Assembly, at its forty-seventh session, should be requested to establish, under the aegis of the General Assembly, an intergovernmental negotiating committee for the elaboration of an international convention to combat desertification in those countries experiencing serious drought and/or desertification, particularly in Africa, with a view to finalizing such a convention in June 1994.

The General Assembly acted swiftly and established the International Negotiating Committee. The committee met five times and was able complete the final draft of the convention in time for the June 1994 deadline; a remarkable achievement given the complexity of the issues and the diversity of opinions represented within the committee (UNEP, 1995; Kassas, 1995b; Williams *et al.*, 1995; Chasek, 1997). Table 2 is a brief outline of the Desertification Convention.

Table 1 Possible impediments to past anti-desertification measures[a]

Planning/policy
Lack of national plans addressing control measures in development planning
Lack of prioritization due to conflicting demands on financial resources
Inadequate identification of social and economic causes
Poorly designed conservation works and over-emphasis on technical solutions
Disincentives of government pricing policies
Over-optimistic goals
Failure to address increases in human and livestock populations, which exacerbates the processes of desertification
Resources
Global economic recession, reduced financial resources and exacerbation of exploitative land use activities
Lack of external funding for UN in fulfilling the plan of action
Inadequate human resources or appropriate education at national levels
Lack of scientific machinery for management and assessment of desertification
Participation
Inadequate involvement or consideration of local knowledge, experience and skills in planning and implementation of management strategies
Forcing management upon local people who may not willingly accept external interference
Inadequate co-ordination or development of integrated approaches from independent government departments
Failure to combine efforts of UN agencies for interdisciplanary approaches
Monitoring
Lack of effective monitoring mechanisms for degradation processes and assessment status
Lack of follow-up observations on management impacts
Lack of maintenance of projects
Inadequate enforcement of stock limitations and control of grazing to maintain balance between livestock and plant cover

[a] *Source:* Williams *et al.* (1995), adapted from Darkoh (1989, 62–71).

STRENGTHS AND WEAKNESSES OF THE DESERTIFICATION CONVENTION

The strengths and weaknesses of the Desertification Convention re ect the very diverse perspectives of those involved in negotiating the convention (Kassas, 1995b; Chasek, 1997) as well as the lack of scientific knowledge relating to the interactions between desertification and climate (Williams and Balling, 1996). The developing countries in particular were disappointed over the lack of any additional financial resources to cope with some of the fundamental issues such as poverty alleviation and food security (Table 3). The developed countries have

Table 2 Outline of the Desertification Convention[a]

Part 1 Introduction: includes terminology, objectives and guiding principles
Part 2 General provisions: addresses obligations of affected country parties; the priority for Africa; and relationships with other conventions
Part 3 Action programs, scientific and technical cooperation and supporting measures
Section 1 covers action programs at national, sub-regional and regional levels; the importance of international cooperation; the need for support and coordination for both elaboration and implementation of programs; and regional implementation annexes
Section 2 focuses on information collection, analysis and exchange; research and development; and transfer, adaptation and development of technology
Section 3 centers on capacity building, education and public awareness; and financial resources and mechanisms
Part 4 Institutions: provides for the conference of the parties, the permanent secretariat and the committee on science and technology; and stresses the importance of networking of institutions, agencies and bodies
Part 5 Procedures: focuses on communication of information; measures to resolve questions and settle disputes; status of annexes; amendments to the convention and the annexes; and the right to vote
Part 6 Final provisions: includes, inter alia, information on signature, ratification, entry into force and withdrawal
Annexes 1–4: Regional implementation annexes for Africa, Asia, Latin America and Caribbean, and the northern Mediterranean

[a] *Source:* adapted from UNEP (1995).

Table 3 Key aims decided at the Desert Margins Initiative, Nairobi, January 23–26, 1995[a]

Overall goal
To contribute to sustainable food security and poverty alleviation
Overall objective
To promote innovative and action-oriented dry land management research to arrest land degradation
Specific objectives
Understanding land degradation
Assessing dry land management practices
Evaluating the role of livestock in the rangeland/arable land continuum
Designing policies, programs, and institutional options
Improving natural resource management
Fostering domestication of tree species
Formulating drought management strategies
Enhancing institutional capacities
Exchanging technologies and information

[a] *Source:* author's notes and International Crops Research Institute for the Semi-arid Tropics, February 1995.

persistently argued that issues as broad as this should not be part of a legally binding convention (Chasek, 1997). The impasse persists.

Perhaps the most useful aspect of the convention is that it has brought desertification into the international arena as a problem transcending national boundaries, and has encouraged individual countries to prepare National Action Plans (China's Agenda 21, 1994).

It seems fitting to conclude with the words of Baba Dioum, a conservationist from Senegal in Africa: In the end we will conserve only what we love, we will love only what we understand, and we will understand only what we are taught.

See also: Deserts, Volume 1; Desertification, Volume 3.

REFERENCES

Aubréville, A (1949) *Climats, Forêts et Déserti cation de l'Afrique Tropicale*, Société d'Editions Géographiques, Maritimes et Coloniales, Paris.

Chasek, P S (1997) The Convention to Combat Desertification: Lessons Learned for Sustainable Development, *J. Environ. Dev.*, 6(2), 147–170.

China's Agenda 21 (1994) *White Paper on China's Population, Environment, and Development in the 21st Century*, China Environmental Science Press, Beijing.

Ci Longjun (1998) The Impacts of Global Change on Desertification in China, in *Sustainable Development in Arid Zones: Assessment and Monitoring of Desert Ecosystems*, Vol. 1, eds S A S Omar, R Misak, and D Al-Ajmi, Balkema, Rotterdam, 45–60.

Darkoh, M B K (1989) *Combating Deserti cation in the Southern African Regions: an Updated Regional Assessment*, UNEP, Desertification Control Programme Activity Center, Moscow.

Dregne, H E, Kassas, M, and Rozanov, B (1991) A New Assessment of the World Status of Desertification, *Deserti cation Control Bull.*, 19, 6–18.

Glantz, M H, ed (1977) *Deserti cation: Environmental Degradation in and Around Arid Lands*, Westview Press, Boulder, CO.

Glantz, M H, ed (1987) *Drought and Hunger in Africa: Denying Famine a Future*, Cambridge University Press, Cambridge.

Grainger, A (1990) *The Threatening Desert: Controlling Desertification*, Earthscan, London.

Hare, F K and Ogallo, L A J (1993) *Climate Variations, Drought and Deserti cation*, World Meteorological Organisation, Geneva.

International Crops Research Institute for the Semi-arid Tropics (1995) *Sustainable Natural Resource Management Options to Arrest Land Degradation in the Desert Margins of sub-Saharan Africa: Revised Proposal for a Systemwide Ecoregional Initiative*, ICRISAT, Patancheru, Andhra Pradesh, India.

Kassas, M (1995a) Desertification: a General Review, *J. Arid Environ.*, 30, 115–128.

Kassas, M (1995b) Negotiations for the International Convention to Combat Desertification (1993–1994), *Int. Environ. Aff.*, 7(2), 176–186.

Mainguet, M (1994) *Deserti cation: Natural Background and Human Mismanagement*, 2nd edition, Springer-Verlag, Berlin.

Middleton, N and Thomas, D S G, eds (1992) *UNEP World Atlas of Deserti cation*, 1st edition, Edward Arnold, London.

Middleton, N and Thomas, D S G, eds (1997) *UNEP World Atlas of Deserti cation*, 2nd edition, Edward Arnold, London.

Thomas, D S G and Middleton, N J (1994) *Deserti cation: Exploding the Myth*, John Wiley & Sons, Chichester.

Tucker, C J, Dregne, H E, and Newcomb, W W (1991) Expansion and Contraction of the Sahara Desert from 1980 to 1990, *Science*, 253, 299–301.

UNCED (1992) *Earth Summit Agenda 21: Programme of Action for Sustainable Development*, United Nations Environment Programme, New York.

UNEP (1990) *Report of the ad-hoc Consultative Meeting on the Assessment of Deserti cation*, United Nations Environment Programme, Nairobi.

UNEP (1995) *United Nations Convention to Combat Deserti cation in those Countries Experiencing Serious Drought and/or Deserti cation, Particularly in Africa*, UNEP Publication CCD/95/1, Geneva.

Williams, M A J (2000) Desertification: General Debates Explored Through Local Studies, *Prog. Environ. Sci.*, 2, 229–251.

Williams, M A J and Balling, Jr, R C (1996) *Interactions of Deserti cation and Climate*, Arnold, London with WMO and UNEP.

Williams, M A J, McCarthy, M, and Pickup, G (1995) Desertification, Drought and Landcare: Australia's Role in an International Convention to Combat Desertification, *Aust. Geogr.*, 26, 23–32.

FURTHER READING

UNEP (1991) *Status of Deserti cation and Implementation of the United Nations Plan of Action to Combat Deserti cation*, Report of the Executive Director, United Nations Environment Programme, Nairobi.

Desertification: the UN Definition

see Land Degradation in the Mediterranean (**Volume 3**)

Distance Learning and Environment

Jane Dougan

Nova Scotia Southeastern University Oceanographic Center, Dania Beach, FL, USA

Distance learning - planned education or learning in which one or more learners are physically separated, or at a

distance, from the person or source providing instruction, and often also from one another - is not new, but it is undergoing a revolution in reach and form, fueled largely by the explosion in use of information and communication technologies (ICTs), particularly the internet and the world-wide web. In 1981, there were an estimated 213 Internet hosts (computers with a direct connection); by January 1989, there were approximately 80 000. January 2000 surveys estimate that there are now over 72 million Internet host sites worldwide (predominantly in North America, Europe and Asia/Paci c), of which more than six million are educational (Internet Software Consortium, • http://www.isc.org•). This brings new opportunities for learning and new challenges.

With ICTs, learners are not so much at a distance but part of a virtual (cyberspace) learning community. Students (and teachers) can learn independently or as part of a virtual classroom; can access and discuss course materials with other learners in synchronous (real-time) or asynchronous learning networks (ALNs); and can collaborate on real-world or computer-simulated activities and materials locally and remotely. Increasingly, the learning need not be distant (i.e., isolated); instead, it may be distributed or connected learning.

While technology is a key driver, arguably the most challenging considerations involve the intent and content of the materials, and how learning at a distance happens. The hoped-for democratization of the Internet has the potential to further the development of learning communities; incorporating dynamic, interdisciplinary, constructivist, learner-centered approaches that transcend borders of geography and time, altering traditional notions of where formal learning begins and ends. Anyplace, anytime delivery is particularly appropriate in meeting the increasing demand for exible lifelong learning. In June 1999, G8 nations adopted the Cologne Charter - Aims and Ambitions for Lifelong Learning, which identi ed the importance of distance learning, particularly through ICT networks, as a key building block. In terms of environmental sustainability, the ability of the teaching materials to travel to the learner reduces the ecological footprint created by necessitating travel to a central learning institution. The increasing mobility is intellectual, rather than fossil fuel dependent. There are environmental, social and economic bene ts; people can learn while they remain in their community, within their bioregion and close to family and work.

At the same time, there are challenges. Effective distance learning requires ongoing supportive infrastructure at the production and delivery level, as well as relevant, appropriate content. Technology and content alone cannot be sustained without supportive and ongoing regulatory frameworks and distance education policies. Technology-intensive learning materials are time-consuming to prepare. As the global reach of distance learning expands, there is a corresponding need for increased local participation in the production of distance learning materials, particularly those involving multimedia. Distance learning materials predominantly originate within an urban, industrialized world, with English as a lingua franca. At the same time, the use of multimedia and ICTs provides increasing opportunities to incorporate and exchange knowledge of the cultural and social perspectives and ecological realities of the learners. In the context of global change, these are critical considerations, especially if learning is to play a role in supporting environmental, social and economic sustainability at a local as well as global level.

HISTORY AND EVOLUTION OF DISTANCE LEARNING

Distance learning has always evolved in response to global change and transition; re ecting economic and social changes in how and where we live and work, and technological changes in how we communicate and learn. The urge to develop and use tools to interpret and communicate how we live within the environment can be traced back more than 22 000 years, to a time when our ancestors recorded their observations of nature on cave walls. Since then, the tools have increased in breadth and sophistication, as has our knowledge base, and the challenges we face in learning to live sustainably.

Before distance learning, there was learning by correspondence, similarly fueled by a combination of technological innovation (a mechanized printing process, the infrastructure of roads and railways, the organization of the post office) and social change. As literacy and public education increased, so did demand for education by correspondence: in some countries, partly in response to a high concentration of people moving to cities; in others, to educate a population highly dispersed throughout remote areas.

The late 1800s were a time of experiment and individual initiative. In England, Isaac Pittman taught far-away students shorthand by postcard; in Sweden, when a student moved away, Hans Hermod taught him by letter and the resulting correspondence program (Hermods-NK1) is now part of the national educational system. Gradually, correspondence learning was incorporated into national formal educational systems. It expanded during World War II (in large part, to educate service personnel, even within Prisoner of War camps). From then on, learning by correspondence increasingly incorporated radio (e.g., Korea Air-Correspondence Junior College; University College in Nairobi's air-correspondence courses for teachers in rural areas; Madureza Project in Brazil.) In the early 1970s, Britain's Open University used prime-time television broadcasts along with radio and print correspondence,

and provided a successful and innovative model for other countries. Educational broadcast television learning programs have been used with great success worldwide, particularly in Latin America; e.g., Telesecundaria (Mexico), Telecurso 2000 (Brazil). Broadcast school education was also used in widely spaced settlements in Australia.

There are thousands of accredited distance learning programs and courses, and their number is increasing exponentially. Emergence of the Internet as a widely available international network has transformed the ability and means to access, process and share information. The boundaries between traditional and distance learning are dissolving, and distance education courses and even online undergraduate and postgraduate degrees are increasingly seen as credible. Distance learning programs and institutions accredited under Generally Acceptable Accreditation Principles (GAAP) are listed in the World Guide to Higher Education, the Commonwealth Universities Yearbook, or at: http://www.degree.net. On-campus courses offer distance learning components; distance learning courses incorporate real-world local learning activities. Virtual universities challenge the need for a physical campus. The African Virtual University (AVU) (www.avu.org) is a World Bank sponsored satellite-based telecommunications education network using videotape and broadcast to deliver lectures to 24 universities in 16 countries of Sub-Saharan Africa, and developing an extensive digital library. The Virtual University of the Monterrey Institute of Technology (Mexico) delivers university-level courses through live and prerecorded television broadcasts, and incorporates Internet communication for faculty and students. Universidad EAFIT (Colombia) is embarking on Project Virtual Classroom, exploring the use of shared virtual environments as tools for distance education. Project CONEXIONES is aimed at creating a virtual space for children to share ideas and learning with others nationally and internationally (http://sigma.eafit.edu.co/• virtualc/main_eng.htm.).

The JASON Project is an award-winning interdisciplinary online distance learning program allowing student interaction worldwide with year-round virtual scientific expeditions (http://www.jasonproject.org/). Digital Earth is a work in progress towards a seamless virtual representation of the planet that will allow people to explore and interact with the vast amounts of natural and cultural information gathered about the Earth (http://www.digitalearth.gov). Among numerous other environmental courses available by distance learning, Britain's Open University (http://www3.open.ac.uk) offers a postgraduate diploma in Environmental Decision Making with an emphasis on helping students clarify their thinking in the face of complex issues and incorporate environmental considerations into everyday decision making . The University of London Wye College External Programme has been running MSc programs in environmental management and environmental assessment, wholly by distance learning, since 1993. These are just a few examples of the increasing range and number of environmental distance learning courses and programs available.

There are expanding numbers of web sites for teacher training materials and knowledge sharing. The World Links for Development (WorLD) program provides training in the use of technology in education for teachers and students in developing countries, and links students worldwide in collaborative on-line learning projects (www.worldbank.org/worldlinks). Education Network Australia (EdNA) is a exible Internet-based Australian schoolteacher education program focusing on new technologies in learning (http://rice.edn.deakin.edu.au/CurrCorp/Currcorp.htm). Global Learning and Observations to Benefit the Environment (GLOBE) is a network of research scientists, students and teachers from over 8000 schools in over 85 countries working together to understand the global environment (http://www.globe.gov/). Leadership for Environment and Development (LEAD) is:

> a global, independent, education foundation [providing] continuing professional education for outstanding mid-career individuals•••introducing them to issues of environment and development•••with a guiding philosophy•••rooted in the ethical value of caring for our planet and all life that it sustains now and in the future (www.lead.org).

Its mission is accomplished in part by linking over 1000 associates and fellows worldwide via the internet-based network LEADnet, and providing a multi-language search engine LEADcat (http://www.leadcat.org) (*see* LEAD (Leadership for Environment and Development), Volume 4).

From an environmental and social perspective, distance learning via information and communication technology (ICT) can change our notions of community, connecting learners across borders and overcoming traditional political and national knowledge hierarchies. Networks of learning are becoming decentralized. The Association for Progressive Communications (APC) is linking over 50 000 non-governmental organizations (NGOs) in 133 countries to build strategic communities and initiatives•••to equitable human development, social justice, participatory political processes and environmental sustainability (http://www.apc.org).

THE DIGITAL DIVIDE

Learning at a distance has always attempted to reach groups and individuals who might be excluded from formal learning; women, low-income, immigrants, adult learners. The impersonality and reach of correspondence learning overcame more than geographic barriers: e.g., those of race, socio-economic level, gender, age, education-level, and

physical disability. Today, a major challenge is to overcome a digital divide in access to learning and information. The typical Internet user is a young (under 35) college-educated urban male, with a higher than average income and uent in English (the language of 80% of web sites). While access to computers and connectivity is increasing, the gap between the haves and the have-nots remains wide. The number of Internet hosts has more than doubled (from • 30 million to • 70 million) within three years, but it may be a considerable distance to the nearest telephone line; an estimated two-thirds of the world's people have never used a telephone. Even if there is access to networks, the cost may be prohibitive or restrictions in bandwidth (the amount of information that can be sent through a channel) may limit access.

Concerted efforts are being made to bridge global gaps in access, quality and delivery of education, with an increasing focus on the potential of distance learning. International meetings and forums on education and knowledge are attempting to resolve these challenges. For example, the Global Knowledge 1997 Conference (Toronto) and the Second Global Knowledge Conference (Malaysia, March 2000) have strengthened the Global Knowledge Partnership (www.globalknowledge.org). Similarly, the 1990 World Conference on Education for All (EFA) led to the establishment of an International Consultative Forum on EFA (www.unesco.org/education/efa), with further global meetings in Paris (1991), New Delhi (1993), Amman (1996), and a World Education Forum (2000) (http://www2.unesco.org/wef). The May 2000 symposium titled, *Shaping the Network Society: The Future of the Public Sphere in Cyberspace*, sponsored by Computer Professionals for Social Responsibility (www.cpsr.org), resulted in the Seattle Statement (www.scn.org/cprs/diac-00/seattle-statement.html), which calls for active, informed citizen participation and has now been signed by over 150 people from 20 countries. The United Nations Scientific and Cultural Organisation's (UNESCO) Education 21 initiative is embracing adult and distance learning within its quest to strengthen education in support of sustainable development, and their Learning without Frontiers (LWF) program (http://www.unesco.org/education/lwf/) is concerned with issues of access and lifelong learning at a community and individual level. The LWF program was pivotal in the Joint Initiative on Distance Education taken by nine high-population countries (Bangladesh, Brazil, China, Egypt, India, Indonesia, Mexico, Nigeria, Pakistan). More recently, a close collaboration has evolved between UNESCO and the Learning Development Institute (http://www.learndev.org.) towards consolidating and expanding the LWF program.

Online newsgroups and electronic mailing lists devoted to distance learning increasingly focus on issues such as equity and democracy in information-sharing and knowledge-building, and environmental sustainability.

For many countries and regions, especially rural areas, wireless technologies offer the best solution to joining a linked world of learning. Innovative methods of unwiring the World are being explored (http://www.media.mit.edu/unwired/) via international research and partnership. The Telecommunication Development Bureau (BDT) of the International Telecommunication Union (ITU) has developed a virtual training center (http://www.itu.int/VTC) as the base for a networked Global Telecommunication University/Global Telecommunication Training Institute (GTU/GTTI): a virtual university based on distance learning principles, with the aim of overcoming the technology training gap in developing countries. Further, ITU/BDT affirms that:

> the protection of our environment is essential for sustainable world development. Application of modern telecommunication technology to remote sensing, meteorology and telemetry programs can assist humanity to ensure efficient surveillance and management of the Earth's limited resources: water, forests, the atmosphere and organisms and plants that constitute biodiversity.

Accordingly, ITU/BDT organized an international symposium (Tunis, 1996) on the potential role of telecommunication technologies in environmental protection, and has launched a series of related pilot projects. Relatively inexpensive methods of wireless communication can increase accessibility for remote regions and poorer communities. Geographic information systems (GIS) technology is a collaborative tool for sharing maps and data related to environmental learning and natural resource management (*see* Geographic(al) Information Systems (GIS), Volume 4). Digital satellite links and high bandwidth satellite networks are on the horizon. Computers alone are not the solution; a reliable telecommunications infrastructure is required, emphasizing community access; e.g., telecenters in easily accessible locations such as libraries, post offices, clinics. These can offer a broad range of support and access to multimedia services, including distance learning.

As the world's information systems and learning become linked, ethical and economic questions arise surrounding ownership of intellectual property in electronic course materials, and there are disturbing indications that some distance learning courses may become tools for marketing products.

This brief article provides merely an overview of issues surrounding distance learning and environment in the context of global change. There are dozens of new and innovative approaches to environmental learning by distance, and no doubt more just over the horizon. The following is a guide to further sources of information on global change

and distance learning activities, in addition to those already mentioned.

INTERNATIONAL DISTANCE LEARNING ORGANIZATIONS

The American Center for the Study of Distance Education (ACSDE) (http://www.ed.psu.edu/acsde/) publishes the American Journal of Distance Education (AJDE) and a monthly electronic journal, The Distance Education Online Symposium (DEOS). DEOS-L is an interactive listserv, and DEOSNEWS serves approximately 5000 subscribers in more than 80 countries.

Center for Research in Distant and Adult Learning at the Open University of Hong Kong (http://www.ouhk.edu.hk/cridal/) (hosted the first Conference on Research in Distance and Adult Learning in Asia, June 2000).

Commonwealth of Learning (COL), (http://www.col.org/), an international organization created by Commonwealth Heads of Government to encourage the development and sharing of open learning/distance education resources and technologies .

The Development Education Program (http://www.worldbank.org/depweb/) provides teacher and student materials, case studies, and outreach activities related to social, economic and environmental sustainable development.

European Association of Distance Teaching Universities (http://www.eadtu.nl/content.htm).

The Global Distance Education Net (DistEdNet) (http://www.globaldistancelearning.com/) is an online multilingual knowledge guide to distance education targeted for developing countries. There is a core site located at the World Bank, with regional sites elsewhere.

The Inter-American Distance Education Consortium (CREAD) (http://www.cde.psu.edu/DE/CREAD/cread.html) is a network of individuals and institutions throughout North, Central and South America, joining resources and expertise to redefine educational partnerships in the 21st century .

The International Center for Distance Learning (http://www-icdl.open.ac.uk/icdl/), affiliated with the Open University UK, promotes distance learning related research, teaching, consultancy, information and publishing activities, primarily in Commonwealth countries, including an extensive database.

South African Institute for Distance Education (SAIDE), (http://www.saide.org.za/) including Southern Africa Global Distance Education Net.

United Nations Development Programme (UNDP) INFO21 site (http://www.undp.org/info21/sectors/s-edu.html) provides information and links to a variety of issues, including distance education and protection of the environment.

Virtual University (http://www.vu.org/), a non-profit worldwide learning community and the largest educational portal on the Internet today .

Virtual-U (http://virtual-u.cs.sfu.ca/vuweb/) incorporates the Global Educators Network, TeleLearning Network, and Virtual Learning Environments.

Women's International Electronic University, (http://www.wvu.ed/), multilingual, promotes ways for women to learn and women's ways of knowing , including cybercourses.

NEWSGROUPS AND ELECTRONIC MAILING LISTS

ACPACOMX is an international learner discussion list (listserv@mitvma.mit.edu).

INTED-L for learners, educators, trainers interested in international education (listserv@uwwvax.uww.edu).

INTER-L for learners, educators, trainers interested in international education (listserv@vtml.cc.vt.edu).

DEOS-L part of DEOS established by the American Center for the Study of Distance Education, DEOS-L is a moderated listserv for discussion of current issues in distance education (listserv@lists.psu.edu).

Distance Learning and Developing Countries (DLDC): a discussion list for researchers, teachers and students. For information, see http://www.egroups.com/group/dldc/info.html.

Global Knowledge for Development (GKD) discussion list is administered by the Global Knowledge Partnership, consisting of public, private and not-for-profit organizations committed to sharing information, experiences and resources to promote•••knowledge and information as tools of sustainable equitable development . See http://www.globalknowledge.org/discussion.html. To subscribe, contact majordomo@mail.edc.org.

For related articles, *see* Environmental Education, Volume 4; Virtual Environments, Volume 5.

FURTHER READING

Bates, A W (1999) *Managing Technological Change*, Jossey-Bass Publishers, San Francisco, CA.

Burkett, I (2000) Beyond the Information Rich and Poor: Futures Understandings of Inequality in Globalizing Informational Economies, *Futures*, 32, 679–694.

Filip, B (2000) *Distance Education in Africa: New Technologies and New Opportunities*, Japan International Cooperation Agency (JICA) – US Office, Washington, DC.

Gibson, C C (1998) *Distance Learners in Higher Education: Institutional Responses for Quality Outcomes*, Atwood Publishing, Madison, WI.

Palloff, R M and Pratt, K (1999) *Building Learning Communities in Cyberspace, Effective Strategies for the Online Classroom*, Jossey-Bass Publishers, San Francisco, CA.

Disturbances Generated by Humans

see Disturbance (**Volume 2**)

Dobríš European Environment Assessment Process

see Dobríš European Environment Assessment Process (*Case study,* **Volume 4**)

Dooge, James

(1922-)

Professor James Clement Ignatius Dooge was educated in his home country of Ireland and in the US, in civil engineering, uid mechanics and hydraulics.

Jim Dooge's life has moved along four intense and parallel streams, each of which complements the others and each of which could easily fill a single life for most human beings. Thus, Jim Dooge's life has been filled with a large and warm family; teaching and research; key positions in his country's political life; and contributions made to a variety of important international organizations. It has been his unique and multiple talents, energy, commitment, and skills as a leader in so many different areas which has allowed this unusual and rich combination to ourish.

Jim Dooge began his career during the war years as a junior engineer on sewage treatment and river improvement schemes in Ireland, and for eight years after 1946 he was the design engineer at the Civil Works Department of the Electricity Supply Board in Dublin. In 1956 he spent two years pursuing academic studies as a research associate at the University of Iowa, and upon returning to his country he began a long and distinguished academic career in Ireland, at University College Cork. Here he was a Professor of Civil Engineering from 1958 to 1970, and then at University College Dublin, where he was Professor of Civil Engineering from 1970 to 1984. He then became consultant hydrologist to several Irish universities as well as a much sought-after consultant and adviser to numerous international organizations.

Jim Dooge's contributions to his country's academic life have been recognized by various honors, including membership and presidency of the Institution of Engineers of Ireland and the Royal Irish Academy, and honorary membership of the Irish Meteorological Society and the Royal Dublin Society.

At the same time (1948) as the beginning of his academic career, Jim Dooge embarked on a political career when he was elected a member of the Dublin County Council of which he became chairman during two different periods in the early 1950s. After his return from the United States, and while pursuing his career as a Professor in Dublin, he was elected a member of the Irish Senate, having been nominated by the Institution of Engineers of Ireland. In 1965 he was reelected to the Senate and was also elected its deputy chairman. In 1973 he was reelected for a third term and was elected chairman of the Senate. During the period to 1977 he was also member of the three-person Presidential Commission, which exercised the functions of the head of state during any presidential absence or vacancy.

In 1977 Jim Dooge attempted to retire from public life in order to devote more time to academic concerns and to international cooperation in science and engineering. This retirement was not to last for more than four years, as in 1981 he was nominated to the Senate by the Taoiseach, who also appointed him minister of foreign affairs. In 1982 he was reelected to the Senate and became leader of the house and leader of the opposition. In 1983 he became a member of the New Ireland Forum established to pursue the quest of a lasting peace and stability in Ireland. In 1984 he was made chairman of the Ad Hoc Committee of Representatives of the EEC Heads of Government on Institutional Affairs and European Union. In 1987 he was finally able to step down from formal participation in public service, although he has never stopped being an in uential member of Irish life, even as a private citizen.

Parallel to all these accomplishments in Ireland, Jim Dooge began to participate in international scientific and engineering activities. In the mid-1950s he became a member of the International Association for Hydraulic Research and member and later president of the International Association for Hydrological Science. In the mid-1970s he began what was to be a long and important participation in the life of the International Council for Science (ICSU), first through the International Union for Geodesy and Geophysics, and later through ICSU itself, as president of the ICSU-UATI Committee on Water Research. He was secretary-general, and finally president of ICSU from 1993 to 1996. During this period Jim Dooge contributed to a large variety of ICSU's actions, especially in the climate research area, and in the involvement of the scientific community in the preparations for the 1992 United Nations Conference

on Environment and Development. During that year Jim Dooge also served as the chairman of the International Conference on Water and the Environment, held in Dublin. He also served as chairman of the organizing committee of the Second World Climate Conference (1990) and member of the United Nations Educational, Scientific, and Cultural Organization (UNESCO's) Scientific Advisory Committee set up in 1997. He has also served on a variety of committees to advise the European Union's scientific programs.

Jim Dooge has been given numerous honors for his international activities, including the prestigious 1983 Prize in Hydrology, membership in numerous scientific societies and academies of other countries and four Doctorates *honoris causa* from prestigious universities.

In the midst of all this activity, as a family man, a prominent figure on his country's political stage and on the international stage, Jim Dooge managed to keep up an active research and lecturing career. He has given lectures on research results in systems hydrology in a large number of countries on all continents and has served as external examiner in many of these. He began publishing his scientific research as early as 1947, and is still productive, publishing on issues ranging from water sciences, natural disasters, global change to climate issues.

It is difficult to imagine the development of Irish political life, the progress made in hydrological sciences and in international cooperation in science without the remarkable contributions during the past 50 years of the exceptional Professor J C I Dooge, whose personal qualities match his professional accomplishments.

Photo: Courtesy of ICSU.

JULIA MARTON-LEFEVRE *USA*

Dowdeswell, Elizabeth

(1944-)

Elizabeth Dowdeswell, came to prominence in international environmental circles during preparations for the United Nations Conference on Environmental and Development (the so-called Rio or UNCED Conference). She was then Canada's principal delegate to the Intergovernmental Panel on Climate Change (IPCC) and she became the co-chair of the working group on mechanisms within the Intergovernmental Negotiating Committee on the Framework Convention on Climate Change.

Ms Dowdeswell was brought up in rural Saskatchewan in the Canadian Prairies. She obtained an undergraduate degree in home economics and a teaching certificate from the University of Saskatchewan before doing a graduate course in behavioral science at Utah State University in the USA. Then she began a career in education as a high school teacher, university lecturer and educational administrator at the local level.

Her rise in Saskatchewan public affairs was meteoric. Ms Dowdeswell achieved the rank of Deputy Minister of Culture and Youth in her early thirties and set national benchmarks on cultural and heritage policy. Ms Dowdeswell then became a management consultant to the Canadian federal government and served as a leader of several public enquiries into politically delicate issues such as Canada's unemployment, benefits scheme. She was also Canadian Chair of the Great Lakes Water Quality Board in the framework of the Canada-US International joint Commission. This led to Ms Dowdeswell being appointed Assistant Deputy Minister at Environment Canada and head of the Atmospheric Environment Service. This position automatically made her Canada's Permanent Representative to the World Meteorological Organization (WMO) and immediately thrust her into global environmental politics. She served two terms on WMO's executive council and became intensively involved in the climate change debate.

In 1992, Ms Dowdeswell was elected by the United Nations Third General Assembly to succeed Dr Mostafa K Tolba of Egypt as the third executive director of the United Nations Environment Programme (UNEP). She took up her appointment in January 1993, and served to the end of 1997. Included in her UNEP legacy were renewed regional networkings and new foci for UNEP's relationships and partnerships within the UN family, particularly the World Bank, UNEP and the Global Environmental Facility.

On leaving UNEP, Ms Dowdeswell returned to Canada. She currently lives in Toronto.

Photo: by Bill Kiely.

REUBEN J OLEMBO *Kenya*

DPSIR (Driving Forces–Pressures–State–Impacts–Responses)

As an acronym, DPSIR refers to a framework for structuring environmental information. It is a further development of the P–S–R (pressures–state–responses) model initially proposed by the Organization for Economic Co-operation and Development (OECD).

Driving forces, or drivers, designate human activities, which consume resources or generate pollution, nuisances and other disturbances of the environment. They may be grouped into economic sectors (agriculture, industry, energy, transport, tourism, other services, etc.), or in social categories (households, governments, private enterprises, etc.). *Pressures* describe, qualitatively and quantitatively, emissions and waste, as well as the utilization of resources. *State* expresses the modifications of the state of the environment (air and water quality, water resources, habitat destruction, etc.) resulting from the pressures. *Impacts* represent the effects of state modifications on human health and well-being, on ecosystems and on natural resources. *Responses* are measures taken to counter these negative impacts, mostly by public authorities (regulations, economic and fiscal instruments, education, etc.) but also by industry (codes of conduct, reporting, environmental management systems, etc.) and non-governmental organizations (NGOs).

The five elements of the scheme are connected in various ways. There is a causal chain from D to R and there are direct links between R and each of the other elements inasmuch as they constitute the target of the measures taken. If R is aimed at one of the D's, the scheme forms a loop: the D thus affected will presumably reduces the pressures it exerted, the state of the environment will improve and the negative impacts will be mitigated.

The DPSIR model is a useful analytic tool. It presents information in an integrated form, bringing out the complexity of the main environmental issues. It is helpful in developing indicators, whether they be descriptive, performance or efficiency indicators. It facilitates the identification of the most efficient and cost-effective measures to take. Applied dynamically, it provides the means to assess the performance of implemented policies. It brings out the advantages of integrating environmental concerns in sectoral policies (considered as driving forces) as the best way to strive for sustainable development.

PHILIPPE BOURDEAU *Belgium*

EF

EA (Environmental Assessment)

Environmental assessment is an activity that operates at the interface between scientific research and policy formulation. EAs range in size from local (e.g., proposed construction of a power station or reservoir) to global (e.g., control of greenhouse gases). In some countries, an assessment of a local development such as a power station is called an *environmental impact assessment* (EIA). Complex scientific issues such as climate change and biodiversity losses are long term, contain considerable uncertainty and involve many stakeholders, including potential winners and losers; particularly in such cases, an EA (when summarized in easily understandable language) helps policy makers to formulate options and trade-offs.

See also: IPCC (Intergovernmental Panel on Climate Change), Volume 1; Cumulative Environmental Assessment, Volume 4; Environmental Assessment of Major Development Projects, Volume 4; EIS (Environmental Impact Statement), Volume 4; Integrated Assessment, Volume 4; Integrated Assessment Models of Global Change, Volume 4; Social Assessment, Volume 4; Social Impact Assessment, Volume 4; Strategic Environmental Assessment, Volume 4; Social Impact Assessment (SIA), Volume 5.

R E MUNN *Canada*

ECA (Economic Commission for Africa)

ECA is the regional arm of the United Nations (UN), mandated to support the economic and social development of its 53 member states, foster regional integration, and promote international co-operation for Africa's development. Established in 1958 and based in Addis Ababa, Ethiopia, ECA is one of five regional economic commissions under the administrative direction of the UN headquarters. It reports to the UN Economic and Social Council (ECOSOC) through the Conference of African Ministers Responsible for Economic and Social Development and Planning, and convenes a number of intergovernmental organs and committees. The commission has a staff of some 850, comprising professionals with a wide range of expertise, as well as general support and field service.

Since 1995, ECA has embarked on major institutional and managerial reforms. This renewal process has resulted in a new strategic focus defined by five core programs and two cross-cutting themes. The core programs are: Facilitating Economic and Social Policy Analysis; Ensuring Food Security and Sustainable Development: Strengthening Development Management; Harnessing Information for Development; Promoting Regional Cooperation and Integration; Fostering Leadership and Empowerment for Women in Africa; and Enhancement of ECA Capacities.

During its 40-year history, the commission has made a number of significant contributions to African development. The Commission was instrumental in setting up the African Development Bank (ADB), which is the continent's leading multilateral development financier today. It facilitated the establishment of a number of sub-regional organizations, including the Economic Community of West African States (ECOWAS) and the Preferential Trade Area for Eastern and Southern Africa (PTA), now the Community of Eastern and Southern Africa (COMESA). It was instrumental in building some 30 technical institutions to buttress the socio-economic development of Africa, encompassing such fields as technology, banking and finance, minerals and remote sensing, and planning and management. Notable among these is the Eastern and Southern African Management Institute (ESAMI) and the African Institute for Economic Development and Planning (IDEP). The commission also played a key role in the articulation of landmark economic strategies, such as the Lagos Plan of Action and the Cairo Agenda for Re-launching Africa's Development. ECA continues to provide member states with technical assistance in many fields and to foster the development of civil society groups, including non-governmental

organizations (NGOs), professional associations and intellectual networks.

MOSTAFA K TOLBA *Egypt*

ECLAC (Economic Commission for Latin America and the Caribbean)

The ECLAC was originally established as the Economic Commission for Latin America (ECLA) by resolution 106 (VI) of the United Nations (UN) Economic and Social Council (ECOSOC) in 1948; it was redesignated as ECLAC in 1984 by ECOSOC resolution 1984/67, and is headquartered in Santiago, Chile with sub-regional offices in Mexico and the Trinidad and Tobago and national offices in Argentina, Brazil, Colombia, Uruguay and Washington, DC.

ECLAC is one of the five regional commissions of the UN established by the UN to coordinate policies, promote economic and social development and foster regional and international trade within their respective regions. ECLAC's initial focus was Latin America, but it has subsequently expanded its geographical scope to include Caribbean countries.

Membership of ECLAC is open to member States of the UN from North, Central and South America as well as the Caribbean. There are a number of European countries, namely, France, Italy, the Netherlands, Portugal, Spain and the United Kingdom, which are also members due to their strong historical, economic, and cultural ties to the region. A number of non-self-governing territories in the Caribbean region have become associated members. In sum there are 41 members and seven associate members.

In 1996, ECLAC's mandate was updated by ECOSOC resolution 553 (XXVI) in an effort to increase the relevance of activities sponsored in the region. In accordance with this new mandate, ECLAC acts as a center of excellence to aid its member States in the integrated analysis of the development process. This includes formulating, evaluating, and following up public policy measures as well as providing assistance in areas of specialized information. ECLAC experts also offer advice, training and support for regional and international cooperation and coordination.

The meetings of the Commission are held every two years, with a Committee of the Whole meeting in between sessions. At each biennial meeting representatives of ECLAC member States debate topics affecting the Latin American and Caribbean region's economic and social development, review its activities for the previous biennium, and approve its work program for the next biennium for execution by its secretariat, which is headed by an executive secretary.

FURTHER READING

ECLAC in a few words, United Nations ECLAC web site, http://www.eclac.org.

Ministry of Foreign Affairs and Trade of New Zealand (1999) *United Nations Handbook*, Wellington, 122–123.

JOHN W ASHE *Antigua and Barbuda*

Eco Taxes

Kevin S Hanna
Wilfrid Laurier University, Waterloo, Canada

By integrating economic and environmental decision making, economic instruments work within the marketplace to make environmentally positive decisions more attractive. Economic theory suggests that if environmental damage is brought into the market and a value is assigned to negative activities, then it is possible to establish a charge equal to the value of the impact. This discourages unsustainable behavior. But the concept can be dif cult to put into operation, and assigning a monetary value to environmental degradation is in itself a contentious issue. Eco taxes, or green taxes, are one way of providing incentive for environmentally positive behavior, and for internalizing some environmental costs.

In the realm of environmental management, economic incentives generally fall into two primary categories, tax instruments and non-tax instruments. Tax instruments can include tools such as environmental charges, incentives, and combinations of charges and incentives. Environmental charges, or taxes, are usually designed to change behavior by placing a monetary charge on activities that impact the environment. *Eco taxes* are often discussed in the context of a Pigovian tax (named after the economist Pigou), which is also called a pollution tax, where a polluting firm pays a tax based on the cost of damage. The term eco taxes, however, has come to encompass a broader range of tax instruments.

Charges can be applied in several ways; in conjunction with regulations by applying a charge to emissions that

exceed a certain level, application to the total amount of discharge, or to actual amounts of a specific ef uent. Or, instead of applying charges to emissions, they can be applied to inputs or products. Tax incentives modify behavior by rewarding environmentally positive activities, or by penalizing poor behavior. They can include preferential income tax treatment for particular producers, exemptions and deductions, or tax benefits to investors in specific projects and activities. By combining incentives and charges, the two approaches work together by providing a tax on an activity, such as emissions, while providing a rebate or tax credit for introducing new technology. Behavior can be modified while allowing policy makers some exibility to develop ways of offsetting any inequities or costs that may be associated with taxes.

Other options that can function like a tax include user charges, deposit refunds, or even compliance bonds. Deposit refund programs are applied to instances where products can be recycled or reused. A surcharge is applied and refunded when the product or packaging is returned to a collection system. The effect, ideally, is to increase the recovery and reuse or recycling of selected products. Compliance bonds are an option where activities that have environmental impacts are subject to regulation. Such bonds are easily applied to resource extraction activities such as logging and mining, and help insure compliance with reforestation or reclamation requirements. In instances of non-performance the bond is forfeited and applied to remediation.

Tax incentives are often discussed in the context of production activities, with an emphasis on particular emissions or by-products. But specific tax instruments can also be related to land use conservation. For example, tax deductions or credits for contributions relating to conservation, tax deductions for specific uses, and tax exemptions for environmentally positive uses.

EXAMPLES OF ECO TAX INSTRUMENTS

Creating new taxes, or raising existing ones, presents a range of implementation challenges. In most regimes, tax instruments are usually not administered by environmental agencies. In order to put such tools into practice, the cooperation of other departments often has to be secured. Most mechanisms also require support from the political level, and in some instances the general *revenue trap* would have to be addressed. The latter refers to the problem that some agencies have with respect to the monies they generate. In most fiscal systems, funds collected by government are often given over to general revenue, and then reallocated through a budget process. In instances where a conservation agency collects revenue for environmental work, it would have to be certain that it could dedicate these funds to environmental projects either through the budget process or through special purpose accounts.

Regardless of the implementation issues, taxes have long been the traditional and preferred way of raising revenue for most governments. Two main criteria should be considered in eco taxes. Economic efficiency dictates that the net benefits to society should be maximized when the funding mechanism is introduced. Equity, in the economic sense, dictates that the charges, taxes, or fines, should directly re ect or be related to the costs that are imposed on society by the action. Both criteria are based on economic principles and can be measured to varying degrees depending on the instrument and the context. Both equity and efficiency are aspects of overall cost allocation principles. The three basic principles that apply to any evaluation of new taxes and fees are the *polluter pay principle*, the *user pay principle*, and the *bene ciary pay principle*.

The *polluter pay principle* is related directly to the deterioration of the environment and natural resources. Simply defined, it holds that the polluter should bear the cost of measures to reduce pollution decided upon by public authorities to ensure that the environment is in an acceptable state (Opschoor and Vos, 1989). Efficiency is served because the cost of polluting is re ected in the taxes or charges applied, and the market can allocate these to consumers and producers alike. Ideally, there is a clear indication from the market of the cost of environmentally unsustainable activity. Equity is served because the cost is borne by the producers and consumers of the product or service causing the problem (*see* PPP (Polluter Pay Principle), Volume 4).

The *user pay principle* simply holds that the consumer of an environmental benefit bears the cost of consumption. Therefore society does not subsidize the recreation or resource using activities of others. Equity and efficiency is served by allocating costs to those who use the environmental resource. The costs imposed on the environment, such as remediation, restoration, maintenance and related services, are borne by those who use a specific service.

The *bene ciary pay principle* allocates costs based on consumption of a benefit. Those who benefit from a resource pay for its maintenance, restoration or provision. The difficulty in this principle is the process of allocation. While user-pay and polluter-pay principles can be directly applied, the beneficiary pay principle is based on the theoretical willingness to pay for a resource. Willingness to pay, ability to pay, and actually paying are also issues inherent in implementing a tax or fee based on a beneficiary-pay principle. Identifying who benefits from a resource presents significant administrative and technical issues in implementation and feasibility; especially when the benefit is indirect

or not explicitly sought by the recipient. User fees and beneficiary fees are discussed in more detail below.

Some of the products that are provided though the work of conservation agencies can also be considered joint products. Joint products benefit not just the general public in a tangible or intrinsic way, but they also constitute a direct benefit to industry or special interest groups. Clean water, for example, benefits firms that need it for manufacturing. Water quality also affects recreation, the commercial fisheries, and other amenities that often have measurable benefits. The preservation of wilderness and habitat benefits hunting, fishing, and other recreational activities. Habitat and general environmental quality can be related to the desirability of a community as a place to live or as a place to locate business. Governments provide many of these amenities through direct program activities, and indirectly through regulation or other incentives.

Taxes in General

Taxes cover a broad range of instruments applicable to the general public. These can include direct taxes such as property taxes (both *ad valorem* and tangible), taxes on business receipts, business income taxes, personal income tax, and death or gift taxes. Indirect taxes are applied to transactions and expenditures. The value-added taxes common in Europe and North America are applied to virtually all goods and services at most stages of production and to retail sales. Sales taxes, applied to goods sold at the retail level, are common in many regimes. The semantics of taxation overlap and the distinction between some labels can be subtle. Most national governments use many of these methods.

In some instances a special surcharge or surtax could be added to support environmental programs, or a portion of existing receipts could be dedicated to conservation work. These charges can be applied selectively only to residents, transactions, or goods within a defined area. Most tax options have to contend with issues of political and administrative feasibility. Public support may depend on designing carefully targeted taxes that are applied directly to activities or entities that have a direct impact on the environment.

Surtaxes and Surcharges

A surtax, or conservation tax, implies a dedicated increase that is applied to an existing tax, fee, permit, or license. The additional revenue is ideally allocated directly to predetermined program activities. Governments at all levels administer common agents such as commercial fishing and hunting licenses, research permits for animal and ora collection, parks fees, resource royalties, discharge permits, and even highway tolls. Such permits and fees lend themselves to a surtax to support environmental objectives.

Surtaxes can realize significant revenues. Surtaxes are also applicable to income tax. There are examples where healthcare objectives and the application of a minimum required tax are implemented through relatively small surtaxes. British Columbia's Forest Renewal Program is supported by a surtax on timber royalties, the surtax raises about CDN$400 million per annum for forest management programs.

Special Sales and Service Taxes

This category refers to taxes that may apply to the general public as a whole but are added to specific products and services. They are selective and are directly applied to goods and services that relate to a well-defined environmental problem, region, or in support of a specific program. For example, new taxes can be added to fees for hotels, campgrounds, and other accommodation, automobile and recreation fuel, garbage and other disposal services, and products such as pesticides and fertilizers.

Implementing such options often requires integration of efforts among different levels of government, and in some contexts local government might have a significant role. Issues such as the geographical extent of such taxes can also be important. But the potential benefits for the environment can be significant. Additional taxes can be dedicated to environmental work, and publicizing this purpose can make the addition of the tax more palatable to the public and decision makers. The public may accept small increases in consumption taxes if the proceeds are dedicated to environmental work, and monetary dedication is consistent and well stated.

Transaction and Other Unique Charges

Transaction charges do not apply to the general population but are applied to exchanges that directly affect environmental quality. Charges can be applied to land severances, and to production materials used to manufacture chemicals, plastics, and wastes. Transaction charges may also be levied on the transport and disposal of wastes associated with environmental degradation. Complementary waste taxes can be applied to storage, volume of production, and level of toxicity of wastes. The issuance of development permits, the carrying out of inspection visits and similar transactions are also amenable to such charges, which can be applied on almost any level to activities that impact the environment.

User Fees and Beneficiary Fees

User fees, or user charges, are applied to those who use a specific resource or environmental service. Beneficiary

fees have the advantage of being selective and relatively precise. The potential application of user fees is broad and can be applied to recreation, such as parks, water, wildlife, or fisheries, or it can be industrial, agricultural or urban such as the withdrawal of water, sewage or storm water treatment, solid or liquid waste disposal, or even infrastructure use. User fees imply a structure that is ideally designed to capture the full cost of service delivery. They help bridge the information gap by showing consumers that there is a monetary cost associated with specific actions. Ideally, user fees provide an incentive for the sustainable use of resources, and can also act to improve revenue collection for the provision of environmental services, or remediation and clean up.

User fees apply directly to those who access a specific service, and beneficiary fees apply to those who directly or indirectly benefit from an initiative. For example, a beneficiary fee could be applied to fishing licenses within a jurisdiction to support fish stocking or habitat programs, because all license holders potentially benefit from such programs. Or a beneficiary fee can be applied to fishing in a specific lake or bay where a restocking program has been initiated, or shoreline cleanup has been initiated. In instances of fees for special areas, a license can be sold that entitles the holder to fish or hunt within a defined area. The proceeds from a location specific fee or license sur-tax can be dedicated to conservation efforts within that area.

User fees now apply to many services provided by government. Many agencies are gradually introducing and increasing such charges to recover the costs of services they provide, and to generate additional revenue. In most locales, fees are applied at national parks, and many governments are implementing or considering cost recovery methods that are, in essence, user fees. In North America, these are being increasingly extended to the use of selected services, facilities, inspections, licenses, and legal services such as property title transfer, many of which have been previously paid for entirely, or have been subsidized, from general taxes.

REFERENCE

Opschoor, J B and Vos, H B (1989) *Economic Instruments for Environmental Protection*, Organization for Economic and Development Cooperation, Paris.

Eco-efficiency

see Eco-efficiency (*Opening essay,* **Volume 4**)

Eco-engineering to Promote Ecological Sustainability in China in an Era of Global Environmental Change

see Eco-engineering to Promote Ecological Sustainability in China in an Era of Global Environmental Change (*Case study,* **Volume 4**)

Eco-labelling

see Environmental (Eco) Labeling (**Volume 4**)

Ecological Dumping

Ecological dumping: in the context of world trade policies, this phrase refers to the practice of achieving lower export prices (and resulting larger sales) for agricultural products by spending less money on environmental protection.

R E MUNN *Canada*

Economic and Social Commission for Asia and the Pacific (ESCAP)

see ESCAP (Economic and Social Commission for Asia and the Pacific) (**Volume 4**)

Economic Commission for Africa (ECA)

see ECA (Economic Commission for Africa) (**Volume 4**)

Economic Commission for Latin America and the Caribbean (ECLAC)

see ECLAC (Economic Commission for Latin America and the Caribbean) (**Volume 4**)

Ecosystem Approach

see Ecosystem Approach (**Volume 2**)

Ecosystem Health

Ecosystem health may broadly be defined as the state or condition wherein ecosystem functions are not compromised by human activity. Ecosystem health may be viewed from ecological (biophysical), socio-economic, human health, and policy perspectives. From an ecological perspective, healthy ecosystems are characterized by the absence of distress syndrome, by their resilience to perturbations, and by their characteristic organization and productivity. The absence of distress syndrome refers to the absence of commonly exhibited signs of ecosystem pathology under stress including: alterations of biotic community structure to favor opportunistic species, exotic (non-native) species, and life forms of smaller size; increased disease prevalence; loss of biodiversity; and alteration of productivity. Resilience refers to the capacity of the ecosystem to recover from natural perturbations (e.g., forest fires, oods, etc.). Resilience can be measured in terms of both the time for recovery and the extent of recovery. Organization is assessed in terms of the structure of the biotic and abiotic components. Productivity is measured in terms of both primary and secondary production.

Healthy ecosystems supply an abundance of nature's services, such as potable water, renewable resources, etc. Curtailing these services often has severe economic and human health consequences. Economic consequences may include: fisheries lost owing to habitat destruction and/or overharvesting; forests lost to acidification, chemical smog, overharvesting, etc. Human health consequences may include: water-borne diseases that thrive in disturbed ecosystems, respiratory diseases enhanced by chemical smog, new diseases resulting from human interference with ecological integrity (e.g., Lyme disease), etc. Ecosystem health can also be assessed in terms of public policy: in healthy ecosystems, safeguards are in place to prevent human activities from becoming a stress on natural environments. Effective policies regulating land use, waste residual discharges to air and water, overharvesting, and the introduction of exotic species are essential for maintenance of ecosystem health.

DAVID J RAPPORT *Canada*

Ecosystem Integrity

A living system exhibits *integrity* if, when subjected to disturbance, it maintains an adaptive organizing capability to recover toward a goal that is normal and good for that system. Some authors reserve the meaning of integrity to relate only to ecosystems that have not lost, to a significant degree, their natural pristine organization, identity and goal-directedness. Ecosystems that have been transformed through human activities to a different, but desirable, self-organizational complex may be said, by them, to be *healthy* (*see* Ecosystem Health, Volume 4) but not to possess natural integrity. Other authors use the term integrity in both cases, but then bear an onus to specify some benchmark state according to which such integrity may be assessed.

Goals or end-states other than the pristine or naturally whole may be taken by some authors to be normal and good from some human perspective.

Disruptions can arise external to an ecosystem and thus impact it from outside. Climatic surprises and invasions by destructive humans are common examples of such external disruptions. But disruptions can also arise internally in an ecosystem, such as eruptions of species that usually have only a sparse presence, or the collapse of an abundant dominant species. Sometimes external and internal disruptions are interrelated, with an internal disruption triggered by an external event. If the disruptions and stresses are too intense, an ecosystem can disintegrate into a disorderly mess mostly of species that we find useless or offensive, or into a barren area with a few tough life forms.

To maintain systemic integrity in order to survive and thrive, a living thing must maintain an appropriate balance between partial openness to its external environment and partial closedness, in order to manage its internal features. Single-celled and multi-celled organisms create and maintain semi-permeable membranes to serve as a selective barrier within which each can maintain identity, self-organization and goal-directedness as well as other attributes of integrity. The barrier permits good things to enter the living system and bad things to be expelled from it. In humans the skin, lungs and gut are part of this membrane which includes as sensing organs our visual, olfactory, gustatory, auditory and tactile senses.

At higher levels of organization, a population, a species or an ecosystem does not create a biotic physical barrier around itself as a living entity, though some may be constrained geographically by some abiotic discontinuity in their external environment. If we knew how to perceive and measure it, we might find a complex organizational boundary at the space–time edges of a population

or ecosystem. And it might have some analogues to the sensing organs mentioned above. However, this might well extend an organismal analogue of an ecosystem too far.

An ecosystem's external environment may change gradually and cyclically, as with the seasons. Alternatively, it may change abruptly and violently as with storms, droughts and human interventions. The internal system also changes incessantly, in part due to its own goal-directed activities that emerge and change through evolution, and in part due to some irreversible adaptations to external in u-ences. Hence maintaining integrity usually does not involve only simple, stereotypical processes; instead exible and reciprocal adaptation occurs between an ecosystem and its systemic environment. For an ecosystem, such adaptation in turn involves reciprocal rearrangements within its network of nested elements of populations and organisms.

Because self-maintenance of ecosystem integrity may involve emergence of new ecological phenomena, it may not be possible to measure an ecosystem's integrity fully in a deterministic way. Some of an ecosystem's structures and processes that are related to its integrity can be measured. From comparative studies it may have been inferred that some species populations are particularly vulnerable to collapse or eruption when an ecosystem is stressed to an extent that may trigger its disintegration or transformation to another phase. Such integrative indicator species can then be monitored to provide partial insight into the status of an ecosystem's integrity or health.

Other indicators and indices of ecosystem integrity are used in practice. In freshwaters the concentration of the phosphate ion in the spring season is helpful in characterizing the water body as oligotrophic with little phosphate or eutrophic with much phosphate, each of which may be taken to be a state of ecosystemic health or integrity. With intermediate levels of phosphate concentration, the ecosystem may be in a turbulent transitional ux exhibiting less integrity than with either the oligotrophic or eutrophic phase.

See also: Phase Shifts or Flip- ops in Complex Systems, Volume 5.

HENRY REGIER *Canada*

EEA (European Environment Agency)

EEA is a legally independent body of the European Community with a mandate to provide timely, targeted, relevant and reliable information to support the environmental policy process and sustainable development. Established by European Community Regulation (1210/90) and revised in 1999 (Council Regulation 933/1999), the Agency started working in Copenhagen in 1994. In carrying out this mandate the Agency has two main objectives. First, to assist the European Community and member countries to identify, frame, prepare and implement sound and effective environmental policy measures and legislation, and to monitor, evaluate and assess environmental achievements and outcomes. Second, to establish and coordinate the European environment information and observation network (EIONET) for the collection, analysis and assessment of data shared with EC services, EEA member countries and international organizations, and through this to work toward a simplified and streamlined production of European environmental information.

All European Union (EU) countries are members of the EEA, but other countries may join who share the concerns of the EU and member states and the objectives of the agency. Non-EU member countries currently include Iceland, Liechtenstein and Norway. Thirteen new members are expected to join the EEA very soon; among them, ten Central and Eastern European countries together with Cyprus, Malta and Turkey.

The EEA works in all areas enabling it to describe the present and foreseeable state of the environment from the point of view of environmental quality and sensitivity, pressures on the environment and the socio-economic dimension.

The EEA works with national monitoring centers and other bodies in a network which the agency was mandated to set up and coordinate. This European environment Information and Observation Network, based on existing national networks, includes so-called Topic Centers (groups of expert institutes in the member countries under contract to the EEA) which the EEA uses to gather and process information in the different themes of interest. To ensure the broad dissemination of this information and to keep the public properly informed, the EEA publishes regular assessments of the state and outlook of the European environment every 3–5 years (e.g., EEA, 1999), and more frequent indicator-based assessments of the environment (EEA, 2000). Information is also made available to the public directly in a Reference Center accessible through the Internet (http://eea.eu.int).

REFERENCES

EEA (1999) *The EU Environment at the Turn of the Century*, Office for Official Publications of the European Communities, Luxembourg.

EEA (2000) *Environmental Signals 2000*, Office for Official Publications of the European Communities, Luxembourg.

DAVID STANNERS *Denmark*

EIS (Environmental Impact Statement)

An EIS is a summary of an environmental (impact) assessment written in plain English for the various non-specialists involved in the environmental assessment process (including citizen interveners at public hearings).

The contents of an EIS usually include:

1. a description of the proposed development or activity (where will it be located/occur? What are the basic engineering and other facts? Who will benefit? Who may lose?;
2. a discussion of alternatives (e.g., different sites, different forms of energy production) that might meet the same development goals;
3. baseline environmental conditions (a factual account of current environmental conditions and recent trends);
4. a description of the scoping and bounding conditions used to reduce the number of variables/factors considered;
5. predictions of future conditions with and without development (during construction, operation and decommissioning);
6. estimates of the significance of the environmental changes expected;
7. recommendations (not applicable in some jurisdictions):
    - for acceptance, provisional acceptance (acceptance with modifications), rejection; or postponement of a decision until more baseline data are collected and analyzed;
    - for approval of proposed monitoring and inspection procedures to be undertaken during the life of the project.

R E MUNN *Canada*

Emerging Environmental Issues

see Emerging Environmental Issues (*Opening essay*, **Volume 4**)

Emissions Trading

Emission trading is a market-based approach to controlling emissions of air pollutants through the buying and selling of allowances to emit a specified pollutant over a specified period. In contrast to the *command and control approach*, where the maximum allowable emission from each source is regulated and is usually based upon the best available technology, in emission trading a limit (i.e., the total number of emission allowances) is placed upon the aggregate emissions from those sources involved in the trading. The sources may then trade amongst themselves. Ideally the only restrictions would be that the total number of allowances is not exceeded and that environmental requirements such as local air quality standards or maximum rates of acidic deposition from the atmosphere are met. Some sources may create surplus allowances by reducing their emissions or even purchasing surplus allowances from other sources. In some trading schemes, surplus allowances may be banked for use at a future, more advantageous time, or for sale. Those selling their surplus allowances can make a profit if the initial cost of creating the surplus allowances (either through control measures or through the initial purchase) is less than the sale price. On the other hand, if other sources anticipate a shortfall of allowances to cover future emissions (perhaps because of an expansion or more stringent environmental requirements), they may wish to purchase allowances if the purchase cost is less than the cost of reducing emissions. Through trading, therefore, the cost of maintaining or reducing emissions may be less than that of the command and control approach, and managers of individual sources may have more exibility in meeting air pollution requirements. Emission trading has been most extensively used in the United States for controlling the lead content of gasoline, chloro uorocarbons to reduce the destruction of the stratospheric ozone layer, and sulfur dioxide to reduce acid precipitation.

RODERICK W SHAW *Canada*

Endangered Species

Jeffrey A McNeely
The World Conservation Union, Gland, Switzerland

An endangered species (EN) is a species of plant or animal whose population is so low that it is in danger of becoming extinct unless urgent remedial action is taken. Many countries have their own lists of EN that are protected under national law, but the international authority is the World Conservation Union's (IUCN's) Red List Programme: see the 1996 IUCN Red List of Threatened Animals, and the

*1997 IUCN Red List of Threatened Plants (*see *World Conservation Union (IUCN) Red List of Threatened Species, Volume 4).*

IUCN's has three levels of endangerment within its structure of categories of threat to species (Figure 1). It uses five general criteria for assessing level of endangerment based on variables closely associated with extinction risk. These criteria include:

1. declining population (past or projected);
2. small distribution and decline or uctuation;
3. small population size and decline;
4. very small population or very restricted distribution;
5. quantitative analysis (such as population viability analysis).

A population of plants or animals is considered *vulnerable* (VU) when it has an observed, estimated, inferred or suspected population size reduction of at least 30% over the past 10 years or three generations, whichever is longer, and where the causes of the reduction are not demonstrably reversible, or not clearly understood, or may not have ceased, or could recur. *Endangered* (EN) are those whose populations have declined by 50% over the same time, while *critically endangered* (CR) species are those that have suffered a population size reduction of 80% or more over the same time period. Table 1 summarizes the current totals of threatened species.

Most experts consider this list to be very conservative, because it is based on IUCN's very strict criteria and only

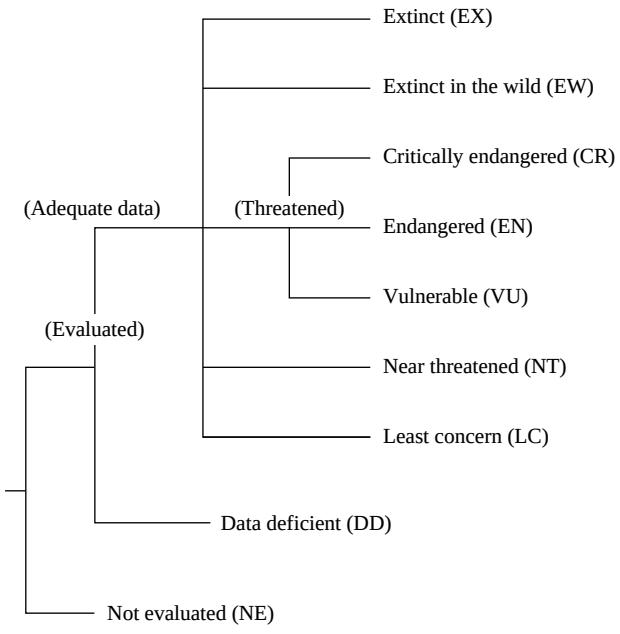


Figure 1 IUCN categories of threat

Table 1 The current totals of threatened species[a]

Taxon	CR	EN	VU	% Threatened
Mammals	169	315	612	22.8
Birds	168	235	704	11.1
Reptiles	41	59	153	3.2
Amphibians	18	31	75	2.6
Fishes	157	134	443	2.7
Insects	44	116	377	
Plants	371	6522	7951	12.5

[a] No percentage is given for invertebrates because the estimate of total insect species numbers is so approximate.

in the case of mammals and birds are there adequate data. A consensus of field biologists working on endangered (EN) has concluded that about 20% of reptiles, 25% of amphibians, and 34% of fishes (mostly freshwater) are under some threat of extinction. The current information on the status of threatened species is heavily biased towards terrestrial vertebrates. Relatively few invertebrates have been assessed, making it difficult to estimate the percentage that are threatened; because the invertebrates are the most numerous species, they are the ones that are projected to lose the greatest number of species.

The IUCN Red List Categories provide an explicit, objective framework that can be applied consistently and widely to classify species according to their extinction risk. This enables comparisons to be made across different taxa, provides a clear understanding of how species are threatened, provides a basis for planning conservation actions, and enables the monitoring of recovery efforts. The criteria are most appropriate when applied at the global level and assigning a taxon to a category may be different when the criteria are applied at the regional or national level. For example, a species that is globally endangered may be common in a country where it is effectively protected; conversely, a species that is common and widespread at a continental level may be endangered in a given country, especially if that country is at the edge of the distribution of the species.

Countries with a strong scientific research capacity often have the largest numbers of EN, especially of the taxa that are difficult to study. For example, Australia, US, and South Africa are among the top four countries with the most EN of reptiles, amphibians, and invertebrates, though none of them makes the top five for mammals or birds. For vertebrates, Indonesia, China, Brazil, and India are among the top five for threatened species of both birds and mammals; but of course these countries are also very rich in species numbers. Some taxa have very large percentages of EN. For example, both species of elephants are endangered, as are all the species of rhinos, a quarter of the species of primates, and seven of the 23 species of crocodiles. Among the birds, albatrosses and grebes have nearly 20% of their species endangered. Among the fish, over half the sturgeon

are endangered. The freshwater mollusks are another group that is highly endangered, especially the species that are endemic to one or a few islands.

EN often are rare, genetically impoverished, of low fecundity, dependent on patchy or unpredictable resources, extremely variable in population density, persecuted, out competed for habitat by alien invasive species, or otherwise prone to extinction as a result of human dominated changes to landscapes. The most significant proximate cause for endangerment is undoubtedly related to habitat, including reduction in area, fragmentation, and overall degradation. Habitat loss or degradation is specifically identified as one of the main reasons for species becoming endangered for 75% of mammals, 44% of birds, 68% of reptiles, 58% of amphibians, 55% of fishes, 52% of crustaceans, 47% of insects, 95% of bivalves, and 50% of gastropods. Many scientists consider these figures to underestimate the proportion of species that are threatened due to impacts on their habitats, but these figures clearly indicate the importance of maintaining intact ecological systems for the conservation of biodiversity.

Another major factor of endangerment is the introduction of alien invasive species that may lead to hybridization, competition, predation, disease, parasites, and the alteration of ecological relationships between species; thus the introduction of these alien invasive species may threaten entire ecological communities, on both islands and mainland areas. Aquatic systems are especially vulnerable to such impacts. Direct exploitation affects at least some species of mammals (such as rhinos and tigers) and reptiles (such as crocodiles).

One objective of putting a species on the endangered list is to foster its recovery, so many governments prepare recovery plans to bring the species back to a population sufficient to remove it from the endangered list. A recovery plan for an EN will accurately diagnose the factors causing the decline of that species and define the measures for counteracting those factors.

Detailed information on globally threatened species can be found on the internet at http://www.wcmc.org.uk/data/database/rl**.** anml**.** combo.html.

See also: Extinctions (Contemporary and Future), Volume 2; World Conservation Union (IUCN) Red List of Threatened Species, Volume 4.

Environment Liaison Center International

see National and Local Responses to Desertification (**Volume 4**)

Environmental Assessment (EA)

see EA (Environmental Assessment) (**Volume 4**)

Environmental Assessment, Cumulative

see Cumulative Environmental Assessment (**Volume 4**)

Environmental Assessment of Major Development Projects

Robert Goodland
World Bank, Washington, DC, USA

Environmental assessment (EA) is that part of economic development planning which seeks to promote sustainability by improving environmental management and decision making. EA is a process which conserves environmental services and prevents delays, while reducing environmental damage and risks, especially to the poor. EA parallels the feasibility phase, often takes at least a year, and feeds into the nal project design. The most important part of EA is a fully budgeted schedule of implementation of the measures designed to prevent, mitigate or compensate for the predicted impacts during construction and operation. Responsibility is assigned to implement the schedule. Capacity strengthening and training is normally essential to implement the mitigation and to monitor results. EA ends with decommissioning and clean up.

EA, an integral part of economic development planning, seeks to promote the sustainability of such development. EA is basket of tools to improve environmental management and decision making. Primarily, EA is designed to conserve or improve environmental services, and to reduce risks. Secondarily, EA reduces and compensates for the adverse environmental impacts of a proposed project. EA reduces costs by preventing damage, especially that harming the poor, and by reducing delays. If the EA is prepared in parallel with the feasibility study of the project, there will be more agreement on the final design, thus reducing

dissent. If a project is prepared without an EA, which has to be added later, the project will be delayed, expensive design changes may be necessary, and delays will be inevitable.

The EA process normally starts as soon as the project has been identified in general terms, but before it has been designed in detail. Typically, to take a highway example, the need for a highway from A to B is identified, but the details (precise routing, width, etc.) have not yet been specified. The process begins with EA *scoping* what will be needed to carry out the full EA, which disciplines will be needed, for how long and when. The main potential impacts are predicted in general terms based on prior experience and a review of the literature about a project of this nature.

The impacts are then screened in order to focus on potentially major impacts and not to waste resources on minor ones. For example, if a proposed highway may run near a hospital or community center, a noise pollution specialist may be needed on the EA team. If the highway may run in or near a National Park, a biologist will be required. Stakeholder views are sought to this stage, often by means of community meetings, requesting comments by publishing an outline of the project in the local newspaper. Radio information and requests for comments are effective for illiterates. Comments received in whatever form are consolidated and the suggestions used to improve design.

A useful part of the early EA process is the ruling out (*screening*) of potential impacts, together with the reasons why they are ruled out. In a reservoir project to site A, for example, if water-borne diseases were to be ruled out, the justification might be that Dr X, an expert from the World Health Organization, with long experience in malaria, writes that neither malaria nor any of its vector mosquitoes, has ever been recorded in this country and that the nearest malaria is 1200 km away at site B, which is sufficiently far for its occurrence at site A to be very unlikely. The water in the proposed reservoir is of such a type that vector breeding is not a big risk. Then similar findings should be sought for other water-related diseases. If a water-borne disease is not mentioned in the EA, stakeholders, especially affected people, rightly ask if it has been overlooked, which means the whole EA is questionable. Such transparency out fosters confidence, hence consensus, by stakeholders. The EA has to be comprehensive at the beginning, gradually eliminating potential issues, so that most attention can be allocated to the more likely impacts. Ruling out of non-issues, and the justification for ruling them out, is a major and important part at the outset of EA.

Following scoping and screening, an interdisciplinary team of environmentalists of appropriate specialties is assembled in the project area and starts work on predicting the nature of impacts, working with the engineering design team to prevent as many impacts as possible by design improvements. The team then devises mitigation for those impacts which cannot be designed out. The EA team and the project design team work closely together for the entire duration of the design phase, often called the *engineering feasibility study*, for a year or more.

At least one year is needed to complete many EAs partly because many environmental impacts may be seasonal (e.g., animal migration such as migrating caribou herds in the Alaska pipeline case), so all seasons need to be monitored. As the frequency of some impacts exceeds annual seasonality, and as many environmental features vary greatly from year to year, more than a single year of monitoring is often necessary.

Implementation of the EA recommendations results in improved project design and a set of safeguard measures to reduce adverse impacts. This preventive approach adjusts the design of the project, such as by re-routing the highway around a sensitive site. The second means to improve the project is the implementation of a series of specified mitigatory measures. The third means is by monitoring during construction and subsequent operation of the development in order to assess the extent to which the design and mitigatory measures are working. Adaptive management adjusts the environmental precautions taken. Escrow accounts or performance bonds may be necessary to pay for cleanup and decommissioning costs after the project, even if it has different owners.

Proper design can prevent most potential impacts, and the remainder can be reduced to acceptable levels, but there will always be a residue of minor impacts which are not predicted, or which are not fully mitigable. If the EA team judges that to be the case, the EA should recommend *offsets*. These are measures to compensate for unmitigated impacts to ensure that the environment is unambiguously better off in net terms with the project. This is in addition to impact prevention and mitigation.

See also: Integrated Assessment, Volume 4; Social Impact Assessment, Volume 4; Valued Ecosystem Component, Volume 4.

Environmental Data Report

Mostafa K Tolba
International Center for Environment and Development, Cairo, Egypt

The UNEP Environmental Data Report series is designed as a critical review of currently available environmental data and related information. More speci cally, through its mix

of tabular and textual presentation of material, it serves a number of functions, namely:

- *identi cation of the principal sources and compilers of environmental data sets;*
- *provision of extracts or summaries of key data sets in tabular form;*
- *highlighting the de ciencies in the available information base, not only in terms of information gaps and the quality of the existing data, but also in terms of relevance to policy-making;*
- *highlighting the ongoing efforts to correct the de ciencies in the information base;*
- *provision of an overview of the signi cant and emerging environmental trends; and*
- *attempts to relate environmental conditions to human activities or pressures on the environment.*

Environmental information (whether as data, statistics, appropriately packaged synthesis or qualitative materials) is becoming increasingly recognized as an essential component of national development plans. National activities, such as transportation, agriculture, energy production or use, waste management, forestry and fisheries have a potential for direct and indirect impacts on the environment. The United Nations Conference on Environment and Development (UNCED) held in Rio de Janeiro in June 1992, stressed the need to integrate environment and development in future policy formulation. Reliable environmental data are therefore necessary for use by governments, as well as by industrial, public and private sectors, to frame policy, to set priorities and to assess results. The improvement and use of available environmental information is a major requirement for the development of national and international policy.

The need for sound information at all levels, from that of the senior decision maker at the international level to the field worker at the national level, was also highlighted at UNCED. The action plan, commonly known as *Agenda 21*, which was adopted at the conference, identified two program areas needing better implementation to ensure that decisions are based increasingly on sound information: bridging the data gap, and improving information availability (UN, 1992).

The absence of data from a number of developing countries, especially in a regional context, hinders national environmental protection and national development and planning, as well as the evaluation of regional and global effects. To achieve cost-effective data collection and assessment, users and their requirements need to be more clearly identified. In addition, ways of transforming existing information into forms more useful for decision-making are essential, together with targeting the information at different user groups.

Governments must have at their disposal not only their own national data upon which to assess the impact of their policies, but also international data to aid their development of sound environmental action plans. In order to disseminate better environmental data at regional and international levels of relevance to management needs, UNEP (United Nations Environment Programme) established, in the mid 1980s, its own biennial Environmental Data Report series. The *UNEP Environmental Data Report* is a source volume of carefully checked, verified and annotated data that provides a basis for detailed environmental assessments at the national and international level. The reports also cover environmentally-related sectors such as transport, energy, human health and industry to satisfy the needs of decision makers. Effective linkages between the different sectors or environmental concerns are made in each report. It is not possible for any one of the reports in this ongoing series, to contain all the data and information which are available for a particular sector or topic. Therefore, in each report, data are chosen to illustrate selected topical issues, geographical concerns, transboundary issues and global data collections.

Examples of data included in the latest editions come from the UN agencies and their cooperating organizations, particularly WHO (World Health Organization), WMO (World Meteorological Organization) and FAO (Food and Agriculture Organization), as well as other international and regional organizations, economic institutions and national State of the Environment Reports (SOE).

Environmental and environment-related data in the Environmental Data Report is organized as follows:

- *The state of the environment*: the first section describes the state of the global environment in terms of environmental contamination, climate change and the status of the natural resource base (Parts 1 to 3).
- *Human and economic activity*: these chapters describe significant trends in the pressures on the global environment. These include population growth and economic development (Part 4), energy production and use (Part 6), industrial activity and transport (Part 7). The problems associated with the generation and disposal of wastes are outlined in Part 8. The impacts of natural and man-made hazardous environmental events (Part 9) and human health issues (Part 5) are also addressed.
- *Policy responses*: the final section outlines some of the regional and global measures taken to protect the global environment and assesses their effectiveness (Part 10).

Each chapter comprises tabular material (presented as a continuous block at the end of the chapter) plus supporting text and graphics. The supporting texts identify the principal data sources, discuss the quality of the existing data sets and summarize key environmental trends, and the footnotes

to the data tables provide information of a more technical nature and full details of the original data sources.

Although each edition of the *UNEP Environmental Data Report* contains a central or core set of data tables which are updated each time (these are identified by a green-colored background), many of the data and supporting information presented in successive editions are new. This material is carefully selected to re ect topical and emerging issues. Owing to space restrictions, it is not possible to provide full coverage of every topic in each edition; some issues are thus covered in alternate editions. However, within the supporting texts, extensive cross-referencing is made to earlier editions and to other chapters. For convenience, references in the text are given after each individual subsection and the key findings are highlighted in a summary paragraph at the beginning of each chapter. Inset boxes provide more detailed discussions of specific issues or a particular study and, where appropriate, examples of environmental indicators.

For more information, consult UNEP, PO Box 30552, Nairobi, Kenya.

Environmental Education

Mohamed Kassas
Cairo University, Giza, Egypt

Environmental education was one of the principal topics on the agenda of the UN Conference on Human Environment held in Stockholm (Sweden) in 1972. Recommendation 96 of the Plan of Action adopted by the Conference required the United Nations Educational, Scienti c and Cultural Organisation (UNESCO) and the United Nations Environmental Programme (UNEP), together with interested international bodies, to establish an international program on environmental education. UNESCO and UNEP collaborated in planning and developing a world cooperative program. The preparatory process included an international workshop in Belgrade, 1975, followed by regional symposia held during 1975- 1977 in: Europe (Helsinki), Africa (Brazzaville and Lagos), Arab States (Kuwait), Asia (Bangkok), Latin America (Bogota), North America (Ottawa), East Europe (Prague). The outcome of these worldwide consultations provided materials for an intergovernmental conference held in Tbilisi in 1977 that launched the International Environmental Education Program (IEEP). In 1987, 10 years after Tbilisi, a second international conference (the Moscow Congress on Environmental Education and Training) produced the International Strategy for Action in the eld of Environmental Education and Training in the 1990s. We note that training was added to education. The Agenda 21 adopted by the UN Conference on Environment and Development (Rio de Janeiro, 1992) contained Chapter 36 on environmental education, which outlined three principal programs: (a) reorienting education towards sustainable development; (b) increasing public awareness; and (c) promoting training.

The UN Commission on Sustainable Development, that monitors progress in implementing Agenda 21, reviewed in its fourth session (1996) progress and constraints related to Chapter 36 of Agenda 21. Environmental education acquired a new form: education for sustainability. For a review of the UN led endeavor in this eld, see Smyth et al. *(1997).*

ENVIRONMENTAL EDUCATION: DEFINITIONS

The following four definitions are not synonymous, they re ect priorities related to various objectives and operational purposes.

1. *Environmental education* is the process of recognizing values and clarifying concepts in order to develop skills and attitudes necessary to understand and appreciate the inter-relationship between humans, their culture and his biophysical surroundings. Environmental education also entails practice in decision-making and self-formulating codes of behavior about issues concerning environmental quality (IUCN, 1972).
2. Environmental education is the process dealing with the relationship between humans and their natural and human-modified surroundings, and includes the relation of population to pollution, resource allocation and resource depletion, conservation, transportation, technology and urban and rural planning and to the total human environment (*US Public Law 91-516, the Environmental Education Act, 1970*).
3. Environmental education is aimed at developing a world population that is aware of, and concerned about, the environment and its associated problems, and which has the knowledge, skills, attitudes, motivation and commitment to work individually and collectively towards solving current problems or prevention of new ones (*International Workshop on Environmental Education, Belgrade, 1975*).
4. Environmental citizenship: environmental education and education for sustainability attempt to create an environmentally responsible population that contributes to sustainable development. The related components of environmental citizenship include information, awareness, concern, attitude, beliefs, education and training,

knowledge, skills, literacy and responsible behavior (Hawthorne and Alabaster, 1999).

Related to environmental education is the concept of *worldview*.

> The expression environmental worldview is typically used to describe beliefs, values and concepts that collectively make up the individual's perception of the environment and humankind's relation to it.
>
> Disignes and Tomsen (1995)

Environmental education aims at: instilling in the learners a positive perception of the environmental worldview, and cultivating in the society, through educational institutions, a broader environmental worldview.

Lucas (1980) recognizes three approaches to environmental education: education in the environment, education about the environment, and education for the environment. While environmental education seeks the goal of understanding the environmental issues, environmental training is to provide selected people with the necessary skills to monitor, assess and manage the environment and the interactions between environment and development.

ENVIRONMENTAL EDUCATION: OBJECTIVES AND GUIDELINES

The Tbilisi Conference (1977) concluded that environmental education was to create and develop the following attributes:

- awareness, to help individuals and social groups acquire awareness and sensitivity to the total environment and its allied problems;
- knowledge, to help individuals and social groups acquire basic understanding of the total environment, its associated problems and the responsible presence of humanity in it;
- attitude, to help individuals and social groups acquire social values, strong feelings of concern for the environment and the motivation for actively participating in its protection and improvement;
- skills, to help individuals and social groups acquire the necessary skills to solve environment–society problems;
- evaluation ability, to help individuals and social groups evaluate environmental measures and education programs in terms of ecological, political, economic, social, aesthetic and educational factors;
- participation, to help individuals and social groups develop the sense of responsibility and urgency regarding environmental problems so as to ensure appropriate action to solve these problems.

The Tbilisi Conference endorsed the guiding principles for environmental education included in the Belgrade Charter (A Global Framework for Environmental Education) produced by the International Workshop of 1975. The principles included:

- consider environment in its totality, including natural and human-modified, technological and social aspects (including economic, political, technological, cultural, historical, moral, aesthetic);
- begin environmental education at the pre-school level and continue through all formal and non-formal stages, making it a continuous lifelong process;
- draw on the specific content of each discipline in making possible a holistic and balanced perspective in order to be interdisciplinary in its approach;
- examine major environmental issues from local, national, regional and international points of view so that students receive insights into the concept of an environmental worldview;
- promote the value of, and need for, local, national and international co-operation in the prevention and solution of environmental problems;
- while taking into account the historical perspective, focus on current and potential (future) environmental situations;
- explicitly consider environmental aspects in plans for development and economic growth;
- enable learners to have a role in planning their learning experiences and provide an opportunity for making decisions and accepting their consequences;
- develop environmental sensitivity, knowledge, problem-solving skills and clarification of values at every age, but with special emphasis in early years on environmental sensitivity to learner's own community;
- help learners discover the symptoms and the real causes of environmental problems, emphasize the complexity of environmental problems and the need to develop critical thinking and problem-solving skills;
- utilize diverse learning environments and a broad array of educational approaches to teaching and/or learning about and from the environment, with due stress on practical activities and first-hand experience.

ENVIRONMENTAL EDUCATION: ETHICS

All through history, humankind was aware of its intimate relationships with nature. The culture of kinship with other creatures on Earth is old, spiritual environmental ethics was part of ancient civilizations that was fostered through teachings of religions. This may have been subdued through societal transformations associated with industrialization and urbanization, processes that have apparently blurred the sense of intimacy with, and dependence on, nature: its

air, water, land and biota. Rural habitats remained closer to nature.

New environmental awareness was initiated in response to environmental damage caused by various forms of pollution and deterioration of life-support systems that became particularly evident during the second half of the twentieth century. Environmental groups of concerned citizens and advances in the sciences of ecology and environment created worldwide feelings that Man has the duty to protect the environment. Environmental education became a tool for nurturing environmental ethics and for instilling prudent attitudes and practices towards nature, its health and its resources.

By the 1990s the concept of sustainable development became prevalent, almost a widely held belief. This has added a future dimension to environmental ethics: responsibility of the present generation towards the well being of the future generations and their environment. *The IUCN-UNEP-WWF (1991): Caring for the Earth: a Strategy for Sustainable Living* calls for worldwide adoption of an ethic for sustainable living, under which every person takes responsibility for his or her impacts on nature, and each generation undertakes to leave to the future a world as diverse and prosperous as the world it inherited. The institutions of religion have an effective role in promoting environmental ethics. For instance, Harvard University initiated a program on religion and environment that included a series of conferences on major religions and environment. See the series of papers in Volume 5 on religion and the environment, for several of the world's religions. Civil society organizations, through initiatives like the green consumer movement, the green-belt movement, etc., are effective tools in cultivating environmental ethics. Professional associations are keenly interested in the roles of its membership in conserving the environment. For instance, the World Federation of Engineering Organizations, through its Committee on Engineering and Environment, produced and adopted in its Annual Plenary Session, New Delhi, 5 November 1985 a *Code of Environmental Ethics for Engineers*. Business, industry and commerce are responding positively to requirements of environmental concerns. Prior to 1980s, industry seemed to regard environmental issues as a peripheral nuisance, but during recent decades the process of greening the corporate boardrooms is spreading. The World Business Council for Sustainable Development is an example. For a review on this subject, see Tolba *et al.* (1992).

ENVIRONMENTAL EDUCATION: IN THE SCHOOL

The formal education system is conceived as comprising four stages: pre-primary, primary, secondary and tertiary. At the pre-primary level, education addresses the child's emotional orientation to nature and to the environment of home and neighborhood. In rural areas in most developing countries, pre-primary schooling is not available, environmental education has no formal structure but is gained through helping at home, in the field and in play; sensitivity to the environment (elements of nature, cycle of the seasons, etc.) comes naturally. In more fortunate urban habitats some nursery and kindergarten schools exist. In developed countries, pre-primary schools are better equipped and trained teachers are available.

Children at primary school can accept integrated knowledge especially in subjects related to health, nutrition and sanitation. Mental perception of the environment seems to develop at the age of 9/10 years. Formal environment education may be introduced as a subject in primary school curricula or as an added element of subjects like natural history, population, hygiene, etc. Adequately trained and motivated teachers are cardinal for success. In large countries with a wide diversity of habitats, teachers need to be capable of framing teaching and teaching materials to suit local environmental conditions. Well-trained and inspired teachers are key elements of success.

The secondary stage (age: 11/12–17/19 years) is in some countries one 6-year school, in others it comprises an intermediate school followed by a secondary school. This is the age of being capable of comprehending the synthetic nature of environmental knowledge. Secondary education is usually oriented to traditional disciplines, and there seems little room in the curricula to accommodate integrative environmental elements. Two approaches are widely debated, but seems to be favoured. (1) Environmental issues may permeate each element (discipline) of the student's learning. (2) Environment with its ramifying aspects may be set as an element in the curricula. In either form, environmental education has become a worldwide feature in secondary schools.

The substance of environment in the school processes may combine in-school basic environmental knowledge (ecology, natural resources, population dynamics, problems of habitat degradation, etc.) and out-of-school activities, involvement in community activity, problem-solving projects, etc. The latter reminds us of Lucas (1980) approach: education in the environment; for this purpose field stations, teaching facilities in biosphere reserves, etc., could provide invaluable support.

Environmental education, as other aspects of environmentalism, does not enjoy easy acceptance. Critics in the USA charge that environmental education is attempting to turn school children into overzealous, unthinking environmentalists ; turning children into environmental crusaders . Two elements are blamed: teaching materials being scientifically weak and politically motivated, and teachers trying to impart their own values to learners. In 1997 an Independent Commission on Environmental Education

(ICEE) carried out a survey and evaluation study on environmental education in US schools; see Disinger (1997) for a review. The ICEE noted that environmental topics were generally infused into other subject areas, mostly science and that environment-focused teaching and learning materials often do not provide a framework for progressive building of knowledge. The importance of the relationship between environmental concerns and development of good citizenship was noted and approved. ICEE recommended, school should consider teaching environmental education as an upper-level multidisciplinary capstone course integrating what students have learned in science, social science, and other upper-level courses . ICEE emphasized that teachers of environmental education need substantive preparation in science, economics and mathematics, and that teaching materials need to be scientifically accurate and well balanced.

ENVIRONMENTAL EDUCATION: TERTIARY LEVEL

Environmental education seeks to provide learners in tertiary institutions with two attributes: environmental literacy and environmental professionalism. Evolution of university programs seems to move from the initial concern with nature conservation (applied ecology) to the wider concern of interactions of development and environment.

Environmental literacy means, apart from the sensitivity required by environmental citizenship, that learners at the tertiary level should obtain basic knowledge of the interactions of society and environment in the processes of resource development and management of human habitat. Here, professionally trained engineers, agriculturalists, physicians, economists, lawyers, etc., need to be fully cognizant of the likely impacts of societal activities and development schemes on the health and quality of the environment. For this purpose, a broadly based general program in environmental sciences may be introduced into the undergraduate curricula. This program may include a first part, available to all students, on the basic interactions within the biosphere and a second part comprising case studies that relate to the professional areas included in the curricula, that is, tailored according to the interest of the class.

Tertiary institutions are required to produce professionals for a variety of job requirements. These include:

1. technicians for environment monitoring;
2. managers of environment schemes;
3. research scholars in fields of environment;
4. school teachers for environmental education;
5. technicians for programs of awareness.

Two lines seem to prevail: institute a university first degree (BSc, etc.) in environmental sciences, or institute postgraduate courses (diploma, MSc, etc.) in environmental sciences. The latter seems to be more widespread, and may be a general (university) degree or a specialized degree (environmental engineering, occupational health, environmental accounting, environmental legislation, environmental education, etc.).

ENVIRONMENTAL EDUCATION: COMMUNICATION AND MEDIA

Success in implementing schemes for conservation of natural resources, protection of the environment, reversing environmental degradation, etc., requires support and positive participation of people. This requires programs of non-formal education that disseminate environmental knowledge, cultivate public awareness and motivate active participation. This component of environmental education has two principal thrusts: cognitive skills that make people understand the environment and appreciate the consequences of action and of inaction and the intimate interdependence between society and environment; and work skills that enable people to participate meaningfully in implementing corrective measures.

The press is an effective tool for education in general. Radio has been and is likely to remain a principal medium for non-formal education in rural areas in developing countries. Television, video and cinema have become widespread media with great potentials for bringing home the environmental message. It can present real environmental scenes or graphics that dramatically mimic natural scenes. These media especially radio and television have the advantages of:

1. presenting messages in series that are regularly transmitted and become part of the daily programs of recipients;
2. combining news (documentaries) with expert comments and explanations;
3. being used for programs of group education (in country clubs, community organizations, etc.) and hence allow for group interactions. They can also have a useful role in the school.

Developments in media interactions with audience and in methods for assessing impacts of media programs provide useful guidance for improving effectiveness and reach.

Non-formal environmental education, through media and community organizations, need to be supported with ow of materials. The International Institute for Environment and Development (IIED, London) created a body (Earthscan) in the early 1970s to provide this support. UNEP and its associates produce a series (OUTREACH) of dossiers, each providing basic information on an environment issue

and variety of means for its narrative and graphic presentation. A dossier is meant to provide raw material that could be remolded to suit various media. OUTREACH is produced in several languages and has now produced (more than 100 issues) a most valuable set of reference materials.

See also: Distance Learning and Environment, Volume 4.

SUMMARY

In the 1970s the concept of environmental education reached the world agenda, in the 1980s it was expanded to environmental education and training, in the 1990s it was transformed to education for sustainable development. This evolution of concepts related to evolution of perception of the role of education in understanding and in managing the relation of mankind and the environment.

Environmental education is a lifelong process that extends from the pre-school stage throughout life. It aims at providing learners with knowledge of, and instilling feelings towards, the environment, and at providing learners with skills that enable them to participate in managing the environment, and in conserving its quality and resources.

Environment education is provided through formal channels (school system with its various grades), and non-formal channels (media, community organizations, etc.).

Environmental education needs trained and inspired teachers and media staff, and teaching and communication materials that are based on correct science and effective presentation.

REFERENCES

Disigner, J F and Tomsen, J L (1995) Environmental Education Research News, *The Environmentalist*, 15, 3–9.

Disigner, J F (1997) Environmental Education Research News, *The Environmentalist*, 17, 153–156.

Hawthorne, M and Alabaster T (1999) Citizen 2000: Development of a Model of Environmental Citizenship, *Global Environ. Change*, 9, 25–43.

IUCN (1972) *European Working Conference on Environmental Education*, Final Report, Morges, Switzerland.

IUCN-UNEP-WWF (1991) *Caring for the Earth: a Strategy for Sustainable Development*, IUCN, Gland, Switzerland.

Lucas, A M (1980) *Science and Environmental Education: Pious Hopes, Self-Praise and Disciplinary Chauvinism, Stud. Sci. Edu.*, 7, 1–26.

Smyth, J, Blackmore, C, and Harvey, T (1997) Education for Sustainability at the United Nations: Making Progress? *The Environmentalist*, 17, 173–179.

Tolba, M K, El-Kholy, O A, El-Hinnawi, E, Holdgate, M W, McMichael, D F, and Munn, R E (1992) *The World Environment: 1972- 1992*, Chapman and Hall, London, 659–692.

Environmental Impact Statement (EIS)

see EIS (Environmental Impact Statement) (**Volume 4**)

Environmental (Eco) Labeling

Osama A El-Kholy
Arabian Gulf University, Manama, Bahrain

Labels serve many purposes. They identify the contents of a package, the manufacturer or supplier of a product, its brand name, composition, certain properties, method of use, warnings, approval or certi cation by a regulatory authority, claims of some advantages, or citation of names of high pro le users.

Labels and trade marks can be registered so as to restrict their use to the original owners. They are protected by law against use by unauthorized persons.

Environmental (Eco) labels provide information, either about a product's, or a service's, eco characteristics (e.g., low energy consumption or low emissions of wastes, absence of certain undesirable ingredients, safer disposal after use, etc.), or its impact on the environment, in general.

Currently, there are three types of eco-labels. Type I labels are voluntary, multiple criteria labels awarded by a third party to products in a particular category. Type II are self-declared claims by a manufacturer, importer, distributor or retailer who stands to benefit from the claims. Type III are based on the results of a process called *life cycle assessment* (LCA).

Experts at a 1991 United Nations Environment Program (UNEP) conference agreed that programs dealing with eco labeling are making the positive statement that a product or service that is to be used for a particular purpose is less harmful to the environment than similar products or services. Consequently, as consumers became more concerned with eco issues, eco-labels became an important marketing tool.

Starting in the late 1980s, eco-labels bearing different names and logos were issued in an increasing number of countries. Since these were being granted using different criteria by a variety of governmental, non-governmental, and business organizations, there were worries that national

labeling schemes could become barriers to international trade, particularly for developing countries. In fact, the prevailing view until quite recently was that eco-labeling was not working. Consumers became skeptical and started to doubt the integrity of eco-labeling schemes. It soon became clear that there was a need for international eco labeling standards. Developing such international standards depends on LCA. This analyses the impact of a product or service on the environment throughout the whole of its life cycle.

Standardizing this process worldwide would lead to internationally acceptable eco-labels. The International Standards Organization is working on developing such international standards for LCA and eco-labeling.

ECO-LABELS WORLDWIDE

The oldest eco-label is Germany's Blaue Engel (Blue Angel), first issued in 1978. Since then the Blue Angel symbol has appeared on a steadily increasing number of German, as well as foreign, products. In 1998, it covered some 4500 products in about 80 product categories from almost 1000 manufacturers. Approximately 13% of the manufacturers, and 15% of the products awarded the label are foreign, coming from 13 European countries together with China, Japan, Korea and the USA.

Authority for the label to appear on a certain product was entrusted, by a private law contract, to the German Institute for Quality Control and Labeling. The label is issued by an eco-labeling jury on the basis of criteria set up by the German Ministry of the Environment, Natural Conservation and Nuclear Safety. The procedures for application and the development of criteria are quite elaborate and involve the Federation of German Industries, Consumers' Association, and the Federal Environmental Agency.

It is worth noting that the Blue Angel is not the only eco-label in Germany. There is competition from smaller programs, but Blue Angel has maintained steady growth over the last few years.

There are national eco-labeling systems in many countries, both developed and developing (e.g., Singapore, India, Zimbabwe). Rather than list them, it is more useful here to dwell on attempts to develop transnational eco-labeling systems that attempt to address the main difficulty with national schemes, viz. the use of different criteria and procedures for issuing an eco-label. The first such initiative came from the Nordic countries. In 1989, Finland, Iceland, Norway and Sweden, later joined by Denmark, launched the Nordic Swan eco-label. Governments subsidize the program which has issued more than 500 licenses covering some 1500 products in just under 50 product groups. The scheme enjoys popularity amongst Nordic consumers not yet matched by other eco-labels in other countries. The Nordic agencies responsible for the program have been appointed as competent bodies for the European Union (EU) eco-label.

The EU eco-label, a ower logo, was started in 1992. This label is intended to replace existing national labels in the member states. Under the EU regulation setting up the system, member states must establish a competent body to run the eco-labeling process. All member states are involved in establishing ecological criteria for product groups. This process is elaborate and goes through six phases before criteria are adopted by a decision of the European Commission. This approach is gathering momentum slowly and some 200 products from under 30 manufacturers have enrolled. Criteria for just over 10 product groups were developed by 1998 and another 10 are in the pipeline. Adopting an EU eco-label means eventually abandoning national labeling schemes in favor of a unified European one.

In North America, the Canadian Environmental Choice Program was established in 1988 by the Federal Government and administered by an environmental services consultancy. The USA Green Seal was established in 1990. The focus of the Canadian program has been in the professional procurement market. Companies that obtained Green Seal certificates in the USA include a number of multinationals. The focus is again on the institutional purchasing sector. The Canadian and US programs joined hands to form the Global Eco-labeling Network (GEN).

GEN, a voluntary not-for-profit association of eco-labeling organizations founded in 1994, sets criteria and certifies products and services. It aims at improving, promoting and developing eco-labeling of products and services. At present, 14 eco-labeling organizations from Europe, Asia, North and South America are members. The GEN secretariat is located in Canada and has an office in Japan.

The year 1999 witnessed growing opposition in Europe to food products made with, or containing, genetically modified organisms. Although such products have been widely consumed in the USA for a number of years, this has changed recently and there seems to be growing opposition to the consumption of such products in the USA as well. This has resulted in the call for eco-labeling of genetically modified food products. There is strong resistance in the USA to unilateral action in this and similar areas. This emphasizes the need for internationally accepted labeling and monitoring procedures (Hasiberger, 2000).

REFERENCE

Hasiberger, A G (2000) Monitoring and Labeling for Genetically Modified Products, *Science*, 287, 21, 431–432.

FURTHER READING

Blue Angel Website (in English), www.blauer-engel.de.
EU Ecolabel Scheme Website, www.ecosite.co.uk.
GEN Secretariat Website, www.gen.gr.jp., GEN Secretariat, 2781, Lancaster Road, Suite 400, Ottawa, ON, Canada K1B1A7.
International Standards Organization, Case Postale 56, CH-1211, Geneva c.20, Switzerland.
ISO 14000 Website, www.ISO14000.com.
RAL (1999) Environmental German Label Blue Angel : Product Requirements, May, Umweltbundesamt, Berlin.

Environmental Policies for the World Oceans

Gunnar Kullenberg
International Ocean Institute, Gzira, Malta

Major ocean issues of great concern to humankind which environmental policies of the world ocean must address include:

- *coastal zone degradation and physical disturbances resulting from multiple, and often con icting, uses by an increasing coastal zone population: coastal erosion, sea level changes, physical alterations in the water and of critical habitats, in particular coral reefs, mangroves, seagrass beds;*
- *marine pollution, mainly of the coastal zones and shelf seas, but also affecting remote areas through long-range atmospheric transport of persistent pollutants; (The major source of marine pollution is land-based activities of society; pollution from shipping has been reduced and is reasonably controlled through environmental policies adopted in the framework of International Maritime Organization (IMO); dumping and incineration at sea have likewise been controlled reasonably well through environmental policies adopted in IMO);*
- *over shing in about 2/3 of all shing areas according to the Food and Agriculture Organization (FAO), and the alterations of marine biological diversity due to disturbances from habitat destruction, pollution, including eutrophication and changes of nutrient balances, shing and transfer of species through shipping and aquaculture developments;*
- *the role of the ocean in global change, including climate change, in the balance of emission and absorption of greenhouse and other gases, e.g., water vapor, carbon dioxide (CO_2), oxygen, changes in ocean circulation and of large-scale perturbations such as El Niño phenomena, frequency and strength of storms and storm-surges;*
- *the exploitation of non-living marine resources, including building material, oil and gas, hydrates, renewable energy through ocean thermal energy conversion, waves, tides, ocean currents;*
- *the exploitation of recently discovered deep-sea ecosystems associated with sea-bottom vents, for medical pharmaceutical and other purposes;*
- *the very limited understanding we have of ocean processes, conditions, resources, biological diversity and ecosystems, with the related considerable uncertainty as to consequences and impacts of changes (e.g., rather sudden changes in the deep ocean conveyor belt), be they natural or human induced;*
- *the transportation of goods by shipping, and the related handling transfers and interfaces, especially in ports, but also safety at sea, and transfer of non-indigenous species;*
- *the recreation and tourism industry, depending on the ocean and coastal zones as major environmental components.*

The importance of the ocean for the global balance of biogeochemical cycles, the climatic variability and stability for the global food production, and thus for global change cannot be over emphasized. A stable sea level over the last 6000–7000 years has been very important for coastal development.

Additional challenges facing environmental policies for the ocean are that oceanic problems are closely inter-related and need to be considered as a whole, that almost all sectors of society in one way or another are dependent upon the ocean or the marine environment, and that currently most national and international institutions are sectorally oriented and are unable to deal with ocean issues in a comprehensive fashion. This means that even in cases where comprehensive international environmental policies for the ocean exist, their applications are fragmented and insufficiently coordinated.

International policy instruments have been adopted which, if they were to be implemented would address several of the issues.

The Third Law of the Sea, the United Nations Convention on the Law of the Sea (UNCLOS) and the results of the Rio Conference, the United Nations Conference on Environment and Development (UNCED) 1992, when taken together, constitute a comprehensive international environmental law for the ocean and a means of its implementation. Chapter 17 of Agenda 21 comprises seven major program areas, which spell out modes of implementation of several parts of the Law of the Sea. The adoption of the Law of

the Sea and its entering into force in 1994 have produced a comprehensive perspective of the ocean, beyond being a body of water and a means of transportation, by focussing attention on the life, production, resources and environmental dimension of the ocean.

Part XIII of UNCLOS establishes a consent regime for marine scientific research, and puts science on an equal footing with economics, law and other major sectors of society.

An important element of the regime for research is cooperation. The need for cooperation is fully recognized and is even made mandatory: states shall cooperate. This, when implemented, will stimulate cooperation also in other circumstances than research. UNCLOS provides stimulus for technology transfer, human resources and institutional development, and capacity building, which is very necessary for the implementation of UNCED-related agreements or any environmental policies for the world ocean.

The global conventions and agreements, signed on the occasion of or resulting from UNCED 1992, addressing climate change, biological diversity, fisheries management, the marine pollution from land-based activities and the desertification issues, are pushing in the direction of developing adequate global observing and early warning systems based partly on the precautionary principle. There is need for adequate data delivered in a timely and reliable fashion for forecasting and compliance monitoring. This is also reason for the development of the Global Ocean Observing System (GOOS), its modules and regional components. Very considerable advances have been made since the inception of GOOS in 1989/1990.

Fundamental pillars of UNCLOS are the Common Heritage Principle and the Exclusive Economic Zone (EEZ). These policy instruments provide a means of achieving greater equity among nations than presently is the case.

Through the EEZ concept, coastal states are given full rights of the resources, but also full responsibility for management, for environmental conditions, pollution, etc. This provides the potential for a focussed and identified management. However, we know that it does not yet work as it could. There is a need to build capacity and provide for education and joint programs. Interdisciplinary and applied research is required; including integration of the results, making integrated assessments, linking natural, social and economical perspectives.

The environmental policies dealing with shipping include the International Convention for the Prevention of Pollution from ships, the Marine Pollution convention (MARPOL) of 1973, with so far six Annexes. Following several tanker accidents in the 1970s, the Conference on Transfer Safety and Pollution Prevention was held at IMO in 1978. It adopted a Protocol on ship design and operation which was incorporated into the Convention on the Safety of Life at Sea (SOLAS) of 1974; and a Protocol related to marine pollution from oil, the 1978 MARPOL Protocol. This Convention is called MARPOL 73/78. The other Annexes of the Convention address various aspects of pollution from ships, the latest being air pollution (Annex VI). The IMO is still working on regulations regarding management of ballast water in order to prevent spreading of non-indigenous species, and of anti-fouling paints harmful to the environment.

Dumping and incineration at sea is regulated through the London Convention of 1972 with various protocols. This is also managed in association with IMO. Special attention has been given to the release of radioactive material into the ocean. The international policies are dealt with through the International Atomic Energy Agency (IAEA).

The complexity of the world ocean makes it attractive to address the environmental concerns through regional mechanisms, which can be harmonized both with global environmental policies and with national ones, and also can address regional priorities. These are often different from region to region.

Since the Stockholm Conference on the Human Environment 1972, the regional approach has been pursued in a systematic fashion with respect to formulation, reaching consensus and creating required mechanisms for implementation of marine environmental policies. This has included the North Eastern Atlantic regions, where the International Council for the Exploration of the Sea (ICES) has served as scientific advisory body external to the regional environmental conventions, which are HELCOM (the Helsinki Commission–Baltic Marine Environment Protection Commission) and OSPARCOM (Convention for the Protection of the Marine Environment of the North–East Atlantic) for the North Sea. The latter took a significant step towards an integrated regional marine environmental policy through its revision in 1992/94, when the Oslo and Paris Conventions, dealing with seaborne and land-based sources of pollution respectively, were combined.

For other major regions of the world ocean, regional environmental policy development has been pursued by the United Nations Environment Programme (UNEP), in its Regional Seas Programme. UNEP has cooperated with all United Nations (UN) Agencies involved in marine affairs in this development. Currently ten Regional Seas Conventions or Agreements are in existence. They are all based on the same principles: overall legal instrument, protocols dealing with specific regional problems, participation of governments through the ministry of the environment and shared costs; with a regional secretariat as coordinating mechanism for implementation of actions agreed by contracting parties. Capacity building including equipment, development of methods for monitoring and surveillance, establishment of early warning systems, training and public awareness creation are important elements of the regional action plans. However, other major components of an integrated policy

Box 1 Acronyms relating to the marine environment, mentioned in this paper but not specifically referenced elsewhere

EEZ:	Exclusive Economic Zone
HELCOM:	Helsinki Commission – Baltic Marine Environment Protection Commission
ICES:	International Council for the Exploration of the Sea
MARPOL 73/78:	Marine Pollution Convention
OSPARCOM:	Convention for the Protection of the Marine Environment of the North–East Atlantic
UNCLOS:	United Nations Convention on the Law of the Sea

for the ocean are not included, e.g., shipping, fisheries, tourism and exploitation of non-living resources.

Only in one region, the Mediterranean Sea, have the contracting parties taken the step towards a comprehensive marine environment policy. This includes establishment of a regional Commission for Sustainable Development.

The fisheries policies are of particular importance (*see* Policies for Sustainable and Responsible Fisheries, Volume 4).

Implementation of environmental policies for the world ocean is governed at the international level through the UN system. The Commission on Sustainable Development, created after UNCED 1992, meets every year and reports through the Economic and Social Council (ECOSOC) to the General Assembly on the implementation of Agenda 21. Interagency mechanisms exist for dealing with the ocean issues, including the Joint Group of Experts on Scientific Aspects of Marine Environmental Protection (GESAMP) created in 1969, and the Administrative Committee on Coordination (ACC) Subcommittee on Ocean and Coastal Areas, created in 1994 as a result of UNCED 1992.

The implementation of UNCLOS is considered annually in a specific agenda item by the General Assembly, on the basis of an UN system-wide report prepared by the Secretary General.

Environmental non-governmental organizations play an important role in focussing attention on issues and in pushing for the establishment of adequate environmental policies for the ocean. Greenpeace International, Friends of the Earth, the International Union for the Conservation of Nature–World Wildlife Fund and the International Ocean Institute must be specifically mentioned for their pioneering efforts (see also Box 1).

REFERENCE

UNCED (1992) *Protection of the Oceans, all Kinds of Seas, Including Enclosed and Semi-enclosed Seas, and Coastal Areas and the Protection, Rational Use and Development of their Living Resources*, Chapter 17, Agenda 21.

Environmental Refugees

Norman Myers
Oxford University, Headington, Oxford, UK

There is a new phenomenon in the global arena: environmental refugees. These are people who can no longer gain a secure livelihood in their homelands because of drought, soil erosion, deserti cation, deforestation and other environmental problems. In their desperation, they feel they have no alternative but to seek sanctuary elsewhere, however hazardous the attempt. Not all of them have ed their countries, many being internally displaced. But all have abandoned their homelands on a semi-permanent if not permanent basis, with little hope of a foreseeable return.

In 1995 these environmental refugees totaled at least 25 million people, compared with 23 million traditional refugees, or people eeing political oppression, religious persecution and ethnic troubles. The environmental refugees total is expected to double by the year 2010, and to increase steadily for a good while thereafter as growing numbers of impoverished people press ever harder on overloaded environments. When global warming takes hold, there could be as many as 200 million people overtaken by sea-level rise and coastal ooding, by disruptions of monsoon systems and other rainfall regimes, and by droughts of unprecedented severity and duration.

INTRODUCTION

Of the 25 million environmental refugees in 1995, there were roughly five million in the African Sahel, where a full 10 million people had ed from recent droughts, only half returning home. Another four million, out of 11 million refugees of all types, were in the Horn of Africa including Sudan. In other parts of sub-Saharan Africa, (where 100 million people are considered to be semi-starving), seven million have been obliged to migrate in order to obtain relief food. In China, with its 120 million internal migrants, at least six million deserve to be regarded as environmental refugees, having been obliged to abandon their farmlands due to shortages of agricultural plots in the wake of decades of population growth. In Mexico there are one million new environmental refugees each year; some become assimilated in cities, and a few return home, leaving

a cumulative total, as a bare minimum, of two million. Finally, there are those people displaced involuntarily by public works projects, notably large dams, increasing in number by 10 million every year (with a cumulative total of 50 million in just China and India). Most of them resettle elsewhere, but the number remaining in a refugee-like situation totals one million.

THE HAITI EXPERIENCE

For years Americans have watched streams of Haitian boat people heading towards Florida. Not only have these people been eeing political oppression, they have been driven by the rundown of the environmental resources on a grand scale (soil, water, trees) that underpin their agricultural economy.

Almost three-quarters of the populace depends directly on agriculture. Only one-third of all land is considered suitable for farming, but population pressures cause three-fifths to be cultivated. Two-thirds of the farmland is on slopes of more than 20°, yet have to support farmers at densities of over 270 km^{-2}, as great as in the most heavily populated agricultural areas of India. The soils are thus exceptionally prone to erosion, half of all farmlands being so eroded that they may be unreclaimable. In some places much of the landscape has lost virtually all its soil, exposing large stretches of bare rock.

The combination of environmental rundown with population growth (Haiti's is the second highest in the Caribbean) means that *per capita* grain production is only a little over half what it was 40 years ago. Many Haitians enjoy only 80% of an acceptable calorie intake: they are chronically malnourished (jargon for semi-starving). One child in ten dies before the age of five. Since 1980, the average cash income has been declining, until today it is just over \$400 $year^{-1}$.

Primarily as a result of all these environmental travails, at least 1.3 million Haitians or one in five of the populace have left their homelands, 300 000 of them heading for the United States. In former times, some of them were driven by political oppression and government corruption, but for most, the predominant factor has been environmental, and to that extent these people deserve to be called environmental refugees.

Even though their country now enjoys a measure of political democracy and economic freedom, many Haitians still lack adequate food. Many more environmental refugees will leave from Haiti for the United States. This migration poses high costs in the destination states. State and local governments in Florida have paid out as much as \$250 million $year^{-1}$ to cater for Haitian aliens, more than 17 times the \$14 million the Agency for International Development has been spending annually for environmental safeguards throughout the Caribbean.

ENVIRONMENTAL REFUGEES WORLDWIDE

The estimate of 25 million for the 1995 total of environmental refugees is cautious and conservative. Scattered throughout the developing world are 135 million people threatened by severe desertification, and 550 million people subject to chronic water shortages. While certain of these people will have been included in the 25 million figure, many could well have been driven to migrate without being counted as environmental refugees.

Of the nearly one billion additional people added to the global population during the 1990s, a good proportion will have been among communities with a cash income of \$1 day^{-1} or less. They include the people most likely to be subsisting, or rather struggling to survive, in environments too wet, too dry or too steep for sustainable agriculture. In sub-Saharan Africa, these environments will have needed to support an extra 150 million people within the 1990s, and a similar total in India.

Poverty serves as an additional push factor associated with the environmental problems displacing people. Other factors include population pressures, malnutrition, landlessness, unemployment, over-rapid urbanization, pandemic diseases and government shortcomings, together with ethnic strife and conventional con icts. In particular, it is sometimes difficult to differentiate between refugees driven by environmental factors and those impelled by economic problems. In certain instances, people with moderate though tolerable economic circumstances at home feel drawn by opportunity for a better livelihood elsewhere. They are not so much pushed by environmental deprivation as pulled by economic promise. This ostensibly applies to many Hispanics heading for the United States. But those people who migrate because they suffer outright poverty are frequently driven also by root factors of environmental destitution. It is their environmental plight as much as any other factor that makes them economically impoverished. This generally applies to those refugees who migrate to areas where economic conditions are little if any better than back home, as is the case with many people who migrate within sub-Saharan Africa and the Indian subcontinent. In this instance, with poverty and life on the environmental limits as the main motivating force, it matters little to the migrants whether they view themselves primarily as environmental or economic refugees.

In short, there is a gradient of factors at work. At one end are those who are driven by environmental problems outright, and at the other end are economic migrants who are voluntary opportunists rather than refugees. In between is a gray zone where one category sometimes tends to merge into the other. The assessment to date is no more than a first-cut effort, albeit preliminary and exploratory, to come to grips with a prominent and fast growing problem that is all too real for those who endure it, however much the purists may argue about final definitions.

On top of all these sub-problems is the lack of official recognition, whether on the part of governments or international agencies, that there is an environmental refugee problem at all.

THE FUTURE OUTLOOK

How many environmental refugees can we realistically anticipate in the future: or rather, how many people are likely to become vulnerable to environmental problems that could force them to migrate? Let us consider the outlook for the year 2010. The population of developing countries is projected to grow by more than one billion people, a 24% increase in just 15 years. Sub-Saharan Africa's total will have expanded by some 240 million, a 42% increase, and the Indian subcontinent's by 377 million, 32%. The numbers of people in absolute poverty are predicted to swell from 1.3 to 1.6 billion. The 135 million people affected by severe desertification could well increase to 180 million. The populations of water short countries, 550 million today, are expected to surge to more than one billion. If the 1985 onwards plateauing of crop yields continues, there will be more, and more widespread, shortfalls in food production (especially in sub-Saharan Africa and the Indian subcontinent), while international tradable stocks will be increasingly unable to keep up with fast burgeoning demand.

The 25 million environmental refugees in 1995 had mostly become obliged to migrate since 1980, when their numbers first started to climb rapidly. In light of patterns and trends of environmental decline and associated problems such as spreading poverty and population increase, it is probable that by 2010 there will be at least another 25 million such refugees if only because the impelling factors will continue to be at least as prominent for the communities concerned. (This supposes too that there will be few preventive measures of sufficient scope.) In fact, the increase could well be more than 25 million because of increasingly degraded environments coupled with growing numbers of impoverished people.

For a specific instance of the problem's scope to expand, consider the prospect for sub-Saharan Africa until the year 2010. This is already the region with half the world's traditional refugees and at least a similar proportion of environmental refugees. Despite some advances in soil conservation (Kenya, Ethiopia), small scale agriculture (Nigeria, Zimbabwe), reforestation (Tanzania, Malawi), anti-desertification (South Africa), and population planning (Kenya, Zimbabwe, Botswana), the outlook is unpromising. The region's population is projected to increase to over 800 million people, fully 42% more than in 1995. Severe desertification may well affect more than 100 million people, half as many again as today. Ten countries are expected to be experiencing chronic water shortages or even acute water scarcity, with collective populations totaling well over 400 million people. Without greatly expanded efforts to tackle the region's lack of development, *per capita* Gross National Product (GNP) will probably stagnate in real terms at around $500.

Most important of all will be the region's incapacity to feed itself. Some twenty countries with a projected collective population of 440 million are expected to experience up to 25% shortfall in food supplies, and a further eight countries with a projected 75 million people face more severe deficits. *Per capita* food production is scheduled to decline even below the meager 1990 level, possibly as much as 20% below. The total number of malnourished people could grow to 500 million, with 100 million destitutes obliged to live for the most part off imported food (in just southern Africa there were 20 million people dependent on relief food in 1992). The food deficit could well rise to as high as 50 million tons. Due to its exceptional poverty the region will be increasingly unable to compete in the global grain market. Food aid worldwide in 1995 was only 7.5 million tons, enough to make up the diets of only 10 million semi-starving people.

On the global scale, there will be problems of climate warming. Due largely to sea-level rise and ooding of coastal zone communities, but also to increased droughts and disruptions of rainfall regimes such as monsoonal systems, global warming could threaten large numbers of people with displacement by 2050 or earlier. Preliminary estimates indicate the total of people at risk of sea-level rise in Bangladesh could be 26 million, in Egypt 12 million, in China 73 million, in India 20 million, and elsewhere, including small island states, 31 million, making a total of 162 million. At the same time, at least 50 million people could be at severe risk through increased droughts and other climate dislocations.

All in all, the issue of environmental refugees promises to rank as one of the foremost human crises of our times. To date, however, it has been viewed as a peripheral concern, a kind of aberration from the normal order of things, even though it is an outward manifestation of profound deprivation and despair. While it derives primarily from environmental problems, it generates problems of political, social and economic sorts. As such, it could readily become a cause of turmoil and confrontation, leading to con ict and violence. Yet as the problem becomes more pressing, our policy responses fall further short of measuring up to the challenge. To repeat a key point: environmental refugees have still to be officially recognized as a problem at all.

At the same time, there are limits to host countries' capacity, let alone willingness, to take in outsiders. Immigrant aliens present abundant scope for popular resentment, however unjust this reaction. In the wake of perceived threats to social cohesion and national identity, refugees can become an excuse for outbreaks of ethnic tension and civil disorder, even political upheaval. This is already the

case in those developed countries where immigrant aliens increasingly prove unwelcome, as witnessed by the experience of Haitians in the United States and North Africans in Europe. Almost one-third of developed countries are taking steps to further restrict immigrant ows from developing countries.

Yet measures to relieve the plight of refugees of whatever kind have drastically diminished in relation to the growing scale of the problem. Although the annual budget of the United Nations High Commissioner for Refugees has recently been boosted to $1.3 billion, the agency is increasingly unable to supply food and shelter for refugees of traditional kind alone, much less to invest in rehabilitation or repatriation of these refugees. Meanwhile, the world's refugee burden is borne primarily by the poorest sectors of the global community. In 1995 the 20 countries with the highest ratios of official (traditional) refugees had an annual per capita income of only $750.

POLICY OPTIONS

There is much scope for preventive policies, with the aim of reducing the need to migrate by ensuring an acceptable livelihood in established homelands. First of all, we need to expand our approach to refugees in general in order to include environmental refugees in particular. We cannot continue to ignore environmental refugees simply because there is no institutionalized mode of dealing with them. If official standing were to be accorded to these refugees, this might help to engender a recognized constituency, e.g., those 900 million people who endure some degree of desertification, four million of whom have become environmental refugees in the Sahel alone. While desertification entrains costs of $42 billion year^{-1} just through the loss of agricultural produce, the United Nations' Anti-Desertification Action Plan would cost no more than $10 billion year^{-1}. Yet the amount subscribed so far is less than $2 billion, ostensibly on the grounds that arid land dwellers have no constituency and hence lack political leverage.

Second, we need to widen and deepen our understanding of environmental refugees by establishing the root causes of the problem; not only environmental causes but associated problems, plus the interplay of the two sets of forces. There are many conceptual gray areas as proximate concerns and ultimate causes, the contributory roles of population pressures and poverty, the linkages to ethnic tensions and conventional con ict, and so forth. Suppose, for instance, there had been a better understanding and hence a better anticipatory response on the part of the Philippines government when the agricultural frontier closed in the country's lowlands during the late 1980s, bringing on a sudden increase in landlessness and an upsurge of migration into steep country in search of land. The problem could have been somewhat relieved through before-the-event measures to redistribute existing farmlands, to improve rural infrastructure and to supply more off-farm employment, thereby reducing the need to migrate out of the lowlands.

Consider too the root causes of famine. If a famine has been human-made, it can be human-unmade, whereas natural factors can only be managed and accommodated. Just as the recurrent droughts in sub-Saharan Africa cannot all be blamed on climate, so the recurrent famines cannot all be blamed on drought. Drought has often served to trigger famines by disrupting the social, economic and political processes that would normally ensure sufficient access or entitlement to food.

Probably most important of all is that there can be little advance except within an overall context of what has come to be known as *sustainable development*. This applies notably to reliable access to food, water, energy, health and other basic human needs, the lack of which is behind many environmental refugees' need to migrate. In big picture terms, sustainable development represents a sound way to preempt the environmental refugee issue in its full scope over the long run. As a prime mode to tackle the issue, then, there would be a handsome payoff on investment to foster sustainable development in developing countries through greater policy emphasis on environmental safeguards, together with efforts to stem associated problems such as poverty, population and landlessness.

However, only select parts of sustainable development measures address the particular problem of environmental refugees. Especially pertinent, for example, would be the Anti-Desertification Action Plan as applied to the Sahel and arid sectors of the Horn of Africa, both being sources of large numbers of environmental refugees, whether present or prospective. Also warranting closely targeted responses are food-short regions within sub-Saharan Africa and the Indian subcontinent.

Much could be achieved too through better targeting of foreign aid. The annual budget of the main source of multilateral aid, the United Nations Development Programme, has declined so much in recent years that it is now little more than that of the United Nations High Commissioner for Refugees ($1.3 billion in 1995), meaning that the United Nations' main response in this regard tends to be as reactive as proactive. Worse, the 10 developing countries with well over two-thirds of the world's poorest of the poor receive only one-third of foreign aid and it is impoverished communities that serve as the source of most environmental refugees. India has 20% of all people in absolute poverty worldwide, yet it receives only five percent of total foreign aid. Were foreign aid to be more closely directed at impoverished people in the main countries and regions concerned, it could help to relieve the problem while it is still becoming a problem, i.e., before it becomes entrenched.

Equally to the point, developing countries and aid organizations alike devote only around 10% of their development expenditures to priority human needs such as water and sanitation, nutrition, and health, the lack of which is often a contributory factor to the environmental refugee problem. The communities most deficient in these essentials are the 1.3 billion people in absolute poverty. The challenge could be largely surmounted if the funding proportion were to be doubled to 20%. The additional annual cost for the United States would be no more than $2 billion, or $7.7 per American (the price of a beer every three months). Alternatively, the sum could readily be made available by restructuring the US aid program of $9 billion year^{-1}.

There is much more that developing countries themselves can do, and at no great cost. According to the United Nations Children's Fund (UNICEF), to eliminate deaths from famine would cost only a little more than $0.5 billion year^{-1}; to cut malnutrition among women and children (who make up a disproportionate share of environmental refugees), less than $2 billion; and to reduce hunger among the poorest households, little over $6 billion. All of these measures would help to reduce prime pressures generating environmental refugees. The total cost would be less than $9 billion, or $7 for each of the 1.3 billion people in absolute poverty: these being the communities that are a main source of environmental refugees. By contrast, developing countries spend an annual average of $40 citizen^{-1} on military activities. In 1991 Ethiopia assigned 13% of GDP to this purpose, four times the global average, even while it featured some of the largest numbers of environmental refugees in proportion to population size.

FURTHER READING

Doos, B R (1995) *Can Large-Scale Environmental Migration be Predicted?* Global Environmental Management, Vienna.

Kane, H (1995) *The Hour of Departure: Forces that Create Refugees and Migrants*, Worldwatch Institute, Washington, DC.

Myers, N and Kent, J (1995) *Environmental Exodus: an Emergent Crisis in the Global Arena*, The Climate Institute, Washington, DC.

Westing, A H (1992) Environmental Refugees: a Growing Category of Displaced Persons, *Environ. Conserv.*, 19, 201–207.

Environmental Responses: an Overview

see Environmental Responses: an Overview (*Introductory essay*, **Volume 4**)

ESCAP (Economic and Social Commission for Asia and the Pacific)

The United Nations (UN) ESCAP is one of the five United Nations Regional Commissions, the others being Africa, East Europe, Latin America and the Caribbean, and West Europe and others.

ESCAP was established in 1947 in Shanghai, China, as the Economic Commission for Asia and the Far East to assist in post-war economic reconstruction. The Commission moved its headquarters to Bangkok in January 1949. The name was changed in 1974 to re ect both the economic and social aspects of development, and the geographic location of its member countries. ESCAP has been the main UN economic and social development center for Asia since 1977.

ESCAP is the largest of the five UN Regional Commissions, comprising 51 members and nine associate members and representing some 60% of the world's population, or 3.5 billion people.

ESCAP members include Japan and Australia; the dynamic tiger economies (Republic of Korea, Singapore, Thailand, Indonesia, Malaysia), and seven new Republics of Central Asia, which are undergoing rapid economic transition. Thirteen of the world's least developed countries make up one-quarter of ESCAP's full members, and 19 small Pacific island developing countries account for over one-third of ESCAP's total membership. The smallest country is Niue with a population of 2000. The largest is China with a population of 1.2 billion people.

ESCAP spawned two of the region's vital institutions – The Asian Development Bank and the Mekong River Commission.

The main legislative organ of ESCAP is the Commission, which meets annually at the ministerial level and reports to the UN's Economic and Social Council. It provides a forum for all governments of the region to review and discuss economic and social issues and to strengthen regional cooperation.

MOSTAFA K TOLBA *Egypt*

EU (European Union)

The European Union (EU) was established by the Treaty of Maastricht, signed on the 7th of February 1992, and further defined by the 1997 Amsterdam Treaty. It currently includes

the 15 countries belonging to the European Communities: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxemburg, Netherlands, Portugal, Spain, Sweden, and the United Kingdom, totaling 370 million citizens.

The Union, which is grounded in the European Communities treaties, has the task to organize, in a manner demonstrating consistency and solidarity, relations between the member states and between their peoples.

The Union's objectives are:

- to promote economic and social progress which is balanced and sustainable, in particular through the creation of an area without internal frontiers, the strengthening of economic and social cohesion and the establishment of economic and monetary union;
- to assert its identity on the international scene, in particular through the implementation of a common foreign and security policy;
- to develop close cooperation on justice and home affairs.

The Union is served by a single institutional framework including the following:

- the European Parliament, directly elected, with legislative power and the task to control the use of executive power;
- the council of the EU, representing the governments of the member states, which sets political objectives and coordinates national policies and shares legislative power with the European Parliament;
- the European Commission, with three basic functions: initiator of proposals for legislation, guardian of the Treaties and manager and executor of Union policies and of international trade relations;
- the Court of Justice, which provides the judicial safeguards to ensure that the law is observed in the interpretation and the application of the Treaties;
- the Economic and Social Committee, representing the various categories of economic and social activity and advising the commission, the council and the European Parliament;
- the Court of Auditors, the European Investment Bank, the Committee of the Regions, the European Central Bank and the European Ombudsman.

The EU is in charge of common policies in the areas of trade, agriculture, fisheries, transport, research and technological development, environment, etc. It sets basic minimum standards in environmental protection, which must be applied in the member states, and it has the duty to integrate environmental concerns in the definition and implementation of sectoral policies in order to promote sustainable development.

The European Commission is headed by a College of Commissioners, one of which is responsible for environment policy and oversees the Directorate General Environment.

The budget allocated to this Directorate General is used for the implementation of action programs, the preparation of new legislation, the funding of projects aimed at improving the environment in the member states (LIFE and Cohesion programs). A special appropriation is made for the European Environment Agency, located in Copenhagen, whose mission is to provide comprehensive information on the state of the European environment and the effectiveness of policies. Environmental research is managed by the Directorate General Research, which funds large multinational projects in support of the relevant EU policies.

PHILIPPE BOURDEAU *Belgium*

European Environment Agency (EEA)

see EEA (European Environment Agency) (**Volume 4**)

European Union (EU)

see EU (European Union) (**Volume 4**)

Exergy (High Quality Energy)

see Monitoring in Support of Policy: an Adaptive Ecosystem Approach (*Opening essay*, **Volume 4**)

FAO (Food and Agriculture Organization)

The FAO was founded in October 1945 with a mandate to raise levels of nutrition and standards of living, to improve agricultural productivity, and to better the condition of rural populations.

Today, the FAO is the largest autonomous agency within the United Nations system with 180 member nations plus the Economic Union (EU) (member organization) and more than 4000 staff members around the world. Following recent efforts to decentralize, the FAO's staff includes almost 2200 people at headquarters and more than 1800 working at decentralized offices and field projects. The organization's 2000–2001 biennial budget is set at $650 million. FAO assists developing countries through agricultural and rural development projects with about US$300 million per year with funding obtained from donor agencies, governments and its own regular programme; half of this assistance is for technical assistance and support and half for emergency assistance.

Since its inception, the FAO has worked to alleviate poverty and hunger by promoting agricultural development, improved nutrition and the pursuit of food security – the access of all people at all times to the food they need for an active and healthy life. The organization offers direct development assistance, collects, analyses and disseminates information, provides policy and planning advice to governments, and acts as an international forum for debate on food and agriculture issues.

The FAO is active in land and water development, plant and animal production, forestry, fisheries, economic and social policy, investment, nutrition, food standards and commodities and trade. It also plays a major role in dealing with food and agricultural emergencies.

A specific priority of the organization is encouraging sustainable agriculture and rural development, a long-term strategy for the conservation and management of natural resources. It aims to meet the needs of both present and future generations through programs that do not degrade the environment and are technically appropriate, economically viable and socially acceptable.

For further information, see telex-room@fao.org or http://www.fao.org/.

HENRI CARSALADE *Italy*

FCCC (Framework Convention on Climate Change)

see United Nations Framework Convention on Climate Change and Kyoto Protocol (**Volume 4**)

Focus Groups for Integrated Assessments

see Integrated Assessment (**Volume 4**); Public-driven Response to Global Environmental Change (*Opening essay,* **Volume 4**)

Food and Agriculture Organization (FAO)

see FAO (Food and Agriculture Organization) (**Volume 4**)

Founex Meeting (1971)

see United Nations Conferences on the Environment (**Volume 4**)

Framework Convention on Climate Change

see United Nations Framework Convention on Climate Change and Kyoto Protocol (**Volume 4**)

G

GATT (General Agreement on Tariffs and Trade)

The GATT was created in 1947 as a draft agreement on rules for trade in goods between then only 23 contracting parties. In order to service the GATT agreement and facilitate trade negotiations, an *ad hoc* organization known as GATT was established alongside the GATT agreement. The GATT was not a fully- edged international organization, having no legal personality, but was rather a revolving contractual agreement. This status persisted through the 1986–1994 Uruguay Round of world trade talks, during which contracting parties negotiated additional agreements for trade in services (General Agreement for Trade in Services, GATS) and intellectual property (the Agreement on Trade-related Aspects of Intellectual Property Rights, TRIPs). An agreement was also made during the Uruguay Round to establish the World Trade Organization (WTO), which came into being as an international organization on 1st January 1995.

The WTO was established as an intergovernmental organization to administer all the trade agreements under its auspices, and accession and adherence to those agreements. The WTO's main decision-making bodies are councils and committees in which all members can participate. With its secretariat in Geneva, and current membership of 140 countries, the purpose of the WTO is to regulate governments' interventions into international trade according to a negotiated rules-based system, to settle trade disputes between parties, and to organize trade negotiations. Administrative and technical support is also provided by the WTO.

While GATT no longer exists as an institution, the GATT agreement continues to exist in an amended form as a principal WTO agreement. Moreover, the GATT agreement's key principles of non-discrimination, transparency and predictability have been incorporated into the agreements on services and intellectual property, and are now widely recognized as key principles in all other WTO agreements. The GATT, GATS and TRIPs agreements, and their related subsidiary agreements, today make up core agreements that are administered by the WTO and are subject to a single system of rules for resolving disputes. All WTO members adhere to core agreements and some adhere to various additional WTO plurilateral agreements.

Although the WTO has no specific agreement dealing with the environment, a number of the WTO agreements include provisions dealing with environmental concerns. Following the Uruguay Round, members created the WTO Committee on Trade and Environment (CTE) to study the relationship between trade and the environment. The CTE has the responsibility, inter alia, to make recommendations to the WTO General Council about any changes that might be needed in WTO trade agreements for environmental purposes.

For further information, see http://gatt.org.

HUSSEIN ABAZA *Switzerland*

GEF (Global Environment Facility)

see United Nations Framework Convention on Climate Change and Kyoto Protocol (**Volume 4**)

General Agreement on Tariffs and Trade (GATT)

see GATT (General Agreement on Tariffs and Trade) (**Volume 4**)

GEO (Global Environment Outlook)

The State of the Environment Assessment unit of the United Nations Environment Programme Division of Early

Warning and Assessment has the core mandate to keep under review the state of the global environment and identify emerging issues that require international attention. A major activity of the unit is the preparation, through a collaborative, participatory process, of the GEO reports. These reports provide comprehensive assessments of the state of the global environment and reviews of policy responses as well as outlooks for the future.

The global environment assessment process, through which the GEO report is produced, is cross sectoral and participatory, incorporating regional views and perceptions and building consensus on priority issues and actions through dialogue among policy-makers and scientists at regional and global levels. Input is solicited from an extensive array of sources throughout the world: regional and thematic collaborating centers, United Nations organizations, and independent experts. The first GEO report (GEO-1) was published in January 1997(OUP/ISBN 0 19 521349 1), the second (GEO-2000) was published in September 1999 (Earthscan/ISBN 1 85383 588 9), and the third (GEO-3) is due to be published in time for the 2002 World Summit on Sustainable Development.

The main GEO report is underpinned by numerous technical and regional reports. Other by-products have been developed out of the GEO process, including overview booklets, which provide a brief synopsis of the main reports. A training manual is being used to train collaborators and others involved in integrated environmental assessment and reporting around the world. A GEO-for-youth report, *Pachamama: Our Earth, Our Future*, was published in October 1999 (Evans Brothers/ISBN 0 237 52119 9). Most of these products are available in more than one language.

GEO reports can be accessed through http://www.unep.org. Further information can be obtained from geo@unep.org.

MARION CHEATLE *Kenya*

Geographic(al) Information Systems (GIS)

Paul Longley
University College London, London, UK

Measurement, simulation and modeling of environmental change entails a multitude of geographic problems that range from the routine to the monumental. Yet the solutions to diverse environmental problems share common characteristics since they almost always involve aspects of location. Thus, for example, a mineral extraction business is concerned with locational problems when it determines where to dig, how to route access roads, and how subsequently to institute land ll and landscaping activities. In similar vein, a farmer needs to solve locational problems in order to make better decisions about the amounts of fertilizer and pesticide to apply to a eld system. Location is also fundamental to environmental impact assessment.

Longley *et al.* (2001), from which this discussion is adapted, classify locational problems according:

- *Scale* (level of geographic detail): the effects of human activity upon the environment may be assessed at a range of scales from the architectural to the global.
- *Intent*: while some problem-solving is driven by human curiosity and the quest for greater scientific understanding, others problems are much more solution-oriented and practical in nature.
- *Timescale*: some geographic problems are manifest over short time periods, while others assume greatest importance over the long term.

Geographic information systems (GIS: often termed geographical information systems in Europe) provide a framework for locational problem solving across all geographic scales, in scientific as well as application-driven contexts, and can support both short- and long-term decision-making. Some aspects of the development of GIS applications may be much more straightforward than others. The handling of geographic information often begs unique, complex and difficult choices, although application is more straightforward if the objects of study are crisp and well defined. The cost of creating reliable and robust geographic information is often great, yet there are few technical obstacles to subsequent reproduction and distribution of it to other users. Many environmental questions cannot be solved with reference to just a single source of information, yet it can be straightforward to process or merge different information sources. If these and other challenges can be met with both rigour in generalization and sensitivity to context, GIS allows us to assess how the world looks, the forms it takes, how it works, and the natural and human processes that fashion it (Longley *et al.*, 2001).

The specific form of the world varies geographically, and every surface (or near surface) environment is unique. However, the effects of processes upon surface environments are much more general – for example, the effects of greenhouse gas emission on the environment are similar irrespective of where on the Earth the source is. GIS is unique in that it allows us to gauge the effects of general processes upon specific environments. Thus, for example, the effects of acid rain upon forests in Scandinavia

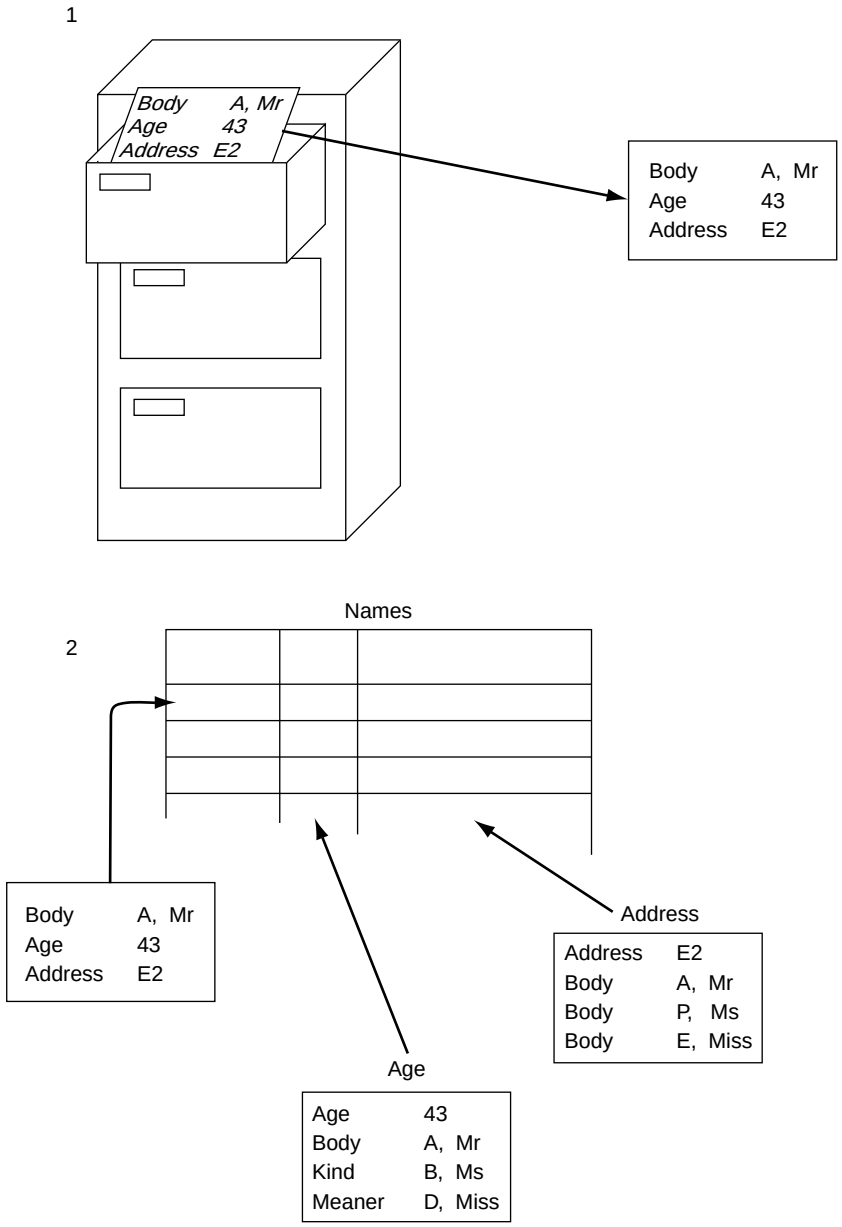


Figure 1 A GIS can be thought of as a digital filing cabinet of maps (Martin, 1996)

might be identified by applying general models of atmospheric dispersion and circulation to the specific air pressure systems prevailing over the North Atlantic and Baltic in a particular study period.

In many geographic problems, there is no simple relationship between a single process and an observed outcome. Rather, a range of factors may conspire to produce an observed outcome in a particular instance, as with the combined impact of climate, bedrock geology, aspect and slope angle upon rates of soil erosion following deforestation. A second unique property of GIS is that it enables the relative importance of a range of different factors to be assessed and incorporated in location-specific models of process. GIS can be adapted to circumstances in which such factors may not be expressible in commensurate terms, or where it is necessary to accommodate multiple goals and objectives.

GIS is conventionally deemed to comprise:

- software, designed to perform well-defined functions (GIS software);
- digital representations of aspects of world (GIS data);
- a community of people who use these tools for various purposes (the GIS community);
- applied problem-solving or scientific investigation (doing GIS).

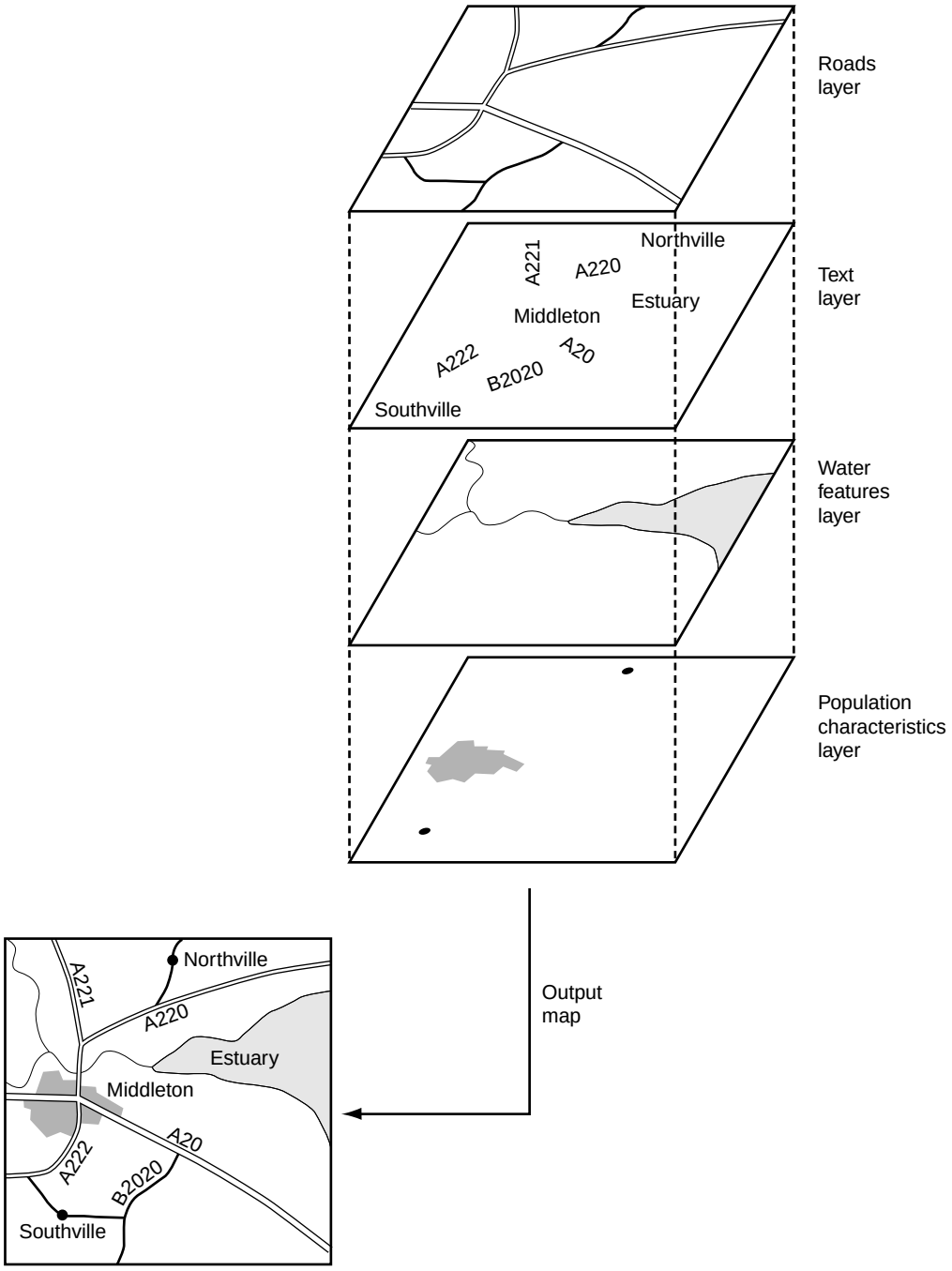


Figure 2 GIS can be used to create composite maps

There are many ways of defining GIS (see Longley *et al.*, 2001). Perhaps most straightforwardly, a GIS can be thought of as a digital version of a filing cabinet of maps (Figure 1: adapted from Martin, 1996). However, GIS is much more than this, because data from different sources can be brought together in order to create composite maps (Figure 2), which can be used directly to inform decision-making. In this sense, a second definition of GIS is as an environmental decision support system. Natural resource managers often need to know where spatial objects are. A third way of considering GIS is as a mechanized inventory of geographically distributed features or objects. A fourth definition of GIS emphasises the integrating role of GIS, as a scientific tool for revealing attributes of composite datasets (such as those shown in Figure 2) that are invisible in any one of the constituent themes. Finally,

GIS may be considered as providing a toolkit of basic techniques for performing operations on geographic data that are too tedious or expensive or inaccurate if performed by hand.

The first GIS was the Canada Geographic Information System, designed in the mid 1960s for natural resource management. In an unrelated development in the late 1960s, the US Bureau of the Census developed dual independent map encoding in order to support automatic referencing and aggregation of census records. It was not long before the early GIS developers identified that common solutions were applicable to the widest range of application areas, and soon generic computer software was being developed for computer mapping (Foresman, 1998). An early distinction was between software written to manipulate a regular grid of cells known as a *raster*, each element of which of which takes a numerical attribute value, and the vector data structure that comprises points, lines, nodes, segments and polygons (Figure 3). The development of raster GIS has also been associated with the development of satellite remote sensing (*see* Remote Sensing, Volume 2). The modern history of GIS dates from the early 1980s, when the price of sufficiently powerful computers fell to the point at which larger numbers of smaller organizations could adopt the innovation.

Today's GIS is a complex of software, hardware, people, institutions, networks, and activities. The major software vendors (such as ESRI Inc., Intergraph Corporation, etc.) provide distinct products to serve very different needs. These range from the geographically enabled spreadsheets that are available as adjuncts to popular word-processing packages, through simple systems for viewing, analyzing, and mapping data, to software for supporting GIS-oriented Web sites and sophisticated software systems that support the widest range of spatial analysis operations.

A GIS will normally comprise some or all of the following component parts.

- *Hardware* to enable the user to interact with the computer software – typically through typing, pointing, clicking or speech. Hardware may be configured as desktops, laptops, personal data assistants, in-vehicle devices, cellular telephones and a range of other devices.
- *Software*, that is the suite of interrelated computer programs that allows a range of spatial data operations to be performed.
- A *database*, which may have been created for a particular purpose, but in many environmental applications will likely have been created to fulfil a range of purposes (generic data, such as framework data: Rhind, 1997).
- *Network* connection for rapid communication or information sharing – typically the Internet and/or the intranets of large organizations. In so-called client-server GIS, the network connects the user's hardware device (the client) to one or more servers of data and/or software.
- *People* who design, program and maintain the GIS, who supply it with data, and who interpret the results of GIS-based analysis.

Huge industries have developed around each of these component parts. The GIS data industry accounts for the greatest proportion of the estimated US$ 7 billion annual sales value of GIS (1999 prices), and is involved with the acquisition, creation, maintenance, dissemination and sale of GIS data. The GIS software industry directly employs several thousand programmers, software designers, systems analysts, application specialists, and sales staff. The provision of GIS hardware is part of the information technology mainstream, and also supports specialized vendors of geographically enabled devices such as global positioning system receivers. The innovation of the Internet also led to the development of the GIS service industry, in which businesses, agencies and the public are provided with remote access to data services in return for subscription and/or advertising revenues. Education in GIS is also a significant business in itself, and GIS is taught through thousands of university programs, through on-line programs (e.g., ESRI Inc.'s Virtual Campus at http://www.esri.com) and through in-service training.

Representative GIS applications in the environmental realm focus on topics such as the conservation of wildlife

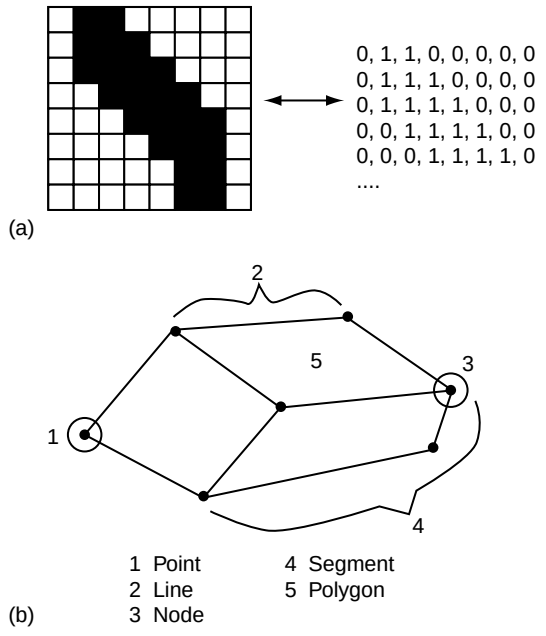


Figure 3 (a) The raster data structure comprises a regular grid of cells; while (b) the vector data structure comprises points, lines, nodes, segments and polygons

and biodiversity, planning and management of nature reserves, interpretation and protection of scenery, and management of relationships between different land uses (Aspinall, 1999). The integrative power of GIS allows the wider geographical context of designated sites to be considered, and approaches based on co-operation have replaced a focus on competition between land uses. This has been given immediate relevance by initiatives related to sustainable development and the maintenance of biodiversity. Sustainability also emphasizes the interdependence of human populations and environmental resources, and GIS is an important tool for evaluating the broad remit of wide-area conservation initiatives. GIS is also increasingly used as a decision support tool to monitor environmental change in real time (Larsen, 1999).

Geographic information systems are undoubtedly useful tools for measurement, modeling and analysis of global environmental change. Yet their usage also raises frustrating and profound questions, such as:

- how does a user know whether apparently significant results are robust and defensible?
- What is the most appropriate spatial and temporal scale for assessing change?
- What principles might help a GIS user to design better maps?
- How may representations of physical, environmental and social systems, and the interactions between them, best be integrated?
- How can user interfaces be made readily understandable by novice users?

Resolution of these questions is central to the developing field of *geographic information science* (GISc) (Goodchild, 1992). GISc studies the fundamental issues arising from the creation, handling, storage and use of geographic information, as a well-defined class of information in general. The related terms *geomatics*, *geoinformatics*, *geocomputation*, *spatial information science*, and *geoinformation engineering* all suggest similar scientific approaches to the fundamental issues raised by the use of GIS and related technologies.

Today's GIS is much more than a single piece of software residing on a single computer, and the practice of GIS has been revolutionized by the development of the Internet. Some applications, including those in the environmental realm, have been criticized for their ethically questionable motives, and it is important to remain cognizant that while any technology may appear neutral, it is always used in a social context. GIS remains a rapidly developing field, yet its core organizing principles, techniques and management practices will be much more enduring than the particular proprietary software systems that are used to put it into practice (see Longley *et al.*, 2001 for details).

REFERENCES

Aspinall, R (1999) GIS and Landscape Conservation, in *Geographical Information Systems: Principles, Techniques, Management and Applications*, eds P A Longley, M F Goodchild, D J Maguire, and D W Rhind, John Wiley & Sons, New York, 967–980.

Foresman, T W, ed (1998) *The History of Geographic Information Systems: Perspectives from the Pioneers*, Prentice Hall PTR, Upper Saddle River, NJ.

Goodchild, M F (1992) Geographical Information Science, *Int. J. Geogr. Inf. Syst.*, 6, 31–45.

Larsen, L (1999) GIS in Environmental Monitoring and Assessment, in *Geographical Information Systems: Principles, Techniques, Management and Applications*, eds P A Longley, M F Goodchild, D J Maguire, and D W Rhind, John Wiley & Sons, New York, 999–1007.

Longley, P A, Goodchild, M F, Maguire, D J, and Rhind, D W (1999) Introduction, in *Geographical Information Systems: Principles, Techniques, Management and Applications*, eds P A Longley, M F Goodchild, D J Maguire, and D W Rhind, John Wiley & Sons, New York.

Longley, P A, Goodchild, M F, Maguire, D J, and Rhind, D W (2001) *Geographic Information Systems and Science*, John Wiley & Sons, Chichester.

Martin, D J (1996) *Geographic Information Systems: Socioeconomic Applications*, 2nd edition, Routledge, London.

Rhind, D W (1997) *Framework for the World*, John Wiley & Sons, New York.

GESAMP (Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection)

GESAMP is an advisory body sponsored by the International Maritime Organization (IMO), the Food and Agriculture Organization (FAO), the United Nations Educational, Scientific and Cultural Organization (UNESCO) and its Intergovernmental Oceanographic Commission (IOC), the World Meteorological Organization (WMO), the World Health Organization (WHO), the International Atomic Energy Agency (IAEA), the United Nations (UN) and the United Nations Environment Programme (UNEP). Members of GESAMP are appointed by the sponsoring organizations and serve in their personal capacity. The activities of GESAMP fall into two broad categories:

1. preparation of periodic reviews on the state of the marine environment; and

2. evaluation of specific scientific issues relevant to the protection of the marine environment.

GESAMP has defined *marine pollution* as introduction by Man, directly or indirectly, of substances or energy into the marine environment (including estuaries) resulting in such deleterious effects as harm to living resources, hazards to human health, hindrance to maritime activities including fishing, impairment of quality for use of seawater and reduction of amenities. This definition provided the basis for a legal definition of marine pollution in most international conventions dealing with the protection of the oceans, including the United Nation Convention of the Law and Sea (UNCLOS). GESAMP actively participated in the preparatory process leading to the United Nations Conference on Environment and Development (UNCED).

Since it was established in 1969, 69 reports and studies have been published in GESAMP reports and studies series. In addition to two global assessments of the state of the marine environment (GESAMP, 2001a), studies published in the last decade include topics such as: review of nutrients as potentially harmful substances; evaluation of the hazards of harmful substances carried by ships; some aspects of pollution modification of atmospheric and oceanic processes and climate; atmospheric input of trace species to the world oceans; long-term consequences of low-level marine contamination; coastal modeling: global strategies for marine environmental protection (GESAMP, 1991); significance of carcinogens as marine pollutants; impact of oil, individual hydrocarbons and related chemicals on the marine environment; environmental consequences of anthropogenically discharged sediment to the coastal zone; guidelines for marine environmental assessment; biological indicators and their use; monitoring of ecological effects of coastal aquaculture wastes; invasion of the ctenophore *Mnemiopsis leidyi* in the Black Sea; sea-surface microlayer and its role in global change; contributions of science to integrated coastal management; patterns, threat and development of a strategy for conservation of marine biodiversity; hazard assessment of ships' cargoes; safe and effective use of chemicals in coastal aquaculture; and land-based sources and activities affecting the quality and uses of the marine environment (GESAMP, 2001b).

For further information see http://gesamp.imo.org/.

REFERENCES

GESAMP (1991) Global Strategies for Marine Environmental Protection, IMO, London.

GESAMP (2001a) A Sea of Troubles, UNEP, Nairobi.

GESAMP (2001b) Protecting the Oceans from Land-based Activities, UNEP, Nairobi.

STJEPAN KECKES *Croatia*

GIS (Geographical Information Systems)

see Geographic(al) Information Systems (GIS) (**Volume 4**)

Global 500 Roll of Honor for Global Achievement

see International Environmental Prizes (**Volume 4**)

Global Change System for Analysis Research and Training (START)

see START (Global Change System for Analysis Research and Training) (**Volume 4**)

Global Environment Outlook (GEO)

see GEO (Global Environment Outlook) (**Volume 4**)

Global Environmental Change and Natural Disasters

Jim Bruce
Ottawa, Canada

Rapidly escalating disaster losses have been the most striking and disturbing consequence of global environmental change. Land use changes and increased exposure to the risks have undoubtedly played the major role, but there is also increasing evidence that intensity and frequency of extreme weather events in a changing climate are contributing to the rising disaster toll.

DISASTER LOSS DATA

The global increases in losses from great natural disasters are truly alarming. In the mid-1960s, losses averaged about $5 billion per year, rising until in the 1990s they were about $50 billion per year in constant dollars (an estimated $93 billion total loss in 1998) (Munich Re). These loss estimates are based on informed extrapolation of insured losses. During this period with a 10-fold increase in disaster loss, global economic production has increased three to four times and the global population has doubled. Thus, the gross global data on population and economic activity exposed to risks cannot begin to explain the increase in economic damages in disasters.

A very important factor, which has contributed to the increasing losses, has been the migration of populations to vulnerable regions, particularly the growth of large urban areas on the world's coasts and along river valleys and growth of squatter settlements near cities. Increasing wealth in many urban communities also has increased the exposure to risks. However, reinsurance industry spokesmen consider that these factors alone cannot account for the rising toll of losses. The Red Cross-Red Crescent Society's report on 1998 disasters also concluded that total losses exceeded $90 billion in 1998, that natural disasters and environmental problems were responsible for 58% of the world's refugees and that some of the increase is due to climate change.

SIGNIFICANCE OF DISASTER TYPES

It must also be recognized that while individual earthquakes and volcanic eruptions can cause significant loss of life and local devastation, e.g., Northridge, CA, 1994, Kobe, Japan, 1995, their global total impacts are far less than that for the more numerous weather or climate related disasters; i.e., storms, floods and droughts. The World Disaster Report of the International Federation of Red Cross and Red Crescent Societies (1996) provides global statistics from 1970 to 1994. Of the 138 000 people killed each year on average, 82% were due to climate/weather disasters; of five million made homeless on average per year, 93% were from floods, droughts and storms; and of the 124 million people affected, 98% suffered through climate-related disasters. The human impact of disasters falls very unevenly on regions of the world. Some 87% of those affected (1965–1992) were in Asia, and more than 90% of deaths were in Asia and Africa. Even on a *per capita* basis, disaster sufferers per unit of population in Asia are six times that of North America and 42 times that in Europe.

On the economic side, *Information Paper 4 for the World Conference on Natural Disaster Reduction, Yokohama,* May 1994, presented a compilation of information from a number of sources, particularly the Center for Research on Epidemiology of Disasters, University of Louvain, on events which were classified as *major* disasters. Major was defined as more than 1% of annual gross national product (GNP) of country(s) affected, more than 1% of population affected, or more than 100 deaths. For events from 1963–1992 causing significant damages (more than the 1% of GNP threshold) 84% were storms floods and droughts, 10% earthquakes and 6% other disasters (landslides, volcanoes, etc.) Thus, by all criteria, on a global basis, climate related disasters have caused and continue to cause by far the most damages, suffering and deaths. However, it must be recognized that for individual disasters, the average amount of direct economic loss is greatest for earthquakes.

Another significant outcome of the analyses provided at the Yokohama Conference was that the number of disasters from 1963–1992 causing significant economic losses has risen three to five times more rapidly for climate-related disasters than for earthquakes. Could this be due entirely to some preferential increase in exposure to the risks for storms, floods and droughts than for earthquakes? This seems unlikely to be the only factor at work since some of the world's most vulnerable coastal zones with population growth are susceptible to both geological and weather/climate disasters (e.g., Japan, USA (CA)).

LAND USE CHANGES

The relationship between floods and landslides and the removal of forests and other vegetation on the slopes of watersheds is well understood. Flood peaks and landslides are substantially increased by loss of vegetation on hillsides. This is especially true in smaller watersheds. In large basins, the effect of land-cover changes tends to be masked by many other factors.

In addition, the expansion of cities with accompanying increases in paved and roofed area results in higher runoff rates and greater floods in urban and suburban areas. For example, in one study in southern Ontario, the percentage of rain and snowmelt that runs off immediately, increased from 10 to 43% with urbanization of the watershed. Thus, with rapidly growing cities, human actions are undoubtedly increasing the percentage of rain or snowmelt that runs off quickly, causing floods.

In the case of droughts, studies in the Sahelian region and elsewhere have shown that while drought periods are caused by large-scale changes in the general circulation of the atmosphere, local actions can prolong the dry periods and make their effects more devastating. Removal of vegetation for firewood, or by grazing animals, changes the albedo of the land surface and the surface roughness, both leading to prolongation of dry weather and its adverse effects. Losses of vegetation can also reduce the penetration of water into the soil, accelerating runoff and leaving little flow for the dry season. It is evident then that human actions in affecting

land uses and vegetation are already having significant effects on severity of floods and droughts.

The growth of cities results in vegetation loss and soil changes in their area of urban sprawl, and large cities can cause devastation far beyond their boundaries. For example, urban demand for fuel wood often leads to clear-cutting of nearby forests that had served local populations for generations. To illustrate, Delhi, India, needs 612 tons of firewood per day, and much of this wood now comes from forests as far as 700 km away (Kreimer and Munasinghe, 1992).

CLIMATE FACTORS

Are climate extremes becoming more frequent and severe, to at least partially explain observed disaster trends? If so, are they likely to continue to increase?

Unfortunately, understanding is far from complete of the natural variability of the climate system or of changes in the frequency of extreme events due to increasing greenhouse gases and aerosols in the atmosphere from human activity. These are matters of active research.

Recent developments in that research tend to support the view that extreme events will increase with continued human induced greenhouse gas forcing of climate and have already begun to do so in some regions. Predictions of future climate are made with global mathematical models. Most recent models have interactive atmosphere–ocean–land surfaces. They are run with realistic assumptions of continued emissions of carbon dioxide and other greenhouse gases and take into account the regional cooling effects of aerosols (especially sulfate particles). It should be noted that commitments to emission limitations in the Kyoto Protocol (1997) to the United Nations Framework Convention on Climate Change, even if fully implemented, would only slow, to a modest extent, the rate of global increase of greenhouse gas emissions. A doubling of carbon dioxide (CO_2) equivalent from preindustrial levels would probably occur in the second half of the 21st century without more vigorous controls.

Such global change models (GCMs) have been used by research groups in a number of countries to project not just the evolution of the warming trends over the coming century, but in some cases changes in extremes. When the models project a change in frequency or intensity of some type of event, researchers then ask whether recent data, with greenhouse forcing to date, show trends consistent with the model projections. Among the most studied of these trends is that for heavy rainfall intensities.

HEAVY RAINFALL INTENSITIES

A number of different GCM model runs with a doubled CO_2 atmosphere have been examined to assess projections of future frequency of heavy rainfall events. A comparison of five such model results concluded that there would be a decrease in return period (i.e., increased frequency) of extreme short period (• 1 day) rainfall events in five regions studied. Similar results were reported from the CSIRO (Australian) GCM and the UK high resolution model, which showed a mean intensity increase of 10–30% at most latitudes, in a doubled CO_2 atmosphere, but with the heaviest 10% of rain events increasing by more than 50%. The Canadian Climate Model projects that 20-year return-period heavy rainfalls will on average become 10-year rainfalls over much of the world (Zweirs and Kharin, 1998).

In a number of studies, down-scaling of the grid size from the GCMs was achieved with limited area models nested within GCMs. Several of these, for Australia, the Central Plains of North America and the Indian monsoon region, other tropical land regions, and northern mid-latitudes, project greater frequencies of heavy one-day rainfalls. The Intergovernmental Panel on Climate Change (IPCC) summarized the situation in 1995 by saying With increasing greenhouse gas concentrations, many models suggest an increase in the probability of intense precipitation. This tends to confirm the earlier IPCC (1992) conclusion that a warmer climate will be one in which the hydrologic cycle will in general be more intense leading to more heavy rain events.

Among the regionally observed trends, well confirmed by long-term reliable data, are increases in the frequency of extreme rainfall events in USA, Japan and northern Australia, and the northwest coast of India. For example, Tom Karl of US National Climate Data Center and colleagues, reported in *Nature* (Karl *et al.*, 1995) that there was a highly significant (20%) increase over the USA in the summer season in the proportion of rain occurring in heavy rain days (• 50.8 mm) over the period 1911–1992. In another paper these authors showed that the proportion of the USA that has had a much greater than normal amount of precipitation in extremely heavy one-day events (• 50.8 mm) has steadily increased over the 20th century. They calculate that this latter result has a less than 1 in 1000 chance of occurring in a quasi-stationary climate and thus is very probably a manifestation of a changing climate. (Karl *et al.*, 1996) These analysts more recently found similar but less pronounced trends over shorter period records to 1996, for China and the former Soviet Union. Thus, trends in observed rainfall data are in several regions consistent with the model projections.

Increasing frequency of high intensity rainfall would result in more flash floods, more frequent overflow of storm drainage facilities, greater soil erosion and higher risk of landslides in unstable areas. These studies tend to support the conclusions of the Red Cross-Red Crescent Society

that a changing climate is contributing to the increases in disaster losses.

OTHER EXTREME CLIMATE EVENTS

Other studies have shown that the frequency of very intense winter storms in the Northern Hemisphere would increase with greenhouse forcing of climate and have done so since the 1970s (Lambert, 1996). Hail, lightning-set wildfires, and tornadoes all are related to warmer conditions in the atmosphere's surface layers, although other conditions must be present as well. Consistent data on these small-scale phenomena are extremely difficult to obtain for trend analyses. However, their frequencies correlate positively with observed higher surface temperatures, especially higher minimum temperatures at night, a phenomenon widely observed in the warming climate of recent decades.

TROPICAL CYCLONES

For tropical cyclones, no clear trends emerge from the data, nor are they projected by models for the future, except for trends related to the El Nino Southern Oscillation (ENSO) phenomenon.

Data from the World Meteorological Organization (WMO) yield no evidence of changes in frequency and intensity on a global basis. However, the frequency of North Atlantic hurricanes declined over the period 1970–1987 while those in the western Pacific increased. These changes are much more likely to be due to the relatively intense and persistent warm El Nino events of the past two decades, than due to greenhouse forcing of climate although these two may be related. ENSO events tend to depress Atlantic hurricane activity and change the distribution of tropical cyclones in the Pacific.

Modeled future tropical cyclone activity under human-induced climate change, is also ambiguous. A comprehensive analysis by Lighthill *et al.* (1994) suggests that present climate models are inadequate to address the problem. On physical grounds, however, it has been argued that while there will likely be little change in numbers or in average intensities of tropical cyclones, it is highly probable that there would be an increase in the intensity of the most severe cyclones (Emanuel, 1996).

ENSO PHENOMENA

Large scale floods and droughts are associated with naturally oscillating modes of the climate system, the El Nino, La Nina (ENSO) events of the equatorial Pacific Ocean, the North Atlantic Oscillation and its close relative the Arctic Oscillation. The most devastating flood and drought conditions in tropical and sub-tropical regions are associated with the ENSO events. Abnormally high sea surface temperatures off the equatorial coast of western South America (El Nino) or lower than normal temperatures (La Nina) both have profound effects on global circulation patterns, but with especially strong signals in subtropical and tropical zones. El Nino conditions are typically associated with storms and floods in western South America, droughts in northeast Brazil and much of the southwest Pacific, floods in parts of Africa and southern China and severe storms in California. La Nina conditions are not a complete mirror image of this pattern but are mostly the reverse.

El Nino events have been increasingly more persistent and intense over the past 80 years and especially the past 30 years. An important question is whether that trend is a natural one, or has been forced by increased surface heating due to greenhouse gas accumulations. Some scientists claim that recent trends to 1995, are part of an ENSO cycle of 100 years or so (Rajagopalan *et al.*, 1997). On the other hand modeling work at both the Princeton Geophysical Fluid Dynamics Laboratory (Knutson and Manabe, 1998) and the Max Planck Institute, Hamburg (Timmerman *et al.*, 1999) gives strong support to the theory that greenhouse forcing has influenced the observed trend. This research also indicates that as greenhouse gas concentrations due to human activities increase over the coming decades, El Nino-like conditions are likely to be more frequent and intense, but still interspersed with La Nina periods. This in turn means that alternating flood-drought cycles would increase in severity in much of the world. These findings are consistent with a growing view among climate researchers that climate change will be manifest mainly through increased intensity of the warm phases of the natural modes of variation of the climate system, rather than through just a steady rise in mean temperatures.

SEA LEVEL RISE

Added to these effects, mean sea level, which has risen 10–20 cm in the last century, is projected to increase two to five times as rapidly this century due to continued expansion of warming ocean waters and melting of glaciers (IPCC, 1995). This will further exacerbate flooding in estuaries and increase the effects of storm surges in low-lying coastal areas. Higher storm surges with tropical cyclones can have particularly devastating effects on many small island states, and countries such as Bangladesh.

SUMMARY

This review of statistics on disasters, land use changes and climate, and projections of future climatic extremes and sea level rise, leads to the following points:

- Global losses, both human and economic, in natural disasters, mostly climate-related, have risen rapidly in the past three decades.
- While increased populations and infrastructure exposed to the risk is likely the main cause, changes in land use and climate appear to have been contributing factors.
- With population growth, land use changes which increase ood and drought impacts will continue.
- Increases in heavy rainfalls in some regions and greater frequency of other severe climatic events have been observed, are consistent with model projections of climate change, and are likely to continue to increase.
- Observed increases in frequency and intensity of El Nino conditions will probably continue with accompanying ood and drought conditions over many regions.
- Observed and projected sea level rise adds to ood and coastal storm damages.

Ironically, as the International Decade for Natural Disaster Reduction (1990–1999) came to a close, disaster loss prevention measures were ever more urgently required. Reduction of the rates of land use and climate changes through adoption of more environmentally sustainable policies could assist in slowing the trend. Disaster loss mitigation measures as an adaptation response to climate change and variability, must receive increasing support.

REFERENCES

Emanuel, K A (1995) Comments on Global Climate Change and Tropical Cyclones, Part I, *Bull. Am. Meteorol. Soc.*, 76, 700–705.

IDNDR Secretariat (1994) *Disasters Around the World - a Global and Regional View*, Inf. Paper #4 World Conference on Natural Disaster Reduction, Yokohama.

IPCC, Climate Change (1995) *Second Assessment Report*, Vol. 1–3, Cambridge University Press, Cambridge.

IPCC, Climate Change (1992) *The Supplementary Report to the IPCC Assessment* (1990), IPCC Secretariat, Geneva.

Karl, T R, Knight, R W, and Plummer, N (1995) Trends in High Frequency Climate Variability in the Twentieth Century, *Nature*, 377, 217–220.

Karl, T R, Knight, R W, Easterling, D R, and Quayle, R C (1996) Indices of Climate Change for the United States, *Bull. Am. Meteorol. Soc.*, 77(2), 279–92.

Knutson, T R and Manabe, S (1998) Model Assessment of Decadal Variability and Trends in the Tropical Pacific Ocean, *J. Clim.*, 9, 2273–2296.

Kreimer, A and Munasinghe, M (1992) *Environmental Management and Urban Vulnerability*, World Bank Discussion Paper No. 168, Washington, DC.

Lambert, S J (1996) Intense Extratropical Northern Hemisphere Winter Cyclone Events 1990–1991, *J. Geophys. Res.*, 101, 21, 219–21, 325.

Lighthill, J, Holland, G J, Gray, W M, Landsea, C, Emanuel, K, Craig, G, Evans, J, Kurihara, Y, and Guar, C P (1994) Global Climate Change and Tropical Cyclones, *Bull. Am. Meterol. Soc.*, 75, 2147–2157.

Rajagopalan, B, Lall, U, and Cane, M A (1997) Anomalous ENSO Occurrences: an Alternate View, *J. Clim.*, 10, 2351–2357.

The World Disaster Report of the International Federation of Red Cross and Red Crescent Societies (1996) Oxford University Press, London.

Timmerman, A, Oberhuber, A J, Bacher, A, Esch, M, Latif, M, and Roeckner, E (1999) Increased El Nino Frequency in a Climate Model Forced by Future Greenhouse Warming, *Nature*, 398, 694–696.

Zweirs, F W and Kharin, V V (1998) Changes in the Extremes of Climate Simulated by CCC GCM2 under CO_2 Doubling, *J. Clim.*, 11, 2200–2222.

FURTHER READING

Henderson-Sellers, A and McGuffie, K (1995) Global Climate Models and Dynamic Vegetation Changes, *Global Change Biol.*, 1.

Lambert, S J (1995) The Effect of Enhanced Greenhouse Warming on Winter Cyclone Frequencies and Strengths, *J. Clim.*, 8 1447–1452.

Global Forest Watch

Dirk Bryant

World Resource Institute, Washington, DC, USA

*Global Forest Watch is an international data and mapping network, combining on-the-ground knowledge with the power of the Internet and advanced digital technology to provide accurate information about the world's forests to a global audience. Global Forest Watch is currently working with more than 75 partners around the world to document activities inside remaining intact forests (who is behind these activities, and how well these actors are abiding by national laws) - in effect, enforcing transparency on the global forest products industry. Global Forest Watch information is available in hardcopy, as well as through the Internet at www.globalforestwatch.org. Global Forest Watch is currently housed within the World Resources Institute (WRI), but will eventually become a separate non-government organization (NGO) (*see *WRI (World Resources Institute), Volume 4).*

CURRENT EVENTS

On February 29th, 2000, Global Forest Watch was officially launched at the National Press Club in Washington, DC. At this event the first three Global Forest Watch reports, *Canada's Forests at a Crossroads: an Assessment in the Year 2000*; *An Overview of Logging in Cameroon*; and *A First Look at Logging in Gabon,* were released as well as the Global Forest Watch Website (see Figure 1 for the Global Forest Watch logo).

As of January 2001, Global Forest Watch is working in 10 countries – Cameroon, Canada, Gabon, Indonesia, Chile, Romania, Russia, Venezuela, the US, and the Democratic Republic of Congo. The Global Forest Watch network of 75 partners represents NGOs, universities, and research groups. Supported by 12 WRI staff who provide Geographic Information Systems (GIS) mapping expertise and backstopping for day-to-day operations, Global Forest Watch works closely with scientists, governments, and industry to validate results through an extensive peer review process before they are published.

During 2001, Global Forest Watch reports will be published for Chile, Indonesia, Russia, and Venezuela with the goal of working in 21 countries over the next three years, covering some 75% of the world's remaining intact forests.

MARKET IMPACT

Recent endangered forest-free purchasing commitments by major US and European wood products companies – including Home Depot, IKEA, Lowes, and others – offer a huge potential to shift the timber industry out of the world's last intact natural forests. Activist groups are pursuing an innovative strategy to protect these areas by pressuring major wood users to stop sourcing from intact forests. This approach will help build demand for well-managed certified forest products, as supplies are squeezed within intact forest areas. Several of these companies have stated they will favor supplies certified by the Forest Stewardship Council (FSC) or equivalent wood suppliers. However, there is a huge shortfall of certified wood. In addition, several companies have committed themselves to purchasing only from those suppliers in compliance with environmental laws and regulations, which will help curb demand for illegally harvested wood. To assure the protection of intact forests, demand must be shifted from cut-and-run and illegal harvesting of these areas to sustainable, certified supplies, increasingly sourced from secondary and plantation forests.

Currently, no mechanisms exist for the implementation of these commitments. Until they are put into place, these commitments will remain empty promises.

Global Forest Watch, through its international network of partners, will provide the tools and information that companies need to turn these commitments into purchasing practices. Within the next three years, with Global Forest Watch information, corporate and other users will be able to track, for the first time, where wood products originate.

Figure 1 Global Forest Watch logo

AUDIENCE AND PRODUCTS

Global Forest Watch's audience includes policy makers, corporations, civil society, and consumers. While Global Forest Watch will publish hardcopy reports documenting forest development in countries throughout the world, many products will be web-based and available through the Global Forest Watch Website. Products will include interactive and static maps, and a searchable database that will allow users to determine the name of a company that is operating in any given forest logging concession area.

GOVERNANCE STRUCTURE

Global Forest Watch is governed by an international steering committee that includes one representative per country, the Global Forest Watch director, and one representative from WRI. The steering committee provides overall guidance on Global Forest Watch's direction. The Global Forest Foundation is a separate organization, serving as a funding entity and a visionary body for Global Forest Watch.

DEVELOPMENT

Global Forest Watch grew out of WRI's 1997 publication, *The Last Frontier Forests: Ecosystems and Economies on the Edge* (Bryant *et al.*, 1997). This publication, conducted in consultation with 90 forest experts from around the world, found that only 20% of the forests that once covered the earth remain as large tracts of intact natural (frontier) forest. These areas provide islands of habitat for ora and fauna, key life support functions critical to human survival (such as storing vast amounts of

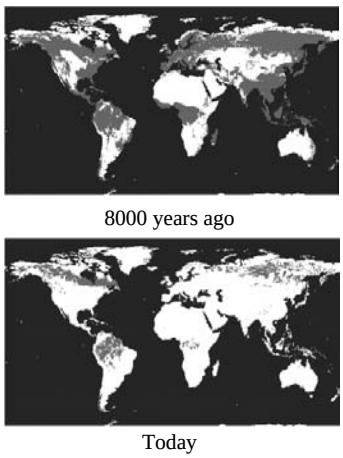


Figure 2 Frontier forest image

carbon (C) to offset the threat of global warming), and are home to the world's traditional indigenous forest peoples. However, outside of far-northern regions, three-quarters of this remaining frontier forest is threatened, and will likely be cleared or degraded in the next 10–20 years at current rates of exploitation. Unsustainable logging practices pose the primary threat to these forests. In tropical countries, much of the timber is harvested illegally (the Brazilian government estimates over 80% of logging in the Amazon Basin is illegal). In Canada and other northern countries, the remaining forest is being clear-cut to feed demands for wood products (Canada – the world's largest wood exporter – harvests 90% of its timber from old growth and primary forests) (see Figure 2).

The analysis for *The Last Frontier Forests* publication was conducted at a coarse scale (1 : 8 000 000) which provided a global perspective of the issue, but was far too coarse for far-reaching discussions about which areas within a country are under threat, and which should be off limits for development. Through this work, however, WRI became acquainted with organizations around the world that were working on forest issues in their countries. At the same time, technologies, such as satellite imagery and GIS mapping, became more available and useful for documenting development activities within forests. Coupled with an expressed interest on behalf of other environmental organizations, corporations and scientists, has led WRI to pursue its mapping to a finer scale, resulting in the creation of Global Forest Watch.

REFERENCE

Bryant, D, Nielsen, D, and Tangley, L (1997) *The Last Frontier Forests: Ecosystems and Economics on the Edge*, World Resources Institute, Washington, DC.

Global Resource Information Database (GRID)

see GRID (Global Resource Information Database, of the United Nations Environment Programme (UNEP) Division of Early Warning and Assessment) (**Volume 4**)

Goldemberg, Jose

(1928-)

Professor Goldemberg had a very uncommon evolution of his career, from purely academic research in experimental nuclear physics to general science to involvement in public affairs and then to administration. The highlight of his career was his efforts to improve support for science activities and the protection of human rights of scientists in Brazil during the period of military government (1964–1984). The future role of nuclear energy in Brazil marked his transition toward a growing interest in energy alternatives and renewables. His management skills were harnessed when he acted as the rector of the University of Sao Paulo and then as a cabinet member in the federal government. He played a very significant role in defusing the nuclear race between Argentina and Brazil.

His grandparents arrived in Brazil at the turn of the century and settled in a small town. In 1935 they moved to Porto Alegre, where public educational facilities were better. He attended a good public high school established in the previous century. The school's approach combined a strong belief in rationalism with hierarchy. There, he developed a craving to find the ultimate causes of the events around us and to understand the structure of matter.

Professor Goldemberg graduated with a Bachelor of Sciences from the Universidade de Sao Paulo, carried out graduate work at the Physics Department, University of Saskatchewan, Canada and the Physics Department, University of Illinois, USA and got his PhD in Physical Sciences from the Universidade de Sao Paulo (1954). He served as Assistant Professor, University of Toronto, Canada, and Professor of Physics, and Director of the Institute of Physics at Universidade de Sao Paulo. He served also as Associate Professor at the University of Paris

(Orsay) (1964); Senior Research Associate (1977–1982) and Visiting Professor (1993–1994) Princeton University, USA; Visiting Professor at the International Academy of the Environment, Geneva, Switzerland (1995) and President of the Brazilian Association for the Advancement of Science (1979–1981).

He received the DSc Honoris Causa from the TECHNION, Israel Institute of Technology (1991); was co-winner of the Mitchell Prize for Sustainable Development, USA (1991); he established the José Goldemberg Chair in Atmospheric Physics at Tel Aviv University in 1994. He is a member of the Brazilian Academy of Sciences and the Third World Academy of Sciences. He is the author of many technical papers and books on nuclear physics, the environment and energy in general.

His best known publications of include: *Energy for a Sustainable World*; *Energy for a Developing World*; *How to Stop Global Warming*; *Energy Needs in Developing Countries and Sustainability*; *Energy, Environment and Development*.

MOSTAFA K TOLBA *Egypt*

Goldman Prize

see International Environmental Prizes (**Volume 4**)

Goodman, Gordon

Professor Gordon Goodman has been active internationally in the fields of environment and development since the early 1960s, initially as a scientific researcher and later as the developer of environment/development institutions. In the 1960s, when he was general secretary of the British Ecological Society, he founded its Industrial Ecology Group, with an international membership, and helped to pioneer the application of ecology to the management of urban industrial environments. Under the auspices of the Scientific Committee on Problems of the Environment (SCOPE), which he helped to establish, he was the senior author and editor of a technical report for the United Nations (UN) Conference on the Human Environment: which was accepted by governments as the basis for the first Global Environmental Monitoring System (GEMS).

As Professor of Applied Biology in the University of London, Gordon Goodman was a member of UK Government environment committees on air pollution, water pollution, waste management, and environmental health. He was the founding director of the international Monitoring and Assessment Research Center (MARC) at the University of London, designed to support the UN Environment Program's Global Environment Monitoring System. He was invited to Sweden to establish the Beijer Institute (Royal Swedish Academy of Science) an international institute for energy, resources, and the human environment. Subsequently he was invited by the Government of Sweden to establish the Stockholm Environment Institute as its founding executive director and later, the chairman of its international board.

Professor Goodman has acted as an adviser on polluted land reclamation for local government and the private sector. He has also advised governments, the UN Specialized Agencies (United Nations Educational, Scientific and Cultural Organization (UNESCO), World Health Organization and World Meteorological Organization), the 1972 Stockholm Conference, the 1987 Brundtland Report, and as special scientific adviser, the 1992 Rio Conference' Preparatory Group, on problems of energy, environment and development.

His research interests include food-web chemistry (atmosphere-soil-plant-animal transfers and ecotoxicology) of environmental degradation and reclamation (techniques for restoring grass and trees to heavy-metal contaminated land in the Lower Swansea Valley, developing a moss-bag technique for the quantitative monitoring of atmospheric heavy-metal pollution in South Wales), energy risk management, global warming/climate change, and energy/resource problems of developing countries. He has authored or edited several books and written numerous scientific and technical papers on these subjects. He played an important role in developing and establishing the Rockefeller Foundation's Leadership for Environment and Development (LEAD) in 12 countries/regions worldwide and was responsible for producing its training curriculum (*see* LEAD (Leadership for Environment and Development), Volume 4). He was also the inaugural president of the International Service for the Acquisition of Agri-Biotech Applications (ISAAA) devoted to agricultural biotechnology training and genetically modified (GM) crop studies in developing countries.

Professor Goodman currently sits on the panel for the Volvo Environment Prize and is a member of the Earth Council. He is a foreign member of the Royal Swedish

Academy of Sciences (RSAS). His honors include the Older Linnaean Gold Medal of this Academy, honorary doctorates in science from the University of York, UK and Clark University, US, and Commander (First Class) of the Swedish Royal Order of the Polar Star.

MOSTAFA K TOLBA *Egypt*

Governance vis-à-vis Management

see Monitoring in Support of Policy: an Adaptive Ecosystem Approach (*Opening essay*, **Volume 4**); Governance and International Management (**Volume 5**)

Great Lakes Region of North America

see Great Lakes Region of North America (*Case study*, **Volume 4**)

Green Investment

Rory O'Brien
University of Toronto, Toronto, Canada

*Green investment refers to an economic investment strategy in which investors make corporate environmental performance the major focus in their role as shareholders in a company. Though the term is often used interchangeably with socially responsible, or ethical, investment, green investment is in effect a sub-set of it, with a speci c emphasis on investment that has mainly environmental, as opposed to social, consequences (*see *Socially Responsible Investment, Volume 4).*

Research has shown that environmentally responsible companies tend to perform better nancially than their industry peers. As a result, green investment has been growing rapidly in the last 10 years. Through the use of screening criteria, stock is purchased or sold by investors in order to push companies toward more sustainable business practices. Increasing numbers of investors are also promoting corporate change through the use of shareholder activism. A key factor in investor empowerment is standardizing and disclosing corporate performance measurements.

CORPORATE ENVIRONMENTAL BEHAVIOR AND FINANCIAL PERFORMANCE

In 1999 the Dow Jones sustainability group index (DJSGI) launched the first global stock index intended to track the performance of 200 major companies selected on the basis of high measures of environmental, social and economic sustainability. Using data on company stock price between 1994 and 2000, the DJSGI showed a small but significant out-performance of the larger Dow Jones global index (DJGI), comprised of 2000 companies throughout the world.

The DJSGI results support research done in the past decade on the relationship between a company's environmental performance and share price. Most studies have shown either a positive correlation between superior environmental management and the price of stock, or at the very least, a neutral effect, thereby putting to rest the unsubstantiated but persistent belief that investors had to accept a lower rate of return on their social investments.

This fact has not gone unnoticed by individual investors, who have been rapidly pouring money into hundreds of newly created socially responsible mutual funds and unit trusts. Of these, many are green funds, with a focus primarily on the environment. Social and environmental funds have recently proven to be strong performers – in the United States, for example, a significantly larger percentage of them was given higher ratings by three major investment tracking agencies than the rest of the mutual fund industry for the period between 1997 and 2000. These institutions, like other green investors, use social and environmental screens in their stock selection as well as screens based on financial performance.

GREEN SCREENS

In 1999, $1.2 trillion, or four-fifths of the $1.5 trillion in the United States estimated to be invested using socially responsible screening criteria, was subject to green environmental screens. In some instances, investors screen out entire industries perceived to be detrimental to the environment, such as the nuclear industry. In many cases, however, a best-of-sector approach is taken, wherein top performers in an industry, those that are exemplary in their environmental practices, are rewarded with investors' money. Green screens promote investing on the basis of positive corporate performance and deter investment in negatively performing cases.

Examples of positive screens include, among others, having enhanced pollution prevention policies and programs, environmentally friendly products and services, energy efficiency programs and alternative energy sources, and re-use and recycling of materials in operations.

Negative screening criteria include such things as repeated violations of environmental laws and regulations, hazardous waste production, nuclear energy, production of ozone depleting substances, excessive emissions of greenhouse gases, and unsustainable practices in resource industries such as deforestation and over-fishing.

A handful of firms such as Ethical Investment Research Service (EIRIS) in the UK, Investor Responsibility Research Center (IRRC) and Kinder Lydenburg and Domini (KLD) in the US, and Michael Jantzi Research Associates (MJRA) in Canada track the social and environmental activities of companies on the basis of screening criteria, maintaining extensive databases on hundreds of publicly traded corporations. Their clients include institutional investors and investment management firms, particularly those who do not maintain in-house research expertise in this area.

SHAREHOLDER ACTIVISM

Not all green investors will sell their shares in a company that is found to be harming the environment. An increasing number are becoming shareholder activists, seeking to engage corporate directors in dialogue and raise awareness of issues. In many cases, such low-key pressure, particularly from larger institutional investors such as pension funds, is sufficient to create change. In cases where corporate leaders are not amenable to such an approach, the more public route of formal submission of a shareholder resolution at the company annual general meeting is the only viable alternative. Such resolutions are growing in number, especially as environmental advocacy groups become increasingly aware of this tactic. In 1999, the environment was the most common category of shareholder proposal in the US. In the UK and Canada, however, such resolutions remain relatively rare, as the legal framework in those countries makes it much more difficult to circulate social or environmental resolutions to shareholders.

CORPORATE ENVIRONMENTAL REPORTING

To adequately evaluate corporate behaviour, green investors must have detailed and reliable sources of information. Government regulation has greatly contributed to the amount of public disclosure of corporate activities, with strict reporting on such things as toxic waste production or pollution emissions providing a basis for judging environmental performance. More progressive companies have gone beyond what is legally required and have adopted even higher standards.

Among green investors, one of the more highly regarded sets of standards is known as the Coalition for Environmentally Responsible Economies (CERES) Principles. Created in the aftermath of the disastrous Exxon Valdez oil spill in 1989, the CERES Principles were developed by a coalition of major US environmental groups and socially responsible investors, including several large public pension funds, in order to provide business with guidance on sound environmental practices. The principles, which have been endorsed by over 50 companies, among them several large multinational corporations, are as follows:

- protection of the biosphere;
- sustainable use of natural resources;
- reduction and safe disposal of wastes;
- energy conservation;
- risk reduction to employees and communities;
- safe products and services;
- environmental restoration;
- informing the public about harmful actions;
- management commitment to these principles;
- audits and reports on environmental practices made public.

A related joint initiative of the CERES coalition and the United Nations Environment Programme (UNEP) that is gaining acceptance worldwide is the global reporting initiative (GRI). The GRI is promoting the use of Sustainability Reporting Guidelines, which are intended to act as comprehensive standards for corporate environmental reporting. Universal deployment will greatly aid investors in evaluating and comparing companies on social and environmental, as well as financial, grounds.

The ISO 14000 family of standards from the International Organization for Standardization is another useful tool. In continuous development throughout the 1990s, these standards were designed to help improve a company's environmental management system, providing guidelines for monitoring, evaluating and communicating corporate environmental performance. The ISO and GRI should be of particular help to management in presenting environmental audit reports to shareholders.

Green investment grew rapidly throughout the latter half of the 1990s, and has shown itself to be both a cause and result of trends toward improved corporate environmental practices and increased public awareness of environmental issues. As more firms and investors become cognizant of the linkage between greener operations and higher share prices, we should see an increase in green investments, fueled in no small measure by enhanced corporate environmental reporting.

FURTHER READING

CERES (muzila@ceres.org) (2000) CERES Coalition for Environmentally Responsible Economies, http://www.ceres.org/. 20/10/2000.

Dow Jones (2000) Dow Jones Sustainability Group Index, http://www.sustainability-index.com/. 20/10/2000.

Kurtz, L (1997) No Effect, or no Net Effect? Studies on Socially Responsible Investing, *J. Invest.*, 6(4), 37–50.

Repetto, R and Austin, D (2000) *Pure Pro t: The Financial Implications of Environmental Performance*, World Resources Institute, Washington, DC.

Social Investment Forum (info@socialinvest.org) (1999) Social Investment Forum: Responsible Investing Trends in US, 5/11, http://www.socialinvest.org/areas/research/trends/1999-Trends.htm. 06/04/00.

Sparkes, R (1995) *The Ethical Investor*, Harper Collins, London.

GRID (Global Resource Information Database, of the United Nations Environment Programme (UNEP) Division of Early Warning and Assessment)

UNEP, as the environmental voice of the United Nations, analyzes the state of the global environment, assesses global and regional trends, and provides early-warning information on environmental threats and policy advice based on the best scientific and technological capabilities available. As part of this mandate, UNEP maintains a network of cooperating centers around the world to provide environmental data and information to enhance the scientific basis for decision-making.

GRID is a system of cooperating centers within UNEP that is dedicated to making environmental information more readily accessible to environmental analysts as well as international and national decision makers. Its mission is to provide timely and reliable geo-referenced environmental information and access to a unique international data service to help address environmental issues at global, regional and national levels in order to bridge the gap between scientific understanding of earth processes and sound management of the environment.

There are 15 GRID centers around the world, some with regional focus and others having a thematic focus. GRID centers have an active program dealing with the integration and analysis of data sets from multiple sources to derive policy relevant information about such pressing issues as the status of the world's freshwater, the loss of biodiversity, changes in land cover and population–environment interactions. Some recent activities include:

- catalyzing the development of global data sets related to land cover, river basin boundaries, topography and population distribution;
- monitoring environmental hot spots around the world using satellite data;
- early warning of emerging environmental threats in Africa;
- assessing risks and threats to human health associated with the degradation of ecosystems;
- assessing the status of the world's remaining closed forests;
- responding to environmental emergencies;
- assessing environmental vulnerability;
- assessing river basins in an integrated manner;
- developing an Internet-based project tracking and mapping system;
- the World Atlas of Desertification;
- the African Elephant Database;
- vector-borne diseases in Africa;
- assessing impacts of climate change.

Following is the list of the GRID network of centers:

GRID–Nairobi (Kenya), UNEP, Box 30552, Nairobi, Kenya, Tel.: • 254 2 62 42 14; Fax: • 254 2 62 43 15; Telex: 22068 UNEP-KE; E-mail: johnannes.akiwumi@unep.org.

GRID–Geneva (Switzerland), 11 Chemin des Anemones, CH-1219 Chatelain, Geneva, Switzerland, Tel.: • 41 22 917 82 94/95; Fax: • 41 22 917 80 29; E-mail: ron.witt@grid.unep.ch.

GRID–Bangkok (Thailand), Outreach Building, Asian Institute of Technology, PO Box 4, Klongluang, Pathumthani 12120, Thailand, Tel.: • 66 2 5162124/5245365; Fax: • 66 2 5162125/5246233; E-mail: grid@ait.ac.th.

UNEP/GRID–Sioux Falls (USA), EROS Data Center, US Geological Survey, Sioux Falls, SD 57198, USA. Tel.: • 1 605 594 6107/6117; Fax: • 1 605 594 6119; E-mail: singh@usgs.gov.

GRID-Arendal, Longum Park, PO Box 1602, Myrene, 4801 Arendal, Norway. Tel.: • 47 37 03 5650; Fax: • 47 37 03 5050; E-mail: grid@grida.no.

GRID–Budapest (Hungary), Department for Environmental Information Systems, Ministry for Environment and Regional Policy, Fo u 44-50, H-1011 Budapest I, Hungary. Tel.: • 36 1 457 33 69; Fax: • 36 1 201 43 61; E-mail: bozo@kik.ktm.hu.

GRID–Canada, Canada Centre for Remote Sensing, Geomatics Canada, 588 Booth Street, 3rd Floor, Room 335, Ottawa, Ontario, Canada K1A 0Y7. Tel.: • 1 613 947 1227; Fax: • 1 613 947 1382; E-mail: ron.brown@ccrs.nrcan.gc.ca.

GRID–Christchurch (New Zealand), c/o ICAIR, PO Box 14-199, Christchurch, New Zealand. Tel.: • 64 3 358 6995; Fax: • 64 3 358 6999; E-mail: smith@gridc.org.nz.

GRID–Denmark, Spangsbjerg Kirevej 111, DK-6700 Esbjerg, Denmark. Tel.: • 45 75 45 45 11; Fax: • 45 75 45 45 87; E-mail: grid_gnv@gridg.unep.ch.

GRID–Moscow (Russia), Federal Centre for Geoecological Systems, FCGS, State Committee for Environmental Protection, GSP, 4/6 B Gruzinskaya Street, 123812 Moscow, Russia. Tel.: (7.095) 176-7560/7520/2675; Fax: (7.095) 176-7560/7520/2675; E-mail: root@ipofcgs.msk.ru.

GRID–Kathmandu (Nepal), International Centre for Integrated Mountain Development (ICIMOD), 4/80 Jawalakhel, GPO Box 3226, Kathmandu, Nepal. Tel.: • 997 1 525316/525313 Ext. 476; Fax: • 997 1 536747/524509; E-mail: menris@icimod.org.np.

GRID–Sao Jose dos Campos (Brazil), Instituto Nacional de Pesquisas Espacias, Av. Dos Astronautas, 1758, Sao Jose dos Campos, SP, 12227-010, Brazil. Tel.: • 55 123 45 65 20; Fax: • 55 123 45 64 68; E-mail: grid@dpi.inpe.br.

GRID–Tsukuba (Japan), Centre for Global Environmental Research, National Institute for Environmental Studies, Environment Agency of Japan, 16-2, Onogawa, Tsukuba, Ibaraki 305, Japan. Tel.: • 81 298 50 2349; Fax: • 81 298 58 2645; E-mail: grid@nies.go.jp.

GRID–Warsaw (Poland), Environmental Information Centre, Ul. Merliniego 9, 02-511 Warszawa 12, Poland. Tel.: • 48 22 6274623; Tel/Fax: • 48 22 488561; E-mail: grid@gridw.pl.

GRID–Tbilisi (Georgia), 1, M. Alexidze St. VI ., #603, Institute of Geophysics, 380093, Tibilisi, Georgia. Tel.: • 995 32 942808/335514; Fax: • 995 32 942808; E-mail: z.jitcharacdze@gridb.org.

ASHBINDU SINGH *USA*

H

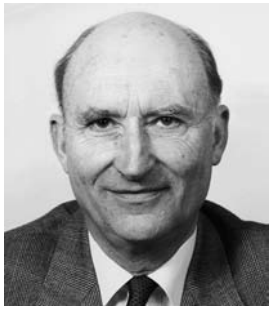
Health, Ecosystem
see Ecosystem Health (**Volume 4**)

Health, Public
see Policy Responses to Public Health Issues Relating to Global Environmental Change (*Opening essay*, **Volume 4**)

Health: the WHO Definition
see WHO (World Health Organization) (**Volume 4**)

Heirarchy Theory
see Monitoring in Support of Policy: an Adaptive Ecosystem Approach (*Opening essay*, **Volume 4**)

Holdgate, Martin



(1931-)

Sir Martin Wyatt Holdgate, CB, PhD, Fl Biol, was born on 14 January 1931, graduated from the University of Cambridge in 1952 with a 1st Class Honors BA in Natural Sciences, followed by an MA and PhD in Insect Physiology in 1955. The years 1955–1966 were spent as a research fellow, Queens College, Cambridge (1955–1956), lecturer in Zoology, Manchester University (1956–1957), lecturer in Zoology, Durham University (1957–1960), assistant director of Research, Scott Polar Research Institute, University of Cambridge (1960–1963) and chief biologist, British Antarctic Survey (1963–1966). These formative years of teaching and research provided Dr Martin Holdgate with scholarly experiences in biology, ecology of Tristan da Cunha Islands and the southernmost parts of Chile, and the biogeography of the Antarctic and Sub-Antarctic (maritime) regions, and a worldwide perception of the global environment.

During the 22 years (1966–1988) he combined and maintained the functions of a chief scientist at home and a chief architect of international endeavors in several environmental fields. At home he was deputy director (research), of the Nature Conservancy (1966–1970); director, of the Central Unit on Environmental Pollution, Department of Environment (1970–1974); director of the Institute of Terrestrial Ecology (1974–1976); director general of research and deputy secretary of the Department of Environment (1976–1981) and chief environment scientist for the Departments of Environment and of Transport (1981–1988). At the international level Dr Holdgate was a leading member of the International Council for Science (ICSU) Scientific Committee on Antarctic Research (SCAR), convened the Second SCAR Symposium on Antarctic Biology (1969) and edited its proceedings. He coordinated international programs of biological research by nations active in Antarctica and drafted the *Agreed Measures for Conservation of Antarctic Fauna and Flora* (adopted in 1974). He was also a member of the Scientific Committee for the International Biological Programme (IBP, 1964–1970), an active member of the ICSU Scientific Committee on Problems of the Environment (SCOPE) and joint author of its first global review (*Environmental Issues, SCOPE 10, 1977*). He led UK delegations to numerous international meetings preparatory to the UN Conference on Human Environment (1972), and to the Governing Council of the United Nations Environment Programme (UNEP).

Between 1983–1984 he was president of the council and part-author and joint editor of the two UNEP comprehensive reports on the state of the global environment (1982 and 1992). Between 1993 and 1996 he was a member and rapporteur of the UN Advisory Board on Sustainable Development.

From 1988 to 1994 he held the post of Director General of the World Conservation Union (IUCN), the world's largest association of states, government agencies and non-governmental organizations concerned with conservation and sustainable use of natural resources. During these years at the helm of this in uential institution, he was the principal administrator and the principal scientist, a combination very rarely attained. When he retired from this international commission he was recognized with special honor at home, being knighted. His post-retirement activities include: chairman of the UK Government Energy Advisory Panel (1993–1998), co-chairman of the UN Intergovernmental Panel on Forests (1995–1997), member of the UN Task Force on Environment and Human Settlements and president of the Zoological Society of London.

Sir Martin Holdgate is the author, co-author and/or editor of some 10 books and numerous research papers.

FURTHER READING

Holdgate, M W (1970) *Antarctic Ecology*, Academic Press, London.

Holdgate, M W (1979) *A Perspective of Environmental Pollution*, Cambridge University Press, Cambridge.

Holdgate, M W (1996) *From Care to Action: Making a Sustainable World*, Earthscan, London.

Holdgate, M W (1999) *The Green Web: a Union for World Conservation*, Earthscan, London.

MOHAMED KASSAS *Egypt*

IAI (Inter-American Institute for Global Change Research)

The IAI was created through an intergovernmental agreement signed by 16 nations of the Americas in May 1992, in Montevideo, Uruguay. Its mission is to augment the scientific capacity of the region, to better understand global change phenomena and their socio-economic implications, and to promote the dissemination of the emerging information. The Institute pursues the principles of scientific excellence, international cooperation, and the full and open exchange of scientific information relevant to global change.

To function as a regional entity and to conduct research that no one nation can undertake on its own, the IAI was designed as a network of research institutions collaborating to implement the Science Agenda (Figure 1). The IAI is governed by the Conference of the Parties, currently composed of 18 countries, which meets every year. Between conferences, the IAI's activities are directed by an Executive Council and managed by an Executive Director, whose office is located at 1758, Avenida dos Astronautas, Sao José dos Campos, SP, 12227-010, Brasil (Fax: • 55-12-341-4410, E-mail: iaibr@dir.iai.int). A Scientific Advisory Committee of 10 scientists makes recommendations regarding the Science Agenda, develops long-term plans and assesses the scientific results obtained by the Institute. The Science Agenda includes four broad themes: (1) understanding climate variability in the Americas; (2) comparative studies of ecosystems, biodiversity, land use, and water resources; (3) changes in the composition of the atmosphere, oceans and fresh waters; (4) integrated assessments, human dimensions and applications.

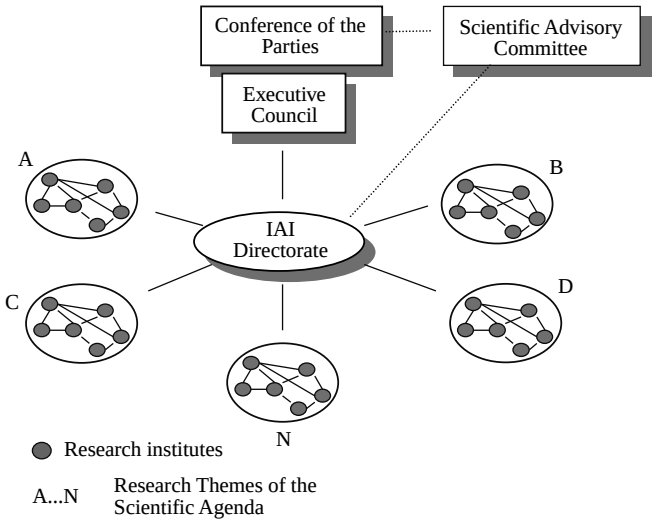


Figure 1 IAI, multinational distributed network of research institutions

IAI's strategy to implement global change research involves a three-step mechanism: (1) funding planning activities through start-up grants, specifically designed to encourage investigators to begin collaboration on proposals for long-term projects that will serve as the core for future research networks; (2) augmenting on-going activities in research, training and education, and data and information collection through initial science programs of up to 3 years duration, involving institutions from three or more member countries; (3) implementing collaborative research networks around themes of the Science Agenda by supporting up to 5-year projects involving institutions from at least four member countries. Proposals are selected based on: (1) their scientific excellence; (2) their multinational/multidisciplinary structure, and (3) the relevance of the research to on-going or future policy actions. Information emerging from IAI's research is published in scientific journals and/or presented in scientific meetings and symposia. An updated list of these publications can be found in the IAI homepage (http://www.iai.int). An IAI newsletter is also published quarterly at the Departamento de Ciencias de la Atmosfera, UBA, Pabellon II, 2do piso, Ciudad Universitaria, 1428 Buenos Aires, Argentina (Fax: • 54-11-4576-3356, E-mail: iainews@at.fcen.uba.ar).

The IAI also has a Summer Training Institute on Interdisciplinary Global Change Science located at the University of Miami, USA (e-mail: http://www.rsmas.miami.edu/IAIUM). The Institute activities are organized each year around a regionally important global change theme of

societal relevance and involve the participation of 20–25 early-career natural and social scientists of the Americas. The IAI has signed and implemented several cooperation agreements with regional and international organizations linked to global change and with National Councils for Research and Technology of IAI member countries.

See also: Inter American Institute for Global Change (IAIGC), Volume 2.

ARMANDO RABUFFETTI *Brazil*

ICREA (International Commodity-related Environmental Agreements)

ICREAs are trade agreements that internalize environmental costs within the export price. They are sometimes called *recompensating duties*, or *recompensating tariffs*.

R E MUNN *Canada*

ICSU (International Council for Science)

The ICSU is a non-governmental organization, founded in 1931 to bring together natural scientists in international scientific endeavor. ICSU has recently (April, 1998) changed its name from International Council of Scientific Unions to International Council for Science , although the acronym ICSU remains unchanged. [Throughout this encyclopedia, the term International Council for Science will be used.] Its national membership comprises 98 multidisciplinary bodies (scientific research councils or science academies) and 26 international single-discipline scientific unions, to provide a wide spectrum of scientific expertise, enabling members to address major international interdisciplinary issues which none could handle alone (ICSU also has 28 scientific associates).

The council seeks to break the barriers of specialization by initiating and coordinating major international interdisciplinary programs, and by creating interdisciplinary bodies which undertake activities and research programs of interest to several members. A number of bodies set up within ICSU also address matters of common concern to all scientists such as capacity building in science, environment and development and the freedom in the conduct of science.

The council acts as a focus for the exchange of ideas and information and the development of standards. Hundreds of congresses, symposia and other scientific meetings are organized each year around the world, and a wide range of newsletters, handbooks and journals are published.

The principal source of ICSU's finances is the contributions it receives from its members. Other sources of income are the framework contracts from the United Nations Educational, Scientific, and Cultural Organization (UNESCO), and grants and contracts from other United Nations bodies, foundations and agencies, which are used to support the scientific activities of the ICSU unions and interdisciplinary bodies. The total budget of the whole ICSU family amounts to over US$ 15 000 000 a year. One of ICSU's greatest strengths, however, is the time contributed freely by the thousands of scientists committed to the objectives of the ICSU.

The ICSU headquarters is located in Paris, France (51 Blvd de Montmorency, Paris 75016, Secretariat@icsu.org).

R E MUNN *Canada*

IDNDR (The International Decade for Natural Disaster Reduction)

In the late 1980s, the United Nations General Assembly (UNGA) was becoming increasingly alarmed about rapidly increasing human and economic impacts of natural disasters. There was also a growing awareness that much could be done to reduce losses and suffering through preventative measures (e.g., strengthened buildings, land use planning to avoid vulnerable areas), and through warning and preparedness systems. Dr Frank Press of the US National Academy of Sciences proposed that a decade be declared, to more fully use scientific and engineering knowledge to reduce losses worldwide. In 1989 the UNGA adopted a resolution declaring the 1990s the IDNDR. The word loss should have been present between disaster and reduction to avoid future confusion, but it would have made the designation even longer than IDNDR.

By some measures, the IDNDR was a failure. Disaster losses increased throughout the Decade. Economic losses have increased some 10 times between the 1960s and the

late 1990s. However, loss of life, in the industrialized world did not increase, probably due to improved warning and preparedness systems during the Decade. Numbers of people affected averaged more than 130 million per year between 1971 and 1996. The Red Cross–Red Crescent Society reported that 1998 global losses exceeded $90 billion and that 58% of refugees in the year were due to environmental and natural disasters. The experience of disaster loss increases is attributable to population increases, increased vulnerability due to occupation of hazardous areas, increased infrastructure at risk, land use change and climate changes (*see* Global Environmental Change and Natural Disasters, Volume 4).

On the other hand, it may be useful to examine whether during the 1990s decade, some mechanisms were put in place that could limit future losses.

In the UNGA Resolution of 1989, a Scientific and Technical Committee (STC) was called for, as well as a Secretariat. All costs were to be supported by voluntary contributions mainly from United Nations (UN) member countries. From the beginning, the UN lodged the secretariat in a succession of agencies responsible mainly for relief and response rather than those that would be concerned with prevention as a measure towards sustainable economic development. This, in part, led to lack of resource commitment by those UN agencies that housed' the secretariat and a general unwillingness on the part of donor countries to make the expected substantial contributions. Some key countries, especially Japan and Germany, UK and Nordic countries, made modest contributions.

The STC in the early part of the decade defined three main targets to guide efforts and as a means of assessing progress. They called for each country to complete by 2000:

1. a comprehensive risk assessment;
2. mitigation measures addressing the risks built into sustainable development planning;
3. access to global, regional and local warnings, and preparedness systems in place.

In addition the STC, which met at least annually, and the secretariat put strong emphasis on establishing national committees or key focal points in each country to achieve these targets; these were formed in more than 100 countries. A major successful effort in securing political commitment for disaster loss prevention was through the 1994 UN World Conference on Natural Disaster Reduction, Yokohama, Japan. The scientific community, mobilized mainly by the World Meteorological Organization, UN Educational, Scientific and Cultural Organization and the non-governmental International Council for Science, undertook many valuable decade projects to develop new capabilities and to apply existing knowledge to disaster loss mitigation.

At the IDNDR Programme Forum, Geneva, July 1999, many countries, international and regional organizations presented success stories. Measured against the targets, how much progress was made?

1. Risk is a product of two factors, probability of the hazard, and vulnerability to the hazard. *Hazard assessment* for all types of natural hazards, climatic, hydrologic and geological, was carried out effectively in many countries, and regionally in several cases (e.g., North America, Mexico, USA and Canada). However, *vulnerability assessment* proved much more difficult with adequate techniques still under development.
2. Mitigation, or pre-disaster preventive and preparedness actions, increased. For example, USA is now committed to provide 15% of disaster loss payouts to such actions. Some other countries have undertaken similar steps. The UN Development Programme and World Bank both established units with a disaster loss prevention responsibility.
3. Warning and preparedness systems were established or strengthened for cyclones in the SW Pacific region, SW Indian Ocean and the Caribbean and for 10 volcanoes. Warnings of eruption of Mt Pinatubo in the Philippines saved many lives.

In addition, the countries of the world signed several important Conventions during the 1990s which when implemented, will reduce disaster losses. These include:

1. The UN Framework Convention on Climate Change which, in requiring adaptation to climate extremes, could address the majority of disaster losses;
2. The Convention on Biological Diversity which in protecting forests and coral reef ecosystems, will also reduce disaster impacts, and;
3. The Convention on Desertification and Drought to help reverse drought disaster impacts especially in Africa.

Thus, for the new millennium, there have been put in place during the IDNDR, at least partly due to the Decade, a number of mechanisms that should assist in reducing natural disaster impacts. At the same time, it must be recognized that human populations and megacities are likely to continue to become more vulnerable (currently an estimated one billion people live in temporary towns at the outskirts of cities). Hazards are likely to continue to increase due to global changes. It will be a major challenge to minimize human and economic losses in future; vigorous efforts are increasingly needed in the post-Decade era.

JIM BRUCE *Canada*

IDWSSD (International Drinking Water Supply and Sanitation Decade)

The Water Decade, officially called the IDWSSD, was conceived during the United Nations Water Conference at Mar del Plata in 1977, and was launched at the United Nations General Assembly in New York during November 1980, at the recommendation of the World Health Organization (WHO). The basic goal of the Water Decade was safe water and sanitation for all. This was an ambitious goal, even if the most basic services for safe water and sanitation were to be provided. In 1980, less than 30% of people in rural areas of developing countries had access to safe potable water supplies (Traoré, 1992, 3.9). In contrast, in urban areas the comparable number was just under 80%. For sanitation facilities, some 35% of rural people and 66% of urban people were satisfactorily serviced. As a result of this situation, the WHO had estimated that diarrhea was causing the death of six million children each year, and that parasitic worms infected nearly 50% of the entire population of developing countries (Traoré, 1992, 3.7).

While these figures indicated that the need to improve drinking water supplies and sanitation was a huge task, they did not reveal the even greater disparities existing in the least developed nations, or the needs of the poorest people in such countries. On a regional basis, the need was greatest for the rural people in countries of Western Asia, where, on average, only 22% of the population was serviced, and in Africa south of the Sahara where 23% of the rural population had access to water supply services. For Africa as a whole, 30% of the people had access to safe water. Regarding sanitation, the poorest levels of service occurred in Southeast Asia and in the eastern Mediterranean region. These patterns indicated that, in 1980, about 1.65 billion rural people at a global scale did not have access to a safe water supply, and about 1.47 billion did not have access to safe sanitation for disposal of human wastes.

The overriding motivation of the Water Decade was to improve health and therefore overall living standards. Indeed, Barbara Ward had stated that the number of drinking water standpipes per person was perhaps the best indicator of quality of life of rural people in developing countries. For her, provision of safe potable water and sanitation facilities was the one initiative which could have the greatest impact at a global scale to improve overall well being. By the end of the Water Decade, significant progress had been made, with 60% more people having access to a safe water supply source compared to the 1970s, and 140% more having access to satisfactory sanitation. In absolute terms, these percentages translate into over 1.5 billion additional people being provided with safe water supplies and 750 million people receiving access to safe sanitation facilities during the 1980s. However, despite these improvements, over one billion people were still without safe water supply and nearly 1.8 billion people did not have adequate sanitation at the end of the water decade (World Health Organization, 1991).

The main explanation for the shortfall was that during the 1980s the global population increased by about 750 million people. As a result, in many regions of the world, but especially in rural areas of developing countries, basic human needs for water use were not being satisfactorily met by the end of the Water Decade. Prospects for the future are not encouraging. For example, population projections at the beginning of the 1990s showed that in the following decade the global population was expected to increase by almost 900 million people. By late 1999, the human population on the Earth had been officially documented by the United Nations as having reached six billion people.

REFERENCES

Traoré, A (1992) Water for the People – Community Water Supply and Sanitation, in *International Conference on Water and Environment: Keynote Papers*, ICWE Secretariat, World Meteorological Organization, Geneva, 3.1–3.22.

World Health Organization (1991) *The International Drinking Water Supply and Sanitation Decade: Review of Decade Progress (as at December 1990)*, World Health Organization Division of Environmental Health, Geneva.

BRUCE MITCHELL *Canada*

IFIAS (International Federation of Institutes for Advanced Study)

IFIAS is an international grouping of leading research institutes, originally established in 1972 on the initiatives of the Rockefeller and Nobel Foundations. It was part of a ourishing of innovative institutions in that period, which included international think tanks, new cooperative ventures in the sciences and social sciences, as well as various institutes for advanced study. IFIAS was designed to pool the resources of the international scientific research community to examine long-term, cross-disciplinary, and controversial issues.

Originally situated in Sweden, IFIAS moved its headquarters to Toronto, Canada in 1986, and is currently hosted by the Institute for Environmental Studies, University of

Toronto. At various periods in its history, it has had offices in Washington, DC and Maastricht, in the Netherlands. Its membership has ranged from 20–40 institutes at various times, covering a wide range of advanced research in everything from ekistics (i.e., the study of building form), to coastal waters management.

Because of its mandate, IFIAS has often been at the forefront of identifying issues that have subsequently become more widely appreciated. In the 1970s, IFIAS worked on early climate change research (contributing to the basic work of the World Climate Programme); energy and self-reliant development at the local village level in developing countries; hazards research including the dynamics of drought; biotechnology; and Islam and the West. IFIAS also took a lead in a number of large-scale international probes, such as the Analyzing Biospheric Change program, initiated in 1981, which examined a whole range of rapid changes, including urbanization, agricultural degradation, and coastal waters pollution. Project Ecoville, in particular, was a path-breaking attempt to apply systems thinking to rapid urbanization around the world.

In the late 1980s and 1990s, IFIAS was one of the originating partners in the founding of the International Human Dimensions of Global Change Programme (HDGCP) along with the International Social Science Council, the United Nations University, and the United Nations Educational, Scientific and Cultural Organization. This was designed to be a social science complement to the International Geosphere–Biosphere Programme; and the HDGCP has undergone various transformations and name changes since that period – now the International Human Dimensions Programme.

As with many other international collaborative scientific institutions, IFIAS has struggled to maintain its activities and presence in an era of cutbacks and the loss of the initial visionary impetus that accompanied the founding of so many innovative networks in the 1960s and 1970s. In spite of this, IFIAS's more recent activities and initiatives have included work on technological change and innovation, an international coastal cities initiative (a project of the Institute for Environmental Studies at the University of Toronto); the Gender, Science, and Development Programme which contributed significantly to the Beijing Women's Conference; and continuing work on the future of international science.

The history of IFIAS can be studied in two works: Stahle *et al.* (1988) and Timmerman (1999). Further information concerning IFIAS can be obtained from http://www.ifias.ca.

REFERENCES

Stahle, N H K, Nilsson, S, and Lindblom, P (1988) *From Vision to Action: Science and Global Development*, IFIAS, Toronto, Canada.

Timmerman, P (1999) *The Place of IFIAS in the World: Retrospect and Prospect*, IFIAS, Toronto, Canada.

PETER TIMMERMAN *Canada*

IHDP (International Human Dimensions Programme on Global Change)

The IHDP is an international, interdisciplinary. social science program to promote and coordinate research aimed at describing, analyzing and understanding the human dimensions of global environmental change (GEC), i.e., the way people and societies (a) contribute to GEC; (b) are in u-enced by GEC; and (c) mitigate and adapt to GEC.

In 1996, the International Social Science Council was joined by the International Council for Science as co-sponsors of the IHDP. While strengthening the IHDP itself, this also facilitated collaboration with the International Geosphere–Biosphere Program (IGBP) and the World Climate Research Program. The IHDP established a strengthened secretariat in Bonn, Germany, and is guided by a Scientific Committee representing different geographic and scientific backgrounds. National Human Dimensions Committees are an essential component of the IHDP's research strategy emphasizing a bottom-up approach. The IHDP collaborates with over 30 of these Committees, and communicates with them through a quarterly newsletter, and a web page.

For the IHDP, the human dimensions of GEC encompass the full range of social and natural sciences disciplines necessary to analyze and understand people's role as both the possible cause and target of GEC. IHDP Science Projects are a key mechanism used to generate IHDP research activities in priority areas, to promote international collaboration of researchers, and to link policy makers and researchers. Projects evolve from proposals by the Scientific Committee or via proposals to the IHDP made directly by one or more National Human Dimensions Committees or programs. There are presently four IHDP Science Projects:

- Land Use and Land Cover Change (co-sponsored with IGBP); *see* IGBP Core Projects, Volume 2;
- GEC and Human Security;
- Institutional Dimensions of GEC;
- Industrial Transformation.

RAMINE V SHAW *Germany*

IIASA (International Institute for Applied Systems Analysis)

The IIASA, located in Laxenburg, Austria, is an independent, non-governmental, interdisciplinary research institution that specializes in natural and social scientific research methods and models for use by policy makers, the scientific community, and the public worldwide. The Institute currently concentrates its research efforts within the following core themes: Environment and Natural Resources, Energy and Technology, and Population and Society. IIASA is renowned for its expertise in transboundary air pollution, energy management and conservation, land use, forest resources management, water management, technological evolution, population dynamics, risk analysis and economic transition. Continued concerns about human welfare have added two related areas to the research program: the future of social security systems and implications of future catastrophic events on society and its institutions. The Institute has also made major inroads in advancing the theories and methodologies used to tackle these applied and policy-relevant issues, including the construction of large databases and sophisticated models. For example, several of the Institute's databases, such as those on carbon dioxide (CO_2) mitigation technologies, European forest resources, global climate and land use and land cover, are among the most comprehensive in the world. Governments and private institutions in Europe, Asia, and the Americas draw upon these data resources, as well as on IIASA-developed software, for modeling such issues as transboundary air pollution and water quality management. Since its inception in 1972, IIASA has fulfilled its strategic goal: to conduct international and interdisciplinary scientific studies to provide timely and relevant information and options, addressing critical issues of global environmental, economic, and social change, for the benefit of the public, the scientific community and national and international institutions. After three decades, IIASA continues as a center for innovative research; international, interdisciplinary collaboration; conferences and workshops; and networking with other organizations of scientists worldwide. IIASA's funding comes from sponsoring member organizations in 15 countries throughout North America, Europe and Asia. In addition, a valuable source of funding and direction for IIASA's research projects comes from various grants, sponsorships, or partnerships in both member and non-member countries. Governments throughout the world must address unprecedented changes in the world's social, economic, and environmental conditions. To fashion effective policies, national leaders must understand the complex problems associated with global change and recognize their interrelationships. Providing this insight is precisely the mission of IIASA.

For further information, email: volker@iiasa.ac.at.

GORDON MACDONALD *Austria*

IMAGE

see Integrated Assessment (**Volume 4**)

IMO (International Maritime Organization)

IMO is a specialized agency of the United Nations (UN) with headquarters in London, UK. It was established in 1948 as the Inter-governmental Maritime Consultative Organization (IMCO). Its name was changed to the IMO in May 1982, following an amendment of its convention.

IMO consists of the assembly, the council, the maritime safety committee, the marine environment protection committee, the legal committee, the technical cooperation committee, the facilitation committee, and a number of sub-committees. It has a secretariat headed by a secretary general.

IMO's principal objective is to

> Encourage the general adoption of the highest practicable standards in matters concerning maritime safety, efficiency of navigation and the prevention and control of marine pollution from ships; and to deal with legal matters related to (these) purposes.

One of its functions is to provide for the drafting of conventions, agreements, or other suitable instruments •••.

IMO has adopted many treaties and other instruments on maritime safety, prevention and control of marine pollution from ships, and liability, and on compensation for vessel-source pollution. Among these are conventions on safety of life at sea; prevention of collisions at sea; training and certification of seafarers; prevention of marine pollution from ships; intervention in cases of marine pollution casualties; civil liability for environmental damage caused by ships; and international compensation funds for pollution damage. Non-treaty instruments include codes for the construction and equipment of ships, international code of maritime signals and traffic separation schemes.

IMO performs secretariat and other functions in respect of the 1972 London Dumping Convention. It is recognized

by the UN Convention for the Law of the Sea as the competent international organization on navigation, including pollution from vessels and by dumping.

Headquarters: 4 Albert Embankment, London SE1 7SR. Tel.: • 44-(0)20-735-7611.

THOMAS A MENSAH *UK*

Indicators on Environmental Quality and Sustainable Development

Allen L Hammond
World Resources Institute, Washington, DC, USA

Indicators are quantitative measures that simplify and help interpret large amounts of data and analysis. For example, the performance of a national economy is usually summarized by indicators such as the gross domestic product (GDP), the in ation rate, and the unemployment rate. Indicators such as the Dow Jones industrial average or the S & P 500 are often reported to summarize the performance of the US stock market. In a complex age, indicators perform an essential communication function. They help decision-makers and ordinary citizens alike to quickly grasp important trends.

So far, however, there are very few environmental indicators in widespread use, especially indicators that would summarize at a national scale whether environmental conditions were getting better or worse. Moreover, there is growing interest in indicators that could summarize progress (or the lack of progress) towards sustainable development, encompassing national performance in economic, environmental, and social conditions.

Such interest is motivated in part by the belief that indicators are crucial to informed decision-making. In the business world, this attitude is often expressed by the adage, " If you can't measure it, you can't manage it." Without appropriate indicators, decision-makers may not focus adequately on environmental or social problems, and public awareness may not generate the pressure that, in democracies, is often crucial for signi cant policy changes.

ENVIRONMENTAL INDICATORS

Over the decade of the 1990s, efforts to develop environmental indicators have proceeded on many different fronts. Work by Environment Canada led to a system of indicators describing environmental pressures such as pollution, the state or condition of the environment, and the social or institutional response to the problem (Environment Canada, 1991). This pressure/state/response framework was applied to a number of specific environmental issues in environmental indicators developed and reported by the Organization for Economic Cooperation and Development (OECD) for its member states (OECD, 1994). The government of the Netherlands developed an integrated set of environmental pressure indicators that could be aggregated or combined to give a single, summary indicator that it reports publicly on a periodic basis (Government of the Netherlands, 1991; Adriaanse, 1993). The US Environmental Protection Agency developed an indicator of toxic releases that aggregates data from more than 100 different chemicals and tens of thousands of industrial facilities (EPA, 1993).

Publication of these indicators has sometimes led to significant changes in policy or behavior. In the Netherlands, for example, publication led to new and stricter regulation of acidifying gases and to voluntary compacts for reduction of toxic pollution with industry. In the US, the Toxic Release Inventory stimulated voluntary commitments to drastic reductions in toxic pollution by many major chemical companies.

The concepts of a framework that shapes a system of indicators and of aggregation to summarize or simplify a set of indicators into a single measure that is more easily communicated are widely used in the indicator literature.

More recent work on environmental indicators has explored additional approaches. Research at the University of British Columbia led to an indicator of the environmental footprint of consumption patterns, measured in terms of the amount of land that would be required to support that consumption (*see* Ecological Capacity, Volume 3). A detailed analysis of the ow of all materials through industrial economies by the Wuppertal Institute and an international collaboration of research institutes led to a number of aggregated indicators – such as total materials requirement and total pollution output – to describe and summarize those ows and their potential environmental consequences (Adriaanse, 1997). Similar materials ow studies have now been completed for many OECD countries; the resulting indicators are sometimes described as *throughput measures*.

In addition to statistical indicators, work by the World Resources Institute and its collaborators developed spatial indicators (digital maps) of the state and the stresses or potential degradation of coral reefs and other ecosystems (Byrant *et al.*, 1998). Such spatial indicators show the patterns of environmental threats across an ecosystem as well as giving quantitative summary measures such as the percent of the habitat at high risk of degradation.

Although adoption of environmental indicators is proceeding slowly, there is some evidence of progress

and a growing richness of analytical work in indicator development.

SUSTAINABLE DEVELOPMENT INDICATORS

Efforts to develop not just environmental indicators but more comprehensive measures of sustainable development are spreading, although relatively few indicators of this type have been published or widely adopted. One impetus for such efforts has been the UN Commission on Sustainable Development, which has actively encouraged the development and testing of indicators. A second impetus is the growing conviction among some researchers that existing indicators of economic growth such as the GDP give a distorted sense of progress – that it is not more economic output but rather increased human welfare that is essential to sustainable development.

One approach to sustainable development indicators has been to modify economic indicators or to group existing economic indicators with environmental and social indicators. The World Bank, for example, has developed and calculated for most nations an experimental indicator of economic, natural, and human capital, all measured in economic terms. Supplementing this wealth indicator is a corrected GDP measure called *genuine savings* that takes into account such dissavings as degradation of natural resources and pollution clean-up expenditures while adding in as investments such things as educational expenditures (World Bank, 1996). An informal group of indicator experts is exploring an approach that would cluster several indicators for each of economic, environmental, and social aspects of sustainability (*see* Natural Capital, Volume 2).

A second approach is to develop new frameworks – and perhaps radically new indicators – that offer fresh insights into what sustainable societies might require. One example of this approach was developed by an interagency task force within the US government. It incorporates such concepts as *endowments* (the inherited stock of natural resources, built capital, and human capabilities) and processes (from industrial processes that transform matter to educational processes that increase human capacity) as well as a tentative list of indicators (US Government, 1998).

Progress towards sustainable development indicators remains slow (Moldan and Billharz, 1997). Part of the difficulty is that the underlying concept of sustainable development lends itself to many different interpretations, and may well be dependent on the state of development of a country as well as being scale-dependent. Indicators useful for industrialized countries, for example, may not be appropriate for developing countries still highly dependent on natural resources; measures applicable at a continental scale may bear little relation to those appropriate for an urban area or a rural community. Furthermore, there are unresolved basic measurement issues: practical ways to measure the social capital of a society – for example, the strength and flexibility of its institutions or the richness of informal social connections or the prevalence of corruption – do not exist. As a result, an extended period of experimentation and research may be necessary before widespread adoption of sustainable development indicators takes place.

REFERENCES

Adriaanse, A (1993) *Policy Performance Indicators*, SDU, The Hague.

Adriaanse, A (1997) *Resource Flows: The Materials Basis of Industrial Economies*, World Resources Institute, Washington, DC.

Bryant, D, Burke, L, McManus, J W, and Spalding, M (1998) *Reefs at Risk: a Map-based Indicator of Potential Threats to the World's Coral Reefs*, World Resources Institute, Washington, DC.

Environment Canada (1991) *A Report on Canada's Progress Toward a National Set of Environmental Indicators*, State of the Environment Report No. 91–1.

Environmental Protection Agency (EPA) (1993) *1991 Toxics Release Inventory*, EPA, Office of Pollution Prevention and Toxics, Washington, DC.

Government of the Netherlands (1991) *National Environmental Program, 92- 95*, Second Chamber of Parliament, The Hague.

Moldan, B and Billharz, S, eds (1997) *Sustainability Indicators, SCOPE 58*, John Wiley & Sons, Chichester.

Organization of Economic Cooperation and Development, OECD (1994) *Environmental Indicators*, Paris.

The World Bank (1996) *Monitoring Environmental Progress*, The World Bank, Washington, DC.

US Government (1998) *Sustainable Development in the United States: an Experimental Set of Indicators*, US Interagency Working Group on Sustainable Development Indicators.

Indus River Basin, Water Management Issues

see Indus Basin: a Case Study in Water Management (**Volume 3**)

Industrial Ecology

Industrial ecology (IE) is an evolving field of study that deals with materials and energy flows and transformations, both within an economic system and also across its

boundaries with the natural world. The term was first used in Japan (1972) but was revived in a well-known article by Frosch and Gallopoulos (1989) as an intentional metaphor to suggest an industrial system of diverse specialized firms that imitates a natural ecosystem to the extent of utilizing material wastes as a natural ecosystem recycles nutrients. IE, which emphasizes transactional relationships among production units (firms) can be distinguished from *industrial metabolism* (IM), which emphasizes materials/energy flows and transformations (Ayres, 1989). However, there is considerable overlap and the two concepts are used somewhat interchangeably. Other related concepts include *life-cycle analysis* (LCA) (sometimes called a *life cycle assessment*), *mass flow analysis* (MFA), *substance flow analysis* (SFA), *recycling*, *remanufacturing*, and *dematerialization*. In reality the IE metaphor is flawed. A natural ecosystem depends upon primary producers, which are photosynthetic green plants. In addition, it includes a hierarchy of animals and microorganisms: plant-eaters (grazers), several trophic levels of carnivores and omnivores, plus a variety of parasites and decay organisms. By contrast, the current industrial system has two kinds of primary producers. One category utilizes organic inputs from agriculture and forestry. The other category of producers utilizes abiotic inputs: minerals, metals and fossil fuels. There is a superficial resemblance to the food chain, in that firms higher on the chain utilize the products of firms lower on the chain.

A natural ecosystem, taken as a whole, generates neither products nor wastes. It merely endures and provides a stable microenvironment that permits its components to reproduce themselves and thus the system. It exchanges raw materials on a continuing basis with natural reservoirs (air, water and soil).

Reprocessors and recyclers in the industrial system are only roughly analogous to parasites and/or decay organisms. In the natural system their function is preprogrammed (behavior). In the industrial ecosystem, by contrast, their role, if any, is determined by economic incentives (prices, demand) and regulations. The industrial system differs fundamentally from an ecosystem in that the former depends essentially on exploiting natural capital (minerals and fossil fuels) put in place eons ago, whereas an ecosystem utilizes only current or recent primary production.

REFERENCES

Ayres, R U (1989) Industrial Metabolism, in *Technology & Environment*, eds J Ausubel and H E Sladovich, National Academy Press, Washington, DC.

Frosch, R A and Gallopoulos, N E (1989) Strategies for Manufacturing, *Sci. Am.*, 261(3), 94–102.

ROBERT U AYRES *France*

Industrial Responses to Stratospheric Ozone Depletion and Lessons for Global Climate Change

see Industrial Responses to Stratospheric Ozone Depletion and Lessons for Global Climate Change (*Opening essay*, **Volume 4**)

Integrated Assessment, Definition of

Integrated assessment (IA) is a reflective and iterative participatory process that links knowledge (science) and action (policy) regarding complex global change issues such as acidification and climate change. IA can be defined as an interdisciplinary process of combining, interpreting and communicating knowledge from diverse scientific disciplines in such a way that the whole cause–effect chain of a problem can be evaluated from a synoptic perspective with two characteristics: (i) it should have added value compared to single disciplinary assessment; and (ii) it should provide useful information to decision makers (Rotmans and Dowlatabadi, 1997).

Assessment comprises the analysis and review of information derived from research for the purpose of helping someone in a position of responsibility to evaluate possible actions, or think about a problem. Assessment means assembling, summarizing, organizing, interpreting, and possibly reconciling pieces of existing knowledge, and communicating them so that they are relevant and helpful to an intelligent but inexpert decision-maker (Parson, 1995).

The cause effect chains that IA aims to evaluate start with socio-economic drivers, leading to economic activity and other practices, leading to emissions and other pressure on the environment, leading to environmental changes, leading to physical impacts on societies and ecosystems, leading to socioeconomic impacts, eventually returning to change the socioeconomic drivers. Therefore, IA needs to integrate insights from a multitude of disciplines to arrive at a synoptic view on the problem at hand.

See also: Integrated Assessment Models of Global Change, Volume 4.

REFERENCES

Parson, E A (1995) Integrated Assessment and Environmental Policy Making, in Pursuit of Usefulness, *Energy Policy*, 23(4/5), 463–476.

Rotmans, J and Dowlatabadi, H (1997) Integrated Assessment Modeling, in *Human Choice and Climate Change*, Vol. 3, eds S Rayner and E L Malone, Battle Press, Columbus, OH, 291–377.

JEROEN P VAN DER SLUIJS *The Netherlands*

Integrated Assessment

Jeroen P van der Sluijs
Utrecht University, Utrecht, The Netherlands

Integrated assessment (IA) can be de ned as an interdisciplinary process of combining, interpreting and communicating knowledge from diverse scienti c disciplines in such a way that the whole cause- effect chain of a problem can be evaluated from a synoptic perspective with two characteristics: (i) it should have added value compared to single disciplinary assessment; and (ii) it should provide useful information to decision makers (Rotmans and Dowlatabadi, 1997). Thus (IA) is an iterative participatory process that links knowledge (science) and action (policy) regarding complex global change issues such as acidi cation and climate change.

Assessment comprises the analysis and review of information derived from research, for the purpose of helping someone in a position of responsibility to evaluate possible actions or think about a problem. In this context, assessment thus means assembling, summarizing, organizing, interpreting, and possibly reconciling pieces of existing knowledge, and communicating them so that they are relevant and helpful to an intelligent but inexpert decision-maker (Parson, 1995).

Providing such a synoptic view required new analytical tools and procedures to integrate a wide range of disciplinary knowledge. As Parson (1995) phrased it: " to make rational, informed social decisions on such complex, long-term, uncertain issues as global climate change, the capacity to integrate, reconcile, organize, and communicate knowledge across domains - to do integrated assessment - is essential." IA has emerged as an approach that is suitable for accommodating the uncertainties, complexities and value diversities of issues such as climate change.

The cause–effect chains that IA aims to evaluate start with socioeconomic drivers leading to economic activity and other practices, leading to stresses on the environment, leading to environmental changes, leading to physical impacts on societies and ecosystems, leading to socioeconomic impacts, eventually returning to cause changes in the socioeconomic drivers. Therefore, IA needs to integrate insights from a multitude of disciplines and stakeholders to arrive at a synoptic view of the problem at hand.

One of the consequences of the novelty of the field of IA is that the literature does not yet show a uniformity in terminology. Even the appropriate term to indicate the research field is subject to debate. Some groups talk about *integrated environmental assessment* (IEA), whereas others use the term IA. The proponents of the term IA argue that adding environmental as the focus detracts from the integrated nature that IA should have. On the other hand, IA makes no reference to complex environmental policy problems, whereas until now, the assessment of complex environmental policy issues has been both the incentive for the development of, and the major concern of IA. There is, however, no *a priori* reason to restrict the application of IA to environmental policy problems. It is obvious that the IA research community has the aspiration that IA be applicable to the entire field of science for policy. IEA can then be seen as the subset of IAs concerned with environmental issues.

In the literature a large number of attempts to define IA can be found. Weyant *et al.* (1996); Van der Sluijs (1997) and Toth and Hizsnyik (1998), give reviews of IA definitions. This multitude of attempts to define and redefine IA re ects the novelty and dynamics of the field. For the purpose of this article we adopt the widely cited Rotmans and Dowlatabadi (1997) definition, describing IA as:

> an interdisciplinary process of combining, interpreting and communicating knowledge from diverse scientific disciplines, in such a way that the whole set of cause–effect interactions of a problem can be evaluated from a synoptic perspective with two characteristics: (i) it should have added value compared to single disciplinary assessment; and (ii) it should provide useful information to decision makers.

To understand the concept of IA, it is useful to distinguish *doing assessment* from *doing research*. Assessment comprises the analysis and review of information derived from research, for the purpose of helping someone in a position of responsibility to evaluate possible actions or think about a problem. Assessment usually does not mean doing new research. Assessment means assembling, summarizing, organizing, interpreting, and possibly reconciling pieces of existing knowledge, and communicating them so that they are relevant and helpful to an intelligent but inexpert policy-maker or other actor involved in the problem at hand (Parson, 1995).

IA addresses three goals (Weyant *et al.*, 1996):

- coordinated exploration of possible future trajectories of human and natural systems;

- development of insights into key questions of policy formulation;
- prioritization of research needs in order to enhance our ability to identify robust policy options.

The IA process overlaps with the policy development process, as is shown in Figure 1.

An important notion in the definition given above is that IA is understood as a process and not a method. In the early period of the development of the field (that is from the early 1980s till the mid 1990s), the label IA was often used to refer to the technical analytical methods used in the process, with integrated assessment models (IAMs) and scenario development as the dominant tools. Nowadays, it is widely recognized that IAM is not a complete IA methodology, it is one set of analytical tools for integration used in conjunction with other methods of integration in a broader participatory assessment process.

Historically, IA is rooted in scientific and public policy efforts to understand and control acid deposition in the 1970s in Europe and North America. Public policy organizations had asked the scientific community to prepare IAs of the acidification problem (Toth and Hizsnyik, 1998). In practice, traditional forms of knowledge integration, such as books, expert panels and advisory bodies combining monodisciplinary assessments were too slow and too in exible to fulfill all the needs of the dynamic and issue-driven policy process. Examples of policy needs where traditional means of integration were weak, are the *ex-ante* evaluation and comparison of the potential impacts of a wide range of policy scenarios, and estimates of the effectiveness of alternative (combinations of) conceivable policy measures. Facilitated by developments in computer technology and systems analysis, IA modeling emerged in the 1980s as a new methodology that was capable of addressing such questions in a fast, exible and consistent way.

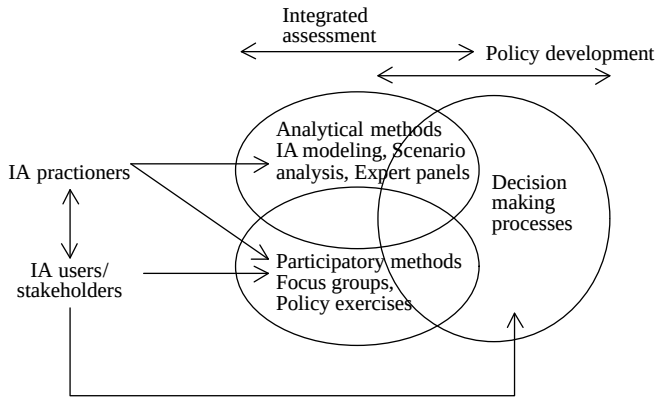


Figure 1 IA involves practitioners and users, uses technical analytical methods and participatory methods and overlaps with the policy development process

IAMs are computer simulation models in which knowledge from many different disciplines is combined in an analytical computational framework to analyze the problem at hand in an integrated fashion. IAMs are used for several purposes, such as scenario analysis, (*ex-ante*) evaluation of the environmental, economic and social consequences of different policy strategies, translation of environmental quality standards into acceptable emission trajectories (*safe landing corridors* or *tolerable windows*), and optimization of key policy variables such as rate of greenhouse gas emission reduction or level of carbon taxes.

Within IAMs, both the natural system and the socioeconomic system are simulated. A perfect policy evaluation IAM would model the complete so-called *causal chain*, including all the feedbacks within this chain. The major innovative characteristic of an IAM, which has made it an essential tool in IA, has been its capacity to dynamically include feedback loops between different stages of the causal chain. There is no other means of integration of insights into complex interlinked cause–effect relationships from multiple disciplines with that capacity.

The first generation of these integrated models focused on acid rain. The Regional Acidification INformation and Simulation (RAINS) is the most obvious example (Alcamo *et al.*, 1990). RAINS was developed in the 1980s at the International Institute for Applied Systems Analysis. The RAINS model played a major role in the international acid deposition negotiations in the framework of the United Nations (UN) Convention on Long-range Transboundary Air Pollution, and became an annex to the UN SO_2-protocol (*see* RAINS (Regional Air Pollution Information and Simulation), Volume 4).

In the second half of the 1980s, the success of the RAINS model led to the belief that integrated modeling would be the optimal way to interface science with policy, regarding environmental issues. Consequently it is not surprising that IAMs were developed all over the world. For the climate issue more than 50 IAMs coexist (Van der Sluijs, 1997). Looking back at the developments in the IA field over recent decades, the role of IAMs in the IA process has been closely linked to the policy development process. The climate change problem is a key example to illustrate this changing role and its consequences for evolution of IA and its toolbox.

The development of international climate policy can be roughly divided into five periods (Table 1). IA started to play a role in the agenda-setting phase. In this process of agenda setting, the 1985 Villach Conference on the Assessment of the role of Carbon Dioxide and other Greenhouse Gases in Climate Variations and Associated Impacts (*see* Villach Conferences, Volume 4) and the 1988 World Conference on the Changing Atmosphere: Implications for Global Security in Toronto were the milestones. These conferences were typical attempts to bring together experts

Table 1 The development of international climate policy

Period	Phase	Major characteristic
1895–1985	Foundational period	Scientific concern about global warming developed
1985–1988	Agenda-setting phase	Climate change was transformed from a scientific into a policy issue
1988–1990	Pre-negotiation period	Governments became heavily involved in the process
1990–1992	Formal intergovernmental negotiations phase	Led to the adoption of the FCCC[a] in May 1992
1992–present	Post-agreement phase	Focus on the elaboration and implementation of the FCCC and the initiation of negotiations on additional commitments; more and more stakeholders became involved in the climate policy debate

[a] UN Framework Convention on Climate Change.

from a multitude of scientific disciplines to arrive at an integrated view on the climate problem. These meetings were successful in putting the issue of climate change on the policy agenda, and were the first steps towards the creation of IPCC (*see* Intergovernmental Panel on Climate Change (IPCC): an Historical Review, Volume 4).

In the pre-negotiation phase, climate IAMs began to be developed. For instance, in 1986 the Netherlands National Institute of Public Health and Environmental Protection developed the Integrated Model to Assess the Greenhouse Effect (IMAGE)-1 model (Rotmans, 1990). IMAGE was a pioneer in the field: the (completely revised) IMAGE-2 model (Alcamo, 1994) is still one of the leading models, at the frontiers of IA modeling. IMAGE was developed at the time when the climate problem appeared on the policy agenda. In that phase, scenario analyses were needed to support the thinking about the question of how much greenhouse gas emission reduction was needed in order to manage the risk. IAMs were also used for scenario analysis in the successive assessments produced by the Intergovernmental Panel on Climate Change (*see* Intergovernmental Panel on Climate Change (IPCC): an Historical Review, Volume 4).

In response to developments in the post-agreement phase of the international climate policy debate, in which quantified emission limitation and reduction objectives were negotiated, IAM results from different modeling groups were fed into the negotiations. A typical IAM application in this phase has been the calculation of so-called safe landing zones and corresponding safe emission corridors (or tolerable windows). *Safe emission corridors* are defined as the allowable lower and higher bounds of greenhouse gas emission scenarios, relating to a set of criteria for climate policy whose purpose is to protect both the environment and the economy from disruption. Ecological constraints (maximum acceptable total temperature increase; maximum acceptable rate of temperature change; and maximum acceptable sea level rise) determine the upper limit and economic constraints (maximum acceptable rate of emission reduction) the lower limit of the corridor. Such calculations have played a role in selecting the emission reduction targets in the Kyoto agreement. Overall over the past decade, IAMs have played an increasingly important role in the legitimization and scientific underpinning of policy targets regarding global environmental risks.

With the Kyoto protocol, a new phase has started. Major changes of global economic structure are being envisaged, as implied by long-term reductions of greenhouse gas emissions. As the 1998 Buenos Aires Conference of the Parties (COP)-4 and the 1999 Bonn COP-5 meetings have shown, however, no such major changes are currently being implemented. Rather, a long-term effort has started to design possible institutions and mechanisms whose implications are gradually being explored by international environmental diplomacy.

This phase is essential to build up the know-how and the trust required to develop effective environmental policies at a global scale. However, it is clearly insufficient if the problem of climate change is to be effectively addressed. If effective policy is to emerge, actions taking place at the level of international environmental diplomacy must be combined with actions involving various stakeholders, who range from peasants to forest managers, from tourist operators to inhabitants of coastal zones, and from financial investors to ordinary citizens. Involving the latter will be necessary because climate mitigation measures will require consumer and worker cooperation, as well as citizen consent, to be successfully implemented.

For IA, this implies a growing need for the integration of social science research, and in particular of participatory techniques into research on global change. Social science is needed to provide knowledge about stakeholders and their ways of opinion forming, and also to provide opportunities for including the knowledge of stakeholders and their judgments about controversial issues in policy making.

An innovative development in the IA field has been the design of *Participatory IA* (PIA) methodologies. A key example and pioneer in PIA is the European Urban LifestYles, SuStainability and Integrated Environmental Assessment (ULYSSES) project (*see* ULYSSES (Urban Lifestyles, Sustainability, and Integrated Environmental Assessment), Volume 4). ULYSSES has developed procedures allowing interfaces between expert models of environmental change on the one hand, and lay participants in focus group discussions on the other hand. ULYSSES conducted IA *focus groups*, in which selected citizens debated climate change and mitigation options, and were supported in their deliberations by access to expert information. IA focus groups undertake group discussions conducted with citizens and guided by a moderator. The citizens are given access to computer models on complex environmental topics to support their debates. For this process, a model moderator supports them. At the end of the process, the citizens give their assessment of the problem and of policy options. The PIA methodology developed in the ULYSSES project is comprehensively described at http://www.zit.tu-darmstadt.de/ulysses/tutorial.htm.

ULYSSES has shown that IA focus groups can increase the possibilities for ordinary citizens to articulate their views on environmental issues in a thoughtful manner. This can help to bridge possible gaps between decisions by policy makers and citizens' views. Exploring alternative explanatory and moral frames in such processes can complement expert assessments on complex environmental issues. In particular, focus group procedures involving open-ended discussions are well suited for PIA. In PIA, the discussions can broaden into directions impossible for the researchers to imagine beforehand. This is vital if we are to learn more about how the diversity of everyday life and experience affects the understanding, values, and objectives that different social groups bring to complex environmental issues.

Analytical IA methods and PIA methods complement each other in the IA process, the one being strong where the other is weak and *vice versa*. The fruitful combination of the two approaches yields a powerful means to interface science and policy with the capacity to cope with the complexity, uncertainty and value diversity that characterize present-day environmental problems.

REFERENCES

Alcamo, J, Shaw, R, and Hordijk, L, eds (1990) *The RAINS Model of Acidi cation: Science and Strategies in Europe*, Kluwer, Dordrecht.

Alcamo, J, ed (1994) IMAGE 2.0: Integrated Modeling of Global Climate Change, *Water, Air, Soil Pollut.*, 76(1/2).

Parson, E A (1995) Integrated Assessment and Environmental Policy Making, In Pursuit of Usefulness, *Energy Policy*, 23(4/5), 463–476.

Rotmans, J (1990) *IMAGE an Integrated Model to Assess the Greenhouse Effect*, PhD Thesis, Rijksuniversiteit Limburg, Maastricht.

Rotmans, J and Dowlatabadi, H (1997) Integrated Assessment Modeling, in *Human Choice and Climate Change*, eds S Rayner and E L Malone, Vol. 3, The Tools for Policy Analysis, Battle Press, Columbus, OH, 291–377.

Toth, F L and Hizsnyik, E (1998) Integrated Environmental Assessment Methods: Evolution and Applications, *Environ. Model. Assess.*, 3, 193–207.

Van der Sluijs, J P (1997) *Anchoring Amid Uncertainty; On the Management of Uncertainties in Risk Assessment of Anthropogenic Climate Change*, PhD Thesis, Utrecht University, Utrecht.

Weyant, J, Davidson, O, Dowlatabadi, H, Edmonds, J, Grubb, M, Parson, E A, Richels, R, Rotmans, J, Shukla, P R, Tol, R S J, Cline, W, and Frankhauser, S (1996) Integrated Assessment of Climate Change: an Overview and Comparison of Approaches and Results, in *Climate Change 1995, Economic and Social Dimensions of Climate Change*, eds J P Bruce, H Lee, and E F Haites, Contribution to Working Group III to the Second Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge MA, 183–224.

Integrated Assessment Models of Global Change

Pierre Matarasso

Centre International de Recherche sur l'Environnement et le Développement, CNRS-EHESS, Nogent-sur-Marne, France

Integrated assessment models (IAMs) of global change are computer models, which integrate sets of equations describing the interwoven evolution of the different sub-systems making up the Earth system:

- *human production systems and their economic counterparts, including exchanges, nancial and monetary issues;*
- *coupled atmospheric and oceanic systems; and*
- *terrestrial and marine ecological systems.*

IAMs are based on a detailed description of geodynamic and ecological phenomena involving general circulation models (GCM), atmospheric chemical models and ecological models. However, economic issues are central to the assessment, through the coupling of these natural science models with macroeconomic models.

The goal of an IAM is to produce a uni ed picture of climate change oriented towards the evaluation of the

consequences of this phenomenon (negative and positive impacts), taking into account the driving forces behind climate change (demography, economy, development •••), possible mitigation processes (nuclear and renewable energies, demand side management (DSM), carbon capture and sequestration) and adaptation processes (changes in production methods in agriculture, protective infrastructures, etc.). IAMs of very different sizes and levels of detail have been developed (Dowlatabadi, 1995; Rotmans, 1998; Janssen, 1998).

It should be stressed that the question raised by IAMs is in essence a problem of decision-making involving a common property resource at a global level in an uncertain context, where the measures against climate change and their consequences are spread over much of the Earth's surface in an extended time frame. The purpose of an IAM is to integrate all these measures to prevent climate change, and its positive or negative impacts in a global framework. The aim of this integration is to de ne an indicator of ef ciency for each possible strategy involving mitigation and adaptation measures, summing up costs and bene ts of these measures at a world level and over the entire time scale of climate change. In other words, IAMs are devoted to climate policy analysis and in particular, to the computation of optimal emission trajectories. This problem concerns the most complex object that science has ever had to cope with: Planet Earth. Therefore, IAMs raise an important philosophical question: what are the scienti c limits to the knowledge of our planet, considered as a complex system? Which parameters of our future action regarding climate change can be selected according to reason and reliable arguments? An IAM constitutes an attempt to transcend providence or intuition in the decision and negotiation processes concerning climate change.

More speci cally, an IAM constitutes an attempt to nd answers to the following questions: (1) what are the best economic compromises involving measures designed to limit climate change and adaptation to it?; (2) what is the optimum distribution of the burden of mitigation and adaptation measures over different parts of the world?; (3) how do we combine institutional measures, such as taxes, tradable permits and subsidies to technical innovation to stabilize greenhouse gases (GHG) emissions at a safe level? A typical problem addressed by an IAM is to analyze when, where, and how to react to climate change in order to satisfy ethical and economical principles such as the precautionary principle, cost bene t analysis or cost ef ciency analysis. More generally, the dynamic aspects of global change are studied in the framework of an IAM with the intention of assessing the relative importance of the various uncertainties pertaining to that phenomenon.

The research community presently gathering around IAMs has reached a broad consensus of opinion on the questions that IAMs should address. However, in contrast, the technical solutions to these problems are extremely diverse, due to the complexity of the system under study. Therefore, the range of problems addressed by IAMs will probably remain stable for some time while model-building techniques will rapidly evolve. We will attempt to sum up this consensus, before introducing some of the general technical aspects of IAMs.

THE PARADIGM OF DECISION-MAKING IN FACING GLOBAL CLIMATE CHANGE: WHEN, WHERE AND HOW TO ACT?

If we decide to combat climate change, we have to answer at least three classes of questions related to the nature of different institutional and technical measures. The questions are:

- when are measures against climate change required?
- where are measures required? and
- what kind of measures are needed? (Manne and Richels, 1995).

In addition, measures should be evaluated in proportion to their consequences in terms of impacts. Impacts are obviously linked to the atmospheric levels of GHGs, which result from each set of measures.

Each of the preceding questions can be subdivided into many subtle problem areas, not the least of which is the fact that measures, as well as their consequences, in terms of impacts must be analyzed at three separate levels: (1) the physical level; (2) the direct economic level; (3) the indirect macroeconomic level of effects induced by mitigation policies and climate change impacts.

When to Act, or the Timing Problem

The elaboration of a set of institutional measures (taxes, quotas, tradable permits, innovation incentives) and related technical measures over time (mitigation and adaptation processes) is determined by the speed and level of climate change, technical progress, research and development expenditure, demography, economic development, etc. For example, an IAM can be used to evaluate and compare a package of precautionary measures and a package of follow-up measures with various temporal horizons associated with different final states regarding GHG concentrations and different levels of climate change. The use of an IAM has shown that many different paths (with various emission and mitigation profiles) can produce the same results in terms of final atmospheric concentration of GHGs. Therefore, decision-makers have to elaborate and compare sets of measures, with contrasting time-scales, which can lead to different global social costs. The determination of an optimal level of action, representing a good mix of mitigation and adaptation measures is sought. This is an

area where the use of an IAM has, directly or indirectly, structured and informed the international debate concerning the timing of policy measures (Wigley *et al.*, 1996).

Where to Act and Spatial Flexibility

The possibilities of reducing the emissions of GHGs are distributed throughout the different geographical regions and nations of the world. If we assess the costs of these reductions, it might appear that some geographical areas present, for technical and economic reasons, potentially cheaper mitigation possibilities than others. In this context a strategy of mitigation can be viewed as a set of technical operations indexed by geographical location and time scale, where these operations are ordered by increasing costs. Determining such strategies is one of the main goals of an IAM.

How do We Act on a Technical Level with respect to Trade-offs in the Mitigation Process?

On a technical level, mitigation measures include a number of potential components:

- at the level of energy demand;
- at the level of energy production (renewable or nuclear energy, more efficient energy conversion, carbon dioxide (CO_2) capture devices, etc.);
- at the level of the creation of CO_2 sinks (tree planting, induced algae production •••);
- mitigation measures could be distributed amongst the different GHGs (CO_2, methane (CH_4), nitrous oxide (N_2O), sulfur hexa uoride (SF_6), per uorocarbons (PFCs), hydro uorocarbons (HFCs)), with possible trade-offs between these gases.

Many IAMs are formulated to suggest least-cost packages of technical mitigation measures integrating various possible trade-offs and compromises between the different technical options.

What Kind of Public Policy will Trigger an Appropriate Association of Technical Options?

In many past studies related to energy issues, energy options were considered on the basis of technical scenarios. The shift from technical scenarios to reality was accomplished through planning and national policy formulation. This view is no longer valid in the present context of competition on energy markets. In negotiations concerning climate change, the discussion has progressively shifted from an approach centered on technical scenarios to the definition of public policies in terms of economic procedures and legal agreements. One of the goals of many IAMs is to test different economic taxes and permit regimes (distribution of entitlement to emission rights, quotas and ceilings of emissions, effects of differing exibility mechanisms such as the *clean development mechanism* (CDM) (*see* Clean Development Mechanism, Volume 4). Simulation of permit trading or of progressive carbon taxes and their interaction with development, growth, economic processes, technical progress, world markets of goods and capital are central to many IAM concerns. Equity issues play a major part in the evaluation of simulation results.

Evaluation of the Climatic Impacts of Different Packages of Measures

New climate patterns can be very diversified and this may, in turn, have various consequences on economic activities. Categories of impacts can be divided into the negative and positive impacts that can affect different societies at different moments in time. Typical forms of climate events producing damages or negative impacts are:

- hurricanes and tornadoes and other types of extreme weather events;
- droughts or oods; and
- rises in sea level.

Possible positive impacts are

- Lengthening of growing seasons;
- CO_2 fertilization of plant growth; and
- improved rainfall in dry areas.

IAMs provide a quantified description of impacts and evaluations of regional consequences through climate patterns, calculated with GCMs fed with GHG concentrations resulting from sets of measures and actions. Particularly important is the study of the impacts of climate change on economic activities and possible adaptations (Tol, 1996).

Evaluation of the Cost and Macro Economic Consequences Induced by Mitigations or Adaptation Measures and Impacts

It should be noted that each pattern of measures and impacts of climate change has a profound in uence on economic issues. A few examples will illustrate this point:

- the reduction of individual car use through switching towards other transportation modes might have a knock-on effect in the entire economy (with, for example, reductions in car production and a slowing down of individual transportation-related activities in services such as garages, gasoline distribution, etc.);
- limiting energy consumption might depress energy prices, which could be detrimental to oil producing countries. On the contrary, measures based on CO_2 capture will not affect energy markets;

- promoting a new international market of permits can have a marked in uence on the mobility of capital; and
- a reduction in agricultural production may result in economic decline and migration.

IAM MATHEMATICAL STRUCTURE

In the absence of a general theory of modeling, only a very simplified description of the technicalities and issues concerning the working of IAMs can be given. An IAM is a computer model, but a model made up of many intricate phenomena belonging to the natural and human domain. It is an assemblage of very heterogeneous processes. In some IAMs, different sub-models are linked, in others equations representing the different phenomena are more truly integrated. Such models may be described from different viewpoints. The first viewpoint is thematic; it consists of an individual description of the rather independent sub-models or models at a lower level coupled in the global model (equations grouped in themes and linked by common variables). The second viewpoint concerns purely methodological considerations independent of the classes of the phenomena described by sets of equations and involves a more fundamental definition of types of equations. The third viewpoint is related to solving procedures. In order to be useful, a model should produce a solution; and yet, different solving procedures are available for a given model, depending on the questions put to the model. From this latter viewpoint, IAMs are very different from natural science models that do not raise the issue of decision-making and hence the study of many alternative scenarios that will never materialize in the real world. The problem of decision-making in a context of indeterminacy and uncertainty is at the core of IAM resolution and computing (Dowlatabadi, 1995; Rotmans, 1998; Janssen, 1998). In order to understand the way in which IAMs work and especially the procedures for solving them, the second viewpoint should be adopted.

General Structure of an IAM

An IAM is a set of equations describing the evolution over time of the Earth System as a whole. Basically an IAM is a dynamic model, which implies the following types of equations, illustrated by simplified examples:

- Equations concerning evolution, linking one period to the next period. For example, the equation which establishes that the level of concentration of CO_2 is dependant on emissions, mitigation and on the different natural processes of absorption of this gas. Other classic equations in the same category involve the evolution of capital goods through investment, obsolescence and dismantling, and descriptions of demographic change according to birth and death rates.
- Equations linking contemporaneous variables; equations linking GHG concentrations to global temperature; the relationships between damages and global temperature. In the economic domain, production expressed as a function of various factors such as available capital and available manpower; emissions calculated as a function of global production; constraints stipulating that consumption and investment should not exceed total production at a given time.

Concerning variables, it is usual, in the first instance, to distinguish between:

- ow variables (emissions, CO_2 captured by forests and biomass, final consumption, investment, death and birth.); and
- stock variables (atmospheric level of CO_2, levels of various capital stocks, population figures).

This is only a rough classification as some variables, which play a major role in quantification of ows, are neither stock variables, nor ow variables. Good examples of such variables are the temperatures in a GCM and the prices of goods in the economic section of an IAM. In a simplified interpretation of these classifications of equations and variables, it can be argued that the equations of evolution describe the evolution of stock variables as a function of the values of the relevant ows.

In the past, models were often expressed in continuous time (and treated as systems of differential equations), especially natural science models such as GCMs. Today many factors favor the expression of models in discrete time (with finite difference equations):

- the models are too complicated to be treated analytically and the calculation of results requires the use of computers that rely on discrete computations;
- at the level of the Earth System, natural discrete periods already exist, such as the seasons or the farming year and budgeting within industry and services;
- the treatment of discrete events implies the use of discrete time periods.

Many IAMs are therefore expressed in a discrete time framework. In some models, different time scales could be used since geodynamic phenomena often involve shorter time periods than economic processes.

Solving Procedures in IAMs

Basically two types of solving procedures are used in IAMs

- recursive solving through sequential simulation; and
- intertemporal solving through optimization.

The principle of recursive sequential simulation is fairly straightforward: with prescribed initial conditions, the state of the system at $t \bullet 1$ is computed from the state at t, by introducing values for ows in the evolution equations. Some of these ows are defined as control variables or policy variables. In simulations these control variables will be given through an exogenous control scenario (for example, a mitigation policy). It should be noted that all ows should be specified to arrive at a determinate evolution of the system represented. The calculation follows the temporal order of the real pattern of events and simulates change.

The principle of inter-temporal solving through optimization is different. A system of simultaneous equations is established by duplicating, for all periods, the group of equations valid for one period (equations of evolution and contemporaneous equations). Initial conditions and, eventually, final conditions are defined. This system of simultaneous equations expressed for n periods, with time indexes for each period, is then solved globally as a whole (and not sequentially as in ordinary simulation), generally through optimization algorithms.

The main difference between the two approaches lies in the treatment of decision-making. In simulation, in order to impose a determinate evolution, all the decision variables (usually ow variables, such as investment decisions in the energy field) have to be defined exogenously by a scenario. This is why simulation models are considered as models of policy evaluation. The evaluation centers on the scenario (mitigation and adaptation policies, emission permit policy) fed into the model and its consequences in terms of climate change impacts. In contrast, in inter-temporal solving procedures, some of the possible decision variables are left free and a determinate evolution is computed in which the algorithm selects an optimal control, that is a complete set of control variables covering the whole time scale. Optimization is achieved by selecting a temporal evolution by computation which maximizes an inter-temporal function, usually an economic welfare function expressed by a discounted sum, over all periods of the utility of consumption at each period. In this framework, running the model results in the selection of a coherent optimal set of decisions throughout the whole time scale (mitigation and adaptation measures, or least-cost mitigation for a given target of final GHG concentrations) including consequences in terms of damages, cost and benefits. This is why this procedure is termed *optimal policy de nition*.

It should be stressed that in the optimization framework, various constraints on the final period (stable final concentrations of GHGs for example) can be specified and computation produces solutions respecting these constraints (or the optimization algorithm yields the result that such a solution is impossible to find). In simulation, on the other hand, the scenario submitted to computation produces end conditions that could not be anticipated. This is why simulation might lead to overshoot and collapse not previously envisaged in the scenario submitted. In other words, simulation is blind or short sighted and optimization in some respects is forward looking and mimics behavior with expectations or perfect foresight, where decisions are taken with a knowledge of their consequences. This last peculiarity is especially important when trade-offs between decisions over a longer time scale are envisaged.

In theory, any dynamic model structure (if by model we mean the core structure of the equations) can be solved either way. In practice, our ability to compute inter-temporal optimization is limited to simplified models for a basic reason: an optimization algorithm explores an infinite number of alternative solutions whilst simulation computes only one solution at a time. Optimization therefore requires much greater computing power. Other model-solving procedures should also be mentioned, especially the inverse mode in which a solution is computed in a reverse temporal order, with a final state taken as initial conditions.

For an excellent example of policy optimization models see Nordhaus (1994) who introduced this paradigm in the field of climate change. An example of a leading simulation and policy evaluation model, IMAGE, can be found in Alcamo (1994).

CURRENT DEBATES CONCERNING IAM STRUCTURE AND USES

Cost Benefit Analysis Versus Cost Effectiveness Analysis

Evaluation of impacts is one of the weakest elements of an IAM. Impact evaluation raises very complex problems due to the many uncertainties of climate models for long-term forecasts.

Some IAMs try to integrate the cause (human activities), modification of the cause (mitigation of emissions), consequences (climate change impacts) and adaptation to consequences (adaptation to climate change) in the framework of cost-benefit analysis (CBA). Costs are related to mitigation, adaptation or damages and benefits to avoided damages or more favorable climate conditions. But due to the huge uncertainties concerning physical impacts, evaluation of adaptation costs, many models of a lower level of integration adopt a cost effectiveness analysis (CEA) framework. In such cases the aim is to reach a given stable final concentration target of each GHG or to satisfy mitigation objectives at least cost (Bruckner *et al.*, 1999).

The Top-Down Versus Bottom-Up Debate

The level of description of economic phenomena, especially energy systems, is a much debated question in IAMs. Traditionally there are two opposing approaches, top-down and

bottom-up modeling. Top-down models synthesize major tendencies in economic trends such as demography, economic growth, investment, consumption and a fairly crude indicator of trade. In a top-down IAM context, geodynamics are aggregated and damages are dealt with through damage functions in a global context.

Bottom-up models involve the detailed description of components of economic systems, usually energy systems, in terms of elementary process analysis. When IAMs are developed from a bottom-up perspective, damages are also usually more detailed and their evaluation takes into account the temporal and geographic peculiarities of climate change and their consequences on agriculture, industry and housing.

When insights regarding timing of action are sought or when a broad geographical distribution of the burden of mitigation is envisaged, order-of-magnitude analysis is often a prerequisite. In this case, top-down models are used. But if the question involves precise mitigation options with trade-offs between different energy technologies (nuclear, renewable), trade-offs between gases (CO_2, CH_4, N_2O, SF_6, PFCs, HFCs), and trade-offs between reductions of emissions and capture of CO_2 by tree plantation, bottom-up models are absolutely necessary. Only a bottom-up model can break down a global mitigation effort into a sequence of precise technical decisions with time and regional indexes. Increasingly, the frontiers between top-down and bottom-up models are vanishing; some models incorporate a bottom-up description of the energy or production sector (for example the MERGE model, and MARKAL family models) within a top-down framework for the whole economy. For the foreseeable future, bottom-up models will raise questions concerning data availability and computation. Even if they seem more satisfactory from many points of view, it will be a long time before they are fully effective (Dowlatabadi, 1995; Janssen, 1998).

Uncertainty

Uncertainty is the thorniest issue in the IAM field; and it does not appear that a consensus has yet been reached. However, some methodological trends can be identified. They range from a strict application of the precautionary principle to more classic applications of decision-making in the framework of uncertainty theory. The treatment of uncertainty causes difficult problems in computation, as uncertainty adds many dimensions to the search for sustainable development solutions. Also the question of whether a reduction of uncertainty through research and development is possible has to be raised. Usually two general categories of uncertainty are defined:

- parametric uncertainty that stems from imperfect knowledge; and
- stochasticity due to natural variability.

Significant uncertainties occur, for example, in demographic behavior, damage functions, energy efficiency trends and in endogenous technical progress. The discount rate (policy choices are very sensitive to the discount rate) and information costs (funds spent to reduce uncertainty) are important issues in the uncertainty debate. On the important subject of discounting, see Portney and Weyant (1999). It should also be noted that uncertainty is a major issue in the debate on the timing of mitigation; uncertainty acts in favor of precautionary action, especially in the case of the high inertia of energy systems (Ha-Duong *et al.*, 1997).

Various ways (with many intermediate levels) of treating uncertainty have been developed:

- a systematic treatment of uncertainty using rigorous mathematical techniques such as stochastic dynamic optimization (Kann and Weyant, 2000);
- application of the precautionary principle and the viability approach through inverse problem methodology, where sustainable approaches in an uncertain context are established through a reverse computation (the *Tolerable Window Approach*) (Bruckner *et al.*, 1999);
- exploring a large number of alternative problem formulations with different sets of risk events (cultural theory) (Rotmans, 1998).

The When and Where Flexibility Debate

In a cost-effectiveness framework, one of the hotter questions is the search for the best allocation of mitigation in space and time. Use of IAMs has shown that the answer to these questions can be far removed from a first guess, which is usually to begin action as soon as possible and in the countries with higher levels of energy consumption. In fact there are many arguments in favor of delayed action. For example, it seems justified to await the effect of technical progress, the end of life of current energy installations, a better understanding of the real effect of climate change and a better discounting effect. Delayed action appears to have some advantages in economic terms when tested through IAM. In space it is best to allow some degree of exibility (with emission permits or CDM) in order to take advantage of the lowest mitigation costs available at a world level. The idea of exibility in time has been called into question when dynamic effects such as the increased cost of mitigation are introduced into the scenario, where rapid mitigation is required (if impacts turn out to be worse than previously expected). The when and where controversy is a typical example of a debate, which has developed from insights introduced by IAMs. Such a debate stemmed from the use of inter-temporal IAMs that showed that the introduction of exibility in time and space favors lower global social costs in climate change (Wigley *et al.*, 1996; Ha-Duong *et al.*, 1997).

IAM IN THE CONTEXT OF THE HISTORY OF SCIENCE

World Dynamics

As previously stated, the idea of constructing models of what is, perhaps, the most complex object ever considered in science, planet Earth, raises many epistemological questions. For some scientists, this attempt is still illegitimate. The idea of comprehensively integrating global phenomena is not new and the first approach of this type was both highly praised and strongly denigrated (Forrester, 1971; Meadows *et al.*, 1972). These two books met with a very mixed response. Widely accepted by the public and the media, they were rejected by many earth scientists and economists. Many people misinterpreted the results as predictions rather than simply answers to what-if questions. Notwithstanding an inadequate appraisal of the Earth's complexity (through the concept of static limits, the use of differential equations and deterministic system dynamics, inadequate economic representation, etc.), these two books should receive due credit as the first attempt to think globally about the future of planet Earth. As such they in u-enced many of the contemporary model builders.

But the criticisms leveled were equally important in developing lines of argument that still apply to the new generation of models:

- the data are insufficient;
- knowledge of the differing phenomena is grossly approximate;
- in running models, many elements, ranging from the inadequate treatment of uncertainty to deterministic chaos and multiple equilibriums, alter the dynamics.

However, although they seem justified at first sight, these critiques do not suggest any alternative. The problem is complex and so are the methods devised to treat it. Therefore, we should make the effort to analyze both the limitations and the merits of IAMs. The models should not be considered as truth machines, but as the best (and only) means we have to elaborate complex arguments related to a complex problem. In this context, models should be viewed as arguments in form, integrated into the framework of provisional knowledge, with the intention of defining action, rather than as a faithful mirror of the essence of the phenomena. At best, IAMs should be seen as providing order-of-magnitude modeling, comparison of policy scenarios and forecasting of trends, but not prediction.

A New Dimension in Science

The attempt to construct a predictive or a decision-making model at the level of an object as complex as planet Earth opens an entirely new field in science. We should first underline the paradox produced by this attempt. Since the 16th century the study of the universe and the cosmos has been a miraculous success. The study of the terrestrial environment and complex ecologies is only just beginning. The *cosmos* is, in some ways, more familiar than the *oikos* (Greek term from which ecology and economy derive). The part seems more complex than the whole. Clearly planet Earth is almost unthinkable in its complexity, especially in the context of decision-making. However, unfortunately, the human impact on the terrestrial environment through climate change has obliged us to focus on this objective.

Such complexity defies the very element which underpins the success of modern science since Galileo: the reduction of the complex to simple components. The scientists of the 17th and 18th centuries provided a new understanding which led to the development of modern science through a straightforward metaphysical principle: the superior being who created the world founded this creation on reason and simplicity, and based it on mathematics. The way to understand this world was to apply a platonic principle: to look for what did not change in the change itself and express this in mathematical form as the laws of change. The quest for invariance, symmetries and conservation principles opened up a new avenue in the discovery of mathematical principles underlying phenomena at a cosmological level. This constituted the origin of differential calculus. Such an approach led to the successful unification of the earth and the heavens under the universal laws of physics. This replaced Aristotelian physics, which preceded modern science, in which the earth and the heavens were governed by different principles.

The approach of the new physics of the 17th and 18th century did not, however, provide an answer to some of the problems which have been raised since Aristotle's time and which are still relevant at a terrestrial level (in ecology and economy). From the 17th to the 20th century, the mathematization of reality borrowed most of its principles and style from a mechanical model, that is to say mathematical modeling of movement, or *kinesis* (and associated systems of differential equations). But at the level of global reality, there is a form of change other than the change of place. This other type of change is designated by Aristotle as *genesis*. In genesis, new beings differing from those preceding are created and the ways of changing constantly change, in part because of the introduction of these new beings. Modern IAMs are still confronted with the major difficulty of the temporal heterogeneity of dynamic phenomena.

Some scientists who paved the way for modern IAMs, such as Von Neumann, Georgescu Roegen or Casti, underlined the fact that models of reality at a

global level should depart from the classical platonic discourse concerning phenomena (search of invariance and mathematical realism) and return to a new form of Aristotelianism (formalization of reasoning through logic on reality with deep qualitative changes). Besides the question of genesis at a terrestrial level (as underlined by Casti), not everything can be explained by a single cause, an efficient cause (a cause entirely defined by ordinary state variables), as in the mathematization of movement. Other forms of cause such as the Aristotelian formal causes, i.e., phenomena guided by programs or internal representation, should be reconsidered. Therefore, some new principles should be introduced in models of the phenomena at the level of the Earth, notably:

- *discrete events*: events which cannot be submitted to an infinite division; a move in a game, the process of formation of an animal and the construction of a factory are examples of such events;
- *formal causes*: the evolution of living creatures implies a program such as a genetic code; the evolution of human societies implies intention, will, anticipation and programs of action founded on symbolic representations by economic actors;
- *indeterminism*: the evolution of social phenomena depends on decisions taken by economic actors or agents.

In this context, classical laws of evolution only partly determine the global phenomena integrating natural events and human activities. In this sense the evolution of the Earth System is indeterminate as it involves decision-making. This new context pervades the IAM attempt to represent global phenomena. More precisely, the problem is no longer to define a phenomenology of the Earth's evolution, which treats this phenomenon as something exterior to humankind, on which we have no action, akin to planetary motions. On the contrary, we have to study very diverse trajectories which will never materialize, apart from the one that will derive from our decisions or actions. In this sense, the main object of an IAM is to set up realistic fictions to facilitate decision-making. This self-contradictory expression raises many questions: what are the realistic criteria in representing a very complex and heterogeneous object and what is the optimum level of detail in the fiction? The prevailing belief in classical science of the quasi-perfect adherence of the mathematical model to the phenomena (such as in mechanics and electromagnetism) no longer applies, and this leads to many epistemological enigmas.

All these peculiarities are responsible for major difficulties in model building (in articulating a deterministic view with alternatives and uncertainty in decision-making), in model calibration (calibration of sequences of events far from known situations), in running models (protocols for running models devoted to policy definition).

Some scientific resources exist to treat these problems, occupying a position between logic and differential calculus. Important sources of inspiration may be provided by operational research and optimal control theory.

REFERENCES

Alcamo, J (1994) *IMAGE 2.0 Integrated Modelling of Global Climate Change*, Kluwer Academic Publishers, Dordrecht.

Bruckner, T, Petschel-Held, G, Toth, F L, Fussel, H-M, Helm, C, Leimbach, M, and Shellnuber, H-J (1999) Climate Change Decision Support and the Tolerable Window Approach, *Environ. Model. Assess.*, 4, 217–234.

Dowlatabadi, H (1995) Integrated Assessment Models of Global Change, An Incomplete Overview, *Energy Policy*, 23(4/5), 289–296.

Forrester, J W (1971) *World Dynamics*, Wright-Allen Press, Cambridge, MA.

Ha-Duong, M, Grubb, M J, and Hourcade, J C (1997) In uence of Socio-economic Inertia and Uncertainty on Optimal CO_2 Emission Abatement, *Nature*, 390, 270–274.

Janssen, M (1998) *Modelling Global Change: The Art of Integrated Assessment 3 Modelling*, Edward Elgar, Cheltenham.

Kann, A and Weyant, J P (2000) Approaches for Performing Uncertainty Analysis in Large Scale Energy/economic Policy Models, *Environ. Model. Assess.*, 5, 29–46.

Manne, A and Richels, R (1995) The Greenhouse Debate: Economic Efficiency, Burden Sharing and Hedging Strategies, *Energy J.*, 16(84), 1–37.

Meadows, D H, Meadows, D L *et al.* (1972) *The Limits to Growth: a Report for the Club of Rome's Project on the Predicament of Mankind*, Universe Books, New York.

Nordhaus, W D (1994) *Managing the Global Commons*, The MIT Press, Cambridge, MA.

Portney, P R and Weyant, J P (1999) *Discounting and Intergenerational Equity*. Resource For the Future, Washington, DC.

Rotmans, J (1998) Method for IA: the Challenges and Opportunities Ahead, *Environ. Model. Assess.*, 3, 155–180.

Tol, R S J (1996) *A Decision-Analytic Treatise of the Enhanced Greenhouse Effect*, Vrije Universiteit, Amsterdam.

Wigley, T M L, Richels, R, and Edmonds, J A (1996) Economic and Environmental Choices in the Stabilisation of Atmospheric CO_2 Concentrations, *Nature*, 379, 240–243.

FURTHER READING

Schellnhuber, H and Wenzel, J (1998) *Earth System Analysis, Integrating Science for Sustainability*, Springer-Verlag, Berlin, Heidelberg.

Integrated Pest Management (IPM) in an Era of Global Environmental Change

Richard A Fleming
Great Lakes Forest Research Centre, Sault Ste. Marie, Canada

Organisms considered as threats to human beings or to their interests are often classi ed as pests. The fact that many pretty garden plants (e.g., purple loosestrife, Lythrum salicaria*), once introduced to new locations around the world, became aggressive weeds emphasizes the subjectivity and context-dependence of the term pest. Besides weeds, organisms considered pests include disease-causing microorganisms (e.g., fungi, bacteria, and viruses), as well as invertebrate (e.g., protozoa, atworms, nematodes, snails, slugs, insects, and mites) and vertebrate (e.g., rabbits, elk, deer, birds, and rodents) animal species. As a group, these organisms threaten humans indirectly by causing losses to food and ber production systems in agriculture, aquaculture, and forestry, and directly by causing or transmitting disease (e.g., malaria, Lyme disease, yellow fever, West Nile virus, bubonic plague, and sleeping sickness).*

Generally, demands for pest management will almost certainly increase as global change progresses. The exact form of these demands is unknown. Global change can be expected to produce unpredictable changes in the nature of pest problems and in the effectiveness and perceived costs of control methods. These uncertainties have a number of important implications for future pest management. They make it important to have a full menu of methods available and to have the knowledge available to integrate these methods effectively. Generating the available knowledge will depend on research, sampling, and monitoring.

Integrated pest management (IPM) ful ls this role. IPM was conceived in the late 1950s as using ecological and economic knowledge to optimally blend biological and chemical methods for effective and sustainable pest control. Thus, IPM integrates different kinds of knowledge and different methods of control. Recently there has also been increasing emphasis on integrating pest management decisions into the context of total resource management. Advances in information technology are greatly facilitating this task. Thus, as global change progresses, one can expect that IPM approaches will increasingly dominate pest management, and that generating knowledge and integrating decision-making into total resource management will become ever more important aspects of IPM.

PEST CONTROL METHODS

Since their initial development in the 18th century, pest control programs have generally been aimed against either insects, pathogens (disease-causing organisms), or weeds. Pest control methods range from practices that reduce pest populations directly to methods that provide resistance and tolerance to pest attack. These methods have been categorized as cultivation, physical, chemical, and biological.

Cultivation Methods

Many cultivation methods involve reducing the suitability of the environment for pest populations. For instance, draining, ooding, or changing the soil's level of acidity or alkalinity are often used to combat soil-borne and sometimes other types of pests. Rotating a susceptible agricultural crop with non-host species or resistant varieties in time and space can reduce pest survival after emergence from resting or dormant stages of the life cycle and during dispersal. Other cultivation methods, such as the surgical removal of infectious parts of infected hosts (e.g., pruning diseased fruit or branches in orchards), directly decrease the size of the pest population. Other examples are sanitization, the elimination of shelters (e.g., crop litter, alternate host species) which pests rely on during critical stages of their life cycles, and roguing, the removal of diseased or infected plants or seeds. Tillage and cultivation (e.g., coverage with straw, leaves, or wood chips) often provide weed control.

Physical Methods

Physical methods generally aim to eliminate the pest entirely from all or parts of the resource. Physical methods can take the form of mechanical or sticky barriers such as netting, screens, or fences on small scales. On much larger scales, quarantine procedures and importation restrictions can support pre-existing geographical barriers. Heat treatments are often applied to crop seeds or greenhouse soil to kill seed-borne or soil-borne pests (e.g., insects, fungi, nematodes, and weed seeds).

Chemical Methods

Chemical methods are largely based on pesticides, materials poisonous to at least one stage of the pest's life cycle. The first recorded chemical pesticides were poisonous plant compounds. In the 1760s, ground tobacco was used to kill aphids in France. The 19th century saw the introduction of other natural products (e.g., creosote, rotenone, turpentine, kerosene, and petroleum) plus some inorganic compounds (e.g., lime sulfur, Paris green, lead arsenate, hydrogen cyanide, and Bordeaux mixture). The relative ineffectiveness of many of these pesticides prompted further development of pest control methods. During the late 19th

and early 20th centuries there was increasing confidence that pest populations could be successfully managed and manufacturing companies began to concentrate on improving pesticides and application technology.

The effectiveness of synthetic organic compounds such as the insecticide dichloro-diphenyl-trichloroethane (DDT) and the herbicide 2,4-dichlorophenoxyacetic acid developed during the first half of the 20th century beckoned in a new generation of chemical pesticides. These new pesticides were very effective at killing pests, but a variety of problems accumulated as their usage expanded. These problems included harm to non-target organisms and human health, poisoning of soil and water resources, and in some cases, a resistant pest population with its natural enemies depleted by pesticide exposure. Consequently, legislation restricting pesticide usage was enacted in many countries. For instance, Canada, Sweden, the United Kingdom, and the United States have each restricted or completely banned DDT applications because DDT accumulates in human body fat, harms certain wildlife species, and persists in the environment. This general situation motivated a search for selective pesticides (pesticides unlikely to harm exposed, non-target organisms) and for methods which accurately target applications on pests and reduce off-target effects.

Most pesticides (e.g., bactericide, fungicide, herbicide, insecticide, miticide, and nematicide) exhibit some selectivity and thus are typically categorized by the pests they kill. For instance, few insecticides kill weeds. The need for ever more selective pest control motivated the development of new pesticides such as mercaptosuccinate, trade named Malathion, which demonstrated reduced human toxicity and they revived interest in biological methods.

Research showed that careful timing of applications (to attack pest populations during vulnerable stages of their life cycle), adjustment of dosages, and advances in application technology could produce better targeted applications and reduce off-target effects. The need for better timing led to programs for monitoring pest activity, and with better timing the same effect on the pest population could be achieved with smaller dosages. Technological advances have been largely directed at reducing spray volumes. Most pesticides are sprayed as solutions or suspensions and this requires applying much inert material along with comparatively little active ingredient. Modern ultra-low volume spraying technologies apply less than 1 liter ha^{-1}, roughly 1% of the volumes applied in the 1950s. Systemic fungicides represent another recent advance in targeting. Crop plants absorb the applied fungicide and then distribute it throughout their tissues so whole plants become effectively resistant to the targeted fungal pathogen.

Biological Methods

Biological control has been practiced in one form or another for almost 10 000 years when farmers, probably unintentionally, began planting seed from resistant plants. Manipulation of natural enemies began to be seriously considered after the successful importation of the ladybird beetle (*Rodolia cardinalis*) from Australia in 1888 to control the cottony-cushion scale (*Icerya purchasi*) threatening California's citrus fruit industry. At about the same time, microorganisms were first used against insect pests. Interest in biological control waned as the effectiveness of synthetic organic pesticides became apparent, but in the early 1960s grave ecological concerns (e.g., pest resistance, risks to non-target organisms) about pesticides arose, and this propelled government research into biological methods and accelerated plant breeding programs for developing and patenting new pest-resistant plant varieties.

Biological methods encompass a broad range of strategies and approaches. The longest and most extensively employed method in agriculture is selection of crop varieties that are tolerant of, immune to, or resistant to pest attack. Many countries support large research programs devoted to incorporating genetic resistance against plant pathogens (and also some insects) into major crops. A difficulty with this approach is that as successful varieties get more widely planted, sometimes creating vast monocultures, selection pressure unavoidably mounts for pathogenic races able to overcome the resistance of these new varieties. Another genetically based method introduces lethal or weakening genes into an insect population so that mating produces what are virtually neutered insects, or an unbalanced sex ratio, in the next generation. The goal is to depress the reproductivity of that generation.

Sterile male release is another strategy aimed at depressing pest reproductivity. Males reared in captivity are sterilized by gamma radiation and then released to mate with wild females who consequently lay no viable eggs. This has been effective in Texas against the screwworm, a major cattle pest.

Some methods focus on the chemical substances which attract pests to food or to the opposite sex. Attractants for the opposite sex (pheromones) generally attract only a specific pest species and can be broadcast during the adult part of the lifecycle to disrupt mating. Traps baited with sex or food attractants can be used for monitoring or reducing pest population densities. In the latter case, traps might include a pesticide or pathogen as well as the attractant.

Natural enemies are the focus of many biological control strategies directed against insect pests. Conservation strategies attempt to maintain the effects of native predators and parasites by ensuring as favorable an environment for them as possible. Typically this is accomplished by habitat manipulation, often in conjunction with cultivation methods. Augmentation strategies generally involve rearing and then releasing native predators and parasites in order to increase the size, and hence the effect of, natural enemy populations. Introducing exotic natural enemies is often

attempted against imported pests. First, the pest's native habitat is searched for natural enemies. Appropriate species are then reared and introduced into the new area to attack the imported pest. Often the introduced natural enemy is unable to establish itself in the new environment. Occasionally the introduced natural enemy does establish but it attacks non-target organisms (as well as the target species) and may thus become a pest itself. Hall *et al.* (1980) estimate that less than 20% of introduction attempts have been economically successful; this demonstrates the complexities involved in manipulating population regulation.

Legislative Measures

Importation restrictions on goods which may harbor exotic pests are common aspects of international trade and travel. Legislative measures can also operate on a more local scale. For example, in some fruit-growing areas where pest control is the most expensive and time-consuming farming operation, quarantine regulations require removal of untended orchards acting as sources of pestilence for neighboring farms.

INTEGRATED PEST MANAGEMENT (IPM) AND GLOBAL CHANGE

In the late 1950s when problems with extensive use of chemical methods such as pesticide resistance and non-target impacts became apparent, and the relative ineffectiveness of available biological methods during pest outbreaks was a concern, the notion of IPM was conceived. Accordingly, chemical methods are used only when needed to supplement, and in ways minimally disturbing to, biological methods. Thus, IPM programs also require ecological and economic knowledge to decide when chemical use is justified and how chemicals should then be applied. Much of the ecological knowledge is gained from sampling programs designed to estimate pest population sizes and life cycle stages as the growing season progresses. Recent pest population sizes are typically used as a basis for predicting population sizes later in the growing season, and hence, future impacts on the resource. Life cycle stage typically affects pest vulnerability to particular control methods and is therefore important to consider in choosing methods. Economic factors such as expected market price and pesticide costs partially determine whether the expected benefits of considered pest control efforts justify the likely incurred costs. Ecological costs (e.g., non-target pesticide impacts) should also be considered; so benefits and costs do not necessarily have monetary measures. A further complication is that pest management is only one aspect of the total management of the resource, so pest management decisions need to be integrated into the broader context of total resource management. For instance, in forestry, the same forest areas are often used for recreation as well as timber and paper production and these multiple uses mean that pest management decisions can no longer always be based on just economic thresholds. The issue of carbon credits is a new climate change-driven use for forests which can be expected to have some effect on pest control decision making, and the ranking of different control techniques. Thus, modern IPM integrates biological, chemical, cultivation, legislative, and physical controls; it integrates economic, ecological, and other knowledge; and it integrates pest management decisions in the broader context of total resource management.

Global Change Impacts on Pests

Global change has long been in uencing the pest status of many species. Global change is occurring because the scale of human activities now exceeds that of all other species and affects biospheric uxes of matter and energy, and it is the ever-increasing human demand for resources which underlies most pest problems. At one extreme, some pest species have maintained the same general population behavior since long before human activity had any significant effect on their environment. Nonetheless, they are considered pests now because expanded human demands for resources have made the pests competitors. For example, the outbreak cycle of the spruce budworm, *Choristoneura fumiferana* has shown little change since long before Europeans had any significant impact on North America's spruce/fir forests, but now this insect is considered a major pest by the forest industry.

Generally, however, the behavior of pest populations has responded to global change. Global change encompasses the direct and indirect effects of changes in atmospheric composition, climate, biological diversity, and land use, and it is particularly through effects on biological diversity and land use that human-caused environmental modifications have inadvertently contributed to the development of many pest problems. For example, the large-scale displacement of natural vegetation with single-crop (monoculture) agriculture has substantially improved the survival rates of some pests by: (a) providing a uniform food source over vast areas which allows easy access to new host plants for pests during the critical dispersal stages of their life cycles (*see* Life Cycles, Volume 2); and (b) by destroying the habitat of, and consequently reducing the impacts of, many competing species and natural enemies (diseases, predators, parasites). Such increases in survival have been large enough to produce major pests from what were once relatively unimportant species (e.g., Colorado potato beetle, *Leptinotarsa decemlineata*) which attacked only scattered native plants. Modern cultivation practices including irrigation, fertilization, and harvesting methods which leave much residual plant litter, often improve pest (e.g., cereal

rust, *Puccinia* spp.) reproduction or survival during critical parts of the life cycle. Crop protection with pesticides (e.g., DDT) toxic to competing species and natural enemies of a pest (e.g., California red scale, *Aonidiella aurantii*) may seriously impede these regulating factors over the long term. The volume of modern transportation of people, merchandise, and organisms, both accidentally and purposefully, introduces many exotic organisms (e.g., potato late blight, *Phytophthora infestans*) into ecosystems where the introduced species lack effective competitors and natural enemies. In some cases, pest problems initiated by human activity have been serious and extensive enough to cause human responses on scales constituting acceleration of further global change. For instance, after the potato (*Solanum* spp.) became widely established as the major food crop in Northern Europe, particularly in Ireland, late blight was also brought back from the New World, presumably in some diseased potatoes. When late blight epidemics destroyed Northern European potato crops between 1830 and 1850, the resulting widespread famine prompted massive shifts in human populations and migrations to North America.

Global Change Impacts on Pest Management

Global change can be expected to affect the practice of pest management in many ways. First, in addition to persisting pest problems, new ones are expected to arise as a result of the ever increasing scope of human activities. As discussed above, these activities have historically included extensive changes in land use (e.g., monoculture), disruption of native biodiversity (e.g., through species introductions and ecosystem simplification), and steadily rising resource usage. The accelerating pace of global industrialization and human population growth give every reason to expect these trends towards increasing pest problems to persist.

Other aspects of global change, such as changes in atmospheric composition and climate, could also have profound in uences on future pest problems. Many pest species (and presently innocuous species which will become future pests) will likely benefit from better survival with warmer climates in their current range and may well be able to invade previously hostile territory once the climate there becomes more suitable for them (*see* Life Cycles, Volume 2). Longer growing seasons may allow some pests (e.g., aphids) to increase the number of generations they have each year. This may lead to increased genetic adaptability of these populations as well as larger population sizes. Some pest species (e.g., certain plant pathogens, slugs) which prefer cool, damp climates may become relatively harmless as their populations wilt in warmer climates. In agriculture and forestry, some crops and trees may grow larger and faster as CO_2 levels rise, but the duration of this benefit is uncertain. On the other hand, long-lived forest trees may become stressed and have difficulty completing their life cycles as the climate warms. Stressed trees are often more vulnerable to pest attack, and if a tree population is already having difficulty completing its life cycle, losses to a pest outbreak could extirpate the population from the area. The complexities of the interactions make it exceedingly difficult to predict how currently established pest problems will respond to global change, let alone forecast the future, presently unforeseen pest problems.

Besides altering the nature of pest problems, global change is also likely to affect approaches to pest management through shifts in the perceived benefits and costs of most control methods. For instance, with increasing pressure for environmentally benign pest control methods, pesticides must become ever more selective, short-lived, and accurately applied. Recent advances include the development of microbial insecticides, chemical formulations based on organisms (e.g., bacteria and viruses) pathogenic to the pest. The pathogenic components of these formulations are not self-sustaining in the environment so microbial insecticides, like chemical insecticides, must be resprayed as required. Biotechnology and genetic engineering promise tools for further developing the effectiveness and selectivity of microbial pesticides and plant resistance, but use of these tools has become hotly contested because of perceived threats to human health and biodiversity. Extensive testing will likely be required for both political and scientific reasons before many genetically engineered pest control products become operationally available. Climate change could have profound effects on biological and cultivation control methods. For instance, changes in the synchrony of seasonal development rates could disrupt predator–prey and host–parasite relationships and thus nullify attempts at manipulating natural enemy effects. In short, the optimal mix of pest control methods in an IPM program is likely to be continually shifting as global change progresses. Successful IPM programs will be those with the capacity to adapt to these shifts.

REFERENCE

Hall, R W, Ehler, L E, and Bisabri-Ershadi, B (1980) Rate of Success in Classical Biological Control of Arthropods, *Bull. Entomol. Soc. Am.*, 26, 111–114.

FURTHER READING

Bellows, T S and Fisher, T W, eds (1999) *Handbook of Biological Control*, Academic Press, San Diego, CA, 1–1046.

Cannon, R J C (1998) The Implications of Predicted Climate Change for Insect Pests in the UK, with Emphasis on Non-indigenous Species, *Global Change Biol.*, 4, 785–796.

Kommedahl, T and Williams, P H, eds (1985) *Challenging Problems in Plant Health*, American Phytopathological Society, St Paul, MN, 1–538.

Southwood, T R E (1991) *Ecological Methods: with Particular Reference to the Study of Insect Populations*, 2nd edition, Chapman and Hall, London, 1–524.

Stern, V M, Smith, R F, van den Bosch, R, and Hagen, K S (1959) The Integrated Control Concept, *Hilgardia*, 29, 81–101.

Inter-American Institute for Global Change Research (IAI)

see IAI (Inter-American Institute for Global Change Research) (**Volume 4**)

Intergovernmental Oceanographic Commission (IOC)

see IOC (Intergovernmental Oceanographic Commission) (**Volume 4**)

Intergovernmental Panel on Climate Change (IPCC): an Historical Review

John Houghton[1], Richard Moss[2], and James P Bruce[3]

[1] *IPCC, Geneva, Switzerland*
[2] *IPCC, Geneva, Switzerland*
[3] *IPCC, Geneva, Switzerland*

The IPCC was established in 1988, under the joint sponsorship of the World Meteorological Organization (WMO) and the United Nations Environment Programme (UNEP), to provide assessments of scienti c and technical information on climate change, its impacts, and potential responses. Through its reports and methodologies, workshops, and expert meetings, the IPCC has become a primary source of current, independent scienti c information on climate change, its potential impacts, and options for response. This article describes the evolution of IPCC and of its working groups.

INTRODUCTION

It has been known for well over 100 years that the presence in the atmosphere of gases such as carbon dioxide (CO_2) leads to a warming of the Earth's surface. In 1896, Svante Arrhenius, a Swedish chemist, first computed the amount of warming that would occur should the amount of CO_2 in the atmosphere double. In the 1960s, observations made by Charles Keeling from the Mauna Loa Observatory on a mountaintop in Hawaii showed that background atmospheric CO_2 concentrations were indeed increasing significantly. However, it was not until the 1980s that it began to be generally recognized that increased human activities involving the burning of fossil fuels (coal, oil and gas) was resulting in a large increase in atmospheric CO_2 and a probable doubling of its pre-industrial concentration before the end of the 21st century. The possibility of serious consequences to the world's climate was increasingly raised. A comprehensive account of global warming and climate change can be found in Houghton (1997).

By the mid-1980s, the issue of climate change induced by human actions which augment greenhouse gases, especially CO_2, was moving on to the agendas of governments and the business community. An important scientific meeting in Villach, Austria in 1985 (*see* Villach Conferences, Volume 4) reviewed the evidence for global warming and concluded *inter alia* that concentrations of all the greenhouse gases including CO_2, methane and nitrous oxide were increasing and significantly changing the natural radiation balance at the earth's surface. The results of the Villach 1985 meeting did much to alert governments to the seriousness of the issue.

UNEP and the WMO, which had co-sponsored the Villach session jointly with the non-governmental International Council of Science (ICSU), then established (1986) a scientific advisory group on greenhouse gases (AGGG). Each of the sponsors named two representatives to the AGGG, all highly respected scientists in this field and Dr F Kenneth Hare of the University of Toronto was appointed as chair. The AGGG's work produced increasing evidence that without remedial action, significant damage might occur from climate change, and/or that costly adaptation measures might be needed. In addition, remedies to reduce greenhouse gas emissions would involve major changes in global energy and industrial systems.

Contributing to rising media attention in the late 1980s about the possibility of irreversible climate change occurring within the lifetimes of our children was the conclusion of an international conference held in Toronto in 1988, which, for the first time, pressed for specific international action to mitigate climate change. It was in that year too that world leaders began to speak out; for instance, Mrs Thatcher expressed her concern about climate change in a speech to the Royal Society of London that was widely publicized. However, many governments, led by that of the

US, became uneasy about a situation in which the main international advisory body on the subject was a group of academic and independent scientists, however eminent. Instead, governments wanted a science assessment process that would involve them, as well as industry and developing countries likely to be most affected by rising sea level. It was therefore timely that in 1988 a new scientific body to address the issue, the IPCC was set up jointly by the WMO and UNEP. Bert Bolin from Sweden, a scientist with a substantial record of contributions to the science of climate change, agreed to chair the IPCC, which established three working groups to facilitate its work.

- Working Group I: the science of anthropogenic climate change.
- Working Group II: the impacts of climate change.
- Working Group III: response strategies.

A secretariat was established at WMO Headquarters, Geneva, under the direction of N Sundararaman. Over the years, the terms of reference of Working Group I have remained unchanged, but those of Working Groups II and III gradually evolved, as shown by the titles of the published assessment reports:

Working Group II:

- 1990 assessment: climate change: impacts assessment
- 1995 assessment: impacts, adaptation and mitigation of climate change
- 2000 assessment: impacts, adaptation and vulnerability

Working Group III:

- 1990 assessment: climate change: response strategies
- 1995 assessment: economic and social dimensions of climate change
- 2001 assessment: mitigation

The First Assessment (1990)

A *First Assessment Report* (FAR) was published in 1990, the initial emphasis being on the natural sciences. The most successful parts of this report were those on the science of the climate system and on the physical impacts of a changing climate. This report was in uential in development of the United Nations (UN) Framework Convention on Climate Change (FCCC), signed at the UN Earth Summit in Rio de Janeiro in 1992 (*see* United Nations Framework Convention on Climate Change and Kyoto Protocol, Volume 4). the IPCC Impacts Assessment, brought together the then-emerging literature on effects of climate change on physical and economic systems worldwide. A third volume, on response strategies, was also prepared for the FAR. However, this part of the assessment became heavily politicized, and while the volume contains valuable insights on means of reducing greenhouse gas emissions and increasing sinks, it was unable to provide the sort of consensus achieved in the other parts of IPCC 1990.

The Relation between IPCC and FCCC

The UN/FCCC, signed by close to 160 countries at the Earth Summit in Rio de Janeiro in June 1992, sets the context in which international discussion regarding appropriate action can be pursued. The development of the Convention's agenda clearly requires continuous scientific and technical input. The Conference of the Parties (COP) to the FCCC set up a subsidiary body for science and technological advice (SUBSTA) to organize this input (*see* Conference of Parties, Volume 4). The IPCC works closely with SUBSTA through a joint working group to ensure that IPCC assessments are geared to provide the detailed scientific and technical input required.

The objective of the FCCC is contained in Article 2. It recognizes the need to prevent continued change of the climate. It reads as follows:

> the ultimate objective of this Convention and any related legal instruments that the COP may adopt is to achieve, in accordance with the relevant provisions of the Convention, stabilization of greenhouse gas concentrations in the atmosphere at a level that would prevent dangerous anthropogenic interference with the climate system. Such a level should be achieved within a time frame sufficient to allow ecosystems to adapt naturally to climate change, to ensure that food production is not threatened and to enable economic development to proceed in a sustainable manner.

The Second Assessment (1995)

The IPCC restructured itself for its *Second Assessment Report* (SAR), which was completed in 1995. While Working Group I continued to focus on the climate system, Working Group II was asked to consider impacts and responses together (both adaptation and mitigation) because human adaptive responses could be expected to reduce the magnitude of the potential impacts. Working Group III focused on economic issues related to both impacts and mitigation.

The Third Assessment (2001)

Background preparations for the *Third Assessment Report* (TAR) started in September 1996 at a workshop for members of the research community convened at Imperial College, London (IPCC, 1996). This workshop developed several ideas for the structure of the report and how to improve coordination across the three working groups, particularly in the use of emissions and climate scenarios. The suggestions were incorporated into a white paper prepared by Robert Watson, IPCC chair, which was circulated for several rounds of review by both experts and government representatives (Watson, 1997). The government

representatives of the IPCC approved general plans and procedures for the TAR in September 1997 (IPCC, 1997). Members of the research community drafted detailed outlines for the chapters in a meeting in late June 1998, and governments approved these outlines and lists of lead authors to prepare them in October 1998 (IPCC, 1998). This represented more than two years of effort before the assessment even started.

The Third Assessment was tasked with covering the complete range of scientific, technical, economic, and social issues associated with the climate system and climate change deemed important by the expert or policy-making communities. Working Group I was again asked to focus on the scientific aspects of the climate system and climate change. Working Group II was given the charge of assessing the scientific, technical, environmental, economic, and social aspects of the vulnerability of ecological systems, socio-economic sectors, and human health to climate change. Working Group III was asked to review aspects of the mitigation of climate change, and a synthesis report was planned to provide a policy-relevant synthesis and integration of the three working group reports. The reorganization of the panel's Working Groups II and III was motivated by a desire to join scientific and technical aspects of impacts and mitigation assessment with economic analysis of these topics.

A series of crosscutting issues were also identified, including development, sustainability, and equity, costing methodologies, decision analysis frameworks, and uncertainties. Four papers were prepared by authors of the TAR on these issues to provide guidance on achieving more consistency across the working groups. For example, the guidance paper on uncertainties sought to develop common approaches for assessing and reporting on key uncertainties, so that users of the three working group reports would find similar terms used to describe the level of confidence associated with the TAR's findings. The papers were also intended to make the preparation of the synthesis report, which would need to draw on findings from all three groups, a process which would be facilitated if information was provided in a common format and based on agreed overall approaches. Working Group II played an important role in these papers, particularly the guidance on uncertainties. (Moss and Schneider, 2000).

THE ASSESSMENT PROCESS

Richard Moss and John Houghton

In its structure, the IPCC followed closely the structure of the assessment panels of the Montreal Protocol, which had been set up in 1987 by UNEP and WMO to address the problem of the depletion of stratospheric ozone by chloro uorocarbons (CFCs) and related chlorine-containing chemicals – another example of global pollution. Through the negotiation of the Protocol, with its arrangements for inputs from scientists and other experts, methods had begun to be developed in the international community through which global pollution problems could be assessed. Although the issue addressed by the Montreal Protocol was a more limited one than that of global climate change, especially in the range and size of the human activities that contribute to it, it was appropriate that the IPCC should build on this experience in addressing climate change issues. The development within the IPCC of ways to involve large numbers of scientists and of formal procedures for peer review in turn uenced the ongoing work of the assessment panels of the Montreal Protocol.

Recognizing the complexity and broad scope of the climate change issue as well as the potential for controversy, the IPCC process was developed around a set of principles meant to ensure accuracy and wide acceptance of its assessments of the state of knowledge. These principles include:

- openness and transparency: make draft reports freely and publicly available to experts and stakeholders for review and comment;
- comprehensiveness: provide information on all aspects of the climate change issue;
- inclusiveness: assess the full range of scientifically-grounded views, establishing consensus when possible, highlighting scientific disagreements when contradictory evidence cannot be reconciled, and characterizing uncertainties;
- policy relevance and neutrality: present information relevant to decision makers without making policy recommendations.

In practice, the IPCC works at two overlapping levels: as a formal intergovernmental body, and as a scientific and technical assessment body. At the intergovernmental level, approximately 120 governments participated in the IPCC process in the year 2000. In addition, several thousand scientists were involved in the drafting and reviewing the panel's reports. The assessment process includes five main steps:

- Commissioning of reports: intergovernmental sessions of government delegates approve the terms of reference of IPCC reports, workshops, and expert meetings, to ensure their relevance to decision-making needs. Scientific experts meet in Working Group sessions to approve detailed outlines and work plans.
- Drafting of technical materials: the chapters of the reports – highly technical in nature – are prepared by members of the research community based primarily on information in the open scientific literature. Contributions to the report are broadly solicited. Any expert can

contribute so long as the basis for his/her contribution is a peer-reviewed scientific article.
- Drafting of summary for policymakers: the summaries for policymakers present the most policy-relevant conclusions from the technical chapters. They are jointly drafted by the government representatives and expert lead authors – whose participation ensures consistency between the summaries and the underlying technical chapters.
- Review: drafts go through a two-stage review process. The first stage is an expert review, after which the drafts are revised and distributed for simultaneous government and expert review. Review editors ensure that all review comments are dealt with objectively by the authors in the review process.
- Acceptance and approval: each summary for policymakers is approved by governments on a line by line basis. The lead authors participate in the meetings to approve the summaries to ensure that any changes are consistent with the findings of the chapters and that the information is balanced. The panel accepts underlying technical chapters if they are judged to have been prepared according to IPCC procedures and represent a balanced assessment of the state of knowledge.

Structuring the process in this way results in the insulation of the evaluation of the scientific literature and drafting of the technical chapters of the report from the political processes of approving the final reports and the summaries for policymakers. Scientists are given freedom to prepare the reports on the basis of standard scientific peer-review principles. Political review and approval of the summaries involve interpretation and negotiation by government representatives, under the watchful eye of the scientific authors. This political review creates buy-in by governments, who are more likely to trust the conclusions and hence to use them in formulating government policies. As a result, the reports are more in uential than they might otherwise be. There are several points of interaction between the two levels, including the initial development of outlines, the preparation of the summary for policymakers, the government review process, and the intergovernmental meetings at which the reports are finally approved.

THE ACTIVITIES OF IPCC WORKING GROUP I

John Houghton
The IPCC Working Group I 1990 Scientific Assessment (IPCC, 1990)

It was agreed at the first meeting of the IPCC that a new scientific assessment of the whole issue of anthropogenic climate change should be prepared. There had, of course, been assessments before of the climate change issue, notably one published in 1986 organized under the auspices of SCOPE (the Scientific Committee on Problems of the Environment) (SCOPE, 1986) of ICSU again under the chairmanship of Bert Bolin. The IPCC saw its task as updating previous assessments but with a difference. Previous assessments had involved relatively few of the world's leading climate scientists. Because of the global nature of the issue that brought with it a large measure of international concern, the IPCC's ambition from the start was to involve as many representatives as possible from the world scientific community in the new assessment.

To assist in the preparation of the Working Group I report, a small technical support unit was set up within that part of the UK Meteorological Office at Bracknell, which was concerned with climate research. The report comprised eleven chapters totaling over 300 pages dealing with different components of the scientific issue together with a policymakers' summary and an executive summary. Twelve international workshops were held to address these different components. One hundred and seventy scientists from 25 countries contributed to the report either through participation in the workshops or through written contributions. A further 200 scientists were involved in the peer review of the draft report. The thorough peer review played an important part in ensuring the achievement of a high degree of consensus amongst the authors and reviewers regarding the report's conclusions.

A policymakers' summary (20 pages) together with its executive summary (2 pages) were prepared for those without a strong background in science, providing a clear statement of the status of scientific knowledge at the time and the associated uncertainties. In preparing the first draft of the policymakers' summary, the lead authors were first involved; it was then sent out for the same wide peer review as the main report. A revised draft of the summary was then discussed line by line at a plenary meeting of the working group attended by government delegates from 35 countries together with lead authors from the chapters. The final wording was agreed at that meeting. A avor of the style and content of the report is given by the first few paragraphs of the executive summary that read as follows:

> We are certain of the following:
>
> - there is a natural greenhouse effect which already keeps the Earth warmer than it would otherwise be;
> - emissions resulting from human activities are substantially increasing the atmospheric concentrations of the greenhouse gases: CO_2, methane, CFCs and nitrous oxide. These increases will enhance the greenhouse effect, resulting on average in an additional warming of the Earth's surface. The main greenhouse gas, water vapor, will increase in response to global warming and further enhance it.
>
> We calculate with confidence that:
>
> - some gases are potentially more effective than others at changing climate, and their relative effectiveness can be

estimated. CO_2 has been responsible for over half the enhanced greenhouse effect in the past, and is likely to remain so in the future;

- atmospheric concentrations of the long-lived gases (CO_2, nitrous oxide and the CFCs) adjust only slowly to changes in emissions. Continued emissions of these gases at present rates would commit us to increased concentrations for centuries ahead. The longer emissions continue to increase at present day rates, the greater reductions would have to be for concentrations to stabilize at a given level;
- the long-lived gases would require immediate reductions in emissions from human activities of over 60% to stabilize their concentrations at today's levels; methane would require a 15–20% reduction.

Based on current model results, we predict:

- under the IPCC business as usual (scenario A) emissions of greenhouse gases, a rate of increase of global mean temperature during the next century of about 0.3 °C decade^{-1} (with an uncertainty range of 0.2–0.5 °C decade^{-1}); this is greater than that seen over the past 10 000 years. This will result in a likely increase in global mean temperature of about 1 °C above the present value by 2025 and 3 °C before the end of the next century. The rise will not be steady because of the influence of other factors.

Later sections of the summary addressed the scientific uncertainties and the question of the degree to which anthropogenic climate change had been observed in the climate record.

Over the period of the preparation of the report, a significant change occurred in the attitudes of the scientists involved. To begin with, many scientists felt that the scientific uncertainty was too large for any useful statement to be made regarding future climate change. However, increasingly the scientists realized their responsibility to articulate carefully and honestly the available knowledge. This responsibility seemed all the more important because many individuals not possessing much expert scientific knowledge were making forecasts of future climate change – often of an extreme kind. Also, it was increasingly recognized that there was enough certainty in the science to provide meaningful information regarding the likely future provided that the uncertainty was also fully explained.

Many of the world's leading scientists in the field of climate and climate change contributed to the report. Inevitably they came mostly from developed countries although a significant number of contributors from developing countries were also involved. That so many of the world's scientists contributed or were involved in the review process meant that there was a genuine feeling of ownership of the report by the world scientific community.

The IPCC process led to a significant degree of consensus. It is sometimes pointed out that consensus amongst scientists is not necessarily a sign of scientific health; argument and disagreement are seen to be more usual building blocks of scientific advance. But the consensus achieved by the IPCC is not complete agreement about everything. It is agreement particularly about what we know and what we do not know – distinguishing clearly those matters about which there is reasonable certainty from those where there remains much uncertainty and where there continues to be lively debate and disagreement. It is this limited consensus that is reflected in the executive summary of the 1990 IPCC Report, which has been widely acclaimed for the clarity and crispness of its presentation.

It was clear from an early stage that not only was the scientific content of the assessment important but also the way in which it was presented. Scientists left to themselves do not always recognize what is relevant to policymakers, or present their material with maximum clarity. Further, the presentation of a scientific document can appear to a policymaker to convey a political message even though none was intended.

It has therefore been helpful in the formulation of summaries of IPCC Reports to involve policymakers or their representatives. For instance, they were full participants in the government review process and in the Working Group Plenary Meeting, which agreed the wording of the summary of the 1990 Working Group I scientific assessment. The report was greatly improved in its relevance and clarity through their participation. In addition because governments participated in the process, they felt a sense of ownership of the report.

The 1992, 1994 and 1995 IPCC Reports of Working Group I

As soon as the 1990 report was complete, the IPCC began work on further reports. In 1992, in time for the Earth Summit, Working Group I produced a report updating what was known about greenhouse gases (Houghton *et al.*, 1992), their sources and sinks, and about observations and modeling of climate change. In addition, the 1992 report developed various emission scenarios for the emissions of greenhouse gases over the 21st century, based on a variety of assumptions regarding factors such as world population, economic growth, availability of fossil fuels, etc.

The 1994 report updated the information and analysis regarding the radiative forcing of the various greenhouse gases (Houghton *et al.*, 1996a). Of particular importance was the new work carried out on the profiles of emissions of CO_2 and other greenhouse gases, which would lead to the stabilization of these gases in the atmosphere at different concentration levels.

By 1995, when the IPCC produced its second comprehensive assessment (Houghton *et al.*, 1996b; Watson *et al.*, 1996; Bruce *et al.*, 1996), 5 years after the first one in 1990, the community of scientists involved with the IPCC had become substantially greater. More scientists from more

countries participated both in the report's preparation (about 480 scientists from more than 25 countries) and in its review (over 500 from 40 countries). The participants at the plenary meeting of Working Group I that approved the summary for policymakers included 177 delegates from 96 countries, representatives from 14 non-governmental organizations and 28 lead authors.

Regarding climate change over the 21st century and its likely impacts, the messages of the 1995 report were essentially the same as those of the 1990 report. Some further detail had emerged during the five years in between, especially regarding the likely contribution to climate change from atmospheric aerosols. Also, there was more confidence amongst scientists that the historical increase in global average temperatures that had occurred since wide-scale industrialization might be the result of anthropogenic change. The Working Group I plenary meeting debated for a considerable length of time how to express this somewhat greater scientific confidence in the interpretation of the recent climate record – although still surrounded by much uncertainty. The delegates unanimously agreed a sentence carefully crafted by the meeting: the balance of evidence suggests a discernible human in uence on global climate.

The IPCC Working Group I 2000 Third Scientific Assessment

The IPCC 2001 Third Scientific Assessment has comprehensively updated the 1995 report. When work began, it had been envisaged that, being an update, it would be shorter than its predecessors. However, it soon became clear that during the five-year period, the amount of research had been rapidly growing as had the volume of published material. Despite efforts to keep it short, it was not possible to do justice to the scientific advances and the report was substantially longer than the 1995 one.

Despite the increasing demands on their time made by the IPCC, many scientists were again involved in the preparation of the report, including many of the world's leading experts: 122 scientists contributed as lead authors, 515 as contributing authors, 21 as review editors, 337 as reviewers in the expert review – other reviewers were involved in the subsequent government review. At the plenary meeting of the working group held in Shanghai in January 2001, delegates from 99 governments attended, as well as about 50 lead authors.

The Third Scientific Assessment of Working Group I contains three new and important conclusions.

1. New observational evidence is presented that the world is warming. In particular, the decade of the 1990s was probably the warmest decade globally since instrumental records began around 1860. In addition, new analyses of proxy data for the Northern Hemisphere indicate that the last decade of the 20th century was the warmest of the last millennium. Further evidence for warming comes from the recent melting of glaciers and of decreases in the extent of snow cover and ice.
2. A stronger statement regarding the detection of anthropogenic climate change in the observational record was agreed compared to that made in 1995. Careful observational and modeling studies of the climate of the 20th century carried out since 1995 have provided new and stronger evidence that most of the warming observed over the last 50 years is attributable to human activities.
3. Projections of the likely change in global average temperature during the 21st century were larger than in 1995, mainly because of the substantial reduction in the expected sulfur dioxide emissions because of the impact of air pollution controls. There would therefore be less cooling occurring by sulfate particles and less offset to the projected warming by greenhouse gases (*see* Climate Change, Detection and Attribution, Volume 1).

THE ACTIVITIES OF IPCC WORKING GROUP II

Richard Moss

The IPCC Working Group II 1990 FAR: Initial Assessment of Impacts

Preparation of the Working Group II contribution to the FAR (1990) involved over 150 specialists from 24 countries. Leading the effort were Yuri A Izrael (Union of the Soviet Socialist Republics) and W J McG Tegart (Australia). The report focused on several broad impacts categories: agriculture and forestry; natural terrestrial ecosystems; hydrology and water resources; human settlements, energy, transport, and industrial sectors, human health, and air quality; oceans and coastal zones; and seasonal snow cover, ice, and permafrost. Overall, the report showed that climate change would likely be of great importance for the future development of mankind and the economy, although it would not lead to any catastrophic socio-economic consequences on global scales. The report did not anticipate any adaptation, technological innovation, or any other measures to diminish the adverse effects of climate change because this was considered to be the responsibility of Working Group III.

Preparation of the report was made difficult by the fact that Working Group II received recommendations on climate scenarios from Working Group I only late in the preparation of the impacts report. The timing of provision of scenarios, coupled with the known shortcomings of models – the extent to which they incorporate water vapor feedbacks, effects of aerosols, sudden doubling rather than transient increase of greenhouse gas concentrations, and high variance across models at regional scales – led

the Soviet co-chair to champion the use of paleoclimate analogues. The paleoclimate analogue method assumes that past warm geologic intervals provide insight into possible future climate conditions. A controversy erupted within the working group on the validity of this approach, focusing on a number of shortcomings of the method (Agrawala, 1998a,b). In addition to the use of paleoclimate analogues, authors made use of some of the climate scenarios published in the literature (Izrael, 1991). For the most part, these scenarios posited a more extreme set of changes than those ultimately suggested by Working Group I's business-as-usual scenario. The impacts based on these more dire scenarios were seen by Working Group II as a warning of the consequences of continued uncontrolled emissions (Working Group II FAR, Scenarios chapter, p. 1).

The IPCC Working Group II 1995 SAR: Integration of Impacts and Responses

The IPCC restructured itself for its SAR, which was completed in 1995. Working Group II was asked to consider impacts and responses together (both adaptation and mitigation) because human adaptive responses could be expected to reduce the magnitude of the potential impacts. (Working Group III focused on economic issues related to both impacts and mitigation.) Working Group II's contribution to the SAR was co-chaired by RT Watson (USA) and MC Zinyowera (Zimbabwe). Over 500 scientists from 55 countries participated as authors – including a greatly expanded set of researchers from developing countries.

The purview of the group was massive: primers on ecological processes and the energy system; 18 impact chapters including several that were focused on predominantly natural ecosystems, several on human activities or systems such as agriculture and human population health, and several on environmental processes such as hydrology and desertification. In addition, the report also included seven chapters on technical options for reduction of emissions in broad sectors of the economy including energy, industry, transportation, human settlements, agriculture, forestry, and cross-cutting issues such as geo-engineering and land use. Finally, the nearly 900-page volume included technical appendices on methods for assessing impacts, mitigation options, and an inventory of technologies, methods, and practices. Responsibility for these efforts was distributed across four sub-groups, each with its own set of co-chairs.

Working Group II introduced several innovations to the IPCC assessment of impacts, drawing on a formal methodology for impact assessment prepared under the IPCC's auspices during 1993–1994 (Carter *et al.*, 1994). Considering adaptive responses as a part of the analysis of impacts was felt to provide a more realistic picture of potential adverse consequences of climate change. In addition, more attention was given to potential beneficial effects of climate change, for example increased water use efficiency in some types of crops as a result of increased atmospheric concentrations of CO_2. To the extent permitted by the literature, authors also attempted to integrate bio-physical changes such as increases or decreases in crop yields or water available with potential consequences for socio-economic systems and resources.

Confidence in regional projections of temperature, precipitation, soil moisture, extremes, and other parameters important to impact assessment remained low. In addition, Working Group II confronted two specific challenges in obtaining state-of-the-art climate scenarios for use in quantitative modeling of impacts. First, regional projections of climate would be affected by aerosols as well as greenhouse gas emissions, but aerosols were only beginning to be incorporated into climate change modeling. Second, models using gradual increases in greenhouse gas concentrations (transient scenarios) rather than a sudden doubling of concentration were just being prepared and it was considered important to use these new scenarios for understanding the effects of a gradual transition in climate. These issues shaped several aspects of the Working Group II SAR. Working Group II, in close collaboration with Working Group I, compiled a set of transient climate scenarios with aerosol forcing developed in conjunction with three modeling centers (Geophysical Fluid Dynamics Laboratory (US), Hadley Centre (UK), and Max Plank Institute (Germany)) and distributed these to authors for use in the report (Greco *et al.*, 1994). To distinguish between those uncertainties arising from unknowns related to responses of systems to climate change from those related to the regional-scale climate projections themselves, Working Group II focused on the vulnerability of different systems and sectors to the level of change posited in these scenarios. The analysis integrated information on the sensitivity of response of systems to climate change in a wide variety of functional forms ranging from quantitative relationships to more qualitative analysis and the adaptive capacity of the systems, i.e., the degree to which adjustments in practices, processes, or structures could reduce adverse effects or take advantage of opportunities. In addition to attempting to distinguish among the different sources of uncertainty, Working Group II also tried to apply a consistent set of confidence levels to re ect the amount of consistent information and the degree of author consensus that supported the main findings of the report.

The Working Group II contribution to the SAR reached several overarching conclusions, especially that human-induced climate change is an important added stress to many natural and human systems already threatened by pollution, increasing resource demands, and practices ill-suited to current climate conditions, including variability. Quantitative estimates were prepared for geographic shifts in ecosystems, for changes in food production, for shifts in incidence of vector-borne diseases, and numerous other

potential effects. General conclusions on adaptation and the conditions necessary to facilitate successful adjustment to changing climate were reached in the summary for policymakers, although the individual chapters did not include as much information on this issue as some expected. Finally, detailed options for reduction of greenhouse gas emissions were reviewed, with the overall conclusion that 10–30% efficiency gains above present levels are available in many parts of the world at little or no economic cost. Integrated energy demand and supply scenarios were also evaluated as part of the report to explore the contributions of different technologies to low emission energy futures.

The report was ultimately seen as an advance (Kates, 1997), particularly in areas such human population health impacts, but was still deficient in several important respects that affected plans for subsequent reports. Attempting to analyze the effects of climate change at a global scale provided important generic information about basic relationships and processes in different systems and sectors, but it did not take account of the regional differences in environmental, social, and economic circumstances around the globe. This was particularly a concern because the circumstances of different countries would lead to very different sensitivities, adaptive capacities, and vulnerabilities, all of which needed to be factored more explicitly into the assessment. In addition, there was insufficient information on adaptation options contained in both the literature and the SAR. While the report presented a vast amount of information on technical options, separating the socio-economic analysis of impacts and mitigation costs into a separate working group (Working Group III) resulted in policymakers receiving seemingly different messages on the magnitude of the social and economic costs of impacts and mitigation measures. Finally, integrating impacts and mitigation options across so many sectors and regions was also seen as overwhelming the analytic framework of the report. The chapter on cross-cutting issues attempted to deal with macro-scale questions such as whether, in the face of climate change, there would be sufficient land, soil, and freshwater resources to support a growing population with changing dietary preferences while simultaneously increasing land use for biomass energy production or sequestration of carbon in forests. But the framework for integrated assessment was insufficient to come to grips with this issue, even in the view of lead authors (Schneider, 1991).

Targeted IPCC Reports: Meeting the Information Needs of UN/FCCC Negotiators

In the period immediately following the completion of the SAR, the IPCC began to respond to the emerging needs of the UN/FCCC, which had entered into force in April 1994 with the first meeting of COP-I in Berlin. The primary point of interaction was the Convention's subsidiary body on scientific and technological advice, which was seen as the conduit through which the IPCC SAR would be provided to the full Convention, and through which subsequent requests for reports, methodologies, and other information would be directed to the IPCC. Additional reports were requested of the IPCC in early 1996 to support the Convention's consideration of whether existing commitments were adequate to meet the Convention's goal of avoiding dangerous anthropogenic interference in the climate system. A series of four technical papers – documents that synthesized information from previous IPCC reports and applied them to a particular question or decision context – resulted. In addition, special reports – assessments that reviewed new literature but had a more specific orientation than the panel's periodic comprehensive assessments – were also prepared on a number of topics, including the impacts of aviation, the regional impacts of climate change, emissions scenarios, and technology transfer.

Lead responsibility for a number of these requests fell primarily to Working Group II. One of the technical papers was an assessment of *Technologies, Policies, and Measures for Mitigating Climate Change* (IPCC, 1996) that married the technical studies of mitigation potential from the Working Group II SAR with the assessment of policies that would affect the economic and market feasibility under different circumstances from Working Group III. A second was a special report on the *Regional Impacts of Climate Change: an Assessment of Vulnerability* (Watson *et al.*, 1997) that explored the potential consequences of climate change and other environmental stresses for 10 continental or subcontinental scale regions. This report represented an important step in the evolution of the impact assessment process for the IPCC in that previous reports analyzed the effects of climate change by sector, rather than by looking at how change would affect all of the important sectors in a particular region in an integrated fashion. Working Group II, principally through its technical support unit, also played an important role in a number of the other technical papers and special reports of this period.

The IPCC Working Group II 2000 TAR: Integrating Sectors and Regions

The Working Group II contribution to the Third Assessment Report provided an update to the sectoral treatment of impacts contained in the SAR (IPCC, 1995), and to the regional focus of the Special Report on Regional Impacts of Climate Change (IPCC, 1998). The Working Group was co-chaired by Dr O Canziani (Argentina) and Dr J McCarthy (USA). The report was compiled by 183 lead authors and 243 contributing authors between July 1998 and February 2001. Comments were received from 440 reviewers, with guidance on assimilating these comments provided by 33 review editors. The revised report was accepted, and the accompanying summary for policymakers

approved in detail, at a session of the Working Group held in Geneva, Switzerland, from 13 to 16 February 2001 (IPCC, 2001).

The literature on impacts grew considerably in the years since publication of the SAR, resulting in a substantial increase in knowledge of vulnerabilities for a variety of ecological and human systems. One of the major advances in the TAR over previous reports was increasingly solid observational evidence of changes consistent with warmer temperatures in many physical and biological systems (e.g., continued glacial melting, earlier onset of leafing in plants, shifts in geographic ranges of both plant and animal species, etc.). There was also emerging evidence that some social and economic systems were affected by changes in climate, but confounding effects of changes in demographic and other conditions made these effects much more difficult to quantify. A number of unique systems were identified as especially vulnerable to climate change, including coral reefs and atolls, mangroves, boreal and tropical forests, polar and alpine ecosystems, wetlands, and remnants of native grasslands. Systems were identified as being especially sensitive to potential changes in climate extremes. Multiple interacting stresses from other environmental changes (e.g., acid deposition) or socio-economic conditions (e.g., increases in human population density) were also identified as crucial in determining how vulnerable systems would be to future changes in climate.

The report confirmed earlier findings that the vulnerability of human populations varies substantially by region due to differences in baseline climate, the potential magnitude of change, and the varied socio-economic characteristics and pressures that give rise to different degrees of sensitivity and adaptive capacity. The report included assessments of adaptive capacity, vulnerability, and key concerns for each of eight different regions or groups of countries. Increased emphasis on the components of adaptive capacity (e.g., access to capital, highly educated populations, adequate economic infrastructure), as well as technological options for adaptation (e.g., new drought-resistant crop varieties), confirmed that developing countries, particularly the least developed, are most vulnerable to climate change, due to their relatively reduced access to resources for adaptation.

A synthesis chapter focused on a set of reasons for concern that was intended to help readers in making their own determinations about what constitutes dangerous human interference in the climate system. The reasons for concern included the relationship between global mean temperature increase and (1) loss of unique systems, (2) distribution of impacts geographically, (3) aggregate economic damages, (4) the probability of extreme weather, and (5) the probability of large-scale singular events such as reduction in North Atlantic thermohaline circulation.

Overview of Working Group II Experience

Over its evolution, Working Group II has contributed to the development of assessment methods and products. As Working Group II adopted the rigorous drafting and review process used by Working Group I (with lead authors of technical chapters largely autonomous and responsible for the content of their chapters and several layers of technical and policy review), the quality of the products of the group improved. While some critics assailed the process as politically motivated or involving a biased group of scientists who were more inclined to see climate change as a serious problem, governments and the scientific community continued to find that the process and its products were useful in assessing the scientific literature and applying it to understanding the potential impacts of human activity on climate, the potential impacts of changes in climate on society and the environment, and of options for reducing emissions to reduce the risks of large-scale global environmental change.

THE ACTIVITIES OF IPCC WORKING GROUP III

J P Bruce

The First Assessment (1990)

The initial emphasis of the IPCC was on the natural sciences. There was a volume on response strategies prepared by Working Group III but as already mentioned herein, this part of the assessment became heavily politicized, and while the volume contains valuable insights on means of reducing greenhouse gas emissions and increasing sinks, the report was unable to provide the consensus achieved in the other parts of IPCC 1990.

The Second Assessment (1995)

For the SAR, it was recognized that there was a large and growing climate change literature in the social sciences, especially economics. Thus, Working Group III was reborn in 1992 to more closely parallel, in the social sciences, the literature assessment focus for the natural sciences conducted in Working Groups I and II. The Bureau for Working Group III was selected at an IPCC plenary session in Harare and consisted of co-chairs, Canada (James Bruce from 1993) and Korea (Hoesung Lee), with vice-chairs, Kenya (Richard Odingo) and Norway (Lorents Lorentsen from 1994). The Canadian government supported the formation of a small technical support unit, based in Toronto, under the leadership of Erik Haites.

At the first plenary session of Working Group III, Montreal, May 1993, with some 80 countries represented, an outline was agreed of the main topics that countries wished to see included in the report. This was approved by the full

IPCC Plenary in June 1993. The outline defined sufficiently the scope of the work, to permit requests to be made to countries for nomination of experts to serve as authors of the various chapters. The response to the call for nominations was very encouraging, but failed to include some of the leading economists and other social scientists that had published papers related to climate change. In these cases, countries were encouraged to consider adding to their nominations some of these prominent experts. In the event, a very impressive group of economists was nominated to the writing teams as well as a more limited number of social scientists, engineers and natural scientists. There was unfortunately an unavoidable dominance of US and other Western authors since at that time (1993) they had undertaken most of the modeling efforts.

While the work on the 11 chapters of the TAR proceeded, a further task was required of Working Group III. That was an evaluation of the IS92 emission scenarios. An IPCC team had developed these scenarios of future greenhouse gas emissions over the coming century in 1992, based on projections of population, economic conditions and technology. They were intended as an input to mathematical modeling of future climate conditions, without policy-driven emission limitations. A writing team led by Joseph Alcamo (The Netherlands), compared the IS92 scenarios to other scenarios in the published literature, many prepared after 1992. They concluded that the six IS92 scenarios covered much of the very wide range of likely future emissions and were suitable for climate modeling work. However, the authors noted the high degree of dependence on input assumptions that are not readily testable, and a number of gaps in knowledge. Thus, IPCC Working Group III decided to undertake construction of a new set of emission scenarios. Working Group III appointed a writing team led by N Nakicenovic, International Institute for Applied Systems Analysis, Austria, to undertake this work. The team developed a number of plausible, and internally consistent story lines about the future and derived emission scenarios from these. They became available in 1999.

In preparing the 11 chapters of the TAR, there were many legitimate controversies amongst the lead authors. Among the more lively debates was that between developing and industrialized country authors on decision-making frameworks (Chapter 2). A difference in view also arose between economists over the discount rate that should be applied in considering climate change actions – a social discount rate to take inter-generational aspects into account, or a market rate (Chapter 4). In addition, the different perspectives of top-down econometric modelers and bottom-up sectoral studies on costs of emission reductions resulted in divergent results (Chapters 8 and 9). Differences of views are re ected in the chapters.

However, the controversy that drew most attention, and spread to a more public debate, was in connection with Chapter 6 on the social costs of climate change: greenhouse damage and the benefits of control. In this chapter, the authors cited literature dealing with the statistical value of a human life since the projected number of deaths was part of the damage estimates for a significant climate change. The economics literature uses proxy values for how much society is prepared to pay for saving a life (e.g., building straighter roads to reduce accidents, vaccination programs). These proxy values are higher for wealthy countries than for poorer ones. While this was not a central matter in the chapter or report, it became a matter of lively debate since developing countries clearly could not accept that a life in their country is worth less than a life in industrialized countries.

A Plenary Session of Working Group III (Geneva, July 1995) was held to consider a draft summary for policy makers drawn from the chapters, to which national delegates had to agree. The debate became very heated over Chapter 6, and resulted in the need to hold a second plenary session in Montreal in October 1995. On that occasion, a special task group, chaired by Ambassador John Ashe (Antigua and Barbuda) was struck, which developed summary language that was acceptable to national delegates, but which left economist lead authors somewhat dissatisfied. The finally agreed text for the summary for policy makers read in part:

> There is no consensus on how to value statistical lives or aggregate statistical lives across countries. Monetary valuation should not obscure the human consequences of climate damages, because the value of life has meaning beyond monetary consideration. Some regard valuation of such impacts (human mortality) as essential to sound decision making, but others reject monetary valuation on ethical grounds.

At the plenary sessions of Working Group III in 1995, a number of important conclusions, based on the chapters, were adopted in the summary for policy makers. This summary and the underlying report were accepted by the full IPCC Plenary in Rome, December 1995. Many of the Working Group III findings were also included in the 1995 IPCC Synthesis Report. Some of the important conclusions in the summary for policy makers of Working Group III were the following.

- Estimates of (aggregate annual) damages from $2-3\,^{\circ}C$ global warming tend to be a few percent of world gross domestic product (GDP), with, in general, considerably higher estimates of damage to developing countries as a share of their GDP. Small islands and low-lying coastal areas are particularly vulnerable to projected sea level rise accompanying global warming.
- Climate change policies should not aggravate existing disparities or attempt to redress all equity issues. Climate policy also raises particular questions of equity among generations. Sustainable development is one approach to inter-generational equity, as is the selection

of a suitable discount rate for actions to limit climate change (see above re Chapter 4).

- The literature indicates that significant no-regrets opportunities for greenhouse gas mitigation are available in most countries. No-regrets measures are those whose benefits, such as reduced energy costs and reduced emissions of local/regional pollutants, equal or exceed their cost to society, excluding the benefits of climate change mitigation. They are sometimes known as measures worth doing anyway (*see* No Regrets Principle, Volume 4).
- The risk of aggregate net damage due to climate change, consideration of risk aversion and the precautionary principle, provide rationales for greenhouse gas mitigation actions beyond no regrets.
- Potential for energy efficiency gains of perhaps 10–30% below baseline trends over the next two to three decades can be realized with net economic benefits at zero net cost for society as a whole (an example of no-regrets measures).
- Two main types of models were used to analyze costs of greenhouse gas mitigation. These are referred to as top-down models of the large-scale economy, and bottom-up models, which incorporate more detailed studies by sector of engineering costs. However, it was recognized that these model structures are converging.
- In the specific case of stabilizing emissions at 1990 levels over the next several decades, most top-down studies estimate annual costs in the range of • 0.5% of GDP (equivalent to a gain of about $60 billion in total for Organization for Economic Co-operation and Development (OECD) countries at today's GDP levels) to 2% of GDP (equivalent to a loss of about $240 billion).
- Although it is difficult to generalize, top-down analyses also suggest that the costs of substantial reductions below 1990 levels in OECD countries could be as high as several percent of GDP. However, studies also show that appropriate timing of abatement measures and the availability of low-cost alternatives may substantially reduce the size of the overall bill. Realization of economic double dividends (through use of carbon taxes to reduce more distortionary taxes), and secondary environmental benefits through reduction of local and regional air pollution, can also significantly offset costs of mitigation in some countries.
- Bottom-up studies are more optimistic about the potential for low or negative cost emission reductions, and the capacity to implement that potential. Such studies show that the costs of reducing emissions by 20% in developed countries within two to three decades are negligible to negative (i.e., economically beneficial). Other such studies suggest that there exists a potential for absolute reductions in excess of 50% in the longer term, without increasing and perhaps even reducing total energy system costs to society.
- Studies indicate that global emission reductions of 4–18%, together with increases in real incomes are possible from phasing out fuel subsidies world-wide.
- Appropriate early warning signals are required to allow producers and consumers to adapt cost-effectively to constraints on greenhouse gas emissions and to encourage research and development. Failure to adopt policies as early as possible to encourage efficient replacement investments at the end of the economic life of plant and equipment (i.e., at the point of capital stock turnover) imposes an economic cost to society. Implementing emissions reductions at rates that can be absorbed in the course of normal turnover are likely to be cheaper than enforcing premature retirement now.
- The (economic) value of better information on predictions of, impacts of, and responses to climate change is likely to be great. Research and development on energy-efficient technologies and non-fossil energy options also offer high potential value.
- A prudent way to deal with climate change is through a portfolio of actions aimed at mitigation, adaptation and improving our knowledge. The appropriate portfolio will differ for each country. The challenge is not to find the best policy today for the next 100 years, but to select a prudent strategy and to adjust it over time in the light of new information.

In addition to the contribution to the SAR, summarized brie y above, IPCC Working Group III (1992–1997) also left a legacy of several important workshop reports (see section on Completed IPCC Reports and Projects, below).

A further legacy of the Working Group III contribution to the IPCC Second Assessment was the active involvement in the IPCC process for the first time, of a number of economists, and a lesser number of other social scientists. A criticism was that nearly all of the economists and other social scientists were in one working group and that they needed to be involved in the full range of impacts, adaptation and mitigation measures considered by IPCC. In the structure for the TAR, this was achieved to a greater extent. Another result of the Second Assessment is that the Working Group III findings and the full IPCC Report have clearly in uenced thinking in the economic and policy communities. Indeed, some 400 economists in North America in early 1997 signed petitions to governments on this issue. They said:

> as economists we believe that global climate change carries with it significant environmental, economic, social and geopolitical risks, and that preventive steps are justified. Economic studies have found that there are many potential policies to reduce greenhouse gas emissions for which total benefits outweigh the total costs.

The Third Assessment (2001)

For the Third IPCC Assessment Report, the division of effort was modified somewhat. Working Group II addressed impacts of and adaptation to climate change in physical, social and economic dimensions. Working Group III did the same for mitigation measures (i.e., reducing greenhouse gas emissions and increasing sinks).

The contribution of Working Group III to the TAR casts the climate change issue in the contexts of development, equity and sustainability. The report affirms that technologies now available could reduce global greenhouse gas emission levels close to, or even below those of year 2000 by 2010 and even lower by 2020. Furthermore, known technological options (energy efficiency, alternative fuels, process changes) when combined with socio-economic and institutional changes could in the long term achieve stabilization of CO_2 concentrations at double pre-industrial levels or lower. Many measures can be achieved at little or no net costs, especially when co-benefits of emission reductions or carbon sequestration are taken into account. These include reduction of local air pollution, and water quality benefits of better land and forest conservation. The range of estimates of the costs to OECD countries of implementation of the Kyoto Protocol are from 0.1–1.1% of GDP if a full range of emission trading is permitted (*see* Tradable Permits for Greenhouse Gases, Volume 4). Without trading mechanisms, costs could be considerably higher. However, the costs of inaction could be very much higher for many vulnerable countries. To address climate change, a transition away from conventional fossil fuels is necessary and will require substantial investment, replacing investment in further developing fossil fuel resources.

Working Group III continues to suggest that a portfolio of early actions in a country, including energy efficiency measures, alternative fuels, carbon sequestration and research support increases exibility for the future. Such exibility will be essential in moving towards appropriate contributions, at minimal costs, to the ultimate goal of stabilization of atmospheric concentrations of greenhouse gases.

CONCLUDING REMARKS

The elements that have been critical to the success of the IPCC have been those of:

1. ensuring the widest possible participation by experts from all relevant disciplines;
2. ensuring a clear separation between scientific assessment and policy determination;
3. through the close involvement of the expert community, of governments and, more recently, of relevant industry, ensuring wide ownership not only by the community of experts but also by those who have a stake in the policy process.

Although anthropogenic climate change is perhaps the largest and most complex problem concerned with the global environment that we face, it is not unique. Models similar to that of the IPCC with its essential elements could be applied elsewhere. The development of the work of the IPCC has demonstrated the enormous capacity of the international community to work together towards the common aims of care for humanity and care for the environment and provides encouragement in the belief that problems as complex as that of climate change are capable of resolution.

For complex environmental issues such as climate change, IPCC's experience demonstrates that an interdisciplinary approach is essential to assessing both natural sciences and social sciences, including economics.

COMPLETED IPCC REPORTS AND PROJECTS

Reports

IPCC (2001) *Climate Change 2001: The Scienti c Basis*, Contribution of Working Group I to the TAR of the IPCC, Cambridge University Press, Cambridge.

IPCC (2001) *Climate Change 2001: Impacts, Adaptation, and Vulnerability*, Contribution of Working Group II to the TAR of the IPCC, Cambridge University Press, Cambridge.

IPCC (2001) *Climate Change 2001: Mitigation*, Contribution of Working Group III to the TAR of the IPCC, Cambridge University Press, Cambridge.

IPCC (2000) *Emissions Scenarios*, IPCC Working Group III Special Report, eds N Nakicenovic and R Swart, Cambridge University Press, Cambridge, 1–599.

IPCC (2000) *Land Use, Land-use Change, and Forestry*, IPCC Working Group II Special Report, eds R T Watson, I R Noble, B Bolin, N H Ravindranath, D J Verardo, and D J Dokken, Cambridge University Press, Cambridge, 1–377.

IPCC (2000) *Methodological and Technological Issues in Technology Transfer*, IPCC Working Group II/III Special Report, eds B Metz, O R Davidson, J W Martens, S N M van Rooijen, and L Van Wie McGrory, Cambridge University Press, Cambridge, 1–466.

IPCC (1999) Aviation and the Global Atmosphere, IPCC Working Group I/III Special Rerport, eds J E Penner, D H Lister, D J Griggs, D J Dokken, and M McFarland, Cambridge University Press, Cambridge, 1–373.

IPCC (1998) *The Regional Impacts of Climate Change: an Assessment of Vulnerability*, IPCC Working Group II Special Report, eds R T Watson, M C Zinyowera, and R H Moss, 1998, Cambridge University Press, Cambridge, 1–518.

The Regional Impacts of Climate Change: Summary for Policymakers, available from IPCC Secretariat or electronically: http://www.usgcrp.gov/ipcc/html/RISPM.html.

IPCC (1996) *Technologies, Policies, and Measures for Mitigating Climate Change*, Technical Paper 1, IPCC, Working Group II, eds R T Watson, M C Zinyowera, and R H Moss, available from IPCC Secretariat or electronically: http://www.gcrio.org/ipcc/techrepI/toc.html.

IPCC (1997) *An Introduction to Simple Climate Models Used in the IPCC Second Assessment Report*, Technical Paper II, IPCC Working Group I, Available from IPCC Secretariat.

IPCC (1997) *Stabilization of Atmospheric Greenhouse Gases: Physical, Biological and Socio-economic Implications*, Technical Paper III, IPCC Working Group I, Available from IPCC Secretariat.

IPCC (1997) *Implications of Proposed CO_2 Emissions Limitations*, Technical Paper IV, IPCC Working Group I, Available from IPCC Secretariat.

IPCC (1995) *Climate Change 1995: the Science of Climate Change*, Contribution of Working Group I to the SAR of the IPCC, Cambridge University Press, Cambridge.

IPCC (1995) *Climate Change 1995: Impacts, Adaptations and Mitigation of Climate Change: Scientific-Technical Analyses*, Contribution of Working Group II to the SAR of the IPCC, Cambridge University Press, Cambridge.

IPCC (1995) *Climate Change 1995, Economic and Social Dimensions of Climate Change*, Contribution of Working Group III to the Second Assessment Report of the IPCC, Cambridge University Press, Cambridge.

IPCC (1994) *Climate Change 1994: Radiative Forcing of Climate Change and an Evaluation of the IPCC IS92 Emission Scenarios*, eds J T Houghton, L G Meira Filho, J P Bruce, H Lee, B T Callander, E F Haites, N Harris, and K Maskell, Cambridge University Press, Cambridge, 1–339.

Climate Change Policy Instruments and their Implications, Proceedings of Tsukuba Workshop, January 1994, eds A Amano *et al.*, Center for Global Environmental Research, National Institute for Environmental Studies, Environmental Agency of Japan, Tsukuba.

Climate Change and Integrated Assessment Models, Bridging the Gap, Proceedings of Asia–Pacific Workshop, UN University, Tokyo, March 1997, eds O K Cameron *et al.*, National Institute for Environmental Studies, Environmental Agency of Japan, Tsukuba.

Equity and Social Considerations Related to Climate Change, Proceedings of Nairobi Workshop, July 1994, eds R S Odingo and A Katama, ICIPE Science Press, Nairobi, Kenya.

Mitigation and Adaptation Cost Assessment: Concepts, Methods and Appropriate Use, 1998, Workshop Proceedings June 1997, eds J Christensen, K Halsnaes, and J Sathaye, UNEP Collaborating Center on Energy and Environment, Riso National Laboratory, Roskilde, Denmark.

The Summaries for Policymakers and the Technical Summary are also available in Arabic, Chinese, French, Spanish, and Russian, available from IPCC Secretariat.

Preparing to Meet the Coastal Challenges of the 21st Century, Conference Report of the World Coast Conference, 1993, available from Netherlands Coastal Zone Management Centre.

Climate Change 1992: the Supplementary Report to the IPCC Scientific Assessment, Report of the IPCC Scientific Assessment Working Group (Working Group I), eds J T Houghton, B T Callander, and S K Varney, Cambridge University Press, Cambridge, 1–200.

IPCC (1992) *Climate Change 1992: the Supplementary Report to the IPCC Impacts Assessment*, Report of the IPCC Impacts Assessment Working Group (WGII), available from Australian Government Publishing Service.

IPCC (1992) *Global Climate Change and the Rising Challenge of the Sea*, Report of the Coastal Zone Management Subgroup of the IPCC Response Strategies Working Group (Working Group III), available from Netherlands Coastal Zone Management Centre.

Proceedings of a Workshop on Assessing Technologies and Management Systems for Agriculture and Forestry in Relation to Global Climate Change, Canberra, 1992, available from IPCC Secretariat.

IPCC (1992) *Climate Change: the IPCC 1990 and 1992 Assessments*, IPCC FAR Overview and Policymaker Summaries, and IPCC Supplement, available from IPCC Secretariat.

IPCC (1990) *Climate Change: the IPCC Scientific Assessment*, Report of the IPCC Scientific Assessment Working Group (Working Group I), eds J T Houghton, G J Jenkins, and J J Ephraums, Cambridge University Press, Cambridge, 1–365.

IPCC (1990) *Climate Change: the IPCC Response Strategies*, Report of the IPCC Response Strategies Working Group (Working Group III), Island Press, Washington, DC, 1–273.

IPCC (1990) *Observed Climate Variations and Change*, Contribution in Support of section 7 of the 1990 IPCC Scientific Assessment, available from IPCC Secretariat.

IPCC (1990) *Strategies for Adaptation to Sea Level Rise*, Report of the Coastal Zone Management Subgroup of the IPCC Response Strategies Working Group (Working Group III), available from Netherlands Coastal Zone Management Centre.

Methodologies

IPCC (1994) *IPCC Technical Guidelines for Assessing Climate Change Impacts and Adaptations*, available from IPCC Secretariat, 1–59.

Carbon Balance of World Forested Ecosystems: Toward a Global Assessment, available from SILMU/the Academy of Finland.

IPCC (1994) *IPCC Guidelines for National Greenhouse Gas Inventories* (3 volumes), available from IPCC Secretariat.

IPCC (1992) *Preliminary Guidelines for Assessing Impacts of Climate Change, 1992*, IPCC Working Group II, available from IPCC Secretariat.

IPCC (1991) *Assessment of the Vulnerability of Coastal Areas to Sea Level Rise - a Common Methodology*, Report of the Coastal Zone Management Subgroup of IPCC Working Group III, available from Netherlands Coastal Zone Management Centre.

REFERENCES

Agrawala, S (1998) Context and Early Origins of the Intergovernmental Panel on Climate Change, *Clim. Change*, 39(4), 605–620.

Agrawala, S (1998) Structural and Process History of the Intergovernmental Panel on Climate Change, *Clim. Change*, 39(4), 621–642.

Bruce, J, Lee, H, and Haites, E, eds (1996) *Climate Change 1995: Economic and Social Dimensions of Climate Change*, Cambridge University Press, Cambridge.

Carter, T R, Parry, M L, Harasawa, H, and Nishioka, S (1994) IPCC Technical Guidelines for Assessing Climate Change Impacts and Adapatations, Department of Geography, University College London, London, and Center for Global Environmental Research, National Institute for Environmental Studies, Tsukuba, Japan.

Greco, S, Moss, R H, Viner, D, and Jenne, R (1994) *Climate Scenarios and Socio-economic Projections for IPCC Working Group II Assessment*, IPCC, Washington, DC.

Houghton, J (1997) *Global Warming: the Complete Brie ng*, 2nd edition, Cambridge University Press, Cambridge.

Houghton, J T, Callander, B A, and Varney, S K, eds (1992) *Climate Change 1992: the Supplementary Report to the IPCC Assessment*, Cambridge University Press, Cambridge.

Houghton, J T, Meira Filho, L G, Bruce, J, Lee, H, Callander, B A, Haites, E, Harris, N, and Maskell, K, eds (1996a) *Climate Change 1994: Radiative Forcing of Climate Change and an Evaluation of the IPCC IS92 Scenarios*, Cambridge University Press, Cambridge.

Houghton, J T, Meira Filho, L G, Harris, N, Kattenberg, A, and Maskell, K, eds (1996b) *Climate Change 1995: the Science of Climate Change*, Cambridge University Press, Cambridge.

IPCC (1990) *Climate Change, the IPCC Scienti c Assessment*, eds J T Houghton, G J Jenkins, and J J Ephraums, Cambridge University Press, Cambridge.

IPCC (1993) *IPCC Procedures for Preparation, Review, Acceptance, Approval, and Publication of its Reports*, approved by the panel at its ninth session (Geneva, 29–30 June 1993).

IPCC (1995) *IPCC Guidelines for National Greenhouse Gas Inventories*, IPCC Secretariat, WMO, Geneva.

IPCC (2000) *IPCC Special Report on Land Use Change and Forestry*, in press.

Izrael, Y (1991) Climate Change Impact Studies: the IPCC Working Group II Report, in *Climate Change: Science, Impacts and Policy: Proceedings of the Second World Climate Conference*, eds J Jager and H L Ferguson, Cambridge University Press, Cambridge.

Kates, R W (1997) Report on Reports: Climate Change 1995 – Impacts, Adaptations, and Mitigation, *Environment*, 39(9), 29–33.

Moss, R H and Schneider, S H (2000) *Uncertainties in the IPCC TAR: Recommendation to Lead Authors for More Consistent Assessment and Reporting*, IPCC, Geneva, Switzerland.

Schneider, S (1991) Three Reports of the Intergovernmental Panel on Climate Change, *Environment*, 33(1): 25–30.

SCOPE (1986) *SCOPE 29, the Greenhouse Effect, Climatic Change and Ecosystems*, eds B Bolin, B Doos, J Jager, and R Warrick, John Wiley & Sons, New York.

Watson, R T, Zinowera, M C, and Moss, R H, eds (1996) *Climate Change 1995: Impacts, Adaptation and Mitigation of Climate Change*, Cambridge University Press, Cambridge.

FURTHER READING

Boehmer-Christiansen, S (1994) Global Climate Protection Policy: the Limits of Scientific Advice, *Glob. Environ. Change*, 4(3), 140–159.

Moss, R H (2000) Ready for IPCC-2001: Innovation and Change in Plans for the IPCC Third Assessment Report, *Clim. Change*, 45, 459–468.

Moss, R H (1996) Avoiding Dangerous Anthropogenic Interference in the Climate System: the Roles of Values, Science, and Policy, *Glob. Environ. Change*, 6(1) 3–6.

Moss, R H (1995) The IPCC: Policy Relevant (Not Driven) Scientific Assessment, *Glob. Environ. Change*, 5(3), 171–174.

Pace Energy Project (1997) Pace University School of Law's Center for Environmental Legal Studies, Global Warming Central *The Great Global Warming Debate*, Released June 16, 1997, http://www.law.pace.edu/env/energy/debateintro.html#Believe, introduction by Ed Smeloff, Director, the Pace Energy Project.

Schneider, S S (2001) What is Dangerous' Climate Change? *Nature*, 411, 17–19.

Schrope, M (2001) Consensus Science, or Consensus Politics? *Nature*, 412, 112–114.

Skodvin, T (1999) *Science- policy Interaction in the Global Greenhouse: Institutional Design and Performance of the IPCC*, Center for International Climate and Environmental Research, Oslo, Norway.

Internalization

The phrase *internalization of costs* refers to an action or instrument for correcting a market failure. The underlying idea is one of the most fundamental in economics: that a perfect free competitive market will result in a Pareto-optimal allocation of resources. *Pareto-optimality* is an ideal state such that each actor in the market, consisting of all buyers and sellers, ends up as well off as possible without making another actor worse off. This ideal result depends on a number of conditions, which need not be discussed in detail here. However, one of those conditions is that the price at which an exchange takes place maximizes gains (and minimizes losses) to both buyers and sellers. This assumes that the exchange between a buyer and a seller does not affect any other party, either beneficially or adversely. Such effects, when they occur, are known as positive or negative spillovers or (more commonly) externalities. The important point is that, in the real world, most kinds of economic transactions do result in *externalities* of one sort or another. For example, production and consumption activities generate wastes and pollution, not to mention noise, congestion, and so forth. Externalities are pervasive, not rare.

It is a fundamental assumption of economics that external effects, as well as goods and services exchanged in the

marketplace, should have monetary value. An important result in economic theory is that a real (i.e., imperfect) market approaches Pareto-optimality if, and only if, the monetary value of externalities is re ected in the market prices of all goods and services that are exchanged. Thus, when a negative externality such as pollution, is generated by the production (or consumption) of a good, the quantity produced is Pareto-optimal only if the resulting damage cost to all affected parties is included in the price of that good or service. This is what is meant by internalization of damage costs. When damage costs are not internalized, the damage-producing good or service is over-produced. Similarly, if a positive externality is neglected (e.g., the contributions of Research and Design (R&D) to general scientific progress) the good or service (e.g., R&D) will be under-produced.

The responsibility for internalizing costs lies with government. Instruments available for the purpose are generally classified as regulation (e.g., by imposing emission standards) and/or taxation. A tax on the carbon content of fuels would be an example of the latter. However, a newer idea for internalizing costs is the creation of exchangeable permits or exchangeable consumption quotas. In this case the role of government is to decide on the allowable pollution burden and to regulate the market for permits, allowing the market itself to set the price.

ROBERT U AYRES *France*

International Commodity-related Environmental Agreements (ICREA)

see ICREA (International Commodity-related Environmental Agreements) (**Volume 4**)

International Council for Science (ICSU)

see ICSU (International Council for Science) (**Volume 4**)

International Decade for Natural Disaster Reduction (IDNDR)

see IDNDR (The International Decade for Natural Disaster Reduction) (**Volume 4**)

International Drinking Water Supply and Sanitation Decade (IDWSSD)

see IDWSSD (International Drinking Water Supply and Sanitation Decade) (**Volume 4**)

International Environmental Prizes

Ian Burton

Meteorological Service of Canada, Downsview, Canada

The creation of a highly visible and prestigious prize is often regarded as an effective way of promoting a particular activity or cause. Perhaps the most famous are the Nobel Prizes offered for outstanding achievement in elds of scienti c research, as well as peace and economics. Such prizes not only reward achievement, but also are thought to challenge others to greater effort, and also perhaps to re ect some credit on the prize donor or sponsor.

In recent times, the rise of the environmental movement, and the recognition of the importance of scienti c understanding of the environment, has given rise to the idea that a prestigious Nobel-type prize for the environment would help to advance the protection of the global environment, and strengthen the reputation of interdisciplinary environmental science and scientists. Three leading environmental scientists have in fact been awarded a Nobel Prize. Paul Crutzen (Max-Planck-Institute for Chemistry in Mainz, Germany), Mario Molina (Massachusetts Institute of Technology in Massachusetts, United States) and Rowland Sherwood (University of California in California, United States) won the prize in 1995 for their work in atmospheric chemistry, particularly concerning the formation and decomposition of stratospheric ozone.

For the environment not one, but many prizes have been established. A few, brie y described in the following text, have attained outstanding international recognition. The list is by no means comprehensive and is necessarily somewhat arbitrary. The selection of prizes is illustrative rather than de nitive.

THE TYLER PRIZE

The World Prize for Environmental Achievement was established by the late John and Alice Tyler in 1973. Prizes are awarded for the protection, improvement or understanding of ecological and environmental conditions anywhere in the world or for the discovery, further development, improvement or understanding of known or new sources of energy.

Laureates receive a US$200 000 prize and are presented with a gold medallion at ceremonies in Los Angeles.

Contact: Dr. Jerome B. Walker, Executive Director, The Tyler Prize, Office of the Provost, University of Southern California, Los Angeles, CA 90089-4019, USA. Telephone: • 1-213-740-9760, fax: • 1-213-740-1313, e-mail: tylerprz@usc.edu.

THE GOLDMAN PRIZE

The prize was established in 1990 by the Goldman Environment Foundation and is given each year to six environmental heroes , one from each of six continental regions: Africa, Asia, Europe, Island Nations, North America, and Central/South America. The prizes are awarded for sustained and important efforts to preserve or enhance the environment, including, but not limited to: protecting endangered species, combating pollution, restoring damaged ecosystems, raising public awareness of environmental issues, or in uencing environmental policies.

Laureates receive a US$100 000 prize.

Contact: Goldman Environment Foundation, 1 Lombard, Suite 303, San Francisco, CA 94111, USA. Telephone: • 1-415-788-9090, fax: • 1-415-788-7890, e-mail: gef@igc.apc.org, URL: www.goldmanprize.org.

THE INTERNATIONAL SAINT FRANCIS PRIZE FOR THE ENVIRONMENT

The prize was established in 1990 by the Franciscan Center of Environmental Studies in Rome, and the General Custody of the Sacred Convent of Saint Francis in Assisi. The prize addresses the global and interdisciplinary character of environmental, spiritual, ethical, cultural and interreligious issues in the search for solutions to ecological problems. The prize is awarded in three categories: education and communications; scientific research; realized projects and effective action. The prize is awarded to persons or institutions that, in a spirit of humility and simplicity, exemplify or promote love of nature. Consideration is also given to conformity to the values and spirit of St. Francis as expressed in his *Canticle of All Creatures.*

In keeping with the vow of poverty taken by St. Francis there is no award of money. Laureates receive a sculpture created by Mario Ceroli and attend the prize award ceremony and symposium in Assisi.

Contact: Franciscan Center of Environmental Studies, via del Serafico, 1, 00142 Rome, Italy. Telephone: • 39-6-519-15-95, fax: • 39-6-504-33-53 or 519-20-67.

THE BLUE PLANET PRIZE

The prize was established in 1992 by the Asahi Glass Foundation of Japan. The prize is an international environmental award given to individuals and organizations who have made outstanding achievements in scientific research and its application and in doing so have helped provide solutions to global environmental problems.

Each year, two award recipients are chosen and each laureate receives 50 million Yen.

Contact: Nobuaki Kunii, Head of Secretariat, Asahi Glass Foundation, Science Plaza, 5-3 Yonbancho, Cliyodaku, Tokyo 102-0081, Japan. Telephone: • 81-3-5275-0620, fax: • 81-3-5275-0871, e-mail: post@af-info.or.jp, URL: www.af-info.or.jp.

THE VOLVO ENVIRONMENT PRIZE

The prize was established in 1988 by the Swedish automobile manufacturer with headquarters in Goteborg. The criteria for receiving the prize are that:

••• the candidate shall have made outstanding innovations or discoveries in the scientific, socio-economic, or technological area, which have had a direct or indirect significance in the environmental field and are of global or regional importance. If a person who meets these standards has also been active in in uencing the decision-making process and in the change of policies, this would be regarded as an additional merit.

Laureates receive SEK 1.5 million.

Contact: The Volvo Environment Prize Foundation, c/o Environmental Affairs, AB Volvo, S-405 08 Goteborg, Sweden. Telephone: • 46-31-59-66-77, fax: • 46-31-53-84-31, e-mail: vhk.1650kh@memo.volvo.se. (*See* Volvo Environment Prize, Volume 4).

THE GLOBAL 500 ROLL OF HONOR FOR ENVIRONMENTAL ACHIEVEMENT

This award was established in 1987 by the United Nations Environment Programme (UNEP). While not strictly a prize, the roll of honor serves a similar purpose. A place on the roll of honor is awarded to those who have demonstrated that environmental issues originate in the daily life, and that individual actions affect the larger community. It recognizes the achievements of individuals and

organizations in the front lines of global environmental action.

Contact: The Secretary, Global 500 Roll of Honour, United Nations Environment Programme, Communications and Public Information, P.O. Box 30552, Nairobi, Kenya. Telephone: • 254-2-62-3401 or 62-3128, fax: • 254-2-62-3927 or 62-3692, e-mail: elisabeth.guilbaud-cox@unep.org or ipainfo@unep.org.

UNEP SASAKAWA ENVIRONMENT PRIZE

The UNEP Sasakawa Environment Prize, sponsored by the Nippon Foundation, was established in 1982. It is awarded every November to individuals who have made outstanding contributions to the management and protection of the environment at the global level.

Laureates receive US$200 000, a crystal trophy and a certificate from the United Nations Secretary-General.

Contact: The Secretary, UNEP Sasakawa Environment Prize, United Nations Environment Program, Communications and Public Information, P.O. Box 30552, Nairobi, Kenya. Telephone: • 254-2-62-3401 or 62-3128, fax: • 254-2-62-3927 or 62-3692, e-mail: elisabeth.guilbaud-cox@unep.org or ipainfo@unep.org.

ZAYED INTERNATIONAL PRIZE FOR THE ENVIRONMENT

The Zayed International Prize for the Environment was founded by the Crown Prince of Dubai to promote sustainable development through various environmental protection initiatives. The Prize of US 1 000 000 dollars is presented every 2 years as follows: first prize, US$500 000; second prize, US$300 000; third prize, US$200 000.

Contact: General Secretariat, Zayed International Prize for the Environment. Fax: • 971-4-332-6777, e-mail: zayedprz@emirates.net.ae.

In addition to this short list of well known prizes there are many more prizes and awards offered on a national and community level far too numerous to identify.

The sponsorship of environmental prizes is not limited to any one type of organization. Private sector companies, government agencies, individual citizens, and religious orders have all become involved. Nor are the prizes limited to science. Both practical results and communication and education are also recognized. No one prize has yet emerged as the preeminent Nobel prize for the environment, and probably never will. The diversity of global environmental prizes re ects the multidimensional nature of global environmental change, and the eclectic and ecumenical character of the global environmental movement.

International Environmental Standards

Scott Vaughan
North American Commission for Environmental Cooperation, Montreal, Canada

A growing list of global, regional, transboundary and shared environmental problems, and state and non-state actors have increasingly turned to a broad range of international environmental standards. Such standards assume several forms, including standards of environmental conduct articulated in such core declarations as the Rio Principles of the United Nations Conference on Environment and Development (UNCED), quantitative emission reduction or threshold standards contained in binding environmental treaties requiring signatories to undertake speci ed domestic measures, the growing emphasis on harmonized environmental and human health assessment procedures, and an increased interest in and adoption of voluntary international environmental management standards. Despite the absence of a precise estimate of the scope and number of such standards, thousands continue to evolve, be reviewed, amended and strengthened. Altogether, they form the backbone of the international environmental agenda.

A number of de nitional issues arise when describing international environmental standards. The rst revolves around the very broad, and at times ill-de ned, parameters of the term environment itself. Experts point to the absence of a universal de nition of environmental policy, such that policies and standards which fall under the environment rubric include measures specifying maximum allowable pollution emission limits for speci c pollutants and chemicals, the speci cation of environmental performance standards for technologies or more general industrial production methods, standards covering the management of renewable and non-renewable natural resources (e.g., forestry or sheries), or standards related to wildlife management and the conservation of biodiversity.

It is worth noting that although standards and regulations often cover identical environmental policies, these two terms - standards and regulations - are not synonymous. Environmental regulations are developed, codi ed and enforced within domestic jurisdictions, at the federal, sub-federal (state/provincial and/or municipal) levels, or through executive or administrative procedures. Regulations developed through such domestic legislative or administrative frameworks tend to be mandatory, and are applied within domestic jurisdictions. By contrast, compliance with international and domestic environmental standards is generally non-mandatory, for several reasons. International organizations mandated to develop standards

are not similarly mandated to enforce the domestic implementation of such measures through supra-national or supra-jurisdictional mechanisms. (However, most environmental treaties contain measures by which other parties can review the extent of domestic compliance with international standards, as well as revert to various non-compliance and more formal dispute settlement provisions.) This absence of a supra-national authority partly stems from the emphasis of international environmental law itself to respect the sovereign right of individual nations to establish their own environmental regulations and standards.

The above distinction between international standards and domestic regulations is not intended to suggest that international environmental laws are not binding upon signatory governments. As illustrated below, many international environmental agreements contain speci c and quanti able targets to which signatory countries are obliged to adhere, primarily through various domestic measures affecting production or consumption of environmental resources.

However, international environmental standards also include a wide range of non-binding measures. As a group, such standards can be divided into four broad categories: (a) standards that are contained in binding international treaties, or non-binding standards that are developed internationally; (b) plurilateral, regional or bilateral standards to address transboundary or shared environmental and conservation problems or priorities; (c) standards setting out best practices in environmental assessment and risk assessment procedures; and (d) voluntary environmental management standards.

STANDARDS CONTAINED IN INTERNATIONAL ENVIRONMENTAL AGREEMENTS, OR DEVELOPED THROUGH NON-BINDING INTERNATIONAL CONSENSUS

Multilateral Environmental Agreements (MEAs)

Approximately 185 MEAs have been negotiated and signed by governments. Obligations and standards contained in international treaties cover a very broad range of both general obligations of environmental protection and the pursuit of sustainable development, as well as specific, quantitative measures that countries must implement through the codification of MEA standards at the national level, through a process of treaty ratification and introduction of national regulations in support of treaty commitments. During this process of national ratification, varying degrees of exibility exist, whereby states can adopt standards that are either higher than those articulated in the MEA itself, or that translate into slightly different national approaches than those suggested in the international standard itself. This exibility in the process of translating international standards into national regulations takes into account important differences, which exist, between countries in domestic regulatory approaches and partnerships in place to pursue environmental policies.

MEAs are often organized, by legal scholars, into environmental media, such as atmospheric protection, the protection of freshwater resources, the protection of oceans and the marine environment, the management of hazardous wastes and chemicals, the protection of wildlife and biodiversity, and the protection of habitats. Each of these broad groupings of MEAs contains different, and often complementary, approaches to environmental management, whereby different types of environmental standards are pursued. For example, under the rubric of the protection of the marine environment and its resources, international standards cover a wide range of activities and obligations contained in different legal instruments: standards related to controlling sources of marine pollution, pipelines and seabed operations are set in the United Nations Law of the Sea Convention (UNCLOS), while standards controlling the operational discharge of pollution from marine vessels, as well as the accidental release of pollutants from vessels, are set out in the International Convention for the Prevention of Pollution by Ships (MARPOL 73/78). UNCLOS also establishes an umbrella for several fisheries management approaches. These include the United Nations (UN) Conference on Straddling Stocks and Highly Migratory Fish Stocks, which establishes limits on commercial fishing in the high seas. To lower damages to fisheries from the use of driftnets in commercial operations, the UN General Assembly adopted a moratorium in 1990 on the use of driftnets. This non-binding resolution by the General Assembly complements standards set out in the Wellington Convention intended to prohibit member countries from using driftnets in their commercial operations.

Although there are important distinctions that exist among different MEAs, re ecting the media-specific, focused and increasingly sophisticated nature of environmental policy, there are numerous examples of common environmental standards that are germane to several international accords. One such example is the principle of international cooperation in support of the on-going exchange of scientific and technical cooperation, coupled with a commitment by signatory governments towards the on-going assessment of specified environmental problems. Another example is the principle of prior informed notification and prior informed consent, contained in several MEAs. These latter principles represent the operational foundation of several MEAs, including, for example, the Convention on International Trade in Endangered Species (CITES), the Basel Convention, and more recently the Montreal Protocol (in which parties are obliged to issue export and import licenses for specified, ozone-destroying substances identified under the Protocol).

Since MEAs represent perhaps the most important source of international environmental standards, examples of specific standards contained in three particularly prominent environmental accords are provided below.

The Montreal Protocol on the Control of Substances that Deplete the Ozone Layer

The Montreal Protocol, originally signed in 1987, establishes concrete provisions that parties to the Protocol are obliged to follow. Two broad types of standards are contained in the Protocol and its amendments: standards covering domestic production and consumption of a clearly defined group of ozone-depleting substances; and standards covering cooperation between parties to the Protocol, as well as standards specifying the treatment of non-parties to the Protocol.

The foundation of the Montreal Protocol rests on the obligations that parties must follow, setting out minimum standards relating to domestic targets and timetables to freeze, reduce and eliminate the domestic production and consumption of ozone-depleting substances. The original (September 1987) Protocol required parties to freeze the domestic consumption and production of chloro uorocarbons (CFCs) at 1986 levels by 1990, to reduce them by 20% by 1993, and to reduce them by 50% by the year 2000, as well as freezing four types of halons by the year 2005. Exceptions to these initial schedules were extended to developing (Article 5) countries for a period of 10 years.

Following scientific evidence that emerged after September 1987, which confirmed the accelerated rate of stratospheric ozone depletion, the Protocol has undergone several important amendments, each of which contains newly revised standards to which signatory governments must adhere. Such amendments began in the London meeting of the Protocol (held in 1990), in which the CFC reduction timetable was moved up to the year 2000. Other standards set out in London included a 50% reduction in halons by the year 2000, the reduction in carbon tetrachloride by 85% by 1995, with a complete phase-out by the year 2000; a freeze of methyl chloroform emissions by 1993, a 30% reduction by 1995, a 70% reduction by 2000, and a complete phase-out by 2005; and a reduction by 20% in all other fully halogenated CFCs by 1993, and a phase-out by 2000. Amendments adopted at the Conference of the Parties following London further strengthened the obligations of parties to protect the ozone layer. For example, in Copenhagen in 1992, the phase-out schedule for most controlled substances was further accelerated, to 1995; substitute chemicals, hydrochloro uorocarbons and hydrobromo uorocarbons, were included in new schedules, and methyl bromide was also included in the list of controlled measures (with exceptions for agricultural quarantine and other specified uses).

As the Montreal Protocol marked its first decade, developing (Article 5) countries have begun the process of implementing domestic controls on the production, consumption and international trade in substances controlled under the Protocol. By 1 July 1999, all Article 5 countries were obliged to implement a freeze on the consumption of five (Annex A) CFCs. A similar freeze on halons and methyl bromide will commence on the 1st January 2002, and other freezes and reduction schedules will begin slightly later. For example, by 2010, Article 5 countries will implement a 100% elimination of all CFCs and halons.

The second grouping of environmental standards set out in the Montreal Protocol comprises methods of cooperation among parties, both in financial and technical assistance extended to developing countries, as well as procedures governing international trade in controlled substances between parties. Such procedures include prior informed notification provisions. Standards also include the prohibition of all international trade in controlled substances between parties and non-parties to the Protocol (with certain exceptions). This latter area, that of international environmental standards specifying trade prohibitions, has been a source of discussion and concern in the World Trade Organization's Committee on Trade and Environment, and highlights the problem of inconsistent obligations and standards that can arise between different areas of public policy and treaties (*see* WTO (World Trade Organization), Volume 4).

CITES

The CITES (*see* CITES (Convention on International Trade in Endangered Species of Wild Fauna and Flora), Volume 4) agreement was completed in 1973, and entered into force in July 1975. Among the key objectives of CITES is to establish an international mechanism to address high exploitation levels and volumes of international trade in wild animals and plants, which present risks to their survival. Two key standards represent the foundation of CITES. The first is the sophisticated system of notification procedures governing all trade in a list of species, which are protected under the Convention. Under the notification system, CITES trade permits (covering both import and export permits) must be exchanged between parties to the Convention if trade in a certain species is permitted. Where all trade in specified species is prohibited, parties are obliged to ensure that domestic authorities enforce all trade in specimens of endangered species, through customs authorities and other measures. The second is the development of science-based procedures relating to species assessments: species classifications are based on consensus decisions taken by parties to the Convention concerning whether to include, remove or alter the status of species listed in one of three CITES appendices. Altogether, approximately 30 000 species are listed in CITES appendices. Taken together, the notification system and CITES

appendices represent important examples of international environmental standards.

Kyoto Protocol

In December 1997, the Kyoto Protocol was adopted under the auspices of the UN Framework Convention on Climate Change. The Kyoto Protocol (*see* United Nations Framework Convention on Climate Change and Kyoto Protocol, Volume 4) established quantified greenhouse gas (GHG) emission limitations for industrialized countries. In addition, the Protocol sets out the initial rules and allowable mechanisms with which governments can pursue domestic and international GHG emission reductions. Under the Kyoto Protocol, industrialized countries that ratify the accord are required to lower GHG emissions with a view to achieving a 6% reduction in total domestic emissions based on 1990 levels, based on average emissions in the years 2008 through to 2012. The Protocol covers six gases – carbon dioxide, methane, nitrous oxide, hydro uorocarbons, per uorocarbons, and sulfur hexa uoride. Those parties that have ratified the targets set out in the Kyoto Protocol can meet domestic GHG reductions either through domestic measures alone, or by including options based on international cooperation identified under the Protocol, and subsequently referred to as the *Kyoto Mechanisms*. Among the mechanisms identified under the Protocol is the clean development mechanism (CDM) (*see* Clean Development Mechanism, Volume 4), which allows industrialized countries to sponsor and share credit for emission offset projects undertaken in developing countries: under rules set out in the Protocol, all transactions involving industrialized and developing countries would be brokered through the CDM, with certified emission reductions achieved through the CDM undertaken between the year 2000 and the 2008–2012 period being used as credits to meet the domestic emission goals of industrialized countries. In addition, the Kyoto Protocol will consider international emission trading schemes as being supplemental to domestic actions for reducing GHG emissions (*see* Emissions Trading, Volume 4). Under the treaty, emission reductions that are achieved with international emissions trading would view the latter approach as being supplemental to any domestic actions taken.

International Environmental Standards Developed through Non-binding Negotiations

In addition to the international standards developed in binding international legal agreements, numerous environmental standards are developed, reviewed and revised through a process of non-binding international procedures. Such standards include, as noted, non-binding international environmental principles. Among the most prominent of these are such provisions as the Common Heritage of Humankind, the Obligation Not to Cause Environmental Harm, the Principle of Intergenerational Responsibility, the Precautionary Principle and the Polluter and User Pays Principle. Each of these principles and concepts articulates different types of minimum standards of acceptable actions that state and non-state actors are expected to pursue in support of environmental protection. These principles are contained both in their own right, such as in the UNCED Rio Principles, as well as in the UNCED Forest Principles commitment adopted by governments at the UNCED. It is important to note that while such environmental principles are non-binding (or are commonly referred to as soft law), they are increasingly important in guiding both national and international environmental laws and regulations. To illustrate, differing interpretations of the operational implications of the precautionary principle, in relation to burden of proof, in denying market access within countries for foods treated with different additives and drugs, was the source of at least one dispute in the World Trade Organization (between the US and the European Union (EU) around the use of growth hormones in beef).

This case illustrates an important area of international standards around the protection of human health and the environment. The International Program on Chemical Safety (IPCS) (*see* IPCS (International Program on Chemical Safety), Volume 4) jointly administered by UNEP, the World Health Organization (WHO) and the International Labor Organization, undertakes comprehensive assessments of the effects of individual chemicals on human health and environmental quality. Such assessments have been undertaken for over 150 chemicals, and conclusions of assessments, which comprise a summary of the type and severity of risks to humans from different exposures of the specific chemical, represent one type of non-binding, information-based international environmental standard. In many cases, results of IPCS assessments are adopted by national environmental or health authorities, and therefore represent one type of international environmental standard.

Other examples of international standards that combine scientific evidence with international consensus include international food standards developed under the Codex Alimentarius Commission (jointly administered by the Food and Agriculture Organization and the WHO). CODEX sets out procedures both to determine international food safety priorities, as well as procedures to be followed by countries and that international body relating to the adoption of draft standards. International food standards combine scientific evidence, gathered through different scientific committees, with other non-scientific policy goals (including, for example, commercial interests related to ensuring fair trade in international food markets). In mid-1999, the Codex Alimentarius Commission established three new task forces to review how international standards will be set for

foods produced with biotechnology, trade in fruit juices, and the production and use of animal feeds.

A third example of international standards that marry scientific consensus with other policy goals are predictions issued by the Intergovernmental Panel on Climate Change regarding probable increases in global mean temperatures.

REGIONAL AND BILATERAL ENVIRONMENTAL STANDARDS

In addition to standards set out in MEAs and through non-binding international negotiations, an estimated 700 bilateral or regional environmental agreements exist. Such agreements are similar to global or multilateral agreements in terms of scope, but are designed to address shared or transboundary environmental problems.

Examples of regional environmental standards include the over 200 binding Environmental Directives which have been adopted by the EU, and with which the 15 EU member countries are obliged to comply. EU Environmental Directives cover such activities as allowable emission levels for different specified pollutants, proposed standards related to environmental performance requirements in the road transportation sector, directives covering the treatment and management of individual hazardous wastes and toxic chemicals, and a wide range of other standards related to community-wide approaches to environmental protection. For example, in June 1998, the EU established a new water framework directive. Among the provisions in the framework include a listing of 30 hazardous chemicals slated for priority action, including eventual phase-out. Other examples of international standards included in EU law include voluntary environmental management systems (EMAS), which are viewed as complementing the ISO 14 000 series of environmental management standards (*see* ISO (International Organization for Standardization), Volume 4), and criteria setting out provisions related to environmental labeling schemes in the EU. The following provides more detailed examples of bilateral or regional environmental standards:

Canada–US Bilateral Air Quality Agreement

Adopted in 1991, the agreement sets out obligations to control sulfurdioxide (SO_2) and nitrogen oxide (NO_x) emissions, the principal sources of acid rain. Among the standards contained in the agreement are specific targets outlining emission reduction levels, as well as the establishment of procedures related to joint cooperation between the two countries. Among the quantitative emission reduction targets set out in the agreement covering the US were a required reduction in annual SO_2 emissions by 10 million tons *per annum* below 1980 levels, by the year 2000 (with limited exceptions); the achievement of a national cap of just under nine million tons of SO_2 emissions per year from electrical utilities by the year 2010; the introduction of additional national regulations if SO_2 emissions from industrial sources exceed 5.6 million tons *per annum*, and the specification of certain technology-based standards intended to lower stationary and mobile sources of NO_x. For Canada, for example, obligations under the agreement included a reduction in SO_2 emissions in the seven easternmost provinces to 2.3 million tons *per annum* by 1994; establishment of a cap at that level of SO_2 emissions until 1999; to achieve a permanent cap of 3.2 million tons by 2000, and other emission reductions. In addition, the Acid Rain Agreement sets out provisions for notification and consultation requirements covering a range of activities, including assessment, notification and mitigation obligations (Article V) covering proposed changes in domestic laws, regulations and policies that may affect SO_2 and NO_x emissions.

Northwest Atlantic Fisheries Organization (NAFO)

NAFO is a regional organization comprising 15 member countries of the Northwest Atlantic area, with the mandate of promoting the conservation and optimum utilization of the fishery resources of the Northwest Atlantic area within a framework appropriate to the regime of extended coastal State jurisdiction over fisheries . The centerpiece of NAFO's work is the annual setting of total allowable catch (TAC) limits for commercially significant species, and the allocation of quotas to NAFO members. The TACs and quotas, while adopted by consensus, are not binding: member countries have the exibility within NAFO to object to the quota allocated within a pre-determined time period, and subsequently declare its own quota.

The Bamako Convention on the Ban of Import into Africa and the Control of Transboundary Movement and Management of Hazardous Wastes within Africa

Fifty-one countries of the Organization of African Unity signed the Bamako Convention in early 1991. The Convention prohibits the import of all hazardous wastes from outside Africa, with no exceptions. The Convention notes that any imports originating outside the continent shall be deemed illegal and a criminal act, although specific legal recourse provisions are not fully articulated. The Convention also prohibits the dumping of hazardous wastes at sea or in the territorial waters of Parties. The Convention also requires Parties to establish national waste monitoring and enforcement mechanisms. Although the Bamako Convention prohibits waste imports from outside of Africa, it allows trade in wastes between African states that are party

to the Convention. However, the Convention requires all parties to pass domestic legislation requiring waste generators to notify all waste shipments prior to their movements. Moreover, the Convention stipulates that all generators will be subject to strict joint and several liability for the release of hazardous wastes. The Bamako Convention arose as a direct result of concern amongst African countries during the final phase of negotiations, which led to the adoption of the Basel Convention on the Transboundary Movement of Hazardous Wastes, in March 1989 (*see* Basel Convention, Volume 4). Specifically, many African countries voiced concern that the original Basel Convention did not ban waste shipments from developed to developing countries.

UN Economic Commission for Europe Protocols on Persistent Organic Pollutants and Heavy Metals

In 1998, environment ministers from Europe (including western, central and eastern European countries), the US, Canada and Mexico signed two protocols to eliminate or severely restrict all emissions of heavy metals and *persistent organic pollutants* (POPs). Standards covering heavy metals and POPs were included as Protocols under the umbrella UN Economic Commission for Europe (ECE) Convention on Long-Range Transboundary Air Pollution. The POPs protocol identifies 16 substances, which are subject to stricter regulations, including the listing of eight substances – aldrin, chlordane, chlordecone, diedrin, endrin, hexabromobiphenyl, mirex and toxaphene – that would be subject to an outright ban upon entry into force of the protocol. Provisions contained in the protocol on heavy metals include emission reduction levels covering cadmium, mercury and lead, setting new and binding emission limits for new as well as existing industrial and energy production sources, the requirement for the use of best available technologies for reducing emissions, and a commitment to phase out leaded gasoline. This regional accord complements work at the global level in the development of an international Convention on POPs, which is being negotiated under the auspices of UNEP.

STANDARDS THAT SET OUT BEST PRACTICES OR GUIDELINES IN ENVIRONMENTAL ASSESSMENT AND ENVIRONMENTAL RISK ASSESSMENT

At the UNCED, the world community committed itself towards the development of harmonized environmental and risk assessment procedures (Agenda 21, 19, 20a). In recent years, considerable attention has focused on developing common environmental assessment procedures, to help ensure that comparable environmental data, baselines and methodologies are used by individual countries. In order to help frame common environmental impact assessment (EIA) procedures, in 1991 the UN ECE facilitated negotiations towards the Convention on EIA in a Transboundary Context. The Convention identifies key sectors that should be subject to environmental assessments, and commits signatories to establish comparable technical and other procedures, including public participation and EIA documentation. Efforts by several international organizations, including the OECD Development Assistance Committee, UNEP, the World Bank and others continue to focus on developing common international procedures and standards for project and policy reviews.

International effort also focuses on developing common approaches to specific environmental issues. For example, the World Meteorological Organization (WMO) has been coordinating activities to harmonize assessment methodologies related to both weather forecasting and climate change. A standard assessment tool that continues to be developed by WMO and the Intergovernmental Panel on Climate Change is the General Circulation Model, which is applied to simulate the rate and impacts of climate change. Under the IPCS, as another example, common scientific concepts and principles related to assessing risks to human health associated with chemical exposure have been developed, and workshops have been held in support of the harmonization of procedures to assess risks from exposure to chemicals. In mid-1999, Canada and the US reached an agreement covering procedures for the sharing of information related to the environmental and health review of new industrial chemicals, including administrative procedures concerning the exchange of data covering new chemicals and polymers.

VOLUNTARY INTERNATIONAL ENVIRONMENTAL STANDARDS

In recent years, interest in voluntary international environmental standards has increased among both state and non-state sectors, including, in particular, the private sector. Examples of international voluntary standards include codes of conduct negotiated and adopted by international industry associations as well as private corporations and companies. It has been estimated that more than 50 such international environmental codes of conduct have been adopted by the private sector, covering a wide range of sectors. Such international codes complement initiatives undertaken at the national level. Examples include:

- the Business Charter on Sustainable Development (1990), adopted through the International Chamber of Commerce;
- the Responsible Care Program and its supporting Guiding Principles and Guidelines and management practice codes;

- the FAO Code of Conduct on the Distribution and Use of Pesticides which was adopted by the International Group of National Associations of Agrochemical Manufacturers International; and
- the Environmental Charter of the International Council on Metals and the Environment.

Another example of an international industry code of environmental conduct is UNEP's Statement by Commercial Banks on the Environment. The statement, which was negotiated by UNEP and an advisory group of banks, commits some 150 bank signatories to best practices in environmental management, undertaking business partnerships with suppliers that follow similarly stringent environmental standards, and obliges signatories to integrate environmental considerations into core credit risk decisions. (A similar UNEP initiative, launched in 1995, involves the insurance and pension fund sectors.)

Another important component of international standards related to the environment is the development of environmental management tools. Such tools are intended to help companies identify, prioritize and integrate environmental considerations into operational systems. The two most prominent tools are the EU's EMAS, and standards developed under the ISO 14 000 series of environmental management standards.

International Federation of Institutes for Advanced Study (IFIAS)

see IFIAS (International Federation of Institutes for Advanced Study) (**Volume 4**)

International Human Dimensions Program on Global Change (IHDP)

see IHDP (International Human Dimensions Programme on Global Change) (**Volume 4**)

International Institute for Applied Systems Analysis (IIASA)

see IIASA (International Institute for Applied Systems Analysis) (**Volume 4**)

International Maritime Organization (IMO)

see IMO (International Maritime Organization) (**Volume 4**)

International Organization for Standardization (ISO)

see ISO (International Organization for Standardization) (**Volume 4**)

International Program on Chemical Safety (IPCS)

see IPCS (International Program on Chemical Safety) (**Volume 4**)

International Register of Potentially Toxic Chemicals (IRPTC)

see IRPTC (International Register of Potentially Toxic Chemicals) (**Volume 4**)

International Saint Francis Prize

see International Environmental Prizes (**Volume 4**)

IOC (Intergovernmental Oceanographic Commission)

The IOC was created by the United Nations Educational, Scientific and Cultural Organization (UNESCO) General Conference in 1960, as a body with functional autonomy within UNESCO, following the recommendation of the First International Oceanographic Conference, convened by UNESCO in Copenhagen, Denmark in 1960. The

mandate of IOC is to develop, recommend and coordinate implementation of international cooperative marine research programs; to develop and provide ocean services to member states; and to develop and help implement training, education and mutual assistance programs for capacity building in marine science, technology, ocean observations and services.

The IOC has a governing body of its own member states in the form of: an assembly of all its member states (presently 126) meeting biannually; an Executive Council, meeting annually; and a secretariat, housed in the UNESCO headquarters in Paris, and headed by an executive secretary, with the rank of assistant director general of UNESCO. The membership is open to any one member state of anyone of the organizations of the United Nations (UN) system. A state becomes a member state of IOC through communicating its wish to the Director General of UNESCO.

The IOC cooperates closely with all bodies and programs of the UN system dealing with various aspects of ocean issues and with relevant non-governmental organizations, e.g., International Council for Science (ICSU), Scientific Committee on Oceanic Research (SCOR), the Scientific Committee on Problems of the Environment (SCOPE), IUCN and the International Ocean Institute (IOI).

The IOC cooperates with and provides advice to the parties of the UN Framework Convention on Climate Change and the Convention on Biological Diversity.

The IOC is recognized in the UN Convention on the Law of the Sea as the competent international organization in the context of scientific research and ocean observations.

The IOC assembly formulates and guides the implementation of the IOC programs. These currently focus on four major themes: international ocean research; establishment of the Global Ocean Observing System; capacity building; efficient and widespread sharing of ocean data and other ocean services. The major program areas are developed and guided by subject area intergovernmental committees. Some of these are joint with other organizations, e.g., the World Meteorological Organization (WMO), the UN Environment Programme, and the International Maritime Organization (IMO).

The implementation of the IOC programs relies on active participation of the member states and financial contributions for international programs from UNESCO, member states directly and donor institutions. In order to stimulate capacity building and promote participation, the IOC has created regional subsidiary bodies in developing regions. The IOC has currently seven such regional, intergovernmental subsidiary bodies governed by the members states from the region in question. The IOC has several series of publications, e.g., reports of its bodies; workshop and training course reports; books; ocean maps; manuals and guidelines.

The IOC co-sponsors all major global ocean programs and the joint Group of Experts on the Scientific Aspects of Marine Environmental Protection (GESAMP). It provides the secretarial services for the Intersecretariat Committee on Scientific Programs Relating to Oceanography, and the Administrative Committee on Coordination (ACC) Subcommittee on Oceans and Coastal Areas. The IOC assembly was the body that proposed the celebration of the 1998 International Year of the Ocean.

The address of the IOC secretariat is: Intergovernmental Oceanographic Commission, UNESCO, 1, Rue Miollis, 75732 Paris Cedex 15, France.

GUNNAR KULLENBERG *Malta*

IPCC (Intergovernmental Panel on Climate Change)

see Intergovernmental Panel on Climate Change (IPCC): an Historical Review (**Volume 4**)

IPCS (International Program on Chemical Safety)

The IPCS was launched in 1980 as a joint undertaking of the World Health Organization (WHO), the United Nations Environment Programme (UNEP) and the International Labor Organization (ILO).

Its role is, on the one hand, to establish the scientific health and environmental risk assessment basis for the safe use of chemicals and, on the other, to strengthen national capabilities for chemical safety.

In this context, it is carrying the following main tasks:

- evaluation of the risk to human health and the environment from exposure to chemicals and establishment of guideline values for exposure: about 200 priority chemicals have thus far been assessed (environmental health criteria documents), and health and safety guides have been published for over 200 chemicals;
- to promote the development and use of methods for laboratory testing and epidemiological studies for the health and environmental assessment of chemicals: about 20 reports in this area, plus ten books by the Scientific Group on Methodologies for the Safety Evaluation of Chemicals (SGOMSEC), co-sponsored by

the IPCS and SCOPE (the Scientific Committee on Problems of the Environment) of ICSU (the International Council for Science);
- to promote cooperation, in particular with developing countries, for capacity building in the area of chemical safety: organization of courses, preparation of training modules, etc.;
- to promote international cooperation for emergencies and accidents involving chemicals;
- to support national programs for prevention and treatment of poisonings involving chemicals: handbooks, poison information database *INTOX*, etc.

IPCS is managed by the WHO in Geneva and works through a worldwide network of Participating Institutions and National Focal Points. It played a major role in the elaboration of Chapter 19 (Chemicals) of Agenda 21, defining an international strategy for action on chemical safety into the 21st century. This strategy calls for accelerating the international assessment of chemical risk, harmonizing the classification and labeling of chemicals, establishing risk reduction programs, preventing illegal international traffic in toxic and dangerous products and building capacities.

The implementation of Agenda 21, Chapter 19 led to the establishment in 1994 of the International Forum on Chemical Safety (IFCS), a non-institutional arrangement where government representatives and NGOs (non-governmental organizations) meet to integrate and consolidate efforts to promote chemical safety, as well as to the setting up of, the Inter-organization program for the Sound Management of Chemicals (IOMC), grouping relevant bodies in the UN family (including UNEP and its chemicals program, *see* IRPTC (International Register of Potentially Toxic Chemicals), Volume 4), plus the OECD (the Organization for Economic Co-operation and Development).

IPCS has made an invaluable contribution to the improvement of chemical risk assessment and the dissemination of good practices in this important area. One can only regret that WHO decided recently to downgrade the program.

PHILIPPE BOURDEAU *Belgium*

IPO

see Integrated Pest Management (IPM) in an Era of Global Environmental Change (**Volume 4**)

IRPTC (International Register of Potentially Toxic Chemicals)

The IRPTC was set up shortly after the United Nations Environment Program's (UNEP) creation, upon a suggestion of the Scientific Committee on Problems of the Environment (SCOPE), which had prepared proposals for the 1972 Stockholm United Nations (UN) conference on the human environment. It was realized that the large and ever-increasing number of manufactured chemical products, some of which were highly persistent, was generating global contamination (e.g., DDT residues in Antarctic fauna), with little knowledge of its health and environmental impacts.

The IRPTC was supposed to build a database on these chemicals and their production, uses, degradation patterns, environmental pathways, toxicity and ecotoxicity, regulatory controls if any, etc. This it undertook and achieved to a certain extent. Its databank with data profiles on chemical substances (over 8000) is available for consultation. It also helps developing countries to establish their own national registers and offers training on the control of hazards posed by chemicals. It operates a query response service and publishes a bulletin on information about hazardous chemicals.

IRPTC has now expanded to become UNEP Chemicals, the mission of which is (1) to catalyze actions to promote chemical safety by helping countries in building their capacity to produce, use, and dispose of chemicals safely, and (2) to facilitate global actions that may be needed to reduce or eliminate chemical risks. In all this it cooperates with the International Program on Chemical Safety (IPCS), a joint program between the World Health Organization (WHO), the UNEP and the International Labor Organization (ILO).

National capacity building is an important part of UNEP Chemicals' work, either through the promotion of access to the relevant information or by direct work with the countries in organizing training and workshops (about 20 per year) and providing hotline support to governments.

UNEP Chemicals has also been quite active recently in its second role by working toward two important global agreements. The first is a legally binding instrument for the application of the prior informed consent (PIC) procedure for the import of hazardous chemicals, which was initially applied on a voluntary basis but was eventually signed in 1998 by over 50 countries. The second deals with the persistent organic chemicals (POPs) 12 of which are targeted for priority action, which would include, among other things, reduction programs. The matter is handled under the auspices of the International Forum on Chemical Safety (IFCS) and it is hoped that an agreement will be adopted in the year 2000.

Further information is available from http://www.unep.org/unep/program/hhwb/chemical/irptc.

PHILIPPE BOURDEAU *Belgium*

ISO (International Organization for Standardization)

The ISO is comprised of national standardization bodies from approximately 130 countries. The mandate of the ISO, which is based in Geneva, Switzerland, is to promote the development of international standards and supporting activities, including the exchange of technical information, technical capacity building and training, the review and revision of standards, and the formulation and adoption of new ones. An underlying goal of ISO standards is to facilitate international trade in goods and services, as well as to support international cooperation in such areas as scientific, technological and economic activity.

Traditionally, ISO standards have been developed to cover merchandise products. Hundreds of technical committees meet regularly to review and update a wide range of product-related standards. A highly successful and visible departure from traditional ISO activities was the launching of the ISO 9000 series of quality management standards. Following the ISO 9000 series, in the early to mid-1990s, the ISO commenced work on the ISO 14000 series of international, voluntary environmental management standards. Established under the Technical Committee 207 of the ISO, the 14000 series are intended to address six broad areas of environmental management: environmental management systems; environmental auditing and related investigations; environmental labels and declarations; environmental performance evaluation; life cycle assessment; and terms and definitions. By mid-1999, more than 5500 organizations had obtained certification worldwide under the ISO 14000 series.

The goal of the ISO 14001 series, completed in late 1996, is to promote better internal management approaches in companies and organizations, by identifying priority areas, establishing internal systems and providing external benchmarks by which to measure progress in improving environmental management. Given differences among companies that can apply for the ISO 14000 series, it is by definition generic and exible, so as to meet differing operational needs of large and small scale enterprises in all sectors, in a way that is cost effective and subject to external verification by third-party certification agencies. Among the concerns of the series is that this exibility implies the absence of clearly defined and quantitative environmental management targets. More information on ISO can be obtained by contacting the ISO, 1, Rue Varembe, Case Postale, CH-1211 Geneve 20, Switzerland.

SCOTT VAUGHAN *Canada*

IUCN (The World Conservation Union)

Founded in 1948, with the original name of the International Union for Conservation of Nature and Natural Resources (IUCN) and now called The World Conservation Union, the IUCN brings states, government agencies and a diverse range of non-governmental organizations together in a unique world partnership. Its 955 members include 76 state members, 111 government agencies, 732 non-governmental organizations, and 36 affiliates. Its six technical, voluntary commissions, with nearly 10 000 members from over 180 countries, include: education and communication; environmental law; environmental, economic and social policy; protected areas; species survival; and ecosystem management. IUCN's mission is to in uence, encourage and assist societies throughout the world to conserve the integrity and diversity of nature and to ensure that any use of natural resources is equitable and ecologically sustainable . IUCN provides information and technical knowledge, based on the latest scientific research. It promotes a common approach to the world's environmental problems, ensuring that lessons learned in one area are available in others. And as a global advocate for the environment, it represents the views of its members on the world stage. IUCN's Environmental Law Centre has drafted or assisted in the drafting of the major international conservation conventions, including CITES, wetlands, world heritage, migratory species, and biodiversity. Its Species Survival Commission provides the authoritative view on threatened species through its Red Data List, its World Commission on Protected Areas produces the definitive UN List of Protected Areas, and its Biodiversity Policy Programme is a leading advisor on implementation of the Convention on Biological Diversity. IUCN has about 900 employees and an annual budget of about US$60 million. With headquarters in Switzerland, IUCN's operations are decentralized and are carried forward by a network of 42 regional and country offices, located principally in developing countries. IUCN's website is http://www.iucn.org.

JEFFREY A MCNEELY *Switzerland*

Izrael, Yuri A

(1930-)

Yuri A Izrael, Russian atmospheric scientist, graduated in 1953 from the Central Asian State University (Tashkent), and received his PhD in 1963, when he became director of the Institute of Applied Geophysics, State Committee for Atomic Energy. In 1970 he became deputy head of the USSR Hydrometeorological Service, and head in 1974. Despite the enormous management and administrative burden associated with these positions, Izrael continued his active research career over the next two decades, working (often on Saturdays) with a small group of scientists interested in the problems of environmental monitoring and assessment. During this period, Professor Izrael sponsored a number of international workshops on the subject, including publication of the proceedings. Izrael developed the concept of integrated monitoring systems, the idea being that the biosphere ought to be treated as a whole, involving coordination amongst meteorologists, hydrologists, oceanographers, limnologists and ecologists. These ideas had a considerable impact on the design of the global environmental monitoring systems that were being developed by the World Meteorological Organization (WMO) and the United Nations Environmental Programme (UNEP) in the 1970s and 1980s.

During these two decades, Dr Izrael was the Soviet representative to the WMO, where he became Second Vice-President, and then First Vice-President. He was also an active member of the Intergovernmental Panel on Climate Change (IPCC) in its early years.

In 1991 Yuri Izrael retired from government service and created his own research institute, the Institute of Global Climate and Ecology. In 1994, he became a full member of the Russian Academy of Sciences, and in 1996, he was chosen as Secretary of the Department of Oceanology, Atmospheric Physics and Geography of the Russian Academy of Sciences.

Academically Izrael has published frequently, in both English and Russian, including a number of books that he has written or edited. His latest work (1998) is an atlas of cesium deposition on Europe after the Chernobyl accident, published by the European Union in Luxembourg.

R E MUNN *Canada*

JK

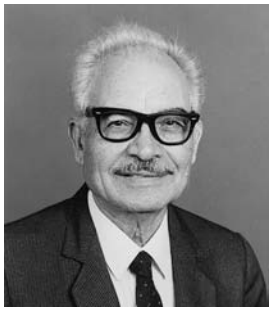
Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection (GESAMP)

see GESAMP (Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection) (**Volume 4**)

Joint Implementation

Joint implementation is the terminology used in the Framework Convention on Climate Change (FCCC) – whereby greenhouse gas reduction projects are implemented in one FCCC party and funded from another; the sponsoring party gains a credit against its own obligations under the convention to reduce its emissions of greenhouse gases (*see* United Nations Framework Convention on Climate Change and Kyoto Protocol, Volume 4).

R E MUNN *Canada*



Kassas, Mohamed

(1921-)

Mohamed Kassas, Egyptian, Emeritus Professor of Botany, University of Cairo, holds a PhD from the University of Cambridge (UK) (1950). He is married and has two children.

At home, in Egypt, he is a member of the Egyptian Academy of Science and the Institute d 'Egypte. Internationally, he is past President (1978–1984) of the World Conservation Union (IUCN), fellow of the Indian National Academy of Science, the World Academy of Art and Science, and the Club of Rome.

In the early 1960s, he wrote reports to the Egyptian authorities on the likely environmental consequences of the Aswan High Dam. These were among the earliest environmental impact assessment studies. Since the preparations for the Stockholm Conference on the Human Environment 1970, he has become the most in uential figure in Egyptian environmental circles.

His principal research interests during the last 55 years have been in plant ecology, especially in arid lands. He and his students (a score of them got their MSc and PhD under his supervision) surveyed plant life in the deserts and coastal lands of Egypt, Sudan, Saudi Arabia, and Iraq, and contributed to basic understanding of the ecology of desert vegetation. At the international level, he was among the pioneers who warned the world against hazards of desertification and contributed to the United Nations Educational, Scientific, and Cultural Organization (UNESCO) Arid Lands Research Program, and to the scientific preparation for the United Nations Conference on Desertification (1977). He was on the panel of senior advisors to the United Nations Environment Programme (UNEP) between 1973 and 1992. He is on the editorial boards of several international journals in the fields of ecology and environmental sciences. He has published profusely in major scientific journals, especially in Europe and US. He was part author and co-editor of the first of two major UNEP environment reports: The World Environment 1972–1982.

In the process of elaborating the World Conservation Strategy (launched 1980), he, as president of IUCN, assisted in drafting and he actively participated in consultations among partners: IUCN, UNEP and World Wildlife Fund and their associates. The final text bears his moderating in uence.

He was a very active member of the (*inter alia*) Egyptian delegation to the 1972 Stockholm Conference. He led

or actively participated in the development of The National Environmental Action Plan and The National Conservation Strategy and led the Egyptian delegation through most of the negotiations of the UN Convention on Desertification. During this period he became the *de facto* science advisor to the Intergovernmental Negotiating Committee for the Elaboration of an International Convention to Combat Desertification (1992–1994).

Professor Kassas has many of national and international honors. He has published more than 100 scientific papers in highly respectable journals, a representative sample of which are:

- *Habitat and Plant Communities in the Egyptian Deserts*, Part I–VII, Journal of Ecology, 1953–1970;
- *Deserti cation Versus Potential for Recovery in Circum-Saharan Territories; Arid Lands in Transition*, AAAS, US, 1970;
- *Ecology and Management of Deserti cation. Earth 88: Changing Geographic Perspective*, National Geographic, 1988;
- *Deserti cation: a General Review*, Journal of Arid Environments, 1995;
- *Fragile Ecosystems in Near East Countries. Ecology Today: an Anthology of Contemporary Ecological Research*, 1998.

Professor Kassas was the recent (2000) winner of the important Zayed International Prize for the Environment.

MOSTAFA K TOLBA *Egypt*

Kyoto Protocol

see United Nations Framework Convention on Climate Change and Kyoto Protocol (**Volume 4**)

L

Lake Victoria

see Lake Victoria (*Case study*, **Volume 4**)

Lang, Winfred

(1941 - 1999)

The late Ambassador Lang played a very significant role in the negotiation and formulation of the Vienna Convention on the protection of the ozone layer and its Montreal Protocol on substances that deplete the ozone layer.

He studied law during the period 1959–1966 at Vienna University (Law School), Oslo University, The Hague Academy of International Law and the Vienna Diplomatic Academy. From 1966, until he died in 1999, he was member of the diplomatic service of Austria. He was posted in Belgium, Paris, Tokyo and Geneva. He was deputy permanent representative and then permanent representative to the United Nations (UN) Office of International Organizations in Geneva; observer to the Committee on Disarmament and head of Austrian delegation to many disarmament related conferences. In 1982, he became Professor of International Law and International Relations at the University of Vienna. He published profusely, books and articles on international law, then on international environmental law.

In 1985, he was elected president of the UN conference that adopted and opened for signature the Vienna Convention. Immediately after the adoption of the Vienna Convention he played an important role towards the formulation of the Montreal Protocol by chairing the Ad Hoc Working Group on the Development of the Montreal Protocol. In 1986, he was the president of the UN Conference on Biological and Bacteriological Weapons (Geneva).

In 1987 he was elected president of the UN conference that finalized and adopted the Montreal Protocol and opened it for signature. In 1993 he was chairman of the council of the European Free Trade Association (EFTA). He served in various capacities within and outside the UN system. He was also since 1993 member of the Board of Trustees of the United Nations Institute for Training and Research (UNITAR).

Ambassador Lang was married with three children. He passed away on 19 May 1999.

MOSTAFA K TOLBA *Egypt*

LEAD (Leadership for Environment and Development)

The Leadership for Environment and Development (LEAD), established in 1991 by the Rockefeller Foundation, is today an independent, non-governmental and non-profit international organization with the basic mission of capacity building for selected future leaders, whose actions will contribute to the decisions made about environment and development issues in most parts of the world.

LEAD associates are generally drawn from five sectors of society: academia, business, government, media, and non-governmental organizations. They are between the ages of 30 and 40 and are fully employed in their organizations and are expected to play key roles in their countries. Their employers contribute to the LEAD program by allowing the associates to be absent from their jobs to participate in the LEAD training which normally takes approximately 80 days (including week-ends) during the 2-year training program.

Once they complete the training phase of the program, the associates become LEAD fellows for whom there exists a special program designed to nurture this growing network of leaders. LEAD ensures that all of the members of the LEAD family (associates and fellows) are kept in close touch with each other and with other, relevant organizations, through LEAD's own electronic network, LEADnet.

LEAD presently has nearly 1000 fellows and associates from over 40 countries through its 13 national and regional member programs in Brazil, Canada, China, the Commonwealth of Independent States, Europe, Francophone Africa, India, Indonesia, Japan, Mexico, Nigeria, Pakistan, and six countries of southern Africa. Studies to expand the LEAD Program to the United States are presently underway.

More information about LEAD may be obtained from its web site: www.lead.org.

JULIA MARTON–LEFEVRE *USA*

Leapfrogging Technology

José Goldemberg
University of São Paulo, São Paulo, Brazil

Growth and development, which are the fundamental aspirations of 70% of the world's population, can lead to disastrous consequences from the environmental viewpoint. It is argued that this disastrous course of events can be avoided by leapfrogging the steps followed in the past by industrialized countries and by incorporating currently available modern and ef cient technologies early in their development process.

Presently, only one-fourth of the world's population has reached a standard of living that can be considered acceptable. In the low income economies of the developing world, gross domestic product (GDP) *per capita* is only one-tenth as high as in the Organization for Economic Cooperation Development (OECD) countries. *Per capita* consumption of raw materials and energy is also about 10 times smaller. Of the three-fourths of the world's population spread among more than a 140 non-OECD countries, only a small fraction has reached a reasonable standard of living. The rest live a little above, or below, the absolute poverty line.

Such disparities will not last forever. Growth and development are fundamental aspirations of the 70% of the world's people who live in the poor countries of Africa, Latin America, the Middle East and South Asia. Indeed, the economies of a number of very populous developing countries are growing rapidly. In some areas, GDP *per capita* is likely to soon approach the level of the developed countries.

Long-term studies of *intensity-of-use curves* (measuring mass of material used per unit of income) have shown that, in general, they rise and then fall with further development. (Mandelbaum, 1978). These curves indicate that:

1. the intensity of use of many materials follows a predictable pattern: as *per capita* GDP increases from 0, intensity-of-use increases, reaches a maximum at moderate levels of *per capita* GDP, and then declines;
2. peak intensities of use of a material by different societies tend to fall over time.

Such behavior is particularly striking for curves describing industrialized countries' intensity-of-use of energy (usually called *energy intensity*, or energy consumption per unit of GDP) (Figure 1).

As Figure 1 shows, energy intensity grows during the initial phase of industrial development as heavy infrastructure is put in place, peaks, and then declines (Martin, 1988). Countries that industrialize later go through the same pattern as their predecessors, but with less accentuated peaks: they do not have to reach such high energy intensities, because they benefit from modern methods of manufacturing and more efficient systems of transportation. This was

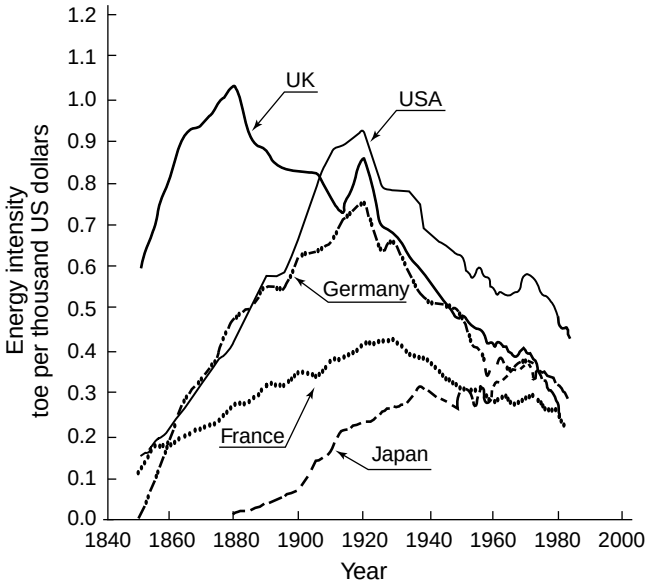


Figure 1 Long-term trends in energy intensity of industrialized countries. Only commercial energy is considered in this analysis. Commercial energy includes all energy that is the subject of monetary transactions (generally coal, oil, gas, nuclear, and hydro). *Source:* Martin (1988)

true even before the oil crisis of 1973, when rising oil prices accelerated the pace of energy efficiency improvements in industrialized countries.

The energy intensity of most industrializing countries is increasing although, in general, it is still lower than in industrialized countries (Goldemberg, 1996). Since most energy growth is in the developing countries, this trend portends a potential energy and environmental disaster.

The strategy of promoting energy efficiency – i.e., performing the same task with less energy – which has been successful in the OECD countries since 1973 is necessary but not sufficient for developing countries. As developing countries build modern economies with attendant industrial infrastructure, transportation systems, and urban development, growth in commercial energy consumption is inevitable.

But developing countries have an alternative: rather than mimicking the industrialized nations, going through an economic development phase that is dirty and wasteful, creating an enormous legacy of environmental pollution, they can leapfrog over some of the steps originally followed by industrialized countries, and incorporate currently available modern and efficient technologies into their development process. Less-developed countries are important theaters for innovation and leapfrogging, especially in energy-intensive, basic material industries such as steel, chemicals, and cement.

1. The adoption of cellular telephones can supplement and sometimes replace the traditional wired telephone systems, such as in Manila or certain regions in China. Although cellular telephones were originally developed for mobile uses or rural areas where wiring is expensive, technical developments have made them economically competitive for regular service.
2. The restructuring of the world steel industry opens new possibilities for developing countries to enhance their comparative advantages. In the past 5 years, since 1995, large conventional steel mills, requiring the use of large blast furnaces, coke ovens and sintering plants, have come under attack for their environmental impacts, including toxic and carcinogenic by-products. In many parts of Europe, licenses for new plant construction are almost impossible to obtain. One result has been an increase in the use of electric furnaces. These furnaces were used in 35% of the world steel production in 1995, compared to 10% in the 1960s. This technology depends on the availability of low-cost electric energy, which is abundant in many of the developing countries to which the modern steel industry is migrating. Another trend has been toward decentralized, small and medium size mills with a production capacity of less than 1 million tons per year.

The potential for leapfrogging is inherent in both process and products. Often, there is synergy between the two, as between the use of renewable energy sources and energy efficiency. Take, for example, lighting in isolated villages, typically supplied by kerosene lanterns, batteries, or candles. Switching to a compact uorescent light bulb, which is four times as efficient as a conventional incandescent bulb, makes it economical to supply power from a solar photovoltaic (PV) panel. Connecting to an electric grid – probably a requirement if inefficient conventional bulbs are used – is unnecessary, generating vast savings in capital equipment.

A second major dividend from leapfrog technologies derives from the avoided costs of long-term environmental clean-up. Most current environmental expenditures (approximately $100 billion per year in the US) go to mopping up old toxic sites, supplying power from a solar PV panel, installing scrubbers in coal-fired power plants, and so on. Connecting to an electric grid – probably a requirement if inefficient conventional bulbs are used – is unnecessary, generating vast savings in capital equipment. A significant fraction of health care costs are linked to environmental pollution and degradation. Using leapfrog energy technologies minimizes these future costs.

Market forces might not suffice to make transportation systems pollution-free, since the benefits are largely public but the costs are largely private. Government incentives and national and international regulations will be necessary to transform this sector. With some assistance, technological revolutions in batteries, fuel cells, and drive systems, as well as in the use of raw materials, could very significantly reduce energy use and accelerate the path to a pollution-free world (Goldemberg, 1998).

REFERENCES

Goldemberg, J (1996) A Note on the Energy Intensity of Developing Countries, *Energy Policy*, 24(12), 759–761.

Goldemberg, J (1998) Leapfrog Energy Technologies, *Energy Policy*, 26, 729–741.

Mandelbaum, W (1978) *World Demand for Raw Materials in 1985 and 2000*, McGraw-Hill, New York.

Martin, J M (1988) L'Intensité Energetique de l'Activité Economique dans les Pays Industrialisés: les Evolutions de Très Longue Periode Livrent-elles des Enseignements Utiles? *Economies et Societés - Cahiers de L'ISMEA*, 22, 9–27.

Life-cycle Assessment

see Environmental (Eco) Labeling (**Volume 4**)

London Dumping Convention

Thomas A Mensah
London, UK

The London Dumping Convention is the short name for the 1972 Convention on the Prevention of Marine Pollution by Dumping of Wastes and Other Matter adopted by an International Conference held in London from the 30th of October to the 13th of November 1972.

The object of the convention is to promote international action to control the pollution of the sea by dumping. *Dumping* is defined as any deliberate disposal at sea of wastes or other matter from vessels, aircraft, platforms or other man-made structures.

The wastes and other matter covered by the convention are listed in three Annexes. Dumping of wastes and other matter listed in Annex I (black list) is prohibited absolutely. Those in Annex II (grey list) cannot be dumped at sea except with a prior special permit, and those in Annex III may only be dumped with a prior general permit. Special and general permits are issued by a contracting state if wastes or other matter to be dumped are to be loaded in its territory or by a vessel or aircraft registered in its territory or ying its ag. Each contracting party is required to take appropriate measures to prevent and punish violations of the convention.

A Protocol was adopted in 1996 to align the 1972 convention to the provisions of the 1982 United Nations Convention on the Law of the Sea. The Protocol amends the definition of dumping to include storage of wastes and other matter in the seabed from vessels, aircraft platforms and other man-made structures and abandonment or toppling at site of platforms and other structures for the sole purpose of disposal. The Protocol is more restrictive regarding the substances that can be dumped at sea. It prohibits dumping of all wastes and other matter except the seven categories expressly listed in Annex 1. It also prohibits the practice of incineration at sea. Other innovations are a new procedure for the settlement of disputes and provisions on compliance, including technical assistance to promote ratification and effective implementation by developing countries.

The original depositaries of the convention are the Governments of Mexico, Russia, the UK and the US. Pursuant to article XIV, paragraph 2, of the convention, the contracting parties designated the International Maritime Organization (IMO) to be responsible for secretariat duties in relation to the convention. IMO provides secretariat and other services to the consultative meetings of the states parties to the convention.

Further information may be obtained from the secretariat of IMO, 4 Albert Embankment, London SE1 7SR, England, UK.

LRTAP (Convention on Long Range Trans-boundary Air Pollution)

The convention on LRTAP was signed in Geneva (Switzerland) on 13th of November, 1979 and entered into force on 16th of March, 1983. By January 2001, 47 parties, including the European Community, had ratified the convention. The United Nations Economic Commission for Europe (ECE) is serving as secretariat for the convention. Fifty-five countries in Europe, North America and Central Asia and Israel are members of ECE.

The convention establishes a broad framework for cooperative action on air pollution and sets up a process for negotiating, adopting and implementing concrete measures to control specific pollutants through legally binding protocols. Eight protocols to the convention had been adopted by January 2001.

The protocol on Long-term Financing of the Cooperative Program for Monitoring and Evaluation of the Long-range Transmission of Air Pollutants in Europe (European Monitoring and Evaluation Program, EMEP) was adopted in 1984 and has been ratified by 38 parties. EMEP has three main components: collection of emissions data, measurement of air and precipitation quality and modeling of the movement of air pollutants.

The protocol on the Reduction of Sulfur Emissions of 1985 (the first sulfur protocol) has been ratified by 22 parties. It required parties to reduce annual emissions by at least 30% from 1980 levels as soon as possible and, at the latest, by 1993.

The protocol concerning the Control of Emissions of Nitrogen Oxides (NO_x) of 1988 has been ratified by 28 parties. It required parties, as a first step, to limit their nitrogen emissions to their 1987 levels by the end of 1994 and to apply best available technology to major new stationary and mobile sources.

The protocol concerning the Control of Emissions of Volatile Organic Compounds (VOCs) of 1991 has been ratified by 20 parties. In the protocol, parties committed themselves to reduce VOC emissions by at least 30% by

1999, using 1988 or any other year between 1994 and 1990 as the base year.

The protocol on Further Reduction of Sulfur Emissions (the second sulfur protocol) of 1994 has been ratified by 23 parties. The protocol was the first to use the *critical loads* concept, assigning to each country a reduction target based on the effects its sulfur emissions have on the region's ecosystems (*see* Critical Load, Volume 3).

The protocol on Heavy Metals of 1998 has been signed by 36 signatories. It aims at cutting emissions of cadmium, lead and mercury, including phasing out of leaded petrol.

The protocol on persistent organic pollutants (POPs) of 1998 was signed by 36 signatories. Its objective is to control, reduce or eliminate discharges, emissions and losses of certain POPs into the environment.

The protocol to Abate Acidification, Eutrophication and Ground level Ozone of 1999 has been signed by 31 signatories. The protocol sets emission ceilings for 2010 for four pollutants: sulfur, NO_x, VOCs and ammonia.

The three latest protocols do not yet have the needed 16 ratifications to enter into force.

According to the latest review of the implementation of the protocols in force (done in 1999) all parties to the first sulfur protocol met the required reductions in 1993. Some countries have reduced their emissions much more than required in the protocol.

Of the 28 parties to the nitrogen oxides protocol, 19 met the requirements in each of the years 1994–1996. Regarding the protocol on VOC seven parties have already achieved the target levels of the protocol. A further four have not yet met the target but appear to be on course to do so. The targets for 2000 in the second sulfur protocol have been met by 15 parties, whilst another four are on course to do so.

None of the protocols has been ratified by all parties to the convention. The reasons for non-ratification of the various protocols differ from case to case. Country replies to questionnaires point to problems of getting agreement through national procedures (ministries and parliament) or to difficulties in achieving the targets set in the protocols. In certain countries overall economic problems make some emission controls a lower priority. A considerable amount of effort is usually required by national civil servants before national procedures enable ratification.

See also: Convention on Long-range Transboundary Air Pollution (LRTAP), Volume 4.

KAJ BARLUND *Switzerland*

M

MAB (UNESCO)

see UNESCO (United Nations Educational Scientific and Cultural Organization) and the Environment (**Volume 4**)

MacNeill, Jim

(1928-)

Jim MacNeill was born on April 22 in 1928. He has enjoyed nearly four decades of experience as a policy advisor on environment, energy, natural resources management and sustainable development to leaders of international organizations, governments and industry. He is an author, speaker, member of several boards, and the recipient of many honors and distinctions, national and international. Currently, he is chairman of the World Bank's Independent Inspection Panel in Washington, DC. He is also chairman of the International Institute for Sustainable Development (IISD) in Winnipeg. He is a member of the board of the Woods Hole Research Center, Woods Hole, MA, the Wuppertal Institute on Climate and Energy Policy in Germany and the Environmental Education and Training Institute of North America (EETINA), Mexico City. He is also editorial chairman of the Earth Times, New York, and member of the jury for the Volvo Environment Prize, Gothenburg, Sweden. Recently, he served as a member of the High Level Advisory Group established by the secretary-general of the Organization for Economic Cooperation and Development (OECD) to advise on the future of that organization.

He began his career in Saskatchewan in 1952 as an economist and later as an engineer/manager with responsibility for coordinating development of the power, irrigation, and other benefits of the huge, multipurpose, South Saskatchewan River Dam Project. Moving to Canada's Department of Energy, Mines and Resources in 1965, he was responsible for policy and later for federal responsibilities for water and renewable resources, including marine resources.

In 1970, during a 2-year stint as an advisor to Prime Minister Pierre Trudeau, he authored the first major book published in Canada on *Environmental Management*. In 1971, the book played a major in uence in shaping the policies of Canada's new Department of Environment. In addition, he played a key role in the establishment of that department, one of the firsts in the western world. He led Canada's preparations for the 1972 Stockholm Conference on the Human Environment, while acting as a senior advisor to the United Nations (UN) Secretary-General of that Conference. Later, as deputy minister of the federal Ministry of Urban Affairs, he played a major role in shaping Canada's urban policy. This led to his appointment as Canada's Ambassador and Commissioner General for the 1976 United Nations Conference on Human Settlements (Habitat), held in Vancouver.

Moving to the international arena, Dr MacNeill served for 7 years in Paris as Director of Environment for the OECD. While at OECD, he negotiated several international agreements on policies and programs to enhance environmental quality.

In the same year, Norway's Prime Minister, Gro Harlem Brundtland, invited him to become a member and Secretary-General of the World Commission on Environment and Development (the Brundtland Commission). He was the principal architect and major author of the Commission's world-acclaimed report *Our Common Future*. From 1989 to 1992, he acted as a special advisor to the Secretary-General of the Earth Summit. He also established and chaired the EcoFund, which provided over $5 million in special funding for the Rio Conference.

MacNeill is the author of a number of articles, publications and books. He is also the recipient of a number of awards including the Order of Canada, his country's highest honor, the highest achievement award of his native Province of Saskatchewan, the City of Paris' Silver Medal, the Merit Award of the Climate Institute, and the Lifetime Achievement Award of Environment Canada.

Born on Earth Day, Jim MacNeill was an ecological refugee at the age of one, when his family moved from the Southern Dust Bowl to the Central Park Belt of the Province of Saskatchewan. He is married with two children and has four grandchildren.

MOSTAFA K TOLBA *Egypt*

Man and the Biosphere (UNESCO)

see UNESCO (United Nations Educational Scientific and Cultural Organization) and the Environment (**Volume 4**)

Management of Acid Sulfate Soils

see Acid Sulfate Soils (**Volume 3**)

Management of Agricultural Nitrate Issues

see Nitrogen: Agricultural Uses and their Impacts (**Volume 3**)

Management of Agricultural Nitrogen Emissions

see Nitrogen: Agricultural Uses and their Impacts (**Volume 3**)

Management of Disturbance-prone Systems

see Disturbance (**Volume 2**)

Marine Environment Protection Committee (MEPC)

see MEPC (Marine Environment Protection Committee) (**Volume 4**)

Marine Pollution: the GESAMP Definition

see GESAMP (Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection) (**Volume 4**)

MARPOL/73

see MEPC (Marine Environment Protection Committee) (**Volume 4**)

Materials Flow Accounting

William M Stigliani
Center for Energy and Environmental Education, University of Northern Iowa, Cedar Falls, IA, USA

Materials flow accounting (MFA) is an invaluable tool for solving problems of chemical pollution. In essence, MFA provides a tracking system for the flow of a pollutant as it progresses through different stages of its lifetime in the industrial economy, and identifies points along the flow where the pollutant is released to the environment. The important stages include: (1) extraction of raw materials, (2) production (manufacturing), (3) storage and transport, (4) consumption, and (5) waste disposal. Recycling can have an important influence on the mass flow because it reduces the quantities of materials in the extraction, production, and transport stages, as well as pollution generated within these stages.

The guiding principle of MFA is the law of mass balance. Materials cannot disappear. When their useful lifetime in the economy is over they are returned to the environment, usually in a degraded form called pollution. The physical laws of nature determine whether the pollutant is stored or mobilized, persistent or decomposed, dissipated or bioaccumulated. Estimations of mass flows in the industrial economy help to insure consistency with estimates

of mass flows in the environment. In fact, MFA can provide invaluable information, often lacking in environmental assessments, in identifying sources of pollutants, quantifying their loads to the environment, and insuring that estimates of such loads are neither too high nor too low.

The MFA technique is flexible with regard to its scale of application. At the factory level, MFA can account for major waste streams and help to discern how they can be reduced via recycling or reuse. Often, a waste from one process is a feedstock for another, either in-house or in a different industry. So called waste exchange networks are already in place in many nations that serve to broker the transfer of one industry's waste to another's feedstock.

On the other end of the spectrum, MFA can be applied at the regional to continental scales, where it can be an instrument for analyzing and quantifying cross-boundary pollution. MFA has an important role to play as cooperating nations devise policies for reducing air, water, and solid waste pollution between and among states, provinces, and nations.

METHODOLOGY

As shown in Figure 1, emissions of pollutants to the environment can occur at all stages of its pathway through the industrial economy. The quantity of an emission released can be calculated as the product of two indices: (1) the quantity of pollutant in a given stage (represented here as Q_i); and (2) an emission factor that is a fraction of the quantity that goes to air, water, or soil (represented as EM_{ic}). The index i is keyed to the five stages indicated in the figure; $i = 1$ for the extraction of raw materials stage, $i = 2$ for the production stage, etc. The index c in EM_{ic} represents the three compartments of the environment. As defined here, $c = 1$ for atmosphere, $c = 2$ for water, and $c = 3$ for soil. For example, the emission of pollutant to the atmosphere during stage 2 (production) can be calculated (Equation 1):

$$\text{emission to air} = Q_2 \times EM_{21} \qquad \bullet 1 \bullet$$

By similar calculations total emissions to air, water, and soil over all stages in the flow can be summed in Equations (2), (3) and (4):

$$\text{total emissions to air} = \sum Q_i \times EM_{i1} \qquad \bullet 2 \bullet$$

$$\text{total emissions to water} = \sum Q_i \times EM_{i2} \qquad \bullet 3 \bullet$$

$$\text{total emissions to soil} = \sum Q_i \times EM_{i3} \qquad \bullet 4 \bullet$$

Typically, the MFA analysis is broken down into three steps:

1. identifying the sources of the chemical being assessed for each of its stages in the industrial economy;
2. assigning values to Q_i and EM_{ic} and quantifying chemical emissions to air, water, and soil;
3. assessing current chemical flows and their pathways into the environment, and estimating how proposed environmental policies would shift mass balances and affect environmental impacts.

Despite the power of the MFA methodology, assigning quantitative values to Q_i and EM_{ic} is often not entirely straightforward. Rather, the data must be pieced together from a variety of sources, making the task extremely labor intensive. Some important references include: reports of various industries and trade associations; state, provincial, and national accounts providing statistics on economic, environmental, and production data; publications of international organizations such as those of the United Nations Environment Programme and the United Nations Industrial Development Organization; articles in the environmental and economic literature; direct environmental monitoring data; and model estimates based on engineering principles.

A system of national accounting that harmonizes economic, environmental, and production/consumption data would be a vast improvement over the current availability of data, and would allow MFA much broader application than it now enjoys. Data availability has slowly improved, however, over the last 10–20 years. One example is the vast database, which began in 1987, of the US Environmental Protection Agency's Toxic Release Inventory.

EXAMPLE OF THE RHINE RIVER BASIN

In contrast to regions defined by political boundaries such as states, provinces, or nations, river basins are perhaps the ideal spatial scale for MFA studies. By definition, a river basin encompasses the land area that drains into the river or its tributaries. Thus, after accounting for pollutant losses to the sea via river transport, and atmospheric transport into and out of the basin (both of which can be monitored and modeled), it can be assumed that emissions of pollutants within the basin remain in the basin. This allows the economic space where the sources of pollution reside to correspond to the environmental space over which the pollutants are discharged and accumulated. Such overlapping of spatial domains greatly facilitates the calibration of economic and environmental data.

The Rhine Basin offers a particularly fertile area of investigation of material flows and environmental impacts because historically it has been one of the most heavily industrialized regions of the world. Although the stream flow of the Rhine comprises only about 0.2% of the flow of all rivers in the world, about 10–20% of the total Western chemical industry (Organization for Economic Cooperation

and Development (OECD) countries) is located in its basin. The area extends over five European nations: most of Switzerland, the northeastern region of France, Luxembourg, most of the Southwestern Lander (provinces) of Germany, and most of the Netherlands. The population of the basin is about 50 million and the area is about 220 000 km^2. About half of the land is devoted to agriculture, one third is forest, and the remainder is urban and suburban land.

The following discussion illustrates the usefulness of MFA as applied to cadmium flows in the Rhine Basin over the period 1970 to 1988. A more detailed discussion can be found in Stigliani and Anderberg (1994).

Step 1: Identification of Sources of Cadmium in the Industrial Economy of the Rhine Basin

Cadmium provides an interesting example because there are two distinct modes by which it enters the industrial economy. One is cadmium's inadvertent interjection into the economy piggy-backed on high volume raw materials that, when finished into consumer products, carry the cadmium as a trace impurity. Phosphate fertilizer is one important example. The ores from which it is made contain concentrations of cadmium in the range of 0.25–80 parts per million. Much of the cadmium is retained in the fertilizer, and is spread on agricultural land.

Table 1 Sources and emissions of cadmium in the Rhine Basin in 1970 and 1988 (tons year^{-1})

Source	Air (1970/1988)	Water (1970/1988)	Land (1970/1988)
Inadvertent sources of cadmium present as impurity			
Natural sources		3.2/3.2	
Coal combustion	45.0/16.2		9.9/19.2
Iron and steel production	48.8/6.8	32.1/3.0	68.3/14.0
Non-ferrous metal refining[a]	29.3/1.5	3.7/0.0	
Zinc product use[b]		4.5/1.1	
Phosphate fertilizer manufacture		27.9/9.8	
Phosphate fertilizer use			47.7/35.9
Cement manufacture	7.8/0.8		
Subtotal	130.9/25.3	71.4/17.1	125.9/69.1
Sources of cadmium related to products containing cadmium			
Zinc/cadmium refining	31.6/3.5	77.0/0.1	60.8/54.6
Products containing cadmium, manufacture	1.7/0.5	30.0/0.5	50.4/11.5
Products containing cadmium, use		2.9/1.5	
Products containing cadmium, disposal	8.4/4.6		675.0/659.0
Subtotal	41.7/8.6	109.9/2.1	786.2/725.1
Grand total	172.6/33.9	181.3/19.2	912.1/794.2

[a] Does not include zinc refining since cadmium is refined as a by-product, and thus is not an inadvertent waste at the zinc refinery.
[b] Refers to zinc products that contain cadmium as an impurity.

As shown in the upper part of Table 1, other important commodities containing trace amounts of cadmium are coal, iron and steel, zinc products, and other non-ferrous metals.

The category, cement manufacture, illustrates the complexity of proper source identification, and highlights the importance of constructing a comprehensive inventory with particular attention paid to sources that at first glance appear to be unrelated to the pollutant in question. There is no intrinsic step in cement manufacturing that depends on the use of cadmium-containing materials. Factories in the Rhine Basin, however, have used old automobile tires as a fuel. In the manufacture of tires, zinc is incorporated as an additive to enhance performance. Trace amounts of cadmium piggybacked on the zinc are thus inadvertently added to the tire. When combusted as a fuel, a tire releases cadmium to the atmosphere.

The lower part of the table indicates the second way cadmium enters the economy. The listed sources are products containing cadmium, where cadmium is added intentionally. The four major products are pigments, nickel–cadmium batteries, cadmium plated steel, and cadmium stabilizer in polyvinyl chloride (PVC) plastic. Cadmium is linked to zinc refining because it is present in trace amounts in zinc ore, typically 1–2%. These concentrations are generally sufficiently high to make separation and refining of the cadmium economically feasible. In fact, it is by this route that all primary cadmium is produced.

Step 2: Estimation of Cadmium Emissions and Deposition to Air, Water, and Soil

Each source indicated in Table 1 was assigned a set of values for Q_i and EM_{ic}. They were multiplied and summed to obtain cumulative emissions to air, water, and land. Changes in emissions are dictated by changes in Q_i and/or EM_{ic}. In general, values of EM_{ic} decline over time in response to increasingly more stringent environmental emission standards, or the adoption of newer and cleaner technologies. Values of Q_i depend on the intensity of production and consumption of materials containing cadmium within the basin. In some cases the materials have increased (e.g., coal), and in others they have decreased (e.g., cadmium plated steel).

The grand total in the table shows that large reductions in air (80%) and water (90%) emissions occurred between 1970 and 1988. These were due to several factors, among the most important being the enforcement of more stringent environmental standards, the shut-down of old factories or the introduction of modern cleaner technologies, and reduction in overall use of cadmium in the Basin.

In contrast, the reduction in deposition to land was only about 13% between 1970 and 1988. One reason for this is that reducing emissions to air and water created an increase in solid wastes. For example, reducing cadmium emissions to the atmosphere from coal burning facilities was attained by collecting more of the y ash that contained cadmium in the smokestack. Thus, reducing emissions to air led

to an increase in fly ash that was sent to landfill. This can be observed in Table 1, which shows coal combustion generated 9.9 tons of emissions to land in 1970, compared to 19.2 tons in 1988. A more important factor is that among the sources of land deposition, disposal of products containing cadmium to municipal landfill (nickel–cadmium batteries, etc.) is by far the largest input, comprising 74% of all landfilled cadmium in 1970, and 83% in 1988. The use of some products containing cadmium increased and others decreased in the 1970s and 1980s, but overall products delivered to landfills declined only slightly over this period.

The entry marked zinc/cadmium refining shows exceedingly large emissions to air and water in 1970, and drastic reductions by 1988. This trend appears surprising given that zinc refineries, even in 1970, were in the business of producing refined cadmium as a lucrative by-product of zinc. In this case, however, there was one antiquated zinc refinery that did not refine cadmium, but rather, viewed it as a waste product. Emission control measures were primitive, and the cheapest option was to unload the cadmium to the air and water. Needless to say, such practices could not survive as clean up of point sources in the basin began in earnest by the late 1970s.

The final point of note is that solid wastes (i.e., emissions to land) dwarf the emissions to air and water. In 1970 when total emissions were around 1264 tons, 14% went to the air and water each, and 72% to the land. Percentage of solid waste to land was even higher in 1988 with a 94% share, compared to 4% and 2%, for air and water, respectively. This is not surprising since products containing cadmium dominate most of the gross cadmium solid flow in the basin.

Step 3: Applying MFA for Assessment of Risk Reduction Policies for Cadmium Pollution in the Rhine Basin: a Hypothetical Example

In comparing the reduced cadmium emissions in 1988 relative to 1970, one might surmise that environmental risks of cadmium pollution in the Rhine Basin are declining. Our MFA model, however, allows us to probe beyond superficial assessments based upon limited information. As an illustration, let us test the effectiveness of a proposed policy that would ban by 2010 all products containing cadmium except for nickel–cadmium batteries, with a target that 50% of the batteries would be recycled, thus averting their disposal in municipal landfills.

Figure 2 shows the flow of products containing cadmium in the Rhine Basin in the mid-1980s, as determined by the MFA model. One may notice that because there was a shortage of cadmium metal needed for the manufacture of the products, 570 tons of refined cadmium was imported. Also, with the exception of industrial batteries, there was very little recycling. As a result of this flow, most of the cadmium (almost 700 tons) ended up as solid consumer waste, of which over 80% went to municipal landfill and the balance to incineration.

Figure 3 shows how the flows would change if the proposed policy were fully enacted. One consequence of the policy would be creation at the zinc/cadmium refineries of a surplus of cadmium in contrast to the scarcity that existed in the mid-1980s, when 570 tons had to be imported.

The problem arises because the supply of cadmium into the Rhine Basin is essentially inelastic. It enters the basin as a by-product of zinc refining, which still maintains a high degree of productivity. The fact that 50% of the batteries would be recycled increases the surplus even further. Thus, the amount of cadmium ending up in municipal landfills would indeed be sharply reduced to around 100 tons. However, cadmium-laden solid waste would increase at refineries from about 30 tons in the mid-1980s to more than 450 tons if the policy came into in force. This analysis suggests that the result of the policy is merely to transfer the solid waste load from the municipal landfill to the landfill into which the zinc refinery wastes are deposited. Thus, the policy would most likely have little positive benefit with respect to cadmium's impact on the environment.

With respect to human health, the major exposure pathway is:

$$\text{cadmium in agricultural soil} \longrightarrow \text{cadmium in vegetables} \longrightarrow \text{cadmium ingested in human diet}$$

Any policy targeted at lowering risks to human health should pay close attention to input of cadmium to agricultural land. Cadmium in safely engineered and maintained landfills, whether municipal or industrial depositories, is less likely to constitute a serious hazard to human health. We can see from the last column of Table 1 that cadmium deposition to land from phosphate fertilizer use differs significantly from other sources. It is the only one that is not delivered to an industrial or municipal landfill. Rather, it is inadvertently applied directly to agricultural land as a trace impurity of phosphate fertilizer. Even though in 1988 its total on agricultural land constituted only about 9% of all cadmium deposited on land, its risk to human health is magnified many times because of its direct linkage to human exposure via crop uptake.

Having identified this key exposure pathway, MFA can carry the assessment to the next level of complexity. It is possible to construct a mass balance for cadmium in agricultural soils by estimating its inputs and outputs. The results are shown in Table 2, which identifies three inputs and three outputs.

Atmospheric deposition is related to the sources identified in the air column in Table 1. Some of these air emissions are eventually deposited on agricultural land, either in wet form with rain or in dry deposition. Sewage

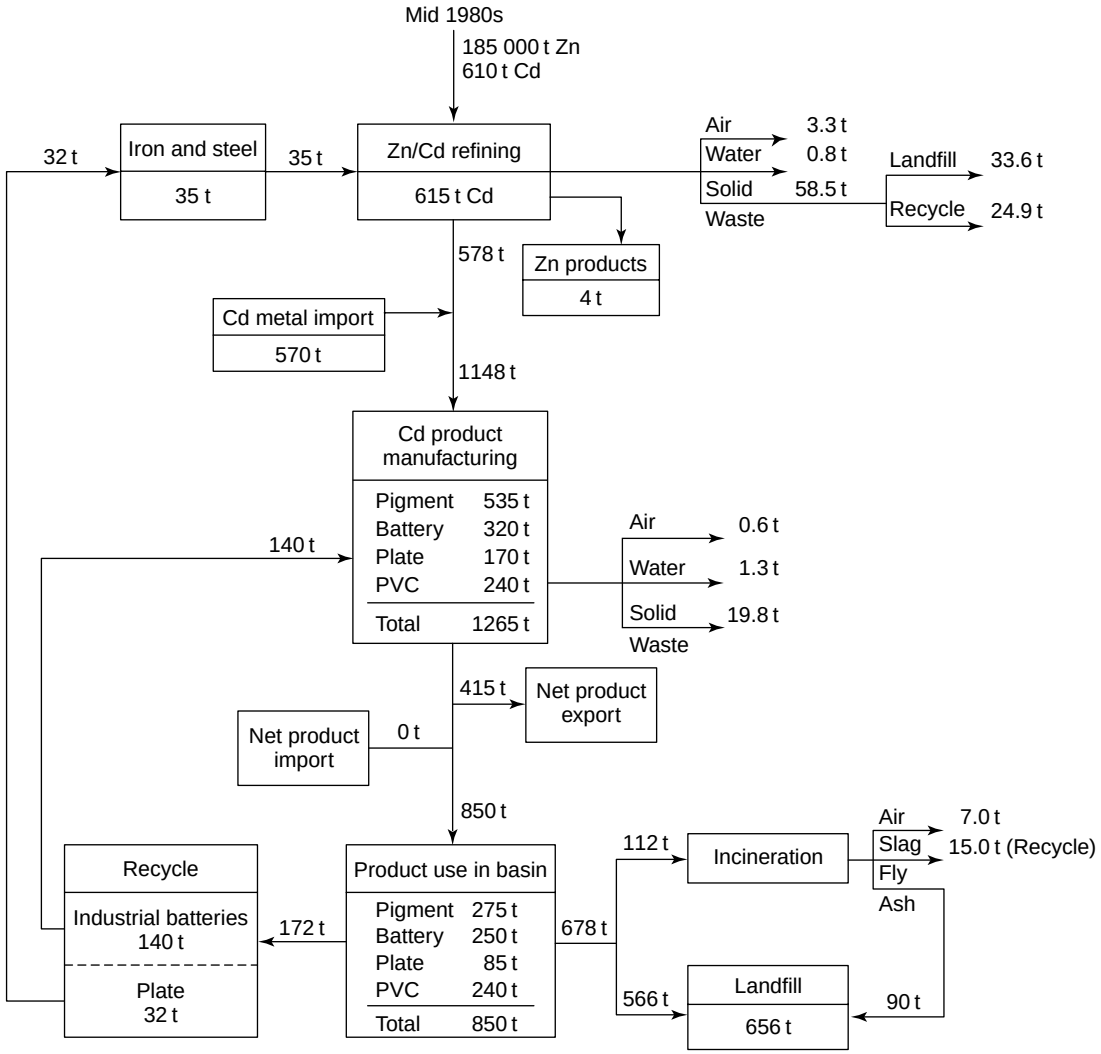


Figure 2 Balance of cadmium product use and disposal in the Rhine River Basin in the mid-1980s

sludge containing cadmium is applied as a fertilizer in some areas, although its overall use basin-wide is rather limited. We can see that 30 years earlier, total inputs of cadmium were almost double what they were in 1988. Also, atmospheric deposition was the major source to croplands in 1970, but by 1988 it was surpassed by phosphate fertilizer.

The three outputs of cadmium from agricultural land are plant uptake, runoff, and erosion. One important observation is that while all cadmium input to the land declined between the years 1970 and 1988, the uptake by crops increased (i.e., compare 2.2 tons in 1970 to 4.1 tons in 1988). The reason for this is that over the entire time period, inputs continued to exceed outputs, even though by 1988 the net positive balance was less than half what it was in 1970. A positive balance indicates that the soils were still accumulating cadmium, and its soil concentration was increasing as well. Since plant uptake of cadmium depends on soil concentration, it is not surprising that uptake would continue to increase.

The above analysis suggests that to reduce the public health risk of cadmium toxification, the most effective step would be removal of cadmium from phosphate fertilizer rather than banning cadmium products. It is curious that to date this has not been accomplished. One reason is that current regulations work against this measure. Removing cadmium from the final product is neither difficult nor too expensive. The expensive and laborious step is managing the cadmium waste after it is removed from the phosphate. Cadmium solid waste, like other toxic materials, is tightly regulated. Environmental enforcement agencies impose stringent engineering standards and a demanding maintenance protocol. To avoid the difficulty of landfilling and safely maintaining a non-leaking depository, it is cheaper and far less bothersome to keep the cadmium in the fertilizer and spread it in diluted

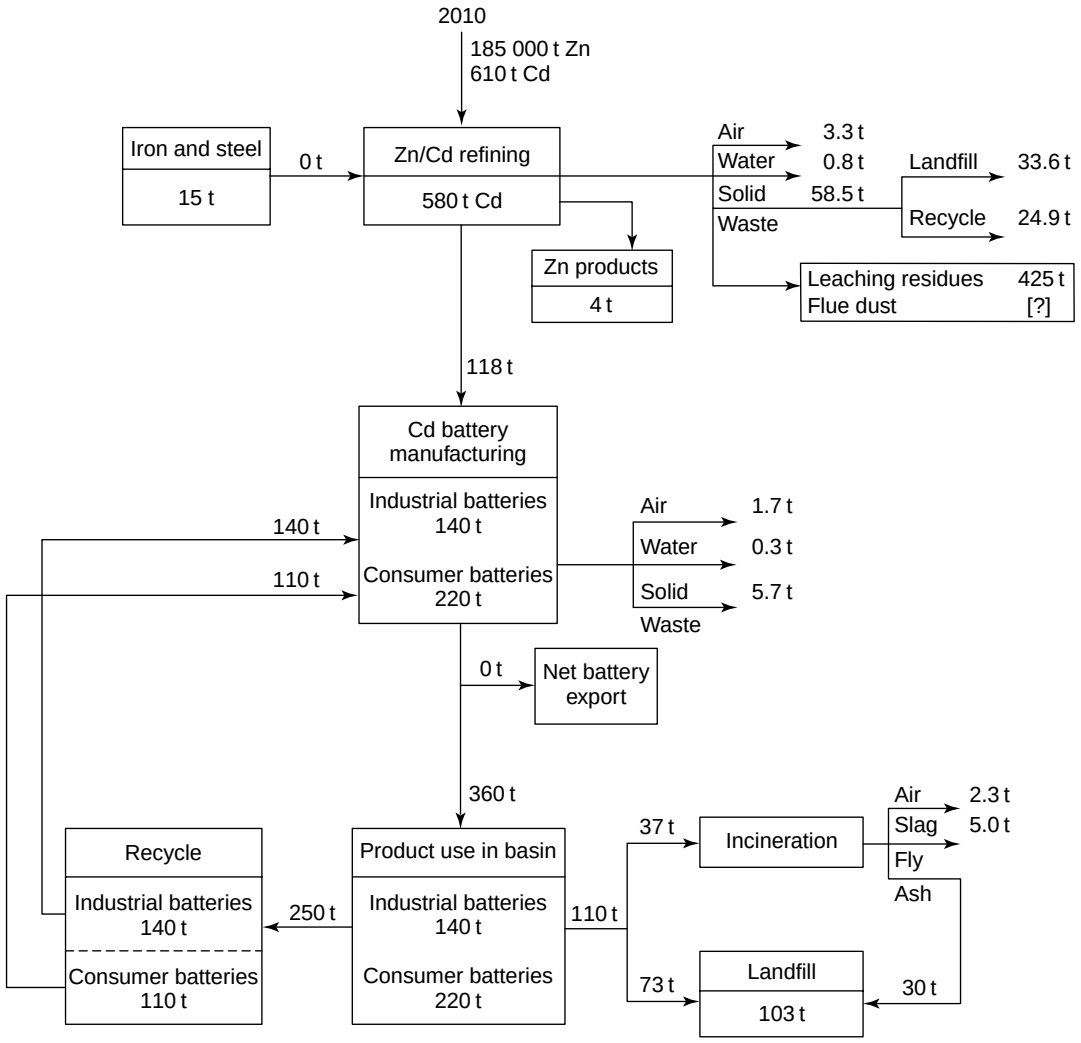


Figure 3 Balance of cadmium product use and disposal in the Rhine River Basin: 2010 Scenario for Cd product ban except for nickel–cadmium batteries

Table 2 Inputs and outputs of cadmium to agricultural soils in the Rhine Basin (tons year^{-1} and percentage of total)

Process	1970 (%)	1988 (%)
Inputs		
Atmospheric deposition	+61.6 (55.2)	+13.2 (26.3)
Phosphate fertilizer	+47.7 (42.8)	+35.9 (71.7)
Sewage sludge	+2.2 (2.0)	+1.0 (2.0)
Subtotal	+111.5 (100)	+50.1 (100)
Outputs		
Plant uptake	−2.2 (20.0)	−4.1 (51.9)
Runoff	−6.8 (61.8)	−1.4 (17.7)
Erosion	−2.0 (18.2)	−2.4 (30.4)
Subtotal	−11.0 (100)	−7.9 (100)
Inputs minus outputs[a]	+100.5	+42.2

[a] Because inputs are larger than outputs, there is net accumulation of cadmium in agricultural soils in both 1970 and 1988.

amounts over agricultural land. It seems that more prudent policies could be enacted to encourage the removal of cadmium during the phosphate production process while at the same time easing the regulatory burden on the manufacturers.

CONCLUSION

MFA can be an invaluable tool in the management of chemical pollution. Identifying sources and quantifying ows of pollutant-laden materials through the industrial economy provides a basis for a comprehensive understanding of the most serious threats to the environment and human heath. It is especially useful for exposing pollutant sources and pathways that might otherwise be hidden from the assessor.

MFA provides a benchmark against which environmental managers can test prospective policies. It sheds light on how

a proposed policy might change a pollutant's mass balance as well as its flows and exposures in the environment. As provided in the above example of cadmium in phosphate fertilizer, we have seen how a relatively minor flow can have a disproportionately high impact when it is unregulated and feeds directly into important exposure pathways.

Holistic solutions to environmental pollution should include MFA as a complement to traditional environmental analysis. Both kinds of studies are needed for comprehensive understanding of societal causes of trends in pollutant emissions and their environmental effects. A great step forward would be the establishment of a national accounting system that harmonizes economic and environmental data with data on production and consumption.

See also: Environmental Cadmium in the Food Chain, Volume 3.

REFERENCES

Stigliani, W M (1993) The Integral River Basin Approach to Assess the Impact of Multiple Contamination Sources Exemplified by the River Rhine, in *Integrated Soil and Sediment Research: a Basis for Proper Protection*, eds H J P Eijsackers and T Hamers, Kluwer Academic Publishers, The Netherlands, 359–384.

Stigliani, W M and Anderberg, S (1994) Industrial Metabolism at the Regional Level: The Rhine Basin, in *Industrial Metabolism - Restructuring for Sustainable Development*, Chapter 7, eds R U Ayres and U E Simonis, United Nations University Press, Tokyo, 119–162.

Maximum Sustainable Yield (MSY)

Starting late in the 19th century, governmental managers with formal responsibilities for regulating the harvests of natural resources that were not privately owned (forests, birds, mammals, fish, shellfish, etc.) came to interpret success in that role as depending on achievement of MSY of the most valued population in the relevant lands and waters. Maximization was as important as sustainability, in that each resource sector was expected to do its best for a nation's economic progress.

In retrospect we note that such an early version of MSY is based implicitly on an ecological concept of an ecosystem's carrying capacity for the valued population. Also, there is a presumption that the target population and its host ecosystem will maintain their integrity by adapting systemically to the external harvesting of that population at the MSY level. Finally, it is to be noted that this is implicitly a high-grading approach, in that the harvester selects only the most valued population available at the time, but may do so carelessly and destructively of other features of the ecosystem (*see* Carrying Capacity, Volume 4; Ecosystem Integrity, Volume 4).

By the 1950s, the MSY goal was being applied to all lands and waters of North America that were not zoned as protected natural parks or private property. By about 1970 the goal had been extended to all the world's oceans. And by 1970 it was becoming apparent to many researchers that this goal was grossly simplistic, at best. For example, various other human uses of these ecosystems were affecting the valued resource populations ever more intensely, and almost always adversely. Also, fisheries were increasingly targeting a mix of several interacting populations of fish and shellfish in an ecosystem. Hence a resource population could not be managed as though it were isolated ecologically, as presumed a century earlier with a simple MSY approach.

MSY was initially interpreted as relating to the harvest in terms of biomass or total mass of an individual population or set of interacting populations of a particular valued species. If the harvested biomass was then sold, a price could be observed and a comparable goal of maximum sustainable economic yield (MSEY) could perhaps be specified, at least if the costs of harvesting could also be included in the calculations. It was quickly learned that MSEY implied that the harvesting goal in terms of biomass was less than MSY and the term *optimum sustainable yield* (OSY), came to be the practical goal, at least in some locales where the political process permitted some rationality.

Simplistically MSY, as the amount of biomass of a population harvested per year, can be equated to the amount of mass that the population can add through reproduction and growth in that year, with the population at a size at which such an addition is maximized. Slowly growing populations of whales, large fish and tree species in cold habitats may add only a few percent of biomass per year under those conditions, so the annual MSY amounts to only a few percent of the biomass of that population. Fast growing populations of small fish and shellfish may double their biomass under MSY conditions, so that the total harvest over one year, if appropriately distributed seasonally, may equal the amount of biomass of the population at the beginning of the year.

An accountant who takes into consideration the opportunity cost of capital invested in a resource might find that earnings attract a higher rate of return elsewhere in an economy than the rate of growth in biomass of some natural resources. Under such circumstances the financial returns to the entrepreneurial harvester of a slowly growing population could be maximized if that population were rapidly over-exploited to the point of mining the resource,

and investing the quick profits at higher rates of return elsewhere. In these circumstances a broadened version of MSEY, which focused on the productivity of economic capital, could justify the unsustainable over-exploitation of a slowly growing resource. This kind of economic development has often occurred with respect to forests, whales and large fish species.

By the end of the 20th century, a simplistic linear approach to management of natural resources, as exemplified by the use of MSY or OSY for a few of an area's most valued species, had fallen into disrepute. The macroeconomic considerations of economic development, simply over-rode any ecological considerations on many occasions. Even where there was still a commitment to ecological sustainability, it had become evident through much bitter experience that the population dynamics models used for estimating the annual or periodic total allowable harvests of valued populations could not be made sufficiently accurate, for annual adjustments to harvest regulations, with any reasonable level of funding available for such information services. An alternative, more systemic regime seems to be emerging, with an ecosystem approach (*see* Ecosystem Approach, Volume 2) to an understanding of the natural resources in their habitats, together with a pluralistic decision-making regime along the lines that had evolved over centuries in many locales the world over with respect to common property resources.

Assimilative capacity is a similar concept to MSY both practically and politically, but relates to loadings of wastes rather than removals of resources. It denotes the maximum amount of a particular pollutant that can be assimilated by a particular ecosystem in a year. Assimilation, in turn, may mean simple dilution, or transport and burial by natural processes, or gradual degassing to the atmosphere, or biological inactivation, or other processes. It too is in disrepute, and new alternative approaches are coming into place, one of which resembles that for natural resources sketched above.

Sometimes the term MSY is also applied to intensive culture of farms of many kinds. Some of a farm's limiting factors (*see* Carrying Capacity, Volume 4) may be progressively relaxed to increase the maximum yield. A question then arises as to which if any of the farm's original natural features are to be maintained in such an altered ecosystem, and how that would then constrain further maximization.

HENRY A REGIER *Canada*

Mediterranean Action Plan

see Mediterranean Sea (*Case study*, **Volume 4**)

Mediterranean Sea

see Mediterranean Sea (*Case study*, **Volume 4**)

MEPC (Marine Environment Protection Committee)

An organ of the International Maritime Organization (IMO), the MEPC was established in November 1973 by resolution A.297 (VIII) of the IMO Assembly as a permanent subsidiary body of the Assembly with the mandate to coordinate and administer ••• the activities of the organization concerning the prevention and control of marine pollution by ships . It was also to perform such functions as are or may be conferred upon the organization under international conventions for the prevention and control of pollution from ships.

By resolution A. 296 (VIII) MEPC was designated as the appropriate body to perform functions concerning the amendment of the International Convention for the Prevention of Pollution from Ships, 1973 (MARPOL/73) and the 1973 Protocol Relating to Intervention in the High Seas in Cases of Marine Pollution by Substances other than Oil, both of which had just been adopted.

Following the amendment of the IMO Convention by resolution A.317 (Es.V) of 1974, MEPC became a principal organ of IMO in 1982 with the mandate to consider any matter within the scope of the organization concerned with the prevention and control of marine pollution from ships. It may also

> consider appropriate measures to facilitate the enforcement of conventions, provide for the acquisition of scientific and technical information on the prevention and control of marine pollution from ships for dissemination to States ••• and consider and take appropriate action with respect to any other matters falling within the scope of the organization which would contribute to the prevention and control of marine pollution from ships •••.

MEPC is responsible for implementation of MARPOL/73 as modified by the Protocol of 1978 relating to the International Convention for the Prevention of Pollution from Ships (MARPOL 73/78), the 1990 International Convention on Oil Pollution Preparedness, Response and Cooperation and major aspects of the activities of IMO on the protection of the marine environment.

See also: IMO (International Maritime Organization), Volume 4.

THOMAS A MENSAH *UK*

Meta-analysis

see Use of Epidemiological Data to Develop Air Quality Standards for Short-term and Long-term Environmental Health Benefits (**Volume 4**)

MINK (Missouri, Iowa, Nebraska, Kansas) Study

see MINK (Missouri, Iowa, Nebraska, Kansas) Study (*Case study*, **Volume 4**)

Mitigating Land Degradation in the Mediterranean

see Land Degradation in the Mediterranean (**Volume 3**)

Mitigation

see Adaptation Strategies (*Opening essay*, **Volume 4**); Natural Hazards: Social and Policy Responses (**Volume 4**)

Monitoring in Support of Policy: an Adaptive Ecosystem Approach

see Monitoring in Support of Policy: an Adaptive Ecosystem Approach (*Opening essay*, **Volume 4**)

Montreal Protocol, Multilateral Fund for the Implementation of

Omar E El-Arini
Multilateral Fund Secretariat, Montreal, Canada

While the original Montreal Protocol on substances that deplete the ozone layer (UNEP, 1987), adopted in 1987, provided for technical assistance to promote participation and implementation, it did not include any arrangement for nancial assistance to developing countries to meet the control measures applicable to them. Such an arrangement - the nancial mechanism - was put in place through the London amendment to the Protocol adopted in 1990 (UNEP, 1990). According to Article 10 of the amended Protocol, the parties are to establish a mechanism to provide nancial and technical cooperation, including the transfer of technologies to parties operating under paragraph 1 of Article 5 of the Protocol to enable their compliance with the control measures and to meet all agreed incremental costs of achieving such compliance. (Developing countries that are party to the Montreal Protocol and whose annual per capita consumption of controlled substances is less than 0.3 kg for Annex A substances are generally referred to as simply Article 5 parties. Other Parties (from developed countries) are referred to as non-Article 5 parties.) This mechanism would include a multilateral fund. Pending the entry into force of the London Amendment, the parties decided in decision 2/8, to set up an interim mechanism, including an interim multilateral fund, which began operation in 1991 (UNEP, 1990).

The London Amendment entered into force in August 1992 and, accordingly, the fourth meeting of the parties, held later that year, formally established the nancial mechanism, including the multilateral fund and transferred to the latter the resources remaining in the interim fund (UNEP, 1992c). The multilateral fund became operative in January 1993. Over the years the protocol was amended four times and adjusted ve times, and 130 developing countries have become parties to the Protocol, 110 of which have received nancial assistance from the multilateral fund amounting to US$1.13 billion to phase out 135 000 ODP (ozone depletion potential, one ODP tonne of a controlled substance equals one metric tonne of the substance multiplied by its ODP; the latter ranges from 0.001 to 10) tonnes in consumption and 37 000 ODP tonnes in production.

Likewise, the business plans for 2001 are to be compliance-driven, with priority given to countries most in need of funding to meet their targets. The implementing agencies involved in institutional strengthening in those Article 5 countries unable to comply with the chloro uorocarbon (CFC) freeze and the 2002 freeze with respect to halons and methyl bromide, have been requested to contact the national ozone units in those countries to assess the activities that might be needed to ensure compliance. Nevertheless, when the fund was replenished in 1999 for the period 2000- 2002, it was accepted that it was important to continue to undertake activities even in those countries that did not strictly require immediate assistance to meet their initial targets, in order to maintain momentum for the phase-out and to ensure that they would be in a strong position to complete

the nal phase-out steps to be achieved in the second half of the present decade.

STRUCTURE

The fund operates under the authority of the parties, who decide on its overall policies and the level of funding, while responsibility for overseeing the operations of the fund rests with an executive committee, supported by the fund secretariat. The projects supported by the fund are implemented by four implementing agencies – the United Nations Development Programme (UNDP), the United Nations Environment Programme (UNEP), the United Nations Industrial Development Programme (UNIDO) and the World Bank; and by a number of developed countries.

Executive Committee

The executive committee was established to develop and monitor the implementation of specific operational policies, guidelines and administrative arrangements, including the disbursement of resources, for the purpose of achieving the objectives of the multilateral fund (UNEP, 1990; UNEP, 1992c). It consists of 14 members, seven from Article 5 parties (from developing countries) and seven from non-Article 5 parties (from developed countries). The members are selected by each group of parties and formally endorsed by the meeting of the parties. The office of chair of the committee is subject to rotation, on an annual basis, between Article 5 and non-Article 5 parties, while the group not providing the chair selects the vice-chair from among its committee members.

The executive committee meets three times a year, each meeting lasting 3 days. The meetings are preceded by two-day meetings of the two standing sub-committees – the sub-committee on project review and the sub-committee on monitoring and evaluation. A small sub-group on the production sector has also been established and meets periodically.

Fund Secretariat

The role of the fund secretariat is to assist the executive committee in the discharge of its functions. The secretariat is based in Montreal, Canada, and consists of a small number of internationally recruited professional staff and local support staff. The secretariat is headed by the chief officer, who is nominated by the executive committee for appointment by the executive director of UNEP and reports directly to the committee.

The terms of reference of the secretariat (UNEP, 1991a) as approved by the executive committee at its third meeting include 23 specific responsibilities, which essentially include: development of the three-year plan and budget and a system of funds disbursement; management of the business planning cycle of the fund; monitoring the expenditures and activities of the implementing agencies; preparation of policy papers and other documents; review and assessment of project-related submissions of the implementing agencies; liaison between the committee, governments and the implementing agencies; and servicing the meetings of the executive committee. Since 1997, the secretariat has also included a monitoring and evaluation function, established expressly by the executive committee.

Implementing Agencies

The roles of the implementing agencies for projects approved under the multilateral fund were outlined broadly in decision 2/8 of the second meeting of the parties to the protocol (UNEP, 1990). More detailed arrangements are set out in the individual agreements concluded between the executive committee and UNDP, UNEP, UNIDO and the World Bank, which cover such matters as management of funds and reporting requirements.

Broadly speaking, UNDP, UNIDO and the World Bank are responsible for the preparations and implementation of investment projects, while the main thrust of UNEP's activities is on information dissemination, capacity-building, institutional-strengthening, networking, and assistance to low-volume consuming Article 5 countries.

The agencies' activities are based on their annual business plans, which contain specific targets for the year and must be approved by the executive committee. The shares of total funding to be received by each agency for investment projects are also decided upon in advance by the committee.

Projects submitted by the agencies for approval of the executive committee are reviewed by the fund secretariat for conformity with the eligibility criteria and are then forwarded to the executive committee.

Treasurer

The UNEP acts as treasurer of the fund, at the invitation of the executive committee. As such, it is responsible for receiving and administering contributions, including bilateral contributions, and disbursing funds to the fund secretariat and the implementing agencies, as agreed by the executive committee. It also reports to the executive committee on the status of the fund, including the availability of resources.

CONTRIBUTION AND REPLENISHMENT

The multilateral fund has been replenished four times for three-year cycles. The interim multilateral fund that was

Table 1 Replenishments of the multilateral fund

Replenishment period	Amount of replenishment (US$ million)
1994–1996	455
1997–1999	466
2000–2002	440

set up by the second meeting of the parties wound up with a total capitalization of US$240 million for the period 1991–1993. Since that time, the parties have undertaken technical studies in 1993 (UNEP, 1993), 1996 (UNEP, 1996c), and in 1999 (UNEP, 1999c), to support their decisions to replenish the fund three more times in the amounts indicated in Table 1.

The level of contributions required from each party is determined by the meeting of the parties in accordance with the UN scale of assessments. The budget is prepared and the assessments made in US dollars.

Bilateral cooperation by a party may be credited as a contribution to the multilateral fund up to a total of 20% of the total contribution due from that party (UNEP, 1990). Requests to offset such cooperation against assessed contributions to the fund must be approved by the executive committee. By the end of 1999, a total of US$39 million had been approved for bilateral cooperation, representing about 4% the total funds approved through the multilateral fund. The range of bilateral projects approved includes demonstration projects, training workshops, project preparation, investment projects and technical assistance.

The terms of reference of the multilateral fund (UNEP, 1990; UNEP, 1992c) state that contributions shall be in convertible currency or, in certain circumstances in kind and/or in national currency. Payment may be made either in cash or through the use of promissory notes, issued with a fixed schedule of encashment with the option of an accelerated schedule in the light of needs. In 1999, the 11th meeting of the parties (UNEP, 1999c) agreed to the use of a fixed-exchange rate mechanism on a trial basis for the 2000–2002 replenishment, in order to ease some of the administrative difficulties caused to parties by having to make commitments in foreign currencies. The implementation of the new mechanism will be reviewed at the end of 2001 to determine its impact on the operations of the multilateral fund and on the funding of efforts to phase out ozone-depleting substances in Article 5 countries.

As of November 2000, the total income to the multilateral fund, including interest and miscellaneous income stood at some US$1.2 billion, while disbursements amounted to US$1.13 billion. Arrears stood at some US$141 million for the period 1991–1999.

OPERATION

Phase One: From Start-Up to the Review of the Financial Mechanism (1991–1995)

The multilateral fund started very much from scratch. Not only was it the first funding mechanism of its kind, but the practicalities of the task ahead were ill defined: even though the use of ozone-depleting substances was virtually ubiquitous, the records of the amounts consumed and the sectoral distribution within Article 5 countries were in general rudimentary.

Much of the early work of the executive committee was devoted to operationalizing the general guidance on incremental costs and project eligibility handed down by the second and fourth meetings of the parties. The start-up period of the multilateral fund saw the establishment of the basic operational procedures for the fund, including the approval of eligibility criteria for project approval and guidelines for setting priorities. Importantly, the eligibility criteria covered projects other than capital investment projects, for example, technical assistance and clearing-house functions.

The period also saw the initiation of the process of preparing country program, which were aimed at, *inter alia*:

- Assessing the consumption of controlled substances and their sectoral distribution within countries.
- Setting out the strategy and an action plan for the implementation of the Protocol.
- Presenting a timetable for each activity and for action plan review, as well as a budget and financing program.

The country program is intended to be flexible, setting out the framework for the country's actions to meet the requirements of the Montreal Protocol. It serves as the basis for project preparation, although, initially, in the interests of expediting the start-up process, the executive committee did allow the funding of individual projects prepared prior to the completion of country programs.

Although institutional strengthening is not explicitly mentioned in the indicative list of categories of incremental costs approved by the second and fourth meetings of the parties to the Protocol, the executive committee did recognize at its fifth meeting, in November 1991 (UNEP, 1991b) that it might, in exceptional cases, be an essential element in achieving the objectives of the fund. The exceptional cases rapidly became the rule and, at its seventh meeting, in June 1992 (UNEP, 1991a), the committee decided that the main objective of institutional strengthening was to provide the necessary resources to an eligible country to enable it to strengthen a mechanism within the country for the speedy implementation of projects and for ensuring effective liaison with the executive committee, the fund secretariat and the implementing agencies.

Once a review is completed, the project is submitted to the executive committee with a recommendation from the secretariat. This stage has frequently seen significant adjustments made to project proposals, particularly in terms of costs and technology choices, to take account of previous policy decisions by the executive committee. Should the issues arising from project proposals not result in a positive recommendation by the secretariat, they are forwarded to the executive committee, together with the project proposal for individual consideration. At meetings of the executive committee, project proposals, together with any related policy issues, are initially considered in the sub-committee on project review, which makes recommendations to the committee as a whole. The sub-committee was originally composed of six members, three from Article 5 and three from non-Article 5 members of the committee. As of the year 2000, a seventh member has been added, with the incumbency alternating between Article 5 and non-Article 5 members on an annual basis. To save time within the sub-committee, the practice was adopted of submitting project proposals under two headings:

- Projects for blanket approval, which, in the view of the Secretariat, meet all the eligibility requirements and may be approved en bloc, without discussion.
- Projects for individual consideration, i.e., proposals that give rise to policy issues that require consideration and resolution by the sub-committee and the executive committee itself.

Once the project proposal has been approved by the executive committee, with whatever modifications or conditions that might be decided upon, the necessary funding may be released by the treasurer to the implementing agency concerned. Each implementing agency is responsible for the implementation and supervision of its projects and is required to report to the executive committee annually on the progress of work, and to submit a final report on completion of the project.

Throughout the first phase in the fund's development, the communication of progress in project implementation was a constant preoccupation of the committee, possibly as a result of frustration at the perceived lack of immediate tangible results of fund activities in terms of actual phase-out of controlled substances. A similar concern was re ected in repeated requests to the implementing agencies to expedite project delivery. At its eighth meeting (UNEP, 1992b), the committee decided that implementing agencies should submit progress reports on project implementation and identify and describe any problem areas. At the same time, the secretariat should assess project implementation status, ascertain problem areas, and offer assistance in their resolution between committee meetings. A uniform format for reporting was adopted in 1995, re ecting the need for closer monitoring by the executive committee and greater uniformity and clarity in reporting by the agencies. By the end of the period, the executive committee had committed itself to a more formal and exhaustive monitoring and evaluation system, as part of the more structured approach toward planning and follow-up that was put in place after 1995 (UNEP, 1997).

Indeed, although much commendable work was completed during the first four years of the fund's existence in terms of the development of policies, guidelines and the processes of project preparation and approval, the whole period was marked by a lack of true strategic planning. Project proposals were submitted and approved on a first come, first served basis, and were submitted from countries without a country program. There was no emphasis on the sectoral distribution of projects or the mix of countries for which projects were prepared. Nor were there any established cost-effectiveness criteria for project eligibility. In addition, most industrial plant conversion was based on total replacement of equipment, and, importantly, the projects were not explicitly linked to the achievement of the compliance targets by Article 5 parties.

Phase Two: From the Review of the Financial Mechanism to the First Control Measure (1995–1999)

The year 1995 was a pivotal one in the evolution of the multilateral fund. It saw four major developments:

- the introduction of a rigorous system of business planning;
- the establishment of cost-effectiveness thresholds for project funding;
- the completion of the review of the effectiveness of the financial mechanism, called for in decision 4/18 of the fourth meeting of the parties (UNEP, 1992c), and the review of the situation of the parties operating under paragraph 1 of Article 5, including the effective implementation of financial cooperation and transfer of technology to them, required by Article 5, paragraph 8, of the amended Montreal Protocol (UNEP, 1990);
- The establishment of the basis for a monitoring and evaluation system for the Fund.

With the second replenishment of the multilateral fund, for the period 1994–1996, there was a significant increase in the number of project proposals being submitted to the executive committee. By late 1994, it became clear that there was a need to address the situation where the total value of proposals would exceed the amount available to the fund. This was initially done through the addition of a procedure known as *cleared for approval* pending availability of funds, which meant that the funding necessary to finance approved projects would be released by the treasurer as and when it became available after the executive committee meeting at which the project was cleared. This departure

from the previous practice of approving projects only up to the limit of resources available at the meeting concerned could not however solve the emerging structural problem of an excess of project proposals over resources allocated to the Fund. At its 16th meeting, in March 1995 (UNEP, 1995a), the executive committee approved a framework for resource allocation, under which resources would be divided into distinct slices for various categories of projects. It also set cost-effectiveness thresholds for the various sectors and sub-sectors, with the proviso that projects from low-volume consuming countries would not be subject to these thresholds.

The experience of the period 1994–1995 highlighted the need for stricter business planning that would link proposed activities to the funding available and to the sectoral priorities. The first such plans were submitted by the agencies in 1996. They were to include all projected activities for the year in question, as well as performance indicators, thus providing a basis for monitoring and evaluation as well as strategic planning. In order to avoid project delays resulting from the emergence of policy issues at the project review stage, agencies were also encouraged to include any policy issues likely to arise so that they could be addressed in a proactive manner by the executive committee. A consolidated business plan would then be prepared by the secretariat of the multilateral fund, together with targets for controlled substances phase-out for each agency. The consolidated plan also sets out the resources to be allocated to each agency in the course of the year. Non-Article 5 Parties also submit business plans for bilateral cooperation activities.

The year 1995 also saw the completion and review of the meeting of the parties of two major studies of relevance to the operations of the fund. The first, on the operation of the financial mechanism, had been called for at the fourth meeting of the parties (UNEP, 1992c), when the fund was established on a permanent basis. As a result of the study, the seventh meeting of the parties (UNEP, 1995b) approved a series of 21 actions to be taken to improve the mechanism, including the development of a systematic approach to policy development, the establishment of a monitoring and evaluation system, the improvement of policy dissemination procedures, and the establishment of firm compensation policies for phase-out in the production sector. The implementation of these actions underlay much of the work of the executive committee in the years to come.

The second study had been provided for in Article 5, paragraph 8, of the Montreal Protocol as amended at London in 1990, which stipulated that a meeting of the parties;

> shall review, not later than 1995, the situation of [Article 5 Parties], including the effective implementation of financial cooperation and transfer of technology to them, and adopt such revisions that may be deemed necessary regarding the schedule of control measures applicable to those parties.
>
> (UNEP, 1990)

At Copenhagen, in 1992, the parties had agreed that any further amendments to the adjustment and amendment of the protocol adopted at London would be applicable to Article 5 parties only after this review had taken place and would be based upon its conclusions. The study itself, as presented to the seventh meeting of the parties, presented a number of possible phase-out scenarios and the related costs to the multilateral fund. As a result of the review, the parties decided (UNEP, 1995b) to retain the London schedule for Article 5 Party phase-out of annex A and B substances and introduced control measures for hydrochloro uorocarbons (HCFCs) and methyl bromide for those countries.

Although a presumption against the use of HCFCs as substitutes and a requirement to fully justify such use had been in place since the 15th meeting of the executive committee in 1994, the decision to make them controlled substances for Article 5 opened the way for a protracted and often acrimonious debate within the committee. Immediately after the meeting of the parties (1995) (UNEP, 1995b), the executive committee decided to request more complete justification for the use of HCFCs as transitional substances and to make it clear that the industrial enterprise concerned would have to bear the full costs of subsequent conversion to non-HCFC substitutes. Subsequent decisions called for even more stringent requirements for the submission and review of information to support the choice of HCFCs, including a letter from the government concerned in which it would verify that it had reviewed the specific situations involved in the project, as well as its relevant commitments under the protocol and that it would state that it had determined that HCFCs were required at the present time and that it understood that no funding would be available for future conversions from HCFCs for the company or companies concerned. The whole question of policy on HCFC use as an interim technology is still being considered by the executive committee.

The inclusion of methyl bromide (used in the agricultural sector for soil fumigations) as a controlled substance for Article 5 countries led to an acceleration of work on the subject in the executive committee, which had already approved three regional methyl-bromide data-collection and demonstration projects in 1995. Methyl-bromide projects are perhaps the most challenging area in which the fund has operated, as there is no one-size-fits-all alternative for every application or crop. In addition, while industrial conversion through the replacement and destruction of existing equipment is a one-off event, there are few technical impediments to a farm reverting to methyl bromide in the years after the completion of a phase-out project. Interim guidelines on methyl-bromide demonstration projects were adopted at the 20th meeting of the committee, in October 1996 (UNEP,

1996), and $10 million for such projects were allocated for the period 1997–1999. A strategy and guidelines for the whole sector, including a priority list of crops for soil fumigation, was approved in 1998, and the first investment projects aimed at phase-out began with the approval of a project in the tobacco sector in Cuba in November 1998. As of July 2000, a total of $35 million had been allocated to demonstration and investment projects on methyl bromide and the total amount of methyl bromide to be phased out stood at 937 ODP tonnes.

As already mentioned, the follow-up of project implementation was a general cause for concern in the first few years of the fund's operation. The period 1995–1999 saw the establishment and refining of a monitoring and evaluation system, as called for in the study on the effectiveness of the financial mechanism. In 1997, a monitoring and evaluation sub-committee was established and a monitoring and evaluation post created within the secretariat. The sub-committee, like its counterpart for project review, meets immediately prior to meetings of the executive committee and also has seven members, drawn from the membership of the committee. An evaluation guide was also prepared, intended to be a dynamic document subject to revision in the light of experience. Henceforth, all project proposals were to include a standardized component on monitoring and evaluation, including milestones for the completion of the various stages of the project that should be able to provide sufficient information to enable a comparison between the approved proposal and the project completion report. The purpose of the process is to ascertain whether the project objectives have been met in terms of the disbursement of funds, the amount of controlled substances phased out, and various other performance indicators. Agencies are to submit progress reports by May each year for all activities approved by the end of the previous year.

The period 1995–1999 saw a greater consciousness of the need to anchor fund activities more tightly to the commitments of parties under the protocol. At its 19th meeting, in May 1996 (UNEP, 1996a), the executive committee decided that all project proposals should contain a brief description of how the project would help the country achieve the 1999 freeze, and this decision was reiterated in one form or another at a number of subsequent meetings.

The period also saw attention turn to the question of production sector phase-out. Although only a handful of Article 5 countries produce ozone-depleting substances, the question of funding phase-out in the production sector has been one of the most complex faced by the executive committee. Work on establishing guidelines began in 1994, and was largely carried out by two small bodies established by the committee – the expert group on the production of ozone-depleting substances and the executive committee's subgroup on the production sector. Agreement was reached that the basis for compensation should be loss of future profits rather than the incremental costs of constructing new facilities. The main difficulty lay in devising a methodology for calculating these losses, as the total arrived at could vary widely depending on the baselines used and the assumption made for elements such as market trends. Donors, in particular, were reluctant to agree to any methodology without having an idea of the possible financial implications once data from actual factories were factored in.

Very little progress was made in 1996 and 1997, pending the submission of data on production facilities from Article 5 countries. In 1998, the executive committee agreed that technical audits of production facilities should be conducted to verify the data submitted and to calculate a possible reasonable level of compensation on the basis of a number of criteria. The following year, the first production phase-out project was agreed after lengthy negotiations with the government of China. A total sum of $150 million was allocated for the closure of the entire CFC production capacity in the country over a period of 12 years (UNEP, 1999a). An annual payment and phase-out schedule was included in the agreement reached with the country, with the release of each installation being made contingent upon the completion and independent verification of the reductions indicated in the schedule. Later in the same year, a similar agreement (UNEP, 1999b) was reached for the production sector in India, and a draft agreement with Argentina is currently under preparation.

Phase Three: Meeting the Control Measures (1999–2015)

As noted above, the first control measure for Article 5 countries – the freeze on consumption of CFCs at the average of the 1995–1997 level – came into effect on 1st July 1999. The focus of work of the executive committee since that time has been on the interim measures between freeze and final phase-out, which include a 50% reduction in CFC and halons consumption by 2005, and 20 and 85% reductions in the consumption of methyl bromide and carbon tetrachloride, respectively, by the same year. An analysis of controlled substances phase-out in Article 5 countries produced by the secretariat in 2000 indicated that the majority of countries were in a position to comply with the freeze control measures and that several already had sufficient phase-out activities approved to enable them to meet the 2005 target.

This period has seen a growing number of investment projects for the phase-out of methyl bromide, particularly in light of the freeze scheduled for 2002. It has also seen the commencement on work in new areas, e.g., on projects to limit controlled substances emissions from process agent applications, which became eligible for funding following a decision adopted by the 10th Meeting of the Parties (UNEP, 1998), at the end of 1998, and on the new subsector covering the assembly, installation and servicing of

refrigeration equipment. The current phase also includes the committee's decision on the refrigerants management plan for low-volume consuming countries (UNEP, 2000).

Strategic planning is becoming increasingly geared toward making the Article 5 countries the main motors for the fulfillment of their compliance obligations. In that respect, the executive committee has instructed all implementing agencies to ensure that project proposals are based on the current strategic planning of the Article 5 country concerned to ensure that the national ozone unit is fully involved in the planning and preparation of projects. The units are regularly provided with information on the progress of project implementation to assist them in improving their capacity to monitor and evaluate projects implemented including their impact at the country level.

Analysis of the most recent data reported by 101 developing countries indicates that 94 of them have been able, or will soon be able, to achieve the mandatory freeze in their consumption and/or production of chloro uorocarbons (CFCs). In addition, several of these countries are well on their way to achieving the mandatory 50% reduction in CFC consumption ahead of the protocol schedule of 2005.

ACHIEVEMENTS

While the most obvious achievement of the multilateral fund is the accelerated phase-out of ozone-depleting substances in developing countries, a less prominent – but in the long term possibly equally important – achievement lies in the successful transfer of technology to developing countries. Through the investment projects, all the new technology for ozone-friendly substitutes has been successfully transferred. The multilateral fund has paid for the costs of patents and technology licensing fees and the design, testing and demonstration of new technologies. This has allowed some developing countries to leap over interim options. For example, there are now more plants using liquid carbon dioxide (CO_2) technology in developing countries in the production of exible foams than in the US and Europe combined. Equally, the fund has been responsible for the increased penetration of hydrocarbon technologies, including adoption by small consumers in developing countries.

The institutional arrangements for project screening have allowed the fund to operate in a highly cost-effective way: many of the projects completed have shown savings over the approved amounts, which themselves were in many cases significantly lower than the original cost estimates submitted by the implementing agencies.

Capacity-building and institutional strengthening are two other areas in which marked progress has been made as a result of the fund. More than 105 countries have institutional strengthening projects financed from the fund. These projects generally involve the establishment and maintenance of a national ozone unit within the country. The fund provides resources for various networking activities, including regional workshops. As a result of the capacity-building activities undertaken with fund support, most medium-size and large countries now implement their own projects using national experts and drawing on the clearing-house facilities maintained by UNEP.

Demonstration projects for the introduction of integrated pest management techniques as an alternative to the use of methyl bromide may also be applicable for the replacement of other agricultural chemicals with more environmentally friendly pest-control technologies (*see* Integrated Pest Management (IPM) in an Era of Global Environmental Change, Volume 4).

Projects for the development of refrigerant management plans all contain such elements as the training of technicians and of customs officers, implementation of recovery and recycling projects and additional institutional support for the development of legislation. Again, this general enhancement of the capacities of the developing countries – in terms of both human resources and institutions – will surely have an impact that goes beyond the immediate objectives of the Fund.

THE ROAD AHEAD

The ultimate test of the effectiveness of the multilateral fund will be whether all Article 5 countries meet their commitments under the Montreal Protocol without disruption in the various sectors that have hitherto been dependent on a supply of the controlled substances. In the near term, up to 2005, there are grounds for optimism that the compliance targets for Article 5 countries will be met, at least by the majority of countries. There are, however, challenges remaining, particularly in achieving phase-out in the production sector in all Article 5 countries, ensuring the successful implementation of refrigerant management plans, handling the residual requirements for servicing existing equipment, combating illegal trade in controlled substances, and dumping of CFC-based appliances in developing countries. The central issue will be to ensure the sustainability of the achievements to date and to prevent backsliding and reversion to the use of controlled substances. On the latter point, there is some evidence that the prevailing lower price of CFCs is encouraging their use in servicing equipment initially designed for use with ozone-friendly substances. This situation may, however, only be a transitional phenomenon that will change as prices of controlled substances increase and that of substitutes decrease. In any event, much will depend not only on the collective resolve of the parties to the Protocol to maintain the momentum toward phase-out but also on the level of vigilance, commitment and capacity at the national level (For other entries on the Montreal Protocol *see* Ozone Layer: Vienna Convention and the Montreal Protocol, Volume 4).

ACKNOWLEDGMENTS

I wish to express my gratitude to Mr John Fry, Editor in the Secretariat of the Convention of Biological Diversity, for his invaluable assistance.

REFERENCES

UNEP (1987) *Montreal Protocol on Substances that deplete the Ozone Layer, Final Act*, UNEP.

UNEP (1990) *Report of the Second Meeting of the Parties to the Montreal Protocol*, (UNEP/OzL.Pro.2/3), 29th June.

UNEP (1991a) *Report of the Third Executive Committee of the Multilateral Fund* (UNEP/OzL.Pro/ExCom/3/18/Rev.1), 17th June.

UNEP (1991b) *Report of the Fifth Meeting of the Executive Committee of the Multilateral Fund* (UNEP/OzL.Pro/ExCom/5/16), 22nd November.

UNEP (1992a) *Report of the Seventh Meeting of the Executive Committee of the Multilateral Fund* (UNEP/OzL.Pro/ExCom/7/30), 27th June.

UNEP (1992b) *Report of the Eighth Meeting of the Executive Committee of the Multilateral Fund* (UNEP/OzL.Pro/ExCom/8/29), 21st October.

UNEP (1992c) *Report of the Fourth Meeting of the Parties to the Montreal Protocol* (UNEP/OzL.Pro.4/15), 25th November.

UNEP (1993) *Report of the Fifth Meeting of the Parties to the Montreal Protocol* (UNEP/OzL.Pro.5/12), 19th November.

UNEP (1995a) *Report of the 16th Meeting of the Executive Committee of the Multilateral Fund* (UNEP/OzL.Pro/ExCom/16/20), 17th March.

UNEP (1995b) *Report of the Seventh Meeting of the Parties to the Montreal Protocol* (UNEP/OzL.Pro.7/12), 27th December.

UNEP (1996a) *Report of the 19th Meeting of the Executive Committee of the Multilateral Fund* (UNEP/OzL.Pro/ExCom/19/64), 10th May.

UNEP (1996b) *Report of the 20th Meeting of the Executive Committee of the Multilateral Fund* (UNEP/OzL.Pro/ExCom/20/72), 18 October.

UNEP (1996c) *Report of the Eighth Meeting of the Parties to the Montreal Protocol* (UNEP/OzL.Pro.8/12), 19th December.

UNEP (1997) *Report of the 22nd Meeting of the Executive Committee of the Multilateral Fund* (UNEP/OzL.Pro/ExCom/22/79/Rev.1), 30th June.

UNEP (1998) *Report of the 10th meeting of the Parties to the Montreal Protocol* (UNEP/OzL.Pro.10/9), 3rd December.

UNEP (1999a) *Report of the 27th Meeting of the Executive Committee of the Multilateral Fund* (UNEP/OzL.Pro/ExCom/27/48), 12th April.

UNEP (1999b) *Report of the 29th Meeting of the Executive Committee of the Multilateral Fund* (UNEP/OzL.Pro/ExCom/29/65), 26th November.

UNEP (1999c) *Report of the 11th Meeting of the Parties to the Montreal Protocol* (UNEP/OzL.Pro.11/10), 17th December.

UNEP (2000) *Report of the 31st Meeting of the Executive Committee of the Multilateral Fund* (UNEP/OzL.Pro/ExCom/31/61), 7th July.

Multi-issue Assessments

R E Munn
University of Toronto, Toronto, Canada

Methods for undertaking single-issue environmental assessments are well established. But sometimes a suite of interrelated issues must be considered. A policy designed to resolve one issue may create or worsen another. In order of increasing complexity, multi-issue assessments may relate to:

1. *Potential impacts of a single environmental stressor (and appropriate policy responses) with respect to a variety of responders, e.g., the impacts of acid deposition on forests, agricultural crops, human health, lakes and groundwater.*
2. *Potential impacts of several environmental stressors, e.g., acid deposition, smog episodes and climate change, on a single type of responder, e.g., forests or lakes.*
3. *Potential impacts of several environmental stressors, e.g., acid deposition, smog episodes and climate change, on a variety of responders, e.g., on both agricultural crop yields and human health.*
4. *Potential impacts of several stressors, some of which are not environmental, on one or a variety of types of responders, e.g., the effects of climate change and human population growth on acid deposition and renewable resource depletion.*

The methodological uncertainties involved in dealing with multi-issue assessments are enormous, and it cannot be claimed that a handbook of solutions to the problems raised is available. Some stressors operate intermittently over decades or longer, and the responses are cumulative. During such long time periods, socioeconomic, cultural and technological conditions may change greatly, sometimes abruptly. These are real-world situations, which are very dif cult to model. Furthermore, the costs involved in undertaking additional scienti c research to resolve uncertainties, in establishing early warning monitoring systems, and in changing the human lifestyle patterns responsible for causing the issues, will vary greatly from issue to issue. It will be dif cult to develop common economic yardsticks that can be used across a range of issues.

SOME EXAMPLES OF MULTI-ISSUES

Taking action to reduce a trend towards some undesired environmental state, or to eliminate the long-term risk of such a state, may have either positive or negative effects

on other environmental components, and on socioeconomic conditions.

Positive bene cial effects: of course, positive effects of a particular policy on other biospheric or socioeconomic compartments would be desirable. For example, planting trees reduces soil erosion, increases the biospheric sink strength of carbon dioxide (CO_2), and provides jobs for students and the unemployed. Policies that foster these benefits are called no-regrets policies (*see* No Regrets Principle, Volume 4). Generally speaking, environmental assessments should try to identify the opportunities that such policies might provide.

Negative and potentially harmful effects: on the other hand, action taken to reduce some environmental threats may cause potentially harmful effects on other environmental components, or on the socioeconomic well-being of society. Here are some examples that fall into this category:

1. Replacing chloro uorocarbons in the manufacturing sector with other gases may reduce stratospheric ozone depletion but the replacement gases may contribute to climate warming.
2. Automobile catalytic converters reduce emissions of tropospheric ozone precursors (NO_x and volatile organic compound (VOC)) but release nitrous oxide (N_2O), a greenhouse gas.
3. Sulphate aerosols cause acidic deposition while at the same time absorbing and scattering solar radiation, thus cooling the surface of the earth. Reducing sulfate emissions will therefore partially compensate for greenhouse gas warming.
4. Electric cars reduce emissions at the point of use, but they operate on lead batteries which must be recharged, increasing emissions from coal-fired and oil-fired power stations; also, increased emissions from lead smelters and from re-processing of lead batteries will cause local pollution. Lave *et al.* (1995) estimate that an electric car will cause the indirect release of 60 times more lead per km^{-1} of use than a car using leaded gasoline.
5. The cumulative deposition of sulfates and nitrates into lakes, the increasing penetration of UV-B radiation into lakes, and the potential drop in lake levels due to a drier climate all contribute to a decline in the health of freshwater ecosystems (Schindler *et al.*, 1996).

GUIDELINES ON HOW TO FORMULATE A MULTI-ISSUE ASSESSMENT

The following guidelines are not prescriptive but may be helpful to practitioners of multi-issue assessments:

General Guidelines

- The ecosystem approach (EA) should be used as the arching framework for multi-issue assessments (*see* Ecosystem Approach, Volume 2; Monitoring in Support of Policy: an Adaptive Ecosystem Approach, Volume 4).
- A common policy framework should be used, i.e., the process of policy formulation should be the same for each issue.
- Winners and losers should be identified with respect to each issue.

Global Multi-issues

Some issues may be well recognized and of immediate concern. Other issues may involve great uncertainty and may be long-term. How should a balance be struck between current threats and long-term sustainability of the human race? Only some general suggestions can be given, as follows:

- The assessment should cover very long time periods – several decades at least, even though the uncertainty is great. One way of designing such an assessment is to develop several internally consistent socioeconomic scenarios, then testing the effects of each on the environmental issues being studied, the objective being to select adaptive robust policies that minimize future environmental risks. Every effort should be made not to foreclose future options.
- The concept of biogeochemical cycling provides a good framework for assessing global multi-issues. Most of the major issues are coupled through these cycles, e.g., climate change, stratospheric ozone depletion, acidic depletion, and nitrogen over-fertilization of the biosphere.
- Relevant results coming from the emerging field of ecological economics should be applied wherever possible (*see* Ecological Economics, Volume 5).
- The precautionary principle should be adopted (*see* Precautionary Principle, Volume 4).
- The possibility of feedbacks, overshoots and ip- ops should be assessed.
- People's perceptions of the differing risks associated with different environmental issues will require identification; a common understanding amongst all stakeholders, of the risks involved will need to be achieved. In some cases, this process will lead to disagreement rather than agreement amongst the parties concerned. However, information of this nature will be useful in the subsequent negotiating process, for which the multi-issue assessment will be a valuable decision-support tool.

Regional Multi-issues

In some cases, a region (e.g., Los Angeles County) is the main cause of its environmental problems. Sometimes, however, regional issues may be amplified by global change, e.g., by the potential damage that would be caused by sea level rise. Nevertheless, regional issues are more likely to be more immediate, and to involve large concentrations of people and industry than are global ones. In this context, the following suggestions are made:

- Regional multi-issues should be assessed within the framework of global change.
- An appropriate spatial boundary for a regional multi-issue assessment is often a watershed or river basin. This may sometimes extend the proposed outer boundary of a conventional environmental assessment, but it may often provide long-term insights into the nature of interactions amongst two or more environmental issues.

CONCLUSION

As a general rule, practitioners ought to look farther afield and further ahead than they normally do whenever they prepare environmental assessments. This is particularly true with regional and global assessments.

As a final example of this broader outlook, Chamiedes *et al.* (1994) have used a simulation model to project that by the year 2025, due to the increasing numbers of motor vehicles, there will be more frequent and more intense surface oxidant episodes in mid-latitudes, placing 10–35% of the world's grain production at risk. The simulation model assumes that all other environmental factors will remain unchanged, but global projections of climate warming suggest that drought may become more prevalent and that the frequency of stagnant anticyclonic conditions may increase. The risk factor for decreased crop yields is therefore greater than that obtained from an analysis of the effect of increasing automobile emissions alone.

REFERENCES

Chamiedes, W L, Kasibhatla, P S, Yienger, J, and Levy, II, H (1994) Growth of Continental-scale Metro-agro-plexes, Regional Ozone Pollution, and World Food Production, *Science*, 264, 74–77.

Lave, L B, Hendrickson, C T, and McMichael, F C (1995) Environmental Impacts of Electric Cars, *Science*, 268, 993–995.

Schindler, D W, Jefferson, P J, Parker, B R, and Stainton, M P (1996) Consequences of Climate Warming and Lake Acidification for UV-B Penetration in North American Boreal Lakes, *Nature*, 379, 705–708.

FURTHER READING

Munn, R E, ed (1995) *Atmospheric Change in Canada: Assessing the Whole as Well as the Parts*, Summary Report, Institute for Environmental Studies, University of Toronto, 1–36.

Munn, R E, ed (1997) Proc. Workshop on Atmospheric Change in Canada: Assessing the Whole as Well as the Parts, *Env. Monit. Assess.*, 46, 1–190.

Munn, R E and Marouf, A (1997) Atmospheric issues in Canada, *Sci. Total Environ.*, 203, 1–11.

N

National and Local Responses to Desertification

Harold E Dregne
Texas Tech University, Lubbock, TX, USA

Deserti cation is the widely used term for land degradation in the world's drylands. Those drylands represent about 40% of the earth's land surface, lying in what climatologists refer to as the arid, semi-arid, and dry subhumid climatic zones. Deserti cation was the subject of the 1977 United Nations Conference on Deserti cation. That conference focused attention on the effect that land degradation in the drylands has on land condition and human welfare.

There are six deserti cation processes of major concern worldwide. They are vegetation degradation of grazing lands, water erosion, wind erosion, salinization, soil compaction, and soil fertility decline. All of these processes are human-induced and are, accordingly, preventable. Drought frequently is associated with deserti cation but it is only a contributing factor, not a cause. Drought is an inevitable climatic condition, which is not caused by humans and cannot yet be predicted. Its impact on livestock and humans can, however, be eased by good land management before the drought strikes and by preparing for drought-induced water shortages.

Most countries, both developed and developing, have government agencies charged with conserving land resources, such as grazing and croplands. Additionally, there are numerous non-governmental organizations (NGOs) that have for their objective the conservation of natural resources and the improvement of human welfare. 'Landcare', for example, is an in uential Australian NGO which is a grassroots coalition of urban and rural people. Its purpose is to generate political, public, and private support for land conservation throughout Australia. Its efforts complement and strengthen the activities of government agencies.

Desertification is an old problem, despite the greatly increased attention focused upon it since the United Nations Conference on Desertification in 1977. Historical evidence demonstrates that serious and extensive land degradation occurred many centuries ago in the drylands, principally in three regions: around the Mediterranean Sea, in the Mesopotamian Plain of Iraq, and on the Loess Plateau of central China.

Water erosion was widespread in the eastern Mediterranean hills and mountains more than 2000 years ago. Deforestation provided wood for fuel, timber for ships and temples, and space for cropland. Cutting the stately cedars of Lebanon is the classical example of how tree removal leads to land degradation. With the loss of the protective cover of trees and leaf litter, soil erosion was greatly accelerated and landscapes altered, giving the false impression that the climate had become more arid. As long ago as 2600 BC Egyptians were known to have imported timber from Phoenicia for the resins used in mummification. Today, only a few sacred groves remain in Lebanon to show how the magnificent cedars looked. Even though the climate is the same as it was when the cedars ourished, the possibility of reforestation now is slight. Bare rocks, continued reliance of people on wood for fuel, and wandering goats free to eat every green plant that manages to germinate are formidable obstacles to the survival of trees.

In a pattern that has been repeated across the world, destruction of forests was followed by expansion of cultivated land ever higher on the slopes. Early settlers learned to hold the soil in place by constructing terraces. In later years, as wars, civil strife and pestilence devastated the countryside, maintenance of terraces and other soil-saving structures often was neglected. The inevitable result was accelerated water erosion, leading to the barren landscapes that now mark many of the hills of the eastern Mediterranean countries. Occasionally, well-maintained terraces demonstrate that the land is capable of sustained use for thousands of years if properly cared for.

The most significant instance of desertification resulting from soil salinization in olden times is found in the lower valleys of the Tigris and Euphrates rivers. Present-day Iraq is still trying to recover from a salt build up that began more than 4000 years ago in the plains of ancient Mesopotamia. Salinity problems in irrigated lands became serious for the first time in about 2400 BC in southern Iraq. Additionally,

irrigation systems were plagued with siltation of canals. The heavy load of sediment coming from eroded uplands made it mandatory that canals be cleaned frequently. That burden and the increasing salt problem eventually led to the abandonment of irrigation over much of Mesopotamia in the 13th century AD when Mongol armies conquered the area.

Wind-deposited soils of the Loess Plateau in China are naturally highly erodible. That susceptibility to water erosion is responsible for the occurrence in the Loess Plateau of probably the most severe land degradation anywhere in the world. The degradation began at least 3000 years ago when forests were turned into cropland as the population grew. The loss of productive cropland to erosion was bad enough but the eroded soil, when it enters the Yellow River, causes serious problems of sedimentation in reservoirs and raises the bed of the river above the surrounding plains. The latter problem threatens the lives of untold numbers of Chinese in the North China Plain if a major flood were to break the dikes that now confine the river to its main channel.

Deterioration of rangelands due to overgrazing and wood cutting seems to have become significant only in the last few centuries. Large numbers of cattle, sheep, and goats are responsible for most of the vegetation degradation. Now, few rangelands anywhere in the drylands are in excellent or good condition. The exceptions are the rare places where the native vegetation is unpalatable to livestock, as in northwestern Australia.

DISTRIBUTION OF DRYLANDS

Drylands are defined by the United Nations as lands in the arid, semi-arid and dry subhumid climatic zones. Extremely dry (hyperarid) regions are excluded from the dryland definition because they are too dry for any agricultural activity, including grazing. The largest climatic desert in the world is the Sahara in Africa. Crop production in the Sahara is possible only in irrigated oases such as the Nile Valley. Within the drylands, there are three dominant land uses: grazing, rainfed cropland, and irrigated agriculture. About 88% of the drylands are used for grazing (rangelands), 9% for rainfed cropping, and 3% for irrigated agriculture.

The United Nations Educational, Scientific, and Cultural Organization (UNESCO) published a map in 1977 showing the world distribution of arid lands. That map is based on the ratio of precipitation to potential evapotranspiration ($P \bullet ET_p$) and is a measure of the degree of water deficiency. A $P \bullet ET_p$ of less than 0.03 means that the climate is extremely dry. A ratio of 0.50–0.75 indicates that the climate is dry subhumid. Arid and semi-arid zones have intermediate $P \bullet ET_p$ ratios.

Among the continents other than Antarctica, Australia has the highest percentage of drylands (84%) and Europe the least (3%). Asia has the largest dryland area (20 100 000 km^2), followed by Africa with 14 000 000 km^2. Europe has 150 000 km^2 of drylands, nearly all around the Mediterranean Sea.

GLOBAL SIGNIFICANCE OF DESERTIFICATION

There is limited reliable information available on the global extent and severity of land degradation in the drylands. Many countries have experimental data on the magnitude of water erosion, salinization of irrigated lands, and soil fertility decline. Fewer have data on rangeland deterioration, wind erosion, and soil compaction. A 1992 Global Assessment of Soil Degradation, funded by the United Nations Environment Programme, produced the first worldwide map of soil degradation in both humid regions and the drylands. No similar map showing the distribution of vegetation degradation in the drylands has yet been published.

Dominant land degradation processes differ among the three major land uses in the drylands. In rangelands, the principal desertification process is vegetation degradation, followed by water and wind erosion. Salinization, usually accompanied by waterlogging of soils, is by far the most economically damaging process in irrigated lands but soil compaction is common also (*see* Salinity and Agriculture, Volume 3; Waterlogging, Volume 3). Water erosion, wind erosion, soil fertility decline and soil compaction constitute the major degradation processes in rainfed croplands. In a few countries, mainly Australia, Canada and the United States, salinization affects local non-irrigated land in certain semi-arid climatic zones. Soil compaction is a widespread and important problem in cultivated lands, especially where field operations are done with heavy machinery. It attracts little attention, usually, because the compaction occurs out of sight in the subsoil. Fortunately, its effects are easily reversed, once it is recognized.

In 1991, the United Nations Environment Programme commissioned a country by country analysis of global desertification. The study was a part of the background material prepared for the United Nations Conference on Environment and Development (UNCED) held in Brazil in 1992. Table 1 summarizes the estimated extent of land degradation in the drylands at that time, by land use.

Since the database upon which the desertification estimates were made was weak, the numbers in the table are only general approximations. The desertified rangeland

Table 1 Global desertification in drylands (1992)

Land use	Area (ha)	Desertified (%)
Irrigated land	145 500 000	30
Rainfed cropland	457 700 000	40
Rangeland	5 160 000 000	70

figure, particularly, could be much too low. It could well be 80% or more, given the increases in livestock populations in the past 50–100 years.

REVERSIBILITY OF DESERTIFICATION

Of the six principal processes of desertification, all except water and wind erosion are reversible using well-known practices. However, the reversibility may not be economic to carry out. The best example of that limitation is reversibility of rangeland deterioration. Restoration of the original vegetative conditions in overgrazed pastures can frequently be done if the rainfall is relatively high. This means that reversing vegetation degradation in dry subhumid climates probably will require 10–20 years, depending upon how badly degraded the land is. In the arid climatic zone, however, restoration may well take 40 or more years. If vegetation degradation has caused severe water or wind erosion, restoration may not be possible at all unless the landscape is drastically altered, at a high cost.

Soil fertility decline and soil compaction are relatively easy to reverse. Salinization usually is costly to reverse but the benefits of doing so usually are also high. Soil loss due to water and wind erosion is not reversible in a human lifetime because soil formation is slow. Erosion damages to soil productivity are a function of soil depth: greater damage on shallow soils, lesser damage on deep soils.

COST OF DESERTIFICATION DAMAGE

Land degradation causes two kinds of damage, on-site and off-site. On-site damage is the impact the degradation process has directly on the place where it happens. Off-site damage is that which occurs elsewhere. For example, on-site water erosion damage refers to soil and crop damage in the field where erosion occurred. Off-site damage refers to the harm that water erosion does elsewhere, such as sediment accumulation in reservoirs, deterioration in river water quality, increases in flooding, and other problems.

Water and wind erosion are the degradation processes that cause most off-site damage. In fact, off-site wind erosion damage is much worse than on-site damage, principally due to air pollution, reduction in visibility, and abrasion of structures and machinery by air-borne soil particles. Asthma problems in the Canary Islands have been attributed, in part, to dust carried on winds from the Sahara. The cost of on-site and off-site water erosion damage is believed to be about equal. In the Loess Plateau of China, however, off-site water erosion damage costs are much higher than on-site costs.

The most comprehensive study of on-site soil degradation costs was conducted in eastern Australia by the state of New South Wales. That survey showed that deterioration of soil structure, principally increased soil compaction, was by far the most important degradation process in terms of loss of income from lower crop yields. Salinization of irrigated land was next in importance. The annual cost of income foregone due to all forms of soil degradation was US$160 000 000 per year. A more generalized analysis in Canada concluded that the annual on-site cost of soil degradation in the entire country was about US$900 000 000.

CAUSES OF DESERTIFICATION

The direct cause of land degradation is improper management of land resources. The indirect cause is far more complex. High on the latter list is poverty and land tenure problems. Contributing factors include lack of incentives to conserve land, absence of credit arrangements for purchasing fertilizers and pesticides, extension of rainfed cropping into climatically marginal lands where droughts become more frequent and more severe, weak research and training programs and poor transportation facilities.

While droughts do not cause desertification, they do magnify the effects of previous overuse of the land. Rangeland vegetation weakened by overgrazing when the weather is good becomes especially susceptible to further degradation when drought strikes. With the reduced vegetative cover, water and wind erosion is more likely to occur. Rainfed croplands also become more susceptible to erosion during droughts because crop cover is reduced. Drought effects become much more pronounced when droughts continue for two or more consecutive years.

COMBATTING DESERTIFICATION

Preventing desertification is more effective and cheaper than reclamation of degraded land. Unfortunately, it is easier to get support for coping with a crisis than for avoiding future crises. Land degradation in the drylands now is so widespread that stopping further degradation is a major challenge to governments, non-governmental agencies, and land operators. Combating desertification cannot be done successfully as an isolated effort directed solely toward that objective. Anti-desertification programs must be part of a broader development policy to bring about improvement of rural conditions of health, education, communication, the economy and human welfare. A technical program of land conservation has little chance of success if it attacks only the symptoms of the problem.

Basic principles of good land management are well known. Their use does, however, require knowledge of the local environment and the adaptations that must be made for the practices to be introduced successfully. Windbreaks, for example, have been shown to reduce wind erosion worldwide. The tree and shrub species to be planted,

however, must be adapted to the local soils and climate. There is no one universal solution that fits all environmental conditions.

Adoption of improved technology appears to be essential if desertification is to be controlled, given the inability of the technology in current use to prevent continued land degradation. The technology needed is what is appropriate for the educational and financial level of the land users. By and large, the technology that is the most simple, cheap, easy to use, and effective is the most appropriate technology. It may be high level or low level technology, depending upon what is needed to fit the circumstances.

Optimally, the use of advanced systems of technology in agriculture would increase crop production on the better lands. That could allow the poorer lands to be returned to less intensive grazing, forestry, wildlife, and recreational uses, as has been done in the humid regions of Europe. The current need to eke out as much as possible from every piece of available land, no matter what its condition, is a severe restraint on land improvement efforts.

Reliance on advanced technology to solve desertification problems may be essential but it has drawbacks. There are certain to be negative as well as positive effects. Constant vigilance is necessary to keep the negative effects under control. Fertilizer use is a prime example. A little may be good but a lot can pollute surface and groundwater supplies.

ORGANIZATIONAL FRAMEWORK FOR COMBATTING DESERTIFICATION

Desertification programs almost always are administered by governmental and non-governmental units of agriculture, the environment, or soil and water conservation. Every country in the world has a national agency dealing with agriculture. Most nations also have a ministry, department or independent agency responsible for environmental affairs. Soil and water conservation commonly is carried out in ministries of agriculture. Desertification, with its central emphasis on the conservation of natural resources, fits easily within already established government agencies, for the most part. Soil and vegetation degradation, the cornerstone of desertification, is the subject of teaching and research in agronomy and range science departments of universities. There is nothing significantly different about desertification that would ordinarily require establishment of a new administrative unit. The absence of special desertification departments or agencies does have the disadvantage of not publicizing the issue or not giving it legitimacy within governments and teaching and research institutions.

The same organizational framework noted earlier applies to hundreds of NGOs. Those environmental organizations have contributed greatly to local, national, and international efforts to combat desertification. They have been particularly effective in Africa, India, and Australia. A coordinating agency, the *Environment Liaison Center International* in Nairobi, Kenya, has a membership of 850 organizations in 107 countries. It is a global communications network for NGOs and community groups working on environmental and development affairs, with desertification a major issue. Funding for the Center comes from United Nations agencies and international donor organizations in developed countries. In 1994, a new international network for NGOs called the International NGO Network on Desertification and Drought (RIOD) was organized. It is operated out of the Environment Liaison Center International.

The United Nations Environment Programme, headquartered in Nairobi, has been the international leader in assisting governments in developing countries to plan national action programs and establish desertification control projects. Funding for projects comes mainly from donor agencies, the United Nations Development Program, and the World Bank. NGOs frequently are especially effective at the local level in generating community support for forestry and soil and water conservation projects. The United Nations Food and Agriculture Organization (FAO) carries out many projects dealing with desertification processes, particularly projects on soil erosion control.

In 1996, an international United Nations convention (agreement) on combating desertification and drought became operative. The official name of the convention is The United Nations Convention to Combat Desertification in Those Countries Experiencing Serious Drought and/or Desertification, Particularly in Africa. It usually is referred to as the Convention to Combat Desertification (CCD). The objective of the convention is to combat desertification and mitigate drought effects by an integrated approach that will contribute to sustainable development in affected areas. The secretariat for the CCD is in Bonn, Germany.

See also: Deserts, Volume 1; Desertification, Volume 3.

FURTHER READING

Journal articles assessing progress since 1977 conference (1984) *Environ. Conserv.*, 11(1,2).

Mainguet, M (1991) *Deserti cation*, Springer-Verlag, London, 1–306.

United Nations (1978) Round-up, Plan of Action, and Resolutions, United Nations Conference on Desertification, New York, 1–43.

United Nations Environment Programme (1992) Status of Desertification and Implementation of the United Nations Plan of Action to Combat Desertification, Report of the Executive Director, UNEP/GCSS, III/3, 88.

United Nations Environment Programme, *Deserti cation Control Bulletin*, published bi-annually, PO Box 30552, Nairobi, Kenya.

National Self-interest: a Major Factor in International Environmental Policy Formulation

Detlef F Sprinz[1] and Tapani Vaahtoranta[2]

[1] *Potsdam Institute for Climate Impact Research, Potsdam, Germany*

[2] *Geneva Centre for Security Policy, Geneva, Switzerland*

The second half of the 20th century witnessed substantial growth in the number and magnitude of international environmental problems, both transboundary and global in scope. In response, a multitude of actors - from local to global - have responded to the challenge of international environmental problems. Egoistic (systematic self-interest) states, represented by governments, are still the major institutions attending to such challenges, both as entities of international law by (not) taking over legally binding obligations and in terms of (not) providing the resources needed to pursue environmental policies. Which incentives lead these countries to negotiate international environmental agreements that mitigate or adapt countries to international environmental problems? The interest-based explanation of international environmental policy (Sprinz and Vaahtoranta, 1994) provides a parsimonious answer by focusing on two crucial factors, namely ecological vulnerability and abatement costs. By classifying countries according to their values of both variables, it can be explained why some countries act as pushers for reducing international environmental problems while other countries act as draggers. Furthermore, it is suggested that technological advances may allow more countries to pursue strict international environmental policies as a result of declining abatement costs.

THE INTEREST-BASED EXPLANATION

The interest-based explanation suggests a benefit–cost approach as the central calculus driving a state's decision to strive for strict international environmental agreements. The degree to which countries are ecologically vulnerable provides incentives to remedy the situation by undertaking beneficial mitigation (e.g., reducing emissions) or adaptation measures (e.g., enhancing ground water supplies in arid areas). Thus, the interest-based explanation suggests that countries which experience high ecological vulnerability (i.e., damages to the environment) will favor strict international environmental regulations as compared to countries which have a resilient environment. However, countries are not solely driven by considerations of the environmental damages avoided by way of international environmental policies. Instead, they also take the abatement costs of such measures into account. Other aspects held constant, the interest-based explanation expects that countries with low abatement costs will be more inclined to strive for stringent international rules as opposed to countries with high abatement costs. Once countries are classified along both dimensions, namely ecological vulnerability and abatement costs, as high or low, four groups of countries can be distinguished (see Table 1).

In particular, we expect countries with high ecological vulnerability and low abatement costs to act as *pushers* in international environmental negotiations, whereas countries with the opposite characteristics are expected to act as *draggers* (see Table 1). Countries with low ecological vulnerability and low abatement costs are expected to act as *bystanders* since they have little ecological incentive to strive for strict regulations, but enjoy low abatement costs. Countries with high ecological vulnerability and high abatement costs are expected to act as intermediates since their pursuit of strict international environmental regulations will be demanding in terms of resource mobilization (see Table 1).

This four-fold typology generates expectations about the behavior of countries. Since the benefit–cost ratio is the most favorable for pushers and the least favorable for draggers – with bystanders and intermediates falling in between – the interest-based explanation also generates expectations about the stringency of the international rules favored: pushers will lead with high demands and draggers will be the least enthusiastic, whereas both bystanders and intermediates will fall in between.

The interest-based explanation of international environmental policy is not static. Technological progress often reduces abatement costs. Should this be the case, formerly less enthusiastic countries (draggers and intermediates) are likely to be able to strive for international environmental regulations that are more strict than those originally favored.

The original explanation (Sprinz and Vaahtoranta, 1994) was tested in two case studies, namely the Montreal Protocol on Substances that Deplete the Ozone Layer (see, e.g., Benedick, 1998) and the Protocol to the 1979 Convention on Long-range Transboundary Air Pollution on

the Reduction of Sulfur Emissions or Their Transboundary Fluxes by at Least 30% (see, e.g., Jackson, 1990; Levy, 1993; Sprinz and Helm, 1999). By using quantitative data and either thresholds or averages for allocating countries to groups according to ecological vulnerability and abatement costs, it could be shown that the predictions were largely borne out in the case of stratospheric ozone. In the case of acid rain, only the group of pushers was well explained, whereas the countries in the remaining three groups mostly behaved passively in regulating sulfur emissions in the 1980s and early 1990s. Furthermore, it could be shown that reductions in abatement costs made countries more inclined to agree to more demanding environmental policies (stratospheric ozone); and economic interests to sell abatement technologies led countries to favor stricter international rules (stratospheric ozone and acid rain).

In the following section, we will summarize how the interest-based explanation was received in the literature (see the section on discussion in the literature) and outline a strategy to extend the explanation towards compliance with international environmental agreements (see the section on re ections and extensions).

DISCUSSION IN THE LITERATURE

The academic debate on the interest-based explanation focused on two aspects: the underlying assumptions about the role of the state, and the variables included in the explanation.

Role of the State

The interest-based explanation assumes that the nation–state plays a central role in international environmental policy. It explains in a parsimonious way how countries reacted to international environmental problems in the latter part of the 20th century (the explanation has been critized for not attending to the causes of international environmental action, but this was never the goal of the explanation, see Frank, 1997) by focusing on the interests of countries (Keohane, 1997). In discussing the interest-based explanation, the literature emphasizes four aspects as being central for international environmental action: (i) public pressure on decision makers resulting from (ii) dramatic instances of environmental degradation and (iii) scientific understanding of the environmental problem; furthermore, (iv) it has been suggested that international institutions have been responsible for the current environmental agreements (see Frank, 1997; Meyer *et al.*, 1997; Ungar, 1995). While providing useful insights into international environmental action, these factors offer an insufficient explanation of international environmental policy since they omit the role of the state and its interests.

If countries were not ecologically vulnerable, there would be no need for environmental protection. Vulnerability by itself may not be able to spur a response. Rather, the scientifically informed perception of an ecological risk can make decision makers act on environmental degradation. (We are grateful to Matthew Auer for drawing our attention to this point (private correspondance, 8th May 1999)). Incorporating ecological risk perception makes the interest-based explanation more dynamic since the perception of ecological vulnerability may change even in the absence of changes in the state of the environment. The same observation applies to abatement costs. Public pressure may also contribute to international environmental action, although it is suggested in the literature that those political leaders having played a key role in pushing for environmental protection have not been driven by public concern but by their personal beliefs about the severity of environmental problems (see Kawashima, 1997).

However, the observable reality shows that states' preferences and policies differ from each other. The divisions cut across more familiar ways of categorizing states. For instance, one cannot necessarily predict states' preferences with reference to Western and developing countries (this categorization is used in Meyer *et al.*, 1997). For example, though western industrialized countries have been the main pushers in the international negotiations to curb the emissions of greenhouse gases, they have not acted as a unified group. The United States has referred to scientific uncertainties as well as high abatement costs and has tended to act as a dragger. Most developing countries have been sceptical about the need to act on climate change, but the judgment has not been shared by the small island nations that are particularly vulnerable to the potential rise in sea level and to greater incidences of tropical storms (for an introduction to the international politics of climate change, see Soroos (1997), 176–207, as well as Luterbacher and Sprinz, 2001). Neither does it seem useful to regard international institutions as the dominant cause of international environmental policy. For a measurement concept of the effect of international environmental institutions and a review of the literature, see Sprinz and Helm (1999), as well as Helm and Sprinz (2000). The emphasis on international institutions fails, in particular, to account for variance in states' reactions to environmental degradation. A more fundamental reason for not regarding the United Nations or other international institutions as the main actors is the fact that states have established international institutions to facilitate environmental cooperation among themselves (Bernauer, 1995).

In short, the assumption of the interest-based explanation is that the main actor responsible for international environmental policy is the state, whose preferences may be shaped by factors such as scientific understanding and public concern. Even though the state is regarded as the main actor in international environmental policy,

international cooperation to protect the environment should not necessarily be perceived as being impossible. The success of international environmental policy depends to a great extent on whether states perceive shared interests or not. Whereas in the case of stratospheric ozone depletion, there has been a sufficient degree of shared interests to enable states to agree on the protection of the ozone layer, we witness a greater diversity of state interests in the case of climate change.

Explanatory Variables

The second major reaction concerns the explanatory variables used in the interest-based explanation. Some authors suggest that the explanation could be even more parsimonious by focusing only on abatement costs (for a strong focus on the economic cost of environment protection as a factor in uencing states' policies, see Biermann (1997), Montes and Magno (1997), as well as Vogel, 1997) whereas others would like to add more explanatory variables. We find it difficult to exclude environmental vulnerability from the explanation without seriously weakening its ability to account for international environmental policy. In particular, it would be difficult to explain why a country would ever accept costly environmental remedies in the absence of ecological vulnerability. There are several examples of how environmental vulnerability or the perception of it shapes the preferences of states. For instance, the policy of the Federal Republic of Germany on acid rain changed in the early 1980s when the Germans learned about the large extent of damage to their own forests; and Finland has been particularly concerned about the environmental problems in the bordering Russian regions and has been willing to pay for controlling and solving these problems which are adverserly affecting the environment of Finland.

A first lesson to be drawn from the interest-based explanation is that economic side payments can be used to change the preferences of states and to in uence their policies on environmental protection. As David Vogel concludes, one way of promoting effective international environmental action is for the more af uent countries not only to change their own policies but also to provide less af uent countries with sufficient incentives to modify theirs as well (Vogel, 1997). The question is how much is required and feasible. From an economic perspective, it would be most effective to protect the international environment if transfers are kept sufficiently small in relation to the donor country's wealth and sufficiently large in relation to the recipient country's wealth.

A second lesson highlights the role of changing the ecological risk perception of laggards and bystanders. China's policy on curbing the emissions of greenhouse gases, for example, should favor relative emission reductions in a pronounced way if China were to change its evalution of the vulnerability of its low-lying coastal areas to the threat posed by the rise of sea level.

Two studies tested the interest-based explanation or used a similar explanatory approach, and deserve more detailed attention. Rowlands (1995) uses the explanation to find out how well it accounts for national climate change policies. He calculates environmental vulnerability and abatement costs indices for 24 countries. To determine the ecological vulnerability of countries, Rowlands uses variables which measure the impacts caused by sea-level rise and measures of the impacts on agriculture. For the abatement costs he relies on three previous studies and uses the projected energy needs as well as the potential for fuel substitution to estimate the abatement costs for countries not included in the studies. After having incorporated secondary benefits, he predicts that there should be three kinds of responses to the limiting of climate change: pushers should be proactive in the efforts to reach an agreement, draggers should resist such efforts, and intermediates should be passive. Rowlands' findings show that 11 out of 24 countries were acting as predicted by the interest-based explanation and concludes that the interest-based explanation may be helpful in understanding states' environmental foreign policies. We concur with Rowlands that our explanation may be linked to a more extensive domestic policy model, and that the explicit operationalization of the two dimensions of the explanation is crucial.

In her interview-based study, Kawashima (1997) explains how five industrialized countries adopted their policies to reduce carbon dioxide (CO_2) emissions. While she does not apply the interest-based explanation, she uses similar explanatory factors to account for the behavior of states. Based on the policies pursued, countries are divided into three groups. The Netherlands and Germany are leaders, the UK and Japan act as followers, and the US is the only laggard in her sample. Kawashima's research suggests that six factors might have affected the preferences of states, namely harm caused by climate change (ecological vulnerability), economic cost of reducing emissions (abatement costs), domestic factors, international politics, the degree to which an international agreement is binding, and the experience of past negotiations on other environmental issues. According to the answers from expert interviews, four factors affected the national policies in particular. Interviewees from all countries except the US were concerned about the ecological impact of climate change, whereas US interviewees stressed the importance of high CO_2 abatement costs. These results are in accordance with the interest-based explanation. Kawashima also found that two additional factors had played a role in shaping the states' preferences. First, the role of political leaders was regarded as one of the most decisive factors in the Netherlands, Germany and Japan. These leaders pushed for emission reductions regardless of the scientific evidence of

the adverse ecological impact or of the abatement costs involved. Second, international politics played a decisive role for the UK and Japan. International pressure within the European Union in uenced British policy, and Japan's policy was affected by its concern for an active international role. Kawashima agrees with Rowlands that domestic policy and international politics may shape states' policies on environmental issues.

In general, we suggest that the present form of the interest-based explanation permits a first order assessment of the likely positioning of countries in negotations on international environmental agreements (for an affirmative assessment of the method used in the interest-based explanation, see Zurn, 1997). Subsequent detailed studies may allow for the incorporation of additional variables which permit detailed explanation of the behavior of a more limited set of countries. However, the purpose of the interest-based explanation is to simplify reality in order to draw attention to major causal relationships in international environmental policy. A more elaborate explanation will have to demonstrate that the inclusion of additional factors leads to substantive increases in explanatory power.

REFLECTIONS AND EXTENSIONS

Re ecting on the interest-based explanation, we will discuss two aspects which merit attention: a variance argument and an intercomparison (or locking) argument. In addition, we will outline how the original explanation can be extended from the stage of negotiating international environmental agreements to the stage of compliance.

First, the usefulness of the classification scheme seems to vary with the magnitude of the variance found across the two variables, namely ecological vulnerability and abatement costs. If there is very little variation in the values of both variables across countries, the classification becomes less robust. Thus, the higher the variance, the less arbitrary the thresholds, and the stronger the classificatory power of the explanation.

Second, the operationalization of the interest-based explanation rests on the relative ecological vulnerability and abatement costs, i.e., it relies on the relative positioning of countries within the same pollution domain. However, the explanatory power of the explanation could be enhanced by pooling the abatement costs and ecological vulnerability scores from a large set of pollution domains across countries – using a common metric which makes such data intercomparable. In such a pooled analysis, an (absolute) average threshold for ecological vulnerability and abatement costs could be determined for the set of countries (locking of the thresholds) and the particular pollution case could now be nested within the larger pooled analysis. This would allow us to assess whether we should expect more pushers in these negotiations as compared to the pooled average, and, subsequently, more stringent regulations. Similar propositions can be made with respect to other quadrants of the classification found in Table 1.

In general, each party to an international treaty is expected to comply with its rules. Some treaties also have specific articles on non-compliance, but still some states do not always comply with treaties. The original interest-based explanation derived hypotheses about the expected negotiation behavior of states. In the remainder of this section, we outline an extension of the explanation so as to capture expected compliance behavior while keeping the basic logic of the original explanation. For reasons of convenience, we do not assume that countries change their classificatory label between both phases (e.g., because of changes in the abatement costs or their ecological vulnerability). Furthermore, we assume that at the end of the negotiations countries are forced to take the decision of whether to sign or not sign the international environmental agreement (signing an international agreement may take various forms such as accession, signature, ratification without signature, etc.). This may lead to the expectation that the signatories will be bound by the legal provisions of the international environmental agreement whereas the non-signatories remain free to pursue any policy they wish. Contrary to this depiction spurred by the tradition of international law, this extension of the interest-based explanation to the stage of compliance provides a much richer explanation.

First, the basic provisions of international environmental agreements may lead to positive abatement costs and therefore divert resources from the pursuit of other public policies. Thus, it is hypothesized that the higher the compliance costs for a country, the lower the degree of compliance.

Second, ecological vulnerability provides countries with positive incentives to actually implement and enforce international environmental agreements domestically. Thus, it is hypothesized that the higher the degree of ecological vulnerability for a country, the higher the degree of compliance with the basic rules of an international environmental agreement.

Pursuant to the classification introduced earlier, we suggest that one should dichotomize countries into those with low versus high compliance costs and those with low versus high ecological vulnerability, and derive expected behavior accordingly. Countries which are signatories of an international environmental agreement will carry the attribute committed', whereas non-signatories will be labeled uncommitted' (see Tables 2 and 3).

Even if a country does not sign an international environmental agreement, high ecological vulnerability provides it with some incentives to undertake actions in line with the provisions of the international environmental agreement (e.g., uncommitted pushers), especially if self-help is possible and efficient. This is also supposed to hold for countries

Table 2 Expected compliance of signatory countries[a]

		Ecological vulnerability	
		Low	High
Relative national Compliance costs	Low	Committed bystander medium compliance	Committed pusher full to (voluntary) over-compliance
	High	Committed dragger no to low compliance	Committed intermediate medium compliance

[a] Upper entries depict classificatory labels, lower entries refer to the degree of compliance. Scale for degree of compliance: none, low, medium, full, over-compliance. Scale for ecological vulnerability and compliance costs: low, high.

Table 3 Expected compliance of non-signatory countries[a]

		Ecological vulnerability	
		Low	High
Relative national Compliance costs	Low	Uncommitted bystander low compliance (voluntary)	Uncommitted pusher medium to full compliance (voluntary)
	High	Uncommitted dragger no compliance	Uncommitted intermediate low compliance (voluntary)

[a] Upper entries depict classificatory labels, lower entries refer to the degree of compliance. Scale for degree of compliance: none, low, medium, full, over-compliance. Scale for ecological vulnerability and compliance costs: low, high.

with low compliance costs. Thus, we suggest that the same rationale holds for signatories as well as non-signatories. However, non-signatories are expected to achieve a lower level of (voluntary) compliance with the provisions of an international environmental agreement (compare Tables 2 and 3). Nevertheless, for most groups we should expect actual steps towards partial compliance rather than inaction. We will highlight this by way of paired comparisons of each group across signatory status.

Five levels of compliance are used (in ascending order): none, low, medium, full, and over-compliance. The lowest rating (none) refers to no compliance behavior or even actions which are directed against reducing ecological vulnerability (e.g., increased emission); low compliance relates to very minor steps taken in the direction of compliance; medium refers to steps short of full compliance; full compliance refers to 100% congruence of the rules in the international environmental agreement and compliance-related behaviors; and over-compliance refers to any substantial exceedance of full compliance. This 5-step scale should not be confused with the two-point scale (low vs. high) for ecological vulnerability and compliance costs.

First, push countries in international negotiations seem to have the greatest incentives to actually sign and comply with international environmental agreements because of the attractive benefit–cost ratio. Thus, we should expect full compliance, at a minimum, and possibly (voluntary) over-compliance. If a push country does not succeed in upgrading international rules sufficiently, it may decide not to sign the international agreement, e.g., for domestic political reasons which would not support weak international rules. However, given their interest configuration, we should expect uncommitted pushers to voluntarily implement the international environmental agreements to a medium or full extent.

Second, draggers are likely to show the reverse behavior because of the low benefit–cost ratio. For committed draggers, signing international environmental agreements only fulfills the useful functions of creating leverage at extracting resources from other countries in order to enhance compliance, or the opportunity to in uence the negotiations of subsequent international environmental agreements. For uncommitted draggers, there are even fewer incentives to undertake any actions towards compliance. In fact, we should expect no compliance at all.

Third, intermediate countries are challenged by the high costs of compliance. Therefore, we should expect committed intermediates to show medium levels of compliance, whereas uncommitted intermediates will only endeavor to show low levels of (voluntary) compliance. Thus, the vigor of compliance behavior falls in between those of pushers and draggers because of the intermediate benefit–cost ratio. A similar calculus applies to bystanders.

In general, this approach to compliance predicts a substantial range of compliance levels – even among signatory countries. While signatory countries are likely to show uniformly higher levels of compliance as opposed to non-signatory countries, even some non-signatory countries are likely to contribute to the alleviation of international environmental problems due to their enduring ecological and economic incentives.

CONCLUSIONS

The interest-based explanation of international environmental policy provides predictions of the likely behavior of countries in international environmental negotiations, by utilizing a combination of ecological vulnerability and abatement costs. Given its parsimonious and operational nature, the interest-based explanation can be empirically tested and extended from the stage of negotiating international agreements to the stage of expected compliance with international environmental agreements. While the critique of the explanation has focused on the role of states in international environmental action, as well as the number of variables to be included in research, we remain confident that the interest-based explanation provides a successful approach to derive initial expectations of the behavior of countries.

REFERENCES

Benedick, R E (1998) *Ozone Diplomacy - New Directions in Safeguarding the Planet*, Harvard University Press, Cambridge, MA.

Bernauer, T (1995) The Effect of International Environmental Institutions: How We Might Learn More, *Int. Organ.*, 49(2), 351–377.

Biermann, F (1997) Financing Environmental Policies in the South. Experiences from the Multilateral Ozone Fund, *Int. Environ. Aff.*, 9(3), 179–218.

Frank, D J (1997) Science, Nature, and the Globalization of the Environment, 1870–1990, *Soc. Forces*, 76(2), 409–435.

Jackson, C I (1990) A Tenth Anniversary Review of the ECE Convention on Long-range Transboundary Air Pollution, *Int. Environ. Aff.*, 2, 217–226.

Kawashima, Y (1997) A Comparative Analysis of the Decision-making Process of Developed Countries Toward CO_2 Emissions Reduction Targets, *Int. Environ. Aff.*, 9(2), 95–126.

Keohane, R O (1997) What Can Political Scientists Contribute to an Understanding of Environmental Policy?, *Deg. Chan.*, 9, 1–4.

Levy, M A (1993) European Acid Rain: The Power of Toteboard Diplomacy, in *Institutions for the Earth - Sources of Effective International Environmental Protection*, eds P M Haas, R O Keohane, and M A Levy, MIT Press, Cambridge, MA, 75–132.

Luterbacher, U and Sprinz, D F, eds (2001) *International Relations and Global Climate Change*, Cambridge University Press, Cambridge, MA.

Meyer, J W, Frank, D, Hironaka, A, Schofer, E, and Tuma, N (1997) The Structuring of a World Environmental Regime, 1870–1990, *Int. Organ.*, 51(4), 623–651.

Montes, M F and Magno, F A (1997) Trade and Environmental Diplomacy: Strategic Options for ASEAN, *Pac. Aff.*, 70(3), 351–372.

Rowlands, I (1995) Explaining National Climate Change Policies, *Global Environ. Change*, 5(3), 235–249.

Soroos, M S (1997) *The Endangered Atmosphere - Preserving a Global Commons*, University of South Carolina Press, Columbia, SC.

Sprinz, D and Carsten, H (1997) *International Governmental Organizations and National Participation in Environmental Regimes: the Organizational Components of the Acidi cation Regime*, Erasmus University, Rotterdam.

Sprinz, D and Vaahtoranta, T (1994) The Interest-based Explanation of International Environmental Policy, *Int. Organ.*, 48(1), 77–105.

Sprinz, D F and Helm, C (1999) The Effect of Global Environmental Regimes: a Measurement Concept, *Int. Polit. Sci. Rev.*, 20(4), 359–369.

Ungar, S (1995) Social Scares and Global Warming: Beyond the Rio Convention, *Soc. Nat. Resour.*, 8(5), 443–456.

Vogel, D (1997) Trading Up and Governing Across: Transnational Governance and Environmental Protection, *J. Eur. Public Policy*, 4(4), 556–571.

Zurn, M (1997) Assessing State Preferences and Explaining Institutional Choice: the Case of Intra-German Trade, *Int. Stud. Q.*, 41(2), 295–320.

FURTHER READING

Helm, C and Sprinz, D F (2000) Measuring the Effectiveness of International Environmental Regimes, *J. Con ict Resolution*, 45, 630–652.

Natural Hazards: Social and Policy Responses

Dennis S Mileti and Mary Fran Myers
University of Colorado, Boulder, CO, USA

Natural hazards and disasters are not isolated phenomena that can be divorced from the societies they affect: a disaster is a symptom of broader and more basic problems. Hazard losses, and the fact that there seems to be an inability in most if not all societies to reduce those losses, are the consequences of narrow and short-sighted development patterns, cultural premises, and attitudes toward the natural environment, science, and technology. A proposal is put

forward to assist people and societies in taking responsibility for reducing hazard losses.

DISASTER LOSSES ARE GROWING

Natural hazards have caused a steady increase in property losses in the past century as the world's population has grown and increasingly occupies hazard-prone areas of the planet.

For example, the dollar losses associated with most types of natural hazards in the United States are rising. A conservative estimate of total dollar losses from 1975 to 1994 is $500 billion (in 1994 dollars). More than 80% of these costs stemmed from climatological events. More revealing is that in that 20-year time frame, the ten most costly disasters (based on dollar losses) in the US occurred between 1989 and 1994, suggesting bigger and more catastrophic losses are on the rise.

While the United States has succeeded in saving lives and reducing injuries from some natural hazards such as hurricanes over the last half-century, recent events such as Hurricane Mitch in Central America, the Venezuelan oods and landslides, and the earthquake in Turkey which claimed thousands of lives indicate that deaths and injuries from natural hazards have not declined substantially world-wide.

WHY DISASTER LOSSES ARE GROWING

Many disaster losses are the predictable results of interactions among three major systems: the global physical environment, which includes climate change and the occurrence of hazardous events; the social and demographic characteristics of the communities and societies that experience disasters; and the buildings, roads, bridges, and other components of the built environment. Growing losses result partly from the fact that the world's capital stock is expanding, but they also stem from the fact that all these systems and their interactions are becoming more complex. Three main in uences are at work.

Global Change

First, the Earth's physical systems are constantly changing. For example, a warming climate is expected to produce more dramatic meteorological events such as storms, oods, drought, and heat waves. Indeed, the greatest impact of climate change on human beings probably will be an increase or decrease in the frequency of different types of extreme events that initiate natural disasters.

Socio-demographic Change

Second, recent and projected changes in the demographic composition and distribution of the world's population mean greater exposure to many hazards. The number of people residing in earthquake-prone mega cities and coastal areas subject to cyclones, for example, is growing rapidly. The immigration of poor people into these mega cities invariably results in their occupying the most vulnerable ood- or landslide-prone areas of these communities. Worsening inequality of wealth also makes many people more vulnerable to hazards and less able to recover from them.

Changes in the Built Environment

Third, the built environment – public utilities, transportation and communication systems, and homes and office buildings – is growing in density, making the potential losses from natural forces larger. The ability of the built environment to withstand the impacts of extreme natural forces plays a direct role in determining the number of lives lost, the number and severity of injuries, and the financial impact of disasters.

Other Issues

Another major problem has become clear over the past 20 years: some efforts to head off damages from natural hazards only postpone them. For example, communities below dams or behind levees may avoid losses from the oods that those structures were designed to prevent. But such communities often have more property to lose when those structures fail. This is due to the fact more development occurred as a result of the structure because people thought they would be safe from the ood risk. Such a situation contributed to catastrophic damage from the 1993 oods in the Mississippi basin. And many of the world's dams, bridges, and other structures are approaching the end of their designed life revealing how little thought their builders gave to events 50 years hence. The availability of insurance in some cases induces people to occupy hazard-prone areas. Similarly, by providing advance warnings of severe storms, it is possible that more people have been encouraged to build on fragile coast lines. Such development, in turn, makes the areas more vulnerable by destroying dunes and other protective natural features.

THE CYCLICAL NATURE OF COPING WITH NATURAL HAZARDS

Societal adjustments to natural hazards have typically been viewed in a cyclical framework comprised of four stages: preparedness, response, recovery, and mitigation. Preparedness involves building an emergency response and management capability in advance of a disaster to facilitate an effective societal response if a disaster occurs. Response refers to the actions taken immediately before, during, and after a disaster in order to save lives, minimize property

damage, and enhance the effectiveness of recovery. Disaster recovery involves short-term activities to restore vital life-support systems to minimum operating standards and long-term activities to return life to normal or improved quality. Mitigation refers to the policies and activities that will reduce an area's vulnerability to damage from natural disasters. It is important to note that the word *mitigation* as used here is equivalent to the word *adaptation*. In this context, mitigation is much broader in scope than it is when referred to in the global climate change literature. Mitigation generally includes those adjustments described below (*see* Adaptation Strategies, Volume 4).

TRADITIONAL APPROACHES TO DEALING WITH HAZARDS

Over past decades an array of techniques and practices has evolved to reduce and cope with losses from disasters.

Land Use

Wise land use planning that limits expansion into sensitive areas can keep people and property out of the way of hazards and can also contribute to promoting environmental quality.

Warnings

Significant improvements in short-term forecasts and warnings (hours to days ahead of a hazardous event) have dramatically reduced loss of life and injury in North America and other parts of the developed and developing world. It is important to remember, however, that warning systems do not significantly limit damage to property, nor do they mitigate economic disruption from disasters.

Engineering and Building Codes

Disaster-resistant construction of buildings and infrastructure is an essential component of any mitigation program. Engineering codes, standards, and practices have been promulgated for natural hazards; now they must be incorporated into local building codes and practices world-wide, with special attention paid to the enforcement of those codes.

Insurance

Insurance does help minimize some disruption caused by disasters by ensuring that people with coverage receive compensation for their losses. However, most property owners do not buy coverage against special perils, notably earthquakes, hurricanes, and oods. For example, in the US only about 20 percent of the homes exposed to oods are insured for them. Further, the industry has problems providing insurance in areas subject to catastrophic losses because many insurers do not have the resources to pay for a worst-case disaster.

New Technology

Computer-mediated communication systems, geographic information systems (GIS), remote sensing, electronic decision-support systems, and risk-analysis techniques have developed substantially during the last two decades and can support hazard mitigation programs. For example, GIS models enable managers to consolidate information from a range of disciplines, including the natural and social sciences and engineering, and to formulate plans for mitigation accordingly (*see* Geographic(al) Information Systems (GIS), Volume 4).

Emergency Preparedness and Recovery

Good mitigation programs will never eliminate the need for plans to address the destruction and human suffering imposed by disasters. As such, creating policies for disaster preparedness and response are needed to save lives and injuries, limit property damage, and minimize disruptions, when extreme environmental events do occur. In addition, such policies that are extended to include plans that focus on recovery and reconstruction for rebuilding in safer ways and in safer places can be effective in reducing future vulnerability to disasters.

A CENTRAL PROBLEM WITH TRADITIONAL APPROACHES

One central problem with the accepted methods described above for coping with hazards is that they have been based on the idea that people can use technology to control nature to make themselves safe. Further, most strategies for managing hazards have followed a traditional planning model: study the problem, implement one solution, and move on to the next problem. This approach suggests that hazards are static and mitigation is an upward, positive, linear process.

But natural disasters are not linear problems that can be solved in isolation. Rather, they are symptoms of broader and more basic problems. Losses from hazards, and the fact that most societies cannot seem to reduce them, result from shortsighted and narrow conceptions of the human relationship to the natural environment.

To redress those shortcomings, a shift is needed to a concept that links wise management of natural resources with local economic and social resiliency and that views hazard mitigation as an integral part of overall sustainable development.

A NEW APPROACH TO HAZARDS

The main guidelines for achieving this shift and improving society's ability to mitigate hazards are described below.

Adopt a Global Systems Perspective

Rather than resulting from surprise environmental events, disasters arise from the interactions among the Earth's physical systems, its human systems, and its built infrastructure. A broad view that encompasses all three of these dynamic systems and interactions among them can enable us to find better solutions.

Accept Responsibility for Hazards and Disasters

Human beings and not nature are the cause of disaster losses, which stem from choices about where and how human development will proceed. There is no final solution to natural hazards, since technology cannot make the world safe from all the forces of nature, and in some cases (e.g., climate change) can even be seen to be increasing these forces. As such, humans must accept responsibility for their decisions that result in future risks.

Anticipate Ambiguity and Change

The view that hazards are relatively static has led to the false conclusion that any mitigation effort is desirable and will in some way reduce future losses. In reality, change can occur quickly and nonlinearly. Human adaptation to hazards must become as dynamic as the problems presented by hazards themselves.

Reject Short-term Thinking

Mitigation as frequently conceived is too short-sighted. In general, people have a cultural and economic predisposition to think primarily in the short term. Sustainable mitigation will require a longer-term view that takes into account the overall effect of mitigation efforts on this and future generations.

Account for Social Forces

Societal factors, such as how people view hazards and mitigation efforts or how the free market operates, play a critical role in determining the extent of future disaster losses. Because such social forces are so in uential, growing understanding of physical systems and improved technology alone cannot suffice. To effectively address natural hazards, mitigation must become a basic social value.

Embrace Sustainable Development Principles

Disasters are more likely where unsustainable development occurs, and the converse is also true: disasters hinder movement toward sustainability because, for example, they degrade the environment and undercut the quality of life. Sustainable mitigation activities should strengthen a community's social, economic, and environmental resiliency, and *vice versa*. By incorporating disaster resistance into community planning and economic development, growth can be managed and can result in lessening the impact that disasters have on communities, businesses, and citizens.

ESSENTIAL GOALS FOR A NEW APPROACH

Sustainability means that a locality can tolerate and overcome damage, diminished productivity, and reduced quality of life from an extreme event without significant outside assistance. To achieve sustainability, communities must take responsibility for choosing where and how development proceeds. Each locality must evaluate its environmental resources and hazards, choose future losses that it is willing to bear, and ensure that development and other community actions and policies adhere to those goals.

Six objectives must simultaneously be reached to mitigate hazards in a sustainable way and stop the trend toward increasing catastrophic losses from natural disasters.

Maintain and Enhance Environmental Quality

Human activities to mitigate hazards should not reduce the carrying capacity of the ecosystem, for doing so increases losses from hazards in the longer term.

Maintain and Enhance People's Quality of Life

A population's quality of life includes, among other factors, access to income, education, health care, housing, and employment, as well as protection from disaster. To become sustainable, local communities must consciously define the quality of life they want and select only those mitigation strategies that do not impede any aspect of that vision.

Foster Local Resiliency and Responsibility

Resiliency to disasters means that a locale can withstand an extreme natural event with a tolerable level of losses. Mitigation actions should be consistent with achieving that level of protection.

Recognize that Vibrant Local Economies are Essential

Communities should take mitigation actions that foster a strong local economy rather than detract from one.

Ensure Inter- and Intra-generational Equity

A sustainable community selects mitigation activities that reduce hazards across all ethnic, racial, and income groups, and between genders equally, now and in the future. The costs of today's advances should not be shifted onto later generations or less powerful groups.

Adopt Local Consensus Building

A sustainable community selects mitigation strategies that evolve from full participation among all public and private stakeholders. The participatory process itself may be as important as the outcome.

RECOMMENDED ACTIONS AND INITIAL STEPS BEING TAKEN

The shift to a sustainable approach to hazard mitigation will require extraordinary actions. Here are several essential steps; many initial efforts that are underway are described.

Building Local Networks, Capability, and Consensus

Today, hazard specialists, emergency planners, resource managers, community planners, and other local stakeholders seek to solve problems on their own. An approach is needed to forge local consensus on disaster resiliency and nurture it through the complex challenges of planning and implementation.

One potential approach is a network of stakeholders in every community that would engage in collaborative problem solving. Each network would produce an integrated, comprehensive plan linking land use, environmental, social, and economic goals, as well as disaster resiliency. The stakeholder network especially needs to determine the amount of damage that those who experience disasters can bear. These plans would enable policymakers, businesses, and residents to understand the limitations of their region and work together to address them. Full consensus may never be reached, but the process is key because it can foster the sense of community required to mitigate hazards.

This kind of holistic approach situates mitigation in the context of other community goals that, historically, have worked against action to reduce hazards. Also, the process ensures that each locality controls the character of its disasters, forcing stakeholders to take responsibility for natural hazards and realize that the decisions they make today will determine future losses.

A problem with this approach is the challenge of generating sufficient interest to create a stakeholder network. However, efforts to overcome this problem are being funded by the US Federal Emergency Management Agency (FEMA) through its Project Impact program. Project Impact is a community based initiative designed to bring key partners together, including the private and non-governmental sectors, to work in concert with local government officials to develop and implement hazard mitigation projects which not only support disaster loss reduction but meet other community goals such as increased economic vitality and environmental quality as well.

Establish a Holistic Government Framework

To facilitate sustainable mitigation, all government policies and programs related to hazards and sustainability should be integrated and consistent. Statutes and regulations should be reexamined in light of the principles of sustainable mitigation. Changes needed might include limiting the subsidization of risk, making better use of incentives, and setting national policies for land use.

New Zealand commenced major reform of its emergency management framework in the mid-1990s. The modifications are designed to transfer the nation's response-focused counter-disaster approach into a system grounded in a sustainable hazard management framework. The reform calls for refocusing action onto the management of risk and the natural environment.

Conduct Nationwide Hazard and Risk Assessments

Not enough is known about the changes in or interactions among the physical, social, and built environments that are reshaping each nation's hazardous future. National risk assessments should meld information from those three environments so hazards can be estimated interactively and comprehensively, to support local efforts on sustainable mitigation. Once completed, these assessments need to be provided to communities at a scale that is sufficient for local decision-making.

Build National Databases

Countries should collect standardized data on losses from past and current disasters, to establish a baseline for comparison with future losses. The information should include the type, location, and cause of losses, taking into account problems of double-counting, comparisons with gross domestic product, and the distinction between regional and national impacts. A database is also needed on mitigation efforts: what they were, where they occurred, and how much they cost in order to provide a baseline for local cost-benefit analysis. These archives are fundamental to informed decision making and should be accessible to the public.

The US FEMA recently funded two panels of national experts at the US National Academy of Science's National

Research Council and the Heinz Center (a nonprofit environmental research organization located in Washington, DC) to address how the US might move forward to accurately count disaster losses. Although not yet implemented, FEMA and other agencies are working toward standardizing how loss data are measured and collected that might soon be brought together into a national data base.

Provide Comprehensive Education and Training

Today hazard managers are asked to deal with problems they have never before confronted, such as understanding complex physical and social systems, conducting sophisticated cost-benefit analyses, and offering long-term solutions. Education in hazard mitigation and preparedness should therefore expand to include interdisciplinary degree programs. University-based holistic programs that move away from traditional disciplines toward education designed to solve the real-world problems entailed in linking hazards and sustainability must be created.

Efforts to move away from traditional educational programs are underway in the United States. FEMA's Emergency Management Institute has established a Higher Education Initiative with a goal of establishing at least one program in a college or university in each of the 50 states that offers a bachelors degree in emergency management. Further, in cooperation with the US Army Corps of Engineers and FEMA, the Natural Hazards Research and Applications Information Center at the University of Colorado is working to create a vision for the higher education needs of the nation's hazards and emergency managers.

Measure Progress

Baselines for measuring sustainability should be established so future progress can be gauged. Interim goals for mitigation and other aspects of managing hazards should be set, and progress in reaching those goals regularly evaluated. This effort requires determining how to apply criteria such as disaster resiliency, environmental quality, intra- and inter-generational equity, quality of life, and economic vitality to the plans and programs of local communities.

Also important is evaluating hazard-mitigation efforts already in place before taking further steps in the same direction. For example, the US National Flood Insurance Program (NFIP), which combines insurance incentives and land use and building standards, has existed for more than 30 years, yet its effectiveness has never been thoroughly appraised. This is about to change as the FEMA has begun to conduct a detailed assessment of the NFIP. Six areas will be examined in detail: the occupancy and use of oodplains, costs and consequences of ooding, insurance rating, oodplain management and enforcement, hazard identification and risk assessment, and communication.

Another way to measure progress is to collect the new knowledge relevant to hazard mitigation and disaster response and recovery that results from every disaster that occurs. This information should be systematically collected and evaluated to assess progress in putting knowledge into practice.

Share Knowledge Internationally

Countries must share knowledge and technology related to sustainable hazard mitigation with one another and be willing to learn from each other as well. Disaster experts worldwide need to collaborate with development experts to address the root causes of hazard vulnerability, including overgrazing, poverty, deforestation, and unplanned development. Disaster reduction should be a part of everyday development processes, and international development projects must consider vulnerability to disaster.

There is progress being made in this arena. The United Nations' International Decade for Natural Disaster Reduction (*see* IDNDR (The International Decade for Natural Disaster Reduction), Volume 4) and the World Meteorological Organization's work in forecasting, preparedness, and mitigation have done much to promote efforts that reduce vulnerability to hazards around the globe. In addition, many of the world's financial organizations such as the International Development Bank, the Inter American Development Bank, and the World Bank now recognize the importance of incorporating mitigation into their lending and relief policies. An example is the launching of the World Bank's ProVention Consortium in February 2000. This Consortium is dedicated to reducing disaster risk in developing countries by making disaster prevention and mitigation an integral part of development initiatives.

A FEW FINAL WORDS

A good deal of progress has been made in the last few decades about understanding the relationship between society and its natural environment. Much of what societies have done in their recent past has been to adapt to environmental change and forces as they occurred. Global climate change, however, demands a shift in perspective, especially in regard to natural hazards. It illustrates the importance of adopting a future-oriented perspective and longer-term approach when dealing with the relationship between society and the natural environment: *See also* Natural Hazards, Volume 3; Adaptation Strategies, Volume 4; Risk Management in an Era of Global Environmental Change, Volume 4; Hazards in Global Environmental Change, Volume 5.

FURTHER READING

Burby, R J, ed (1999) *Cooperating with Nature: Confronting Natural Hazards with Land-use Planning for Sustainable Communities*, Joseph Henry Press, Washington, DC.

Burton, I, Kates, R W, and White, G F (1993) *The Environment as Hazard*, 2nd edition, The Guilford Press, New York.

Godschalk, D R, Beatley, T, Berke, P, Brower, D J, and Kaiser, E J (1999) *Natural Hazard Mitigation: Recasting Disaster Policy and Planning*, Island Press, Washington, DC.

Kunreuther, H and Roth, Sr, R J, eds (1999) *Paying the Price: The Status and Role of Insurance Against Natural Disasters in the United States*, Joseph Henry Press, Washington, DC.

Mileti, D S (1999) *Disasters by Design: a Reassessment of Natural Hazards in the United States*, Joseph Henry Press, Washington, DC.

Nile River

see Nile River (*Case study*, **Volume 4**)

Nitrate Sensitive Areas (NSAs)

see Nitrogen: Agricultural Uses and their Impacts (**Volume 3**)

Nitrate Vulnerable Zones

see Nitrogen: Agricultural Uses and their Impacts (**Volume 3**)

No Regrets Principle

The *no-regrets principle:* because of public apathy, uncertainties in possible outcomes, and/or the costs involved, it is sometimes difficult for a government or industry to do all that it should to prevent a long-term environmental situation (like climate warming) from developing. Thus, it is often urged that at least a start should be made through the promotion of a no-regrets policy, i.e., a policy which can be justified because it is likely to be beneficial, not only with respect to the issue under consideration, but also with respect to other environmental and/or socio-economic issues. Examples with respect to greenhouse gases include the following (Munn *et al.*, 1996):

1. Promotion of increased energy efficiency to reduce carbon dioxide emissions, thus reducing emissions of oxides of sulfur and nitrogen while conserving fossil fuel supplies for future generations.
2. The extension of industrial product cycles by reusing and recycling wastes.
3. Innovations in energy conversion, reducing environmental impacts without affecting the benefits currently enjoyed by high-energy users.
4. Tree planting not only to absorb carbon dioxide from the atmosphere, but also to provide shade during hot weather, to increase the available habitat for wildlife, and to reduce soil erosion.

However, the no-regrets principle should not be invoked as an excuse for not taking drastic action if there is evidence that a dangerous trend may otherwise become irreversible.

REFERENCE

Munn, R E, la Riviere, J W M, and van Lookeren Campagne, N, eds (1996) *Policy Making in an Era of Global Environmental Change*, Kluwer, Dordrecht, 1–106.

FURTHER READING

Hassol, S (1998) Clear Need to Act on Global Warming, *Nature*, 396, 210.

R E MUNN *Canada*

Non-equilibrium Thermodynamics

see Monitoring in Support of Policy: an Adaptive Ecosystem Approach (*Opening essay*, **Volume 4**)

O

Obasi, G O P

(1933-)

Godwin Olu Patrick Obasi, a Nigerian citizen, has been the Secretary-General of the World Meteorological Organization (WMO) since 1 January 1984. He holds a Doctor of Science degree (1963) in Meteorology from the Massachusetts Institute of Technology, USA. He won the Carl Rossby Award for the best doctoral degree thesis.

He served as Dean of the Faculty of Science, Professor and Chairman of the Department of Meteorology at the University of Nairobi, Kenya and Visiting Research Associate at the Florida State University, Tallahassee, FL. He has published over 100 scientific and technical papers. He is a Consulting Editor to many international journals in meteorology.

Professor Obasi was instrumental in the initial planning of the Global Atmospheric Research Program (GARP) Atlantic Tropical Experiment. He also served as Vice-President of the WMO Commission for Atmospheric Sciences.

As Secretary-General of WMO, Professor Obasi raised the profile of WMO and the National Meteorological and Hydrological Services (NMHSs) worldwide in their role as major contributors to the sustainable development of nations in climate-related areas including climate change, improved weather forecasts and seasonal prediction, mitigation of natural disasters, e.g., combating desertification, increasing support to food security, water resources assessment and management, and the monitoring and prediction of the state of the stratospheric ozone layer.

In support of the NMHSs, he meets regularly with Heads of State and Governments, Ministers and senior Government officials worldwide. He has also fostered active collaboration between WMO and the organizations of the United Nations (UN) as well as with other international and regional governmental and non-governmental organizations in areas of interest to the NMHSs. Professor Obasi has initiated a number of reforms within the secretariat to enhance its efficiency and cost-effectiveness. His vision is to continue to develop a respected, forward-looking and effective organization that will meet the challenge of present and future generations.

In recognition of the outstanding achievements of the WMO under his able leadership as Secretary-General, the initiatives he has launched for the development of NMHSs in support of the sustainable development of nations, as well as for this continuing active involvement and achievements in the fields of meteorology, operational hydrology and general environmental matters, Professor Obasi has been honored by the governments of several member countries, universities, academic institutions and meteorological scientists worldwide, including Romania, Philippines, Nigeria, Russian Federation, Paraguay, Kenya and Czechoslovakia.

MOSTAFA K TOLBA *Egypt*

OECD (Organization for Economic Co-operation and Development)

The OECD, founded in 1961, is an intergovernmental organization of 29 of the world's most industrialized countries. As of early 1999, the OECD member countries are: Australia, Austria, Belgium, Canada, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Republic of Korea, Spain, Sweden, Switzerland, Turkey, the UK and the US. The European Commission also takes part in the work of the OECD. It addresses economic, social and environmental policy concerns in

an interdisciplinary way, and has sustainable development as one of its strategic priorities. The OECD Environment Policy Committee (EPOC), supported by the Environment Directorate, has been working on environmental policy issues, especially those at the interface of the economy and the environment, for nearly three decades. It cooperates extensively with other parts of the organization and with outside partners.

Two recent developments have provided a new impetus and direction to the Environment Program of work for which EPOC is responsible. First, the OECD Environment Ministers' Shared Goals for Action, which OECD Environment Ministers agreed on at their meeting in April 1998, promotes strong national policies and effective regulatory structures on the protection of the natural environment and human health. The goals are to:

- promote an integrated policy approach which encourages coherence among economic, environmental and social policies;
- strengthen international cooperation in meeting global and regional environmental commitments;
- strongly support participation, transparency, information and accountability in environmental policy-making by public authorities at all levels;

The second development is the implementation of a major horizontal program on sustainable development throughout the organization, which will culminate in a report to the OECD Ministers in 2001. Issues which will be addressed in this include climate change, the environmental impact of reforming subsidies and taxes, technology, the management of natural resources, governance for sustainable development, and indicators.

The integration of environment and economic policies, central to achieving sustainable development in OECD member countries and non-member countries alike, remains at the center of the Environment Program. It will promote compatibility and mutual reinforcement between economic and environmental policies, by: identifying market and intervention failures and how to remove them; analyzing policy instruments for achieving environmental objectives at least cost; and developing tools and strategies for integration in such sectors as agriculture, energy or transport. Key issues include the use of economic instruments, especially environmental taxes and charges, and the impact of subsidies on the environment.

The Environmental Outlook and Strategy is a major new activity. The outlook will provide an economy-based vision of environmental conditions in 2020 in key sectors and in relation to major environmental problems. It will be the basis for a new OECD Environment Strategy to be discussed when OECD Environment Ministers next meet in 2001.

Two new crosscutting activities in the program deal with *resource ef ciency* and *sustainable consumption*. Each is of growing importance for both OECD and non-OECD countries attempting to manage renewable and non-renewable resources more sustainably, and to in uence the ever rising demands for goods and services in such areas as transport, food, housing and tourism.

The Environment Program also examines the links between the globalization of trade and investment, and the environment. There is active engagement of non-member countries in policy analysis and dialogue on this and many other parts of the program. The Environment Directorate is also the secretariat for the Environmental Action Program Task Force, which is part of the Environment for Europe process.

On climate change, the Environment Program is contributing substantially through wide-ranging analysis in support of the development of domestic and international policy by Annex I parties to the United Nations (UN) Framework Convention on Climate Change. Work on environmentally sustainable transport, to explore the policies needed to achieve sustainability in this sector by 2020, is another element of the program.

Chemicals and biotechnology are also addressed, as is the management of the transfrontier movement of waste. The prevention and reduction of risks to health and the environment from chemicals, the harmonization of policies and regulations, and the promotion of integrated approaches to chemical testing, data and management are the main focus. Several OECD instruments, or Council Acts, deal with chemicals and waste management – as well as other environmental subjects – and member countries are committed to implement these.

Environmental data and indicators are essential to underpin the other work, and the OECD regularly publishes a core set of environmental indicators. It also measures countries' environmental performance with a program of Environmental Performance Reviews of all the OECD countries; this is nearing the completion of its first cycle. These peer reviews, in which other OECD member countries as well as the Secretariat are actively involved, have also been undertaken for several non-member countries in central and Eastern Europe, and most recently for Russia.

JEREMY EPPEL *France*

Optimum Sustainable Yield

see Maximum Sustainable Yield (MSY) (**Volume 4**)

Organization for Economic Co-operation and Development (OECD)

see OECD (Organization for Economic Co-operation and Development) (**Volume 4**)

Ozone Layer: Vienna Convention and the Montreal Protocol

Mostafa K Tolba
International Center for Environment and Development, Cairo, Egypt

The atmosphere has evolved over geological time into a mixture of some 80% nitrogen and 20% oxygen. It also contains many trace gases, in concentrations ranging from a few points of a percent down to a few parts per trillion. Their biochemical, geochemical and geophysical importance is much greater than their relatively low concentrations might suggest. This is particularly true of ozone, which is produced and destroyed at a wide range of altitudes in the atmosphere. Near the ground, ozone is a potentially harmful pollutant, formed when emissions from cars and industry react with sunlight. But between 25 and 40 km up in the stratosphere, ozone is formed naturally from molecular oxygen and is a blessing. Although its concentration is only about 10 parts per million by volume, it acts as a natural filter which absorbs and blocks the dangerous short-wave ultraviolet (UV) radiation (UV-B) of the sun.

Ozone exists in equilibrium in the natural stratosphere, being formed from molecular oxygen and being destroyed by UV radiation. Reactive chemicals, such as the oxides of hydrogen, nitrogen and chlorine, can accelerate the process of ozone destruction and thus upset the natural balance, leading to a net reduction in the amount of ozone. Each molecule of these chemicals acts as a catalyst, destroying many (even thousands) of molecules of ozone before being removed from the stratosphere through interaction with other free radicals, and sinking down into the troposphere.

Concern that human activities were destroying stratospheric ozone began in the late 1960s and concentrated on high flying supersonic aircraft. Their engines become so hot that they convert atmospheric nitrogen and oxygen into nitrogen oxides (NO_x) and deposit these 17 to 20 km up in the stratosphere. The NO_x then destroys ozone. But when plans to build fleets of commercial supersonic aircraft were abandoned, this concern faltered.

In 1974, two scientists working at the University of California, Professors Sherwood Rowland and Mario Molina, advanced the hypothesis that chlorofluorocarbons (CFCs) which are inert gases near the ground, gradually rise into the stratosphere, where they cause a catalytic reaction which destroys ozone (Molina and Rowland, 1974). As CFCs were used in homes and industry, for example, the suggestion set alarm bells ringing. Rowland and Molina subsquently became Nobel Laureates as a result of their 1974 paper.

Early in 1977, United Nations Environment Programme (UNEP) convened a meeting of experts, which adopted a 21 point World Plan of Action on the Ozone Layer and gave UNEP the responsibility for collating, disseminating and encouraging research. The UNEP and the World Meteorological Organization (WMO) established a Co-ordination Committee on the Ozone Layer (CCOL) made up of representatives of specialized agencies, national, intergovernmental and non-governmental organizations, and scientific institutions. This committee was to produce semiannual assessments of ozone layer depletion and its impact.

At this stage there was no prima facie case against CFCs. But some governments were already anxious. In 1976, Canada had announced that it would order the progressive elimination of the most common CFCs (11 and 12) and prohibit their use in non-essential spray cans. In 1978, the USA banned the use of CFCs as aerosol propellants. Some large chemical companies began to look into developing safe substitutes.

OZONE OBSERVATIONS 1970–1987

Statistical analyses of ground-based total-column ozone measurements show that between 1970 and 1980 there was no statistically significant decrease in the annual global average content of ozone (i.e., the decrease was less than 1%). Model calculations, taking into account increases in nitrous oxide and halocarbons, predicted that total-column ozone should have decreased by less than 0.5% at all latitudes in the summer, and between 0.5 and 1% in the winter, with the largest depletions predicted at high latitudes. Thus the model results were not inconsistent with the observations.

By 1986 three mechanisms (two involving changes in natural processes (1 and 2) and one involving human activities (3) had been proposed to explain the observations of ozone loss:

1. *Solar cycle mechanism*: periodic increases in the abundance of nitrogen oxides in the lower Antarctic stratosphere due to variations in solar output.

2. *Dynamic mechanism*: a change in the circulation pattern from downwelling of ozone-rich air from the upper stratosphere to upwelling of ozone-poor air from the troposphere.
3. *Halogen mechanism*: several different theories invoking the catalytic destruction of ozone due to CFCs and halons.

A ground-based study based at McMurdo, Antarctica, in 1986 showed that:

1. The abundances of nitrogen oxides were exceptionally low, which largely disproved the solar cycle mechanism.
2. The levels of long-lived tracers were such that a dynamical mechanism was unlikely.
3. The elevated levels of active chlorine species suggested that chlorine catalysis was the most probable mechanism.

In 1981, accumulating scientific information led the UNEP Governing Council to establish an *ad hoc* working group of legal and technical experts to work out a framework convention on the protection of the ozone layer; the group met four times between January 1982 and March 1985. At the first meeting (January 1982) it was recognized that although the threat was still distant and the issue shrouded in scientific uncertainties, there was an urgent need for an international effort to deal with it. Excerpts from the opening statement by Mostafa Tolba, then executive director of UNEP, provide an overview of the situation.

> We are meeting to lay the groundwork for a convention that some argue may not be needed, for we cannot say with certainty that the ozone layer is being depleted. However, all the most reliable scientific evidence points to the fact that the earth's protective ozone layer has been, is being, and, more importantly, will continue to be depleted by CFCs and other chemicals unless the international community takes preventive action. The urgent task of this meeting is to create a framework that will make that action effective.
>
> Less than a decade ago, the problem of ozone depletion caused by chlorine atoms penetrating the stratosphere was hardly known. Now, following substantial research and monitoring, we are aware of a potential pollution problem of a scale and consequence never before faced.
>
> The costs of continued CFC pollution are not primarily ours to bear. If scientific observations over the next few years turn the theory of ozone depletion into unchallengeable fact, then the hazard of increased UV light exposure due to ozone depletion is a legacy we will pass on to future generations. Not only Man but also most other living things are susceptible to UV-B exposure. Nor is the problem one of limited area. Atmospheric transport ensures that the risk is distributed to all parts of the globe.
>
> If ozone depletion does reach levels where it becomes measurable and at the same time CFC emissions remain at their present levels, then, because of the stability of these chemicals in the troposphere, their imperviousness to degradation, their ability to remain in the atmosphere for long periods, the absence of any significant sink, there exists a certainty of a major accumulation of CFCs in the stratosphere. There, by catalytic action of chlorine atoms, photochemical dissociation of ozone molecules is predicted to occur. Such is the life time of these chemicals in the stratosphere that even if future releases of CFCs were stopped or severely limited today, ozone depletion would probably still continue well into the next century.
>
> We are dealing then with a problem that has yet to be proven conclusively. It is one that is out of sight, and one that could so easily be out of mind. The problem of ozone is not qualitative but quantitative. Uncontrolled release of CFCs other than CFCs 11 and 12 and of chemicals such as methyl chloroform and carbon tetrachloride, which may also affect the ozone layer, obscure the issue and limit our ability to predict with any confidence the future of the ozone layer. And our imperfect understanding of latitudinal, seasonal and other natural variations compounds this problem.
>
> Although the possible consequences of ozone depletion will not be experienced until well into the future, perhaps beyond our own lifetimes, time is not on our side. We have to act now if we are to ensure that the more severe penalties from upsetting the ozone balance are never incurred. These may include consequences for agricultural production, fisheries and human health. Recent research shows that many terrestrial plants, including important crops such as wheat and rice, and aquatic organisms such as fish eggs and larvae, undergo damage when exposed to increased levels of UV-B. The link between exposure to solar UV-B and skin cancer is well established and there are indications that sunlight may also be a causative factor of malignant melanoma.

In retrospect it is impressive how many of the above speculations would be proven to be true. Chemicals causing ozone depletion did indeed include carbon tetrachloride and methyl chloroform, which would be included in the controlled substances list in 1990, and in the same year the suspected effects of UV-B were confirmed by the assessment panels.

THE VIENNA CONVENTION, MARCH 1985

The negotiations that began in 1982 brought together representatives of the governments of both industrialized and developing nations, spokesmen for the scientific community, and industry representatives, notably those of the Chemical Manufacturers Association (CMA); although non-governmental organizations (NGOs) and the media were also present, they did not initially play an important role.

From the beginning, nations formed groups advocating very different approaches. Western nations fell into two factions: The Toronto Group, comprising Canada, the United States, Finland, Norway, Sweden, and Australia, who favored a worldwide ban on the use of CFCs as aerosol propellants but opposed other CFC restrictions; and the European Community (EC) nations, who advocated eventual limits on total production but opposed cuts from current production levels. Developing countries were concerned about the potential of the convention to impose stipulations that might inhibit their own development. Some

nations were anxious to address the problem by any means necessary, whereas others, such as the United Kingdom, France and Japan, were reluctant to agree to any regulatory measures. Among the industrial representatives, the European Council of Chemical Manufacturers' Federations (CEFFIC) took the position that even with significant CFC usage the threat was only a distant possibility and could be met with some future action; in short, they opposed cuts in the production and use of CFCs.

In spite of these disagreements, there was a general acknowledgment that although scientific uncertainties would exist for some time; it was necessary to consider the consequences of awaiting complete certainty, whereas action taken now might prevent irreversible damage to the ozone layer (*see* Precautionary Principle, Volume 4).

Major differences among Western countries remained unresolved, and only a draft General Framework Convention stressing the need for more research was ready for adoption by the Vienna plenipotentiary conference in March 1985.

However, before the Vienna meeting could take place, news from the scientific community electrified the delegates. Joe Farman of the British Antarctic Survey announced that in comparing measurements taken in the Antarctic spring (September and October) of 1984 with records compiled since the 1950s, he had found that the stratospheric ozone layer over the continent had been in sharp decline since the late 1970s (Farman *et al.*, 1985). Since 1957 the austral spring ozone layer had thinned by about 40%; in effect, a hole had developed in the ozone layer over the Antarctic as big as the United States and as deep as Mt. Everest, spreading north over Argentina and New Zealand. After Joe Farman's paper had been published, National Aeronautics and Space Administration (NASA) reviewed its long-term data banks and discovered that they too showed a downward trend in stratospheric ozone over Antarctica: the analysis technique had been programmed to discard outliers.

The plenipotentiary conference convened in Vienna in March 1985, under the auspices of the UNEP and the WMO, adopted a treaty to protect the ozone layer. It committed its signatories to take appropriate measures to protect human health and the environment from human activities that potentially could have adverse effects on the ozone layer. The community of nations had decided not to wait for incontrovertible proof of cause and effect, which could come too late to avert irreparable harm, but to take action against a future threat, in what is the first global application of anticipatory and precautionary principles.

The conference also requested that the UNEP continue work on a protocol on ozone-depleting substances, and the UNEP Governing Council delegated this task to its executive director, at the same time authorizing him to convene a diplomatic conference, if possible in 1987, for the purpose of adopting such a protocol, clear evidence of the sense of urgency that prevailed.

THE MONTREAL PROTOCOL, SEPTEMBER 1987

Negotiation began in 1986 in a changed perspective on the issues involved. Scientific information continued to pour in that strengthened the case for an effective protocol, but economic factors had taken on additional importance.

In 1986 NASA, whose satellites had produced the spectacular infrared maps of the Antarctic ozone hole, estimated that a 3% increase in CFC emissions could deplete the ozone layer by about 10% by the year 2050. The US Environmental Protection Agency (EPA) said that if CFC emissions continued at recent rates, the resulting extra UV-b radiation could cause 40 million more skin cancer cases and 12 million more eye cataracts over the next century in the United States alone. At about the same time, Swiss scientists reported evidence of an Arctic hole over Spitzbergen, with the ozone layer thinning as far south as Switzerland. The distant threat was clearly coming nearer. Pressure for global regulations and limits on CFC emissions was growing.

Some governments, thinking action premature, resisted, but most governments, most scientists, and some manufacturers, unwilling to gamble with the composition of the atmosphere and the future of the planet, advocated action. Scientists were firmly in favor of controls, without which, they said, changes to the ozone layer would certainly increase skin cancer and would probably affect agricultural production, atmospheric pollution, and the climate. Before the political and legal negotiations began in Geneva in April 1987, the Toronto Group, led by the United States, changed its position and supported a global freeze on CFC production followed by a series of reductions that would lead to a complete ban. The EC countries wanted a cap on production but no cuts; the Soviet Union and Japan were reluctant to accept any control measures; the developing countries feared that the new protocol would hamper their industrial development.

Trade factors dictated much of the differences between the Toronto Group and the EC: European chemical industries saw US industries as being a step ahead of them in developing substitutes for the CFCs, which could endanger their markets. Because the United States had already banned the use of CFCs in aerosol containers, it was concerned that further controls would affect their use in refrigeration and foam production unless the measures were carefully drafted. Negotiations went on almost non-stop between April and September 1987.

When the final text was presented to the Plenipotentiary Conference in Montreal, the feeling of triumph was general. This was the first truly global environmental treaty;

moreover it dealt with an issue that posed a threat, not immediately, but far into the future, one that potentially affected everyone on earth today. It was a monument of collective action, a masterpiece of compromise. It had the advantages of ease of implementation, exibility due to its mechanism permitting adjustments to meet scientific, technological, and socioeconomic changes, and the clearly applied principle of common but differentiated responsibility (*see* Common but Differentiated Responsibility Principle (Stockholm/Rio), Volume 4). It was also the first treaty to set for itself, subject to conditions, a date for its own entry into force: January 1, 1989, barely fifteen months after the treaty had been signed. There would follow a number of meetings to fine-tune its details, but the date (September 16, 1987) went down as a landmark in the history of international negotiation.

Advances in Scientific Understanding after Montreal

There has been tremendous advance in scientific understanding of the ozone layer since Montreal. These advances, based on laboratory studies, special field campaigns, theoretical studies and re-assessment of trends in stratospheric ozone, were reviewed during the 1989 and 1991 international scientific assessments which were required by the Montreal Protocol. The key conclusions of these two international assessments are summarized in Boxes 1 and 2.

London, Helsinki, Copenhagen and Beyond

Six months after the Protocol was signed, a major report was released by the NASA-National Oceanic and Atmospheric Administration (NOAA) Ozone Trends Panel, demonstrating conclusively that human activities were causing atmospheric concentrations of chlorine to increase around the world. Even more disturbing was the finding that from 1969 to 1986, a small but significant depletion of the ozone layer, amounting to 1.7 to 3.0% depending on latitude, had already occurred over heavily populated regions of the Northern Hemisphere including North America, Europe, the Soviet Union, China, and Japan. The announcement made headlines around the world.

London, 1989

Media pressure for more concrete action was intense. Two months after the entry into force of the Protocol, and only a month before its contracting parties were to meet in Helsinki, British Prime Minister Margaret Thatcher convened an international ozone conference in London, in March 1989.

More than 120 governments participated in the London meeting, more than two-thirds of them at the ministerial level, a far cry from the meager two dozen governments represented at the start of negotiations in 1982 and the sixty in Montreal in 1987. More than 90 environmental organizations, as well as media from all over the world were there, all calling for a stern approach to the problem. Several countries announced in London that they would ratify the Protocol; developing countries reiterated their demand for assurance that financial resources and technology transfer would be forthcoming.

Helsinki, 1989

The deluge of scientific information that had preceded the opening of the Helsinki 1989 First Meeting of the Parties to the Montreal Protocol assured that stronger control measures would be enacted. The difficult areas were finance and technology transfer. China, supported by India and other developing countries, had proposed establishment of a special fund for the Protocol. Neither China nor India was a party to the Protocol, in fact only 31 of the contracting parties were in Helsinki, but they were accompanied by 51 non-parties, which showed the global importance attached to this issue.

Box 1 Key findings of the Scientific Assessment of Stratospheric Ozone: 1989

- *Antarctic ozone hole*: the weight of scientific evidence strongly indicates that chlorinated (largely Man-made) and brominated chemicals are primarily responsible for the recently discovered substantial decreases in stratospheric ozone over Antarctica in the Southern Hemisphere springtime.
- *Perturbed arctic chemistry*: while at present there is no measurable ozone loss over the Arctic comparable to that over the Antarctic, the same potentially ozone-destroying processes have been identified in the Arctic stratosphere. The degree of any future ozone depletion will probably depend on the particular meteorology of each Arctic winter and future atmospheric levels of chlorine and bromine.
- *Long-term ozone decreases*: the analysis of the total-column data from ground-based Dobson instruments shows measurable downward trends from 1969 to 1988 in the Northern Hemisphere (30 °N to 64 °N) in the winter months which cannot be attributed to known natural processes.
- *Model limitations*: these findings have led to the recognition of major gaps in theoretical models used for assessment studies. Assessment models do not simulate adequately Polar Stratospheric Cloud (PSC) chemistry or polar meteorology. The impact of these shortcomings for the prediction of ozone-layer depletion at lower latitudes is uncertain.

Box 2 Advances in the Understanding of Stratospheric Ozone: 1991

- *Antarctic ozone losses continue*: strong Antarctic ozone holes have continued to occur. Large increases in surface UV radiation have been observed in Antarctica during periods of low ozone. While no extensive ozone losses have occurred in the Arctic comparable to those observed over Antarctica, localized Arctic ozone losses have been observed in winter concurrent with observations of elevated levels of reactive chlorine.
- *Larger global ozone decreases observed*: ground-based and satellite observations continue to show decreases of total-column ozone in winter in the Northern Hemisphere. For the first time, there is evidence of significant decreases in spring and summer in both the Northern and Southern Hemispheres at middle and high latitudes (about 3.5% at 45°N in summer for the period 1979–1991) as well as in the southern winter. No trends in ozone have been observed in the tropics. The downward trends were larger during the 1980s than in the 1970s by about 2% loss per decade. The observed ozone decreases have occurred predominantly in the lower stratosphere.
- *Observed ozone losses due to industrial halocarbons*: recent laboratory research and an extended interpretation of field measurements have strengthened the evidence that the Antarctic ozone hole is primarily due to chlorine-(and bromine-) containing chemicals. In addition, the weight of evidence suggests that observed middle (and high) latitude ozone losses are largely due to chlorine and bromine. Therefore, as the atmospheric abundances of chlorine and bromine increase in the future, significant additional losses of ozone are expected at middle latitudes and in the Arctic.
- *Future levels of ozone depletion*: even if the control measures of the amended Montreal Protocol (London, 1990, were to be implemented by all nations, the current abundance of stratospheric chlorine (3.3–3.5) parts per billion volume (ppbv) is estimated to increase during the next several years, reaching a peak of about 4.1 pphv around the turn of the century. Since the mid-latitude ozone losses are apparently due in large part to chlorine and bromine, the increased level of chlorine and bromine that are estimated by the year 2000 are expected to result in additional ozone losses during the 1990s comparable to those already observed in the 1980s. There is also the possibility of incurring widespread losses in the Arctic. Reducing these expected and possible ozone losses requires further limitations on the emissions of chlorine (and bromine) containing compounds.

London, 1990

The Second Conference of the Parties in London in 1990 was attended, mostly at the ministerial level. Against all odds, the London meeting succeeded. Governments agreed to phase out CFCs by the end of the century and to include a series of other chemicals in the list of controlled substances. They agreed to establish a special fund to help developing countries comply with the Protocol and devised rules for contributions by industrialized countries as well as a management plan. Language was found to provide developing countries with alternative chemicals and technologies.

These amendments to the Montreal Protocol required ratification, and ratification requires time. So that valuable time would not be lost, the parties agreed by a simple decision to establish an interim fund of $240 million for 1991–1993, with allocations to be determined by size of country: $5 million to a small country, $20 million to a medium sized country, and $40 million to a big country, such as China or India. These financial incentives led China to join in 1991 and India to follow soon after.

Montreal was offered and accepted as the seat for the new fund. It is still there.

Copenhagen, 1992

At Copenhagen the Contracting Parties agreed to reach the 100% phase-out of the substances that deplete the ozone layer by 1996 and established the fund to replace the interim one.

Since then, the process is continuing to control other substances that deplete the ozone and to replenish the Multilateral Fund.

Lessons from the Montreal Protocol

The sense of urgency shared by the negotiators of the Montreal Protocol not only led to the conclusion of an international treaty through an innovative process, but also resulted in the establishment of a number of precedents. One was that once a two-thirds majority of the delegates had approved the adjusted control measures, they became binding on its signatories without the lengthy process of formal ratification. Another was the establishment of the Interim Multilateral Fund by a simple decision of the Conference of the Parties, although there was no legal basis for this step in either the Vienna Convention or the Protocol; the fund's establishment under such circumstances was considered by some legal experts to be an unprecedented step. The non-compliance provisions were similarly unprecedented. Other decisions that were made by consensus, which would have under normal legislative rules required years of formal negotiation, included the requirement that any production increase of controlled substances be on the condition that it was not for export. The data-reporting provisions were similarly streamlined to permit swift action.

An important principle was developed and crystallized in the Montreal Protocol and its amendments, that of *common but differentiated responsibility* (*see* Common but Differentiated Responsibility Principle (Stockholm/Rio), Volume 4). Although it had existed since the 1972 Stockholm Conference, it had not previously been incorporated into regulatory framework. This was a strong encouragement to the developing countries to remain whole-heartedly in the negotiating process.

Probably the most important lesson of the ozone treaty negotiations was the value of informal consultation, away from the negotiating table (*see* Industrial Responses to Stratospheric Ozone Depletion and Lessons for Global Climate Change, Volume 4). Although the negotiators' goal remains the same as in the formal sessions, because they are not committing their governments in these off-the-record conversations, they can be more relaxed and more open to recognizing the merits of other points of view in reaching common solutions and to making compromises. They become friends working for a common cause.

None of these factors, however, would have been enough to bring about these international environmental agreements had it not been for the basic change in national attitudes away from the primacy of national sovereignty and toward international cooperation. During the years of negotiations, the issue of sovereignty was never raised: each country was in a sense interfering in the internal affairs of all the others for safeguarding the environment of all.

There is still at this writing some work remaining to achieve the goal of protecting the ozone layer. Governments until now have been unwilling to strengthen the non-compliance procedures and to carry them out vigorously. Yet, it can safely be said that the legal obligations established under the Ozone Convention and Protocol are preventive rather than remedial. There is also cause for optimism: The two instruments were designed to be flexible and adaptable to changing conditions, and the mandated periodic assessments will assure their relevance and usefulness well into the future.

See also several articles in Volume 1 on the scientific issues involved, and in Volume 4 on the policy issues.

FURTHER READING

Assessment Chairs for the Parties to the Montreal Protocol (1991) Synthesis of the reports of the Scientific, Environmental Effects, and Technology and Economic Assessment Panels.

Farman, J C, Gardiner, B G, and Shanklin, J D (1985) Large losses of total ozone in Antarctica reveal seasonal ClO_x/NO_x interaction, *Nature*, 315, 207–210.

McFarland, M (1989) Chloro uorocarbons and ozone, *Environ. Sci. Technol.*, 23(10), 1203–1208.

Molina, M J and Rowland, F S (1974) Stratospheric sink for chloro uoromethanes: chlorine atom catalyzed destruction of ozone, *Nature*, 249, 810–814.

NASA (1992) *Interim Findings: Second Airborne Arctic Stratospheric Expedition*, Press briefing, NASA Headquarters, 3 February 1992.

UNEP (1987) *Montreal Protocol on Substances that Deplete the Ozone Layer*, UNEP, Nairobi.

UNEP (1989) *Environmental Effects Panel Report*, UNEP, Nairobi, 64.

Watson, R T, Geller, M A, Stolarski, R S, and Hampson, R F (1986) *Present State of Knowledge of the Upper Atmosphere: an Assessment Report*, NASA Reference Publication 1162, 20.

P

Participatory Integrated Assessment (PIA)

see Integrated Assessment (**Volume 4**)

Policies for Sustainable and Responsible Fisheries

John Caddy[1] and Henry Regier[2]

[1] *Imperial College, London, and CINVESTAV, Merida, Mexico*

[2] *University of Toronto, Toronto, Canada*

For millennia, human populations have changed planetary ecosystems; and hydrological basins and adjacent coastal seas have accumulated the consequences of human upstream alterations to terrestrial and aquatic ecosystems within these watershed or marine catchment basins. In addition to direct and indirect effects of shing, communities or assemblages of sh species must adapt to the aquatic effects on sh habitats of ecologically stressful human practices on land and along shore. Some sh species cannot tolerate these habitat changes. Climax assemblages of native species, or even the persistence of species components locally, cannot be sustained under intense stress to them and their habitat.

Some efforts to limit ecologically harmful practices date back many centuries. They are re ected in religious taboos, cultural mores, traditional knowledge and legal regimes. But as humans have became more numerous, and especially with industrialization, urbanization and impoverishment of shers, most attempts to limit adverse ecological effects have eventually failed. Contemporaneously, the stress regime on ecosystems generated by humans has grown inexorably. Degradation of aquatic ecosystems may not be apparent to the casual observer but can be inferred from sheries and other ecosystemic data.

Over recent millennia, humans have harvested sh, mostly for food for humans and their domestic animals, and to a much lesser extent for recreational and ceremonial purposes. When the demand for sh was less intense, shers selectively removed taxa that were highly valued locally and in trade, i.e., premium sh or target species. Taxa that were not valued for human food, i.e., coarse sh, were often discarded where they were caught, or transformed into animal feed. A main conclusion is that a major challenge in governance of sheries is managing the shers and not just the sh stocks.

CURRENT ISSUES IN THE GROWING DEPLETION OF WORLD FISH STOCKS

In recent decades, a progressively broader spectrum of fish species has been harvested by commercial fishers. New technologies in preserving and processing fish for human consumption, animal feed and pet food have resulted in increases in the value of incidentally captured species in the by-catch that were once discarded. Anglers have diversified to target a broader array of sport fish. Poor people, with access to fish near human settlements where diets are short of protein, may harvest all species almost unselectively.

The destructiveness to fish habitat of some fishing methods, such as trawling, has become increasingly evident. Ecosystem stresses from fishing have generally increased in phase with impacts on fish habitat from other human activities. Empirical study has shown that some fish species populations are suppressed and even extirpated by such a combination of stresses while other populations may react positively and increase. Both types of anthropogenic stresses, fishing and indirect impacts of human activities on fish habitat, may act synergistically, and there is often confusion as to proximate causes in any given situation. Qualitative and quantitative characteristics of a fish association or community are altered further by human-induced changes to the association of higher vertebrate populations (reptiles, birds and mammals) which depend on fish for their diet. Thus, humans have long acted in ways that

were unsustainable from both a fisheries and an ecosystem perspective.

From scattered evidence we infer tentatively that fish population extirpations occur more frequently through habitat destruction than through direct over-fishing, although extreme vulnerability, longevity and/or low fecundity exacerbate the risks of extinction by capture. The occurrences of fish population or stock extirpation, and much less frequently of species extinction, apparently increased progressively during the 19th and 20th centuries. The emphasis in management now seems less on the actual extinction of fish species than on the more immediate loss of genetic diversity through elimination of local stocks or races, especially for freshwater, estuarine and anadromous species, and those occupying specialized and restricted habitats such as estuaries, coral reefs or sea mounts. Nonetheless, currently few of the world's ecological associations of fish populations are not affected adversely by habitat abuses and/or improper fishing. Fishers can now take fish anywhere in the world's waters; for example, the recent ability to exploit economically extremely slow-growing, long-lived deep water species of oceanic slopes, ridges or sea mounts such as the orange roughy, is a deep concern. Given the few restraints that exist on overharvesting in international waters, many deep water stocks may be effectively eliminated before they can be studied.

Technological progress in industrial fishing of ocean resources has outstripped the ability of international regulatory bodies, hampered by inadequate powers allocated to them by coastal states, to counter the negative effects of overfishing in an environment where fishing power has grown continuously over recent decades. New technologies such as synthetic fibers for netting, echo location of schools, satellite navigation, and improved processing, transportation and marketing of fish, have all contributed to a frightening efficiency of harvesting and a greater range, versatility and speed of redeployment of fishing eets.

Atmospheric, hydrological and ecological processes carry hazardous chemicals everywhere, and especially to higher latitudes. Climate change is altering the current patterns and temperatures of water bodies. Stratospheric ozone depletion is changing the spectrum of the sun's irradiation at the water surface, particularly off the Antarctic coast. Exotic species are transported widely in the ballast water of ships and through the fish culture and aquarium trades. The genie of genetically modified organisms is also now out of the bottle as far as fish culture is concerned, with likely effects of escaped cage-reared salmon strains on reproduction of native stocks. In developed countries, local and regional ecosystemic stresses may be abating while global stresses are intensifying. In less developed countries all stresses may be expanding and intensifying.

PROSPECTS

So what might a new global commitment to sustainable and responsible fisheries mean in practice? Pessimists may fear that the sustainability concept relates to an historic ideal that has become progressively less realistic over the centuries. Late in the 18th century, Condorcet and Malthus initiated a debate that has continued since then between technological optimists and ecological pessimists. Two centuries later, the optimists may appear to be winning this debate; with each passing decade, natural features of ecosystems are less capable of satisfying the growing human wants and needs, and we are relying on increasingly artificial ecosystems using increasingly artificial techniques. Pessimists question if these artificial systems, with expanding loss of natural features, can exist independently of a healthy biosphere with the integrity of its natural life-support capabilities persisting unchallenged.

In the 1880s European fisheries experts debated whether or not serious over-fishing of oceanic fish populations did, or even could, occur. That debate was not resolved finally until the 1990s, when the collapse of several well-studied and supposedly well-managed groundfish stocks in the north Atlantic cast doubt on the efficacy of current risk-adverse type of science-based management regimes.

We note in retrospect that during the intervening 20th Century, much of the science and politics of fishing was captured by this debate about the reality of overfishing. Rational resolution may have been hindered because commercial interests and their client nations sought to maintain the freedom of the sea in order to exploit commercial opportunities. At the same time, scientific advisors working for governments placed too much emphasis on the concept of equilibrium conditions' and the mistaken idea that fisheries were sustainable' at levels at or around *maximum sustainable yield* (MSY) (*see* Maximum Sustainable Yield (MSY), Volume 4).

The politically weaker conservationists or ecological pessimists bore the onus of proof, in a case-by-case basis. Most management measures tended to be short-term, palliative, and fraught with problems due to unresolved differences about allocating catches between eets and states. In consequence, management was not sufficiently rigorous or sustained to lead to stock rebuilding. Powerful interests, i.e., exploitive technological optimists, dictated that fishing practices should not be constrained unless and until clear evidence of over-fishing was presented to, and accepted by the fishing industry. Exploiters often resisted reporting data that could prove useful to the conservationists, or reported false data on landings or their provenance.

In the 1990s a different debate was joined at a global level, with a revaluation of the proposition that commercial fisheries or other forms of exploitation of wild natural resource can be managed sustainably, and if so with what constraints? This debate has led to the extension of

the precautionary principle (*see* Precautionary Principle, Volume 5) from its original application in pollution science, to fisheries management. An attempt is now beginning to reverse the burden of proof to rest on exploiters rather than on conservationists, and require an assessment of fisheries impacts before authorizing new technological applications or increases in exploitation rate.

New interrelated global initiatives in ethics, law and science, applied to the shared use of fish and other ecosystem goods and services, began in the 1990s. There is as yet only scant empirical evidence that these noble commitments, which the two of us helped to foster, are being successfully put into practice. Optimists may argue that organisms that resemble wild fish, from a commercial perspective, can be reared economically in fish culture facilities. They admit that there will be adverse effects on the natural habitat, on the wild fish genome, etc. But it is better, they may say, to rear standardized, genetically-modified fish in abundance than to rely on stocks of variable wild fish that will likely fade away in any case. While the issues of fish culture, restocking and sea ranching are not addressed further in this essay, we have noted the dangers of genetic modifications to the fitness of wild stocks that they may pose. Among currently unresolved dangers of intensive and semi-intensive fish culture are: importation of diseases; local pollution of inshore environments and lagoons; and dependence on a limited supply of foodstuffs derived from harvesting wild stocks.

EARLIER SCIENTIFIC AND TECHNICAL APPROACHES

In this essay we mainly consider fisheries on fish (family Pisces) as such, but the aquatic realm provides somewhat similar opportunities and constraints to the ecological production, harvesting and relevant policies for other kinds of aquatic animals. Thus, much of our text is directly relevant to policies on harvesting of invertebrates and marine mammals.

During the 20th century a number of different approaches to population dynamics were employed for understanding, modeling and managing of fisheries. From roughly the 1950s onwards, these approaches were being applied in a rather widespread fashion to the management of commercial fisheries by industrialized nations. The primary focus was on the maximization of physical yield, using MSY as a target reference point sanctioned eventually by the Law of the Sea. Much later, and still inadequately, the focus shifted to the optimization of economic returns.

A problem with setting targets for catch or fishing rate is that fisheries data are rarely adequate to ensure a precision of greater than • 30% in optimal yield estimates, hence overshoots were frequent with subsequent recovery difficult. As a result, fishing effort tends to continuously increase, following the ratchet principle', aided by hidden effects of technological improvements not contemplated during the formulation of fisheries regulations. Much more recently, the specification and avoidance of the risk of overfishing has begun to be given priority in some fisheries, associated with the concept of precaution. As an example, the use of MSY as a limit reference point, rather than as a target, requires that managers take into account the variance in the data when seeking to avoid exceeding MSY, and prenegotiate an immediate effort reduction, should such an overshoot occur.

Within a particular ecosystem – river, lake, coastal sea, ocean – attention was focussed mostly on populations of the fish species favored by the fishers, with little concern for the inevitably high discards and other impacts on species caught incidentally by relatively unselective fishing gear. With commercial fisheries, such target species brought increasingly higher prices on an international market. Since the 1960s that market, in turn, has been increasingly efficient at linking hard-currency consumers anywhere, with fish resources harvested anywhere else in the world. Valued target species have shown an apparently unconstrained upward mobility in price, in turn increasing incentives for overharvesting. With ease of harvest and rudimentary regulatory systems, target species tended to be the first depleted in the local fish community. A bioeconomic model for a particular fish species might even have been fitted to data on catch and cost of fishing, but rarely was the concept of *economically optimal yield* used as a benchmark for determining harvest rate.

The abstractions from reality provided by simple single species models, that are often the best possible, given the limited information available, offer an illusion of certainty to fisheries managers. However, there is a growing awareness that aspects of the complex ecosystems in which the fish live, and of the societal and market complexities within which fishers operate, may better be described by chaos theory. Through time, a single species model's drift from realism is difficult to spot, and the inadequacies of information often means that decisions are made on yesteryear's information.

Numerous empirical generalizations or rules of thumb have been inferred from meta-analyses of population dynamics analyses over the years. Some examples are:

- The MSY of target species occurs when the total mass of the exploited fish population in the ecosystem is of the order of 50% of the biomass of that population in its historic pristine or virgin' state; unlike a virgin stock dominated by slow-growing older fish, an exploited population is dominated by more productive younger fish, posing a danger that excessive exploitation will cut into the population's spawning potential.
- The mean age in a population drops as fishing rate in a conventional fishery increases, so that the main

possibilities for regulating fishing are either postponing age at first capture by gear regulations (e.g., using larger meshes in the net), by avoiding fishing on juvenile nursery or adult spawning aggregations, and/or by reducing the fishing intensity, eet size and the individual fishing power of boats.

- As fishing intensity increases, longer-lived species (e.g., cods, halibuts, sturgeons) are replaced by shorter lived species (such as low-value forage fish and invertebrates), with consequent impacts on ecosystem balance.
- Species show an innate preferred temperature range; populations near the latitudinal mode of that temperature range are most productive while those at either the cold, high latitude or warm, low latitude extremes are smaller and produce less surplus for fisheries harvest perhaps due to irregular recruitment.
- In monetary units, net economic returns from commercial fisheries are usually maximized when harvests of individual populations are some 10–20% less than the MSY in mass units; this level of exploitation often corresponds roughly to the *optimal sustainable yield* or the *Maximum Economic Yield* in particular.

Current regulatory approaches often assume that the resource can be exploited throughout the year and throughout its range. Evidence is accumulating however, that such simplicity involves dangers because of the high seasonal vulnerability of some stocks due to a lack of effort control engendered by unresolved allocation issues. A more precautionary approach would be to start from the perspective that exploitation is prohibited, except for those time–area windows when established access holders are authorized to take a safe proportion of the stock, using selective, and tested harvesting methodologies with minimal discarding.

Because of the intractable unpredictability of nature and human activities that in uence exploited fish populations, in practice the population dynamics approach can mainly be used to construct accountancy models for fish currently in the waters, as a basis for allocation. Such models have little explanatory or predictive power with respect to what yield can reliably be expected some years in the future.

The population dynamics approach alone has seldom provided appropriate information directly relevant to the management of recreational fisheries. Generally anglers appreciate a natural setting and enjoy the prospect of capturing a large fish, at least occasionally. This implies that a low overall rate of harvest must apply in order to allow survival of that statistically improbable phenomenon, the large, older or record' fish. The gustatory pleasure and dietary benefit of actually consuming the fish may not now rank as highly as in Isaac Walton's day, and catch-and-release angling has gradually become more popular than angling for food. Aesthetic more than utilitarian considerations come into play with respect to considerations of sustainable angling pressure.

Artisanal or domestic fishers, who are often the poorer people of a region, tend to be opportunistic with respect to fish to which they have access. In North America, old treaties with some Native Peoples recognized usufructuary rights to harvest fish and other natural features in large areas beyond their immediate core reserves. Some of these treaties have been interpreted recently to imply that a Tribe is entitled to half the optimal sustainable yield of valued target species within its Treaty area. Often an artisanal fishery does not specialize just on a few valued species in its locale. Instead, fishers may remove fish of any size of numerous species. In many parts of the world a concept like sustainable fishing has seldom been applied to such artisanal fisheries. The poorest of the poor capture what they can where, when and however they can do so. Commercial and angling fishers may complain that artisanal fisheries are unsustainable, perhaps because some immature fish may be taken, but in general, artisanal fisheries provide more employment per ton of harvest, low wastage, and much of the catch goes for human food in local communities where fishers play an important economic role, as opposed to industrial eets which may not employ local labor, and where the owners are not necessarily vessel operators.

Aboriginal peoples who live in ecosystems that retain some features of wilderness may capture fish in ritualized ways for ceremonial purposes. In North America, for example, Aboriginal peoples who were not assimilated fully in Western culture have been re-instituting ancient ceremonial practices that relate to the spiritual interrelationships of humans with nature. Relatively few fish may be taken, so that there is no direct threat to sustainability. But ceremonial fishing may be deemed to exacerbate harm done by commercial fishers who are, or may be fishing such populations intensely. Or anglers may resent the removal of the scarce large fish valued both for recreational and ceremonial purposes.

Within an interdisciplinary ecosystem approach, there may often be a role for a population dynamics approach within a broader ecosystem perspective, but a scientist should take care to relate the more limited assessments to a more comprehensive ecosystem-based approach (*see* Ecosystem Approach, Volume 2). Sustainability of fisheries then becomes a facet of the joint sustainability of a mix of uses, both direct and indirect, of an aquatic ecosystem.

Integrated Coastal Area Management and watershed management are becoming more relevant, linking fisheries with other users of the aquatic environment through concepts such as zonation by priority use of subareas, and the employment of tools such as Geographical Information Systems now being widely used for planning user interactions and access rights (*see* Geographic(al) Information Systems (GIS), Volume 4).

A multi-stakeholder commitment to sustainability relates, in turn, to some politically agreed vision of selected features

of an ecosystem which are to be sustained. Any such vision will be tentative and subject to periodic revision as an ecosystem evolves in ways that are not fully predictable. Heretofore, few such shared visions have been negotiated and thus few are available to serve as case studies to guide further efforts to achieve joint sustainability. Any fishery of more than low intensity causes the ecosystem and the fish association to adapt irreversibly to some extent; thus a pristine state cannot provide a practical vision for a sustainable fishery. A debate between biodiversity preservationists and those favoring sustainable use in all waters continues, with the former slowly gaining in some countries through the adoption of marine parks or conservation areas in which commercial fishing may be banned.

OWNERSHIP REGIMES AND SUSTAINABLE FISHING

In many contexts, the management and transaction costs that come with private ownership of wild fish and their habitat may exceed the benefits to the owner of such ownership. In such cases, effective exercise of such ownership may not be worth the effort. Quasi-ownership regimes, involving a less formal balance of privileges and obligations, may be less expensive to administer. Common property resources may be formally owned in common by groups of people. The term may also be used loosely where access by fishers is open because no person or group holds formal ownership rights to the fish. The recent history of fishery policies is full of accounts of unsustainable exploitation of valued fish, in such open-access situations where no specific allocation of user access rights seems to mean that no duty of care for the resource is exercised.

A kind of Gordian knot is perceived by some fisheries analysts. It involves: both habitat and fish that are a commons; open access situations for fishers; fishing anywhere at any time of the year; unrestricted use of any kind of capture gear; rent-free harvesting; and willing consent of all harvesters prior to imposition of any constraints on extant fishing conventions and practices. This policy syndrome may seldom occur in full bloom in the real world, but is an interesting abstraction of what could happen where people have no commitment to intelligent joint action or to civic ethics.

Poorly-managed fisheries, with some features of the common property syndrome, may be complicated further by subsidies from governments to some fishers to help them to compete more effectively in such free enterprise. These subsidies, often rooted in political patronage, may be justified as efforts to assist poor fishers by supplying new technology to make their fishing more effective, or to reduce the costs of fishing. In fact, this may lead to further reduction of the level of fish abundance at which an economic breakeven point for fishing is realized, thus further increasing stock depletion. When external subsidies are not accompanied by compensatory constraints on fishery practices, they may well result in the intensification of existing unsustainable fishery practices, and promote further impoverishment.

Historically, many kinds of policy arrangements have emerged to guide the harvest of fish populations in situations where legal ownership, with a broad roster of enforceable rights, is not in place. In fact, in marine and some freshwater fisheries, ownership of a fish does not precede capture, and even where rights-based systems such as Individual Transferable Quota systems apply, it is more usual to speak of stewardship by the stakeholders over the resource, than ownership of a portion of the stock. As such, there may be few if any fish populations in the wild – other than in ponds, lakes and rivers on private estates – for which legal ownership of wild fish stocks can be invoked and enjoyed in practice. Ecosystems are by nature open systems, thus local, regional and global cultural stresses cannot be prevented from impacting at the local level. Thus, indirect effects of distant interests cannot be excluded from local ecosystems, however clear local legislation on exclusive ownership may appear to be.

Locally, some form of negotiation may lead to a balanced combination of privileges and obligations for each legitimate fisher and other users of the ecosystem. All stakeholders who enjoy some privilege of access are then constrained by some shared view of a preferred state of particular ecosystemic features, including those of the fish community or assemblage to be sustained. All local stakeholders' interests are also constrained by an ecosystem's subsequent responses to continuing regional and global stresses acting on it.

As is also the case with some mammals and birds, some fish species have populations that are strongly migratory with an annual cycle or space–time path. Policy issues, such as those concerning sustainable use, as related to highly migratory animals, are usually more difficult than those related to sedentary populations, since there are many points of harvest along the migration route. These are often in different jurisdictions or in international waters where legal constraints to overharvest are minimal. To limit harm to the habitat of highly migratory fish (as tunas and other large migrators are termed in the Law of the Sea), and to limit harvests to sustainable levels requires prior agreement among all those political entities scattered along the migratory path or with access to it.

Many inter-jurisdictional agreements have been negotiated over the decades to share sustainable yields of designated migratory stocks and to limit the harm done through abuse of habitats and bad fishing practices. Few of these agreements have served to protect the relevant stocks for long. In particular, enforcement of regulations in a multi-jurisdictional species range is a major weak point:

the negotiation of a sustainable *Total Allowable Catch*, and fair allocations from it are other obstacles where agreement is not easily reached.

In ocean waters outside of national extended economic zones (EEZs), neither the fish nor their habitats are owned by any group or nation. Currently the global regime of governance is not developed to a stage at which such ownership in the open ocean has been legitimized through fisheries commissions, special (e.g., bilateral) arrangements, or by the United Nations and its specialized agencies. A 1995 agreement on how the Law of the Sea provisions will apply to Straddling Fish Stocks (i.e., those lying across boundaries between EEZs and the High Seas), and Highly Migratory Fish Stocks, was not yet ratified at the turn of the century. Some measure of law and order with respect to hopes for sustainability may however be exercised jointly in specific areas through multinational or bilateral treaty arrangements among powerful fishing nations, expressed in the form of fishery commissions set up to provide for a discussion between member nations. Common provisions for monitoring, control and surveillance (MCS) and enforcement capabilities are rarely ceded by states to these commissions. Responsibilities for MCS and prosecution of fishers who offend outside the EEZs remain largely with the ag states of an offender's vessel. Non-tariff barriers with respect to trade are beginning to be used, but are not especially effective given the magnitude of unsatisfied demand for fish internationally. More recently, ecocertification with labeling of fish products from properly managed fisheries has begun, and is aimed at giving the consumer a role in choosing products from areas where healthy fish stocks are maintained by proper management.

SUSTAINABLE FISHERY REGIMES FROM AN ECOSYSTEM PERSPECTIVE

Every natural aquatic ecosystem together with its cultural system of human users is unique in some important ways. A universal analysis that ignores such uniqueness would not likely suffice for practical and political purposes within a particular ecosystem.

With ecosystems generally, intensification of some mix of ecological stresses has proceeded more rapidly than the creation of appropriate science to understand what then happens in particular ecosystems. Current efforts to innovate scientifically, with an ecosystem approach that takes account of all stresses, have mostly focussed separately on rivers, large lakes, enclosed coastal seas and shelf seas. Comparative studies or ecosystemic meta-analyses have been initiated within all four, with beginnings of studies interrelating all four types. The new science relies more on narrative and contextual case studies that have testable features, rather than on time-independent universalistic analyses also with testable features.

A favorite place for human settlements has always been along a river, near its mouth with a lake or sea. The waters near such settlements are often a particularly important habitat, at least seasonally, for valued fish species. With many exceptions, fishers have generally preferred to exploit large individuals of near-shore sedentary populations or of populations that performed predictable migrations through waters near settlements where they could be accessed easily during known seasons. Unless careful environmental stewardship was practiced, human settlements as onshore centers of cultural organization induced disorganization of ecosystems in adjacent nearshore waters. In terms of economic accounting, the environmental costs and consequences of settlements were externalized to the natural aquatic system and its human users. Such degraded waters have often been zoned tacitly as being closed to fishers, in part because of the high concentration of hazardous chemicals and pathogenic organisms in the fish esh. Increasingly strong programs to rehabilitate such locales are being initiated, with some successes.

One feature of the generally high fecundity of exploitable marine species is that a high level of attrition of juveniles naturally occurs due to predation, before the mortality of individuals drops to a much lower steady rate for adults. This means that a fish becoming large and mature even in a virgin population is a statistically rare occurrence, with probability of survival from egg to spawner of the order of one in a million in some cases. Thus, fisheries managers must make difficult decisions between: (a) harvesting juvenile fish (for which there is a good market in many countries) and conserving the few adult survivors; or (b) conserving the juveniles through closures of nursery areas and large mesh size and limiting harvesting of adults such that an adequate spawning stock survives. A fishing strategy which harvests all age groups intensively is rarely sustainable.

Centuries ago, fishing did not need to become very intense before some adverse effects of highgrading by discarding juveniles (which were dead in most cases) and by-catch became apparent, even in the absence of any noticeable habitat degradation. Fewer large fish were caught; annual recruitment of young fish failed in many years, more effort was needed to take the usual harvest, etc. If the fishing method involved barring the way to spawning habitat, then the population waned over the years due to inadequate recruitment.

Aristocrats or wealthy angler naturalists have traditionally acted to regulate commercial fishing locally and to protect the aquatic habitat in order to preserve near-pristine sports fishing opportunities. A common method was to have such habitat and riparian lands declared a preserve and to make capture of the preferred fish illegal other than in small numbers by the owners. Anadromous salmonids are prime examples of such conserved sports fish resources.

Nevertheless, of the uncounted thousands of anadromous salmonid stocks that existed three centuries ago, only a few percent have retained optimal sustainability.

Unintended habitat changes due to moderately intensive human activities in the ecosystem's hydrological basin or in the aquatic habitat itself may lead to increases in populations of small pelagic fish and other low-value forage fish, and thus to increases in their MSYs. The opposite may occur with large fish of demersal/benthic populations. With poorly managed or inappropriately regulated fisheries, fishing down the food web often occurs. Populations of apical predators – the large fish at the peak of the food chain such as groupers, pike, lake trout and halibut – were over-fished, with a reduction in annual catches leading to a reduction of predatory pressure and expansion of populations of smaller prey species, which in turn attracted fishing pressure.

The effects on an aquatic ecosystem of habitat degradation, together with overexploitation, interact in a common direction to produce a system dominated by lower trophic levels and small fish. This has often been labeled loosely with the code word *cultural eutrophication*, since the combined effects show some similarities to those caused by true eutrophication as with excessive nutrient runoff from land to aquatic systems. In turn, the combined effect of remediating a suite of cultural abuses has been termed *cultural oligotrophication*: an ecosystem is rehabilitated to a state showing some similarities to that which applied in pristine water bodies with low in ows of nutrients. Pristine oligotrophic systems contained clean water, low plankton density and dominance by apical or climax species in the food web (*see* Oligotrophic/Heterotrophic/Eutrophic, Volume 2). For both cultural eutrophication and cultural oligotrophication, a strongly artificial ecosystem may result, as for example in large lakes. Ecosystems have their ways of self-organizing beyond our full control. Thus the distinction between a culturally-stressed aquatic system and extensive aquaculture is blurred, now that some 30% or more of the products of photosynthesis in marine coastal waters eventually reach the human table.

An ecosystem adapts to an externally-applied stress regime, always. It may however be capable of self-organizing into more than one phase. Thus, the ecosystem may appear in an oligotrophic or eutrophic state, depending on the external regime of natural factors and cultural stresses. In such cases – in the past confined mostly to fresh waters, but now increasingly evident in coastal lagoons and semi-enclosed seas – the human community in the region may then take terrestrial and aquatic subsystems into account and opt politically for one of these states of nature. The stakeholders or managers might then formulate a qualitative/quantitative vision with benchmarks including preferred features of a past state of nature which are to be sustained in the face of the mix of legitimated human activities, that are themselves always evolving. They may use selected features of a fish association typical of the preferred state, e.g., oligotrophic or eutrophic, as an integrative indicator appropriate for assessing the state of sustainability.

During the 19th and 20th centuries, a common approach to mitigating the effects of habitat and fishery abuses has been to introduce selected exotic species that may have some innate tolerance of the ecological effects of extant ecological abuses. Or it may have been inferred that the reproductive capabilities of diminished native or introduced species have been severely harmed, which might justify the use of fish hatcheries to counteract such inferred effects. These may be perceived as examples of artificially upgrading a currently degraded fish association. This approach has sometimes been spectacularly successful, at least from a perspective of some anglers. Mostly it doesn't work very well, or has unexpected side effects.

The use of exotics and hatcheries has usually brought with it a new set of difficulties, some of which may be perceived to be disastrous. Some jurisdictions are now interpreting sustainability, in the context of rehabilitation from past abuses, as relating mostly to natural reproduction of valued native fish species, and are discontinuing reliance on exotics and hatcheries.

If the stress regime is extremely intense, the adaptation may involve systemic disintegration into a near-chaotic state of a few tolerant species, and ultimately death of all. The concept of sustainability as commonly applied in practice seems to presuppose that the ecosystem under consideration has not been forced to sacrifice many of its desirable features in adapting to the extant stress regime.

Current use of the sustainability concept also presupposes that the living part of an aquatic ecosystem can and does self-organize to a state of health or integrity, and contributes some degree of homeostatic capability with respect to the incessant uctuations characteristic of a stressed regime (*see* Monitoring in Support of Policy: an Adaptive Ecosystem Approach, Volume 4). In other words, sustainability implies that an ecosystem and fish association can take care of itself, to an important degree. Management costs can be constrained as a result of healthy ecosystem processes.

The shared vision of a preferred ecosystemic state, with its fish association, usually implies discontinuation of the more egregious ecological abuses and some rehabilitation to correct unacceptable effects of past abuses. In such cases it is this vision of a rehabilitated state to which a sustainability policy must relate and move towards through a long-term integrated-use policy. It may seldom be a matter of restoring a previous state, or preserving a current state which is already unacceptably degraded. The latter condition is still manifest in many parts of large lakes and coastal seas near industrialized cities.

Early in the 21st century there may be no fish population in the world that is not stressed by humans in one or more ways. Hence the concept of sustainability must always relate to a complex of natural and cultural processes. Relevant policies to foster sustainability must in turn exhibit an appropriate or requisite measure of complexity.

REFORMS OF THE 1990s

A global movement to reform human practices in our biosphere and be more responsible has been growing since the late 1960s. It was then that Sweden put in motion activities that led to the 1972 United Nations Conference on the Human Environment convened in Stockholm. Grossly unsustainable practices that degraded aquatic habitats and overfished valued stocks, though discussed at Stockholm, were generally neglected in the oceans until two or three decades after that.

In the 1990s, several initiatives by the United Nations or its specialized agencies have created a formal set of inter-related documents that is providing the conceptual framework for the emergence of the new global regime on fish and fishing. These include:

- the United Nations Conference on Environment and Development in Rio de Janeiro in 1992 from which came Agenda 21 with chapters on Aboriginal rights, biodiversity and its sustainability, the oceans and climate change;
- the United Nations Conference on Straddling and Highly Migratory Fish Stocks of 1995 which closed a major gap in the 1982 United Nations Law of the Sea Convention;
- the United Nations Food and Agriculture Organisation's Code of Conduct for Responsible Fisheries of 1995 with some six specialized appendices.

A strong theme in these initiatives was on the ocean but the agreements, suitably interpreted, are also relevant to closed seas, lakes and rivers.

For a particular ecosystem, although the transition has only begun, these reforms imply a change in emphasis from:

- maximal sustainable and independent use of each resource in the ecosystem by its own set of opportunistic stakeholders, to
- optimal responsible and shared use of the whole ecological web and aquatic habitat by an interacting network of sets of legitimate stakeholders.

Note that one aspect of this change is the shift in emphasis from *sustainability of use* to *responsibility of use*.

Responsibility by an ecosystem's users, including fishers, implies that each user and group of users bears an onus to take a precautionary approach to that use. A fisher's rights/privileges are balanced by a set of responsibilities/obligations to provide evidence *a priori* that the fisher's use will not induce ecosystemic changes that are impermissible with respect to the ecosystemic vision shared by all the ecosystem's users. During use, the fisher dutifully reports correct data relevant to accountability commitments and submits to sanctions for improper practices. This may be because the fisher is driven by considerations of intergenerational equity', and the wish to leave his children's generation an ecosystem that will provide the same livelihood, food potential and aesthetic pleasure as he has enjoyed. Or the fisher may be driven by deeper ethical commitments. From this it is evident that the search for sustainability inevitably takes us into the field of civics and ethics in a search for responsible behavior on the part of users, emphasizing the importance of this subject in future educational curricula from the earliest ages.

As a governance paradigm, the reform process is generating compatible information services and implementation features at all the levels in the nested hierarchy from individual users within groups of similar users, to the relevant level of formal governance within a nation, to a regional international commission.

Experts in international law were much involved in the 1990s' negotiations, together with experts in large-scale economic, ecological and technological systems. An emphasis is emerging on the methods of transdisciplinary discourse appropriate for negotiating and accountability processes, with respect to this new interactive set of international commitments. Versions of such appropriate discourse include: natural justice; communicative action; ethical discourse; post-normal participatory science (*see* Post-normal Science, Volume 5); and action research. A broadened scientific approach may still have a place for the reductionist population dynamics approach that dominated the earlier regime as it has for bioeconomic modeling, but that role is clearly subservient now. In fact, the application of games theory, developed by the military, to modeling competing fisheries makes it clear that the major challenge in governance of fisheries is managing the fishers and not just the fish stocks. Games theory also outlines the formidable obstacles in the face of reaching a sustainable harvest regime in situations where competition for resources occurs and allocations are not agreed to early on in the process.

Nonetheless, in recent decades, numerous innovators in the science and management of fisheries and other uses of aquatic ecosystems in various parts of the world have already begun local and regional reforms that have some of the features of the those mandated by the 1990s United Nations reforms. These include developing tools and concepts such as integrated coastal area management, large marine ecosystem approaches, appropriate scaling of human actions through marine catchment basins, and

ecosystem approaches in large freshwater basins. Some innovators are now meshing their own regional efforts within global initiatives to foster feedback mechanisms that might expedite implementation of the reforms.

Once proper management policies are developed, inefficiencies in traditional governance institutions are a serious impediment to the successful implementation of the policies. One fundamental weakness is that such institutions and their old mandates rarely encompass fully the real problem in all its dimensions. Fishery-induced and other anthropogenic causes are often intertwined in such a way that Ministries of Agriculture, Environment, Rural Affairs and others would have to cooperate in arriving at solutions which would then have repercussions for other Ministries, not least of which would be the Ministry of Finance.

People practicing governance in obsolescent ways may be offered few if any incentives to effect change. If an interdepartmental committee is struck to deal with fisheries, in the context of the health of aquatic environments, it may: be dominated by non-innovative persons; be given inadequate resources; be allowed an extended time frame that may invite postponement of urgent problems; and may eventually provide advice that is weakened with compromises. Another approach is to allocate rights to the private sector or to devolve responsibility to lower governmental entities such as provinces, municipalities and co-operatives. Here the incentive of ownership, and ready recourse to the courts that this implies, may provide a more prompt means of correcting imbalances. By following this latter course, the number of jurisdictional disputes may in turn become more frequent and intense.

In the international arena, the problem of shared stocks and their management has been assigned by the Law of the Sea to Fisheries Commissions and Arrangements. To date, states have rarely turned over to these bodies enough of their jealously guarded rights to allow them to conduct surveillance and exercise control effectively, especially with respect to allocations of catch. Multinational entities at the government–industry interface could be established with full powers to jointly manage shared stocks and national fleets, providing, for example, joint marketing services and dividends to the countries whose fleets are participating in the fishery.

The challenges to governance with respect to responsible fisheries – in freshwaters, in EEZs, in coastal marine waters, and in international oceanic waters – are daunting. Deliberate efforts directed toward quick learning would seem to be in order. Attractive incentives to innovate might speed such learning.

We fear that it may be decades before these 1990s responsibility reforms will be well in place in much of the world. Sustainability of some desirable ecosystemic features that will be extant at those times may then be more certain. Perhaps our responsible successors will then have transcended the centuries-long conflict between ecological pessimists and technological optimists.

Regretfully, the nature of fisheries for wild resources that will be conducted sustainably a few decades hence cannot now be predicted. Without a dramatic paradigm shift towards a revaluation of renewable resources through precautionary thinking, the prospect is for further ecosystem degradation and loss of biodiversity through human impacts. A necessary, but not necessarily adequate, precondition for avoiding further declines is that the fine words, principles and agreements developed laboriously in the 1990s be implemented swiftly.

FURTHER READING

Caddy, J F (1993) Towards a Comparative Evaluation of Human Impacts on Fishery Ecosystems of Enclosed and Semi-Enclosed Seas, *Rev. Fish. Sci.*, 1(1), 57–95.

Caddy, J F (1999) Fisheries Management in the Twenty-first Century: will New Paradigms Apply? *Rev. Fish Biol. Fish.*, 9, 1–43.

Caddy, J F and Cochrane, K (2001) A Review of Fisheries Management Past and Present and Some Future Perspectives for the Third Millennium, *Ocean Coast. Manage.*, (in press).

Hilden, M (2000) The Role of Integrating Concepts in Watershed Rehabilitation, *Ecosys. Health*, 6(1), 39–50.

Kay, J J and Regier, H A (1999) An Ecosystemic Two-Phase Attractor Approach to Lake Erie's Ecology, in *State of Lake Erie (SOLE) - Past, Present and Future*, Ecovision World Monograph Series, eds M Munawar, T Edsall, and I F Munawar, Backhuys, Leiden, The Netherlands, 511–533.

Pauly, D, Christensen, V, Dalsgard, J, Froese, R, and Torres, Jr, F (1998) Fishing Down Marine Food Webs, *Science*, 279, 860–863.

Regier, H A, Whillans, T H, Christie, W J, and Bocking, S A (1999) Over-Fishing in the Great Lakes: the Context and History of the Controversy, *Aquat. Ecosyst. Health Manage.*, 2, 239–248.

Policies for Sustainable Forests: Examples from Canada

Peter N Duinker
Dalhousie University, Halifax, Canada

In some parts of the world, forests were at one time the dominant land cover, but agricultural and urban land uses have

reduced forest area dramatically. Examples would include Ireland, and some countries in the Mediterranean area. Forest policies in such countries need to focus on maintaining what little forest remains, and on ways to increase forest area. In other parts of the world, forests still dominate the landscape. Examples include Canada and Russia. In these countries, forest policies are concerned perhaps less about the forest area and more about forest condition. Forest ecosystems have become degraded in many countries due to exploitive practices, and restoration strategies should dominate forest policies in these countries.

Forests are important to people in many ways. Humans value forests for a wide range of ecological functions that pertain to conservation of biodiversity, soil, water and air. We also value forests for their economic functions, many of which are associated with wood products like lumber, panels and paper, but also for other goods such as fuel, foods, furs, medicines and craft materials. In addition, we value forests for their social functions, including recreation, solitude, study, aesthetic satisfaction, existence, and others. In a word, forests are extremely important to humans around the globe. Forest policies need to address all these values.

The main theme in all forest debates and discussions today is sustainability; how can we manage and use forests so that their myriad bene ts are available for future generations just like they are today, or better? We speak of sustainable forestry, forest sustainability, and sustainable forest management (SFM). In this article we shall examine how forest policies embrace the concept of sustainability. First we shall explore the meanings of forests and policies, and then we shall look at some forest policies to discover their sustainability content. Most of the policy examples are taken from Canada, where about 10% of the world's forests (by area) exist.

POLICIES

Policies are written statements that purvey the intent of the authoring organization and guide the behavior of persons and groups under the authority of that organization. Thus, the natural resources department of a national government may issue a policy that describes its goals, objectives, principles and strategies for forests. That policy would guide the behavior of the government itself as well as the persons and organizations over which the government has authority (e.g., recreationists, wood-using industries, loggers and silvicultural contractors). Policies basically try to convey what the authors are trying to accomplish, and how the aims are to be accomplished.

Policy documents range from vision statements to detailed operational guidelines. For governments, policy documents may include high-level policy statements, legislation, and regulations and guidelines under the laws. Policy-oriented documents related to forests are of three types. One is a policy declaration, where the organization issues a policy that dictates its own intent and behavior regarding forests. Another is a policy proposal, where the authors suggest what the policies should be for other persons or organizations. Finally, there are policy analyses, where the authors lay out the costs and benefits of implementing alternative policies on the same theme or issue. In this article, the emphasis is on policy declarations.

FORESTS

Let us define forests as tree-dominated ecosystems. Ecosystems have no predetermined scales or boundary definitions; people define them in a way that is appropriate to their own needs. Where then do forests stop, and other ecosystems start? In 1993, Ontario's Forest Policy Panel wrestled with this question, and offered the following insights. We said that a forest is:

- a great expanse of trees as far as the eye can see; or, in more technical terms, a landscape of stands;
- a group of trees in part of the agricultural landscape: a woodlot, a windbreak and a soil stabilizer;
- trees in the urban environment, lining residential streets, in parks and as fragments of original forests in valleys, ravines and naturalized areas;
- the tree cover found in (northern) barrens ••• where the trees may be sparse and small but they are old, intrinsic parts of the ecosystem;
- all ages and types of trees: young seedlings to mature trees, natural and artificial regeneration; and
- more than just trees: shrubs, plants, animals, insects, microscopic creatures and soil are all part of a forest ecosystem.

Thus, forests are wherever trees are, and span the range from urban to rural, young to old, sparse to dense, fragmented to unbroken, degraded to pristine, very small to very large. It is important to develop policies for all conditions and kinds of forests (*see* Forest: the FAO Definition, Volume 2).

EXAMPLES OF POLICIES PROMOTING SUSTAINABLE FORESTS

Government Policies

Intergovernmental Examples

In 1992 in Rio de Janeiro, at the Earth Summit (or, the UN Conference on Environment and Development (UNCED)), several international cooperation agreements

relating to forests were signed. These include the UNCED Statement of Forest Principles, the UN Convention on Biological Diversity, the UN Framework Convention on Climate Change, and Agenda 21. Each one of these has a strong forest component. The Statement of Forest Principles, a non-binding document endorsed by many nations, strongly encourages countries to conserve forests by adhering to a broad set of guidelines on how sustainable forest use and management should be carried out. Recent international forest policy initiatives (e.g., the World Commission on Forests and Sustainable Development) have discussed whether to pursue a fully- edged international convention on forests.

The UN Convention on Biological Diversity makes little specific mention of forests, but we know that forest ecosystems around the world are critical in the overall theme of conserving biodiversity. This underscores the point that forest management can be significantly affected by policies on other topics. We also know that, were it not for steady and increasing pressures from the expanding human population, forests would cover a much greater proportion of the Earth's land surface. Therefore, we can be sure that, while the convention is equally applicable to ocean, grassland and desert ecosystems too, much of the stimulus behind its creation came from concerns about the incessant erosion of forest biodiversity. If the convention is actually implemented around the world in the way it is intended to be, great strides will be made to reverse the decline of forest biodiversity globally.

The UN Framework Convention on Climate Change is mainly focussed on ways to slow down the build up of greenhouse gases in the atmosphere. This, in turn, should slow down the rate at which the world's climate is expected to change as a result. Carbon dioxide, an inevitable waste product from combustion of fossil fuels, is the main target for reduction. Forests take up carbon from the atmosphere through photosynthesis, and release it to the atmosphere through respiration. Many people hope that changes in the way we manage forests will make them, on the whole, net sinks of carbon, which means that forests would take up more carbon from the atmosphere than they give back. Parties to the Convention are supposed to develop and implement national policies that address mitigation of climate change. Countries with substantial forest areas are now examining ways to manage forests specifically to sequester more carbon from the atmosphere.

Agenda 21 emerged also from UNCED as a major action agenda on worldwide sustainable development. It calls on governments to become aggressive about sustainable development and adopt appropriate policies. One of Agenda 21's many chapters addresses the problem of deforestation. The Agenda paints a broad picture of how to improve forest management, and calls on all governments to create national action programs for sustainable forest development.

The efforts discussed above were all initiated during the early 1990s. Significant additional efforts have been made since that time to advance global forest policies. Two are particularly notable. For one, after a series of hearings on various continents, the World Commission on Forests and Sustainable Development released a report in 1999 entitled *Our Forests, Our Future*. The report calls on all forest stakeholders immediately to implement SFM, and to take bold political decisions and develop new civil society institutions to improve governance and accountability regarding forest use. The other initiative is that of the Intergovernmental Forum on Forests (IFF), set up by the UN Commission on Sustainable Development to succeed the Intergovernmental Panel on Forests (IPF). Both groups have been working toward comprehensive new understandings on how to achieve forest conservation worldwide, and have advanced global thinking tremendously on the need and prospects for an international, legally binding instrument (e.g., convention) on SFM.

National-level Examples

Each country has its own unique way of creating and implementing national forest policies. In some countries, forest policy is a main preoccupation of the national government (e.g., Sweden, Finland), but in other countries, many forests are under the control of sub-national governments such as states or provinces (e.g., Canada, US, Germany). Thus, forest policies vary widely among nations. We shall focus on Canada as an example.

Canada is a federation of ten provinces and three territories. The constitution puts publicly owned forests, which comprise about 94% of the forest area in the country, mostly under the control of the provincial governments. However, Canada has a national forest service (the Canadian Forest Service, or CFS), and implements policy for matters in which the national government has a clear role (e.g., information systems, international affairs, foreign trade, national forest land, and so on). Perhaps more visible, though, is the federal government's role in assisting forest stakeholders pursue SFM. The Canadian Council of Forest Ministers (CCFM), comprising ministers responsible for forests from the federal, provincial and territorial governments, is particularly active in this regard.

The CCFM recently facilitated renewal of both the Canada Forest Accord and the National Forest Strategy. The Accord is a high-level statement containing a vision, a goal, and some principles and commitments to action, which must be adhered to by all its signatories. To give a avor of the Accord's content, here is the goal statement:

> Our goal is to maintain and enhance the long-term health of our forest ecosystems, for the benefit of all living things

both nationally and globally, while providing environmental, economic, social and cultural opportunities for the benefit of present and future generations.

The Accord commits its signatories to prepare public and measurable action plans . These plans are guided by the Accord's sister document, the National Forest Strategy. There have been four such strategies in Canada, the latest produced in 1998 through a stakeholder consultation process. The Strategy calls on the forest sector stakeholders to make substantial progress in improving conditions related to nine strategic directions:

- forest ecosystems;
- forest management;
- public participation;
- the forest industry (meaning mainly the wood-using industry);
- forest science and technology;
- communities and the workforce;
- Aboriginal peoples;
- private woodlots; and
- global commitments.

Accord signatories, which include governments and a wide range of NGOs, have produced action plans that address each of these nine themes, according to their respective jurisdictions and interests.

Sub-national Level

We shall focus here on Ontario, a large province in central Canada, which is home to some 11.5 million people. Ontario has some 58 million ha of forest, 88% of which is owned by the provincial government.

Substantial changes were made to Ontario's forest policies during the 1990s. All these policy changes have focussed on making the forests and the forest sector more sustainable. In 1994, the Ontario Ministry of Natural Resources adopted a policy framework for sustainable forests. The framework includes a goal for the province's forests, much like the national goal in the Canada Forest Accord. It also provides guiding principles and strategic objectives for sustaining: (a) forest ecosystems (focussing on biodiversity, protected areas, and conservation of soil, water and air); (b) forest uses (focussing on employment, wood supplies, non-timber products, forest investments and revenues, and tourism); and (c) forest social values (focussing on cultural and spiritual values, recreation, and knowledge and understanding). In addition, the framework provides principles for forest decision-making, and promotes ecosystem management as the dominant management approach.

Hot on the heels of the framework came new legislation to govern the management of Ontario's public forests. The Crown Forest Sustainability Act of 1994 replaced the old Crown Timber Act. The new Act repeated some key principles from the Framework, as follows:

1. Large, healthy, diverse and productive Crown forests and their associated ecological processes and biological diversity should be conserved.
2. The long-term health and vigour of Crown forests should be provided for by using forest practices that, with the limits of silvicultural requirements, emulate natural disturbances and landscape patterns, while minimizing adverse effects on plant life, animal life, water, soil, air and social and economic values, including recreational and heritage values.

Major changes embodied in the Act included establishment of trust funds to ensure adequate funding for regeneration silviculture for each forest management unit. Previously, money for regeneration activities was granted annually from the provincial legislature, and forest management was usually a poor cousin in negotiations where health care, education and roads to compete for money. The Act also established a number of regulations, themselves to be considered as policy documents. Four key manuals are regulated under the Act: (a) forest management planning manual; (b) forest information manual; (c) forest operations and silviculture manual; and (d) scaling manual. The forest management planning manual was issued in 1996, and compared to its predecessor of 1986, requires extensive stakeholder consultations and rigorous technical analyses of forest sustainability.

Non-government Policies

International Level

Various international NGOs have their own forest policies. A well-known example today is the Forest Stewardship Council (FSC). The FSC was founded in Toronto in 1993, and has headquarters in Mexico. Environmental NGOs, aboriginal groups, consumer organizations and independent professional foresters are prominent among the members. The mission of the FSC is to support environmentally appropriate, socially beneficial, and economically viable management of the world's forests. The FSC, operating as a kind of certifier of certifiers, has created an eco-label that can be applied only on products made of wood coming from forests certified by an FSC certified auditor. The criteria applied by the auditor must be consistent with the FSC Principles and Criteria for Forest Stewardship. There are ten principles, each of which is amplified by some 4 to 10 criteria. The 10 principles cover the following topics:

1. compliance with laws and FSC principles;
2. tenure and use rights and responsibilities;
3. indigenous peoples' rights;
4. community relations and worker's rights;

5. benefits from the forest;
6. environmental impact;
7. management plan;
8. monitoring and assessment;
9. maintenance of natural forests;
10. plantations.

In essence, the FSC is trying to improve forest management throughout the world using a market-driven mechanism. The hope is that consumers will prefer to buy goods stamped with the FSC eco-label, even if it means higher prices. At the moment, the FSC is the only operational eco-labeling scheme for forest products available around the world. Regional international schemes are also developing, such as the Pan-European Forest Certification initiative.

National Level

In most countries there are national forest-related NGOs which develop and declare forest policies of their own. For example, there is a forest certification scheme operating nationally in Canada. Late in 1993, representatives of the Canadian forest products industry requested the Canadian Standards Association (CSA) to undertake development of a system standard for SFM. Using a consultative, consensus-seeking process, the CSA developed the required standard to be consistent with the ISO 14 000 standard series on environmental management systems.

The standard is voluntary, and focuses on the following four components: commitment; public participation; management system; and continuous improvement. The management system has five elements: preparation, planning, implementation, measurement/assessment, and review/improvement. The CSA standard requires forest managers to forecast and monitor all indicators chosen locally to represent the status of 21 themes associated with the six nationally accepted criteria published in 1995 by the CCFM:

1. conservation of biological diversity;
2. maintenance and enhancement of forest ecosystem condition and productivity;
3. conservation of soil and water resources;
4. forest ecosystem contributions to global ecological cycles;
5. multiple benefits to society;
6. accepting society's responsibility for sustainable development.

Any forest management agency or company wishing to become registered as compliant with the CSA standard adopts the standard as its operating policy, and is required to develop specific and detailed policies of its own for SFM.

Sub-national Level

Examples of sub-national forest NGOs in Canada include the Ontario Forest Industries Association (OFIA), the Council of Forest Industries of British Columbia, and the Nova Scotia Forest Products Association. As an example partly in response to the stimulus of the 1992 National Forest Strategy, the OFIA created its Guiding Principles and Code of Forest Practices. The document is binding upon all members of the organization, which represents about 20 firms which, collectively, account for most of the industrial wood processing in Ontario. Each year, member companies undertake an audit of their performance against the principles and the code, and report confidentially to the OFIA. Summary reports are made public. While some companies may find the principles and code redundant with their own internal policies, others have made significant improvements in their management practices because of their required compliance with the document.

WHAT TO LOOK FOR IN POLICIES FOR SUSTAINABLE FORESTS

With continued improvements in public participation in forest decision-making, more and more people are becoming involved in looking at forest policies and discussing ways to improve them. When people get involved like this, what should they seek in a good policy for sustainable forests? Here is a generic list, drawn from a combination of forest policy textbooks and examples of up-to-date strong policies:

- a firm grasp of the main forces driving change in forest ecosystems and the forest sector; these forces either constrain or give opportunity to the fulfillment of forest values; examples are demographic changes, climate change, globalization of trade;
- a comprehensive goal or vision statement, consistent with modern thinking and putting priority on forest sustainability;
- a broad array of forest values with a balance among social, economic and ecological themes, and each with its own strategic objective;
- a broad definition of forests to include most or all situations where trees are a dominating feature of the ecosystem;
- a small but strong set of broad principles to guide people's thinking about and behaviors toward forests;
- a firm recognition of how forest policies in uence policies on other topics (e.g., agriculture, water, wildlife), and how other policies affect forests;
- a package of regulations and incentives that promote forest sustainability on private lands;

- innovative arrangements for private use and management of public forest resources that ensure conservation of ecosystems and resources;
- a firm commitment to the use of protected areas as a key element of biodiversity conservation; and
- requirements for effective and efficient public participation in forest decision-making at all levels.

CONCLUSIONS

Forests have faced tremendous pressures down through the ages as humans have exploited them and, in so many instances, degraded them. The prospects for forests in the 21st century may be grim, with ever-increasing pressures to be faced. One main pressure is the continuing rise in the global human population. Each new person born will place demands on all the things that forests can offer to support a good quality of life. Another pressure is our constantly expanding expectation on forests; the more we know about them and interact with them, the more we want forests to do for us. Finally, global atmospheric change may put tremendous stresses and strains on forests as pollutants continue to be pumped into the air and as the climate continues to change.

We should realize, though, that these undesirable trends could be reversed. However, this depends on whether nations, international governing institutions, companies, NGOs and citizens will mobilize their collective wills to implement forest policies that can help forests survive the continuing pressures they face. We certainly have the know-how and technical capacity, not to mention the rhetoric; do we have the political will too?

See also: Global Forest Watch, Volume 4.

FURTHER READING

Aplet, G H, Johnson, N, Olson, J T, and Sample, V A (1993) *De ning Sustainable Forestry*, Island Press, Washington, DC.

Cubbage, F W, O'Laughlin, J, and Bullock, Jr, C S (1993) *Forest Resource Policy*, John Wiley & Sons, New York.

Ellefson, P V (1992) *Forest Resources Policy: Process, Participants, and Programs*, McGraw-Hill, New York.

FAO (1997) *State of the World's Forests*, Food and Agriculture Organization of the United Nations, Rome.

Hummel, F C, ed (1984) *Forest Policy: a Contribution to Resource Development*, Martinus Nijhoff/Dr W Junk, The Hague, Netherlands.

Maser, C (1994) *Sustainable Forestry: Philosophy, Science, and Economics*, St. Lucie Press, Delray Beach, FL.

Murray, C H (1993) *The Challenge of Sustainable Forest Management: What Future for the World's Forests?* Food and Agriculture Organization of the United Nations, Rome.

Repetto, F and Gillis, M (1988) *Public Policies and the Misuse of Forest Resources*, Cambridge University Press, New York.

World Commission on Forests and Sustainable Development (1999) *Our Forests ••• Our Future*, Cambridge University Press, Cambridge.

Policies that Promote Sustainable Energy Futures

José Goldemberg
University of São Paulo, São Paulo, Brazil

Fossil fuels account for approximately 80% of human consumption of energy. Such dependence is unsustainable due to: (1) expected ultimate exhaustion of resources; (2) the impurities they contain; (3) emissions of CO_2. The Framework Convention on Climate Change (FCCC) might be an important driver in reducing fossil fuel use and thus open the way to a more sustainable energy future.

Energy policies to reduce local and regional environmental impacts have been in effect for a long time. Local pollution has to do with local governments, since it deals with clean air, fresh supplies of clean water, the removal and disposal of solid wastes and liquid ef u-ents, street cleaning, etc. However, in many developing countries, a large fraction of the population still lives among the rubble and residues that it produces, due to the lack of resources to remove waste and to build sewers and engineering works needed for the supply of water.

Regional pollution is caused mainly by automobiles, energy production and heavy industry which are inherent to more prosperous societies. Large cities and adjoining areas, such as Los Angeles, Mexico City and Sao Paulo, have long been suffocating under the pollution caused by the emissions and smog resulting from the burning of fossil fuels. Sometimes the amount of pollution produced is large enough to cause regional and even transborder problems, such as acid rain, some of which originates in the US but is responsible for the destruction of life in Canadian lakes. The same has happened to lakes in Scandinavia, due to industrial activities in surrounding countries. Regional pollution has to be dealt with at the state or national level and eventually among a number of countries.

Global environmental problems and concerns are more recent, the most conspicuous of them being the destruction of the stratospheric ozone layer by chloro uorocarbons (CFCs) and the greenhouse effect.

These problems result from changes in the composition of the atmosphere and have little to do with national borders. The causes of such global problems are gases originating anywhere in the world and the problem can only be tackled at the international level.

The ozone layer problem was successfully resolved by a concerted effort of a few countries which led to the Montreal Protocol. Subsequent decisions are effectively leading to the phase-out of CFCs from the face of the Earth. Other substances have been developed to replace them.

The outstanding current environmental issue is the greenhouse effect, which is proving to be much harder to solve since the most important greenhouse gas is carbon dioxide (CO_2), the unavoidable result of the combustion of fossil fuels, which is responsible for 80% of all energy consumed in the world today.

The basic policy in this case is to reduce CO_2 production through improved efficiency in energy use and to search for proper replacements such as renewable energy sources.

Appreciable success has been achieved in industrialized countries to reduce CO_2 emission mainly through rationalization in the use of energy. CO_2 emissions in those countries have leveled off since 1980 or are growing very slowly. In developing countries, however, emissions are growing at a fast rate.

The main policy instrument to solve this problem is the FCCC signed in 1992 during the Rio Summit and the Kyoto Protocol (1997), which adopted a binding resolution according to which industrialized countries, as a whole, would reduce their CO_2 emissions by 5.2% below the 1990 level in the period 2008–2012.

The US accepted a reduction of 7% to their 1990 level. Since emissions have been growing by approximately 1% since 1990, the reductions required would be much higher. Some estimates indicate that the reduction required will be approximately 25%.

In order to make it easier for industrialized countries to reduce their emissions, the Kyoto Protocol, in addition to setting caps on emission for the industrialized countries, has introduced features that could induce industrialized and developing countries to cooperate in reducing CO_2 emissions.

These proposals are the exibility mechanisms by which emission reductions in developing countries (or economies in transition, i.e., the former Soviet Union and countries of East Europe) can be counted as reductions in industrialized countries. In this category are emissions trading (ET), Joint Implementation (JI) and The Clean Development Mechanism (CDM).

ET is the more general, market-based instrument. It enables achievement of limits at least cost by taking advantage of marginal cost differentials in emissions abatement among countries. Since greenhouse gases are uniformly mixed at the global scale, both the damages from emissions and the benefit from emission reductions are independent of their origin. In order to minimize the cost of given global emission reductions, abatement should take place where the costs are lowest. This instrument is restricted to trading among the Annex I (industrialized) countries.

JI limits emissions trading to developed countries that have accepted emission reduction commitments under the Kyoto Protocol, as listed in its Annex I. Countries that are certified to be below their national limits are allowed to sell greenhouse gas offsets to other Annex I countries who are above their own limits. The latter's motivation is the opportunity to meet their commitments at lower cost than through purely domestic actions.

CDM is an instrument for joint action on emissions reduction (and therefore trading) between the developed countries listed in Annex I of the Kyoto Protocol and developing countries, which are not listed in Annex I. Article 12 of the Kyoto Protocol was the result of a political compromise. It conciliated two sharply opposing views that almost wrecked the Kyoto Conference in 1997. For some countries, an essential element of the Kyoto Protocol was emission trading based on adoption of mandatory reduction commitments by all countries. This feature was contained in a United States Senate Resolution, but was unacceptable to the developing countries. The creation of the CDM emerged as the solution.

In Annex I countries, there will be competition among the several mechanisms defined in the Kyoto Protocol to meet emission reduction requirements. The portion of resources allotted to each of them will prove to be an important internal problem to be resolved. The Kyoto Protocol left undefined what part of the emission reduction commitments could be met outside national borders through use of the exibility mechanisms. A small percentage would encourage domestic action in Annex I countries, reducing the financial ow to developing countries. A larger one would discourage domestic actions and eventually tend to lead to an inordinate increase of emissions in Annex I countries. In all likelihood, the market will determine the share for the CDM compared with other mechanisms.

The key industrial countries have not yet ratified the Kyoto Protocol, but – if and when they do – it might prove to be as effective as the Montreal Protocol in solving one of the most difficult global environmental problems of our time.

REFERENCE

Skinner, R (1994) Effects of CO_2 Reduction Policies on Energy Markets, in *The Economics of Climate Change*, Proceedings of an OECD/IEA Conference, OECD, Paris.

Policies to Achieve Sustainable Agriculture

John Lingard
University of Newcastle upon Tyne, Newcastle upon Tyne, UK

Sustainable agricultural development has the goal of ensuring adequate levels of current farm production, whilst preserving the land and natural resource base over time in order not to compromise or reduce development opportunities for future generations. It conserves land, water, plant and animal genetic resources, is environmentally non-degrading, technically appropriate, economically viable and socially acceptable. Agenda 21 of the Earth Summit in Rio in 1992 called for the development of sustainable agriculture and land management and the establishment of the systems necessary to monitor its achievement. Since then, this has led to a wide range of activities to de ne sustainable agriculture, land management and forestry and to develop monitoring methodologies, sustainability indicators and thresholds (Pretty, 1995).

The linkages between sustainable agricultural production and the natural resource base are strong, complex and, as yet, poorly understood. Economic constraints and government policies are important impediments to the adoption of sustainable agricultural practices, particularly in LDCs where the pressure to meet population-induced, food supply demands is likely to hinder the development of sustainable agriculture. Better methods must be found to measure and monitor the natural resource base in order to avoid counterproductive policies and then to design policy packages which are sustainable in the face of ever-increasing population and food demand pressures. Policies for sustainable agriculture must ensure the efficient use of available resources, develop sound and far-reaching investment strategies, and address the technical problems of natural resource degradation within a framework of appropriate national and international institutions. It won't be easy!

Non-sustainable agricultural practices, brought about by short-term mismanagement and policies, can result in a dramatic deterioration in the potential productivity of soils with a resulting decrease in the technical efficiency of inputs applied to the soil to make it more productive. As a consequence, there will be a decrease in the exibility of farming practices, a limit on innovative cropping and farming systems, and an inevitable increase in the costs of food production. Non-sustainability and natural resource degradation result from the interactions between economic and biophysical systems and are largely brought about by human behavior, directly by farmers and land users, and indirectly by policy makers and traders who partly determine prices in domestic and international markets.

Changing land use and agricultural practices worldwide are shaped by economic forces and demographic pressures, and often, non-sustainable systems may result. As populations and food demands grow, many farmers in the Third World are often forced to cultivate smaller and smaller plots, where the soil eventually becomes depleted, or they expand onto marginal lands: fragile hillsides, semi-arid areas and cleared forest land. India, with a recent annual population growth rate of 2.1%, a *per capita* income growth rate of 1.8% per annum and an income elasticity of demand for wheat of 0.5, must increase its wheat production by 3%, each and every year, if wheat prices are not to rise and threaten food security.

Biophysical land degradation and non-sustainable practices have their roots in household and social behavior. Access to the capital resources of human, natural, physical and financial capital by individual households acts as a constraint on their behavior as they seek to maximize incomes or defend their livelihoods. Cropping patterns, livestock enterprises, agro-forestry, aquaculture and other livelihood activities, and the associated soil, water and natural resource-using practices are the outcomes of decision-making of farm households attempting to earn an income and living. Often, these are made in response to market prices of agricultural inputs and outputs. In fragile environments in many resource-poor, less developed countries (LDCs) throughout the world, the economic pressures on households for merely short-term family survival are paramount. There, farmers degrade resources, not out of greed and ignorance, but out of short-term necessity. Peasant producers are often only loosely connected to poorly functioning, distorted and incomplete markets. There are few immediate rewards to such farm households for enhancing soil resilience, ecological sustainability and biodiversity. The interests of global society and individual households may thus diverge due to short-termism and market failure (Syers and Lingard, 1996).

AGRICULTURAL POLICY

Agricultural policy is concerned with the methods used by governments to change the decision-making environment within which agricultural production takes place; e.g., by altering farm input and output prices, introducing new institutions and regulations (e.g., food safety, animal welfare or environmental standards) or by promoting novel technologies. Sometimes policies are further sub-classified as price, marketing, input, credit, mechanization, irrigation, land reform and research policies (Ellis,

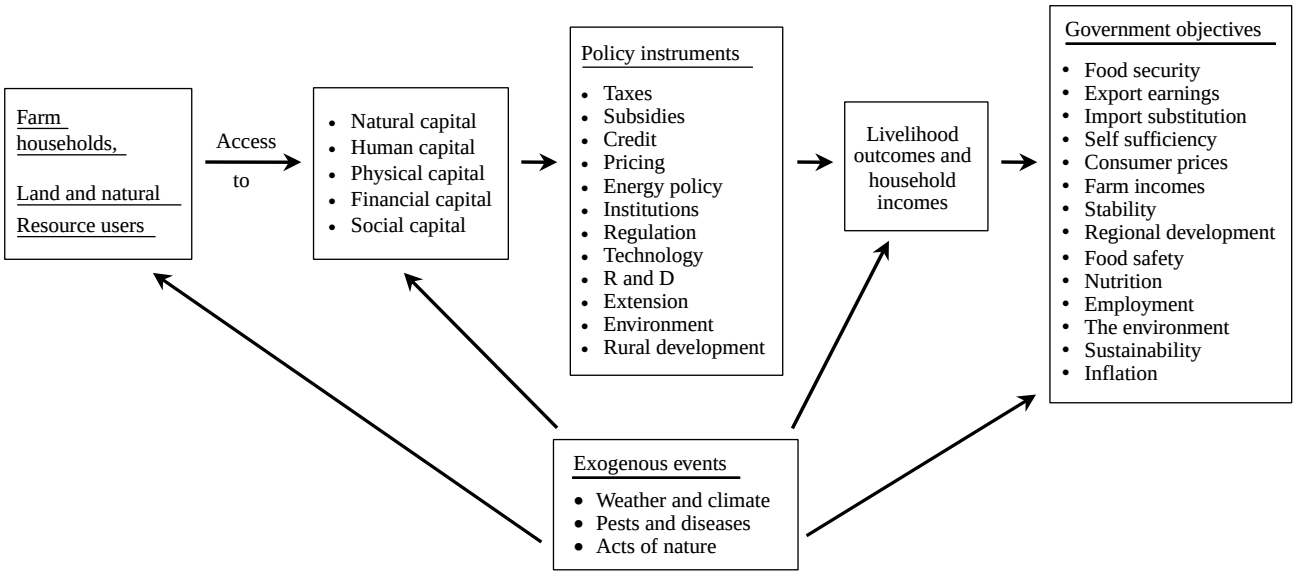


Figure 1 A schematic representation of agricultural policy

1992). Recently, a range of agri-environmental policies has been introduced in many, more af uent developed countries.

Figure 1 illustrates one form of the agricultural policy process whereby stated government goals, targets or objectives are achieved via the selection of the best or most appropriate policy instruments which cause farm households to modify their behavior subject to the available resource constraints and exogenous events over which the policy process has no control (e.g., climate and natural events). If high government or state priority is given to sustainable agricultural production, and agriculture is recognized as an important contributor to, and source of, pollution, then appropriate policy instruments could be identified and enforced to increase the sustainability of the agro-ecological system and to reduce agriculturally produced pollution. The chosen instruments (e.g., taxes, subsidies or regulations) should guide and steer farm producers and natural resource users to change their practices in such ways that society achieves a higher level of sustainability and welfare from within the farming system. It may, however, be necessary to trade off some loss in another policy objective, say foreign exchange from agricultural export earnings from e.g., cotton, to achieve this higher level of sustainability.

Economic incentives, therefore, have a key role to play in establishing agricultural development which is both economic and sustainable (Schuh and Archibald, 1996). Getting the prices right is critical to obtaining an efficient allocation of resources over both time and space. Given the importance of agricultural policy in transmitting society's wants and needs back to farmers, we need open-minded, intelligent and informed policy makers and governments who are not easily swayed by self-interest and narrow political pressure groups.

With the increasing globalization and integration of international agricultural and food markets, a macro-economic approach should be adopted to establish the price incentives for sustainability, with domestic prices of outputs and inputs being firmly based on prevailing trade parity prices (*see* Agricultural Subsidies and Environmental Change, Volume 3). That is, agricultural prices should be near to their corresponding international market prices with little explicit subsidization or taxation of domestic producers. International trade policies and managed exchange rates are often a major cause of the lack of agricultural sustainability. In other words, distorted prices create unsustainable development. In developed countries, agricultural subsidization leads to the excessive use of fertilizers, pesticides, irrigation water and modern inputs, whilst in LDCs, discriminating against agriculture and artificially depressing agricultural prices causes land to be undervalued, thereby reducing incentives to careful soil husbandry. Land managers and users there have neither the means nor the incentive to invest in terracing or contour cultivation to arrest soil erosion on steep slopes. A top-down, macro-economic approach should provide the initial framework for agricultural price policy before the more common, bottom-up microeconomic instruments are used.

PRICING-POLICY EXAMPLES

In many parts of the world, current pricing-policies for farm output, natural resources and inputs unfortunately discourage sustainable agriculture.

Water Policy

Salinization and waterlogging problems in irrigation systems which are associated with overuse would improve significantly if water pricing mechanisms, which currently greatly undervalue water, were reformed (*see* Salinity and Agriculture, Volume 3; Waterlogging, Volume 3). Farmers should be expected to meet the recurrent, full operation and maintenance costs of water provision, although metering and charging by the actual volumes taken is probably still not feasible on both technical and cost grounds in LDCs, but is in California. Over 90% of India's water use is for agriculture, mostly for irrigation, yet the revenues from irrigating farmers cover less than 10% of the total costs of operating and maintaining the irrigation systems. Subsidies cost Indian taxpayers dearly and there is not enough money to repair and desilt canals, so the whole irrigation network deteriorates. Such systems encourage misuse and over-use of irrigation water at the head of the system and under-use at the tail, and have led to much water-logging and salinization.

Fertilizer Policy

Subsidies on chemical fertilizers should be carefully evaluated. Fertilizers can run off into water courses, contaminating aquifers and groundwater and polluting public water supplies, thus causing health and environmental hazards. The challenge is to develop fertilizer technologies and farming methods to reduce or mitigate these deleterious effects, whilst at the same time, reaping the considerable yield-enhancing and productivity benefits from balanced fertilizer use. New ways must be found to create more efficient use of fertilizers, so that smaller applications are needed.

Subsidies often divert supplies away from smallholders and are a disincentive to other soil nutrient management techniques. In any case, they are often fiscally unsustainable for the limited government budgets in LDCs. Subsidized fertilizer exports from developed countries discourage the start-up of local fertilizer industries and build up long-term trade dependencies. Fertilizer subsidies should be part of, and tied to, an overall land improvement investment program, and where fertilizer costs are unacceptably high for farmers due to transport constraints, the subsidies should be removed once the transport infrastructure improves. As an aid to accelerate farmer education and the adoption of inorganic fertilizers, subsidies have a limited and initial role to play, but should be rapidly phased out once farmer learning and widespread adoption have taken place.

Pesticides Policy

Pesticides, another chemical additive to agricultural production, often produce environmental and human health externalities; that is, unwanted and unpriced interactions on other parties. Many governments subsidize pesticide use despite these externalities and the capacity of insect pests to evolve and develop resistance to pesticides. Pesticides can affect groundwater, fisheries and wildlife, and pesticide deaths worldwide may be as high as 250 000. Integrated pest management strategies (IPM), allowing a limited use of pesticides as part of an overall strategy involving mixed crops bred for disease resistance, staged planting and the biological control of insects using natural predators, have proven successful in both reducing rice pests and the damaging externality effects in rice production in Indonesia. The use of pesticides by farmers has declined and rice yields have been maintained following a drastic reduction in input subsidies and outright bans on some of the more dangerous pesticides. Sustainable IPM strategies have since been adopted elsewhere in world agriculture. New taxes should be considered for the most polluting and dangerous pesticides to discourage their use and to fully re ect their environmental and human health externalities. *See* Integrated Pest Management (IPM) in an Era of Global Environmental Change, Volume 4.

Other Pricing-Policies

In Latin America and South-east Asia, logging concession prices should re ect their true economic values to reduce deforestation on lands which are inherently unsuitable for intensive agriculture. In Latin America, land tax systems which currently encourage extensive, low input farming should be changed and enforced. Governments worldwide should reconsider agricultural taxes, duties, price subsidies, overvalued exchange rates and other mechanisms which reduce farmers' incomes.

RESEARCH AND TECHNOLOGY POLICY

A high priority to promote sustainable agriculture is well funded research aimed at the holistic understanding of natural resource degradation and the development of new resource-conserving and yield-enhancing technologies. The key to sustainable increases in food production from a static and fixed natural resource base is knowledge. Research is needed to better understand the socio-economic and policy conditions that support investment in land and water management and to design the appropriate policy instruments that promote such conditions. Priorities might be soil nutrient improvement through green manuring and the use of locally available products, control of soil erosion, water harvesting and recycling techniques, and the use of salt-tolerant plant varieties in saline conditions, etc. Yield losses and the financial costs of natural resource degradation must be calculated and the private and public returns to investment in land improvement established. Low-cost strategies

for technology development with local experimental input, indigenous knowledge and participative approaches need funding to develop appropriate new technologies for hillside, dryland and other semi-arid farming environments. There is an urgent need for a wider range of site-specific, on-farm research with greatly improved knowledge inputs and feedback from farmers.

There are some reassuring developments in the way in which technologies for promoting sustainable agriculture are conceived, evaluated and implemented. The use of participatory techniques, a wider recognition of indigenous knowledge and a closer collaboration amongst researchers, extension workers and small rural households potentially lead to enhanced adoption and improved livelihoods, primarily because they recognize the on-farm and near market constraints involved. This work provides an important interface between social scientists and science and technology, and a interdisciplinary research agenda for the future.

Funding for agricultural research could come from a mixture of private and public sources, including non-governmental organizations (NGOs). The development and use of new technology is a key policy instrument in attaining sustainable economic growth and development. However, policies must also foster and finance agricultural research if the world is to feed itself; the omens here are not promising. For example, the budget of The Consultative Group on International Agricultural Research (CGIAR) fell during the 1990s, seriously jeopardizing the ability of its network of 14 International Research Centers to maintain their research, development and testing activities. There is now a great need to urgently bolster national and international agricultural research funding in the quest for sustainability.

CREDIT POLICY

Many sustainable farming systems will entail substantial investment in land improvement and soil remediation to support more intensive input use. These include the planting of trees and hedges on steep slopes, terracing, intercropping and mulching to build up soil organic matter. Credit may be needed to support interventions; this requires, large, upfront, initial lump sums. However, repayment capacities need to be established at the outset and only activities which are inherently profitable should be supported. In this way, governments, NGOs and land user associations can promote land investments via local farmer experimentation to develop and test new land management systems.

One should not, however, underestimate the shortfalls in funding capacities for credit schemes, particularly at the required local government and regional levels. In any case, government credit provision which supplies credit to farmers at subsidized rates of interest is not without its critics. Frequently, such schemes are themselves a source of bureaucratic and market failure: the funds provided are often diverted away from their intended agricultural investment use; institutions fail to stimulate savings, and as well as the lending behavior, are expensive on limited government budgets and do not create lasting and sustainable rural financial markets.

PROPERTY-RIGHTS POLICY

Secure and assured property rights and rights of access to natural resources are essential for long-term investment by land users in land conservation and improvement, and the development of private and community management strategies. A lack of secure tenure is a major constraint on investment in natural resources; improvement and open access to common property resources results in overgrazing, over-fishing, etc. Property rights for women farmers is a significant problem in many parts of the world, despite the fact that women are often the principal land managers and farmers. Tenure and rental arrangements in many regions often militate against sustainable agricultural management and discourage investment. Property rights reform can bring long-term benefits, but governments are often reluctant to resolve the inevitable con icts and dislocation associated with such redistribution issues.

AGRI-ENVIRONMENTAL POLICIES

A more recent development of policies, particularly in developed countries, is a range of policy measures to provide incentives to farmers to safeguard and conserve the environment and produce environmental goods and services rather than crop and animal products. These services include non-traded public goods, such as landscape, amenity, open space, habitats for ora and fauna, wetlands, etc. In the US, set-aside programs have been used to divert erodible farmland from crop production in order to protect the topsoil. Set-aside programs have been widely used in recent years in the US, European Union and Japan to control the over-production of subsidized cereal production, although sometimes an environmental component has been added to the supply reduction objective.

Since the mid 1980s, European Union farmers have been compensated to reduce the intensity of their (subsidized) farming methods to encourage conservation compatible with sound farming practice. The *environmentally sensitive area* (ESA) scheme operating throughout the EU uses voluntary agreements and compensation payments to encourage farmers to adopt environmentally friendly practices in designated areas. In the UK, farmers are offered a 5-year agreement with annual compensation payments in exchange for following a contract of prescribed farming practices which include restrictions on fertilizer use and stocking densities, prohibitions on the use of herbicides,

pesticides and lime and on the installation of new drainage. Farmers also agree to maintain landscape features, such as hedges, ditches, ponds, woodlands, dry-stone walls and field barns. Adherence to the ESA prescriptions increases the supply of public environmental goods at the expense of a fall in agricultural output and profits, and farmers may end up being over-compensated for the income losses incurred (Lingard, 1999).

REFERENCES

Ellis, F (1992) *Agricultural Policies in Developing Countries*, Cambridge University Press, Cambridge, MA.

Lingard, J (1999) A Linear Programming Model of a Pennine Dales Farm in an Environmentally Sensitive Area , *Farm Manage.*, 10(6), 318–333.

Pretty, J (1995) *Regenerating Agriculture: Policies and Practice for Sustainability and Self-reliance*, Earthscan, London.

Schuh, G E and Archibald, S (1996) *Integration of Sustainable Agriculture and Rural Development; Issues in Agricultural Policy*, Winrock International, Morriston, AR.

Syers, J K, Lingard, J, Pieri, C, Ezcurra, E, and Fause, G (1996) Sustainable Land Management for the Semi-arid and Sub-humid Tropics , *Ambio*, 25(8), 484–491.

Policies to Achieve Sustainable Tourism

Eugenio Yunis
World Tourism Organization, Madrid, Spain

Tourism has been one of the fastest growing industries in recent decades, the importance of which is likely to increase even more in the 21st century, with signi cant impacts, positive and negative, upon the environment.

Tourism refers to all movements of people, irrespective of their motivation, but involving at least one overnight stay and not more than twelve consecutive months in a location that is not their usual place of residence. Tourist movements can be domestic or international.

The World Tourism Organization (WTO OMT) conducted a detailed study (Tourism Vision 2020, WTO OMT, Madrid, 1999) to examine how the tourism industry, and particularly tourism demand, is likely to evolve in the next 20 years. The results indicate that international tourist arrivals are likely to increase at an annual average rate of 4.1%, to reach nearly 1.6 billion by the year 2020; that is almost three times the number of arrivals recorded in 1998.

The impacts that these hundreds of millions of tourists moving around the globe may cause upon the environment need to be anticipated, prevented if possible and continuously monitored, if tourism is to contribute to sustainable development. This is so because of the complex and close relationships existing between tourism and the environment.

THE GROWTH OF TOURISM: PAST PERFORMANCE AND FUTURE TRENDS

Tourism has undoubtedly been one of the outstanding social and economic phenomena during the second half of the 20th century. International tourism movements in the world at large increased from 25 000 000 in 1950, to an estimated 650 000 000 in 1999, that is, an average growth rate of 7.3% per annum. This is perhaps the strongest and most steady growth rate observed after World War II among all indicators used to measure progress in social and economic development.

Since tourism became accessible to most consumers in the developed world, by the mid-1960s, world tourism receipts have been growing even faster than tourist arrivals, indicating that consumers are devoting a higher percentage of their income to tourism. Thus, in the period 1975–1980, while arrivals grew at an average annual growth rate of 5.2%, the corresponding income generated increased by an annual average of 20.9%. Similarly, during the 1985–1990 period, arrivals grew at 7.0% annually, while tourism income grew at 17.9% per annum.

The above figures refer only to international tourism. Yet it is equally important to consider domestic tourist movements, since the impacts of both types of tourist ows on the physical environment are basically the same. Detailed accurate figures for domestic tourism are, however, difficult to obtain. Some countries, such as France, Spain and the United States, have estimated that hotel occupancy by nationals is between 8 and 10 times higher than by foreigners.

All signs indicate that tourism will continue to grow, as a result of a number of factors affecting both demand and supply. Demographic trends, increased disposable incomes and free time, better and faster communications, better access to information by consumers, and an increased interest in nature and culture, seem to be the main elements pushing tourism demand. On the supply side, the openness and globalization of national economies, competition among large hotel groups, airlines and tour operators, cheaper transport, the creation of new generations of tourism products (e.g., thematic parks, sports tourism, etc.), are some of the main factors that contribute to the growth of tourism.

THE RELATIONSHIP BETWEEN TOURISM AND THE ENVIRONMENT

Unlike most other economic activities, tourism has a complex, many-sided relationship with the natural and man-made environment. First and foremost, tourism needs a clean, well-preserved, healthy environment to ourish. Indeed, the environment is fundamental to most forms of tourism: natural beauties and landscapes attract holiday visitors, as do pristine areas with rich biodiversity, well-preserved archaeological sites, architecturally rich urban areas, historic routes and the like. Clean, orderly cities are preferred by those attending conferences or trade fairs, while business and other tourists will generally avoid polluted cities or contaminated natural areas.

The elements that make up the environment are, in fact, the raw materials for the tourism industry. But unlike other industries, the tourists should not deplete these inputs, since such depletion would reduce their capacity to attract tourists.

Environmental disasters or substantial changes to the natural environment, such as intensive and indiscriminate felling of forests, shipwrecks of oil tankers, industrial pollution of rivers, etc., can drastically affect the possibility of developing tourism in areas endowed with natural or cultural assets.

Climate change, caused by many human activities including tourism-related air transport, is proving to be a major threat to many tourist destinations. Sea level rise resulting from global warming in coming decades will jeopardize not only the tourist industry of many small islands and small island states, but also their entire existence, since tourism is in most cases their main, and frequently their sole economic activity. Similarly, global warming and the resulting reduced snowfall are already affecting ski resorts in several mountain areas.

Secondly, tourism can significantly contribute to the preservation and improvement of natural areas, places of historic and cultural interest, and even urban environments. By giving economic value to the landscape and to archaeological sites, tourism generates income for their conservation and appreciation by people. Similarly, tourism can serve to improve people's environmental awareness. Adequately conducted eco-tourism and other forms of nature-based tourism are powerful tools for educating the general public about the fragility of ecosystems, endangered species and their conservation, and environmental matters in general.

However, for these benefits to occur, tourism must be developed and managed in a manner that does not exhaust the resource, while provisions need to be made for ensuring that part of the income generated is actually spent for conservation. This can be done through one or a combination of several instruments, such as taxing revenues of tourism operators, entrance fees, accommodation and food charges, direct sales to tourists, etc.

A third aspect that needs to be considered is that inadequately planned or improperly managed tourism can result in very negative environmental impacts. Uncontrolled growth of tourism facilities, be it hotel, transport or catering infrastructure, particularly in coastal zones, has already destroyed the resource base of many destinations around the world. Furthermore, the excessive number of summer tourists in certain beaches, that are scarcely populated during the off-season, has resulted in the contamination of marine or lake environments, due to inappropriate, or even the total absence of sewage systems, or of garbage collection and related treatment infrastructure.

Similarly, the use of national parks and protected natural areas by tourists, if uncontrolled, may degrade the plant and animal life of the areas, instead of contributing to their conservation.

In summary, the environmental impacts of the tourism industry can be numerous. This means that a careful, balanced analysis of such impacts – actual or potential – is normally required before reaching any conclusions regarding the development of tourism at any particular site, area, town or country. Such analysis is specially recommended prior to the implementation of tourism investment projects, particularly when these are of a large size, or when a whole area or community will be transformed as a result of such investments.

THE REACTION OF THE INTERNATIONAL COMMUNITY

It is not surprising, therefore, that several international organizations – intergovernmental, non-governmental and from the tourism and travel sectors – have addressed the subject of environmental sustainability in tourism.

The WTO OMT, as the largest intergovernmental agency in this sector, has been promoting an environmentally conscious approach to tourism planning since the early 1980s, with a number of technical publications and assistance missions to developing countries. The more recent publications of WTO OMT in this field are detailed in the bibliography at the end of this article.

The WTO OMT, jointly with the World Travel and Tourism Council, acting on behalf of the private sector, and the Earth Council, together formulated an Agenda 21 follow-up for the travel and tourism industry in 1995, being the only economic sector to have used the global Agenda 21 – emanating from the Rio Earth Summit in 1992 – to develop its own agenda for action.

Agenda 21 for the travel and tourism industry is necessarily general in nature, since it has been prepared by organizations which have a global coverage and a universal responsibility. Yet, the fact that it is global and general does

not diminish the validity of its recommendations. Indeed, in its part two, there is a very detailed framework for action, describing the responsibilities of government departments, national tourism authorities and travel trade organizations on the one hand, and of individual travel and tourism companies on the other. Of course not all these guidelines are equally applicable to all countries and destinations, but most can be easily adapted and applied, if there is the political will to do so.

Joint actions by the WTO OMT with the United Nations Environment Program, the World Conservation Union and other institutions, as well as independent actions carried out by WTO OMT and the United Nations Environment Program, have resulted in a number of guidelines for the development and management of specific types of tourism destinations, such as national parks, historic cities, archaeological sites, coastlines, small islands and others.

The United Nations Commission on Sustainable Development also took an interest in the tourism sector and the need for it to be developed in a sustainable manner. At its 7th session (April, 1999), it focussed on tourism, made a number of recommendations to governments and to the private tourism sector, and took a number of decisions in this respect (UN-CSD7 report on Tourism and Sustainable Development, UN, New York, 1999).

MAKING TOURISM MORE SUSTAINABLE: POLICIES AND GUIDELINES

There is often confusion about the concept of sustainable development when applied to tourism. Some authors and commercial organizations tend to equate sustainable tourism with green tourism or eco-tourism, that is, with some particular forms of tourism related to nature. It should be clearly stated, however, that this is not correct; all tourism activities, whatever their motivation (beach holidays, cultural circuits, business trips, conferences, congresses or fairs, health-related trips, adventure travel or eco-tourism, etc.) must be sustainable.

This means that the planning and development of tourism infrastructure, its subsequent management and also its marketing should be based on environmental, social, cultural and economic sustainability criteria; these criteria are meant to ensure that neither the natural environment nor the socio-cultural fabric of the host communities will be impaired by the arrival of tourists; on the contrary, local communities should benefit from tourism, both economically and culturally, while increased resources should be given for conservation and enhancement of the surrounding natural environment. Sustainability implies that not only tourism enterprises, but also the communities in which they operate, have something to gain from tourism. Sustainability implies, to paraphrase a recommendation of the Rio Summit of 1992, that tourism resources and attractions are utilized in such a way that their subsequent use by future generations is not compromised.

The sustainability of tourism should not be viewed as a static concept, referring only to the planning and development phases. Ensuring sustainability requires a continuous process of adaptive management and monitoring, since as destinations develop, grow in size and become more popular among tourists, new problems will arise and new solutions will have to be found to maintain or even improve their sustainability level.

A number of policy and planning approaches for achieving a sustainable tourism development have been proposed by the WTO OMT and other international agencies and academic institutions. The key elements of such approaches can be summarized as follows:

- A tourism policy is required at the national, regional and local levels. Tourism can no longer grow as a result of spontaneous, individual and uncoordinated actions by isolated actors; governments at all levels are required to formulate policies for its orderly development.
- The tourism policy should be the result of a participatory process, in which all interested parties, and particularly the local community, are involved.
- A comprehensive approach should be adopted, whereby all aspects of tourism development are considered in the planning process, including the tourist attractions, the basic infrastructure and services, accommodation and catering facilities, transportation, management aspects, human resource development, as well as the institutional elements.
- Similarly, an integrated approach is required. Tourism cannot be planned in isolation; instead, it must be part of the overall development efforts of a country or area, while local tourism plans must be integrated into national tourism strategies.
- Environmental impact assessment techniques must be applied from the very beginning of any tourism development plan or project. Efforts should be made to ensure that the cost of avoiding any potential damages to the environment as a result of such development has been taken into account in the economic and financial calculations.
- Tourism companies and destinations must be managed with the environment in mind. This needs appropriate environmental training of tourism managers and employees at all levels, to enable industry-wide application of environmental management systems in all tourism companies.
- The management of tourism at both the company and the destination levels requires continuous monitoring, particularly with respect to environmental variables. Monitoring allows managers and public authorities

to take corrective actions when needed and before negative impacts become irreversible.
- As an indispensable element for the above monitoring process, indicators for sustainable tourism (WTO, 1995) must be developed and introduced as a normal practice at all tourism destinations. Indicators also serve to guide consumers about the environmental and social quality of the destination, and are good benchmarks to stimulate destinations to compete on sustainability grounds rather than on price.

An effective tool for promoting conservation of natural areas and their biodiversity, rich archaeological sites, and even historic cities, and for ensuring that a sustainable tourism use is made of such places is their designation as protected areas. In this particular case, the following essential guidelines can be suggested for ensuring that tourism contributes to the preservation and improvement of such areas, while not generating negative impacts:

- *Zoning*: it is fundamental to define, in all protected natural areas, zones with different uses and use intensity. Tourists cannot normally be allowed in all parts of the protected area, or at least not all of them at the same time or in the same number.
- *Tourism demand management*: tourism demand for natural areas must be managed, and this can be done through a combination of various instruments: (a) determining the total maximum number of tourists/visitors that can be accepted in the area, per day and/or *per annum*, also known as *tourist carrying capacity*; (b) establishing a pricing policy in line with the acceptable maximum number of visitors; (c) defining marketing, distribution and promotion policies in line with the pricing policy and with maximum demand numbers.
- *Design for tourism management*: management infrastructure as well as tourist facilities in the area should be designed keeping in mind that they are to be used by tourists. Local materials and local styles (ecodesign) should be used in all buildings within the area, incorporating all environmentally friendly technologies available (solar energy, natural ventilation and illumination, recyclable materials, classified waste disposal, etc.) Pedestrian paths should be built and the surface of intensive use areas should be hardened. A comprehensive waste management system should be established.
- *Crowd management*: tourist routes within the area should be designed in order to manage tourists' movements. Timing and sequencing programs should be established and implemented with appropriate information and sign posting. Norms should be set up concerning the use of vehicles within the area. Access and stop points should be established in line with timing and sequencing, etc.
- *Continuous monitoring of impacts*: a comprehensive site inventory should be established with the double purpose of: (a) selecting species of greatest interest to tourists, and (b) undertaking periodic inventories or counts of species. Regular ecological audits should be carried out to measure any impacts from tourism. A battery of indicators and a corresponding data collection system should be established.
- *Behavior management*: rules should be established, disseminated and enforced for the behavior of tourists. A code of practice should be set up for tour operators marketing the area and also for the tourists. Neighboring communities, protected-area staff, and tourists should be educated on the importance of preservation in general and on specific endangered species that may be present in the area. Ample information should be provided to tourists, before and during their visit to the area, including descriptive information about the area and its habitat, dangers and risks, endangered species, do's and don'ts within the area, etc.
- *Pro t sharing for local conservation and improvement*: mechanisms should be established to ensure that part of the financial proceeds resulting from tourism activities within the protected area are re-invested in the same area.

THE ROLES OF DIFFERENT STAKEHOLDERS

The Tourism Private Sector

It is commonly stated that, in a free market economy, the responsibility for ensuring the sustainability of tourism products and services should rest basically with individual companies. It is further argued that the industry will listen to the market, adjusting its behavior to become more environmentally friendly; this can be done either by each company adopting and applying its own environmental management system, or by following industry-wide, commonly agreed but voluntary codes of conduct *vis-à-vis* the environment (several examples can be found among major hotel chains, the International Hotels Environment Initiative, Green Globe 21, the Costa Rica based Certification System for Sustainable Tourism, and many others (see the United Nations Environment Program, 1995)).

Several international hotel chains, as well as many independent hotels have incorporated into their operations these environmental management systems, or at least a number of concrete measures to save energy and water, to reduce waste, to recycle some of this waste, etc. In so doing, they have achieved considerable environmental improvements as well as important cost savings and financial gains. Furthermore, their public image has also improved, thus enhancing their marketability, at least among environmentally conscious tourists.

The Public Sector

The role and responsibility of private tourism companies cannot be understated. However, it might be naive to think that only with such measures would environmental sustainability in tourism be assured. Public authorities, particularly those at the local or municipal level have also various and very important roles to fulfil. Their main functions and responsibilities are:

- setting up environmental, socio-cultural and economic norms, rules, regulations and minimum standards in general or in particular for the tourism sector;
- supervising and controlling that such norms are being enforced at the appropriate levels, and that companies and individuals are satisfying the standards;
- disseminating information among tourism and hotel companies with respect to the technologies available, nationally and internationally, to reduce or arrest the negative environmental impacts which typically occur in tourism and hotel companies;
- establishing loan schemes, subsidies, fiscal and other measures and incentives to encourage the introduction of the latest environmental technologies into the day-to-day operations of tourism companies;
- taking appropriate measures to manage and maintain, in a sustainable manner, those public components of tourism facilities, such as natural and cultural attractions, public beaches, roads and other forms of transport infrastructure, parking grounds, public parks, etc.;
- promoting the establishment of voluntary certification schemes, and their related accreditation mechanisms, for stimulating sustainability in tourism companies and destinations.

CONCLUSION

National, regional and local authorities need to fully take up their responsibilities in tourism. The commercial tourism sector, on its part, must follow the above principles and guidelines in a more rigorous fashion than at present. This will not only ensure environmental sustainability of the tourism industry, but also strengthen its capacity to generate economic benefits for companies and host communities.

Most importantly, the tourism sector will then make a stronger contribution to the overall sustainability of economic development of destination countries, enhancing at the same time the image of the tourism industry, and as a result, obtaining greater financial and other support from governments.

REFERENCES

United Nations Environment Programme (1995) *Environmental Codes of Conduct for Tourism*, UNEP.

World Tourism Organization (1995) *What Tourism Managers Need to Know: a Practical Guide to the Development and Use of Indicators of Sustainable Tourism*, WTO, Madrid, Spain.

FURTHER READING

Bosch Camprubí, R, Pujol Marco, L, Serra Cabado, J, and Vallespino, R F (1998) *Turismo y Medio Ambiente*, Editorial Centro de Estudios Ramón Areces, S A Madrid, Spain.

Capellà, D, Munoz, F, and Ullastres, J A (1998) *Turismo Sostenible en el Mediterráneo: Gu´a para la Gestión Local*, Brau Ediciones, Spain.

Cater, E and Lowman, G, eds (1994) *Ecotourism: a Sustainable Option?* Royal Geographical Society and John Wiley & Sons, Chichester, 1–218.

Center for the Promotion of Imports from Developing Countries (1997) *Quick Scan Sustainable Tourism*, CBI, The Netherlands.

Cooper, C and Wanhill, S (1997) *Tourism Development: Environmental and Community Issues*, John Wiley & Sons, Chichester.

Gortazar, L and Marin, C (1999) *Tourism and Sustainable Development: the Island Experience*, Gobierno de Canarias, INSULA (International Scientific Council for Island Development), Canary Islands, Spain.

Hall, C M and Lew, A A (1998) *Sustainable Tourism: a Geographical Perspective*, Longman, London.

Harrison, L C and Husbands, W, eds (1996) *Practising Responsible Tourism: International Case Studies in Tourism Planning, Policy and Development*, John Wiley & Sons, New York.

Inskeep, E (1991) *Tourism Planning: an Integrated and Sustainable Development Approach*, Routledge, London.

Mowforth, M and Munt, I (1998) *Tourism and Sustainability: New Tourism in the Third World*, Routledge, London.

Stabler, M, ed (1997) *Tourism and Sustainability: Principles to Practice*, CAB International, UK.

Tourism Concern/WWK UK (1992) *Beyond the Green Horizon: Principles for Sustainable Tourism*, Tourism Concern/WWF, UK.

United Nations Environment Program/World Tourism Organization (1998) *Final Report of the Lanzarote Conference on Sustainable Tourism Development in Small Islands States and Other Islands*, UNEP/WTO, *Lanzarote, Spain*.

World Tourism Organization (1998) *Guide for Local Authorities on Developing Sustainable Tourism*, WTO, Madrid, Spain.

World Tourism Organization (1999) *Guide for Local Authorities - Supplementary Volume on Sub-Saharan Africa*, WTO, Madrid, Spain.

World Tourism Organization (1999) *Guide for Local Authorities - Supplementary Volume on Asia and the Paci c*, WTO, Madrid, Spain.

World Tourism Organization (1999) *Guide for Local Authorities - Supplementary Volume on Latin America and the Caribbean*, WTO, Madrid, Spain.

World Tourism Organization (1999) *Sustainable Development of Tourism - an Annotated Bibliography*, WTO, Madrid, Spain.

World Travel and Tourism Council/World Tourism Organization/Earth Council (1995) *Agenda 21 for the Travel and Tourism Industry: Towards Environmentally Sustainable Development*, WTTC/WTO/Earth Council.

Policy Exercises: a Tool for Solving Environmental Policy Dilemmas

William M Stigliani
Center for Energy and Environmental Education, University of Northern Iowa, Cedar Falls, IA, USA

Environmental problems, particularly those related to global change such as climate change, deforestation, species extinction, and urbanization, are intimately linked to socio-economic activities. Solutions to these problems involve complex trade-offs between economic, social, and technological considerations. Often there are no easy remedies, and policy makers are confronted with the dilemma of choosing from a set of policy options, none of which is appealing for a number of reasons, e.g., high scienti c uncertainty, high-stake economic and political risks, potential for social upheaval.

Moreover, no matter what policy is eventually adopted, there are likely to be winners and losers. For example, real estate developers in coastal areas would be losers in any policy curtailing coastal settlement to reduce the impacts of sea level rise. In contrast, real estate developers of lakeside properties are likely to be winners if the focus of settlement were to shift away from coastal marine areas to scenic freshwater resort areas. How is it possible to secure consensus amongst stakeholders when some will win and others will lose?

There is general agreement that standard means for factoring scientific information into long-term strategies for managing the environment are inadequate. One major problem with tackling policy dilemmas is the large information gaps in scientific knowledge that currently preclude the development of rigorous, quantitative analyses for predicting environmental change. Moreover, for the information that is available, we do not yet know how to use it optimally in the formulation of prudent policies that both protect the environment and foster economic development. Another perhaps more fundamental problem is that currently there is no established mechanism by which scientists and senior policy people can exchange ideas on the long-term management of the environment. As a result, policy makers are often disappointed by the lack of policy relevance in much of the scientific research, and scientists are often discouraged because their research appears to have no impact on public policy.

Historically, two common means of science/policy synthesis have been: (1) Blue Ribbon panels; and (2) large-scale computer simulation models for environment-economic analysis (Brewer, 1986). Although useful under many circumstances, neither is entirely adequate as a stand-alone policy tool for managing the kind of uncertain

Table 1 Environmental issues as policy dilemmas

Issue	Policy dilemma
Managing water resources in an era of climate change	Reservoirs, dams, and municipal water works are built to last at least 50 years under assumption of a stable water supply. For new construction, should planners factor in costly safety factors to protect against fundamental changes in water availability caused by changes in climate?
Long-term forestry management in an era of global deforestation	Currently there is an abundance of harvestable wood for paper pulp, but global deforestation and other factors could cause a severe wood shortage in the future. Increasing wood supply requires planting of new forests decades in advance. Should forestry managers plan to increase forest lands at a time when there is no crisis, and future supply and demand is too uncertain for precise planning?
Development of coastal areas in era of climate instability and sea level rise	Settlement in coastal areas is booming globally, but climate change may cause irreparable damage stemming from sea-level rise and an increase in the frequency and force of hurricanes. Should plans for coastal development factor in concerns about climate change?
Transportation needs versus air quality and greenhouse gas emissions	The rate of increase in the world passenger car fleet is four times higher than the rate of population growth. Much of this increase is in developing countries, causing serious urban air quality problems and large increases in the emission of the greenhouse gas CO_2. Should automobile traffic be regulated to reduce urban air pollution and mitigate against climate change?

environmental futures confronting us over the next 100 years. Table 1 provides examples of such possible problems.

Several authors (Wack, 1985a,b; Beck, 1983) have argued for a scenario approach to managing an uncertain future. They note that the role of scenarios is to enhance the decision-maker's understanding of the future by providing perceptions of several possible future environments against which decisions can be tested. The goal is not to predict the future, but rather to manage the present by learning to live with uncertainty, factoring it into the decision process, and building bridges between decision-makers and scientists.

In the late 1980s, a group at the International Institute for Applied Systems Analysis combined scenarios and computer simulations in the development of a new methodological framework named *policy exercises* (Brewer, 1986; Toth, 1986, 1988, Stigliani *et al.*, 1989). Similar in approach to war games developed half a century ago by the US military, the process utilizes alternative scenarios of possible future environments as the centerpiece for structuring appropriate policy responses. Environmental scientists and senior policy makers are confronted with scenarios dealing with a specific environmental problem posed as a policy dilemma. The full time span of the scenarios is not revealed initially. Rather, the scenario is presented in time slices (e.g., 2005, 2015, 2025). At each time the participants are asked to describe policies for appropriate management of the environment. The suitability of the policies is then tested as the future events in the scenario unfold. In fact, the future evolves partly in response to the succession of policy moves made by the participants. Moreover, the participants are invited not only to partake in the exercise, but also to challenge the validity of the scenario presented. However, the challenger must then defend his or her vision of the future. Such interplay is encouraged to ensure that the scenarios challenge the decision-maker's preconceived ideas about what will happen, and accordingly may teach him or her to think in terms of an uncertain future.

The objective of a policy exercise is not to develop a firm set of environmental policies for the next 50 years but:

- to sensitize policy makers to uncertainties due to natural and human factors and to alert them to the possibility that some of their current policies, even though well intended, may lead to negative consequences over the long term;
- to train policy people in how to use adaptive management strategies, which are continually updated as new information is received;
- to convince policy people of the need for regular policy exercises with scientists, as a preferred method of setting environmental agendas and priorities.

See also: Futures Research, Volume 5; Scenarios, Volume 5.

REFERENCES

Beck, P W (1983) Forecasts: Opiates for Decision Makers, in *Third International Symposium on Forecasting*, Philadelphia, PA, June 5–8, 1983.

Brewer, G D (1986) Methods for Synthesis: Policy Exercises, in *Sustainable Development of the Biosphere*, eds W C Clark and R E Munn, Cambridge University Press, Cambridge, 455–473.

Stigliani, W M, Brouwer, F M, Munn, R E, Shaw, R W, and Antonovsky, M (1989) Future Environments for Europe: Some Implications of Alternative Development Paths, *Sci. Total Environ.*, 80, 1–102.

Toth, F L (1986) *Practicing the Future: Implementing the Policy Exercise Concept*, Working Paper 86–23, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Toth, F L (1988) *Practicing the Future Part 2: Lessons from the rst Experiments with Policy Exercises*, Working Paper 88–12, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Wack, P (1985a) Scenarios: Unchartered Waters Ahead, *Harv. Bus. Rev.*, September/October, 71–89.

Wack, P (1985b) Scenarios: Shooting the Rapids, *Harv. Bus. Rev.*, November/December, 139–150.

Policy Responses to Public Health Issues Relating to Global Environmental Change

see Policy Responses to Public Health Issues Relating to Global Environmental Change (*Opening essay,* **Volume 4**)

POP (Persistent Organic Pollutants)

see International Environmental Standards (**Volume 4**)

PPP (Polluter Pay Principle)

The PPP holds that the price of a good or service should include all the costs incurred in its production. Thus, the costs imposed by emissions and resource use are internalized. In essence, the polluter pays the cost of pollution. This is akin to a Pigovian tax (named after the economist

Pigou), which in theory is equivalent to the marginal cost of damage at the optimal pollution level. Ideally, if a firm pays the full cost of pollution, then there is an incentive to make improvements in technology and reduce emissions. The PPP is often envisioned in the form of a tax (for an example of ecotax instruments *see* Eco Taxes, Volume 4).

In 1972, the Organization for Economic and Development Co-operation, and the European Community (EC) adopted the PPP as a component of environmental policy. The EC further endorsed the PPP in 1975 in the Single European Act, though the extent of implementation has been variable.

KEVIN S HANNA *Canada*

Precautionary Principle

The *precautionary principle* is both a support for, and a criticism against, the science of global environmental change. It places the onus of proof on those who wish to change the *status quo*, and to show beyond reasonable doubt that they are following a duty of care for the planet and its inhabitants.

This dictum carries two important messages. The first is that science has to operate on the assumption that the lack of evidence of cause and effect is not a sufficient condition to adopt the provisional hypothesis that there is no link. In essence, the scientific prognosis must focus on the worst-case set of assumptions and also attend to the possible consequences for the most vulnerable people. At the societal level, *vulnerability* is the incapacity to avoid danger, or to be uninformed of impending threat, or to be so politically powerless and poor as to be forced to live in conditions of danger. On a personal level, vulnerability relates to such a physical condition as to be unable to withstand stress to a weakened immune system. In essence, precaution carries with it both the principle of playing safe, and paying due attention to the physically weak and the politically powerless.

The second quality of precaution is that it relies on a more participatory form of science for its effectiveness. This notion is called *civic science*, and involves a more deliberative, and stakeholder involved, process of analysis and consensus building. The key to this aspect is that outcomes are not given: they are shaped by the very processes of participation that in uence their interpretation, by those whose consent is vital if decisions are to be taken.

Take, for example, the issue of traffic futures for a city, in the face of uncertain evidence over climate change, the role of offsets for car-produced carbon dioxide (CO_2), and the public health effects of vehicle exhausts under changing future climates (possibly longer, drier summers with more sunshine, inversions and concentrations of photochemical smog). Any decision to require the vehicle user to pay for these social costs will require their understanding of what might happen to climate and to public health under differing combinations of vehicle movement over the next 50 years.

The precautionary principle, therefore, encourages interactive (i.e., user-friendly) and integrated (i.e., comprehensive) assessments of various future states. These may be in uenced, in turn, by assumptions in the science of global environmental change and in the likely outcome of certain policy measures (such as eco-taxation, using the revenue for socially constructive investments).

Invoking the precautionary principle is also a re ection of growing distrust of science as a competent and authoritative arbiter of complex problem solving. A majority of citizens in Western countries do not believe that science is free of commercial and political pressures. This loss of faith is evident in the dispute in Europe over the introduction of genetically modified food, though it applies to food safety generally. In the United States, however, precaution is less of an issue, as the American culture supports innovative technology and capitalistic risk taking. Indeed, world trade regulations heavily in uenced by American business interests, do not allow too strict an interpretation of the principle to affect the spirit of free trade. This will remain an area of notable tension between North America and Europe for some time to come.

The precautionary principle, therefore, re ects degrees of trust in the institutions of science, and the culture of risk and innovation in nation states. It is both socially and politically framed, but has significant implications for the future role of environmental science in a global age.

See also: Precautionary Principle, Volume 5.

FURTHER READING

Cable, S and Cable, C (1998) *Environmental Problems: Grassroots Solutions: The Politics of Grassroots Environmental Conicts*, St Martin's Press, NY.

Foster, K R, Vecchia, P, and Repacholi, M H (2000) Science and the Precautionary Principle, *Science*, 288, 979–981.

O'Riordan, T, Cameron, J, and Jordan, A, eds (1994) *Reinterpreting the Precautionary Principle*, Cameron May, London.

Tinkner, J and Raffensperger, C, eds (1999) *Protecting Public Health and the Environment: Implementing the Precautionary Principle*, Island Press, NY.

TIM O'RIORDAN *UK*

Q

Quotas in International Environmental Negotiations

Amanda Wolf
Victoria University of Wellington, Wellington, New Zealand

A quota in international environmental negotiation refers to a limit or target, designated in units directly related to an environmental problem, chosen by consensus or near-consensus. Limits may be presented as a single gure, or as shares of the total limit. Allocations of shares may be given as percentages of the total or as concrete numbers. Quotas may stand alone as instruments to achieve environmental objectives, or they may form a basis for or be combined with other instruments. Not surprisingly, over time, quotas have become more sophisticated and the range and style of their use have expanded. The most recent instances of quotas underpin key integrated management concepts.

In some theoretical literature, quotas are criticized as crude measures relative to incentive based measures for environmental management. However, particularly at the international level, quotas endure for reasons associated both with their inherent qualities and with the fact that management strategies using quotas, like all other strategies at the international level, must be negotiated. Quotas can be simple, yet exible, and thus suited to the inevitable accommodation of different interests and circumstances.

A balanced assessment of how quotas contribute to better long-term environmental policies requires a review of the various forms of quotas and their historical applications, a consideration of quotas against the key criteria of effectiveness, ef ciency, and equity, and an examination of the features of negotiated decision making that may favor quotas.

Negotiation is the principal means of policy making at the international level. It is a painstaking, incremental process, complicated by the complexity and uncertainty of many environmental problems. What is being negotiated is not some state of nature or environmental quality, but the mechanisms through which it is hoped environmental objectives will be realized. Parties negotiate over rules, standards, and practices, with environmental objectives as indirect in uences. Many factors – science, economics, politics, practice, beliefs, values, behavioral theories – play a part in the negotiation of quotas. A great deal of attention is directed to the features of a basic environmental management framework.

Negotiators of international environmental agreements know that their decisions will have implications in terms of environmental effectiveness, productive, and allocative efficiency, and ethical and distributional features of effects and costs (equity) (Organization for Economic Cooperation and Development, 1994). These criteria form a basis for deciding what instruments to use. In general, instruments are classified on the basis of several features: whether the instrument is direct or indirect, freeing or constraining, process- or effect-focussed, and quantitatively or qualitatively based. This last point distinguishes instruments directed to the quantum of an activity from those directed at the qualities, such as rate and the use of best available technology, of an activity. Bearing in mind that environmental instruments are mostly shades of gray, quotas typically are direct, constraining, effect-focussed, and quantitatively based. In comparison with other instruments, quotas have been thought to give greater weight to the implications for environmental effects and less to costs.

Basic quota instruments have two parts. The first part designates the unit of management and the second, a rule. The unit of account is the focus of limits in activity or behavior. Examples of units of account are elephant tusks over 10 kg taken by non-native hunters in a given year or tonnes of chloro uorocarbons (CFCs) emitted by a developed country in a given year. The second part is a statement of what shall be done and may include conditions and eligibility requirements. A corresponding rule in the tusk example might specify that no more than 50 be internationally traded in the given year between approved buyers and sellers. How a quota is to be implemented, monitored and enforced are key issues for the success of the instrument and will have bearing on the choice of units of account and rules.

The simplest type of quota sets a single total limit, and the management rule requires all participating parties to jointly ensure that the limit is respected. Early whaling regulations used this device to set yearly numbers, by species, for whales that could be taken by members of the International Whaling Commission. A single limit may also be indexed. The whale quota evolved into a total allowable catch denominated in blue whale units. This measure was based on the different amounts of oil that could be extracted from different species of whales. For example, two and one-half humpback whales were considered equivalent to one blue whale. In 1945, the quota was set at 16 000 blue whale units. Similarly, rules based on baskets of chemicals are used in ozone layer and climate change agreements. However, in these cases, the baskets are only a basis for more elaborate rules.

Indeed, while total limits on their own are now rare, many environmental agreements make use of total limits in conjunction with allocations of the total among participating parties. Allocations may be expressed either as a percentage of the total or as a discrete number. Many fisheries treaties use a base-year catch, as a proportion of total catch that year, to allocate shares of successive years' total allowable catch. Occasionally, as in the Convention on Future Multilateral Cooperation in the Conservation of Northwest Atlantic Fisheries (1978), parties will agree on a total catch and divide it up without recourse to a principle or formula. In addition to historical use, formulas may draw on geographic proximity, relative economic development or economic hardship, and traditional use.

Some recent agreements still make use of rather simple quotas. For instance, the Convention for the Conservation of the Southern Bluefin Tuna (1993), essentially set a total allowable catch at the recent catch level in the expectation of subsequent ecologically justified allocations, which, however, have not been achieved. Parties are thus loosely bound to respect an allocation based on their historical catch despite changing conditions. It can be argued that in this instance, the simple quota represents an achievement over open harvest and provides a basis for incremental improvement. In a similar vein, in putting ambient standard pollution agreements into operation approximate measures may be used, due to technology limitations and lack of information. Thus, across the board 50% reduction in total inputs from rivers were called for in the 1987 amendments to the Convention on the Protection of the Marine Environment of the Baltic Sea Area (Helsinki Convention, 1974).

Despite the instances already cited, it is sometimes claimed that pollution-related treaties do not make use of quotas. This claim is likely based on the absence of explicit total limits. Many agreements (in pollution and conservation cases) simply set an allocation formula or otherwise assign limits directly to participants. An example is found in the 1985 sulfur dioxide protocol to the Convention on Long-Range Transboundary Air Pollution (LRTAP, 1979) through which countries agreed to curb their emissions of sulfur dioxide by 30% from a base year. In this case, the effect of the rule on the environment is directly linked to the number and emissions of participating countries.

The definition of a quota easily covers not only temporary bans – a quota limit set at zero until some aspect of the ecosystem recovers – but also some rules associated with limit values. Limit values come in the form of emission or ambient (environmental quality) standards for pollutants in a given period and medium. If the intent is to hold aggregate amounts of the pollutant to some set level, that level is a quota. For example, when Mexico and the United States concluded a treaty for managing the salinity of the Colorado (Brownell and Eaton, 1975), a salinity standard was established. Based on a reference salinity at Imperial Dam, and taking into account a required minimum ow of water, water entering Mexico is to have an annual average salinity of no more than 115 • 30 ppm. The rolling reference value meant that the standard was readily applicable in drought and normal conditions. In addition, in support of this salinity standard, the treaty limits the pumping of groundwater to 160 000 acre feet year$^{\bullet 1}$ country$^{\bullet 1}$ (an acre foot is 1233.5 m^3). The salinity standard is a quota-based standard, since there is a limit value set. The groundwater quota is a resource quota in support of a pollution goal.

Quotas are an essential part of tradeable permit schemes and joint implementation. In both cases initial allocations are made in the form of limits, but bilateral arrangements are permitted to enhance the cost-effectiveness of meeting the established limits.

Given the variety of treaty classification systems in use, it is difficult to estimate the overall incidence of quotas in environmental management. Some evidence is provided in a comprehensive review of the management provisions of 72 treaties concerned with fisheries and the conservation of other resources and 55 air and water pollution agreements (Wolf, 1997). This review showed that quotas were used in 33 conservation agreements. In these applications, quota units are typically individuals of a species, such as whales or elephants, or other biomass measures, such as a tonne of fish. Since agreements often serve several objectives, quotas are frequently combined with other measures. For example, fish size limits may be used in conjunction with quotas in order to protect reproductive-size stock preferentially.

Habitat protection measures may make use of a different sort of quota. The 1991 framework Convention for the Protection of the Alps suggests that imposing limits on leisure and tourist activities that harm the environment is one possible measure. The Convention on Biological Diversity (1992) calls for regulations to limit the introduction of living modified organisms. Such regulations could be expressed as quotas.

Quotas have had a prominent role in air pollution management. The acid rain and ozone layer agreements make extensive use of phased reductions, essentially providing each country, or source emitter, with an emission limit. The Framework Convention on Climate Change (FCCC, 1992) is underpinned by limits on greenhouse gases in the atmosphere. The 1997 Kyoto Protocol to the FCCC sets out a legally binding commitment for countries to meet country-specific quantitative emission limitation and reduction objectives for six greenhouse gases. The limits range from an increase of 10% to a decrease of 8% of 1990 levels to be achieved between 2008 and 2012.

In agreements to reduce water pollution, quotas in the form of ambient standards are used to allocate indirectly the costs of pollution abatement, rather than directly allocating a resource (as in conservation uses). This indirect approach contrasts with the direct allocation of resources in conservation applications. The 1991 revision to the Convention on the Protection of the Rhine Against Pollution by Chlorides establishes an ambient standard for chloride loading in the river ($15\,kg\,s^{-1}$), which may be increased whenever the concentration at the German-Dutch border exceeds $200\,mg\,l^{-1}$ of chloride ion (Bernauer, 1995). Ambient standards are also prominent in marine pollution agreements. Lists of banned and restricted substances form the basis of environmental management. Restricted substances (the gray list) in the Convention for the Protection of the Mediterranean Against Pollution (1979) and protocols, for instance, are to be managed by establishing ambient standards at various locations, depending on factors such as effects on human health, availability of waste technologies and characteristics of the receiving environment. Ambient standards, however, are costly to establish, and in practice may be approximated by rules on emissions. In these cases, the quota basis is not rendered irrelevant, since ambient standards remain an essential component in judging the effectiveness of a management regime. Moreover, because ambient and emission standards draw upon different normative assumptions, the negotiations surrounding a choice of mechanism may make use of the different bases. Emission standards, while technically and administratively simple, cannot take account of varying ambient conditions which may be linked to relevant variables, such as economic development, as the Mediterranean negotiations made clear (Bliss-Guest, 1981).

Over the last quarter of a century, environmental agreements have favored a multi-criterion approach as negotiators seek to balance the costs and environmental and distributional impacts of agreement terms. Once parties agree to jointly seek an environmental outcome (not an insignificant achievement) emphasis shifts to cost and distributional matters. It is in this context that the issue of instrument choice rises to the surface. If quotas are then judged to have a role, either alone or in conjunction with other regulations and incentives, negotiators must address the details of the quota in order to achieve an acceptable balance of effectiveness, efficiency, and equity.

Assuming quotas work as intended, are fully implemented and universally respected, and are not combined with other instruments, then some conclusions are possible about the various types of quotas with respect to the key criteria. Some of the conclusions hinge on the fact that quotas in international agreements must be negotiated giving rise to transaction costs and certain strategic behaviors with respect to setting up terms of agreements. Quotas that set a total limit are more effective than those that do not, since the limit guarantees an environmental effect. In this manner whale quotas (set in numbers of whales) are distinguished from the sulfur dioxide 30% club, since in the latter case the environmental effect is contingent on the number of participating countries. In the absence of an explicit total limit, negotiators' efforts to revise rules may be hampered by the difficulty in projecting the impacts of the existing rules. For instance, the impact of the Montreal Protocol on Substances that Deplete the Ozone Layer (1987) was contingent on the trend in CFC emissions in developing countries, where overall use was allowed to rise to a threshold 0.3 kg *per capita*.

The efficiency of a quota derives from the degree to which it both facilitates the rational use of resources and minimizes the costs of coming to a joint allocation of the regulated units. Thus, a total limit without allocated shares is inefficient, as parties overcapitalize in an effort to exploit the regulated resource before the limit is reached. If the quota mechanism establishes rules or a formula for allocating shares of the limit that stand from year to year, bargaining costs and the costs of delays are reduced. Parties are better able to anticipate their allotment for future years and plan accordingly. They are not required to bargain repeatedly to maintain or increase their share. Of course, if the initial allocation formula contains embedded inefficiencies, then it may ultimately be more efficient to bargain until shares find their most efficient owner.

It is not possible to generalize about the equity objective. This objective may be best met if the mechanism eschews global rules in favor of adjustments in allocations for relevant differences in parties' situations. Allocation formulas, for instance, may make it more difficult to make exceptions on equity grounds. On the other hand, if powerful parties are overly self-serving, the exceptions may ow in the wrong direction. Furthermore, equity is tied closely with effectiveness and efficiency. In the final analysis, only the specific context and details of a given quota mechanism can provide a basis for equity judgements. This is not to say that equity is ambiguous or that it is a wild-card in negotiations.

Experience shows that equity is increasingly integral to successful negotiations. A close examination of a sample of some of the most important treaties negotiated in the 1990s

reveals an emphasis on the precautionary and sustainable development principles (*see* Precautionary Principle, Volume 4; Sustainable Development Policy, Volume 4). Both principles require a balance of attention to the main criteria of effectiveness, efficiency, and equity. In addition, it is also important diplomatically to build in the highest extent of exibility feasible, in order to allow individual countries to best tailor their obligations to their domestic circumstances. A final feature of these recent exercises is the move to independently ground quotas (limit values) on the basis of an explicit balancing of criteria. These criteria cover ecology, economic development, and allocation of costs and benefits.

Most importantly, recent quotas exhibit a far greater degree of differentiation in units of account than earlier examples. The critical loads modeling used to help negotiators reach allocations of allowable sulfur emissions is the most clear-cut instance of this trend. There are three general features: (1) the establishment of management cells, such as a grid over space or the determinations of stocks and sub-stocks of living resources; (2) the establishment for each of these cells of a limit based, to the extent possible, on scientific knowledge; (3) the translation of the scientifically justified targets into management targets, for instance by setting a concrete target on the basis of a scientific range or by adjusting for technical feasibility. The management targets for different cells need not all be based on the same principles. The final determination of targets, which may be thought of as local quotas, is settled in a negotiation, which takes into account costs and distributions of costs. The end result is that scientific targets, modified for pragmatic reasons are approached as closely as possible in light of constraints set by factors such as economic development, costs of various technologies, as well as differences in parties' views of fairness. Essentially management targets are finalized only after significant debate on the relevance and effect of constraining factors.

Experienced negotiators will recognize the generic format of this procedure. Advances in information and modeling have, at least since the well-documented law of the sea (Sebenius, 1992) enabled the approach to be conducted transparently. Further refinements are possible with devices such as indexing (as used in setting ozone depleting potentials for various chemicals and allowing countries to assemble their own baskets). The approach also has conservation applications. Proposals have, at various times, been discussed for managing whales and elephants by subspecies stock and for establishing criteria for the sustainable management of tropical timber, which would allow for quota-based harvests where criteria are met.

In summary, quotas are an integral part of the international management of global environmental change. They are a focal solution in negotiations, based on precedent, on the direct links to environmental impacts, and on the lack of compelling alternatives. When the key features of quotas are borne in mind – rather than a crude caricature – quotas are revealed as essential components of a rich variety of environmental management instruments. As quantities, quotas are relatively simple to work with – akin to bargaining over a wage rate in labor negotiations – even if agreement is difficult. Quotas facilitate reporting and monitoring in a manner relevant for assessing environmental impacts. In the current globalized economy, quotas are ready currency, needing no exchange rate calculations. State parties to a quota agreement remain free to use a wide range of instruments to meet their obligations. Finally, quotas fit with the incremental and evolutionary process of global environmental management.

REFERENCES

Bernauer, T (1995) The International Financing of Environmental Protection: Lessons from Efforts to Protect the River Rhine Against Chloride Pollution, *Environ. Polit.*, 4, 369–390.

Bliss-Guest, P (1981) The Protocol Against Pollution from Land-based Sources: a Turning Point in the Rising Tide of Pollution, *Stanford J. Int. Law*, 17, 261–279.

Brownell, H and Eaton, S (1975) The Colorado River Salinity Problem with Mexico, *Am. J. Int. Law*, 69, 255–271.

Organization for Economic Cooperation and Development (OECD) (1994) *Managing the Environment: The Role of Economic Instruments*, OECD, Paris.

Sebenius, J (1992) *Negotiating the Law of the Sea*, Harvard University Press, Cambridge, MA.

Wolf, A (1997) *Quotas in International Environmental Agreements*, Earthscan, London.

R

Regional Air Pollution Information and Simulation (RAINS)

see RAINS (Regional Air Pollution Information and Simulation) (*Case study*, **Volume 4**)

Regional Responses to Global Environmental Change

see Regional Responses to Global Environmental Change (*Case study*, **Volume 4**)

Risk Management in an Era of Global Environmental Change

Howard Kunreuther
University of Pennsylvania, Philadelphia, PA, USA

Society faces large challenges as to how to deal with the increasing losses from natural, technological and environmental hazards. Residents in hazard-prone areas are reluctant to protect themselves before the event occurs. The insurance and reinsurance industry is concerned that insurers cannot provide protection against these catastrophic risks without exposing themselves to the danger of insolvency or signi cant loss of surplus. There has also been a concern that these losses will increase due to population growth in hazard prone areas and climate change creating more disasters than in the past.

This article explores the importance of insurance coupled with other policy instruments, such as regulations and standards, and can be used by interested parties (e.g., banks and nancial institutions) to encourage loss reduction measures and provide nancial protection against these hazards. New advances in information technology and risk assessment coupled with the emergence of new nancial instruments for covering large losses provide the ingredients for rethinking the way society deals with catastrophic risks. To encourage investment in cost effective loss-reduction measures, there is a need for well-enforced building codes as well as nancial incentives such as long-term loans tied to an insurance policy.

INTRODUCTION

There is grave concern, by the property and casualty insurance and reinsurance industry, that insurers cannot continue along their current path of providing protection against certain risks or taking on additional risks, without exposing themselves to the danger of insolvency or significant loss of surplus, which will threaten the availability of future coverage (Kunreuther and Roth, Sr, 1998).

More specifically, since 1989 insurance and reinsurance firms have suffered losses from disasters that have wreaked havoc with their balance sheets. Figure 1 depicts the magnitude of the catastrophic losses experienced by the insurance industry in the United States from 1949 to 1999. The drastic change from 1989 to 1999 is obvious. Prior to Hurricane Hugo in 1989 (where insured losses were over $4 billion), the insurance industry had never suffered any loss of over $1 billion from a single disaster. Since that time Hurricane Andrew in 1992 caused over $15 billion in insured losses and the Northridge earthquake in California resulted in over $12 billion in insurance claims Between 1989 and 1999 another 15 disasters exceeded $1 billion in insured losses (Swiss Re, 2000). Commercial development has followed the population's movement to coastal areas, and this has increased the potential economic losses from natural disasters in this part of the country.

The purpose of this article is to explore how insurance can be combined with other policy instruments, such as mitigation measures, building codes and well specified standards, to better manage safety and reduce losses from

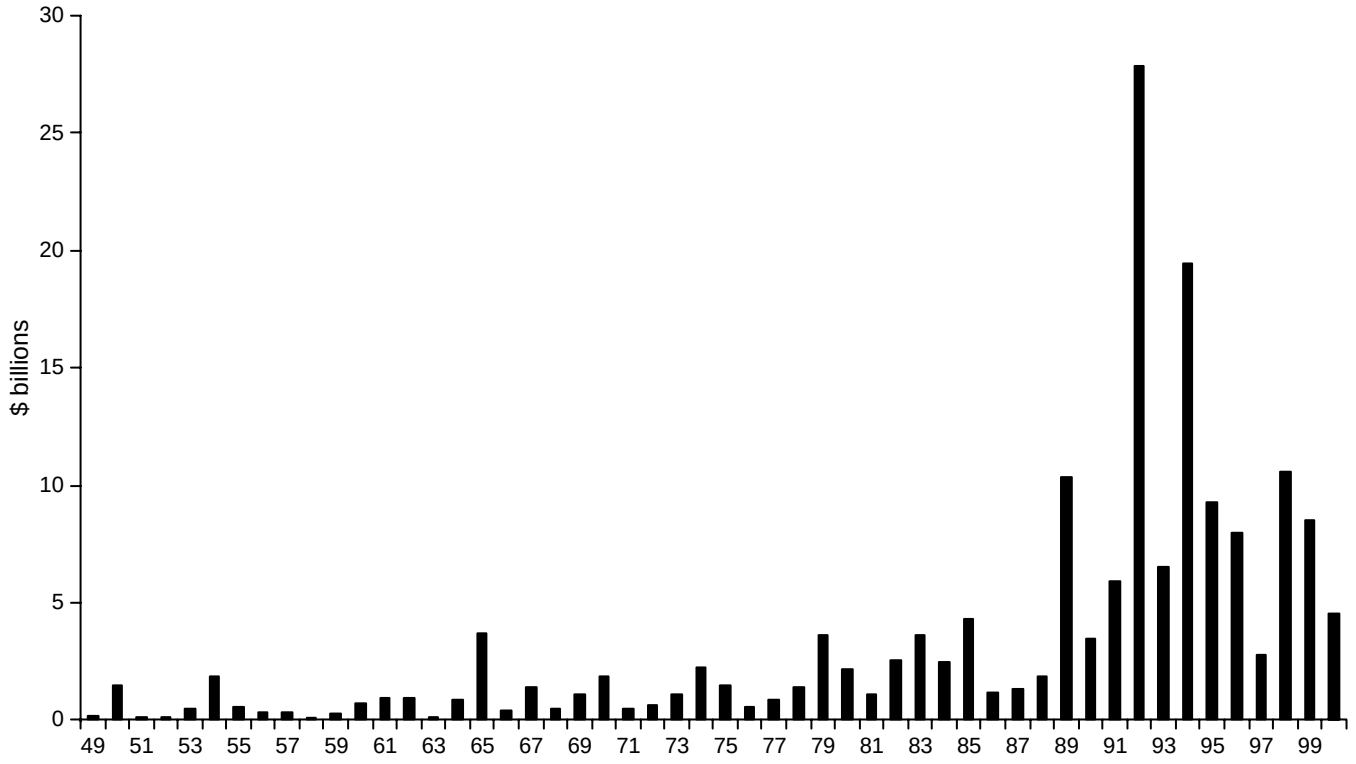


Figure 1 Insured catastrophe losses (1949–2000) (in 2000 dollars). (*Source:* Insurance Services Office)

catastrophic risks from technological and natural hazards. In particular, there is a need for a public–private partnership for dealing with these challenging issues. The principal focus will be on reducing losses from natural disasters, but the concepts and principles discussed below are relevant for dealing with safety systems for technological hazards. Three recent developments provide the ingredients for rethinking the way society deals with these risks:

1. There has been an emergence of new capital market instruments, such as catastrophe bonds for dealing with losses from mega-disasters. Although the volume of business to date in these instruments is relatively small, they offer promise for protection in the future, both in the developed countries as well as emerging economies (Insurance Services Office, 1999; Standard and Poors, 2000).
2. New advances in information technology (IT) and risk assessment offer an opportunity to estimate the chances and potential losses of future disasters and catastrophic events more accurately than in the past. On the IT side, the development of faster and more powerful computers enables one to examine extremely complex phenomena in ways that were impossible even five years ago. Scientific advances in risk assessment have reduced the uncertainty associated with predicting the chances and consequences of these low probability–high consequence (LP–HC) events.
3. Mitigation measures promise to reduce losses from natural disasters and catastrophic accidents. In particular, there are many benefits, which have traditionally not been considered as part of standard benefit cost analysis. These benefits may make mitigation extremely attractive to all of the concerned parties (Heinz Center, 1999; Mileti, 1999).
4. Global mean surface temperature has increased by between about 0.3 and 0.6 °C since the late 19th century. The change is unlikely to be entirely natural in origin and suggests a noticeable human in uence, although there are uncertainties in key factors, including the magnitude and patterns of long-term natural variability. Global sea level has risen by between 10 and 25 cm over the past 100 years and much of the rise may be related to the increase in global mean temperature (Intergovernmental Panel on Climate Change, 1995). In its Third Scientific Assessment (IPCC, 2001), the Intergovernmental Panel on Climate Change (IPCC) concluded that there has been a discernible human in uence on global climate . This is the strongest official pronouncement yet that human-induced warming is real (Kerr, 2000) (*see* Intergovernmental Panel on Climate Change (IPCC): an Historical Review, Volume 4).

FRAMEWORK FOR ANALYZING CATASTROPHIC RISKS: THE CASE OF NATURAL DISASTERS

To analyze the impact of catastrophic risks on society it is important to consider the following elements:

Population and Property at Risk

One way to characterize who is at risk is to construct a community or region consisting of homes, businesses and other properties, which are subject to future disasters. One needs to know the design of each structure, whether specific mitigation measures are in place or could be utilized, and the property's location in relation to the hazard (e.g., distance from an earthquake fault line or proximity to the coast in a hurricane-prone area) as well as other risk-related factors.

There has been a large in ux of residents and businesses into hazard prone areas in recent years. By the year 2010 the population in regions of the Southeast Atlantic Coast of the United States, having annual hurricane probabilities of 10% or more, is projected to increase by 23% from 64 to 73 million people compared with 14% for the entire United States (Mills *et al.*, 2000). The reasons often given for moving to these areas is because they are seen as desirable places to live. This may be true in many developed countries. In emerging economies, such as Bangladesh, people are forced to live in these high-risk areas because there is no other place for them to go (Etkin, 1999).

Risk Assessment and Potential Damage

The potential damage to a community is determined by estimating the probability that disasters of specific magnitudes will occur and specifying the resulting losses to structures in harm's way. Relatively sophisticated models of catastrophic losses have been developed in recent years to model these distributions (Dong *et al.*, 1996). There is still considerable uncertainty as to what impact climate change will have on future disasters and their magnitudes. For example, with respect to hurricanes, Pielke and Landsea (1998) found no trend in damages when the data were corrected for in a- tion, population in ux to these areas and changes in wealth. On the other hand, climate models suggest there is likely to be an increase in ooding events due to more precipitation and greater atmospheric absolute humidity (White and Etkin, 1997).

Key Stakeholders

There are a number of interested parties concerned with natural disaster damage. Government agencies such as the Federal Emergency Management Agency (FEMA) in the United States, have stressed the importance of building codes and enforcement of regulations to reduce losses from natural disasters. The primary insurance companies provide protection to residential and commercial sectors for losses such as those caused by fires (including those resulting from earthquakes) and wind damage from tornadoes and hurricanes. In the United States this coverage is offered through the standard homeowners' policies normally required as a condition for a mortgage, and through commercial multi-peril policies. The types of policies offered and the nature of the requirements vary from country to country.

Reinsurers relate to insurers in the same manner that insurers do to property owners. They provide protection to primary insurers by insuring a portion of their claims in exchange for a premium. For all but the largest insurance companies, reinsurance is a prerequisite to offering insurance against natural disasters when there is a potential for catastrophic losses. As pointed out above, the capital markets have recently provided private insurers access to funds in the form of catastrophe bonds. The insurer borrows from investors or an institution at higher than normal interest rates to cover extreme losses from hurricanes and earthquakes that exceed a trigger amount. If this amount is exceeded, then the interest on the bond, the principal, or both, are forgiven. For more details on these catastrophe bonds and other financing and hedging instruments, see Doherty (1997).

A number of other interested parties play key roles in the design and enforcement of insurance and damage mitigation requirements. These parties include financial institutions, through specific conditions for obtaining a mortgage; the construction industry, by designing safer structures; state and local governments, through ordinances and building codes; and the real estate sector through provision of information on hazards to potential buyers and owners.

Decision Processes

An understanding of the decision processes of the property owner and insurer provides guidance for developing a set of strategies for reducing losses from future natural hazards. If residents in hazard-prone areas are reluctant to voluntarily adopt mitigation measures that are cost effective, then this provides a rationale for developing and enforcing building codes. The involvement of municipalities, contractors, third-party inspectors and other key stakeholders is essential if this measure is to work. Similarly if insurers are willing to give premium discounts for adopting mitigation measures, then there is an opportunity for banks and financial institutions to partner with the insurers to develop a financially attractive mitigation loan/insurance package.

Policy Implications

One question that needs to be addressed in developing strategies for managing catastrophic risks is the appropriate

roles of the private and public sectors, in financing the cost of large-scale disasters. To the extent that private insurance markets provide protection against catastrophe risk, policymakers must decide how these markets should be regulated. They must also determine the role of regulations (e.g., restricting developments through land-use restrictions) and standards (e.g., well-enforced building codes) and the extent to which private choice and incentives will guide hazard mitigation efforts. Within the realm of public choice, decisions also must be made with respect to the delegation of authority among the different levels of government and its agencies. In evaluating these options, policymakers must consider how various governmental actions affect the behavior of firms and individuals in responding to catastrophe risk.

WHY IS THERE LIMITED INTEREST IN MITIGATION?

The decision on whether to adopt risk mitigation measures (RMMs) to reduce losses from catastrophic accidents or natural disasters can be framed in a very straightforward manner. There is an upfront cost of the mitigation measure which will either reduce the probability of the accident or disaster occurring and/or the magnitude of the loss should the event take place. The benefits of the mitigation measure will be reaped for the length of time that the property is in place. In reality there are several reasons why a firm or individual will decide not to invest in such measures. Some individuals may perceive the probability of a disaster causing damage to their property as being sufficiently low that the investment in the protective measure will not be justified. Individuals may have relatively short time horizons over which they want to recoup their investment. Even if the expected life of the house is 25 or 30 years, the person may only look at the potential benefits from the mitigation measure over the next 3 to 5 years.

If people have budget constraints, then they will be averse to investing in the upfront costs associated with protective measures simply because they feel they cannot afford these measures. It is not unusual for one to hear the phrase We live from payday to payday when asked why a household has not invested in protective measures. Individuals may have little interest in investing in protective measures if they believe that they will be financially responsible for only a small portion of their losses should a disaster occur. For example, if residents anticipate liberal disaster relief from the government should they suffer damage, they would have less reason to invest in an RMM.

The empirical data on studies of mitigation adoption in hazard-prone areas of the United States suggest that individuals are not willing to invest in RMMs despite the rather large damage that either they and/or their friends and neighbors suffered from recent disasters. For example, after Hurricane Andrew in Florida in 1992, most residents in hurricane-prone areas appear not to have made cost effective improvements to existing dwellings that could reduce amount of damage from another storm at relatively low cost. A July 1994 telephone survey of 1241 residents in six hurricane-prone areas along the Atlantic and Gulf Coasts revealed that 62% of the respondents had not installed hurricane shutters, used laminated glass in windows, installed roof bracing, and/or made sure that side walls were bolted to the foundation either before or after Hurricane Andrew (Insurance Institute for Property Loss Reduction, 1995).

Measures, such as strapping a water heater with simple plumbers tape, can normally be undertaken by property owners at a cost of under $5 in materials and one hour of their own time (Levenson, 1992). This RMM can reduce damage by preventing the heater from toppling during an earthquake, creating gas leaks and causing a fire. Yet these and other mitigation investments are not being adopted by residents in earthquake-prone areas in the United States. A 1989 survey of 3500 homeowners in four California counties subject to the hazard reported that only between 5 and 9% of the respondents reported adopting any loss reduction measures (Palm *et al.*, 1990). This behavior suggests that individuals do not believe that investing in the RMM will increase their residence's property value, or that they have either short time horizons and/or severe budget constraints, which either reduce their perceived net benefits from RMMs or simply prevent them from making the investment (Kunreuther, 1996).

Turning to the relationship between insurance and mitigation, some interesting findings emerge from recent surveys undertaken by Palm and Carroll (1998) who report that those who had adopted mitigation measures were also more likely to buy earthquake insurance than those who had not taken these loss reduction measures. This raises the interesting question as to whether certain types of individuals and managers want protection for reasons that have less to do with their perception of the risk than their intrinsic worries and concerns.

IMPROVING THE ESTIMATES OF THE RISK

In setting rates for catastrophic risks, insurers have traditionally looked backwards by relying on historical data to estimate future risks. Such procedures are likely to work well only if there is a large database of past experience, which forms the basis for extrapolation into the future, and if the time series does not contain a trend. LP–HC events generally have a relatively small historical database. In addition, many technological and environmental risks are associated with new processes, so that past performance data are lacking. One thus has to rely on scientific modeling and epidemiological data to estimate these risks.

Fortunately there is considerable scientific work undertaken in the areas of natural, technological and environmental hazards to provide estimates of the probabilities and consequences of events of different magnitudes. A discussion of recent advances in seismology and earthquake engineering can be found in Federal Emergency Management Agency (1994) and Office of Technology Assessment (1995). Regarding technological hazards, the Wharton Risk Management and Decision Processes Center is now compiling a comprehensive database on the impact of large-scale catastrophic accidents on health and safety risks (Kleindorfer *et al.*, 1997). With respect to environmental risks to health, such as groundwater contamination, databases have been assembled which open up opportunities for providing insurance protection on risks that recently had previously been considered uninsurable by firms in the industry (Freeman and Kunreuther, 1997). As we better understand the effects that climate change is likely to have on the frequency and magnitude of natural hazards, then these impacts need to be incorporated in future risk estimates (see Mills *et al.*, 2000 for the current perspective of the US Insurance Industry on climate change).

Today there are growing numbers of catastrophe models that have been utilized to generate data on the likelihood and expected damage to different communities or regions from disasters of different magnitudes or intensity. Each model uses different assumptions, different methodologies, different data and different parameters in generating its results. Hence, there is need for a better understanding as to why these models differ and for attempts to reconcile these differences in a more scientific manner than has been done up until now (see Insurance Services Office, 1996).

BROADENING PROTECTION AGAINST CATASTROPHIC LOSSES

New sources of capital from the private and public sectors could provide insurers with funds against losses from catastrophic events, which would alleviate insurers' concerns that the next major disaster might leave them insolvent. Recently investment banks and brokerage firms have shown considerable interest in developing new financial instruments for protecting against catastrophic risks. Their objective is to find ways to make investors comfortable trading new insurance-linked secured instruments covering catastrophic exposures, just like the securities of any other asset class. In other words, catastrophe exposures would be treated as a new asset class (Insurance Services Office, 1999). A description of recent trends in this market can be found in Lane and Beckwith (2000) and Standard and Poors (2000).

Turning to the role of the public sector, Lewis and Murdoch (1996) developed a proposal that the federal government offer catastrophe reinsurance contracts, which would be auctioned annually. Another option is for the government to provide reinsurance protection against catastrophic losses. Private insurers would build up the fund by being assessed premium charges in the same manner that a private reinsurance company would levy a fee for excess loss coverage or other protection.

POLICY IMPLICATIONS: NEED FOR A PUBLIC–PRIVATE PARTNERSHIP

The key to a successful hazard management program is to design ways for reducing losses while providing protection should a disaster occur. The empirical data suggest that there is a need for new public–private partnerships to address this problem. Today most property owners have limited interest in investing in loss reduction measures. Insurers have little reason to encourage mitigation in hazard-prone areas since they feel that the rates they are allowed to charge are inadequate. Hence they want to charge the same premiums with and without mitigation and hope that some policyholders will decide not to renew their insurance policies (Kunreuther and Roth, 1998).

Three types of public–private partnership programs are proposed that can reduce losses from future disasters based on our understanding of the decision processes of the relevant interested parties. In order to successfully implement these programs, there needs to be a better understanding of the nature of the risks associated with natural hazards.

Importance of Building Codes

Building codes mandate that property owners adopt mitigation measures. Such codes may be desirable, when property owners would otherwise not adopt cost-effective RMMs because they either misperceive the benefits from adopting the RMM and/or underestimate the probability of a disaster occurring.

Cohen and Noll (1981) provide an additional rationale for building codes. When a building collapses, it may create externalities in the form of economic dislocations and other social costs that are beyond the economic loss suffered by the owners. These may not be taken into account when the owners evaluate the importance of adopting a specific mitigation measure. For example, if a building topples off its foundation after an earthquake, it could break a pipeline and cause a major fire that would damage homes not affected by the earthquake in the first place. In other words, there may be an additional annual expected benefit from mitigation over and above the reduction in losses to the specific structure adopting this RMM.

One way to encourage the adoption of cost-effective mitigation measures is for banks and financial institutions to provide a seal of approval to each structure that meets

or exceeds building code standards. Upon receipt of that certificate, there would be a set of incentives provided by banks (e.g., lower mortgage rates) as well as insurers (e.g., lower premiums). The success of such a program requires the support of the building industry and a cadre of qualified inspectors to provide accurate information as to whether existing codes and standards are being met. Insurers may want to limit coverage only to those structures that are given a certificate of disaster resistance. For more details on ways to make communities disaster-resistant, see Central US Earthquake Consortium (1997).

Premium Reductions Linked with Long-term Loans

Premium reductions for undertaking loss prevention actions can be an important first step in encouraging property owners to adopt these measures. The basic rule in this case is a simple one: if the premium reduction is less than the savings in expected claim payments due to mitigation, it is a desirable action for the insurer to promote. If homeowners are reluctant to incur the upfront cost of mitigation due to budget constraints, then one way to make this measure financially attractive to the property owner is for the bank to provide funds for mitigation through a home improvement loan, with a payback period identical to the life of the mortgage. The cost of the loan will then be less than the reduction in insurance premiums if the mitigation measure is a cost-effective one with respect to reducing property damage (Kunreuther, 1997).

Many poorly constructed homes are owned by low-income families, who cannot afford the costs of mitigation measures on their existing structure or the costs of reconstruction should their house suffer damage from a natural disaster. Equity considerations argue for providing this group with low interest loans and grants for the purpose of adopting cost-effective RRMs or for them to relocate to a safer area. Since low-income victims are likely to receive federal assistance after a disaster, subsidizing these mitigation measures can also be justified on efficiency grounds.

Lower Deductibles Tied to Mitigation

An alternative way to encourage consumers to adopt mitigation measures is to change the nature of their insurance coverage rather than reducing the premium. More specifically, the insurer could offer a lower deductible to those who adopt mitigation at the same or lower price than if they had decided not to invest in the RMM. Such a program is likely to be very attractive given the empirical and experimental evidence that suggests that consumers appear to dislike deductibles even though they offer considerable savings in premiums (Johnson *et al.*, 1993).

CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH

This article makes the case for a new approach for dealing with catastrophic accidents and disasters that takes advantage of recent developments in IT and the emergence of new financial instruments to deal with non-diversifiable risks. These two major changes open up opportunities for residents and firms to undertake cost-effective loss protection measures while at the same time providing a financial cushion to insurers concerned with catastrophic losses that will lead to insolvency.

A strategy for undertaking research in this area suggests analyzing the impact of disaster or accidents of different magnitudes on a set of structures, industrial plants or their equipment. In order to determine expected losses and the maximum probable losses arising from worst case scenarios, it may be necessary to undertake long-term simulations. For example, one could examine the impacts of earthquakes or hurricanes of different magnitudes and intensities on the losses to a community or region over a 10 000 year period. In the process one could determine expected losses based on the probabilistic scenarios for each of these disasters as well as the maximum credible loss during this period based on a worst-case scenario. By constructing large, medium and small representative insurers with specific balance sheets, types of insurance portfolios, premium structures and a wide range of potential financial instruments, one could examine the impact of different disasters and accidents on the insurer's profitability, solvency and performance through a simulation. Examples of the application of such an approach to a model city in California facing an earthquake risk and one in Florida facing a hurricane risk can be found in Kleindorfer and Kunreuther (1999a,b).

There is considerable uncertainty in estimating the probability of disasters of different magnitudes occurring and the magnitude of the resulting losses. Future studies could examine how one can incorporate these uncertainties in an analysis as to which mitigation measures are cost effective and how to price new financial instruments. Will mitigation reduce the uncertainty of future losses, so that these new financial instruments could be more easily marketed to investors?

This is a very exciting time for key stakeholders in the public and private sectors to explore new opportunities for dealing with catastrophic risks. If insurance can be used as a catalyst to bring other interested parties and new financial instruments to the table, it will have served an important purpose in helping both the insurance industry and society deal with the critical issue of reducing future losses, and providing protection against natural, technological and environmental disasters.

See also: Natural Hazards, Volume 3; Natural Hazards: Social and Policy Responses, Volume 4; Hazards in Global Environmental Change, Volume 5.

ACKNOWLEDGMENTS

Support from NSF Grant #CMS97-14401 to the Wharton Risk Management and Decision Processes Center at the University of Pennsylvania is gratefully acknowledged. Portions of this paper incorporate material from Kunreuther (2000). Kunreuther is the Cecilia Yen Koo Professor of Decision Sciences and Public Policy, Department of Operations and Information Management, The Wharton School, University of Pennsylvania.

REFERENCES

Central US Earthquake Consortium (CUSEC) (1997) *Disaster Resistant Communities*, CUSEC, Memphis, TN.

Cohen, L and Noll, R (1981) The Economics of Building Codes to Resist Seismic Shocks, *Public Policy*, Winter, 1–29.

Doherty, N (1997) Financial innovation for financing and hedging catastrophe risk, presented at the Fifth Alexander Howden Conference on Disaster Insurance, Gold Coast, Australia.

Dong, W, Shah, H, and Wong, F (1996) A Rational Approach to Pricing of Catastrophe Insurance, *J. Risk Uncertainty*, 12, 201–218.

Etkin, D (1999) Risk Transference and Related Trends: Driving Forces Towards more Mega-disasters, *Environ. Hazards*, 1, 69–75.

Federal Emergency Management Agency (1994) *Assessment of the State-of-the Art Earthquake Loss Estimation*, National Institution of Building Sciences, Washington, DC.

Freeman, P and Kunreuther, H (1997) *Managing Environmental Risk Through Insurance*, American Enterprise Institute, Washington, DC.

Heinz Center for Science, Economics, and the Environment (1999) *The Hidden Costs of Coastal Hazards: Implications for Risk Assessment and Mitigation*, Island Press, Washington, DC.

Insurance Institute for Property Loss Reduction (IBHS) (1995) *Homes and Hurricanes: Public Opinion Concerning Various Issues Relating to Home Builders, Building Codes and Damage Mitigation*, IBHS, Boston, MA.

Insurance Services Office (1996) *Managing Catastrophic Risk*, Insurance Services Office, New York.

Insurance Services Office (1999) *Financing Catastrophe Risk: Capital Market Solutions*, Insurance Services Office, New York.

Intergovernmental Panel on Climate Change (IPCC) (1995) *IPCC Second Assessment Synthesis of Scienti c-Technical Information relevant to interpreting Article 2 of the UN Framework Convention on Climate Change*, IPCC, Geneva, Switzerland.

Johnson, E, Hershey, J, Meszaros, J, and Kunreuther, H (1993) Framing, Probability Distortions, and Insurance Decisions, *J. Risk Uncertainty*, 7, 35–52.

Kerr, R A (2000) Draft Report Affirms Human In uence, *Science*, 288, 589–590a.

Kleindorfer, P, Lowe, R, and Rosenthal, I (1997) *Major Event Analysis in the United States Chemical Industry: Proposed Studies Using the EPA RMP• Info Data Base*, Wharton Risk Management and Decision Process Center, University of Pennsylvania, PA.

Kleindorfer, P and Kunreuther, H (1999a) The Complementary Roles of Mitigation and Insurance in Managing Catastrophic Risks, *Risk Anal.*, 19, 727–738.

Kleindorfer, P and Kunreuther, H (1999b) Challenges Facing the Insurance Industry in Managing Catastrophic Risks, in *The Financing of Property/Casualty Risks*, ed K Froot, University of Chicago Press, Chicago, IL.

Kunreuther, H (1996) Mitigating Disaster Losses Through Insurance, *J. Risk Uncertainty*, 12, 171–187.

Kunreuther, H (1997) Rethinking Society's Management of Catastrophic Risks, *Geneva Pap. Risk Insur.*, 83, 151–176.

Kunreuther, H (2000) Insurance as a Cornerstone for Public–Private Sector Partnerships, *Nat. Hazards Rev.*, 1, 126–136.

Kunreuther, H and Roth, Sr, R, (1998) *Paying the Price: The Status and Role of Insurance Against Natural Disasters in the United States*, Joseph Henry Press, Washington, DC.

Lane, M and Beckwith, R (2000) *Trends in the Insurance-Linked Securities Market*, Lane Financial L L C, Trade Notes, see http://www.lanefinanciallc.com.

Levenson, L (1992) Residential Water Heater Damage and Fires Following the Loma Prieta and Big Bear Lake Earthquakes, *Earthquake Spectra*, 8, 595–604.

Lewis, C and Murdock, L (1996) The Role of Government Contracts in Discretionary Reinsurance Markets for Natural Disasters, *J. Risk Insur.*, 63, 567–597.

Mileti, D (1999) *Disasters by Design: A Reassessment of Natural Hazards in the United States*, Joseph Henry Press, Washington, DC.

Mills, E, Lecomte, E, and Peara, A (2000) *US Insurance Industry Perspectives on Global Climate Change*, Lawrence Berkeley National Laboratory Report No. 45185, University of California, Berkeley, CA (mimeo).

Office of Technology Assessment (1995) *Reducing Earthquake Losses*, USGPO, Washington, DC.

Palm, R and Carroll, J (1998) *Illusions of Safety: Cultural and Earthquake Hazard Response in California and Japan*, Westview Press, Boulder, CO.

Palm, R, Hodgson, M, Blanchard, R D, and Lyons, D (1990) *Earthquake Insurance in California: Environmental Policy and Individual Decision Making*, Westview Press, Boulder, CO.

Pielke, R and Landsea, C (1998) Normalized Hurricane Damages in the United States, *Weather and Forecasting*, 13, 351–361.

Standard and Poors (2000) *Sector Report - Securitization*, June.

Swiss Re (2000) *Natural Catastrophes and Man-made Disasters in 1999*, Sigma Report No. 2/2000, Swiss Re, Zurich.

White, R and Etkin, D (1997) Climate Change, Extreme Events and the Canadian Insurance Industry, *Nat. Hazards*, 16, 135–163.

S

Salinity, Waterlogging and Agriculture: Policy Perspectives

see Salinity and Agriculture (**Volume 3**)

Saskawa Environment Prize

see United Nations Environment Programme (UNEP) Saskawa Environment Prize (**Volume 4**)

SCEP (Study of Critical Environmental Problems)

International environment assessments have become important policy tools in recent years. See, for example, the various Intergovernmental Panel on Climate Change (IPCC) assessments. The SCEP (1970) and the Study of Man's Impact on Climate (SMIC) (1971) (*see* SMIC (Study of Man s Impact on Climate), Volume 4) were forerunners of this approach with respect to large-scale long-term issues, and are of considerable historical interest. A principal motivation for these particular studies was to provide delegates to the Stockholm Conference on the Human Environment with syntheses on these emerging scientific issues.

Recognizing the need for focussing international attention on the global environment, Professor Carroll Wilson of the Massachusetts Institute of Technology (MIT) organized a conference to consider some of the more critical environmental problems of the world. The month-long meeting took place in Williamstown, MA, in July 1970, and involved some 40 scientists and professionals. The 319 page report of the conference was published by the MIT Press (SCEP, 1970).

The focus of the study was on global problems, and was primarily concerned with the effects of pollution on mankind through changes in climate, ocean ecology, and in large terrestrial ecosystems. In order to assess these effects it was clear that an extensive data base should be prepared covering all types of data – scientific, technical, economic, industrial, and social. Such a data base was assembled, and (though it was considered deficient in many respects) it marked a major step forward in this kind of study.

The final report summarizes, among other things, the climatic impacts of mankind's activities, some of the ecological effects, the impacts of industrial products and pollutants, problems of disposal of domestic and agricultural wastes, and energy consumption. The part that seemed to have attracted the most public attention was the possibility of weakening the protective ozone layer in the stratosphere by large eets of high ying (supersonic) aircraft.

Throughout the report there is an expression of the dire need for more good information and further analysis, and it stresses the importance of monitoring the many factors that govern mankind's existence on our planet.

REFERENCE

SCEP (1970) *Man's Impact on the Global Environment: Report of the Study of Critical Environmental Problem (SCEP)*, sponsored by the Massachusetts Institute of Technology, MIT Press, Cambridge, Massachusetts, and London, 1–319.

WILLIAM W KELLOGG *USA*

Scientific Responses in an Era of Global Environmental Change

see Scientific Responses in an Era of Global Environmental Change (*Opening essay*, **Volume 4**)

SCOPE (Scientific Committee on Problems of the Environment)

SCOPE, is one of the scientific committees of the International Council for Science (ICSU) (*see* ICSU (International Council for Science), Volume 4). The mandate of SCOPE is to assemble, review and assess the information available on human-induced environmental changes and the effects of these changes on people and the biosphere; to assess and evaluate the methodologies of measurement of environmental parameters; to provide an intelligence service on current research; and by the recruitment of the best scientific information and constructive thinking, to establish itself as a corpus of informed advice for the benefit of centers of fundamental research and of organizations and agencies operationally engaged in studies of the environment. SCOPE acts as the interface between the science and the decision-making spheres, providing analytical tools to promote sound management and policy practices. Many of the SCOPE syntheses of major environmental issues were published by John Wiley & Sons Ltd., Chichester, UK. In the context of global environmental change, SCOPE 29 (The Greenhouse Effect, Climatic Change and Ecosystems, 1986) provided the intellectual framework for many of the policy debates on greenhouse gas emissions that have taken place over the last decade. Similarly, SCOPE 55 (Functional Roles of Biodiversity: A Global Perspective, 1996) has been a basic reference in the development of the United Nations Convention on Biological Diversity.

At present, representatives of 40 member countries and 22 International Unions and Scientific Committees participate in the work of SCOPE, with scientists from around the world contributing their time and expertise to the projects in the scientific program. SCOPE is governed by a General Assembly, which meets every 3 years. Between Assemblies, SCOPE's activities are directed by an Executive Committee, and managed by an Executive Director, whose office is at 51 Boulevard de Montmorency, 75016 Paris, France (Fax: +33 1 4288 1466, E-mail: secretariat@icsu-scope.org).

R E MUNN *Canada*

Seismic Risk (and Risk Assessment)

Chen Yong
China Seismological Bureau, Beijing, China

A key question that must be addressed in earthquake disaster reduction is: how much loss might a city or a region suffer in future earthquakes? There are two components comprising the basic structure of a loss estimate study: (1) the seismic hazard analysis, involving the identi cation and quantitative description of strong ground motion caused by the future earthquakes, and (2) the seismic risk analysis, involving the vulnerability analysis of buildings and other facilities to earthquake damage and the losses that may result from this damage.

Seismic hazard (Box 1) describes the potential for dangerous, earthquake-related natural phenomena such as ground shaking, fault rupture, or soil liquefaction. These phenomena could result in adverse consequences to society, such as the destruction of buildings or the loss of life. *Seismic risk* is the probability of occurrence of these consequences. The output of a seismic hazard analysis could be a description of the intensity of shaking of a nearby magnitude-eight earthquake, or a map which shows the level of ground shaking in various parts of the country that have an equal chance of being exceeded. The output of a seismic risk analysis could be the probability of damage (in some unit of currency) from a nearby magnitude-eight earthquake, or the probability of fatalities due to seismically induced nuclear power-plant accidents. Seismic risk is a probabilistic expression of the product of seismic hazard

Box 1 Definitions of seismic hazard, vulnerability and seismic risk

Earthquake hazard is defined as the probability that a certain value of a macroscopic intensity or of a ground motion parameter (i.e., particle acceleration, velocity and displacement) will not be exceeded at a specific site in a specific period of time.

Vulnerability is the expected degree of loss within a defined area resulting from the occurrence of earthquakes, vulnerability is expressed on a scale from 0 (no damage) to 1 (full damage).

Seismic risk is the expected degree of losses caused by earthquakes, and therefore the product of seismic hazard and vulnerability:

$$\text{Risk} = \text{Hazard} \times \text{Vulnerability}$$

Elements at risk include the number of persons, the value of properties, the level of economic activities such as public service *et al.*, and so on in a given area.

and its consequences. One needs to know the seismic hazard in order to calculate the seismic risk. If not already known, defining the seismic hazard becomes part of the risk estimation process.

Quantification of earthquake disasters is the basis for the reduction of natural disasters and the risk management of which the seismic hazard analysis and the seismic risk assessment are the key scientific problems (Figure 1).

Most earthquake loss studies use an inventory approach in which predicted damages in various categories of structures and facilities in a concerned region are estimated and added together to obtain the total estimated loss for particular intensity ranges (ATC, 1985). Such an approach requires a detailed inventory database of the structures and facilities in the region, which is not always readily available in many regions of the world. There is an alternative means of estimating earthquake losses based on several macroeconomic indices such as the gross domestic product (GDP) and population (Chen *et al.*, 1998). Based

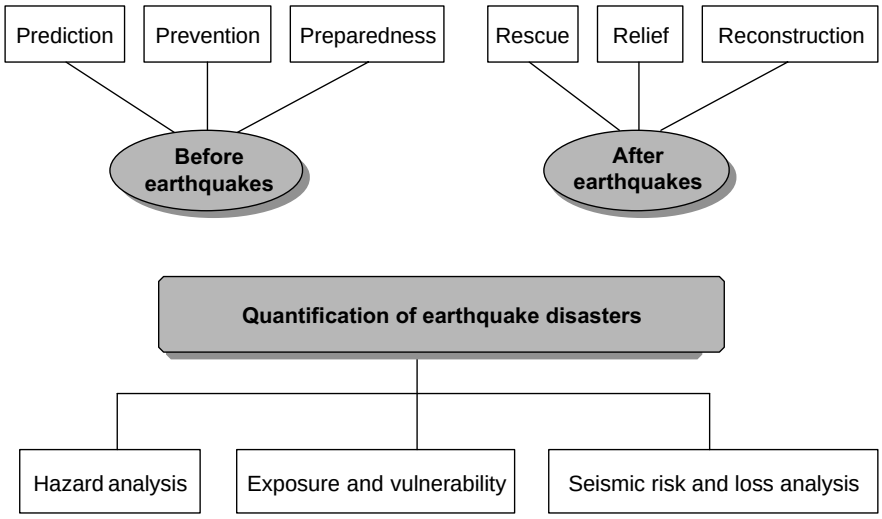


Figure 1 Quantification of earthquake disaster is essentially a scientific problem

Table 1 Twenty countries or regions of the world with the highest expected loss in the next 50 years (exceeding 10% probability) (population and GDP data are from a report of the World Bank (1998))

Region	Population in 1996 (million)	GDP in 1996 (billion US$)	Loss (billion US$)	Loss per year (billion US$)
World	5678	28 584	1181	22.6
Japan	126	4600	354	7.1
China (Mainland)	1218	805	125	2.5
United States	265	7342	98	2.0
Indonesia	198	226	49	1.0
Italy	57	1208	40	0.8
Greece	10	123	40	0.8
Mexico	97	335	29	0.6
Philippines	72	84	22	0.4

Table 1 (*continued*)

Region	Population in 1996 (million)	GDP in 1996 (billion US$)	Loss (billion US$)	Loss per year (billion US$)
Turkey	63	181	19	0.4
Chile	14	52	18	0.4
Argentina	35	294	18	0.4
India	939	356	18	0.4
Colombia	34	64	17	0.3
Pakistan	134	65	16	0.3
Russia	148	440	16	0.3
Peru	24	61	15	0.3
Iran	61	111	11	0.2
Venezuela	23	67	9	0.2
Afghanistan	21	10	7	0.2

on the published earthquake loss data during 1980–1995, the relations between GDP and earthquake loss have been formulated empirically for several intensity ranges. The world's land surface was divided into unit cells $0{\bullet}5 \times 0{\bullet}5^{\circ}$ in size; then the GDP of each cell was apportioned based on its population and the GDP and population of the region to which it belongs. The predicted seismic loss of the cell was then estimated from the seismic hazard probability function, its GDP, and the empirical relation between GDP and seismic loss. Employing readily available socioeconomic data as the basis for the vulnerability analysis, the method enables us to obtain seismic loss estimates for regions without a detailed inventory of exposed structures or the required collateral geological information. Seismic loss estimates can also be upgraded easily with socioeconomic data for the rapidly developing areas of the world.

Table 1 shows the expected earthquake loss in selected countries or regions. In the last two decades, property damage is estimated at \$25–100 billion resulting from natural disasters (including earthquakes). Table 1 indicates that the average earthquake loss rate in the future 50 years is about US\$20 billion year^{-1}, which is much higher than the loss in the past two decades. This is because the world today is becoming increasingly vulnerable to the threat of natural hazards. Rapid growth is bringing with it increasing complexity, more extensive systems of infrastructure, dense populations and excessive demands on the immediate environment and essential natural resources.

REFERENCES

Applied Technology Council (ATC) (1985) *Earthquake Damage Evaluation Data for California (ATC-13)*, Applied Technology Council, Redwood City, CA.

Chen, Y, Chan, L S, Chen, Q, Liu, J, Dong, W, and Shah, H (1998) Assessment of Global Seismic Loss Based on Macroeconomic Indicators, *Nat. Hazards*, 17, 269–283.

World Bank (1995) *World Development Report 1994: Infrastructure for Development*, Oxford University Press, New York, 1–254.

Self-organizing Phenomena

see Monitoring in Support of Policy: an Adaptive Ecosystem Approach (*Opening essay*, **Volume 4**)

SIDS (Small Island Developing States)

see Sustainable Development of Energy Resources in Small Island Developing States (SIDS) (**Volume 4**)

SMIC (Study of Man's Impact on Climate)

International environment assessments have become important policy tools in recent years. See, for example, the various Intergovernmental Panel on Climate Change (IPCC) assessments. The Study of Critical Environmental Problems (SCEP) (1970) (*see* SCEP (Study of Critical Environmental Problems), Volume 4) and the SMIC (1971) were important forerunners of this approach with respect to large-scale long-term issues, and are of considerable historical interest. A principal motivation for these particular studies

was to provide delegates to the Stockholm Conference on the Human Environment (held in 1972) with syntheses on these emerging scientific issues.

Following the highly successful SCEP in 1970, it was decided to hold a more in-depth study of a special problem that was of growing concern, inadvertent climate modification due to mankind's increasing activities. It was organized by Professor Carroll I Wilson of the Massachusetts Institute of Technology (MIT) and hosted by the Royal Swedish Academy of Sciences at its conference center in Wijk, near Stockholm. It took place for three weeks during July 1971, and involved 30 scientists from 14 countries.

The 308 page report of the meeting, published by the MIT Press (SMIC, 1971), covers what we knew at that time about the Earth's climate system, a brief description of how the climate has changed in the past, and how mankind is changing the climate now. Among the activities that are most likely to affect the global climate are the changes of the land surface due to agriculture and deforestation, the addition of smoke and other particles (which cause cooling), and the addition of vast amounts of CO_2 to the atmosphere from the burning of fossil fuels (which cause warming due to the greenhouse effect).

Although there was considerable discussion about which of the last two opposing effects might dominate, the final report does not answer the question. It states in part: It is clear to us that without additional research and monitoring programs, the scientific community will not be able to provide the firm answers which society may need if large-scale, and possibly irreversible, inadvertent modification of the climate is to be avoided .

REFERENCE

SMIC (1971) *Inadvertent Climate Modi cation: Report of the Study of Man's Impact on Climate (SMIC)*, Sponsored by the Massachusetts Institute of Technology, MIT Press, Cambridge Massachusetts, and London, 1–308.

WILLIAM W KELLOGG *USA*

Social Assessment

Robert Goodland
World Bank, Washington, DC, USA

Social assessment, a process designed to elucidate the societal relationships of a country and the structure of society, embraces social appraisal (SAp) and social impact assessment (SIA). SAp is broader than and overlaps with SIA. SAp seeks to understand the socio-cultural and demographic characteristics of populations likely to be affected by a project, SIA predicts the negative social impacts likely to accrue from a proposed project, and seeks to prevent or mitigate them.

Social assessment (SA) seeks to understand the societal relationships of a country, the structure of its societies, their cultures, the wishes of their people. This is a subset of *national social analyses*, akin to the series of National Environmental Action Plans and similar national environmental diagnoses, which, despite the name, first identify national environmental priorities, and second, seek to address them.

SA looks beyond project-induced impacts, and towards development effectiveness of projects. SA seeks to ensure that positive social development outputs are achieved by the project. Environmental assessment (EA) remains more impact oriented than SA, but is moving toward achieving a net improvement of the environment with the project, beyond mere mitigation of known impacts.

SAp (social appraisal) is defined as the socio-cultural and demographic characteristics of populations likely to be adversely affected by a proposed project, the social organization of productive activities in the project area, the cultural acceptability of project design, its compatibility with intended beneficiaries, and the social strategy for project implementation. Therefore, the terms SA, social analysis and SAp overlap and are partly focused on identification of national social characteristics, or those of the affected people, and national social priorities, and by implication their fulfillment. These terms also contain the sense of identification of the potentially adverse social impacts of project design, but are not confined only to impact mitigation. SA tends to be more at the individual project level; social analysis is more macro or even national in scope.

SIA is a narrower concept, which is falling into disuse. SIA used to be the identification, prevention, minimization, and mitigation of potentially adverse impacts of a proposed project on human society, carried out as part of project design, closely parallel to EA. SA is becoming a tool to monitor the social impacts of a project during implementation. SA as a monitoring tool allows adaptation or restructuring to ensure that intended social development impacts are indeed achieved. This contrasts with EA which rarely leads to project restructuring or even adaptation of the project after construction. Clearly, effective SA can be achieved only after knowing as much as possible about the socio-cultural characteristics of the country, or at least of the project area.

The reason for overlap of these terms is that social dimensions are often specific to the society surrounding a project. Environmental impacts are biophysical and often less site specific. SO_x air pollution, for example, arises

from specific sources (e.g., from a high-sulfur coal thermal electricity generating plant) and is more independent of the immediate environment into which the sulfur gases are released. Human lungs have similar tolerances for inhaling sulfur dioxide, irrespective of their social characteristics. Social analysis does not focus only on the prediction of adverse impacts and their mitigation to the extent that EA still does. That is why SIA has become rare in the terminology of social analysis. Social analysis takes the broader approach of understanding the social fabric, diagnosing the social map, and assessing the willingness of stakeholders to participate in the proposed project.

Sectoral SAs are clearer; they address the whole sector, ranking all potential projects according to their social acceptabilities, well before a single project is identified. Only the best project from the sectoral ranking is then later taken up and designed. Sectoral SA assesses the social impacts of sectoral changes.

The main topics comprising the sociological dimension of economic development include poverty alleviation, social welfare, social justice, equity, vulnerable groups, gender balance, human and social sustainability, culture, education, resettlement, unemployment, land tenure, common property and open access resources, abolition of slavery, debt peonage, United Nations International Labor Organization (UN ILO) conventions, and slum upgrading.

See also: Social Impact Assessment, Volume 4.

Social Impact Assessment

Frank Vanclay
Charles Sturt University, Wagga Wagga, Australia

Social impact assessment (SIA) is the process of analyzing and managing the intended and unintended consequences of planned interventions on people so as to bring about a more sustainable and equitable biophysical and human environment. SIA is more than a technique or tool, it is a philosophy about development which considers the negative effects of development, the goals of development, and the processes of development. The objective of SIA is to ensure that development that does occur maximizes the bene ts and minimizes the costs, especially those borne by the community and often not adequately taken into account by decision-makers, regulatory agencies and developers. By identifying impacts in advance, better decisions can be made about which projects should proceed, and mitigation measures can be developed to minimize the harm from a speci c project or policy.

SIA

SIA has three different levels of meaning: (1) a paradigm, field of research and practice, or discipline; (2) a methodology or environmental management instrument; and (3) a discrete step (or specific task) within a methodology. The concept of SIA is applied at different levels, often with some people failing to appreciate that other meanings or levels exist.

As a field of research and practice (or paradigm), SIA consists of a body of knowledge, techniques, and values. Various individuals identify themselves as SIA professionals, or list SIA as one of their disciplines or specialty areas. There is a community of scholars engaged in research and practice of SIA. These people practice the methodology of SIA and undertake associated social and environmental research to inform the practice of SIA.

As a methodology or instrument, SIA is the process (series of steps) that SIA professionals follow in order to assess and manage the social impacts of planned interventions. That process will be described in this entry, and is perhaps the most common understanding of the term, SIA. It should be noted, however, that the discipline of SIA would consider that SIA practitioners have a professional responsibility, or duty of care, that extends beyond the mere undertaking of the SIA process.

In its narrowest meaning, SIA refers to the prediction of the likely social impacts of a proposed project – one of the tasks or steps within an environmental assessment framework. SIA professionals tend not to define SIA this way, although frequently this is how SIA is seen by other environmental professionals. While the prediction of impacts is an important step, it should not be seen as defining of SIA because SIA, as a paradigm or process, is of greater use to governments and communities (and developers) in its role of managing change and mitigating impacts, than in the simple prediction of impacts. Many negative social impacts can be easily lessened or averted by effective participatory processes, implementation of remedial measures, or improved design.

There has been debate over the precise meanings of terms such as SIA, social analysis, social assessment, social appraisal, social monitoring, and social soundness analysis. Much of this debate has been ill-informed, in particular failing to appreciate the three levels of meaning of SIA as highlighted above. New terminology has developed partly in response to changing ideologies, to promote a particular ideological position, or to escape previous poor practice. But a rose is a rose by any other name. Inventing new terms doesn't resolve the situation and, worse, it inevitably creates confusion. It is probably better to decry poor practice, and to argue for a better understanding of the concept as a whole.

Despite the plethora of names, albeit with their nuances, and although there is not a generally-accepted definition, there is widespread agreement within the SIA community about the concept in principle. Such an understanding can be defined as:

> SIA is the process of analyzing (predicting, evaluating and re ecting) and managing the intended and unintended consequences on the human environment of planned interventions (policies, programs, plans, projects) and any social change processes invoked by those interventions so as to bring about a more sustainable and equitable biophysical and human environment.

The important features of this understanding of SIA are that:

1. The goal of impact assessment is to bring about a more ecologically and socially sustainable environment. Impact assessment, therefore, must be about promoting community development and empowerment, strengthening capacity building skills, and developing social capital.
2. The focus of concern of SIA must be better outcomes of development, not just the identification or amelioration of negative or unintended outcomes. Assisting communities to identify development goals, and ensuring that positive outcomes are maximized, are perhaps more important than minimizing harm from negative impacts.
3. The good practice of SIA must include consideration of the second and later order impacts. In order to do so, the discipline of SIA needs to develop a good understanding of impact pathways – that is, the chains of events that lead to impacts. The discipline needs to be aware of the iterative processes that stimulate social change processes that cause social impacts as well as the processes by which changes to the biophysical environment cause social impacts.
4. It is implicit that social and biophysical impacts (and the human and biophysical environments) are interconnected.
5. In order for the discipline of SIA to learn and grow, there must be analysis of the actual impacts that occurred as a result of past activities – SIA must be *re exive* (self critiquing) and evaluative of its theoretical bases as well as of its practice.
6. The methodology of SIA can be applied to a wide range of planned interventions, and can be undertaken on behalf of a wide range of actors, and not just within a regulatory framework.
7. The process of SIA includes the adaptive management of impacts, projects and policies and therefore needs to be involved (at least considered) in the planning of the project or policy from inception.

THE FOCUS OF SOCIAL IMPACT ASSESSMENT

SIA arguably originated with the 1969 National Environmental Policy Act of the USA (NEPA) (see Interorganizational Committee, 1994; Vanclay, 1999). However, the NEPA understanding of SIA is narrowly conceptualized, tending to consider SIA only at the project level, and only considering first-order impacts, the role of SIA being limited to predictive assessment within the context of a regulatory framework. This limited understanding of SIA pervades the legislation (and agencies) of other nations that modeled their legislation on NEPA.

In contrast, nowadays most SIA professionals consider that SIA is more than a technique or step, rather, it is a philosophy about development and democracy. As such, ideally it considers pathologies of development (i.e., harmful impacts), goals of development (such as poverty alleviation), and processes of development (e.g., participation, capacity building) (based on an idea from Robert Goodland).

An important feature of SIA is the professional value system that a practitioner should uphold. While all impact assessment practitioners should have a commitment to sustainability as well as to scientific integrity, they should also uphold an ethic that advocates openness and accountability, fairness and equity, and defends human rights. The role of SIA goes far beyond the *ex-ante* (in advance) prediction of impacts – SIA, in relation to planned interventions, also encompasses: empowerment of local people; enhancement of the position of women, minority groups and other disadvantaged members of society; development of capacity building; alleviation of all forms of dependency; increase in equity; and a focus on poverty reduction. SIA needs to counterbalance the econocratic and technocratic models that characterize the thinking of many development professionals and their agencies.

SIA can be undertaken in different contexts, and for different purposes, creating some difficulties in defining or evaluating SIA. The nature of an SIA done on behalf of a multinational corporation as part of that company's internal procedures (some sort of social environmental management system or environmental audit) may be very different to an SIA undertaken by a consultant in compliance with regulatory agency or funding agency requirements, or an SIA undertaken by development agency officials interested in ensuring best value for their country's development assistance. These, in turn, may be very different to an SIA undertaken by staff or students at a local university on behalf of the local community, or an SIA undertaken by the local community itself. Each of these roles of SIA is worthwhile, and none must pretend to be the definitive statement. Evaluation of an SIA needs to consider the purpose or role that was intended by the SIA.

It could be argued that NEPA-based conceptualizations of SIA are related to protecting individual property rights, with clear statements of impacts required to ensure that individual rights are not transgressed. Where these rights are violated, SIA could be seen as contributing to mitigation and compensation mechanisms. To that extent, SIA tends to concentrate on negative impacts. In other contexts, perhaps there is less need to be concerned about negative impacts on small groups of individuals or on individual property rights. Instead there should be greater concern with maximizing social utility and development potential, while ensuring that such development is generally acceptable to the community, equitable and sustainable. The improvement of social well-being of the wider community should be explicitly recognized as an objective of development projects and plans, and as such should be an indicator considered by any form of impact assessment. However, awareness of the differential distribution of impacts among different groups in society, and the impact burden experienced by vulnerable groups in the community is of prime concern.

WHAT ARE SOCIAL IMPACTS?

Those who perceive SIA as meaning only the prediction step within an environmental assessment framework tend to have a limited view of what social impacts are. They tend to see social impacts only in terms in demographic changes, job issues, financial security, impacts on family life, etc. This limited view of SIA also creates demarcation problems about what the social impacts to be identified by SIA are, versus what is considered by health impact assessment, cultural impact assessment, heritage impact assessment, aesthetic impact assessment, or gender impact assessment.

The SIA community of scholars, on the other hand, considers that all issues that affect people, directly or indirectly, are pertinent to SIA. For example, changes to the environment that affect people are social impacts. In order to address this issue and to escape the confusion about the limited understanding of SIA and of social impacts, there has been some discussion about whether there needs to be a new field called human impact assessment. However this would only lead to more terminology proliferation.

A convenient way of conceptualizing social impacts is as changes to one or more of the following (based on an idea by Audrey Armour):

- people's way of life – that is, how they live, work, play and interact with one another on a day-to-day basis;
- their culture – that is, their shared beliefs, customs, values and language or dialect;
- their community – its cohesion, stability, character, services and facilities;
- their political systems – the extent to which people are able to participate in decisions that affect their lives, the level of democratization that is taking place, and the resources provided for this purpose;
- their environment – the quality of the air and water people use; the availability and quality of the food they eat; the level of hazard or risk, dust and noise they are exposed to; the adequacy of sanitation, their physical safety, and their access to and control over resources;
- their health and wellbeing – where health is understood in a manner similar to the World Health Organization definition: a state of complete physical, mental and social wellbeing and not merely the absence of disease or infirmity ;
- their personal and property rights – particularly whether people are economically affected, or experience personal disadvantage which may include a violation of their civil liberties;
- their fears and aspirations – their perceptions about their safety, their fears about the future of their community, and their aspirations for their future and the future of their children.

A FRAMEWORK FOR UNDERSTANDING IMPACT PATHWAYS

Figure 1 provides a conceptual framework for understanding the ways social and biophysical impacts are interdependent and how all impacts are ultimately social impacts. People experience impacts and/or they give a value to the environment and to the preservation of species. The major difference between Environmental Impact Assessment (EIA) and SIA lies in the pathways through which impacts are generated, and the different levels of attention given to those pathways or impacts by different groups of impact assessment professionals. Impacts are produced either through the physical environment (pollution that threatens health; noise that reduces the quality of life) or

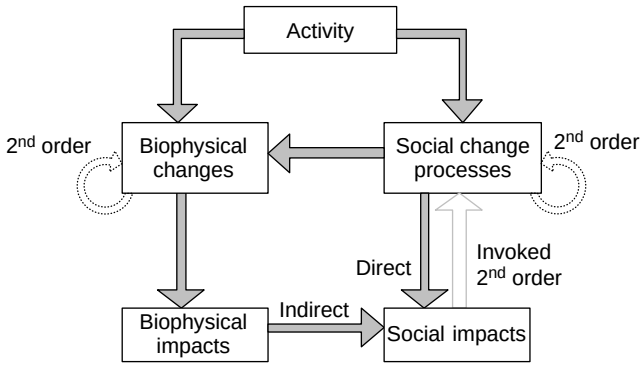


Figure 1 Linkages between biophysical and social impacts. (Reproduced in slightly modified form, with permission of Slootweg *et al.*, 2001)

through the social environment (in ux of foreign tourists creating anxiety in a local community, creation of employment that raises the standard of living, etc.).

In this framework, which is developed more fully in Slootweg *et al.* (2001), direct and indirect social impacts are distinguished. Direct social impacts result directly from an intervention, and may be the intention of specially designed activities to in uence the social setting (intended impacts), or may unintentionally result from the activities. Indirect social impacts are a result of changes in the biophysical environment. Biophysical changes can affect the functions the environment provides for people. For example, if an activity causes land degradation, one of the biophysical impacts can be that the productive capacity of the land will decrease. The resulting reduction of income from farming activities is an indirect social impact. Or, dredging activities may lead to increased turbidity of water, which may alter the ecology of a river ecosystem that may reduce the fish population that forms the nutritional basis for the local people. Other physical changes may have effects on disease organisms or disease vectors such as mosquitoes, rats and snails, which may lead to health impacts. It should be noted that social change processes also cause biophysical changes, which may lead to further biophysical and social impacts.

Figure 1 also differentiates *social change processes* from *social impacts*. Confusion exists in the SIA literature in relation to the lack of distinction between the social change processes that are caused by projects or policies, and the social impacts that are actually experienced. Strictly speaking, an impact must be experienced (felt) by an individual, family or household, or a community or society, whether in corporeal (physical) or perceptual terms. To a large extent, social impacts are socially constructed – they are perceived differently by different people through socially-mediated understandings of what is normal or natural, or to be expected or even acceptable. A social change process, on the other hand, is a process of change that may lead to the experience of impacts in certain circumstances depending on the social context (groups, nations, religions, etc.). Thus, resettlement (relocation of a community), for example, is a social change process, not a social impact. However, it may well *cause* social impacts such as anxiety and stress, uncertainty, disruption to daily living, potential change to family structure, as well as reduction in the feeling of homeliness. Similarly, demographic changes such as a large population increase (or decrease), the presence of seasonal workers, and/or weekend residents, should be considered as social change processes rather than as social impacts. They are not impacts in themselves, but rather they cause other impacts,

Demographic processes	**Institutional processes**
In-migration	Globalization (the incorporation of the local into the global) loss of autonomy of decision making at the local level
Out-migration	Decentralization
Presence of newcomers	Institutionalization and bureaucratization
Presence of (temporary) construction workers	
Presence of seasonal residents	**Empowerment Processes**
Presence of weekenders	Democratization
Presence of tourists	Capacity building
(Involuntary) resettlement	
Rural to urban migration	**Socio-cultural processes**
Urban to rural migration	Globalisation (the incorporation of the local into the global) loss of cultural identification; macdonaldization, coca-cola development; cultural hegemony
Economic processes	Marginalization and exclusion (the process of creating marginal groups in society, which as a result are denied access to services)
Diversification of economic activities	Segregation (the process of creation of social difference within a community)
Concentration of economic activity	
Impoverishment	**Other processes**
Inflation	
Fluctuation in currency	
Job loss/retrenchment	
Tourism	
Geographical processes	
Conversion of landuse	
Urbanization	
Increased transportation and rural accessibility	
Physical splintering (such as caused by major roads)	

Figure 2 Categories of social processes with examples

Individual and household level	Community and institutional level
1. Death, death of family member 2. Arrest, imprisonment, detention, torture, intimidation, or other abuse of human rights inflicted on an individual 3. Reduced availability of food and adequate nutrition 4. Reduced control over fertility (loss of availability of contraception, and gender dis-empowerment to make decisions about birth control) 5. Reduced level of health and fertility (ability to conceive) 6. Reduced mental health, increased stress, anxiety, alienation, apathy, depression 7. Uncertainty about impacts, development possibilities, or about own life or the lives of those one is responsible for, as a result of social change 8. Changed responsibilities 9. Loss of aspirations about the future for self or children 10. Reduced personal safety, increased hazard exposure 11. Experience of stigmatisation and deviance labelling 12. Reduction in perceived quality of life, subjective well-being, self esteem, self image 13. Reduction in standard of living, level of affluence 14. Worsening of economic situation, level of income, property values 15. Decreased autonomy, independence, security of livelihood 16. Change in employment, or becoming unemployed 17. Decrease in occupational opportunities, potential diversity, flexibility in employment 18. Moral outrage, blasphemy, religious affront, violation of sacred sites 19. Upsetness (objection/opposition) to project, NIMBY 20. Dissatisfaction due to failure of a project to achieve heightened expectations 21. Annoyance (dust, noise, strangers, more people) 22. Disruption to daily living, way of life (having to do things differently) 23. Reduction in environmental amenity value 24. Reduced perception of communityness, community cohesion, integration 25. Loss of community identification, connection to place (do I belong here?) 26. Changed attitude towards local community, level of satisfaction with the neighbourhood 27. Disruption to social networks 28. Alteration in family structure, family stability, divorce 29. Increased family violence 30. Deteriorating gender relations within the household 31. Changed cultural values 32. Worsening perceptions about personal health and safety, risk, fear of crime 33. Reduced leisure opportunities 34. Reduced quality of housing 35. Reduced feeling of homeliness 36. Increased density and crowding 37. Reduced aesthetic quality, outlook, visual impacts 38. Increased workload, amount of work needed to be undertaken to survive/live reasonably 39. Inability to meet obligations to living elders or ancestors	1. Death of people in the community 2. Violation of human rights, freedom of speech 3. Reduced adequacy of physical infrastructure (water supply, sewerage, services and utilities) 4. Changed demographic structure 5. Reduced adequacy of community social infrastructure, health welfare education libraries etc. 6. Reduced adequacy of housing in the community 7. Increased workload on institutions, local government, regulatory bodies 8. Diminished cultural integrity (continuation of local culture, tradition, rites) 9. Loss of rights over, and access to, resources 10. Destruction of, or other negative influences on, heritage and other sites of archaeological, cultural, or historical significance 11. Loss of local language or dialect 12. Profanization of culture 13. Increased inequity (economic, social, cultural) 14. Increased concern about social justice issues in relation to minority or indigenous groups 15. Worsening gender relations in the community 16. Decreased economic prosperity 17. Increased dependency, reduced autonomy, reduced diversity, decreased viability of the community 18. Increased unemployment level in the community 19. Loss of other options (opportunity cost) 20. Increased actual crime 21. Increased actual violence 22. Increased social tensions, conflict or serious divisions within the community 23. Increased corruption, decreased credibility or integrity of government 24. Decreased level of community participation in decision making, loss of empowerment 25. Impact on the social values about heritage and biodiversity 26. Burden of national debt *Source*: Vanclay (1999) (updated)

Figure 3 A listing of potential social impacts

such as a breakdown of the social fabric of the community, or changed perceptions about a community by local residents, and they may place too much pressure on the physical and social infrastructure in a community. Alcohol or other drug use is not a social impact but a process, which, depending on the context, may cause social impacts such as family violence and economic hardship.

Figure 1 displays not only how the physical environment and social impacts are integrated, but also the iterative processes affecting social change processes, and between the experience of impact and further social change.

Part of the reason why many SIA scholars have confused issues of social change processes and social impacts, is because demographic processes such as changes in the size and/or composition of the population are more easily measured than the impacts themselves. Because they provide empirical data, documentation of likely social changes has been more acceptable to technically oriented regulatory agencies. But, social change processes do not always cause negative social impacts. The ultimate experience of an impact is determined by the interaction between the characteristics of the proposed project or policy, the characteristics of the affected community, the nature of any mitigation processes, as well as the characteristics of specific individuals. To assist in understanding the differences between social change processes and social impacts, listings of social impacts and social change processes are presented (see Figures 2 and 3).

Figure 2 presents only a few examples of social change processes, because the number of social change processes is potentially infinite, and the processes vary widely according to the activity being planned. The list of impacts in Figure 3, however, is reasonably complete, although some disagreement may exist about the precise wording and categorization of impacts. In Figure 3, impacts are classified according to the level of experience of impact – that is, whether they are experienced at an individual or household level, or whether they are experienced by the community or society as a whole. Of course, impacts on society as a whole will translate into impacts that will ultimately affect individuals as individuals.

All of the social impact variables must be understood in their sociological context, and, of course, in their local cultural context. Homeliness, for example, does not mean the physical quality of the house, but the social relationships among the occupants of the building, and between them and the building. It is a subjective concept, based on people's perceptions and relating to the meaning and experience people attach to the place where they live and build their home.

The list of impacts in Figure 3 is provided to display the full range of potential social impacts, and to assist people not familiar with SIA to become aware of the full extent of it. In order to identify all the likely impacts of a planned intervention, it is necessary to develop an understanding of the impact chains (causal processes) so that the second and subsequent order impacts will be identified. Figure 3 should not be considered a checklist because a checklist approach tends only to identify first-order impacts.

THE METHODOLOGY OF SIA

There are a number of different models about how SIA should be implemented, and about the order in which the various steps should be undertaken. While there is no universal standard for SIA (although perhaps the US Interorganizational Committee 1994 document comes close), it is appropriate to state that a good SIA does the following (Vanclay, 1999):

- identifies interested and affected peoples;
- facilitates and coordinates the participation of stakeholders;
- documents and analyses the local historical setting in which the project will occur so as to be able to interpret likely responses to the project, and to assess cumulative impacts;
- gives a rich picture of the local cultural context, and develops an understanding of local community values, particularly how they might relate to the planned intervention;
- identifies and describes the activities which are likely to cause impacts (scoping);
- predicts likely impacts and how different segments of the community are likely to respond;
- assists in evaluation and selection of program alternatives (including a no-development option);
- assists in site selection;
- recommends mitigation measures;
- provides suggestions about compensation;
- describes potential con icts between stakeholders and advises on resolution processes;
- develops coping strategies in the community for dealing with residual or non-mitigatable impacts;
- contributes to skill development and capacity building in the community;
- advises on appropriate institutional and co-ordination arrangements for all parties;
- assists in the devising and implementation of monitoring and management programs;
- collects baseline data (profiling) to allow evaluation and audit of the impact assessment process and the project itself.

SIA can be applied at the policy level, at the program level and at the project level. Any planned intervention has a so-called life-cycle of four stages: planning, construction (or implementation), operation and decommissioning (phasing-out). Each of the four phases has its unique

impacts and should be considered separately. Importantly, because of the potential presence of fear, anxiety and speculation, social impacts can also occur not only at the level of a plan, but even when an intervention is just an idea. So, not only is an SIA of a plan required to predict what will happen in the implementation of a plan, an SIA can be desirable in the creation of the plan.

CONCLUSION

SIA, at its ideal, is a process that actively brings about sustainable development that consciously considers social issues. In that respect, it should be considered a process of navigation rather than as a process of prediction. Because of the interrelationship between environmental and social issues, ultimately all impacts, including biophysical impacts, can be considered to be social impacts because they are experienced or valued by humans. In effect, then, SIA is environmental assessment. SIA recognizes the dependency of the human world on the biophysical world, and expresses this dependency in terms of the impacts on humans. Effective SIA requires thorough EIA.

SIA is a process which includes aesthetic impacts, economic impacts, health impacts, heritage impacts, psychological impacts, as well as impacts on political and institutional systems. In short, it covers everything, but reports impacts in terms of what matters, or will likely matter, to people as a result of a planned intervention. For thorough analysis of the impacts, the input of specialists from separate disciplines is required.

SIA can be applied in the context of national environmental regulation or because of the requirements of international donor agencies, but it can also be undertaken by a community to ensure optimality of social goals, or for a corporation to ensure triple bottom-line accounting – financial soundness, environmental soundness and social soundness.

ACKNOWLEDGMENT

This entry draws on material previously published as Vanclay (1999), and Slootweg *et al.* (2001).

REFERENCES

Barrow, C J (1997) *Environmental and Social Impact Assessment, an Introduction*, Arnold, London.

Becker, H A (1997) *Social Impact Assessment: Method and Experience in Europe, North America and the Developing World*, UCL Press, London.

Burdge, R J (1999) *A Community Guide to Social Impact Assessment*, (revised edition), Social Ecology Press, Middleton.

Burdge, R J and Vanclay, F (1995) Social Impact Assessment, in *Environmental and Social Impact Assessment*, eds F Vanclay and D A Bronstein, John Wiley & Sons, Chichester.

[US] Interorganizational Committee on Guidelines and Principles for Social Impact Assessment (1994) Guidelines and Principles for Social Impact Assessment, *Impact Assess.*, 12, 107–152.

Slootweg, R, Vanclay, F, and van Schooten, M (2001) Function evaluation as a framework for the integration of social and environmental impact assessment, *Impact Assessment Project Appraisal*, 19, 19–28.

Taylor, C N, Bryan, C H, and Goodrich, C G (1995) *Social Assessment: Theory, Process and Techniques*, 2nd edition, Taylor Baines and Associates, Christchurch.

Vanclay, F (1999) Social Impact Assessment, in *Handbook of Environmental Impact Assessment*, Vol. 1, ed J Petts, Blackwell Science, Oxford.

Socially Responsible Investment

Rory O'Brien
University of Toronto, Toronto, Canada

In 1993, upon being elected president of South Africa, Nelson Mandela called for an end to the global economic boycott of his country. The boycott, which was instrumental in pressuring the white regime to open elections to black citizens, was spearheaded since the mid-1970s by a movement to curtail investment in South African industries. At the time, a broad coalition of church groups ghting apartheid promoted the idea that it was not ethical to invest in corporations doing business that supported an unethical political regime. This was not a new idea - Quakers had been refusing to invest in military suppliers, and temperance proponents had been advocating the divestment of stock in alcoholic beverage producers, for over a hundred years. It was, however, the rst time that investment as a tool for social change began to be applied systematically on a large scale.

Over the past several decades, the socially responsible investment (SRI) movement has grown considerably, from being practiced by only a handful of church-based investors, to entering the mainstream of the investment community. In the US, for example, currently one in every seven dollars in the stock market is invested according to the criteria and practices of social investment. This has, and will continue to have, a growing impact on the behavior of corporations, making them more likely to adopt environmentally-sustainable business practices, promote diversity and equality in the workplace, and deal

more fairly with employees, customers, and community stakeholders.

TYPES OF SRI

There are three general manifestations of SRI – *screened investments*, *shareholder activism* and *community investment*, each used in particular circumstances, and often in conjunction with one another.

Screened Investments

Regular professional investors judge whether or not to invest in a stock, based on the economic performance of that company. They use certain criteria, known as screens, which tell them how risky the investment is compared to the potential return. A high debt-to-equity ratio, for example, may be a screen that precludes the investor from buying that stock, since it indicates that the company is burdened with debt servicing and is less exible in its strategic options. So, too, do social investors use screens. The difference is that once social investors finish with the economic screening, they then apply other screens that are based on the social or environmental performance of that company.

There are positive and negative SRI screens. Whenever social investors wish to put their money into companies that are doing good things from a values perspective, they apply positive screens. A particular company might be targeted because it has a good track record of employee relations, or produces a product that is beneficial to the environment, such as solar heating panels. Negative screens, on the other hand, can exclude companies doing bad things from the investor's point of view. One of the more commonly applied negative screens are for the so-called sin stocks, in which conscientious investors will not put their money into companies that deal with tobacco, alcohol or gambling. Others include such activities as child labor, weapons production, and nuclear power.

Shareholder Activism

Ultimate control of a corporation rests with its shareholders, who have the right to submit and vote on proposed resolutions at the corporate Annual General Meeting. Traditionally, most resolutions are formulated by a company's management, and relate to election of directors or major financial decisions such as potential mergers. Increasingly, however, socially responsible investors, particularly coalitions of institutional investors such as progressive pension funds, are likely to present resolutions that direct management to change current policies and practices related to both corporate governance issues (such as executive compensation) and issues of social responsibility (such as the environment). A majority of shareholder votes need not be attained to exert a positive in uence – frequently managers will begin creating improvements based on less than 10% of ballots cast in favor of such resolutions. Indeed, the essence of shareholder action is to enter into constructive dialogue with management, and many changes are induced without resorting to formal shareholder proposals.

Community Investment

Investing in one's own community is a socially responsible way to keep the local economy healthy. Institutions and programs such as credit unions and community loan funds provide the capital for housing and small business development for local citizens, particularly those unable to obtain loans from more conventional sources. Communities with a strong base of local enterprise are less likely to suffer the social disruptions resulting from massive layoffs in one-industry towns.

GROWTH OF SRI

Social investment has developed mostly in the US, Canada and Britain (where it is known as *ethical investment*). Though it is widely acknowledged that SRI investments in Britain and Canada have undergone significant growth in the past several years, only in the US have researchers systematically tracked increases in each of the three aspects of the SRI industry. Figure 1 shows the results of comprehensive surveys of the SRI industry done in 1995, 1997 and 1999. Screened investments have grown the most rapidly, due both to the increase in the number and size of screened mutual funds and also to the success of the media-fueled anti-tobacco campaign. In 1999, over $2 trillion in total was socially invested in the US, accounting for roughly 13% of the $16.3 trillion under professional management in the US. SRI investments, up from $695 billion in 1995, have increased at twice the rate of non-SRI investments.

The growing popularity of SRI can be primarily attributed to the debunking of the popular myth that people investing with social concerns had to be content with a lower rate of return. Recent studies have shown that there is no significant difference between SRI and non-SRI investment returns. In fact, there is evidence that companies practicing social responsibility actually perform better than others over the long term. Since its inception in 1991, the Domini 400 Social Index, an American index for tracking socially-screened corporations, has had a slightly higher rate of return than the unscreened Standard and Poor 500 Index.

EFFECT ON CORPORATE PERFORMANCE

While it is clear that values-based screening has become an important factor in the purchase of stock, its impact on

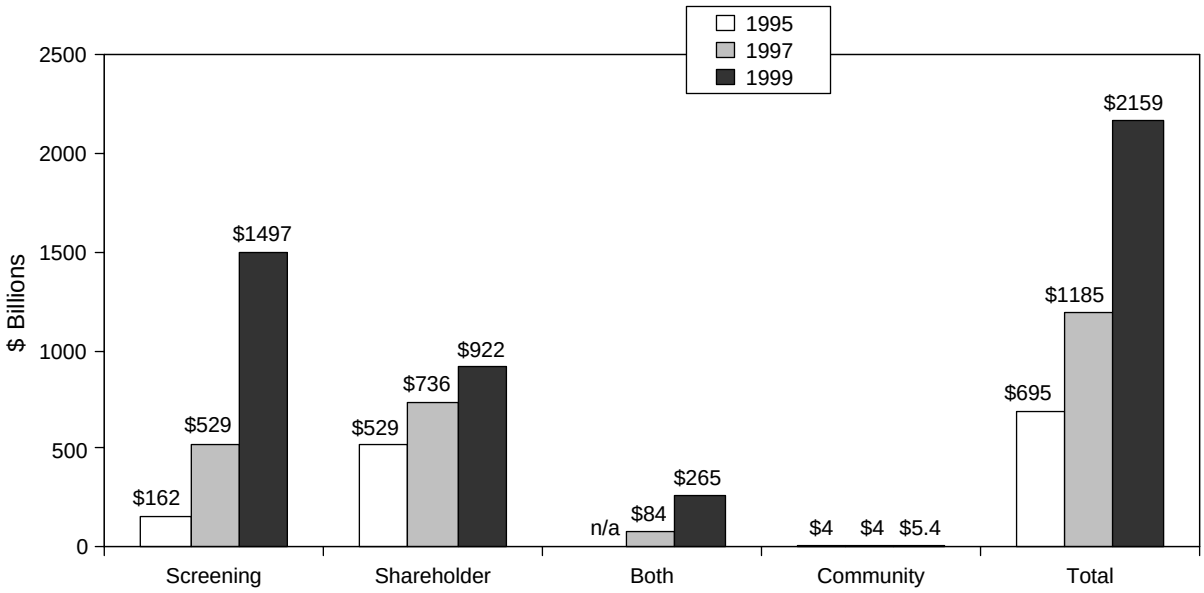


Figure 1 Growth in SRIs in the US 1995–1999 (reproduced from Social Investment Forum, 1999 with permission). Portfolios that both screen and conduct shareholder advocacy include $84 billion of assets in 1997, and $265 billion in 1999. In 1995, any portfolios involved in both screening and shareholder advocacy assets are included in screening. The calculation for total is: total = screening + shareholder + community – both

corporate social performance to date is less clear. There are increasing indications, however, that the effects are beginning to be realized. SRI screens have become valuable benchmarks for corporate leaders; companies are now adopting corporate codes of conduct, such as the Coalition for Environmentally Responsible Economies principles (*see also* Green Investment, Volume 4), which often form the basis for screens. A growing number of advocacy groups, such as Greenpeace, are starting to ally themselves with shareholder activist organizations as part of their issue campaigns, successfully targeting larger institutional investors, such as pension funds, for support. To date, the changes most frequently negotiated involve greater disclosure on activities and their effects, such as reporting on pollution.

REFERENCE

Social Investment Forum (info@socialinvest.org) (1999) Social Investment Forum: Responsible Investing Trends in US, 5/11, http://www.socialinvest.org/areas/research/trends/1999-Trends.htm. 06/04/00.

FURTHER READING

Brill, H, Brill, J A, and Feigenbaum, C (1999) *Investing with Your Values: Making Money and Making a Difference*, Bloomberg Press, Princeton, NJ.

CERES (ceres@igc.apc.org) (1998) The CERES Principles, 29/12, http://www.ceres.org/about/principles.html. 06/04/2000.

Kinder Lydenberg Domini & Co (webemail@kld.com) (1999) The Domini 400 Social Index, http://www.kld.com/wdomi.html. 06/04/00.

Pratt, R (1997) *In Good Faith: Canadian Churches Against Apartheid*, Wilfred Laurier Press, Waterloo, Ontario.

Soil Erosion

Harold E Dregne
Texas Tech University, Lubbock, TX, USA

There is good presumptive evidence that soil erosion is a long-continuing problem around the world. There also is much anecdotal evidence, particularly of off-site damage due to erosion, such as increased ooding and frequency of dust storms. Direct quantitative measurements of erosion are dif cult and costly to make, which accounts for the lack of good data in many countries. Better understanding of soil erosion processes has allowed construction of good predictive models of soil loss. Progress is being made on developing models relating soil erosion to reduction in soil productivity. High erosion rates do not necessarily mean high losses in productivity. Effective erosion control practices are known for most situations. They usually are not pro table for farmers for the rst few years.

INTRODUCTION

Accelerated water and wind erosion are widely believed to be a serious global problem. There is no doubt about the increase in water erosion that occurred more than 2000 years ago in the eastern Mediterranean countries. Plato noted the brown plumes of muddy water that Grecian rivers carried into the blue Mediterranean Sea. He lamented the loss of soils from the rocky hills of Greece and the threat it posed to food production. The effects of that early erosion still limit crop production around the Mediterranean basin. Human induced (accelerated) wind erosion also has a long history but usually for small areas in arid regions. One common example is villages along the Silk Road in China which have been covered by moving sands from nearby cultivated or overgrazed fields. Wind erosion also occurs in humid regions, such as in coastal Denmark.

EXTENT AND SEVERITY OF SOIL EROSION

Water and wind erosion are, by far, the most important soil degradation processes. Not only does accelerated erosion reduce soil fertility and increase ooding, along with other damage, but it brings an irreversible loss of potential soil productivity. Once soil is gone, it requires hundreds or thousands of years for nature to restore it to its original condition. The consequence, then, is an irreversible wearing away of the soil as long as erosion continues. That irreversibility is why it is better to stop erosion before it has damaged productivity than to wait until significant damage has been done. Water erosion is the principal type of erosion in humid regions. Wind erosion is dominant in arid regions but water erosion can be significant in cultivated land and overgrazed rangelands in semi-arid climates. Of the approximately 1.5 billion ha of cropland in the world, probably 60% or more has experienced some degree of erosion. Sheet erosion is easy to ignore because it does relatively little damage each year. Gully erosion is much less prevalent than sheet erosion. Gullies are proof that severe damage has been done to the land. Mobile sand dunes and more frequent dust storms provide proof that wind erosion has become severe.

Field measurements of water erosion have been made by many countries. Rather costly techniques of catching runoff water and eroded soil from plots are used. A major drawback to using field plots, beside the expense, is the requirement to maintain records of rainfall and runoff for several years. The principal uncontrolled variable is the difference in rainfall from year to year. Even three years is too short to give high confidence in the results. Wind erosion presents an almost insurmountable difficulty in measuring the amount of soil loss. With water erosion, all the water and soil running off a plot can be measured. With wind erosion, no device yet made can collect all of the soil particles in a vertical cross-section of the air at all heights above an eroding field.

EROSION EQUATIONS

Erosion equations based on experimental data and model development are coming into wide use. The equations use data on crops, soils, rainfall, land slopes, and conservation practices to calculate predicted water erosion rates. Similar data, plus wind direction, are part of the wind erosion equation. The equations are used in the United States, where they were developed, to estimate erosion at about 800 000 sites for 5-year erosion inventories. Since the equations were developed under the conditions that exist in the United States, they must be used with caution in places where soils, crops, and weather conditions are different. Local experimental data are needed to check the accuracy of the equations. A few simpler and less accurate soil erosion equations have been developed in other countries.

A very important question that arises among users of erosion data is the effect that levels of erosion have on potential soil productivity. A water erosion rate of 40 tons ha^{-1} does not affect productivity the same on all soils and with all crops. On shallow soils, 40 tons ha^{-1} of soil loss would be considered high because of the thinness of the soil. On a deep soil, 40 tons ha^{-1} would be moderate or even slight erosion. Erosion control practices are needed much sooner on shallow soils than deep ones to protect productivity.

An erosion-productivity impact calculator (EPIC) has been developed to relate erosion by wind or water to potential soil productivity loss. The equation does reasonably well but needs additional validation of the reliability of the results. EPIC requires considerable input data, which is sometimes not available. Improved erosion model development is an objective of many research groups in many countries.

ON-SITE AND OFF-SITE IMPACTS

Soil loss brings two main types of damage: on-site and off-site. On-site damage by water erosion consists primarily of reduction of soil productivity by removal of the nutrient-rich topsoil. Water ow can also wash soil away from plant roots or cover small plants with soil. Off-site damage includes many things, such as increased ooding downstream, sedimentation of reservoirs, pollution of streams and lakes, eutrophication of lakes, and others. For wind erosion, on-site damage is largely done by strong winds and air-borne soil particles shredding of leaves and stems of young plants, called sand blasting. Removal of soil around plant roots and burying field plants also occurs. Off-site damage from the air-borne sand pollutes air downwind,

abrades buildings and other structures, corrodes machinery, lowers air visibility, covers roads, causes respiratory problems, etc.

There are few studies of on-site versus off-site costs of erosion. Those that have been made indicate that those costs are about equal for water erosion. Wind erosion off-site damages were estimated to be about 45 times greater than on-site costs in one large-scale analysis.

EROSION CONTROL

Control practices for water and wind erosion are based on one principle: slow down the rate of water and wind movement over the soil surface. For water erosion, gently sloping or level (bench) terraces are widely used to control soil and water runoff. Terracing has been practiced successfully for more than 1000 years in China, Yemen, Peru, and elsewhere. In addition to controlling erosion, terracing produces more or less level land that is excellent for crop production, especially in drylands, where water retention is important. Once terraces have been constructed, continued maintenance is essential to prevent gully formation and terrace destruction by runoff during heavy rains.

The best way to reduce water and wind erosion is by having a good vegetative cover on the land. Vegetation serves to protect the soil from the direct impact of raindrops, which causes splash erosion and increases susceptibility of the soil to further erosion. It also slows down runoff and holds soil particles in place. Healthy grasslands and forests usually experience very little erosion by wind or water.

Unfortunately, most croplands are barren for extended periods between growing seasons. For that situation, several land management techniques have been devised. Strip cropping, crop rotations, agroforestry, stubble mulching, intercropping, gravel mulching, no-tillage, and minimum tillage represent various ways to reduce both water and wind erosion on croplands. Gravel mulching (spreading gravel over the soil) stops erosion immediately and reduces evaporation but is so labor intensive that the practice is common only in a small area in central China. No tillage and minimum tillage are high technology practices that generally require use of herbicides and a high level of management.

Certain tillage practices are effective in reducing wind erosion for short periods of time by roughening the soil surface. Throughout the world, wind erosion control relies heavily on windbreaks of trees and shrubs on borders of fields. Windbreaks should be oriented at right angles to the dominant wind direction.

See also: Soil Deterioration and Loss of Topsoil, Volume 3; National and Local Responses to Desertification, Volume 4.

FURTHER READING

Morgan, R P C (1979) *Soil Erosion*, Longman, London, 1.
Troeh, F R, Hobbs, J A, and Donahue, R L (1980) *Soil and Water Conservation*, Prentice Hall, London, 1–718.

Speth, James Gustave

(1942-)

On 21st May 1997 the United Nations (UN) General Assembly confirmed James Gustave Speth for a second 4-year term as Administrator of the United Nations Development Program (UNDP), following his nomination by UN Secretary-General Kofi Annan.

Speth served as Special Coordinator for Economic and Social Affairs under Secretary-General Boutros Boutros Ghali and now serves as Chair of the United Nations Development Group, created as part of the United Nations reforms initiated by Secretary-General Kofi Annan.

Speth, known as Gus, founded the World Resource Institute (WRI) in 1982 and served as its president until January 1993. WRI, based in Washington, DC, is a center for policy research and technical assistance on environment and development issues. Speth was then a senior advisor to President-elect Clinton's transition team, heading the group that examined the US role in natural resources, energy and the environment.

Before founding WRI, Speth served from 1977 to 1981 as a Member and then for 2 years as Chairman of the Council on Environment Quality (CEQ) in the Executive Office of the President. As President Carter's CEQ chairman, Speth was a principal advisor on matters affecting the environment and had overall responsibility for developing and coordinating the President's environmental program. In 1981 and 1982 he was Professor of Law at Georgetown University Law Center in Washington, DC, teaching environmental and constitutional law.

From 1970 to 1977 Speth was senior attorney with the Natural Resources Defense Council, an environment organization he co-founded.

Among Speth's awards are the National Wildlife Federation's Resources Defense Award, the Natural Resources Council of America's Barbara Swain Award of Honor,

the Keystone Center's National Leadership Award, a 1997 Special Recognition Award from the Society for International Development, the 1998 Leadership Award of the Alliance for UN Sustainable Development Programs, the *Ordre National du Lion* of Senegal, and the *Grand Cordon du Wissam Alaouite* of Morocco.

Prior to joining UNDP, Speth organized numerous leadership initiatives on environment and development. In 1991 he chaired a US task force on international development and environmental security which produced the report *Partnership for Sustainable Development; A US Agenda*. In 1990 he led the Western Hemisphere Dialogue on Environment and Development which produced the seminal report *Compact for a New World*. Among his writings is A Post-Rio Compact , which appeared in the Fall 1992 issue of *Foreign Policy*.

Speth graduated *summa cum laude* from Yale University in 1964, attended Oxford University as a Rhodes Scholar and graduated from Yale Law School, where he was a member of the *Yale Law Journal*, in 1969. In 1995 he received an honorary Doctor of Laws degree from Clark University. He served in 1969 and 1970 as a law clerk to US Supreme Court Justice Hugo L Black. He is married to Cameron Speth and they have three children. He was born in Orangeburg, South Carolina in 1942.

MOSTAFA K TOLBA *Egypt*

START (Global Change System for Analysis Research and Training)

START is an international non-governmental framework for building indigenous capacity in developing regions to cope with the scientific and science-related policy aspects of environmental change. It is co-sponsored by the three international global change science programs, the International Geosphere–Biosphere Program, the International Human Dimensions Program on Global Environmental Change and the World Climate Research Program, all of which are programs of the International Council of Science.

START's challenge is to help meet the needs of developing countries with respect to global change science. It does this through the mechanism of regional research networks established in developing regions, which are intended to:

- develop human and institutional capacities in global change science;
- conduct research on regional aspects of global change;
- assess the impacts of regional findings;
- provide information in appropriate forms to policymakers.

Six regional networks and coordinating offices are now operational:

- Pan-Africa (covering Sub-Saharan Africa);
- Mediterranean (including Northern Africa);
- South Asia;
- Southeast Asia;
- Temperate East Asia;
- Oceania.

START also closely collaborates with the existing Inter-American Institute for Global Change Research on activities in the Americas.

START regional networks address the following themes:

- regional climate variability and change;
- changes in composition of the atmosphere;
- land use/cover change and land degradation;
- impacts of global change on terrestrial ecosystems and biodiversity;
- global change and coastal zones, land–ocean interactions and international waters;
- regional assessments of global change impacts.

A major part of START's initial effort has been capacity building in its broadest sense involving a fellowship and visiting scientist program, recognition of talent in developing regions, development of collaborative research networks, and partnerships between regional scientists and the international scientific community. All are research-linked, recognizing that capacity building is best achieved through continual involvement in research.

For further information, see http://www.start.org.

ROLAND FUCHS *USA*

Stockholm and Beyond

Mostafa K Tolba
International Center for Environment and Development, Cairo, Egypt

The 1972 Stockholm Conference on the Human Environment was a turning point in the history of environmental awareness and action. Its Action Plan for Human Environment, the establishment of the United Nations Environment Program that year, and the enthusiasm of non-governmental organizations (NGOs), both at and after

the conference, gave further impetus to the environmental movement and gave it effective expression in the international community.

Before Stockholm, many developing countries were suspicious of environment as a public issue on a global scale. This was partly due to the way environment was de ned in those days - as a question of pollution and its abatement. It was seen as a technical question, amenable to technological, legal and administrative answers. Few today would be satis ed with this restrictive interpretation, which stemmed from the preoccupations of industrialized societies.

Developing countries were, of course, aware of the threats which uncontrolled pollution posed to the world and to themselves. They realized too that their own societies would increasingly have to face problems of pollution. Some leaders, at least, felt that they should avoid the mistakes already committed by the industrialized countries.

They also recognized that environmental concern in the North might have positive economic spin-offs for the South. For example, developing countries might bene t from the resiting of polluting but productive industries, and from replacement of environmentally suspect synthetic products with the natural ones they produced themselves.

Yet many developing countries had misgivings. Some saw environment as a ploy of the rich to prevent developing nations from industrializing, and to keep developing countries supplying the raw materials on which the development and prosperity of the North depended. Some argued that if pollution meant industry, they would welcome it wholeheartedly. More realistically, perhaps some feared that their development ambitions might suffer as industrialized countries diverted aid to the new problem of the environment. If anyone was going to lose out through a concern for the environment, they argued, it should be the rich countries, who were largely responsible for pollution.

Developing countries also feared that environmental concerns would impose extra burdens upon them - for instance, that stringent standards might put up the cost of imported technology and machinery. They suspected that their export trade would be harmed by the environmental regulations industrialized countries were likely to adopt.

All in all, therefore, environment and development were seen as being in con ict. It was feared that a commitment by the international community to environment was likely to detract from its commitment to development.

During early times, people learned that their actions could damage the natural resources by which they lived. Tree cutting, overgrazing and soil erosion undermined agricultural productivity around the Mediterranean, in southern and central China, in India and Central America. Historical records give evidence of early attempts of conservation, religious taboos protecting some species of animals, some forest groves and plants; the use of organic fertilizer and other practices to maintain soil fertility and prevent erosion; the creation of wildlife or natural reserves. But in spite of these efforts, ourishing civilizations disappeared.

The industrial revolution of the latter 18th and early 19th centuries and the 20th century, and scientific and technological revolutions dramatically increased humanity's need for natural resources and its pressure on the environment. Overwhelmingly, this impact has been negative.

Today – as we stand at the beginning of a new millennium – we face an ecological crisis that has assumed planetary dimensions. The problems are so great that they threaten to derail economic development.

The technological advances in the first half of the 20th century raised fundamental questions about the future. Could the globe continue to support its rapidly growing population? And how appropriate was the technology itself? These questions were first debated in scientific circles, but soon caught the attention of the public.

By the late 1950s public concern was rising: people had died in smog episodes in Belgium, the United States and the United Kingdom; lives had been shattered by mercury pollution in Minamata and Niigata, Japan; acidification was killing lakes in Scandinavia and North America; birds were dying from the side effects of pollution; and oil spills had polluted the sea. This concern gathered momentum in the 1960s and culminated in the convening of the United Nations Conference on the Human Environment in Stockholm in 1972.

Founex 1971: A Stepping Stone to Stockholm

In 1971, a panel of experts from both developed and developing countries met in Founex, Switzerland, to clarify the issues and point out the links between development and environment. They identified many serious environmental problems which were largely concentrated in the Third World. The environmental problems of the South are very often the result of extreme poverty and the lack of economic and social development.

For example, it is in the developing countries that water quality, housing standards, sanitation and nutrition are lowest, disease most prevalent and natural disasters most destructive. It is there too that the destruction of vital natural resources – such as soil, vegetation and wildlife – is most alarming, and the impact of a fast growing population most apparent.

The panel also agreed that problems already distressingly familiar to richer countries would emerge in developing ones, perhaps sooner than many realized, as their industry, agriculture, transport and communications expanded. There were differences between the problems of rich and poor countries, but also many shared concerns.

The panel established beyond doubt that developing countries faced vast environmental difficulties of their own.

STOCKHOLM 1971

Growing world concern for the future of the global environment, as most eloquently exemplified by Racheal Carson in *Silent Spring*, was given focus at the United Nations Conference on the Human Environment held in Stockholm 1972. This created the United Nations Environment Programme (UNEP) to act as a catalyst to the nations of the world.

New understanding of the relationship between environment and development emerged from the Stockholm Conference. Firstly, the concept of the environment was expanded. A more holistic approach, encompassing the socio-economic factors which lie behind many of the problems, replaced the old technocratic approach, with its heavy emphasis on the natural sciences and on consequences rather than causes. The link between development and environment was clarified.

The aims of development were also redefined, to make quality of life, rather than a limitless pursuit of material possession, the main criterion of success. Developing countries came to appreciate the need to incorporate environmental considerations into their development efforts. Finally, it was clear that an already complex issue had become even more complex and that the environmental agenda must be expanded.

At the time of Stockholm, Cocoyoc and the World Commission on Sustainable Development there was enthusiasm and progress. In the 1980s various environmental assessments were made, and several environmental conventions and other binding legal instruments were painstakingly developed and agreed.

COCOYOC 1974

Thinking advanced further at the 1974 Cocoyoc Symposium on Patterns of Resource Use, Environment and Development Strategies, which was jointly organized by UNEP and the United Nations Conference on Trade and Development. It agreed on six main points.

First, economic and social factors were often the root cause of environmental degradation. Patterns of wealth, income distribution and economic behavior, both within and between countries, obstructed development and led to inequities.

Second, meeting the basic needs of the world's population was a chief goal for the international community and nation states. The needs of the poorest strata of mankind were particularly urgent, but they must be met without impinging upon the outer limits of the Earth's carrying capacity.

Third, different nations, and groups with them, placed widely differing demands on the biosphere. Access to many cheap natural resources was preempted by the rich. However, it was not only the rich who wasted natural resources: the poor were often left with no option but to destroy them.

Fourth, developing countries must develop their own self-reliant approach to development, rather than following in the footsteps of the industrialized countries.

Fifth, the search for alternative patterns of development and lifestyles, in both developed and developing countries, were the principal means of achieving environmental and development goals.

Finally, this generation must look ahead and not so preempt the planet's limited resources or pollute its life support systems that it jeopardizes the wellbeing – even existence – of future generations.

SUSTAINABLE DEVELOPMENT

At the beginning of the 1970s, the Club of Rome (*see* Club of Rome, Volume 4) formulated the limits to growth' theory. The idea was that the planet's physical limits were being reached. What had seemed infinite to a relatively small human population was now alarmingly finite. If we pushed beyond those limits, the Earth's environment would break down, lose its ability to sustain us and our industries and our agriculture. We would be the end losers.

Of course, the theory is only partly correct – as was shown by later studies, including those of the Club of Rome itself. But it was an eye opener. For we now have the ability to destroy the environment completely. In so doing we would write off the economic machine which feeds six billion people.

The Club of Rome initially assumed that all resources were limited. They pointed to the oil crisis and to the impending dearth of all petroleum products. They pointed out that resources could only be exploited at the margin, and that in most cases we had already crossed the margin of increase. They pointed out that many of the world's fish populations were declining due to overexploitation. They pointed out that more species had been extinguished during the 20th century than in the last million years. Resources that could have sustained a smaller population forever were being snuffed out.

They were right. But they ignored the fact that the Earth still has on essentially infinite supply of sunlight, and that that resource is still largely untapped except through natural means such as photosynthesis and the cycling of water between the oceans and atmosphere. They also ignored the fact that many of Man's economic activities can be effectively altered to make them work with the biosphere instead of in competition.

WORLD COMMISSION ON ENVIRONMENT AND DEVELOPMENT

The World Commission on Environment and Development gave a powerful example of resource depletion in its report *Our Common Future*, published in 1987. The Commission pointed out that most of the new wealth created in the 1980s was created in non-polluting industries. It was the service industries such as information and data which were the hub of present growth.

This growth which works in tandem with the environment is what we call sustainable development'. Sustainable development will be the foundation of economic and social planning in the next century. If it is not, Malthusian laws will take over again. What we see in the Horn of Africa will spread and will take its toll on the whole Earth. Famine in Africa, ooding in South Asia, and the proliferation of small wars are signs of what could happen to a world which fails to regulate its growth according to the constraints of the biosphere.

The 1987 report of the World Commission on Environment and Development, was released at around the same time as UNEP released its own report, *Environmental Perspective to the Year 2000 and Beyond*. The reports are complementary; both were called for by the 1983 Governing Council of UNEP. Both elaborated environmental strategies for agriculture, health, transport, urban planning and other key economic sectors.

At the core of the concept of sustainable development is the requirement that current practices should not undercut future living standards. In other words, present economic systems should maintain or improve the resource and environmental base, so that future generations will be able to live equally well or better. Sustainable development does not require the preservation of the current stock of natural resources or any particular mix of human, physical and natural assets. Nor does it place artificial limits on economic growth, provided that growth is both economically and environmentally sustainable.

Two points, however, are crucial. First, there is nothing new about the link between environment and development and the criteria for sustainable development. We have grappled with them for some time. Sometimes we have succeeded. But so often development has proceeded as usual, becoming the cause – instead of the cure – of ecological destruction.

Secondly, we remain at the starting gate of policy implementation. Economic priorities and ecological goals remain a dialogue of the deaf, both sides presenting generalities, critiques and simplistic alternatives. Too much rhetoric; too much lip service; too little action.

Of course, there are no quick fixes. Sustainable development is not a ready-made policy menu. It is a demanding series of concrete, costed and draconian reforms which confront failed economic policies and instigate new structural adjustment programs. These reforms work to alleviate poverty, meet basic human needs and lead to economic conditions that promote improved environmental quality. Throughout the global South, a rigidly unfair, protectionist international economic order, commodity price volatility, debt and poverty all tighten the stranglehold of natural resource destruction, chronic pollution and skyrocketing population.

THE UNITED NATIONS CONFERENCE ON ENVIRONMENT AND DEVELOPMENT (UNCED) (1992)

The environmental problems facing our planet and its peoples are still very great. Human populations continue to climb worldwide. Larger and larger numbers of people mean ever greater demands on the Earth's natural resources. More land is taken for agriculture and with it go the wilder spaces. As a consequence, planetary biological diversity drops. As standards of living rise, consumer demand increases and so industry spreads. Vast amounts of industrial wastes are released into air and water and so onto the land; as industry spreads so the wastes increase. People, their actions, and their wastes are undermining the physical and chemical basis of global life support systems. Never before in the long history of Earth has a single species ever threatened the entire planet. Life on earth is in danger and humans are the main cause.

Modern scientific understanding of the world in which we live began in the 17th century but comprehension of the true complexity of the interactions between environmental forces that enable the planet to function as a home to life has only emerged during the last five or six decades. Increasingly, it is realized that many of these interacting systems must be looked at on a planetary scale. This means that more attention must be given to investigating how the planet functions as a whole.

It is no longer sufficient to say simply This is what happens.' The quest now is to find out why and how it happens. Only when we have this knowledge will it be possible for us to put environmental and resource management on a sound long-term footing. Only then will true sustainable development be possible. For this we need better data and information. It is often said that we do not need more environmental data, that we have enough, even too much. We now realize that this simply is not true. We need many more data, not fewer – but data of the right kind and for specific purposes, carefully chosen and carefully collected. In the meantime, we must press ahead regardless, using every scrap of existing knowledge and experience that we have; development and people cannot wait for the results of lengthy research and analysis.

To this end scientists are increasingly considering the developmental applications of their research, and have realized that sustainable development is actually both a desirable and an attainable goal. The Ivory Tower approach is no longer the norm.

Most scientists know that they will have to work more and more closely with national planners and managers to translate their findings into practical actions. This is a major and important change in philosophy that will certainly have far reaching benefits for us all.

Nevertheless, the environment continues to worsen. Deforestation and land degradation spread and their rates increase. Poverty is identified as one of the major causes of environmental deterioration as people struggle just to survive in a world where it is becoming increasingly obvious that most natural resources are used, squandered some would say, by a fortunate few at the expense of the unfortunate majority. At the same time, scientists identify whole new categories of global environmental threats whose implications are enormous, devastating – and expensive to evaluate.

More and more NGOs are founded all over the world and they begin to act as the voice of human consciousness alerting people everywhere to what was happening. The larger of these, such as *Friends of the Earth*, *Greenpeace*, the *World Watch Institute* and the *World Resources Institute* attract global followings. However, to many, it begins to seem that everything done by society is shown to have adverse and expensive consequences for the environment.

Those in power in rich countries are appalled at the implied costs and time scales of suggested remedies. Most people in poor countries are already too devastated to notice; their long-term concern is, as always, to keep their families alive. NGOs trying to pressure governments shout constantly and ever louder about what is happening and what is needed.

Not surprisingly, in the rich countries reaction has set in and people have begun to turn away from the environment. Their local environment looks much the same as it always has so the scientists, and environmental prophets of doom are probably wrong. In the poor countries reaction also sets in, but of a different kind. Here remedial measures to control environmental degradation proposed by aid donor countries are often viewed as attempts to hold back national development. Aid programs to help assess natural resources are seen by some governments as attempts by the rich to get specialist knowledge for their own economic gain.

Remote sensing of these resources from satellites is viewed with particular suspicion. This attitude, coupled with the constant accounts of how huge amounts of aid money and supplies are being siphoned off by some developing country leaders for personal benefit, adds in the industrialized nations to the growing sense of disillusionment with environment and development concerns.

This disillusionment reached its peak at UNCED held in Rio de Janeiro in June 1992. Fortunately, the main outcome of the conference was a carefully thought out and well-structured program for the future which would take environment and development safely through the 21st century. Agenda 21 as it is called, idealistic though it may be, is a worthy and genuine blueprint for world action, setting out what needs to be done to ensure that humanity has a long-term future. The postulated cost outlay of Agenda 21 is staggering, although in the long run the returns and benefits will be many times that outlay. What it immediately resulted in, however, is that before the conclusion of the conference, developing countries called upon the rich nations for billions of dollars of additional aid for national development, stressing Agenda 21 and the need to alleviate poverty. There is no doubt that this money is needed, but the size of the demand, coming when it did, coupled with a growing dissatisfaction with what is considered, probably more wrongly than rightly, as an overly-bureaucratic and costly United Nations system made most developed-country governments back away from worries about the world environment and almost all of them from responding properly to requests for sizeable additional financial resources, particularly from the Official Development Assistance (ODA).

UN SPECIAL ASSEMBLY (1997)

Five years after the Rio conference, the UN General Assembly held a Special Session to review the implementation of Agenda 21. The conclusion of the Special Session was that while some improvement was achieved, overall the situation in 1997 was worse than in 1992. What annoyed developing countries most is that instead of increasing ODA as promised at Rio, it fell from 0.34% in 1992 to 0.29% in 1996.

However, people think, learn and, above all, can reason. These are features that make them unique among Earth's life forms. Because of this, there has to be hope. There are signs that people everywhere are starting to take a fresh look at the world. This time it is a quieter, more restrained, less belligerent and more realistic view. There is a growing awareness that environment and development really are linked – in all countries. Environmentalists (and even conservationists) are increasingly recognizing that life without industry is very difficult in the modern world; at the same time industrialists have become aware that environmental care by them may have long-term cash benefits, as well as being good for customer relations. Several developing countries have successfully become industrialized nation states and are starting to realize that environmental regulations and controls really are needed and that they do bring practical benefits to their peoples. Other less developed countries are looking for leadership to these

newly industrialized nations and are starting to emulate their growing environmental awareness and concern.

Many ordinary folk have begun to realize that all peoples on Earth are but slightly different genetic and cultural variants of the same basic human model.

FURTHER READING

Brown-Weiss, E (1988) *In Fairness to Future Generations: International Law, Common Patrimony and Intergenerational Equity*, Transnational, Dobbs Ferry, NY.

IUCN/UNEP/WWF (1980) *The World Conservation Strategy*, IUCN, Gland, Switzerland.

IUCN/UNEP/WWF (1991) *Caring for the Earth: a Strategy for Sustainable Living*, IUCN, Gland, Switzerland.

Meadows, D H, Meadows, D L, Randers, J, and Behrens, W W (1972) *The Limits to Growth*, Earth Islands, UK.

South (Nyerere) Commission (1990) *The Challenge to the South Commission*, Oxford University Press, Oxford.

Tolba, M K and El-Kholy, O A, eds (1992) *The World Environment 1972 - 1992, Two Decades of Challenge*, UNEP, Chapman and Hall, London.

UNDP (Annual Series) *Human Development Report*, Oxford University Press, Oxford.

UNEP (1987) *Environmental Perspective to the Year 2000 and Beyond*, UNEP, Nairobi.

World Bank (Annual Series) *World Development Report*.

World Commission on Environment and Development (WCED) (1987) *Our Common Future*, Oxford University Press, Oxford.

Strategic Environmental Assessment

Robert Goodland
World Bank, Washington, DC

Strategic environmental assessment (SEA) is the systematic and comprehensive process of evaluating the environmental and social impacts of a policy, plan or program and its alternatives. SEA identi es potential impacts, designs measures to mitigate them, then ranks - on environmental and social grounds - all potential projects in that sector, to be used in decision making by stakeholders. The implication is that only the projects at the top of the social and environmental rankings will be taken up, assuming they are similarly good on economic, engineering and other criteria.

Traditional, reactive, project-level EA is necessary but not suf cient to exploit opportunities which exist today but which may be gone tomorrow. Reactive EA must be transformed into proactive or strategic EA. SEAs improve investments over the long term and should be fed by long term projections. Thus, SEA is a process by which environmental implications are integrated into decision making above the project level. SEA extends EA into nontraditional areas such as entire sectors, policies and programs. Sectoral EAs are a subset of SEAs, which began almost a decade ago to great effect. SEA has started internationally, and has been used most successfully in the power and transport sectors, and in water supply.

SEAs, those that look at programs, policies, treaties and other nontraditional areas, are the newest types of EA. SEA is the application of EA above the project-level. Although already promulgated by the European Union and several countries, and although provided for by United States National Environmental Protection Agency (NEPA), SEAs are still not yet routine. For example, during the decades of trade negotiations between USA and Canada, the North American Free Trade Agreement (NAFTA) was not subject to SEA procedures.

SECTORAL SEAs

Sectoral SEA, the commonest form of SEA, is the process of examining potential environmental and social implications of all or most of the potential projects proposed for the same sector. Sectoral SEA can in uence project selection, which project level SEA almost entirely cannot. Sectoral SEAs provide an environmental ranking of all proposed projects in one sector before prefeasibility. This helps in project selection (e.g., gas vs. coal vs. hydro vs. nuclear energy in the power sector; or road vs. rail vs. air in the transport sector).

Sectoral SEA begins with a development objective or goal and then evaluates the numerous possibilities of meeting agreed, desired results. Instead of beginning with, say, a preconceived proposal of a 200 MW coal-fired power plant at spot x, a Sectoral SEA would begin with the premise of meeting projected power needs by optimal methods, including energy conservation and development of renewable energy. Sectoral SEAs reduce the costs of subsequent project level EAs, but do not obviate the need for them. Sectoral SEAs have started to be used, especially in improving traditional least-cost economic analyses into economic, social and environmental least-cost sequences. Power sector expansion sequences increasingly have environmental and social rankings alongside the conventional economic ranking, and this constitutes a form of sectoral SEA. Sectoral SEAs are most frequent in the power sector, less so in transport sector, and rare so far in the agricultural sector.

A strategic EA seeks to integrate environmental considerations into decision-making above the project level, while sectoral SEAs are viewed as a subset of SEAs. In the power sector for example, once the case for new generation

capacity has been made, demand side management (DSM), such as improvements in efficiency, reduction of losses, and conservation, is well in hand, and long-run marginal cost is being achieved, then sectoral SEAs help decide first what sort of new capacity would be preferable: DSM and conservation vs. coal vs. hydro vs. gas vs. oil vs. nuclear energy (*see* Demand Management, Volume 4).

Another value of sectoral SEAs is in the detection of gaps in existing data, thus allowing timely collection of relevant additional data required to make informed decisions about the selection of the next project in the sector. Sectoral EA makes project-specific SEAs much faster, cheaper and more robust because data are already assembled into information management systems and data gaps are already identified. The great power of a sectoral SEA is that it helps rank potential projects in an environmental sequence, so that environmentally sound projects are taken up before environmentally weaker ones. Environmental ranking of all potential projects exposes society to tradeoffs. The ranking itself should largely be as a result of participation by civil society. By such rankings, SEA fosters transparency into long range plans. This helps acceptance by taxpayers and affected people. It also decreases the likelihood of purely political decisions. Pork barrel selection of a project in a politician's home area is less likely because it becomes clear that better or cheaper sites or technologies are available elsewhere.

For example, if an EA team conducts an EA on a nuclear power plant, the team should first explore the possibility that demand management may reduce demand enough to allow postponement of the need to increase capacity. Secondly, the team should determine if other power generation alternatives are economically and environmentally feasible, such as hydro or natural gas. As a general rule, the power sector has long-term least-cost sequence plans which sometimes include environmental considerations, such as human involuntary resettlement and wildlands, but these often get lost in the hurry to complete engineering and economic analyses. These environmental and social costs then reemerge after heavy expenditures on detailed engineering studies have been completed and the proposed project has become more firmly entrenched in national development plans. Sectoral SEAs should identify the real options at a stage before expenditures for design have become too great.

In a recent case, preparation of Nepal's Arun 201 MW hydro was criticized because not enough alternatives sites had been socially, environmentally and economically compared with the single proposed big one. Often, dams are proposed as a part of a cascade series with one large dam controlling the ow to several smaller downstream dams. The SEA must consider the entire proposed sequence when assessing the first dam of the series. This holds even if dams subsequent to the first one are not immediately envisaged.

The same holds for highway projects. If a good SEA team is asked to assess the impacts of a new highway, they may well recommend a railroad, not a highway. Such a response would not please the National Highway Authority that requested the highway SEA and would likely result in dismissal of the SEA team. This might not be the best course for the environment nor for the nation.

STRUCTURAL ADJUSTMENT SEAs

A structural adjustment SEA of a country or group of countries has economic development as its primary goal – through deregulation, privatization of national industries, subsidy removal or free trade, for example. At the March 1995 UN Social Summit, there were major differences over the social and environmental impacts of structural adjustment. Some claimed poverty and equity had decreased after structural adjustment, others the contrary. Some claimed that growth is a prerequisite for reducing poverty and that for growth, structural adjustment is needed. The poor are the first to be harmed by environmental damage. Most subsidy removal harms the poor more than other groups. The environment does not have needs; it is the life support system of all life, including humans. Systematic SEAs of structural adjustments during the design phase would eliminate most major impacts and could readily mitigate other significant ones.

While the precise details of the environmental and social effects of structural adjustment loans (SAL) are too complicated to be teased out, the major impacts are clear, easy to identify, and relatively easy to mitigate. Unless major potential impacts are designed out or otherwise mitigated, SALs impose major social and environmental costs – usually on the poor. Most SALs have much in common although precise details differ from country to country. SEA should analyze the potential impacts of the common elements in each case.

- *Devaluation* (aimed at fostering exports): Will imported food prices rise? Will the poor expand small-scale agriculture into marginal lands?
- *Removal of subsidies* (on food, fuel and agricultural inputs): In general, environmentalists encourage removal of subsidies, because it reduces waste. In specific cases, will higher prices for food and kerosene and other domestic fuels result in impacts on the poor?
- *Removal of fuel subsidies*: Will they lead to increased fuelwood use, with resulting deforestation, loss of biodiversity and carbon fixing capacity, soil erosion, and possible trends in desertification?
- *Removal of agricultural input subsidies*: Will fertilizer and biocide prices rise? Will agricultural intensification increase or decrease? Will extensification increase in developing/developed countries?

- *Developing compared with Developed Country Impacts*: Will the situation in industrial countries be the opposite of that in developing countries? In Western Europe, for example, would removal of price and land set-aside subsidies cause intensification of production on the same plots?
- *Extensi cation vs. intensi cation*: In poorer countries, instead of seeking productivity gains by intensifying production on the same-sized plot, are farmers more likely to extensify into marginal areas, hoping to get the same production from a larger area farmed with fewer inputs? What are the risks of small farmers expanding the agricultural frontier? Will output productivity and biodiversity suffer? Will long-term soil productivity decline?
- *Natural resource depletion*: Will the SALs accelerate natural resource depletion (e.g., soil fertility, fuelwood, timber, fish)?
- *Exchange rate adjustment*: Will natural resources depletion accelerate if the exchange rate falls because the exporting country has to export more, but import the same amount?

SEA of structural adjustment can prevent environmental damage by:

1. Systematic monitoring of environmental trends and anticipating emerging problems when policy reforms are being prepared.
2. Identification and assessment of serious potential environmental impacts of proposed economy-wide reforms, which should begin at the design stage of structural adjustment.
3. Following identification, potential impacts need to be prevented by design changes. For example, in subsidy removal, the first step is to start with the subsidies benefiting mainly the rich (e.g., air travel, limousines, mansions, non-staple foods, electricity and water consumption over a minimal level per household), leaving subsidies benefiting the poor until later or not at all (staple foods, cooking oil and fuel, public transit). Where design changes are not possible, mitigation measures are necessary, such as social safety nets.

SEA of Privatization

Corporations owe allegiance to shareholders; government's first priority is the well being of their citizens. While these two groups overlap to a limited extent, there remains a major role for government regulation following privatization. Privatization should proceed only after an SEA has identified the impacts, such as major unemployment, when a government agency sheds workers as part of privatization. In addition, government regulatory capacity needs to be in place so it becomes adequate as soon as privatization begins. The government needs to ensure that the private sector is well regulated in order to balance shareholder interests with those of the citizens.

SEA of International Treaties

Many international treaties create major impacts, such as enabling polluting industry to relocate to countries where social, labor and pollution standards are less expensive. Such treaties need to be environmentally assessed during the drafting stage because once negotiated it is too late to mitigate, as was the case in the NAFTA. Also, if social and environmental impacts are assessed and agreed upon in advance, signing and ratification is faster.

ENVIRONMENTAL ASSESSMENT OF NATIONAL BUDGETS

The national budget is arguably the most important statement of environmental priorities that any government ever makes. A budget containing two orders of magnitude more money for the military than for education reveals a government's priorities. For example, the US Federal budget shows that the government thinks military threats are 15 times more important to national security than threats from poverty, population growth or the environment. At least at a superficial level, it is relatively easy to identify anti-environmental expenditures in a budget. The most detailed analysis of any national budget so far is that of the US federal budget by Friends of the Earth, which also shows how to cut environmentally harmful expenditures.

The first need is a systematic SEA of the budget to seek to reduce anti-environmental budget expenditures such as subsidies to grazing, mining, water and timber interests. Tax breaks for oil drilling means the oil industry is spreading its own private risk across the public domain, which does not equally participate in any future profits. Subsidies and other inducements to civilian nuclear power are strongly asymmetrical at best. About 75% of total research and development expenditures have been consumed by the nuclear industry over the last four decades, although it generates only about 3% of global commercial energy.

The second need is to identify the pro-environmental expenditures in the budget and compare them with other expenditures to force taxpayers to ask if those are the priorities for which they voted. Also, where defence expenditures are deemed to be excessive, the SEA should highlight this for voter's consideration. The social side of the assessment will address the employment implications of the budget proposals. Fortunately, more environmental attention almost always increases employment, rather than reducing it. Annual reviews of public expenditures can include environmental and social analysis too.

FURTHER READING

Thisson, W, ed (2000) Special Issue on Strategic Environmental Assessment Planning and Policy Making, *Impact Assessment Project Appraisal*, 18, 174–251.

Strong, Maurice

(1929-)

Maurice Strong was born in Manitoba, Canada. He is married with two sons and two daughters and a foster daughter. In recognition of his achievements over the last 30 years, Strong has received honorary Doctorates from 41 universities in Canada, the US and Europe. He has also become a fellow of the Royal Society (Great Britain), the Royal Society of Canada, and the Royal Architectural Society of Canada. Apart from these academic honors, Strong has received close to 30 awards from various bodies in different parts of the world, including the Blue Planet Prize in 1987.

Strong is one of a small number of godfathers of the new environmental movement. He is the architect of the two major United Nations (UN) Conferences, which dealt with environmental issues. As secretary general of the UN Conferences on the Human Environment held at Stockholm in 1972, he toured the globe twice to preach the need for quick action to stop environmental deterioration for the benefit of the human race. He met ominous resistance from developing countries who felt that his call was a plot by the developed countries to stop developing countries from pursing their development activities. Strong did not budge and kept at it until the successful completion of the Stockholm Conference with the adoption of the Stockholm Declaration and Action Plan. The latter was a blueprint for action that needed to be taken by all governments, intergovernmental bodies and non-governmental organizations.

The Stockholm Conference also recommended that the UN establish a new global body to implement the Stockholm Action Plan and to become the environment conscience of the world. This was accepted by the UN General Assembly; and the United Nations Environment Programme (UNEP) was established on 1st January 1972. Strong became its first executive director (under-secretary general of the UN). Under his leadership during the first three years of its life, UNEP set a very strong course forward.

Twenty years later, Strong organized, again as secretary general, the much more difficult UN Conference on Environment and Development held in Rio de Janeiro in 1992. The Rio Conference advocated integration of environmental considerations into all development plans. After more than two years of difficult negotiations during the preparatory process for the Conference, Strong and his secretariat prepared the conference documents, the Rio Declaration, Agenda 21 and the Guidelines for Forest Management. At the Conference two major global agreements: the UN Convention on Climate Change and the Biodiversity Convention were signed and a third, the Convention to Combat Desertification was launched.

These two major achievements are more than enough for anybody to be recognized with high respect in the environment field. However, Strong did not retire then. He now leads and chairs a huge number of environmental institutions including Earth Council, Stockholm Environment Institute, Technology Development, World Resources Institute and is director of the Environment Capital Corporation.

MOSTAFA K TOLBA *Egypt*

Study of Critical Environmental Problems (SCEP)

see SCEP (Study of Critical Environmental Problems) (**Volume 4**)

Study of Man's Impact on Climate (SMIC)

see SMIC (Study of Man's Impact on Climate) (**Volume 4**)

Sustainable Cities Policies

Rodney R White
University of Toronto, Toronto, Canada

Of the many changes in the nature of the human habitat, the process of urbanization is one of the most dominant. Without any encouragement from government, people

worldwide continue to move into cities, such that we are now approaching a situation where half of the world's population is urban. The trend is continuing strongly, especially in poorer countries where this movement started much more recently than in the older industrialized countries. In richer countries, cities now account for nearly 80% of the total population.

As people become urban, they tend to use more resources and produce more wastes. Yet their very concentration provides opportunities to conserve resources and reduce wastes. Until recently this potential has been seized by very few cities and it has been largely ignored by national governments (City of Toronto, 1991). This lack of connection is very evident in documents such as those produced by the Inter-governmental Panel on Climate Change where urban systems receive little or no attention. For their part, only a minority of urban governments have addressed global issues such as climate change. It is as if, by becoming arti cial environments, cities and their inhabitants have decided to ignore the biosphere on which they are utterly dependent.

CONCEPTS

The Global Scale

The concept of sustainable development has evolved as a response to the numerous challenges posed by global environmental change. Given the problems, there must be solutions. If sustainability of the human species and its material wants is in question, there must be an alternative path to lead us out of trouble. Sustainable urban development is an assumption added on top of the first assumption. As half the world's population now lives in cities, there must be a sustainable variant of urban life (Stren *et al.*, 1992).

This is not the place to debate the wider issue of the long-term sustainability of the human species; instead we can examine some of the policy options that might make cities more sustainable than they currently are, without making any guarantees that this way of life can last for ever. In place of the absolute notion of sustainable (for all time) cities, let us consider the encouragement of ecological cities – meaning built environments that operate in harmony with the ecosystems on which they depend.

Impacts on a global scale are a product of the magnitude of the world's population and the throughput of materials used to support their lifestyle (Daly and Cobb, 1990). As people become wealthier, this throughput, or impact, increases inexorably. So far we have not discovered a lifestyle which brings greater longevity and material comfort without increasing the impact of the individual on the biosphere. The following argument relates to the richer countries of the world and the lifestyle associated with their inhabitants.

The Urban Scale

Our present condition of unsustainability is a function of the deepening divorce of our actions (such as buying groceries and driving an automobile), from knowledge of their physical consequences for the biosphere. Improved awareness of this connection is an essential step in developing ecological cities. Several concepts have been developed to make this connection clearer to urban inhabitants and to urban decision makers beginning with Abel Wolman's model of urban metabolism which produced measurements of the main physical inputs – water, food and fuel and outputs – mainly sewage, solid waste, and air pollutants (Wolman, 1965; White, 1994). The measurement of these ows provides an approximate notion of how a city functions as an ecosystem, interacting with the bio-geochemical cycles on which it depends. Valuable as this model is, it has had little impact on urban planning or the individual's awareness of the environmental consequences of his/her daily actions. Perhaps its usefulness was limited by the need to collect data at the level of the entire urban system.

The ecological footprint has been more successful in improving awareness of the link between behavior and action. Developed by William Rees and Mathis Wackernagel, the ecological footprint is an accounting tool of the land requirements for the individual which he or she can measure based on a simple audit of his/her activities for a period as short as one or two weeks (Wackernagel and Rees, 1996) (*see* Ecological Capacity, Volume 3; Ecological Footprint, Volume 3). Repeated assessments at this individual scale have produced estimates of how much land is required to support an individual in a city in one of the rich countries of the world. The range lies between four and eight hectares per person (Onisto *et al.*, 1998), depending on the degree of automobile dependence and the individual's diet, hobbies and commitment to conservation. If all the world's people made similar demands, we would need between three and four Planet Earths to support us. The footprint demonstrates the extent to which our present trajectory is not sustainable.

The ecological footprint provides a strong link between behavior and impact, because it measures the extent of the need to rebalance our demands with the physical possibilities in the long term. Scenarios, based on the footprint concept, give us useful indications of the new directions that we should consider for urban policy. Thus, urban policy has the potential to act as a link between individual behavior and global environmental change. This provides a possible mechanism for repairing the rupture between the behavior of modern urban individuals and those persons'

understanding of their impact on the planet that supports them.

UNDERSTANDING HOLISM

There are several barriers to understanding the linkage between the individual, urban and global scales of behavior and response. One major barrier is the difficulty we have in conceptualizing the city as a singular physical entity, like a lake or a forest. Urban dwellers – if they think about the city as a whole at all – might see it as a mark on a map, or from an aircraft, but the way in which the city responds to the physical environment and, in turn, affects that environment is difficult to envisage.

One opportunity to see the city as a whole in a dynamic context is when the city is stressed or threatened by a physical catastrophe such as a hurricane, an earthquake or, more rarely, an ice storm. One recent example of this response was the landfall of hurricane Georges in September 1998. As for all hurricanes the path of Georges was carefully plotted as it made its way westwards; hourly estimates were made for its probable landfall. Twelve hours before it was expected on America's Gulf coast it seemed that it would strike New Orleans. Nearly all the property damage from hurricanes is caused by the storm surge and accompanying rainfall, to which downtown New Orleans was exceptionally vulnerable, being below sea level. Most of the nearly three million inhabitants left the city by automobile and bus; others were sheltered in a major sports stadium. As it approached land, the hurricane veered to the north-east, just missing New Orleans but nonetheless causing US$10 billion of damage, of which $3.3 billion was insured. The near miss demonstrated the vulnerability of a major city located in a delta and lying in a hurricane zone. Yet people continue to move to the city and industry continues to invest in the region.

An extreme weather event of similar magnitude, in January 1998, had a very different impact. The January ice storm that fell on the Ottawa Valley, Eastern Ontario and parts of New England completely blanketed Montreal and downed most of the power lines into the city. At one time only one line was still functioning and only four hours' supply of water remained in the city. Households without power moved into nearby shelters, but the city was not evacuated because the blocked roads and bridges made mass evacuation impossible and there was nowhere nearby to move to. This eventuality had not been foreseen. Fortunately none of the small fires caused by downed power lines went out of control. Damage amounted to US$2.5 billion (of which $1.15 billion was insured) and generated more than 800 000 insurance claims – the largest number of claims ever made for a single event. Most of the damage resulted from loss of power to household refrigerators and freezers causing food spoilage as four million people were left without power in their homes for several weeks (Munich Re, 1999).

Both events affected entire cities and both could have been very much worse if circumstances had been slightly different. In one case vulnerability resulted from a location below sea level on a major storm track, in the other from dependence on a few above-ground power lines in a Northern city in winter. Once the entire urban system is stressed simultaneously, the vulnerability of modern technology based societies is suddenly exposed – we are not separate from nature after all, but still very much part of it. This revelation might help us to reexamine the conditions necessary for developing more sustainable urban systems, in terms of making preparations to reduce the impact of natural disasters.

POVERTY AND URBAN VULNERABILITY

The above examples of system-wide urban response to extreme weather events are taken from the richest countries of the world and even there, the capacity to respond to sudden stress was weak. Yet the poorer countries contain more than 200 cities with more than a million people, including 20 with more than 10 million (Dowdeswell, 1996). Seven of the world's ten largest cities are in the poorer countries (N'Dow, 1996). The inhabitants of these cities are even more vulnerable to extreme weather events than people in the cities of the rich, due to scarcity of infrastructure of all kinds. They are also vulnerable to the everyday risks associated with widespread poverty.

> Pollution particularly affects those already suffering from malnutrition and infectious disease, which lower their ability to resist chemical pollutants. For most children in the large cities of developing countries, breathing the air may be as harmful as smoking two packs of cigarettes a day.
>
> (World Bank, 2000, 141)

Although the risks of climate change are great, the solution is very obvious. Humanity must adopt technologies and lifestyles that are compatible with a stabilization of greenhouse gas emissions in the very near future. It will be difficult. Yet choices are available to people in those countries, who emit more than a ton of carbon per person per year. The solution to the problem of poverty is far less obvious. Currently the development model on offer is that of the West, based on urbanization, industrialization, heavy use of fossil fuels, and overuse of potentially renewable resources, water and other materials.

In the face of this complex set of problems the tentative offer of a clean development mechanism in a traditional technology transfer format is unlikely to do much either for the long-standing problems of poverty, or for the stabilization of greenhouse gas emissions (Grubb *et al.*, 1999). If even the cities of the richer countries are vulnerable

to climate change, then they are hardly in a position to pass on clean development to cities in the poorer countries. Only when cities in rich countries have implemented such technology and lifestyles at home, will they be able to assess the relevance of this experience for other parts of the world.

THE ECOLOGICAL CITY – OPPORTUNITIES AND RISKS

An ecological city, being one that operates in harmony with the ecosystems on which it depends, should reduce some of the system-wide risks caused by extreme weather events. Thus, a city in New Orleans' location would discourage further development in areas exposed to storm surges or lying below sea level. In Montreal the power supply could be protected by laying power lines underground or by providing more transmission lines into the city. Alternatively, households could provide their own protection by installing backup power supply.

Most cities would enjoy multiple benefits from encouraging the development of the urban forest from a few scattered trees and parks to a denser canopy that would provide energy savings from shade in the summer and protection from the wind in a northern winter. Trees could also be useful for reducing the adverse impact that the city has on the hydrological cycle, whereby stormwater runoff is intensified and infiltration to the groundwater is reduced. Yet increased tree cover would also bring some additional risks. In Montreal many ice-laden trees fell over, or limbs broke off, blocking the roadway and causing property damage. In cities like Toronto with a large population of raccoons, more trees would provide more habitats for a pest that is likely to become a reservoir for a rabies infestation moving up from the United States. The Northeastern United States is also threatened by the spread of Lyme disease which is carried by ticks, hosted by deer which live in the forests, which in turn are being penetrated by urban development (Longstreth and Wiseman, 1989). In drier regions, like the American Southwest, urban forests provide tinder for wildfires which have become recurrent events in recent years (Insurance Services Office, 1997).

The advantages of restoring some natural ecosystem functions are thus, to some extent, balanced by attendant dangers. In the above examples, extending the urban forest cover might encourage the return of the urban threats of disease and fire, just as in the previous century when inadequate water supply led to outbreaks of cholera, typhoid and yellow fever, and wooden buildings were very susceptible to fire. Modern urban risk management is thus an integral element for developing cities that are more in harmony with the physical environment, including both local and global weather events.

MOVING TOWARDS SUSTAINABILITY

The local issues of urban vulnerability are intertwined with the global issues of climate change, stratospheric ozone depletion and acid deposition. The fossil fuel basis of modern urban and industrial technology contributes to many local, regional and global problems. Some urban citizens and governments are aware of these connections and understand that action to reduce one problem may also reduce others. Thus a reduction in fossil fuel use will improve air quality and reduce a community's contribution to climate change. The reduction will also bring immediate economic benefits through fuel savings and a reduction in damage to human health and to buildings and infrastructure. For those cities which are members of Climate Protection Network managed by the International Council for Local Environmental Initiatives (ICLEI), it is the combined benefits that form the basis for policies designed to make their communities more sustainable. So far the number of urban governments with a broad perspective on these issues, and the environmentally aware citizens needed to support their efforts, are few in number. Even in cities like Toronto, which made an early start on reducing carbon dioxide emissions, support for a drive towards a more ecologically harmonious urban system is not extensive and can easily be disrupted by changes in the political situation and external economic factors, such as a fall in the price of petroleum.

In general, Western European cities are much further ahead on this issue than cities in the rest of the industrial world where dependence on the automobile is growing, not decreasing (Nijkamp and Perrels, 1994; Lafferty and Eckerberg, 1998). Hard-won gains against point source polluters like power stations are being swamped by uncontrolled increases in automobile use and urban sprawl. One city which is held up as an example of what can be done to reduce dependency on the automobile is Curitiba, Brazil, which reversed the worldwide tendency toward automobile dependency and actually brought drivers back to public transport, based on buses running in priority lanes. The watchword for this 20 year effort has been that a city must know where it is growing, how and why (Rabinovitch, 1992; Rabinovitch and Leitman, 1996). In Northwest Europe new communities are being built, in which private automobile ownership is excluded. Instead, communal ownership allows shared use and less dependence. Such schemes are even making a foothold in North America. However, these initiatives are the exceptions. Generally, powerful forces still make the shift towards sustainability quite difficult.

One reason why the worldwide trend towards urbanization has not captured the potential environmental benefits over the last 20 years is that the price of petroleum has fallen steadily in real terms, so that it is now cheaper than it was in the early 1970s at the time of the OPEC

(Organization of Petroleum Exporting Countries) oil price hike. In Europe gasoline taxes are significant, but in North America they are very much lower. Cheap oil encourages suburban sprawl which planners have been generally powerless to oppose.

In the face of the mandatory greenhouse gas emission reductions instigated by the Kyoto Conference of the Parties (to the Framework Convention on Climate Change) the industrialized countries (especially those outside Europe) have vigorously searched for ways to avoid major lifestyle changes (such as automobile dependence) and still move towards the Kyoto target. They champion options such as joint implementation and carbon trading, which suggests that they are prepared to pay financially for the privilege of continuing to emit four or five tons of carbon per person per year. One way to encourage them to change their lifestyle is to apply this willingness to pay to a carbon tax, which would then provide support for urban planners to develop more energy efficient land use patterns which are already well understood (Owens, 1986, 1991). Higher residential density makes public transport viable. It is not a complicated concept.

No mention has been made so far in this article about the role of the private sector, although many companies have taken initiatives that urban governments have been reluctant to follow. Some major fossil fuel companies have publicly stated their intention to restructure their operations for a post fossil fuel world. They anticipate that the threat of climate change will eventually force national governments to mandate reductions of greenhouse gases and these companies intend to position themselves to prosper under a new regime (Browne, 1997).

Even if the price signals could be changed, a further obstacle blocks concerted urban action in favor of the adoption of more efficient land use and transportation systems, reduction in other energy uses, fuel switching and full-cost life cycle accounting. There is a sense that the actions of one city will make little difference to a globally distributed and still growing problem. Within the richer world one response to this objection is that there is a powerful demonstration effect from one successful, ecological city which respects the ecosystems on which it depends. Like an antidote to a disease, once it has been discovered, it can be replicated and applied throughout the world – at least to those parts of the world that can afford it.

For the poorer countries, the simple existence of more appropriate technology is not a solution in itself. At the Kyoto conference, there was much discussion of the role of joint implementation (of greenhouse gas reduction technologies) and the development of a clean development mechanism, with rich and poor countries acting in partnership. However, there is deep suspicion among the poorer countries that these proposals are simply a means of richer countries getting credit for reducing emissions without having to implement any lifestyle changes at home. One way – perhaps the only way – to counter this suspicion is for the richer countries to adoption clean development technology themselves. Without such a commitment, the invention of friendlier technologies and the provision of financial and training assistance to poorer countries will fail to bridge the gap between rich and poor countries (White, 1992, 1993).

If the richer countries fail to clear this hurdle, then a paradox may emerge for those individual cities that embrace more sustainable policies. They may become islands of quasi sustainability in a chaotic world of ecosystem breakdown. That is why it is essential that urban governments and their citizens who have developed a more sustainable lifestyle, to get the message out to those societies that either cannot, or will not, adapt to environmental change.

See also: Anthropogenic Metabolism and Environmental Legacies, Volume 3; Greening of Cities, Volume 3.

REFERENCES

Browne, J (1997) Climate Change: The New Agenda, A Presentation to Stanford University, California, 19 May.

City of Toronto (1991) *The Changing Atmosphere: Strategies for Reducing Carbon Dioxide Emissions*, City of Toronto, Toronto.

Daly, H E and Cobb, J B (1994) For the Common Good: Redirecting the Economy towards Community, the Environment and a Sustainable Future, Green Books, London.

Dowdeswell, E (1996) Editorial in Human Settlements, *Our Planet*, 8, 1.

Grubb, M, Vrolijk, C, and Brack, D (1999) *The Kyoto Protocol, a Guide and Assessment*, Royal Institute of International Affairs and Earthscan Publications, London.

ICLEI. http://www.iclei.org.

Insurance Services Office, Inc. (1997) The Wildland/Urban Fire Hazard, ISO Insurance Issues Series, Insurance Services Office, New York.

Lafferty, W M and Eckerberg, K, eds (1998) *From Earth Summit to Local Agenda 21, Working towards Sustainable Development*, Earthscan Publications, London.

Longstreth, J D and Wiseman, J (1989) The Potential Impacts of Climate Change on Patterns of Infectious Disease in the United States, in *The Potential Effects of Global Climate Change in the United States*, Appendix G: Health, eds J B Smith and D A Tirpak, Environmental Protection Office, Washington, DC.

Munich Re (1999) Annual Review of Natural Catastrophes 1998, Munich Re: Munich.

Nijkamp, P and Perrels, A (1994) *Sustainable Cities in Europe, a Comparative Analysis of Urban Energy - Environmental Policies*, Earthscan Publications, London.

N'Dow, W (1996) Change of Habitat. Human Settlements, *Our Planet*, 8, 1.

Onisto, L, Kraus, E, and Wackernagel, M (1998) How Big is Toronto's Ecological Footprint? Using the Concept of Appropriated Carrying Capacity for Measuring Sustainability, City of Toronto, Toronto.
Owens, S (1986) *Energy, Planning and Urban Form*, Pion, London.
Owens, S (1991) *Energy-conscious Planning, The Case for Action*, Council for the Protection of Rural England, London.
Rabinovitch, J (1992) Curitiba: Towards Sustainable Development, *Environ. Urban.*, 4(2), 62–73.
Rabinovitch, J and Leitman, J (1996) Urban Planning in Curitiba, *Sci. Am.*, March, 46–53.
Stren, R E, White, R R, and Whitney, J B, eds (1992) *Sustainable Cities. Urbanization and the Environment in International Perspective*, Westview, Boulder, CO.
Wackernagel, M and Rees, W (1996) *Our Ecological Footprint. Reducing Human Impact on the Earth*, New Society Publishers, Gabriola Island, British Columbia.
White, R R (1992) The International Transfer of Urban Technology, Does the North have Anything to Offer for the Global Environmental Crisis? *Environ. Urban.*, 4(2), 109–121.
White, R R (1993) *North, South and the Environmental Crisis*, University of Toronto Press, Toronto.
White, R R (1994) *Urban Environmental Management. Environmental Change and Urban Design*, John Wiley & Sons, Chichester.
Wolman, A (1965) The Metabolism of Cities, *Sci. Am.*, September, 179–188.
World Bank (2000) *Entering the 21st Century, World Development Report 1999/2000*, Oxford University Press, New York.

Sustainable Development

Sustainable development was popularized as a concept by the World Commission on Environmental and Development (WCED) (WCED, 1987), which defined it as development which meets the needs of the present without compromising the ability of future generations to meet their own needs. Despite the elusiveness of a robust and practical definition (Munn, 1992), scores of definitions have since been proposed. For example, the National Task Force on Environment and Economy (1987) of Canada defined sustainable economic development as development which ensures that the utilization of resources and the environment today does not damage prospects for their use by future generations. In *Caring for the Earth* (IUCN/UNEP/WWF, 1991), sustainable development is seen as development that provides real improvements in the quality of human life and at the same time conserves the vitality and diversity of the Earth. Some scholars have worked vigorously to develop comprehensive definitions. Rees (1990) defined sustainable development as positive socio-economic change that does not undermine the ecological and social systems upon which communities and society are dependent. Its successful implementation requires integrated policy, planning, and social learning processes; its political viability depends on the full support of the people it affects through their governments, their social institutions, and their private activities.

Discussions about sustainable development have been complemented with debates on the meanings of the general concept of sustainability (e.g., Brown *et al.*, 1987; Dovers, 1990; Shearman, 1990), with no clear, widely accepted definition in sight. Robinson *et al.* (1990) offered the following: sustainability is defined as the persistence over an apparently indefinite future of certain necessary and desired characteristics of the socio-political system and its natural environment. Brown *et al.* (1987) provided a useful characterization of sustainability through a set of demands: (a) continued human life on earth, (b) long-term maintenance of biotic resources and agricultural productivity, (c) stable human populations, (d) limited-growth economies, (e) emphasis on local-scale self-reliance, and (f) continued environmental quality. Both sustainable development and sustainability will doubtless remain much discussed concepts, each defying precise, broadly acceptable definition but both providing strong conceptual frameworks in support of debate and action design for improving human welfare.

REFERENCES

Brown, B J, Hanson, M E, Liverman, D M, and Merideth, R W (1987) Global Sustainability: Toward Definition, *Environ. Manage.*, 11, 713–719.
Dovers, S R (1990) Sustainability in Context: an Australian Perspective, *Environ. Manage.*, 14, 297–305.
IUCN/UNEP/WWF (1991) *Caring for the Earth: a Strategy for Sustainable Living*, IUCN, UNEP, and WWF, Switzerland.
Munn, R E (1992) Towards Sustainable Development, *Atmos. Environ.*, 26A, 2725–2731.
National Task Force on Environment and Economy (1987) *Report of the National Task Force on Environment and Economy*, Canadian Council of Resource and Environment Ministers, Ottawa, 18.
Rees, W E (1990) *De ning " Sustainable Development"*, Center for Human Settlements, University of British Columbia, Vancouver.
Robinson, J, Francis, G, Legge, R, and Lerner, S (1990) Defining a Sustainable Society: Values, Principles and Definitions, *Alternatives*, 17(2), 36–46.
Shearman, R (1990) The Meaning and Ethics of Sustainability, *Environ. Manage.*, 14, 1–8.
WCED (1987) *Our Common Future*, Oxford University Press, Oxford.

PETER N DUINKER *Canada*

Sustainable Development of Energy Resources in Small Island Developing States (SIDS)

Anilla Cherian
American Foundation for the University of West Indies, New York, NY, USA

A 1999 United Nations review of the development of sustainable energy resources in SIDS, points out that little or no progress has been made in the majority of SIDS in increasing the availability of energy services. For SIDS, the nexus between energy services and economic development has been complicated by two factors: (a) heavy dependence on imported petroleum for commercial energy needs and (b) the problem of global climate change and the attendant growth of fossil fuel related emissions of carbon dioxide. Although SIDS are among those that contribute the least to the problem of global climate change in terms of low per capita and aggregate greenhouse gas emissions, these states would suffer most from the adverse effects of global climate change (such as sea-level rise, coastal zone inundation, and escalations in the frequency and intensity of hurricanes and typhoons) which threaten their very existence. The sustainable development of energy resources in SIDS is therefore an urgent priority.

ENERGY RESOURCES IN SMALL ISLAND DEVELOPING STATES

Small island developing states (SIDS) share a number of challenges that affect the provision of energy services such as: limited internal markets; lack of economies of scale; high transportation costs; grave vulnerability to natural disasters and to the adverse effects of global climate change, limited availability of human and institutional capacity to manage and use available energy resources on a sustainable basis, and ever increasing pressures on biomass resources.

To provide a better understanding of the linkages between energy and sustainable development in SIDS, two key issues need to be highlighted: energy consumption and the impacts of energy dependence and expanding access to renewable energy technologies (RETs).

Energy Consumption and Impacts of Energy Dependence

For the majority of SIDS, imported petroleum (mainly end-use products) is the chief source of primary commercial energy and is used largely for transportation and electricity generation. It is estimated that petroleum will remain the most important commercial energy source for SIDS in the foreseeable future, despite ongoing efforts to develop alternative energy sources. While the total petroleum consumption of SIDS increased by an average of 3.7% during the period 1992–1995, a number of SIDS registered negative or no growth in their total oil consumption and also experienced declines in *per capita* consumption of commercial primary energy. The ratio of petroleum imports to total merchandise exports in most of SIDS remains comparatively higher than in other developing countries and has remained almost unchanged since 1992 (see Table 1).

The almost total dependence of SIDS on imported petroleum for their commercial energy needs continues to cause severe imbalances in trade and the rising costs of petroleum imports have put a serious drain on limited national financial resources. The recent economic devastation wrought by tropical hurricanes has worsened existing fiscal stresses on SIDS. Prices of petroleum products in SIDS are among the highest in the world. Prices of petroleum fuels landed in the Pacific SIDS for instance, are typically 200–300% higher than international values. The cost of electricity generation using petroleum products is much higher in SIDS given added costs of fuel distribution for small-scale generation systems.

Most SIDS remain heavily dependent on traditional forms of energy such as fuel wood and bagasse (dry pulp residue from sugar canes), particularly in rural and remote areas where biomass products (fuel wood from natural forests, coconut shells, husks and stem wood, residues from crops such as coffee, cocoa, maize, cassava, peanuts and rice) are used primarily for cooking purposes. In SIDS with large-scale sugar cultivation, bagasse is used for fuel in sugar mills and for electricity generation. Unfortunately, the energy secured from biomass usage in rural areas is neither efficiently produced nor efficiently used. Savings accruing from a reduction in SIDS' dependence on imported petroleum or from increased use of energy efficiency and conservation measures and RETs could contribute significantly to improving the economic well-being of many SIDS.

Expanding Access to RETs

Many SIDS have significant renewable energy resources but these resources vary significantly across countries. SIDS have a high and relatively constant supply of solar energy. Direct solar energy is currently used in many SIDS for heating water, and to some extent in the Caribbean for crop drying and processing. In a number of SIDS, the use of small-scale solar photovoltaic (PV) power to provide electricity in rural areas and remote islands with

Table 1 Production, trade and consumption of commercial energy for selected SIDS (thousand metric tons of coal equivalent and kilograms per capita)[a]

Country	Year	Total primary energy production	Imports	Exports	Per capita consumption	Total consumption
Barbados	1993	125	511	24	1973	511
	1994	118	387	29	1582	411
	1995	125	467	30	1707	446
	1996	110	520	33	1742	455
Cuba	1993	1457	12 210	146	1184	12 241
	1994	1863	12 260	126	1182	12 288
	1995	2138	12 266	103	1193	13 377
	1996	2149	13 302	106	1196	13 380
Fiji	1993	48	569	171	480	364
	1994	50	565	171	477	368
	1995	53	583	171	493	386
	1996	53	594	172	497	396
Jamaica	1993	11	3957	87	1586	3842
	1994	14	3984	72	1716	4196
	1995	15	4356	80	1790	4418
	1996	16	4571	81	1843	4591
Mauritius	1993	13	1034	0	657	1334
	1994	9	1012	0	624	1342
	1995	17	1121	0	642	1348
	1996	13	1142	0	730	1361
Papua New Guinea	1993	7879	1067	7624	302	1239
	1994	8022	1047	7767	292	1227
	1995	7315	1035	7088	284	1222
	1996	5888	1035	5659	278	1224
Samoa	1993	3	62	N/A	401	65
	1994	3	59	N/A	380	62
	1995	3	63	N/A	402	66
	1996	3	63	N/A	400	66
Seychelles	1993	N/A	236	N/A	1033	74
	1994	N/A	245	N/A	1099	80
	1995	N/A	240	N/A	1059	77
	1996	N/A	240	N/A	1084	80

[a] *Source:* United Nations (1998).
N/A, not applicable.

isolated pockets of low load densities appears to have been successful but more work on financing and institutional arrangements needs to be done to promote further commercialization.

Wind power and biomass resources vary significantly with location, both within and between countries while hydroelectric power is available only in a few countries. Wind energy does not appear to offer great potential for many SIDS, as wind energy generation is seasonal, site specific and unsuited for atolls where trees block wind paths. Hydropower resources for primary electricity production are extensive only in a few islands for example; Fiji, Jamaica, Solomon Islands, Samoa and Vanuatu, but many island countries have mini hydropower capacity of a few megawatts.

Although traditional biomass fuel usage in SIDS is both inefficient and unsustainable, biomass energy offers tremendous potential for RET applications (see Table 2). Since the management and disposal of wastes is a major concern for SIDS, waste-to-energy or biogas systems, need to be seriously considered for their role in converting organic wastes into sustainable energy and organic fertilizer. Comprehensive technical feasibility assessments for geothermal and wind energy are still to be undertaken. However, feasibility assessments are being seriously considered and were discussed at a recent meeting convened by the Scientific and Technical Advisory Panel of the UN Global Environment Facility in Barbados.

Despite some growth in the use of modern RETs, particularly solar PV systems in SIDS, the wide-scale application and use of RETs has not occurred in SIDS. One of the principal reasons has been the lack of technical and policy-related knowledge concerning RETs within SIDS. Currently most SIDS lack appropriate institutions and do not have adequate technical capacity to evaluate RETs and implement their usage. Policy makers in SIDS need to be able to make informed decisions on the economic and technical viability of appropriate RETs and need to

Table 2 A brief overview of RET applications in selected SIDS

Biomass	Bagasse: Mauritius has initiated a Bagasse Energy Development Program. Bagasse is used as fuel for sugar mills in Barbados, Fiji and Jamaica
	Biogas: biogas systems (differing in size, design and feedstock) have been built in Barbados, Guyana, Jamaica, St. Lucia, St. Vincent and the Grenadines and Trinidad and Tobago
	Biofuels: used to run small diesel generators in Fiji, Tuvalu and Samoa
Solar	PV systems: increased usage in rural areas and remote islands particularly in Pacific SIDS such as Kiribati and Tuvalu, also in Cape Verde and Jamaica
	Solar thermal: most widely used RE application in the Caribbean. Solar water heaters offer exciting market opportunities for the growing tourist and hotel industry, used widely in Barbados, Mauritius and Seychelles
	Solar drying: research on this application being conducted in the Caribbean, particularly in Jamaica and Trinidad and Tobago
Hydro-electric	There is a wide range of hydropower capacity in SIDS. Several islands have mini-hydro power capacity of a few megawatts and others like Fiji, Solomon Islands, Samoa and Jamaica have more extensive installed capacity
Wind	Wind turbines have been used in Cape Verde and Jamaica. The potential for wind energy and in particular small diesel hybrid power systems is being studied by several Pacific SIDS
Hybrid and other new technologies	Wave power, PV and OTEC are some of the other technologies that offer interesting opportunities for SIDS. The Scientific and Technical Advisory Panel of the Global Environment Facility has recommended that OTEC be studied more closely for application in Caribbean and Pacific SIDS

OTEC, ocean thermal energy conversion; PV, photovoltaic.

have a detailed understanding of how best to encourage technology transfer and the national/regional adaptation of RETs.

The wider use of RETs has also been held back because of lagging capital investments and financing of renewable energy systems in SIDS. Sustainable energy development requires affordable credit financing but many SIDS do not have the available resources or the institutional capability to afford the relatively high up-front costs of RETs. Expanding access to RETs in SIDS will require political commitment and financial investments to build appropriate institutional and human capacity, to develop the necessary regulatory frameworks, energy policies, and financing mechanisms and to catalyze public sector involvement.

POLICY OPTIONS FOR THE SUSTAINABLE DEVELOPMENT OF ENERGY RESOURCES IN SIDS

Energy policies are an effective mechanism for promoting integrated economic and sustainable development goals. This section will examine three options for improving the accessibility of sustainable energy services in SIDS.

Improving Energy Efficiency

Energy efficiency measures play an important role in the overall economic productivity of SIDS. One of the crucial components of a sustainable energy system is the improved efficiency of energy end-use, particularly in the tourism, agriculture and industry sectors. Targeted technical options for energy efficiency currently exist within specific target sectors including commercial and residential construction, transportation, manufacturing and agriculture. Given the unique characteristics of SIDS energy consumption patterns and the small internal markets, a detailed understanding of energy efficiency options based on regional needs is vital.

Policy options that support the adoption of energy efficiency measures in SIDS include the following:

1. Identification and adoption of low and no-cost energy efficiency improvements in key sectors such as tourism and agriculture;
2. development of new regulatory and market frameworks that encourage improvements in energy efficiency;
3. establishment of comprehensive national/regional energy policies with clearly defined energy and material efficiency goals;
4. technical cooperation efforts particularly among other SIDS to promote energy efficiency, especially in the areas of technological innovation and adaptation, local capacity building and increased training and information; and

5. coordinated public awareness campaigns to emphasize importance of economic and environmental gains of energy and material efficiency.

Development and Use of Cost-effective RETs

Given the decentralized nature of RETs, renewable energy can clearly provide reliable, cost-effective energy services for populations living in rural and remote areas in SIDS. However, specific policy problems associated with the implementation in RETs in SIDS need to be urgently addressed, such as variations in renewable energy resources availability, variations in local and national production processes, institutional barriers, financial constraints and human capacity limitations that prevent the wide-scale application and adoption of RETs.

In order to increase the accessibility of modern and sustainable energy services, the following policy options should be considered:

1. Introduction of innovative institutional and financing mechanisms such as micro-financing, cooperative credit and leasing arrangements that allow access to energy services especially for low-income rural populations.
2. Building appropriate institutional mechanisms that promote increased investments in decentralized rural energy systems.
3. Creation of an enabling regulatory and financial environment to encourage increased private sector investment in relevant RETs applications.
4. Training of personnel associated with rural energy service systems.
5. Establishment of networks at the regional and subregional levels to encourage cooperation in adapting RETs for use in SIDS and renewable energy related technology transfer.

Institutional Strengthening and Capacity Building for Sustainable Energy Systems

Limitations in institutional and human capacity at national and regional level seriously impair the ability of SIDS to meet the dual challenge of developing sustainable energy resources and responding to global environmental problems. The United Nations Secretary General's Report on the status of the implementation of the Barbados Programme of Action (BPOA) for SIDS highlights two key constraints related to institutional capacity building, namely, the limited availability of human resources and a lack of financial resources for developing and strengthening institutions/mechanisms. Both these constraints restrict the extent to which effective action on sustainable energy issues and global environmental challenges such as global climate change can occur in individual SIDS. The central challenge is therefore to build regional capacity (technical and institutional) that would promote effective policy responses within the Caribbean, develop data and monitoring networks to provide early warning of changes in climate and sea-level, and strengthen legal and institutional mechanisms for addressing the challenge of developing sustainable energy resources.

The specific issues that need to be addressed are:

1. Development of human and institutional capacity for technological assessment and effective adaptation of relevant options in areas such as energy efficiency and renewable energy;
2. Establishment of training programs that would develop technical capacity among SIDS decision makers in the areas of integrated energy planning, energy financing, energy efficiency, rural energy development and renewable energy.

CONCLUSION

Under the aegis of the Alliance of Small Island States (AOSIS) whose member states are most vulnerable to the adverse effects of climate change, SIDS' negotiators have played an active role in international negotiations on global climate change. Focusing on the linkages between the climate change negotiations and energy issues, AOSIS convened a workshop in January 2001 to prepare for the April 2001 meeting of Ninth session of the Commission on Sustainable Development (CSD-9) in which energy was a principal sectoral theme. It is important to note that a key document under consideration by CSD-9 entitled, Energy for Sustainable Development – Options and Strategies , recognizes that one of the challenges to improving the accessibility of energy includes the unique and diverse set of conditions that impinge on SIDS energy supply and energy security concerns. The AOSIS workshop assisted in the preparation and development of SIDS-specific options and issues related to the goal of energy sufficiency for sustainable development. AOSIS negotiators played a key role in the CSD-9 negotiations on energy for sustainable development and were able to ensure that key SIDS concerns such as an emphasis on development of relevant capacity and the promotion of renewable energy resources were included in the final CSD-9 decision on energy for sustainable development. In order to continue to address the dual challenge of securing sustainable energy resources and addressing global climate change concerns, SIDS negotiators need to focus on the most appropriate mechanisms and strategies that could promote energy efficiency, enhance institutional and human capacity in energy management, and promote effective diffusion of RETs.

See also: Sustainable Development Policies in Small Island Developing States, Volume 4.

REFERENCES

United Nations (1998) *Energy Statistics Yearbook: 1996*, United Nations, New York.

United Nations (2000) *Energy and Sustainable Development - Options and Strategies: Report of the Secretary General (E/CN.17/ESD/2001/2)*, United Nations, New York.

FURTHER READING

Ashe, J, Van Lierop, R, and Cherian, A (1999) The Role of the Alliance of Small Island in the Negotiation of the United States Framework Convention on Climate Change, *Nat. Resource. Forum*, 23, 209–220.

United Nations (1994) *Barbados Programme of Action for the Sustainable Development of Small Island Developing States*, United Nations, New York.

United Nations (January 1999) *Status of Implementation of the Barbados Programme of Action for the Sustainable Development of Small Island Developing States: Report of the Secretary-General to the Seventh Session of the Commission on Sustainable Development (E/CN.17/1999/6)*, United Nations, New York.

United Nations (January 1999) *Sustainable Development of Energy Resources in Small Island Developing States: Addendum to the Report of the Secretary General (E/CN.17/1999/6/Add.12)*, United Nations, New York.

United Nations Environment Programme (1999) *Global Environment Outlook*. Earthscan, London.

United Nations Development Programme (1999) *Energy as a Tool for Sustainable Development for African, Caribbean and Paci c Countries*, United Nations, New York.

United Nations (January 1999) *National Institutions and Administrative Capacity n Small Island Developing States: Addendum to the Report of Secretary General (E/CN.17/1999/6/Add.6)*, United Nations, New York.

Sustainable Development Policies in Small Island Developing States

John W Ashe

Permanent Mission of Antigua and Barbuda to the United Nations, New York, NY, USA

Small island developing States (SIDS) share a number of disadvantages: vulnerable economies; dependence upon narrow resource bases and on international trade; lack of economies of scale; high transportation and communication costs; susceptibility to natural disasters, including long-term threats due to sea-level rise in the 21st century; scarce land resources, inordinate pressures on coastal and marine environment and resources; and a limited availability of the means to exploit natural resources on a sustainable basis.

In recognition of these special circumstances and particular vulnerabilities, which are embodied in Agenda 21, which in itself re ects a global consensus and political commitment at the highest level on development and environment co-operation, the United Nations convened a Global Conference on SIDS to translate Agenda 21 into speci c policies, actions and measures to be taken at all levels (national, regional and international) to enable SIDS to achieve sustainable development. At this Conference, held in Bridgetown, Barbados in 1994, the international community, inter alia, concluded that SIDS are a special case for both environment and development because they are ecologically fragile and vulnerable, and because they face particular constraints in their efforts to achieve sustainable development. The Conference adopted the Barbados Program of Action (BPOA) which contained a blueprint for the sustainable development of SIDS.

THE 1999 REVIEW OF THE BARBADOS PROGRAM OF ACTION (BPOA)

At the 1999 review of the 1994 BPOA, by the United Nations Commission on Sustainable Development (CSD), almost 5 years to the date of the Global Conference, one of key areas highlighted for action by SIDS in their quest for sustainable development was the formulation of national and regional sustainable development strategies and policies which would allow for a more effective use of national and regional human, institutional, financial and natural resources, as well as co-operation at the regional and inter-regional levels. In the context of the review of the BPOA, it was agreed that the development of these comprehensive and collaborative national and regional strategies can also provide a solid basis for the more efficient and cost-effective implementation of donor-assisted programs and projects. This was deemed to be especially the case if these strategies were conceived as action-oriented, allowing for a process of step-by-step improvements and adjustments, and as an instrument to promote broader participation of all relevant groups and civil society.

It was also recognized and agreed that the formulation of sustainable development strategies and policies, through transparent and participatory approaches, would require an appropriate national policy framework that would, *inter alia*, set out clear indicators and benchmarks of progress, which, while re ecting the circumstances of individual countries, would also re ect wider goals, including regional objectives, and that such indicators would also provide a

framework for measuring and evaluating the effectiveness of national implementation strategies and international co-operation.

The challenges that SIDS face in formulating and developing policies for sustainable development remain daunting, given their existing political, socio-economic and environmental conditions. Despite commendable efforts to date by them, difficult decisions, complex trade-offs and arduous work need to be undertaken in the policy arena. For SIDS policy planners, planning for sustainable development will require a clear definition of national needs, a better fit between resources and needs, and a more realistic appraisal of the implications of any proposed development, in terms of the demand for resources which remain invariably scarce and the balancing of frequently competing demands of various stakeholders.

In highlighting the progress made to date, SIDS have emphasized their considerable efforts in the pursuit of sustainable development at the national and regional levels, complemented by international support. These efforts include *inter alia*, respective measures which have been undertaken to formulate national development strategies, action plans and policies and to strengthen regulatory frameworks; the establishment and/or strengthening of national institutions to co-ordinate implementation of the BPOA at the national level; and the establishment or strengthening of regional institutions for effective regional co-ordination.

SIDS have also identified significant constraints to the implementation of the BPOA, including the lack of financial support from the international community, inadequate human resources and a lack of appropriate training, inadequate institutional capacity, inadequate capacity for the enforcement of environmental legislation and regulations, and investment resources. Attention was also drawn to: long-term constraints emerging from the globalization of production, distribution and finance, which are based on the ability – in terms of skilled human resources, technology and natural resource endowments – to take advantage of economies of scale and achieve a reduction of costs; and to the short-term constraints emerging from trade liberalization, which have eroded the trade preferences from which SIDS have hitherto benefited.

Assessments of country experiences in institutional and administrative capacity building have highlighted a number of constraints in SIDS. Those shared by most SIDS include a limited availability of human resources, and a lack of training and financial resources. The majority of SIDS have very small populations, which seriously restricts their ability to staff specialized resource and environmental management institutions. Additionally, the brain drain from many SIDS, and the traditional dependence in some SIDS on specialized expertise from expatriates, have hindered efforts at institutional and administrative capacity building. A lack of adequate training has further weakened the local capacity to deal with complex issues. In several SIDS that have established or upgraded institutions charged with resource conservation and environmental protection, inadequate national budgets have seriously limited the scope of their operations. Even in situations where local expertise exists for dealing with environmental protection and resource management problems, shortages of funds hinder action.

NATIONAL POLICIES FOR SUSTAINABLE DEVELOPMENT

While continuing to stress the importance of continued international support, including financial support, to assist them with their efforts to achieve sustainable development, SIDS have begun the process to attain this goal and simultaneously address the constraints through the introduction of policies at the national level. Table 1 summarizes the national actions taken by a selected number of SIDS from the Caribbean, Pacific and Indian Ocean regions.

Sector-specific Policies

Tourism

Tourism is often identified as a promising growth sector in SIDS. Tourism offers one of the few opportunities for economic diversification and has many linkages with other economic sectors, and if integrated into national development plans with adequate provisions for intersectoral linkages, it can contribute to the growth of all tourism-related activities in all of the major economic sectors: agriculture, including fishing; industry; and services, including transportation. At present, the extent of tourism activities in SIDS varies widely between geographical regions, as well as among countries within regions. Likewise, the economic benefits derived from tourism are diverse. In some SIDS (Barbados, Antigua and Barbuda, Seychelles), tourism has become the major contributor to the gross domestic product (GDP), while in others, it is still relatively undeveloped.

For SIDS, the main objectives of the promotion of tourism as a growth sector are to accelerate the growth of national incomes, provide employment opportunities, and generate foreign exchange earnings and government tax revenues. In view of the serious constraints faced by SIDS in promoting export-led development in the context of an increasingly competitive and fast-changing global economy, a large number of SIDS are giving increasing importance to tourism as a source of increasing foreign exchange earnings. Tourism is a labor-intensive activity that does not require a high level of specialization, and if effectively managed, it could exert a strong

Table 1 National actions taken by a selected number of SIDS from the Caribbean, Pacific and Indian Ocean regions

Country	National strategies and policies for sustainable development	Legislative framework plans	Initiatives taken and achievements	Focal areas
Barbados	Established a national CSD to develop a national sustainable policy and plan	Legislative instruments have been put in place to ensure coastal protection and conservation, and all-encompassing environmental management legislation has been proposed	Establishment of a separate Ministry of Environment, Energy and Natural Resources; a solid waste management project is near completion; a coastal zone unit has been created	Ensuring the input and involvement of the private sector and civil society; coastal zones and fragile marine ecosystems
Jamaica	National Environmental Action Plan and numerous other policy documents: national land policy (1997); national industrial policy and the policy for a system of protected areas (1997)	Natural Resources Conservation Act (1991); drafted natural irrigation master plan (1998) watershed management plan, national environment education action plan for sustainable development (1999–2000)	Creation of the Natural Resources Conservation Authority (1991); additional initiatives include the establishment of the Sustainable Development Council (1996) and a Sustainable Network Program (1997)	Water resource management, tourism, taxes, disposal of non-biodegradable materials, economic instruments in environmental management, sustainable indicators
Fiji	Parliamentary approval of a strategic plan for 1999–2001	Parliamentary consideration of a bill to ensure the sustainable use of resources	Establishment of a national council for sustainable development; environmental management units within government ministries, departments and agencies; an environmental trust fund; a tribunal; the strengthening of the focal environmental agency in government and the establishment of an ozone depleting substance and climate change unit, energy conservation unit and natural resource management unit	Development plan highlights the strategies to achieve sustainable economic and social progress at the end of the 20th century and points to the positive indicators for sustainable economic recovery; provision of infrastructure and other essential utilities, provision of core social services, education, health and housing

Solomon Islands	National environment management strategy; incorporation of environmental issues and strategies into 3-year medium-term development strategy	Wildlife Protection Act (1998) and Environment Protection Act (1998) have been approved by parliament	Formulation of a national biodiversity strategy action being pursued as part of the national response to the Convention on Biological Diversity	
Maldives	First National Action Plan (1990); second National Action Plan (1997)	Formulation and enactment of the Environmental Protection Act; ratification and implementation of various global conventions and agreements	Introduction of EIA procedures; establishment of an environment research center; designation of protected areas and species; restructuring of national commission for the protection of the environment	Climate change and sea-level rise, coastal zone management; biological diversity, conservation; integrated reef resources management; integrated water resource management; management of solid waste and sewage; pollution control and hazardous wastes; sustainable agriculture; human settlements and urbanization
Seychelles	Environmental management plan (1990); new plan will be articulated for the period 2000–2010		In the process of finalizing the 1997–2007 Tourism Master Plan; completed national biodiversity action plan for submission to the Convention on Biological Diversity; National Biodiversity Center has been launched	Focus on strengthening the legal and institutional arrangements for environmental protection and natural resources management

Table 2 Some of the major initiatives undertaken by some SIDS

Country	Goals	Measures adopted	Outcomes
Mauritius	Enhance cohesion among economic development, socio-cultural change and ecology	Establishment of a Ministry of the Environment and Quality of Life; Master Tourism Plan to ensure sustainability; legislative measures for conserving the environment in tourism areas	Intense tourism development; diversification of economy
Maldive Islands	Rapid tourism development: planning of physical plants and incentives for foreign capital investment and reduction of environmental impact	Partial moratorium on tourist accommodations; regulations strengthened; Environmental Protection and Conservation Act enacted; environmental impact studies carried out and work on development indicators initiated; measures put in place for the protection of coral reefs and coastal systems and the conservation of ecosystems	Rapid development of the sector; major contributor to all sectors of the economy
Dutch Antilles	Promotion of tourism without overloading carrying capacity	Tourism Master Plan; measures to protect ecosystems; 7-year moratorium on new construction; legal, institutional and organizational changes	Environmental impact minimized

attractive force on the rest of an island's economic system, including the conservation of certain traditional activities, such as agriculture and fishing, and the preservation of an island's cultural heritage and natural assets. However, the sudden, rapid development of tourism can cause significant social and environmental disruptions in some SIDS. Among other things, potential long-run effects could be a reduction in the living standards of high proportions of island populations and a sense of alienation brought about by their lack of access to limited land resources.

More importantly, the fragile ecosystems of SIDS and their generally more limited scope for pursuing alternative development strategies make concerns for the environmental impact of tourism very acute. These are particularly important concerns because the tourism sector, which is almost totally dependent on the natural environment, is viewed by many SIDS as a fast route to social and economic development. The environmental impacts of tourism are attributable to the development of tourism infrastructures and facilities and to the impacts of tourist activities. Rising prices of land increase the pressure to build on agricultural land, land under forest cover, and land close to sandy beaches and coral reefs. Deforestation, and intensified or unsustainable land use, cause erosion and loss of biological diversity.

As more and more SIDS seek to make their tourism industry sustainable, many have begun to develop measures for coping with the adverse social and cultural impacts of tourism. These include: (1), efforts to keep the ratio of the visitor population to the local population at an acceptable level through appropriate diversification of the tourism product and the tourist market, particularly if mass tourism has been promoted for a long time in the past; (2), effective educational programs to raise the awareness of the local population about the good and bad aspects of tourism; and (3), countrywide, effective security measures against crime and drugs, particularly in areas most frequented by tourists.

Some SIDS have also introduced measures for coping with environmental impacts, including: (1), the development of a multidisciplinary approach for rigorous vetting of tourism development proposals, taking into account prospective cumulative impacts of tourism development, and the establishment of environmental standards for approval of projects; (2), appropriate review and revision of existing taxes and subsidies to ensure that they are in fact conducive to sustainable development, and the development of additional economic instruments to better internalize environmental costs; (3), the imposition of appropriate user fees for the use of national terrestrial and marine parks and reserves, and earmarking the accrued receipts for the purposes of improvement and management of sensitive areas; (4), inducing the tourism industry, through appropriate user charges, to meet the full cost of freshwater that it uses, and to contribute adequately to solid waste management and the cleaning of beaches; (5), implementing a sustainable marketing strategy in co-operation with all partners involved in the tourism industry; and (6), encouraging voluntary initiatives and adherence to codes of conduct by the business community.

Table 2 provides a summary of some of the major initiatives undertaken by some SIDS.

INSTITUTIONAL STRENGTHENING AND ADMINISTRATIVE CAPACITY BUILDING

Most SIDS have realized that their inherent characteristics (small size, small populations, ecologically fragile ecosystems, limited natural resources) make it all the more necessary for them to develop effective institutions and an administrative capacity in order to meet the multitude of challenges arising from the imperatives of resource conservation and environmental protection. Recent country assessments reveal that since the 1994 Global Conference, a substantial number of SIDS have taken a number of policy initiatives aimed at institutional strengthening and administrative capacity building. These initiatives include the following:

1. The establishment of a high-level policy-making or co-ordinating body. A considerable number of SIDS have set up high-level bodies to guide and co-ordinate sustainable development policies and measures. These bodies often assume the institutional structure of a high-level commission or council, headed by the head of government or a senior minister, with a broad membership of government departments and, in a number of cases, major groups and private sector representatives. While not all such bodies have policy-making mandates, most have advisory and co-ordination functions among their principal responsibilities. Their establishment and leadership have provided a catalyst for energizing action.
2. The establishment of specialized bodies. Some SIDS have set up new offices or units within government departments, and have entrusted them with responsibilities for specific sectors, such as waste management or fisheries resources. In other SIDS, existing structures have been upgraded and strengthened in terms of administrative status, manpower and funding. The reinforcement has enabled them to deal with the emerging problems more effectively.
3. The enactment of new legislation. A large number of SIDS have enacted legislation to chart a new course for economic growth and development while complying with regional and international legal instruments. Such steps are often followed by the formulation of a national strategy addressing different sectors and issues in the broader context of sustainable development. The enactment of new laws and the subsequent promulgation of executive rules and regulations have provided an overall legal and policy framework for strengthening institutions and building administrative capacity.
4. The introduction of environmental impact assessment (EIA) and economic instruments. In some SIDS, EIA has been applied to an increasing number of projects, making it an important environmental management tool. In others, economic instruments, such as user fees, deposit refunds and tax incentives have been introduced on a limited scale with a view to promoting behavioral changes in favor of sustainable development practices, while at the same time, generating revenues to help strengthen the institutional and administrative capacity. In such instances, the appropriate use of economic instruments is becoming an effective administrative tool for environmental management and resource conservation.

AREAS REQUIRING FURTHER ACTION

As noted earlier, the challenges SIDS face in formulating and developing polices for sustainable development remain daunting. As SIDS respond to the challenges of sustainable development, key areas that will need to be addressed from the policy perspective include:

1. Inadequate enforcement of legislation. For those SIDS which have enacted new legislation, weakness of enforcement has emerged as a major problem; the new laws on resource conservation and environmental protection, essential as they are, take time and substantial resources to enforce. Among other things, enforcement requires the training of personnel and upgrading of law enforcement structures; there are indications that suggest a weakening of the commitment to enforce the laws in recent years in several SIDS. In some SIDS, enforcement has become sporadic; individual instances of resistance to enforcement, often associated with economic considerations, have compounded the problems.
2. Lack of co-ordination. Despite the fact that most newly-created high-level bodies in SIDS are entrusted with the functions of co-ordination, plenty of problems remain in this area. By its very nature, co-ordination is a difficult undertaking; this is certainly made more so in SIDS by the variety of inter-sectoral issues involved, the different types of inter-disciplinary expertise required and the large number of agencies involved. Country assessments suggest that most SIDS remain at the lower end of the performance curve in the co-ordination of sustainable development policies and measures.
3. Lack of pro-active strategy. Recent experience in SIDS shows that government action in institutional and capacity building is often taken in response to emerging needs and/or new regional/international obligations; while emergency responses are necessary, SIDS governments need to adopt a more pro-active stance in institutional and administrative capacity building through a vision of the country's long-run socio-economic development and its institutional and administrative capacity requirements. Such a pro-active stance is necessary for SIDS to improve planning, and to enhance their capacity for tackling future problems.

4. Lack of recognition. Despite some strides in environmental awareness, the problem of inadequate public recognition of the importance of institutions dealing with resource conservation and environmental protection has impeded progress in strengthening these institutions and their capacity; the lack of public recognition has also affected the effectiveness of the co-ordination of sustainable development policies and measures, as the co-ordination mechanisms are not always accorded the authority and recognition due to them.
5. Sea level rise. Simulations obtained from global climate change models indicate the possibility that sea level will rise significantly in this century. For SIDS such as the Maldives, which are not very far above sea level at their highest point, this is a major concern and a serious policy question (Ashe, 1999) (*see* Sea Level, Volume 1).

NOTES

The author served as the Chairman of the Committee of the entire 22nd Special Session of the United Nations General Assembly for the comprehensive review and appraisal of the implementation of the BPOA for the sustainable development of SIDS.

The United Nations Department of Economic and Social Affairs, which monitors the progress made in the implementation of the BPOA for the Sustainable Development of SIDS, adopted at the Global Conference, currently includes 41 SIDS: Africa (Cape Verde, Comoros, Mauritius, Sao Tome and Principe, Seychelles); West Asia (Bahrain); Asia and the Pacific (Cook Islands, Fiji, Kiribati, Maldives, Marshall Islands, Micronesia, Nauru, Niue, Palau, Papua New Guinea, Samoa, Singapore, Solomon Islands, Tokelau, Tonga, Tuvalu, Vanuatu); Europe (Cyprus, Malta); Latin America and the Caribbean (Antigua and Barbuda, Aruba, Bahamas, Barbados, Cuba, Dominica, Dominican Republic, Grenada, Haiti, Jamaica, Netherlands Antilles, St Lucia, St Kitts and Nevis, St Vincent and the Grenadines, Trinidad and Tobago, United States Virgin Islands.

REFERENCES

Addendum to the Secretary-General's Report on the Barbados Program of Action – National Institutions and Administrative Capacity (E/CN.17/1999/6/Add.6), 19–30 April, 1999.

Addendum to the Secretary-General's Report on the Barbados Program of Action – Sustainable Tourism (E/CN.17/1999/6/Add.11), 19–30 April, 1999.

Ashe, J W (1999) Small Island Developing States and Global Climate Change: Overcoming the Constraints, in Natural Resources Forum, Special Issue on Small Island Developing States, Vol. 23, No. 3, August, 1999.

Barbados Program of Action for the Sustainable Development of Small Island Developing States (1994) Adopted at the Global Conference for the Sustainable Development of Small Island Developing States, Bridgetown, Barbados.

Decision 6/4 of the Commission on Sustainable Development on the Review of the Program of Action for the Sustainable Development of Small Island Developing States, April, 1998.

Report of the Meeting of Representatives of donors and small island developing States, 24–26 February, 1999, Report of the Secretary-General of the United Nations to the 7th Session of the Commission on Sustainable Development, E/CN.17/1999.18, 19–30 April, 1999.

State of Progress and Initiatives for the Future Implementation of the Barbados Program of Action (BPOA) for the Sustainable Development of Small Island Developing States (SIDS) (1999) Draft Chairman's Compilation on the Commission on Sustainable Development to the 22nd Special Session of the United Nations General Assembly, April, 1999.

Status of Implementation of the Barbados Program of Action for the Sustainable Development of Small Island Developing States (1999) Report of the Secretary-General to the Seventh Session of the Commission on Sustainable Development, E/CN.17/1999/6, 19–30 April, 1999.

Sustainable Development Policy

John Middleton
Brock University, St. Catharines, Canada

Sustainable development is an approach that seeks to provide quality of life for all people. It is based on the notion that economic growth (particularly in poorer countries) and social equity are essential, but that they alone are not adequate; life in environmental harmony with the planet is also essential, even from a purely human point of view. The well-known de nition of the Brundtland Commission (World Commission on Environment and Development, 1987) is widely accepted in principle: " sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their own needs."

*The global community invested heavily in negotiations on sustainable development in the 1990s, trying to agree on ways to put this principle into practice. The Rio Earth Summit in 1992 (*see *UNCED (United Nations Conference on Environment and Development), Volume 4) was a mammoth event, and included the largest meeting of national Heads of State and of Government on any subject up to that time. Post-Rio negotiations on parts of the sustainable development agenda have seen even larger of cial summits.*

As a result, our collective understanding of the relationship between sustainable development and global change, and our ideas on how to proceed towards solutions, are much more sophisticated now than they were 10 years ago. On the other hand, the critical link between principle and practice, or between theory and action, still requires much work.

The global negotiations on sustainable development (Box 1) sometimes resemble an exotic ritual theatre with words (many, many words) bearing little resemblance to the everyday reality of our planet. On the surface, there is little relation to the biophysical understanding of environment, and even less to the complex science of atmospheric chemistry and global climate change. Nevertheless there is a compelling logic, albeit not primarily a scientific one. Sustainable development policy often makes more sense when organized along the following four dimensions.

1. *Ecology, economy, society*: the sustainable development approach to global change, as to other issues, is based on the idea (or theory, or article of faith), that development and environment need each other. That is, considered properly, there is no necessary trade-off between environmental and human values, with one or the other winning at the expense of the other. Instead, any rational, sustainable policy will be a seamless integration of ecological, economic, and social factors, sometimes expressed as the three E's, i.e., ecology, economy, and equity. Climate change policy offers a clear example. It would be a foolish waste of time to try to deal with global climate change by trying to ignore the enormous economic forces at stake linked by way of carbon to the fossil fuel economy. Equally, it would be politically impossible (even before entering into questions of global justice) to negotiate a global climate change policy without recognizing that some countries will have to take more responsibility for action than others, based on their respective levels of socio-economic development – *common but differentiated responsibilities* (*see* Common but Differentiated Responsibility Principle (Stockholm/Rio), Volume 4), according to the standard jargon. By this logic, a sustainable development approach incorporating ecological, social, and economic factors is not the best approach, it is the *only* feasible approach to a global climate policy. This is a technical, not a moral conclusion.
2. *Scale*: sustainable development makes explicit statements about scale, both spatial and temporal. The environmentalist agenda once focused on ensuring a healthy environment for future generations (temporal or intergenerational equity). The sustainable development agenda has evolved to give equal attention to the environmental problems faced by the current generation (spatial inequalities). This understanding was key to getting all countries, not just the rich minority, to the negotiating table at the same time.

   The nature of proposed actions has also evolved considerably through the course of the sustainable development negotiations. Once it was assumed that most actions would be by national governments, because the biophysical reality of global change demands action at the national and global scales, and because national governments are the traditional actors in global diplomacy. Now it is increasingly accepted that real change will also require action closer to the ground, by municipalities, communities, private companies, and local non-governmental organizations (NGOs). For example, one of the most useful products of the sustainable development negotiations is the Best Practices Database produced by the UN Centre for Human Settlements in conjunction with the Habitat II conference. The database lists dozens of practical (albeit incremental) initiatives relevant to climate change, in categories such as land use management and environmental management, for emulation elsewhere, most at the municipal or community (rather than national) scale.

Box 1 The principal sustainable development negotiations

- United Nations Conference on Environment and Development (UNCED); Rio de Janeiro, 1992, and subsequent annual meetings of the United Nations Commission on Sustainable Development (CSD) since 1993;
- Convention on Biological Diversity (CBD); Rio de Janeiro, 1992, and subsequent meetings of the Conference of the Parties (COP) to the convention in Nassau (1994), Jakarta (1995), Buenos Aires (1996), Bratislava (1998) and Nairobi (2000);
- Framework Convention on Climate Change (FCCC); Rio de Janeiro, 1992, and subsequent meetings of the COP in Berlin (1995), Geneva (1996), Kyoto (1997), Buenos Aires (1998), Bonn (1999) the Hague (2000) and Bonn (2001);
- World Conference on Sustainable Development in Small Island Developing States; Bridgetown, Barbados, 1994;
- International Convention to Combat Desertification and Drought Particularly in Africa; Paris, 1994, and subsequent meetings of the COP in Rome (1997), Dakar (1998), Recife (1999) and Bonn (2001);
- International Conference on Population and Development (ICPD), Cairo, 1994;
- World Summit for Social Development (WSSD), Copenhagen, 1995;
- Fourth World Conference on Women (FWCW), Beijing, 1995;
- United Nations (UN) Conference on Human Settlements (Habitat II), Istanbul, 1996.

Local scale actions can contribute to sustainability only if they are widespread and general; otherwise, improvements in one area may be at the expense of increased problems elsewhere. For example, a rich country could reduce its emissions of greenhouse gases by moving some heavy industry to developing countries, but the net impact on global climate would be nil. Calculation of net impact requires comparable data from both countries, in other words, a monitoring and reporting system with agreed standards among all countries. Unfortunately this remains a very controversial topic, tied up with perceived (or real) threats to national sovereignty. In the short term, progress may be easier (albeit still difficult) by way of bilateral agreements between two countries, under the Actions Implemented Jointly (AIJ or JI – Article 6.1) provisions of the Kyoto Protocol.

3. *Reactive, anticipatory, radical*: responses to environmental problems such as climate change can be thought of on a continuum from reactive to anticipatory to radical (or fundamental). The proposal to store liquefied carbon dioxide in ocean trenches, for example, is reactive: it attempts to undo, partially and temporarily, a symptom of global change rather than deal with causes. An anticipatory policy, such as for example more concentrated urban design, aims to reduce emissions in the first place, working within existing mechanisms and structures of the economy. A radical (in the word's original sense of root) policy, bioregionalism or deindustrialization for example, questions the fundamental assumptions of the basic social economic model. Policies at all points on the continuum can be useful. In general, the sustainable development negotiations have tended to shy away from decisions at the extremes of the policy continuum. The poorer countries tend to be suspicious of reactive options, as these leave unquestioned the unsatisfactory status quo that keeps them in a position of weakness. On the other hand most parties to the negotiations are suspicious of radical options that are impossible to apply given the fragile political consensus that keeps all parties at the negotiating table. Judgments about what is too radical have evolved over the course of the 1990s, especially as NGOs have been allowed a larger and more official role in the negotiations. For example, resolutions accepted unanimously by the world community to eliminate poverty (World Summit on Social Development, 1995), and to provide adequate shelter for all (Habitat II, 1996) are considered fundamental breakthroughs in the context of the UN system.
4. *Values, ideology, strategy*: all thoughtful actions are ultimately grounded in some set of basic values. The values are typically crystallized in an ideology, a short form summary of basic values that eliminates the need to engage in deep philosophical investigations every time action is required. In turn, the ideology is put into place via some coherent strategy linking theory and practice (Macdonald, 1991). The basic ideology of the sustainable development negotiations is very explicit: Human beings are at the centre of concerns for sustainable development. They are entitled to a healthy and productive life in harmony with nature (Rio Declaration on Environment and Development, Principle 1, 1992). The remaining 26 principles of the Rio Declaration, and the many follow-up negotiations (Box 1) have expanded and elaborated this fundamental statement, but have remained notably consistent with it. In this sense, sustainable development is clearly not just another word for environmentalism. The links among values, ideology, and strategy are incomplete until the results of actions are compared against basic values, to check whether the right results are being attained in practice. In the end, sustainable development policies will have to be judged by their real impacts on the world, just as will competing ideologies such as environmentalism or global free trade.

International decisions cannot be implemented in the same way as national ones; the UN is not a super government. Indeed member countries of the UN are obsessive in avoiding any language that would compromise national sovereignty. Instead, the in uence required for making decisions and taking actions at the global scale comes from two different sources.

The first source of in uence is by way of the traditional mechanism of diplomacy and negotiation among national governments for mutually beneficial goals. Sometimes reluctant nations feel compelled to enter such negotiations based on fear of economic pressure, although other kinds of in uence (for example military, or neocolonial) are still sometimes in play as well. In other circumstances national self-interest is clearly advanced by forming alliances. A classic example is the Association of Small Island States (AOSIS). Individually the nations involved are tiny and almost powerless, whereas together they form an important lobby within the UN's one country-one vote system. AOSIS has become an important force in support of quicker action on global climate change, based on the reasonable fears of some of its members that their whole countries may disappear as a result of modest increases in sea level.

The second source of in uence available for implementing sustainable development policy comes in the form of pressure from members of civil society on their own national and sub-national governments. This is a relatively new development in most parts of the world, but one that has the potential to change significantly the actions of national governments. In the preparatory meetings for

the Rio conference, security guards were routinely posted outside committee rooms to ensure that only official government delegates were permitted to enter. At Rio the large parallel meeting of NGOs caught the world's attention. By the time of the Desertification negotiations, NGOs were participants at the subcommittee level, and this changed the philosophy of how to attack desertification, stressing community-scale approaches. By the time of the FWCW (Fourth World Conference on Women) in Beijing, NGO complaints about how they were being treated by the organizers very nearly forced national governments to cancel their own participation. In the end almost 10 000 non-governmental delegates participated. By the Habitat II conference in Istanbul in 1996, non-governmental and community-based organizations, including municipal governments, were participating on an equal legal footing with national delegations in some circumstances, and were beginning to be included as part of national delegations. It is no coincidence that this blossoming of the in uence of civil society coincided with the rapid expansion of the Internet as a routine means of communication beyond most direct government control.

As national governments were engaged in the official sustainable development negotiations, they were simultaneously coming to an in uential agreement on more open world trade through the General Agreement on Tariffs and Trade (GATT) and the World Trade Organization (WTO) process. These parallel trade negotiations also use the phrase sustainable development, but have used a strikingly different rhetoric, and different ideological assumptions about means and ends. Sustainable development argues that ecological, economic, and social factors are and must be inextricably linked. By contrast the WTO produced a Singapore Declaration on the occasion of its first summit in 1996 that mentioned environment only once, in paragraph 16, referring to the existence of a subcommittee. The same document argued strongly against any attempt to use discrepancies in income or working conditions as grounds for decisions about international trade, that is, against linking economic with environmental or social issues. The differences between the two sets of negotiations were stark, even though the national governments participating were almost the same.

Despite the differences, there are some interesting signs of movement towards a common understanding that may well provide the most promising tool for a practical sustainable development approach to global change. On the one hand, the traditional sustainable development community realizes that it must of necessity come to terms with, and harness, the tremendous power of the private sector and of the international market if it is to make practical changes in the world. The Global Environmental Facility, a joint program of the World Bank, the UN Development Programme, and the UN Environment Programme, has provided useful experience. The official participation of the CBD (Convention on Biological Diversity) in activities of the WTO, linking, for example, forestry practices, timber export markets, carbon cycles, and global change, is an even more promising development.

On the other hand, the WTO is learning that an exclusively technocratic, market-based economic approach is not adequate even when there is near unanimity among national governments. The spectacular failure of the Seattle summit of the WTO (December, 1999) was a very public demonstration of this new reality. The experience with the Multilateral Agreement on Investment was also educational. This agship initiative, negotiated mainly behind closed doors by the world's most economically powerful countries, was in the end defeated, primarily by the concerted action of non-governmental and community-based organizations, using the experience and tools learned during the sustainable development negotiations, among others, on how to use the internet as a sophisticated global-scale organizational tool. Since that embarrassing episode, the WTO rhetoric has evolved quickly to bring social and environmental considerations closer to the heart of the issue. The WTO Secretariat's Trade and Environment Report of October 1999 is essentially indistinguishable from the view of the sustainable development lobby:

> ••• [T] here is no basis for the sweeping generalizations that are often heard in the public debate, arguing that trade is either good for the environment, or bad for the environment. The real world linkages are a little bit of both, or a shade of grey. Win-win outcomes can be assured through well designed policies in both the trade and environmental fields.
>
> (World Trade Organization, 1999)

Actions rather than words will determine whether sustainable development policy is a breakthrough in decision-making at the global scale, or a costly waste of time and resources. Great progress has been made at the conceptual level, but progress towards implementation of significant changes remains very slow. It is now difficult to imagine action that does not incorporate the logic of sustainable development policy, but it remains unclear whether this framework is sufficient to bring about changes of the scale necessary to confront the social, economic, and ecological impacts of global environmental change.

REFERENCES

Macdonald, D (1991) *The Politics of Pollution*, Mc Clelland and Stewart, Toronto.

World Commission on Environment and Development (1987) *Our Common Future*, Oxford University Press, Oxford.

World Trade Organization (1999) Trade and Environment Report, http://www.wto.org/wto/english/tratop_e/envir_e/stud99_e.htm.

FURTHER READING

International Institute for Sustainable Development, http://www.iisd.ca.

United Nations Centre for Human Settlements (UNCHS) Best Practices Database, http://www.bestpractices.org.

United Nations Commission on Sustainable Development, http://www.un.org/esa/sustdev.

United Nations Environment Program, http://www.unep.org.

United Nations Framework Convention on Climate Change. Report of the Conference of the Parties on its Third Session, Held at Kyoto from 1 to 11 December 1997, http://www.unfccc.de/resource/docs/cop3/07.htm.

Sustainable Energy Policies

see Sustainable Energy Policies (*Opening essay*, **Volume 4**)

Sustainable Transportation

Richard Gilbert
Centre for Sustainable Transport, Toronto, Canada

Travel and the movement of freight have been part of human experience since the migrations of our distant ancestors out of Africa, first to Europe and Asia and then to Australasia and the Pacific islands. Among the most remarkable journeys have been those to the Americas: from Asia in the millennia before history - across what is now the Bering Strait to as far south as Terra del Fuego - and from Europe and Africa during the last millennium and perhaps before.

Societies across the world have prospered in military conquest, trade, and economic progress to the extent they have rationalized the movement of people and freight. Over the years, effective transport brought advantage to numerous peoples: the Phoenicians, Romans, Mongols, Venetians, Incas, Dutch, British, and Americans, among others. During the last 200 years, the links between transport and economic progress have become especially tight.

Until the 19th century, travel everywhere was uncomfortable, dangerous, and enormously time-consuming. Freight movement posed even greater difficulties. Rail transport made the difference. The linking of two earlier inventions - wheels on smooth iron rails and the steam engine - allowed widespread motorized transport across land, and the beginning of a new era in the mobility of people and goods. Also important was the linking of the steam engine to the paddle wheel and propeller to provide motorized water transport.

Rail transport began to give way to road transport in the first part of the 20th century, although the main expansion in the use of road vehicles has occurred since 1945. Air transport arrived soon after motorized road transport, allowing high-speed travel over great distances and ready access to remote places. Ocean freight still dominates the carriage of products and raw materials.

Motorized transport has facilitated and even stimulated just about everything now regarded as progress. It has helped expand intellectual horizons and deter starvation. Comfort in travel is now commonplace, at a level hardly dreamed of in former years even by royalty, as is ready access to the products of distant places.

The growth of personal road transport - chiefly the automobile - has been closely associated with two of the major phenomena of the 20th century: growth in material well-being and extension of voting rights. Ownership and use of an automobile - usually the most expensive of consumer purchases - have assumed in rich countries the status of democratic rights. As tokens of passage into adulthood, they can be more important than the ability to vote.

The evident benefits of mass motorization have begun to pall, and a strong case can be made that the disadvantages it brings are beginning to outweigh the numerous advantages. This case has been put well by Greene and Wegener (1997) "Near the end of the 20th century, the belief in the desirability of perpetual growth in mobility and transport has started to fade. In metropolitan areas, the myth that rising travel demand will ever be satisfied by more motorways has been shattered by reappearing congestion. People have realized that the car has not only brought freedom of movement but also air pollution, traffic noise and accidents. It has become obvious that, in the face of finite fossil fuel resources and the need to reduce greenhouse gas (GHG) emissions, the use of petroleum cannot grow forever. There is now broad agreement that present trends in transport are not sustainable, and many conclude that fundamental changes in the technology, design, operation, and financing of transport systems are needed."

TRANSPORT'S RESOURCE USE IS UNSUSTAINABLE

Figure 1 shows the remarkable increases in the motorized movement of people and freight during the last 150 years, particularly since 1945. Worldwide averages are shown; they hide up to 100-fold differences between countries in the amounts of movement of both people and freight.

Transport makes use of large portions of the world's renewable and non-renewable resources. Vehicle manufacture is a major user – and re-user – of steel,

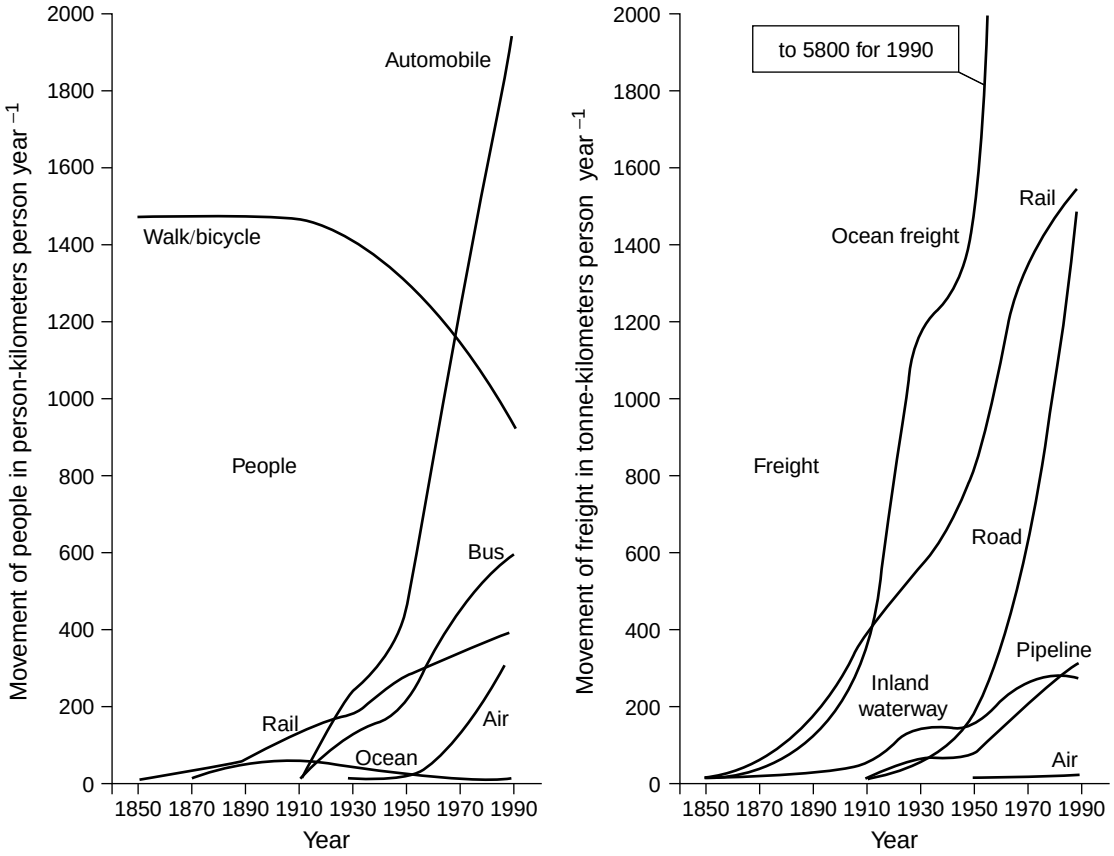


Figure 1 Worldwide *per capita* annual movement of people and freight by motorized transport since 1850, showing for comparison estimates of average distances moved by walking/bicycling. The graphs comprise estimates based on several sources, notably Mitchell (1992, 1993, 1995)

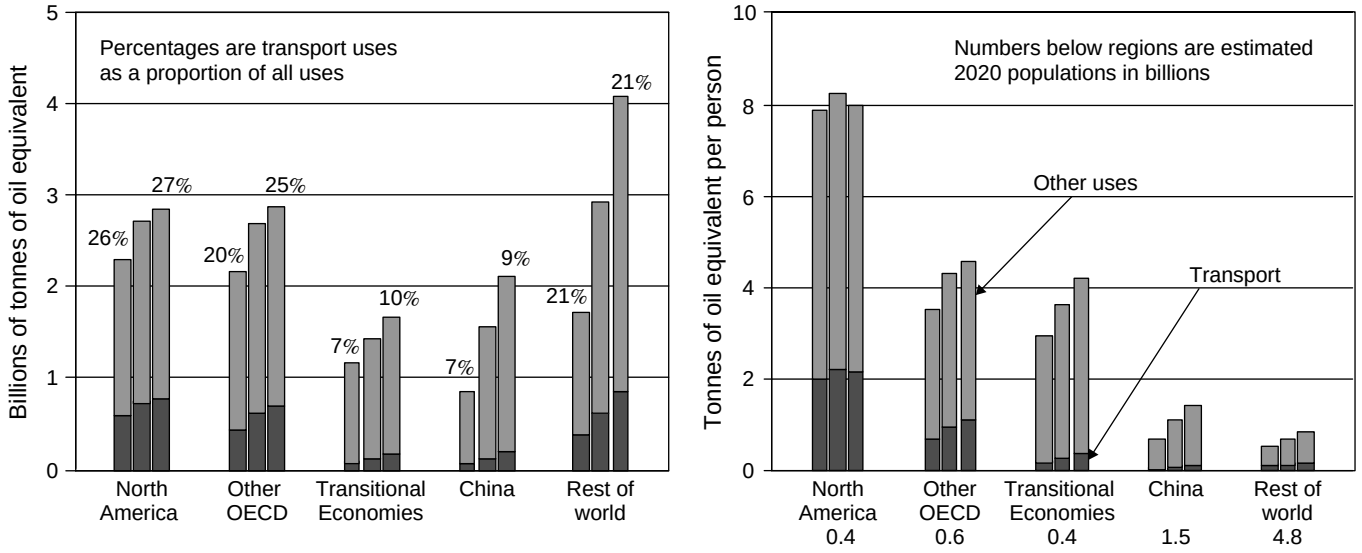


Figure 2 Total and *per capita* energy use for transport and other activities, 1995, 2010, and 2020. (Based on data presented in Part 4 of IEA, 1998)

aluminum, numerous minerals and rarer metals, and plastics. Highway construction involves between one-third and one-half of all material ows for buildings and infrastructure. Above all, transport is a user of petroleum (crude oil), whose refined products provide the energy for more than 95% of the motorized movement of people and freight. Transport accounts for about 60% of the total use of oil products. (These products also heat buildings, fuel industrial processes, and serve as basic ingredients for over 6000 products including numerous plastics and pharmaceuticals.)

Figure 2 shows that in 1995 the average North American was responsible for the use of about two tonnes of oil for transport. About two-thirds of this was for personal automobile use; most of the remainder fueled freight trucks and trains. If indirect energy use for transport is counted – including the energy involved in the manufacture of vehicles, the construction and maintenance of infrastructure (mostly roads), and the production of fuels – the average North American's transport energy use in 1995 was closer to the equivalent of three tonnes of oil.

The projections for 2010 and 2020 in Figure 2 are the business-as-usual (*see* Business-as-usual Scenarios, Volume 5) projections of the International Energy Agency (IEA, 1998). Without special intervention, energy use for transport is set to increase dramatically in all regions and at a higher rate than energy use for all other purposes (by 78% vs. 6% from 1995 to 2020). Transport activity is expected to increase even more than transport energy use, because the IEA's projections assume some increase in the energy efficiency of transport.

The rates of increase in use of energy for transportation are projected to be greater in the poorer countries of the world, represented by the transitional economies, China, and the rest of the world groupings in Figure 2. Their energy use for transport is projected to increase by 142% between 1995 and 2020, compared with an average increase of 46% in the richer countries of the Organization for Economic Cooperation and Development (OECD). *Per capita* rates in the poorer countries will remain relatively low.

While demand for oil continues to grow, production of conventional oil, i.e., cheap oil, is projected to peak within the next 15 years (IEA, 1998). Consequent price rises will increase the economic feasibility of production of other forms of oil and of other fuels. The most promising early replacement for conventional oil is oil from oil sands, which already comprises a quarter of Canada's oil production, and is set to comprise more than half by 2005. Large deposits of material from which oil can be processed, including shale and bitumen, are also found in Russia, the United States, Venezuela, and other countries.

Along with many other replacements for conventional oil, oil from oil sands has a high energy cost of production. The use of this extra energy, plus local adverse impacts, means that the environmental cost of transport activity will increase, other things being equal (*see* Oil Shales and Tar Sands, Volume 3).

There are readier supplies of two other transport fuels, coal and natural gas. Coal cannot be burned cleanly and efficiently in vehicles. Natural gas can, but it is difficult to transport and its use – as with all fossil fuels – results in globally acting emissions. The cleanest, readily available transport fuel is electricity from renewable sources (solar, wind, moving water), but its use in freely moving vehicles is limited by battery technology. The cleanest fuel for freely moving vehicles is hydrogen, but its potential is limited by challenges in production, storage, and conversion to motive power.

TRANSPORT HAS MANY ADVERSE EFFECTS ON THE ENVIRONMENT

Transport makes a substantial adverse contribution to global climate change. The largest part comprises emissions of carbon dioxide (CO_2), the most important GHG. CO_2 emissions are almost precisely correlated with fossil fuel use. Transport activity results in release of several other GHGs, chie y methane (natural gas), hydrochloro uorocarbons (HCFCs) used as the coolant in vehicle air-conditioners, and nitrous oxide (N_2O, produced in vehicle pollution-control devices). Further GHGs are produced by aircraft at cruising heights, effectively doubling or tripling the contribution of aviation to climate change (Intergovernmental Panel on Climate Change, 1999). In total, transport contributes directly and indirectly well over a third of the current human contribution towards climate change. This is the main reason that transport is presently unsustainable.

Stratospheric ozone depletion is another global impact of transport, mostly through release of HCFCs from the air-conditioning systems that feature in most of the world's vehicles. Their more potent predecessor coolants – chloro- uorocarbons (CFC) – are being phased out. As a consequence, estimation of the current contribution of transport to ozone depletion is difficult.

Transport has numerous regional and local effects that are unsustainable to the extent that they result in cumulative, irreversible changes in ecosystems or human health. Most of these effects involve emissions into the atmosphere from fossil fuel combustion. In OECD countries, such emissions of harmful substances from road transport have mostly declined in recent years in spite of the large increases in transport activity indicated in Figures 1 and 2.

At the leading edge of these changes – both in increased transport activity and in reduced environmental impacts – is the United States, whose trends for emissions from transport of the major local and regional pollutants are illustrated

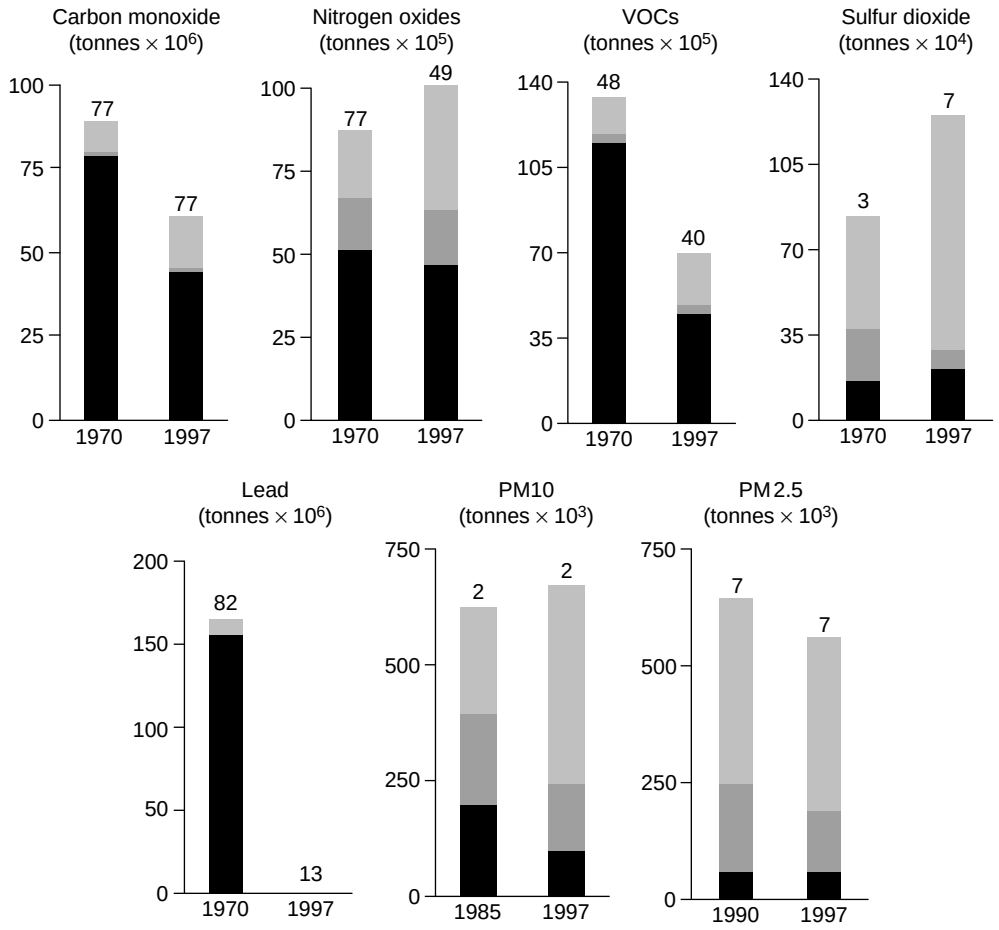


Figure 3 Trends in emissions of local and regional pollutants from on-road and other transport activity in the United States. The bar graphs are based on data provided in US EPA (1998). Scales, units, and base years vary according to extent of emissions and availability of data. Black parts of bars show emissions from gasoline-powered road vehicles; dark grey parts show emissions from diesel-powered road vehicles; light grey parts show emissions from non-road vehicles and engines, including aircraft, commercial and pleasure watercraft, agricultural and construction vehicles, and several other categories. VOCs are VOC; PM10/2.5 is particulate matter of diameter less than 10/2.5 μm. Numbers above bars show on-road and non-road transport emissions as a percentage of all such emissions from transport and other sources for the respective years

in Figure 3. Total amounts emitted from road vehicles, particularly gasoline-powered vehicles, have declined on account of improved vehicles (notably through the use of three-way catalytic converters in the exhaust systems) and improved fuels (notably through the use of lead-free gasoline). Figure 3 shows, however, some large increases in emissions from non-road vehicles and engines – enough in the cases of several pollutants to offset the reductions from on-road vehicles. Construction vehicles and lawn mowers contribute the most to the non-road category.

It is the concentrations of harmful substances in the air that affect ecosystem and human health. They do not have a straightforward relation to transport emissions. This is because of local differences in dilution and because there are many non-transport sources of pollution. Relevant data on air quality are surprisingly few, notwithstanding much public concern. Available data suggest that concentrations of transport-related harmful substances in the air are mostly declining in the richer countries, but that levels of these compounds near traffic and within vehicles continue to be hazardous. However, the rates of decline are often regarded as insufficient to remove the risk to health and ecosystems in a timely manner. For many pollutants, reductions to below 10% of current levels are considered necessary for sustainability. At current rates of decline, many pollutants will still be at unacceptable levels in 2050.

Residents of poorer countries are mostly experiencing increasing levels of these pollutants. In the industrializing countries of the world, notably China, transport is replacing industry as the major source of pollution in urban areas (Gilbert, 1997).

Other effects of concern are noise – a more frequent source of complaint about nearby traffic than air pollution – and the occupation and contamination of land and water. Traffic noise is often sufficiently loud to impair human health through sleep disturbance and in other ways; some 70% of urban and suburban residents in Europe and North America feel they are adversely affected in this way.

The use of land for transport has many serious consequences. Transport infrastructure sterilizes land that could be used for other purposes. It disrupts natural drainage patterns and presents barriers to species movement that can reduce local biodiversity. Transport activities, particularly refueling facilities, are associated with contamination of land and water. A special concern is contamination of water systems by fuels, lubricating oil, and volatile organic compounds (VOC) from water-borne transport, whether ocean freight carriers or small pleasure craft.

THE NATURE OF SUSTAINABLE TRANSPORTATION

The World Bank has proposed a comprehensive approach to sustainable transportation that speaks to three kinds of sustainability. Economic and financial sustainability is put first. It is a matter of requiring that resources be used efficiently and that assets be maintained properly. For the Bank, environmental and ecological sustainability is a matter of ensuring that the external effects of transport be taken into account fully when public or private decisions are made that determine future development. As well, the World Bank speaks to social sustainability, which requires that the benefits of improved transport reach all sections of the community.

Canada's Centre for Sustainable Transportation has defined a sustainable transportation system as one that:

- allows the basic access needs of individuals and societies to be met safely and in a manner consistent with human and ecosystem health, and with equity within and between generations;
- is affordable, operates efficiently, offers choice of transport mode, and supports a vibrant economy; and
- limits emissions and waste within the planet's ability to absorb them, minimizes consumption of non-renewable resources, reuses and recycles its components, and minimizes the use of land and the production of noise (Centre for Sustainable Transportation, 1998).

The OECD has defined environmentally sustainable transport (EST) as follows:

> Transportation that does not endanger public health or ecosystems and meets mobility needs consistent with (a) the use of renewable resources at below their rates of regeneration and (b) use of non-renewable resources at below the rates of development of renewable substitutes.

This definition has been elaborated in terms of six quantitative criteria whose attainment would involve major reductions in the adverse impacts of transport. The criteria values are rooted in scientific assessments and other formulations as to what is required for sustainability from OECD countries. For example, the CO_2 criterion speaks to an 80% reduction from 1990 levels in emissions of CO_2 from transport in OECD countries (OECD, 1999). This is derived from what is required to stabilize atmospheric CO_2 levels, which appears to be a worldwide reduction in the order of 50% (Intergovernmental Panel on Climate Change, 1996). The scientific assessment is then augmented by consideration of the need to provide some room for growth in transport activity in poorer countries (United Nations Environment Programme, 1999).

The OECD work has involved articulation of three visions or scenarios of EST for 2030 as well as a business-as-usual projection. In one of the three scenarios, the EST criteria are met through technological change alone. In another, the criteria are met through activity change alone. In the third scenario, the criteria are met through changes in both transport technology and transport activity.

The attainment year of 2030 and the number of criteria – and to a lesser extent their character – are somewhat arbitrary. Both shorter and longer timeframes incur difficulties, the former because the pace of change itself could be overly disruptive, the latter because what has to be attained could become more unreachable as time passes. Criteria concerning several more adverse impacts could be applied. However, the OECD work suggests that the number of criteria could be reduced to three – concerning carbon dioxide, noise, and the use of land – because other criteria are usually met if the CO_2 criterion is met.

HOW TRANSPORTATION MIGHT BECOME AND REMAIN SUSTAINABLE

A central consideration in planning for and moving towards the attainment of EST concerns the relative effort to be placed on technological factors and on other factors such as shifting transport modes and reducing transport activity. Figure 4 illustrates the relative effort assigned to the different factors in the OECDs EST project and in a parallel project concerning countries in Central and Eastern Europe (Central European Initiative, 1999).

The broad strategies developed during the two studies illustrated in Figure 4 suggest that about 45% of the movement towards EST – whether in the richer OECD countries or the poorer Central European Initiative (CEI) countries – will have to be achieved through improvements in the technology of vehicles, fuels, and infrastructure. This implies reductions in the full life cycle fuel

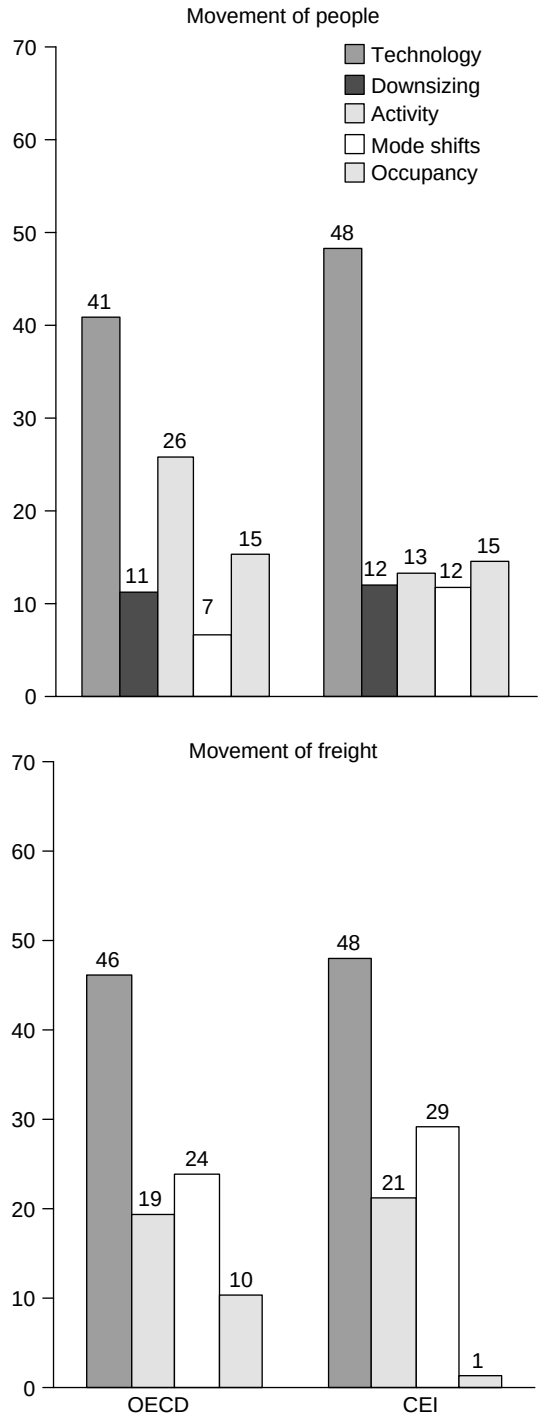


Figure 4 Balance of effort required for the attainment of sustainable movement of people and freight in OECD and CEI countries. The bar graphs are based on information in OECD (1999) and Central European Initiative (1999). Bars and numbers above bars show estimates of the percentage shares of the effort involving different means required to achieve EST. The various means are described in the text

intensity of vehicles in the order of 50%, which is an attainable objective. Private automobiles used in OECD countries – responsible for more than 80% of person-kilometers – now require on average about 10 l of gasoline for each 100 km driven. Hybrid gasoline–electric vehicles already on sale in Japan and North America require less than half as much fuel, as do diesel-fuelled vehicles on sale in Germany. The studies suggest further that a stronger emphasis on gains through technology would likely require widespread use of hydrogen as a fuel for combustion engines or for fuel-cells. Even if the in-car technology becomes adequate and affordable by 2030, the supporting infrastructure will not likely be in place. Achieving comparable gains for aviation will be much more difficult.

The other means of attainment of EST all involve changes in behavior that in many cases will be highly unacceptable at first. Figure 4 suggests that about a quarter of the result will come from reductions in the overall amounts of motorized transport activity. Substitution by walking and bicycling will be a factor, as will reductions in trip length facilitated by increased heterogeneity and density of settlements. Reductions in automobile ownership, discussed further below, may have a special role in reducing transport activity.

Downsizing of vehicles will play a significant role, particularly for automobiles. Recent rates of weight reduction per unit of engine power will likely be sustained and engine sizes will drop. Mode shifts, chie y from automobiles to various forms of public transport, will play a relatively minor role in changing the movement of people. In some areas, nevertheless, quite large expansions of public transport will be required, for environmental reasons and to prevent congestion associated with denser settlement patterns. Mode shifts, from road to rail or water, could play a much larger role in the movement of freight.

Increased occupancy has a large role to play in the movement of people, whether for private automobiles or for public transport vehicles. Indeed, public transport occupancy levels have become so low in the US, journeys by car are now on average more environmentally sound. Typically a bus uses six times as much fuel per kilometer as a car, but currently carries only five times as many people – while being capable of carrying 15–20 times as many people as a car (Centre for Sustainable Transportation, 1998). The comparable factor for freight movement – increased loading – was considered to require less emphasis.

When buses are well occupied, they are clearly environmentally superior to cars. Their better performance in terms of emissions per passenger-kilometer when well occupied is not the only source of their advantage. Perhaps the most important feature of good public transport is that it makes ownership of a car unnecessary or less necessary. As discussed below, car ownership is a major cause of car use. The mere availability of a bus may in uence whether or not a car is bought. If the proximity of the bus route prevents

some car ownership, it may be making a strong although not obvious contribution towards sustainable transportation.

CAUSES OF THE TRENDS IN TRANSPORT ACTIVITY

If the strategies developed in the course of the OECD and CEI studies were followed, there would have to be substantial emphasis on behavioral change as well as on technological improvement. The necessary changes in behavior will be achieved more readily if the causes of patterns of transport activity are better understood. These causes are numerous, ranging from circumstance (people who live in remote suburbs often have no choice but to drive often and alone) to distaste for public transport (buses are loser-cruisers).

A prominent but overlooked cause of automobile use is automobile ownership. Figure 5 shows how closely ownership and use have been linked in Canada during the last few decades. Similar data are available for several other countries, notably the US. The figure suggests a very tight link between ownership and use, and a possible causal relationship whereby purchase of a car inevitably leads to what happens in Canada to be an average of some 16 500 km driven annually. It follows that curbing ownership may be among the most effective ways of curbing use. Certainly in places where ownership is limited by government action (e.g., Singapore, Tokyo, and countries with command economies) automobile use is consequently at a relatively low level.

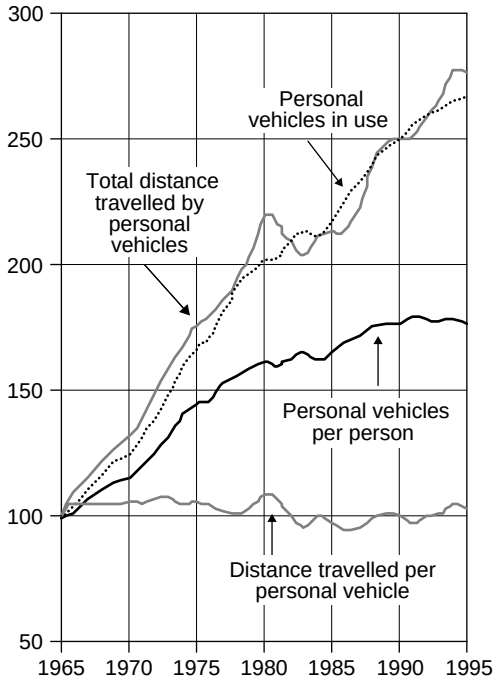


Figure 5 Distances traveled and ownership of personal vehicles in Canada, 1965–1995 (normalized so that 1965 = 100 for each variable). From Centre for Sustainable Transportation (1999). Personal vehicles include light trucks, vans, and sport-utility vehicles used as personal automobiles as well as regular automobiles

Acknowledgement of the tight link between the ownership and use begs questions as to the causes of ownership, but also broadens the scope for investigation of causes. As important, it broadens the selection of instruments for reducing use to include those for reducing ownership.

Curbs on ownership are often considered to be more unacceptable – and undemocratic – than curbs on use, although logically there is no case to be made for the democratic superiority of one over the other. Attempts to limit ownership need not be coercive. Land-use planning policies can ensure equal advantages for non-ownership by mandating higher densities, a mix of activities, much public transport, and services supporting non-ownership – in short, all of the attributes of successful downtown areas that have many residents (Gilbert, 1999).

INSTRUMENTS FOR CHANGE

There are basically five kinds of policy and other instruments available to governments for achieving the changes in transport technology and transport activity required for sustainable transportation. Two key types of instrument are regulatory instruments and fiscal instruments. An example of a regulatory instrument is the setting of mandatory performance requirements for new automobiles. Charges for road use and incentives to scrap polluting vehicles are fiscal instruments. As well, governments can invest funds, for example in bicycle lanes and public transport. They can engage in education and other persuasive activities. Finally, they can change administrative or governance arrangements so that effective bodies with appropriate objectives have the means and the authority to ensure progress towards sustainable transport.

Economists prefer fiscal instruments. Indeed, some economists argue that sustainability is a matter of getting the prices right. (The World Bank's view, noted above, comes close to this position.) One of many problems with this argument is that transport activity seems to be relatively inelastic, at least in the short term: large increases in prices produce only small changes in behavior. Given that sustainability may require relatively large changes in transport activity, price increases may be insufficient.

Price increases appear to be more effective in the longer term, i.e., raising gasoline prices today will have a more noticeable effect five or ten years from now. This suggests that fuel price increases cause people to buy more fuel-efficient cars rather than to drive less.

Regulatory instruments have a stronger record of effectiveness than fiscal instruments; nevertheless, they are less favored by economists and increasingly less favored by

policy-makers. The best-known regulatory instrument may be the fuel-efficiency requirements for new automobiles introduced by the US government in the 1970s (the so-called Corporate Average Fleet Efficiency or CAFE standards). The precise extent of their effect is debated, but these seems general acknowledgement of their responsibility for at least part of the major improvements in fuel efficiency recorded in North America in the 1970s and 1980s, and for maintaining the gains in the 1990s while fuel prices fell steeply in real terms. Meanwhile, without CAFE-like instruments, the fuel efficiency of European cars has worsened, to the point that the efficiencies for North America and European cars are now similar.

More fuel is used in North America because there are many more personal vehicles per person, because vehicles are driven farther on average, and because well over one-third of the personal vehicles on the road are the larger-engined light trucks, vans or sport-utility vehicles, which have been subject to less strict CAFE standards. These vehicles may soon have to meet the CAFE standards of regular automobiles.

Instruments are generally more effective if – such as the CAFE standards – they are tied to clear performance goals with regular monitoring and penalties for non-compliance.

Perhaps the most promising instrument in terms of moving towards sustainable transportation would be a sophisticated form of fuel rationing that involves equal distribution to each citizen of tradable entitlements to emit CO_2 from transport. The total of all distributed entitlements in a year would be equal to what the country in question had agreed would be its total CO_2 emissions from transport for the year. For example, each person could be entitled to emit a tonne of CO_2 each year from transport. Persons who required more, for business or personal use, could purchase entitlements from persons who required less, through a readily accessible entitlements market. Heavy users of freight transport would be among those who would ensure an attractive market price. Such a hybrid fiscal-regulatory instrument would be the ultimate expression of the principle of tying instruments to sustainability goals.

The essential feature of any rationing scheme is that proper enforcement ensures a cap on what is rationed. Marketable entitlements provide for exibility while maintaining the integrity of the cap. Indeed, market-based rationing of one kind or another seems inevitable if transportation is to be made more sustainable, either in the form of a separate scheme for transport or as part of a broader-based scheme covering emissions of all GHGs from all sources. Rationing of automobile ownership is already in place in Singapore, where there are monthly auctions for entitlements to purchase new vehicles. Rationing of automobile use has been proposed for Mexico City, involving marketable permits to drive on a particular day (Goddard, 1997) (*see* Tradable Permits for Greenhouse Gases, Volume 4).

SUSTAINABLE TRANSPORTATION IN DEVELOPING COUNTRIES

Projections for China and for the rest of the world in Figure 2 suggest that in 2020 developing countries will have more than 80% of the world's population but be responsible for less than half of the energy used for transport, and thus contribute less than half of the overall environmental impacts of transportation. Issues of transport sustainability will thus continue to be largely those of OECD members and other richer countries. Nevertheless, already in most of the major urban areas of developing countries, transport is a major environmental problem and often a cause of social and economic distress.

A major part of the current problem in the urban areas of developing countries is the widespread use of highly polluting vehicles and poor quality fuels. Improvements in vehicles and fuels would still leave the basic challenge, which is unrelenting growth in transport activity. Moreover, reliance on technology is an expensive option for a developing country and its beneficial effects are evident only after several years. Focusing on reducing use can be more cost-effective, especially in the short term.

This is the logic behind the best known of the measures taken in a developing country to address the environmental problems posed by transport: Mexico City's *Hoy no circula* (HNC) system, which requires that private cars be off the road for one working day each week. The Mexico City region is the largest urban region in the developing countries, and among the areas of the world with the worst air quality. Goddard (1997) has observed:

> for many years Mexico City residents have complained about the unlivability of the city. These complaints have centered principally on the high levels of air pollution, congestion and noise, problems cause principally by private automobiles. ••• One recent estimate for Mexico City estimates that the total economic losses associated with the unrestricted use of the private car exceed $7.1 billion annually. ••• Of the 29.5 million trips made daily, 39% are in private automobiles.

Other evidence suggests that these private automobiles are owned almost exclusively by the wealthiest 20% of households, and that the burden of the resulting pollution falls primarily on the remaining 80%.

The HNC system may have had the perverse effect of increasing pollution levels as numerous car owning households purchased an additional car – often an inexpensive, old, polluting model – just to have a vehicle available every day. One already noted solution is to transform the HNC system into one involving marketable permits.

In Beijing, China's most populous city, average household incomes and levels of car ownership are lower than in Mexico City, but transport has nevertheless become the major cause of poor air quality in the urban region. A policy emphasis on encouraging bicycle use – the number of

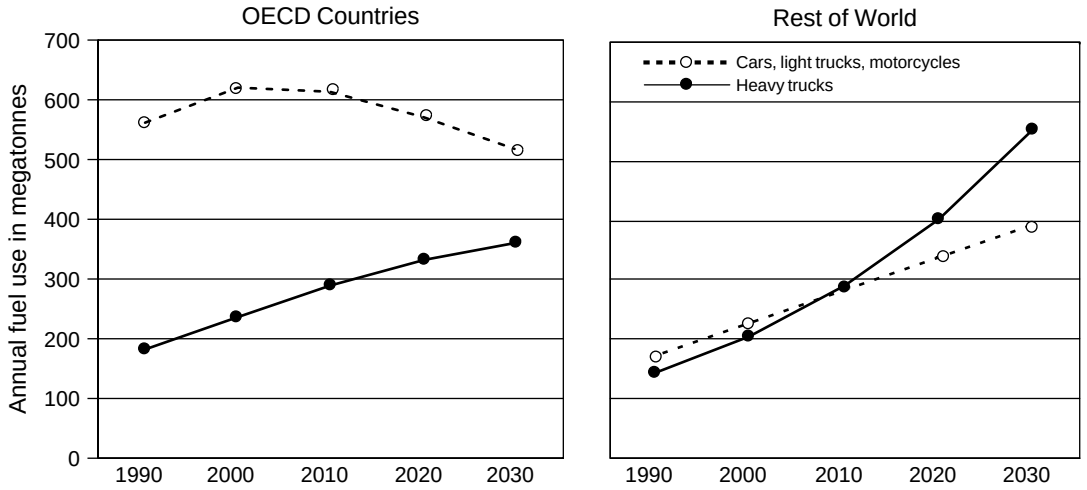


Figure 6 Projected fuel use for transport in OECD and other countries, 1990–2030

bicycles in use in Beijing tripled in the 1980s – appears to be being replaced by encouragement of private automobile ownership. Projections of the resulting environmental impacts are alarming (Mackett, 1999).

Also alarming have been business-as-usual projections for India that speak to a 5.5-fold increase in energy consumption for transport between 1990 and 2020. Proposals have been made to halve the rate of increase in order to reduce local and global pollution and avoid importation of petroleum-based fuels (Ramanathan and Parikh, 1999). The main features of the proposals are discouragement of private automobile use and truck use, expansion of public transport and rail freight capacity, and discouragement of air transport. The proponents argued that the resulting extent of avoidance of emissions of GHGs would make the Indian rail system a strong contender for investment under Joint Implementation programs (*see* Joint Implementation, Volume 4). Such programs allow richer countries to meet part of their commitments to reduce emissions of GHGs at lower cost by contributing to appropriate developments in poorer countries.

The transport and environment challenges of India, Mexico City, and Beijing are variants of those found throughout the developing world, most especially but not only in large urban areas. It might seem that rapid growth in automobile use is the major problem. Figure 6 suggests that in the poorer countries of the world, heavy truck traffic may already have as much environmental impact as private cars, and is set to have more.

Many of the solutions proposed for both richer and poorer countries call for more public transport. This is already used more in developing countries but almost everywhere has the potential for further expansion as an alternative to further growth in private automobile use. However, public transport can ourish alongside the private automobile only if it is given priority, which usually means imposing more onerous restrictions on automobile ownership or use than most governments – in rich or poor countries – are ready to impose.

CONCLUDING COMMENTS

Major increases in motorized transport activity were among the defining characteristics of human existence in the 20th century. Accompanying these increases were profound adverse environmental impacts, largely a consequence of the dependence of transport on the combustion of fossil fuels. During the 21st century, technological improvements will provide part of the response to the growing disadvantages posed by trends in transport activity. The major part, however, will likely have to come from changes in transport behavior.

Developing countries pose particular challenges, especially if their projected growth in transport activity continues beyond the next few decades. Preventing poorer countries from eventually achieving the amount of transport activity found in what are now the richer countries would be inequitable and perhaps unfeasible. On the other hand, dependence of all humanity on motorized transport to the present extent of rich countries would be unsustainable.

The only feasible solution may be massive restraint on transport activity in rich countries. This would be part of a new development strategy directed towards helping people in the South (developing countries) become as free from want as people in the North and having people in the North use as few non-renewable resources, particularly fossil fuels, as people in the South. Accordingly, a goal of sustainable development should be as much to make the North like the South, as it is to make the South like the North. Indeed, protection of the global environment may

require that in terms of the overall amount of transport activity, the North be made more like the South than *vice versa*.

See also: Transport: Global Freight and Passenger Flows, Volume 3; Transport Infrastructure, Volume 3.

REFERENCES

Central European Initiative (1999) *Towards Sustainable Transport in the CEI Countries*. United Nations Environment Program (UNEP); Organization for Economic Cooperation and Development (OECD); Federal Ministry for the Environment, Youth and Family Affairs, Austria.

Centre for Sustainable Transportation (1998) *Sustainable Transportation Monitor No. 1 and 2*, Centre for Sustainable Transportation, Toronto.

Centre for Sustainable Transportation (1999) *Sustainable Transportation Monitor No. 1 and 2*, Centre for Sustainable Transportation, Toronto.

Gilbert, R (1997) *Reducing Urban Air Pollution in APEC Economies: Transportation and the Roles of Local Government*, Asia Pacific Foundation of Canada, Vancouver, Canada.

Gilbert, R (1999) Reducing Automobile use in Urban Areas by Reducing Automobile Ownership: the EANO Principle, in *Proceedings of the Onzièmes Entretiens du Centre Jacques Cartier, December 1998*, Centre Jacques Cartier, Lyon, France.

Goddard, H C (1997) Restrictions on Vehicle use in Mexico City, *Int. J. Environ. Pollut.*, 7(3), 357–374.

Greene, D L and Wegener, M (1997) Sustainable Transport. *J. Transp. Geogr.*, 5(3), 177–190.

IEA (1998) *World Energy Outlook*, International Energy Agency, Paris.

Intergovernmental Panel on Climate Change (1996) *Climate Change 1995: The Science of Climate Change*, eds J Houghton, L G Meira Filho, B A Callander, B A Callander, N Harris, A Kattenberg, and K Maskell, Cambridge University Press, Cambridge.

Intergovernmental Panel on Climate Change (1999) Summary for policymakers, in eds J E Penner, D H Lister, D J Griggs, D J Dokken, and M McFarland, *Aviation and the Global Atmosphere*, Cambridge University Press, Cambridge, 1–12.

Mackett, R (1999) Towards the Solution of Urban Transport Problems in China, *J. Environ. Sci.*, 11(3), 334–338.

Mitchell, B R (1992) *International Historical Statistics: Europe, 1750- 1988*, Macmillan, London.

Mitchell, B R (1993) *International Historical Statistics: The Americas, 1750- 1988*, Macmillan, London.

Mitchell, B R (1995) *International Historical Statistics: Africa, Asia & Oceania, 1750- 1988*, Macmillan, London.

OECD (1999) *Scenarios for Environmentally Sustainable Transport*, Organization for Economic Cooperation and Development, Paris.

Ramanathan, R and Parikh, J K (1999) Transport Sector in India: an Analysis in the Context of Sustainable Development, *Transp. Policy*, 6, 35–45.

United Nations Environment Programme (1999) *Global Environment Outlook 2000*, Earthscan Publications, London.

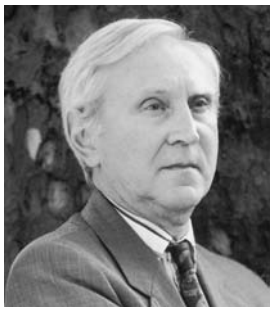
US EPA (1998) *National Air Pollutant Emissions Trends Update, 1970- 1997*, United States Environmental Protection Agency, Washington, DC.

FURTHER READING

OECD (1995) *Motor Vehicle Pollution: Reduction Strategies Beyond 2010*, Organization for Economic Cooperation and Development, Paris.

World Bank (1996) *Sustainable Transport: Priorities for Policy Reform*, The World Bank, Washington, DC.

T



Tickell, Crispin

(1930-)

Sir Crispin Tickell, the British diplomat, was educated at Westminster School and Christ Church Oxford. He received a BA with 1st class honors in modern history 1952. He was a fellow at the Center for International Affairs, Harvard University, 1975–1976, and a visiting fellow of all Souls College, Oxford 1981.

Tickell comes from the diplomatic community, probably the reason for his outstanding diplomatic skills that prepared him for his leadership role. He joined the British Diplomatic Service in 1954 and served at the Foreign Office, and British Embassies at The Hague, Mexico and Paris (1954–1970). At the Foreign and Commonwealth Office, he was private secretary to successive Chancellors of the Duchy of Lancaster as ministers responsible for the negotiations for British entry into the European Community 1970–1972 and then head of the Western Organizations Department (1972–1975) responsible for North Atlantic Treaty Organization (NATO) matters, European security problems and political military affairs generally.

Tickell is probably one of a very limited number of those active in the field of the environment to chair or lead an enormous number of environmental and conservation institutions. Since 1990 he has been, president of the Marine Biological Association, chairman of the Climate Institute, Washington DC, director of the Green College Center for Environmental Policy and Understanding, chairman of the Advisory Committee on the Darwin Initiative for the Survival of Species, convenor of the UK Government Panel on Sustainable Development, president of the Earth Center, chancellor of the University of Kent at Canterbury, president of the UK National Society for Clean Air and Environmental Protection, and president of the Gaia Society. Recently he has also been president of the Royal Geographical Society (1990–1993), chairman of the International Institute for Environment and Development (1990–1994), warden of Green College, Oxford (1990–1997) and chairman of Earthwatch Europe (1991–1997).

Tickell has received a huge number of honors including eleven honorary decorate degrees from academies and universities in Mexico and UK and four honorary fellowships.

Tickell has been a very active environmental advocate particularly since 1990, meaningfully addressing global environmental and sustainable development issues. He has authored many books and reports and has appeared on several radio and television programmes. Among his major contributions is his book on *Climate Change and World Affairs* (Tickell, 1977).

REFERENCE

Tickell, C (1977) *Climate Change and World Affairs*, Pergamon Press 1978, revised edition Harvard University and University Press of America, 1986.

MOSTAFA K TOLBA *Egypt*

Toepfer, Klaus

(1938-)

Klaus Toepfer is the current Executive Director of the United Nations Environment Programme (UNEP). Toepfer

graduated from the University of Munster Germany in 1964, and obtained his doctorate in economics in 1968. Toepfer began his academic career at the same institution in the economics department, with an emphasis on regional planning and development. He has also lectured at universities in Hagen, Bielefeld, Speyer, Hanover, Mainz, and most recently, as an Honorary Professor at Tongji University in Shanghai, China.

Toepfer began his 4-year term as the Executive Director of UNEP on 1 February 1998. He is well known in United Nations (UN) circles from his tenure as Chairman of the United Nations Commission on Sustainable Development, from May 1994 to May 1995. In addition to his duties at UNEP, Toepfer is Director-General of the United Nations Office at Nairobi.

Toepfer is a member of the Senior Management Group of the UN. He also chaired a Task Force, set up by the Secretary-General, to review existing structures and arrangements through which environment and environment-related activities are carried out within the UN, and to prepare proposals on reforming and strengthening UN activities in the environment and human settlement areas.

The professional life of Toepfer has combined extensive work in government and politics with a concurrent career in academia. His government service began in 1971, as the Head of the Department for Planning and Information in the State Chancellery of the Saarland. Subsequently, from 1978 to 1985, he served as the State Secretary and in 1985–1987 as Minister for Social Affairs, Health and Environment of Rhineland-Palatinate. He was elected to the German Bundestag in December 1990.

Toepfer held the post of Federal Minister for the Environment, Nature Conservation and Nuclear Safety from May 1987 to November 1994. From November 1994 until joining UNEP, he was the Federal Minister for Regional Planning, Building and Urban Development and Coordinator for the Transfer of the Bundestag (Parliament) and Federal Government to Berlin.

Toepfer has also held various positions in a national political party, (the Christian Democratic Union (CDU)) at local, regional and national levels. He has been a member of the CDU's presiding committee since October 1992.

Toepfer is the author of numerous articles and publications on economics, development, the environment, human settlements and international affairs. In 1986 he was awarded the Cross of the Order of Merit of the Federal Republic of Germany; and, in 1989, the Commander's Cross of the Order of Merit, for exceptional service to his country.

Born on 29 July 1938 in Waldenburg, Silesia, Klaus Toepfer is married with three children.

MOSTAFA K TOLBA *Egypt*

Tolba, Mostafa

(1922-)

Mostafa Tolba was born in the town of Zifta, north of Cairo in 1922. He graduated with first class honors from Cairo University in 1943, and obtained his PhD from the Imperial College, in 1948. He returned to Cairo to eventually become Professor in the Faculty of Science, Cairo University, where he established his own school in microbiology and is currently Emeritus Professor. He was also Professor in Bagdad University, 1954–1959.

After this distinguished academic career, Dr Tolba joined the Egyptian civil service as Undersecretary of State for Higher Education and Minister of Youth, and on the international scene as an alternate member of the United Nations Educational, Scientific and Cultural Organization (UNESCO's) Executive Board. He became the first president of the newly established Academy for Scientific Research and Technology in 1971.

In 1972, he led Egypt's delegation to the Stockholm conference on the Human Environment, thus starting a lifetime commitment to environmental issues. In the preparatory meetings leading to the conference, as well as during the conference itself, he was among the first representatives of developing countries to realize and promote the view that there are no irreconcilable con icts between the environment and development. He argued that concern for the environment should extend to the social as well as the physical aspects. At the time, quite a number of developing countries saw environmental issues merely as problems of pollution that were the concern of developed countries. It was no wonder that he was nominated, immediately after Stockholm, as Deputy Executive Director of the newly established United Nations Environment Program (UNEP).

Within 2 years, he became the Executive Director – a post he held until retiring at the end of 1992.

Under his leadership, UNEP became the core organization within the United Nations (UN) family acting as the catalyst spurring governments, business, academia, international governmental organizations (IGOs) and non-government organizations (NGOs) to meaningful action. UNEP, one of the small members in the UN family, could leverage on the average four times its modest budget to carry out more than a thousand projects.

Following up on his position during Stockholm, he diligently promoted his philosophy of development without destruction . Its implications are clearly re ected in his speeches, books and in UNEP's programs, in many fields, and at many levels.

It is difficult to single out Tolba's most significant qualities and contributions. He is by nature an optimist who pursues his vision and objectives with remarkable diligence combined with communication skills. His obvious mastery of the subject, sincerity, and dedication to the causes he is pursuing, command the respect of those who deal with him, be they heads of states or ordinary people. As a negotiator he is respected by all parties who usually hold con icting and, what seem to many, irreconcilable views. He patiently leads them to agree to compromises that many have despaired of reaching.

Combined with this, is his uncanny vision of emerging environmental problems. As early as the mid-1970s, he concentrated on the issue of stratospheric ozone layer depletion as meriting careful monitoring on the scientific front. At the right moment, he could manage, whether in Vienna, Montreal and later on in London, to formulate the prototype model for dealing with global environmental issues and the effective mechanisms for the transfer of technology and funds to developing countries, as and when necessary. Other examples of his skills as a diplomatic negotiator on the international arena are the Basel, and the Biodiversity Conventions, which were concluded after tortuous and tricky negotiations.

Dr Tolba has received many awards and prizes, both from academic institutions, governments and NGOs in many countries. These include honorary doctorates, awards, medals, and high decorations. He cherishes in particular the DSc degrees from Moscow State University, and the University of Guadalajara; DLL, Williams College; fellowship of Imperial College, London; the Sasakawa Prize; the Only One Earth Award of the René Dubos Center; the Distinguished International Service Award of the Regents of the University of Minnesota; and the First Order Decoration of the Arab Republic of Egypt.

Back in Egypt, he established in 1994 the International Center for Environment and Development (ICED), a not for profit organization financing environmental projects in the less developed countries through an endowment fund administered by an independent board of trustees.

OSAMA A EL-KHOLY *Bahrain*

Tradable Permits for Greenhouse Gases

Christiaan Vrolijk and Michael Grubb
The Royal Institute of International Affairs, London, UK

In December 1997 over 180 countries agreed with the Kyoto Protocol to take action against the threat of global climate change. This Protocol sets the aggregate target for industrialized countries to reduce their emissions to 5% below the 1990 base year level by the commitment period of 2008- 2012. However emissions of greenhouse gases (GHGs) are projected to grow substantially, over 30% for some Organization for Economic Cooperation and Development (OECD) countries. To ease the burden of the Kyoto Protocol, several measures have been included, such as differentiated targets across the industrialized countries, including Eastern Europe and Russia, and spreading the effort over a basket of GHGs and sinks (such as forests). The Protocol further introduced the so-called " Kyoto Mechanisms" , international emissions trading, joint implementation and the clean development mechanism (CDM). These mechanisms allow international transfers of emission allowances and reduction credits between the Parties to the Protocol. Even though this Protocol is the most likely regime for trading GHGs, a trading system can also be used in other circumstances, such as domestic regulation.

Tradable permits can reduce substantially the costs of emission abatement (and compliance with the Kyoto Protocol). Experience with various schemes, such as the US Acid Rain Program, has shown cost reductions of over 75%. Some modeling work on emissions trading under the Kyoto Protocol projects even greater advantages. Because of emissions trading, emission reductions will take place where they cost least, anywhere in the world and in any of the GHGs targeted. However, some problems do

exist that could undermine the legitimacy of emissions trading.

The climate is changing. Ten years ago, that was a very controversial claim. Now it is widely accepted that changes are being observed, that human emissions of GHGs are a possible cause, and that such emissions may lead to far greater climatic changes during the coming decades. In 1990, governments began to negotiate what to do about it, and in the mid-1990s they started to debate more substantive commitments.

*The Framework Convention on Climate Change (FCCC) was negotiated at the 'Earth Summit' in Rio de Janeiro in 1992 and entered into force in 1994. Under the Convention the Parties committed themselves to stabilize greenhouse gas concentrations 'at a level that would prevent dangerous anthropogenic interference with the climate system' (FCCC, Article 2). The headline commitment for the industrialized countries (the Annex I Parties) was to return greenhouse gas emissions to 1990 levels, and to show a reversal in the trend of growing emissions before the year 2000. The Conference of Parties (COP) meets annually as the supreme body of the Convention, dealing with various issues related to it (*see *Conference of Parties, Volume 4). At the first COP in 1995 in Berlin, the Parties agreed that these commitments were inadequate and they began 'a process to enable it to take appropriate action for the period beyond 2000, including the strengthening of the commitments of Annex I Parties' (Decision 1/CP.1, 7 April 1995, the 'Berlin Mandate'). Specifically it was decided to 'set quantified emission limitation and reduction objectives within specified timeframes'. The negotiating process based upon this 'Berlin Mandate' culminated at the Third Conference of Parties (COP-3) in Kyoto, Japan, December 1997.*

The Kyoto Protocol that was negotiated at the COP-3 sets out renewed, legally binding emission reduction commitments for the Parties listed in Annex B (Annex B of the Kyoto Protocol is virtually equivalent to Annex I of the Convention, the industrialized countries, including Eastern Europe and Russia.) of the Protocol in the period of 2008- 2012. The overall commitments sum up to a 5.2% emission reduction from 1990 for a basket of GHGs, including carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O), some industrial gases (hydrofluorocarbons (HFC), perfluorocarbons (PFC), sulfur hexafluoride (SF_6)) and emissions and removals from land-use change and forestry (LUCF or 'sinks').

TRADING GREENHOUSE GASES

The main greenhouse gas, carbon dioxide (CO_2), is released when burning fossil fuels, which supply most of the industrialized countries' energy needs. A Protocol to limit these emissions, and thus the use of fossil fuels, would therefore limit the potential for economic growth. Some early studies suggested substantial losses of Gross Domestic Product (GDP) if limitations on greenhouse gas emissions were imposed. However, recent studies all show losses of under 1%, even for domestic implementation without any flexibility, equivalent to forgoing a few months of GDP growth and [it] is hardly discernible compared to the projected overall growth and uncertainties therein' (Grubb *et al.*, 1999). As a result many have argued for flexible commitments as the most cost-effective way of limiting emissions.

In the late 1980s tradable permits were proposed as an efficient solution, which was both practical and equitable (Grubb, 1989). Increased flexibility with emissions trading made it more plausible for governments to agree to greater emissions control, at lower cost. International emissions trading was seen as the most efficient and direct route to the international flexibility sought by the US and some other countries – and was viewed with suspicion by most of the rest of the world.

The principal idea behind emissions trading is simple; two entities exchange their emission quota if that is in their common economic interest. One Party reduces its emissions further than necessary and sells its spare quota to another Party that has not been able to reduce enough, or would not be able to do so for the present market price.

In the late 1980s and the beginning of the 1990s, interest in emissions trading for controlling GHGs had grown, particularly in the US, for several reasons. First, the US experience with its tradable permit system to control sulfur dioxide (and ozone-depleting substances) was widely seen to have been a success: emissions were reduced at much lower cost than originally predicted, because of the flexibility of the permit system and the incentives it gives for innovation. Second, emissions-trading is more attractive to industry than other economic instruments (such as carbon taxes), because the system creates a tradable asset: the permit or allowance. Even if a company has to buy permits to cover all its emissions, it still acquires the value of those permits as an asset, which can in principle be sold in the future if emissions can be reduced. This contrasts with a tax, which extracts revenue from the firm without adding any compensating value. In addition, governments may give out some or all of the permits initially, reducing the cost to industry. In many countries, taxes have proved extremely controversial and difficult to implement, primarily because of industrial opposition, so that other market-based approaches are attracting growing interest.

Emissions trading in principle separates who has the reduction commitments and bears the costs, from where and by whom the emission abatement actually take place. Where emissions have a direct regional impact, the split between the reductions and the location is not trivial and free trading of emissions is impossible, because it could lead to local pollution 'hot spots'. Although there will

be large regional differences in the effects of climate change – some regions will get warmer, others colder, wetter or drier – this does not depend on regional differences in the concentrations of the gases, the problem is global. Unlike many other cases, such as acid rain and smog, local climatic changes cannot be stopped by local measures: the location of emissions is unimportant. From a climate point of view, GHG emissions are homogenous: a ton emitted in New York is the same as a ton emitted in New Delhi. GHGs, therefore, are ideally suited for emissions trading.

The FCCC indicates that Parties can act on their commitments jointly with other Parties. The Convention itself did not establish a mechanism, then controversial, for joint implementation, or any other form of international transfers of emissions. However, an activities-implemented-jointly (AIJ) pilot phase was set up in Berlin, as a first step towards such a mechanism. The Kyoto Protocol, strengthening the commitments of the Convention, established several forms of international transfers of emission allowances and credits.

The COP decided that no credits shall accrue to any Party as a result of greenhouse gas emissions reduced or sequestered during the pilot phase from AIJ (Decision 5/CP.1, 7 April 1995, In this decision the COP recognized that AIJ will not be seen as fulfillment of current commitments), it can therefore not be considered as emissions trading in order to meet commitments. However, the experience with AIJ, gathered before Kyoto, showed the vast potential for cheap emission reductions, with projects mainly situated in the economies in transition (Central and Eastern Europe and the former Soviet Union.) and some developing countries. This pilot phase is now extended to gain further experience.

Other trades' have also taken place between countries and companies. The Nordic countries set up a shadow market in 1996 between the four governments, to meet their commitments under the Framework Convention. Some companies have traded emission reductions or invested in sequestration to meet voluntary emission targets.

However, the most important prospect for a market for greenhouse gas emissions was established with the Kyoto Protocol. The Protocol identified three different mechanisms for international transfers, discussed in the next section. It also opens the opportunity for domestic industrial emissions trading schemes.

MERITS AND PROBLEMS OF TRADING

Some of the merits of international exibility through emissions trading and other mechanisms of international transfers have been discussed previously. These arguments were used throughout the Kyoto negotiations by many of the non-European Union (EU) countries of Annex I, the major proponents of the exibility mechanisms. Because it is politically much easier to take cheap actions than expensive ones, cost-effectiveness is very important for the success of the Protocol. The establishment of an emissions trading market will automatically create a price for emission allowances. A decision not to take emission reduction measures would then forego the opportunity cost of selling these allowances. This way a market gives incentives to reduce emissions. The international transfers could also meet another objective of the Convention, technology transfer, and it establishes a ow of financial resources to developing countries and economies in transition, that could be used to fund sustainable development and adaptation measures.

However, some problems also occur from international transfers, often brought to the table by non-governmental organizations (NGOs). Emissions trading and the other mechanisms of Kyoto are very innovative and their effectiveness is still uncertain. Before Kyoto, it was also argued that negotiating these mechanisms would take more time than was available, time that could better be spent on other issues. In particular the monitoring and compliance systems of the Protocol are now substantially more complicated and need more time to be negotiated and put in place. Apart from this, the large financial resource ows from international transfers, resulting from trades in allowances necessary to gain substantially from the mechanisms, will also be politically difficult to implement.

Since Kyoto, the debate has changed. The Kyoto mechanisms are seen as an integral part of the deal of the Kyoto Protocol; without this exibility, many Parties would not have been able to agree. In particular the environmental NGOs are seeking restrictions on the use of the mechanisms, to ensure that the mechanisms do not diminish the environmental gains of the Protocol and to restrict trade in the excess allowances granted to some of the Economies in Transition (sometimes called hot air'). Also the EU has submitted proposals for ceilings on the use of the international transfer mechanisms, for similar arguments as the NGOs, to encourage domestic action in all Annex B Parties, and other political reasons.

Proponents of the trading mechanisms argue that it gives incentives for innovation. Opponents, however, argue the opposite. Trading encourages those with low marginal abatement costs to reduce their emissions further, while reducing the reductions taken by the high-abatement cost polluters. Overall, trading reduces the pressure to innovate for those with high abatement costs, and increases reductions from the other sources, who may be able to do this with conventional technology, since their abatement costs are relatively low. Consequently, short-term costs are reduced, but less innovation takes places that reduces costs in the long term (Driesen, 1998).

A more philosophical argument is that trading would imply rights to emit', as opposed to obligations to reduce

emissions'. Particularly for developing countries, this is an important distinction. The Convention and Protocol clearly state common but differentiated commitments' and developed country leadership' (in particular Articles 3 (Principles) and Article 4 (Commitments) of the Convention). The international transfer mechanisms could lead to a violation of these principles, since the developed countries could buy their way out of taking action and maintain growing emissions. Violation of the leadership principle could seriously undermine the further evolution of the climate regime.

MECHANISMS FOR INTERNATIONAL TRANSFER UNDER THE KYOTO PROTOCOL

Before Kyoto, some industrialized-country Parties were pushing for exibility in the commitments. This took the form in many proposals of differentiation of emission targets between Parties and in the development of some form of international trading of emissions quota. The Kyoto Protocol now includes several forms of exibility:

1. The commitments include a basket of GHGs and sinks, not only CO_2.
2. The commitments vary between countries, according to their opportunities and willingness to reduce emissions, in theory leveling the effort for all Parties.
3. The bubbling provision (Article 4) makes it possible to share the commitments with a group of countries (the EU uses a bubble').
4. Three mechanisms are available for international transfers of emission allowances.

The Kyoto Mechanisms' for international transfers of emission allowances are *joint implementation* (JI, Article 6) (*see* Joint Implementation, Volume 4) the *clean development mechanism* (CDM, Article 12) (*see* Clean Development Mechanism, Volume 4), and international emissions trading (Article 17). The Protocol only provides a framework for the mechanisms, and referred the design of rules and guidelines to later sessions of the COP. The Buenos Aires meeting resulted in a work program for these mechanisms, with a view to taking decisions on all the mechanisms under Articles 6, 12 and 17 of the Kyoto Protocol at its sixth session [in 2000 in The Hague]' (Decision 7/CP.4, 14 November 1998).

Emissions trading is defined in Article 17 of the Kyoto Protocol: The Parties included in Annex B may participate in emissions trading for the purposes of fulfilling their commitments under Article 3 [i.e., the emission reduction targets]. Any such trading shall be supplemental to domestic actions for the purpose of meeting quantified emission limitation and reduction commitments under that Article. The Article gives no detail of how this trading would work; in Buenos Aires the COP decided on a deadline for these issues by the Sixth Conference of Parties (COP-6). Article 17 states that Parties may trade. It is therefore questionable whether a domestic emissions trading scheme or an international industrial emissions trading scheme will be governed by Article 17.

It is possible that a domestic trading scheme would fall under Article 2: Each Party ••• shall implement and/or further elaborate policies and measures in accordance with its national circumstances ••• . Intergovernmental transfers must be conducted in units of assigned amounts, i.e., measured in the basket of gases and sinks during the commitment period. Private sector trading, however, may be denominated in any subset of these gases and periods, and will probably begin with CO_2 emissions from power generation and industrial processes.

Article 6 of the Protocol on joint implementation says, any Party included in Annex I may transfer to, or acquire from, any other such Party emission reduction units resulting from projects . Some provisions are made in the Protocol, and the COP can define additional rules. Interestingly the Protocol allows Parties to authorize legal entities (e.g., industry) to participate in this project-based mechanism. The traded emission reduction units include all gases and sinks.

The CDM is defined in Article 12 of the Kyoto Protocol. It has three main purposes: to assist Parties not included in Annex I in achieving sustainable development and in contributing to the ultimate objective of the Convention, and to assist Parties included in Annex I in achieving compliance with their quantified emission limitation and reduction commitments under Article 3 . Projects under the CDM can generate certified emission reductions. The inclusion of sinks in this mechanism is still uncertain; sinks are not mentioned (as under joint implementation), but also not excluded from the CDM. As under joint implementation, legal entities can be authorized to undertake CDM projects. Article 12.10 explicitly states, in contrast to emissions trading or joint implementation, that credits can be obtained from 2000 onwards.

The Buenos Aires Plan of Action set a negotiating track for deciding on the issues of all three of the Kyoto Mechanisms by the COP-6, at the end of the year 2000 in The Hague, Netherlands.

The Fourth Conference of Parties (COP-4) in Buenos Aires also decided to continue the pilot phase of AIJ, for continued capacity building for the (project-based) Kyoto Mechanisms – especially in the Parties of Annex B with economies in transition and the developing countries that lacked experience so far (decision 6/CP.4, 14 November 1998). The review process should be finished at or before COP-6 in 2000. This would mean that the final conclusions of the pilot phase experience would be drawn up before the start of crediting of CDM projects. However, of the 95 projects in the second synthesis report (FCCC/CP/1998/2), only 27 took place in the developing countries, the other 68 were situated in the Annex I economies in transition.

Therefore, experience for developing countries (CDM-type projects) has been limited.

The Kyoto Mechanisms are perceived to be crucial for reaching the targets set by the Kyoto Protocol. Some studies suggest that most of the reductions compared to a business-as-usual emissions path would come through one of the mechanisms. According to a study by the Massachusetts Institute of Technology (MIT), and supported by other studies, emissions trading would deliver up to 25% of all reductions for limited trading, or nearly 75% with worldwide trading (Ellerman and Decaux, 1998).

Apart from the options of international transfers and a domestic trading system, taxes, technology standards and emission caps can be used. In the UK, e.g., the government has released its plans for an energy tax, despite protests. How this tax and any other measure will actually interact with a possible emissions trading market is still being discussed, but some compatibility problems may arise. Some other countries too have introduced carbon taxes, but the lessons learned are limited because the taxes are low and the main emitters are excluded. The Kyoto Mechanisms will have to compare favorably with these other measures in order for this novel, still relatively unknown instrument to be implemented.

All three of the international transfer mechanisms of the Kyoto Protocol are variations on the simple emissions trading model. Article 17 emissions trading allows transfers between Parties, governments, of Annex B. Joint implementation, Article 6, involves investments in specific projects that will reap the emission reduction credits. The CDM of Article 12 is similar to joint implementation, because it is based on investments in specific projects, but for political reasons it carries tougher conditions on sustainable development and developing-country adaptation to climatic changes. Both project-based mechanisms seem to be more credible, steering investment to climate-friendly technology, where emissions trading leaves the possibility of large trades in assigned amounts without action being taken.

THE IMPACT OF THE KYOTO MECHANISMS

A first indication of the task facing industrialized countries in implementing the Kyoto commitments may be observed simply by considering the reductions required from the levels already attained during the 1990s. Emissions data available up to 1995–1996 show that the aggregate emissions of Annex I countries has gone down by about 5% since 1990. This would mean that the Kyoto commitment would stabilize emissions roughly at this level (of the mid-1990s) fifteen years hence (Grubb and Vrolijk, 1998).

Depending on the projections used, business-as-usual emissions in the commitment period for the OECD countries could be between 20 and 30% above the Kyoto target. Emissions of all of Annex I countries, including those of the Economies in Transition, are expected to exceed the target only by about 10–20% (Ellerman and Decaux, 1998; Grubb *et al.*, 1999).

In terms of distribution, the story is radically different. In the EU, emissions have declined – albeit not primarily due to climate change policies – and the recent projections suggest that member states would just achieve the Convention's aim of holding emissions of CO_2 and other GHGs below 1990 levels in the year 2000 (in aggregate). The US, Japan and Canada are committed to bigger reductions from the higher levels they had reached in the mid-1990s, while the Economies in Transition have scope for a big increase. Figure 1 illustrates the specific changes from 1995 levels needed to reach the Kyoto targets for different groups within Annex I. This confirms that the Kyoto commitments represent a big reduction from 1995 levels for the OECD countries, and a big increase for all the Central and Eastern European Countries (CEEC), especially the ones less advanced in their economic transition, that are not currently on track for EU membership.

Quantitative Impact of the International Transfer Mechanisms

These large regional differences in commitments are based on political will, the limitation and reduction opportunities perceived by the various parties and also on development indicators. This and the large difference in projected emissions open up a large potential for trade between the Parties.

The exact regime of the mechanisms is still under debate. Under this political uncertainty, industry is not willing to invest and only a few trades have taken places so far. It is, however, industry that eventually has to make the reductions. The likely impact of emissions trading and the other mechanisms can be estimated from experience from other pollution-trading regimes, such as the Acid Rain Program of the US and modeling studies.

The estimates for the aggregate mitigation costs are still wide-ranging. A study by the MIT (Ellerman and Decaux, 1998) estimated the global costs of the Kyoto commitments to be around $120 bn, without using any exibility. Including the other GHGs of the Protocol could reduce costs by a quarter. Other studies (Grubb *et al.*, 1999) estimate reductions in GDP in the range of 0.2–2.0% by the end of the compliance period, or 0.3–1.4% neglecting the outlying studies, with an average around $115 bn, for domestic implementation. A few of these studies analyze the impact of the trading mechanisms, they indicate aggregate cost reductions from trading of between 50 and 75%. The estimates of cost reductions from domestic abatement only through international transfers between

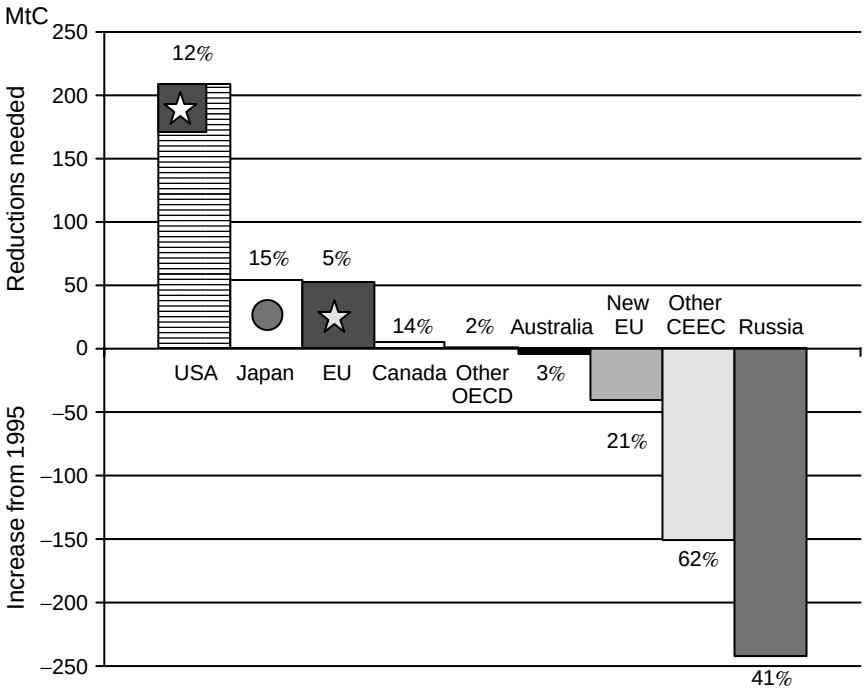


Figure 1 Gap between Kyoto commitments and 1995 emission levels in principal Annex I groups. The bars show the difference between reported emissions in 1995 and the Kyoto commitments in absolute terms; negative data indicate scope for increase from 1995 levels. The percentage difference between reported emissions and the target is also indicated next to the bars. The group the ''Other OECD'' consists of Iceland, New Zealand, Norway, and Switzerland. ''New EU'' is the countries set to join the EU. Other countries are the ''CEEC'', including Ukraine and the other Baltic countries

Annex B Parties range from around 10–75% for the US and the EU, on average nearly 50%; cost-saving in Japan is slightly higher. For full global trading (through the CDM), cost reductions are much higher: 60–90% for the US and the EU (ignoring one outlying estimate of only 33%); again cost-saving in Japan is higher, around 80–95% (Weyant, 1999; Ellerman and Decaux, 1998; Grubb *et al.*, 1999).

Taken at face value, these predictions of global cost are not very high, even without any international exibility. Compared to the economic growth and uncertainties about growth in the coming decade, the impact of Kyoto' might not be discernible.

The total scope of exibility' (including the mechanisms and the basket of gases) in the Kyoto Protocol has a profound effect on the emission distribution between the Parties. According to some studies, the OECD Parties will import on average about a quarter of their reductions, and with full global trading nearly three quarters, much of it through trading with developing countries. The Kyoto agreement could leave CO_2 emissions in the US and Japan around 10% above 1990 levels (and even much more for some other countries) without breaching the letter of the formal agreement – and that is before considering additional exibility from sinks and some further exibilities sought after. Figure 2 displays the results of a study by the Authors for the US, as a large importer.

Some Practical Problems

The international transfer mechanisms of the Kyoto Protocol as presented in the previous section are an attractive option for implementing the Kyoto commitments. Apart from the problems discussed earlier, that have to be addressed in the negotiations, there are numerous practical problems to be solved before a market could develop.

The Kyoto Protocol defines three mechanisms for transfers of emission allowances. The extent to which the credits from the mechanisms are equivalent will be crucial. The important questions are: who will be able to trade in the various mechanisms, e.g., will there be an industry-equivalent of Article 17 emissions trading; what gases will be covered by the various mechanisms; will the quota' from the three mechanisms carry the same risk, compliance procedures and therefore price?

The specific design of the CDM faces dilemmas of integrity and size. If the rules on additionality, contributions to sustainable development, etc., are rigorous, then governance could be very complex and administration costs high, thereby deterring investment. Weak rules and governance

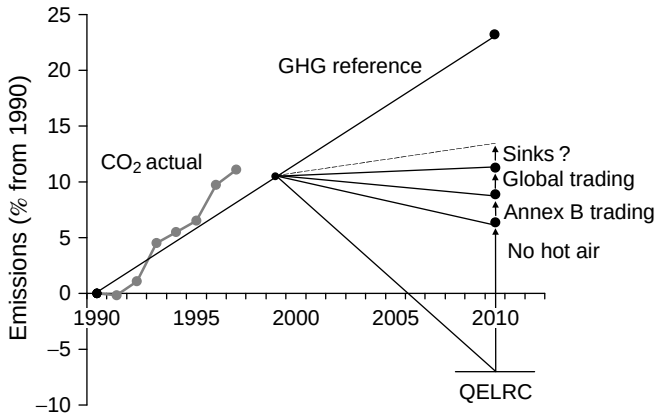


Figure 2 Assigned amounts and possible greenhouse gas emissions under the Kyoto Protocol, United States. Least cost trading in the commitment period, for the base case used in the emissions trading model developed at the Royal Institute of International Affairs, would leave US GHG emissions about 11% above 1990 levels. CO_2 emissions would be around 2–3% higher assuming greater abatement in other GHGs; the use of sinks would add to this difference. Excluding 'hot air' from the emissions market (''No hot air'' in the diagram) would result in additional domestic abatement of about 2.5%

may attract more investment at the risk of corrupting and discrediting the system, as well as undermining the criteria of sustainable development and the objectives of the Convention, since maximizing emission credits' under the CDM is equivalent to minimizing action within Annex I'. However, for both better or for worse this mechanism as generally conceived is likely to be constrained by the inherent limitations of credit-based, project-based systems, and for some countries it may be primarily a bridge towards adopting commitments and joining the international trading system.

The project mechanisms explicitly (and necessarily) involve private entities. Emissions trading will also be most effective when conducted by private entities acting internationally. However, this requires either new legal procedures that transfer part of governmental obligations – assigned amounts – to private entities, or procedures for automating simultaneous transfers of private sector permits alongside governmental assigned amounts.

The Danger of Inflation of Commitments

The political and economic circumstances in which the Kyoto Protocol was negotiated led to some perverse emission allocations, and an overall pattern that may induce large-scale emissions trading, principally from Russia and Eastern Europe to the US, Canada and Japan. The allocations of Russia and the Ukraine in particular could represent surpluses amounting to 5–10% of total Annex I assigned amounts. Transferring this surplus to other Parties would undermine the legitimacy of emissions trading and the further evolution of the whole climate regime.

The emission restrictions of the Kyoto Protocol are not trivial politically, but are clearly inadequate to lead efficiently towards atmospheric stabilization below a doubling of pre-industrial CO_2 levels, or to induce necessary technological changes. Indeed under some scenarios we find real resource abatement costs to be negligible, though Russia and Ukraine might still extract a price and they could act as monopoly suppliers. The implied scale of West–East transfer of both emissions and finance (if the price does stay up) would also raise serious political debate. Extensive use of sinks or the CDM would further reduce the level of effort required in Annex I.

If emission trends and further analysis do point towards such outcomes, debate will be reinforced about various kinds of restrictions on the use of the transfer mechanisms. Excluding trading of surplus, allocations could reduce national and aggregate emissions by several percentage points (as compared with no restrictions), and increase the value of the trades with developing countries (and of trading for most Parties within Annex B). Achieving this is politically and procedurally problematic. The EU has proposed to introduce caps on the amounts traded, stimulating domestic action, quantifying the term supplemental in the Protocol, reducing trades in reductions that would have taken place anyway (hot air'), etc. However, these discussions do not seem to lead to a quick politically feasible solution. Other approaches could focus upon assessing the adequacy of domestic actions.

Also, significant use of the CDM or sinks would violate the collective aim of the Protocol to reduce industrialized-country emissions to at least 5% below 1990 levels; it would allow their emissions to grow from the 5% reduction already achieved by 1995. Excluding trading of surplus entitlements could be one way of making room for the use of sinks and the CDM without violating the collective Annex I commitment, while shoring up the value of emission credits.

A classical economic perspective would suggest that, providing the reductions are real and not fraudulent, this is not a problem: it simply re ects the fact that the exibilities can uncover a huge potential for low-cost reductions. However, this perspective is static, it takes no account of induced innovation or a host of other issues associated with developing long-term efficient trajectories towards atmospheric stabilization. There is, in short, a potential tension between cost minimization in the first commitment period, and the generation of sufficient and efficient pressures required to change course towards long-term stabilization. These complex dynamic questions thus combine with political debates about the appropriate degree of exibility in implementing the Kyoto Protocol, in which some parties fear that *too much* exibility would enable the richest countries to

avoid making the changes and innovations that are central to establishing leadership and long-term resolution of the climate problem.

PROSPECTS AND CONCLUSIONS

The outcome of the Buenos Aires conference focuses most key decisions, including those on the Kyoto Mechanisms, upon the COP-6. Many countries will seek ratification based on the outcome of this conference and re ect other aspects of meaningful participation' of developing countries. Breaking the trend of continuously growing emissions, stabilizing or even reducing emissions of the OECD countries (the Worldwatch Institute, of Washington, DC, reported in late July 1999 that its latest emission estimates show a global decline in emissions (0.5%), while the economy is growing (2.5%)), and entry into force of the Protocol would plausibly establish leadership by the industrialized countries, and create a stronger basis for discussing subsequent global commitments, that are to start in 2005. However, there is also reason to believe that some countries will not ratify the Protocol, and consequently it might never enter into force. Entry into force of the Kyoto Protocol, requires countries emitting 55% of industrialized country emissions to ratify. None of the industrialized Parties on its own could block entry into force. However, the Parties with a large share of emissions, such as the US (over 35%), EU (nearly 25%) and Russia (nearly 20%), are of crucial importance.

Apart from resolving the outstanding issues of the Protocol, experience in the use of the Kyoto Mechanisms has to be built in coming years: experience is growing slowly. Without properly working trading mechanisms it will be politically impossible to keep the Kyoto Protocol from collapsing. Just over 100 projects are reported under the AIJ pilot phase (the second synthesis report on AIJ, FCCC/CP/1998/2, reported 95 projects by June 1998, with total emission reductions of 44 million tons of carbon and annual reductions of 1.91 million tons. One year later the *Joint Implementation Quarterly*, June 1999, reported in total 129 AIJ projects.)

The total emission reductions through these projects is still modest, around 2 million tons of carbon-equivalent, only a tiny fraction of what the forecast of demand is. The Nordic countries have run an emissions trading shadow market with good result. And some companies have started with emissions trading.

One of the most promising examples of industrial involvement in emissions trading is the internal emissions trading scheme of British Petroleum (now BP-Amoco) in September 1998. In the early stages of this experiment, a small number of trades have taken place for a price of $17–$22 per ton of CO_2. The too-perfect allocation of permits between the participating business units may have substantially reduced the number and size of the trades. Incorporation of methane and the other business units, including the newly acquired Amoco units, will broaden the basis of trading.

Many other initiatives have been taken, including more internal trading regimes in industry, industrial trades in emission reduction options, pilot trading schemes in some industrialized countries and (emission) brokers and exchanges diversifying into carbon trading.

The Kyoto Protocol set emission reduction targets for industrialized countries. In environmental terms these targets are modest, but compared to present emission trends the targets that the countries set themselves in Kyoto are tough. For economic and political reasons the Kyoto mechanisms, various forms of exibility through international trade in emission allowances, are an essential tool for reaching the target, since experience from other environmental regimes shows that the potential for cost-savings through trading is substantial.

However, the Protocol settles few of the core issues about how such mechanisms might actually work and be governed in the international context. That task is complicated by the inconsistencies between the mechanisms. The commitments are defined for 2008–2012, but companies need incentives now if they are to find it profitable to take the actions that will be required to meet them. Early and cumulated crediting is allowed for the CDM, the most complex and novel of the mechanisms, which will involve investment in regions where typically projects may take many years to get going. Early and cumulated crediting is not specified for joint implementation among Annex I countries, where governments and companies are already more geared for action. Sinks are allowed under joint implementation, but remain disputed in the CDM. Monitoring and certification procedures are different between the two project-level mechanisms; emissions trading, in theory the simplest mechanism but one on which a great deal hangs politically, remains wholly undefined. The unfinished business of Kyoto's mechanisms for international transfer is immense, as the subsequent Buenos Aires conference illustrated starkly. There is some, but still too little, experience with the trading mechanisms to date.

This article re ects developments up to early 2001.

REFERENCES

Driesen, D M (1998) Free lunch or cheap fix? The Emissions Trading Idea and the Climate Change Convention, *Environmental Affairs Law Review*, 26(1), Boston College, MA.

Ellerman, D and Decaux, A (1998) Analysis of Post-Kyoto CO_2 Emissions Trading using Marginal Abatement Curves, Joint Program on the Science and Policy of Global Change, Massachusetts Institute of Technology (MIT JPSCGC) Report No. 40, October 1998.

Grubb, M (1989) *The Greenhouse Effect: Negotiating Targets*, RIIA, London.

Grubb, M and Vrolijk, C (1998) The Kyoto Protocol: Specific Commitments and Flexibility Mechanisms, *EEP Climate Change Brie ng No. 11*, April 1998.

Grubb, M, Vrolijk, C, and Brack, D (1999) *The Kyoto Protocol: a Guide and Assessment*, RIIA/Earthscan, London.

IEA/OECD (1998) *World Energy Outlook 1998* Paris, *International Energy Outlook 1999: With Projections to 2020*, DOE/EIA-0484(99), Washington, DC.

Vrolijk, C (1999) The Buenos Aires Climate Conference: Outcome and Implications, *RIIA Brie ng Paper No. 2*, RIIA, London.

Weyant, J P (1999) The Costs of the Kyoto Protocol: a Multi-model Evaluation, *Energy J., Special Issue*, 1999.

Tradable Permits for Limiting Stratospheric Ozone Depletion

Rory O'Brien
University of Toronto, Toronto, Canada

Tradable permits are being increasingly used as a policy instrument to encourage industries to reduce their pollution emissions. In 1987, they were rst used in a global context when they were written into the Montreal Protocol on Substances that Deplete the Ozone Layer as the main option for regulating international trade in ozone-depleting substances (ODS) among its signatories. Internationally, production quotas for ODS, such as chloro uorocarbons (CFCs), were allowed to be transferred between Parties to the Protocol, with the stipulations that the total combined levels of the Parties involved not exceed the agreed-on production limits, and that the United Nations (UN) Environmental Programme Secretariat be noti ed of the transfer. Trade restrictions were also effected between Parties and non-Parties.

The trading regime was most useful, however, when it was applied at the domestic level, helping achieve national reduction targets among the major ODS producers, notably the US, through inter-corporate trades. Trading has not been extensively used at an international level, with only one transfer of production quotas between countries to 1998, and only 20 international trades with US companies from 1992 to 1995. This does not mean it has been a failure, however. Its main success lies in being an integral part of a precedent-setting political and economic system implemented between and among governments and industries on a worldwide scale. It also has been instrumental in encouraging universal participation in the Protocol and discouraging participants from relocating operations to non-Party countries.

At a domestic level, as evidenced by the United States, permits have been used in conjunction with other economic incentives to successfully ensure that the phase-out of ozone depleting chemicals was accomplished by 1995 year-end, in accordance with the schedule set out in the Protocol. Over 238 million kg of ODS in the US were transferred in 494 inter-company trades between 1992 and 1995.

One of the main issues arising out of the implementation of the global trading system is that of ensuring compliance. One manifestation of this problem is the rise in illegal imports and exports, mainly between developed and developing nations. This has led to an increased emphasis on licensing systems and customs monitoring. Another situation, wherein certain countries not in compliance with the Protocol were exporting new ODS under the legal provision for exporting recycled ODS, was resolved in 1997 by prohibiting those countries from exporting any ODS.

OVERVIEW OF POLICY INSTRUMENTS

There are two main types of policy instruments available to government seeking to in uence the environmental performance of polluters. These are *command-and-control regulations* and *market-based incentives.*

Command and control regulation has usually been the predominant public policy instrument to deal with the problem of corporate environmental pollution. In many cases, legislation is used to set limits on the amount of pollution a company is allowed to create. It also can ensure compliance through a system of mandatory reporting. Besides emission controls, other examples include: bans, production input controls, consumption controls, price controls, rate of return regulation, technology standards, licensing and certification, and standards of environmental impact. Anyone failing to comply faces penalties.

Market-based instruments (also called economic instruments) place financial burdens on polluters, to the extent that these costs provide an incentive to reduce pollution. Though still generally requiring a regulative framework, market-based instruments allow more scope for initiative, financially rewarding those who make the most effort to comply.

Examples of market-based instruments include: charges (to industry and consumers to cover costs of pollution), market creation (tradable permits, market/price intervention, liability insurance), subsidies (grants, soft loans, tax allowances), deposit-refund systems, and financial enforcement incentives (non-compliance fees, performance bonds).

Final policy choices are frequently a mixture of several options, depending on the circumstances. In general, industry prefers market-based incentives to command

and control regulations, as it gives more autonomy and exibility.

TRADABLE PERMITS

One of the policy options used in the case of the Montreal Protocol was that of a regime of tradable permits. The basic operation of a permit-trading system is fairly simple – an overall limit on the total emission of a pollutant is established, followed by a schedule for periodically reducing that total amount. A reasonable unit of pollution is set, divided into the total amount, and allocated to the individual countries, or firms in an industry, on some kind of equitable basis (such as an historical baseline, or an auction). These units become permits to pollute, and the owners treat them as property rights which, like other assets, they are allowed to sell or use as they see fit.

To illustrate the utility of such a system of tradable permits, we can look at US efforts to reduce the level of toxic lead in the environment. In 1973, the US Environmental Protection Agency (EPA) instituted a program of mandated reductions in the amount of lead added to gasoline. In 1982, to provide small refiners with more exibility in meeting new, lower lead limits, the EPA set up a system for allowing refineries to trade their rights to add specified quantities of lead to their gasoline. The difference between the amount a refiner was allowed, and the reduced amount it actually added, could be sold to other refiners, or, beginning in 1985, banked for future use or sale. The program, which ended in 1987 with the complete phase-out of lead additives, was considered to be quite successful. It resulted in the elimination of lead in gasoline within a short timeframe, and did so in a cost-effective manner – it was estimated that the program saved refiners $226 million (*see* Tradable Permits for Sulfur Emission Reductions, Volume 4).

TRADABLE PERMITS AND THE MONTREAL PROTOCOL

In 1985, the Vienna Convention for the Protection of the Ozone Layer was put in place as an international framework convention to deal with the problem of atmospheric ozone depletion. The Convention's Montreal Protocol on Substances That Deplete the Ozone Layer, which came into effect in 1989, regulated the details of policies and implementations (*see* Ozone Layer: Vienna Convention and the Montreal Protocol, Volume 4).

The Montreal Protocol was the first multi-national agreement to use tradable permits as part of a major policy package addressing a global environmental problem. The system that was put into place had to address the problems of encouraging a fair, equitable, cost-efficient reduction in ODS, while at the same time ensuring inclusivity and compliance among nations.

Designating Participants and Terms of Trade

The trading participants are those countries that have signed the Protocol (173 states in 1999). Each participant, however, is bound only to adhere to the initial Protocol, unless they are also signatories to its subsequent amendments.

Parties are considered developing countries if they have less than an annual consumption of the controlled substances of 0.3 kg *per capita*. To avoid the economic hardship of having to quickly adopt new and expensive ODS substitutes, developing countries were given extensions of approximately 10 years beyond that of the developed countries to meet reduction targets. In setting baseline amounts of ODS, developing countries have the choice of using the average of its national annual consumption between 1995 and 1997, or the 0.3 kg *per capita* consumption.

The actual traders of permits have been those companies in developed nations that produce ODS. While there were only 17 producing companies in the world in 1985, in 2000 there are almost 80, with most ODS being produced in India and China.

In the US, the federal government allocates production and consumption allowances that can be traded with domestic or foreign companies. Companies in the European Union (EU) can trade with any company that is part of the Union, or with any other Party, provided their combined quota levels do not exceed the pooled limits of the EU as a whole.

ODS trading with countries not party to the Protocol is restricted. Imports of ODS from non-Parties have been prohibited since 1990, and exports to non-Parties banned since 1993. This is designed to prevent leakage, i.e., reduction strategies pursued in one region that could result in an increase in emissions in another region (such as when an ODS producer finds it decreases expenses by transferring operations to a country that has not set emission limits and therefore does not impose costly compliance measures). The Parties also encourage non-Parties to sign the Protocol.

Setting Quotas and Schedules

The quotas for production and consumption of substances and their schedule for phase-out were set according to a complex formula for various substances grouped according to their molecular structures. The groups, or baskets, of ODS are weighted according to their ozone-depleting potential (ODP). This means that one chemical in the group can be replaced with another, enabling the Party to find other chemical substitutes in the least costly manner. In the Protocol, *production is* defined as the amount of controlled substances produced, minus the amount destroyed by approved technologies and minus the amount entirely used as feedstock in the manufacture of other chemicals. The amount recycled and reused is not to be considered as

production. *Consumption* means production plus imports minus exports of controlled substances.

Production and consumption limits are based on an allowed proportion of production and consumption by a Party in a particular base year. In post-Montreal meetings of the Parties, this proportion and the time limits changed. CFCs, the most important type of ODS, for example, were originally supposed to be cut to 50% of 1986 levels by 1998, but were later set to be 100% eliminated by the end of 1995 as a result of new scientific evidence on the rate of ozone destruction. Stockpiling of surplus ODS produced within production limits is allowed, but unused production quotas are forfeited.

Ensuring Compliance

The Ozone Secretariat in Nairobi acts as the Administrator of the system. Information on baseline year amounts, periodic reports on ODS production and consumption, and documentation on all trades, including imports and export, must be filed with the Secretariat in a timely manner. Such reporting has shown a need for improvement – in 1997, 38 of 151 countries had not filed their 1995 data, and only 43 had reported their 1996 data.

Illegal imports/exports, i.e., smuggling, is becoming more of a problem as the price differentials increase between developed and developing countries and demand remains for inexpensive supplies. Such trade is reported to be over 20 000 tonnes of ODS per year in 2000. Protocol signatories have begun implementing harmonized' customs codes and tracking procedures, and now require import/export licensing systems to be in place. Individual Parties are also increasing enforcement efforts, with strict penalties being imposed on smugglers – in the US, for example, 20 individuals have received jail sentences and fines totaling $60 million.

Several states, including Latvia, Lithuania, and the Russian Federation were found to be in non-compliance in 1996, meaning they were continuing to produce a controlled substance past its phase-out date. This was accompanied by notice of some instances of illegal exports of newly produced ODS falsely labeled as recycled and reused substances (which are allowed under the Protocol). As a consequence, a prohibition against all exports of non-complying substances by those countries went into effect in 1997.

A number of signatories to the Protocol have used tradable permits as a policy option, including Canada, members of the EU, New Zealand, Singapore, and recently China, but it has been the United States that has used it most extensively. In 1985, the United States, the European Community, and Japan accounted for 70% of the global production of CFCs. Over 40% was produced by the US alone, with a single company, Du Pont, controlling half the North American market. As the US was the largest producer and user of CFCs and other ozone destroying chemicals, it is worthwhile reviewing its experience in implementing the permits system.

THE TRADABLE PERMIT SYSTEM IN THE UNITED STATES

In 1978, regulations were enacted in the US to prohibit the use of CFCs as aerosol propellants. Anticipating that further regulation would be needed to resolve the ozone issue, the EPA in 1980 solicited responses from industry regarding other possible CFC control measures. Such measures included mandatory controls (such as requiring the use of alternative technologies), taxes and marketable permits.

The EPA considered economic incentives primarily for two reasons. First, CFCs are not of concern near emission sources (no direct health or biospheric impacts there) but they have long residence times, disperse widely through the global atmosphere and ultimately participate in stratospheric chemical reactions. So a single national tax or permit system, which is relatively simple to set up, can be applied uniformly.

The second reason had to do with the large numbers of small firms that used CFCs in their production processes. Any attempt to come up with acceptable regulations covering the gamut of uses, as well as monitoring and enforcing control measures would be almost impossible. A general limit on CFC production and use, or a tax, however, allows the voluntary use of a wide range of economic and technological options.

Economic studies commissioned by the EPA also indicated that a permit system was a cost-effective solution, particularly if it were applied to the small number of producers and importers of CFCs. In 1988, regulations were put into place to create a permit system based on the Montreal Protocol (and in 1990, Congress passed an amendment to the Clean Air Act specifically enacting the provisions of the Protocol).

Tradable allowances were allocated to a handful of firms (eight producers and 20 importers) using 1986 as the baseline year. These firms were allowed to trade their allowances within their own companies and with other firms, both within and outside of the US. It is interesting to note that providing allowances to a limited number of large corporations produced windfall profits in the billions of dollars for those companies. Congress took the opportunity to institute an excise tax on ODS transactions, both to cover costs of administering the EPA's program as well as providing another incentive to switch to untaxed chemical substitutes.

There were several other pertinent regulations:

- to help speed up the retirement of allowances, and hence quicken ODS reductions, an offset discount'

was imposed on each trade, reducing the transferor's stock of permits by 1% of the amount traded;
- in accord with the Protocol's formulaic definition of production and consumption, producers of ODS were required to hold both production and consumption allowances, while importers needed only to have consumption allowances;
- producers were allotted an additional 10% of their quota to be used for exporting to developing nations, in the period prior to 1993, provided records were kept and the recipient country did not re-export the materials; and
- allowances were provided for the destruction of controlled substances.

Data on the number, size and frequency of trades reveals that the system was well used. In 1992, 171 trades totaling 73 million kg were made, representing 13% of the allowances available for that year. The great majority of these trades were intra-company trades, and international trades were almost negligible.

Was the US permits regime successful? It appears it was. From 1989 to 1994, CFC production and consumption in the US remained well below that required by the Protocol, with full phase-out occurring on schedule by 1996. It is difficult to separate the effects of tradable permits from those of the many other related regulatory initiatives, but there are some assertions that can be made. One is that it was cost-effective, due mainly to the decision to regulate only a few large producers and importers, which resulted in low administrative overhead. Another is that the overall simplicity of the regulatory environment enabled the exibility to quickly adapt to the changing adjustments to the Protocol. A third is that it contributed to marketplace efficiency, stemming from the fact that no specific technologies were mandated to achieve targets. This allowed industry to rapidly develop and adopt new technologies in order to realize cost savings.

CONCLUSION

The drafting of the Montreal Protocol and its subsequent amendments has been the result of some enlightened experimentation in the fields of politics, science, and economics. Such experimentation has produced, in a relatively short time span, a workable system of production, trade and consumption that, along with other policy instruments, has succeeded in reducing the amounts of ozone-depleting substances in the world. One of the mainstays of this system is the allocation of permits to produce and consume a variety of controlled substances, as well as the option to transfer them to other Parties. This permit trading model, which operates simultaneously at a supra-national level, and, with somewhat more vigor, at the national level, has been very good at promoting reductions in ODS in the developed nations, but has not yet had much opportunity to prove itself a good mechanism for reducing ODS in the developing world.

Tradable permits have proved to be a good choice as a policy instrument. They are accepted by business because they allow for exibility in operations, and promote cost-efficient methods of complying with the required abatement in pollution. Governments tend to prefer them as well, as they often require less costly and bureaucratic infrastructures to implement than traditional command and control measures. But this is not to say that they can and should be used in all circumstances. They are a tool that holds promise, surely, as shown so far in the case of the Montreal Protocol. However it remains to be seen whether they will prove to be as versatile or as effective when applied to the remaining accumulating stocks of ozone depleting chemicals, or to the far larger problem of reducing greenhouse gas emissions.

For other examples of emission trading, *see also* Emissions Trading, Volume 4; Tradable Permits for Greenhouse Gases, Volume 4; Tradable Permits for Sulfur Emission Reductions, Volume 4.

FURTHER READING

Cassils, J (1991) *Exploring Incentives: an Introduction to Incentives and Economic Instruments for Sustainable Development*, NRTEE, Ottawa.

Hahn, R and Hester, G (1989) Marketable Permits: Lessons for Theory and Practice, *Ecol. Law Q.*, 16(2), 361–406.

Lee, D (1996) Trading Pollution, in *Ozone Protection in the United States*, ed E Cook, WRI Publications, Baltimore, MD, 31–38.

Mullins, F (1998) Lessons from Existing Systems for International Greenhouse Gas Emission Trading, 06/08, http://www.oecd.org//env/cc/freedocs.htm#emis10, in *Climate Change Documents*, 28/07/00.

OECD (1997) Experience with the Use of Trade Measures in the Montreal Protocol on Substances that Deplete the Ozone Layer, 22/12, http://appli1.oecd.org/olis/1997doc.nsf/linkto/ocde-gd(97)230, in *Trade and Environment*, 28/07/00.

OECD – Joint Working Party on Trade and Environment (1999) Trade Measures in Multilateral Environmental Agreements: Synthesis Report of Three Case Studies, 15/02, http://appli1.oecd.org/olis/1998doc.nsf/LinkTo/com-env-td (98)127-final, in *Trade and Environment* 28/07/00.

Ozone Secretariat – UNEP (2000) *Handbook for the International Treaties for the Protection of the Ozone Layer*, http://www.unep.org/ozone/Handbook2000-full.htm, 28/07/00.

Shapiro, M and Warhit, E (1983) Marketable Permits: The Case of Chloro uorocarbons, *Nat. Resour. J.*, 23(5), 577–591.

Stirpe, D (info@arap.org) (1999) *Background on Illegal Imports of CFCs*, http://www.arap.org/textonly/pr/illegalcfcimports.html, 28/07/00.

Tradable Permits for Sulfur Emission Reductions

Roderick W Shaw
Rodshaw Environmental Consulting Inc., Nova Scotia, Canada

Acidic deposition from the atmosphere (more commonly known as acid rain) has been considered to be an environmental problem for several decades in various regions of the world, including eastern North America, Europe and Asia. Emission trading is a market-based approach to controlling emissions of acidic deposition through the buying and selling of allowances to emit acidifying pollutants such as sulfur and nitrogen. In contrast to the command and control approach where the maximum allowable emissions from each source are regulated, in emission trading a limit (i.e., the total number of emission allowances) is placed upon the total emissions from those sources involved in the trading. The sources may then trade amongst themselves. For example, if a source could reduce its emissions at a cost per tonne which is less than that it could receive from selling its excess emission allowances to another source, it has the potential to reduce its emissions and make a pro t. On the other hand, a source which nds itself in need of additional emission allowances (for instance because it is planning an expansion, or because the number of allowances to which it is entitled has been reduced for environmental reasons by the regulating body) may purchase the necessary emission allowances if they are available at a cost which is less than its cost of reducing emissions. Through trading, therefore, the cost of maintaining or reducing emissions in the trading area may be less than that of the command and control approach, and managers of individual sources may have more exibility in meeting air pollution requirements. Ideally the only restrictions would be that the total number of allowances is not exceeded and that environmental requirements such as local air quality standards or maximum rates of acidic deposition are met.

By far, the most experience with emission trading to reduce acid-forming emissions has been in the US. The US Environmental Protection Agency (USEPA) estimated in 1994 that compliance costs in the US could be reduced by 3.1 billion dollars per year as compared to conventional regulatory approaches. In 1996, emissions of sulfur dioxide were 2.9 million tons, or 35% below the emissions cap for that year. The agency also claims that these emission reductions resulted in a decrease of 10- 25% in deposition of sulfur by precipitation. Emission trading is being discussed in Canada and Europe but has not been adopted to any great degree there.

INTRODUCTION

The need to Reduce Acidic Deposition from the Atmosphere

Acidic deposition from the atmosphere (more commonly known as acid rain) has been considered to be an environmental problem for several decades in various regions of the world including eastern North America, Europe and Asia. This deposition originates with the emission to the atmosphere of sulfur and nitrogen oxides from both natural and human sources, with the latter being dominant in populated areas. Human sources of sulfur oxides include thermal–electric generating stations fueled by fossil fuels containing sulfur, and industrial sources such as non-ferrous ore smelters and oil refineries. These emissions are transported hundreds or even thousands of kilometers through the atmosphere, converted to acidic precipitation and fog, and deposited on receptors of varying sensitivities. Where fresh water bodies and forest soils are relatively sensitive to acidic deposition, this deposition is believed to result in adverse environmental effects such as decreases in fish populations and increased stress upon forests.

Approaches to Reducing Emissions of Acidifying Air Pollutants

The traditional approach has been command and control which sets rigid upper limits to the release of pollutants by a specified type of emission source, such as power plants or smelters. These limits are based upon generally accepted technologies. To allow for factors such as the general health of the local economy, some leeway may be given in the emission limits, or in the timetable for implementing them. While the command and control approach appears to be fair and relatively simple from an administrative point of view, it does not take into account the costs of controlling emissions at a particular installation (even if those costs are much higher than average). Nor does it take into account whether the emission reductions will result in levels of acidic deposition that are within the assimilative capacity of the environment.

In contrast, with emission trading, the regulating body sets a limit to the total emissions by all sources involved in the trading, and distributes an emission allowance to each source involved in the trading. If a source could reduce its emissions at a cost per tonne, which is less than that it could receive from selling its excess emission allowances to another source, it has the potential to reduce its emissions and make a profit. On the other hand, a source which finds itself in need of additional

emission allowances (for instance because it is planning an expansion, or because the number of allowances to which it is entitled has been reduced for environmental reasons by the regulating body), has a choice. It can reduce its emissions (usually through technological means such as abatement equipment) or, alternatively, it can purchase the necessary emission allowances if they are available at a cost, which is less than its cost of reducing emissions.

Therefore, it can be seen that emission trading will not only keep the total emissions among the trading sources within the prescribed limit, but it will also give individual sources the exibility of how they go about managing their emissions. Studies have shown that the total cost of keeping emissions within the total allowance can be less than that incurred through the command and control approach. Individual sources can benefit by having the potential to make a profit from creating surplus emission allowances for sale, or by purchasing emission allowances that cost less than the emission abatement that would be otherwise needed.

HOW EMISSION TRADING WORKS

There are several steps involved in setting up a scheme to trade emission allowances.

1. The area within which trading will occur is delineated. Trading can occur at many levels:

   - within a given site such as a plant;
   - within several plants owned by a particular firm;
   - among several firms;
   - among jurisdictions or;
   - within a specified geographical area.

   As will be shown in the section on taking the environment into account, the trading area must be large enough to include enough sources such that trading will bring about the required reductions in deposition.
2. The total number of emission allowances is determined. One possibility is that the initial total allowance should be approximately equal to the total emissions of the trading sources at the present time. This assumes that the present rate of acidic deposition be approximately maintained, although a shift of emissions through trading could conceivably concentrate emissions in a particular area, and result in undesirable increases in acidic deposition in certain parts of the trading area. If the present rate of acidic deposition is already too high (as it often is), then the total number of allowances may have to be gradually reduced with time.
3. The emission allowances are allocated among the trading sources. One simple approach is to initially distribute the allowances in proportion to the sources' emissions in a given reference year. This could be potentially unfair because sources which have not in the past reduced their emissions receive more than their fair share of the allocation. In addition, before the trading scheme has been initiated, it is important not to give sources the opportunity to increase their emissions artificially, so as to gain an unfair advantage in the initial allocation of emission allowances.
   Some proponents of emission trading suggest that the initial allocation of emission allowances be done by auction. There is a danger, however, that larger firms could artificially drive up the price of emission allowances, thereby sti ing competition from smaller firms. In the US, initial auctioning of all allowances has been rejected, although a small fraction is auctioned each year to provide opportunities for trading.
   It should be noted that an emission allowance gives the holder the right to emit a specified amount of the air pollutant not for perpetuity but for a fixed period, usually a year. It would be only fair that a source permit would be reissued each year with the same number of allowances that it held the previous year although, if the acid deposition situation requires it, a certain fraction of a source's emission allowances could be withdrawn from one year to the next, to reduce the overall emissions. In the US, to maintain a competitive market, 2.8% of a source's sulfur dioxide allowances each year are set aside to make them available in a sealed-bid auction.
4. An organization is established for keeping track of the ownership of emission allowances, ensuring that sources possess enough allowances to cover their emissions, and disciplining those who emit in excess of their allowances.
   Monitoring of emissions will be important to reduce the chances of cheating. In the command and control approach, the very existence of an installed piece of abatement equipment would be a strong indication that emissions are likely within the required limits. (If an emitter goes to the considerable expense of installing equipment, there is a good chance that they might operate it properly). In contrast, emission trading would require closer monitoring of emissions, perhaps on a continuous basis.
   Enforcement of the emission limits is similar for the command and control and for the emission trading approaches. Financial penalties for emissions exceeding the allowances should, of course, be greater than the cost of abating emissions to be within the allowances. Revenues from fines could be used at least to partially recover the costs of emission monitoring, and the administrative costs of the trading scheme.

More detailed descriptions of various aspects of emission trading may be found in Klaassen and Forsund (1994); NERA (1991) and UN (1995).

TAKING THE ENVIRONMENT INTO ACCOUNT

Trading emission permits for gases that lead to acidic deposition has to take into account where the trades take place. This is because these gases have relatively short atmospheric lifetimes (hours to days to at most weeks) and do not have time to become uniformly mixed. As a result, emissions of acidifying gases from a given source will lead to more acidic deposition at some receptor points than at others, because of patterns of prevailing winds and precipitation.

The relationships between emissions at given sources and deposition at given downwind receptors are called source–receptor relationships which are estimated through the use of computer models of the long-range atmospheric transport of air pollutants. Usually, these models track a unit or puff of emissions, which is emitted from a source and transported with the observed wind. During successive time intervals (typically one or several hours) the sulfur dioxide and nitrogen dioxide in the puff are transformed chemically to sulfates and nitrates, and incorporated into cloud droplets and precipitation, which are acidified. In addition these pollutants are deposited to the ground by dry deposition (in the absence of precipitation), and by wet deposition (e.g., precipitation).

The computer model repeats the above calculation for all time intervals over an extended period (typically months or years), and for all sources. It is obvious that generating source–receptor relationships can be a very computer intensive (and expensive) undertaking. Nevertheless, once the task has been completed, one has an estimate not only of the total amount of acidic deposition at a given receptor point, but how much of it originated from each source. Given the great expense of installing pollution control equipment, spending money to generate the best possible source–receptor relationships to aid in devising more efficient air pollution control strategies is a good investment.

A model that has been extensively used for developing an assessing control strategies for acidic deposition in Europe for the past decade or so is the regional acidification information and acidification (RAINS) model, developed at the International Institute for Applied Systems Analysis (IIASA) in Laxenburg, Austria. The RAINS model consists of many modules describing energy use, emissions of acidifying pollutants, atmospheric transport and deposition using source–receptor relationships, and environmental impacts. The model has been a valuable tool for the United Nations Economic Commission for Europe (UN–ECE) in developing sulfur and nitrogen emission protocols under the aegis of the Convention on Long-range Transboundary Air Pollution (*see* RAINS (Regional Air Pollution Information and Simulation), Volume 4).

The importance of source–receptor relationships has several implications for acidic deposition control strategies, including emission trading schemes. On the assumption that we want to control deposition at certain sensitive, ecologically valuable receptor sites,

1. It is essential that as many sources as possible affecting the sensitive receptor sites in a substantial way be included in the trading scheme. A study of the feasibility of trading sulfur dioxide emissions within the four Atlantic provinces of Canada to reduce acidic deposition there found that, to bring deposition to within ecologically acceptable ranges, the trading area would have to be expanded beyond the four provinces to include parts of the eastern US and central Canada.
2. Conversely, apart from considerations of local air quality, emissions from sources that do not affect sensitive receptor areas should receive less priority in any control scheme.
3. The effects on acidic deposition of any trading of sulfur and nitrogen emissions must be assessed, and the best way of doing this is through the use of source–receptor relationships. Both the command and control and the emission trading approaches can assure that the total emissions from a group of sources stay within a pre-determined limit, but they do not ensure that acidic deposition will be reduced. In fact, emission trading, if carried out strictly on the basis of profitable buying and selling, could result in emissions being shifted geographically such that acidic deposition would *increase* at some receptor points.

In the US, sulfur dioxide emission trading among power plants has not taken into consideration the effects upon reducing acidic deposition; the assumption has been that reducing emissions will result in an overall reduction in deposition. Indeed, a study by the USEPA concludes that, by 2010 when the emission trading program will have been implemented, there will be no more than a 10% difference in acid deposition with or without trading. This suggests that trading will perform as well as command and control in preventing hot spots of acidic deposition.

In the above mentioned study in Atlantic Canada, a computer model maximized trading within the region but with no restraints upon changes in acidic deposition; acidic deposition at several sensitive receptor areas was calculated to actually *increase*. Another experiment was then carried out, also maximizing trading but not allowing deposition to increase anywhere. In this second experiment, it was still possible to carry out the maximum number of trades, but deposition actually *decreased* in several receptor areas.

SOME QUESTIONS AND ANSWERS ABOUT EMISSION TRADING

Is Emission Trading not just Trading Licenses to Pollute?

Anyone operating a plant or device, whose pollutant emissions are regulated in some way, has in effect a license to pollute. For example, an operator of an automobile is allowed to emit so many grams of carbon monoxide per kilometer travel. Therefore, there is no difference between command and control and emission trading with respect to being licensed or entitled to emit air pollutants. With command and control the entitlement is attached to a single source whereas in emission trading the entitlement is held by a group of sources that may share it in ways determined by market forces. One advantage of emission trading is that the entitlement is not increased as new sources are added, while with command and control this could happen, to the detriment of the environment.

Can Emission Allowances be Stockpiled and could this not be Detrimental to the Environment?

An emitter can create surplus allowances (either by reducing emissions or by purchasing allowances on the market), declare them surplus and, rather than sell them right away, formally register them in a bank. This formal banking of allowances has advantages over informally held surplus allowances that are lost at the end of their valid period. Formally banked surpluses retain their validity for a longer period and can either be used by the holder in the future when it is more advantageous from an operational point of view (e.g., when undergoing an expansion). Alternatively, they can be sold at a profit in the future if the holder anticipates that the market price will rise because of a scarcity (e.g., the emission cap or total number of allowances is expected to shrink because of tighter environmental restrictions, or because new sources are expected to join the market).

Banking can in the short run actually benefit the environment because it may give emitters an incentive to reduce emissions and create bankable surpluses. However, the unrestricted use of banked allowances at a future time could be detrimental to the environment by leading to a spike of emissions unless, as was discussed in the section on taking the environment into account, their use is restricted by considerations of local air quality, or maximum allowable rates of acidic deposition.

It is important, however, that sources contemplating the banking of allowances feel confident that the banked allowances will not be devalued or confiscated. This could happen, e.g., when the authorities feel at some point in the future that emissions must be reduced at any cost for environmental reasons. Should this be required, it would be fairer to reduce the value of *all* allowances (banked and in use), so that everyone and not just those holding banked allowances shares the burden of reducing emissions.

WHERE EMISSION TRADING HAS BEEN USED FOR SULFUR AND NITROGEN EMISSIONS

United States of America

Early steps

By far, the most experience with emission trading to reduce acid-forming emissions has been in the US where various types of trading have been carried out since the 1970s. The 1970 US Clean Air Act amendments required each area of the US to attain ambient air quality standards for oxides of nitrogen and oxides of sulfur, among other pollutants. The principal means of improving air quality was through the traditional command and control approach. However, following this amendment, various policies were introduced to complement the command and control approach and in particular to accommodate economic growth and to provide the managers of facilities with more exibility. In 1976, the USEPA adopted its *offset* policy. This was the first in a series of policies that eventually evolved into a full- edged emissions trading policy.

The basic currency of any US emission trading scheme is an emission reduction credit (ERC), which is any emission reduction created when a company controls its emissions to a degree greater than the law requires. This may be done through changing its process (including changing fuels), installing pollution control technology, curtailing production or, in an extreme case, shutting down production.

Offsets are required for companies wishing to build or expand their operations in areas where air quality standards have not yet been attained. Normally, approval of new or expanded plants would have been impossible to obtain in these non-attainment areas. The offset policy allowed proposed new or expanded sources to purchase offsetting emission reductions from another source in the non-attainment area. In this way, the new or expanded source would not result in a net increase of emissions in the area.

Over the next 10 years, the emissions trading policy in the US evolved into a more sophisticated four-pronged approach, involving offsets, netting, bubbles and banking. *Netting* allows a firm to balance expected emission increases at some stacks at a given plant with emission decreases at other stacks at the same plant.

Bubbles provide exibility to managers of sources in that they can cover a single plant or a set of plants owned by the same firm. A manager can make emission control decisions in aggregate for all sources covered by the imaginary bubble rather than by each emission point within the firm's bubble.

For example, it might be much cheaper to control emissions at one stack than at another within the bubble; it would make economic sense, therefore, to over-control emissions at that one stack and under-control at others.

Banking, as has already been described, gives firms a legal means of storing surplus emissions (ERCs) for later use.

Recent Experience in the United States

In 1990 the Clean Air Act in the US was amended again; Title 4 of the Amendments set three major goals to combat acidic deposition:

- by the year 2000, a reduction in sulfur dioxide emissions of 10 million tons per year from 1980 levels, primarily from coal-fired electric generating stations. (In the US, English tons (equal to 2000 lbs or 908 kg) are used, rather than the metric tonne (equal to 1000 kg).
- a nationwide cap of sulfur dioxide emissions, beginning in the year 2000.
- a two million ton per year reduction in nitrogen dioxide emissions.

These goals were to be met through a two-phase program. In Phase I, beginning in 1995, part of the sulfur and nitrogen oxide reductions were to be achieved through emission reductions at the largest, highest-emitting plants. During Phase II (beginning in 2000), the emission reductions are to be attained through more stringent measures at virtually all fossil-fuel power plants.

Title 4 of the Clean Air Act Amendments allows for the creation of tradable *emission allowances*, an authorization to emit one ton of sulfur dioxide (or nitrogen dioxide) for a specific year. At the end of each year, each unit is to hold allowances at least equal to its annual emissions. If a unit emits more than it is allowed, a penalty of $2000 must be paid per excess ton of emissions (this fine was set to make it uneconomical to emit in excess of allowances). Moreover, the excess emissions must be offset (deducted) in the following year. Regardless of the number of allowances a unit holds, it is never entitled to exceed the local air-quality standards.

Beginning in 1995, allowances were allocated for each year but could be stored (banked) for future use. The distribution of allowances was based upon average fossil fuel consumption from 1985–1987. Allowances can also be obtained from three US reserves. The major one is a setting aside of 2.8% of the annual allowances for an annual auction and direct sales at a cost of $1500 per ton. Allowances can also be purchased on the secondary market from those sources that have created excess emission allowances for sale. New sources (those that begin operating in 1996 or later) do not receive any allowances but must buy them from the secondary market or at EPA auctions.

Allowances may be bought, sold or banked by any person, not only electric utilities, but also individual corporations, brokers, environmental groups or private citizens. For example, an environmental group, if it can raise the necessary funds, can buy emission allowances and then not use them, thus improving local air quality, and decreasing acidic deposition.

The USEPA estimated in 1994 that compliance costs in the US could be reduced by 3.1 billion dollars per year as compared to conventional regulatory approaches. In 1995, the largest one-year drop in emissions since 1970 occurred. In 1996, emissions of sulfur dioxide were 2.9 million tons, or 35% below the emissions cap for that year. The agency also claims that these emission reductions resulted in a decrease of 10–25% in deposition of sulfur by precipitation. Some researchers believe, however, that the environmental benefits and costs savings from emission trading have yet to be demonstrated. Concentrations in ambient air of sulfur dioxide also decreased by 17% between 1994 and 1995. However, it may still be somewhat early for a full evaluation. At least one researcher suggests that there is as yet very little hard evidence that trading has provided any increase in environmental benefits, or encouraged innovation and technological change, or economic efficiency over what might have occurred under a command and control program with appropriately designed incentive structures.

Trading among sulfur and nitrogen oxides (both of which contribute to acidic deposition) and other pollutants could theoretically take place, but this has not yet occurred in the US or elsewhere.

Canada

In Canada, experience in the trading of acid-forming air pollution emissions has been limited to within-firm trading of sulfur dioxide and nitrogen dioxide emissions by several industries in Ontario, and of sulfur dioxide emissions by power utilities in the province of Nova Scotia. In Ontario's Countdown Acid Rain program, a schedule was imposed upon the four large industries (three copper–zinc smelters and the provincial electricity generating utility) to reduce emissions between 1986 and 1994. All four companies are now operating under their own emission bubble. Although there is no trading permitted among the four companies, within the electric utility, trading of emissions can take place among the 28 boilers involved; it is estimated that trading has resulted in substantial cost savings. In Nova Scotia, the power utility operates within a sulfur dioxide bubble of 145 000 tonnes per year; trading of emissions among plants of the utility is permitted.

Europe

To date, experience with trading emissions within Europe has been relatively limited as reliance is still placed mainly

on the command and control approach. However, as will be described below, some aspects of emission trading such as offsets and bubbles have been used in several countries.

In Germany and Austria, it is possible for a new or expanded plant to be built in an area where air quality standards are not met, if the new plant replaces an existing plant of the same type, or if the existing plant is renovated such that its emissions are reduced. The offset would, however, have to result in a reduction of the annual average pollutant concentration in the area, and the new plant would have to meet state-of-the-art emission standards.

As early as 1984 in Denmark, a bubble of 116 000 kilotonnes of sulfur dioxide (to be attained in 1995) was imposed upon emissions from power plants. In 1987, an additional bubble for nitrogen oxides of 85 000 kilotonnes was set. Data suggest that actual emissions of both sulfur and nitrogen oxides were around 10% less than the allowed bubbles. It has not been possible to estimate what cost savings the bubble approach might have brought about to the power producers.

In the Netherlands, the state and provincial governments signed an agreement with the association of electricity producers to reduce sulfur dioxide emissions from 41 kilotonnes in 1989 to 18 kilotonnes by 2000, and nitrogen dioxide emissions from 74–30 kilotonnes during the same period, on the basis of a shrinking bubble. It is estimated that the trading will save the electricity producers about half the cost of 500 million guilders (about $250 million) of using uniform emission standards.

Despite the above use of offsets and bubbles, there has been no evidence in Europe of a market where permissions to emit were transferred from one party to another in exchange for money. Researchers suggest that there are no sound economic and environmental reasons for European reluctance to adopt more aspects of emission trading. One difference between the US and Europe is that there is a greater geographical differentiation of sensitivity to acidic deposition in Europe. This means that, e.g., credits or allowances obtained by reducing emissions at sources, which affect Scandinavia (a relatively sensitive area) might be considered to be worth more than those from sources, which affect southern Europe (a relatively less sensitive area). *See also* Tradable Permits for Greenhouse Gases, Volume 4; Tradable Permits for Limiting Stratospheric Ozone Depletion, Volume 4.

There may also be a cultural reason for the differences in degree of acceptance of trading in the US and Europe. In the US, government intervention is less than in Europe, and more reliance is placed upon a market-based approach to solving problems, including air pollution. In Europe, government intervention and regulation is more accepted. Furthermore, there is more uniformity of environmental policy in North America than there is in Europe; e.g., in southern Europe, environmental policy in still in the inception stage. One should also keep in mind that emission trading has been evolving in the US for more than 25 years and might take a similar period to do so in Europe.

REFERENCES

Klaassen, G and Forsund, F R, eds (1994) *Economic Instruments for Air Pollution Control*, Vol. 19, Kluwer Academic Publishers, Dordrecht, Netherlands, 281.

NERA (1991) *Market-Based Approaches to Managing Air Pollution in Alberta*, Discussion paper prepared for Alberta Departments of Energy and Environment by National Economic Research Associates, Cambridge, MA, 227.

United Nations (1995) *Controlling Carbon Dioxide Emissions: The Tradable Permit System*, Report UNCATD/GID11, Geneva, 40.

Transboundary Water Resource Management: The Example of the Canada/US IJC (International Joint Commission)

see Transboundary Water Resource Management: The Example of the Canada/US IJC (International Joint Commission) (*Case study,* **Volume 4**)

Transferable Fisheries Quotas (TFQs)

Ross Shotton
Food and Agriculture Organization, Rome, Italy

Quotas are used in sheries management to control the amount of harvest. Thus, they can be used to prevent the biological productivity of a sh stock from being excessively reduced so that catches are less than they can otherwise be. Most commonly, quotas determine how much catch can be harvested during a shing season - in these cases, the quotas are used as a form of output control. Fisheries managers also use quotas in the form of limitations on the amount of shing that is permitted, e.g., number of traps, eet size or number of vessels permitted to operate

in the shery. These regulations are a form of input control. Commonly, both measures may be used in the same shery.

The amount of quota that may be caught - the total allowable catch (TAC) - is determined from stock assessments (the current amount of stock biomass) and analyses of the state of the stock, e.g., its spawning biomass - how much of the stock is mature and can reproduce - the age structure, and factors affecting its productivity, such as the oceanographic conditions. In traditionally managed sheries, once the quota is determined, shermen compete among themselves to maximize their share of the xed amount permitted. To be competitive in doing this, they increase the shing power of the vessels by replacing them with newer, larger, more expensive vessels and installing the most modern technology to aid shing. To pay for these, incentives to cheat are increased and the pro tability of the industry is reduced by the extra capital costs involved.

The solution that has developed over the last 20 years, at least in industrial sheries, is to assign to shermen, vessel owners or investors, rights to harvest a speci c amount of sh (in output controlled sheries), or to use a certain amount of shing gear (in input controlled sheries). Thus, once the operator is awarded his individual quota, he can then harvest it in the manner that he nds most economically ef cient. No longer is there any incentive to invest in a vessel that is larger than required to physically take the harvest; his individual quota assures him of his share.

Quotas, when assigned, may be granted to individuals or companies. In several management jurisdictions, they have been assigned to the vessel that has been responsible for past harvests (e.g., Alaska, Iceland and the UK). Once assigned, when the management regime permits, these harvest, or access, rights may be sold or leased to other sheries operators; hence, the term transferable.

WHY IS THERE INTEREST IN THIS FORM OF FISHERIES MANAGEMENT?

Granting of transferable harvest rights implies a form of property rights. Just as in other endeavors, when there is ownership, there is greater incentive to look after the resource. This is particularly so in fisheries because of the way in which transferable quota systems are characteristically designed. Harvest rights are usually assigned as a certain fraction of the TAC (Total Allowable Catch). If the resource is managed well, e.g., if nursery areas are protected and small fish are left to grow, the TAC can be increased, and an initial quota of 100 tons may be increased to 110 tons if there is a 10% increase in the TAC. Experience shows that this is changing the way fishermen view, and participate in, the management of their fisheries. Instead of reacting to government managers, they now participate in the decision-making and often fund activities that bring about an increase in stock abundance, and hence, in future TACs. As a consequence of their property rights, industry has an incentive to ensure that research and stock assessments are adequately funded (and has an incentive to minimize the costs of doing so).

WHAT ARE THE FEATURES OF TFQs?

Not all TFQ schemes are alike; neither are the biological, social, economic, cultural and political situations in which fisheries operate. However, one of the underlying characteristics is that to be able to transfer a catch right, it must be owned in the first place, i.e., it has the characteristics of property. In the case of fishing entitlements, these may be seen from the perspective of their durability, security, exclusivity and transferability.

Durable fishing rights are those that holders can possess for long periods; in New Zealand, fishermen hold their rights in perpetuity. In Namibia, the period varies between seven and 14 years, depending on a government assessment of the rights holder's success in contributing to national policy objectives. Clearly, more durable rights are worth more to the holder, though they provide the government with less exibility to change policy if they wish.

Exclusive fishing rights exclude others, who are not bound by the same rules of ownership, from participating in the fishery. Weakness in the exclusivity of quota rights may occur in a stock if it is shared with another country or migrates to the high seas where it is also fished. In a national context, weak exclusivity can arise where recreational fishermen share a commercially exploited stock or where two different sectors, e.g., artisanal and industrial fisheries, target the same stock. Usually recreational fishermen, though often individually restricted to daily limits, can collectively catch as much as they can harvest. Government policy may allocate the remaining to the commercial sector. Clearly, this lack of exclusivity affects the value of the commercial sector's harvesting rights.

Secure fishing rights are those that are not easily removed from a fisherman. Of course, unless constitutionally guaranteed, laws can always be changed to allow appropriation of the rights. In many countries, the constitution may specify that the government must act as a trustee on behalf of the people and it is impossible to grant ownership of fisheries rights to individuals. In these cases, access rights are commonly considered to be privileges. Noteworthy though, is the common practice that even where governments can legally appropriate fishing rights, they never do so without compensating the holders, even when there is no legal requirement to do so.

The right of quota transfer is the most important criterion in the use of property rights in fisheries management. Fishermen often hold licenses that provide a fair degree of exclusivity, durability, and security, but do not have the

right to sell them. It is the right of transferability that leads to most of the benefits that can be derived from this form of management. There are several aspects of transferability. Usually there is a minimum holding that can be transferred, e.g., 100 kg, though fishermen may need to own more than this to obtain a license to participate in the fishery. There may be a limit on the maximum amount of quota that a single owner may hold, e.g., from 0.5% in some Pacific Halibut fisheries to 35% in deepwater fisheries in New Zealand. There may be restrictions on the nationality of the rights owner – in New Zealand, only citizens or New Zealand owned companies can buy quota. In the US, the rule depends on the fishery. For surf clams, no such restrictions exist, but only US fishermen are entitled to physically harvest the clams. Quota holders may be able to lease their holdings to other operators. Alternatively, rules may require that the quota holder is present on the vessel that harvests the catch.

WHERE ARE TFQs USED?

TFQs have been in use for about two decades. In New Zealand, they were introduced as a result of changes to national legislation, while in other management jurisdictions they have been implemented within the scope of existing national legislation, e.g., Iceland, Holland and Canada. In Canada, participants in several fisheries, with the support of the Department of Fisheries and Oceans, have created their own legal agreements through civil contracts, whereby fishers agree to abide by rules that, in effect, create property rights in the fisheries. In addition, property can only exist when it is respected and the associated rights are enforceable.

Most marine fisheries on the Pacific coast of Canada operate on an individual transferable quota (ITQ) basis, though salmon and herring are exceptions. Fisheries for whitefish in lakes on the Canadian prairies and many fisheries in the Maritimes of Canada also function using ITQs. On the east coast, many fisheries have mixed management, with different fleet sectors using different management methods. For example, the offshore scallop fleet has long used enterprise allocations but the inshore Bay of Fundy sector has used boat limits or competitive quota. In the US, the major fishery to use individual vessel quotas is that for sablefish and Pacific Halibut in Alaska. There is a surf clam and ocean quahog fishery based on a similar approach in the mid-Atlantic states, and a small fishery for wreckfish in the southern US that also uses transferable quota. Iceland started its individual fishing quota program with herring and then expanded it to include most important demersal species. Holland also started its transferable quota program for just one species group – plaice and sole – and then expanded to cod and related demersal fishes. The UK has a form of transferable quota management, run through national producers organizations, bodies that formerly were primarily concerned with marketing.

In the Southern Hemisphere, New Zealand operates a quota management system; ITQs have been issued for 42 species, or groups of species, managed as 257 separate fish stocks within this system. In Australia, nearly 100 fish stocks are managed using transferable quotas. Two of Australia's most important fisheries, that of the Western Australian lobster fishery and the Northern Prawn Fishery (NPF), are managed using input controls. For the lobster fishery, the input units are individual traps (or pots). For the NPF, the tradable units were initially a function of vessel engine power and gross tonnage, but were later changed to units (12 inches–1 foot) (3.05 cm) of trawl headrope length. Chile has had some fisheries managed through an ITQ approach and Namibia has a rights-based program, which is unique in Africa, though its conditions of quota transferability are restricted relative to other such programs.

WHAT DO TFQs NOT DO?

TFQs do not solve all the problems of fisheries management. Where the right is to harvest a part of the TAC, the TAC must still be properly calculated and the decision on the level of the TAC must withstand risk-prone lobbying efforts to increase them. Thus, the administrative resource requirements (staff and funding) for data collection and analyses remain.

TFQs do not decrease the amount of administration that is required to manage a fishery; in fact, the contrary. Catches must be recorded against those who have the entitlement (either by ownership or lease); registries of quota owners, including those who have leased some or all of their entitlement must be as accurate as the records of a bank. In some cases, the TFQs can be securitized, are liable to liens and have been declared marital assets in divorce proceedings.

Provisions are often necessary to handle situations where fishermen have inadvertently caught more than their holding permits, and some or all of this excess must be debited against the following year's allowance. Where the fishery exploits several species simultaneously, fishermen may need to balance their quota holdings across all the species they have caught – and when the species mix taken does not reflect their holdings, trading is necessary to achieve the required balance.

TFQs are often tied to particular stocks or fishing areas, so enforcement is needed to ensure that catches are taken from the stock area of the specified rights. If there are seasonal closures or closed areas for conservation reasons, marine surveillance will be required, as with conventional forms of (quota) management.

TFQs reduce the incentives, and frequency of fishermen who cheat, as it is clear that those who are robbed are

the other fishermen themselves. In this way, peer pressure facilitates the success of the programs. However, TFQs do not abolish the free rider problem, and some fishermen or fishing companies may exceed their quota if they believe it can be done without detection and punishment. Because the usual limit on landings is an amount of fish, as with traditional forms of quota-controlled fisheries, incentives will remain for fishermen to discard the less valuable fish they catch (called hi-grading). These fish are usually smaller and often immature, so that in landing their quota fishermen may have caught significantly more than their entitlement. The result is that fishing mortality is higher than the management plan calls for. A measure that has attempted to resolve this problem is to require that all fish caught are landed, i.e., fishermen are not allowed to discard fish at sea. Another approach is to allow such fish within reason, to be sold at a lower price to the administrating authority.

WHAT ARE THE CONCERNS REGARDING TFQs?

While there are concerns that TFQ systems may not achieve their biological, resource-oriented objective, the main concerns relate to their social and community implications. Because one common policy objective of TFQs is to increase the economic efficiency of a fishery, not surprisingly, when quota entitlements are sold, the most profitable operators usually buy them. These new owners may wish to land and process this fish in different locations than before. If this process continues, communities may become impoverished as the economic activity associated with the fishing activity is moved elsewhere.

Another concern arises from the success of this form of management. Often, TFQ management systems are introduced to remedy severe structural problems in a fishery, e.g., too many fishing boats, excessive capitalization, lack of profitability and other deficiencies. To promote TFQ management and because of the industry conservatism, but more to the point, because of the absence of any value in the fishing rights arising from the unprofitable nature of the fishery, governments commonly gift the future rights to the fishermen. As the scheme starts to work – as intended – rights that had no major value when granted, may come to assume great value. Many find the gain by individuals of such windfall profits from the privatization of a public asset morally wrong.

Others are concerned that TFQs may prevent young fishermen from entering the industry, because as quota rapidly assumes great value, the capital entry requirements become beyond the reach of a new fisherman to secure. Consequently, the structure of the fishery may change from an owner-operated industry to one where fishermen are essentially salaried employees. This situation is viewed with great aversion in many traditional fishing communities.

Another concern relates to the benefits that society should expect in the way of resource rents from those who have gained ownership of a resource that was recently public property (albeit usually with little access for those not already owners of a fishing license). Should rights-holders pay society some rent for their entitlement beyond funding the management costs of the fishery? Or should they, as the rights-owners usually claim, simply pay their contribution to society through normal taxes on profits realized?

On an operational basis, many believe that the way these systems have been introduced in the past have not appropriately recognized the rights of the many, created through their past participation in the industry as workers. In many applications, the rights have been assigned to the vessel, i.e., implicitly to the vessel's owners. Many argue that crews too should participate in receiving rights. But, the complication in doing this is in reaching agreement on the basis by how this could be done, not least because the available records of employment may not permit such a procedure without excessive legal challenges to the reliability of the information that is to be used.

HOW ARE TFQs IMPLEMENTED?

One of the major concerns of those with the option of entering a TFQ management program is determining the share of the future TAC they will receive. The most common method has been to base the initial allocation on the past catches of the operator. Different management regimes have used different formulae, but usually an earlier fishing period is taken, e.g., the last five years, and then the option may be taken of selecting, e.g., the three best years in terms of catch taken. The total of the selected catches represents the number of shares in the fishery, which are then divided into the TAC to give a catch per share.

In other management jurisdictions, a participant's current investment in the fishery may be considered with, e.g., 30% of the new allocation based on past catch and 70% based on an assessed value of the operator's investments. Where rights are being assigned to a new fishery, and there are no claims to precedence, the administrator may choose to auction the rights to the highest bidder. If there are concerns about quota eeing an area, regulations may tie the ownership of the quota to those who are members of a particular community. Thus the term in these situations is called *community quota*. Once allocated, sale of the quota may be restricted for a specified period to ensure that the new owners appreciate the value of the asset they have acquired. But, what these variations mainly show is

the exibility that exists in the way TFQ systems can be designed and implemented.

PROSPECTS FOR THE 21ST CENTURY

TFQ systems of fisheries management increase, rather than decrease, the administrative requirements of fisheries management. Thus, to the extent that current fisheries management is constrained, owing to inadequate commitment of human and financial resources, then so will the potential for expansion of this form of management be constrained. A pessimistic view is that little expansion is likely of the existing club of countries benefiting from this form of management (Canada, Iceland, US, Australia, New Zealand, Holland, Namibia and perhaps, Chile and the UK) during the next one to two decades.

An optimistic view is that, in view of the undeniable benefits of TFQ management, fishermen, industry and other stakeholders will successfully pressure governments into allowing this form of management. This process will be facilitated as recognition of governments grows that granting of property rights in fisheries does not diminish their sovereign rights or responsibilities.

See also: Quotas in International Environmental Negotiations, Volume 4.

FURTHER READING

Arnason, R and Gissurason, H H (1999) *Individual Transferable Quotas in Theory and Practice*, University of Iceland Press, Reykjavik, 1–215.

Bader, H (1998) Who Has the Legal Right to Fish, *U. Alaska Sea Grant Coll. Prog. Mar. Advisory Bull. No. 49*, 1–24.

Crowley, B L (1996) *Taking Ownership. Property Rights and Fishery Management on the Atlantic Coast*, AIMS, 1–323.

Gimbel, K L (1994) *Limiting Access to Marine Fisheries. Keeping the Focus on Conservation*, Center for Marine Conservation and World Wildlife Federation USA, 1–316.

Hanna, S and Munasinghe, M, eds (1995) *Property Rights and the Environment: Social and Ecological Issues*, World Bank and Beijer Institute, Stockholm and Washington, DC.

Neher, P A, Arnason, R, and Mollett, N (1988) *Rights Based Fishing*, NATO ASI Series, Kluwer, 1–541.

Organization for Economic Co-operation and Development, (1993) *The Use of ITQ in Fisheries Management*, Paris.

Pálsson, G and Pétursdóttir, G, eds (1997) *The Social Implications of Quota Systems in Fisheries*, Tema Nord, Nordic Council of Ministers, 1–353.

Pétursdóttir, G (1997) *Property Rights in the Fishing Industry*, FRI, University of Iceland, 1–101.

Shotton, R, ed (2000) Use of property rights in fisheries management. *Proceedings of the FishRights99 Conference*. Fremantle, Western Australia, 11–19 November 1999. Minicourse lectures and Core Conference presentations. FAO. Tech. Rep. No. 404/1, FAO, Rome, 1–342.

Shotton, R, ed (2000) Use of property rights in fisheries management. *Proceedings of the FishRights99 Conference*. Fremantle, Western Australia, 11–19 November 1999. Workshop Presentations. FAO. Tech. Rep. No. 404/2, FAO, Rome, 1–462.

Trends in Environmental Management in the Last 40 Years

see Trends in Environmental Management in the Last 40 Years (*Introductory essay*, **Volume 4**)

Tyler Prize

see International Environmental Prizes (**Volume 4**)

U

ULYSSES (Urban Lifestyles, Sustainability, and Integrated Environmental Assessment)

ULYSSES has been an innovative European research program (commissioned by EC DG XII, part of the Fourth Framework program), which has developed a procedure for citizen participation in integrated assessment (IA) (*see* Integrated Assessment, Definition of, Volume 4). The procedure comprises a process based on the *focus group* method. A group of citizens debates climate policy and sustainable development, having been given access to state of the art IA models to support their debates.

Between 1996 and 1999 ULYSSES has conducted and analyzed group discussions with over 400 citizens in eight European cities. Participants approached the climate change issue from a broad range of criteria including fairness, justice, and personal values and beliefs, and they related climate change to their own experiences and lifestyles. Participants usually preferred strong reductions of energy use in the next decades to business-as-usual trends. They were often in favor of actions to be taken in order to reduce the risks of climate change, even if there were substantial scientific uncertainties. The issue of trust in science and policy communities is essential in this context.

ULYSSES has demonstrated that IA–Focus Groups can increase the possibilities for citizens to articulate their views on environmental issues in a thoughtful manner. This can help to bridge gaps between decisions by policy makers and citizens' views. Exploring alternative explanatory and moral frames in such processes can complement expert assessments on complex environmental issues. Focus group procedures involving open-ended discussions add a richer dimension to the traditional quantitative survey methods. In IA–Focus Groups, the discussions can broaden into directions impossible for the researchers to imagine beforehand. This is vital if we are to learn more about how the diversity of everyday life and experience affects the understanding, values, and objectives that different social groups bring to global change issues (*see also* Integrated Assessment, Volume 4).

JEROEN P VAN DER SLUIJS *The Netherlands*

UNCED (United Nations Conference on Environment and Development)

UNCED was held at Rio de Janeiro in June 1992; 20 years after the United Nations (UN) Conference on the Human Environment held at Stockholm. UNCED is also commonly known as the Rio or Earth summit/conference. The UNCED event re ected a growing concern among all nations about the deterioration of the global environment. Global problems such as climate change, ozone depletion, and the loss of biological diversity had become significant issues in the public conscience, and this was gradually re ected in interest at the political level. The importance of integrating environmental and economic objectives into public policy had also gained prominence. The growing use of the terms *sustainable development* and *sustainability* is, in part, a re ection of this realization. The UNCED was intended to recognize the importance of a healthy environment to sustainable economic and social development.

UNCED was the culmination of several years of negotiation which led to the signing of the non-binding UNCED (Rio) Declaration, Agenda 21, and the Statement on Forest Principles. Two formal binding treaties, the Framework Convention on Climate Change and the Convention on Biological Diversity, were negotiated outside the preparatory process for the Rio Conference, but signed during the Rio Conference. The Conference involved representatives of UN member nations and a large number of non-government organizations (NGOs) concerned with environment and conservation. In addition, parallel NGO events were held to critique the UNCED. The Rio Declaration lists 27 principles

of sustainability. Several themes are particularly worth noting: the principle of common but differentiated responsibility for environmental degradation and improvement; the importance of integrating economy and environment in the development of policy and strategies; attention to private and intellectual property rights (including technology transfers); and the need for new and innovative approaches to environmental regulation (including economic instruments). Though signatory nations recognize the Declaration's Principles, adoption into national policy and legislation is not legally binding.

See also: Agenda 21, Volume 4; Common but Differentiated Responsibility Principle (Stockholm/Rio), Volume 4; Convention on Biological Diversity, Volume 4; United Nations Conferences on the Environment, Volume 4; United Nations Framework Convention on Climate Change and Kyoto Protocol, Volume 4; Sustainability: Human, Social, Economic and Environmental, Volume 5.

KEVIN S HANNA *Canada*

UNCHS (United Nations Centre for Human Settlements)

The UNCHS (Habitat) was established in 1978, 2 years after the United Nations (UN) Conference on Human Settlements in Vancouver, Canada. Based in Nairobi, Kenya, UNCHS (Habitat) is the lead agency within the UN system for coordinating activities in the field of human settlements.

Currently the structure of the center comprises:

1. A governing body called the UN Commission on Human Settlements composed of 53 governments elected by the general assembly of the UN from its membership on a rotational basis and representing the five UN recognized regional groups: Africa, Asia and the Pacific, East Europe, Latin America and the Caribbean, West Asia and West Europe and others.
2. A secretariat headed by an executive director with the rank of under secretary general of the UN and appointed by the secretary general.

The center was the secretariat for the second UN Conference on Human Settlements (Habitat II), held in Istanbul, Turkey, in June 1996. This conference formulated the Habitat Agenda and the Istanbul Declaration which together form Habitat's Global Plan of Action. In this document governments commit themselves to the goals of adequate shelter for all and sustainable urban development. Habitat is the focal point for the implementation of these commitments.

Habitat's operational activities focus on the following priorities: promoting housing for all, improving urban governance, reducing urban poverty, improving the living environment and managing disaster mitigation and post-con ict rehabilitation. Acting as a catalyst in the mobilization of technical cooperation, the center is supporting the implementation of the Habitat Agenda at local, national and regional levels.

During the biennium 2000–2001, Habitat will undertake two significant global campaigns, one for secure tenure and the other on urban governance. The aim of both campaigns is to reduce urban poverty through policies, which emphasize equity, sustainability and social justice. Strategic and operational partnerships with different levels of government, local authorities, non-governmental and community based organizations, the private sector and agencies within the UN system are crucial to the success of these campaigns.

UNCHS (HABITAT) ACTIVITIES

The Urban Management Programme

This is a long-term technical cooperation program designed to strengthen the contribution that cities and towns in developing countries make towards human development. The program is a partnership of the international community, involving a broad range of actors in developing countries at the national, regional, municipal and community levels. UNCHS (Habitat) is the executing agency, the World Bank is an associated agency and the UN Development Programme provides core funding and overall monitoring.

The Sustainable Cities Programme

This is a joint program of UNCHS (Habitat) and the United Nations Environment Programme (UNEP) dealing with the urban environment; uncollected garbage, poor drainage and impassable roads, unsafe water supply and inadequate infrastructure, polluted rivers and air. The program supports urban environmental planning and management at city and national levels through system-wide capacity building and networking. More than 20 cities are now primary participants in this program with several thousand professionals directly and regularly associated with it in over 40 cities and 30 partner organizations around the world.

Localising the Agenda 21 Programme

UNCHS (Habitat) is the UN task manager for the human settlements aspects of Agenda 21, adopted by the United

Nations Conference on Environment and Development (UNCED) in Rio de Janeiro in 1992. To reinforce the process of localization, this project offers a multi-layered support system to selected medium-sized cities in Kenya, Morocco and Vietnam.

The Safer Cities Programme

This joint Habitat–UNEP program was launched in 1996 at the request of African cities who want to address urban violence by developing a prevention strategy at city level. The objective is to establish a community-wide consultation process to reduce crime and address its causes through the development of crime prevention measures, methods and strategies.

The Local Leadership and Management Training Programme

This program supports national capacity-building institutions to provide training in local leadership and management to local government elected officials and other urban managers. Some of the program's main activities include preparing innovative training manuals and management tools as well as the development of assessment needs methodologies.

The Women and Habitat Programme

The objective of this program is to empower women within human settlements and urban development by improving their participation and decision-making power at all levels. The program works on capacity building, monitoring and research, particularly in the role of women in leadership and governance, management and negotiation, construction skills and finance for urban development.

The Community Development Programme

This is designed to build capacity among urban managers, practitioners and community organizations to improve the physical, social and economic infrastructure in low-income settlements through participatory concepts and approaches.

The Global Urban Observatory (GUO)

The GUO monitors the implementation of the Habitat Agenda through the development and application of policy-oriented urban indicators (The Urban Indicators Programme) and the sharing of lessons learnt from best practices at local and national levels (The Best Practices Programme). Urban indicators cover essential issues such as poverty, employment, infrastructure and transport, environmental management, local government performance and housing. The Best Practices database allows users to identify and access information, knowledge and expertise and provides solutions to some of the most pressing challenges facing an urbanizing world.

REGIONAL AND INFORMATION OFFICES

To improve the efficiency of in-country program preparation and management, as well as to raise the profile of Habitat's work worldwide, the center has established three regional offices, in Rio de Janeiro, Brazil (covering Latin America and the Caribbean), Fukuoka, Japan (covering Asia and the Pacific) and Nairobi (covering Africa and the Arab States). The center also has active information offices in Budapest, Geneva, New York, Beijing, Moscow and others.

WORLD HABITAT DAY AND PUBLIC AWARENESS

Through its publications, websites, exhibitions and public events, Habitat has played a leading role in raising global awareness about shelter and urbanization issues. World Habitat Day is celebrated on the first Monday of October every year and each year embraces a different theme. 1998s theme was Safer Cities; 1999s theme took on the idea of Cities for All.

Habitat produces two agship publications: *The Global Report on Human Settlements* provides a review of human settlements conditions, including their development and continuing evolution. *Habitat Debate* is the center's quarterly magazine, designed to be a forum for dialogue and debate on urban issues. It is disseminated to over 8000 readers worldwide, including non-governmental organizations, local authorities, governments, academic institutions, libraries and the media. It is also available on the Internet via the center's website http://habitat.unchs.org/home.htm.

ROBYN LYONS *Kenya*

UNCLOS (United Nations Convention on the Law of the Sea)

The Convention was adopted by the Third UNCLOS convened by General Assembly resolution 2750 C (XXV) of the 17th of December 1970. The conference worked for more than a decade before it was able to agree on a Convention that was acceptable to the majority of the participating

states. The Convention adopted by the conference was finally opened for signature on the 10th of December, 1982 at Montego Bay, Jamaica. On the 28th of July 1994 the General Assembly, by resolution 48/263, adopted the agreement relating to the implementation of the UNCLOS of the 10th of December 1982 (the 1994 Agreement). The 1994 Agreement is to be interpreted and applied together with Part XI (of the Convention) as a single instrument. Accordingly, UNCLOS refers to the 1982 Convention as modified by the 1994 Agreement.

UNCLOS establishes a comprehensive framework for the regulations of all ocean space. The 17 parts and nine annexes contain provisions governing, *inter alia*, the limits of the territorial sea, contiguous zone, exclusive economic zone and continental shelf of individual states; the right of access to the sea, freedom of navigation and other lawful uses of the sea in various maritime zones; exploitation, conservation and management of living resources of the sea; deep sea mining in the area beyond the limits of national jurisdiction; marine scientific research; protection and preservation of the marine environment; and the settlement of disputes. The provisions set out the rights and obligations of coastal and other states, including landlocked states, in the use and management of the seas and their resources.

UNCLOS establishes new international bodies to perform specific functions in connection with the implementation of its provision. Among these are the International Seabed Authority, which is responsible for regulation and control of deep-sea mining; the Commission on the Limits of the Continental Shelf; and the International Tribunal for the Law of the Sea for the peaceful settlement of disputes concerning the interpretation and application of the provisions of the Convention.

UNCLOS entered into force on the 16th of November 1994. (As on the 30 June 2001, there were 135 states parties to the Convention. This includes the European Union, which is a state party' in its own right, pursuant to Annex IX to the Convention.)

Further information on the Convention is available on the website of the United Nations: http://www.un.org/depts/los.

THOMAS A MENSAH *UK*

UNCTAD (United Nations Conference on Trade and Development)

The UNCTAD was established in 1964 as a body of the United Nations (UN) General Assembly. Its objective is to promote international trade, particularly between countries at different stages of development, in order to accelerate economic growth in developing countries. In light of their particular economic structures, most of these countries are very trade-dependent. UNCTAD provides policy advice and technical assistance to developing countries in the areas of institution-building and policy-making so that they may take full advantage of their enhanced participation in international trade and the global division of labor.

UNCTAD was founded at a time when many developing countries were in the process of gaining their political independence and thought their interests were underrepresented by other multilateral organizations, particularly by the General Agreement on Tariffs and Trade. In the first two decades of its existence, UNCTAD focused on the shaping of a new international economic order, the stabilization of commodity prices, and the achievement of fair terms of trade. More recently, it has shifted emphasis towards issues of globalization and structural reform, based on the new consensus that market-oriented economic policies, supplemented by active government support and founded on political pluralism, are the basis for fostering development.

UNCTAD is now recognized as a central forum for North–South dialogue. Conferences of UNCTAD are held every four or five years. It also conducts a number of analytical activities to advise developing countries in multilateral negotiations on liberalization of trade within the framework of the World Trade Organization (WTO).

UNCTAD increasingly views environmental issues as an integral part of the relationship between trade and development, as manifest in the important role it was assigned by Agenda 21 to make trade and environment mutually supportive in favor of sustainable development. To support this mission, UNCTAD undertakes analytical work on trade, environment and development, including in the field of competitiveness, market access, eco-labeling, multilateral environmental agreements, enabling measures, and trade liberalization and sustainable development. It also assists developing countries in the area of trade and environment through country studies, technical assistance programs and training to enhance the capacity of the public and private sectors of developing countries to address, in a participatory framework, trade and sustainable development issues.

Many of UNCTAD's research and training activities on trade and environment issues are coordinated with other international organizations, in particular the United Nations Environment Programme, the United Nations Development Program, the WTO and the Organization for Economic Cooperation and Development (OECD).

For further information, see http://www.unctad.org.

HUSSEIN ABAZA *Switzerland*

UNDP (United Nations Development Program)

UNDP, The United Nations Development Programme is part of the United Nations. Its mission is to help countries in their efforts to achieve sustainable human development by assisting them to build their capacity to design and carry out development programs in poverty eradication, employment creation and sustainable livelihoods, the empowerment of women and the protection and regeneration of the environment, giving first priority to poverty eradication.

UNDP also acts to help the United Nations family to become a unified and powerful force for sustainable human development and works to strengthen international cooperation for sustainable human development. UNDP, at the request of governments and in support of its areas of focus, assists in building capacity for good governance, popular participation, private and public sector development and growth with equity, stressing that national plans and priorities constitute the only viable frame of reference for the national programming of operational activities for development within the United Nations system.

UNDP acts through a unique network of 145 country offices in 140 countries and territories. Each of these offices is headed by a resident representative. UNDP resident representatives normally serve as resident coordinators of the operational activities of the United Nations system, supporting at the request of governments the coordination of development and humanitarian assistance. Resident coordinators also help to orchestrate the full intellectual and technical resources of the United Nations system in support of national development.

UNDP receives voluntary contributions from nearly every country in the world. It seeks to ensure a predictable ow of resources to support its programs. It provides grant funds through criteria on universality that strongly favor low-income countries, particularly the least developed.

MOSTAFA K TOLBA *Egypt*

UNEP (United Nations Environment Programme)

The establishment of UNEP was recommended by the Stockholm Conference. The recommendation was endorsed by the general assembly of the United Nations (UN) in or through its resolution 2997 (XXVII) of December 1972. Accordingly, four entities were established:

- a 58-nation governing council to act as the UN's intergovernmental organ for the environment;
- a small secretariat to provide a focal point for environmental action and coordination within the UN system;
- a voluntary environment fund to finance new environmental initiatives in whole or in part including those suggested by the Stockholm Action Plan, particularly integrated projects, and other activities agreed by the governing council;
- an Environment Coordination Board, to ensure cooperation among all UN bodies involved in environmental programs.

By resolution 3004 (XXVII) the general assembly decided to locate the small secretariat, in Nairobi, Kenya. UNEP was thus the first UN agency to be based in a developing country. It also has offices and UNEP-administered convention secretariats in Bangkok, Bonn, Geneva, Manama (Bahrain), Mexico City, Nairobi, New York, Washington, DC, Osaka, Paris, Montreal, The Hague, Athens, Kingston (Jamaica) and other cities.

UNEP's governing bodies include:

- the governing council composed of 58 governments, from five geographical regions, elected by the general assembly for four-year terms;
- the committee of permanent representatives;
- the high-level committee of ministers and officials.

UNEP works closely with UN agencies, international scientific and professional communities, non-governmental organizations, civil society and, in particular, governments, to ensure that important environmental issues facing the world are properly addressed.

UNEP's activities cover a wide range of environmental issues and environmentally related problems and sustainable development. Its present program of work focuses on five areas:

- environmental information, assessment and research, including environmental emergency response capacity and strengthening of early warning and assessment functions;
- enhanced coordination of environmental conventions and development of environmental policy instruments;
- fresh water;
- industry and technology transfer;
- support to Africa.

UNEP's functional programs are:

- environmental early warning and assessment;
- environmental policy development and law;
- environmental policy implementation;
- technology, industry and economics;
- regional cooperation and representation;

- cooperation with environmental conventions;
- communications and public information.

The resources available to UNEP for the administration and implementation of its program come from four distinct sources:

- the regular budget of the UN;
- the Environment Fund;
- trust funds;
- counterpart contributions.

For further information, contact United Nations Environment Programme, PO Box 30552, Nairobi, Kenya, http://www.unep.org.

NAOMI POULTON *Kenya*

UNESCO (United Nations Educational Scientific and Cultural Organization) and the Environment

Although the concept had not been brought forward when the UNESCO was set up in 1946, a number of activities clearly focusing on the environment were introduced from the beginning in the programs of this new agency. One of these first steps was the convening by UNESCO in October 1948, jointly with the French government in Fontainebleau, of a conference for the creation of what is today known as the World Conservation Union (IUCN).

A particularly important effort took place, from 1948 onwards, aiming at the scientific study of the problems of the arid lands of the world. This research program was placed under the guidance of an expert advisory committee, which held its first session in Algiers in 1951. The last session, held in Jodhpur (India) in 1964, marked the end of this long-term venture which had considerable in uence on the future environmental activities of UNESCO, focusing on world wide cooperation of specialists, interdisciplinary, problem oriented research and capacity building.

While many countries at the end of the 1950s were gaining their independence and were looking for economic development, based in particular on better knowledge of their natural resources, UNESCO was invited to extend its scientific programs into other areas. This included the humid tropics, with the launching of a modest international program in 1957, and the vast domain of marine sciences, with a first major operation, the International Indian Ocean Expedition (1959–1965), and the establishment of the Intergovernmental Oceanographic Commission (IOC) (*see* IOC (Intergovernmental Oceanographic Commission), Volume 4).

UNESCO is an intergovernmental organization and it appeared at this period that certain scientific issues related to the environment and its natural resources, called for an intergovernmental cooperative approach because of their geographic nature in the context of national sovereignty.

Thus, following a US proposal, UNESCO launched in 1965 a 10 year worldwide effort to understand and analyze the Earth's water cycle, to evaluate surface and ground water resources, to establish a basis for their sound management, to train the appropriate specialists and to enhance the place of hydrology among scientific disciplines. During this International Hydrological Decade (IHD) (*see* IHP (International Hydrological Program), Volume 1) a common program of research and monitoring was defined and its coordination ensured by a council, where a limited number of countries were represented at expert level, together with representatives from other United Nation (UN) or non-governmental organizations (NGOs) concerned. In all countries, a National IHD Committee was set up to arrange for the participation of the country in the world program.

The success of IHD, and the fact that water related issues were becoming more acute in most parts of the world, led to the establishment of a longer-term cooperation program. This International Hydrological Program (IHP) continues today, with successive four-year planning phases and national committees in some 140 countries. Particular stress is placed on concrete problems in water resources management. However, more purely scientific issues, including for instance the in uence of climatic change on these resources or the relations between hydrology and ecology, are also covered.

The development of UNESCO's environmental programs has always been closely linked to research on natural resources. While in the UN system scientific matters concerning the atmosphere are the responsibility of World Meteorological Organization (WMO), those concerning the hydrosphere, the lithosphere and the terrestrial biosphere have fallen largely within UNESCO's purview. This has given rise to an important production of thematic continental or world maps such as the *Soil Map of the World* (with Food And Agriculture Organization (FAO)), the continental climatic atlases (with WMO) and a host of maps on vegetation or on geology and related fields. The study of natural hazards (earthquakes, eruptions, oods, tsunamis, etc.) has been developed, particularly through emergency scientific missions. An International Geological Correlation Program has been in operation since 1973.

It is, however, in the vast area of interactions between human activities and the different eco-systems of the planet that UNESCO has pioneered a major innovative

program. In September 1968, it convened in Paris, a world conference on the scientific basis for the rational use and conservation of the resources of the biosphere. Four years before the UN Stockholm Conference on the Human Environment, this Biosphere Conference represented the first intergovernmental forum looking explicitly at reconciling environment and development, or in other words at promoting what would be called 20 years later sustainable development.

A major outcome of that Conference was the launching by UNESCO in 1970 of an intergovernmental and interdisciplinary program on Man and the Biosphere (MAB). This program was devised along the same principles as IHD, with a coordinating council and national committees, which operate today in 120 countries. However, the subject matter of MAB is more diversified since it deals practically with the entire spectrum of economic and social activities in the various eco-systems of the world, from tropical forests to urban systems, from coasts and islands to mountains. Until 1985, the different MAB projects attempted to cover exhaustively all domains. Since then, while these projects are maintained actively in many countries, a more selective approach has been followed.

One of the original MAB projects has emerged as a backbone for most other activities and is providing an exceptional tool for conservation of biodiversity as well as for ecological research and monitoring or for experimental management of resources towards sustainable development. This project concerns the establishment of a World Network of *Biosphere Reserves* (*see* Biosphere Reserves, Volume 2). Each Biosphere Reserve (there are at the moment 356 in 90 countries) consists schematically of a core area (terrestrial or coastal) legally protected for its biodiversity, surrounded by a well-defined buffer zone where only activities compatible with the conservation objectives of the core are permitted, itself included in a transition area where local communities are invited to participate in the management of the system and share in its economic and social benefits.

While a number of the existing Biosphere Reserves are not yet fulfilling adequately their multifunctional role, the network, which they constitute, and which is governed by a statutory framework calling for periodic review, is progressively improving. In 1995, an important conference held in Seville adopted a global strategy for Biosphere Reserves and formulated this statutory framework. Biosphere Reserves allow local communities to participate and benefit from conservation measures, thus ensuring long term protection of biodiversity, offering a tool for land use planning and constituting a large set of sites available for environmental research and monitoring on a variety of issues.

This brief review of natural and social science activities of UNESCO in the field of environment should not omit two paramount dimensions of its mission, namely education and culture. Environmental education has been for 25 years a major concern of the organization, where it has actively cooperated with United Nations Environment Programme (UNEP), including joint sponsorship of a series of world conferences (Tbilissi (1997), Moscow (1987), Toronto (1992)). Culture is associated with nature in the 1972 UNESCO Convention concerning the protection of the world heritage and many sites listed under this Convention show symbolically the link between these two fundamental poles of the human destiny.

MICHEL BATISSE *France*

UNESCO MAB

see UNESCO (United Nations Educational Scientific and Cultural Organization) and the Environment (**Volume 4**)

UNESCO Man and the Biosphere (UNESCO-MAB)

see UNESCO (United Nations Educational Scientific and Cultural Organization) and the Environment (**Volume 4**)

UNFPA (United Nations Population Fund)

UNFPA, the world's largest internationally financed source of assistance regarding population issues, became operational in 1969. The fund concentrates its programs and resources in three priority areas: reproductive and sexual health, including family planning; population and development strategies; as well as advocacy. It seeks to ensure universal access to reproductive health services to all couples and individuals by the year 2015, support population and development strategies that enable capacity building in population programming, promote awareness of population and development issues, and advocate for the political will and mobilization of resources necessary to accomplish its mission. It is guided by, and promotes, the principles of the Program of Action adopted by 179 countries and entities at the International Conference on Population and

Development (ICPD, 1994). The fund currently extends assistance to some 145 developing countries and those with economies in transition, to enable them to address population and development concerns in the context of their respective priorities and cultures.

UNFPA began its mission to promote awareness of population concerns, and of the social, economic and environmental implications of national and international population problems in both developed and developing countries at a time when understanding of the linkage between population, socio-economic development and the environment was limited. Recognizing the complex relationship between population and the environment, and aware that meeting the basic human needs of growing populations is dependent on a healthy environment, UNFPA played an active role in all the recent United Nations (UN) conferences, including the UN Conference on Environment and Development (Earth Summit) held in Rio de Janeiro in 1992 and the World Food Summit held in Rome in 1996. At each of these Conferences, UNFRA advocated the integration of population, development and environment issues, pointing out that such factors as population, health, education, poverty, patterns of production and consumption, the empowerment of women, and the environment are closely interconnected and should be considered through an integrated approach. Consistent with Agenda 21 adopted by the Earth Summit and the ICPD Program of Action, the fund seeks to promote the eradication of poverty and the reduction of both unsustainable production and consumption patterns, as well as negative impacts of demographic factors on the environment in order to meet the needs of current generations without compromising the ability of future generations to meet their own needs.

As part of the follow-up activities to these major Conferences, UNFPA seeks to ensure that population, development and environment concerns are adequately addressed in all its program activities. UNFPA assists countries to incorporate population, development and environmental factors in planning and decision-making at all levels, and integrate these dimensions in national development policies and programs. UNFPA promotes cooperation and collaboration among UN system organizations, bilateral agencies, governments, non-governmental organizations, and the private sector in addressing issues of population, development and the environment. It works with a broad range of partners, including civil society, to promote sustained and sustainable social and economic development that meets human needs, ensures well-being and protects the natural resources upon which life depends.

UNFPA has funded research and analysis on the relationship between population pressures, poverty and environmental degradation; the population dimension of the food supply system, including the causes of food insecurity and imbalances in supply and distribution. Support has also been provided for research on the population dimension of desertification; population factors related to water-scarce and water-short regions, including effective strategies to conserve and utilize the resource in a sustainable manner; the causes and consequences of environmental refugees; and the roles of women and men as resource stewards.

UNFPA has prepared a guidebook on population, resources and environment in collaboration with the World Conservation Union (IUCN) and has provided support for activities with the United Nations Environmental Programme (UNEP). UNFPA has worked with the Food and Agriculture Organization of the UN and the International Institute for Applied Systems Analysis (IIASA) on the preparation of a major report on the carrying capacity of agricultural land in the developing world. UNFPA also published an advocacy booklet entitled, *Food for the Future: Women, Population and Food Security*, as input to the World Food Summit. It co-organized, together with UNEP and the IUCN, an International Workshop on Population–Poverty–Environment Linkages to review case studies from various regions that would serve to identify the nature of these linkages, as well as the policy and the programmatic responses that could address them.

To promote awareness of the linkages between population, development and the environment, UNFPA, together with the International Planned Parenthood Federation, the Swedish International Development Cooperation Agency, the World Wide Fund for Nature and the IUCN, publishes a joint quarterly publication entitled, *People and the Planet*. This publication informs a wider audience on how demographic factors, combined with poverty and lack of access to resources in some areas and excessive consumption and wasteful production patterns in others, can exacerbate environmental degradation and resource depletion.

UNFPA has recently signed a Memorandum of Understanding with UNEP, the leading global environmental agency that sets the global environment agenda, to ensure that both population and environmental factors are integrated into policy and program planning, development, monitoring and implementation. Areas of collaboration include technical guidance and research, advocacy, public awareness, education and training, and strengthened coordination.

UNFPA, which became operational in 1969, is a subsidiary organ of the UN General Assembly and has as its governing body the Executive Board of the UN Development Program and of UNFPA. As of the 1st January 2001, the total number of staff was about 840. These included some 200 at headquarters in New York. The rest are distributed between Geneva and various field offices. The percentage of women on UNFPA's professional staff at headquarters and in the field was 50%, one of the highest percentages among UN agencies and organizations.

Incumbents to the post of executive director, held by Thoraya Obaid since 1 January 2001, are appointed by the UN Secretary General.

Among the most important publications issued by the UNFPA are:

- *POPULI*, the UNFPA magazine (two issues a year);
- *The State of World Population* (annual).

Contact Information: United Nations Population Fund (UNFPA), 220 East 42nd Street, New York, NY 10017-5880, USA, Tel.: +1-212-297-5020, Fax: +1-212-557-6416, E-mail: ryanw@unfpa.org, Website: www.unfpa.org.

Contacts: Stirling Scruggs, Tel.: +1-212-297-5011, E-mail: scruggs@unfpa.org; Alexander Marshall, Tel.: +1-212-297-5020, E-mail: marshall@unfpa.org.

NAFIS SADIK *USA*

UNIDO (United Nations Industrial Development Organization)

UNIDO is the specialist United Nations (UN) agency helping developing countries and transition economies to pursue sustainable industrial development.

The UNIDO secretariat carries out programs and activities approved by the UNIDO's policy-making organs. UNIDO's current membership consists of 168 countries. They meet once every two years at the General Conference, which approves the budget and work program of UNIDO and appoints the director general. The Industrial Development Board (53 member states) reviews the implementation of the work program and the budget, which is prepared by a subsidiary body, the Program and Budget Committee (27 member states).

UNIDO headquarters are at the Vienna International Center in Vienna, Austria. UNIDO Offices in New York and Geneva maintain on-the-spot contacts with member states and the other UN organizations based there.

UNIDO field offices are maintained in the following countries:

- Africa: Cameroon, Cote d'Ivoire, Ethiopia, Ghana, Guinea, Kenya, Nigeria, Senegal, Tanzania, and Zambia;
- Arab countries: Algeria, Egypt, Lebanon and Sudan;
- Asia and the Pacific: China, India, Indonesia, Iran, Pakistan, and Vietnam;
- Latin America and the Caribbean: Argentina, Bolivia, Colombia, Honduras, Jamaica and Mexico.

UNIDO operates a network of Investment Promotion Service (IPS) offices which promote business contacts between industrialized and developing countries and economies in transition. The UNIDO IPS offices are financed by the countries in which they are located: Austria, Bahrain, Belgium, China, Egypt, France, Greece, India, Italy, Japan, Poland, Republic of Korea, Russian Federation and the United Kingdom.

UNIDO's financial resources come from regular and operational budgets, as well as contributions for technical cooperation activities. Derived from member states' assessed contributions, the regular budget of UNIDO for 1998–1999 amounted to about US$130 million.

UNIDO's new integrated package of services was developed and approved by its member states. The 16 service modules were designed to meet specific needs and requests of developing countries and economies in transition. Once adapted and customized to the specific needs of individual countries, these services promote the three Es of the UNIDO's Business Plan: competitive economy, productive employment and sound environment.

In the areas of sound environment, UNIDO concentrates on:

- environmental policy framework;
- climate convention and the Kyoto Protocol;
- energy efficiency;
- rural energy development;
- cleaner production;
- pollution control and waste management;
- Montreal Protocol.

For further information, see http://www.unido.org/.

MOSTAFA K TOLBA *Egypt*

United Nations Centre for Human Settlements (UNCHS)

see UNCHS (United Nations Centre for Human Settlements) (**Volume 4**)

United Nations Conference on Environment and Development (UNCED)

see UNCED (United Nations Conference on Environment and Development) (**Volume 4**)

United Nations Conference on Trade and Development (UNCTAD)

see UNCTAD (United Nations Conference on Trade and Development) (**Volume 4**)

United Nations Conferences on the Environment

Peter M Haas
University of Massachusetts, Amherst, MA, USA

Global United Nations (UN) conferences on the environment are an institutional innovation of the 1970s. With mounting concern about the degradation of the physical environment, governments approached the UN to convene a number of global conferences to address the host of human activities with transboundary and global environmental consequences. These environmental conferences were part of a broader effort at global problem-solving for a new class of global problems associated with international interdependence. As global interdependence became increasingly politicized in the 1970s, the UN system turned to global conferences as a way to highlight the interconnections between issues that had previously been treated in isolation. The topics of the global conferences were new to the international agenda, as previous multilateral conferences had principally addressed international economic topics. The UN, as the only venue with global participation, was the logical forum for such meetings to include all the affected countries.

The 1972 UN Conference on the Human Environment, and the 1992 UN Conference on Environment and Development directly addressed the subject of environmental protection, but UN Conferences devoted to specialized aspects of human impacts on the environment were commonplace in the 1970s. Conferences were convened to address climate, food, population, water, deserti cation, and other instances of global environmental threats. The frequency with which such global conferences were held diminished in the 1980s and 1990s. Other than the decadal meetings of conferences on population, women, and food, the follow-up annual reviews to the United Nations Conference on Environment and Development (UNCED) of the UNCED commitments, and the more comprehensive and high pro le UNCED + 5 meeting in 1997 and UNCED + 10 in 2002, states have been less willing to support specialized UN global conferences.

These global conferences were intended to mobilize concern about new problems, and to coordinate national actions to study and monitor environmental quality and human activities with environmental consequences. They also were intended to develop joint measures to prevent various sources of environmental degradation and attenuate the effects of human actions on the environment. Economic and equity concerns cut across most of the other specialized conferences.

A full list of global environmental conferences is presented in Table 1, of Global Environmental Conferences, along with their formal results.

Typically, global UN conferences on the environment last for several weeks, with high level diplomatic attendance for the last two or three days to overcome political deadlocks and to sign legally binding resolutions and other commitments developed at the Conference. Decisions are generally reached by consensus, so negotiations are slow. Preceding the Conferences are often several rounds of preparatory meetings (Prep Coms), often spread over one or two years, at which national delegations are presented with background papers and preliminary negotiations are conducted on the documents intended to be approved at the Conferences themselves. Most of the arduous work of reconciling political differences occurs during the preparatory conferences.

Generally the global UN conferences on the environment have produced declarations and action plans. The most successful conferences endorsed new policy doctrines and policy targets for the international community, and also authorized the creation of new international organizations, approved legal commitments and generated new financial resources. Others have failed to spark international concern or to catalyze robust international commitments and action, and have remained virtual dead letters, such as the 1977 Desertification Conference, the 1979 Conference on Science & Technology for Development, and the Conferences on Human Settlements. The most successful, in terms of their administrative accomplishments, have been the 1972 UNCHE, the 1974 World Food Conference, 1992 UNCED, and the 1994 International Conference on Population and Development (Weiss and Jordan, 1976).

The 1972 UNCHE, held in Stockholm, Sweden was the first major global environmental conference. Sponsored by the UN, it convened 113 countries to discuss contemporary environmental issues. UNCHE adopted the Stockholm Declaration, establishing 26 principles of behavior and responsibility to serve as the basis for future legally binding multilateral accords; the Action Plan for the Human Environment that specified 109 recommendations in the areas of environmental assessment, environmental management, and supporting institutional measures (Schmidt,

Table 1 Global Environmental Conferences[a]

Year	Name/location	Product/outcome
1972	United Nations Conference on the Human Environment (UNCHE) (Stockholm)	Declaration of Principles Action Plan UN Environment Programme
1974	World Food Conference (Rome)	Universal Declaration on the Eradication of Hunger and Malnutrition World Food Council International Fund for Agricultural Development
1974	World Population Conference (Bucharest)	World Population Plan of Action
1977	UN Water Conference (Mar del Plata)	International Drinking Water Supply and Sanitation Decade (1981–1991)
1977	UN Conference on Desertification (Nairobi)	Plan of Action to Combat Desertification
1978	UN Conference on Human Settlements (Vancouver)	UN Center for Human Settlements Global Strategy for Shelter to the Year 2000
1979	UN Conference on Science and Technology for Development (Vienna)	Vienna Program of Action on Science and Technology for Development
1979	Agrarian Reform and Development (Rome)	
1979	World Climate Conference (Geneva)	
1981	UN Conference on New and Renewable Sources of Energy (Nairobi)	Nairobi Program of Action for the Development and Utilization of New and Renewable Sources of Energy
1984	World Conference on Agrarian Reform and Rural Development	Program of Action on Agrarian Reform and Rural Development
1984	Second World Population Conference (Mexico City)	
1985	Third World Conference on Women (Nairobi)	
1990	Second World Climate Conference (Geneva)	
1992	UN Conference on Environment and Development (Rio de Janeiro)	Rio Declaration Agenda 21 Framework Convention on Climate Change Convention on Biodiversity Forest Principles UN Commission on Sustainable Development
1994	International Conference on Population & Development (Cairo)	Program of Action
1995	Fourth World Conference on Women (Beijing)	Beijing Declaration and Platform of Action
1996	Habitat II (Istanbul)	The Habitat Agenda and Istanbul Declaration on Human Settlements
1996	World Food Summit (Rome)	Rome Declaration on World Food Security and World Food Summit Plan of Action
1997	UN General Assembly Special session on Sustainable Development (UNCED + 5)	Towards the Further Implementation of Agenda 21

[a] *Source:* Fomerand (1996); Weiss *et al.* (1997) especially Chapter 9; Caldwell (1996). I am grateful to Zuhre Aksoy for help in compiling this table.

1973). Implementation was intended for governments and international organizations.

The 1992 UNCED, held in Rio de Janeiro, marks the high water mark of these institutional outputs. UNCED adopted the Rio Declaration, Agenda 21, and the Forestry Principles. The Convention on Climate Change and the Biodiversity Convention were adopted by their respective diplomatic conferences at the time of UNCED. The Rio Declaration contains 287 principles to guide subsequent action. Agenda 21 is a sweeping action plan to promote sustainability, with 2509 specific recommendations elements, applying to states, international institutions, and members of civil society. The Commission on Sustainable Development was created to ensure effective follow-up of UNCED; to enhance international cooperation and rationalize intergovernmental decision-making capacity; and examine progress in Agenda 21 implementation at the local, national, regional and international levels (Grubb *et al.*, 1993).

The variation in the degree of success of conferences is due to a number of factors. The more successful

conferences were free of profound political schisms or geopolitical tensions amongst major parties including Cold War tensions between the US and the USSR and profound North–South cleavages. The issue at hand was of immediate concern to the industrialized countries that would be willing to commit resources.

UNCHE, for instance, was held at a fortuitous moment. Domestic environmental movements were becoming particularly active in the US and Europe. Potentially profound North–South disagreements were avoided by prior high-level discussions that rejected the conceptual dichotomy between economic growth and environmental protection, extended the international agenda to include environmental concerns of the South, regarded natural resource policy as well as the pollution concerns of the industrialized countries, and provided a notional commitment to additional financial assistance on behalf of the North. Environmental protection was not seen as being inconsistent with other established goals in international negotiations, including national security and economic liberalization. UNCHE also provided the first opportunity for China to stake a position in international diplomacy following US recognition. North–South relations became more acrimonious with the New International Economic Order (NIEO) discussions in the late 1970s, and it proved harder to forge consensus at international conferences. Even with these factors, Cold War divides still modestly in uenced the Conference, as the Soviet Union and the Eastern bloc countries withdrew at the last minute over the participation of West Germany, yet because the superpowers were in a period of détente, such tactical linkages were not perceived as provocative with cold war overtones.

Effectiveness

It is difficult to evaluate the effectiveness of many of these conferences, in part due too weaknesses and gaps in our ability to monitor subsequent progress. The record is generally mixed, at best, in terms of achieving the targets and aspirations expressed in the action plans and declarations of the Conferences. It is difficult to directly measure effects on the environment, and the record of states in complying is mixed or uncertain. At UNCED + 5 the Commission on Sustainable Development tried to evaluate overall progress achieved since UNCED. It determined, amongst many observations, that production and consumption patterns had become more energy efficient in industrialized countries; land use con icts were more acute in developing countries between competing demands for agriculture, forest cover and urban uses; and that water scarcity remained a major threat to development and human health in developing countries (UNCED, 1997).

In short, it is difficult to evaluate the effectiveness of the conferences on states' policies and on observable environmental impacts. This is because it would be unreasonable to expect such conferences to yield lasting and clear effects on states' policies and on the environment. It is equally unreasonable to assign blame to conferences for failing to reverse environmental decline.

Ultimately, international conferences are modest institutional features of international relations and are unlikely to induce profound changes by themselves or exercise sustained in uence on states. Because of their one-off frequency, there is limited access to top-level officials, and it is difficulty to maintain long-term pressure on governments through national reporting, information circulation, oversight, or lobbying. Thus it is difficult for transnational policy networks to sustain consolidated political in uence on states through global environmental conferences. Such conferences generally lack significant political or financial resources for inducing change, and lack any lock-in mechanisms by which decisions become deeply institutionalized within the legal and political systems of attending countries.

Functions of Conference Diplomacy

Global environmental conferences also have a number of indirect effects with longer-term effects on national policies affecting the prospects for sustainable development. While you cannot directly stop human activities that degrade the environment through universal declarations or conferences, global UN conferences have served a number of distinctive political and social purposes that in uence governments' concern about the environment and their willingness to commit scarce political and financial resources to its protection. The effects of the most successful conferences have been to increase national concern, and to increase government capacity to politically and technically address problems by means of agenda setting, consciousness raising, expanding participation, monitoring, knowledge generation and diffusion, target setting, norm development and diffusion, and administrative reforms (Hass *et al.*, 1993; Reinecke and Deng, 2000; Fomerand, 1996).

Agenda Setting

Global environmental conferences can place new issues on the global agenda and galvanize national concern by publicizing new issues. The conferences often have the effect of reframing issues for decision-makers. For instance, environmental protection was firmly placed on the international agenda at the UNCHE conference, and the preliminary Founex meeting effectively reconciled North–South differences about the priority accorded to environmental considerations in economic planning, establishing the principal that the two goals could be compatible, especially with concessionaire finance from the North to pay for incremental pollution control costs in the developing countries. UNCHE also helped inform Northern governments of Southern countries' concern with an alternate

agenda that would supplement the North's primary focus on industrial pollution, waste management, and transboundary environmental threats: with a focus on resource deterioration, deforestation and water quality, and the underlying problems of insufficient money for sewage treatment and effective resource management.

The 1994 International Conference on Population and Development shifted the public debate on population issues to a focus on the underlying social and political and economic forces that in uence population growth. The Program of Action marked a distinctive shift in population policy towards cooperation to eradicate poverty; encouragement of universal access to health care services; and empowerment of women.

Popularizing Issues and Raising Consciousness

Conferences provide a brief window of opportunity for educating mass publics and government officials about environmental issues. Conferences spawn publicity about the declarations and statements of principles issued at them. Because many journalists attend the conferences, they provide an opportunity for NGOs and the media to publicize issues in their national media as well as for educating members of the media about environmental issues.

Generating New Information and Identifying New Challenges for Governments

Preparation for conferences often generates information for countries about their environmental problems, the array of policies available for addressing such issues, and the political coalitions organized around them. States are invited to submit national reports about conditions in their countries in advance of the conference, and this process can lead states to learn of new problems, clarify their recognition of their national interest, and identify the potential for political compromise. These reports are often synthesized by the secretariats for subsequent dissemination.

Early Warning of New Threats

Conferences help focus attention on new problems, and also help to identify institutional gaps and needs in addressing such problems. The *Assessment of the World Food Situation*, presented to 1984 World Food Conference, helped focus attention on the world food gap that threatened developing countries. UNCHE helped to identify the urgency of addressing land-based marine pollution, as well as identifying the institutional need of creating a global environmental monitoring system (GEMS) that subsequently became one of UNEP's core activities.

Galvanize Administrative Reform

Conferences also prompt governments to create or reform national bodies responsible for forms of environmental protection. National administrative bodies serve as the nodes of transnational environmental policy networks. At the time of UNCHE only 26 governments had administrative agencies responsible for environmental protection (15 in the developing countries, 11 in developing). The preparation for UNCHE led many governments to recognize the need for creating national environmental agencies. By 1982 the total number was up to 144 (34 in Developed, 110 in developing), UNCED led to the establishment of Sustainable Development committees and bodies in nearly 150 countries.

Adopt New Norms, Certify New Doctrinal Consensus and Set Global Standards

Global conferences are sites of doctrinal contests. UNCHE developed new principles of soft law that have been interpreted and applied by international lawyers to inform a generation of international environmental law. Specific programmatic action, such as the 2,509 specific proposals in Agenda 21 set the stage for legitimate responses to international conferences. The identification of numbers of people at risk from malnutrition, and targets for Official Development Assistance (ODA) and hunger reduction stipulated at World Food Conferences similarly established standards and aspirations for subsequent governmental practices.

Mass Involvement of New Concerns

International environmental conferences contribute to the participation of new concerns in international environmental politics through the invitation of new groups to attend international conferences. Environmental conferences have been leaders in the introduction of NGOs to international diplomacy. These meetings developed the practice of parallel NGO conferences to the governmental conferences, as well as allowing NGO participation as observers at the governmental conferences. Roughly 178 NGOs participated at UNCHE (Thompson Feraru, 1982). Over 1400 were represented at UNCED.

Despite the vast increase in numbers of NGOs attending international environmental conferences, the participation is still heavily tilted towards the North, where NGOs have greater financial support and are better able to find resources to attend conferences. At UNCED, 70% of the registered NGOs came from industrialized countries.

Conferences provide the potential for networking and developing transnational issue networks to coordinate international campaigns, and NGOs may subsequently provide information to governments and apply pressure on governments.

Conferences often invite participation from major nonstate groups, including NGOs, the transnational scientific community (*see* Environmental Movement the Rise of Non-government Organizations (NGOs), Volume 5), and, since UNCED, multinational corporations (Schmidheiny, 1992). Such groups are invited to attend expert group meetings in advance of the conference as well as parallel NGO

events and even as observers at governmental meetings. Participation is often, particularly in preliminary meetings, by umbrella groups such as GESAMP and ICSU, or the ICC.

Global environmental conferences may be deliberately designed to foster new coalitions more generally, and to build support for environmental protection at the national level by including the political in uence of transnational policy networks. Maurice Strong, the secretary general of UNCHE and UNCED coined the phrase the process is the policy to capture the idea that through conference diplomacy more concerns and perspectives could be introduced to international environmental policy making.

There is still wide variation in the extent of NGO in uence at conferences. The rules of participation remains set by states' decisions in ECOSOC and the organizations are continually constrained (if not totally hamstrung) by state choices to allocate resources and set rules of behavior for the organizational dealings with NGOs.

International Conferences and Governance

Global environmental conferences play a part in a broader shift in global environmental governance. It is now widely accepted that governance occurs in a decentralized manner, though loosely tied networks of multiple states, functional state agencies, and non-state concerns. The environment is no different.

Accumulated conferences over the last 30 years have contributed to an aggregate shift in international politics by extending participation and access to environmental diplomacy to national environmental agencies and to NGOs and networks of scientists: a process that Jacques Fomerand (1996) describes as a large-scale process of social mobilization . The most successful conferences have promoted broader processes of social learning and the construction of new, more comprehensive conceptual frameworks for global environmental governance through issue clarification, popularization of issues, and the introduction of new approaches to environmental policy making. Through this institutionalized process of participation and education, new environmental norms of environmental protection have been diffused, and participating states have been encouraged to endorse them and to apply them nationally.

Global environmental conferences have contributed to aggregate substantive changes in environmental governance. The Founex preparations for UNCHE contributed to transcending the environment/development dichotomy in the framing of international environmental policy. The focus of the environmental conferences has shifted from concern about resource scarcity and depletion to efforts to understand and protect ecosystem integrity. Scientific consensus has crystallized around comprehensive forms of ecological management doctrines such as Sustainable Development: and a new consensus was forged at the 1994 Population Conference in Cairo that population growth could not be considered in isolation from social issues shaping family planning choices, including women's roles in society. Jacques Fomerand (1996) in a comparative study of UN world conferences, concludes that Today it is recognized that population issues must be viewed within the framework of the close links between population, economic growth, sustainable development, and the promotion of women's condition in all its aspects as well as greater gender equality in general. (See also Sanwal, 1993.)

Later, UNCED's Agenda 21 was organized and designed around a matrix of issues so that policies would be developed to address the interconnections among human activities (i.e., industry, agriculture, styles of decision making, consumption patterns, and technology) and the environment, as well as amongst global ecosystems (e.g., the atmosphere, freshwater, oceans and land): with chapters of Agenda 21 designed to capture the intersections located in each cell of the matrix (UN, 1991). The earlier UNCHE framework had been organized around the more traditional tripartite administrative framework of environmental assessment (evaluation and review, research, monitoring, information exchange), environmental management (goal setting and planning, international consultation and agreements), and supporting measures (education and training, public information, organization, financing, technical cooperation) (UNEP, 1981). UNCHE had been organized around a conceptual framework of traditional administrative functions, whereas UNCED was set up to capture the newly appreciated analytic attributes of the issues to be addressed.

REFERENCES

Caldwell, L (1996) *International Environmental Policy*, Duke University Press, Durham, NC.

Fomerand, J (1996) UN Conferences; Media Events or Genuine Diplomacy, *Global Governance*, 2(3), 361–375.

Grubb, M, Koch, M, Thompson, K, Munson, A, and Sullivan, F (1993) *The Earth Summit Agreements a Guide and Assessment*, Earthscan and Royal Institute of International Affairs, London.

Haas, P M, Keohane, P O, and Levy, M A (1993) *Institutions for the Earth*, MIT Press, Cambridge, MA.

International Conference on Population and Development (1994) *Synthesis of National Reports on Population and Development*, UN Document A/49/489.

Reinecke, W H and Deng, F M (2000) *Critical Choices Toronto*, IDRC.

Sanwal, M (1993) Sustainable Development, the Rio Declaration, and Multilateral Cooperation *Colorado Journal of International Environmental Law and Policy*, 4(1), 45–68.

Schmidt, E (1973) *The Results from Stockholm*, Verlag, Berlin.

Schmidheiny, S (1992) *Changing Course* MIT Press, Cambridge MA.

Thompson Feraru, A (1981) Stockholm and Vancouver: The Role of ISPAs at UN Conferences, in *Knowledge and Power in a Global Society*, ed W M Evan, Sage Publications, London.
UN (1991) Structure and Organization of Agenda 21 A/CONF.151/Pc/42 9 July 1991.
UNCHE (1972) Report of the United Nations Conference on the Human Environment, A/Conf.48/14 3 July 1972, 23.
UNCED (1997) *Overall Progress Achieved Since the United Nations Conference on Environment and Development*, E/CN.17/1997/2.
UNEP (1981) *In Defense of the Earth*, UNEP, Nairobi.
Weiss, T G, Forsythe, D P, and Coate, R A (1997) *The United Nations and Changing World Politics*, 2nd edition, Westview Press, Boulder, CO.

United Nations Convention on the Law of the Sea (UNCLOS)

see UNCLOS (United Nations Convention on the Law of the Sea) (**Volume 4**)

United Nations Convention to Combat Desertification (UNCCD)

Hama Arba Diallo
Secretariat of the Convention to Combat Desertification, Bonn, Germany

The issue of deserti cation was rst addressed by the international community at the multilateral level, as a follow up to one of the greatest environmental catastrophes of the past decades: the unprecedented African drought that struck the Soudano- Sahelian region from 1968 to 1973. Within that context, the General Assembly of the United Nations (UN) convened in Nairobi in 1977, the rst UN Conference on this issue which adopted the UN Plan of Action to Combat Deserti cation.

Fifteen years later, the UN Conference on Environment and Development, held in Rio de Janeiro, con rmed the urgent necessity to tackle this major environmental issue and recommended therefore, within Chapter 12 of Agenda 21, the elaboration of an International Convention to Combat Deserti cation.

The UNCCD (United Nations Convention to Combat Desertification) in those countries experiencing serious drought and/or desertification, particularly in Africa was adopted on 17th June 1994 in Paris, France. On that occasion, it was signed by 114 countries and the European Community. Resolutions on Urgent Action for Africa and on other activities to be undertaken in the interim period, before the Convention came into force, were also adopted.

June 17th was subsequently declared the World Day to Combat Desertification and Drought by the General Assembly of the UN (Decision GA 49/115).

The Convention entered into force on the 26th of December 1996, upon the deposit of the 50th instrument of ratification. As of June 2001, the Conventions had 174 Parties. It is legally binding to its members and represents a major follow-up to the 1992 UN Conference on Environment and Development, which called Parties to build upon the sustainable development paradigm and to provide a balanced legal framework for increased international cooperation in combating desertification.

KEY PRINCIPLES OF THE CONVENTION

According to the Convention, desertification means land degradation in arid, semi-arid and dry sub-humid areas resulting from various factors, including climatic variations and human activities (Article 1.a).

Desertification is recognized as a major global threat, which puts at risk the livelihoods of over one billion people in more than 110 countries, particularly in Africa. Causes include overcultivation, overgrazing, deforestation and poor irrigation practices.

Desertification does not affect developing countries alone. North America and European countries also suffer from it. Countries which are not directly affected also bear some of its consequences, for desertification contributes to poverty, food insecurity, political unrest and the resulting migrations.

In this framework, the Convention represents an innovative document with the following characteristics:

- it gives priority to Africa;
- it builds on a democratic, bottom-up approach, emphasizing popular participation and the creation, both nationally and internationally, of enabling environments for local populations to help themselves reverse land degradation;
- it takes a long-term approach that fully integrates the physical and biological dimensions of desertification with social and economic realities, while giving the latter full weight;
- it contains precise commitments, detailed in the regional implementation annexes, for the preparation by affected country Parties of National Action

Programs, fully integrated with national policies for sustainable development, as well as Sub-Regional Action Programs and Regional Action Programs;
- it provides for consultative processes leading to partnership agreements spelling out the contributions of both affected country Parties and the international community to these programs; in particular, developed-country Parties are called to actively support, individually or jointly, the efforts of affected developing-country Parties, particularly those in Africa and the least-developed countries;
- it gives an unprecedented role to non-governmental organizations in the preparation and implementation of Action Programs.

THE STRUCTURE OF THE CONVENTION

The objectives of the Convention are to combat desertification and mitigate the effects of drought, through effective actions at all levels, supported by international cooperation and partnership arrangements, in the framework of a full participatory and integrated approach which is consistent with Agenda 21, with a view to contributing to the achievements of sustainable development in affected areas.

The decision-making body of the Convention is the Conference of the Parties (COP), which shall keep under review and evaluate implementation of the Convention, harmonize policies, establish subsidiary bodies, and undertake additional actions. The COP met in Rome, Italy (October 1997), then in Dakar, Senegal (December 1998), Recife, Brazil (November 1999) in Bonn, Germany (December 2000) and will meet in Genera, Switzerland in its Fifth Session (October 2000).

The Permanent Secretariat of the UNCCD is established to serve the COP. The Secretariat also facilitates assistance to affected developing-country Parties in their consultative process and compilation of information required under the Convention, and promotes awareness raising on the issue of Desertification. It has been based in Bonn, Germany, since February 1999.

A Committee on Science and Technology (CST) appointed by the Parties was established under the Convention as a subsidiary body. The Committee provides the COP with information and advice on scientific and technological matters relating to combating desertification and the effects of drought and meets in conjunction with the ordinary sessions of the COP. The COP may, as necessary, appoint *ad hoc* panels to provide information and advice through the Committee on specific issues. The COP has established an *ad hoc* group on benchmarks and indicators, on traditional knowledge on early-warning systems and on survey and evaluation. It has also indicated that early warning systems in the broadest sense of the term are one of the issues for consideration of the CST.

The Convention contains four regional implementation Annexes – for Africa, Asia, Latin America and the Caribbean, and the Northern Mediterranean and a fifth annex has been adopted at COP4 to cover Central and Eastern Europe and should enter into force in September.

The African Annex is the most elaborate. It comprises 19 articles, and addresses a broad range of issues, including commitments and obligations of both African and developed-country Parties. Explicit reference is made with regard to technical assistance and cooperation to ensure that preference is given to the utilization of less costly local experts. Emphasis is placed on the need for increased coordination among the key players involved in desertification activities, including donors, national governments, non-governmental organizations and local populations. The annex also contains provisions for financial mechanisms and resources, and coordination, partnership and follow-up arrangements.

The other annexes are shorter and re ect the different priorities of the regions:

- The Asian annex is relatively general in scope. It foresees cooperation and coordination activities among country Parties, including networking on specific areas of interest such as the promotion of scientific, technical, technological and financial cooperation.
- The Latin American and Caribbean Annex mentions the important links between desertification and loss of biological diversity, as well as debt issues, unfavorable international economic trade practices, and other socio-economic factors. It also emphasizes traditional knowledge, know-how and practices.
- The Northern Mediterranean Annex is more scientifically oriented. It stresses urbanization and agricultural practices as economic causes of desertification and provides for collaboration with other regions in the preparation and implementation of action programs. It is also unique in that it clearly disqualifies the region from eligibility for funds raised through the Convention.

The Convention foresees the establishment of a Global Mechanism (GM) to promote actions leading to the mobilization and channeling of substantial financial resources, including money for the transfer of technology, to affected developing-country Parties.

On the occasion of the first COP, the Parties agreed to invite the International Fund for Agricultural Development to house the GM in Rome, Italy, and to cooperate with the World Bank and UN Development Program in this framework.

More information on the Convention can be obtained from the UNCCD web site at http://www.unccd.int.

See also: Deserts, Volume 1; Desertification, Volume 3.

United Nations Development Program (UNDP)

see UNDP (United Nations Development Program) (**Volume 4**)

United Nations Educational Scientific and Cultural Organization (UNESCO) and the Environment

see UNESCO (United Nations Educational Scientific and Cultural Organization) and the Environment (**Volume 4**)

United Nations Environment Programme (UNEP)

see UNEP (United Nations Environment Programme) (**Volume 4**)

United Nations Environment Programme (UNEP) Regional Seas Programme (RSP)

Stjepan Keckes
Borik, Rovinj, Croatia

The UNEP RSP, initiated in 1974, consists of a series of regional action plans adopted by high-level intergovernmental conferences. The action plans are designed to protect and develop the marine and coastal environment and their resources through coordinated joint efforts of the states participating in the plans. In a number of regions, the action plans are associated with regional conventions that provide their legal frameworks. Presently, the RSP is UNEP's major program for the implementation of Chapter 17 of Agenda 21 and other global programs and conventions relevant to the marine and coastal environment.

Nine action plans have been developed under UNEP's auspices and adopted, as part of the UNEP RSP, for the following geographic regions: the Mediterranean (1975), Persian/Arabian Gulf (1978), Wider Caribbean (1981), East Asian Seas (1981), West and Central Africa (1981), South-East Pacific (1982), South Pacific (1982), Eastern Africa (1985) and North-West Pacific (1994). The development of action plans for the North-East Pacific and the Upper South-West Atlantic is currently underway. Seven of these action plans are supported by legally binding regional conventions adopted in Barcelona (1976), Kuwait (1978), Abidjan (1981), Lima (1981), Cartagena (1983), Nairobi (1985) and Noumea (1986).

Regional programs adopted for the Red Sea and Gulf of Aden (1976), the Black Sea (1993) and the South Asian Seas (1995), or being developed for the Caspian Sea, albeit developed without major inputs from UNEP, are closely associated with and considered also as parts of the RSP.

All adopted conventions are in force. The conventions are of framework type and are supplemented by protocols dealing with specific issues, such as cooperation in cases of pollution emergency, control of pollution from dumping and land-based sources, protection of special geographic areas and biodiversity.

Most action plans are structured in a similar way, although the specific activities for a given region re ect the priorities and needs as defined by the governments of each region. Typically they are related to:

- environmental assessment, i.e., identification of the causes and sources of environmental problems, their magnitude and effects on the region;
- environmental management, focusing on integrated coastal zone management and control of land-based activities affecting the marine and coastal environment;
- environmental legislation through development of regional conventions and specific protocols, as well as development of national legislation in conformity with regionally and globally agreed norms;
- institutional arrangements for the coordination of activities agreed to be carried out under the action plans; and
- financial arrangements supporting the agreed activities; in most regions special trust funds are established to which governments make annual contributions.

Strengthening national institutional capabilities through training of national experts and technical cooperation, transfer of information and technologies and raising public awareness are important elements of all action plans.

While the focus of the first action plans was on the protection of the marine environment from pollution, the subsequently adopted action plans shifted their priorities to all issues relevant to the development and protection of the marine environments and their resources. The periodic

revisions of the action plans broadened their scope in order to emphasize issues related to integrated management and use of coastal and marine environment along the lines recommended by Agenda 21 adopted at the United Nations (UN) Conference on Environment and Development. In some regions, determined efforts are being made to pay more attention to the problems of small island states, to the management of the associated river basins and to the potential effects of climate change. Protection and management of fisheries resources is the only major issue not covered by most action plans in the framework of the RSP. A further positive development was the establishment, in one of the regions (Mediterranean), of a regional commission for sustainable development, as a subsidiary body of the contracting parties to the Barcelona Convention (*see* Blue Plan, Volume 4; Mediterranean Sea, Volume 4). The establishment of a similar body, as a joint subsidiary body of two African regional seas conventions, is under consideration.

The periodic meetings of the contracting parties to the regional seas conventions or, when action plans are not supported by such conventions, periodic high-level intergovernmental meetings, represent the highest authority guiding the action plans, determining the priorities that should be dealt with by the plans, and allocating the financial resources of the trust funds to these priorities.

UNEP provides the secretariat with six action plans and their respective legal instruments, either directly through its headquarters, or through semiautonomous regional coordinating units managed by UNEP. Three of the action plans have secretariats that are independent of UNEP.

More than 140 countries participate in the RSP through the involvement of about 400 national institutions and their experts. Active support to the implementation of the RSP is provided by a large number of international, intergovernmental and non-governmental organizations.

UNEP provided the funds needed to launch the action plans of the RSP, with the understanding that the governments will gradually assume the financial responsibility for their sustained implementation. However, the present straitened financial circumstances of most countries participating in the RSP are among the major impediments for a more vigorous progress in the development of the RSP.

Whenever appropriate, the RSP is used as the main mechanism to implement the various ocean-related activities of UNEP, such as the Global Plan of Action for the Protection of the Marine Environment from Land-Based Activities, and the Global International Waters Assessment, as well as the relevant parts of global conventions, such as the UN Convention on the Law of the Sea, and the Convention on Biological Diversity. UNEP's overall coordination of the RSP should ensure that such activities, although implemented regionally, remain essentially global in nature.

However, the horizontal cooperation between regional action plans and conventions that constitute parts of the RSP is relatively weak, although direct exchange of information, data, experience and expertise may be of mutual benefit to them. Likewise, there is little substantive cooperation between the RSP and the regional programs of similar nature, such as those carried out under the auspices of the Oslo/Paris Commission or the Helsinki Commission.

For further information on RSP, see http://www.unep.ch/seas/.

United Nations Environment Programme (UNEP) Saskawa Environment Prize

The *Sasakawa Environment Prize*, originally known as the Pahlavi Prize, was first awarded in 1976. The Prize was then funded by Iran and was administered by UNEP.

In 1982, the UNEP Governing Council accepted an endowment of US$1 million from the Japan Shipbuilding Industry Foundation to finance the Sasakawa International Environment Prize which would be administered by UNEP.

Now known as the UNEP Sasakawa Environment Prize, it is awarded annually to leading environmentalists and recognizes the work of these individuals at the global level.

Since its inception, interest in the award has increased significantly as attested by the growing number of nominations. After serious deliberations and in the light of the kinds of nominations received over the years, the Selection Committee decided to re-evaluate the process and has recommended that all nominations be considered on an annual basis and that the Prize be awarded to individuals who have made an outstanding global contribution to the management and protection of the environment . The Prize also aims to encourage environmental achievement in any field of the environment.

The annual award of $50 000 was increased to $200 000 in 1990.

Candidates to the Prize can be associated with any field of the environment. Those eligible to make nominations include, but are not limited to, specialists in environmental sciences, academies of science, engineering and research, members of the United Nations system, governments, intergovernmental organizations, trade unions and non-governmental organizations.

The Prize is named after Mr Ryoichi Sasakawa, who passed away in July 1995, and who was the founder and Chairman of the Sasakawa Foundation for more than three

decades. He donated more than US$400 million to social and public works projects both within and outside Japan.

The first Chairman of the Prize Selection Committee was His Excellency Dr Misael Pastrana Borrero, the late former President of the Republic of Colombia. For more than a decade Dr Pastrana Borrero led the committee with great commitment and dedication to the field of environment.

The inaugural UNEP Sasakawa Environment Prize was awarded posthumously to Dr Aurelio Peccei, who had died just two months before the Prize Selection Committee announced its decision to recognize the outstanding contribution that the former industrialist had made to defend the environment.

Nominations and curriculum vitae can be sent to: The Secretary, UNEP Sasakawa Environment Prize, United Nations Environment Programme, Information and Public Affairs Branch, PO Box 30552, Nairobi, Kenya. Tel.: +254-2-6-3401/62-3128; Fax: +254-62-3692/62-3927, E-mail: rajinder.sian@unep.org, elisabeth.guilbaud-cox@unep.org.

MOSTAFA K TOLBA *Egypt*

United Nations Framework Convention on Climate Change and Kyoto Protocol

Michael Zammit Cutajar
United Nations Framework Convention on Climate Change, Bonn, Germany

The United Nations Framework Convention on Climate Change (the Convention) is the centerpiece of intergovernmental action to combat the threat of human-induced climate change. Negotiations on the Convention were launched by the United Nations (UN) General Assembly in December 1990, in response to warnings from scientists over rising emissions of greenhouse gases from human sources, such as fossil fuel combustion and land clearance. Scientists warned that increased concentrations of greenhouse gases in the atmosphere could lead to irreversible climate change, with potentially dramatic negative impacts on human health, food security, economic activity and infrastructure.

The UN General Assembly established an Intergovernmental Negotiating Committee (INC) to conduct negotiations on the Convention, under the Chairmanship of Jean Ripert (France). After 15 months of negotiations, the Convention was adopted on 9 May 1992, and opened for signature that same year at the UN Conference on Environment and Development (the Earth Summit) in Rio de Janeiro. The Convention entered into force on 21 March 1994, after receiving the required 55 instruments of rati cation, acceptance, approval or accession. One hundred and eighty-three countries, plus the European Community, were Parties to the Convention (as of August 2000), so that it now enjoys virtually universal membership.

THE CONVENTION

Objective and Principles

The ultimate objective of the Convention is to stabilize atmospheric concentrations of greenhouse gases at a level that would prevent dangerous human interference with the climate system. The Convention does not specify what is meant by dangerous, but it does state that ecosystems should be allowed to adapt naturally, food production should not be threatened, and economic development should proceed in a sustainable manner.

The Convention defines a set of principles to guide Parties in their efforts to achieve this objective. A key principle is that of equity and common but differentiated responsibilities (*see* Common but Differentiated Responsibility Principle (Stockholm/Rio), Volume 4). This principle recognizes that all Parties share a common obligation to address climate change, but that developed countries bear a particular responsibility to take the lead in modifying longer-term emission trends. This is because developed countries have historically contributed the most to climate change, their *per capita* emissions remain above those of most developing countries, and they have greater financial and institutional resources to tackle the problem.

Commitments

For the purpose of assigning commitments, the Convention identifies two main groups of countries. Countries listed in its Annex I include those that were members of the Organisation for Economic Cooperation and Development (OECD) in 1992, the European Community, and countries with economies in transition (EITs) from the former Soviet Union and Eastern Europe (Box 1). All other Parties, basically, developing countries, are known as non-Annex I Parties.

General Commitments

All Parties, both Annex I and non-Annex I, are subject to general commitments to address climate change and adapt to its effects. These general commitments include:

- preparing national climate change programs, promoting the development and transfer of climate-friendly technologies;

Box 1 Countries listed in Annex I to the Convention

Australia	Liechtenstein*
Austria	Lithuania
Belarus	**Luxembourg**
Belgium	Monaco*
Bulgaria	**Netherlands**
Canada	**New Zealand**
Croatia*	**Norway**
Czech Republic*	Poland
Denmark	**Portugal**
Estonia	Romania
European Economic Community	Russian Federation
Finland	Slovakia*
France	Slovenia*
Germany	**Spain**
Greece	**Sweden**
Hungary	**Switzerland**
Iceland	Turkey (not yet a Party to the Convention)
Ireland	Ukraine
Italy	**UK**
Japan	**USA**
Latvia	

Notes: Bold indicates countries also listed in Annex II to the Convention. An asterix (*) indicates countries whose names were added to Annex I at COP 3.

- promoting sustainable forest management (forests act as *sinks*, absorbing carbon dioxide, the main greenhouse gas, from the atmosphere);
- preparing for adaptation to climate change (e.g., by developing coastal zone management plans);
- employing methods such as impact assessments to minimize any adverse effects (e.g., on the economy) of climate change mitigation or adaptation measures;
- cooperating in scientific research;
- promoting public awareness of climate change.

The Convention recognizes that certain groups of developing countries are especially vulnerable to the adverse effects of climate change (e.g., countries with low-lying coastal areas) and/or to the impact of the implementation of response measures (e.g., countries whose economies are highly dependent on fossil fuel imports). A process is currently underway to consider how best to address the specific needs and concerns of these Parties (see also the section on financial assistance).

Specific Commitments

In accordance with the principle of equity and common but differentiated responsibilities, only Annex I Parties were subject to the Convention's specific, but non-legally-binding aim to return greenhouse gas emissions to 1990 levels by 2000. The EITs (Economies in Transition) are allowed a certain degree of flexibility in this regard, due to the recent economic and political upheaval experienced by these countries. Several EITs have invoked this clause to take on a baseline other than 1990.

The most recent emissions data indicate that few individual Annex I Parties will meet this specific aim. Except for the EITs, whose emissions fell dramatically as a result of their economic collapse following the break-up of the Soviet Union, the emissions of most Annex I Parties may have continued to rise since the adoption of the Convention. As a group, however, the Annex I Parties may have indeed returned their collective emissions to 1990 levels by 2000, due principally to the falling emissions of the EITs and cuts achieved in Germany and the United Kingdom (emissions data for the year 2000 were not yet available at the time of writing).

Finance and Technology Transfer

In order to effectively address climate change and adapt to its effects, the Convention recognizes that developing countries will need access to financial resources and climate-friendly technologies. The OECD members of Annex I (who are also listed in the Convention's Annex II) are therefore committed to providing new and additional financial resources for this purpose. This includes a requirement to assist developing countries that are particularly vulnerable to the adverse effects of climate change in meeting the costs of adaptation. Financial assistance to developing countries is mostly channeled through the Convention's financial mechanism; currently operated by the Global Environment Facility (GEF) (see the following section). In addition, Annex II Parties are required to promote the transfer of climate-friendly technology to both developing countries and EITs, while greater emphasis is now being placed on the need for capacity-building in both these groups.

Some financial support is also available for eligible developing countries to enable them to attend sessions of the Convention bodies (see the following section). This support is funded by voluntary contributions from Parties, and is channeled through a special trust fund under the Convention. Approximately $1.8 million a year was earmarked for this purpose in 2000 and 2001.

Reporting and Review

All Parties are required to report on the action they are taking to implement the Convention. The level of detail required, however, is greater for Annex I Parties, and the time frame for submitting their reports, known as *national communications*, is tighter. Most Annex I Parties have now submitted two national communications, and their third was due on 30 November 2001. These national communications must include information on the climate change policies that

Annex I Parties have implemented, along with the projected effects on their emissions. Annex II Parties must also report on the financial assistance they have provided to developing countries and on their technology transfer initiatives. The national communications of Annex I Parties are subject to an individual in-depth review by a team of experts, whose reports are made publicly available. In addition, Annex I Parties must submit annual inventories of their greenhouse gas emissions. These are subject to a technical review for a trial period starting in 2000.

The information required in national communications from non-Annex I Parties is less specific, in line with the more general commitments of these Parties under the Convention. The Convention allows more time to non-Annex I Parties to prepare their national communications, including their emission inventories, and the time frame is also contingent on the receipt of financial resources from the GEF. For this reason, non-Annex I Parties have only recently begun to submit their initial national communications. As of August 2000, 27 had been received. Non-Annex I Party national communications are not subject to individual in-depth review.

Institutions

Conference of the Parties

The supreme decision-making body of the Convention is the Conference of the Parties (COP) (Figure 1). The COP, which brings together all Parties to the Convention, usually meets on an annual basis. It has now met six times in cities around the world; its latest session at the time of writing, COP 6, was held in The Hague (Netherlands) in November 2000.

The COP has responsibility for reviewing the Convention's implementation and for assessing the progress made by Parties in meeting their commitments. At each of its sessions, the COP takes decisions aimed at promoting the effective implementation of the Convention. For example, at its second session (Geneva, 1996), the COP adopted revised guidelines to be used by Annex I Parties to prepare their national communications. It further updated these guidelines at its fifth session (Bonn, 1999). Also at its fifth session, the COP established a consultative group of experts on non-Annex I Party national communications. At its fourth session (Buenos Aires, 1998), the COP launched a technology transfer consultative process aimed at giving new impetus to efforts to promote climate-friendly technology transfer. Decisions taken by the COP, around 20 at each session, make up a detailed rulebook for the implementation of the Convention.

The proceedings of the COP, and those of the subsidiary bodies discussed below, are governed by written rules of procedure. Although Parties have not yet formally adopted these rules, due to disagreements over the voting majority needed to take decisions, the draft rules of procedure are applied at each session of the COP, except for the rule on voting. As there is no agreed voting rule, almost all decisions under the Convention must be taken by consensus, which is taken to mean that there is no stated objection to a decision. Amendments to the Convention and its Annexes, however, may be adopted by a three-quarters majority vote, because the Convention itself specifies this voting rule. No amendments have yet been adopted to the Convention itself, although its Annex I was amended at COP 3 (see Box 1).

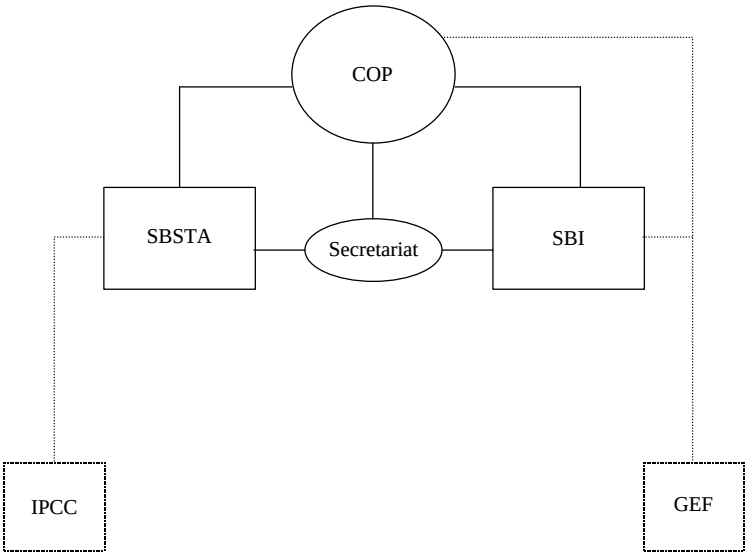


Figure 1 Convention bodies. Note: The GEF formally reports to the COP, hence the direct link shown in this figure. COP is the Conference of Parties; SBSTA is the Subsidiary Body for Scientific and Technological Advice, and SBI is the Subsidiary Body for Implementation

Observer organizations, such as representatives of UN bodies, inter-governmental organizations (IGOs) or non-governmental organizations (NGOs) (e.g., business and industry groups, environmental NGOs, local authorities), are allowed to attend meetings of the COP and the subsidiary bodies, unless at least one-third of Parties object. There has been considerable interest in the Convention on the part of observer organizations, and their representatives routinely outnumber Party delegates at sessions of the Convention bodies. Almost 400 NGOs and approximately 35 IGOs are currently accredited to the Convention process.

Subsidiary Bodies

The Convention also established two permanent subsidiary bodies, the Subsidiary Body for Scientific and Technological Advice (SBSTA) and the Subsidiary Body for Implementation (SBI) (see Figure 1). Both bodies meet at least twice a year to carry out preparatory work for the COP. They are open ended, which means that any Party can participate in their sessions.

The SBSTA has responsibility for providing advice to the COP on scientific and technological matters. This includes, for example, work on suitable methodologies for compiling emission inventories or on how best to promote the transfer of climate-friendly technologies. The SBSTA also serves as the link between information provided by scientific bodies outside the climate change regime, such as the Intergovernmental Panel on Climate Change (IPCC), and the policy needs of the COP.

The SBI advises the COP on issues related to the implementation of the Convention, in particular on the review of national communications. It also has specific responsibility for liaison with the GEF, as well as for financial and administrative matters, including consideration of the proposed program budget (see the section on the secretariat).

Financial Mechanism: the Global Environment Facility (GEF)

The Convention established a financial mechanism to channel financial assistance to developing countries and entrusted its operation to the GEF on an interim basis. A memorandum of understanding with the GEF was adopted at COP 2, and COP 4 confirmed that the GEF would continue to operate the financial mechanism, subject to review every four years. The GEF is accountable to the COP for its climate change activities, and the COP provides it with regular policy guidance.

The GEF, a joint initiative of the UN Development Programme, the United Nations Environment Programme (UNEP) and the World Bank, was launched in 1991 (initially as a pilot program) to fund developing country projects resulting in global environmental benefits (the GEF also funds projects on ozone protection, international waters and biodiversity). Since its establishment, the GEF has allocated a total of US$5.7 billion to climate change activities.

IPCC

Although the IPCC is not an institution of the Convention, it has played a crucial role in its development. The IPCC was established in 1988 by UNEP and the World Meteorological Organization to provide policy-makers with authoritative and up-to-date information on climate change science. The First Assessment Report of the IPCC, published in 1990, confirmed that climate change was indeed a threat and recommended the launch of negotiations on a treaty to address the problem. The IPCC's Second Assessment Report (SAR), finalized in 1995, presented the landmark conclusion that the balance of evidence suggests that there is a discernible human in uence on global climate . The SAR estimated that global average temperatures would increase by 1–3•5 °C by 2100, the fastest rate of change since the end of the last ice age, while sea levels would rise by 15–95 cm over the same time period. The SAR's conclusions were critical in accelerating the momentum of negotiations on what was to become the Kyoto Protocol. The Third Assessment Report was released in 2001.

The IPCC also prepares more concise Special Reports and Technical Papers on specific issues, usually at the request of the SBSTA. For example, it recently finalized Special Reports on aviation and the global atmosphere, technology transfer and land-use, land-use change and forestry. This latter Special Report will be particularly helpful to Parties as they consider how to treat carbon sinks under the Kyoto Protocol (see the section on specific commitments under the Kyoto Protocol) (*see* Intergovernmental Panel on Climate Change (IPCC): an Historical Review, Volume 4).

Secretariat

The COP and its subsidiary bodies are serviced by a secretariat, currently staffed by some 100 international civil servants. The main functions of the secretariat are to make practical arrangements for sessions of the Convention bodies, provide support for on-going negotiations (e.g., by drafting background documents), circulate information submitted to it by Parties, and ensure the necessary cooperation with secretariats of other bodies, such as the GEF and the IPCC.

The secretariat operates autonomously within the administrative framework of the UN. It is headed by an executive secretary, at the level of assistant-secretary-general of the UN, who is accountable to the COP and to the secretary-general of the UN. The current executive secretary, Michael Zammit Cutajar (Malta), has headed the secretariat since it was set up in 1991. Since August 1996, the secretariat has been located in Bonn (Germany), having moved there from its previous location in Geneva, following an offer from Germany to host the secretariat. The offer was accepted at COP 1.

The secretariat works according to a biennial program budget, which is proposed every two years by the executive secretary. The current program budget for the biennium 2000–2001 amounts to around $12 million per year. It is funded by contributions from Parties, with their shares calculated according to the UN scale of assessment.

THE KYOTO PROTOCOL

When governments adopted the Convention, they knew that its commitments would not be sufficient to achieve its ultimate objective. They therefore included provisions to ensure the review of the Convention at regular intervals, with the aim of strengthening their commitments as scientific evidence of climate change hardened and political will increased. The first review of the Convention took place at COP 1 (Berlin, 1995) where, in a decision known as the Berlin Mandate, Parties agreed to launch negotiations to strengthen commitments for Annex I Parties through the adoption of a protocol or another legal instrument. The Berlin Mandate specified that this round of negotiations should not set any new commitments for developing countries. Negotiations took place in the specially convened Ad Hoc Group on the Berlin Mandate (AGBM), under the Chairmanship of Ambassador Raúl Estrada-Oyuela of Argentina. The AGBM met eight times over a period of 31 months, before forwarding the results of its work to COP 3 (Kyoto, 1997), the deadline for concluding negotiations set by COP 1. Intense negotiations continued at COP 3 before the Kyoto Protocol was eventually adopted on 11 December 1997.

The Protocol will enter into force when it has received instruments of ratification, acceptance, approval or accession, from 55 Parties to the Convention, including Annex I Parties accounting for 55% of the 1990 carbon dioxide emissions of that group. As of August 2000, 22 Parties had ratified the Protocol, all of them developing countries. However, almost all Annex I Parties were among the 84 Parties who signed the Protocol during the one-year period when it was open for signature, thus indicating their intent to ratify in the future.

The Kyoto Protocol shares the objective and principles of the Convention. It will also make use of the Convention's institutions, such as the subsidiary bodies, secretariat and financial mechanism, while the COP to the Convention will serve as the meeting of the Parties to the Protocol (the COP/Meeting of the Parties, MOP).

Emission Commitments

The cornerstone of the Kyoto Protocol is its emission targets for Annex I Parties, which, in a significant step forwards from the Convention, are legally binding. The emission targets for individual Annex I Parties are listed in the Protocol's Annex B (see Box 2). Together, they add up to a total reduction in Annex I Party emissions of around 5.2% from 1990 levels. However, because the emissions of most Annex I Parties have increased since 1990, this collective target amounts to a real reduction of over 20% from business-as-usual projections. If this target is met, it will at last reverse the upward trend in greenhouse gas emissions that has prevailed in the industrialized world over the past two centuries.

Box 2 Individual emission targets under the Kyoto Protocol

Party	Percentage change	Party	Percentage change
Australia	+8	Liechtenstein	−8
Austria	−8	Lithuania	−8
Belgium	−8	Luxembourg	−8
Bulgaria	−8	Monaco	−8
Canada	−6	Netherlands	−8
Croatia	−5	New Zealand	0
Czech Republic	−8	Norway	+1
Denmark	−8	Poland	−6
Estonia	−8	Portugal	−8
European Community	−8	Romania	−8
Finland	−8	Russian Federation	0
France	−8	Slovakia	−8
Germany	−8	Slovenia	−8
Greece	−8	Spain	−8
Hungary	−6	Sweden	−8
Iceland	+10	Switzerland	−8
Ireland	−8	Ukraine	0
Italy	−8	UK	−8
Japan	−6	USA	−7
Latvia	−8		

Notes: Belarus and Turkey, while both included in Annex I to the Convention, were not assigned emission targets under the Kyoto Protocol and therefore not included in Annex B, as they were not Parties to the Convention when the Protocol was adopted.

The individual targets of Annex I Parties range from −8% for the 15 members of the European Union, Switzerland and several EITs, to an increase of +10% allowed to Iceland. Under the terms of the Protocol, the EU is permitted to redistribute its target among its 15 member States, so long as the overall −8% reduction is achieved. The EU has already reached agreement on how it will do this.

The emission targets cover the six main greenhouse gases (see Box 3) taken together as a basket, and are calculated against a baseline of 1990 (except for the EITs, who may request a different baseline). However, an alternative baseline of 1995 may be chosen for HFCs, PFCs and sulfur hexa uoride (SF_6). This is because, in some Parties, emissions of these gases increased considerably in the early

Box 3 Greenhouse gases covered by the Kyoto Protocol

Carbon dioxide (CO_2)
Methane (CH_4)
Nitrous oxide (N_2O)
Hydrofluorocarbons (HFCs)
Perfluorocarbons (PFCs)
Sulfur hexafluoride (SF_6)

1990s, as they were used as substitutes for ozone-depleting substances being phased out under the 1987 Montreal Protocol on Substances that Deplete the Ozone Layer. Certain activities that enhance carbon sinks, notably afforestation, reforestation and tackling deforestation, may be used to offset emission targets, although the uncertainty and complexity surrounding carbon sinks means that the details of these provisions remain to be worked out (*see also* Boreal Forest Carbon Flux and it s Role in the Implementation of the Kyoto Protocol Under a Warming Climate, Volume 4). Annex I Parties must meet their emission targets by the commitment period of 2008–2012. A target period was chosen rather than a single year to smooth out annual uctuations in greenhouse gases. The maximum amount of greenhouse gas emissions that an Annex I Party may emit over this five-year commitment period is known as its *assigned amount*.

In addition to these emission commitments for Annex I Parties, all Parties (including non-Annex I Parties) are subject to a set of general commitments that reaffirm the general commitments in the Convention and seek to advance their implementation, in order to achieve sustainable development. These general commitments cover the preparation of greenhouse gas inventories and national climate change programs; the development and transfer of technology; scientific research; education, training and public awareness on climate change; and reporting. Regarding national climate change programs, the Protocol specifies that these would include consideration of the energy, transport and industry sectors, as well as agriculture, forestry and waste management.

Implementation Mechanisms

The Kyoto Protocol does not specify how Annex I Parties should implement their emission targets. It does, however, include a list of possible policies and measures that Parties could adopt, including enhancing energy efficiency, developing new and renewable forms of energy, and phasing out market imperfections (such as subsidies and tax exemptions) that go against the objective of the Convention. The Protocol also requires Annex I Parties to strive to minimize the adverse effects of their climate change mitigation policies on other Parties, especially developing countries. Fossil fuel exporters, for example, are particularly concerned about knock-on effects on their economies that may arise from measures implemented by Annex I Parties to reduce their greenhouse gas emissions from fossil fuel consumption.

The Kyoto Protocol also includes three innovative mechanisms that Annex I Parties may use to help them meet their emission targets as cost-effectively as possible.

1. Under Article 6 of the Protocol, Annex I Parties may invest in emission reducing (or sink enhancing) projects in other Annex I Parties, and credit the resulting emission reduction units to their own assigned amounts. A corresponding debit is made from the assigned amount of the host country of the project. This mechanism is often known as *joint implementation* (*see* Joint Implementation, Volume 4). It is expected that joint implementation will mostly be carried out through investments by the OECD members of Annex I in the EITs, where climate change mitigation options are often cheaper and more plentiful.
2. Article 12 of the Protocol defines a *clean development mechanism* (CDM) (*see also* Clean Development Mechanism, Volume 4), whose purpose is not only to help Annex I Parties meet their emission targets, but also to assist non-Annex I Parties in achieving sustainable development. Under the CDM, Annex I Parties may use certified emission reductions achieved through climate change mitigation projects in non-Annex I Parties to offset their own emission targets. Because non-Annex I Parties do not have emission targets or baselines of their own, it is particularly important to ensure that emission reductions achieved through CDM projects are additional to what would otherwise have occurred in the absence of the project. With this aim in mind, emission reductions must be certified by separate operational entities, and also be subject to independent auditing and verification. The CDM will be supervised by an executive board. Developing countries that are particularly vulnerable to the adverse effects of climate change will gain added benefit from the CDM, as a share of the funds generated by the mechanism will be diverted to help these countries meet adaptation costs. The Kyoto Protocol envisages that participation in joint implementation and the CDM will prove attractive not only to governments, but also to private entities, such as businesses. Both mechanisms certainly hold considerable potential to harness private sector investment for climate-friendly projects in EITs and developing countries, thus increasing the ow of much-needed financial resources to these countries, while promoting the transfer of environmentally friendly technologies. Joint implementation and the CDM could thus create unprecedented opportunities to tackle climate change

in an economically effective way, while also promoting sustainable development.

3. The third Protocol mechanism, outlined in Article 17, is *emissions trading* (*see also* Emissions Trading, Volume 4). Emissions trading allows an Annex I Party that finds it relatively easy to achieve its emissions target to transfer parts of its assigned amount that it may not need to another Annex I Party that is experiencing greater difficulty in meeting its commitments. Article 17 states, however, that use of emissions trading by Annex I Parties to meet their targets must be supplemental to domestic climate change mitigation action. Although the Protocol does not mention the private sector, many Parties see opportunities for emissions trading, both domestically and internationally, as an important incentive for engaging private businesses in climate change mitigation efforts.

Credibility Mechanisms

The legally binding nature of the Protocol's emission targets places a premium on accurate emission inventories, rigorous reporting by Parties and thorough review of both inventories and reports, to ensure that compliance can be authoritatively assessed and cases of non-compliance identified. The Protocol therefore provides for more stringent guidelines for the preparation of both annual emission inventories and national communications, and for a more rigorous in-depth review process than that under the Convention. These provisions can be understood as the Protocol's credibility mechanisms, which seek to uphold its environmental effectiveness.

Key to securing the Protocol's environmental credibility is the compliance provisions outlined in Article 18. These call on the COP/MOP, at its first session after entry into force, to approve procedures and mechanisms to determine and address cases of non-compliance with the Protocol. However, any non-compliance procedures and mechanisms that entail binding consequences can only be adopted by amending the Protocol.

Unfinished Business

When Parties adopted the Kyoto Protocol at COP 3, they left a large burden of unfinished business to be worked out in subsequent negotiations. In particular, the details of how joint implementation, the CDM and emissions trading would operate in practice remain unresolved. In addition, more technical work is needed to iron out provisions on carbon sinks and to draw up guidelines for the Protocol's credibility mechanisms, especially the all-important compliance procedures. A further outstanding issue is how to minimize adverse effects, of both climate change and mitigation action, on developing countries.

The Kyoto Protocol cannot enter into force while so much unfinished business is still pending; Annex I Parties need to have a clearer idea of the shape of their commitments under the Protocol before ratifying it. Recognizing this, at COP 4, Parties adopted the so-called Buenos Aires Plan of Action for finalizing the Protocol's unfinished business, and set COP 6 (The Hague, November 2000) as a deadline for adopting decisions on it. (The Plan of Action also covered important on-going Convention issues, notably the financial mechanism and technology transfer.)

Since COP 4, work on the Protocol's unfinished business has proceeded at an intensive pace in both the SBSTA and the SBI, and also in a Joint Working Group established by COP 4 to negotiate the Protocol's compliance procedures. Numerous workshops and informal meetings have also been held in between sessions.

Concluding Remarks

The Kyoto Protocol has already promoted much greater awareness of climate change among businesses and the public worldwide. Its adoption sent a clear signal that governments take the threat of climate change seriously, and are committed to tackling it. This signal has triggered a noticeable response; many businesses are already implementing energy efficiency measures, increasing their investments in renewable energy, accelerating the development of low-carbon fuel cells or exploring emissions trading opportunities. Even in developing countries, which have no formal emission targets, there is strong evidence that environmentally friendly policies are slowing the growth in greenhouse gas emissions. These policies are usually implemented for reasons other than climate change, demonstrating that tackling climate change can also bring economic and social side-benefits.

Despite these encouraging signs, much more aggressive action is needed to move emission trends onto a downward path and avoid the worst effects of climate change. Combating climate change is a long-term task; a new round of negotiations to determine the next set of commitments post-2012 is due to start no later than 2005, and may be launched earlier. Targets for Annex I Parties will need to be tightened and, eventually, commitments must be extended to a wider group of countries. For now, however, the next crucial step is entry into force and implementation of the Kyoto Protocol. The Convention and Protocol have laid the foundations for an effective response to climate change; it is now up to the Parties to implement their provisions and finally mark an end to the business-as-usual trends that threaten the world's climate.

Further information on the Convention and the Kyoto Protocol, including all official documents and many national communications, may be obtained from the secretariat web site (www.unfccc.de).

United Nations Industrial Development Organization (UNIDO)

see UNIDO (United Nations Industrial Development Organization) (**Volume 4**)

United Nations Population Fund (UNFPA)

see UNFPA (United Nations Population Fund) (**Volume 4**)

UNSCEAR (United Nations Scientific Committee on the Effects of Atomic Radiation)

The UNSCEAR is a Committee of the General Assembly of the United Nations. The Committee was established in 1955 in response to widespread concerns regarding the effects of radiation on human health and the environment. At that time, nuclear weapons were being tested in the atmosphere, and radioactive debris was dispersing throughout the environment, reaching the human body through the intake of air, water and food. Treaties now ban atmospheric tests, but many other exposures to radiation occur from other sources, both natural and man-made.

The Committee collects and assesses global data on the sources and effects of ionizing radiation. Detailed reports to the General Assembly are produced every few years and are made widely available as United Nations sales publications. The latest editions were 1988, 1993, 1994, 1996 and 2000. The reports review exposures from natural radiation sources, from man-made practices and events such as nuclear power production, nuclear tests, Chernobyl and other accidents, from medical radiation uses and occupational radiation exposures. Methods of exposure evaluation and extensive data compilations are included. The reports also review results of epidemiological, radiobiological and genetic effect studies in order to estimate risks of exposures. Other biological topics are addressed to improve understanding of mechanisms involved in radiation effects and to interpret modified features of exposures, e.g., dose rate, age/tissue sensitivities, adaptive responses, interactions with other agents. The assessments are recognized to be scientifically accurate and authoritative, and they serve as the basis for national and international standards for radiation protection.

The Committee is comprised of 21 member countries. Its Secretariat is located at the Vienna International Center, P.O. Box 500, A-1400 Vienna, Austria.

BURTON BENNETT *USA*

Urban Lifestyles, Sustainability, and Integrated Environmental Assessment (ULYSSES)

see ULYSSES (Urban Lifestyles, Sustainability, and Integrated Environmental Assessment) (**Volume 4**)

Use of Epidemiological Data to Develop Air Quality Standards for Short-term and Long-term Environmental Health Benefits

Jeffrey R Brook[1], Richard T Burnett[2], and Marjorie Shepherd[1]

[1] *Meteorological Service of Canada, Toronto, Canada*
[2] *Health Canada, Ottawa, Canada*

Air pollutant emissions need to be managed and ideally reduced to avoid environmental and human health problems. This can be achieved by setting and enforcing ambient air quality standards as part of a program of long-term national and international environmental policies. Standards must be based upon science before they are accepted by the public and industry. Epidemiological studies, which include the analysis of how the health within a population responds to environmental hazards, such as air pollution, are a useful tool for developing air quality standards. To demonstrate why and how information from epidemiological analyses can be applied, a case study based on a recently developed Canadian air quality standard will be discussed.

Table 1 Common air pollutants associated with urban smog

Direct products of combustion *Primary Pollutants*	Products of atmospheric chemical reactions *Secondary Pollutants*	Other
Carbon monoxide	Ground level ozone	Particles[a] (e.g., re-suspended dust)
Nitrogen oxides		Volatile organic compounds (e.g., unburned fuel, solvent escape)
Particles[a] (PM_{10}, $PM_{2.5}$)		
Sulfur oxides	Particles[a] (PM_{10}, $PM_{2.5}$)	

[a] Particles or suspended particulate matter (PM) are tiny solid/liquid particles, which are made up of a range of chemical compounds, many of which are produced during combustion. Particle concentrations, as well as gaseous pollutant concentrations, are often measured as a mass per cubic meter of air. The size of the particle included in a measurement should also be specified, where particle diameter (d) in micrometers is indicated after the abbreviation; PM_d (e.g., $PM_{2.5}$ means all suspended particulate matter below 2.5 μm in diameter). Sulfate, nitrate, ammonium and organic carbon are the most prevalent chemical compounds in $PM_{2.5}$ while larger particles will contain more soil-based compounds like silicon, aluminum and calcium.

THE NEED FOR AIR QUALITY STANDARDS

Air pollution or smog is an urban and regional-scale problem with the air pollutants having their greatest impact on people and the environment in the source area (i.e., an urban setting) and up to at least 500 km downwind (regional scale). Inhalation of the common pollutants associated with urban smog, which are listed in Table 1, can lead to a range of cardio-respiratory effects from exacerbation of asthma in children to premature mortality of people with severe, but not necessarily untreatable, pre-existing diseases. The amount of research documenting these effects has increased significantly during the past 10 years. In particular, several epidemiological studies of populations in different cities found associations between daily mortality rates and day-to-day variations in fine particle concentrations at levels below current ambient air quality standards. This consistency among studies lends strength to the case for more stringent standards for fine particles. The fact that acute effects other than mortality, such as increases in hospital admissions, emergency department visits, asthma symptoms, changes in heart rate variability, and changes in blood composition, have been found, as well as evidence of chronic effects, strengthens the scientific rationale for a new fine particle standard (i.e., coherence among effects). However, particles are not the only pollutant of concern. There continues to be strong evidence that ozone has a significant impact on public health, including mortality. Nitrogen dioxide, carbon monoxide and ultra fine particles ($PM_{0.1}$) have also been implicated in recent studies. Thus, many pollutants in the smog mixture are of concern and their combined effect on public health is greater than the effect of any individual pollutant (see Figure 1).

The pollutants in smog should be managed in a coordinated fashion, optimizing the benefits to public health, especially given their common link to fossil fuel combustion and their interactions in the atmosphere. For example, lowering greenhouse gas emissions would lead to direct decreases in air pollutants. There is also the potential for indirect benefits if changes in global climate and subsequent changes in prevailing regional weather patterns, which could cause more smog episodes and/or overall increases in pollutant levels, are avoided.

Figure 1 The Toronto skyline: (a) obscured by the fine particles found in urban smog and; (b) a relatively particle-free day. The CN Tower, which is ≈5 km away, and the other buildings are much less visible due to high particle levels. The hourly average $PM_{2.5}$ concentrations in (a) and (b) were 40 and 6 μg m^{-3}, respectively. Hourly concentrations of 50–100 μg m^{-3} are not uncommon in many of the world's major cities. The annual average 24 h $PM_{2.5}$ concentration in Toronto is about 16 μg m^{-3}, just above the new US Environmental Protection Agency annual standard

Management of air pollutants is achieved through *source-based* (*emission controls*) and *receptor-based* (*air quality standards*) strategies. Ideally, only a source-based approach should be needed, requiring emissions to be kept at an absolute minimum. This is not achievable due to technological limits and economic constraints. Even if every source was optimally controlled, an ever-increasing number of sources would still lead to increasing smog. Thus, air quality standards are needed to ensure that when there are too many sources or emissions are too large, there are mechanisms prohibiting smog levels from getting too high. Standards or knowledge of safe levels are also important because as global emissions increase, which is likely as developing countries increase their energy consumption and economic growth continues worldwide, the area affected by air pollution can be expected to increase.

Air quality standards are management targets to reduce impacts on health and the environment to acceptable levels. They have been established in many countries and guidelines have been recommended by the World Health Organization (WHO). Although the standards, objectives or guidelines (hereafter referred to as standards) are similar among countries, and human health protection typically serves as the rationale, there are differences regarding enforcement and compliance. If data exist, then environmental protection, particularly for crops, forests and aquatic ecosystems, may also provide rationale and useful information for establishing standards. Most jurisdictions review standards on a regular basis ensuring that they re ect current knowledge of health and environmental effects, so that they continue to be effective in promoting progress towards cleaner air.

USE OF EPIDEMIOLOGICAL DATA FOR HEALTH-BASED STANDARDS

Setting the level and the form (e.g., hourly average, 24 h average, annual average, etc.) for a standard is not straightforward. The health effects data may be inconsistent due to variations in study design and type and geographic location. Several approaches to studying effects exist, but the three basic approaches for obtaining data for setting a standard are animal toxicology, controlled human exposure studies, and epidemiological studies (observational). Each of these is necessary to demonstrate the biological plausibility of an effect and to build a solid scientific case for new standards. However, observational epidemiological studies involving the population at large are particularly useful for standard setting because:

- they provide a basis for determining exposure and concentration response functions (CRFs). For examples of CRFs see Figure 2(a);
- they are the most direct way of assessing the health outcomes of real-world mixtures of pollutants to which people are exposed;
- they are the only way to determine or predict health effect responses among a human population because, unlike laboratory animals, populations are heterogeneous, containing individuals with varying susceptibilities, disease status and exposures. A population-based response cannot be predicted from animal toxicology and all the necessary information is not likely available from controlled human exposure studies due to ethical reasons;
- they are based on large administrative databases and thus demonstrate the impacts on public health and even enable estimates of the costs to society;
- they are the only direct source of information on long-term cumulative effects.

Observational epidemiological studies build a link directly between air quality monitoring site data, which are ultimately used to evaluate implementation programs, and health impacts. When such links are detected, as has been demonstrated for many air pollutants in smog, it is important only to know that there is a correlation between the ambient monitor concentrations and the personal exposure to pollutants of outdoor origin.

There are some difficulties with relying on epidemiological data to establish CRFs for individual air pollutants. There are significant potential confounders to the interpretation of the epidemiological studies, including weather, co-varying air pollutants and possibly indoor air exposure. Most recent studies have used techniques which reduce confounding, but it is not possible to completely eliminate this concern. Nonetheless, given the variety of locales, pollutant mixtures and exposure levels in the epidemiological studies and the growing body of evidence from other types of studies, it is clear that confounding does not negate the association observed between ambient concentrations and impacts.

Another difficulty is limited information on the mode of action; and causal mechanisms are only now being elucidated, especially in the case of particles. Consequently, objective assessment of the coherence and consistency of the body of evidence, as was discussed above for the case of particles, is an important step that must be undertaken before accepting and applying the epidemiological data.

DEVELOPMENT OF CONCENTRATION-RESPONSE FUNCTIONS AND THEIR APPLICATION IN SETTING A STANDARD

Achieving an air quality standard has traditionally implied that there will be no adverse health effects if concentrations are kept below that level. Recent epidemiological analyses

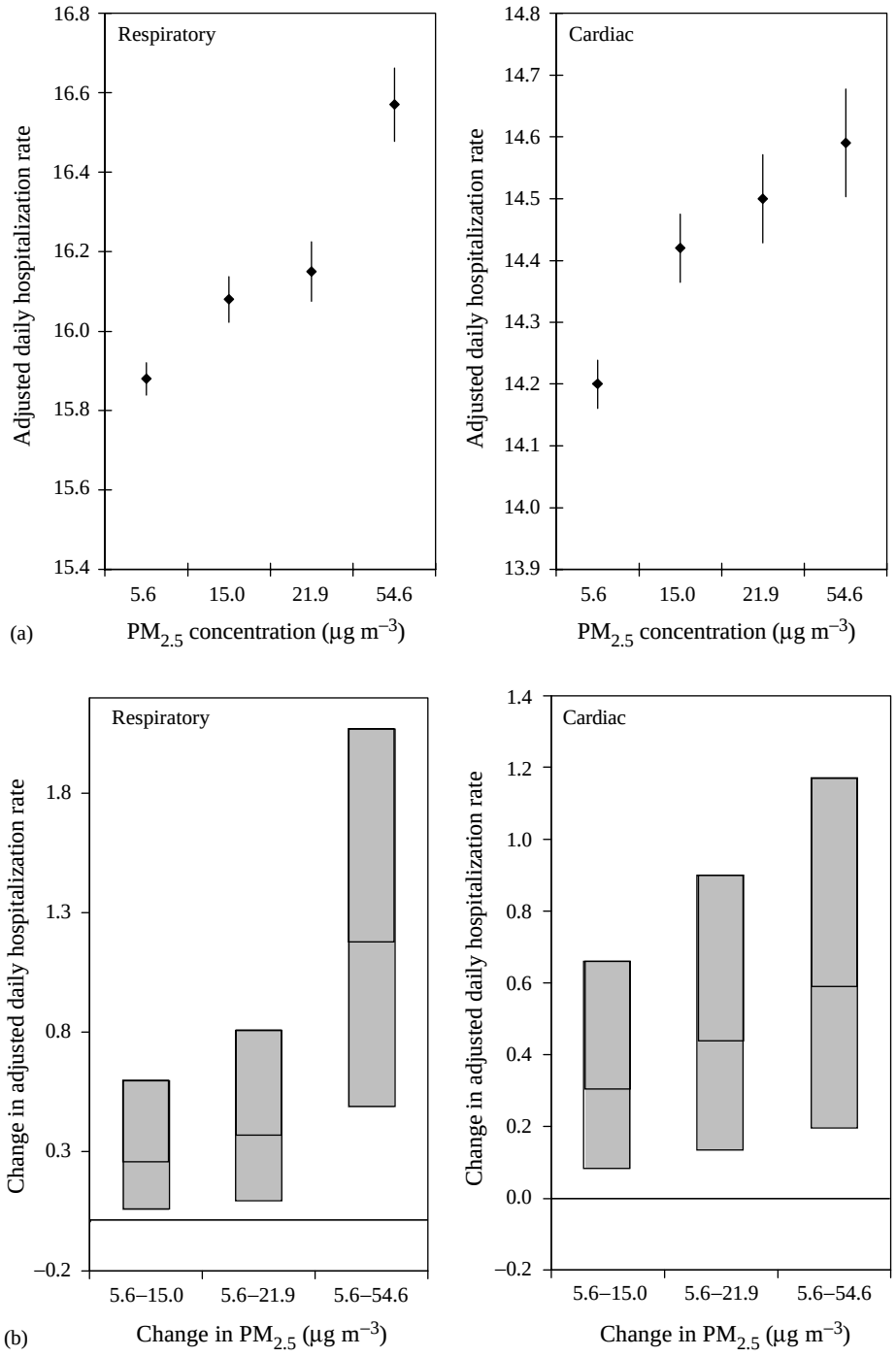


Figure 2 (a) Relationship between the mean adjusted respiratory and cardiac hospitalization rates (±1 standard error) and ambient $PM_{2\cdot5}$ concentrations in southern Ontario, Canada. The four groups represent observations within the lower two quartiles (mean $PM_{2\cdot5}$ of 5.6 μg m^{-3}), the third quartile (mean $PM_{2\cdot5}$ of 15.0 μg m^{-3}), the 75th to 90th percentile (mean $PM_{2\cdot5}$ of 21.9 μg m^{-3}) and the 90th to 100th percentile (mean $PM_{2\cdot5}$ of 54.6 μg m^{-3}). (b) Statistical significance of the increases in the mean adjusted respiratory and cardiac hospitalization rates associated with an increase in the ambient $PM_{2\cdot5}$ concentration. Solid bar shows the mean value of the increase in rate and shaded areas above and below the bar indicate ±2 standard errors (≈95% confidence interval). An overlap with the zero line indicates a non-significant change. The lowest concentration at which a significant change was detected was for an increase from 5.6 to 15.0 μg m^{-3}

focusing on ozone, particles (PM) and carbon monoxide, however, have not found conclusive evidence of a threshold below which there are no effects. The CRFs from studies involving relatively large populations have mostly been found to be linear or near linear throughout the range of ambient air concentrations (FPWG, 1999).

The lack of thresholds for air pollutant impacts from the observational epidemiological studies does not imply that individuals do not have a threshold below which they are not affected. There is a broad distribution of people with pre-existing conditions within any population and each could potentially have his/her own threshold. Furthermore, this threshold could vary in time depending on a person's current health. The present observational studies do not provide information on specific individuals so the entire range of ambient concentrations is associated with some health response within the population.

Even if a threshold can be identified, it would have been derived from currently available data and methods, which generally have limitations. New information could reveal that the most relevant threshold is different than the currently accepted value. For example, there is presently very little data on the chronic or long-term effects of air pollutant exposure. Consequently, the CRFs and subsequently identified standards are based on impacts of acute exposures. Future data from chronic effect studies could lead to different CRFs and threshold estimates that may be more relevant for minimizing or eliminating the effect on the population. Thus, the process of standard setting is an ongoing process of re-evaluation as knowledge increases. Epidemiological studies play an important role in this process.

Without evidence of a threshold, setting a defensible air quality standard is difficult. Risk analysis/management is one option to be considered, but it is necessary to define an acceptable level of risk to public health. Should the most susceptible members of the population be protected or should the average person be of most concern? Should the cost of emission controls be an issue? Can such costs be reliably compared to the cost to society arising from the burden on public health? Should a margin of safety be considered? There are no simple answers to such questions. However, to be able to undertake a quantitative, scientifically defensible analysis, a CRF is required for each health endpoint to be considered (e.g., mortality, hospital admissions, incidence of bronchitis).

The two main approaches for deriving a CRF are to:

1. Quantitatively combine the results from the most robust epidemiological studies to establish a relatively broad-based CRF (referred to as a *meta-analysis*);
2. Rely on the most relevant study for the area/country of interest.

The meta-analysis approach is useful because published data from multiple studies can be considered so there is typically a greater degree of confidence in the derived CRF. In addition, this CRF should be more generally applicable for a broad range of geographic areas, even those not represented by the original epidemiological studies. The second approach directs significant attention at one study and thus, it is important to critically review its strengths and weaknesses before applying the results.

In many epidemiological studies, particularly for rare events like mortality, the air pollutant effect is expressed as a relative risk per given change in the ambient concentration. Relative risk values are useful because they can be translated into a percentage change in the incidence of a health effect for a baseline population. To apply risk estimates, it is helpful to standardize them to a unit change in concentration yielding a CRF. The final step in applying the CRF is to obtain the baseline rate ($\overline{Rate}_{outcome_i}$) for the corresponding health outcome, i, (e.g., cardio-respiratory mortality) expressed as number of occurrences per given number of people (e.g., per million population).

CONSIDERING THE OPTIONS FOR A PM STANDARD: A CASE-STUDY BASED UPON A RECENT CANADIAN EXERCISE

In Canada, a group of air quality and health experts from governmental agencies at the federal and provincial/territorial levels, with support from outside experts as needed, provide guidance to air quality managers for development of ambient standards and management strategies. Involvement from various levels of government and geographic regions ensures that a broad range of perspectives is considered. In addition, this group consults with stakeholders, such as environmental and health organizations and industry. This process was followed to provide guidance for setting national ambient standards for fine PM.

Three approaches for developing health-based recommendations for the standard were considered:

1. Base the standard on the *lowest observable adverse effect level* (LOAEL), which is the lowest concentration at which one is confident that there are adverse health effects;
2. Base the standard on the LOAEL incorporating a margin of safety based on uncertainty factors;
3. Base the standard on an analysis of the health benefits of achieving various arbitrarily defined $PM_{2.5}$ concentration levels.

These approaches were the only justifiable options given the lack of evidence of a health effect threshold or a no-observable adverse effect level (NOAEL: the highest concentration at which one is confident there are no adverse health effects), for PM_{10} and $PM_{2.5}$. Following any of these approaches necessitated identification of CRFs to be able to identify the LOAEL and/or to estimate health benefits.

Daily mortality rate and hospital admissions were selected as the health endpoints to be considered in the standard setting exercise. Mortality was selected due to the large number of studies reporting consistent and similar effects. Hospital admission rates were considered because several Canadian studies used this endpoint and because hospitalization studies capture a larger segment of the affected population compared to mortality studies. Hospital data in Canada are extensive and reliable due to the nationalized health care system. In general, the data for these two endpoints should be much more reliable than subjective measures of $PM_{2\cdot5}$ impact such as absenteeism, restricted activity and prevalence of respiratory symptoms (e.g., asthma), which are often obtained through surveys.

The CRF for mortality was determined after review of many studies, including those involving PM_{10} or other particle-related measures such as particulate sulfate concentration and British smoke shade or black smoke (BS). Due to the availability of concentration measurements, there were more studies involving PM_{10}. Thus, a meta-analysis approach was used to determine the CRF. Though $PM_{2\cdot5}$ is suspected to be a stronger indicator for human health effects, there were a limited number of studies. Therefore, the $PM_{2\cdot5}$ CRF for mortality came from one publication summarizing/combining the associations (mortality vs. $PM_{2\cdot5}$) among six US cities (Schwartz *et al.*, 1996). Since this study included multiple cities, it was expected to provide a representative CRF, applicable to Canadian populations. The reliability of this study was demonstrated by its PM_{10} results, which were consistent with a wide range of other studies.

The CRF for hospital admissions was derived from one Canadian study (Burnett *et al.*, 1995), which involved a large population residing in Ontario. Like the selected mortality study, the PM_{10} and $PM_{2\cdot5}$ relationships derived from this study were consistent with the results found in other investigations for other locations. The values of the CRFs used for the Canadian exercise are given in Table 2.

As discussed above, a NOAEL could not be determined because the epidemiological studies found an increasing effect throughout the entire range of PM concentrations. However, from the studies providing the CRFs it was possible to identify a concentration below which data limitations lead to less statistical confidence in the health effect relationship. This involved examining the relationship at the lower end of the concentration distribution by determining average mortality or hospitalization rates (suitably adjusted for temporal and weather factors) for selected ranges of PM concentrations such that these average rates increased with increasing concentration. Statistical tests were then used to determine the concentration at which the rates were different from the rates associated with the next lower concentration range. The lowest concentration at which a statistically significant increase in admission rate was observed was denoted as the LOAEL. Thus, the LOAEL is the level above which there are demonstrated effects on human health. It is scientifically based and provides a basis for establishing goals for long-term air quality management. This *statistical threshold* does not represent a threshold in the classic sense and the shape of the relationship could be linear below the LOAEL, but the benefits of achieving air concentrations below that level can not be calculated with confidence.

Table 2 Concentration-response functions (% change) associated with a $1\,\mu g\,m^{-3}$ change in 24 h average $PM_{2\cdot5}$ concentration

Health endpoint	CRF (95% confidence interval)
Mortality	0.14 (0.18, 0.10)
Respiratory hospital admissions	0.074 (0.099, 0.049)
Cardiac hospital admissions	0.07 (0.10, 0.036)

Figures 2(a,b) provide an example of how the LOAEL was identified from the Canadian hospitalization data (Burnett *et al.*, 1995). First all PM measurements were grouped into concentration ranges. In Figure 2(a), which shows 24 h average $PM_{2\cdot5}$ calculated from sulfate, four groupings were found to best represent the relationship because of the clear increases in respiratory and cardiac hospitalizations from group to group. The difference in the hospitalization rate between each group is shown in Figure 2(b). The rates within the second group, which has a 24 h mean $PM_{2\cdot5}$ of $15\,\mu g\,m^{-3}$, are significantly different from the first group because the confidence interval (± 2 standard errors) does not include zero. Sub-dividing the lowest concentration group further and comparing hospitalization rates amongst these groups showed that there were no statistically significant changes in rates at these lower concentrations. Thus, the average concentration of the second group was taken as the LOAEL (i.e., $15\,\mu g\,m^{-3}$). Similar analyses involving the concentration-mortality functions derived from the six US cities, yielded a similar value for the LOAEL. The first standard setting approach listed above was therefore based upon a LOAEL of $15\,\mu g\,m^{-3}$ (24 h average). The second approach, which will not be discussed further, was also based upon this number, but with adjustments allowing for a margin of safety.

The third approach for development of standards involved examining the health benefits, or avoided impacts, of achieving various 24 h $PM_{2\cdot5}$ concentration targets. The steps involved in this process are outlined in Figure 3. First, target levels are identified and the current/recent ambient measurements in the region of interest (e.g., Canada) are compared to this value. New distributions of ambient concentrations are predicted using various rollback algorithms until all or a given percentage of the measurements are below the proposed target (Figure 3a). In Figure 3(a), the $PM_{2\cdot5}$ concentration on all days with levels above the

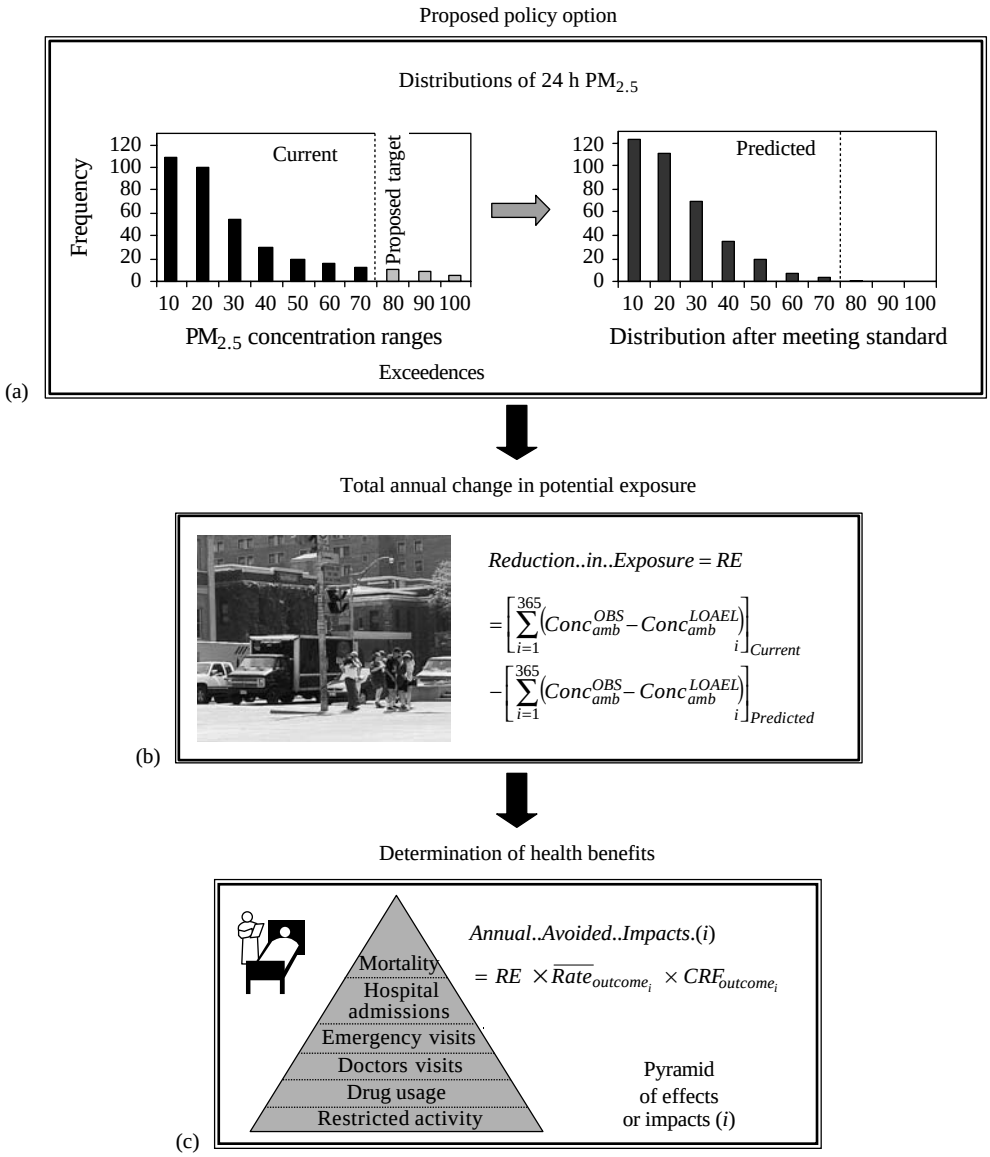


Figure 3 Steps involved in estimating the health benefits of a proposed air quality standard for $PM_{2\cdot5}$. (a) The current observed annual frequency distribution in 24 h concentrations ($\mu g\,m^{-3}$) is modified in a realistic fashion so that concentrations remain below the proposed standard (predicted distribution). (b) The total reduction in exposure (RE) over an annual time period is calculated by summing all concentrations above the LOAEL from the current and predicted $PM_{2\cdot5}$ distributions. (c) The RE is multiplied by the CRF (percent change in health outcome per 1 $\mu g\,m^{-3}$ change in $PM_{2\cdot5}$) and the baseline rate in the outcome (per a given population size) to estimate the annual health benefit (avoided outcome) for the given population size represented by the baseline rate

estimated background concentration decrease (i.e., not only high days with concentrations above the target levels are affected). All days with concentrations greater than the background concentration are assumed to have an impact on health, thus reducing those concentrations is expected to result in Avoided Impacts (Figure 3c).

Given a linear CRF throughout the concentration range, the total annual reduction in $PM_{2\cdot5}$ exposure can be calculated (Figure 2b) separately and then multiplied by the CRF to determine the Annual Avoided Impacts. By definition, it is not possible to reliably calculate benefits when concentrations are below the LOAEL ($Conc_{amb}^{LOAEL}$). Thus, the reduction in exposure (RE) in Figure 3(b) is calculated relative to the LOAEL (a negative difference is set to zero).

Annual Avoided Outcomes can be calculated for a range of health effects, such as those in the Health Effects Pyramid in Figure 3(c). The Canadian exercise involved mortality and hospitalization using CRFs from Table 2. This could be done for additional outcomes such as emergency department visits if reliable CRFs are available.

SUMMARY

Urban air quality is generally thought of as a local human health issue. However, pollutants emitted and building up in urban areas spread throughout the surrounding environment having impacts beyond the local area, ultimately changing the mix of trace gases and fine particles on a global scale. These changes lead to global environmental change such as climate warming due to increasing carbon dioxide and regional-scale cooling due to the direct and indirect radiative effects of aerosols. Combustion is a key source of both carbon dioxide and aerosols, as well as the main smog pollutants. Therefore, the consequences of policy decisions to minimize climate change and to reduce urban air pollution (or lack of such decisions) are strongly linked.

The key policy tool to address urban air quality is the air quality standard. Setting a standard is a complex and iterative process. Application of epidemiological information can be used to support this process. This scientific input to standard setting is weighed along with socio-economic information, technical issues and politics. For this reason it is very important that decisions regarding urban air quality policies are seen in the context of long-term changes in global climate as well as their effectiveness in avoiding or reducing adverse public health impacts.

REFERENCES

Burnett, R T, Dales, R, Krewski, D, Vincent, R, and Brook, J R (1995) Associations Between Ambient Particulate Sulfate and Admissions to Ontario Hospitals for Cardiac and Respiratory Diseases, *Am. J. Epidemiol.*, 142, 15–22.

Burnett, R T, Cakmak, S, and Brook, J R (1998) The Effect of the Urban Ambient Air Pollution Mix on Daily Mortality Rates in Canadian Cities, *Can. J. Publ. Health*, 89, 152–156.

FPWG (Federal/Provincial Working Group on Air Quality Objectives and Guidelines) (1999) National Ambient Air Quality Objectives for Particulate Matter: Addendum to the Science Assessment Document, Ottawa, Public Works and Government Services Canada, www.hc-sc.gc.ca/ehp/ehd/catalogue/bch.htm

Schwartz, J, Dockery, D W, and Neas, L M (1996) Is Daily Mortality Associated Specifically With Fine Particles? *J. Air Waste Manage. Assoc.*, 46, 927–939.

Taubes, G (1995) Epidemiology Faces its Limits, *Science*, 269, 164–169.

FURTHER READING

Aunan, K (1996) Exposure-response Functions for Health Effects of Air Pollutants Based on Epidemiological Findings, *Risk Anal.*, 16, 693–709.

Bass, B and Brook, J R (1997) The Derivation of High-resolution Meteorological Data Sets: a Review of Downscaling Procedures for Integration of Multiple Air Issues, *Environ. Monit. Assess*, 46, 151–174.

Gold, D R, Litonjua, A, Schwartz, J, Lovett, E, Larson, A, Nearing, B, Allen, G, Verrier, M, Cherry, R, and Verrier, R (2000) Ambient Pollution and Heart Rate Variability, *Circulation*, 101, 1267–1273.

Working Group on Public Health and Fossil-Fuel Combustion (1997) Short-term Improvements in Public Health from Global-climate Policies on Fossil-fuel Combustion: an Interim Report, *Lancet*, 350, 1341–1349.

World Health Organization (1999) Air Quality Guidelines, http://www.who.int/peh/air/airqualitygd.htm.

User Pay Principle

see Eco Taxes (**Volume 4**)

V

Valued Ecosystem Component

The phrase *valued ecosystem component* (VEC) first appeared in the environmental assessment (EA) literature in a Canadian work by Beanlands and Duinker (1983, 1984). It was defined as environmental attributes considered to be important in project decisions (Beanlands and Duinker, 1983). Synonymous is the term *indicator* as used earlier by Holling (1978) who defined indicators as measures of system behavior in terms of meaningful and perceptible attributes. Also equivalent in meaning is the term *environmental performance indicator* as used by Duinker and Baskerville (1986). VECs, like the related concept of *ecosystem values* (Duinker and Doyon, 1999), can refer to physical things (both biotic, such as songbirds or moose, and abiotic, such as water or air), ecological processes (e.g., nutrient ow, carbon sequestration), or ecosystem conditions (e.g., degree of soil compaction, biodiversity). The VEC concept was developed in recognition that ecosystems are comprised of too many components for EAs to be assimilated successfully, considering our poor understanding of most ecosystems, budget constraints and the short timeframes imposed on assessors. Thus, EA practitioners need to set priorities on which few components will garner most or all of their analytical attention. Criteria for setting such priorities should include a balance of social, economic and ecological factors, with input from all affected and interested parties (Beanlands and Duinker, 1983). Some EA practitioners prefer to use the term *valued environmental component* (with the same acronym) in a effort to apply the concept equally to ecological, social and economic elements in a broad interpretation of environment.

REFERENCES

Beanlands, G E and Duinker, P N (1983) *An Ecological Framework for Environmental Impact Assessment in Canada*, Institute for Resource and Environmental Studies, Dalhousie University, Halifax and Federal Environmental Assessment Review Office, Hull, 132.

Beanlands, G E and Duinker, P N (1984) An Ecological Framework for Environmental Impact Assessment, *J. Environ. Manage.*, 18, 267–277.

Duinker, P N and Baskerville, G L (1986) A Systematic Approach to Forecasting in Environmental Impact Assessment, *J. Environ. Manage.*, 23, 271–290.

Duinker, P N and Doyon, F D (1999) *Biodiversity Assessment in Forest-management Planning: What are Reasonable Expectations*? Proceedings of a Symposium on Ecosystem Management of Forested Landscapes, UBC Press, Vancouver, in press.

Holling, C S, ed (1978) *Adaptive Environmental Assessment and Management*, John Wiley & Sons, Chichester.

PETER N DUINKER *Canada*

Vavroušek, Josef

(1944- 1995)

Josef Vavroušek was a Czech environmentalist, scientist and politician, and founder of the Environment-for-Europe process. As a student in Prague in the early 1960s, Vavroušek was one of leaders of the student opposition movement. After the Soviet occupation of Czechoslovakia in 1968 he became Vice-president of the Czech and Slovak young people's organization S.O.S. in Paris (August–September 1968). He obtained his PhD (equivalent) from the Czech Technical University in Prague in 1975 then worked for 15 years (1975–1990) as a scientist in the field of cybernetics, general systems theory, science of science, and human environment, specializing in macro-ecology of Man and environmental policy. During this period he was a member of the executive body of the Ecological Section of the Biological Society, Czechoslovak

Academy of Sciences (1985–1991), one of the founding members of the first executive body of the Circle of Independent Intelligentsia (1988–1989) and a member of the Club of Rome. He was also one of the founders of the Civic Forum and later of the Civic Movement in Czechoslovakia and a member of their executive bodies.

In 1989 Vavroušek was one of the leaders of the Czechoslovakian Velvet Revolution. In April 1990 he became vice-chairman of the State Commission for Scientific and Technical Development (responsible for the environment), and then in June 1990 he became the first (and last) environment Minister of the Federal Government of Czechoslovakia. It was in this role that Vavroušek proposed and organized the first pan-European conference of environment Ministers. In Vavroušek's own words:

> ••• there are not many fields where effective co-operation in continental or even global scale is so necessary, like in the field of environmental protection and restoration. Not a single group of countries can solve the severe legacy of environmental deterioration and find ways towards a sustainable future alone. Environmental co-operation is thus not only the necessary condition for substantial improvements of our ecological situation but also the most natural first step of our European integration.

The conference was held at Dobříš Castle near Prague in June 1991 and was attended by environment Ministers or deputies from almost all European countries together with representatives from many international bodies and non-governmental organizations. The conference initiated an Environmental Program for Europe and a pan-European environmental reporting process (*see* Dobr s European Environment Assessment Process, Volume 4). Vavroušek's initiative and vision thus set the basic framework of long-term European environmental cooperation which has continued with Ministerial Conferences in Lucerne (April 1993), Sofia (October 1995), Aarhus (June 1998) and Kiev (2003).

As environment Minister he headed the Czechoslovakian delegation to the Rio Summit in 1992, but then political changes and the division of the country brought his ministerial career to a premature end. Returning to science he joined the Institute of Applied Ecology at Charles University and then went on to found and become president of the Society for Sustainable Living (established in October, 1992). Vavroušek's contribution to European environmental cooperation was to bring the interdisciplinary experience and understanding of how human systems and the environment work into mainstream political parlance across the new Europe at the time of its birth and to emphasize the importance of human values and environmental ethics in the search for sustainable ways of living. Vavroušek was killed with his daughter Petra by an avalanche in March 1995 while hiking in the Slovak mountains.

DAVID STANNERS *Denmark*

Villach Conferences

In the 1980s several scientific conferences were held, in the pretty resort town of Villach, Austria. These meetings of experts were organized jointly by the World Meteorological Organization (WMO), the United Nations Environment Programme (UNEP) and the non-governmental International Council for Science (ICSU).

The first was in November 1980 under the framework of the then new World Climate Programme (WCP). Its focus was do we have a problem due to rising atmospheric concentrations of carbon dioxide? The scientists concluded that continued consumption of fossil fuels could increase global mean temperature by 1 °C or more early in the 21st century, and continue thereafter. They anticipated that regional ecosystems, agricultural production and water supply will be materially affected . The main recommendation was that a major interdisciplinary research program should be undertaken urgently, and a five-point research agenda was proposed.

Probably the most important and in uential Villach Conference was that held in October 1985, sponsored by the same three organizations. WMO Publication #661 contains the *Report of the International Conference on the Assessment of the Role of Carbon Dioxide and of Other Greenhouse Gases in Climate Variations and Associated Impacts*. This conference had the benefit of a new report prepared by the International Meteorological Institute (IMI), Stockholm and led by Professor Bert Bolin. Among the key findings were that other greenhouse gases, particularly nitrous oxides, methane, ozone and chloro uorocarbons, were estimated to have added one half again as much temperature change as carbon dioxide (CO_2) alone. In addition, the concentrations of some of these gases were increasing more rapidly than CO_2. The IMI report stated that the expected global mean temperature increase due to a double CO_2 equivalent, expected in the 21st century, is of about the same magnitude as the change of global temperature from the last glacial period to the current interglacial . The potential sea level rise in such a scenario was estimated at 20–165 cm. Likely consequences for agriculture and forestry were emphasized.

The experts from 29 countries at the Villach Conference reviewed and debated the IMI findings and produced their own hard-hitting statement. It said that combining the other greenhouse gases with CO_2 meant that, it is now believed that in the first half of the 21st century, a rise of global mean temperature could occur which is greater than any in Man's history.

The conference statement goes on to say:

> Many important economic and social decisions are being made today on long-term projects – irrigation, hydro-power, drought

(management), agricultural land use, structural design, coastal engineering and energy planning – all based on the assumption that past climatic data, without modification, are a reliable guide to the future. This is no longer a good assumption.

These conclusions reverberated in a number of national capitals and stimulated a far greater concern on the part of environment and energy agencies as well as key business sectors. As follow-up, the Conference requested the WMO, UNEP and ICSU to establish a small task force on greenhouse gases. This was done in the form of an international Advisory Group on Greenhouse Gases of seven eminent scientists. Subsequent developments in international scientific assessments are discussed in the article Intergovernmental Panel on Climate Change (IPCC): an Historical Review, Volume 4.

The scientific community through the Villach conferences, coordinated by UNEP, WMO and ICSU, thus set the stage for the formation of IPCC in 1988, the Second World Climate Conference 1990, the signing of the UN Framework Convention on Climate Change in 1992, and the Kyoto Protocol to the Convention in 1997. While environmentalists may feel that action on this issue has been slow, others may see the rate at which the world has responded as remarkably swift for such a socially and economically profound issue.

The author was chair of Villach 1985.

JIM BRUCE *Canada*

Volvo Environment Prize

Mostafa K Tolba
International Center for Environment and Development, Cairo, Egypt

Volvo established the Volvo Prize Foundation to promote - without pro t motive by an individual or organization - scienti c research in the eld of environmental protection through the annual award of a prize, designated the Volvo Environment Prize, to an institution, organization, individual or other recipient that has demonstrated outstanding scienti c innovativeness in the area of environmental protection.

The Volvo Prize is one of the major environmental prizes. The Volvo Corporation saw that the threats posed by pollution of the biosphere and the loss of its resources are evident today and all human activity has an impact on the environment. The landscape is changed, the quality of air and water deteriorates, and ef uents and noise increase. Over the past few decades, a consensus has been spreading across the social, industrial and political spectrum about the urgent need to show greater care for Earth's ecosystems and their management.

Volvo also felt that the reality of environmental protection demands a system approach involving the entire life cycle of a product from conception to design, production, use and, finally, recycling or disposal. It is only by taking such a global view of the problems that we will be able to solve them effectively. Deep and current knowledge is essential to be able to choose the right priorities in the environmental field. The new results that research is presenting are therefore a necessity for both industry and society in general. By presenting an annual environmental prize, Volvo wants to attract attention to extraordinary research efforts and encourage scientists whose work and discoveries are of decisive importance to our future. The creation of the Volvo Environmental Prize was meant to be an expression of Volvo's awareness of the interdependence of all actions, inventions and processes that sustain and protect the environmental base.

A comprehensive nomination and selection procedure has been developed to ensure that those who are awarded the Volvo Environment Prize meet the requirement of having made outstanding innovations or discoveries – scientific, socio-economic or technological – which have direct or indirect significance in the environmental field and are of global or regional importance.

The Prize rewards achievement, and priority is given to an individual or group of named individuals rather than to an institution.

Employees or companies within the Volvo Group, including parties who performed research on behalf of the Volvo group, are not eligible to receive the award.

The Foundation capital now totals SEK 20 million allowing a total annual prize sum of SEK 1.5 million. The Volvo Environment Prize is awarded annually to one or more individuals or organizations that the international Prize Committee deems to have made an outstanding innovation or discovery in the environmental field, e.g., in areas such as environmental control, pollution, hazardous waste, food, agriculture or nature and wildlife.

Candidates for the Volvo Environment Prize are nominated by universities, research institutes and their staffs, by scientists and engineers in the fields of environmental protection, life sciences, earth sciences and social sciences, by national academies of science and their staffs as well as by other persons and organizations, all over the world.

The Scientific Committee makes an initial screening and evaluation of candidates and presents its findings and a number of options to the Prize Committee. The

responsibility for selecting the laureate or laureates of the annual Volvo Environment Prize for recommendation to the Board of the Foundation lies with an international Prize Committee.

Matters relating to the Volvo Environment Prize are handled by Mr Claes Beyer, Acting Member of the Board of the Volvo Environment Prize Foundation, and Dr Lena Gevert, Director Environmental Affairs AB Volvo.

Further information on nomination procedures, as well as nomination forms, can be obtained from: The Volvo Environment Prize Foundation, Ms Kikki Hugestrand, Volvo Event Management, S-405 08 Goteborg, Sweden. Tel.: • 46-31-59-66-77; Fax: • 46-31-53-84-31, E-mail: vhk.huge@memo.volvo.se.

See also: International Environmental Prizes, Volume 4.

WZ

Ward, Barbara

(1914- 1981)

A British economist and author, Barbara Ward was among the first persons to integrate a concern for international development and poverty reduction with a concern for environment and sustainable development. From the late 1940s, she was among the most persuasive supporters of international development through her books, her writing for *The Economist*, her teaching at Harvard University and later as the Albert Schweitzer Professor of international economic development at Columbia University. From the late 1960s, she brought to her concern for development a growing concern about the environment. Amongst her many books, perhaps *Only One Earth: Care and Maintenance of a Small Planet* (co-authored with René Dubos) and published in 1972 was the most in uential. Written to bring the latest environmental knowledge to a popular audience, the book describes the concerns that had led to the organization of the United Nations Conference on the Human Environment in Stockholm in 1972. This is among the first books to draw together a concern for care and maintenance of the environment with a commitment to meeting human needs. Its contribution to the definition of sustainable development can be seen in its statement, 15 years before the Brundtland Commission's report *Our Common Future*, that the charge of the United Nations to the (Stockholm) Conference was to clearly define what should be done to maintain the Earth as a place suitable for human life not only now, but also for future generations (page 25). Her later books (*The Home of Man* published in 1976 just before the United Nations Conference on Human Settlements in 1976 and *Progress for a Small Planet* published in 1979) amplified and developed this theme. She also chaired the symposium in Mexico in 1974, organized by the newly formed United Nations Environment Programme and the United Nations Council on Trade and Development (UNCTAD) which produced the Cocoyoc Declaration; this declaration and her report of the Cocoyoc meeting were in uential in emphasizing the need for development (and international development agencies) to be more environmentally responsible (*see* UNCTAD (United Nations Conference on Trade and Development), Volume 4). At that time, few development assistance agencies had made any provision to consider the environmental implications of what they funded.

For the last decade of her life, Barbara Ward devoted much of her time to building the International Institute for Environment and Development (IIED) with David Runnalls. She saw IIED as the means of ensuring that her concerns about the environment and her strong commitment to social justice would outlive her. She made IIED an independent policy research institute whose task was to help set the policy agenda for environment and development for governments and for international agencies. Within this institute, she helped set new precedents for North–South partnerships, much of its work being of the form of collaborative research with groups in Africa, Asia and Latin America. She also recognized the need to combine rigorous research with strong public advocacy, as can be seen in the creation of *Earthscan* in 1976, initially as a media information unit within IIED, before developing into a leading independent publisher of environment and development books.

Barbara Ward's in uence was not only through her writing but also through her public speaking, her contributions to radio and television programs and her capacity to cajole and persuade those she advised. She was a very gifted public speaker. Her obituary in *The Economist* noted that she had been

> a conscience and provider of fresh ideas, a spur and enthuser, a phrasemaker, provoker and counsellor to one pope and handfuls of cardinals and archbishops, to two American presidents and handfuls of their senior advisors, to British and countless other prime ministries and ministers of rich and poor governments alike, to successive presidents of the World Bank •••

She was knighted and also given a life peerage, becoming Baroness Jackson of Lodsworth, DBE.

DAVID SATTERTHWAITE *UK*

Watson, Robert T

(1948-)

Robert T Watson is Director for Environment and Head of the Environment Sector Board at the World Bank. In September 1997, he became the Chairman of the Intergovernmental Panel on Climate Change (IPCC).

In May 1996, Dr Watson joined the World Bank as Senior Scientific Advisor at the Environment Department. Before joining the Bank, he was Associate Director for the Environment in the Office of Science and Technology Policy in the Executive Office of the President in the White House. Prior to joining the Clinton White House, Dr Watson was Director of the Science Division and Chief Scientist for the Office of Mission to Planet Earth at the National Aeronautics and Space Administration (NASA).

Dr Watson has played a key role in the negotiation of global environment conventions and the evolution of the Global Environment Facility (GEF). At the critical stages of the negotiations of the Montreal Protocol he led a meeting of Ozone Modeling Groups which settled the argument over whether the stratospheric ozone is being depleted or not. He then chaired the extremely important and powerful Montreal Protocol assessment panels during the period that saw huge participation in the Protocol and the establishment of the Multilateral Fund.

From 1991 to 1994, he served as Chairman of the GEF's Scientific and Technical Advisory Panel. He was Chair or Co-Chair of a number of international scientific assessments, including IPCC Working Group II, the United Nations Environment Program-World Meteorological Organization (UNEP-WMO), International Scientific Assessment of Ozone, and UNEP's Global Biodiversity Assessment. Dr Watson has testified in the US Congress on numerous occasions regarding global environmental issues.

Dr Watson received his PhD in Chemistry from London University in 1973 and has received many national and international awards and prizes for his contributions to science, including the American Association for the Advancement of Sciences Award for Scientific Freedom and Responsibility in 1993.

Photo by Nicholas Sundt.

MOSTAFA K TOLBA *Egypt*

WBCSD (World Business Council for Sustainable Development)

The WBCSD (see Figure 1 for logo), is the pre-eminent business voice on business and sustainable development issues. It is a coalition of some 150 international companies united by a shared commitment to sustainable development. Its mandate is to develop closer cooperation between business, government and all other organizations concerned with the environment and sustainable development; to encourage high standards of environmental management in business itself; and to ensure that the right framework conditions are put in place so that business can operate efficiently and make faster progress toward sustainable development.

The WBCSD focuses on: eco-efficiency; sustainability through the market; corporate social responsibility; climate and energy; innovation, technology, society and sustainability; and natural resources. It also benefits from a learning-by-sharing cross-cutting program. Further, a scenario program specializes in synthesizing alternative futures, helping companies develop strategic responses to the sustainable development issues that may affect their business.

With respect to global environment change, the WBCSD has helped create a paradigm shift in the way in which business does business – from a fractured view of environment and development issues to a holistic view of business and sustainable development. Further, the WBCSD has changed the perception of society toward business which is now recognized as a solution provider for sustainable development. Importantly, the WBCSD has initiated a culture of stakeholder dialogues with all constituencies of society. It is also contributing to sectoral industry initiatives to generate momentum for a change in policy and behavior.

The WBCSD currently has five member-led sectoral initiatives: mining and minerals, cement, sustainable mobility, forestry, and sustainability in the electricity utility industry. Members are drawn from 30 countries and 20 major industrial sectors, and the WBCSD also benefits from a global network of national and regional business councils and partner organizations located in the developing world. The WBCSD is governed by a council, composed of the

World Business Council for Sustainable Development

Figure 1 The logo for the WBCSD

chief executive officers of its member companies, which meets once a year. Between council meetings, WBCSD's activities are managed by an Executive Committee and administered by the President, whose office is at 4, chemin de Conches, CH-1231 Conches-Geneva, Switzerland. Fax: • 41 22 839 3100; E-mail: info@wbcsd.org.

BJORN STIGSON *Switzerland*

WCED (World Commission on Environment and Development)

How far have we advanced in achieving a common understanding of environmental issues? Where are we heading? And what are our ultimate goals? These were some of the questions raised at the celebrations of the tenth anniversary of the Stockholm conference that took place at the headquarters of the United Nations Environment Program (UNEP) in Nairobi in 1982. In an endeavor to provide answers to these questions, the Governing Council (GC) of UNEP decided to prepare an Environmental Perspective to the Year 2000 and Beyond. For this purpose, the Council established its own intergovernmental preparatory committee. A Special Commission, comprised of eminent international figures, would assist the GC in this exercise, in addition to its work in addressing the broader issue of environment in the global *problematique* (GC decision 11/3 of 1983). The General Assembly (GA) of the United Nations endorsed the recommendations of the GC (GA resolution 38/161 of the 19th of December 1983), and suggested that the Special Commission (which became known afterwards as the WCED) should focus mainly on the following terms of reference for its work:

1. To propose long-term environmental strategies for achieving sustainable development to the year 2000 and beyond.
2. To recommend ways in which concern for the environment may be translated into greater cooperation among developing countries and between countries at different stages of economic and social development and lead to the achievement of common and mutually supportive objectives which take account of the interrelationships between people, resources, environment and development.
3. To consider ways and means by which the international community can deal more effectively with environmental concerns, in the light of the other recommendations in its report.
4. To help to define shared perceptions of long-term environmental issues and of the appropriate efforts needed to deal successfully with the problems of protecting and enhancing the environment, a long-term agenda for action during the coming decades, and aspirational goals for the world community, taking into account the relevant resolutions of the session of the special character of the GC of UNEP in 1982.

The WCED was composed of 21 members from different countries appointed in their individual capacities. Mrs Gro Harlem Bruntland of Norway was appointed chairperson of the Commission, and Dr Mansour Khalid of Sudan was appointed vice-chairman.

The work of the WCED began in May 1984 and in July 1984 a WCED Secretariat was established in Geneva. Dr Jim MacNeill (Canada) was appointed as secretary general, and the secretariat consisted of 15 professional staff members. In October 1984, the WCED selected eight key issues for analysis during the course of its work:

- perspectives on population, environment, and sustainable development;
- energy, environment and development;
- industry, environment and development;
- food security, agriculture, forestry, environment and development;
- human settlements, environment and development;
- international economic relations, environment and development;
- decision support systems for environmental management;
- international cooperation.

The WCED held a number of deliberative meetings and/or public hearings in all regions of the world to get a first-hand view of environment and development issues in those regions. The work of the WCED was carried out by special working groups and advisory panels. The commission also engaged experts, research institutes, and academic centers of excellence from around the globe to prepare more than 75 studies and reports relating to the above-mentioned eight issues for the Commission's review and consideration. In April 1987, the WCED published its report under the title *Our Common Future*, familiarly known as the Bruntland report (WCED, 1987).

REFERENCE

WCED (1987) *Our Common Future*, Oxford University Press, Oxford.

ESSAM EL-HINNAWI *Egypt*

WFC (World Food Council)

The WFC was established on 17 December 1974 by the General Assembly resolution 3348 (29) on the recommendation of the World Food Conference held in Rome in November 1974, as a United Nations (UN) organ reporting to the General Assembly through United Nations Economic and Social Council (ECOSOC). Its mandate was to review periodically major problems and policy issues affecting the world food situation and the steps being proposed or taken to resolve them by governments, by the UN system and its regional organizations and to recommend remedial action as appropriate. The Council consisted of 36 members nominated by ECOSOC and elected by the General Assembly with due consideration to balanced regional representation.

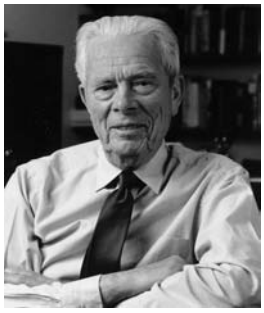
Until 1993, the WFC had its main headquarters in Rome and was serviced by an executive director and a small secretariat. Following the adoption of General Assembly resolution 47/212 (B) on May 6th 1993, the post of executive director was disestablished and the functions and activities previously carried out by the WFC Secretariat were transferred to the UN Department for Policy Coordination and Sustainable Development in New York.

On July 1st 1996 General Assembly resolution 50/227, implementing measures for the restructuring and revitalization of the UN in the economic, social and related fields, was adopted. Paragraph 72 concerns WFC: The functions of the World Food Council shall be absorbed by the Food and Agriculture Organization of the United Nations and the World Food Program, and consequently the World Food Council is discontinued .

HENRI CARSALADE *Italy*

White, Gilbert

(1911-)



Gilbert F White (born 1911) is one of the world's best known and most admired environmental scientists. For over half a century he has initiated new thinking across a wide range of fields including water supply, ood hazards, river basin development, arid zone problems, energy policy, natural resource management and decision-making, environmental perception and behavior, and global environmental change. His geography PhD dissertation *Human Adjustment to Floods* (University of Chicago, 1942) set in motion a radical reassessment of public policy to encourage more appropriate human use of hazardous areas as opposed to the reliance on engineering works to protect them. The concept of human adjustment has served as a practical guide to action; as for example in the new disaster mitigation policies of the US Federal Emergency Management Agency, as a widely applied research paradigm, and as an early aspiration for human reconciliation with the natural world.

White's first positions were as geographer with the Mississippi Valley Committee, the National Resources Committee, and the National Resources Planning Board in President Franklin Roosevelt's Administration. As a Quaker and a conscientious objector in the Second World War he served with the American Friends Service Committee in France and Germany, where he was interned and spent 1 year in detention. The program in which he participated was awarded a Nobel Peace Prize.

After the war White served as President of Haverford College until 1956 when he returned to Chicago as Chair of the Department of Geography. He remained at Chicago until moving in 1969 to the University of Colorado at Boulder to assume the Directorship of the Institute of Behavioral Science. At Colorado, he is currently the Gustavson Distinguished Professor Emeritus of Geography and continues to be actively and widely engaged in global environmental issues, and to be involved with the Natural Hazards Research and Applications Information Center, which he founded in 1975.

Some of White's most important international contributions to environmental science were made during his two-term presidency (1976–1982) of the Scientific Committee on Problems of the Environment (SCOPE), a body formed in 1969 as part of the International Council for Science (ICSU) (*see* ICSU (International Council for Science), Volume 4). Under White's guidance, SCOPE began to address environmental problems at a global scale, and established itself as a reliable source of scientific advice for governments and organizations concerned with the quality of the global environment. With White's insight and leadership, SCOPE laid the scientific groundwork for the worldwide interest in global environmental issues. In particular, SCOPE's work with biogeochemical cycles provided the scientific basis for the International Geosphere Biosphere Program that has engaged scientists from many countries. White has also been especially in uential in fostering a partnership of effort between social scientists and natural scientists. During the Cold War years, Gilbert White maintained strong links with Russian scientists which helped to keep open the doors of communication through some difficult times. Russian collaboration was important

in the International Geographical Union's Commission on the Environment and its international project to study and compare adjustments with natural hazards.

Gilbert White is also known for his tireless efforts to apply his scientific knowledge to the promotion of peace and the resolution of con icts. Such motivation lay behind his chairmanship of the Ford Foundation Project on the Lower Mekong River (1964) and subsequent attempts to bring warring nations to the table in the interests of improved river basin management. As President of SCOPE, White raised foundation funds and organized a coalition of world scientific organizations to study (between 1983 and 1988) the environmental effects of nuclear war (the so-called *nuclear winter hypothesis*) and served on the international Steering Committee throughout the study. Most recently, he has helped to bring scientists from Israel, Palestine, and Jordan together in a commonly agreed and accepted study of the waters of the Jordan River.

Gilbert's late wife, and frequent collaborator, Anne White joined with him and David Bradley, of the London School of Tropical Hygiene and Medicine, in a landmark study of the decisions made in the procurement and use of domestic water supplies in East Africa. Published under the title *Drawers of Water* in 1972, the research changed government attitudes towards the quantity and quality of water investments for millions of people. It was certainly the most in uential research work I have done in terms of affecting the lives of people, says White. This study is now being repeated to assess the changes a generation later.

The range and depth of White's public and scholarly contributions is evident not only in his own over 300 publications, and in the large number of leadership roles he has played in many diverse organizations (including the US National Research Council, UNESCO, ICSU, the International Geographical Union, The International Social Sciences Council, Resources for the Future, and the American Friends Service Committee), but also in his receipt of many prizes and awards, and in what others have written in honor of his work. A *festschrift* (Kates and Burton, 1986) includes 13 essays by authorities in fields to which he has contributed, and a companion volume includes 29 of White's most in uential papers.

Anne once wrote of her husband: What makes him keep at this? I think it is all tied together by his Quaker faith in the ability of humans to marshal their inner resources to deal competently and lovingly with the outer world and with their fellow human beings. Then there is the real fun he gets out of the exchange of ideas with others, and the challenge of making real friends, not just co-workers, out of those others. And not least, there is his innate and humble desire to leave the world a bit better place than he found it.

In May 2000, Gilbert White was awarded the Public Welfare Medal of the US National Academy of Science, and in December 2000 President Clinton presented him with the National Medal of Science.

REFERENCE

Kates, R W and Burton, I, eds (1986) *Geography, Resources, and Environment*, University of Chicago Press, Chicago, IL.

FURTHER READING

Three recent publications provide insights from White's long carrer.

White, G F (1994) Re ections on Changing Perceptions of the Earth, *Annu. Rev. Energy Environ.*, 19, 1–13.

White, G F (1996) Emerging Issues in Global Environmental Policy, *Ambio*, 25(1), 58–60.

White, G F (1998) Re ections on the 50-Year International Search for Integrated Water Management, *Water Policy*, 1, 21–27.

IAN BURTON *Canada*

WHO (World Health Organization)

The World Health Organization (WHO) was founded in 1948, as a specialized agency of the United Nations (UN). WHO is the directing and coordinating authority on international health work, with the aim of the attainment by all peoples of the highest possible level of health. It has 191 member states, whose interests are represented in the World Health Assembly (WHA), which is the decision-making authority, and the secretariat. The WHA meets every year, in Geneva, to decide policy matters, and approve the operational budget of WHO. The WHA is guided in decision making by an executive board composed of 32 individuals technically qualified in the field of health. Each member of the board is nominated by a designated member state and serves a three-year term. The Board meets at least twice a year. The key functionary of WHO is the director-general. Any member state may nominate a candidate, who meets a defined set of criteria, to the position of director-general. The candidate subsequently makes a presentation before the WHA. The position is filled by the candidate elected by simple majority (secret ballot) by members of the WHA.

WHO carries out its mandate through a highly decentralized network comprising six regions. Only 30% of WHO staff work in the headquarters at Geneva. One-quarter of WHO staff work in the regional offices. Each region has a regional committee and a regional office. At the head

of each regional office is a regional director. The regional offices formulate WHO policies to meet their specific needs. The offices function under the following acronyms: PAHO (Pan American Sanitary Bureau), EURO (Regional Office for Europe), AFRO (Regional Office for Africa), EMRO (Regional Office for the Eastern Mediterranean), SEARO (Regional Office for South-east Asia) and WPRO (Regional Office for the Western Pacific).

Within each region, member states frequently have a WHO representative, to support the government in the planning and management of national health programs, and to carry out the offices of WHO. Almost half of all WHO staff work is at the country level.

Health, as defined in the WHO constitution, is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity. Within this context, WHO strives to:

- stimulate the eradication of epidemic, endemic, and other diseases by promoting improved nutrition, housing, sanitation, working conditions, and other aspects of environmental hygiene and by promoting research in the field of health and cooperation among scientific and professional groups which contribute to the enhancement of health;
- provide information, counsel, and assistance in the field of health; and to establish and maintain such administrative and technical services as may be required, including epidemiological and statistical services;
- propose international conventions and agreements on health matters and promote technical cooperation, furnish appropriate health services, technical assistance and, in emergencies, necessary aid, upon the request or acceptance of governments;
- promote improved standards of teaching and training in the health, medical and related professions; establish international standards for biological, pharmaceutical and similar products, and standardize diagnostic procedures;
- foster activities in the field of mental health, especially those activities affecting the harmony of human relations.

WHO leads the world alliance for health for all, an initiative to assure that global resources for health are evenly distributed and that essential health care is accessible to everyone.

Among its major accomplishments have been the eradication of smallpox and yaws, reduction in mortality and increase in life expectancy, delivery of essential drugs, introduction of environmental sanitation measures, and provision of guidelines for healthier cities. WHO is at the threshold of controlling other major diseases such as leprosy and poliomyelitis. WHO monitors diseases with epidemic and pandemic potential such as tuberculosis (TB), human immunodeficiency virus (HIV)/acquired immunodeficiency syndrome (AIDS) and viral hemorrhagic fevers like Ebola. WHO takes an integrated, coordinated approach to prevent, treat, rehabilitate, and where possible, cure major non-communicable diseases such as cardiovascular disease, cancers and respiratory diseases. This involves a systemic approach to disease-specific interventions, by the application of cost-effective methods of detection and management and major sustained global campaigns to encourage healthy lifestyles, healthy public policies and acceleration of research. WHO helps safeguard ethical standards by stimulating discussion and setting global guidelines on topics of biomedical concern, such as cloning, and human drug-testing.

WHO works closely with other organizations in the UN system, and with bilateral agencies, intergovernmental and non-governmental organizations, and with health-related institutions around the world, officially designated as WHO collaborating centers.

WHO recognizes that health, and the pursuit of health for all, can only be assured in healthy environments. As such, much of its work is focused on environmental elements vital to the protection and promotion of good health. These include air quality, drinking water, solid and hazardous wastes, water supply and sanitation. WHO pays special attention to climatic change and such by-products of industrialization as ultra-violet and other radiation, chemical waste production, and electromagnetic fields (EMF). Through its Healthy Cities project, WHO collaborates with local governments and other UN agencies, such as the United Nations Environment Programme (UNEP) and the Food and Agriculture Organization (FAO) (Joint FAO/WHO Meeting on Pesticide Residues (JMPR), 1990), to conduct research into and implement suitable programs for the protection of environmental health. With its partners, WHO produces methodology workbooks for professionals (WHO, 1994), textbooks and guides (Lwanga *et al.*, 1999) for school use, as well as summary reports of its findings (WHO, 1999), from meetings, conferences, research and other professional consultations. These are useful for enhancing public awareness of and investment in the protection of the whole living environment, but also in expanding the knowledge base of professionals across sectors in the interrelation of their work, and in the need for inter-sectoral solutions to health, environmental and other problems.

REFERENCES

Joint FAO/WHO Meeting on Pesticide Residues (JMPR) (1990) *Principles for the Toxicological Assessment of Pesticide Residues in Food*, WHO, Geneva, 1–117.

Lwanga, S K, Tye, C-Y, and Ayeni, O, eds (1999) *Health Statistics: Lesson and Seminar Outlines*, 2nd edition, WHO, Geneva, xii • 1–230.

WHO Environmental Health Series (1994) *No. 170: Assessing Human Health Risks of Chemicals: Derivation of Guidance Values for Health-based Exposure Limits*, Geneva, 1–73.

WHO (1999) *Tobacco-Free Initiative International Consultation on Environmental Tobacco Smoke (ETS) and Child Health*, 11–14 January, WHO, Geneva.

ALEYA EL BINDARI HAMMAD *USA*

WMO (World Meteorological Organization)

see WMO (World Meteorological Organization) (**Volume 1**)

World Business Council for Sustainable Development (WBCSD)

see WBCSD (World Business Council for Sustainable Development) (**Volume 4**)

World Commission on Environment and Development (WCED)

Jim MacNeill
Ottawa, Canada

On October 1st, 1984, the 23 members of the newly appointed WCED sat down together for the first time in Geneva to confront the interlocking issues of development, growth, population, resources and environment. As Secretary General, I thought of it this way: If the world's population doubles over the next 40- 50 years, can we increase food production by a factor of four, energy use by six, and income by about eight (much more evenly distributed, of course), without at the same time pushing the planet beyond certain critical thresholds that we are only now beginning to understand? Or, to put it another way, can growth on the scale needed to meet future needs and aspirations be managed on a basis that is sustainable? I called that the sustainability question.

Three years later, on October 19th, 1987, the Commission presented its report, entitled Our Common Future, *to the United Nations (UN) General Assembly in a special plenary debate (The World Commission on Environment and Development, 1987; MacNeill, 1989). When the debate was scheduled, no one knew that October 19th was to become known as Black Monday, the day the New York Stock Exchange dropped by 502 points. That knocked the debate off the front pages, but it also dramatically underscored the Commission's answer to the sustainability question. The answer was: "Yes ••• but •••." The 'but' conveyed a large number of proposals for fundamental changes in the way we manage the planet.*

One of the familiar quotations of our times is to be found in Our Common Future. *It is the de nition of* Sustainable Development: "*meeting the needs of the present without compromising the ability of future generations to meet their own needs.*"

The Commission's proposals went well beyond the expectations of the General Assembly when, 4 years earlier, in the fall of 1983, it debated and, by consensus, reluctantly adopted a resolution calling for the establishment of a special commission (Resolution of the General Assembly 38/161, United Nations, 1983). The resolution, proposed by the United Nations Environment Programme (UNEP) Governing Council, had been under consideration for more than 1 year. At its 1982 Special Session in Nairobi, the Governing Council looked back 10 years since the landmark Stockholm Conference of 1972 and congratulated itself on a decade of achievement. Many noted, however, that the achievements were concentrated mainly in the richer countries of the Organisation for Economic Cooperation and Development (OECD). They could afford the often high costs of react-and-cure environmental protection strategies. The poorer countries of the developing world could not and, during the decade since Stockholm, they had experienced a steady decline in environmental and resource trends.

Others noted that while environmental protection treated the symptoms of the underlying trends in both rich and poor countries, the sources of the trends in business-as-usual economic development were growing much more rapidly and were reinforcing and accelerating the downward spiral of environmental and resource degradation. The production and consumption patterns of the rich, combined with the pervasive poverty and exploding populations of the poor, put pressures on the environment of an intensity never seen before. Projecting the trends to 2000, it was clear that, without a change of course, environmental degradation, poverty and misery in many countries would get steadily worse.

The General Assembly debate in the fall of 1983 revealed that leaders in many governments had serious reservations

about the establishment of a special commission. Some felt that to support it was to admit that current approaches to managing the environment (and, perhaps, even development) were awed. They were right, of course. Some feared that a commission would come out with yet another study predicting inevitable doom and bewailing growth. They were wrong. Others feared that an independent commission would make recommendations they would not want to support but could not oppose without political cost. They too were right, as the experience of the past 16 years has shown.

The resolution finally adopted by the General Assembly was long, even by UN standards. It was marked by vague have-it-both-ways language, and riddled with convoluted and contradictory compromises. Since any attempt to follow it would have doomed the commission to failure, one of the first things that the Chairman, Gro Harlem Brundtland and I agreed was that we would together reinterpret the resolution to provide the commission with a name and with meaningful terms of reference. We later agreed that the commission would be known as the WCED. We also agreed that the General Assembly had, in effect, asked the Commission to propose a new global agenda for change. The WCED would have three objectives: to take a fresh look at the critical issues of environment and development and bring in some practical recommendations for dealing with them; to propose new forms of international cooperation on these issues; and to raise the levels of understanding and commitment to action by individuals, voluntary organizations, businesses, institutes, and governments. And, finally, we insisted that the General Assembly had agreed that the Commission was an independent body linked to, but outside the control of, governments and the UN system.

GETTING STARTED

In December, 1983, Secretary General Perez de Cuellar invited Mrs Gro Harlem Brundtland (then Leader of the Opposition, later Prime Minister of Norway, and now Director General of the World Health Organization) to be Chairman of the Commission, and Dr Mansour Khalid, one time foreign minister of the Sudan, to be vice chairman. A month later, in January, Mrs Brundtland called me in Paris, where I was Director of Environment for OECD, and asked me to become Secretary General. I agreed to consider it and, in the meantime, to advise her on the organization of the Commission. Subsequently, in consultation with me, she and the vice-chairman appointed the other members of the Commission. We were 23 in all, from 22 different countries, regionally balanced, including all of the major power groups, and with a strong majority of 14 from developing countries.

Before agreeing to accept the Chairman's invitation to become Secretary General, I had to persuade myself that there was a fighting chance to achieve a positive result. Following an OECD meeting in London in March 1984, I retreated to a quiet hotel for five days to think it through. I had already consulted with colleagues in OECD and with top officials in environment, foreign affairs and many other sectors in more than two dozen countries. To clear my mind, I drafted a statement on the key issues confronting the Commission, as I saw them, and on the substance and strategy of the inquiry it should pursue. In the process, I became convinced that, with a bit of luck, a lot of hard work, and about $8 000 000 in voluntary contributions, the task was do-able. After confirming that the concepts in my statement, which I entitled *Mandate for Change*, would enjoy the Chairman's broad support, I agreed to become Secretary General.

MANDATE FOR CHANGE

I put my statement before the Commission at its first full meeting in Geneva in October 1984 (The World Commission on Environment and Development, 1984). Concerning the work program, the essential choice was clear. We could take the easy route and adopt a standard agenda, basically the traditional list of environmental problems, or we could pioneer an alternative agenda integrating development and environment.

Concerning the method of inquiry, the choices were also clear. We must do all of the normal things one would expect of a commission of inquiry – contract the best advice, establish high-level panels, and invite distinguished people to meet with us. In addition, I proposed two things that had not been done before: first, that we hold open public hearings on every continent; and second, that after completing our report, we extend the life of the Commission by nine months to ensure that the recommendations were presented to governments, industry, non-government organizations and the media in a series of meetings around the world.

The proposal for public hearings met with some difficulty. It arose out of my experience with OECD, where I introduced the idea of public meetings in the OECD 1979 review of New Zealand's environment policy. Some Commissioners pointed out that, under the UN resolution, we had no authority to conduct public hearings. In fact, they argued, to consult civil society on issues that were highly sensitive and controversial could and would be seen by governments hosting us, as an infringement on their national sovereignty. We had already received invitations from several governments to hold our meetings in their countries and I suggested that we make our acceptance conditional on agreement by the host government to allow us to organize such hearings. The Commission agreed. The hearings

went on to become the Commission's hallmark and, in my view, the primary source of the hard-won consensus on the issues before it.

The proposal for a post-report advocacy phase met with even greater difficulty. It arose from what I had learned about the fate of the recommendations of earlier Commissions, such as Brandt and Palme, which had simply published their reports. Many commissioners argued for a similar policy. The UN Resolution gave us no authority to extend our life, they said. Our obligations ended with the publication of our report. Nevertheless, they did agree to allow the proposal to remain in the work program and, when I secured the necessary funds, it was fully embraced by all commissioners. In the event, this proved to be a major reason why the report was so well received and so successful in changing the terms of the environment/development debate.

In deciding not only to adopt the alternative agenda, but also to include public hearings in its strategy of inquiry, the commission gave itself a chance to take a fresh look at the issues and to come up with some meaningful recommendations for change. With a budget of around $8 000 000, which it also adopted, it gave itself a license to find the resources needed. My statement, *Mandate for Change*, was published with few modifications as its point of departure for global consultations.

A GLOBAL CONSULTATION

Some 900 days later, at the end of February 1987, the Commission concluded its deliberations on *Our Common Future* at its final meeting in Tokyo. A full consensus had been reached on a diagnosis of the key issues of environment and development, and on a global agenda for change to manage the growing risks and opportunities challenging the world community. A political document, *Our Common Future* was to be hailed by major economic and environmental journals and by leaders in the government, corporate, scientific and non-governmental world as the most hopeful and useful report on environment and development to appear in two decades.

But the Commission had gone down to the wire on a number of issues. Energy, with its many unresolved dilemmas, and with the emotions surrounding nuclear energy reinforced by Chernobyl, had resisted several attempts at consensus. In Tokyo it kept Commissioners going several nights and provided the stuff of drama. Population, security, Antarctica, international economic relations and, of course, institutional reform proved almost as difficult.

Between its second meeting in Jakarta and the final moments in Tokyo, the Commission had achieved agreement from initially wide and, on some issues, fundamental disagreement. Composed of political leaders – present and former ministers of foreign affairs, finance, energy, population, environment and many other portfolios – industrialists, scientists and senior administrators, the Commissioners represented almost every shade of ideology, academic background and personal experience (for a list of Commissioners, see The World Commission on Environment and Development, 1987, Annex 2). They served in a personal capacity, but national, cultural and group loyalties were strong and manifested themselves in many ways. Tackling some of the most complex issues facing humankind, issues loaded with questions of power, equity and justice, the group had a built-in potential to blow itself apart. But it did not. Why?

The question remains and the answers vary depending on one's perspective. From my perspective as Secretary General, a few things stand out. The agreement reached at its first meeting on an alternative agenda and a strategy for the inquiry was fundamental. The capacity of commissioners to rise above their differences and to search for and ultimately find common ground was evident from the beginning. It was aided by a compelling sense of the urgency of the issues of global change, and a growing conviction that the approaches to management then in place were deeply awed and that change was essential. The brilliant reports of our expert panels and the work of a dedicated secretariat and excellent consultants and advisers provided basic support. All of these were essential to the consensus, but they were not sufficient.

Commissioners skirmished with each other for months before we finally broke through the diverse cultural and ideological skins that we all brought to the table. And, in my view, we broke through finally only because of the public hearings and site visits that we organized in every part of the globe.

The hearings enabled Commissioners to hear the testimony of nearly 1000 experts and concerned citizens on five continents. They gave us direct exposure to the issues on the ground and direct contact with the people living the issues. They provided us, almost complete strangers at the beginning, with a common base of information and a common set of experiences. For two days, sometimes three, Commissioners listened to testimony delivered with conviction and often with great emotion. Dialogues with ministers, experts and ordinary people on the front line of the interface between environment and development, helped to tear down the normal barriers of communication in the subsequent, closed meetings of the Commission.

I recall many incidents in Brazil, Indonesia, Norway, USSR, Zimbabwe and Canada. We not only invited groups and individuals to come to our hearings, but where necessary we took our hearings to them. We went to East Kalamantan to see ancient tropical forests being mined for timber and to talk to local groups struggling with problems that most commissioners could not even imagine. We ew to the highlands of Zimbabwe where the land had been

sheared of trees, compromising prospects for development. In Sao Paulo, the leaders of several Indian nations were present, their homelands threatened by massive deforestation. Electricity seemed to jump from the eyes of the first who spoke when he declaimed: I am the son of a small nation ••• dying and as we die insisting that there is only one place for us to live. Two days later, when the Chairman tried to close the hearing long after the scheduled hour, they crowded the railing shouting You must not stop. We have so much more to say, and you will not be back!

The public hearings and site visits revealed the heavy contradictions between the reality of environment and development totally interlocked in the daily lives of people, industries and communities, and the unreality of the distinctions between them made in our academic, economic and political institutions. They brought a new dimension to the information and analyses provided by our consultants and expert panels. They also provided a context within which to understand the series of disasters that grabbed the world's headlines during the 900 days of the Commission's work: the drought-triggered human and ecological catastrophe in Ethiopia, the Sahel and elsewhere in Africa; the great fire of East Kalamantan; the chemical disasters in Bhopal, Mexico City and the Rhine; the growing rate of forest death in Europe; the fresh warnings on climate change; the stratospheric ozone hole and Chernobyl.

Environmental crisis was seen to be a euphemism for what in reality is a failure of economic and other development policies, and the institutions behind those policies. Following each hearing, I observed commissioners with an environmental background and those with a development background, moving closer to an ever-widening consensus based on a shared understanding of actual situations and the issues arising from them.

SUSTAINABLE DEVELOPMENT

Looking forward to the 21st century, the Commission saw a world filled with opportunities and risks without parallel in human history. Human ingenuity in science and technology have resulted in enormous gains, and have provided the tools to shape an even more promising future. But these same forces have also given rise to a range of new risks and many of them represent clear and ominous threats to our common future.

These opportunities and risks, said the Commission, stem from two colliding sets of trends. First, the world's population and its economic and technological systems are on the greatest growth track in all of history. Second, even with current levels of population, business-as-usual economic growth threatens to drive the earth's ecological systems past critical thresholds essential for future development and even survival. These colliding trends are accentuating rich–poor differences within countries and among nations, and they are compounding the problems of cooperatively managing the risks to our common future.

Growth on the scale required to meet present and future needs would entail a further rapid expansion of the world economy, perhaps by as much as five to ten times over the next half century. A five- to ten-fold increase over 50 years sounds enormous, but it actually re ects annual growth rates of only 3.2 and 4.7%. In many developing countries, that is hardly enough to keep up with projected rates of population growth, let alone reduce levels of poverty.

The Commission recognized that such growth could translate into a colossal new burden on the environment. Concerning energy alone, for example, the Commission found that if nations were to employ current forms of energy use, energy supply would have to increase by a factor of five, just to bring developing countries, with their present populations, up to the level of consumption now prevailing in the industrialized world. Critical life-support systems would collapse long before reaching those levels.

A new development path was required, one that sustained human progress not just in some areas for a few years, but everywhere for the indefinite future. The Commission was convinced that humanity has the ability to make development sustainable, a concept that the Commission defined in various ways, most famously as development that meets the needs of the present without compromising the ability of future generations to meet their own needs.

With the publication of *Our Common Future*, the concept of sustainable development spread rapidly through the corridors of government, industry, media and civil society. Today, it trips off the tongues of Presidents and Prime Ministers, editorial writers, academics and priests, as well as protest marchers in Seattle, Washington and elsewhere. The concept goes far beyond sustainability in the environmental sense. It is a broad concept for social and economic progress that has the potential to provide new insights for institutional reform and inspiration for international cooperation (see Box 1).

Sustainable development, the Commission found, requires strong measures to change both the content and the quality of growth and equally strong measures to ensure a more equitable distribution of the proceeds of growth,

Box 1 Strategic imperatives for sustainable development (WCED, 1987)

The Commission identified a number of broad directions that development must take if it is to be sustainable. It called these the strategic imperatives for sustainable development and it focussed on seven of them: reviving growth; changing the quality of growth; meeting essential needs for jobs, food, energy, water, and sanitation; ensuring a sustainable level of population; conserving and enhancing the resource base; reorienting technology and managing risk; and merging environment and economics in decision making.

nationally and internationally. *Our Common Future* has a number of recommendations in this regard, in the fields of international economic relations, energy, industry, agriculture, the management of the global commons, and new concepts of security which go beyond military security.

In the Commission's view, the principal development challenge is to meet the needs and aspirations of an expanding developing world population, especially those who live in absolute poverty and are unable to satisfy even their most basic needs. Meeting essential needs requires not only a new era of economic growth for nations in which the majority are poor, the Commission said, but an assurance that those poor get their fair share of the resources required to sustain that growth. Such equity would be aided by political systems that secure effective citizen participation in decision making and by greater democracy in international decision making.

Urgent steps are needed to limit population growth. Sustainable development can be pursued more easily when population size is stabilized at a level consistent with the productive capacity of the ecosystem. The Commission pointed out, however, that the issue is not simply one of global population size. Providing people with facilities and education that allow them to choose the size of their families is essential and, for women, a basic human right of self-determination. Population policies must be integrated with other economic and social development programs such as health care and the provision of livelihoods.

Repeatedly, the Commission stressed that sustainable development involves more than growth. It requires a change in the content of growth to make it less material- and energy-intensive and more equitable in its impact. In economic terms, it means living off the earth's interest, without encroaching on its capital. It also means investing to sustain and even enhance the earth's present capital stocks of environmental resources, so that future dividends can be ensured and enlarged. Policies to this effect are required in all countries as part of a package of measures to maintain the stock of ecological capital, to improve the distribution of income, and to reduce the degree of vulnerability to economic crises.

If needs are to be met on a sustainable basis, the Earth's natural resource base must be conserved and enhanced. Major changes in policies will be needed to cope with the industrial world's current high levels of consumption, the increases in consumption needed to meet minimum standards in developing countries, and expected population growth. However, said the Commission, the case for the conservation of nature should not rest only with development goals. It is part of our moral obligation to other living beings and future generations.

Sustainable development also requires a reorientation of technology – the key link between humans and nature. Technology development must be changed to pay greater attention to environmental factors. And the capacity for technological innovation needs to be greatly enhanced in developing countries so that they can respond more effectively to the challenges of the future.

The common theme throughout *Our Common Future* is the need to integrate economic and ecological considerations in decision making. Echoing the OECD's finding that the environment and the economy can be made mutually reinforcing, the Commission noted that the compatibility of environmental and economic objectives is often lost in the pursuit of individual or group gains, with little regard for the impacts on others, with a blind faith in science's ability to find solutions, and in ignorance of the distant consequences of today's decisions. Institutional rigidities add to this myopia.

Those responsible for managing natural resources and protecting the environment are institutionally separated from those responsible for managing the economy, whose focus is narrowly concerned with production and growth. Energy utilities produce power, while the acid rain they also produce is left to the ministries of environment. Treasury departments subsidize unsustainable practices in agriculture, forestry, fishing, industry, transport and other sectors, but ignore the consequences for resources and the environment. The challenge is to give the central economic and sectoral ministries the responsibility for the quality of those parts of the human environment affected by their decisions, and to give the environmental agencies more power to cope with the effects of unsustainable development.

The integration of economic and ecological factors into the law and into decision-making systems within countries has to be matched at the international level. Economic interactions through trade, finance, investment, and travel will also grow and heighten economic and ecological interdependence. Hence in the future, even more so than now, sustainable development requires the unification of economics and ecology in international relations.

The ability to anticipate and prevent environmental damage requires that the ecological dimensions of policy be considered at the same time as the economic, trade, energy, agricultural, and other dimensions. They should be considered on the same agendas and in the same national and international institutions. The real world of interlocked economic and ecological systems will not change; the policies and institutions concerned must.

TIME FOR CHANGE

Seven weeks after Tokyo, on April 27th, 1987, the Chairman, in the presence of the Commission, presented *Our Common Future* to the world at a major event in London. Ten years later, it had sold over 1 000 000 copies in nearly 25 languages, surely a record for a UN report.

After the London launch, Commissioners discussed their recommendations with over 100 governments, often directly with the President or Prime Minister. We presented the recommendations to intergovernmental organizations, including the heads of OECD and virtually all of the UN agencies. We discussed them with the heads of the World Bank and the regional development banks. We held seminars, formal and informal, with industrial and non-governmental leaders on every continent. Presenting our report to a head of state or chief executive officer had a great advantage. It meant that their senior officials had to read it in order to prepare a briefing note. As a result, leaders from government, industry and civil society were ready when we presented the report to the UN General Assembly.

The UN commenced consideration of *Our Common Future* in a plenary debate on October 19th, 1987. Prime Minister Gro Harlem Brundtland presented our recommendations in the presence of the UN Secretary General, Prime Minister Gandhi, and many other political leaders who had come to support them.

A dramatic highlight was provided by the President of the small island nation of the Maldives. He galvanized the Assembly with a plea for action to prevent the physical disappearance of his nation. The highest point in the Maldives is less than 2 m above sea level. Most of that nation of 1190 islands is below 1 m, so a 1 m rise in sea level would spell extinction for his nation. To endangered species, humans have added a new category of global risks, endangered nations (see the Commonwealth Group of Experts, 1989). The report speaks of ••• far reaching social and economic effects on low-lying areas, as in Guyana ••• Kiribati, Tuvalu, and other Commonwealth countries ••• They face the prospect of ooding from the sea and greater risk of storm surges, but deeper ooding on inland plains ••• There are potential implications – ••• for agriculture; fresh water supply threatened with salinization; the sitting of towns, factories, power plants and airports; and hazardous waste disposal.

Noting that the next few decades are crucial, the Commission concluded *Our Common Future* with a call for action to break out of past patterns. Attempts to maintain social and ecological stability through old approaches to development and environmental protection will increase instability, it said. Security must be sought through change. To achieve the needed changes, the Commission proposed that the UN General Assembly transform *Our Common Future* into a United Nations Programme on Sustainable Development. An international conference could be convened for this purpose, it said, to promote follow-up arrangements that will be needed to set benchmarks and to maintain human progress. While the General Assembly endorsed all of the Commission's recommendations in principle, on that point, as the world knows, it did act. The result, five years later, was the Rio Earth Summit.

See also: Agenda 21, Volume 4; UNCED (United Nations Conference on Environment and Development), Volume 4.

NOTES

From 1984–1987 Jim MacNeill was Secretary General of the WCED. Previously Director of Environment at the OECD and a Deputy Minister in the Canadian Government, he is now Chairman of the World Bank's Independent Inspection Panel.

REFERENCES

Commonwealth Group of Experts (1989) *Climate Change: Meeting the Challenge*, Commonwealth Secretariat, Marlborough House, Pall Mall, London SW1Y 5HX, September.

MacNeill, J (1989) Strategies for Sustainable Economic Development, *Sci. Am.*, 261(3).

The World Commission on Environment and Development (1984) *Mandate for Change*, WCED, Geneva, August.

The World Commission on Environment and Development (1987) *Our Common Future*, Oxford University Press, New York.

World Conservation Union (IUCN)

see IUCN (The World Conservation Union) (**Volume 4**)

World Conservation Union (IUCN) Red List of Threatened Species

As a way of instituting a standard for establishing the status of threatened species, IUCN has designed and promulgated the Red List of Threatened Species. Based on the work of some 7000 experts who are members of the IUCN Species Survival Commission, the Red List is based on a consistent and objective quantitative system of categories and criteria for listing. The categories include extinct, extinct in the wild, critically endangered, endangered, vulnerable, lower risk, data deficient, and not evaluated. These categories provide an easily and widely understood method for

highlighting those species under a high risk of extinction, helping to focus attention on the conservation measures designed to protect them.

The criteria for listing include population reduction, extent of occurrence, size of population, and probability of extinction, with different parameters leading to the allocations to categories. The 1996 Red List assessed the conservation status of all species of birds and mammals. While coverage for fish and plants is less complete, and for invertebrates is focused on only a few groups (for example, butter ies) coverage of these latter groups is constantly improving.

Red List Authorities (RLA) are being established for all taxonomic groups included on the list, usually the specialist group of the IUCN Species Survival Commission responsible for the species; an exception is birds, where BirdLife International is the designated RLA. The role of RLA is to ensure that all species within their jurisdiction are correctly evaluated against the categories and criteria at least once every 10 years and, if possible, every five years. The Red List Programme has adopted the use of a commercial software package called RAMAS.

The Red List grew out of IUCN's *Red Data Book*, first published in 1996 (for mammals). Dozens of subsequent Red Data Books have been published at global, regional, and national levels, helping to stimulate laws to address the problem of endangered species in many countries. The Red List has been an essential element to communicate the idea of species threatened by extinction, and has generated a global response to the extinction crisis.

The IUCN Red List is now widely recognized as the most comprehensive, balanced, and apolitical global approach for evaluating the conservation status of species of plants and animals. For further information, see http://www.redlist.org.

See also: Extinctions (Contemporary and Future), Volume 2; Endangered Species, Volume 4.

JEFFREY A MCNEELY *Switzerland*

World Food Council (WFC)

see WFC (World Food Council) (**Volume 4**)

World Health Organization (WHO)

see WHO (World Health Organization) (**Volume 4**)

World Resources Institute (WRI)

see WRI (World Resources Institute) (**Volume 4**)

World Trade Organization (WTO)

see WTO (World Trade Organization) (**Volume 4**)

World Wide Fund for Nature (WWF)

see WWF (World Wide Fund for Nature) (**Volume 4**)

Worldwatch Institute

Founded in 1974 by Lester Brown, Worldwatch Institute is an independent, public policy research institute widely respected for its pioneering work on sustainable development and the connections between the natural world and human welfare. The institute is one of the world's leading sources of environmental analysis and information.

The institute's mission is to foster the evolution of an environmentally sustainable economy, one in which human needs are met in ways that do not threaten the prospects of future generations. Worldwatch seeks to achieve this goal through conducting interdisciplinary research on global environmental issues and disseminating the results worldwide. The results of its research are distributed through a worldwide publishing network, through close working relationships with the world's major news organizations, and through individuals working in various ways on their own to achieve the institute's goals.

Worldwatch's interdisciplinary approach is highly valued by decision makers in both the public and private sectors, ranging from elected members of national legislatures to corporate planners. The combination of the institute's cutting-edge analysis and accessible writing style means that specialists in particular fields look forward to each new publication. Over the years, the institute has evolved a stable of five publications. They are, in order of launching: the *Worldwatch Papers*, Worldwatch books, *State of the World* (annual), *World Watch magazine* (bimonthly), and *Vital Signs* (annual). The institute also publishes a *Worldwatch Database Disk*, and in

2000 it launched an electronic publication: *Worldwatch Issue Alerts*. Written by Lester R Brown, the Alerts are released via email, fax, and the institute's website • http://www.worldwatch.org/alerts/indexia.html• .

The institute's research is translated through a global publishing network. At present, it has some 175 publishing contracts in effect in over 30 languages, including all of the major languages, such as Arabic, Chinese, French, Japanese, Portuguese, Spanish, and Russian, and many minor ones (see • http://www.worldwatch.org/foreign/index.html•).

The institute has developed a close working relationship with the world's major news organizations and global wire services. The institute and its researchers are cited in, or are bylines of, some 40 stories a day in newspapers and magazines (and associated websites) worldwide. In addition, the electronic media disseminate the results of its research.

REAH JANISE KAUFFMAN *USA*

WRI (World Resources Institute)

The WRI is an independent center for knowledge on global environmental and development issues. WRI provides objective information and practical proposals for policy and institutional change that will foster environmentally sound, socially equitable development. Its stated mission is to move human society to live in ways that protect Earth's environment and its capacity to provide for the needs and aspirations of current and future generations.

WRI works to put ideas into action. Activities are centered around four themes:

1. *Biological resource stewardship*: WRI focuses on halting the degradation of biological resources, particularly in forests and agriculture, and enhancing their contribution to economic development.
2. *Climate protection*: WRI seeks to prevent human-caused climate change by promoting cooperation among businesses and nations, rich and poor, in implementing international agreements, fostering economic opportunity, and developing cleaner energy sources and technology.
3. *Development paths*: WRI works with a worldwide network of partner organizations to help promote development policies that reduce poverty, improve livelihoods, and protect the environment.
4. *Sustainable enterprise*: WRI fosters sustainable enterprises by helping to provide the management strategies and incentives that will enable firms to build future profitability around environmentally friendly products and processes.

Based in Washington, DC, WRI was established in 1982 by James Gustav Speth. It has an international staff of 125, who work with advisors, collaborators, and partner institutions in more than 50 countries. WRI has a Board of Directors of 34 prominent scientists, international policy makers, and corporate managers.

In all of its policy research and work with institutions, WRI tries to put ideas into action, meshing the insights of scientific research, economic and institutional analyses, and practical experience with the need for open and participatory decision-making.

For further information contact WRI at: World Resources Institute, 10 G Street NE, Washington, DC 2002, www.wri.org.

DONNA WISE *USA*

WTO (World Trade Organization)

The WTO created on 1st January 1995, is the institutional body established as a result of the successful completion of the Uruguay Round of multilateral trade negotiations. The WTO is an intergovernmental organization, succeeding and incorporating the framework of the General Agreement on Tariffs and Trade (GATT).

The central mandate of the WTO is trade liberalization. To that end, it incorporates key trade liberalization principles such as most-favored nation, national treatment and non-discrimination, provisions constraining the use of various trade restrictions and distortions (such as tariffs, quotas, subsidies, non-tariff measures and other trade instruments), and the general exceptions clause (GATT Article XX), which allows members to deviate from WTO obligations in pursuit of several public policy goals, including the protection of plant, animal or health life, or the conservation of exhaustible natural resources. WTO obligations are supported by binding international dispute settlement procedures, which represent the most stringent dispute arbitration mechanisms in support of international law.

From the outset, the WTO has broad-based membership representing industrialized and developing countries, in addition to countries with economies in transition. As of November 2000, 140 countries were WTO members.

WTO rules set out rights and obligations among states covering trade in merchandise goods, including sectors that remained largely excluded from trade negotiations in the GATT (notably, agriculture and textiles). In addition, the WTO sets out rules covering both trade in services, as well as trade-related intellectual property rights. Among the emerging issues on the new trade agenda are competition policy and investment policy, and these may be part of future multilateral trade negotiations.

Since its creation, the WTO Committee on Trade and Environment has examined links between trade rules and environmental policy. Three main clusters continue to be examined: the relationship between multilateral environmental agreements and trade rules; market access issues arising from environmental standards; and the role of civic society in trade policy. A number of WTO Dispute Settlement Panel and Appellate Body reports have involved environmental, human safety and food safety-related issues. The fact that the WTO finds itself arbitrating public policy issues traditionally outside of trade policy, coupled with a perceived lack of transparency and public accountability, have prompted concerns among the public about the mandate and capacity of the WTO to address international environmental issues. The implications of the WTO on environmental policy and environmental quality were the subject of public protests during the 1999 WTO Ministerial Meeting, held in Seattle, Washington.

More information on WTO can be obtained by contacting the WTO Secretariat, Rue de Lausanne 154, CH-1211 Geneva 21, Switzerland or at www.wto.org.

SCOTT VAUGHAN *Canada*

WWF (World Wide Fund for Nature)

The WWF is a leading independent conservation organization headquartered in Switzerland, with National Organizations in 26 countries and associates in five countries. Founded in 1962, it was originally known as the World Wildlife Fund and retains this name in Canada and the United States of America. It has 4.7 million supporters and 22 program offices in countries where it is focusing its activities. Its ultimate goal is To stop, and eventually reverse, the accelerating degradation of our planet's natural environment and to help build a future in which humans live in harmony with nature . WWF combines practical, solution-oriented field projects with education, capacity building and policy work. It seeks to demonstrate what can be done to address conservation problems, and then to help others acquire the necessary knowledge and skills to carry out long-term conservation for themselves. Through its policy work, WWF helps create the laws and agreements needed to conserve biodiversity. WWF's national organizations carry out conservation activities in their own countries and contribute technical expertise and funding to WWF's International Conservation Programme. The WWF network invests well over US$ 200 million per year in its conservation activities in some 100 countries, with most of its income coming from personal donations. Roughly half is used for domestic conservation by WWF National Organizations and half by the International Conservation Programme. WWF has recently developed a new approach to conservation, identifying 200 ecoregions that are critical to the conservation of the world's biological diversity, and seeks to focus its activities in these regions. Its Forest Programme seeks to halt the loss and degradation of forests and all kinds of woodlands, particularly old growth forests, and to establish a network of protected areas encompassing 330 million hectares that will include at least 10% of each of the world's major forest types. Its Oceans Programme seeks to create incentives for sustainable fishing, to reduce or eliminate destructive fishing practices and the by-catch of marine wildlife, and to establish effective recovery plans for key species of threatened marine fishes. The website for WWF is http://www.wwf.org.

JEFFREY A MCNEELY *Switzerland*

Zayed International Prize

see International Environmental Prizes (**Volume 4**)

Case Studies

Regional Responses to Global Environmental Change

Osama A El-Kholy
Arabian Gulf University, Manama, Bahrain

This contribution sets the stage for a series of essays on regional responses to global environmental change. It distinguishes between two major groups of global change: the systemic and the cumulative. Systemic global change requires worldwide responses, while cumulative global change must be dealt with by regional responses. Furthermore, in reviewing the case studies that follow, we need to bear in mind the differences between responding regionally to problems of conserving and sharing of natural resources, and addressing problems of regional environmental pollution.

PROBLEMS AND SOLUTIONS

A number of environmental problems are arising that require tackling global change on a multinational scale. To start with, is there a known solution to the problem? If there is, can it be implemented at acceptable financial and social costs? How long does it take to implement it? How urgent is it to act? And finally, how can the socio-economic burden be equitably shared by all the parties involved, given their different contributions to the problem and their abilities to bear the cost of remedial action?

Two things have become clear. Global problems cannot be solved without the active participation of all countries, developing and industrialized. To be effective, remedial action should be based on the precautionary principle. We need to act on the basis of scientific evidence, rather than on direct unambiguous observation of global change. If we wait that long, the damage may become irreversible.

These are the questions that have shaped, and continue to shape, responses to global change.

TWO TYPES OF GLOBAL CHANGE

Manifestations of global change fall under one of two major categories: the systemic and the cumulative. Systemic changes cover the whole globe, even though their impacts vary from place to place. Depletion of the stratospheric ozone layer, and global climate change due to global warming are well known examples of global systemic change.

Cumulative global changes, however, start on a local scale affecting specific areas. As these changes proliferate, they gradually acquire global dimensions. Examples are the various forms of land degradation, freshwater shortage, air and marine pollution.

SYSTEMIC CHANGES CALL FOR GLOBAL RESPONSES

Responses to systemic global change cannot be described meaningfully on a strict delineation of regions, geographical or political. The fact is that a number of countries from different regions have played prominent roles in building common multinational responses that cut across geographic and political boundaries. Common responses have sometimes come from countries that are widely separated geographically. This has happened both within developing, as well as amongst industrialized, countries, and even across this duality of grouping.

The three most important manifestations of systemic global change that have elicited global responses so far (stratospheric ozone layer depletion, loss of biodiversity and climate change) provide interesting examples of differences in our experience in responding to systemic change. The first was a pioneering exercise in formulating a global response to systemic change. It provided a prototype model for tackling such problems. The new approaches and mechanisms devised in this case were copied in addressing other phenomena of systemic change on a global scale later on. The second looked at first as if it had created a new balance of power between the industrialized and the developing countries. The third dealt with a problem that was shrouded in more scientific uncertainly than usual, while at the same time calling for solutions that are too drastic socially and economically to implement.

REGIONAL RESPONSES TO CUMULATIVE GLOBAL CHANGE

In the case of cumulative global change, the local manifestation of such change is primarily the concern of the countries in the locality of change. It is here that regional responses are well defined. They re ect the regional understanding of the problem, as well as the level and experience of cooperation in dealing with regional problems. This is obviously conditioned by the political climate in the region, the balance of power between the countries involved, and sometimes even the attitude of powerful players from outside the region.

The case studies of regional responses to global change highlight two major types, as outlined next.

TWO TYPES OF CUMULATIVE CHANGE

One major consideration that defines how certain factors shape the regional response is the nature of the cumulative

change problem; whether it concerns the conservation and sharing of natural resources, or addresses environmental protection. Examples of the first are the regional initiatives for the basins of the Zambezi and the Danube rivers, Lake Chad, and the Mekong delta. Examples of the second are the United Nations Environment Program (UNEP) regional seas program, the North African Green Belt and the European Air Pollution projects.

Strictly speaking, pollution abatement on a regional scale has a direct impact on natural resources, whether they come from land, waterways, or open seas. However, it also has more direct and significant benefits to human health and well-being. The planning and implementation of regional responses to pollution problems face the countries in a region with a situation different from that of projects for the sustainable development of shared natural resources.

SUSTAINABLE DEVELOPMENT OF SHARED NATURAL RESOURCES

As might be expected, projects for the conservation, sustainable development and equitable sharing of common natural resources raise highly sensitive political issues. They usually come up against practices that have been in place for considerable periods of time and that were established under very different political climates from those currently prevailing. The rational utilization and equitable sharing of common resources in particular presents sometimes highly complex challenges that render it difficult to achieve common regional responses. An international treaty providing a framework for dealing with this particular problem recently drafted in a United Nations (UN) conference (the UN Treaty on Non-navigational Uses of International Waters) was not accepted by some of the major player countries in some of the world's major river basins.

Conservation of marine resources in some areas has also resulted in disputes, even within reasonably well-established regional set-ups, such as the European Union (Spanish boats fishing off the British coast). The problem becomes more complicated when the dispute draws in countries that are outside the region, as in the case of the disputes between the European Union and Canada or Spain and two Maghreb countries (Morocco and Mauritania).

REGIONAL RESPONSES TO THE POLLUTION OF AIR, LAND AND WATER

In joint regional efforts to manage the pollution of one or more of air, land, water, certain conditions have to be satisfied. First there needs to be a reliable regional monitoring system to determine the type and scale of pollution, its geographic distribution, and its sources. When reasonably well-developed national monitoring systems exist, the task becomes that of integrating these into one coherent regional system. The situation is more difficult and costly when such a monitoring system has to be established almost from scratch. This has to be followed by an environmental assessment of the impacts on the different environmental media. Next, the nature and cost of optimum remedial action needs to be worked out. Assuming that these are also known, the equitable sharing of costs is the next step. This is perhaps the most difficult stage in working out a regional program for environmental protection. Finally, there is need for periodic reviews of the issue in the light of changing global and local conditions: the so-called *adaptive assessment process*.

Perhaps one of the more successful examples is the European effort to control air pollution across the continent, initiated by the UN Economic Commission for Europe. There are also two cases of dealing with land-based sources of marine pollution (the Black Sea and the Mediterranean), where an outline action plan has been worked out; but funding for its implementation has yet to be obtained. Several other examples exist in combating desertification in some regions.

It can, however, be said in conclusion that responding to global change on a regional basis is gathering momentum and gradually engaging the attention of decision-makers. While the industrialized countries are ahead in this process, developing countries are moving in the same direction with help from donor countries and inter-governmental organizations.

This contribution sets the stage for selected case studies on regional responses to regional stresses in the context of global change. (It is not possible to provide more than a sampling of such studies in this volume.)

These case studies have been brought together following this introduction at the end of Volume 3 – see Table of Contents for Volume 3 for a detailed listing.

Seas

Baltic Sea

Bengt-Owe Jansson
Stockholm University, Stockholm, Sweden

The semi-enclosed Baltic Sea is a poor receiving area for the wastes of the increasing human population in the large watershed. Nutrients and toxic substances from land runoff and atmospheric fallout create turbid waters, dead bottoms and decline of fish-eating birds and mammals. Nine countries with different policies and stage of development make strong measures of redevelopment difficult to take. Still, success stories of recovering birds and seals, thanks to the development of institutional tools and citizen involvement, give hope for the restoration of ecosystem health.

This semi-enclosed sea is surrounded by nine countries with a total population within the watershed of ca 90 million (Figure 1). It is geologically young, having passed four stages: the Baltic-Ice Lake, the Yoldia Sea, the Ancylus Lake and the Littorina Sea, alternating between fresh and saltwater since the recession of the last glaciation 10 500 years ago. Its mean depth is 55 m (Table 1) and the narrow connections to the North Sea are shallow, restricting the total water exchange to 25 years. The freshwater runoff from numerous rivers in the drainage area generates an inflow of the salty water from the North Sea, creating a gradient of the surface salinity from ca 10 parts per thousand (ppt) inside the entrance to nearly freshwater in the northern Gulf of Bothnia. The surface water temperatures show a similar gradient northwards with fast ice cover during 60% of the year in the northernmost part. The sills create a vertical stratification of the brackish water with a stable halocline at 50–70 m depth, between a low salinity surface layer, slowly moving out of the Baltic, and a saltier, and therefore heavier bottom water flowing inwards along the eastern side of the basin due to the earth's rotation. Due to the minimal tidal movements, the stratification of the water makes the Baltic deep water very sensitive to processes that consume oxygen (O_2), like decomposition of organic material of dead plants and animals settling on the bottom. The decomposing bacteria assembly also contains hydrogen sulfide-producing species, the deeper parts of the Baltic Proper alternating between oxic conditions with higher animals like worms, crustaceans and clams, and periods of hydrogen sulfide dominated by billions of sulfur bacteria. Denitrifying bacteria during periods of low O_2 convert nitrogen (N) in the sediment to nitrogen gas, which leaves the sea for the atmosphere. This process may export half of the annually deposited nitrogen load in the Baltic. An important renewal mechanism for the bottom water is large pulses of North Sea water, rich in salt and O_2, caused by prolonged periods of westerly gales which transport water in the order of tens of km^3 into the Baltic basin. In between, stagnant conditions with low O_2 and hydrogen sulfide may dominate for decades (Figure 2, map) These unpredictable events are responsible both for maintaining the salinity of the whole Baltic and for the amount of economically important fish (Figure 2c).

The organism assembly in the Baltic is a mixture of hardy saltwater and freshwater species, immigrating from ocean and freshwater, and therefore not adapted to the adverse conditions found in the Baltic. Of the 1500 animal species seen with the naked eye off the southern coast of Norway, some 50 are left in the Baltic at the latitude of Stockholm. This means that the number of species performing the basic processes of an ecosystem like fixing solar energy, decomposing dead organic mater, and filtering water, is small and a decrease in species diversity due to stressing environmental conditions will restrict the production capacity of the system. Still, the Baltic has worked well when undisturbed. Cod and flounder can be caught in the same gill net as pike and perch and the species of plants and animals are found in large individual numbers although usually of smaller body sizes. At optimal conditions, the fisheries have landed close to 1 million tons of fish per year, mainly herring, sprat and cod. This is 1% of the world catch from a sea 1 per mille of the surface of the world ocean (Figure 2a).

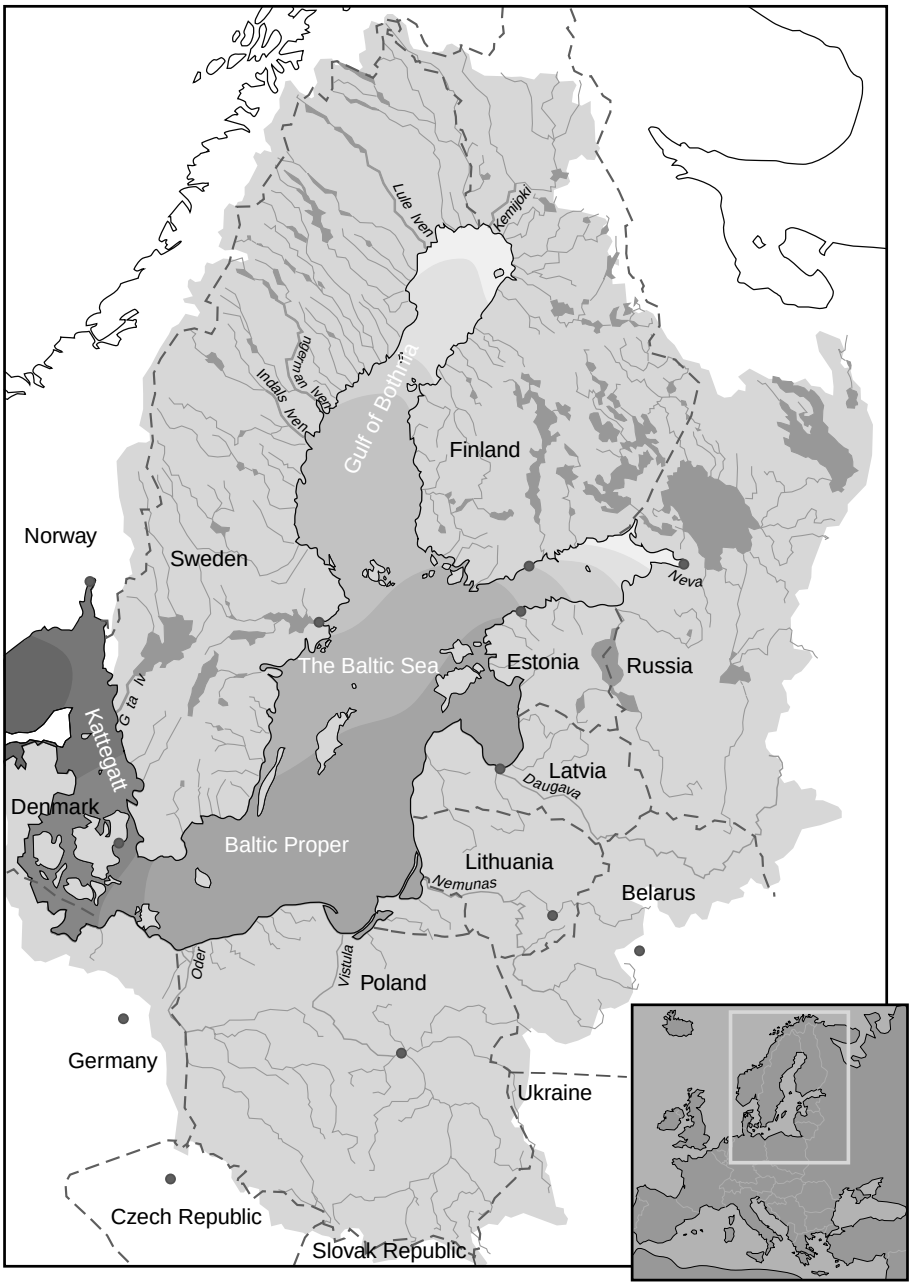


Figure 1 A map of the Baltic Sea

A typical Baltic habitat is the rocky archipelagos of the Swedish east coast and the Finnish west coast. Tens of thousands of islands and islets, with bottoms covered with the brown seaweed bladder wrack or *Fucus*, the blue mussel *Mytilus*, and during the summer green and brown filamentous algae form important spawning and nursery areas for fish. The great potential for sports fishing and boating and the beauty of the scenery have made these areas of the Baltic targets for an increasing tourist industry and summer residence for the coastal urban human population.

Some of the economically important fish like herring need the coastal areas for spawning but spend most of their time in the open waters, feeding on the pelagic food web. Based on solar energy and nitrogen and phosphorus (P) in the water, the microscopic plants show several blooms during the summer half of the year, among them a conspicuous, toxic bloom of cyanobacteria at the end of summer, taking its nitrogen from the atmosphere. This is nature's way of solving a lack of nitrogen due to consumption by previous algal blooms. The settling of dead plants and animals on the sea oor forms both

habitat and food for the bottom animals like crustaceans and the Baltic clam, *Macoma*, important food for the bottom fish.

The southern coasts of the Baltic Sea consist of transported sand and silt from the receding inland ice. Poland and the eastern Baltic states have good soil for agriculture, compared to the northern part of the Baltic drainage basin, dominated by forests and forestry. Since the 19th century, the population of the Baltic basin has increased four times and the use of fossil energy about 30 times. The urban centers, especially around the coasts, have expanded, and act as large consumers of energy and material, exporting large amount of wastes to sea and atmosphere. The emissions of phosphorus and nitrogen from agriculture, the transport sector and industry have, since 1900, increased eight and four times, respectively (Figure 2e). The increased consumption of O_2 of the deeper bottom waters due to increased primary production have caused extended periods of hydrogen sulfide. Decreased water transparency by 2–3 m since the 1910s has reduced the areas of *Fucus* vegetation because of less light. In the shallow parts of the sea, the amount of fish food of worms, crustaceans and clams has increased, and good food also in the open waters has favored the growth of pelagic fish like herring and sprat. Cod, spawning in deeper water, suffers from the stagnation periods (Figure 2c) and is, like the other species, overfished. Toxic substances from industry and agriculture have accumulated in the upper parts of the food chain, e.g., seals and fish-eating birds. The persistent polychlorinated hydrocarbons like polychlorinated biphenyls (PCBs) and dichlorodiphenyltrichloroethane (DDT) brought seals and sea eagles close to extinction during the 1970s (Figure 2b).

The suite of environmental incidents: the mercury crisis during the 1960s, the DDT and PCB problems during the 1970–1980s, and the increasing eutrophication of the coastal waters, escalated public pressure on existing responsible institutions. Theoretically it led to an increase of the resilience of the total Baltic system by developing institutions at different levels to prepare the system for environmental crises. One of the main catalysts for this process was the International Stockholm Conference on the Environment 1972. Among the lobbyists behind the conference were two informal organizations: the Baltic Oceanographers and the Baltic Marine Biologists, both with biannual conferences and workshops for standardizing methods, planning field studies and exchanging data. The critical state of the Baltic Sea discussed at the Stockholm Conference initiated the Convention on the Protection of the Marine Environment of the Baltic Sea, signed by the Baltic states in Helsinki, Finland in 1974. It is the first convention of its kind, including detailed prescriptions on how to combat pollution of all kinds in a multinational, marine area. Its commission, the Helsinki Commission (HELCOM), got an extended mandate in 1992, through including nature conservation and coastal zone management, already mentioned in a ministerial declaration of 1988. There, the seven states also announced a 50% reduction of the main pollutants to be achieved by 1995. At a prime ministerial conference in Sweden in 1990, a Baltic Sea Declaration was adopted, implemented by a task force within HELCOM to an action-oriented program – Joint Comprehensive Program. The investment priorities were based on effects on human health, ecology and economic feasibility. The most cost-effective hot spots have been identified, and of the 132 locations nearly two-thirds are located in the Eastern countries. In 1991–1992 the estimated costs amounted to 18 billion ECU over a 20-year period. The goal set in the ministerial declaration of 1988 has not yet been reached, however.

Table 1 Oceanographic data of the Baltic Sea. (For fuller description see Voipio, 1981)

Statistics	Measure
Area, total km^2	415 000
Area, Baltic Proper km^2	267 000
Length (N–S) km	1300
Width (W–E) km	1200
Average depth (m)	55
Maximum depth (m)	459
Sill depth (m)	17
Volume (km^3)	21 700
Residence time of water (years)	25
Tides (cm)	2–3
Total drainage area (km^2)	1 641 650

A regional Agenda 21 for the Baltic Sea Region (Baltic 21) was adopted by the environmental ministers in 1996. It comprises goals, scenarios and action programs for sustainable development for seven sectors: agriculture, energy, fisheries, forestry, industry, tourism and transport. The work shared by the different countries was completed in 1999.

HELCOM is the basic institutional instrument for governing the *natural capital* (*see* Natural Capital, Volume 2) of the Baltic Sea. It also comprises non-governmental organizations (NGOs). Individually, or through NGOs, the people have activated existing organizations and forced the birth of new ones because of their wish for clear water, pure food and clean air. During the Soviet era, environmental action was checked in the eastern countries. The Nature Conservation Society in Estonia was established in 1966 as the first one in Soviet Union. Like the Ecological Club in Poland a decade later, it demonstrated the bad state of the environment and fought for democracy. The environmental movement in the eastern countries played an important role in starting the transition process. Also, the World Wildlife Federation

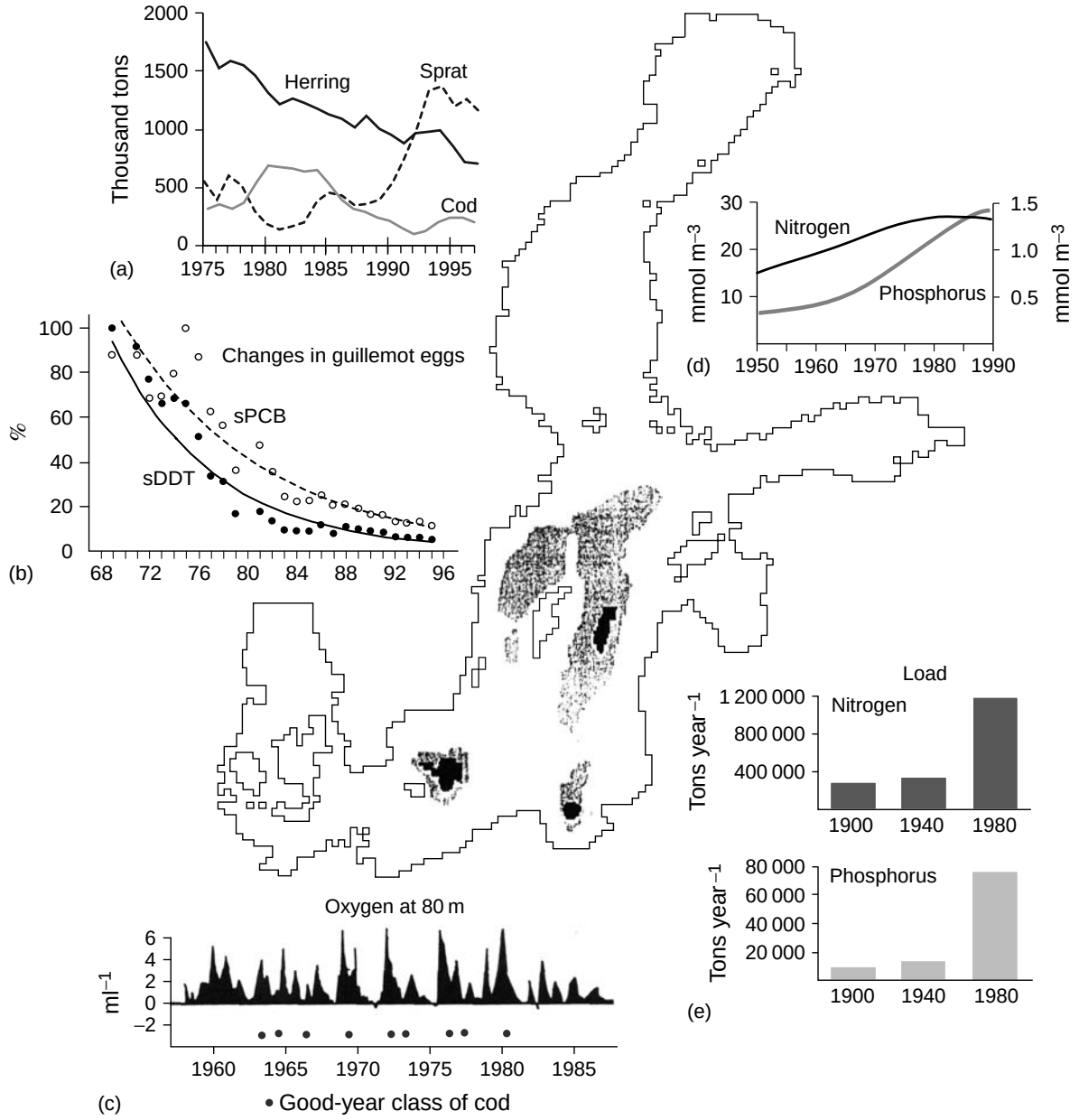


Figure 2 The changing Baltic Sea. Map in the middle of the figure shows the O_2 conditions of the bottom water in August 1998, gray shading corresponding to bottom areas with less than 2 ml l^{-1}, black areas with hydrogen sulfide (F Wulff, personal communication). (a) Trends of spawning stocks of main commercial fish species (ICES, 1998). (b) Concentrations of DDT and PCB in egg yolk of alks from central Baltic compared to levels in 1968 (Bignert, 1997). (c) O_2 content in main cod spawning areas compared to good year classes (Bernes, 1989). (d) Trends of nitrogen and phosphorus concentrations in open Baltic waters (computer simulations by F Wulff, personal communication). (e) Total annual load (input) of nutrients to the Baltic Proper since 1900 (Jansson and Dahlberg, 1999)

and Greenpeace International are active, and during the 1990s the Swedish Society of Nature Conservation and its Finnish counterpart created an umbrella organization for the many NGOs in the region – the Coalition Clean Baltic. Several of the organizations have observer status in governmental boards and management organizations and constitute important bridges between the public and the policy makers by increasing the spatial distribution of interactions.

The Lake Peipus Project is a good example of what citizens' stamina and hard work can achieve. This lake, fourth in size in Europe, is now shared by Estonia and Russia. Before the transition, the area was supported by Soviet Union as an important producer of high-quality

fish and agricultural products. When this market collapsed, people, especially on the Russian side, suffered, agriculture and management of the fishery sources declined, and water quality decreased. Thanks to a few concerned and idealistic people in the area, a transboundary commission was created, monitoring was restarted and participatory planning of the common resource use was developed by attracting foreign capital. A new future seems to be opening up through tourism, based on the pristine nature of the area.

The effects of the liberation of the occupied Eastern countries are far reaching, as environmental problems are regional, integrated and long-term. In 1994, the ministries for spatial planning formulated a framework for spatial planning in the Baltic Region: Vision and Strategies around the Baltic Sea 2010. The goal is to develop a strong infrastructure based on combinations of railways and roads, linking peripheral and central areas. West–East exchanges are especially considered.

Although the environmental goals have not been reached, nutrient concentrations in the open waters have leveled off since the mid-1980s (Figure 2d). This is probably due to the recession in economic development in Poland and the eastern countries. Due to bans in the mid-1970s, the levels of PCB and DDT in eggs of fish-eating birds and in gray seals are now down to 10 and 5% of 1968 levels, respectively, and the northern gray seal population is rapidly recovering (Figure 2b). Growing problems are the expanding urban centers, however, needing products and services of the surrounding ecosystem: freshwater, food, material, and processing of wastes. Calculated on an area basis, quantifying the ecological footprint in the late 1990s, the 29 biggest cities use 75–150% of the total Baltic Basin area.

The present environmental status of the Baltic Sea is probably close to the border between that of a less productive and healthy stability domain. These should include a clear reduction in the human resource use in the western countries, in order to create a more sustainable Baltic Sea area.

A description of the changing conditions of the Baltic Sea is found in Jansson and Dahlberg (1999).

See also: Water Resources: Baltic, Volume 3.

REFERENCES

Bernes, C (1989) *Monitor 1989. Climate and the Natural Environment*, National Environmental Protection Board, Stockholm, Sweden.

Bignert, A, ed (1997) *Swedish Marine Monitoring Data*, Swedish Museum of Natural History.

ICES, C M (1998) *Report of Baltic Fisheries Assessment Working Group*.

Jansson, B O and Dahlberg, K (1999) The Environmental Status of the Baltic Sea in the 1940s, Today, and in the Future, *Ambio*, 28, 312–319.

Voipio, A, ed (1981) *The Baltic Sea*, Elsevier Oceanography Series, Amsterdam.

Black Sea

Henri J Dumont
Ghent University, Ghent, Belgium

The surface area of the Black Sea is ca 423 000 km^2, and its volume 547 000 km^3. Maximum depth is 2212 m. The abyssal zone is situated in the south and is distinctly east- west oriented (Figure 1), while extensive shallow shelves (less than 200 m deep) occur in the north. The sea is currently a net exporter of water towards the Mediterranean, with the surface out ow about twice the deep and more salty in ow (about 130 km^3 year^{-1} ow in, against about 300 km^3 year^{-1} owing out, according to one conservative estimate). This excess is due to the contribution of a few major rivers, with the Danube (ca 210 km^3 year^{-1}) particularly prominent, followed by the Dnepr (50 km^3 year^{-1}). The in ow through the Strait of Kerch, mainly composed of water of the Don and Kuban rivers, varies from 20- 95 km^3 year^{-1}). All major rivers empty in the shallow northern zone, causing a salinity gradient from north to south. This northern, fresher water may freeze in winter, while the more saline water in the south does not. At least as important is a vertical gradient in water chemistry. The heavier salt water of Mediterranean origin sinks to great depths and does not mix with the overlying lighter surface water. This produces a gradient in salinity, with total salinity at the surface (0- 300 m) 17- 21 mille^{-1}, and below that ca 22 mille^{-1}. Because of the lack of exchange between layers, nutrients and other compounds sinking out to the bottom are lost to the dynamics of the surface ecosystem. Moreover, the deep layers have become totally depleted in oxygen. Only reduced compounds (e.g., H_2S) occur there (see Figure 2), and, as a result, more than 80% of the volume of the Black Sea is virtually devoid of eukaryotic plant and animal life.

BRIEF GEOLOGICAL HISTORY OF THE BASIN

The contemporary Black Sea, including the Sea of Azov, is currently a true hybrid between a sea and a lake. However, it was a brackish-water lake for most of its existence. Like the (south) Caspian, it is a remnant of the Tethys Sea, which gradually became isolated from the

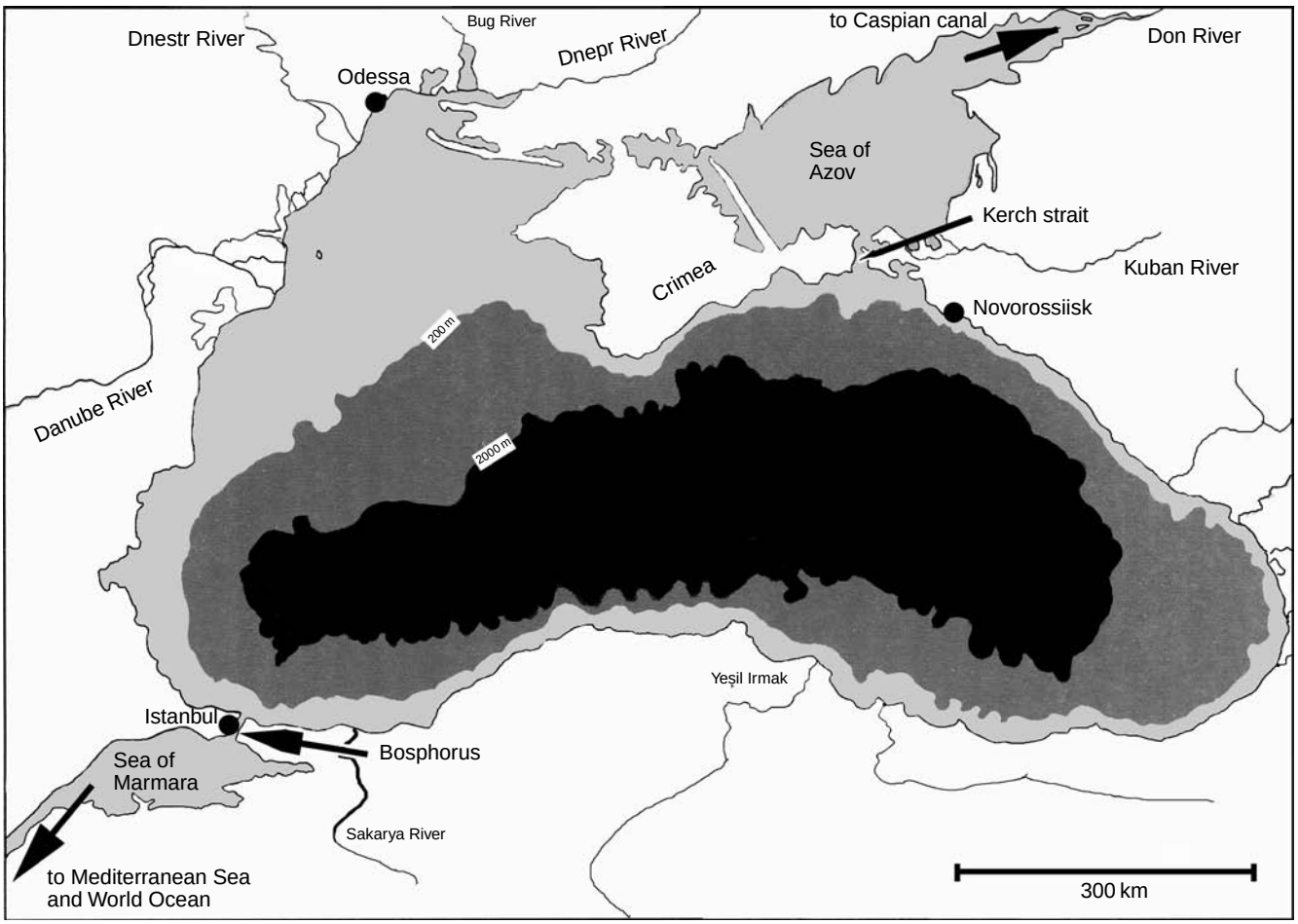


Figure 1 Map of the Black Sea, with major bathymetric divisions and major tributaries indicated

World Ocean as the Miocene orogenesis uplifted the Alps, Carpathians, Pontic, Balkan and Caucasus mountains. There has been no connection with the ocean from the Pliocene onwards (at least 3×10^6 years BP) until the mid-Holocene (ca 7000 BP, possibly even as recently as 5000 BP), save for an interval during the Riss-Wurm interglacial, some $1{\bullet}5-1 \times 10^5$ years BP.

Throughout the Pleistocene, there was a complex suite of recessions and high levels, corresponding with glacial/deglaciation cycles. The Black Sea and the Mediterranean were not connected, however, and the Black Sea uctuated between saline and almost fresh conditions. Around the mid-Holocene, when the Mediterranean broke through the Bosphorus, it stood over 100 m below its present level. The opening of the Bosphorus had dramatic effects: the entire north-west shelf and Sea of Azov were ooded at a rate that probably outpaced that of a walking person (see Figure 1: the ooded zone was not much smaller than that currently below 200 m in depth). Due to the fact that this at area was likely rather densely populated, and possibly a site of early agriculture and domestication of cattle, the ood event was remembered in legends and myths, like the Gilgamesj epos. It was also recorded in history as the Biblical ood.

BIOLOGICAL DIVERSITY

The original biota of the Black Sea were shared with the Caspian Lake, and included large numbers of locally evolved crustaceans. Some of these endemics still survive, but many of them have been pushed back to the estuaries and coastal wetlands or limans, where salinity is low enough to suit their osmotic requirements. The entering Mediterranean water raised the salinity of the open water zone sufficiently to allow a large number of marine species and groups to successfully colonise the Black Sea. These include algae, and animals from protozoans and cnidarians to marine mammals such as dolphins and seals. Many truly marine species are still incapable of adjusting to the salinity of the Black Sea, however, and as a result, the species richness of the sea is only between 0 and 50% of that of the Mediterranean, depending on the group considered. The

deep hypoxic zone furthermore limits the number of species to those that can live in the littoral and in the surface layers of the pelagic.

In spite of this impoverished assemblage, many species became quite abundant and exploitable. Such is, or was, the case of numerous fish, mollusks (mussels, oysters), and of the sessile seaweeds *Cystoseira* (a brown alga) and *Phyllophora* (a red alga) that covered a surface area of thousands of km^2 until the 1960s, when they were heavily exploited as a source of agar.

DECLINE AFTER 1950

The history of the biota of the Black Sea after 1950 is invariably one of decline, due to eutrophication by nutrient-loaded surface and river runoff, overharvesting, introduction of alien species, and a combination of these, with an acceleration since the 1970s. A specific threat is intense shipping activity, including the traffic of crude oil from the harbour of Novorossiisk, endpoint of the Caspian oil pipeline, to the world market via the Bosphorus where at least one major accident occurred in the 1990s.

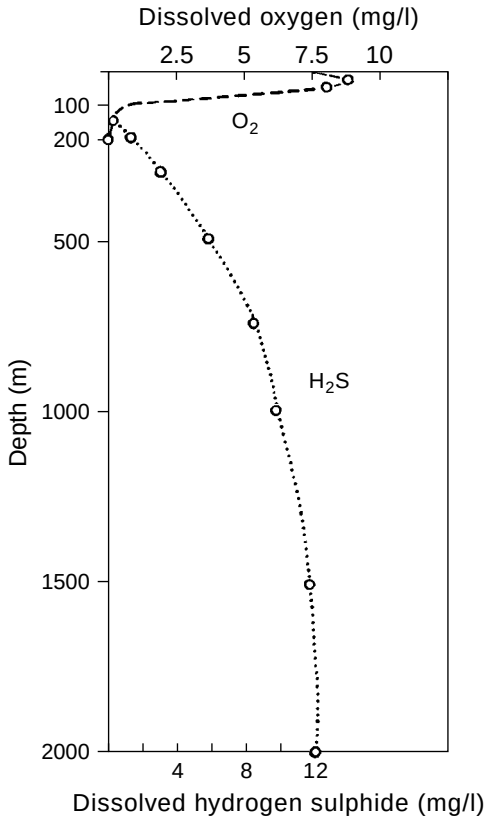


Figure 2 Vertical profile of dissolved oxygen (dashed line) and of hydrogen sulphide (dotted line) in the Black Sea. Both gases are mutually exclusive, and indicate the depth to which the water masses mix (modified from Mandych *et al.*, 1995). (Reproduced with permission of Backhuys Publishers, Leiden, The Netherlands)

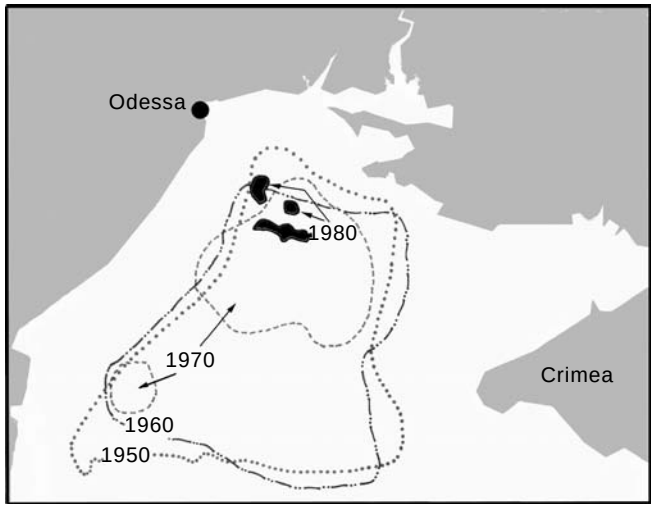


Figure 3 An example of the environmental degradation of the Black Sea and its biota: progressive reduction of the *Phyllophora* fields after 1950 (modified from Zaitsev and Mamaev, 1997). (Reproduced with permission of UN Publications, New York)

All biota have decreased, but the trend has been particularly spectacular in the north, where the enormous *Cystoseira* and *Phyllophora* algae-fields have shrunk to the point of almost complete annihilation (Figure 3). These algae were true keystone species, providing food and shelter to a complex community, ranging from protists to fish, and all these species declined or disappeared along with the algae. Eutrophication, progressive bottom deoxygenation encroaching the shelf areas since the 1970s, and stirring up of sediment by trawl fishing have all accelerated their demise.

Overfishing, in its turn, has caused the collapse of the populations of several pelagic fish like tuna, mackerel, swordfish and others, although in certain cases, the destruction of spawning grounds (sometimes situated in the Mediterranean) also contributed. The fisheries gradually shifted towards smaller species such as anchovy and sprat but these, in their turn, collapsed after the introduction – presumably with ballast water from the east coast of North America – of the fist-sized comb jelly *Mnemiopsis leidyi*, a voracious carnivore. This introduction may have occurred in the late 1970s, and was first spotted in 1982. By 1988, a fresh-weight biomass of 1×10^9 tons had developed, which wiped the sea clean of zooplankton, pelagic fish eggs, and fish larvae. Pelagic fisheries dropped by at least an order of magnitude and a large part of the fishing eet went out of business. Numerous other unintentional introductions took place, but none was so destructive as that of *Mnemiopsis*. After 1988, a relative decline occurred, perhaps due to shortage of prey, or an increase in *Beroe*,

another ctenophore, and one that preys on *Mnemiopsis* itself. The introduction of milkfish, a natural predator of *Mnemiopsis* in its natural range along the coasts of the east Atlantic, has also been proposed. In addition to containing the intruder, it might itself develop into an exploitable resource.

ACTIONS

The regional political situation was not conducive to coordinated action between the six littoral countries until the 1980s. It was therefore not until the deployment of large-scale, international initiatives in favour of the environment, that serious attempts at addressing the pressing problems of the Black Sea were undertaken. The Global Environmental Facility (GEF) of the United Nations (UN) launched the Black Sea Environmental Program in 1993. As a result of the regional cooperation that thus developed, a Black Sea Action Plan (BSAP) was agreed at a ministerial conference in Istanbul in 1996, and a permanent Black Sea secretariat was set up. This project worked out detailed assessments of the state of valuable biological resources, and plans (including required budgets) for their restoration or conservation. Clearly, in addition to donor funding, a substantial effort by the littoral countries themselves will be needed; success or failure of the BSAP is therefore now in political hands. A similar GEF programme on the Danube River, the single most important contributor of nutrients, organic residues, and occasionally highly toxic compounds to the Black Sea, has been carried out in parallel. A third one deals with the Dnepr River. Pollution abatement along the Danube, including its upper reaches in Germany and Austria, will be needed to significantly improve the situation of the Black Sea. Because the Black Sea traps most chemicals in its anoxic monimolimnion after not more than 1 year, the effect of such measures should rapidly become visible. It is therefore to be expected that the future of the Black Sea as an ecosystem will be decided within the first decade of the new millennium.

REFERENCES

Mandych, A F, Ryasintseva, N I, and Zaitsev, Y P (1995) The Black Sea, in *Enclosed Seas and Large Lakes of Eastern Europe and Middle Asia*, ed A F Mandych, SPB Academic, Amsterdam.

Zaitsev, Y and Mamaev, V (1997) *Marine Biological Diversity in the Black Sea. A Study of Change and Decline*, UN Publications, New York.

FURTHER READING

Dumont, H J (1999) *Black Sea Red Data Book*, UNOPS, New York and Kenobi Productions, Ghent.

Humborg, C, Ittekkot, V, Cociasu, A, and van Bodungen, B (1997) Effects of Danube River Dam on Black Sea Biogeography and Ecosystem Structure, *Nature*, 386, 385–388.

Sorokin, Y I (2001) *The Black Sea*, Backhuys, Leiden.

Caspian Sea

Henri J Dumont
Ghent University, Ghent, Belgium

The Caspian Sea, the largest lake on earth, is situated east of the Caucasus and north of the Elburz mountains. It is the endorheic terminus of the Volga River, which forms a large delta at its at north end, and contributes more than 80% of its yearly in ow, 237 km^3 year^{-1} on average (range 200- 450 km^3). More than 100 other rivers contribute the remaining 20%, with only the Kura and Ural signi cant. Lake volume is c. 78 000 km^3 (about 40% of the world's continental surface water). Lake surface area, situated about 28 m below ocean level, is close to 400 000 km^2. The shape is elongated (c. 1500 km) and is almost perfectly north- south oriented, thus stretching from a continental to a Mediterranean climate type. In addition, there is a steep precipitation gradient between the wet western shores, and the desertic eastern shores. The lake is composed of three basins: a shallow, freshwater north basin, and two deep and brackish central and south basins (Figure 1). Salinity in the deep basins reaches 12- 14 mille^{-1}, except in the salt-pan of Karabogaz, which may completely dry out, and usually contains a saturated brine. It is exploited for salt and for brine shrimp. In contrast to the Black Sea, the lake is oxygenated to the bottom, but at times of high lake stands, when much freshwater ows in, vertical circulation may become compromised, and anoxic conditions may develop in the two deep basins.

GEOLOGICAL HISTORY

The lake formed in the Miocene when the Tethys sea closed, as India and the Arabian peninsula hit the Asian continent, pushing up major mountain chains, and creating a number of landlocked basins, including the Black Sea, the South and Middle Caspian troughs, and the Amu and Syr Darja Rivers.

From the very beginning, the Caspian was a brackish lake, but it shrank and expanded on numerous occasions, and its salinity uctuated accordingly, from a nearly saturated brine at low lake stands, to almost freshwater at high

stands. At high lake stands, contacts were established with the Black Sea and with the Aral lake.

IMPACTS

About half of the heavy industry of Russia is situated in the Volga valley, and its effluents, washed to the Caspian, cause toxic pollution and eutrophication, which was particularly strong in the delta and the shallow north during the 1980s. After the economic collapse of Russia, the situation greatly improved in the 1990s. Other impacts have a more local character, and consist of the effluents of major cities, which may have strong effects on water quality if they are discharged to semi-enclosed, shallow bays, such as that of Krasnovodsk (presently Turkmenbashi) (Figure 1). Sloppy exploitation of oil fields, whether on land or offshore, is also a major source of pollution, for example, along the Kazakh coast. The bay of Baku is an example of the combined impact of a major city (Baku), its oil fields, and the industrial sewage of the industrial complex of Sumgait. As a result, Baku Bay is said to be biologically dead.

Leakage of radioactive wastes through a limestone bottom to the lake has been claimed to occur near the Gur'evskaia nuclear power plant in the north east corner of the lake.

WATER LEVEL FLUCTUATIONS

The hydrological balance of this closed basin is determined by the algebraic sum of river inflow, direct precipitation, and evaporation (neglecting groundwater inflow). Dry years in the Volga basin should thus be reflected in a negative balance, and a falling lake level. Things are not that simple, however, and the actual relationship between Volga inflow and lake level is statistically weak. This indicates that evaporation is a major factor too. In fact, in historic times, there has been an alternation of maxima (around the -26 m level) and minima (around the -30 m mark). In the 20th century, one maximum occurred in the 1930s, and a second one in 1995, separated by a low in 1977. A major man-made impact, the construction of a series of reservoirs in the Volga basin, took place in the 1940–1960s, coinciding with a receding phase. Little effect of this damming was felt by the littoral population, and by the time the lake started rising again, the reservoirs were all full.

One of the striking features of the 1930–1977 recession was a loss of ecological memory: many old taboos on shoreland use were broken and, as the lake fell, various constructions, including heavy industry, were built on previously flooded grounds. In the post-1977 rise, problems of flooding arose. These were comparatively mild in Iran, but involved oil fields in Azerbaijan, Turkmenistan, and Kazakstan, and caused severe pollution problems along the Azeiri shores, where major deposits of toxic solid waste were washed out from the industrial zone of the city of Sumgait. Corrective actions have included polder and dike building, and the main lesson learnt is clearly that history repeats itself, and that certain coastal zones should not be developed.

BIOTA

Biologically, the lake shows a percentage endemism in its fauna and flora comparable to that of Lake Baikal, although its total species richness is only about half that of its Siberian counterpart. The biota of the Caspian have been selected for euryhalinity, a capacity to adapt to wide changes in salinity, and are probably derived from estuarine ancestors with a freshwater origin. True marine groups are absent.

Because of the conspicuous absence of marine groups, and in order to increase the number of harvestable products from the lake, Soviet scientists embarked on a program of species introductions – almost all marine – as early as the 1930s. Some of these acclimatisations were successful, others were not. In addition, the commissioning of a canal between the Volga and the Don opened the Caspian to maritime traffic, allowing direct contact with the ocean through the Black Sea. Via this route, a number of additional species (especially those capable of settling on a ship's hull) entered the Caspian (e.g., balanids), while others escaped to the outside world, eventually creating problems of mass development elsewhere. A case in point is the bivalved mollusk *Dreissena polymorpha*, which caused great economic loss in the Great Lakes of North America.

The current use (and abuse) of ballast water in shipping has added another hazard. In the Black Sea, a brackish-water comb jelly, *Mnemiopsis leidyi*, originally found in estuaries along the east coast of North America, was thus introduced in the 1980s. This voracious predator eliminated zooplankton and larval fish to a point where, in 1989, it had reached a wet biomass of 10^9 tons. As a result, the Black Sea pelagic fishery, already stressed by eutrophication, pollution, and overexploitation, collapsed. If this species were to successfully invade the Caspian, it might rapidly destroy the local sardine or kilka fisheries, with possible fatal repercussions for the kilka-eating top predator, the endemic caspian seal.

In 1996, the first reports of fishermen sighting strange jellies in their nets started circulating; in 1999, scuba divers off the Sjevsjenko peninsula brought in the first certified video images of *Mnemiopsis* in the Caspian. A population explosion, with biomasses of up to $0.5\,\mathrm{Kg\,m^{-2}}$ (wet weight), was reached in late summer 2000 (Ivanov *et al.*, 2000).

Another valuable product of the Caspian is caviar, the roe of the four local species of anadromous sturgeon fish. At

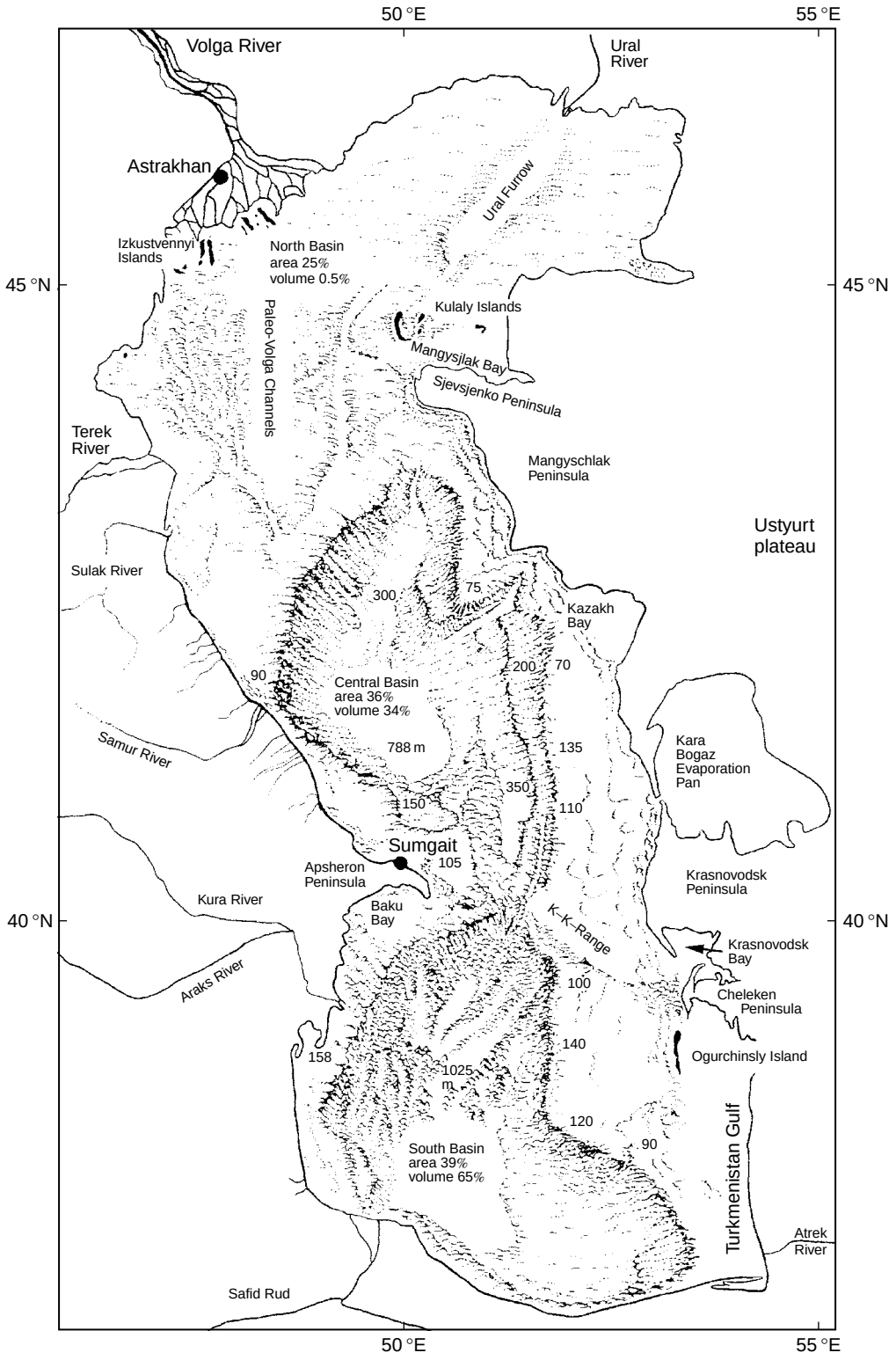


Figure 1 Map of the Caspian basin, identifying the major rivers, the three basins and their characteristics, and some localities cited in the text. The K–K range is a sub-lacustrine threshold that separates the central and south basins (Modified from Dumont, 1998 with permission)

certain times, over 90% of the world production of sturgeon was in the Caspian basin, with a peak of 39×10^3 tons in 1903. It should be noted that 7–10% of a female's body weight is caviar.

A first collapse of the fisheries occurred in the 1930s, because of simultaneous riverine and lacustrine fishing, destroying mature individuals migrating up the rivers (mainly the Volga) to spawn, and immatures in the lake. Legislation prohibiting pelagic fisheries resulted in a recovery after World War II. However, the damming of the Volga (and later of the Kura) soon caused a new problem: the spawning grounds of the sturgeon were ooded, and reduced to 12% of their original extent.

This situation, in its turn, was corrected by the construction of a series of sturgeon hatcheries in the surroundings of Astrakhan. The yearly release of fingerlings to the Volga reached a maximum of c. 100×10^6 specimens in the mid-1980s, but declined after the collapse of the Union of the Soviet Socialist Republics (USSR) because of inadequate funding and deterioration of the hatcheries. The collapse of the USSR also caused a beneficial decrease in the pollution levels of the Volga, which had caused massive muscle-tissue deterioration in sturgeon, peaking in 1987–1988, but there was an antagonistic increase in illegal fisheries, riverine as well as pelagic. As a result, the stocks of sturgeon began to decline once more. In the case of the largest sturgeon, the Beluga (*Huso huso*), declines are such (Figure 2) that extinction may be reached around 2003.

Attempts to revitalise the sturgeon aquaculture in the lower Volga by the World Bank and other international agencies have not, so far, produced tangible results, even though the technology for rearing sturgeon to adulthood is well worked out. Cultured caviar facilities, using non-Caspian or hybrid sturgeon species, are currently springing up around the world, and will likely provide the market with a high-quality product in the near future. Tragically, Russia will have lost one of its most traditional resources in the process. Once at par in economic value with oil production, it has now plumetted to only a tiny fraction of its former value. Rather than protecting all sturgeon species under the Washington convention, the best possible measure for keeping Caspian species from becoming extinct may be sturgeon-for-caviar farming elsewhere, since it may relieve pressure from the natural populations, and strongly reduce the demand of the world market for Caspian-derived caviar.

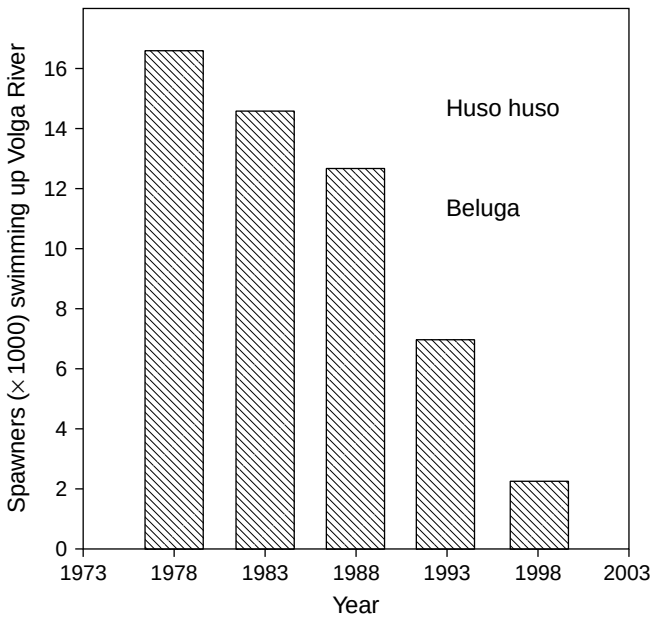


Figure 2 The decline of the Caspian population of Beluga in the late 20th century, as measured by the number of adults swimming up the Volga River to spawn

INTERNATIONAL EFFORTS AT REHABILITATION

The disintegration of the Soviet Union, the dominant player in the region, created a new and complex political situation, with five instead of two riparian nations. Much previous legislation became obsolete, and cooperation between the new states was initially wrought with problems of distrust and con icting interests. Kazakhstan and Azerbaijan gained a sudden strategic importance because of their proven or expected oil reserves, most of which lie below the bed of the lake. A legal tangle developed around the status of the Caspian: lake (in which case it would be divided into five sectors), or sea (with the law of the sea allowing each coastal state a limited zone of exclusive economic rights, besides a vast common zone)? This delayed the development of a common policy in environmental matters, in spite of efforts since 1994 by the World Bank, the United Nations Development Programme, and the United Nations Environment Plan to start a major Global Environment Facility (GEF) project addressing the most pressing issues of the lake. Shortly before the close of the century, a GEF and a TACIS (European Union funded) program was finally launched, aiming at mitigating environmental degradation, and regulating the exploitation of the living (mainly fish and seal) and non-living (mainly oil) resources of the basin.

Among the initiatives taken by this programme, a meeting in Baku, in late April 2001, should be cited, in which experts and government officials tried to work out a strategy to counteract the *Mnemiopsis* invasion, which already showed signs of affecting the sardine populations, and even the seals. An agreement was reached that, as soon as *Mnemiopsis* biomass would reach $1\,\mathrm{Kg\,m^{-2}}$ (wet weight), the procedure for an intentional introduction of *Berae cf. ovata*, a jelly predator of *Mnemiopsis*, should be initiated immediately.

REFERENCE

Dumont, H J (1998) The Caspian Cradle, in *The Predatory Cladocera of the World*, ed I K Rivier, Backhuys Publishers, Leiden.

FURTHER READING

Dumont, H J (1998) The Caspian Lake: History, Biota, Structure, and Function, *Limnol. Oceanogr.*, 43, 44–52.

Ivanov, V P, Kamakin, A M, Ushivtsev, V B, Shiganova, T, Zhukova, O, Aladin, N, Wilson, S I, Harbison, G R, and Dumont, H J (2000) Invasion of the Caspian Sea by the comb jellyfish *Mnemiopsis leidyi* (Ctenophora), *Biol. Invasions*, 2, 255–258.

Kaplin, P (1995) The Caspian: its Past, Present and Future, in *Enclosed Seas and Large Lakes of Eastern Europe and Middle Asia*, ed A F Mandych, SPB Academic, The Netherlands.

Kosarev, A N and Yablonskaya, E A (1994) *The Caspian Sea*, SPB Academic, The Netherlands.

Rodionov, S N (1994) *Global and Regional Climate Interaction: the Caspian Sea Experience*, Kluwer, The Netherlands.

Mediterranean Sea

Michel Batisse

Centre D'activitees Regionales du Plan Bleu, Sophia Antipolis, France

The word Mediterranean stands for 'in the midst of lands'. Indeed the Mediterranean Sea - a large, deep body of water extending from Gibraltar to the Dardanelles and the Suez Canal - is totally surrounded, and even encroached, by its bordering lands and innumerable islands. While these lands and islands have long been referred to as the Mediterranean region, the actual limits of this region are hard to de ne on a map. It enjoys a characteristic climate, with a dry summer season and irregular winter rains, but this climate extends to non-bordering countries like Portugal and Iraq. It is the land where the olive tree grows but this is dif cult to delineate. The entire territory of the bordering countries, some of which are only partly Mediterranean like France, Morocco or Turkey, should be considered when discussing environmental policies and regulations. The catchment basin - excluding the upper Nile - is a useful frame for water-related issues. The administrative coastal regions, with their nearby rugged hinterlands, provide the most useful set of truly Mediterranean data. The coastal fringe has to be considered for urbanization, tourism, etc. Thus, what is known as the Mediterranean region has a variable geographic dimension depending on the issue under study.

CONTROLLING MARINE POLLUTION

Environmental concern in the Mediterranean was originally focused on marine pollution. Twenty-five years ago, this resulted in the adoption, under the auspices of the United Nations Environment Programme (UNEP), of the Barcelona Convention for the protection of the Mediterranean Sea against pollution. This convention has given birth to a number of related protocols and is accompanied by a Mediterranean Action Plan to facilitate its implementation in a broader context of cooperation. The most remarkable feature of this regional effort is that it associated all bordering States, without exception, in spite of their various and well-known political disputes, around common objectives.

Contrary to what has been somewhat over-stated, the Mediterranean Sea as a whole is not much more polluted than other parts of the world ocean. What is true is that it constitutes a major navigation route, with some 300 oil tankers at sea at any one time. The Mediterranean is also a major touristic zone, with millions of vacationers and their myriad of sail boats. As result, the high sea is subjected to an unreasonable amount of oating debris–particularly plastics – as well as by tar balls originating from illegal deballasting and tank washing operations. These latter practices are now more or less under control, thanks to the implementation of MARPOL 73/78 (*see* MEPC (Marine Environment Protection Committee), Volume 4) and of the protocols of the Barcelona Convention relating to dumping from ships and to cooperation in case of emergencies. It is worth noting that the Mediterranean waters have also a relatively high mercury content. This, however, is due to natural sources and would present some health hazards only if a very large amount of seafood were consumed.

The most serious problem lies in the pollution of coastal waters. This pollution comes mainly from domestic and industrial ef uents from the coastal regions, which are the home of some 150 million people to which some 120 million visitors each year should be added, and from the discharge of all the rivers in the basin, which carry solid and liquid waste from the hinterland, including agricultural pesticides and fertilizers, which have caused important algal blooms in the Adriatic.

Much has been done over the last two decades to reduce local sources of pollution, particularly along the northwestern industrialized coast and also in some critical areas such as Kastela Bay in Croatia and Izmir in Turkey. More than 50% of the Mediterranean coastal towns with a population over 10 000 are now served by wastewater treatment plants and half of these have secondary treatment. Yet the others, as well as many industries and isolated tourist resorts, still discharge their ef uents directly, and some 9 million cm^3 of untreated wastewater go to the sea each day. Some 5 million tons of urban solid waste further compound the overall pollution problem along the coastline

every year. However, the Mediterranean countries have now made the commitment to establish sewage and solid waste disposal systems for all urban agglomerations exceeding 100 000 people by 2005.

The impact of the range of domestic, industrial and agricultural pollutants has been extensively studied under the Mediterranean pollution program (MEDPOL), a component of the Mediterranean Action Plan (UNEP-MAP, 1997). The most important protocol concerning land-based sources of pollution, signed in 1980, was updated in 1996. The time is ready for a major operational effort to reduce remaining pollution hot spots.

FROM THE SEA TO THE LAND

When the Barcelona Convention and the Mediterranean Action Plan were under negotiation in 1975, it was already clear to some people that most of the pollution affecting the sea resulted from the way certain activities were carried out on land and that therefore the effective control of marine pollution would depend on changes in the development process taking place in the bordering countries. On these grounds, it was decided to undertake an in-depth analysis of the environment–development interactions in the entire region, to study how these could evolve and in what way the overall situation of tomorrow would depend on what was taking place or decided today. This component of the Mediterranean Action Plan, based on system-analysis and on future-oriented studies, came to be known as the Blue Plan, a symbolic reference to the color of the sea that the convention was meant to protect (*see* Blue Plan, Volume 4).

The method chosen by the Blue Plan was to define the key elements of the Mediterranean system, to evaluate their current interactions, and then to build scenarios on their possible future evolution. The elements considered at appropriate geographic levels were the following: population dynamics, urbanization, agriculture, industry, energy production, tourism, transport, soils, water resources, vegetation cover and biodiversity, coastal land use, and marine pollution.

This Mediterranean system – particularly when considering the coastal regions – is unique. This results from its long common history, which has forged parallel ways of life, and from its typical geographical features, characterized by the presence of the sea, a unique climatic pattern, a rich endemic biodiversity and a rugged hinterland. However, three major facts lead to strongly contrasting situations between the northern European shore and the southern and eastern shores spreading from Morocco to Turkey.

The first contrast is in demography and in the related issue of urbanization. Populations tend to concentrate in the coastal regions on both sides, due to attraction to the sun and the sea in the industrialized North, and due to rural migration from poor hinterlands in the South. However, in spite of a marked decrease in fertility rates, the southern and eastern countries continue to see their population grow rapidly and become younger, while in the North population will at best remain stable and become older (Figure 1). These two opposite trends have major consequences in terms of employment, food, urbanization, land and water use, steel and cement production and pollution. Thus, the growth of urban population in the south will be difficult to manage in social and environmental terms (Figure 2).

This demographic contrast is the corollary of an equally strong north–south disparity in levels of economic development and quality of life. Even when corrected by purchasing power parity, gross national product (GNP) *per capita* remains seven times higher in France than in Egypt. Twenty Mediterranean countries are currently parties to the

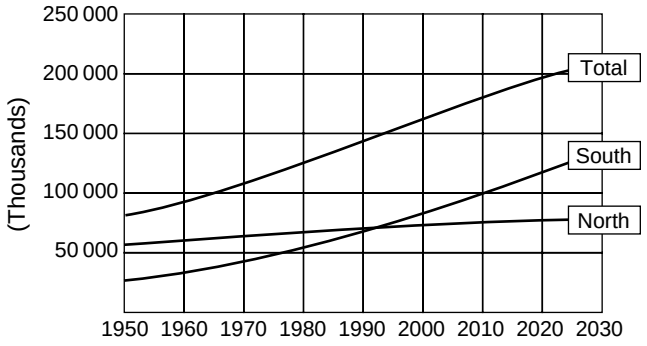


Figure 1 Mediterranean coastal population

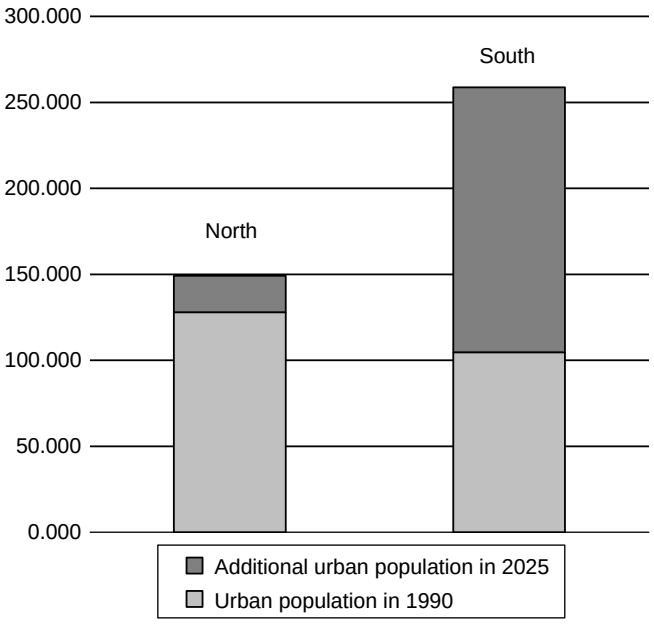


Figure 2 Urban population of the countries of the north and of the south and east of the Mediterranean Basin 1990–2025

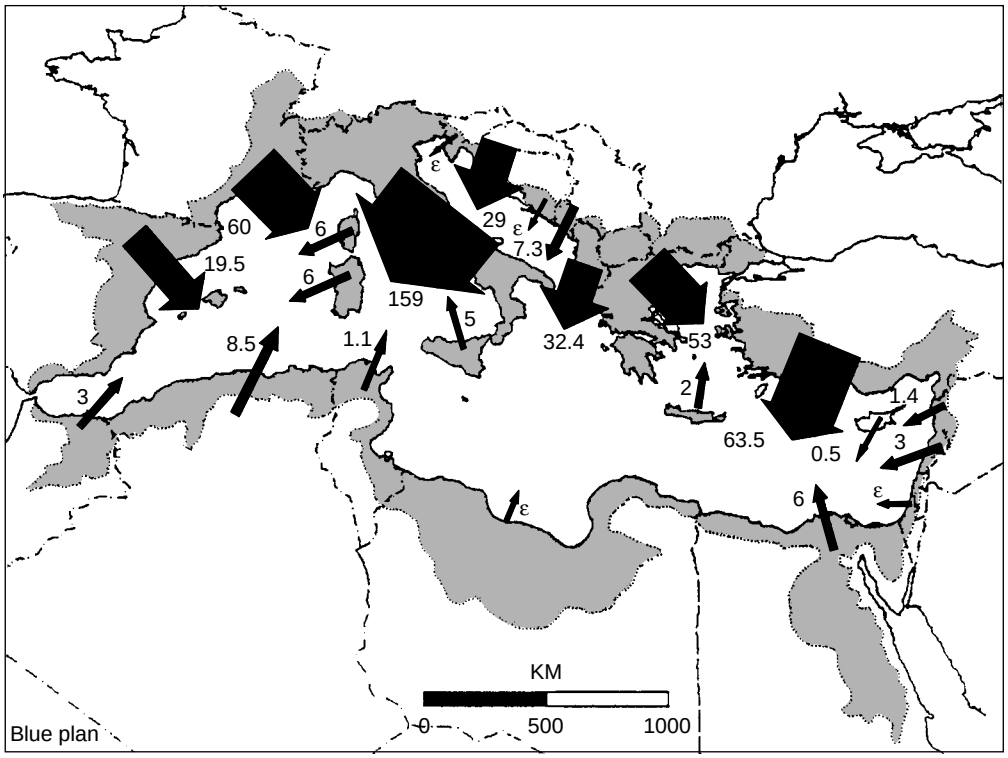


Figure 3 Real annual flows from each country to the Mediterranean (km^3). The grey area represents the Mediterranean catchment basin. The figures given are in billion m^3

Barcelona Convention. For 1998, according to the World Bank, their GNP *per capita* is as follows (US$): France and Monaco 24 940, Italy 20 250, Israel 15 940, Spain 14 080, Greece 11 650, Slovenia 9760, Malta 9940, Croatia 4520, Lebanon 3560, Turkey 3160, Tunisia 2050, Algeria 1560, Egypt 1290, Morocco 1250, Syria 1020, Albania 810. Corresponding data for Bosnia, Cyprus and Libya are not available. When corrected by purchasing power parity, disparities between countries are somewhat smaller.

A third and equally significant contrast results from the difference in water availability. While the term *Mediterranean climate* suggests a common pattern, rainfall is far lower in the South than in the North. In addition, the northern shores (including Turkey) are backed by humid regions providing important river ow, while the southern countries are backed by deserts, with negligible supply. The Nile was once a major exception, but its out ow has almost totally disappeared since the Higher Aswan Dam was constructed (Figure 3). In terms of climate change, the studies conducted so far seem to indicate a possible increase in aridity and in rainfall variability.

The Blue Plan scenarios, built upon coherent sets of assumptions for the period extending from 1985 to 2025, illustrate the likely consequences of different development, environment and cooperation strategies on the future of this Mediterranean system. This *prospective* approach was applied for the first time to such a vast region and could possibly be used in other parts of the world. It amply demonstrates that trend scenarios, even favorable ones based on strong economic growth, lead to serious inequities and more or less irreversible environmental damage due to excessive pressure on resources for the southern and eastern countries – particularly on the coast – and are therefore not acceptable. Only scenarios that take into account environmental protection and social equity, combining appropriate national strategies with enhanced Mediterranean cooperation, lead to what could be called sustainable development for the entire region and could therefore avoid major domestic and international tensions (Grenon and Batisse, 1989).

Clearly, sustainable development in the Mediterranean cannot be achieved independently of what will take place elsewhere, since the region is not isolated economically and technologically from the rest of the world. However, it represents in many ways a microcosm of the entire planet and its future prosperity depends upon a number of imperatives which may be of relevance to other areas. These key imperatives, derived from the conclusions of the Blue Plan scenarios and endorsed by the Mediterranean countries, can be summarized as follows:

1. to incorporate *a priori* (rather than through corrective measures) the protection of the urban, rural and natural

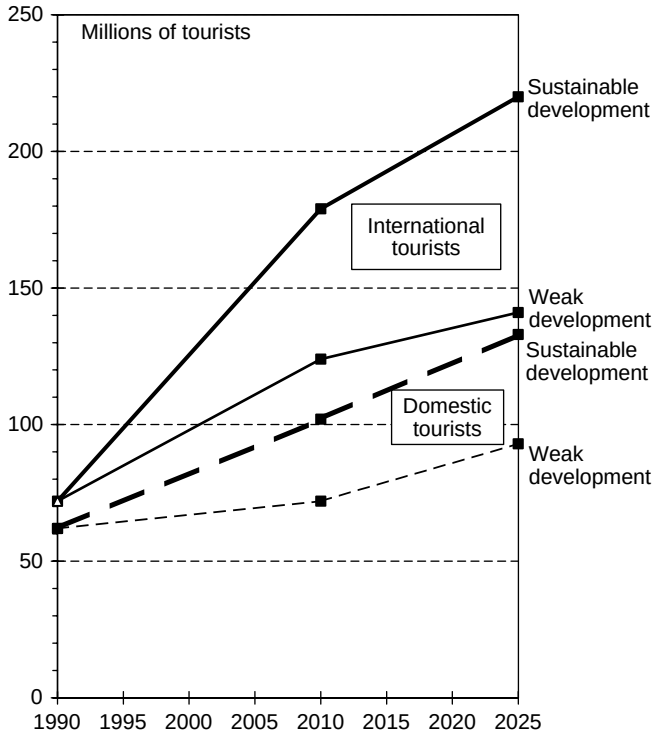


Figure 4 International tourists and domestic tourists in the Mediterranean coastal regions (according to two scenarios)

environment in all development strategies and projects in the region;

2. to ensure integrated management of coastal areas and of the coastal fringe (on both the terrestrial and marine sides), which constitute major natural resources for all Mediterranean countries, notably for tourism; the development of international tourism is essential for most countries in the region for their balance of payments; at the same time, sustainable development increases also domestic tourism (Figure 4);
3. to ensure integrated management of water resources, stressing improved management of water demand, particularly for irrigation; water resources *per capita* have already fallen below the level of stress of $1000\,m^3\,year^{-1}$ for all countries between Morocco and Syria except Lebanon (Benblidia *et al.*, 1997);
4. to promote through education and the media a better perception by all concerned – including decision-makers and the general public – of the economic, social and environmental interactions in the region;
5. to develop and effectively implement economic cooperation between the countries of the European Union and the countries south and east of the Basin, and also amongst the latter countries themselves, favoring bilateral and regional programs and projects of sustainable development.

CHANGING THE SCENARIO

While the Mediterranean countries have not yet taken sufficient measures to meet the above imperatives, a number of positive developments indicate progress is being made. The question is rather whether the pace of progress in policies and actions is sufficient to counter the negative trends, which have to be reversed.

In this respect, a first requirement is to properly assess and monitor the environmental situation. While demographic and economic data are usually available, environmental data are often missing or very localized. The Blue Plan, which is acting as a Mediterranean Environment and Development Observatory has stimulated the establishment of corresponding national observatories in such countries as Lebanon, Morocco, Syria, Tunisia and Turkey, while official bodies in the European countries are fulfilling the same functions.

As a follow-up of the United Nations Conference on Environment and Development (UNCED), the Tunisian authorities convened in 1994 a ministerial conference on sustainable development in the Mediterranean, which formulated an Agenda MED 21, as an adaptation of Agenda 21 to the specific features of the region.

In 1995, 20 years after its adoption, the Barcelona Convention was amended in order to extend its scope beyond marine pollution and in particular to include protection of the coastal regions (as defined by each country within its own territory). At the same time, the protocols of the Convention concerning dumping from ships and specially protected areas were strengthened.

The Mediterranean Action Plan itself entered a second phase, referring explicitly to the sustainable development of the coastal areas of the region, including activities aiming at the integration of environment and development in such sectors as agriculture, industry, energy, tourism, transport, urban management, water resources, soils, living marine resources, forest and plant cover, conservation of nature, landscapes and sites, in addition to the previously carried out measures relating to the assessment, prevention and elimination of marine pollution.

The budget collectively fixed by the bordering countries and the European Commission for the implementation of the Mediterranean Action Plan hardly exceeds 6 million dollars per year, which could appear to be totally disproportionate relative to the ambitious tasks to be covered in this vast region. It should be stressed, however, that these very limited funds are primarily meant to catalyze individual actions by the various countries and to harmonize such actions in the region. While it would be difficult and somewhat artificial to provide a figure for the total effort of Mediterranean countries to respond to the challenge of environmental issues, it is widely recognized that the existence of the Barcelona Convention and of the Mediterranean

Action Plan has begun to change the situation in the right direction.

It should be also noted that the Convention and the Action Plan have triggered important sustainable development projects supported by international financing organizations. Thus, the World Bank, the European Investment Bank, the United Nations Development Programme (UNDP) and the European Commission have joined forces since 1990 in a Mediterranean Environment Technical Assistance Program (METAP) which has sponsored a full array of projects in the southern and eastern countries of the region, concerning such issues as environmental performance indicators, urban water treatment, solid waste disposal, coastal areas management, capacity building, establishment of cooperative networks, etc.

In the same vein, the Mediterranean region provides an ideal place for implementing projects relating to international waters and to biodiversity under the Global Environment Facility (GEF). Thus, on the basis of the work carried out under MEDPOL, a very important scheme is underway for effective control of land-based pollution hot spots. A Strategic Action Program has been prepared and a first year project amounting to some 13 million dollars (including a contribution from the French GEF) allows concrete actions to begin. Similarly, UNDP/GEF is committed to a project aimed at the protection of coastal and wetland ecosystems amounting to some 13 million dollars and covering Albania, Egypt, Lebanon, Morocco, Tunisia and Palestine. Here again, it should be stressed that METAP and GEF funds require counterpart financing by the recipient countries, usually of a far greater magnitude, thus demonstrating their commitment towards environmental improvement.

SUSTAINABLE DEVELOPMENT, SECURITY AND SOLIDARITY

Now that environmental issues have been identified and important steps have been and are being taken to solve them, the path towards sustainable development must be defined. In this respect, responding to the 1995 ministerial meeting in Tunis, the contracting parties to the Barcelona Convention have decided to set up a Mediterranean Commission on Sustainable Development (MCSD). This is an interesting body, which brings together representatives from all bordering countries and the European Commission together with representatives from a number of local authorities, of socio-economic actors and of non-government organizations. While this commission has only a consultative status, its in-depth reports on certain issues like sustainable development indicators, management of water demand, sustainable tourism, etc., provide a common framework for all actors concerned in the region and a basis for formulating projects to bring to the attention of financing bodies.

At the same time, at the national level, most Mediterranean countries as well as a number of regions, provinces and cities in the Basin are preparing their own Agenda 21 and their own plans in accordance with the recommendations of the UN Commission on Sustainable Development and of MCSD. It is interesting to note for example that all countries recognize the need to limit and plan to reduce the share of irrigation in their total water demand which, difficult as it may be, represents the best approach to avoid situations of scarcity (Figure 5).

Much more, however, is needed. As shown by the Blue Plan scenarios, responding to environmental change in the Mediterranean cannot be dissociated from the question of balanced and equitable development in the region, which in turn implies deliberate solidarity between North and South. This part of the world has too long been cursed by international or domestic con icts, be it in North Africa, in the Near East or in the Balkans and the time has come for an effective partnership.

It is encouraging to note that the European Union has become aware of the mounting dangers which the current demographic, economic, social and environmental trends would bring to the region if a serious cooperation effort was not undertaken. As put by French President Chirac in

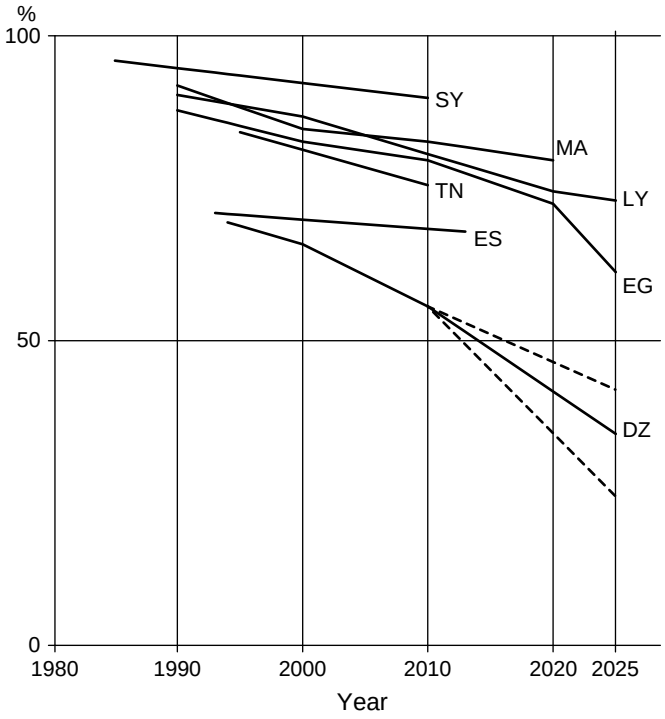


Figure 5 Projections of irrigation shares in the total water demands in some Mediterranean countries (as indicated in national plans). DZ, Algeria (three scenarios); EG, Egypt; ES, Spain; LY, Libya; MA, Morocco; SY, Syria; TN, Tunisia

an address to students in Cairo: after breaking a wall to the East, Europe must now build a bridge to the South . In November 1995, the Ministers of Foreign Affairs of the European Union and the southern and eastern Mediterranean countries met in Barcelona to adopt a declaration of major significance. Its aim is to guarantee peace, stability and prosperity in the region and in particular to set it up as a free-trade zone by 2010. In order to reach these objectives, the decision was made to carry out from 1997 onward a number of cooperative projects between the European Commission and the other countries concerned on a bilateral or regional basis. A mechanism and a fund called MEDA were set up to implement this program. The fund amounts to some 5 billion dollars for the first 3 years and currently being renewed.

Under the economic cooperation component of this program (the other two focussing on security and on culture), it is important to note that agriculture is excluded, but that both environment and water are given a significant place. Thus, under MEDA a basin-wide information system on technical know-how concerning water, Euro-Mediterranean Information System (Emwis), is already in place, a short- and medium-term action program for the environment has been approved (Short and Medium Term Priority Environment Action Program, SMAP) and the Blue Plan has been entrusted with a project for developing environmental statistics in the region.

Like elsewhere in the world, the objective of sustainable development in the Mediterranean is ambitious and complex and requires major changes in the behavior of all economic and social actors. This region, however, may well offer some comparative advantages in this respect. The environmental issues have been thoroughly analyzed, mechanisms for north–south cooperation have been established, bordering countries demonstrate growing commitment, and the European Union has the means to provide the fuel for change. Much will however depend on political stability and on the strengthening of the democratic process in the south. Much will also depend on the continued support of the people from the European countries which do not border the Mediterranean. Many of these people indeed come each year to enjoy its unique natural and cultural heritage and should contribute in some ways to its conservation.

REFERENCES

Benblidia, M, Margat, J, and Vallée, D (1997) *Water in the Mediterranean Region*, Plan Bleu, Sophia Antipolis, 91.

Grenon, M and Batisse, M (1989) *Futures for the Mediterranean Basin: the Blue Plan*, Oxford University Press, Oxford, 279.

UNEP-MAP (1997) Transboundary Diagnostic Analysis for the Mediterranean Sea (Draft). UNEP (OCA) MED IG 11/Inf.7, UNEP, Athens.

Aral Sea

Michael Glantz

National Center for Atmospheric Research, Boulder, CO, USA

The Aral Sea is a terminal inland sea that is sandwiched between two Central Asian deserts. The existence of the sea depends on the inflow from two major rivers: the Amudarya and the Syrdarya. Their combined flow at one time equaled 110 km^3 per year. In 1960, the sea was the fourth largest inland sea on the globe. Today, just four decades later, it has fallen to sixth place. There is little doubt as to the cause of the shrinking of the Aral Sea, both in volume and in surface area. Decisions to divert water from the two major rivers that have fed the sea for millennia have resulted in less water reaching the sea. Evaporation from the sea surface no longer balances the annual inflow from the rivers. In this short span of time the level of the sea has dropped from 53 m in 1960 to 36 m in the late 1990s. This has caused the sea to divide into two parts: the Little Aral to the north fed by the Syrdarya and the Big Aral to the south fed by the Amudarya. Its surface area decreased from 66 000 km^2 in 1960 to 32 000 km^2 in the mid-1990s. Water salinity also increased in this period, changing from up to 10 g l^{-1} in 1960 to more than 35 g l^{-1} in the 1990s, the same level of salinity as the open ocean. If sea level continues to drop another 5 meters or so, the sea will break up into 12 small saline ponds.

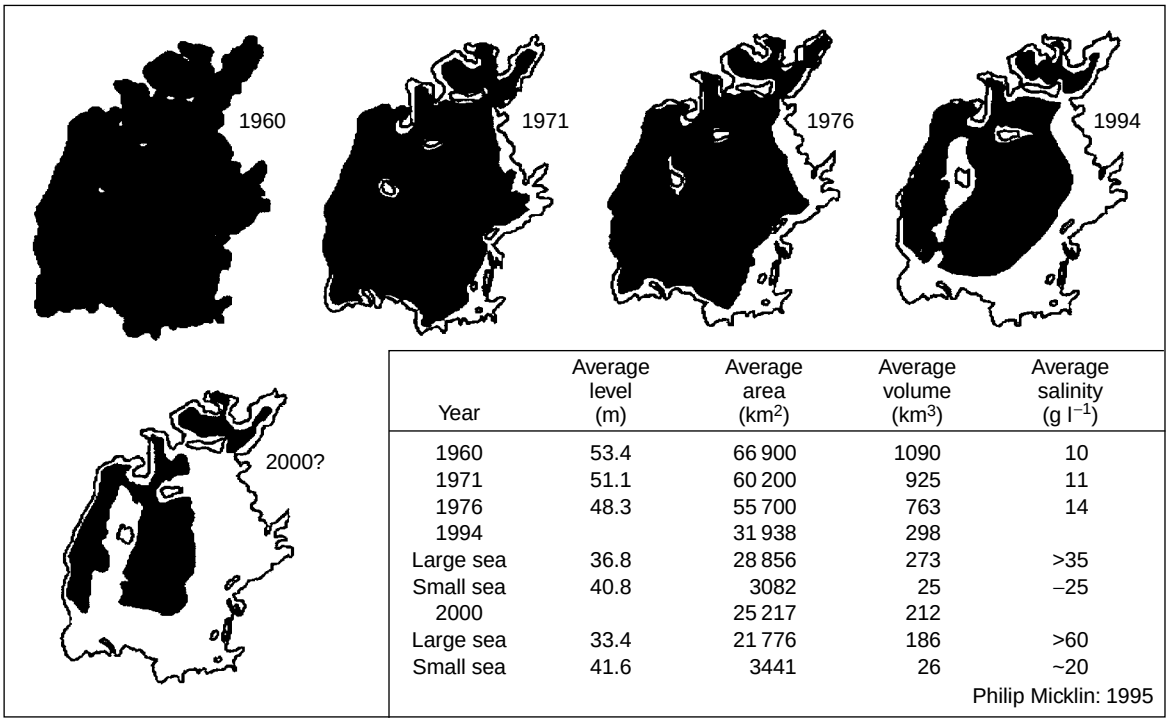
The demise of the Aral Sea (Figure 1) has been called one of the worst human-caused environmental disasters of the 20th century. A review of historical information about the Aral region shows that the drop in sea level was not an unanticipated result of decisions to divert water for irrigation purposes into the desert landscape. If and when the sea disappears, as will likely happen in the next couple of decades, it will have followed a scenario that had been suggested in 1927 by Russian scientist V V Tsinzerling. About 50 years later, Victor Kovda, Professor of Soil Science at Moscow State University spoke frequently of the plight of the Aral Sea when he was the President of the International Council for Science (ICSU)'s Scientific Committee on Problems of the Environment (SCOPE).

The problem started in the 1950s. It was then that decisions were made in the Kremlin by Communist leaders of the Soviet Union to expand the area under irrigation in Central Asia's Karakum and Kyzylkum deserts. The goal was to increase sharply cotton production in the arid Central Asian Republics (CARs), a region with almost perfect conditions for growing cotton, except for the lack of water. That agricultural boundary constraint could be remedied, however, by massive water diversions to the fertile but dry desert sands. As a result, the area under irrigation in the region was expanded from 4 million ha to more than 7 million ha in just a few decades. In the mid-1950s the construction of the Karakum Canal was begun. The canal diverts an estimated 10–15 km^3 per year of water from the Amudarya into Turkmenistan. Today, this canal is the longest human-built canal in the world with a length of about 1400 km.

In the late 1970s the Syrdarya, and in the late 1980s the Amudarya, failed to reach the sea at all. In the early 1990s, the rivers' flows reached the sea once again, as a result of abundant snow pack in the Pamir and Tien Shan mountains, the source of water for these rivers.

Water diversions from the rivers were only part of the region's environmental problems that were related to the rapid expansion of irrigated cotton production. In the process of increasing cotton yields, massive amounts of fertilizers, pesticides and herbicides were used, based on the belief that if a small amount is beneficial, then a big amount would be much better. Large portions of these chemicals eventually made their way into drainage ditches, the groundwater and finally back into the rivers and ultimately the sea. Eventually, they ended up in the fish and in the drinking water.

As the sea dried and its shoreline receded, the newly exposed seabed, containing salts and other agricultural chemicals, became exposed to wind action as early as the mid-1970s. Dust storms were generated that were large enough to be seen from space by Soviet cosmonauts. The storms filled the air with toxic substances, depositing them on agricultural fields at a distance of a few hundred



Year	Average level (m)	Average area (km^2)	Average volume (km^3)	Average salinity ($g\ l^{-1}$)
1960	53.4	66 900	1090	10
1971	51.1	60 200	925	11
1976	48.3	55 700	763	14
1994		31 938	298	
Large sea	36.8	28 856	273	>35
Small sea	40.8	3082	25	–25
2000		25 217	212	
Large sea	33.4	21 776	186	>60
Small sea	41.6	3441	26	~20

Philip Micklin: 1995

Figure 1 The changing profile of the Aral Sea 1960–2000. (Reproduced with permission from Springer-Verlag Micklin and Williams, 1996)

kilometers to the south and west of the sea. The combination of air and water pollution had major adverse effects on the health status of the population living immediately around the Aral coastline. This area, known now as the disaster zone, has deteriorated sharply, leaving many inhabitants of the region with birth defects, short life expectancy, anemia, and several other serious illnesses.

Although various Russian scientists bravely cautioned Soviet political leaders for decades about the adverse environmental impacts on the sea and the fishing sector that depended on the sea, the poor decision-making was not exposed until Mikhail Gorbachev became the Secretary General of the Communist Party of the Soviet Union in the mid-1980s. At that time his new policy of *glasnost* (openness) exposed the Aral crisis to the Soviet public as well as to the rest of the world.

Since then, there have been many national and international workshops and conferences to discuss the future use of regional water resources and the fate of the sea and its surrounding territory. One of the more popular ideas put forth initially by World Bank consultants was to let the sea disappear but to preserve the deltas of the Amudarya and the Syrdarya. The deltas are important wetlands that are biologically productive for both ora and fauna native to the region. Another less popular earlier suggestion had been to divert water from Russia's north- owing rivers to the south to replenish the Aral Sea and to provide additional water for cotton and rice production in the region. Yet another more recent proposal has been to transfer water (uphill) from the rising Caspian Sea to the declining Aral Sea, 400 km to the east (Dukhovny and Sokolov, 1997).

At one time the sea had supported a robust fishing industry that captured upwards of 50 000 metric tons per year of various commercially favored species and employed tens of thousands of inhabitants in the region. As the water quantity declined, the Aral's seaports of Aralsk and Muynak became separated from the receding coastline by tens of kilometers. The desiccated seabed was strewn with abandoned and rusting fishing vessels (Figure 2), an image which has since become a symbol of the degradation of the sea and human health as a result of deliberate political decision-making.

As the water quality declined, the fish population declined with it. Eventually, the Soviet government imported fish into the region from the Pacific Ocean and from the Baltic Sea to keep the fish processing factories operating. Finally, this became too costly and the practice was discontinued.

For a few years after the breakup of the Soviet Union in December 1991, the newly independent CARs – Kyrgyzstan, Turkmenistan, Kazakhstan, Uzbekistan, Tajikistan – voiced their common goal as Saving the Aral Sea . However, it became clear almost immediately that it was a goal that they could neither attain nor one they had the desire to reach. Sustaining cotton production remained a primary objective as the way to generate sorely needed foreign exchange to be then used for

Figure 2 The skeleton of a boat stands in a sandy desert that once was the port of Muynak, now tens of kilometers from the Aral Sea (photo by M H Glantz)

economic development. There was no voluntary reduction in irrigation, or in the acreage devoted to cotton production. Irrigated land, however, had to be abandoned because of poor yields or extreme soil salinization. There has also been a sharp reduction in the use of agricultural chemicals, not because of their adverse impacts on human settlements and ecosystems, but because of their high cost.

Despite the political rhetoric to the contrary, there is still competition among the CARs for control over the limited water resources in their region. With newly found independence, the Republics' leaders have shown reluctance to compromise on their perceived national water needs, although there are local, national and regional attempts to make more efficient use of that water. Problems related to water sharing are likely to continue in the future as upstream countries (Kyrgyzstan, Tajikistan, Afghanistan) seek to increase the use of water within their borders.

An important characteristic of the Aral Sea that has apparently been overlooked by the five CARs and donor nations is its symbolic value. Environmental and other groups around the globe have shown considerable interest in the fate of the sea since the mid-1980s. If the CARs were to let the sea break up into a dozen or so small saline ponds, it is highly likely that many of those groups will turn their attention, interest and resources elsewhere to other serious environmental problems for which their involvement could make a difference.

In the absence of major changes either in the national policies of the five CARs or in a natural increase in river flow to the sea, the sea will likely disappear in the not too distant future. As it is today, the remnants of the sea could serve as an example of what happens when societies do not respect the limits on the use of their technologies and the limits of Nature to absorb the insults of misapplied human activities on a fragile environment.

To protect the Aral Sea from further degradation, it has been suggested that the sea in its current state be given World Heritage Site status by UNESCO (United Nations Educational, Scientific and Cultural Organization). This designation would mean that the countries in the Aral Sea basin would have a responsibility to preserve the sea at least at its present level; and it would also oblige the international community to support the CARs in their efforts to protect the sea. As a heritage site, the Aral Sea would serve as a constant reminder of the failure of a political system to protect the fragility of the natural environment.

REFERENCES

Dukhovny, V A and Sokolov, V I (1997) Development of the Caspian–Aral Seas Program: An ICWC Perspective, in *Scientific Environmental, and Political Issues in the Circum-Caspian Region*, eds M H Glantz and I S Zonn, NATO ASI Series, Kluwer, Dordrecht, 185–190.

Micklin, P and Williams, A T (1996) *The Aral Sea Basin*, Vol. 12, NATO ASI Series, The Environment, Springer-Verlag, Berlin.

FURTHER READING

Bortnik, V N (1999) Alteration of Water Level and Salinity of the Aral Sea, in *Creeping Environmental Problems and Sustainable Development in the Aral Sea Basin*, ed M H Glantz, Cambridge University Press, Cambridge, 47–65.

Glantz, M H and Figueroa, R M (1998) Does the Aral Sea Merit Heritage Status? *Global Environ. Change*, 7, 357–380.

Glantz, M H, ed (1999) *Creeping Environmental Problems and Sustainable Development in the Aral Sea Basin*, Cambridge University Press, Cambridge.

Ruffin, M H and Waugh, D, eds (1999) *Civil Society in Central Asia*, University of Washington Press, Seattle, WA.

Zonn, I S (1999) The Impact of Political Ideology on Creeping Environmental Changes, in *Creeping Environmental Problems and Sustainable Development in the Aral Sea Basin*, ed M H Glantz, Cambridge University Press, Cambridge, 157–190.

Great Lakes Region of North America

Henry Regier[1] and Linda D Mortsch[2]
[1] *University of Toronto, Toronto, Canada*
[2] *Waterloo, Canada*

Region-wide efforts to correct particular abuses of modern industrialization and urbanization began late in the 19th century, and have expanded and intensified gradually

since then. Regional manifestations of some global problems came into focus in the 1980s; preventive and adaptive efforts then began slowly. Particular regional initiatives are described concerning several global issues: hazardous contaminants; climate change and stratospheric ozone depletion; and water diversions and allocations as related to commodi cation of water in free trade.

INTRODUCTION, THE GREAT LAKES REGION

The waters of the Great Lakes are an integral part of the natural and cultural heritage of eastern North America, and have global recognition as valued ecosystems. Scores of small nations of Aboriginal peoples of the Anishinabek and Haudenaushaunee cultures, as well as many people of the two larger nations of USA and Canada, live here with a total population of over 33 million people. Spanning some 1200 km from west to east, the chain of lakes, with connecting and outlet rivers, is fed by many short rivers, wetlands and groundwater aquifers.

Starting in the 19th century, the southerly part of the Great Lakes Region gradually became strongly industrialized with steel and petrochemical works and related manufacturing as well as intensive agriculture and related processing. The sparsely-settled northerly hinterland of the region provided natural resources for the southerly industrial heartland. Fresh water for manufacturing and domestic use was inexpensive and in such abundance that cheap hydroelectric energy from its ows and ship transport on its surface contributed greatly to the modern development.

Interjurisdictional difficulties with sharing of beneficial uses (e.g., fisheries, water) and with limiting of harmful abuses (e.g., dumping human wastes with their pathogens into boundary waters) became region-wide issues late in the 19th century. Formal intergovernmental arrangements to resolve those and numerous other difficulties gradually came into place, issue by issue, during the 20th century. In 1987, Canada and the USA entered into a formal commitment to address the many interactive issues in an ecosystemic context, without intending to defocus efforts to correct priority problems like contaminants. In the following years, many governance institutions at all levels from local to basin-wide then came to accept this ecosystem approach. Implementing it in practice progressed slowly.

In the 1980s it was apparent to opinion leaders here as elsewhere, that some worrisome issues were regional manifestations of continental or global processes (e.g., acid rain, organohalide contaminants, introduction of harmful exotic species). Then some global issues that had not yet attracted much notice in the region came to be recognized here, too (e.g., climate change, loss of biodiversity, spread of harmful exotic species, commodification of water within free trade, stratospheric ozone depletion, urban sprawl in the coastal zone with in ux of many humans).

GENERAL FEATURES OF RESPONSES TO GLOBAL CHANGE

Until the mid-1980s, the basic approach to resolving region-wide problems in the Great Lakes involved primary leadership and funding by the two federal governments, in formal cost-sharing partnerships with the states and province. In effect, these governments had accepted the primary onus to correct matters, and mostly at their expense. The countries created a number of binational commissions to help with this work. Within jurisdictions, there evolved numerous kinds of regulations, with each set delegated to matching agencies in different jurisdictions and with transboundary connections among agencies. Higher levels of government helped to fund clean-up efforts of several kinds. Meanwhile, some of the worst industrial polluters abandoned their obsolescent plants to become derelict brownfields or gradually rebuilt them with better technology. By the mid-1980s this command and control approach was recognized to be less than optimal politically, economically and environmentally. This realization emerged at about the time that a series of global issues were recognized to have important regional and local aspects.

Pessimistic realists perceive events since about 1985 as a retreat by the two higher levels of government from environmental responsibilities. Optimistic idealists perceive that redemocratization is occurring with governance at a more appropriate level. The latter involves, in turn, a shift of primary onus to that level, under appropriate formal guidance by senior levels of governance. Overall, this process did not proceed smoothly.

With environmental uses and abuses generally, it has come to be accepted, in theory, that the person, corporation or agency intervening in the natural environment must bear a strong onus to act appropriately and to provide evidence of proper behavior. Costs of environmental stewardship have increasingly been internalized within the interveners' financial accounts. Industry associations have formulated codes of ethics for participating corporations, and these have been enforced occasionally. Professions have updated their codes a little. Governments have been considerate of failings where some evidence of good faith has been offered. These reforms, proceeding haphazardly, had not gotten far in the Great Lakes Region by the end of the 20th century, and the prognosis was uncertain. Again, pessimists feared a relapse into severe further degradation, while optimists hoped for continued gradual rehabilitation of the harm done by past abuses as well as successful prevention or counter-action of new abuses.

A COMBINED TECHNICAL AND POLITICAL INITIATIVE ON CONTAMINANTS

The 1978 binational Great Lakes Water Quality Agreement had as an objective zero discharge and virtual elimination

of hazardous contaminants such as certain organohalides. Large funds were committed toward this end by different levels of government and corporations. Hundreds of citizens participated in creating and implementing *Remedial Action Plans* concerning contamination and other problems in 43 severely degraded *Areas of Concern* around the shores of these lakes and their connecting rivers. Areas of Concern included Green Bay (WI), Saginaw Bay (MI), the Cuyahoga River (OH), Buffalo Harbor (NY State) and Burlington and Toronto Bays in Ontario. Many other degraded locales up the tributary streams and along shores were not addressed specifically as part of the binational program.

By the turn of the century, major remediative progress was documented for most of the Areas of Concern and the task of remediation was judged to have been completed in one of the 43 locales; the delisted locale is Collingwood Harbor in Ontario. Monitoring of various organisms that accumulate contaminants provided ambiguous data on the rates of decline and the eventual endpoints attainable. Some researchers inferred that levels of contamination were declining at encouraging rates for most hazardous chemicals, due in part to burial in sediments and to natural degassing of these fugitive substances from the waters to the atmosphere (to be carried mostly northward and eastward to cause harm there). Optimists extrapolated the rates to forecast that most of these chemicals would eventually fall to non-detectable levels later in the 21st century. Assumptions for optimistic forecasts included: that no large deposit of contaminants on the shores of these waters would break out of its burial site, that no major industrial accidents involving spills would occur in the region or upwind of it, and that atmospheric loadings from outside the region would also fall steadily. Pessimistic researchers saw no firm basis in the available data and information for such optimism.

With partial success on the contaminant and other chemical issues, concern about them abated in the region, at least among politicians and civil servants. Experts concerned about newly perceived risks of endocrine disruption urged increased efforts toward zero discharge and virtual elimination. In both countries, the level of governance at which this issue was being addressed appeared to have shifted from the federal level: in part to the maker and user of such a substance interacting with the local community; and in part to the international level with respect to guidelines on the trade in these materials. At both the local and global levels the governance institutions were evolving slowly.

PRECAUTIONARY APPROACH TO DIVERTING AND EXPORTING FRESH WATER

The quantity and quality of Great Lakes waters have been a recurring concern at all levels of governance since the 19th century. Waters naturally flowing out of this chain of lakes northeastward into the Gulf of St Lawrence and the Atlantic Ocean were perceived by some to be wasted. Ecologists have found that these outflows into the marine system have major beneficial effects there; thus there is no waste.

The city of Chicago has diverted water from Lake Michigan to flush effluents from poorly-treated sewage into the Mississippi River system. Large agricultural and urban users of water to the southwest of the Great Lakes, with costs often subsidized by governments, have long sought further governmental subsidies to help negotiate and divert water to them from the Great Lakes.

Regional assessments of the effects of the various scientific scenarios of global climate change on the levels and flows of water in the Great Lakes are not consistent. The precautionary principle requires an emphasis on preventing strongly adverse possibilities (*see* Precautionary Principle, Volume 4; Precautionary Principle, Volume 5). Partly for this reason, and partly because fresh water in bulk quantities may become a commodity under free trade, the various levels of government in the region turned to a binational commission for advice on the issue. The policy recommendation invokes a precautionary approach and urges against further diversions and exports of water in bulk quantities. Protection of the integrity of regional wetland ecosystems, with the integrity strongly dependent on a natural hydrologic regime with marked fluctuations, is a key consideration.

Aboriginal peoples are particularly concerned about the issues of water removal from this region. Their cultures are closely linked to aquatic realities of regional ecosystems and they may perceive their treaty rights to be violated by diversions and bulk withdrawals of water. This issue may provide a stimulus for the Aboriginal peoples to organize as federations to participate effectively in regional and global issues.

A NO-REGRETS TECHNICAL INITIATIVE IN TORONTO

In the older, larger cities of the region, the urban infrastructure had aged late in the 20th century. Water supply, sewage collection and treatment, heating and cooling of buildings all were old and inefficient, and becoming more so. Cooling devices still used halomethane gases some of which escaped into the atmosphere to cause stratospheric ozone depletion. Heating was done in part with steam and electricity from the combustion of fossil fuels with the production of carbon dioxide, a greenhouse gas. In the 1990s the city of Toronto, motivated in part by concerns about climate change and stratospheric ozone depletion, embarked on a phased process to address its local issues in a coordinated way.

Toronto is located on the shore of Lake Ontario, which is deep with a large mass of cold water year round at depths below about 25 m. Toronto withdraws domestic water from this deep cold water. With domestic water supplies, there is no particular advantage to cold temperatures as such. So incoming cold water could be used as a heat sink in a heat exchanger; the ice slurry produced in the exchanger could then be piped to buildings for cooling purposes. (Some large buildings in the city need to be cooled even in the dead of winter, because of the buildings' insulation against summer temperatures and the heat given off by electronic equipment and respiring humans.) The warmer domestic water might lead to warmer sewage and more rapid decomposition of sewage wastes.

The ecological effects, on land and in water, of such deep lake-water cooling are expected to be complex. No disastrous ecological effects have been forecast. Among the desirable effects are reductions in the release of gases that cause stratospheric ozone depletion and climate warming.

EFFORTS RELATED TO CLIMATE CHANGE

In the 1970s some meteorologists in the Great Lakes Region feared that global cooling was underway. There had been controversy between the protagonists of cooling and of warming in that decade. By the mid-1980s however, the concern about warming came to be taken more seriously by increasing numbers of researchers; and numerous scientific activities have been mobilized. Most assessments of ecological and other effects took one or more of the global climate simulations as plausible scenarios, and then sought to infer likely climatological and other consequences for this region. At the end of the 20th century, there remained important inconsistencies and uncertainties among the regional interpolations of different global simulations, and these contributed to political confusion concerning the urgency of practical actions aimed at reversing climate warming. The general approach taken was to assign somewhat higher priority to existing commitments on extant environmental abuses that were also likely to contribute to climate change, i.e., a no-regrets approach (*see* No Regrets Principle, Volume 4). Some of these initiatives have been sketched above.

Versions of an ecosystem approach are coming to be standard with environment-related planning and decision-making in the region. Global issues, including climate change and stratospheric ozone depletion, may be included explicitly in the larger biospheric context for a particular ecosystem within the region. The various principles of the United Nations' Agenda 21 (precautionary principle, sustainable development, equitable sharing, sensitivity to aboriginal peoples) may also be taken into consideration. Admittedly, this is an optimistic interpretation.

See also: Water Resources: Great Lakes Case Study, Volume 3; Transboundary Water Resource Management: The Example of the Canada/US IJC (International Joint Commission), Volume 4.

Lake Victoria

G O Obasi
World Meteorological Organization, Geneva, Switzerland

Lake Victoria, the second largest freshwater lake in the world, has unique hydrological features. It straddles the Equator over Kenya, Uganda and Tanzania, and the catchment area extends to Rwanda and Burundi. The Lake has a surface area of about 67 580 km^2. The mean and maximum depths are about 40 m and 92 m, respectively. Lake Victoria is one of the principal sources of the river Nile which ows out of the Lake north of Jinja (Uganda), for about 6650 km before entering the Mediterranean Sea at Alexandria in Egypt. Its drainage basin encompasses about 3 349 000 km^2 (IDAEL, 1996; Lehman, 1998; UNDP and WMO, 1974). The Nile is the lifeline for the downstream semi-arid countries of the Sudan and Egypt, and has been so ever since the time of the Pharaohs when it sustained the great ancient civilizations.

Millions of people depend on Lake Victoria for their food, water, employment, transport, energy and other socio-economic activities. Some of the challenges to the sustainable utilization of Lake Victoria resources include the rapid growth of population and urbanization in the Lake basin; pollution by urban sewage, industries, and agricultural activities; signi cant competition and potential con icts among various sectoral water-use activities at national and regional levels; and climate variability including droughts and oods (Obasi, 1997a; Obasi, 1996).

HYDROLOGY AND CLIMATE OF LAKE VICTORIA BASIN

The annual rainfall over the Lake area ranges from 800 to over 2000 mm in the Lake interior and accounts for about 80% of the total Lake water input of 15–18 milliards (cubic km) per annum. One-third of the remainder comes from the river Kagera with a drainage basin area of 60 000 km^2 and two-thirds from many streams and tributaries. The mean annual out ow at the only outlet (river Nile) measured over 80 years is about 12.2 milliards. The highest rise in the Lake level of 1.25 m was recorded in 1961, while the lowest level

was recorded between May 1922 and April 1923 (Flohn and Burkhardt, 1985; Hastenrath and Kutzbach, 1983; Indeje *et al.*, 1998; UNDP and WMO, 1974). There are, however, no reliable estimates of sediment loads into and out of the Lake.

The climates of the Lake basin region range from equatorial to arid types. There is, however, plentiful moisture near the Lake due to swamps and other large lakes. The Lake is also known to generate its own climate (Okeyo, 1987; Ogallo, 1988; Sun *et al.*, 1999). For example, a semi-permanent low pressure system that is often referred to locally as the Lake Victoria trough migrates into and out of the Lake during the night and day time respectively, and is associated with significant diurnal rainfall variability.

There are strong seasonal and inter-annual variabilities in the rainfall patterns which are often re ected in stream ow and lake-level uctuations. Recurrences of oods and droughts are quite common in the Lake Victoria basin, and often result in loss of life and property, mass displacement of the Lake inhabitants, destruction of crops, erosion, siltation, and the increase in water-borne diseases. Some of the seasonal variations and extreme weather and climate anomalies in the Lake Victoria basin have been associated with the El Nino/Southern Oscillation (ENSO) phenomenon (Indeje *et al.*, 1998; Mukabana and Pielke, 1996; Okeyo, 1987; Ogallo, 1988; Sun *et al.*, 1999).

Paleoclimatological studies indicate that climate variability and change have been the major cause of the past deterioration of Lake Victoria (IDAEL, 1996).

Climate models indicate a rise in the mean global temperatures of $1-3.5\,^{\circ}C$ by the end of the 21st century (IPCC, 1995). Such warming is expected to give rise to thermal expansion and changes in the level of Lake Victoria. However, the current climate models are unable to provide consistent and skilful regional climate scenarios that are required for projecting the impacts of climate change on the hydrological cycle at the river basin level. This is a major challenge for climate modelers and hydrologists (Obasi, 1997b). Nevertheless, policy decisions can be derived from the current models using a hypothetical range of climate change scenarios.

WATER QUALITY AND SOME ENVIRONMENTAL PROBLEMS

Lake Victoria is experiencing rapid environmental damage. The water hyacinth is the most visible (although perhaps not the most damaging) of all the problems. The plant's ability to absorb pathogens from water is a much-exploited trait utilized in water treatment facilities around the world. However, over Lake Victoria it forms thick mats which clog dams, choke and deoxygenate the waters, destroy the shores and render inaccessible the traditional fishing areas and landing beaches.

Other environmental problems relate to over-fishing, and to the introduction of some health-hazard fishing methods such as fish poisoning. The large Nile perch, which is a relative newcomer, has caused havoc to the ecosystem, decimating certain native fishes and invertebrates. Some species have completely disappeared. These developments raise concern about the sustainability of fish resources. Deforestation around the lakeshores and beyond is another major problem. Improper cultivation of deforested land areas has accelerated the runoff of top soil into the Lake. Industrial pollution is also on the increase. There are no effective waste management programs in place. Thus, the ecological health of Lake Victoria has been significantly affected by the dumping of untreated ef uents by several industries. For example, phosphorous concentrations have risen by a factor of two (from about $0.05\,mg\,l^{-1}$ in the 1960s), algae bloom have increased by a factor of 8–10, the deep waters of the lake are more depleted in oxygen for longer periods of time each year, and filamentous and colonial blue-green algae now dominate the algae community. While the ecosystem changes in the Lake are well documented, the causes of the changes are uncertain due to data limitations.

SOME RELEVANT PROGRAMS OF THE WORLD METEOROLOGICAL ORGANIZATION (WMO) AND OTHER ORGANIZATIONS

Various programs in support of member countries and their regional development plans of Lake Victoria Basin have been implemented, including a WMO/UNDP hydrometeorological survey of the catchments of Lakes Victoria, Kyoga and Mobutu Sese Seko in 1967. Under the project, 25 meteorological stations, 200 rainfall stations and 45 stream-gauging stations were installed.

Several initiatives have also been undertaken by the individual countries, and other UN agencies and bodies including the Food and Agriculture Organization, the United Nations Educational, Scientific and Cultural Organization (UNESCO) and the United Nations Environment Programme. The Lake Victoria Environmental Management Project is one such project undertaken by the Kenya, Uganda and Tanzania through the Global Environment Facility and the International Development Agency funding.

Global on-going programs relevant to the water resources of Lake Victoria basin (Ogallo, 1988; Sun *et al.*, 1999) include:

- Enhancement of real-time and near real-time monitoring, and improvement of the database at national, regional and global levels, for weather and hydrological forecasting and climate prediction. The World

Hydrological Cycle Observing System in particular, will assist in hydrological monitoring and forecasting (WMO, 1998).

- The Hydrological Operational Multipurpose System for the transfer of technology in operational hydrology (WMO, 1998).
- Some research programs implemented jointly by WMO and other international bodies such as the Intergovernmental Oceanographic Commission of UNESCO and the International Council for Science are meant to improve climate prediction and the understanding of the impacts of climate anomalies on water resources. These include the Global Energy and Water Cycle Experiment, and Climate Variability and Predictability study.

SUSTAINABILITY OF LAKE VICTORIA

Sustainable utilization and management for the Lake Victoria water resources in this century will require, among others:

1. An improved observation network including remote sensing capability for monitoring all components of the hydrological cycle and water quality within the basin.
2. An improved capacity for efficient and timely exchange of meteorological and hydrological data and information, and services.
3. An improved database for the whole catchment, including past, present and projected sectoral water use activities. This can enable the planning of integrated and sustainable water policies to be developed in order to minimize con icts over water use.
4. Global warming could result in thermal expansion and lake level rise. It could also result in enhanced evapotranspiration over the lake catchment. Enhanced interdisciplinary research is therefore required in order to properly understand the processes that have been responsible for the past and present hydrometeorological evolution in the Lake Victoria catchment, including the in uence of the ever-increasing human effects on the basin hydrological characteristics.
5. Enhancement of capacity for the development of climate/hydrological models for timely early warning and disaster preparedness.
6. Assessment of the potential impacts of climate change on lake basin hydrology, and the implications of such changes on current and projected future catchment water use activities.
7. Enhanced capacity building at national and regional levels. Most of the countries in the region still lack skilled interdisciplinary human resources, equipment and other basic facilities that are required for optimum application of current scientific and technological knowledge in operational hydrology.
8. The need to promote an appropriate mechanism for closer coordination of water use activities and the associated policies at local, regional, and international levels, including the harvesting of the natural resources of the Lake, and a reduction in the release of wastes into the Lake and rivers.

REFERENCES

Flohn, H and Burkhardt, T (1985) Nile Runoff at Aswan and Lake Victoria. A Case Study of a Discontinuous Climate Time Series, *Z. Gletscherkunde Glazialgeol.*, 21, 25–130.

Hastenrath, S and Kutzbach, J E (1983) Paleoclimatic Estimates from Water and Energy Budgets of East African Lakes, *Q. Res.*, 19, 141–153.

IDAEL (1996) *Limnology, Climatology and Paleoclimatology of the East African Lakes*, Gordon Breach Publication, Amsterdam.

Indeje, M, Semazzi, F H M, and Ogallo, L J (1998) ENSO Signals in East African Rainfall and their Prediction Potentials, *Int. J. Climatol.*, (in press).

IPCC (1995) *Climate Change 1995, The Science of Climate Change*, Summary for Policy Makers, Working Group I Report, WMO/UNEP Intergovernmental Panel on Climate Change, 1–56.

Lehman, J T (1998) *Environmental Change and Response in East African Lakes*, Kluwer Academic Publishers, Amsterdam.

Mukabana, J R and Pielke, R A (1996) Investigating the In uence of Synoptic-scale Monsoonal Winds and Mesoscale Circulations and Diurnal Weather Patterns over Kenya using a Mesoscale Numerical Model, *Mon. Weather. Rev.*, 124, 224–243.

Obasi, G O P (1996) Climate, Climate Change, Variability and Predictability, RGICS Paper No. 36, Rajiv Gandhi Institute for Contemporary Studies, New Delhi, 1–35.

Obasi, G O P (1997a) *Climate Change and Freshwater Management. Proceedings of the First World Water Forum*, Marrakesh, Morocco, World Water Council, Marseilles.

Obasi, G O P (1997b) Implications of the IPCC Projections on 21st Century Climate – a View from WMO, Eighth Symposium on Global Climate Change Studies, California, preprints American Meteorological Society, MA, 1-8.11.

Ogallo, L J (1988) Relationships Between Seasonal Rainfall in East Africa and the Southern Oscillation, *J. Climatol.*, 8, 31–43.

Okeyo, A E (1987) The in uence of Lake Victoria on the Convective Activities over the Kenya Highlands, *J. Meteorl. Soc. Japan*, 689–695.

Sun, L, Semazzi, F H M, Giorgi, F, and Ogallo, L J (1999) Application of the NCAR Regional Climate Model to Eastern Africa. Part I: Simulations of Autumn Rains of 1988, *J. Geophys. Res.*, 104(D6), 6529–6548.

UNDP and WMO (1974) *Hydrometeorological Survey of the Catchments of Lakes Victoria, Kyoga and Albert*, UNDP/WMO, 769, 1–48.

WMO (1998) WHYCOS, World Hydrological Cycle Observing System, *WMO*, 876, 1–12.

Nile River

Mohamed Kassas
Cairo University, Giza, Egypt

The Nile Basin is shared by 10 countries, the river spans 35 degrees of latitude with a south to north alignment from the Equator to the Mediterranean. The catchment area is large but the discharge is not: the Sudd swamps of southern Sudan consume most of its equatorial waters, and its northern reaches traverse arid territories that contribute very little water. The geological and cultural histories are complex, the climatic systems that prevail are very different, and climate futures need to be studied within a collaborative program shared by the 10 countries of the basin.

The course of the River Nile runs in a south to north alignment for 35 degrees of latitude: from Lake Tanganyika (Lat. 4 °S) to the Mediterranean (Lat. 31 °N). Its basin is shared by 10 countries: The Congo Democratic Republic, Burundi, Rwanda, Tanzania, Uganda, Kenya, Eritrea, Ethiopia, Sudan and Egypt, and spans a wide spectrum of climates (from humid tropics to deserts), vegetation types, landforms, land uses and human cultures. This basin is one of two biogeographical corridors that link the equatorial and tropical realms in the south with the palaearctic (Mediterranean) realms in the north. The other is the Red Sea. For details see Rzòska (1976). This river has a long and complicated history that relates tectonic–geological events, climate changes and related rise and fall in the Mediterranean water level. Its rudiments go back some six million years, but its present form is recent. Said (1994) recognizes three terminal phases: the beta Neo-Nile (26 000 before present (BP), the gamma Neo-Nile (20 000–12 000 BP), and the delta Neo-Nile (10 000 BP). For details of this history, see Said (1993).

One can conceive the extant Nile as a recent connection of three principal systems: the system that drained waters of the equatorial plateau and formed an inland delta that remained juvenile (the marshlands of the Sudd region of southern Sudan), the system that drained the Ethiopian highlands (rivers Sobat, Blue Nile, Atbara, etc.) and poured onto what was an extensive lake (now the Gezira and Butana alluvial plains of eastern Sudan), and the northern (Egyptian) system that collected waters from several tributaries (now desert wadis) emanating from the Red Sea coastal mountains. The present river collects waters from two sources: the equatorial plateau and the lower plains of southern Sudan and the Ethiopian highlands. The former receives year-round annual rainfall of more that 1000 km^3, but the volume of water it contributes to the northward owing White Nile is about 15 km^3. The lower plains embrace the 700 km long swamp (The Sudd), which seems to consume considerable volumes of water; it receives little water from a western tributary (Bahr-el-Arab) with extensive catchments that comprises parts of southern Sudan, Chad and the Central Republic of Africa. In general, this southern section seems wasteful and invites engineering schemes for improving its efficiency. The Jonglei Canal scheme is one example (a canal that bypasses the Sudd and carries part of the equatorial waters directly to the White Nile). A collaborative program of hydrometeorological studies on the equatorial lakes was initiated in 1967 with the participation of Tanzania, Kenya, Uganda and Egypt and support of the World Meteorological Organization; between 1971 and 1974 Ethiopia, Rwanda, Burundi and Zaire joined the program. This program built a considerable database on the hydrology of these equatorial sources of the Nile.

The Ethiopian sources seem more efficient, only some water is lost in the Machar marshes of the Sobat. The 84 km^3 per year of water reaching Aswan (Egypt) are mainly (83%) Ethiopian waters (13% from Sobat, 58% from Blue Nile and 12% from Atbara); only 17% come from the equatorial sources via the White Nile. Though the Nile is one of the longest rivers of the world, its yearly discharge to its mouth is modest compared with other African rivers (Table 1).

Two features seem to explain this relatively small ow of water; firstly, the considerable water losses of the equatorial sources and secondly its downstream part, from the mouth of Atbara in the Sudan to the Mediterranean (c. 2700 km), crosses an extremely arid region (The Sahara) and contributes almost nothing to its waters. The countries that share this international river, especially the two downstream countries (Sudan and Egypt), built several reservoirs on the Atbara (Khashm-el-Girba), the Blue Nile (Sennar and Roseires), the White Nile (Jebel-Awlia) and the Aswan Reservoir and the Aswan High Dam (Egypt). There is still room for schemes that conserve and tap the considerable volumes of water of the equatorial Nile. The hydropower resources are only partly (c. 10%) developed. There is considerable potential for development in the Equatorial and Ethiopian tributaries and in the principal course of the Nile.

Table 1 Comparison of yearly discharge of the Nile with other African rivers

	Length (km)	Basin (km^2)	Discharge (km^3 $year^{-1}$)
Nile	6700	2 960 000	84[a]
Congo	4700	3 830 000	1248
Niger	4100	1 220 000	192
Zambezi	2700	1 200 000	223

[a] Inter-annual variability is considerable: 120 km^3 in 1916 and 42 km^3 in 1984.

The human history associated with the River Nile provided part of the world cultural heritage: the Pharaonic civilization that ourished in its downstream part (Egypt), and the stories of adventurous travels of the explorers searching for its Ethiopian and equatorial origins. Though the seasonal rise and fall of the River was generally regular (of the 820 oods recorded on the Roda Nilometer between the 7th and 15th centuries, 73% were normal oods, 22% were low and 5% were destructively high (Said, 1993)), the origin of the water remained a mystery. During the middle and new Kingdoms of Ancient Egypt, the empire was extended southward to Nubia and at one time up to Khartoum. Trading missions, via the Red Sea, established posts in the land of Punt (the present East African Horn including Somalia and Ethiopia). Christianity entered Ethiopia (330 AD) by missionaries from Alexandria. All through ancient and medieval history, Egypt believed that oodwater came from highlands of Ethiopia; in years of very low ow, emissaries were sent to ask the king of Ethiopia to let the ood ow to Egypt.

The narratives and reports of expeditions sent to explore the equatorial sources of the Nile by the then rulers of Egypt (Mohammed Ali and his sons) between 1821 and 1840, by the Royal Geographical Society (Great Britain) and the Khedival Geographical Society (Egypt) between 1850 and 1870, and subsequently by British and German governments in East Africa, are contained in a rich series of volumes. The most southerly source of the Nile was discovered in 1937: the two rivers Luvironzo and Nlavarango join and form the Ruvuvu tributary of the River Kagera, which ows into Lake Victoria. The discovery was commemorated by the words Caput Nili inscribed on a small hill to the north of Rutana village (Lat. 4 °S) in Northern Burundi. See Hurst (1952) for a summary and Moorehead (1960) for the story. The story of the exploration of the Ethiopian sources was told by Moorehead (1962). Two names are associated with the 1770s: the French geographer D'Anville who produced a map of Ethiopia and the Blue Nile based on reports of Portuguese missionaries, and the British explorer James Bruce.

Five principal climate systems prevail within the Nile Basin: the equatorial climate of the lakes plateau (rainfall c. 1200 mm year^{-1} with weak bimodal seasonal distribution); the Ethiopian monsoonal climate (summer rainfall c. 750 mm year^{-1}, mainly in June–August); the Central Sudan semi-arid climate (summer rainfall c. 399 mm year^{-1}); the Saharan climate in Northern Sudan and almost all of Egypt; the Mediterranean coastal arid climate with winter rainfall (c. 150 mm year^{-1}). The likely manifestations of global climate change are uncertain. The IPCC (1998) report states that The global change model (GCM) scenarios provide widely diverging pictures of possible future river ows, from 30% increase to 78% decrease •••. There is need for a regional climate modeling effort over the Nile to help reduce this uncertainty . The 10 countries that share the river basin are at present engaged in formulating a Nile Basin Initiative of regional studies towards sustainable management of its resources. A climate research program could be a valuable element of this endeavor.

See also: Lake Victoria, Volume 4.

REFERENCES

Hurst, H E (1952) *The Nile*, Constable, London.

IPCC (1998) *The Regional Impacts of Climate Change: an Assessment of Vulnerability*, Cambridge University Press, Cambridge.

Moorehead, A (1960) *The White Nile*, Harper and Row, London.

Moorehead, A (1962) *The Blue Nile*, Harper and Row, London.

Rzòska, J, ed (1976) *The Nile: Biology of an Ancient River*, W Junk, The Hague.

Said (1993) *The River Nile*, Pergamon Press, London.

Said (1994) Origin and Evolution of the Nile, in *The Nile, Sharing a Scarce Resource*, eds P P Howell and J A Allen, Cambridge University Press, Cambridge.

General

Climate Change Assessment, United States

Michael C MacCracken
Lawrence Livermore National Laboratory, Livermore, CA, USA

In 1997, the US Global Change Research Program (USGCRP) initiated the National Assessment of the Potential Consequences of Climate Variability and Change for the United States. This national assessment included analyses of the importance of climate variability and change in 20 regions around the US, in ve cross-cutting sectors focused around natural resources and public health, and for the US as a whole. Climatically, the hundred-year projections that were used indicated that the climatic conditions of the northern US would become like the central US, the conditions of the central US like those across the southern US, and conditions across the southern US like tropical and subtropical regions. The results of the analyses indicated that there will be signi cant regional variability in both climate change and resultant consequences across the US. Although overall forest productivity was projected to increase for the next several decades, as a result of the carbon dioxide fertilization effect, speci c natural ecosystems were projected to be at signi cant risk because of the very limited potential for their movement and reestablishment. Although the US has abundant water resources, widespread concerns about water resources arose in every region. As a whole, US agricultural production was projected to increase, although those farming marginal lands are likely to be impacted by further reductions in their economic competitiveness. The extensive coastal regions in the US were projected to be exposed to rising sea levels, increasing the risk to extensive coastal infrastructure, barrier islands, and coastal wetlands. Rising temperatures and the accompanying rise in the absolute humidity, were projected to lead to much more uncomfortable summer weather conditions across the southern and eastern US; analyses indicated, in general, that health consequences could be minimized by taking a wide array of public health measures.

As the issue of climate change gained attention in the 1980s as a result of scientific advances and early international conferences, the US Congress passed The Global Change Research Act of 1990. This public law recognized the early scientific findings that human activities were starting to change the global climate, asserting that:

(1) industrial, agricultural, and other human activities, coupled with an expanding world population, are contributing to processes of global change that may significantly alter the Earth habitat within a few generations;
(2) such human-induced changes, in conjunction with natural uctuations, may lead to significant global warming and thus alter world climate patterns and increase global sea levels. Over the 21st century, these consequences could adversely affect world agricultural and marine production, coastal habitability, biological diversity, human health, and global economic and social well-being.

To address these issues, the US Congress established the USGCRP and instructed US federal research agencies to cooperate in developing and coordinating

> a comprehensive and integrated US research program which will assist the nation and the world to understand, assess, predict, and respond to human-induced and natural processes of global change.

The USGCRP has led to coordination of the global-change related research efforts of 10 federal agencies, including the departments of Agriculture, Commerce (specifically the National Oceanic and Atmospheric Administration [NOAA]), Defense, Energy, Health and Human Services, and Interior along with the agency-based research programs in the Environmental Protection Agency, National Aeronautics and Space Administration, the National Science Foundation, and the Smithsonian Institution. While interagency coordination has led to a number of organizational challenges, each agency contributes its particular strengths in order to address the very broad issues raised by concerns about global change.

One section of the congressionally approved Act indicated that the USGCRP

> shall prepare and submit to the president and the congress an assessment which:
>
> 1. integrates, evaluates, and interprets the findings of the program and discusses the scientific uncertainties associated with such findings;
> 2. analyzes the effects of global change on the natural environment, agriculture, energy production and use, land and water resources, transportation, human health and welfare, human social systems, and biological diversity; and
> 3. analyzes current trends in global change, both human-induced and natural, and projects major trends for the subsequent 25 to 100 years.

Fulfilling this mandate for the issue of potential climate change has taken several forms. At the international level, the USGCRP coordinates US participation in the assessments of the Intergovernmental Panel on Climate Change (IPCC). The IPCC assessments in 1990, 1996, and 2001 have, at the global and continental scales, carefully reviewed the scientific literature concerning climate change, its potential consequences, and options for mitigation. With the IPCC finding in 1996 that human activities were having a discernible in uence on the global climate, the USGCRP undertook a national level assessment intended to provide detail about potential consequences within the US.

Given this charge, an overall goal of the USGCRP has been to analyze and evaluate what is known about the potential consequences of climate variability and change for the US in the context of other pressures on the public, the environment, and the nation's resources. At the same time, assessment activities have been designed to initiate an active dialogue with those who would be affected by the potential consequences in order to help prepare them to deal with climate variability and change more effectively and to sharpen the research focus concerning coping options. Thus was born the National Assessment of the Potential Consequences of Climate Variability and Change for the United States, commonly called simply the *National Assessment*.

The US National Assessment has had three major components:

1. Regional analyses: Regional workshops and assessments were used to characterize the potential consequences of climate variability and change in regions spanning the US. A total of 20 workshops were held around the country during 1997 and 1998. Regions holding workshops included: New England and upstate New York; Metropolitan New York; Mid-Atlantic; Central and Southern Appalachians; Southeast; South Atlantic Coast and Caribbean; Gulf Coast; Great Lakes; Eastern Midwest; Northern Great Plains; Central Great Plains; Southern Great Plains; Rocky Mountain/Great Basin; Southwest-Rio Grande basin; Southwest-Colorado River basin; California; Pacific Northwest; Alaska; Pacific islands; and Native peoples and homelands. Based on the issues identified, most of the regions went on to prepare assessment reports that would address the issues identified in the workshops; the remaining regions are likely to be the subject of subsequent studies.
2. Sectoral analyses: Workshops and assessments were also organized to characterize the potential consequences of climate variability and change for five major sectors that cut across environmental, economic, and societal interests. The sectors included agriculture, forests, human health, water, and coastal areas and marine resources. As US assessment activities continue, the expectation is that examination of additional sectors such as transportation, energy, business and trade, wildlife, urban communities, international couplings, and many other sectoral perspectives will be needed to provide a more comprehensive perspective on US vulnerability.
3. National overview: A 13-member National Assessment Synthesis Team (NAST) was appointed as a federal advisory committee to provide the national overview. The NAST co-chairs were Dr Jerry Melillo of the Marine Biological Laboratory in Woods Hole, MA, Dr Tony Janetos of the World Resources Institute in Washington, DC, and Tom Karl of NOAA's National Climatic Data Center in Asheville, NC. The NAST was given responsibility for providing a national perspective that summarized and integrated the findings of the regional and sectoral studies and that then drew conclusions about the importance of the consequences of climate change and variability for the US. Both an overview report intended for decision makers and the public and a foundation report intended for experts were prepared (NAST, 2000, 2001). Among the innovations in the NAST report was use of a well-defined lexicon for conveying the relative level of likelihood and certainty of particular consequences. Thus, consequences were described as being either very likely, likely, possible, unlikely, or very unlikely in a consistent manner, based on the expert judgment of the NAST in evaluating the available literature and commissioned studies.

There were a number of unique organizational approaches used in the conduct of the assessment. First, responsibility for the National Assessment process was widely distributed and broadly inclusive, with each of the regional, sectoral, and synthesis activities being comprised of experts from universities, government agencies, public and private sector organizations, and from other stakeholder communities. Second, the assessment process was financially supported in a shared manner by the set of USGCRP

agencies, with each agency assuming responsibilities for those activities in which they had particular interests. Third, the assessment is viewed by the agencies as the start of a long-term engagement, bringing together the scientific community and those affected by climate change. Together, these steps have been designed to promote broad understanding of the issue and its importance and to encourage early recognition of the need to start adapting to climate change.

Conduct of the assessment focused on addressing issues of importance to people in particular places (the regions) and with particular interests (the sectors). Starting with the broad array of public concerns about the environment, the Assessment explored the degree to which existing and future variations and changes in climate might affect issues that people really care about. A short list of questions guided the process:

- What are the current environmental stresses and issues that form the backdrop for potential additional impacts of climate change?
- How might climate variability and change exacerbate or ameliorate existing problems? What new problems and issues might arise?
- What are the priority research and information needs that can better prepare the public and policy makers for reaching informed decisions related to climate variability and change?
- What coping options exist that can build resilience to current environmental stresses, and also possibly lessen the impacts of climate change?

In analyzing the potential consequences, the National Assessment teams utilized three approaches to generate plausible future climatic conditions. Historic data sets were assembled and used to evaluate the potential consequences of a reoccurrence of past variations in the climate, recognizing that there would also be an underlying warming trend. Computer-generated scenarios from the Canadian and Hadley (UK) climate modeling centers were used to provide a self-consistent set of plausible future conditions. These model scenarios spanned the range from warm/moist to hot/dry conditions that are typical of the results from wider set of climate models. Should these conditions occur, the US would experience about a $5-10\,^{\circ}F$ (about $3-6\,^{\circ}C$) warming and modest to strong precipitation increases across the US. Such changes would be equivalent to shifting warm climatic zones northward, such that, over the 21st century, the future climate of the central US would become like the current climate of the southern US. The third approach, only used in a few instances in this initial assessment phase, was to consider the amount of change that would significantly increase existing vulnerability by, for example, crossing a threshold or triggering a non-linearity. In addition to the various approaches to treating prospective climate variability and change, ecosystem models were used to generate prospective changes in vegetation cover, and socio-economic scenarios were generated to consider possible changes in population and economic activity.

The various regional and sectoral studies provided a diverse picture of potential consequences, with a mixture of beneficial and detrimental consequences. While potential adaptation measures were identified that could moderate adverse consequences for many societal activities, disruptions were found that were likely to affect the complex distribution of ecosystems across the US, with some types of rare ecosystems (e.g., alpine meadows, coral reefs) being particularly impacted because there was little potential for natural or assisted adaptation. Due to the likelihood of intensified evaporation and changes in the amounts and duration of mountain snowpack, water resources arose as an issue in regions across the US, from the dry regions of the southwest to the more water-rich regions of the Pacific Northwest. Timber and agricultural production were likely to increase due to carbon dioxide (CO_2) fertilization, but, should warming occur near the upper bounds of estimates, increased water stress and potential disruptive threats (e.g., fires for forests, pests for agriculture) could create significant vulnerability. Coastal and permafrost regions were found to be likely to suffer increasing stress, creating problems for buildings and other infrastructure. While the warmer, moister conditions would raise the heat index, reducing the quality of life, the health threat from such conditions is currently being reduced through increased reliance on air-conditioning. More generally, sustaining the health of the US population was found to require intensified attention to public health practices and community infrastructure and policies (e.g., building codes).

Quite clearly, many important questions and issues remain to be addressed. Given the initial nature and limited scope of the assessment, it was also recognized that uncertainties, including possible surprises, are also likely. For this reason, it is expected that assessment activities will become an ongoing part of the USGCRP.

Extensive information about the assessment, participants on the various assessment teams and groups, and links to the activities of the various regions and sectors are available over the Web at http://www.nacc.usgcrp.gov or by inquiry to the Global Change Research Information Office, PO Box 1000, 61 Route 9W, Palisades, NY 10964.

REFERENCES

National Assessment Synthesis Team (NAST) (2000) *Climate Change Impacts on the United States, Overview*, US Global Change Research Program, Cambridge University Press, Cambridge, UK, 154.

National Assessment Synthesis Team (NAST) (2001) *Climate Change Impacts on the United States, Foundation*, US Global Change Research Program, Cambridge University Press, Cambridge, UK, 612.

Dobříš European Environment Assessment Process

David Stanners
European Environment Agency, Copenhagen, Denmark

Dobr´s Castle near Prague in the Czech Republic was the location of the rst pan-European Environment Ministers conference held in June 1991. This conference, held under the auspices of the United Nations Economic Commission for Europe (UNECE), launched a new European cooperation in the eld of the environment and a pan-European assessment process to analyze for the rst time in the same format Europe's environment and its relationship and contribution to regional and global change. The assessment process has resulted in two reports to date: Europe's Environment: The Dobříš Assessment *(EEA, 1995, 1998). A third report had been requested for the 5th pan-European conference of Environment Ministers to be held in Kiev in 2003.*

Concerns about environmental changes in the countries of central and Eastern Europe were a common and sometimes a dominant theme in the struggles that led to the revolutionary changes that took place in Europe in the early 1990s. This, combined with a desire for greater European integration, was the catalyst for the Dobříš initiative which was proposed in 1990 by the then new Environment Minister of the Federal Government of Czechoslovakia, Josef Vavroušek (*see* Vavrousek, Josef, Volume 4) The initiative gave birth to the so-called Environment for Europe process which, following the Dobříš Castle conference, was followed by Ministerial Conferences in Lucerne (April, 1993), Sofia (October, 1995), Arhus (June, 1998) and Kiev (2003). In addition to the assessment process, the initiative has given birth to: an environmental action plan for Central and Eastern Europe, coordinated by the Organization for Economic Cooperation and Development (OECD); a pan-European biological and landscape diversity strategy, jointly coordinated by UNEP (United Nations Environment Programme) and the Council of Europe; and many diverse initiatives tackling different aspects of European environmental problems and global environmental change.

The environmental assessment process initiated at the Dobříš Castle Ministerial conference completely changed environmental reporting at the European level. For the first time Ministers directly requested an assessment to serve their purposes: to support the development of a comprehensive environmental program for Europe; to provide a sound basis for effective measures, strategies and policies; and to inform the public and raise awareness about environmental problems. The initiative dovetailed with and complemented developments in the European Union (EU) with the proposed establishment in 1990 of the European Environment Agency (EEA) (*see* EEA (European Environment Agency), Volume 4), a body designed to supply sound environmental information to support policy development in the EU to respond to and to help avert environmental changes.

Although environmental assessment at the EU level had EU environmental policy as a reference, no single common policy frame yet exists at the pan-European level against which the Dobříš European environment assessment process can refer. Thus, the aim of the first report was to set a comprehensive baseline and reference against which future progress, or deterioration, could be measured. Except for the global assessments of UNEP (see example, UNEP, 1992), hitherto, no comprehensive environmental assessment on such a regional scale had been attempted. This was partly due to the lack of a common policy framework, but it was also due to the lack of a common information base of comparable and reliable data on which to base the assessment. While there is much data available in the different countries, the coverage and scope vary enormously and there are many information gaps. Since an assessment was required and requested by the policy makers, a novel exercise began: assessing in parallel the environmental problems and the information required to make such an assessment, each helping to improve the other. This new dynamic approach to environmental assessment and reporting now underlies the technique followed by the EEA to produce the best available policy-relevant assessments for EU policy-making to address regional and global environmental change.

The European Commission accepted the task to produce the first report in cooperation with the UNECE, UNEP, OECD, the Council of Europe, the International Union for the Conservation of Nature (IUCN), the World Health Organisation (WHO) together with individual European countries. Produced by a task force set up by the European Commission to prepare the EEA, the report was published by the Agency itself in September 1995 and presented to the third environment Ministers conference held in Sofia, Bulgaria in October 1995.

Given the broad remit and lack of specific policy framework, the approach taken in this report was to assess the environment from four different angles: the environmental media; the pressures; human activities; and the environmental problems themselves. This approach provided a rich, interweaving set of perspectives with which to view the environment, and established the basis for many future assessments performed at different levels. It also set the way for the development of the DPSIR model of assessment now adopted by the EEA in its reporting on regional and global environmental change (D, driving forces; P, pressure; S, state; I, impacts; R, response) (*see* DPSIR (Driving Forces Pressures State Impacts Responses), Volume 4).

The second pan-European assessment was presented to the 4th Environment for Europe Ministerial conference held in Arhus, Denmark in June 1998. This conference requested the EEA to continue to measure progress with environmental problems in Europe and to report to the 5th conference to be held in Kiev in 2003 with a report based on indicators.

REFERENCES

EEA (1995) *Europe's Environment: The Dobr´s Assessment*, eds D A Stanners and P Bourdeau, Office for Official Publications of the European Communities, Luxembourg.

Europe's Environment: The Second Assessment (EEA) (1998) *Europe's Environment: The Second Assessment*, Office for Official Publications of the European Communities, Luxembourg.

UNEP (1992) The World Environment 1972–1992, eds M K Tolba, O A El–Kholy, E El–Hinnawi, M W Holdgate, D F McMichael, and R E Munn, United Nations Environment Programme, Chapman and Hall, London.

Eco-engineering to Promote Ecological Sustainability in China in an Era of Global Environmental Change

Rusong Wang
Chinese Academy of Sciences, Beijing, China

Rapid social, economic and environmental changes have taken place in China since its opening to the world and its transition from a planned to a market economy. Almost all kinds of environmental pollution found in early-industrialized countries, and ecological deterioration in poor developing countries, have emerged in China, exerting heavy impacts on people's health and local and regional life-support ecosystems.

Fortunately, a positive way - ecological engineering - has been found in China for alleviating the multiple ecosystem risks. Grounded in ancient human ecological philosophy, ecological engineering in China encourages an environmentally sound, economically productive, and systematically responsible way of alternative development. It uses the human ecosystem approach through ef - cient technology, appropriate institution and self-reliant technology. Here, the key is an integration of hardware (technological innovation and integrative design), software (institutional reform and system optimisation) and mindware (behavioural inducement and capacity building) (Wang and Yan, 1998).

The main characteristics of ecological engineering are to use:

- systematically coupled technology*: horizontal and vertical combination of various frontier and traditional technologies in one eco-complex, having the function of primary, secondary and tertiary production, and natural services;*
- ecologically adaptive technology*: closing the open loop of material metabolism at the ecosystem scale, and enhancing the structural and functional exibility so as to have high adaptability to the changing environment; and*
- economically efficient technology*: encouraging deeper processing and higher ef cient use of local resources, and decentralized and cost-effective wastes treatment so as to turn negative environmental impacts into positive economic bene t.*

Four kinds of ecological engineering are described below.

RURAL ECO-ENGINEERING: ECO-FARMING, ECO-AGRICULTURE, ECO-VILLAGE AND ECOPOLIS

Based on the carrying capacity of the local ecosystem, eco-farming in China encourages efficient use of material, energy and other eco-factors in farmland through intercropping, interplanting, multi-layer planting, water conservation, organic fertilizing and integrative pest management. This traditional technology has sustained Chinese agriculture for more than 5000 years. Unfortunately, it has been disappearing recently in some industrialising areas such as the Peal River Delta and the Yangtze Delta. Eco-technologies such as mulberry grove – fishpond, nightstool manure recycling and biogas systems used to be popular in

every household before the area became prosperous in the 1980s.

Based on ecological economics, eco-agriculture in China encourages competition, symbiosis and self-reliance among farming, forestry, husbandry, fishery and rural industry. There are more than 2000 eco-agricultural experimental villages or counties in China covering 123 000 km^2 and including 23.92 million people. By 1995, due to the development of this eco-agriculture, the annual increase of crop yield was over 15%, the yield per hectare increased by 10% compared with 1980, which is 4.5 and 9.2 times higher, respectively, than that of those counties without eco-agricultural development. The *per capita* income in the experimental areas was 12% higher on average than that of neighboring regions. It was 32% higher in 105 eco-agricultural experimental sites in Hubei province than the provincial average (Wang *et al.*, 1998).

Besides natural and economic symbiosis, rural sustainability has to rely on farmers' social behavior, local governance, rural–urban relationships and large-scale markets and the environment. The eco-household, eco-village and ecopolis are just this kind of development. There are more than 1000 eco-villages, 51 national eco-agricultural counties, a few eco-prefectures and one eco-province (Hainan Island) in China characterized by integrative development with many social, economic and environmental benefits.

ECO-INDUSTRY: CORN ENGINEERING IN ZHAODONG CITY

Corn industrial engineering was initiated in Zhaodong in 1996, a county in Heilongjiang province with a yield of more than two million tons of corn annually. This approach combines agriculture, industry and community development into an integrated system. It consists of four parts: corn processing with zero emission engineering; stalk engineering for forage, fertilizer and fuel engineering; comprehensive use of corn cobs engineering; and roots and residues recycling. The production chain of maize processing is prolonged, and wastes are used for producing furfural and xylose alcohol. Maize stalks are used to produce fodder for livestock, biodegradable plastics and other industrial materials. Corn roots are treated physically and biologically to fertilize cropland. Suspended protein and carbohydrates are retrieved from the wastewater with high chemical oxygen demand, which is then used for aquaculture and irrigation. Cow and pig manure is used to produce biogas. The biogas slurry is processed for pig and fish feeds. The solid wastes are used for culturing edible fungus and earthworm breeding. The residue and detritus from the earthworm beds, together with livestock and poultry manure, are used to fertilize maize fields. The liquid waste residue from biogas fermentation is a kind of eco-pesticide after being processed. In this way, starting from a small corn our mill, 17 factories for conversion and utilization of maize have been constructed, and the total production has increased by 110 times in four years while realizing clean production. This eco-industry is spontaneously initiated by enterprises rather than the academic world, and its main gain is economic but it also has ecological benefits (Lu *et al.*, 1998).

ECO-SETTLEMENT: ECOLOGICALLY INTELLIGENT CONSTRUCTION IN JINGHUA CITY

The public toilet in most Chinese cities is dirty, ugly and stinking. But in Jinghua, a middle size city in Zhejiang province with 300 000 habitants, the toilet is a kind of green, clean and productive house with self-purification, production and maintenance, called an eco-toilet. In fact, the eco-toilet is only one kind of an eco-building being developed in this town since the 1990s. Compared with normal building, an eco-building is a kind of ecologically intelligent building characterized by:

- decentralized waste treatment and regeneration;
- water conservation and recycling;
- wall and roof greening to restore the ecological service and production function of the occupied land;
- renewable energy exploitation such as solar and bio-gas use;
- eco-material use based on local alternative resources;
- eco-healthy indoor building and decorative materials;
- well insulated structure; and
- eco-compatible landscape design to embody Chinese Feng-Shui (wind and water) and man and nature in one principles.

A demonstration area of more than 30 000 m^2 of eco-buildings has been constructed in the town since the 1980s for residential, business, public and civil service use.

All organic wastes in these buildings go to an underground bio-gas pond, the roof is covered by a garden or a solar energy collection system. The eco-roof and green wall have significant benefits of thermal insulation, air freshening, sanitation improvement, economic products and reduced load on the municipal sewage system. The waste water is cleaned and reused for gardening and partly discharged, meeting the national water discharge standard of grade 2–3. The indoor temperature of the eco-building is 3 °C higher in winter and 3 °C lower in summer than that of a normal building. Furthermore, greening of the roof has restored the ecosystem capacity of 70–128% of the built-up areas on average. It is indeed possible to recover the ecological service and production

function of the land occupied by the building (Yu and Liu, 2001).

URBAN-RURAL ECO-SERVICE: GUANGHAN PILOT ECOLOGICAL ENGINEERING FOR MUNICIPAL SOLID WASTES (MSW) RECYCLING

MSW treatment is an urgent issue faced by almost all cities in China. Around half a million tons of household garbage is currently produced in Chinese cities and towns every day, most of which is simply dumped into the suburbs, causing severe water, air and soil pollution, land degradation, and human health problems in many regions. To search for a feasible way for MSW recycling, a pilot ecological engineering has been designed and implemented in Guanghan, a small city in Sichuan province with 90 000 residents producing 100 tons of MSW day$^{\bullet}$ ^{1}. It aims at waste reduction (to encourage waste marketing, residue-free agro-goods for sale, packing control, residue reduction, classified collection, etc.), disinfection (cleaner collection, toxicants separation, bacteria sterilization, sanitary landfill, water recycling, air quality control), regeneration (paper, glass, plastics, iron, steel, organic matter, heat generation), industrialization (mechanical industry for MSW, eco-fertilizer industry from MSW, crop residues and human and animal excrements, eco-industry of fully recycling production) and systematisation (planning and design, legislation and policy, monitoring and supervision, capacity building and institutional reform).

The municipal government enacted a series of regulations and legislation in institutional reform and integrated management for MSW. The sole responsibility of government for MSW has been changed to responsibility-sharing among policy makers, private enterprise and the public, where local government is responsible only for policy making and institutional reform, private companies are given incentives to earn profits from MSW through eco-tech, citizens are responsible for voluntarily reducing, separating, recycling, monitoring and inspecting wastes, and scientists and technicians are responsible for providing ecologically sound methodologies and technology. Economic incentive measures have been introduced to strengthen the responsibility of society. After a 2-year pilot study, the Shan-Feng company began to make a profit. The environmentally sound composting fertilizer that was produced not only met the demand for green food production and garbage disposal, but also partly solved the relevant environmental problems of livestock wastes, crop residues and human excrements. And the most important finding is the new sense of self-reliance that the MSW problem can be solved through ecological engineering. A national MSW conference was held in Guanghan in 1999 to disseminate the results to other counties (Hu *et al.*, 1998).

REFERENCES

Hu, D, Wang, R, Yan, J, Xu, C, and Wang, Y (1998) A Pilot Ecological Engineering Project for Municipal Solid Waste Reduction, Disinfection, Regeneration and Industrialization in Guanghan City, China, *Ecol. Eng.*, 11, 129–138.

Lu, B, Yan, J, and Wang, R (1998) Integrated Ecological Engineering of Corn Utilization in Zhaodong County, *Ecol. Eng.*, 11, 139–146.

Wang, R and Yan, J (1998) Integrating Hardware, Software and Mindware for Sustainable Ecosystem Development: Principles and Methods of Ecological Engineering in China, *Ecol. Eng.*, 11, 277–290.

Wang, R, Yan, J, and Mitch, W J (1998) A Promising Approach Towards Sustainable Development in Developing Countries, *Ecol. Eng.*, 11, 1–16.

Yu, Y and Liu, S (2001) Eco-building Design in Jinghua City, in *Sustainable and Adaptive Technology for Middle and Small Sized Cities' and Towns' Development: Methods of Planning and Management*, eds R Wang and Z Ouyang, Meteorological Press, China.

MINK (Missouri, Iowa, Nebraska, Kansas) Study

Norman J Rosenberg

Pacific Northwest National Laboratory, Washington, DC, USA

There is concern that the increasing concentrations of CO_2, CH_4, N_2O and other radiatively active trace gases will lead to global warming and climate change. How global climate change will play out in speci c regions of the world cannot yet be predicted with con dence. However, a number of the general circulation models (GCMs) used to project climate change agree that the mid-continental portions of Northern Hemisphere continents will become warmer and drier. For this reason, a group of researchers at Resources for the Future and the Paci c Northwest National Laboratory undertook to study the implications of a change to a warmer- drier climate on the total economy of a portion of the central United States - Missouri, Iowa, Nebraska and Kansas (hereafter the MINK region or MINK).

The MINK region (Figure 1) is primarily agricultural except on the Ozark Plateau in southeast Missouri where forestry is practiced and in the Nebraska Sandhills, used primarily for grazing. The economy of the region is more dependent on agriculture than any other in the country.

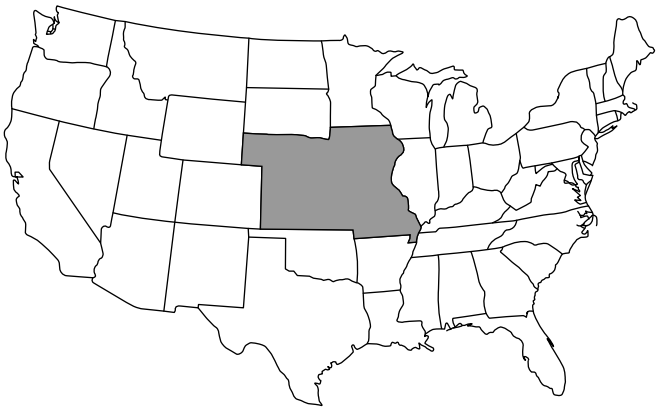


Figure 1 The MINK region (Missouri, Iowa, Nebraska and Kansas) of the central United States

The major rivers in MINK – the Missouri, Arkansas-White-Red, upper and lower Mississippi – support irrigation and navigation and provide hydropower. The MINK study focused on agriculture, forestry and water resources. Implications for the energy sector were also analyzed. Details of the study can be found in Rosenberg (1993).

Rather than use GCM projections of climate change, the researchers developed a surrogate scenario based on weather records for the subject region for the period 1931–1940 – the dirty thirties'. The most severe and protracted drought in the recorded history of the United States was experienced in this decade. These records and records for the period 1951–1980 (by definition of the World Meteorological Organization the normal climate at the time of the study – 1988–1992) were used to drive mathematical models that simulate crop and forest productivity under the normal (current) and 1930s climatic conditions. It was necessary in these simulations to consider that the rising atmospheric carbon dioxide (CO_2) concentrations that drive greenhouse warming are also of direct benefit to plant growth. Rates of photosynthesis increase and water use decreases in response to a so-called CO_2 fertilization effect. A different approach was taken to estimate how rivers in the MINK region would fare under the more stressful climate of the 1930s. Stream flow in the 1930s was compared to that during the 1951–1980 period in 15 watersheds that were then and are still essentially unmanaged. Further, since measurable climate changes may be decades off in the future, the 1930s' climate (with and without CO_2 fertilization) was also imposed on the farms and forests as they might be in the year 2030 when greater regional and global populations will increase demand for food, wood products and water, and new production methods will have been deployed. Finally, the efficacy of adaptation to climate change was studied by comparing crop and forest productivity and river flows with and without attempts by farmer, forester and water manager to adjust to change.

Assuming no attempt is made to adapt, the models indicate that the 1930s climate would lower current yields of dryland corn, sorghum, and soybean by about 25% for the region as a whole; CO_2 fertilization would reduce these losses to between 13 and 17%. With CO_2 fertilization region-wide yields of dryland wheat were 10% higher than at present. Aggregate loss of crop yield for the region as a whole is only 4–5% if farmers make use of currently available methods for adaptation such as changing planting dates and using new adapted crop varieties. By 2030 anticipated technologies would raise today's crop yields by 70% or more, assuming no climate change. However, with the dirty thirties' climate imposed on 2030 yields of these more productive crops would still be subject to losses of about 25%. If CO_2 fertilization were to take hold, these losses would be largely offset.

The economy of the region as a whole would not suffer greatly from imposition of the surrogate climate change, either today or in 2030. Allowing for sectoral adjustments to the new climate, the decline in regional income and production would likely not exceed 1–2%. But these aggregate regional estimates obscure the likelihood that certain sectors could be severely disrupted. Irrigation would not be sustainable in western Nebraska and Kansas and would shift to the more topographically varied eastern portions of the region. Long-term effects on forestry would be severe. With little potential for adaptation in the already marginal forests of the Ozark region, the models show that productivity would be lower still and many of the current species would vanish. Impacts on water resources would be severe, with current stream flow decreasing by 28% in the Missouri and upper Mississippi and 12% in the Arkansas-White-Red river basins. Water-based recreation would suffer and navigation on the Missouri would become uneconomic in these circumstances. The energy sector would not be greatly affected. Increased demand for air conditioning would be largely offset by reduced demand for space heating and crop drying.

MINK was the first study to consider the impacts of a climate change scenario on the total economy of a region and to draw policy guidance from its results. The effects on agriculture are most easily quantified (see above). Unless demand for biomass energy warrants investment in tree plantations, the only management that would be economically justified in the region's forests would be the harvest of dead and dying trees. Removal of subsidies for irrigation would preserve more water to sustain income-generating wildlife habitat, fisheries and water-based recreation. Diverting water from navigation for greater hydropower production would be not only a necessary adaptation but would also reduce the amount of fossil fuels consumed in the region. Maintenance of a

strong research establishment will be essential if the new technologies needed to permit adaptation are to be there when needed.

REFERENCE

Rosenberg, N J, ed (1993) Towards an Integrated Impact Assessment of Climate Change: The MINK Study, 1993, in *Climatic Change*, Vol. 24, Kluwer, Dordrecht, 1–173.

RAINS (Regional Air Pollution Information and Simulation)

The prototype of the Regional Air Pollution Information and Simulation (RAINS) model was developed at the International Institute for Applied Systems Analysis (IIASA) in Austria in 1984. The model has evolved considerably over the years and currently has two versions: one for Europe and one for Asia. The model contains modules for pollution generation (from industry, power sector, transport and agriculture), long range transport of air pollutants, and environmental effects of acidifying compounds.

RAINS is a tool for analyzing alternative strategies to reduce acidification, eutrophication and ground-level ozone in Europe. It combines a variety of information relevant for the development of cost-effective emission control strategies in Europe:

- projections of future economic, agricultural and energy development in 38 European countries;
- the present and future emissions of sulfur dioxide, NO_x, volatile organic compounds (VOC) and ammonia resulting from these activities;
- the options for reducing emissions and the costs of these measures;
- the atmospheric dispersion characteristics of sulfur and nitrogen compounds and the formation of ground-level ozone;
- the environmental sensitivities of ecosystems towards acidification, eutrophication and ground-level ozone.

RAINS integrates latest scientific findings, up-to-date databases and advanced systems analytical (optimization) tools.

As far as technically possible, the RAINS software package is operational in a standard PC environment and/or on the Internet. In early 2001 the following versions were publicly available:

1. PC version 7.2 (released early 1997) (http://www.iiasa.ac.at/• rains/index.html)
2. RAINS web version
   - RAINS PM Web Module giving access to particulate matter databases (http://www.iiasa.ac.at/• rains/index.html)
   - On-line ozone/acidification scenario analyses (http://www.iiasa.ac.at/• rains/index.html)
3. For Asia, the RAINS-Asia PC software (http://www.iiasa.ac.at/Research/TAP/rains_asia/docs/index.html)

Work is underway to transfer, as far as technically possible, the full RAINS model to the Internet and to enable permanent general access to up to date databases. Furthermore, RAINS will be extended to include fine particulate matter, its sources, abatement options, transport and effects.

The RAINS model is the outcome of a multi-year scientific effort of an interdisciplinary and international team of researchers at IIASA.

RAINS has been at the center of major international environmental negotiations. Twice it has been central to renegotiation of the Convention on Long-range Transboundary Air Pollution, the umbrella convention regarding air pollution across all Europe (*see* LRTAP (Convention on Long Range Trans-boundary Air Pollution), Volume 4). RAINS also underpins the European Union policy and directives on air pollution. Countries in Southeast Asia are turning to the model for help with their growing air pollution problems.

With some 900 copies of RAINS delivered to governments, private firms, academic institutions, and environmental non-governmental organizations (NGOs) in more than 50 countries on four continents, RAINS is one of the most successful integrated assessment tools yet devised.

The RAINS prototype model for Asia was developed in 1992 and is currently focused on sulfur dioxide. The RAINS-Asia model offers the opportunity to assess sulfur deposition and ecosystems protection levels resulting from different energy pathways and different emission control strategies. The costs of various control options are also provided.

FURTHER READING

Alcamo, J M, Shaw, R, and Hordijk, L, eds (1990) The RAINS Model of Acidification, Kluwer Academic, Dordrecht, The Netherlands.

Schopp, W, Amann, M, Cofala, J, Heyes, C, and Klimont, Z (1999) Integrated Assessment of European Air Pollution Emission Control Strategies, *Environ. Model. Software*, 14(1), 1–9.

Tuinstra, W, Hordijk, L, and Amann, M (1999) Using computer models in international negotiations, *Environment*, 41, 33–42.

LEEN HORDIJK *The Netherlands*

Transboundary Water Resource Management: The Example of the Canada/US IJC (International Joint Commission)

A P Lino Grima
University of Toronto, Toronto, Canada

Over the last nine decades, the Canada- US IJC has established a reputation as a credible international model for dealing with transboundary water issues. The IJC was set up in 1912 under the provisions of the 1909 Boundary Waters Treaty. It consists of three members appointed by the US and three members appointed by Canada, all of whom are expected to act as members of the IJC rather than as appointees of the two signatories. Over the last decades the reports of the IJC have been unanimous and the members of the commissions have hardly ever voted along national lines.

The IJC has given a lot of attention to Great Lakes water quality and quantity (Ashworth, 1986; Weller, 1990) and its work on the Great Lakes best exemplifies both its strengths and weaknesses. Because of the IJC's responsibilities for the Great Lakes, the IJC has a Windsor Great Lakes office in addition to the two offices in the national capitals. However, its responsibilities extend to the whole Canada-US border. The IJC has two functions. First, it has quasi-judicial powers to approve applications for the use or diversion of boundary waters that would affect their ow or level. It is a measure of the success of this international body (and of the independence of its members over the years) that the decisions on these applications have never been challenged in court. Second, the IJC may receive a reference, i.e., a request to study and report on an issue and make recommendations for its resolution. The IJC has dealt with numerous references, the first one in 1912, regarding the high level of bacterial pollution in the Great Lakes. Other issues that the IJC has studied as references over the last nine decades include the Trail Smelter air pollution issue, the Columbia River, the allocation of Niagara hydroelectric power, the ooding of the Red River in Manitoba and North Dakota and the pollution from land-based activities in the Great Lakes Basin. It should be noted that on several issues (e.g., Columbia River, Niagara power), the two governments signed a new treaty rather than adding new responsibilities to the IJC. There have been several references about the uctuating levels of the Great Lakes (Grima, 1993; Grima and Wilson-Hodges, 1977) and most recently the two federal governments have asked the IJC to study and make recommendations for protecting the integrity of the boundary waters in the face of expected threats from climate warming and bulk exports of water (IJC, 1999). These references have given the IJC a well-deserved reputation for putting together a common factual and technical base (Spencer *et al.*, 1981). This is an essential first step for moderating debate among the competing publics and also of providing direction for action by the US and Canadian governments. In order to discharge this function, the IJC requires technical information about the hydrological–chemical–biological–economic system and information about what the various publics want.

One of the most innovative features of the IJC (especially since its rules of operation were established 90 years ago!) is its obligation to consult the public about its references. The IJC notifies the public at the start of its reference in order that interested groups of individuals may submit opinions and information to the IJC or the technical study group set up by the IJC to conduct an investigation. At the end of study the IJC shall hold public hearings to allow the public to respond to the findings of the study and the recommendations. In this respect, the IJC has been a pioneer in establishing public participation in the policy making process! (Grima and Mason, 1983).

One of the major successes of the IJC was the signing of the Canada-US Great Lakes Water Quality Agreement (GLWQA) by President Nixon and Prime Minister Trudeau, in 1972, when the most pressing and visible problem was the elevated concentrations of phosphates and consequent proliferation of algae. The GLWQA is the most significant instrument for protecting and restoring the integrity of the Great Lakes ecosystems. The phosphorus loads were lowered across the basin especially by discontinuing the use of phosphate detergents and improvements in point sources of nutrient loading, particularly the sewage treatment plants across the Great Lakes Basin. For example, the total phosphorus load from municipal sources to Lake Erie was reduced from about 15 000 ton year$^{\bullet 1}$, in 1972, to about 2500 ton year$^{\bullet 1}$ by 1983 (Colborn *et al.*, 1990). On the other hand much less progress has been reported with respect to non-point (i.e., diffuse) sources such as agricultural runoff and the clearing of ground cover during the building phase of housing subdivisions. The reduction in eutrophication that originated largely in phosphate-rich detergents and the consequent restoration of water quality in the Great Lakes Basin is an example of successful environmental management and international cooperation.

The GLWQA was revised in 1978, to re ect the less visible pollution from the use of chlorine-based chemicals such as dioxin and other organochlorines (Colborn *et al.*, 1996). Among the principles included in the agreement, one could note the use of the ecosystem approach

(*see* Ecosystem Approach, Volume 2), and zero discharge of persistent toxic substances. The GLWQA was further revised in 1987 (US and Canada, 1987). The 1987 revision of the agreement commits the two governments to develop *remedial action plans* (RAPs) for the 43 *areas of concern* (AOC) (since reduced to 42 with the cleanup and delisting of the Collingwood, Ontario, AOCs) and also lakewide management plans for each of the five Great lakes. The RAPs were to be developed with public advisory committees (PAC) (called CAG or citizen advisory groups in the US), which were set up as stakeholder groups for each of the AOCs. In addition, the RAPs shall embody an ecosystem approach to restoring and protecting the beneficial uses of the water in each AOC. The ecosystem approach refers to the comprehensive treatment of issues such as urban runoff and industrial discharges as well as to the involvement of a wide array of agencies and institutions at the local, state/provincial and federal level, private companies and non-governmental organizations (NGO). The RAP program is significant because cleaning up the most polluted parts of the Great Lakes is the necessary first step to demonstrate the commitment of the governments of the US and Canada to achieve the objectives of the GLWQA. The narrow objectives of the RAPs is to restore the beneficial uses of the 43 AOCs. These beneficial uses may be summarized as drinkable, fishable and swimmable waters. The beneficial uses have been reduced by at least 14 specified impairments including beach closings to swimming, taste and odor problems in drinking water and reproductive problems in birds, mammals and fish.

On the 10th anniversary of the formal start of the RAP process, (which also coincided with the 25th anniversary of the signing of the GLWQA), the Great Lakes Water Quality Board (WQB) of the IJC published its assessment of the progress of the RAPs (WQB, 1997). They selected 21 examples of successful RAPs that included the delisting of Collingwood, Ontario as an AOC, and considerable expenditures ($1 billion) in the Rouge River, MI, to reduce combined sewer over ows. The other 19 examples record much more modest accomplishments and there are another 21 RAPs that were not included at all in this list of selected examples of successful RAPs! The WQB's assessment also records the budget cutbacks and reduced support for the RAPs and the associated public advisory committees. Other observers, such as the International Association of Great Lakes Research (1997), while describing the RAPs as an unprecedented collaboration of international significance, also emphasize that reduced funding levels and changing government priorities place the process in serious jeopardy. The pre-eminent NGO in the region, Great Lakes United (1997), is even harsher in its assessment. Its statement points out that the major shortcomings are the removal of funds, particularly in Ontario and the fact that the RAPs have not been incorporated into existing institutional planning frameworks, which one would expect in a comprehensive ecosystem approach. (See also the more balanced set of comments by four academics in the same commentary). The cheerleading stance of the WQB is expected since the members of the WQB are also the administrators in the federal, state and provincial agencies that have to make the RAP process work! In addition, as the WQB assessment quite correctly points out, greater emphasis should be placed on celebrating and marketing successes achieved over the last 10 years in order to maintain public interest in the RAPs and build a constituency to increase pressure for restoring budgets.

One of the reasons for revising the GLWQA, in 1978, was the realization that one very important Great Lakes issue had not been addressed in 1972, viz., the high levels of organochlorines and heavy metals in the ecosystem and the consequent effects of exposure to these synthetic chemicals on humans and wildlife. Deformities such as crossed bills and unhatched eggs in fish-eating birds, reduced mink populations, lesions in lake trout and other top aquatic predators are ascribed to these persistent pollutants which are found in concentrations of 1 parts per billion (ppb) or so in the water but which eventually bio-accumulate and bio-magnify as they are ingested up the food chain (Colborn *et al.*, 1990). In addition to being suspected carcinogens, these persistent toxins also disrupt the endocrines system in wildlife (i.e., they disrupt embryonic development even in healthy adults) (Colborn *et al.*, 1996). The IJC and the Great Lakes policy makers have not registered much progress in controlling these chlorine-based chemicals, which include pesticides, polychlorinated biphenyls (PCB), dioxins and furans. And progress is likely to be slow since these chlorine-based chemicals are used in numerous items in every day use such as water pipes, building materials, plastic containers, car parts and so on. In addition to the continuing use of such chemicals, there are large quantities buried in sediments and landfills (the most notorious example was the Love Canal), in most of the AOCs. The RAPs are supposed to be dealing with these reservoirs of pollution. However, the RAPs themselves are at a crossroad, perhaps in crisis, largely because of budget cutbacks and also because they lack an appropriate administrative structure for developing and implementing a clean-up plan.

Indeed, the RAPs could serve as a good example of the strengths and weaknesses of the IJC as a whole. The strengths of the IJC include its quasi-judicial tasks in dealing with water level issues; it acts as a fact-finder when it is dealing with references for studies of complex issues; it brings together teams of international experts and provides an international perspective to bi-national issues; it has gained respect because of its reputation for impartiality; and it allows for regular meetings of

middle and senior managers. Its weaknesses stem from its utter dependence on the agendas and priorities of the two federal governments; for its operations it depends on the two governments for budget and personnel, which reinforces its fragility as an advisory body. The use of agency personnel creates a situation where the experts who advise the IJC are the same administrators of the agencies that are responsible for the state-of-the-lakes in the first place. It is most unlikely that the IJC's functions will be expanded beyond its present considerable powers to study, recommend, cajole and persuade. It is almost certain that the IJC will continue to play a lead role in the management of water resources shared by two friendly nations and therefore will continue to stand out as a successful example of international environment management.

See also: Water Resources: Great Lakes Case Study, Volume 3.

REFERENCES

Ashworth, W (1986) *The Late Great Lakes: an Environmental History*: Alfred Knopf, New York.

Colborn, T A, Davidson, A, Green, S N, Hodge, R A, Jackson, C I, and Liroff, R A (1990) *Great Lakes, Great Legacy*? The Conservation Foundation and the Institute for Research in Public Policy, Baltimore, MD.

Colborn, T, Dumanoski, D, and Myers, J P (1996) *Our Stolen Future*. Penguin Books, New York.

Great Lakes United (1997) Position Statement on RAPs, *Commentary on the Status of RAPs, J. Great Lakes Res.*, 23(2), 224–225.

Grima, A P (1993) Enhancing Resilience in Great Lakes Water Levels Management, *Int. J. Environ. Stud.*, 44, 97–111.

Grima, A P (1997) Ten Years of Remedial Action Plans (RAPs): Re ections and Suggestions, *Commentary on the Status of RAPs, J. Great Lakes Res.*, 23(2), 232–233.

Grima, A P and Mason, R (1983) Apples and Oranges: Towards a Critique of Public Participation in Great Lakes Futures, *Can. Water Resour. J.*, 8(1), 22–50.

Grima, A P and Wilson-Hodges, C (1997) Regulation of Great Lakes Water Levels: The Public Speaks Out, *J. Great Lakes Res.*, 3(3–4), 240–257.

International Association of Great Lakes Research, (1997) Position Statement on RAPs, *Commentary on the Status of RAPs, J. Great Lakes Res.*, 23(2), 221–223.

IJC (1999) Protection of the Waters of the Great Lakes: *Interim Report to the Governments of the US and Canada*, Washington, DC and IJC, Ottawa.

Spencer, R, Kirton, J, and Nossal, R, eds (1981) *The International Joint Commission: 70 Years On*. University of Toronto Press, Toronto.

US and Canada (1987) *Protocol to the 1978 GLWQA*. IJC, Windsor, Ontario.

Water Quality Board (1997) Position Statement on the Future of the RAPS, September 1996, Commentary on the Status of RAPs, *J. Great Lakes Res.*, 23(2), 212–221.

Weller, P (1990) *Fresh Water Seas: Saving the Great Lakes*, Between the Lines, Toronto.

FURTHER READING

Colborn, T, Dumanoski, D, and Myers, J P (1997) Hormone Impostors, *Sierra*, 82(1) 28–35.

Encyclopedia of
Global Environmental Change

egec

The volumes in this Encyclopedia are:

Volume One
The Earth system: physical and chemical dimensions of global environmental change

Volume Two
The Earth system: biological and ecological dimensions of global environmental change

Volume Three
Causes and consequences of global environmental change

Volume Four
Responding to global environmental change

Volume Five
Social and economic dimensions of global environmental change

Encyclopedia of
Global Environmental Change

Editor-in-Chief
Ted Munn
Institute for Environmental Sciences, University of Toronto, Canada

5

egec

Social and economic dimensions of global environmental change

Volume Editor
Peter Timmerman
IFIAS, Toronto, Canada

JOHN WILEY & SONS, LTD

Copyright © 2002 John Wiley & Sons, Ltd, Chichester
Baffins Lane, Chichester
West Sussex PO19 1UD, UK

National: 01243 779777
International: (• 44) 1243 779777
e-mail (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on http://www.wiley.co.uk
or http://www.wiley.com

Copyright Acknowledgments
A number of articles in the Encyclopedia of Global Environmental Change have been written by government employees in the United Kingdom, Canada and the United States of America. Please contact the publisher for information on the copyright status of such works, if required. In general, Crown copyright material has been reproduced with the permission of the Controller of Her Majesty's Stationery Office. Works written by US government employees and classified as US Government Works are in the public domain in the United States of America.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except under the terms of the Copyright Designs and Patents Act 1988 or under the terms of a license issued by the Copyright Licensing Agency, 90 Tottenham Court Road, London, W1P 0LP, UK, without the prior permission in writing of the publisher.

Other Wiley Editorial Offices

John Wiley & Sons Inc., 605 Third Avenue,
New York, NY 10158-0012, USA

Wiley-VCH Verlag GmbH, Pappelallee 3,
D-69469 Weinheim, Germany

Jacaranda Wiley Ltd, 33 Park Road, Milton,
Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01,
Jin Xing Distripark, Singapore 129809

John Wiley & Sons (Canada) Ltd, 22 Worcester Road,
Rexdale, Ontario M9W 1L1, Canada

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library.

ISBN: 0 471 97796 9

Typeset in 10pt Times by Laser Words Private Limited, Chennai, India
Printed and bound in Great Britain by Antony Rowe, Chippenham, Wiltshire

Editorial Board

Editor-in-Chief

Ted Munn
Institute for Environmental Studies, University of Toronto, Canada

Volume Editors

Volume One ***The Earth system: physical and chemical dimensions of global environmental change***
Dr Michael C MacCracken, *Lawrence Livermore National Laboratory, California, USA* and
Dr John S Perry, *formerly, National Research Council USA*

Volume Two ***The Earth system: biological and ecological dimensions of global environmental change***
Professor Harold A Mooney, *Stanford University, Stanford, USA* and
Dr Josep G Canadell, *GCTE/IGBP, CSIRO Sustainable Ecosystems, Canberra, Australia*

Volume Three ***Causes and consequences of global environmental change***
Professor Ian Douglas, *University of Manchester, Manchester, UK*

Volume Four ***Responding to global environmental change***
Dr Mostafa K Tolba, *International Center for Environment and Development, Cairo, Egypt*

Volume Five ***Social and economic dimensions of global environmental change***
Mr Peter Timmerman, *IFIAS, Toronto, Canada*

International Advisory Board

Dr Joe T Baker
Commissioner for the Environment, ACT, Australia

Professor Francesco di Castri
CNRS, France

Professor Paul Crutzen
Max Planck Institute for Chemistry, Germany

Professor Eckart Ehlers
IHDP/IGU Germany

Professor José Goldemberg
Universida de São Paulo, Brazil

Dr Robert Goodland
World Bank, USA

Professor Hartmut Grassl
WCRP, Switzerland

Professor Ronald Hill
University of Hong Kong, Hong Kong

Dr Yu A Izrael
Institute of Global Climate & Technology, Russia

Professor Roger E Kasperson
Clark University, USA

Professor Peter Liss
University of East Anglia, UK

Professor Jane Lubchenco
Oregon State University, USA

Mr Jeffrey A McNeely
IUCN, Switzerland

Professor Thomas R Odhiambo
Hon. President, African Academy of Sciences and
Managing Trustee, RANDFORUM, Kenya

Sir Ghillean T Prance
University of Reading, UK

Professor Steve Rayner
Columbia University, USA

Contents

Preface to the Encyclopedia of Global Environmental Change ix

Preface to Volume Five xv

The Human Dimensions of Global Change 1
The Changing Human–Nature Relationships (HNR) in the Context of Global Environmental Change 11
Economics and Global Environmental Change 25
Ecological Economics 37
Environmental Politics 49
Global Environmental Change and Environmental History 62
Globalization in Historical Perspective 73
Technological Society and its Relation to Global Environmental Change 86
Relating to Nature: Spirituality, Philosophy, and Environmental Concern 97
Social Science and Global Environmental Change 109
The Emergence of Global Environment Change into Politics 124
The Environment and Violent Conflict 137
Development and Global Environmental Change 150

Anthropology and Global Environmental Change 163
Art and the Environment 167
Attenborough, David 175

Baha'i Faith and the Environment 176
BAT (Best Available Technology) 183
Bateson, Gregory 183
Brent Spar 184
Brower, David 185
Buddhism and Ecology 185
Business-as-usual Scenarios 191

Carson, Rachel Louise 192
CBA (Cost Benefit Analysis) 193
Chipko Movement 193
Christianity and the Environment 194
Circulating Freshwater: Crucial Link between Climate, Land, Ecosystems, and Humanity 201
Commons, Tragedy of the 208
Cousteau, Jacques 209

Deep Ecology 211
Demographic transition 211
Discounting 214

Earth Charter 216
Earth Day 216
Earth First! 217
Ecocentric, Biocentric, Gaiacentric 217
Ecofeminism 218
Eco-socialism 224
Ecosystem Approach 225
Ecosystem Services 226
Emergy 228
Encyclopedias: Compendia of Global Knowledge 228
Enlightenment Project 229
Environmental Defense Fund 230
Environmental Economics 230
Environmental Ethics 231
Environmental Movement – the Rise of Non-government Organizations (NGOs) 243
Environmental Philosophy: Phenomenological Ecology 253
Environmental Psychology/Perception 257
Environmental Security 269
Environmental Sociology 278
Equity 279
Exxon Valdez Oil Spill 283

Francis of Assisi 284
Friends of the Earth 285
Futures Research 285

Gaia Hypothesis 287
Gandhi, Mohandas Karamchand 290
Governance and International Management 292

Hazards in Global Environmental Change 297
Hinduism and the Environment 303
Homocentric 311
Human Body, Immediate Environment 312

Indigenous Knowledge, Peoples and Sustainable Practice 314
International Environmental Law 324
ISEW (Index of Sustainable Economic Welfare) and GPI (Genuine Progress Indicator) 331

Islam and the Environment 332
ISSC (International Social Science Council) 339

Jains and the Environment 341
Judaism and the Environment 349

Kelly, Petra 355
Kondratyev, Nikolai Dmitrievich 356

Landscape, Urban Landscape, and the Human Shaping of the Environment 357
Leopold, Aldo 367
Life Style, Private Choice, and Environmental Governance 368
Literature and the Environment 370
Love Canal 382

Malthus, Thomas Robert 384
Man and Nature as a Single but Complex System 384
Modeling Human Dimensions of Global Environmental Change 394
Modernity vs. Post-modern Environmentalism 408
Muir, John 411

Nasr, Seyyed Hossein 412
National Environmental Law 413
Nature 419
New Ageism 420
NIMBY (Not In My Backyard) 421
Noosphere 421

Phase Shifts or Flip-flops in Complex Systems 422
Political Movements/Ideologies and the Environment 429
Political Systems and the Environment: Utopianism 443
Post-normal Science 451
Precautionary Principle 455
Property Rights and Regimes 457

Religion and Ecology in the Abrahamic Faiths 461
Religion and Environment Among North American First Nations 466
Roosevelt, Theodore 475

Scenarios 476
Seveso 482
Sierra Club 483
Small is Beautiful 483
Social Ecology 484
Social Impact Assessment (SIA) 484
Social Learning in the Management of Global Atmospheric Risks 485
Soft Energy Paths 487
Sovereignty and Sovereign States 487
Sustainability: Human, Social, Economic and Environmental 489

Theology 492
Theories of Health and Environment 492
Thoreau, Henry David 502
Three Mile Island 503
Torrey Canyon 503

UNU (United Nations University) 504
Virtual Environments 505
Waldsterben 508
WCC (World Council of Churches) 509

Alphabetical List of Articles 511

List of Contributors 523

Selected Abbreviations and Acronyms 553

Index 569

Preface to the Encyclopedia of Global Environmental Change

THE ENVIRONMENT

The word *environment*, whose dictionary meaning is simply that which surrounds, has in the last few decades become a buzzword , encompassing an exceedingly diverse array of elements and social issues. Taking the original meaning as a point of departure, it is clear that we humans depend totally on the environment provided by planet Earth for the food we eat, the water we drink, and the air we breathe. Thus changes in this environment must be of vital concern. Will Earth continue to sustain humans in a way that also encourages the ourishing of the other living things with whom we share the planet? This question has loomed ever larger as it has become more evident that human activities have been inducing major changes in all of the compartments of the global environment. We have converted forests and savannas to farms and cities; we have exhumed ancient treasures of fuels and minerals; we have used the rivers and winds as convenient sewers; and we have released entirely new chemical compounds and organisms into the environment. In the 1960s, the scienti c community began to use the word *environment* in this new non-specialist sense. Soon too, Departments of the Environment were created by many governments, and new scienti c journals began publication while others were re-named. For example, the *International Journal of Air Pollution* became *Atmospheric Environment*! In the ensuing decades, the world community has come to see the environment in many different ways, as a life-support system, as a fragile sphere hanging in space, as a problem, a threat and a home.

GLOBAL ENVIRONMENTAL CHANGE

A broader and deeper understanding of the global aspects of environmental concerns emerged in the 1970s and 1980s, and a new phrase *global environmental change* acquired popular currency. Paleoresearch had revealed that environmental change was far from new, and by no means the sole result of the actions of heedless humans. Since the planet s formation, virtually every element of its environment has been undergoing massive changes on all space and time scales. Oxygen waxed and carbon dioxide waned in the atmosphere. Continents moved about the planet s surface like scum on a soup kettle. Great ice sheets grew and shrank. Above all, the force of life (the biosphere) emerged as a dominant driver of planetary change.

However, another vital insight began to emerge about 1980: the inescapably interlinked nature of these many environmental changes. On the very longest time scales, continental drift moved lands into different climates but also changed the climate of the globe itself. Photosynthetic life changed the atmosphere but also made possible more advanced life forms that could take advantage of the new environment. On shorter time scales, atmosphere and ocean often interact to produce the massive changes in the Southern Paci c that we term El Niño and La Niña, whose consequences extend across the planet, and profoundly affect even our socioeconomic systems. Indeed, we have come to realize that human-induced perturbations in the environment are becoming increasingly large, and are potentially coming to dominate the natural workings of the complex and interdependent global system that sustains life on Earth. Humans and their global environment are no longer independent; they are ever-increasingly becoming interdependent components of a single global system.

Thus, the term *global environmental change* has come to encompass a full range of globally signi cant issues relating to both natural and human-induced changes in Earth s environment, as well as their socioeconomic drivers. This implicitly includes concerns for the capacity of the Earth to sustain life that have motivated the development of studies of global change and sustainable development in the last few decades. Analyses of global environmental change therefore demand input from the social sciences as well as the natural sciences (and indeed also from the engineering and health sciences) necessitating an inescapably interdisciplinary approach.

Scientists from many disciplines have been attracted to this growing eld of global environmental change. This is particularly noticeable in the biological sciences through the encouragement of IGBP (International Geosphere Biosphere Programme), which invites ecologists to expand their eld of vision from the plot and landscape scales to the regional, continental and global ones, and to interact with scientists from other disciplines in exploring environmental change at these larger scales. Indeed this trend has encouraged publishers and scienti c societies to introduce new journal titles, and these publications from all accounts appear to be ourishing.

At the same time, human social, economic, and cultural systems are rapidly changing under the in uence of growing globalization. In the economic sphere, for example, today s discourse centers on multinational corporations, the Global

Economy, and the Globalization of Trade. Thus, the agendas of most of the physical, life, and social sciences increasingly focus on global-scale changes.

THE EVER-CHANGING ENVIRONMENT

Since the launching of Earth-viewing satellites, we have been able to see a constantly updated moving picture of our planet. Great weather systems sweep around the globe, waves of green and brown ebb and flow over the continents with the seasons, as do white waves of ice and snow. Nevertheless, a remarkable feature of the Earth system has been its relative stability. Since the dawn of life, the planet's environment has remained within a range of conditions that has supported life. Moreover, with some notable and perhaps cautionary exceptions, changes from year to year, decade to decade, millennium to millennium, have been modest, particularly during the last 10 000 years. Many features of human society exhibit similar behavior. Civilizations and cultures evolve slowly for the most part in response to environmental, economic, and social driving forces over their long lives. Great cities like Rome endure through the ages. A citizen of ancient Babylon probably would have little difficulty understanding the politics of modern Toronto.

But today we appear to be entering an era of change unprecedented since Babylon was a cluster of mud huts. There are many reasons to believe that changes greater than humankind has experienced in its history are in progress and are likely to accelerate.

Virtually every measure of human society – numbers of people and automobiles, airplanes, energy consumption, generation of waste – is increasing exponentially. While the values of a few indicators are leveling off, the magnitudes of annual increases of others remain immense.

Major changes are becoming evident in many critical elements of the environment – increasing carbon dioxide concentrations, stratospheric ozone depletion (not to mention the stratospheric Antarctic ozone hole and the possibility of a similar Arctic event), rising sea level, declining productivity of soils, widespread collapses of fisheries, dramatic declines in biodiversity, destruction of tropical forests, etc. Within living memory, the countryside around large cities has been swallowed up by suburban developments and highways. Humans are stepping on the gas pedal of the planet's environment, and perhaps recklessly breaking the survivable speed limit of global change.

Paleoclimatological studies have shown that the stable environment that we take for granted has not always prevailed. Indeed, the relatively brief period in which human civilization has developed is somewhat unusual in its equable stability. Neolithic hunters in Europe some 13 000 years ago saw their climate shift from temperate to glacial in a single short lifetime. The evolution of human society is punctuated by wars, pestilence, and technological revolutions. Major – perhaps even catastrophic – change could occur in the future, because we know it has occurred in the past.

Rapidly advancing understanding of both natural and human systems – and above all the ability to translate that understanding into quantitative models – has enabled us to explore the future of our global society and its global climate with unprecedented realism. Although prediction of the single future path that we will follow is inherently unpredictable, it is possible to map a broad range of future environmental trajectories that we might take, each completely consistent with our understanding of how the system works. Such scenario-building exercises amply confirm our concerns that the changes of the 21st Century could be far greater than experienced in the last several millennia. Business-as-usual for human society appears to imply business-as-highly-unusual for the global environment.

THE INTERLOCKING BIOGEOPHYSICAL – SOCIOECONOMIC SYSTEMS

Recognition came in the 1970s that many of the environmental issues are inter-connected through the biogeochemical cycling of trace substances, especially carbon, sulfur, nitrogen and phosphorus. In fact, in a prescient statement on the main environmental research priority of the 1980s, Mostafa Tolba (then Executive Director of the United Nations Environment Programme) and Gilbert White (then President of the Scientific Committee on Problems of the Environment) drew the attention of both the scientific and the science-policy communities to the need to understand the major global biogeochemical cycles in order to maintain the global life support systems in a healthy state (Tolba and White, 1979). Quoting from that statement,

> We draw attention to the fundamental scientific importance of understanding the biogeochemical cycles which link and unify the major chemical and biological processes of the Earth's surface and atmosphere. The results will have practical significance for all of us who inhabit an Earth with limited resources and who, by our actions, increasingly affect the quality of the human environment.

So it is that many of the global environmental issues – acid rain, stratospheric ozone depletion, climate change, nitrogen over-fertilization – are inter-related through the global biogeochemical cycles.

One of the interesting results of the study of the Earth's history has been the discovery of the global teleconnections of the Earth system, with some major climatic shifts occurring simultaneously in the two hemispheres. In an analogous way, there has been recognition for a long time that human social, economic, and cultural systems are globally inter-related. That these are connected in turn to the

Earth system was implicitly recognized as early as the 1972 Stockholm Conference on the Human Environment.

To manage the human responses to this enormously varied but at least moderately coupled world system in an era of increasing global change through the diverse array of local, national and international organizations is indeed a daunting challenge. An essential first step is to describe past and present states, then to explain the various phenomena observed, and finally to develop predictive models (or at least a range of scenarios) describing the future behavior of this total Socioeconomic–Cultural–Environmental system. In this process, environmental scientists must learn how to assimilate better the new information constantly being received, although uncertainties, often large, will continue to exist. Within this mix, rational and effective policies must be developed that will balance the risks and costs of global environmental change in an adaptive way. Only then will we be able to even formulate, much less start to implement, rational and effective policies that cope effectively with global change. The prospect of change tempts us to think in terms of winners and losers. However, such analyses often do not play out in simple ways. For example, a modest increase in rainfall would cause farmers to rejoice while vacationers would despair. However, greater farm production can lead to lower commodity prices, thereby reducing farm incomes while making vacation food purchases less expensive. Human society is pretty well adapted to the present environment, so change is necessarily a challenge.

In the longer-term, much depends on the ability of societies to respond, to adapt. Societies will differ in the resources – natural, human, and technological – that are available to them. They will also differ in terms of the values and priorities they attach to physical, social, and environmental goals, and in the social and political mechanisms that they employ to seek these goals. These differences in the human world of generations yet unborn may be as great – and as significant – as the changes in the global environment. A major challenge for our times is to develop frameworks for understanding complex interdisciplinary issues of this complexity.

No discussion of change and the future can be complete without consideration of risk. Projections of the future, however imaginative or soundly based, necessarily center on plausible, surprise-free scenarios – population will increase, economies will advance, climate will change – all typically slowly and smoothly. However, such projections made at the beginning of the 20th Century would have missed two World Wars, the automobile, aviation, space travel, television, and McDonald's in Beijing. By definition, genuine surprises cannot be predicted. However, an understanding of the impacts of past surprises may help us to make our society and our world somewhat more robust in the face of the unknown surprises that await us.

THE HUMAN DIMENSIONS OF GLOBAL ENVIRONMENTAL CHANGE

The human dimension of global change has been defined as the ways in which individuals and societies contribute to global environmental change, are influenced by it, and adapt to it. Many empirical studies have been undertaken that describe human activities in physical terms, based on various kinds of indicators. Rates of deforestation, urbanization, and changing levels of emissions of greenhouse gases are only some measurable contributors to environmental change. But the dynamics of human activities and global change are much more complex, and they reflect the complexities of the human–nature relationship.

While the sheer burden of human activity on the planet is important, we know that the major forces at work are human beings operating together through political systems, corporations, interest groups, and beliefs that sway whole peoples. These raise questions about the nature of choice, about the hopes, dreams, and frustrations that impel people forward, or block them. At the moment, for instance, the dream of the good life – which has been a critical element of every religious tradition in the world – is increasingly being defined in terms of material possessions and powerful images that are shown on global media. Can this particular version of human happiness be sustained on a limited planet? Some people say that it can; others say that it cannot.

The central questions about the role of human beings in global environmental change revolve around social, cultural, economic, ethical, and even religious issues. These are becoming more and more pressing, and more and more foundational as human beings deliberately or inadvertently modify more and more of the planet. It is also obvious that in this generation the modification of organisms and ecosystems may well extend to the modification of human beings themselves. Among the fundamental questions are: What motivates us towards saving or harming the environment? How do we see ourselves with regard to nature? What is our responsibility to this and future generations? Who do we think we are, and who would we like to become? Appropriately enough, Volume Five of this Encyclopedia wrestles, in many different voices, with these ultimate questions that remain intimately linked with the sweep of physical, chemical, biological, geographical, and institutional changes documented and discussed throughout the earlier four volumes.

The Brundtland philosophy urges us not to reduce options for future generations. Implementation of this idea is, however, difficult. For example, it often seems more compelling to alleviate current poverty than to protect the environment and renewable resources for future generations. Many scientists agree that new approaches are needed to meld the social with the natural sciences in the policy arena. Some of the new methodologies that go beyond the physical sciences

are described in Volumes Four and Five of this Encyclopedia, e.g., post-normal science, integrated environmental assessment, and the precautionary principle.

In a commentary in which he described the 21st Century as the century of the environment, Edward O. Wilson began by referring to the growing *consilience* (the interlocking of causal explanations across disciplines) so that the interfaces between disciplines become as important as the disciplines themselves. He then stated that this interlocking amongst the natural sciences will in the 21st Century also touch the borders of the social sciences and humanities. In the environmental context, environmental scientists in diverse specialities, including human ecology, are more precisely defining the arena in which that species arose, and those parts that must be sustained for human survival (Wilson, 1998).

This is a major challenge for environmental scientists. Already through DNA techniques, for example, it is possible to trace back connections amongst prehistoric peoples through the last Ice Age to modern times.

THIS *ENCYCLOPEDIA OF GLOBAL ENVIRONMENTAL CHANGE*

More than two million papers are published every year in science and medicine, a twentyfold increase since 1940 that taxes the resources of concerned citizens, scientists, university departments, research institutes and libraries. At the same time, many policy-motivated organizations find it difficult to draw together the necessary expertise for resolving the newly emerging environmental issues. The scientific literature relating to the environment is burgeoning. However, research syntheses are in general still scattered across a broad spectrum of journals and books, and information on global environmental change is not readily available in an inter-related way. In particular, it is quite uncommon for contributions from the natural sciences to appear in the same journal or workshop proceedings as contributions from the social sciences and the humanities.

This *Encyclopedia of Global Environmental Change* is a comprehensive and integrated reference in the broad area of global environmental change that will be conveniently accessible and productively usable by students, managers, administrators, legislators, and concerned citizens. The Encyclopedia consists of five volumes of interrelated material:

Volume 1 The Earth system: physical and chemical dimensions of global environmental change (Michael C MacCracken and John S Perry, editors)

Volume 2 The Earth system: biological and ecological dimensions of global environmental change (Harold A Mooney and Josep G Canadell, editors)

Volume 3 Causes and consequences of global environmental change (Ian Douglas, editor)

Volume 4 Responding to global environmental change (Mostafa K Tolba, editor)

Volume 5 Social and economic dimensions of global environmental change (Peter Timmerman, editor)

The first four volumes cover the broad issues of the science and politics of global environmental change. Volume Five adds an often understated but extremely important aspect, linking as it does, global environmental change to the socioeconomic, cultural and ethical dimensions of human societies. Here will be found a rich panoply of writings by people who are not natural scientists but who have thought deeply about the environment. It places global environmental change in a refreshing historical, sociological and cultural context. In many contributions, the time horizons of most interest are the last hundred and the next hundred years. However, some contributions dip backward millions of years. Throughout, the emphasis is upon the dynamics of the various processes discussed – how and why did they change? A second recurrent theme is the interconnection of processes and changes – What produces change? What is impacted by change? Finally, we attempt to deal even-handedly with natural and human-induced change, and with impacts on both the natural world and human society. From the numerous diverse articles in the Encyclopedia, we believe that the user can obtain a coherent picture of this complex and dynamic system of which we all are a part.

To assist in promoting this coherence, each volume begins with a group of extended essays on major topics that embrace the field covered in that volume. These are intended to provide an introduction to the topic, a convenient road map through cross-references for exploring the Encyclopedia. Then there follows, in alphabetical order, shorter articles on a variety of scientific topics, descriptions of scientific programs, definitions, acronyms, and biographies of leading contributors to the field – from Charles Darwin, through the Russian ecologist Vernadski, to the three most recent environmental Nobel Laureates – Crutzen, Rowland and Molina. Indeed, these definitions, biographies and acronym definitions are, we believe, a uniquely valuable feature of the Encyclopedia. In the case of acronyms of international organizations and programs, it is no exaggeration to state that a young environmental scientist requires not only a good understanding of science, but also a good knowledge of acronyms, if they are to follow the discussions taking place at many international meetings! Also included in the alphabetic listings are abundant cross-references to related topics in the same or other volumes.

The substantive scientific articles that comprise the meat of the Encyclopedia are original contributions by active scientists from around the world, and thus

represent authoritative and up-to-date summaries of the state of current knowledge, direct from the producers of this knowledge. A number of these articles break new ground in synthesizing and summarizing our understanding from novel viewpoints and in unconventional ways. Thus, readers will find a wide variety of styles and approaches within the articles, reflecting in a unique way the rich diversity of today's world science. The articles have also had the benefit of careful reviews, particularly by our Editors.

The scientific essays and some of the program descriptions begin with a few italicized paragraphs written for non-specialists. These are not intended to be abstracts of the paper to follow, but rather are aimed at providing the reader with an introduction into why the topic is important and where it fits into the broader aspects of global change – a kind of encouragement to read on. Reading on should not be too difficult a task, since most of the scientific essays are written at the level of journals such as *Scienti c American* and *AMBIO* that are intended for the non-specialist. The Encyclopedia will be a valuable source of information for everyone with a general interest or a need-to-know in the various environmental fields (the natural sciences, socio-economics, engineering, the health sciences, and policy analyses) particularly as they relate to global-scale environmental change, its drivers, and its consequences. It is also expected that among the audiences for each volume will be practitioners and researchers in the fields covered by the other four volumes. We believe that this rather unusual – indeed unique – Encyclopedia will be used in a variety of ways.

1. Some people will employ it simply as a convenient source of information on specific topics – What has been happening recently to the sea ice in the Arctic Ocean? Who was Roger Revelle? What is soil mineralization? What is ICSU or IGBP? What is the Kyoto Protocol? What is deep ecology?
2. Serious students in the environmental sciences may begin by reading one or all of the introductory essays, which present highly compressed crash courses in a range of central topics.
3. We believe that the substantive specialized articles that constitute the bodies of the volumes are productively and enjoyably readable in their own right.
4. Finally, a case has been made by one of the volume editors (Peter Timmerman) for encyclopedia browsing as an enjoyable and productive pastime (*see* Encyclopedias: Compendia of Global Knowledge, Volume 5).

We believe that these five volumes are in the tradition of the human aspiration towards the compilation of global knowledge that sparked the preparation of the first encyclopedias in the 18th Century.

Our Editors have had the benefit of a marvellous postgraduate education in the course of the Encyclopedia's evolution! We are immensely grateful to the many, many authors in our virtual university who have contributed their wisdom to this project.

ACKNOWLEDGMENT

Many helpful comments and draft paragraphs were received from the Volume Editors during the preparation of this Preface.

REFERENCES

Tolba, M K and White, Gilbert F (1979) Global Life Support Systems, SCOPE Newsletter, No. 7, October. 2 pp.

Wilson, Edward O (1998) Integrated science and the coming century of the environment, SCIENCE 279, 2048–2049.

Ted Munn
Editor-in-Chief

Preface to Volume Five

Volume Five is devoted to the social, political, economic, and spiritual dimensions of global environmental change, and as such represents a departure from the conventional focus on the purely physical aspects of global change. It highlights the profound shifts in human thinking and awareness that are required to wrap our minds around the advent of globalization, and our increasing ability to affect natural systems, sometimes to our own benefit, sometimes not. This powerful role cannot simply be captured by treating human beings as if they were another natural force. Social, cultural, and economic ideas and institutions shape the desires and hopes, the conflicts and resolutions of conflict that are central to the human dimension of global change. Yet, at the same time, human beings are incontestably part of nature – as other volumes in this Encyclopedia demonstrate. Volume Five overlaps with, and complements, these other volumes.

Because of the complex weave of interaction between humanity and the environment, this volume contains many essays and articles that are more in the realm of probes than fixed descriptions of their topics. Powerful words and powerful ideas, metaphors, myths, beliefs, images and artefacts – these are all vehicles for the creation and shaping of meaning among human beings. Topics covered include the great political and economic theories, the most influential views of nature from Plato to Rachel Carson, and the historic and literary seedbeds for the rise of environmental thought and practice in our time.

Of particular importance in this volume are the introductory essays from leading figures in the field, and special efforts have been made throughout to give space to alternative voices and ideas. Dialogue and diversity are essential to human development, and we hope that the reader is stimulated by this volume towards his or her own thoughtful response to the increasing responsibility for the future of the Earth that has come upon us in our time.

Among the voices that we are privileged to present in this volume, the Editors would like to single out that of Ester Boserup, whose pathbreaking work as a student of the social dynamics of technological change, and of the role of women in economic development, make her an exemplar of the kind of interdisciplinary thinker so often proclaimed as necessary, and so seldom found. She died before the Encyclopedia could go to press, but we would like to dedicate this volume to her memory.

Peter Timmerman
Editor of Volume Five

The Human Dimensions of Global Change

PETER TIMMERMAN

University of Toronto, Toronto, Canada

The central icon of the environmental movement is the earth hanging in space, that mysterious and unexpected cargo brought back from the voyages to the moon. *See* Figure 1 in **The Earth System**, Volume 1. It appears on bedroom walls, refrigerators, and annual reports by multinational food companies. Blue-green and spherical against the stark blackness of space, it speaks of all things natural, all things green, the vision of one earth, one world.

Yet, part of its paradox is that this environmental symbol was the result of the most advanced technological project in human history (at that time), whose roots are easily traceable to the challenge of the Cold War, the long-range missile projects beginning in World War II, the atomic bomb, the militarization of space, and even the elegant camera technology that enabled the pictures to be taken.

But the image is even more iconic and mysterious than that. At rst glance, it appears a lonely image the only home we have devoid of human reference points, at least at the distance of the moon. Yet a nagging question is: where are we in this picture? The easy answer is that we are in there somewhere, maybe waving, or drowning. The more intriguing answer is that we are the picture: that it took all of science and technology that is, all of the ability to provide a temporary earth in a plastic suit in deep space to keep a human alive and all of history, that is, not just evolution, but all the social and cultural reasons that would make a being turn around at a precise moment oating in space, pick out the earth from the surrounding darkness as something worth photographing, and then lift a camera to his eyes, focus, and take a picture with some purpose in mind.

The literary critic, Marshall McLuhan, made a typically prophetic remark in 1970, in an obscure inaugural piece in an early environmental magazine (since defunct):

> whereas the planet had been the ground for the human population as gure; since Sputnik, the planet has become gure and the satellite surround has become the new ground •••. Once it is contained within a human environment, Nature yields its primacy to art.

McLuhan here invokes the familiar image of the gure-ground reversal the faces that turn into vases, and back again; or the duck that ips into a rabbit and ducks back again to suggest that while up until the arrival of the image from space, human beings saw themselves as gures on a ground, the environment is now a mere gure on human grounds.

The gure-ground reversal suggests that this is a sudden perceptual shift; which may be the case, but its elements have been arriving for some time.

One early element can be found in the idea that the higher you go off the earth, the closer you approach the realm of God. This is obviously exhilarating, but it is also blasphemous and unsettling. An early source is the temptation of Jesus, who, after being baptized by John, goes into the wilderness where he is threatened by the devil. Among the trials he undergoes is that of the potential for total earthly power; and the devil takes him to the highest point of the world and shows it all to him before Jesus rejects it. Peaks are for falling off of, as well as for gaining perspective.

In later medieval literature, airborne journeys of a more lighthearted sort take to the skies, the hero drawn by geese or lifted up by air currents, or in a dream. Yet when Dante leaps off the earth and moves into the higher realms in *The Divine Comedy*, it is clear that this is unsettling. A later appearance of the earth from space, in Milton s *Paradise Lost* is ominously seen from the perspective of Satan, newly released from the prison of Hell, and ying around the universe looking for a new home away from home.

This Godlike stance, epitomized by the image of the earth from space, is, as literature suggests, powerful. The arrival of global maps and globes in the 16th century was part of the great surge of Western power that began with the age of discovery and has only intensi ed its reach in the age of the Internet. The ability to hold the world in our minds is the forerunner of the ability to hold it in our hands, and *vice versa*.

Today we continue the same process. Again, it was McLuhan who said that in the age of satellites there is no more wilderness. There is hardly anywhere on the earth where one can hide, for example, from cruise missiles that, guided by satellite, can drop on a designated square meter of ground. The Hubble telescope, of which we have heard so much, has sister scopes whose magnifying powers are turned in to gaze on the details of our lives. It is clear that the American Armed Forces believe that the next high ground for military superiority is the satellite surround. Evidence of this could be found during the Gulf War in 1991, when the first action of the Allied powers was to cut off access to international satellite pictures for the Iraqis, the result being that they were forced to fight essentially blind.

Another related element, contributing to the figure-ground reversal, is the gradual increase in humankind's influence over the physical systems of the planet. Other volumes in this encyclopedia highlight this encroachment, where human beings are creaming off and rearranging vast parts of the earth's primary production for our own purposes. Certainly human beings have affected large chunks of the earth's natural terrain before, but the ability to affect whole physical and chemical cycles of the planet surely represents a watershed (airshed, earthshed) in earth's history. Our fingers and fingerprints are now and will continue to be all over the genetic future of a myriad species. The ground plans in which they will figure are human influenced, human inspired.

THE HUMAN DIMENSION OF GLOBAL CHANGE

A volume of an encyclopedia on global environmental change that deals with the human dimension of global change is thus necessarily confronted with fundamental issues, of which the most fundamental remains: what does it mean to be human?

Tightly related questions, but more obviously tractable (though one need not get too carried away with tractability) are: what is the relationship between human meanings and goals, and the tools, such as science and technology, that we are using to make those meanings and meet those goals? What does the humanizing of the planet mean? Is it manageable, possible, agreeable, delightful, horrifying?

In 1996, the philosopher Thomas Nagel published a book called *The View From Nowhere* (Nagel, 1996), which describes in detail the stance which modern science wishes to take. Another philosopher, Hilary Putnam, actually refers to this stance as *God's Eye View* (Putnam, 1995). The earliest intimation of that stance may be Archimedes' famous remark, that if he were allowed to place his lever anywhere, he could move the world. This hints at the later stance, which is that the earth is not a privileged vantage point, and in fact there is no privileged vantage point. Science seeks (or sought) to eliminate all the effects and influences of special locations, investigators, etc., in favor of universal laws. This has been undermined – at least at the popular level – somewhat by the paradoxes of the observer and the observed at the quantum mechanical level of science; but the ethos remains intact.

The social critics (Max Horkheimer, Theodore Adorno, and others) of the Frankfurt School in the 1930s were among the first to worry about the possibility that the search for this particular kind of enlightenment – what we might call the universe lying spreadeagled on the dissecting table – was not itself neutral. They suggested that the rhetoric of expanding human knowledge was harnessed to a very powerful unexamined Puritanical drive towards a pure objective truth, in the face of which ordinary human experience would wither and die. They also suggested that there is a complex connection between what we could call objectivity and the treating of the world as if it were an object.

The experience of the Nazi years (which, among other things, led to the fleeing of the Frankfurt School to America), and the arrival of the atomic bomb and the Cold War, substantially undermined the claims of what has been called the *Enlightenment Project* or the *Age of Modernity* (*see* Enlightenment Project, Volume 5). The claim that reason and science could improve the lot of humankind became juxtaposed against the claim that it could also destroy humankind. It was not enough to say that evil is always possible, and reason and science could be misused; critics began to question if there was something unreasonable in the very framework of modernity.

The famous (or notorious) outcome of this was the arrival of *postmodernism* whose antagonism to modernity was described succinctly by the French philosopher, Jean-Francois Lyotard, as a suspicion of all grand narratives (Lyotard, 1984). These grand narratives included the spread of reason over the world, universal human rights, the redescription of the natural world in scientific terms, the increasing recasting of natural resources by technology, and globalization.

It is worth pointing out in this context that virtually all the international organizations and initiatives represented in this encyclopedia subscribe to the grand narrative tradition, simply because they were either generated in the 19th century explosion of scientific networks and societies; or in the post-World War II heyday of the United Nations.

Postmodernist thought is aligned to what is sometimes called *postnormal science* (*see* Post-normal Science, Volume 5), which is an expression of a similar form of skepticism about the rhetoric of neutral science. Postnormal science interrogates this rhetoric, but also examines the work practices of scientists, their worldview assumptions,

the agenda-setting devices (e.g., why are scientists looking at this particular part of the world and not another?) and their communities.

Rayner and Malone, in their overview essay for this volume, describe in detail a number of the tensions between the universalizing tendencies in the physical sciences and the less powerful (but often similarly universalizing) counter-strategies in the social sciences and the humanities (*see* Social Science and Global Environmental Change, Volume 5).

To report on some personal experience, it was while I was involved with the project to establish the International Human Dimensions of Global Change Programme in the late 1980s that the project leaders found themselves practically, as well as theoretically, faced with the implications of the split between the worldviews of the physical sciences and those of the social sciences and humanities. Roughly speaking, however, the main dividing line is not between the sciences and the social sciences; in fact, the dividing line is within the social sciences, between those social sciences whose practitioners see themselves engaged in some form of quasi-physical science enterprise, and those that see themselves engaged in (what for want of a better phrase) what could be called meanings and frameworks.

The practical implications of this were (and are) that models of global change generated by physical science institutions treat human beings as if they were another element in a diagram of boxes with wires. The output of certain social sciences – parts of physical geography, anthropology, demography, etc. – can fit reasonably congenially into these wiring diagrams. Social behavior, communications, politics, and so on, are marginalized or added on as rhetorical flourishes in the form of feedback loops. On the other side of the dividing line, other social scientists and humanists assault the whole enterprise, i.e., that it is a symptom of the global situation we are in (the rationalized objectification of the world) and not a solution. This kind of standoff has complicated efforts to create international research programs that could usefully integrate the natural and the social sciences.

THE RISE OF THE ENVIRONMENTAL MOVEMENT

Traditionally, histories of the environmental movement (especially in North America) begin with conservation efforts and egregious examples of early industrial pollution, and trace its evolution through the creation of national parks, the upsurge of interest in the 1960s through Rachel Carson's (1963) *Silent Spring*, and the arrival of environmental legislation in the 1970s and 1980s. This fits in with a progressive model of a set of problems, problems identified, an orderly political response, and so on.

In light of the above, however, I believe that a more powerful history links environmentalism to some central concerns of the Romantic movement in the late 18th century about the fundamental relationship between human beings and Nature, in particular the nature of Human Nature in a world increasingly subjected to human influence. Moreover, the most compelling parts of works like Rachel Carson's are not about this or that pesticide, but about threats to what could be called the fabric of life, i.e., threats.

As discussed more fully in Art and the Environment, Volume 5 and Literature and the Environment, Volume 5, the Romantic movement was driven by a desire to overturn and reconstruct what the artists, poets, and advanced political thinkers of the time believed to be an outmoded, and indeed repressively dead, cosmology. This cosmology – patriarchal, hierarchical, and rational (in the sense of a machine-like logic) – was increasingly being used to cage and tie down new expressions of human well-being. Because it was outmoded, it was like a snake skin or an exoskeleton that was too small – externally applied, not internally generated.

Among the metaphors applied to this cosmology were the mask and the machine. The mask – ultimately derived from the critique of civilization fomented by the French philosopher Jean-Jacques Rousseau – symbolized the hypocrisy and public demeanor of ancient powers, as well as the emerging urban populace alienated from those powers, and from each other. The opposite of the mask was the Man of transparent virtue, and part of the increasingly vicious theatre of the French Revolution (1789–1795) was the relentless stripping off of, first, the masks of rank and station, and then, the stripping off of the masks of personal life down to the bone.

The machine came to symbolize both the newfound energy that could stimulate a new world of wealth creation, but also the turning of people into creatures of the machine. The early Industrial Revolution appeared to be like the maw of some great creature, driving rural people away from their farms and villages, and into hideously polluted factories where they worked inexorably, hour after hour, day after day. The fact that incomes and life-expectancies began their also inexorable rise upward, and the uncomfortable fact that supposedly happy rural life before industrialization would not bear close examination, had less impact than the immediately obvious relationship between the dismantling of a seemingly organic world and the interim chaos that ensued. Capitalism turned everything into a resource for exploitation.

To the Romantic writers and poets – William Blake, Henry David Thoreau, Friedrich Schiller – Nature becomes many things in the wake of the extraordinary eruption of industrialization into the world. Most obviously, it becomes a refuge from the ills of modernizing life, which is one of the sources for the conservation and parks movement. Less

obviously, it becomes a kind of comrade, or surrogate-like human beings, Nature is under threat. Some social ecologists see this as a central thread in modern environmental politics. Another aspect of Nature is its presence in the world as an alternative system or structure to that being raised by human ingenuity. The fact that it has a creative autonomy of its own, and is supremely inventive and complex in its own right, challenges the arrogance of technological man.

For some people, Nature becomes an entry into a spiritual realm, and interestingly enough, it is critics like John Ruskin and poets like Gerard Manley Hopkins, who fasten on the intricate details of Nature's design as the entry way; and not, as earlier, on Nature as a grand symbol. In this, one can see the growing influence of natural science, initially on the Victorians, and later in the synthesis that would become the environmental movement.

However important this history may be, the real history of the modern environmental movement, to my mind, begins with Hiroshima, and is followed very quickly by the revelations about the Nazi death camps. It was these events that concentrated a growing unease about the capacity of human beings to destroy themselves, and perhaps the world around them. We can begin to see a change of tone in environmental writing (still nature writing) in the 1950s (Rachel Carson's first successes appeared then); and it is little remembered that the writing of *Silent Spring* coincided with the first great international grassroots environmental action to stop the above-ground testing of nuclear weapons. Part of what fueled that successful political activity was the unsettling prospect of having what had hitherto been seen as a most benign metaphor, mother's milk, poisoned; and poisoned at a distance. Carson's pesticide poisoning at a distance amplified and echoed that other concern.

These threats to the fabric of life – to the possibility that rents in the fabric might be caused by human ignorance – were (in my mind) the spiritual dynamic behind the environmental movement. In the 1970s, it went somewhat underground, as the nuts and bolts of environmental management took center stage. I say somewhat, because it could always be found, for example, in the concern for the potential loss of endangered species. But at various points, it has returned in full force to revitalize environmental activism. In the late 1980s, the prospect of a rent in the fabric of the sky – the ozone hole – regenerated a whole array of environmental actors and institutions. Similarly, if more slowly, the prospect of irreversible climate change has begun to concentrate the minds of many activists.

The current activist environmental movement – in part because of the political dynamic associated with global environmental issues (symbolized by the Rio Earth Summit in 1992) – appears to be more inclined to return to the question of the role of modern industrialization and its technological imperatives. Globalization as a phenomenon is becoming widely seen as the interim step towards the ultimate transformation of the Earth into a human enterprise. Since multinational corporations and the rhetoric of global free trade are propelling much of this, environmentalists have begun making more and more sophisticated critiques of mainstream economics.

ECONOMICS AND GLOBAL CHANGE

A central fact about contemporary life is that globalization is rooted in a range of theories and practices that derive from economics. This is contested by those who would prioritize political structures such as the nation state, or strategies of imperialism, or even social or legal or philosophical forces deriving from the spread of Western models of individualism across the globe. Nevertheless, if one sees in economics a complex web of all of these, ultimately justified by an almost religious belief in the virtues of the market system whose every whim needs to be tended, then its strength as a belief system for at least the elites in the West and their acolytes can be hardly disputed.

Looked at historically, one of the most important aspects of modern economic theory and practice is that it represents the first and most successful form of systems theory as applied to social systems. This is part of its power; and as will be discussed shortly, is essential to its ethical dimension.

Economics, like the modern social sciences generally, originated in various attempts to do for the social world what Newton had done for the physical world, that is, find a set of simple laws that would explain and underpin a complex set of phenomena. The Scottish and French enlightenment thinkers were haunted by this possibility; and among the first, and among the most controversial, of the attempts at fulfilling this vision, was the provocative work by the British writer (actually born in the Netherlands) Bernard de Mandeville (1714), *The Fable of the Bees*.

In the *Fable*, Mandeville breaks with the moral and spiritual traditions of the world, and argues that private vice should be encouraged, because spending on luxury goods, prostitutes, race horses, etc., generates many public benefits through employment. Here in embryo is the central pillar of neo-conservative thought in our own day. What made Mandeville's work so influential was the linking of local phenomena to global phenomena with opposite characteristics. The social thinker is either to trace the impacts of the local phenomena to their global outcomes from a neutral stance (the view from nowhere), or, even more influentially, realize that the local phenomena can be described quite differently when seen from a higher vantage point. Moreover, the allocation of goods and bads can be dissociated from any notion that the Deity bestows them.

This powerful systems model sat in the armory of the early economists, awaiting further refinement. In the meantime, these economists, including the physiocrats in France (mid-18th century), and the classical economists – Smith, Malthus, and Ricardo (1750s–1830s) began the task of tracing out the transformation of Nature, people, and investments/finance/invention/buildings into land, labor, and capital. With each tracing, the relationship between economic activity and the physical world diminished, and was replaced by the intricate workings of the market.

Perhaps one of the hardest things for a non-economist (actually, for most economists as well) to recognize is that economics does not deal with the physical world. This became particularly true after the first neo-classical revolution in the mid 19th century associated with J S Mill, William Stanley Jevons, and others; and continued into the 20th century (with variations). Instead, it deals with the subjective desires of individuals, as expressed through competing demands signaled by prices, and adjudicated by a market whose other component is supplies, drawn to the market by the price signals, which are re-determined by the interaction of demand and supply. More modern theory speaks of rational allocation of scarce resources, but again, the real world only appears in the tenuous link supplied by an assumption of scarcity.

As a result, in contemporary economic theory and practice, the natural world enters only as it influences prices. If there is no price on koala bears, i.e., no one wants or values them, they do not exist. To deal with this problem (and others) is the domain of the emerging discipline of *environmental economics* (*see* Economics and Global Environmental Change, Volume 5; Environmental Economics, Volume 5). Among the tasks of this discipline are: to find shadow values for environmental goods that do not show up in market prices; to develop surrogate measures for environmental value; subject environmental goods that currently have no legal or proprietary owners to privatization; and generally to internalize within the market what are called externalities.

This environmental economics approach can be compared to the approach of *ecological economics* (*see* Ecological Economics, Volume 5). Ecological economics repudiates the entire superstructure of modern economic thought polemically described above; and rather proposes to return to the kinds of approaches characteristic of the French physiocrats, and others. Specifically, ecological economists focus on a variety of physical aspects of human use of the resources of the earth to sustain life.

This approach considers, among other things, what are the physical requirements to feed human beings, sustain agriculture year by year, harness energy, process mineral and chemical feed stocks for industrial production, and assess the ability of the earth systems to handle the waste products of our activities. There are a number of problems with this alternative approach, of which perhaps the least serious is that there is no consistent set of measuring tools and common language among the variety of competing versions of ecological economists. Economists such as Daly and Cobb (1993) have been working for some years on alternative measures of gross national product and monitoring systems other than price mechanisms.

More serious is the difficulty of internalizing the other economic approach, which, whatever one's views, does currently drive the international economic system. The price mechanism is ingrained into the workings of virtually all decision-making processes, and acts as a fundamental purveyor of information in current society.

Most serious of all are the intangible benefits of the standard economic approach as a moral and ethical system. A lot that could be said about this, but I would like to discuss two issues on this topic briefly.

The first, returning to where this section began, with Mandeville's vision, is that one benefit of the standard economic approach is it allows people to repudiate compassion; or to put it another way, to feel good about being self-interested. Obviously, technical economists (welfare economists and the like) would protest this kind of remark; but I am speaking here about a general cultural influence. Part of the neo-conservative agenda that has been so powerful in recent years is an argument that people who act compassionately towards the poor are acting inefficiently. Indeed, they are, in fact, standing in the way of the improvement in the lot of the poor that would be produced by allowing the market to operate more efficiently. Everything that interferes with the market – social safety nets, public insurance, public health care – is by definition inefficient, and thus drags down the potential for the economy. It is the hard-hearted economic realist who is really the benefactor of the poor.

A related moral aspect of this system is that it is supposedly morally neutral. Since it derives from individual wants, whatever those wants may be, it says nothing about whether some wants are better than others. This classic attack on the amorality of utilitarians is, in fact, a source of pride to economic liberals; though they tend, paradoxically, to bemoan the loss of conservative values like family life that their system is happily destroying.

The second (and related) issue is that not only are the origins of those wants now subjected to gross manipulation by corporate advertising; but at the heart of the system – for a variety of obscure historical and sociological reasons – one can find a powerful model of the human which is being propounded. In this view, contrary to virtually all ethical and spiritual traditions in the world, human beings are fundamentally self-interested, and possessed of infinite desires which can never be fully satisfied; and that rather than try and eliminate these desires, they should be encouraged. Similarly, consumption patterns are now fully seen

as expressions of one s personality. The rise of fashion, the need for status differentiation in an increasingly homogenized world, and the transformation of the citizen into a consumer rather than a producer, have all contributed to the seemingly unlimited growth in consumption in developed countries.

As has been stated by numerous writers, the spread of this in nitely desiring consumer model to the developing world is doubly troubling. Analysts of the so-called ecological footprint the amount of extra resources needed to service developed country growth already argue that we will need three or four new earths to service a world population carrying around in its heads the lifestyles of North America.

MANAGING THE EARTH?

In the meantime, if we look at actual production patterns on the earth, designed to service this population, we can see very important trends in the technical means being deployed in the reordering of the planet for human ends.

Approximately 40% of land-based net primary production (that is, plant-generated material), and over 25% of marine net primary production is now being diverted or rearranged for human use. Much of this is taken by the domestication of livestock, the harvesting of higher animals and sh, and agriculture. This is completely unprecedented.

What is happening around the world is essentially a restructuring of ecosystems in order to maximize production, primarily through creating systems according to at least ve rules.

1. organize for harvesting of natural products at the peak of productivity. The productive peak for many natural ecosystems is at the end of the juvenile phase;
2. concentrate for easier planting/monitoring/harvesting;
3. replace/buffer/enhance ecosystem processes;
4. add value at, and beyond, the farm gate;
5. suppress extra-market externalities.

Here are two examples of some of these rules in operation.

One example is sh farming. Worldwide natural sh stocks are in deep trouble, with something like 34% of sh species threatened with extinction. One solution is intensi ed sh farming in captive tanks. It is calculated that one in four sh that reaches market today comes from a sh farm. The problem is that a sh farm is a monoculture, and all the supporting services have to be brought in from the outside sh farmers are now competing with poultry and pork producers for grain and protein meal supplements, such as soybeans. These farms also produce high levels of waste, they suffer from outbreaks of disease, and forms of chemical pollution.

This is exactly the same situation as in monocultural agriculture. The problem with the average managed eld is that it is trying to turn back into a meadow or a forest. Today s agriculture requires immense inputs of pesticides, herbicides, gasoline for tractors, etc., in order to create an outdoor factory. The most recent types of high-yield corn require that all the microorganisms in the surrounding soil be killed off. This kind of agriculture is becoming more and more expensive, in part because the weeds and the bugs, i.e., the ecological precursors of a return to a meadow, are being chemically and biologically selected for resistance to whatever the latest crop strategy is.

A similar process has been underway for many years in forestry.

Rule one means that crops, trees and sh must be harvested in the juvenile phase and before they level off as adults, at which point it is economically inef cient for them to just be sitting around. This means that there are no older age cohorts left, which also means that the regeneration through decomposition disappears. If you have no old trees, you have no carbon being returned to the soil, no place for insects and other decomposers to nestle in, and so on.

Rule two, which is essential to agriculture and mining, concentration and simpli cation means, in living systems, that there is no diversity, and that makes the system vulnerable to catastrophic attack by virus, bug, plague, weed, or genetic mutation or rapid climate change.

Rule three means that all the processes of energy cycling, nutrient creation, removal of wastes, and so on, have to be imported into the simpli ed ecosystem from outside. In order to do that, energy is required, in addition to resources from somewhere else, which means that some other high-quality ecosystem (or stock of fossil fuels) is being tapped into and degraded. It is based on the assumption that there is always somewhere else left to go to get the supplies needed to feed the arti cial ecosystem.

Rule four, which is well known to farmers, means that the primary production of the ecosystem is virtually worthless it is essentially a host for what will happen next. A potato is worth nothing compared to barbacue potato chips. That is where the money is. A particularly ironic example of this came recently during the collapse of the East Coast Fisheries. One of the largest Canadian sh processors, which was instrumental in destroying the sheries, and has now moved on to the South China Sea, turned its attention to sh species that it had never bothered to sh for before, ground them up, and then rebuilt them as sh ngers in the shapes of star sh and seahorses, and was selling them at great pro t to Colonel Sanders and other fast food outlets in the US.

Rule ve, which is perhaps the most important, is an attempt to eliminate the environment as a part of economic costs of production and marketing. The best example of this the one that is causing global warming is the suppression of geography to enable the smooth running of free trade. Free trade, or if you like, international free markets,

depend upon the ability to ship goods to any place in the world without friction or at very low friction. Distance from one place to another is no longer to be a consideration. The reason this can be done – more or less – is the low price of fossil fuels. Cheap transportation means that California produce can compete, for example, with Ontario produce virtually year round.

The reason why fossil fuels are cheap is that the earth subsidizes carbon dioxide and other emissions. The ecological costs of these fuels are not internalized: if they were, free trade would die tomorrow. As it is, we are awaiting global warming.

POLITICS AND THE WANING OF TRUST

Global change presents us with a number of issues – climate change, the extinction of species, and the disposal of nuclear waste – that require us to think in the very long term, over hundreds and maybe thousands of years. One reason why people have turned to ethicists and spiritual leaders to articulate new environmental ethics, and others to new environmental politics, is a sense that this kind of task is beyond the normal reach of our current ways of thinking, and requires something different. An important recent set of discussions has revolved around the concept of trust', what Niklas Luhmann (1979) referred to as one of the ways in which we try personally or politically to cope with uncertainty over time – through vowing to be trustworthy, trusting others, entrusting ourselves to them, accepting trusteeships, etc.

It is a little known fact that one central element of Western representative democracy grew out of a debate (actually a war) about whom to trust, how to entrust, and for how long, with decision-making power. The British Crown sought (under Charles I in the 1630s) to retain the power of taxation without representation' (to use an anachronistic phrase) by claiming emergency powers to deal with supposed external threats to the nation. Charles wanted to build ships – and get some extra pocket money. The ensuing Ship Money' struggle and the Civil War that followed led the next generation of politicians and philosophers to work on the conditions for declaring emergency powers'. The political theory and practice that evolved further clarified the nature of the entrusting relationship between citizen and government as a whole, first in Britain, and then elsewhere. As outlined by John Locke (among others), the contract between government and citizen became seen to be – in part – based on the entrusting of decision-making power to a representative body for a period of time subject to eventual review by the electorate.

This historical interlude is important because it points up the fact that pivotal to the entrusting of decision-making to representatives was the idea of time: that the government could not always refer back to citizens for every decision and indeed it was important not to do so, when issues were so complex or in need of such speedy resolution that constant referral back through simple referenda, etc., was unlikely to produce a better outcome than focused debates by representatives, who were free to be convinced by their opponents of the wisdom or folly of their position. Obviously, the legitimacy of this entire setup – which is not exactly an accurate description of how things turned out in representative institutions around the world – is now under great stress. Not only are even democratic governments currently more executive and less legislative, for a variety of complex reasons, than the ideal supposes, but citizens are now more educated and unwilling to defer to elected representatives than they may have been; and there has also been an evolution towards at least the rhetoric of more active participation by unelected individuals in the decision-making process. This has been aided and abetted by new communications technologies like the Internet.

Representative democracies have not been able to come up with appropriate ways to respond to this desire for increased participation, and what is in place is somewhat haphazard and flawed; and this has affected the ability of governments and institutions to respond to underlying public anxiety about fundamental environmental questions, and the whole range of longer term concerns in any other way than by bureaucracy. On the one hand, there is an increasing use of single issue referenda, which are subject to manipulation and distortion, but give people a voice' – or at least a gesture. On the other hand, there has arisen over the past while, a whole array of quasi-participatory committees, NGOs, and investigative bodies like the Environmental Assessment Review Panels and Royal Commissions. These usually report to Cabinet or Parliament, and they are supposed to engender trust that the pros and cons of the longer term are being weighed and assessed properly by experts as well as citizens in a deeper way.

These new approaches seem, however, to do little to address some of the more fundamental distrusting going on out there. This is tied to the phenomenon that this section began with referring to: the increasing recognition that decisions being made (or avoided) are extending over larger and larger stretches of time and space, and having effects that are not part of the original mandate' of the political institution making the decision. A classic example of this is the repudiation by outraged citizens of agreements made by town councils, and other local bodies, to site hazardous or radioactive waste facilities – and the casting about for an appropriately legitimate level of government that can make that kind of decision. We didn't elect you to make that kind of decision , is the rallying cry. But democratic countries have not yet found any way of legitimating decisions about where to put high-level nuclear waste for a million years or more, in spite of a multitude of commissions, reports, inquiries, and so on.

Associated with this is an easily awakened mistrust of the elite and professional apparatus of rational management'. Part of the legitimacy of government in our time is in its professed ability to manage any change rationally. The rise of the modern bureaucratic state is based on the premise that utilitarian calculations mixed with expert projection can cope rationally with most of the situations that will arise. So powerful has this rationalization been that on the occasions where symbols of managed rationality have broken down – Three Mile Island, Chernobyl, Vietnam – they have threatened the legitimacy of the entire system of governance. Some philosophers suggest that the special form of rational management associated with the modern world was fatally crippled at Auschwitz. Others, less radically and more hopefully, suggest that people are now less prone to do things just because their government tells them to – like walking unprotected into atom bomb test sites.

Certainly one of the interesting facts of our time is that the citizenry and the environmental radical are often now arrayed on the side of prudential' behavior with regard to proposed future risks; while governments, administrative experts, and corporate elites are essentially risk-prone'. Their trust in their own capacities is undisturbed, in part because they are so locked into short-term forecasting. It is everyone else who finds themselves forced to inject pessimism and uncertainty back into the process; and to mandate the consideration of worst-cases and the very long term.

It is usually said that politicians and the system' are unable to think past the next election, and that this is somehow due to the greediness and short-sightedness of their masters, the electorate. While there is something to this, it is not everything. There are endless examples of the short-sightedness of the expert and the forecaster, whose models accentuate the present and discount the future. The expert can miss the forest for the trees, and the fact that there are no more trees may not show up as a loss on the balance sheet, if the property values can be cashed out elsewhere. Counter to these balance sheets and snazzy models is often something much humbler, more conservative (in the original sense) and simpler: the uneasy sense among the ordinary person that something has gone wrong, and that if we keep going the way we are going, things can only get worse.

Common sense is not all that common, and I do not want to exalt the common man or woman as the paragon of all wisdom. Yet it was the philosopher Karl Popper who made the case that one of the great virtues of democracies was that sometimes the citizenry were the only people who could break the stranglehold of some grand stupidity that was mesmerizing the all-knowing powers-that-be.

In this category – a feeling for the overall wisdom of what is going on – I would put the feel for the longer term. Perhaps it is just living a life, or the fact of having children, or being the latest in a long story line that stretches back in time. Traditional cultures developed methods of thinking ethically over the longer term, and this is one of the reasons they were, in large part, sustainable.

One crucial element of thinking about the longer term, facing the future, is facing towards the past. To develop a sense of responsibility over time, traditional cultures have relied upon ancestor worship and long-standing rituals. Thinking like an ancestor' is a familiar strategy: respect for one's living elders, and one's long-dead ancestors provides one with a compass for what one should do now – and also what one should do to keep faith with those people into the distant future. If we continue to do what has worked well in the past, then the future is secured.

This kind of approach to risks is, of course, threatened when something new and unique comes along like nuclear power or PCBs. It is also threatened across the board by a society that, in cutting itself free from the cycles of the seasons and of ecosystems, no longer values the wisdom that comes from having been through the cycles many times. We value the innovations of the young; the old we discard and stuff away in old people's homes, their wisdom useless to us.

Another attempt to extend ethical or moral sentiments across time is associated with those cultures that believe in reincarnation. In Tibetan Buddhism, for example, people are encouraged to treat animals properly because they were one's mother in a previous lifetime (and there is always the possibility that becoming an animal is one's own fate). All beings are thereby linked across time by familial ties which engender respectful treatment.

A third traditional approach, perhaps more familiar in what derives from a Christian culture, is to see things sub specie aeternitas', that is, under the gaze of eternity or God. One's own actions, or the actions of others, can then be seen as part of a larger pattern or narrative. Related to this is the idea of the moral absolute or imperative.

Moral absolutes are associated particularly with the work of the German philosopher Immanuel Kant, who expanded on a rich tradition going back at least to Plato. Moral absolutes (such as thou shalt not kill) are supposedly free from historical or geographical specificity. This means that they stretch throughout time: it will always be wrong to do this, whatever the change in the situation. This kind of absolute moral stance can also reflect a belief that the consequences of one's acts are not important – that one should do the right thing, whatever happens.

Long-term issues tend to foster absolutist positions, partly because the sense of a problem stretching out for a long time comes to equal forever'.

A moral and ethical consideration which seems to be new to our time is the idea of not doing anything to undermine the sustainability of life on earth – we might say to

fray fatally the fabric of things'. This new consideration derives from our considerably enhanced power to affect earth's biological and physical systems, which we have always previously been able to take for granted. Seedtime and harvest shall not fail was God's pledge to Noah after the flood. Hans Jonas (in his 1984 book, *The Imperative of Responsibility*), referred to this as the ontological imperative'. Essentially, human beings are now subject to a physical imperative which is simultaneously a moral imperative: the basic rule is not to threaten the continuation of life systems on the Earth.

Many people already subscribe to the most drastic version of this imperative, which shows up as a revulsion towards extinction of species. It is a widely shared belief (supported by surveys) that since extinction is forever', endangered species require special treatment. It is not clear how old this sentiment is: that is, previous generations likely lacked the ecological knowledge (as do we still in many cases) to predict the ultimate consequences of the wanton destruction of individual members of species to the point where reproduction became impossible. A widespread assumption was that there were always more Xs where X came from.

So there are perhaps gestures towards some new moral imperatives based on a new awareness of our capacities for destruction; and the broadening of a new moral horizon, based on our anxiety that there should continue to be dawns over the old horizon for as far as we can envision the future.

THREATS TO THE INTIMATE

The concerns over the threats to the fabric of life implied by certain facets of global environmental change have begun to return – as in Rachel Carson's day – not just to the local impacts and expressions of global change, but to what could be called *threats to the intimate*.

An ordinary threat to the intimate is the issue of endocrine disruptors – the possible disruption of the human endocrine system by some combination of the bath of chemicals within which people in modern societies operate. But a more extraordinary threat is posed by the prospect of genetic engineering, including ultimately the engineering of human life.

Because genetic engineering obviously involves getting down to the guts of things, it is obvious that it provokes great fear and concern, some of which may well be technically unjustified. However, the underlying fear and concern is connected to the themes I have been discussing. We are on the verge of having human beings determine what the nature of the human is going to be. This presents us with a kind of moral vertigo, rather like sitting in a barber's chair looking into an infinite set of retreating mirrors.

For environmentalists, the spread of uncontrolled bioengineering is certainly of concern; but it would be foolish to deny that there is also a visceral reaction to the whole enterprise, controlled or uncontrolled. This reaction is part of the often-unnoticed, deeply conservative strain in environmentalism; and it is in part a reaction to the often-unnoticed, deeply radical strain in the modern economic enterprise. The fact that critics of protestors against bioengineering use a variety of economic and utilitarian arguments reinforces the belief that the experiment with the natural world is part and parcel of the growing push to experiment on human beings, who are, after all, themselves part of the natural world. So one is forced back onto some very old and familiar debates: is there more to human beings than their material parts, than their genetic makeup, etc? If there is, what is its basis?

At the end, we return to the image of the earth from space. The French existentialist philosopher, Gabriel Marcel once made a distinction between what he called *problems* and *mysteries*. Problems – from the Greek word *problemata*, to throw in front of – are things that one can stand back from and attempt to solve from the outside. They can be cracked open, like nutshells. Mysteries, on the other hand, are things or questions that cannot be handled in this way: the more you try and solve a mystery, the more it pulls you in, the more it involves you personally in its solution. Murder mysteries are really problem novels – there is a problem to be solved, and the detective moves on to the next one. True mysteries simply expand, deepen, the more you probe into them; and usually bring into question your own deepest beliefs and concerns.

The earth from space looks like a problem – How to manage planet Earth? is a familiar phrase, brought into currency by a special issue of *Scientific American* in 1989 – but it is in fact a mystery, since it encompasses us, and calls into question the deepest wellsprings of our current World Project – the transformation of the ground out of which we came into something more akin to our own wishes and desires.

REFERENCES

Daly, H and Cobb, J (1993) *For the Common Good*, Beacon Press, Boston, MA.

Jean-Francois Lyotard (1984) *The Postmodern Condition*, A Report on Knowledge, Manchester University Press, Manchester, xxiii–xxv/3–11/37–41.

Jonas, H (1984) *The Imperative of Responsibility*, University of Chicago Press, Chicago, IL.

Luhmann, N (1979) *Trust and Power*, Wiley, Chichester, UK.

Nagel, T (1996) *The View From Nowhere*, Oxford University Press, New York.

Putnam, H (1995) Comments and Replies, in *Reading Putnam*, eds P Clark and B Hale, Basil Blackwell, Oxford.

The Changing Human–Nature Relationships (HNR) in the Context of Global Environmental Change

BERNHARD GLAESER

Social Science Research Center, Berlin, Germany

In his 'Metamorphoses or Transformations', completed in the year 8 AD, the great Roman poet, P Ovidius Naso tells the story of changes in nature, mythology, and human history. In particular, the philosophy lecture in Book 15 gives an amazing account of natural and social change that include such themes as global change (GC), the Gaia hypothesis, environmental destruction, and human and political development. Ovid's depiction demonstrates the broad scope of natural and social change; it is the starting point for this attempt to discuss a variety of perspectives concerning the changing relationship between humans and nature in the context of global environmental change (GEC).

The overall objective of this introductory essay is to convey a broad view of social and cultural aspects of GEC. It represents a Western, social science perspective, and re ects on today's discourses as in uenced or characterized by the turn from the second to the third millennium. It is concluded that the international scienti c community can and should play a vital role in nding solutions to sustain the environmental conditions for the sake of global livelihood, including social justice.

NATURAL VERSUS SOCIAL CHANGE

Relationships between humankind and nature are first of all physical and material: they determine the very substance of and conditions for human survival. But, at the same time, humans interpret and construct relationships between themselves and nature. It is this complex interplay that constitutes and complicates consideration of the social and cultural aspects of GEC. Shifting perceptions relate to physical, economic, social, and cultural changes over time. The changing relationship between humans and nature is an open issue; the cultural and social sciences attempt to interpret it in terms of the following questions: What did the human nature relationship (HNR) look like before and after the start of GEC? What is the specific difference? When did GEC occur and with what consequences for the HNR? We do not really know the answers to these questions for several reasons.

- Natural as well as social systems are determined by perpetual change . When change ceases, the system stops functioning and perishes.
- There is disagreement as to the novelty of GC, because there is some dispute over the meaning of the term globalization . One view is that globalization indicates Western modernity; that it was initiated by Columbus discovering and exploiting the New World, followed, in turn, by the rise of capitalism and the age of technology. This view is based upon the idea that the Renaissance combined wisdom and the power of Mediterranean antiquity with northern European modernity, which really means globalization on a temporal, historical scale. We may note that the dominant view of globalization is entirely Eurocentric.
- Opinions differ as to the values attached to nature and the nature–human relation. These values depend on the degree of an individual's exposure to nature, the needs he or she attaches to nature, the capacity for self-reflection contained in these various concepts and the willingness to abstract from nature. Individuals, ethnic groups and cultural systems construct their own concepts of nature on which they rely and to which they relate. Stakeholders shape the respective HNRs.

These reasons suggest the following point of departure and focus. Since natural and social systems are characterized by perpetual, uninterrupted change, we shall focus on what appears to be its present mode, at the turn of the 21st century, and its corresponding interpretations, concepts and constructs. Humans and nature: which is central, which is peripheral? The age-old question is again raised. It has ethical and political consequences; these are labeled respectively as the *anthropocentric* and, conversely, the *bio-ecocentric* approach. What is the dominant concept of nature? Nature may mean cosmos, the origin of life, an object for philosophical contemplation or artistic production – not to speak of its practical form as a natural resource serving as a vital factor of economic production. Are there different cultural outlooks, such as a typically oriental or typically occidental perspective? Ex oriente lux, enlightenment from the Orient, was the belief or romantic fashion at different ages, including contemporary New Age culture.

What is the importance attached to the relationship-to-nature topic and who are the winners and the losers of GEC? In Europe, during the 1990s, environmental issues lost a lot of appeal to issues of job security and the labor market; whereas in the rural developing world, environmental degradation and loss of soil fertility are often synonymous in that they determine income, livelihood and survival chances – those of rural women more than anybody else (Kiely and Marfleet, 1998; Klingebiel and Randeria, 1998).

Finally, how does the sustainability concept serve in this context and at this point: What is its social dimension and its political outlook? Holistic, interdisciplinary or transdisciplinary approaches have won recognition. Consequences are felt by the research community who have become stakeholders in this field. The mode of knowledge production will change and this implies yet another relationship: that between the natural and the social sciences as well as that between the sciences and the participatory involvement of other (non-scientific) stakeholders affected by GEC. In the political arena, global environmental politics is becoming a strategic issue as conflicts over water have already appeared; environmental security (Brennan, 1999) follows military (Aspen Institute, 1999) and food security (Schulz, 1999). Structural policy is needed to give GC a proper shape (Schellnhuber and Pilardeaux, 1999).

NATURE: IS IT PERIPHERAL TO HUMANS?

As we use ever bigger telescopes to observe the universe, such amplification technology also enables us to literally view the past. We are not too far from witnessing the birth of the universe: Empirical evidence will tell us which hypotheses and theories are to be falsified. Light travels at a velocity of approximately 300 000 $km\,s^{-1}$, which means, for instance, that it takes sunlight a full eight minutes to reach the earth. These simple and well-known physical facts have theoretically interesting implications as scales increase. Imagine that we can see and photograph a galaxy at a distance of one billion light years away! Since the light we've thus observed has been traveling for so long a time, namely one billion years, we are, from a human standpoint, literally looking into the remote past without even knowing whether what we observe still exists.

The Copernican revolution revealed the fact that the Earth was not the center of the universe, or even the center of the solar system. Our earth is simply a minute and by no means unique particle somewhere at the periphery of our galaxy, the Milky Way , and also of the presently known universe. This revelation could not be publicized freely at the time, because it hurt the interests of an important 16th century global player and stakeholder, namely, the Catholic Church. Throughout most of Europe, the Church held a monopoly on the interpretation of not only all metaphysical matters, but also the view of the world and all worldly matters. The Church of Rome represented God the Almighty on Earth, so it claimed, as an institutionalized human trustee. It intended to expand the domain of its monopolistic trust around the globe, along with the worldly powers that used fire and sword to subdue the rest of humanity. So this interpretative and explanatory competence meant power, in a subtle way perhaps. The promulgation of the new heliocentric theory by the Polish astronomer Nicolaus Copernicus (1473–1543) and the rise of science as a whole thus challenged the authority of the Church and its cultural imperialism.

Today, half a millennium later, we witness a second Copernican Revolution (Schellnhuber, 1999) which calls our attention back to planet Earth. The Earth System as a whole is to be analyzed. Understanding it is the basis on which to develop concepts of and for global environmental management. Earth-System analysis (Schellnhuber and Wenzel, 1998) is the holistic and transdisciplinary attempt to model and simulate the *ecosphere*, i.e., the geosphere–biospheric complex, including its climatic history as far back as half a million years ago. This is an endeavor that takes the Gaia hypothesis seriously – the paradigm that interprets the Earth system as a cybernetic whole endowed with a self-regulating capacity (Lovelock, 1979). The basis of this theory is hardcore empirical data that demonstrate that self-regulating biospheric mechanisms have kept the Earth's crust stable and its environment habitable. Biospheric evolution eventually produced humankind.

After four and a half billion years of natural and eventually cultural evolution we have learned two things. First, we are capable of continuously undermining the conditions for our own survival. Global environmental change in the destructive mode can be seen, for instance, in human activities such as oil spills and toxic waste proliferation, perforation of the stratospheric ozone layer (thus exposing us to more UV–B radiation), or nuclear warfare. Second, we have developed – perhaps to a lesser extent – the ethical and managerial tools to protect and safeguard the global environment against forces of destruction like those just mentioned (Parry and Livermore, 1999; Pearce, 1999).

Again, interpretative and explanatory competence has a significant role. Unlike 500 years ago, when enlightenment and the rationality of science freed humankind from religious obedience to strive toward self-determination, it may now be the interpretative dialectics of mythology that reveal the destructive mode of science and technology: Humankind has become simultaneously Shiva and Vishnu, the Hindu gods who represent the concepts for Destruction and Preservation, respectively.

THE MILLENNIAL SHIFT: GNOSTICISM AND THE ENVIRONMENT

Today, at the beginning of the third millennium (AD in our Christianity-centered temporal accounting) environmental management has to cope with GEC. Fears of global destruction and extinction witnessed the earlier transition from the first to the second millennium. According to the doctrines of Gnosticism, salvation could be attained only through the pursuit of spiritual truth and the transcendence of matter. Jesus Christ was considered by Gnostics to be noncorporeal. In keeping with this doctrine was the belief that the beginning of the second millennium would bring about the spiritual age. Gnosticism incorporated an apocalyptic and chiliastic vision of nature's decline and God's ultimate reign over Earth, 1000 years after the birth of Christ. Chiliasm is the doctrine stating that Christ will reign on Earth for 1000 years; ancient gnostic knowledge and belief originated in Mesopotamia, Syria, and Egypt, around the Mediterranean, and were revived by Scotus Eriugena in the 9th century (Voegelin, 1952, 1959).

At present there seems to be a lack of interest in environmental (including GEC) issues, specifically among social scientists. In the following I draw a parallel between the lack of interest in the natural world on the part of many social scientists and the lack of interest in the physical world by Gnostics. My thesis is that some concept-oriented or social-constructivist social scientists are, or behave as if they were, Gnostics. As proof of this, I briefly review some current journals in the field that devote special attention to the recent millennial shift.

Independent of theories that explore such parallelism, the review provides an insight into which topics academics and editors of prominent journals consider of paramount importance during an historically important period. Some historical dimensions and explanations concerning the development of sociology and the motives of its representatives are presented as these relate to the HNR.

Is GEC and/or the changing HNR of interest to social scientists? A few international academic journals focused their last issues in 1999 (or the first issues in 2000) on millennial aspects: the transition from the second to the third millennium, its meaning and relevance for the development of social or interdisciplinary science. Let us look at four of these journals and their topical issue: two interdisciplinary publications and two specialized publications in sociology. I begin with the interdisciplinary publications, *The*

Ecologist (1/2000) and *Universitas* (12/1999), a German-language journal whose readership is mainly among university instructors and students. Both millennium issues concentrate on gnostic themes.

The Ecologist (1/2000) deals with the cosmic covenant re-embedding religion in society, nature and the cosmos . Emphasized is the role that religion can and must play in saving what remains of the natural world , as Edward Goldsmith puts it in his editorial. Knowledge and values that attributed the utmost priority to the preservation of the creation were once propagated by the various religious groupings or cultures. The rediscovery and revival of ecological themes and cosmic theologies appear to be an environmental priority. The theological underpinnings of most religions relate the individual to society, the natural world, and the cosmos. Mainstream science committed the ultimate blasphemy in that Homo scienti cus has dei ed himself ; mainstream religion has lost its way and needs to return to its roots. Noah s ood symbolizes the forces of chaos. Historical storms and oods in Orissa (India) or Vietnam may remind us of this archetype human failure to observe the cosmic covenant, that is, to ful ll our contract to live in harmony with the laws of nature and the cosmic order (Goldsmith, 2000). Religious inspiration and perennial beliefs are to be found among the primal creeds or religions because these derive from the cosmic covenant the universal revelation given to humans (Grif- th, 2000). Tribal stories maintain such wisdom and its ecological message (Wilson, 2000). So did classical mythology and the ecological worldviews of ancient societies such as those of Greek, Chinese, Egyptian, Indian, or Persian antiquity, by using the notion of the path that must be taken to maintain the cosmic order on which human welfare depends (Chaitanya, 2000; Goldsmith, 2000). Related cosmic and ecological insights are quoted for the Islamic and the Judeo Christian traditions, to some extent in u- enced by the ancient civilizations (Barker, 2000; Echlin, 2000; Murray, 2000; Nasr, 2000; Rossi, 2000; Roth, 2000). The cosmos was embedded in the Church until western scienti c thought replaced the term creation with environment , thus separating human from non-human nature. The message of the special issue of *The Ecologist* is that such desecration of the cosmos ought to be reversed in the third millennium.

Universitas (December 1999) entitles its special millennium issue Endzeiten, neue Zeiten? (Final age, new age?) and focuses on topics of transition (Geissler, 1999) between the two millennia. Western industrial societies have eliminated many rites of passage and transition; the symbolism of reworks, for instance, whose original intention was to vanquish evil spirits , has been lost. Instead, the transition from the second to the third millennium has been marked by spectacular events not necessarily of universal importance or interest such as who (in somebody s town) gives birth to the rst millennium baby (Hilgers, 1999). In a similar way, the great issue of the apocalypse has changed. Originally, St. John the Divine pictured the apocalypse in The Book of Revelations as the vision of salvation coming about after a transitory collapse. This transitory collapse is itself frequently referred to as the *apocalypse*. The visionary apocalypse represented the advent of the millennial God s reign on Earth, the New Jerusalem as it was called. In the 20th century, Hitler and Stalin were associated with the advent of the apocalypse. The transitional period during which both leaders were in power was an extremely violent and bloody episode, characterized by massive internment in concentration camps and gulags, and massive human slaughter, in particular genocide. All of this was designed to purify human blood or to convince people to adopt the right doctrine , for the purpose of achieving some perfect millennial Third Reich or ideal post-historic age of communism. Yugoslavia has witnessed a sad revival of apocalyptic nationalism. The cyberspace apocalypse, fortunately somewhat less deadly, has led us into the third millennium. Virtual reality represents the New Jerusalem ; the cybernaut, as the new human , exists independently of bodily needs and achieves immortality as part of the permanent memory of a computer network. St. John s millennial Third Reich becomes the age of knowledge beyond the ages of agriculture and industrialization (Vondung, 1999). The fear of collapse was stirred up by the Y2K problem , the possibility of a global computer network breakdown. This was originally a technology problem that was heavily and massively mythologized perhaps because billions of dollars were at stake (Csef, 1999). Among the many predictions concerning the year 2000, the 1972 Club of Rome report on the limits to growth was misunderstood as a model to predict the real breakdown of the global economy due to resource depletion, environmental degradation, and population explosion (Schmid, 1999). To summarize, Gnosticism plays a vital role in the *Universitas* millennium issue. Global environmental change and changing HNRs are included but do not feature centrally.

Globalization is featured outside the main section of that special journal issue in an interview (Reif, 1999) conducted with Harold James, a Princeton University historian, who speaks of globalization beginning in the 19th century and suggests that while the process may be deplorable for some, it is nevertheless irreversible. James cites earlier historical events and occurrences as evidence for globalization, such as the rst transatlantic cable in 1866, the New York stock exchange crashes of 1906 and 1907 and their immediate repercussions on the European stock markets, or unifying global trends in fashion, including the Japanese adoption of Western dress and fashion, and similar trends in the ne arts since the age of the Renaissance. As one reaction, globalization has also provoked some outcry such as the one referring to it as the

globalization trap (Martin and Schumann, 1996) or critics in the developing world who say, globalization equals imperialism . Earlier (just as they are today) protective tariffs were designed to bar international competition. Welfare and social legislation was designed to strengthen the nation state. Visionary third-ways , such as an attempted balance between the market and planned economy, are viewed by some as roads leading directly into Third World underdevelopment. GEC and the HNR do not appear to be connected with globalization in the sense it is used here.

The sociologist's world-view, similarly, seems in general to be less environment related and nature minded. We shall now turn to two sociology journals and consider their millennium special issues.

The British Journal of Sociology (1/2000) and *Current Sociology* (4/1999) take up the millennial problematique, the latter focusing on the future of sociology and the social sciences , the former on sociology facing the next millennium . This could be an indicator of the importance attached to the issue of GC and the nature–human relationship by eminent mainstream sociologists. The latter journal invited its contributors to consider what the Millennium might indicate about the history of human societies' and especially how sociology' is facing up to the challenges and opportunities posed , and to provide analyses of such transformations (Urry, 2000). Manuel Castells opens the volume by proposing a grounded theory of the network society as the social structure of the information age . The networks are empowered by the new communication technologies and reshape the relations of production, consumption, power, experience and culture. Ecologism is an example of an alternative network in opposition to dominant networks (Castells, 2000). Since the 1960s, globalization (along with complexity analysis and cultural studies) has transformed the context of sociology, according to Immanuel Wallerstein. He proposes a unified or re-unified historical social science as a truly global exercise (Wallerstein, 2000). Goran Therborn documents the shift from a universal to a global sociology in the second century after sociology became a discipline. He forecasts a comparative and competitive focus among the neighboring disciplines – sociology, political science, and economics – rather than along the social-versus-natural-sciences divide (Therborn, 2000). Cross-cultural, empirical comparison of societies in transformation, without worrying too much about theoretical concerns, is advocated (Esping-Anderson, 2000). The conflict between the patchwork quilt of nation-states and the cosmopolitan order of human rights may open the door to the second age of modernity (Beck, 2000). Contributions on urban sociology (Sassen, 2000), cultural diversity and the internet (Featherstone, 2000), and mobile sociology (Urry, 2000) conclude the volume.

Dissenting from the social sociology orientation, a natural tilt can be detected in subsequent contributions. Science and technology studies and the social explanation of natural scientific facts work towards a physical sociology and its epistemology (Latour, 2000). The focus on socioenvironmental theory and the case of genetic modification of food reveal that the social sciences' relationship to nature and environment matters (Adam, 2000). Timescape is conceived as the temporal equivalent of landscape. The timescape analysis of socio-environmental matters brings contextualized temporal complexity to the heart of social theory . Thus the time aspect is central to understanding sustainability and its emphasis on nature's regenerative capacity. Intergenerational equity and cultural equity are at stake – the ownership of reproduction has been transferred to transnational companies. A time-sensitive scholarly enterprise is ••• the task that confronts social theory at the beginning of the new millennium (Adam, 2000).

Current Sociology (Volume 4, (4) October 1999) reports the results of the symposium on The Future of the Social Sciences in the 21st Century which was the concluding session of the 14th World Congress of Sociology in Montreal in August 1998. The whole journal issue dwells on the pros and cons and various aspects of interdisciplinarity, of opening up to and collaborating with neighboring disciplines. Disciplinary boundaries should be negotiated, not simply closed down. To think in space-time and touch geography and history is recommended (Massey, 1999; similarly Allardt, 1999). An active interdisciplinarity is needed, in particular with economics, and also a transdisciplinary or intercultural approach (although the latter terms are not used explicitly, the idea they encompass is implied) to link up with different regions or countries in the age of globalization (Boyer, 1999). Finally, all fragmented perspectives, including those of singular, exclusive disciplinarity, ought to be abandoned in favor of a theoretical unitary reconstruction of the social sciences, if we wish to avoid both the barbarism of economist reductionism and the conservative nihilism of post-modernism (Boron, 1999).

Are the millennial issues representative of social science and interdisciplinary thinking? Do they address the pressing environmental themes substantively and sufficiently? To take up the scope of interdisciplinarity is certainly laudable *per se* and represents present-day avant-garde scientific development. The scientific base, though, appears a little narrow as the natural sciences have been left out, by and large. The HNR in the context of GEC was obviously not considered to be of major concern in the wake of a millennial shift. A few environmentally based and theoretically challenging contributions were competently put forward in the *British Journal of Sociology*. They do not, however, constitute anything close to a social science mainstream movement. To construct a concept of nature and the HNR is hardly even attempted. What are the conclusions as to

the relevance of the HNR and GEC for the social science mainstream? Has there been visible change?

Let us briefly review some specific aspects of the environmental sociology discourse as it began in the 1970s (we set aside the Chicago School of sociology in the 1920s) along with the global environmental movement. We argued earlier (Glaeser, 1996) that environmental sociology became established as a sociological subdiscipline in the US within the short time sequence from 1978–1980. The theoretically ambitious goal at that time was to acknowledge physical and biological factors as independent variables that influence the dependent variables of social structure and social behavior. This was intended to become a new paradigm within the sociological knowledge canon to turn the sociological mainstream towards an HNR concept.

The new disciplinary paradigm was defined through transforming the human exemptionalism paradigm (earlier called the exceptionalist paradigm) into the new ecological paradigm (Dunlap and Catton, 1979). Traditionally, the dominant world-view had been to accept humans as the one unique and superior creature on Earth, capable of quick adaptation to environmental change for cultural rather than genetic reasons. The new ecological paradigm also deviated from a specific sociological tradition established by the early French sociologist Emile Durkheim (1858–1917), who postulated that social facts can only be explained by other social facts. In a most authoritative assessment, Buttel concluded in 1987 that environmental sociology had found recognition as a speciality within sociology. It did not succeed, however, in redirecting mainstream sociology (Buttel, 1987). In our view, Buttel's evaluation was still valid in the 1990s (Glaeser, 1996); it provides the historical background and a broader base for the millennial focus discussed above.

There was one innovation, however, that gradually altered the social science outlook. During the preparation phase of the United Nations Environment Conference on Environment and Development (UNCED) Rio summit in 1992, GEC came up as a new political and scientific paradigm: In 1987, the theme had not yet been included in Buttel's (1987) state-of-the-art review and agenda for environmental sociology. Once again sociology was slow in the uptake. An early exception to this was a contribution by Buttel and Taylor (1992) who advocated in favor of the society-nature relation as a social construct; they argued as well that the GEC topic was a social construct simultaneously serving as a scientific concept and as a way of mobilizing the community of scholars. In short, environmental science and environmental movements are complementary. To date, we might interpret that proposal as an early attempt to integrate environmental stakeholders in a transdisciplinary scientific approach. Even at present, the HNR theme within the GEC context is still not social science mainstream, but it has gained momentum as will be demonstrated below. This is especially the case for the sustainability discourse in the context of the social situation in the early 21st century. The remaining parts of this contribution will examine somewhat more closely the social dimension of sustainability and present some consequences with respect to historical, theoretical, behavioral and political aspects of the HNR within the GEC context.

SUSTAINABILITY: THE SOCIAL DIMENSION

What matters about GEC and sustainable development is the human dimension . But obvious difficulties arise when operationalizing this idea (Rotmans, 1998; Rockwell, 1998). The concept of GC is broader than that of GEC. GC refers to the totality of changes on the planet Earth, including all human intentions and alterations (Rotmans, 1998). It involves both the biophysical and the human system, whereas GEC refers to the human-induced biophysical changes only. To disentangle the natural from the anthropogenic changes within the GEC framework represents a major exercise fraught with difficulty.

The sustainability concept, according to Merle Jacob, is ambiguous; its utility diminishes when one tries to operationalize the concept (Jacob, 1996). The ambiguity, oscillating between an anthropocentric orientation that focuses on the needs of future generations, and an ecocentric view that concentrates on living within the carrying capacity of supporting ecosystems, owes much to the normative character of the sustainability concept, as Rotmans argues, and depends on the cultural perspective of the actors using it. Hence different cultural perspectives would have to be elaborated and translated into different preferences so as to arrive at an operational definition of sustainable development that is linked to the notion of GC (Rotmans, 1998). The categorization of cultural perspectives and biased preferences could be linked to Dunlap's and Catton's (1979) paradigmatic shift and dichotomy between the human exemptionalist (exceptionalist) and the new ecological paradigm.

Sustainability emerged as a new development paradigm out of the concept of ecodevelopment , its predecessor. The term was popularized by the Brundtland Report, Our Common Future , from 1987. The goal was to reconcile environment and development, yet there was a strong bias in favor of environmental sustainability, which, of course, was necessary to counterbalance the strong emphasis on economic growth.

What is *social sustainability*? When we discuss social, or perhaps more appropriately societal, sustainability we can build on a relative consensus by saying that we are searching for the criteria to explain why and how societies are sustained. It would then be possible to make some reasonable predictions about the future. We use data from six to eight thousand years of human civilization composed of, say, 200–300 generations of people. Biologically speaking, this

is minute even though we tend to think of it as a relatively long period in historical, political and sociological terms.

What are the factors that maintain or help maintain a social entity for a longer period of time, such as a social group or society at large? Putting the question like this, hints at a quick and simple answer. On closer investigation, however, we see that it might involve the totality of social science theory including those parts that are yet to be written. With all of these constraints in mind, Serbser (2000) in a contribution to *Gaia* (*see* Gaia Hypothesis, Volume 5) – a Swiss–German, multidisciplinary journal covering ecological perspectives in science, the humanities, and economics – suggested the following procedure. Let us identify those or some of those necessary conditions without which social (societal) survival would not be feasible. We talk about formal requirements in a pragmatic sense: they should be independent of each other, yet in combination constitute the societal context in a total bio-cultural sense. Six preconditions (three cultural and three ecological) for a sustainable human society are suggested. The cultural order contains social action, constitution of social groupings, and their transformation in the sense of social change. These conditions for societal sustainability are contained in and derived from established contributions to sociological theory. The ecological order includes social space, social metabolism and dominance.

Social action is intentional and linked to symbolic systems such as our language, which is sufficiently imprecise to enable us to deal with complex issues. Social groups or social units as subsystems are constituted by the notion of self-identity and, complementarily, by a sharp outline defining other, competing social units. These groups undergo a constant process of reorganization, that is, transformation or social change. Social groups need a spatial environment as a constituting frame. Social space defines the situation of social action; it determines symbols and language. Social metabolism refers to the activities and interactions of social groups: they produce, they consume, and they reproduce themselves, under a regulatory framework of legal and ethical principles. Dominance, finally, refers to social control, the power structure and governance (including the system of checks and balances).

All six features – social action, social groupings, social change, social space, social metabolism, dominance – according to Serbser (2000), work in combination and, as an interacting set, they determine the survivability and sustainability of a society. But (as Serbser notes with regret) these features and their interaction have hardly been taken into account in the social sustainability discourse.

As these ideas are very recent, they are necessarily still somewhat vague. The yardstick to measure the degree to which they are analytically concrete and applicable could be as follows: First, they could serve as a tool to facilitate the determination, perhaps on a quantitative scale, of whether or not any given social (or societal) situation is sustainable. Second, in a more dynamic sense, they could produce recommendations for the implementation of measures as an incremental approach to societal sustainability.

The examples of integrated and sustainable coastal management or the deep sea commons, to take extreme and topical examples for regulated social processes, show that inequalities or simply competing interests need to be negotiated in a process of mutual control and bargaining. Even if some disagreement remains, consensus must be reached on the degree of disagreement accepted. If this does not happen, social exclusion will occur, either voluntarily or as a result of external force. The state or process of social exclusion is certainly not a sustainable one because there will always be group members, be it individuals or states, that will aim to reverse a dissatisfactory situation. It may thus be concluded that consensus building by negotiation indicates a state or process of social sustainability. We may term the absence of such indicators negative social sustainability .

Global inequality is a topic that was discussed extensively in the 1970s – on an international scale, as early as 1972, during the United Nations Conference on the Human Environment (UNCHE I) in Stockholm (Glaeser, 1997). The theme took on a new dimension during the GEC debate. More explicitly than ever before are North–South relations and inequalities linked to environmental problems that have been intimately connected with livelihood concerns. Following the two oil crises of 1973 and 1979, and the lost decade of the 1980s which was characterised by huge public debt and structural adjustments in the South, GEC is no longer defined within a social or cultural context (Redclift and Sage, 1998).

Environmental problems are viewed differently in the South than they are in the North. Poverty ecology gives priority to livelihood concerns (Agarwal and Narain, 1991); the focus is on the impacts of trade liberalization, the repayment of international debts, and structural adjustment. These aspects of globalization differ markedly from the lifestyle concerns characteristic of Northern wealth ecology . Here the environmental problems associated with GEC are climate change, acid rain, ozone depletion, loss of biodiversity and the collapse of fisheries in various parts of the world. Two decades after the United Nations General Assembly had agreed in 1980 on a drinking water decade (Glaeser and Phillips-Howard, 1983), the absence of clean drinking water is still considered to be the major environmental problem for millions in the South, causing disease and epidemics as recently illustrated in Mozambique in 2000. The International Monetary Fund (IMF) policy and other adjustment policies may be responsible for environmental degradation when resource needs increase or

additional pressures are exerted, for instance, on marginal soils or common goods. Two striking examples of this are forest depletion due to firewood extraction and fish stock depletion in coastal waters due to over-fishing and eutrophication.

Equity conflicts are social sustainability conflicts. The South cannot sustain itself in social terms if social inequalities exist. To reduce inequalities means to come closer to intragenerational equity which, on the other hand, could contradict intergenerational equity, that is, the care for future generations. Both equity issues could contradict the policy of environmental efficiency using energy tax incentives or tradable pollution permits. The latter issue, equity versus environmental efficiency, is based on the value we as a global society attach to nature, concretely to the service functions of the global ecosystem, such as the environmental cleanup function – nature's own regenerative capacity. Global commons, such as the oceans, are global sinks; they absorb, in particular, many carbon-based pollutants. Forests, too, act as global sinks in the sense of pollutant absorbers but, unlike the oceans, they are nationally owned. Social justice in the sense of intragenerational equity, for instance, in the case of forests, would thus call for compensation to be paid to the owners of common goods by the polluters. According to Redclift and Sage (1998), a global contract would mean that Southern development concerns would have to be met before Northern global environmental issues are contended with. So we have come full circle back to where we started 1972 in Stockholm! Environment versus development.

From the developing world point of view, poverty reduction, redistribution of wealth and socioeconomic equality or at least equity may be the main issues of social sustainability – environmental sustainability being its prerequisite. Some term it participation , consensus building through negotiation and social competition , or the absence of social exclusion . The overall theme is social justice. In a useful synopsis, Goodland (1995) produced a table in which social, economic, and environmental sustainability are juxtaposed. Following a widely accepted definition, sustainability means maintenance of capital. Capital embraces a social, a human and a natural form. Social (or moral) capital is constituted, among other things, by tolerance, compassion, patience, cultural identity, community cohesion, honesty standards, laws/institutions, and discipline, and requires maintenance in terms of shared values, equal rights, community participation, and religious and cultural interaction. The creation of social capital is needed to achieve social sustainability. Conversely, there can be no social sustainability without environmental stability – the latter providing the basic conditions for the former.

Environmental sustainability (as a pre-requisite for social stability) means that natural capital must be maintained. Human welfare is to be improved by protecting raw material sources and by ensuring proper sinks for human wastes. Renewables are to be kept within regeneration rates; the depletion rate for non-renewables must be kept within the renewable substitution rate. The amount of wastes produced should not exceed the assimilative natural-environmental capacity. Scale is thus added to efficiency and allocation as a third economic criterion. It constrains the throughput of energy and material through the economic subsystem and thus controls the use of natural capital from environmental sources to sinks.

Economic sustainability means holding the scale of the economic subsystem within the biophysical limits and includes production and consumption. Economic sustainability devolves on consuming interest, rather than capital (Goodland, 1995). As economists value goods and services in monetary terms, major problems arise when natural capital is to be valued. This is especially true for resources of common access such as the oceans or air. The problem becomes all the more complicated, once we realize that more than just economic values are attached to natural goods and services. Our understanding of individuals' values is limited; to understand and evaluate choices is a cross-disciplinary exercise. Ideally, economic values are represented by market transactions. If there is no market, as in the case of common access resources, transactions may be replaced by consumers' willingness to spend time or money to gain access, or by their willingness to reveal a preference for a site or a style of recreation (Lockwood, 1999). The willingness-to-pay approach to elicit environmental values may be complemented by participatory social deliberation as a means to inform the environmental decision-makers (Brouwer *et al.*, 1999).

There are obvious limits to the economic valuation of nature and the environment. Economics can certainly not cope with what philosophers term *intrinsic values* of ecosystems, non-human species or the inanimate nature. Whether intrinsic values are justified and necessary for an environmental ethic is still a matter of dispute. The top value in a value hierarchy may be assigned to human survival or human quality of life on the one side, or to landscapes and the natural world on the other. These values will eventually be incorporated into decision processes and depend on (often conflicting) positions held and adopted by stakeholders (Lockwood, 1999). Values and norms must be activated by cognitions specific to the environmental context, for instance, by adverse consequences anticipated for humans (anthropocentric value orientation) or for natural ecosystems (biocentric value orientation). Socio-psychological models are used to measure environmental values and attitudes, and to research whether these have been processed and transformed into corresponding behavior (Dunlap and Van Liere, 1978; Stern *et al.*, 1993, 1995, among others). Held values reflect

social settings and norms and interact with policy issues. To incorporate environmental values and behavior into political decision-making would certainly enhance social sustainability.

It appears that there has been a shift in the logical hierarchy of sustainability whereby the need for social sustainability is a motivating factor to pursue environmental sustainability, which, in turn, depends upon economic sustainability. In this sense, the approach followed appears to be anthropocentric, notwithstanding the fact that the notion of sustainability is inspired by the care for nature and the preservation of the environment, which may be termed enlightened anthropocentrism (Summerer, 1989). Environmental care can be integrated into authentic human development. Authors who agree with this somewhat normative statement will still disagree as to what is included in human development. If it is a rather qualitative concept in the sense of human well-being, measured in terms of health, knowledge, social order or community, it may not correlate closely with the quantitative concept of economic growth and material affluence (Dower, 2000). The question of values shows up again in the problem of alignment or non-alignment of social and natural change, in the Western world-view and enlightenment imperative to control and use nature, and in the dilemma between development and environment.

It is fairly obvious that social sustainability and negative social sustainability to a large extent depend on the theory (or theories) involved. Sustainability and its converse are theoretically constructed but this does not mean that they are untrue or unreal: The notion of social construction simply implies that social action relies on social concepts, and that social concepts depend on perspectives, some of which are scientific. Social constructs determine not only the notion of social sustainability, they also determine the notion of environmental sustainability and, as a consequence, the relationship between humans and their social and natural environment – between societies and nature. These relationships change as the underlying concepts change. The most recent of such transformations occurred in the context of what is most commonly termed, accepted, and constructed as GEC. The natural and the social sciences frame their variegated perspectives to deal with methodologically proper aspects of the HNR. Again, these perspectives are constructed in terms of scientific convention, and from the historical and methodological standpoint represented by a given discipline. Dissidents in different disciplines agreed on a dissenting perspective: that is, a unifying approach. Their scope has been integrative as opposed to analytical – compartmentalized. The holistic approach to dealing with HNRs of various kinds and in a multitude of facets has been called human ecology . *Human ecology*, enriching the synthetic and systems oriented ecological outlook with the social and cultural human actor approach, eventually became a new academic eld and subject of its own.

CONCLUSION: GLOBALONEY OR WHAT?

What conclusions can we draw with respect to HNR and their changes within the GEC context and at the historic opening of the third millennium? We may sort and distinguish between aspects related to history, theory, ethics and behavior, and policy.

History

The public in most countries seems to be poorly informed about global risk issues. This is the outcome of an international survey of public awareness and concern about environmental problems conducted in 1992 by the Gallup International Institute in 24 nations, diverse in terms of their geography, economics and social settings. Yet, even if laypersons have a limited understanding of global warming in a more technical sense, the issue s appearance as a visible social problem has surely heightened the public s general sense that humans are having a detrimental impact on the environment (Dunlap, 1998). These concerns extend from the local to the global environment and include the various aspects of the environmental problematique. Ecological awareness has evolved over decades, beginning in the 1960s, and has developed into an ecological worldview. Public attention may oscillate in accord with uctuating media attention, but it is unlikely to disappear altogether. Environmentally oriented social movements and organizations have mushroomed in the North as well as in the South. To what extent public awareness translates into behavioral changes is a matter of dispute and ongoing research.

The ndings on varying degrees of public awareness, on a basically high level, are supported by a media survey covering the American mass media in the ten-year period from 1987 to 1996 (Mazur, 1998). Until the late 1980s, environmental attention had focused either on biospheric issues exclusively, such as the destruction of rainforest or species extinction, or atmospheric hazards exclusively, mainly global warming, acid rain, and ozone depletion. Theses issues began to cluster during the 1987 1990 period, showing up as global problems. This was also a period of rising media coverage resulting in widespread public attention. Some of these hazards were connected; for instance, the greenhouse effect, the ozone hole, and the extinction of dinosaurs were interrelated in the public consciousness. A drop in the coverage of GEC can be observed for the period from 1992 to 1996, following the Rio summit. There are several explanations for this decline, none of which seems to be wholly satisfactory (Mazur, 1998). One plausible explanation, however, is that new

story lines such as the downfall of the Soviet Union and the breakup of satellite countries in the early 1990s, or the Gulf War in 1991 had greater appeal for journalists and news agencies in an evolving age of news as entertainment: infotainment .

Theory

GC issues and the HNR are polarized between epistemological idealism and epistemological realism. The distinction between the two positions appears clearer than before despite the fact that there has been a certain tendency to tackle conventional environmental problems subjectively in the social sciences and objectively in the natural sciences: environmental awareness and reflexivity versus hard facts and measured observation in environmental reality . Philosophers today may recall a comparable situation in the late 18th and the early 19th century when the German philosopher, Immanuel Kant, resolved a similar academic conflict.

Contemporary social science investigations have either evolved along the path of social constructionism (or constructivism), representing neo-idealism, or they have followed an orientation that pre-supposes a material world independent of percipient human actors (Rosa and Dietz, 1998) – neo-realism.

The neo-realism guides the social and scientific analysis of environmental changes as well as the political economy interactions between environment and society. A famous example is the still influential IPAT model which was proposed in the early 1970s and assumed that environmental impact I is a function of population size P, affluence *per capita* A, and technological development T (*see* **DPSIR (Driving Forces Pressures State Impacts Responses)**, Volume 4). The systems approach in world modeling simulates similar relationships on the basis that there are crucial driving forces that regulate the system and that are probably influenced by policy and politics. Social scientists have often criticized such concepts for being too simplistic. The approach, they argue, is too mechanistic and does not reflect human agency to the extent that it is actually present, that is, the complexity and reflexivity (reflexiveness) of social action and political response (Glaeser, 1995; Rosa and Dietz, 1998).

The neo-idealist orientation towards problems and research on environmental change issues highlights two aspects:

1. the uncertainties in the body of knowledge and the scientific knowledge claims, and
2. the attempt to provide explanations for scientific and public recognition of the environmental change problem as influenced and shaped by historical, social and political forces (Rosa and Dietz, 1998).

In this approach, the emergence of scientific concern and the rise of public awareness are scrutinized; these issues eventually become more important than the environmental problem under dispute. Environmental threats to the global ecosystem or human health are perceived only to the extent that they attract media attention and are publicized accordingly. To a great extent, the social constructivist approach is reflexive, and it is applied as a science of science – meta-theory. Constructivist methodology is useful in detecting critical shortcomings in realist models, which may be based on or entail false (or at least uncertain) assumptions. Social constructionism, on the other hand, has been criticized for neglecting real world problems and concerns in that human–nature relations and environmental change issues are constructed or conceptualized, that is, produced or created rather than extracted or mapped .

It seems wise, then, to re-adopt the Kantian position in the sense that the strongholds of epistemological idealism and realism are to be combined. The critical potential of social constructionism should be retained without forgetting that the survival or livelihood problems facing humankind do not disappear when we turn our attention away from them. We may cautiously suggest that the issue of GEC has largely been adopted by natural scientists, who view themselves as realists. The underlying models and assumptions ought to be scrutinized by reflecting which construction represents which stakeholders' interests. The issue of HNR, on the other hand, has been taken up through idealist social science and theory of science. What is at stake here are HNR changes over time, space, and culture; the social construction of GEC themes is among the relationships under consideration. Recent examples of this include the identification of driving socio-political forces behind GEC, closely related to factors of modernity (Spaargaren and Mol, 1992; Mol and Spaargaren, 1993; Wilenius, 1999); an heuristic reading of classical sociology texts to provide theoretical insights into GEC studies (Prades, 1999), or the sacralization of nature and cosmocentric mythology (Giner and Tabara, 1999).

Ethics and Behavior

We take a closer look at the mythology issue since it takes up behavioral aspects and the ethical complex with respect to ecological rationality. Religious responses to GEC may inspire modifications in how we treat the world environment, this in turn having repercussions for how we conceive human actions and what we deem to be a rational social order. These conceptions have shaped – for certain ecology-minded groups in society, at least – the present cultural situation which interprets GEC as risky.

It is argued that eco-religion is a necessary condition to implant ecologically rational behavior. Environmental anxiety coincides with the chiliastic movement (see the section on the millennial shift: gnosticism and the environment) which views (eco)religious disobedience as cause for apocalypse. GC and growing scarcity are seen as a consequence of our environmental misconduct. Religious responses shift from monotheism to pantheism, whereby nature as God (*deus sive natura*) becomes the object of worship. Examples like Lovelock's Gaia hypothesis or the deep ecology movement illustrate well how this potential can grow into a firm religious belief. Global ecological rationality will emerge as a new form of rationality and induce new cultural contexts of action affecting personal individual behavior and action to mitigate environmental destruction. Whereas eco-religion will eventually die, its behavioral aspects may still remain ••• with future generations. The ecological ethic will thus survive the spirit of the corresponding religious beliefs that inspired it ••• (Giner and Tabara, 1999).

While the notion of ecological rationality considers the intricate relationship between religion and science, the mechanisms and processes for social change as a means to achieve global ecological rationality are not revealed. Still, it should be noted in our context of HNR that the metaphysical and ethical components of transformation are linked to social behavior and policy formation as they relate to GEC (Giner and Tabara, 1999), and as they are finally incorporated into everyday culture, manners and lifestyles of the world population, forgetting their eco-religious roots.

Policy

Environmental policy and management represent the human action approach to GEC in the HNR. Globalization includes global environmental policy and management. Global environmental problems cannot be solved at the national level but the national level will still retain its importance and role in the development of environmental policy. The *subsidiarity principle* is pursued by the European Union. It means that a higher policy level replaces a lower policy level only if the lower level cannot appropriately take care of the problems under consideration. The *subsidiarity principle* is expected to be adopted on a global scale: Local environmental problems should be tackled locally; global problems should be dealt with globally. The pre-requisite that there be competent global scale actors to manage this, has thus far not been fulfilled. There is no World Environment Organization that has the power and the standing of the World Health Organization or the World Trade Organization. The global environment is regulated by Multilateral Environmental Agreements, such as the 1997 Kyoto Protocol, which regulates the reduction of greenhouse gases. The agreement acknowledges global commons to which open access is denied. To date there are more than 170 multilateral agreements to regulate environmental protection on a global scale. National participation is voluntary; not all countries sign or participate in such agreements. Whether an agreement is supported or opposed depends very much on the international epistemic communities that represent expert networks and act across national boundaries (Petschow and Droge, 1999; Swanson and Johnston, 1999).

The international research community on the human dimensions of GC includes natural and social science scholars with common interests, working in universities, research institutions and government laboratories. They communicate with each other at conferences or workshops, through scientific journals, and via the Internet. Major changes in academic development, as induced by global environmental research, include the growing volume of inter- and transdisciplinary research to address human and environmental problems, the commitment to public policy and management, and the rapid diffusion of information, via the electronic media, among the research community and other stakeholders. This new structuring and production of knowledge is sometimes addressed and discussed as the electronic invisible college (Brunn and O'Lear, 1999). The members of the GC community have research interests in various overlapping areas within the human–environment interface: human ecology. They include cross-disciplinary clusters, working on specific topics such as environmental and climate change, land use and resource management, integrative coastal zone management and eco-tourism, sustainable agriculture, environmental protection and food security. The electronic invisible college promotes virtual conferences on specialized topics of GC among leaders, students, and practitioners.

Scientific expertise provides important inputs into environmental policy, planning, and the decision-making process, from global level to local level. A major difficulty, apart from the implementation deficits, which have been widely discussed by political scientists, may be termed here (adopting macro economics terminology) the magic triangle of sustainability . Social (societal), environmental and economic sustainability represent conflicting goals that must be optimized. Optimization means not only a participatory bargaining process between the stakeholders involved; it also requires scientific information, hard or at least fuzzy data with respect to changes in ecological carrying or absorbing capacity for a given region, and resulting social and economic impacts. Sustainability is very often not accountable when it comes to concrete cases. We live in a state of uncertainty regarding scientific data. The ethical and political consequence will then be to act cautiously, in a value-conservative mode: If the precise limits to sustainability are unknown, it is imperative that we not bar the path to increasing or at least maintaining sustainability. Just such a precautionary principle has, to some extent,

been adopted in environmental legislation (*see* Precautionary Principle, Volume 4). Its implementation depends on the HNR that we as the 21st century GEC society wish to realize. Reviewing the politico-environmental agendas since Stockholm 1972, we may state cautiously that ecological modernization , which applies high-tech efficiency and ecological taxes within a growing market, has won acceptance in many countries; whereas ecological structural change , which builds on the social reorganization of society to achieve consumerist modesty and lifestyle self-limitation, is far from becoming a political option.

In summarizing and concluding let me raise a final question: Is all talk of globalization just a lot of globaloney as a number of journalists and critics seem to believe or are there substantive issues that must and can only be dealt with on a worldwide scale – globally; and, if so, what does dealing with them globally really mean (for Occident and Orient, North and South, women and men, rich and poor)? The HNR has been a major theme in natural philosophy for centuries, if not millennia. The aspect of change was added more recently with respect to concepts of nature. The global environment was hardly a human concern before the 1970s – a decade that has witnessed the limits-to-growth discourse, the first UN Conference on the Human Environment, and two oil crises reminding the global society of the simple fact that global natural resources are finite. The *avant-garde* of the social sciences dealt with all of the above issues; by interpreting them as social constructs, to reveal some different perceptions. The scientific community realized that the search for GEC solutions needs inter- and transdisciplinary (including non-scientific stakeholder) synthesis and policy related cooperation (among others: Brewer, 1986; Ravetz, 1986; Committee On Research Opportunities and Priorities For EPA, 1997).

GC is reflected in regional and local development. Development and change in different parts of the world or in different segments of society create winners and losers . Perceptions are usually considered to be a function of culture and development. The inverse tendency is also true: Perceptions such as humans are the masters of the universe (*dominium terrae*) or exploitation of nature enables us to upheave the social order (*dominium hominis*) induce or determine new developments. Globaloney ? This is pretty serious business. Let us work together to turn globaloney into meaningful efforts to provide every citizen of the globe with prosperity, security, and a healthy environment in which to thrive. This is what is intended by globally sustainable politics, policies, and livelihoods.

ACKNOWLEDGMENTS

I wish to thank Wolfgang van den Daele, Rainer Dobert, Ted Munn, and Peter Timmerman for their helpful comments and constructive criticisms of the manuscript in its earlier stages.

I am also grateful to Mary Kelley-Bibra and Emma Aulanko for technical assistance and other invaluable help in the final preparation of this article.

REFERENCES

Adam, B (2000) The Temporal Gaze: the Challenge for Social Theory in the Context of GM Food, *Br. J. Sociol.*, 51, 125–142.

Agarwal, A and Narain, S (1991) *Global Warming in an Unequal World: a Case of Environmental Colonialism*, Center for Science and Environment, New Delhi.

Allardt, E (1999) The Future of the Social Sciences in the 21st Century – a Comment, *Curr. Sociol.*, 47(4), 13–17.

Aspen Institute (1999) *A New Transatlantic Agenda for the Next Century*, Aspen Institute, Berlin.

Barker, M (2000) The Book of Enoch and Cosmic Sin, *Ecologist*, 30(1), 30–34.

Beck, U (2000) The Cosmopolitan Perspective: Sociology of the Second Age of Modernity, *Br. J. Sociol.*, 51, 79–105.

Boron, A A (1999) A Social Theory for the 21st Century?, *Curr. Sociol.*, 47(4), 47–64.

Boyer, R (1999) Le Paradoxe des Sciences Sociales: Les Vues d'un Economiste Dissident', *Curr. Sociol.*, 47(4), 19–45.

Brennan, E M (1999) Population, Urbanization, and Security: a Summary of the Issues, *Environmental Change & Security Project Report No. 5*, 4–14.

Brewer, G D (1986) Methods for Synthesis: Policy Exercises, in *Sustainable Development of the Biosphere*, eds W C Clark and R E Munn, IIASA-Book, Cambridge University Press, Cambridge, 455–473.

Brouwer, R, Powe, N, Turner, R K, Bateman, I J, and Langford, I H (1999) Public Attitudes to Contingent Valuation and Public Consultation, *Environ. Values*, 8(3), 325–347.

Brunn, S D and O'Lear, S R (1999) Research and Communication in the Invisible College of the Human Dimensions of Global Change, *Global Environ. Change*, 9, 285–301.

Buttel, F H (1987) New Directions in Environmental Sociology, *Am. Rev. Sociol.*, 13, 456–488.

Buttel, F H and Taylor, P J (1992) Environmental Sociology and Global Environmental Change: a Critical Assessment, *Soc. Nat. Res.*, 5, 211–230.

Castells, M (2000) Materials for an Exploratory Theory of the Network Society, *Br. J. Sociol.*, 51, 5–24.

Chaitanya, K (2000) Man, Nature and Cosmos in Vedic India, *Ecologist*, 30(1), 21–24.

Committee on Research Opportunities and Priorities For EPA (1997) *Building a Foundation. For Sound Environmental Decisions*, National Research Council, National Academy Press, Washington, DC.

Csef, H (1999) Die Millenniums-Panik, Zeittypische Angste an der Schwelle zum neuen Jahrtausend, *Universitas*, 54(642), 1137–1149.

Dower, N (2000) Human Development – Friend or Foe to Environmental Ethics? *Environ. Values*, 9(1), 39–54.

Dunlap, R E (1998) Lay Perceptions of Global Risk: Public Views of Global Warming in Cross-national Context, *Int. Sociol.*, 13, 473–398.

Dunlap, R E and Catton, W R (1979) Environmental Sociology, *Ann. Rev. Sociol.*, 5, 243–273.

Dunlap, R E and Van Liere, K (1978) The New Environmental Paradigm, *J. Environ. Educ.*, 9, 10–19.

Echlin, E (2000) An African Church Sets the Example, *Ecologist*, 30(1), 43.

Esping-Andersen, G (2000) Two Societies, One sociology and No Theory, *Br. J. Sociol.*, 51, 59–77.

Featherstone, M (2000) Archiving Cultures, *Br. J. Sociol.*, 51, 161–184.

Geissler, K A (1999) Der Ubergang im Ubergang, *Universitas*, 54(642), 1162–1173.

Giner, S and Tabara, D (1999) Cosmic Piety and Ecological Rationality, *Int. Sociol.*, 14, 59–82.

Glaeser, B (1995) Environment, Development, Agriculture, Integrated Policy Through Human Ecology, UCL Press, London.

Glaeser, B (1996) Sociology of the Environment: a German-American Comparison, *Human Ecol. Rev.*, 3(1), 32–42.

Glaeser, B (1997) Environment and Developing Countries, in *The International Handbook of Environmental Sociology*, eds M Redclift and G Woodgate, Edward Elgar, Cheltenham, 101–118.

Glaeser, B and Phillips-Howard, K D, eds (1983) Clean Water for Third World Development, *Vierteljahresberichte: Problems of International Cooperation 94*, 317–426, special issue.

Goldsmith, E (2000) Religion at the Millennium, *Ecologist*, 30(1), 4–8.

Goldsmith, E (2000) Archaic Societies and Cosmic Order – a Summary, *Ecologist*, 30(1), 15–17.

Goodland, R (1995) The Concept of Environmental Sustainability, *Ann. Rev. Ecol. System.*, 26, 1–24.

Griffith, B (2000) Return to the Center, *Ecologist*, 30(1), 9–11.

Hilgers, M (1999) Zum Jahrtausendwechsel, *Universitas*, 54(642), 1125–1127.

Jacob, M L (1996) *Sustainable Development: a Constructive Critique of the United Nations Debate*, Department of Theory of Science and Research, Goteborgs Universitet Goteborg.

Kiely, R and Marfleet, P, eds (1998) *Globalisation and the Third World*, Routledge, London.

Klingebiel, R and Randeria, S, eds (1998) *Globalisierung aus Frauensicht, Bilanzen und Visionen*, Dietz, Bonn.

Latour, B (2000) When Things Strike Back: a Possible Contribution of Science Studies to the Social Studies, *Br. Jr. Sociol.*, 51, 107–123.

Lockwood, M (1999) Humans Valuing Nature: Synthesizing Insights from Philosophy, Politics and Economics, *Environ. Values*, 8(3), 381–401.

Lovelock, J E (1979) *Gaia: a New Look at Life on Earth*, Oxford University Press, Oxford.

Martin, H P and Schumann, H (1996) *Die Globalisierungsfalle. Der Angriff auf Demokratie und Wohlstand*, Rowohlt, Hamburg.

Massey, D (1999) Negotiating Disciplinary Boundaries, *Curr. Sociol.*, 47(4), 5–12.

Mazur, A (1998) Global Environmental Change in the News: 1987–90 vs. 1992–6, *Int. Sociol.*, 13, 457–472.

Mol, A P J and Spaargaren, G (1993) Environment, Modernity and the Risk-Society: the Apocalyptic Horizon of Environmental Reform, *Int. Sociol.*, 8, 421–459.

Murray, R (2000) The Cosmic Covenant, *Ecologist*, 30(1), 25 – 29.

Nasr, S H (2000) The Spiritual and Religious Background of the Environmental Crisis, *Ecologist*, 30(1), 18 – 20.

Parry, M and Livermore, M, eds (1999) A New Assessment of the Global Effects of Climate Change, *Global Environ. Change*, 9, supplementary issue.

Pearce, D (1999) *Economics and Environment. Essays on Ecological Economics and Sustainable Development*, Edward Elgar, Cheltenham.

Petschow, U and Droge, S (1999) Globalisierung und Umweltpolitik. Die Rolle des Nationalstaates, *Aus Politik und Zeitgeschichte*, B23/99, 23–31.

Prades, J A (1999) Global Environmental Change and Contemporary Society: Classical Sociological Analysis Revisited, *Int. Sociol.*, 14, 7–31.

Ravetz, J R (1986) Usable Knowledge, Usable Ignorance: Incomplete Science with Policy Implications. In *Sustainable Development of the Biosphere*, eds W C Clark and R E Munn, IIASA-Book, Cambridge University Press, Cambridge, 415–432.

Redclift, M and Sage, C (1998) Global Environmental Change and Global Inequality: North/South Perspectives, *Int. Sociol.*, 13, 499–516.

Reif, A (1999) Der Prozess der Globalisierung ist unumkehrbar, Interview with H. James, *Universitas*, 54(642), 1212–1220.

Rockwell, R C (1998) From a Fictional Globe to Poetic Ecosystems: Modeling Human Interactions with the Environment, in *Earth System Analysis. Integrating Science for Sustainability*, eds H J Schellnhuber and V Wenzel, Springer, Berlin, 461–487.

Rosa, E A and Dietz, T (1998) Climate Change and Society: Speculation, Construction and Scientific Investigation, *Int. Sociol.*, 13, 421–455.

Rossi, V (2000) Sacred Cosmology in the Christian Tradition, *Ecologist*, 30(1), 35–39.

Roth, S (2000) The Cosmic Vision of Hildegard of Bingen, *Ecologist*, 30(1), 40–42.

Rotmans, J (1998) Global Change and Sustainable Development: Towards an Integrated Conceptual Model, in *Earth System Analysis. Integrating Science for Sustainability*, eds H J Schellnhuber and V Wenzel, Springer, Berlin, 421–453.

Sassen, S (2000) New Frontiers Facing Urban Sociology at the Millennium, 51, 143–159.

Schellnhuber, H J (1999) Earth system Analysis and the Second Copernican Revolution, *Nature*, 402, supplement, *C*19–*C*23.

Schellnhuber, H J and Pilardeaux, B (1999) Den Globalen Wandel Durch Globale Strukturpolitik gestalten. *Aus Politik Zeitgeschichte*, 52–53(99), 3–11.

Schellnhuber, H J and Wenzel, V, eds (1998) *Earth System Analysis. Integrating Science for Sustainability*, Springer, Berlin.

Schmid, K P (1999) Wahrlich, Wahrlich, Ich Sage Euch Was Uns Fur 2000 Alles Vorhergesagt Wurde und Warum Fast Nichts Davon Eintraf, *Universitas*, 54(642), 1150–1161.

Schulz, M and Kracht, U, eds (1999) *Food Security and Nutrition. The Global Challenge*, St. Martin s Press, Berlin.

Serbser, W (2000) Societale Nachhaltigkeit, *Gaia*, **9**, 71 75.

Spaargaren, G and Mol, A P J (1992) Sociology, Environment, and Modernity: Ecological Modernization as a Theory of Social Change, *Society Nat. Res.*, **5**, 323 344.

Stern, P C, Dietz, T, and Guagnano, G A (1995) The New Ecological Paradigm in Social-psychological Context, *Environ. Behav.*, **27**, 723 743.

Stern, P C, Dietz, T, and Kalof, L (1993) Value Orientations, Gender and Environmental Concern, *Environ. Behav.*, **25**, 322 348.

Summerer, S (1989) Vorsorge contra Nachsorge Ist Umweltqualitat planbar? in *Humano kologie. Grundlagen praventiver Umweltpolitik*, ed B Glaeser, Westdeutscher Verlag, Opladen, 272 285.

Swanson, T and Johnston, S (1999) *Global Environmental Problems and International Environmental Agreements. The Economics of International Institution Building*, Edward Elgar, Cheltenham.

Therborn, G (2000) At the Birth of Second Century Sociology: Times of Re exivity, Spaces of Identity and Nodes of Knowledge, *Br. J. Sociol.*, **51**, 37 57.

Urry, J (2000) Editor s Introduction: Sociology Facing the Millennium, *Br. J. Sociol.*, **51**, 1 3.

Urry, J (2000) Mobile Sociology, *Br. J. Sociol.*, **51**, 185 203.

Voegelin, E (1952) *The New Science of Politics*, The University of Chicago Press, Chicago, IL.

Voegelin, E (1959) *Wissenschaft, Politik und Gnosis*, Kosel, Munich.

Vondung, K (1999) Die Apokalypse am Ende des zweiten Jahrtausends, *Universitas*, **54**(642), 1128 1136.

Wallerstein, I (2000) From Sociology to Historical Social Science: Prospects and Obstacles, *Br. J. Sociol.*, **51**, 25 35.

Wilenius, M (1999) Sociology, Modernity and the Globalization of Environmental Change, *Int. Sociol.*, **14**, 33 57.

Wilson, D (2000) Grandfather s Story, *Ecologist*, **30**, 12 14.

Economics and Global Environmental Change

EBAN GOODSTEIN

Lewis and Clark College, Portland, OR, USA

Confronted with any environmental problem (whether toxic emissions from a local incinerator, or the vast conundrum presented by global warming) the question " how much is too much?" necessarily forms the starting point for an economic analysis. Since pollution and resource degradation are a by-product of material production, and human welfare depends upon both material production and a livable environment, what is the right trade-off? Of course, such a question is a normative one; as such it has no objectively right or wrong answer. Although economists generally agree both on a common ethical framework, and a particular de nition of sustainability as a desirable social goal, there is signi cant disagreement about how much pollution and resource degradation society should tolerate if we are to achieve a sustainable economy. Some support the use of bene t- cost (BC) analysis to guide resource development and pollution decisions; others prefer a stricter precautionary principle.

As economists move to more positive analytical ground, however, broader agreement emerges, though on some points signi cant debate remains. First, why do laissez faire market economies generate unsustainable outcomes? Certainly because of negative externalities. These are dif cult to internalize through private negotiation owing both to the predominance in the economy of open access common property resources, and the diffuse nature of environmental harm. An additional argument is that markets encourage investments with rapid paybacks, at the expense of longer-term investments.

Second, are there ways to reduce the costs of environmental regulation? Certainly by shifting away from the current prescriptive forms of regulation, to more exible, incentive-based (IB) approaches. A second approach supported by many environmental economists is to promote direct investment in clean technology.

Finally, how can we promote environmental quality in poor countries? First, raising incomes for the poor majority is a critical step; in the face of desperate poverty, the environment will become rapidly degraded. Second, progress can be made in reducing population growth by recognizing the economic motives underlying fertility decisions. Third, the complex links between rich country consumption, the movement towards free trade and investment, and global environmental quality are important factors, if still poorly understood. Fourth, effective global environmental agreements are hard (but not impossible) to achieve because they are public goods, subject to free riding.

How much pollution is too much? How can we do better? And how can we resolve pressing global environmental issues? Economic approaches to answering these questions share the goal of shedding light on the trade-offs implicit in different environmental policy choices.

THE ETHICAL FRAMEWORK OF ECONOMICS: UTILITARIANISM

Economists are in the contentious business of recommending what they perceive to be the mix of government policies and private market institutions needed to maximize, in Adam Smith's words, The Wealth of Nations . Wealth in the economist's context should be interpreted broadly to mean all those things that bring people a high quality of life. Societies are poor, not wealthy, if they have polluted air and water; if they lack wilderness areas for recreation and solitude; and if they are unable to satisfy a moral desire to protect other species from extinction.

It is important to emphasize that economic analysts are concerned with human welfare or happiness. To an economist, saving the blue whale from extinction is valuable only insofar as doing so yields happiness (or prevents tragedy) for present or future generations of people. The existence of the whale independent of people is of no importance. This human-centered (or anthropocentric) moral foundation underlying economic analysis, which has

as its goal broad human happiness or utility, is rooted in the 19th century ethical system known as *Utilitarianism*.

Some environmentalists are hostile to utilitarian arguments for protecting the environment. Indeed, an economic perspective on nature is often viewed as the primary problem, rather than part of the solution. The philosopher Mark Sagoff (1995, 618) puts it this way:

> ••• the destruction of biodiversity is the crime for which future generations are the least likely to forgive us. The crime would be as great or even greater if a computer could design or store all the genetic data we might ever use or need from the destroyed species. The reasons to protect nature are moral, religious and cultural far more often than they are economic.

The focus on anthropocentric, utilitarian arguments by economists is not meant to discount the importance of other ethical views. Indeed, over the long run, non-utilitarian moral considerations will largely determine the condition of the planet which we pass on to our children and theirs. But economic analysis is firmly wedded to this utilitarian ethical framework; moreover, arguments about human welfare invariably crop up in debates over environmental protection, often playing dominant roles.

Within this utilitarian worldview, sustainability takes on a particular meaning. Sustainable outcomes require that we reduce pollution (or stop depleting resources) if doing so, on balance, prevents the decline of living standards below their current level for the typical (median) member of any future generation (Pezzey, 1992).

MARKET SYSTEMS AND SUSTAINABILITY: EXTERNALITIES

Economists largely agree that pure free market systems are unlikely to generate sustainable outcomes (Goodstein, 1999a, Chapter 3). *Laissez faire* systems fail because of the existence of negative externalities: costs generated in the production or consumption of a good which are not borne by the producer or consumer of the good. For example, when I buy a liter of gasoline and burn it in my jet ski, I impose a variety of costs on society as a whole, ranging from the emission of local air and water pollutants, to global warming. Many of these costs are not reflected in the price of gasoline. They are thus external to the buyer and seller.

If all resources in an economy were privately owned, and environmental damages could be proved easily, then externalities would be internalized through private negotiation or litigation. For example, if Bill Gates owned the Mississippi River, and my jet ski was polluting it via emissions of soluble hydrocarbons, then Bill could sue me and force me to internalize the external costs imposed on his river. However, many important resources are not privately owned: the air, rivers and streams, oceans, most forests, deserts and other natural habitats. Moreover, environmental damages are often spread among many parties, and difficult to prove in specific cases, creating high transactions costs for private resolution of the externality problem. Given these two features (the widespread presence of common property resources, and the high transactions costs associated with court claims for damages) economists agree that free market systems will generate excessively high levels of pollution and resource degradation.

The conventional solution to this problem is to force companies and consumers to internalize externalities through government regulation. Regulation can be either prescriptive (command-and-control, CAC), in which firms are required to install particular types of clean up technologies, or it can be incentive based (IB). The latter approach includes both pollution taxes, and marketable permit systems. These IB methods internalize externalities by forcing firms to pay for pollution, but leave the specific method of pollution reduction up to the companies themselves. (These systems are discussed more fully below).

If government forces firms to pay for pollution through regulation, how much should firms pay? The conventional view is that businesses should pay an amount per unit of pollution equal to the damage done by the last unit of pollution emitted (the *marginal unit of pollution*). Thus, if the last ton of sulfur dioxide coming out of a power plant stack leads to economic damages of $200, than the plant should pay $200 for every ton that it emits. In competitive markets, setting pollution prices equal to marginal economic damage in this way balances the costs of pollution control (higher prices for products like electricity) against the benefits (less damage from pollution like acid rain), leading to maximum total monetary benefits to society.

Economists have developed a variety of techniques to determine monetary values for pollution damages (Goodstein, 1999a, Chapter 8). These range from the mundane (fewer sick days for workers) to the highly controversial (placing a dollar value on stroke deaths prevented, reduction in child IQ avoided, or biodiversity preserved). How can these latter types of valuations be done?

As one example, the value of a life saved in benefit–cost (BC) studies typically ranges from $2–10 million. These numbers comes from studies that examine the wage premium for risky employment. Put overly simply, economists have found that workers like police officers receive a wage premium of around $500 for accepting a 1 in 10 000 increase in the risk of premature death on the job. 10 000 police officers thus exchange one of their lives for about $5 million.

One can quarrel with this line of reasoning on many grounds. And clearly, if considerations of this type dominate, then uncertainty in dollar valuations of the benefits of environmental protection can loom large. Nevertheless, economists have developed a number of tools for monetizing important components of quality of life traditionally left out of measures like gross domestic product (GDP).

This allows society to put a monetary value on pollution prevention: for example, global warming damage estimates from carbon dioxide emissions from coal fired power plants range from around half a penny per kWh to three cents per kWh (Krupnick and Burtraw, 1996). And while numbers in this range may ultimately turn out to be wrong, defenders of this kind of estimation of the dollar benefits of environmental protection respond that even the wrong dollar value is better than the default price of zero.

MARKET SYSTEMS AND SUSTAINABILITY: SHORT TIME HORIZONS

While the problem of externalities can, in theory, be dealt with via regulation, free market economies have another feature which may lead to unsustainable outcomes. In modern market systems, firms expect profit rates of return on the order of 15 to 20% in order to induce them to either conserve resources or to invest in new technology. With such high discount rates, projects with payoffs occurring much more than five years in the future seldom look profitable. Thus, from a business perspective, it seldom makes sense to preserve rainforests on the grounds that in 50 years the biodiversity it contains will be highly valued; neither are energy companies much interested in solar power as long as cheap oil and coal have at least five year lifetimes.

It is easy to show that these short time horizons create a situation in which future generations are not as well off as we could possibly make them. If we invested more today in conservation of certain resources or in basic research and development (R&D), then we could with a high probability raise the welfare of our descendants. This is the general rationale for government support of science and technology. However, it is not clear whether our short-term bias is unsustainable, that is, whether it actually reduces the welfare of future generations below the level that we enjoy today. Even with their short-term biases, market systems are incredibly dynamic in terms of the development of new technologies. It may be that the five-year time horizon (coupled with government support for longer-term research) can still insure that the well-being of future generations will not decline. This depends in turn on the degree to which new technologies can substitute for depleted or degraded natural resources and ecosystem services (Goodstein, 1999a, Chapters 6 and 7).

SUSTAINABILITY: THE NEOCLASSICAL VIEW

The sustainability debate hinges on this issue of human-created substitutes for natural capital: inputs (raw materials) and waste sinks serving the economic system. On this point, two camps have developed: *neoclassical* and *ecological*. The neoclassical tradition in economics is based on a vision of broad substitutability of inputs and outputs: labor can substitute for capital in production; oranges can substitute for apples in consumption. Neoclassical economists acknowledge that the natural world supplies important inputs and waste sinks for the economic system. However, incremental degradation of this natural capital, it is argued, seldom leads to the loss of unique services; all have reasonable substitutes.

Recall that sustainability was defined earlier as non-declining welfare for the typical member of a future generation. The neoclassical argument is that investment in created, human-made capital can (and tends to) generate a flow of services improving quality of life which at least offsets the loss of services provided by degraded natural capital. This concept has been dubbed *weak sustainability*.

Neoclassical economists often use the example of whale oil or copper telephone wires. As these resources became relatively scarce and their prices rose, petroleum and fiber optics emerged as substitutes. What about a harder case like the loss of climate stability due to global warming? Neoclassical economists respond, first that as greenhouse gas emissions are regulated, the externality is internalized, and the price of carbon dioxide (CO_2)-based services (like gasoline-powered automobiles) rise, then new low-CO_2 technologies will emerge. Moreover, advances in agricultural techniques will insure adaptation to a changing climate, and dikes can be built to hold back sea level rise. Clearly there will be losers from climate change, but on balance, the argument goes, living standards for the typical person on the planet will continue to rise even in the face of moderate climate change.

Neoclassical economists in this context are technological optimists. This means that they believe that in well behaved market economies (those in which negative externalities have been internalized via IB regulations like pollution taxes or marketable permit systems), substitutes will emerge for scarce (and increasingly expensive) natural resources and environmental waste sinks. System sustainability is thus assured by assumption.

Technically, this means that the economy will display a positive rate of growth of *per capita* net national welfare, the familiar GDP measure of economic output adjusted downwards to reflect the costs of economic growth. If net national welfare, which balances the benefits of increased material consumption against the attendant environmental and social costs, is growing on a *per capita* basis, than by definition, the well-being of future generations is also rising (Goodstein, 1999a, Chapter 6).

While (weak) system sustainability is generally assumed in the technologically optimistic neoclassical framework, sustainable decisions at the micro level should be made using BC analysis. For example, when considering whether

or not to build a coal plant, one should compare the discounted stream of measurable monetary costs (material and environmental) over the expected life of the plant, against the discounted stream of measurable monetary benefits (cheaper power). If the benefits exceed the costs, than building the plant would be sustainable in the weak sense, because doing so would, on net, raise the welfare of the typical member of future generations.

Future costs and benefits are discounted in BC analysis to reflect the opportunity cost of productive foregone investment. For example, the value of $100 received in 10 years (ignoring inflation) is a lot less than the value of $100 received today. If the interest rate were 10%, I could invest that $100 today and have $260 in 10 years. Conversely, to have $100 on hand in 10 years I would need to bank only $39 today. This figure of $39 is called the *Present Discounted Value* of $100 received in 10 years, if a 10% interest or discount rate is available. While discounting makes sense for individual financial decisions, at the social level it can generate perverse outcomes (*see* Discounting, Volume 5).

To illustrate, suppose we are considering installing a filter on our coal plant that will generate a one-time-only benefit of $100 million 100 years from now. At a discount rate of 10%, a $100 million benefit gained 100 years from now is worth only $7200 today. The logic of BC analysis thus suggests that we should not spend a measly seven thousand dollars and change today to yield a benefit of $100 million to our great grandchildren. Why? Because if we put the money in the bank, future generations would have more than $100 million on hand.

The problem is that, for real world projects subject to BC analysis, we do not intend to put net savings in a trust fund; the alternative is instead to spend it elsewhere in the economy (say, to reduce energy prices from the coal plant) where it may or may not yield a 10% rate of return to society as a whole. While discounting makes obvious sense on a personal level for short time horizons, it breaks down for actions yielding a stream of social benefits far into the future.

On the flip side, not discounting at all also generates perverse outcomes. Consider a project that yields $1 worth of benefits every year, forever. Such a project has infinite value, and would thus seem to be worth sacrificing the entire planet's output to finance. Given these two extremes, there is a lot of debate over the proper approach to discounting future benefits and costs (Portney and Weyant, 1999).

If indeed the analysis is sensitive to the choice of a discount rate, or else tries to weigh controversial benefits such as lives saved or biodiversity preserved, then uncertainties can easily grow large enough to render determination of a (weakly) sustainable outcome at the project level difficult, if not impossible.

SUSTAINABILITY: THE ECOLOGICAL VIEW

In contrast to neoclassical economists, ecological economists are technological pessimists (*see* Ecological Economics, Volume 5). Tracing their intellectual lineage explicitly to Malthus, ecological economists have no faith that technological substitutes for important natural resources, such as climate stability, will be forthcoming from real-world market systems. Thus, ecological economists would reject a coal plant following the logic that CO_2 emissions are already destabilizing the global climate. Essentially, ecological economists argue that for important resources such as fresh water, ultraviolet protection, biodiversity and climate stability, the future consequences of current resource degradation are too uncertain to countenance the use of BC analysis.

Instead ecological economists argue for what they call the Precautionary Principle (*see* Precautionary Principle, Volume 4). To preserve the welfare of future generations, therefore, unique forms of natural capital should be protected *unless the costs of doing so are prohibitively high* (Ciriacy-Wantrup, 1968; Daly, 1996). This strategy of protecting natural capital has the advantage of not relying on the development of uncertain substitutes to insure (weakly) sustainable outcomes. Thus following the Precautionary Principle promotes strong sustainability from an ecological perspective, by insuring that future generations will not be impoverished by the loss of natural capital. Moreover, the argument is often simultaneously advanced that the cost of protection will not be so high as is frequently claimed (Laitner *et al.*, 1998).

Along this line, ecological economists point to survey evidence suggesting that, in developed countries, broad growth in material consumption in fact buys very little increases in societal happiness. This occurs, it is argued, because beyond a basic level, utility from consumption depends on relative rather than absolute levels of consumption. To the extent that this is true, environmental quality is being sacrificed only to feed a rat race in which human welfare does not rise with increased material consumption, clearly an unsustainable state of affairs (Mishan, 1968; Howarth, 1996).

Which view is correct? Neoclassical economists argue that the ecological position is too extreme. They insist there are trade-offs, and that we can pay too much for a pristine environment. Resources and person-power invested in reducing small cancer risks or preserving salmon streams, for example, are resources and people that cannot be invested in schools or health care, also of value to future generations. BC analysis is needed to obtain the right balance of investment between environmental protection and other goods and services. Moreover, in the sustainability debate, neoclassical economists argue that history is on their side: Malthusian predictions have been discredited time and time again.

Indeed, in recent years, some scientists have made stunningly bad forecasts regarding looming resource shortages. The problem with these predictions, as with Malthus's original one, was that they dramatically underestimated the impacts of changing technologies. Neoclassical economists point to these failed predictions to support their basic assumption that natural and created capital are indeed good substitutes: we are not running out of natural resources or waste sinks.

Ecological economists respond that history is not a good guide for the future. One hundred and fifty years after the beginning of the industrial revolution, the argument is that accumulating stresses have begun to fundamentally erode the resilience of local and global ecosystems upon which the economy ultimately depends. Indeed, ecological economists have largely shifted their 1970s concerns about running out of nonrenewable minerals and oil to concerns such as: biodiversity, fresh water, environmental waste sinks, and productive agricultural land. While ecological economists stretching back to Malthus have indeed done their share of crying wolf, this does not, of course, mean the wolf is not now at our door.

The neoclassical and ecological perspectives differ dramatically in their assessment of the likelihood of sustainable outcomes from real world market economies. Nevertheless, both perspectives are economic: both accept the common definition of sustainability offered above; both are grounded in broadly utilitarian philosophy, and both deal in the currency of trade-offs. In addition, this is a debate about ends, not means. Ecological economists prefer that resource development or environmental degradation be regulated according to a precautionary principal; neoclassical economists prefer a BC test. When it comes to means, however, economists from both schools would prefer to see greater reliance on incentive systems than is provided by existing environmental and resource regulations (see below).

THE CURRENT REGULATORY SYSTEM

Perhaps the most striking aspect of environmental regulation at the national level is its brief history. As recently as 1970, for example, the US had no major federal legislation controlling the discharge of pollutants into the air and water, no national regulations covering the disposal of hazardous waste onto land, no process for reviewing new chemicals, only a limited procedure for registering new pesticides, and no protection for endangered species.

Beginning in the early 1970s, countries around the world began adopting national environmental protection legislation. Most of these laws required government agencies to develop detailed prescriptive regulation for several different types of industrial polluters. Economists have since dubbed this approach CAC regulation. CAC involves two parts. The first is uniform emission standards. All similar sources are commanded to meet identical emission levels. The control part of the name refers to the technology-based regulatory approach employed by much current regulation. Regulators specify, at a high level of detail, precisely which technologies firms must employ to reduce emissions.

While this initial effort to internalize externalities was effective in many regards, CAC regulation is generally not cost-effective, either in the short run or the long run. (By cost-effective is meant achieving a given pollution reduction goal at the lowest possible cost.) Uniform emission standards raise costs in the short run, since all plants must meet the same reduction targets, regardless of differences in the cost of doing so. Technology-based regulation raises costs in the long run by stifling incentives for innovation in new pollution control technology. These factors have led most economists to advocate a greater use of IB regulation like emission charges or marketable permit systems.

More broadly, regulation has been criticized as ineffective on a variety of grounds:

- regulatory design and enforcement is subject to significant political influence;
- regulation addresses symptoms in a piecemeal way, rather than underlying causes in a comprehensive way;
- regulations must be continually tightened or else they will be overwhelmed by economic and population growth;
- regulation has already picked off the easy targets: large, stationary pollution sources. Increasingly, pollution is arising from mobile sources like cars, and non-point sources like run-off from streets and farms.

These factors have led some economists to call for greater government support for the development of technologies which are much cleaner in the first place.

Before discussing calls for reform, however, it is important to note that CAC systems have significant achievements to their credit (Goodstein, 1999a, Chapter 14). In the United States for example, in spite of the fact that economic activity more than doubled from 1970 to 2000, the general urban air pollution picture has actually improved, dramatically for lead. In addition, large reductions in air toxins have been recorded over the last decade. Industrial emissions of some water borne pollutants have dropped dramatically and, overall, water quality has improved a bit beyond 1970 levels. Regulation of hazardous waste is likely to prevent the development of future *Love Canals*. And the rising cost of disposal, along with the right-to-know regulations, and the potential for Superfund liability is beginning to focus corporate attention on reducing waste through pollution prevention. Particularly nasty new pesticides are not likely to make it through the EPA's initial screen, and under prodding from Congress, the agency will restrict the use of the

worst of the existing pesticides in the first few years of this century. Finally, only a few listed species have slipped into extinction. The fact that regulation has managed to hold the line against economic growth is in itself an impressive accomplishment.

DOING BETTER: INCENTIVE-BASED REGULATION

It is straightforward to see how uniform standards and government mandated technology requirements under CAC raise the costs of regulation. In the case of uniform standards, consider two neighboring oil refineries. Refinery A, which has the ability to reduce emissions of say, nitrogen oxides, at very low-cost, is nevertheless expected to meet the same standard as next door refinery B, a firm with very high costs of reduction. The reason this is not cost-effective is simple: the same overall emission reduction (and local air quality) could be achieved at lower cost by having the low-cost firm meet a tighter emission standard, while relaxing the standard for the neighboring high-cost firm. Similarly, in the short run, a single technological mandate is unlikely to provide the cheapest pollution control solution for different firms all over the country. This lack of flexibility inherent in centralized, technology-based regulation raises costs (Goodstein, 1999a, Chapter 16).

However, long-run cost effects are more important. Technology-based regulation works against technological improvement in several ways. First, once a firm has installed the government-mandated technology it has no incentive to do better. Second, the firm actually has a positive incentive not to do better. If, for example, it discovered a better technique for pollution reduction, and the EPA then decided to deem this new technology to be the standard, the firm might legally be required to upgrade its pollution control technology at other new facilities. Firms have a distinct incentive to keep the agency from upgrading the state-of-the-art technological standard. Finally, if in spite of these obstacles, firms do seek an innovative, non-approved approach, they must first obtain regulatory clearance to do so.

IB regulation provides an alternative to CAC. There are two kinds, closely related. The first is a pollution tax (also known as an effluent or emission charge or fee). For example, to reduce power plant emissions of carbon dioxide, which fuel the greenhouse effect, society could institute a tax on emissions of, say, $25 per ton of CO_2. Alternatively, one might achieve a roll-back through a marketable permit system (also known as a tradable permit or cap-and-trade system). Here CO_2 permits would be issued or auctioned to polluters only up to a certain target level of overall emissions. These permits could then be bought and sold by others, again putting a market price tag on pollution (*see* **Tradable Permits for Greenhouse Gases**, Volume 4).

The first effect of IB systems is to eliminate uniform emission levels at different sites. Plants facing high costs of reducing pollution will cut back a little and either pay the emission charge or buy additional permits. Plants with low costs will cut back a lot, and either save on pollution taxes, or make money by selling some permits. Overall pollution levels can be set the same as under CAC regulation, but the compliance costs will be less as firms are given greater flexibility, as well as incentives to make emission cuts at the lowest cost sites.

The second beneficial aspect of IB regulation is that it provides a flexible framework for innovation. Firms can install any pollution control technology they desire. Moreover, IB regulation also establishes continuous incentives for technological progress. CAC regulation does not penalize firms for any pollutants they emit once they have installed the technology mandated by the government; by contrast under IB regulation firms pay for every pound of pollution coming out of the stacks. They do this directly under an emission charge system, but pollution bears a similar price under a cap-and-trade system. By reducing pollution towards zero, firms can make money by selling their excess permits.

Textbook emission charge systems are not often found in the real world, perhaps because, as new taxes they are politically difficult to institute. By contrast, cap-and-trade systems are increasingly popular. The biggest success to date has been the acid rain program in the US. Initiated in 1995 to control the emission of sulfur dioxide from power plants, the program has functioned largely according to plan. SO_2 emissions have been reduced by about half, and the program has generated significant cost savings (Schmalensee *et al.*, 1998).

A couple of caveats are in order regarding IB systems. First, they only work well for pollutants that do their damage regionally or globally. If pollutant damage is local, than a simple IB system will generate hot spots: high pollutant damages around high cost-of-reduction plants. Second, the government must have good monitoring and enforcement capabilities. Firms must either pay the effluent charges due, or face stiff fines for any pollutants emitted above their permitted levels. In some cases, CAC regulation may be easier to monitor and enforce than an IB approach (Goodstein, 1999a).

DOING BETTER: PROMOTING CLEAN TECHNOLOGY

Regulation, whether CAC or IB, is designed to internalize externalities, and force firms and consumers to bear the social costs of pollution. Most economists agree that a shift towards greater use of IB regulation would help reduce

the costs of environmental protection, and thus make it politically easier to achieve. But is getting the prices right through regulation (whether IB or CAC) enough to insure a sustainable economy?

A second strategy is to directly promote the development and diffusion of clean technology: fuel-cell vehicles, solar and wind powered electricity, energy and water efficiency technologies, alternative agriculture, recycling (Goodstein, 1999a, Chapter 18). Rather than rely solely on controlling pollutants at the end-of-the-pipe, advocates of this approach argue that government should promote the use of technologies which reduce polluting inputs and processes in the first place. Technologies can be advanced through policies like R&D, consumer subsidies, and public procurement contracts. So called technology-forcing regulations can also promote new options; examples in the US include the Corporate Average Fuel Efficiency (CAFE) standards, energy efficiency standards for appliances, and the California zero emission vehicle requirements.

From a theoretical point of view, government efforts to influence the direction of technological progress can be justified by what economists call *path dependence* (Arthur, 1991). This theory maintains that current production technologies (for example, US reliance on private automobiles for urban transportation) represent only one possible path of development. A potentially cost competitive alternative in this case might be mass transit, dominant in many European and Japanese cities.

The path a society chooses depends on a variety of factors, including the relative political strength of the conflicting interests, chance historical circumstance, and, of course, consumer preferences and relative production costs. In the auto example, the US government's decision to construct the interstate highway system beginning after World War II, which in turn promoted suburban development, provided the decisive advantage to private transport.

However, once a path has been chosen, other paths are closed off. This happens for three reasons. First, infrastructure and R&D investments are increasingly directed towards supporting the chosen technology, and diverted from the competing path. Second, the chosen technology is able to exploit economies of scale to consolidate its cost advantage. Third, complementary technologies develop which are tailored to the chosen path, further disadvantaging the competing path. In the transportation example, this would include the sprawling retail and housing patterns of US cities, which now virtually require a private vehicle for access.

Path-dependence theory suggests that once a path is chosen, there is no easy way to switch tracks. However, in retrospect, we can see that technological choices have social consequences: the adoption of private transport, for example, has borne a substantial environmental cost. Thus, the role of government is to try and influence the current market driven process of technological development towards a path consistent with a sustainable future. Note that this theory assumes that governments in modern capitalist societies already necessarily play a major role in shaping technological change through infrastructure decisions and subsidy policies; the key here is to insure that the role is a positive one.

Under a clean technology strategy, promotional efforts should be limited to technologies which are close to commercial development, generate a quality of service and have long-run production costs comparable to existing technologies, and are environmentally superior. Each of these requirements must be met, otherwise, the technology will not spread rapidly, and little environmental benefit will result. But given these conditions, why aren't private entrepreneurs developing them in the first place?

The answer lies in the short-term bias of market actors discussed above. Simply because clean technologies are potentially cost competitive does not mean that they are more profitable than existing technologies. Entrepreneurs tend to introduce products which fill a market niche and provide at least temporary monopoly profits (computers, VCRs, cell phones). Clean technologies, by contrast, generally are not offering a new product; rather they go head-to-head to with an existing, well-established technology in a mature industry: electricity or transportation, for example. Thus they must enter an already competitive field, where only normal profits can be expected. The only clear-cut advantage that clean technologies have is in their environmental impact. While this may provide some marketing leverage, it generally will not guarantee high profitability.

Under normal market circumstances, new technologies often take a substantial time to develop a widespread following. This is due to consumers' lack of knowledge about the advantages of the new technology, as well as differences in consumer needs. The transition to any new technology requires a marketing commitment to overcome this lack of information. Marketing expenses are sunk costs, that is, costs that cannot be recovered if an investment fails. The higher the sunk costs associated with an investment, the riskier it becomes.

Clean technologies face particularly high sunk costs (and thus high risk) because they do not market themselves by offering a service that consumers do not already have. Instead, clean technologies need to woo consumers from the use of the existing technology. While clean technologies offer comparable services, they also tend to require users to learn new consumption habits. This requires a big investment in marketing, which cannot be recovered if the business fails.

In summary, many clean technologies are unlikely to be highly profitable in the short to medium term; as a consequence, private sector investors show little interest. Government policy can change this equation by absorbing

some of the risk associated with the introduction of these new technologies, nurturing infant industries.

There is disagreement among economists as to how far government should move in this direction. Basic scientific R&D generates positive externalities; not all of the benefit of innovation is captured by the inventor. For that reason, most economists agree that support of basic science is an important governmental role. Beyond that, however, while some argue that government should support significant investment in clean technologies, others maintain that government should do nothing. In part this has to do with disagreement over the extent to which the short-term bias of market economies, discussed above, has on sustainable outcomes. Those opposed to government support for technology feel that, if the prices are right (that is, if externalities have been internalized), then the private market place will bring new technologies on line soon enough to avoid unsustainable levels of environmental damage.

The disagreement also has to do with political economic views on the efficacy of government intervention into the economy. Opponents of clean technology promotion argue that the government is likely to promote the wrong technologies, and that these decisions are better left to a decentralized market process.

THE ECONOMICS OF THE GLOBAL ENVIRONMENT

This essay has so far outlined the theoretical relationship between *laissez faire* markets and sustainability, and discussed policy options at the national level for promoting sustainability. We now turn our attention to global environmental challenges: the complex links between widespread poverty in poor countries (working in part through high population growth rates), high levels of consumption in rich countries, and environmental degradation at both the global level and in the less developed world. Finally, we end with a look at the economics of global environmental agreements.

In rich countries, it is commonly assumed that environmental quality can only be improved by sacrificing material consumption. In poor countries, by contrast, a broad-based growth in income will, in many respects, tend to improve the quality of the environment. There are four close connections between poverty and the environment (World Bank, 1992).

1. For poor people, many environmental problems are problems of poverty. The biggest environmental health threats facing most people in poor countries are unsafe drinking water, compounded by inadequate sewage facilities, and urban air pollution. Exposure to indoor air pollution (smoke) from cooking and heating sources ranks close to industrial and auto air pollution as a concern.
2. Poor people cannot afford to conserve resources. Out of economic necessity, poor people often put an unsustainable burden on the natural capital in their immediate environment. Urban residents scour the immediate countryside for fuel (firewood or animal dung) leading to deforestation or the elimination of fertilizer sources. Landless farmers are pushed into over farming small plots, farming on steep mountain slopes that soon wash out, or farming in clear-cut rain forests incapable of sustaining agriculture.
3. Richer people demand more pollution control. As *per capita* income rises in a country, people begin to express a more effective demand for pollution control. Partly this has to do with education. As income rises, so do levels of awareness regarding environmental threats. Partly, it has to do with expanding democracy. As income rises, political participation tends to increase. As a result, people are provided with the opportunity to express a political demand for pollution control. Partly, it has to do with a shift in industrial composition: wealthier countries rely more on services and other relatively clean industries, while less developed countries have more basic manufacturing and mining. Finally, the relationship can be explained by relative risk considerations: when life spans are short due to inadequate nutrition or access to basic health care, concerns about respiratory, neurological and reproductive diseases, or cancer from pollution, are dampened.
4. Population growth slows with increased income. The final link between poverty and the environment lies in income growth as a means of population control, discussed later.

These four interconnections between poverty and the environment reveal that, in poor countries, one need not trade-off rising material living standards for improved environmental quality. In fact, the only effective way to improve environmental conditions is to alleviate the tremendous poverty faced by many of the people in these nations.

POPULATION AND DEVELOPMENT

Higher incomes tend to be associated with lower population growth rates. While the relationship is not hard and fast, for example, welfare families in America are not significantly larger than the average, it holds as a general rule, both across countries, and within countries. Many factors go into determining family size; religious and cultural norms obviously play a prominent role. Yet the strength of the poverty–fertility relationship suggests that economic factors are quite important as well (Dasgupta, 1995).

In poor countries, children provide two important economic functions: health and old age insurance, and after early childhood, a potential supplement to family income.

Given limited resources, a family could potentially obtain comparable economic benefits following one of two approaches: a high investment strategy with one or two children, or a low investment strategy with many children.

The high investment strategy would involve focusing all available resources on one or two children: insuring that they survive infancy, are provided with a healthy diet, and go on to attend school. Such a child might well land a good job, and increase family income substantially. For poor people with limited resources and restricted access to public services, the high investment strategy is very difficult to pursue. Moreover, it is tremendously risky. Suppose that after five years of intensive investment, a single child dies. Yet, as income rises, the high investment strategy begins to be more attractive. Resources for the long-term (15–18 year) investment are more readily available, and access to better health care means that the chances of premature death decline.

This BC model of family size is a very crude one, omitting as it does very important issues of culture, kinship, religion, love and affection. Yet it does a good job explaining the strong relationship between poverty and fertility observed in many places around the world. In particular, because there are *good economic reasons* for poor people to have large families, simple provision of low-cost birth control is unlikely to solve the population problem (though it would certainly help). Can we use this BC model to suggest other effective policies for controlling population growth?

Balanced Growth and Provision of Social Services

The poverty–fertility relationship seems to provide a straightforward way to control population: eliminate poverty! Unfortunately, if nations were able to do this easily, they would have done so already. It does remains true, however, that balanced and rapid economic development which raises the material welfare of the poor majority is one of the most effective population control measures available. In addition to balanced growth, another way to reduce poverty and thus population growth rates is through the form of publicly provided social services: education, health care, and public pension plans.

Reduced Infant and Childhood Mortality

Improved public health is an important factor in reducing long-run population growth rates. The BC approach to family size suggest that improved health care would help slow population growth for one principal reason: the risk associated with investing in a child's health and education would be reduced, encouraging families to substitute quality for quantity.

Education

The economic model of family size tells us that population growth will slow when parents follow a strategy of high investment in their children. A key element making such a strategy possible is access to education. This is true for three reasons. First, the availability of education directly lowers the cost of pursuing a high investment strategy to all parents, educated and uneducated alike. Second, as parents become educated themselves, a strategy of substituting quality for quantity also becomes easier, since the parents can provide guidance to the children.

Finally, as parents become educated, their wages tend to rise. This increases the opportunity cost of parents' time, making a low-investment strategy less attractive. This is true because a low investment, quantity strategy requires a bigger commitment of time devoted to child-rearing than does a high investment, quality strategy. Another way to look at this is that, as the parents' wages rise, family income and economic insurance are better served by parents working than by raising more children.

For this reason, education is particularly important for women, since they do most of the child rearing. One of the best ways to control fertility is to have women participating in the *modern sector* of the economy. All poor women work. However, employment in agricultural or household labor does not appear to reduce fertility. As the opportunity cost of the woman's time rises, a quantity strategy becomes less and less attractive for the family as a unit.

Education for women appears to have a strong impact on fertility control for other reasons as well. To this point, we have treated the family as a homogeneous unit, yet men and women play different roles in the family, and may have different ideas about the desirability of limiting family size. In the developing world, as in rich countries, the responsibility for taking birth control measures (whether abstinence, prolonged breast feeding, or a technological approach) generally falls on the woman. Thus the direct consumers of birth control devices and fertility control information are generally women, providing another argument for improved female education (*see* Ecofeminism, Volume 5).

More importantly, however, most of the world's societies are male dominated, patriarchal cultures. In such societies, women often must obtain approval from their husbands in order to control their fertility. This distinction is important, because holding all things equal, it is probably true that women prefer to have fewer children than do men. This is likely for a number of reasons: women do the vast majority of the hard work involved in child rearing; child rearing interferes with women's other economic activities; and child birth represents a significant health risk for many women. Thus, policies which strengthen women's status and bargaining position within the family itself, for example, better education or access to paid employment, are likely to have important effects on fertility decisions, independent of their impact on overall family income and economic insurance.

Family Planning

Having argued above that poor people may have good reasons for having large families, it is also true that a big obstacle to pursuing a high investment strategy is lack of access to family planning services. There remains a large unsatisfied demand for birth control worldwide (World Bank, 1992).

GLOBAL CONSUMPTION AND TRADE

In discussions of population pressures on the environment, it is important to recognize that the environmental damage people do depends not only on absolute numbers, but also on the natural capital depleted and waste products generated by each person. As an upper middle class citizen of the US, I consume more than *100 times the resources* used by the average individual from a poor country. The point here is not to induce guilt about af uent lifestyles, but rather to recognize that one needs to focus as much attention on reducing the environmental impact of high consumption in rich countries, as is paid to reducing population growth in poor countries.

Moreover, although the population problem is concentrated in poor countries, through legal and illegal immigration, poor-country population pressure spills over into af uent countries. And while today the consumption problem is centered in rich countries, this too will begin to change as a growing number of poor country residents aspire to the consumption levels found in rich countries.

Because inhabitants of rich countries are responsible for over two-thirds of global economic activity, at least two-thirds of global pollution can be laid at our doorstep. The western European countries, the US and Japan comprise about 15% of the world s population, but they consume 71% of the world s output (World Bank, 1992, Table A.2). As a result, rich-country consumption has to date been responsible for most of the global atmospheric pollution: global warming, ozone depletion, acid rain, and radioactive contamination. Rich countries have had by far the biggest impact on polluting and over- shing the oceans.

In addition to generating global pollution problems, there is a second element to the consumption pollution link. High levels of consumption demand in rich countries have been responsible for an unsustainable draw-down in environmental quality and the stock of natural capital in many poor countries. Many poor countries rely on the export of primary resources or agricultural commodities to earn money for imported fuel, food, consumer goods and weapons. In principal, this trade is environmentally sustainable. However, weak sustainability requires that any draw down in the stock of natural capital (whether oil or mineral reserves, rainforest resources, top soil, or air and water quality) be compensated for by investment in created capital.

The overwhelming demand for resources in rich countries (ranging from gasoline, to steel, to bananas, to beef) has in many cases depleted the natural capital stock in poor countries, without commensurate investment in created capital. Why has this occurred? Historically, colonial governments tended to drain resource-generated wealth from their colonies, investing little in human capital or infrastructure. In the post-colonial period, falling relative prices for primary resources, low taxes on politically powerful resource-based industries, and high levels of spending on military and imported consumption goods by the ruling elite have constrained investment in created capital. Finally, the burden of debt repayment has led to a ow of created wealth from poor to rich countries over much of the last 20 years.

This is not to say that trade in agricultural commodities, or any other goods, is on balance, bad for the environment. Rather, to insure sustainability, the gains from trade must be invested in created capital suf cient to offset the draw-down in natural capital. If they are not, than rich-country consumption leads to unsustainable development in poor countries as surely as does rapid population growth.

What is the relationship between trade and the environment? The pro free trade argument is that trade leads to economic growth, economic growth reduces poverty, and as we saw above, a reduction in poverty is a critical precondition for environmental protection in poor countries (Bhagwati, 1993). Poor countries often support trade liberalization through multilateral agencies like the World Trade Organization (WTO) in order to keep the rich countries from unilaterally imposing tariffs or other trade restrictions. International trade and investment can also facilitate the transfer of cleaner technologies from rich to poor countries.

These positions clearly have merit. Yet critics respond, rst, that trade does not always lead to growth because, as argued above, suf cient pro ts from trade are not reinvested in poor countries to compensate them for their loss of natural resources. Second, even if economic growth occurs, the environmental bene ts which accompany growth can be overwhelmed by the increase in pollution from export production, leading to a decline in environmental quality in poor countries.

Finally, critics charge that environmental regulations in rich countries will be weakened by liberalized trade and investment, as business mobility increases, and legal challenges are lodged by international competitors (Daly, 1993). On the former point, economists have found that in fact, very few businesses have ed from developed to underdeveloped countries to escape stringent environmental regulations (Goodstein, 1999b). This appears true in part because even for heavily regulated rms, environmental costs seldom rise above 2% of total business costs. Moreover, much pollution control technology is imbedded in modern plant design. A new chemical plant built by a multinational in Chile will in fact look a lot like one built in Texas. When

firms do head to the South, the reason is, overwhelmingly, low wages.

On the second point, free trade agreements can indeed bind the hands of environmental regulators, since they give foreign governments the ability to challenge certain regulations as trade barriers. In 1999, the US acting on behalf of Ford and Daimler-Chrysler, challenged a Japanese government initiative designed to reduce CO_2 emissions from automobiles. The law required that new mid-sized cars sold in Japan have engines as efficient as the most efficient on the market, which were then produced by Mitsubishi. Under the rules of the WTO, the US sought to have the law overturned as a barrier to trade. If the WTO rules in favor of the US, Japan will have to rescind the law or face trade sanctions (Wallach and Sforza, 1999).

Is free trade good or bad for the environment? More liberal trade is generally championed by economists because it is believed to increase efficiency, and in the long run, raise material living standards across the board. However, the process of globalization is crafting a new set of rules that will benefit some more than others, and it is also certain to generate many losers. The environmental challenge is to harness the efficiency gains generated by trade into domestic investments in poor countries, thereby helping create a sustainable future. From this perspective, agreements promoting more trade should be viewed as a means to an end (stabilizing population growth, enhancing food security, transferring sustainable technology, and conserving resources) not as an end in itself. Yet at this point, trade agreements have only just begun to acknowledge environmental and sustainability concerns.

THE ECONOMICS OF GLOBAL AGREEMENTS

Beyond poverty, overpopulation, and the growth of environmental impacts from global over-consumption, transboundary pollution problems present serious threats to sustainability. Global warming, ozone depletion, acid rain, and exploitation of the oceans, Antarctica, and even outer space, are all processes occurring outside the jurisdiction of a single nation. There is no international environmental protection agency with the authority to regulate exploitation of the global commons.

To address these problems, nations have begun to negotiate international agreements. However, economic theory suggests that effective international agreements are hard to develop. From an economic point of view, an international pollution control agreement is a public good, which must be provided voluntarily by the private nations of the world. Public goods are goods which are enjoyed in common; a classic example, if dated, is the warning service provided by a lighthouse. Technically, economists describe public goods as non-excludable. Once in operation, it is impossible to exclude any passing boat from utilizing the warning beacon provided by the lighthouse. Public goods tend to be under-supplied in a free market economy precisely because many of those who benefit can free ride if the good is provided at all. This than justifies government in taxing the beneficiaries, and then subsidizing the provision of public goods: for example, basic scientific research.

Reducing global warming is a public good, because there is no way to exclude free riders from enjoying the climate stabilization benefits provided by a treaty. However, in this case, there is no international government to insure that the sustainable level of protection is provided.

The public-good nature of environmental treaties has two implications. First, treaties will tend to be too weak from a BC point of view, since signatory nations are reluctant to reveal and commit their true willingness-to-pay in the bargaining process. Second, once signed, nations will have a strong incentive to cheat on the agreement. Unilateral cheating is another way to free ride on the pollution control efforts of others. Of course, if everyone cheats, the agreement will collapse.

Beyond free riding, agreement on burden sharing is difficult to achieve. In principle, each country might contribute their true willingness-to-pay for a treaty. This willingness-to-pay in turn would depend on both the benefits received, and ability to pay. Both of these might vary widely between nations. For example, low-lying Bangladesh, has a tremendous stake in slowing global warming, and preventing a devastating sea-level rise. Yet the country is poor, and would have a difficult time financing strong measures to reduce CO_2 emissions. On the other hand, a land-locked, wealthy country like Switzerland has a high ability to pay, but may have fewer immediate interests at stake.

A poor country's willingness-to-pay to join an agreement will typically be much smaller than that of a rich country, simply because it has a much lower national income. Yet poor-country participation will often be vital. For example, if China were to further industrialize using its vast coal reserves, global warming would accelerate considerably. Given China's low willingness to pay for a reduction in global warming (as a result of its low income), a compensation fund would have to be established for China to sign a greenhouse treaty. Those with a high willingness-to-pay (typically rich countries) would have to pay for China to adopt less polluting and expensive energy sources. Otherwise, it would not be in China's interests to sign the treaty.

If each nation did contribute its true willingness-to-pay, then the agreement process would generate an efficient level of global pollution control. However, a country's underlying willingness-to-pay for an agreement is not a well-defined thing in the first place, and is certainly not transparent to negotiators from other nations. Therefore, each nation's bargainers will have an incentive to understate

its true interest in a treaty in the hopes that others will shoulder more of the burden. In the extreme, a country might not sign a global warming or other environmental treaty at all, but still benefit from the efforts of other nations, acting as a pure free rider.

While the theory of public goods predicts that international environmental treaties will be both weak and susceptible to enforcement problems, agreements nevertheless do get signed and efforts are made to insure compliance. The 1986 Montreal Protocol to protect the ozone layer, in spite of the theory sketched out above, has been a remarkable success story. This was due to the perception on the part of many countries of a clear and present danger; relatively few pollutant sources; and the realization (after the treaty was signed) of relatively low compliance costs (Goodstein, 1999a, Chapter 22).

By contrast, the struggle to implement the 1997 Kyoto global warming accord reflects much more of the difficulty in assigning burden sharing, and the free rider phenomenon. This agreement was signed by all of the developed countries, including the US, and calls for a reduction in greenhouse gas emissions to below 1990 levels. Following the model of the Montreal Protocol, poor countries were not required to participate at the outset. The rationale for this was two-fold: developed countries had caused the problem, and they had the resources to develop the alternative energy technologies needed for large-scale reduction in carbon emissions. However, there has been strong opposition in the US Senate and from President Bush to treaty ratification: not surprising, since the US is the biggest greenhouse gas polluter. The US Senate argues that poor countries should also be included in the treaty and required to reduce emissions at the outset. This position, if it is maintained, threatens to scuttle the entire agreement (Goodstein, 1999b).

CONCLUSION

This essay has provided a brief introduction to the economic analysis of global environmental change. For economists, the goal of sustainability (an improved quality of life for most people) is well defined, if hard to operationalize. Within this utilitarian framework, an economist will ask three questions. First, given that pollution and resource degradation are a by-product of economic output, how much pollution and resource degradation are consistent with a sustainable future? Second, once the pollution reduction or resource preservation target is set, how can we achieve our goal at the lowest possible cost? And finally, how can we address global environmental problems, including the pressures placed on the environment by widespread poverty; high population growth rates; unsustainable trade patterns; and the need for more effective global agreements? As indicated, there is often lively debate among economists regarding the right answers. But what we do agree on is the centrality of these three questions.

ACKNOWLEDGMENT

Portions of this essay were first published in *Economics and the Environment* (John Wiley and Sons) and are used with permission.

REFERENCES

Arthur, B W (1991) Positive Feedbacks in the Economy, *Sci. Am.*, February, 92–99.

Bhagwati, J (1993) The Case for Free-Trade, *Sci. Am.*, November, 41–49.

Ciriacy-Wantrup (1968) *Resource Conservation: Economics and Policies*, 3rd edition, University of California Press, Berkeley, CA.

Daly, H (1993) The Perils of Free Trade, *Sci. Am.*, November, 41–49.

Daly, H (1996) *Beyond Growth: The Economics of Sustainable Development*, Beacon Press, Boston, MA.

Dasgupta, P (1995) The Population Problem: Theory and Evidence, *J. Econ. Literature*, 33, 1879–1902.

Goodstein, E (1999a) *Economics and the Environment*, 2nd edition, John Wiley & Sons, New York.

Goodstein, E (1999b) *The Trade-off Myth: Fact and Fiction about Jobs and the Environment*, Island Press, Washington, DC.

Howarth, R B (1996) Status Effects and Environmental Externalities, *Ecological Economics*, 16(1), 25–34.

Krupnick, A and Burtraw, D (1996) The Social Costs of Electricity: Do the Numbers Add Up? *Resour. Energy Econ.*, 18, 423–466.

Laitner, S, Bernow, S, and DeCicco, J (1998) Employment and other Macroeconomic Benefits of an Innovation-Led Climate Strategy for the United States, *Energy Policy*, 26(5), 425–432.

Mishan, E J (1968) *The Costs of Economic Growth*, Staples Press, London.

Pezzey, J (1992) Sustainability: An Interdisciplinary Guide, *Environ. Values*, 1(4), 321–362.

Portney, P and Weyant, J (1999) *Discounting and Intergenerational Equity*, Resources for the Future, Washington, DC.

Sagoff, M (1995) Carrying Capacity and Ecological Economics, *Bioscience*, 45(9), 610–619.

Schmalansee, R, Joskow, P, Ellerman, A D, Montero, J P, and Bailey, E M (1998) An Interim Evaluation of the Sulfur Dioxide Trading Program, *J. Econ. Perspect.*, 12(3), 53–68.

Wallach, L and Sforza, M (1999) *Whose Trade Organization? Corporate Globalization and the Erosion of Democracy*, Public Citizen, Washington, DC.

World Bank (1992) *World Development Report 1992: Development and the Environment*, Oxford University Press, New York.

FURTHER READING

See various articles in Volumes 1 and 4 concerning climate change.

37

Ecological Economics

RICHARD B NORGAARD

University of California at Berkley, CA, USA

Economists, ecologists, and scholars from a variety of other disciplines have organized an intellectual community known as ecological economics. The International Society for Ecological Economics (ISEE) was established in 1989, in response to a concern that the dominant ways of economic thinking have fostered particular forms of global change with detrimental social and environmental impacts. The constitution of the Society states:

The purpose of the Society is the advancement of our understanding of the relationships among ecological, social, and economic systems and the application of this understanding to the mutual well-being of nature and people, especially that of the most vulnerable, including future generations.

Note that ecological economists broadly stress the interaction of multiple systems and the "mutual well-being of nature and people", putting the emphasis on the interaction, rather than people or nature alone. The concern, however, is also focused on the "most vulnerable", again whether people or other species.

Expanding on this general statement, ecological economists share three general concerns that are neither adequately recognized in dominant academic and professional economic circles, nor appropriately considered in political economic discourse. These are: (1) that economies must be understood as operating within the larger biogeochemical Earth System receiving energy from the sun; (2) that environmental sustainability depends on viable social systems; and (3) that progress toward the mutual well-being of nature and people is being thwarted by excessive material consumption on the one hand, and excessive material inequality on the other. Ecological economists are methodologically eclectic, using ecological reasoning and both dominant and alternative economics in their search for better understanding of the interrelations between people and their environment, for indicators of sustainability, and for ways of bringing individual human behavior into conformity with collective human goals. While the concern with excessive consumption and equity clearly makes ecological economics an alternative community for economists and ecologists in rich countries, these same concerns make ecological economics a very comfortable community for scholars addressing sustainable development in poor countries.

THE NATURE OF ECOLOGICAL ECONOMICS

Many have tried to define ecological economics, few have reached the same explication (see, for example, Martinez-Alier, 1987; Costanza, 1989; Costanza *et al.*, 1997; Soderbaum, 2000). This difficulty, of some concern both within and beyond the field, arises for a variety of reasons. First, many researchers who self-identify as ecological economists use the models and methods of more conventional environmental economists. This has led some environmental economists to argue that there is nothing new in ecological economics. Yet the difference is clear. Ecological economists, as a whole, also use other models and methods, and actively discuss the strengths and weaknesses of environmental economic models in the contexts of the others. Second, since its modern beginnings, ecological economists have been searching for effective ways to understand and convey the importance of justice in the context of environmental sustainability and human dignity. Third, ecological economists have been grappling with how environmental problems require both new ways of understanding science and new ways of joining knowledge distributed among scientists, practitioners, and lay people to effect collective action. These latter two broad facets of ecological economics perplex both new entrants to the field and external observers who try to logically fashion ecological economics out of ecology and economics as narrowly defined. Fourth, and stemming from the first three, there is mild disagreement over whether (a) ecological economics should be striving for a new paradigm, a new set of integrative models, systems assumptions, and empirical methods, or (b) it should, or must, remain a large umbrella under

Box 1 Working Assumptions of Ecological Economics

- economies and the global economy overall require energy, materials and diverse genetic/species/ecosystem patterns and reproductive capacities provided by healthy natural systems;
    - current levels of energy use, material flows, and genetic/species/ecosystem loss are both unnecessary to a good life and a threat to the health of natural systems;
- economies and the global economy overall require healthy social systems;
    - current levels of material inequity are immoral, do not support human dignity, and are a threat to healthy social systems;
    - current rates of economic change, especially the drive for globalization, are outpacing our abilities to adapt and maintain the health of social systems.

which economists, ecologists, and others scholars compare and contrast different models and learn together. These issues are central to this description of the field.

While the scope and direction of ecological economics are actively discussed, there is considerable agreement among ecological economists with respect to their working assumptions. Ecological economic reasoning is not unique in its dependence on working assumptions, general beliefs about society and nature. Each research community focuses on a portion of reality and makes assumptions about larger systems, or takes a very broad view and makes assumptions about the details. No discipline understands the whole of reality, so assumptions about the relations between the parts studied and the whole are a necessity. Economist Karl Schumpeter referred to the combination of assumptions and choice of model as a scholar's preanalytic vision (Daly, 1996).

The following working assumptions (see Box 1) are broad and highly generalized in nature. Relatively few ecological economists work from all of them, yet most ecological economists acknowledge the importance of each of them and how they work together. This recognition and respect demonstrate that ecological economists see the field as a collective effort.

Threats to natural and social systems are a threat to people overall, but they are especially a threat to today's poor and to future generations. Ecological economists put social and natural systems at the center of their thinking, rather than individuals. The term social systems' encompasses all forms of economic and social organization, the legal and informal institutions that set the rules, and the knowledge and shared beliefs that allow people to understand each other. The metaphor of health helps elaborate the meaning of the mutual well-being of nature and people . Yet desirable properties of these systems remain vague. When empirical research is conducted, specific favorable qualities of social and environmental systems are adopted. Some ecological economists argue that energy use or material flows or biodiversity are sufficient indicators of biogeochemical system health. These effectively serve as their working assumptions when conducting specific analyses. With respect to social systems, justice, participation, shared-learning, human dignity, and other terms are often invoked.

Ecologists and natural scientists largely hold the assumption that the economy works within the larger biogeochemical system generally, hence the appropriateness of ecological as a modifier to economics. The concern with human communities, and how these are threatened by inequity and the speed and character of current economic change, however, are nurtured by ecological economists' broader roots in the social sciences and applied work in developing countries. By identifying environmental and social systems as central to their thinking, ecological economists are explicitly arguing that they are important, i.e., that they have value in themselves. The intertwining of *natural system* and *social system* concerns into one preanalytic vision owes much to the rich mix of scholars, from developed and developing countries, interacting through the ISEE during its first decade.

The sub-points in the working assumptions with respect to materialism, growth, and inequity are even more clearly value judgments. There are corresponding values integral to the working assumptions of mainstream economists. However, the values remain implicit, rather than explicit. By assuming, for example, that science allows us to rise above nature, and that material progress helps us all to live a better life, conventional economists are implicitly valuing these approaches. Whether one *a priori* presents people, as in nature or rising above nature, is not simply an issue of fact but a statement of a preference or goal as well. Similarly, ecological economists' moral concern with the distribution of happiness, and their strategic concern that inequity threatens the health of social systems, are difficult to separate. Such value judgments help ecological economists select and frame their research and what subsequently appears in the ecological economics literature, just as surely as historic beliefs about progress and how it helps all affect the research and literature of mainstream economists.

HISTORICAL BACKGROUND

An appreciation for the historical path of economics is essential to understanding the ways in which ecological economists have parted from that path and what they seek through exploring alternatives (Christensen, 1989; Costanza *et al.*, 1997, Chapter 2). The conclusions economists reach

are determined by the structure of their formal models, the quality of their data on the state of the economy, and very general assumptions about the relationships between economies, society, and biogeochemical reality. The particular assumptions that economists have come to share over their history differ from those acquired by ecologists, for example, and set them apart as a separate academic tribe and professional culture. Some assumptions are embedded in the ways in which economic models simplify complexity to a manageable form. Some are rooted in the social issues and liberal political philosophy of the formative years of economics in the 19th century. Additional premises were embraced as economists assumed roles as technical experts in national governments and international agencies during the 20th century to evaluate public investments, assure full employment, promote development, and, most recently, facilitate trade and international capital mobility. While there has always been diversity in formal models and assumptions among economists, particular combinations of models and assumptions have clearly dominated at different points in time, complementing the problems at hand and the political economy of power affecting economic change within nations and internationally.

Classical Economic Origins of Ecological Economics

Ecological economists have an af nity with the way many classical economists, from the late 18th through most of the 19th century, combined moral reasoning and systems thinking to describe problems and point toward solutions. Many of the most important economists in this period also knew that environmental systems and resources were central to human well-being. The following vignettes of early economists provide critical background.

Francois Quesnay (1694 1774) and the Physiocrats argued that economic activity should be governed in keeping with the laws implanted in Nature by Providence . Quesnay, an admirer of physics, named the school of economic thought he established to re ect the scienti c principles on which economic wealth was hypothesized to depend. Quesnay and his disciples developed a simple input output table of the economy, categorized by economic class, with all wealth owing from agricultural labor. A disciple of Quesnay, Anne Robert Jacques Turgot, rose to the position of comptroller general in France in 1776, began to impose physiocratic reasoning to tax policy, and was soon relieved of of ce after offending too many political interests. While the arguments of Adam Smith soon surpassed those of the physiocrats, they in uenced many 19th century economists including Karl Marx. The input output table was a precursor to modern material ow models of economies. The physiocrats were early ecological economists in the sense that the laws of natural science, to the extent they understood them, were integral to their understanding of economics.

Adam Smith (1723 1790), widely recognized as the founder of modern economics, was a moral philosopher who argued that shared moral standards develop because individuals empathize with each other. Concerned with whether individual economic decisions summed to the common good, he reasoned that if two people choose to enter into an exchange, it is because the exchange makes each of them better off. So long as the individual decisions made by parties to an exchange did not adversely affect other people, society as a whole was better off through the exchange. Appealing to Judeo-Christian images of God, Smith invoked the metaphor of the market being guided as if by an invisible hand. Smith was not simply arguing for free markets in the abstract. *Mercantilism*, a system wherein the state granted sole trading rights to particular companies, favored a few over the many. Smith argued that freer markets would be better for people overall. Thus the rst modern economist fought the socioeconomic power system of the time with ethical and economic reasoning. His plea for the market and against government intervention was in opposition to the tight relations that had developed between companies and the state, not against government *per se*. The relations between corporations and government have changed many times since mercantilism, and they are again of some concern.

Thomas Malthus (1766 1834), a clergyman turned economic philosopher, explained the prevalence of war and disease as material phenomena rather than as acts of God. He argued that human populations were capable of increasing exponentially and would do so as long as suf cient food and other essentials of life were available. He further hypothesized that people could expand their food supply only arithmetically through new technologies and expansion into new habitats. Given the geometric potential of population growth, and the arithmetic food constraint, population would periodically surpass food supply. At these times, people would ravage the land, go to war over food, and succumb to disease and starvation. Human numbers would consequently drop to sustainable levels, but the process would repeat itself again. Both Charles Darwin and Alfred Wallace acknowledged that Malthus model of population pressing against an environmental constraint was key to their formulation of the theory of natural selection. Limits to growth are also central to many ecological economic models.

David Ricardo (1772 1823) introduced a second model of how economic activity relates to the environment, not because he was concerned with environmental degradation or human survival, but rather to justify why landlords received a rent from their ownership of land. Ricardo argued that people would initially farm the land that

produced the most food for the least work. As population increased, farming would extend to less-fertile soils requiring more labor. His model shows how increasing population drives people to farm in previously undisturbed areas and how higher food prices lead to intensified effort on land already in production. His model of how humans extend and intensify land use underlies many debates over development and conservation.

John Stuart Mill (1806–1873), the son of social philosopher James Mill (1773–1836), argued that competitive markets complement individual liberty. Mill also wrote on both the immorality and the waste of subjugating women to men. He neither saw material prosperity as an end in itself nor thought continuous growth in material well-being was possible. Mill envisioned economies becoming mature and reaching a steady state, in which people would be able to enjoy the fruits of their earlier savings. The idea that economies would reach a steady state was logical. Change was common, unceasing growth was not, and relatively steady states rather than random change were perceived as natural. A century after Mill, Herman Daly, one of the founders of modern ecological economics, elaborated on the necessity and advantages of a steady-state economy (Daly, 1973).

Karl Marx (1818–1883) criticized capitalism on the grounds that it left workers at a subsistence wage while capitalists thrived. Marx, an admirer of Darwin and biological thinking, was concerned with how the introduction of fertilizer inputs produced an agricultural revolution during the 19th century that not only detrimentally transformed social relations but irreversibly disturbed ecological relations as well (Foster, 1999). He had strong progressive beliefs, arguing that capitalism was but a phase in a larger inevitable history of progress toward socialism and eventually communism. While his long-run predictions still inspire resistance movements, the logic of his critiques, his moral stance, and historical approach are an integral part of many schools of social thought to this day. The capitalist economic system was tamed during much of the 20th century, partly in response to how Marx had framed its problems. Capitalism keeps changing and has become especially virulent again at the opening of the 21st century while inequity within and between nations has increased.

Alfred Marshall (1842–1924) synthesized the work of the classical economists, retained their philosophical voice, and incorporated the findings of the mathematicians who formalized economics. Among Marshall's many theoretical contributions to economic analysis can be found frequent references to the beauty of nature, and the importance of clean air and water to a decent city life. He argued that markets, driven by the possibilities for individual gain, would not provide sufficient public goods. Lastly, he argued that while economists had drawn very successfully on the mechanical models of physics, the future of economics was in joining with biology.

Natural Science Origins of Ecological Economics

The classical economists were broadly read, aware of developments in the natural sciences, and thereby prone to address questions of natural limits. They were not, however, scientists themselves. Scientists, well aware of the rapidity of economic change during the 19th century, periodically asked how the laws being discovered in the natural sciences might affect the possibilities for economic progress (Martinez-Alier with Schlupmann, 1987). Serhii Podolinsky (1850–1891), writing on energy and economics during the 1880s, explored the implications of the efficiencies of energy use as it flows from the sun, through agriculture, on through human consumption and labor output, and hence into the broader economy. Today, agricultural energetics and the possibilities for greater dependence on renewable energy are central to ameliorating climate change. Patrick Geddes (1854–1932) argued in the 1880s that economists inappropriately blended the physical processes of economies with monetary exchanges. Much like modern ecological economists, he argued that more could be revealed if we modeled the physical and the monetary processes separately. Leopold Pfaundler (1839–1920) argued in 1902 that physical principles needed to be applied to calculate the carrying capacity of the earth to derive realistic expectations for the future. Frederick Soddy (1877–1956), a Nobel laureate in chemistry, argued in the 1920s that economists did not understand how our economies are driven by current and stored energy from the sun, that stored energy was the only real wealth, and that the conception of monetary wealth implied that economists believed people could eat their cake and have it too . A J Lotka (1880–1949) borrowed from the mathematical developments in economics to outline a theory of value based on ecological fitness (Kingsland, 1985). For reasons partly elaborated in the next section, economists did not respond to the early concerns expressed or alternatives provided by natural scientists. Today, many of these same concerns and alternatives drive the biophysical approaches of ecological economists.

The Rise of Modern Economic Culture

During the later decades of the 19th century, the market model became formalized mathematically and key concepts became much better understood. During roughly the same period, economics and the other modes of social analysis began to assert themselves as separate sciences, assuming an objective stance while actively trying to shed their ethical baggage. Economists turned increasingly inward, staying less and less in touch with advances in the natural sciences

as well as with the other social sciences, while they pretty much dropped out of philosophical discourse all together. The less economists read beyond economics, the more natural their own assumptions appeared to them. Later, during the Great Depression, economists were brought into national governments to help stabilize economies and to undertake cost-benefit analyses of alternative investments, of water projects first, and eventually almost everything. Earlier arguments about achieving the greater good were modified to justify ignoring equity issues as economists concentrated on efficiency in public policy analysis. Soon after, economists became significant players in the new mission of international development, advocating that the best way to reduce poverty was through growth. As economics turned inward academically, the shared assumptions of economists narrowed and hardened even while economists' new professional roles required that assumptions change with the problems and politics of the times.

It is important to understand that while the models of economics have become better understood through their formal development, the assumptions inherent to the models have not changed. What changed during the 20th century was how economists used their models. Early economists compared the results of their economic reasoning with that of other forms of reasoning, playing them off each other to maximize understanding. As economists assumed greater roles in public decision-making, they needed to justify that the economic way of thinking was the best way of thinking. During the last decades of the 20th century, academic economists began exploring new approaches and broadening their thinking. Professional economists, however, still have had to defend established ways and particular answers.

Environmental Economics

Environmental economics is a sub-discipline of economics. Its strongest roots are in the US where Richard T Ely, the founder of the American Economic Association at the end of the 19th century, also founded a school of land economics and utility regulation economics at the University of Wisconsin. Early environmental economists – Ely, George Wehrwein, Marion Clawson, Maurice Kelso, and Siegfried Ciriacy von Wantrup – raised fundamental questions from the 1920s into the 1960s about property, organization, environmental degradation, and resource conservation when much of the profession was engaged in mathematical refinement of perfectly functioning markets. Environmental economists in the 1950s and 1960s – Karl William Kapp, E J Mishan, John Krutilla, Allen Kneese, and Jack Knetsch – questioned growth and its effects on the environment, when their colleagues were advocating greater resource use in developed and developing countries in an economic race with the former Soviet Union. With the rise in popular environmental concern in the early 1970s followed by the energy crisis, environmental economics flourished, and the concept of market failure became central to economic understanding. The Association of Environmental and Resource Economists was founded in 1979.

Though their roots include fairly radical thinkers, environmental economists have generally extended existing economic patterns of thinking to environmental issues rather than question whether economic thinking is a part of the problem. The professional role of economists in environmental agencies has reinforced the tendency to rally behind dominant approaches rather than raise fundamental questions. The assumptions embedded in the culture of economists, and largely assumed by environmental economists, were well established before the rise of ecology as a full-fledged science in the second half of the 20th century. Economists' assumptions about nature and possibilities for the future are rooted in an earlier time when technological change seemed to only solve problems, not create new ones as fast as it solved the old. In rich countries where the sub-discipline started, environmental economists tend to advocate modest corrections to the existing development paths (*see* Environmental Economics, Volume 5).

The Rise of Modern Ecological Economics

In response to the narrowing of economic thought during the 20th century, a few scholars more open to thinking across natural and social science boundaries provided the foundations for modern ecological economics. Kenneth Boulding (1910–1993) worked with philosophers and natural scientists on a general systems theory in the 1950s, introduced the metaphor of spaceship earth to stress how economies must work within limits (Boulding, 1966) and subsequently wrote on ecodevelopment in a broad evolutionary framework (Boulding, 1978). Kneese *et al.* (1970) building on the first law of thermodynamics, constructed material balance economic models to better understand the limits of markets. Odum (1971), an ecologist, developed an energetic systems view that he applied across ecological and economic systems. Georgescu-Roegen (1906–1994) argued that economic activity needed to be understood as an entropic process (Georgescu-Roegen, 1971). Systems theorists Meadows *et al.* (1972) wrote *Limits to Growth*, convincingly arguing that unlimited growth was impossible. Daly (1973) reintroduced Mill's steady-state economy with arguments for limiting throughput of energy and materials. Entomologist turned energeticist David Pimentel documented the energy costs of alternative forms of agriculture (Pimentel, 1974).

Karl-Goran Maler of Sweden explored some of the implications of materials balance to a general equilibrium economic approach (Maler, 1974). Passet (1979) in France began to combine economic and thermodynamic thinking, while Pillet teamed up with Odum (1987) to write on energy

and economics. Many had already suggested corrections to the system of national accounts (SNA) so that, for example, expenditures correcting new environmental problems were not treated as benefits, Roefie Hueting of the Netherlands pushed the corrective process the furthest (Hueting, 1980). In 1982, about 40 economists and ecologists met in Stockholm, hosted by ecologist Ann-Mari Jansson, to explore how economics and ecology could be integrated (Jansson, 1984). At that meeting, Robert Costanza used the term ecological economics in a paper suggesting how energy analysis and economic valuation might be merged (Costanza, 1984). The term gained further prominence as the title to a book on how economists had resisted responding to challenges from natural scientists (Martinez-Alier, 1987).

While the Stockholm meeting emphasized natural constraints, ecologist William Murdoch argued that distributional issues were absolutely central to understanding economies and environments in developing countries (Murdoch, 1984). In 1987, Joan Martinez-Alier hosted a second key meeting of ecologists and economists in Barcelona. At this meeting, the nature of our understanding of complex systems and what this meant for the policy process was clearly added to the agenda of ecological economics (Norgaard, 1989; Funtowicz and Ravetz, 1990). Following discussions at Barcelona, the journal *Ecological Economics* was started by Robert Costanza and Herman Daly who also took the initiative to found the International Society for Ecological Economics (ISEE) soon after.

THE METHODS AND ISSUES OF ECOLOGICAL ECONOMICS

Ecological economists are methodologically pluralistic (Norgaard, 1989). The field includes economists using market, macro, institutional, Austrian, and Marxist forms of analysis. There are ecologists who do energetic analyses, population biology, evolutionary ecology, hierarchy theory, food web modeling, landscape ecology, and global biogeochemical modeling. Environmental historians, ecological anthropologists, development sociologists, systems theorists, and epistemologists and social philosophers, among others, also participate in ecological economics. Such a diversity of backgrounds is a strength, for each of the methods of the individual disciplines only spotlights a portion of the problem. Where models in economics and ecology share the same mathematical form, they can be combined into a broader, brighter light. Population models from ecology mesh nicely with production models from market economics. Input–output analysis is used in macro-economics and in food web analysis and some energy accounting models, hence can be readily combined. To the extent that evolutionary thinking has developed in economics, it combines comfortably with the approach of evolutionary ecology. Most of the models of ecology and economics, however, do not fit together. Thus ecological economists contend with different, and often somewhat incongruent, findings from different patterns of thinking. When the understandings gained from separate models can be aligned, confidence is gained. When separate understandings are in juxtaposition, conclusions are withheld and new questions are raised. How methods are used can best be described by looking at several particular issues ecological economists address.

Environmental Valuation

The relationship between market prices and value has long been at the center of controversy. Economists put a very specific meaning to the word *value*, defying the rich heritage of the term and constraining its meaning in discussions of the future. The early critiques of economics by natural scientists were driven by a concern that economists' understanding of value was divorced from natural limits and processes. The idea that values are inherent to nature, an idea that fuels debates among environmental ethicists today, were explored under the concept of natural value, for example, by Lotka. During the latter half of the 20th century, economists colonized this realm of value too by developing a variety of techniques for putting monetary values on environmental goods and services, the values they would have if they were traded in markets. Ecological economists use these new techniques but interpret the results with greater caution than environmental economists for three major reasons.

The first concern of ecological economists with environmental valuation comes from within economic theory itself. How much people are willing to pay (WTP) for clean air, for example, assumes that people do not already have a right to clean air. When asked how much they would be willing to accept (WTA) to give up clean air, which assumes people have the right to clean air, people give answers that are two or three times WTP, and in some studies 10 times or more. Environmental economists have generally adopted the polluter pay principle, implying that people have the rights to a healthy environment (*see* PPP (Polluter Pay Principle), Volume 4). Yet in the case of contingent valuation research, environmental economists have adopted the convention that WTP gives more appropriate answers than WTA. Ecological economists have not adopted the same convention. Rather, they see the difference between WTP and WTA as a crucial example of an underlying problem in environmental valuation.

The second concern is that market prices reflect the existing distribution of rights more generally, not simply between polluters and pollutees, but between the rich and the destitute, wealthy nations and poor, those who choose traditional ways and those who embrace the market, and present and future peoples. Existing techniques for simulating how an environmental market would work, implicitly

assume that the existing distribution of rights to resources and environmental services is the correct one. Howarth and Norgaard (1995) built an overlapping generations model of a hypothetical economy and showed that with all markets working perfectly, whether the economy was running at all for future generations depended on their having rights. With different assignments of rights across generations, resources have different prices over time. And when non-market environmental services were included in the model, environmental services have different values depending on the rights of future generations. Ironically, a motivating factor for doing environmental valuation has been to justify changing the law so that a wilderness area or an endangered species is saved for future generations. Yet neoclassical economics itself indicates that economic reasoning depends on the rights of future generations in the first place. So, there is a major problem of circularity in standard environmental economic reasoning.

The only way out of this circularity is to accept that questions about the distribution of rights must be addressed through some method of moral reasoning beyond the simple individual utility calculus behind economic reasoning. While such complications have long been recognized by utilitarian philosophers and economic theorists, acknowledging that economic reasoning alone does not result in unique answers, with respect to the public good, defeats the efforts of economists to provide simple answers to policy questions. Practicing economists, and many academics today as well, will respond that questions about the distribution of rights need to be addressed separately by others in other arenas, and that economics is just one input to the collective decision-making process. Yet the appropriate response within the economic framework would be to present policy makers with analyses indicating how different assignments of rights would affect people and how economic and non-economic activity would unfold over time. This would entail a deeper approach to economic analysis, more work for economists, and a new style of interaction with policy makers. Ecological economists have begun to take on this challenge.

Third, in the economic model, individuals have predetermined preferences and the problem of valuation is one of aggregating them to a collective preference. Individuals, however, are not born with preferences. They acquire them interacting in society and through advertising and other commercially funded media. Acknowledging that individual preferences are socially constructed opens up values analysis to the processes of their construction, how they have changed over time, which interests have influenced the changes, and whose ends are being served. It is well known that people respond differently as individuals than when as members of a group. This can be due to the social pressure of being within a group. It can also be due to the fact that many things, such as clean air and biological diversity, are very difficult to obtain acting as an individual, and so individuals, when choosing between alternatives, do not think about these things as options.

Ecological economists see individuals and groups as being able to draw on diverse multiple types of values and ethics depending on the context. Individuals have values and ethics rooted in their religion, sometimes more than one. They have values and ethics rooted in their particular cultural histories. And individuals and groups learn from their particular experiences and thereby develop values and ethics. Different values and ethical criteria are called forth depending on the setting, whether one is acting as an individual, a member of a family, within a profession, as a church member, as a citizen, etc. Because these categories are not exclusive and one is accountable in multiple, and not always foreseen, settings, weighing alternatives, individually and collectively, is considerably more difficult than a simple utilitarian calculus. It is also much more interesting (O'Connor, 2000).

In opposition to the western fact – value dichotomy, ecological economists are a little more prone to see facts and values as interlocked. There were no predefined preferences for biodiversity for example, before ecologists began to study the importance of biodiversity, and new technologies allowed us to better understand genetics in the last quarter of the 20th century. To be sure, we now recognize particular species and genetic traits as being important, as we have come to realize how something of prior value, say a medicinal drug, is obtained and must be protected. But such particulars are not biodiversity, though efforts to value biodiversity have certainly been attempted through aggregating particulars. Rather, we now value biodiversity *per se* even though there are multiple definitions of the term and we only know the number of species within an order of magnitude. In effect, biodiversity is a term like health, something that carries with it connotations of value as vague as the term itself, but very important to people nonetheless.

Ecological economists are interested in how social systems produce and reproduce some types of values and not others at different times. They are interested in how new mechanisms for making collective choice can be established that take advantage of expert advice, economic calculations, and group learning. Values juries seem like an interesting possibility. When ecological economists do an analysis that aggregates values in a single numeraire, they explore its strengths and weaknesses from the perspective of values having multiple realms.

Correcting the System of National Accounts

The primary creator of the SNA, Simon Kuznets, warned that gross domestic product (GDP) simply measures market and government activity, not well-being. Karl William

Kapp, an early ecological economist, documented how the accounts failed to consider natural resources appropriately as well as household labor (Kapp, 1950). Various economists, in response to the environmental concerns of the early 1970s, looked into the shortcomings of the SNA. Ecological economists, including Herman Daly, Robert Goodland, Roofie Hueting, Richard Norgaard, and Henry Peskin, participated during the 1980s in efforts hosted by the United Nations Environment Programme and the World Bank to correct the SNA. The goal was to treat the depletion of resource stocks like the depreciation of capital and to treat expenditures to correct expanding environmental problems as intermediate rather than final products. These were seen as simply better accounting practices. Some also hoped, however, that sufficient corrections could be made so that GDP, or some new aggregate measure, not only made a serious effort to measure economic well-being but also to account for whether that well-being could be sustained.

Working with Herman Daly, Clifford Cobb devised and undertook the empirical work to derive an *Index of Sustainable Economic Well-Being* (ISEW) (Daly and Cobb, 1989). The ISEW shows sustainable well-being increasing until around 1970, and then declining since. The ISEW, in different stages of evolution, has been calculated for many countries by non-governmental organizations (NGOs) with similar results. The ISEW has evolved further in the hands of *Rede ning Progress*, a US NGO, into the *Genuine Progress Indicator* (Cobb *et al.*, 1999). Even though rates of growth increased dramatically during the 1990s, the GPI declined because of increasing inequality, increased defensive environmental expenditures, and growth that has been at the expense of future generations (*see* ISEW (Index of Sustainable Economic Welfare) and GPI (Genuine Progress Indicator), Volume 5).

Since the 1980s, governments around the world have begun to incorporate some of the suggestions to correct their SNA. They have consistently taken the approach, however, of developing satellite accounts of environmental expenditures and stock resource depletion, and not used this information to adjust GDP or devise a new aggregate indicator. The rationale for developing but not incorporating environmental and resource accounts has been explicated by a study of the US National Research Council (Nordhaus and Kokkelenberg, 1999).

Whether one is devising a new aggregate indicator or simply establishing environmental and resource satellite accounts in monetary units, one must determine unique, defensible values for non-market goods and services. Since the 1980s, however, ecological economists have enriched their ranks and acquired a more complex understanding of values. This development makes it increasingly dif cult for ecological economists to recommend and defend any particular formal method for deriving non-market environmental values for standardized accounting purposes. Accounting for resource depletion using values derived from current resource prices, for example, may grossly under-represent the value of that depletion, for new extraction technologies may be keeping current prices low while their value to future generations remains high. There is also a concern that no one indicator will ever be adequate. Herman Daly is fond of saying that looking at GDP to plan for the future is like driving forward looking into the rearview mirror. While looking out the windshield is a major improvement, those working on indicators of human well-being, environmental well-being, and sustainability are increasingly talking about a dashboard of indicators. Nevertheless, most ecological economists support environmental accounting as an improvement over not accounting for the environment. Ecological economists also hope that the process of deriving and disseminating environmental accounting information will help lead people to ask deeper questions.

Biophysical-Economic System Analysis

The efforts to link the natural sciences with economics are numerous, with none being dominant. At best, a cursory summary of the work of ecological economists can be provided. First, as methodological pluralists, ecological economists rely on the natural sciences in order to understand the limits of particular economic models and assumptions, and they accept the ndings of biophysical analyses on their own terms, as additional information of equal importance as that stemming from economic analysis. Second, ecological economists have linked, and in some cases integrated ecological and economic models. Third, the concept of joint production from economics has been developed as a link between economics and the natural sciences. Fourth, ecological economists have extended models from the natural sciences to economics, dispensing with traditional models of economics altogether. Each of these needs elaboration.

Ecological economists often compare the differences in the insights derived from natural science and economics. Gowdy (1994), for example, has explored the concept of coevolution in biology, how it relates to cultural ecology in anthropology, restated institutional economics in coevolutionary terms, and compared all of this with conventional market analyses. Malte Faber and others have further developed the ideas of Nicholas Georgescu-Roegen, arguing that, as we rely less and less on current solar energy and more and more on fossil hydrocarbon and other mined resources, economic activity needs to be understood as moving toward a closed system and subject to the second law of thermodynamics (Faber *et al.*, 1996). This means that we need to understand economic activity as an entropic process, as taking high quality energy and materials and reducing them to waste heat and dispersed minerals. Looking at new technology from this perspective, we see that it allows economic

activity to accelerate, for entropy to increase faster, leaving less for the future, not more. Thus, to the extent that we rely on stock resources, especially of energy, progress is not possible, indeed even sustainability is not possible. The alternative, of course, is to return to relying on current solar energy, in part by working with natural systems. This, however, raises further issues.

Humans are co-opting approximately 40% of the primary energy captured by terrestrial ecosystems (Vitousek *et al.*, 1986). Ecological economists use this study as an indication of how human activity has transformed the environment, and as an indicator of the limits to further transformation. For Herman Daly and other ecological economists, the analysis of Vitousek stresses the urgency of moving to a steady state while there is still some margin left (Daly, 1996).

Wackernagel and Rees (1996) have used biophysical analysis to calculate the *ecological footprint* of a population, the area of productive land and water needed to produce the resources it consumes and assimilate the wastes it produces. Ecological-footprints range from approximately 0.5 hectares per person for people in very poor countries to 10 or more hectares for average Americans; the rich use even more (*see* **Ecological Footprint**, Volume 3). Biophysical analyses provide critical insight into our use of stock resources (Cleveland and Kaufman, 1991), into carrying capacity, and into the difficulties of moving toward life-styles consistent with carrying capacity (Folke, 1991; Kaberger, 1991; Jannson, 1991). While biophysical indicators create considerable controversy within ecological economics when their supporters argue that they are suitable for making policy directly, they contribute to how ecological economists understand the relationships between people and nature.

Second, Volterra's logistic growth curve has long been incorporated in the economics of forests and fisheries. Going beyond this, but still building on the heuristic models of population biology, the field of bioeconomics arose with the mathematical models of Colin Clark (Clark, 1976; Conrad and Clark, 1987). Ecological economists have pushed further in this direction. The relationship of biological concepts to ecological economics have been explored: ecological stability (Common and Perrings, 1992); ecological resilience (Brown and Roughgarden, 1995), and biological diversity (Norgaard, 1987; Perrings *et al.*, 1995). Empirical research has also linked ecological systems and economic choices (Costanza *et al.*, 1989).

Third, some ecological economists argue that the idea of joint production from economics provides an adequate bridge to the natural sciences. A petroleum refinery, for example, produces a range of products from jet fuel to tar, jointly in the sense that it is difficult to get one without the other. All production processes, however, jointly produce good things and bad things. Economists have simplified their models by only looking at the good things produced rather than the total set of desired and undesired outputs that show consistency with the first and second laws of thermodynamics. While economists have downplayed joint production as a relatively rare situation, from an ecological economic perspective, the concept of joint production provides a critical connection between economics and understanding from the natural sciences (Baumgartner *et al.*, 2000).

Fourth, a few ecological economists have dispensed with economic models altogether and used methods rooted in the natural sciences to understand the human predicament. For example, Howard Odum (1996) has argued that the concept of emergy , defined as the energy used up in an ecological or economic process, can be used to trace costs through ecological and economic systems, providing better measures of value and wealth. Odum argues that emergy analyses provide a scientific basis for choosing between alternative technologies and development projects (*see* **Emergy**, Volume 5). On the other hand, Richard Norgaard pursued a coevolutionary explanation of change to a conclusion that substantially changes how we treat expert knowledge. He argued that social and environmental change can be understood as a coevolutionary process between value, knowledge, organizational, technological, and environmental systems. Characteristics within each system select upon the fitness of characteristics in the other systems such that they coevolve together, reflecting each other, and becoming tightly interlocked. But change does occur through biological mutations, cultural innovations, and introductions from other ecosystems and cultures (Norgaard, 1994). Thus, collective action can best be based on a process of shared knowledge between scientists, those with traditional and experiential knowledge, and those who will actually be implementing the change.

THE FUTURE OF ECOLOGICAL ECONOMICS AND RELATED ALTERNATIVES

The future of ecological economics depends on three broad issues. The first is whether ecological economics will evolve into a new overarching paradigm or whether it will remain an umbrella under which multiple, contradictory analyses are undertaken. This has important ramifications with respect to how ecological economics is put into practice. The second issue is whether ecological economics will become a field that is embraced by scholars in the developing world as it has historically developed around scholars in Europe and North America. The third issue is how the scope of ecological economics will be affected by the development of other transdisciplinary fields. Of particular importance are the developments in agroecology, conservation biology, ecological engineering, and industrial ecology starting from the natural sciences, and the study of

the commons and political ecology starting from the social sciences. At the same time, the fields of economics and ecology themselves continue to evolve and will also thereby affect the future of ecological economics.

New Overarching Paradigm or Continued Methodological Pluralism?

Among ecological economists, there is already considerable agreement on underlying assumptions and on the need to learn across models and disciplinary cultures. This agreement itself indicates ecological economics is a new overarching process, if not a paradigm. Many ecological economists from its beginnings, however, have pleaded for greater unity (Daly, 1984; Proops, 1989). To them, an overarching model, or perhaps a small set of broad models, is needed to understand, not simply describe, the big picture and to guide practitioners to appropriate particular models. Ecological economists need to establish a new language. Simply incorporating the languages of multiple disciplines, with the same words having different historical meanings, makes it difficult to communicate to the public and policy makers. Taking ecological economics to public and policy arenas has, thus far, entailed hiding its richness and concentrating on correcting the SNA or making arguments for reduced material and energy flows. The US Environmental Protection Agency made a small effort to understand how taking a broader approach to valuation might work (Bingham *et al.*, 1995), but has not implemented any of the recommendations. Conventional economic thinking is well entrenched in political discourse and agency decision-making in part because it is simple and consistent. It is difficult to imagine how a field as eclectic as ecological economics can have an impact without having comparable coherence.

On the other hand, continuing the methodological pluralism of ecological economics is not a default position for its advocates. Rather, the pluralists have compelling arguments too. First, many scholars have sought unity within the natural sciences and within the social sciences without success, so the chances of unity across them are slim indeed. Methodological pluralism is a reality. At least for the foreseeable future, ecological economists should not pin their hopes on accomplishing something so difficult as unifying either the natural or the social sciences, let alone the combination of the two. Second, agreeing on less than a truly overarching model would entail reducing the scope of inquiry of ecological economics. This would make the field more vulnerable to mistakes and surprise in the future. Third, reducing ecological economics so that it fits into the existing policy process is accepting defeat from the beginning. The existing political discourse is dominated from the top by those who have money. It needs reinvigorating with a new paradigm that is more democratic. Ecological economists, as pluralists, are more open to discussing a multiplicity of values and incorporating the voices of those with traditional and experiential knowledge. With respect to the policy process, the existing agencies are set up to act on advice from experts speaking separately. Inter-agency task forces, science juries, and other interactive consensus-building innovations are opening up this process, and it is through these and further evolutions (Irwin, 1995; Stern and Fineberg, 1996), perhaps to full discursive democracy (Dryzak, 1990; Fishkin, 1991), that the richness of ecological economic understanding can be effective. Fourth, different methods inherently stress different values. By focusing on different things, each model says, this is important , and, by implication, that is not . The multiple languages, especially of the social sciences, also reproduce and transmit values from different historical and cultural settings. In this sense, the plea for cohesion in models and language is a plea for historical denial and cultural extinction.

Ecological Economics as a Global Enterprise

Ecological economics is becoming the preferred base for economists and other scholars working to implement sustainable development in economically poor countries. Scientists trained in western ways of thinking must overcome strong cultural barriers as they work in poor rural communities. From this perspective, disciplinary boundaries are but a minor nuisance. And implementing sustainable rural development clearly requires input from multiple disciplines. Scholars from the economically poorer countries are becoming stronger and more assertive as their numbers increase and as the economic gap between rich and many, though certainly not all, poor countries increases. The shared vision of development, with each nation developing a full complement of economic sectors and public services, has collapsed to a northern vision of globalization. North–South tensions over the amelioration of climate change and biodiversity loss have arisen within the dominant conceptual and institutional frameworks. Ecological economics, with its emphasis on equity and respect for the diversity of ways of knowing, has a tremendous opportunity. Ecological economics could become a global network of scholars learning together, building a global vision for the future that respects diversity while demonstrating how that vision could be achieved.

Other Alternatives to Economics

In addition to the work of ecological economists, there are other efforts to affect how a narrow pattern of reasoning from economics has dominated modern policy and politics. Like ecological economics, most of the other alternatives also span multiple disciplines. Social scientists have joined in the study of *common property management*

regimes. Economists and sociologists have joined to understand *socio-economic systems* together. Political economy, the study of power that has evolved across the social sciences since Marx, is now complemented by the emerging field of *political ecology*. There are similar efforts centered in the natural sciences that are reaching across to the social sciences. For example, biologists concerned with the conservation of biological diversity are anxious to work with economists. Academics and practitioners are organizing around the ideas of *ecosystem health*, *industrial ecology*, and *ecological engineering*. Many ecological economists also participate in these other interdisciplinary efforts. And like ecological economics, most of these efforts are questioning the conventional assumptions of how science relates to the policy process. While it is clear that ecological economics is clearly focused on the interplay between ecological and economic understanding, all of the emerging interdisciplinary efforts ask related questions to some extent. Thus the future of ecological economics depends on how numerous parallel and overlapping interdisciplinary efforts evolve in the context of each other.

REFERENCES

Baumgartner, S, Dyckhoff, H, Faber, M, Proops, J, and Schiller, J (2000) *The Concept of Joint Production and Ecological Economics*, Discussion Paper Series No. 319 (June), Department of Economics, Universitat Heidelberg.

Bingham, G, Bishop, R, Brody, M, Bromley, D, Clark, E, Cooper, W, Costanza, R, Hale, T, Hayden, G, Kellert, S, Norgaard, R B, Norton, B, Payne, J, Russell, C, and Suter, G (1995) Issues in Ecosystem Valuation: Improving Information for Decision Making, *Ecol. Econ.*, 14, 73–90.

Boulding, K E (1966) The Economics of the Coming Spaceship Earth, in *Environmental Quality in a Growing Economy*, ed H Jarrett, Resources for the Future, Johns Hopkins University Press, Baltimore, MD.

Boulding, K E (1978) *Ecodynamics: A New Theory of Societal Evolution*, Sage Publications, Beverly Hills, CA.

Brown, G and Roughgarden, J (1995) An Ecological Economy: Notes on Harvest and Growth, Chapter 5 in *Biodiversity Loss: Economic and Ecological Issues*, eds C Perrings, K-G Maler, and C Folke, Cambridge University Press, Cambridge.

Christensen, P P (1989) Historical Roots for Ecological Economics – Biophysical Versus Allocative Approaches, *Ecol. Econ.*, 1(1), 17–36.

Clark, C W (1976) *Mathematical Bioeconomics: The Optimal Management of Renewable Resources*, Wiley-Interscience, New York.

Cleveland, C J and Kaufman, R K (1991) Forecasting Ultimate Oil Recovery and Its Rate of Production: Incorporating Economic Forces into the Models of M. King Hubbert, *Energy J.*, 12(2), 17–46.

Cobb, C, Goodman, G S, and Wackernagel, M (1999) *Why Bigger is Not Better: The Genuine Progress Indicator - 1999 Update*, Redefining Progress, San Francisco, CA.

Common, M and Perrings, C (1992) Towards an Ecological Economics of Sustainability, *Ecol. Econ.*, 6(1), 7–34.

Conrad, J M and Clark, C W (1987) *Natural Resource Economics: Notes and Problems*, Cambridge University Press, Cambridge.

Costanza, R (1984) Natural Resource Valuation and Management: Toward an Ecological Economics, in *Integration of Economy and Ecology: An Outlook for the Eighties*, ed A M Jansson, Proceedings from the Wallenberg Symposia, Stockholm.

Costanza, R (1989) What is Ecological Economics? *Ecol. Econ.*, 1(1), 1–7.

Costanza, R, Farber, S C, and Maxwell, J (1989) Valuation and Management of Wetland Ecosystems, *Ecol. Econ.*, 1, 335–361.

Costanza, R, Cumberland, J, Daly, H, Goodland, R, and Norgaard, R B (1997) *An Introduction to Ecological Economics* (an Intermediate Level College Text), International Society for Ecological Economics and St. Lucie Press, FL.

Costanza, R, Perrings, C, and Cleveland, C J (1997) *The Development of Ecological Economics*, Edward Elgar, Cheltenham.

Daly, H E (1973) The Steady-State Economy: Toward a Political Economy of Biophysical Equilibrium and Moral Growth, in *Toward a Steady-State Economy*, eds H E Daly and W H Freeman, San Francisco, CA.

Daly, H E (1984) Alternative Strategies for Integrating Economics and Ecology, in *Integration of Economy and Ecology: An Outlook for the Eighties*, ed A M Jansson, Proceedings from the Wallenberg Symposia, Stockholm.

Daly, H E (1996) *Beyond Growth: The Economics of Sustainable Development*, Beacon Press, Boston, MA.

Daly, H E and Cobb, J B, with contributions by Cobb, C W (1989) *For the Common Good: Redirecting the Economy Toward Community, the Environment, and a Sustainable Future*, Beacon Press, Boston, MA.

Dryzek, J S (1990) *Discursive Democracy: Politics, Policy and Political Science*, Cambridge University Press, Cambridge.

Dryzek, J S (2000) *Deliberative Democracy and Beyond: Liberals, Critics, Contestations*, Oxford University Press, Oxford.

Faber, M, Manstetten, R, and Proops, J (1996) *Ecological Economics: Concepts and Methods*, Edward Elgar, Cheltenham.

Fishkin, J S (1991) *Democracy and Deliberation: New Directions for Democratic Reform*, Yale University Press, New Haven.

Folke, C (1991) Socio-Economic Dependence on the Life-Supporting Environment, Chapter 5 in *Linking the Natural Environment and the Economy: Essays from the Eco-Eco Group*, eds C Folke and T Kaberger, Kluwer, Dordrecht.

Foster, J B (1999) Marx's Theory of Metabolic Rift: Classical Foundations for Environmental Sociology, *Am. J. Socio.*, 105(2), 366–405.

Funtowicz, S and Ravetz, J R (1991) A New Scientific Methodology for Global Environmental Issues, in *Ecological Economics: The Science and Management of Sustainability*, ed R Costanza, Columbia University Press, New York.

Georgescu-Roegen, N (1971) *The Entropy Law and the Economic Process*, Harvard University Press, Cambridge.

Gowdy, J (1994) *Coevolutionary Economics: The Economy, Society and the Environment*, Kluwer, Boston, MA.

Howarth, R B and Norgaard, R B (1995) Intergenerational Choices Under Global Environmental Change, Chapter 6 in *The*

Handbook of Environmental Economics, ed D W Bromley, Basil Blackwell, Oxford.

Hueting, R (1980) *New Scarcity and Economic Growth: More Welfare through Less Production?* Amsterdam. North Holland.

Irwin, A (1995) *Citizen Science: A Study of People, Expertise and Sustainable Development*, Routledge, London.

Jansson, A M, ed (1984) *Integration of Economy and Ecology: An Outlook for the Eighties*, Proceedings from the Wallenberg Symposia, Stockholm.

Jansson, A M (1991) Ecological Consequences of Long-Term Landscape Transformations in Relation to Energy Use and Economic Development, Chapter 6 in *Linking the Natural Environment and the Economy: Essays from the Eco-Eco Group*, eds C Folke and T Kaberger, Kluwer, Dordrecht.

Kaberger, T (1991) Measuring Instrumental Value in Energy Terms, Chapter 4, in *Linking the Natural Environment and the Economy: Essays from the Eco-Eco Group*, eds C Folke and T Kaberger, Kluwer, Dordrecht.

Kingsland, S E (1985) *Modeling Nature: Episodes in the History of Population Ecology*, University of Chicago Press, Chicago, IL.

Kneese, A V, Ayres, R U, and d'Arge, R C (1970) *Economics and the Environment - A Materials Balance Approach*, Johns Hopkins University Press, Baltimore, MD.

Maler, K G (1974) *Environmental Economics: A Theoretical Inquiry*, Johns Hopkins University Press, Baltimore, MD.

Martinez, A J with Schlupmann, K (1987) *Ecological Economics: Energy, Environment, and Society*, Basil Blackwell, Oxford.

Meadows, D H, Meadows, D L, Randers, J, and Behrens, III, W W (1972) *The Limits to Growth*, Universe Books, New York.

Murdoch, W W (1984) Ecology and Economics: Integration and a Common Theoretical Problem, in *Integration of Economy and Ecology: An Outlook for the Eighties*, ed A M Jansson, Proceedings from the Wallenberg Symposia, Stockholm.

Nordhaus, W D and Kokkelenberg, E C, eds (1999) *Nature's Numbers: Expanding the National Economic Accounts to Include the Environment*, National Academy Press, Washington, DC.

Norgaard, R B (1987) Economics as Mechanics and the Demise of Biological Diversity, *Ecol. Model.*, 38 (September), 107–121.

Norgaard, R B (1989) The Case for Methodological Pluralism, *Ecol. Econ.*, 1(1), 37–57.

Norgaard, R B (1994) *Development Betrayed: The End of Progress and a Coevolutionary Revisioning of the Future*, Routledge, London.

Norgaard, R B (1997) A Coevolutionary Environmental Sociology, in *The International Handbook of Sociology*, eds M Redclift and G Woodgate, Edward Elgar, Cheltenham, UK.

O'Connor, M, ed (2000) Social Processes of Environmental Valuation, Special Issue, *Ecol. Econ.*, 34(2).

Odum, H T (1971) *Environment, Power, and Society*, John Wiley & Sons, New York.

Odum, H T (1996) *Environmental Accounting: Emergy and Environmental Decision Making*, John Wiley & Sons, New York.

Passet, R (1979) *L'economique et le Vivant*, 2nd edition 1996, Economica, Paris.

Perrings, C, Maler, K-G, Folke, C, and Holling, C S (1995) *Biodiversity Loss: Economic and Ecological Issues*, Cambridge University Press, Cambridge.

Pillet, G and Odum, H T (1987) *Energie, Ecologie, Economie*, Georg Editeur SA, Geneva.

Pimentel, D (1974) Energy Use in World Food Production, *Environ. Biol. Rep.*, 74–1, Cornell University, Ithaca, NY.

Proops, J L R (1989) Ecological Economics: Rationale and Problem Areas, *Ecol. Econ.*, 1(1), 59–76.

Soderbaum, P (2000) *Ecological Economics: A Political Economics Approach to Environment and Development*, Earthscan, London.

Stern, P C and Fineberg, H V, eds (1996) *Understanding Risk: Informing Decisions in a Democratic Society*, National Academy Press, Washington, DC.

Vitousek, P M, Ehrlich, P R, Ehrlich, A H, and Matson, P (1986) Human Appropriation of the Products of Photosynthesis, *BioScience*, 36(6), 368–73.

Wackernagel, M and Rees, W (1996) *Our Ecological Footprint*, New Society Publishers, Gabriola Island, British Columbia, Canada.

Environmental Politics

ROBERT PAEHLKE

Trent University, Peterborough, Canada

Many analysts date contemporary environmental politics, especially in North America, from the 1992 publication Silent Spring, or from the first Earth Day (April 22, 1970). The historic roots of environmental politics of course run deeper, but it is first important to distinguish environmental politics from all other forms and aspects of politics. One way to do this is to draw a distinction between environmental politics and distributive (also called redistributive) politics. Distributive politics is about the involvement of government in the distribution and redistribution of economic and social benefits and disbenefits among the individuals, groups, interests, and regions of any given polity. It is also about the social and economic management inherent in the maximization of the total economic pie to be distributed. Environmental politics, in contrast, primarily focuses on the unintended negative consequences of this latter pursuit, on the protection of human health, on the well-being of other species and on the long-term future of human generations, or on such other non-monetary values as wilderness.

Given this focus, which favors not so much capital or labor, or rich or poor, environmentalism is sometimes said to be neither left, nor right (Paehlke, 1989). There is much truth to this claim despite the fact that many on the political right see environmentalism as a variant of the left, notwithstanding that are environmentalisms of the left, right and center. Left environmentalism looks to regulate corporate polluters and to democratize within the public sphere significant segments of economic decision making. Right environmentalism rather favors what are called market-based policy tools, worries a great deal about population (though it may simultaneously reject choice regarding abortion), and believes that the market is the best means of protecting the environment (through green products and other means of feedback to private economic decisions). Centrist environmentalism is inclined to some mix of these polar perspectives.

Environmental politics has, essentially, placed three distinct value sets into competition with other societal objectives such as profit maximization or short-term economic growth. The three value sets are: (1) wilderness, biodiversity, habitat and nature; (2) pollution reduction and the protection of human health from environmental hazards; and (3) sustainability (Paehlke, 1997). Each of these has a deep history preceding the beginnings of contemporary environmental politics; those beginnings in effect identify when environmental politics became more integrated and rose to a significant place within the wider political agenda-setting process. After 1960 environmental issues came to have real resonance for significant numbers of people, were regularly reported in the media, and soon thereafter led to a wide array of relevant legislative initiatives and administrative points of contact. Long before that point, however, many of these values were advanced by an earlier conservation movement or put forward by thoughtful individuals.

Arguably, broad-based environmental politics only became possible when and where a certain level of material prosperity had been achieved. As Samuel Hays (1987), the US environmental historian, noted:

Evolving environmental values were closely associated with rising standards of living and education.… But with rising incomes something beyond necessities and conveniences now lay within the reach of many; they may be called amenities.… Environmental quality was an integral part of this new search for a higher standard of living.

(Hays, 1987)

Thus, from the 1950s onwards the amenities associated with prosperity came more and more to include outdoor recreation, air and water quality, and the protection of the broader-than-economic well-being of future generations.

ENVIRONMENTAL VALUES: THE BASIS OF ENVIRONMENTAL POLITICS

Concern for the rst of the three dimensions of environmental values, the preservation of wilderness and the conservation of nature, has deep roots, especially in North America, the US in particular. Conservation moved from the realm of ideas into the realm of environmental politics and policy in the late 19th century with the campaigns for and the creation of national parks, in Yellowstone (1872) and Yosemite National Parks in the US and Banff National Park in Canada (1885). Indirectly related to those events were the important writings of George Perkins Marsh, John James Audubon, and, perhaps especially, Henry David Thoreau. John Muir, a founder of the Sierra Club in 1892, contributed enormously to the development of the idea of wilderness and to the campaign to preserve it.

Marsh (1801 1882), a Vermont lawyer and politician who traveled extensively, was one of the rst to systematically study the many ways in which humans have disrupted nature (*see* **Marsh, George Perkins**, Volume 3). Late in life, he produced a remarkable book that in many ways foreshadowed by a century the modern environmental movement. The book, (Marsh, 1965) *Man and Nature* in 1864, argued that humans had gained great power to alter nature and were doing so to the detriment of both nature and themselves. Marsh cited and documented the effects of deforestation, erosion, loss of biodiversity (he did not, of course, use that modern word), altered water regimes, and general loss of biological productivity. He cited examples from his native Vermont, from Europe and Turkey (where he had served as Ambassador) and argued that the destruction of wilderness in the pursuit of human gain was imposing too many costs. Parallel, but less documented, expressions of concern were advanced at the time by artist John James Audubon (1785 1851) and Henry David Thoreau (1817 1862).

These early expressions of concern were heard, but the time was not ripe for environmental politics in practice which could lead to effective public action. At the time most North Americans were doing all they could to survive through agricultural pursuits which in turn required the clearing of forests, not their preservation. The writings of these authors are revered today as profound early insights, providing historic depth and richness to the contemporary environmental movement. The works of all three, from Audubon s paintings of birds to Thoreau s literary masterpiece *Walden*, served to inspire those who followed.

Among those inspired by Thoreau s life and writings was John Muir (1838 1914). Muir took things further politically than did Thoreau. Also a gifted writer about nature, the Scottish born Muir, having explored vast tracts of America, including Florida and California on foot, published widely-read essays in the leading magazines of the day. In the 1880s Muir successfully campaigned for the establishment of Yosemite National Park in California. Muir s Sierra Club remains today one of the largest and most in uential of North American environmental organizations. Earth Day is celebrated on a date near to John Muir s birthday. In the early part of the century, conservationist in uence in the US reached even into the White House, especially during the presidency of Theodore Roosevelt.

Leading European gures in the early history of environmental thought include Gilbert White of Selbourne, England who provided, in the mid-18th century, one of the earliest detailed accounts of the complex associations and interactions within nature, a precursor to the modern science of ecology. Notable in the 19th century, as precursors of conservationism and as promoters of a poetic and aesthetic appreciation of nature, were Wordsworth, Ruskin, and Morris. This sensibility informed the formation of Great Britain s National Trust and the efforts to protect both the built and the natural environment. Equally important to the eventual development of environmental politics were political philosophers including John Stuart Mill, who was the rst to discuss steady state economics and Edmond Burke, whose organic vision of community and tradition even today lends support to political thinking regarding the environment. Finally here one must mention Svante Arrhenius, the Swedish chemist, who hypothesized in 1896 that the carbon dioxide produced by burning coal could ultimately contribute to warming the climate of the planet.

The growth in human numbers and the spreading and sprawling of human settlement associated with industrialization and motorized transportation throughout the 19th and 20th centuries continuously and dramatically reduced the quality of habitat for wild plant and animals. Humans came to occupy and utilize a very high proportion of what was once wild space. Much of North America (excluding Canada s North), and virtually all of Europe and the rest of the world, is now a human-controlled landscape. Telecommuting and other new communications-based technologies may accelerate the dispersion of human habitation. As this process has advanced through the centuries, many have come to realize that only conscious conservation and careful land use policies can keep any signi cant portion of the planet open to wild non-human species and spaces. Two additional developments in the idea of conservation came with Aldo Leopold s land ethic and, later, the widely in uential ideas of deep ecology, especially the thinking of the Norwegian philosopher Arne Naess.

Leopold, an early forestry graduate from Yale University, worked as a forester and wildlife manager. As he grew older he doubted more and more the utilitarian conservation philosophy which he had been taught and had practiced. Leopold developed a philosophy rooted in the science of ecology, in the interconnections between animals, plants, water and soil; he envisioned humankind as an ordinary member, rather than a conqueror, of the community of the

land. Naess, and other deep ecologists writing in the 1970s, continued the development of an ecologically based ethic. They distinguished between a human-centered (anthropocentric) approach to human-nature relationships from an eco-centric (deep ecology) approach. Essentially, the latter views ethical issues from non-human viewpoints and has formed the basis for the animal rights movement, the radical defense of wilderness and the rejection of those forms of conservation based in resourcism wherein all nature is valued solely in terms of the needs of humans (*see* Leopold, Aldo, Volume 5).

The ideas underlying the conservation movement prior to the 1960s were primarily focused on the conservation and preservation of resources, wilderness and nature's diversity. However, in the early days of the environmental movement the primary emphasis within environmental politics shifted to pollution and the effects of pollution and toxic chemicals on human health. These new concerns had their own history, but prior to the 1960s did not often capture the public imagination or a place on the wider political agenda. In contrast, in the 1960s, many of what had come to be called environmentalists saw traditional conservation and wildlife protection organizations as too conservative, as not sufficiently aware of the more important dimensions of the environmental crisis. The environmental movement looked more to urban concerns and did thereby reach, politically, a broader segment of the by then urbanized population than had the conservation movement. In effect, only some people care about nature and other species, but everyone can be aroused by concern with the air humans breathe, the food we eat, and the water we drink.

Historically, public, medical and political concern regarding air or water pollution can be traced to the ancient Persians and Romans and is even mentioned obliquely in the Christian Bible at Deuteronomy 23: 12–13. In 1273 restrictions were placed on coal burning in the city of London and in 1661 the Englishman John Evelyn submitted a tract entitled *Fumifugium* regarding air quality in that city to Charles II. In 1713 the Italian physician Bernardino Ramazzini published a compendium of diseases related to toxic exposures and other hazards in the workplace and in 1775 Percivall Pott discovered an epidemiological link between coal tar exposure in British chimney sweeps and a higher incidence of scrotal cancer. In the 19th century there was limited legislative action regarding pollution in both Great Britain and the US (in mid-19th century in Britain, in 1876 in St. Louis and in 1881 in Chicago). Smokestacks were required, though pollution abatement as we now know it was not an available option. At the same time some water pollution in rivers was the result of the removal of open sewage from public streets and lanes.

Only in the 20th century did pollution abatement progress significantly, but that progress was in part the result of an even greater need. Human population swelled, pesticides were more widely utilized, industrial capacity swelled enormously, and the array of toxic industrial chemicals broadened (especially with the massive post-WWII expansion of petroleum-based chemical industries). The environmental politics of pollution first took hold with large numbers of people in the 1920s with regard to the risk of acute poisoning from pesticides such as lead arsenate and other arsenic-based agricultural chemicals. These concerns ran high in Europe and in the US and grew larger with the publication in the 1930s of two sensational best-sellers: *One Hundred Million Guinea Pigs* and *American Chamber of Horrors*. As a result, residues of arsenic-based insecticides on foods were regulated in the US in 1932, but those standards of protection were reduced later in the decade as overt public fears subsided. Soon afterwards these toxic pesticides were replaced with 1,1,1-trichloro-2,2-*bis*-(p-chlorophenyl)ethane (DDT), hailed at first as a miracle chemical.

It was DDT and related pesticides that played a significant role two decades later in the development of the contemporary environmental movement. With DDT and related compounds, the concern was less with direct toxicity of residues for human health, but rather with bioaccumulation within food chains and the resultant ecological disruptions. Many organic chemicals are persistent in the environment and accumulate in the bodies of virtually all animal life, in ever higher concentrations the higher one goes in the food chain (from insects to small fish to larger fish to eagles, for example). As Rachel Carson pointed out in *Silent Spring* anti-mosquito use of 1,1-dichloro-2,2-*bis*-(p-chlorophenyl)ethane (DDD) at 0.02 parts per million (ppm) resulted in lake plankton with 5 ppm, small fish with 40–300 ppm, and large fish with up to 2500 ppm. At 1600 ppm the fish-eating birds around the lake died out, at a time when the chemical was no longer detectable, even in trace amounts, in the lake. Contemporary environmental politics has been very much about the sciences of toxicology, epidemiology and ecology.

What makes Rachel Carson a founding figure in environmental politics is her ability to integrate diverse small scientific findings and to array that data and present it to a wider public of non-experts in a careful and forceful way. Carson was not afraid to draw conclusions and to do so in a very public way. Despite massive attacks on her science and her credibility by industry representatives, her views held up and a wide range of policy initiatives ensued (and a new dimension of politics was born) (*see* Carson, Rachel Louise, Volume 5). Concern with pollution in the public mind was also fueled by two acute air pollution disasters (in Donora, Pennsylvania in 1948 and in London, England in 1952), and perhaps as well by general fears related to then, widespread above ground nuclear weapons testing. Environmental politics arose out of these factors and soon eclipsed conservation concerns in terms of public attention.

The third strand of environmental ideas and values, sustainability concerns, has also had a long history. The focus on sustainability values goes back at least to the writings of Thomas Robert Malthus (1766–1834). Malthus was a moralist who concluded that human population would inevitably outstrip humankind's ability to produce sufficient food; as he saw it, the former increased geometrically (or exponentially), while the latter increased arithmetically (or incrementally). Malthus was the ultimate naysayer to the boundless optimism in the liberal intellectual circles of the time wherein it was imagined that progress and increasing bounty were natural characteristics of the human condition. Where liberals believed in the perfectibility of humans, education, science and democracy, Malthus was a pessimist who only hoped that human depravity and overpopulation could be checked sufficiently to avoid the worst of resource shortfalls and misery. Misery for the poor, however, was seen as one of three necessary checks on excessive human population growth; Malthus opposed the Poor Laws of the time as overly generous.

Few contemporary sustainability advocates are eager to embrace Malthus as an intellectual heir, neither are they enthusiastic about W Stanley-Jevons, who published *The Coal Question* in 1865. This book presented a calculation of the rise in coal consumption in Great Britain between 1854 and 1864 (at 3.5% annually) and concluded, given known domestic reserves, that the burgeoning nascent industrial society was not likely to be long-lived. In his words:

> we cannot long maintain our present rate of increase of consumption ••• we can never advance to the higher amounts of consumption supposed ••• the check on our progress must become perceptible within a century ••• the cost of fuel must rise, perhaps within a lifetime, to a rate injurious to our commercial and manufacturing supremacy; and the conclusion is inevitable, that our present happy condition is a thing of limited duration.
>
> (Jevons, 1965, 271)

Jevons did not allow for the rise of an industrial economy based on other fossil fuels, traded globally and extracted even from under the sea.

As off the mark as both Malthus and Jevons were, they were right that resources are available in only finite amounts, that there is some limit to the human population that can be supported by nature's capacities, and that resource limits can restrain economic development and social well-being. Where they were wrong was in assuming that, inevitably, limiting outcomes would follow from shortages in one particular resource at one given level of technological development. They were also too early to appreciate the ecological and human health implications associated with continuous and simultaneous growth in human population, resource extraction and industrial production. It is not so much that the human population and total economic output cannot possibly grow, but that they can and that the price paid is not always worthwhile it in net terms; and without technological selectivity and careful planning. That is a contemporary view of sustainability, though there is far from unanimity, even among environmentalists, regarding such matters.

Two other, more modern, precursors of the sustainability debate are the American conservationists Fairfield Osborn and Samuel Ordway. Osborn spoke of the goal of humanitarianism being, not the quantity but the quality of living (Osborn, 1953, 226). Population restraint and resource conservation in the name of a higher quality of life were central objectives for both Osborn and Ordway. Ordway (1953, 31), feared that without careful use:

> ••• basic resources will come into such short supply that rising costs will make their use in additional production unprofitable, industrial production will cease, and we shall have reached the limit of growth. If this limit is reached unexpectedly, irreparable injury will have been done to the social order.

The solution: restraint and perhaps redirection of human material wants. In Osborn's words, we ••• must temper (our) demands and use and conserve the natural living resources of the earth (Osborn, 1948, 201). In Ordway's Our needs can be supplied if our wants are bridled ••• The false ideology which worships unlimited expansion must go (Ordway, 1953, 39).

These views, penned and not widely influential in the hyperexpansionist 1950s, when each year's new automobiles were larger and less fuel efficient, would have had very great resonance within the environmentalist politics of the 1970s. Only then did these three strands of environmental thought, ecology, health, and sustainability, come together in an integrated way under a banner of environmental politics. Only perhaps from the late 1960s were these values widely important for more than fleeting periods of time. Only after Earth Day (1970) could one speak of a significant environmental movement and environmental politics as a important force within the whole of society in a number of nations.

SINCE EARTH DAY: HIGHLIGHTS OF RECENT ENVIRONMENTAL POLITICS

Following the publication of *Silent Spring*, public concern with pesticides, pollution and environmental protection continued to grow though these concerns were all but overwhelmed at the time by concern regarding civil rights (especially in the US) and opposition to the war in Vietnam (throughout the world). However, by the late 1960s concerns such as the accelerating growth of the human population worldwide and incidents such as the massive oil release from the Torrey Canyon off England in 1967 and the Santa Barbara offshore oil blowout in 1969, opposition to airport noise and the proposed Concorde and supersonic transport (SST) jetliners, threats to whale populations

from heavy hunting with advancing technologies, concerns regarding nuclear power and nuclear testing, combined with a number of conservation issues, especially the proposed construction of a ski resort in Mineral King Canyon (in California) by the Disney Corporation and the proposed construction of a dam at Lake Pedder in the Australian state of Tasmania, accelerated the development of environmental politics. One important outcome of this new concern was the passage in 1969 of the pioneering piece of legislation in the US, the National Environmental Policy Act (NEPA) which introduced environmental impact assessment, a procedure soon widely adopted.

Another result of, and a further stimulus to, the growing salience of environmental politics was the first Earth Day (April 22, 1970). US Senator Gaylord Nelson, a Democrat from Wisconsin, proposed the application of the teach-in techniques of the peace movement to rising environmental concerns. The results were remarkable. *Time Magazine* estimated that 20 million Americans took part, others that Earth Day events involved individuals in 1500 universities and 10 000 schools. There were simultaneous demonstrations in major cities and new organizations and publications arose in the year leading up to Earth Day and the year following. New, still significant, activist organizations included Pollution Probe in Canada, Environmental Action in Washington, DC, and the Natural Resources Defense Council in Washington, DC. Also newly created within this two-year period were the (US) Environmental Protection Agency and Environment Canada. The early 1970s witnessed an array of strong new legislation within many jurisdictions.

Following on NEPA and Earth Day, the US passed a new Clean Air Act in 1970 and a Clean Water Act in 1972. The Safe Drinking Water Act was passed in 1974 and the Resource Conservation and Recovery Act in 1976. Canada passed its Clean Water Act and Fisheries Act amendments in 1970; its Clean Air Act in 1971, Ocean Dumping Control Act in 1975 and Environmental Contaminants Act in 1976. The European Council adopted its First Action Program (on the environment) in 1973 and issued directives on quality objectives for drinking water in 1975 and water quality in 1976. Generally, Europe was slower off the mark than North America in this period, but pioneering efforts such as the establishment of the publication *The Ecologist* and its 1972 publication of *Blueprint for Survival* are particularly notable. In Australia, following some initiatives at the state level, in 1974 the Whitlam Labor government passed an Environmental Protection Act requiring advance consideration of the potential environmental consequences of proposed resource initiatives.

Whereas pollution was the focus of environmental politics on Earth Day, and in the ensuing period, sustainability issues also came to the fore, beginning in 1972 with the United Nations (UN) Stockholm Conference on the Human Environment. The conference emphasized a wide range of issues related to environment and development, as well as trade in endangered species, the protection of cultural and natural heritage and the prevention of marine pollution. Stockholm was crucial to the early establishment of environmental politics on the global stage. That year also saw the publication of the Club of Rome sponsored *Limits to Growth* (Meadows *et al.*, 1972), which gripped public attention, in part because it was based on the then-new notion of a computer simulation. This best-selling book polarized public opinion between those who saw a need for radical reform of industrial society to avoid resource exhaustion, pollution and human overpopulation and those who defended the *status quo*. The long-term projections from trends put forward in the study were widely attacked, and doubtless flawed in the details and in a number of the assumptions made. However, those flaws were soon overrun by immediate real world events. The Organization of Petroleum Exporting Countries (OPEC) slowed exports and radically escalated oil prices in 1973 (and again in 1979) and the world entered into the period of the energy crisis.

The energy crisis accelerated and altered the course of development of environmental politics throughout the world. Oil prices rose rapidly and the price of other energy sources and most commodities, including food, rose with them. Sustainability attained an intensity of public concern easily equal to that of pollution and the protection of nature. As well, for the first time, large numbers of people reflected on the long-term viability of industrial society. Prior to 1973 few had, since the arrival of the age of oil, imagined a world without readily available and inexpensive energy. This new concern was particularly challenging to any who doubted the viability, or desirability, of nuclear energy. The energy crisis stimulated sales of fuel-efficient automobiles, energy efficient appliances and home insulation, but it did not straightforwardly advance support for all aspects of environmental politics.

Through much of the 1970s, environmental concern with the dangers associated with oil tankers, pipelines and offshore oil exploration, opposition to nuclear power, and calls for the regulation of air pollution from coal seemed to some to stand in the way of energy independence and continuing prosperity. Environmentalists and OPEC alike seemed to challenge both the way of life and the economic wellbeing of wealthy individuals and societies. Environmentalists appeared to some to be saying no, or go slow, to every viable alternative to OPEC's political and economic power. However, two factors soon lessened that political tension: the notion of soft energy paths and the reality of nuclear accidents (as well as the high and rising costs of nuclear construction and the unknown cost of the long-term management of nuclear wastes).

A young physicist/environmentalist named Amory Lovins coined the concept of a *soft energy path* – an energy future that combined significant increases in

energy efficiency with the development of relatively environmentally benign, and renewable, energy sources (Lovins, 1977), solar power, wind power, geothermal energy, biomass, and small-scale hydroelectric generation. Detailed country-by-country studies were undertaken by Friends of the Earth and other environmental organizations to demonstrate the technical and economic feasibility of these combined options to both fossil fuels and nuclear power. Lovins himself was perhaps most concerned to undermine the view that nuclear electricity was a reasonable future alternative to fossil fuels; he emphasized the variety of risks associated with nuclear-generated electricity including the possible proliferation of nuclear weapons. Opposition to nuclear electricity was strong throughout the 1970s, especially in Germany, Sweden, Austria and the US, but notably not in France where Green politics had a much slower start than elsewhere in Europe and where there was also relatively wide nationalist-rooted support for a national nuclear weapons capability.

The Nuclear industry's problems accelerated following the accidents at Three Mile Island in Pennsylvania in 1979 and the far more serious mishap at Chernobyl in 1986. Nuclear energy also suffered from two other problems. Especially punishing to nuclear power were high interest rates put in place in the early 1980s to counter high inflation (resultant in turn in large part from high fossil fuel prices). Construction of nuclear power plants is slow and thus high interest rates resulted in an uneconomic level of costs before many nuclear plants on order or under construction were completed. New plant orders dried up almost everywhere and the newly forming German Green Party concentrated on issues related to nuclear power stations and nuclear weapons. Improvements in the efficiency of electricity use (in electric drive motors, lighting fixtures and appliances of all kinds) also slowed the demand for electricity despite a continuing increase in the number of uses. Economics, technology and environmental politics seemed to collectively conspire against nuclear energy for the remainder of the 20th century.

The late 1970s and early 1980s continued to witness shocking new environmental disasters, in the massive release of dioxin in Seveso, Italy in 1976 and the 1978 revelations regarding the massive burial in the 1940s and 1950s of toxic waste in the Love Canal, New York. However, this period also saw the first significant political backlash against environmental protection advocacy. Economic interests hurt by conservation measures and anti-pollution regulations became more aggressive in the defense of their interests. Neo-conservatism (usually called neo-liberalism in Europe) came to power in Great Britain, the US and elsewhere and was often determined to deregulate and to shrink government. Ronald Reagan opposed the protection of remaining remnant redwood forests and campaigned openly against air quality regulations in Ohio and other US coal mining regions. In office he appointed those who would roll back regulatory enforcement and who favored resource extraction interests. Environmental protection was no longer an uncontested motherhood issue; it had, in effect, come of age politically.

During the Reagan years pro-environmental opinion waned somewhat, but the commitment of those concerned grew as evidenced by an expansion of membership in US environmental organizations such as the Sierra Club. At the same time Australia saw a massive mobilization to oppose the Franklin Dam in Tasmania and green parties were created in many European countries. While environmental politics could not quite maintain the surge of the late 1960s and early 1970s, environmental activism continued throughout the period of low governmental responsiveness during the early 1980s. By the mid- to late 1980s a series of events, new concerns and new scientific findings brought a resurgence of wider public concern which led in turn to a new set of legislative and policy initiatives in the early 1990s.

The most dramatic and visible events of the period were: the catastrophic release of toxic chemicals in Bhopal, India (1984), the Chernobyl nuclear disaster in the Ukraine (1986) and the massive oil release from the Exxon Valdez off the pristine Alaska coast (1989). New or growing concerns included an array of problems associated with the disposal of municipal solid waste and the siting of landfills and incinerators, widespread concern with species loss especially in relation to the rapid cutting of tropical rainforests, and continuing concern with acid precipitation and with the disposal of toxic and radioactive wastes. New scientific findings also accelerated the renewed wave of environmental concern. Most notable of these were stratospheric ozone-layer depletion and climate warming resulting from the ongoing release of greenhouse gases. Of particular concern in this regard were the possibility of rising sea levels and possible threats to ocean phytoplankton and other basic life forms. By the late 1980s environmental matters were topping public opinion polls as issues of greatest concern. A new wave of environmental politics had begun and, as we will see below, both the environmental movement, and the response to it, was more sophisticated the second time around.

TWO WAVES IN CONTEMPORARY ENVIRONMENTAL POLITICS

In many nations interest in environmental issues, and in non-governmental organization (NGOs) and environmental movement activity, was higher in the periods 1968–1975 and 1985–1992 than at any other time. When one compares the two periods, both the most important issues and certain characteristics of the movement were quite different. The first wave saw the creation of some of the

leading environmental organizations and the environmental movement was at the time motivated primarily by new ndings, perspectives and concerns regarding pollution, the potential for future resource shortages (especially energy), nuclear power, and rapid human population growth. Politically, the rst wave was often highly confrontational and many early environmental advocates stood well apart in manner and viewpoint from mainstream society. The second wave raised some challenging new issues, but was generally more accepted by the wider population and in some cases, accommodation was sought by economic and political elite.

The early environmental movement in the rst wave sometimes displayed a signi cant apocalyptic and apolitical dimension. Modern industrial economies were seen by some to be on the brink of collapse, in danger of being suddenly overwhelmed by the pollution, diminishing resources (and the agrant waste of those resources), and out-of-control population growth. Only some of the information advanced in this period was carefully grounded in scienti c ndings, but the assertions struck a public nerve nonetheless and provoked governments, universities, and scienti c organizations to expand their environmental research capabilities, to reorganize research priorities and even to develop new interdisciplinary approaches.

As noted, it was in the early 1970s that many governments created environmental agencies and ministries. At the same time many universities, sometimes in surprising places, saw individual faculty and students direct research and teaching attention to a wide range of environmental issues. The locus for consideration of matters environmental could be as varied as departments of geography, political science, chemistry, biology, history, economics, or urban planning. Some interdisciplinary environmental studies programs grew very rapidly only to suffer enrollment declines as public interest waned somewhat (between the waves) in the late 1970s and early 1980s. Those declines con rmed for many critics within the universities that these issues and new approaches were but a ash in the pan and not, as claimed, a new and yet enduring approach to scholarship. Yet the declines were temporary and new sub- elds arose in everything from environmental ethics and economics to conservation biology and air pollution chemistry.

This altered intellectual and scienti c terrain in universities, and the wider learning within society as a whole, signi cantly altered the environmental politics of the second wave as compared to the rst. Throughout the 1970s and 1980s, graduates in environmental sciences and policy studies moved into ever increasingly professionalized environmental NGOs and policy consulting rms. Relevant government departments, previously predominantly populated by planners and engineers, added ecologists and other environmental scientists to their staffs and many engineering and law faculties, for example, added environmental components to their degree programs. Individuals with extensive NGO experience were hired by government ministries, or returned to graduate study and/or joined professional consulting rms or university faculties. By the time of the second wave (1985 1992), both the issues and the fundamental character of the environmental movement had changed. The movement itself had become highly professionalized (a change that was not without its critics from within and without the NGO community).

The second wave saw the re-emergence of preservationist and biodiversity concerns central to the conservation movement (dating back into the 19th century) and an increased interest in global issues. This contrasted to the rst wave where pollution and sustainability concerns dominated. The second wave also saw a more professionalized environmental movement provide ecological, economic, and legal analysis to government and to society at large. Environmentalism was much less on the fringe: environmental ideas were widely acknowledged and seen to be accepted by economic and political elite. A sophisticated array of technical options and policy tools was advanced rather than the vague notions of withdrawal from society and/or advocacy of crude end-of-the-pipe pollution controls so common in the rst wave. Environmental organizations discussed the importance of various taxation regimes, environmental user fees, governmental procurement patterns, environmental audits of industries, state-of-the-environment reporting, economy environment integration and sustainable development.

Society itself had become more environmentally sophisticated through changes over the years in public school curricula, nature and conservation television programming, and the sheer numbers of individual citizens directly affected by environmental concerns from smog to the siting of nuclear power plants, from the loss of a favorite woodland to the safety of drinking water or food. In the second wave, in Canada for example, environment economy round tables were established at all levels of government bringing together environmental activists, industry leaders, labor unions, and government of cials. Less well-known were the long-established links between labor unions and environmentalists in both the US and Canada dating back into the late 1960s through the efforts of the United Auto Workers Union and the Oil, Chemical, and Atomic Workers Union. In the 1990s the interactions, and sometimes, shared agreements were wider. Policy advances were sometimes made through these cooperative processes, especially at local level. More important perhaps, by the end of the second wave few participants saw the differences and distinctions between conservation and environmental organizations as important any longer.

More dramatic as a characteristic of the second wave was the introduction by industries and retailers in Europe, North America and elsewhere of a wide variety of so-called

green products. Many 1970s environmentalists would have been appalled at the very notion (and more than a few 1990s environmentalists had their doubts as well). Nonetheless, suddenly products with environmental claims were available from less polluting household cleaning products to energy-efficient lighting devices to road asphalt made from used tires to recycled paper and boxes. In Germany, Greenpeace entered directly into the business of producing energy-efficient refrigerators. Whatever the limits of green product practices, the introduction and wide acceptance of green products speaks to a considerable environmental capacity within both industry and society, and a very different style of environmental politics than in the more confrontational first wave. Also typical of this shifting approach were environmental audits and life cycle analysis, both tools used directly by industry to assess the environmental effects of their operations and products. Much changed in terms of the way products were made, recycled, and sold to environmentally sophisticated consumers, but as well the centrality of government regulation of industry had declined.

One way to understand these developments as a whole (from the late 1960s through the mid-1990s) is to see environmental capacity as having broadened and spread from environmental advocacy groups and universities into government agencies and the economy and society as a whole. Another way to understand the change is to see economic and political elite as becoming on the one hand more accommodating of some change, but also more sophisticated at delaying and at redirecting any calls for radical change. That is, the elite have become willing to shift products and processes where such changes are relatively inexpensive (and perhaps even profitable), but seek to deflect change from policies which would impact severely on profit margins or access to resources. Regardless of the overall accuracy of such interpretations, the second wave was different from the first and it too faded in the face of the economic difficulties of the 1990s. These recent challenges to environmental politics will be considered below, but first it is important to identify the principal institutional structures of environmental politics: environmental organizations and Green Parties.

ENVIRONMENTAL ORGANIZATIONS: GREEN PARTIES VERSUS NGOs

In many locations the environmental movement has gone from a political movement of individuals and NGOs to one which also includes adherents to a Green Party. Most European countries, as well as Australia and New Zealand (which can boast the first nationwide Green Party, the Values Party founded in 1972), now have Green Parties strong enough to have elected members of the legislature at the provincial (state) or national level. In the Australian state of Tasmania the Green vote has exceeded 30% of the total on occasion. In Germany and elsewhere, Greens have held the balance of power at the state level, entered the government on the national level and governed outright in the city of Frankfurt. But in few other countries have Green Parties been so successful; in North America the Greens are all but absent despite the long-standing existence of a strong environmental movement. Why have Green Parties not been more widely successful, one might ask? Why in North America and elsewhere has voting for Green Parties never approached the level of support for environmentalism that exists in terms of public opinion polls or in terms of membership in environmental organizations?

One reason for this limited success within electoral processes is that environmentalists have no consistent position on many political issues of great importance to many people. What is the Green position on gender or regional issues? What is the position of environmentalists on inflation, unemployment, drunk driving, abortion or appropriate levels of taxation? Clearly there is none in most of these cases. It could be said that environmentalism is a truncated ideology: socialists, conservatives and liberals have consistent views on most questions, environmentalists do not. Some environmentalists are quite conservative on economic and social issues, others are quite progressive. It is hard to hold a political party together when its members simply do not agree, or even in some cases care about, some of the leading issues of the day. Many voters wonder how could a Green Party govern; in government a party must make decisions one way or the other.

In addition, the road is difficult for any new political parties; the advantages of incumbency or recent incumbency are considerable. As well, Green Parties have succeeded almost exclusively within multiparty political systems, usually where there is some form of proportional representation (party voting, i.e., a percentage of the votes producing a set number of seats on a national or regional basis). In such a context Green Parties can more easily emphasize some issues and ignore others (or trade support on issues of lesser importance for gains on the environment). Only the Minister of the Environment in the coalition government of France formed in the late 1990s has been a Green. Green Parties work better in some systems than others. Where there are single-member constituencies, as in Canada, the US and Great Britain, Green Parties have not gained much ground, even where many people like *green* ideas. Usually, existing parties take up green ideas, whether there are successful Green Parties or not, so long as there is support for green ideas. These ideas have played an important political role for more than 30 years within widely varied institutional arrangements: clearly,

then, environmental politics does not always require stand-alone Green Political Parties.

In broad terms, political movements seek to influence society and/or government; in contrast, political parties seek to become the government themselves or to be directly a part of it. Both Green Parties and NGOs influence the way society and government function, altering some of the collective choices that are made and participating in a societal process called *agenda setting*. Agenda setting determines which issues are priority ones, which concerns get the most attention and resources, which are dealt with first and most frequently by decision makers. All interest groups including environmental organizations, the media, business, political parties and many other organizations are involved in the process of agenda setting.

A movement is not so formally organized as a political party or an interest group (NGO). Individuals in some roles and activities may be part of a movement without necessarily being members of NGOs or even seeing themselves as members of that movement. For example, a group of neighbors may oppose the siting of a landfill or some other facility with environmental dimensions, though some are opposed to many objectives of the major environmental organizations. Nonetheless, their opposition to the particular project in question should be understood as contributing to the wider environmental movement. That movement includes, as well the larger, established environmental organizations such as Greenpeace and the Sierra Club, thousands of smaller organizations, as well as many unaffiliated individuals. A formal organization has members, regular meetings, likely a constitution of some sort, a budget, and so forth. A movement does not, it is much looser and more variable than that, but it certainly is identifiable and visible.

The heart of the environmental movement and environmental politics has, in most jurisdictions, been environmental organizations; environmental public interest groups or NGOs. Some are single issue oriented and fleeting others are very large, even global in their scope of operations and some have continuous organizational histories of a century or more. It is not possible to present here anything like a full nation-by-nation list. There are many thousands of environmental organizations in North America alone. In the nations of the European Union more than 10 million individuals are members or active supporters of environmental organizations. One of the earliest European groups still in existence is England's Commons, Open Spaces and Footpaths Preservation Society founded in 1865. The organizations with the greatest global reach include Greenpeace (founded by Canadians, but with at one point approaching million members in, for example, the Netherlands), the Worldwide Fund for Nature (WWF) with activities on every continent, the International Union for the Conservation of Nature (IUCN) based in Switzerland, and Friends of the Earth International (FOEI).

Perhaps equally important are large conservation and environmental organizations based primarily within one nation: the Sierra Club, the Natural Resources Defense Council, the Wilderness Society, the Environmental Defense Fund, the National Audubon Society, the Nature Conservancy, and the National Wildlife Federation in the US. In Canada, in addition to Greenpeace and important parallel organizations of US or other origin, are Pollution Probe, the Canadian Nature Federation, the Toronto Environmental Alliance and the Western Canadian Wilderness Committee. Other important environmental organizations in the richer nations include: in Germany, the Bund fur Umvelt und Naturschutz Deutschland (BUND); in Denmark, the Danish Society for the Conservation of Nature (founded in 1911); in Australia, the Australian Conservation Foundation and the Tasmanian Wilderness Society; and in Great Britain, (in addition to large branches of Friends of the Earth and Greenpeace) a number of large, long-established, conservation organizations such as the National Trust and the Royal Society for the Protection of Birds.

Finally here it is important to identify several radical challenges to established environmental organizations. These challenges are ongoing and arise from many quarters, they continually prod larger organizations that inevitably become just somewhat more cautious as memberships and the size of professional staffs increase (and the need to raise funds and maintain membership size grows). Friends of the Earth was originally a split-off from the Sierra Club. The environmental justice movement, which emphasizes racial bias in the siting of environmentally undesirable facilities, has been very important in the US since the late 1980s. Numerous environmental justice grassroots organizations have been highly critical of the larger, more established environmental organizations for under representing visible minorities and their concerns. Also highly significant since the 1970s has been *ecofeminism*, a movement within a movement which focuses on environmental issues which disproportionately affect women, especially but by no means exclusively women within developing nations (Warren, 1995). As well, the ideas of *deep ecology* helped to encourage an animal rights movement, and such radical wilderness protection organizations as Earth First!, and in so doing restored something of a confrontational style within second wave environmental politics.

Also sometimes associated with a more radical strain of environmental politics is the simultaneous consideration of environmental concerns and the problems of poorer nations; here the movement has been pushed by such international groupings as the Rainforest Alliance.

THE COMBINED CHALLENGE OF ENVIRONMENT AND DEVELOPMENT

Combined environment and development concerns came to prominent attention in the 1960s with a focus on rapid human population growth, especially within poorer nations. At the time this attention often had a neo-Malthusian tone and frequently contained warnings regarding imminent and widespread famine. Radical and direct interventions to slow birth rates were seen by some as appropriate and necessary initiatives. Paul Ehrlich (1975), in his widely read *Population Bomb* (first published 1968), called for taxes (in the US) on cribs, diapers, and toys, followed by an aggressive anti-natalist foreign policy abroad. External policies were to include tying food aid to the adoption of strong population control initiatives by aid recipient nations. Ehrlich supported proposals in India for compulsory sterilization of males with three or more children. Garrett Hardin advocated state-issued licenses to reproduce, licenses which could be bought or sold on an open market. It is difficult to imagine assertions of this sort being put forward now and, needless to say, there were strong responses at the time as well.

Some decades later those anticipated-to-be-imminent famines have thus far been avoided through improved food production and non-coerced population restraint (especially in Europe and North America, but in other nations as well). Nonetheless, human population may well already be well past the environmental optimum, but public attention has turned more to development than to population considerations. Barry Commoner (1972), responding to Ehrlich, argued that environmental damage was more a function of technology choices than of human population levels (or even of affluence). He further argued that human population restraint would follow from advances in economic wellbeing. Other analysts have noted strong correlations between population restraint and social policy advances such as equitable education and work opportunities for women, rising economic expectations, wider availability of old age pensions, and access to health care for infants. In other words, family size declines with social and economic advance; people will have fewer children when there is hope that those they do have will survive and that they themselves will not, in old age, be dependent on their children.

Thus, while 1990s environmental advocates would still favor the universal availability of affordable birth control opportunities, the discourse on environment and development has changed very much since the 1960s. Population remains a concern, but the emphasis has turned to development, economic policy and equitable global access to resources. Following Commoner's response to Ehrlich were a series of other studies, which further altered our sense of these issues. In the 1970s one important shift in tone was visible in the differences between *Limits to Growth* (1972) and a follow-up Club of Rome study *Mankind at the Turning Point* (1974). Where the former emphasized the need to restrain human population, economic growth and resource use to protect the future, the latter emphasized the need to see growth as necessary within poor nations and resource and economic restraint as more appropriate in rich nations. Hungry people with few material possessions do not readily embrace blanket calls for restraint; global environmental politics required greater sophistication than that.

Malthusian approaches to environmental protection waned further still with the insights of Amory Lovins and advocates of appropriate technology, and especially with the work of the World Commission on Environment and Development (the Brundtland Commission). The 1970s and 1980s saw many insights into more environmentally benign housing, agricultural and energy technologies which also promised economic advance, especially within low-development contexts. Lovins' promotion of solar, wind and biomass energy production was an important part of this changing focus. The Brundtland Report, *Our Common Future* (World Commission on Environment and Development, 1987), however, helped to further shift the focus from a more or less exclusive focus on the environmental costs of economic advance to a more balanced approach which included as well the environmental costs of failures to advance economically.

Our Common Future advocated a balanced consideration of environmental protection and economic development (sustainable development) and asserted that both were necessary, more mutually supportive than mutually exclusive. In the language of the report:

> In most countries, environmental policies are directed at the symptoms of harmful growth; these policies have brought progress and rewards and must be continued and strengthened. But that will not be enough. What is required is a new approach in which all nations aim at a type of development that integrates production with resource conservation and enhancement, and that links both to the provision to all of an adequate livelihood base and equitable access to resources (39–40).

The goal of the Brundtland Commission was to help to create a global environmental politics which was large enough to contain development advocates (mostly from poor nations), environmental protection advocates (mostly from rich nations) and enlightened representatives of government and business.

It is not universally agreed that this attempt was successful, or that sustainable development is not an oxymoron. Nonetheless the discourse around the politics of environment and development has shifted a great deal. It is now widely known that some of the gravest environmental damage is linked to economic desperation. Few environmental advocates would deny any longer that economic growth restraint is equally appropriate everywhere or that there are not some opportunities for simultaneous economic growth and environmental protection, even in rich nations. Few

would any longer ignore the importance of economic and social advance to the achievement of human population restraint or suggest that birth rate coercion was either moral or effective. The larger question remains however, as to what economic activities need to be restrained where and how and whether or not it is possible to simultaneously give a free hand to the forces of global economic growth and to achieve either environmental or social advance. Concern regarding environment and development has now largely merged into a larger debate surrounding economic globalization.

GLOBALIZATION AND SUSTAINABILITY

In the wake of the Cold War, global trade expansion and global trade regimes have rushed to fill the vacuum and to make the entire world secure for mobile capital investment. In this new context, global investors can bring enormous pressures to bear on governments by even implicitly threatening to withhold or transfer investment. This new reality bears in turn on an important question in environmental politics; the centralization or decentralization of environmental protection. In general, the smaller the unit of government and the less economically diverse the political jurisdiction, and the higher the local unemployment rate, the greater the relative power of mobile corporations. Thus, in recent years environmental protection rollbacks have been widespread and at the same time much environmental politics and decision-making has moved to the international, treaty-based, level (*see* Environmental Movement – the Rise of Non-government Organizations (NGOs), Volume 5).

Before discussing environmental initiatives at the international level a distinction needs to be drawn between the fact of increasing global economic integration and the negotiated trade regimes, such as the North American Free Trade Agreement (NAFTA) or the General Agreement on Tariffs and Trade (GATT). When a firm can pay skilled manufacturing workers a dollar or two a day in Indonesia, China or Mexico, it can easily afford to pay quite high tariffs to sell into the US, Japan, or Europe. This is especially true if, at the same time, firms avoid legal protection for trade unions, workplace health and safety regimes, and enforced environmental regulations. Trade treaties do not cause these problems. They may help to speed the trading process, but trade treaties also provide a potential opportunity to harmonize (average) upwards, rather than downwards, on labor and environmental standards. As with the NAFTA Side Agreement on Labor and the Environment, there is no technical (as distinct from political) reason why trade treaties could not contain progressive social and environmental provisions. There is also nothing but a lack of political will (obviously no small matter) which prevents common (and enforced) environmental regulations, as is virtually now the norm within the European Union.

An impressive array of multilateral and bilateral environmental treaties have recently been signed; indeed, since the late 1970s environmental politics seems to be as often conducted at the diplomatic level as at the national level. Major treaties include: the Convention on International Trade in Endangered Species of Wild Flora and Fauna (1973, 1979, 1983); the Montreal Protocol on Substances that Deplete the Ozone layer (1987); the Basal Convention on the Control of Transboundary Movements of Hazardous Wastes and Their Disposal (1989); the UN Biodiversity Convention (1992); and the UN Framework Convention on Climate Change (1992, followed by the additional agreement in Kyoto, Japan in 1998). Quite clearly it is the case that many global environmental problems can only be resolved with the active participation of many, if not most, nations. The larger question, however, is how effective is signing any number of international environmental agreements when there is no effective international enforcement?

Moreover, the new realities of global corporate mobility and widespread employment vulnerability have enormous implications for environmental politics. Even in the richest nations, as well as in local government circles and among employees everywhere, there is now a thoroughgoing political timidity. Every government now fears losing employment, investment, and tax revenues. Almost any social behavior (even including child labor) is somewhere justified in terms of global competitiveness and employees and firms everywhere must compete with that new standard. No level of profit, or level of executive or product endorsement compensation, is seen any longer as unacceptable. Further, what would be seen in Europe or North America as environmentally intolerable, even outrageous, behavior, is almost expected in Nigeria, Indonesia, or Guyana, even when good corporate citizenship prevails on the environmental front in the Netherlands, Canada and/or the US.

Interestingly, some of today's billionaires, not a reborn political radicalism, are often the very people raising questions about the need to find ways to balance corporate power in the age of globalization. George Soros (1997, 47) argued:

> Laissez-faire capitalism holds that the common good is best served by the uninhibited pursuit of self interest. Unless it is tempered by the recognition of a common interest that ought to take precedence over particular interests, our present system is liable to break down.

James Goldsmith (1994, 17), the perhaps eccentric British/French billionaire is more direct:

> ••• 47 Vietnamese or 47 Filipinos can be employed for the cost of one person in a developed country, such as France. Until recently, ••• 4 billion people were separated from our economy by their political systems, primarily communist or socialist, and because of a lack of technology and of capital. Today all that has changed. Their political systems have been transformed,

technology can be transferred instantaneously anywhere in the world on a microchip, and capital is free to be invested anywhere the anticipated yields are highest.

Many in the world of environmental advocacy would conclude from such views that global economics requires global environmental and social standards; that is, that an economically integrated globe needs more political integration, not less. They conclude that if Southeast Asian forest workers earned even half the wages of those in the Pacific Northwest, the rate of destruction of tropical rain forests might well slow considerably, and simultaneously development would advance. Similarly, if Caribbean bauxite miners were paid half what a miner would earn in Germany, there would be less aluminum that was not recycled. There is now broad agreement that the economic, the social and the environmental are tied together in numerous ways; and that environmental advocates must work towards a consistent position regarding all three. Employment insecurity, it is agreed, undermines environmental concern, in both poor countries and rich. Environmental politics in the 1990s and beyond may well be moving toward greater integration within the broader political, and toward including in its expanded purview many aspects of social and economic policy.

Expanding the purview of environmental politics at the intellectual level revolves around the rising importance of sustainability. This shift within environmental thought is especially pronounced in European environmental politics which are increasingly in the 1990s leading the way worldwide (and where wilderness issues have never had the salience that they have had in North America). One new approach to sustainability analysis is *ecological footprint analysis*, which concludes that it would take three Earths of resources to bring the whole of the human population to present North American living standards (Wackernagel and Rees, 1996) (*see* Ecological Capacity, Volume 3). This analysis, however, assumes that the gains in the poor nations will be based on the energy and materials consumption patterns of today s richest nations. That need not be true. One can envision a gross domestic product (GDP)/energy and materials ratio three or four times what is achieved at present, but this might still leave us several planets short should human population indeed double again as now seems all but inevitable.

In the end, if one prefers a world that is both more equitable and environmentally sustainable, it is hard not to be open to the idea that the rich nations are rich enough, at least in terms of energy and material consumption. It would seem, though, that the poor everywhere could hardly use less of anything than they presently do without nding ways to gain in quality of life while holding their ground in terms of energy and materials use. Environmental politics must stress such things as the introduction of high quality, low cost public transportation, especially as replacements for broken down, fuel inef cient automobiles. Also hopeful are demand-side management (DSM) initiatives by energy utilities directed at the needs of low-income customers. Nonetheless, it may well remain the case that the rich in the rich countries cannot be easily and sustainably supported even at present consumption levels and patterns if the more numerous poor everywhere are to advance signi cantly in the same terms. Politically, this view would seem all but impossible in the present context, but one might speculate that if there is to be a successful third wave in environmental politics it will come at the global scale and focus on sustainability and equity issues.

Thus, two of the best prospects for future environmental politics in a globalized world are: (1) accelerating the adoption of technologies such as those outlined by von Weisacker *et al.* (1998) in *Factor Four* and (2) developing a global politics which would promote the distribution of fair shares of ecological space as outlined in Carley and Spapens (1998). The rst of these works contends that we must, through ef ciency gains and the de-materialization of GDP double global wealth while simultaneously halving resource and energy use. The second work advocates calculating resource use at the national level and comparing that use level to a national fair share of sustainable resources calculated on a global or regional basis. The educational power of such calculations could be considerable. The political challenge then becomes to nd ways to accommodate development based on that fair share without a loss in quality of life for those currently well off.

Environmental politics would seem to have come a very long way in four decades: from local pollution issues to global multidimensional concerns, from seeing environmental issues as largely separate from distributional issues to seeing the two as bound up together. A long road has been traveled and progress has been made, but the challenges have not gotten smaller.

REFERENCES

Carley, M and Spapens, P (1998) *Sharing the World*, Earthscan, London.

Carson, R (1962) *Silent Spring*, Houghton Mif in, Boston, MA.

Commoner, B (1972) *The Closing Circle*, Bantam, New York.

Ehrlich, P (1975) *The Population Bomb*, Rivercity Press, Rivercity, MA.

Goldsmith, J (1994) *The Trap*, Macmillan, London.

Hays, S P (1987) *Beauty, Health and Permanence: Environmental Politics in the US, 1955 1985*, Cambridge University Press, Cambridge.

Jevons, W S (1965) *The Coal Question*, Augustus M Kelley, New York.

Lovins, A B (1977) *Soft Energy Paths*, Ballinger, Cambridge.

Marsh, G P (1965) *Man and Nature: Physical Geography as Modified by Human Action*, Harvard University Press, Cambridge.

Meadows, D H, Meadows, D L, Randers, J, and Behrens, III, W W (1972) *The Limits to Growth*, Universe, New York.
Ordway, S H (1953) *Resources and the American Dream*, Ronald Press, New York.
Osborne, F (1948) *Our Plundered Planet*, Little Brown, Boston, MA.
Osborne, F (1953) *The Limits of the Earth*, Little Brown, Boston, MA.
Paehlke, R C (1997) Environmental Values and Public Policy, in *Environmental Policy in the 1990s*, eds N J Vig and M E Kraft, Congressional Quarterly Press, Washington, DC.
Paehlke, R C (1989) *Environmentalism and the Future of Progressive Politics*, Yale University Press, New Haven, CT.
Soros, G (1997) The Capitalist Threat, *Atlantic Monthly*, **279**, 45–8.
von Weisacker, E, Lovins, A B, and Lovins, L H (1998) *Factor Four: Doubling Wealth, Halving Resource Use*, Earthscan, London.
Wackernagel, M and Rees, W E (1996) *Our Ecological Footprint*, New Society, Philadelphia, PA.
Warren, K J (1995) Ecofeminism, in *Conservation and Environmentalism: An Encyclopedia*, ed R C Paehlke, Garland, New York.
World Commission on Environment and Development (1987) *Our Common Future*, Oxford University Press, Oxford.

FURTHER READING

Kallett, A and Schlink, F J (1976) *One Hundred Million Guinea Pigs*, Ayer, Salem, NH.
Lamb, R D (1976) *American Chamber of Horrors*, Ayer, Salem, NH.
Mesarovic, M and Pestel, E (1974) *Mankind at the Turning Point*, Dutton, New York.

Global Environmental Change and Environmental History

IAN G SIMMONS

University of Durham, Durham, UK

To try and extract some of the ideas which intertwine with the environmental changes wrought by humans in the last 10 000 years, an agreed empirical historiography of change is needed. The article suggests that the major world-wide periods of human history for the present purpose can be determined by the type of economy pursued. In effect, these are periods which are characterised by access to energy sources and hence to the technologies available: these properties allow a society to change its surroundings, providing it has occupancy of the land and water for a suf cient time. Materialist historiography has always taken very seriously the notion that the ways in which people made a living were central to their other values and that technology drove history. The story of what has happened, where, and when, is well-documented at global, regional and local levels but less well developed is: (a) the relationships of such metamorphoses to changes in human thinking; and (b) whether all these alterations present history simply as 'one damn thing after another' or whether there are emerging narratives of process in which we can see patterns emerging with a 10 000 year relevance.

THE EMPIRICAL NARRATIVE

The first item in Johann Sebastian Bach's will is his share in a Thuringian coalmine. That fact can be seen as a nexus between a local environmental change, a global shift in economy, a family in its household and his searing representations in music of order and faith which comprise one of the high achievements of Western culture. But is there any connection between these statements other than that of being symbols on a piece of paper? How does the economy of a family, a region or a country or the world link to their ideas about the world or do all these aspects float free of each other, with only coincidence in time in common?

To try and extract some of the ideas, which intertwine with the environmental changes wrought by humans in the last 10 000 years, an agreed empirical narrative of change is needed. Much of this is provided in other sections of this Encyclopaedia and only a summary is given here. Table 1 suggests that the major world-wide periods of human history for the present purpose are determined by the type of economy pursued. In effect, these are periods which are characterised by access to energy sources, and hence to the technologies available: these properties allow a society to change its surroundings, providing it has occupancy of the land and water for a sufficient time. In this article the emphasis will be on the relations of those observed alterations: (a) to the metamorphoses of human thought that accompanied them; and (b) to emerging narratives of process which we can construct with hindsight. A coda ponders whether there are lessons for today and tomorrow in this kind of hybrid-knowledge fabrication.

Human history is reduced to economy in Table 1. This provides a starting point in the sense that materialist historiography has always taken very seriously the notion that the ways in which people made a living were central to their other values and that technology drove history. In the other direction, it is obvious that the environmental impact of a hunter-gatherer is likely to be less than that of a bulldozer, except that the one economy has in fact taken up about 90% of human evolutionary time and the other is (even loosely interpreted as any industrial economy based on hydrocarbons) only 200 years old; the bulldozer is largely a post-1940 development. Thus although the access to energy and technology of solar-powered agriculturists was restricted, their tenure of the earth has been long, and they have penetrated into some environments whose resilience in the face of their impacts was limited. The use of tabular format rather hides the fact, of course, that not all hunter-gathers have disappeared, and that there are some farmers who have little access to industrially produced materials. In both cases, a revival of the older ways is being promoted in some regions of the world. Nevertheless, the changes wrought by the industrial way of life and its

Table 1 Global environmental history: a periodization[a]

	Dates: approximately	Energy sources; Energy consumption	Technology	Impacts
Hunter-gatherer	10 000–8000 BCE	Solar energy captured recently as biomass 5000 kcal cap^{-1} day^{-1}	Stone, bone, basket, skins, wood	Localized but significant in marginal species and habitats
Agriculturalist	8000 BCE–CE 1750	Solar energy captured recently as biomass, wind and water 26 000 kcal cap^{-1} day^{-1}	Addition of metals plus sophisticated construction methods, specialised tools; domesticated biota	Global impact of agriculture on soils, water and land-forms: e.g., the terrace transfer of species on all scales
Industrialist	CE 1750–1950	As above + stored solar energy as fossil fuels 77 000 kcal cap^{-1} day^{-1}	Intensification of material use in cities, as chemicals, as effluents, as machines	Intensification of impact and spatial extension
Post-industrialist	CE 1950–	As above + nuclear energy and alternative sources of solar-based energy 230 000 kcal cap^{-1} day^{-1}	Electronics; genetic manipulation	New chemicals and new genetics create uncertainties on larger scale

[a] All rows contain the influences in the rows above them unless explicitly excluded; cap, *per capita*.

post-industrial successor (which has the ability to generate large amounts of capital and spin it off into almost any part of the globe) cannot be gainsaid. An electricity-powered economy with a high proportion of services is sometimes labelled *post-industrial* but although it may call in resources from a low-energy, low-information periphery, it exerts many strong influences upon the world's ecology. It is fundamentally responsible for much of the appropriation of Net Primary Productivity by human societies, which now amounts to about 40% of the global terrestrial total. If the pre-agricultural terrestrial biomass was about 842×10^{15} g of carbon, the present figure is, according to Matthews (1984), 737×10^{15} g of carbon, a reduction of 12.4%. In spatial terms, this means that although 51.2% of the terrestrial surfaces can still be described as undisturbed', if rock, ice and barren land is taken out of the total, then that category totals only 27.0%. As calculated in economic terms, ecosystem services are worth about $33 trillion per year, which compares with the gross domestic product (GDP) of all the world's nations at $18 trillion. Employing energy use as a surrogate for environmental impact, the current level of impress is 13.1 terawatts (TW); a human population of 12 billion people with a more evenly distributed access to energy might produce a footprint equivalent to 90 TW.

The story of what has happened, where, and when, is well documented at global, regional and local levels. What is less well developed is: (a) the relationships of such metamorphoses to changes in human thinking; (b) whether there are emerging narratives of process in which we can see patterns emerging with 10 000 year relevance. So this essay will attempt to discuss both an *ideational narrative* in which a history of ideas is interrogated for its relation to what actually happened in the world and a *narrative of process* in which longer-term processes are plucked out of the multifariousness of history for further examination.

THE NARRATIVE OF IDEAS

Two avenues will be followed here. The first is the discussion of ideas, which seem to be connected to economies quite closely. If they are connected to economies: then they are bound up in the webs which link ecology, human use on the ground and what humans perceive in their minds. The second is the opposite: those actions and attitudes, which seem to float free of the economies and ecologies of their time.

Any reasonably accurate long-term perspective on human–environment relations can only be made by the historically well-informed. Those who relied on especially

partial stories are likely to have constructed different views of the unfolding of society from those who had extensive assemblages to consult. An early symbol of the desire for inclusive compilations of knowledge is the model globe. The significance of early examples such as that made by Martin Behaim in 1492 and annotated with the nature of commodities and business opportunities throughout the known world can scarcely be underestimated. Behaim's globe, for example, was used to sell Magellan's projected voyage to sponsors (Jardine, 1996). The gatherings of the 18th and 19th centuries take on a special significance, starting perhaps with the *Encyclopédie* of 1751–56. A fundamental example is the ambitious aim of Alexander von Humbolt (1769–1859) to depict in a single work the entire material universe which led to his multi-volume *Cosmos* of 1845–1862. The detailed results of Charles Darwin's global voyage, and his later observations and experiments also typify the desire to accumulate comprehensive bodies of knowledge, though Darwin (1809–1882) lacked the cosmic aspirations of Humboldt.

No study of material-ideational links can ignore the centrality of Karl Marx (1818–1883). The environmental aspects of Marx are implicit rather than explicit and urban–industrial rather than global–historical. Nevertheless, he viewed the integrated development of a political economy as a humanising, socialising and indeed internalising of natural substances and forces in the environment. According to Marx, therefore,

> The mode of production in material life determines the general character of the social, political and spiritual processes of life. It is not the consciousness of men that determines their existence but, on the contrary, their social existence which determines their consciousness.

So one environmentally-oriented interpreter, Parsons (1977), is able to say that for Marx all human values, whether economically-linked like food, or higher values like music, have their ultimate origin in non-human nature. Yet Marx does not attempt to define the equivalent linkages for pre-industrial economies in the detail that he applies to the separating and alienating effects of industrialisation in the context of the social relations of capitalism. Followers and explainers of Marx have tried to work out a deeper historical sequence. Ted Benton, for instance, has asserted that the theory of nature–society relations which is developed in *Capital* is incapable of conceptualising today's population–resource–environment interactions. Benton's reconstruction (1996) of the intellectual resources of Marx and Engels brings forward the notion that development is not a unilinear quantitative expansion of production but rather a range of qualitatively different ways of realising human social possibilities. Marxian analysis of pre-industrial societies may need to build upon *Capital* rather than simply apply it.

Another attempt to interpret material history is espoused by Karl Kautsky (1854–1938). He attempted to show parallels between the processes of nature and those of human history. The non-human environment was in fact human-created or at least much modified in many aspects and therefore its evolution was social. Nevertheless it had material connections since for him every social innovation originated in the last analysis from a new technology. Thus technology is a predominant mediating influence between the natural environment and human societies (Kautsky, 1988).

A more recent re-statement of some of these notions has come from Max Oelschlager (1991), from whom there are assertions such as:

> once the agricultural turn was made, philosophy and theology sprung forth with a vengeance ••• the crucial interaction between existential and conceptual materials – lying at the heart of cultural process and the ecological transition – increased in both frequency and pervasiveness ••• ideological reconstruction was inevitable. And ••• The idea of nature as a machine was deeply rooted in the experience of the Industrial Revolution and the pervasive influence of machines on life.

The first of these statements traces its connections in part at least to the surpluses created by irrigated agriculture: philosophy may bake no bread, but an assured supply of bread enables philosophers to live. The second reinforces the Kautsky directive that technology is a key mediator between humans and all else. The development of fossil fuels has clearly allowed profound transformations of social structures to take place. New work groupings and new social groupings with specialised interests are examples. Goldblatt (1996) argues that industrialism does not cause capitalism (nor its environmentally degrading effects) but it does unleash it. It would be generally agreed that access to energy sources and to technology are principal factors but not *the* principal factor; they must be considered in relation to all other aspects of the social order.

In Smil's (1994) collection of work on business cycles in the West, there is a confluence of material change with economic life. This can be interpreted to show a coincidence between new energy sources and innovative prime movers on the one hand, and accelerated investment on the other. Innovation in energy technology and access falls in the middle of a business down-cycle and its adoption then coincides with the rise of the new upswing. The sequence suggested comprises: (a) the upswing of 1787–1814 is related to the development of coal and the stationary steam engine; (b) that of 1843–1869 is connected to moving steam engines in the forms of the steamship and the railroad, together with improvements in iron smelting; (c) that of 1898–1924 with the introduction of commercial electricity in the form of electric motors in factories; (d) the period around 1937 which coincides with the gas turbine, fluorescent light and the discoveries in nuclear physics;

and (e) the phase that was ended in 1973 by oil producing countries (OPEC), which had cheap oil, and the extensive and unquestioned use of the motor vehicle and nuclear fission power.

The ideational effects of these surges are difficult to trace directly, but there can be little doubt that each of them helped to strengthen the Promethean aspects of the western world-view. Grubler (1994) suggests parallel developments in western agriculture: innovation up to the mid-19th century, mercantilism from 1850–1935 and industrialization from 1935 to the present. Each of these was a cluster of biological, mechanical and social/organizational change, with the social sometimes preceding the new crops and new techniques. The idea of social structures founded upon the dissipation of energy is a meta-narrative that takes off from these abstractions and turns them into a general model.

The classic work on the long-term relations of the natural environment and all facets of human societies is that of the Annales school of historians, and in particular the central figure of Fernand Braudel (1902–1985). He wished to understand patterns of nature as well as history, so that a radical inter-discipliniarity might be accomplished, using the insights of other social and natural sciences. If a long enough period of time was inspected, then not only the history of events (*histoire evenmentielle*) might be better linked to the world in which it was situated, but the *longue duree* might the better include environmental change, which was seen as proceeding along a much slower time path than changes located solely in human actions. Braudel (e.g., 1972) saw environmental transmutation as the slowest vector in the whole society-nature complex, a point on which later appreciations of the rapidity of e.g., a climatic change or the onset and progress of soil erosion might well take a different view. In the end, though, the Annales view is of a complex rhythmical web in which environment has to be given a place but cannot crassly determine the nature of a society. If environment is often critical (in terms of the juxtaposition of mountain and sea in the Mediterranean basin, for example), it is rarely as direct as in the example of snow being brought down in summer from the Alps to Italian cities being the origin of the making of ice-cream. The tone is more of environmental than economic determinism.

The French attempts to make long-term connections between economy and society are also exemplified in the work of Georges Dumezil (1898–1986). He had the idea that the diverse cultures of the Indo-Europeans (from Iran to Scandinavia) could be constituted in terms of the three functions of sovereignty, war and production. This last was, according to one commentator (Littleton, 1982), the least important (comprising the herder-cultivator stratum) and so perhaps accounts for its lack of detailed treatment compared with the other two, in his major outputs. In view of the endpoint of Dumezil's analysis of the extent to which a common linguistic ancestry is necessarily accompanied by a common social and cultural heritage, this seems like an opportunity missed. The same is true of Toynbee's great compilations (1976) of historical change which are heavily weighted in their treatment towards early civilisations. In these the causation runs from environment to religion and economy and the values do not stem in an obvious fashion from economy; periods of breakdown relate to political and social events, not rooted in an economy's structure. The advances of Islam into Europe in the 16th century or the Hellenic demise of 431 BC are little connected in any direct fashion with the imbalances between population, resources and environment in those regions at those times.

One of the most complete attempts at linking economy and the non-material has come with the work of Thompson (1989). He articulated a number of global phases of an economic type, beginning with early agriculture along the margins of the mountains of south-west Asia and ending with a Paci-c Shift which corresponds largely with a post-industrial society. For each of these, he posits a series of intellectual and social characteristics, among which science and religion are numbered. The sequence does not include a speci-c hunter-gatherer era but it is not dif-cult to add one. Nor is it dif-cult to add some other environmental linkages, such as dominant environmental pathologies and rates of population growth. One implication is that these are world-wide categories which transcend the varieties of culture found in, say, Europe, Africa and Asia. The descriptions given by Thompson, when translated to tabular form, make it necessary to remember that the transition from one category to another is regionally asynchronous. However, the work is certainly suggestive: the tying-in of a speci-c type of religion to each economic phase raises interesting questions and taken together, each row might be said perhaps to constitute a worldview, if the cells are verbally connected and scanned for their emergent attributes.

The notion that each stage of social change corresponds to its own types of law, government, religion (and other characteristics) was a major achievement of Giambattista da Vico (1668–1744) but who did not develop very far the notion of economy as a propellant of such change. His cycles of civilisation, in which divine, heroic and humanist phases went round more than once, did not attach themselves to, e.g., hunter-gatherers, irrigation agriculture and Mediterranean pastoralism. He did however suggest that Latin was a language which had evolved in a forest environment, making a case for the evolution of words like lex, ilex, aquilex, legumen and legere from a sylvan environment. But as Berlin (1992) points out, Vico's main contribution to thought is his distinction between the natural world as knowable but not intelligible, and the human world as knowable and intelligible. Because we made it; the suggestion, contra Descartes, that a reconstructive imagination

(fantasia) was a third kind of knowledge needed by students of history, seems to have a great deal to commend it.

The summation of this type of material seems clear: that in a few cases there seems to be a direct connection between the economy and some at least of the non-material values of a society. The attachment seems to be closer in solar-powered, pre-industrial economies, despite the assertion of Marx that all values arose from economics and class structure. In a sense this parallels the arguments for environmental determinism, which has always held more appeal in pre-industrial contexts, though it is clearly not absent in a post-industrial world.

DISSOCIATED TRENDS

If the links between economy and environmental values cannot be traced easily for the apparently obvious examples by the well-known analysts discussed above, then it is not surprising that a number of less obvious connections can be aired. One result of ranging across 10 000 years of change is the provision of examples to suit almost any taste in narrative. Rather like the Bible, something can be found to bolster any point of view. It is scarcely surprising then to be able to find examples of environmental thought which bear no perceptible relationship to the material circumstances of their origins and development. Culture, say Harvey and Reed (1994), can take on the features of a chaotic cascade of symbols and ideas that have a life of their own. Hence humans appear to make themselves.

To take one example: if we accept the magnitude of the Classical Greek achievements in both chronicling their own past as well as in thought about humanity and nature, then it is difficult to tie much of this to the particular agricultural economies of the time. Those economies had been in largely the same condition for centuries before the efflorescence of thought that we associate with the golden age of Greek philosophy and history. There is one possible exception to this: in the soil erosion of Attica, the Epicureans found confirmation for their view that the earth was senile, overpopulated and generally on its way to perdition. If indeed their outlook can be traced to the combination of fire and pastoralism prevalent on the mountains of Attica, there is a powerful link (Hughes, 1994). Even so, the actual course and intensity of the soil erosion has been disputed and its general role may have been exaggerated. But a false report is not necessarily any less influential than a true one.

A parallel ambiguity can be seen for the Renaissance period in Europe. Again, it is possible to interpret this as a flowering of a particular culture, which bears no relation to major changes in production from the land. It is true that nature is relevant, as can be seen in the loving depictions of landscape which now become a part of so many paintings. Initially of the background but in the case of Piero di Cosimo and the school of Giorgione, both around the turn into the 16th century, nature occupies much more of the frame, yet once more, no obvious connection with regional economic change is apparent. The contextual factor to be considered, perhaps, is that of increasing wealth, pride in ownership and in security made possible by trade and technology. Medieval Europe may not have been static technologically though less developed than China in terms of inventions, it had acquired a quantitative approach to matter, according to Crosby (1997). It is possible to suggest that advances in technology during the medieval centuries enabled societies to flourish in which the developments of thought, art, and consumption which could be later constituted as The Renaissance, were cultivated. But the really revolutionary ideas of the period and its aftermath, such as the new cosmologies of Copernicus (1473–1543) and Galileo (1564–1642) where, as John Donne (1572–1631) put it,

> ••• new philosophy calls all in doubt,
> The element of fire is quite put out;
> The sun is lost, and th'earth, and no man's wit
> Can well direct him where to look for it.

are certainly little related to energy efficiency, climatic change, or soil type.

Later centuries show some similarities. Keith Thomas's book (1983) on the re-valuation of nature in England in the years 1500–1800 argues that in many ways the men and women of those years developed a willingness to be more tender towards the living but non-human components of their world. But nowhere does he hold that these are strongly related to material change. He associates a new kindness to animals with the oncoming of industrialization (both in Britain and elsewhere in Europe) but thinks that the conditions for it were in place by 1700, rather before any strong manifestations of the new order. Yet as he says, if the prerequisites were latent in the Judaeo–Christian tradition, why did they emerge at a particular time and place, if it were not for changes in social and economic conditions? The point of nexus for him is the greater use of working animals in the sight of well-to-do townsmen or the clergy, though he embeds this in a wider context of cultural and social practices, mostly of a pre-industrial type. Possibly the growth of city living is more important than the primary economic type; there appears to be no major technological changes of a mediating kind.

In her account of western ideas from the hunter-gatherers of the Americas through to sustainability, Merchant (1995) constructs an inclusive history in which the story of Western civilization since the 17th century can be conceptualized as a grand narrative of fall and recovery. The story is carried by the three subplots of religion, modern science, and capitalism. None of these, we may notice, enforces a particular economy, though capitalism and industrialization seem to be umbilically linked outside the former Communist nations. Individual values therefore are attached

to this narrative as a whole rather than major changes in material circumstances. The same might possibly apply to other meta-narratives; in a different interpretation of Western ideas, Arran Gare (1993) details the withdrawal of philosophy from any attempt at totalizing of knowledge: it no longer deals with the whole world's interactions of nature, history and society. In large measure, this is caused by a drift down the centuries towards the dominance of nihilism. Its current end-point is a highly individualistic society, which involves an extreme detachment from, and instrumentalization of, the world, whether that world be of nature or of other people. People in particular are not participants in the stream of life and in the becoming of the world but transcendent consciousness in a world of things which are only externally related to each other. The watchwords are reason, foresight and efficiency. The connections to material conditions are at most secondary.

This last point seems to emerge from many studies of the possibility of technological determinism, where a considerable body of opinion declares that technology constrains but does not determine. Even highly-coupled technological explanations of socio-economic change cannot account for all the variance in the historical picture: Bimber (1994) argues that even Marx was not really a technological determinist. But there are hard versions of technological determinism in which a sequence of technological development prescribes an evolutionary path for societies irrespective of their social and cultural practices. A softer version would apply this to industrial capitalism but not previous periods of history. Overall, technological determinism looks more convincing over big sweeps of time and place, but the picture seems more contingent at smaller scales. At both, the alternative to a kind of total freedom for technology is its social production. It is becoming increasingly clear that the gender relations of a society will be important in these kinds of connections.

AGENCY AND STRUCTURE

In any summary of worldviews, therefore, the self-image of the human being is intimately bound up with what can be technologically achieved but which then feeds back in a self-aware view of him/herself that must surely affect human–environment relations but at a remove. These images might include: (a) the economic man (sic) of the post-Enlightenment period who is individualistic, materialistic and utilitarian; (b) the competitive individual beloved of the Social Darwinists; (c) the people who were successful because they could deal with the base instincts revealed by Freud; (d) the reductionist product beloved of psychologists who think that behaviour will someday be totally explicable at the molecular level; and (e) the notion that humans are essentially *spirit* in a contingent animal casing. None of these appears to be closely coupled to the evolution of different types of economy. Yet neither are they irrelevant to the construction of the Western worldview, which is now so dominant in all its aspects, including that of environmental thought. The bringing forward of the fragment and of the instant moment seems not to have diminished a belief in an ordered cosmos. Uncertainty within limits seems to be the only reasonable judgement upon the apparent free-floaters: like a kite in unskilled hands, there is a lot of swooping about even if the strings are finite in length.

FRAGMENTATION

Any world chronicle furnishes examples of social changes, which can be interpreted as producing societies in which the demands of the individual are addressed before those of the polity as a whole. Some of these metamorphoses also lead to instrumentalities where the non-human world comes very low in any list of priorities.

In such an account of the Holocene, the starting point can be taken as our reading of hunter-gatherer societies which live within the compass of a unified mythology. The worldview thus engendered may well encompass all aspects of life: the relations between individual humans as well as between the group and its food sources are all referable to the same mythic structure as the social agglomeration and its place in the cosmos. It seems unlikely that such myths always provided for harmony among people or for an unfluctuating relationship with resources (and so should not therefore be romanticised unduly) but nevertheless it provided a coherent explanation of uncertainties at most temporal and spatial scales.

Very similar mentalities may have pertained in the earliest agricultural communities of the hill-lands of the Levant but once successful irrigation agriculture was established in the basins of great rivers (such as the Nile, Tigris–Euphrates, Indus and Huang He), then stratified societies might emerge. They contained classes of people who were devoted to e.g., government, religion, or artisanship and so were no longer in daily physical contact with soil, plants and animals in the way that is inevitable for a hunter-gatherer. Such trends are obviously, intensified in urban societies, though the small absolute size of the pre-19th century city must not be forgotten. Thus both for the townspeople and the farmers, the natural world is potentially an instrument of a cultural set of ordinances and not an intuitively perceived flow through the quotidian existence of all (Shepard, 1982). For some interpreters, this can be seen as the beginning of the regarding of nature as part of the *Other*. The relationships of this type of economy become closer after the development of fossil fuel energy in the 19th century. Modern technology has proved to be a fragmenter of myths, economic systems and land cover systems alike.

Yet between say 5000 BCE and 1800 CE, the world was dominated by economies which depended upon recently-fixed biomass for their existence (Simmons, 1996) and within this pre-industrial world, some trends of social individuation can be discerned. For example, the communal life of the manor in medieval Europe was altered by the withdrawal of the lord and his lady to private apartments behind the hall, a change which eventually resulted in the 17–18th century development of the corridor, allowing genuinely private rooms, especially bedrooms. Stronger movements towards social privatization – from openness to seclusion – start in the later Middle Ages (in Europe at least) and accelerate around 1800. Hence they partly coincide with and partly precede the development of industrialisation, with its reliance upon technology and its subsidy of current biomass by fossil fuels. Instances might include the banishment of (normal) death in public from the middle of the 18th century, death by execution in public disappears (mostly) between 1770 and 1870, and around 1800 prisons and madhouses cease to be public spectacles for the curious. Further, ca. 1800, the private enclosed space of the nuclear family begins to dominate in the west (Spierenburg, 1991).

It is not surprising, then, that further and different kinds of drawing apart should come to pass in the 19th century, given all the innovations of all kinds that the century provided. Steiner (1989) proposes that the 1870s were the critical time, with the poetry of Stephan Mallarme as the emblematic event, since it first explicitly dissociated word from meaning (his words reached out in all directions towards other words and images, being fluid, ambiguous and suggestive, according to one critic) clearing the path for modern semiotics and eventually, an idealist view of the world as being only language. Yet, Mallarme can be seen to have written at a time when many connectivities were being severed. Examples include the pointillisime of Seurat, the making of moving films as images per second, the pictures without object of Kandinsky, the loss of tonality in Schoenberg's music, the atomistic science of Cajal and the invention in Cuba (then a Spanish colony) of the concentration camp (Everdell, 1997). The shifts of 1870–1910 can perhaps be summed up in the transitions of the paintings of Piet Mondrian in those years or given a coda in the anxious splinters of T S Eliot's *The Waste Land* of 1922.

More recent developments include the individuation of entertainment and communications via the Walkman, the VCR (video cassette recorder) and the mobile telephone, the display of the dissected innards and body parts of animals as high art, the withdrawal of public broadcasting in favour of owned channels and its parallel in transport. There has also been a rise in the fragment rather than the prolonged: concert promoters speak of the increased popularity of shorter pieces (so performances of Dimitri Shostakovitch's 7th Symphony [80 min] are getting rare). TV programs gravitate to the magazine format rather than any extended treatment, with children's programmes becoming a total whirl of short spans and desperate camera angles. Most significant of all has been the replacement in more than one national culture of the 5-day cricket game as the norm by various shorter and more colourful versions.

Is there any reflection of these long-term events in environmental connectedness? To quote a set of parallel examples is certainly not to establish a causal relationship. But it is not very difficult to see the setting out of different land cover and water use systems, which differentiate between production, pleasure and nature. In the case of the first, common property resources have undergone constant attrition through several thousand years: the open fields of Europe have steadily become privately owned in the name, largely, of more intensive production. Much open and commonly held land was enclosed and reclaimed as soon as legal and technological systems allowed, examples can be found from the Fens of East Anglia to the American West. The second category can be seen in the way special places are set aside for recreational use in rural areas: the right to roam of e.g., Sweden is a rare phenomenon. Eventually there is the nature reserve and the wilderness in which wild nature (even pristine ecosystems if these can be proven) is set apart in a legal and often physical sense from the rest of the region in which it is situated. The UNCLOS (United Nations Conference on the Law of the Sea) findings attempted to apply the same trends to the oceans. The end-point of this group is found in the assertions of environmental conservatives that only total private ownership of resources and environment will prevent degradation.

COALESCENCE

For many discernible long-term trends in human affairs, a counter-current can be found. If we look at 10 000 years of human history and its relationship to natural history, comings-together can be discerned as well as drawings-apart.

One of the earliest was the formation of empires, which covered large areas and facilitated within them certain uniformities. The Roman Empire, for example, imposed a largely uniform legal system from the Levant to Wales. Thus as early as the 15th century it would have been possible to trade a small and valuable item from Ireland to Japan. Similar processes were carried further by later imperia based on industry; the coalescence of these and earlier world systems has been chronicled by Wallerstein and his critics (Frank and Gills, 1993). The present-day equivalent seems to be the globally instantaneous electronic telecommunications that spin a digital web round the earth rather like candy-floss round a fairground stick. The difference between the 15th century and today is perhaps the way

in which that system is, above all, devoted to processing the symbols of capital and allowing gobs of them to spin off and alight in one place or another to be the sweeteners of various developments. As well as the spinning of money, there is now the transfer of verbal information and image on an instantaneous basis via the Internet. It is easy to over-estimate the influence of such a device since its distribution is rather limited and its driving forces, initially anarchic, seem to be coming under the control of big players. But as an embodiment of Vernadsky's conception (1945) of the *noosphere* (a mind-sphere equivalent to the biosphere and atmosphere; *see* Noosphere, Volume 5), it comes closer than anything so far experienced, more so even than cheap travel by jet aircraft. For capitalistic modernism, the metaphor of the machine for both humanity and nature seems appropriate; the post-modernist world of virtual money seems not yet to have developed a comparable meta-metaphor.

The binding systems with environmental linkages are often of a hybrid human-nature type like the oceans and atmosphere. Not only were these globally interactive in their natural state but the human impacts upon them have the potential to be spread globally. Some influences are coalescent: that of the greenhouse gases is well documented and there is a persistent worry that there could be an equivalent in the oceans since they receive all the substances put upon the land and among these are many compounds toxic to life. The couplings revealed by intensive study of the temperature regime of the Pacific Ocean carry the implicit suggestion that any human-induced global warming may affect the frequency and intensity of for instance El Nino events. All these add up to a world in which the driving force of a population growth, which arrived at 6 billion in 1999, cannot be ignored at any scale. Demographic change can be seen as a non-reversible binding force which determines the size of the gyre which links humanity and nature, as indeed does the degree to which societies have access to commercial energy (Adams, 1988).

Can these trends bring about coalescence and bindings within the organic and human spheres? Without doubt, there are material changes consequent upon empires and trade. The Roman Empire was the cause of the production of wheat throughout much of its extent since it desired that particular cereal for bread. It also affected nature in the pursuit of the circus: North Africa and Asia Minor for example were depleted almost entirely of their lion populations in order to obtain animals for these spectacles; Scotland contributed a bear to one such festive occasion. Trade and discovery encouraged the transfer of species from one continent to another, so that European weeds became as familiar in the tropics as potatoes in industrial cities of northern Europe. This process started as early as the spread of Neolithic crops from the Levant to Western Europe but received enormous impetus during the 16th century conquests by European powers and again in the 19th century's age of empire. The latter, for instance, brought about great ecological similarities in the tropics when different ecosystems were converted to the ubiquitous plantation producing rubber or plant oil.

In today's environments, there are effects in the type of food production encouraged by the Green Revolution and its successor developments. Large amounts of fertilizers, biocides and water produce a uniformity of production methods, which the more firmly installed by the conditions of uniform produce demanded by industrial-nation supermarkets. The general uniformity of several kinds of tourist destinations (is this fenced-off and guarded resort hotel in Thailand, Kenya or the Caribbean?) is another outcome of the globalizing forces. World-wide, the International Monetary Fund (IMF) is the institution which most produces this uniformity since it enforces the conception that every region must do what it does best without regard to local culture or the carbon-loadings brought about by export-oriented economies.

The way in which the application of natural sciences seem most likely to bind globally is via biotechnology, in which genetic uniformity is the most strongly aligned to the demands of world commerce. Such as dictated by global trade organisations and development institutions with their you must do a lot of what you do best [as we judge it]' policies, following the lead of international agencies. The reduction in the genetic diversity of crop plants during the 20th century as a result of commercial pressures to home in on one or two varieties or strains is well known. In that light, the advent of cloning from a single cell, as happened with a sheep in 1997, seems a logical development and full of business potential. It seems unlikely that it will produce a higher biodiversity in the world outside the laboratory, unless by some chance its search for genetic resources allies it to the conservationist movement.

IMPROVISATION AS A VIRTUE

The sociologist Niklas Luhmann (1989) points out that humans cannot communicate directly with the non-human world via speech. Though some New Age enthusiasts would dispute the assertion, most people in the Western world would, after a brief but regretful pause, agree. Luhmann suggests that in fact we talk to each other about the environment and since it is such a complex set of discourses, the discussions are separated out into a set of specialized channels in each of which there is resonance. Each of these channels corresponds to some recognised organizational conduit through which a discussion can flow. Thus we can have environmental economics, law, sociology, politics, religion, education, poetry, music, each containing a set of resonances about what is essentially the same thing but which has been represented in different ways.

One more important cultural facet can be added: in the West there has long been a tendency to dualism. Thus in each channel, the thrust is towards a 0/1 dichotomy: legal/illegal, relative/absolute, scientific/nonscientific, economic/uneconomic, and in the last analysis, good/evil. Again relating to much earlier histories, there is the concomitant likelihood that the 1' state will be Us and the 0' state, the Other. Thus one outcome of attempts to represent the complexity of human–environment relations is a series of separations of both fact (in the sense of the facts of politics, economics and law, for example) and value in the connotation of what is acceptable and what is not. An appealing image for this analysis is that of the pipe-organ. Looking at an organ case, we see a number of pipes of different length which each sounds only one note but which can be on or off: the correspondence to Luhmann's idea of resonances is quite strong. In some organ cases, the pipes are grouped rather as our approach to knowledge is clustered: the sciences, the social sciences, humanities and similar groups. Pursuing the analogy, we remember that the instrument can be played to an exact score as laid down by a composer, or the player may improvise.

Another image that could be helpful is that of the DNA (Di-ribo Nucleic Acid) molecule. Models show a double helix with the strands linked by base pairs. One strand could represent the actual changes in the world and the other our ideas about those transformations. The base pairs then look like the ways in which information about the one is transferred to the other. Some of the pairs will embody the best possible information flow whereas others will necessarily be fragmented or distorted. We can imagine a state in which information about an environmental process is not collected because it is too expensive or where known data are concealed because of the effect they might have on a company's share value. The model could be taken further to suggest the increasing magnitude of the interactions by widening the gyre of the helix and by colouring the strands to show the historic sequence. One problem is that it gives, as do tabular constructions, an air of inevitability: the possibility of contingent change looks unlikely. On the other hand, a creative piece of computer graphics might encompass all these aspects in a way several thousand words cannot.

TRENDS IN ENVIRONMENTAL HISTORIOGRAPHY

The historian William Cronon (1992) has remarked that we have had studies of ecology and economy, or studies of nature, but too rarely have we had all three together. Apart from the actual difficulties of bringing into one focus the different kinds of information provided in such approaches, there is perhaps the point that environmental history has tended to be directed towards certain ends. There are many examples of histories, which are prophetic in the sense of being warnings against a future, which is either inevitable or avoidable by prompt action. In Classical Greece, for instance, the Epicureans thought that the land was senile, overpopulated and that generally both it and the population had declined in their capabilities since an earlier Golden Age when all was in an optimal condition. In the period from 1965 onwards, roughly, there has been a plethora of books on environment, which have traced trajectories of the decline of nature and the rise of overpopulation and pollution. Almost every period in recorded history has been seen as the turning point towards perdition: the Ancient World's eventual materialism, the desacralization of Medieval Europe, the anthropocentrism of the Renaissance, the mechanical model of the world of Newtonian physics as adopted enthusiastically by the industrial revolution, the relativism thought to be made possible by the discoveries of uncertainties of quantum mechanics in the 1920s, and the triumph of capitalism inherent in the fall of the Berlin Wall, have all been seen as signal events and processes on the way to hell. Some histories indeed suggest that a short but pyrotechnic existence is the evolutionary fate of the human species and that we might as well have a hell of a good party even if we are sick in the bushes afterwards.

Within such over-riding themes, shorter periods attract more diverse agendas. In an era of the discovery of the ills of colonialism, the adverse environmental effects of imperial foresters, farmers and white hunters are bound to attract attention though some counterbalance has been offered by the role of colonial scientists in their prescient studies of the relation of rainfall, land use and environmental stabilities. Then again, reaction against some particular aspect of the present is often seminal. In the United Kingdom, the urban sprawl of the 1920s and 1930s reinforced myths about the rural idyll of (especially) the English village and provoked books, poetry and paintings that implanted the more acceptable faces of the pre-1914 landscape in people's minds as the desirable piece of heritage that might have to be fought for. It was not the images of industrial Halifax, or bungaloid Peacehaven that were supposed to inspire Tommy in his trench in Flanders or his tank outside Tobruk, but the combination of thatched cottage, medieval church and friendly pub. In the race for the tourist dollar and yen, the environment of Scotland is a powerful adjunct, no matter that much of it results from centuries of over-exploitation by rich landlords eager for profits from timber, sheep and deer or that the tartan cult is a 19th century invention by an English cloth maker with some spare capacity and an eye for the purchasing power of the military.

Romanticism, on the other hand, was largely provoked by the valuation of the individual over the mass. While it often attracted environmental icing (such as Wordsworth's anathema to the coming of the railways to the Lake District), it

is linkable to changes in the natural world only indirectly: by the growth of human population numbers, possibly. Yet the break with previous attitudes was absolutely critical in the revaluation of for example the mountain and other wild landscapes of Europe and North America, with incalculable consequences for the founding and management of national parks and for the rise of tourism. So the writing of environment history is bound to look out for examples of contingent turning points of an analogous character and the present time of turmoil of thinking and opinion-forming can thus be posited as a turning point from the pre-existing era governed by the machine metaphor to another phase, for which the appropriate language has not been born, but which will share in the renewed valuation of emotion, community and communication: a sort of environmental post-Diana syndrome.

ABANDONING BALANCE

The death in 1997 of the historian of ideas Isaiah Berlin recalls the relevance of some of his thought. He notes that the attainment of Utopias is usually accompanied by coercion on a large scale. The techno-centrics are accused of many sins in their exploitation of primary resources in developing countries, for example; the eco-centrics are labelled eco-fascists and associated with unsavoury regimes that were practitioners of some ecologically acceptable ideas. So Berlin (1991) would not favour the provision of a score with a de nite end in view; rather he is with the improvisers. He also reminds us of the relevance of the fragment of the Greek Archilochus, The fox knows many things, the hedgehog knows one big thing , for it is the Utopians that know one big thing and they are usually prepared to convert everybody to their point of view, like it or not. In our present context, then, Berlin s words seem germane. He suggests that a

> ••• perfect solution is, not merely in practice, but in principle, impossible in human affairs, and any determined attempt to produce it is likely to lead to suffering, disillusionment and failure.

Among other reasons, this is because there are incommensurables to be dealt with, there are choices between paths of equal good, and indeed some ultimate values may be incompatible with one another, then,

> ••• the best that one can do is to try and promote some kind of equilibrium, necessarily unstable, between the different aspirations of differing groups of human beings ••• [this] ••• is not a passionate battle-cry to inspire men to sacri ce and martyrdom and heroic feats. Yet if it were adopted, it might yet prevent mutual destruction, and, in the end, preserve the world.

The phrase crooked timber is from Immanuel Kant, who observed that out of the crooked timber of humanity no straight thing was ever made and in this context, no straight thing has ever been made nor is it worth trying. Fudge, muddle, crabwise-movement and indeed improvisation are the ways in which most attempts to construct a *modus vivendi* are made and which produce the greatest and most lasting successes.

This raises the dif cult issue of the relation between *pluralism* and *relativism*. Many thinkers might agree that the independence of the non-human world imposes in the end some limits, as does the nite nature of the capabilities of the human organism, as indeed do the empirical possibilities of human organisation. Further, to echo Kierkegaard, relativism is possible for history but is much more dif cult in making decisions for the future (Lukes, 1998). There are, naturally, notional confrontations which we may be asked to consider, which are too far removed from the everyday to matter to us as individuals, though they may matter a great deal to others. Then there are real options, in which, as Manuel Castells (1998) exhorts:

> ••• liberation is for people to free themselves from uncritical adherence to theoretical or ideological schemes, to construct their practice on the basis of their experience, while using whatever information or analysis is available to them, from a variety of sources.

In this we can recognise not only the need for pluralism but also that analysis may include the best moral judgement that a shared, locally felt, democracy can produce. No place for Utopia, therefore; neither for existential hand-wringing. So organists need to follow J S Bach all the way: they need to be able to understand the scores and to leave some of them behind as inspiration for those to follow but the real genius is shown in the quality of the improvisations (*see* **Virtual Environments**, Volume 5).

REFERENCES

Adams, R N (1988) *The Eighth Day, Social Evolution as the Self-Organisation of Energy*, University of Texas Press, Austin, TX.

Benton, T (1996) Marxism and Natural Limits, in *The Greening of Marxism*, ed T Benton, Guilford Press, New York, 157 183.

Berlin, I (1991) *The Crooked Timber of Humanity*, ed H Hardy, Fontana, London.

Berlin, I (1992) *Vico and Herder*, The Hogarth Press, London.

Bimber, B (1994) Three faces of Technological Determinism, in *Does Technology Drive History?* eds M R Smith and L Marx, MIT Press, Cambridge, MA, 79 100.

Braudel, F (1972) *The Mediterranean and the Mediterranean World* in *the Age of Phillip II*, Harper Collins, London.

Castells, M (1988) *End of Millennium. The Information Age*, Vol. 3, Blackwell, Malden, MA.

Cronon, W (1992) A Place for Stories: Nature, History and Narrative, *J. Am. Hist.*, **78**, 1347 1476.

Crosby, A W (1997) *The Measure of Reality. Quantification and Western Society, 1250 1600*, Cambridge University Press, Cambridge.

Everdell, W R (1997) *The First Moderns. Pro les in the Origins of Twentieth-Century Thought*, University of Chicago Press, Chicago, IL.

Frank, A G and Gills, B K (1993) *The World System. Five Hundred Years or Five thousand?* Routledge, London.

Gare, A (1993) *Nihilism Incorporated. European Civilization and Environmental Destruction*, Ecological Press, Bungendore, NSW.

Goldblatt, D (1996) *Social Theory and the Environment*, Polity Press, Cambridge.

Grubler, A (1994) *Technology, Changes in Land Use and Land Cover: A Global Perspective*, eds W B Meyer and B L Turner, Cambridge University Press, Cambridge, 287–328.

Harvey, D L and Reed, M H (1994) The Evolution of Dissipative Social Systems, *J. Soc. Evol. System.*, 17, 371–411.

Hughes, J D (1994) *Pan's Travail. Environmental Problems of the Ancient Greeks and Romans*, The Johns Hopkins University Press, Baltimore, MD.

Jardine, L (1996) *Worldly Goods. A New History of the Renaissance*, Macmillan, London.

Kautsky, K (1988) *The Materialist Conception of History*, ed J H Kautsky, Yale University Press, London.

Littleton, C S (1982) *The New Comparative Mythology. An Anthropological Assessment of the Theories of Georges Dumézil*, 3rd edition, University of California Press, Berkeley, CA.

Luhmann, N (1989) *Ecological Communication*, Polity Press, Cambridge.

Lukes, S (1998) Berlin's Dilemma. The Distinction between Relativism and Pluralism, *Times Literary Supplement*, 4956, March 27, 8–10.

Matthews, E (1984) Global Inventory of Pre-Agricultural and Present Biomass, *Prog. Biomet.*, 3, 237–246.

Merchant, C (1995) Reinventing Eden: Western Culture as a Recovery Narrative, in *Uncommon Ground. Toward Reinventing Nature*, eds W Cronon and W W Norton, New York, 132–159.

Oelschlager, M (1991) *The Idea of Wilderness*, Yale University Press, New Haven, CT.

Parsons, H L (1977) *Marx and Engels on Ecology*, Greenwood Press, Westport, CN.

Shepard, P (1982) *Nature and Madness*, Sierra Club, San Francisco, CA.

Simmons, I G (1996) *Changing the Face of the Earth*, 2nd edition, Blackwell, Oxford.

Smil, V (1994) *Energy in World History*, Westview Press, Boulder, CO.

Spierenburg, P (1991) *The Broken Spell. A Cultural and Anthropological History of Preindustrial Europe*, Rutgers University Press, New Brunswick, NJ.

Steiner, G (1989) *Real Presences*, Faber and Faber, London.

Thomas, K (1983) *Man and the Natural World. Changing Attitudes in England 1500- 1800*, Allen Lane, London.

Thompson, W I (1989) Pacific Shift, in *Nature in Asian Traditions of Thought: Essays in Environmental Philosophy*, eds J B Callicott and R T Ames, SUNY Press, Albany, NY, 25–36.

Toynbee, A (1976) *Mankind and Mother Earth. A Narrative History of the World*, Oxford University Press, Oxford.

Vernadsky, V I (1945) The Biosphere and the Noosphere, *Am. Sci.*, 33(1), 1–12.

Globalization in Historical Perspective

JOHN H BODLEY

Washington State University, Pullman, WA, USA

Globalization can be understood as a cultural process in which peoples, cultures, natural resources, and ecosystems from throughout the world are drawn into a single, vast, elite-directed network of production and consumption. Such a network was certainly in place by 1600, but its human and cultural roots lie much deeper. From a broad historical perspective, globalization is a scale phenomenon in which growth occurs in great waves up to the limits set by particular cultural organizational structures, in particular social and natural environments. When a growth threshold is reached, either some form of perhaps cyclical equilibrium is established, or a breakthrough beyond the limit is achieved to a larger scale. Another cultural transformation must take place before another growth wave can occur, and so on. Globalization is simply the most recent phase of a humanly driven cultural process designed to increase the scale of culture. The principal outcome of this growth process is greater concentration of social power. This growth process is fundamentally different from biological evolution, which is concerned with species-level survival, and may not be directed toward a goal of greater organizational scale and complexity.

THEORETICAL PRELUDE

Globalization resembles earlier cultural transformations, but it has produced a profoundly different cultural world because of the absolute quantities of people, energy flow, and materials involved, and the distinctive cultural institutions that globalization requires. The historical evidence suggests that beyond a certain minimum scale, cultural growth naturally becomes an elite-directed process that disproportionately benefits the elite and shifts costs to the majority. This makes growth a difficult process to sustain, because elite decision-makers who enjoy increased benefits, but are relatively unaffected by increased costs, have little incentive to accept natural or cultural limits. As used here, *scale* refers to the absolute size of population, economic values, or any quantitative variable related to human well-being. *Culture scale* refers to specific order of magnitude differences in the size of social systems, and the corresponding cultural features that support them. The theory is that scale-increases occur as an elite-directed process designed to disproportionately concentrate social power under elite control. In this respect, *big* may not be better for everyone, but small economies may be beautiful as Schumacher (1973) has observed.

Scale theory, by emphasizing the role of human decision-making and decision-makers, addresses the theoretical shortcomings of orthodox cultural evolutionary theory, and offers a more useful way to understand globalization. Evolutionary, historical, or sociological explanations of cultural development that take society as the unit that adapts itself under the influence of natural selection are theoretically unsatisfactory, because such explanations make the growth process appear both natural and inevitable. These approaches also tend to ignore, or obscure, the role of human decision-making agents, and the differential distribution of the costs and benefits of growth. Scale theory takes individuals and households as the active, decision-making units, and treats socially expressed culture, ideology, technology, and institutions as the means, or survival vehicles used by particular human agents to achieve their objectives and improve their life chances . This assumes that humans are primarily driven to survive and reproduce, and to ensure that their children will be able to do the same. In support of this primary biological objective, individuals may have an innate, although variable disposition to dominate and exploit others whenever possible, and they will seek to minimize their own effort by getting others to work for them. However, the existing institutional structure of culture sets limits on competitive striving by individual decision-makers.

Scale theory is theoretically related to the evolutionary materialism approach outlined by sociologist Stephen Sanderson (1990, 1999), and the concept of dynamic materialism developed by economic historian Graeme Snooks (1996, 1997, 1998). The common assumption of these approaches is that globalization was not pre-ordained, is not

inevitable, and is perhaps not even an adaptive, or beneficial process for humanity as a whole, in contrast to the claims of orthodox evolutionary theory, and many current advocates of globalization. Robert Wright (2000) for example, claims that globalization is the product of natural cooperation, not competitive struggle.

It is inappropriate to apply a purely biological model of evolution to cultural development because culture and cultural information are transmitted much more easily than genes. Culture change is an intentionally-directed process, that occurs through purposeful invention, diffusion, and/or imitation, rather than through random mutation and natural selection. However, because human genes and culture affect each other, it is more accurate to acknowledge that human development is a co-evolutionary process involving both culture and biology (Boyd and Richerson, 1985). Furthermore, civilizations do not reproduce and multiply, rather, they have historically grown larger and fewer. Carneiro (1978) and Naroll (1967) graphically demonstrate the long-run decline in the number of polities in the world, and Taagepera (1978a,b, 1997) shows how they have increased in territorial size. Civilizations may appear to compete with other civilizations as systems, but their human leaders are the actual competing units. When evolutionary success is measured by either adaptability or reproductive success, civilizations fare poorly as evolving organisms , because their diversity, numbers, and longevity decline over time. Culture itself was the key to human adaptive success, but the growth of human cultural *systems* has been curiously maladaptive for humanity as a whole because it has reduced cultural diversity, made patently less-sustainable cultures dominant, and actually increased the absolute number of people living under suboptimum physical conditions.

GLOBALIZATION AND SCALE THEORY

If globalization is fundamentally a scale phenomenon, it is important to consider the physical significance of size. The scale of human society and culture is important theoretically because it is related to four crucial variables:

1. the level and distribution of household well-being, standard of living, or human costs and benefits;
2. the distribution of decision-making power;
3. the institutional complexity and technological organization of society;
4. the trajectory of growth and the dynamics of change.

Scale-increase is a disruptive, increasingly expensive process that changes all of these variables, and disturbs pre-existing equilibria. It is difficult to imagine why natural selection would ever favor growth in the scale of society and culture in general, or the development of a global-scale culture in particular. Size is a crucial variable in nature, where large things are understandably rare. Small animals can survive on fewer resources than large animals. They reproduce more rapidly, recover more quickly, and can adapt to change more easily than larger animals. This may be the reason why bacteria are the most successful organisms on earth and perhaps in the universe, considering total number of individuals, diversity of species, total biomass, and overall adaptability as the measure of success. Likewise, people living in small cultures can respond more quickly to environmental changes, and can make decisions more democratically than people in larger societies. Energy-intensive communications and transportation technology make it possible for global-scale societies to respond quickly to environmental crises, but bureaucracy and power inequalities bias such response. Because of their relative social equality, people in smaller societies are more able to share the benefits of their success. Scale itself may be the most important variable, because basic human nature can be assumed to be a constant.

The physical laws of gravity, and geometry mean that growth is an allometric process, in which internal relationships change disproportionately. For example, surface area increases geometrically by the square of length, and volume and mass increase by the cube of length. These principles also apply to cultural growth and development, and especially to globalization. The most critical disproportion for human societies is in the distribution of social power and decision-making abilities. Social power naturally becomes disproportionately concentrated in the elite as societies grow larger, unless specific cultural countermeasures are applied by the majority. This helps explain why cultural growth occurs, and calls attention to who directs change, and who most benefits from it.

Like biological organisms, the size and form of human societies occur within relatively narrow ranges determined by thresholds set by physical laws and the functional limits of culture. The population of autonomous small-scale, domestically organized prehistoric societies averaged only 500 people, and typically ranged within magnitudes of under a threshold of 10^3 (1–9 thousands), although there were much larger tribal world systems. Pre-modern politically-organized city-states, and agrarian empires did not normally exceed population magnitudes of 10^7 (10–99 millions), also organized into agrarian world systems. Commercial organization produced a global economic world system that is unlikely to ever exceed a magnitude of 10^{10} (10–99 billions).

Human settlements also display remarkable, mathematically-predictable scale regularities. For example Roland Fletcher (1995, 71–81) found that settlement size appears to be limited by the increasing difficulty of interpersonal communication and the stress of interaction that can only be overcome by organizational, technological, or design changes. The interactions amongst people in

a settlement increase exponentially as population grows, such that a 10-fold increase in population would theoretically produce a 100-fold increase in potential interaction. Domestically organized villages typically remain below a threshold of 200–500 persons.

The urbanization process, or the transformation of villages into cities, was the first crucial step toward globalization. Cities have always been at the center of socio–cultural transformation and economic growth. Urbanization began when power-seeking elites successfully developed new political and economic institutions to overcome the limits to societal growth that maintained domestic-scale villages in equilibrium. Fletcher (1995, 82–90) concludes that the total area of urban settlements, and the rate of urban growth are limited primarily by the technology of interpersonal communication. As cities grew larger, commercial exchange became a virtual necessity, because political power can only effectively command tribute payment within relatively narrow limits. The earliest cities became possible, some 6000 years ago, with the development of bureaucratic political organization, supported by writing, or its functional equivalent. These cultural changes permitted a 100-fold increase in the area of the earliest cities, above the area of pre-literate, domestically organized villages, and a 1000-fold increase in areal growth rate.

Andre Gunder Frank (1990, 1991) and Gills and Frank (1991) argue that globalization is the culmination of a growth process that began with the world system some 5000 years ago based on the earliest Mesopotamian city-states. The first cities of 25–45 000 people required lengthy and expensive exchange networks, and imply the existence of at least a regional world trade system for their material support. Five thousand years of continuous growth in this world system required massive transformations in cultural organization and technology. Cities of 100 000 people or more, which appeared by perhaps 500 BC, required monetary systems, and the large-scale commercialization of food systems which ultimately prepared the way for the capitalism world system. During the 4000 years up to 1500 AD, the number of such large cities in the world steadily increased from perhaps 8 in 2000 BC to 75 by 1500 AD. Their average population size increased from 30 000 to 100 000 people, with the largest over 500 000. By 1500 AD the total number of people living in such cities may have exceeded 7 million people (Chandler, 1987; Sanderson, 1999, 112; Wilkinson, 1992, 1993), and would have required the vast exchange system that became the modern global-scale world system. This new world economy initially grew out of a European foundation between 1450 and 1640, and became a truly global-scale world system by 1815 (Wallerstein, 1974). Globalization is an urban-centered phenomenon. Since approximately 1850, the development of fossil fuels, railroads, electronic communication, mechanized printing, and the emergence of large-scale corporate businesses permitted a 100-fold scale increase in the size of cities beyond the previous threshold of 100 km^2 and 10 million people, and a 1000-fold increase in urban growth rates.

THE DYNAMICS OF CULTURE SCALE

During the past million years proto-humans and humans have experienced eight great transformative waves of cultural development (Table 1), each producing societies of successively larger scale, and a larger global population. In this development process, the dynamics of culture change itself changed, becoming more rapid, and more narrowly directed. Globalization is only the most recent such transformation. The first transformation may have been completed 100 000 years ago. This coincided with the appearance of physically modern humans, and fully human culture based on shared, socially-transmitted, symbolic information, and speech. The second transformation beginning perhaps as early as 50 000 years ago saw the development of highly efficient foraging technology which allowed people to occupy virtually the entire globe. Growth to the limits of efficient foraging produced an equilibrium world of perhaps 10 million people, with leadership widely dispersed. The third transformation was the change from mobile to settled village life based on domesticated plants and animals during the Neolithic, which began at the end of the Pleistocene 10–15 000 years ago. The Neolithic wave produced a world of tribal villagers with perhaps 85 million people.

Cultures can be grouped into three broad-scale types according to whether social power is organized domestically, politically, or commercially. Social power is the way individuals and households secure the material, social, and cultural resources that they need to maintain themselves and successfully reproduce. It is assumed that everyone is naturally driven to seek social power, although the specific form that power may take is an important cultural variable. When social power was domestically organized, all households

Table 1 Great cultural transformations

Domestic scale cultures
Hominoids • Humans, 100 000 BP
Humans • Forager cultures, 50 000 BP (speech, art)
Foragers • Village farmers, 10 000 BP (domestication)
Political scale cultures
Villages • Chiefdoms, 8000 BP
Chiefdoms • Agrarian civilizations, 6000 BP (urbanization)
Commercial scale cultures
Agrarian civilizations • mercantile capitalism, 1450 AD
Mercantile capitalism • industrial capitalism, 1800 AD
Industrial capitalism • financial capitalism, 1900 AD

had relatively equal access to crucial resources, and there was no incentive for growth in either the scale of society or economic productivity. Domestic cultures displayed demographic equilibria in general balances with their resources that could be measured in millennia. However, they were not totally static. It is a striking fact that most important domestic technologies including weaving, ceramics, and domestication were produced in domestically organized societies, and the benefits were universally distributed.

An important cultural organization and scale line was crossed some 8000 years ago, when the first aspiring power elites succeeded in transforming the characteristically weak, temporary, and informal village *headman* position into the coercive political of ce of *chief* that could be transmitted cross-generationally within a ranked descent line system. In this *politicization* process, village chiefs were then able to extend political power over more than a single village, producing simple, and then larger, and more complex chiefdoms that encompassed thousands of people. As soon as it became possible for individuals to consolidate their power gains, and transmit them to their children, the move up the scale to city-states, kingdoms, and agrarian empires occurred relatively quickly. The key difference between this kind of elite-directed transformation, and the previous domestically directed developments, is that the bene ts and costs were not equally shared. In this respect the politicization process is not fundamentally different from the contemporary commercial globalization process. Under politicization, elites assume life or death power over commoners, and extract surplus energy from them, treating them much like domesticated animals. Elites enjoy superior life chances, because they control vast material surpluses and can insulate themselves from natural shortages. Elites live in larger, more secure residences, and they surround themselves with domestic servants, and multiple wives and concubines. The scale of elite power corresponds directly with the scale of society, thus producing a natural incentive for elites to promote further growth. Commoners must feed themselves, pay taxes, ght wars, suffer famines, and become human sacri ces, either in temples, royal tombs, or on battle elds. Elite-directed cultural developments such as urbanization, intensive agriculture, monumental architecture, monotheism, writing, and military technology are designed to support political power and exclude non-elites.

How this great political transformation occurred cannot be fully explained, but power-seeking individuals had to overcome the limits that the majority placed on their leaders. It is likely that this occurred when the majority was forced to choose among the lesser of evils under crisis conditions. One of the paradoxes of large-scale, politically-organized societies is that they are conspicuously unstable. Social inequality and continuous expansion are dif cult and costly to sustain using political power. Ancient agrarian civilizations displayed a persistent pattern of growth and collapse, with a long-term trend toward increasing total area, population, and urbanization. Equilibrium was an exception, and few civilizations lasted more than 130 years, and only rarely did they exceed 300 years (Taagpera, 1997). The threshold for politically-organized empires was probably reached by the British Empire which brie y controlled one-quarter of the world s population and land area. Growth beyond that level called for a different form of organization, and this requirement was met by an expansion of commercial enterprises.

THE CAPITALIST ROOTS OF GLOBALIZATION

Globalization is a cultural process in which economic elites use commercial exchanges to create a supra-national, global-scale social system, a world-economy , that operates without a single centralized political power. As Wallerstein (1974) observes, an economically-based world system makes possible a much larger, more ef cient, and apparently more resilient system than could be integrated into a world-empire by political power alone. Globalization was created by the great cultural transformation that accompanied the rise of capitalism in Europe. Thus, capitalism itself is the key to understanding globalization. There have been many formal de nitions of capitalism, but its most critical feature is that it supports the rise to dominant decision-making power of commercial elites, who are able to concentrate social power by co-opting, and at times even directing the politicization process in order to promote their commercial projects. Capital is social power, whether in the form of nance capital, xed capital in buildings, equipment, and inventory, or land. Capital accumulation, rather than consumption, is the primary objective of a capitalist economy. Under capitalism, the commercial elite is dominant, but political rulers may sometimes also take a direct entrepreneurial role. Rulers support and defend commercial activities, they supply crucial infrastructure, and help to maintain and develop human capital by supporting public education, health, and welfare. The capitalist commercial elites are politically free to develop new institutions and organizational forms that allow them to accumulate capital by commercial business enterprises.

The minimal formal elements of capitalism include developed markets for land, labor, and money. In classical Marxist terms, capitalism can be readily de ned as an economic system in which money is primarily exchanged for commodities, which are in turn exchanged for more money. This is production for commercial exchange, rather than for domestic consumption, or use value. The dominant purpose of exchange is to obtain a monetary pro t. The social structure of capitalism is distinctly unequal. Adam Smith (1776) described this inequality as *specialization*, or a division of labor between landlords, manufacturers, and laborers,

arguing that laborers naturally needed masters. When viewed as a total economy (Snooks, 1994), a national capitalist economy has three sectors, with a household economy at the center, which produces, maintains, and allocates labor, and receives wages, salaries, interest, and goods and services from the private commercial sector and from the public, or political sector. The national economy is not a closed system, but gives and receives economic flows to and from other national economies. The basic organizational elements of capitalism include a world-economy with a geographic and hierarchical functional division of labor, and with strong national states at the core. The global hierarchy is expressed as an unequal international distribution of costs and benefits in which the holders of capital are rewarded more generously than those who have only their labor to exchange. Maldistribution of power thus seems to be an intrinsic feature of this system, especially in the absence of a strong world-empire government.

A long-running intellectual debate over the origins of capitalism and globalization concerns the relative importance of ideological factors versus material conditions. Anthropologist Marshall Sahlins (1996) places capitalism within a broader European cosmology that views people as naturally evil and selfish. From this perspective, economic scarcity is also a natural and perpetual human condition, and it is a given that societies are controlled by hierarchical power structures and coercive governments. These ideas are sustained by the salutary beliefs that everyone will benefit from economic growth driven by free market capitalism. These fundamental beliefs also explain the quest for increased social power and technological progress, which are thought to be natural and inevitable.

Many anthropologists are cultural materialists, arguing that economic practices are primarily determined by material conditions. However they acknowledge that cultural ideas, or beliefs, are an important part of human motivation. The beliefs underlying capitalism can be treated as the *ideological* causes of globalization, because they rationalize the cultural transformations that produces globalization. From a materialist and a scale perspective, elite power-seeking is the primary motivation for globalization. However, the opportunities for power-seeking are limited by both environmental and cultural conditions as shaped by historical circumstance. Diverse belief systems can be mobilized to rationalize many different power-seeking strategies, and in that respect must be considered secondary to the personal motivation of power-seeking. To some extent, globalization was a consumer-driven process, but consumer demand is highly constrained by the cultural organization of power, political economy, and political ecology. Globalization was also not just a product of the market economy or monetization, because for millennia organized markets and money have linked villages, towns, and cities in all the major civilizations. However, local markets have primarily traded in domestic staples, and benefited society as a whole. It was long-distance trade in luxury goods for the elite that made it possible for merchants to take advantage of profit-producing price differentials and offered the initial step toward globalization.

THE EUROPEAN ORIGINS OF GLOBALIZATION

In order to explain the origins of the modern capitalist world, social scientists at first focused on European science, exploration, and technology. Initially, they considered the Industrial Revolution to be the principal architect of material progress. However, 20th century historians demonstrated that capitalism preceded industrialization by centuries, and they argued that the long-term process of capital accumulation gradually produced widespread improvement in human welfare. For example, Snooks (1993, 4) ecstatically proclaimed that industrial growth since 1750, ••• massively shifted the distribution of income in favor of the lower socioeconomic classes . At the same time many postmodern anthropologists argued that the commodities that were at the heart of globalization could be understood by their culturally unique symbolic meanings for individual consumers, and their use in the construction of individual self-identity (McCracken, 1988; Miller, 1987). Some even described commercialization as a democratic process in which markets, commodities, consumption patterns, and their cultural meanings came to be mutually constructed by marketers and consumers. Like earlier theories of cultural evolution, these postmodern interpretive approaches make capitalism and globalization seem both natural and inevitable, and avoid crucial issues of the distribution of material benefits and social power.

The capitalist global system was developed by merchants and investors centered successively in Venice, Antwerp, Genoa, Amsterdam, London, and New York. A growing number of elites in these urban trade centers enjoyed ever-higher living standards and consumption levels, whereas wageworkers, serfs, or slaves gained little control over the conditions of daily life. Many Europeans did experience some improvements over the depressed living conditions and chronic insecurity they had experienced under feudalism.

The economic elites who created the modern world system were responding to a historically unique set of circumstances. The crisis of feudalism from 1230–1400 that peaked in 1350 was the immediate precursor of the push for European commercial expansion overseas. This prolonged crisis was an unfortunate conjunction of bad weather, and overpopulation produced by gross inequality and over-consumption by the nobility. This situation created a century of hyperinflation, peasant rebellions, wars between the nobility, famine, and disease. The nobility

raised rents to stay ahead of inflation, but the less fortunate majority saw their absolute wages decline. The political and economic turmoil made it advantageous for the merchant and political elite, who were the only people capable of strategic economic planning, to turn to more capital-intensive technology, seek foreign resources, and invest in mercantile capitalism. Furthermore, the existence of autonomous market towns under the politically decentralized European feudal system created an obvious opening for enterprising merchants to expand their power (Sanderson, 1999). Economic crises spurred the growth of financial capital because it forced governments to borrow from private lenders. Another round of inflation and turmoil from 1470–1610 corresponded with the beginnings of long-distance overseas expansion, which was followed quickly by the establishment of the Bank of England in 1694, and the London stock exchange in 1697.

Historian Fernand Braudel (1977) argued that successful commercial families were able to manipulate political and economic power within very hierarchical, but not totally closed European societies. A key element in their success was that they managed to transmit and accumulate their economic power across generations. Capital accumulation by a few households in an intensely competitive cultural environment confers a decisive advantage to the successful, as Snooks (1998) has observed. Accumulation would not be advantageous in a culture that distributed resources equitably. The fact that the European population had recovered from the 14th century devastation of plague and war by 1650, and began a long upward curve of exponential growth by 1800, suggests that most people were gaining some material benefits from capitalism and globalization. Many ordinary households intensified their labor, producing more for the market, and engaging in wage labor as opportunities arose in the expanding cities. In this respect Europeans of all social strata participated in globalization, and benefited from it to some degree. However, very few were able to accumulate investment capital and secure economic independence.

Some historians suggest that a consumer revolution or a democratization of consumption began in England and America as early the 18th century, and regard this as the first appearance of mass consumption (McKendrick *et al.*, 1982). Such an outcome would be evidence for a widely shared benefit of globalization. However, even long after the beginning of globalization, living standards remained extremely inequitably distributed, and the benefits of higher consumption levels spread very slowly. Consumption by the relatively powerless poor in the vast lower class was limited to the necessities of food and domestic durables. The shopkeepers, merchants, and professionals in the small middle class, could afford the comforts and decencies of house, groceries, bedding and tableware. The costly luxuries of the leisured lifestyle were reserved for the handful of households in the upper gentry and aristocracy, who engaged actively in competitive emulation. As the global trade network of mercantile capitalism developed, newly imported commodities were generally consumed first as luxuries by those in the highest social ranks, and were then very gradually diffused to the lower classes. By the 1680s, English merchants were importing different colors and qualities of cloth to appeal to specific social ranks, and giving out free samples to the nobility, hoping to both promote and profit from competitive emulation. The earliest imported consumer goods to become objects of mass consumption in England were pepper and dried fruit by 1559, tobacco by 1650, sugar by 1690, and tea by 1730 (Shammas, 1990, 1993). Wills (1992) stresses that new consumer markets did not just happen. They were created by intense competition between rival international trade monopolies.

Theodore Rabb (1967) provides an excellent picture of the initiators and prime beneficiaries of globalization who were centered in London between 1575 and 1630. In a given year, probably only 2500 individual investors owned the 33 joint-stock companies that actually funded and directed England's mercantile-capitalism overseas expansion and thereby transformed the world. These economic elites represented no more than 0.25% of all households in the kingdom of some 5 million people. Three quarters of these investors were wealthy merchants who picked the only route open to them for material advancement, given their non-aristocratic social position. Only one quarter of the investors were drawn from the hereditary landed aristocracy, who represented the top 2% of the social hierarchy. Historian Lawrence Stone (1965, 1973) shows that the aristocrats were diversifying their investments in order to maintain their privileged social status which was threatened by the unsettled conditions accompanying the collapse of feudalism.

European economic growth was certainly boosted by the 10-fold increase in the existing stock of gold and silver represented by imports to Europe between 1500 and 1680. The degree to which bullion was narrowly controlled is a truly astounding reflection of the elite domination of the emerging global system as a whole. At this time, the Spanish hub of the New World silver trade was in Mexico city, where in an average year a mere 44 large investors handled more than 60% of total cargo value in trans-Atlantic commerce (Hoberman, 1991). The largest English development enterprise of the era, after privateering, was the British East India Company, which was founded in 1600. The East India Company proved to be an excellent investment. Seven of its trading expeditions conducted from 1601–1612 produced an average annual income of £132 for each investor. In England at that time 98% of households lived on less than £120 a year. Of course many of the larger investors owned shares in several different companies, and

also received rents and other income. According to Gregory King's (1936) careful estimates, by 1688 some 2000 large merchant traders were earning £400 a year, putting them economically almost equal with knights and esquires in the gentry class (King, 1936). Exploitation of low wage workers, slavery, trade monopolies, and privateering produced very high rates of returns on investments, but the very highest returns were enjoyed by aristocrats at the highest level of government who supplemented their investments with royal favors, bribes, and abuse of public office (Stone, 1973, 3–15). A few very high incomes undoubtedly accounted for much of the economic growth during this period, and not all was invested in further growth. Between 1600 and 1688 the total wealth of England more than tripled, and much of the new wealth financed a frenzy of luxury building and consumption by the elite.

AGRARIAN TRANSFORMATION AND LONDON POVERTY, AD 1500–1700

The initial development of the modern world economic system is mirrored in the transformation of the English agrarian system between 1500 and 1700 AD This transformation was carried out simultaneously with overseas expansion, and was directed and financed by many of the same few individual investors. The commercialization of grain removed a feudally-organized agrarian food system in which a land-owning aristocracy extracted surplus grain from a peasantry that was largely self-provisioning. At least 15 regional grain markets had emerged by 1500, and some 800 small towns grew up around local marketplaces. Similar conditions and transport factors meant that grain could be sold at the same price in each distinct market region (Gras, 1915). Even though local farmers could sell in any market town, the regional markets remained relatively isolated. Few specialist merchants were able to take advantage of price differentials in different markets. Local municipal authorities imposed numerous regulations on grain handling and marketing, all apparently designed to protect the consumer from price manipulations, poor quality, or dishonest measures. Middlemen grain traders, or wholesalers were regarded with suspicion, and widely assumed to be disreputable.

The city of London's population grew 10-fold between 1500 and 1700, as the city became the center of the expanding global economy. This order of magnitude population growth soon overwhelmed Britain's medieval system of local and regional grain distribution. In 1250 London drew on a local grain market within a radius of 40 km, but by 1500 the city had grown to some 50 000 people, and the local market proved totally inadequate. London needed a much larger distribution system that could draw from all over England. In response, the volume of coastal grain shipments increased 10-fold between 1580 and 1680, and a single national market emerged. After 1500 imported grain became increasingly important in the London market. Previously, the foreign grain trade was handled by foreign merchants, and largely in response to famines. The grain supply was so problematic that from 1514 to 1678 London's municipal authorities bought and sold grain, established public granaries, and urged the guilds to buy and store grain for resale to consumers at reasonable prices during shortages. By 1700, London grain merchants were trading with Scandinavia, central Europe, Holland, Spain, the Caribbean, New England, Virginia, and Canada. As scale theory would predict, the new global food market disproportionately benefited a few very large merchants. Just four merchants handled half of London's wheat export business during three trading seasons between 1676 and 1683 (Gras, 1915, 197).

London's population reached a phenomenal 700 000 people in 1700, and the most devastating plagues and famines seemed to be over. Most of the national population was no longer in control of their food distribution system, and with the large land-holders fencing off the open fields that the peasants had used for subsistence, and converting them to pastoral use, most rural households were also losing control over food *production*. Most foodstuffs, including grains, had become commodities that everyone had to purchase with cash. England's national grain market was absorbed into a single global market. One could of course argue that the subsistence-level peasants ultimately benefited from these changes in the food system. However, no one should be forced to choose between rural, subsistence-level servitude, and urban poverty and wage-servitude. By the late 19th century conditions had clearly worsened for the majority. The New Domesday Book shows that 75% of the rural land in England and Wales in 1875 was held by just 14 000 large landholders, who represented a mere 0.28% of all households in a population now grown to 25 million people (Bateman, 1883). This was not an improvement over feudalism. 95% of households were left with less than 0.5% of all privately owned land. At the same time Charles Booth (1892–1903) found that 1.2 million people, more than 30% of London's population, were living in poverty as defined by inadequate income and crowded housing conditions. The 18 000 wealthy, upper class people represented the top 0.5%. These data suggest that the previous four centuries of economic progress increased the absolute number of poor in England, and left most people with less control over the conditions of their daily life.

INDUSTRIALIZATION AND URBANIZATION IN AMERICA, 1790–2000

Between 1790 and 1990 the American population increased 65-fold, and national income expanded an incredible 1000-fold. Even though America was widely perceived of as

a land of opportunity, the actual distribution of income and wealth changed little and remained very inequitable throughout this period. Vast areas of Native American land were opened for settlement, but much of this land was difficult and expensive to develop, or it was controlled by land speculators. Economic growth did produce a large middle class, but the absolute number of very poor also increased, whereas the elite became inconceivably powerful. The primary cultural changes from mercantile capitalism to industrial and financial capitalism that drove America's economic growth were reflected in changes in the agrarian system, much as in the earlier transition from feudalism to mercantile capitalism. In 1790 America was a rural nation in which 95% of the population lived on small, self-provisioning farms that produced limited marketable surpluses. By 1990 America was 75% urban, and agriculture was large-scale, capital-intensive, industrialized, and fully commercialized. This elite-directed cultural transformation was largely carried out within the space of four decades between 1850 and 1890, and clearly disadvantaged small-scale agricultural producers (Berry, 1977).

America's transformation required a series of massive changes in transportation, communication, production, distribution, and in the organization of business enterprise itself, that collectively expanded the speed and scale of commercial transactions (Chandler, 1977). The most important change was the creation of bureaucratically-organized, corporate businesses, that combined multi-unit enterprises and could amass capital and grow to a scale never before possible. Remarkably, few people actually financed and directed these changes. Historian Peter Hall (1973) argues that America's great national expansion was initiated in the 1850s by a tiny group of some 200 Boston families. These families were each worth at least $100 000 in 1848, owned 37% of the personal property in the city, and were the major shareholders in Boston's 50 largest banks and insurance companies (Pessen, 1973). At that time, Boston was a city of 137 000 people, much like London in 1580, and was the banking center for the entire country. The Boston elite accumulated trans-generational family fortunes from a mercantile base by investing prudently, marrying carefully, and forming dynastic trusts (Friedman, 1964; Marcus, 1992, 60–70; Farrell, 1993). According to Hall, these wealthy New Englanders financed the first great American business corporations, and used their philanthropy to create the elite private foundations and colleges, such as Harvard and Yale, that trained the professionals who then built the institutional and ideological infrastructure for a commercially focused, elite-directed American nation.

As America grew, the number of millionaire investors centered in Boston, New York, and Philadelphia expanded from an estimated 40 in 1860, to 2500 in 1870, and 3800 in 1897 (Soltow, 1975, 112–113; Beard and Beard, 1934, 383–384). At the same time more than half of the adult male population remained landless, in spite of the Homestead Act of 1862 and the westward expansion movement. By 1900 there were perhaps 1500 industrial corporations with capital assets of $1 million or more, directed by some 18 500 corporate board members and major stockholders (Roy, 1983). Collectively these 22 000 financiers and corporate directors formed an economic elite that represented less than 0.15% of American households in a nation of 76 million people. Their decisions to expand the scale and scope of commerce had transformed the nation and the world by the end of the 19th century.

After 1900 small farmers were steadily marginalized by the technological changes promoted by the land grant colleges, the Department of Agriculture, and the Extension Service which promoted the interests of the largest agricultural producers, in response to corporate interests (McConnell, 1953; Domhoff, 1996). These new institutional structures in combination with the railroads, financiers, farm machinery manufacturers, the petroleum and chemical industries, and mass-merchandisers account for the rapid dominance of the large-scale factory farm system. By 1987 the 100 largest corporate farms accounted for more than 10% of all agricultural production (Krebs, 1992, 28) and by 1992 14% of total farms accounted for 45% of all farm product sales (US Department of Commerce, 1996). This is the top-heavy agricultural model that was later promoted overseas as the Green Revolution, where it had the same marginalizing effect on small farmers.

By the 1920s New York City had become the center of the evolving modern world system. Europe's immigrant poor crowded into the slums of New York, where by 1900 some 2.3 million people, 70% of New York's population, were jammed into unsanitary tenements (DeForest and Veiller, 1903). America's commercial-elite had already accumulated the world's greatest concentration of economic power under the control of a few wealthy families, giant industrial corporations, and financial institutions. Many New York families, financiers, and corporate directors and their descendants played a leading role in creating the national and international institutional structures that produced a vast increase in the scale of the global economy. For example, John D Rockefeller's personal fortune of $200 million gave him enormous influence. However, financier J P Morgan's effective power was much greater. By 1912 Morgan exerted direct or indirect control over every sector of the national economy, with directorships or controlling interests in 34 large financial institutions, 10 insurance companies, 32 transportation systems, 12 public utilities, and 135 other corporations, with combined assets of $45 billion (Brandeis, 1914). In 1994, one individual directed five giant corporations with combined capital assets of $549 billion, and ten people directed 37 companies with combined assets of $2 trillion, representing nearly 10% of all corporate assets in the country (Bodley, 2000). In 1998

there were three executives who each controlled financial assets of more than a trillion dollars in the three largest Wall Street investment firms (Califano, 1998).

TRIBAL VICTIMS OF GLOBALIZATION, 1450–2000

The globalization process included several phases of colonialism in which peripheral and external areas were incorporated into the expanding world system. Millions died during the initial mercantile-capitalist phase from 1450–1800 when many of the world's great agrarian civilizations were brought under European domination. During the following industrial phase from 1800–1900, much of the rest of the world was added to the periphery of the world system. Europeans settled in the most favorable areas, and turned other peripheral areas into sources of raw materials and labor. However, dispersed colonies proved too costly and difficult to administer, and external colonialism came to a formal end in most of the world shortly after 1950. Internal colonial conquest continued throughout the 20th century until virtually all of the most isolated autonomous domestic-scale peoples and cultures were incorporated in the world system.

At the beginning of the Industrial Revolution, Japan had completed its independent transition to capitalism, but ancient civilizations still existed in China and Africa, and there were numerous traditional kingdoms and/or chiefdoms in India, the Middle East, and the Pacific. The great Western colonial powers claimed 55% of the world's land area, but exercised effective political control over only a third of the world (Clark, 1936, 5). Perhaps 200 million indigenous people, roughly 20% of the global population, still lived in autonomous and largely self-sufficient domestic-scale, tribal cultures, and still effectively controlled approximately half the world. Over the 150 years between 1800 and 1950 virtually all of this territory was conquered and incorporated into the world system in a process of internal colonialism in which as many as 50 million indigenous people may have died (Bodley, 1999a). Natural resources and labor flowed into the world system from these formerly external indigenously controlled territories. This final conquest produced a fully global modern world system, but at an enormous cost in ethnocide, genocide, and ecocide suffered by the peoples and territories that were forcibly incorporated.

The conquest and incorporation of indigenous people and territory not only destroyed the economic autonomy and political independence of millions of indigenous peoples, it also had a devastating demographic impact. Demographic disruption often initially took the form of drastic depopulation, when people were enslaved and killed by invaders, and when they died from newly introduced diseases to which they had little immunity. Many cultural groups that relied on fertility-limiting cultural practices to maintain low population densities were quickly exterminated by sudden increases in mortality. Many indigenous peoples died as a direct result of violence perpetrated by outsiders. Many governments organized military campaigns against tribes, and a few deliberately pursued genocidal policies in an effort to totally exterminate certain tribal groups. The largest loss of life probably resulted from the actions of individual colonists who acted on their own before governments established formal political control. Colonial governments sometimes intentionally left certain regions as uncontrolled frontiers, where their citizens were allowed to systematically kill and exploit tribal peoples. In areas where tribal people were not needed as labor, they were sometimes classified as non-human , and were killed with impunity. As early as 1837, under pressure from humanitarians, the British government officially acknowledged the human cost of colonialism in a lengthy Imperial Blue Book on aborigines (House of Commons, 1837), but the imperialism went unchallenged.

As soon as indigenous peoples lost control over their territories and subsistence resources, it often became impossible for them to remain economically self-sufficient. Depopulation also reduced the economic viability of tribal groups, but competition with colonists over resources, especially when the tribal land base was reduced by government decree, was often the decisive factor compelling native peoples to participate in the market economy, whether as wage laborers or cash croppers. There was also a desire to secure manufactured goods, such as metal tools and factory clothing. However, the desire for trade goods was usually not sufficient to force tribals into full-time, permanent participation in the market economy. As long as indigenous groups retained a viable subsistence economy, they were often poorly motivated target workers, and colonial administrators used special taxes and laws to force them into the market economy. Involvement in the market economy became a self-reinforcing cycle, because wage labor disrupted traditional subsistence activities, and cash-cropping depleted natural resources making self-sufficiency even more difficult. The problem for most indigenous people was that they were forced into the global system at the very lowest level, without the economic resources to operate effectively in a capitalist world, and without the political power to protect their interests. Where reservations were set up for conquered indigenous peoples, or treaties were negotiated, as in the United States, lands that were originally communally held were often converted to individual ownership and acquired by outsiders. Reserves were often too small to support people, or valuable minerals rights were controlled by the state. In the last quarter of the 20th century, beginning in Australia, many indigenous groups formed political organizations and successfully regained more effective control over many of their territories. The

struggle for land rights and economic independence has been the central issue for indigenous organizations.

RICH AND POOR IN THE GLOBAL SYSTEM, 1950–2000

The global-scale world system that was already in place by the beginning of the 20th century has gone through a further organizational transformation since the global crisis of war and depression from 1929 1945. This transformation involved the construction of an elite-designed framework of international governmental institutions that included the International Bank for Reconstruction and Development (the World Bank) and the International Monetary Fund (IMF), both products of the United Nations Monetary and Financial Conference held at Bretton Woods, New Hampshire in 1944, the United Nations (UN, 1945) organization itself, and the General Agreement on Tariffs and Trade (GATT, 1947), and the GATT s successor, the World Trade Organization (WTO, 1995) (*see* **GATT (General Agreement on Tariffs and Trade)**, Volume 4; **WTO (World Trade Organization)**, Volume 4). According to the preamble of the UN Charter, these new institutions were utopian visions designed to save succeeding generations from the scourge of war and for the promotion of the economic and social advancement of all peoples . The outcome of this transformation was a world system dominated by nance capital. By the end of the 20th century, money and the exchange of nancial instruments became more important than the movement and production of physical goods and services (Barnet and Cavanagh, 1994).

The modern world system is not a self-organized system operating under the invisible hand of a global free market; it changes in response to pressures from speci c interest groups with unequal power. Governments continue to exercise sovereign rights over territories and are the primary agents in international political society, but as we have seen, a centralized global government would be too costly to be effective. Various non-governmental organizations (NGOs) have become new actors in an emerging transnational civil society. New institutions are being designed by the interaction of NGOs and governments to deal with a multitude of speci c global issues (Young, 1997). The great transformation from industrial to nancial capital made commercial elites more in uential than political rulers. Those who bene ted the most from the global ow of nance capital at the end of the 20th century were an estimated six million individuals who each possessed investable assets of more than a million dollars. In 1997 the world s multimillionaires collectively controlled $17.4 trillion (Gemini Consulting 1998), which was more than half of global gross domestic product (GDP), yet they constituted less than 0.1% of global population. The world is now directed by economic decision-makers who are guided by the utopian ideals of laissez-faire capitalism that call for aggressive reductions in the regulatory power of governments, the removal of all trade barriers, and global management by negotiations among private, self-selected interest groups with very unequal power. These changes are designed to further accelerate global economic growth, on the assumption that growth will bene t everyone (the *trickle-down theory*). Unfortunately, in spite of the utopian objectives of the UN, the problems of scale and power increased after 1945. Although wealth continued to accumulate in the core of the world system, by the end of the 20th century much of the world was more impoverished and the global system was less sustainable than ever before.

Accurate assessment of human welfare during the postwar period is impossible because there are no accurate gures for worldwide wealth and income distribution. However, the World Bank estimated in 1990 that more than 1 billion people were living in absolute poverty with *per capita* incomes of under $370 dollars. Three billion people lived in the 45 low income countries, with *per capita* incomes under $660. The average infant mortality rate in these poor countries was 8-fold greater than in the 24 high income countries where 812 million people lived. Raising the poverty line, and allowing for lognormal income inequality within each country suggests that in 1965 there were 2.8 billion (85%) poor and 50 million (1%) rich in the world (Bodley, 1999b, 2000). By 1997 there were more than 4 billion (70%) poor people, and 310 million (5%) rich people. Thus, economic growth reduced the *proportion* who were poor, and slightly increased the proportion who were rich. However, the absolute number of poor increased dramatically, far outpacing the absolute increase in the rich. This suggests that if we accept this global standard of economic well-being, growth since 1965 impoverished far more people than it enriched.

This ampli cation of poverty during the 20th century duplicates the similar impoverishment process during the earlier transition from feudalism to mercantile capitalism, and again during the transition from mercantile to industrial capitalism in Europe and America. Both of these great transitions did little to change the pre-existing base of social inequality, but because of scale effects, they also produced tremendous increases in wealth and power at the top of the social hierarchy. At the same time, given high levels of inequality, even small improvements in the material level for people at the bottom of the hierarchy, in combination with new economic conditions, created new incentives for people to produce larger families. These changes created an unprecedented era of population growth. However, inequality, not population growth, was the primary cause of poverty. Amartya Sen (1981) pointed out that people died in modern famines not because of food shortages, but because inequality denied them access to income and productive resources.

One of the guiding justifications for the globalization of capital-intensive agriculture was that it was the only way to feed the world. However, malnutrition has proven to be an intractable social problem, rather than a technological problem. The Food and Agriculture Organization of the UN (FAO, 1999) concluded that there were more hungry people in the world in 1997 (830 million) than the entire population of the world (720 million) in 1750! Many under-nourished people were marginalized by more competitive capital-intensive enterprises, or by degraded environments. The chaotic transition from state socialism to new forms of capitalism in Eastern Europe produced 26 million undernourished people by 1997. When the concept of malnutrition was broadened to include vitamin and mineral deficiency, and over-consumption, as well as calorie and protein deficiency, it was estimated that three billion people, half of the world's population, were malnourished by the year 2000 (Gardner and Halweil, 2000). This is equal to the entire population of the world in 1960! Forty years of economic growth may have actually magnified malnutrition.

The power-concentrating effects of economic growth can be illustrated in Washington State, in the Pacific Northwest of the United States, where between 1960 and 1995 both population and *per capita* personal income doubled (in constant dollars). Washington benefited enormously from global trade in aircraft, information technology, and agricultural commodities, but these benefits were very narrowly distributed and economic power became highly concentrated. A close look at the distribution of property ownership showed that as individual urban places grew larger, home ownership rates declined, and the holdings of the largest property owners steadily increased (Bodley, 1999b). Villages with populations under 2500 people displayed the most equitable distribution of property assets and business revenues. Corporate business and remotely owned businesses became increasingly important in larger urban places, and larger businesses captured a progressively disproportionate share of total sales volumes. In cities with populations of 100 000 or more, the top 3% of businesses received more than 80% of business revenues. As wealth became more concentrated, the elite created exclusive urban places, where ultra high property values formed barriers against non-elites. By 1998 individual social power was becoming grotesquely distorted by the phenomenal growth in global information technology corporations. The collective net worth of Washington's billionaires was already nearly twice the value of all deposits in all the commercial banks in the state. Measuring their income as increase in net worth, in 1998 the income of Washington State's 11 billionaires was already 28% of the total personal income in the state, and was growing at 35% a year. Greater Seattle's economic elite occupied the same power position in the world as London's economic elite in 1600, or Boston's elite in 1850, and they too were personally transforming the world by their decision-making. Disproportionate power of this magnitude is clearly the product of scale, and it has profound implications for the future development of humanity.

The global system seems to be approaching, and may have already exceeded, a scale threshold at which elite-directed, capital intensive commercial development becomes unsustainable, even assuming further technological miracles. It would be reasonable to predict that the power-concentrating potential of globalization under financial capitalism will soon be restrained, and perhaps reversed as new democratic political movements in the currently most developed centers of the global system press an effective argument for economic democratization. As more and more people feel disempowered by global markets and giant, monopolistic corporations, there will be strong incentives for political and economic decentralization. This decentralization of power may be accelerated by persistent political and economic collapses in impoverished peripheral areas that will force radical rethinking of the international development process. Economic democracy will require a restoration of economic power to households, small-scale producers, local and regional markets. This change could be supported by new local currency systems, and stronger local governments, even as large businesses and national governments give up power. Just as disenfranchised 17th century Englishmen developed the first modern political party, the Levellers, that helped replace the powerful concept of the divine right of kings with the radical new concept that political power originated with the people, today people can form new political parties and take back the power they have granted to giant commercial corporations. It will become increasingly obvious that information technologies make it feasible for the majority to use less capital-intensive production and distribution systems. That will strengthen local communities, and produce a very high quality of life for most people. The economic efficiencies of small scale will be quickly realized as local production and distribution systems replace the costly long-distance movement of materials, so that overly rewarded corporate elites and great private investors become unnecessary. The idea of economic globalization driven by all-powerful commercial corporations and great concentrations of finance capital will be a quaint anachronism, like the idea of divine kings.

REFERENCES

Algaze, G (1989) The Uruk expansion: cross-cultural exchange in early Mesopotamian civilization, *Curr. Anthropol.*, 30, 571–608.

Barnet, R J and Cavanagh, J (1994) *Global Dreams: Imperial Corporations and the New World Order*, Simon and Schuster, New York.

Bateman, J (1883) *The Great Landowners of Great Britain and Ireland*, 4th edition, Harrison, London.

Beard, C A and Beard, M R (1934) *The Rise of American Civilization*, Macmillan, New York.

Berry, C J (1994) *The Idea of Luxury: A Conceptual and Historical Investigation*, Cambridge University Press, Cambridge.

Berry, W (1977) *The Unsettling of America: Culture and Agriculture*, Sierra Club Books, San Francisco, CA.

Bodley, J H (1999a) *Victims of Progress*, 4th edition, Mayfield, Mountain View, CA.

Bodley, J H (1999b) Socioeconomic Growth, Culture Scale, and Household Well-Being: A Test of the Power-Elite Hypothesis, *Curr. Anthropol.*, 40(5), 595–620.

Bodley, J H (2000) *Cultural Anthropology: Tribes, States, and the Global System*, 3rd edition, Mayfield, Mountain View, CA.

Booth, C (1892–1903) *Life and Labour of the People in London*, Macmillan, London.

Boyd, R and Richerson, P J (1985) *Culture and the Evolutionary Process*, The University of Chicago Press, Chicago, IL.

Brandeis, L D (1914) *Other People's Money and How the Bankers Use It*, Frederick A Stokes, New York.

Braudel, F (1977) *Afterthoughts on Material Civilization and Capitalism*, The Johns Hopkins University Press, Baltimore, MD.

Braudel, F (1982) *The Wheels of Commerce. Civilization and Capitalism 15th- 18th Century*, Harper and Row, NY.

Braudel, F (1992) *Civilisation Matérielle, Économie et Capitalisme. Civilization and Capitalism, 15th- 18th century*, University of California Press, Berkeley, CA.

Califano, D (1998) The Battle for Assets, *Worth*, 7(11), 106.

Carneiro, R L (1978) Political Expansion as an Expression of the Principle of Competitive Exclusion, in *Origins of the State: the Anthropology of Political Evolution*, eds R Cohen and E R Service, Institute for the Study of Human Issues, Philadelphia, PA, 205–223.

Chandler, Jr, A D (1977) *The Visible Hand: The Managerial Revolution in American Business*, The Belknap Press of Harvard University Press, Cambridge, MA.

Chandler, T (1987) *Four Thousand Years of Urban Growth*, St. David's University Press, Lewiston, New York.

Clark, G (1936) *The Balance Sheets of Imperialism: Facts and Figures on Colonies*, Columbia University Press, New York.

DeForest, R W and Veiller, L, eds (1903) *The Tenement House Problem: Including the Report of the New York State Tenement House Commission of 1900*, Macmillan, New York.

Domhoff, G W (1996) *State Autonomy or Class Dominance? Case Studies on Policy Making in America*, Aldine de Gruyter, New York.

Farrell, B G (1993) *Elite Families: Class and Power in Nineteenth Century Boston*, State University of New York, Albany, NY.

FAO, Food and Agriculture Organization of the United Nations (1999), *The State of Food Insecurity in the World*, Rome.

Fletcher, R (1995) *The Limits of Settlement Growth: A Theoretical Outline*, Cambridge University Press, Cambridge.

Frank, A G (1990) A Theoretical Introduction to 5000 Years of World History, *Review*, 2, 1–28.

Frank, A G (1991) A Plea for World System History, *J. World Hist.*, 2, 1–28.

Friedman, L M (1964) The Dynastic Trust, *Yale Law J.*, 73, 547–592.

Gardner, G and Halweil, B (2000) Nourishing the Underfed and Overfed, in *State of the World 2000: A Worldwatch Institute Report on Progress Toward a Sustainable Society*, eds L R Brown, C Flavin, H French, and L Starke, W W Norton, Worldwatch Institute, New York, 59–78.

Gemini Consulting (1998) *World Wealth Report 1998*, Gemini Consulting, New York.

Gills, B K and Frank, A G (1991) 5000 Years of World System History: the Cumulation of Accumulation in *Core/periphery Relations in Precapitalist Worlds*, eds C Chase-Dunn and T D Hall, Westview Press, Boulder, CO.

Gras, N S B (1915) *The Evolution of the English Corn Market: From the Twelfth to the Eighteenth Century*, Harvard University Press, Cambridge.

Hall, P D (1982) *The Organization of American Culture, 1700- 1900: Private Institutions, Elites, and the Origins of American Nationality*, New York University Press, New York.

Hoberman, L S (1991) *Mexico's Merchant Elite, 1590- 1660*, Duke University Press, Durham.

House of Commons (1837) Report from the Select Committee on Aborigines (British Settlements) *Imperial Blue Book*, British Parliamentary Papers, VII, 425.

King, G (1936) *Two Tracts: (a) Natural and Political Observations and Conclusions upon the State and condition of England, (b) Of the Naval Trade of England A° 1688 and the National Pro t then arising thereby*, ed G E Barnett, The John Hopkins Press, Baltimore, MD.

Krebs, A V (1992) *The Corporate Reapers: The Book of Agribusiness*, Essential Books, Washington, DC.

Marcus, G E and Hall, P D (1992) *Lives in Trust: The Fortunes of Dynastic Families in Late Twentieth Century America*, Westview Press, Boulder, CO.

McConnell, G (1953) *The Decline of Agrarian Democracy*, University of California Press, Berkeley, CA.

McCracken, G (1988) *Culture and Consumption: New Approaches to the Symbolic Character of Consumer Goods and Activities*, Indiana University Press, IN.

McKendrick, N (1982) The Consumer Revolution of Eighteenth-century England in *The Birth of a Consumer Society: The Commercialization of Eighteenth-Century England*, eds N McKendrick, J Brewer, and J H Plumb, Indiana University Press, Bloomington, IN, 9–33.

McKendrick, N, Brewer, J, and Plumb, J H (1982) *The Birth of a Consumer Society: The Commercialization of Eighteenth-Century England*, Indiana University Press, Bloomington, IN.

McMahon, T A and Bonner, J T (1983) *On Size and Life*, Scientific American Library, New York.

Miller, D (1987) *Material Culture and Mass Consumption*, Basil Blackwell, Oxford.

Naroll, R (1967) Imperial Cycles and World Order, *Peace Res. Soc.*, 7, 83–105.

Pessen, E (1973) *Riches, Class, and Power Before the Civil War*, Heath, Lexington, MA.

Rabb, T K (1967) *Enterprise and Empire: Merchant and Gentry Investment in the Expansion of England, 1575- 1630*, Harvard University Press, Cambridge, MA.

Roy, W G (1983) Interlocking Directorates and the Corporate Revolution, *Soc. Sci. Hist.*, 7(2), 143–164.
Sanderson, S K (1990) *Social Evolutionism: A Critical History*, Basic Blackwell, Oxford.
Sanderson, S K (1999) *Social Transformations: A General Theory of Historical Development*, Rowan and Littlefield, Lanham, MD.
Schumacher, E F (1973) *Small is Beautiful: Economics as if People Mattered*, Harper and Row, New York.
Sen, A (1981) *Poverty and Famines: An Essay on Entitlement and Deprivation*, Clarendon Press, Oxford.
Shammas, C (1990) *The Pre-industrial Consumer in England and America*, Oxford University Press, Oxford.
Shammas, C (1993) Changes in English and Anglo-American consumption from 1550 to 1800 in *Consumption and the World of Goods*, eds J Brewer and R Porter, Routledge, London, 177–205.
Smith, A (1776) *An Inquiry into the Nature and Causes of the Wealth of Nations*, Vol. 1, Strahan and Cadell, London.
Snooks, G D (1993) *Economics Without Time: A Science Blind to the Forces of Historical Change*, Macmillan, London.
Snooks, G D (1994) New Perspectives on the Industrial Revolution in *Was the Industrial Revolution Necessary?* ed G Snooks, Routledge, London, 1–26.
Snooks, G D (1994) *Portrait of the Family within the Total Economy: A Study in Longrun Dynamics, Australia 1788- 1990*, Cambridge University Press, Cambridge.
Snooks, G D (1996) *The Dynamic Society: Exploring the Sources of Global change*, Routledge, London.
Snooks, G D (1997) *The Ephemeral Civilization: Exploding the Myth of Social Evolution*, Routledge, London.
Snooks, G D (1998) *The Laws of History*, Routledge, London.
Soltow, L (1975) *Men and Wealth in the United States 1850- 1870*, Yale University Press, New Haven, CT, 112–113.
Stone, L (1965) *The Crisis of the Aristocracy 1558- 1641*, Oxford University Press, London.
Stone, L (1973) *Family and Fortune: Studies in Aristocratic Finance in the Sixteenth and Seventeenth Centuries*, Oxford University Press, Oxford.
Taagepera, R (1978a) Size and Duration of Empires: Systematics of Size, *Soc. Sci. Res.*, 7, 108–127.
Taagepera, R (1978b) Size and Duration of Empires: Growth-Decline Curves, 3000 to 600 B.C., *Soc. Sci. Res.*, 7, 180–196.
Taagepera, R (1997) Expansion and Contraction Patterns of Large Polities: Context for Russia, *Int. Stud Q.*, 41, 475–504.
United States Department of Commerce, Bureau of the Census (1996) *Statistical Abstract of the United States 1996- 1997*, 116th edition, Government Printing Office, Washington, DC.
Wallerstein, I (1974) *The Modern World - System: Capitalist Agriculture and the Origins of the European World-Economy in the Sixteenth Century*, Academic Press, New York.
Weber, M (1930) *The Protestant Ethic and the Spirit of Capitalism*, George Allen and Unwin, London.
Wilkinson, D (1992) Cities, Civilizations, and Oikumenes: I, *Comp. Civ. Rev.*, 27, 51–87.
Wilkinson, D (1993) Cities, Civilizations, and Oikumenes: II, *Comp. Civ. Rev.*, 28, 41–72.
Wills, Jr, J E (1993) European Consumption and Asian Production in the Seventeenth and Eighteenth Centuries, in *Consumption and the World of Goods*, eds J Brewer and R Porter, Routledge, London, 133–147.
World Bank (1995) *World Development Report*, Oxford University Press, New York.
Young, O R (1997) *Global Governance: Drawing Insights from the Environmental Experience*, MIT Press, Cambridge, MA.

FURTHER READING

Durham and London, The Tudor Sumptuary Laws, *Engl. Hist. Rev.*, 30(119), 433–449.
Robbins, R H (1999) *Global Problems and the Culture of Capitalism*, Allyn and Bacon, Boston, MA.

Technological Society and its Relation to Global Environmental Change

ESTER BOSERUP (Deceased)[†]

[†] *Independent Consultant*

The influence of technology on environment, and vice versa, is a relatively new subject in scientific discussion, but it has been crucial in human history. Historians divide human history in technological stages: the first stage when small human groups subsisted on wild food, then the invention of agriculture, then urbanization, industry, and now an ongoing change from the industrial society to the electronic age. Since the industrial revolution in Western Europe a couple of centuries ago, there have been more and more examples of negative effects of technological change on the environment, fauna and flora, and also on human life.

ADAM SMITH AND MALTHUS

The basic relationship between technology and environment is strongly influenced by a third factor: population size, as was pointed out two centuries ago, by the classical British economists Adam Smith and Malthus. They agreed on the crucial role of population change for human development, but while Smith thought population growth was a positive factor, Malthus thought it had negative effects. Smith's view was the generally accepted one: child mortality was very high and the family wanted to have many children to ensure that some would survive. Rulers also favored a larger population. Many inhabitants meant many productive workers and many soldiers to win wars and conquer additional territory. But because he observed the beginning of the industrial revolution, Smith added that population growth would not only increase the number of workers, but also their production per head, because it would make further specialization of labor possible.

Malthus' essay was a great sensation. He applied an ecological view, arguing that population growth reduces average incomes, because he considered the globe with its natural resources a constant, so population growth would reduce natural resources per head and increase mortality due to insufficient food production. The carrying capacity of the globe could not be exceeded.

In retrospect, we can see that they both had a point. Population growth can have both positive and negative results, depending partly on the relative strength of the positive and negative factors, but also on the strength of a number of other factors, which economists and others, who use simplified models, fail to take into account – first of all social and cultural factors. Therefore, we have neo-Malthusians who put great emphasis on the negative results of technological changes, which pollute air, fresh water, and soil, and threaten human life and health and the survival of fauna and flora. And we have anti-Malthusians, including many neoliberalists, who downplay the negative effects of modern technology on the environment, while focusing on the positive effects on average incomes. What we lack is an accepted theory of how all the relevant factors play together in the complicated process of economic and human development (Boserup, 1996).

Malthus overlooked that carrying capacity is not a constant because there can be substitution between different resources, be it labor or capital investment as replacement for land or one metal for another, and the type of substitution which is possible depends on the particular resource endowment and other conditions for resource transfer or invention in the region under investigation. In other words, it is necessary to use a historical approach to answer the question of effects of major population change on technology and environment (technology is defined as including both material and method).

Specialists from many disciplines have studied the relationship between changes in technology and in environment in addition to ecologists and natural scientists. One broad group consists of economic, medical, cultural, and general historians, another of archaeologists, anthropologists, and demographers. These groups have focused on historical changes in technology and environment in different parts of the world at different periods. But the problem has also been studied in a different way, concentrating on recent changes in the developing countries. Such studies have been

made by researchers from many disciplines in universities and other institutions that focus on aid to developing countries. In this essay both methods have been combined, for instance by comparing the process of ancient urbanization in Asia and the Middle East and its relation to environmental change with the much later process of urbanization and industrialization in Western Europe under different environmental conditions.

THE DYNAMICS OF LAND USE

Ten thousand years ago, or earlier, some human groups changed from hunting and food gathering in forested land to shifting cultivation, felling the natural vegetation by fire, sowing or planting seeds or roots, and after a few years when yields declined, repeating the operation in another spot leaving the preceding one fallow for two decades or more, until soil fertility had been fully regained. However, as their population continued to increase, the cultivators had to shorten the fallow and only bushes had time to invade the field. With further growth of population, recultivation had to be made when only grasses had invaded the land. Forest fallow, the first step of shifting cultivation, is still in use in thinly populated parts of many developing countries in all continents, for instance in the Indian Himalayas and in tropical Africa (*see* **Shifting Cultivation and Land Degradation**, Volume 3).

If this type of gradual intensification of land use is allowed to go on, without the cultivator family carrying green manure or dung to the field to preserve soil fertility, shifting cultivation is *unsustainable* and the land becomes sterile, but whether the process is *renewable* depends on the climate. With too little rain the land may become a desert, but with enough rain the original vegetation may come back, and the whole process may be repeated. However, if the cultivators have preserved soil fertility by manuring the fields with green manure or dung, and their populations continue to increase, they could go on from bush fallow, to more and more intensive short fallow systems, to annual cropping and, if the climate allowed it, to multicropping. This was, in fact, the way most of world agriculture developed before the industrial revolution created substitutes for the natural manure, and the human labor needed to apply it. In other words, the intensive systems of agriculture, which are used today in densely populated countries that still use traditional systems of agriculture, are repetitions of systems applied millennia ago. Land use is not static, but can be, and often was a dynamic adaptation to changes in population size and environment, and the process seems to have been similar in many parts of the globe (Boserup, 1965).

The historical process replaced less land per head by more labor for current operations and investment in land improvement, including terracing of mountain slopes and leveling of land, irrigation and draining. Gradually, over centuries or millennia, the environment changed. The landscape was transformed from a large forest with a few hunter-gatherer families to one of villages surrounded by arable fields, grassland for domestic animals and some forest for gathering of fertilizing matter, timber and fuel-wood.

These processes motivated the change of tools by transfer of technology from more advanced neighbors or invention by trial and error and speculation, from digging sticks to hoes and to ploughs drawn by oxen or horses. Technological change is seldom exogenous, i.e., accidental or made by a so-called genius, but endogenous changes may be made if the motivation becomes strong, due to population change, or some other factor, often a lost war in the case of weapons. In agriculture, increasing local population with need for additional food production promoted an endogenous process of technological change, which augmented the carrying capacity of the land by using the increased labor force for current operations and labor investment.

With intensification of agriculture more use was made of female and child labor. Already among hunter-gatherers, women and children took part in food gathering, but with shorter fallow the share of work performed by women and children increased until the system required very hard labor by the whole family all year round, as was documented by Lossing Buck in his research on China in the 1930s (Buck, 1937).

The increasing use of family labor had, sometimes at least, a profound influence not only on family life, but also on culture, and even religion. Forest people believe in forest spirits and use sorcerers and sacrifices to ask for protection. Sorcerers seem to be the first professional group in human history; and in the Indian caste system the priests, the Brahmans, constitute the highest caste. But many agriculturists, for instance in Southern India, sacrifice to local female fertility goddesses to obtain a good harvest, while pastoralists sacrifice to male gods to obtain good results from their flocks of animals. The later disappearance of goddesses in most of the world seems related to the reduction of the role of women in the agricultural economy when the plough came into use, and to her lower status compared to that of men.

The patriarchal system, where the male head of household can dispose of the labor power of other family members, led to intensification of food production through labor-intensive technologies. The right of adult men to dispose of the labor of other family members does even today encourage a pro-natalist attitude, polygamy, and keeping children away from school, even when there is free access to schools in the neighborhood.

The increasing intensity of land use changed not only the use of family labor, but also land tenure. As long as the local population is small, there may be free access to forest, grazing and cultivable land which has regained fertility by

fallow; with a larger population, only the local population has full rights to land, while foreigners must pay crop-shares or work as slaves, tenants, or wage workers, with or without the right to graze animals in village grazing areas, and gather fuel and food in the forests surrounding the village.

Until the middle of the 20th century diseases, including epidemics, to which the local population had not developed resistance, had kept long-term rates of population growth very low in developing countries, and – similar to Europe in earlier centuries – there had been periodic large declines of population, caused by epidemics and wars, often followed by forced migration and hunger. But during and after the Second World War a scientific effort was made to combat disease by means of modern technology also in many non-industrialized countries and colonies – the result being a population explosion, due not least to a large reduction of child mortality.

Because of the general acceptance of Malthusian theory in the industrialized countries, this acceleration of population growth created widespread fear of future world famine. So, similar to the transfer of Western medicine to developing countries, it was decided to start research aimed at transferring other modern Western technologies to densely populated Asian countries, where the problem was expected to become worst. Two types of research were started: one was aimed at improving the technology of contraception, the other was agricultural research aimed at raising crop yields by use of Western techniques, such as the use of chemical fertilizers and development of better seeds – the so called *Green Revolution* (*see* Green Revolution, Volume 3).

Green Revolution technologies can only be successfully introduced in areas with a reasonably good infrastructure: good transport facilities for both inputs and outputs, storage facilities, access to agricultural advisers and other extension services, and perhaps local research stations. All this is unavailable in regions of subsistence production and is not likely to be provided in the least developed countries, or in subsistence regions of large countries with much better prospects for agricultural development in more advanced and more densely populated areas. Therefore, the Green Revolution was promoted in densely populated Asian countries, simultaneously with improvement of the rural infrastructure in those areas, if necessary. The research was highly successful in stimulating production. However, both in the Asian areas for which it was designed, and in the Western industrialized donor countries, production increased so rapidly that the feared appearance of world shortages became one of embarrassing food surpluses. It also contributed to these unexpected surpluses, in that the rapidly increasing rural labor force in areas that continued to use traditional labor intensive technologies made it possible either to expand the area under cultivation or to increase cropping frequency.

When food surpluses increased in the industrialized western countries, their governments decided not to reduce government support to agriculture, but instead to give food aid to developing countries, and to lower export prices for food in order to be more competitive. Food aid and subsidies to food exports were granted not only to densely populated, but also to sparsely populated developing countries. But in many cases, a small population is not due to especially unfavorable natural conditions for food production, but it is the small home market, and lack of access to export markets, which force rural populations to remain subsistence producers, or to migrate to mines, towns or foreign owned plantations.

Many governments in developing countries welcomed the chance to supply their increasingly populated towns with cheap food, or to sell the food with a large profit. In the first case the effect was a discouragement to local agriculture, and food exports from the industrialized to the developing countries became very large.

In Asia, the Green Revolution technology, the increasing rural labor force, and the large food imports, combined to satisfy the effective demand for food without the need for investments in expansion of the cultivated areas. But it would be wrong to interpret this as a sign of lack of reserves of uncultivated, but cultivable land. If it should become impossible to increase crop yields much further, the Green Revolution technology can be modified with more emphasis on land improvement by investment in hitherto uncultivated land and on further cropping intensity. The possibilities for further multiplication of food production are often overlooked because hunger and malnutrition among the poor are wrongly interpreted as lack of the capacity of world agriculture to produce sufficient food. Evans rightly talks about well-intentioned scientists who seek technological solutions to problems of social and economic inequity (Evans, 1998).

Like other major technological changes, the Green Revolution had profound social and cultural effects. The income differences between areas with and without a Green Revolution became very large, and in the former there were important changes in family and fertility. When chemical fertilizer, pesticides, and mechanized inputs reduce the demand for the type of operations that were hitherto performed by women and children, the motivation for the male head of the household to be polygamic, and have a large number of children, is reduced or eliminated. Therefore, modernization of agriculture is accompanied by reduction of polygamy and fertility, except where strong religious inner and outer resistance against birth control delays the process, perhaps for a long time. Also, when children are less needed for agricultural or other work, they are more likely to be sent to school, but in many cases to schools, which are hostile to cultural change. In all cases, however, the reduction of fertility and agricultural toil create

a very large improvement in the health and position of women.

These effects are visible by a comparison between Asian areas with Green Revolution, and sparsely populated African countries. In the former, fertility is declining not only in urban, but also in rural areas. By contrast, there is much less fertility decline in the sparsely populated African countries, and it seems usually to be limited to urban areas.

MOBILITY AND URBANIZATION

After food production, the creation of urban centers is usually considered the next important step in technological development. Nearly all ruins of ancient towns are found in Asia and the Near East, a region with a dry and warm climate different from the cold and rainy climate of Western Europe. Both their early towns and their large cities antedate those of Western Europe by millennia, although insufficiency of rain only allowed agriculture at river sites, where water was seeping in, or with irrigation. It seems that we have here one of the many examples where a handicap – in this case shortage of water – proved to be an advantage, because it stimulated invention, in this case of irrigation technology, both for food production and for transport.

As long as their population was small, the cultivators lived on the shores of the river, or farther away from it, if there was natural seasonal overflow of the river. When the population increased, the people helped the natural overflow to spread more by digging small terraces on the mountain slope beside the river; and leading the water along these, they spread it by removing a few stones from the edge of the terrace.

Another step in the development of irrigation was to dig canals in the plain below the mountains or hills from which the flood came, to spread the water more widely, thus increasing the area under cultivation. This was followed by a crucial technological change. Construction served to lift the water either by gates or dams, or by water-powered machines, i.e., shaduffs and waterwheels. In other words, the use of water power had been invented, and it had the additional advantage that it benefited both food production and transport of products and people, which now could move up-river even when unassisted by sails and wind power. So, also in these natural conditions, the increase of population motivated technological change.

Not only in recent technological development, but throughout human history, changes in mobility and communication have been of crucial importance. In early times, many human groups depended solely upon human legs and perhaps primitive boats for mobility, but many pastoral peoples had large mountable animals (horses, dromedars, camels) and were therefore much more mobile than were the hunter-gatherers and cultivators discussed above.

The climatically dry and semi-dry zones of Asia and North Africa have for millennia been inhabited by pastoral peoples, either because their habitats have always been dry for climatic reasons, or perhaps have become so by human maltreatment of fragile land. Pastoralists subsist mainly by the products of their animals, supplemented by some vegetable food. If the group increased in size, they might make over-intensive use of the natural grazings and spoil the land, but sometimes they would avoid it by spreading more and more widely, and in large semi-dry areas they might become nomads, moving with their animals and tents, if possible by mounting the animals.

The diet of pastoralists is richer in proteins than that of cultivators, and they are usually larger and stronger than cultivators, who subsist mainly on vegetable food. This, together with their much larger mobility, made them a formidable enemy for the cultivators in the river valleys. The nomads attacked their villages, enslaved their inhabitants, and made themselves rulers of the area, using the villagers to provide themselves and their families with vegetable food and domestic, sexual and sometimes military services. If a pastoral people came to rule a number of villages, their descendants might become hereditary nobility between the ruler and the subservient population of peasants.

The family system and other cultural patterns of pastoralists are different from those of cultivators. Large animals belong to men, and mounted animals are a status symbol for male pastoralists. One may ask if a similar type of stratification had emerged in Europe, where the horses became a symbol of the high rank of the hereditary nobility that went to war on horseback, held tournaments and horse races and monopolized the right to hunt on horseback in the forests.

Carneiro has suggested that it was a condition for early urbanization that the population lived in circumscribed areas, i.e., areas surrounded by deserts, steep mountains or seashores, which forced the inhabitants to intensify food production if they must feed an increasing population in that type of environment (Carneiro, 1998). If the circumscribed area was as large as in Mesopotamia and the large river-valleys in India and China, a sufficiently large and dense population could produce and transport enough food for creating large urban centers at an early date. Existence of large towns and cities depends on transport networks for food to the non-agricultural population living in the urban centers. Road transport was inefficient over long distances but a dense network of canals made supply to large towns possible – for instance in China. But later, in the railway age, the Chinese canals proved to be a handicap because of the many bridges that had to be built. Thus, from being a technological leader until around 1500 AD, China, after

the industrial revolution in Western Europe, turned into a non-industrialized developing country.

In river valleys of Asia and the Near East, nomadic peoples from the Central Asian mountain range managed to create huge empires by conquering many hundreds of villages belonging to many different peoples, and using the inhabitants as soldiers. Wars were the means to get soldiers for further conquest, and construction workers for fortifications and embellishment of towns.

There were also productive infrastructural investments: large canals for transport of supplies and persons, and large dams for irrigation and expansion of the network of canals. In old times, large dams were made by an army of manual workers, and this technique continued to be used in independent India. The Indian government wanted to provide work for both local people and poor immigrants from other parts of India. It was a striking contrast to the simultaneous road building by the Americans in Vietnam, who used ultra-modern American mechanized equipment operated by American personnel.

At the time of the old empires, there were violent fluctuations in population, and the cruelty of the empires, both in war and peace, has been assumed to be a major explanation of these fluctuations, but, as already mentioned, it seems likely that a major cause of the fluctuations was epidemic diseases. Of course slaughter and famines among the fugitives contributed to the catastrophe.

Today there are lorry transport routes all over India and AIDS has followed them. In Southern and Central Africa the epidemics threaten to be even worse than the old killers from earlier centuries. World population was not stationary before the demographic transition, as Malthus believed, but was slowly increasing with intermittent sharp declines, where epidemics are supposed to have been a larger killer than warfare. The means to avoid an epidemic was to flee to another area, thus spreading the epidemic further.

SUPPLY OF POWER FOR THE INDUSTRIAL REVOLUTION

Western Europe had little urbanization before the industrial revolution. Because of the abundant rain, irrigation and water power never attained the importance it had in Asia, but Western Europe imported the techniques of water-wheels, on its small rivers, for milling of cereals and other purposes. In the 11th century, England had more than 5000 watermills in operation.

By contrast, Western Europe with its long coastal lines from the Mediterranean to the Baltic Sea made more use of wind power for long-distance transport of products and persons than did Asia. With strong motivation to improve shipping technology, Western Europe succeeded in dominating the whole world in the 19th century, by means of its colonization policy in the other continents. This was a big advantage for a region, which was relatively poor in natural resources.

From an early age there were caravan routes inland between China, the Black Sea and the Russian rivers and lakes, and later trading routes along the Asian coast between East Asia and Europe. At the European end the trading routes became longer and longer, until European coastal settlements on the Mediterranean, the Atlantic coast, the North Sea, and the Baltic had become active participants in the Asian trade, with large government-owned or government-supported Asiatic Companies.

At the first stage of this development, the boats and small ships could only carry light goods, and the trade consisted of luxuries (silk, fine textiles, spices, precious stones and metals, processed or not) for the use of rulers and rich merchants in the areas served by the trade. But when improvements in shipbuilding made long distance transport of bulky and heavy goods economical, the Europeans expropriated deposits of ores, fertile agricultural land and forests behind their trading stations in other continents, and used them to produce raw materials or semi finished products for exports to Europe. When improvements in navigation made Atlantic crossing safer, the Europeans established themselves in America, produced tobacco, coffee, cotton and other crops in monoculture, partially with use of slaves bought or caught in America and Africa. The Atlantic crossing was followed by exploration of the coasts of other continents, and many of the areas where the trading stations were located ended by becoming formal colonies of European countries.

Because most of the products exported from the colonies were owned by European companies and produced by cheap dependent non-European labor, colonial trade was very unequal, with large imports to Western Europe, and small exports – of which a share was for consumption by the European civil and military personnel and businessmen who lived in the colonies. By continuing for centuries to deliver ores, timber and crops for a fraction of the real costs, the colonies actually financed a great part of the Western European industrial revolution. Moreover, as Needham has shown, a large share of what has been considered proof of Western European inventiveness, was in fact technology transfer from China, either directly or via the colonies (Needham, 1954). But technological change promoted inventiveness, and the imports of new techniques were followed by important inventions by European scientists and businessmen, including the steam-engine and the railway, which linked the European countries together physically and economically, while the steamship linked them better to their colonies in the East, South and West.

Because Western Europe made less use of water power than China, and only made use of coal much later, it eradicated much of its forests and imported timber from Sweden and Russia. Therefore, the Wooden Age lasted very

long in Europe, but gradually more and more industries shifted their fuel supplies from wood to coal or water power, though the iron industry still used charcoal to melt ore. As late as 1750, per head output of iron in Europe was lower than in the least developed countries today, and the military sector took most of the supplies. Iron production in Western Europe is a classical example of a crucially important demand-induced technological change, and for a time Western and Central Europe became technological leaders on a world scale, owing to their large deposits of coal and ore in locations that were favorably positioned to attract both raw materials and market-oriented industries (Boserup, 1981).

The Coal-Railway Age became a period of extreme centralization, both regionally and locally, with large rural–urban and east–west migrations in Europe. The industrial importance of Italy, which had neither coal nor ore, was reduced until the 20th century, when its deposits of gas were discovered and exploited.

Because of their access to raw materials produced by cheap colonial labor and expropriated land and mines, British and other Western European industrialists were very efficient exporters of industrial products and solidly in favor of free trade in the world market. But that did not prevent them from protecting their local manufacturing industries by means of high custom duties and later by quotas, because they had to pay much higher European wages, when the colonial products were processed in Europe. The result became the well-known pattern of colonial trade where the colonies delivered food and raw materials to the industrial countries, while Britain and other colonial powers delivered manufactured goods to the rest of the world.

The industrial revolution promoted investment in private and government infrastructure: roads from railway stations and harbors to customers and other ancillary urban or rural services. Moreover, military technologies were improved and new armament factories built. In the military sector, as in the civilian, there was a beginning of the coming change from reliance on large numbers of soldiers and unskilled labor to more reliance on equipment and skilled specialists. Labor shortage was no more a permanent feature – from now on it alternated with unemployment.

The industrial revolution did not only create the business cycle, it also created the serious problem of pollution of the environment. Although there were accidents, and sometimes salinity, the Asian type of reliance on water power for industry and transport was a non-polluting technology, but in Europe the change from wood to coal had serious consequences because of the smoke. Though the burning of wood also produced smoke, that of coal became much worse and promoted tuberculosis, especially in mining and industrial districts, and in urban areas where the households used coal for heating. Also the new railways produced smoke in the densely populated areas, which they served. Moreover, the rapid growth of both the mining and the processing industries resulted in concentrations of population in the most polluted areas, which were becoming both unpleasant and unhealthy. Because of the rapid urbanization the immigrants had to live either in old houses or in self-built huts, and the towns became surrounded by huge slums, which contributed to an increase in mortality and made room for employment of further immigrants.

The natural landscape was transformed by the new ugly construction of mines and factories. This gave birth to the first protest movement: the *romanticist movement* at the end of the 18th century, which regretted the damage made to landscape, fauna, and flora. The environmental movement was born of the industrial revolution.

Like earlier structural changes in the economy, the industrial revolution in Western Europe was accompanied and followed by major political, demographic, occupational, and cultural changes. Organized labor movements appeared, due to the frequent periods of unemployment with downward pressure on wages. The new industries helped to put downward pressure on male wages by offering employment to young unmarried women with the result that the marriage age increased, reducing the birth rate, but for the poorest the increasing recruitment of child-labor in industry provided an inducement to continue high fertility and the decay of family life of industrial workers. Moreover, as a result of the industrial revolution, a new middle class of technicians and other specialists appeared, both in private enterprise and in government, and became a more important part of the labor force and of society in general. Following the French Revolution, secularization spread among intellectuals and the feminist movement was born, giving rise to employment of middle class women.

In the 19th century, more and more industrializing countries introduced suffrage for men but, except in New Zealand, where women had to wait until the next century. The concept of human rights was until recently defined more narrowly than today, and the declaration of human rights by the French Revolution and the American Constitution covered only men. Not only were women excluded from the human race in the 19th century, but also colored natives in the colonies. Not only most Europeans in the colonies and at home, but also much of European science, agreed with the racist theory of European superiority in relation to other races, both blacks, reds, and yellows, and tried to explain racial difference by hereditary differences in, e.g., the shape of the skull. The embarrassing problem of technology transfer from Asia was avoided by classifying the ancient Greeks as Europeans and focusing higher education on the writings of Plato and Aristotle, forgetting about their oriental predecessors.

The attitudes of European missionaries in the colonies were of course different, but usually they considered and treated the natives as children. Many whites in the colonies

regarded only white people as real human beings, while members of other races were either considered sub-standard humans or not humans at all, a good excuse for killing or enslaving them in spite of the attitude of the European churches. In conformity with these racial theories, the Europeans arriving in Canada and Australia expropriated the whole territory declaring the land empty; these actions are today occupying the courts and generating public debates in both countries.

Before the 19th century the loyalty of European populations had usually been to their ruler or to their local region, whether independent or not, but from the 19th century onwards, it changed in many cases to strong feelings of attachment to the Nation, i.e., the area under the same government, which had become strengthened by the appearance of national railways, mining centers, large industries, and increasing national government bureaucracies, civil and military.

After the industrial revolution, Western European governments had made efforts to transform the European part of their possessions into national states, partly by propaganda and other support for what was labeled *national culture*, and partly by propaganda against or suppression of minority cultures. A strong means to obtain this was the increasing school attendance by compelling both teachers and pupils to use the national language only. The growth of armaments and other national industries and establishments, and the competition between the European countries in exports and acquisition of additional colonies, also contributed to a feeling of national unity. Not only were literature and art strongly influenced by the national movements, but also science.

DECOLONIZATION AND NEOLIBERALISM

The industrial revolution had important long-term effects on military technology and world politics. Having helped France and England to win two world wars, the United States became the uncontested technological world leader, a change that was related to the environmental differences between these areas. Railways had been of central importance in the European process of industrialization, as already mentioned. On the North American Atlantic coast they were built early, but not over the enormous stretches of near-empty land, in the interior of the continent. Transcontinental railways were built with French and Russian capital, using local wood instead of coal, but they were unprofitable, and the foreign investors lost their money before the railways could make both ends meet under American ownership.

Once more, however, a handicap became an advantage, because the poor profitability of many railways induced American business and scientists to develop better means of power and transport, that could serve thinly populated areas over very long distances better than the railway. The result became the American dominance in the automobile and aircraft industries, which derive power from oil, a product that was available within their own territory. Automobiles offer more flexible transport than the railway, and aircraft contributes to globalization. For a time their influence on the environment was positive, because they replaced the bulky and polluting coal by oil, but later both industries became serious polluters.

After the Second World War, the United States used its influence as a superpower to broker the liberation of the colonies of the European countries. The United States wanted this liberation, not only because of their own past history, but also because the colonial preferences in international trade discriminated against US exports. England and some other colonial powers accepted a peaceful liberation of their colonies, because many had become public burdens due to increasing expenditure on military and administration, and because liberation did not imply any change in ownership of private enterprises. To enable the former colonies to take over the costs of military and civil administration and other public infrastructure, the United Nations and some national governments created organizations of international aid to developing countries, and the World Bank provided long term finance for large scale investments in infrastructure and other major projects.

An international boom followed the process of liberation of the colonies. The colonial powers had wanted to keep the colonies as suppliers of food and raw materials, and as markets for industrial products, but after independence, many former colonies wanted to industrialize, and change their pattern of trade. Many obtained grants and loans from donors to cover the expected deficit. The American business community used the improved access to direct trade with former European colonies to create large multinational companies, followed by an increasing number of other countries.

Multinational companies apply the most advanced technologies in their production of goods and services for the world market. Usually, they have their headquarters and top managers in the United States, Europe or Japan, with dozens or hundreds of enterprises spread over several continents. Unskilled labor is recruited where it is cheap, and qualified labor, technicians and computer services where wages are low, but qualifications high – for instance in India.

The liberation of the colonies has allowed many developing countries to increase employment and incomes by becoming seats for one or more of the enterprises belonging to a multinational company or a sub-deliverer of materials or services. The multinationals are efficient producers in the world market, because they are cost-reducing for their customers, but often they are even worse despoilers of the environment than local industries, because they can choose

to operate in the countries where they pay the least for protection of land and people, as was the case in Bhopal in India in 1984, where more than 5000 members of the local community were killed by a chemical leakage.

Modern technology has saved millions of human beings, but also destroyed millions. Many modern inventions threaten human health or survival, while others contribute to the deterioration of the physical and biological environment and some do both. More often than not, enterprises, government bodies, and municipalities dispose of their waste, polluted by dangerous metals and chemicals, into rivers and lakes, or ugly mine tailings, reducing the access to fresh water and making the Green Revolution poisonous.

Heavy agricultural machinery may despoil large areas of fragile soil. The oceans are polluted by in ow from poisoned rivers, cleaning of ships in open seas and many other polluting practices. In large cities, air pollution by motor traf c and domestic cooking and heating has become as unpleasant and dangerous to health as burning of coal was earlier. And nuclear energy, once considered a clean replacement for coal, has proved to be so controversial that calls for its abandonment are becoming strong and widespread. However, now it seems that better alternatives are on the way: more use of solar energy, perhaps combined with liquid hydrogen.

Pollution of the water in rivers, lakes, and oceans can be avoided or repaired, and municipalities and governments often try to achieve this, by charging payments for use of clean water and for misuse of fresh water, but these payments are usually so low that it is cheaper for the consumers to misuse the resources. Not only too low penalties, but also too high premiums can cause disasters. Because of the high premiums paid in the Soviet Union to producers who exceeded their production plan, collective farms in Uzbekistan neglected to fallow their irrigated elds, bringing desalinization.

To avoid or limit the misuse of natural resources, the producers should pay the costs of avoiding misuse, which is in most cases a fraction of the costs of repair, if that is at all possible. In many cases there are regulations which should induce or force producers to delay using new technologies, until suf cient scienti c investigations of the effects have been made, either by the producer or by government, but this is unlikely to be the case in countries that are eager to house subsidiaries of multinational companies.

The problem of who should pay, the multinational or the country where one or more of its enterprises are located, concerns not only environmental costs, but also costs of public infrastructure, for instance roads, harbors, airports, and unemployment insurance. However, neoliberal theory undervalues the importance of infrastructure in modern economies.

We meet the problem who should pay? whether we discuss environmental problems or the welfare state. Many American economists have an easy answer: the neoclassical school tacitly assumes that infrastructure costs are negligible, so what is right for a business enterprise is also right for the economy. The economists are right in pointing out that, if the local, national or multinational company is charged *more* than its indirect costs, the price structure cannot be the correct guide for the market and the investments, but they conveniently forget that this is equally true if the company is charged *less* than the true costs, including the replacement costs of necessary infrastructure, be it private or public. For the same reason, it distorts the structure of world production, when virtually all governments in the industrialized countries, whether they are members of World Trade Organization (WTO) or not, give subsidies to non-competitive national producers and pose restrictions on imports from effective foreign producers of goods and services.

The colonies of European countries had been handicapped in industrialization by the colonial trade system, but after they had become independent, many of them used the large grants of foreign aid in the rst period of independence to start ambitious programs of industrialization by investment in the types of physical and human infrastructure, which are needed for industrialization. Large infrastructures may take decades to complete and even longer before countries can begin paying back their loans; so many of the new projects proved to be more costly and took much longer to construct than foreseen, while in the meantime the 30 good years (1950 1980) were over. The neoliberal governments in the United States and England had reduced foreign aid, part of which was replaced by private loans or credits from international organizations especially the World Bank.

As a result, many of the former colonies became heavily indebted and had to cut down on their investment plans, sometimes by arresting half nished developments. Some nanciers, getting nervous, tended to make the loans shorter and shorter, until a crisis broke out in one developing country after another. Production and employment were sharply reduced, and the large informal sector broke down, because its customers lost their incomes. The informal sector in developing countries consists of small enterprises, often family industries that produce traditionally-made goods and services for worker families and the lower-middle class, who prefer the cheap prices in this sector to better quality and design in the modern industrial and service sector. Because it uses traditional technology, the informal sector provides a large share of the employment and self-employment in developing countries.

When the sharply reduced activities in both the modern and the informal sector made it impossible for large numbers of the urban population, and some of the rural too, to subsist, many, for instance in Africa, chose to return

to subsistence production in the villages they had left for the town, but this was not possible, or was less attractive, for earlier rural-urban migrants from areas where the Green Revolution had made access to land more difficult or impossible. Many others chose to migrate to the United States, Canada or Europe, many of them to larger towns and cities, which more than ever became melting pots for people of different races, nationalities, languages, beliefs, and past experiences.

In many of the developed western countries, to which not only economic, but also increasing numbers of political refugees turned in desperation, immigrants from less developed countries became more and more unwelcome, or were refused access, because of increasing resistance to immigration. This was so not only because the labor market was deteriorating, but also because of increasing resistance against having to live together with foreigners , especially if they belonged to a different national or religious group and had contradictory ideas on the role of women in the family.

To adapt to the international crisis, which created overcapacities in many national and multinational companies, more and more of these have chosen to merge in order to reduce their costs – especially by large reductions of personnel, both in developed and developing countries. For example, oil companies that had competed by constructing service stations next to each other discovered that the true costs of competition could be high. However, by cutting down on these expenses they added to the unemployment crisis in many of the countries where they operated.

The mergers were not a means to avoid or reduce competition, but were aimed at sharpening the competition between the remaining larger and stronger multinationals and between them and national, private or government owned companies. Instead of aiming at a strengthening of the world economy, by cooperation both within the private sectors and between these and the government sectors, national and international, the neoliberalists continued to destabilize both the national economies and the world economy by their continuing efforts to destroy their competitors, be they national or international, instead of focusing on the common interest in stable growth of the world economy.

The present prolonged crisis in most of the world markets is a repetition of the situation in the 1930s, when both mass consumption and investments continued to decline, until the new Keynesian policies increased mass consumption and promoted profitable investments. But today, capital-exporting countries prolong the crisis by their sharp cuts in development-aid in a period where sharply increasing income inequalities also reduce mass consumption, thus reducing the profitability of many investments.

GLOBALIZATION AND THE MODERN CASTE SYSTEM

From ancient times, large technological changes have been accompanied by and followed by fundamental changes in the economic and political organization of society, in religion, and secular culture. Adam Smith identified the process as more and more specialization of labor, while Marx interpreted it as a gradual change from an original one-class society, which after a long period of war between two classes, a rich and a poor one, ended with a lasting victory for the lower class, that had become the industrial working class.

But Marx's interpretation of human history was wrong. Neither before nor after the time of Marx, was human history a war between two classes; it is much better interpreted in terms of the Indian multicaste system, if this is modernized by adding the scientific community with its increasing claim to be accepted as the first caste. The ancient Indian system had three upper castes, the priest, the secular power (army), and the trader, with the rest of the Hindu population as sudhras , the fourth caste, and the adherents of other religions as non-caste foreigners .

If we interpret this caste system as a stylized description of world history, the first caste: the priest, replaces the sorcerer as the highest authority, after the shift from the gatherer to the producer of food, as mentioned earlier. The next step is a contest of power between the first and the second caste, the priest and the king with his army, a fight for power that, for instance in Europe, lasted many centuries. It became more complicated when the third caste, the trader (later: financier or capitalist) appeared, with urbanization and monetization, visiting villages as foreign tradesmen and lending money to the higher castes. The trader was often himself a foreigner; in Asia, Chinese or Indian; in Europe, Jewish. Sometimes he was massacred by caste members. In this period of history the European church was split into Protestants and Catholics, as explained by Max Weber.

The next step in the contest between the castes was *democratization*. The sudhras became the electorate, thus becoming legal contestants for power, creating alliances (nearly everywhere precarious) with one or two of the higher castes. Finally, we have the struggle between the four castes and the non-caste foreign group, be that a minority or a subjugated majority, as was the case in the colonies, where discriminatory treatment of castes and religions was a means to retain power.

The distinction between the members of the four castes and the non-caste group is even more divisive than the distinction between caste-members, and it explains why virtually all human societies have been more or less incessantly at war either with neighbors or in international or intercontinental alliances. For a particular people, wars were the means to become castes if they won, but non-castes

deep below all others if they lost. The latter brought a catastrophic decline of status and income, if not of life.

In addition to the three high-caste groups: the church; the government with its army; and the business world, modern societies have an additional caste, which like other castes tends to be hereditary: the scientific community – distinguished by high-level education and employment in universities and other scientific or modern business institutions and enterprises. When the new caste appeared there was a struggle, as in Iran today, between the new caste and the first caste, but it ended most often with a compromise.

Recently, the scientific caste has become much more numerous and stronger as a result of the development of computer technology. It is the beginning of the fundamental technological change from the industrial to the electronic age, which has been supported and promoted by the multinational companies and by many governments and scientific institutions.

The future change to the electronic age will bring mass-communication with English becoming the world language. This will make it possible for most of the future world population to be informed directly about conflicting points of view of different countries and castes, instead of receiving only or mainly the propaganda of its own group.

Already today, the spread of knowledge benefits from information in the media from private and public international and regional organizations, including the ones focusing on protection of the environment (*see* Global Forest Watch, Volume 4). The explosive growth of electronic information is of particular importance for schools in thinly populated and in poor regions, where illiterate children (and adults) can come in direct contact with the world by pictograms (icons) and images instead of being cut off from globalization by illiteracy.

Another group, which has benefited from the recent technological changes, are women from the new caste who have entered high positions by means of personal achievements or helped by heredity and education.

By its achievements, especially in the natural sciences, members of the scientific caste have become not only respected, but also feared by the other castes, the reason for the latter being that the scientific caste has obtained spectacular results not only concerning nature, but also relating to humans, their health and survival. A widely read book speaks of the *risk society*, where both rich and poor are threatened by the risks of modern scientific interference with human health and food (Beck, 1992).

The temptation is great to reduce cost and time, cutting short the time spent on research and development. The scientists, the laboratory, the producer, and even the government would like to be first to come to market; thus gaining in profit and honor. But in the past, there have been disastrous results of insufficient research, like the tragedy of the deformed Thalidomide-children, and, more recently, the lack of control with AIDS-contaminated blood. The damage is often discovered years later, and is in most cases irreparable. So the general public is roused to the risks of insufficient research, especially within genetic technology for food production and medical care.

The public case for strict control both with existing and new potentially life-threatening technologies is strong, and it is imperative that concerned scientists, producers, and governments combine their efforts to prevent misuse and thoughtless behavior.

In the international business community reaction to increasing globalization of the economy is positive, but that is not the case with most national governments, which are hamstrung by strong nationalism, which emerged in the industrialized countries in the 19th century and in the liberated colonies in the 20th century. With this background of increasing nationalism, it is not surprising that the meetings of the United Nations and other international organizations are often used to voice conflicting national interests, instead of promoting international cooperation as a replacement for wars.

Many adherents of economic globalization have a negative attitude to its cultural aspects. People with different cultures, religions, and customs have come much closer to one another, both because of immigration, and through increasing use of computer technology. For many, the reaction to this is negative. Because of nostalgia for the past and fear of the future, they revive abandoned old beliefs and customs, be they beliefs in spirits and fortune telling, customary dress, the patriarchal family pattern, or fundamentalist religious practice.

The revival of many shadows of the past, which have been sources of serious conflict, contributes to making the process of globalization more difficult. But it is not by returning to old customs of family and co-village preferences, or of racial prejudices, that the inescapable process of globalization can be carried out peacefully. In fact, we live in a period where increasing income differences, both within and between countries and continents, increase the risk that desperate individuals or governments will use the proliferating, more and more efficient, mass-killing technologies, and that they might succeed in eradicating all or a large share of humanity, deliberately or by error.

REFERENCES

Beck, U (1992) *Risk Society, Towards a New Modernity*, Sage Publications, London.

Boserup, E (1965) *The Conditions of Agricultural Growth. The Economics of Agrarian Change Under Population Pressure*, Earthscan, London.

Boserup, E (1981) *Population and Technological Change. A Study of Long-term Trends*, University of Chicago Press, Chicago, IL.

Boserup, E (1996) Development Theory: An Analytical Framework and Selected Applications, *Population and Development Review*, 22, 505–515.

Buck, J L (1937) *Land Utilization in China*, University of Nanking Press, Nanking.

Carneiro, R L (1998) *Slash and Burn Cultivation Among the Kuikuru*, Anthropologia, Carracas.

Evans, L T (1998) *Feeding the Ten Billion. Plants and Population Growth*, Cambridge University Press, New York, 145.

Needham, J (1954) *Science and Civilization in China*, Cambridge University Press, Cambridge.

Relating to Nature: Spirituality, Philosophy, and Environmental Concern

FREDERICK FERRÉ

University of Georgia, Athens, GA, USA

Classical Greek philosophy had its origin about 2600 years ago. Human spirituality is much older. No one knows exactly when our early ancestors began to experience religious awe or fear or ecstasy, but paleontologists nd evidence of spirituality along with the tools and implements of the earliest remnants of Homo neanderthalensis as well as Homo sapiens. Burial rites and cave painting strongly suggest that with the rise of our species also rose deep feelings and comprehensive beliefs, shaping our relations to nature.

The details of prehistoric spirituality must remain a subject for speculation, but it seems likely that tribal adoption of totem animals in uenced hunting practices, and that religious taboos and fertility festivals were involved in the invention and carrying on of agriculture. Because the great shift, about 10 000 years ago, from the mobile hunting gathering life to the settled existence of agriculture, and eventually cities, represents the single most signi cant change in the human relationship to nature in the entire history of our species; the spiritual attitudes guiding these practices had enormous signi cance for the environment. Earth worship, Sun worship, beliefs in spirits inhabiting elds, streams, and trees; ritual inhibitions on eating and mating; and the rise of shamans and priests, guiding group practices and keeping calendars: all these laid the foundation for deeply felt attitudes on what was permissible or impermissible in human dealings with the natural environment. Some of these attitudes may continue to resonate in the present day, almost too deep for notice or discussion.

*Vigorous discussion, however, of the alleged huge burden of guilt born by Christianity for sponsoring damaging environmental attitudes was prompted by a 1967 article by Lynn White, Jr, an eminent historian, in the magazine, Science. White accused Christian spirituality of undermining reverence for nature by isolating the holy in a transcendent creator God alone, thus leading to a sharp divide between the sacred and the profane. European attitudes based on such a spiritual dualism, applied from the middle ages to the present, White asserted, reduce nature to a mere resource for human use and development. His prescription was to counter a dangerous spirituality with an environmentally more benign one, also present in the Christian tradition, particularly that of St. Francis, who saw the sacred in nature all around him and wrote hymns to Brother Sun. As can be imagined, a variety of theological responses greeted this landmark critique. The debates continue to reverberate (*see also ***Christianity and the Environment****, Volume 5;* ***Ecofeminism****, Volume 5;* ***Religion and Ecology in the Abrahamic Faiths****, Volume 5).*

Philosophical as well as theological worldviews deeply shape attitudes toward nature. Classical Greek philosophy struggled with the de nition of the human position in nature, therefore in uencing the sorts of attitudes deemed appropriate for humans to take toward the environment. These philosophical issues were intimately (and interestingly) entangled with theological ones in the rst fteen centuries of the Christian era, separating again in the modern era, during which they have continued to shape the norms of human relations with other creatures, with matter in general, and with other persons in society. The dominant philosophical modern world picture was dualistic, with humanity and all values on one side, and the rest of the world, valuable, if at all, only because humans might value it, on the other. Alternative philosophical views have challenged this dominant picture, but in the atmosphere generated by the alliance of philosophical dualism with the early modern science of Galileo and Newton, these challenges have found it extremely dif cult to ourish.

The environmental realities of late 20th and early 21st century modern industrial societies have shown the urgency of nding ways of relating to nature that will allow both human beings and the rest of nature to thrive.

Hence there are calls for reexamination of the religious and philosophical grounds for the attitudes which contemporary men and women express in their treatment of nature and each other. The alienation between culture and nature, mind and matter, value and fact, sacred and profane, transcendent and immanent, needs to be healed. Assuming that most of the resources for changing any culture from within need to be found among the de ning achievements of that culture, resources from ancient Greek philosophy, biblical religion, and science will be especially valuable in any step toward reforming western postmodern environmental attitudes and practices. But since environmental changes and concerns are global, it will be important that foundations be laid for a wider ecumenical dialogue with spiritualities and philosophies drawn from non-western cultures as well.

RELIGIOUS POLARITIES OF IMMANENCE AND TRANSCENDENCE

Although the spiritual beliefs and practices of prehistoric humans pose many unknowns, there are rich sources of clues to be found in artifacts left by ancient peoples as well as in the observed beliefs and practices of recent or contemporary peoples living in preliterate (in this sense prehistoric) societies. The latter societies often share much in common with ancient practices of food production, hunting, housing, and the like, and may be surmised to have at least analogous worldviews and attitudes toward nature.

The earliest religiously significant artifacts left by ancient humans and contemporary preliterate societies both indicate strong attitudinal connections to nature. The sacred is felt as pervasive through, dwelling fully within, the natural environment. At one pole, spiritualities of this sort, in which the sacred is perceived as inside rather than outside the natural world, can be called spiritualities of *immanence* (in-dwelling the opposite of *transcendent*). Animals are revered. Hunting and agricultural tribes alike relate in totemic identi cations with characteristic powers of familiar species. Statuettes, the Venus gurines dating from the Upper Paleolithic Period (approximately 25 000 BC), depict women with exaggerated secondary sexual characteristics, breasts and buttocks, and are taken to indicate widespread fertility worship over a large geographical area, including Europe and Russia. Thousands of gurines of animals, and stone plaques of animals and people, show how important nature loomed in prehistoric consciousness. Also from the Upper Paleolithic (c. 15 000 BC) are cave paintings (e.g., at the Lascaux Grotto in South-central France) of animals, mainly in France and Spain but also in the Urals, probably created for ritual and magic involved with sustaining herd populations and with preparations for hunting. These show intense human interest in and close observation of the natural environment.

After the rise of agriculture and civilization in the fertile crescent, earlier Great Mother and Earth Mother myths, preserved in writing by the earliest civilizations of Sumer and Akkad, trace all natural gods to dei ed female generosity. But male fertility gods were also of great importance in ancient times. In many Middle Eastern societies the god Baal played a central role in relating humans to the natural order. In the mythology of Canaan, he is lord of storms, thunder, lightning, springs, and moisture in general. One of his tasks is to ght a major battle with Mot, the god of death, every seven years. If he wins, humans are assured seven years of plenty; if he loses, seven years of drought and famine result. In addition, in an obvious seasonal cycle, Baal must ght an annual battle with Mot toward harvest time. Baal always loses this encounter and winter comes, but his sister, Anath, annually revenges Baal by killing Mot, thus allowing Baal, and a new growing season, to return to life. Astarte, sister of Anath and Baal, represents another important Middle Eastern deity of sex and fertility, as well as of war. She was worshipped as the primary divinity of Tyre, Sidon, and Elath but also had important followings in Akkad, Egypt, and Europe. In Akkad she was known as Ishtar; in Egypt her name became assimilated to Isis; in Greece and Rome she was associated with Aphrodite (Venus), Artemis (Diana), and Hera (Juno); all aspects of the Great Mother. Aphrodite was patroness of human sexuality; Artemis was mistress of the animals, both wild and domestic; and Hera was matron of all things female, patroness of breast feeding, comforter of women.

The opposite pole in relating to nature is that of *transcendence*, in which the sacred (envisioned as dwelling outside or above the natural universe) is carefully defended from contact with all this familiar landscape of crops, animals, war, and babies. In the cultural and religious melting pot of the eastern Mediterranean region in the second and rst millennia BC, nomadic Hebrew tribes partly assimilated, partly held themselves apart, in the struggle for supremacy between these two great attitudinal modes of relating to nature. As strongly re ected in the Bible, some of the Hebrews, on adopting agricultural and urban ways of life, were tempted to add also the worship of Baal and his fertility-oriented consorts to the worship of El or Yahweh, a deity associated with the more ascetic life of tents, sheep herding, and desert storms. Much to the offense of the stern supporters of tradition, many Hebrews tried to have it both ways, seeing the sacred in natural processes and indulging even in the ritual prostitution and sexual promiscuity accompanying agricultural festivals of planting and harvest.

Hosea, an early Hebrew prophet during the reign of King Jeroboam II (786–746 BC), had a wife who may herself have been a temple prostitute, to his great personal anguish and moral displeasure. Israel, the northern kingdom of the Hebrew tribes, is condemned by Hosea for likewise prostituting herself in observing the feasts of Baal, sharing her favors with paramours other than her true husband, Yahweh. Hosea foresees destruction of the northern kingdom, as punishment for the spiritual promiscuity illustrated by the fertility rituals he found disgusting. But in a striking oracle which he acts out, he redeems (buys back) his adulterous wife from prostitution to indicate that Yahweh, as faithful husband, will ultimately take back at least a cleansed remnant of the chosen people on whom he has established his covenant love.

Other prophets, like Amos, Hosea's southern contemporary, thundered against the dilution of religious commitment to Yahweh represented by Baal-inspired attitudes toward the immanence of the sacred, and drew primary attention back to issues of transcendent principles of social justice instead. Caring for the poor, the weak, the vulnerable, not sacrificial ritual or mere lip service, is where the sacred (what God ultimately cares for) must be located. Hebrew contempt for the fertility-related modes of spirituality showed itself from the time of Hosea and Amos in language as well. Earlier, Baal-terms had been accepted without discomfort. King Saul (ruled c. 1021–1000 BC) had a son, Ishbaal. Still earlier, Gideon, the judge, was known also as Jurubbaal. But Queen Jezebel, a princess of Tyre and Sidon who married King Ahab (ruled c. 874–853 BC), campaigned vigorously to make worship of Baal exclusive and mandatory, aiming to replace the worship of Yahweh entirely. This polarized the atmosphere and afterward led the very name Baal to be replaced sometimes by the Hebrew word for shame, *boshet*. Ishbaal became Ishboshet and Baal's sister goddess, Astarte, was renamed Ashtoreth, a compound of Astarte and boshet through which the Hebrews expressed their stern disapproval for her followers' excessive immersion in sacred immanence.

The textual analysis of the Hebrew Bible is complex and many layered. But eventually (by around 500 BC) the priestly tradition, whose authors were responsible for the creation story with which Genesis begins, absorbed and affirmed in stirring imagery the prophetic insistence on the transcendence of God, far above and beyond the realm of nature, formed as it was out of the chaos of the waters over which God's spirit brooded. In later interpretation, this story was taken to depict the creation of the material universe out of nothing , so utterly distinct must the creation be from its transcendent Creator. In parallel, in keeping with the prophetic ideal, the opposition expressed in the Ten Commandments against other gods and especially against graven images , or any likeness of anything that is in heaven above, or that is in the earth beneath, or that is in the water under the earth , strongly reinforced in early Hebrew spirituality the refusal to find the sacred anywhere in nature but only in the Lord your God, who brought you out of the land of Egypt, out of the house of bondage (Exodus, 20: 2–17).

This stern stress on history and justice made for striking contrast with the morally easygoing emphasis, in competing spiritualities, on finding the sacred immanent within nature, whether in animals or in storms or in the Earth's fertility. If the sacred in any worldview is understood as whatever is considered ultimately valuable, whatever is finally worthy of worship, then Hebrew consciousness came, over time, to locate the sacred exclusively in the transcendent God, eternal source of righteousness, standard of justice for human community, and paradigm of faithfulness to a covenant made and sustained despite the vicissitudes of history. Nature was significant not in itself but mainly as the non-moral stage upon which the moral dramas of history could play out.

Coming to this view of nature was a gradual process, and not without alternatives in the development of Hebrew spirituality. In Genesis itself stands an alternative creation story, older than the more familiar priestly story of Genesis 1. This older story of Genesis 2 depicts a different order of creation and a much closer relation between human beings, animals, and the material order. God creates the first man not out of nothing but, rather, out of dust from the ground (Genesis, 2:7), and creates the animals afterward as helpers for man, so that he would not be alone (Genesis, 2:18). The animals all get their names from their human fellow creature. The power to give names does recognize the unique gift of language enjoyed by humanity, but it does not by itself suggest alienation or domination. Finally, to make a more fitting helper, God makes the first woman from organic tissue removed from the man. *Differentiated mutuality* (different parts in positive relation) is the mood of this older creation story.

The later, principal creation story of Genesis 1, in contrast, does explicitly encourage domination. The first humans are commanded both to multiply their numbers and to subdue all the rest of nature for human use and profit (see Genesis, 1:28). The moral tone expressed in the verb *subdue* by its Hebrew original is not gentle. The same verb, kābash, is found in violent contexts such as conquering enemies on the field of battle, or even assaulting a woman by force (see Esther, 7:8). In this context, the various animals and plants are explicitly given to humans for rule and consumption. Humans, uniquely made in the image of God are intended to take dominion (Genesis, 1:26) by the Creator God himself, standard giver of justice and model of righteousness. Nature, far from being a locus of immanent sacred value, is seen instead as made up of things to be controlled and used.

Within the Judaism and Christianity that grew out of early Hebrew spirituality, internal tensions can be found between the attitudes of dominion and those of differentiated mutuality. In Jewish law, animals were sacrificed with good conscience, but, in contrast, farmers were forbidden to muzzle an ox when it treads out the grain (Deuteronomy, 25:4), suggesting attitudes of mutuality and fair play with animals. In Christianity, Jesus is shown withering a tree with a curse (see Mark, 11:2–26) and using a herd of swine as vehicles to send evil spirits over a cliff into the sea to perish (see Matthew, 8:30–33), but also, in a more tender picture, he is quoted as appreciating the value of ravens and lilies, though differentiating their value from that of human beings, which he holds to be far greater.

> Consider the ravens: they neither sow nor reap, they have neither storehouse nor barn, yet God feeds them. Of how much more value are you than the birds!

Also,

> Consider the lilies, how they grow; they neither toil nor spin; yet I tell you, even Solomon in all his glory was not arrayed like one of these. But if God so clothes the grass which is alive in the field today and tomorrow is thrown into the oven, how much more will he clothe you, O men of little faith?
>
> (Luke, 12:24, 27–28)

The spiritual development of Western Europe over the following millennia is complex, influenced from many sources, but Christian ambivalence about the natural environment continued, with a gradual loss of the sense of differentiated mutuality. Mystical world denial, influenced by strains of philosophical Platonism, drew attention away from the world of the senses and toward radically transcendent realms of the sacred. In this, Christianity took a different trajectory from Judaism, which (despite neoplatonic subthemes of its own) remained more of the earth, earthy, with less sharp distinctions between soul and body than came to typify the medieval Christian worldview. Even in Christendom, of course, practical work needed to be done, whether on feudal estates or in monasteries. But such secular interests were clearly distinct from what was ultimately important. Practical inventions abounded in the middle ages, released from restraints that might have been imposed by religions of immanence. Christian opposition to the Druids, for example, released the oak forests of Europe from the protection of Druidic veneration, transforming sacred groves into sources of lumber.

These developments, strengthening the dominion attitudes of European Christians, have continued until recent times, with important reinforcement from modern philosophy and science (see the following section), and they are still reverberating throughout the industrial civilization (now worldwide) that traces its roots to Christian Europe. In America, in the 1960s, some theologians embraced and celebrated our culture's liberation from entangling spiritualities of immanence, a liberation provided to modern secular civilization by Biblical stress on the exclusive transcendence of the sacred (Cox, 1966).

THE LYNN WHITE CONTROVERSY

One of the most important single challenges to Christian self-congratulation on environmental attitudes came in 1967 from a seminal, wide-ranging article published by the historian, Lynn White, Jr, in the journal, *Science* (White, 1967). White, a noted expert on the thought and technology of medieval Europe, pressed a case leading squarely to the need for deep religious reform as a necessary condition for any long-term solution to the massive environmental destruction that humankind, led by western civilization, is wreaking on nature.

White's argument begins by noting the enormous impact the human race has had, and is increasingly having, on the natural environment, from taming the banks of the Nile six millennia ago, to the recent Aswan Dam. By irrigating, overgrazing, deforesting, and the like, humans have changed the face of the earth from time immemorial. But now there is a qualitative change. Since the onset of scientifically-led technology, the impact of our race upon the environment has so increased in force that it has changed in essence (White, 1967, 1203).

Science and technology, having such explosive powers when combined, need themselves to be understood if they are to be controlled. The linkage between them is a striking historical phenomenon, since for most of human history science was an elite, aristocratic activity and technology a lower-class function. Their recent marriage is the extraordinary outcome of the democratic revolutions of Western Europe, which, by reducing social barriers, tended to assert a functional unity of brain and hand (White, 1967, 1204). White continues: Our ecologic crisis is the product of an emerging, entirely novel, democratic culture. The issue is whether a democratized world can survive its own implications. Presumably, we cannot unless we rethink our axioms (White, 1967, 1204).

The first step toward this necessary rethinking, for White, is recognizing that science and technology as we know them are distinctly occidental. Though science has inherited much of value from other cultures, today, around the globe, all significant science is western in style and method, whatever the pigmentation or language of the scientists (White, 1967, 1204). Equally important, though less often noticed, is the fact that this leadership (especially in technology) is not a new phenomenon. In Western Europe, by 1000 AD, at the latest, water power was used in a variety of industrial processes besides grinding grain; within 200 years, wind power was harnessed

as well. The technologies of Europe did not excel those of other civilizations in delicacy but in power. Science, too, began to flourish in Western Europe much earlier than the more visible scientific revolutions of Nicolaus Copernicus (1473–1543) and Galileo Galilei (1564–1642). Eager translation of scientific texts from Arabic and Greek into Latin began in the late 11th century and gained momentum. By the late 13th century, Europe had seized global scientific leadership from the faltering hands of Islam (White, 1967, 1204).

How shall we account for this extraordinary (historically unprecedented) scientific and technological dynamism in Western Christendom? White proposes that the answer lies in the worldview of Christianity itself, particularly as interpreted by the theologians and applied by the peasants of Western and Northern Europe. In the terms of the section on religious polarities of immanence and transcendence, Christian exclusion of the sacred from immanence in nature freed Europeans to exploit the earth in unprecedented ways. White illustrates from the technology of plowing. Early plows, developed in Near Eastern and Mediterranean regions, where soil is light and dry, did not turn the sod but merely scratched it. Cross plowing was necessary, but two oxen were sufficient for such relatively gentle preparations for planting. In northern Europe, however, where soils are sticky and moist, another plow was invented as early as the seventh century. It was

> equipped with a vertical knife to cut the line of the furrow, a horizontal share to slice under the sod, and a moldboard to turn it over. The friction of this plow with the soil was so great that it normally required not two but eight oxen. It attacked the land with such violence that cross plowing was not needed, and fields tended to be shaped in long strips.
>
> (White, 1967, 1205)

This violence changed the relations between humanity and the soil in a profound and ominous way.

> Formerly, man had been part of nature; now he was the exploiter of nature. Nowhere else in the world did farmers develop any analogous agricultural implement. Is it coincidence that modern technology, with its ruthlessness toward nature, has so largely been produced by descendants of these peasants of Northern Europe?
>
> (White, 1967, 1205)

White rejects the idea of coincidence. What people do about their ecology depends on what they think about themselves in relation to things around them. Human ecology is deeply conditioned by beliefs about our nature and destiny – that is, by religion (White, 1967, 1205). The attitudes behind our current environmental crisis, encouraging exploitation of nature with no regard to non-human values is the residue of Christianity, White declares. Especially in its western form, Christianity is the most anthropocentric religion the world has seen (White, 1967, 1205). Victory over Druidism and other pagan religions of immanence turns out to have been double edged: By destroying pagan animism, Christianity made it possible to exploit nature in a mood of indifference to the feelings of natural objects (White, 1967, 1205). Linking such attitudes to a passionate interest in natural science, motivated by a desire, as was often said, to think God's thoughts after him, meant that modern western science was cast in a matrix of Christian theology (White, 1967, 1206). That theology was one of transcendence, not immanence. Thus White draws the following conclusions:

> first, that viewed historically, modern science is an extrapolation of natural theology and, second, that modern technology is at least partly to be explained as an occidental, voluntarist realization of the Christian dogma of man's transcendence of, and rightful mastery over, nature. But, as we now recognize, somewhat over a century ago science and technology – hitherto quite separate activities, joined to give mankind powers which, to judge by many of the ecologic effects, are out of control. If so, Christianity bears a huge burden of guilt.
>
> (White, 1967, 1206)

Assessing guilt is one thing, suggesting solutions is something else. White does not believe that superficial approaches to environmental concerns, mechanical or cosmetic clean-up fixes, will suffice, in view of the depth of the origins of these problems. But even science and technology, given their attitudinal and ideological makeup, are weak reeds. More science and more technology are not going to get us out of the present ecologic crisis until we find a new religion, or rethink our old one (White, 1967, 1206). To this end, White proposes that we ponder the greatest radical in Christian history since Christ: St. Francis of Assisi . St. Francis was a Christian who emphasized immanence.

> The key to an understanding of Francis is his belief in the virtue of humility – not merely for the individual but for man as a species. Francis tried to depose man from his monarchy over creation and set up a democracy of all God's creatures.
>
> (White, 1967, 1206)

He held a belief in animal souls: souls with intrinsic value and worthy of religious salvation.

> His view of nature and of man rested on a unique sort of panpsychism of all things animate and inanimate, designed for the glorification of their transcendent Creator, who, in the ultimate gesture of cosmic humility, assumed flesh, lay helpless in a manger, and hung dying on a scaffold.
>
> (White, 1967, 1207)

With this in mind, and in the conviction that we shall continue to have a worsening ecologic crisis until we reject the Christian axiom that nature has no reason for existence save to serve man (White, 1967, 1207), White proposes that Francis of Assisi should be the patron saint for ecologists.

Theological response to White's critique was voluminous, and at times sulphurous. There is no need here to

cover the blow-by-blow reaction of Christians to the ideas outlined above, but various general positions can be identified. It is at least controversial, for example, whether St. Francis is such a supporter of immanence as White claims. Are entities in nature sacred in themselves or simply as precious as gifts from God? Often such arguments are not so much direct refutations of White as meditations on Christian spirituality in relation to nature: meditations expressed in the context of a new sensitivity to environmental concerns stirred in the 1960s and focused by the Lynn White controversy.

One position, taken by Cox (1966), preemptively celebrated all that White deplored. Cox embraced and welcomed the secularization of nature represented by Judeo-Christian desacralization of the sacred groves. In so doing, he recognized that the environment would lose a measure of protection, at least for a time.

> Yahweh, the Creator, whose being is centered outside the natural process, who calls it into existence and names its parts, allows man to perceive nature itself in a matter-of-fact way. It is true, as some modern writers have pointed out, that modern man's attitude toward disenchanted nature has sometimes shown elements of vindictiveness.
>
> (Cox, 1966, 20)

But this vindictiveness is a passing phase, Cox argues, and should not be seen as a necessary part of mature secularization.

> Like a child suddenly released from parental constraints, he [modern man] takes savage pride in smashing nature and brutalizing it. This is perhaps a kind of revenge pressed by a former prisoner against his captor, but it is essentially childish and is unquestionably a passing phase.

The norm, rather, is quite different. The mature secular man neither reverences nor ravages nature. His task is to tend it and make use of it, to assume the responsibility assigned to the man, Adam (Cox, 1966, 20).

Since this was written shortly before the publication of White's essay and cannot be taken as a direct answer to it, Cox does not attempt to deal with the long, historical time frame of domination alleged by White. On its defensive side, his position is more focused on recent, obvious violence done to nature, by the union between secularized theory and practice that marks late modern western civilization. But, in addition, Cox launches an offensive, in keeping with the ancient prophetic stress on social rather than environmental concerns, against ancient and modern terrors created by the illicit sacralization of anything natural. This means that:

> we should oppose the romantic restoration of the sprites to the forest. It may seem pleasant at first to reinstate the tribal spirits, but (as Hitler made all too clear) once the Valkyries return, they will seek a bloodthirsty revenge on those who banished them.
>
> (Cox, 1966, 31)

Interpreting destructive environmental attitudes as simply the passing product of human immaturity makes no use of the Christian doctrine of sin, which could, if deployed here, account both for the human tendency to ravage nature and for the revenge of the Valkyries. An alternative position to Cox's, making sin central, does just this. The French social critic and theologian, Jacques Ellul, has long condemned technology (Ellul, 1964). More recently, in the midst of the Lynn White controversies, though not in direct reference to it, Ellul returns to the theological fundamentals of the human relation to nature (Ellul, 1984). The relation, he argues, was intimate and free in paradise, but after the primal fall of humanity into sin, the relation to nature is in all respects distorted and under the rule of necessity. Both concern for the earth and concern for social justice are of equal importance to God.

> The comparison reveals how much God loves creation with all that it includes, its variety, its blossoming; and it means that the ecological destruction is of the order of sin: as considerable as war, genocide, the exploitation of man by man, injustice. There is no scale of sin; there is only the immensity of the love of God from which man claims to escape, the love which still endures for this creation that we ravage.
>
> (Ellul, 1984, 152)

And ravage we must. For Ellul,

> it is impossible to detach the relationship of man with the world from his relationship with God. It is no accident that from the moment when man challenged God (the 18th century), the technique of exploiting the world has taken off. It is no accident that the place where man decided that there was no God and that nature was purely natural is the place where the exploitation of the world developed (the west).
>
> (Ellul, 1984, 153)

Since God and scientific technology are mutually incompatible,

> conversely, it is no accident that the first great goal of the science oriented toward technology was to shake off the restrictions of divine tutelage. The effort to affirm science by itself, without limit, as judge of everything and carrying its own legitimization *inevitably* involved, as the other side of the coin, the devastation of the world, the squandering of possibilities, the frenzy of destruction.
>
> (Ellul, 1984, 153)

The faithlessness of Jews and Christians, who were in a position to understand this and respond with repentance by setting limits to human aggressiveness, has sealed the world's fate. It means that they, and they alone, are responsible before God, for the disaster in which we are beginning to live (Ellul, 1984, 154).

Ecological disaster, an inevitability for Ellul, may indeed be in store, say other theologians, but perhaps a middle way averting disaster can be found: a way in which humanity can

play a more positive role tending the Earth with constructive rather than destructive interventions. The stain of sin, no matter how deep, must be balanced by the readiness of God to redeem any situation, no matter how bleak. For Egbert Schuurman, a Dutch theologian, only this additional assurance makes the Christian message a Gospel, good news . God's grace holds the key to human repentance. This means that the stubborn pride that currently distorts human attitudes toward nature and society can at any moment be melted and a new order instituted. Humans need above all to pray for a widespread change of heart. Guided by the right motive, man in his cultural activity can be a blessing for nature (1 Kings, 4:33–34) and at the same time enter into an open way toward the future (Schuurman, 1984, 117).

Theologians of this middle way, regarding the proper human attitudes and actions on Earth may not look to St. Francis as the patron, as Lynn White suggests, but may prefer another outstanding figure in Christian history, St. Benedict. St. Francis celebrates unity with nature, but an ideal Franciscan democracy of all God's creatures allows little scope for the sort of human leadership, manipulation, and control, that complex civilization requires.

> Compared with St. Francis's deep feeling and sense of union with the natural world, St. Benedict's response was more practical, using nature with care and respect. The Benedictine monasteries combined work and contemplation. They developed sound agricultural practices, such as crop rotation and care for the soil, and they drained swamps and husbanded timber all over Europe. Benedictines were creative in practical technologies related to nature.
>
> (Barbour, 1980, 25)

PHILOSOPHICAL POLARITIES OF EXCLUSION AND INCLUSION

It is all very well for theologians to debate whether the natural environment should be regarded with respect, reverence or brotherly love, but if there are no fundamental categories available for general public understanding of how natural objects can be thought to have value at all, then the link between spiritual guidance and public policy is severed. This has been the case for most of the modern period. The outstanding philosophers of the age have tended to exclude values from any firm locus in nature, and the civilization whose modern conscience and institutions were shaped in the general worldview of its intellectual leaders reflects this exclusion in its practices.

Two feuding philosophers, both founding fathers of the modern worldview, agreed on little else, but did close ranks against the possibility of values subsisting, in any way, in nature. Thomas Hobbes (1588–1679) was a monist, a materialist who denied that the human soul is anything more than the reverberations of fine atoms; Rene Descartes (1596–1650), in contrast, was a dualist who held that the soul is absolutely immaterial and free from contamination by material categories. Both, however, accepted the distinction, propounded earlier by Galileo Galilei, between primary and secondary qualities. The former, primary qualities are the properties of natural objects open to mathematical description, such as size and shape, weight and momentum, location and motion. These are held to be characteristics, actually belonging to the thing in itself. Since they can be quantified, they can be grasped with precision, understood objectively. The latter, secondary qualities are the colors, smells, tastes, sounds, and feels of things. They are changeable, relative to the setting and the condition of the perceiving organism. They cannot be quantified but can be described only vaguely by eliciting subjective memories not everyone shares. They are the source of countless illusions, optical, auditory, and tactile. In extreme cases they obviously do not belong to things but are merely in the perceiver.

Galileo gives as an example the capacity of a feather to tickle. For some people it may have great tickle-power, but that is to say something about the people, not the feather. Now this tickling is all in us, and not in the feather, and if the animate and sensitive body be removed, it is nothing more than a mere name (Galileo, 1954, 86). But for Galileo, Hobbes, and Descartes, and through them for most modern philosophy, this same dependence on the presence of our animate and sensitive body is manifested by all the sensory qualities. As Galileo put it, Of precisely a similar and not greater existence do I believe these various qualities to be possessed, which are attributed to natural bodies, such as tastes, odors, colors, and others (Galileo, 1954, 86). From this comes the first of the great cliches that virtually every school child has learned from modern philosophy: If a tree falls in the forest and no one is there, does it make a sound? The standard modern answer, of course, is no . Physical vibrations in air and earth will be caused, but without ears and minds to translate those quantifiable vibrations into qualitative crackles and thumps, there is no sound, strictly speaking.

But if the hues of autumnal leaves are not really in nature but only in animate and sensitive lookers, and if the tones of birds are not really in nature but only in animate and sensitive listeners, then how much more obviously subjective must be the feelings of approval and enjoyment that human beings experience while looking and listening within their environment! From this comes the second modern philosophical cliche: Beauty is in the eye of the beholder . This saying can, and must (in this philosophical framework) be generalized to all values. Just as beauty is generated by human appreciators, so must goodness be generated by human judges, and truth be generated by human thinkers. Nature itself possesses no such values. It is odorless, colorless, soundless and tasteless

in itself. The gardenia's smell, the sunrise's hues, the brook's burble, and peach's flavor are dependent on human presence. Therefore, it makes no sense to preserve them from human exploitation or even to respect them along with humans. It is conceptually unthinkable for them to have intrinsic value apart from humans; it would be sheer philosophical confusion to imagine that they could have standing against human interests, could be addressed with reverence, or treated with brotherly love.

The immense prestige of the mechanical world picture created by Isaac Newton (1643–1727), which incorporated these key assumptions on the exclusion of qualities, values, and experience from the basic stuff underlying the natural world, greatly reinforced earlier views on the inert, intrinsically valueless character of nature in itself. This exclusionary world picture was then brought to its logical climax in the philosophical explosion lit by Immanuel Kant (1724–1804), who gave powerful deductive arguments for the absolute separation of the world we experience, the world in which we live and in which science works, from the world as it is in itself. Freedom, morality, knowledge, God, our own true selves (everything that is genuinely valuable) resides in the latter realm, which Kant called the *Noumenal*. Whatever can be experienced or even positively thought belongs to the realm of appearance, which Kant called the *Phenomenal* (Kant, 1966). Dignity is deserved only by moral agents, in Kant s view, and such agents must exclusively be rational, since morality itself consists in willing actions out of respect for moral lawfulness, and moral laws can only be recognized and obeyed by rational beings. Nothing in nature, therefore, can have dignity or (on the opposite side of the coin) rights. If we refrain from shooting an old, faithful dog which has served us well, it is not because the dog has a right to a comfortable retirement, for the dog cannot judge , but because we should cultivate tender feelings as part of our duty to humanity (Kant, 1963, 240).

Later, in the 20th century, philosophers convinced by Kant (and the main thrust of modern philosophy) that rights-language has no place in connection with natural things, but unpersuaded that it is any more possible to demonstrate by reason alone the foundations of morality even among human beings, dismissed all values as mere emotional expressions of preference and aversion (Ayer, 1946, Chapter 6). On this view, such sayings as nature is intrinsically valuable or we should all respect the dignity of nature are literal nonsense. There is no way to verify or falsify such utterances if they are taken as attempting to make claims about matters of fact that could possibly be true or false. Any and all experienceable facts would be equally compatible with them. But if taken instead as expressions of a speaker s feelings of warmth toward nature or a wish that others might share such attitudes, the utterances can be seen to have a function. What they do not do, however, is describe anything about nature itself, which stands extralinguistically as a vast realm of actual or possible appearances capable of con rming or confuting our hypotheses, and nothing more.

Despite enormous differences in thought-worlds, interesting similarities link this recent philosophical view of nature to Plato s in the fourth century BC. Things and the appearances of things are low on the hierarchy of values for Plato (c. 428 348 BC). When sense experiences can be uni ed into formal concepts of things, and concepts can be linked into theories (mathematical and otherwise) of increasing inclusiveness, there is a corresponding rise in worth, for Plato, both epistemological and metaphysical. The natural world of constant change ranks only on the lower order of mere becoming; the intelligible Forms that give understanding, in contrast, qualify for the higher order of true being: changeless, eternal, beautiful. Finally all are ideally uni ed through the supreme form of the good. Finite things are not valuable in themselves as changing particulars bound to the cycles of generation and decay, but gain value only to the degree that they exemplify the eternally perfect forms that give them meaning.

Plato s rmly exclusionist position, locating all genuine value outside nature in a higher, immaterial domain, was an easy t for Christian doctrine in need of a philosophical framework. There are Platonic themes inside the New Testament itself, especially in the Fourth Gospel, and during the rst three centuries AD, the early Church Fathers found the transcendence motifs of Judeo-Christian traditions hospitable to Platonic interpretations (Wolfson, 1970). Here, as noted in section one, no immanent sacredness in groves, neither Greek nor Druid, could nd legitimacy. The interplay between Christian spirituality and Platonic philosophy dominated European consciousness for more than 1000 years, encouraging other-worldly attitudes where ultimate values were at issue, bifurcating nature into a picture album lled with iconic reminders of eternal realities, on one level, and a mere warehouse for human exploitation, on the other.

The exclusion of value from nature is the lead story of western philosophy, dovetailing nicely with the transcendence of value in dominant western spirituality. Still, there has all along been an undercurrent of dissent, not only in religion (as in St. Francis), but also in philosophy. Plato s most distinguished student, Aristotle (384 322 BC), dissented at once. Instead of banishing value from the substances in nature to an independent realm of eternal forms, Aristotle proposed that we understand natural things as the inextricable combination of forms with individualizing, spatially locating matter, a hylomorphic (matter form) way of including all that Plato found valuable, but fully within the burgeoning domain of change. Living things can then be seen as working out their own internal Forms, their natures or essential characters. Their intrinsic teleology

(explanation of something in terms of the purpose it serves rather than in terms of its causes) can be fulfilled or frustrated. A tree can be chewed down before it has a chance to bear fruit, but a beaver can use its trunk and branches to build a dam in which to house and nurture precious offspring. Nature is a domain of competing and cooperating cycles of individual subjects of lives with intrinsic ends, norms of development, experiences, and satisfactions. All living things, for Aristotle, have souls, understood as the functioning of their innate characters, or forms. Plants have the powers of nutrition and reproduction and thus have vegetative souls; animals have these powers with the additional capacities of locomotion and sensation, giving them sensitive souls. Human beings have all the previous functions plus the conscious abilities to calculate and reflect, which is what is meant by *rational souls*. In every case it is the whole organism that is functioning in the ways characteristic of its essential nature: matter and form working together. Particular beings do not merely exemplify something higher, but manifest and embody the values of their forms concretely.

Aristotle's dissent from Plato was lost to Europe for the first millennium of Christian history, but it had been kept alive in Muslim thinking, and during the 12th century it was recovered by the West. This recovery, and its assimilation in the 13th century through the work of great scholastics like Albert the Great (1200–1280) and St. Thomas Aquinas (1224–1274), severely shook Platonic orthodoxy, but Aristotelianism prevailed, overwhelming the universities of the 14th and 15th centuries with the power of its new synthesis. Unfortunately for Aristotelian inclusionism, however, the very success of its worldview lured it into triumphalism. It became rigid and its science, scholastic in the familiar pejorative sense, the epitome of what modern thinkers like Galileo, Hobbes, and Descartes scorned as old-fashioned, reactionary word play. The value-excluding counterattack of modernity felt fresh and liberating in contrast.

Still, in modern philosophy, as a persistent undercurrent, value including positions were voiced despite the inhospitable climate. Baruch Spinoza (1632–1677), in lonely isolation, worked out a brilliant counter to the dualism of Descartes and the materialist reductionism of Hobbes. His solution was to find God in nature and nature in God, a pantheism in which everything should be considered to have infinite attributes, of which mind and matter are only two; though two of great importance to human beings. Nature, comprising all there is or could be, should be seen both as determined and created, *natura naturata* (nature natured), and as free and self-creating, *natura naturans* (nature naturing). In both respects, every manifestation of nature is a mode of manifestation of the underlying divine substance, the infinitely perfect one. Spinoza, having been expelled from his synagogue and enjoying few academic connections as a lens grinder, left no disciples, though his thought gained popularity in the 18th and 19th centuries among romantic and counter-cultural thinkers like Johann Wolfgang von Goethe (1749–1832) or Samuel Taylor Coleridge (1772–1834).

Gottfried Wilhelm von Leibniz (1646–1716) also argued against the grain of dominant western philosophy for an inclusionist position that took a very different shape from Spinoza's. Leibniz, a universal genius who invented symbolic logic and co-discovered the calculus with Isaac Newton, developed a highly pluralistic worldview in which the elements of everything in nature are to be seen (as in Spinoza's universe as a whole) with dual aspects of materiality and mentality. These ultimate *monads* (any elemental unit, like a soul) have experience and value, even if only dimly, for themselves. There are many different grades of monad, including human consciousness but in principle stretching high above and far below throughout the universe of things. Each manifests its own internal essence, much as Aristotelian living substances do. On Leibniz's theory of substance, there can be no direct relations among the monads, but all are coordinated by the creator God, whose perfection entails that only the best one of all the possible worlds that might have been created was actually realized. Leibniz in his time made few disciples. Excluded from the intellectual establishment represented by the Royal Society of London, he, too, had to await the 19th century for rediscovery and appreciation.

Other philosophers, notably Georg Wilhelm Friedrich Hegel (1770–1831) and Henri Bergson (1859–1941) labored to develop conceptual frameworks, in which nature could be thought to include its own rich palette of values, and these philosophers and others developed strong followings, especially outside the circles of thought in which the methods and presumptions of modern science are taken for granted as authoritative. But these have always been views struggling to breathe in the atmosphere created by modern western culture, whose dominant religion, science, and mainstream philosophical tradition all work to smother ways of thinking that would include values directly in nature. The western industrial machine, if it is to be countered or transformed by modes of thought and spirit, has not yet met its match.

BALANCING THE POLARITIES

Although machines are the embodiment of ideas and attitudes, knowledge and goals, it is true that once they are embodied, they are not easily stopped by ideas and attitudes alone. So it is that the rapacious western industrial machine, now worldwide, is likely to require a special convergence of circumstances, both physical and mental, before it can be tamed to the requirements of environmental good citizenship. In part, this convergence actually began in the latter part of the 20th century, when industrial civilization began to choke on its own runaway excesses. Pollution of

air, soil, and water, depletion of nonrenewable resources, desecration of precious beauty, distortion of social fairness between have's and have-nots: all these can be seen as plainly predictable outcomes of the modern system, not flukes, and thus they urgently require rethinking of the ideas, and adjusting the attitudes, that went into designing the industrial machine in the first place. In this spirit Lynn White cast around for an alternative religious hero for Christians, offering St. Francis as inspiration. Had there been no environmental crisis, no pollution disasters, no species extinctions, no fears of still greater horrors ahead, there would have been no need to search for blame or to seek new spiritualities. But practical exigencies have in fact arisen, preventing reliance on the *status quo*. They call in question even powerful orthodoxies, which in the absence of crises might have remained unchallenged. The prestige of science and technology, now recognized as contributing to the problem, is not enough to quiet the challenge. Something profound has gone wrong. Perhaps the atmosphere of the 21st century may be more hospitable to deep ideational reforms offering more comprehensive thinking, finer spirituality, and healthier ways of conducting practical life on the Earth.

The practical cash value of the polar clash between religions of transcendence and those of immanence comes down to the question whether the human race should conduct its business with nature as from above or from within the environmental order. The spiritualities of transcendence offer justifications for the former; spiritualities of immanence provide support for the latter. Each speaks with power to shape attitudes and policies. Is there any hope of finding a position of balance where the dynamics offered by the two poles can both be incorporated? Such balance has not often been sought. Even if this position were to be found, actual religions would need to make potentially painful internal adjustments to incorporate it. But perhaps the urgencies of the present historical situation could make the early 21st century an era of rare openness to just such religious growth and change.

Doing justice to the deep intuitions supporting human transcendence over nature without opening the door to heedless species chauvinism should be possible if actual human capacities are affirmed without indulging in the prejudicial snobbery that has too often distorted the affirmation. Human beings actually are different in profoundly significant ways from anything else found in nature. Nothing else in the known natural universe can approach human capacities for conceptual freedom. Normal adult humans can manipulate linguistic symbols to take account of possible or actual states of affairs far remote from actual circumstances. Language allows humans to consider distant places (across the mountains) and remote times (when Grandfather was a little boy , or next spring, at planting season) and weave detailed tapestries of thought out of insubstantial ideas. This capacity allows the development of long-term purposes and the fashioning of implements designed to achieve them. In this quite obvious way, human beings transcend the here and now of nature. Even highly intelligent non-human species do not come close to exhibiting human freedoms of dealing thoughtfully with what is not real and may never be real. In fact, one of the crucial ways in which humans use their transcendence over nature's *here-and-now* is to prevent the actualization of unpleasant possibilities envisaged in thought and therefore countered. An equally crucial way of using transcendence is by long-range planning of complex desired outcomes that could never have become real without the steady application of purpose.

Dreaming of distant futures and choosing between them, effectively harnessing present resources to achieve ideas and actualize outcomes: these are the basis for another sort of transcendence of nature, the transcendence of moral responsibility. Rocks slide and cause damage but are not responsible because they were only following the laws of physics and could have done nothing else under the given circumstances. Likewise predators kill but are not responsible. What they do is not deliberately chosen. Instincts refined by natural selection come into play as the animal responds, sometimes intelligently but always in ways restricted to the *here-and-now* of environmental circumstances. Human beings, in contrast, free to consider alternative possibilities in some detail and to choose which one to bring into actuality after (at least sometimes) extensive thought and sustained effort, transcend the rest of nature by carrying the burden of responsibility for outcomes deliberately achieved. The rest of nature operates by caused causes and cannot be blamed. Only human beings are influenced by norms and ideals (for better or worse) and can be held to account for their actions. If humans did not transcend nature in this way, there would be no point in appealing for improved environmental policies. One does not appeal to stratospheric ozone to reverse trends for the good of the climate, or urge deer to stop reproducing for the good of their habitat; but one may appeal to humans, on behalf of a larger envisaged good, to sacrifice some of their short-term interests or even to curb their appetites. And sometimes these appeals may be heeded, translated into action. Those who say that humans are no different from the rest of nature would in logic need to quiet their preachments. But of course preachments and appeals to conscience and ideals continue, appropriately, since humans do transcend nature in being morally responsible agents.

In these special capacities, for flying conceptually above the here and now, for taking account of the remote future and considering which of various envisaged possibilities should be realized, for implementing deliberate purposes, for knowing the self as morally responsible agent and as precious center of awareness destined to die, we discover the special joys and pains of personhood. The spiritualities

of transcendence, it seems, give voice to the justified awareness of the special value, the unique dignity, that goes with being a human person. So understood, this is far from a license to rape the rest of nature. It is, instead, the acknowledgment of special burdens and an invitation to unique satisfactions.

Though unique in these ways (the ways in which human beings are persons) our human species is by no means unique in the multitude of ways in which we are organisms. The spiritualities of immanence, stressing the intense and comprehensive values of immersion in nature, are obviously correct in stressing continuities and their profound importance. Our bodies, like those of most living species, both animals and plants, are sexually reproduced; and this is good. The fertility festivals and orgiastic rites of early religions of immanence celebrated the sacredness of nature's creativity, the urge to multiply and swarm, fructifying and filling every niche. Human bodies are made of meat and need nourishment. Metabolism reigns in us, as in all other forms of life. Food, water, air in adequate quantity and quality, are necessities for us as for all other living things in nature. The harvest festivals and hunting rituals associated with the spiritualities of immanence celebrate sacredness in domains of undeniably vital significance.

The Hebrew prophets, stressing the personalistic values above all (moral faithfulness, purity, justice) were offended by the organismic values they saw being venerated in prosperous agricultural Canaan. There were cultural and historical reasons for the revulsion of the prophets against the city-dwelling Canaanites. But there seems no logical requirement that we adopt the either/or attitudes they preached. Rejoicing in fertility need not be done unjustly, to the detriment of widows and orphans or against the interest of the powerless. On the contrary, acknowledging the continuities between our organic constitution and those of other animals, affirming the networks of relations between all animals and plants and the inorganic environment that supports the lives of all; these tributes to organicism are compatible with recognizing the unique values of personhood.

Perhaps, if we were to speculate about the spiritual and ecological needs of the future, we should try to work out a new position, we could call it personalistic organicism, a viable spirituality that could live creatively in the dynamic tension between the poles of transcendence and immanence. If we could establish such a point of balance, then ecologically sensitive moral agents, enabled to feel and appreciate vital natural connections between humanity and all other life forms (immanence) through uniquely powerful personal norms (transcendence), might be liberated to seek social and ecological justice without sacrifice of either aspect to the other.

This is not the place for the detailed working out of this project, which has been undertaken elsewhere (Ferre, 1996, 1998, 2001). There is no reason in principle, however, against hoping that the polar opposition between images of transcendence and immanence can be balanced in some yet more inclusive mode of spirituality. Can a synthesizing worldview, such as personalistic organicism, also hope to balance the theoretical tensions between the philosophies of exclusion and inclusion?

The great strength of philosophies arguing for the exclusion of values from nature has been the argument from the necessary subject-dependence of value. Beauty is in the eye of the beholder , or some variant of this modern mantra, is hard to deny. What could beauty be, if entirely divorced from the possibility of any appreciator? Just as pain (or Galileo's tickle) makes no sense apart from subjective awareness, so beauty and like values cannot be understood without ultimate reference to some center of appreciation or valuation. No purpose without a purposer; no love without a lover; no value without a valuer. Values have to be valued by some valuer to be real. The case is strong. Objective circumstances may be perfectly suited for grounding a joyful experience of beauty, but if the experience never occurs, those unrequited circumstances alone will not be enough to count as beauty itself.

This much can be agreed. Such agreement, however, does not force the conclusion that nature has no values of its own. The inclusion of values in nature can simultaneously be affirmed by expanding the scope of subjectivity. Kant was right in stressing the centrality of mind, but wrong in defining mind too narrowly in terms of rationality, in which human persons stand out from the rest of nature. Modern philosophy in its main tradition has been deeply but mistakenly anthropocentric in assuming that only human beings are subjects in the relevant sense of the word. No one present , in this constricted thought frame, tends to mean no human at hand. If the old saw about the tree falling in the forest were reframed, however, in terms of sentient organisms, there would be a much richer and more nuanced answer. Squirrels, bears, birds, rabbits, deer (perhaps spiders and earthworms, too) would detect the crash and react, each in its own way. Similarly, if subjects (defined as at least dimly aware centers of experience) were to be acknowledged throughout nature, there could be no doubt that qualities too are firmly woven into nature's fabric. Hummingbirds and bulls distinguish and appreciate the quality, red, in different ways. These appreciations, positive and negative, are facts about nature but also facts about how valuing is active in nature. Since butterflies and bees are drawn to attractive visual appearances and odors, for example, natural selection among flowers has run riot in sponsoring competitive beauty contests. Throughout – in the air, on the land, deep in the sea – the attractions and repulsions of positive and negative valuations are an obvious dynamic of nature.

Thus, since practical ethics finally rests on the felt obligation to show appropriate respect to quality and value, of

whatever kind, wherever found, a worldview recognizing subjective valuings beyond the human, will gain a firm basis for thinking through and proclaiming human duties, owed directly to entities in nature as well as to fellow human centers of value. Broadening the scope of subjectivity can balance the polarities between the truths of exclusion (only subjects, not inert objects, can ground value) and the truths of inclusion (subjects are everywhere). But this can happen only if the great divide, mistakenly fostered by Kantian anthropocentrism, is successfully challenged.

As one such challenger, personalistic organicism, stressing the continuities beneath the differences between entities in nature, will recognize alliances with other philosophies of continuity. Aristotle's stress on the innate teleologies of things, that they seek goods of their own and have their own fulfillments, will be an important insight to retain. Spinoza's and Leibniz's view that all things have internality, a mind-like character, will be another. Also alert to new data and theories, personalistic organicism will stress a postmodern physical understanding of the energetic, self-organizing universe that increasingly, through much of the 20th century, was revealed by science. Not the inert particles of Newton and Kant, but the dynamic wave-events of 21st century physics, interpreted as moments of extremely simple experience, will offer personalistic organicism an inspiring model for the basic elements of the vibratory universe (Whitehead, 1925). These, as they creatively wrestle actuality out of potentiality, are nature's minimum units of value, both in themselves and for others (Whitehead, 1976).

On such a postmodern speculative vision, responsive both to dynamic science and to balanced spirituality, a postmodern ecumenical dialogue is conceivable between religions and philosophies, East and West, on new ways to cooperate in the quest for social justice among persons and for moral consideration of the Earth.

REFERENCES

Ayer, A J (1946) *Language, Truth and Logic*, 2nd edition, Victor Gollancz, London.

Barbour, I G (1980) *Technology, Environment, and Human Values*, Praeger, New York.

Cox, H (1966) *The Secular City: Secularization and Urbanization in Theological Perspective*, 2nd edition, Macmillan, New York.

Ellul, J (1964) *The Technological Society*, translated by John Wilkinson, Vintage, NY, originally published as *La Technique ou l'Enjeu du Siècle*, (1954), Librarie Armand Colin, Paris.

Ellul, J (1984) The Relationship between Man and Creation in the Bible, translated by W Deller and K Temple, in *Theology and Technology: Essays in Christian Analysis and Exegesis*, Vol. 73, eds C Mitcham and J Grote, University Press of America, Lanham, originally published (1974) as Le Rapport de l'Homme a la Creation Selon la Bible, *Foi et Vie*, 137–155.

Ferre, F (1996) *Being and Value: Toward a Constructive Postmodern Metaphysics*, State University of New York Press, Albany, NY.

Ferre, F (1998) *Knowing and Value: Toward a Constructive Postmodern Epistemology*, State University of New York Press, Albany, NY.

Ferre, F (2001) *Living and Value: Toward a Constructive Postmodern Ethics*, State University of New York Press, Albany, NY.

Galileo (1954) *Opere*, in *Metaphysical Foundations of Modern Physical Science*, Vol. IV, ed E A Burtt, Doubleday, Garden City, 333.

Kant, I (1963) *Lectures on Ethics*, translated by Lewis Infield, Hackett, Indianapolis, IN.

Kant, I (1966) *Critique of Pure Reason*, translated by M Muller. Doubleday, Garden City, New York.

Schuurman, E (1984) A Christian Philosophical Perspective on Technology, in *Theology and Technology: Essays in Christian Analysis and Exegesis*, eds C Mitcham and J Grote, University Press of America, Lanham, originally published (1980) as *Technology in a Christian-Philosophical Perspective*, Potchefstroom University, Potchefstroom, South Africa.

White, Jr, L (1967) The Historical Roots of Our Ecologic Crisis, *Science*, 155(3767), 1203–1207.

Whitehead, A N (1925) *Science and the Modern World*, Macmillan, New York.

Whitehead, A N (1976) *Process and Reality: An Essay in Cosmology*, corrected edition, eds D R Griffin and D W Sherburne, Free Press, New York.

Wolfson, H (1970) *The Philosophy of the Church Fathers*, 3rd edition Harvard, Cambridge, MA.

Social Science and Global Environmental Change

STEVE RAYNER[1] AND ELIZABETH L MALONE[2]

[1]*Columbia University, New York, NY, USA*
[2]*Pacific Northwest National Laboratory, Richland, WA, USA*

The descriptive *tradition in the social sciences, which tends towards quantitative methods of tracing stocks and flows of social data through time and space, has been seen as more useful in mainstream global environmental research that relies on natural science methods and models. The* interpretive *tradition has been relegated largely to the sidelines, with its concerns to understand motivations, ideas, and values, and its frequent analytic focus on narrative and text.*

The descriptive approach derives much of its strength from simplifying assumptions about complex behaviors and tends to assume homogeneity of human responses to price signals. It therefore is often used for large-scale analyses. In contrast, the interpretive approach emphasizes variation and complexity, sometimes to the point of resisting simplification on principle. It therefore tends to produce local and disaggregated analyses. Descriptive researchers tend to see themselves as outside observers and to make utilitarian assumptions; interpretive researchers see themselves as participants in their studies and often assume individual rights-based ethics.

These are differences in kind that are both essential to research in global environmental change. For example, descriptive approaches have revealed much about what would happen under various scenarios of climate change, but interpretive approaches can provide value-related parameters as a basis for choosing among candidate policies.

Collaborative research that includes both descriptive and interpretive approaches has been carried out in studies of actual decision-making, which is a social-political process, not a rational, utility-maximizing one; and in mutual learning systems that incorporate models and negotiation. Further collaborations among researchers from both traditions and policymakers can help to bridge the current gap between industrialized and less industrialized countries, to expose assumptions that can improve decision making under uncertainty, and to make policy decisions more implementable and effective (see Box 1).

TWO STYLES OF SOCIAL SCIENCE

Social science in general is characterized by a long-standing gulf between two traditions or styles of analysis. One is the *descriptive* tradition, that identifies and traces stocks and flows of population, goods, services, wealth, and information, through time and space. Statistical modeling is an important tool in this approach. The other is the *interpretive* tradition, that seeks to understand the motivations, formation of ideas, sources of values, the creation and interpretation of information, and the development of shared understanding as the basis for the conduct of discourse and human actions in society. Focus on text and narrative is a frequent focus of this approach.

Because the two approaches are so different in kind, there has traditionally been little cross-fertilization between them. Models in both kinds of social science are simplifications intended to enable analysis of more complex cases. Simplifications tend to be more accepted in quantitative models; in qualitative models, they may be seen as mere stereotypes. Descriptive models, which tend to be mathematical or quantitative, are incomplete in that they cannot account for the full variety of behavioral variations among diverse human populations. Interpretive models, which tend to be verbal or qualitative, are also incomplete in that they are seldom capable of powerful generalization from case studies.

Public policy makers and private sector decision-makers tend to favor the descriptive style of analysis because it usually provides quantitative information that can be

Box 1 The Two Approaches to Social Science Analyses of Global Environmental Change

The *descriptive tradition:*

- Quantitative book-keeping of stocks and flows of social data
- Assumes homogeneity in human responses over large areas
- Models global change as a series of interconnected 'big' boxes
- Research agendas usually set by natural scientists

The *interpretive approach:*

- Emphasis on defining problems, rather than solving them
- Seeks to understand motivations, sources of values, interpretation of information
- Studies are most frequently undertaken at small spatial scales and results are often disaggregated
- Research agendas set by social scientists

represented as an objective basis for an organization s decision making, independent of the judgment of any particular decision-maker.

In the social science of global environmental change, the gulf between descriptive and interpretive approaches is exacerbated by the heavy reliance of global environmental science on models designed to represent the stocks and ows of resources and pollutants in an environmental medium, such as land, the atmosphere, or the ocean. These models can be made to interface with the models of economic growth and resource use that are typical of social science in the descriptive mode.

Opportunities for social science collaboration in major international scienti c research programs addressing global environmental change, such as the Intergovernmental Panel on Climate Change (IPCC), have been largely focused on extending the framework already established by the natural sciences. For climate change, that framework consists of a four-box conceptual model: quanti ed emissions of greenhouse gases, atmospheric chemistry, climate and sea level, and ecosystems. Within this framework (Figure 1), the social sciences provide highly aggregated data on human activities leading to greenhouse gas emissions. These data can be used to drive natural science models of global atmospheric chemistry and physics. In turn, the natural sciences aim to model climatic impacts on managed and unmanaged ecosystems upon which humans depend. At this point, social scientists are invited to project the outcomes of these changes for large-scale patterns of agricultural and industrial activity, stimulating macroeconomic and technological responses, which, in turn, may eventually alter anthropogenic emissions estimates. The outputs of such research are presented as data: grist to the decision-maker s mill. The same framework shapes the bulk of research undertaken within the international social science programs of climate and other global environmental change research, such as the Human Dimensions Programme of the International Social Science Council, which emphasizes stocks, ows and driving forces of change, particularly in relation to land use and the industrial metabolism of society.

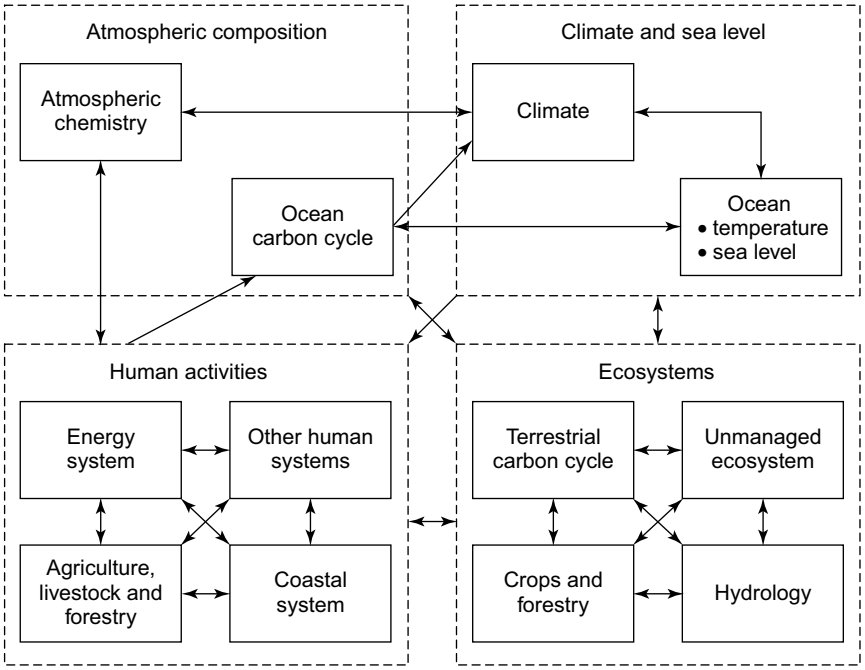


Figure 1 Elements of the climate change problem (*Source:* Watson *et al.*, 1996)

Even though the style of social science emphasized in these programs seems quite compatible with the cyclical framework postulated by the natural sciences, there has been significant concern about the issue of bridging the gap between natural and social sciences, a gap that is invariably taken for granted and associated with much hand wringing. Much is often made of the differences between the two intellectual traditions, exemplified in the distinction between the experimental tradition (including bench or laboratory science) and the study of human nature and the history of ideas.

The comparisons inevitably find the social sciences wanting with respect to the characteristics of theoretical consensus about fundamentals, measurability and phenomena and computability of relationships among them, and precision and accuracy of prediction. The explanations for these differences range from the charitable observation that social behavior is inherently more complex than the behavior of natural systems to the less charitable assertion that the social sciences are younger and, thus, less mature than natural science.

Since Snow's (1959) classic essay on the two cultures, we have become accustomed to the idea that the intellectual landscape bequeathed to us by the Enlightenment is bisected by a major fault line. But, the most significant fault does not, in fact, fall between the natural and the social sciences. Indeed the conceptual and methodological approaches of modeling the behavior of demographic and macroeconomic systems are essentially indistinguishable from those employed to model changes in the atmosphere or in ecosystems. The significant intellectual rupture falls right in the middle of the social sciences, one side of which blends seamlessly with natural science while the other side ultimately merges into the humanities. In debates about global environmental change, this rupture means that some social science research has been used while other research, though relevant and potentially useful, has not been employed in either formulating the issues or addressing them.

THE DESCRIPTIVE APPROACH

The descriptive approach primarily uses inventories and accounting systems for the stocks and flows of people, money, raw materials, commodities, and pollutants. For example, Figure 2 depicts human driving forces, land use, and land cover as discreet units that interact with each other in much the same ways that software modules interact. The descriptive tradition in social science derives from the empirical philosophy underlying natural science descriptions, specifically Newtonian science. Newtonian science is based on the belief that the mind learns everything from what is evident to the senses. English empiricist philosophers, following Locke (1623–1704), postulated that the human mind is a *tabula rasa* (clean slate) upon which the external world is accurately reflected (Locke, 1690). The radical empiricist David Hume (1711–1776) claimed that all knowledge derives from sensory experience. It has become almost a cliche to observe that Hume's philosophy is a Newtonian science of Man. Newtonian thermodynamic determinism is the background of the descriptive tradition in the social sciences.

This approach developed among scholars concerned with wealth, trade, and population in the century following Newton's (1642–1724) formulation of the laws of thermodynamics and Harvey's (1578–1657) description of the circulation of the blood. Indeed, Francois Quesnay (1694–1774), the key figure among the physiocratic economists of prerevolutionary France was also the royal physician. He drew explicitly upon Harvey's physiology in describing the economy. His *Tableau Economique*, published in 1758, is generally considered to be the first schematic description of an entire economy. The physiocratic models departed from previous thinking about the economy in that they did not invoke human choice, desire, and intention to explain the behavior of the economic system. Instead, attention focused on the characteristic operations of the system, i.e., what can be observed and measured.

A generation later, Ricardo's (1772–1823) model of the economy also elides human intentions (Gudeman, 1986, 45). His concept of the individual is not explicitly elaborated, but the actors in his model, whether landowners, capitalists or laborers, lack intentionality. Ricardo's view of the human is exemplified by his difference with Malthus (1766–1834) over Say's Law, according to which supply creates its own demand. For Malthus, demand is not a mechanistic response to supply, because it is made up of both the power and the will to purchase. But Ricardo (1814, VI : 133), foreshadowing the declaration of the apocryphal Madison Avenue advertising executive, If we make it, they'll buy it! maintained that the will is very seldom wanting where the power exists . Variable tastes and changing desires were not part of Ricardo's image of the human. Ricardo's model moved even further away from concern with human intentions than those of the physiocrats by his use of a mathematical schema, which emphasizes form at the expense of content. After Ricardo, economic models took on the mechanical characteristics of closed thermodynamic systems or mass-balance equations.

Following in the steps of the physiocrats and Ricardo, the descriptive approach relies heavily on the ability to count, weigh, and measure things such as numbers of people, industrial productivity, or degrees of wellbeing. This aspect has led some commentators (e.g., Robinson and Timmerman, 1993) to describe the descriptive method as a physical flows approach. However, less tangible elements, such as beliefs and values, can also be represented in the descriptive method as counts of the number of people

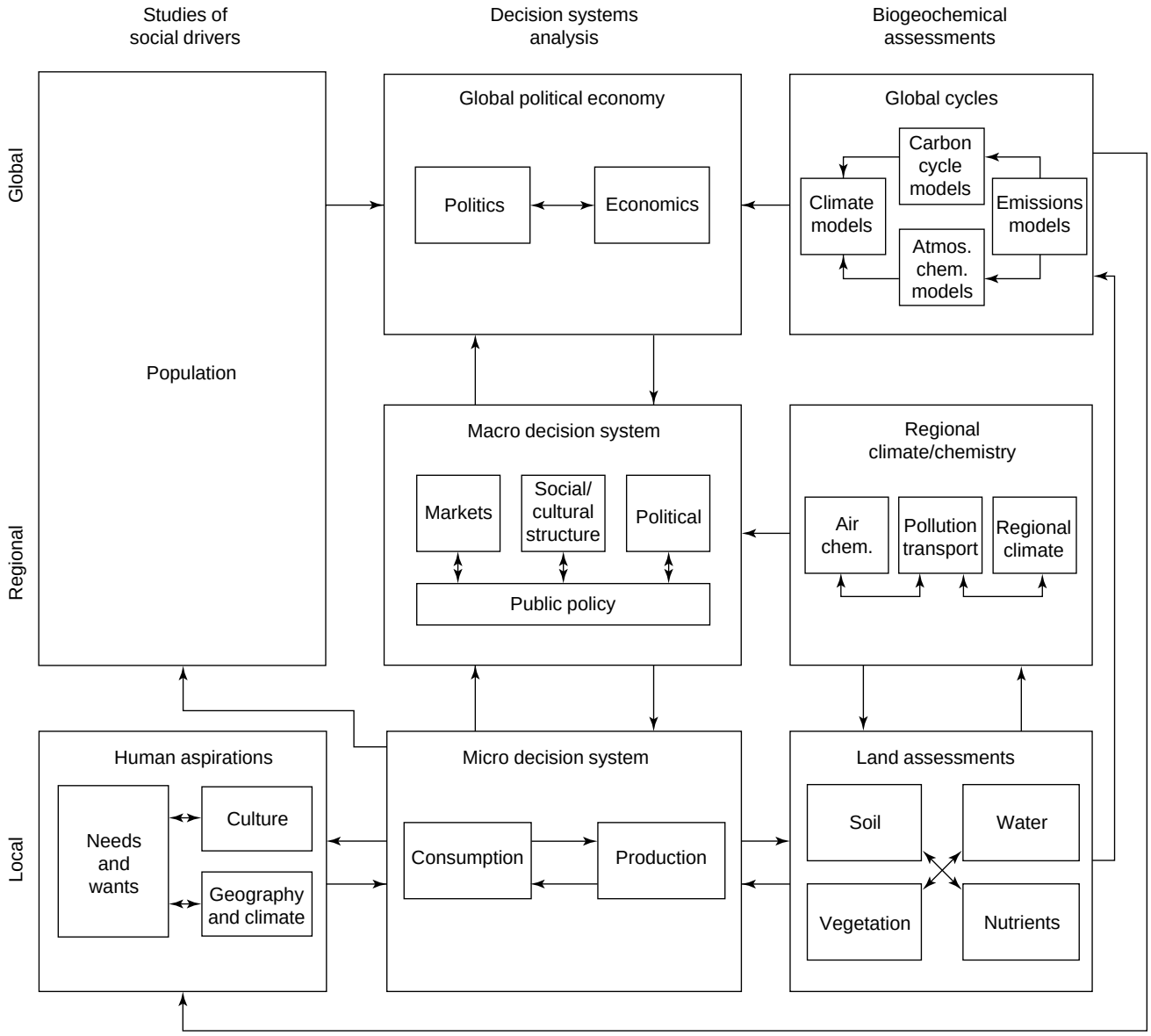


Figure 2 The stocks and flows approach of the descriptive paradigm (*Source:* Rayner *et al.,* 1994)

professing them, e.g., mass survey research or analyses of voting behavior. Knowledge and values converted to quantities in this way can therefore be accounted for as stocks and flows.

The descriptive approach has several advantages. The results of descriptive social science investigations are able to draw on the legitimacy of the natural sciences, by mirroring the scientific method and the thermodynamic paradigm. Counting and measuring can provide insight into the scope and severity of a problem or locate probable causes among many candidates. Descriptive analysis can compare various pathways from the standpoint of economic efficiency. Its simplifying assumptions and equations can lay out large problems and render them more tractable. And, finally, because the descriptive approach deals primarily with tangibles, it provides a basis for control; e.g., counting tons of fossil fuel used can provide a basis for a tax scheme that will limit the number of tons burned. This approach is typical of economics and demography as well as much of (American) quantitative sociology, behavioral psychology and quantitative political science.

The descriptive approach in social science and its pre-eminent place in policy analysis arose as part of the larger intellectual effort to devise technologies that could enhance the rational management of nature and society by the emerging modern state. Headrick (1990, 59) points out that, As early as the seventeenth century ••• (governments) encouraged the increase in knowledge of and power over

nature . Headrick lists examples, such as the patent system introduced in England in 1623 and the founding of scientific societies throughout the 17th and 18th centuries. This was the prelude to a massive increase in the rate of global resource extraction. He continues, Insofar as technology is knowledge of nature and how to manipulate it for human ends, it follows that countries with the most information are also the most able to develop their economies and transform the land (Headrick, 1990, 59). The social science technologies of the same period included the new field of statistics, which emerged in the mid-18th century as the practice of inventorying the resources of the state. Merchant (1990, 682) eloquently describes the contribution of the descriptive paradigm to the development of modern political, commercial, and industrial power:

> ••• When mapped by explorers and cartographers, catalogued and inventoried by militarists and computer scientists, ••• (nature) is controlled by an eye of power and subject to unlimited surveillance. Foucault's model of the panopticon of Jeremy Bentham, in which an entire prison can be surveyed from a single central tower, translates to the concept of a cultural overseer.

This stance has been variously described as the separation of the subject from the object, Man/Nature dualism, and instrumentalism. As Latour (1986, 29) observes:

> ••• we can work on paper with rulers and numbers, but still manipulate three-dimensional objects out there •••. Distant or foreign places and times can be gathered in one place in a form that allows all the places and times to be presented at once.

Thus, the descriptive tradition in both natural and social science separates us from nature and society while providing us with a significant level of control over them. For example, this power is highly concentrated in the computer models that we currently use for tracking materials flows and their consequences in both environmental science and macroeconomics, for example, general circulation models of climate and general computable equilibrium models of the economy.

Like all sources of power, this one comes with a price tag; separation of knowledge from the knower. The development of the modern industrial state relied on strategies of control, and modes of mapping, tabulation, recording, classification, and demarcation. These increasing levels of control over nature meant also increasing separation from experiences that should be part of the decision-making process. Faced with disembodied parcels of descriptive facts, decision-makers may become more and more isolated from experience of how myriad decisions transform production and consumption at the level of the community, the firm, and the family. For example, in debates about the cost effectiveness of climate change mitigation measures, the hardships of poor people in environmentally marginal areas (deserts, coasts) may be hidden in world gross domestic product (GDP) statistics. For climate change policies, this has been a source of the disjoint between cost effective strategies and effective implementation.

Furthermore, the descriptive method is epistemologically realist. The meaning of belief and value structures and the ways in which they are constituted, reproduced, transformed, and translated into action in social life are treated as unproblematic in this approach. Inquiry into these is firmly located outside of the descriptive method within what we refer to as the interpretive method.

THE INTERPRETIVE APPROACH

In contrast, the interpretive method focuses on understanding the meaning that human agents create in the conduct of social life, upon which they build their understanding of their world, and through which they seek to act upon that world. Thus the interpretive method focuses on the nature of experience, the structure of perceptions, the recognition of interests, and the development of frameworks for collective action.

Because these factors often elude quantification, the interpretive method has often been characterized as qualitative social science, contrasted with quantitative approaches that we describe as the descriptive method. However, we eschew this particular distinction between the two traditions for three reasons:

1. Interpretive social science need not be qualitative (some practitioners having devised sophisticated methods to quantify phenomena that are not as obviously susceptible to counting as heads or barrels).
2. The quantities in descriptive analyses are, in any case, often a numerical representation of a qualitative judgment on the part of the analyst rather than an empirical quantity.
3. The distinction between quantitative and qualitative measurement is actually rather fuzzy (for example, dresses may be sold as small, medium, and large or as sizes 2–16).

The distinctions between descriptive and interpretive approaches can be rather fuzzy, as well. In fact, within the interpretive approach a researcher can take data of all kinds, quantitative and qualitative, and make coherent meanings that can be used to devise workable arrangements, to effectively address problems, and to set up usable process frameworks. For example, anecdotal and case study evidence can be used to supplement or supplant statistical data. And, of course, the descriptive approach usually includes some interpretive analysis, e.g., in the discussion following the presentation of results and in the conclusions and recommendations of technical reports.

The interpretive tradition in social science emphasizes the essentially social character of the operation of the

human mind. The interpretive method is strongest in history, cultural anthropology, and (European) qualitative sociology. Interpretive analyses of various kinds of global environmental change have addressed the framing of the problem as well as issues of stakeholder involvement, sociocultural values, the nature and production of knowledge, and policy implementation (research and development investment, technology selection and diffusion, etc.).

The interpretive tradition derives from philosophical discussions about how human beings acquire knowledge and form (or are formed into) social structures. Although the interpretive tradition can be traced back to classical Greece, the modern interpretive tradition derives from Immanuel Kant s (1724 1804) disagreement with the English empiricists. Kant speci cally rejected the radical empiricist claims of Hume, that all knowledge derives from sensory experience, insisting instead that reason was itself a capacity of the human mind that shaped experience. Thus, in contrast with the descriptive approach, the interpretive tradition has always distinguished itself by its emphasis on human choice and intentionality. The mind s structuring of sensory experience derives from three sources: the physiology of the human nervous system, the unique historical experience of the individual, and the social and cultural categories and rules that are provided by language and by membership in a family and a larger community.

> Kant introduces a new conception of knowledge. Knowledge does indeed have as a source the Humean element of impressions, the sensory element in which the mind is passive •••. But Kant continues, there is another element in our knowledge, which is derived not from sensory experience •••. The second element comes from the mind itself. The human mind is not a blank tablet or an empty cupboard as the empiricists Locke and Hume claim. It is equipped with its own pure concepts by means of which it organizes the ux of sensory impressions into substances, qualities, and quantities, and into cause and effects.
>
> (Lavine, 1984, 193 194)

However, while seeking to avoid imposing a deterministic framework on the study of human choice, the interpretive social sciences, exempli ed by the philosopher and historian Wilhelm Dilthey (1833 1911) retained their commitment to provide systematic understanding of its manifestations.

> Dilthey wanted to uphold the humanist component of historical science without compromising ideals of rigor and objectivity. He thought of interpretation as a disciplined quest for orderly connections in an observable world; a way of doing science, not of escaping science. Yet he often expressed the difference between the human and natural sciences by drawing a sharp distinction between understanding and explanation: We explain nature, we understand mind .
>
> (Selznick, 1992, 76)

One of the primary foci of the interpretive method is the understanding of how humans draw meanings out of the data presented to them, emphasizing the sociocultural and institutional (including political) processes involved. The systematic study of such processes yields understanding that can help us make critical choices, knowing what assumptions and decision elements underlie those choices.

As Ryle (1949) suggested, thinking does not take place in the head, but all around us. What we think with is not a private metaphysical mind, but words, pictures, gestures, actions, and both natural and manufactured objects. Indeed we assign symbolic meaning so as to impose some sort of order and coherence on the stream of events. In so doing, we sift and lter our sensations of the world. Contemporary studies on the physiology of the mind (Edelman, 1994; Damasio, 1994) reinforce the physical basis for thinking that can be conjoined with individualistic and social/cultural theories of knowledge and action.

> In the process of making the whole business comprehensible, some perceptions are admitted, some rejected, and others combined or broken down. If we did not lter experiences in this way or make use of public symbols for organizing perceptions and communicating them to others, then we would be likely overwhelmed by the variety of possible interpretations that could be assigned to events. We would have to abandon intellect and discourse and thereby be forced, like the lower animals, to rely on instinct. Mankind would be reduced, as Geertz (1973) has observed, to mental basket cases .
>
> (Gross and Rayner, 1985, 3 4)

Different people, of course, organize and assign meanings to their experiences differently. Though we have the same general brain structures and turn our eyes to the same world, there are variations in what we see. As Kuhn (1970, 193) says:

> research shows that very different stimuli can produce the same sensations; that the same stimulus can produce very different sensations; and, nally, that the route from stimulus to sensation is in part conditioned by education. Individuals raised in different societies behave on some occasions as though they saw different things.

The interpretive method insists that no standpoint exists outside of history or of society from which either can be independently observed. All organized knowledge, by de - nition, depends upon socially construed conventions for its organization. In this sense, the interpretive method treats knowledge as having no objective existence. However, most interpretive social scientists use comparison of interpretations to generalize human behaviors.

STYLE AND SCALE

The descriptive approach derives much of its strength from simplifying assumptions about complex behaviors. It tends to assume homogeneity (or at least a tendency to homogeneity) of human responses to price signals. It

therefore is capable of high levels of aggregation in its analyses.

In contrast, the interpretive approach emphasizes variation in human motivation. It explores complexity, sometimes to the point of resisting simplification on principle, regardless of how useful or appropriate it might seem. Indeed, attempts to simplify within interpretive models are often criticized as stereotyping. The interpretive approach therefore tends to produce highly disaggregated analyses.

As a result of their different approaches to aggregation, the two kinds of social science tend to be comfortable at different extremes of scale. The descriptive approach tends towards large-scale analyses, with the nation state (which grew up with the descriptive approach) often accepted unquestioningly as the appropriate level for many descriptive analyses. The descriptive approach may be more suited to initial analysis of problems at the global scale; the stance of the observer and tendency toward aggregation can facilitate analysis at that level.

The interpretive approach, on the other hand, tends to focus on the local scale of face-to-face discourse. Here, the human individual is often regarded as being the natural, self-evident unit of analysis. It is often difficult to demonstrate the relevance of social science at this scale to global-level research agendas.

However, the case is frequently made Not only that the human activities that drive and mitigate environmental change vary significantly by region or place, but also that historical or local contextual factors are so influential that understanding must be grounded in the specifics of the case (Turner *et al.*, 1990, 18). In the last analysis, the activities that drive land-use change or anthropogenic greenhouse gas emissions are always local, and whatever the impacts of global environmental change on human populations, they will be experienced at the local level.

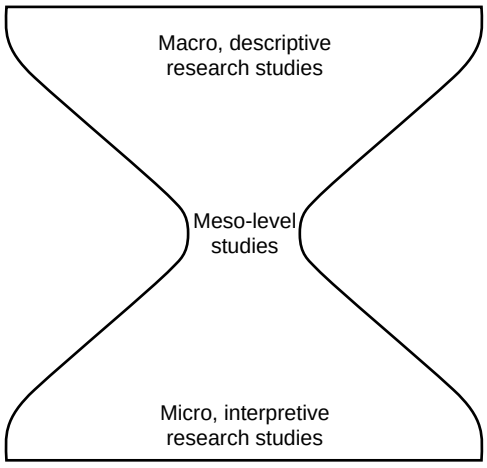


Figure 3 Social science research concentrated at macro and micro scales

The issue for the social sciences is not merely the assertion of the importance of local level analysis. As a result of the polarization of the descriptive and interpretive methods, research areas in the social sciences are shaped like an hourglass (Figure 3). That is, theories, methods, and data are underdeveloped at the middle scales between the individual and the nation state.

Furthermore, correspondence between the macro and micro scales is often treated in a simplistic way. The nation state or the firm is often treated as if it were one large individual, with a unitary intelligence and decision-making capacity. Or the resources of real individuals may be represented merely by dividing total resource by the number of people, resulting in the creation of a mere *per capita*; the anonymous average person from whom *per capita* statistics are derived.

> The image is of a male person, sometimes a homunculus inside each of us, sometimes a gigantic system incorporating the whole of society or the world. Our pervasive microcosm to whose outlines all our explanations are drawn to fit is a quintessential stranger; he has no family or friends, no personal history, his emotions are not like ours, we don't understand his language, still less his purposes.
>
> (Douglas and Ney, 1998)

For example, in the field of energy modeling, bottom-up analysts project individual behavior onto the national scale (one large individual) while top-down analysts assume that the behavior of individuals reflects that of the national population in proportion to their numbers (*per capita*). Although the names of the modeling approaches seem to indicate macro and micro scales, in fact neither approach is based on a theory of micro–macro scale articulation; both are descriptive approaches that assume undifferentiated individuals.

Local-level analyses, on the other hand, often focus on individuals and small, specialized institutions without regard to their generalizability (more studies are needed to determine whether similar results can be found in neighboring villages); or their generalizability is taken for granted, i.e., village level solutions are recommended for settings the world over.

Linking the local and the global is frequently cited as one of the most challenging aspects of climate and other manifestations of global change. Few attempts have been made to validate studies at one end of the scale with those at the other, and the research studies are sparse at the middle levels, especially those that articulate connections among individuals, groups of various sizes and cultural backgrounds, and global level players in global environmental change issues. From the perspective of the social sciences, meeting this imperative requires, as a first step, linking the interpretive and the descriptive approaches, using the research approaches and strengths of each, and expanding the reach of study, as appropriate, into various middle

scales, with appropriate integration of overlap. However, as discussed in the next section, the viewpoints cannot and should not be merged; each makes unique contributions to scientific knowledge.

STYLE AND STANDPOINT

Throughout the social sciences, there is poor understanding of the articulation of human behavior at the local level to the behavior of the global social and economic system. Furthermore, available knowledge of local behaviors has not yet penetrated very deeply into the global environmental change assessment and research agendas of the IPCC, the IGBP and the Human Dimensions Programme of the International Social Science Council.

Linking the local and the global cannot be achieved simply by increasing the scale and quantifiability of interpretive analysis to meet a more thickly textured descriptive analysis as it attempts to accommodate lower levels of aggregation. The gap between the two approaches is not merely spatial but raises fundamental issues of what kinds and sources of knowledge we value as analysts. Contrasting 16th century iconography with contemporary satellite photographs, Ingold (1993) illustrates the situation by tracing the change in the standpoint of human inquiry from the Enlightenment to the present (Figure 4). In Maffei's *Scala Naturale* of 1564, the scholar is shown at the center of the environment consisting of 14 concentric spheres envisaged to form a giant stairway, the ascent of which, step by step, affords more comprehensive knowledge of the world through experience within it. In modern satellite imagery, the scholar experiencing the world from within, is displaced by an observer viewing the world from without.

This descriptive standpoint favors observation over experience. Local (interpretive) knowledge, originating in experience, is downgraded as partial, parochial, and ultimately unreliable. Global knowledge is treated as universal, total, and real.

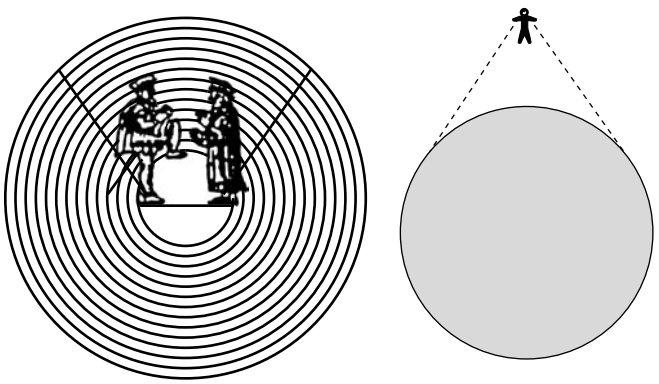


Figure 4 Local and global standpoints. (Reproduced by permission of Routledge in Ingold, 1993)

Against this view, Ingold argues that:

> The difference between them (local and global perspectives) is not one of hierarchical degree, in scale or comprehensiveness, but one of kind. In other words, the local is not a more limited or narrowly focused apprehension than the global, it is one that rests on an altogether different *mode* of apprehension; one based on an active perceptual engagement with components of the dwelt-in world, in the practical business of life, rather than on the detached, disinterested observation of a world apart.
>
> (Ingold, 1993, 40–41)

If the two standpoints are so different in kind, how can a potential for complementarity be applied to a global environmental change issue? The descriptive approach provides an absolutely vital link between the social sciences and natural science analyses of global environmental change. Collaboration between natural science and the descriptive approaches has revealed much about what would happen under various scenarios, but it does not help in choosing among the probable or feasible scenarios. To understand these issues, we also require interpretive approaches that can introduce qualitative and value-related parameters. For example, both approaches are highly relevant to current negotiations about how to implement the Framework Convention on Climate Change (FCCC).

The FCCC states that The ultimate objective ••• is to achieve, in accordance with the relevant provisions of the Convention, stabilization of greenhouse gas concentrations in the atmosphere that would prevent dangerous anthropogenic interference with the climate system . Further, policies and measures to deal with climate change should be cost effective so as to ensure global benefits at the lowest possible cost . A particular ceiling for greenhouse gas concentrations can be maintained in a variety of ways. Some will be more costly than others. Economic analysis in the descriptive tradition can show that it is more expensive for all nations to stabilize emissions individually than for the world to stabilize emissions jointly. But some regions might be better off under an individual stabilization scheme. Descriptive analysis can identify conditions under which high population countries would have an incentive to drop out of an international agreement to control fossil fuel carbon emissions in the post-2000 period, even if emissions rights were allocated on an equal *per capita* basis. Descriptive analysis can also highlight the need to construct dynamic international agreements, capable of being modified as global changes occur and societies evolve. But the descriptive approach cannot successfully explore social and ethical dimensions of different political options and has difficulty dealing with tradeoffs between equity and efficiency.

Redressing the disconnection in style, scale, and standpoint within the social sciences will at least require bringing together the descriptive and interpretive methods. Using and broadening established mechanisms, researchers can adopt

either a *multidisciplinary* or an *interdisciplinary* approach. In the former, scientists from various fields work together on a problem that has been defined within the framework of one particular discipline. A truly interdisciplinary approach involves people from different fields working on a problem that they have defined together in a way that it cannot be defined from within any single discipline. Interdisciplinary research often yields insights not otherwise attainable.

Perhaps the strongest reason for exploiting the complementarity between the two approaches is the nature of global environmental issues. The dimensions of global environmental change simply cannot be adequately addressed without using both approaches. Furthermore, strong linkages between the two can provide a further bridge to rich resources within the humanities that can also help to deal with fundamental issues such as social justice.

For example, the development of eco-feminism (*see* Ecofeminism, Volume 5) represents just one linkage between research in the humanities and social sciences, centering on an environmental debate. Arguments about the right relation between humans and non-human nature often have their basis in spiritual and moral values which rely on evidence from theology, moral philosophy, and literary imagery as much, if not more than, scientific data. Eco-feminism, which has been linked with neo-paganism (specifically, goddess worship), draws its rationale from a sociological critique of the twin oppressions of the domination of women and nature (Merchant, 1992, 185). Sociological, political, and image-rich arguments are blended, for example in *Woman and Nature: The Roaring within Her* (Griffin, 1978). Drawing on such diverse notions as the reconstitution of the subject and elemental divinity, Griffin and others seek to demonstrate the importance of women (and women's work) and non-human nature. Part of the pro-feminist argument revolves around the idea that women, as the carrier of children, are the first environment and thus have a special link with other nurturing systems in nature. Another part of the argument relies heavily on statistics about the type and value of women's work in agricultural and other economic systems.

LINKING SCIENCE AND HUMAN VALUES

At the human level, global environmental change is ultimately an ethical issue. Just as technical analyses and arguments prove insufficient to persuade communities to use imported technologies, human decisions about the threat of global environmental change are not merely technical. They are decisions about equity, what is fair in our relationships with each other and about natural ethics, what is right with respect to our relationships with nature.

While the analysis and description of ethical systems as social phenomena belong in the social sciences, the normative exercise of ethical reasoning is traditionally the province of philosophy and theology. Thus, bridging the gap between the descriptive and interpretive traditions in social science (and even exploiting the commonalties between interpretive approaches in the social sciences and the humanities) is an imperative for effectively linking the practice of scientific and moral reasoning in confronting global environmental change.

A further practical imperative to integrate the descriptive and interpretive methods is the need to understand human choice in social change as well as aggregating market behavior and mapping demographic change. Following in Ricardo's footsteps, the descriptive method assumes continuity of human preferences and consistency of human behavior over time. The descriptive models do not seek to explain the values that humans hold. They therefore cannot anticipate sudden ruptures in social behavior resulting from changes in values and the institutional arrangements that embody those values, or to assess the potential for changing human motivations through political or other forms of social intervention. Because a vast array of human behavior involved is effectively compressed into drivers and responses, the descriptive approach encourages the adoption of behaviorist kinds of explanations for the actions of social systems and the individuals within them. Hence, policy prescriptions based solely on the descriptive paradigm are limited to instrumental tinkering with technology and prices.

But, social systems involve human choice; unlike a billiard ball shot across a table or an electron orbiting the nucleus of an atom, a human being has the ability to make conscious decisions about the directions in which he or she is moving and at what speed. Although social decision-makers are often constrained by their own paradigms and the initial conditions of the problem they are attempting to resolve, the existence of human choice means that their actions cannot be accounted for in purely deterministic terms.

VALUES IN SCIENCE

Human choice also raises a methodological problem for the construction and operation of descriptive social science. Science, whether natural or social, is conducted by humans and is thus itself subject to human choice. Humans choose what to study and what to ignore, what methods to use in their analysis, and what criteria to apply in determining the validity of the data gathered. For instance, do we study energy conservation or geoengineering? Do we assess quality of life or longevity? Must we have an untreated control population or can we learn from studies that treat all subjects?

In making these choices, researchers and decision-makers inevitably make value judgments. However, when the value-based assumptions embedded in the theory or model

disappear into the background, they come to be seen as natural and are uncritically accepted, often without any conscious thought about either their presence or their implications. Indeed, the more that scientists reduce human choice within descriptive models, the more that they interfere as agents. Excluding human agency from applied social science research is itself a powerful and constraining assertion of the researcher s own choice upon society. Since we humans cannot stand outside of history and society to observe and describe them, we must use the insights of interpretive social science to make the underpinnings of our analysis as explicit as we can.

Interpretive social science brings to the conduct of scienti c inquiry an awareness of the impact of human choice, by introducing *re exivity* into the research and policy making processes. Re exivity is the self-conscious examination of the implicit assumptions that are inevitably embedded in any analytical approach. In the descriptive approach, assumptions may be systematically laid out but are typically unquestioned within the research study. This is consistent with the outside observer stance of the descriptive researcher. If, however, these assumptions are made explicit, the researcher has an opportunity to question them, rather than taking them as given. The interpretive contribution of keeping assumptions visible adds meaning to the research results by providing clear boundaries and caveats. Thus, if research in the two paradigms can be integrated, a more complete analysis can be accomplished.

The descriptive approach only works effectively when scientists take their assumptions for granted, so that they can structure experiments that will have meaning within those assumptions. Acting scienti cally means acting on the assumption of a determinate nature waiting to be described by a neutral observation language (Fish, 1994). In the natural sciences, Kuhn (1970, 163 164) notes:

> that once the reception of a common paradigm has freed the scienti c community from the need constantly to re-examine its rst principles, the members of that community can concentrate exclusively upon the subtlest and most esoteric of the phenomena that concern it. Inevitably, that does increase both the effectiveness and the ef ciency with which the group as a whole solves new problems.

Humans must reproduce and extend cultural conventions that are unquestioned in everyday life. As Alfred Whitehead has said in another connection,

> it is a profoundly erroneous truism ••• that we should cultivate the habit of thinking what we are doing. The precise opposite is the case. Civilization advances by extending the number of important operations, which we can perform without thinking about them. This is of profound signi cance •••. We have developed these practices and institutions by building upon habits and institutions which have proven successful in their own sphere and which in turn become the foundation of the civilization we have built up.
>
> (Hayek, quoted in Koford and Miller, 1991, 22)

However, while in daily life the ability to do more things with less deliberation may be a sign of technical and social development, it also increases the danger of unwelcome surprise. The interpretive social sciences have the capability to make the implicit explicit, to provide society with the tools of re exivity and to enhance society s resilience to shocks.

Combining the re exivity of the interpretive with the instrumental knowledge of the descriptive method seems an obvious course for improving our understanding of the human dimensions of global environmental change and use in policy making. A re exive social science drawing effectively on both the descriptive and interpretive methods could help identify the multiplicity of characteristics needed to achieve successful solutions, both immediate and longer term, associated with varying alternative global environmental management strategies.

SOCIAL SCIENCE RESEARCH AREAS IN GLOBAL ENVIRONMENTAL CHANGE

Global environmental change may not be the deciding factor in whether humanity as a whole ourishes or declines. The resilience of human institutions and their ability to monitor and adapt to changing conditions seems to be more important. Therefore, the social sciences have much to contribute towards illuminating issues such as pollution, climate change, and biodiversity. This section provides a brief discussion of issues taken up by social scientists in climate change research (Rayner and Malone, 1998) to indicate the breadth of social science contributions to global environmental change research.

Insight into how scientists choose to study global environmental change and how they form a scienti c consensus demonstrates the basis on which issues are taken up: and on what basis they could be taken off the agenda. To avoid promoting unrealistic public and policy-maker expectations of scienti c prediction and control over nature, the scienti c community thus must work for acceptance and public authority through patiently constructing communities of belief that provide legitimacy through inclusion, participation, and transparency.

Similarly, knowing why and how people decide that global environmental change is worthy of attention reveals the variety and roots of human values and concerns. Forms of social solidarity appear to be a crucial variable in shaping people s sensitivities to risk in general and threats to environmental and climate change in particular. Furthermore, equity is a crucially important issue, but at least three equity principles can be applied: proportionality, priority, and parity (Thompson and Rayner, 1998). Recognizing the broader debates inherent in environmental change issues (development is another) may allow policy makers to address a number of important problems concurrently.

Sorting out causal explanations and prescriptions for change puts the debates about environmental change in a new light and provides new ways of finding common concerns and areas of agreement. For example, in climate change debates different groups cite population, consumption, or the market as the villain; since all of these diagnoses are partial explanations, debates among groups can lead to stalemate. When confronted by such a dilemma, social scientists can examine alternative framings of the problem and solution that will be helpful in resolving the issues.

The rational-choice model of decision-making is an inadequate guide in global environmental change issues. Social science research can identify and delineate alternative decision-making frameworks that include quality and meaning issues. A broadly based framework may help to reveal win-wins that have been overlooked.

Resource management practices (for land, water, energy, etc.) are the immediate cause of environmental degradation. Social science research can describe the human activities that give rise to environmental concerns, identify possible mitigation actions, indicate where adaptations will be necessary, and illuminate how institutional and cultural structures and abilities to change will both constrain and open up possibilities to make and implement policy.

Social science research demonstrates that the process through which choices articulate across scales is not a linear mechanism that consistently produces the most rational alternative at the next highest scale. Rather, it is a social as well as a knowledge process that requires a high level of trust and agreement (including standardization of methods and results) to gain recognition at another scale. Research that describes choice processes is important; equally important is research that describes processes by which individuals can be persuaded to conform with new normative requirements of corporations and governments, as implemented by the decision-makers who are their officials.

ENVIRONMENTAL JUSTICE: UTILITY- AND RIGHTS-BASED APPROACHES

The descriptive and interpretive approaches embody quite different normative imperatives for decision-makers. The tendency of the descriptive paradigm towards high levels of aggregation gives rise to a top-down decision-making rationality. The quantitative aspect of the paradigm leads to an essentially utilitarian perspective on decision-making. That is, the practice of inventorying the stocks and flows of goods and bads creates the conditions for a decision framework based on a technique for calculating societal happiness measured by their distribution (Bentham, 1748–1832). The rise of *utilitarianism* as an explicit decision-making principle in the 18th and 19th centuries paralleled the development of systems of national accounting and statistics designed to assess and facilitate the development of industrial capitalism but which also held out the prospect of civil or corporate leaders being able to rationally assess the impact of decisions on the wellbeing of the statistical population.

It is but a short step to turn the possibility of such a means for calculating what would contribute to the greatest happiness of the greatest number into the imperative that one should pursue that goal. The solution that provides the greatest happiness of the greatest number also must be an efficient solution, since any departure from efficiency, also by definition, reduces the amount of good available for distribution. A bias toward the value of efficiency is inherent in the methods of policy analysis and utilitarianism has been the ideological position most forthrightly incorporating this standard as a central value (Heineman *et al.*, 1990, 38). However, it can be seen that efficiency is not merely a technical issue or an indication of rational behavior within utilitarianism, but is also an intrinsically moral imperative that arises from the descriptive paradigm itself.

Thus, the utility principle domesticated moral diversity for decision-making authorities by offering the capability to measure and monitor the stocks and flows of societal good, the usual proxy for good being wealth in some form. By the same process, decision-maker awareness of alternative ethical considerations was systematically attenuated. The imperative to provide for societal good at the highest level of aggregation provides no guidance for securing the happiness of minorities and individuals, even of those individuals in the happy majority.

> The guiding criterion for policy is the greatest good for society, quantitatively defined. But contemporary utilitarians, primarily economists and theorists of public choice, like Bentham, still have no principle for distributing this social good according to manifest principles of equity.
>
> (Heineman *et al.*, 1990, 40)

In contrast, the tendency of the interpretive tradition to focus on the individual rather than the nation state directs the attention of scholars working in that paradigm to the particular circumstances of decision-making rather than to the aggregate outcome. Attention to disaggregated particulars, combined with the dominance of methodological individualism, articulates smoothly with an orientation towards bottom-up, rather than top-down decision-making. The insights of the interpretive paradigm are more likely to be of interest to those who espouse libertarian or egalitarian ethics, emphasizing the rights of the individual or the minority in the face of majority preferences, as, for example, in Kantian ethics and the Jeffersonian political tradition in the US. Hence, the interpretive paradigm is often associated with a critical stance towards the *status quo*, and is often equated with and labeled as the critical tradition in social science, which marginalizes its significance by definition.

Increasing insight into the diversity of motives, values, and preferences of individuals, however, actually tends to frustrate utilitarian social accountancy, which depends on blending out such distinctions in the process of aggregation. It is hardly surprising therefore, that the insights of the interpretive paradigm are not merely considered irrelevant to, but actually have to be excluded from utilitarian decision-making in order to preserve the rationality and legitimacy of the utility principle.

Hence, the distinction between two social science paradigms is not merely an artifact for the history of ideas or a scholarly distinction of mere academic interest. It actually lies at the heart of the crisis of governance that pervades the local, national, and global communities at the start of the 21st century. That is, the tension is very real between interdependence and independence, between pursuit of the greatest happiness of the greatest number, and the assertion of individual, local, or ethnic rights that ought not to be violated even at the expense of the aggregate good. Whereas Kant's principle that every person is to be regarded as an end in him or herself is generally recognized as a form of the doctrine of human rights, Bentham dismissed the concept of rights as plain nonsense and the imprescriptible rights of man as nonsense on stilts (Russell, 1946, 742). He denounced the articles of the *Declaration des droits de l'homme* as falling into three classes: (1) Those that are unintelligible; (2) those that are false; and (3) those that are both (Russell, 1946, 742).

Similar vituperation for the social inefficiency of rights-based ethics is not unheard of among contemporary utilitarians. For example, in response to proposals by sociologist Robert Bullard that current inequities in the distribution of environmental burdens on minorities and the poor should be addressed on an environmental rights basis, rather than according to risk-based criteria, economist Albert L. Nichols responded:

> This framework has considerable popular appeal, but it ultimately is counterproductive from the perspectives of both society as a whole and even the specific groups it tries to champion. Moreover, it provides little practical guidance to environmental decision-makers trying to set priorities. ••• Bullard's proposed environmental justice framework makes continued inequities in protection more likely •••.
>
> (Nichols, 1994, 267)

Bullard replied that his proposals

> are no more regressive than the initiatives taken in the 19th century in eliminating slavery and Jim Crow measures in the US. This argument was a sound one in the 1860s when the 13th Amendment of the Constitution was passed despite the opposition of proslavery advocates, who posited that the new law would create unemployment (slaves had a zero unemployment rate), drive up wages (slaves worked for free), and inflict undue hardship on the plantation economy (loss of absolute control of privately owned human property).
>
> (Bullard, 1994, 260)

Clearly these are not merely technical arguments about the best way to clean up the environment. Similar clashes between the utilitarian- and rights-based views arise over the projected costs of climate change. In response to damage estimates that climate change will result in a decline of global productivity of less than 1% over the course of this century, utilitarians have expressed the view that only low-cost mitigation measures can be justified. On the other hand, those who espouse a rights-based approach point out that even less than 1% of global productivity over 100 years may translate into considerable suffering and premature death for millions of poor people in vulnerable regions of the less industrialized world.

The issues of risk and justice provide good loci for probing inextricable links between analytic methodologies and underlying social commitments to the Kantian individual in him or herself or the Benthamite aggregate good. Often such issues seem to be intractable when descriptive research and analysis provide the only framework for addressing them.

Typically, policy makers do approach risk characterization as a technical issue: exposure pathways, dose, response, etc. However, the technical information is unlikely to influence citizens living in an area where environmental risk is present. If the science–policy relationship remains squarely in the descriptive mindset, the response to citizen objections is likely to be collecting more technical data; the inclusion of values and worldviews will not be an option. However, if the science–policy relationship is collaborative and includes interpretive data, science will be more integrated into the policy context, more contextual and openly value laden, less oriented to mastery over natural and social processes, and more accessible to the public at large (Robinson, 1992b, 249). To the extent that scientists, policy makers, and the public can learn about each other's positions and preferences, the solution space for a problem of risk analysis becomes larger and a solution more possible.

USE OF SOCIAL SCIENCE RESEARCH

In the same way that collaboration is needed between researchers who use descriptive and interpretive approaches, researchers and policy makers need to forge more collaborative relationships. An important element of this collaboration must be the use of both approaches and multiple tools to illuminate many facets of the issues at hand. Current arrangements and policy discourses tend to favor the descriptive approach, although this was not always the case. The policy analysts of the Reformation were the hermeneutic (interpretive) religious scholars who advised both Catholic and Protestant kings and princes about all aspects of policy based on the interpretation of holy texts and secular precedents.

The descriptive approach to policy making is usually considered to be more appealing to policy makers because of its immediate potential for instrumental use. Quantitative analyses of responsiveness to tax rates or the effectiveness of regulation can, in principle, be directly translated into a set of policy choices about whether to implement a carbon tax or appliance efficiency standards, and even at what level taxes or standards should be set. In other words, descriptive research usually manages to come up with a bottom line. Generally, interpretive social science tends to be dismissed by policy makers and their social science advisors as lacking this potential to provide practical guidance. How can social scientists provide powerful, practical research findings sufficiently mindful of research limitations? Can we replace or redefine the model of scientific truth being directly used by policy makers? And what would such a research–policy relationship look like?

First, we can recognize that our model does not actually mirror reality, i.e., policy makers do not frequently make direct use of descriptive research's bottom lines. Empirical research in the US (where the instrumentalist emphasis on the bottom line is probably most strongly emphasized) shows that, despite a generally positive attitude to such analysis, it is seldom acted upon in any directly identifiable fashion (Rich, 1977; Weiss and Bucuvalas, 1980; Whiteman, 1985; House and Shull, 1988).

These studies indicate that the actual impact of descriptive policy analysis is much more diffuse. Indeed, Weiss (1982) suggests that the real usefulness of policy analysis may lie in enlightenment rather than instrumental purposes, less as a tool for solving specific problems than as a way of orientating people towards issues. And much of this is not deliberate, direct, and targeted, but a result of long-term percolation of social science concepts, theories, and findings into the climate of informed opinion (Weiss, 1982, 534).

The apparent concreteness of descriptive social science information in practice seldom makes it any more usable by decision-makers for obtaining enlightenment than the results of interpretive studies. Both may contribute equally. However, once a course of action is chosen, descriptive data may be more frequently invoked for the purposes of rationalization and persuasion (Patton, 1978; Whiteman, 1985).

Speaking (scientific) truth to (policy making) power is thus revealed as a coercive illusion worthy of Frank Baum's Wizard of Oz. The answer to what will we do if we abandon speaking truth to power? is that we will do what we have always done but with greater awareness of what we are doing . Abandoning illusions is the first step on the path to authentic empowerment of individuals and communities.

Once that step is taken, we find that paradigms do exist for the relationship of science researchers and policy makers. For example, Robinson, in a series of articles (Robinson, 1982, 1988, 1990, 1991, 1992a) has argued that the Dragnet view of science (named for the US television series whose no-nonsense detective hero would prompt witnesses to deliver Just the facts ma'am) is at the root of perhaps the most fundamental misconception underlying standard views of the science/policy relationship: that it is a one way flow of objective information from science to policy (Robinson, 1982). He proposes instead a model relationship in which researchers, policy makers, and the public form mutual learning systems (Robinson, 1992a) that use modeling tools to explore alternative futures (backcasting) rather than trying to predict the future (forecasting). Elements of this model include the explicit recognition that policy questions are not essentially questions of fact but of value, and that both a physical flows perspective and an actor-system perspective are needed to provide a usefully integrated approach to policy questions (Robinson, 1991).

Such a collaboration would also help researchers and policy makers cope with both uncertainty and ignorance about how the social and biogeophysical systems interact. Indeed, uncertainty about these interactions is of such magnitude as to be better characterized as indeterminacy. Furthermore, this indeterminacy is likely to persist well beyond the timeframe in which actions would need to be taken to prevent, mitigate, or manage potential undesirable aspects of the human nature interactions.

Making assumptions explicit is essential to make natural and social science more relevant and more effective in policy making. When applying a model or theory to a particular situation, the reflexive researcher considers carefully whether or not the assumptions embedded in the approach actually match the policy context in which the knowledge is being used. If they do not, then the information obtained will not be valid for that context; knowledge is thus conditional. Policy decisions that ignore this conditional aspect often meet with strong opposition from people who do not feel that the assumptions are valid for the case at hand. For example, in democratic societies, if political institutions base their policy decisions on assumptions about human behavior that seem irrelevant to the public's experiences of itself and the world around it, they risk eroding the very legitimacy they rely upon to implement their policies.

An important effect of a collaboration between (both kinds of) science and policy would be the inclusion of social science analyses from the industrializing world. The Southern sensibility has emerged from an experience with weak institutions of governance (Banuri, 1993). Consequently, Southern critics have challenged the modern descriptivist features of instrumentalism, impersonality, and legitimization of impersonal violence. Further, the Southern sensibility sees environmentalism as based in an integrated perception of humans and nature, placing emphasis on issues of justice, equity, institutions of governance, and property rights. Including these perspectives as data in the global environmental change debate acknowledges the

genuine interdependence of rich and poor nations and broadens the solution space for global policy making.

Re-forming the relationship between environmental science research and environmental policy making would make space for the useful research that exists within interpretive science, for the alternative world views of less industrialized peoples, and for policy making strategies that account for uncertainty and indeterminacy. The environmental science and policy community would then include needs, wants, and beliefs in its data sets alongside measurements and statistics. Policy decisions and agreements would explicitly include social organization data and cultural assumptions as well as market and consumption data in global environmental policy decisions to make them more implementable and effective.

REFERENCES

Banuri, T (1993) The Landscape of Diplomatic Conflicts, in *Global Ecology: A New Arena of Political Con ict*, ed W Sachs, Zed Books, London.

Bullard, R D (1994) Unequal Environmental Protection: Incorporating Environmental Justice in Decision-making, in *Worst Things First: The Debate Over Risk-based National Environmental Priorities*, eds A D Finkel and D Golding, Resources for the Future, Washington, DC.

Damasio, A R (1994) *Descartes' Error: Emotion, Reason, and the Human Brain*, G P Putnam's Sons, New York.

Douglas, M and Ney, S (1998) *Missing Persons: Personhood in the Social Sciences*, University of California Press, Berkeley, CA.

Edelman, G M (1994) *Bright Air, Brilliant Fire: On the Matter of the Mind*, Basic Books, New York.

Fish, S (1994) *There's No Such Thing as Free Speech (and It's a Good Thing, Too)*, Oxford University Press, New York.

Geertz, C (1973) *The Interpretation of Culture*, Basic Books, New York.

Griffin, S (1978) *Woman and Nature, the Roaring Inside Her*, Harper and Row, San Francisco, CA.

Gross, J and Rayner, S (1985) *Measuring Culture: A Paradigm for the Analysis of Social Organization*, Columbia University Press, New York.

Gudeman, S (1986) *Economics as Culture: Models and Metaphors of Livelihood*, Routledge, London.

Headrick, D R (1990) Technological Change, in *The Earth Transformed by Human Action: Global and Regional Changes in the Biosphere over the Past 300 Years*, eds B L Turner, W C Clark, R W Kates, J F Richards, J T Mathews, and W B Meyer, Cambridge University Press, Cambridge.

Heineman, R A, Bluhm, W T, Peterson, S A, and Kearny, E N (1990) *The World of the Policy Analyst: Rationality, Values, and Politics*, Chatham House Publishers, Chatham, NJ.

House, P W and Shull, R D (1988) *Rush to Policy*, Transaction, NB.

Ingold, T (1993) Globes and Spheres: The Topology of Environmentalism, in *The View from Anthropology*, ed K Milton, Routledge, London.

Koford, K J and Miller, J B (1991) Habit, Customs, and Norms in Economics, in *Social Norms and Economic Institutions*, eds K J Koford and J B Miller, University of Michigan Press, Ann Arbor, MI.

Kuhn, T S (1970) *The Structure of Scienti c Revolutions*, 2nd edition, The University of Chicago Press, Chicago, IL.

Latour, B (1986) Visualization and Cognition: Thinking with Eyes and Hands, *Knowledge Soc.*, 6, 1–40.

Lavine, T Z (1984) *From Socrates to Sartre: The Philosophical Quest*, Bantam Books, New York.

Locke, J (1690) *An Essay Concerning Human Understanding*, T Basset, London, reprinted in 1990 by Scolar Press, Aldershot.

Merchant, C (1990) The Realm of Social Relations: Production, Reproduction, and Gender in Environmental Transformations, in *The Earth Transformed by Human Action: Global and Regional Changes in the Biosphere over the Past 300 Years*, eds B L Turner, W C Clark, R W Kates, J F Richards, J T Mathews, and W B Meyer, Cambridge University Press, Cambridge.

Merchant, C (1992) *Radical Ecology: The Search for a Livable World*, Routledge, New York.

Nichols, A L (1994) Risk-based Priorities and Environmental Justice, in *Worst Things First: The Debate over Risk-based National Environmental Priorities*, eds A D Finkel and D Golding, Resources for the Future, Washington, DC.

Patton, M Q (1978) *Utilization Focused Evaluation*, Sage Publications, Beverly Hills, CA.

Rayner, S and Malone, E L (1998) *Human Choice and Climate Change, Volume 4: What Have We Learned?* Battelle Press, Columbus, OH.

Rayner, S, Bretherton, F, Buol, S, Fosberg, M, Grossman, W, Houghton, R, Lal, R, Lee, J, Lonergan, S, Olson, J, Rockwell, R, Sage, C, and van Inhoff, E (1994) A Wiring Diagram for the Study of Land-use/Cover Change: Report of Working Group A, in *Land-Use and Land Cover: A Global Perspective*, eds W B Meyer and B L Turner, II, Cambridge University Press, Cambridge.

Ricardo, D (1814) *The works and correspondence of David Ricardo*, ed P Sraffa (1952), Cambridge University Press, Cambridge.

Rich, R F (1977) Uses of Social Science Information by Federal Bureaucracies, in *Using Social Research in Public Policy Making*, ed C H Weiss, Lexington Press, Lexington, KY.

Robinson, J B (1982) Apples and Horned Toads: On the Framework-Determined Nature of the Energy Debate, *Policy Sci.*, 15, 23–45.

Robinson, J B (1988) Unlearning and Backcasting: Rethinking Some of the Questions We Ask about the Future, *Technol. Forecast. Soc. Change*, 33, 325–338.

Robinson, J B (1990) Futures under Glass: a Recipe for People Who Hate to Predict, *Futures*, 22(9), 820–43.

Robinson, J B (1991) Modelling the Interactions between Human and Natural Systems, *Global Environ. Change*, 1, 629–647.

Robinson, J B (1992a) Of Maps and Territories: the Use and Abuse of Socioeconomic Modeling in Support of Decision Making, *Technol. Forecast. Soc. Change*, 42, 147–164.

Robinson, J B (1992b) Risks, Predictions, and Other Optical Illusions: Rethinking the use of Science in Social Decision-making, *Policy Sci.*, **25**, 237 254.

Robinson, J and Timmerman, P (1993) Myths, Rules, Artifacts, and Ecosystems: Framing the Human Dimensions of Global Change, in *Human Ecology, Crossing the Boundaries*, eds S D Wright, T Dietz, R Borden, G Young, and G Guagnano, Society for Human Ecology, Fort Collins, CO.

Russell, B (1946) *History of Western Philosophy*, George Allen and Unwin, London.

Ryle, G (1949) *The Concept of Mind*, Hutchinson, London.

Selznick, P (1992) *The Moral Commonwealth: Social Theory and the Promise of Continuity*, University of California Press, Berkeley, CA.

Snow, C P (1959) *The Two Cultures*, Cambridge University Press, Cambridge.

Turner, B L, Clark, W C, Kates, R W, Richards, J F, Mathews, J T, and Meyer, W B (1990) *The Earth Transformed by Human Action: Global and Regional Changes in the Biosphere over the Past 300 Years*, Cambridge University Press, Cambridge.

Watson, R T, Zinyowera, M C, and Moss, R H, eds (1996) *Climate Change 1995: Impacts Adaptations, and Mitigation*, Cambridge University Press, Cambridge.

Weiss, C H (1982) Policy Research in the Context of Diffuse Decision-Making, in *Social Science Research and Public Policy-Making*, ed D B P Kallen, NFER-Nelson, Windsor.

Weiss, C H and Bucuvalas, M J (1980) Truth Tests and Utility Tests, *Am. Soc. Rev.*, **45**, 302 313.

Whiteman, D (1985) The Fate of Policy Analysis in Congressional Decision-Making, *West. Political Q.*, **38**, 294 311.

The Emergence of Global Environment Change into Politics

JOHN DUNN

University of Cambridge, Cambridge, UK

Unlike most major elements of contemporary politics, global environmental change entered the political eld, not under the peremptory impress of immediate political incentives, but through a slow, vague, diffuse and uneven shift in belief. The beliefs in question, moreover, were not about the properties or potential conduct of fellow human beings, but about the residually non-human setting of human life in its entirety. Local environmental change, apparent or real, has been seen as politically signi cant throughout the history of our species, in myth since Noah's ood, and in recorded reality since at least ancient Egypt. Even today most environmental changes which are seen as politically important are relatively local in incidence or highly differentiated in their spatial signi cance (Elliott, 1998; Held et al.*, 1999). Global warming has very different medium-term implications for the inhabitants of the Maldives or the Sahel than it does for those of Berkshire. The holes in the ozone layer, when they appear, cover some bits of the humanly inhabited map and not others. To see global environmental change as a factor and focus within contemporary politics is not to ignore (still less to deny) this drastic variation in impact and signi cance. It is simply to register the cumulative effect, and the awesome prospective signi cance, of a novel practical preoccupation and a fresh way of seeing the setting and destiny of human political life as a whole. It is easier, naturally, to plot the cumulative effect thus far than to assess the prospective signi cance (Dunn, 1998). But since the impact so far in practice is relatively modest, and the prospective signi cance at the limit genuinely eschatological (the termination of human life through the consequences of human action), it would be an error of judgement to concentrate principally on the more tractable topic.*

Since global environmental change emerged into politics in the first place through a hesitant change in belief and not an abrupt alteration in the matrix of political incentives in any definite political location, it can only be understood adequately by identifying where that shift came from. At present we do not really know. We still have no compelling and thorough intellectual history of the interactive dynamics of the large range of natural sciences whose combined cognitive resources give us our best assessment, either of how much damage the natural setting of human existence has already undergone, or of how rapidly we can expect it to continue to deteriorate, or even of what could, in principle, be done effectively to retard, halt or reverse some of these processes of deterioration. We have no coherent synthetic study of the often very idiosyncratic and obsessive personal odysseys of individual scientists or publicists who picked out major sources of continuing damage to habitat, and conveyed them eloquently to mass audiences in countries which permitted such communication. Either of these would be an exacting technical exercise in the contextual history of ideas (Dunn, 1996). Each, if it was done convincingly, would carry important implications. Because the major capitalist countries today still make the primary contribution to global environmental deterioration through the continuing momentum of their productive systems, and since they also, at present, continue to provide much the best access to public media and the least obstructed channels for mass communication, some of the main political ambiguities of their emerging preoccupation are already writ large in both of these two histories. As yet, however, none of us can do much more than guess intelligently at the contours of the histories themselves. It is extraordinarily difficult in principle to explain why any important passage of intellectual history should have gone as it did, not least because these passages necessarily contain episodes of real intellectual creativity (Quinton, 1999). Intellectual creativity, for reasons made familiar by Karl Popper's supposed proof of the impossibility of predicting the future of any natural science (Popper, 1957), may just in detail be beyond human explanation.

What certainly is important in this sporadic and amorphous shift in belief, however, is its timing, its

location, and its dependence on rapid diffusion of images and judgements, on the relative independence of such diffusion from the immediate purposes of holders of political authority, and perhaps, too, on a certain dedicated obstinacy and crankiness on the part of some of its main carriers. For major shifts in belief to be practicable at all, it is necessary for their protagonists to be able to endure loneliness and unpopularity, and to be willing to seem and go on seeming damagingly eccentric.

Many very different things happen as global environmental change emerges into politics, but one disconcerting and inevitably elusive thing which does so is that quite bizarre and apparently personal obsessions of the far periphery move insistently towards the centers of power, responsibility, judgement and eventual choice in world politics. This is not a process which anyone's personal or professional experience of politics can equip him or her to chronicle with much ease or assurance. It is a standing temptation to confident misjudgements from virtually any angle. But it is every bit as important for the human future to learn to see it more accurately as it is to learn to judge more accurately just how deeply and in what ways we are still damaging our habitat, and what we can do to improve our judgement adroitly and at some speed. If we are to do so, we must improve our collective assessment of whom or what to trust (or distrust) to instruct us on the risks which we are running, on how these risks can be minimized, and against which of them we can, to any significant degree, insure ourselves by other courses of conduct which remain open to us. All of these are political tasks as much as cognitive puzzles. None of them is a matter about which any individual human or group of humans can yet be said to know exactly what is the case (let alone what could become the case). Some of them are matters about which the very idea of exact knowledge is probably confused in the first place.

When social scientists or politicians now speak of the emergence of global environmental change into politics, what they have in mind is not merely this morass of political suspicion and bemusement. It is also the appearance, in a readily specifiable set of institutional sites, of a relatively determinate agenda of novel concerns (Hurrell and Kingsbury, 1992). The concerns themselves may have come originally from quite personal and historically contingent acts of judgement on the part of particular individuals, but they have become an agenda through an increasingly definite and insistent political process. They have imposed themselves first on the attention, and then on the purposes, of those in authority (insofar as they have any) by at least threatening to limit or were remove the power of those who were previously unaware of them or were more than prepared to ignore them. By this stage, plainly, they had indeed altered the matrix of political threats and rewards faced by professional politicians in and out of office, and made their inroads precisely by doing so. In this process personal change of belief broadens out into the modulation of public opinion, and public opinion, in turn, in due course condenses into an effective political force pressing political leaders in novel directions.

To map this diffusion accurately it would be necessary to relate together very variegated sorts of changes, from the creation and development of political parties, non-governmental organizations (NGOs) or pressure groups, through the impact of each of these on the dynamics of party competition in different nation states, and on to the impact of all of them together on state bureaucracies, inter-state negotiations, treaties and their enforcement, and the creation or modification of international organizations, initiated by state decision but exerting, for some purposes, at least a persuasive or normative authority over the states which have created them. As global environmental change emerges into politics, it does not merely create a new agenda and a novel political field centered on that agenda. It also forges a multiplicity of fresh institutions and practices which take that agenda for their own and press different elements of it to what they take to be their own (and the globe's) advantage: this provides both a milieu of existence and an endless series of actual or potential teams of action, usually competing more or less acrimoniously for attention, financial resources, influence, and even power.

While the agenda and many of the participants are novel enough, many of the constituents of this new field are altogether more familiar. They were there before; and they depend neither for their power nor for their *raison d'être* on this novel field. This is uncontroversial in the case of nation states and much of their public bureaucracies. It is uncontroversial for the main legal and organizational frameworks of the national economies, from whose interactive global dynamics the environmental changes largely emanate. It is even uncontroversial in the case of the principal international sites of political or legal authority – above all the United Nations (UN) itself.

The new institutional sites are at most subordinate agencies of older institutional sites which still clearly hold greater authority, or (in the case of Green parties in a number of countries as yet principally in Western Europe) competitors of limited efficacy with older and better entrenched institutions in the same settings. Only in the field of pressure groups or NGOs (Greenpeace, Friends of the Earth) has the novelty of the agenda yet bred correspondingly novel types of agent with the vitality and innovative capacity to press it home in essentially unprecedented ways.

All of these sites offer an opportunity for careers open to talent, energy and luck, and redirect political initiatives accordingly. There is no *a priori* reason why some of them should be more important, more dependably benign, or more irredeemably corrupt than others. Some analysts or participants put such trust as they can muster in NGOs

or in the conceptually still more diffuse substance of civil society, global or national. Others not merely focus on, but presume the essential practical rationality of, at least some existing states or institutional processes of inter-state cooperation. Others still place altogether greater confidence in the heuristic delicacy and epistemic reliability of market processes, confining their political attention to the question of how to minimize impediments to the latter's functioning by designing an optimal regulatory regime.

None of these viewpoints is simply inane. But none offers a privileged perspective on the bewilderingly complex and still largely opaque field which the agenda of global environmental change is attempting to address. This is not as yet a field which can be surveyed as a whole with any assurance from any given human vantage point. There is no strong reason to believe that it will ever become such. To improve our understanding of it, we must divide it up into more manageable components, and postpone asking how these components in the end bear upon one another until we have mustered a reasonably complete and accurate inventory of each of them.

It is best to begin with the immediate sources of the most drastic global environmental changes, which plainly fall in the realm of economics. Some important, and often environmentally highly destructive, forms of economic production are closely tied to the logic of military competition today (Held *et al.*, 1999), as they have been for several centuries in the past (McNeill, 1983; Kennedy, 1988). But the principal dynamic source of environmental degradation is still a process of acutely competitive expansion in the production and exchange of goods and the diffusion of services, driven on the supply side by the quest to accumulate capital and on the demand side by the quest for greater and more rewarding consumption. There is little agreement, unsurprisingly, on whether the global economy forms in any real sense a unity, or whether it is governed by a single overarching logic (Brenner, 1998; Wallerstein, 1995; Little, 1982; Lal and Myint, 1994). There is still less agreement as to whether the process of economic development itself implies a predestined direction for the future evolution of economies and polities across the world, or even on where that direction is most likely to lie. Is the world trading system likely to be more liberal or less so in 25 years time? Are there likely to be many more, or far fewer, or roughly the same number of clearly consequential sites for major political or economic choice in the world in several decades time? But while there is sharp dispute over the degree, scope and significance of present globalization, and the inevitability or otherwise of future globalization, and some variation in optimism about the possibility of generating more sustainable patterns of development within the next generation or two, virtually no coherent analysis of the politics of global environmental change assumes marked and protracted retardation of growth in the production and consumption of goods and services, the velocity of economic transactions, or the flow of international trade. This may be more an index of cowardice or failure of imagination than of rationally secured belief. But the continuity of assumption across sharply contrasted personal dispositions and political tastes is an important political factor in itself, and has done much to shape the sequence of international diplomatic negotiations on response to global environmental change which have centered on the new agenda and provided its most prominent and potentially consequential entree into politics.

To focus in this way on global economic dynamics is not to presume that all political action is prompted in the last instance by a narrowly economic causality. But it certainly is to accept the decisive role of economic struggle and effort in setting the limits to political possibility in this domain, as it does throughout modern politics (Dunn, 1990b). Politics sets its own limits, in return, on how economic activity can be and is organized (Dunn, 1999). But in a human world so comprehensively reshaped by the history of capitalism, it does so throughout on terms which are obtrusively set by the requirements for competitive success on markets which, however doctored they may be, can never be fully controlled by political choice. One very important practical question which is already being tested in the domain of global environmental change is whether (and how far) collective political intelligence and can hope to re-establish control over the blind, convulsive, irresponsible dynamics of competitive economic effort. This has been under spirited dispute for several hundred years; and the answer remains disconcertingly open, not least because a less discerning review of political will, examined over time, makes it all too clear how much easier it is to disrupt or obstruct large-scale economic activity than to bring it under firm control. This is as much a truth about economic organizations themselves as it is about the interaction of formally political with formally economic institutions. But its political significance is every bit as momentous as its significance within narrowly economic activities.

What made it politically impossible to ignore global environmental change any longer was the transformation brought about by the great economic boom after the Second World War which came to an end in 1973 (Brenner, 1998). The next quarter of a century may have seen less rapid growth in the more advanced capitalist economies, and some degree of regression for extended periods of time in state socialist economies and their uncomfortable successors (Brenner, 1998). But, virtually throughout, it also saw a continuation of net global economic growth, and growth from an initial level so high that the pace and scale of global environmental degradation became at least blearily apparent to the most myopic and irresponsible holders of state authority. (See Held *et al.*, 1999, 390–396 for an account of the progress of global environmental degradation and its economic embedding.) As it did

so, it set up and intensified a steadily more obtrusive and extensive network of diplomatic interactions, institutional invention, and attempts to modify economic and political practices to diminish the harm which they were so conspicuously inflicting. At stake throughout this political and economic adjustment were both the urgency and relative priority of minimizing particular threats, the range of expedients through which the diminution might most effectively be undertaken, and the necessarily vexed question of who precisely was to pay for it and at whose expense. The relative urgency of minimizing particular threats was initially the preoccupation, principally of political amateurs, with a high degree of sensitivity to environmental vulnerability. But even it penetrated the world of professional high politics with some rapidity from the early 1980s, reaching as unlikely targets as Britain's Prime Minister Margaret Thatcher. The assessment of how best to meet these threats clearly required diplomatic and economic expertise but she soon convinced career politicians and public officials of the importance of addressing it. The allocation of responsibility for meeting the bill, still more clearly, was the stuff of high politics and intense interstate diplomatic activity throughout.

The core of this process of recognition and incipient political response was a series of major international conferences, running from the UN Conference on the Human Environment in Stockholm in 1972, to the UN Conference on Environment and Development in Rio in 1992 (Elliott, 1998, Chapter 1). Before 1972, international agreements which bore directly and explicitly on the protection of the environment, focused principally on the preservation of wildlife, the law of the sea (alongside the law of war, the first major focus of efforts to develop a truly international law), and from the 1970s onward the attempt to contain the obvious and acute threat of nuclear armaments, like the menace of nuclear winter (Elliott, 1998, 8–11).

The danger posed by nuclear weapons, for obvious reasons, was a focus of very high and distinctly arcane politics from the point at which the scale of their potential for destruction first became clear to anyone but nuclear physicists – in effect from the Manhattan Project onwards. An intensely controversial issue throughout this history has been just what did in fact determine the political and military choices made by state powers, not merely over the deployment or use of nuclear or thermonuclear weapons, but even over the development of civil nuclear power as a basis for national electricity supply (Alperovitz, 1995; Clark and Wheeler, 1989; Freedman, 1981; Gowing and Arnold, 1974; MccGwire, 1987). Because it has been so intensely politicized and so evidently consequential throughout its history, both within particular states and still more so in the more conflictual relations between them, this has been the last area of global environmental degradation in which public opinion has made a clear political impact. Even in countries like Sweden, where popular attitudes have been deeply affected by Green awareness for decades, the politics of transforming a domestic electricity industry, largely based on nuclear power, have been as slow as they have proved painful. Not even the disastrous explosion in the Chernobyl nuclear reactor in 1986, the pollution which this scattered so rapidly over such immense distances, and the dramatic impact of this on European, and indeed, world public opinion, proved sufficient to accelerate the shift from nuclear to other energy sources for electricity production. What it may well have done in retrospect, however, is to halt the formidable momentum of the nuclear industry up to that point.

The concern for the preservation of wildlife had a protracted imaginative past in the slow modulation of European conceptions of the proper relations between human beings and other living creatures (Thomas, 1983; Dunn, 1969; Passmore, 1974). But it had, of course, a very different imaginative past in other areas of the world, whose religious traditions had long taken a less anthropocentric view of the relative significance of humans and other creatures (Gombrich, 1971; Carrithers *et al.*, 1985). In European colonial or imperial territories, and above all in British India, the practices of preservation were often more closely linked to the preservation or extension of elements in the lifestyle of a land-owning aristocracy or gentry than they were linked to the systematic commercial exploitation of natural resources, or the attempt to contain the damage which this was always likely to inflict.

In the peculiarly ecologically vulnerable continent of Africa, interstate agreements between the European powers to preserve wildlife were signed as early as 1900, virtually as soon as the scramble for Africa was completed, and well in advance of the more far-reaching Convention on Nature Protection and Wildlife Preservation of 1940 which covered the western hemisphere as a whole. In the African case, it seems clear, the 1900 agreement was not a response to extensive damage already plainly done, but rather a conscious prophylactic exercise in stewardship, reflecting a deliberate acknowledgement and assumption of human responsibility for non-human nature (cf. Passmore, 1974). This may have had very limited effect on the dynamics of subsequent ecological degradation. But it certainly confirms that concern for such degradation is neither a recent prerogative nor necessarily an index of deep political disaffection in the face of existing political or economic power or privilege.

A purer case of eventual cultural revulsion at the environmental destructiveness and brutality of the exploitation of land resources over a far longer period – in fact several centuries – is given by the territorial subjugation and ethnic displacement which marked Europe's impact on the North American continent (Cronon, 1983, 1992). In the United States, popular organizations for the preservation

of wilderness areas and bird and animal species (the Sierra Club, the National Audubon Society) helped both to prompt and also to respond the opportunities provided by the National Parks network (Runte, 1997), and furnished a potent basis of organization and experience for the very much larger numbers of private citizens who began to sense the ecological crisis which faced their country in the 1960s and 1970s.

One way of seeing what has happened in the emergence into politics of global environmental change is to see this as the rapid and disconcerting confluence of two startlingly different sorts of political setting and two starkly contrasted types of human meaning. One setting is the exercise of statecraft from the summits of highly organized and always prospectively coercive pyramids of power and authority. The interpretation of the goals of statecraft and the constraints upon it is always open to fierce dispute; but over time it strongly favors a focus on the distinctive rationality and the often singularly perturbing *mores* of states as such: on *raison d'êtat* (Meinecke, 1957). The other setting is the movement of the individual soul or sensibility, increasingly importuned within the cultural history of the West, to respond with Romantic excitement and veneration to the wonder of the rest of nature, the non-human setting of human life, and to measure itself by the depth and energy of its response (Taylor, 1989, 1992; Unger, 1999).

A third setting and type of meaning, the economic rationalization of production and exchange under an increasingly globally pervasive capitalism, has penetrated both ruling elites and popular sensibilities quite deeply, but in very different ways. It has moved the imperative of economic efficiency ever closer to the center of *raison d'êtat* (Dunn, 1990b; Baldwin, 1992; Keohane, 1984; Gilpin, 1987). But it has also inflamed the post-Romantic sense of spiritual self-regard in the face of the dizzy commodity fetishism of modern mass consumption, and has intensified revulsion at the cumulative damage which this inflicts. In the earlier stages of these twin developments, it was easy to doubt that their simultaneous occurrence could be of any enduring political importance. How could the personal, the aesthetic, and the often self-consciously unworldly, converge effectively with the carefully self-protected and heavily armed sites of ultimate coercive authority and organizational control in collective human life (in their own vision, largely impersonal, deeply inured to ugliness, at least at a safe distance, and worldly to the last degree)? This was an encounter for which neither element could genuinely and clear-headedly wish, especially in its earlier stages: at best a collision which was bound to deform many of the purposes which came together in it, at worst a nightmare chaos which threatened all that each held dearest. For state elites, even in a representative democracy, this perspective cannot readily change. For them, the shifting preoccupations of their fellow citizens are always more likely to prove an irritation or inconvenience than a steady source of gratification. But the citizens, consumers and producers, who give these elites their power as well as their legitimacy, have their own lives to live and naturally prefer their own wills and sentiments to bear on the ways in which they are governed. In representative democracies, moreover, they are assured that this is both a right and a genuinely available opportunity (Dunn, 1992).

The increasingly hazardous and alarming momentum of global environmental change was thus a brusque threat to the format of existing politics and to the imaginative habits and practical routines fostered and reinforced by that format. Above all, it was a challenge to the implicit assumption that these were in good working order and could be confidently trusted to take effective care of the ongoing destinies of modern human populations on the latters' behalf, and without too close invigilation from outside the ranks of their habitues. It was also thus a challenge to the existing institutional division of political, social and economic labor, and to the professionalization of political and public bureaucratic life. What it suggested, above all, was not that the existing incumbents of high political and bureaucratic roles should be cashiered and promptly replaced with human alternatives with different tastes, dispositions or purposes, in the manner suggested by the development of the routine organizational apparatus of modern representative and administrative politics since the French Revolution (Finer, 1997; Manin, 1997; Dunn, 1992, 2000). Instead, it was that these roles, as a whole, should somehow be deinsulated, and opened up to far more diffuse, and often instrumentally pretty unreasonable, movements of taste and surges of anxiety, dispersed across the population at large. This was a prospect calculated to dismay any self-respecting political class or professional stratum; and it was obvious enough that it could readily and rapidly do far more harm than good (as it may already have done in a number of prominent instances, like the Brent Spar oil platform disposal decision). What was not obvious was how it could well be avoided, or what more promising model of overall political and institutional transformation was in principle available.

As it emerged into politics, the hazards of global environmental change underlined, above all, a need for rapid large-scale transformation of many prominent human practices. A case can be made for seeing the key site for this change in three distinct ways – as lying in the realm of culture (in the conceptions of the goals of human life and the consequent living priorities of given human populations); as lying in the main organizing logic of economic activity; or as lying in the genesis and strengthening of new institutions of coordination and control, which can determine how every human population will in fact respond to the newly recognized, and ever less comfortably parochial, context of its physical existence. These different locations of the

need for change clearly imply very different strategies for fomenting change. Judging their current plasticity or potential responsiveness is not merely largely the prerogative of different academic disciplines. It also requires, at least in the first instance, a hypothetical presumption of relative stability in each of the other two.

A world of generalized and indeterminate flux cannot be assessed coherently from any particular angle, and precludes firm judgement, even in principle, on how best to respond to it. Such a world is no more tractable in practice; but at least it does not rule out comprehension in principle. Seen over time, moreover, it is distinctly less discouraging even in practice. Thirty years ago, very little of our present scientific basis for assessing the more salient elements in global environmental change was yet available. Much of what we can now determine with relative accuracy was a matter for the purest (and most projective) speculation. Today many of the most consequential judgements of the degree and imminence of the peril in which we stand remain alarmingly unclear, and hence continue to be fiercely contested even by the ablest scientists directly concerned. Some of the most devastating potential scenarios are likely to remain effectively imponderable in any foreseeable future. But we have no reason whatever for seeing our world (as a whole) as a domain of generalized and indeterminate flux, and every reason to hope that we may succeed in determining what, within it, is modifying what, and in what ways, far more accurately and sensitively, even in the next few decades. To do this, we need above all to learn to take the formats of our existing fields of academic or cognitive specialization more lightly and less literally – to see these formats more as precarious and politically driven artefacts, and less as cultural treasures or dependable deliverances of the obvious. When and if we learn to do this, the hypothetical bracketing of adjacent cognitive domains will become both easier and more natural, and we can begin to learn how to judge how each given domain bears on the others. In place of a categorical assumption of external fixity grounded mainly in intellectual habit, or an epistemically gratuitous presumption of infinite social and political plasticity grounded solely in personal temperament, we can hope to develop a controlled heuristic strategy of open inquiry. Such a strategy could (and would need to) interrogate all elements of the motivation, interactive dynamics and institutional context of human behavior, in face of a still rapidly deteriorating habitat, and address throughout the question of how that deterioration can be arrested or reversed.

If we think of these three potentially decisive sites for adjustment, and see the adjustment itself as a single collective task for our species over time, it is clear at once how hard it is to judge how much of the task can sanely be assigned to each site. If the crucial site proves to be the domain of culture, the task is one for the human spirit: this is because the fundamental, but endless choice of all human beings is what sort of persons we wish to be and what sort of lives we wish to live. Assessing grounds for hope (or despair) in this domain is an ancient human preoccupation: a task in the first place for religious practitioners or moralists, and then perhaps, more academically, for theologians and philosophers, historians, sociologists or psychologists, especially, in the end, perhaps for psychologists. The inductive record here is certainly discouraging. Grounds for despair come at least as readily as grounds for hope. And hope itself, when it does come, seems often to come less from grounds of any specifiable kind than from sheer biological vitality. Can psychologists learn to judge the fundamental properties of the human spirit? Can they in principle ever tell us, even over a very long span, whether or not a human being is a kind of creature which could learn, morally, to live far more responsibly, temperately and delicately than ever before in any numbers? Can a person learn a very different level of patience, tolerance and generosity towards immense numbers of fellow humans quite far away, of whom very little is known? Modern capitalist civilization is founded on the focusing and intensification of the effort to get what we want and to live, within its ever shifting contours, as happens to please us: on the pursuit of modern liberty (Constant, 1988, 313–328). Can this, in principle, be transposed into a tissue of tasteful and dependable self-inhibition in the service of a threatened global ecology? On balance the historical odds seem clearly against it. Perhaps no psychologist (perhaps not even psychologists as a profession in their entirety) can be said to know this to be a task beyond the unmediated resources of the human psyche. But insofar as it is cumulative at all, the intellectual history of psychology as a would-be science leads us to expect it to prove so. And however limited their present contribution to telling us whether we collectively possess the spiritual resources to preserve our global habitat, it is at least less paltry than the advice which they can yet offer on the question of how best to mobilize the resources which we do have for the purpose.

If the crucial site is the organizing logic of economic activity, the task is principally for economists, for those who design and implement governmental economic policy, and above all for those who lead and manage very large economic enterprises. The key problem here is how to modify structures of incentives so that at least as much continues to be produced and exchanged indefinitely, but with altogether less attendant damage. The task is to alter the grounds for action of economic agents, to limit negative externalities, and to do so without impairing the dynamism, or flexibility of adjustment, of the overall economic activity off which, through which, and within which, human populations all now live. There has been very extensive study of the economics of global environmental problems over the last three decades, and some definite intellectual

headway in analyzing both their sources and the range of means through which it might be possible to handle them effectively (Dasgupta and Maler, 1997; Mabey *et al.*, 1997; Swanson, 1996; Haugland *et al.*, 1998).

What is already clear from this study is that, whilst there are no rigid economic obstacles to lessening the pace and scale of environmental damage, the costs of halting or reversing such damage, even where it could in principle be reversed, are vast enough to preclude its reversal in many fields in the imaginable future. Equally clearly, the allocation of these costs between different groups of human beings over time is a matter of politics through and through, and certain to remain just as bitterly contentious for the imaginable future as it has already proved over the last quarter of a century.

If the crucial site of decision is neither in culture nor in the technical design of economies, but in the devising and implementation of novel institutions of coordination and control through which to recapture some degree of control over our cumulative damage to our habitat, the challenge in the end is to our capacities for reflexive prudence and political cooperation (Dunn, 1990a, conclusion). The task is, above all, one of political understanding and political action.

A pessimistic view would see all three of these tasks as already unmistakably beyond us. The spiritual assignment of chastening greed, enhancing temperance and composure, and guiding us back into a more serene harmony with the rest of the world which we have so deformed and defiled, suggests familiar Platonic remedies for spiritual corruption: a harsh censorship, a formidable closure of experience, a rigid structure of authority, made politically credible only by uninhibited cultivation of dubiously noble lies. This defies the logic of modern politics and modern economics more or less in their entirety. There is no reason whatever to believe that the power could still be mustered to impose it, and every reason to fear that the harm which it would certainly do would massively outweigh any possible good that might also flow from it.

The economic assignment of internalizing the negative externalities suggests an endless battery of technocratic expedients, of which the only aspects about which we can yet be reasonably confident is that most of them will not work in practice, and that, in their entirety, they will largely reproduce the drastic inequalities in life chances everywhere embodied in existing economies, and reinforce the acute resentment which these already arouse. The political assignment of judging together what our predicament really is, choosing together how to respond to it, and cooperating with one another to ensure that these choices are implemented, suggests familiar oscillations between the invention and parading of international institutions for cooperation, and sustained haggling and mutual intimidation within these (along with other older sites) to impose the costs even for the outcomes which we do, on balance, favor on others who are even worse placed to afford them: an unrelenting struggle in which the strongest members of society will still take what they can and the weak still yield what they must (Thucydides, 1921, 158).

Neither the spiritual, the economic nor the political task, however, can safely be written off. In all three, some degree of contribution is almost certain to prove indispensable, if any secure headway is to be made in such a devastatingly complicated, slow, unobvious and sequentially uninviting endeavor. The least rewarding to focused thought in the short run is the direct assault on the psyche, not least because the focus in question is so extraordinarily hard to maintain. Much more promising is a careful reconsideration of the relative elasticity of technocratic economic expedients and more overtly open-textured political conceptions and practices. The economic expedients (above all the design of regulatory regimes and tax systems) can articulate with some precision with the still more plainly technical appraisal of the dimensions and rectifiability of environmental degradation, to make up a joint agenda of what most imperatively needs to be modified, and how it can best be modified. This would certainly not show that we can so modify it in practice. But it would make it as clear as it can be made that we must possess some good reasons for wishing to modify it. The agenda of global environmental change over the last quarter of a century has been an attempt to capture such a joint assignment (Brenton, 1994; Vogler and Imber, 1996; Tolba and Rummel-Bulska, 1999; McCormick, 1995). The diffuse but sometimes very vigorous political processes which came to bear on the major international environmental meetings from the Stockholm summit of 1972 to the Rio Summit of 1992 and since (Thomas, 1994), and which now permanently surround all governmental and intergovernmental environmental policy-making, have been a sustained attempt to interpret and implement that agenda in practice as their participants judged best.

The UN Conference on the Human Environment, held in Stockholm in June 1972 (Elliott, 1998, 12) in the presence of the Prime Ministers of Sweden and India, set the pattern for subsequent diplomatic and organizational initiatives in limiting environmental damage by issuing a declaration, adopting an Action Plan, and spawning the new UN Environment Programme (UNEP), founded in 1973, with its headquarters in Nairobi, under the leadership first of the Canadian Maurice Strong and then of the Egyptian Mostafa Tolba (Tolba and Rummel-Bulska, 1999). From its foundation, this organization represented a precarious (and very poorly funded) compromise between the competing anxieties of the wealthier and poorer countries, and a clear concession to the fears and jealousies of existing UN agencies. But its subsequent impact on a string of influential meetings to assess and respond to global environmental

change in many different fields (Elliott, 1998, 109) was not merely a testimony to the skills and determination of its own leadership, but also an index of the increasingly evident urgency of the issues at stake, and the diplomatic and political advantages of approaching these, not through the political apparatuses of individual sovereign states, or the cumbersome machinery of the UN General Assembly, but through a specialized institution fully devoted to the purpose. In the two decades between the Stockholm meeting and the much larger and more ambitious UN Conference on Environment and Development held at Rio de Janeiro in June 1992, the agenda of global environmental deterioration became much more dense and more differentiated and acquired a quite new political momentum. It also generated a wide range of new institutions, practices and agreements to monitor or respond to the many different dimensions of environmental degradation. By the time of the Rio meeting, the national delegations accredited to the conference were larger and more numerous, and in many cases led by their respective Presidents or Prime Ministers. In addition they were officially supplemented by over 1400 accredited NGOs; and the conference itself was paralleled by a Global Forum with more than 30 000 participants, held under NGO auspices and organization (Elliott, 1998, 19). Both the conference itself, and the activities surrounding it, were the focus of unprecedented media attention. By its close, the conference had adopted three wide-ranging formal agreements and drawn up two major new conventions, one on climate change and the other on biodiversity.

Even in retrospect it is hard to distinguish the ultimate effect of this flurry of diplomatic activity from the rhetorical vigor and invention lavished upon it, and the individual and institutional careers advanced by it. A skeptical assessment would see the main clear gains as lying in the strengthening and extension of the global apparatus of scientific monitoring, and the political affirmation of the urgency and priority of the goal of limiting further ecological degradation, rather than in the relatively modest and still largely unimplemented provisions for forest protection embodied in the Rio Statement of Forest Principles (Elliott, 1998, 85–89), the precariously coherent Framework Convention on Climate Change, the practical impact of which remains very much in contention (Elliott, 1998, 68–73), or the Convention on Biological Diversity, which has little immediate prospect of slowing the rate of species loss in marine or forest environments (Elliott, 1998, 73–89). In all of these domains, what does decide the ultimate outcome of such initiatives is not the personal political taste of a small range of powerful political leaders in wealthy countries (let alone of a much larger range of leaders of much poorer ones). It is the balance of political and economic force in intensely fought distributive struggles across the economies and societies of the world. It is never, in principle, possible to see what that balance is at any point in time, although it sometimes becomes quite easy to judge in retrospect what it must have been, with several decades of hindsight.

In interpreting the pace and vigor of political response to perceptions of environmental deterioration, it is most illuminating to consider separately the relative technical tractability of the project of arresting or reversing the damage, and the structures of interest involved both in suffering the damage itself and in bearing the cost of averting it. Where great concentrations of wealth or economic power have a large stake in things as they are (the eating habits of Japanese consumers at the height of the bubble economy, the energy industries of western powers, or indeed of the former Union of Soviet Socialist RepublicsUnion Of Soviet Socialist Republics (USSR) and its successor states), the political assignment of altering them rapidly is virtually impossible. Where the great majority of the world's present population have a large (and in the short term an inextricable) stake in things as they are, it is also normatively uninviting. Assigning the costs for arresting global warming with any rapidity is not merely a matter of balancing the power of very powerful interests against the human needs of very large and poor populations; it is also a matter of balancing huge short-term costs against potentially even vaster long-term gains, in a political context in which the losses are devastating and impossible to overlook, while the gains cannot in principle be guaranteed to accrue at all, and will accrue, if they eventually do, only to persons who are in many cases quite different from those who now bear the costs (Parfit, 1984). Ecological degradation harms the interests, and arguably violates the rights and threatens the needs, of future generations of human beings. But any allocation of the costs of seeking to forestall it is bound to entrench on the interests of some present or future human beings, and quite likely in practice to impair the rights, or even encroach on the needs, of many groups of human beings who have no idea of what is at stake and no means of defending themselves against the measures in question. The politics of global environmental change cannot therefore be validly understood as a struggle between virtuous, clairvoyant and deft NGOs, consciously devoted to world salvation, and vicious, obtuse and clumsy states or corporations, more or less inadvertently committed to world destruction. Nor, self-evidently, can it be validly understood as the mirror image of such a struggle, with the normative signs reversed. What it is above all is a very maladroit and confused exploration of just how dangerous and destructive global environmental change really is, and an understandably acrimonious range of quarrels about who is to blame for its grimmest aspects, and who has either the responsibility or the power to begin to cope with them. Anyone who finds this surprising would simply show that they do not understand what politics is really like or why it is as it is (Dunn, 2000).

The speed and clarity with which environmental damage is identified and challenged in the first place has depended, throughout human history, both on the freedom, confidence and political effectiveness of its immediate victims, and on the salience of the damage itself. The pace at which that damage is arrested and reversed depends at least as much on the technical tractability of the process to intentional human intervention. When the Iraqi occupying forces set fire to the oil wells of Kuwait, their political master fully intended the scale of destruction which he unleashed. But most important environmental damage is an unintended side effect of activities pursued for quite other reasons; and a very large proportion of it is still as unwelcome as it is unanticipated.

If we consider six of the major dimensions of global ecological degradation in relation to one another – cross-border acid rain, stratospheric ozone-depletion, deforestation, loss of biological diversity, and global warming (with its attendant threat of massive climate change), it should by now be possible to explain much of the pattern of political response by relating together their relative salience, the political strength or weakness of their immediate victims, the power of those with a major stake in the activities which principally cause them, and the costs and technical tractability of beginning to remedy them. Unsurprisingly, no one has yet attempted an overall interpretation of this ambition. (For a thoughtful and well-informed synoptic view see Held *et al.*, 1999). But it is not difficult to suggest its main outlines.

Marine pollution has been one of the earliest foci of interstate environmental concern, both because of the obtrusiveness of some of its aspects, and because it is so plainly, in most cases, beyond the authority or power of any particular state to control or rectify. In the case of more vulnerable seas in areas with relatively wealthy populations (the Mediterranean, for example), the dangers were identified relatively early, and attempts made to remedy them. But the main threat of marine pollution has come from the vast expansion in world water-borne trade since 1945 and the hectic development of the world oil industry, especially from off-shore fields in the last quarter of a century. The starkest destruction thus far has come in the case of huge inland water bodies in the former Soviet Union (Lake Baikal, the Aral Sea), where the immediate victims were effectively politically impotent, and from major oil spills. Particularly vulnerable major sites at present include the Arctic and Antarctic regions (where the international attempt at environmental protection has been especially strong, not least because initial territorial appropriation was still minimal), and the Caspian Sea, which is both a fully enclosed water body and the site of vast oil and gas reserves in a populous and still relatively impoverished region.

Marine pollution remains extraordinarily difficult to police, with the relation between those who inflict and incur its costs in most instances impossible as yet to turn into structures of clear mutual accountability. But it is reasonable to expect much further effort to make them more so, because the interests in marine pollution cannot in principle, over time and under accurate interpretation, outweigh the interests in preventing it. Since the interests in its continuance for the most part also lack tight and firm links to major state powers, and the technical possibilities for enhanced surveillance must be virtually limitless, in the end this effort is likely to win some success. Only in the case of the oil industry is it hard to imagine rapid progress in arresting marine damage; and here the explanation is less the sinister political linkages of major oil producers than the virtually worldwide mass interest in continued access to low cost energy supplies, especially for transport, cooking, and over much of the world, heating. It is not an accident that the most prominent and entrepreneurially dynamic executive of a major oil company over the last few years, Sir John Browne of BPAmoco, should also be the most prominent exponent of the imperative for the industry to respond effectively to the cumulative threat of its environmental damage.

In Western Europe cross-border acid rain was the first major political focus for environmental concern and resentment in the post-war epoch, with the forests and lakes of Scandinavia as the most evocative site of harm, and the principal perpetrators seen as the cars, trucks and power stations of the UK and Germany (Held *et al.*, 1999, 404–405). Due to the relative diplomatic and economic intimacy of Western Europe, some degree of accountability for these harms was established with comparative ease, and sustained effort has been made for at least a decade to modify the technologies of power stations and road transport to cut levels of nitrogen and sulfur dioxide emissions. A more recent and alarming example of cross-border atmospheric pollution was the month-long smog which engulfed much of the Malay peninsula and Northern Indonesia in the summer of 1997 as a result of Indonesian forest fires. Here, the prospect of rapid adjustment to avert further damage seems poor, since the level of cooperation and mutual confidence embodied in the South–East Asian Nations (ASEAN) is still distinctly lower than in the case of the European Union, the immediate economic circumstances of the majority of the populations in question already straitened, the dynamics of economic growth severely disrupted, and the political coherence and efficacy of the Indonesian state still far below those of any Organization for Economic Cooperation and Development member. An even less promising site for prompt rectificatory action is the atmospheric pollution induced by traffic, power stations, and in many instances wood-burning, in the vast metropolitan urban agglomerations of the world's poorer countries, from Delhi to Mexico City.

Paradoxically, the most encouraging instance of political response to global environmental damage has come

from one of the more recently identified and unnerving examples – the damage to the ozone layer, with its sharp prospective increase in harmful radiation. The source of that damage has been the use over the last half century of chlorofluorocarbons (CFCs), principally in refrigeration (Held *et al.*, 1999, 396–397). The decisive evidence for the scale and immediacy of this threat came from scientists of the British Antarctic Survey (Rajan, 1997, 44–45), gathered between 1982 and 1984 and published in 1985. The complex process of initial international political response to this evidence, commercial haggling over, and subsequent enterpreneurial adaptation to, its implications by major producers of CFCs, and political bargaining between wealthier and poorer countries over distributing the costs of adjustment, was bound to take some years (Rajan, 1997, Chapters 3 and 4). But in retrospect, the time-span between the publication of the British Antarctic Survey team's article in Nature in May 1985, the signature of the Montreal Protocol restricting CFC use in September 1987 (Rajan, 1997, 54), and its coming into effect in an amended version late in the summer of 1992, in a form which took better account of the interests of the poorer countries, was remarkably brief. We do not know quite why the response was this speedy. The startlingly grim implications of the Antarctic data? The unambiguous source of the damage? The relatively small number of major producers of CFCs and the clear (if expensive) availability of means to replace them with a far less damaging alternative? But the combination of shock and available remedy does offer some insight into how we might hope to alleviate other equally clear but less imminent menaces at a less hectic pace, where the costs of doing so are not too forbidding.

Deforestation, and loss of biodiversity through this and many other means, have proved far harder to arrest. Assessing responsibility for each, and allocating the costs of even attempting to bring them to an end, have been bitterly politically contentious, both within many individual states, and between groupings of wealthier and poorer states across the world. The complexities of perception and ambivalences of sentiment evoked by each reach back as far as civilization itself, and are inextricable from the history of class or ethnic conflict in most human societies (Thompson, 1975; Cronon, 1983). With the massive human population growth of the last half century and its prospective extension into the next few decades, levels of deforestation, to which the English or Italians became accustomed several centuries ago, and which the subcontinent of South Asia has undergone over the last century, now prospectively threaten the great rain forests of West and Central Africa, and even in the end the Amazonian basin itself. The extension of agriculture and stock-raising at the expense of forest cover, so long seen as the least equivocal index of human advance, now appears not as a single unambiguous good, but a potentially ever less coherent sequence of increasingly wanton actions. The political impetus to persist with each remains extremely powerful; but the overall practical rationality of doing so is steadily harder to defend. No area of world trade today is as dominated by the political power of sinister and consequentially damaging interests as the agricultural exports of North America and the European Union, with their massive internal subsidization. The relation between markets and food supplies has been morally fraught throughout modern history; and even the most compelling case for laissez-faire in the field of food is hard to reconcile with the levels of market distortion locked into the routine politics of most long-established capitalist democracies. This is not an aspect of human social existence in the world today which any extant ideology articulates clearly and evaluates compellingly, though there is, of course, widespread and sometimes sophisticated understanding of the harms and benefits that flow from the bewildering variety of individual practices.

The most effective arguments for the imperative of preventing further deforestation are prudential, and center on its prospective impact on global warming, and hence on climate change. The gap between recognizing the force of the arguments and halting (or even reversing) deforestation is immense, and certain to remain for a long time as economically perturbing as it is politically challenging. But at least deforestation is a relatively salient phenomenon, something which could, in principle, in due course be tracked accurately by satellite camera (*see* Global Forest Watch, Volume 4). The loss of biodiversity is inherently more nebulous, and far harder to measure until it is already too late. The prudential arguments for the urgency of arresting it involve far more speculative judgements, many of them about future scientific advances, while the normative arguments for the same conclusion, stressing the enormity of eliminating natural differentiation itself, have little solid foundation in the great majority of extant human cultures, and receive scant tendential support from the spread of the industrialized economies and ever more commoditized forms of life. It is hard to be confident, in the case of plant, marine or animal life, that net loss of biodiversity will slow in the near future, and just as hard to believe that the domestic political judgement and diplomatic finesse of states today equip them to focus effectively on this issue and cooperate firmly and deftly with one another in responding to it (Rajan, 1997, Chapters 7 and 8).

The most politically interesting and potentially consequential dimension of ecological degradation so far identified is the process of global warming through the combustion of fossil fuels, and its prospective impact on climate across different human habitats. Whilst the scale and timing of disruption are thus far impossible to estimate accurately, it is striking how rapidly the levels of risk ultimately at stake, have come to be accepted, even by those with the clearest immediate stake in continuing

to deny them (Houlder, 1999a). In marked contrast with British governmental vacillation over the hazards presented by genetically modified food crops (Houlder and Wigton, 1999), and the collapse of the UN Cartagena meeting in February 1999 on the biosafety protocol in the face of US insistence on pressing its entitlement to export genetically modified crops under the rules of the World Trade Organization (1999), there now appears to be little coherent governmental opposition in international negotiations to the need to lower emissions of greenhouse gases and enhance the efficacy of energy use with some rapidity. This is certainly not an index of the limited scale or political importance of the interests which will suffer directly and immediately from any such adjustment. In each case, these could hardly be more extensive. What it reflects, above all, is the universality of the interest in having a viable habitat, the imponderability of the scale of risk in question, and the relative simplicity and obtrusiveness of the mechanism which will initiate the harm (however complicated its subsequent mediation proves to be).

Unlike older and starker preoccupations like desertification (where there has been extensive international coordination to limit the threat, under the auspices of the UNEP (Elliott 1998, 90–94), or shrinking or degenerating water resources (Houlder, 1999b), what drives home the significance of global warming is not its immediacy or concreteness, but the potential limitlessness of its incidence and the comprehensiveness of its scope. It is these two together which explain the density and tenacity of the diplomatic negotiations which led up to the UN Framework Convention on Climate Change of May 1992 and its signature at Rio, and which overrode the very limited practical gains of the Convention itself (Elliott, 1998, 68–69) and the sharp divisions of interest which had ensured these, to establish a far more definite (and perhaps even prospectively effective) set of mechanisms for addressing the issue a mere five years later in the December 1997 meeting which drew up the Kyoto Protocol (Grubb *et al.*, 1999). A core provision of the Kyoto Protocol is the system for trading permits for the provision of greenhouse gases, which is intended to redistribute the physical burden of limitation in such a way as to meet the agreed limits, rapidly and at the lowest net cost in individual welfare or economic growth foregone. Any process of bargaining and coordination which involves parties with highly unequal assets is likely to generate results which are disquieting on grounds of justice (Barry, 1989). But no large-scale political process of this kind can be conducted behind a veil of ignorance. To produce any measure of concrete agreement, and still more to secure agreements which stand any real chance of being implemented in practice, it is essential to identify outcomes which reflect a clear common interest, which is evidently more important than the necessarily even clearer conflicts of interest over how the costs of securing it are distributed between the parties. Only extended experience of trading emissions permits, in the context of an evolving international trading regime and a world economy whose dynamics are certain to remain highly erratic for the foreseeable future, will show whether the Kyoto Protocol has established a formula for adjusting global production over time to the need to lower fossil fuel consumption in the longer run.

What is already clear is that the Kyoto Protocol, even if observed, will at most slow the accumulation of greenhouse gases for many decades to come, and that the process of coordination is certain to remain arduous across the different forms through which it must be pursued – from individual national industries or nation states, through economic blocs like the European Union, the North American Free Trade Agreement (NAFTA) or ASEAN, to the global panoply of UN organizations. The unintended effects of policy instruments like national energy taxes on the international competitivity of high energy use industries such as steel show the indispensability of coordinating widely. The laborious and often opaque activity of diplomatic haggling throughout the global negotiations on climate change of the last two decades, shows the proliferating costs in time, patience and analytic effort in every further attempt to coordinate policies. Strong *a priori* grounds for doubting whether coordination in face of sufficiently discrepant structures of interest can be either possible or rational (Olson, 1965; Hardin, 1982, 1995; but cf Axelrod, 1997) must be offset by weaker inductive grounds for recognizing that a great deal of coordination does in fact sometimes occur. In the end, what does and will decide the degree of coordination is the internal process of political judgement itself, a process led by active cognitive interest, but also distorted throughout by personal folly and sometimes far from benign passions (Dunn, 2000).

What has emerged into politics with the agenda of global environmental change is an overwhelming new political challenge. The most myopic of us can readily see that this challenge is bound to intensify many old political problems, and not unlikely to impair some of our existing resources of organization and mutual confidence for dealing with many components of our familiar political predicaments. What is less easy to see is whether the challenge itself might, or could not, generate fresh resources, whether of organization or of mutual confidence, with which to try to confront it. Hardest of all is to judge how the balance between sharpening challenge, faltering institutions and weakening mutual confidence, and tentative genesis of new institutional structures and changed sensibilities and dispositions, is likely to come out over time. But even this is best seen in the end, less as a question about how to interpret the human world, than about how to change it for the better. It is not

a scientific question with an, as yet unknown, scientific answer, but a practical question which can be answered only in practice.

It will, of course, be answered by what we and our descendants prove to do, and fail to do. There is every reason to fear that it may be answered very badly, as, in a sense, it now turns out, it has been answered for most of recorded human history. Since, for the first time in that history, we now have good reason to believe that we hold the fate of our own species fully in our own hands, there is clear urgency to learn to answer it far better. This is a challenge not just to our moral imaginations and limited self-control, but to the very core of our practical good sense: our will and capacity to preserve ourselves individually and collectively over time. Above all, today and for the imaginable future, it is a challenge to our political judgement (Dunn, 1998).

REFERENCES

Alperovitz, G (1995) *Atomic Diplomacy*, 2nd edition, Pluto Press, London.

Axelrod, R (1997) *The Complexity of Cooperation: Agent-based Models of Competition and Collaboration*, Princeton University Press, Princeton, NJ.

Baldwin, D A (1992) *Economic Statecraft*, Princeton University Press, Princeton, NJ.

Barry, B (1989) *Theories of Justice*, Harvester Wheatsheaf, London.

Brenner, R (1998) The Economics of Global Turbulence: a Special report on the World Economy 1950 1998, *New Left Rev.*, **229**, May June.

Brenton, T (1994) *The Greening of Machiavelli*, Earthscan, London.

Carrithers, M, Collins, S, and Lukes, S, eds (1985) *The Catgeory of the Person*, Cambridge University Press, Cambridge.

Clark, I and Wheeler, N (1989) *The Origins of British Nuclear Strategy*, Oxford University Press, Oxford.

Constant, B (1988) *Political Writings*, ed B Fontana, Cambridge University Press, Cambridge.

Cronon, W (1983) *Changes in the Land: Indians, Colonists and the Ecology of New England*, Hill and Wang, New York.

Cronon, W (1992) *Nature s Metropolis: Chicago and the Great West*, Norton, New York.

Dasgupta, P and Maler, K G, eds (1997) *The Environment and Emerging Development Issues*, 2 Volumes, Clarendon Press, Oxford.

Dunn, J (1969) *The Political Thought of John Locke*, Cambridge University Press, Cambridge.

Dunn, J (1990a) *Interpreting Political Responsibility*, Polity Press, Cambridge.

Dunn, J, ed (1990b) *The Economic Limits to Modern Politics*, Cambridge University Press, Cambridge.

Dunn, J, ed (1992) *Democracy: The Unfinished Journey*, Oxford University Press, Oxford.

Dunn, J (1996) *The History of Political Theory*, Cambridge University Press, Cambridge.

Dunn, J (1998) Democracy, Globalization and Human Interests, *Il Politico*, **43**, 353 374.

Dunn, J (1999) How Politics Limits Markets: Power, Legitimacy, Choice, in *The Politics and Economics of Power*, eds S Bowles, M Franzini, and U Pagano, Routledge, London, 85 100.

Dunn, J (2000) *The Cunning of Unreason: Making Sense of Politics*, Harper Collins, London.

Elliott, L (1998) *The Global Politics of the Environment*, Macmillan, Basingstoke.

Finer, S E (1997) *The History of Government*, 3 Volumes, Oxford University Press, Oxford.

Freedman, L (1981) *The Evolution of Nuclear Strategy*, Macmillan, London.

Gilpin, R (1987) *The Political Economy of International Relations*, Princeton University Press, Princeton, NJ.

Gombrich, R (1971) *Precept and Practice: Traditional Buddhism in the Rural Highlands of Ceylon*, Clarendon Press, Oxford.

Gowing, M and Arnold, L (1974) *Independence and Deterrence: Britain and Atomic Energy 1945 1952*, 2 Volumes, Macmillan, London.

Grubb, M, Vrolijk, C, and Brack, D (1999) *The Kyoto Protocol: a Guide and Assessment*, Earthscan, London.

Hardin, R (1982) *Collective Action*, Johns Hopkins University Press, Baltimore, MD.

Hardin, R (1995) *One for All: the Logic of Group Con ict*, Princeton University Press, Princeton, NJ.

Haugland, T, Bergesen, H O, and Roland, K (1998) *Energy Structures and Environmental Futures*, Oxford University Press, Oxford.

Held, D, McGrew, A, Goldblatt, D, and Perraton, J (1999) *Global Transformations: Politics, Economics and Culture*, Polity Press, Cambridge.

Houlder, V (1999a) BPAmoco Chief Sets Out Case for Energy Taxes, *Financial Times*, 13th January, 6.

Houlder, V (1999b) UN Warns of Water Wars Next Century, *Financial Times*, 19th March, 4.

Houlder, V and Wigton, D (1999) UK Moratorium on Modi ed Crop would Break EU Law, *Financial Times*, 19th February, 20.

Hurrell, A and Kingsbury, B, eds (1992) *The International Politics of the Environment*, Clarendon Press, Oxford.

Kennedy, P (1988) *The Rise and Fall of the Great Powers*, Unwin Hyman, London.

Keohane, R (1984) *After Hegemony*, Princeton University Press, Princeton, NJ.

Lal, D and Myint, H (1994) *The Political Economy of Poverty, Equity and Growth: a Comparative Study*, Clarendon Press, Oxford.

Little, I M D (1982) *Economic Development: Theory, Policy, and International Relations*, Basic Books, New York.

Mabey, N, Hall, S, Smith, C, and Gupta, S (1997) *Argument in the Greenhouse: the International Economics of Controlling Global Warming*, Routledge, London.

MccGwire, M (1987) *Military Objectives in Soviet Foreign Policy*, Brookings Institution, Washington, DC.

McNeill, W (1983) *The Pursuit of Power: Technology, Armed Force and Society since AD 1000*, Basil Blackwell, Oxford.

Manin, B (1997) *The Principles of Representative Government*, Cambridge University Press, Cambridge.

Meinecke, F (1957) *Machiavellism*, translated from the German by D Scott, Routledge and Kegan Paul, London.
Olson, M (1965) *The Logic of Collective Action*, Harvard University Press, Cambridge, MA.
Parfit, D (1984) *Reasons and Persons*, Clarendon Press, Oxford.
Passmore, J (1974) *Man's Responsibility for Nature*, Duckworth, London.
Popper, K R (1957) *The Poverty of Historicism*, Routledge and Kegan Paul, London.
Quinton, A (1999) *My Son the Philosopher (review of Randall Collins, the Sociology of Philosophies)*, New York Review of Books, 46, 8/4/99.
Rajan, M G (1997) *Global Environmental Politics: India and the North- South Politics of Global Environmental Issues*, Oxford University Press, Delhi.
Runte, A (1997) *National Parks: the American Experience*, University of Nebraska Press, Lincoln, NE.
Swanson, T M, ed (1996) *The Economics of Environmental Degradation: Tragedy for the Commons?* Edward Elgar, Cheltenham.
Taylor, C (1989) *Sources of the Self*, Harvard University Press, Cambridge, MA.
Taylor, C (1992) *The Ethics of Authenticity*, Harvard University Press, Cambridge, MA.
Thomas, C (1994) *Rio: Unraveling the Consequences*, Frank Cass, London.
Thomas, K (1983) *Man and the Natural World: Changing Attitudes in England 1500- 1800*, Allen Lane, London.
Thompson, E P (1975) *Whigs and Hunters*, Allen Lane, London.
Thucydides, V (1921) *History of the Peloponnesian War*, translated by C F Smith, Harvard University Press, Cambridge, MA.
Tolba, M and Rummel-Bulska, I (1999) *Global Environmental Diplomacy: Negotiating Environmental Agreements for the World*, MIT Press, Cambridge, MA.
Unger, R B (1999) *Democracy Realized*, Verso, London.
Vogler, J and Imber, M (1996) *The Environment and International Relations*, Routledge, London.
Wallerstein, I (1995) *After Liberalism*, The New Press, New York.
World Trade Organization (1999) *Financial Times*, 26th February.

FURTHER READING

McCormick, J (1995) *The Global Environmental Movement*, 2nd edition, John Wiley & Sons, Chichester.

The Environment and Violent Conflict

DANIEL M SCHWARTZ

University of Toronto, Toronto, Canada

Increasingly, violent con ict - and the potential for such con ict - is being linked to environmental problems such as water pollution, over- shing, deforestation, and global warming. Research suggests that it is developing nations that are most vulnerable to this environmentally induced con ict.

Environmental stress is rarely considered to be the sole factor in precipitating con ict, which can take the form of both con ict between and within nations. In many cases environmental stress is a relatively distant factor, acting in combination with other economic and social factors such as poverty and weak governments. In other cases, con ict breaks out when rival nations, or rival groups within a nation, battle for diminishing supplies of environmental resources. Although environmental stress is usually only one cause of con ict among many, the evidence suggests that it can play an important role, and that violence may be avoided by addressing environmental problems.

Interest in the linkages between environmental stress and violent con ict has grown rapidly in the last two decades. In part, this is because the end of the Cold War has shifted discourse on the issue of security in both academia and policy circles. Although traditional threats to security - such as conventional and nuclear war - still exist, there has been an increasing recognition that security is also tied to less apparent dangers such as chronic poverty, infectious disease, and the depletion and degradation of environmental resources.

INTRODUCTION: THE CHANGING NATURE OF SECURITY

In 1977, Lester Brown of the Worldwatch Institute published *Rede ning National Security*, a pioneering work, which argued that the most significant security dilemma facing humankind was an ever-widening gap between the supply of, and demand for, environmental resources (Brown, 1977). This idea challenged mainstream conceptions of security and launched a debate on the topic of redefining security that has grown increasingly sophisticated over the last two decades.

In the mainstream schools of thought that focus on the issue, security is (and always has been) largely synonymous with state security. In a world comprised of self-interested and calculating individuals, the state is seen as the purveyor of authority and security. Without the state, individuals would live in what 17th century philosopher Thomas Hobbes termed a state of nature, and life would be solitary, poor, nasty, brutish, and short. Moreover, mainstream security analysts believe that the greatest challenge to state security comes from outside the state in the form of military force or organized violence. The most effective means for a state to repel an outside threat of force is to engage the enemy with military force of its own.

Academics and policymakers who agitate for a redefinition of security, contest these assumptions made by mainstream analysts. Some proponents of redefining security argue that the focus of security should be shifted away from the state, and nearly all contend that threats to security should include non-military ones – especially environmental stress. These proponents champion the concept of environmental security to challenge the near monopoly that mainstream security analysts have held in the past.

Environmental security analysts that encourage a move away from state security, favor either a shift upwards to the global level, or downwards to the individual level. A shift upwards to the global level is encouraged by those who contend that environmental problems such as global warming and ozone depletion transcend national boundaries and highlight the oneness of a world divided along arbitrary and increasingly meaningless political boundaries. Those who favor a shift downward to the individual (or human) level, assert that environmental problems such as urban pollution often affect human wellbeing more than they affect an abstraction such as the state. Other environmental security analysts, however, do not favor a shift away from state security. They insist that as long as the state continues to be the central organizing principle in global politics, state security should continue to be the focus.

Although some environmental security analysts do not agitate for a movement away from state security, most maintain that non-military threats to security are at least as important as military threats. They point out, for instance, that the annual toll of individuals suffering from infectious diseases, such as the Ebola virus and AIDS, far outnumbers the global number of battlefield casualties in the last decade (Pirages, 1996, 13). In short, these analysts contend that security must not be measured only in terms of military might, but also in terms of economic wellbeing, social stability, and environmental health.

Environmental security analysts have made some tangible headway in their attempt to redefine security. At the national level, for instance, the US Department of Defense, in 1993, created the position of Deputy Under Secretary for Environmental Security. At the international level, the North Atlantic Treaty Organization and the United Nations Environment Programme have explicitly recognized the issue of environmental security (Schwartz and Singh, 1999; Gleditsch, 1997).

But efforts to redefine security have also been met with suspicion and outright scorn. Skeptics argue that proponents of environmental security are mere opportunists, dressing up environmental concerns in security garb, with the intention of shifting resources away from military budgets to environmental budgets. Moreover, critics contend that too broad a definition of security renders the concept meaningless for all intents and purposes.

But there is yet another way in which environmental stress and security may be inextricably linked. Recent research has discovered various ways in which environmental stress can lead to the outbreak of violent conflict. Environmental stress can contribute to resource shortages, which can provoke conflict between nations and induce conflict within nations. Environmental stress can also engender and interact with economic and social phenomena – such as poverty and weak governments – to create conflict between and within nations. These linkages are discussed in detail below. For our immediate purposes, however, it is important to note that even if we hold to a mainstream conception of security – in which military threats prevail – the environment-conflict nexus suggests that the environment can be an important component of security. If environmental stress can in fact contribute to violent conflict within and between nations, then even the staunchest of mainstream security analysts must recognize the importance of environmental issues.

ENVIRONMENT AND VIOLENT CONFLICT

The notion that disputes and violent conflict can erupt over access to environmental resources appears commonplace: international wars and civil conflicts have been fought over access to land and resources since time immemorial. Three factors, however, make modern-day analyses on environment and conflict novel. First, there has been a shift in focus from non-renewable resources such as oil and minerals, to renewable resources such as water, cropland, forests, and species. Most experts concur that conflicts over non-renewable resources will continue to erupt. The Gulf War, for instance, was fought, at least in part, over access to oil. But for the first time, renewable resources are being considered as systematic factors in the explosion of hostilities.

Second, modern-day analyses on environment and conflict emphasize the role of population growth in fostering environmental stress. Throughout this essay, I use the term environmental stress (or environmental problems) to indicate the degradation and depletion of renewable resources. Although I will not make explicit reference to it on every occasion, I consider population growth to be intimately connected to environmental stress. As a population grows, the demand on environmental resources increases; and as the demand on environmental resources increases, the abundance of the resource itself shrinks while the existing supply is often over-taxed. Although some demographers have recently predicted that population growth will actually begin to decline worldwide in approximately 50 years, there is also an overwhelming consensus that the global population will hit between 7.5 and 10.5 billion in the year 2050. With such large increases predicted for the near future, population growth will continue to be intimately connected to environmental stress.

Finally, as will be discussed in detail below, modern-day analyses on environment and conflict often focus on the manner in which environmental stress can combine with other economic and social factors to produce conflict. Whereas past analyses focussed solely on conflict over environmental resources, modern-day analyses also assess how the degradation and depletion of renewable resources can act in conjunction with social and economic conditions to produce conflict between and within nations.

In what follows, I examine the various ways in which environmental stress can precipitate violence. Before examining these linkages, however, I distinguish between wide-ranging environmental stress and localized environmental stress. By wide-ranging, I refer to environmental problems that are affecting the entire globe or large areas of the world simultaneously. Two prominent examples are global warming and depletion of the stratospheric ozone layer. By localized environmental stress, I refer to environmental problems that are affecting small areas of the globe at different times. Examples include deforestation and water pollution.

This distinction notwithstanding, it is important to emphasize the global nature of all environmental stress. The interconnectedness of the environment makes it likely that even the effects of localized environmental problems will spread beyond their initial boundaries. Water

pollution, for instance, can easily spread beyond the immediate boundaries of its inception through groundwater and tributaries. Moreover, global economic and social forces make it likely that the roots of localized environmental problems can be traced back to forces outside of the area immediately affected. Deforestation in many developing nations, for instance, is partly the result of growing demand for lumber and wood by-products in developed nations.

INDIRECT INTERNAL CONFLICT

Evidence to date indicates that environmental stress is most likely to result in conflict when it engenders and interacts with social and economic factors to produce conflict within a nation. Because the evidence for this type of conflict is abundant, and because the processes identified here will shed light on all of the various conflict types discussed below, I will spend a good deal of time explicating indirect internal conflict.

I consider this type of conflict to be indirect, because conflict does not arise over a particular environmental resource. Rather, the conflicts are often typical ones, including ethnic conflicts, class conflicts, revolutions, insurgencies, etc. But the cause of these conflicts can often be traced back, in part, to environmental stress.

Conflict is internal when environmental stress produces violence within the boundaries of a nation. The process that leads to conflict can often have cross-border dimensions, such as when cross-border migrations that are induced by environmental stress play a role in creating conflict. And, as noted above, local environmental stress itself is often the result of global environmental change and economic forces. Nevertheless, if the conflict itself is confined to within the perimeters of a particular nation, we can consider the conflict to be an internal one.

There are at least five pathways to indirect internal conflict that involve environmental stress: economic decline, migrations, social segmentation, erosion of civil society, and curtailment of the state. These pathways rarely work in isolation from one another. In most cases, the various roads to conflict are very complex and involve elements from a number of pathways working in synergy. To better understand the role that environmental stress plays in engendering conflict, however, it is useful to parcel out these pathways and examine each one individually before considering the ways in which they criss-cross one another to form the broader roads to conflict.

Economic Decline

Perhaps one of the most evident and widely cited social effects of environmental stress is economic decline. There are many types of environmental stress that can contribute to economic decline, including wide-ranging environmental problems such as global warming and stratospheric ozone depletion, as well as more localized environmental problems such as deforestation, water pollution, and over-fishing. Although environmentally induced economic decline can lead to general economic malaise in society, it is more likely to affect the poorer segments of society and thus widen the differential between rich and poor. This is because it is the poorer segments of society that are more likely to be highly dependent on environmental resources, and less able to buffer themselves from the effects of environmental stress (Homer-Dixon, 1999, 88).

It is essential to keep in mind that environmental stress is only one cause among many contributing to economic decline and widening income gaps. Nevertheless, it appears that it can be an important factor. Global warming, for instance, could lead to a rise in sea level and produce more frequent extreme weather events such as hurricanes and tornadoes. These events, in turn, could cause major property damage to coastal regions and result in the injury to, and/or loss of, human life. Global warming may also lower precipitation in some regions and yield pest infestations. These effects would strain the economic resources of local regions and nations. Myers (1993, 174), for instance, contends that global warming could cause the United States' grain belt to become unbuckled, with as much as half of the farmlands in the Southeast failing, California running out of water to irrigate its huge areas of cropland, and many of the country's woodlands and forests fading from the scene. Most analysts, however, including Myers, note that it is developing nations that will likely bear the brunt of global warming. For example, regions that are already plagued with fragile environments, such as the Sahel and the Indian subcontinent, could face severe droughts that significantly undermine their agricultural productivity.

Stratospheric ozone depletion can also induce conditions that lead to economic decline. For instance, by depressing immune systems, stratospheric ozone depletion can increase disease rates in humans and animals. Some experts suggest that recent increases in the occurrence of skin cancers and eye cataracts in the Southern Hemisphere are already attributable to depletion of the stratospheric ozone layer. Stratospheric ozone depletion can also adversely affect agricultural productivity. Studies suggest, for instance, that stratospheric ozone depletion could reduce crop yields by anywhere from 5% for wheat to 90% for squash. Finally, and perhaps most ominously of all, stratospheric ozone depletion can wreak havoc on entire ecosystems. Phytoplankton-based food chains in the sea, for instance, could be disrupted. If phytoplankton are affected, higher levels in the food chain – such as zooplankton and fish – would likely be affected as well (Myers, 1993, 166–167).

If these impacts from stratospheric ozone depletion materialize – increased disease rates in humans, declining

agricultural productivity, and ecosystem damage – government coffers would inexorably feel the pinch between increasing demand on national and regional health care systems, and decreasing revenues from agricultural and fishing sectors. Although developed nations would hardly be immune to the impact of stratospheric ozone depletion, it is again developing nations that would likely suffer the direst economic consequences because they often lack the infrastructure and economic flexibility to mitigate the impacts.

If the potential ramifications from global warming and stratospheric ozone depletion are sound in theory, there is little documented evidence to date that would support the connection between these wide-ranging environmental problems and economic decline. The socio economic effects of global warming and stratospheric ozone depletion, quite simply, have not yet begun to play themselves out in any systematic way, although there are indications that this is beginning to happen.

The link between more localized environmental stress and economic decline, conversely, has been evident for years. Many regions, especially in the developing world, are already feeling the financial sting of localized environmental stress.

Deforestation, for instance, has had particularly pernicious consequences for regional and national economies. Large-scale unsustainable logging practices may reap short-term rewards for a nation, but these logging practices can also sow the seeds of long-term economic decline once initial supplies are exhausted. Barber (1997, 53–54), for example, argues that Indonesia's unsustainable deforestation practices over the last three decades created huge amounts of wealth, but also sacrificed future profits that are now sorely missed by the nation's economy. Deforestation can also reduce the available supply of fuel wood – the wood that many people in developing nations are highly dependent upon for their daily energy requirements. As the supply of wood dwindles, prices rise, forcing poor families to spend an increasing portion of their budget on this resource. In Chiapas, Mexico, for instance, rapid deforestation of the Lacandon rainforest over the last two decades has deprived local citizens of much needed wood. The lack of fuel wood, in turn, has contributed to widespread poverty in the region (Homer-Dixon and Percival, 1996, 17–20). Finally, deforestation can foster silting, soil erosion, and desertification; change regional hydrological cycles and precipitation patterns; and decrease the land's ability to retain water during rainy periods. All of these effects limit agricultural productivity, cause damage to pivotal infrastructure such as roads and bridges, and reduce the hydroelectric capacity of rivers. Homer-Dixon (1999, 81–93) notes that in China, South Asia, Southeast Asia, and the Middle East, much cropland has been lost or degraded by erosion. And on the Filipino Island of Palawan, only half of the potential farmland is irrigable because of the hydrological effects of decreases in forest cover.

Water pollution has also created economic decline by limiting agricultural activity and fostering disease. These effects have been particularly acute in the Middle East. Homer-Dixon (1999, 90–92), for instance, contends that in Gaza, water aquifers have been salanized from over-pumping and severely polluted from agricultural run-off. The result has been damaged soil, declining crop yields, and severe reductions in potable water. Routine consumption of polluted water has led to widespread health problems in Gaza's Palestinian population, straining health care budgets and retarding economic development.

Finally, over-fishing has induced economic decline. Some developed nations have seen their economies eroded by the collapse of fisheries. Canada, for instance, has at various times imposed moratoriums on commercial cod fishing on the East Coast and severely restricted commercial salmon fishing on the West Coast. But once again, it is developing nations – especially regions that are highly dependent on their fisheries – which have been most affected by the depletion in fish stocks. For example, Lang (1995) traces regional and national economic decline in Kenya back to a combination of water pollution and over-fishing in Lake Victoria – a key body of water for the nation and a pivotal source of revenue for many citizens.

Migration

A second pathway from environmental stress to conflict is through migration. Wide-ranging and localized environmental stress, in conjunction with other social and economic factors, can induce movements of people within national borders as well as between nations. Sometimes large and spontaneous migrations are caused by environmental factors, such as floods, droughts, and locusts. Flooding in Mozambique and Madagascar during February and March, 2000, for instance, caused large and rapid movements of people in these African nations. The degradation and depletion of renewable resources, however, is often a distant cause of migration. It may, for instance, contribute to economic decline, which subsequently contributes to migration. Further, this type of environmentally induced migration rarely involves large and spontaneous upheavals of people. Rather, the migration occurs slowly over long periods of time. These two caveats notwithstanding, migration induced by the depletion and degradation of renewable resources can be an important pathway to conflict, especially in developing nations where people's livelihood is often tied intimately to these resources.

The availability of arable land is likely one of the most important factors contributing to migration. This is especially evident in developing nations where agriculture can constitute a sizeable portion of local economies. Of

course, it is not the availability of arable land *per se* that induces migration, but rather the economic prospects that correspond to this availability. A scarcity of arable land will create economic decline, which acts as a push factor and decreases the quality of life, thereby inducing migration. Moreover, the prospect of migrating to verdant regions, where land is not scarce, can create a pull factor by offering migrants an enhanced quality of life. Large in uxes of people can often lead to a host of new environmental problems in the migrants destinations, thereby perpetuating a cycle of environmental stress and migration.

A scarcity of arable land is often caused, in part, by environmental stress. Deforestation, in particular, can induce ooding, soil erosion, and deserti cation, which can then degrade the nutrient base needed for agriculture. Hazarika (1993) identi es just such a process in Bangladesh during the 1980s. By deforesting upstream areas in the Himalayas, downstream soil was exposed to rain and winds. Consequently, the ooding which naturally occurs in this area was signi cantly exacerbated, and land degradation ensued. Combined with rapid population growth, land degradation forced farmers to divide the remaining arable land into ever-smaller portions. As the prospects for farming became increasingly grim, Bengalis began to look across the border to the relatively plush lands in Northeast India. The result was a gradual movement of over 10 million Bengalis to Northeast India. Similarly, Renner (1996, 106) contends that massive deforestation and soil erosion played an important role in decreasing the agricultural productivity of peasants farming the highland areas in Ethiopia. As a result, large numbers of peasant farmers migrated to the Ogaden area in southeastern Ethiopia where the prospects for farming were perceived to be substantially better.

Environmentally induced land scarcity is not the only factor that can lead to economic decline and migration. Smil (1992, 17 19) posits that water and air pollution have engendered migrations from the interior provinces of China to the country s eastern and southern coastal regions especially coastal cities where prospects for earning a decent wage are widely considered to be far superior to those in the interior provinces. Similarly, Wegemund (1996, 307 311) points to a shortage of fuel wood (caused primarily by rapid deforestation) as well as a scarcity of water (caused in part by water pollution and in part by increasing demand on the resources) as two key environmental problems that have yielded signi cant migrations in Africa s Senegambian region.

Social Segmentation and Erosion of Civil Society

While studies on the linkages between environmental stress, economic decline, and migrations are quite common, the impact of environmental stress on social segmentation and civil society has yet to receive widespread attention. Homer-Dixon (1999, 96 98) demonstrates, however, that environmental scarcity can produce social segmentation by deepening divisions among ethnic, religious, and linguistic groups. Homer-Dixon also shows how environmental stress can fray the web of networks in society made up of nongovernmental organizations, unions, and community service organizations. These societal networks form the core of civil society, and are key to mediating the relationship between individuals and the state. Just as migrations can occur as a result of both environmental stress and economic decline, social segmentation and the erosion of civil society can occur as a result of a complex interaction between environmental stress, economic decline, and migrations. Again, environmental stress in these cases is a distant but potentially important cause.

Homer-Dixon argues that environmental stress can contribute to a scarcity of resources, which can then deepen the fault lines between those who pro t from those scarce resources and those who are impaired by the scarcity. Environmental stress, he argues, encourages competition among groups for control of resources critical to survival and prosperity, and it encourages resource-dependent groups to turn inward and to focus on narrow survival strategies. This competition over resources often takes place against a backdrop of economic decline, which is also, in part, the result of environmental stress. Moreover, the vicious cycle between poverty and migration, discussed above, can also play a role by increasing competition for ever-scarcer resources in recipient nations.

Critical to his analysis of social segmentation, is what Homer-Dixon has called *resource capture*. Resource capture occurs when environmental stress and population growth combine to encourage powerful groups within a society to shift resource distribution in their favor. By shifting the resource distribution in their favor, these powerful groups capture the resource and prosper from speculation and monopolistic pro ts. Although this process increases the wealth of powerful groups in society, it also produces a signi cant scarcity of environmental resources for weaker and poorer groups in society. As competition for resources increases, tensions between rival groups within the poorer segment of society are heightened. Moreover, tensions are often in amed by powerful group leaders, who use divide and conquer tactics to solidify their clout.

Social segmentation is intimately tied to the erosion of civil society. When a society segments and the various factions turn inward, the myriad societal networks that are crucial to the functioning of a society can quickly disappear. Researchers have argued convincingly that these networks are crucial to the functioning of a society because they help to build bonds of trust between people of differing ethnic, religious, and linguistic backgrounds. Moreover, as these networks erode, civil society quickly weakens, thereby

reducing the ability of marginalized people to articulate their demands on the state.

These processes of social segmentation and erosion of civil society are illustrated by Homer-Dixon's analysis of events in South Africa from the mid 1980s to mid 1990s. With the collapse of Apartheid, large numbers of blacks began to migrate from their homelands – which often suffered from severe environmental deterioration and poverty – to the country's urban centers. The migrants settled in peripheral and ecologically sensitive lands in the urban centers, such as hillsides and river valleys. The huge influx of people onto these lands induced further environmental problems such as water pollution and deforestation. These environmental problems, in turn, promoted economic hardship and provided a window of opportunity for powerful warlords – local leaders who control their own paramilitary – to seize control of the remaining resources. With scarce resources firmly in their hands, warlords were able to illicit the support of their local groups, which were often divided along ethnic lines (e.g., the Zulu and Xhosa) and family-based lines. The warlords exploited these societal divisions in order to promote their own interests. Consequently, competition between local groups increased, social segmentation solidified, and the societal networks that form the core of civil society were severed.

Curtailment of the State

The fifth and final pathway from environmental stress to conflict is through curtailment of the state. Environmental stress, in conjunction with other social and economic variables – including economic decline, migration, and social segmentation – can curtail key institutions of the state. The state generally consists of the group of institutions in society that wield authority and make binding rules for the nation, or regions within a nation.

Homer-Dixon (1999, 98–102) identifies at least two ways in which environmental stress can curtail key state institutions. First, environmental stress may increase the financial burden on the state. Environmentally induced economic decline, for instance, may compel the state to increase the size of subsidies; and environmentally induced migrations into urban centers can greatly augment infrastructure requirements such as sewers and water piping. For instance, Homer-Dixon argues that to deal with acute water scarcity in China – induced by a combination of pollution and population growth – the state must spend huge sums on new infrastructure such as wells, dams, canals, pipelines, and irrigation systems; and it must build large facilities to control industrial and municipal pollution.

Second, environmental stress can affect state elites, such as high-ranking officials and business people with close connections to the upper echelons of decision-making in the state. In particular, environmental stress can threaten the incomes of elites that depend on the extraction of an ever-degrading and depleted resource base. Consequently, these elites may turn to the state for compensation, or they may act to block institutional reforms that would distribute more fairly the costs of rising scarcity. Alternatively, environmental stress can increase the scarcity and value of resources, thereby generating opportunities for powerful coalitions of elites to gain control of these resources in a process similar to that of resource capture, described above. As these elites become increasingly powerful, they can challenge the authority of the state. Increasing land scarcity in the southern Chinese province of Guandong, for instance, has created huge profits from land sales, inducing developers and corrupt officials to illegally control these revenues and ignore dictates from the state in Beijing.

The Roads to Conflict

The five pathways from environmental stress to conflict – economic decline, migrations, social segmentation, erosion of civil society and curtailment of the state – rarely work in isolation from one another. The pathways are often complex, and involve shared elements. For example, social segmentation can involve both economic decline and migrations, both of which can be induced by environmental stress.

But how do these pathways actually interact to produce conflict? Here the complexity gets ratcheted up another level. There simply are no necessary or sufficient causes of conflict, and there is no single pathway to conflict that one can consistently discern. Nevertheless, by mapping out the criss-crossing pathways and their complex interactions, one can begin to make out some broader roads to conflict. These roads can be traversed along many different combinations of pathways. Figure 1 illustrates the various combinations. Below I discuss a number of these possible combinations and provide some examples to illustrate the processes.

One possible road to conflict involves both economic decline and curtailment of the state. As noted above, environmentally induced economic decline rarely affects an entire society or nation uniformly. Rather, it usually widens the gap between rich and poor, which then fosters grievances among the poorer segments of society. If this element of society perceives the state to be the over-arching cause of the economic injustice they are experiencing, they might initiate conflict with the state in an attempt to reverse their plight. These challenges to the state can assume various forms, including armed rebellion, guerrilla warfare, terrorism, and *coups d'état*.

But challenges to the state are likely to occur only if challengers perceive the state to be vulnerable. Conflict with the state, then, is most likely to take place when a segment of society is aggrieved and when the opportunity for reversing

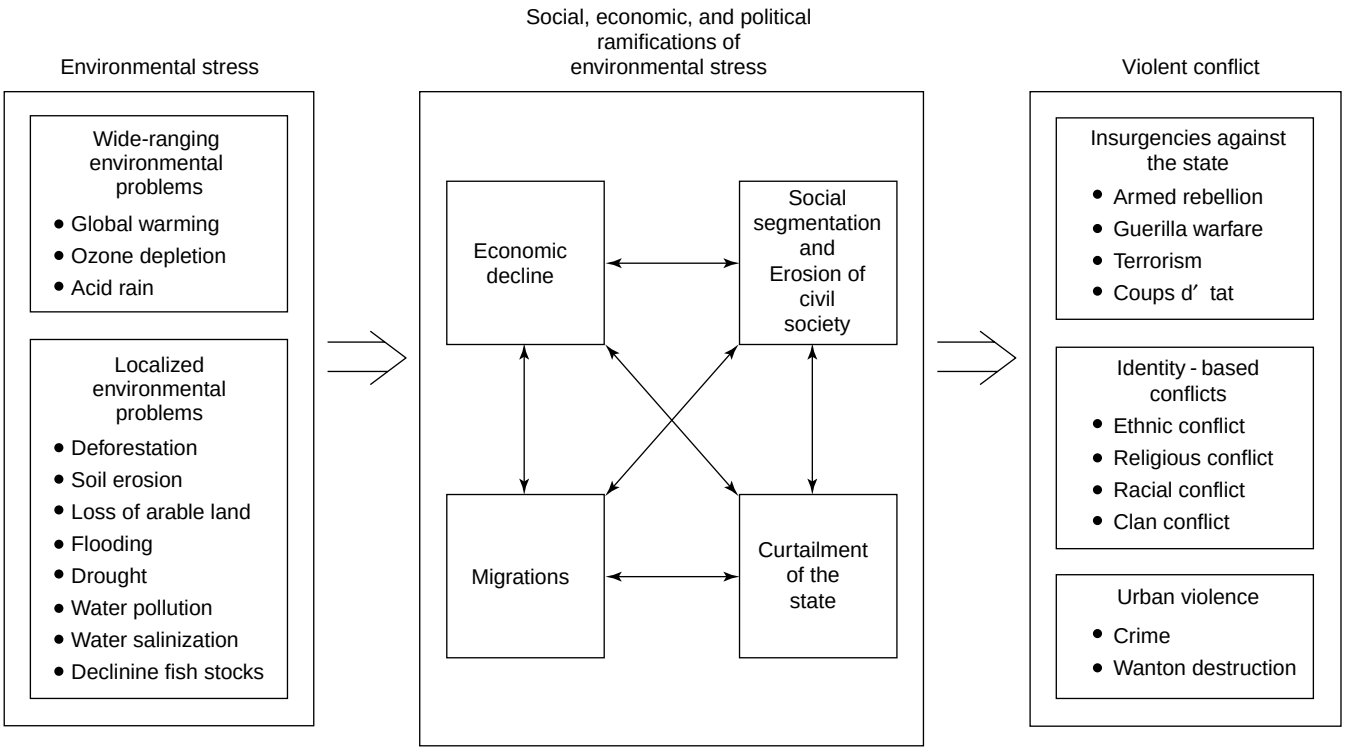


Figure 1 The roads to conflict

their status is present. Environmental stress, as discussed above, can influence both of these factors by contributing to economic decline and curtailment of the state.

Suliman (1992) contends that the road to conflict in Sudan has followed these pathways. Since 1983, violent attacks by the Sudan People's Liberation Army have been taking place against Sudanese state institutions, as well as businesses associated with the state. A majority of analysts trace the roots of this conflict to development projects conducted by the state – controlled largely by elites in Northern Sudan – in which resources are extracted from the South. Many Southern Sudanese consider these development projects to be a form of economic exploitation that would enrich only the state and its cronies.

Suliman, however, traces the roots of this conflict back further than most analysts. He contends that state-led development projects in Southern Sudan were initiated, in part, because deforestation and an increase in large-scale mechanized farming had induced desertification and land degradation in the northern regions of the country. Economic decline in Northern Sudan not only compelled the state to look south, but also strained their financial resources. Given the grievances of Southern Sudanese, and the compromised power of the state, many Sudanese have supported and participated in violent attacks against state institutions.

Another possible road to conflict involves economic decline, migration, and social segmentation. As discussed above, environmental stress can contribute to economic decline which, in turn, can engender small-scale but persistent migration. When migrants enter a new territory, they often upset the existing political balance and consequently deepen divisions in society. The result can be a society consumed by conflict.

The type of violence that this combination of events can trigger is often different from the conflicts discussed above, which featured attacks directed against the state. When social segmentation sets in, conflicts between identity-based groups are more likely to result. The identities upon which these groups are based include race, language, religion, and lineage.

Events in Northeast India during the 1980s illustrate this process. As discussed above, Northeast India received millions of Bengali migrants during the 1980s. These Bengali migrants were driven to the Indian states of Assam and Tripura, in part by environmental stress and economic decline. In Assam, Bengali migrants were perceived as economic competitors and religious outsiders. Moreover, Bengali migrants shifted the balance of political power by developing significant influence in local institutions. As a result of these processes, social segmentation hardened between the predominantly Hindu Assamese and the primarily Muslim Bengalis. More than 4000 lives were lost in ensuing violence between 1979 and 1985. In Tripura, a similar series of events created small-scale but persistent

violence between identity-based groups that continues to the present (Homer-Dixon, 1999, 141–142).

A final example of a possible road to conflict involves all of the pathways identified above – economic decline, migration, social segmentation, erosion of civil society, and curtailment of the state. This combination of processes can be especially pernicious when the migration involves large movements of people from rural areas to urban centers. Violence in these cases may not only take the form of insurgencies against that state and identity-based conflict; it can also take the form of urban violence that includes crime and wanton destruction. Because rapid urbanization is a characteristic common to many developing nations today, this process is especially salient.

Events in South Africa during the 1980s and 1990s illustrate this process. As described above, environmental stress in South Africa induced rural–urban migration, which combined with further environmental stress and economic decline to engender social segmentation. As a result of this social segmentation in South Africa's urban centers, societal networks were eroded or simply unable to take hold in the first place. Compounded by increasing demands on the state and a decreasing ability of the state to meet these demands, a number of urban centers in South Africa – most notable in the Johannesburg and KwaZulu–Natal regions – began to experience widespread political and criminal violence that persists today.

Homer-Dixon (1999, 164–166) contends that Pakistan has traveled a similar road to conflict. In Pakistan, water pollution, deforestation, and soil erosion have engendered economic decline and large movements of rural–urban migration. In Karachi, Pakistan's largest urban center, the massive influx of people has created urban slums that suffer a chronic shortage of basic urban amenities such as potable water, garbage disposal, and electricity.

This squalor and shortage of basic amenities, has induced resource capture – the process identified above whereby local group leaders gain control of a resource and prosper from speculation and monopolistic profits. In Karachi, this process has manifested itself in local mafias that obtain water from illegal hydrants or from poorer districts in the city and then sell it for exorbitant profit – often to the very people from whom the water was pilfered. This process of resource capture has not only furthered economic decline, but has also hardened social segmentation between Karachi's many competing ethnic groups, and eroded many of the societal networks that make up the city's civil society.

Underpinning all of these factors are crippled state institutions that cannot meet the increasing demands of the urban populations. Environmental stress has not only reduced Pakistan's state revenues but has spurred rural–urban migration that has simultaneously increased financial demands on local and national governments. Moreover, corruption and cronyism plague both national and local governments. With institutionalized channels of protest and action stymied, Pakistani loyalties and allegiances have become increasingly localized and social segmentation has further solidified.

The result of the interaction of all these pathways to conflict has been increasing urban violence. This violence has taken the form of all three types of conflict described above, including urban violence (e.g., rising crime rates), ethnic-based violence (e.g., conflict between the Punjabis, Pathans, and Sindhi ethnic groups), and violence directed at the state (e.g., attacks on the Karachi Electricity Supply Corporation).

INDIRECT INTERNATIONAL CONFLICT

Just as environmental stress can engender and interact with social, economic, and political factors to produce conflict within nations, a similar process can lead to conflict between nations. I call these conflicts indirect international conflicts.

International conflict between El Salvador and Honduras in 1969 involved some of the processes identified above. On July 14th, 1969, El Salvador launched an invasion of Honduras, commencing a battle that has been coined the Soccer War, because hostilities were launched immediately following a series of hotly contested World Cup matches. The conflict lasted only 100 days, but claimed the lives of several thousand individuals, turned 100 000 into homeless and jobless refugees, and inflicted severe economic losses on both nations.

Although many pundits have blamed the outbreak of conflict on a bitter rivalry cultivated during the World Cup soccer matches, Durham (1979) shows that environmental stress was one of the key underlying factors contributing to this international war. Deforestation and soil degradation, in conjunction with rapid population growth, forced landless peasants in El Salvador to migrate to Honduras, where many poor Hondurans were already struggling with their own scarcity of arable land. Large landowners in Honduras, who had previously been held responsible by poor Hondurans for their economic plight, successfully redirected blame onto the newly arrived El Salvadoran migrants. Hondurans reacted by expelling the El Salvadorans, which then prompted the government of El Salvador to officially close its borders to Honduran refugees and eventually declare war on its neighbor.

Molvoer (1991) contends that a similar process of environmental stress and migration combined to produce conflict between Somalia and Ethiopia in 1977–1978. Overgrazing induced widespread deforestation and desertification in Somalia, which prompted large migrations of Somali pastoralists into Ethiopian territory. The migrations brought the Somali pastoralists into increasing competition with Ethiopian groups dependent on local resources.

The bitter competition between these groups fuelled tensions between Ethiopia and Somalia that eventually found its outlet in armed conflict between the nations.

These two cases suggest that environmental stress – especially localized environmental problems such as deforestation and soil degradation – can contribute to conflict between nations in much the same way that environmental stress can contribute to internal conflicts. Just as with the internal conflicts described above, it is developing nations that are most at risk of succumbing to these forces, because these nations are often highly dependent on environmental resources and lack the ability to mitigate the effects of environmental stress.

Developing nations, however, may not be the only nations at risk of environmentally induced international conflict. With the onset of wide-ranging environmental problems such as global warming, acid rain, and stratospheric ozone depletion, the spectre of international conflict that involves developed nations are portrayed as increasingly likely scenarios.

Lipschutz and Holdren (1990, 126–129) argue that, to date, the connection between environmental stress and international conflict has been weak, because localized environmental stress tends to affect the offending country more than any other nation. The damage incurred from this localized environmental stress has only rarely been sufficient to evoke a significant shift in a nation's conduct in the international arena. With the onset of environmental problems such as global warming, however, these dynamics may shift. Wide-ranging environmental stress threatens to produce unprecedented magnitudes of change: change that could foment large social and economic upheavals that inexorably drag developed nations into the fray. For the time being, of course, the potential for indirect international conflict that involves developed nations is based on reasoned conjecture rather than empirical evidence. Nevertheless, the scenarios conceived by some analysts are not beyond reason.

Perhaps the most apocalyptic scenario sees hemispheric conflict between developing nations in the South and developed nations in the North. The most common underlying environmental cause in this scenario is global warming. Although the impact of global warming may be gradual, the possibility of non-linear effects remains. If global temperatures pass a certain unknown threshold, for instance, changes in the planet's physical systems may be rapid and advance at a greatly increased magnitude. The grave scenarios involving hemispheric conflict sometimes draw on the potential for these non-linear effects of global warming.

Gleick (1989, 338) notes that global warming could affect key goods and services such as freshwater availability and food productivity. The societal impacts of these climatic changes, he writes, will be widely distributed, but they are likely to be felt far more severely by poorer nations, posing important and still unresolved questions about equity, fairness, and international environmental ethics. Similarly, Lunde (1991) contends that diplomatic disputes over responsibility for the generation of greenhouse gas emissions, as well as disputes over the potential ramifications of global climatic change, are likely to play into already-existing tensions between southern and northern nations. And Renner (1996) asserts that the politics of global warming is quickly shaping up as a dispute between the industrialized North, accounting for the bulk of emissions, and the South, which perceives itself to be a victim of global warming. While all three aforementioned analysts reject the notion of an all-out international conflict between North and South, they caution that global warming is an issue that can cause tensions between these hemispheric counterparts to rise steeply, increasing the likelihood of such violence.

Another route to conflict between northern and southern nations is through large-scale migrations and dislocations of people. Localized environmental problems, such as soil degradation and over-fishing, can cause small-scale but persistent migrations in developing nations. Some analysts contend that wide-ranging environmental problems such as global warming, coupled with rapid population growth and increasing disparities in wealth, could induce large-scale migrations from the developing to the developed world. Although small-scale migrations from the developing to the developed world already exist – migrations from North Africa to Europe, for instance, are common – wide-ranging environmental problems could exacerbate this process. Large-scale migrations could increase tensions between northern and southern citizens, and eventually induce hemispheric violence.

Renner (1996, 109–112) notes that citizens in many developed nations have become fearful that they are now being invaded, not by armies and tanks, but by migrants who speak other languages, worship other gods, belong to other cultures, steal their jobs, and threaten their way of life. Many industrialized nations, including the United States, Germany, and France, have witnessed significant domestic movements to curb the growing tide of refugees pounding at their borders on a daily basis. The spectre of massive influxes of people from the developing nations to developed nations elevates the possibility of the rest of the world against the West.

Other scenarios for international war envisage regional wars between developing nations and developed nations, as well as between developing nations. Again, wide-ranging environmental stress is predicted to be the catalyst of such wars, especially given the potential for non-linear effects. Myers (1993, 118–121), for instance, forecasts a potential downside scenario for the Indian Subcontinent, where global warming produces chronic shortages of food and water. These conditions aggravate tensions between India

and Pakistan, and eventually escalate into nuclear threats and hostilities. Similarly, Myers (1993, 146–148) downside scenario for the Americas sees huge numbers of Mexicans illegally migrating into the United States because of resource shortages, disease, and chronic poverty caused by global warming. The massive influx of illegal migrants then compels the United States to invade its southern neighbor in an effort to cut off the routes traveled by the migrants.

DIRECT INTERNATIONAL CONFLICT

If environmental stress may contribute indirectly to international conflict, it is the potential for direct international conflict that has garnered the most attention from scholars, government officials, pundits, and the general public alike. By direct conflict, I refer to violent hostilities that arise when two or more parties vie for control and/or access to environmental resources. In these instances, two or more nations engage in conflict over particular renewable resources in the same way that nations have fought wars in the past over non-renewable resources such as minerals and oil.

There are a number of renewable resources that could potentially elicit international conflict, such as cropland, fisheries, and forests. Indeed, international incidents over fisheries have already taken place. In 1995, for instance, a Canadian warship fired shots across the bow of a Spanish trawler, in a feud between the two nations over fishing rights along Canada's east coast.

It is the increasing competition over access to water resources, however, that constitutes the greatest threat of direct international conflict. Access to freshwater resources—lakes, rivers, and groundwater aquifers—is perceived by many to become the most important *casus belli* of the 21st century. Government officials have long made portentous claims about the advancing threat of water wars. In 1979, for instance, Egyptian President Anwar Sadat declared, "The only matter that could take Egypt to war again is water." In 1990, Jordan's King Hussein made similar declarations. Academics studying the issue have made equally portentous claims about the potential for conflict over water. Cooley (1984, 3), for example, contends that "Long after oil runs out, water is likely to cause wars, cement peace, and make and break empires." And Starr (1991, 19) argues that "Water security will soon rank with military security in the war rooms of defense ministries."

Water attracts such widespread attention as a potential *casus belli*, because this resource is a mainstay of life. Gleick (1998, 3) notes that freshwater is integral to all ecological and societal activities, including the production of food and energy, transportation, waste disposal, industrial development, and human health. Yet the availability of freshwater resources is declining significantly in many regions in the world. As population growth heightens the demand on water resources, water pollution and climatic change are combining to simultaneously decrease the supply. If annual water availability drops below 1700 m^3 per person, domestic self-sufficiency becomes almost impossible. Gleick estimates that there are already over two dozen nations that fall below this threshold, including nations in Africa, the Middle East, Asia, and the Caribbean.

Moreover, freshwater resources do not respect the political boundaries created by humankind. Gleick estimates that more than half the land area of the world, and perhaps 70% of inhabitable land areas, is in an international watershed where river flows or lakes are shared. And he notes that over 220 river basins are shared by two or more nations. Furthermore, these water resources are neither distributed nor shared equally. In the Nile river basin, for instance, Egypt is endowed with a renewable supply of just 1.8 $km^3\ year^{-1}$, but extracts as much as 55.1 $km^3\ year^{-1}$—far more than their renewable supply. In the same river basin, Ethiopia is endowed with a renewable supply of 110 $km^3\ year^{-1}$, but extracts only 2.2 $km^3\ year^{-1}$.

Conflicts over water resources have already been documented. Indeed, Gleick notes that water wars are nothing new to human history. Forty-five hundred years ago, the control of irrigation canals was the source of conflict between the states of Umma and Lagash in the ancient Middle East. Twenty-seven hundred years ago, Assyria and Arabia clashed over the control of water wells. In the modern era, Israel and Syria have engaged in international conflict over water resources on at least two occasions. In 1965–1966, Syria tried to divert headwaters of the Jordan away from Israel, prompting Israel to attack the Syrian diversion facilities. And the 1967 Yom Kippur War in the Middle East was initiated, in part, because of disputes between Israel and Syria over the Jordan River.

With increasing environmental stress, there is a fear that conflict over water will become much more frequent in the near future. There are already many disputes over water resources that threaten to boil over into conflict. In the Middle East, the issue of water scarcity remains largely unresolved. An agreement between Israel and Syria over access to the waters of the Banias River, for instance, is still on the bargaining table. Jordanian concerns about Syrian dams on the Yarmouk River also remain unresolved.

But the spectre of conflict over water is not unique to the Middle East. In Europe, Czechoslovakia and Hungary nearly came to blows in 1992, because of a dispute over the construction of a dam and the diversion of waters on the Danube River. In Africa, Bostwana, Nambia, and Angola have disputed access to the shared waters of the Chobe River since the 1980s. In Asia, Cambodia, Laos, Thailand, Vietnam, and China have all engaged in disputes over the waters of the Mekong River. In South America, Ecuador and Peru engaged in small-scale armed skirmishes in 1995, in part because of disagreement over the control of the

headwaters of the Cenepa River. And in North America, Mexico and the United States continue to vie for control over the waters of the Rio Grande.

Despite growing fears that environmental stress could exacerbate water resource conflicts, some experts on environment and conflict maintain that these scenarios are based more on hyperbole than theoretical and empirical strength. Homer-Dixon (1999, 139), for instance, contends that international wars over river water are likely only when a narrow set of conditions hold: The downstream nation must be highly dependent on the water for its well-being; the upstream nation must be able to restrict the river's flow; there must be a history of antagonism between the two nations; and the downstream nation must be militarily much stronger than the upstream nation. He argues that these restrictions explain why international disputes over river water have only rarely escalated to international armed conflict.

Others argue that social and economic forces mitigate the threat of water wars. The potential for international conflict over water resources is eased, for instance, by global awareness of the interconnectedness of environmental issues, the expansion of economic interdependence, the involvement of international organizations, the evolution of international law in the field of transboundary water resources, and the potential for new technologies that relieve environmental stress.

DIRECT INTERNAL CONFLICT

Just as nations may clash over renewable resources, populations within nations may likewise come into conflict. Although direct internal conflict may occur over water, other renewable resources such as fisheries and arable land may also be sources of tension.

Compared to international conflicts over water, the threat of direct internal conflict has received very little attention in academic and policy circles. This is not entirely surprising, given that nations often contain relatively robust legal/institutionalized mechanisms for resolving disputes that occur within national boundaries, whereas these mechanisms are often relatively impotent in the international system.

Nevertheless, conflict over resources can occur (and has occurred) within national boundaries. The mechanism that ignites these conflicts is much the same as that for direct international conflict. When a resource becomes increasingly scarce within a nation, tensions may rise between groups that are competing for these resources. Environmental stress can contribute to the scarcity of resources. Population increase, for example, may increase the pressure on available arable land, while deforestation and soil degradation may decrease the supply of this arable land.

Groups that compete for these resources may be divided in a familiar way along ethnic, religious, or linguistic lines. Alternatively, these groups may be divided along vocational or professional lines. Loggers, for instance, may come into conflict with naturalists over the preservation of forests. Conflict can also occur between competing groups within a given vocation. Indigenous loggers, for example, could clash with non-indigenous loggers vying for the same forestry resources.

Developing nations are once again the most vulnerable to direct internal conflict. This is because these nations are often characterized by deep fault lines between competing groups, and because they often lack sufficient legal/institutionalized mechanisms for dealing with disputes between competing groups. Gleick (1998, 5–6) notes, for example, that for much of the 1970s and 1980s, India's Karnataka and Tamil Nadu peoples clashed over irrigation rights to the Cauvery River. A 1991 Interim Order handed down by the Cauvery Waters Tribunal – an institution set up by the state of India to resolve the ongoing dispute – only exacerbated tensions between these two groups, resulting in the deaths of over 50 people.

Although developing nations are the most susceptible to these types of internal conflicts, developed nations are not entirely immune. In the United States, for instance, the Los Angeles Valley aqueduct was repeatedly bombed in 1924 by a group of farmers perturbed that water was being diverted from their communities to urban centers in Southern California. The same region almost witnessed conflict in 1935, when the state of Arizona came to the brink of conflict with the state of California over plans for construction of the Parker Dam and subsequent diversion of water from the Colorado River (Gleick, 1998, 5). More recently, in 1999, native and non-native fishers on the East Coast of Canada clashed on several occasions, after the Supreme Court of Canada granted natives the right to fish year round for lobsters.

CONCLUSIONS: RESEARCH DEBATES AND POLICY IMPLICATIONS

The evidence presented above appears to highlight a disturbing trend in the international system. Increasingly, violent conflicts are being linked to the depletion and degradation of renewable resources. These conflicts may take the form of violent clashes between and within nations over particular renewable resources such as freshwater and fish stocks. Or the degradation and depletion of renewable resources may engender and interact with other economic and social factors – such as poverty and weak states – to foster conflict between and within nations. Environmental stress is almost never the only factor leading to conflict. Nevertheless, it can be an important one.

Although copious evidence exists to support the environment-conflict linkages, a number of theoretical and methodological criticisms have been leveled at the research, calling into question the validity of existing evidence on the environment-conflict nexus. I will not offer a comprehensive review of the controversies here. Rather, I will briefly highlight a few criticisms that my colleagues and I have dealt with in detail (Schwartz *et al.*, in press). These criticisms deal with issues that I consider the more pertinent ones under contention.

One of the most important theoretical criticisms of environment-conflict research is that linkages between environmental stress and violence are overstated, because humanity shows astonishing capacity to adapt to environmental problems. Markets stimulate human inventiveness and commerce that open up new sources of scarce resources, encourage conservation, and create technologies that allow substitution of relatively abundant resources for scarce ones. Critics contend that environment-conflict researchers have failed to consider these adaptive processes in their theoretical models.

Although adaptive processes certainly operate in many cases, societies often fail to adequately adjust to environmental stress. This is especially true for many developing nations, where institutions and markets can be weak and inflexible. Moreover, environmental stress itself can have debilitating effects on some economies, societies, and social groups. It is for these reasons that developing nations are more susceptible than developed nations to the social manifestations of environmental stress.

Environment-conflict research is also criticized on theoretical grounds for neglecting to consider the role that social phenomena play in the creation of environmental stress. Critics contend that the causal arrow that traces a route from environmental stress (e.g., soil degradation) to social and economic phenomena (e.g., migration and poverty) through to conflict (e.g., ethnic strife) is misleading, because environmental stress itself is caused in large part by social and economic factors. They argue, therefore, that the causal arrow should begin with social and economic phenomena.

There are at least three reasons why the causal arrow can begin with environmental stress. First, environmental stress is not only influenced by social variables like institutions and policies; it can itself affect these institutions and policies in harmful ways. This is the case when shortages of a renewable resource, such as cropland or forests, motivate elites to seize control of the resource's remaining stocks (the process of resource capture, explained above). In other words, we should not assume that social variables are completely independent and external starting points in the causal chain; it turns out that they can be affected by environmental scarcity, sometimes negatively. Second, the degree of environmental stress a society experiences is partly a function of the particular physical characteristics of the society's surrounding environment. These characteristics are, in some respects, independent of human activities. For instance, the vulnerability of coastal aquifers to salt intrusion from the sea and the depth of upland soils in tropical regions are physical givens of these environmental resources. Third, once environmental stress becomes irreversible – as when most of a country's vital topsoil washes into the sea – then the stress is, almost by definition, an external influence on society. Even if enlightened reform of institutions and policies removes the underlying social causes of the stress, it will remain a continuing burden on society – because the environmental stress itself is irreversible.

Aside from these important theoretical criticisms, environment-conflict researchers have also faced methodological criticisms. At issue in these methodological critiques are key issues pertaining to causation – i.e., the notion that one or more factors can cause an outcome to occur. The concept of causation has been debated in philosophical circles for hundreds of years. Although some consensus has emerged on how social scientists can show causation, the issue has hardly been resolved with any finality.

In order to show causation, mainstream social scientists favor quasi-experimental methods, in which potentially confounding variables are controlled and cases are selected in a random manner. The single case-study method neither controls for confounding variables nor selects cases in a random manner. Consequently it is widely perceived to be flawed by mainstream social scientists. Critics contend that because environment-conflict researchers have used the single case-study method rather than the quasi-experimental method to gather evidence, they are not able to make legitimate claims about causation.

My colleagues and I argue that the single case-study method in fact plays an important role in complementing the quasi-experimental methods preferred by mainstream social scientists. In order to demonstrate this, we draw on a distinction between causal effect (i.e., the impact that one variable has on another variable) and causal mechanism (i.e., the linkage between two or more variables). We argued that both causal effect and causal mechanism are essential to showing causation. Although quasi-experimental methods can estimate causal effect, they say little about causal mechanism. The single case-study method, on the other hand, can trace casual mechanisms but can not estimate causal effect.

We argue that research to date on environment and conflict has demonstrated the causal linkages between environmental stress and conflict (i.e., causal mechanism), but has not yet determined the degree to which environmental stress plays a role relative to other factors in the generation of conflict (i.e., causal effect). We conclude that a next crucial step in environment-conflict research is to estimate the causal effect of environmental stress.

While theoretical and methodological debates surrounding environment-conflict research continue in academic circles, at least some policymakers in foreign policy circles have begun to incorporate the research into their agenda. The US Central Intelligence Agency (CIA), for instance, has conducted a State Failure Task Force Report on the relationship between environmental stress and the outbreak of violent conflict such as revolutions, ethnic strife, and genocides (Esty *et al.*, 1999). And the Canadian counterpart to the CIA, the Canadian Security Intelligence Services (CSIS) has also recognized the linkages between environmental stress and violence (Gizewski, 1997).

For some, interest by the CIA and CSIS in environment-conflict research has only confirmed their worst fears: that by including environmental issues in debates over security, actions pertaining to the environment may become increasingly militarized. Although these fears may have some basis, environment-conflict research does suggest that if policymakers want to mitigate conflicts and enhance security in the future, they must pay attention to the state of the environment.

REFERENCES

Barber, V (1997) *Environmental Scarcities, State Capacity, and Civil Violence: the Case of Indonesia*, University of Toronto Press, Toronto.

Brown, L R (1977) *Rede ning National Security*, Worldwatch Institute, Washington, DC.

Cooley, J K (1984) The War Over Water, *Foreign Policy*, 54, 3–26.

Durham, W H (1979) *Scarcity and Survival in Central America: Ecological Origins of the Soccer War*, Stanford University Press, Stanford, CA.

Esty, D C, Goldstone, J A, Gurr, T R, Harff, B, Levy, M, Dabelko, G D, Surko, P T, and Unger, A N (1999) State Failure Task Force Report: Phase II Findings, in *Environmental Change Security Project Report*, Woodrow Wilson International Center for Scholars, Washington, DC, 5, 48–68.

Gizewski, P (1997) Environmental Scarcity and Conflict, *Commentary: a Canadian Security Intelligence Service Publication*, 71, Spring 1997.

Gleditsch, N P (1997) Conflict and the Environment, *Proceedings of the NATO advanced research workshop*, Kluwer Academic Publishers, Dordrecht.

Gleick, P H (1989) Climate Change and International Politics: Problems Facing Developing Countries, *Ambio*, 18, 333–339.

Gleick, P H (1998) *The World's Water: 1998- 1999*, (Advanced Copy), Island Press, Washington, DC.

Hazarika, S (1993) Bangladesh and Assam: Land Pressures, Migration and Ethnic Conflict, *Occas. Pap. Ser. Proj. Environ. Change Acute Con ict*, 3, 45–63.

Homer-Dixon, T F (1999) *Environment, Scarcity, and Violence*, Princeton University Press, Princeton, NJ.

Homer-Dixon, T F and Percival, V (1996) *Environmental Scarcity and Violent Con ict: Brie ng Book*, University of Toronto, Toronto.

Lang, C (1995) Environmental Degradation as a Cause of Political Conflict, Social Stress, and Ethnic Tensions, Occasional Paper No. 12, *Environ. Con icts Proj.*, Swiss Peace Foundation, Berne.

Lipschutz, R D and Holdren, J P (1990) Crossing Borders: Resource Flows, the Global Environment, and International Security, *Bull. Peace Proposals*, 21, 121–33.

Lunde, L (1991) North/South and Global Warming – Conflict or Cooperation? *Bull. Peace Proposals*, 22, 199–210.

Molvoer, R K (1991) Environmentally Induced Conflicts? *Bull. Peace Proposals*, 22, 175–88.

Myers, N (1993) *Ultimate Security*, W W Norton, New York.

Pirages, D (1996) Microsecuirty: Diseases, Organisms, and Human Well-Being, *Environ. Change Secur. Project Rep.*, 2, 9–17.

Renner, M (1996) *Fighting for Survival: Environmental Decline, Social Con ict, and the New Age of Insecurity*, W W Norton, New York.

Schwartz, D M and Ashbindu, S (1999) *Environmental Conditions, Resources, and Con icts*, United Nations Environment Program, Nairobi.

Schwartz, D M, Deligiannis, T, and Homer-Dixon, T F (in press) Environmental Scarcity and Violent Conflict: A Response to Gleditsch, in *Environmental Con ict*, eds P F Diehl and N P Gleditsch, Westview Press, New York.

Smil, V (1992) Environmental Change as a Source of Conflict and Economic Losses in China, *Occas. Pap. Ser. Proj. Environ. Change Acute Con ict*, 2, 5–39.

Starr, J R (1991) Water Wars, *Foreign Policy*, 82, 17–36.

Suliman, M (1992) Civil War in Sudan, Occasional Paper No. 4, *Environ. Con icts Proj.*, Swiss Peace Foundation, Berne.

Wegenmund, R (1996) Ethnic Transborder Conflicts in the Senegambian Region Caused by Environmental Degradation, in *Environmental Degradation as a Cause of War*, eds G Bachler and K Spillman, Verlag-Ruegger, Zurich.

Development and Global Environmental Change

WOLFGANG SACHS

Wuppertal Institute for Climate, Energy, Environment, Wuppertal, Germany

The notion of development, offering language to the desire of many nations for justice and dignity, has been a powerful idea forcing global environmental change. The article traces the rise and decline of the development idea over the past 50 years. It attempts to uncover the concept's many layers of meaning as they built up during post-war history; it exposes the bitter legacies - socially, culturally, environmentally - which have been left behind by this period; and it points to the recent transition from development to globalization. Finally, some prospects are sketched out for achieving more equity outside the framework of conventional development.

HISTORY

The rise of epochs often go unnoticed, but the dawning of the development age occurred at a precise date and time. On 20 January 1949, President Harry S. Truman, in his Inaugural Address to Congress, dubbed the home of more than half the world's people underdeveloped areas. This was the first time that the word *underdevelopment* (OED, 1989, XVIII, 960), which was later to become a key category ordering global relations, was used by a prominent political figure. Truman, after drawing a sharp line between democracy and communism in the first part of his speech, directed the attention of his audience to the Southern Hemisphere with the following words:

> Fourth, we must embark on a bold new program for making the benefits of our scientific advances and industrial progress available for the improvement and growth of underdeveloped areas. More than half the people of the world are living in conditions approaching misery. Their food is inadequate. They are victims of disease. Their economic life is primitive and stagnant. Their poverty is a handicap and a threat both to them and to more prosperous areas. For the first time in history, humanity possesses the knowledge and the skill to relieve the suffering of these people.
>
> (Truman, 1949)

With these high-flying words of the Point Four Program, the development age, the particular period in world history following the colonial era, was opened (to be superseded some forty years later by the globalization age). In this period, the relationship between Europe/America and the rest of the world was being shaped by specific assumptions about time, geographical space, and relevant social actors. These assumptions came to frame the development discourse; they stand in continuity and contrast with the assumptions of both the earlier colonial and the later globalization periods.

Linear Global Time and the Rule of Gross National Product (GNP)

In the light of the concept development , all peoples on the globe appear to move along one single road. The lead runners show the way; they are at the forefront of social evolution, indicating a common destination even for countries which had highly diverse trajectories in the past. Many different histories merge into one master history, many different time scales merge into one master time scale. The imagined time is linear, only allowing for progressing or regressing; and it is global, drawing all communities worldwide into its purview. In contrast to cultures which may embrace a cyclical view of time or which live out stories enshrined in myths, the linear view of time privileges the future over the present, and the present over the past. As the concept of linear global time spreads, indigenous peoples like the Rajasthani in India or the Aymara in Peru, for instance, are compelled to put aside their particular chronographs. They are inevitably pulled into the perspective of progress.

Of course, the belief in progress predates the development age by almost two hundred years. The European Enlightenment was already able to interpret the multiplicity of cultures in space as a succession of stages in time, viewing history as a never-ending process of improvement. Taken from biology, the metaphor development constructed history as a process of maturation; society is likened to, say, a flower which develops according to inner laws, in a continuous and irreversible fashion, towards a final stage of bloom. However, since about 1800,

development has only been used as an intransitive concept (Wieland, 1979); authors such as Hegel, Marx, and Schumpeter conceived it as a process of history's own making, but not yet as a project to be carried out under the direction of human will and reason (Lummis, 1996). This changed with the advent of the development age. Development took on an active meaning; it turned into a project of planners and engineers who set out to systematically remodel societies to accelerate maturation; a project to be completed within several decades, if not years.

Such optimism was facilitated by the fact that economic performance had become the all-encompassing measure of a country's excellence only after 1945. Sir Frederick Lugard, the inspirer of British colonial theory in the 1920s, had still conceived the task of a colonial power as a dual mandate; as both the economic development of colonial territories for the benefit of industrial countries and as moral concern for the native populations (Lugard, 1922). Economic progress and the welfare of the natives had been considered as two distinct duties; the civilizing mission had comprised both developing resources such as land, minerals and timber, and elevating the natives to a higher level of civilization. It was only at the time of Truman that the double mandate collapsed into one: development. The former distinction between an economic and a moral realm vanished, a sign of a conceptual shift. From then on, not only resources figured into the development formula, but people as well. Inversely, the moral concern for people was eclipsed by the economic concern for growth. This shift indicated that a new worldview had come to the fore: the degree of civilization of a country can be measured by its economic performance.

As it happened, this measure of excellence has been available only since 1939, when Colin Clark for the first time compiled national income figures for a series of countries, revealing the gulf in living standards between rich and poor (Arndt, 1987, 35). GNP *per capita* provided a ready-made indicator for assessing the position of countries moving along the road of development. Informed by an economic worldview and aided by a statistical toolkit, experts for decades to come de ned development as growth in output and income per head. A developing country is one with real *per capita* income that is low relative to that in advanced countries like the United States, Japan, and those in Western Europe (Samuelson and Nordhaus, 1985, 812).

Hierarchic Global Space and the Imperative of Catching-up

As already indicated, the chrono-politics of development was, and is, accompanied by a particular geo-politics (Di Meglio, 1997). Through the prism of development, the confusing diversity of nations across the globe appears as a clear ranking order shaping the orientation of both the powerful and the less powerful. After all, the very metaphor of development implies predominance. Just as immature fruit can be only recognized by comparing it to a mature fruit, the stages of underdevelopment can only be recognized by displaying particular societies as examples of maturity. Development without predominance is therefore like a race without direction; assigning the positions of leaders and followers is part and parcel of a developmentalist construction of history.

While inequality between nations in colonial times was understood within an authoritarian framework, which can be likened to a father-child relationship; in the development age it was seen in an economic framework, which can be likened to a race between differently endowed runners. Indeed, staying with that image, it can be said that development policy basically had two objectives: rst, bringing countries onto the racetrack, i.e., into the orbit of the world market; and secondly, turning them into competent runners, i.e., putting them on a path of sustained growth. However, it was only after the diversity of living conditions had been reduced to a hierarchy of aggregate national income gures that the enormous distance separating rich and poor countries leaped into view. No matter what ways of life the Kikuyus in East Africa or the Gujaratis in India cherished, no matter what ideals they aspired to, in the development mindset their diversity is crammed into one single category; they are underdeveloped. Thus, entire peoples found themselves locked into being de cient, de ned not according to what they are and want to be, but according to what they lack and are expected to become.

As is so often the case, the de nition of the problem already implied the solution. When low income is considered the salient problem, raising incomes is the key solution. Any desire for change in Southern countries was thus reinterpreted as demand for economic development, casting aside other possible interpretations: for example, that oppression or dependency might be the problem which would call for liberation or autonomy as a solution. Nor was there a notion that cultures might in the rst place aim for non-economic ideals, be it the integrity of the clan or the celebration of religious rituals. On the contrary, the race in the economic world arena, like any race, was seen to be dominated by the imperative of catching up. Indeed, the chrono-politics and geo-politics particular to the development idea engendered a monumental historical promise: the promise that, at the end of the day, all societies would be capable of bridging the gap to the rich and sharing in the fruits of industrial civilization. It may be in this promise that the foundation of the development creed in Christian thought becomes most palpable (Rist, 1997). Development can be understood as a secular salvation story, constituting an ecumenical community, which places its trust in the good works of providence and faithfully follows the path of predestination (Tenbruck, 1989).

Rising Nation States and the Social Contract

After World War II it was incumbent upon the United States to project a New World Order. Germany and Japan had been defeated; France and Britain seriously wounded. As the colonial powers lost their grip on the world, independence movements sprang up in the South and claimed the right to nation states. A power vacuum opened which threatened to be filled either by violence or communist takeovers. In this situation, the US launched (following its own self-image) self determination, free trade, democracy, and international cooperation as the core values of a future order. A world was presented which is held together by economic interdependence and not any longer by political dominion. Economic strength had taken the place of military power. As colonialism had occupied overseas territories, the doors to political freedom would be thrust open: provided, at least, that trade could flow freely. The US, having itself emerged as a nation from an anti-colonial struggle, readily backed the process of de-colonization, proclaiming at the same time economic development as an overarching goal. Development was thus a conceptual vehicle for American dominance with a liberal face; it allowed for heralding national independence while expanding predominance. A new model of power appeared on the horizon: anti-colonial imperialism.

Decolonization implied the building of numerous nation states in Asia and Africa. As in 19th century Latin America, the nationalist leaders aspired to that model of political organization, which had been forged into a European norm through centuries-long struggles against the Pope, local lords, and foreign intruders. The nation, after all, is an imagined community, residing within the confines of a particular territory and governed by a sovereign state (Anderson, 1983). Establishing the nation as the relevant community (as opposed to family clans or religious affiliations), carving out governable territorial units, and imposing the authority of a bureaucratic state, was an arduous task which called for a mobilizing narrative. The development idea provided legitimation; it became the *raison d'être* for the emerging states. The new governments largely internalized the image thrust upon them by the countries perceived as advanced; they saw in the fight against underdevelopment the mission of their nations. In a certain sense, the right to self-determination had thus been acquired in exchange for the right to self-definition (Rist, 1997, 79).

Moreover, the perspective in which one would eventually catch up with the rich countries restored self-respect and pride to countries, which had been humiliated by colonialism. Such a perspective promised the new countries an equal standing among nations; the demand for development expressed the desire for recognition and justice. What fuelled the determination to catch up was a double asymmetry in power between the South and the North. Culturally, the West had become the intimate enemy (Nandy, 1983) of the indigenous elite, giving shape to their imagination of success; and politically, the might of the North had become so formidable that pure survival instinct forced the South to seek similar economic and technological means. As this asymmetry widened during the decades which followed, the demand for development grew stronger and more desperate up to the point of codifying a right to development in the United Nations General Assembly of 1986. Then and before, the emerging countries reinterpreted the power gap as a development gap. Apart from some exceptions, they saw no other choice than to join the race.

However, the rich nations also had a stake in making development a global project, in whose name a social contract of cooperation could be forged between North and South. After the horrors of war, it was commonly held belief in the UN that peace could only be preserved by launching economic development on a worldwide scale. Virtually all countries rejected at that time the supremacy of the market and believed in the active management and planning of the economy by the state (Hobsbawm, 1994). In particular the US, attributing the outbreak of war in Europe to economic disorder, remembered their own successful management of the crisis during the New Deal, when J M Keynes had recommended state action to counter unemployment and underproduction. Freedom from fear and want had been already held out by President Roosevelt in the Atlantic Charter of 1941; it is from this point that development became a cornerstone of the mission of the UN. For the sake of stability and peace, the need for economic growth (steered along through public intervention) was projected upon the world. In this sense, development can be seen as an exercise in global Keynesianism to keep disorder at bay. Both the hegemonial needs of the North and the emancipatory needs of the South converged upon the prospect of development.

The Poor as Target and the Dethronement of GNP

The situation was rather straightforward in the two post-war decades. Despite differences in approach, development had been identified with economic growth, if not plainly with industrialization. No matter if shortage of physical capital or shortage of human capital was seen as the major deficit, both answers (capital formation or schooling) had aimed at increasing the flow of goods and services. However, the concept of development began to be contested in the 1970s when attention shifted to the poor, exposing the failure of growth to benefit the large majority of people. Robert McNamara, incoming President of the World Bank, put a finger into the wound:

> Growth is not equitably reaching the poor ••• Rapid growth has been accompanied by greater maldistribution of income in many developing countries ••• We should strive to eradicate absolute poverty by the end of the century. That means in practice the elimination of malnutrition and illiteracy, the reduction

> of infant mortality and the rising of life expectancy standards to those of the developed nations.
>
> (McNamara, 1973)

Merely watching the rate of income growth was not enough anymore, the social content of development now mattered. For whose benefit? emerged as the benchmark question. As a consequence, the scope of attention expanded; not only the middle classes boosting market output, but also the populations left out by growth, and even the victims of growth, all became targets for specifically designed interventions. Development was thus redefined as something transcending growth, as economic growth plus redistribution, plus participation, or plus human development. In this vein, throughout the years that followed in quick succession, areas like employment, equality, poverty eradication, basic needs, informal sector, and women were established as fields of development action, each of them bringing a new set of tools and a new tribe of experts to the fore.

With these extensions, a conceptual inflation set in. Soon, development meant everything and nothing, the concept ceased to denote anything in particular, it just connoted good intentions. It had no content, but retained a function: it justified any action in the name of some higher evolutionary goal. However, what the concept lost in semantic precision, it gained in political versatility. Opposing camps both claimed to promote development; the struggle over meaning reflected from now on the struggle over policy. In particular, the controversies turned again and again on the role of economic growth, with a focus on GNP growth lined up against a focus on social or (later) environmental quality. While the first focus cherished the positivism of growth, the latter centered on non-economic wealth. While the first pushed output, the latter cured the consequences. It thus follows that development can be both the injury as well as the therapy.

At any rate, this dichotomy continued to shape the development debate for subsequent decades (Nederveen Pieterse, 1998). A straight line runs from the International Labour Organization (ILO's) world employment strategy in 1970 to the basic needs approach and finally the Human Development Reports of United Nations Development Programme (UNDP) in the 1990s. Similarly, a line can be drawn from the export promotion of the Asian Development Bank in the 1970s to structural adjustment in the 1980s and finally the bailing-out policies of the International Monetary Fund (IMF) in the 1990s. Some growth skeptics have subsequently redefined development as enlarging people's choices and capabilities, a formula which provided the foundation for the Human Development concept with its emphasis on literacy, health and participation. Its essence is to place development at the service of people's well being rather than people at the service of development (Banuri *et al.*, 1994, 16). Even this view, however, (along with more recent creations like the social capital approach) cannot escape the shadow of the development creed; the Human Development Index is, much like the GNP, a deficit index; it ranks countries hierarchically, assuming that there is only one best way of social evolution.

A New International Order and the Rise of Third Worldism

The states, which had emerged out of the crumbling colonial empires, did not stay long without a voice in international affairs. After the Nonaligned Movement had formed at Bandung in 1955 so as to put some clout behind the de-colonization process, the G 77, set up in 1963 at the eve of the first United Nations Conference on Trade and Development (UNCTAD) Conference, aimed at articulating collective bargaining power in the world economic arena. Development ceased to be seen as an objective to be achieved in individual countries, but was thought to require a less hostile international economic structure. Obstacles to development are not just to be found in domestic habits and institutions, as suggested by the modernization discourse; they now present themselves as well in the detrimental terms of international trade. This framing of the development problem was intellectually prepared and further elaborated by the so-called dependency theory , which identified structures of unequal exchange and in general the gradient of economic power as the source of continuing underdevelopment. Responding to their structurally peripheral position, Southern countries, perceiving themselves as sharing the same destiny, formed a coalition demanding a new international economic order. The South opposed itself to the North: a constellation which reached its high-point in the 1970s as the oil exporting countries successfully displayed their collective market power with regard to the affluent economies.

Although the demand for redefined rules in the world economy weakened as the oil cartel collapsed and the debt crisis exploded in the 1980s, the G 77 continued to be a major player in UN politics, in particular after the rise of East-Asian countries as serious competitors. In 1991 the South Commission restated the call for a fair world order (South Commission, 1991). More specifically, the moment the North began to call upon the South for cooperation in environmental matters, the Southern coalition, all internal differences notwithstanding, renewed their grievances about the asymmetry of power. Since then, in environmental negotiations, the claims to greater economic space are bargained against demands for a wiser use of the biosphere. Until today, however, this quest for equal power and recognition remains focussed on relations between states, leaving the question of inequality within countries in the dark. Just as the Declaration on the New International Economic Order of 1974 had failed to mention domestic inequality even once, the calls for greater justice still sounded

hollow in the nineties because the ongoing marginalization of social majorities remained hidden behind the veil of national sovereignty.

Disrupted Nature and the Conservation of Development in Time

In the 1980s, the promise of development received a second blow. While a decade earlier the persistence of poverty had begun to undermine the concept's social viability, now the emerging natural limits to growth cast doubts on its long-term viability. Combustion based on fossil energy, the core of industrial metabolism, threatened to overburden the atmosphere and the growing world economy's voracity for living resources threatened to destabilize forests, water quality and soils across the globe. Against the backdrop of emerging biophysical limits to economic growth, the development idea underwent yet another round of conceptual inflation. Following the logic of turning victims into clients, development had to be reworked to allow for both infinite growth and preservation of nature. Again, a qualifier was attached, defining *sustainable development* as development that meets the needs of the present without compromising the ability of future generations to meet their own needs (World Commission on Environment and Development (WCED), 1987, 8). The formula pointed to the future, but to a bleak future of scarcities rather than a bright future of progress. It called for development choices that would not drastically restrict the environmental space available for future generations.

Although highlighting justice in time, the canonical formula underemphasized justice in social space. Constraints imposed by the present generation on future generations were given prominence over constraints imposed by powerful groups on less powerful groups within one generation. Needs and generations, after all, are socially neutral terms; they do not allow for vertical distinctions. Yet such distinctions are crucial when it comes to intra-generational equity. Whose needs and what needs are supposed to be met? Is sustainable development supposed to meet the needs for water, land and economic security or the needs for air travel and bank deposits? Is it concerned with survival needs or with luxury needs? Leaving these questions up in the air, the acceptance of sustainable development in circles of privilege and power was facilitated, obfuscating the point that there may be no sustainability without restraint on wealth.

Furthermore, by linking *sustainable* to development, a terrain of semantic ambivalence was created. The new concept subtlely shifted the locus of sustainability from nature to development; while sustainable previously had referred to renewable resources, it now referred to development. With that shift, the perceptual paradigm changed. The meaning of sustainability slid from conservation of nature to conservation of development. Given that development had conceptually become an empty shell, which may cover anything from the rate of capital accumulation to the number of latrines, exactly what should be kept sustainable was unclear and contentious. Hence all sorts of political players, even fervent protagonists of economic growth, have in subsequent years been able to couch their intentions in terms of sustainable development. The term had soon become self-referential, as a definition offered by the World Bank neatly confirms: What is sustainable? Sustainable development is development that lasts (World Bank, 1992, 34).

LEGACY

For fifty years, development has been much more than just a socio-economic endeavor; it has been a perception which models reality, a myth which comforts societies, and a fantasy which unleashes passions. However, perceptions, myths, and fantasies rise and fall independent of empirical results and rational conclusions. They rise when they are pregnant with promises, and they fade when they turn into sterile commonplaces. Achievements such as the average rise of *per capita* GNP in Southern countries from 1960 to 1997 at a rate of 2.1% per year, the success story of a number of East-Asian economies, or the declining rates of infant mortality even in the group of low-income countries, are not indicative enough to fortify the faith in development. In the same way, the enumeration of failures, such as the rising absolute number of poor people (living on less than $1 per day) worldwide, and their slowly increasing relative number in Sub-Saharan Africa as well as in Latin America, is not necessarily forceful enough to undo that faith. Instead, a world-view, like development, loses appeal when the implicit promises cease to command credibility. This is what has happened in the last 10–15 years; the three founding promises: that economic development will spread across global space, improve human destiny, and continue for ever in time, have turned stale.

Social Polarization

After fifty years of development, the promise of greater justice in the world has largely evaporated. In the international arena, the notorious gap in income between North and South has not been bridged; on the contrary it has widened to an extent that it is now unimaginable that it could ever be closed. In 1996, the 20% of the world population living in affluent countries had an income at their disposal 82 times higher than the poorest 20% of the world population; in 1960, it was just 30 times higher (UNDP, 1998, 29). Upon closer inspection, to be sure, the picture is far from homogeneous because these relative figures hide, for instance, that *per capita* income in oil-exporting and in East Asian countries has sharply risen during the last 20 years.

But absolute impoverishment proceeds at the same time; *per capita* income in more than 80 countries is today lower than it was 10 or more years ago (UNDP, 1999, 3). Social polarization among countries advances, while the size and the composition of the poles may change. The world might have developed, indeed: but in two opposite directions. Looking at the over-all picture, it is probably no exaggeration to say that the aspiration of catching up with the rich has ended in a blunder of planetary proportions.

This is even more true if one considers the destiny of large majorities of people within most countries; the polarization between nations (often, but notably not in India and a number of East Asian countries) repeats itself within each country. In Brazil, for example, the share of the poorest 50% of the population in national income amounted to 18% in 1960, while it fell to 11.6% by 1995 (UNDP, 1998, 29). Economic growth, it turns out, has often failed to reach its most heroic objective; to alleviate the burden of the poor. Investments in ports, roads, steel mills and fertilizer plants might have fuelled national income, but they have rarely trickled down to the poor. In fact, it has taken decades of misplaced development assistance to discover that there is only a loose relationship between levels of economic growth and levels of poverty. Growth, it turns out, is certainly not sufficient for mitigating poverty; land rights, community coherence, and self-organization have been shown to be at least as important. Yet, precisely these conditions of livelihood have often been undermined in the pursuit of growth. Dams displace people, machines substitute rural workers, cash crops replace subsistence crops, urban migration follows the loss of self-affirmation. More often than not, the natural and social base of livelihood economies have thus been exploited for building the base of a market economy. Against this backdrop, it is not astonishing that the spread of misery has often accompanied economic growth.

The development perception, having for a long time seen poverty as simply a lack in income, failed to appreciate life-enhancing, non-market resources for the poor, such as rights, social capital, and natural resources. As a consequence, the application of growth recipes had a polarizing effect; it turned frugality into misery, making a minority better off in the process. Moreover, in staying true to its bias that only income growth matters, the development perception remained blind to the effect of unequal power relations. However, as access to money and non-money resources is determined by power, growth without redistribution of power enabled the urban middle classes, manufacturers, and big farmers to corner the gains of prosperity, shifting the cost to small farmers, indigenous people, and the urban proletariat. And still today, a totally disproportionate share of economic gains in Southern countries goes to the rich and politically well connected (Ayres, 1998, 126). Through development, furthermore, countless people have been drawn into the money economy, a transition which modernized both poverty and wealth (Illich, 1971). Where the gradient of power and prestige is calibrated according to purchasing power, expectations tend to explode while opportunities remain limited. Given that satisfaction is relative to expectations, fuller integration into the money economy may therefore even increase the feeling of poverty. Television sets in shacks have become symbols for the unbridgeable gulf between means and expectations.

Unsettling of Cultures

The most dramatic and far-reaching social change of the second half of the 20th century, the one which separates the modern world forever from the past, is the death of the peasantry (Hobsbawm, 1994, 289). This marked the end of several thousand years of cultural evolution when the overwhelming majority of the human race lived by agriculture, raising livestock or harvesting the sea as fishers. After the peasants of rural Europe and Japan had more or less stopped tilling the land by the 1960s, Latin America, large tracts of Asia, and North Africa followed suit in the last decades of that century. Only three regions of the globe remained essentially dominated by their villages and fields: Sub-Saharan Africa, South and continental South East Asia, and China. However, while in the Northern countries the population of the shrinking agrarian world was largely absorbed by the expanding industrial world, in the Southern countries only a minority in this transition found a dignified living in towns and cities. Nevertheless, urbanization has continued to change the human condition on an ever more massive scale. Just during the last 25 years, the share of the world population which is living in urban areas went from slightly more than one third to one half, a figure which is expected to rise to two thirds by 2025 (World Bank, 2000, 46). Development policy, indeed, had set out to propel agrarian societies into the urban-industrial age. It sought to replace traditional Man by modern Man, an endeavor which, however, ended in fatal success; while traditional Man has vanished, modern Man has by no means arrived. Living in a no-man's land, exiled from tradition and excluded from modernity, has become the destiny of most of the world's people.

Nothing less than turning entire societies upside down has been the intent of development planners right from the start. For example, the 14-person mission to Colombia, the first of its kind sent out by the Bank for Reconstruction and Development (later the World Bank) in 1949 arrived at the following conclusion:

> Piecemeal and sporadic efforts are apt to make little impression on the general picture. Only through a generalized attack throughout the whole economy on education, health, housing, food, and productivity can the vicious circle of poverty ••• be

> decisively broken. But once the break is made, the process of economic development can become self-generating.
>
> (Escobar, 1995, 24)

In this engineering spirit, experts set out to remake societies from a host of locally based subsistence communities into nationally integrated economies. They aimed at reorganizing social ties in functional terms, as called for by the ambition to set up machinery geared towards producing growing amounts of material wealth. Under the expert s gaze, time-honored ways of living and knowing faded into oblivion, reduced to obstacles to development . Instead, decontextualized production models, depicting people and nature as abstract objects to be changed, were projected upon in nitely diverse human realities. People rarely g-ured as agents in the framework of rationalist planning (Hobart, 1993); their interests, passions, and knowledge hardly mattered against the backdrop of grand schemes for resource mobilization. Small wonder that development strategies based on such models failed time and again; they were too much in dissonance with the dynamics of a given community. It is for this reason that development has always produced both order and chaos.

At any rate, development has often failed to grasp the rich complexity of non-economized societies. It could not appreciate, for instance, that such settings can be regarded as symbolic sites (Zaoual in Rahnema, 1997; Apffel and Marglin, 1998) where communities live out narratives that link them to their divinities or where social energy is primarily invested in the upkeep of a network of friends, relatives or clan members. In such circumstances, any modernization will run quickly into communitarian constraints, as relations to divinities or to fellow citizens are likely to collide with the requirements of functional performance. To put it in more general terms, development has aimed at achieving that decisive shift which distinguishes modern civilization from all others: primacy is not given any longer to the relations between people, but to the relations between people and things (Dumont, 1977). In the rst case, events are evaluated in the light of their signi cance with regard to neighbors, relatives, ancestors, and gods: whereas in the second, they are judged according to what they contribute to the acquisition and ownership of things. This impersonality postulate (Banuri in Apffel and Marglin, 1990) according to which impersonal relations are inherently superior to personal relations can well be regarded as speci cally Western; to make it prevail is what development as modernization was all about.

It is probably true that this shift has been a mixed blessing for large majorities of the world s people. On the one hand, it moved many regions and classes into the modern world with its liberties and conveniences; on the other, it disembedded countless persons from their cultures, sending them to join the global majority of underconsumers. As long as cultures (both large and small ones) are con ned to themselves, people everywhere have tended to view the corner of the world they inhabit as particularly favored and their own ways of life as good: that is, quintessentially human (Tuan, 1986, 1). However, as all cultures are drawn into the maelstrom of global interaction and assimilation, reinforced by the transborder ow of images, this self-con dence can hardly be maintained. As borders limited the space of comparison, they facilitated limited, but attainable satisfaction, while a borderless world, making the space of comparison explode, offers unlimited, but often unattainable satisfaction. This goes a long way in explaining why both excitement and dissatisfaction grow along with globalization.

Nature's Predicament

After the Second World War, the US, along with other industrialized nations, could still feel it was at the forefront of social evolution. After 50 years this premise of superiority has been fully shaken (if not shattered) by the ecological predicament. Using the racetrack metaphor, development may have been a race, which was conducted unfairly, and which has driven the majority of runners to exhaustion, but its demise is imminent before the tribunal of history because the horses appear to be running in the wrong direction. From the local to the global level, many experiences have shown that the sources (water, timber, oil, minerals etc.), sites (land for mines, settlements, infrastructure), and sinks (soils, oceans, atmosphere) for the natural inputs of economic growth have become scarce or have been thrown into turbulence. As a consequence, the promise that development will continue for ever has collapsed. For instance, if all countries followed the industrial example of emitting *per capita* on average 11.4 tones of CO_2 annually, the emissions of six billion people would amount to roughly 68.4 billion tones; more than ve times the 13 billion tones the earth is capable of absorbing. In other words, to bring all countries up to the present standard of living in af uent countries, ve planets would be needed to serve as source for the inputs and sinks of economic progress. Against this backdrop, development has moved into an impasse. With biophysical limits to economic expansion (hard to pin down and ever contestable, but nevertheless real) emerging in the last quarter of the 20th century, the North cannot be held up as a model any longer; the trail-blazers are without a compass. The recognition of the niteness of the earth has been a fatal blow to the idea of development as envisaged by Truman.

In fact, ecological constraints represent only half the story; they are compounded by the fact that approximately 20% of the world population consume 80% of the world s resources. The majority of the global consumer class roughly those that have a bank account, some career prospects, and access to a car can be found in the North.

It is those 20% who eat 45% of all meat and fish, consume 68% of all electricity, 84% of all paper, and own 87% of all cars (UNDP, 1998, 2). They can be called the omnivores (Gadgil and Guha, 1995), namely those who are capable of cornering environmental resources to their benefit at the cost of other groups. In the global context, the industrialized countries tap into the patrimony of nature to an excessive extent; they draw on the environment far beyond their national boundaries. The ecological footprint (Wackernagel and Rees, 1996) (*see* **Ecological Footprint**, Volume 3) that they produce is larger (and in some cases much larger) than their own territories; a great deal of the resources and sinks they utilize are not available for other countries anymore. In fact, the Organisation for Economic Cooperation and Development (OECD) countries surpass the (in terms of ecology and equity) admissible average size of such a footprint by a magnitude of about 75–85%; as matters stand today, the wealthy 25% of humanity occupy a footprint as large as the entire biologically productive surface area of the earth. In a closed environmental space, the question of how much is tolerable therefore becomes intermingled with the question, who is getting how much?

But also within countries, in particular Southern nations, the consumer classes often succeed in sequestering themselves against environmental burdens, leaving the noise, the dirt, and the ugliness of the industrial hinterland at the doorsteps of less advantaged groups. In 1994 13% of Southern urban citizens lacked access to clean drinking water, and almost twice as many did not even have the simplest latrines (World Bank, 2000, 140). Contrary to the *fata morgana* of development, health conditions for the poor in cities are today worse than in rural areas. Moreover, just as in the case of the citizens of the industrial world, the Southern middle classes live off the resource base which supports the ecosystem people (Gadgil and Guha, 1995), that third of the world population (UNDP, 1998, 80) which derives their livelihood directly from free access to land, water, and forests. Building large dams and extracting ore, drilling groundwater wells and capitalizing agriculture for the benefit of the urban classes often degrades the ecosystems on which they depend. As the appropriation of resources can often only proceed after the rights of inhabitants have been denied, human rights violations frequently go hand in hand with resource conflicts (Johnstone, 1994). Such kinds of pressures, adding to others like unequal landholdings or a growing population, may turn ecosystem people into landless and rootless squatters who have no other choice than exhausting fragile lands and woods. It is only a slight exaggeration to say that development in these cases deprives the poor of their resources in order that the rich may live beyond their means.

In sum, environmental degradation arises from two contradictory settings: one of success and domination, the other of marginality and powerlessness. In the first instance, corporations and consumers of the affluent world dispose of the economic power to mobilize, if necessary over long distances, huge amounts of resources, producing pollution, devastation, and turbulence in the process. In the second instance, poor people without purchasing power degrade their habitats, after having lost their traditional rights or any other kind of entitlement to secure sufficient sources of livelihood. Both the degradation by the affluent and the degradation by the poor can be largely considered the outcome of one and the same process of economic development. Resource voracity on the part of the powerful, and resource scarcity on the part of the powerless, combine in pushing the planet to the brink.

TRANSITION

It is not just its own fading promises, but also shifts in the world economy that have contributed to the decline of the development age. Since the mid-1980s the accelerated rise of globalized markets along with the arrival of an information-based economy has profoundly changed the post-war international order, a transformation which played itself out fully the moment that the East–West division of the world collapsed in 1990. In essence, studies on development had concentrated on the transition of nation states from agrarian to industrial. With globalization, the coordinates of modernization have changed; the agenda is now dominated by the shift of power from nation states to transnational markets and from industrial structures to informational structures. Deterritorialization and dematerialization (*see* **Dematerialization of the Economy**, Volume 4) have emerged as powerful trends, which escape the categories of development, without however, canceling the aspirations behind it. For the hopes which nourished the development creed are more alive than ever – the hope of the poor for a life in dignity as well as the hope on the part of Southern elite to be finally on an equal footing with the affluent North.

Globalization Instead of Development

In the course of globalization, what can be called the Westphalian constellation (Menzel, 1998) is reaching its end. For it was after the Peace of Westphalia in 1648, which established the principle of territorially bounded sovereignty that a particular form of polity, the European nation-state, came into existence. In its idealized version, the nation-state circumscribed a territory upon which a polity, an economy, a nation, and a culture rose. Like a container, it was supposed to hold society in all its layers within a demarcated space, creating a self-enclosed entity which in turn engaged with other such entities on the international level (Beck, 1997). Though reality never conformed to this

conception entirely, these containers finally burst open with globalization. Goods, money, information, images, people flow across borders, giving rise to a transnational social space where interactions occur over long distances, sometimes even in real time. As a consequence, the former (though always partial) integration of economy, polity, and culture within a territory breaks apart, turning states into just one player amidst transnational networks of exchanges in many spheres of life.

Against the background of these changes, development loses both its object and its agent. As a matter of course, the development discourse had focussed on the transformation of territorially bounded societies; they were thought to be the units by which social evolution proceeds. States were thus the privileged sites of development. However, as societies are perforated by transborder flows (be it through foreign capital, satellite television or migrants), the objective of development planning begins to dissolve. Attention no longer focuses on developing national economies, but either on inserting certain players successfully into the world market or on securing livelihoods for local communities. Likewise, the agents of development change. While previously the state was expected to be the engine of development, several new agents, all of them moving largely irrespective of borders, now diminish the role of the state. In this vein, private foreign investment has overtaken public assistance, television imagery has superseded national narratives, and non-governmental organizations (NGOs) have shouldered many development projects. At the present time, the development concept looks strangely out of place in the era of globalization.

Furthermore, the development creed had been embedded in a conception of linear worldwide progress, which fails to resonate with a globalist mindset. During the heyday of industrial modernity, history had been conceived as a movement with a direction, a universal process whose pointer was called rationalization or liberation (Bauman, 1992). This predominance of time over space in ordering worldviews, however, has been turned around by the shift in consciousness, which is connected with globalization. For what captures the attention of the post-modern mind is not a clear universal sequence of social change, but the simultaneous presence of a plurality of differences across geographical or virtual space. In perception, space gradually gains the upper hand over time; it is not the sequence of things that matters, but their possible combinations. The present change in guiding metaphors can be taken as an illustration; the road of progress is being replaced by the connectivity of networks. While the former saw societies as progressing along a time scale, the latter sees shifting patterns of flows between non-contiguous locations. In the transnational and digital world where success depends on being inserted into relevant, ever shifting circuits and not primarily on the position of one's country on the racetrack, the development idea ceases to express the excitement of the day.

The New Divide

Among other things, *globalization* tends to undercut social solidarity, both nationally and internationally. As societies are less and less contained within nation states, the reciprocal links between social classes, which constitute a polity, become weakened. After all, the nation state, in particular as long as social Keynesianism had been on the agenda, was capable of rebalancing the relations between the rich and the poor, be it as welfare state in the North or as developmentalist state in the South. Under the pull of the transnational economy, however, the social contract which lay at the core of the redistributive state has begun to unravel. As the elites aspire to catch up with the vanguards of the international consumer class, their old-style sense of responsibility for the disadvantaged sections of their own society withers away, because they themselves, instead of feeling superior with respect to their countrymen, feel now to be inferior with respect to their global reference groups. Following this drift, governments are inclined to ally themselves with the globalizing forces and increasingly show disregard for the majority of their citizens who live outside the global circuit (Kothari, 1993). Committed to promoting the insertion of their industries and middle classes into global markets, they consider the non-competitive social majority a liability rather than a boon. As a result, in many societies a split opens up between the globally oriented middle class on the one side and (in terms of the world market) superfluous populations on the other. While globalization removes barriers between nations, it thus erects new barriers within nations.

Equally, the social contract between rich and poor nations, which, all counter forces notwithstanding, had laid the base for international development policy after the Second World War, did not survive the onslaught of transnational competition. Neither within nor between states is there much concern any longer for redistribution. Already in the 1980s the politics of structural adjustment largely replaced the development consensus, giving priority to macroeconomic stability in favor of an unhampered transborder mobility of capital. Deregulation and liberalization were supposed to bring indebted countries up to the standard of a free-trade player: yet in many cases the less advantaged sections were just brought to their knees. Disregarding social and environmental costs, currency stability became the entrance ticket to the circuit of transnational capital flows, and thus became the overriding objective of the IMF and the World Bank. This implied a shift of focus; the concern of dominating development agencies was now creating a stable playing field for transnational corporations and was no longer interested in improving the welfare of

the people. However, the process by which development institutions could become agents of the world market rather than agents of societal welfare was facilitated by the success of about ten emerging markets in export-led growth. Along with the former rise of oil producing economies and the dissolution of the Eastern Block, this success finally split the Third World as a somewhat homogenous group of nations. South Korea, for example, which in 1960 had still been on a par with Bangladesh, produced in 1996 as much output as the entire Sub-Saharan African region. Globalization, in other words, made a number of Southern economies (or regions therein) into players on the world market, at the price, however, of driving a deeper wedge between them and the large majority of countries in the South.

The more that shifting transnational space encompasses not entire countries, but only larger or smaller sections of a country, the more obsolete the North–South division becomes. Indeed, the dividing line, if there is any, in the world of globalization does not primarily run between Northern and Southern countries, but the line separates the global middle class on the one side and the excluded social majority on the other. The global middle class is made up of the majority in the North and a smaller elite in the South; its size equals roughly that 20% of the world population which has access to an automobile. Globalization accelerates and intensifies the integration of this class into the worldwide circuit of goods, communication and travel. But an invisible border separates in all nations, in the North as well as in the South, the rich from the poor. Entire categories of people in the North, like the unemployed, the elderly and the competitively weak, just as entire regions in the South, like rural areas, tribal zones and urban settlements, find themselves excluded from the circuits of the world economy. Even information-based capitalism, by linking valuable players and places in a non-contiguous pattern, turns people and territories across countries into black holes of informationalism (Castells 1998, 161). At any rate, the major rift today appears to be between the globalized rich and the localized poor; the North–South divide, instead of separating nations, runs through each society, albeit in different configurations.

Security Instead of Development

As the development consensus faded away, two themes have emerged in its wake. The first is globalization and is concerned with the stability of the transnational economy. Its aim is the expansion of global markets leading to greater welfare. The so-called Washington Consensus of 1986, which declared structural adjustment to be the highest form of development, can be considered its take-off point, and the IMF its guardian. The second theme cherishes security and is concerned with protection against risks. Its aim is the need for prevention in the face of threats to human survival and dignity. UNDP's Human Development Report, which annually explores the state of human security, may exemplify this current, although the debate on *environmental security* (Mathews, 1989) is part of this mode of thought as well. To some extent, the two themes are merely a reincarnation of the 1970s conflict between top-down, pro-growth on the one side, and bottom-up, pro-poor approaches on the other.

The concern for security crystallized in the 1990s after the promises of development had lost credibility. With the high-flying optimism that once powered development withering away, the perceptions changed; the South ceased to be seen as young and full of potential, as in the time of Truman, but as the breeding ground of social and environmental turbulence. In particular, no one still clung to the ideal of a radiant future for the social majority, which has become superfluous in terms of the global economy. The best that could be achieved is survival in decency. Moreover, as globalization not only provides comforts, but many woes as well, the North felt increasingly threatened by immigration, civil wars, and environmental competition. As a consequence, the South is no longer considered with hope, but with suspicion; developing countries turn into risk zones, and their less advantaged citizens are primarily perceived as risk factors. Development policy shed its skin again. It has now largely adopted a security agenda where prevention replaces progress as the objective of development. Catching up is out of the question, assistance now aims at preventing the worst scenario from happening. Securing people's livelihood is the noble concern, projects for clean water, market access, woodless stoves, or community organization, are typical examples: a far cry indeed from reaching a modern paradise. But also the notion of security is a contested terrain; the question looms large: whose security – that of vulnerable people or the security of the OECD-dominated economy? While, for instance, many NGO's work to enable the less advantaged to protect themselves, on the diplomatic level the stakes are often different. International negotiations (in particular environmental negotiations), in a world risk society, deal implicitly with the defense of the stronger against the risk presented by the weaker. The redistribution of risks and not any longer the redistribution of economic opportunities is their hidden agenda. For both grassroots movements and governments, security has thus become a key concern because this is all that is left of the development idea after the belief in progress has vanished (*see* Environmental Security, Volume 5).

PROSPECTS

The development age might have withered away, but its core agenda is still unfinished. For decades, development

had provided language for voicing the desire for justice. Yet justice in both its basic forms, as redistribution of opportunities and as recognition of otherness, fails to have made much progress, at least beyond the confines of the global middle class. However, the agenda is not only unfinished, it also faces new constraints. For the meaning of justice is bound to change in an era of biophysical limits. As long as limits were not on the horizon, justice could be identified with growth. The famous metaphor of the growing cake which eventually offers larger pieces for everyone without imposing smaller pieces on anybody illustrates how the idea of open-ended growth could nicely sidestep hard questions on equity. But in a closed environmental space, the claim for justice cannot be reconciled any longer with the promise of material intensive growth, at least not for the world's majority. For this reason, the quest for justice will need to be decoupled from the pursuit of development with a capital D.

Contraction and Convergence

Under conditions of finitude, there is only a limited part of the global environmental space available for each country. Therefore, the concept of development as a racetrack without a finishing line is historically outdated. In order to envisage sustainability scenarios for this century, in particular in the context of climate change, it is helpful to distinguish two distinct trajectories beginning from two opposite poles and spanning a variety of initial conditions. Northern countries start their trajectory towards a low-risk and equitable level of fossil energy flows from high consumption levels, reducing them over time until they reach levels sustainable in terms of both ecology and equity. This may be called the *trajectory of contraction*. Southern countries, on the other hand, start from relatively low levels of fossil energy flows, increasing them over time until they approach the trajectory of industrial countries at sustainable levels of resource throughput. This may be called the *trajectory of convergence*. Both trajectories pose related, but different, challenges. For industrial countries, the challenge consists in bringing down resource flows without a decline in human well-being and in social justice. For Southern countries, however, the challenge consists in raising levels of resource consumption at a much smaller gradient than industrial countries did historically; increasing human well-being concurrently with equity.

A similar logic applies to inequalities within countries. Given that the omnivores are not confined to the North, retreating from excessively occupied environmental space will be expected from Southern middle classes as well. After all, the elite in countries like Mexico, China, and Brazil rival the population of many OECD countries in numbers. Therefore, the trajectories of contraction and convergence also apply to the development paths of different social classes.

Towards Resource-light Economies

The models of wealth brought about by the spectacular growth in the OECD countries in the last fifty years are structurally oligarchic; they cannot be generalised across the world without putting everyone's life chances in jeopardy. Chemical agriculture, the automobile society or meat-based nutrition are all cases in point. For this reason, the move towards models of frugal use of wealth among the affluent is a matter of equity not just of ecology. However, conventional development thinking implicitly defines equity as a problem of the poor. But designing strategies for the poor, developmentalists worked towards lifting the bottom, rather than lowering the top (Goodland and Daly, 1993). The wealthy and their way of producing and consuming weren't under scrutiny, and the burden of change was solely heaped upon the poor. In future, however, justice will be much more about changing the lifestyles of the rich than about changing those of the poor.

In order to move towards resource-light economies, two broad strategies can be distinguished. The first is the attempt to gradually decouple economic output from resource flows. For instance, enhancing the ecological efficiency of technologies and organisational structures, aims at reducing the volume of resource input per unit of economic output. In all likelihood, the efficiency of resource use can be enormously increased; examples for eco-intelligent production and services abound (*see* Dematerialization and Sustainable Development, Volume 4). The second strategy is the attempt to decouple quality of life from economic output. Indeed, quality of life has many sources beyond purchasing power; it derives from non-monetary assets as well, such as access to nature, participation in community, or the wealth in public goods. What is at stake here is not the efficiency, but the sufficiency in resource use. Such an orientation aims at the art of sustaining higher qualities of life out of a given set of material inputs; it ponders how much is needed for attaining welfare, value, beauty, and meaning. In brief, the transition to resource-light economies is likely to require a dual-track strategy: a reinvention of means (efficiency) as well as a prudent moderation of ends (sufficiency). In other words, it is about doing things right and about doing the right things.

Leapfrogging into the Post-fossil Age

For Southern economies the challenge is to embark upon growth patterns that are both pro-environment and pro-poor, without going through all the stages of industrial evolution as Northern countries did. At the present moment when the fossil-fuel age is on the decline, economies that once were said to be lagging find themselves in a favourable position. Not yet being locked into old-style industrialisation, they have the prospect of *leapfrogging* into a post-fossil

age, skipping the resource intensive styles of production and consumption of the industrial world (*see* **Leapfrogging Technology**, Volume 4). For instance, Southern countries face important decisions about introducing and designing infrastructures such as energy, transport, sewage, and communication systems, the introduction and maintenance of which in industrial countries have caused the earth s resources to dwindle. Today, many Southern countries are still in a position to avoid this unsustainable course, opting without further detours for infrastructures which would allow them to embark upon a low emission and resource-light trajectory. Investment in infrastructure such as ef - cient rail systems, decentralized energy production, public transport, grey-water sewage, surface irrigation, regionalized food systems, dense urban settlement clusters etc., could set a country on the road towards cleaner, less costly, more equitable, and less emission intensive development patterns. It goes without saying, however, that such a choice is in the rst place not a technical, but a cultural one; it requires envisaging models of wealth different from those in the North.

DEMOCRACY AND ECOLOGY

Environmental resources are valued as a source of livelihood, by groups as diverse as the shermen of Kerala, the forest dwellers of the Amazon, the herders of Tanzania, and the peasants of Mexico. Over the centuries many of these communities had developed complex and ingenious systems of institutions and rules regulating ownership and use of natural resources in such a way that an equilibrium between resource extraction and resource preservation could be achieved (*see* **Indigenous Knowledge, Peoples and Sustainable Practice**, Volume 5). However, particularly under the pressure of the resource needs brought forth by the omnivores, their basis of livelihood has been undermined, degrading their dignity and sending many of them into misery. In such a context, sustainability in the rst place means ensuring the rights of communities both to their resources and to their culture.

Democratic rights and resource productivity are particularly intertwined when it comes to ecosystem people. For their livelihood such communities need to undertake efforts to increase the productivity of all components of the village ecosystem; from grazing and forest lands to croplands, water systems and animals (Agarwal and Narain, 1989). After all, they often suffer from a shortage of biomass rather than from a shortage of cash. However, only conferring a substantial degree of control over their resource base to communities will ensure the degree of power and participation necessary for cultivating forests, elds and waters according to local rules and customs. Democratic rights and entitlement to resources are thus prerequisites for building a biomass-based, non-carbon economy. Ensuring sustainable livelihoods (the ability of an individual or a family to meet their basic needs in a manner that is digni ed, but does not undermine the natural resource base on a large scale) will therefore largely require policies which put democracy, equity, and environmental care before the quixotic pursuit of monetary economic growth.

See also: **Sustainable Development**, Volume 4; **Sustainability: Human, Social, Economic and Environmental**, Volume 5.

REFERENCES

Agarwal, A and Narain, S (1989) *Towards Green Villages*, CSE, New Delhi.

Anderson, B (1983) *Imagined Communities*, Verso, London.

Apffel-Marglin, F and Marglin, S A (1990) *Dominating Knowledge. Development, Culture, and Resistance*, Clarendon, Oxford.

Apffel-Marglin, F (1998) *The Spirit of Regeneration. Andean Culture Confronting Western Notions of Development*, Zed, London.

Arndt, H W (1981) Economic Development: A Semantic History, *Econ. Dev. Cult. Change*, **26**, 463 484.

Arndt, H W (1987) *Economic Development. The History of an Idea*, Chicago University Press, Chicago, IL.

Ayres, R U (1998) *Turning Point. An End to the Growth Paradigm*, Earthscan, London.

Banuri, T, Hyden, G, Juman, C, and Rivera, M (1994) *Sustainable Human Development*, UNDP, New York.

Bauman, Z (1992) *Intimations of Postmodernity*, Routledge, London.

Beck, U (1997) *Was ist Globalisierung?* Suhrkamp, Frankfurt.

Castells, M (1998) *End of Millenium*, Blackwell, Oxford.

Dumont, L (1977) *From Mandeville to Marx*, University of Chicago Press, Chicago, IL.

Di Meglio, M (1997) *Lo sviluppo senza fondamenti*, Asterios, Trieste.

Escobar, A (1995) *Encountering Development. The Making and the Unmaking of the Third World*, Princeton University Press, Princeton, NJ.

Gadgil, M and Guha, R (1995) *Ecology and Equity*, Routledge, London.

Goodland, R and Daly, H (1993) Why Northern Income Growth is not the Solution to Southern Poverty, *Ecol. Econ.*, **8**, 85 101.

Hobart, M (1993) *An Anthropological Critique of Development*, Routledge, London.

Hobsbawm, E (1994) *The Age of Extremes. A History of the World, 1914 1991*, Pantheon, New York.

Illich, I (1971) *Celebration of Awareness*, Boyars, London.

Johnston, B R (1994) *Who Pays the Price? The Sociocultural Context of Environmental Crisis*, Island Press, Washington.

Kothari, R (1993) *Growing Amnesia. An Essay on Poverty and Human Consciousness*, Penguin, Delhi.

Lugard, F (1922) *The Dual Mandate in British Tropical Africa*, Blackwood, Edinburgh.

Lummis, C D (1996) *Radical Democracy*, Cornell University Press, Ithaca, NY.
Mathews, J T (1989) Redefining Security, *Foreign Affairs*, 68, 162–77.
Menzel, U (1998) *Globalisierung versus Fragmentierun*, Suhrkamp, Frankfurt.
McNamara, R (1973) *Address to the Board of Governors*, Nairobi.
Nandy, A (1983) *The Intimate Enemy. Loss and Recovery of Self under Colonialism*, Oxford University Press, Delhi.
Nederveen Pieterse, J (1998) My Paradigm or Yours? Alternative Development, Post-Development, Reflexive Development, *Dev. Change*, 29, 343–373.
OED (1989) *The Oxford English Dictionary 1989*, entry underdeveloped , Vol. XVIII, 960.
Rahnema, M and Bawtree, V (1997) *The Post-development Reader*, Zed, London.
Rist, G (1997) *The History of Development*, Zed, London.
Samuelson, P and Nordhaus, W (1985) *Economics*, McGraw Hill, NY.
South Commission (1990) *The Challenge to the South*, Oxford University Press, Oxford.
Tenbruck, F H (1989) Der Traum der Sakularen Okumene. Sinn und Grenze der Entwicklungsvision, in *Die kulturellen Grundlagen der Gesellschaft: der Fall der Moderne*, Westdeutscher Verlag, Opladen, 291–307.
Truman, H (1949) Public Papers of the President of the United States: Harry S. Truman, US Government Printing Office, Washington, DC.
Tuan, Y F (1986) *The Good Life*, University of Wisconsin Press, Madison.
UNDP (1998) Human Development Report 1998.
UNDP (1999) Human Development Report 1999.
Wackernagel, M and Rees, W (1996) *Our Ecological Footprint*, New Society, Gabriola Island.
Wieland, W (1979) Entwicklung, in *Geschichtliche Grundbegriffe*, eds O Brunner, W Conze, and R Koselleck, Klett-Cotta Stuttgart, 199–228, Vol. II,.
World Bank (1992) *World Development Report 1992*, Oxford University Press, New York.
World Bank (2000) *World Development Report 1999/2000*, Oxford University Press, New York.
World Commission on Environment and Development (WECD) (1987) *Our Common Future*, Oxford University Press, Oxford.

FURTHER READING

Sachs, W (1992) *The Development Dictionary*, Zed London.
Sachs, W (1999) *Planet Dialectics. Explorations in Environment and Development*, Zed London.

A

Adaptation
see Adaptation Strategies (*Opening essay*, **Volume 4**)

Animal Liberation
see Environmental Ethics (**Volume 5**)

Animal Rights
see Environmental Ethics (**Volume 5**)

Anthropocene
see Anthropocene (**Volume 1**)

Anthropocentrism
see Environmental Ethics (**Volume 5**)

Anthropology and Global Environmental Change

John H Bodley
Washington State University, Pullman College, WA, USA

The most striking general conclusion to be drawn from the cultural ecological data in the anthropological record is that the speed and scale of resource depletion and environmental degradation accelerates with increases in the scale of culture and the concentration of social power (Bodley, 1994; 2000; 2001). It is signi cant that people living in small-scale cultures were better able to maintain long-term, relatively resilient relationships with the natural environment than peoples living in larger scale cultures. Cultural resiliency is a key feature of successful long-term human adaptability. It is the cultural ability to minimize human-caused detrimental impacts to the environment, while smoothing out the human impact of natural environmental uctuations. Cultural resiliency also means minimizing destabilizing uctuations in human population and human demands on the natural environment. The concept of resiliency is more than that of balance or equilibrium, because it emphasizes the dynamic aspects of human and natural system. These conclusions are contrary to long established beliefs about evolutionary progress, and they challenge the popular ideology that unlimited growth, especially economic growth, is a natural process, and the best way to improve human well-being.

Anthropologists have always been concerned with the relationship between the people they studied and the natural environment. Since the late nineteenth century, the first professional anthropologists focused their research on tribal peoples who were directly dependent on natural resources for their survival. The environment was obviously too important to be ignored. Initially, anthropologists studied how people exploited the environment, then they asked how culture might in turn be shaped by the environment, or by the nature of the human–environment relationship. Nineteenth century anthropologists were often natural historians, and it was common then, and still today, for them to be very broadly interdisciplinary, with interest, and sometimes formal training, in geography, geology, zoology, and botany. During the first half of the twentieth century the anthropological interest in environmental issues led to the identification of cultural ecology as a distinctive sub-discipline. Contemporary prehistoric archaeologists often maintain a focus on environmental issues.

Most early academic anthropologists rejected any simple environmental determinism, but they grouped similar cultures into large culture areas that often reflected underlying

common cultural adaptations to generalized natural areas. Before 1950, pioneer cultural ecologist Julian Steward, cautiously observed that over time people tended to organize themselves in similar ways to exploit similar environments, thus producing broadly similar cultural types in widely separate areas of the world. By the late 1960s, cultural ecologists began to apply equilibrium models to small-scale cultures, treating them as human systems in balanced, adaptive relations with natural ecosystems. These functionalist approaches were largely abandoned because they could not explain how systems developed and changed, and they often treated cultures as artificially closed systems. Since the 1980s, cultural ecologists have increasingly situated the peoples they study within the national and international political economy, and interest has shifted to approaches that will help local peoples defend their resources and subsistence economies. Many contemporary anthropologists continue an environmental focus, but it has broadened to include contemporary commercially organized, and global scale cultures, as well as existing indigenous peoples with small-scale societies.

CULTURAL EVOLUTION AND HUMAN ADAPTABILITY

One of the longest-running theoretical debates in anthropology concerns the nature of cultural evolution. The orthodox view as originally developed by Morgan (1877), and refined by White (1949, 1959) and others (Sahlins and Service, 1960), was that general cultural evolutionary progress was an inevitable and beneficial process of growth and development leading stage by stage to more complex societies, using more energy more efficiently, and with greater adaptability and security. Cultural evolutionists thought that evolutionary progress was such a self-evident human benefit that they did not demonstrate it. White (1949) emphatically declared that culture evolved as the amount of energy harnessed per capita per year increased. The assumption was that cultural evolution gave people greater control over nature, and this was good for people. However, it is significant that the most enthusiastic anthropological exponents of these measures of cultural progress wrote before the energy shortages of the 1970s, and before the United Nations Brundtland Commission (WCED, 1987) report on sustainable development. Thoughtful reconsideration of all the standard measures of cultural evolutionary progress suggests that it must be a maladaptive process.

Leslie White attributed the most recent step in cultural development to the switch to fossil fuels and nuclear energy, but it is now obvious that the advantages of this may prove illusory. Cultures can be arranged in a progressive general evolutionary sequence of increased energy use running from 5000 to 12 000 kcal per capita per day in small-scale band and village societies, 26 000 kcal in pre-capitalist agrarian civilizations, and 230 000 kcal consumed by Americans in the 1970s (Cook, 1971). However, critics pointed out that this sequence does not account for the eventual depletion of fossil fuels, the serious deficits in using non-renewable energy to produce food energy, or the costs of the waste by-products of energy production and consumption. Hubert (1969) predicted that global petroleum production would peak in 1995, and that the supply would be virtually exhausted by 2075. The most telling criticism was the calculation that if energy consumption increased at 5% a year for 200 years, the waste heat produced would equal the heat of incoming solar radiation and the earth would burn up. Fortunately, before such a *Sun Day* the polar ice caps would melt and rising sea levels would flood out most of the power plants (Luten, 1974). Human reproductive success would also be an unattractive measure of human evolutionary progress. If the production of larger, more complex cultural systems is the measure of evolutionary success, then the process still seems maladaptive for long-term human survival, because fewer, larger, more homogenous, and less durable cultures have replaced more numerous, more diverse, and more durable small cultures. Greater cultural complexity creates a human survival problem in part because larger, more complex cultural systems incorporate and subordinate smaller systems.

Some of the evolutionary theorists recognized that small-scale cultures might actually be better adapted to particular local ecosystems than large-scale cultures. White (1949) observed that foragers, with the simplest cultures, may have had the most satisfying kind of social environment that man has ever lived in. Sahlins (1968) noted that foragers were successful because they limited their wants to the consumption levels that their environment and technology could support and they lived satisfying and affluent lives. As the contemporary environmental crisis began to unfold, many anthropologists commented on the connections between self-sufficient small-scale societies and greater social equality, and relative equilibrium with the environment (Bodley, 1975). This suggested the possibility that increased social scale caused inequality, poverty, and global environmental change.

More recent biocultural evolutionary theory shows how cultural evolution could become a maladaptive process that would undermine the resiliency of both natural and human systems. From the biocultural perspective, cultural evolution is produced by changes in culture, which is conceived from the ideas that human behavior is directed in the same way that genes create biological organisms. Cultures change as shared symbolic information changes. The important difference between genes and culture is that individuals intentionally produce and transmit culture, and they can borrow from many sources (Boyd and Richerson, 1987). However, the scale of culture influences how it is created and transmitted. Cultural transmission is frequently

biased, because people emulate the beliefs and behavior that appear to be most successful. Cultural emulation is easy and efficient, but it can lead to maladaptive runaway economic growth and power aggrandizement, which in turn cause global environmental change.

In small, domestically organized societies the members of each household are daily making cultural decisions about technology, production, and consumption. In larger scale, politically organized societies, cultural evolution becomes a political process in which a single ruler can direct the actions of thousands, or even millions of people (Durham, 1991). Household-level decision making is inherently more responsive to local social and environmental conditions, but it can be over-ridden by political rulers who may be far removed from the environmental consequences of their decisions (Rappaport, 1977a,b). Thus, it is not surprising that large scale agrarian civilizations directed by political elites, frequently collapse because they exhaust the resource base, generate social conflict, or become too costly to maintain.

THE UNIQUENESS OF THE CONTEMPORARY ENVIRONMENTAL CRISIS

Bennett (1976) describes a seemingly inevitable ecological transition from environmental equilibrium to disequilibrium. From this viewpoint there is nothing particularly unique about global environmental change. Bennett placed small-scale, tribal societies at the equilibrium end of the continuum, but argued that all people had the same behavioral propensities that would lead to drastic environmental change. However, this mixes human means, ends, and secondary consequences. Biocultural theory maintains that all people are driven by a human nature that seeks domestic security and the future welfare of one's children. Small-scale societies achieve these human ends cooperatively by remaining small, consuming resources sustainably, and resisting aggrandizing individuals who would promote security-reducing growth in consumption. Some people in commercially organized cultures competitively elevate their consumption levels in order to obtain these same ends for themselves. Bennett argues that equilibrium cultures are only pauses in the overall historical tendency toward exponential increases in environmental use and impact. However, historical tendencies are not inevitable; they are the outcome of particular events and individual decision-making.

By the 1980s, as global economic growth and environmental problems intensified, some anthropologists began to argue that people have never been in equilibrium with the environment. For example, Rambo (1985) called Malaysian shifting cultivators primitive polluters, because they put carbon dioxide into the atmosphere and lived in smoke-filled houses. He declared that primitive and civilized societies interacted with the environment in essentially the same way. In a similar vein, Krech (1999) maintains that Native Americans were not always conservationists, and were in the process of exterminating the bison before Europeans arrived. These revisionist anthropologists, and many others (Headland, 1997) were reacting against the widespread tendency of some environmentalists and deep ecologists to attribute a mystical oneness with nature and an ecological nobility to tribal peoples. The mistaken implication of such romanticism of the human past, would be that our only salvation is a return to the Stone Age. This is a misleading issue because anthropologists have abundant evidence that tribal peoples were pragmatic materialists who burned on a large scale, sometimes killed more animals than they needed, felled trees to harvest fruit, and were not always guided by the spiritual sanctity of nature. There is also evidence that prehistoric humans hunted some animals to extinction, and may have contributed to the extinction of the Pleistocene megafauna, such as New World elephants, although there is much controversy on the details (Martin, 1984).

These facts need to be viewed in a larger perspective. The reality is that until 8000 years ago, when the world was still domestically organized and inhabited by only 8 million people living mostly in mobile bands at extremely low densities, there was no need for deliberate conservation practices. Likewise, optimal foraging theory suggests that intentional over-hunting would have quickly proven unproductive, because as prey species become scarce, hunting them becomes inefficient (Smith, 1983). Archaeological evidence that Australian aborigines successfully lived as hunters and foragers for at least 60 000 years (Roberts *et al.*, 1990), and that aboriginals in Southern Africa successfully survived as a people for 130 000 years (Klein, 1979) leaves little doubt that people living in very small-scale societies produce resilient cultural systems able to maintain very long-term balances with their natural resources. The archaeological record needs to be compared with the estimates at the end of the twentieth century showing a high proportion of plants and animals threatened with extinction, or nearing threatened status (Baillie and Groombridge, 1996; Tuxill, 1999). Nothing in human prehistory or history compares with the present rate of global environmental change.

It is remarkable that the scale of global change corresponds directly with increases in the scale of culture. This was dramatically confirmed by the ARCHAEOMEDES project, an interdisciplinary investigation of 30 000 years of environmental change in the Mediterranean region initiated by the European Union in 1992. Researchers found that the measurable rate at which land degradation could be observed from land clearing, erosion, and dessication increased progressively by orders of magnitude from tens of millennia during the Paleolithic, to millennia during the

Neolithic, to centuries during the Roman era, and to decades under industrial capitalism since 1850. In one region of Spain, researchers found that half of all the erosion over the past 10 000 years had occurred in the past 500 years, and it had accelerated over the past 150 years (Leeuw, 1997).

THE DRIVING FORCES BEHIND GLOBAL ENVIRONMENTAL CHANGE

The most widely accepted anthropological explanation for the rise of cultural complexity and intensi ed environmental problems has been population pressure on carrying capacity. Carrying capacity must be de ned in relation to particular environments and technologies, and is not a constant (*see* **Carrying Capacity**, Volume 4). Population pressure is also a variable that can be different under different consumption demands and distribution patterns. Nevertheless, population pressure remains a powerful explanatory model. It is a variation on Malthus (1895) observation that population has the potential to grow at a faster rate than food production. Population pressure and subsistence intensi cation have been used to explain the Mesolithic to Upper Paleolithic transition (Hayden, 1981), the domestication of plants and animals leading to the transition from the Upper Paleolithic to the Neolithic (Cohen, 1977), increased energy input and technological innovation in agriculture (Boserup, 1965), and the origin of the state and civilization (Carneiro, 1970; Steward, 1949). All of these changed the environment. Population pressure has also been explicitly linked to contemporary environmental problems (Ehrlich, 1968; Homer-Dixon, 1991).

The problem with all of these population pressure explanations is that often they do not explain population pressure itself, and they do not deal with the social inequality that also produces scarcity. The archaeological record demonstrates that for most of human prehistory population growth was minimal, not because of high mortality, but because fertility was culturally limited (Hassan, 1981). In domestic-scale cultures in the absence of political pressures women opted for small families, because extra children were a disadvantage. Only in politically organized and commercial societies are there strong externally imposed incentives for population growth. When anthropologists have looked at extreme examples of supposed population pressure producing environmental stress and human misery, whether in Bangladesh (Hartmann and Boyce, 1982), Brazil (Scheper-Hughes, 1992) or El Salvador (Durham, 1979, 1995), they have found that social inequality, not population, was the problem.

Viewed from a culture scale and biocultural evolutionary perspective, the driving force behind global environmental change is the natural human desire of individuals to improve the material security of their households under cultural conditions of economic scarcity produced by social inequality and competitive striving. The important point is that economic scarcity and environmental problems are produced culturally by social inequality, they are not natural conditions.

ANTHROPOLOGISTS, INDIGENOUS PEOPLES, AND ENVIRONMENTAL CHANGE

Cultural ecological researchers have often helped indigenous people defend their ecosystems and natural resources against deterioration caused by the uninvited intrusion of outside commercial interests. The most important anthropological support, from both archaeologists and cultural anthropologists, has been in helping indigenous communities document their long-term use of particular places and resources, in order that extensive, traditionally owned and used territories can be legally titled to communities and protected. In some cases it has been useful to document that traditional uses were sustainable. Anthropologists have also helped demonstrate that indigenous communities have highly developed knowledge of their ecosystems, including the names and natural histories of plants and animals. Some indigenous communities have asked researchers to help them with the dif cult problem of managing natural resources for both subsistence and commercial uses. Other communities have sought to protect portions of their territories for eco-tourism (*see* **Indigenous Knowledge, Peoples and Sustainable Practice**, Volume 5).

REFERENCES

Baillie, J and Groombridge, B, eds (1996) *1996 IUCN Red List of Threatened Animals*, World Conservation Union (IUCN), Gland, Switzerland.

Bennett, J W (1976) *The Ecological Transition: Cultural Anthropology and Human Adaptation*, Pergamon Press, New York.

Bodley, J H (1975) *Victims of Progress*, 1st edition, Cummings, Menlo Park, CA.

Bodley, J H (1994) A Cultural Scale Perspective on Human Ecology and Development, in *Advances in Human Ecology*, Vol. 3, ed L Freese, JAI Press, Greenwich, CT, 93 112.

Bodley, J H (2000) *Cultural Anthropology: Tribes, States, and the Global System*, 3rd edition, May eld, Mountain View, CA.

Bodley, J H (2001) *Anthropology and Contemporary Human Problems*, May eld, Mountain View, CA.

Boserup, E (1965) *The Conditions of Economic Growth*, Aldine, Chicago, IL.

Boyd, R and Richerson, P J (1985) *Culture and the Evolutionary Process*, University of Chicago Press, Chicago, IL.

Carneiro, R L (1970) A Theory of the Origin of the State, *Science*, **169**, 733 738.

Cohen, M N (1977) *The Food Crisis in Prehistory: Overpopulation and the Origins of Agriculture*, Yale University Press, New Haven, CT.

Cook, E (1971) The Flow of Energy in an Industrial Society, *Sci. Am.*, 224(3), 134–144.
Durham, W H (1979) *Scarcity and Survival in Central America: Ecological Origins of the Soccer War*, Stanford University Press, Stanford, CA.
Durham, W H (1991) *Coevolution: Genes, Culture, and Human Diversity*, Stanford University Press, Stanford, CA.
Durham, W H (1995) Political Ecology and Environmental Destruction in Latin America, in *The Social Causes of Environmental Destruction in Latin America*, eds M Painter and W H Durham, University of Michigan Press, Ann Arbor, MI, 249–264.
Ehrlich, P (1968) *The Population Bomb*, Ballantine, New York.
Hartmann, B and Boyce, J (1982) *Needless Hunger: Voices from a Bangladesh Village*, Institute for Food and Development Policy, San Francisco, CA.
Hassan, F A (1981) *Demographic Archaeology*, Academic Press, New York.
Hayden, B (1981) Research and Development in the Stone Age: Technological Transitions Among Hunter-Gatherers, *Curr. Anthropol.*, 22(5), 519–548.
Headland, T N (1997) Revisionism in Ecological Anthropology, *Curr. Anthropol.*, 38(4), 605–630.
Homer-Dixon, T F (1991) On the Threshold: Environmental Changes as Causes of Acute Conflict, *Int. Secur.*, 16(2), 76–116.
Hubert, M K (1969) Energy Resources, in *Resources and Man*, National Academy of Sciences, ed W H Freeman, San Francisco, CA, 157–242.
Klein, R G (1979) Stone Age Exploitation of Animals in Southern Africa, *Am. Sci.*, 67(2), 151–160.
Krech, III, S (1999) *The Ecological Indian: Myth and History*, W W Norton, New York.
Leeuw van der, S E (1997) *ARCHAEOMEDES: a DG-XII Research Programme to understand the Natural and Anthropogenic Causes of Land Degradation and Deserti cation in the Mediterranean Basin*, University of Paris, Paris.
Luten, D B (1974) United States Requirements, in *Energy, the Environment, and Human Health*, ed A Finkel, Publishing Sciences Group, Acton, MA, 17–33.
Malthus, T R (1895) *An Essay on the Principle of Population (Parallel Chapters from the [1st] and [2nd] Editions)*, Macmillan, New York.
Martin, P S (1984) Prehistoric Overkill: The Global Model, in *Quaternary Extinctions: a Prehistoric Revolution*, eds P S Martin and R G Klein, University of Arizona Press, Tucson, AZ, 553–573.
Morgan, L H (1877) *Ancient Society*, Holt, New York.
Rambo, A T (1985) *Primitive Polluters: Semang Impact on the Malaysian Tropical Rain Forest Ecosystem*, Anthropological Papers no. 76, Museum of Anthropology, University of Michigan, MI.
Rappaport, R A (1977a) Maladaptation in Social Systems, in *The Evolution of Social Systems*, eds J Friedman and M J Rowlands, Duckworth, London, 49–71.
Rappaport, R A (1977b) Normative Models of Adaptive Processes: a Response to Anne Whyte, in *The Evolution of Social Systems*, eds J Friedman and M J Rowlands, Duckworth, London, 79–87.
Roberts, R G, Jones, R, and Smith, M A (1990) Thermoluminescence Dating of a 50 000-year-old Human Occupation Site in Northern Australia, *Nature*, 345, 153–156.
Sahlins, M and Service E R, eds (1960) *Evolution and Culture*, University of Michigan Press, Ann Arbor, MI.
Sahlins, M (1968) Notes on the Original affluent Society, in *Man the Hunter*, eds R B Lee and I DeVore, Aldine, Chicago, IL, 85–89.
Scheper-Hughes, N (1992) *Death Without Weeping: the Violence of Everyday Life in Brazil*, University of California Press, Berkeley, CA.
Smith, E A (1983) Anthropological Applications of Optimal Foraging Theory: a Critical Review, *Curr. Anthropol.*, 24, 625–651.
Steward, J H (1949) Cultural Causality and Law: a Trial Formulation of the Development of Early Civilization, *Am. Anthropol.*, 51, 1–27.
Steward, J H (1955) *Theory of Culture Change*, University of Illinois Press, Urbana, IL.
Tuxill, J (1999) Appreciating the Benefits of Plant Biodiversity, in *State of the World 1999: a Worldwatch Institute Report on Progress Toward a Sustainable Society*, eds L R Brown, C Flavin, H French, and L Starke, W W Norton, New York, 96–114.
WCED (World Commission on Environment and Development) (1987) *Our Common Future*, Oxford University Press, Oxford, New York.
White, L A (1949) *The Science of Culture*, Grove Press, New York.
White, L (1959) *The Evolution of Culture*, McGraw-Hill, New York.

FURTHER READING

Politis, G G, Prado, J L, and Beukens, R P (1995) The Human Impact in Pleistocene-Holocene Extinctions in South America: the Pampean Case, in *Ancient People and Landscapes*, ed E Johnson, Museum of Texas Tech University, Lubbock, TX, 187–205.

Archimedes

see Anthropology and Global Environmental Change (**Volume 5**)

Art and the Environment

Peter Timmerman
University of Toronto, Toronto, Canada

This artistic dimension of global environmental change is complex and underexplored. It ranges from the roots of the

artistic impulse in human beings that impels them to explore, shape, and 'make their mark' on the world; to the artefacts in art, music, architecture, gardens, etc., that result from this impulse. If art is often what human beings make when they are at their freest and most creative, it also thrives on the transformation of the real, the material, the tightly constrained into works of art that capture some otherwise inexpressable element of the nature of being in the world. We can see that the artistic expression of the environment is always happening; but it seldom expresses itself directly. The dynamic pressures of global environmental change are similarly nding their way into the arts of today's world. Today's artist is a direct inheritor of the Romantic idea that the great artist (or the great work of art from the great artist) is some kind of 'antenna of the race': and apart from the elemental human delight in art, we can also explore art for its extraordinary capacity to intimate the deeper undercurrents within what has happened, is happening, and is to come in the world around us.

The following article deals primarily with the artistic impulse in general (primarily in painting). See ***Landscape, Urban Landscape, and the Human Shaping of the Environment****, Volume 5 for more details on a related topic.*

ELEMENTS OF ART AND THE ENVIRONMENT

The earliest artworks we have are symbolic of many aspects of the arts and the environment, including global environmental change. The cave paintings in Altamira, Spain, Lascaux, France, and elsewhere, which date from roughly 35 000–15 000 years ago, depict on their walls large herds of animals that were soon to be extinct; and the movements of those herds can be linked to the progress and retreat of the ice sheets of the last ice age. Perhaps of even more relevance is the fact that we cannot say what these exquisite works of art were for. We can call them art, but there are various theories of their purpose; charms to increase the herds, sympathetic magic to give the hunter some control over the beasts, an expression of spiritual kinship with the animal world, or simple delight in being able to represent them.

This mysterious entrance into the relationship between the arts and the environment persists. The most obvious aspect of this relationship is representation, the picture of a tree or an animal that impresses us because it looks like, or at least reminds us, of that which is being represented. There seems to be a basic human delight in the power or playfulness of drawing or painting an image of something else on a different medium. But there are other aspects of the impulse towards art that are at least as important; and, given that many people are now familiar with abstract act, it is clear that art is not all about lifelike pictures. Among the main aspects of the artistic impulse worth considering are:

- the medium, that is, what is the basic material in which the artist is working, its characteristics, its limitations, its possibilities.
- The content, that is, the subjects, themes, or ideas that the artist is using the medium to express.
- The role played by the work of art, that is, is it a picture for sale, a building for living in, a ritual mask for a dance, or whatever.
- The audience, who the work is for, if anyone; and, lastly,
- the artistic impulse and expression, the capacity to create.

THE ARTIST AND THE ENVIRONMENT

In considering the arts and the environment, we can look at the last of these aspects first, since it will give us a perspective on the tensions and dynamics of artistic expression. In his pathbreaking book, *Playing and Reality*, the child psychoanalyst D W Winnicott (1963) argued that when infants begin to deal with the world, they are initially immersed in an non-separate environment, since they have not yet created a sense of themselves as individuals. The separation into selves (primarily mothers and children) takes a long time, and the fears and tensions involved are often handled through play. Play is the crossing backwards and forwards across boundaries that are in part created and discovered through play. A child learns how to move, what things break and what things don't, and what actions will get punished (or rewarded) and by whom. Play is the process whereby this can happen without too much danger.

For Winnicott, the ability to play is central to the creative process. He sees that, among the problems children (and artists) face are the times when their play is somehow constrained, or becomes truly dangerous, or the transition to reality is somehow thwarted. In many ways, the artist is someone who seems to be able to keep the impulse to play with the materials, boundaries, and environment around him or her into adulthood. It is this which helps to explain their ability to make connections that have not been made by others; to be humorous about or threatening to the *status quo*; and occasionally to provide insights into worlds that some believe to be beyond human boundaries. At the same time, this ability has often been seen as unholy or dangerous; and there is a consistent thread running through the history of artistry as being to madness near allied.

EARLY ENVIRONMENTAL ART

This initial focus on the artist is somewhat misleading, since not only do we have no records of individual artists for many centuries; but the whole idea of the artist is a

somewhat modern or romantic casting of the topic towards the individual. Nevertheless, the idea that the creative impulse is associated with the playing with or negotiation of boundaries in one's surrounding environment can be very fruitful.

Many of the cave animal paintings can be seen to respond to the cracks and crevasses of the walls, humps of animals being painted on outcroppings of walls, etc. In other animal and spirit paintings from around the world, we can also see that there is a kind of intimacy with the physical material supplied by the environment. For instance, in the Peruvian desert south of Lima, there are vast geoglyphs (earthmarks) believed to be about 4000 years old including animal and geometric patterns, hundreds of meters in size. Many aboriginal communities engaged in making petroglyphs (stone marks) and other paintings both above and below ground.

This *earth art* (which has recently been resurrected by contemporary artists seeking to express concerns about the environment) is, as already mentioned, possibly linked to various kinds of spiritual ritual or magic. Anthropological research in the last 200 years has shown that in many kinds of spiritual ritual or magic, a symbolic structure is established linking the patterns created by the tribe or priesthood or shaman on the earth with the spiritual realm. It is very important to recognize that the strong boundaries between the human, the animal, and the spiritual worlds that are characteristic of contemporary Western life would be seen as bizarre to virtually all other peoples in the history of the world. Rather, the boundaries between these worlds or realms were seen as permeable and changeable and transitory.

The sacred spaces and sacred times set up by rituals were often designed to mimic the underlying order of the cosmos, and the artistic artifacts (masks, images, music, dancing) as aspects of that order; or as a part of the continuing conversation with the spiritual world. This role of the art object persists, not only in the familiar paintings of the Madonna and Child over the altars of many Christian churches around the world, but also in the residual feelings of people when they look at a Cezanne painting and somehow sense that they are in touch with some deeper aspect of reality as portrayed by a dish of apples.

When we look at the origins of architecture, we can see that the megaliths of ancient Europe (like Stonehenge), the great temples of the Mayans, and the vast Pyramids, all have connections and orientations towards the sacred and the spiritual. Even in domestic architecture, the elementary forms of building (walls, roofs, spaces for entrances, exits, etc.) mingle the utilitarian with the orientation towards meaning; the hearth, the shrine, the basic axis of the village or town.

Much of the official art that survives, associated with the large kingdoms of Egypt, Sumer, Assyria, etc., of the ancient world, depicts the natural world in accordance with some aspect of the ritual needs of the societies involved, though there is obviously some delight in the excellence of the craftsmanship and artistry involved in these depictions. For example, in Egyptian art, such as the palette of King Narmer, *circa* 3000 before the Christian era (BCE), or mural panoramas of voyages down the Nile (or the river of Death), one finds a sensitivity to the lush river world, and to the dignity of animals (many of which were worshipped as gods). Assyrian art, depicting the hunting and killing of lions and other animals, glorifies the powers of the all-conquering king, but also depicts the throes of the animal as well.

ART AND THE ENVIRONMENT IN THE CLASSICAL AND MEDIEVAL PERIODS

Art historians are confronted with some severe difficulties in dealing with representations of the environment in the classical arts of Greece and Rome, because so little painting remains apart from mosaics and some paneled rooms. These tend to favor the decorative and the pastoral (the pastoral being an art form that evokes the rural retreat of the wealthy from the increasingly dirty town). Much of the rest of landscape representation is symbolic or serves as a backdrop to human endeavors.

In Greek culture (epitomized by the flourishing of Athens from 600–300 BCE), landscape as a theme is practically non-existent, except in two areas: the continuing presence of the pagan gods of nature in certain forms of nature (e.g., the home of the gods on Mount Olympus) including the local gods of place; and the appearance in the Hellenistic period (after 300 BCE) of the beginnings of proto-sciences such as geography and epidemiology that obliquely made reference to natural forms and processes. From later Roman paintings (often copies of earlier Greek artworks), we can recover a variety of versions of pastoral, the rural scene, and make-believe landscapes illustrating mythological scenes from Homer or other legends.

Through the period of the Roman Empire, there continues this somewhat haphazard interest in nature. The arrival of Christianity had a twofold effect: first of all, as Christianity became more powerful, it systematically destroyed as much of the previous pagan tradition (including pagan shrines and holy groves) as possible; and second, it emphasized even more strongly the peripheral role of the natural world. What was now central was the spiritual struggle of Man towards God, and the natural environment was replaced by the spiritual environment. Images of animals or plants (when they did not appear as marginal decorations to manuscripts) were saturated with symbolic reference, exemplified by the sheep and the shepherd that can be found portrayed around the cupolas of Byzantine churches.

While there were breaks in this artistic tradition (most notably with the reevaluation of the natural world associated with the figure of St. Francis of Assisi *circa* 1200 AD), it is not really until the arrival of the Renaissance in the 15th century that the natural world by itself is seen as anything other than a space of symbols with no inherent interest of its own, though it was occasionally connected to the power and splendor of an all creating Deity. Indeed, in the evolving tradition, Nature became more and more seen as a wild space, symbolic of the spiritual dangers of wilderness and irrationality.

If we look brie y at the neighboring world of Islam, which ourished in the Middle East from the 7th century, and provided a strong challenge to the medieval West, we nd that while images of nature were banned from representation, the whole tenor of Islam in this period led to extraordinary advances in science, geography, history, astronomy and chemistry. Islam was much more favorable to the exploration of nature as an expression of the signs and symbols of Allah than Christianity in this period. Unfortunately this does not appear overtly in the artifacts that survive from that time, which are almost completely made up of abstract decoration (derived, in part, from the example of beautiful cursive patterns of Arabic writing of passages from the Koran). Later developments (in the 14th century and after) in, for example, Persian miniature paintings that do contain landscapes were in uenced by Chinese paintings (Islam connecting with China down the long trade routes).

ART AND THE ENVIRONMENT IN THE ORIENT

Some reference should be made to artistic expression and Nature in the Oriental world, speci cally China (Chinese principles in uenced the arts of Korea and Japan, and make it an appropriate single example of a vast topic). There are two main ideas at work in those Chinese arts associated with environment, which include such arts as *geomancy* (known as *feng-shui* or the appropriate placing of buildings and parks), poetry, and landscape painting.

The rst idea, which eventually becomes associated with the spiritual tradition of Taoism, but which goes back to the dawn of Chinese civilization, is that there is no separation between the order of the natural world and the order of the moral or social worlds. There is a kind of natural moral order in the world, called *the Way*, *the Tao*. Human dif culties come from straying away from that order. Similarly, proper balance of natural forces will help to improve the moral world. For instance, mountains and water are like interacting opposites (what the Chinese refer to as *yin* and *yang*) (*see* **Land Transformation in China: 3000 Years of Human Ecological History**, Volume 3); the height of the mountain balances off the depth of the water; the movement of water contrasts with the stillness of the mountain. The placement of dwellings, and the contemplative isolation of the scholar or monk, all rely on the balancing energy of natural elements.

The second idea, associated with the philosopher Confucius (*circa* 600 BCE) and his in uential descendents, is that the natural, moral, social, and political orders should all be in harmony, and that the central pivot of this harmony is not necessarily the natural order, but rather the cultivated order of the nobleman (or ruler). Confucius occasionally uses natural imagery, one of his sayings is that The wise delight in water; the good delight in mountains , referring to the balance between appropriate action and the stability of the ancient state and rules of justice, but his emphasis is on the cultured order. This approach was particularly in uential in the extremely elite cultures of the T ang (618 907 AD) and later Chinese dynasties saturated with Confucian thought.

The artistic expression of these ideas is most familiar in the Chinese landscape scrolls, which unfold vast misty landscapes expressing Taoist environmentalism; in which tiny meditative gures or rustic shermen can occasionally be picked out. Many landscapes (and later gardens associated with this artistic impulse) also gesture in the direction of the Taoist Lands of the Blessed: the future Paradise within which all will dwell. The tiny gardens that can be found around Chinese and Japanese Buddhist (and Taoist) monasteries and temples attempt to reproduce in stones, mosses, sand, and miniature trees, the vision of these lands.

In painting, the poetic and the theoretical meet in artists like Wang Wei (699 761 AD) whose long poetic sequences unfold along with extensive landscapes that the viewer wanders through. Further, the very materials and acts of painting are resonant with orientations toward aspects of the natural order whose penetration is essential to the greatest art. Often the artist is exhorted, through disciplined exercise, to submerge or eradicate the self, allowing the larger natural order to speak through the brush. This contrasts with the very different stance taken by the Western artist (at least until the modern period).

ART AND THE ENVIRONMENT IN THE RENAISSANCE

The Renaissance period (1450 1600) in Western culture saw the beginnings of a reevaluation of the natural world, in part due to a series of struggles over the legacy of Aristotelian science which began reappearing in the West in the 12th and 13th centuries (many of the scienti c writings of Aristotle and other Greek philosophers with an interest in the workings of nature had in fact been preserved by Islamic scientists). The ferment and cross-fertilization between this Aristotelian science and the already existing Christian

tradition (based in large part on a neo-Platonic base) would eventually lead to the growth of an experimental tradition that overturned the whole medieval world view.

Curiously enough, the artistic signal for this ferment came about in part as a glorification of the perspective of the individual observer, thanks to the invention of perspective in the 1400s. Previous to this, the presentation of space in paintings was shaped by meaning (for instance, more important people would be higher in the paintings) and not by any desire for accurate representation. Perspective introduced the eye view of the spectator as a standard from which, by simple geometric rules, a consistent space could be generated within which landscapes, buildings, cities, etc., could be organized. The irony is that this space, which provided a methodology for representing nature in such a convincing way that it has become totally natural to us, is in fact an abstracting of space. Every space is treated the same: special or sacred places no longer influence the world of representation within which they are portrayed.

We can link this development of a framework of abstracted space to the similar and simultaneous development of abstracted time with the arrival of the mechanical clock. In the architecture and painting of the Renaissance, we see the twin drives of clarifying and ordering of spatial representation. This new power provided artists with the capacity to explore, almost as if they were early scientists, the phenomena of light and shade, as well as the dynamic portrayal of weather, water, and the atmospherics of distance perception (for instance, the bluing of color as a landscape retreats away from the viewer). Leonardo da Vinci is of course the most famous representative of this, and in his paintings as well as his manuscripts, we can see the probing mind of the artist who uses his art as an exploratory tool into different fields, like anatomy, mechanical engineering, and ballistics.

For environmentalists, this new power can be seen as a double-edged sword. For it is in this period that the new methods of representation become allied to the emerging forces working towards the control and reshaping of the natural environment. This was part of the restructuring of what could now be called the early modern world, which wrested control from the Church and vested it in nation states. These states embarked on a quest for power, which has not ended yet.

The most obvious of the ways in which the arts were associated with this quest included the arts of mapmaking (essential to the management of the new lands associated with the voyages of discovery; see e.g., Delano-Smith, 2001), theatrical and painterly spectacle (ranging from the glorifications of the church in the various projects in the Vatican to the various historical allegories by Rubens for the French and English kings), and the architectural grandeurs of St. Peter's Basilica in Rome.

Less obviously, the art of the period increasingly focuses on the autonomous individual who walks through a world of his own creation; or at least his own substantial manipulation. Even that seemingly most natural of all painting genres, the landscape, which originates in the Netherlands (a new state of the era), and which is characterized by the absence of human beings in a central role in the depictions, cannot get away from the fact that the picture is designed to be seen according to the rules of perspective, which require the observer to hold the whole scene together.

Towards the end of the 17th century and into the 18th–usually known as the Age of Reason or the Enlightenment–a further tension was added to the depiction of nature with the arrival of the powerful abstract notions of Newtonian time and space, and the emerging image of a physical world subjected to absolute laws of physics. This coincided with the rise of the absolutist state, whose most familiar artifact was the formal garden at Versailles, home to the Sun King', Louis XIV of France.

The great landscape artists of the period, such as Claude Lorrain (1600–1682) and Nicholas Poussin (1594–1665), depicted landscapes derived from classical history and mythology, which provided the viewer with images from an idealized memory of a Roman or Greek past that might be reproducible again. As the Age of Enlightenment continued, the ideal emotions (of virtue, duty, obedience, reverence to noble ancestors) symbolized by these paintings began to be superseded by less restrained emotions and sentiments. In hindsight, we can see that these emotions, like a breeze fluttering through the leaves of pastoral scenes, were the forerunners of the greater storms that would usher in *Romanticism* (1750–1850).

In the run up to Romanticism (which is the seedbed for modern *environmentalism*) the role of nature as depicted in paintings and articulated in poetry and literature, oscillates between the emerging picture of a universe structured like a vast clockwork operated by an increasingly distant maker; and a related, but different picture of human beings connected to each other and to nature by sympathetic and empathic emotions or sentiments; captured in poems like James Thomson's (1700–1748) *The Seasons*, where each season influences the sensitive soul. With hindsight, we can see that this oscillation is an uneasy solution to the deeper problems associated with the imaginative consequences of the final breakup of the old world view that had buttressed the old world for so many centuries.

ART AND THE ROMANTIC ENVIRONMENT

This previous world view, which had supported the social and political order of things through an ascending pyramid of order and control which was felt to reflect a divine order of increasing rationality and access to higher truths, came

to be seen increasingly as bogus or at least out of date when confronted with newly restless (and increasing) populations, the redrawing of the maps of the world and the cosmos, and the surging potential for scientific enlightenment. Artists found themselves not only profoundly influenced positively by the vast increase in human capacities to harness energy through science, technology, and industrial organization; but also profoundly worried about the consequences of aspects of that harnessing in the disruptions of the Industrial Revolution. The Romantic Revolution was, for artists, simultaneously an exultation of what was now possible for the individual or the awakened community, throwing off its ancient fetters; and a dark warning about the repressive possibilities of new powers in new hands.

Politically, this helped propel the revolutionary struggles of the French Revolution (1789), an essential moment in the rise of modernity. Artistically, the Romantic artist fastened onto the idea of the artist himself (and now for the first time practically, herself) as the embodiment of the alienated human being in a world being transformed by factories and machines. The creative artist was that special form of the alienated human being whose antennae or capacities were capable of bridging (after creative or political struggle) the emerging gap between the self and the outside world, being given over increasingly to industrialization. The allies of the alienated human being in his or her struggle could most obviously be found in the similarly struggling realm of nature, since nature was herself becoming more and more subjected to forces of mechanization and alienation. At the same time, however, natural places could provide a refuge from the increasing power of human technology.

The continuing landscape tradition in art was obviously an important vehicle for these emerging themes. In 1757, the English philosopher Edmund Burke wrote his immensely influential treatise on aesthetics, *A Philosophical Enquiry into the Origin of Our Ideas of the Sublime and the Beautiful*, in which he crystallized an emerging theme of the period, which was that the natural world was, among other things, potentially a place of emotional (sublime) power; for example, the sense of overpowering awe at the sight of Niagara Falls. This was, as it were, the most powerful way to overcome the split between human and nature: by the shock of brute natural experience.

Landscape as a symbol or a carrier of emotion was central to the 18th and 19th century paintings of Gainsborough, Constable, and Turner in England, Caspar David Friedrich in Germany, the Barbizon School in France, and others. Some of the themes of these paintings include:

- vast mountains, waterfalls, landslides, or immense forests within which the human presence is dwarfed or muted into meditative silence (Turner; Friedrich);
- nostalgic scenes of ordered country life, before the onslaught of industrialization (Constable);
- Oriental, historical, or classical landscapes that are far from contemporary life (Corot; Bonington; Delacroix);
- the simple rustic life of peasants and farming (Millet; the Barbizon school).

The whole range of Romantic nature portrayal in art was simultaneously reinforced by similar themes in Romantic poetry, music, novels, etc. The sensitivity towards nature was a barometer of one's personal sensitivity; and often of one's politics.

The most important English writer on art and the environment in the 19th century was unquestionably John Ruskin. Ruskin was a central influence on a number of continuing themes, such as the return of Gothic style architecture to public buildings, the emphasis on handicraft work (that would be picked up by William Morris later), and various campaigns to save monuments (such as St. Marks in Venice). His greatest claim is probably that he taught a couple of generations of artists and writers to engage in the intense study of natural forms and details. Ruskin believed that in the details of natural forms could be found spiritual patterns and truths, and that they were portals of self (and social) awakening.

The descendants of Ruskin include not only the Pre-Raphaelite painters of his own day, some of them who were committed to Ruskinian detailing in their work, but much later photographers like Ansel Adams who specialised in intensely grained portraits of natural objects like rocks, trees, and waterfalls. The mention of photography also reminds us that in Ruskin's day artists were being challenged by the new phenomenon of the photograph that could capture details of the world in an instant, details that might take days or weeks to obtain on a canvas.

The French Impressionists, though in many ways quite different than photographers in their aims, nevertheless found themselves often engaged in kinds of work allied to at least some of the tendencies that could be found emerging (over time) in photographic aesthetics. Among these were the emphasis on the immediate response to a scene; the organizing of a scene by light, shade, and color rather than by underlying physical structures or the rules of classical perspective; and the increasing use (particularly by Degas) of seemingly accidental frames around scenes so that pictures began to look as if they were sudden glimpses into reality rather than long-meditated artifacts.

In a sense, the Impressionists were the culmination of the Perspective Project, except that their observer was acutely sensitive to the immediate atmospheric environment, so that form begins to dissolve into light, color, and seasonality, as in Monet's late studies of Rouen cathedral and haystacks at different times of year. Almost immediately, this dissolution was countered by returns to certain kinds of classical formalism (as in the paintings of Cezanne),

and the deploying of forms and visual elements generated, not by the physical world, but by symbolic or emotional expressiveness. The most famous examples of these are the late works of Vincent Van Gogh, whose Sunflowers are simultaneously visionary piercings into the primal energy of the natural world, and also expressions of the artist's own emotional vocabulary of line and color. Cezanne, too, in his intense study of the modeling of natural objects, finds himself pressed into creating a painting language remote from the previous conventions of representation.

ART AND THE MODERN ENVIRONMENT

It is a truism of art history that the late 19th century pressed the traditional norms of representing the natural environment to their limits; and that Picasso (drawing on the work of Cezanne and others) broke those norms completely. *Modernism* in art, literature, and other areas, was exemplified by the collapse of a primary logic of correspondence. In a logic of correspondence, x and y in the artwork correspond to X and Y in the world, and that is how representation works. This is replaced by a logic of coherence, where the internal coherence of the elements of the artwork is more important than their potential correspondence to an outside world. A novel like James Joyce's *Ulysses* (1922) gets much of its power from its one-to-one correspondence to a day in Dublin, it is the last gasp of the traditional novel. Joyce's last novel (modernist or postmodernist), *Finnegans Wake* (1941) has only the vaguest external reference: its structure is internal (and in fact, the novel, if it is a novel, has no beginning or end, but is circular).

Similarly, the most avant-garde artistic expression moves away from simple representations of the natural world, just at the moment when photography makes such representations normal, and in fact begins to provide representations of parts of the natural world unrepresented before, thanks to time lapse photography, microphotography, satellite photography, etc. Typically, a modernist painter like Marcel Duchamp, in his *Nude Descending A Staircase No. 2* (1912) will use images derived from the work of Eadward Muybridge's photosequences of movement. But he will use them to pattern the movement of descent of a figure so that it breaks up what would (in traditional art) have been a frozen moment on a stair.

There may be some relationship (though this is extremely controversial) between the coherence model of modern art and the virtual disappearance of landscape from the artist's repertoire for about 50 years following the arrival of Picasso on the scene. It survives overtly in the warped or dreamlike images of the Surrealist painters, and the cool urban spaces of the American painter Edward Hopper. Nevertheless, the art historian is aware that *abstraction*; which appears to be the most unrepresentative art of all; abstraction (associated with Wassily Kandinsky, Piet Mondriaan, and others), originated in a complex desire by these artists to abstract from the real world deeper spiritual and emotional forms than could be captured simply by representational pictures. The abstract painting often can be seen as a formal or emotionally expressive landscape (an interior landscape of the mind, or simply of the possibilities of the basic elements of paint itself) canvas, color, light, frame, etc. Certainly, throughout the modern period there have been artists engaged in representative art, but until recently, they were completely marginalized. Interestingly enough, it is in marginal forms like watercolor that many artists of nature (Winslow Homer, John Singer Sargeant, Andrew Wyeth) and many nature artists (that is, wildlife artists) worked.

One critical fact about the modern artist is his or her sophistication with regard to all previous forms of representation. The way in which artists will subvert photographic realism, for instance, or tease the viewer with false hopes that x will indeed represent X, all serve to highlight, or undermine the norms of visual representation. This kind of approach, pioneered by the Dada artists during the First World War as an attack on the kind of normalcy that could have generated a world war, fixed on photomontage and collage as possible subversive forms. Pop art, in the 1960s, worked variations on this. Today, on many fronts, artists are engaged in a kind of guerilla warfare against the juggernaut of stale or compromised images that are constantly regurgitated through the marketing machines of the economic system. A kind of anti-art can be seen wherever environmental activists are to be found; posters, parodies of advertising campaigns, cartoons, expressive collages, etc.

THE RETURN OF EARTH ART

The advent of Earth Art or Eco-Art was in many ways a reaction to certain aspects of modernism (particularly those parts of modern art that seemed to be complicit in those parts of modernity that are threatening to the natural world), but are also rooted in an artistic sensitivity to the increasing burden of global environmental change.

One distinct element of modern art has been a devotion to abstract or formal structures whose relationship to an external world are not necessarily immediate or conventional. Among the results of this in the buildings associated with the International Style (from the late 1920s) e.g., was a kind of unrooted geometricization of form and structure; i.e., you could build a perfectly formal skyscraper anywhere in the world. The best architects would, of course, relate their efforts in subtle ways to their surroundings; but the vast burgeoning of big box buildings, and the complacent hubris of post World War 2 urban planning generated (at least

among certain artists and architects) a revulsion towards the whole ethos associated with what was increasingly referred to as The Modernist Project . This Project was seen as deeply rooted in a kind of unstoppable juggernaut of capitalist development, mingling engineering, planning, market economics, and the rhetoric of democracy together into a completely seductive paradigm.

The revulsion against this has taken many forms, from the guerilla art forms already mentioned, to attempts to resuscitate a local dimension to art, architecture, and sculpture; to recover a sense of place, as well as space. This has been embodied in different forms in different arts. For example, in sculpture, some people have taken their cue from Henry Moore and Barbara Hepworth, who worked in organic forms, and were very sensitive to the relationships of human beings in the environment; while others have explored the long tradition of the monument in the landscape, the obelisk, temple, pillar. The mutual transformation of monument, landscape, and human observer has been explored in Environmental Sculpture (sculptures that are large enough to create an environment within which the observer moves) and in the very influential Earth Art.

Earth Art is in many ways a return to the glyphic projects of early humankind, exemplified by the earthworks in deserts, on mountain tablelands, and in caverns. Robert Smithson's great earthworks (like Spiral Jetty on the Great Salt Lake in Utah) and Christo's wrapping of landscapes and buildings, are different ways of highlighting the tensions created by our contemporary abilities to reshape the earth for our own purposes.

As Lee (*see* Landscape, Urban Landscape, and the Human Shaping of the Environment, Volume 5) further discusses, there are numerous variations in what contemporary artists have seized upon in our situation. To give only one example of something that might contrast with the grand scale of a Smithson or a Christo, an artist like the Chilean Cecilia Vicuna collects shells, wood, pieces of wool and vicuna (!) in what she has called precarious art; tiny pieces of nature rescued for a moment (Merewether, 1992).

It is worth also pointing out in this context that there is a widespread greening going on, most notably in architecture. With new interests in: local or vernacular buildings that were always sensitive to its environment out of necessity; in new technologies for building, heating, cooling, and otherwise improving the efficiency of buildings; and in living buildings with roof gardens, greenhouses, and other natural process devices (*see* Greening of Cities, Volume 3). There are also basic concerns with recycling, the use of chemicals in painting, and so on, that are now part of the consciousness of most artists.

CONCLUSIONS

This review of the vast array of artistic responses to environmental change has focussed on the main line of artistic endeavor. An equally brief, but completely different review, might focus instead on the art of what Westerners would consider to be marginal peoples (traditional crafts, indigenous arts, and unofficial art of all kinds) whose arts both speak of different relations to their environments, but are also more and more subject to the complexities of living in a world becoming saturated with the visual and aural products of Western commercial culture.

A third review would ignore the past almost completely, and report from the artistic trenches of today, exploring the experiences and expressiveness of artists, still acting, as the poet Percy Bysshe Shelley once described them, as the unacknowledged legislators of the race. Their ability to capture and portray aspects of reality that we have not seen before, is one reason why they are so important to us. But it is also easy to overplay that, as well; as if artists have to be spokespeople for anything. It is not clear (if it ever was) what the role of the artist is (or what art is for, if anything) but, as an artist friend of mine once said: You know that canary in the mine? Everyone always says that it was there as a warning for the miners. I think they had it there for that purpose; but I'll bet they liked hearing it sing, too.

REFERENCES

Delano-Smith, C (2001) The Hidden Meanings of Maps, *Nature*, 411, 133–134.

Merewether, C (1992) Walking on the Wild Side: Savage Paradigms, in *Eco-Art: Elaboration, Coordination and Execution of the ECOART Project*, Spala Editora, Rio de Janiero.

Winnicott, D W (1963) *Playing and Reality*, Penguin Books, Harmondsworth.

FURTHER READING

Appleton, J (1996) *The Experience of Landscape*, John Wiley, Chichester.

Beardsley, J (1989) *Earthworks and Beyond: Contemporary Art in the Landscape*, Abbeville, New York.

Janson, H W and Janson, A F (1995) *History of Art*, 5th edition, Harry N Abrams, New York.

Treagar, M (1980) *Chinese Art*, Oxford University Press, New York.

Warnke, M (1994) *Political Landscape: Art History of Nature*, translated by D McLintock, Reaktion Books, London.

Attenborough, David

(1926)

David Frederick Attenborough, British naturalist, author and broadcaster, has done more than any other person to popularize and create an interest in natural history in the UK and in many parts of the world through his many television programs and series produced by BBC television. After serving in the Royal Navy from 1947 1949, he joined the staff of an educational publishing house. In 1952 he moved to the BBC as a trainee producer and soon established his reputation through his famous *Zoo Quest* series, aired from 1954 1964, which took the viewers to many remote places Sierra Leone (1954), Guyana (1955), Indonesia (1956), Papua New Guinea (1957), Argentina and Paraguay (1958), the Southwest Paci c (1959), Madagascar (1960), northern Australia (1962) and down the Zambesi from source to mouth in (1964). His next series was *Eastwards with Attenborough* set in Southeast Asia followed by *The Tribal Eye* which examined sculpture (one of his hobbies), weaving, metal casting and other tribal activities around the world. He received worldwide acclaim and the British Academy of Film and Television Arts Desmond Davis Award for his 1979 13-part series *Life on Earth*, then the most ambitious series ever produced by the BBC Natural History Unit. This was followed by *The Living Planet* series in 1984, and *The Trials of Life* in 1990, which dealt with animal behavior. His series *The First Eden* was a survey of the impact of humankind on the lands around the Mediterranean, which was followed by *Lost Worlds, Vanished Lives*, a series on fossils. A popular series was *Life in the Freezer*, about wildlife in the Antarctic (1990). He treated plant life in *The Private Life of Plants*, which was aired in 1995. His most recent series in 1998 was on birds of the world. In this long record of natural history broadcasting, Sir David has treated many aspects of biodiversity and has drawn attention to the fragile nature of natural ecosystems, which has stimulated interest amongst many different audiences. Sir David was knighted in 1985 for his services to broadcasting, was made a commander of the Royal Victorian Order in 1991, and a companion of honor in 1996. He has received honorary degrees from many universities including Cambridge, where he studied for his rst degree in natural sciences from 1945 to 1947. He was elected a fellow of the Royal Society in 1985, the year in which he also received the Founders Medal of the Royal Geographical Society. He is the recipient of the International Cosmos Prize 2000. He has received many other awards and medals including an International Emmy Award in 1985. He has served as a trustee of the British Museum, the Royal Botanic Gardens, Kew and the Worldwide Fund for Nature and a President of the British Association for the Advancement of Science. Sir David s hobbies include music, tribal art and, of course, natural history.

His books include: *The Private Life of Plants*, BBC (1994), and *The Life of Birds*, BBC (1998).

GHILLEAN T PRANCE *UK*

B

Baha'i Faith and the Environment

Richard M Landau
Georgetown, Ontario, Canada

This article explores the basic principles and beliefs of the members of the Baha'i Faith and then examines how these can and are being applied to environmental and development challenges worldwide. Starting with a brief exploration of the basic spiritual tenets enunciated by Baha'u'llah, the prophet-founder of the Baha'i Faith, this article examines the Baha'i prescription for resolving the dif cult challenges before humanity. The Baha'i Faith began in the nineteenth century in Persia and today numbers over six million adherents worldwide. Baha'is believe that all world religions originate from a common divine source and that Baha'u'llah was the messenger from God for this era. Contained in his teachings are speci c measures to usher in a new world order based on spiritual principles.

The article concludes with an exploration of the participation of the Baha'i International Community (BIC) in the United Nations (UN)-sponsored initiatives as well as a range of development and environmental projects undertaken by national and local Baha'i communities.

The author has been a member of the Baha'i Faith in Canada since 1973 and is a graduate of Carleton University and the University of Ottawa. A leader in interfaith dialogue, Mr Landau is author of "The Willing Suspension of Belief: How the World's Religions Can Work Together."

The Baha'i Faith, which was founded in 1863, is the world's second most geographically widespread religion with more than six million adherents living throughout the world's nations, territories, islands and outposts. Following the example and teachings of their prophet-founder Baha'u'llah (AD 1817–1892), the world's Baha'is consider themselves to be the citizens of one country. Baha'is regard the world as one organic unity.

The Baha'i Faith considers the monotheistic world religions part of an ever-advancing continuum that has a design. Each religion, they assert, has its origins in a common source or Godhead. A covenant exists between God and humanity whereby God reveals his plan gradually through his messengers. This is the fountainhead of human progress. Thus, from time to time, God sends forth prophets with revelations appropriate for a specific people at a specific period of human development. In keeping with the idea of this progressive revelation, often the laws and customs of preceding revelations are abrogated with the advent of each succeeding religion. For Baha'is, Baha'u'llah has revealed God's message to humanity for the current age; an age which will be characterized by world unity.

Baha'u'llah (translated as The Glory of God), who was born in Persia, revealed numerous volumes of scriptures and laws upon which the Baha'i Faith is founded. He lays claim to being the most recent in a line of chosen messengers from God that includes his immediate precursor known as The Bab (translated as The Gate), Mohammed, Jesus, Moses and Abraham as well as Zoroaster, Buddha and Krishna. The Bab (1819–1850) who was born in Shiraz, Persia revealed in 1844 that he was the gate for one greater than himself who would begin his mission to humanity in 1863. Ecclesiastical and civil authorities in Persia, alarmed by the rapid growth of the Babi movement and The Bab's claim to a revelation from God, persecuted his adherents and martyred The Bab on July 9, 1850. Likewise, because of his teachings, Baha'u'llah spent his adult life in prison and exile in various outposts of the Ottoman Empire, finally living out his last days under house arrest in the port city of Akka, near Haifa, Israel.

While the Baha'i Faith has its origins in Islamic Persia, it is a discrete and independent faith that claims to represent the fulfillment of prophecies in the sacred texts of the preceding world religions. Baha'is hold all revealed scriptures in highest regard as the word of God, believing that the teachings of Baha'u'llah, by virtue of the fact that they are the most recent revelation from God, are the most relevant for today.

The many teachings revealed by Baha'u'llah cover every aspect of life and relations between humanity and creation. Among the most basic tenets is a belief in the unity and

interconnectedness of all things: the singularity of God; the equality of the races, sexes and all humanity; and that the chief task facing humanity is the construction of a just and merciful world-embracing civilization.

The pursuit of unity is reflected in the Faith's administrative order which includes elected local, regional, national and international administrative bodies. The worldwide headquarters of the Faith is located on Mount Carmel in Haifa, Israel – the final resting place of the remains of The Bab.

Baha'is believe unity should also characterize the relationship between humanity and the natural environment created by an all-powerful God. In the words of Baha'u'llah (1976: section CXXXII, 288), Ye are all the fruits of one tree, the leaves of one branch. From the Baha'i perspective, humanity is both physically and metaphorically linked to the world. In a letter, Shoghi Effendi (1933), a direct descendant of Baha'u'llah and known as the Guardian of the Baha'i Faith, wrote:

> We cannot segregate the human heart from the environment outside us and say that once one of these is reformed everything will be improved. Man is organic with the world. His inner life moulds the environment and is itself also deeply affected by it. The one acts upon the other and every abiding change in the life of man is the result of these mutual reactions.

Baha'is recognize that the world is undergoing rapid socio-economic transitions that make the protection of the environment and sustainable development both critical and challenging. Baha'is believe that only an integrated, balanced and comprehensive world view with a belief in a divine creator and unity of purpose will resolve environmental and development challenges. For example, when science and technology don't serve a divinely ordained purpose, they will actually contribute to the erosion of the planet's biodiversity. Materialistic civilization that replaces the idea of citizen' with consumer' cannot concern itself with the long-term viability of life on Earth.

THE INTERCONNECTEDNESS OF HUMANITY AND THE EARTH

The Baha'i view on environmental conservation and sustainable development holds that: (a) because the natural universe is a reflection of the majestic qualities and attributes of the Supreme Being, it inspires and should be accorded the utmost respect; (b) all of creation is interconnected, and (c) that the unity of humanity is the essential truth and compelling force in this age. Of this, Baha'u'llah (1976: section CXVII, 250) wrote: The Earth is but one county, and mankind its citizens. The concepts of world citizenship, prudent stewardship of the Earth, and the interconnectedness of all things is the essence of the Baha'i Faith.

Abdu'l Baha (1982) (translated as Servant of the Glory), the son of Baha'u'llah amplified this point:

> For every part of the universe is connected with every other part by ties that are very powerful and admit of no imbalance, nor any slackening whatever •••

In another reference, he remarked:

> Cooperation and reciprocity are essential properties which are inherent in the unified system of the world of existence, and without which the entire creation would be reduced to nothingness.
>
> (Abdu'l Baha, from a hitherto untranslated tablet)

At the very heart of the Baha'i view of the relationship between humanity and the natural universe is the belief that all of creation is an expression of the many names and attributes of an all-powerful God. Like the many different attributes of God, the natural realm has diverse causes or ideal environments in which it flourishes and expresses itself. Life is tenacious and can adapt itself to such diverse climates as polar, temperate, tropical and desert.

> Nature in its essence is the embodiment of my name, the maker, the creator. Its manifestations are diversified by varying causes, and in this diversity there are signs for men of discernment. Nature is God's will and is its expression in and through the contingent world.
>
> (Baha'u'llah, 1982)

> Every man of discernment, while walking upon the earth, feeleth indeed abashed, in as much as he is fully aware that the thing which is the source of his prosperity, his wealth, his might, his exaltation, his advancement and power is, as ordained by God, the very earth which is trodden beneath the feet of all men. There can be no doubt that whoever is cognizant of this truth, is cleansed and sanctified from all pride, arrogance, and vainglory.
>
> (Baha'u'llah, 1979)

Yet, while nature is seen as the repository of the many attributes of God, Baha'is are not pantheists, i.e., they do not worship nature or hold it in high esteem for its own sake. The natural realm exists to serve a humanity that has as its task the carrying forward of an ever-evolving divinely ordained world order that will usher in universal peace and harmony. As such, Baha'is believe that humanity must act as a wise steward of the natural realm, though neither nature nor humanity is at the core of the universal design. Rather, it is God.

THE ENVIRONMENTAL CHALLENGE AND SOLUTIONS

Just as humanity, the environment and spirituality are all interconnected, so too are the factors that have led to the environmental challenges. Speaking on behalf of the worldwide community of Baha'is, the BIC office at the UN issued a statement making the point:

> None of these problems – the debilitating inequities of development, the apocalyptic threats of atmospheric warming and

> ozone depletion, the oppression of women, the neglect of children and marginalized peoples, to name but a few – can be realistically addressed without considering all the others. None can be fully addressed without a magnitude of cooperation and coordination at all levels that far surpasses anything in humanity's collective experience.
>
> (Baha'i International Community, 1998a)

According to the BIC, the unfettered exploitation of planetary natural resources is one symptom of a sickness of the human spirit. Thus, any lasting solution to the environmental and developmental challenges will need to recognize the spiritual nature of each human, the interdependency of all humans, and their relationship with the environment. In other words, development will need to be more than simply for short-term economic advantage; it must also further and benefit the minds and spirits of all humanity.

Clearly, cooperation between all peoples, governments and agencies will be required to effect lasting solutions to the environmental challenges. However, the BIC points to certain trends in the world which tend to undermine the very foundations of collaboration. Among these it includes:

> ••• the widespread lack of moral discipline, the glorification of greed and material accumulation, the increasing breakdown of family and community, the rise of lawlessness and disorder, the ascendancy of racism and bigotry, and the priority given to national interests over the welfare of humanity – all of which destroy confidence and trust, the foundations of collaboration.
>
> (Baha'i International Community, 1997a)

It is the Baha'i position that only the abandonment of these destructive trends will create the necessary setting in which the spiritualization of humanity can be realized and the consequent unity and cooperation between humans can develop solutions to meet the environmental challenges.

> Such qualities include love, compassion, forbearance, trustworthiness, courage, humility, cooperation and willingness to sacrifice for the common good – qualities of an enlightened citizenry, able to construct a unified world civilization.
>
> (Baha'i International Community, 1997a)

THE NATURE OF SUSTAINABLE DEVELOPMENT

Development, in the Baha'i view, is an organic process in which the spiritual is expressed and carried out in the material (Abdu'l-Baha, 1995). As with the environmental challenge, the Baha'i view calls for broad-based organic answers that are consistent with the development of the spirituality of all people.

For example, community growth and development will need to respond to the genuine need of all people to have close contact with the natural world. This will influence all aspects of development – from design and engineering to community and land-use planning. Primary among these will be the need for carefully planned maintenance of agricultural lands.

Baha'is believe that science and technology can only provide answers to sustainable development when they take into account the needs of the human soul. For example, there is little value in building high-efficiency vast networks of concrete urban roads if the style of architecture blocks sunlight, prevents people from walking and generally leaves the human being dwarfed.

> The vast forces of science and technology must be harnessed to serve the material, intellectual, emotional and spiritual needs of the entire human family. This will require that all peoples be involved in generating scientific knowledge and determining its applications. As participation increases, technologies which have tended to desensitize and alienate, to make satisfying work and crafts redundant, to destroy the environment, and to cause sickness, infirmity or death, will, no doubt, be reconsidered, redesigned or abandoned.
>
> (Baha'i International Community, 1997b)

Stewardship, from a Baha'i point of view means that the value of nature and its preservation cannot be expressed in sheer economic terms. A more balanced approach to sustainable development can only result when planners have a deep understanding of the significance of the natural realm in the material and spiritual development of all humanity.

Consequently, good stewardship and prudent management of the Earth's resources is not merely an *add-on* that is developed in response to a paucity of the resources, but rather an essential and fundamental responsibility that must be given fullest consideration at all times. Good stewardship doesn't involve rescuing nature from environmental disasters. It involves long-term planning that minimizes any possibilities of such emergencies occurring.

Material development which serves solely an economic master is not a model favored by the Baha'is. They believe that the diverse peoples of the world will be more inclined to support development policies and programs if they are based on spiritual principles and the inherent dignity of the human being. As such, Baha'is have proposed that spiritual indicators be applied to measure the value of development in terms of its impact on the spiritual, cultural and social advancement of humanity.

These indicators are drawn from the essential teachings of Baha'u'llah. For example, one of the main tenets of the Baha'i Faith is that men and women are equal. Baha'is believe that just and sustainable development will only be possible when women worldwide are welcomed as equal co-partners in every field of endeavor.

> For Baha'is, the commitment to the emancipation of women is not a recent development nor is equality of the sexes a vague ideal. It is our conviction that the unification of the human race depends on the establishment of the equality of men and women.
>
> (Baha'i International Community, 1998b)

Another of the Baha'i development indicators concerns the equitable distribution of wealth. One of the basic tenets of the Baha'i Faith is the need to redress the extremes of wealth and poverty whereby absolute impoverishment and lavish luxury are virtually side by side. Experts tells us that there are enough resources in the world to meet the needs of all humanity. Therefore, to eliminate poverty, we will need to find more equitable methods of distribution and we will need to moderate excessive and sometimes wasteful consumption and the accumulation of wealth for its own sake. At the same time, nations will need to develop fair and equitable trade relations built on the principle that the trading partners are true equals.

If development is to be sustainable, the Baha'is suggest the following:

> Wealth is most commendable, provided the entire population is wealthy. If, however, a few have inordinate riches while the rest are impoverished, and no fruit or benefit accrues from that wealth, then it is only a liability to its possessor. If, on the other hand, it is expended for the promotion of knowledge, the founding of elementary and other schools, the encouragement of art and industry, the training of orphans and the poor – in brief, if it is dedicated to the welfare of society – its possessor will stand out before God and man as the most excellent of all who live on Earth and will be accounted as one of the people of paradise.
>
> (Abdu'l Baha, 1990)

Universal education is one of the requirements that will speed the advent of a world united to promote common cause. Education that promotes a world consciousness and the understanding that there is an integral connection between every human being will create the conditions in which humanity is united to meet environmental and developmental challenges.

Unity is a prerequisite for any effort to safeguard the Earth's habitat. The type of unity envisioned by the Baha'is encompasses much more than just geography, climatology or biology. Rather, it is the outgrowth of an undying belief that humanity is a one world community. In such a community, it seems only logical that matters of economic relations and sustainable development must be addressed with a balanced universal perspective that takes into account the world's many cultures and resources.

PROPOSED COURSES OF ACTION

Calling on principles enunciated in the revelation of Baha'u'llah, whom Baha'is regard as the messenger and prophet from God for this age of humanity, the Baha'is of the world have proposed specific courses of action that will protect the environment and define the parameters of sustainable development. Over 100 years ago, Baha'u'llah called for an international legal system, sharing of the world's resources, a realignment of the world's economic and governmental relations, and reform in the behavior and patterns of human consumption.

Drawing on these teachings, the BIC prepared a statement for the proposed Earth Charter for the United Nations Conference on Environment and Development (UNCED) in 1992. In the document, the Baha'is outlined the process for achieving universally acceptable standards:

> It is our conviction that any call to global action for environment and development must be rooted in universally accepted values and principles. Similarly, the search for solutions to the world's grave environmental and developmental problems must go beyond technical–utilitarian proposals and address the underlying causes of the crisis. Genuine solutions, in the Baha'i view, will require a globally accepted vision for the future, based on unity and willing cooperation among the nations, races, creeds, and classes of the human family. Commitment to a higher moral standard, equality between the sexes, and the development of consultative skills for the effective functioning of groups at all levels of society will be essential.
>
> (Baha'i International Community, 1997c)

The document proposed that representatives of the world's religions be assembled, possibly under the auspices of the World Bank or the UN Development Programme, to consult about spiritual principles and their impact on both the individual and the progress of society (Baha'i International Community, 1998a). Such an assemblage, the Baha'is believe, could reach common agreement on a limited number of spiritual principles and how these would provide a basis for developing policy priorities. Based on this agreement, goals and benchmarks for progress would be established and monitored by the organization under whose auspices the assemblage is convened.

The Baha'is believe that the world religions can take the initiative and collaborate because of the common thread that unites all of the world's major religious traditions.

> The changes required to reorient the world toward a sustainable future imply degrees of sacrifice, social integration, selfless action, and unity of purpose rarely achieved in human history. These qualities have reached their highest degree of development through the power of religion. Therefore, the world's religious communities have a major role to play in inspiring these qualities in their members, releasing latent capacities of the human spirit and empowering individuals to act on behalf of the planet, its peoples, and future generations.
>
> (Baha'i International Community, 1997c)

Furthermore, the cooperation of an international development agency would signal their recognition of the significance of the spiritual dimension of human nature. Already the internationally accepted Agenda 21 and The Habitat Agenda have acknowledged that the spiritual needs of the individual and of society are significant factors in human progress and are inseparable from ecological, economic, social, and cultural development.

The next step would involve the development of consultative processes on both the national and local

levels whereby communities would be encouraged to utilize and develop their own independent spiritual measures for action, derived from the larger plan. Such plans and policies would likely have the backing of many and would receive the formalized support of religious authorities and institutions.

In a reflection of the Baha'i administrative order, which has no clergy but devolves responsibility on each individual right down to the local community, the Faith stresses the importance of local action in any initiatives.

> Development must be decentralized in order to involve communities in formulating and implementing the decisions and programs that affect their lives. Such a decentralization need not conflict with a global system and strategy, but would in fact ensure that developmental processes are adapted to the planet's rich cultural, geographic, and ecological diversity.
>
> (Baha'i International Community, 1997c)

Baha'is believe that the individual has a key role in the unfolding of a planetary system of sustainable development. Therefore, acknowledging the spiritual dimension of humanity and providing for the moral, emotional, physical and intellectual development and education of each person will be a building block toward a new vision of planetary society.

To meet the environmental and development challenges, Baha'is affirm that the top-down model of community development will need to give way to a more participatory, knowledge-based and values-driven process of governance. When people view the decision-making process as something they own – not as a remote and Byzantine system of laws – they will accept their responsibility for shaping a new world.

At the very core of the environmental and developmental crises facing humanity, Baha'is believe, there is a lack of moral leadership that pervades every level of decision making – from the highest levels of government to the family unit itself. This is evidenced by the constant discovery of political scandals that reveal a bankruptcy of ethical leadership. Humanity may have even lost its ability to define and identify morality in leadership because of the barrage of messages that obfuscate and confuse the issue.

While the world's religions, development organizations, governments and individuals are all called upon to play a role in sustainable development, long-term solutions will require a new and integrated vision of global society. This vision will have as its underpinnings and its charter, a new set of values based on the belief that all of humanity is one. For Baha'is, the very bedrock and hope for a sustainable new world order is the acceptance of the oneness of humanity. This principle will cause the restructuring of the world's administration to reflect the fact that the world is one nation. This does not mean that any culture or nation must abandon its distinctive identity. In fact, the entire principle of unity in diversity, which the Baha'i Faith champions, supports and actively encourages each people's right to maintain, protect and uphold its distinctiveness in the face of the homogenizing influences of international capital.

In the Baha'i view, world unity is not mutually exclusive of cultural diversity and national autonomy. Each person can legitimately have a balanced sense of pride in his or her culture and national identity. However, every person is called to a broader notion of loyalty: the uplifting and progress of the human soul, of every human being and the entire world civilization. The Baha'i approach emphasizes that the world is one nation and it calls for a universal auxiliary language, which may in the future prove to be English. The yet to be determined auxiliary language will facilitate inter-cultural communication and will not replace peoples' own mother tongues. Each individual maintains the right to preserve his or her cultural identity and mother tongue.

> In the view of the BIC, acceptance of the oneness of humanity is the first fundamental prerequisite for this reorganization and administration of the world as one country, the home of humankind. Recognition of this principle does not imply abandonment of legitimate loyalties, the suppression of cultural diversity, or the abolition of national autonomy. It calls for a wider loyalty, for a far higher aspiration than has so far animated human efforts. It clearly requires the subordination of national impulses and interests to the imperative claims of a unified world. It is inconsistent not only with any attempt to impose uniformity, but with any tendency towards excessive centralization. Its goal is well captured in the concept of unity in diversity.
>
> (Baha'i International Community, 1997d)

The BIC believes that the change in consciousness that would be represented by the adoption of the term *world citizenship* is a prerequisite before the peoples of the planet can accept and promote a coordinated and reasonable approach to global sustainable development. The entire idea of world citizenship can only take hold when one accepts the inter-relatedness of all human beings, of the impact of their actions upon each other. It means that the world is no longer constituted of billions of discrete beings and scores of disconnected governments and trans-national corporations. With the advent of world citizenship, each accepts that his or her actions in any part of the globe is likely to have impacts well beyond the local or regional spheres of influence.

The Baha'i understanding of the implications of world citizenship extends beyond simply a new passport or slogan:

> World citizenship encompasses the principles of social and economic justice, both within and between nations; non-adversarial decision making at all levels of society; equality of the sexes; racial, ethnic, national and religious harmony; and the willingness to sacrifice for the common good.
>
> (Baha'i International Community, 1997e)

The BIC believes that the most effective method for promoting sustainable development is logically through adoption of world citizenship. The full meaning and import of world citizenship will have an impact on the way nations conduct themselves with each other. When humanity and its economic, social, and political orders are preoccupied with disunity, antagonism and rigid provincialism, the Baha'is submit that there is no room for a concerted worldwide strategy on sustainable development. In other words, any effort to realize sustainable development can only be marginally successful without the animating principles of world citizenship and one-world homeland. The prerequisites clearly call for harmony and unity amongst all peoples and nations of the world.

In a paper entitled *World Citizenship: a Global Ethic for Sustainable Development*, presented to the first session of the UN Commission on Sustainable Development, 14–25 June 1993, the BIC reflected at length on the necessary first step of establishing the concept of world citizenship. They went so far as to lay out a plan for its introduction and the requisite actions of the world leaders.

> They should foster an ethic of service to the common good and convey an understanding of both the rights and the responsibilities of world citizenship.
>
> Using the concept of world citizenship as an integrating theme, the UN should publicize its ideals, activities and goals, so that people come to understand the unique and vital role the UN plays in the world and, therefore, in their lives. Similarly, the UN should promote world citizenship in all its public activities, including celebrations of its historical milestones and tours of UN headquarters. Every UN document that deals with sustainable development should also include this principle—beginning with the preamble of the proposed Earth Charter. World citizenship must become the single most important point of ethical reference in all UN activities.
>
> The services of the advertising industry should be enlisted to promote world citizenship.
>
> (Baha'i International Community, 1997e)

THE PROMISE OF A BETTER FUTURE

The Baha'is believe that there are dual processes at work in the world: the one best characterized as spiritualizing, embryonic, and beneficial to humanity; the other is the decaying and destruction of institutions and ways of thinking that no longer serve an evolving worldwide civilization.

The Baha'is are optimistic that humanity will survive the serious environment challenges and development issues facing it. They believe that the covenant God made with Abraham and Noah and has renewed with every Messenger sent to humanity is evidence of the long-term viability of humanity. This does not, however, allow humanity to abdicate its stewardship responsibilities nor the huge commitment to persevere and make sacrifices and changes that will transform the world. Shoghi Effendi (1980) looked forward to this renewal of civilization:

> In such a world society ••• [t]he economic resources of the world will be organized, its sources of raw materials will be tapped and fully utilized, its markets will be coordinated and developed, and the distribution of its products will be equitably regulated ••• The enormous energy dissipated and wasted on war, whether economic or political, will be consecrated to such ends as will extend the range of human inventions and technical development, to the increase of the productivity of mankind, to the extermination of disease, to the extension of scientific research, to the raising of the standard of physical health, to the sharpening and refinement of the human brain, to the exploitation of the unused and unsuspected resources of the planet, to the prolongation of human life and to the furtherance of any other agency that can stimulate the intellectual, the moral and spiritual life of the entire human race.

According to the BIC, it is the actions of governments, non-governmental organizations, the forces of capital, society in general, and significant individuals that will determine how quickly humanity arrives at a universal consensus for sustainable development. The onus is on every party to consciously and deliberately give a thorough evaluation to the meaning of the goals toward which they are working. This will ensure that all parties can be effective partners in progress. The BIC says that "clear goals, meaningful policies and standards, identified programs, and agreed upon indicators of progress are necessary if advancement toward humanity's common future is to be charted and regular corrections to that course determined and carried out" (Baha'i International Community, 1998a).

BAHA'I ACTION ON SUSTAINABLE DEVELOPMENT

Representing the world's Baha'is, the BIC office at the UN has played a prominent role in the various UN-sponsored summits on the environment and sustainable development. Exemplary among these, was the participation of the BIC in the Earth Summit—the Rio de Janeiro Conference in June 1992.

The Baha'is focused on the Earth Charter (*see* **Earth Charter**, Volume 5) which they felt was potentially the most significant document under consideration at Rio de Janeiro. In numerous languages they circulated nearly one million copies of the environmental and development statements of the BIC. In the opening paragraph of its presentation to a preparatory working group of the UNCED, the BIC wrote of the Charter:

> It could offer a unifying vision for the future and articulate the values upon which a peaceful, prosperous and harmonious world society could practically be constructed. In so doing, the Earth Charter could lift the context of deliberations on humanity's future to a new level—to the level of principle. Only discourse at the level of principle has the power to invoke a moral commitment, which will, in turn, make possible the discovery of enduring solutions to the many challenges confronting a rapidly integrating human society ••• the Earth

> Charter can tap a powerful source of individual and collective motivation, which will be essential for the reorientation of the world toward a sustainable future.
>
> (Baha'i International Community, 1997f)

The Baha'i presentation to UNCED urged that the idea of the oneness of humanity should be proclaimed in the preamble to the Charter, which should then be taught in the world's schools and communicated worldwide in preparation for the organic change in the structure of society which it implies (Baha'i International Community, 1997f).

In fact, the landmark Peace Monument unveiled at the conclusion of the 1992 Earth Summit in Rio de Janeiro was an initiative of the BIC. At the Summit's closing ceremony, soil from some 40 nations was deposited by children into the 5-m high monument. Each year since the Summit, World Environment Day in Rio includes a ceremony at the monument during which soil from other nations is added. The inscription on the monument is the words of Baha'u'llah: The Earth is but one country, and mankind its citizens.

Both Baha'i communities and individual Baha'is are in the forefront of activities aimed at furthering conservation and sustainable development. The following list is a small sample of projects Baha'is are involved with worldwide:

1. The establishment of a Baha'i office of the environment as an adjunct of the BIC office at the UN.
2. Issuing a 1989 compilation of Baha'i writings – *Conservation of the Earth's Resources*. The text has been studied by Baha'i communities worldwide.
3. Numerous national and local Baha'i communities have established their own environmental offices and committees, often in cooperation with like-minded organizations. In Japan, Canada, Brazil, Taiwan, Colombia, the Philippines and other nations, Baha'i communities have established curricula for education about the environment.
4. Nur University in Santa Cruz, Bolivia, an institution of higher learning, established on Baha'i principles, offers a Masters degree in Development.
5. A range of publications dealing with environmental and developmental issues is now published by Baha'is. This includes: *One Country*, a quarterly newsletter of the BIC; *Ecologia Y Unidad Mundial*, an Argentine Baha'i newspaper; and others.
6. The Baha'i Vocational Institute for Rural Women, located in India and the Clean and Beautiful Swaziland campaign founded by a Baha'i – Dr Irma Allen – have both received Global 500 Awards from the UN Environment Programme.
7. An organic farming project by the Baha'i community of Japan teaches how to grow food without artificial fertilizers or pesticides.
8. In rural Kenya, a Baha'i-sponsored development project encourages and empowers village women to develop their own entrepreneurial weaving businesses.
9. In Bolivia and Malaysia, Baha'i communities have launched fish farming projects.
10. Baha'is in the UK, the Philippines, Singapore and Taiwan have all organized and/or sponsored arts and educational activities geared to creating awareness of the fragile environment and conservation.
11. Local Baha'i communities in the UK have become active proponents of Local Agenda 21, working with partner groups and with local authorities.
12. Working in collaboration with other organizations, the BIC hosted two World Forestry Charter Gatherings in 1989 and 1994.
13. The BIC made a formal presentation to the World Faiths and Development Dialogue hosted by the President of the World Bank and the Archbishop of Canterbury at Lambeth Palace during the Lambeth Conference, February 1998.

Finally, as if to address directly the very issue of environmental biodiversity and sustainable growth, the architecture and landscaping of each of the Baha'i Holy Sites around the world is a model of the blending of natural and architectural beauty, efficiency and diversity. Each of these sites features a diverse range of flora to reflect the Faith's teachings about diversity and the buildings are designed to complement and augment their surroundings.

For more information, visit:

- www.bahai.org – the official website of the worldwide Baha'i community;
- www.bic-un.bahai.org/i-e-env.htm – a page of the website of the BIC office at the UN, which lists and links to environmental papers released by the BIC.

REFERENCES

Abdu'l Baha (1982) *Selections from the Writings of Abdu'l Baha*, Baha'i World Centre, Haifa, Section 137, 157.

Abdu'l Baha (1990) *The Secret of Divine Civilization*, Baha'i Publishing Trust, Wilmette, 24–25.

Abdu'l Baha (1995) *Paris Talks*, 12th edition, Baha'i Publishing Trust, London, 9.

Baha'i International Community (1997a) *Sustainable Development and the Human Spirit*, UN.

Baha'i International Community (1997b) *Sustainable Communities in an Integrating World*, UN.

Baha'i International Community (1997c) *Earth Charter*, UN.

Baha'i International Community (1997d) *International Legislation for Environment and Development*, UN.

Baha'i International Community (1997e) *World Citizenship: a Global Ethic for Sustainable Development*, UN.

Baha'i International Community (1997f) *The Earth Charter/Rio De Janeiro Declaration and the Oneness of Humanity*, UN.

Baha'i International Community (1998a) *Valuing Spirituality in Development*, UN.
Baha'i International Community (1998b) *Women and Men: Partnership for a Healthy Planet*, UN.
Baha'u'llah (1976) *Gleanings from the Writings of Baha'u'llah*, Baha'i Publishing Trust, Wilmette, IL.
Baha'u'llah (1979) *Epistle to the Son of the Wolf*, Baha'i Publishing Trust, Wilmette, 44.
Baha'u'llah (1982) *Tablets of Baha'u'llah*, Baha'i World Centre, Haifa, 142.
Shoghi Effendi (1933) Letter to an Individual Baha'i, Through his Secretary, 17 February.
Shoghi Effendi (1980) *The World Order of Baha'u'llah*, Baha'i Publishing Trust, Wilmette, 203–204.

BAT (Best Available Technology)

The term BAT has undergone a significant change in meaning since its introduction into environmental policy debates in the early 1970s. Initially, BAT meant regulation through a technological standard, under which the relevant government agency requires firms to employ the pollution control or abatement technology that yields the lowest level of pollution. The US Clean Air and Clear Water Acts also instructed regulators to consider feasibility and cost factors; BAT was interpreted to mean designating of the most effective currently demonstrated and cost-effective technology. These trade-offs were highlighted in various elaborations, including BAT economically achievable, BAT not entailing excessive costs, best available demonstrated technology, and best practicable technology. The actual practice of regulation by technological standard was soon criticized on two grounds. By the mid-1970s, it was widely viewed as having the effect of stalling or even freezing the introduction of new technology; though requiring use of particular technology created a stable base level of environmental protection, it also inhibited further progress by discouraging innovation. In the 1980s, critics contended that actual technological standards were emphasizing end-of-pipe treatment rather than pollution prevention. Greater acceptance of economists' arguments that market incentives would yield more pollution control at less overall cost, and environmentalists' arguments for preventing or minimizing pollution through redesign of goods and industrial processes produced a very different policy debate in the early 1990s. BAT came to mean the production technology yielding the greatest reduction in pollution rather than a regulatory approach.

M J PETERSON *USA*

Bateson, Gregory

(1904- 1980)

Gregory Bateson was a British anthropologist, trained at St. John's College, who built his career in the US, and who also made significant and innovative contributions in the fields of communication, cybernetics, information, psychiatry and learning theory. He was the son of geneticist William Bateson who is regarded as the founder of the field of genetics, and grandson of William Henry Bateson, who had been a master at St. John's College where he brought about significant reforms. Gregory was first married to Margaret Mead with whom he pioneered the use of visual methods in ethnographic research in Bali (Lipset, 1980). They were parents of Mary Catherine Bateson, an anthropologist and contemporary author who collaborated with him and who carries on some elements of his work.

Beginning with the publication of the ethnography *Naven* in 1936 (Bateson, 1958), a consistent underlying theme in all of Bateson's work was a concern with science as a *process* of knowing rather than an accumulation of facts, and with the social implications of errors in scientific thought that occur as a result of what Whitehead referred to as the fallacy of misplaced concreteness, or that of regarding scientific explanation as a description of external and concrete phenomena, rather than as a product of interaction between the observer and that which is observed. A second fallacy he observed was that of seeking to control an interactive system of which one is a part, through quantitative measurement. An example of these types of errors is found in the characterization of ecosystems at a single level of analysis, as composed of discrete entities that respond mechanically to inputs and outputs of energy. According to Bateson, reliance on this type of analysis would only increase the likelihood of runaway ecological degradation, because the false sense of an ability to predict and control the factors of interest would only make a pathological system more efficiently pathological. This would lead to more rapid self-destruction, since it does not address the false premises upon which the model is based. Organizing society and technology around this false sense of control would also reduce flexibility and thus the capacity to respond to ecological degradation (Bateson, 1979; Harries-Jones, 1995). The most famous example of this in Bateson's work is the alcoholic who attempts to assert his mastery over his drinking problem by resorting to a stiff drink to bolster himself.

Instead, Bateson stressed the importance of relationships that provide the basis for organization, and that are a greater limiting factor than energy. Relationships, which are sustained through communication of information rather

than by energy flows, are also important as a source of information about context and meaning. Recursive relationships in turn lead to pattern, which cannot be characterized through linear logic and quantitative analysis, which are therefore inadequate for characterizing living systems. An example of this error is found in Darwinian theories of evolution, in which evolution is characterized as a linear process and a force of progress, which fails to account for organization and denies interdependent relationships among organisms and their environment (Bateson, 1979; Harries-Jones, 1995).

Another important contribution was his concept of the double-bind, a paradoxical and incoherent situation in which contradictory rules appear simultaneously relevant, such as when a child receives conflicting cues from a parent, or in which a playful situation is understood literally. In other words, given that meaning changes with context, behavior is constrained by a perceived context or definition of a relationship that is no longer relevant, which leads to nonsensical beliefs. Context is taken to include relationships as well as fundamental premises and habitual behaviors that are seldom questioned, all of which constrain action, and that are normally taken as a given. In clusters, such beliefs have been linked to schizophrenia as well as to forms of common madness (Bateson, 1979; Jaeger, 1994). Associated with this is the concept of deutero- or double-loop learning, which refers to learning about context which, in contrast with rote-learning, provides a frame of reference and meaning to a given situation. Papers addressing these and related concepts were brought together and published in his best-known book, *Steps to an Ecology of Mind* (Bateson, 1972).

Bateson alluded a number of times to a new unnamed science, for which, in his final book, *Mind and Nature* (Bateson, 1979) he offered a set of epistemological principles. In sum, he is widely regarded as having provided a new perspective on the ecological predicament. In retrospect, many of his ideas can be found embedded in the concepts of post-normal science and adaptive management of ecosystems (Tognetti, 1999).

REFERENCES

Bateson, G (1958) *Naven: A Survey of the Problems Suggested by a Composite Picture of the Culture of a New Guinea Tribe Drawn from Three Points of View*, 2nd edition, Stanford University Press, Stanford, CA.

Bateson, G (1972) *Steps to an Ecology of Mind*, Chandler, San Francisco, CA.

Bateson, G (1979) *Mind and Nature: A Necessary Unity*, Bantam Books, New York.

Harries-Jones, P (1995) *A Recursive Vision: Ecological Understanding and Gregory Bateson*, University of Toronto Press, Toronto.

Jaeger, C C (1994) *Taming the Dragon - Transforming Economic Institutions in the Face of Global Change*, Gordon and Breach, Yverdon.

Lipset, D (1980) *Gregory Bateson: The Legacy of a Scientist*, Beacon Press, Boston, MA.

Tognetti, S S (1999) Science in a Double-bind: Gregory Bateson and the Origins of Post-Normal Science, *Futures*, 31, 689–704.

SYLVIA S TOGNETTI *USA*

Best Available Technology (BAT)

see BAT (Best Available Technology) (**Volume 5**)

Biocentrism

see Environmental Ethics (**Volume 5**)

Brent Spar

The public protest against the dumping of an oil-contaminated derelict off-shore petroleum platform was a watershed in the clout of non-government organizations. Transnational giant, Royal Dutch Shell, backed down in the face of massive public outrage organized by Greenpeace.

The Brent Spar, a 4000-tonne oil installation, was to be towed from the North Sea and dumped in the North Atlantic, pursuant to a permit issued by the government of the UK on February 16th, 1995. Greenpeace launched protests in May, 1995 including the occupation of the rig off-shore by scientists and others. Greenpeace argued that dumping a rig at sea would set a dangerous precedent and would undo the progress against ocean dumping. North Sea environment ministers were to meet in June in Denmark on the issue of ocean discharges of hazardous substances within the Oslo and Paris Commission (OSPARCOM). The Brent Spar issue caught fire in a dramatic way as Greenpeace catalyzed a massive public outcry against Shell. Shell, its reputation already severely damaged by allegations of environmental and human rights abuses in the Niger Delta, initially rejected its critics' complaints. It had never backed down in the face of public protest, even in the high-profile and international controversy that followed the death of author and Shell critic Ken Saro-Wiwa at the hands of Nigerian dictator General Abacha.

But the growing momentum in the movement opposed to dumping the Brent Spar increased. By June, Shell reversed itself and, although the towing of the installation was already underway, announced it would not dump the Brent Spar at sea but would tow it into a Norwegian fjord, where it remains today. Subsequently, the controversy is remembered as one of exaggerated claims about the degree of contamination of the oil installation by Greenpeace, which has admitted its estimates subsequently proved false. Greenpeace, however, maintains that despite its errors, the precedent against ocean dumping of oil installations was important and one which was subsequently endorsed by 11 of the 13 North Sea nations within OSPARCOM.

ELIZABETH MAY *Canada*

Brower, David

(1912- 2000)

David Brower was one of North America's leading environmentalists. Over his extremely active life, he founded and steered some of the world's most influential environmental organizations. In all things, he was supported by Anne Brower, his equally ecologically aware wife.

An accomplished climber and ski mountaineer enthusiast, David Brower joined the Sierra Club in 1933, and was elected to its national board in 1941 (*see* Sierra Club, Volume 5). He became the organization's first executive director in 1952, a position he held until 1969. Under his leadership the organization grew from 2000 to 77 000 members. In 1969, he left the club staff on request. Despite a number of policy differences, Brower continued to be perennially elected to the Sierra Club Board (in 1983, 1986, 1995 and 1998), resigning in 2000 complaining that the organization lacked a sense of urgency in its campaigns. He died 6 months later.

In 1969, Brower turned his attention to the creation of a grassroots, activist organization, Friends of the Earth (FoE) (*see* Friends of the Earth, Volume 5). Brower realized his vision of a truly global organization with member groups in 63 countries. But, once again, as executive director of FoE (US), he experienced conflict with the board of directors. Brower believed FoE should pursue nuclear disarmament. The culminating issue was the board directive to relocate national headquarters from San Francisco to Washington, DC. Brower's intransigence led to his termination in 1982. When Brower went to court to challenge the FoE Board's action, the news made the front page of the *New York Times*.

In 1982, Brower founded Earth Island Institute as well as the biennial Fate and Hope of the Earth Conferences, working to integrate peace, environmental, human rights and social justice groups. His biography, *Encounters with the Archdruid* by John McPhee, is in its 27th printing.

Brower is best remembered for his successful campaigns to keep dams out of Dinosaur National Monument, the Yukon and the Grand Canyon. He also played a significant role in the creation of national parks and seashores in Kings Canyon, the Redwoods, the North Cascades, Alaska and Cape Cod, among many others.

ELIZABETH MAY *Canada*

Buddhism and Ecology

Stephanie Kaza
University of Vermont, Burlington, USA

During its 2500 year history, Buddhism has evolved across a wide range of physical and cultural geographies. From the Theravada traditions in tropical South and Southeast Asia, to the Mahayana schools in temperate and climatically diverse China and Japan, to the Vajrayana lineages in mountainous Tibet, the Buddhist teachings have been received, modi ed, and elaborated in many ecological contexts. Across this history, Buddhist understandings about nature and human-nature relations have been based on different teachings, texts, and cultural views. These have not been consistent by any means; in fact, some views directly contradict each other.

Scholars debate whether or not Buddhist philosophies of nature led to any recognizable ecological awareness among early Buddhists. Most members of early Buddhist societies, including many monks, preferred the comforts of village life over the threats of the wild. Only forest ascetics chose the hermitage path with its immersion in wild nature. The word (nature) itself has many different meanings in various Asian languages. Concepts and attitudes towards nature vary across time and place as well. Indian Buddhist literature, for example, shows relatively little respect for wild nature, preferring tamed nature instead; Japanese Buddhism reveres the wild but engages it symbolically through highly developed art forms.

Even with these distinctions, Buddhist texts do contain many references to the natural world, both as inspiration for teachings and as a source for ethical behavior. For those exploring Buddhist teachings in the context of the environmental crisis, Buddhist traditions are potential sources for philosophical and behavioral guidelines towards nature. From the earliest guidelines for forest monks to

the hermitage songs of Milarepa, from the Jataka tales of compassion to Zen teachings on mountains and rivers, the inheritance is rich and diverse.

EARLY BUDDHIST VIEWS OF NATURE

Buddhism developed as a major world religion in the 5th century before the common era (BCE) in north India, where the historical Buddha lived and taught. At the time, the region was undergoing ecological upheaval through the growth of urban centers and political centralization. Previously uninhabited forests were cleared for agricultural expansion and town development. Wild areas that were home to rhinos, elephants, tigers, and large snakes, were seen as lled with threat but also available as marginal land. The Buddha and his followers took advantage of some of these lands but adapted their teaching and retreat times to accommodate the farming seasons and to draw on local community support.

In the canonical story of the Buddha, there are many references to nature, especially trees. The Buddha to be, Siddharta, was born the son of a king of the Shakya tribe in the foothills of the Himalayas. His birth took place under a tree, and as a young man, Siddharta spent many years wandering in the forests and mountains of India in pursuit of spiritual understanding. At the time of his enlightenment, he had been sitting at the foot of a large bodhi tree (*Ficus religiosa*) for seven days. According to the story, the Buddha vowed not to move from the tree until he gained some understanding of the source of human suffering. For many years after, he taught in shaded groves to large gatherings of monks and lay people.

After six days and nights of sitting in meditation, in a state of great concentration, he perceived all his previous lives in a continuous cycle of birth and death, and then saw the vast universe of birth and death for all beings. From this he came to understand the law of cause and effect or *karma* and the universal existence of suffering due to impermanence. This was the First Noble Truth. Then he realized the driving force behind birth and death, craving or attachment to existence: the Second Noble Truth. As the Buddha saw that all phenomena, i.e., all of nature, arise from complex sets of causes and conditions, he realized that liberation from suffering lay in this very insight: the law of mutual causality or dependent origination (in Sanskrit *pratityasamutpada*, in Pali *paticca samuppada*). This was the Third Noble Truth: the path to release from suffering. He further laid out a prescription for practice in the Fourth Noble Truth of the Eight-fold Path. He named right livelihood, right practice, and right speech, among others, as methods for achieving release from suffering.

The Buddha s profound appreciation for the universal existence of suffering engendered in him a lifetime compassionate response, expressed in the form of sharing his teachings with others. He saw that compassion (*karuna*) and loving kindness (*metta*) for all beings arises directly from this understanding of the nature of suffering. Early Buddhism was strongly in uenced by the Hindu and Jain principle of *ahimsa* or non-harming. Buddhist teachings urged monks not to harm any living thing; killing animals for food was against the monastic code. The classic Jataka Tales of India recount the many former lives of the Buddha as an animal or tree when he showed great compassion to others who were suffering. Over 500 tales have been handed down from the oral tradition, ranging from simple animal fables to fragments of heroic epics. It is said that the verses and stories were told by the Buddha himself as a way of commenting on particular life situations challenging his students. In each of the tales, the Buddha-to-be sets a strong moral example of compassion for plants and animals in his many lifetimes before his rebirth as the Buddha (Beswick, 1956).

The Agganna Sutta from the Pali Canon collection of early Buddhist texts relates the Buddhist counterpart of a creation story in which human activities clearly affect their environment. The original beings are described as self-luminous, subsisting on bliss and freely traveling through space. At that time it is said that the earth was covered with a avorful substance much like butter from which greed originated. The more butter that the beings ate, the more solid their bodies became. Difference of form appeared and the more beautiful ones developed conceit and looked down on the others. Self-growing rice appeared on the earth to replace the butter and before long, people began hoarding and then stealing food. As people erred in their ways, the richness of the earth declined. The point of the sutta is to show that environmental health is bound up with human morality (Ryan, 1998). Other suttas also spell out the environmental impacts of greed, hate, and ignorance, showing how these three poisons produce both internal and external pollution. In contrast, the practice of generosity, compassion, and wisdom can reverse such environmental decline and produce health and purity.

The early monastic code, the *Vinaya*, contained a number of guidelines for caring for the environment (De Silva, 2000). Monks were not to travel during the rainy season for fear of killing the worms and insects that come to the surface in wet weather. Similarly, monks were not to dig in the ground or drink unstrained water. Even wild animals were to be treated with kindness. Plants too were not to be injured carelessly but respected for all they give to people. Buddhists adopted a reverential attitude toward large trees, carrying on the Indian tradition regarding *vanaspati* or *lords of the forest*. Some of these huge trees were thought to be former Buddhas; protecting trees and preserving open lands were considered meritorious deeds. The Buddha constantly urged his followers to choose natural habitats to engage

in meditation, free from the influence of everyday human activity.

NORTHERN VIEWS

As Indian Buddhism developed into many strands of philosophy and practice, teachings were carried north to China. Each sect emphasized particular texts, principles, and practices, each with varying degrees of application to environmental concerns. The Hua Yen school of Buddhism, which arose in 7th century in China, placed particular emphasis on the law of interdependence or mutual causality. Ecological understanding of natural systems fits very well within the Buddhist description of interdependence (Cook, 1989). Throughout many cultural forms of Buddhism, nature is perceived as relational, each phenomenon dependent on a multitude of causes and conditions. These causes include not only physical and biological factors but also historical and cultural factors, i.e., human thought forms and values.

The Hua Yen Avatamsaka or Flower Ornament Sutra used a teaching metaphor, the Jewel Net of Indra, to communicate the infinite complexity of the multi-causal universe. This cosmic net contains a multifaceted jewel at each of its nodes, with each jewel reflecting all the others. If any of the jewels becomes cloudy (toxic or polluted), the less clearly it reflects the others. To extend the metaphor, stresses on any of the net lines, e.g., through loss of species or habitat, affect all the other lines. Likewise, if clouded jewels are cleared up (rivers cleaned, wetlands restored), life across the net is enhanced. Because the net of interdependence includes not only the actions of all beings but also their thoughts, the intention of the actor becomes a critical factor in determining what happens. This, then, provides both a principle of explanation for the way things are, and a path for positive action.

The law of interdependence suggests a powerful corollary, sometimes noted as emptiness of separate self. If all phenomena are dependent on interacting causes and conditions, nothing exists by itself, autonomous and self-supporting. This Buddhist understanding (and experience) of self directly contradicts the traditional Western sense of self as a discrete individual. Philosopher Alan Watts called this assumption of separateness the skin encapsulated ego, the very delusion that Buddhist practices seek to cut through. Interpreting the Hua Yen metaphor, modern American poet Gary Snyder suggests the empty nature of self provides a link to the wild mind, or access to the forces that determine the nature of life (Snyder, 1995). These forces act outside of human influence, setting the historical, ecological, and even cosmological context for life.

Tien-t'ai monks in 8th century China believed in a universal Buddha nature that dwelled in all forms of life. Sentient (animal) and non-sentient (plant) beings and even the earth itself were seen as capable of achieving enlightenment. This concept of Buddha nature is closely related to Chinese views of *chi* or moving energy that is ever changing, always taking on new forms. Thus their views of nature reflect a dynamic sense of flow and interconnection between all beings, with Buddha nature arising and transforming constantly.

Northern or Mahayana schools also came to emphasize the path of the *bodhisattva*, one who vows to serve others until all the world's suffering is extinguished. Where earlier Theravada schools emphasized achieving enlightenment and leaving the world of suffering, the northern schools influenced by Confucian social codes, placed great value on service to others. Tibetan schools reinforced this vow by encouraging people to treat all sentient beings as possibly having been their mother in a former life. Bodhisattva acts of service are thus personally motivated, creating a foundation for a kind of virtue ethic. This directly applies to relations with plants and animals as well as people, encouraging environmental protection and kindness as important to enlightenment.

As Zen Buddhism became established in Japan, monastic temples were often built in mountainous or forested places. The strong tradition of *haiku* and other classic verse forms cultivated a sense of oneness with nature, from insects to landforms. Dogen (1200–1253), founder of the Soto sect of Zen, spoke of mountains and waters as *sutras* themselves, the very evidence of the Dharma arising (Dogen, 1986). He taught a method of direct knowing, experiencing the dharma of nature with no separation. The goal of meditation was non-dualistic understanding, complete communion or transmission between two beings. Dogen showed how much of human suffering described in the First Noble Truth generates from egoistic views based in dualistic understanding. Enlightenment is then the breakthrough or liberation from these views to experience the self and myriad beings as one interpenetrating whole.

MODERN BUDDHIST ECOLOGICAL VIEWS

Buddhist environmental teachers and writers at the beginning of the 21st century emphasize five primary arenas of practice and philosophy which support an environmental view: interdependence, compassion, mindfulness, non-dualistic views, and detachment from self (Kaza and Kraft, 2000). At the heart of the Buddha's path is reflective inquiry into the nature of reality. Some people experience interdependence in its more ecstatic forms of communion with plants and animals or sacred places. But engaging interdependence in today's environmental context also means undertaking rigorous examination of conditioned beliefs and thought patterns regarding the natural world. This would include such challenges as objectification of plants and animals, stereotyping of environmentalists, dualistic

thinking of enemyism, the impacts of materialism, and environmental racism.

The law of interdependence is based on an understanding of the nature of the many relations at play in a situation. This could mean, for example, assessing who's who in an environmental conflict from a context of historical and geographical causes and conditions. Such investigation includes learning about ecological relationships under siege as well as observing the distribution of power across human political relationships.

The practice of *ahimsa* or *non-harming* arises from a true experience of compassion. Buddhist precepts or ethical guidelines are based fundamentally on reducing the suffering of others. The first precept, *not killing*, has been applied to environmental dilemmas related to food, land use, pesticides, pollution, and cultural economic invasion. The second precept, *not stealing*, suggests examining the implications of global trade and corporate exploitation of resources. *Not lying* brings up issues in advertising and consumerism. *Not engaging in abusive relations* covers a broad realm of cruelty and disrespect for non-human others. Vietnamese Zen teacher Thich Nhat Hanh interprets the precept prohibiting drugs and alcohol to include the toxic addictions of television, video games, and junk magazines. Non-harming extends to all beings, not merely to those who are useful or irritating to humans. It has also been applied to environmental oppression of plants, animals, rivers, rocks, and mountains as well as to human oppression based on race, class, or gender discrimination. This Green Buddhist teaching is congruent with many schools of ecophilosophy, which respect the intrinsic value and capacity for experience of each being.

Mindfulness practice, a natural support to Buddhist environmentalism, is being taught in a range of contexts. The basic teachings of the Satipatthana Sutta or the mindfulness text, cultivate awareness of breath, body, feelings, and mind. Walking and sitting meditation are used to generate a sense of centered presence and alertness. Such mindfulness generates appreciation and respect toward the natural world, with practices related to food and eating, time and place, and personal wellbeing. Those practicing mindfulness are encouraged to slow down, consider their actions carefully, and make every effort not to cause suffering to others, including plants and animals.

Most political battles play out as confrontations between apparent enemies: loggers verses spotted owl defenders, housewives verses toxic polluters, bird lovers verses pesticide producers. From a Buddhist perspective, this kind of dualistic hatred destroys spiritual equanimity. Thus it is much better to work from an inclusive perspective, offering kindness to all parties involved, even while setting firm moral boundaries against harmful actions. A Buddhist orientation to non-dualism can help to stabilize a volatile situation and establish new grounds for negotiation. Buddhist texts emphasize a strong relationship between intention, action, and karmic (actions in a previous existence) effects of an action. If an environmental campaign is undertaken out of spite, revenge, or rage, that emotional tone will carry forth into all the ripening of the fruits of that action (and likely cause a similar reaction in response). However, if an action is based in understanding that the other party is also part of Indra's Jewel Net, then things unfold with less antagonism.

Perhaps the most significant teaching of the Dharma relevant to Buddhist activism is the practice of *detachment from the ego-generating self*. Thus, a Buddhist approach to environmental activism would be non-heroic, i.e., not motivated primarily by the need for ego identity or satisfaction. Strong intention with less orientation to the self relieves the activist from focusing so strongly on results. One does what is necessary in the situation, not bound by the need to reinforce one's ideas or to have it turn out a certain way. Small b' Buddhists have been able to act as bridge builders in hostile or reactive situations by toning down the need for personal recognition. Cautioning against the self-serving ego, Buddhist teachers emphasize the power of *kalyana mitta*, or spiritual friendship: acting together in mutual support to help others practice the Dharma and take care of this world.

The Buddhist path of liberation includes the practice of physical, emotional, and mental awareness. Such practice can increase appreciation for the natural world; it can also reveal cultural assumptions about privilege, comfort, consumption, and the abuse of nature. Scholar Alan Sponberg suggests that a Buddhist environmental ethic is a *virtue* ethic, based fundamentally on development of consciousness and a sense of responsibility to act compassionately for the benefit of all forms of life. Through the practice of Green virtue ethics, modern teachers encourage students to be environmentally accountable for their actions, from eating food, to using a car, to buying new clothes. Through following the fundamental precepts, environmentally oriented Buddhists can practice moderation and restraint, simplifying needs and desires in order to reduce suffering of others.

RECENT HISTORY OF BUDDHISM AND ECOLOGY

In the last few decades, Buddhists around the world have responded creatively to environmental problems, drawing on principles of Buddhist philosophy and practice. Forest monks in Thailand lead meditation walks around polluted areas; Tibetans collaborate with western nongovernmental organizations (NGOs) to expose recent environmental destruction in Tibet. American Buddhists have worked on issues of consumerism, wilderness protection, and animal rights. One of the earliest voices for Buddhist

environmentalism was Gary Snyder, North American Zen student and poet, who illuminated connections between Buddhist training and ecological activism. In the 1950s and 1960s, members of the Beat generation and the 1960s counterculture explored these links further, saying that spiritual leadership was necessary to halt planet-wide ecological destruction.

In the 1970s the environmental movement swelled, and Buddhist centers became well established in the West. Some retreat centers confronted ecological issues head on. The Zen Mountain Monastery in New York faced off with the Department of Environmental Conservation over a beaver dam and forestry issues. The Green Gulch Zen Center in Northern California worked out water use agreements with its farming neighbors and with the Golden Gate National Recreation Area. At a time when vegetarianism was not a popular choice, most American Buddhist centers refrained from meat eating, often with awareness of the associated environmental problems. Several Buddhist centers made some effort to grow their own organic food.

By the 1980s Buddhist leaders were explicitly addressing the eco-crisis and incorporating ecological awareness in their teaching. In his 1989 Nobel Peace Prize acceptance speech, His Holiness the Dalai Lama proposed making Tibet an international ecological reserve. Vietnamese peace activist and Zen monk Thich Nhat Hanh spoke of *interbeing* using ecological examples. Zen teachers Robert Aitken in Hawaii and Daido Loori in New York examined the Buddhist precepts from an environmental perspective. The US-based Buddhist Peace Fellowship, founded in 1978, gave prominence to environmental concerns on its activist agenda.

In Thailand, village priests took the initiative to perform ritual ordination of significant trees as a symbolic gesture of solidarity with threatened forests. Other monks engaged in efforts to question economic development and its environmental impacts. Plastic bags, toxic lakes, and nuclear reactors were targeted by Buddhist leaders as detrimental influences on people's physical and spiritual health. Similarly, in Tibet and Burma, Buddhist environmentalists drew attention to oil pipelines, hunting of endangered species, and threats to unique habitats.

In the Western world, the topic of Buddhism and ecology was picked up by Buddhist publications, conferences, and retreat centers. The Buddhist Peace Fellowship produced a substantial packet and poster for Earth Day 1990. The first popular anthology of Buddhism and ecology writings, *Dharma Gaia*, was published by Parallax Press that same year, following the more scholarly collection, *Nature in Asian Traditions of Thought*. World Wide Fund for Nature brought out a series of books on five world religions, including *Buddhism and Ecology*. Buddhist magazines such as *Tricycle*, *Shambhala Sun*, *Inquiring Mind*, *Turning Wheel*, and *Mountain Record* devoted whole issues to the question of environmental practice.

In 1990 Middlebury College in Vermont hosted a conference on spirit and nature where the Dalai Lama stressed his commitment to protection of the environment. At the 1993 Parliament of the World's Religions in Chicago, when Buddhists gathered with Hindus, Muslims, pagans, Jews, Jains, and Christians from all over the world, a top agenda item was the role of religion in responding to the environmental crisis. Parallel interest in the academic community culminated in ten major conferences at Harvard University, defining a new field of religion and ecology. The first of these conferences, convened by Mary Evelyn Tucker and John Grimm in 1996, focused on Buddhism and Ecology (Tucker and Williams, 1997).

The academic community, however, did not address the *practice* of Buddhist environmentalism. This was explored by socially engaged Buddhist teachers such as the Dalai Lama, Sulak Sivaraksa, John Daido Loori, Philip Kapleau, Bernie Glassman, Christopher Titmuss and Thich Nhat Hanh. Joanna Macy developed a transformative model of experiential teaching designed to cultivate motivation, presence, and authenticity. Her methods were strongly based in Buddhist meditation techniques and the Buddhist law of codependent arising. Working with John Seed, Buddhist Australian rainforest activist, she developed a ritual Council of All Beings to engage people's attention and imagination on behalf of all beings (Seed *et al.*, 1988). Thousands of councils have now taken place in Australia, New Zealand, the US, Germany, Russia, and other parts of the Western world.

BUDDHIST ENVIRONMENTAL ACTIVISM

Examples of Green Buddhism on the front lines are still relatively rare. They reflect three major types of activism, which characterize environmentalism today: (1) holding actions of resistance; (2) structural analysis and alternatives; and (3) cultural transformation. Holding actions aim to stop or reduce destructive activity, buying time for more effective long-term strategies. In Northern California, a small group named the Ecosattvas has been protesting the logging of old-growth redwood groves. For one demonstration, they created a large prayer flag covered with human handprints to serve as a testimony of solidarity for those participating in the resistance actions. Later, several Ecosattvas made a pilgrimage into the heart of the Headwaters Forest, carrying a Tibetan treasure vase with gifts and prayers on behalf of the redwoods.

Moved by the suffering of animals in research cages, factory farms, and export trade stores, two Zen students in the San Francisco area formed a Buddhist animal rights group. Drawing on principles of non-harming, they educated Buddhists about the plight of monkeys, beef cattle,

and endangered parrots. One of them has continued this work in Europe, focusing on the cruelty of large-scale pig farming. Addressing the dangers of nuclear waste, a Berkeley-based study group protested the storage of nuclear waste below ground. As an alternative they developed a vision of nuclear guardianship for storage containers above ground that was based in Buddhist spiritual practice. In a parallel action, Zen student and artist Mayumi Oda helped to organize Plutonium Free Future and the Rainbow Serpents to stop shipments of deadly plutonium to Japan.

The second type of activism, structural analysis and the creation of alternative visions, has also engaged modern Buddhists. In 1997 the Soka Gakkai affiliated group, the Boston Research Center for the 21st Century, held a series of workshops addressing the people's Earth Charter, an internationally negotiated set of ethical guidelines for human–Earth relations. The center published a booklet of Buddhist views on the charter's principles to stimulate discussion before adoption by the United Nations (Morgante, 1997) (*see* Earth Charter, Volume 5). Members of the International Network of Engaged Buddhists and Buddhist Peace Fellowship started a Think Sangha to undertake structural analysis of global consumerism. Collaborating between the US and Southeast Asia, they held conferences in Thailand on Alternatives to Consumerism, pressing for moderation and lifestyle simplification.

The third type of activism, transforming culture, is barely underway and is sometimes met with resistance. Buddhist centers in rural Northern California, the Green Gulch Zen Center and Spirit Rock, and the Zen Mountain Monastery in New York already demonstrate a serious commitment to the environment through vegetarian food practices, land and water stewardship efforts, and ceremonies which include the natural world. In the Sierra foothills of California, Gary Snyder has been a leader in establishing the Yuba River Institute, a bioregional watershed organization working in cooperation with the Bureau of Land Management. Members have undertaken survey work, controlled burns, and creek restoration projects with the local community. Planning ahead, Snyder estimates the level of commitment necessary to reinhabit a place and build a community that might eventually span generations. The Zen Mountain Center in Southern California is beginning similar work, carrying out resource management practices such as thinning for fire breaks, restoring degraded forest, and limiting human access to preserve areas. Applying Buddhist principles in an urban setting, Zen teacher Bernard Glassman has developed environmentally oriented small businesses which employ local street people, selling products to socially responsible companies such as Ben and Jerry's.

Several Buddhist centers have developed lecture series, classes, and retreats based on environmental themes. The Zen Mountain Monastery offers Mountains and Rivers retreats based on the center's commitment to environmental conservation. These feature backpacking, canoeing, nature photography, and *haiku* (poems), as gateways to Buddhist insight. The Ring of Bone Zendo offers backpacking *sesshins* in the Sierra Mountains. The Green Gulch Zen Center co-hosts the Voice of the Watershed, a series of talks and walks across the landscape of the *two* valleys. The Manzanita Village in Southern California includes deep ecology practices, gardening, and nature observation as part of their mindfulness retreats.

Most of these examples represent social change agents working within Buddhist or non-Buddhist institutions to promote environmental interests. In addition to these organizational initiatives, individual Buddhists are taking small steps to align their actions with their Buddhist practice. Many people, Buddhists included, are turning to vegetarianism and veganism (not eating animal products) as more compassionate eating choices for animals and ecosystems. Others are committed to eating only organically grown food, as a way to support pesticide-free soil and healthy farming. Sulak Sivaraksa in Thailand insists that the Western standard of consumption is untenable if extended throughout the world. Some Buddhists have joined support groups for reducing credit card debt, giving up car dependence, and creating work cooperatives. For many students, environmental awareness and personal change flow naturally from a Buddhist practice commitment.

CONCLUSION

How might Buddhism and ecology affect the larger environmental movement and how might it influence Western Buddhism in general? Will Buddhist environmentalism turn out to be more environmental than Buddhist? The answers to these questions must be largely speculative at this time, since Green Buddhism is just gaining a footing. It is possible that this fledgling voice will be drowned out in the brownlash' against environmentalists, or in Western resistance to engaged Buddhism. Environmental disasters of survival proportions may overwhelm anyone's capacity to act effectively. The synergistic combination of millennialism and economic collapse may flatten Green Buddhism as well as many other constructive initiatives.

But if one takes a more hopeful view, it seems possible to imagine that Green Buddhism will grow and take hold in the minds and hearts of young people who are creating the future. Perhaps some day there will be Ecosattva chapters across the world affiliated with various practice centers. Perhaps Buddhist eco-activists will be sought out for their spiritual stability and compassion in the face of extremely destructive forces. Buddhist centers might become models of ecological sustainability, showing other religious institutions ways to encourage ecological culture. More Buddhist teachers may become informed about environmental issues

and raise these concerns in their teachings, calling for moderation and restraint.

That being said, Buddhism is not the only or necessarily the best path for dealing with the environmental crisis. Moral leadership and community organizing from all religious traditions are needed to stop the downward spiral of planetary ecological devastation. Committed practitioners of non-harming can inspire others who are trying to resist destructive practices. Ecologically articulate Buddhists can advance inter-religious dialogue to meet the challenges of global warming, over-consumption, and other systemic ills. Drawing on a Buddhist perspective, academics, policy analysts, and poets can bring fresh insights to once intractable problems.

What happens next lies in the hands of those who are nurturing this wave of enthusiasm for Green Buddhism. Religious leaders, teachers, and scholars as well as younger generations, full of energy and passion for protecting the home they love, will determine the shape of Buddhism and ecology in the future. As the rate of environmental destruction continues to accelerate, many forms of dialogue and activism are sorely needed. Buddhism and its environmentally supportive teachings have much to offer. Like other world religions, this ancient tradition is being called once again to offer its wisdom and teachings in yet another context.

REFERENCES

Beswick, E (1956) *Jataka Tales: Birth Stories of the Buddha*, John Murray, London.

Cook, F (1989) The Jewel Net of Indra, in *Nature in Asian Traditions of Thought*, eds B J Callicott and R T Ames, State University of New York Press, Albany, NY, 213 229.

De Silva, L (2000) Early Buddhist Attitudes toward Nature, in *Dharma Rain: Sources of Buddhist Environmentalism*, eds S Kaza and K Kraft, Shambhala Publications, Boston, MA, 91 103.

Dogen, E (1986) *Shobogenzo: Zen Essays by Dogen*, translated by T Cleary, University of Hawaii Press, Honolulu, HI.

Kaza, S and Kraft, K (2000) *Dharma Rain: Sources of Buddhist Environmentalism*, Shambhala Publications, Boston, MA.

Morgante, A (1997) *Buddhist Perspectives on the Earth Charter*, Boston Research Center for the 21st century, Cambridge, MA.

Ryan, P D (1998) *Buddhism and the Natural World: Toward a Meaningful Myth*, Windhorse Publications, Birmingham.

Seed, J, Macy, J, Fleming, P, and Naess, A (1988) *Thinking like a Mountain: Towards a Council of All Beings*, New Society Publishers, Philadelphia, PA.

Snyder, G (1990) *The Practice of the Wild*, North Point Press, San Francisco, CA.

Tucker, M E and Duncan, R W (1997) *Buddhism and Ecology: the Interconnectedness of Dharma and Deeds*, Harvard University Press, Cambridge, MA.

FURTHER READING

Batchelor, M and Brown, K (1992) *Buddhism and Ecology*, Cassell, London.

Callicott, J B and Ames, R T (1989) *Nature in Asian Traditions of Thought*, State University of New York Press, Albany, NY.

Chapple, C K (1993) *Nonviolence to Animals, Earth, and Self in Asian Traditions*, State University of New York Press, Albany, NY.

Habito, R L F (1993) *Healing Breath: Zen Spirituality for a Wounded Earth*, Orbis Books, Maryknoll, NY.

Hunt-Badiner, A (1990) *Dharma Gaia*, Parallax Press, Berkeley, CA.

Kapleau, P (1982) *To Cherish All Life: a Buddhist Case for Becoming Vegetarian*, Harper and Row, San Francisco, CA.

Macy, J (1991) *Mutual Causality in Buddhism and General Systems Theory: the Dharma of Natural Systems*, State University of New York Press, Albany, NY.

Macy, J and Molly, Y B (1998) *Coming Back to Life: Practices to Reconnect Our Lives, Our World*, New Society Publishers, Gabriola Island, British Columbia.

Nhat Hanh, T (1993) *Love in Action*, Parallax Press, Berkeley, CA.

Queen, C (2000) *Engaged Buddhism in the West*, Wisdom Publications, Cambridge, MA.

Schmidthausen, L (1997) The Early Buddhist Tradition and Ecological Ethics, *J. Buddhist Ethics*, **4**, 1 42.

Sivaraksa, S (1992) *Seeds of Peace*, Parallax Press, Berkeley, CA.

Sivaraksa, S (1999) *Socially-engaged Buddhism for the New Millennium*, Suksit Siam, Bangkok.

Snyder, G (1995) *A Place in Space*, Counterpoint Press, Washington, DC.

Business-as-usual Scenarios

A *business-as-usual scenario* is also called a *surprise-free scenario*. The expression was invented by Herman Kahn to describe a baseline scenario from which alternative scenarios can be created. It is based on current trend data assuming that no unusual events change the direction of the trends. Kahn also said that the biggest surprise would be if the surprise-free scenario actually occurred. Contrary to popular belief, the business-as-usual scenario does not imply that current conditions will remain unchanged; instead, these conditions change in the forseeable future based on current dynamics.

JEROME C GLENN *USA*

C

Capitalism

see Globalization in Historical Perspective (*Opening essay*, **Volume 5**)

Carson, Rachel Louise

(1907- 1964)

Rachel Louise Carson, scientist, author, and heroine to millions, launched the modern day environmental movement with warnings that humanity was poisoning itself and the biosphere. Her best selling book *Silent Spring* is her lasting legacy.

Rachel Carson was born in 1907 in Pennsylvania, US. From her landlocked childhood, her affinity for the oceans began with graduate work in biology at Johns Hopkins University in Baltimore, MD. Financial constraints prevented Rachel Carson's pursuit of a doctorate, and she assumed a teaching position at the University of Maryland. She continued her marine field research, teaching in the summer sessions at Johns Hopkins for 7 years. Her postgraduate research included studying at Woods Hole, MA. She became a government marine biologist and in the 1930s began her popular writing career with a series of articles in the *Baltimore Sunday Sun*. One expanded article was published in *The Atlantic Monthly*. That article ultimately was expanded into her first book, *Under the Sea Wind* (Carson, 1941). The publication 10 years later of *The Sea Around Us* (Carson, 1951) gave Rachel Carson commercial success and fame. She resigned her job as a government marine biologist to more fully dedicate herself to writing about the growing threats to the natural world she loved. Carson next published in 1955, *The Edge of the Sea* (Carson, 1955), her third book described as a poetic biography of the ocean. It was quickly followed by *The Sense of Wonder* (Carson, 1956).

The next phase of her life's work was to collect and research the evidence of the damage caused by chemical pesticides, unleashed as a new arsenal of synthetic toxic substances at the end of the Second World War. Her work on the subject was originally published in *The New Yorker* magazine and then in book form as Silent Spring in 1962 (Carson, 1962).

Silent Spring caused a storm of controversy. It became a best-seller, catalyzing environmental groups calling for the banning of dichlorodiphenyltrichloroethane (DDT) and other chemicals linked in Carson's work with widespread damage to wildlife and to human health. The chemical industry lobby was well prepared for *Silent Spring* and launched a merciless attack on Carson. Although she had expected denials and debate over her claims, she was unprepared for the highly personal nature of the charges. As her obituary in *The Baltimore Sun* noted, industry spokesmen tried to depict her as a neurotic, frustrated, hysterical personality of no scientific standing, a mere popularizer and hobbyist. Unmarried and childless, the industry argued that such a woman could not legitimately claim concern for future generations. Rachel Carson defended her well researched work with appearances before Congressional committees. What the public did not know was that even as *Silent Spring* was published, Carson was already fighting breast cancer, the disease that would ultimately take her life just 18 months after publication of her most critical work. Before her death, she fulfilled a lifelong dream, to see the California Redwoods with Sierra Club executive director, David Brower (*see* Brower, David, Volume 5). David Brower attributed her success by telling her publisher, Paul Brooks, she did her homework, she minded her English and she cared. (Brooks, *The House of Life*.)

Silent Spring remains a vitally important book 40 years after its publication. Far from stale, the book accurately anticipated many of the more subtle health risks of chemical pesticides, including endocrine disruption and xenooestrogenic effects. In Rachel Carson's own words, As crude a weapon as the cave man's club, the chemical barrage has been hurled against the fabric of life, *Silent Spring* (Carson, 1962). Despite her efforts and the movement she sparked, pesticide sales continue to soar and are now valued at tens of billions of dollars annually around the world.

REFERENCES

Carson, R L (1941) *Under the Sea Wind*, Oxford University Press, New York.
Carson, R L (1951) *The Sea Around Us*, Oxford University Press, New York.
Carson, R L (1955) *The Edge of the Sea*, Houghton Mifflin, Boston, MA.
Carson, R L (1956) *The Sense of Wonder*, Harper and Rowe, New York.
Carson, R L (1962) *Silent Spring*, Houghton Mifflin, Boston, MA.

FURTHER READING

Lear, L (1997) *Rachel Carson: Witness to Nature*, Holt, New York, 1–634 (see also Zuk, M (1997) *Science*, 298: 1897 for a book review).

ELIZABETH MAY *Canada*

CBA (Cost Benefit Analysis)

CBA is a method for evaluating public sector investments. It ideally considers all benefits (gain in utility) and costs (loss in utility) due to a project regardless of to whom they might accrue. CBA is also ideally timeless, in that all costs or benefits are considered, irrespective of when they may occur. The method simply weighs, in monetary terms, benefits versus costs. If benefits are greater than costs, then the project is potentially acceptable. There are three main challenges associated with the approach: measurement, valuation, and discounting.

The measurement of intangible costs and benefits may be difficult or inappropriate in some contexts. Aesthetic, cultural, ecological, or social utility may not be easily amenable to quantification. Beyond the problem of measuring a utility, the assignment of a monetary valuation can also be difficult. The use of shadow prices (imputed value for an item that has no clear market value) or surrogate markets (where a shadow price can be developed from a marginal rate of substitution or transformation) are sometimes used to address valuation problems. An important problem in valuation is the accuracy or representativeness of the value assigned to a benefit or cost, especially those that are inherently intrinsic.

Discounting (*see* Discounting, Volume 5) is the process of finding the value today of future benefits and costs. A rate of interest is applied to a sum (discounting is the reverse of compounding) to calculate present or future value. Future benefits and costs are discounted because of the belief in time preference (people prefer benefits today rather than in the future, thus a dollar today is worth more to an individual than a dollar receivable in one year), and because of the existence of a positive interest rate in the market. Discounting has the effect of reducing the scale of benefits and costs that accrue far into the future. A social discount rate is often used. It does not reflect a market rate, but is instead based on a social time preference rate or the social opportunity cost of capital. Discounting is criticized on the basis that it treats future generations unfairly. Depending on the choice of a discount rate, CBA may favor immediate consumption or diminish the value of environmental costs incurred far into the future.

KEVIN S HANNA *Canada*

Changing Human–Nature Relationships (HNR) in the Context of Global Environmental Change

see The Changing Human–Nature Relationships (HNR) in the Context of Global Environmental Change (*Opening essay*, **Volume 5**)

Chipko Movement

The Chipko Movement is celebrated globally by the environmental movement, heralded as a glorious example of eco-feminism, forest preservation, and local control over natural resources. The movement began spontaneously from a single incident in the Himalayan region of Uttar Pradesh, India in April 1973. Forested slopes around a village had been sold to logging concessions. Village women, fearing a loss of fresh water and destabilized slopes once the forests were gone, literally hugged the trees to prevent contractors from felling them. The word *chipko* means to embrace. Village women, Dhoom Singh Negi and Bachni Devi, among many others, spread the word with the Chipko slogan What do the forests bear? soil, water and pure air.

The movement has its roots in Gandhian theory and many prominent writers and philosophers became its leaders. The movement was embraced as a spiritual one, fitting the context of ancient Sanskrit scriptures, as well as other religions' invocation of the unity and oneness of life, by prominent philosopher and Chipko adherent, Dr Indu

Tikekar. The movement spread through many parts of the Himalayan region of Uttar Pradesh. In 1980, India's Prime Minister, the late Indira Gandhi, responded to public appeals with a 15-year moratorium on logging in the Himalayan forests of Uttar Pradesh.

The movement also included prominent men as leaders. Sunderlal Bahuguna, a Gandhian activist, whose appeal to Mrs Gandhi had proved successful, subsequently embarked on a two year, 5000 km trek by foot thorough the Himalayan regions, spreading the Chipko message beyond Uttar Pradesh. Bahuguna expressed his message as Ecology is permanent economy .

The expanded movement succeeded in stopping clear-cut logging in the Western Ghats and the Vindhyas. The Chipko movement continues to lobby for a new natural resources policy, more sensitive to local people's needs and ecosystem sustainability.

ELIZABETH MAY *Canada*

Christianity and the Environment

David G Hallman
World Council of Churches, Toronto, Canada

Western Christianity is a relative latecomer in recognizing and expressing concern about contemporary human assaults on ecosystems. The path of awakening for Christians has included not only growing awareness of the seriousness of the ecological crises but also the role that their own tradition has played in the relationship between human societies and their environments.

Some of the major tenets of western Christian thought and practice contributed to a perception of an exalted place of the human within the broader natural world which provided some of the legitimacy for scienti c, economic and technological exploitation of nature. But throughout the history of Christianity, there are also inspiring examples of individuals and schools of thought which championed a concern for the well-being of all creatures.

In recent times, Christians have become conscious of that checkered history and have become convinced that caring for creation needs to be a fundamental part of how they live out their faith in the world. In addition to the work of theologians reconceptualizing Christianity to respond to the ecological challenge, individual Christians, community parishes, and national and international ecumenical institutions have become active in addressing many of the threats to creation. However, the degree of Christian engagement in environmental concerns should not be exaggerated. It still represents a minority concern within the broader framework of theological, spiritual and social justice preoccupations.

Several communities of experience, often from the margins of the Christian mainstream, have had a particularly important impact on theological understandings of the relationship of humans to the broader natural world. In particular, eco-feminist theologians, indigenous peoples, and activists and ethicists from developing nations of the Economic South *have challenged the dominant approaches and have contributed insights that have deepened and, in some cases, redirected the focus of Christian involvement in environmental issues (see Note on Terminology, at the end of this article).*

Out of the plethora of environmental challenges facing the world, a number have become priorities for the engagement of the wider Christian ecumenical community and will likely grow as major challenges for the future. These include climate change, the exploitation of both humans and non-humans by the forces of the global economy and trade, and the spiritual values that can underpin a more sustainable relationship between human societies and the rest of the natural world.

HISTORICAL BACKGROUND

Christianity played a role in the development of the western industrialized economic model that is at the root of many of our current ecological problems. Hallman (1992) gives an analysis of the interconnections among science, religion and economics in the genesis of today's ecological crises and an identification of new models from the disciplines for living more sustainably. Through various sources, Christianity provided support for the scientific and industrial revolutions. From early in the Middle Ages, monasteries preserved and encouraged scholarship in classical Greek and Roman culture. Without this data base, modern intellectual development and scientific progress would have been immeasurably slowed. When Protestantism emerged, it emphasized the distinction between the Creator and the creature. In effect, the reformers desacralized the natural world and focused attention on humanity's standing and condition before God. This contributed to the study and manipulation of nature by the budding sciences. Francis Bacon contended that all Creation had meaning only in relation to humanity.

> Man, if we look to final causes, may be regarded as the center of the world insomuch that if man were taken away from the world, the rest would seem to be all astray, without aim or purpose.
>
> (Thomas, 1983)

Ian Barbour notes that: The desacralization of nature encouraged scientific study, though it also (along with other

economic and cultural forces) contributed to subsequent environmental destruction and the exploitation of nature (Barbour, 1990).

The industrial revolutions that followed the scientific revolution also benefited from religious attitudes of the day. Early Protestants believed in industrious, disciplined work as they shifted the religious focus from otherworldly to innerworldly asceticism. The blessing of economic success could be seen as an outer sign of inner faith and assurance of salvation. The virtues of frugality and simplicity, coupled with hard work and saving, converged with the needs of early capitalism. People worked hard and saved, but didn't spend on material indulgence: that was just the ethic needed to establish a capital base.

There was a sense of religious fervor among those involved in the early scientific and economic pursuits: God was beckoning humanity to use all of its intelligence to usher in a new day of enlightenment and prosperity. The natural world was seen as yielding its long-held secrets to the rigor of scientific experimentation. Economic theories were formulated to describe the unseen laws that governed supply and demand. There was a genuine conviction among the theologians, scientists and economists that God's will was being realized in their efforts.

Theological attitudes which viewed the natural world primarily as a God-given resource whose utility related to exploitation for human purposes continued as a dominant perspective until relatively recently. A document from the 1961 Assembly of the World Council of Churches (WCC) states that the Christian should welcome scientific discoveries as new steps in man's domination of nature (Granberg-Michaelson, 1994).

The pastoral constitution *Gaudium et Spes* of the Second Vatican Council expressed a general expectation that:

> ••• (human beings), created in God's image, received a mandate to subject (to themselves) the earth and all that it contains ••• thus, by the subjugation of all things to (humanity), the name of God would be wonderful in all the earth.

Though these broad streams of religious intellectual thought have contributed over the past several hundred years to a perception of the natural world as exploitable without consequence, there have been Christian spiritual voices offering a different perspective. Mystics such as Hildegaard of Bingen and St. Francis of Assisi articulated a reverence and valuing of all life for its own sake that contrasts with the more utilitarian approach of later Christianity. There have also been communities of religiously motivated persons over the centuries such as monastic orders, which practiced a conservation ethic in relation to nature around them. These earlier religious voices and practices concerned about the wellbeing of all creation have found new relevance during the past several decades as Western Christians have come to see the seriousness of the ecological problems facing the Earth.

THEOLOGICAL CHALLENGES

An American historian, Lynn White Jr, was one voice that did much to stimulate public awareness of the critique of Christianity in relation to environmental problems. In a famous 1967 article in the journal *Science* entitled *The Historical Roots of our Ecological Crisis*, White maintained that Judeo-Christian scriptures, theology and traditions must accept some responsibility for societal attitudes that allowed for exploitation of the environment (White, 1967).

The environmental criticisms leveled at Christianity by White and others shook theologians and lay Christians. Fortunately the response was not in the main defensiveness but rather the launching of an extended period of reflection, reanalysis of scripture and theology, and reconceptualization of models for the interrelationship of God as creator, the human species, and the rest of creation.

There are three main areas that have been identified in Christian theology and scripture as being particularly problematic. The most frequently cited is the concept of God giving humans authority over all creation:

> Let us make the human in our image, after our likeness; and let them have dominion over the fish of the seas, and over the birds of the air, and over the cattle, and over all the earth, and over every creeping thing that creeps upon the earth.
>
> (Genesis 1:26)

Theologians and biblical scholars have helped to place this seeming blanket mandate for domination into a context which highlights a range of conditionalities tied to this authority. In *Genesis*, the story of creation precedes the arrival of evil or the fall. Thus, the giving by God of this authority over other creatures occurred at a point where humans were without sin and hence would be expected to exercise that dominion in ways pleasing to God who created and loves the world. Secondly, there is another creation story in *Genesis* 2, which uses very different imagery for the human relationship to the rest of nature. The human is to till and keep the garden. This *dominion theology* perspective challenges an unrestricted exploitive authority that could be read into the first creation story.

Canadian theologian, Douglas John Hall, was an early exponent of *stewardship theology*, which sought to transform Christian approaches away from the unfettered dominion model toward one of responsible caring for the Earth. Hall (1982) argues that stewardship means that humans are given responsibility to care for the world as God's stewards. That is, humans do not own the Earth but are God's surrogate caretakers: The Earth is the Lord's and the fulness therof; the world and all that dwell therein (*Psalm* 21:1). To care for the Earth as God would care for it points to a much more responsible attitude since we believe that God loves the Earth. Further, as stewards, humans will be held responsible by the Master, for the care which they have exercised, or failed to exercise, on that which belongs to the Master.

A second theme in Christian theology which has been seen as posing problems for our relationship to the environment is the dichotomy which is identi ed between the spiritual and the physical. In the New Testament Epistles, we hear a number of references that would seem to counsel rejection of this physical world including the following:

> Do not love the world or the things of this world. If anyone loves the world, love for the Father is not in him.
>
> (*I John* 2:15)

> Do you not know that friendship with this world is enmity with God? Therefore, whoever wished to be a friend of the world, makes himself an enemy of God.
>
> (*James* 4:4)

However, biblical scholarship that gained added relevance as environmental awareness grew, has helped to place these letters to communities of early Christians into historical context. At a time when the general expectation was that the end of the world was at hand and that Christ would be returning soon, the writers were encouraging their readers not to place their faith in material wealth. While there has been a persistent thread over the centuries stressing the religious af rmation of creation s goodness, there have also been interpretations of these passages strongly emphasizing otherworldliness and rejection of this world. Environmentalists have worried that such theology could undermine a social attitude of concern for the wellbeing of the Earth.

A third concept that has been identi ed in the critique of Christian theology is the emphasis on personal moral will as the ultimate de ning characteristic of the human. While there are philosophical traditions that have also focused on moral agency, Christian theology has given it a particular priority. Though an emphasis on moral will does not of necessity lead to a negation of the value of the natural world, some of its implications seem to point in that direction. Since only humans exercise moral choice and judgement via rational processes, it was believed that higher value was attributed to the human than to the creatures lacking these capacities. The elevation of the rational capacities also implied the denigration of other attributes such as emotions, intuition, and sense perception itself.

NEW THEOLOGICAL DIRECTIONS

The critiques of Christianity by Lynn White Jr and others helped to unleash a robust period of theological reconceptualization regarding the relationship of humans to the rest of creation. In addition to these challenges from outside the religious community, there were also those within the faith who were shaking traditional approaches. The creation-centered work of Thomas Berry and Matthew Fox from within the Roman Catholic tradition provided a radical reorientation to understandings of the interrelationship of God, humanity and the rest of creation. Their writings would prove quite unsettling to the church hierarchy. *Stewardship theology* referred to above was another of the initial responses and did much to af rm that biblically sound and intellectually rigorous alternate understandings to *dominion theology* were possible.

Process theology has been a further approach that has widely in uenced the evolving Christian thinking about the ecological crisis. Based on the work of philosopher Alfred North Whitehead, process theologians such as Cobb and Grif n (1977) developed a systematic application of Whitehead s ideas with the emphasis on God s loving interrelatedness with creation. Cobb in particular has played a signi cant role through his writings in articulating the ecological implications of process theology (Cobb, 1972; Cobb and Birch, 1981; Cobb and Daly, 1989). One of his principal arguments has been the inherent interconnection between God, humanity and the natural world. ••• for process theology, as an ecological theology, human beings are part of nature ••• Humanity is seen within an interconnected nature (Cobb, 1972) .

German theologian Jurgen Moltmann comes to similar conclusions within the framework of his own theological analysis. In a book based on his Gifford Lecture series, Moltmann emphasizes the relationship of God to all creation as the starting point.

> An ecological doctrine of creation implies a new kind of thinking about God. The center of this thinking is no longer the distinction between God and the world. The center is the recognition of the presence of God in the world and the presence of the world in God.
>
> (Moltmann, 1985)

Through biblical analysis and systematic theological argumentation, Moltmann describes various manifestations and implications of the presence of the Spirit of God within creation. A consistent theme is the centrality of relationship as the most fundamental de ning characteristic of existence. Theological understandings such as Moltmann s and those of process theology echo and in some ways enhance scienti c discoveries and reconceptualizations of the 20th century which have also moved in the direction of an emphasis on the primacy of relationships within matter and the natural world.

Eco-feminist theologians have articulated perceptive analytic critiques of traditional Christian theology in terms of its destructive contribution to the ecological crisis and have posited new models for understanding the relationship between humanity and the rest of creation. While recognizing and recovering much rich insight in the Biblical scriptures that can help ground a new environmental ethic, eco-feminist theologians are nonetheless more prepared than some other theologians to acknowledge the ambiguity of the Bible. This stems in part from the broader feminist analysis that recognizes the cultural conditioning

of many of the scriptural writers working as they were in very patriarchal societies.

Eco-feminist approaches to reflection on ecological theology and ethics draw on a broad range of human capacities. Their methodology and the content of their analysis are interconnected in the recognition that our ecological problems are due in part to a hierarchical valuing of creatures and the attributes associated with them. Men have been seen as the most valued and the traditional male-oriented emphasis on rationality, control and exploitation has dominated both in human relationships and in terms of the Earth. Eco-feminists insist that intuition and emotion are indispensable as elements in the repertoire of human response if we are to make significant progress in solving our environmental problems. They argue that a reorientation toward the Earth will require a foundation of love and caring for life in its fullness not just rationalistic analyses and technological fixes (*see* **Ecofeminism**, Volume 5).

There are creative options proposed by eco-feminist theologians for new theological approaches to our relationship with creation. The *web of life* is one image that eco-feminists use to emphasize the interrelatedness and interdependence of the various elements of creation.

Sallie McFague has suggested reimaging God not only as father but also as mother (the nurturer of all life), as lover (the most intimate of relationships), and as friend (a constant and supportive presence) (McFague, 1993). Such models of God provide a deeper appreciation of the role of God as creator and sustainer of life and can contribute to values that would be more environmentally sensitive: e.g., as protection of life, joy and love in our relating to the Earth, and solidarity with the oppressed members of creation.

Chung Hyun Kyung (1994) has done much to popularize eastern religious and cultural appreciations for the unity of all life and the sacredness of creation. Bringing such insights into dialogue with more traditional Christian approaches has provoked considerable controversy. It is nonetheless an essential contribution if Western Christianity is to honestly acknowledge its cultural role in the economic development model which has come to dominate the global economy and to seek partnership with other faith traditions in building a value base for an Earth ethic.

Equally threatening to some Christians have been the insights being offered by indigenous peoples around the world. Struggling back after several centuries of cultural oppression by colonizing powers, many indigenous peoples are recapturing their traditions that include important understandings of ways in which to relate sustainably to the natural world around them. Western societies seeking to reconceptualise their relationship to the Earth could benefit from the rich contributions of indigenous creation stories, the understanding of nature as being members of one's family, rituals used to relate to the various seasons and creatures in the natural world, and concepts of community organization for long-term sustainability (Rajotte, 1998). Particularly helpful have been the writings of people of indigenous heritage who have connected native spirituality and Christian theology such as the American Indian George Tinker, Stan MacKay (a Canadian of Cree origin) and Rob Cooper (a Maori from Aotearoa, New Zealand) (Tinker *et al.*, 1994).

There has been a risk in the rejuvenation of native spiritual ecological insights that Christians and others might appropriate those concepts and rituals into their own practices in a romanticized manner without recognizing the historical context out of which those indigenous understandings emerged. In part to respond to this risk, indigenous peoples have raised tough political and economic issues related to the ongoing oppression of their cultures and societies in addition to sharing elements of their native spirituality. Christian communities have been challenged to acknowledge and repent of their role over the past centuries as agents of cultural genocide. Some churches and ecumenical organizations have acted in solidarity with indigenous peoples in their struggles in local, national and international forums.

Theologians, ethicists and activists involved in environment and development issues in countries of the economic South have been another source of critical social, economic and political critique of the western development model and its links to Christian traditional understandings of the relationship of humans to the rest of creation.

Theological contributions from the economic south on issues of ecotheology have been relatively recent. Theologians and ethicists from these countries have understandably been more preoccupied with the economic struggles of their countries which have suffered 500 years of exploitation by imperialistic nations, colonializing powers and now global financial institutions. Countries in Latin America produced a vibrant movement of liberation theology that played a pivotal role in articulating the ethical basis for a systemic social, economic and political critique of the western economic model and its historical exploitation of the south. However, many of the liberation theologians were as anthropocentric and utilitarian in their orientation toward the non-human elements of the natural world as were the tradition Christian theologians of the North.

However, as the depth of the ecological crises became more apparent as well as their interlinkages with the western economic development model, theologians and ethicists from countries of the *Economic South* began to make important connections between the historical struggles of their countries and environmental problems. Moreover, justice for the poor had always been the primary focus of liberation theologians and it was becoming clear that the poor were also the most vulnerable to the health and social effects of environmental problems.

Leonardo Boff from Brazil is among the more prolific and insightful theologians who has sought to develop an integrated approach to socio-economic issues and environmental concerns. While recognizing the contributions of various critical responses to the prevalent social model such as liberation movements, pacifist and non-violence groups, and ecological movements, Boff argues for the need for a more integrated approach.

> It is important today to articulate these different critiques of the dominant system. However, we must urgently seek to develop a new paradigm for society that does not repeat the mistakes of the old but integrates all human beings in a more humane way and establishes more benevolent relationships with the environment.
>
> (Boff, 1994, 1997)

The insights and analyses of theologians, ethicists and activists from countries of the *Economic South* have had a significant impact on the work on environment and development issues by churches and ecumenical organizations such as the WCC (*see* WCC (World Council of Churches), Volume 5). Theologian Jesse Mugambi from Kenya has enriched ecumenical understandings of the social, economic and environmental impacts of colonialization in Africa and has articulated nuanced analyses of so called globalization and the homogenization processes where a few power centers dictate the rules for the rest of the world. Aruna Gnanadason from India who is working currently as a staff member of the WCC has provided leadership in the integration of gender justice with ecology and economic development.

It is impossible to note all of the influential theologians and ethicists who are contributing today to deepening Christian reflection on the implications of the ecological crisis and what the faith has to offer as constructive responses. Mary Evelyn Tucker, John Grim, Dieter Hessel and Rosemary Ruether, all of whom have done writing in their own right, have made a very important contribution through the organizing of the April 1998 Conference on Christianity and Ecology on behalf of the Harvard University Center for the Study of World Religions. The resulting book includes contemporary reflections by many of the most important theologians and ethicists concerned today about the ecological crisis.

Larry Rasmussen and Dieter Hessel have produced a book that looks at implications for the church of engagement in ecological issues after bringing together a wide range of theologians, ethicists, and practitioners for a 1998 Conference on Ecumenical Earth, at Union Seminary in New York. Rasmussen (1996) has also written an ambitious integrative book *Earth Community, Earth Ethics* which combines analyses of the interrelated crises of environmental degradation and global economic injustice with the articulation of new visions that integrate biblical reflection, theological rigor, and inspiring spirituality.

I would also add some of my own writing to this list of attempts to take the richness of the theological and ethical debates and make it more popularly available to a wider Christian audience, for instance, *A Place in Creation - Ecological Visions in Science, Religion and Economics* (Hallman, 1992) and *Ecotheology - Voices from South and North* (Hallman, 1994).

CHURCH ENVIRONMENTAL RESPONSES AT LOCAL, NATIONAL AND INTERNATIONAL LEVELS

To write about Christianity and the environment is not just to discuss the theological and ethical reflections that have propelled the faith into serious confrontation with the ecological crisis. That rejuvenated faith, having had its eyes opened to the seriousness of the problems, its own historical complicity, and most importantly, new models of engaging theology and spirituality in relevant responses, is being expressed in action.

Church denominations in many parts of the world have prepared educational and action resources to assist their members learn more about environmental problems within an ethical and spiritual context and to offer opportunities for activities to make a constructive difference:

- US churches circulated information packets and lobbying suggestions to congregations at Earth Day to encourage them to press for effective endangered species legislation.
- German churches did a study of energy use in churches as a prelude to a campaign to become more energy efficient.
- Philippine churches ran educational programs on forestry and collaborated with other social groups in a campaign for a moratorium on wide scale logging.
- Tanzanian environmental workshops were held for church and community leaders and included work on specific projects such as dam building to preserve badly needed water for irrigation.
- Dutch churches encouraged members to reduce their carbon dioxide emissions from transportation 3% per year.
- Zimbabwean churches have integrated tree planting into their liturgies.
- Japanese Christians collaborated with Buddhists, Shintos and members of New Religions to host an interfaith gathering and march during the 1997 session of the UN negotiations on climate change in Kyoto.
- Argentinean Catholic, Orthodox, Protestant, and Episcopal churches worked together to host an ecumenical service during the 1998 session of the UN negotiations on climate change in Buenos Aires.
- Swedish churches bought parcels of forestland to ensure that they are harvested in a sustainable manner.

- Canadian churches advocated against plans for the premature and potentially dangerous disposal of high level nuclear wastes.
- The Ecumenical Patriarch of Constantinople has sponsored environmental education programs for youth and a special seminar focused on environmental problems around the Black Sea.

The list could include many more examples of efforts by churches at national and local levels to translate their faith into practice in caring for the well being of creation. Persons involved in organizing such efforts will be the first to acknowledge that, though progress is being made, it remains an ongoing struggle to convince the churches as institutions and the individual members of the importance of engagement in environmental issues.

At the international level, several ecumenical organizations, especially the WCC, have played a role in both fostering theological and ethical reflection on ecological concerns as well as facilitating the active engagement of their member churches in specific issues. An important consultation was organized by the WCC in 1974 in Bucharest and brought together scientists, economists and theologians to discuss implications of the recently published study of the Club of Rome *Limits to Growth*. One of the important contributions of this event was the articulation of the concept of sustainability, the idea that the world's future requires a vision of development that can be sustained for the long-term, both economically and environmentally. During the 1970s, the WCC had a program on the just, participatory and sustainable society (JPSS), which included the Energy for My Neighbor project which was intended to sensitize churches about energy problems faced by developing countries and to activate practical steps to ameliorate the energy situation of those in need, and the Massachusetts Institute of Technology (MIT) Conference in 1979 on Faith, Science and Technology, which became known primarily for its controversial position on nuclear power.

In 1983, the WCC Assembly in Vancouver adopted a process-focused action plan on justice, peace and integrity of creation (JPIC) through which churches were encouraged to work together on these interrelated themes. Many churches became increasingly attentive to environmental concerns during this period, adopting policy statements and initiating education and advocacy activities on specific issues. The JPIC process culminated in a World Convocation on Justice, Peace and Integrity of Creation in Seoul, Korea in 1990, at which a series of ten theological affirmations and specific covenants for action were approved. They provide a description of the interrelatedness of economic inequity, militarism, ecological destruction, and racial injustice and the theological, ethical and spiritual basis for affirming and sustaining life in its fullness.

There have also been a number of occasions of interaction between Christian theologians and leaders of other living faiths, focused on rediscovering the important contributions from within the traditions and sacred writings of each of the faith systems, which could help move human societies toward greater respect for the natural world. One of these events was an interfaith consultation hosted by the WCC in August 1991 to develop proposals for inclusion in an Earth Charter.

The 1992 United Nations Conference on Environment and Development in Rio de Janeiro provided an opportunity for witnessing to the spiritual dimensions of the ecological crisis. Many faiths were represented at the Rio Earth Summit and held joint vigils, ceremonies and workshops. The WCC sponsored a major ecumenical gathering bringing to Rio 150 representatives of churches from over 100 countries for two weeks of prayer, worship, study and involvement in the Earth Summit. Particularly important were the connections made with many other non-governmental organizations representing environmental groups, development bodies, and women's networks.

During the 1990s, the WCC work on environment-related issues focused primarily on global climate change, monitoring the work of the UN Commission on Sustainable Development, and beginning a more significant engagement in ethical issues raised by the growth in the biotechnology industry.

CHALLENGES FOR THE FUTURE

Climate change has become one of the major foci for international ecumenical activity on ecological issues and is likely to continue to be so as the threat increases. Not only does climate change represent a threat to the well-being of God's Earth but it is also a profoundly ethical issue since it is being precipitated largely by the rich industrialized countries while the consequences will be suffered disproportionately by the poorer developing countries and by future generations.

The ecumenical community through the WCC has participated in the UN negotiations on climate change treaties since 1989. In 1996–1997, the WCC sponsored an international petition campaign to build greater public pressure on the governments of industrialized countries to take action to reduce their emissions as a lead up to the Kyoto Summit in December 1997. The WCC made a major statement on climate change as an issue of justice at the Kyoto Summit (World Council of Churches, 1997). Churches in many countries have organized educational and advocacy activities and have sponsored ethical reflections on climate change within the context of models for sustainable societies. These programs have been accompanied with resources to assist individual members take practical steps in their own homes, lives and communities to reduce energy use and contribute to limiting the emission greenhouse gases.

A second major challenge which will likely increase in importance for the Christian community in the future is the interconnection of the environment and the global economic system. These are well illustrated in the climate change issue but they also go beyond and encompass global trade, the role of multinational corporations and international financial institutions, and the scandalous economic inequities between peoples of the world. The ecumenical community has a long history in advocacy and action for economic justice and a more recent history with environmental issues. These two become interconnected when dealing with the implications of increasing economic globalization. Addressing globalization will pose a challenge to the churches in terms of their capacity for analysis of complex systems, resources to be engaged from local to global levels and the strength to persevere despite intense pressure that can be mounted by the vested power structures being critiqued.

A third challenge for the future is the reconceptualizing of the churches' important environmental activity in a spirituality that can both inspire the work and sustain those involved. Throughout western culture generally, there is a widely perceived thirst for spiritual nourishment to counter the bareness of contemporary consumerism and materialism. Institutional religion is aware that many people find its forms of worship and engagement do not adequately satisfy their spiritual needs. Thus, both Christians active in social and environmental issues within churches and non-Christians outside the church, feel the need for greater spiritual nourishment. Concern for the well being of the Earth is one area where a nurturing of spiritual values can be linked intrinsically with a justice agenda.

NOTE ON TERMINOLOGY

Terminology is very complex when trying to describe different regions of the world today. In United Nations (UN) parlance, reference is made to developed and developing countries to distinguish between the richer primarily northern and western industrialized nations and the poorer primarily southern nations. However, many people object to that vocabulary because it implies a higher level of technological and social sophistication for the developed countries. Third World was another term that was current for some time but that has also fallen out of favor because again it suggests an ordering of value or worth in comparison with the First World. A distinction between the North and the South has become more utilized because most of the poorer nations are south of the equator and south does not carry any particular value connotation. Not all of the poorer countries are in the geographic south and there are some industrialized nations in that region. Hence, for my purposes in this paper and in my other recent writings, I have come to use the phrase countries of the *Economic South* to denote the poorer nations. While I recognize the clumsiness of the phrase and the increasing difficulties in making generalizations between the North and the South in our increasingly complex world, this is the designation that I find most satisfactory to date.

REFERENCES

Barbour, I (1990) *Religion in an Age of Science*, Harper & Row, San Francisco, CA.

Boff, L (1994) Social Ecology: Poverty and Misery, in *Ecotheology - Voices from South and North*, ed D G Hallman, WCC Books, Geneva, and Orbis Press, New York.

Boff, L (1997) *Cry of the Poor, Cry of the Land*, Maryknoll, Orbis Press, New York, 114.

Cobb, Jr, J (1972) *Is it too Late? A Theology of Ecology*, Bruce Publications, Beverley Hills, CA.

Cobb, Jr, J and Birch, C (1981) *Liberation of Life*, Cambridge University Press, New York.

Cobb, Jr, J and Daly, H (1989) *For the Common Good: Redirecting Economy Toward Community, the Environment, and a Sustainable Future*, Beacon Press, Boston, MA.

Cobb, Jr, J and Griffin, D (1977) *Process Theology: An Introductory Exposition*, Christian Journals, Belfast.

Granberg-Michaelson, W (1994) Creation in Ecumenical Theology, in *Ecotheology - Voices from South and North*, ed D G Hallman, WCC Books, Geneva, and Orbis Press, New York.

Hall, D J (1982) *The Steward: A Biblical Symbol Come of Age*, Friendship Press, New York.

Hall, D J (1986) *Imaging God: Dominion as Stewardship*, Erdmans/Friendship Press, Grand Rapids, MI.

Hallman, D G (1992) *A Place in Creation - Ecological Visions in Science, Religion and Economics*, United Church Publishing House, Toronto.

Hallman, D G (1994) *Ecotheology - Voices from South and North*, WCC Books, Geneva, and Orbis Press, New York.

Kyung, C H (1994) Ecology, Feminism and African and Asian Spirituality, in *Ecotheology - Voices from South and North*, ed D G Hallman, WCC Books, Geneva, and Orbis Press, New York.

McFague, S (1987) *Models of God: Theology for an Ecological, Nuclear Age*, Fotress Press, Philadelphia, PA.

McFague, S (1993) *The Body of God: an Ecological Theology*, Fortress Press, Minneapolis, MN.

Moltmann, J (1985) *God in Creation: an Ecological Doctrine of Creation*, SCM Press, London.

Rajotte, F, ed (1998) *First Nations, Faith and Ecology*, United Church Publishing House, Toronto and Cassell, London.

Rasmussen, L (1996) *Earth Community, Earth Ethics*, Orbis Books, Maryknoll, NY.

Thomas, K (1983) *Man and the Natural World: Changing Attitudes in England 1500- 1800*, Allen Lane, London.

Tinker, G, MacKay, S, and Cooper, R (1994) *Ecotheology - Voices from South and North*, ed D G Hallman, WCC Books, Geneva, and Orbis Press, New York.

White, Jr, L (1967) The Historical Roots of our Ecological Crisis, *Science*, 155, 1203–1207.

World Council of Churches (WCC) Statement to the High Level Segment of the Third Session of the Conference of the Parties (COP3) to the UN Framework Convention on Climate Change (1997) Kyoto, Japan, available from the author (c/o The United Church of Canada, 3250 Bloor St. W. Toronto, M8X 2Y4, fax 1-416-232-6005 e-mail: dhallman@uccan.org).

Circulating Freshwater: Crucial Link between Climate, Land, Ecosystems, and Humanity

Malin Falkenmark
Natural Science Research Council and Stockholm International Water Institute (SIWI), Stockholm, Sweden

A simplified holistic view is presented of humanity s present environmental predicament as seen from a freshwater perspective. Like the doctor who bases much of his diagnosis of a patient on his blood system, its circulation and quality characteristics, the author bases her diagnosis of the world ecosystem of which Man is a part on the bloodstream of the biosphere, i.e., the water cycle. The aim is to identify the most crucial actions in future-oriented decision-making. In contrast to a widespread but false belief that water is very simple and can be handled within the water sector, many parallel functions and manifold appearances have to be taken into account when addressing the interaction between human society and the physical landscape hosting its life-support system. Fundamental regional differences in terms of both vapor ow of green water involved in biomass production, and liquid ow of blue water available for societal use in aquifers and rivers are highlighted. It is shown that the developing world at present experiences a threefold escalation: of water stress, of water dispute proneness, and of water pollution load. Analysis suggests that water accessibility constraints will increasingly disturb global food security. Escalating food needs make it particularly important to find ways of reconciling upstream/downstream con icts of interest especially as intensified upstream biomass production (agriculture, forestry) easily leads to downstream river depletion. Such depletion will generate increasingly severe problems for water-dependent societal activities and aquatic ecological services downstream.

TIME TO LOOK AT THE BLOOD SYSTEM OF THE BIOSPHERE

One thing characterizing our time is the predicament in which environmental decision makers nd themselves: squeezed between two shields. On the one hand, the public expects them to repair the environmental damage that has already materialized; on the other, the public expects them to be forward-looking and ensure that growing demands on socio-economic development will not make the problems even worse.

A necessary basis for future-oriented decision-making is a minimum of basic understanding of fundamental relationships between humans and the natural landscape hosting their life-support system (freshwater, biomass production, presence of minerals and energy sources, etc.). Past studies on environmental effects of human activities have not produced such an understanding but often got caught up in the considerable complexity of ecosystem behavior. Moreover, one fundamental component the water cycle acting as the bloodstream of the biosphere tends to be absent in many such studies (Falkenmark, 1997). Equally remarkable is the lasting land/water dichotomy and the considerable conceptual barriers to realizing the water dependence of land use and biomass production. In the water community, the alertness to water scarcity problems (CFWA, 1997) tends to be concentrated to allocation problems within the so-called water sector (water supply for household, industry, and irrigation). This attention is, however, quite recent. Most countries have probably not started to deal with it but pursue supply strategies, as in the past. Not much attention is being paid to upstream/downstream con icts of interest.

It is getting increasingly clear that humanity s genuinely water-dependent future cannot be discussed without attention to water s deep involvement in many different functions in both nature and society. For example, potential water-related constraints to global food production are starting to send serious concerns through the scienti c community (Falkenmark, 1997b; Falkenmark *et al.*, 1998 and references therein). The nite character of the freshwater circulating on the planet has made it essential that the water requirements for food production be properly entered into global water resources assessments.

This situation has made it necessary to expand our attention from the conventional water sector focused on societal use of the liquid ow of blue water in aquifers and rivers to include also the vapor ow of green water involved in rainfed biomass production in agriculture, forestry and natural vegetation systems (Figure 1) (Falkenmark and Rockstrom, 1993).

Evidently the plants do not mind if the water available to the roots is naturally in ltrated rainwater or external irrigation water that has been added from groundwater or surface water. Since the land cover is instrumental in the rainwater partitioning process at the ground surface, land

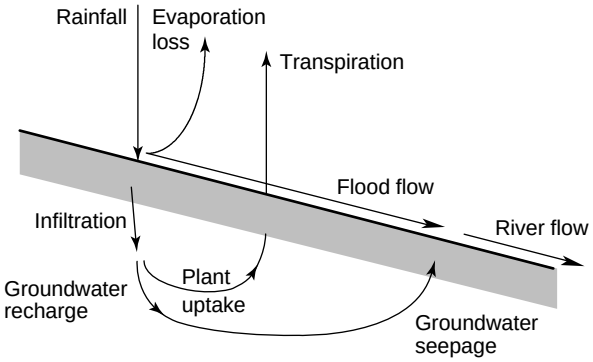


Figure 1 The rainfall partitioning into the vertical green water branch, encompassing productive and non-productive components, and the semi-horizontal blue water branch, encompassing the water in groundwater aquifers and rivers

use changes tend to alter the partitioning of the incoming rainfall between these two water flows, and therefore influence the blue water that is available for allocation within the water sector.

The water cycle sets a large-scale exterior constraint to both human and natural life. The aim of this article is to tie together a whole set of water-related phenomena within a joint conceptual framework. This framework starts from the basic interactions between human society and the life-supporting landscape in which we live. It demonstrates the enormous management problems that population growth generates due to water's deep involvement in these interactions, and identifies a set of urgent actions.

TECHNOSPHERE AND BIOSPHERE SHARE THE SAME BLOOD SYSTEM

It is probably not much of an exaggeration to state that water is the most misunderstood component of life. It is mistakenly taken as quite simple, since it is part of everybody's daily life. In reality it is extremely complex with many parallel functions in both nature and society: health function, habitat function, psychological and religious functions, carrier functions (solutes, silt), and production functions (green-water dependent biomass production, blue-water dependent socio-economic production). Successful environmental management therefore critically depends on our ability to address such complexity.

Water fills all living beings, it has vital functions on all scales – from the planetary scale down to the cell scale. It is a universal solvent, carrying nutrients to the cells and waste products away from the cells. In the life-support system, it acts as a carrier of dissolved matter, as a key food substance for both humans and fauna, as a key raw material in plant production, and as a habitat for aquatic flora and fauna. In society it functions almost as a lubricant of industrial development through all its key functions in industrial production (carrier, solvent, washer, cooler). At the same time, water is a key operator in different collapse-driving processes: as carrier of microbiological disease agents to humans, in its tendency to form floods and inundations, as eroding agent, and as contaminant carrier in general.

Since all ecosystems are water-dependent, water is fundamental as a silent messenger in generating ecosystem degradation and biodiversity disturbances from human manipulations of the landscape.

To facilitate communication on these complex issues, the following simple metaphor may be helpful (Falkenmark, 1997a). Human activities in the landscape can be illustrated as involving interaction between a social sphere and a landscape sphere. The imperative in the social sphere is to satisfy the basic needs of the population of water, food, energy, etc. and meet their aspirations in terms of goods and services. Performing this involves manipulations of land and water in the landscape for harvesting the natural resources there (water, biomass, energy, minerals). These manipulations are physical (clearing, tilling, drainage, well drilling, pipes and canals) and chemical (agricultural chemicals, waste produced in the social sphere and disposed of in the landscape: to the air, to the land, or directly to the water bodies). Since the natural laws operate in the landscape, side effects are produced, building up into air pollution, land degradation, water degradation and as higher order effects, i.e., ecological degradation. If the needs and aspirations of the inhabitants are not satisfied or as a response to frustration raised by the exacerbating environmental side effects, reactive responses occur in the social sphere: passive ones like famine, poverty, diseases, disputes, suffering, or active ones like riots, environmental migration, policy changes, etc.

Water's multifunctional involvement in the interaction between human society and the physical landscape is reflected in its manifold appearances in terms of:

- the provision of water supply for various societal activities;
- water as a resource manifested as availability in rivers and aquifers, water involved in the production of plant substance from carbon dioxide from the air and water from the soil; streaming water as a source of energy; and water bodies as habitats for aquatic biota;
- side effects of human landscape manipulations manifested as water quality degradation; altered seasonality in the rivers; escalating floods, river depletion; and aquatic ecosystem degradation;
- water-related diseases, societal damage by inundations, etc.

The sectorial organization of society implies that also the political driving forces on demands on the water that

flows through the landscape are sectored and unconnected although they are all dealing with the same water (Falkenmark, 1997a):

- the health authorities are interested in water supply and sanitation to protect against water-related diseases, high morbidity and mortality;
- the agricultural authorities are responsible for crop production to keep famines away, generating raised water requirements, often also land degradation;
- the environmental authority is responsible for habitat protection to avoid ecosystem degradation and problems with protein nutrition;
- the economic development authority is responsible for industrial production, thereby generating increasing water requirements, often also pollution loads.

The tunnel view tendency in each of these sectorial bodies introduces an incoherence in decision-making that explains much of the past difficulties in coping with emerging environmental problems.

A CLOSER LOOK INTO THE UNDERLYING WATER PROCESSES

Landscape manipulations, that are necessary to harvest the natural resources, generate side effects through:

1. the water partitioning of incoming precipitation in its interaction with soil and vegetation;
2. water's function as a universal solvent continually moving down the landscape in the river basin;
3. the integrity of the water cycle circulation, transferring impacts of human manipulations along the chain from atmosphere to land and terrestrial ecosystems to groundwater and rivers to water bodies/lakes and aquatic ecosystems to coastal waters and marine ecosystems (Figure 2).

The first of these links can be further clarified by referring to the distinction earlier indicated between two complementary types of water flows (cf. Figure 1):

- the invisible vapor flow back to the atmosphere – referred to as green water flow which is part of the plant production process;
- the liquid flow passing above and below the land surface – referred to as blue water flow – and consisting of the surplus or non-evaporated precipitation.

Food, fodder, fiber, fuelwood and timber are all produced by green water, and thereby involve a consumptive use (Lvovich and White, 1990) which is of the order of $600-6000\,m^3$ of water per ton biomass produced, depending on the hydroclimate. Human societies (municipalities, industries, irrigation schemes, etc.) are run by blue water withdrawn from rivers and groundwater. Such uses tend to add a pollutant load to the water when returned to the river.

In its contact with the land surface, incoming precipitation passes two partitioning points (Falkenmark, 1986):

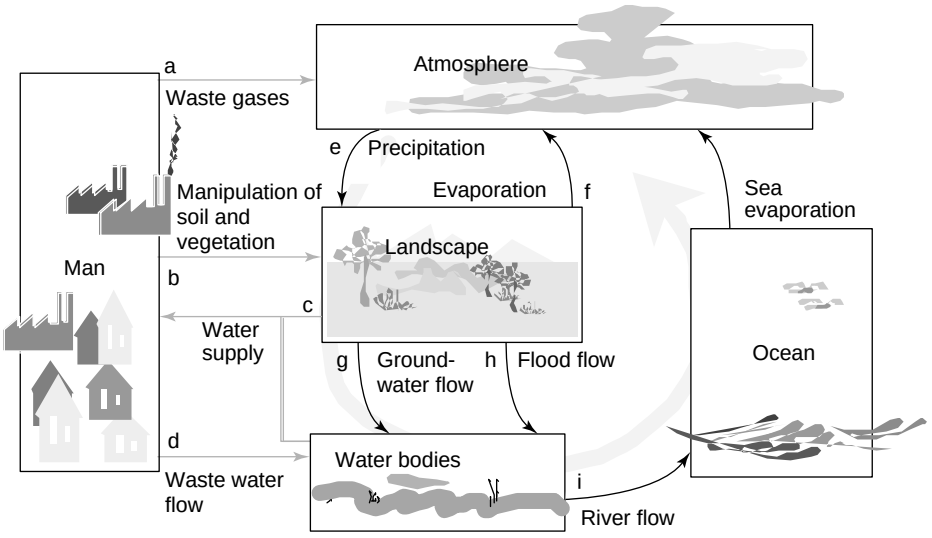


Figure 2 Mental continuity-based image of the water cycle, linking the net sea evaporation, the wetting of continental landscapes, the recharge of aquifers, the flood flows in rivers, and the river discharge back to the sea. Man's interactions are indicated by the four horizontal arrows to the left: (1) waste gas output to the atmosphere; (2) manipulation of soil/vegetation – physically and chemically; (3) water withdrawals from aquifers and rivers; (4) wastewater output to water bodies. When the population grows, these arrows will grow in size, whereas the water-flow arrows in the system will change only when climate changes

- at the soil surface, dividing the precipitation between flood flow and infiltration;
- in the root zone, dividing the infiltrated water between plant uptake and groundwater recharge.

Since land use activities modify soil and vegetation, they also influence this partitioning and therefore both the amount of river flow and its seasonality (fluctuations between low flow and high flow). Land conversions that involve major vegetation changes tend to impact this water partitioning, and result in green/blue water redirections, traditionally spoken of as water balance changes.

Deforestation typically reduces the green water return flow to the atmosphere and therefore increases the flow in the blue water branch (a typical result is water logging (*see* **Waterlogging**, Volume 3), as happened on the Hungarian Plain, increased river ow, or ood ows). In South Africa, a major campaign is now employing over 42 000 people to eradicate invasive alien plants, that are more water consumptive than the natural vegetation. The campaign has double objectives: on the one hand to increase the blue water ow that is available to support society s water needs, on the other to restore degraded biodiversity.

Reforestation works the opposite way, easily producing river depletion. In South Africa afforestation and reforestation after more than 5 years is allowed only with issued permits (van der Sel, 1997). Due to the importance of this land/water linkage, in the South African government water issues are handled by a joint Ministry of Water Affairs and Forestry. The frequent recommendation of forest plantations to protect vulnerable soils from crusting and erosion easily neglects the lower partitioning point in Figure 1 while concentrating all its interest on the upper one. This neglect may lead to surprising river depletions developing as unexpected side effects of the plantations.

The second link is related to the combined effect of water both as a solvent and as a substance involved in key chemical interactions between the in ltrating water and the geological surroundings taking place in the root zone (Falkenmark and Allard, 1991). It is widely known that the water-divides formed by the hilltops in the landscape de ne the water ow modules, i.e., the drainage basins, each enclosed by its own water divide. Therefore, the precipitation that does not return to the atmosphere as green water, forms blue water and either feeds the river with rapid ood ow or passes underground as groundwater ow, seeping back to the surface in springs, local hollows and valley bottoms.

A slope in the landscape is typically divided into uphill recharge areas where the rainwater moves downwards through the root zone, recharging the groundwater and downhill discharge areas where groundwater returns back to the land surface. In the recharge areas, bogs may develop, fed by precipitation. In the local hollows, groundwater-dependent wetlands develop; and in the valley bottoms the groundwater seeps back to the surface and either evaporates, feeds riparian forests or recharges the water courses with a time-stable ow. In the riparian zone, groundwater and ood dependent riparian wetlands may develop.

The chemical composition of the natural water develops as the in ltrating water passes from recharge to discharge areas (Falkenmark and Allard, 1991). It rst interacts with two systems: the organic root zone system which adds carbon dioxide and humates, making the water aggressive, and the soil system from where available minerals are picked up. As the water arrives down into the groundwater zone it has acquired two cardinal characteristics: alkalinity/acidity and redox potential (*see* **Redox Potential**, Volume 2). These two in combination determine what the water picks up along its continued underground pathways through deep soil and bedrock. The resulting chemical composition in the seepage zone, where the rising water meets oxygen again, characterizes the biodiversity there.

The third link relates to the continental part of the water cycle as such. The global water cycle links the ocean, the atmosphere, the land and its terrestrial ecosystems, the water bodies and their aquatic ecosystems, and the marine water and coastal and marine ecosystems. The water cycle is the mega-desalinization plant of the planet on which life here depends. It is energized by the energy from the sun. On the continents water is moving above and below the ground surface under the law of gravitation.

The third link is illustrated in Figure 2 where man s interventions with the cycling water has been distributed among the different compartments as indicated by the four arrows to the left:

- waste gas output to the atmosphere;
- manipulation of the soil/vegetation physically and chemically;
- water withdrawals from aquifers and water bodies;
- waste water output to water bodies.

When population grows these arrows grow in size, whereas the circulating water ow will change only when climate changes.

FUNDAMENTAL REGIONAL DIFFERENCES IN CHALLENGES

Now that this basic mental image of the interaction between society and the physical landscape hosting the life-support system has been explained and discussed, the next step is to explore fundamental regional differences in the principal system: on the one hand those related to the hydroclimate in uencing the green water ow, on the other the differences in the blue water ows.

It is thought-provoking to note that most industrialized countries are located in the temperate zone. There, water blindness is widespread and people tend to take water for

granted. Unfortunately, it seems that water blindness is widespread also, however, in the South. The poorest countries are in the dry climate tropics and subtropics where the everyday question number one is where to find the water and food needed to survive. The fact that many colonial powers in the tropical and subtropical countries based their understanding on natural conditions in temperate climates invites the question whether their decisions were environmentally deleterious and introduced rather than eliminated environmental problems.

Turning next to fundamental blue water differences (renewable water only), Figure 3 gives a global level overview of the water resources situation in the world today. The present level of water withdrawal (use to availability ratio or technical scarcity) is shown against the population pressure on water (number of individuals sharing each flow unit of water or demographic water scarcity) (Falkenmark and Lundqvist, 1998). In this diagram, different levels of *per capita* water use (in cubic meters per person and year) show up as a diagonal line.

Moving up the vertical scale is costly and means more reservoirs, canals and pipelines but also more administration. Experience suggests that when withdrawals exceed 20% of the availability, investments for reservoirs, transfers, irrigation schemes, etc. start to be costly as seen in the national economy. Moreover, in dry climates it is difficult to mobilize more than 50% due to evaporation losses from reservoirs. One hundred percent is only possible to reach by large scale waste water reuse, by underground storage, or where a large river passes right through a country so that the water is easy to access. Moving along the horizontal scale, on the other hand, means more water competition and disputes, and increased pollution load since there are more people polluting each flow unit of water.

In this diagram, the world regions form five rather distinct clusters with radically different water resources predicaments:

- truly water-scarce regions with high population pressures on water resources, technically close to the hydrological ceiling – altogether 1.6 billion inhabitants in 1995 and rapid population expansion;
- the water spenders, i.e., regions with a moderate to high water mobilization level, moderate population pressure, and very high *per capita* water use due to large scale, low-efficiency irrigation – altogether a rather stable population of 0.3 billion;
- an average-level set of regions, moderate in all three senses – mobilization levels, population pressure and *per capita* use – altogether 2.5 billion and some of them with ongoing population expansion;
- a group of water rich regions with low mobilization level, low population pressure, and moderate *per capita* use – altogether a rather stable population of 0.8 billion;
- an outlier group with low mobilization level and *per capita* use – a rapidly growing population – currently 0.4 billion people.

The diagram suggests that population growth is pushing large parts of the world towards the right, i.e., a situation

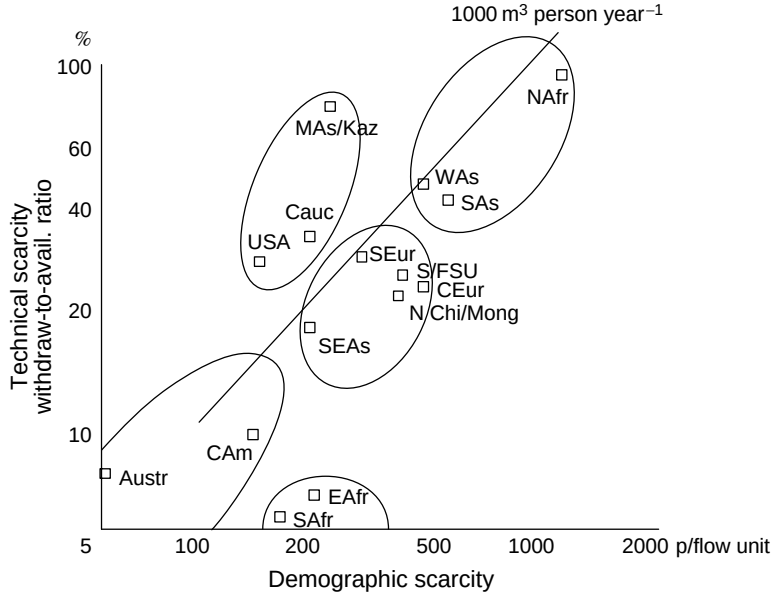


Figure 3 Fundamental regional differences in the water scarcity predicament. The figure shows characteristics in terms of demographic and technical water scarcity, respectively, for Shiklomanov's 26 global regions. The diagonal line shows *per capita* water withdrawal of 1000 m^3 person^{-1} year^{-1}, needed for food self-sufficiency in semiarid tropics and subtropics. (Data from Shiklomanov, 1996)

of growing dispute proneness and pollution loading. Ultimately, a growing proportion of available water will no longer be usable due to too much pollution (Lundqvist, 1998). This is evidently a real danger to socio-economic development in the developing world since it may lead to societal stagnation wherever industries employing large groups of the population are closed down by court orders, due to excessive water pollution, or where pollution-driven frustration generates riots and social unrest. There are several cases, for example in India where this process is starting to develop.

Figure 3 can also be used to see the implications of population growth for food production (Falkenmark, 1997b). Most critical is the extra blue water needed to secure food supply in dry climate regions, where yields are low due to water shortages and plant damage from dry spells. The situation was recently explained in a Round Table on water scarcity and global food insecurity in *Ambio* (Falkenmark *et al.*, 1998). It was assumed that, in semi-arid climates (where most of the rapidly growing populations are), maybe 50% of crop water requirements may be provided from infiltrated precipitation, while irrigation has to contribute the remaining 50%. This adds up to a blue water need of $800\,m^3\,year^{-1}$. Adding another $200\,m^3\,year^{-1}$ for households and industry leaves us with a water need of $1000\,m^3\,year^{-1}$ altogether. Many regions are moving to under the food self-sufficiency line, and will increasingly become dependent on food imports.

This analysis suggests that water will increasingly limit food self-sufficiency possibilities in dry climate regions with rapid population growth. Countries that will not be able to stay self-sufficient, due to lack of accessible water, should study their comparative advantages in order to find out what to export in exchange for food imports.

FROM HYDROEGOISM TO HYDROSOLIDARITY

What are then the key challenges of the near future when it comes to finding out what some of the crucial actions would be? The overriding task is to successfully cope with environmental preconditions while satisfying societal needs. Evidently, water's flow through the river basin links human activities upstream with opportunities and problems downstream. Therefore, downstream societies are in a sense the prisoners of the upstreamers (Figure 4).

Human interactions with the water can be divided in two categories: water-dependent activities and water-impacting ones. Man's water-impacting activities are basically of three kinds: land use conversions (influencing blue/green water partitioning, as already indicated in the above discussion of deforestation and reforestation); water withdrawals/flow management (influencing river flow and low/high flow seasonality); and the addition of pollution loads.

It has in particular to be realized that water use may involve consumptive water use (Lvovich & White, 1990); most of the water withdrawn for irrigation is evaporated into the atmosphere. The result is downstream river depletion (e.g., Aral Sea basin, Yellow River in China). After non-consumptive use, on the other hand, where water is used in a through-flow based manner, the water used goes as a return flow either to groundwater or to the river, generally carrying a pollution load of agricultural chemicals, industrial and/or municipal pollutants.

Downstream water users and ecosystems will suffer from these changes in water quantity and quality. Protection of downstream water interests for wetland protection and conservation of biodiversity will demand conservation of a certain minimum flow. Some experts say that at least 25 percent of the river flow has to remain for downstream ecosystems, others that 75 percent has to remain. Such criteria are evidently equivalent to posing constraints on acceptable upstream consumptive use and pollution load respectively, i.e., upstream food production. The conflicts of interest here, in other words, involve existentional issues for humanity that will now have to be honestly faced.

It is extremely important that these upstream/downstream conflicts of interest be adequately realized. True negotiating

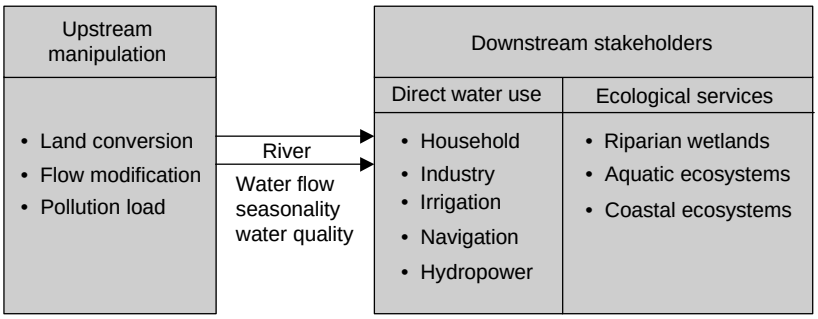


Figure 4 Conflicting stakeholder interests in terms of water-impacting activities and water-dependent activities, respectively, for which reconciliation methods will have to be developed

arenas will have to be developed so that the unavoidable side effects of manipulations, required for the satisfaction of societal needs, can be analyzed and discussed in an intentional way rather than arriving as unintentional surprises as has often been the case in the past (Rockstrom *et al.*, 1999).

THREE CRUCIAL TASKS FOR THE NEXT 30 YEARS

It is clear from the above that the present predicament, particularly of the developing world, includes three phenomena:

- escalating water competition;
- escalating dispute proneness;
- escalating pollution load.

On the regional scale, the upstream/downstream dependencies within river basins involve major challenges related to evident con icts of interest, whether the river basins are national or transnational, shared amongst several countries. Fortunately, there is probably also a proneness to create institutions that facilitate cooperation.

A set of fundamental regional differences have to be kept in mind (all regions suffer from recurrent oods):

- In western countries the key problem is water pollution (basically chemical) reducing water usability, threatening human health and aquatic ecosystems downstream.
- In newly industrialized countries (post-communist countries and monsoon-climate countries in the humid tropics) the key problem may be water shortage during the dry season (lack of dilution water) and/or a combination of microbiological and chemical pollution.
- In the lowest income countries the key problem is dif culties in making water accessible for use (small rivers go dry most of the year, large rivers are international), water competition is escalating, and dispute proneness and pollution loads are rapidly growing as water demands for socio-economic development increase.

This overview suggests that three key challenges have to be urgently addressed:

1. Supplying food in water-short countries, unable to feed their rapidly growing populations. This may be foreseen to involve expectations of a six-fold export from temperate zone countries (the breadbasket of the world) to help feed these water-short countries unless continued forest clearance for rainfed agriculture be accepted as a vital weapon in a global strategy to feed humanity.
2. Ensuring that water remains usable for all the societal sectors depending on it. This points to the critical need to go from today s willful neglect of water pollution to a situation where both industry and agriculture take water pollution seriously, and try to minimize their contributions (Lundqvist, 1998).
3. Developing mechanisms for con ict reconciliation between upstream land and water use, and downstream water use and ecosystem services (Falkenmark, 1998). This points to the crucial need to develop a water ethic, which has to be up to date and based on an understanding of how the life-support systems function.

If society has to manipulate land and water and those manipulations inevitably produce side effects, then the creation of a sustainable development policy which does not undermine its own resource base must be an issue of balancing the manipulations needed against the side effects produced (Falkenmark and Suprapto, 1993). Criteria are needed for this balancing: both human-ecological (groundwater must remain drinkable, land must remain productive, crops and sh must remain edible); ecological (biodiversity and crucial ecological services must be preserved) and socio-economic (poverty and hunger must be eradicated, and income generated).

HUMAN RIGHTS AND HUMAN OBLIGATIONS: BROADENING THE PERSPECTIVE

As already indicated the general belief that the water issue is very simple and can be handled within a water sector is altogether false. In reality it is as complex as the ow of blood in the human body. A doctor would never make his diagnosis of a sick patient without considerable attention to his blood system.

The past neglect in environmental sciences of water s many different roles in the interaction between development and environment has seriously delayed a proper understanding of so-called environmental problems and led to much lip service and redundancy in statements at international meetings. The price paid in terms of human suffering cannot be quanti ed but it can be noted that since the Stockholm Conference on the Human Environment, the world population has expanded by some 2500 million, mainly in the developing countries where the environmental problems have only gotten larger.

Much stress is presently being put on human rights to water; an example of our careless language. What is tacitly being referred to is not water as such but the provision of safe household water. The fundamental importance for humanity s future of nding ways for peaceful sharing of the precipitation falling over a joint river basin between

those living upstream and those living downstream, however, suggests that there is an even larger need for human water solidarity. In the future, human water obligations have to be given equal weight to the right to safe household water. Given a situation where upstream and downstream countries tend to have tremendous problems in agreeing on issues on the sharing of transboundary water systems, this indicates the need for seeking support from religious and philosophical circles in the search for a water ethics.

See also: **GEWEX (Global Energy and Water Cycle Experiment)**, Volume 1; **Hydrologic Cycle**, Volume 1; **Hydrology**, Volume 1; **Limnology**, Volume 1; **Soil Moisture**, Volume 1; **Buffering Capacity**, Volume 2; **Eutrophication**, Volume 2; **Hydrology**, Volume 2; **PET (Potential Evapotranspiration)**, Volume 2; **Salinization**, Volume 2; **Fisheries: Pollution and Habitat Degradation in Tropical Asian Rivers**, Volume 3; **Groundwater Withdrawal and the Development of the Great Man-made River Project, Libya**, Volume 3; **Indus Basin: a Case Study in Water Management**, Volume 3; **Inter-basin Transfer (IBT) for Water Supplies**, Volume 3; **Irrigation: Environmental Impacts**, Volume 3; **Irrigation: Induced Demise of Wetlands**, Volume 3; **Marshes, Anthropogenic Changes**, Volume 3; **River Regulation**, Volume 3; **Salinity and Agriculture**, Volume 3; **Water Resources: Baltic**, Volume 3; **Water Resources: Great Lakes Case Study**, Volume 3; **Water Use: Future Trends, and Environmental and Social Impacts**, Volume 3; **Waterlogging**, Volume 3; **Yellow River, China: a Case Study in Water Resources**, Volume 3; **Aral Sea**, Volume 4; **Great Lakes Region of North America**, Volume 4; **Lake Victoria**, Volume 4; **Nile River**, Volume 4.

ACKNOWLEDGMENTS

This paper is a revised version of my Volvo Environment Prize lecture, published in *Ambio*, Vol. 28, No. 4, June 1999.

REFERENCES

CFWA (1997) *Comprehensive Assessment of the Freshwater Resources of the World*, UN and SEI, World Meteorological Organization, Geneva.

Falkenmark, M (1986) Freshwater – Time for a Modified Approach, *Ambio*, **15**(4), 192–200.

Falkenmark, M (1997) Society's Interaction with the Water Cycle: a Conceptual Framework for a More Holistic Approach, *Hydrol. Sci.*, **42**(4), 451–466.

Falkenmark, M (1997) Meeting Water Requirements of an Expanding World Population, *Philos. Trans. R. Soc. Biol. Sci.*, **352**, 929–936.

Falkenmark, M (1998) Dilemma when Entering 21st Century – Rapid Change but Lack of Sense of Urgency, *Water Policy*, **1**, 421–436.

Falkenmark, M and Allard, B (1991) Water Quality Genesis and Disturbances of Natural Freshwaters, *The Handbook of Environmental Chemistry*, Vol. 5, Part A, ed O Hutzinger, Springer-Verlag, 45–78.

Falkenmark, M, Klohn, W, Lundqvist, J, Postel, S, Rockstrom, J, Seckler, D, Shuval, H, and Wallace, J (1998) Water Scarcity as a Key Factor Behind Global Food Insecurity – Round Table Discussion, *Ambio*, **27**(2), 148–154.

Falkenmark, M and Lundqvist, J (1998) Towards Water Security: Political Determination and Human Adaptation Crucial, *Nat. Resour. Forum*, **21**(1), 37–51.

Falkenmark, M and Rockstrom, J (1993) Curbing Rural Exodus from Tropical Drylands, *Ambio*, **22**(7), 427–437.

Falkenmark, M and Suprapto, R (1993) Population-landscape Interactions in Development. A Water Perspective to Environmental Sustainability, *Ambio*, **21**(1), 31–36.

Lundqvist, J (1998) Avert Looming Hydrocide, *Ambio*, **27**(6), 428–433.

Lvovich, M M and White, G F (1990) Use and Transformation of Terrestrial Water Systems, in *The Earth as Transformed by Human Action*, eds B L Turner, II, W C Clark, R W Kates, J Richards, J T Mathews, and W Meyer, Cambridge University Press, Cambridge, 235–252.

Rockstrom, J, Gordon, L, Falkenmark, M, Folke, C, and Engvall, M (1999) Linkages among water vapor flows, food production and terrestrial ecosystem, *Conserv. Ecol.*, **3**(2), 5.

Shiklomanov, I A (1997) *Assessment of Water Resources and Water Availability of the World*, Background Report to the Comprehensive Assessment of the Freshwater Resources of the World, Stockholm Environment Institute and World Meteorological Organization.

van der Sel, D W (1997) Sustainable Industrial Afforestation in South Africa under Water and Other Environmental Pressures, in *Sustainability of Water Resources under Increasing Uncertainty*, Vol. 240, IAHS Publication, 211–216.

Civic Science

see Precautionary Principle (**Volume 4**)

Commons, Tragedy of the

Tragedy of the commons is used to refer to a situation in which people who share in the use and benefits of a resource are inclined to damage or over-use it. Examples include

public lands, fisheries, waterways and public byways, the airwaves, and the atmosphere. Those who use this phrase typically assume that the word commons means that access is open to all; from that premise the argument is that the rational user of a common resource has no incentives for self-restraint (Hardin, 1968). There is no way to assure that other users will not take advantage and that the one who tries to conserve will be able to reap the benefits in the future. This idea is often used as an explanation for local, regional, and global environmental degradation and as a rationale for either government intervention in or privatization of resources used in common. However, it is important to recognize that in many cases access to resources held in common is not open access; and that they are managed through rules agreed upon by the users, as well as user-generated systems for monitoring and enforcement (McCay and Acheson, 1987; Ostrom, 1990). This is true for many local-level communities using forests, grasslands, fisheries, and other common resources, and it is also true for many international environmental regimes. Consequently, a second meaning of tragedy of the commons is the situation in which government intervention, privatization, and other factors diminish the effectiveness of local and user-based institutions for common resource management, creating environmental problems. Related are situations in which resettlement schemes, violence, or privatization programs result in the displacement of people dependent on commonly held access rights, intensifying their poverty and, in many cases, forcing them to intensify their use of marginal and vulnerable natural systems or to become part of urban-industrial systems.

See also: Property Rights and Regimes, Volume 5.

REFERENCES

Hardin, G (1968) The Tragedy of the Commons, *Science*, 162, 1243–48.

McCay, B J and Acheson, J M, eds (1987) *The Question of the Commons*, University of Arizona Press, Tucson, AZ.

Ostrom, E (1990) *Governing the Commons*, Cambridge University Press, Cambridge.

BONNIE J MCCAY *USA*

Cost Benefit Analysis (CBA)

see CBA (Cost Benefit Analysis) (**Volume 5**)

Cousteau, Jacques

(1910- 1997)

Jacques-Yves Cousteau was a French oceanographic explorer and filmmaker who was best known for his prime time television series in the 1960s and 1970s, his role in the development of the first self-contained underwater breathing apparatus (SCUBA) equipment, and his advocacy for conservation of the marine environment.

Cousteau had no formal education in science or engineering. However, later in his career he was awarded honorary degrees (DSc) from California, Harvard, and Ghent. He entered the French Naval Academy in 1930, and eventually reached the rank of captain. His initial intention was to be a navy pilot but he seriously injured one of his arms in a car accident, ending his flying career. He started swimming to strengthen his arm and developed his love for swimming in the sea. Cousteau subsequently experimented with goggles, snorkels and other diving equipment. Cousteau and Emile Gagnan, an engineer from Paris, developed the Aqua-Lung (i.e., the first SCUBA gear) in 1943. This revolutionized diving since it freed divers from heavy diving suits requiring cumbersome air hoses and lifelines. The invention allowed free swimming and facilitated greater exploration of the ocean.

During World War II, Captain Cousteau conducted espionage for the French Resistance. He received decorations (e.g., Croix de Guerre) and was made Knight of the Legion of Honour for his efforts. After the war and with money from a British philanthropist (Thomas Loel Guinness), he acquired a decommissioned American-constructed minesweeper that had served in the British navy. He used this ship, the *Calypso*, for ocean exploration until it sank in Singapore Harbor in 1996. Cousteau and Jean Mollard, an engineer, developed a miniature submarine known as the Diving Saucer (1959) that allowed deep sea observation. He also developed other underwater equipment, conducted studies on underwater living and helium–oxygen diving techniques, and improved color photography by perfecting the underwater camera.

Cousteau greatly raised public awareness of ocean issues as a prodigious author and filmmaker. His book, *The Silent World*, published in 1953, was enormously

successful. It sold five million copies and was translated into 22 languages. His subsequent extensive number of highly successful books included *The Living Sea*, *The Whale - Mighty Monarch of the Sea*, and *Jacques Cousteau's Calypso*. His 1956 film, *The Silent World*, won the Academy Award for best documentary. He also won Academy Awards for *The Golden Fish* (1959) and *World Without Sun* (1964). Cousteau's series of television documentaries, *The Undersea World of Jacques Cousteau* (1968–1976), popularized marine issues and educated the public. By the end of his career he had written or contributed to more than 50 books and at least 150 films.

Jacques Cousteau earned numerous honors during his 70 year career. He was the director of the Musse Oceanographique of Monaco for 31 years (despite lacking a science degree), a member of the United States National Academy of Sciences and a member of the Academie Francais. Other honors include the Special Gold Medal of the National Geographic Society (1961), placement on the United Nations Environment Programme's Global 500 Roll of Honour for Environmental Achievement (1988), and the James Smithson Bicentennial Medal from the Smithsonian Institute (1996). Despite these honors, he has stated that he felt his success in increasing public awareness and concern for the ocean was his greatest achievement. The efforts of Jacques Cousteau for advocacy and promotion of ocean conservation are continued by the Cousteau Society Inc., which he founded in 1973.

Photo: from Columbia and The Silent World.

FURTHER READING

Earle, S (1997) Cousteau Remembered, *Popul. Sci.*, 251(4), 80.

Glasgow, E (1997) Jacques Cousteau and the Discovery of the Sea, *Contemp. Rev.*, 271(1580), 135–138.

Madsen, A (1986) *Cousteau: An Unauthorized Biography*, Beaufort, New York, 1–270.

Marden, L (1998) Master of the Deep, *Natl. Geogr.*, 193(2), 70–79.

Schick, E A, ed (1997) Obituary of Jacques-Yves Cousteau from *New York Times* A p. 1 (June 26, 1997). *Current Biography Yearbook*, H W Wilson, New York, 1–699.

The Cousteau Society, Accessed May 3 (2000) Jacques-Yves Cousteau: Sailor and Explorer, http://www.cousteausociety.org/aboutjycc.html.

GLYNN GOMES *Canada*

D

Deep Ecology

Deep Ecology is the most fundamental form of ecocentrism. Its holistic view considers that humans are not separate from nature and that the earth, Gaia, is one organism, whose every (interdependent) part has intrinsic value. Current abuse of nature ultimately stems from inappropriate belief systems, most particularly Christianity, and the classical scientific worldview developed in the West during the Enlightenment. These are held to encourage anthropocentrism and attitudes of separation from and superiority to nature. Such attitudes are termed *shallow*, whereas Deep Ecology (the distinction was made in 1972 by the philosopher Arne Naess) is considered so because it fundamentally challenges such Enlightenment Project assumptions. *See* Enlightenment Project, Volume 5.

Deep Ecology celebrates richness, diversity and equality among all Earth's creatures, holding that humans have no right to reduce this richness and diversity except to satisfy vital needs. The lifestyles and practices stemming from such a principle emphasize simplicity, frugality, self-reliance, steady-state economics and limiting human population size. Invoking Aldo Leopold's land ethic, Deep Ecology regards human intervention in nature as inherently destructive. It seeks instead to preserve and expand wilderness areas.

Deep Ecology criticizes classical science's contemporary monopolization of respectable approaches to knowing the world and nature, seeing the predominance of rational thought as endemic to the anthropocentrism that has produced ecological crisis. By contrast, deep ecologists advance intuition as an equal or even superior form of cognition. Through intuition, they argue, the continuity between the human self and the rest of the cosmos may be properly apprehended and appreciated. Hence, drawing on diverse influences such as Heidegger, Buddhism, native American insights and Western romanticism, Deep Ecology calls upon people to develop a quasi-mystical ecological consciousness by which they will feel themselves part of the natural world, as a self-in-self.

Social Ecology and Eco-socialism roundly criticize Deep Ecology for displaying reactionary and anti-humanist tendencies.

DAVID PEPPER *UK*

Defense Fund, Environmental

see Environmental Defense Fund (**Volume 5**)

Demographic transition

Wolfgang Lutz and Anne Goujon
International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria

The global demographic transition was the salient feature of world population changes throughout the 20th century. It is the reason underlying the population explosion that brought world population size from 2.5 billion in 1950 to six billion in 2000. But it is also the explanation why birth rates have been declining around the world and are expected to further decline where they are still high. As a consequence, population growth has already leveled off in many developed parts of the world and is expected to level off in all parts over the course of the 21st century. A decline in birth rates does not immediately result in the stabilization of population size, due to the fact that a very young population - a consequence of past high fertility - results in an increasing number of potential parents over several decades. This phenomenon is called the momentum of population growth.

In 1945, Frank Notestein developed the paradigm that all populations undergo a process of demographic transition at some point in their history. The *demographic transition theory* has evolved as a generalization of the sequence of events observed in what are now developed countries. The model distinguishes four phases in the evolution of populations. In the initial phase of the transition, the combination of high birth and death rates produces a slow or no growth equilibrium. That happened throughout most of human history: The world population grew slowly because birth and

death rates uctuations cancelled each other (see Figure 1). In the second phase, death rates begin to fall because of exogenous factors such as rise in living standards (better nutrition, hygiene) and improvements in health. This phase is a period of rapid population growth. The demographic transition theory asserts that in the third phase, and after some delay, the decreases in mortality will be followed by a decrease of fertility rates. During this phase, the population growth rate will decrease. The latter phase is when there is a quasi equilibrium between low birth and death rates.

Different societies have experienced transition in different ways, and today various regions of the world are following distinctive paths (Tabah, 1989). Many developed countries are in the last, fourth stage, or even beyond that, as many European countries have reached below replacement fertility levels and seem to stay at those levels. All world countries have gone through the second stage of the transition (mortality decline) and most are in the third stage of fertility decline or at a post-transition stage.

The demographic transition began in more developed countries (MDCs) in the 18th century and spread to less developed countries (LDCs) in the last half of the 20th century (Notestein, 1945; Davis, 1954, 1991; Coale, 1973). In MDCs, mortality rates declined comparatively gradually beginning in the late 1700s and then more rapidly in the late 1800s; fertility rates declined, as well, after a lag of 75 100 years. The demographic transition started much later in LDCs. The rst sign of a decline in mortality rates only appeared during the period between the two world wars. Average annual population growth rates doubled to reach 1% year^{-1}. Nevertheless, it was only in the 1950s and 1960s that the diffusion of medical techniques and progress in public health allowed the global population growth rate to increase dramatically. It reached 2.4% during the 1960 1970 period for the developing world.

The broad result has been a gradual transition from a small, slowly growing population with high mortality and high fertility, to a large, slowly growing population with low mortality and low fertility. On the theoretical level there are two different ways to explain demographic transition. In one view, the fertility decline is a direct response to the mortality decline. This so-called homeostasis argument claims that societies tend to nd an equilibrium between births and deaths. When death rates decline due to progress in medicine and better living conditions, the equilibrium is disturbed and the population grows unless birth rates decline in response to the new mortality conditions. The fact that fertility tends to decline many years after a decline in mortality may be explained by a perception lag. The equilibrium is supposed to be attained by a country when it reaches replacement-level fertility; that is an equilibrium in which each generation exactly replaces itself. The replacement fertility level depends on the sex ratio at birth, as well as on the mortality between birth and the end of reproductive life. In countries that are through the stage of mortality decline (all MDCs and some LDCs), the level of replacement fertility is about 2.1 children.

The alternative to the homeostasis argument assumes that modernization of society acts as a driving force of both declining mortality and declining fertility. Fertility decline lags behind mortality decline, according to this view, because fertility is embedded in the system of cultural norms more strongly than mortality, and therefore changes more slowly. The historical record of Europe where fertility sometimes declined simultaneously with mortality and population growth was generally much lower than in today s high fertility countries tends to support the second explanation. But the two arguments are not necessarily mutually exclusive.

Figure 2 illustrates the example of demographic transition in two countries: Austria, a developed country in

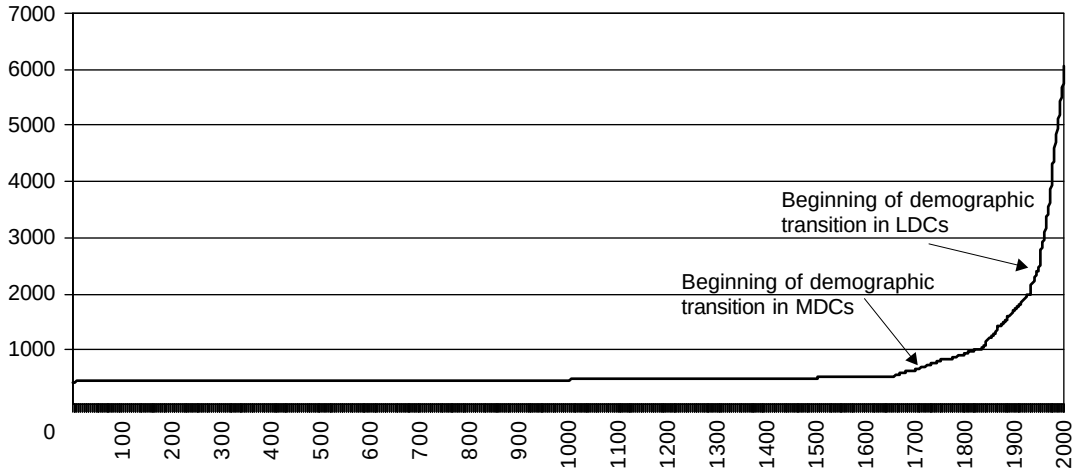


Figure 1 World population, 1 AD to 2000 AD

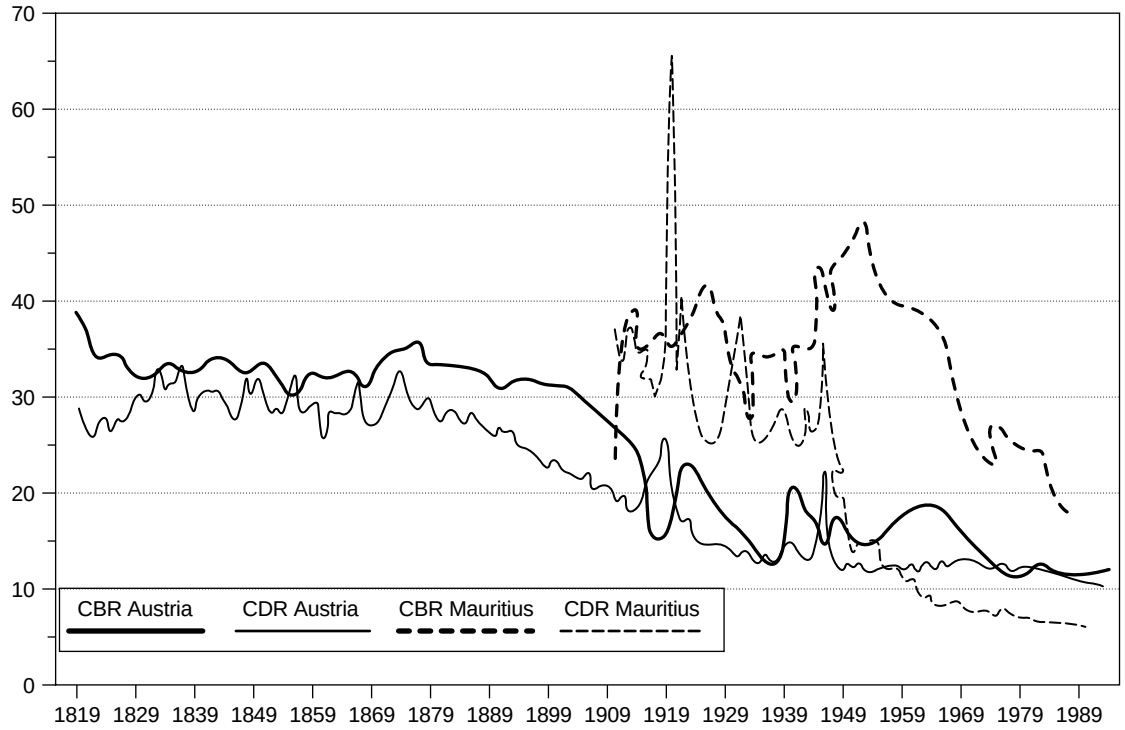


Figure 2 The demographic transition in Austria and Mauritius, crude birth (CBR) and death (CDR) rates

Central Europe and Mauritius, a developing country in the Indian Ocean. Both nations have good records for birth and death rates going back more than a century. Until about 1880 in Austria, annual birth and death rates fluctuated widely. These fluctuations can also be observed in Mauritius until 1945, due to epidemics, the changing severity of endemic diseases (notably malaria), and changing weather conditions (cyclones). Whenever birth rates are consistently above death rates, as was the case in Mauritius during the late 19th century, the population grows.

After 1880 in Austria, death rates started to decline until they had reached approximately their present level after World War II. The decrease in mortality rates started 70 years later in Mauritius but the decline was much more abrupt than in Austria. Death rates in Mauritius declined precipitously due to malaria eradication and the introduction of modern medical technology. Although the speed of the mortality decline and original levels of mortality were different in the two countries, the two developments were of the same nature. However, in the Austrian case, birth rates started a slow decline soon after the death rates, which is somehow typical of European countries, thus supporting the validity of the modernization argument mentioned above.

On the contrary, in Mauritius, birth rates remained high or even increased somewhat due to the better health status of women (a typical phenomenon in the early phase of demographic transition). By 1950 this had resulted in a population growth rate of more than 3% year^{-1}, one of the highest in the world at that time. Birth rates subsequently declined, with the bulk of the transition occurring during the late 1960s and early 1970s when the total fertility rate declined from more than six children per woman to less than three within only seven years, probably the world's most rapid national fertility decline. This decline occurred on a strictly voluntary basis and resulted from the combination of high female educational status and well-implemented family planning programs (Lutz, 1994). Because of the still very young age structure of the Mauritian population, current birth rates are still higher than death rates and the population is growing by about 1% year^{-1} despite fertility at around replacement level.

Empirically observed trends in all parts of the world have overwhelmingly confirmed the relevance of the concept of demographic transition to LDCs. With the exception of pockets where religious or cultural beliefs are strongly pronatalist, fertility decline is well advanced in all regions except sub-Saharan Africa, and even in that region many signs of a fertility transition can be perceived. In Southeast Asia and many countries in Latin America, fertility rates are on par with rates seen in MDCs only several decades ago, and in several countries, such as China, Taiwan, and Korea, fertility is at sub-replacement levels. In Southern Asia and North Africa, the demographic transition has not followed the pattern of a clear lag between the

decrease in death rates preceding a decrease in birth rates. The two rates declined almost simultaneously, somehow resembling the historical European pattern over the last 40 years.

The biggest difference between the demographic transition processes in what are now the MDCs and LDCs has been the speed of mortality decline. Mortality decline in Europe, North America, and Japan came about over the course of two centuries as a result of reduced variability in the food supply, better housing, improved sanitation, and, finally, progress in preventive and curative medicine. Mortality decline in LDCs, in contrast, occurred very quickly after World War II as a result of the application of Western medical and public health technology to infectious, parasitic, and diarrheal diseases. Life expectancy in Europe rose gradually from about 35 years in 1800 to about 50 in 1900, 66.5 at the end of World War II, and 74.4 in 1995. In LDCs, it shot up from 40.9 at the end of World War II to 62.1 in 1995. The increase was particularly impressive in Eastern Asia, where life expectancy at birth increased from around 40 to more than 70 years over that period. In Africa it only increased to slightly above 50 years, with life expectancy now falling in several countries due to AIDS. The increase that took MDCs about one and a half centuries to achieve came to pass in LDCs in less than half a century. As a result of the speed of the mortality decline, populations in LDCs are growing three times faster today than did the populations of present-day MDCs at the comparable stage of their own demographic transition.

Major social and economic changes, often summarized by the notion of modernization, are at the heart of the demographic transition theory. The changes in fertility trends are only one aspect of the radical social changes that many developing societies are experiencing. Cultural change, urbanization, and mass education have spread the image of the smaller nuclear family. It is to be expected that within the first decades of the 21st century, the final stages of demographic transition will be reached in all parts of the world.

See also: Demographic Change: Indonesian Transmigration, Volume 3; Demographic Change: Peopling of the Paci c Islands, Volume 3; Demographic Change: the Aging Population, Volume 3; Global Population Trends, Volume 3.

REFERENCES

Coale, A J (1973) The Demographic Transition, in *Proceedings of the International Population Conference*, Vol. 1, International Union for the Scientific Study of Population, Liège, Belgium.

Davis, K (1954) The World Demographic Transition, *Ann. Am. Acad. Polit. Soc. Sci.*, 237, 1–11.

Davis, K (1991) Population and Resources: Fact and Interpretation, in *Resources, Environment and Population: Present Knowledge*, eds K Davis and M S Bernstam, Oxford University Press, Oxford, 1–21.

Lutz, W, ed (1994) *Population Development Environment: Understanding their Interactions in Mauritius*, Springer-Verlag, Berlin.

Notestein, F W (1945) Population – the Long View, in *Food For the World*, ed T W Schultz, University of Chicago Press, Chicago, IL, 36–57.

Tabah, L (1989) From One Demographic Transition to Another, *Popul. Bull. UN*, 28, 1–24.

Development

see Development and Global Environmental Change (*Opening essay*, **Volume 5**)

Discounting

Discounting is a method of weighing monetary values that occur in different time periods (Portney and Weyant, 1999). The construction of a dam, for example, entails large expenditures on labor and materials while the dam is being constructed; revenues from hydropower accrue over the next 100 years, say, with relatively minor operating costs. Will the value of electricity over the next 100 years pay for the construction costs and provide investors a market return? Investors convert values at different time periods to a present value equivalent by discounting to the present. One dollar put in the bank at 10% interest is worth \$1.10 in a year, thus \$1.10 (a year from now) has a present value of \$1. If the \$1 were left in the bank for two years, it would be worth \$1.21 (1•1 × 1•1), or \$1 (two years from now) has a present value of 1•1•21 = 0•826. More generally, investors are interested in the net present value (NPV), i.e., whether the present value of the benefits is greater than the present value of the costs.

$$\mathrm{NPV} = \sum_{t_0}^{n} \frac{B_t - C_t}{\bullet 1 + r \bullet^n} \qquad \bullet 1 \bullet$$

where B_t is the benefit or revenue in year t, C_t is the costs in year t, and r is the rate of interest.

Economists justify discounting future benefits and costs in public decisions because both investors and consumers appear to discount future returns and costs. Furthermore, future harms can be offset by investing pennies today.

For example, suppose one tonne of carbon dioxide (CO_2) emissions in the year 2000 imposes $1000 of costs on individuals living in the year 2100. With investments earning 8%, those generating the emissions could offset the $1000 loss by investing a mere $0.45 today.

A very important caveat is in order. Discounting is inappropriate when society determines on ethical grounds that future generations should have the right to particular environmental or other conditions. Howarth and Norgaard (1995) have shown that when rights are assigned to future generations, the economy can still operate efficiently. For example, people may determine today that it is unethical to subject future generations to the unknown risks of global warming. Climate change would be averted on ethical grounds. The economy can still function efficiently given the new understanding of the rights of future generations, but those in the present generation who benefit from the prior distribution of rights would be worse off.

REFERENCES

Howarth, R B and Norgaard, R B (1995) Intergenerational Choice Under Global Environmental Change, Chapter 6, in *The Handbook of Environmental Economics*, ed D W Bromley, Blackwell, Cambridge, MA.

Portney, P R and Weyant, J P, eds (1999) *Discounting and Intergenerational Equity*, Resources for the Future, Washington, DC.

RICHARD B NORGAARD *USA*

Dominion Theology

see Christianity and the Environment (**Volume 5**); Theology (**Volume 5**)

E

Earth Art

see Art and the Environment (**Volume 5**); Landscape, Urban Landscape, and the Human Shaping of the Environment (**Volume 5**)

Earth Charter

The success of such global efforts as the International Declaration on Human Rights led planners of the UNCED to initiate negotiations for an Earth Charter. Beginning in August 1990, the Road to Rio process of preparatory meetings aspired to draft an inspiring, meaningful and relevant call to a new relationship between humanity and the biosphere. The effort engaged many non-governmental organizations (NGOs), particularly those within the faith community. Various drafts were circulated by the Society of Friends, the Buddhists and the Franciscan Institute (devoted to the life of Saint Francis of Assisi).

But negotiations fell apart before the 1992 Earth Summit. In an exchange that captured the differing visions of delegations, north and south, Canadian Ambassador Arthur Campeau urged that the statement be free of United Nations (UN) bureaucrat language; that it be simple, clear and inspiring, the sort of document you could hang on a child's bedroom wall. To which a delegate from a developing country responded, Our children don't have bedrooms.

The failure to achieve an Earth Charter saddened a number of its promoters, notably secretary general of UNCED, Maurice Strong. It was Strong who subsequently ensured that the fifth anniversary of Rio (Rio Plus 5, March 1997) launched a renewed effort to draft an Earth Charter. Strong co-chaired an Earth Charter Commission, with former President of the USSR, Mikhail Gorbechev, and included an impressive roster of past and current heads of government, celebrities, indigenous people, parliamentarians, religious leaders, youth and activists. The Commission released a benchmark draft in March 1997 and held a three year process of public outreach and debate. The chief drafter, Steven Rockefeller of Middlebury College, Vermont, traveled the world collecting thousands of interventions. Difficult balancing and negotiation took place to develop a document that could be acceptable across various and very diverse cultures, religions and social norms. In March 2000, the group met to finalize and approve the document. It will be presented to the UN at the special session marking 10 years since the Earth Summit in 2002.

The Earth Charter attempts to set out a framework of ethics and values to re-define the way humanity interacts with the natural world. Unlike previous environmental statements, such as the Stockholm and Rio Declarations, the Earth Charter has a profoundly spiritual basis. It acknowledges the concern for individuals within a species to be protected from cruelty. It enunciates the need for the practice of non-violence and not merely the absence of conflict.

The Earth Charter's fundamental commitments are to:

> respect the Earth in all its diversity; care for the community of life with love, understanding and compassion; build democratic societies that are just, sustainable and peaceful; secure Earth's bounty and beauty for future generations.

The intent is to secure a place for the Earth Charter as soft law through seeking widespread public support from around the world.

ELIZABETH MAY *Canada*

Earth Day

The first *Earth Day*, April 22nd, 1970, was nearly exclusively an event observed only within the US, but by its 13th anniversary, April 22nd, 2000, the event was global and involved half a billion people in over 100 countries.

Earth Day's humble beginnings grew out of the anti-Vietnam War movement's effective use of teach-ins on college campuses in the 1960s. US Senator Gaylord Nelson and Denis Hayes organized the first Earth Day as a consciousness-raising and educational event. The US

was reeling from news that Lake Erie was dead due to eutrophication, as proclaimed on the cover of *Life* magazine, and the chilling warnings of Rachel Carson's *Silent Spring*, less than a decade before. Earth Day gained rapid acceptance and was marked by a Presidential Proclamation signed by the then US President Richard Nixon. Large US corporations, such as General Motors and Coca Cola, joined in declarations of corporate environmentalism.

The success of the 1970 event can be tied to the passage of key environmental protection legislation in the US, such as the Clean Air Act, as well as the creation of governmental environment departments, a trend that was solidified by the first United Nations Conference on the Human Environment in Stockholm in 1972.

Earth Day's founders had initially not intended the day to be celebrated annually. Subsequent Earth Days have reflected the ebb and flow of public environmental support. Earth Day 1971 was a much smaller affair than that in 1970; while Earth Day 1990 was huge, Earth Day 1991 was nearly invisible. Denis Hayes was once again the chief organizer of Earth Day 2000, which focused on climate change. As a millennial bash it succeeded in gaining media attention, with a full special edition of *Time* magazine devoted to it. April 22nd has emerged after three decades as having a firm hold on the name *Earth Day*.

ELIZABETH MAY *Canada*

Earth First!

Earth First! (EF!) is the brainchild of an angry environmentalist and self described redneck, Vietnam veteran Dave Foreman. Al Gore, in *Earth in the Balance* (published in 1991), quotes Foreman as saying, It's time for a warrior society to rise up out of the earth and throw itself in front of the juggernaut of destruction, to be antibodies against the human pox that's ravaging this precious beautiful planet.

Seeing so much destruction of vanishing wilderness in the US, Foreman established a group under the banner No compromise in defense of the earth. Unlike other environmental groups working within the system to achieve change, EF! adopted tactics outside the law, espousing monkey wrenching and direct action. EF!, which has since disavowed the practice, spiked trees to discourage logging (a spiked tree can cause serious injury to a logger whose chain saw meets with unexpected metal).

In 1989, Foreman and three other EF! members were arrested by the Federal Bureau of Investigation (FBI) on charges of conspiracy to blow up a dam. The FBI had infiltrated the group and all discussions of explosives had originated with the FBI agent. The EF! members were acquitted.

Other EF! activists have experienced similar suspicious acts and violent sabotage. Pacific Northwest anti-logging activist Judy Barry was blown up in her car. She survived, but was seriously disabled. She maintained until her death years later that she had not been transporting the explosives as claimed by the authorities.

EF! continues to attract primarily young eco-activists. EF! is currently active in 13 countries. The group disavows hierarchies and conventional organizational structure. Its UK web page proclaims, EF! has no central office, no paid workers, no decision-making bodies, and is not even an organization in the normal sense of the word.

Its approach has led to a number of similar direct action groups, such as the Ruckus Society.

ELIZABETH MAY *Canada*

Ecocentric, Biocentric, Gaiacentric

Ecocentrism is an ideology or worldview that centers on and prioritizes the whole planetary ecosystem as the prime source of value and the proper focus of our attention. The term is virtually synonymous with *biocentric* (centering on the whole biosphere) and *Gaiacentric* (centring on Gaia – the earth, whose systems form one living organism). Such foci distinguish this worldview from *anthropocentrism*, whose humanistic priorities regard human perceptions and interests as the source of all value in nature.

Ecocentric/biocentric/Gaiacentric ideology underlies *deep ecology* environmentalism. Its central ethic assigns intrinsic value to all animate and inanimate elements in the cosmos – value, that is, which would reside in these elements even if human beings were not there to attribute their own values (for instance, economic utility) to nature. Intrinsic value implies that the role of humans is simply that of equal members of the ecological community, with no preemptory rights over other species or other parts of the system. Ecocentrism also emphasizes the importance of balance and interconnectedness in nature and of retaining the unity, stability, diversity, and harmony of all ecosystems.

Ecocentrism reflects the influences of early 12th century organicist philosophy, 19th century Romanticism and anarchism, and of an arcadian tradition in ecological science and in literature. This last is manifest in Aldo Leopold's Land Ethic (1949, *A Sand County Almanack*), which holds that A thing is right when it tends to preserve the integrity,

beauty, and stability of the biotic community. It is wrong when it tends otherwise. Some other ecocentric principles derived from Leopold's land ethic include: everything is connected to everything else; the whole is greater than the sum of the parts (synergy); human and non-human nature are a unity (*see* Leopold, Aldo, Volume 5).

From these principles, and from the notion of a whole ecological community of which animals, plants and humans are all equal members and interdependent parts, it is suggested that the environment has direct rights, that it qualifies for moral personhood and that it is deserving of a direct duty. Ecocentrism then clearly places the environment on a moral par with humans, giving it a biotic right. If applied in practice, such a right would foster environmental protection and preservation over the approach of conservation – the latter implying managed change in accordance with considerations of economic utility. It would also confer legal standing on wildlife and other elements in nature, enabling environmentalists to defend them in court against the damaging effects of potential development, as has happened in the US.

DAVID PEPPER *UK*

Ecocentrism

see Environmental Ethics (**Volume 5**)

Ecofeminism

Karen J Warren
Macalester College, St Paul, MN, USA

The term ecofeminism was coined in 1974 by Francoise d'Eaubonne to call attention to women's potential to bring about an ecological revolution (d'Eaubonne, 1974). Unlike earlier feminisms, which focused on sex-roles, equal rights, equity in the workplace, educational opportunities for women (e.g., liberal feminism), socioeconomic conditions of women as workers (e.g., Marxist feminism), ecofeminism understands feminism to be a movement to end all systems of domination, including the domination of non-human animals and nature. Since 1974, ecofeminism has surfaced throughout the globe in the form of both women-initiated, grassroots environmental actions and interdisciplinary perspectives on the inextricable interconnections among human systems of unjusti ed domination - both of humans and earth others. The distinctiveness of ecofeminism, then, is that it is a feminist environmentalism and an environmental feminism. Just as there is not one feminism, there is not one ecofeminism. There is no one, uni ed ecofeminist perspective on environmental issues. What ecofeminists have in common is a commitment to the historical interconnections among human systems of domination (including the unjusti ed domination of non-human nature), and to replacing oppressive practices, policies, and philosophies with ones that are not. What makes ecofeminism feminist is its starting point: It begins with women and sex/gender analysis to call attention to the interconnected exploitations and liberations of women, other 'human others' (e.g., people of color, poor people, children, colonized peoples) and 'earth others'. What makes ecofeminism ecological is that non-human animals and nature are integral to the practice and theory of ecofeminism. Ecofeminism not only makes visible the ways in which the well-being of the natural environment is linked to the well-being of diverse groups of humans; it also expands on the notion of a human self as an ecological self. Using insights from both the environmental movement and ecology in the framing of ecologically responsible and sustainable policies, practices and philosophies, ecofeminism has emerged as a political movement among all racial/ethnic, socioeconomic, gendered, and geographical locations.

Ecofeminism is represented by a variety of feminist perspectives and practices. It nds expression in a wide range of (mostly) women-initiated, community activist initiatives around such issues as air and water pollution, biodiversity and cultural diversity, deserti cation and deforestation, indigenous agriculture and food production, toxins and hazardous waste disposal, global warming and acid rain, human overpopulation and habitat destruction, species preservation, development and biotechnology. The slogan of ecofeminism might well be: nature is a feminist issue. At the heart of ecofeminist practice and theory are the various interconnections - historical, causal, empirical, economic, political, scienti c, symbolic and literary, linguistic, conceptual, ethical, religious and philosophical - between systems and practices of domination. As a movement of global environmental change, the main project of ecofeminism is two-fold: to make visible the ways in which the exploitation and destruction of the non-human environment is tied to the exploitation of women, people of color, children, and the poor; and to develop policies, practices, and philosophies which undo and overcome these interconnected dominations. There is literally no area of environmental concern that falls outside the interest of ecofeminist activism and theory. This is not surprising, considering the intimate connections between the health, status, and well-being of the non-human environment and the health,

status, and well-being of women and other 'human others' cross-culturally - now and in the foreseeable future. According to ecofeminists, any plausible, workable, responsible agenda for the 21st century will include ecofeminist insights into and solutions to the globally interlocking systems of exploitation and domination.

EMPIRICAL WOMEN–NATURE INTERCONNECTIONS

Since the starting point of ecofeminism is women, at the outset ecofeminism focuses on women–nature interconnections. But since all women are of some race/ethnicity, socioeconomic status, geographic location, sexual orientation, and marital status, ecofeminism is also about race–nature, class–nature, geographic location–nature, sexual orientation–nature, and marital status–nature interconnections. For simplicity, this essay subsumes these varieties of interconnections under the generic rubric *women- nature interconnections*.

One way to reveal the sorts of concerns that characterize ecofeminism is to consider empirical data on women–nature interconnections. Such data help to clarify why nature is a feminist issue.

Trees

Tree shortages are a problem in many parts of India. Historically, first-world development policies that replaced indigenous, multispecies forests with monoculture eucalyptus and teak plantations solved the problem of too few trees, but in a way that disproportionately affected the livelihoods of local women. Traditionally, women in rural India have gendered responsibilities to maintain domestic economies, which are highly dependent on trees and tree products for food, fodder, fuel, herbs, dyes, medicines, household utensils and building materials, and income-generating activities. Monoculture eucalyptus is unpopular among local women because it provides no food, fodder, herbs, dyes, and medicines and is a problematic fuel. Replacement of indigenous forests with eucalyptus thereby affects women's abilities to maintain household economies, especially when these women also must take on the labor tasks of the local men who move to city areas to work in the eucalyptus plantations.

Food

Women farmers globally grow at least 59% of the world's food; in parts of Africa, women grow 80–90% of the food. Women farmers tend to work longer hours, have fewer assets and lower incomes than men farmers, primarily because their access to credit is limited. Agricultural policies and practices exported from the Northern Hemisphere (the North) to the Southern Hemisphere (the South) often do not recognize the full extent of women's contributions to all aspects of agricultural work (e.g., in plowing, planting, harvesting, weeding, processing and storing crops, and caring for livestock), involve technological improvements which are not aimed at training women or which develop products inappropriate from the standpoint of local women (e.g., solar stoves for women in Africa who cook before dawn and after dusk, or maize shellers that are more time-consuming to use than when women shell by hand).

Water

Less that 50% of the population of the South has a source of potable water or facilities for sewage disposal. Since women and children in the South perform most of the water collection and distribution work, it is women and children who are disproportionately harmed by the presence of unsanitary water. Also, the time spent by women and children in water collection in Africa and Asia (from 5–17 h each week) is time unavailable for other activities (e.g., attending school). Droughts and floods exacerbate the situation. Droughts and floods are among the most serious natural disasters in the South, contributing to desertification, deforestation, and soil erosion on a large scale. But it is poor people, especially women and children, who are most seriously affected by these natural disasters, since they constitute the largest group who live in those areas.

The North also has water-related problems that disproportionately affect women, communities of color, and the poor. The presence of lead and leaking chemicals, as well as the improper disposal of hazardous wastes, has contributed to groundwater contamination of municipal water supplies in all parts of the US. Ground water, the drinking water source for nearly half of the US population, disproportionately affects inner city poor, especially pregnant women, women of color, children, and the elderly, more than other groups.

Toxins

The United Church of Christ's Commission on Racial Justice published an already classic report in 1987 entitled *Toxic Waste and Race in the United States*. The commission found that race is the most important determinant in the location of hazardous waste in the US. Three out of every five African and Hispanic Americans, and more than half of all Asian Pacific Islanders and American Indians live in communities with uncontrolled toxic waste sites. Emelle, Alabama, which is almost 80% African American, houses the nation's largest hazardous waste landfill, receiving toxins from 45 states. Probably the greatest concentration of hazardous waste sites in the US

is on the predominantly African American and Hispanic south side of Chicago. Navajo Indians are the primary workforce in the mining of uranium in the US. However, the dumping of uranium tailings on Navajo (and other American Indian) lands has resulted in higher percentages of health risks (e.g., miscarriages for pregnant women, gynecological cancers, cleft palate and other birth defects) than non-Indian populations.

According to ecofeminism, awareness of these empirical interconnections between the status and treatment of women and other human sub-dominant groups, and the status and treatment of non-human nature, must be part of any informed environmental practice, policy, or philosophy. But ecofeminism doesn't stop there. Awareness of other sorts of women–nature interconnections must also be made visible, addressed, and, where unjustified domination is involved, replaced by life-affirming practices, policies, and philosophies.

ADDITIONAL WOMEN–NATURE INTERCONNECTIONS

These examples provide concrete illustrations of ways in which the exploitation of women is intimately connected to the destruction of the earth. They also make visible why sensitivity to the plights and expertise of women and other human sub-dominants is important to any informed feminism, environmental activism, and environmental policy. But these empirical connections are just one sort of women–nature interconnection of interest to ecofeminists. Consider some of the other ones.

Some ecofeminists have focused on historical interconnections between the domination of women (and other humans others) and the domination of Earth. Ecofeminist philosophers, for example, have been among those who trace patterns of domination to the Western philosophical tradition of rationalism, according to which the mental trait of reason or rationality is both the essence (or hallmark) of humanness and that which elevates (at least some) humans above the inferior, physical, bodily, corporeal realm of nature (Plumwood, 1993). Drawing on the dualisms of mind/body, reason/emotion, culture/nature, that which is associated with mind, reason and culture is claimed to be superior to that which is associated with body, emotion, and nature. In systems of domination, the subordinated others (e.g., women, slaves, children) have historically been identified with the inferior physical world of nature; the justified domination of these other human others' is based on their alleged association with what is biological, natural, bodily, or physical.

Ecofeminist historian Carolyn Merchant provides a different account. Merchant argues that the death of nature occurred between 1500–1700, when an older world order was replaced by a reductionist, mechanistic world view of modern science, which sanctioned the exploitation of nature, unchecked commercial and industrial expansion, and the subordination of women. According to Merchant (1980, 2)

> The metaphor of the earth as a nurturing mother was gradually to vanish as a dominant image as the Scientific Revolution proceeded to mechanize and to rationalize the world view. The second image, nature as disorder, called forth an important modern idea, that of power over nature. Two new ideas, those of mechanism and of the domination and mastery of nature, became core concepts of the modern world.

This change in controlling imagery was directly related to changes in human attitudes and behavior toward the earth. Whereas the nurturing Earth image can be viewed as a cultural constraint restricting the types of socially and morally sanctioned human actions allowable with respect to Earth, the new images of mastery and domination functioned as cultural sanctions for the denudation of nature. According to Merchant, the change in dominant nature imagery from organism to machine permitted the removal of classical moral constraints on how one treats mother nature, a living organism. The mechanistic metaphor thereby helped sanction the exploitation of inert, passive nature as object of study.

These historical interconnections raise important conceptual issues about social systems of oppression and domination, conceptions of the human self, and the nature of human responsibilities to the non-human world. For example, ecofeminist philosophers have provided analyses of fundamental features of oppressive conceptual frameworks that underlie systems of domination (Warren, 1990). A conceptual framework is a set of basic beliefs, values, attitudes, and assumptions about oneself and one's world. It is a historically located, socially constructed thought world – a lens through which one sees and interprets the world. When a conceptual framework is oppressive (e.g., a sexist, racist, or classist conceptual framework), the beliefs function to justify the unjustified domination of the other (e.g., women, people of color, the underclass).

There are five noteworthy features of an oppressive conceptual framework: (a) value-hierarchical, up–down, thinking; (b) oppositional and mutually exclusive value-dualisms (*either- or* thinking); (c) conceptions of power which advantage the power of ups over downs; (d) conceptions of privilege which systematically advantage the ups over downs; and (e) a logic of domination: a moral premise which justi es the subordination of downs by ups on the grounds that the ups have a morally relevant trait which the downs, as downs, lack. Whether the conceptual framework is sexist, racist, classist, heterosexist, colonialist, or ethnocentric, insofar as they are oppressive conceptual frameworks, these ve features seem to be shared in common (Warren, 2000).

These basic conceptual interconnections among social systems of domination are important to ecofeminist philosophy. Oppressive conceptual schemes justify relationships of domination and subordination through a logic of domination, which asserts that whatever is up deserves to be up and is morally justified in treating as inferior whatever is down. When women, people of color, the poor or children are downs in *up down* relationships of domination and subordination, their justified subordination often has rested historically on their alleged association with the inferior, non-mental realm of animals and nature. In this way, sexism is conceptually tied to *naturism*, i.e., the unjustified domination of animals and non-human nature.

There are two basic ways this conceptual link occurs. First, sexism and naturism are linked through an oppressive conceptual framework characterized by a logic of domination. Since all feminists must oppose the logic of domination (required by patriarchal arguments for sexism), and since the arguments for the domination of non-human nature presuppose the same logic of domination, all feminists must oppose the arguments for the domination of non-human nature. The logic of traditional feminism thereby requires extending feminism beyond a concern with sexism, or even a concern with all forms of human subordination, to a concern with naturism; naturism must also be included about the systems of domination that feminism opposes (Warren, 1990).

The second conceptual reason sexism and naturism are linked has to do with the concepts of gender and nature. The concept of gender is socially, historically, and materially constructed. Both the concept *women*, and the identities of particular women, are constructed from and reflect such factors as race/ethnicity, class, age, affectional orientation, ability, geographic location, and religion; to be a woman is to be a woman of a certain race/ethnic, class, age, affectional orientation, ability, geographic location, and religion. This means that there is no pure gender; gender always occurs in conjunction with race, class, age, affectional orientation, geographic location.

Similarly, the concept of nature is socially and humanly constructed; it is not a fixed, static, self-evident, given or absolute concept. What is meant by nature, and even what counts as a natural object for humans, is constructed from and reflect such factors as the race/ethnicity, class, age, affectional orientation, geographic location, and religion of humans who name, describe, judge, understand, and interact with nature. The meanings of nature are constructed out of human values, beliefs, attitudes and assumptions – different and differing human conceptual frameworks.

Obviously, the claim that women and nature are social constructions does not deny that there are actual humans and actual non-human natural entities. One can hug a human, climb a tree, swim a river, cultivate a plant, destroy an ecosystem, or enjoy being in nature. These entities (e.g., humans, trees, nature) are physical realities. But what is meant by human, tree, or nature, and what constitute the identities of humans, non-human natural entities, and nature are social realities as well. These social realities are social constructions. What constitutes country for Australian aboriginal peoples is not what Europeans mean by country, land, landscape, property, or wilderness. In the words of anthropologist Deborah Bird Rose, the closest European and American-derived expression is nourishing terrain, a place that gives and receives life. Not just imagined or represented, it is lived in and lived with. Country is sacred, alive, conscious, the law, place, managed, owned, land. Country knows, hears, smells, takes notice, takes care, is quiet, dangerous, lonely, sorry or happy or sad, is healthy, sick, or not so good. Country is criss-crossed with song lines and the tracks of the Dreamings: walking, slithering, crawling, flying, chasing, hunting, weeping, dying, giving birth (Rose, 1996, 35). The relationships between people and their country is a a kinship relationship, and like relations among kin, there are obligations of nurturence. People and country take care of each other (Rose, 1996, 49).

Such linguistic differences about what something is called are important to ecofeminism. As many philosophers (e.g., Wittgenstein) have argued, the language one uses is deeply embedded in conceptions of oneself and one's world – one's conceptual framework. When that language is sexist or naturist, it reflects and reinforces conceptions of women and non-human nature as inferior to, having less status, value, or prestige than, that which is identified as male, masculine, or human (i.e., male).

Such is the case with the language used historically in Western societies to describe women and nature. Women are described in animal (nature) terms as pets, dogs, cats, pigs, cows, sows, foxes, chicks, serpents, bitches, beavers, mares, old mares, old bats, old hens, mother hens, pussies, pussycats, cheetahs, bird-brains, hare-brains, elephants, and whales. Language that animalizes or naturalizes women in a patriarchal culture where animals are seen as inferior to humans (men) thereby reinforces and authorizes women's inferior status. Similarly, language that feminizes nature in a patriarchal culture where women are seen as subordinate or inferior reinforces and authorizes the domination of nature. Mother nature is raped, mastered, conquered, mined, controlled, poked, prodded, and pried into; her secrets are penetrated and her womb is put into service by the man of science. Virgin timber is felled, cut down; fertile (not virile) soil is tilled and land that lies fallow is barren (not impotent) and useless, like a woman who cannot conceive a child. The exploitation of nature and animals is justified by language which feminizes them; the exploitation of women is justified by language which naturalizes women.

Just as the justification for the dominations of women and non-human nature is reinforced by sexist–naturist language,

so too the justification for the domination of people of color is reinforced by racist-naturist language. For example, in the US, the justification for dominating African American men was bolstered by language and images of sub-dominant children, animals, and nature. Slaves were mere brutes, animals, dogs, coons, apes. African American men were denied the rights of citizenship as a subordinate and inferior class of beings, who had been subjugated by the dominant race ••• (Stone, 1974; n 19). The legal status of slaves as things, mere property or chattel was accepted law in the 1850s.

Languages of domination are crucial to effective implementations of the logic of domination by colonizers. Standard European and Anglo-American land-use attitudes and descriptions of wilderness involve the peculiar notion that if one cannot see traces of one's own culture in the land, then the land must be natural or empty of culture (Rose, 1996). Terra nullius (land that was not owned) was the official doctrine of (white) Australia until overturned by the High Court's Mabo Decision in 1995. That decision marks the first time Aboriginal and Islander people's occupation and ownership of country was formally and publicly acknowledged.

The view that land with no ownership is uninhabited wilderness is familiar in the US as well. Wilderness is often conceived as wild, idle, worthless frontier until tamed and cultivated through the white settler's agriculture. Land that just lies there, barren, useless, uncultivated, for example, untouched virgin prairie, has no value until domesticated by the white man's plow. White racism is not just a social construct around skin color and ethnicity; it is also an ecological construct, one that reflects and reinforces ecological as well as social factors.

Many ecofeminists argue that cultural transformation must occur if these interlocking systems of domination are to be eliminated. Since historically there are important symbolic interconnections between the domination of women (and other human others) and the domination of non-human nature, one must examine a culture's images, rituals, rites, ceremonies, stories, myths, songs, literature, theologies, religions and spiritual practices.

Some ecofeminists have examined the religious and theological symbols that affect our cultural views about humans and nature. For example, ecofeminist theologian Elizabeth Dodson Gray argues that the Christian biblical tradition is a hierarchical order of creation, a pyramid of dominance and status, where woman comes after and below man, children come after women, and pets, animals, plants and nature itself follow, in descending order of status and worth. Gray argues that we need to think ourselves out of such a patriarchal conceptual trap by revisioning a theology of creation that is not a hierarchical pyramid of dominance and subordination (Gray, 1979).

Other ecofeminists (spiritual ecofeminists) agree with Gray. Despite their very vocal critics, including ecofeminist critics, spiritual ecofeminists (e.g., Spretnak, 1982) argue that non-patriarchal, earth-based spiritualities are not only necessary to any healing relationship between humans and the earth, but to any liberatory ecofeminist politics. The politics of women's spirituality must be a Green politics.

Some ecofeminists turn to literature, particularly women's nature writing, to affect a cultural transformation. Paula Gunn Allen, Margaret Atwood, Rachel Bagby, Annie Dillard, China Galland, Sally Gearhart, Susan Griffin, Linda Hogan, Winona LaDuke, Ursala Le Guinn, Marge Piercy, Leslie Marmon Silko, Alice Walker, and Terry Tempest Williams are just a few of the women who have explored these interconnections through their stories, poems, novels, and essays. Ecofeminist literature is important because cultural images reflect not only who we are but help shape who we become. Attitudes toward animals, plants, mountains, rivers – what Leopold (1949) refers collectively to as *the land* – reflect deeply seated values about culture, nature, wilderness and the wild, and human nature. Some nature writers romanticize and uncritically idealize wilderness. Others imagine a wild landscape inhabited by humans who occupy a unique place as co-members of the ecological community. Still others challenge the very concepts of nature and wilderness, and the historical construction of a nature/culture dualism, which puts (some) humans in oppositional superiority over nature. The emerging field of ecofeminist literary criticism examines the political dimensions and ramifications of these concepts within ideologies of domination, while seeking to develop literature and criticism that emphasizes appropriate and edifying human relationships within the ecological community.

When one challenges cultural sources of systems of domination, epistemological claims also get challenged. Whose knowledge? and Knowledge according to whom? become important questions, since ecofeminist theory and practice always examines the social context in which epistemological claims are generated. Women in the so-called developing world who have gender responsibilities for managing forests, for example, may have invaluable technical knowledge about multiple uses of trees, based on their hands-on, daily, lived experience as forest managers. Their experiential ways of knowing – their epistemology – may challenge mainstream, Western conceptions of the objectivity, value-neutrality, and impartiality of genuinely scientific knowledge claims.

Epistemological worries often interact with psychological ones. Ecofeminist sociologist Ariel Salleh, for example, challenges ethical and epistemological positions that embody a distinctly masculine sensibility, which desires to transcend nature and culture. Salleh argues that since ecofeminism reveals the patriarchal conceptual foundation

of the continued destruction of the natural environment, what is needed is that the unconscious, psychological connection between women and nature needs to be made conscious, and the hierarchical fallacies of the great chain of being acknowledged, in order for any real growth towards a sane, humane, ecological future (Salleh, 1984).

Given the nature of some of these interconnections between the treatment of women and the treatment of non-human nature, it is not surprising that perhaps the most visible ecofeminist theory and activism has been in the political arena. Conferences, workshops, direct actions, and grassroots organizing around environmental issues have been initiated globally by a new radical core of low-income women environmentalists. These women are motivated by the irrationalities of capital-intensive growth and the destruction of their domestic economies. Whether it is the Chipko movement in India, the Women's Pentagon Actions in the US, the Seminar on Ecofeminism at the United Nations Conference on Environment and Development in Rio de Janeiro (1992), or the varieties of global women-initiated actions for better health care, reproductive rights, and environmentally safe neighborhoods, ecofeminist activism has an international presence as a distinctly feminist and environmental movement.

Lastly, much of the scholarly literature of ecofeminism has focused on environmental ethics. Minimally, the goal of ecofeminist ethics is to develop theories and practices concerning humans and the natural environment that are not male-biased and which provide a guide to action in the pre-feminist present. This has involved the development of a variety of ecofeminist ethical positions and practices, where notions of care, loving perception, friendship, sharing, appropriate reciprocity, and kinship emerge as morally relevant and important values. One needs only to look at the range of ecofeminist positions regarding such practices as factory farming, vivisection, worldwide traffic in wild animals, violence toward women, health care, reproductive technologies, the location of hazardous waste, and environmental conservation and preservation to see the impact of ecofeminist moral theory in moral practice.

Since its inception as a distinct feminism, environmentalism, and form of feminist practice in 1974, ecofeminism has emerged as an important, timely, powerful movement at local, national, international, and global levels. As a social and political movement, ecofeminists have taken the lead in exploring the roles of science, development and technology in the development of environmental policies (e.g., on biodiversity, global warming), practices (e.g., the development of biotechnologies), and decision-making structures (e.g., for implementing Agenda 21), which make the expertise of women and indigenous peoples central to both the analysis and resolution of pressing environmental problems. Ecofeminist activism has produced change through numerous kinds of action, including direct action, civil disobedience, strikes, boycotts, demonstrations, lobbying, and other more traditionally recognized forms of political action (Sturgeon, 1997). Ecofeminist activists have engaged in coalition building among and across different constituencies (e.g., peace, environmental, women's, anti-racism, anti-colonialism, animal rights movements). And ecofeminist activists have raised awareness about the interconnections among all political and economic structures of domination and exploitation.

As a theoretical position, ecofeminist scholarship has broadened and deepened Western and non-Western views in such different arenas as the arts, economics, ethics, geography, history, literature, philosophy, politics, religion, sociology, and science. It has helped shape a new sense of our human selves as ecological, relational selves who are both similar to and different from non-human nature in distinct and important ways. It has been a leader in the environmental justice movement, and has helped white Western feminism to expand its self-conception to recognize not only the interconnections among forms of human domination and liberation based on gender, race, class, affectional orientation, age, and geographic location, but among these and the domination and liberation of non-human animals and nature.

Perhaps the most important theoretical contribution of ecofeminism is simply the starting point of its environmental analysis: using concrete empirical data and women's felt, lived experiences to show the very real interconnections between the plight of women, other human others, and the domination of non-human nature. This starting point makes any adequate theory or practice of feminism, environmentalism, or human liberationism intimately and inextricably interconnected.

The remarkable cross-cultural flourishing of ecofeminism internationally since d'Eaubonne (1974) coined the term is testimony to its power and promise. Ecofeminism is visible in the practical, policy, and theoretical spheres. It is grassroots, multiracial/multiethnic, local and global. It is action-oriented, giving voice to local women-initiated environmental practices. It is forward-looking and transformative in focus, moving us beyond harmful social systems of domination, which endanger all life on earth. According to ecofeminism, any feminism, environmentalism, or environmental policy which fails to include ecofeminist insights into the unjustified and interconnected dominations of women, other human others, and non-human nature will simply be inadequate for the needs of the 21st century.

REFERENCES

d'Eaubonne, F (1974) *Le Feminisme ou La Mort*, Pierre Horay, Paris.

Gray, E D (1979) *Green Paradise Lost*, Roundtable Press, Wellesley, MA.

Leopold, A (1949) The Land Ethic, in *A Sand County Almanac and Sketches Here and There*, Oxford University Press, New York.
Merchant, C (1980) *The Death of Nature*, Harper and Row, San Francisco, CA.
Plumwood, V (1993) *Feminism and the Mastery of Nature*, Routledge, London.
Rose, D B (1996) *Nourishing Terrains*, Australian Heritage Commission, Canberra.
Salleh, A K (1984) Deeper than Deep Ecology: The Eco-feminist Connection, *Environ. Ethics*, 6, 339–345.
Spretnak, C, ed (1982) *The Politics of Women's Spirituality*, Doubleday, Garden City, NY.
Stone, C (1974) *Should Trees Have Standing?* William Kaufmann, Los Altos, CA.
Sturgeon, N (1997) *Ecofeminist Natures*, Routledge, London.
Warren, K J (2000) *Ecofeminist Philosophy: A Western Perspective on What it is and Why it Matters*, Rowman and Littlefield, Lanham, MD.
Warren, K J, ed (1997) *Ecofeminism: Women, Culture, Nature*, Indiana University Press, Bloomington, IN.

FURTHER READING

Adams, C, ed (1993) *Ecofeminism and the Sacred*, Continuum, New York.
Anderson, L, ed (1991) *Sisters of the Earth*, Vintage Press, New York.
Cuomo, C J (1998) *Feminism and Ecological Communities*, Routledge, London.
Gaard, G, ed (1993) *Ecofeminism: Women, Animals, Nature*, Temple University Press, PA.
Kolodny, A (1984) *The Land Before Her*, University of North Carolina Press, Chapel Hill, NC.
Mellor, M (1997) *Feminism and Ecology*, New York University Press, New York.
Mies, M and Shiva, V (1993) *Ecofeminism*, Zed Books, Atlantic Highlands, NJ.
Murphy, P D (1995) *Literature, Nature, and Other*, State University of New York, Albany, NY.
Ruether, R R (1974) *New Woman, New Earth*, The Seabury Press, New York.
Shiva, V (1988) *Staying Alive*, Zed Books, London.
Warren, K J, ed (1996) *Ecological Feminist Philosophies*, Indiana University Press, Bloomington, IN.
Warren, K J (1990) The Power and the Promise of Ecological Feminism, *Environ. Ethics*, 12, 125–146.
Zahava, I, ed (1988) *Through Other Eyes*, The Crossing Press, Freedom, CA.

Ecological Economics

see Ecological Economics (*Opening essay*, **Volume 5**)

Economics and Global Environmental Change

see Economics and Global Environmental Change (*Opening essay*, **Volume 5**)

Economics, Environmental

see Environmental Economics (**Volume 5**)

Eco-socialism

Eco-socialism is a radical homocentric application of socialist analysis and prescriptions to environmentalism. It also modifies traditional socialism to take account of environmental issues and perspectives. Its critical analysis of history, social change and economics draws particularly on Marx's (early) writings, especially as interpreted by William Morris. Its prescriptions often revive utopian socialist traditions of decentralization, direct economic democracy, communal ownership of the means of production, etc. Hence the brand of socialism represented is close to anarchist-communism, although there are some major differences between eco-anarchists and eco-socialists concerning analysis and strategies. Eco-socialism's historical materialist analysis locates the causes of contemporary environmental abuse in the workings of the economic mode of production of capitalism, and the institutions and world view necessary to its functioning. Eco-socialism argues that environmentally unsustainable development is inherent to capitalism, therefore to end the former, the latter must be abolished and replaced by socialism. In socialism, it is argued, people can end the alienation from nature and from each other that causes environmental degradation, yet production and industry in pursuit of Enlightenment Project ideals could continue. Such production, with distribution, would be rationally planned – perhaps by an enabling state, or, in more anarchistic visions, by bodies representing federation of local communities and regions.

Eco-socialist society would rediscover and express our real relationship to nature – neither separation and superiority, as contemporary capitalism presupposes, nor of mere equality, as ecocentrism believes. Rather, society and nature are dialectically related, so that each is a manifestation of the other. Nature is socially produced, and what humans do is natural. Socialist communities would recognize that they are not intrinsically determined by nature's limits, as deep ecology believes. But they are likely to want to steward,

protect and wisely manage relationships with nature, for the benefit of all community members.

DAVID PEPPER *UK*

Ecosystem Approach

What constitutes valid science and a valid scienti c method? That question has never been fully resolved to the satisfaction of every informed scientist who has employed a scienti c method that had been duly legitimated within some network of expert and responsible peers. There have always been critics to challenge any claim that some particular kind of science was the correct or the best one. Perhaps a view shared most widely among scientists is that honest and modest skepticism is a necessary ingredient within science. An opposing view might be termed fundamentalist.

The concept that species including humans have emerged historically within evolution, in which natural selection plays a strong role, was presented scienti cally by Charles Darwin and others in the mid-19th century. Some Christian scholars took exception to the whole notion of natural evolution. Skeptical scientists and theologians may enjoy debating but fundamentalistic scientists and theologians clash.

The con ict within science concerning evolution is not as well known as that between some scientists and some theologians. In the mid-19th century the physicists and chemists were using experimental methods in laboratories to discover many laws of nature. These scientists preferred to view the world and cosmos as a kind of clockworks in which every phenomenon had a prior cause; cause effect linkages could be described and explained fully in linear deterministic terms. In principle, a fully informed scientist could draw diagrams, de ne terms, write formulae and insert numbers that altogether explained something completely. Such a scientist could achieve conceptual closure, which presupposed that nature itself was closed to seminally new happenings.

But the concept of evolution implied that nature was open to the emergence of new species, say. So some aspects of nature must remain open even though linear laws like those of Isaac Newton, which appear to be deterministic, may also play a role. Are universal laws only partially successful in constraining the contextual processes apparent in evolution? Or may there be ways of perceiving contextual phenomena that are consistent with universal laws, or *vice versa*? Generally, should a scientist start with accepted linear universal laws and then study apparent non-linear exceptions in particular contexts, option A. Or should he/she start with non-linear contextual speci cs and try to infer quasi-linear universal generalities, option B? Perhaps a dialectical approach can be taken in an attempt to be fair to both options. Eventually some scientist may nd or has found a way to synthesize and transcend such a debate.

The concepts of ecology, ecological systems and ecosystems arose within geological, biological and geographic studies that addressed the tension between these two mindsets, or what Magoroh Maruyama calls *mindscapes*. Thus Ernst Haeckel, a contemporary of Darwin, worked at the interface of evolution through natural selection and an approach that has come to be called *environmental determinism*. Later ecologists such as Arthur Tansley deliberately chose option A, more or less. Other ecologists such as Frederick Clements chose option B. An early stage of a synthesis may now be emerging with a focus on complexity as in the innovations of Lynn Margulis, James Lovelock and many others, option C. Each approach continues to have its protagonists entering the 21st century, though the ranks of option C appear to be swelling at the expense of options A and B.

According to this simplistic classi cation, three somewhat different versions of *ecosystem approach* can now be encountered in the scienti c and popular literatures that focus on biological phenomena. Within much of contemporary biology there may still be an underlying presupposition that the disciplined, disinterested scientist can make objective observations, conduct logically correct analyses and infer valid generalities about some features of reality. Such a presupposition may be termed *deterministic* or *positivistic*. Thomas Kuhn apparently intended his notion of a paradigm to relate to a properly justi ed inference obtained by using an approach based on such a presupposition.

On the other hand, these terms, and especially ecosystem approach, are coming to be used with perceived phenomena that include what is formally biological and much more. Some include global atmospheric and hydrological processes, the current manifestations of which are strongly dependent on historical biological events. Or cultural economic and governance processes may be included. An ecosystem approach may even be invoked with respect to cosmogenesis, or the evolution of the cosmos during the past 10+ billion years. With some of these wider applications of the term, positivistic presuppositions may be replaced by a set of constructivist presuppositions that relate primarily to the working hypotheses within a network of scienti c peers and not to some prior notion of external reality. Within such a mindscape, a particular version of an ecosystem approach may be perceived as a dynamic movement rather than a Kuhnian paradigm.

Within current governance processes related to environmental issues, a linear deterministic version, as in option

A, may still be found in some research and management, related to utilitarian interests in natural resources, say. Option B may dominate with preservationist interests as in biodiversity. Option C may dominate in studies and practices concerning natural/cultural systems in which the residents are committed to a policy of responsible reciprocity or caring sharing among all the kinds of creatures, as with the role of Aboriginals in environmental matters. Innovators with an option C kind of ecosystem approach may find useful roles for contributions by option A and B kinds, so long as the latter do not act to prevent their transcendence as in option C.

FURTHER READING

Caley, M T and Sawada, D, eds (1994) *Mindscapes: the Epistemology of Magoroh Maruyama*, Gordon and Breach, London.

Fuller, S (2000) *Thomas Kuhn: A Philosophical History for Our Times*, University of Chicago Press, Chicago, IL.

Kay, J, Regier, H, Boyle, M, and Francis, G (1999) An Ecosystem Approach for Sustainability: Addressing the Challenge of Complexity, *Futures*, **31**, 721–742.

Keller, D R and Golley, F B (2000) *The Philosophy of Ecology: from Science to Synthesis*, University of Georgia Press, Athens, GA.

Lovelock, J E (1979) *Gaia: a New Look at Life on Earth*, Oxford University Press, Oxford.

Margulis, L (1998) *Symbiotic Planet: a New Look at Evolution*, Basic Books, New York.

HENRY A REGIER *Canada*

Ecosystem Integrity

see Ecosystem Integrity (**Volume 4**)

Ecosystem Services

Ecosystem services are the goods and life support functions provided by natural ecosystems and their species that sustain and fulfill human life. Ecosystem services are the profit that we as humans reap from the Earth's natural capital.

Table 1 Ecosystem services and functions adapted from Costanza *et al.* (1997) with permission from *Nature* (http: //www.nature.com/)

Ecosystem services	Examples
Regulation of atmospheric chemical composition	Carbon dioxide (CO_2)/oxygen (O_2) balance, O_3 for UVB protection and SO_x levels
Regulation of global temperature, precipitation, and other biologically mediated climatic processes at global or local levels	Greenhouse gas regulation, DMS production affecting cloud formation
Capacitance, damping and integrity of ecosystem response to environmental fluctuations	Storm protection, flood control, drought recovery and other aspects of habitat response to environmental variability mainly controlled by vegetation structure
Regulation of hydrological flows	Provisioning of water for agricultural (such as irrigation) or industrial (such as milling) processes or transportation
Storage and retention of water	Provisioning of water by watersheds, reservoirs and aquifers
Erosion control and soil retention within an ecosystem	Prevention of loss of soil by wind, runoff, or other removal processes, storage of stilt in lakes and wetlands
Soil formation processes	Weathering of rock and the accumulation of organic material
Storage, internal cycling, processing and acquisition of nutrients	Nitrogen fixation, nitrogen, phosphorous and other elemental or nutrient cycles
Recovery of mobile nutrients and removal or breakdown of excess or xenic nutrients and compounds	Waste treatment, pollution control, detoxification
Pollination and movement of floral gametes	Provisioning of pollinators for the reproduction of plant populations
Biological control (trophic dynamic regulations of populations)	Keystone predator control of prey species, reduction of herbivory by top predators

(*continued overleaf*)

Table 1 (*continued*)

Ecosystem services	Examples
Refugia; suitable habitats for resident and transient populations	Nurseries, habitat for migratory species, regional habitats for locally harvested species, or overwintering species
That portion of gross primary production extractable as food	Production of fish, game, crops, nuts, fruits by hunting, gathering, subsistence farming or fishing
That portion of gross primary production extractable as raw materials	The production of lumber, fuel or fodder
Sources of unique biological materials and products	Medicine, products for material science, genes for resistance to plant pathogens and crop pests, ornamental species (pets and horticultural varieties of plants)
Providing opportunities for recreational activities	Eco-tourism, sport fishing and other outdoor recreational activities
Providing opportunities for cultural and other non-commercial uses	Aesthetic, artistic, educational, spiritual and/or scientific values of ecosystems

Ecosystem services range from regulation of the chemical composition of the atmosphere, to water supply to pollination to recreation. Prugh (1995) categorizes these services into four main areas, regulation, carrier, production, and information services. *Regulation services* are provided, for example by the atmosphere in protecting the Earth's inhabitants from meteorites, and filtering UVB radiation. Ecosystem processes such as atmospheric chemical composition, climate, water, nutrient cycling, soil and wastes are also regulated by nature for life to exist on Earth. *Carrier services* from nature provide the space and a suitable substrate for humans to live and play, and as a habitat for all of the Earth's plants and animals. Nature provides *production services* by supplying the oxygen, water, food, medicines, raw materials and other resources to humans and other living creatures. And *information services* provided by nature include aesthetics, artistic inspiration and scientific education and information.

Table 1 adapted from Costanza *et al.* (1997) lists a variety of ecosystem services.

See also: Ecosystem Services and Costing, Volume 2; Natural Capital, Volume 2.

REFERENCES

Costanza, R, d'Arge, R, de Groot, R, Farber, S, Grasso, M, Hannon, B, Naeem, S, Limburg, K, Paruelo, J, O'Neill, R V, Raskin, R, Sutton, P, and van den Belt, M (1997) The Value of the World's Ecosystem Services and Natural Capital, *Nature*, 387, 253–260.

Prugh, T (1995) *Natural Capital and Human Economic Survival*, ISEE Press, White River Junction, VT.

FURTHER READING

Baskin, Y (1997) *The Work of Nature*, Island Press, Washington, DC, 1–263.

Daily, G C (1997) *Nature's Services: Societal Dependence on Natural Ecosystems*, Island Press, Washington, DC, 1–392.

ADAM FENECH AND ROGER HANSELL *Canada*

Ecosystem Value

see Ecosystem Services and Costing (**Volume 2**)

Ecotheology

see Christianity and the Environment (**Volume 5**); Theology (**Volume 5**)

El Niño and La Niña: Socio-econimic Impacts

see El Niño and La Niña: Causes and Global Consequences (**Volume 1**)

Emergence of Global Environment Change into Politics

see The Emergence of Global Environment Change into Politics (*Opening essay*, **Volume 5**)

Emergy

Emergy is a term developed by Howard T Odum (Odum, 1986, 1996) for use in environmental and economic accounting. Odum has been an important contributor to energetic analysis of ecosystems and has a long history of applying energy analysis to questions of environmental and human well-being (Odum, 1971). Emergy measures both the work of nature and of humans previously required, directly and indirectly, to generate a product or service. While natural ecosystems run on current flows of solar energy moderated by storage systems with fairly rapid turnover, the human economy also draws heavily on very long-term solar storage in the form of fossil hydrocarbons and has begun to tap nuclear energy. Odum does not ask how much energy it took to produce the fossil fuels or nuclear energy, but he does argue that the net energy, the difference between the energy in the resource and the energy required to mine, transport, and process the resource into a usable form, is the important measure. Odum argues that all energy should be converted to solar equivalents, or *solar emergy*, for accounting consistency and because of the dominant role of the sun.

Odum (1971) believes that there is a fourth law of thermodynamics which he has labeled the maximum power principle, or more recently the *maximum empower principle*. In the competition among self-organizing processes, network designs that maximize empower' will prevail (Odum, 1996). Thus, he argues that by selecting choices that maximize emergy production and use, policies and judgments can favor those environmental alternatives that maximize real wealth, the whole economy, and the public benefit (Odum, 1996). Odum and his co-researchers have used emergy analysis to explore specific issues such as the relative desirability of different fuels, to broad questions such as the sustainability of alternative modes of development.

Howard Odum has established a school of energy analysis with a sizeable number of followers, largely trained with Odum, who advocate that emergy analysis should replace, or at least complement, other forms of analysis, especially economic analysis as currently practiced. Critics argue that energy flows do not fully characterize all of the critical aspects of environmental systems, that human values cannot be reduced to energy determinism, and that it is highly unlikely that society will eliminate the diverse institutions through which decisions are currently made, and reduce the power of those interests supporting current institutions, and replace them with institutions for emergy analysis. At the same time, the insights from emergy analysis may prove sufficient to extend existing institutions so as to incorporate yet another approach and reduce somewhat the dominance of economic interests and reasoning.

REFERENCES

Odum, H T (1971) *Environment, Power, and Society*, Wiley, New York.

Odum, H T (1986) *Emergy in Ecosystems. Ecosystem Theory and Application*, ed N Polunin, Wiley, New York.

Odum, H T (1996) *Environmental Accounting: Emergy and Environmental Decision Making*, Chapters 1 and 16, Wiley, New York.

RICHARD B NORGAARD *USA*

Encyclopedias: Compendia of Global Knowledge

Peter Timmerman
University of Toronto, Toronto, Canada

" ••• to collect all the knowledge scattered over the face of the earth, to present its general outlines and structure to the men with whom we live," was Diderot's description of his purpose in the Encyclopedie ou Dictionnaire Raisonne des Sciences, des Arts et des Metiers *(1751- 1772). This encyclopedia was the symbol (and a central driving force) in the French Age of Enlightenment, and had at least two purposes: to wrestle knowledge free from the hierarchical institutions that held onto it as a source of their power and prestige; and to promote the idea that reason could penetrate into every aspect of human activity. Among the most famous articles in the encyclopedia were those on the arts and crafts, including du Quesnay's path-breaking work on the systematic analysis of land as the basis of economic development.*

The growth of encyclopedias in the 18th century (including the earlier *Chambers Cyclopaedia, or Universal Dictionary of Arts and Sciences* (1728), the immediate forerunner of the *Encyclopedie*, and the later British (Scottish) *Encyclopedia Britannica* (1768–1771) fostered by Andrew Bell, can be seen as the expression of a number of different forces at work. First, there was a powerful sense that the emerging sciences of the previous 100 years were beginning to generate substantial amounts of new knowledge that needed to be synthesized for a general reading audience. Second, there was an emerging new middle class in Europe who had – or could be persuaded to have – a need for enlightenment. Third, both Britain and France were now extensively involved in exploration and colonial developments that were beginning to encompass the entire globe.

When Germany became a colonial power, and an aspirant to global power, it too generated its lexicons and the classic Lobel-Brockhaus encyclopedias.

The vast *Soviet Encyclopedia* (1926–1947 and later editions) is a famous example of the symbolic power of encyclopedias, since the creation of the Encyclopedia unfortunately coincided with a whole sequence of ups and downs in official versions of everything, and caused innumerable headaches (and not a few dangers) for compilers and revisers through the years. This was in the grand tradition of Diderot's earlier work, in which a number of entries had to be penned anonymously.

Behind these specific creations, there also lurks the perennial dream of having all knowledge at one's fingertips (and possibly the whole world under one's command). As Collison (1966) notes, it was Francis Bacon (1561–1626) who wanted to re-organize all human knowledge into a 130-part hierarchical study divided into main areas (external nature, Man, and Man's actions upon nature), below which would be sub-areas, and so on. The impulse behind this was, in a sense, to discern or bring order into a world whose previous structural meaning (the medieval Christian world view) had been profoundly shaken.

The potential structure, topics, and outlay of encyclopedias have (since Bacon) been the subject of many tendentious discussions (McArthur, 1986), ranging from whether alphabetizing entries (neutral, handy, but unimaginative) is best, to the potentially infinite possibilities being envisaged (and beginning to be created) by online storage and retrieval systems.

The explosion of knowledge has not been matched, but has certainly been accompanied, by an explosion of specialized encyclopedias in music (Grove), the social sciences (Seligman), and innumerable science encyclopedias, dictionaries, lexicons, etc. Virtually all of them wrestle with the same issues: comprehensiveness versus selectivity; organizational structure; issues of balance; timeliness, accuracy, appropriate assessment.

Global environmental change represents a significant challenge to the encyclopedist, for all the reasons cited above. The impulse to assess that change has perhaps been best captured in the series of books spanning over a century inspired by George Perkins Marsh *The Earth As Modi ed by Human Action* (1885, first published as *Man and Nature*, 1864), most recently *The Earth As Transformed By Human Action* (Turner, 1990), which, if not absolutely encyclopedic, are at least inspired by the possibility of a great global perspective on our situation. Similarly, there are extensive international scientific and social scientific projects whose products could be seen as informal encyclopedias of global environmental knowledge.

There are post-modern voices that consider encyclopedias and similar projects to be significant examples of the total globalizing mentality of the continuing enlightenment project, that is, the use of homogenized and rationalized information as a tool for power. Others – of whom this author is one – rather remember days as a child sitting in a deep armchair with a huge volume of some encyclopedia, diving into distant worlds I might never visit. Encyclopedias are probably still big enough to serve both purposes, and others.

REFERENCES

Collison, R (1966) *Encyclopedias: Their History Throughout the Ages*, Hafner, New York.

McArthur, T (1986) *Worlds of Reference*, Cambridge University Press, Cambridge.

Turner, B L (1990) *The Earth as Transformed by Human Action*, Cambridge University Press, Cambridge.

FURTHER READING

Diderot, D (1967) *The Encyclopedia: Selections*, ed and translator S J Gendzier, Harper and Row, New York.

Enlightenment Project

The *Enlightenment* was an 18th century European philosophy emphasizing reason and individualism rather than tradition. The attempt to apply scientific rationalism throughout human knowledge is sometimes termed the *Enlightenment Project*. This term has come to represent all the ideals associated with the modern period of the 18th to 20th centuries. These ideals form a grand narrative which is optimistic about human history, seeing in it the working out of progress – particularly in terms of increasing material welfare and universal justice – in which all of humanity can share.

This is to be achieved partly through scientific and technological advancement, following Francis Bacon's creed that scientific knowledge equals power over nature, and that this power is to be used for the benefit of Man's estate, freeing all humanity from scarcity and protecting it from nature's caprices. Additionally, the Enlightenment Project would emancipate societies from the irrational yolk of myth, religion and superstition, by deriving moral and legal principles from rational thought alone. Such principles – e.g., the universal declaration of human rights, social justice, sanctity of human life, freedom of speech, equality of opportunity and/or outcome – should be applied everywhere.

This thinking has formed the basis for the spread of Western-style values, including democracy and capitalism, in global modernization. Ostensibly, it champions rational enquiry as the basis for all knowledge, together with the beliefs that only intellectual progress could bring world order and happiness, and that objective science and universal morality and law are attainable as David Harvey says in his 1989 book, *The Condition of Postmodernity*, the Enlightenment Project s axiom was that there was only one possible answer to any question. Postmodernists now challenge this on empirical grounds that the pursuit of grand narratives made the 20th century in particular one of massive and widespread human oppression and suffering.

DAVID PEPPER *UK*

Environment and Violent Conflict

see The Environment and Violent Conflict (*Opening essay*, **Volume 5**)

Environmental Defense Fund

One of the United States rst and most respected environmental groups, *Environmental Defense Fund* (EDF) was founded in 1967 by residents of Long Island, NY, who were pressing for a ban on the pesticide dichlorodiphenyltrichloroethane (DDT). In 1972, EDF members felt that their efforts, and those of others, were rewarded by the decision by the US government to ban DDT. From EDF s small origins, it grew quickly within the United States. (It has no international branches). By 2000, it had over 300 000 members and 170 professional staff members, including more than 75 full time scientists, lawyers and economists. EDF headquarters are in New York City, but it has substantial of ces in Washington, DC, and seven other US cities. Its 1998 budget was $27.8 million.

The organization s hallmark is professionalism. Its members provide nancial support, but are not engaged in governance nor are they active participants in election campaigns. In terms of its organizational niche, EDF represents policy research and the extolling of market mechanisms to nd environmental solutions. As a caricature of organizational culture, if Greenpeace hangs banners from smoke stacks, Earth First! spikes trees, and Sierra Club lobbies congress, EDF is in the boardroom. EDF s Alliance for Environmental Innovation, in collaboration with Pew Charitable Trusts, represents one of the only environmental organizational efforts at direct engagement with private corporations to facilitate higher environmental corporate performance.

EDF has actively promoted such solutions as emissions trading, both to address local air pollution problems as well as a global approach to reduce greenhouse gases. EDF continues to be active on issues of chemical releases and recently targeted the pollution problems created by expanding intensive livestock factory farms.

ELIZABETH MAY *Canada*

Environmental Economics

Environmental economics addresses how environmental problems arise, how market mechanisms can be used to improve environmental quality, and the level of quality that most ef ciently balances preferences for material goods and for environmental services (*see* **Ecological Economics**, Volume 5).

From the perspective of neoclassical economic theory, environmental problems exist because markets for environmental quality are incomplete or totally absent. Markets fail because rights to the services of the environment are unclear or cannot be assigned to individuals and must be held in common. One person cannot obtain cleaner air without everybody obtaining cleaner air, so clean air cannot be sold to individuals and must be obtained collectively. Government action is justi ed by market failure, and environmental economics addresses what governments should do.

Economists argue that markets maximize the difference between the bene ts and costs of providing a good. Environmental economists thus argue that the level of environmental quality should be determined so as to also maximize the differences between the bene ts and the costs. The costs of obtaining environmental quality are typically relatively easy to measure for they are the value of the market goods forgone to install pollution control equipment or take other steps. Environmental valuation is the art and science of estimating the bene ts that do not have market equivalents. Valuation techniques include looking at defensive and corrective expenditures that could be avoided if environmental quality were improved, analyzing how differences in environmental quality affect land values, deducing from individual behavior what people would be willing to pay,

and simply asking people what they would be willing to pay for improved environmental quality. The efficient level of pollution is achieved when the additional costs of clean-up just equal the additional benefits.

The dominant approach is to regulate the level of pollution allowed by each polluter, adjusted for size or output of the production unit. Environmental economists, however, argue that it is more efficient to determine the appropriate level of total pollution, distribute this right among polluters, and then allow polluters to trade pollution rights. With tradable permits, polluters who can reduce pollution more cheaply do so and sell their permits to those whose costs of pollution reduction are high (*see* Tradable Permits for Sulfur Emission Reductions, Volume 4). This results in an efficient reduction of pollution in that the level of environmental quality sought is achieved at the least cost. Alternatively, this same efficient result could be achieved by taxing pollution or subsidizing pollution reduction.

Environmental economics also addresses the harvesting of biological and other renewable resources and the mining of minerals, or non-renewable resources, over time. The rate of interest plays a central role in the efficient rate of use, along with the rate of reproduction or replenishment for renewable resources, and the expected future demand for the resource. For both renewable and non-renewable resources, the pattern of use over time that maximizes net present value is the economically efficient rate of use (*see* Discounting, Volume 5).

FURTHER READING

Chapman, D (2000) *Environmental Economics: Theory, Application, and Policy*, Addison-Wesley, Reading, MA.

RICHARD B NORGAARD *USA*

Environmental Ethics

J Baird Callicott
University of North Texas, Denton, TX, USA

An environmental ethic is a set of behavioral rules, or a set of principles or precepts for governing human behavior toward the natural environment. Environmental ethics may mean more than one such ethic - that is, simply the plural of environmental ethic. More usually it refers to the philosophical subdiscipline devoted to constructing and justifying an environmental ethic. This article is about environmental ethics in both senses. It reviews a number of the most notable environmental ethics that philosophers have recently constructed and tried to justify and thus it surveys the new philosophical subdiscipline devoted to that enterprise.

In response to a popular perception of an environmental crisis in the late 1960s, the two words in the title of this article were rst put together early in the next decade. Environmental ethics (the philosophical subdiscipline) made its debut as a formal area of study with a college course in the subject at the University of Wisconsin, Stevens Point in 1971. It gradually developed and solidi ed in 1979 with the advent of Environmental Ethics, *the journal which has been continuously published ever since. During the 1980s, anthologies, monographs and textbooks in environmental ethics began to appear. During the 1990s, several more journals devoted to environmental ethics began publication and the literature in the eld continues to grow exponentially.*

In retrospect, an earlier informal environmental ethics literature can be identi ed. Doing so is to some extent a matter of philosophical bias and taste but the following authors would probably be recognized as seminal by all authorities: White (1789); Thoreau (1854); Marsh (1864); Muir (1916); Leopold (1949) and Carson (1962). The concept of an environmental crisis was probably seeded in the Zeitgeist by a popular book, The Quiet Crisis, *by Udall (1963), US Secretary of the Interior in the John F Kennedy administration. (The loud crisis was presumably the Cold War threat of nuclear holocaust). Also in uential was* The Closing Circle *by Commoner (1971).*

What happened to precipitate a quiet environmental crisis seems to have been this. In the spirit of beating swords into plowshares, technologies developed for warfare during the 1940s were adapted for peaceful use in the age-old human war against nature. The high combustion air polluting engines developed to power bombers, ghters and tanks powered postwar bulldozers, tractors and harvesters. Dichloro-diphenyl-trichloroethane (DDT), made infamous by Rachel Carson in Silent Spring, was rst developed to delouse allied soldiers and then applied as an agricultural pesticide. Most spectacularly, atomic ssion developed for the war-ending superweapon, was harnessed to produce electricity. Through the 1950s no downsides to such technologies were widely noticed or publicized. After a decade's time lag, however, their detrimental effects began to be evident to the naked eye, nose and ear. Air over major cities became dirty and smelly and pulsed with the ubiquitous noisy throb of gasoline and diesel motors. Rivers became open sewers, murky with municipal and industrial waste. A whole North American Great Lake appeared to be dying. Soil erosion accelerated. Spilled oil fouled beaches. Mankind was poised to realize the long held dream of conquering nature, only to nd that "to the victor belong the spoils" was in the case of this triumph an ironic

proverb. The cartoon character Pogo's malapropism was often quoted: " we have met the enemy and he is us."

This great awakening to an environmental crisis occurred in the most hyper-energetic and hyper-critical decade of the century in which the rhetoric of revolution was so commonplace that it began to be exploited commercially. (Advertisers, for example, urged Americans to join the Dodge Revolution). To address the monumental environmental crisis, nothing short of a revolution in ethics seemed adequate. Historically in the West, ethics appeared to be not just incidentally but militantly anthropocentric. A major pastime of Western philosophers had been to ruminate on what it was about human beings that made us unique, superior to all other creatures and exclusively worthy of moral treatment. In the Judeo Christian tradition of thought, it was our being created in the image of God; in the Greco Roman, it was our innate rationality. By whatever device, Western philosophy had with virtually one voice insisted upon a great gulf separating human nature from that of the lower animals, to say nothing of plants and the abiotic aspects of the environment. A revolutionary environmental ethic would have to de ate human pretensions and rehabilitate a philosophically denigrated nature.

ANTHROPOCENTRISM

Scarcely before the call by Richard Sylvan (Routley, 1973) for a new, environmental ethic had been heard, there appeared a staunch defense of the resources of anthropocentrism and the Western tradition of ethics to address the environmental crisis. Passmore (1974) argued that anthropocentric Western ethics would adequately protect the environment but only if actually implemented, if practiced as well as preached. He saw, furthermore, a dire threat to the achievements of Western civilization – its traditions of rationalism and humanism – in the call to develop a wholly new non-anthropocentric ethic and supporting metaphysics or to borrow one from Asian traditions of thought, as some had suggested (White, 1967).

Strong Anthropocentrism

The anthropocentric strain of environmental ethics has remained vigorous and persistent. A kind of unapologetic, uncritical anthropocentrism – sometimes characterized as strong anthropocentrism – still dominates environmental economics and most of public environmental policy. The environment represents a pool of natural resources or environmental goods, and a sink for wastes or environmental harms. The best allocation of environmental goods and harms is measured in units of human utility, quantified in units of currency such as dollars or pounds sterling and determined by conventional economic calculations (Krutilla and Fisher, 1985). Ethics enters the equation only when goods extracted from the environment by one person or group of persons results in harm suffered by another person or group. For example, if logging steep slopes results in siltation of streams which prevents salmon from successfully reproducing, a strongly anthropocentric environmental ethic might censure the loggers or the timber barons for whom they work, only if commercial salmon fisherman are materially injured. And then the moral matter might be settled by a transfer of funds from the injuring to the injured parties. In their own right, the soil, the salmon and the stream have no moral standing whatever.

The contribution of specialists in environmental ethics to strongly anthropocentric environmental policy is minimal in comparison with that of environmental scientists and economists. Most of the expertise informing such debates is hydrological, geological, biological and economic, etc. The specific contribution of environmental ethicists, under such strongly anthropocentric circumstances is to apply – off the rack, as it were – familiar, well-established ethical theories to the novel environmental situation (Pojman, 1999). The strongly anthropocentric environmental ethicist will, for example, come to an issue prepared to apply the latest model of Kantian ethics, utilitarianism, virtue ethics or rational choice theory generated in the ivory tower by the most celebrated philosophical pooh-bahs of the moment.

Environmental Justice

Because economics is more sensitive to considerations of efficiency than of justice, some strong anthropocentrists argue that environmental economics needs to be augmented by considerations of environmental justice (Wenz, 1988). For example, in the United States a greater percentage of black people than their percentage in the total population live uncomfortably close to a toxic waste dump or incinerator. Although the use of DDT has long been banned in the United States, it is still manufactured there, and exported for use in less developed countries. These and similar unequal distributions of environmental goods and especially harms appear *prima facie* to be unjust. Strongly anthropocentric theories of environmental justice are designed to critique and mitigate instances of environmental injustice.

Sometimes included within the purview of environmental justice is consideration of the welfare or rights of future human generations (if any there be) to natural resources and environmental amenities (Partridge, 1981; Sikora and Barry, 1978). Morally enfranchising future generations requires less application than creative modification of standard Western moral philosophy because new and challenging philosophical conundrums arise. How, for example, can non-existent persons have rights? Contemporary utilitarians define happiness not in terms of pleasure and pain as did classical utilitarians but in terms of preference satisfaction. How, thus, can we know what sort of environment

future human generations may prefer? They may prefer Disney-like simulations to real nature and virtual to actual reality. Furthermore, shifts in environmental policy to take into account the welfare and rights of future human generations might alter the actual composition of future generations. Imposing, say, a stiff energy tax on the present generation as a means to ensure that future generations can have their fair share of fossil fuels might, as one consequence, induce more people of the present generation to stay home at night, turn off the lights and go to bed early. That in turn might change the timing or frequency of the sexual activity of the present generation thus changing which gametes unite to form the embryos of future people. If so, a policy that was designed to benefit the people who would exist in the future, before the policy was adopted, would not benefit them at all – indeed it would have cost them their lives – and would benefit other future people instead. I do not suggest that these problems are fatal to the moral enfranchisement of future generations; I only suggest that they pose a distinctly philosophical challenge to a strongly anthropocentric environmental ethic that includes future human generations. Where future generations are concerned, therefore, environmental policy makers may feel as great a need to consult philosophers as to consult economists.

Weak Anthropocentrism

Weak anthropocentrism typically also morally enfranchises future human generations (Norton, 1984). In addition, it considerably broadens the notion of what counts as a human interest in the environment. People value nature, weak anthropocentrists point out, not only as a pool of material resources (food, fiber, fuel, and so on) but as a pool of psycho-spiritual resources as well. People recreate in unspoiled nature; take aesthetic pleasure and spiritual solace in it; they even take comfort in just knowing that bits of unspoiled nature exist, although they may never directly experience them. Indeed, some weak anthropocentrists suggest that nature might be a moral resource for people (Partridge, 1984).

Resource economists work diligently at the contingent valuation – or shadow pricing – of these natural amenities, because they rarely have an actual market price. People are asked what they would be willing to pay for a clean river in which to canoe or for a view of a mountain range unobscured by smog (Cummings *et al.*, 1986). Also among popular economic methods of calculating the monetary value of such is the travel-cost method in which the money people spend (on gasoline, food, lodging, outdoor gear and so on) to visit unspoiled natural areas is calculated to value (in the economic sense) a national park or designated wilderness area (Peterson and Sorg, 1989).

Unfortunately, an old growth stand of cathedral redwoods, say, will usually be worth far more money as a timber resource than as a psycho-spiritual resource, despite the best efforts of environmental economists to maximize the shadow monetary value of the latter. Weak anthropocentrists, therefore, typically insist that some human values should not be expressed in monetary terms (Norton, 1988). As Kant (1785) once remarked, some things have a price, others a dignity. People are not entirely unanimous in their opinion about what things have a dignity, any more than about what things are worth spending money on. The appropriate place for things with a price to compete is the market; the appropriate place for things with an alleged dignity to compete is the political arena (Sagoff, 1988). People vote with their money in the marketplace – for Rovers, say, but not for Yugos. For those things for which a price is inappropriate, people vote with their choices at the ballot box, albeit usually indirectly. They vote for representatives at various levels of government who, among other things, stand for certain values – for a community free of pornography shops, or for environmental quality: clean air and water, protection of wilderness areas and endangered species, reduction of greenhouse gases, and so on. Further, values settled politically should, ideally, trump market values. For example, in most industrialized societies, the political process has strongly affirmed the value of human emancipation: human beings should not be sold for slaves at any price, no matter what the market will bear. Similarly, weak anthropocentrists insist that humanly valued aspects of nature – designated wilderness areas, national parks, endangered species – should not be for sale at any price (Sagoff, 1988).

In addition to environmental goods and harms, attention has increasingly been directed to free ecological services performed by unspoiled nature, such as soil building and stabilization, water retention and purification, nitrogen fixation, and crop pollination. These too have been valued monetarily by economists (Costanza *et al.*, 1997). The monetary valuation of ecological services has also been criticized from a weak anthropocentric point of view (Sagoff, 1997).

Mention of the Kantian dictum that some things have a price, others a dignity, reminds us that Kant grounded the dignity of persons in their intrinsic value. By parity of reasoning, if some aspects of nature also have dignity, must they have it because they also have intrinsic value? However, anthropocentrism, by definition, weak as well as strong, restricts intrinsic value to human beings and acknowledges only the instrumental value of nature. The difference between strong and weak anthropocentrism may be summed up then as follows. Strong anthropocentrists recognize only the commodity, amenity and service value of nature and express all such values in monetary terms. The economic reduction of all environmental values, also severely short-changes future generations because a Euro presently in hand is worth more than a future Euro, the discount rate roughly corresponding to the current interest rate.

Thus, if an environmental economist were to compare the present monetary value of undeveloped potential commercial real estate in a metropolitan area, on the one hand, to its green-belt amenity value expressed in Euros for people living 100 years hence (estimated by contingent valuation and travel cost), on the other, the future environmental amenity value would be discounted at say 6% per year. Weak anthropocentrists recognize the commodity and service values of nature, but are reluctant to lump the wide range of psycho spiritual resources into the amorphous category of amenities. Most important, weak anthropocentrists oppose the expression of the psycho-spiritual values of nature in monetary terms. Although anthropocentric, all environmental values, in other words, are not preferences. One may prefer chocolate ice cream to strawberry and fish and chips to steak and kidney pie but the way one values Lake Windemere is not of the same order. Therefore, the most basic difference between strong and weak anthropocentrism is that the former treats all values as preferences, which are expressible in monetary terms. The latter treats preferences as a subset of human values and recognizes in addition a suite of environmental values – aesthetic, moral, religious, demeaned by strong anthropocentrists as mere amenities that transcend preferences and that are not appropriately expressed in monetary terms.

NON-ANTHROPOCENTRISM

Non-anthropocentrists seek to provide intrinsic value for non-human natural entities and/or nature as a whole.

Animal liberation and animal rights are not the same as environmental ethics. Indeed, in practice, concern for animal welfare and environmental concerns are often opposed. People for the ethical treatment of animals (PETA), for example, are often in the front lines of resistance to plans to protect endangered species of native plants by exterminating the bloated populations of feral goats, pigs, burros or what have you threatening them. Nevertheless, one approach to non-anthropocentric environmental ethics is based on the theoretical foundations of animal liberation and animal rights.

Contemporary anthropocentric environmental ethicists, strong and weak, typically take anthropocentrism for granted and do not bother to defend the intrinsic (or non-instrumental) value of human beings. Classically, however, anthropocentrists defended the intrinsic value of human beings by positing an intrinsic value conferring property, alleged to be possessed by human beings alone. One such candidate property is the image of God alleged to characterize all and only human beings. Such a property, however, is philosophically maladroit because it is non-empirical. Not on argument, but on faith, one must accept that human beings are in fact created in the image of God, for that image cannot be found by autopsy or inference. For those outside the specific communities of faith in which the image of God is a tenet – the Judaic, Christian and Muslim communities – such a proposition is either meaningless or dubious. Moreover, defending anrthropocentrism by appeal to a non-empirical intrinsic value conferring property, such as the image of God, would dangerously open up a Pandora's box of other, more noxious non-empirical properties that might even more narrowly restrict the class of beings possessing intrinsic value. Racists, for example, might find in the same religious sources a non-empirical property that morally disenfranchises people of color who allegedly bear the mark of Cain. Sexists, similarly, might find in the same religious sources a non-empirical property that morally disenfranchises women, who were created from Adam's rib and who are, thereby, made subservient to men. Thus, modern anthropocentric philosophers have overwhelmingly preferred to follow Kant (1785) and claim that reason is the intrinsic value conferring property possessed by all and only human beings.

The theoretical fulcrum leveraging the intrinsic value of non-human animals is the argument from marginal cases (Regan, 1979). If reason is not to serve as a mere sanitized and secularized substitute for the image of God, it must be an empirically detectable property. A minimally rational being must be able to do certain things – pass certain tests that evidence the possession of this property, such as to perform simple arithmetic calculations, solve simple practical problems, or use language meaningfully. Most human beings can do all such things, but some cannot. These are the marginal cases from which the argument takes its name. They are usually identified as infants, the severely mentally impaired and the abjectly senile. Because they are not rational, they do not have intrinsic value; and because they do not have intrinsic value, they fall beyond the moral pale and may be treated as we treat non-human animals. And how is that? We use animals as subjects for harmful medical experiments and tests, raise them in miserable conditions and slaughter them for food, hunt them for sport, display them in zoos and so on. The tails side (pun intended) of the argument is that many animals that conventional Western ethical theory and practice morally disenfranchise are minimally rational, according to these objective measures. Non-human primates have been observed to make and use tools in order to solve simple problems, have learned to use American sign language (to say nothing of the complex, but so far undeciphered social vocalizations of whales, dolphins, and elephants), to scheme and plan, and so on.

Animal Liberation

But back to the heads side of the argument from marginal cases coin. If we are to morally enfranchise the marginal cases, we must identify a capacity that the marginal cases possess as the intrinsic value conferring property. Bentham

(1789), a contemporary of Kant, argued that the capacity to experience pleasure and pain, called *sentience*, ought to be that property. Also, sentience is arguably more relevant to the bene t that it confers than reason. For why are we concerned how others treat us except that we can suffer? But many kinds of animals are also sentient. They too, therefore, possess intrinsic value and should thus be morally enfranchised equally.

Animal Rights

As Bentham knew full well, however, it is perfectly possible to raise animals comfortably and slaughter them painlessly. If we also hide their fate from them, so that they do not dread it, we can have our moral cake (or meat), so to speak, and eat it too. For that reason, Regan (1983) has proffered an alternative intrinsic value-conferring property being the subject of a life, that is, being a self with a capacity to enjoy a rich subjectivity shared by us and those animals that people most care about (mammals) as the conceptual foundation of animal rights. To be a subject of a life, according to Regan, implies, among other things, having desires, preferences, beliefs, memory of the past, anticipation of the future, and a sense of personal identity through time. In Regan s opinion all mammals one year old or more are subjects of a life, so understood, and thus should also enjoy the basic rights of life, liberty, and the pursuit of happiness.

Subjects of a life represent only a tiny fraction of the natural environment. Many more animals are sentient than are subjects of a life. But even if we stipulate that sentience is the proper intrinsic value conferring property and that all animals are sentient including those without a central nervous system animals, nevertheless, still represent only a small part of the environment. Hence, even if all animals were granted full moral standing, only a small part of the environment would be morally enfranchised. Therefore, some environmental philosophers, concerned not exclusively or even primarily about animals, but about the environment more generally, have tried to build a more inclusive non-anthropocentric ethic on the theoretical platform of animal liberation and animal rights theory.

I shall turn to that construction project in a moment. Before leaving the present matter, I should note that some partisans of animal liberation/animal rights argue that animal-oriented ethics, *per se*, may double dip as environmental ethics. At rock bottom, the argument runs parallel to the argument that strong anthropocentrism may serve as an environmental ethic (Jamieson, 1998). Wild animals can survive and ourish only in appropriate habitat. To ensure the welfare or protect the rights of animals would, therefore it seems, require us to preserve their habitat. We could call this a strong zoocentric environmental ethic. Undeveloped, the environment is instrumentally valuable to a wide variety of animals, as well as to us human beings as a pool of natural resources. The fatal problem with strong zoocentrism, as it seems to me, emerges when we consider that different kinds of animals have different habitat requirements. Thus preserving the habitat of one kind of animal will evict another kind. And if, as Singer (1977) reasons, all animals are equal, we have no theoretical justi cation for being biased in favor of one kind of animal instead of another. How thus can we say that habitat for endemic, native salmon spawning is to be preserved at the cost of displacing habitat for cosmopolitan, exotic carp; or that woody habitat for white tailed deer is to be preserved at the cost of displacing grassy habitat for domestic cattle?

BIOCENTRISM

Just as animal liberation and rights theorists challenge reason as the intrinsic value conferring property, so biocentric environmental ethicists challenge sentience or being a subject of a life as the intrinsic value conferring properties. The capacity to experience pleasure and pain, to say nothing of the suite of capacities comprehended by the phrase *subject of a life*, did not come into existence gratuitously or for their own sakes. Rather, sentience and most other capacities of consciousness evolved to foster the preservation of the organisms that possess them (Goodpaster, 1978). What is of ultimate value to organisms is to live and ourish. Pain and pleasure are powerful motivators toward that end, rewarding life-fostering behaviors and punishing life-threatening ones. The ability to remember the past and imagine the future enables organisms to learn from experience and to anticipate eventualities, which enhance their ability to survive and ourish. But if life itself is the be all and end all of organic existence, then maybe we should think that all living beings have intrinsic value.

The rst biocentrist seems to have been Schweitzer (1923) who insisted upon a reverence for life. He was a professional physician and musician, but only an amateur philosopher. A close study of Schweitzer s rhetoric reveals that he was primarily in uenced by Schopenhauer (1818), a neo Kantian philosopher who argued that the unknowable *Ding an sich* (thing in itself), which Kant believed to lurk in all phenomena, was the will to live. The will to live is one and the same in all things, inanimate as well as animate, according to Schopenhauer. Schweitzer, however, seems not to have appreciated the full sweep and grandeur of Schopenhauer s philosophical vision and himself thought the will to live inhabited only living beings. Therefore, just as we human beings experience it vividly and poignantly when our lives are threatened, so too do all other kinds of living beings. A biocentric ethic, Schweitzer believed, could be based on the capacity that we cognitively advanced living beings have for empathy and compassion with all

the others, who are, in essence, identically the same as ourselves.

Such a frankly metaphysical grounding of intrinsic value in nature is now quite out of philosophical fashion, though Schweitzer remains popular. Independently of Schweitzer, Feinberg (1974) suggested that having interests is what entitles a being to basic moral rights and that interests are compounded of *conations*, which consist of drives, aims, goals, latent tendencies, directions of growth and natural fulfillments – whether conscious or not. Conative plants and other living beings, although lacking consciousness of any kind, may nevertheless have interests, in this sense, and thus intrinsic value, and moral rights. Feinberg's suggestion was first taken up by Goodpaster (1978), then developed by Taylor (1986) into a full fledged biocentric theory of environmental ethics. Although Taylor carefully distinguishes between inherent worth and intrinsic value, this and other such conceptual niceties are too fine grained for presentation here. The following must suffice. Parallel to Regan's notion of a subject of a life, Taylor argues that all organisms are teleological centers of life. The operative term, here, *teleological* has long been in the philosophical lexicon and is formed from the Greek *telos*, meaning end, goal, or purpose. All organisms have innate drives, aims, goals – *teloi* in a word – which can be realized or frustrated. The most basic ones are to grow, to flourish, and to reproduce. Therefore, all organisms have inherent worth (intrinsic value) and all have it equally. Taylor avoids going so far as to claim that all teleological centers of life have rights, but he does argue that they are owed respect. A moral agent should respect all organisms and leave them alone to realize their own ends.

How then can a moral agent realize his or her own *teloi*? One must eat, be clothed, sheltered, educated, cured, all at the expense of other teleological centers of life. Taylor (1986) attempts to resolve morally the inevitable conflicts of interest arising between equally valuable teleological centers of life. He begins with the right of each person to self-defense. Taylor argues that when, for example, a disease-causing bacterium invades one's body, one is entitled to defend oneself with lethal antibiotics. Then to this unambiguous paradigm of self defense he assimilates the more problematic cases – taking the life of another organism to feed oneself, to clothe and shelter oneself, and so on.

ECOCENTRISM

Taylor's arguments justifying human exploitation of other teleological centers of life seem inconsistent with the severe constraints of his biocentric theory of environmental ethics, and are therefore not very convincing. But even if they were, biocentrism is beside the point. Environmentalists are simply not concerned about the welfare of individual shrubs, bugs and grubs. The medicinal eradication of millions of disease-causing bacteria in the intestines of a human being is not an environmental concern – except if the improvident use of antibiotics breeds resistant strains of bacteria; and then the concern would be less distinctly environmental than one of public health. Nor are environmentalists morally concerned about killing trees in a forest for the purpose of improving the overall health of the forest, or about culling excess deer on an overpopulated range. Environmental concerns typically focus on levels of biological organization beyond individual plants and animals – populations, species, communities, ecosystems, biomes, the biosphere – and on non-organic aspects of the environment, such as rivers, lakes, bays, gulfs, seas, oceans, soils and the atmosphere. Environmentalists are morally concerned about the extirpation of populations of certain species, such as the Bengal tiger from the Indian sub continent, about the threatened global extinction of species, such as the blue whale, about the integrity of biotic communities, especially temperate prairies and tropical rainforests that are so rich in variety of species, about the health of ecosystems, such as the Oostvaardersplassen or Chesapeake Bay, about anthropogenic acid deposition, atmospheric pollution, global warming and such.

To develop a relevant environmental ethic, one that directly addresses actual environmental concerns, we must step outside what may be called the modern classical paradigm of ethics. Goodpaster (1979) identified the common argument form of the two dominant schools of thought in modern ethics – one going back to Bentham (1789), the other to Kant (1785) – as a generalization of egoism. I (*ego*) demand that you consider my welfare or respect my interests when your actions affect me. I justify that demand by appeal to some capacity I possess – rationality, sentience, having a rich inner life or consciousness, having interests – which makes me intrinsically valuable and therefore entitles me to moral considerability. By the same token, however, I am obliged to consider the welfare or respect the rights of all those who possess the same capacity. This argument form underlies anthropocentrism, animal liberation/animal rights, and biocentrism. They differ only in the choice of the intrinsic value-conferring capacity and thus in the set of morally considerable entities they recognize. In the case of anthropocentrism, the intrinsic value-conferring property is reason; in that of animal liberation, it is capacity to suffer; in animal rights, experiencing a rich subjective life; in biocentrism, having interests. At the end of the day, however, the modern classical paradigm leaves us with only individuals – individual human beings, individual animals, individual organisms being morally considerable. Environmental concerns, however, are almost always holistic. Thus, we must go back to the theoretical drawing board and start afresh.

THE LAND ETHIC

Of all the environmental ethics so far devised, the land ethic, first sketched by Leopold (1949), is most popular among environmentalists. In his essay, *The Land Ethic*, Leopold begins by alluding to Darwin's account of the origin and development of ethics in the *Descent of Man*. The Leopold land ethic is thus grounded in evolutionary biology therefore making it more consonant than any alternative so far devised with the sciences that inform contemporary environmental concern (Callicott, 1989).

The existence of ethics presents a problem for Darwin's attempt to show how all things human can be understood as having gradually evolved by natural (and sexual) selection, from traits possessed by closely related species, his project in *The Descent of Man*. An ethic demands that moral agents selflessly consider the interests of others in addition to their own. The theory of evolution would seem to predict, however, that the selfish would out-compete the selfless in the struggle for existence, and thus survive and reproduce in greater numbers. Therefore, greater and greater selfishness, not selflessness, would seem to be selected for in any population of organisms, including those ancestral to *Homo sapiens*. But history indicates the opposite: that our remote human ancestors were more callous, brutal, and ruthless than we are. At least so it seemed to a refined English gentleman who while serving as naturalist on the round the world voyage of the HMS Beagle had observed first hand what he and his contemporaries regarded as states of savagery and barbarism similar to those from which European and Asian civilizations were believed to have emerged. Without a convincing evolutionary explanation of its existence and progressive development, Darwin's pious opponents might point to ethics among human beings as a clear signature by the hand of providence on the human soul.

To the conundrum presented by the existence and allegedly progressive development of ethics, Darwin's resolution is straightforward and elegant. For many kinds of animals and especially for *Homo sapiens*, life's struggle is more efficiently prosecuted collectively and cooperatively than singly and competitively. Poorly armed by nature, as solitaries, hominids would fall easy prey to their natural enemies or starve for lack of the wherewithal to obtain food. Together our primate ancestors might stand some chance of fending off predators and attacking prey larger than themselves. Like many other similarly situated species, evolving human beings thus formed primitive societies; or put more precisely, those hominids that formed primitive societies evolved. But without some rudimentary ethics, human societies cannot stay integrated. As Darwin (1871) puts it,

> No tribe could hold together if murder, robbery, treachery, & c., were common; consequently such crimes within the limits of the same tribe are branded with everlasting infamy'; but excite no such sentiment beyond these limits.

Darwin's speculative reconstruction of the evolutionary pathway to ethics begins with altruistic parental and filial affections which motivate parents (perhaps only the female parent in many species) to care for their offspring and their offspring to desire the company of their parents. Such affectionally bonded nuclear families are small and often ephemeral societies, lasting, as in the case of bears, only until the next reproductive cycle. But the survival advantage to the young of being reared in such social units is obvious. Should the parental and filial affections chance to spill beyond the parental-filial relationship to that between siblings, cousins and other close kin, such plurally bonded animals might stick together in more stable and permanent groups and defend themselves and forage communally and cooperatively. In which case there might also accrue additional advantages to the members of such groups in the struggle for life. Thus do mammalian societies originate in Darwin's account.

By themselves, the social impulses and sentiments are not ethics. An ethic is a set of behavioral rules, or a set of principles or precepts for governing behavior. The moral sentiments are, rather, the foundations of ethics, as Hume (1751) and Smith (1759) argued, a century or so before Darwin considered the matter. In addition to the social sentiments and instincts, *Homo sapiens* evolved to a high degree of intelligence and imagination and uniquely possessed a symbolic language. Hence, we human beings are capable of generally representing those kinds of behavior which are destructive of society (murder, robbery, treachery, & c.) and articulating prohibitions of them in emotionally colored formulae – thou shalt not kill, steal, bear false witness, etc. – which formerly we called commandments and today call moral rules.

So much then for the origin of ethics. Darwin goes on to account for the development of ethics. As human social groups competed with one another for resources, the larger and better organized out-competed the smaller and less well organized. Hence clans, firstly, merged into tribes; tribes, next, into nations; and nations, eventually, into republics. The emergence of each of these levels of social organization was attended by a corresponding expansion of ethics. Darwin (1871) sums up this parallel growth of ethics and society as follows: As man advances in civilization, and small tribes are united into larger communities, the simplest reason would tell each individual that he ought to extend his social instincts and sympathies to all the members of the same nation, though personally unknown to him. This point being once reached there is only an artificial barrier to prevent his sympathies extending to the men of all nations and races. Further, with the emergence of each new level in the social hierarchy – the clan, the tribe, the nation, the republic, the global village – the content of the moral code changed or was supplemented to reflect and facilitate the novel structure of each newly emerged level.

During Darwin's lifetime, a universal ethic of human rights was only dimly visible on the horizon. By the mid-twentieth century, when Leopold was gestating the land ethic, a universal human rights ethic may have seemed more nearly attainable. In any case, Leopold, often called a prophet, looked farther ahead than did Darwin himself, indeed farther ahead than Darwin could have looked in the absence of a well developed ecological world view. Leopold (1949) summarizes Darwin's natural history of ethics with characteristic compression: All ethics so far evolved rest upon a single premise: that the individual is a member of a community of interdependent parts. Then he adds an ecological element, the community model of the biota espoused most notably by Elton (1927). Ecology simply enlarges the boundaries of the community to include soils, waters, plants and animals, or collectively: the land (Leopold, 1949). When we all learn to see land as a community to which we belong not as a commodity belonging to us (Leopold, 1949) that same simplest reason, of which Darwin speaks, might kick in. And when it does, what results will be a land ethic that changes the role of *Homo sapiens* from conqueror of the land community to plain member and citizen of it (Leopold, 1949).

Basically, what Leopold did to cook up the land ethic was to take over Darwin's recipe for the origin and development of ethics, and add an ecological ingredient, the Eltonian community concept. According to Leopold (1949), a land ethic implies respect for fellow members and also for the community as such. The land ethic, in other words, has a holistic dimension to it that is completely foreign to the mainstream modern moral theories going back to Bentham and Kant. The holistic dimension of the land ethic – respect for the community as such, in addition to respect for its members severally – is, however, not in the least foreign to the Darwinian and Humean theories of ethics upon which it is built. Darwin (1871) could hardly be more specific or emphatic on this point. Actions are regarded by savages and were probably so regarded by primeval man, as good or bad, solely as they obviously affect the welfare of the tribe, – not that of the species, nor that of an individual member of the tribe. This conclusion agrees well with the belief that the so called moral sense is aboriginally derived from the social instincts, for both relate at first exclusively to the community. Hume (1751) insists that we must renounce the theory which accounts for every moral sentiment by the principle of self love. We must adopt a more public affection and allow that the interests of society are not, even on their own account, entirely indifferent to us.

That is not to say that Hume, certainly, and even Darwin make no theoretical provision for a lively concern for the individual members of society as well as for society *per se*. Darwin (1871) writes of the all important emotion of sympathy. Sympathy means *with feeling* and so can extend only to feeling – that is, sentient – individuals. It cannot extend to levels of social organization beyond the individual, which have no feelings *per se*. Hume and Darwin, however, recognized other moral sentiments than sympathy, some of which – patriotism, for example – relate as exclusively and specifically to societies as sympathy does to sentient individuals. In the Leopold land ethic, in any event, the holistic aspect eventually eclipses the individualistic aspect. Toward the beginning of *The Land Ethic*, Leopold (1949), as noted, declares that a land ethic implies respect for fellow members of the biotic community, as well as for the community as such. Toward the middle of *The Land Ethic*, Leopold (1949) speaks of a biotic right to continue but such a right accrues, as the context indicates, to species, not to individual specimens. Toward the end of the essay, Leopold (1949) writes a summary moral maxim, a golden rule, for the land ethic: A thing is right when it tends to preserve the integrity, stability and beauty of the biotic community. It is wrong when it tends otherwise. In it, there is no reference at all to fellow members. They have gradually dropped out of account as the *The Land Ethic* proceeds to its climax. Because of its focus on the biotic community as such, the Leopold land ethic may be called ecocentric.

Ultimately traceable, as it is, to the theory of moral sentiments developed in the eighteenth century by David Hume and Adam Smith, the several objects of moral respect recognized in the ecocentric land ethic, do not so much have intrinsic value as they are valued intrinsically. The difference is not merely semantic. Intrinsic value, as conceived by Kant and his contemporary intellectual descendants, is what philosophers now call a *supervenient property* – that is, an objective property that piggybacks on another objective property. For Kant, as noted, that latter property is reason, for Singer it is sentience, for Regan it is being the subject of a life, and for Taylor, being a teleological center of a life. For the contemporary intellectual descendants of Hume and Smith, however, value is first and foremost a verb, not a noun that designates an objective (albeit supervenient) property. Something is, in other words, valuable only if it is valued by some valuing subject. And we human valuing subjects value things in at least two different ways: instrumentally and intrinsically. One values one's house, automobile and other useful artifacts instrumentally. One values one's self, spouse, children, other relatives, friends, neighbors, fellow citizens, fellow man intrinsically. One also values intrinsically one's family, municipality, country and the human race, *per se* – that is, as collective wholes. One may value intrinsically a wide array of other things as well. Indeed, one can value anything under the sun intrinsically – the Eiffel Tower, for example, or even a worn-out old shoe. But we usually do not value any and everything intrinsically; we usually do so for a reason. A

fundamental reason is a sense of belonging to a community of interdependent parts.

The Problem of Ecofascism and the Land Ethic

Holism is the land ethic's principal strength, because actual environmental concerns are holistic, but also its principal liability. According to Leopold (1949), we *Homo sapiens* are but plain members and citizens of the biotic community. Then it would seem that the summary moral maxim of the ecocentric land ethic applies to *Homo sapiens* no less than to the other members and citizens of the biotic community. A human population of more than six billion individuals is a dire threat to the integrity, stability, and beauty of the biotic community. Thus the existence of such a large human population is ethically wrong. To right that wrong should we not do what we do when a population of white tailed deer or some other species irrupts into and threatens the integrity, stability, and beauty of the biotic community? We immediately and summarily reduce it, by whatever means necessary, usually by randomly and indiscriminately culling the members of such a population – respectfully, of course – until its numbers are optimized.

It did not take its critics long to draw out the apparently vitiating implication of the ecocentric land ethic. According to them, the land ethic is a case of environmental fascism because the good of ecological wholes takes precedence over the good of the component individuals, including human individuals, according to Aiken (1984), Regan (1983), Ferre (1996), and Shrader-Frechette (1996). If so, it must be rejected as monstrous. Happily, it does not. To think that it does, one must assume that Leopold proffered the land ethic as a substitute for, not an addition to, our venerable and familiar human ethics. But he did not. Leopold refers to the various stages of ethical development – from tribal mores to universal human rights and finally, to the land ethic – as accretions. Accretion means an increase by external addition or accumulation. The ecocentric land ethic is an addition to our several accumulated social ethics – now more than ever, strongly individualistic – not something that is supposed to replace them. With the advent of each new stage in the accreting development of ethics, the older stages are not erased or replaced. I, for example, am a citizen of a republic, but I also remain a member of an extended family, and a resident of a municipality. And it is quite evident to us all, from our own moral experience, that the duties attendant on citizenship in a republic (to pay taxes, to serve in the armed forces or in the Peace Corps, for example) do not cancel or replace the duties attendant on membership in a family (to honor parents, to love and educate children, for example) or residence in a municipality (to support public schools, to attend town meetings). Similarly, the duties attendant upon citizenship in the biotic community (to preserve its integrity, stability, and beauty) do not cancel or replace the duties attendant on membership in the human global village (to respect human rights).

As members of multiple communities, each generating ethical obligations, how do we prioritize when duties to one community or its members conflict with duties to another? Naess (1989) suggested two complementary priority principles for resolving such conflicts. The first we might call the inverse scale principle (ISP). One should give priority to obligations to one's nearest and dearest community members and one's nearest and dearest community as such – to one's family, municipality, etc., in that order. Hence, if in a time of general scarcity, one faced a choice between sharing one's severely limited resources with family members or with unrelated neighbors, the right thing to do would be to share with family members and refuse neighbors. The second we might call the degree-of-interest principle (DIP), which counters the ISP. If the interests at stake are of unequal strength, the lesser interest should give way to the greater. For example, in circumstances not of general scarcity but of unequal distribution of plentiful resources, some are needier than others. If one is relatively wealthy and one's neighbor is desperately poor, then the strong interest in having the wherewithal for survival of one's neighbor should take priority over the weaker interest of satisfying the superfluous consumer preferences of oneself and one's family members.

Applying these priority principles to moral conflicts arising between obligations relative to a biotic community and those relative to a municipality, we might conclude that if our choice were between the destruction of a regional biotic community (and the extinction of its endemic species), on the one hand, and the starvation of the citizens of local human communities, on the other, then priority should be given to our obligations relative to the municipality, in accordance with the ISP. However, in the famous case of the conflict between the endangered spotted owl and the threatened old growth forest in the Pacific Northwest of the United States, on the one hand, and the region's logging/milling communities, on the other, that is not the conundrum we face. Other employment alternatives are available to loggers and mill workers. The biotic community interests at stake are much greater – the destruction of a unique biome and the extinction of its endemic species. Hence in this case, the DIP countermands the ISP. Unfortunately, in some parts of the world, the ethical dilemma does involve the very livelihood, not just the lifestyles of local peoples. For example, some people face the choice of feeding themselves and their families and by doing so contributing to the destruction of tropical forests and a global mass extinction event, on the one hand, or starving, on the other. For them the choice is clear and unambiguous. They must slash and burn. Because they are not my neighbors or fellow nationals, however, it is not entirely

clear what national and international policies, with respect to such practices, I should support.

Change and the Land Ethic

Leopold penned the land ethic at mid century. Ecology then represented nature as tending toward a static equilibrium, and portrayed disturbance and perturbation, especially those caused by *Homo sapiens*, to be abnormal and destructive. In view of the shift in contemporary ecology to a more dynamic paradigm (Botkin, 1990), and in recognition of the incorporation of natural disturbance to patch and landscape scale ecological dynamics (Pickett and Ostfeld, 1995), we might wonder whether the land ethic has become obsolete. Has the paradigm shift from the balance of nature to the ux of nature in ecology invalidated the land ethic? I think not, but recent developments in ecology may require revising the land ethic. From Leopold s original summary moral maxim or golden rule of the land ethic, we may be left now with only the preservation of a biotic community s beauty as a criterion of a thing s land-ethical rightness or wrongness.

Despite all the idle talk about beauty being in the eye of the beholder, ecological beauty might not be an altogether inadequate or relativistic criterion by means of which to judge a thing s environmental rightness or wrongness. Hume (1751) notes that broad shoulders, a lank belly, rm joints, a taper leg; all these are beautiful in our species, because signs of force and vigor. Natural beauty is far less relative and ckle a judgment of taste than we are wont latterly to acknowledge; it is, as it were, an aesthetic registry of robust good health. And indeed Leopold (1999) linked ecological beauty to land health. Land was healthy and therefore beautiful to those with a re ned taste in natural objects in Leopold s estimation, when its ecological processes were being carried out normally. Such ecological processes include photosynthesis; energy transfer through lengthy food chains and tangled food webs, capped by long lived, large bodied carnivores; the recruitment, retention, and recycling of nutrients; the stabilization of soils; the modulation of water ows; and the resistance to and rapid recovery from perturbation by wind, re, ood, pest, pathogen, and other disturbances.

Leopold was, moreover, aware of and sensitive to natural change. He knew that conservation must aim at a moving target. How can we conserve a biota that is dynamic, ever changing, when the very words *conserve* and *preserve* especially when linked to *integrity* and *stability* connote arresting change? The key to solving that conundrum is the concept of scale. Scale is a general ecological concept that includes rate as well as scope; that is, the concept of scale is both temporal and spatial. And a review of Leopold s *The Land Ethic* reveals that he had the key, though he may not have been aware of just how multiscalar is change in nature. There, Leopold (1949) writes, Evolutionary changes ••• are usually slow and local. Man s invention of tools has enabled him to make changes of unprecedented violence, rapidity, and scope. Leopold was keenly aware that nature is dynamic, but, under the sway of mid-century equilibrium ecology, he conceived of natural change primarily in evolutionary, not in ecological terms. Nevertheless, scale is equally normative when ecological change is added to evolutionary change, that is, when normal climatic oscillations and patch dynamics are added to normal rates of extinction, hybridization, and speciation.

As noted, *Homo sapiens* is, in Leopold s opinion, a part of nature, a plain member and citizen of the land community. Hence, anthropogenic changes imposed on nature are no less natural than any other. Nevertheless, because *Homo sapiens* is a moral species, capable of ethical deliberation and conscientious choice, and evolutionary kinship and biotic community membership add a land ethic to our familiar social ethics, anthropogenic changes may be land-ethically evaluated. But by what norm? The norm of appropriate scale.

Let me rst recount Leopold s use of the temporal scale of evolutionary change as a norm for evaluating anthropogenic change. Consider the current episode of abrupt, anthropogenic, mass species extinction which many environmentalists intuitively regard as the most morally reprehensible environmental thing going on today. Episodes of mass extinction have occurred in the past, though none of those has been attributed to a biological agent. Such events are, however, abnormal. Normally, speciation outpaces extinction which is the reason why biological diversity has increased over time. So, what is land-ethically wrong with current anthropogenic species extinction? Species extinction is not unnatural. On the contrary, species extinction anthropogenic or otherwise is perfectly natural. But the current rate of extinction is wildly abnormal. Does being the rst biological agent of a geologically signi cant mass extinction event in the 3.5 billion years tenure of life on Planet Earth morally atter us *Homo sapiens*? Doesn t that make a mockery of the self-congratulatory species epithet: the sapient, the wise species of the genus *Homo*?

Now let s apply this model to a quandry that Leopold himself never considered. Earth s climate has warmed up and cooled off in the past. So, what s land-ethically wrong with the present episode of anthropogenic global warming? We are a part of nature, so our recent habit of recycling sequestered carbon may be biologically unique, but it is not unnatural. A land-ethical evaluation of the current episode of anthropogenic climate change can, however, be made on the basis of temporal scale and magnitude. We may be causing a big increase of temperature at an unprecedented rate. That s what s land-ethically wrong with anthropogenic global warming.

Temporal and spatial scale in combination are key to the evaluation of direct human ecological impact. Long before *Homo sapiens* evolved, violent disturbances regularly occurred in nature. And they still occur, quite independently of human agency. Volcanoes bury the biota of whole mountains with lava and ash. Tornadoes rip through the countryside, leveling houses and trees. Hurricanes erode beaches. Lightning sets fires that sweep through forests and savannas. Rivers drown floodplains. Droughts dry up lakes and streams. Why, therefore, are analogous anthropogenic disturbances – clear cuts, beach developments, hydroelectric impoundments, and the like – environmentally unethical? As such, they are not. Once again, it's a question of scale. In general, frequent, intense disturbances, such as tornadoes, occur at small, widely distributed spatial scales, while spatially more extensive disturbances, such as droughts, occur less frequently. And most disturbances at whatever level of intensity and scale are, mathematically speaking, stochastic (random) and chaotic (unpredictable). The problem with anthropogenic disturbances – such as industrial forestry and agriculture, suburban development, drift net fishing – is that they are far more frequent, widespread, and regularly occurring than are non-anthropogenic disturbances. They are well out of the spatial and temporal range of disturbances experienced by ecosystems over evolutionary time.

Proponents of the new flux of nature paradigm in ecology agree that appropriate scale is the operative norm for ethically appraising anthropogenic ecological perturbations. For example, Pickett and Ostfeld (1995) note that the flux of nature is a dangerous metaphor. The metaphor and the underlying ecological paradigm may suggest to the thoughtless and greedy that since flux is a fundamental part of the natural world, any human caused flux is justifiable. Such an inference is wrong because the flux in the natural world has severe limits ••• two characteristics of human-induced flux would suggest that it would be excessive: fast rate and large spatial extent.

Among the abnormally frequent and widespread anthropogenic perturbations that Leopold (1949) himself censures in *The Land Ethic* are the continent-wide elimination of large predators from biotic communities in Europe and North America; the ubiquitous substitution of domestic species for wild ones; the ecological homogenization of the planet resulting from the anthropogenic worldwide pooling of faunas and floras; the ubiquitous polluting of waters or obstructing them with dams.

The summary moral maxim of the land ethic, however, must be rendered more dynamic in light of developments in ecology over the past quarter century. Leopold acknowledges the existence and land-ethical significance of natural environmental change, but seems to have thought of it primarily on a very slow evolutionary temporal scale. Even so, he thereby incorporates the concept of inherent environmental change and the crucial norm of scale into the land ethic. In light of more recent developments in ecology, we can add norms of scale to the land ethic for both climatic and ecological dynamics in land-ethically evaluating anthropogenic changes in nature. One hesitates to edit Leopold's elegant prose, but as a stab at formulating a dynamized summary moral maxim for the land ethic, I will hazard the following: A thing is right when it tends to disturb the biotic community only at normal spatial and temporal scales, so as not to mar its beauty (or health). It is wrong when it tends otherwise.

SUMMARY

In conclusion, then, there are two fundamental genera of environmental ethics, anthropocentrism and non-anthropocentrism. Of the former there are two species, strong and weak. Strong anthropocentrism reduces all environmental values to preferences capable of economic expression and valuation in monetary terms. All environmental decisions can be made by weighing benefits against costs, tempered by considerations of environmental justice. Weak anthropocentrism agrees that all environmental values are instrumental, but insists that some are not preferences. Human environmental values that conflict with preferences are more properly institutionalized politically than made to compete economically.

There are several species of non-anthropocentrism. Two, *animal liberation* and *animal rights*, are not, properly, environmental ethics; and indeed, are often at odds with environmental concerns and values. A third, *biocentrism* – the proposition that all living beings possess intrinsic value and merit respect – builds upon the theory of animal rights. Though proffered as an environmental ethic, biocentrism is hardly relevant because it morally enfranchises only individuals, while actual environmental concerns are almost always holistic.

Ecocentrism, a fourth species of non-anthropocentrism, is holistic in focus. In its original articulation by Aldo Leopold, the integrity, stability, and beauty of the biotic community is the standard of right environmental conduct. Allegations to the contrary notwithstanding, ecocentrism does not entail environmental fascism. For, as Leopold conceives it, a *land ethic* is an addition to, not a substitute for, our more individualistically inclined social ethics and all other things being equal, our more venerable social ethics take precedence over the land ethic when they conflict. Contemporary ecology, however, posits a more dynamic than stable nature and more loosely associated than tightly integrated biotic communities, existing at multiple scales, subject to a wide variety of periodic disturbances. A more dynamic ecocentrism would morally limit anthropogenic change and disturbance to normal temporal and spatial scales, thus to preserve the land's health and beauty.

REFERENCES

Aiken, W (1984) Ethical Issues in Agriculture, in *Earthbound: New Introductory Essays in Environmental Ethics*, ed T Regan, Random House, New York, 274–288.

Bentham, J (1789) *An Introduction to the Principles of Morals and Legislation*, Clarendon Press, Oxford.

Botkin, D (1990) *Discordant Harmonies: A New Ecology for the Twenty First Century*, Oxford University Press, Oxford.

Callicott, J B (1989) *In Defense of the Land Ethic*, State University of New York Press, Albany, NY.

Carson, R (1962) *Silent Spring*, Houghton Mifflin, Boston, MA.

Commoner, B (1971) *The Closing Circle*, Knopf, New York.

Costanza, R, d'Arge, R, de Groot, R, Farber, S, Grasso, M, Hannon, B, Naeem, S, Limburg, K, Paruelo, J, O'Neill, R V, Raskin, R, Sutton, P, and van den Belt, M (1997) The Value of the World's Ecosystem Services, *Nature*, 387, 253–260.

Cummings, R, Broockshire, D, and Schultz, W (1986) *Valuing Environmental Goods: a State of the Art Assessment of the Contingent Valuation Method*, Rowman and Allenheld, Totowa, NJ.

Darwin, C R (1871) *The Descent of Man and Selection in Relation to Sex*, J Murray, London.

Elton, C (1927) *Animal Ecology*, Sidgwick and Jackson, London.

Feinberg, J (1974) The Rights of Animals and Unborn Generations, in *Philosophy and Environmental Crisis*, ed W Blackstone, University of Georgia Press, Athens, 43–68.

Ferre, F (1996) Persons in Nature: Toward an Applicable and Unified Environmental Ethics, *Ethics and the Environment*, Vol. 1, 15–25.

Goodpaster, K E (1978) On Being Morally Considerable, *J. Philos.*, 75, 308–325.

Goodpaster, K E (1979) *From Egoism to Environmentalism. Ethics and Problems of the Twenty- rst century*, eds K E Goodpaster and K E Sayre, University of Notre Dame Press, IN, 21–35.

Hume, D (1751) *An Enquiry Concerning the Principles of Morals*, Clarendon Press, Oxford.

Jamieson, D (1998) Animal Liberation is an Environmental Ethic, *Environ. Values*, 7, 41–57.

Kant, I (1785) *Grundlegung zur Metaphysik der Sitten*, Johann Friedrich Hartknoch, Riga.

Krutilla, J and Fisher, A (1985) *The Economics of Natural Environments. Studies in the Valuation of Commodity and Amenity Resources*, Resources for the Future, Washington, DC.

Leopold, A (1949) *A Sand County Almanac and Sketches Here and There*, Oxford University Press, New York.

Leopold, A (1999) *For the Health of the Land. Previously Unpublished Essays and Other Writings*, eds J B Callicott and E T Freyfogle, Island Press, Washington, DC.

Marsh, G P (1864) *Man and Nature*, Charles Scribner, New York.

Muir, J (1916) *A Thousand Mile Walk to the Gulf*, Houghton Mifflin, Boston, MA.

Naess, A (1989) *Ecology, Community, Life Style*, translated by D Rothenberg, Cambridge University Press, Cambridge.

Norton, B G (1984) Environmental Ethics and Weak Anthropocentrism, *Environm. Ethics*, 6, 131–148.

Norton, B G (1988) Commodity, Amenity, and Morality: The Limits of Quantification in Valuing Biodiversity, in *Biodiversity*, ed E O Wilson, National Academy Press, Washington, DC, 200–205.

Norton, B G (1996) Environmental Ethics and Weak Anthropocentrism, *Environ. Ethics*, 6, 131–148.

Partridge, E P (1981) *Responsibilities to Future Generations*, Prometheus, Buffalo, NY.

Partridge, E P (1984) Nature as a Moral Resource, *Environ. Ethics*, 6, 101–130.

Passmore, J (1974) *Man's Responsibility for Nature*, Charles Scribner's Sons, New York.

Peterson, G and Sorg, C, eds (1989) *Toward a Measurement of Total Value*, USDA Forest Service, Rocky Mountain Forest and Range Experiment Station General Technical Report, Fort Collins, CO.

Pickett, S T A and Ostfeld, R S (1995) The Shifting Paradigm in Ecology, in *The Ecology of Natural Disturbance and Patch Dynamics*, eds R L Knight and S F Bates, Academic Press, Orlando, FL, 261–278.

Pojman, L P (1999) *Global Environmental Ethics*, Mayfield, Mountain View, CA.

Regan, T (1979) An Examination and Defense of One Argument Concerning Animal Rights, *Inquiry*, 22, 189–219.

Regan, T (1983) *The Case for Animal Rights*, Berkeley University of California Press, CA.

Routley, R (1973) Is There a Need for a New, an Environmental Ethic, in *Proceedings of the 15th World Congress of Philosophy*, Vol. 1, Bulgarian Organizing Committee, Sophia Press, Varna, 205–210.

Sagoff, M (1988) *The Economy of the Earth*, Cambridge University Press, Cambridge.

Sagoff, M (1997) Can We Put a Price on Nature's Services, *Report from the Institute of Philosophy and Public Policy*, 17(3), 7–12.

Schopenhauer, A (1818) *Die Welt als Wille und Vorstellung*, Hesse and Becker, Leipzig.

Schweitzer, A (1923) *Civilization and Ethics*, translated by A Naish, Black, London.

Shrader-Frechette, K S (1996) Individualism, Holism, and Environmental ethics, *Ethics Environ.*, 1, 55–69.

Singer, P (1977) *Animal Liberation: A New Ethic for Our Treatment of Animals*, Avon, New York.

Sikora, R I and Barry, B (1978) *Obligations to Future Generations*, Temple University Press, PA.

Smith, A (1759) *Theory of the Moral Sentiments*, eds A Millar, A Kinkaid, and J Bell, Edinburgh.

Taylor, P W (1986) *The Ethics of Respect for Nature*, Princeton University Press, Princeton, NJ.

Thoreau, H D (1854) *Walden*, Tichnor and Fields, Boston, MA.

Udall, S (1963) *The Quiet Crisis*, Holt, Rinehart and Winston, New York.

Wenz, P (1988) *Environmental Justice*, State University of New York, Press, Albany, NY.

White, G (1789) *The Natural History and Antiquities of Selborne in the County of Southhampton*, reprinted by Macmillan (1900), London.

White, L (1967) The Historical Roots of Our Ecologic Crisis, *Science*, 155, 1203–1207.

FURTHER READING

Callicott, J B (1999) *Beyond the Land Ethic*, State University of New York Press, Albany, NY.

Environmental Movement – the Rise of Non-government Organizations (NGOs)

Elizabeth May
Sierra Club of Canada, Ottawa, Canada

Although active in promoting environmental policies since the 19th century, environmental NGOs became a more profound force in the latter decades of the 20th century.

Environmental groups operate at local, community, provincial/state, national and international levels. Generally, they exist based on public nancial support and involve millions of citizens around the world. Typically and culturally, NGOs are more robust in democratic societies, but emerge as countries transit from totalitarian to democratic governance.

The rise of NGOs has corresponded to a declining role of the nation state in an era of increased global corporate rule. In response, environmental NGOs are themselves increasingly globalized and acting in signi cant ways on international issues, from United Nations (UN) fora to the World Trade Organization (WTO).

The role of NGOs has been accepted by the UN organizations with increasing openness through the preparatory process for the United Nations Conference on Environment and Development (UNCED), and the June, 1992 Rio Earth Summit. Resistance to NGO participation in other institutions, such as the World Bank and the WTO, has spilled over into the streets of Washington, DC for a number of years and in Seattle in November 1999 at the third ministerial meeting of the WTO.

NGOs, as a signi cant force within civil society, will likely continue to occupy an increasing area of political space as the challenges of global environmental change weaken existing institutions.

Reformers of all kinds have long organized themselves into groups and associations to press for change. Successful movements put themselves out of business, such as the slavery abolitionists and the women's suffrage movement. The first successful mass mobilization campaign can be traced back to the UK abolitionists' campaign in the 18th century. NGOs became an accepted part of society dating from the Franco–Prussian war and the creation of the Red Cross.

Although the term *environment* did not come into common usage until the 1960s, concern for nature, clean air and water and protection of animals led to the establishment of the earliest environmental groups toward the end of the 19th century.

A case can be made for the inclusion of public health activists as early environmentalists. Florence Nightingale, for example, bemoaned that the Thames was used as an open sewer and worked to improve standards of public health through improved sanitation:

> It did strike me as odd sometimes, that we should pray to be delivered from plague, pestilence, and famine,' when all the common sewers ran into the Thames and fevers haunted undrained land, and the districts which cholera would visit could be pointed out. I thought that before cholera came that we might remove these causes, not pray God would remove the cholera.
>
> *Suggestions for Thought* (1860)

The killing fogs of London's smoke choked air also led to public protest and movements for change focused on the deleterious effects on public health. Such concerns had been voiced for hundreds of years, but most notably grabbed attention in 1952 when 4000 people died in an acute air pollution event in London. Similarly, the development of British common law remedies for nuisance and riparian rights were intimately tied to protection of one's right to clean air and clean water. Still, the existence of organized environmental groups is seen as a more recent phenomena.

THE NATURE LOVERS

In North America, early environmental groups were organized as hiking clubs, such as the Williamstown Club (1863), the White Mountain Club (1873) and the Appalachian Mountain Club (1876). In 1875, people concerned with the state of the nation's forests organized the American Forestry Association. Unlike its current incarnation, it initially championed forest protection, urged tree planting, lobbied for parks and successfully established Arbor Day.

As the industrial revolution and expanding populations in the Eastern United States pressed against shrinking wilderness, these societies expressed the romantic longing for wildness, influenced by Henry David Thoreau and his 1854 book *Walden* (*see* Thoreau, Henry David, Volume 5).

In this period, both the Audubon Society and the Sierra Club were formed. The Audubon Society and the American Ornithological Society were formed in the 1880s, with a primary mission of the study and protection of wild birds. The Sierra Club (*see* Sierra Club, Volume 5) was founded in 1892 with the broader goal of protecting large tracts of wilderness in the western US. Although the word *environment* was not used at the time, protection of wildness and the urgent imperative to protect nature from ruthless exploitation were common themes in the group's early literature. Founder John Muir (*see* Muir, John, Volume 5) wrote of government attitudes to American forests:

> (the government) is like a rich and foolish spendthrift who has inherited a magnificent estate in perfect order, and then has left his fields and meadows, forests and parks, to be sold and plundered and wasted at will, depending on their inexhaustible abundance.
>
> (Turner, 1985)

The Sierra Club was founded with the goal of introducing urban dwellers to the magnificence of the Sierra Nevada. Muir believed that by experiencing and enjoying the natural world, people would, through transformative and transcendental experience, become converts to the cause of preserving it. This formula proved very successful. Prior to the establishment of the Sierra Club, Muir himself helped persuade outdoorsman, adventurer and hunter US President Teddy Roosevelt to protect critical wilderness (*see* Roosevelt, Theodore, Volume 5). The first national park had been Yellowstone in 1872, but public reactions had been adverse.

In a strategy that foreshadowed Sierra Club's tactics for the next century, Muir plotted a campaign to make Yosemite a national park. The plan included an insider Washington strategy with a Yosemite bill the locus of lobbying, a media strategy with articles in popular magazines extolling the virtues of Yosemite and the urgency of protecting it from development, penned by the now famous Muir, and mobilizing individuals by taking concerned people into the Yosemite wilderness. The Washington and New York lobbying was primarily the work of Muir's friend and collaborator Robert Underwood Johnson, whose role as associate editor of the nation's leading literary monthly, *The Century*, proved crucial in mobilizing public support. The Yosemite bill passed into law in 1890.

But the wilderness of Yosemite was incompletely protected with key areas, particularly the Hetch Hetchy Valley, remaining in state, rather than federal, control. Moreover, the magnificent sequoia forests north of the park were unprotected. These and other wilderness campaigns became the focus for the Sierra Club.

In the spring of 1903 Muir accomplished the ultimate in lobbying coups; a camping trip into Yosemite with President Teddy Roosevelt. By the end of the trip, Teddy Roosevelt declared it would be a shame to our civilization to let them (sequoias) disappear. They are monuments unto themselves , (Turner, 1985, 826). It took several years more effort to get the bill expanding protection of Yosemite to include the Hetch Hetchy Valley through the California legislature, but ultimately, Muir and the Sierra Club prevailed. But the victory was not to be permanent.

THE CONSERVATION MOVEMENT

Across North America another species of environmentalism was on the rise. If Muir and Sierra Club could be described as *preservationist*, the movement for scientific forestry and wise use of natural resources was christened *conservationist*. The leading practitioners and advocates for conservation were Gifford Pinchot and Bernard Fernow. Pinchot was the son of a wealthy Connecticut family, who studied forestry in Germany and became the first dean of the Yale School of Forestry, created through a large Pinchot family donation. German born and trained, Fernow was already established as a forester when Pinchot came on the scene. Pinchot replaced Fernow as the director of the Division of Forestry within the Department of Agriculture and Fernow became the first dean of forestry at the University of Toronto.

Both believed in the efficient use of forests and favored human use and benefits over the love of and protection of wilderness in its own right. Pinchot and Fernow established forestry principles that persist to this day, favoring even-aged forest stands and the elimination of old growth. They championed sustained yield forest management.

Initially, Muir and Pinchot were in league. In 1895, Muir even wrote:

> The forests must be, and will be, not only preserved, but used; and ••• like perennial fountains ••• be made to yield a sure harvest of timber, while at the same time all their far reaching (aesthetic and spiritual) uses may be maintained unimpaired.

But the unity of preservationists and conservationists was not to last. Two bitter disputes permanently split the American conservation movement. The first was the question of the purpose and level of protection for National Forests. Pinchot and his followers believed that scientific forestry and logging should take place in these forests, while Muir and the Sierra Club had succeeded in having 21 million acres (eight million hectares) of forest set aside in 1897 as forest reserves with no mention of a commercial purpose.

Although Muir originally tried to keep both forces unified in order to ensure the success of the cause of forest reserves, the tension between preservation and wise use ultimately snapped coalition efforts. Pinchot's forces succeeded in having Congress pass a Forest Management Act (1897) which stated that the central purpose of reserves was to furnish a continuous supply of timber for the use and necessities of citizens of the United States . The Act also included mining and grazing as acceptable activities within the forest reserve system.

Pinchot's colleague Bernard Fernow summed up the conservationist view, the main service, the principal object of the forest has nothing to do with beauty or pleasure. It is not except incidentally an object of aesthetics, but an object of economics .

The conversion of national forest reserves to logging concessions was a blow to Muir, but it was nothing compared to the heart-breaking loss in the fight for the Hetch Hetchy.

Muir and the Sierra Club had succeeded in having the dramatic Hetch Hetchy Valley added to the Yosemite National Park. One hundred and fifty miles from the growing urban center of San Francisco, the Yosemite wilderness was seen through two very different lenses. To Muir, in wildness lay spiritual salvation:

> thousands of tired, nerve shaken, over civilized people are beginning to nd out that going to the mountains is going home; that wildness is a necessity; and that mountain parks and reservations are useful not only as fountains of timber and irrigating rivers, but as fountains of life.
>
> (Nash, 1967)

For others, the Hetch Hetchy Valley held another form of salvation, an abundant source of water for a parched city. The proposed damming of the Hetch Hetchy was sacrilege to Muir, but sensible to the wise-use conservation movement. As Pinchot wrote:

> As to my attitude regarding the proposed use of the Hetch Hetchy by the city of San Francisco ••• I am fully persuaded that ••• the injury ••• by substituting a lake for the present swampy oor of the valley ••• is altogether unimportant compared with the bene ts to be derived from its use as a reservoir.

San Francisco s engineers had proposed damming the Hetch Hetchy in 1882, long before its inclusion in the Yosemite National Park. In 1890, even after its status as a protected area within a national park was established, the Mayor of the day requested permission to dam the river and use the water. The Secretary of the Interior, Ethan A Hitchcock refused, noting that a dam was inconsistent with the national park designation.

But in 1906, San Francisco suffered the double blow of a devastating earthquake and re. The city s need for water aroused public sympathy far more than the preservationists love of wilderness. In 1908, Hitchcock s successor, James R Gar eld, approved the plan. The aftermath was an historic point in the growth of environmental groups as Muir, Johnson and the Sierra Club launched the rst nationwide environmental protest campaign in US history. The Hetch Hetchy became a national issue.

The ght was bitter and the rhetoric brutal. Muir, who had long written of wilderness in terms of religious reverence, now wrote of its despoilers as Satanic. Johnson also took up the cause of protection of beauty and nature against the materialistic followers of Mammon in the context of a larger morality play. While Muir and Johnson urged that other sources be found for the city s water needs, even sympathetic President Teddy Roosevelt was convinced that there was no other suitable source for water.

In 1912, Muir called those promoting the dam scheme:

> temple destroyers, devotees of ravaging commercialism ••• having a perfect contempt for nature ••• instead of lifting their eyes to the God of the Mountains, lift them to the Almighty Dollar. ••• Dam Hetch Hetchy! As well, dam for water tanks the people s cathedrals and churches, for no holier temple has ever been consecrated in the heart of man.

In what may have been most painful for Muir, dam supporters included friends and residents of San Francisco within the Sierra Club. Warren Olney, one of the Club s founders, led a Club faction in support of the dam. Muir and Johnson acted through the California Branch of the Society for the Preservation of National Parks, to avoid the internal con ict. In 1910, through a referendum within the Sierra Club, Muir established that the Club also favored preservation over damming the Hetch Hetchy, by a vote of 589 161. Still, having to ght within the organization he had founded was a bitter pill. The San Francisco media lampooned Muir as a cranky old fanatic who cared more for trees than people. And there was no doubt his rhetoric had become more fanatical. But nationally, Muir and the preservationists had widespread support, with articles in the *New York Times* and many popular magazines. The rift within the conservation movement was profound. In 1913, the head of the National Conservation Association, Harry Slatterly, wrote to the Secretary of the Interior, Unfortunately, our good friends the nature lovers are still unreasonable in their attitude. There is grave danger they will again be able to block this most necessary legislation (Hays, 1972).

Muir and the Sierra Club organized a broader coalition, pulling away groups, such as women s clubs, traditionally supportive of Gifford Pinchot and the utilitarian conservationists to the cause of Hetch Hetchy. The Hetch Hetchy issue galvanized a new wilderness preservation movement, with effective support from the women s clubs, hunter and sportsmen s groups, outing clubs, scienti c societies and academic institutions. One Senator noted that he had received 5000 letters urging him to save the Hetch Hetchy. National support was strong and Muir and his supporters felt the protests would succeed in thwarting the bill to dam the valley. In November 1913, Muir wrote to Johnson, we re bound to win, enemy badly frightened, Up and smite em!

But the quiet and effective lobbying of in uential San Franciscans had done wonders for the city that wanted water. On December 6, 1913 the Senate vote favored damming the Hetch Hetchy, and despite best efforts, no presidential veto followed.

Muir and the preservationists accepted their defeat on the Hetch Hetchy, nding solace, as Muir wrote, in the fact that the conscience of a whole country has been aroused from sleep . It was no exaggeration to say that the Hetch Hetchy controversy had awakened the nation.

The wilderness movement in the United States continued after Muir s death in December 1914. The appreciation for the spiritual aspect of wilderness was most notably

continued by Aldo Leopold (*see* **Leopold, Aldo**, Volume 5). In 1908 Leopold graduated from Yale School of Forestry, pursuing Gifford Pinchot's philosophy, only to reject his utilitarian view of nature. In summarizing the American approach to nature, he wrote ••• a stump was our symbol of progress . Although Leopold became president of the Ecological Society of America (founded in 1915), his largest contribution to the environmental movement was through his writing. His most influential legacy, *A Sand County Almanac and Sketches from Here and There* was published in 1949.

WILDLIFE PROTECTION MOVEMENT

While the environmental movement took root in North America through wilderness campaigns, the movement also expanded in Europe through the latter part of the 19th century. Lacking vast tracts of wilderness, the focus of nature lovers was the protection of animals. As early as 1824, the first organization concerned with non-human life was established in England by philosopher Jeremy Bentham. The British Society for the Prevention of Cruelty to Animals championed the protection of domesticated animals. It was not until 1903 that the fate of wild species found an organizational champion, the Society for the Preservation of the Wild Fauna of the Empire.

In 1900, a treaty was negotiated by the foreign ministers of colonial powers controlling the continent of Africa and signed in London the Convention for the Preservation of Animals, Birds and Fish in Africa. *The Times* of London's coverage of the conference included a decidedly environmental message

> The advance of civilization, with its noise and agitation, is fatally disturbing to the primitive forms of life. Commerce, moreover, discovers continually some new demand for the trophies of the chase. The horns, the skins, and the plumage of the beasts and the birds have an increasing market value. It is not surprising, therefore, that men of science have become alarmed at the prospect of the extinction of the most interesting and characteristic types of zoological development.
>
> (Bonner, 1993)

But it was not only scientists who worried about the perilous fate of Africa's wildlife. The colonial dukes and lords and princes had come to regard the hunting of African game as their private domain. In 1903, the colonial big game hunters founded the Society for the Preservation of the Wild Fauna of the Empire. In the media, they were dubbed penitent butchers. These conservationists were not discouraged from killing lions under the terms of the 1900 treaty. The treaty actually called for protection only of those species which are either useful to man or are harmless . Hunting lions, leopards, hyaenas and wild dogs was consistent with the treaty. Conservationist US President Teddy Roosevelt was one of the big game hunters hosted by British residents of East Africa. In a remark that highlights the different strands of the fabric of the environmental movement, when hiking through Yosemite, Muir had chastised Roosevelt for his love of hunting: Mr Roosevelt, when are you going to get beyond the boyishness of killing things ••• are you not getting far enough along to leave that off?

The hunter conservationist remains an important part of the modern environmental movement, found in groups such as the National Wildlife Federation in the US, its related Canadian Wildlife Federation, and Ducks Unlimited. Although now quite removed from hunting, the original founders of the World Wildlife Fund (WWF), which came into being in the 1960s, had come from similar origins.

After the devastation of European human environments, Briton E M Max Nicholson worked to establish the first international conservation organization with goals beyond Africa. The International Union for the Conservation of Nature (IUCN) came into existence in 1948, through the work of Nicholson and others (*see* **IUCN (The World Conservation Union)**, Volume 4). Staffed by scientists, it established a reputation for credible and authoritative conservation research. Working with government and non-government members, IUCN (now known as the World Conservation Union) had a unique ability to influence government policy. Still, it lacked the wherewithal to raise funds for its program. To meet this need, the WWF was founded.

Public sentiment had been moved through the influential writings of Sir Julian Huxley, scientist and first head of the United Nations Educational, Scientific and Cultural Organization (UNESCO). Huxley wrote a three-piece series in *The Observer* in 1960 describing his travels though Africa, the magnificent wild creatures and inexorable push toward their destruction. It moved Nicholson and other committed entrepreneurs to organize a fund-raising and campaigning arm for IUCN. IUCN and WWF were in partnership until WWF broke away to do its own work in the mid-1980s. (In the 1990s, the WWF officially changed its name to the Worldwide Fund for Nature.) The early connections between wildlife appreciation and royalty are maintained by WWF through the active roles of Their Royal Highnesses, Prince Philip and Prince Bernhard of the Netherlands.

Key elements of an international environmental movement were in place by the middle of the 20th century, with factions concerned with wilderness preservation, conservation of natural resources, including wildlife for sustained use, and wildlife protection, fighting to keep species from going the way of the dodo. The movement had extended itself through much of the globe, although headquarters and impetus for action largely came from the source of the damage; the industrialized world.

THE MODERN ENVIRONMENTAL MOVEMENT

Most writers date the modern environmental movement's birth as 1963 when Rachel Carson's classic *Silent Spring* was published (*see* next section and Carson, Rachel Louise, Volume 5). It is no doubt true that a new type of organization was required to tackle a new host of environmental threats.

The by-products of warfare were changing the very nature of the environment. World War I gave the world chlorine compounds; World War II yielded chlorinated hydrocarbons, organophosphates, and increased radiation from the new technological ability to split the atom, for armaments and electricity. Pollution from industrial enterprises was compromising regional air quality and killing life in aquatic systems. The new chemicals brought new health risks. While the chlorine compounds and other toxic chemicals found commercial application as pesticides, the arms race led to global contamination with fallout from atmospheric testing and increased radiation from the nuclear power industry. It is clear that a prime motivation for widespread public protests against atmospheric nuclear weapons testing was the impact on public health, particularly on children's health, around the world.

A new type of international organization was born out of the movement to stop atmospheric testing. Although the campaign is usually characterized as part of the peace movement, it had all the hallmarks of environmental concerns. Nuclear testing was distributing tons of toxic radionuclides, including strontium-90, all around the world. Strontium-90's pathways mimic those of calcium. Taken up in grasses, consumed by dairy cows, the radiation made its way to milk. Once consumed, it was stored in bones and teeth, increasing the risk of childhood leukemias and later adult cancers.

In response, the world saw its first truly global grassroots movement in the late 1950s. Many of the groups were established by mothers concerned about nuclear weapons testing and the threat to their children. Although national groups of many different names were established, the largest umbrella, the Campaign for Nuclear Disarmament (CND) organized the key global events. The organization was aided by the prominence of one of its founders, Lord Bertrand Russell. The ubiquitous nature of its incorporation into popular culture can be seen to this day in the use of its symbol as the peace sign. The circle with a semaphore was the symbol for the letters CND.

In the US, the campaign against atmospheric nuclear testing was organized through the Committee for a Sane Nuclear Policy, generally abbreviated to SANE. Key board members included the leading liberal thinkers of the day: editor of the *Saturday Review* Norman Cousins, repeat candidate for president, Norman Thomas, and the author of the best-selling book, *Baby and Child Care*, Dr Benjamin Spock.

The Ban-the-Bomb movement held large marches, launched petition drives, and enlisted celebrities and Nobel laureates in its campaign. In the UK, an annual Easter-time march built public support and momentum. The march started in 1956 when a single Quaker walked from London to a military research facility at Aldermaston, a distance of 54 miles (87 km). The following year, 100 people marched. By 1959, there were 3000 marchers. But, the most celebrated and successful march was in 1960 when over 100 000 marchers made the journey towards London from Aldermaston, culminating with a giant rally at Trafalgar Square. Marchers came from national groups from around the world, with committees from Ghana, South Africa, Greece, Israel, Cyprus, the US, Canada, Ireland and France. The line of the march stretched a distance of ten miles (16 km), Marchers were applauded by supporters along the sidelines, and Musicians played along the route. An enormous volunteer effort orchestrated lodging en route, mostly on school and church floors, and canteen wagons served three meals a day. The youngest marcher, Martin Dowling, was seven weeks old. The eldest, in a wheel chair, was nearly 100. The march was led by Canon Collins, the canon of St. Paul's Cathedral and by Jacquetta Hawks, historian, wife of J B Priestley. Many Labour members of parliament marched as well. The Aldermaston March of 1960 was a major media event, covered by television and radio news around the world. If Woodstock had happened first, it would have been dubbed the Woodstock of the Ban the Bomb movement.

The grassroots political movement was rewarded when in September 1963, the governments of the United States, the Soviet Union and the UK entered into the Nuclear Test Ban Treaty. The treaty banned the testing by detonation of nuclear weapons in the atmosphere. While the treaty is often characterized as part of the developing movement for non-proliferation and peace, it spoke most immediately to ending the widespread dispersion of nuclear contamination through the atmosphere.

The Test Ban Treaty was not the first multilateral agreement to deal with an environmental concern. The early part of the 20th century also saw a number of binational treaties between Canada and the United States; first the 1909 Boundary Waters Convention, which created the International Joint Commission, and the Migratory Birds Convention. But the Test Ban Treaty was arguably the first created in global fashion and resulting from the efforts of NGOs.

SILENT SPRING

The modern environmental movement is usually traced to the publication of *Silent Spring* by Rachel Carson (Carson,

1963) (*see* Carson, Rachel Louise, Volume 5). *Silent Spring* chronicled the health and environmental threats of the escalating and widespread use of chemical synthetic pesticides. The book was both well written and well researched and became a bestseller. Despite the vociferous efforts of the chemical lobby to attack the credibility of Rachel Carson, she became a hero to many Americans. Her warnings led to the creation of new citizen groups working to have certain pesticides banned. Efforts in Long Island, NY to ban dichlorodiphenyltrichloroethane (DDT) led to the creation of the Environmental Defence Fund (*see* Environmental Defense Fund, Volume 5). Traditional groups, such as the Sierra Club, now under the leadership of its first executive director, David Brower (*see* Brower, David, Volume 5), broadened the organization's mandate to tackle pollution and pesticides.

Through the late 1960s, a huge increase in environmental awareness burgeoned in the US. In some ways, it could have been seen as an extension of the movement against the war in Vietnam. But within the political debate taking place in the US, the new environmental movement was often accused by the anti-war movement of being safe, white, and middle class,' detracting from the tougher issue of predominantly black US soldiers drafted to fight an unpopular war in Southeast Asia. Still, both movements grew in strength.

In the fall of 1969, the US government planned to conduct an underground nuclear test in Amchitka, Alaska, and suddenly a new protest movement was sweeping the US to stop the test. Scientists feared that the blast might actually destabilize an area with significant fault lines, with the potential for a catastrophic earthquake. The US Secretary of Defense took his whole family to Alaska to demonstrate his faith in the safety of the nuclear test, but opposition mounted. US President Richard Nixon was besieged with telegrams and letters urging that the test be cancelled. The grassroots opposition also came from north of the border, where a small group in Vancouver organized themselves as Greenpeace to oppose the tests. The test proceeded, but the movement for environmental protection was growing.

In 1970, *Life* magazine ran a cover story, Lake Erie is Dead. The Cuyahoga River in Ohio had caught fire from the chemicals that had been dumped there by local industries. The emblem of the US crest, the mighty bald eagle, was near extinction due to the pernicious effects of DDT as it moved through the food chain. The reproductive success of the bald eagle, as top predator, was dangerously compromised.

The growing movement was harnessed by Denis Hayes and US Senator Gaylord Nelson in a national teach-in called Earth Day (*see* Earth Day, Volume 5). April 22nd, 1970 galvanized a new and modern movement for environmental protection. Earth Day was officially recognized by Presidential Proclamation from US President Richard Nixon. Large corporations seeking public favor, from Coca Cola to Proctor and Gamble, endorsed the event. Tens of thousands of people participated from coast to coast, with limited activities in Canada as well.

Meanwhile, the environment movement was increasing in Europe as well. In the UK, the International Institute for Environment and Development (IIED) was formed and championed by the visionary Barbara Ward (*see* Ward, Barbara, Volume 4). David Brower, having left Sierra Club, founded Friends of the Earth (FoE) in 1969 (*see* Friends of the Earth, Volume 5) and affiliate FoE groups were formed in Europe and elsewhere. Greenpeace, also, was forming national groups around the world. As well, local grassroots efforts for environmental protection were being organized in many countries.

STOCKHOLM

The gathering strength of public concern for the environment was recognized by the UN in the first Conference on the Human Environment, in 1972 in Stockholm. The UN conference is remembered for its adoption of the Stockholm Principles for Environment as well as for the North–South split. Brazil boycotted the conference, believing that the industrialized world's environmental agenda was code for stalling development in the developing world. Yet, the conference was, by any measure, a success.

Secretary general for the conference, Canadian Maurice Strong, immediately sensed the need to include the environmental movement. Under UN terminology, such citizens' groups were called NGOs. The organizing of a parallel NGO conference for Stockholm was a first of its kind. Maurice Strong also commissioned a popular non-official book for the conference. *Only One Earth: The Care and Maintenance of a Small Planet*, by Ward and Dubos (1972), contributed to a growing global environmental consciousness. Surprisingly, it became a bestseller and was printed in twelve languages. Speaking to the NGO conference at Stockholm, US anthropologist Margaret Mead, in what would become a motto for environmental groups for generations to come, said Never doubt that a small group of committed people can change the world. Indeed, it is the only thing that ever has .

The years leading to the UN Summit in Stockholm and those immediately following saw the mainstreaming of environmental concerns. Governments at all levels created departments of the environment. New pollution control legislation was brought forward in countries around the world. In the US, legislation included the National Environmental Protection Act which established environmental assessment, the Clean Air Act, the Clean Water Act and the Endangered Species Act.

The first major wave of environmental activism receded in response to the apparent acceptance by government and corporations of the demands of NGOs. Alternatively, the oil shock of the early 1970s has been credited with dampening demands for environmental protection. Still, the oil crisis instilled an ethic of energy conservation, if briefly, in the psyche of Americans and Canadians. Major auto manufacturers moved to produce more energy efficient cars. Amory Lovins promoted soft energy paths and demand-side energy management. The energy conservation movement led to the creation of a new type of environmental NGO, one concerned with technological innovation for conservation and renewable energy sources.

Environmental groups grew around the world in response to increased pressure on natural spaces and growing pollution. In Australia, the Conservation Society and other groups arose to stop the flooding of Lake Pedder, protect the Great Barrier Reef, the forests of Tazmania as well as to oppose uranium mining. Penan and Dayak indigenous people in Borneo organized with Malaysian environmental groups to call for protection of the rainforest, while in Brazil, rubber tappers, known as *seringueiros*, organized for economic rights as well as protection of the Amazonian rain forest. In December 1988, seringueiro leader, Chico Mendes, was murdered just outside his Xapuri home, despite the presence of state-provided bodyguards. In India, women organized to protect forests through the Chipko movement (*see* Chipko Movement, Volume 5), while in Kenya, Wangaari Mathai organized the planting of trees through the Greenbelt Movement. Even within the totalitarian system of the USSR, voices of environmental activists were heard.

New environmental threats through the next few decades created focused campaigns for change. In the mid-1970s, the thinning of the ozone layer led first to consumer campaigns to avoid those cosmetic products using ozone-depleting clorofluorocarbons (CFCs). As consumers, primarily in the US and Canada, stopped buying products containing CFCs, the manufacturers stopped using them, hailing their commitment to environmental protection. It was not for another decade that citizens realized CFC use had continued and expanded through use in other products, from air conditioning to blown foam insulation. Governments responded to powerful environmental NGO campaigns in global negotiations leading to the Montreal Protocol in 1989. Friends of the Earth International (FoEI) was particularly associated with this campaign as was the Washington, DC, based Ozone Action. By that time, the role of NGOs within high level negotiations had grown. NGO representatives were included within the delegations of many industrialized country governments. NGOs spoke from the floor in plenary, as well as dominating the streets in demands for change.

BRUNDTLAND TO RIO

In the last years of the 1980s, the second major wave of public environmental activism was cresting. It happened to coincide with the hearings of the UN sponsored the World Commission for Environment and Development (WCED or more commonly known as the Brundtland Commission). The WCED was chaired by the leader of the Norwegian parliamentary opposition, Dr Gro Harlem Brundtland. But, when the government in power fell, newly anointed Prime Minister Brundtland did the unexpected: she remained chair of the commission. Other commission members included the former secretary general of the Stockholm Conference Maurice Strong, Ambassador Mohamed Sahnoun of Algeria, Guyana's Shridath Ramphal, the secretary general of the Commonwealth of Nations and the Deputy Prime Minister Mansour Khalid of the Sudan, to name a few of the 22 members.

Unlike other groupings of prestigious and powerful personages pulled together in major commissions, the Brundtland Commission held public hearings around the world. In Indonesia, Brazil, Norway, Zimbabwe, Kenya, the USSR, Canada, Japan, they sought out the views of indigenous peoples, peasant farmers, and NGOs of all shapes and sizes. Hundreds of individuals and groups presented their views on environment and development to the Commissioners. In the process, the role of NGOs and the political space they occupied was expanded. The WCED report, *Our Common Future*, was released in 1987, many of its pages reflect a recognition of the role of NGOs as important catalyst for the emerging concept of sustainable development.

One of the key recommendations of the Brundtland Commission was for a major global conference on the linked problems of environment, development and militarism. The WCED recommended that a UN Summit be convened in 1992, marking 20 years since Stockholm. That advice led to acceptance by the UN General Assembly of a global conference focusing on environment and development. Militarism was dropped from what the Brundtland Commission termed a three-legged stool. The UNCED was held in June 1992 in Rio de Janeiro, Brazil; thus improving prospects for bridging the North–South divide that led to Brazil's 1972 Stockholm boycott.

The Rio process expanded the role of NGOs in multinational issues. The negotiations at the first Preparatory Committee (Prep Com) meeting, held in Nairobi, Kenya in August 1990 were delayed for several days until the issue of the scope of NGO participation rights could be resolved. Hard-line opposition to opening the process of negotiations nearly stymied the effort for greater transparency. But, in the end, the issue was resolved in favor of previously unknown opportunities for NGO contribution to global policy development. NGOs, whether those traditionally accepted within the UN Economic and Social

Council (ECOSOC) bureaucracy or not, would be permitted to speak in plenary at the discretion of the chair and attend most working group negotiations. The advent of electronic communication, via the internet, engaged many more NGOs around the world. Experimental efforts at early Prep Coms evolved by the time of the Earth Summit into a well respected daily bulletin of the progress of negotiations, distributed in paper form, but simultaneously, around the world. Earth Negotiations Bulletin was a flagship project of Canada's International Institute for Sustainable Development.

NGO engagement in the Rio process led to a new level of policy awareness as well as international networking between and among NGOs. Targeted effort was made by northern NGOs to ensure that voices from the developing world were heard. Prominent and respected southern analysis from activists like India's Vandana Shiva, Malaysia's Martin Khor and Walden Bello of the Philippines group, Global South, reached a larger and wider audience. Martin Khor of the Third World Network in Penang had long worked under the spectre of potential political harassment and arrest. Through the course of the Rio preparatory process, his keen political analysis came to be appreciated by the Malaysian government. In time, even Malaysia welcomed NGO participation on its delegation.

By the time the Earth Summit was held in Rio in June 1992, thousands of NGOs from around the world had become engaged in the process. The NGOs, with the assistance of Maurice Strong who was once again secretary general of the conference, occupied substantial political space at the conference as well as holding a global forum, or peoples' response to the governmental process, parallel to UNCED. Strong also recruited Ecuadorian NGO leader, Yolanda Kakabadze, who went on to head the IUCN (now known as the World Conservation Union), to coordinate NGO relations for UNCED.

The women's movement became prominently involved in the Rio process through the efforts of the Women's Environment and Development Organization (WEDO), led by the indefatigable feminist and former US Congresswoman, Bella Abzug. Youth organizations became involved, as did aboriginal peoples.

Within the non-binding agreements for future action, known as Agenda 21, there was explicit recognition of the role of major groups. These were identified as environmental groups, development organizations, women's groups, youth and indigenous peoples.

RIO TO SEATTLE

Progress made in identifying NGOs, increasingly with other sectors labeled civil society, as legitimate agents of change deserving of political space in governmental affairs, was solidi ed in the ongoing work of the UN. The follow-up to the Earth Summit commitments: the Framework Convention on Climate Change, the Convention for the Protection of Biological Diversity, the Rio Declaration and the voluminous, 40 chapter, Agenda 21 all of these important documents were relegated to a new institutional body, set up within the UN backwater, the ECOSOC. The Commission for Sustainable Development (CSD) held annual reviews of progress in the spring of every year. NGOs continued to have substantial rights of participation. But little progress was being made in meeting the Rio targets.

The Rio Earth Summit produced a weak set of goals, and still weaker political will to achieve them. Annual CSD meetings offered ministers of environment the opportunity to share their frustration over their expulsion to the political wilderness. Every government had identi ed competing priorities that left the Earth Summit goals in the vast heap of the marginal and unloved. The reality was that Rio had emerged as the environmental wave of public awareness was already receding. Heads of government issued their Rio promises, went home and, for the most part, forgot them.

Other global summits continued to include signi cant NGO participation. In 1994, the International Conference on Population and Development (ICPD) met in Cairo. Issues of population and reproductive rights that had been pushed to the back burner in Rio were nally dealt with in Cairo. Large numbers of NGOs, many veterans of the Rio process, brought forward environmental concerns within the context of the political mine eld of population pressures. The Cairo Programme of Action explicitly embraced the work of the Earth Summit, two years before. It spoke to issues of unsustainable production and consumption, the need to ensure sustainable levels of consumption as well as the need to balance population numbers.

The year after Cairo, world leaders met in Copenhagen in March 1995 to negotiate commitments to improve the human condition. The World Summit on Social Development (WSSD) also recognized extensive rights of participation for NGOs. Within months, the UN was again hosting a conference (Beijing, September, 1996) the UN Fourth World Conference on the Rights of Women. This brought thousands of NGOs both to the of cial and, a now expected, parallel NGO conference. The documents from the Beijing Conference, the Beijing Declaration and the Beijing Platform for Action, were heavily in uenced by the lively North South debate of NGOs, from environmental, development, human rights and women s groups from around the world.

It can be argued that the seriousness of purpose of governmental enterprise is directly, and inversely, proportional to the extent of NGO participation. While NGOs received credentials and became accepted players in the Rio process, the real confrontation was in negotiations under the General

Agreement of Tariffs and Trade (GATT) within the decade long Uruguay Round. Aggressive trade liberalization was supported by abundant political will. By 1996, the Uruguay Round had succeeded in establishing a powerful, centralized enforcement and governance body, the WTO.

In retrospect, it is interesting how the two sets of global agreements, one to protect the planet's life support systems from the ravages of human activity, the other to ensure that human activity would not be impeded by any barriers such as troublesome environmental measures, could proceed side by side. With the same governments negotiating, there were no discussion within either process of the other. But there is no question which global vision had the upper hand. The WTO had teeth, the Rio agreements an impressive set of gums. Moreover, there were scattered references to the path to sustainable development through trade liberalization and the free flow of foreign direct investment throughout Agenda 21. In addition to targeting new areas for trade liberalization (the removal of trade barriers designed to protect domestic production from cheaper imports) the Uruguay Round set out to create an institutional home for the GATT, a global umpire of trade rules with extremely effective sanctions for any country violating the GATT.

Into this arena, there was no room for NGO engagement. The first ministerial gathering of the WTO, in Singapore in December 1996, offered NGOs observer status to the opening plenary of speech making. But the actual negotiations, in the WTO's own decision-making culture, left out not only NGOs, but most governments. Key agreement was reached on issues within small informal talks between the Europeans, the United States, Canada and Japan. Developing country diplomats could be found wandering the corridors and asking Martin Khor of the Third World Network if he had heard what was taking place behind the closed doors.

The Singapore meeting can be noted for one important decision. Developing countries managed en masse to block the introduction of an investment treaty into the WTO's deliberations. The efforts to accomplish a multi-lateral treaty to protect investors' rights intensified within the Organization for Economic Cooperation and Development (OECD). The treaty under negotiation was dubbed the Multilateral Agreement on Investment (MAI). Unexpectedly, opposition to what seemed an obscure bit of business in a little known international agency spawned a new and powerful international force, the anti-globalization movement.

In January of 1997, the text of the draft MAI was obtained by NGOs and placed on the internet. Public Citizen, a Ralph Nader founded organization in Washington, DC, played a prominent role in spreading information about the MAI from its web site. The text made it clear that, going beyond the position of dominance that trade had achieved through the GATT, the MAI would provide corporations with potential compensation for reduced expectations of profit. Word spread through the internet and through global gatherings, such as the corridors of the CSD meeting at the UN, that something new was afoot to further undermine environmental protection and domestic regulation. The OECD, which had promised a deadline of completed negotiations for spring 1997, moved the deadline to spring 1998.

Within governments, it appeared that little consensus had been reached between trade ministries and less powerful departments. In particular, ministries of culture expressed doubts about the MAI. The French Minister of Culture, in particular, openly criticized the homogenizing terms of treating culture like a product. The many aspects of society perceiving threats in the draft MAI text banded together in a new and broader coalition of environment groups, labor unions, development organizations, health advocacy groups, cultural organizations, educators, women's groups, and aboriginals. Opposition to the MAI spread globally in an unprecedented and grassroots fashion, aided by the internet.

OECD governments increasingly found themselves on the defensive domestically over support for a treaty to decrease nation state powers of regulation and increase private corporate rights. By spring, 1998, the announced deadline for completing the MAI, the negotiations collapsed. A new global movement experienced the power of democratic protest.

Parallel efforts had been under way for years to reform or shut down international financial institutions, such as the World Bank and the International Monetary Fund. Through the late 1990s the spectre of Third World Debt was also subject to a global campaign. Jubilee 2000 was initiated by church groups in the UK, but spread globally as well. Its call for reduction and removal of debt loads on heavily indebted, poor nations, led to improvements in debt cancellation by governments.

Into the simmering and now well-organized network of NGOs from around the world came the news in 1998 that the third ministerial meeting of the WTO would be held in the US, in Seattle, Washington. The news was a gift to a movement aware of its gathering momentum. Previous WTO ministerial sessions had been in Singapore and Geneva. Neither location is conducive to attracting large numbers of protesters. Seattle was an ideal location and groups around the world began to organize. Within the US, the key to fulfilling promises of massive demonstrations in Seattle was the support of the trade union movement, the American Federation of Labor-Congress of Industrial Organizations (AFL-CIO). Environmental groups, from Sierra Club to Greenpeace, animal welfare groups such as the Society for the Prevention of Cruelty to Animals, pro-democracy and trade groups, such as Ralph Nader's Public Citizen, began working with the labor movement, churches and a growing coalition of interests concerned about the

impact of economic globalization. Voices from the South also prepared for Seattle, and key thinkers and visionaries, like Martin Khor and Vandana Shiva, as well as hundreds of organizations began planning for Seattle. European activists, as well, made Seattle a focal point. Canadian anti-free trade campaigner Maude Barlow of the Council of Canadians, played a key role. By November 1999, over 50 000 people filled the streets of Seattle.

The reality is that the WTO was in trouble well before Seattle. The extent to which its rulings exceeded valid trade concerns led directly to a process of political delegitimization of the institution. There was a high level of public distrust as the WTO struck down environmental and health regulations. The trade ministers could no longer count on negotiating under a shroud of secrecy. Unlike the Uruguay Round, the political interest in the WTO process had now expanded beyond trade negotiators. The defeat of the MAI awakened ministers of other portfolios to the reality that trade negotiations were too important to be left to trade ministers. Globally, environment and culture ministers, in particular, were paying attention. They had to pay attention because the public and the news media no longer viewed trade matters as arcane. And because, after the defeat of the MAI, it was clear to a growing citizens' movement that global trade fights were winnable.

The negotiations in advance of the third ministerial meeting had not gone well. No progress had been made on a draft declaration for the ministers in over a year. Two weeks before the Seattle meetings, the preparatory talks in Geneva ended in failure. Intractable issues included agricultural subsidies, with the EU in violent disagreement with US plans to target its subsidies, as well as implementation issues, dealing with complaints mostly against developing countries that they were too slow in implementing the Uruguay Round. On these and other complex issues, there would be no prepared declaration. Everything would have to be negotiated on the ground in Seattle.

Two weeks before Seattle, invitations to world leaders and heads of government were quietly withdrawn. Of heads of governments, only US President Bill Clinton would attend. And, as it turned out, he embraced the protesters' demands for core labor standards. It was the *coup de grace* that ensured no agreements would take place in Seattle.

The street demonstrations and police over-reaction will be the lasting images of Seattle. But as an exercise in citizen political strength, Seattle is much more significant for the growing strength of NGOs. The NGOs involved had developed a strong critique of the current threats of increased corporate rule. They coordinated across huge distances and raised awareness within their own countries. The culminating days of teach-ins and protests in Seattle further increased media coverage and global public identification of a new and previously obscure institution, the WTO.

CONCLUSION

For more than a century, environmental groups have played a major role in evolving governmental and societal responses to the human-caused degradation of our natural world. NGOs have scored many significant victories: from national parks and protection for endangered species, to global treaties to protect the ozone layer, and end nuclear testing. Individual toxic substances have been banned through NGO campaigning and new technologies with reduced environmental impacts have been embraced more quickly.

In the process, many environmental groups have become themselves bureaucratized and seen as part of the mainstream. Despite victories in individual cases, ecological integrity is everywhere threatened. Global environmental health is significantly worse now than at any time in human history.

NGOs have, however, continued to provide an essential watchdog and public education function in every nation on earth. It can be argued that as the role of governments and the nation state itself is shrinking, the role of NGOs is expanding. Civil society, at a global level, is organizing in new ways. Democratic expression as a counter balance too, increasingly powerful, private sector interests is a different role for NGOs than lobbying parliaments for governmental action. NGOs will likely continue to occupy an increasing area of political space as the challenges of global environmental change weaken existing institutions.

REFERENCES

Bonner, R (1993) *At the Hand of Man*, Alfred A Knopf, New York.

Carson, R (1963) *Silent Spring*, Fawcett World Library, New York.

Hays, S (1972) *Conservation and the Gospel of Efficiency*, Athenaeum, New York.

Nash, R (1967) *Wilderness and the American Mind*, Yale University Press, New Haven, CT.

Turner, F (1985) *Rediscovering America: John Muir in his Time and Ours*, Viking Penguin, New York.

Ward, B and Dubos, R (1972) *Only One Earth: The Care and Maintenance of a Small Planet*, W W Norton, New York.

WCED (1987) *Our Common Future*, Oxford University Press, Oxford, New York.

Environmental Politics

see Environmental Politics (*Opening essay*, **Volume 5**)

Environmental Philosophy: Phenomenological Ecology

Ingrid Leman Stefanovic
University of Toronto, Toronto, Canada

In many respects, industrialized society perceives problems of ecology and human dwelling to be technical matters of environmental and architectural management. No less important, however, are assumptions that shape the way that we frame environmental research, policies and programs. How we perceive and implicitly evaluate our built and natural environments in uence the questions we ask as well as the solutions we propose in matters of ecology and human settlement planning.

Environmental philosophy explores precisely such issues of how we understand and value our built and natural worlds. Within this eld, phenomenology - the investigation of hidden structures of experience - has a distinctive role. Rather than construct abstract, speculative ethical theories, the phenomenological method aims to illumine tacit paradigms and worldviews that underlie our relationship with the environment. For example, studies of the impact of the automobile on the environment typically investigate technical matters such as frequency of car use or degree of degradation of agricultural lands. The phenomenologist's contribution to the discussion arises from a concern to uncover tacit elements of human experience that subtly but effectively inspire our strong emotional ties to the car. In the absence of an understanding of peoples' values, loves and modes of behavior, technical solutions may be dif cult to implement. To explore implicit patterns of meanings that condition our everyday behavior becomes a central concern of environmental phenomenology or, as some have coined the eld, phenomenological ecology.

PHENOMENOLOGY IN A CALCULATIVE ERA

Of pivotal and major importance to the field has been the work of German phenomenologist Martin Heidegger, who distinguished between two kinds of thinking in the Western, metaphysical tradition: calculative and originative (Heidegger, 1966). *Calculative thinking* is linear, goal-directed and essentially quantitative, seeking maximum efficiency in its control and mastery of discrete, empirically defined entities. *Originative thinking* examines the foundations of calculation, by investigating not just things but relations between things. Such thinking requires openness to the deeper mysteries of existence that are more than mere logical puzzles to be calculatively ordered.

We have become particularly expert in calculative thinking and, in most cases, we equate calculation with genuine thinking *per se*. Such thinking computes ever new, ever more promising and at the same time more economical possibilities, racing from one prospect to the next (Heidegger, 1966, 46). Specifically result-oriented, calculation seeks to organize individual facts and entities with a goal of maximum efficiency, dexterity and control. In such a scenario, the world appears as a collection of discrete objects subject to human command. Nature becomes a gigantic gasoline station that we envision is ours to tame and direct (Heidegger, 1966, 50–51).

Such a reductionist interpretation of the meaning of thinking arises from within a historical tradition that primarily concerns itself with empirical, quantifiable realities, rather than some foundational, qualitative issues that exceed such quantitative measures. An atomistic analysis of beings (referred to by phenomenologists as *ontic analysis*) takes precedence over any inquiry into the mystery of Being itself (that is, ontological inquiry). Calculative thinking assumes that which is subject to measurement and quantification is real and actual. Material entities, as well as ideas and concepts that can be categorized and delineated in the clear and distinct fashion of bounded entities, are implicitly valued above all else. At the same time, *positivism* – the theory that knowledge is based on natural objects and properties that can be verified by the empirical sciences – gains prominence in a calculative era that assumes that only clearly bounded, materially evident entities are real.

When positively existing entities acquire such precedence, nebulous phenomena such as relationships or negations that cannot be seen as fully present consequently are not subject to serious management. In its fascination with things, our era becomes oblivious to the significance of any phenomena that cannot be secured through measurement, or categorical control. Favoring calculative parameters, originative thinking is obscured and forgotten as a genuine possibility of reflection.

Why is this forgetfulness of relations, negations and voids at all important to environmentalists? Consider first how the modern, positivist worldview interprets a particular negative phenomenon of absence, in a case of blindness or not seeing. St. Thomas Aquinas articulated an interpretation of this privation that is symptomatic of the calculative paradigm (Barrett, 1958). Aquinas appropriated an Aristotelian perspective that distinguished between *ens reale* (real Being) and *ens rationis* (conceptual Being). On this understanding, the cataract on the eye is real, because it can be empirically measured and it is an actual, positive entity. On the other hand, while blindness can be said in some sense to exist, as the *absence* of sight, it *is* only conceptually.

The problem with this interpretation, however, is that it downplays the experiential significance of the phenomenon of blindness. The philosopher William Barrett points out that for the man whose life has suddenly been thrown into

complete darkness, non-seeing, a privation, has descended on him with more crushing effect than brick from a rooftop (Barrett, 1958, 288). To reduce the depth of this experiential significance to the status of mere conceptual Being denigrates the human cost of such loss of sight and, to this extent, has to be philosophically misplaced as well. Nevertheless, strange as it may seem, in a tradition that assigns reality only to positively existing entities, the absence of sight can only be defined in terms of a conceptual truth. The alternative is to subject the phenomenon of blindness to a reductionist paradigm whereby it is circumscribed wholly within the presence of a material object – a cataract.

Analogously, one might imagine how the calculative worldview assigns value to a clear-cut forest. What matters is the timber resource, extracted and manufactured into positively existing entities for human use. Proceeding within a management scenario driven by economic utility, of far less value are the nebulous and too-complex-to-measure impacts on ecosystem functions, processes and relationships. The original forest matters only to the extent that it is a positively existing resource, or else that it exists conceptually in the mind as a fond but overly romanticized image or memory.

The philosophical paucity of the calculative paradigm arises once again when one considers the human impact of involuntary displacement. My home is expropriated and I am required to move to another location. Technically speaking, as my house is bulldozed, it is no longer. Is its significance merely to be relegated to the level of a fleeting memory or conceptual reality?

The trauma of involuntary displacement clearly can indicate more than the physical destruction of a house. Instead, it may mean the end of a whole way of life and a radical de-centering of my very being. The fact that Western society no longer engages in the large-scale displacements of the 1960s' urban renewal programs indicates that, implicitly, we know that the absence of home means significantly more than the destruction of a material object. Rather, the loss affects us deeply in terms of our very sense of place. In fact, we are beginning to sense that we interpret our built and natural environments, not merely in terms of reductionist categories but, more primordially, from a more complex, holistic perspective.

POSSIBILITIES OF ORIGINATIVE THINKING

It may appear that the obvious alternative to reductionist, calculative thinking would be a form of holism. Environmentalists themselves increasingly advocate precisely such an alternative. For instance, the Brundtland Commission emphasizes the need to integrate economic, socio-cultural and environmental concerns, building on the ecologist's widely accepted recognition that the whole is greater than the sum of its parts .

While these appeals can be quite compelling, ecoholism is hardly unproblematic. It may well be true that everything is ultimately connected in some way to everything else. Philosopher David Cooper has a point, however, when he notes that it is without interest and moral implication that a falling tree in Australia has some connection with the birth rate of flies in Alaska; and it is false if it is meant that such connections must be detectable, let alone significant (Cooper, 1992, 166). How broadly is the holistic net to be legitimately thrown if it is to be practically and morally relevant?

Other problems with the holistic alternative also arise. If the whole is greater than the sum of its parts, does that conclusion relegate the individual parts to a lower status on the moral hierarchy? Environmental philosopher Tom Regan files this charge of eco-fascism against holistic thinking. Holistic philosophers who argue that one is justified in sacrificing individuals in order to preserve other species or the integrity of ecosystems are subject to this charge of fascistic rule of whole over parts.

Another major difficulty with advocating holistic thinking as an alternative to calculation relates to the problem of definition. Amidst the growing sentiment that environmental problems require holistic, interdisciplinary solutions, there are calls for the development of more comprehensive indicators of sustainability. The World Bank proposes an Index of Sustainable Economic Welfare (ISEW) (*see* ISEW (Index of Sustainable Economic Welfare) and GPI (Genuine Progress Indicator), Volume 5) to measure gross domestic product but to include natural and social parameters as well. President Clinton's Council on Sustainable Development called for the creation of a number of new economic, social and environmental quality indicators. Again, however, the question arises: does holism mean simply throwing a wider net around discrete entities to accumulate an ever-larger inventory of indicators?

From a phenomenological viewpoint, these efforts to describe a holistic alternative to positivism remain philosophically problematic because they remain immersed within the calculative paradigm. To say that everything is connected to everything else remains focussed on individual, discrete entities that may (or as Cooper points out, may not) be meaningfully interconnected in some way. These efforts do not call for a genuinely different kind of thinking than the calculative that simply seeks to amass a great many individual things in a broader web of significance.

The same problem occurs with accumulating a larger array of indicators, under the guise of interdisciplinary thinking. Genuinely holistic approaches must do more than simply collect an ever-increasing number of entities, if the phenomenologist's call for an alternative to calculation is to be heeded. Finally, in deference to Regan's criticisms, holistic thinking must accomplish more than a grand synthesis or construction of a Super-Part, bigger and better than the individual parts beneath it.

How do we genuinely move beyond calculation, once we recognize that reductionist versions of positivism fail to capture the increasing complexity of environmental problems? Phenomenology offers an alternate vision. The originative, ontological thinking that Heidegger proposes is grounded in *hermeneutics*, the art of interpretation of texts. If we try to read as a word a conglomeration of letters, e.g., *rgonaci* – and then rearrange the letters as *organic*', the second reading generates more than a mere re-configuration of individual letters. On the contrary, in the latter case, an implicit order now recognized among the letters allows for a shift in the act of interpretation that reveals the meaning of the word as a whole.

In fact, when we read a text, the understanding consistently interprets holistically, although we may not be explicitly aware of this fact. Consider how reading requires something different than the accumulation of individual words on a page. The experience of reading is more than merely additive, in that I do not simply store up a linear series of individual words so that, at the end of the sentence, I compute the series into a meaningful totality. Rather, the hermeneutic experience reveals an interplay between the significance of the text as a whole, and its progressive illumination through its parts – words, sentences and paragraphs. Because reading is other than merely summative or sequential, it offers a particularly revealing, everyday example of how understanding is holistic, where the whole means something intrinsically other than a superpart (Bortoft in Seamon and Mugerauer, 2000).

In this respect, Heidegger distinguishes between the totality that we analytically or sequentially construct and the wholeness of the world that is more subtly, tacitly revealed to the understanding. The world itself is much more than something subsequent that we calculate as a result from the sum of all beings. The world comes not afterward but beforehand, in the strict sense of the word ••• Instead, it is so self-evident, so much a matter of course, that we are completely oblivious to it (Heidegger, 1982, 165).

The task for ontological thinking, then, becomes one of illumining how we are in-the-world immediately, prethematically – meaning prior to the reductionist move of defining and segregating discreet entities from the referential context within which they find their place. Originative thinking will aim to uncover the underlying conditions of the possibility of calculation and the scientific analysis of discreet entities. In that case, the question to ask is: how might such a phenomenological undertaking affect our inquiries into the meaning of our built and natural environments?

RETHINKING SENSE OF PLACE

When Heidegger urges us to re-evaluate the calculative paradigm, he invites us to think meditatively or originatively, which is to say to raise anew the question of the meaning of Being. Rather than investigating individual entities in isolation–either the human subjectivity or objective environmental conditions – he urges us to reflect upon the relation between human beings and the places wherein they dwell. We are more than simply isolated subjectivities but are in-the-world, the hyphenation emphasizing an essential belonging of human beings to their environments. To inquire about how we are in-the-world, prior to artificially separating and reifying subjective and objective entities, entails a different kind of ontological thinking that is less concerned with beings than Being itself.

How might we ontologically and phenomenologically describe the way in which we *are* in the world? Heidegger believes that, fundamentally, the way in which you are and I am, the manner in which we humans *are* on the earth, is *dwelling*. To be a human being ••• means to dwell (Heidegger, 1971, 147). When we build our settlements, we do not simply erect physical constructions of particular proportions. Nor is it enough to say that buildings can be more meaningfully described as cultural symbols or that they possess social, economic or technological significance on top of their material substantiality.

Phenomenologist Gaston Bachelard points out that the lived experience of our environment is of more than mere physical artifacts. A house that has been experienced, he writes, is not an inert box. Inhabited space transcends geometrical space (Bachelard, 1964, 47). Similarly, in describing human beings as dwellers, Heidegger wants to say that built spaces are neither mere external objects nor social constructs but rather, spaces open up by the fact that they are let into the dwelling of man ••• Dwelling is *the basic character* of Being in keeping with which mortals exist (Heidegger, 1971, 157, 160). Noel Arnaud's simple but pointed remark captures these insights succinctly: I am not merely contained within space, but on the contrary, Je suis l'espace où je suis. (I am the space where I am) (Bachelard, 1964, 137).

One's fundamental comportment in-the-world is defined by Heidegger in terms of a spatializing and temporalizing activity of understanding and a primordial engagement with a world that is itself more than a static entity. Why, then, is the relationship between humans and their world not simply one that can be defined as a correspondence between two static, present-at-hand entities? Heidegger asks his readers to consider a common example of how we relate to our environment. As I head toward the door of a lecture hall, I am already there and I could not go to it at all if I were not such that I am there. I am never here only, he argues, as this encapsulated body; rather, I am there, that is, I already pervade the room and only thus can I go through it (Heidegger, 1971, 156–157).

Poet Pierre-Jean Jouve reiterates this thought when he writes: Car nous sommes où nous ne sommes pas. (For we are where we are not) (Bachelard, 1963, 211). Certainly, we exist among positively existing entities but our way of being among those entities includes a holistic engagement with a broader, referential context that, as nothing particular, nonetheless opens up the condition of the possibility of entities appearing as entities. This relation of human beings in-the-world moves beyond an engagement with speci c things to include the more mysterious, incalculable ontological phenomenon of dwelling and what others beyond Heidegger have come to refer to in terms of a sense of place (Mugerauer and Relph in Seamon and Mugerauer, 2000).

Philosopher Edward S. Casey maintains that to exist at all as an event is to have a place to be implaced ••• The point is that place, by virtue of its unencompassability by anything other than itself, is at once the limit and the condition of all that exists (Casey, 1993, 13, 15). How we are in the world, in the most primordial sense, is as emplaced. Architect Christian Norberg-Schulz suggests that spirit of place is a qualitative, total phenomenon that cannot be reduced to any single property without losing its concrete signi cance.

On such an interpretation, natural and built places are more than mere geographical locations along geometrically de ned parameters. Phenomenologically understood, the experience of dwelling is much more rich, diverse and engaging than can be captured in any reductionist catalogue of entities. A key element of phenomenological inquiry relates to how entities belong together in place as well as how we might better encourage that sense of place through more sensitive design and policy-making. Heidegger maintained that the Western, metaphysical tradition has slipped into a forgetting of the question of the meaning of Being and, at the same time of necessity, this forgetfulness has compromised our understanding of the meaning of dwelling and being-at-home in our natural and built worlds. The task remains to see how we might further investigate this forgetfulness and seek a more balanced relationship with the places that we inhabit.

PHENOMENOLOGICAL METHODS OF INVESTIGATING NATURAL AND BUILT PLACES

Within a calculative frame of reference, one interprets the notion of method as a logical, linear process of reasoning that demands that speci c steps be followed, in order that de nitive results are achieved. In marked contrast to the calculations of positivist science, Heidegger tells us that the phenomenological, scienti c method is never technique. As soon as it becomes one, it has fallen away from its own proper nature (Heidegger, 1982, 21). Originative, phenomenological interpretations will involve something other than a how-to manual or a transparent, technical delineation of discrete rules. What announces itself within a phenomenological notion of method is a more iterative, process-oriented approach to investigating entities and the underlying ontological conditions of their appearance.

Yet Heidegger does concede that there are three basic components to the phenomenological method: reduction, construction and destruction (Heidegger, 1982, 19). For us, phenomenological reduction means leading phenomenological vision back from the apprehension of a being ••• to the understanding of the Being of this being (Heidegger, 1982, 21). Since ontology is de ned as an inquiry into the meaning of Being, phenomenology is ontology. Methodologically, a key concern will be to nd ways to illumine a holistic, ontological vision of the phenomenon under investigation, rather than to merely catalogue individual, discrete moments.

At the same time, phenomenologists know that Being does not become accessible like a being. We do not simply nd it in front of us (Heidegger, 1982, 21). More than a merely negative methodological measure that opposes beings, phenomenology requires that we should bring ourselves forward positively toward Being itself in a free projection and in engaged, phenomenological construction . That such a move is more than a merely arbitrary, subjective projection is ensured by a methodology that includes a destruction , that is, a critical process in which the traditional concepts are deconstructed down to the sources from which they were drawn (Heidegger, 1982, 23). Underlying traditions and taken for granted patterns of understanding are thereby laid bare.

In short, it seems that the phenomenological method will seek to move from an investigation of beings, to a reduction to ontological grounds. Such a move will do more than relinquish individual entities altogether, but instead will require a special comportment and constructive engagement with the world an engagement that demands a free projection that is both creative and originative at the same time. Finally, as a destruction or dismantling of those arti ces that conceal originary meanings, the phenomenological method will seek to uncover the tacit sources of the traditions that sustain our ways of thinking and the holistic context wherein our questioning is grounded in the rst place.

Accepting such an understanding of phenomenology, one is still drawn to ask: how do these components translate into actual investigative procedures for the study of built and natural environments? How do we study concretely and holistically the manner in which we are emplaced in-the-world, without compromising the ontological endeavor?

One strategy for illumining ontological presuppositions of dwelling consists of a deconstruction of place-based narratives. While common methods of investigating perceptions and values include such tools as standardized questionnaires, the phenomenologist believes that such thematic approaches by definition cannot accommodate the complexity of prethematic presuppositions that are taken for granted before we conceptualize specific responses. Rather than reify peoples' perceptions within preconceived frameworks and parameters constructed in advance by researchers, phenomenological interview techniques lean toward less structured approaches to uncovering perceptions of place. In the words of philosopher Jim Cheney, the deconstruction of spontaneous narratives encourages us to avoid totalizing discourse and the top-down imposition of universal, preconceived theories as we collectively seek to tell the best stories we can to uncover our communities storied residence and local, bioregional truths. The phenomenological task is to illumine from the ground up, so to speak, holistic visions of place that emerge from spontaneous, unstructured narratives of peoples' lived experiences.

Other phenomenological strategies seek to describe more holistically our modes of dwelling in natural and built environments. From phenomenological readings of actual places to interpretations of artistic and literary accounts, to thoughtful, first-hand observation of particular settings, the primordial concern is to illumine taken-for-granted paradigms that condition environmental perceptions and interpretations of place (Stefanovic, 2000).

At times, the way of phenomenology can be frustrating, to the extent that methods challenge traditional attitudes and require a new way of seeing and of asking questions. Nevertheless, properly employed, the methods are rigorous and high standards of qualitative research are demanded. Whether engaged in a phenomenological reading of place or in-depth interviews, the task is to illumine taken-for-granted attitudes and investigate the ontological grounds of judgments that, in every case, inevitably frame the process of environmental decision-making by scientists, public policy makers and laypersons. It is for this reason that phenomenology is described by some philosophers as nothing less than foundational ecology.

REFERENCES

Bachelard, G (1964) *The Poetics of Space*, trans. Maria Jolas, Beacon Press, Boston, MA.

Barrett, W (1958) *Irrational Man*, Doubleday, New York.

Casey, E S (1993) *Getting Back into Place: Toward a Renewed Understanding of the Place-World*, Indiana University Press, Bloomington, IN.

Cooper, D (1992) The Idea of Environment, in *The Environment in Question: Ethics and Global Issues*, eds D E Cooper and J A Palmer, Routledge, London.

Heidegger, M (1966) *Discourse on Thinking*, Harper and Row, New York.

Heidegger, M (1971) Building Dwelling Thinking in *Poetry, Language, Thought*, trans. Albert Hofstadter, Harper and Row, New York.

Heidegger, M (1982) *Basic Problems of Phenomenology*, Indiana University Press, Bloomington, IN.

Seamon, D and Mugerauer, R, eds (2000) *Dwelling, Place and Environment: Towards a Phenomenology of Person and World*, Krieger, Melbourne, FL.

Stefanovic, I L (2000) *Safeguarding Our Common Future: Rethinking Sustainable Development*, State University of New York Press, Albany, NY.

Environmental Psychology/Perception

Charles Vlek

University of Groningen, Groningen, The Netherlands

Global environmental problems have socio-behavioral causes for which technology offers only one (partial) solution strategy. Focusing on fundamental mechanisms of environmental degradation, this entry is about models and methods for social behavior change to reduce collective environmental risks. The exposition develops around an analysis of societal driving forces, economic system characteristics, and individual determinants human needs, opportunities and abilities underlying stressful behavior patterns. The common-resource (or: commons) dilemma paradigm is utilized as a general model for understanding and managing con icts between (short-term) individual and (long-term) collective interests. Following this model, any collective environmental problem and its behavioral causes first need to be diagnosed. Secondly, an evaluation and weighting of individual benefits versus collective risks is required. And thirdly, well-tuned policy interventions must be consistently planned, executed and evaluated. Various behavioral science ingredients are introduced for a sources and causes analysis of collective environmental problems and for specifying effective ways to limit environmental risks within sustainable levels.

Particular attention is devoted to human needs and values, and the relative importance of material wealth for people s overall quality of life. It is hoped that, hereby, physical and economic scientists as well as policy makers may be

inspired to conduct more fundamental problem analyses and to design more effective policies than is customarily done.

MACRO-LEVEL DRIVING FORCES OF UNSUSTAINABILITY

More than 20 years ago Ehrlich and Holdren (1971) introduced a simple formula to estimate the total environmental impact of a particular society, country or region. The present popular version of this equation is: $I = P \times A \times T$, or: environmental impact (I) equals the product of population size (P), the degree of affluence (A) per person, and the quality of the technology (T) used to produce one unit of affluence. The formula implies that there are three different fronts at which the battle for sustainable development is to be fought, viz. population, affluence (production and consumption) and technology. The formula also reveals the substitutability of one component by another. To illustrate, total environmental impact might remain constant under considerable population growth, as long as average personal affluence and/or the resourcefulness per unit of affluence is/are proportionally reduced. Also, while total environmental impact stays constant, the degree of affluence per person may well increase significantly, provided that the number of people and/or the resourcefulness per unit affluence is/are proportionally reduced.

Under an IPAT-perspective, for example, several Organizations for Economic Co-operation and Development countries (like the Netherlands) have little reason to feel comfortable about their sustainability: their population keeps growing – partly through immigration – their affluence is increasing – they are set on economic growth, and the resource needs of their production and consumption technology is hardly decreasing. Also from an IPAT-perspective, (Winter, 1996, 23) even nominates the US for the world's most overpopulated country, not per km^2 but given its citizens' affluence and the resource demands of its technology.

The IPAT formula enables one to explain, predict and eventually manage the size and seriousness of environmental impact, for different geographic regions or countries of the world, to determine the most important impact growth factor(s) and to draw conclusions on optimal environmental management policies. Goodland *et al.* (1994) have systematically examined the potential for change in the three areas covered by the IPAT formula: (1) limiting population growth; (2) limiting affluence and consumption growth; and (3) reducing the resource needs of production and consumption technology. Like Corson (1994), these authors generally focus on a number of policy priorities, which are different in character for high-income and low-income nations of the world. For example, high-income nations are advised to work on:

- transforming the culture of consumerism ••• into an ethics of sufficiency and environmental sustainability;
- internalizing environmental costs in energy prices and accelerating the transition to renewable energy sources (Goodland *et al.*, 1994, 153).

In contrast, the authors advise low-income nations to give priority to: accelerating the transition towards population stability •••, supporting technologies which provide increased employment opportunities for unemployed and underemployed individuals •••, and improving efforts towards poverty alleviation (p. 154).

Goodland *et al.* (1994, 154) generally conclude: technological change and population stabilization cannot suffice to move the world towards an environmentally sustainable future. Instead, a reduction in per capita consumption in high-income nations and a decrease in environmental throughput are required .

In a mid-term report of a multidisciplinary project about sustainable household metabolism, Noorman and Schoot Uiterkamp (1998, 252) conclude:

> the challenges posed by climate change and the desire for equity ••• are formidable. It is clear that to achieve the reduction goals indicated above will require major changes in the metabolism of households and of production. Specific redesign rather than evolutionary change is essential, although this may appear to be at odds with the actual trend of ever-growing development, production, consumption, and disposal of products and services. This is especially true in areas where the fastest growth in consumption is in high energy intensive items (such as cars).

If we search for the socio-behavioral causes of population growth, increasing affluence and an ever more resource-hungry technology, we hit upon other driving forces of unsustainable development, viz. institutions as vehicles for governing human societies, and culture as the conglomerate of socially shared beliefs, values and attitudes. Embedding and shaping the latter two forces are historical developments such as industrialization, division of labor, urbanization and market competition; see Røpke (1999) who indicates two fundamental pre-conditions for the historical generation of wealth in the North, viz. the availability of cheap fossil fuels and the immense transfer of resources from the South to the North (Martinez-Alier, 1995). This is in line with Dietz and Rosa's (1997) argument that the technology component in the IPAT-formula encompasses much more than was originally suggested. Actually T incorporates cultural, social, technical and infrastructural factors, which together determine how much environmental impact an activity is producing. Adding culture and institutions to the IPAT-formula might yield a CIPAT-equation. After rewording just two factors we obtain TEDIC, to indicate a complex of technological, economic, demographic, institutional and cultural developments in industrial society, which did and do stimulate societal (global, or regional) unsustainability. In Dutch, te dik, pronounced similarly as TEDIC,

means too big or too fat. In a different wording and order, the five driving forces have also been distinguished by Stern (1992); see also Vlek (2000).

The several driving forces of unsustainable development are schematized and clarified in Figure 1, which also indicates that (un)sustainability involves economic security and social well-being as well as environmental quality. What Dietz and Rosa (1997) noted about technology (see previous paragraph) may be similarly argued about population as well as affluence: *P* and *A*, too, incorporate cultural, social, technical, infrastructural and other factors driving their (geographically different) growth, particularly during the last 50 years. For population, Hilderink (2000) has computer-simulated the effects of various relevant factors such as fertility, diseases, birth-control and health policies, environmental conditions and economic growth. His conclusions are that in most regions population growth will come to a standstill, that world population in 2055 will stay below 9 billion after which it will decrease to 7.6 billion in 2100, and that United Nations population projections tend to be exaggerated. For affluence, to be further discussed herein, basic behavior determinants are people's needs, opportunities and abilities (NOA) to produce and/or consume, as also indicated in Figure 1.

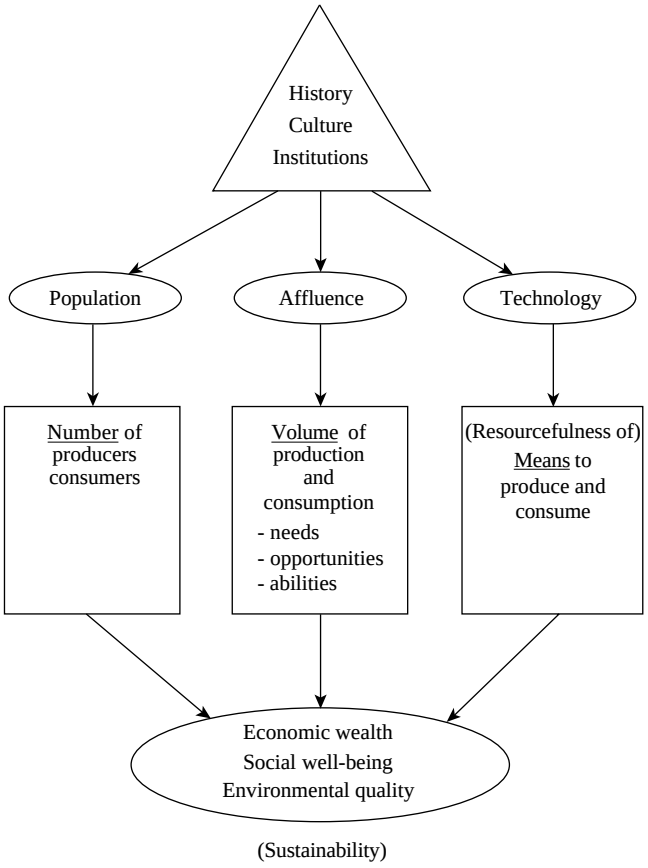


Figure 1 Schematic representation of driving factors underlying economic, social and environmental sustainability dimensions, as embedded in history, culture and institutions

In terms of Figure 1 one may conclude that the battle for sustainable development pertains not only – directly – to population, affluence (production/consumption) and technology, but also – indirectly – to institutions and culture. Thus, on the basis of an intense series of experts' and stakeholders' meetings, Vellinga *et al.* (1995) distinguish five strategic policy options for reducing the societal risks of climate change. These are, respectively:

1. *no regrets*: the climate problem is too uncertain, priorities will be given to policy instruments serving other (socio-economic and environmental) objectives;
2. *least regrets*: a serious climate risk problem is acknowledged, but policies will be aimed at a balancing of the risks of intervention and non-intervention;
3. *acceleration*: climate change is addressed along with other, shorter-term environmental problems, making use of synergies and positive feedbacks in society;
4. *technological innovation*: technology is the only way to meet growing demands and provide environmental security;
5. *institutional and cultural change*: technology being insufficient, major cultural and institutional changes are required to create a sustainable society.

Somewhat to their surprise, the authors noted that among project participants there was less agreement about the five strategic policy options above than about various fields of action that might be addressed (more or less strongly) under all five directions, e.g., introduction of ecotaxes, a low-carbon transport system, renewable energy sources, and the stimulation of technological and cultural innovations.

MESO-LEVEL PRODUCTION–CONSUMPTION CYCLES

The driving forces represented in Figure 1 are characteristic of a complex socio-economic system in which capital, labor and environmental resources are being used for the production of goods and services to meet the needs and desires of numerous consuming individuals, groups and organizations. To understand this complex metabolism of society vis à vis the natural environment, it is necessary to appreciate the interwovenness of household consumption and industrial production. Figure 2 represents what may be called the production-consumption cycle, as institutionalized in a social, i.e., government-regulated market economy. A more elaborate version of Figure 2, also containing an inner consumer self-production cycle, is presented in Vlek *et al.* (2000).

Figure 2 reflects the simple truth that consumers and producers need each other for different reasons, and

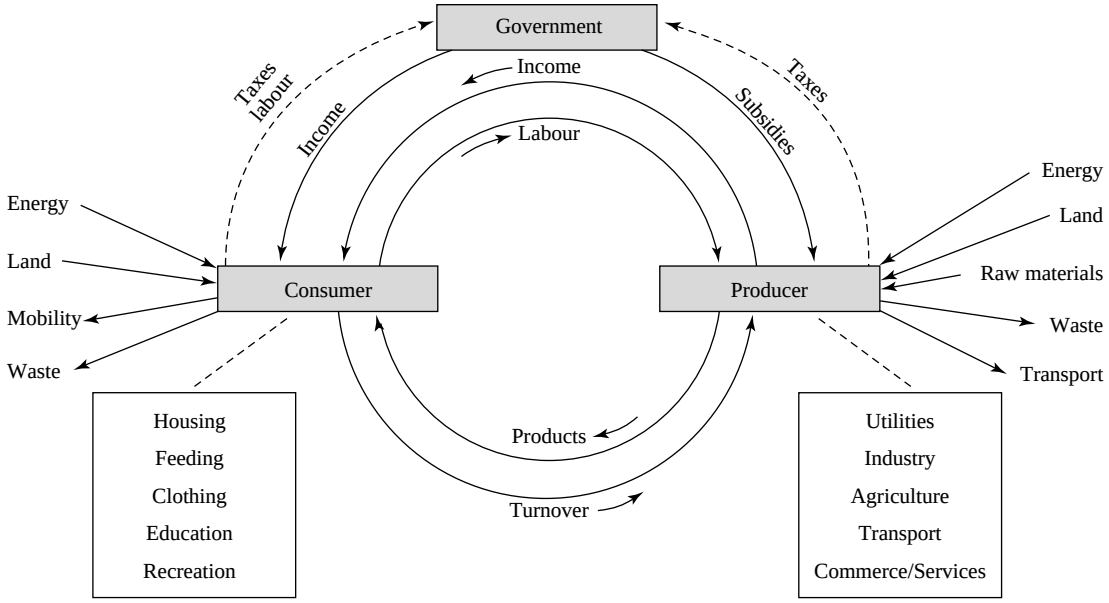


Figure 2 Model of the production-consumption cycle in a government-regulated market economy

that both parties need some government regulation, for which government in turn needs consumers and producers, again for different reasons. The relationships among the three groups are expressed in flows of money, products, labor, taxes and subsidies. The main system functions for consumers are feeding, clothing, housing, education and recreation. The major functions for producers are energy provision, industrial production, agriculture and stock-breeding, product distribution and services. Inputs from outside the socio-economic system come from various environmental resources such as energy, raw materials and land area. External outputs or derivatives occur in the form of various kinds of waste, mobility and noise.

A second observation about Figure 2 is that the whole (Western) production-consumption system is steadily accelerating. This is because it operates under the economic goal of ef ciency, because ef ciency breeds ef ciency and because most people involved cherish ef ciency for reasons of social comparison, competition and status. If ef ciency were less dominant and if material wealth were de-emphasized as one quality of life variable (see further herein), social market economies could decelerate as well as dematerialize (*see* **Dematerialization and Sustainable Development**, Volume 4), to the bene t of other important qualities of life, including social justice, environmental security and natural biodiversity.

Figure 2 can also illustrate a third, not so obvious point. This is the mutual interdependence of producers eco-ef ciency and consumers suf ciency. Eco-ef ciency is the producer s strategy of reducing the overall resourcefulness per unit of production goods or services (*see* **Eco-ef ciency**, Volume 4). However, the bene cial effects of eco-ef ciency may be undone by (further) consumption growth and rebound (or take-back) effects. Hence, eco-ef ciency on the producer s side needs a counterpart on the consumer s side. The latter must necessarily be a strategy of suf cient consumption or suf ciency (Durning, 1992, Chapter 10; Goodland *et al.*, 1994; Reisch and Scherhorn, 1999). This stands in contrast to the current (highly-industrialized) norm of ever-growing, continually maximizing consumption patterns (never enough, says homo oeconomicus).

MICRO-LEVEL DETERMINANTS OF RESOURCEFUL BEHAVIORS

For understanding individual producer and consumer behaviors, and as a prelude to the planning of needed social behavior change, some well-established theories and models of human behavior must now be considered.

Following the NOA-model as presented in Figure 3 (translated from Vlek *et al.*, 1997; inspired by Olander and Thøgerson, 1995), individual consumer behavior is considered as dependent upon the needs (Ns), the opportunities (Os) and the abilities (As) to undertake a particular activity. Needs and opportunities interact to shape people s motivation to perform an activity, while opportunities and abilities together determine subjects behavioral control. Hence, changing resourceful consumer behavior (or producer behavior, for that matter) would involve changing people s (or organizations) needs, their physical, technical or social opportunities and/or their physical, mental or nancial ability or capacity to engage in the relevant behavior. The NOA-model may be linked to the model of

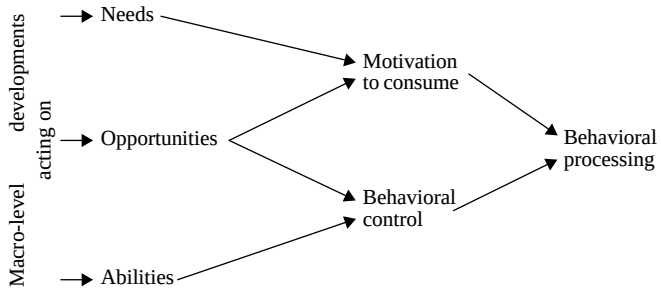


Figure 3 The NOA model of individual consumer behavior. The 'implicit and explicit forces'. ••• at the left are taken to be cultural, institutional, economic and other types of influential societal developments

the social market economy as represented in Figure 2, in order to identify the various actors and factors underlying any party's NOA to undertake resourceful behaviors.

Geller (1989) has proposed an integration of applied behavior analysis (ABA) – involving immediate rewards or punishments, following instrumental and social learning theory (see Geller *et al.*, 1982) – and a pragmatic social marketing approach (Kotler and Zaltman, 1971). Geller's (1989) four-stage ABA-marketing model sequentially comprises: (1) market analysis; (2) market segmentation and selection of target behaviors; (3) design of market strategy and ABA; and (4) evaluation via self-reports and behavior observations. Documentation and dissemination of results plays a key role in the model's practical application.

Stern *et al.* (1999) have proposed and tested a value–belief–norm (VBN) theory for understanding environmentalism in people's actions and behaviors. Their VBN theory links conceptions of moral norms, personal values and environmental attitudes into a coherent set of practical measurements. Their empirical study was aimed at explaining people's degree of environmental citizenship, their private-sphere behaviors, and their willingness to sacrifice for the support of government policies. Stern *et al.*'s study revealed that VBN theory is a far better predictor of environmental behavior indicators than other, less coherent conceptions of people's environmental motivation. This demonstrates that social and personal norms and values play crucial roles in changing behaviors.

The three models just discussed may each be considered as a general matrix for approaching human behavior change. Their consistent message is that effective policies should be aimed at changing fundamental determinants of unsustainable behaviors. In working with such models it is useful to distinguish four basically different types of behaviors. These may be categorized following two major dimensions under which specific behavior determinants can be specified. The first dimension ranges from deliberately planned or reasoned behavior on the one hand, to automatic behavior involving well-established habits, on the other. The second dimension ranges from private to public behavior determinants. The resulting two by two scheme is depicted in Table 1, whose quadrants may be conveniently labeled as deliberation, social comparison, repetition and imitation, respectively. For each behavior type two relevant, well-established behavior theories are indicated.

This simple psychological taxonomy may help us find an appropriate behavior theory or model and methods to understand and possibly influence environmentally impactful behaviors. The message here is that the four behavior types should each be addressed in a specific well-tuned way, in order for behavior change to come about. For example, repetitive (habitual) behaviors may best be approached via changes in a person's physical–technical environment. In contrast, deliberation may be influenced via argumentative information about optimal choices. Likewise, imitative behaviors may best be changed via social example setting and the pressure of social norms. Alternatively, social comparison is rather sensitive to argumentative information about other people's behavior. Disregarding the behavioral distinctions in Table 1 may yield perhaps well intended (and costly), but possibly ineffective policy strategies.

Exploiting the taxonomy of Table 1, Jager *et al.* (2000) have computer-simulated consumer behaviors to clarify which individual and social behavior mechanisms would lead to a earlier or later depletion of a common resource. In several simulation experiments contrasting *Homo psychologicus* and *Homo oeconomicus*, Jager *et al.* demonstrated that resource depletion was accelerated by an optimism effect, an imitation effect and an adaptation effect, which all flourished under uncertainty about the current size of the common resource.

SHORT-SIGHTEDNESS: COGNITIVE LIMITATIONS AND BIASES

Whatever behavior theory applies, there always is the human condition of psychological limitation, which makes environmental risk management rather difficult. Modern societies are confronted by huge discrepancies between the complexity, the uncertainty and the temporal extension of major environmental costs on the one hand, and the relative simplicity, certainty and immediacy of social and economic benefits on the other. By their very nature, the short-term, concentrated benefits are cognitively more available and can be better appreciated than the long-term, widespread costs. Bjorkman (1984) uses the term proximal cognition to explain this unevenness, which tends to make large parts of modern society to be the prisoner of an us, here and now' trap (Vlek and Keren, 1992). This precludes prudent long-term planning and decision making about developments involving major environmental impacts. Historically, such a psychological imprisonment is fairly unique, since earlier societies did not (yet) have the knowledge and the technical means for such large-scale

Table 1 Four different behavior types in two dimensions (translated from Vlek *et al.*, 1997)

	Private determination	Public determination
	Deliberation:	Social comparison:
Reasoned behavior	Decision and choice theory[a] Theory of planned behavior[b]	Social comparison theory[c] Relative deprivation theory[d]
	Repetition:	Imitation:
Automatic behavior	Classical conditioning theory[e] Instrumental learning theory[f]	Social learning theory[g] Theory of normative conduct[h]

The theory of planned behavior[b] (Ajzen, 1991) also incorporates a social component, the subjective (social) norm; hence it partly fits also under public determination. References for the other seven theories are: [a]Hogarth (1987), [c•d]Masters and Smith (1987), [e]Pavlov (1927), [f]Skinner (1953), [g]Bandura (1977), and [h]Cialdini *et al.* (1991).

and multidimensional environmental exploitation as has occurred particularly during the past 50 years.

Pawlik (1991) has indicated five fundamental reasons for most people's lack of appreciation of large-scale environmental risks such as global climate change. First, the signals of climate change, as exemplified by atmospheric temperature changes, are relatively weak compared to the background noise of normal temperature variations. Second, cause–effect gradients of global environmental change are strongly masked and extended in time; it is very difficult to tell which effects are the (indirect, long-term) result of which causes. Third, low-probability (extreme weather) events tend to be seen as impossible and their occurrence as accidental; this hinders the design of effective risk management strategies. Fourth, there is a large social distance separating actors and victims of global environmental change; this naturally suppresses actors' feelings of responsibility. Fifth, there is a low perceived cost-effectiveness of conservation behavior; often wasteful behavior is more profitable here and now. In an international review of studies about human responses to environmental stress, Jager and Vlek (1993) conclude that people are inclined to deny and remain passive about those kinds of environmental nuisance and risk that they believe to be uncontrollable.

In a critical psycho–cultural analysis Gladwin *et al.* (1997) wonder: why is the northern elite mind biased against community, the environment and a sustainable future? These authors identify four principal and interrelated origins (pp. 238–240): (1) a cognitively bounded biological mind; (2) an obsolete worldview mind; (3) an addicted contemporary mind; and (4) a delusional psychodynamic mind. They then formulate four sets of hypotheses about the different minds and provide a research agenda for investigating the conditions for developing a sustainable mind. For example, hypothesis two (p. 241) reads: the biomind is adapted for proximity rather than distance . Hypothesis 10 (p. 248): the viewmind conceives reality according to individualism rather than communitarianism . Hypothesis 11 (p. 250): the contempmind is programmed to favor market efficiency rather than social justice . Their hypothesis 19 reads: the psychomind protects the self from anxiety via rationalisation rather than accurateness .

COMMONS' DILEMMAS IN PRODUCTION–CONSUMPTION ACTIVITIES

The cognitive–motivational factors and tendencies just described are at the basis of what nowadays is known as the common-resource (or commons) dilemma (Hardin, 1968; Platt, 1973; Dawes, 1980; Vlek, 1996). In such social situations a collective (environmental) cost or risk is incurred or generated via the combined negative externalities of numerous individual benefit-seekers who act independently from one another. Vivid examples are the exploitation of fishing grounds by various companies, metropolitan air pollution through massive use of motor vehicles, and large-scale damage to natural ecosystems by expanding road infrastructure. Note that societal meso-level processes in production–consumption cycles (cf. Figure 2) operate as revolving doors for collective risk generation and management, as they are rooted in often powerful organizations of resourceful activities.

In our times, the classical conflict between individual and collective rationalities has acquired frightening proportions. Regarding highly industrialized society the Belgian philosopher Vermeersch (1988, 29) dramatically formulates this conflict as follows (translated from the Dutch):

> the whole forms a system which rushes on autonomously, and nobody can guarantee that somewhere at the end of the route there is a goal waiting which is still meaningful for people. ••• The aimlessness, the irrationality of the total system is being obfuscated by the utter rationality of the system's separate components.

The situation Vermeersch refers to represents a class of market failures that Kahn (1966) has called the tyranny of small decisions. The environmental and social consequences of such social traps (Platt, 1973) are also discussed by Hirsch (1976).

The examples above also indicate that many commons dilemmas have a layered structure comprising various contributing actors and risk managers, who may be identified at the micro-, meso- and macro-levels of society. Individuals may be embedded in groups, which are embedded in organizations, which are embedded in countries. This may go up to the international level where diplomats are facing complicated dilemmas concerning, e.g., the abatement of global atmospheric warming. Due to their layered structure, such commons dilemmas are hard to manage, since there is ample room for external cost-shifting behaviors as well as great difficulty to reach workable control agreements.

For social policy makers, a common-resource dilemma – if recognized as such – constitutes a permanent (dynamic) contrast between a collective risk and an often large collection of individual benefits; minimization of the risk and maximization of the benefits are incompatible social goals between which a trade-off is to be made. Thus, the solution of any commons dilemma consists of a sustainable balance between numerous individuals' (and/or groups', organizations', countries') benefits and the collective costs and risks for all.

What happens in a commons dilemma usually exceeds the physical, cognitive and motivational scope of individual actors. Therefore, the basic question is how the collective cost or risk can be validly assessed, effectively evaluated and acceptably managed so as to stay within sustainable limits (not easy to assess in advance). Collective risk management is a matter of decision-making about risk acceptance and the selection of practical strategies for risk control via social behavior change. Such risk management may be most effective if it links up with the problem diagnosis concerning the behavioral processes whereby the risk is being generated or enhanced. Each major risk management step may be unfolded into three different policy tasks. Thus, understanding commons dilemmas and managing collective risks revolves around nine distinct policy tasks, as listed in Table 2.

Table 2 Nine policy tasks for managing common-resource dilemmas (after Vlek, 1996)

I. Problem diagnosis
1. Analysis and assessment of the collective-risk and risk generation process
2. Promoting social risk perception and communication
3. Analysis and assessment of (numerous) individual benefits
II. Decision making
4. Weighing of collective risk against total individual benefits (need for change?)
5. If risk unacceptable: specification of safer behavior alternatives
6. Setting risk reduction objectives and translation into behavior goals
III. Risk control
7. Design and selection of policy instruments for behavior change
8. Programmatic application of various strategies for behavior change
9. Monitoring, evaluation and feedback of effects of risk reduction policies

The three divisions of Table 2 indicate the key problems in understanding and managing collective risks in commons dilemma situations: raising awareness and appreciation of the collective risk, weighing the risk against (total) individual bene ts, and promoting social behavior change (or restraint). This partly overlaps with three well-known keys to resolving social dilemmas (Dawes, 1980) knowledge, morality and trust. In a thorough analysis of commons dilemma problems Edney (1980) emphasized territorialization (or regionalization, to increase users responsibility) and trust (to reduce competition) as essential solution strategies.

STRATEGIES FOR SUSTAINABLE BEHAVIOR CHANGE

Commons dilemmas re ect persistent con icts between many individual (producer and consumer) interests on the one hand, and a small number of (large-scale) collective interests on the other. As dilemmas, they may be resolved only by the achievement of a safer, sustainable balance of individual and collective bene ts and risks. The nature and the effectiveness of various solution approaches have been investigated in a great number of laboratory and some eld experiments (e.g., Dawes, 1980; Messick and Brewer, 1983). Most of these approaches may be categorized under seven general strategies for social behavior change (Vlek, 1996; see also Cook and Berrenberg, 1981; De Young, 1993; Gardner and Stern, 1996, Chapter 7), as given in Table 3 along with exemplary speci cations.

Strategies 1 (PhAA) and 3 (FES), and certain (physical) forms of strategy 6 (OCh) would initiate so-called structural (or: hard) solutions to a commons dilemma, whose basic nature or type would thereby be altered. Strategies 2 (RaE), 4 (IEC), 5 (SMS), certain other (mental) forms of strategy 6 (OCh), and strategy 7 (CVM) would imply cognitive-motivational (or: soft) solutions (Wilke, 1989). Through the latter, individual actors would be induced to behave in a cooperative (i.e., collectively rational) manner, while the basic nature and payoff structure of the commons dilemma would be maintained. Structural solution strategies are generally more effective, but they are often not available or not easily implemented. Speci c cognitive motivational solution-strategies (RaE, IEC, SMS and some OCh) are more easy to design and apply, but their effectiveness is generally lower; in many cases, however, they are the only thing one can rely on. CVM stands relatively by itself as

Table 3 Strategies for collective risk management in commons dilemmas (Vlek, 1996)

1. Provision of physical alternatives, (re)arrangements (PhAA)	Adding/deleting/changing behavior options, enhancing efficacy
2. Regulation and enforcement (RaE)	Enacting laws, rules; setting/enforcing standards, norms
3. Financial–economic stimulation (FES)	Rewards/fines, taxes, subsidies, posting bonds
4. Provision of information, education, communication (IEC)	About risk generation, types and levels of risk, others' perceptions and intentions, risk reduction strategies
5. Social modeling and support (SMS)	Demonstrating cooperative behavior, others' efficacy
6. Organizational change (OCh)	Resource privatization, sanctioning system, leadership institution, organization for self-regulation
7. Changing values and morality (CVM)	Appeal to conscience, enhancing 'altruism' towards others and future generations, reducing 'here and now' selfishness
By default: 8. Wait and See (WaS):	Do nothing, the quay will turn the ship

a cultural solution (cf. Section 2) on which much behavior change might come to rest.

The (first) seven strategies listed in Table 3 may be applied in different variants and they may be directed at different target behaviors. Because each strategy has its own strengths and weaknesses – which may be summarily expressed in its behavioral elasticity (a dose–effect relationship) – it is generally wise to apply well worked-out combinations of strategies. Obviously, any strategy for social behavior change should be carefully tuned towards the intended target individual, group or organization. The latter should be studied to some extent beforehand, so as to obtain basic information necessary to design an effective approach.

One solution by default to a collective environmental risk problem is nicely expressed in a Dutch saying translated as the quay will turn the ship. That is, number eight in Table 3: WaS will automatically elicit behavioral (catastrophe) responses from causal agents, but these behavior changes then cannot be but inappropriate, too little and too late. Making such unsustainability scenarios palpable – e.g., via scenario studies – would mean that individual actors are (better) enabled to live under the shadow of the (common) future and duly undertake preventive action. Disaster scenarios may thus initiate self-destroying prophecies, as society tries to steer away from them.

The barriers to sustainable behavior change are manifold. In fact, like solutions, barriers may also be listed under the (first) seven rubrics of Table 3. Some examples are: absence of physical or technical alternatives (PhAA), insufficient and/or ineffective law enforcement (RaE), inconsistency of financial incentive systems (FES), unawareness of one's own causal role and possible contribution to solutions (IEC), absence or invisibility of model behaviors by opinion leaders (SMS), organizational goals biased towards short-term survival (OCh), and importance of social status in spending capacity (CVM). Specific variants of the seven strategies may be used to overcome such difficulties.

TRADABLE EXPLOITATION RIGHTS

One special combination of strategies RaE (as government regulation), FES (as payment according to use), IEC (as information provision) and OCh (as privatization) may prove to be an effective as well as efficient approach towards managing common goods. This is the management system known as tradable emission permits, more generally to be labeled as tradable exploitation rights (*see* **Tradable Permits for Greenhouse Gases**, Volume 4; **Tradable Permits for Limiting Stratospheric Ozone Depletion**, Volume 4; **Tradable Permits for Sulfur Emission Reductions**, Volume 4; **Transferable Fisheries Quotas (TFQs)**, Volume 4).

The idea here is that, for the collective as a whole, some central authority determines the maximum sustainable exploitation of some social or environmental resource (to be consumed, enjoyed or polluted), and then distributes (equally, shall we say) among its constituents a collection of tickets each representing a fixed limited exploitation right. Tickets are tradable, so that they may be purchased by those who need a greater than average resource use, from those whose demand is less than average. Ticket prices are established under free-market laws of demand and supply, but the total resource exploitation, corresponding to the ownership of a certain number of tickets, may not be exceeded. If necessary, the central authority may reduce the extent of exploitation per ticket. Alternatively, it may allow for increased exploitation, if the total resource regenerates more quickly than was originally estimated. See Koutstaal and Nentjes (1995) for an economic and legal analysis of the possibilities and limitations of tradable emission permits. The system contains several elements of the alternative commons dilemma games as discussed by Ostrom (1990, Chapter 1).

Tradable exploitation rights as a political–economic risk management system raises a number of fundamental questions. These are mainly concerned with the assessment

of the maximum tolerable overall exploitation of a given resource, the initial distribution (or perhaps: auctioning out) of exploitation tickets, and the central authority's monitoring and control over actual resource exploitation and possible violations of individual purchased rights. Effective functioning of such a system obviously requires collective rules and a sanctioning system. It also necessitates efficient operation of demand and supply pricing mechanisms. It also demands adequate provision of information and communication about exploitation rights, their availability and the monitoring and control of their actual utilization.

If appropriately implemented, tradable exploitation rights would prohibit free riding, whilst the unfairness of a fixed exploitation charge to both small and large exploiters would be avoided (see the discussion of demand revelation by Yamagishi, 1995, 313). How tradable exploitation rights as a system actually works under variations in certain basic characteristics, is a matter of controlled experimentation and creative field study. Practical examples are the distribution of tradable quota-rights for sea-fishing, dairy milk production and animal manure production, respectively, in the countries of the European Union. A promising application would involve transportation kilometers, to limit total environmental impacts from motorized transport.

In view of the common-resource dilemma model of collective environmental problems, tradable exploitation rights is an almost ideal risk management system, because it is aimed at the necessary balance between individual and collective interests, between government and market and (thus) between individual freedom and social equality.

A related risk management system, also involving a combination of various Table 3 strategies, is bond pledging or hostage posting for ensuring collectively responsible behavior. The basic idea is that individual actors promise to be cooperative, and that they commit themselves by pledging a bond (which serves as a hostage). The bond – or hostage – is lost when they defect and behave against the collective interest. Raub and Keren (1993) and Keren and Raub (1993) discuss the theoretical nature and implications of such a system, and they offer empirical evidence for its effectiveness. An interesting practical example from environmental policy is couched in a Swiss proposal to drastically increase fuel prices for motorcars, and then at the end of each year pay a bonus out of a national fuel fund to those who would prove to have driven less than a certain number of car-kilometers (VCS, 1990).

CONDITIONS FOR SOCIAL BEHAVIOR CHANGE

Global environmental problems offer good reasons for social behavior change. Common-resource dilemmas confront subjects with necessary risk–benefit trade-offs and the prompting of direction of needed change. The risk management strategies and systems discussed above can be instrumental in bringing about social change. From such policy foundations as well as from various empirical research (e.g., De Young, 1993; Oskamp, 1995; Gardner and Stern, 1996; Steg and Vlek, 1997) we may distill the following essential conditions for social behavior change, considered from the subject's point of view; see Table 4.

Fulfilling these conditions is like answering obvious questions by subjects of policy intervention campaigns, such as: what is the problem? What is my own responsibility? What (else) could I actually do? Will others cooperate as well? Can we trust risk experts and managers? Where would this all lead to? The latter question in particular is rarely addressed in environmental policy debates. Sustainable futures have much to do with meeting basic human needs and values (cf. WCED, 1987), to be discussed herein, in the section on human needs and values.

HUMAN NEEDS AND VALUES

Individuals, groups and organizations consume raw materials, energy, products and services because they are driven by basic needs and wants, because they cherish cultural beliefs and values, and because they are guided by social and economic institutions. Together, need satisfactions and value manifestations make up people's quality of life. The latter is a multidimensional concept comprising variables such as health, family relations, work, income, housing accommodation, safety and environmental quality (see, e.g., Ormel *et al.*, 1997; Gatersleben and Vlek, 1998; Diener, 2000). Quality of life is not necessarily affected by additional commodities that consumers are urged to desire through clever advertising. The reason is that material goods can be associated with the promise to fulfill needs they are not suited to fulfill.

Human needs and values are numerous and have been the subject of various taxonomies (McDougall, 1932; Maslow, 1954; Max-Neef, 1992; Schwarz and Bilsky,

Table 4 Essential subject-conditions for social behavior change

1. Sufficient awareness of and insight into the collective environmental problem
2. Appreciation of one's own contribution to the problem and one's co-responsibility for solutions
3. Visibility and consistency of strategic policy goals
4. Availability of feasible behavior alternatives
5. Personal (extrinsic and/or intrinsic) incentives for behavior change
6. Basic trust in cooperation of fellow subjects
7. Basic trust in relevant management agencies and authorities
8. An acceptable image of the future resulting from social behavior change

1987). Maslow's well-known needs hierarchy relates physiological, safety, belongingness, esteem and self-actualization needs. Maslow holds that the lower needs are more elemental, and hence override the higher needs in case of deprivation.

Max-Neef (1992) has proposed a taxonomy of nine human needs, which successively involves subsistence, protection, affection, understanding, participation, leisure, creation, identity and freedom. The author separates these general needs from the various specific ways in which they may be satisfied. Four existential categories of needs satisfaction are distinguished: being, having, doing and interacting, respectively. For example, a person's need for identity may be satisfied by being outspoken, having a German shepherd dog, regularly doing hard work, and interacting (meeting) with friends in the local cafe.

Being, having, doing and interacting to satisfy certain needs or to manifest certain values often involve raw materials, energy, products and services. A general question, therefore, is: to what extent may valid (general) human needs be fulfilled in a far less energy- and material-intensive way so as to better fit into society's sustainable development? Another question is: to what extent and in what way might resourceful need-satisfying states (being), possessions (having), actions (doing) or interactions be reinterpreted such that they are going to be seen as actually far less need-satisfying than they originally were considered to be? The latter question touches upon the possible illusory nature of culturally normal and socially accepted ways of needs satisfaction, and it would entail a discussion about more authentic and meaningful ways of satisfying our real needs.

After surveying various theories and taxonomies of human needs and values, Vlek *et al.* (1998) used a shortlist of 22 different quality of life variables to explain people's feelings and judgments about prospective changes in their country's environmental and economic conditions. It appeared that (Dutch) questionnaire respondents expected more positive changes from environmental quality improvement along with economic recession, than from (further) environmental deterioration along with continued economic growth. In a household scenario evaluation study, Poortinga *et al.* (2000) requested 450 respondents to rate each of the 22 variables for its importance to their overall quality of life. After factor analysis it appeared that, with one exception (social justice), the 22 variables could be compactly summarized into seven factors, together explaining 60% of the variance in people's importance judgments. These factors, together with the variables that they include, are given in Table 5, in top–bottom order of averaged importance. Note that maturity, openness to change and material wealth were judged to be less important than family, health and safety, personal freedom and achievement.

Jackson and Marks (1999) scrutinize the question: to what extent does economic growth actually deliver increasing welfare, as is being supposed in neo-classical economics? They refer to Max-Neef's (1995) threshold hypothesis which holds that, in Western-industrial growth patterns, the many economic goods are increasingly being counterbalanced by three basic economic bads, viz. depletion of natural resources, degradation of the natural environment, and various costs to the quality of human life. After considering various data on increased consumer spending on material as well as nonmaterial needs (such as leisure), Jackson and Marks (1999, p. 439) appear to find it difficult to infer increased needs satisfaction from increased expenditure. They also state that material consumption might well provide pseudo-satisfaction of nonmaterial needs or even inhibit or violate those needs.

Ever since Adam Smith (1776), economists have argued that human welfare depends on material wealth, and even today they neglect the contrary experience suggesting that their assumption is valid only when people live in material scarcity. In an affluent society, welfare loses its correlation with wealth.

> The reason why economic growth no longer brings a sense of greater well-being, why the pleasures our new possessions bring melt into thin air, is that at the level of affluence of the American middle class what really matters is not one's material possessions but one's psychological economy, one's richness of human relations and freedom from the conflicts and constrictions that prevent us from enjoying what we have.

In debates about quality of life dimensions such as health, family relations, work and leisure time we often underestimate the fundamental human need for variety and change (not so important following Table 5): we generally don't want to be, have, do and interact with the same (person, goods, jobs or other people, respectively) every day or week of the year. Most people (want to) vary

Table 5 Seven quality-of-life (QoL) factors summarizing 21 out of 22 specific QoL-variables (given in brackets)[a]

6.	Family health and safety (health, partner and family safety)
3.	Personal freedom (freedom, privacy, leisure time)
7.	Achievement (education, work)
2.	Environmental quality (environmental quality, nature/biodiversity, aesthetic beauty)
5.	Maturity (identity/self-respect, security, spirituality/religion)
4.	Openness-to-change (social relations, change/variation, challenge/excitement)
1.	Material wealth (money/income, comfort, status/recognition, material beauty)

[a] From top to bottom factors bundle from (roughly) more to less important QoL variables. Factor numbers indicate the order of eigenvalues, i.e., the percentage of total variance explained by that factor.

in their behaviors between: rationality and emotionality, security and challenge, stability and change, uniformity and variety, tradition and innovation, comfort and hardship, inconspicuousness and social distinction, communality and individuality, and between norm-adherence and norm-transgression. Any kind of consumption (or production) pattern sustainable or not should therefore incorporate reasonable variations along the existential polarities indicated above. If sustainable behavior would (only) be rational, secure, stable, uniform, traditional, comfortable, inconspicuous, communal and norm-adherent, it would fundamentally deter most people for literally all too human reasons. Scitovsky (1976) might summarize this by saying: such a life would provide much comfort all right, but it would not give enough pleasure .

CONCLUSIONS AND SUGGESTIONS

Environmental damage and risks local, regional and global result from a complex whole of diverse behavior patterns making up a lively society. Harmful effects, in the short and/or the long term, can only be prevented if one addresses the causal mechanisms and the sources of environmental over-exploitation. Therefore, collective environmental problems are socio-behavioral problems for which technology offers only one (partial) solution strategy. Other strategies should be focused on population, economic processes, societal institutions and cultural beliefs, norms and values (cf. Figure 1).

In several respects, global environmental problems may be too far away for us, here and now, to appreciate and care about them. The psychological trick for managing them nevertheless would be to coordinate global policy measures with more proximate regional and local policies. This can only be accomplished via causes-and-sources oriented policy approach. Taking again the example of mass motorization, policies could be aimed at reducing global carbon dioxide emissions, regional air pollution and land use, and local noise and safety problems simultaneously, if they acted comprehensively on the private motorcar system as underlying a spectrum of environmental pollutants.

In democratic countries, effective environmental policies tend to be politically controversial, if not unacceptable. This can be understood in terms of the common-resource dilemma, where individual freedom stands in permanent opposition to social equality. In their tragic choice theory, Calabresi and Bobbitt (1978) even hold that freedom and equality are incompatible social goals, whilst a democratic system by nature cannot warrant equal allocation of opportunities and wealth. However, the ideal solution nding a sustainable balance between freedom and equality requires powerful statesmanship, which may be based on the principle that individual liberties should be subordinate to collective social and environmental qualities essential for all.

Finally, global environmental policy makers are facing huge differences in sustainability pro les between different regions of the world. The United Nations Human Development Report 1998 (UNDP, 1998, 1) states: the 20th century growth in consumption, unprecedented in its scale and diversity, has been badly distributed, leaving a backlog of shortfalls and gaping inequalities .

Shortly after the appearance of *Our Common Future* (WCED, 1987), William Ruckelshaus (1989), a former director of the US Environmental Protection Agency, wrote that sustainable development should mean: (a) economic development within ecological limits; and (b) worldwide equitable economic development (Ruckelshaus, 1989). From a psychological point of view, these two international problems rest upon very similar foundations. Both global environmental degradation and widespread human poverty challenge human intelligence, international trust and worldwide morality, so that we can be aware of the problems, design creative solutions and work towards a common future of economic security, social well-being and environmental quality. If we can y around the world within 48 hours, shouldn t we be able to understand and manage the global problems that modern life entails?

REFERENCES

Ajzen, I (1991) The Theory of Planned Behavior, *Organ. Behavior Human Decision Processes*, **50**, 179 211.

Bandura, A (1977) *Social Learning Theory*, Prentice-Hall, Englewood Cliffs, NJ.

Bjandorkman, M (1984) Decision Making, Risk Taking and Psychological Time: Review of Empirical Findings and Psychological Theory, *Scand. J. Psychol.*, **25**, 31 49.

Calabresi, G and Bobbitt, P (1978) *Tragic Choices*, Norton, New York.

Cialdini, R B, Kallgren, C A, and Reno, R R (1991) A Focus Theory of Normative Conduct, in *Advances in Experimental Social Psychology*, ed L Berkowitz, **24**, 201 234.

Cook, S W and Berrenberg, J L (1981) Approaches to Encouraging Conservation Behavior: A Review and Conceptual Framework, *J. Soc. Iss.*, **37**, 73 107.

Corson, W (1994) Changing Course: An Outline of Strategies for a Sustainable Future, *Futures*, **26**, 206 223.

Dawes, R M (1980) Social Dilemmas, *Ann. Rev. Psychol.*, **31**, 169 193.

De Young, R (1993) Changing Behavior and Making it Stick. The Conceptualization and Management of Conservation Behavior, *Environ. Behav.*, **25**, 485 505.

Diener, E (2000) Subjective Well-being: The Science of Happiness and a Proposal for a National Index, *Am. Psychol.*, **55**(1), 34 43.

Dietz, T and Rosa, E A (1997) Environmental Impacts of Population and Consumption, in *Environmentally Significant Consumption; Research Directions*, eds P C Stern, V W Ruttan,

J L Sweeney, and R H Socolow, National Academy Press, Washington, DC, 92–99.

Durning, A T (1992) *How Much is Enough? The Consumer Society and the Future of the Earth*, WW Norton and Earthscan, New York.

Edney, J J (1980) The Commons Problem; Alternative Perspectives, *Am. Psychol.*, 35, 131–150.

Ehrlich, P R and Holdren, J P (1971) Impact of Population Growth, *Science*, 171, 1212–1217.

Gardner, G T and Stern, P C (1996) *Environmental Problems and Human Behavior*, Allyn and Bacon, Boston, MA.

Gatersleben, B and Vlek, C (1998) Household Consumption, Quality-of-Life and Environmental Impacts: A Psychological Perspective and Empirical Study, in *Green Households? Domestic Consumers, Environment and Sustainability*, eds K J Noorman and T Schoot Uiterkamp, Earthscan, London, 141–183.

Geller, E S (1989) Applied Behavior Analysis and Social Marketing: an Integration for Environmental Preservation, *J. Soc. Iss.*, 45, 17–36.

Geller, E S, Winett, R A, and Everett, P B (1982) *Preserving the Environment: New Strategies for Behavior Change*, Pergamon, Elmsford, NY.

Gladwin, T N, Newburry, W E, and Reiskin, E D (1997) Why is the Northern Elite Mind Biased Against Community, the Environment and a Sustainable Future? in *Environment, Ethics and Behavior. The Psychology of Environmental Valuation and Degradation*, eds M H Bazerman, D M Messick, A E Tenbrunsel, and K A Wade-Benzioni, The New Lexington Press, San Francisco, CA, 234–274.

Goodland, R, Daly, H, and Kellenberg, J (1994) Burden Sharing in the Transition to Environmental Sustainability, *Futures*, 26, 146–155.

Hardin, G (1968) The Tragedy of the Commons, *Science*, 162, 1243–1248.

Hilderink, H (2000) *World Population in Transition*, Doctoral Dissertation, Department of Spatial Sciences, University of Groningen, Netherlands, 1–213.

Hirsch, F (1976) *The Social Limits to Growth*, Harvard University Press, Cambridge, MA.

Hogarth, R M (1987) *Judgment and Choice; The Psychology of Decision*, 2nd edition, Chichester, UK and New York.

Inglehart, R (1996) The Diminishing Utility of Economic Growth: From Maximizing Security Toward Maximizing Subjective Well-being, *Crit. Rev.*, 10, 509–531.

Jackson, T and Marks, N (1999) Consumption, Sustainable Welfare and Human Needs – with Reference to UK Expenditure Patterns Between 1954 and 1994, *Ecol. Econ.*, 28, 421–441.

Jager, W, Janssen, M A, De Vries, H J M, De Greef, J, and Vlek, C A J (2000) Behavior in Commons Dilemmas: Homo Economicus and Homo Psychologicus in an Ecological–Economic Model, *Ecol. Econ.*, 35, 357–380.

Jager, W and Vlek, C (1992) Hinder, Stress en Gedrag als Reactie op Milieuverontreiniging, *Gedrag en Gezondheid*, 20, 113–127.

Kahn, A E (1966) The Tyranny of Small Decisions. Market Failures, Imperfections, and the Limits of Economics, *Kyklos*, 19, 23–47.

Keren, G B and Raub, W (1993) Resolving Social Conflict Through Hostage Posting: Theoretical and Empirical Considerations, *J. Exp. Psychol.: General*, 122(4), 429–448.

Kotler, P and Zaltman, G (1971) Social Marketing: an Approach to Planned Social Change, *J. Market.*, 35, 3–12.

Koutstaal, P R and Nentjes, A (1995) Tradable Carbon Permits in Europe; Feasibility and Comparison with Taxes, *J. Common Market Stud.*, 33(2), 219–233.

Martinez-Alier, J (1995) Political Ecology, Distributional Conflicts and Economic Incommensurability, *New Left Rev.*, 211, 70–88.

Maslow, A H (1954) *Motivation and Personality*, Harper and Row, New York.

Masters, J C and Smith, W P, eds (1987) *Social Comparison, Social Justice and Relative Deprivation: Theoretical and Policy Perspectives*, Volume 4, Erlbaum, Hillsdale, NJ.

Max-Neef, M (1992) Development and Human Needs, in *Real-life Economics: Understanding Wealth Creation*, eds P Ekins and M Max-Neef, Routledge, New York, 91–108.

Max-Neef, M (1995) Economic Growth and Quality of Life – a Threshold Hypothesis, *Ecol. Econ.*, 15, 115–118.

McDougall, W (1932) *The Energies of Men*, Methuen, London.

Messick, D M and Brewer, M B (1983) Solving Social Dilemmas: a Review, in *Review of Personality and Social Psychology*, Volume 4, eds L Wheeler and P Shaver, Sage, Beverly Hills, CA.

Noorman, K J and Schoot Uiterkamp, T, eds (1998) *Green Households? Domestic Consumers, Environment and Sustainability*, Earthscan, London.

Olander, F and Thøgerson, J (1995) Understanding of Consumer Behavior as a Prerequisite for Environmental Protection, *J. Consumer Pol.*, 18, 345–385.

Ormel, J, Lindenberg, S, Steverink, N, and Vonkorff, M (1997) Quality of Life and Social Production Functions: A Framework for Understanding Health Effects, *Soc. Sci. Med.*, 45, 1051–1063.

Oskamp, S (1995) Resource Conservation and Recycling: Behavior and Policy, *J. Soc. Iss.*, 51, 157–178.

Ostrom, E (1990) *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge University Press, New York.

Pavlov, I P (1927) *Conditioned Re exes*, Oxford University Press, New York.

Pawlik, K (1991) The Psychology of Global Environmental Change: Some Basic Data and an Agenda for Cooperative International Research, *Int. J. Psychol.*, 26, 547–563.

Platt, I (1973) Social Traps, *Am. Psychol.*, 28, 641–651.

Poortinga, W, Wiersma, G, Steg, L, Vlek, C, Noorman, K J, Moll, H, and Schoot Uiterkamp, T (2000) *Expected Quality of Life Impacts of Experimental Scenarios for Sustainable Household Energy Use*, Departments of Psychology and Environmental Science, University of Groningen, Netherlands, report (in press).

Raub, W and Keren, G B (1993) Hostages as a Commitment Device: A Game-theoretic Model and an Empirical Test of Some Scenarios, *J. Econ. Behavior Organ.*, 21, 43–68.

Reisch, L A and Scherhorn, G (1999) Sustainable Consumption, in *The Current State of Economic Science*, ed S B Dahiya, Spellbound, London and India.

Ropke, I (1999) The Dynamics of Willingness to Consume, *Ecol. Econ.*, 28, 399–420.
Ruckelshaus, W D (1989) Toward a Sustainable World, *Sci. Am.*, September, 114–120B.
Schwartz, S H and Bilsky, W (1987) Toward a Universal Structure of Human Values, *J. Person. Soc. Psychol.*, 53(3).
Scitovski, T (1976) *The Joyless Economy: An Inquiry Into Human Satisfaction and Consumer Dissatisfaction*, Oxford University Press, New York and London (2nd edition 1992).
Skinner, B F (1953) *Science and Human Behavior*, Macmillan, New York.
Smith, A (1776) *The Wealth of Nations*, Penguin Books, Harmondsworth, UK, reprinted 1981.
Steg, L and Vlek, C (1997) The Role of Problem Awareness in Willingness to Change Car Use and in Evaluating Relevant Policy Measures, in *Traf c and Transport Psychology*, eds T Rothengatter and E Carbonell Vaya, Pergamon, London, 465–475.
Stern, P C (1992) Psychological Dimensions of Global Environmental Change, *Ann. Rev. Psychol.*, 43, 269–302.
Stern, P C, Dietz, T, Abel, T, Guagnano, G A, and Kalof, L (1999) A Value–Belief–Norm Theory of Support for Social Movements: the Case of Environmentalism, *Human Ecol. Rev.*, 6, 81–97.
UNDP: United Nations Development Programme (1998) *Human Development Report 1998*, Oxford University Press, New York (see also UNDP: *Human Development Report 1999* (idem)).
VCS: Verkehrs-Club der Schweiz (1990) *Oekobonus: ein Durchführbarer Weg zur Reduktion des Strassenverkehrs* (Ecobonus: a Feasible Way Toward the Reduction of Road Traffic), Postfach, CH-3360 Herzogenbuchsee, Switzerland.
Vellinga, P, Hisschemoller, M, Klabbers, J, Berk, M, Swart, R, and Van Ulden, A (1995) Climate Change, Policy Options and Research Implications, in *Climate Change Research: Evaluation and Policy Options*, eds S Zwerver, R van Rompaey, M Kok, and M Berk, Elsevier Science, Amsterdam, Volume A, 199–132.
Vermeersch, E (1988) *De Ogen van de Panda; een Milieuloso sch Essay*, Marc van de Wiele, Brugge.
Vlek, C A J (1996) Collective Risk Generation and Risk Management; the Unexploited Potential of the Social Dilemmas Paradigm, in *Frontiers in Social Dilemmas Research*, eds W B G Liebrand and D M Messick, Springer-Verlag, New York, 11–38.
Vlek, C (2000) Essential Psychology for Environmental Policy Making, *Int. J. Psychol.*, 35, 153–167.
Vlek, C, Jager, W, and Steg, L (1997) Modellen en Strategieen voor Gedragsverandering ter Beheersing van Collectieve Risico's, *Ned. Tijdschr. Psychol.*, 52, 174–191.
Vlek, C and Keren, G B (1992) Behavioral Decision Theory and Environmental Risk Management: Assessment and Resolution of Four Survival' Dilemmas, *Acta Psychol.*, 80, 249–278.
Vlek, C, Skolnik, M, and Gatersleben, B (1998) Sustainable Development and Quality of Life: Expected Effects of Prospective Changes in Economic and Environmental Conditions, *Zeitschrift für experimentelle Psychologie*, 45(4), 319–333 (with German summary).
Vlek, C, Reisch, L A, and Scherhorn, G (2000) *Transformation of Unsustainable Consumer Behaviors and Consumer Policies*, A Research-programming Report for the International Human Dimensions Programme on Global Environmental Change (Industrial Transformation), Department of Psychology Report no COV 00–04, 1–52, University of Groningen, Netherlands, (also in *IHDP Industrial Transformation: Research Directions*, ed P Vellinga, Bonn (G), IHDP, in press).
Wachtel, P (1989) *The Poverty of Af uence. A Psychological Portrait of the American Way of Life*, New Society Publishers, Philadelphia, PA.
WECD: World Commission on Environment and Development (1987) *Our Common Future*, Oxford University Press, New York.
Wilke, H (1989) Promoting Personal Decisions Supporting the Achievement of Risky Public Goods, in *Social Decision Methodology for Technological Projects*, eds C Vlek and G Cvetkovich, Kluwer, Dordrecht and Boston, 39–59.
Winter, D (1996) *Ecological Psychology: Healing the Split between Planet and Self*, Harper Collins, New York.
Yamagishi, T (1995) Social Dilemmas, in *Sociological Perspectives on Social Psychology*, eds K S Cook, G A Fine, and J S House, Allyn and Bacon, Boston, MA.

Environmental Refugees

see Environmental Refugees (**Volume 4**)

Environmental Security

Steve Lonergan
University of Victoria, Victoria, Canada

With the demise of the Cold War, there has been a rethinking of our traditional perspectives on security. Numerous authors have posited that our concentration on military security may be out of date, and that factors such as disease, human rights, and environmental degradation may pose equal, or even greater, threats to national security. At the same time, there is an increasing sense that states may not be able to guarantee the security of communities and individuals, particularly in reference to these new threats. Global environmental change poses a threat not only to the natural systems on which life depends, but may pose risks to the security of individuals directly. The purpose of this article is to present these views, and provide a new way of thinking about the human dimensions of global change.

THE EVOLUTION OF ENVIRONMENT AND SECURITY RESEARCH

National security has focused traditionally on protecting the territorial integrity and political sovereignty of the state from military aggression from other states. This generally has involved forming alliances and investing in military assets in order to deter potential adversaries and use force effectively when required. In recent years, there has been increased emphasis placed on expanding the traditional conception of security to include so-called non-conventional threats such as resource scarcity, human rights abuses, outbreaks of infectious disease, and environmental degradation caused by toxic contamination, ozone depletion, global warming, water pollution, soil degradation and the loss of biodiversity (Ullman, 1983; Renner, 1989; Westing, 1989). These discussions, in turn, have stimulated research on examining the specific relationship between environment and security.

The literature on environment and security emerged in the 1970s. However, prior to this, as early as the 1950s, discussions on the issue of environmental change and security occurred without explicit use of the term environmental security (Osborn, 1953; Brown, 1954; Ophuls, 1977). Following from these discussions, in 1977, the US Central Intelligence Agency established an environmental center to assess the relationships between environment and security. The US military's use of defoliants in Southeast Asia during the Vietnam War focused international attention on both the intentional and unintentional environmental damage caused by war. The Additional Protocol I to the 1949 Geneva Convention on the Protection of Victims of International Armed Conflicts (1977) was the first of two treaties with major environmental importance that stemmed from international concern over excessive environmental degradation in Vietnam (Diederich, 1992). This primarily humanitarian agreement has to date not been ratified by a number of major powers, including the US, France and the UK, although most objections do not center on the environmental issues contained in the agreement.

Efforts to develop more stringent definitions for the prohibition of widespread, long-term and severe damage to the natural environment continued with the 1977 Convention on the Prohibition of Military or Any Other Hostile Use of Environmental Modification Techniques (the ENMOD Convention), the second of the post-Vietnam treaties.

By the early 1980s, various institutions and writers began addressing security beyond strictly military concerns that affect the state. The United Nations (UN) Commission on Disarmament and Security chaired by Olaf Palme of Sweden, made a distinction between *collective security* and *common security*. The former implied the more traditional interstate military security issues, while the latter reflected the growing array of non-military threats, including economic change, resource scarcity, population growth and environmental degradation. This was followed by the new political thinking of Russian President Mikhail Gorbachev that promoted the concept of comprehensive security as a cornerstone of international politics. Comprehensive security included various threats, including nuclear war, poverty, and global environmental issues.

Coincidentally, numerous writers addressed the issue of expanding the definition of security to include non-military threats. Richard Ullman, for example, offered the following definition of threats to security:

> A threat to national security is an action or sequence of events that; 1) threatens drastically and over a relatively brief span of time to degrade the quality of life for the inhabitants of a state, or 2) threatens significantly to narrow the range of policy choices available to the government of a state or to private, non-governmental entities (person, groups, corporations) within the state.
>
> (Ullman, 1983, 133)

While still circumscribing security within state boundaries, Ullman sought to expand the range of threats to security beyond the traditional military concerns.

The suggestion to broaden the definition of security to include environmental threats was by no means limited to American sources. Although the report of the World Commission on Environment and Development, *Our Common Future*, is best known for its definition of sustainable development, the commission also called for recognition that security was partly a function of environmental sustainability. The Commission highlighted the causal role environmental stress can play in contributing to conflict while also stating that a comprehensive approach to international and national security must transcend the traditional emphasis on military power and armed competition (World Commission on Environment and Development, 1987, 290). Westing (1989) elaborated on this statement by noting that comprehensive security has two intertwined components: political security, with its military, economic and humanitarian sub-components, and environmental security, including protecting and utilizing the environment.

The 1986 nuclear meltdown at Chernobyl and its attendant devastation for neighboring human populations and ecosystems placed health considerations squarely within a security framework for many people. The next year, President Gorbachev proposed ecological security as a top priority that *de facto* would serve as a forum for international confidence building.

The initial phase of environment and security research concluded at almost the same time as the end of the Cold War. Articles by Matthews (1989) and Myers (1989) summarised much of the debate on broadening conceptions of security. Like earlier contributions, these efforts addressed two key issues. First, there was a need to redefine security to include a new range of threats. Such threats included

population growth, resource scarcity, and environmental degradation. Second, there was an acceptance that the object of security was no longer simply the state, but ranged to levels above and below the level of the state. Myers (1993), for example, equated security with

> Human wellbeing: not only protection from harm and injury but access to water, food, shelter, health, employment, and other basic requisites that are due of every person on Earth. It is the collectivity of these citizen needs, overall safety and quality of life, that should gure prominently in the nation s view of security.
>
> (Myers, 1993, 20)

Similar to the arguments made by the World Commission on Environment and Development, Myers argues for moving from security as a freedom from various threats to security as a freedom to access environmental services.

Today, critiques of this early literature often emphasize its ahistorical and alarmist character. Some historical analyses suggest that environmental change has been a key determinant of the rise and fall of civilizations: and, ultimately, of progress in many areas of endeavor. In short, there is nothing new about this relationship. Moreover, some critics argue, the claim that human activity is hastening a catastrophe becomes less in uential as the years and decades pass. In spite of its aws, however, the early literature initiated an important debate and laid the foundations for the more rigorous research programs of the 1990s.

ENVIRONMENT AND CONFLICT

Assessing the nature of linkages between environment and security has proven dif cult. The complexity of multiple interactions and feedback poses tremendous empirical and methodological hurdles. The ambiguous and contested nature of the term *security* also complicates research and policy in the area of environment and security (Dokken and Gr ger, 1995; Lipschutz, 1995; Deudney and Matthew, 1999). As noted previously, the meanings attached to the term range from a narrow state-based de nition of safety from armed con ict, to a much broader conception of security as synonymous with human well-being. In the 1990s a number of researchers have tried to circumvent this discussion by ignoring the word *security* and concentrating speci cally on the role of environmental change and resource depletion as potential causes of violent con ict (Homer-Dixon, 1991, 1994; Libiszewski, 1992). Such con- ict, in turn, could pose a serious threat to the security of individuals, regions and nation states. The general discussions on the nature of security and the role of environmental degradation as a contributor to insecurity and con ict are labeled by Levy (1995a) as the rst wave of environment and con ict research. The empirical research that attempted to prove a link between environment and con ict has been labeled by Levy (1995b) as the second wave of environment and con ict studies.

Work at the Peace and Con icts Studies Program at the University of Toronto (Homer-Dixon, 1991, 1994; Homer-Dixon *et al.*, 1993), the Environment and Con icts Project (ENCOP) in Zurich and Bern (e.g., Libiszewski, 1992; Spillman and Bachler, 1995), and the International Peace Research Institute, Oslo (e.g., Molv r, 1991; Gleditsch, 1992; Lodgaard and Hjort af Ornas, 1992; Gr ger and Smith, 1994; Dokken and Gr ger, 1995; Hauge and Ellingsen, 1997), among others, all contributed towards this effort (see also, Durham, 1979; Westing, 1986; Gleick, 1989, 1991; National Academy of Science, 1991; Lonergan and Kavanagh, 1991). These empirical studies have been crucial, not only in terms of advancing the scholarly discussion of the links between environmental change and violent con ict, but also in publicizing the potential role environmental degradation may play as a contributor to violent con ict. Although many studies focused on the somewhat muddled concept of environmental scarcity rather than on environmental degradation *per se*, the conclusion by Homer-Dixon *et al.* (1993) was clear: ••• scarcities of renewable resources are already contributing to violent con- icts in many parts of the developing world . Subsequent work by Bachler (1998) demonstrated that environmental degradation and resource depletion may play a number of different, and sometimes subtle, roles in affecting security and contributing to con ict. These include environment as background to the tensions, as a channel leading to tension, as a trigger, as a catalyst or as a target.

Some scholars have been critical of this deterministic perspective on environment and con ict (e.g., Deudney, 1991; Dalby, 1992; Conca, 1994; Levy, 1995a,b; Hartman, 1998). Despite the range of case studies undertaken, the evidence for a direct causal link between environmental degradation and violent con ict (implied by Homer-Dixon s statement above) remains speculative and anecdotal.

CLARIFYING THE RELATIONSHIP BETWEEN ENVIRONMENTAL CHANGE AND CONFLICT

The environment and con ict literature posits that several types of environmental threats may have the capacity to contribute to insecurity and to produce con ict as well. Constraints on resources are a crucial factor that is often discussed in the literature (Choucri, 1991). Rapid industrialization and population growth in many regions have resulted in an increased demand for both renewable and non-renewable natural resources, and as Ullman (1983) and others have noted, competition for resources has historically been a major cause of con ict. This simple statement seems intuitively reasonable; however, there are some who feel it overstates the importance of resources and the environment as contributors to con ict (Lipschutz, 1995). At

first glance, the availability of water in the Middle East; the depletion of fish stocks off the east coast of Canada; and deforestation in Brazil, Thailand and elsewhere have all been, or have the potential to be, the source of conflict. It has further been suggested that atmospheric change (both global warming and stratospheric ozone depletion) has the potential to cause significant societal disruption, according to the US National Academy of Science (1991) and Myers (1993). In addition, land degradation (or land use change in general) may directly affect society's ability to provide food resources for a growing population, or may indirectly affect other changes, such as global warming. Homer-Dixon (1994) provides some evidence of these relationships and concludes that environmental scarcity (which includes environmental change, population growth, and an unequal distribution of resources) causes violent conflict. While this contention remains open to debate, it is increasingly accepted that environmental degradation is at least a contributor to conflict and insecurity.

Several other authors have attempted to clarify the possible relationships between the environment and conflict or security as well. Wallensteen (1992), for example, proposes a seven-point classification of the connection between environmental destruction and conflict and/or security.

1. Environmental destruction leading to reduced resources available to society and thus resulting in more contention in society at large.
2. Environmental destruction leading to a shift in power between already existing parties.
3. Environmental destruction leading to the formation of new parties, as part of a reaction to environmental destruction.
4. Environmental destruction leading to environmental issues becoming important for established parties.
5. Environmental destruction leading to environmental issues becoming more central in political affairs than other issues in society.
6. Environmental destruction leading to conflict behaviour involving environmentally based groups.
7. Environmental destruction leading to environmentally based conflict behaviour involving environmentally based groups.

It is clear that the environmental variable is perceived to play a number of different roles in its links to violent conflict. This is reflected in Dokken and Gr ger's (1995, 38) definition of environmental conflict as a conflict that involves environmental stress or degradation, whether as cause, consequence or intervening variable; perhaps in combination with social, ethnical or political elements . Libiszewski provides the following operative definition of an environmental conflict:

> Environmental conflicts manifest themselves as political, social, economic, ethnic, religious or territorial conflicts, or conflicts over resources or national interests, or any other type of conflict. They are traditional conflicts induced by environmental degradation. Environmental conflicts are characterised by the principal importance of degradation in one or more of the following fields: overuse of renewable resources, overstrain of the environment's sink capacity (pollution); or impoverishment of the space of living (1992 : 13).

The types of environmental issues examined as potential contributors to conflict are highly varied in terms of; (1) where they occur geographically (2) at what level they occur (local, national, regional, global) (3) the speed with which they occur and (4) the sources for their occurrence. The types of social, political and economic problems produced or exacerbated by environmental scarcity are equally diverse.

A number of researchers prefer the term *environmentally induced con ict* to *environmental con ict* to describe the environment's causal roles. Environmentally induced conflict leaves more latitude for incorporating the multiple causes that are characteristic to all conflicts. Initial evidence points to the environment as an underlying, distant, or background variable that influences or intensifies the conflict (Dokken and Gr ger, 1995). The term *environmentally induced con ict* also assists in the necessary distinction between the causes of conflict and the issues that are being fought over. Unlike conflicts over non-renewable resources or resource wars where the resources themselves are the alleged object of contention, environmentally induced conflicts often are not viewed as a unique kind of conflict. Instead environmental variables contribute to social effects that are the more traditional grievances precipitating conflict (ethnic differences, relative deprivation) (Gurr, 1993; Homer-Dixon, 1994). The environmental variables do not directly cause the conflict *per se* but instead make more salient the variables that can precipitate conflict (Libiszewski, 1992).

Many researchers do expect environmentally induced conflict from renewable resource scarcity to become increasingly frequent. Unlike non-renewable resources, technological innovation and the market have only achieved limited success in developing substitutes for renewable resources (clean air, fertile top soil) according to this line of argument. Therefore, history may not be a good indicator of future potential for conflict. Increasing environmental degradation and resource depletion, the result of increased population, higher use of natural resources *per capita*, and a constant or decreasing supply of environmental amenities, are pushing ecosystems into a highly uncertain and complex ecological future. This expectation of increasing environmental scarcity encourages many observers to predict an accompanying increase in environmentally induced conflict (Homer-Dixon, 1994; Winnefeld and Morris, 1994; Bachler and Spillmann, 1996; Kahl, 1997; Bachler, 1998).

WHAT TYPES OF ENVIRONMENTAL CHANGE AFFECT SECURITY?

The environmental forces that have been presented as contributing to insecurity are many. Environmental calamities such as earthquakes, volcanic eruptions, floods and drought have always presented a threat to human existence, and their human impact has increased considerably in scale as people have moved into disaster-prone areas. The pace of other, human-induced forms of environmental degradation and resource depletion (e.g., deforestation, desertification, land degradation, erosion, salinization, siltation, climate change), while often more gradual, has been growing rapidly in recent decades owing to a combination of increasing demand, improving technological means of exploitation, and the lagging pace of conservation and control. Meanwhile the ability and perhaps also the inclination of people to adapt to environmental stress is increasingly challenged, particularly where resources and environment provide the principal basis of their livelihood, as is the case in much of the developing world.

Types of environmental change/degradation, which may affect security, include the following:

Natural Disasters

Natural disasters include floods, volcanoes and earthquakes. They are usually characterized by a rapid onset, and their impact (destructiveness) is a function of the number of vulnerable people in the region rather than the severity of the disaster, *per se*. Poor people in developing countries are the most affected because they are the most vulnerable. (Droughts, despite a slower onset, are also included in this category.) Recent earthquakes in Pakistan, and flooding in many regions of the world indicate not only the destructiveness of disasters, but their ability to affect large numbers of people.

Cumulative Changes or Slow Onset Changes

Cumulative changes are generally natural processes, occurring at a slower rate, which interact with, and are advanced by, human activities. The processes include deforestation, land degradation, erosion, salinity, siltation, waterlogging, desertification and climate warming. Human-induced soil degradation is one factor that directly affects economic sufficiency in rural areas. Water shortages also may affect security, and Table 1 notes countries which are experiencing (or will soon experience) conditions of water scarcity, where water scarcity is generally considered to be less than $1000\,m^3$ *per capita* per year (this is a rough estimate only; many countries are able to supplement their water supply through expensive alternatives such as desalination (e.g., Kuwait) or importing water (e.g., Singapore)).

Table 1 Annual water availability, 1995 (selected countries)

Country	Annual water availability, 1995 (cubic meters per capita)
Algeria	528
Burundi	563
Egypt	923
Israel	362
Jordan	314
Kuwait	103
Libya	111
Rwanda	792
Saudi Arabia	254
Singapore	211
Tunisia	443
Yemen	359

[a] *Source:* World Resources, 1996–1997 (1996) Oxford University Press, NY.

Do factors such as water scarcity and human-induced soil degradation in and of themselves affect security? The linkage is much more indirect; in most cases, one or more of the following conditions are also present: rapid population growth, economic decline, inequitable distribution of resources, lack of institutional support and political repression.

Accidental Disruptions or Industrial Accidents

This category includes chemical manufacture and transport and nuclear reactor accidents. The two most obvious examples are the nuclear accident at Chernobyl in 1986, and the Union Carbide accident in Bhopal, India, in 1987. Between 1986 and 1992, there were over 75 major chemical accidents which killed almost 4000 people worldwide, injured another 62 000, and displaced over two million. Most of these displacements, however, were temporary.

Development Projects

Development projects such as dams and irrigation systems often involve forced resettlement and affect many aspects of human security. In India, for example, it has been estimated that over 20 million people have been uprooted by development projects in the past three decades. The Three Gorges Dam project in China, expected to displace 1 million people and the Sardar Sarovar Dam project in India are the most notable present examples. Rapid urbanization in some regions of the world is also forcing people from their land; conversion of agricultural land to urban uses has long been a phenomenon in the north, and increasingly this is the case in the South as well.

Conflict and Warfare

Environmental degradation is considered by many to be both a cause and an effect of armed conflict. Although the evidence of wars being fought over the environment is weak (except, of course, over the ownership of land), there is an increasing use of the environment as a weapon of war or, as Gleick (1990) notes, as a strategic tool. One obvious example in this category is the threat by Turkey to restrict the flow of the Euphrates to Syria and Iraq in order to pressure Syria to discontinue its support of Kurdish separatists in Turkey. Other examples include the purposeful discharge of oil into the Persian Gulf during the Gulf War and the destruction of irrigation systems during conflicts in Somalia. Such activities have similar and, indeed, more immediate consequences as do the slow onset changes noted above. But in these cases, it seems clear that the environment is merely a symptom of a larger conflict, and the root cause of any insecurity is the conflict itself, and the reasons behind it.

CRITICAL PERSPECTIVES ON ENVIRONMENTAL SECURITY

Throughout the environment and security debate, many writers have been critical of linking the two terms. Researchers and analysts from a traditional security perspective tend to discount the role that environmental degradation or resource depletion plays in precipitating violent conflict. They further argue that broadening the definition of security to include a laundry list of modifiers (environmental, ecological, economic, food, human, comprehensive, common) undercuts the term's utility by making it mean something different to each of multiple constituencies. Military critics of tying environment and security together also claim that performing environmental missions takes time and resources away from preparations for the traditional war fighting mission and therefore undermines preparedness and effectiveness in battle.

Environmental critics also claim that there is little evidence to support the argument that environmental degradation or resource depletion has a significant role in causing violent conflict, and especially interstate conflict. Furthermore, the methodological shortcomings of much of the research undermine the findings that do support a case for linking environment and violent conflict. Critics also fear that propagating the term *environmental security* could lead to the militarization of the environment rather than the greening of security. Military institutions, instead of undergoing fundamental change to re ect new security priorities, are more likely to co-opt and weaken the non-statist, non-threat based, cooperative ethic of environmental rescue. This criticism is reinforced by the perception that security institutions are searching for new missions to justify their high Cold War funding levels. Environmental critics also decry the conception of environmental security that has developed as an uniquely Northern and Western term; it is viewed as unacceptable to the South as a paradigm for environmental problems.

CRITICISMS OF REDEFINING SECURITY IN ENVIRONMENTAL TERMS

Critics commonly point out the greater likelihood of the militarization of the environment than the greening of security as northern security institutions search for new missions in the post Cold War period. The term as it is used in this context is Wand ver s (1995). Others share the concern expressed by the term with their own distinct critiques (Deudney, 1990, 1991; Dalby, 1992, 1994; Finger, 1991, 1994; Conca, 1994; Lipschutz, 1995; Gleditsch, 1997). With environmental security being used as a political slogan to gain attention for the environment, the risk is that the historically powerful military institutions will co-opt the green rhetoric rather than willingly giving up resources to more effectively address the new threats to environmental security. Some observers suspect that security may be redened to include environmental considerations merely at a rhetorical level (i.e., national security strategies) but would fail to produce a simultaneous reorientation or dismantling of security institutions and mindsets (Conca, 1994; Finger, 1994; Kakonen, 1994).

More broadly, some critics charge that environmental security encompasses too many problems and threats; for example, problems associated with infectious disease, global warming, environmental damage during war, deforestation, water scarcity, and nuclear waste. With such diverse problems included as the focus of environmental security, the term loses meaning and utility as an analytical tool because there is no delineation of what is included and what is not (Deudney, 1991; Dokken and Gr ger, 1995; W ver, 1995). Instead, linking environment and security merely represents a normative slogan conveying the urgency of addressing global problems in determining the priority of political battles (Levy, 1995a,b).

This criticism illustrates how environmental security is held to different standards for different purposes. Those in academia criticising environmental security as a normative political slogan are asking that the term perform as a sharpened theoretical tool. They discount the early calls for rede ning security, sometimes termed the rst wave of environmental security literature, as undeveloped, a conceptual trick or minimally useful (Levy, 1995; Gleditsch, 1997). From a policy perspective, the rhetorical use of the term is less troubling than the failure by its adherents to suggest speci c policy priorities and interventions which would accompany any rede nition.

The above criticisms arise from authors who assign a high priority to the importance of coming to grips with global problems. However, others continue to find considerable utility in the purely statist and militaristic security assumptions and therefore oppose widening the purview of security to new and different threats (Walt, 1991). They argue that while the Cold War has ended and the dangers of a bilateral stand off have abated, emerging military threats demand a traditional definition of security with continued priority support for the military. Nuclear, chemical and biological weapons proliferation, terrorism, and ethnic conflict remain reasons enough not to dilute the definition of security with peripheral, non-military concerns.

There are also criticisms of environmental security, which are based on the perspective of the South. Egyptian diplomat Saad (1991), for example, argues that invoking the term *environmental security* represents a new Northern justification for continuing the inequitable power relationship between North and South. She worries that wealthy countries of the North can afford to care about the environment and will undermine the international legal principle of sovereignty in the name of a higher goal called *environmental security*. The principle of sovereignty, from the perspective of the South, provides some defense against exploitation by recognizing each state, no matter how weak in capabilities, as the legitimate authority for control over the resources within its borders.

By this reasoning, Northern states may be tempted, in the name of environmental security, to try to dictate the patterns of natural resource usage, development priorities and population policies to developing countries. The stability and welfare of some states rest on sets of social power relationships surrounding the utilization of natural resources (large, politically empowered landowners in Brazil, for example). The elite in certain countries may therefore find an alteration of past social bargains, for the sake of environmental conservation, to be a larger threat to state security than the environmental destruction itself (Conca *et al.*, 1995). Such perspectives raise barriers to obtaining the cooperation of the South with respect to addressing global environmental problems under the guise of environmental security. This argument implicitly recognizes the importance of national or regional perspectives in defining or operationalising environmental security. The content and meaning of environmental security varies across nations and regions. These differences present difficulties when trying to mobilize action on a global scale under the label of environmental security.

> Twenty years ago, the emphasis was on ending the pollution that the industrialized North had been inflicting on the nations of the South. The goals were clean air and water and arable land, the requisites of a decent life; and the modality was international cooperation. Today, however, the North has seized hold of environmental issues by using them to cloak its own security concerns.
>
> (Saad, 1995, 273)

The prominent focus on environmental stress and violent conflict in the environmental security literature also presents an additional cause for Southern suspicion of the term. The North's concern with environmentally induced conflict can easily be viewed as a convenient means to distract attention from Northern environmental problems. High rates of consumption in the North or the historical depletion of resources do not figure prominently in causal models, yet they are integral elements in the larger environmental picture. Global issues such as climate change and stratospheric ozone depletion are not often recognised as salient issues in environmentally induced conflict because their long time lines guarantee marginal relationships to violent conflict. The sources for these global problems tend to emanate disproportionately from the North. Furthermore, Northern interest in environment and conflict linkages often extends only to a concern for regime stability and international security implications. The operationalization of environmental security within the traditional security institutions may stop short of fundamental interest in Southern problems of resource degradation and depletion, poverty, and the inequitable distribution of wealth.

ENVIRONMENT, HUMAN SECURITY AND SUSTAINABLE DEVELOPMENT

The World Commission on Environment and Development (WCED, 1987) and the United Nations Conference on Environment and Development (UNCED) held in Rio de Janeiro in 1992, have reinforced the notion that economic development is necessary to improve quality of life and well being. However, to be sustainable, this development must incorporate environmental, social, political and economic considerations consistent with a strong commitment to equity and a concern for resource limitations and the viability of ecosystems. Indeed, it is well accepted that environmental, social, political, and economic systems are interlinked; any actions involving one of these necessarily affects the others. Even our perspective on environmental change is constrained by social, economic, political and cultural factors; for example no universal definitions of degradation, scarcity or even pollution exist. These terms are defined by context and value systems.

This multi-dimensional characteristic implies that the question of whether environmental degradation is a cause of conflict is misguided. What we should be asking is whether environmental change (linked to a range of other factors) contributes to social tensions, and whether these tensions

may result in instabilities (and, quite possibly, conflict). The past decade of discussions on sustainability have further taught us that any assessment of how environment and security are related must take into account how these other factors influence sustainable development and, in particular, how growth must acquire both environmental and social meaning.

A particularly relevant criticism of the existing literature is that the debate on environment and security has largely ignored a discussion of how environmental security fits into the broader context of international development, social welfare and sustainable development. Although there have been calls for attention to comprehensive security (Westing, 1986) and a multilevel approach to security (Gr ger, 1996), most authors fail to recognize the one overwhelming argument in favor of linking environment and security: environmental problems must always be presented from within a broader perspective that encompasses various forms of equity, including world poverty . This is because poverty and inequity are two of the key factors contributing both to environmental change and to tension and insecurity throughout the world. For research on environment and security to gain social legitimacy, it must incorporate elements of social justice. In turn, for development to be sustainable, it must include the protection of human rights. These concerns are embodied in the broader concept of environment and human security. What is human security?

A further issue that has received little attention in the literature is that space matters. Some authors have recently called for a reshaping of our spatial perspective on linking environment and security. There has been a strong plea for regionalism (sub-national and supra-national) which should be decided on the basis of eco-geography (Dokken and Gr ger, 1995). Here eco-geography implies that the region of concern is defined by ecological boundaries (e.g., watershed or eco-region) and not simply political boundaries. Regionalism need not be confined to ecological regions; it may incorporate ethnic or other factors as well. Perelet (1994) also argues that environmental security problems must focus on the ecosystem level. It has also been suggested that research must focus on the local level or at least at the lowest possible spatial scale for the insecurity/conflict being studied. This has been termed the *subsidiarity principle* (Mische, 1989; Dokken and Gr ger, 1995). For example, the relationship between environment and poverty must be viewed locally and in the context of broader human security dimensions. This is absolutely crucial if we desire to incorporate local knowledge and make the process of development a truly participatory one. These two trends have stimulated further questions on the relationship between environmental security and the goal of a sustainable society.

ENVIRONMENTAL CHANGE AND HUMAN SECURITY

So where does this leave the link between environment and security? Coincidental with the claim that environmental degradation may be a threat to conflict/security has been the move to interpret security more broadly to encompass notions of human welfare, where there is no immediate prospect of armed conflict. A better way of framing the issue might be in terms of insecurities (Gr ger, 1996; Lonergan, 1997). As noted above, the appeal of a broader notion of security, framed in terms of insecurities, is obvious. It allows one to include the key issues of equity and impoverishment, along with other non-conventional threats to security. It also is multi-dimensional (and multi-level according to Gr ger, 1996); this allows us to introduce into the debate issues of human, environmental and social rights, acting at all spatial levels, from the individual to the global. However, as noted above, many authors have expressed significant concerns with using a broader definition of the term *security*.

But even accounting for these concerns, it is clear that there remains a need to reassess our traditional perspectives of security and focus more on the insecurities posed by non-conventional threats. We must also recognize that society is in a profound period of rapid social, economic and environmental change. While continued growth in gross world product may imply improvement in social welfare, there are many disturbing indications that this is not the case for much of the world's population. The number of wars has increased dramatically over the past three decades. However, since 1960, these have primarily been intrastate conflicts, where inequity is a key feature.

This leads to the inescapable conclusion that while environmental degradation and even resource depletion as causes of conflict may be overstated, it is undeniable that increasing inequities in society are likely a major source of these (and other) non-conventional threats to individual security. At the same time, there is recognition that traditional approaches to national security may not, in turn, ensure the security of individuals and communities. This applies to global efforts aimed at curbing global warming and ozone depletion to local sustainable development initiatives. Again, this calls for a broader notion of security, one that focuses on the human element, despite the ambiguities and difficulties this may cause.

See also: Environmental Changes Driven by Civil Conflict and War, Volume 3; The Environment and Violent Conflict, Volume 5.

REFERENCES

Bachler, G (1998) Why Environmental Transformation Causes Violence, in *Environmental Change and Security Project Report 4*, Woodrow Wilson Center, Washington, DC, 24–44.

Bachler, G and Spillman, K R (1996) *Environmental Degradation as a Cause of War* Vol. 2–3, Ruegger Verlag, Zurich.

Brown, H (1954) *The Challenge of Man's Future*, Viking, New York.

Choucri, N (1991) Resource Constraints as Causes of Conflict, *Ecodecision*, 2, 52–55.

Conca, K (1994) In the Name of Sustainability: Peace Studies and Environmental Discourse, *Peace Change*, 19, 91–113.

Conca, K, Alberty, M, and Dabelko, G (1995) From Ecological Conflict to Environmental Security? in *Green Planet Blues: Environmental Politics from Stockholm to Rio*, eds K Conca, M Alberty, and G Dabelko, Westview Press, Boulder, CO, 239–244.

Dalby, S (1992) Security, Modernity, Ecology: The Dilemmas of Post Cold War Security Discourse, *Alternatives*, 17, 95–134.

Dalby, S (1994) The Politics of Environmental Security, in *Green Security or Militarized Environment*, ed Jyrki Kakonen, Dartmouth Publishing, Brookfield, 25–54.

Deudney, D (1990) The Case Against Linking Environmental Degradation and National Security, *Millennium: J. Int. Stud.*, 19(3), 461–476.

Deudney, D (1991) Environment and Security: Muddled Thinking, *Bull. At. Sci.*, (April), 22–28.

Deudney, D and Matthew, R A (1999) *Contested Ground: Security and Con ict in the New Environmental Politics*, SUNY Press, Albany, NY.

Diederich, Jr, M D (1992) Law of War' and Ecology – a Proposal for a Workable Approach to Protecting the Environment Through the Law of War, *Military Law Rev.*, 136, 138–160.

Dokken, K and Gr ger, N (1995) *The Concept of Environmental Security - Political Slogan or Analytical Tool*? International Peace Research Institute, Oslo (PRIO), Report 2.

Durham, W H (1979) *Scarcity and Survival in Central America: Ecological Origins of the Soccer War*, Stanford University Press, Palo Alto, CA.

Finger, M (1991) The Military, the Nation State and the Environment, *Ecologist*, 21(5), 220–225.

Finger, M (1994) Global Environmental Degradation and the Military, in *Green Security or Militarized Environment*, ed Jyrki Kakonen, Dartmouth Publishing, Brookfield, 169–192.

Gleditsch, N P (1992) *Conversion and the Environment, Proceedings of a Seminar in Perm, Russia, November 1991*, PRIO Report No. 2, International Peace Research Institute, Oslo.

Gleditsch, N P (1997) *Con ict and the Environment*, NATO ASI Series Vol. 33, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Gleick, P (1989) The Implications of Global Climate Changes for International Security, *Climate Change*, 15, 303–325.

Gleick, P (1990) Environment, Resources and International Security and Politics, in *Science and International Security: Responding to a Changing World*, ed E H Arnett, American Association for the Advancement of Science, Washington, DC, 501–523.

Gleick, P (1991) Environment and Security: The Clear Connections, *Bull. At. Sci.*, (April), 17–21.

Gr ger, N and Smith, D (1994) *Environment Poverty and Con ict*, International Peace Research Institute, Oslo (PRIO), Report 2.

Gr ger, N (1996) Environmental Security? *J. Peace Res.*, 33(1), 109–116.

Gurr, T R (1993) *Minorities at Risk: A Global View of Ethnopolitical Con icts*, US Institute of Peace, Washington, DC.

Hartman, B (1998) Population, Environment and Security: A New Trinity, *Environ. Urban.*, 10(2), 113–127.

Hauge, W and Ellingsen, T (1998) Beyond Environmental Scarcity: Causal Pathways to Conflict, *J. Peace Res.*, 35(3), 299–317.

Homer-Dixon, T F (1991) On the Threshold: Environmental Changes as Causes of Acute Conflict, *Int. Secur.*, 19(1), 5–40.

Homer-Dixon, T F (1994) Environmental Scarcities and Violent Conflict: Evidence from Cases, *Int. Secur.*, 16(2), 76–116.

Homer-Dixon, T F, Boutwell, J, and Rathjens, G W (1993) Environmental Change and Violent Conflict, *Sci. Am.*, 268(2), 38–45.

Kandakandonen, J (1994) *Green Security or Militarized Environment*, Dartmouth Publishing, Brookfield.

Kahl, C H (1997) States, Scarcity, and Civil Strife in the Developing World, *IWPS Environment and Security Project*, September, 1–60.

Levy, M A (1995a) Is the Environment a National Security Issue? *Int. Secur.*, 20(2), 35–62.

Levy, M A (1995b) Time for a Third Wave of Environment and Security Scholarship? *The Environmental Change and Security Project Report*, 1, 44–46.

Libiszewski, S (1992) *What is an Environmental Con ict?* Environment and Conflicts Project (ENCOP), Occasional Paper No. 6, Center for Security Studies and Conflict Research, Zurich.

Lipschutz, R D (1995) *On Security*, Columbia University Press, New York.

Lodgaard, S and Hjort af Ornas, A (1992) *The Environment and International Security*, International Peace Research Institute, Oslo (PRIO), Report 3.

Lonergan, S (1997) Water Resources and Conflict: Examples from the Middle East, in *Con ict and the Environment*, ed N P Gleditsch, NATO ASI Series Vol. 33, Kluwer Academic Publishers, Dordrecht, 375–384.

Lonergan, S C and Kavanagh, B (1991) Climate Change, Water Resources and Security in the Middle East, *Global Environ. Change*, (September), 272–290.

Mathews, J T (1989) Redefining Security, *Foreign Affairs*, 68, 162–177.

Mische, P M (1989) Ecological Security and the Need to Reconceptualize Sovereignty, *Alternatives*, 14(4), 389–428.

Molv r, R K (1991) Environmentally Induced Conflicts? A Discussion Based on Studies from the Horn of Africa, *Bull. Peace Proposals*, 22, 175–188.

Myers, N (1989) Environment and Security, *Foreign Policy*, 74(Spring), 23–41.

Myers, N (1993) *Ultimate Security: The Environmental basis of Political Stability*, WW Norton, New York.

National Academy of Sciences (US) (1991) *Policy Implications of Global Warming*, National Academy of Sciences, Washington, DC.

Ophuls, W (1977) *Ecology and the Politics of Scarcity: Prologue to a Political Theory of the Steady State*, W H Freeman, San Francisco, CA.

Osborn, F (1953) *Our Plundered Planet*, Little Brown, Boston, MA.

Perelet, R (1994) The Environment as a Security Issue, in *The Environment: Towards a Sustainable Future*, Dutch Committee for Long-term Environmental Policy, Kluwer Academic Publishers, Boston, MA.

Renner, M (1989) *National Security: The Economic and Environmental Dimensions*, Worldwatch Paper 89, Worldwatch Institute, Washington, DC.

Saad, S (1995) For Whose Bene t? Rede ning Security in *Green Planet Blues: Environmental Politics from Stockholm to Rio*, eds K Conca, M Alberty, and G D Dabelko, Westview Press, Boulder, CO, 273 275.

Spillman, K R and Bachler, G (1995) *Environmental Crisis: Regional Con icts and Ways of Cooperation*, Environment and Con icts Project Occasional Paper No. 14, Center for Security Studies and Con ict Research, Zurich.

Ullman, R H (1983) Rede ning Security, *Int. Secur.*, **8**(1), 129 153.

Wallensteen, P (1992) Environmental Destruction and Serious Social Con ict: Developing a Research Design, in *The Environment and International Security*, eds S Lodgaard and A Hjort af Ornandas, (PRIO Report No. 3, 47 54), International Peace Research Institute (PRIO), Oslo.

Walt, S M (1991) The Renaissance of Security Studies, *Int. Stud. Q.*, **35**(2), 211 239.

Wand ver, O (1995) Securitization and Desecuritization, in *On Security*, ed R Lipscutz, Columbia University Press, New York, 46 86.

Westing, A H (1986) *Global Resources and International Con ict: Environmental Factors in Strategic Policy and Action*, Oxford University Press, Oxford.

Westing, A H (1989) The Environmental Component of Comprehensive Security, *Bull. Peace Proposals*, **20**(2), 129 134.

Winnefeld, J A and Morris, M E (1994) *Where Environmental Concerns and Security Strategies Meet*, Rand, Santa Monica, CA.

World Commission on Environment and Development (WCED) (1987) *Our Common Future*, Oxford University Press, Oxford.

Environmental Sociology

Amar Wahab
University of Toronto, Toronto, Canada

Environmental sociology is the study of reciprocal interactions between the physical environment, social organization and social behavior (Sydenstricker-Neto, 1997). Unlike a sociology of environment perspective, which applies traditional social theory to the study of environmental issues, it focuses on the interaction between social and environmental systems. According to Hannigan (1995), the field of environmental sociology still lacks a seminal work which could lift it into the mainstream of debate in the broader field of sociology. Sociologists therefore engage in environmental issues with the aim of enriching mainstream sociological theory, rather than attempting to ambitiously transform the discipline. A defining feature is the work of Dunlap and Catton to replace an anthropocentric approach (Human Exemptionalism Paradigm) with their ecocentric, New Ecological Paradigm. Although this approach has been viewed as over-ambitious, it has stimulated and catalyzed a diversity of paradigms and theoretical approaches to the study of environmental sociology.

Buttel (1987) has cited ve key areas of environmental scholarship:

1. Dunlap and Catton s new human ecology;
2. environmental attitudes, values and behaviors;
3. the environmental movement;
4. technological risk and risk assessment;
5. the political economy of the environment and environmental politics.

Alternatively, Hannigan (1995) has identi ed two distinct problems as central to the existing literature on environmental sociology: (1) the causes of environmental destruction; and (2) the rise of environmental consciousness and movements.

The new ecological paradigm explanation of environmental destruction focuses on the competing functions of environment as supply depot, living space and waste depository within an assumed global carrying capacity (Dunlap and Catton, 1979). This model is concerned with the increasing globalization of environmental problems and the increasing production of waste, but is criticized for its limited incorporation of social de nition. An in uential alternative explanation, deriving from political economy, emerges out of Marxist ideology which identi es industrial capitalism, in which the corporation and the state can be seen as being in opposition to its own citizenry, as the major cause of environmental degradation (Hannigan, 1995). In this explanation, capitalism is said to generate *the treadmill of production* (Schnaiberg, 1980), which operates regardless of the carrying capacity of the ecosystem.

The second problem currently tackled by environmental sociology, looking at environmental consciousness and movements works with an array of hypotheses. Perhaps one of the more in uential of these is the new middle-class thesis which has focused on the conceptualization and constitution of new social movements (NSMs), i.e., activism by individuals and groups who are deemed to possess post-materialist values. According to Hannigan (1995), there have been environmental sociologists such as Steninmetz (1994) who have contested the class basis of NSMs

activism and others such as Habermas (1987) and Beck (1992) who have suggested that NSMs have emerged in marginal spaces, in a culture of political closure organized by the corporate state (which may help to explain their relative ineffectiveness). Hannigan has also employed a social constructionist perspective, which focuses on the social, political and cultural processes by which environmental conditions are defined as risky and actionable. He views the construction of environmental problems and conflicts as hinged on the processes of assembling, presenting and contesting claims.

It is important to note though, that the discourse on environmental sociology has been primarily focused on environmental problems in the economic North, with very little reconceptualization based on the works of third world scholars and integration of development discourses in the South. Criticism surrounding some of the silences regarding interlocking systems of race, space, gender, sexuality, class and other social axes of power that are embedded in the rubric of alternative discourses is less than forthcoming. Some of the themes which have been popularized by environmental sociology involve environmental problems related to agriculture, energy and fuels, hazards and risks, leisure/recreation, natural resources, social impact assessment, and sustainable development. Some central works in the field of environmental sociology include Dunlap and Canton (1979), Hannigan (1995) and Schnaiberg and Gould (1994).

REFERENCES

Beck, U (1992) *The Risk Society*, Sage, London.

Buttel, F (1987) New Directions in Environmental Sociology, *Annu. Rev. Sociol.*, 13, 465–488.

Dunlap, R E and Catton, Jr, W R (1979) Environmental Sociology, *Annu. Rev. Sociol.*, 5, 243–273.

Habermas, J (1987) *The Theory of Communicative Action*, Vol. 2, Polity Press, Cambridge.

Hannigan, J (1995) *Environmental Sociology: A Social Constructionist Perspective*, Routledge, New York.

Schnaiberg, A and Gould, K A (1994) *Environment and Society*, St. Martin's Press, New York.

Steinmetz, G (1994) Regulation Theory, Post-Marxism, and the New Social Movements, *Comp. Stud. Soc. Hist.*, 36(1), 176–212.

Sydenstricker-Neto, J (1997) *Environmental Sociology: a Resource Page*, http://trochim.human.cornell.edu/gallery/Neto/Envsoc1.html.

Environmentalism

see Art and the Environment (**Volume 5**); Literature and the Environment (**Volume 5**)

Equity

Henry Shue
Cornell University, Ithaca, NY, USA

Intense debates about the requirements of equity have been at the very heart of the international negotiations about what to do about rapid climate change, ever since those negotiations began in the 1980s. In this context, equity, distributive justice, and fairness are alternative terms for the same set of requirements. Economists and lawyers tend to use the term equity, and philosophers normally discuss distributive justice, but these disciplinary preferences in terminology have no systematic signi cance. In both cases, the professionals are talking about what ordinary people call fairness.

Equity, or fairness - the terms are employed interchangeably in this article - is a feature that the division of an aggregate or total among multiple parties often ought to have. Fairness is an ethical requirement to be met by speci c, concrete allocations, divisions, or distributions of bene ts and burdens. The fundamental point about fairness is that what matters is not the average per capita distribution of some aggregate, which tends to be the focus of economic analysis and which, after all, is a merely theoretical distribution, but the relative sizes of the actual shares of the aggregate that various parties each have.

Multiple questions usually arise about fairness. What is a fair result (substantive fairness)? What is a fair process (procedural fairness)? What connection, if any, is there between substantive and procedural? In the case of climate change, the multiple tasks required can each be performed by means of fair or unfair processes, and with fair or unfair results; these tasks include the allocation of the costs of adaptation to actual climate change, the allocation of the costs of the mitigation of potential climate change, and the allocation of the limited total of emissions of greenhouse gases compatible with mitigating climate change. Winners and losers will be spread both across the geographical dimension, including the international, and along the temporal dimension, including the intergenerational. How they will be spread is uncertain but will be deeply affected by the political decisions made under the Framework Convention on Climate Change. Many of those decisions can consequently be assessed as fair or unfair. Since the maximum severity of climate change will depend upon the date of the technological transition away from the energy technologies that worsen the problem, the fundamental question of intergenerational fairness is whether policy options, like the exibility mechanisms in the Kyoto Protocol, advance or delay this date of transition.

MULTIPLE QUESTIONS

Fairness can be a feature of an outcome: the question is then whether the various parties among whom some total is to be divided each receives a fair share. That is the question of *substantive fairness*. On the other hand, the question of fairness sometimes applies to the process that produces the outcome: this question is whether the manner by which the total is divided among the various parties is fair. This is the question of *procedural fairness*.

Complex and important questions arise about the relation, if any, between procedural fairness and substantive fairness. Sometimes, in one kind of case, the basis for judgments of substantive fairness and the standards for judgments of procedural fairness are independent of each other. Substantive fairness and procedural fairness can each be specified separately. For example, if a murder has been committed and someone is to be punished, the only substantively fair outcome is for the punishment to be inflicted upon the murderer, the guilty person. Most societies have elaborate standards for the procedural fairness of murder trials – standards like a presumption of innocence for the accused, and adequate opportunity for the accused to make a defense. Yet, tragically, it is entirely possible for a perfectly fair trial – a trial satisfying all the standards of procedural fairness – to lead to the conviction of an innocent person, which is a substantively unfair outcome: the punishment of the innocent. Even the procedurally fairest systems of criminal law are fallible and sometimes reach substantively unfair outcomes by convicting innocent persons. Here procedural fairness does not guarantee substantive fairness.

By contrast, in other kinds of cases there is no basis for judging the fairness of outcomes that is independent of the fairness of the procedure that produces the outcome. A lottery is the clearest example. Suppose a prize is to be awarded by having everyone who wants the prize place a piece of paper with his/her name on it in a hat and having someone who is blindfolded pick one name. It makes no sense for someone whose name was not selected to say after the drawing that the outcome is unfair because the wrong person won the prize. There is no wrong person if the lottery was conducted fairly, as there is when an innocent person is convicted. The only standard of fairness in the case of a lottery is procedural fairness: the right outcome is whatever outcome results if the process is conducted fairly. Here fairness is randomness. In the case of a lottery, by contrast with a criminal trial, there is no such thing as a person who, in advance of the drawing, somehow ought to be selected for the prize, analogous to the person who, in advance of the criminal trial, ought to be chosen for the punishment.

Sometimes, of course, when the prize in a lottery goes to someone who does not need it, and other participants in the lottery do need it, objectors say too bad, the prize should have gone to someone who needed it . If this is to have any solid basis, it must be construed as an objection to this use of a lottery process. If the prize genuinely ought to have gone to someone needy, then no lottery should have been conducted at all (or participation in the lottery should have been restricted by the rules for the lottery to people who needed the prize). The objectors, then, are contending that what had been treated as a prize in a lottery ought not to have been made a lottery prize because, in fact, a relevant basis for judgments of substantive fairness, namely need, was available; therefore, instead of having held a lottery that awarded the prize randomly, a completely different kind of process should have been used to allocate the prize to someone who needed it. Defenders of the lottery might be able to reply that although some things sometimes ought to be allocated according to need, this lottery prize is not one of them, and nothing is wrong with using a fair lottery to award it. This is now a debate about a choice among processes and about the relevance, if any, of a particular substantive basis (need) to that choice of process (see Box 1).

MULTIPLE TASKS

The global effort to deal with climate change simultaneously faces several challenges that can be distinguished from each other in theory but must all be handled in practice by means that are at least complementary, if not mutually supportive. Each of these tasks raises all three kinds of questions about equity. Since climate change is already occurring, and since current concentrations of greenhouse gases in the atmosphere constitute a commitment to additional climate change, one task will obviously be adaptation to climate change. Adaptation will require expensive adjustments, and one cluster of questions about equity concerns the division of the costs, and the benefits, of adaptation (see Box 2).

MULTIPLE DIMENSIONS

The negotiations about whether to take effective action concerning climate change are conducted by national governments; and measurement and reporting of sources, sinks, and net emissions are all categorized in national terms. Consequently, it is difficult to overlook the fact that equity issues have dimensions that are spatial: the international distributions of costs and of emissions. Many debates about the international division of the benefits and burdens have been pure conflicts of national interest, even when sometimes clothed in the rhetoric of equity. International equity will be attained only if powerful national states aim explicitly at it; nothing supports any hope that the pure clash of national interest will somehow lead to international equity. Nevertheless, even the nations that are too weak to defend

Box 1 Multiple questions

Substance

Procedure

Relation between first two

In the case of climate change, some of the fundamental issues about equity concern disputes of the kinds illustrated in the text. Many economists assume that the use of efficient markets is analogous to the use of a fair lottery, in that whatever distribution results is unquestionably acceptable; this is equivalent to assuming that there are no well-grounded principles of substantive fairness by which market outcomes can be assessed. Much of the support for the flexibility mechanisms in the Kyoto Protocol (Clean Development Mechanism [Article 12], Joint Implementation [Article 6], and Trading in Allowed Amounts of Emissions [Article 17]) appears to rest on such an unargued assumption that whatever consequences for the allocations of costs and of emissions are generated by fair processes are unchallengeable on grounds of substantive fairness. This flies in the face of the solidly grounded view that, for as long as the generation of greenhouse gas emissions is essential to human subsistence, every person needs to benefit from at least a minimal level of emissions. If the global ceiling on total emissions is sufficiently low, the only distribution that could possibly guarantee everyone the necessary minimum would be an equal distribution. To guarantee that minimum, market processes need to be restrained by universal inalienable rights to benefit from a minimum level of emissions.

One needs to determine for each situation not only: (a) whether grounds for judging substantive fairness are available – and if so, what they are; and (b) whether standards for procedural fairness are available – and if so, what they are; but also (c) in which cases internally fair processes are adequate, meaning that whatever outcome they produce is therefore acceptable, and in which cases substantive fairness should be the basis for adopting one kind of process rather than another, meaning that a process, even if internally fair, is fully adequate only if it is the best means to fair outcomes.

themselves successfully against unfair treatment at least have official representatives (see Box 3).

INTERGENERATIONAL EQUITY AND INNOVATIVE TECHNOLOGY

A deeper point about technology is buried in the second tendency. The basic decisions concerning climate change are choices about energy technology. Choices about which nations are to bear the current costs of technological transitions obviously implicate international equity; and choices about the rate of technological innovation deeply affect intergenerational equity in a less obvious way. The second tendency ingrained in current approaches to climate change is not to abandon fossil-fuel technologies until their continued retention becomes too expensive for the current generation, and meanwhile to avoid investing very large amounts in research or development of alternative energy technologies. This clearly favors the current generation strongly, although the issue of fairness is whether this favoritism is unreasonably strong. This tendency and the first tendency (to select lowest-cost-first solutions) are related; the first supports the second – it is believed to be cheaper to continue to spread into the less-developed states the current form of industrialization that depends on fossil-fuel than to introduce alternatives. However, many other factors, including entrenched political and economic interests and lack of imagination, also underwrite retention of the familiar fossil-fuel technology that is the most important contributor to climate change. The primary issue of intergenerational equity here is not, however, merely the obvious one of deferring costs so that they fall upon generations later than one's own, although that is an issue too.

Sooner or later, human societies will make a transition from reliance dominantly upon fossil-fuel-driven technologies to reliance on one or more alternative energy technologies. The issues are: at what point in history should this transition be made (substantive intergenerational equity)? how should this date be determined (procedural intergenerational equity)? The current tendency is to allow the date to be determined by relative prices in energy markets. The general idea, according to abstract assumptions in economic theory, is that at some time the prices of fossil fuels will have risen high enough, and the price of some alternative energy will have dropped low enough, that rational consumers will switch to the alternative energy source. The price of alternative energy will have dropped, relative to fossil fuel, because at some point entrepreneurs will have realized that investments in this kind of technological innovation could then be profitable in the relatively short-term. The date of the transition, according to most economists, ought to be determined by the prices – prices are the right basis given the moral assumptions underlying economic theory.

Consideration of intergenerational equity, on the other hand, might suggest that, depending upon many complexities that cannot be explored here, political action ought to be taken to advance the transition to an earlier date. One critical issue is whether human societies can continue to rely predominantly upon fossil fuels for a significantly longer time without making the severity of climate change greater than it otherwise would have needed to be. In other words,

Box 2 Multiple tasks

Costs of adaptation

Costs of mitigation

Allowed emissions

Relation among first three

One means toward controlling the costs of adaptation is abatement, or mitigation, of climate change. Adaptation tackles the effects of climate change; mitigation attempts to slow the rate of climate change itself. Obviously mitigation and adaptation are not separate matters; for example, the less mitigation accomplished, the more adaptation required, other things being equal. Nevertheless, the allocation of the costs and the benefits of mitigation need to be considered independently from the costs and the benefits of adaptation because, to mention only one consideration, any reasonable basis for dividing the costs of mitigation might be different from any reasonable basis for dividing the costs of adaptation. At least in theory it is possible that the parties who ought to bear most of the costs of adaptation are different from those who ought to bear most of the costs of mitigation.

These first two tasks fundamentally concern the fair sharing of financial burdens and benefits. The substantive question, in both cases (adaptation and mitigation), is: who pays? The procedural question in both is: how is it to be determined who pays? The third profoundly important allocative task is the distribution of greenhouse-gas emissions themselves, and, thus, of the activities that produce the emissions. Who is allowed to emit what portion of the global total of allowable emissions? If rapid climate change is to be avoided, an enforceable ceiling must be imposed by the international community on the global aggregate of emissions. That total must, in turn, be divided somehow among the people of the world. For as long as economic activity depends upon the burning of the fossil fuels that emit greenhouse gases, the fair distribution of this limited total has unsurpassed importance. One crucial consideration is that powerful states are unlikely to restrict themselves to the shares they are allocated unless their allocation seems reasonably fair. The simplest way for the distribution to seem fair is for it transparently to be fair. Equity may be politically essential to effectiveness. A second crucial consideration is that, with the overall total severely limited, those who exceed their own fair share will, in fact, be depriving others of their minimum entitlement, unless those others make up for the encroachment by exceeding the mandated total and thereby frustrate the ultimate purpose of the entire effort.

Finally, just as the realization that questions of equity must be analyzed into substantive and procedural leads to the further question about how the two are to be related, the realization that the tasks of equity need to be distinguished into fair allocation of adaptation costs, fair allocation of mitigation costs, and fair allocation of the emissions themselves leads to the further question: how, and/or on what basis, are these three tasks to be related? Is there a fair, coordinated solution to all three problems?

Box 3 Multiple dimensions

Spatial (International)

Temporal (Intergenerational)

Relation between first two: date of technological transition

The temporal dimension of equity is more easily forgotten, but nothing affects more human beings than the intergenerational distributions of costs and of emissions. Indeed, future generations are usually the unrepresented, and, therefore, most vulnerable parties of all, in spite of the fact that the unnoticed other side of the coin in most of the decisions made about how much burden the current generation will handle, is how much is consequently left for succeeding generations to bear. For example, every decision about the relative emphasis to be placed on (present) mitigation efforts as opposed to (future) adaptation efforts is also a decision about the intergenerational allocation of effort and resources. Two tendencies deeply engrained in the current approaches to climate change shift the burdens toward our grandchildren, and their grandchildren, especially strongly.

The first tendency is the largely unquestioned assumption that it is only rational always to pursue least-cost-first solutions to mitigation. While this may often be rational if no interests are given full weight except the interests of the current generation, it is far from evident that such solutions are rational in light of the interests of future generations or are equitable to them. The danger to the future has two aspects. First, if the level of expenditures were fixed independently, then the choice of a least-cost-first approach would at least mean that, for the given expenditure, the most would be accomplished. In reality, however, nothing keeps expenditures fixed, and consequently, many parties will choose to accomplish the same amount of mitigation for a lower expenditure rather than accomplishing more mitigation for the same expenditure. The result is not that more is accomplished to slow climate change but only that less is spent now. Second, least-cost-first can mean higher-cost-later, unless technological advances that reduce the costs of the currently more expensive tasks at least keep pace with the passage of time. Spontaneous technological advances might occur in time to contain costs for future generations, but simply to assume net technological progress would be a largely groundless gamble for which future generations will pay if it turns out badly.

is it the case that the more distant the date of the transition away from fossil fuels, the more severe climate change will become before it stops worsening (because of the greater maximum atmospheric concentration of greenhouse gases)? This depends upon whether, during the period of the continued predominant reliance on fossil fuels, society is willing and able to cap its greenhouse gas emissions at a level that does not make climate change progressively worse by continuing to expand the atmospheric concentration. This is an open question, the answer to which will be determined by decisions yet to be made.

However, those decisions could make matters much worse for future generations than they would be if a political initiative was taken to move the date of transition to an earlier time. If: (1) the choice is made to take no political action to advance the date even though the price of alternative technologies becomes competitive with the price of fossil-fuel technology; and (2) the choice is made, meanwhile, to allow the annual rate of emissions to remain high enough to continue to increase atmospheric concentration, then in order to avoid mitigation costs for the current generation, much higher mitigation costs will have been imposed upon future generations. This appears to be a compound injustice toward our descendants.

FURTHER READING

Agarwal, A, Narain, S, and Sharma, A, eds (1999) *Green Politics*, Centre for Science and Environment, New Delhi.

Att eld, R (1999) *The Ethics of the Global Environment*, Edinburgh University Press, Edinburgh.

Brown, P G (2000) *Ethics, Economics and International Relations*, Edinburgh University Press, Edinburgh.

de-Shalit, A (1995) *Why Posterity Matters*, Routledge, London.

Franck, T M (1995) *Fairness in International Law and Institutions*, Oxford University Press, Oxford.

Grubb, M (1995) Seeking Fair Weather: Ethics and the International Debate on Climate Change, *Int. Affairs*, **71**, 463 496.

Grubb, M, Chapuis, T, and Duong, M H (1995) The Economics of Changing Course, *Energy Policy*, **23**, 417 431.

Grubb, M, Vrolijk, C, and Brack, D (1999) *The Kyoto Protocol: A Guide and Assessment*, Royal Institute of International Affairs, London.

Jones, C (1999) *Global Justice*, Oxford University Press, Oxford.

Paterson, M (1996) International Justice and Global Warming, in *The Ethical Dimensions of Global Change*, ed B Holden, Macmillan, London, 181 201.

Pogge, T W (1998) A Global Resources Dividend, in *Ethics of Consumption*, eds D Crocker and T Linden, Rowman and Little eld, Lanham, MD, 501 536.

Shue, H (1999) Global Environment and International Inequality, *Int. Affairs*, **75**, 531 545.

Ethics, Environmental

see Environmental Ethics (**Volume 5**)

Ethnoscience

see Indigenous Knowledge, Peoples and Sustainable Practice (**Volume 5**)

Exxon Valdez Oil Spill

The spill of crude oil from the Exxon Valdez is probably the world s most infamous tanker disaster, although it was not the world s largest oil spill. In the minutes after midnight of March 24th, 1989, the Exxon Valdez, its captain intoxicated, ran its single hull onto a shoal in the northeastern portion of Prince William Sound, off Alaska. The tanker was carrying nearly 1.3 million barrels of North Slope crude oil. Fully one- fth of that cargo spilled into the sound: 11.2 million gallons of crude (42 million liters). Initially, there were hopes the crude could be contained due to the relatively calm seas. But three days after the Exxon Valdez ran aground, strong northeasterly winds blew up and dispersed the oil hundreds of kilometers from the accident site. The mixing of oil and water, agitated by wind and wave action, created a substance known as *mousse* which was carried 750 km (470 miles) from the grounding site. The oily mixture washed ashore along a huge expanse of Alaskan coastline, from Prince William Sound to southern Kodiak Archipelago and the Alaska Peninsula. Thousands of sea birds and mammals were killed by the oil. Hundreds of volunteers attempted the rescue and recovery of oiled birds.

Exxon spent tens of millions of dollars in attempted remediation, including sand blasting of beaches. The cost of clean-up had a positive economic effect on the state s gross national product (GNP), generating more economic activity than if the cargo had safely reached its destination. Within Prince William Sound, 790 miles (1271 km) of shoreline were oiled, as well as another 2400 miles (3862 km) of shoreline in the Kenai Peninsula Kodiak Archipelago region. Long-term studies continue about the impact of the spill.

ELIZABETH MAY *Canada*

F

Factor-four Scenario (Double the Output with Half the Resources)

see Ecological Economics (*Opening essay,* **Volume 5**)

Fernow, Bernard

see Environmental Movement – the Rise of Non-government Organizations (NGOs) (**Volume 5**)

Flip-flop of a Complex System

see Phase Shifts or Flip-flops in Complex Systems (**Volume 5**)

Focus Group

see ULYSSES (Urban Lifestyles, Sustainability, and Integrated Environmental Assessment) (**Volume 4**)

Francis of Assisi

(1181- 1226)

Francis of Assisi, Saint, mystic, cleric, was born at Assisi (Italy) to a wealthy merchant family. It was assumed that he would enter the family business, although apparently for a time he was a troubadour and soldier. In 1201 he took part in an attack on Perugia where he was taken hostage and held for a year. It was during this period that he began to turn to religion, although after his release he enlisted in another military venture. It was at this time that he had a dream in which God called him to do His works. He returned to Assisi and began to care for the sick. In 1206, he had a vision in which Christ called on him to repair his church, which Francis interpreted as being the Church of San Damiano, near Assisi.

He devoted himself to repairing the church. Francis gave up his rights and possessions, and then began to preach. He was joined by others and his group soon numbered eleven, it was then that he gave them the short Rule (a set of rules of conduct for his new order). Pope Innocent III granted approval for a brotherhood, called the Friars Minor. Although based at Assisi, they preached across central Italy calling for the return to a life of Christ, one that emphasized simplicity, poverty, and a faith in God's care, rather than possessions. In 1212 Clara Sciffi, inspired by his devotion, joined Francis and formed the Poor Clares sisterhood. In 1219 Francis journeyed with crusaders to the Near East.

The brotherhood grew into an order, and larger numbers were attracted to Francis' message of simplicity in life, poverty, piety, and humility before God. He wrote a more detailed Rule, which in time was further revised by new leaders of the Franciscans. In this time he also became known for his walks in the woods and hills, and his sermons to birds and animals. What is more likely is that he preached about nature rather than to it. In nature Francis saw the work of God, and to abuse or defile God's creatures was an affront to God and an act of debasement. Francis saw nature as an extension of God. Understanding the beauty and spirituality of nature was for Francis an integral part of his devotion and piety. His vision of God's world was based on a trust in God's love and he believed that poverty (taking no thought in tomorrow) would enable him to experience that love most fully. This linked him to the poor and disadvantaged and remains a strong part of his appeal today.

Later, he gave up leadership of the Order and went to the mountains to live a life of secluded prayer. For Francis, his life in the hills and forests brought him closer to the essence and presence of God. It was in the wilderness

that he received the stigmata (wounds of Christ). He did return to visit the Franciscans, and also had a small group of followers. In the fall of 1226 he died at Porziuncula, near Assisi.

Regardless of Francis' original beliefs, he is today considered by many to be the Patron Saint of ecology. For example, his name served as the rallying point for efforts such as the Assisi Declaration from the World Wildlife Fund's anniversary gathering at Assisi (in September 1986).

KEVIN HANNA *Canada*

Friends of the Earth

Friends of the Earth (FoE) was founded in 1969 in the United States by former Sierra Club Executive Director, David Brower (*see* Brower, David, Volume 5). Although the organization and Brower ultimately had a well-publicized and acrimonious parting, it lived up to its founder's vision of an effective family of national groups within an international organization.

By 1988, the organization had established relatively autonomous member organizations in 61 countries, all within an international federation; FoE-International (FoEI). One of FoE's clearest successes has been the spread of the network to countries in the developing world, and a relatively democratic decision-making structure between and among national FoEs. Expansion to the south took off after the 1986 annual general meeting, which was hosted by the Malaysian FoE, Sahabat Alam Malaysia. FoEI was also one of the first international environmental groups to recognize the need to support developing civil society in the former USSR. Although founded in the US, FoE's international headquarters is in Amsterdam. Combined global membership in 1988 was close to one million, and the combined annual budget of FoEI was nearly $200 million (US), with 700 full time staff members.

FoE, like Greenpeace and the Sierra Club, is prepared to tackle a wide range of concerns, from forests to pesticides, climate change and ozone depletion to nuclear energy, and the damming of rivers. FoE's US head office in Washington, DC has developed a global leadership role in non-governmental organization (NGO) campaigns to reform the environmental policies of Export Credit Agencies and international financial institutions. Its US Executive Director, Brent Blackwelder, was arrested in the Washington protests against the World Bank and the International Monetary Fund in the spring of 2000. FoE (UK) has been in the lead in protesting genetically modified foods. FoE's style is campaign driven, not corporate. It aspires to grassroots activism buttressed by effective national and international campaigning.

ELIZABETH MAY *Canada*

Futures Research

Allen Tough
University of Toronto, Toronto, Canada

Futures research is a small but vibrant field of inquiry, theory, and practice. It does not predict or promote one particular future. Instead, it studies a wide range of potential or alternative futures. Sometimes these are spelled out in the form of stories about the future, usually called future scenarios or images of the future. Often one or two of these will be very positive, one or two very negative, and the rest somewhere in between. Some of these scenarios are more likely than others, of course, but often these are the negative ones. This exercise can help us choose our preferred futures and work toward making them actually happen.

Long-term global environmental change plays a key role in almost every future we can imagine. Climate patterns, water, food, species and oil affect almost every scenario. Similarly, various possible socio-economic and political scenarios will have major effects on the global environment. Various futures for science and technology, too, could greatly affect the environment. If nanotechnology succeeds in building things atom by atom, molecule by molecule, it may enable us to transform toxic wastes and improve the environment cheaply and easily – a scenario that sharply differs from the usual long-range forecast.

For understanding global environmental change, futures research is most useful when it focuses on a horizon between 30 and 100 years from now. Shorter term scenarios can more easily pretend that the environment is not an important factor. At the other extreme, one can be fairly sure that global environmental change will not be high on the agenda for public policy 1000 years from now: either it will have been solved much earlier or we simply will not be here.

The field of futures research arose from the need for post-war planning in the late 1940s. In its early decades, it dealt largely with positive images of the future. The largest conference of futures researchers in history was held in Toronto in 1980; about 5000 attended. Even at that stage, very few of the papers mentioned war, environment, or population growth as major problems.

Now the field has matured, although it is still remarkably small in light of humanity's needs for thoughtful long-term planning. The major international organization for futures researchers is the World Futures Studies Federation. The Future Generations Program at the University of Malta holds regular conferences and in 1997 encouraged the United Nations Educational, Scientific and Cultural Organization General Conference to adopt a declaration on the responsibilities of the present generations towards future generations. The Foundation For the Future studies the major factors affecting society over the next 1000 years. Although various university departments offer single courses on futures research, under a variety of titles, only a few graduate programs in the world focus on futures research.

Because it focuses on humanity's long-term future, or on future generations, futures research can provide a crucially important perspective to our understanding of global environmental change. By forcing us to look far ahead, futures research makes us realize the crucial need for achieving a sustainable society (and an appropriate size of population) that does not exceed the long-term carrying capacity of our environment and resources. In this way, we can give future generations an opportunity equal to ours instead of a devastated planet crowded far beyond that of today. Ideally we would like to see human well-being and the quality of life improve around the world over the next few decades, but at least we can try to ensure that they do not markedly decline. If we fail to make the needed changes – if massive permanent deterioration of our physical environment and resources occurs – then future generations could be much worse off than we are.

At times, individuals and societies make certain choices because of immediate benefits. At other times people look ahead and wisely give up some immediate benefits because of greater long-term benefits. Futures research encourages this longer-term or future generations perspective. Within such a perspective, longer-term decisions make excellent sense, both at the individual and the societal levels.

Futures research employs a variety of approaches. Three of the most common are extrapolation, scenario-building, and the Delphi method. Computer simulation is not so common, partly because of its complexity, but it is particularly relevant for forecasting global environmental change.

Extrapolation involves simply continuing a series of trends into the future. The variables could be temperature, population, stock prices, or yield per hectare. In the short run, extrapolation can be very successful because the recent track record of a variable is often the best indicator of the next few months. But in the long run extrapolation becomes risky since many trends eventually reverse.

Scenario construction is a more creative futures methodology than extrapolation. At one extreme, creating a set of scenarios is quite similar to writing science fiction or future fantasy. It can be a deeply personal effort to build richly detailed stories of how things might unfold between the present and some future date. Alternatively, scenarios can be constructed by a group of managers or citizens in order to reflect on various possible futures for their company or region. They can be asked to construct the most likely futures, their preferred futures, the worst and best possible futures, or some combination. A global think-tank called the *Millennium Project* of the American Council for the United Nations University uses questionnaires, interviews, and scenario contributors from around the world to build its scenarios, presented in their annual State of the Future report (Glenn and Gordon, 2000). Many scenarios elicit strong emotions, ranging from Let's do everything we can to avoid that horrible possibility through to I'd like to devote my life to achieving that beautifully positive image of the future (*see* Scenarios, Volume 5).

The *Delphi method* involves repeated surveys of a panel of experts on likely future events and achievements, sometimes even on the most likely date that each will occur. Because the results of each round are fed back to the respondents in their next round, the expert opinions often move toward a consensus position. In order to take into account the ways in which variables and contingencies may influence the outcome, a Delphi study may add cross-impact analysis by constructing a matrix to show their interdependencies (Bell, 1997).

Computer simulation can be used to develop and manipulate models of the atmosphere, the biosphere, the economy, or a mix of all three. This approach usefully forces the futures researcher to explicitly and precisely state many assumptions. The resulting model can be used to answer questions about the likely outcome if certain changes occur in key variables or policies. What will happen to world resources and population, for instance, if we continue with the *status quo*, often called the *business-as-usual scenario*? (*see* Business-as-usual Scenarios, Volume 5) What if we truly shift to a policy of sustainability (Meadows *et al.*, 1992)?

REFERENCES

Bell, W (1997) *Foundations of Futures Studies*, Transaction, New Brunswick, NJ.

Glenn, J C and Gordon, T J (2000) *State of the Future at the Millennium*, American Council for the United Nations University, Washington, DC.

Meadows, D H, Meadows, D L, and Randers, J (1992) *Beyond the Limits: Confronting Global Collapse, Envisioning a Sustainable Future*, Chelsea Green, Post Mills, VT.

G

Gaia Hypothesis

G R Williams
Guelph, Canada

Gaia was the name given in Greek mythology to Mother Earth. Thus, it is etymologically related to all our words pre xed with geo-, such as geography and geology. In 1972 James Lovelock appropriated the name Gaia for his concept that " the total ensemble of living organisms which constitute the biosphere (and) can act as ••• an active adaptive control system able to maintain the Earth in homeostasis" (Lovelock and Margulis, 1974a). This imaginative suggestion has gained considerable popular support but has been the subject of trenchant scienti c criticism.

The concept of Gaia arises as an attempt to answer two questions. First, how is one to account for the anomalous composition of Earth's atmosphere? The atmosphere of Planet Earth is anomalous in two respects. The mixture of gases in our atmosphere is far from chemical equilibrium. For example, over geological time, nitrogen and oxygen would be expected to react to form oxides of nitrogen. The coexistence of methane and oxygen in the atmosphere is an even more convincing example, departing from thermodynamic equilibrium by a factor of 10^{30}. The atmosphere of Earth is also anomalous in comparison with that of the other terrestrial planets, Venus and Mars. On these planets the composition of the atmosphere is close to chemical equilibrium. There is little free molecular oxygen (O_2) and the predominant gaseous species is carbon dioxide (CO_2). Lovelock suggested that the aberrances on Earth are caused by the activities of the global biota and that, more generally, the detection of such anomalies could be used as a criterion for the presence of life on a planet (Lovelock, 1975). The spectroscopic detection of thermodynamic anomalies in the atmospheres of planets orbiting other stars is now becoming a possibility. Such evidence would be a remarkable vindication of Lovelock's suggestion. The control experiment has already been performed: a fly-by of Earth by the spacecraft Galileo found profound departures from chemical equilibrium indicative of the presence of life (Sagan *et al.*, 1993)!

The second question to which the Gaia concept seeks to provide an answer is, how has life been able to persist on Earth for a period of at least 3.5 billion years? For this length of time the environmental conditions of the planetary surface must have remained within limits that life in its various forms could tolerate. As Lovelock and Margulis (1974b) noted:

> Stable temperature, pH and element cycling requirements for life must have been met on this planet for the entire recorded history of the Earth. Organisms found in extreme environments are highly specialized. If the Earth had frozen for even a few tens of thousands of years, or if hot acid springs had been widely distributed for even a single epoch, these occurrences would have been discerned from the fossil record.

This stability suggests a control system similar, for example, to those that maintain constancy in such characteristics as body temperature and the concentration of such blood constituents as hydrogen ions and glucose. This constancy of the internal environment, homeostasis, forms the model for the Gaian idea that the conditions of the external environment have been held within limits favourable to life since its origin 3.5–3.8 billion years ago.

The key point to Lovelock's use of the word Gaia is that it brings together two ideas. First is the idea of an adaptive control system analogous to the thermostatic control of room temperature, involving a sensor (thermostat) and effector (furnace or air conditioner). Such a system is inferred from the persistence of life through the geological record. Second is the idea that this control is exercised through the collective activities of the biota. Such control may be inferred, he suggested, from the anomalous chemical composition of the atmosphere. The two ideas are connected (for instance global temperature is profoundly affected by atmospheric composition) but it is essential to the Gaian notion that both be present. In Lovelock's view Gaia is not just a global control system (inorganic control mechanisms can be postulated) nor is it just the global biota. Rather, Lovelock suggests that the global biota, are responsible for maintaining the steady state (but non-equilibrium) properties of the surface of the Earth (Williams, 1996).

The idea of a living Earth, given a name that has religious overtones, has acquired considerable popular appeal. However, some concern has been expressed that an undue emphasis on the adaptability and resilience of the environment (and even its ability to heal itself) might encourage insouciance with respect to the deleterious effects of human activities. That Lovelock s concept is not consonant with those who would interpret Gaia as allowing unfettered human disruption of the environment is clear. However, within the Gaian scheme of things, the species *Homo sapiens* has no privileged position and a degree of stability consonant with the persistence of many forms of life might permit a range of environmental conditions incompatible with human existence. Environmental activists have focused attention on the central role ascribed to the global biota and on the name Gaia, which conjures up a sense of reverence for life. For instance, there is a *Missa Gaia* commissioned by and performed annually in New York City s Cathedral of St. John the Divine. A search of the World Wide Web for Gaia alone displayed 36 339 pages, and 1453 pages for the association of Gaia and Lovelock.

Within the scienti c community, the analysis of the Gaian hypothesis by geochemists and biologists has focused on a number of issues. The notion that a biota can play a catalytic role in determining planetary processes has become more fully appreciated. It has been recognized for many years that the oxygen of the atmosphere is derived from water through the biological process of photosynthesis. More recently, the seasonal uctuations in atmospheric carbon dioxide, that are such a conspicuous feature of the Mauna Loa record, have been recognized as testimony to the ability of the biota to in uence atmospheric composition. But scientists studying the important inorganic cycles of the planet have been reluctant to assign a signi cant role to the biota (Holland, 1984). More recently, as geochemists have started to analyze planetary issues from a systems analysis point of view (Kump *et al.*, 1998) focusing on biogeochemical cycles, Lovelock (1986) has suggested the term *geophysiology*, to bring together those focusing on such cycles with those searching for biotic control systems but the word lacks the mythological resonance that gives the earlier Gaian idea its impact.

It is this latter notion (of a global biotic control system) that has led to the strongest criticism of the idea of Gaia (Williams, 1996, Chapter 5). In early formulations of the idea, Lovelock did not hesitate to write both of control of the global environment by the biosphere (which might be uncontroversial) and of control for the biosphere . Insofar as the latter has teleological implications, i.e., it implies that the Gaian system exhibits purposiveness, this runs counter to mainstream contemporary biological thinking.

Indeed, if one wished to speak of a living Earth or Earth as an organism, it would be almost mandatory to present evidence of such purposiveness. Thus the real problem lies not with the implication of a purposeful Gaia, but with the absence of any account as to how purpose would be built into the Gaian organism. For those entities that biologists recognize as organisms, apparent purposiveness is accounted for by natural selection. Recently, evolutionary accounts have even provided an explanation for cases of cooperative activities of organisms in communities. In addition, cooperation among biotic components is seen in all multi-cellular organisms, which depend upon cooperation between the genes expressed differentially in the differentiated tissues. More relevantly, a similar cooperation can exist between genes in different organisms. However, such cooperation has been demonstrated only within communities of closely related individuals. A global organism would require intergenomic interactions among organisms of many very different species separated by long distances. To most biologists the possibility of such large-scale symbioses seems improbable to impossible.

The problem may be expressed in another way. Apparent purposiveness arises in living systems because the random mutation of genes throws up occasional variants that can survive and reproduce better than the original strain or can survive and reproduce in ecological niches inaccessible to the original strain. Such an evolutionary scenario for the *Origin of Species* has two prerequisites, time and genetic potential variation. There has been plenty of time, four billion years, over which action could take place. What the Gaian concept appeared to lack was a means of explaining the source of genetic potential variation (Ehrlich, 1991) (the grounds for selective success are not important if a means can be identi ed). Symbiotic cooperation may be as important as interspeci c competition, but there must be a vast reservoir of variation, and it must be heritable variation, on which cumulative selection can operate. (Note that the idea of interplanetary selection has never been proposed other than as a *reductio ad absurdum* (Dawkins, 1982)).

Watson and Lovelock (1983) attempted to meet the objections of biologists by constructing a model of a planetary world that was receiving radiation from a star, the output radiation of which was increasing monotonically. In their formulation, the temperature of the planet was set by the balance between incoming and outgoing radation; the latter comprising both re ected and re-radiated energy. If the planetary albedo was held constant, the surface temperature of the planet would rise as the radiative output of the star increased. The expected increase in temperature could be readily calculated using the Stefan Boltzmann relationship between radiation and temperature. Lovelock and Watson then seeded their imaginary planet with daisies of two types: black and white. At lower stellar output the black daisies ourished and decreased the planetary albedo, thus increasing the surface temperature; as the stellar output increased, the white daisies competed successfully. This increased the planetary albedo and the surface

temperature was modulated below the value expected for a lifeless planet. Thus, the presence of these two variants (black or white) of a single species on Daisyworld produced a remarkable stabilization of a global environmental parameter (*see* **Daisyworld**, Volume 2).

Like the original concept of Gaia, the Daisyworld model has produced both interest and controversy. It has been elaborated to accommodate multi-species interactions and it has been demonstrated that some variations of the model are destabilizing. Nonetheless, the original model can be said to have fulfilled its original purpose. It provides an existence theorem that shows that competition between the populations of black and white daisies can produce global environmental stabilization without any teleological presuppositions being built into the model. The problem with Daisyworld is not that the model did not work but that it was so remote from the real world to make its applicability uncertain. The daisies have no physiology or biochemistry, they are simple geometrical black or white discs. And they are introduced into Daisyworld at the will of the modeller. Once again, the problem of evolution appeared to have been evaded. If stability depends upon the daisies having some particular optimum temperature for growth, how is that value achieved? In the model it is inserted by the modeller; whereas in the real world temperature optima are phenotypic characters arrived at by Darwinian selection.

While the specific example had shortcomings, it did prompt scientists to search for similar feedback in the real world. The regulation of the temperature of the real planet Earth, as opposed to the imaginary planet Daisyworld, does in fact provide some possible areas in which to look for supportive evidence for the Gaian idea. One focus of effort was to try to understand the relationship of global biological activity to the well documented ice ages of the past million years (the Pleistocene). Analyses of air bubbles entrapped in ice cores from the Arctic and Antarctic provide important information about the chemistry of the atmosphere during the more recent glacial–interglacial cycles (Lorius *et al.*, 1990). These records indicate that the concentration of the most important greenhouse gas, carbon dioxide, varied in approximate synchrony with the temperature record. From the Gaian perspective, in its most direct sense, it might be suggested that a biotic thermostat should be evident in which the onset of an interglacial–glacial transition (global cooling) would be accompanied by an increased level of carbon dioxide that would offset or ameliorate the fall in temperature. In fact the opposite is evident from the observations; global cooling occurs concomitantly with a decrease of carbon dioxide and, correspondingly, there is an association between global warming and increased carbon dioxide. The precise causes of the changes in carbon dioxide are not understood but such effects appear to suggest a positive feedback rather than the negative feedback that one might expect as a characteristic of Gaia. That is, if the changes in carbon dioxide are driven by biological activity, then the biota are acting to destabilise global temperature such that a relatively minor response to cyclic changes in insulation is thus amplified rather than attenuated. Proponents of the Gaia concept might point to the fact that life persisted through the ice ages (e.g., through the mechanism eventually acting to offset the cooling) or even that the range of temperature variation allowed might have contributed in some way to later sustaining more abundant life (e.g., by usable soil).

Another suggestion of a Gaian connection to the regulation of the surface temperature of Earth is provided by an intriguing relationship between the production of dimethylsulphide (DMS), $(CH_3)_2S$, by marine algae and cloud formation over the open ocean (Charlson *et al.*, 1987) (*see* **Dimethylsul de (DMS)**, Volume 1). Cloud formation is limited by the availability of cloud condensation nuclei. In a marine environment, remote from land, one of the principal sources of such particles arises from the atmospheric oxidation of DMS. DMS is an end product of algal metabolism, predominantly formed when the algae are ingested by zooplankton. Atmospheric oxidation of the DMS occurs principally by reaction with the OH radical (*see* **OH Radical: is the Cleansing Capacity of the Atmosphere Changing?**, Volume 2). The oxidation products, $CH_3SO_3^{\bullet}$ and $SO_4^{2\bullet}$ can act as cloud condensation nuclei. The particles formed in this manner appear to be of more significance in cloud formation than those formed from sea salt. Cloud formation, in turn, can increase global albedo and affect the global radiation budget. Thus, the metabolic activity of marine algae can modulate global surface temperature. Lovelock regards this as an important instantiation (representation by on example) of the kind of processes that one would expect to encounter when looking at Earth from a Gaian perspective. However, to qualify as confirmatory evidence, such processes should be involved in a negative feedback loop. Further, to meet the Darwinian argument, the beneficial effects of such a loop have to outweigh the metabolic costs to the organisms responsible for DMS production. Since Lovelock drew attention to this important climate/biota relationship in 1987, much effort and ingenuity have gone into suggesting how such a feedback might occur but no model has been generally accepted. At the observational level, measurements of $CH_3SO_3^{\bullet}$ in ice cores associate the low temperatures of the last ice age with high $CH_3SO_3^{\bullet}$ and therefore with increased cloud formation. As in the case of carbon dioxide, the activity of the biota appears to have been destabilizing. The decrease in temperature would be more pronounced as the result of both a drawdown of atmospheric carbon dioxide, leading to a reduced greenhouse effect, and an increase in DMS production, leading to an increase in global albedo.

The concept and notion of Gaia is intriguing, it has provoked interesting areas of scientific inquiry and it continues to provoke much debate. As analysis of the Earth as a system continues, the Gaian contribution has been an enhanced recognition of the role of the biosphere. Two effects on biology are likely to persist: a new interest in very large-scale symbioses and a new consciousness of the importance of microbial processes in global ecology. Thus, some acceptance has been given to that part of the Gaian view which sees the functioning of the planetary biosphere as being dominated, today as it has been for 3.5 billion years, by microorganisms (*see* Microbial Diversity, Volume 2).

Is Gaia merely a description (fanciful, useful, or both) of the state of affairs on one particular planet, Earth? Or is *Gaianness*, to use computer jargon, platform-independent? If Lovelock's suggestion that habitation of a planet will always result in an anomalous planetary surface were ever to be substantiated, then Gaia theory would take its place among the major imaginative leaps of scientific thought. One might particularly liken it to the 19th century formulation of the Laws of Thermodynamics, rules that apply to assemblages of matter, independent of structure or location. While further investigation of the Earth will be fruitful in expanding understanding, the search for life elsewhere in the Universe, which was Lovelock's original impetus for development of the concept, will provide its most stringent test.

REFERENCES

Charlson, R J, Lovelock, J E, Andreae, M O, and Warren, S G (1987) Oceanic Phytoplankton, Atmospheric Sulfur, Cloud Albedo, and Climate, *Nature*, 326, 655–661.

Dawkins, R (1982) *The Extended Phenotype*, Oxford University Press, Oxford.

Ehrlich, P (1991) Coevolution and its Applicability to the Gaia Hypothesis, in *Scientists on Gaia*, eds S H Schneider and P J Boston, MIT Press, Cambridge, MA.

Holland, H D (1984) *The Chemical Evolution of the Atmosphere and Oceans*, Princeton University Press, Princeton, NJ.

Kump, L R, Kasting, J F, and Crane, R G (1998) *The Earth System: An Introduction to Earth Systems Science*, Prentice Hall, New York.

Lorius, C, Jouzel, J, Raynaud, D, Hansen, J, and Letreut, H (1990) The Ice-core Record: Climate Sensitivity and Future Greenhouse Warming, *Nature*, 347, 139–145.

Lovelock, J E (1975) Thermodynamics and the Recognition of Alien Biospheres, *Proc. R. Soc. London B*, 189, 167–181.

Lovelock, J E (1986) Geophysiology: A New Look at Earth Science, *Bull. Am. Met. Soc.*, 67, 392–397.

Lovelock, J E and Margulis, L (1974a) Atmospheric Homeostasis By and For the Biosphere: The Gaia Hypothesis, *Tellus*, 26, 1–10.

Lovelock, J E and Margulis, L (1974b) Homeostatic Endencies of the Earth's Atmosphere, *Origins of Life*, 5, 93–103.

Sagan, C, Thompson, W R, Carlson, R, Gurnett, D, and Hord, C (1993) A Search for Life on Earth from the Galileo Spacecraft, *Nature*, 365, 589–593.

Watson, A J and Lovelock, J E (1983) Biological Homeostasis of the Global Environment: The Parable of Daisyworld, *Tellus*, 35(B), 284–289.

Williams, G R (1996) *The Molecular Biology of Gaia*, Columbia University Press, New York.

Gandhi, Mohandas Karamchand

(1869- 1948)

Mohandas Karamchand Gandhi, also known as *The Mahatma* (The Great Souled One), is most famous as the non-violent force behind the independence of India from British rule in 1947. His spiritual stance on the struggle for truth (*satyagraha*) also led him into developing his influential teachings and practices on economics and what would now be called alternative living. The idea of *small is beautiful*, based on the book of the same name by Ernest Schumacher, derives ultimately from Gandhi's example.

Gandhi was born in 1869 in Western India, and as a young man traveled to London to study law. This gave him insights into British culture that were invaluable in later life; for example, well placed letters to *The Times*, or appeals to the British sense of fair play. While in London in the late 1880s, he became caught up in various alternative groups, made up of spiritualists and vegetarians, who were interested in Eastern thought. Ironically, it was these Westerners who awoke Gandhi to the strengths of his own Indian traditions. Gandhi also absorbed the writings of Leo Tolstoy on non-violence, and was particularly influenced by John Ruskin's book, *Unto This Last*, which attacked the classical economists' views of human nature as selfish and competitive. This became hybridized with his own Indian traditions, a mixture of Hinduism and the non-violent religion of Jainism.

As a result, beginning with his career in South Africa as a barrister, Gandhi evolved a sophisticated strategy for pressuring governments and companies to change their policies. This combined exhaustive legal work on the issues at stake with a general mobilization of the affected parties in non-violent protest. Gandhi believed, quite early on, that the main reason for the maltreatment of the Indian communities (and others later on) was that these communities had themselves internalized the vision of their current masters. By non-violent protest, these communities were

empowered, and were morally strengthened by their very refusal to resort to violence. This gave them superiority over their supposed superiors, who inevitably resorted to violence, violence that would in the long run undermine the official claims to legitimacy.

Accompanying this non-violent approach to political conflict, Gandhi further explored the personal and social aspects of the pure in heart, through experiments in simple living in *ashrams*, that is, small communities of believers. These experiments in weaving one's own cloth, cleaning out latrines (seriously threatening in a caste-ridden society where only untouchables were supposed to do this), milking goats, etc., were eventually made part and parcel of Gandhi's politics. He argued that before one could rule the nation (*the raj*) one had to be able to rule oneself (*swaraj*) and that this entailed, among other things, making one's own products, like clothing (*swadeshi*). This was a direct attack on the vast British import/export businesses centered in India.

In the 1930s, as the struggle for Indian independence heated up, Gandhi became world famous, both in prison and out on the march. His contests against the British government mobilized millions of followers in extraordinary scenes of non-violent resistance. Simultaneously, Gandhi's constant effort to understand and develop the implications of the search for spiritual truth, in the personal as well as political spheres made him a beacon for some, and a headache for others. An escalating series of fasts and non-violent protests brought the British government to the reluctant admission that they would have to grant some form of independence to India: a process that was delayed by the onset of World War II. Throughout this period, the political party associated with Gandhi (the Congress Party, headed by a disciple of Gandhi's, Nehru) continued to advocate independence. This was further complicated by the gradual dissolution of a Hindu-Muslim partnership, which would eventually lead to the bitter partition of India into India and Pakistan.

Throughout the run up to Independence (in August 1947) and the rioting and fighting that accompanied it, Gandhi struggled against the tide of violence. He was himself shot by a Hindu militant on January 30, 1948. Near universal mourning ensued.

Gandhi's economic and environmental teachings were immediately repudiated by the new Indian government, which embarked on widespread industrialization and quasi-social central planning. Some of Gandhi's disciples continued to press his teachings and practices subsequently, to marginal effect.

Nevertheless, it can be seen that Gandhian non-violence as a concept and a practice has been one of the most influential weapons in the struggle for justice over the last 50 years. The Reverend Martin Luther King Jr was only one of those for whom Gandhi's model was central. The peaceful revolutions in Eastern Europe, the Philippines, and most recently in Yugoslavia owe much of their inspiration and tactics to Gandhi's spiritual pressure. The core idea, that both sides in any struggle are seeking a truth, which the seemingly weaker side can impress upon the seemingly stronger through its commitment to the justness of its cause in the face of violence, has become widely practiced. Martin Luther King's idea that segregation is a degradation of whites as much as, if not more than, blacks, also derives its moral suasion from a Gandhian vision.

In recent years, Gandhism has maintained its immediate political influence among environmentalists, notably in the mobilization against the Narmada dam, and the Chipko women's movement in Northern India. His writings on self-rule at the village level remain central to participatory models of development. Finally, over the longer term, it is his attempt to lead a spiritually holistic life, where non-violence includes non-violence against the things of the Earth that makes Gandhi a continuing figure to be reckoned with in the environmental movement worldwide.

REFERENCES

Gandhi, M K (1927) *An Autobiography: or, The Story of My Experiments with Truth*, Penguin Books, Harmondsworth.

Iyer, R (1986) *The Moral and Political Writings of Mahatma Gandhi*, Clarendon Press, Oxford.

PETER TIMMERMAN *Canada*

Garden, Landscape

see Landscape, Urban Landscape, and the Human Shaping of the Environment (**Volume 5**)

Global Environmental Change and Environmental History

see Global Environmental Change and Environmental History (*Opening essay*, **Volume 5**)

Global Forests: Citizen Monitoring

see Global Forest Watch (**Volume 4**)

Globalization

see Development and Global Environmental Change (*Opening essay*, **Volume 5**)

Globalization in Historical Perspective

see Globalization in Historical Perspective (*Opening essay*, **Volume 5**)

Governance and International Management

Nazli Choucri
Massachusetts Institute of Technology, Cambridge, MA, USA

In the absence of world government, the matter of global governance assumes particularly important proportions. This is especially true in the environmental domain as human activities generate signi cant problems of global scale and scope. At issue are formalizing modes of moderating con icts and enhancing prospects for cooperation among nations in a context where the proverbial anarchy prevails, modulated by rapid change and increased complexity. Signi cant among these complexities is the variety of actors and agents whose interactions are subject to international management; nation-states, international corporations, non-governmental groups, local actors, and regional agencies are among the most salient of these actors. And, as international rules for collaboration and management consolidate, new actors appear to be emerging with new demands and new propensities for con ict or cooperation as the case might be.

GLOBAL REALITIES

The reality of global politics, and its legal underpinnings or lack thereof, is that only states are enfranchised to act on behalf of individuals. Only states represent populations; only governments represent constituencies with legal standing. Individuals *per se* have, as of yet, no formal mechanism of representation other than the state. While there is evidence of some movement toward recognition of alternative voices in international forums, only the state is recognized as voter.

Therein lies the heart of the dilemma of international management: The state represents a range of demographic constituencies that organize around territorial conglomerations but international problems transcend territoriality and even their mere definition (let alone solutions) cannot be readily mapped over the global arrays of national sovereignty. Moreover, the state is notorious for its porousness (just about everything leaks across national borders) thereby compounding the dual tasks of discharging responsibility and the effective execution of authority.

Challenges to Governance

The essence of the management challenge worldwide is how to reduce the critical disconnects between the demands for coordination and governance, on the one hand, and the supply of legitimate or legitimized mechanisms, on the other – given current realities and evolving complexities.

Modes of Collaboration

Few international realities are as readily discernable and ephemeral, at the same time, than the diversity of collaboration potentials. These range from issue-specific bilateral agreements (formal or informal) to broad ranging, comprehensive global accords (again, formal or informal). In general terms, the logic of international management rests on three key questions, generally known as the *why*, *when* and *how* of global accord:

1. Why collaborate?
2. When to collaborate?
3. How to collaborate?

Of the many conditions that necessitate global accords, two are particularly apt in reflecting the cluster of reasons driving the quest for accord. Countries collaborate (a) in the pursuit of common interests, or (b) in the management of common aversions. In the first instance, states seek collaboration in order to jointly pursue some objective that they might not be able attain individually. In the second, the quest for collaboration is driven by the awareness that they face common adverse conditions that require coordinated action for effective management.

This general logic presumes that countries can identify their specific preferences and objectives, as well as vulnerabilities and sensitivities, and that countries are able to identify the conditions under which unilateral action is not appropriate or bilateral operations will not be effective.

By definition, collaboration involves self-imposed internal or external constraints on national sovereignty. Internal constraints mean refraining from taking actions that have

national consequences. External constraints mean refraining from generating adverse effects outside territorial boundaries. In the environmental domain, broadly defined, as well as in the domain of sustainable development, global accords involve both sets of constraints.

IMPERATIVES OF INTERNATIONAL MANAGEMENT

Underlying the development of modes of international management is the evolution of shared understandings, common formulation of problems at hand, and a joint search for modes of coordinated action. In those terms, the accords themselves are the end product of one type of process; namely that of reaching an agreement. The follow-up requirements involve an entirely new and often complex set of processes at both national and international levels.

The duality, framed in terms of joint pursuit versus common aversion, begs the question of contents, i.e., what is it that is being pursued or needs to be avoided? Who is it that is engaged in avoidance or in pursuit of aversions? What would be the impacts of success or failure (assuming clarity on criteria for each)? How would any of this make a difference? When might this matter most?

Collaboration Imperative

Collaboration becomes an imperative when the problem is recognized, when it is pervasive, and when it eludes unilateral solution and, above all, when no action inevitably means a worsening outcome. These properties are well illustrated in the environment domain characterized by uncertainty, irreversibility, and complexity.

Collaboration becomes a global imperative when:

- *Nature is a player*. This means that environmental effects of human action may take on unanticipated forms, whose uncertainties are suf ciently great as to insert a random element of strong proportions that cannot be contained by human action.
- *Damage is due to legitimate action*. Far from re ecting pathology and deviance, environmental damage is often due to the most normal, routine, and legitimate behavior, whose very nature may be condoned if not lauded worldwide.
- *Force cannot work*. In such contingencies, the deployment of troops, the most conventional instrument of force, is a singularly ineffective, if not a remarkably useless course of action, in that the response is irrelevant to the nature of the challenge.
- *Compliance is imperative*. The pervasiveness of environmental dislocation means that no one can be immune from attack so to speak and that everyone s security is contingent on compliance by everyone else.
- *Doing damage by doing nothing*. Simply by choosing not to take a stand, nations can accentuate prevailing environmental problems; thus, the costs of not participating in evolving environmental accords will be equivalent to overt opposition.

Once collaboration takes place in an institutionalized context, one might ask: why does institutionalization happen?

Institutional Necessity

To simplify, the conventional understanding is that institutionalization at the global level takes place in order to: (a) consolidate and pursue new norms; and (b) coerce states that resist the new norms, and pressure states breaking norms into conformity with the collective understanding; (c) reduce uncertainty in process, outcomes, and information; and in some cases (d) generate and maintain shared modes of communication, understandings, and explanations. Finally, institutionalization is believed to (e) facilitate mediation among con icting actors, and (f) enhance overall prospects for problem solving. All of the above is contingent on some minimum degree of shared understanding of the challenges at issue, in combination with the dual principles of participation and representation in a world based on the principle of national sovereignty.

Key Trends

In this connection, three trends have contributed to the increasing importance of international environmental management. The rst pertains to the increasing role of scientists in shaping the policy agenda. By drawing attention to environmental dislocations due to human activity, by voicing views expressed by a critical mass of scientists playing an important catalytic role by pressing for political recognition of the problems at hand, they exert pressure at national and international levels for policy response. The second trend relates to matters of trial and error, precedent, and renewed efforts in the domain of environmental policy formulation and coordination. The international community has gradually developed a set of principles that reinforce the fabric of international environmental agreements. These include the *precautionary principle* and the *polluter pays principle*, among others. The third trend is the increasing attention at the national and international levels to matters of capacity building for environmental management as well as increased investments in national environmental reporting.

MODES OF INTERNATIONAL MANAGEMENT

In the domain of environmental management and sustainable development, the spectrum of management

modes ranges from formal agreements with legally binding commitments (i.e., the conventions) to informal agreements and soft law (like Agenda 21), on the one hand, and a set of varied instrumentalities and modalities, on the other. These patterns (framed in Table 1 below) are incorporated in the knowledge-based global networking system entitled, the *Global System for Sustainable Development* (GSSD). Used for examining the implications of global accords at local, national, and international levels, GSSD is currently operating in multiple languages and in different parts of the world.

While items (1) and (2) in Table 1 are those most commonly mentioned when referring to global accords, it is often useful to consider that these modalities are often supported by a combination of modes (3)–(14) as relevant. Conceived as legally binding international commitments, the Conventions represent the strongest form of accord and require the most intrusive forms of institutional supports. This latter consideration is due to the fact that formal commitment entails a corollary commitment to put in place enabling mechanisms: nationally or internationally, as the case may be. This corollary is in the nature of an insurance policy.

To reduce potentials for non-compliance due to capacity constraints (i.e., countries may be unable as opposed to being unwilling to comply) a range of capacity building measures increasingly accompanies formal commitments. This is especially the case in the domain of environment and sustainable development. Many of these measures are in the nature of institutional innovations, which can often impinge on traditional conceptions of sovereignty.

INTERNATIONAL RESPONSES AND INSTITUTIONAL INNOVATIONS

When environmental issues are rooted at the local level and travel to the global commons, when individual actions generate global dysfunctionalities, and when states are unable to manage their own environmental agenda, then national governments and their instruments are placed in a precarious position. On the one hand is the situation discussed below (within state borders) that often eludes governmental management. On the other hand are cases where competing sovereignties (each subject to similar elusiveness) seek common principles and modalities of convergence.

Evolving Modalities

Given that states remain the only voters in global forums, nonetheless, there has been a discernable move toward expanded forms of international management involving both participation and representation of non-state actors.

Table 1 Modes of governance and international management as represented in the Global System for Sustainable Development (GSSD)

(1) *Agenda 21*
(2) *Formal conventions, spanning a range of general and comprehensive issues, as:*
- Atmospheric
- Hazardous wastes
- Marine environment
- Terrestrial resources
- Freshwater resources
- Nuclear, safety, and related technology
- Others

(3) *New development and investment mechanisms*
- Cleaner development mechanism (CDM)
- Joint implementation
- Activities implemented jointly
- Innovative investment system

(4) *Technology agreements, such as:*
- Technology transfer
- Training and implementation

(5) *Monitoring performance*
- Limits on hazardous activities
- Benchmarking systems
- Compliance records

(6) *Codes of conduct and voluntary agreements*
(7) *Financial and investment codes*
- Environmental conduct
- Voluntary agreements
- Human rights issues
- Other

(8) *Conflict management and peace strategies*
- Dispute resolution
- Peace keeping
- Conflict prevention
- Other

(9) *Population management*
- Population policies
- International migration strategies
- Resettlement initiatives
- Other

(10) *Economic adjustments and agreements*
(11) *Trade regimes and agreements*
(12) *Environment agreements*
(13) *Corporate responses*
- Organizations
- Agreements
- Private-public partnerships
- Social responsibility

(14) *Strategies of international institutions*

Source: http://gssd.mit.edu.

This move is driven by the realization that environment and sustainability are characterized by rather novel properties distinguishing this realm of evolving international law from other domains of international interaction. For example, the United Nations Framework Convention on Climate Change (UNFCCC), the Kyoto Protocol and the follow-up

measures, all targeted specifically to environmental factors, serve to illustrate the institutional innovations as well as the attendant disconnects (*see* United Nations Framework Convention on Climate Change and Kyoto Protocol, Volume 4).

Among the distinctive features is the framework strategy, rather than precise commitments; process orientation coupled with target specifications; provision for internal as well as international institutional supports; and commitment to long-term adjustment processes based on expectations of innovations in policy instruments.

Less formal than the UNFCCC process, but no less pervasive, is the Agenda 21 process; including the formulation of the contents of Agenda 21, and institutionalization of the follow-up mechanisms via the United Nations Commission on Sustainable Development. Central to these novelties is the tendency for various interests and interest groups to organize networks of communication that eventually become networks of influence. As a result, the increased use of observer status in international forums, combined with the provision of space and the allocation of time for various non-governmental groups to convene during the formal negotiation processes has given non-state actors new venues for voicing their concerns, if not their demands. While none of these innovations alters the fundamental constitution of global politics and none of this transforms the basic sovereignty of the state, it does render the state more transparent and its formal actions subject to greater scrutiny.

These processes are designed to accommodate non-state actors and interests only to the extent that these can be incorporated within the formal structures of deliberation among sovereign entities. But the consolidation of these new institutional processes invariably serves as conduits through which the demands of local constituencies can be reflected in global forums. Invariably, also, they point to the fundamental disconnect, for example, that the mayors of mega cities are not (and cannot be) at the negotiation table nor is there any legal provision for their presence. However, they now have access to transmission mechanisms that may ensure that their views may make their way through the global policy formulation process.

Transcending or Incorporating the State?

What happens next if the views do make their way through the formulation process? In essence the same challenges of disconnect remain. If, and only if, the national government is willing (or agrees) to engage in a dialogue with the mega city institutions (or mayors), can authority devolve from the state to sub-national units and to extra-territorial institutions? Both instances reshape the responsibilities of the state and the scope of its leverages.

In the absence of any serious potential for redesigning the global state system, the most plausible approach might involve rendering more robust some of the existing instrumentalities. Among these are: (a) international pressures on national government to formally account for urban representation in global forums (as part of the national delegation or in nongovernmental contexts); (b) formal arrangements among sub-national governments (such as mayors or functional equivalents) to represent their own interests directly in global forums (as nongovernmental agencies); (c) organized interest articulation of sub-state (or trans-state) representatives to obtain new incentives from national institutions or international institutions, (d) potentials for direct negotiation on projects that arise in the context of CDM (cleaner development mechanism) discussions, and other.

As an example, much of the deliberations and reporting of outcomes for UNFCCC – 5th Conference of Parties (COP/5), has stressed two major venues of deliberation – the state centered forums and the non-governmental organization (NGO) based deliberations. However, even when diversity of interests is evoked, it is almost always with reference to diversity of interests across states. The disconnects at hand pertain to the global implications of sub-national interests and of their potential disjunction with national interests. The data requirements (within the context of national reporting as well as the Intergovernmental Panel on Climate Change (IPCC)) are nationally defined, sectorally formatted, and activity framed (as observations permit).

In addition, we can discern the foundations for four new venues, any one of which might have some merit, but altogether, they may create significant next steps in the question for viable modes of international management. These can be characterized as: (a) consolidation of cross-national political networks supporting collaboration; (b) combining resources for generating and marshalling needed knowledge buttressing the demands for collaboration; (c) maintaining a sustained presence in international forums to influence the evolution of accommodation to non-state actors; and (d) exerting organized and mounting pressures on national representatives toward greater international management.

Many of these developments are converging around the notions of civil international society and of civic global responsibility, where both qualifier c terms reflect a new recognition of non-state agencies, and the potential legitimization of the international civil (in contrast to state centered) society.

Finally, it is important that such venues may contribute significantly to the consolidation of the two critical conditions for effective management, at any level, local or global. These are: (a) viable and relevant modes of representation; and (b) effective and standardized mechanisms for participation. The international community is in the

midst of a grand experiment, namely developing modes of international governance and global management in a still largely sovereign centered world, without imposing formal government and institutions of governance that mirror those of the state system.

See also: International Environmental Law, Volume 5.

FURTHER READING

Brown-Weiss, E and Jacobson, H K, eds (1998) *Engaging Countries: Strengthening Compliance with International Environmental Accords*, MIT Press, Cambridge, MA, http://gssd.mit.edu.

Choucri, N, ed (1993) *Global Accord: Environmental Challenges and International Response*, MIT Press, Cambridge, MA.

The Fridtjof Nansen Institute (1999) *Yearbook of International Co-operation on Environmental and Development*, Earthscan Publications, London.

Green Parties

see Political Movements/Ideologies and the Environment (**Volume 5**)

H

Hazards in Global Environmental Change

Kenneth Hewitt
Wilfrid Laurier University, Waterloo, Canada

*The prospect of disasters related to, or magni ed by, global environmental change is increasingly debated. The problem arises in at least four ways: (i) Dangerous trends: environmental damage or transformations themselves threaten extreme harm; for example, ozone-layer depletion, bioconcentration of toxic contaminants or the emergence of drug-resistant epidemic diseases. (ii) Magni ed risks: there may be increases in the incidence and scale of disasters associated with age-old threats such as droughts, re or pests. Examples include the role of deforestation in oods and wild res, and of climate change in weather hazards or an expanded range of insect-borne epidemic disease (*see ***Natural Hazards***, *Volume 3). (iii) Novel threats: new or magni ed disasters may arise from the development and deployment of technological innovations, or from deliberate environmental modi cations and megaprojects. Some identify such risks with, for example, the introduction or escape of genetically engineered organisms into the biosphere, very large water and power schemes, or ocean oor mining.*

(iv) Increasing social vulnerability and adaptive capacities. This class of problem is related to the fact: that the occurrence and form of a disaster always depends both upon dangerous conditions and the state of human communities exposed to them: increasing social vulnerability and adaptive capacities. The likelihood and impacts of disasters are transformed by demographic, organizational, work and life style changes - even in relation to environmental hazards that do not themselves change in severity or incidence. Endangerment arises with urbanization, transformations in industry, agriculture, transportation, and leisure activities worldwide. Only where public safety measures keep pace, will dangers be kept in check.

Modern innovations and organized responses also counteract such dangers. Research, forecasting, standards, protections, insurance and other safety nets, have dramatically improved public security or the ability to survive damaging events - in some places. Societal vulnerability and adaptability apply, of course, in all four of these disaster contexts. And they involve another issue of global change: changing approaches to disasters. In addition to new disasters and changing risks, there are emerging constituencies, technologies, agencies and agendas relating to disasters. They demand or promote re nements in understanding, or a recasting of the problem. A strong impetus comes from changing expectations of, and dif culties for, national or international emergency preparedness, for humanitarian assistance and activism by, or on behalf of, disaster victims (IFRCRCS, 1998).

Most disaster-related investigations still focus upon environmental and technological agents that may cause harm, whether floods and earthquakes, oil spills, or viral diseases. There is a continual flow of new findings and refinements regarding these hazards. Few will question the importance of such knowledge for risk assessments in a changing world. Nevertheless, the importance of social or cultural understanding, and the dangers of neglecting it, are increasingly recognized (Hewitt, 1983; Blaikie *et al.*, 1994; Quarantelli, 1998; Enarson and Morrow, 1998; Oliver-Smith and Hoffman, 1999).

To grasp the meaning and impetus of these concerns, we must consider their relation to entrenched and contested views of the disaster problem. Most of us may feel we know a disaster when we see one. Yet, like any other general problem – development, poverty, war, ruined habitats – it is much harder to find a widely agreed characterization of disaster, adequate for scientific and humanitarian purposes. Approaches tend to split, firstly, between those preoccupied with what is unique about disasters, and those that link them to broader conditions of environment and society.

DISASTER IN AND OF ITSELF

Disaster or calamity identi es a primary form of experience; a condition of great fear, life-threatening harm and loss. There is also a sense of something untoward, violent, tragic. Everyone dies, but mass deaths by re or ood, in a plane crash, an avalanche, a famine or terrorist attack, seem specially terrible misfortunes. There is not just untimely death and thwarted life. The capacities a person might otherwise have to escape or survive are overwhelmed as in a nightmare. There is the agony of abandonment by, or being unable to protect, loved ones. Also conveyed is the sense of singular occasions or historic(al) turning points (Elliot, 1972; Scarry, 1985).

At their worst, there is little to distinguish the disasters of war, famine, natural and technological catastrophe or pandemic disease. People have always lumped them, or their imageries, together: ruined homes, uprooted families, hungry, exhausted and desperate folk eeing the devastation or huddled in makeshift shelters, the injured and the heaped up dead. Most disasters share this common and dreadful content (Sorokin, 1942). Hence, it is argued they are, above all, humanitarian problems demanding all reasonable efforts to prevent or mitigate them, especially where human-induced changes threaten to make them worse.

Some technical or of cial de nitions also accept and characterize disasters as distinctive and special problems, though they distance themselves from victims feelings and startling imagery. A recent, clear example de nes disasters as:

> ••• non-routine events in societies or their larger subsystems (e.g., regions, communities) that involve social disruption and physical harm. Among the key de ning properties of such events are (1) length of forewarning, (2) magnitude of impact, (3) scope of impact, and (4) duration of impact.
>
> (Kreps, 1995)

Establishing dimensions that seem to bracket disasters as extreme events, and separate them from other times and places, has been a concern of contributing disciplines. There is a preference for impersonal dimensions or mechanisms, in contrast to the experiential or vernacular views, described earlier:

> ••• a state/condition destabilizing the social system that manifests itself in a malfunctioning or disruption of connections and communications between its elements or social units; partial or total destruction/demolition; physical and psychological overloads suffered by some of these elements •••
>
> (Por riev, 1998)

Modern ideas de ne disasters, especially, in relation to, and as the opposite of, predictable, normal, and stable conditions. However, there are other professionals who reject or strongly qualify the idea of disaster as a unique or separate problem. Rather, they see disasters as endemic to, or symptomatic of, the social systems in which they occur (Morren, 1983).

DISASTER MYTHS AND THE PRINCIPLE OF CONTINUITY

Extraordinary terror, trauma, destruction and death surely occur in disasters. Few can argue against paying special attention to them and the worst hit areas. However, comparative and general studies reveal a very different balance of conditions. In most events, most property and persons within a designated disaster areas, are only slightly harmed, or unscathed physically. Even where they are in small numbers, survivors who are not completely disabled, usually carry out most or all of the primary life-saving and harm-reducing activity. In the crucial hours or days before outside help arrives, if it arrives, they rescue those who would, otherwise, be lost. Much, or most emergency relief, food and shelter, is provided by households within or near the devastated areas.

However, popular or media images not only focus on the more or less heroic rescue efforts of outside agencies, the armed forces, international relief and modern technologies. They also portray the disaster in terms of wall to wall devastation, chaos and panic; or concentrate on the more pathetic and prostrate victims, anti-social behavior such as looting, and lucky or miraculous survivals (Ploughman, 1995). Disaster research has shown that these are usually exceptional cases, or re ect mistaken assumptions, stereotypes and selective reportage. Instead, most survivors act, as far as is humanly possible, in terms of the concerns and responsibilities they had before, whether in families and the community, their institutional position or professional skills. This suggests the need for what Quarantelli and Dynes (1977) call a principle of continuity with, rather than a radical separation of disaster from, preexisting societal conditions and normal human responses.

If so, how and how far preexisting conditions and resources in uence peoples responses to disaster are important questions. These, no less than the severity of ood or re, decide the capacities of those at risk to survive and cope. In turn, this suggests that the key questions of global change are how it alters people s vulnerability; and how it erodes or bolsters their adaptive capacities.

INTERPRETATIONS OF CALAMITY

In the broadest terms, explanations of disaster involve four sets of factors:

1. Dangerous agents or hazards, such as earthquake or re, viruses or toxic chemicals, equipment failure or collisions, that initiate the destructive event, or cause most of the damage.

2. Vulnerability of exposed persons, property or activities to dangerous conditions, (discussed in detail below).
3. Risk-reducing and disaster-mitigating measures, which may be directed at particular hazards, or vulnerabilities, and include organized emergency preparedness.
4. Contexts and intervening conditions, factors that indirectly, but sometimes decisively, influence the above three. For example, changes in vegetation cover by forest industries or clearance for agriculture affect flood-generating runoff. Population density, health and well-being which will change due to economic development or urbanization, alter risks of epidemic disease. Technological innovations increase risks though directed at, say, productivity or consumer demand. Such conditions intervene between dangerous forces and vulnerable people to modify their interactions, whether for good or ill, but arise or vary independently. Global change is as likely to affect risk through a range of societal and habitat transformations, as in directly altering flood hazards, safety equipment or high risk technologies.

Actual disasters involve all four sets of factors, but most disaster studies and related activities focus on just one of them. An agent-specific or hazards perspective remains the prevailing perspective. Moreover, this encourages a view of disasters as originating outside, or beyond social control:

> ••• Disaster means the impinging upon a structured community of an external force capable of destroying human life or its resources for survival, on a scale dire enough to excite public alarm, to disrupt normal patterns of behavior •••
>
> (Lemons, 1957)

With respect to technological and social hazards, this idea of disaster converges with that of the accident (Green, 1997) in which danger is not merely identified with unforeseen and unwanted occurrences, or human error, but a domain of uncertainty:

> ••• were there perfectly accurate predictions of what would occur and when it would occur in the intricate web of atmospheric, hydrological and biological systems, there would be no hazard •••
>
> (White, 1974)

This view of disasters as uncertain, unscheduled or accidental events, serves to detach them from preexisting conditions, and the control of on-going social life, and place them in a realm of extreme and rare natural or technological processes. The only possible social responses are then seen as specialized technical efforts to model and counteract such extreme forces and events (Hewitt, 1983).

There is a radical contrast, once again, with interpretations in terms of conditions of everyday life:

> ••• Disaster is the actualization of social vulnerability ••• we interpret the notion of social vulnerability as an independent factor (predictor) of risk •••
>
> (Carlo Pelanda, 1981)

Here, danger is seen to arise from styles of settlement and land use, occupations, technological innovations, wealth and political influence. To accept this means that disaster reduction and emergency preparedness would have to focus much more on the risks embedded in existing social systems, and on global changes that alter peoples' exposure, capacities and everyday developments.

SOCIALLY DISCRIMINATE OR INDISCRIMINATE EVENTS?

The external force/extreme event view also tends to accept that death and loss in disasters are a function of the type, intensity and distribution of the damaging agent, but socially indiscriminate – affecting old and young, rich and poor, men and women at random and, in effect, equally. The opposing view emphasizes the evidence that who lives and dies, the scale and severity of losses, are strongly prefigured in the social order. They clearly identify certain persons or groups, places or activities, as much more vulnerable. Again, this directs us to conditions and developments in society. Social vulnerability, and adaptive capacities must be examined, and the roles of cultural and environmental contexts. The result is a view of the origins of disaster that departs more or less strongly from mainstream, agent-specific investigations that emphasize exceptional or disaster-specific responses.

VULNERABILITY OR THE HUMAN ECOLOGY OF ENDANGERMENT

A vulnerability perspective considers especially how communities are exposed to dangers, the ways in which they are readily harmed, and the protections that they lack. When disasters do occur, it directs attention to peoples' capacities to withstand, cope with, mitigate and recover from damages. It examines whether, or how well, they are served by organized relief and rehabilitation measures (Table 1).

Here, the problem of disaster is seen to depend primarily upon on-going societal conditions or developments. It depends upon many aspects of society that, by design or default, can decrease or exaggerate the impact of damaging agents. However, if particular forms of vulnerability relate to particular hazards, peoples' vulnerability can arise more or less independently of where flood, storm, epidemic disease or explosion may occur, and their severity. Vulnerability is largely created within, distributed or offset by, the social order. It depends upon how society treats people according to their age, class, or religion; on gender and

Table 1 The main forms and conditions of vulnerability (after Hewitt, 1997, 144)

1. Exposure to dangerous agents or conditions through occupation, location near a dangerous facility, lifestyle
2. Physical weakness or susceptibility through predisposition, age, disability, impairment of persons, poor design of buildings, insecure practices
3. Disadvantage, or structural weakness involving lack of resources, capabilities or access to facilities, often identified with poverty, dependency, exploitation and social injustice
4. Defenselessness or lack of protection to counter weaknesses, to prevent or avoid dangerous situations, emphasizing protection otherwise available in the society concerned
5. Lack of resilience or capacities, where the means and ability to respond are limited and impaired, options are not recognized, and readiness for creative responses missing
6. Powerlessness, an inability to gain access to, or influence safety arrangements; exclusion from and inability to get the attention of the relevant and influential social agencies

other divisions of labor; on cultural values and legal rights (O'Riordan, 1986; Enarson and Morrow, 1998).

Of course, everyone is vulnerable, at least in an absolute or worst case sense. Some persons, activities or places may be vulnerable because neither they nor anyone else has the means to make them safer. Moreover, people in all situations can, and usually do, accept some risks.

Yet, a majority of those who actually suffer harm and loss in most disasters are found to be vulnerable in ways that some, if not most, members of their own, or other societies, are not. The homes which collapse in earthquake are not merely structurally unsafe and badly sited, but usually house people disadvantaged in cultural and economic ways (Blaikie *et al.*, 1994). In almost all famines, it is the already hungry or malnourished, who suffer starvation and death (Sen, 1981). Many if not most refugees have been displaced after long-term struggles with livelihood insecurity or coercive treatment (Zwi and Ugalde, 1991). Some people seem driven into, or trapped by, vulnerability syndrome. Rather than a single weakness or category, they are victimized by a whole social context (Wisner, 1993). However, even in less severe cases, vulnerability is a relative condition, and can only be defined and assessed in relation to the safety which others actually enjoy.

HUMAN ADAPTABILITY

Although vulnerability mainly suggests weakness and absent protections, studies of it find that people's active capabilities or resilience in relation to dangers, are at least as important. Only the most impaired or injured are not, in some measure, active agents in their own survival, if not in larger community needs.

There is no form of vulnerability which some or all societies do not deliberately work to reduce and offset. Social safety nets, defenses, insurance, and emergency preparedness can and do mitigate harm. They reduce and offset weakness, or offer protections for the specially vulnerable, for children, the elderly or disabled.

Direct and longer-term involvement with communities at risk, shows how survival turns upon community values and arrangements (Watts, 1983; Vaughan, 1996). Adaptive and coping capacities are also embedded in on-going activity. Patterns of everyday living prefigure and constrain the readiness, capacities, resources and values that will apply in a crisis. Resourcefulness and creative responses may be seriously undermined where social conditions promote submissiveness and dependency, or where individuals, groups and communities find their autonomy constantly opposed and suppressed by the interests of powerful others. Indeed, people who are relatively more endangered, less well-provided with protections and lacking in adaptive capacities, are typically found to be politically weak as well. Defenselessness most often reflects powerlessness (O'Riordan, 1986; Watts and Bohle, 1993). The victims of natural and technological disasters have often, already, been victimized by an unjust, exploitative or coercive social order (Corradi *et al.*, 1992; Kroll-Smith, 1995).

In these views too, the forms of vulnerability and adaptive behavior are not taken as peculiar to what happens during an earthquake or fire. Rather, they direct us to domestic and local spaces of risk, and everyday behavior and conditions. However, as global change studies show, everyday life is nowadays subject to unprecedented changes and risks.

SOCIAL RISK AND DISASTERS IN A TECHNOCRATIC AGE

Today, vulnerability and its changing forms relate, increasingly, to the emergence and reach of late modern society and its institutions. The geographies and contexts of risk relate mainly to whether people live within or outside such societies, and the degree of their access to, or exclusion from, modern safety and security measures.

In the wealthier, late modern countries and enclaves, permanent arrangements, infrastructures and organizations play an overwhelming role in vulnerability, and in the origins of, as well as responses to, disaster. This applies, in particular, to routine public safety and everyday, widely dispersed or chronic dangers. They include institutionalized arrangements to deal with traffic accidents, household fires, illnesses of childhood or the elderly, crime or product safety. These have become the special responsibility of

government, private or non-governmental organizations, oversight agencies or service industries. They involve the largest expenditures on public safety.

Vast sums of private and public money are allocated to insurance and welfare; to health, policing and the courts; to fire fighting, transportation or product safety; to the armed forces, factory or mines inspections, and so on. Whole disciplines and professions have developed to research and carry out these functions, or to train staff for the relevant organizations.

Although, by definition, disasters overwhelm routine arrangements, in modern societies responses are largely dependent upon large, complex organizations, albeit from outside the disaster zone. Such organizations, as Perrow (1995) expresses it, •••have in effect absorbed society . Moreover, the permanent large organizations dealing with chronic dangers, play decisive roles both in limiting and in responding to disasters. In most countries this applies, especially, to the role of the armed forces. Meanwhile, large organizations, their policies, priorities and performance substantially determine each person's safety or vulnerability. Nearly all jobs are in large public and private organizations, and a person's job and status in them, is the main determinant of exposure to, and protection from, risks.

On the other hand, to a growing extent, disasters arise from the fallibility, excesses, or deliberate actions, of these organizations. This is so for many of the graver dangers identified with global change: accidents that arise within or from complex, high-risk organizations. Examples include the Three-Mile Island and Chernobyl nuclear disasters, Bhopal, and the Exxon Valdez oil spill (Sagan, 1993).

The German sociologist Beck (1992) has identified such predicaments with what he calls the risk society. He believes we now confront a new modernity in which the manufacture of risks, or technoscientific accidents, increasingly overshadows productive and progressive forces. He sees people's risk positions, essentially another way of describing vulnerability, as related not just to hazardous processes but to work, food, a place to live, even their leisure choices. Yet, these positions are embedded in, and change with, the more or less extensive risk-producing, risk-averting and risk-redistributing strategies of modern institutions. In relation to global change, he describes the epochal problem of the risk society in this way:

> ••• As nature becomes permeated by industrialization and as tradition is dissolved, new types of incalculability emerge ••• [specifically] ••• the production of risks is the consequence of scientific and political efforts to control or minimize them.
>
> (Beck, 1998)

Meanwhile, social risk positions and the geography of disasters have been transformed according to access to, or exclusion from, the benefits of modern safety arrangements and large organizations. There can be sharp variations in vulnerability within cities and countries, even between members of the same family or neighborhood (Enarson and Morrow, 1998). Disasters, or possible responses, in different regions of the earth, depend upon developments and connections involving those same large organizations. At least, they are involved in massive humanitarian relief efforts in areas remote from cities and industrial development. Meanwhile, the world social space of vulnerability is increasingly a direct or indirect consequence of conditions related to national or global economic, technological, environmental and political influences (Blaikie *et al.*, 1994; Oliver-Smith and Hoffman, 1999).

KNOWLEDGE WORKERS

This condition also places the problem of knowledge center stage. Mainly through the role of large organizations, computerized and centralized controls, thought and action are increasingly in the hands of, or dependent upon, knowledge workers – most of whom are not teachers, academics or intellectuals! They are persons in public and private organizations who work with and through formal systems of information gathering, assessment and communication. They rely upon training in, and facility with, standard technical practices. Even those called practitioners – ambulance dispatchers, traffic policemen, volunteers in public charities – spend much of their time writing things down, exchanging and reading what (official or non-official) others communicate, developing or reading other people's web pages, analyzing and assessing data or undergoing training in such matters. Knowledge workers include virtually all the people who help to generate plans, consultants and inspectors' reports, write field manuals, sit in committees or dispense orders.

To be sure, there are still huge numbers of people, worldwide, whose work is largely manual, whether producing food or raising children, who do not, or cannot participate in the textual and computer literate world of modern knowledge. But ever fewer of them act independent of some technology, procedures and goals, designed and monitored by the knowledge workers. All of this is also part of large organizations and extends their roles.

However, disaster is one of those problems – others are violence, madness, poverty in a world of plenty, crime, addiction, abuse or dropping out – that are specially intractable for, and uniquely threatening to, the rationalized, systematized social order of large institutions (Hewitt, 1983; Burchell *et al.*, 1991). Indeed, part of the current debate concerns how far our approaches to disaster, and their failings, reflect the preoccupations of technocratic organization more than the objective dangers or the predicaments of those at risk (Ericson and Haggerty, 1997).

WHOSE DISASTERS? THE REVERSAL OF EXPERTISE

The contrast between the experience or witnessing of disaster, and technical knowledge work, is not merely one of time and place, but of differing concerns and perspectives. If this seems obvious, it is often ignored in technical disaster work. This not only tends to treat issues from a distance, abstractly and in terms of governmental or other centrally organized responses. It is well known that large organizations can be quite insensitive to, or unaware of, the concerns not merely of vulnerable communities and those trying to cope with and recover from catastrophe, but of other organizations (Turner, 1978). Part of the problem for responsible agencies and professionals, is the extent to which knowledge and understanding of vulnerability and disaster involve a rethink of their normal ideas.

It is not only that suffering and witnessing extraordinary violence justify the oft-quoted feeling of survivors that anyone who was not there could not possibly know or understand. In addition, to the extent that disaster depends upon personal and societal vulnerability or resilience, then it becomes a matter of grasping the circumstances, needs and concerns of those at risk. Vulnerability is always closely bound up with where people live, their work and specific relations to habitat and the larger society, their values and expectations. The idea that someone from outside, however sophisticated in, say, flood hydrology, engineering safety or radiation ecology, can recognize and give a balanced assessment of risks and needs, is socially absurd – especially when there is also an inability to speak the language of those at risk, or being new to the culture, group, and habitat involved. Such is often the plight of the visiting expert; an urbanite speaking to rural peasants, a well-off person trying to understand those living in abject poverty, a survey team of men where women are the most vulnerable, or play a decisive role in family safety.

Without doubt there are things the expert is uniquely able to say and do. The impacts of modern technologies, practices and organizations reach into every corner of the earth, if they do not aggravate or cause the disasters that occur. Experts have privileged opportunities to represent the vulnerable to the powers that be, and *vice versa*. Since large organizations control important resources and actions, it matters how the professionals state the case.

But social understanding requires listening to and recognizing the cultural and psychosocial condition of those at risk. Again, a vulnerability perspective directs us to their situation and predicaments, rather than just impersonal dimensions, processes and organizational procedures. Or rather, it confronts us with the problem of translating between technical understanding into what are variously termed local, gendered, vernacular and indigenous languages or knowledge (Becker, 1991).

CONCLUDING REMARKS

To summarize, social vulnerability and adaptability depend, in large measure, upon on-going material and social life – or their transformations. For that reason, they are surely of decisive importance for the growing concern over sustainable human communities and habitat relations. To adapt to global changes in a sustainable way is, above all, to avoid disastrous collapses. It requires that developments should not put people and property on a collision course with great hazards, or make them less able to withstand or cope with existing risks. It suggests that technological innovations or projects, even consumer goods, that may carry high risks should be subject to exceptional constraints.

REFERENCES

Beck, U (1992) *Risk Society: Towards a New Modernity Translated*, ed M Ritter, Sage Publications, London.

Beck, U (1998) Politics of Risk Society, in *The Politics of Risk Society*, ed J Franklin, Polity Press/Blackwell, Oxford.

Becker, A (1991) A Short Essay on Languaging, in *Research and Re exivity Sage*, ed F Steier, London, 228–235.

Blaikie, P, Cannon, T, Davis, I, and Wisner, B (1994) *At Risk: Natural Hazards, People's Vulnerability and Disasters*, Routledge, London.

Burchell, G, Gordon, C, and Miller, P, eds (1991) *The Foucault Effect: Studies in Governmentality*, Harvester-Wheatsheaf, London.

Corradi, J E, Fagen, P W, and Garreton, M A, eds (1992) *Fear at the Edge: State Terror and Resistance in Latin America*, University of California Press, Berkeley.

Elliot, G (1972) *Twentieth Century Book of the Dead*, Penguin, Harmondsworth.

Enarson, E and Morrow, B H, eds (1998) *The Gendered Terrain of Disaster: Through Women's Eyes*, Praeger, Westport, CT.

Ericson, R V and Haggerty, K D (1997) *Policing the Risk Society*, University of Toronto Press, Toronto.

Hewitt, K (1983) The Idea of Calamity in a Technocratic Age, in *Interpretations of Calamity from the Perspective of Human Ecology*, eds Hewitt, K, Allen, and Unwin, London, 3–32.

Hewitt, K (1997) *Regions of Risk: a Geographical Introduction to Disaster*, Addison Wesley Longman, London.

IFRCRCS (1998) *World Disasters Report*, International Federation of Red Cross and Red Crescent Societies, Martinus Nijhoff, Dordrecht.

Kreps, G (1995) Disaster as systemic event and social catalyst: a clarification of subject matter, *Int. J. Mass Emerg. Disasters*, 13, 255–284.

Kroll-Smith, S (1995) Toxic Contamination and the Loss of Civility: 1994 MSSA Plenary Address, *Sociolog. Spect.*, 15, 377–396.

Lemons, H (1957) Physical Characteristics of Disasters: Historical and Statistical Review, in *Disasters and Disaster Relief*, ed De Witt Smith, *Ann. Am. Acad. Political Soc. Sci.*, 309(January), 1–14.

Morren, Jr, G E B (1983) A General Approach to the Identification of Hazards and Response, in *Interpretations of Calamity*

from the Perspective of Human Ecology, ed K Hewitt, Allen and Unwin, London, 284–297.
Oliver-Smith, A R and Hoffman, S M (1999) *The Angry Earth*, Routledge, New York.
O Riordan, T (1986) *Coping with Environmental Hazards, in Geography, Resources and Environment: Theses from the Work of Gilbert F. White, II*, eds R W Kates and I Burton, University of Chicago Press, Chicago, 212 309.
Pelanda, C (1981) Disasters and sociosystemic vulnerability, *Preliminary Paper #68*, Disaster Research Center, University of Delaware, Newark, DE.
Perrow, C B (1995) Society at Risk in a Society of Organizations, in *Populations at Risk in America: Vulnerable Groups at the End of the Twentieth Century*, eds G J Demko and M C Jackson, Westview Press, New York, 19 35.
Ploughman, P (1995) The American Print News Media Construction of Five Natural Disasters, *Disasters*, **19**(4), 308 326.
Por riev, B N (1998) Issues in the De nition and Delineation of Disasters, in *What is a Disaster?* ed E H Quarantelli, Routledge, London, 56 72.
Quarantelli, E H, ed (1998) *What is a Disaster?* Routledge, London.
Quarantelli, E L and Dynes, R (1977) Responses to Social Crisis and Disaster, *Annu. Rev. Sociol.*, **3**, 23 49.
Sagan, S D (1993) *The Limits of Safety: Organizations, Accidents and Nuclear Weapons*, Princeton University Press, Princeton, NJ.
Scarry, E (1985) *The Body in Pain*, Oxford University Press, New York.
Sen, A (1981) *Poverty and Famines: an Essay on an Entitlement and Deprivation*, Clarendon Press, Oxford.
Sorokin, P (1942) *Man and Society in Calamity*, E P Dutton, New York.
Turner, B A (1978) *Man-made Disasters*, Wykeham Publications, London.
Vaughan, D (1996) *The Story of an African Famine: Gender and Famine in Twentieth Century Malawi*, Cambridge University Press, Cambridge.
Watts, M J (1983) *Silent Violence: Food, Famine and Peasantry in Northern Nigeria*, University of California Press, Berkeley, CA.
Watts, M J and Bohle, H G (1993) Hunger, Famine and the Space of Vulnerability, *GeoJournal*, **30**(2), 117 125.
White, G F, ed (1974) *Natural Hazards: Local, Regional, Global*, Oxford University Press, New York.
Wisner, B (1993) Disaster vulnerability, geographical scale and existential reality, in *Worlds of Pain and Hunger: Geographical Perspectives on Disaster Vulnerability and Food Security*, ed H Bohle, Freiburg Studies in Development Geography, University of Freiburg, Freiburg-im-Briesgau.
Zwi, A and Ugalde, A (1991) Political Violence in the Third World: a Public Health Issue, *Health Policy and Planning*, **6**(3), 203 217.

FURTHER READING

Green, J (1997) *Risk and Misfortune: the Social Construction of Accidents*, University College London Press, London.
Kreps, G R (1998) Disaster as Systemic Event and Social Catalyst, in *What is a Disaster?* ed E H Quarantelli, Routledge, London, 31 55.
Pelanda, C (1982) Disaster and Socioeconomic Vulnerability, in *Social and Economic Aspects of Earthquakes: Proceedings, Third International Conference, Bled, Yugoslavia, 1981*, eds B G Jones and M Tomazevic, Cornell University, New York, 67 91.
Perrow, C B (1984) *Normal Accidents: Living with High-risk Technologies*, Basic Books, New York.
Waterstone, M, ed (1992) *Risk and Society: the Interaction of Science, Technology and Public Policy*, Kluwer, Boston, MA.

Health and the Environment: Theoretical Approaches

see Theories of Health and Environment (**Volume 5**)

Hetch Hetchy Valley

see Environmental Movement – the Rise of Non-government Organizations (NGOs) (**Volume 5**)

Hinduism and the Environment

O P Dwivedi
University of Guelph, Guelph, Canada

The Hindu religion offers a challenging perspective on the subjects of respect for nature (including the principle of the sanctity of all life on this planet and elsewhere) and environmental conservation. In Hinduism, human beings are not given any absolute dominion over their own lives or over the life of any non-human being; only the Supreme God has absolute sovereignty over all creatures, including humanity s life and death. The sacredness of God s creation requires that no damage be in icted on other species without adequate justification. People are not supposed to act as viceroys of God over the earth; nor should they assign degrees of relative worth to other species. All life forms (human and non-human alike) are of equal value and all

have the same right to existence, but only humans have special obligations to care for the environment. In addition, only humans, unlike any other species, are obliged to live in harmony with nature by exercising control over their cravings and obsessions, mainly because they have the God-given capability of free will.

THE VEDIC HERITAGE

Hinduism is the name given to the most widespread religion on the Indian subcontinent. It has had a long history, extending back at least to the arrival of what are now thought to be Indo–Iranian tribes in northern India about 2000 before the common era (BCE) – considered the Vedic period – whose beliefs were mingled with, or layered over, the original indigenous religions in the region. Concerning the dates of the Vedic period, there remain differences of opinion among some Western Indologists and Indian scholars. For example, Griffith (1889) mentions that the Vedic period occurred between 1500 and 1400 BCE, while Indian scholars such as Tilak (1936) and Kane (1966) suggest about 5000–4000 BCE. Perhaps a more realistic date could be about 2000 BCE, see Kak (1994). After the literature associated with the Vedic period (known as the Vedas), there appeared the later Upanishads (addenda to the Vedas, and considered to be datable around 600 BCE, (see Hume, 1977, 6 and Deussen, 1980), and an array of stories and epic poems like the Ramayana and the Mahabharata, that together generated the vast diversity of greater and lesser gods, personal and public rituals, and all the elements that populate today's Hinduism.

The ancient Hindu literature is divided into the Shrutis and Smritis. Shrutis denote the entire body of sacred Vedic Scriptures, which are considered to include the four Vedas, the Upanishads, and some other related texts. Smritis are a number of other Hindu sacred Scriptures created after the Shrutis, including the Ramayana and the Mahabharata, the Dharma-shastras (religious books on Hindu law and behavior), and the Puranas (18 books on Hindu mythologies).

The relationship between human beings and nature attracted the seers of the early Vedic period who contemplated the mysteries of the Creation, the place of heaven and earth, and even beyond. Those Vedic seers would not accept as final what they saw around themselves; instead they asked many penetrating questions, not only about life but also about death and destruction. They were equally interested in the mystery of Creation and the establishment of this universe. Through their deep thinking, guesses, conjectures and postulations, they came to acknowledge that the material causes of this creation happened to be the Panch Mahabhutas (Five Great Elements): earth/Prithivi, air/vayu, space/akash, waters/Apah, light–fire/Agni (Aitareya Upanishad, Chapter III, Verse 3).

These five Mahabhutas are cosmic elements that create, nurture, and sustain all forms of life, and after death or decay, they absorb what was created earlier; thus they play an important role in preserving and sustaining the environment (Dwivedi, 1997a). It should be noted that these Mahabhutas have been deified in the Vedic, and also later in the Puranic literature (starting from 300 BCE to 900 CE). Further, they, all together, are regarded as all-pervasive and omnipresent elements having great creative potency; in addition, they together constitute Brahman (the Almighty God, the all-pervading One, the Supreme Godhead) who manifested the universe and whose manifestation goes on forever. As mentioned in the Shvetashvatara Upanishad:

> The Brahman by whom this entire universe is engulfed. ••• this creation is governed by Him as well as the five great elements: earth, air, space, water, and light.
>
> (Shvetashvatara Upanishad, Chapter 6, Verse 2)

Prithivi (Earth)

The Rig Veda describes Prithivi as a divinity as well as one of the Mahabhuta elements. She is the mother and upholder of all (Rig Veda, Book X, Hymn 18, Verse 10; and Book 1, Hymn 155, Verse 2). Prithivi is also identified with the goddess Aditi: a mother and protector of the holy cosmic law; she is also regarded as a divine ship, full of life-sustaining harvest. She, along with the four other Mahabhutas, sustains our universe. This relationship between earth and humans is superbly depicted by Rishi Atharva (a rishi is a priest/seer), in the Prithivi Sukta (Hymn to Mother Earth) of the early Atharva Veda. (A Sukta is a short composition of verses which relate to a specific subject under discussion. Each Veda consists of several Suktas). This hymn exemplifies the relevance of environmental sustenance and biodiversity to human beings (Satvalekar, 1958). It consists of several prayers including one that promotes the concept of Vasudhaiv Kutumbkam, or the family of Mother Earth (Dwivedi, 1998).

Vayu (Air)

Vayu is the bond and the thread that keeps the universe together. Vayu is also praised like Prana (the life-sustaining breath) – it is the germ of the world and a transformer of seed; without Prana, nothing survives.

Akash (Space)

The term Akash denotes space rather than sky or ether as mentioned in some translated Vedic literature. It is not a material or physical element. In Chhandogya Upanishad, a discussion takes place among Rishis; they finally agree that it is the Akash (space) out of which all beings

(their souls) come, and it is where they go back after their death; thus, the Akash is the final resting place of all beings (Chhandogya Upanishad, Book I, 9th Part, Verse 1).

Apah (Water)

According to the Vedas, water was the first of the cosmic elements. The Rig Veda says that in the beginning, all was water, and there was darkness which engulfed it (Rig Veda, Book X, Hymn 129, Verse 3). Further, the Rig Vedic Hymn 23 (Book I, Verses 16–21) considers water to be the reservoir of all curative medicines and of nectar. Waters are the mothers of all beings, and they are the foundations of all in the universe (Lal, 1995, 9). The genesis of the Universe takes place in the primeval water. The following verse from the Atharva Veda (Book IXX, Hymn II, Verses 1 and 2) illustrates the place of water in our lives:

> may the waters from the snowy mountains bring health and peace to all people. May the spring waters bring calm to you. May the swift current be pleasing to you; and may the rains be a source of tranquility to all. May the waters of the oasis in the desert be sweet to you; and so be the waters of ponds and lakes. May the waters from wells dug by humans be good to them, and may the healing powers of water be available to all beings.

Hindus consider water to be a powerful medium for purification and also a source of energy. Sometimes, just by the sprinkling of pure water in religious ceremonies, it is believed that purity is achieved. That is why, in the Rig Veda, prayer is offered to the deity of water:

> the waters in the sky, the waters of rivers, and water in the well whose source is the ocean, may all these sacred waters protect me.
>
> (Rig Veda, Book VII, Hymn 49, Verse 2)

Agni (Light/Fire)

Agni is considered in the Vedas as the spring of our life because it creates life on earth. Agni, in later Vedic description, is known as the sun and light. In the form of the sun, Agni is regarded as the soul, and also as the ruler and preserver of the world (Maitrayana Upanishad, Chapter 6, Verse 35).

These five Great Elements, taken together, denote the deep bond between human beings and creation; as such, Hindus are urged to show proper respect and due care for these essential elements.

RESPECT FOR FAUNA IN HINDUISM

The most important aspect of the Hindu religion pertaining to the treatment of animal life is the belief that the Supreme Being was himself incarnated in the form of various species. This Lord says, in one Purana:

> this form is the source and indestructible seed of multifarious incarnations within the universe, and from the particle and portion of this form, different living entities, like demi-gods, animals, human beings and others, are created.
>
> (Srimad-Bhagavata Mahapurana, Book I, Discourse III:5)

This is further demonstrated by a famous series of divine incarnations, as pertinently enunciated by Dr Karan Singh (1986) in the Assisi Declaration. (These were declarations on religion and nature made by religious leaders representing Buddhism, Christianity, Hinduism, Judaism and Islam at the Basilica di S Francesco, Assisi, Italy, 29 September 1986 at the time of the 25th Anniversary of the World Wide Fund for Nature):

> the evolution of life on this planet is symbolized by a series of divine incarnations beginning with fish, moving through amphibious forms and mammals, and then on into human incarnations. This view clearly holds that man did not spring fully formed to dominate the lesser lifeforms, but rather evolved out of these forms itself, and is therefore integrally linked to the whole of creation.

Among the series of incarnations of Vishnu (Vishnu is one of the Hindu Trinity who incarnates on earth to protect righteousness and good), was his first appearance in the form of Matsya (a giant fish), when due to the continuous rain, the entire earth was going to be submerged in water. This giant fish appeared with a mysterious boat into which a human pair went along with samples of all herbs, seeds, and all kinds of living creatures. After the rains stopped, a new creation began. The second incarnation was in the form of Kurma (a tortoise), when because of the churning of the Ksheer Sagar (Milky Ocean), a tortoise was needed to hold Mount Mandarachal, which was being used as a churning pestle. Without churning the sea, the demi-gods would not obtain the nectar to make them invincible when fighting with demons. The third incarnation was in the form of Varaha (a boar): when a demon king Hiranyaksha captured the earth, the boar hid it under massive garbage with unendurable stench and filth; and thus the earth was rescued. The fourth incarnation was in the form of Narasingha (half-man half-lion) to fight the demon king Hiranyakashyap, who after defeating all demi-gods had declared himself to be the God of universe. Since the demon King was indestructible by man or animal, the god Vishnu had to be incarnated in the form of a man–lion. The fifth incarnation was in the form of Vamana (a dwarf) when another demon-king Bali became so powerful that he defeated all demi-gods and conquered the entire universe. So Vishnu appeared in the form of a dwarf brahman, visited the court of Bali, and retrieved the entire universe, and restored demi-gods to their place in heaven. The sixth, seventh, eighth

and ninth incarnations were in the form of humans to protect the earth from oppressors and tyrants. But as Rama (the seventh incarnation), he was closely associated with monkeys, and as Krishna (the eighth incarnation as a cowherd) he was surrounded by cows, peacocks, deer, etc. The ninth incarnation was in the form of the Buddha who preached non-violence towards all creatures. So, since Lord Vishnu was incarnated into various living forms, Hindus accord reverence to species.

The Srimad Bhagavatam Mahapurana further states:

> one should look upon deer, camels, monkeys, donkeys, rats, reptiles, birds, and flies as though they were their own children.
> (Book 7, Discourse XIV:9)

In addition to this, many animals and birds are revered by Hindus as these are consort of various gods, goddesses, and incarnations, as illustrated in Table 1.

Almost all the Hindu scriptures place a strong emphasis on the notion that God's grace cannot be received by killing animals or harming other creatures. That is why not eating meat is considered both appropriate conduct and one's Dharma (Dharma is considered as the Cosmic order, duties and obligations based on social and religious laws handed down by tradition and practices).

Further, in the Manusmriti, a warning is given that:

> a person who kills an animal for meat will die of a violent death as many times as there are hairs of that killed animal.
> (Manusmriti, Chapter 5, Verse 38)

Table 1 Animals and Wildlife Associated with Gods/Goddesses[a]

Animals/Birds	Associated with Gods and Goddesses
Lion	Durga
Wild goose	Brahma
Elephant	Indra, Ganesha
Bull	Shiva
Rat	Ganesha
Swan	Saraswati
Eagle	Vishnu
Serpent	Shiva, Sun
Fish	Kama
Monkey	Rama
Horse	Sun
Peacock	Kartikeya, Saraswati
Parrot	Kama
Owl	Lakshmi
Vulture	Shani
Crocodile	Ganga River
Tortoise	Yamuna River
Tiger	Katyayani
Dog	Bhairava, Dattatreya
Deer	Vayu
Donkey	Shitla
Makar	Kama, Varuna

[a] *Source:* Dwivedi and Tiwari (1987).

Similarly, the Yajnavalkya Smriti warns of hell-fire (Ghora Naraka) to those who are killers of domesticated and protected animals:

> the wicked person who kills animals which are protected has to live in hell-fire for the days equal to the number of hairs on the body of that animal.
> (Yajnavalkya Smriti, Acaradhyayah, Verse 180)

Further, in the Narsingha Purana, it is mentioned that a person who roasts a bird for eating will surely be a sinner.

In Mahabharata, there is a story concerning Rishis and God debating the merits of offering grain or the sacrificial lamb (goat) at the end of Yajna; the Rishis insisted that according to the Vedas, the sacrificial material ought to be grain only. It seems that the practice of grain sacrifice might have started from that time. In summary, Hindu scriptures enjoin its followers from hurting animals.

RESPECT FOR FLORA IN THE HINDU RELIGION

As early as in the time of the Rig Veda, tree worship was quite popular and universal in India. The tree symbolized the various attributes of God to the Vedic seers. The Rig Veda regarded plants as having divine powers, with one entire hymn devoted to their praise, chiefly with reference to their healing properties (Rig Veda, 10.97). During the period of the great epics and Puranas, the Hindu respect for flora expanded further. Trees were considered as being animate, feeling happiness and sorrow. Green trees were likened to a living person, as illustrated by the following verse from the Brihadaranyaka Upanishad (3.9.28):

> just like a tree, the prince of the forest, so the man is, in truth; His hairs are the leaves, his skin resembles the external bark. Out of his skin streams forth the blood, like the juice or the sap out of the tree.
>
> > It flows out from the wounded man, like the sap of the tree, when it is cut.
> > The flesh is comparable to the wood (of the tree), the sinews are like the inner bark;
> > The strong bones are like the inner core of the wood, the marrow resembles the marrow (pith) of the tree.

The Hindu worship of trees and plants has been based partly on utility, but mostly because of the divinity assigned to them. Ancient Hindus considered it their duty to save trees, and in order to do so they attached to every tree a religious sanctity. The following deities, for example, are considered to have made their abode in these trees/plants: Lakshmi in Tulasi (*Ocinlum sanctum*), Shitala in Neem, Vishnu in Pipal/Bodhi (*Ficus religiosa*), Buddha in Ashoka, Shiva in Bilva/Bela (*Aegle marmelos*), Brahma and Vishnu in Vata (*Ficus indica*), Van Durga (tree deity) in Asvathha (*Tropbis aspera*). Table 2 further illustrates the names of

Table 2 Trees and Plants Associated with Gods/Goddesses and Spirits[a]

Trees/plants	Associated with Gods/Goddesses
Asvathha	Vishnu, Lakshmi, Ancestors
Vata	Brahma, Vishnu, Hari, Kuber
Tulasi	Lakshmi, Vishnu, Ancestors
Soma	Moon
Bela	Maheshwar (Shiva)
Ashoka	Buddha, Indra
Amalaki	Lakshmi, Kartikeya
Tala	Buddha, Spirits
Mango	Lakshmi, Govardhan
Kadamba	Krishna
Neem	Shitala, Mansa
Palasa	Brahma, Gandharva
Pipal	Vishnu, Krishna, Ancestors
Fig	Vishnu, Rudra
Narikel (coconut)	Moon
Sheora	Van-Durga, Lakshmi
Sij	Shitala
Tamal	Krishna

[a] *Source:* Dwivedi and Tiwari (1987).

certain trees and plants that are closely associated with various deities.

In India, Nepal and Bhutan, there are many sacred forests and groves where green trees and plants are protected. Also, for Hindus, the planting of a tree is still a religious duty. Thus, sanctity has been attached to many trees and plants. Through such exhortations and various writings, the Hindu religion has provided a system of moral guidelines towards environmental preservation and conservation.

Perhaps this rationale impelled the Hindu law-giver Rishi Manu when he declared that people should not fell green trees even for fuel (Manusmriti, Chapter 11, Verse 64). The cutting of trees and destruction of flora was considered a sinful act. Later, Kautilya's Arthasastra prescribed various punishments for destroying trees and plants (Kautilya's Arthasastra III 19:197):

> for cutting off the tender sprouts of fruit trees, or shady trees in the parks near a city, a fine of six panas shall be imposed; for cutting of the minor branches of the same trees, twelve panas, and for cutting off the big branches, twenty four panas shall be levied. Cutting off the trunks of the same, shall be punished with the first amercement; and felling shall be punished with the middlemost amercement.

Even in modern times, it is popularly believed by Hindus that many trees have their own deity, Vriksha-devata or Van-devi, who is worshipped with prayers and offerings of water, flowers, sweets, and encircled by sacred threads (Dwivedi and Tiwari, 1987). That is why for Hindus, forests were never seen as places of darkness and unknown dangers (akin to what Europeans thought), instead these were places of worship, meditation and intimate harmony.

POLLUTION AND ITS PREVENTION IN HINDU SCRIPTURES

Hindu scriptures reveal a clear conception of what can be called an eco-care system. On this basis, a discipline of environmental ethics developed that formulated codes of conduct (Dharma) and defined humanity's relationship to nature. An important part of conduct is maintaining proper sanitation. In the past, this was considered to be the duty of everyone and any default was a punishable offence. As Kautilya wrote (Kautilya's Arthasastra, Book II, Chapter 36, Verse 145):

> the punishment of one-eighth of a pana should be awarded to those who throw dirt on the roads. For muddy water one-fourth pana, if both are thrown the punishment should be double. If latrine is thrown or caused near a temple, well, or pond, sacred place, or government building, then the punishment should increase gradually by one pana in each case. For urine the punishment should be only half.

Hindus considered the cremation of dead bodies and the sanitary maintenance of the human habitat to be essential acts. When, in about 200 BCE, Charaka wrote about Vikriti (pollution) and diseases, caused by pollution of air and waters, specifically about air pollution (Charaka Samhita, Book Vimanastanam, Chapter III, Verse 6.1), he said:

> the polluted air is mixed with unhealthy elements. The air is uncharacteristic of the season, full of moisture, stormy, hard to breath, icy cool, hot and dry, harmful, roaring, coming at the same time from all directions, bad smelling, oily, full of dirt, sand, steam, creating diseases in the body and is considered polluted.

And about water pollution, he wrote (Charaka Samhita, Book Vimanastanam, Chapter III, Verse 6.2):

> water is considered polluted when it is excessively smelly, unnatural in colour, taste and touch, slimy, not frequented by aquatic birds, aquatic life is reduced, and the appearance is unpleasing.

The healing properties and medicinal value of water have been universally accepted – provided it is pure and free from all pollution. Ancient Indian thinkers were aware of the reasons for polluted water. That is why Manu advised:

> One should not cause urine, stool, cough in the water. Anything which is mixed with these un-pious objects, blood and poison, should not be thrown into water.
>
> (Manusmriti, Book IV, Verse 56)

Today, many rivers are still considered sacred. Among these, the river Ganga is considered by Hindus as the most sacred body of water. In Pravascitta Tatva (1.535), disposal of human waste or other pollutants there has been prohibited since time immemorial (Dwivedi and Tiwari, 1987, 84):

> one should not perform these 14 acts near the holy waters of the river Ganga: i.e., putting excrement, brushing and gargling,

> removing cerumen from body, throwing hairs, dry garlands, playing in water, taking donations, performing sex, having attachment with other sacred places, praising other holy places, washing clothes, throwing dirty clothes, thumping water and swimming.

Needless to say, this prohibition has been more honored in the breach than the observance nowadays.

On pollution in general, the great epic, Mahabharata mentions that:

> from pollution, two types of diseases occur in human beings. The first, which is related with body and the other with mind, and both are inter-connected. One follows the other and none exists without the other.
>
> (Mahabharata, Rajdharmanusasanparva, Chapter 16, Verses 8 9)

The above-listed examples show that during ancient times, eco-care was proclaimed, and probably practiced where possible. Through these injunctions and religious duties, a mechanism was developed to create a proper respect for nature a mechanism based on cosmic ordinance and divine law which provided a base for eco-spirituality.

KARMA, THE CYCLE OF BIRTH AND REBIRTH AND THE ENVIRONMENT

Karma means all deeds/actions performed and their consequences, which govern a person s life, and which may continue in the next life because of the philosophy of the migration of soul from one birth to another. People often confuse the Law of Karma with the Law of destiny. The Law of Karma may be said to mean that each act, willfully performed, leaves a consequence in its wake. These consequences, also called Karma-Phala (fruits or effects of action) remain always with us, although their impact may not be felt immediately. It also means that every action performed creates its own chain of reactions and events, some of which are immediately visible, while others take time to surface. Environmental pollution is but one example of the Karma of those people who thought that they could continue polluting the environment without realizing the consequences of their actions for future generations. As a matter of fact, there is a boomerang effect of our Karma; or to put it another way, whatever action we do on others recoils, although not always by a direct reaction and not immediately but often by unconnected ways, on our lives, and sometimes in its own exact measure (Dwivedi, 1994).

Once the cycle of Karma starts, it continues without a break; and although a person may be dead, yet Karma survives in the form of a memory on to the next life of a departed soul. As it is mentioned in Mahabharata:

> an action which has been committed by a human being in this life, follows him again and again (whether he wishes it or not).
>
> (Shanti Parva, 139:22)

It is difficult to say when a human being first got involved in the cycle of Karma, but once a person is in it, he/she cannot escape, even if incarnated in new and different forms in this world. The main reason for its continuation is the indestructible nature of the energy of Karma; it is this energy that may appear today under one name and form, but will reappear under another name and form when the former has died. In this way, Karma is related to the Hindu concept of re-birth (Punarjanma) because the reappearance in various forms is the cycle of birth and death. That is why, in Hindu scriptures, the Atman (soul) never dies, it is eternal, but at the same time it is involved in the cycle of Karma. It is for this reason that Hindus believe that one must suffer tomorrow for what one does today; and the day after tomorrow, for what one does tomorrow; and finally, one suffers in the next life for what one does in this life; and in this way the cycle of Punarjanma continues. This is what was also said by Bhisma to King Yudhisthira after the Great War of Mahabharata:

> O King, although a particular person may not be seen suffering the results of his evil actions, yet his children and grandchildren as well as great grandchildren will have to suffer the results of those actions.
>
> (Mahabharata, Shanti Parva, 129:21)

The environmental crisis facing humanity today relates to the concept of Karma, which becomes even more significant. For in Hindu scripture, the rewards or punishment of Karma are described as events that will happen in the future, and sacrifice is described as Apurva (futuristic) Karma-Phala, the results of which have not yet been seen. We need to realize that the destruction that we are inflicting on our natural surroundings will have future repercussions, which will be felt not only by our children, but also by ourselves due to the Law of Karma, which causes us to return to earth in a subsequent birth.

If we recognize and act on this philosophy of life, then and only then will people start paying due respect to nature as well as taking a stake in the care for the environment. Finally, the concept of Karma envisions an interrelatedness between what we do in this world and what may result now and/or in the future. This appreciation is the key to the point being made here. All of our actions are interrelated with and interconnected to what eventually happens in this world. Although we may not face the consequences individually, nevertheless, someone else will be burdened with or benefit from our actions. It is in this context that the concept of Karma as a guiding force to protect the environment becomes meaningful.

The Hindu belief in the cycle of birth and rebirth, wherein a person may come back as an animal or a bird, gives these species not only respect, but also reverence. This provides a solid foundation for the doctrine of Ahimsa non-violence (or non-injury) against animals and human beings alike. Hindus have a deep faith in the doctrine of non-violence. It

should be noted that the doctrine of Ahimsa presupposes the doctrine of Karma and the doctrine of rebirth (punarjanma). The soul is continually reborn in different life forms such as birds, fish, animals, and humans. Based on this belief, there is a profound opposition in the Hindu religion (as well as in Buddhist and Jain religions) to the institutionalized killing of animals, birds and fish for human consumption. Not only, as mentioned, do almost all of the Hindu scriptures place strong emphasis on the notion that God's grace can be received by not killing his creatures or harming his creation (Dwivedi, 1997a); but further, the pain a human being causes to other living creatures will eventually return to that person later, either in this life or in a later rebirth. It is through the transmigration of the soul that a link is provided between the lowliest forms of life with human beings.

HINDU ECO-SPIRITUALITY TODAY?

Hindu eco-spirituality, in ancient and medieval times, has provided a system of moral guidelines towards environmental preservation and conservation. Environmental ethics, as propounded by ancient Hindu scriptures and seers, were practiced not only by common persons, but even by rulers and kings. They observed these fundamentals, sometimes as religious duties, often as rules of administration or obligations for law and order, but always as principles properly knitted into the Hindu way of life. That way of life enabled Hindus, as well as other religious groups residing in India, to carefully use natural resources, but not to have any divine power of control and dominion over nature and its elements.

If such has been the tradition, philosophy, and ideology of Hindu religion, what then are the reasons behind the present state of environmental crisis facing India; and more specifically, the seeming indifference about environmental protection and conservation among the masses? As we have seen, ethical beliefs and religious values influence people's behavior towards others, including our relationship with all creatures and plant life. If, for some reason, those noble values become displaced by other beliefs, that are either thrust upon the society or transplanted from another culture through invasion (as happened later in India), then the faith of the masses in its earlier spiritual tradition is shaken.

Given about nine centuries of foreign religious and cultural domination by Islamic and Christian cultures, which penetrated all levels of Hindu society, appropriate answers and leadership did not come from its religious leaders and priests; consequently, the masses became ritualistic and inward looking. Added to this are the forces of materialism, consumerism, individual and corporate greed, and the capriciousness and corruption among politicians, bureaucrats and business people. Under such circumstances, religious values that acted as sanctions against environmental destruction have been sidelined as those insidious forces inhibit the religion from continuing to transmit the ancient values which encouraged respect and due regard for God's Creation (Dwivedi, 2000b).

How can the original ancient values and wisdom be transmitted into practice? Can there be an approach for a Dharmic ecology? Can there be a practical Dharmic ecology? And can it be practiced in India?

Many people wrongly believe that India is a Hindu country, not realizing that it is a secular state where secularism, as practiced nowadays, means a disregard for all religions, especially the mainstream Hindu religion and culture. As such, it will take time for any Dharmic ecology to be operationalized in India. Besides, it is not sufficient to examine and extol the ancient wisdom of Hindu seers, to dwell on the Vedic and Puranic heritage, and then simply hope that a self-correcting process and automatic implementation of those noble ideas will take place to remedy those environmental problems.

What is more important is how to put into practice an ecological vision and to make it relevant to modern times. A common strategy for environmental stewardship needs to be developed so that, in the end, people will subscribe to an environmentally caring world. That strategy requires that people show their commitment and care, both individually and collectively, for Mother Earth, and for her natural gifts, which sustain our life-system. Because unless people are able to inculcate eco-care vision, there will be no sustainable future.

It is worth pointing out in this context that we can draw on the powerful concept of *satyagraha*, which literally means persistence and endurance for the truth. This concept excludes the use of violence, because it is believed that, since humans are not capable of knowing absolute truth, we are not competent to punish. The word *satyagraha* was initially coined in South Africa by the most famous modern Indian, Mahatma Gandhi (*see* **Gandhi, Mohandas Karamchand**, Volume 5) to distinguish the non-violent resistance of the Indians of South Africa from the contemporary passive resistance of suffragettes and others. Non-violent resistance embodies five basic elements that ought to be followed by those who wish to use this technique. These five elements are purity of: (a) motive; (b) means to be used; (c) suitability of place; (d) time; and (e) the mental status of the doer/agent. An action is a vishuddha dharmic ashtra (a righteous and proper weapon) if it meets those five conditions. Thus, satyagraha is not conceived as a weapon for the weak only; furthermore, it is different from civil disobedience, passive resistance, and non-cooperation.

Satyagraha differs from passive resistance in that the latter has been regarded as a weapon of the feeble. While it avoids violence, passive resistance does not

exclude the use of violence if, in the opinion of the passive resister, the occasion demands it. However, it has always been distinguished from armed resistance, and its use was at one time confined to the Christian martyrs.

Civil disobedience, although being a branch of Satyagraha, is still different to non-violent resistance. Civil disobedience is a response to immoral statutory enactments, and signifies a person's displeasure with some statutory law (it was first used to show displeasure at paying taxes). Civil disobedience presupposes the habit of willing obedience to laws without fear of sanctions; therefore, it can be practiced only as a last resort and by a select few in the first instance at any rate (Gandhi, 1961, 325). Furthermore, it should be noted that non-cooperation predominantly consists of the individual withdrawing cooperation from the state that has become corrupt or immoral, but excludes the use of civil disobedience which results in state-inflicted punishment. Civil disobedience is a branch of Satyagraha, for it can be practiced by children and the masses, but it does differ in fundamentals.

Carried to its utmost limit, Satyagraha is independent of monetary or other material assistance, and definitely independent of physical force or violence. As Gandhi said:

> Indeed, violence is the negation of this great spiritual force, which can only be cultivated or wielded by those who will entirely eschew violence.
>
> (Gandhi, 1961, p. 34)

It is a force that may be used by individuals and communities, and which may be used in domestic as well as political affairs. This universal applicability is thus an example of its permanence and invincibility. It can be used by men, women and children, as long as they are not capable of meeting violence by violence. Therefore, Gandhi believed satyagraha to be an effective tool in the fight for truth and a basis for non-violent resistance.

In relation to the environmental crisis, this Gandhian philosophy has been effectively used in India. The example of the Chipko movement is an illustration that non-violent resistance can work toward environmental protection (Dwivedi, 1997b) (*see* Chipko Movement, Volume 5). Another example is the Narmada Bachao Andolan (the movement to save Narmada River from damming), which has forced the World Bank to reconsider its loan to India for that massive hydroelectricity project.

We therefore need to consider a satyagraha for the environment saturated with the concepts of eco-care and Dharmic ecology; and we need to become effective advocates and practitioners of this approach, rather than staying on the sidelines. Hindu religious leaders should:

- take initiative and help secular institutions by providing timely and appropriate advice to encourage greater integration of eco-care approaches into the educational curricula;
- strengthen the capability of secular institutions to meet their goals of sustainable development and environmental conservation;
- promote the important concept of Sarva bhoot hitey ratah (serving all beings equally);
- take the lead in promoting the concept of Vasudhaiv Kutumbakam, i.e., the family of Mother Earth, and the obligation of humanity to accept a world of material limits;
- protect and restore places of ecological, cultural, aesthetic and spiritual significance; and
- build partnerships across social, economic, political and environmental sectors, including dialogue with other religions and spiritual traditions (Dwivedi, 2000a).

The choice before the Hindu religion (as well as before all other religions) is either to care for the environment or be a silent participant in the destruction of planetary resources. Partnership with secular institutions must be forged and cooperation fostered at local, regional, national and international levels. An environmental and sustainable development strategy, based on the lines suggested above, could offer a way of bridging the gap and making the essential link between secular, scientific and spiritual forces.

NOTES

Verses quoted in English from various Upanishads are from Deussen (1980); verses in Sanskrit for Isavasya Upanishad, Aitareya Upanishad and Shvetashvatara Upanishad are from Upanishad Anka of Kalyan (1949).

REFERENCES

Deussen, P (1980) *Sixty Upanisads of the Veda*, translated from German by V M Bedekar and G B Palsule, Motilal Banarsidass, Delhi, 2 Volumes.

Dwivedi, O P (1994) Our Karma and Dharma to the Environment, in *Environmental Stewardship: History, Theory and Practice*, ed M A Beavis, Institute of Urban Studies, University of Winnipeg, Winnipeg, Canada, 59–74.

Dwivedi, O P (1997a) Vedic Heritage for Environmental Stewardship, *Worldviews: Environ., Cult. Religion*, 1(1), April, 25–36.

Dwivedi, O P (1997b) *India's Environmental Policies, Programmes and Stewardship*, Macmillan Press, London.
Dwivedi, O P (1998) *Vasudhaiv Kutumbakam: A Commentary on Atharvediya Prithivi Sukta*, 2nd edition, Institute for Research and Advanced Studies, Jaipur, 1st edition published in 1995.
Dwivedi, O P (2000a) Dharmic Ecology, in *Hinduism and Ecology: the Interaction of Earth, Sky and Water*, eds C K Chapple and M E Tucker, Harvard University Press, Cambridge, MA, 4–22.
Dwivedi, O P (2000b) *Classical India, A Companion to Environmental Philosophy*, Blackwell, New York, 37–50.
Dwivedi, O P and Tiwari, B N (1987) *Environmental Crisis and Hindu Religion*, Gitanjali Publishing House, New Delhi.
Dwivedi, O P and Tiwari, B N (1999) Environmental Protection in Hindu Religion, in *Ethical Perspectives on Environmental Issues in India*, ed G A James, APH, New Delhi, 172–173.
Gandhi, M K (1961) *Non-violent Resistance*, Schocken Books, New York.
Griffith, R T H (1889) *The Hymns of the Rig Veda*, ed J L Shastri, 1973, Motilal Banarasidass, Delhi.
Hume, R E (1977) *The Thirteen Principal Upanishads*, Oxford University Press, New York.
James, G A (1998) *Ethical Perspectives on Environmental Issues in India*, APH Publishing Co, New Delhi.
Kak, S (1994) The Evolution of Writing in India, *Indian J. History Sci.*, 28, 375–388.
Kane, P V (1963) *Dharmashastra Ka Itihas* (in Hindi and Sanskrit languages), Uttar Pradesh Hindi Sansthan, Lucknow, India.
Kautilya's Arthasastra (1967) ed R Shamasastry, Mysore Publishers, Mysore.
Lal, S K (1995) Pancamahabhutas: Origin and Myths in Vedic Literature, in *Prakrti: An Integral Vision*, ed S Narayanan, Indira Gandhi National Centre for the Arts, New Delhi, 5–21.
Satvalekar, P S D (1958) *Prithivi- Sukta: Atharvaveda*, Book 12, Swadhyay Mandal, Surat.
Singh, K (1986) The Hindu Declaration on Nature, *The Assisi Declarations* (Messages on Man and Nature) made at Assisi, Italy, 29 September.
Srimad Bhagavata Mahapurana (1982) Translated by C L Goswami and M A Shastri, Gita Press, Gorakhpur.
Tilak, B G (1936) *Gita Rahasya*, Tilak Brothers, Poona.
Upanishad Anka of Kalyan (1949) Gita Press, Gorakhpur, 23(1).

FURTHER READING

Atharva Veda (1982) Translated by Devi Chand, Munsiram Manoharlal Publishers, New Delhi.
Charaka-Samhita (1983) Translated by Priyavrat Sharma, Varanasi, Chaukhambha Orientalia.
Mahabharata (1988) Translated by M N Dutta, Parimal, Delhi.
Mahabharata (1994) Gita Press, Gorakhpur, Gorakhpur, India in Sanskrit with Hindi translation.
Manusmriti (The Laws of Manu) (1975) Translated by G Buhler, Motilal Banarsidass, Delhi.
Rigveda (1974) Commentary by M D Saraswati, 12 volumes, Sarvadeshik Arya Pratinidhi Sabha, New Delhi.

Holon

see Phase Shifts or Flip-flops in Complex Systems (**Volume 5**)

Homocentric

Carolyn Merchant (1992, *Radical Ecology*) applies the term *homocentric* to an anthropocentric position that fosters care and stewardship for nature – for humanistic as distinct from ecocentric reasons. Since the environment is crucial to human well-being and survival, then we have an indirect duty towards it – one derived from communal human interests. This means assuring that the Earth remains environmentally hospitable to human life, and that its beauty and resources are preserved so human life may continue to be pleasant. As such, *homocentrism* is to be further distinguished from *egocentrism*, where the latter is an anthropocentrism based on prioritizing the desires of individuals rather than the communal good. Egocentrism underpins laissez-faire capitalism, with its inherent tendencies to exploit the environment, whereas homocentrism would, at the very least, regulate the market for the sake of wider social and environmental goals – or even replace capitalism by socialism. Both utilitarianism and Marxism espouse a homocentric ethic, hence eco-socialism is a homocentric variety of radical environmentalism, as is, according to Merchant, social ecology.

Homocentric ethics focus on society, therefore emphasizing the social good and social justice flowing from the (Christian) ethic of environmental stewardship. Unlike ecocentrism, which would protect nature for its intrinsic value, homocentrism's utilitarianism would allow managed transformation of nature where the resulting sum of human welfare was to be increased over that accruing by leaving it alone.

As Merchant describes it, homocentrism's ethics combine the mechanistic approach of egocentrism with the organicism of *ecocentrism*. Effectively this means that

homocentrism would apply science and technology derived from analytical and reductionist thinking in the cause of furthering what is seen as an intimate, organic and dialectical relationship between society and nature (*see* Ecocentric, Biocentric, Gaiacentric, Volume 5).

DAVID PEPPER *UK*

Human Body, Immediate Environment

Catriona Sandilands
York University, Toronto, Canada

The human body is a site of environmental change at a variety of levels. Perhaps most obviously, the impacts of many environmental problems are tangibly experienced by human beings as bodily symptoms; from sunburns to cancers, environmental degradation becomes part of everyday knowledge for many people through questions of bodily harm and danger. In this view, the body is a site for research concerning the consequences of particular environmental changes on human life; it is also a powerful trope for forms of political mobilization that link human health and environmental protection, including breast cancer and anti-nuclear activism.

Broadly speaking, the human body is the site of our primary physical experience and knowledge of environments; we interact with our environments through (at least) the bodily processes of touch, taste, smell, sight and hearing. These sensory sources of knowledge are influenced by our species physiology (e.g., human beings have differently organized senses than bees), but also by our specific cultures and activities (e.g., automobile travel, horseback riding and walking lead to very different views of a landscape). Viewed as sites where culture and physiology interact, bodies are important for understanding both ontological and epistemological dimensions of environmental change.

According to geographer Neil Smith, the body is a socially and physiologically constituted scale of place in which varied changes occur and interact, including biological maturation and decay, social identification and differentiation, and daily and intergenerational reproduction (Smith, 1991). In this sense, the human body is an environment, a site in and over which multiple conflicts and struggles take place: access, control, pleasure and pain, health, meaning, interpretation, mobility. The physical/social constitution of the body can thus be seen as a site of environmental change. Larger-scale environmental processes can also be understood as having specific manifestations in the environment of the body; the human body is understood not just as an effect of other scales of change but as a unique site where changes occur in particular ways, at least in part because of particular social and cultural relations of that scale. The prevalence and distribution of breast cancers in industrialized nations, for example, cannot be understood without reference both to the physiological impact of environmental hazards on women's bodies (e.g., the presence of organochlorines in water and consumer products) and to the complex social relations in and through which women's bodies have different histories (e.g., of such risk factors as exposure to toxins, access to health care and experiences of childbearing, variously affected by class, race, ability and sexual orientation). Altogether, these factors highlight the complexity of body/environment politics and the importance of an approach to environmental change that includes social and cultural elements; we are not, in fact, all in the same boat.

In a related vein, Elizabeth Grosz argues that bodies and environments are mutually defining (Grosz, 1995). Environments – including both their biological and cultural elements – orient and organize bodily senses and perceptual (spatial) information; they also leave traces on and in the body in the form of symptoms but also memories and desires. These impacts inspire particular forms of bodily action, in which the body becomes a site or agent of intervention into that environment. For example, a hot and humid day may inspire some people to use air conditioning and others to enjoy the heat; the more people that choose the former, the more energy use and air pollution, thus having an impact on the atmosphere, etc. Far from there being a single, natural reaction to an environmental change, bodies are encultured (e.g., one's preference for cooler or warmer temperatures is, at least in part, a question of memory and desire); they are also active participants in the environments that shape them.

Although this understanding of mutual influence between bodies and environments may seem commonsensical, it is important to note its implications. First, as Grosz (1995, 109) writes, if bodies are not ••• pregiven, ••• environments cannot alienate the very bodies they produce . In other words, there is no single ideal environment for the body, and the appropriateness of any given one has a great deal to do with social and cultural factors (this is not to say that there are no limits to possibility, only that there is certainly more than one). Second, if bodies and environments are involved in a dialectical relationship, and if there

is no single and biologically *a priori* template for the ideal form of this relationship, then it becomes of paramount importance for environmentalists to investigate a variety of different axes of in uence. Speci cally, as is suggested in much contemporary environmental thought, forms of social, political, and economic power have an enormous effect not only on bodies and environments individually, but on their modes of interaction.

In these respects, the ontological question of mutual body/environment in uence is also an epistemological question; bodies shape knowledge of environments, and *vice versa*. Although sensory experience is clearly part of the creation of this knowledge, it is also the case that more socially and culturally sedimented in uences are at work. Bodies and environments can, in the terms of social theory, thus be understood as sites of *inscription* (meaning that they are partly de ned by culture) and as elements of *discourse* (meaning that they circulate as signs within the realm of language, artifact and culture). Such understandings are signi cant because they point to the fact that representations of the human body as environment are, in fact, themselves an important part of the process of environmental change. For example, the metaphorical understanding of the earth as a human body, often (as in the case of some Gaia imagery) a female one, has led historically (and may continue to lead) to particular, often damaging assumptions about the nature of each.

REFERENCES

Grosz, E (1995) *Space, Time and Perversion: Essays on the Politics of Bodies*, Routledge, London.

Smith, N (1991) Homeless/Global: Scaling Places, in *Mapping the Futures: Local Cultures, Global Change*, eds J Bird, B Curtis, T Putnam, G Robertson, and L Tickner, Routledge, London.

Human Dimensions of Global Change

see Scientific Responses in an Era of Global Environmental Change (*Opening essay*, **Volume 4**); The Human Dimensions of Global Change (*Introductory essay*, **Volume 5**)

Human Ecology

see The Changing Human–Nature Relationships (HNR) in the Context of Global Environmental Change (*Opening essay*, **Volume 5**)

I PAT Equation

see Modeling Human Dimensions of Global Environmental Change (*Opening essay*, **Volume 5**)

Index of Sustainable Economic Welfare (ISEW) and Genuine Progress Indicator (GPI)

see ISEW (Index of Sustainable Economic Welfare) and GPI (Genuine Progress Indicator) (**Volume 5**)

Indigenous Knowledge, Peoples and Sustainable Practice

Douglas Nakashima[1] and Marie Roué[2]

[1] *UNESCO, Paris, France*

[2] *Centre National de la Recherche Scientifique, Paris, France*

Indigenous knowledge is entering into the mainstream of sustainable development and biodiversity conservation discourse. Article 8(j) of the Convention of Biological Diversity (Rio, 1992) has contributed to this process by requiring signatories to: " respect, preserve and maintain knowledge, innovations and practices of indigenous and local communities embodying traditional life-styles relevant for the conservation and sustainable use of biological diversity" . As the potential contribution of indigenous knowledge to key items on the global agenda gains widening recognition, an increasing number of scientists and policy-makers are calling for the integration of indigenous and science-based knowledge.

While indigenous peoples who have been lobbying for such recognition have reason to be satis ed, there are also reasons for concern. Are scientists serious enough about this emerging issue to go so far as to question the construction of their own knowledge? Or at the end of the day, will they do little more than add a veneer of traditional ecological knowledge (TEK) and then carry on business as usual? For the time being, the scienti c and the development communities views indigenous knowledge rst and foremost as a resource to be appropriated and exploited. Integration with (or more accurately into) science implies the application of a validation process based on scienti c criteria that purportedly separates the useful from the useless, objective from subjective, indigenous 'science' from indigenous 'beliefs'. Through this process, knowledge corresponding with the paradigm of Western science is extracted, and the rest is rejected. While this cognitive mining may be pro table to science, it threatens indigenous knowledge systems with dismemberment and dispossession.

INDIGENOUS, LOCAL OR TRADITIONAL ECOLOGICAL KNOWLEDGE (TEK)?

The challenge of understanding *indigenous knowledge* begins with the perplexing task of deciding how it should be named. On this matter, few persons agree. Or to be more precise, everyone recognizes that existing terms are for one reason or another unsatisfactory. Should one speak of TEK, the term coined when the field emerged in the public arena in the 1980s? Or abandon this designation in favor of the term indigenous knowledge? TEK has the advantage of immediately evoking the temporal depth of these sets of knowledge and of clearly flagging their domain, that of nature, while at the same time raising the question of their relationship to the ecological sciences. Associating the term traditional, however, with societies that, in the not so distant past, were dismissed as primitive or savage raises the phantom of social Darwinism. The term is all the more inappropriate, if our penchant for constructing binary oppositions leads us to conceive traditional

as a counterpoint to modern, and to think of traditional knowledge as fixed in the past and not susceptible to change. Finally, it may be misleading to apply the term ecological to sets of knowledge whose content and nature extend well beyond the confines of one scientific discipline to include such knowledge as the movements of constellations, the strength of ocean currents or the elasticity of sea ice, and to encompass not only empirical knowledge, but also practices and know-how, value systems, ways of life and worldviews. The fact is that indigenous peoples do not share the dichotomous occidental worldview that separates the material from the spiritual, nature from culture, and humankind from all other life forms, and as a result they do not do science in isolation from other pursuits.

While the label *indigenous knowledge* has the advantage of explicitly designating one of the groups of peoples most concerned – those who designate themselves as indigenous peoples, it raises other concerns. Who is indigenous and who is not? What peoples wish to be designated by this term and which do not? Indeed, this label is poorly adapted to the realities of Asia and Africa, where all peoples are native and any attempt to designate one group as indigenous but not another, provokes confusion. In these regions, unlike in the Americas or Australia, histories of human occupation have not followed a pattern whereby a wave of colonizers coming from abroad has supplanted or dominated a population that, due to its earlier and lengthy presence, is clearly identifiable as aboriginal or indigenous. Furthermore, in several African countries, particularly former colonies of francophone nations, the label indigenous has retained a strongly negative connotation due to its past history of use by colonial oppressors. For this reason, the expression *farmers knowledge* and *local knowledge* are sometimes preferred. The term local knowledge has the additional advantage of not excluding non-indigenous farmers, fishers, health practitioners and others whose extensive knowledge of the natural milieu is also a product of resource-based livelihoods extending across many generations. Its major weakness is a lack of specificity, as most knowledge can be labeled local.

Faced with the complexities of defining indigenous, special rapporteur J R Martinez Cobo provided a report in 1983 to the United Nations (UN) Sub-Commission on Prevention of Discrimination and Protection of Minorities, in which he stressed the right of indigenous peoples themselves to define what and who is indigenous. Further to this principle of self-identification, criteria retained to define the quality of being indigenous include common ancestry with original occupants, a distinct ethnic identity and shared patterns of vulnerability. These definitions are clearly of a political nature. They offer peoples encapsulated within states that may be reluctant to recognize territorial rights or the right to pursue a distinctive way of life, an opportunity to assert their difference and their collective identity. The debate on who and what to define as indigenous is far from closed, and it is unlikely that any single definition will satisfy all parties. Today, the UN Working Group of Indigenous Peoples estimates the world population of indigenous peoples as some 300 million, belonging to some 5000 groups in more than 70 countries, the largest population residing in Asia.

WHAT IS INDIGENOUS KNOWLEDGE?

Indigenous knowledge systems are the complex arrays of knowledge, know-how, practices and representations that guide human societies in their innumerable interactions with the natural milieu: agriculture and animal husbandry; hunting, fishing and gathering; struggles against disease and injury; naming and explaining natural phenomena; and strategies for coping with changing environments. It is through this fine-grained interplay between society and environment that indigenous knowledge systems have developed diverse structures and content; complexity, versatility and pragmatism; and distinctive patterns of interpretation anchored in specific worldviews. Whereas knowledge is conceived in Western culture as an abstract entity independent from practice (e.g., science as opposed to technology), such a compartmentalized view is alien to indigenous societies. It would be self-defeating to consider farmers' knowledge of rain patterns, soil types and crop varieties apart from the ways in which this information is put into practice in their fields. In other words, indigenous knowledge includes not only knowledge but also know-how. Transmission is not only oral, but also in the context of doing (Figure 1). Finally, unlike science, indigenous knowledge does not oppose the secular to the spiritual, and therefore does not separate the empirical and objective from the sacred and intuitive. In indigenous societies, such boundaries are permeable. On the one hand, much knowledge of nature falls within the empirical realm. Hunters have detailed knowledge of the habitat, behavior, diet and migration patterns of their prey. Farmers know how crops should be rotated to maintain soil fertility and which plant products have insecticidal or medicinal properties. This science of the concrete', however, blends imperceptibly into the metaphysical realm. For the hunter, the success of the hunt is as much due to assistance from spirit helpers, as it is to skillful tracking and steady shooting. The continued flow of water for the farmers' field is attributed as much to their respect for the deity of the sacred headwater forests, as to the water-drawing properties of the trees themselves. The concrete and the spiritual co-exist side by side, complementing and enriching rather then competing and contradicting.

Figure 1 Knowledge transmission through observation and practice: young Inuk boy helping his father skin and butcher a caribou (*Rangifer tarrandus*) on the banks of the Kuujjuaq River (Arctic Quebec, Canada). (Photo by D Nakashima)

ON THE NATURE OF THE ''SCIENCE OF THE CONCRETE''

While the current wave of interest in indigenous knowledge dates back no more than a few decades, the knowledge systems themselves have accompanied humankind through countless millennia of environmental change and cultural adaptation. The societies that developed them did not wait for official recognition before forging cultural landscapes and domesticating plants and animals that continue to play a major role in today's global economy. In 1962, Claude Levi-Strauss, a philosopher and renowned anthropologist, published his seminal work on *The Savage Mind*. In his opening chapter entitled the *The Science of the Concrete*, he reflects on the nature and character of indigenous knowledge, which he also qualifies as mythical thought (Levi-Strauss, 1962). More recently, Berkes (1999), a human ecologist, has dedicated a book to indigenous knowledge entitled *Sacred Ecology*. Between the early work of Levi-Strauss and Berkes' recent overview, numerous observers have pondered over these knowledge systems. While some accord them great value and others dismiss them as nonsense, their dual nature, combining the material and the spiritual, has been evident to all.

The first scientist to dedicate himself to the contemporary study of indigenous knowledge was Harold Conklin, who completed his thesis at Yale University in 1954 on *The Relations of Hanunoo Culture to the Plant World* (Conklin, 1954) (the Hanunoo are indigenous peoples of the Philippines). Pioneering a new discipline, ethnoscience, which would later come to be called the new ethnography, he dedicated himself to the study of a society's knowledge of its natural environment, through the examination of indigenous semantic categories. It is true that studies of human relations with plants and animals existed well before the 1950s, and that this intellectual tradition continues to this day under various labels composed of the prefix ethno-, followed by a qualifying term: ethnobotany, ethnozoology, ethnominerology. Ethnoscience, however, introduced a radically different approach that focused upon the knowledge of nature possessed by local peoples themselves. Unlike ethnobotany or colonial economic botany, which are basically applied Occidental sciences carried out in exotic locations and among exotic' peoples, ethnoscience strives to understand the view from within (an indigenous worldview). As a branch of cognitive anthropology, it focuses upon elucidating a cultural grammar, structures of thought built into the local language. For the first time, the recording of indigenous taxonomies was to reveal not only the breadth of a people's knowledge, but also insights into its very nature (cf. Friedeburg, 1979; Bulmer, 1979; Berlin, 1992; Ellen, 1993).

The classificatory prowess of indigenous peoples is on a par with that of scientists. As Levi-Strauss has amply demonstrated, the breadth of indigenous taxonomic knowledge is impressive: a single Seminole informant can identify 250 plant varieties and species, Hopi Indians recognize 350 plants, the Navajo more than 500. The Subanun of the Philippines use more than 1000 botanical terms and the Hanunoo, close to 2000. The diversity and sophistication of plant and animal use also attests to the subtlety of indigenous observations and the capacity to systematize knowledge and know-how. The Buriyat of Siberia, for example, derive from a single species, the bear, no less than 52 therapeutic applications: seven distinct uses for the

meat, five for the blood, nine for the fat, 12 for the brain, 17 for the bile, and two for the fur, etc.

As impressive as these figures may be, for the Hanunoo, plants are not the stuff of lists, but rather objects of passionate and prolonged debate. Having rigorously recorded Hanunoo discussions during and after the evening meal, Conklin concluded that plants, especially cultivated varieties, were the focus of a majority of exchanges. Combining the expertise and interests of systematic and economic botanists, Hanunoo conversations centered upon the hundreds of characteristics which differentiate plant types and often indicate significant features of medicinal or nutritional value (Conklin, 1954, 97). This knowledge is acquired very young and expanded throughout an entire lifetime. One morning, a seven-year-old Hanunoo girl surprised Conklin by suggesting that he show her the pictures of Brown's three-volume guide to useful plants of the Philippines. On her own initiative, she began to comment on the text, assigning each plant a hanunoo name, or solemnly declaring that she had not seen that plant before.' At the end of her remarkable performance, she had identified 51 out of 75 plants with only two errors!

While a science of the concrete' might be expected to have practical applications as its sole preoccupation, Levi-Strauss argues that indigenous knowledge is motivated by considerations other than utility. He cites Speck, for example, who has documented for the Indians of northeastern US and eastern Canada, an extensive herpetological taxonomy that is comprised of distinct terms not only for each genus, but also for species and certain varieties. Yet this knowledge is not practical, as reptiles and amphibians have no economic utility for these groups. Similarly, not only do the Pygmies distinguish between a phenomenal number of plants and animals, but they have also developed detailed knowledge of the habits and behavior of several species, including bats, that cannot be considered to be useful to them. These and other examples, drawn from numerous observers in diverse societies, have led Levi-Strauss to conclude that it is not, as one might expect, use that leads to knowledge. Of the endless trials conducted for the sake of knowing, only a few end up providing useful results:

> ••• animals and plants are not known because they are useful, they are found to be useful or interesting because they are first known .
>
> (Levi-Strauss, 1966, 9)

THE EPISTEMOLOGY OF SCIENCE AND OF INDIGENOUS KNOWLEDGE

The image of the Other held by the West shifts from one extreme to another. Since the time of the first encounter, indigenous peoples have always been a source of fascination. For many Occidentals they embody the ultimate otherness and, as such, they have been the object of both admiration and scorn. Are the savages' the embodiment of our unattainable dreams of freedom, courage, purity, simplicity and social equality? Or are they, as we imagine we were before becoming civilized' – cunning, cruel, ignorant, without faith or order, heeding only the law of the jungle? Until the development of cultural ecology in the 1960s (Lee and Devore, 1968), hunters and gatherers were portrayed as our poor backward cousins, struggling through great efforts and pains to maintain themselves just at the brink of survival. Subsequently, they became the distinguished representatives of what Sahlins (1972) has termed the original affluent society, rich not in material wealth, but luxuriating in time freed by the efficacy of their adaptation to the natural milieu and the adoption of a non-consumptive (Zen) lifestyle. Given these enduring disparities between our views of indigenous peoples, the deep ambivalence of our contemporary discourse on indigenous knowledge should come as no surprise. It is through pronouncements about the Other that Western society plays out its own doubts and fears. In this period of global change, have advances in science and technology brought well-being or alienation to our societies?

Whenever the issue of indigenous knowledge is discussed, sooner or later the recurrent question of validity is raised. Is it scientific? Or, being composed of odds and ends, does it combine the best with the worst, juxtaposing without distinction empirical observations and obscure superstitions?

As a first response, we must understand the persons behind the question. Science is a social construction of our own society, which being modern and Occidental is determined both historically and geographically. It presupposes a separation of the spiritual from the material, of religion from knowledge, and of culture from nature. Such a dissected worldview is not shared by indigenous cultures whose philosophy is better characterized as a cosmovision or as holistic. For this reason, qualifying the knowledge of indigenous people as science is both misrepresentation and reductionism. Nevertheless, this fallacious and futile approach of seeking ourselves in the Other has provided the basis for an entire critique of indigenous knowledge. It is formulated in most cases as negative definitions, i.e., what indigenous knowledge is not. For example, science defines itself as experimental (deductions from hypotheses are tested), systematic (results can be replicated) and universal (results are independent from context, as variables are isolated and controlled). In contrast, indigenous knowledge is often said to be practical (determined by immediate need and utility), local (only applicable in the setting in which it was developed) and contingent (context dependent).

Some even speculate whether this knowledge might lie somewhere between perception and conception, and better be considered as much skill, as concept (Sillitoe, 1998, 229). How do New Guinea highlanders assess the quality of their soil? Even if the results of their analysis do not differ from that of the agronomist, they are reprimanded for not being able to offer up on a platter the analytical discourse that the scientist requires. When asked to explain why or how, they reply: If one is a farmer, one just knows . But in this dialogue between the deaf, where are the barriers and to whose limits are they due? Scientists conclude that indigenous people lack the analytical capacity to raise their skills from the plane of the local and pragmatic to that of the systematized and universal (Sillitoe, 1998). But as Sillitoe's comments suggest, highlander farmers are undoubtedly equally frustrated by the difficulty of communicating their knowledge to scientists, and bewildered by the constraints that the latter impose on how knowledge must be packaged. Instead of opening all intellectual and sensorial pathways (touch, smell, taste, sight, sound) and accepting the farmer's invitation to feel the texture of the soil, smell its composition or taste its acidity, scientists narrow-mindedly insist that explanations pass by the single limited channel of an intellectual discourse compatible with the Occidental literate tradition. For the scientist, if a farmer's knowledge cannot be reduced to the written word, then it is not knowledge, but remains merely skills.

All of these criticisms (or more politely, definitions of what indigenous knowledge is not) arise from our own ethnocentrism and an inability to extricate ourselves from our own culturally-embedded point of view.

KNOWLEDGE AS A BASIS FOR SUSTAINABLE PRACTICE

Overshadowed by the pomp and promise of modern science and technology, indigenous knowledge systems have been disregarded until recently. While scientists and development agencies are only beginning to acknowledge their significance, their enduring role as the mainstay of local food production and health care in the developing world cannot be questioned. In sub-Saharan Africa, for example, local knowledge guides the decisions and practices of small-scale farmers who represent 70–90% of agricultural producers and more than 60% of the population. Artisanal fishers, who represent more than 90% of the fisheries work force world-wide, rely on their own knowledge and skills to locate fish, navigate safely at sea and bring home the catch. Similarly, it is estimated that some 80% of the world's population fulfills their primary health needs through the use of traditional medicines. Even in industrialized countries, local knowledge accumulated across generations continues to play a fundamental role in sustaining localized resource use practices whether they be small-scale farming, fishing, trapping or the gathering of wild produce.

Since the 1992 UN Conference on Environment and Development, and in particular the coming into force of the Convention on Biological Diversity, increasing attention has been drawn to the contributions that indigenous knowledge can make to global biodiversity conservation objectives. This emerging role for indigenous knowledge holders has flowed quite naturally from the recognition that most remaining regions of the world that are biodiversity rich are also homelands for traditional and indigenous peoples. This is no simple coincidence, as in sharp contrast with the biodiversity impoverishment that accompanies the processes of modernization and industrialization, the persistence of traditional ways of life has gone hand in hand with the maintenance of local ecological systems and the conservation, and even enhancement, of biodiversity. In these areas represented as wilderness or virgin in the Western mind, indigenous knowledge participates in the transformation of nature to create cultural landscapes, the products of social system–ecosystem interaction. One dramatic, but by no means unique, example of this process is the use of fire by Aboriginal peoples in Australia to create landscapes that are ecological mosaics (cf. Case Study 1 below). This sustainable practice, which is vital for maintaining biological diversity, has now been integrated into National Park policy in certain parts of Australia through directives that explicitly call for the reinstatement of traditional Aboriginal burning regimes.

Inventories of local biodiversity can also benefit from knowledge encoded in local languages in the form of indigenous categories of natural objects. Finally, as indigenous peoples retain within their knowledge systems an inter-generational memory of fluctuations, trends and exceptional events in relation to the local environment, they can contribute importantly to understanding processes of change, whether these might be long-term, global transformation processes or circumscribed local events. The invaluable contribution of indigenous knowledge to environmental and social impact assessment processes, for example, has been convincingly demonstrated.

Modernization and uniformization of agricultural practices has also triggered a severe loss of biological diversity of domestic plant and animal stocks. Here again, pastoral and peasant communities that have maintained traditional modes of production have today become the major custodians of the world's crop and domestic animal diversity. In tropical agroecosystems in Thailand and Indonesia, for example, peasants commonly maintain more than 100 domestic plant species, as well as harboring in their paddies, rice varieties adapted to a range of environmental conditions (Altieri, 1999). Slash and burn agriculture as practiced by the Karen of northern Thailand offers one example of a sustainable farming system based upon indigenous knowledge

and practice, that enhances biological diversity (cf. Case Study 2 herein).

CASE STUDY 1: ABORIGINAL FIRE REGIMES AND THE CREATION OF CULTURAL LANDSCAPES

Gagadju, indigenous hunter gathers of Australia s Northern Territory, focus their subsistence activities on a coastal fringe penetrating no more than a few kilometers inland (Lewis, 1989; Nakashima, 1998). Two habitat types dominate this area of atlands: tall-open eucalyptus forest and eucalypt woodlands. Interspersed between these habitats are freshwater oodplains, paperbark swamps and rain forest stands. A steep sandstone escarpment dominated by Spinifex and other grasses as well as by shrubs, marks the inland boundary of the coastal atlands. This tropical savannah ecosystem has been and continues to be shaped by the Gagadju through the judicious application of re.

Burning practices are carefully orchestrated, paying close attention to season, habitat, wind direction, state of vegetative growth, moisture levels, previous burn locations and accumulation of debris. The Aboriginal calendar is divided into six seasons. Three of these constitute the wet period of the year: rst, a season of pre-monsoon storms; followed by the monsoon season; and nally the period of knock em down storms. The dry period follows, also divided into three seasons, beginning with a cool, humid season, then a cold weather season and nally a period of hot, dry weather. The Gagadju begin their annual cycle of burning towards the end of the wet period and conduct most burning activities during the ensuing cool and cold weather seasons of the dry period. During the hot, dry season when vegetation is extremely ammable and res are dif cult to control, burns are only conducted under very limited circumstances.

Burning Patterns on the Floodplains

With the end of the monsoon season, water levels on the oodplain begin to recede. Sedges and grasses, which spring up on the exposed ats, quickly mature and are dry before the coming of the knock em down storms. The Gagadju set their rst res in these dry grasses, knowing that the burn will die out along the edges of the rain forests and paperbark swamps (much too moist at this season to burn) and will continue outwards into the oodplains no more than some 2 10 m. These habitat types differ markedly in their sensitivity to re. When dry, the grasses of the oodplains, as those on the escarpment, burn easily and recover quickly. In contrast, the vegetation of rain forests and paperbark swamps is re-sensitive, recovering from a burn only with considerable dif culty. At intervals of a few weeks, the Gagadju ignite and burn successive strips of grasses as they mature and dry, proceeding in this fashion farther and farther out onto the oodplain. On each occasion, previous burn areas and fresh vegetative growths on moister soils serve to contain the re.

This burning of the oodplains serves several purposes. First, the early res around the perimeters of rain forests and paperbark swamps create re barriers that shelter these fragile habitats from res occurring later in the season. Second, the razing of mature, dry stands of grasses triggers a regrowth, whose tender leaves lure kangaroos and wallabies out from adjacent rain forest stands in the early morning and late evening, thus offering the Gagadju hunting opportunities. Finally, the Gagadju note that res also favorably alter the oodplain habitat for the nesting of waterfowl, particularly magpie geese.

Creating Fire Mosaics

Beginning during the cool season and at the start of the cold season, the Gagadju set res in the tall-open forest and eucalypt woodland habitats that dominate the coastal atlands. These habitats are highly susceptible to burning and at the same time, well adapted to re. In the understory, leaf litter accumulates quickly and has high oil content, while sorghum grasses provide a dense cover. While these burn easily when dry, the vegetation is nevertheless quick to recover. Thick bark shelters eucalyptus trees from understory res, and high canopies reduce the likelihood of res spreading into the forest crown. Wind intensity and direction are important allies. The Gagadju use strong midday winds to aim the re towards areas of moister vegetation or previous burn sites to limit the extent of the burn. Furthermore, strong winds prevent heat and ames from licking up into the forest canopy, thus protecting trees in ower that will subsequently provide humans and animals with edible fruit. Early in the cold season the res go out after a few hours and burn relatively small areas. As the season progresses, the vegetation becomes successively drier, and res burn longer and cover larger areas.

Fires in these wooded habitats serve a number of functions. Circles of re are used as a hunting technique, driving animal prey to waiting hunters. By removing surface vegetation and debris, burning facilitates the gathering of yams and roots and the hunting of goannas (Figure 2). Burn sites attract kangaroos and wallabies that come to graze the fresh re-growth of grasses. Furthermore, these serial burnings create an ecological mosaic composed of sites at different stages of recovery from burning. This heightens biological diversity and provides the Gagadju with a choice of subsistence opportunities. Finally, regular bouts of controlled burning impede the growth of thick tangles of underbrush and prevent the dangerous build-up of leaf litter and debris that can result in uncontrollable wild res.

The central role of traditional Aboriginal burning practices in maintaining biological diversity, and in creating

Figure 2 Shaping landscapes with fire: Aboriginal peoples use controlled burning to maintain a habitat mosaic that offers a variety of foraging opportunities. Here a woman sets off to hunt goannas on a newly burned site (Northern Territory, Australia). (Photo by B Glowczewski)

and maintaining the cultural landscapes of Australia, has gained official recognition by National Park authorities in certain parts of the country. In Kakadu and Uluru National Parks the re-establishment of traditional Aboriginal burning regimes has become an explicit Park management goal.

CASE STUDY 2: SWIDDEN CULTIVATION AS SUSTAINABLE PRACTICE – THE KAREN OF NORTHERN THAILAND

Slash-and-burn (*swidden*) agriculture continues to be an important mode of agricultural production in many parts of Africa, Asia, South and Central America and the Pacific (*see* **Shifting Cultivation and Land Degradation**, Volume 3; **Swidden**, Volume 3). It has also been at the heart of one of the most prolonged and vociferous debates over the sustainability of this traditional practice. While the debate is far from closed, there is growing awareness today that swidden agriculture as practiced by many indigenous peoples is not only sustainable, but also contributes to other pressing global concerns including conservation of domestic and wild biodiversity, as well as carbon storage.

Land Use Strategies of the Karen

The Karen land use strategy is based upon the maintenance of several major categories of land that can be distinguished by their use, location and the pattern of ownership: rice paddy fields, areas of swidden cultivation, community forest and watershed forest (Figure 3) (Rerkasem and Rerkasem, 1994; Nakashima, 1998). Rice paddies are owned by individual families and tend to be established close to the village, often at the base of slopes where sediment fans have been created by the erosion of soils from swidden fields above. Lands used for rotational swidden agriculture are much more extensive, lying in a broad band surrounding the village. Families divide their holdings into plots among which cultivation is rotated on a 7–10 year basis. Fields are cultivated for only one or two seasons before being returned to fallow. The different types of forested areas, unlike rice paddies and swidden fields, are held as common property. The extensive community forests provide the Karen with wood and other forest products, as well as serving to graze livestock. Watershed forests harbor the headwaters of major water sources and are denoted as sacred.

Creating and Maintaining a Biologically Diverse Landscape

Through a judicious mixture of tradition, adaptation and innovation, the Karen people shape the landscape to create and maintain a high level of heterogeneity. From the resultant rich biological diversity, they extract a multitude of materials and products that fulfill nutritional, socioeconomic, aesthetic and spiritual needs. Within swidden fields under cultivation, major crops of upland rice, maize and, increasingly today, cash crops such as cabbage, are intercropped with an astounding variety of traditional swidden products such as beans, eggplants, cucurbits, sesame, chilies, yams, gourds, pumpkins, various greens, and spices. Fruit trees, cotton and medicinal herbs are also planted in these fields (Rerkasem and Rerkasem, 1994).

Further to this diversity of cultivated species, Karen farmers also foster the rapid recovery of secondary growth in the fallows, not only to maintain soil fertility and limit

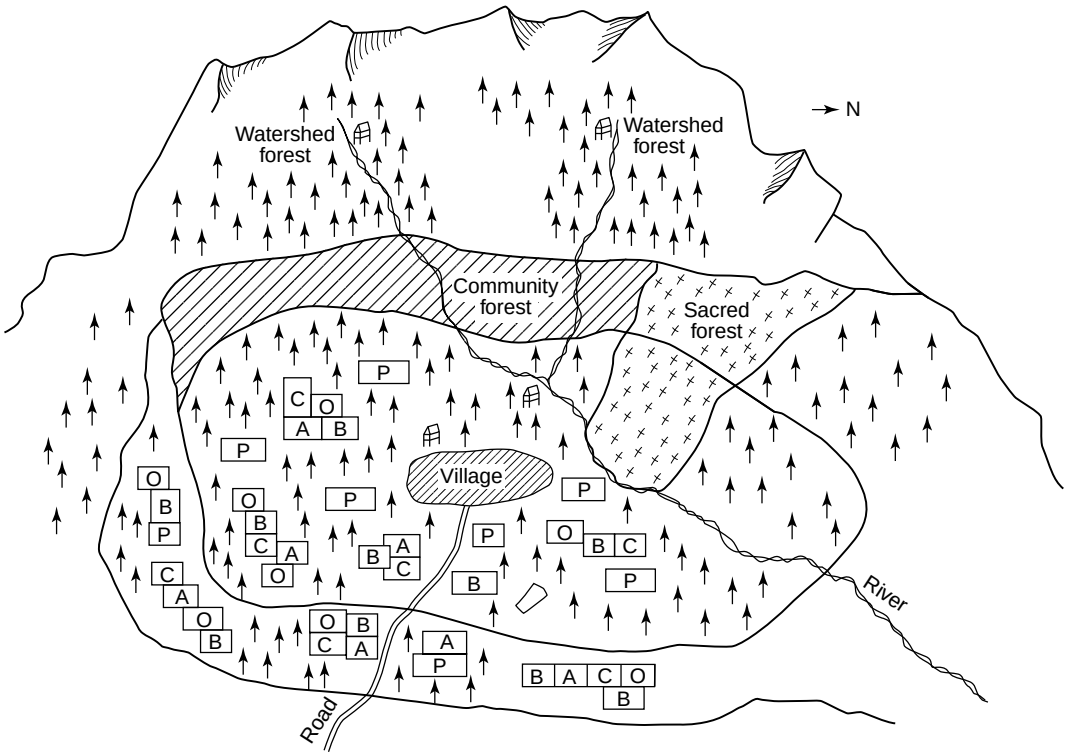


Figure 3 Organization of land use by the Karen of Nong Taw (Chiang Mai Province, Thailand) showing paddy fields (P), swidden fields (O – under cultivation; A – first year fallow; B – 3–4 year fallow; C – 5–7 year fallow), community forest, watershed forest and sacred forest for rituals associated with birth. (Map by J Odochao and C Vaddhanaphuti. Reproduced by permission of Chayan Vaddhanaphuti)

erosion, but also to re-establish a botanical community of which they make full use. When clearing a field for cultivation, branches and brush are left on the ground and burned, providing the soil with nitrogen and carbon-rich ash. Trees are cut well above ground level, leaving roots intact and stumps rising some tens of centimeters above the ground (Figure 4). These stumps sprout quickly, speeding recovery of the forest after cultivation and thus maintaining soil fertility.

Furthermore, plots are distributed so as to ensure that recently used fields juxtapose old fallow plots whose advanced secondary growth speeds the recovery of cultivated lands by providing an abundant supply of seeds. From the first year onwards, the fallow fields provide important supplies of medicinal plants, food in the form of mushrooms, tubers and shoots, as well as areas for the grazing of livestock. After several years, the number of plant species re-established on these sites numbers in the 100s, approaching the biological diversity found in adjacent uncut forest areas, such as the community and watershed forests. The Karen agro-ecosystem also encompasses the use and management of permanently forested lands. Community forests provide wild food products such as mangoes, jackfruit, bamboo shoots, mushrooms, greens and edible herbs, as well as medicinal herbs, firewood, and timber and bamboo for construction. Watershed forests, as vital sources of water for swidden and rice paddy fields, are maintained in a state close to forest climax and strictly protected.

Today the loss of traditional lands to logging interests or commercial plantations and state policies encouraging permanent and intensive agriculture are posing major threats to many swidden cultivators. Less land means shortened fallows and depleted soils, jeopardizing the sustainability of traditional practices. Moving from shifting cultivation to permanent agriculture entails major losses in both biological and cultural diversity, and raises the specter of the millions of hectares of Southeast Asia covered by species-poor *Imperata cylindrica* grasslands, a direct result of the failure of permanent agriculture on phosphorus-poor tropical soils. Rather than import traditional Occidental models of resource exploitation, such as permanent cash monocropping or large-scale non-selective logging, the sustainable practices of small-scale swidden farmers like the Karen should be recognized and supported. They offer numerous benefits including long-term sustainability of local environments and lifestyles, as well as the conservation of domestic crop varieties for many species, and wild plant and animal communities whose biodiversity approaches that of unexploited forest areas and remains far superior to that of permanently cultivated fields. Finally, recent

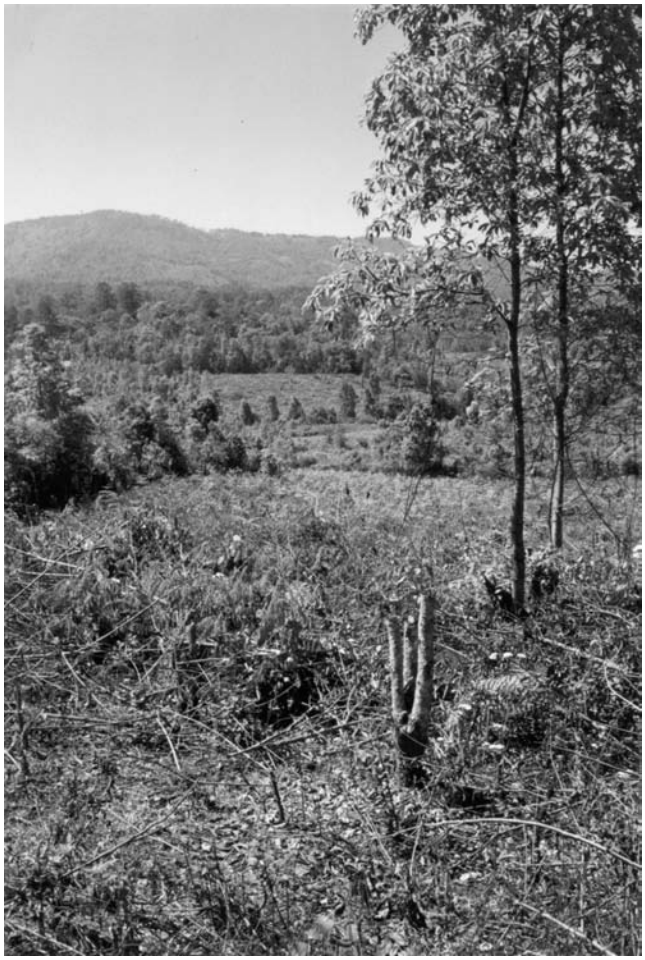


Figure 4 Karen swidden fields showing various stages of regeneration. When clearing fields for cultivation, swidden farmers encourage re-growth by cutting certain trees well above ground level and leaving others intact (Chiang Mai Province, Thailand). (Photo by D Nakashima)

research suggests that agroforests in general, and shifting cultivation in particular, can contribute significantly to carbon sequestration, by replacing conventional monocropping or *Imperata* grasslands with cyclical systems that include forest production.

INTELLECTUAL PROPERTY RIGHTS AND INDIGENOUS KNOWLEDGE: THE THREAT OF COMMODIFICATION

While biodiversity managers or development practitioners are only beginning to define a role for indigenous knowledge, its utility in another domain has been fully recognized and is being vigorously acted upon. Bio-prospecting, the quest for natural products to be exploited for commercial gain, is explicitly targeting traditional knowledge holders. In this arena, the pharmaceutical, agricultural and biotechnological industries have been particularly active. They recognize that the accumulated knowledge and traditional practices of indigenous communities are a powerful resource that can greatly facilitate the task of identifying useful new varieties of domestic plants or animals, isolating novel biological components, or developing innovative technologies and techniques. Recognition, however, has not automatically led to acknowledgement and the need to share benefits with the community is regularly overlooked.

The patenting of domestic plant varieties, traditional medical products and other biological resources whose identification and use are embedded in traditional knowledge has been a source of grave concern for developing countries and indigenous communities. By conferring upon foreign companies or individuals the exclusive right to exploit for commercial benefit certain inventions' that are in fact based upon indigenous knowledge, the patenting process transforms persons who do what they have always done into patent infringers. In one infamous case, the wound-healing capacity of turmeric, known and used in India for centuries, became a patented invention in the US. As a result, it became illegal for Indians residing in the US to use turmeric for this purpose. Similarly, a European patent on the fungicidal properties of the neem plant privatizes the botanical knowledge of Indian farmers who, for generations, have used this natural pesticide in their fields. In the end, both of these patents were revoked, but only after long and costly challenges mounted by the government of India (Dutfield, 1999). Similarly, the Bolivian National Association of Quinoa Producers has mounted a challenge against a US patent relating to a unique male sterile variety of the quinoa plant, a characteristic long known to Andean peasants and used by them to control the development of new hybrids. As Gari (1999, 6) astutely points out this patent does much more than just hijack indigenous knowledge for the eventual material benefit of a developed country. It also strengthens and spreads the Western paradigm of nature and science , by presenting knowledge pillaged from Bolivian peasants as the cutting edge of biotechnology.

Patents are only one of several legal instruments that constitute the current regime of intellectual property rights (IPRs). Others include trademark, copyright, geographical indication and plant variety rights. For indigenous knowledge holders, IPRs have been more often a source of problems, rather than solutions. Certainly they serve to protect the interests of companies that engage in bio-piracy, the unauthorized exploitation of biological resources and indigenous knowledge. A key debate today is whether contemporary IPR regimes can be adapted to also defend the interests of those who are custodians of indigenous knowledge.

To counter the granting of patents prejudicial to indigenous peoples, actions have been taken before the World Trade Organization under the agreement on trade-related aspects of intellectual property rights (TRIPS).

While providing a forum for overturning inappropriate patents, nothing in the TRIPS Agreement has been expressly designed for the protection of indigenous knowledge. Nevertheless, existing IPR arrangements may offer some viable options, and efforts are being made to explore these possibilities, including the development of sui generis (of its own kind) systems. The World Intellectual Property Rights Organization, for example, has recently created a commission on genetic resources, traditional knowledge and folklore, whose mandate includes the investigation of innovative measures to accommodate the exceptional characteristics of indigenous knowledge systems.

The challenge is a sizable one. IPRs have evolved within narrow socio-economic and political contexts. Designed to protect individuals whose inventions require safeguarding in view of their potential commercial value, they remain largely incompatible with indigenous knowledge, which is collectively owned; whose invention extends across generations; and whose raison d etre is not pro t, but ecological understanding and social meaning. Due to this incompatible nature, legal rights may have impacts quite other than those intended. By protecting select elements in isolation from the larger cultural context, they encourage the fragmentation of cultural systems. By designating knowledge owners, they may trigger social dissension between those recognized as proprietors and others who are excluded. Finally, as IPRs protect knowledge by setting rules for their commercial exploitation, ironically they may in the end merely facilitate the appropriation of traditional knowledge by the global marketplace.

RELATIONS OF POWER: KNOWLEDGE APPROPRIATION OR INDIGENOUS EMPOWERMENT?

The scienti c and development community views indigenous knowledge rst and foremost as a resource to be appropriated and exploited. Even scientists with the best of intentions may accelerate the demise of other knowledge systems by valorizing components resembling science and implicitly (or explicitly) casting dispersions on other elements that in science s view are mere superstition and belief. Such a process captures and instrumentalizes indigenous knowledge, strengthening the hand of those in positions of power at the expense of indigenous peoples.

This is not to argue against the exchange and sharing of knowledge between scienti c and indigenous systems. However, as Agrawal (1999) underlines, it would be irresponsible and a disservice to indigenous peoples to ignore the power relations that de ne such processes, and in particular, to fail to appreciate the implications of the severe power asymmetry between indigenous peoples and proponents of science.

To seek a way forward, it is important to come back to the questions of culture and worldview. For scientists, culture is a foreign element whose consideration falls outside the bounds of their profession (though they would probably agree to associate the cultural factor with the indigenous component of the equation). They are quite reluctant to admit that the culture of science is, in itself, a valid object of study. Cultural constructions such as the opposition of Nature (environment) and Culture (society), and the differentiation of rationality from spirituality, the empirical (science) versus the symbolic (religion), have provided science with its very foundations, and remain today an everyday reality of scienti c thought and practice. These tenets are such an integral part of the scienti c worldview that natural scientists are not aware of them as speci c cultural interpretations of the world. For them, they simply represent reality. Scienti c reality, however, differs from that lived by indigenous knowledge holders, whose conceptions of the world include pathways between natural and societal realms and whose spirituality infuses everyday objects and everyday acts.

In other words, there is no sound basis for deciding that one worldview offers a superior reference point for reality than another. We can of course arbitrarily choose and given science s institutional power, it is not surprising that the objective and rational scienti c method is repeatedly called upon to judge other knowledge systems. But it is important to recognize that this is a societal choice, not one defensible from any neutral or extra-cultural perspective. Consequently, the encounter between scienti c and indigenous knowledge must be understood as a meeting of cultures, with the cultural component as prominent in one camp as the other.

Full appreciation of this perspective alters our approach to articulating (a better term than integrating) scienti c and indigenous knowledge. Emphasis must be placed on leveling the playing eld and appreciating indigenous knowledge not as static sets of information to be conserved *ex situ* and integrated into science, but as dynamic components of contemporary indigenous societies and cultures. Accordingly, the protection of indigenous knowledge may better pass through pathways such as conserving indigenous language (as knowledge is encoded in language), ensuring knowledge transmission within the societies themselves, empowering indigenous societies so as to increase their control over processes of change, and ensuring continued access to the environments upon which their ways of life depend (*see* **Development and Global Environmental Change**, Volume 5).

REFERENCES

Agrawal, A (1999) Ethnoscience, TEK and Conservation: on Power and Indigenous Knowledge, in *Cultural and Spiritual*

Values of Biodiversity, ed D A Posey, UNEP, Nairobi, Kenya, 177–180.

Altieri, M A (1999) The Agroecological Dimensions of Biodiversity in Traditional Farming Systems, in *Cultural and Spiritual Values of Biodiversity*, ed D A Posey, UNEP, Nairobi, Kenya, 291–297.

Berkes, F (1999) *Sacred Ecology: Traditional Ecological Knowledge and Resource Management*, Taylor and Francis, Philadelphia, PA.

Berlin, B (1992) *Ethnobiological Classi cation: Principles of Categorization of Plants and Animals in Traditional Societies*, Princeton University Press, Princeton, NJ.

Brush, S B and Stabinsky, D, eds (1996) *Valuing Local Knowledge: Indigenous People and Intellectual Property Rights*, Island Press, Washington, DC.

Bulmer, R N H (1979) Mundane and Mystical in Kalam Classification of Birds, in *Classi cations in their Social Context*, eds R Ellen and D Reason, Academic Press, London, 57–80.

Conklin, H C (1954) *The Relation of Hanunoo Culture to The Plant World*, Yale University, University Microfilms Inc., Ann Arbor, MI.

Dutfield, G (1999) Rights, Resources and Responses, in *Cultural and Spiritual Values of Biodiversity*, ed D A Posey, UNEP, Nairobi, Kenya, 503–515.

Ellen, R (1993) *The Cultural Relations of Classi cation: an Analysis of Nuaulu Animal Categories from Central Seram*, Cambridge University Press, Cambridge.

Feit, H (1973) Ethno-ecology of the Waswanipi Cree; or How Hunters Can Manage Their Resources, in *Cultural Ecology*, ed B Cox, McClelland and Stewart, Toronto, 115–125.

Friedberg, C (1979) Socially Significant Plant Species and Their Taxonomic Position Among the Bunaq of Central Timor, in *Classi cations in their Social Context*, eds R Ellen and D Reason, Academic Press, London, 81–101.

Gari, J A (1999) Biodiversity Conservation and Use: Local and Global Considerations, *Science, Technology and Development Discussion Paper* No. 7, Center for International Development and Belfer Center for Science and International Affairs. Harvard University, Cambridge, MA.

Johannes, R E (1978) Traditional Marine Conservation Methods in Oceania and Their Demise, *Ann. Rev. Ecol. Syst.*, 9, 349–364.

Lee, R B and De Vore, I, eds (1968) *Man the Hunter*, Aldine, Chicago, IL.

Levi-Strauss, C (1962) *La pensée sauvage*. Librairie Plon, Paris [English translation (1966) *The Savage Mind*, University of Chicago Press, Chicago, IL].

Lewis, H T (1989) Ecological and Technical Knowledge of Fire: Aborigines Versus Park Rangers in Northern Australia, *Am. Anthropol.*, 91, 940–61.

Nakashima, D (1998) Conceptualizing Nature: the Cultural Context of Resource Management, *Nat. Resour.*, 34(2), 8–22.

Rerkasem, K and Rerkasem, B (1994) *Shifting Cultivation in Thailand: Its Current Situation and Dynamics in the Context of Highland Development*, Forestry and Land Use Series No. 4, International Institute for Environment and Development (IIED), London.

Roue, M and Nakashima, D (1999) The Discourse of Ecological Correctness: of Dam Builders Rescuing Biodiversity for the Cree, in *Cultural and Spiritual Values of Biodiversity*, ed D A Posey, UNEP, Nairobi, Kenya, 406–412.

Ruddle, K and Johannes, R E, eds (1985) *The Traditional Knowledge and Management of Coastal Systems in Asia and the Paci c*, UNESCO, Jakarta, Indonesia.

Sahlins, M (1972) *Stone Age Economics*, Aldine, Chicago, IL.

Sillitoe, P (1998) The Development of Indigenous Knowledge: a New Applied Anthropology, *Curr. Anthropol.*, 39, 223–252.

Indigenous Peoples

see Globalization in Historical Perspective (*Opening essay*, **Volume 5**)

Interdisciplinary vs Multidisciplinary

see Social Science and Global Environmental Change (*Opening essay*, **Volume 5**)

Internalization of Damage Costs

see Internalization (**Volume 4**)

International Council of Scientific Unions (ICSU)

see ICSU (International Council for Science) (**Volume 4**)

International Environmental Law

M J Peterson
University of Massachusetts, Amherst, MA, USA

International environmental law is the branch of international law addressing human interactions with nature,

with particular focus on protecting the biosphere, including its subsystems, natural processes, and species diversity from serious harm caused by human activities and regulating the human use of natural resources. Its substantive and procedural principles, norms and rules are expressed in a growing number of global, regional, and bilateral treaties, declarations and action programs issued by international conferences and intergovernmental organizations, and emerging rules of customary international law. These specify goals towards which governments should strive, environmental standards they should incorporate into their cooperative and individual efforts to protect the environment, and certain procedures they should follow in cooperative and individual efforts to protect the environment from signi cant harm.

Because international environmental law only addresses states (acting through their governments), it forms only one part of the legal structure designed to promote greater environmental protection. The activities of individuals, households, groups, and private rms remain subject to the national legal system of the state where they act or of which they are a national. Thus, effective protection of the environment requires that international and national environmental law be aligned. This process has been promoted by the extensive borrowings of principles and norms between both levels, international programs that spread regulatory expertise and increase the administrative capacity of developing states, and the efforts of a highly active transnational environmental movement.

One institutional difference between national and international law makes developing, revising, and enforcing international law more complex: the international system lacks the central legislative, executive, and judicial bodies that maintain national legal systems. Development and implementation of international rules depend on a more complex process of bargaining and mutual in uencing between states equal in legal status as sovereigns but unequal in terms of size, wealth, and level of industrialization. This process is increasingly affected by popular pressures within states and the transnational activities of interest groups and social movements.

Both the general international law applying to all states and particular treaties negotiated between two or more states have long re ected concern about aspects of nature. Perceptions of the abundance of sh stocks in uenced positions in the 17th century debates between proponents and opponents of the principle of freedom of the seas. Sharing navigational and other uses of rivers and lakes straddling the territory of more than one state promoted development of a distinct law on international watercourses. Disease control, protection of plants, migratory birds, and animal species, distribution of shing rights and prevention of over shing on the high seas became the subjects of multilateral and bilateral treaties in the late 19th and early 20th centuries. Pollution at sea, particularly from oil tanker operations and waste dumping, transport of hazardous substances, radiological hazards from military and civilian uses of nuclear energy, and transboundary water and air pollution umes were added to the subjects of treaty making after World War II (Ruster et al., *1983; UNEP, 1990).*

Concerns about protecting nature or natural resources were gathered under the common heading of international environmental law, only after ecological worldviews highlighting the interconnection of natural systems and the deleterious consequences of their alteration by human conduct were diffused widely among governments during the preparations for and discussions at the 1972 United Nations (UN) Conference on the Human Environment (Stockholm Conference) (Schachter, 1991; Caldwell, 1998). The 1992 UN Conference on Environment and Development (Rio Conference), with its emphasis on the strong interconnections among environmental protection, economic development, and participatory governance at both national and international levels ensured that the integration of international environmental law with other branches of international law would be taken as seriously as its further development.

The process of developing international environmental law has been shaped by several particular characteristics of environmental issues. It also involves development or wider use of several procedural innovations in negotiating, revising, and implementing international agreements meant to speed those processes and adapt them more fully to contemporary needs. So far it has yielded considerable, though not complete, consensus on a set of overarching principles and policy norms to guide the handling of speci c environmental problems as well as an ever increasing set of global, regional, and bilateral agreements regarding a wide range of environmental problems.

THE SPECIAL CHARACTERISTICS OF ENVIRONMENTAL ISSUES

International environmental law is distinct from other branches of international law not because it is relatively new, but because international environmental issues have several distinct features that give international law-making and law application in that field a distinct character.

- Relatively large and persistent uncertainty about impacts of human action or inaction on the biosphere is the most important of these characteristics. This uncertainty is a compound of changing scientific understanding of natural biosystems, changing natural system responses to human-induced stresses, and changing degrees of human-induced stress as technology alters, global human population increases, and consumption patterns shift. Uncertainty is not unique to

environmental issues, but it has more sources and appears to operate within wider bounds than for other issues.

- Heavy reliance on expert advice, not just from scientists but also from engineers, resource managers, and economists, marks the formulation and implementation of regulatory decisions because of the breadth and pervasiveness of uncertainty. Many environmental problems are very subtle in their early stages, and their identification depends on the enhancement of human perceptual abilities provided by scientific instruments. Assessing alternate solutions generally involves fairly complex assessments of causal relations among multiple factors as well as comparative cost-benefit analyses.
- Lawmakers must consider impacts over a relatively long time horizon because of the timelags between introductions of new technology or other changes in human activity and alteration of natural systems in response.
- Ensuring national compliance with environmental commitments imposes heavy administrative burdens on governments since most of the activity creating the need for explicit environmental protection is undertaken by individuals, households, business firms, and other private entities rather than by government agencies or officials. In many respects, international environmental law is a set of mutually agreed mandates directing governments to develop and use their administrative capacities in particular ways.
- The activities regulated are characterized by widespread externalities: both the bad effects of neglecting the environment and the good effects of improving environmental protection extend beyond the property of the private entity involved and often beyond the territory of the state where the activity occurs. This means that the full costs or benefits are not borne or enjoyed by the private entity or even the country creating harms or providing benefits to others, with the result that law must often include incentive schemes or other mechanisms for inducing actors to take the full range of costs and benefits into account.

At the same time, international environmental law shares two significant characteristics with other branches of international law.

- Necessity for integration or coordination of rules with other branches of international law. Integration will be particularly strong with those other branches of international law regulating use of physical spaces (law of the sea, law of international watercourses) or the exploitation of resources (much of international economic law). The less extensive overlaps with human rights law and the law of warfare are also sufficient to pose coordination problems, requiring similar processes of balancing the values promoted by the different branches of international law and ensuring that their rules are compatible with each other.
- Entanglement in wider debates about the future structuring of international, national, and local governance. International law will be affected by the results of current debate concerning whether the world will remain divided into some 200 states claiming sovereignty over their individual territories and populations. On the one side are governments and publics resistant to giving up sovereignty, whether by making promises to other states or by transferring significant governing authority to an intergovernmental organization like the UN Environmental Programme or a supranational organization like the European Union. These fears are particularly strong in developing states, where any significant centralization of authority is seen as an effort by the industrial states to maintain dominance of the global economy. The fears have been shared by segments of public opinion on both populist or nationalist right and socialist left in many of the industrial states, who see the process as one of yielding to distant and unaccountable elites. On the other side are two strong currents of opinion within the environmentalist movement regarding national sovereignty as an anachronism: one because it stands in the way of the greater global centralization needed to ensure a sustainable world, and the other because it stands in the way of shifting governance to a smaller scale promoting more sustainable lifestyles through application of local knowledge and face-to-face mutual accountability. Similar debates have raged in the fields of international human rights and international trade. All three have intensified as critics have demanded that the World Trade Organization pay more attention to environmental and labor issues. These critics see the interrelated debates as amounting to a choice between globalization on terms decided by a multinational corporation-dominated elite, and participatory political institutions seeking creation of a just world by regulating markets.

PROCEDURAL INNOVATIONS

The widespread and growing sense of urgency in addressing environmental problems has encouraged government officials, legal advisers, and diplomats involved in developing international environmental law to adopt a number of procedural innovations. These are aimed at speeding up the traditionally slow processes of negotiating multilateral treaties

or developing new rules of customary international law, accommodating the need for a wide range of information in decision-making, and providing greater mutual pressure for implementing agreed policy measures than has traditionally existed at the international level (see, e.g., Brown-Weiss *et al.*, 1998; Kiss and Shelton, 1999; Victor *et al.*, 1998 for greater detail.)

Framework Treaty Protocol Form

The older practice of distinguishing between the main portion of a treaty and annexes, protocols, or schedules containing detailed rules that can be modified as needed through an expedited amendment process, has been developed more fully in the framework treaty-protocol form used in many of the major global environmental treaties negotiated since 1985 (Kiss, 1993). The framework treaty identifies the environmental problem to be addressed and specifies procedures for the meetings of the parties that determine the regulatory measures to be adopted, monitor their implementation by participating national governments, and assess the state of the natural systems involved and the efficacy of the regulatory measures with an eye to their revision. The protocols stipulate the particular regulatory measures selected and may also elaborate on the procedures for joint decision, implementation review, and results assessment. Most protocols are developed later as governments gain better understanding of the problem, develop clearer conceptions of possible solutions, and work out the political deals necessary to assure participation by enough governments to provide good chances of success. By making the protocols relatively easy to add to, delete, or amend (whether by delegating that task to the meeting of the parties or by using rules specifying a deadline for opting out), they can be revised as new information or new understanding of policy options develops. Governments thus take an incremental approach to regulation, rather than attempting to work out all the details in a single negotiation, as attempted at the Third UN Conference on the Law of the Sea in 1973–1982 (*see* Ozone Layer: Vienna Convention and the Montreal Protocol, Volume 4; United Nations Framework Convention on Climate Change and Kyoto Protocol, Volume 4).

Periodic Situation Assessment

All law-making and law revision rests on assessments suggesting that a problem is sufficiently important to warrant attention, can be solved or abated through consciously directed human action, and will likely be solved or abated by adopting a particular set of regulatory measures. International environmental law has formalized situation-specific assessments to an unusual degree because of the need for scientific investigation to demonstrate the existence and causes of an environmental problem and the complexities of coping with uncertainty as regulatory responses are formulated. Scientific assessments of the dimensions of the environmental issue at hand often precede or form an important part of the initial negotiation of an environmental treaty. The major environmental treaties also provide for periodic assessment of both the state of the natural environment and of relevant technology or regulatory measures through parallel scientific and technical advisory committees. These periodic assessments create what Sand (1996, 792) calls a pre-ordained learning process in which governments are kept informed of changes in or newly developed knowledge about natural systems, technology, and the efficacy of regulatory responses.

Ongoing Implementation Review

Since the perceived successes of the effort to protect the ozone layer, environmental treaties have included: provisions for reporting policy actions and relevant statistics regarding production or emission of pollutants, joint review of efforts in a multilateral implementation committee separate from but reporting to the *Conference of the Parties* (COP) (*see* Conference of Parties, Volume 4), and dispute resolution. The considerable success of non-adversarial approaches to encouraging better implementation relying on consultation, persuasion, and provision of assistance for administrative capacity building have inspired debate among international lawyers and political scientists regarding the relative importance of compliance – promotion and a more traditional enforcement approach in promoting compliance with treaty commitments (Chayes and Chayes, 1995 and Downs *et al.*, 1996 provide clear statements of the contrasting views.)

Extensive Use of Soft Law

Building on the traditional international law distinction between the legally non-binding *voeux* or recommendations of international conferences and the legally binding obligations of written or oral promises made to other governments, international lawyers have developed a new concept, *soft law*. It covers statements of desired goals and agreed policy directions forming a political or moral commitment stronger than summarized in the traditional understanding of recommendations but still falling short of legal obligation (Dupuy, 1992). Most soft law, like the Stockholm Declaration (1972) and Rio Declarations (1992) and Agenda 21 (1992), is the product of agreed statements adopted at international conferences, but some treaty provisions may also express soft law commitments. The use of soft law is particularly extensive in international environmental law because of the widespread belief that environmental problems must be addressed more

rapidly than the traditional forms of treaty-making allow. Advocates of soft law approaches claim that they help overcome the hesitations induced by uncertainty by allowing governments to make preliminary commitments that can be adjusted later and permit negotiations to get beyond least common-denominator policy choices. Critics worry that heavy reliance on soft law encourages papering over disagreements and creates confusion about what governments have or have not committed themselves to do, making it harder for others to call them to account for lagging in implementation.

Explicit Opening of the International Law-Making Process to Participation by Non-state Entities

Domestic democratization in many parts of the world and the vast increase in transnational activism by organized social movements have combined to increase demands for broader participation in the formulation and revision of international environmental law and the monitoring of governments compliance with it. Environmental groups have had considerable success lobbying international conferences, inspiring similar activity by other groups while the scienti c and technological assessment panels generally consist of persons drawn from the scienti c, industry, labor, and environmental communities as well as government bureaucracies (*see* **Integrated Assessment, De nition of**, Volume 4; **ULYSSES (Urban Lifestyles, Sustainability, and Integrated Environmental Assessment)**, Volume 4; **Post-normal Science**, Volume 5).

EMERGING PRINCIPLES OF INTERNATIONAL ENVIRONMENTAL LAW

Each branch of international law contains certain broad principles that express a reason to prefer or avoid a particular result or behavior. These principles do not specify how to act in particular situations; rather, they orient action by indicating the goals and other considerations to be emphasized when deciding how to act. The struggle to de ne these principles has been intense because reorienting international law to take environmental protection more seriously has required rede ning the central international legal principle of state sovereignty to highlight responsibilities to protect the natural environment. Some of this rede nition was accomplished by expanding traditional doctrines of good neighborliness and abuse of rights, and reinterpreting the traditional legal maxim *sic utere tuo ut alienum non laedas* (use what is yours in a manner that does not harm others). While successfully expanding state responsibilities towards other states, these adaptations of traditional doctrine fell short from an ecological point of view because they failed to provide a basis for obligation to protect nature, natural processes, or natural resources when the environmental impact of human activity is con ned within the territory, airspace, or maritime jurisdiction of a single state. Efforts to establish that broader obligation have been troubled by two persistent disagreements. The rst involves different emphases on the balance to be drawn between the generally accepted principle of permanent sovereignty over natural resources and wealth, highlighting the right of each state to adopt its own program of economic organization and regulation, with the emerging rules of environmental protection. The second stems from disagreements between those who believe that the basic purpose of international environmental law is to enhance human existence by promoting living within the limits of nature and those believing that its basic purpose is to protect the intrinsic value of the biosphere. These have occasionally inspired strong controversy on particular policy questions, but have played out more generally in the reluctance of many governments to accept the notion of an environmental right. Governments of developing states in particular have worried that accepting such a right might erode the also controversial right to development that they have been advocating for many years.

Recent intergovernmental declarations, multilateral treaties, and controversies regarding their application in particular situations have yielded considerable convergence on a set of principles of international environmental law establishing obligations to weigh protection of the environment heavily in decision-making and action. Though the exact legal status of some principles remains in doubt, and their generality often leads to controversy about their application in speci c situations, these broad principles provide the basic orienting statements of current international environmental law.

- The principle that states must ensure that activities within their jurisdiction or control do not harm the environment in other states or global common areas. This principle has the strongest connection with traditional international law principles concerning the rights and duties of sovereign states, as well as the clearest expression in international environmental declarations and treaties.
- The principle of sustainable development. This is sometimes treated as part of international environmental law and sometimes as a principle mandating the closer integration of international environmental and international economic law (e.g., Sands, 1996). It was rst formulated in its current terms by the UN World Commission on Environment and Development (1987) (Bruntland Commission), which de ned it as development that meets the needs of the present without compromising the ability of future generations to meet their own needs . The International Court of Justice stopped short of calling sustainable development a legal principle in the Gabcikovo-Nagymaros Case

(1998), but did characterize it as an objective that states are legally obligated to take into account in decisions on economic policy and particular projects. Third World governments have been particularly insistent that the principle means that economic development will not be subordinated to environmental concerns; environmentalists have been equally insistent that it means environmental protection cannot be subordinated to economic development. Yet this mandate to balance and serve both concerns provides little guidance for striking that balance in particular situations. It is clearest on questions of time horizon, strongly favoring adoption of longer-term perspectives in decision-making.

- The principle of intergenerational equity. This is closely linked with that of sustainable development through reciprocal reference. Sustainability suggests maintaining natural systems over time spans longer than a single generation, while intergenerational equity is defined in terms of acting in such a way that future generations have as many alternatives in use of ecosystems and resources as are available to the current generation. The notion that future generations have legally enforceable rights has inspired considerable controversy, but there is no serious dissent from the underlying claim that governments and other actors must shift away from a fixation on the immediate future (whether this be the next profit or loss statement for businesses or the next election for governments) if they are to provide adequate environmental protection. The controversy about how to do so is considerable, and fuelled by the way longer-term thinking rubs against well entrenched economic (the discounting of future returns) and political (constituencies are those persons alive today) thinking.
- The principle of common concern. This articulates the belief that environmental problems are the concern of all states and is an effort to institutionalize two standards of conduct. The first is that environmental protection, like more traditional duties to provide effective administration of territory or newer duties to respect human rights, is one of the standard obligations of sovereigns, one of the duties that, paired with a set of rights to political independence, territorial integrity, and nonintervention by other states, are integral to the status of sovereign. Applied strictly, the principle means that failure to adopt adequate domestic environmental law is itself a violation of international obligations, a reading that governments have been reluctant to take very far. The second is that obligations deriving from customary or treaty-based international environmental law are owed to the international community as a whole, not simply to the state or states directly harmed by failures to protect the environment. Ambitious efforts to establish all failures to comply with an environmental agreement as violations *erga omnes* (matters about which any state may complain) still inspire controversy, but the principle has supported the institutionalization of the right to point out implementation failures by other states on a problem-by-problem basis in the new system of ongoing implementation review. This principle is reinforced by the traditional principles of good neighborliness and cooperation among states, which are not unique to international environmental law though they have environmental expressions.
- The related principle of common but differentiated responsibility. This expresses the general terms of ongoing bargaining between developing states and industrial states regarding who will provide what part of the effort and financial resources needed for international environmental cooperation. It rests on moral claims that the industrial states' greater role in causing environmental problems (by industrializing without the burden of environmental protection measures and continuing patterns of profligate resource consumption and waste generation) plus their greater ability to redress them (through possession of more advanced technology and greater financial resources) mean that they should bear most of the burden of environmental protection. In some respects the principle can be seen as a generalized interstate expression of the *polluter pays principal* (*see* PPP (Polluter Pay Principle), Volume 4). Though triggering intense bargaining whenever invoked in a particular environmental negotiation, the principle has promoted use of several methods of differentiating obligations in recent environmental treaties. These include: allowing grace periods that give developing countries more time to meet emission standards; defining different baselines that allow developing countries to increase pollution levels to accommodate a measure of economic growth; providing financial assistance; and even accepting that developing country implementation is contingent on provision of financial assistance. However, the current emphasis on a binary categorical distinction between industrial and developing states will need modification in coming years, partly because at least some developing countries are or soon will be significant contributors to pollution problems and partly because the distinction ignores the highly varied situations and economic performances of developing states. Shifting from the current focus on differential norms and rules distinguishing amongst named groups of states to a focus on contextual norms and rules based on differences in situation (Magraw, 1990) would permit better accommodation to the environmental implications of differences in size, geography, climate,

and economic circumstances of different countries (*see* Common but Differentiated Responsibility Principle (Stockholm/Rio), Volume 4).

POLICY NORMS (PRINCIPLES) OF INTERNATIONAL ENVIRONMENTAL LAW

International environmental law also includes a number of policy norms that provide guidance regarding the forms or objectives of regulatory measures to be adopted at the international and national levels. Most of these policy norms (frequently called *principles* to emphasize their importance, a reflection of the lack of standardized terminology for the various types of legal prescription in international law) originated in the national law of a particular state, but were adopted more widely as their usefulness in promoting more effective environmental protection was perceived. The policy norms of international environmental law thus serve as a summary of and mandate to adopt what governments agree are the best national regulatory practices. More specific guidance regarding the precise sort of regulatory measure to adopt (direct control, emissions standards, ambient standards, product standards, process standards, recycling or return rules) and the sort of related economic incentive schemes to adopt is specified on a problem-by-problem basis in the growing body of multilateral, regional, and bilateral treaties and codes of conduct developed through intergovernmental organizations.

- The polluter pays principle (norm) which asserts that the costs of protecting the environment should be borne by those whose activities pose dangers to it. It developed primarily in national discussions of how to pay for cleaning up or avoiding pollution, and rests on rational choice assumptions that changing incentives by altering the costs of activity or inactivity will trigger changes in behavior. It was developed in the national environmental law of several industrial states to justify imposition of pollution taxes, direct regulations requiring polluters to install specified pollution abatement equipment, and to signal that the days of free use of the environment as a sink for all manner and quantities of waste were at an end. It serves similar functions at the international level.
- The precautionary principle, initially elaborated primarily in West Germany, endorsed in Principle 15 of the Rio Declaration, and incorporated into several later multilateral treaties, mandates that action be taken to avoid long-term consequences of some present-day activity when there is significant uncertainty about the extent or material causes of damage from an activity. Traditionally such uncertainty was used as an argument for deferring the imposition of limits on an activity. The precautionary principle shifts the terms of argument by placing the burden of proof that an activity is safe on those who would undertake an activity rather than on those who would impose regulations. In the strictest readings, precaution would require regulators to allow only those activities posing no danger of harm to the environment. Implementing the precautionary norm raises three interrelated controversies: 1) what standard of proof should guide assessments of anticipated environmental effects, 2) what constitutes a harm, 3) what constitutes a danger of harm. Defining danger requires answering two questions: how much harm could result from an activity and what is the probability of that harm occurring. Disagreement on all these matters has made the norm controversial among those who regard it as requiring elaboration of a generally applicable standard. The level of disagreement declines when the norm is understood as mandating an approach to policy that favors erring on the side of caution in authorizing or regulating an activity (*see* Precautionary Principle, Volume 4).
- The norm that governments should promote use of environmentally benign technologies. It first developed in Western Europe but has won wide endorsement. Any controversy it inspired is focused on how far technology choices may be based on the comparative costs of adopting different technologies, leading to a range of formulations ranging from *best available technology* (BAT) to *best available technology not entailing excessive costs* (BATNEEC). Developing county governments have invoked the norm to support their claims that they should enjoy access to technologies on more favorable terms than currently prevail in international commerce.
- The norm of prior assessment of environmental impacts of proposed activities, first developed in the USA and later adopted in other parts of the world. It supplements the precautionary principle by requiring serious efforts to understand environmental consequences before major activities proceed, and has been linked to efforts to develop more transparent and accountable government through arguments that it should include opportunities for comment from the general public.
- The prior notification norm, which is both a continuation and an extension of international rules with a long history. Multilateral treaties regarding particularly hazardous activities, such as maritime transport of nuclear materials, operation of nuclear power plants, transport of toxic or hazardous substances, and incidents of oil pollution at sea, have long included obligations of notification. Most of these obligations involve notification of accidents likely to have effects abroad; a few

also mandate the sharing of data regarding hazardous or toxic chemicals or the securing of prior consent from the government of another state before shipments of certain substances to addressees on its territory occur.

- The public information norm, which is well developed in the domestic legal systems of western industrial states, but only beginning to be defined in international law. It is given particularly strong emphasis by those seeking to increase the direct participation of non-governmental organizations and other non-state entities in environmental governance, but arriving at a consensus international definition is difficult because of the variety of national practices on the subject (*see* Governance and International Management, Volume 5).

For treaty texts, see the Fletcher School of Law and Diplomacy Multilateral Treaty Database, www.tufts.edu/fletcher/multilaterals.html and CIESIN List of Multilateral Treaties relating to the Environment, www.sedac.ciesin.org/entri/texts-home.html web sites; for current developments see the *Yearbook of International Environmental Law* and the *Yearbook of International Cooperation on Environment and Development*.

REFERENCES

Agenda 21 (1992) *Agenda 21: Program of Action for Sustainable Development*, United Nations, New York.

Brown-Weiss, E, McCaffrey, C, Magraw, D B, Szasz, P, and Lutz, R E (1998) *International Environmental Law and Policy*, Aspen Law and Business, Gaithersburg, NY.

Caldwell, L K (1998) *International Environmental Policy*, 3rd edition, Duke University Press, Durham, NC.

Chayes, A and Chayes, A H (1995) *The New Sovereignty*, Harvard University Press, Cambridge, MA.

Downs, G W, Rocke, D M, and Barsoom, P N (1996) Is the Good News about Compliance Good News about Cooperation? *Int. Organ.*, 50, 379–407.

Dupuy, P-M (1992) Soft Law and the International Law of the Environment, *Michigan J. Int. Law*, 12, 420–435.

Gabcikovo-Nagymaros (1998) Case (Hungary v. Slovakia) (1997), *Int. Leg. Mater.*, 37, 162.

Kiss, A (1993) Les Traites-cadres: Une Technique Juridique Characteristique du Droit International de l'Environnement, *Annuaire Francaise de Droit International*, 39, 792–797.

Kiss, A and Shelton, D (1999) *International Environmental Law*, 3rd edition, Transnational Publishers, Ardsley, NY.

Magraw, D (1990) Legal Treatment of developing Countries: Differential, Contextual, and Absolute Norms, *Colorado J. Int. Environ. Law Policy*, 1, 69–99.

Rio Declaration (1992) *Rio Declaration on Environment and Development*, United Nations, New York, also in *Int. Leg. Mater.*, 31, 874–880.

Ruster, B, Simma, B, and Bock, M (1983) *International Protection of the Environment: Treaties and Related Documents*, Oceana Publishers, Dobbs Ferry, NY.

Sand, P H (1996) Institution-building to Assist Compliance with International Environmental Law, *Zeitschrift für auslandisches offentliches Recht und Völkerrect*, 56, 773–793.

Sands, P (1996) *Principles of International Environmental Law I: Institutions, Standards, and Implementation*, Manchester University Press, Manchester.

Schachter, O (1991) The Emergence of International Environmental Law, *J. Int. Aff.*, 44, 457–493.

Stockholm Declaration (1972) *Int. Leg. Mater.*, 11, 1416–1421.

UNEP (1990) *International Register of Environmental Treaties*, United Nations Environment Program, Nairobi.

UN World Commission on Environment and Development (1987) *Our Common Future*, Oxford University Press, New York.

Victor, D, Raustiala, K, and Skolnikoff, E B (1998) *The Implementation and Effectiveness of International Environmental Agreements*, MIT Press, Cambridge, MA.

International Social Science Council (ISSC)

see ISSC (International Social Science Council) (**Volume 5**)

ISEW (Index of Sustainable Economic Welfare) and GPI (Genuine Progress Indicator)

Aggregate economic indicators, such as gross domestic product (GDP), are derived from the system of national accounts (SNAs) that are now standardized across nations. The indicators measure the rate of market and government activity of an economy and are used for comparisons over time and between nations. Economic goods and services produced in households, in subsistence economies, and through the contribution of non-market environmental services are not included in the SNAs and hence are not included in the aggregate indicators. The missing information also results in misrepresentations and misinterpretations of the market data that are included (*see* Ecological Economics, Volume 5). These concerns have led to the development of expanded accounting systems and alternative indicators, the most famous of which are the ISEW and its evolutionary offshoot, the GPI.

Economists frequently argue that it is inappropriate to represent GDP *per capita*, or other indices derived from the SNA, as measures of well-being, or even economic

well-being. Nonetheless, economists and others frequently use these terms in exactly this way. In light of this, many have argued that reasonable aggregate measures of well-being need to be developed. The ISEW was conceived by Herman Daly and John Cobb and developed by Clifford Cobb (Daly *et al.*, 1989) to be a superior indicator of economic well-being. They also hoped that the index would indicate what the sustainable level of well-being was. The ISEW corrects, at least partially, for the absence of the household and environmental sectors in the SNA. It also takes the additional step of addressing equity, arguing that adding $1000 to the annual consumption of a rich person does not improve welfare nearly as much as it would for a poor person. Further, the ISEW does not include government expenditures on national defense, health care, prisons, etc., because they are largely partial corrections for problems that have grown worse over time. An ISEW has been calculated for many European countries.

The GPI in the US evolved from the ISEW and is being promoted by a non-governmental organization, *Rede ning Progress*. GDP *per capita* in the US nearly tripled during the second half of the 20th century. The GPI shows well-being increasing more modestly from 1950 into the early 1970s and then drifting downward. GDP and ISEW show similar patterns in Europe. The ISEW, GPI, and other alternative indicators have played a modest role in environmental debates about economic growth in developed countries and have prodded the World Bank (Lutz, 1993) and national governments (Nordhaus and Kokkelenberg, 1999) to consider amending the SNAs. However, alternative indicators have yet to be incorporated in the debate over North–South equity and global environmental change.

REFERENCES

Daly, H E, Cobb, Jr, J B, and Cobb, C W (1989) *Appendix for the Common Good: Redirecting the Economy toward Community, the Environment, and a Sustainable Future*, Beacon Press, Boston, MA.

Lutz, E, ed (1993) *Toward Improved Accounting for the Environment, an UNSTAT-World Bank Symposium*, The World Bank, Washington, DC.

Nordhaus, W D and Kokkelenberg, E C, eds (1999) *Nature's Numbers: Expanding the National Economic Accounts to Include the Environment*, National Research Council, Washington, DC.

FURTHER READING

Uno, K and Bartelmus, P, eds (1998) *Environmental Accounting in Theory and Practice*, Kluwer, Dordrecht.

RICHARD B NORGAARD *USA*

Islam and the Environment

Fazlun M Khalid

Islamic Foundation for Ecology and Environmental Sciences, Birmingham, UK

It has often been observed that Islam cannot ordinarily be described as a religion and that it prescribes a way of life that goes beyond the performance of rituals. The word for religion (d¯n) is found in the Qur'¯an (anglicized spelling: Koran; this is Islam's sacred text on which much of this essay is based). The word d¯n, which appears in 90 different places, often in contexts that place it outside the purely ritual. D¯n in essence describes an integrated code of behavior which deals with personal hygiene, at one end of the spectrum, to our relationships with the natural order at the other. It provides a holistic approach to existence, it does not differentiate between the sacred and the secular and neither does it place a distinction between the world of mankind and the world of nature.

However, this Islamic mode of expression is now severely attenuated, having been swept aside by the forces of history, like the other older traditions, into a domain which treats the natural world exclusively as an exploitable resource. As what we now understand by modernity advanced, as the secular ethic progressively seeped into the Muslim psyche and as industrial development, economic indicators and consumerism became the governing parameters of society, there has been a corresponding erosion of the Muslim perception of the holistic and a withering of its understanding of the sacred nexus between the human community and the rest of the natural order: The creation of the heavens and the earth is far greater than the creation of mankind. But most of mankind do not know it (Qur ¯an 40:56).

For these and other reasons, Muslims in various parts of the world have in recent times sought the reversal of these trends through the re-establishment of Islamic governance based on the Islamic code known as the Shar¯ ah. Deeply embedded in its matrix are detailed and sometimes complex rules, which lay down the basis for Islamic environmental practice. Islamic jurisprudence contains regulations concerning the conservation and allocation of scarce water resources; it has rules for the conservation of land with special zones of graded use; it has special rules for the establishment of rangelands, wetlands, green belts and also wildlife protection and conservation. Much of the traditional institutions and laws associated with sound environmental practice in Islam have now fallen into disuse.

*Although the re-establishment of Islamic governance is an aspiration held by an increasing number of Muslims, there is now a growing movement amongst them, led by thinkers like Seyyed Hossein Nasr (*see **Nasr, Seyyed Hossein**, Volume 5*), who place an immediate priority in dealing with*

the intractable problems that the human race is creating for itself by overexploiting and degrading the planet beyond repair. Credit is due to Seyyed Hossein Nasr for almost single handedly rekindling the consciousness of not only Muslims, but of adherents of other traditions, to the dimension we have come to describe as faith and nature. His first exposition on this subject was a series of lectures in May 1966, delivered at the invitation of the University of Chicago, which was eventually published (Hossein, 1968). He has written voluminously on this theme since then; see, for example, Hossein, 1996. For Muslims this crisis calls for a fresh evaluation of the teachings of Islam and its practice in the present globalized order.

A GLANCE AT THE BASICS

Islam is the name of the religion discussed in this contribution. A Muslim (called a Musselman in the 18th century) is a follower of Islam. The word Islam literally means submission or surrender, and a Muslim is one who surrenders his will to the will of God. A Muslim country is one in which the majority of people are Muslims – there are about 60 such countries in the world.

The roots of Islamic environmental practice are to be found in the Qur'ān and the guidance (sunnāh) of Prophet Muhammad. Prophet Muhammad is the Prophet of Islam and is usually referred to by Muslims as the Messenger of God (Rasūlullāh). He was born in 570 AD in Mecca and died in 632 in Medina, in what is now Saudi Arabia. Over the centuries, Islamic practice has been elaborated by a succession of scholars and jurists responding to real problems experienced by the growing community of Muslims in various parts of the world.

The Islamic worldview is based on the belief in the existence of an all powerful creator who is the same God (Allāh in Arabic) of the other monotheistic faiths, Judaism and Christianity. Muslims learn from the Qur'ān that God created the universe and every single atom and molecule it contains and that the laws of creation include the elements of order, balance and proportion: He created everything and determined it most exactly (25:2) and It is He Who appointed the sun to give radiance and the moon to give light, assigning it in phases ••• Allāh did not create these things except with truth. We make the signs clear for people who know (10:5).

What may be described as the Islamic creed has three aspects. The first is the core value system that Islam establishes, which is islām itself and in this context is taken to mean submission. The second is faith (imān) and the third is good personal conduct (ihsān) traditionally described as righteousness or piety.

Islam, in its verb context, is elaborated in what is commonly described as the five pillars. The first of these is a two-part declaration (shahadā) bearing witness that there is no God but God and that Muhammad is the Messenger of God; the second is the performance of the five daily prayers (salāt); the third is the annual payment of a specific tax (zakāt); the fourth is to fast during the month of Ramadhan (sawm); the fifth is to make the pilgrimage (hāj) to Mecca at least once in a lifetime for those who have the means to do it.

Imān was described by the Prophet Muhammad as knowing with the heart, voicing with the tongue and expressing with the body. This requires the profession of faith in God, His angels, His books, His messengers, the final day and acceptance that life in all its expressions emanates from the divine source.

Ihsān is described as the act of worshipping Allāh as if you see Him, knowing that even if you do not see Him, He sees you. This goes much beyond the ritual prayers; and every good action performed by a believer is seen as an act of worship. This is commonly expressed by Muslims as doing what is pleasing to Allāh, who is ever present, ever watchful.

Taken as a whole as it is intended to be, caring for planet Earth, our only home, is integrated within the framework of this value system. This is then an everyday concern for the Muslim, which the Qur'ān reminds him by saying, We have not omitted anything from the Book (6:39) and He said Our Lord is He Who gives each thing its created form and then guides it' (20:49). We will now look at how this is reflected in daily life.

THE ETHICAL DIMENSION

One of the stories often told by Muslims concerning the environment is the instruction by Abu Bakr, the first Caliph (Khalif) of Islam to his armies. In addition to telling them not to harm women, children and the infirm, he ordered them not to harm animals, destroy crops or cut down trees. There were two elements present in this decree: the first, to establish justice even as the Muslim armies fought, and the second, to recognize the value of nature. It should also be noted that the environment was not an issue or subject for separate treatment in life as it flowed onwards in both war and peace. The human condition was never separated from the natural order. It was a matter to be reckoned with at every moment of existence like the very air we take into our lungs.

Abu Bakr was the first of four rightly guided caliphs who succeeded the Prophet after his death. They were known as such because of the lengths to which they went to incorporate the instructions in the Qur'ān and the Prophet's example into their rule. The rule of the rightly guided caliphs lasted from 632 to 661 (AD) and became the time, including the time of the Prophet, which all Muslims to this day seek to emulate. In attempting to reproduce the Prophetic model what the rightly guided

caliphs did was to set a pattern that would evolve into the matrix that formed the basis of the Islamic legal system known as the Shar¯ah and which very soon came to circumscribe Muslim life. The word Shar¯ah has an interesting etymology with strong environmental connotations. It means a way or path in Arabic and its derivation refers to the beaten track by which wild animals come to drink at their watering places. It is the road that leads to where the waters of life flow inexhaustibly (Eaton, 1994; 180).

As Muslim populations grew and expanded territorially, their requirements in government became increasingly complex and the Shar¯ah accordingly became more sophisticated. To the Qur'ān and Sunnah were added two other elements: the consensus (ijmā) of scholar jurists and the process of reasoning by analogy (qiyās). Islamic law (fiqh) evolved out of this process and there are two other traditional instruments which were incorporated into this system and which could usefully serve the purpose of formulating environmental law in the Muslim world today. The first of these is interpretation in context (ijtihād) and the second is custom and practice (urf wa adāt).

The Shar¯ah expanded and evolved within this framework to set defining standards for Muslim behavior within the divine decrees of the Qur'ān, including *inter alia* family law, civil law, commercial law and environmental law. From the time of the Prophet, law has taken precedence over theology; this is stated to be the case because Islam's concern is about humankind and its relationship with the rest of creation, beginning, of course, with itself. The law is about defining these relationships and the following are examples of how the Qur'ān sees it (4:134).

> You who have imān (faith); be upholders of justice, bearing witness for Allāh alone, even against yourselves or your parents and relatives. Whether they are rich or poor, Allāh is able to look after them. Do not follow your own desires and deviate from the truth. If you twist and turn away, Allāh is aware of what you do

Truth takes precedence over love, even love for one's parents, and love is known habitually to conceal the truth. There is no compromise with truth and justice,

> so call and go straight as you have been ordered to. Do not follow their whims and desires but say, I have imān (faith), in a book sent down by Allāh and I am ordered to be just between you ••• (42:13)

The Qur'ān also asks us to be just to our natural surroundings, We did not create the heavens and earth and everything between them, except with truth (15:85).

Like the Muslims who had succeeded the Prophet and had attempted to give expression to the divine decree, the scholar jurists approached this matter with great diligence and formulated an ethical base derived from the imperatives laid down in the Qur'ān. These imperatives come under numerous headings but they could be distilled into just three categories for our purposes, bearing in mind that public good (al masalih al mursalah) is the ultimate objective (Doi, 1984). Muslims are to do what is right, forbid what is wrong and act with moderation at all times: let there be a community among you who call to the good, and enjoin the right and forbid wrong. They are the ones who have success (3:104). The Qur'ān again uses an environmental theme in exhorting humankind to be moderate; (6:142):

> it is He who produces gardens, both cultivated and wild, and palm trees and crops of diverse kinds and olives and pomegranates both similar and dissimilar. Eat of their fruits when they bear fruit and pay their dues on the day of their harvest, and do not be profligate. He does not love the profligate.

The Shar¯ah also evolved within the guidelines set by three principles agreed upon by scholar jurists over the centuries. They are:

1. the interest of the community takes precedence over the interests of the individual;
2. relieving hardship takes precedence over promoting benefit;
3. a bigger loss cannot be prescribed to alleviate a smaller loss and a bigger benefit takes precedence over a smaller one. Conversely a smaller harm can be prescribed to avoid a bigger harm and a smaller benefit can be dispensed with in preference to a bigger one.

THE NATURAL ORDER IN ISLAM

As the Islamic tapestry unfolded in its expression over the centuries, we discover that there are no references to the environment, as we understand it today. The word *nature*, which is an abstraction, cannot be found in the Qur'ān and the closest modern Arabic usage is the word b¯'a which connotes a habitat or a surrounding. The word nature will continue to be used in this essay for linguistic convenience. For a further explanation on terminology, see Khalid (1999). The Qur'ān speaks of creation (khalq) and it contains two hundred and sixty one verses where this word is used in its various grammatical forms derived from the root kh l q. These verses contain references to the human world; to the natural world of the Earth, from trees to turtles, from fish to fowl; and to the sun, stars and skies. The very first revelation of the Qur'ān to the Messenger used this word in its verb form to dramatic effect, Recite in the name of your Lord who created, created man from clots of blood (96:1). Creation is the fabric into which the tapestry of life is worked.

Creation or nature is referred to as the signs (āyāt) of Allāh and this is also the name given to the verses of the Qur'ān. Āyāt' means signs, symbols or proof of the divine.

As the Qur'ān is proof of Allāh so likewise is His creation. The Qur'ān also speaks of signs within the self and as Nasr (1993) explains,

> ••• Muslim sages referred to the cosmic or ontological Qur'ān ••• they saw upon the face of every creature letters and words from the cosmic Qur'ān ••• they remained fully aware of the fact that the Qur'ān refers to phenomena of nature and events within the soul of man as āyāt ••• for them forms of nature were literally āyāt Allāh.

As the Qur'ān says, there are certainly signs in the earth for people with certainty; and in yourselves. Do you not then see? (51:20, 21).

Writing from another perspective, Faruqi and Faruqi (1986) say:

> Its (nature's) goodness is derived from that of the divine purpose. For the Muslim, nature is ni'māh, a blessed gift of Allāh's bounty ••• to transform in any way with the aim of achieving ethical value ••• Since nature is Allāh's work, His āyāt or signs, and the instrument of His purpose which is absolute good, nature enjoys in the Muslim's eye a tremendous dignity.

The Qur'ān says about these matters (16:65–69):

> Allah sends down water from the sky and by it brings the dead earth back to life. There is certainly a Sign in that for people who hear. There is instruction for you in cattle. From the contents of their bellies, from between dung and blood, We give you pure milk to drink, easy for drinkers to swallow. And from the fruit of the date palm and the grapevine you derive both intoxicants and wholesome provision. There is certainly a Sign in that for people who use their intellect. Your Lord revealed to the bees: Build dwellings in the mountains and the trees, and also in the structures which men erect. Then eat from every kind of fruit and travel the paths of your Lord, which have been made easy for you to follow'. From inside them comes a drink of varying colours, containing healing for mankind. There is certainly a Sign in that for people who reflect.

The universe we inhabit is a sign of God's creation as is the environment of our innermost selves. They both emanate from the One Source and are bonded by only one purpose, which is to serve the divine will. This bonding of the cosmic to the subatomic is the deep ecology of Islam but it is not a relationship of equals as we can see in the hierarchy of the food chain dominated by Man. Whilst the primary relationship is that between the Creator and the rest of His creation, the Creator Himself determined a subsidiary one, that between Man and the rest of His creation which the Qur'ān defines as follows: It is He Who created everything on the earth for you ••• (2:28); We did not create heaven and earth and everything between them as a game ••• (21:16); We did not create heaven and earth and everything between them to no purpose (38:26); ••• He wanted to test you regarding what has come to you ••• (5:48).

The Qur'ānic view holds that everything on the earth was created for humankind. It was God's gift (ni'mah) to us, but a gift with conditions nevertheless and it is decidedly not something that one runs and plays with. The earth then is a testing ground of the human species. The tests are a measure of our acts of worship (ihsan) in its broadest sense. That is living in a way that is pleasing to Allāh, striving in everything we do to maintain the harmony of our inner and outer environments.

As our interaction with the environment evolved, it manifested itself in a range of rules and institutions. As the Muslim community expanded out of its sparse desert environment, it was confronted by many challenges, one of which was relative abundance. This brought about other problems like over exploitation and waste. Muslims applied themselves to these problems assiduously and it would be salutary to look at this legacy. For more information on this subject see Bagader *et al.* (1994); Llewellyn (1992); Dien (2000). The following is a brief summary of how the Shari'ah developed in this area over the past 1400 years.

Legislative Principles

1. Allāh is the sole owner of the earth and everything in it. People hold land on usufruct that is, for its utility value only. There is a restricted right to public property.
2. Abuse of rights is prohibited and penalized.
3. There are rights to the bene ts derived from natural resources held in common.
4. Scarce resource utilization is controlled.
5. The common welfare is protected.
6. Bene ts are protected and detriments are either reduced or eliminated.

Institutions

1. People who reclaim or revive land (ihyā al māwāt) have a right to its ownership.
2. Land grants (iqtā) may be made by the state for reclamation and development.
3. Land may be leased (ijārā) for its usufruct by the state for its reclamation and development.
4. Special reserves (h¯mā) may be established by the state for use as conservation zones.
5. The state may establish inviolable zones (al-har¯m) where use is prohibited or restricted. Every settlement has a right to create such zones managed by the people and where use is severely restricted. Additionally, it is permitted to establish these zones adjacent to sources of water and other utilities like roads and places of public resort.
6. Makkāh and Mad¯nah are known as the two inviolable sanctuaries (al-haramain) where trees cannot be cut down and animals are protected from harm within their boundaries. They serve as examples of best practice.

7. Charitable endowments (awqāf) may be established with specific conservation objectives.

Enforcement

The primary duty of the Islamic state is to promote the good and forbid wrong doing. As part of these functions, it has the mandate to protect land and natural resources from abuse and misuse. From its earliest years the Islamic state established an agency known as the hisbā whose specific task was to protect the people through promoting the establishment of good and forbidding wrong-doing. This agency was headed by a learned jurist (muhtasib) who functioned like a chief inspector of weights and measures and chief public health officer rolled into one. He was also responsible, among other similar duties, for the proper functioning of the h¯mā and al har¯m zones and he acted as what one might describe as an environmental inspectorate.

The development and application of these principles and institutions have seen a decline over the past two centuries, as another worldview based on the exploitation of natural resources for profit gradually overtook this model. We are experiencing the consequences of this now. However, there are clear indications as to how this Islamic heritage has been and could again be put to good use in the modern context.

AN EMERGING RESPONSE

The Muslim world was gradually co-opted into the new world order by force of arms and by force of economics. The first is history and the second present day reality and what the Muslims lost in between these twin processes was the sap of Islam and their unique way of perceiving the universe. For Muslims, their *d¯n* is a whole, an organic reality, where every element has a function as a part of this whole. For example, Islam law does not make sense without the ethical dimension of the divine revelation. Law in the West is amoral. It deals with human ends for human purposes. The Muslim idea of the highest form of civilization is that it is the one that is pleasing to Allah. In today's Western-dominated global order of which Muslims are a part, conspicuous consumption has become the highest imperative. Muslim nation states of which there are now about 60, are willing co-optees to this consumer ethic.

These are however, superficial considerations; and deep in the interface between Islam and the West are two irreconcilable factors that call for some examination. The first is the matter of existence itself and how it is to be defined and the second is the matter concerning money and how it is to be used.

The traditional worldview, which includes that of the West, was challenged by what we have come to know as the Enlightenment (*see* **Enlightenment Project**, Volume 5), having its origins in 16th century Europe. These events are usually seen as a time in which science began its ascendancy over religion. Tarnas (1996) observes that this movement achieved its maturity in the 19th century, finally resulting in a radical shift of psychological alliance from the divine to humankind. Descartes (the French philosopher, mathematician) finally breached the floodgates of the old order by splitting mind from body and proclaiming a dualistic worldview in his well known statement I think, therefore I am (*cogito ergo sum*). One result of the Enlightenment was science, including the scientific capacity for rendering intelligible certain aspects of the material world and for making mankind (in Descarte's own words) Master and possessor of Nature .

This view is on a collision course with how Islam teaches the Muslim to view the world. There is only one master and possessor of nature and that is the One Who created it, Allāh. This is unequivocally expressed in the very first line of the Qur'ān Praise be to Allāh, the Lord of all the worlds (1:1) and the last verse in the Qur'ān Say: I seek refuge with the Lord of mankind, the King of mankind, the God of mankind' (114:1–3) and repeatedly in between. After Descartes, Isaac Newton's world view led to the well known mechanistic conception of the universe and totally away from the holistic and organic interpretation of things. The result was after the 17th century science and religion became totally divorced (Nasr, 1990).

People would find it impossible to live in today's world without money, but one increasingly comes across interesting appraisals of it like the following for example: ••• in spite of all its fervid activity, money remains a naked symbol with no intrinsic value of its own and no direct linkage to anything specific (Kurtzman, 1993). Money has come to be recognized as mere tokens and:

> there is something quite magical about the way money is created. No other commodity works quite the same way. The money supply grows through use; it expands through debt. The more we lend, the more we have. The more debt there is, the more there is.
>
> (Kurtzman, 1993)

These tokens of value that we create from nothing and use everyday grow exponentially ad infinitum. But we know that the natural world, which is subject to drastic resource depletion, has limits and is finite. This equation is lopsided and the question is for how long can we continue to create this infinite amount of token finance to exploit the real and tangible resources of a finite world. Looked at from this perspective, money, as the modern world has contrived it, assumes the characteristics of a virus that eats into the fabric of the planet. The consequences of this become visible as global environmental degradation.

This magical system underwent a metamorphosis in 1971 when President Nixon of the US unilaterally abandoned

the gold standard. This is not the place to go into the background of this event, suffice it to say that, by abandoning the gold standard ••• he also moved the world into a new standard: the interest standard . (Kurtzman, 1993). It is generally known that Islam prohibits usury or the taking of interest and the term used in the Qur'ān for this is riba. (For an appraisal on interest see Diwaney, 1997) This term has wide connotations and simply put, it means one cannot have something out of nothing. Thus, riba is also seen as prohibiting the free creation of credit. The Qur'ān denounces these practices vehemently and we can see why from the foregoing discussion: those who practise riba will not rise from the grave except as someone driven mad by shaytan's (satan's) touch (2:274); also, you who have imān (faith)! have taqwā (awe) of Allah and forgo any remaining riba if you are muminūn (believers). If you do not, know that it means war from Allāh and his Messenger (2:277, 278). No other proclamation in the Qur'ān matches this degree of trenchancy.

The Muslim world today is comprised of nation states (not an Islamic concept) that are wholly or mostly Muslim and Muslim minorities, from the very large to the very small, living in non-Muslim countries. Muslim countries may be divided into two categories. The first, into those that describe themselves as Islamic states but who will freely admit that the way they run their affairs is a far cry from the example laid down by the Prophet, the Right Guided Caliphs and the paradigm of the Qur'ān. The second group consists of secular Muslim states where Islam plays a ritualistic role in varying degrees of intensity. However, both these groups, as willing or unwilling co-optees to the new world order, resort to the methods and institutions that are part of it, in order to trade for instance. One such institution is the banking system, and every Muslim state, Islamic or secular has banks with or without the approval of the scholar jurists who did much to build the Shar¯'ah over the centuries into a potent force for just government.

It should be obvious from this that it becomes almost impossible for Muslims whether individuals or nation states, to give expression to a normative Islam under the present globalized system. A system that is in direct conflict with two fundamentals that are part of the Islamic worldview. There now prevails a schizoid tendency in Muslim society whereby it strives to maintain its deep attachment to Islam while it persists in tasting the fruits of the current order.

However, people are increasingly taking a serious look at the current model of endless growth that supplies a contrived demand for consumer goods, which appears to be insatiable and which is, in the end, impacting on the biosphere with drastic consequences. This concern is reflected in the West by the growth of activism and protest organizations. These same trends (differently expressed) appear in the Muslim world. The form this usually takes is the demand for the establishment of the just order under the Shar¯'ah to replace the socialist and capitalist models, which are now seen as failures. It is difficult to see where these demands for change will take Muslims but it is increasingly felt that the current order is untenable in terms of keeping the planet in good repair.

Muslim minorities, particularly those living in the West, have an important role to play. Living in the belly of the beast gives them the advantage of perspective and their understanding of events could lead to an unique contribution to the melting pot of ideas that could lead us out of this impasse. For example, Muslim groups in the West are re-examining alternative currencies and ways of trading that are based on the fiqh of Islam and not on falsehood (Douthwaite, 1996). Corruption has appeared in both land and sea because of what people's own hands have brought about so that they may taste something of what they have done, so that hopefully they will turn back (30:40). In other words learn from your mistakes.

One could say with a reasonable degree of certainty that the environmental problems we see today would not have happened in a society ordered in accord with Islamic principles because its world view defined limits to human behaviour and contained excess (Khalid, 2000). We have seen how this was done in the realm of environmental protection although its stated aim was not precisely that. Rather it came about in the course of establishing public good, one of the basic principles of the Shar¯'ah. Safeguarding against human excess had the effect of protecting the natural world.

The Shar¯'ah evolved holistically and new situations were dealt with through the processes we have discussed above. However, there is nothing to stop this from continuing. However there are important impediments to its proper application today:

1. The Shar¯'ah is no longer supreme even in Islamic states because of the dominance of the global system now in place. The influence of international trade and finance is a case in point.
2. The Hisbah which was once the environmental enforcement agency is now virtually non existent.
3. Increasingly civil administration is separating itself from the body of the people who are coming to be known as the religious authorities, i.e., a clergy, which is not recognized in Islam. This is the case in the Sunni tradition of Islam, which accounts for about 85% of the world's Muslim population. The Shia tradition has an established clergy. The concept of *d¯n* is wearing thin and as a consequence the holistic approach suffers.
4. Following the Western model, Muslim states increasingly function in watertight compartments. As a mirror of what is happening in the West, Muslim economists and environmentalists are two separate species with opposing perspectives.

5. The nation state model, which all Muslim countries have adopted, has economic development as its highest priority. Coping with environmental change is much lower on the scale.

The following is a crystallization of what we consider to be the essentials which will bring into focus the dimensions of change that are needed today from an Islamic perspective. These ideas have been developed by us over the past decade and discussed at various international forums. However, we are at a disadvantage because the principles we are discussing have been plucked out of the Islamic tapestry so assiduously woven over the past 14 centuries. They work best as part of the whole, but how they fit together in today's context and how the whole can be improved, applied and made sense of are matters for open discussion.

The planetary system, the earth and its ecosystems, all work within their own limits and tolerances. Islamic teaching likewise sets limits to human behavior as a control against excess and it could be said that the limits to the human condition are set within four principles. They are the unity principle (Tawh̄d); the creation principle (Fitra); the balance principle (Mizān); and the responsibility principle (Khal̄fa).

The Unity Principle

Tawh̄d is the foundation of Islamic monotheism and its essence is contained in the declaration (Shahadā) which every Muslim makes and is a constant reminder of faith. It is there is no God but God (lāilāha illa-llāh) and is the foundational statement of the Unity of the Creator from which everything else flows. The second half of this declaration is Muhammad is the Messenger of Allah (Muhammadur Rasulullah). He is Allah, Absolute oneness, Allah the Everlasting Sustainer of all (112:1, 2). It is the testimony to the unity of all creation and to the fabric of the natural order of which humankind is an intrinsic part: what is in the heavens and the earth belong to Allah. Allah encompasses everything (4:125). This is the bedrock of the holistic approach in Islam as this affirms the interconnectedness of the natural order.

The Creation Principle

The Fitra principle describes the primordial nature of creation: Allah's natural pattern on which He made mankind (30:29). Mankind was created within the natural pattern of nature and being of it, its role is defined by this patterning itself. For a discussion on Fitra, see Yasien Mohammed (1996). Fitra is the pure state, a state of intrinsic goodness and points to the possibility that everything in creation has a potential for goodness and the conscious expression of this rests with humankind.

The Balance Principle

In one of its more popular passages, the Qur'ān describes creation thus (55:1–5):

> The All-Merciful taught the Qur'ān, He created man and taught him clear expression. The sun and moon both run with precision. The stars and the trees all bow down in prostration. He erected heaven and established the balance

Allāh has singled out humankind and taught it clear expression – the capacity to reason. All creation has an order and a purpose and is in a state of dynamic balance. If the sun, the moon, the stars did not bow themselves, i.e., serve the purpose of their design, it would be impossible for life to function on earth. This is another way of saying that the natural order works because it is in submission to the Creator. It is Muslim in the original, primordial sense.

The Responsibility Principle

This principle establishes the tripartite relationship between the Creator, humankind and creation. God created everything for humankind and appointed it the vice-regent (Khalif) on this earth: it is He Who appointed you Khalifs on this earth (6:167). This role was one of trusteeship (amānāh) which imposed a moral responsibility, We offered the Trust to the heavens, the earth and the mountains but they refused to take it on and shrank from it. But man took it on. (33:72). This assumption of responsibility made humankind accountable for their actions, Will the reward for doing good be anything but good? (55:59).

We can deduce in outline from these four principles that creation that is both complex and finite is yet exact. It emerged from one source and was designed to function as a whole. Humankind, like the rest of the natural world, was, as part of the natural patterning of creation, in a state of goodness with potential for good actions. It is inextricably part of this pattern, but is the only element of it with choice, that can choose to act against the divine Will using the very gift of reasoning bestowed upon it by the Creator. Submission to the divine will, the natural law that holds in check the instincts of the predator, is the way to uphold our responsibilities as the Creator's Khalif. Humankind are the guardians of the natural order.

CONCLUSION

The position set out in this essay is that of Islam and as we have observed earlier, there is a clear issue of conflicting paradigms. The question is how to implement the teachings and practices of one worldview into an institutional framework devised by another that has a diametrically opposite outlook. Although it may be possible to incorporate the principles of Islamic environmental law into the legislative

programs of Muslim countries, the problem will remain one of implementation unless appropriate institutional arrangements are made to replace the old and now virtually defunct ones. It is also interesting that efforts to meet the challenge of environmental change in Muslim countries are made by secular agencies; this is only to be expected, as this is generally not seen as a Muslim problem. The role of the Muslim scholar jurists has been redefined into a religious one as the nature of the Muslim state has changed in its attempt to cope with modernity. The Muslim scholar jurists were unwilling partners to this change and were replaced by a secularized administration, which gradually weakened the scholar jurists' influence as it began to take charge. However, there is an awakening amongst Muslims to the realities of environmental change and therein lies a paradox. There is just an off chance that this may lead Muslims to the discovery that they may be on the wrong side of the tracks.

Note

The translation of the Qur'ān used is that by Bewley and Bewley (1999). In subsequent references to the Qur'ān in this essay, a quotation is followed by the chapter and verse numbers in parentheses. The actual names of the chapters have been left out for convenience.

Transliteration of Arabic vowels

a pronounced as *a* in cattle, ā pronounced as *a* in father, *i* pronounced as *i* as in n, ī pronounced as ee as in sheep, *u* pronounced as oo in foot, ū pronounced as oo in soon.

REFERENCES

Bagader, A A, El-Chirazi El-Sabbagh, A, As-Sayyid Al-Glayand, M, Samarrai, M U I, and Llewellyn, O A (1994) Environmental Protection in Islam, *IUCN Environmental Policy and Law Paper*, **20**, second revised edition, Gland, Switzerland.

Bewley, A and Bewley, A (1999) *The Noble Qur ān*, Bookwork, Norwich.

Dien, M I (2000) *The Environmental Dimensions of Islam*, Lutterworth, Cambridge.

Doi, A R I (1984) *Shariah: The Islamic Law*, Taha, London.

Douthwaite, R (1996) *Short Circuit*, Green Books, Totnes Devon.

Eaton, G (1994) *Islam and the Destiny of Man*, Islamic Texts Society, Cambridge.

El Diwaney, T E (1997) *The Problem With Interest*, Ta Ha, London.

Faruqi, I R al and Faruqi, L L al (1986) *The Cultural Atlas of Islam*, MacMillan, New York.

Hossein, N (1968) *Man and Nature The Spiritual Crisis in Modern Man*, Allen and Unwin, London.

Hossein, N (1996) *Religion and the Order of Nature*, Oxford University Press, New York.

Khalid, F M (1999) *Qur ān, Creation and Conservation*, Islamic Foundation for Ecology and Environmental Sciences, Birmingham.

Khalid, F M (2000) Islamic Pathways to Ecological Sanity An Evaluation for the New Millennium, Ecology and Development, *J. Inst. Ecol.*, **3**, Bandung, Indonesia.

Kurtzman, J (1993) *The Death of Money*, Little, Brown & Co, Boston, MA.

Llewellyn, O (1992) *National Legal Strategies for Protected Areas Conservation and Management*, Paper delivered at IVth World Congress on National Parks and Protected Areas, Caracas, Venezuela.

Nasr, S H (1990) *Man and Nature, The Spiritual Crisis in Modern Man*, Unwin Hyman, London.

Nasr, S H (1993) *The Need for a Sacred Science*, Curzon Press, Surrey.

Tarnas, R (1996) *The Passion of the Western Mind*, Pimlico, London.

Yasien, M (1996) *Fitra*, Ta-Ha, London.

ISSC (International Social Science Council)

ISSC (Conseil International des Sciences Sociales) was founded in October 1952 on the basis of a resolution adopted by the 6th United Nations Educational, Scienti c and Cultural Organization (UNESCO) General Conference in 1951, with its secretariat located in UNESCO s headquarter in Paris, the ISSC is an international non-pro t scienti c organization. It has as it aims and objectives the interdisciplinary coordination of research and the promotion of the understanding of human society in its environment. As such, and in line with its primarily organizing functions, ISSC concentrates on the promotion and coordination of social sciences on a global scale. The ISSC, at present is comprised of 14 scienti c international associations/organizations (member associations), six national academies and research councils (member organizations) and a number of associate members, most of them social science associations.

ISSC s ultimate goal is to encourage social scientists and social science organizations to engage themselves in scienti c research of global-scale problems with a special focus on social, cultural and economic issues. In line with this mission statement, ISSC describes its main tasks and functions as follows:

1. to encourage and promote research in the social and behavioral sciences for the bene t and well-being of humanity and its environment, including global issues of concern to the world community;

2. to encourage and pursue global, international and inter-organizational collaborative scientific activities in the social sciences at a multi-disciplinary and inter-disciplinary level among the members of the ISSC;
3. to stimulate, develop, coordinate, enhance and promote inter-disciplinary social scientific programs across institutional, national and regional contexts;
4. to collaborate with the United Nations and its agencies, and particularly with other governmental and non-governmental organizations to further these aims;
5. to provide, through suitable channels, information about its activities and the world's social science community to the public and to other interested parties within the public at large.

In pursuit of these aims and goals, ISSC maintains, at present, a limited number of both interdisciplinary and international research programs. These include the Comparative Research Program on Poverty, and the Conflict Early Warning System Research Program. Together with the International Council for Science, ISSC also co-sponsors the International Human Dimensions Program on Global Environmental Change, a program of special importance for the cooperation between natural and social sciences.

More information on ISSC can be obtained from www.unesco.org/ngo/issc, ISSC/CISS, UNESCO-1, rue Miollis, 75732 Paris Cedex 15, France. Tel.: • 33 (0)1 4568 2558; Fax: • 33 (0)1 4566-7603.

ECKART EHLERS *Germany*

J

Jains and the Environment

Kerry Brown[1•2]
[1]*Harvard University, Cambridge, MA, USA*
[2]*Oakland, CA, USA*

The Jain Declaration on Nature:

> *All the Venerable Ones of the past, present and future discourse, counsel, proclaim, propound and prescribe in unison: Do not injure, abuse, oppress, enslave, insult, torment, torture or kill any creature or living being (Acharanga Sutra 1.4.1.1-2).*
> *translated by Singhvi (1990)*

For nearly three millennia, the Jain people have inhabited a universe that is both inter-galactic and microscopic in its proportions, and bustling to bursting with souls, all of whom are to be treated non-violently (while not proposing harm to microscopic one-sensed beings, Jainism does allow that the practical limit of ahimsa (universal non-violence) is life-forms with two or more senses).

Ahimsa is the first principle of Jainism, and it is a tall order, as the Jains well know. What do you eat? Where do you sit?

The Green consciousness now trespassing on global consumer culture presents a critical challenge and opportunity for a faith that, of all the so-called world faiths, is most obviously and self-consciously saddled with being environmental. If these people can't set an example, who can?

Jainism is of course an indigenous religion, indigenous to India, and most of the seven million Jains still live there with concentrations in Maharashtra, Rajasthan, Gujarat, Madhya Pradesh and Karnataka (Jain, 1986) and perhaps 200 000 Jains in East Africa, the United Kingdom (UK), North America and lightly scattered around the Pacific Rim. More precisely, it is indigenous in the sense that, as well as its 2500 year old philosophical and socio-political roots shared with Buddhism that provide the ascetic, atheistic flavor, and the metaphysical credentials of world religion, there is an older root that is distinctly animist. Together they generate in Jainism one of those invaluable irresolvables that keep a religion alive and tussling with itself. While at one level the Jains are or should be, according to their scriptures, working as hard as they can to get out of the material bondage and heading for eternal salvation, this is a religion that has busied itself with mind-boggling catalogues of the universe's geography, mathematics, biology and physics in which time, space, matter, motion, rest, and souls are all wrapped up together. It is a religion that believes in an infinity of souls that have never existed separately from matter, nor have they created it. Each soul has its own beginningless, endless autonomy, is beleaguered by *karma* (spiritual guilt) that exists as physical particles, and, even after eternal release, will maintain shape and size. It is a religion that has its feet dug firmly, and rather comfortably, in a very real earth while advocating some of the most severe physical hardships imaginable to unearth oneself.

So this is the grain of sand rubbing on the religious order (like the middle-eastern religions' problem of good and evil with an omniscient and beneficent God or Buddhism's compassion for an ultimate emptiness), that either produces a pearl of great value or a very sore oyster. The inner tension has probably contributed to Jainism's long-standing durability and dynamism at both the cutting edge of Indian society and in its courts of power. It is hard to be swept away when the ground you stake is wide, when you've got the skinniest monks and the most well padded magnates. At the most obvious level, the rise of vegetarianism and the demise of animal sacrifice in the sub-continent owe much to the Jains. Mahatma Gandhi's non-violent revolution had Jain fingerprints on it and he acknowledged his Jain friend and mentor Rajchandra Mehta, along with Tolstoy and Ruskin, as the three people who had most deeply influenced him. (Gandhi, 1958). But the modern demands on environmental sensibility and practice are more extensive than even the Jain ascetics' whisk-broom, sweeping tiny creatures out of danger of their tread. And while over 7000 Jain monks and nuns are probably causing less harm than any one on the planet, what about the rest of the Jains?

At a practical level, many Jains are almost outfaced by their faith's complex cosmology and demanding ethics. It is not surprising that they rest on personal vegetarianism, the base-camp of Jainism, and up the ante by removing figs, garlic and onions (thought to have a density of microcosmic life on them), resort to only fruits and leaves of plants, or

just go on a fast. Supplementary actions include sponsorship of one of the Jain animal sanctuaries in India, tree planting, public promotion of vegetarianism and protest on matters such as the killing of millions of cattle with mad cow disease in Britain, culling of kangaroos by Australian farmers, and veal production in Europe. And in the meantime the modern lifestyle of a successful and affluent community has developed which begs all sorts of questions. Nowadays Jain businessmen are part of the industrial complex, jetting the world in planes that dispense more pollutants in one flight than their Mercedes manages in a year, but they would probably still hesitate to burden a horse and ride to London (in Jainism, animal transport is considered an infringement of the animal's autonomy and ease).

Taking the letter of the law expressed a 2000 years ago, identifying its spirit, and creating updated guides for living that are commensurate with where we are now, is a work in progress. This work by one of the oldest guards of the environmental movement has some interesting indicators for us all.

HISTORICAL BACKGROUND

Like most latter-day world religions (with the exception of the body of traditions we call Hinduism), Jainism began as a challenge to assumed spiritual values and, by implication, the socio-economic order.

The first Jains were nirgrantha (the unattached ones). Vardhamana (599–527 BCE) who became Lord Mahavira (great hero), divested himself of the princely trappings of his kshatriya caste, right down to stripping off his clothes and pulling out his hair. He was the 24th and last Jain Tirthankara (ford-builder) in the most recent of the endless cycles of Tirthankaras or Jinas (spiritual victors).

Mahavira was an historical contemporary of Gautama Buddha in the Magadha region, near modern Patna, in northern India. Both men were leaders of anti-brahmanical (and rival) sects in the region, teaching a theory and praxis that had universal law rather than god at the helm, applied itself without caste privilege or the need for brahmanical mediation, and herded souls through the material plane of suffering toward something infinitely more desirable and everlasting. These reform movements were part of the late Vedic period (*circa* 1000–500 BCE) in which meditation and ascetism became paths to a oneness beyond the hurly burly of Indian gods and goddesses. They were also rattling the caste system's cage. Buddhism and Jainism are the only two reform sects from this period to have survived and become religions in their own right, albeit with very different histories, the former fading from India to become one of the largest religions in the world, the latter finding a minority niche in India which it has frequented to this day.

From the start, Jainism, even more than Buddhism, or indeed any other religion, has seen each being's spiritual progress as their personal business, but their physical safety in the meantime as a matter of shared concern. Everybody, however small and simple their consciousness, has a right to be here for their own sake, no one else's, and stay as long as they can, without harm from others. It appears that Mahavira or his followers in some way absorbed a strand in the elaborate braiding of Indian traditions that had a closer eye on the physical here and now. It may have come from previous teachers of the nirgranthas: Pashva, the 23rd Jain Tirthankara (*circa* 850 in Varanasi), or Nemi, the 22nd Tirthankara, thought to have flourished in Saurashtra and been contemporary with Krishna (see Renou, 1953, 114; Jacobi, 1884, 276–279). But from wherever this commitment to understanding the nature and unshakeable physicality of the world came, it has stuck.

> First knowledge of the world, then compassion for it (*Dashavaikalika* 4.10).
>
> translated by (Dundas, 1992, 138)

By the third century BCE, Jainism had divided into two sects, following the sojourn of a group of Jains during famine in the region of modern-day Karnataka. When they returned 12 years later, they were naked and their co-religionists were not. The Jains who had gone south, the Digambaras (sky-clad) sect maintained that complete nakedness was a prerequisite of moksha (eternal release) which also meant that you had to be a man since the female nuns did not take up this practice. This matter and some difference in the details of omniscience have at times led to bitter acrimony between the Digambaras and the Svetambaras (cotton or whiteclad). For the social and doctrinal implications of the controversy, see Jaini (1991).

The Svetambaras are much larger in number and include two reform sects. The Sthanakavasis (hall-dwellers) began in the 15th century and, like protestant movements in Europe and India, decried their religious hierarchy's laxity of practice and trappings of worldly power. The icons of the Jinas and the Jain temples that had become such a spectacular and expensive aspect of Jainism were declared inimical to the process of salvation. Building temples was also reviled for the destruction it wreaked on small life forms in the soil. Sthanakavasi monks and nuns wear a cloth over their mouths, similar to a surgical mask, to prevent themselves inadvertently swallowing small airborne beings.

A further Svetambara split in the eighteenth century generated the Terapanthas (path of the thirteen) who advocated ahimsa (non-violence) as total non-interference, to the good or bad, in the lives of other beings, with the exception of giving alms to ascetics. Both reform sects remain small but are well respected and provide a benchmark of good practice that has influenced the wider community. Despite the Terapanthas non-interference, they are today the most socio-politically active Jain sect.

The other noteworthy matter in Jain history is their survival. What with Hinduism's absorptive powers and the waves of Mughal religious intolerance, it is little short of a miracle to still be there with a distinct identity some three millennia later. The reasons for this as identified by Jaini (1979, 1994), boil down to four general principles which in many ways hold true for Jains' socio-political activity today, environmental and otherwise:

1. *Encourage community participation*: Jain leaders quickly developed a spiritual path and texts for a laity who did not want to renounce everything and take up the life of an ascetic. This laity unobtrusively held their ground when the holy people were under fire from Mughal or Brahman militancy.
2. *Build contacts with the establishment*: Jains had considerable success at converting local maharajahs or establishing an influential presence in their courts. They were seldom completely out of fashion or favor. A Jain monk even managed to persuade Akbar, the Mughal emperor (who reigned 1556–1605) to release prisoners and caged birds and to prohibit the killing of animals on certain days ••• Akbar renounced his much loved hunting and restricted the practice of fishing. However, his enthusiasm took a more predictable turn the following year when those orders were extended and disobedience to them was made a capital offence. (Smith, 1917)
3. *Pick your battles*: in times of strong Hindu resurgence that could have swept them away, the Jaina acharyas (spiritual leaders) kept up their philosophical scrapping with the Brahmans but did not do battle with social and cultural mores. They pegged Jainism tightly to a somewhat rarified spiritual practice. This did of course also weaken their role in social change, as for instance in the acceptance of the caste system as the cultural (but not cosmic) order.
4. *Co-opt popular culture for your own message*: Jain acharyas took the hugely popular arts and literature around Rama and Krishna and produced their own revalued Jain versions. Hindu deities were not so much dismissed from the Jain cosmos as demoted.

The Jain philosophy of anekanta (many pointedness) was no doubt a contributing factor in this diplomatic virtuosity and has remained as an ability to look at an issue and find another way of coming at it.

COSMOLOGY AND PHILOSOPHY

For Jains the physical world is not dismissible either as maya (illusion) or the temptations of Satan, nor is it imminently lovable as the creation of a Creator god. It never started, it will never end and it did not require an eternal spirit to think it up. It just is. Emotions such as love and/or fear of nature that many other religions and the environmental movement bring to it are not appropriate. Maitri, friendliness, with its connotations of detached well-wishing; samyaktva, equanimity; abhaya, fearlessness; kshama, forgiveness for harm done; and daya, empathy, are the key words in the recommended relationship with the rest of the living and non-living world. There is in Jainism a strong tone of pragmatism and a wariness of anything that has a whiff of emotion, of losing one's independence in passion over another, animate or inanimate.

Living and Non-living Entities: Jiva and Ajiva

The second century CE Jain scripture, the *Tattvartha Sutra* (Manual of That Which Is', the only scripture accepted by Digambara and Svetambara sects, translated by Tatia, 1994), explores seven categories of truth. The first category is the existence of souls (jiva), the second is the existence of non-sentient entities (ajiva), namely, matter, time, space, the media of movement and of rest. Both categories are equally real and have been mixed together forever with the aid of subtle karmic matter. The Jain word for a living thing – jiva – is the same word as for soul. To save souls, you save life. And this is not the Neo–Platonic world soul but that of a real person. Every living thing has its own insoluble and unique soul, the knower. It is a horizontal rather than vertical ordering, that recognizes the first principle of ecology, the singularity of each being. It reminds us perhaps not to get so tied up in species salvation that we forget that species is a human taxonomy (using procreative compatibility as the primary criterion) to make order out of a world which in fact never repeats itself. Every individual life is a one-off and has value because it matters to the one who has it. It is not a resource. For this reason an environmental program that included selective culling of species would struggle for Jain support.

> Whatever beings there are, whether moving or non-moving you shall not hurt, whether knowingly or unknowingly. All beings desire to live, no one wants to die. Therefore, a nirgrantha refrains from all acts of injury (*Dasavaikalika Sutra*, 4, 11).
>
> translated by (Jaini, 1979, 66)

The *Tattvartha Sutra* offers the most detailed description of the Jain universe, including mathematical calculations and enumeration of the smallest units of time and space and matter, the qualities and cohesion of atoms, the rings of oceans and islands of galactic proportions and quantity that make up the earthly plane, the categories of life forms according to sensory development, the various classes of karmic particles, the plunging and ascending layers of hell and heaven and their denizens, and even descriptions of black gaps in the sky that are described as dense masses of matter which will suck in a passing god and his spaceship should he veer too near. There is to be sure the ultimate goal of spiritual liberation, but liberation is not so very

distant. It is a state of permanent, disembodied omniscient bliss, still within the cosmos and still within a measurable space. Furthermore, for Jains who are not ascetics, such a vaulted ambition is not to do with this lifetime. They will be reborn.

Like the other eastern religions, Jainism sees material reality as a never ending cycle of creation and destruction, so Jains cannot be so easily mobilized by apocalyptic visions of environmental disaster, as can minds marinaded in the judgement day of the middle-eastern religions. They are mobilized by individual suffering.

Non-violence: Ahimsa (The Philosophy)

> There is nothing so small and subtle as the atom nor any element so vast as space. Similarly, there is no quality of soul more subtle than non-violence and no virtue of spirit greater than reverence for life (*Bhagawati Sutra*).
>
> quoted and translated by (Singhvi, 1990, 6)

Ahimsa is the headline and the subtext of Jain philosophy and practice. It is an intention, a state of mind in which one restrains and ultimately eradicates one s passions, namely anger, pride, deceit and greed thereby restraining the violent acts that arise from these passions. Unintentional harm, even if it kills someone, is not violent. This does leave open the possibility of deciding what is harmful. But the scriptural rider to ahimsa is the proscription against laxity.

Although there are beautiful prayers such as the Kshamapana (prayer of forgiveness) from the Jain scriptures that speak of universal friendship and of deep regret for harm to other beings, the sense of Jain ahimsa is more of non-interference than fussing after. We are not here to protect and coddle other beings or mismanage their affairs, but to let them be. In the business of existence, beings are going to get hurt. So action must be taken with this awareness in mind. Non-action may be preferable. In a hyperactive global society, that thought is almost unthinkable. And of course non-action has its own problems when others are taking actions that are harmful. The live and let live approach, which is often the corollary by which Jains explain ahimsa can be laissez-faire.

Independence: Parasparopagraho Jivanam

Parasparopagraho Jivanam

The popular sutra parasparopagraho jivanam (*Tattvartha Sutra* 5.21) is often translated as all life is interdependent, but it does not have the same meaning as in Buddhism, where the idea of an individual self has only a nominal reality. The individual self is very real for the Jains and one is stuck with it for eternity, so interdependence is interdependence. It is not oneness dressed up as twoness. True,

> One who neglects or disregards the existence of earth, air, re, water and vegetation disregards his own existence which is entwined with them.
>
> quoted and translated by (Singhvi, 1990, p 7)

But ultimately we are all alone.

> Man it is you who are your only friend, why do you want a friend other than yourself? (*Acharanga Sutra* 1.1.3.4. A Svetambara scripture).
>
> translated by (Dundas, 1992, 37)

Action and Reaction: Karma

The principle of action and reaction is native to most religions, and to physical science, and the very basis of our environmental troubles, but astonishingly absent from political and economic policy. For Jains, it is fundamental that your actions *will* have results that will affect *you* as well as others. The mechanism of karma is a major preoccupation in the *Tattvartha Sutra*. Following the categories of sentient and non-sentient beings, the next four categories of truth take up the mechanism of karmic particles, their in ux, binding, cessation of in ux, and falling away, respectively. There is communality about the results, although not to the point that we can quietly unload the results of our misdeeds on someone else and slink off unharmed. Independence is at least as important a tenet as interdependence. The instinctively scienti c mind of Jains is also apparent here. Retribution is not a matter of interference from another dimension. Karmic particles of different varieties exist as subtle matter which ow and stick to us according to our thoughts, words and deeds. They are a missing link between what are usually considered two dimensions, spiritual and physical.

Finally, we must note the seventh category of truth, liberation from worldly bondage that the cessation of karmic in ux makes possible. Much of Jain practice is promoted in the scriptures on the basis of personal bene t on the path to spiritual enlightenment (and tolerable reincarnations in the meantime) rather than for ecological or social bene t. But as we shall see, in the socio-political climate of today, a shift in perspective is evident.

Anekantavada: The Doctrine of Many-sidedness

The Jain doctrine of anekanta, literally, not-onesided, states that all knowledge, except the omniscience of a liberated soul, is only partial truth from a particular point of view. Every individual, not just human, has his or her own unique worldview. All are incomplete but valid views of reality with a personal mix of truth and delusion (courtesy of deluding karmas). Anekanta sweeps away anthropocentrism and dogma. To deny the validity of other s views or to see one as total truth leads to the blindness of dogma.

In Jainism, empirical science is one of five types of incomplete – and until there is wisdom – distorted knowledge. Although there is not the space to explore them here, the Varieties of Knowledge, Gateways of Investigation, Philosophical Standpoints and the Sevenfold Predication of Reality (as both permanent and changing) provide the intellectual grist of anekantavada, and are mechanisms for clearing or at least seeing a few blind spots before trying to find solutions (Tatia, 1994, 9–27, 135–140). It produces an instinctive non-exclusivity in the mind that has done much for Jain survival and which Jains bring to any situation, including environmental problems.

Wisdom, Knowledge, Action: Samyag-darsana-jnana-caritra

The doctrine of anekantavada is downstream from the Three Jewels of Jainism in which wisdom comes before knowledge which comes before action and together they take us to enlightenment. Truth or wisdom is not about the number of practical facts that we have under our belt but the values which inform our selection and the use of these facts in our actions. It seems so obvious (and the Jains are not alone in pointing it out) yet more than 2000 years later information technology is tipped to save the world. And knowing for instance that fossil fuels are harmful, has not changed the fact that our homes and businesses are emitting greenhouse gases as if there were no tomorrow. Modern society is living proof of the Jain philosophy that there are many views, much knowledge, but they do not equal wisdom.

PRACTICE

Ahimsa and the Five Vows

Non-violence in Jainism is essentially restraint. Restraint is related to scarcity of resources but the opposite, the seething abundance of the world and the damage that can be done when you start crashing around in a crowded place.

The Jain laity have been given identity and direction by a set of anuvrats, little vows, that are the same principles but with less strict practices than those of the mahavratas, great vows, of the monks and nuns.

> Abstinence from violence, falsehood, stealing, carnality and greed, these are the vows (*Tattvartha Sutra* 7.1).
>
> translated by (Tatia, 1994)

Ahimsa, non-violence, the first of the five Jains vows is described in the scriptures as the fence of the other vows. All are based on restraint. They are a quintessentially passive position, and therefore deeply antithetical, horrifying even, to globalizing Western culture, yet an increasingly necessary component of effective environmental preservation. How far do you go?

Vegetarianism is the start and sometimes the finish of ahimsa for Jains. No flesh or eggs, and there are many other reductions of diet beyond this; all the way to fasting, which tends to occur in the Jain calendar when other people have feasts. Food and sex are the first ports of call, but restraint in movement of oneself and one's goods, in the accumulation of possessions are all prescribed. Stealing is defined as taking anything that is not given, which, in a world where there is no creator god handing out pastoral rights, puts a very tight reign on our use of natural resources.

Ahimsa and the supporting vows challenging the idea that we should express ourselves as widely and noisily as possible. Just because you can do something doesn't mean you should or that it is worth doing. Self-restraint sits next to the Jain injunction for self-reliance.

JAINS AND THE ENVIRONMENT TODAY

There is among many Jains today, especially the youth, the feeling that saving the environment is their call. Contemporary Jain identity is increasingly identified with nature conservation as a social responsibility rather than primarily a path to personal liberation. Our Environment, Our Responsibility' (front page of Ahimsa June–December, 1997).

Despite the unmatched extremes of their ascetics, *per capita*, the Jains would number among the richest religious communities. Their indomitable interest in materiality has generated fabulous arts and crafts, as well as an economically and politically successful community. Their worldly muscle is beyond their numbers. Professionally, they are concentrated in disciplines and arenas which have direct sway on the post-industrial juggernaut. The mechanism that drove Jains into businesses and professions which these days include the motor industry, agricultural commodities, textiles and their associated chemical industries, information and communications technology, pharmaceuticals, diamond cutting, accountancy, politics and publishing, was the ancient prohibition on professions recognized to be harmful to other life such as animal husbandry and its by-products, forestry, hunting, fishing, quarrying, mining, and the military. Even crop farming is a moral minefield for a good Jain. It's not easy to grow things without killing other things. Once upon a time, business was a relatively harmless undertaking. Now the Jains find themselves in an environmental quagmire. Modern business practice and its growth into sustainable environmental practice is where, potentially, they can contribute most.

The articulation of modern Jain identity is exemplified in two publications of the early 1990s, and in the activity they spawned. The publications reveal developments in Jain thought and social organization as well as some of the processes by which environmental activism is taking hold.

The Right Literature

The Jain Declaration on Nature is a small booklet. The Jains were the eighth faith community to produce such a declaration when they joined the World Wide Fund for Nature (WWF) Network on Conservation and Religion in 1990. (The work of that network has since been taken over by the Alliance for Religion and Conservation (ARC)).

The scale of international input from members of all Jain sects including businessmen, religious leaders, scholars and political figures was unmatched by any other faith. This tiny faith community also produced a declaration that had the biggest circulation. The first print run was 5000 and it has been reprinted twice in English and translated and printed in Gujarati and Hindi. It has operated almost as The Jain Declaration, appearing as basic information at non-environmental events such as the exhibition of Jain art at the V&A Museum in London and the 1993 World Parliament of Religions. It provided text and set a tone for other publications.

The presentation of the Declaration and Jain entry into WWF's Network was, for the Jains, an historic event marked by, among other things, the inaugural souvenir edition of the quarterly Ahimsa with articles on conservation from Jains all over the world. Ahimsa has continued to produce news for the international Jain community with a strong environmental emphasis. The existing Jain press in India, including one of its largest media groups, the Times of India, owned by Jains, also picked up the event and recognized its historic importance.

Refocusing

It is by looking at *The Declaration on Nature* and a contemporaneous translation of the almost 2000 year old *Tattvartha Sutra* and its classical commentaries that we see the dynamism of Jainism in response to its environmental grain of sand. Both works were produced under the auspices of the London-based Institute of Jainology with the collaboration of the India-based Bhagwan Mahavir Memorial Samiti. At one level, there is nothing in the lucid ten pages of *The Jain Declaration of Nature* that is not somewhere in the dense 370 pages of *That Which Is: Tattvartha Sutra*. But although nothing has changed, everything has changed. In *Tattvartha Sutra*, the principles exist in a dizzying cosmological order through which the reader is marched for the purpose of understanding how to free oneself from the tenacious grip of diverse karmic particles that block one's spiritual liberation. In *The Jain Declaration on Nature* these principles and practices are stripped down to the core and described not in terms of the individual's spiritual gains but in terms of the good of the whole natural world. The two are not, of course, mutually exclusive, that is the Jain point, but it is a matter of emphasis.

To take an even more specific example, the all important sutra 5.21 of *Tattvartha Sutra*, parasparopagraho jivanam, which was quoted and elaborated upon in the Declaration, and also appears on its dedication page. The literal translation is most closely expressed as Souls impact one another through their will'. In *That Which Is: Tattvartha Sutra*, the translation is Souls render service to one another' (these services are then identified as both beneficial and harmful). The Declaration translates it as All life is bound together by mutual support and interdependence'. It is a subtle shift of language, a slight change of anekanta, but it is a move from accepting the interference of others whether you like it or not, to actually needing them. Again, as anyone who has family knows, these things are not mutually exclusive.

The variants in translation of this maxim provide an example of how environmental issues are realigning and refreshing humanity's various orthodoxies, not just across what had been a growing scientific–religious divide, but between spiritual traditions. To return to our particular example, the shift in emphasis from independence' to interdependence' by Jains and by the modern *zeitgeist* owes at least as much to Buddhism as to the ecological sciences.

Intra-faith and Inter-faith

There is nothing like a shared crisis to bring people together. Like most religions the Jains have managed to fight more vehemently among themselves than with anyone else; the 2300 year old dispute between the Digambaras and the Svetambaras over the requisites for spiritual enlightenment is probably one of the longest running domestic quarrels' in history. *The Jain Declaration of Nature* and the translation of *Tattvartha Sutra* marked an historic collaboration across all Jain sects whose only precedent had been the celebration of the 2500 anniversary of Mahavir's *nirvana*. Jains from all over the world and all sects came to present the Declaration in London and in the process discovered that amongst them they were doing, and could do, more than they realized for conservation. It opened up channels of individual and organizational communication, which continue to function.

For Jains, joining the WWF also offered the possibility of collaborating with other religious communities on environmental programs. As we have seen, tolerance and collaboration with other faiths is Jain, both philosophically and historically. The journal *Ahimsa* has included positive ecological attitudes of other faiths to highlight the shared ground (*Ahimsa*, vol. 7 no. 4, June–Dec. 1997, 8). This touches on one of the time-bombs of our escalating environmental troubles, the scramble for diminishing resources that will breakdown along community lines, which in the first instance is usually religious. Interfaith work becomes part of a preventive program.

The Establishment and Social Change

The Jain Declaration was the entry document into the Network of the largest environmental non-governmental organization in the world whose president, His Royal Highness (HRH) the Duke of Edinburgh, has championed collaboration with religions. The Institute of Jainology (IOJ) was the driving force behind the Jain Declaration. The IOJ's board of trustees, primarily businessmen, called in Jain political figures in India and abroad, the acharyas of the four Jain sects, the only other all-sect Jain organization (based in Bombay and again dominated by businessmen) and international Jain scholars. The main author of the Jain Declaration was India's High Commissioner to Britain through much of the 1990s.

The Jain Declaration was presented to HRH Prince Philip at Buckingham Palace. Prominent Jains from India, Africa, America and the UK came to present it. The event was surrounded by community celebrations and prayers and spun into the first Jain Environment Day organized by Jain youth in London. It was taken very seriously as a commitment to work with WWF. The Declaration was subsequently presented to the Prime Minister of India.

The environmental outreach to Jains and other religions by WWF and more recently by the World Bank's Vice Presidency for Environmentally and Socially Sustainable Development, taps naturally into that hybrid phenomenon of religion as a world institution. The World Bank heard very clearly from the Jains that agricultural practices such as the feeding of vegetarian animals with the pulped offal of their own species and then the animals' slaughter by the millions in Britain when they subsequently developed mad-cows' disease' is not only totally unacceptable but deeply distressing to Jains.

Jain Environmental Action

What does this mean in practical terms?

Vegetarianism

Vegetarianism is, for Jains, the first vow of any true environmentalist. They continue to be tireless proponents of it throughout India. In the UK where 10% of the population are now vegetarian, the Young Indian Vegetarian Association is Jain dominated and they also have an active role in the national Vegetarian Society. In the process of validating and promoting their practice through public campaigning, young Jains are also being educated ecologically; they are learning about the macro-economic choices for agriculture that raise questions not only about what you produce but how and where you produce and distribute it.

From Animal Sanctuaries to Nature Reserve

Jain animal sanctuaries and hospitals for old and sick animals are a feature of India. In a land where human habitats are under threat by overpopulation, it has been hard to think of, let alone, protect the habitats of other beings. So the first Jain nature reserve was in the UK and was one of the earliest endeavors after the presentation of the Declaration. With the advice of the local council's environmental department, it was established at the largest UK Jain center on the outskirts of London with 32 ha of oak trees, some 300 years old, valuable wetland, and many species including the endangered badger. The reserve was opened on the first Jain Environment Day in 1992 organized by the Jain youth and attended by 1000 local Jains. Local schools as well as the Jain community are actively encouraged to visit it. Every summer the youth association works to ensure its vitality. It provides a small example of a shift in Jain understanding of their ahimsa for other beings – habitat matters.

Reforestation

In India and abroad, tree planting is a popular Jain environmental activity both as a sponsorship by the adult community and as something that children do. Schoolchildren at Bahubali Vidyapith in Karnataka have reforested the hillside behind their school.

Holy places are now also part of this activity. The distinctively designed and superbly engineered Jain temples are among the most beautiful and beautifully maintained in India. And they continue to be built on a scale that has not been seen in the West since the days of the great European Cathedrals. They are important centers of community life. With the exception of the Sthankvasi and Terapanths sects, funding a temple is recognized as a karmically meritorious act as well as one which affords status in the community. Palitana in Saurashtra is the site of the largest Jain (Svetambara) temple complex, more a small town, a stunning cluster of white pinnacles which, until not so long ago wandered over the brown hills of Saurashtra with only one tree left. It is said to be where the first Tirthankara of this epoch, Rushabhdeva, gave his first sermon. Reforestation of these hills with native trees was a direct response to joining WWF. For some years, the IOJ has been working with local Jain trusts, the Gujarat government and their foresters, the local community and ARC. Water systems are in place and planting is being done in batches of 1000 trees. The project has included a 9-day environment camp for 200 students and 10 professors who, along with instruction, planted 2000 trees.

Greening of Palitana' is a model and an attempt to direct some devotional energies and resources swirling around temples toward the natural world.

Jain Youth in Rural India

Over several decades, *Veer Seva Dal* (Brave Service League') in the states of Karnataka and Maharashtra has had youth participation numbering in the hundreds of thousands. The activities have been traditionally Jain, including

strong campaigns for vegetarianism among the rural villages and also native tree planting, much needed in a land where basic human needs for fuel are still tied to wood. Over 100 000 trees have been planted. Hillsides and roadsides to Jina shrines have done particularly well.

This pool of committed young villagers is a potential resource that can be tapped into for a number of tasks, including helping to make the transition to solar energy from wood fuel, eliminating the use of pesticides, and other environmentally sustainable village-level projects.

Western Jain Youth Working the Five Vows

Giving up all possessions, most forms of food, curtailing your movements and, when the time comes, fasting to death, is not the sort of stuff that attracts people in droves any more than telling people they have to give up their fossil fuelled, agri-business lifestyle. As the Jains have discovered, there is a fine line between moderating one's extremes and faking it.

The Jain laity have traditionally used their five vows as a basis for evolutions in their lifestyle, not as a great leap into goodness that never happens. The Young Jains group in the UK took up this approach in their dealings with the environment – identifying modern causes of harm to nature that could become part of one's vows of restraint against violence, those tricky negotiations with the self we all face around the use of cars, levels of consumerism, etc. The principle is basic, but you have to build the habit, and start to create a new normality.

Business: The Hardest Nut to Crack

Given that business was the practice that once took Jains away from acts of *ahimsa*, its role in generating environmental harm is not an easy thing for Jains to face and goes straight to the heart of their current life in the world. As noted earlier, Jains have avoided any professions or businesses where there is use, or traditionally recognized destruction, of the bodies of living beings, but knowledge about how we destroy other life is growing daily. The fact that some fertilizers sold to the third world' are bone meal (often banned in the West), has recently come to the attention of Jain business people who are buying agricultural products such as sugar cane and groundnuts for their factories. If it is confirmed that their suppliers are using these fertilizers, then alternative sources or fertilizers will be found. Yet Jains are the diamond merchants of India, they will cut diamonds for jewelry but they will not mine them because of the damage to earth-dwelling life forms caused by mining. Someone else does that. The importance of highlighting interdependence' in Jain thinking becomes clearer. In modern business it is possible to destroy other life without direct contact with their bodies. You just have to destroy their habitat.

The Ahimsa Environment Award was launched under the auspices of the IOJ at the Indian High Commission in London in 1993 and promoted and judged with the assistance of WWF. The Award took the five Jain vows as the basic questions for competitors to answer:

- *Non-violence*: How has your organization/group reduced its production of harmful waste?
- *Truth*: How open is your organization/group about the impact of its practices on the environment?
- *Non-theft*: How much use does your organization/group make of sustainable and non-depleting resources?
- *Restraint*: How much has your organization/group reduced its use of natural resources?
- *Non-possession*: How does your organization/group implement sharing of the Earth's resources?

The winner in the business category, announced at a dinner attended by three hundred UK Jains, was a recycling paper manufacturer and the runner-up was the creator of ecological balanced plant macrocosms for offices. The Award has not continued but, like so much of the work, it opened a door. Especially for the younger generations, business practice is now linked to environmental practice and to the Jain vows. How far the Jains go with that is still to be seen.

The business and political wherewithal of the Jains and their religious teachings that do not need to have environmental principles wrested from an anthropocentric grip, offer this ancient religion an extraordinary opportunity to be a leader in one of the most important challenges of modern life.

> The Jain tradition ••• enthroned the philosophy of ecological harmony and non-violence as its lodestar •••.
>
> (Jain Declaration on Nature)

> Jains should be leading examples of environmentally-friendly citizens.
>
> (Ahimsa Environment Award promotional leaflet)

REFERENCES

Dundas, P (1992) *The Jains*, Routledge, London.

Gandhi, M K (1958) *Collected Works of Mahatma Gandhi* (62 vols), 1958–76. Government of India Publication Division, Delhi, 32.4, 32.4.

Jaini, P S (1979) *The Jaina Path of Puri cation*, University of California Press, Berkeley, CA.

Jaini, P S (1991) *Gender and Salvation: Jaina Debates on the Spiritual Liberation of Women*, University of California Press, CA.

Jaini, P S (1994) The Jaina Faith and its History, in *Umasvati, That Which Is: Tattvartha Sutra*, translated by Tatia, Harper Collins, San Francisco, CA.

Jain, M K (1986) A Demographic Analysis on Jains in India, *India J.*, 21.2, 33–50.

Renou, L (1953) *Religions of Ancient India*, Schochen Books, New York.
Singhvi, L M (1990) *The Jain Declaration on Nature*, Institute of Jainology, London.
Smith, V A (1917) The Jain Teachers of Akbar, in *Essays Presented to Sir R. G. Bhandarkar*, Poona.
Tatia, N (translator) (1994) *Tattvartha Sutra: That Which Is*, Harper Collins, New York.

FURTHER READING

Jacobi, H (1884) *Jaina Sutras*, Oxford (Sacred Books of the East, 22), Reprinted by Dover Publications, New York, 1984.
Umasvati (1994) *That Which Is: Tattvartha Sutra*, translated by Tatia, Nathmal, Harper Collins (& ISLT), San Francisco, CA.

Judaism and the Environment

Aubrey Rose CBE
Barnet, Hertfordshire, UK

Judaism is the religion of the Jews. It has a continuous history of almost 4000 years and, like Hinduism, is not only one of the oldest of faiths, but one which has been the source of other major religions.

Although particularistic in being linked to one people, it is universal in its concepts and beliefs. Its uniqueness is that it has survived whether its adherents dwelt in their natural home, the land of Israel, or whether scattered over the face of the earth.

Due to the history and the multi-millennial experience of the Jewish people, Judaism has relevant teachings that are global, that relate to the environment, as well as to the process of change. In fact because it emphasizes learning and action rather than dogma and theology, it has not been in con ict with developing scientific ideas, such as evolution.

Thus, it has been possible for a leading modern ecologist, Professor David Bellamy, to state:

> *It is indeed a sobering thought that the early writings of the Jewish people encompass all the basic recommendations of the World Conservation Strategy.*

It is likewise no coincidence that individual Jews have played a notable role in the advance of knowledge and of science through the ages, and especially in the last two centuries when Jews were gradually freed from institutional discrimination and persecution that had affected them in the past.

> *The earth is the Lord s and the fullness thereof The world, and they that dwell therein.*
>
> *(Psalm 24)*

These words encapsulate the Jewish approach to the environment.

The sense of wonder at the world around us pervades the history and the writings of the Jews. Not only in the biblical creation story of the Garden of Eden, through the founding fathers of the tribe, Abraham, Isaac and Jacob (*circa* 1800 before current era (BCE)), on to the development of the nation in the time of Moses and Joshua (*circa* 1200 BCE), but throughout the thousands of years of Jewish history.

This amazement at the manifold expressions of nature went hand in hand with a realization that there was an eternal force moving through all things, whom the Hebrews called God, and other names, but, unlike all other peoples, insisted that divine force or being could not be reproduced in any physical form. Thus was the purest form of monotheism born.

With such a basic belief, and a relationship of thanksgiving to the Eternal, it is not surprising that the people developed many teachings as to the preservation of man s home, planet Earth, and all that it contained.

Nor is it surprising that, once centuries of clerical and legal oppression were thrown off, and the modern age of scienti c discovery began, Jews, unhindered by theological dogma, played a notable role in the advance of science, the ever-widening systematic discovery of the nature of God. There was no con ict.

Thus, what follows will endeavor to describe in more detail Jewish teachings and practice which, in relation to the environment, were most advanced. Those teachings were based on a belief that we are the recipients of God s bounty, and that the moral law, as to the sanctity of life and living things, is the inevitable corollary of such basic faith.

Judaism is based on both faith and reason, on the soaring idealism of the biblical prophets and on the practical application of basic teachings by the rabbis, over millennia, adapting to changing circumstances. The expression *rabbi* bears no comparison to priest. It simply means one acknowledged to have a degree of knowledge and learning and, hopefully, wisdom, that enables him (or now her) to carry out teaching, and judicial and pastoral duties.

Whilst there is belief in a future life, Judaism is very much a this-world faith, seeking, through a sense of balance and moderation, justice and reason, to advance mankind s physical and spiritual condition, and thus to become closer to the eternal spirit, in fact as the Hebrew Tikkun Olam indicates, to repair the world.

It is a remarkable story, which, despite emanating from a tiny people, has been a major influence on world history. Hence its approach to the environment is worthy of consideration.

The future of our planet and all in it depends largely on Man's own actions. If those actions are guided by a moral law balancing needs, then the planet can survive. If, however, there is no guiding law, and material exploitation proceeds apace, regardless of consequences, then environmental change will be so drastic that we may well destroy the globe, the environment, and ourselves. Judaism is one of the world faiths that try to bring the moral law into our treatment of our home, our planet.

A brief word, therefore, about Jewish history is necessary to indicate the background to environmental teachings.

Whilst there are various theories as to the origins of the Hebrews, the general view is that they emanated from a tribe based at the confluence of the rivers Tigris and Euphrates (now in southern Iraq), their leader being Abraham, born in Ur of the Chaldees. His wanderings took him eventually to the land of Canaan, a land promised to him and his descendants.

Among those descendants were his son Isaac and grandson Jacob, the latter producing a progeny later identified as the 12 tribes of Israel. The group prospered agriculturally, especially when Jacob's son Joseph held a position of authority in Egypt, where most of the Hebrews dwelt for over 400 years.

Persecution and slavery led to the exodus from Egypt of what had now become a nation in about 1220 BCE under the inspired leadership of Moses and thence to the conquest of the promised land under Moses' successor Joshua. Subsequent government by Judges was followed by the appointment of monarchs, the most notable being David who established a centralized authority in Jerusalem.

By 930 BCE on the death of David's son Solomon, the kingdom split in two, the northern half, Israel, succumbing to the Assyrians in 721 BCE, whilst the southern part, Judea, flourished until 586 BCE, when most of its population were carried off to Babylon.

Before the century was out, however, Jews began to return to Judea, to reestablish the Temple as their national shrine, and to continue their religious life. Greek influence later pervaded the land, although sovereignty was regained after the revolt of the Maccabees in about 140 BCE, though Roman control was established in 63 BCE.

Efforts to restore national freedom continued for decades, exploding in the Jewish revolt of 66–70 Current Era (CE), brutally suppressed by Rome, though flickering again in 132 CE, another vain uprising against the dominant Roman power. Jewish sovereignty in the land was not restored until 1948 with the establishment of the modern state.

Perhaps almost as remarkable as the story in the land itself is the Jewish Diaspora (dispersion of the Jews). The Babylonian community flourished for over 2500 years until 1960, whilst, at the time of the birth of Christianity, Jews were settled throughout Mediterranean lands. In fact those communities enabled the Christian message – Christianity being then a sect of Judaism – to bear fruit amongst pagan peoples.

There was also a Jewish Diaspora in Arabia, as indicated in the story of the Prophet Mohammed, as well as in India. The main Diaspora developed in western Europe in the first millennium, developing extensively from about 1500 in eastern Europe. Similar communities arose in parts of the British and Dutch empires, where religious tolerance reigned, extending to South America, the Caribbean, Australia and parts of Africa.

However, the bulk of world Jewry in 1850 lived in eastern Europe. Russian persecution and pogroms led to a massive movement of Jews, from 1880 to 1914 westwards, resulting in the present communities of western Europe and, especially, the United States. Also, from about 1880 began the Jewish return to the land of Israel, though Jews had always lived there throughout the centuries. This movement accelerated with the advent of Nazism in the 1930s, as well as the movement from Arab lands after 1948.

This continued existence of a people largely separated from its native land for such periods of time has continually perplexed historians, as it appears to have no parallel. Yet the sources of this survival are clear, and have a distinct bearing on the Jewish approach to ecology. Most of Jewish history for the first 2000 years of the people's existence derives from the Bible, confirmed frequently by archaeological discoveries, as well as some writings by non-Jewish sources.

But the last 2000 years has seen an immense burst of Jewish creativity, particularly rabbinic Judaism. The first five books of the Bible, known as the *Torah*, and the teachings of the Prophets, set out, not only beliefs as to God and morality, but also highly detailed instructions as to treatment of the land, animals, trees, an ecological blueprint as David Bellamy recognized.

For example the destruction of fruit trees was forbidden, even in time of war, according to the book of Deuteronomy. This was enlarged upon by rabbinical teachings, first in the Mishnah, a codifying operation completed in about 200 CE as well as in the Talmud, compiled during the subsequent 300 years in Jerusalem and Babylon, containing over 3 000 000 words in 63 books.

From the simple admonition not to cut down fruit trees the rabbis and teachers developed the doctrine of *bal tashchit* – thou shalt not destroy – enlarged on by subsequent teachers to prevent people committing waste, the basis of modern recycling processes. The great Jewish philosopher and teacher, Maimonides (1135–1204), in turn enlarged on this basic principle as follows:

> It is not only forbidden to destroy fruit-bearing trees but whoever breaks vessels, tears clothes, demolishes a building, stops up a fountain or wastes food, in a destructive way, offends against the law of thou shalt not destroy'. Judaism affirms without reservation that the world is God's creation and that whoever helps to preserve it is doing God's work.

Clearly most, if not all, major religions developed in pre-industrial times, when populations were small, and when the land, its ownership and use, dominated society. It could be argued therefore that these religions are not relevant to our present problems. But that would be shortsighted and superficial. What religions did, and certainly what Judaism did, was to lay down certain essential principles, often based on moral values, which could be adapted as society evolved.

It is worth looking at a few fundamental beliefs and values of Judaism to see how they affect environmental practice. The creation story in Genesis is instructive. Our environment, our world, all creation, was the work of God. This labor took six days. On each day God looked at what he had achieved and saw that it was good . After the creation of man in his own image , God saw that it was very good .

This story, be it accepted literally or metaphorically, reflects Jewish optimism that life is good, but also that the unity of God is reflected in the unity of all creation. Hear, O Israel, the Lord our God, the Lord is one, is the famous statement of Jewish belief.

Albert Einstein, in pursuit of that unity, refers to the manifestations of the profoundest reason and the most radiant beauty. Enough for me, the mystery of eternity of life and the inkling of the marvelous structure of reality. He reflects the glowing words of Isaiah and the Psalmist as to the wonders of creation, when he refers to a rapturous amazement at the harmony of natural law, which reveals an intelligence of such superiority that, compared with it, all the systematic thinking and acting of human beings is an utterly insignificant reflection .

Evolution presents no problem. As a modern commentator, Chief Rabbi Dr Hertz, states, there is nothing inherently un-Jewish in the evolutionary conception of the origin and growth of forms of existence from the simple to the complex, and from the lowest to the highest .

Judaism regards man, made in God's image, as being at the apex of this structure, enjoined not merely to subdue nature, but also, as stewards or trustees, to till it and to preserve it (Genesis).

Out of this creation story two important environmental teachings emerge. First the institution of the Sabbath day, a uniquely Jewish idea. On each seventh day not only must human beings rest, but also the land, and all the animals, perhaps the first animal right in history. This idea of rest for the land is in complete accord with modern agricultural theory. It extends also to every seventh year (Shemittah), and to the 50th or Jubilee year. Thus, in a way, Jews recognized that man had a duty to respect the rights of the land and of animals, all part of creation.

The second teaching relates to biodiversity. Again and again in the creation story reference is made to of its own kind – grass, seeds, fruit, trees, sea and land creatures, birds, each after its kind. This acknowledgement of biodiversity is reflected also in the story of Noah and his ark, into which went many species, with a view to procreation and replenishing the earth.

This respect for the existence of species, their right to exist, is the basis of the Jewish view of conservation. There is also reflected concern for the individuality of the species. The Bible sets out in great detail which species may or may not be eaten, which species may not be yoked together, e.g., the ox and the ass, as the stronger will harm the weaker.

Indeed concern for animals is reflected throughout biblical and rabbinical literature. A righteous man has regard for the life of his animal (Proverbs) led on to the Talmudic injunction that a man should feed his animals before himself. The Biblical book of Exodus enjoins one to take care of stray animals; even animals of one's enemy, and especially of injured animals. There is a like prohibition against killing an animal together with its young, to prevent one witnessing the death of the other. Maimonides again:

> The panic of animals under such circumstances is very great. There is no difference in this case between the pain of humans and the pain of other living beings, since the love and tenderness of the mother for her young is not produced by reasoning but by feeling and this faculty exists not only in humans but in most living things.

The Talmud, with its diverse opinions from 2000 contributors, is agreed, however, that unnecessary pain should never be caused to animals or living creatures (*tsa'ar ba'alei chayim*). Thus, Judaism is opposed to fox or big game hunting, bullfighting, any use of animals for sport which causes pain. Also abhorrent is the trapping of animals for luxury furs, the mowing down of elephants for ivory or merciless hunting of whales.

For self-defense or for food, animals may be killed, and *shechita*, the process of ritual slaughter, is an attempt to lessen pain caused to the animal. In fact the ideal diet was intended to be vegetarian (Genesis 1.29), but meat-eating was permitted after the time of Noah, provided the blood, the life of the animal, was excluded before eating. This contrasted vividly with the practices of surrounding peoples.

These few points illustrate the extraordinary respect Judaism has for life, in Hebrew, *chayim*. There is sanctity to life, which must be respected. Thus, in Talmudic times capital punishment was rarely carried out. In modern Israel, the death penalty was abolished except for Nazi criminals. Only one person has been executed in 50 years, the Holocaust murderer Eichmann. Even the strict injunction to

desist from work or other acts on the Sabbath is overruled in the interests of preserving life (*pikuach nefesh*).

Modern Israel has become an ornithologist's delight. There is a mass annual migration of birds, of many varieties, northwards and southwards. These birds are cherished and sanctuaries are built for them. Five hundred species of birds have been observed. In almost all Mediterranean countries, there is a mass slaughter and shooting of these migrating birds. Not so in Israel, the birds are living beings. Whilst culling may at times be necessary, to kill birds as a sport is not part of Judaism or the Jewish tradition.

Perhaps Judaism's view of the environment is best illustrated by its attitude to trees. The tree is a thing in itself as well as a symbol. It appears in the first chapter of Genesis. By the second chapter it has become an allegory, the tree of life. The tree in Genesis is instructive. In 1901 Jews established as one of the oldest environmental bodies, the Jewish National Fund, primarily to plant trees in the arid Holy Land. Every Jewish household worldwide has, since then, had a little box in which pennies, cents and centimes were deposited to raise money for trees.

Since 1945 over 200 000 000 trees have been planted in Israel, to prevent soil erosion, create shade, preserve the water level, each tree, of its kind, planted according to the nature of the soil and climate, pines near Jerusalem, cypress on coastal plains, and eucalyptus and acacia in saline soil in the southern Negev and Arava.

In what in early Talmudic times was a well-wooded country, restoration in the 20th century has resulted in 40 000 ha of natural forest, 40 recreation areas and parks, 700 picnic and recreation sites, while scientific planting in the desert-like Negev, savannahization, is creating a new and hopeful situation that could be of benefit to all desert areas in the world.

Whilst Judaism celebrates both a secular and religious New Year, it also has a special New Year for Trees (*Tu B'Shevat*). As winter departs and spring approaches, this festival is celebrated worldwide by a burst of tree-planting, especially by children.

The extraction of oil from land and seabed was a major environmental feature of the 20th century, with prominent political and economic consequences. There is a view that, in the 21st century, water will replace oil as a dominant issue. Seas and rivers, lakes and streams, have suffered from chemical pollution, a significant global change. World health authorities estimate that 80% of disease arises from impure, affected water.

As with trees, Judaism has a similar view as to water. Its purity is vital. It is also a symbol. The Bible, and the rabbis, constantly refer to living waters (*mayim chayim*). But in actual terms, the same principles apply. Water should not be wasted but conserved. Hence, innovative procedures were developed such as trickle drip system of watering plants, desalinization, irrigation procedures, transportation of water by a water carrier, use of solar energy for heating and hot water, accumulating and diverting flood water through dams, creating a marine nature reserve on the Red Sea coast, as well as a process for recycling water and funding steps to prevent marine pollution.

It is instructive to cite an example of, how with the best of intentions, mistakes can be made. The Huleh lake wetland in Galilee was drained, as other wetlands have been drained. The result was not what was expected. The wetland had operated as a natural filter and cleansing mechanism for the Sea of Galilee, slowing down water movement, allowing nutrients to be used by the wetland species. The drainage resulted in problems, causing silting and algae blooms, as well as a loss of species.

So the whole process had to be reversed, and now the Huleh wetland is again a source of joy, for species, birds and people. Another vital Jewish teaching is thereby illustrated, the interdependence of all phenomena in creation, which we disturb at our peril. There is a typical Talmudic story, or Midrash, of God taking Adam by the hand in the Garden of Eden and saying, Look around you. See what I have given you. Look after it well, because if you destroy it, there will be no one to put it right. This illustrates another Jewish belief that we have free will to choose between good and bad, and, environmentally, between positive change and negative change, between sensible conservation and ruthless destruction.

It is possible that just as the plague of human immunodeficiency virus (HIV)/acquired immunodeficiency syndrome (AIDS) burst on mankind in the second half of the 20th century, multiple chemical sensitivity syndrome (MCS) may become the 21st century affliction. It is estimated that one in one hundred people, especially in industrialized countries, suffer from this disease. It afflicts those who are sensitive to all the impurities poured into the air; petrochemicals, pesticides, air fresheners, sprays for home and farm, disinfectants, fungicides, toxic substances, and other seemingly benign substances that actually add to the greenhouse effect and stratospheric ozone depletion. For an increasing number of people air pollution represents a potent, ever-present threat to their health, if not to life itself. Many are prisoners in their own homes. Researchers have now also discovered an alarming increase in bronchial illness, especially asthma. What we are doing is simply poisoning the air, and ourselves.

The problem, even in pre-industrial times, was considered in the Bible and by the rabbis of old. Deuteronomy enjoined the efficient clearance of excreta. The Talmud even stated that one should not live in a place that has no privy (or even no vegetable garden or flowers). How surprising to find in the book of Leviticus advocacy of substantial green belts (*migrash*) around certain cities, a principle developed by later rabbis to apply to all cities.

Maimonides' description of polluted air over eight hundred years ago, cannot be bettered:

> The quality of urban air compared to the air in the deserts and forests is like thick and turbulent water compared to pure and light water. And this is because in the cities, with their tall buildings and narrow roads, the pollution that comes from the residents, their waste, cadavers, and offal from their cattle, and the stench of their adulterated food, makes the entire air malodorous, turbulent, reeking and thick, and the winds become accordingly so, although no one is aware of it.

Maimonides might almost have had some modern cities in mind.

He continues with advice as to where to live so as to avoid the polluted air. All these biblical and rabbinical provisions emanate from the view that the people should be holy, for which purpose cleanliness of mind, body and environment, was essential. One rabbi commented, physical cleanliness leads to spiritual purity . Hands had to be washed before eating, on rising from bed, and after going to the privy. To emphasize the significance or the soul of each act or sight, a blessing (*Berachah*) was to be intoned. Hence, Judaism is full of blessings, for every possible occasion, the gift of food, the gift of beauty, the gift of wisdom, and the gift of life.

The rabbis of Mishnah times, 1800 years ago, even set out views on limitation of noise pollution. Incidentally, *à propos* the population explosion, a major factor in environmental change, the rabbis stated the biblical injunction (*mitzvah*), be fruitful and multiply was satisfied if parents had at least two children.

It is not surprising that the Bible and the Talmud were so concerned with environmental matters. Land and livestock were the chief forms of wealth. Jews were the people of the land as much as of the Book. Their main festivals, New Year (*Rosh Hashanah*), Passover (*Pesach*), Pentecost (*Shavuot*) and Tabernacles (*Succot*), although overlaid with religious significance, were all basically harvest festivals which encouraged men and women to appreciate the cycle and produce of the natural year.

Also, what Judaism did was to bring moral teachings into each festival so that, for example, the harvester is enjoined to leave part of the land's produce for the poor, the widowed and the orphan. This is a biblical injunction, one of the many commandments (*mitzvot*) of the faith. The Bible, apart from its historical and religious content, is also, in effect, a textbook of the environment.

It contains the most wonderful and detailed descriptions of landscapes, animals, and flowers. The Psalms, Song of Songs, Isaiah, Ruth, are full of environmental allusions, decked out in the most vivid and beautiful language.

The Bible has been an immense help to those administering the 280 Nature Reserves in Israel, guiding them in the restoration of animals to the land such as the oryx, which had long disappeared. By chance, Israel, and Jerusalem itself, are at the meeting point of three climatic zones. In the northern half of the country, with its Mediterranean climate, as Professor Safriel, formerly Chief Scientist of the Nature Reserves Authority points out, not less than 25 species of wild progenitors of food plants live in natural habitats. They constitute an unprecedented rich genetic source; a bank of genes for resistance for use by farmers the world over.

He points out that seeds are normally stored in gene banks, whereas the wild plants in their natural habitats are continually exposed to the changing environment and their general constitution is constantly molded by the forces of natural selection . An echo perhaps of the words of Genesis each of its own kind .

Yet another echo relates to Professor Safriel's description of the vital thin layer of microscopic, unicellular algae that carpets the desert surface of the Negev, the southern half of the land. Professor Safriel compares this to the surface in Sinai where the desert is overstocked with goats and camels who tear off the delicate algal mesh and the desert's soil crust breaks and disappears. Interestingly, it was the rabbis of Talmudic times who warned against the damage done by allowing sheep and goats to feed without control.

Professor Safriel adds that Israel as a whole, being at a global climatic transition zone, has major world significance in monitoring environmental change, particularly global warming due to the greenhouse effect. Who would have thought that, when medieval cartographers placed Jerusalem at the center of the world, they may have had a point!

One benefit from being so sited is that the Hebrew University Botanical Gardens in Jerusalem, in its 24 ha, is able to grow plants, bushes and trees native to the five world continents, a treasure of botanic diversity.

Thus, love of land and its produce has been a golden thread running through Judaism and Jewish life from the earliest times. In this century, Jews have developed arboreta, herb gardens, experimental botanic centers for plants and trees, worldwide, from Exbury Gardens in Britain to a plant resource station in Hong Kong, from remote farms in Nova Scotia, Canada, to the hillsides of Brittany, France, and of course, Biblical gardens in Israel and elsewhere.

The following illustrates this attachment to nature. In the eighteenth century in Eastern Europe, there arose an exciting, exuberant, religious grouping called *Hasidism*. They believed not only in study and learning, but in joyful prayer and song. Their founder was to be seen invariably in fields and forests. In 1760, one of the foremost figures of that group, Rabbi Shneour Zalman, married. He received a substantial dowry. With his wife's agreement, the whole of that dowry was placed in a fund to help

Jews develop agricultural settlements and to learn the arts and crafts connected with farming and the land. Thus, though urbanized, largely through external pressures, Jews have never lost their attachment to nature and the land, rediscovering in recent years, particularly with the impact of the return to Israel, the basis in nature for much of their Jewish heritage.

The land of Israel is today beset with numerous and serious environmental problems. These are largely the result of the sudden influx of people from many lands in the last half century, but there are dedicated people imbued with the spirit of the past who actively pursue the causes of conservation and balance between the needs of nature and of society. For example, the high aims and clear philosophy of the Nature Reserves Authority reflect well the principles enunciated by both Bible and Talmudic rabbis, adapted to the modern scientific world. Those aims are summarized by the Authority as follows:

> All living organisms and natural phenomena have a right to exist – nature is beauty and a fundamental component of our physical and cultural existence – nature is quality of life – nature is the main resource for scientific progress which improved the wellbeing of humanity – human progress depends upon a positive relationship with nature – we must practice the wisest use of nature's richness without diminishing its reserves – protection of genetic diversity of wild species found in Israel could be critical for the future of human agriculture and animal husbandry – we are obligated to preserve this life-support heritage for future generations.

Neither Bible nor rabbis could have expressed the Jewish view better.

There are many examples of Jewish teachings on the environment. Only the main points can be touched on in this review. Much of value, especially detail, is necessarily omitted. There is an understandable and perennial problem for scientists, who deal in provable facts, experimentation and observation, in recognizing the religious impulses enshrined in many faiths. Yet modern quantum physics may be moving closer to the insights expressed or hinted at in the teachings of world religions.

Certainly Judaism is not opposed to science, as witness the disproportionate number of Jewish scientists, provided the essence of humanity, of individuality, is not thereby abused. Reverence for life and creation is essential, as also is humility.

When Job rails against fate, God says to him,

> Where were you when I laid the earth's foundations? Do you know the laws of heaven? Or impose its authority on earth?
> (Ch. 38)

The Jewish view of the world is God-centered, not man-created. In Abraham Joshua Heschel's words, the world is not a gigantic tool-box for the satisfaction of man's needs . Rather does Judaism say, in humility,

> How manifold are Thy works, O Lord. In wisdom has Thou made them all.
> (Psalm 104)

These words were echoed thousands of years later by Albert Einstein when, pursuing the idea of a unified field force, he continued to be amazed at the mystery of life, the marvelous structure of reality, and the search to comprehend a portion, be it never so tiny, of the reason that manifests itself in nature .

Another ethical value in the Jewish outlook is moderation. Judaism, from Genesis and the Ten Commandments onwards, through rabbinical law (*Halacha*), has always imposed behavioral limits on individuals. Moderation implies a self-limitation on consumption and accumulation of wealth or as Isaiah described it adding house to house, eld to eld .

On the contrary, the rabbis of old taught Who is rich? He who is happy with his portion .

Hence the ruthless pursuit of wealth, development that takes no account of ecological cost, of the assault on nature, on forests, seas, land, atmosphere, witnessed in the 20th century, is contrary to the spirit and teachings of Judaism as developed over millennia as society itself evolved.

As the pace of scienti c discovery increases, and may perhaps develop in ways we cannot as yet foresee, so the challenge to Judaism, and all other religious faiths, will grow. There will be evolution, adaptation, advance, yet the basic principles enunciated so long ago by the Hebrew prophets and developed, century after century, by the rabbis, will continue to be the guiding light of Judaism and of the Jewish people.

K

Kelly, Petra

(1947 1992)

Petra Kelly was a founder of the German Green Party and a forceful leader of environmental and eco-feminist causes. Petra Karin Kelly was born in Guenzburg, Bavaria, in what was then West Germany. In 1960, she moved to the United States with her mother and stepfather, US Army Colonel John E Kelly, her biological father having left the family in 1954.

Living in Columbus, Georgia, young Petra became involved in the civil rights movement. She went on to study at American University (BA 1970), and won a Woodrow Wilson National Fellowship, serving as a teaching assistant for one year. Kelly was active in the student movements against the Vietnam War, nuclear energy, and for civil rights, social justice and women's rights. She worked as a volunteer in the presidential campaign of the late Senator Robert F Kennedy.

In 1970, her 10 year old sister, Grace, died of eye cancer. Petra responded by founding, with her grandmother a cancer research support foundation in her sister's name. In 1971, she received an MA in political science from the University of Amsterdam. She worked as a civil servant within the European Community headquarters in Brussels. In 1973, she formally adopted an orphaned Tibetan foster daughter, Nima. In 1979, she helped found the Green Party, committed to principles of non-violence and environmental protection. When in 1980 she became party leader, she was the first woman in German history to head a political party. She held the post of chairperson and speaker for the Green's executive board until 1982. On March 6, 1983, Kelly was among 27 Green Party members elected to the German parliament, the Bundestag. She served in the Bundestag until the election of December 1990, failing to win the requisite percentage of the votes, 5%, to ensure parliamentary representation.

Kelly was honored with the Swedish Parliament's Right Livelihood Award in 1982, and the Peace Prize of the US group *Women s Strike for Peace* in 1983. She was on the international board of the World Women's Congress in 1990–1991, within the Women's Environment and Development Organization.

In the early 1980s, she became involved with a fellow Green Party MP, retired Major General Gert Bastien, 25 years her senior. They were inseparable and a powerful force, jointly calling for accountability for superpowers' possession of nuclear and other weapons of mass destruction, opposing German economic ties with South Africa, organizing tribunals to examine human rights in Tibet, and China's treatment of dissidents. She and Bastien were arrested in 1983 for blockading US military bases.

In October 19, 1992, Kelly and Gert Bastien were found in their home dead, both of gunshot wounds. The Bonn police concluded that Bastien had killed Kelly and then himself. No suicide note was left, and Bastien appeared to have been interrupted, stopping in mid-word in a mundane letter, his electric typewriter still on. This and the unlocked upstairs door have left some to suspect an unknown assailant was responsible. Regardless of the theory, two things are clear. Kelly did not die by her own hand, and the world lost an inspiring and powerful voice.

ELIZABETH MAY *Canada*

Kondratyev Cycles

see Kondratyev, Nikolai Dmitrievich (**Volume 5**)

Kondratyev, Nikolai Dmitrievich

(1892- 1938)

Nikolai Dmitrievich Kondratyev was an outstanding Russian economist of the 20th century. His work is widely recognized internationally, and the term Kondratyev Long Wave is named after him. (In the US, for example, he became a member of the American Academy of Social Sciences, the American Economic Association, The American Statistical Society, and the American Association on Agricultural Economic Problems.)

Nikolai Kondratyev conducted fundamental investigations into the theory of long-wave economic dynamics; he is the author of works on predicting and long-term planning of economic development and on mathematical statistics and agrarian sector development. The idea of economic long-wave cycles was based on Kondratyev's analysis of a large body of statistical data on the dynamics of the average level of commodity prices, capital interest, wages, foreign trade turnover, and the extraction of cast iron and lead, covering a • 140-year period. Kondratyev not only predicted the great world depression of the late 1920s–early 1930s but he also believed in its inevitability.

Orthodox Marxists attacked Kondratyev's concepts and he was persecuted for his beliefs. He died in one of Stalin's prison camps but he was finally rehabilitated posthumously in the Russian restructuring period.

LIUDMILA ROMANIUK *Russia*

L

Land Ethic

see Environmental Ethics (**Volume 5**)

Landscape Garden

see Landscape, Urban Landscape, and the Human Shaping of the Environment (**Volume 5**)

Landscape, Urban Landscape, and the Human Shaping of the Environment

Brenda Lee
Montréal, Quebec, Canada

For ve centuries, landscape has been used to describe domesticated nature (historically, the British sense), the apparently untouched wild (the American sense) and the pictoral representation of both. It is a complex concept within which the natural and cultural, visual and functional, real and reproduced, interpretive and descriptive, bounded and boundless have been very tightly interwoven. As a technical term, landscape is believed to have been rst used to relate natural and cultural elements in landform by Carl Sauer Morphology of Landscape (Sauer, 1925). In the last two decades its capacity to connect has been exploited by an increasingly wide range of historical, cultural and feminist geographers, and by researchers outside the geographic eld.

Sympathetic branches of archeology, cultural anthropology, culture studies, literary criticism, sociology, philosophy and art history, to link cultural perception of place within society and nature, to the forms that comprise the physical world. Next to place, landscape has become the most important medium for exploring culture - nature interactions. Metaphorically, landscape has replaced context to portray the sweep of signi cant features that meets a focused gaze. Landmarks of its modern use might begin with J B Jackson's founding of Landscape Magazine *in 1951, and include Hoskins (1958); Nairn (1965); Lynch (1972); Williams (1973); Meinig (1979); Watkins (1982); and Cosgrove and Daniels (1988).*

In the natural sciences, landscape ecology developed from the recognition that the proper study of ecology must include Humankind's very signi cant interventions. Applied ecology, although de ned much earlier in the German scienti c literature, was given form in Europe and North America in the early 1930s; it wasn't until the early 1980s, however, that the term landscape was widely accepted as a functionally bounded unit of those patchy ecosystems that Humankind's activities produce (e.g., see Tjallingii and de Veer, 1982, and for an overview of the consolidating disciplines, Schreiber, 1990 and Forman, 1990). The International Association for Landscape Ecology was founded in 1982 to unite the increasingly proli c European and North American work in the biological, geological and geographic sciences. This work has selective links to, but is not integrated with, related activities elsewhere in the humanities and social sciences.

LAND INTO LANDSCAPE: THE MEANING OF THE TERM

The meaning of landscape has therefore never been more complex. Its ambiguity was marked by Meinig (1979) through an identification of 10 of the scenes that it was possible for the scanning eye to behold: nature (the skin of mother earth); habitat (man's domesticated home on earth); artifact (evidence of man's intention); system (of interconnected elements, hydrological to highway); problem (condition of deterioration demanding remedial action); wealth (via a financial appraisal of its worth, existing and value-added); ideology (symbol of the values, the governing ideas, the underlying philosophies of a culture); history (the complex cumulative record of the work of nature and

man in this particular place); place (the unique constellation of characteristics and experiences of each locale); and aesthetic (function subordinated to the painterly laws of intention, form, light and balance).

Two versions of its original and essential meaning exist. The first reaches back beyond the first English usage of the Dutch painters' technical term *landschap* (*landskip* recorded in 1598; the correct *land-scape* in 1603) to the Old English *landscipe*, interpreted as the district or region inhabited and worked by a particular society of people. J B Jackson anchors the old English term in the original neolithic relation of land and its first farming populations. The Northern European Germanic word *land* referred to a clearing and *scape* to a collection of like objects. Similarly, in the southern regions, the Latin *pagus* or *pact*, established the essential socio-political nature of the derivative in modern Romance languages. Landscape was the worked territory of farming peoples. Perceptual definitions are, importantly, derivative.

The second account places more significance on the break in usage (no record exists of the term in middle English) and the cultural development of self that accompanied the exploration of nature as a significant subject of art. Since antiquity, nature had played important decorative, iconographic and supportive roles for the mind, keyed to various symbolic associations; and to medieval and early Renaissance minds, nature was a key medium of Christian symbolism and allegory. It was painters from Switzerland (Konrad Witz – 1400–1447 AD – being the first fully to embed his figures within a topographically identifiable countryside), the Low Countries (Joachim Patinir, Hubert and Jan Van Eyck, Pieter Brueghel the Elder (1400s–1600s), Italy (Giovanni Bellini, Giorgione – Brunelleschi having earlier worked out scientific perspective – 1460s–1500s) and Germany (Albrecht Durer, Albrecht Altdorfer (1500s) who constructed the bases in spatial form and light for the first landscapes that were painted for the value that they had themselves: the 17th century Dutch explorations of nature-based reality that gave interpretive primacy to the individual experiencing subject. This was a time when the new perspective allowed a more realistic and enigmatic nature to emerge on paper, and when the new Renaissance fascination with things-in-themselves began to enliven nature, human and wild. This new nature had a life of its own, and the tension between emblematic projection and expressive interpretation characterizes all early landscape art. The transition from emblem to expression – i.e., from a landscape mirroring cultural associations to that evocative of the personal experience of nature, has been often explored in theatre (Zimbardo, 1986), poetry (Wasserman, 1968), literature (Abrams, 1958) and landscape design (Hunt, 1992). Landscape here has become, in Malcom Andrews' words, worldscape (1999): the myriad of compositions that can be formed from interacting natural and cultural lives-in-themselves. From political to scientific, each picks out its field of significant objects and relationships.

While it is difficult not to use the term *landscape* for all historical treatments of humanized nature, this second understanding is founded in the belief that the concept's modern English formulation accompanies, and depends upon, the individuation and internalization of the analytic Western self. Landscape is, in this sense, essentially a perceived composition of meaningful objects. It is landscape that most accurately represents Cooper's (1992) definition of environment as a field of significance. Landscape, which encompasses the changing patterns of sky, land use and seasonal weather and vegetation, is thus a momentary individual phenomenon, which, like the connoted word, is generalized to a series of widely accepted denotations. This second understanding is sometimes portrayed as purely visual or pictorial; but the aesthetics of land is argued to be much more active, supporting bodily orientation and the related phenomenological experience of being expressly here; doing expressly this (Berleant, 1997; Cooney, 1999) in some space.

The meaning of landscape to those professions dedicated to its design and management has a particularity of its own.

For practitioners of landscape architecture, landscape is narrative: a historically produced composition of biogeochemical forces and the artifacts of human use, which cannot properly be a site of possible intervention until it is understood (see Spirn, 1998). In practice, since conflicts must be resolved and indeterminancies settled in making a design or management plan, only the most practicable outputs of the natural and social landscape sciences can be drawn upon.

For environmental managers, the historical narrative exists, but is normally superseded by the structural detail, which is so difficult to integrate into landscape design: i.e., the characteristics and interactions of its composite units. These units are typically the bounded areas of interest that form patterns within a landscape. Landscape to an environmental or ecological manager is thus the largest functional unit within which smaller sub-units operate, both horizontally and vertically: i.e., an ecosystem. Its study as an (open) system has, since the early 1950s – and more actively, since the early 1980s – gathered together the disciplines previously isolated in specialized studies of ecosystem components (air/weather, surface/subsurface, rock, soil, plants, animals, etc.) so as to assess comprehensively their reactions to, and effects on, human action (see for example, Williams, 1993). *See also* Landscape Ecology, Volume 2.

CHANGING HUMAN PATTERNS ON THE LAND: THE EARLIEST CULTURES

If we move to the earth's landscape itself, the earliest human effects can be traced to the African Old World

tropical forest, whose primates are thought to have given rise to the hominids that first moved out into the grassland and savanna. Through three major glacial periods (500 000–8000 BC) Paleolithic hunting societies survived in small encampments, of which evidence has been found in the north African coastal countries, the east Mediterranean, Italy, France and the British Isles. With the recession of the Wisconsin glaciation, these societies advanced out into the temperate forests of every continent. Their general patterns of movement and settlement (population at that time of around five million) are interpretable through pollen analysis and the relative locations of settlement remains. Those unearthed are small, temporary shelters, the oldest of which lie within the French Terra Amata beach camp (400 000 BC): a grouping of oval huts, sized to hold about 15 people, formed of branches or small trees supported by exterior rings of stones. Palaeolithic sites in general show evidence of hearths (fire is thought to predate the period), wooden spears, hand axes and the separation of living and ritual quarters. It is the paintings in the ritual caves of France and northern Spain that appear to mediate the powerful spiritual relationship between the earliest human societies and their world, where nature appears as a force rather than a place itself.

The agricultural revolution producing the Neolithic landscape (8000–4000 BC) introduced the dynamic force of the human reorientation of natural purpose. Not only were wild grains cultivated (wheat and barley); but both plants and animals (dogs, goats, sheep and pigs) were domesticated – i.e., they were genetically selected wittingly or unwittingly, and thus induced to develop in ways that depended upon human care and will. Control of nature was here being taken into human hands. Originating in the area of the Mesopotamian river valleys, cultivation and domestication were extended west along the shorelines of Mediterranean and Atlantic Europe, changing forest coverage patterns through clearing, exhausting soils and simplifying ecosystems through substitution of the myriad forest and grassland species. Irrigation, which may have started in Egypt about 3050 BC, or perhaps earlier in Iraq, allowed intensification of both agriculture and settlement (development of the hydraulic civilizations of the Tigris, Euphrates and Indus valleys) and brought further changes in the form of salinization, a primary contributing factor in the Sumerian abandonment of land and settlements that peaked about 1700 BC.

In terms of built form, the oldest known Neolithic settlement is Jericho (7500 BC); the largest and most complex, Turkey's Catalhoyuk (6000 BC), which was fed by the new urban function of trade. These are both highly compacted sites, of mud-filled timber frame construction, which turn their backs on the landscape, opening inward and upward. The most haunting remnants however are the Neolithic ritual structures – sacred monuments, temples and tombs – the best known of which is Stonehenge (c. 2000 BC). Built and rebuilt over a millennium, its purpose may have been to understand and celebrate the astronomical geometry of the sky (but this remains under active debate).

Another expression of landscape was perfected in the Egyptian civilization (4000–100 BC) taking its cue from the forms and rhythms of the Nile: 800 km of fertile plain bisected by the river and bounded by desertified grassland. Memorial structures paralleled the river course, the pyramids in Upper Egypt and the temples, incised into the Lower Egypt mountains themselves, to the north. Settlements, although similarly orthogonally disposed, were clearly subsidiary to the monumental landscape, constructed to memorialize what became after the unification of Upper and Lower Egypt (about 3000 BC) the absolute ruler: in the Old Kingdom he was the son of Re, the Sun God; in the Middle Kingdom, he became God himself (Re and the transcendent Amon).

ANTIQUITY

When we turn to Greek culture (a primary root of Western culture), we find the relation between attitudes to nature and patterns of settlement and exploitation to be less clear. Aristotelian *ontology*, or the study of the nature of existence, places Man and his endeavours firmly within the natural realm. On the one hand, Nature was, in general, the theatre of the Gods; and the notion of a pastoral retreat, although immortalized in the Roman poet Virgil's *Eclogues* and *Georgics*, actually originated with the Greek school founded in Alexandria by Theocritus (c. 300–260 BC). On the other hand, the various understandings of civilized human advancement were each tied in some way to the control of natural forces. Greek cities had no treatment of waste; and its agriculture – goats and sheep, vineyards, olive groves and cereal crops – was notoriously hard on the soil cover, particularly on the more ecologically fragile mountain slopes. Coates (1998) remarks on Plato's (unheeded) account of the environmental effects of clearing and overgrazing in his dialogue *Critias*, likening the bared Attic mountainsides to the emaciated form of a once flourishing man.

Coates also points out that, in any case, perception of the natural world for the Greeks was not as a place or locale but a constellation of forces, albeit with human personalities. It was these forces and their biogeological effects that supplied some of the most important design determinants for Greek buildings. City design began as an intuitive response to distinct land formations, the practical art of city building not formally constituted until the Roman technical writings of Pliny, Frontinus, Varus and Vitruvius. In contrast to the static sculptural forms of Egypt, early Greek city building was organic, asymmetrical and

sensitive to the perceptual rhythms of sequential colonnaded space.

The late or Hellenistic Greek city-states formed a foundation for Roman reconstruction in the second century BC. In contrast to the Greeks, the Romans were engineers and administrators, intent on extending Roman law over the colonizable world. This was accomplished through construction of the renowned roads, bridges, aqueducts and lines of fortification, the orthogonal division of land into farming units and the founding of settlements also along orthogonal grids tied most often to shoreline geometry. By the third century AD, Rome had stretched its imperial infrastructure completely around the Mediterranean Sea, north to Scotland, and east to Trabzon on the southeast coast of the Black Sea, founding, apart from most of the important Italian cities – Paris, Nimes, Trier, Lyons, London, Vienna and Cologne.

Roman law made manifest the divine will. The monumental impulse was, as a result, foundational to Imperial construction, which extended the reach of its built form through development of the arch. From the enormous cadenced aqueducts to the massive amphitheatres, stadiums and baths, Roman city design compressed its components into axial routes, moving large quantities of earth to do so. As Coates (1998) again notes, it was a resource-intensive civilization: the Roman baths alone must have consumed enormous tracts of forest. At its most extensive, the Empire reduced the Mediterranean temperate forest to the upper slopes of much of the Middle East and North Africa, stripping the soil of its fertility and flood control capacity as well. Agricultural land was further reclaimed through large-scale drainage of marshlands in central Italy and the Tiber estuary. With an AD 250 population of 1.2 million, Rome itself was a significant source of human and animal waste. The environmental contribution to the Empire's disintegration, although not traditionally recognized, was argued as early as 1864 by George Perkins Marsh to be significant (Marsh, 1970). *See also* **Marsh, George Perkins**, Volume 3.

Although Imperial structures were designed to dominate the land completely, the Jellico and Jellico (1975) point out that design of the smaller places such as the delicately poised Temple of Vesta at Tivoli (205 AD) demonstrated a capacity to balance sensitively the natural and human orders. This capacity was exploited in the Roman gardens, which expanded in importance with the size and luxury of the Roman villa, and the public parks that lined the Tiber.

MEDIEVAL LANDSCAPES, THE ENGLISH EXAMPLE

The disintegration of the Roman Empire into Germanic states in the West and the various components of the Byzantine Empire in the East was accompanied by a decay of its vast urbanized network and of the cities it had fed. Apart from the few that remained as active trading centers, such as London and Milan, and the founding of Rome s Christianized capital, Constantinople (330 AD), most cities disappeared as vital centres. This occurred through losses in population (famine, disease, lowered birth rates) and desertion to the rural areas, where increasingly autocratic estates could provide the protection and livelihood that the Empire under siege could not. In contrast to the classical, the medieval landscape was therefore essentially a working countryside.

Its pre- and post-Roman plant composition and coverage, previously thought to be substantially native, is now argued to have been impacted by intense competition for all land types woodland, peat, forage pasture amongst the peasantry, and between these and the hunting landed gentry. By 1086, 85% of England s native woodland and wood pasture was gone, and many of its large indigenous mammals. Population density then varied by county but, for example, it averaged 13.5 per 100 hectares in the ve east midland counties assessed by the Medieval Settlement Research Group (Lewis *et al.*, 1997). Expansion of cultivation onto the poorer soils subsequently speeded land conversion to the unwooded pasture associated later with England s rolling hills.

The earliest and structurally most important forti cations were the monastic colonies. It was here that the ordered life now tted to the tasks of the new spirituality was maintained until it passed to the urbanizing settlements, complete with clock, account book and structured day (Mumford, 1961).

The later forti ed estates and towns were organized to support the structures and routines of feudal privilege (defense, industry, trade and capital accumulation); at the same time, the unforti ed towns, villages, hamlets, farm clusters and individual farmsteads were organized to meet the needs of the agricultural sector. The commonly imagined medieval town was walled off, having fractured and in lled the Roman grid to create a tight maze of narrow lanes bounded by unbroken building walls, relieved by irregular green spaces, dominated by town and/or guild hall, gothic cathedral, church or chapel, and perhaps a palace or castle. New towns and additions to old ones were normally built along a exible grid, however, and unwalled settlements were often arranged around a broad delineated eld system, communal grazing area or cattle route to upland pasture. Settlement, in general, expanded with the 12th century feudal revolution and died back in the 14th century following the plague (1348 1349) and its recurrences, associated labour shortages, the deterioration of particular soil types that forced a switch from grain cultivation to pasture, and the conversion of whole villages and their eld systems to enclosed pasture (Lewis *et al.*, 1997).

RENAISSANCE LANDSCAPES: ITALIAN AND FRENCH CLASSICISM

A central aspect of modernity is the secularization of many of the elements of the medieval religious framework. The forces that led to this include:

- the spread of a proto-market economy in the 15th century, its move to the outer edges of towns, specialization and increasing trade across Europe, Asia and the Americas;
- the development of the new science, particularly the 16th century challenges to the Scholastics' hold on truth;
- the relocation of divine spirit within the laws and manifest reality of the natural world itself;
- the insight that the inner core of the individual was a realm that was not fixed, but was worthy of being explored;
- the challenge presented by Copernicus and the discovery of the New World putting all previous knowledge in doubt, and fostering a stance of scepticism.

Before the widespread use of the efficient manufacturing technologies that this new philosophy unleashed, the great releases and achievements of the Renaissance remained within its arts. Among the most important of these was the secular garden, the product of a revival of the classical appreciation of nature, classical building form (principally via Alberti) and the increasing attractiveness to the Florentine and Roman aristocracy of the rural retreat. This retreat was enclosed, as was its medieval predecessor, but, since it was set into the Italian hillside, it necessarily offered spectacular views into the countryside beyond.

Of the most accomplished, Alberti's Villa Quaraccchi, Tribolo's Villa Medici and Boboli Gardens, Raphael's Villa Madama, Ligorio's Villa d'Este, Vignola's Villa Lante, the Villa Bernadini, Santi di Tito's Villa Bombicci, and the Villa Gamberaia, the most complicated might be the 1652 Baroque Villa Garzioni, in Collodi, Tuscany. Its seven levels connect a small open-air theatre in the rear, down a sloped parterre, through a series of stair-connected landings, to a central heraldic tapestry of boxed hedge and reflecting pools. What was newly valued here was the relevance of the Roman pastoral ideal: the belief that exposure to the countryside could be a powerful urban antidote, a source of spiritual and aesthetic renewal. And of course from here, the commodification of landscape was just a short step away.

In France, Andre le Notre (1613–1700) (amongst many others) brought the villa garden to its ultimate expression for Louis XIV in Versailles, by wrapping it around the residence and extending its constellations of grand avenues out through the forest to form a great imperial landscape. Views within this extraordinarily complex composition are directed infinitely outward through the use of the long penetrating allee and the recapitulation of sky in its numerous reflecting pools. As Eisenberg (1998) reminds us, this was a landscape in which power was expressed – not vertically as with medieval castles and cathedrals – but horizontally, outward with the expansive aspirations of its Sun King. Its astronomical cost, in materials, water provision for its pools, canals and baroque fountains (a full display consumed more water daily than the 600 000 residents of Paris), potted plants, trees, hedges and the labour for installation and upkeep, not to mention in human lives lost in construction, aggravated publicly that of upkeep of the court. It contributed to the subsequent shift in emphasis to those landscapes which nature could (largely) maintain herself.

THE ENGLISH LANDSCAPE GARDEN

The deepening and extension of the pastoral ideal was accomplished in Britain. The plea for nature as the wellspring of human freedom was first made at the turn of the 18th century by Addison, Shaftesbury, Pope and Switzer, although a garden that would follow nature's irregularity was described as early as 1685 (Sir William Temple's *Garden of Epicurus*). Characterizing the transition from Renaissance formalism to the point at which the English country house was unmoored from its axial attachments and set aloft in a sea of grass, amid archipelagos of wood (Eisenberg, 1998) is Sir John Vanbrugh's Castle Howard (1701–). Himself a set designer, Vanbrugh's achievement was to work out the spatial arrangement of the Palladian redesigned medieval castle and its 16th and 17th century grounds to incorporate the principles and literary references just retrieved from the classical tradition. He alluded as well to his own Gothic past. What he produced was a carefully choreographed blend of ancient, medieval and new: a rolling English meadow partitioned with lake and woodland marked by Gothic references (gates, walls) that recontextualized a series of classically referenced architectural forms (temple, bridge, obelisk, pyramid). In 1770, Thomas Whately wrote that it must be examined, compared, perhaps explained, before the whole design ••• is well understood (Hunt, 1992).

This was in fact Hunt's comment on the landscape at Stowe, the naturalization of which by William Kent (c. 1735) produced one of the most influential (and best maintained) examples of emblematic landscape. Kent's Temple of Ancient Virtue literally engages in dialogue with the Temple of British Worthies across the dip of a lagoon bisecting his Elysian Fields, the grounds overall containing over 30 such referential structures.

Hunt places the transition to a garden that evokes an interior personal, rather than an exterior social, reaction in Kent's Grecian Valley at Rousham (1738). He notes: the subtle varieties of the valley afford a landscape that seems

to answer our moods, that allows a unique and individual response by each visitor to its unobtrusive character (Hunt, 1992). We are here at the genesis of modern landscape. Its arrangement of visual elements and evocative depths are traditionally believed to have been drawn from the paintings of Claude, Gaspard and Salvator Rosa; these may have been more influential in the literary picturesque, however. For spatial structure, argument is made for Renaissance stage design, particularly the court masque scenery of Inigo Jones. The elements of Kent's expressive landscapes were codified by Lancelot (Capability) Brown, whose popularity gave him a tremendous amount of English land to clear, carve, plant and irrigate in his characteristic pastoral mould.

Studies over the past two decades in English art/architectural history and historical geography have pointed to the impacts of the enclosure movement that privatized these grand estates. By transforming the working countryside into genteel recreational grounds, blocking views of distracting functions and aestheticizing the poverty of those remaining on or visible within its frame, those contributing to the picturesque movement not only appropriated land for landscape, but created a universe that eliminated and/or instrumentalized its rightful historic users and use. The shift in ownership also created increasingly concentrated and intensified farms, feeding a growing population and sending increasing numbers of the formerly landed peasantry, into the cities. The pastoral ideal contributes independently of these very real effects, however, and it is possible also to say that an eminently debatable Romantic, expressive view of nature was here taking hold. As Hunt (1992) points out, the picturesque tradition of these designers gave each a frame for admitting a particular amount and form of educated wilderness into the domestic country scene. It is this tradition that has structured landscape design ever since. (See later section on Parks Movement.)

THE EUROPEANIZATION OF NORTH AMERICA

It is now understood that just as the Neolithic agricultural landscape was more domestic than wild, that of the pre-Columbian New World contained complex cultures interacting in equally complex ways with their own native environments. North American ecologists are moving toward their Scandinavian counterparts in the rejection of wild or primitive ecosystems altogether and their substitution by the concept of gradients of human impact (Williams, 1993). Estimates of Amerindian densities in the early 17th century have been raised from sparse to those approaching Western Europe in the American east coast forests. Each Amerindian Nation developed its own means of controlling and exploiting resources: cereal crops and corn were extensively cultivated, fish and game harvested; lines of communication and trade established; territory (and its patterns of use) taken, lost and recovered. Human use is suggested to have produced a landscape mosaic prior to European settlement. The fire used to produce the clearings and shoots that attracted game, and to keep open cultivated land, for example, may have produced much of the midwestern prairie; the fish and game harvested may have reduced species diversity, and certainly limited populations of large game such as elk and buffalo. Relations with nature have been described as a matter of contractual agreement based on pragmatism, fear, and respect born of self-interest (Coates, 1998); but it is clear that they were also tuned to a life within nature for which the West had no appropriate theoretical model.

European conquest replaced indigenous common property conventions with a system of privatization and exploitation geared to commodity production and trade. It famously brought disease, a change in religious orientation, dispossession, and relocation; and environmentally, much more timber clearing, more intensive and destructive patterns of cultivation, much larger permanent settlements and a significant upgrading of resource use. It was depletion that pushed the fur trade – originating along Quebec's Saguenay and St Lawrence routes through the Great Lakes, for example – west to the Rockies. Replacement of fur in the early to mid 19th century with the timber and charcoal industries stripped the US of half its forest cover by the 1920s. Fire further introduced large populations of aspen and jack pine into the then simplified native forests, particularly around the Great Lakes and areas of the Rocky Mountains.

Urbanization began in port towns along the Atlantic and St Lawrence coastlines and remained a port phenomenon until its westward expansion in the 19th century. Until then, people actually moved away from the urban centres to small resource-based communities and farms. Urban society was limited to the few large port/administrative cities in both Canada and the US; if rural, it remained a thin elite layer over discrete areas of countryside. Into the second third of the 19th century, just 8% of Americans and 3% of Canadians lived in towns over 5000. It was not until the colonial economy was superseded by factory production in the last half of the century that significant movement from country to city began, and its specialization at the turn of the century, that the cities began to explode. By 1930 in the US and 1960 in Canada, the urban–rural ratio was 50–50; by 1990 in both countries it was just over 70%.

That life took place within a rural landscape for most North Americans until some time in the mid-20th century did not prevent the shaping of both cultures by views of nature that were strongly idealized. In fact, both nations were deeply split. On the one hand, there was the pastoralism identified with nation in Britain and the US, and the wilderness that turned much of the early Canadian culture inward. The ideal of nature as a cleansing source, as a

retreat from urban ills, is, in all Western cultures, itself deeply ambiguous: it is at once a source of personal/national moral strength and of the elite exclusionism discussed in the previous section (see Buell, 1995 for its American implications).

Practically speaking, its rejuvenative potential was the source of the early Parks Movement (see later section on Parks/Garden Cities Movements); its economic potential was mined unabashedly by the railway industry on both sides of the border in the last quarter of the 19th century, to promote the touristic picturesque. On the other hand, there was another Nature, the Nature that (literally) fed colonial mercantile exploitation and the vigorous industrial capitalism that was established when mercantile structures finally gave way. Here, the dignity of humankind lay in its capacity for control; and John Locke's interpretation of Nature provided the (God) given resources with which to begin.

19TH CENTURY INDUSTRIALIZATION

Industrialization transformed agriculturally based economies and landscapes to economies and landscapes devoted predominantly to resource extraction, refining and manufacture: a development with massive global consequences, both rural and urban. In Britain (the first nation to undergo industrialization), as across Europe, the textile industry, coal mining, and pig iron production, were intensified through introduction of the mechanized loom, the replacement of water with steam power and large-scale iron foundries, thereby tripling growth in industrial towns in the first half of the 19th century.

Factories, coexisting with home manufacture, took over in the last half of the century as the available technology upgraded expectations of what kinds of output could be forthcoming in an increasingly dynamic capitalism. Factories were redesigned to transmit power and reduce the ever-present hazard of fire; they were iron-framed, initially to support the heavy machinery used to spin flax; and made vastly larger and more complex as the scale of operations increased and processes became more diverse and dangerous. Concrete was introduced toward the second half of the century, first in ports and then, as it was cheaper than iron-frame, into construction of the factory buildings themselves. Factory complexes walled in vast courtyards around their central power source(s) creating environments of noise, heat, dust, smoke, refuse and noxious by-products.

Power production depended on, and tended to concentrate around, the coal mines (themselves centres of human and environmental devastation). All this was integrated and increasingly concentrated, first by the systems of engineered canals and then by the new networks of railroads. While there clearly existed many industrial landscapes, those prior to the 1830s stimulated more fascination and awe than disgust; while those during and after the 1830s could, for the vast majority of residents, be characterized by Mumford's phrase, factory, railroad and slum . The pattern of New World industrialization tended to be led by resource extraction, but the industrial cities on all continents were increasingly tooled to produce and distribute the products intended to increase quality of life: the urban environment of production, now severed from that of consumption, was thus unserviced, unrelieved, bleak, filthy and crowded beyond description.

THE PARKS/GARDEN CITIES MOVEMENTS

The appalling condition of 19th century western industrial cities fed the view that urban life, if productive of a nation's finest culture, was in principle bad for the soul and destructive to one's health. Reform was pushed from the bottom up, by women's societies, church groups, physicians, architects, landscape architects, engineers and town planners, particularly in Germany, Britain and the US. Following the 1893 Chicago World's Fair, impetus from North American sources committed to design and civic improvement coalesced into a *city beautiful movement*, producing European inspired designs to ennoble the city's core. These featured the classically designed, monumentally scaled institutions and urban components that industrial cities lacked: grand treed avenues, plazas and parks, parliamentary buildings and city halls, colleges, museums, libraries, etc. What was sought was coherence, visual variety and civic grandeur (Van Noos, 1977). However, the state of the city's basic sanitary, service and transportation infrastructure, fragmentary, often privately (and terribly) run or missing altogether – and of its housing stock – was a concern as early as in the 1860s.

A key component that predated the beautification movement and survived its destruction was the redemptive insertion of Nature into the city as parkland. Its most powerful North American advocate, landscape architect Frederick Law Olmstead, was concerned primarily with integrating the park fully into the large-scale town plans then being initiated and he developed many of the techniques by which this could be done. His designs (along with his partner Calvert Vaux), among them New York's Central Park (his first, in 1858), Brooklyn's Prospect Park, Washington's Capitol grounds, the park systems of Buffalo and Boston, and Montreal's mountain park, drew explicitly on England's pastoral garden design: landform and planting were arranged to isolate the park from the city and to create experiences both of enclosure and of the vastness of nature itself.

Water was wound through meadow and forest; texture was increased through insertion of rocky outcrops and planting variety; experiential variety and surprise were,

overall, inserted into a serpentine formation designed to maximize tranquility. Facilitating this was the separation of vehicular from pedestrian traffic, the roads crossing the park sunk below, thus protected from, the level traversed on foot. Olmstead's strenuous attempts to keep out other reform activities (political as well as recreational) were consonant with his belief that the park must remain an antidote to city life and that political conflict would dissipate within a commonwealth of free and diverse people, contained in an orderly public environment, governed by a benign and sanitary administrative state (Blodgett, 1996).

The obvious problems with this kind of politics in an increasingly structured, and multiethnic, population, combined with increasing pressure for the infrastructure supporting active recreation and team sports, pushed parks design into an interactive, urban phase by the 1920s. Particularly during the post-war boom of the 1950s, parks were created to house the modern recreation facilities: pools, courts, gymnasia, community centres, playing fields and tracks. The new park was a community hub. Further rounds of evaluation were brought about by the city planning exercises of the 1960s and again in the early 1980s, when smaller neighbourhood parks and children's playgrounds were added to increasingly integrated networks, and play equipment was introduced, standardized, or rejected for adventure playgrounds and finally reintroduced.

The redemptive ideal of Nature survives, however, not only in the large open parkland that remains in most Western cities, but – equally importantly – in the now controversial villa in the countryside concept resident in the winding, spacious suburb. What only kings could demand once, was now the prerogative of every commoner who could get hold of the land itself (Mumford, 1961). The standards for urban living laid down in such suburban communities as Olmstead's Riverside (near Chicago, 1869) had been codified particularly by Ebenezer Howard and Clarence Perry into requirements for residential living. Howard's *Garden Cities of Tomorrow* (1902), outlining a cooperative joint-ownership alternative to capitalist land development, placed a network of new urban settlements of fixed population size (32 000) *into* rather than *over* a natural landscape, protecting residential areas from industrial encroachment and both from growth by an encompassing greenspace and the agricultural land beyond. Howard and the members of his 1899 Garden City Association were responsible for England's first new towns, Letchworth (designed by Unwin and Parker, 1902) and Welwyn garden cities (1917); and suburbs (Unwin and Parker's Hampstead Garden, north-west London, 1905–1909). Garden city concepts underlay 13 other British new towns, those in the Union of the Soviet Socialists Republic, Holland, Sweden, Italy, Australia and later variants across both the old and new worlds. They have remained at the core of low-density residential construction (see for example, Hall and Ward, 1998). Clarence Perry's *The Neighbourhood Unit: a Scheme of Arrangement for the Family-Life Community* (1929) systematized conditions he had enjoyed in the Long Island suburb of Forest Hills Gardens. Residential units called neighbourhoods were laid out around community centres and elementary schools, the walk to which would not exceed 15 minutes. In both concepts, all homes were furnished with their own tiny remnant of estate life: the grassed yard.

In this way, the urban fringe that had previously been home both to the rural estates and the dispossessed became accessible to the middle class. The tight walking/horsecar city of the 1850s–1880s, expanded by the radiating corridors of the electric streetcar (to the 1920s) and the local road networks of the 1930s and 1940s finally consumed the vast suburbanizing areas on its periphery, with the construction of the freeway networks of the late 1940s, 1950s and 1960s.

The most memorable and difficult legacy of early 20th century civic design may be that of the wholesale creation of new national capitals: Edwin Lutyen's monumental Edwardian New Delhi (1912–1913) for example; the garden city/City Beautiful-inspired Canberra, work of American architect and associate of Frank Lloyd Wright, Walter Burley Griffith (1913); Brasilia (1960s), the astonishing work of modern art by artist/landscape architect Roberto Burle Marx and a team of architects originally associated with Le Corbusier.

WHERE WE ARE: THE MODERN/ POSTMODERN LANDSCAPE

Modern urban landscapes are clearly identified by post-war efforts to disentangle and systematize land uses (separating home, commerce and industry; street from major artery or freeway); to upgrade use-intensity (achieving a higher capital return, particularly on the new commercially integrated office towers and the interior commercial malls themselves); to highlight the structural monuments to commerce (downtown's ever taller and more distinctive corporate buildings/symbols); and to transform the post-industrial city into a site of consumption and entertainment.

All of these developments have been highly controversial, and attempts have been made since the early 1960s to counteract the separation of city components, the reduction and privatization of public space (streets, squares, parks), the intense commercialization of downtown areas and functions, and the ever-expanding use of adjacent farm and wetland (perhaps started by Jane Jacobs' groundbreaking 1961 work, *The Death and Life of Great American Cities*). There has existed, from this time, a movement that itself is scattered, internally diverse and conflictual to re-instill the best of the pre-industrial city. The new urbanism, in its many manifestations, seeks to reintegrate home and

commerce, restore people and function to public space, compact and diversify the spread-out, single-use areas now comprising modern cities.

To move forward, to the earth's millennial landscape, we might pick up Keekok Lee's distinction between the *natural* and the *artifactual*. The natural is that which directs itself, which contains its own teleology or goal-oriented mechanisms; the artifactual is that which has been willfully produced by Man, the material embodiment of human intentionality (Lee, 1999). Today, the frontiers of human control extend far and run deep. Human-induced change is now, in most of the signi cant realms of the environment affected, of unprecedented size. Many forms of human alteration have either become signi cant fractions of natural change or surpassed it altogether (Meyer, 1996).

It is clear that rates of change, their impact, recoverability and bene ts vary signi cantly across the globe, and that it is no longer possible to label all change to the natural landscape bad. It is, in fact, no longer possible in most areas to distinguish a natural landscape from another kind.

A globalizing economy has brought more of the earth s land surface under a capital maximization regime; at the same time, an increasingly sophisticated microtechnology has allowed human penetration into the smallest of organic processes, taking over genetic control of an increasing array of organisms hoped to provide economic bene t in some way. Some of these organisms have wide-ranging impacts when released into natural environments, willfully or by mistake. The effects of introduced species are controversial, but past experience with known cases has shown that the temporary species enrichment often spells impoverishment in the long term (Meyer, 1996).

> Furthermore, technologies of the rising future, such as molecular nanotechnology, in synergistic combination with biotechnology and microcomputer technology, could intensify this tendency to eliminate natural kinds, both biotic and abiotic as well as their natural processes of evolution or change.
>
> (Lee, 1999)

In sum,

> it can at least be asserted that an overall picture is apparent of unsettlingly rapid, sizable, and escalating change in the global environment. If regions differ in their immediate vulnerability to it, their increasing interdependence in a world economy more tightly knit than ever before may make them unprecedently vulnerable in the long term.
>
> (Meyer, 1996)

We have moved from the inscription of our gods onto the landscape to the inscription of the landscape itself. The number of texts dealing with our understanding and treatment of nature has exploded in the last ve years. They deal with the questions: what can we now call Nature? landscape? and what now, can we experience as our gods? These have become not simply timely, but fundamental questions to a growing number of people.

THE REHABILITATION OF EARTH THROUGH ART

Questions about land, power, form and meaning have been asked and answered most directly in the realm of *land art* (the term *land art* itself is more common in Europe; *Earthworks* and *Earth Art* tend to be used in the US). Thrust into the ow of American art genres in 1968 by a few American gallery renegades, its antecedents explicitly reach back to the earliest human landmarks. The rst of these were clearly spirit-related. In the most extensive and celebrated of the neolithic painted caves, the Lascaux Caves in the Dordogne (25 000 10 000 BC), the placing of animal and human images upon the rock surface seems to have meant to bring into being the forces surrounding the progress of life. The reality of the images literally animates the cave enclosures and structures the experience of its progressively powerful rooms. Neolithic megaliths later made these forces manifest on the land s surface, the tombs and monuments transforming select sites and landscapes into sacred places of ancestral respect. The apogee of these monumental forms however, the later Stonehenge in Wiltshire (2000 BC), seems to have been constructed for the Gods themselves, as were the pre-Roman land scrapings (e.g., Berkshire s White Horse of Uf ngton). These produce images that can only be fully appreciated from above, by a transcendent, celestial eye.

The making of land into monument, paradise, arcadia, temple, tapestry, stage, emblem, canvas and sculpture from Babylon s elaborate hydraulic gardens to the small perfect residential creations of contemporary landscape architects Garrett Eckbo and Michael Hough is land art. The movement that brought the earth s surface back as an explicit medium in the 20th century, however, grew out of the minimalist sculpture of the 1960s, itself created as a series of stark environmental statements. In moves designed to break free of the commodi cation of art, its modernist gallery commercialism, artists sought more democratic, spiritually valid media and viewing sites. Once a small number of sculptors in the US, Germany, Japan and England had discarded the pedestal and the gallery, they began to produce works which formed their own sites. The use of land had been pre gured by Isamu Noguchi s (unbuilt) World War II designs for forms that spoke the unspeakable: *Tortured Earth* (1943) reproduces the bodily ravages of war in large slashed earth mounds; *Sculpture To Be Seen From Mars* (1947) would have mounded the face of a desert into the blank, shocked geometric features of post-atomic Man.

Practitioners were drawn to land art not only by the state of art, but also by the earth, and the practice of landscape

architecture. Michael Heizer sought to found a uniquely American art on the excavation and emplacement of a human geometry on the landforms of southwest desert sites.

I think the earth is the material with the most potential because it is the original source material (Bourdon, 1995).

Walter de Maria began in Germany, but produced his most memorable work in the American southwestern as well. His 1974–1977 Lightning Field near Quemado, New Mexico planted 400 stainless steel poles in a grid aligned for height (averaging just over 6 meters) on the desert floor. The work tends to disappear in the glinting sun of the day and, in the occasional storm, grounds the spectacular play of lightning it is designed to attract. The Bulgarian-born Christo and his wife Jeanne-Claude then outdid the others for spectacle (and socio-economic organization) by (among other similar projects) covering a bay of Australian coastline (*Wrapped Coast, Little Bay, One Million Square Feet*, Sydney 1969), joining two Colorado mountain peaks with 381 m of orange flame and fold (*Valley Curtain*, Grand Hogback, Rifle 1970–1972) and most famously, bringing 5.5 m of gleaming white sail down 40 km of California countryside into the Pacific for two weeks in 1976 (*Running Fence*, Sonoma and Marin Counties). It is Robert Smithson, however, who produced the prose to articulate the movement's soul. Smithson was attracted to the expressive and rehabilitative potential of industrial sites. He is best known for his spiral of limestone and black basalt curving off the shore of Utah's Great Salt Lake (1970).

This site was a rotary that enclosed itself in an immense roundness. From that gyrating space emerged the possibility of the Spiral Jetty (in Beardsley, 1998).

British land art has tended to the more subtle and ephemeral. For Richard Long (*Walking a Line in Peru* 1972, 1975 *Line in the Himalayas of White on Grey Stone*, 1981 *A Line in Scotland*) the walks themselves, the arrangement of stones and artifacts of the actions photographed are his works of art. Hamish Fulton photographs his walks through environments visibly unmoved by human intention. Andy Goldsworthy's *Cracked/Broken Pebbles* (Lancashire 1978) and his gold-leaf wrapped rock (*Yellow Elm Leaves over a Rock, Low Water* Oct. 15 1991 Dumfriesshire, Scotland) set up small beautiful acts of Man in nature. David Nash's *Wooden Waterway* 1978, and *Running Oak Table* 1978 lay odd and distinctly human arrangements of natural materials into a forest clearing. Ian Hamilton Finlay has created a small, densely allegorical neo-classical turf and cut-stone inlay landscape, *Little Sparta* in the Pentland Hills, Scotland. Finlay's return to socially reverential emblems is explicit and changes with his intervention through time.

Land art has also produced the environments more commonly associated with landscape architects. Robert Morris, Herbert Bayer and Andy Goldsworthy have worked with reclaimed industrial and urban waste sites. Michael Heizer used the material and topography of a reclaimed strip mine overlooking the Illinois River to create the massive mounded forms of five indigenous water creatures (*Ef gy Tumuli* 1983–1988, Ottawa Illinois, for the Ottawa Silica Company Foundation). The newest urban land art, however, is the garden, particularly in the hands of designers such as Peter Walker (e.g., *Tanner Fountain* on the Harvard Campus, Cambridge, Mass. 1984, where a sphere of mist is projected over concentric circles of stones placed within a shallow grassy bowl) and Martha Schwartz, whose 1979 *Bagel Garden* turned the front of her Boston terrace house into a tapestry of delicately planted black soil, purple aquarium gravel overlain with shellacked bagels and framing box hedge. Schwartz typically uses unconventional materials and a combination of formal historical and minimal/pop art forms to turn landscape into its own particular art (*see* **Art and the Environment**, Volume 5).

CONCLUSIONS

The very fruitful integration of the natural and social sciences, and the arts, in the study of landscape and cityscape, has paralleled the globalization of human-induced change. This change is an extremely complex, locally distinct phenomenon, sensitive not only to culture and population growth and decline, but also to policy initiatives at every level and the powerful programs of international finance, development and aid.

The compelling attraction of nonhuman nature and of the landscape that reflects its formative forces has, in part, been that of the *essential other*: of the powers and beings beyond our control that surround and contextualize our human life. The modern humanization of the planet itself threatens to recontextualize human life with the products of its own making, that is, to make perfectly narcissistic an existence once shared with its originating forces and beings. What this would do to the conditions of life that have inspired fear, awe, love and creative dialogue since Man has been able to contemplate his own small size is yet to be understood.

REFERENCES

Abrams, M H (1958) *The Mirror and the Lamp: Romantic Theory and the Critical Tradition*, Norton, NY.

Andrews, M (1999) *Landscape and Western Art*, Oxford University Press, New York.

Beardsley, J (1989) *Earthworks and Beyond*, Abbeville, NY.

Berleant, A (1997) *Living in the Landscape: Toward an Aesthetics of Environment*, University of Kansas Press, KS.

Bourden, D (1995) *Designing the Earth: the Human Impulse to Shape Nature*, Harry N Abrams, New York.

Buell, L (1995) *The Environmental Imagination: Thoreau, Nature Writing and the Formation of American Culture*, Harvard University Press, Cambridge, MA.

Coates, P (1998) *Nature: Western Attitudes Since Ancient Times*, University of California Press, CA.
Cooney (1999) Social Landscapes in Irish Prehistory, in *The Archeology and Anthropology of Landscape*, eds P J Ucko and R Layton, Routledge, New York, 46–64.
Cooper, D E (1992) The Idea of Environment, in *The Environment in Question*, eds D E Cooper and J A Palmer, Routledge, London, 163–180.
Cosgrove, D and Daniels, S (1988) *The Iconography of Landscape*, Cambridge University Press, Cambridge.
Davis, K (1965) The Urbanization of the Human Population, *Sci. Am.*, September, 213(3).
Eisenberg, E (1998) *The Ecology of Eden*, Alfred A Knopf, New York.
Forman, R T T (1990) The Beginnings of Landscape Ecology in America, in *Changing Landscapes: an Ecological Perspective*, eds I S Zonneveld and R T T Forman, Springer-Verlag, New York, 35–41.
Hall, P and Ward, C (1998) *Socialable Cities: The legacy of Ebenezer Howard*, John Wiley and Sons, Chichester.
Hunt, J D (1992) *Gardens and the Picturesque: Studies in the History of Landscape Architecture*, MIT Press, Cambridge, MA.
Hoskins, W G (1958) *The Making of the English Countryside*, Hodder and Stoughton, London.
Jellico, G and Jellico, S (1975) *The Landscape of Man: Shaping the Environment from Prehistory to the Present Day*, Thames and Hudson, London.
Kostof, S (1995) *A History of Architecture: Settings and Rituals*, Oxford University Press, New York.
Lee, K (1999) *The Natural and the Artifactual: The Implications of Deep Science and Deep Technology for Environmental Philosophy*, Lexington Books, New York.
Lewis, C, Mitchell-Fox, P, and Dyer, C (1997) *Village, Hamlet and Field: Changing Medieval Settlements in Central England*, Manchester University Press, Manchester.
Lynch, K (1972) *What Time is This Place?* MIT Press, Cambridge, MA.
Marsh, G P (1970) *The Earth as Modi ed by Human Action*, Arno, New York (first published in 1874 under the title: *Man and Nature*).
Meinig, D W, ed (1979) *The Interpretation of Ordinary Landscapes*, Oxford University Press, New York.
Meyer, W B (1996) *Human Impact on the Earth*, Cambridge University Press, Cambridge.
Mumford, L (1961) *The City in History: Its Origins, Its Transformations, and Its Prospects*, Harcourt, Brace and World, New York.
Nairn, I (1965) *The American Landscape: A Critical View*, Random House, New York.
Schreiber, K F (1990) The History of Landscape Ecology in Europe, in *Changing Landscapes: An Ecological Perspective*, eds I S Zonneveld and R T T Forman, Springer-Verlag, New York, 21–34.
Spirn, A W (1998) *The Language of Landscape*, Yale University Press, New Haven, CT.
Tjallingii, S P and de Veer, A A, eds (1982) Perspectives in Landscape Ecology, *Proc. Int. Cong. Netherlands Soc. Landscape Ecol.*, PUDOC, Wageningen, The Netherlands.
Van Noos, W (1977) The Fate of City Beautiful Thought in Canada, 1893–1930, in *The Canadian City*, eds G A Stelter and A F Artibise, McClelland and Stewart, Toronto, 162–185.
Wasserman, E R (1968) *The Subtler Language*, The Johns Hopkins Press, Baltimore, MD.
Watkins, D (1982) *The English Vision: The Picturesque in Architecture, Landscape Architecture and Garden Design*, Harper and Row, New York.
Williams, M (1993) An Exceptionally Powerful Biotic Factor, in *Humans as Components of Ecosystems: The Ecology of Subtle Human Effects and Populated Areas*, eds M J McDonnell and S T A Pickett, Springer-Verlag, New York, 24–39.
Williams, R (1973) *The Country and the City*, Oxford University Press, New York.
Zimbardo, R A (1986) *A Mirror to Nature: Transformations in Drama and Aesthetics*, The University Press of Kentucky, KT.

Leapfrogging Technology

see Leapfrogging Technology (**Volume 4**)

Leopold, Aldo

(1887- 1948)

The name Aldo Leopold has become something of a mantra among environmentalists. The use of this man's name evokes, from his writings and his personal example, a sensing of the scientific complexity of the biological world, combined with a sensing of the elegant beauty which can be discerned in natural science. It is associated with a code of ethics: a code that offers a moral baseline by which our social system could exist constructively within the limits of biological productivity.

At his birth in 1887 in Burlington, Iowa, there seemed to be no reason to anticipate the intellectual impact that this person would have. At the time of his boarding school at Lawrenceville School in New Jersey, he was noted as a bright and somewhat reclusive lad, who had developed a remarkable ability to see and to interpret nature, with an insight that others could not see. At the time of his graduation from Forestry School at Yale University

(1909), his biographers note evidences that he may have appeared to his contemporaries as quite self-assured: a feature, which might have contributed to some initial management problems during his employment with the United States Forest Service (1909–1924). But during this time, his extraordinary perceptive skills with nature led this young forester to discern two impending catastrophes that others could not see: the destructive effects of runaway deer populations, and the impending debacle of soil erosion. At this early age, he was able to foresee the need for preservation of wilderness, and he originated the establishment of the Gila Wilderness area in the state of Arizona.

When Aldo Leopold published his book on *Game Management* (1932), and became of Professor of Wildlife Ecology at the University of Wisconsin (1933 until his death in 1948), he had earned a reputation as a fine scholar, a writer of impressive power, and an intellectual leader (albeit in a non-classical academic subject). Yet in his lifetime, his perceptiveness and his grasp of ecological issues was not always appreciated. While Aldo Leopold was recognized as the leader of the profession of wildlife management, his efforts to provide scientific leadership in resource management in the State of Wisconsin were voted down. His authorship of some of the most perceptive ecological and philosophical essays focused on the beauty that can be perceived in nature, and led to ethical generalities that were recognized as stunning pieces of writing. Yet four major publishing houses rejected his manuscripts that later led to the book, *A Sand County Almanac*. At the time of his death, in 1948, the power of the mantra of Aldo Leopold was only making its first emergence.

But then, *A Sand County Almanac* was accepted by a publisher just days before he died: note the powerful metaphor of a major leader of ecological and ethical thinking who died while fighting a grass fire on a neighbor's farm. It is probable that his blend of scientific knowledge about natural history, ecology, and management, with the essentials of ethical behavior that can serve as the baseline for sensible ecology, was the powerful pulpit from which he could speak about mankind's relationship to the natural world. I personally doubt that he felt that his writings and teachings would have a very widespread impact. But then the full tenor of his post-mortem voice began to rise in the last half of the 20th century. The readership of *A Sand County Almanac* rose continually during the 50 years after its publication (1949, Oxford University Press, New York), breaking the usual performance of literary/scientific books which ordinarily have a half-life of about six years. Instead of fading, this book has been reprinted more than thirty times, and sold more than two million copies in the last 50 years. It has been translated into nine foreign languages. The demand has been sufficiently great that a second book of his essays was published in 1953 (*Round River*, Oxford University Press, New York). Several other collections of his early writings have appeared (*Aldo Leopold's Wilderness*, edited by D E Brown and N B Carmony, Stackpole Books, Harrisburg PA, 1990; *The River of the Mother of God and other essays by Aldo Leopold*, edited by S L Flader and J B Caldicott, University of Wisconsin Press, Madison, WI, 1991; *For the Health of the Land*, edited by J B Callicott and R E T Freyfogle, Island Press, Washington, DC, 1999; and *The Essential Aldo Leopold*, edited by C Meine and R L Knight, University of Wisconsin Press Madison, WI, 1999). And, most important, a definitive and sensitive biography, *Aldo Leopold, his life and work*, has been written by Curt Meine (University of Wisconsin Press, Madison, WI, 1988).

Today, the mantra of Aldo Leopold has become a byword for environmentalists around the world; this man is considered to be the father of the science of *restoration ecology*. His central role in terms of ethical life-styles has stimulated the formation of a new branch of ethics and morality centered on the health of the Earth. One example of his concise definitions of a land ethic: A thing is right when it tends to preserve the integrity, stability, and beauty of the biotic community. It is wrong when it tends otherwise. People come regularly to visit his farm in Wisconsin as a site of inspiration and intellectual renewal. Aldo Leopold has become a powerful mantra for ethical life-styles.

Being introduced at meetings of environmentalist groups, I have had lovely youngsters shriek, Oh, you must be the son of Aldo Leopold followed by embraces and kisses. At another meeting recently, the chairwoman suggested a standing ovation for the son of Aldo Leopold. It is difficult to define the magnitude of the stimulus that this man has left to the world of environmentalism.

Photo: courtesy of the Aldo Leopold Foundation Archives.

A CARL LEOPOLD *USA*

Life Style, Private Choice, and Environmental Governance

Peter Timmerman
University of Toronto, Toronto, Canada

One of the great emerging issues that society faces is the relationship between lifestyle choices and environmental

burdens. The most obvious way of describing this governance problem is the shift from point-source problems, which characterized many jurisdictional responses in the 1970s- 1990s, to current non-point source problems. But it is much more complicated than that, and poses signi cant social, political, and economic concerns for the 21st century.

DEFINING ENVIRONMENTALISM

A central issue in contemporary environmentalism is how to deal with life-styles and private choices. The environmental movement has until now not been able to convince people that many looming environmental problems are rooted in the larger production and consumption patterns of modern society. There is a lot of lip service given to life-style changes, but the underlying patterns are tenacious. To deal with this – many environmentalists believe – will require a transformation of the economic/industrial production system, on the one hand; and on the other, a lifestyle transformation amoung individuals and families. There has obviously been some movement on both these fronts. Industries have been altered; and individuals and households have blue boxes; schools have green curricula; the environment is everywhere. But the environmental agenda is persistently driven by a purity model and not a sustainability model.

The purity model emphasizes clean air, pure water, and the elimination of waste products. It fits in very well with the residual Puritanism of many Anglo Saxon countries, and the whole range of related middle class values. It is naturally associated with personal health, and the general fears about pollution and toxic chemicals. For Ministries of the Environment, this has meant a strong focus on environmental health, on toxin abatement, on end-of-pipe elimination of pollutants, and so on. The environment enters as a carrier or pathway for the delivery of toxic chemicals.

Of course, the other aspect of the environmental agenda is present. People have a broad concern about endangered species and some endangered spaces. There is also broad support for doing something, as long as it is not too onerous (e.g., blue boxes). There are also fears about global climate change and other global issues. But it is clear that after 30 years of the environmental movement that a broad reconsideration of lifestyles by the vast majority of the population in developed countries (let alone developing countries) has not happened.

If it is assumed that consumption of goods and services, especially with the advent of consumerist lifestyles in the developing world, will drastically increase, there are usually two safety valves proposed. The first safety valve is that the greening of industrial production and materials and energy use is rapidly improving, and that will mean that we will be able to continue our consumptive lifestyle without increasing the ecological burden. The second safety valve (related to the first) is that a knowledge and service economy will gradually virtualize a substantial range of current commodity flows.

Neither of these assumptions is completely plausible. While there have been substantial savings made in the elimination of many metals, materials, and energy, the economic growth patterns seem to have simply begun filling the new space available. There is also little evidence that virtualization is going to replace the need for physical resources any time soon. For example, teleworking seems to be a factor in increased car use in suburban belts.

FUNDAMENTAL DRIVERS OF THIS EMERGING ISSUE

Perhaps the fundamental driver of this issue is the model of human well being encapsulated in middle-class ways of life. This is now being promulgated through the media, a model which was embryonic in Southern California and in early suburban New York in the late 1940s, but which has now spread around the world. This lifestyle carries with it a set of intrinsically woven choices about urban and suburban housing, garden structures, transportation biases, and so on. These seem to be only marginally separable, though some separation has been made – for example, smoking and hard liquor consumption patterns have changed drastically over time.

Behind this model is a second driver, namely the powerful assumption of choice as the core element in freedom and freedom of expression. This assumption is historically rooted in a set of religious practices and ideals (concerning free will and sin in Christianity), but has more recently become incorporated within economic theory. In the 19th century, economics was transformed away from a classical model of analyzing supply and demand curves based on resource, labor, and land use, into a neo-classical model which developed the notion of the rational allocation of scarce resources by individuals seeking to maximize their personal utility. This model dovetailed effectively with the rise of consumer society. We now have a cultural model of self-fulfillment encapsulated in the phrase: shopping is freedom. The objects that we buy are now the carriers of self-expression and creativity.

This consumerist ethos is most overtly widespread in children and teenagers, who are acutely conscious of label, product, and status. The school system is currently under assault by corporate marketers and others

attempting to break down the barriers between the cultural space of school and the cultural space of commercialism. At the heart of this struggle is the struggle over who will control the myths, stories, and images of society.

HOW WILL THIS ISSUE FURTHER EVOLVE?

The place of government in this situation is very dif - cult, for a variety of reasons, not the least of which is a long history of misguided government intrusion into trying to reshape human behavior. Nevertheless, the women s movement and other activities prove that governments do have a clear role in fostering and enabling change through legislation, advice, and leadership roles. A simple example of this in the case of consumerism has been government support for alternative automobile eets, recycling, and using pilot activities as examples of best practices.

The big problem, as already mentioned, is the disjunct between personal choice, personal lifestyle and the recognition of those choices and the burden on the environment. I suspect that this is about to alter, in part because the environmental movement is undergoing a signi cant shift in its areas of concern (*see* Editor s Introduction and **Environmental Politics**, Volume 5). It will become ever clearer that the relationship between personal well-being and environmental well-being is much subtler and more powerful than has been credited hitherto. The best example is the extraordinary case of the stratospheric ozone hole, where people suddenly found themselves threatened with the violation of a basic right that they did not even know they had before the right to walk outside on a sunny day.

It will be clearer as the 21st century continues that these kinds of threats threats to the fundamental fabric of life through species loss, climate change, biotechnology, and the rest, will threaten the very de nition of the human. This is already unsettling, but it will become even more unsettling as violations of immune systems, private genetic information, and global deterioration in certain ecosystems accelerate. One can predict that the relationship between these issues and fundamental issues of human well-being will force a recasting of what constitutes the environment and environmental well-being, and that personal sustainability and environmental sustainability will be more tightly drawn together.

This development will allow governments, citizens, non-governmental organizations (NGOs), and industries to renegotiate the boundaries of choice, freedom, and the burdens that current patterns of choice are placing on the environment. This will be very dangerous, but very necessary.

Literature and the Environment

Peter Timmerman
University of Toronto, Toronto, Canada

The relationship between the arts, in general, and environmental change is complex, and certainly not restricted to the presence of images or descriptions of nature as content. Literature poetry, plays, novels, non-fiction is, in part, an alternative environment where the powerful tools of literary form can work within a set of internally generated imaginative rules to explore communicable experience. Literature may be said to have its own ecologies, and has certainly generated worlds of its own.

This article focuses, through the examination of selected pieces of literature, on the Western literary tradition; and for a variety of reasons, including space restrictions, deals with the mainstream tradition, particularly in works of the Romantic period (1750 1850). One of the other reasons is that Romanticism is a pivotal moment in modern literature, as well as culture more broadly; and it is a direct in uence on the rise of modern environmental thought and practice. To elucidate the importance of Romanticism, I draw upon the myth-critical school, associated with the work of the Canadian literary critic, Northrop Frye (1947, 1968), which, whatever its limitations in other areas, provides a very powerful apparatus for entering into some of the more revolutionary aspects of literary practice as it applies to cosmologies and mythologies.

In particular, this article argues that a central role of literature in the context of environmental change is to be an expression and sometimes a fomenter of changes in world views. This is one powerful way in which literature interacts with environmental change: that is, not just recording people s feelings about changes in the natural environment, but challenging and restructuring their mental maps of the world by means of story, character, style, language.

INTRODUCTION

In the summer of 1816, a group of English hippies including Lord Byron and his friend Percy Shelley went to Lake Geneva in Switzerland. Coming after the battle of Waterloo, 1816 was the rst time that anyone from England had been allowed to travel in Continental Europe for a decade, and tourists ocked across the Channel. Lake Geneva was special because it had been the site of a famous romantic novel by the French philosopher Jean-Jacques Rousseau, *Julie: or The New Heloise*, and

the little party of travelers was looking forward to a summer of wandering around the shores, soaking up the atmosphere.

Unfortunately, the summer of 1816 was the year without a summer thanks to the explosion of Mount Tambora in the Southeast Pacific, which cast a pall over the skies of the world, and cooled off or deranged weather patterns everywhere. Lord Byron and his party were forced to stay inside their rented house and watch the rain, day after day. On one such day, they all decided to make up ghost stories. Shelley's wife, Mary, had a dark dream, which she turned into what everyone agreed was the best of the ghost stories, about a scientist who decides to create a human being. Thus, *Frankenstein* (Shelley, 1818) was born.

This is an interesting example of the relationship between the environment and literature, and highlights one way in which much of the creative interplay between the constructed world of human culture, and the natural world within which that culture is created, works. It is almost a parody of the not very creative idea that the environment might be a cause of literature, the way people used to argue that geography causes culture. It is only marginally less creative than the conventional notion that the environment appears in literature as the content of the novel, poem, or play.

A more useful way of thinking about literature and the environment derives from the Romantic movement of the 18th and 19th centuries – though not completely beholden to it – which focused on seeing literature as in part an expression of human exploration of the relationship between the self (consciousness, mind, imagination) and the outside world (nature, environment, materiality); but that the nature of this expression was not just the content, but the quality of style, the voice of the author or narrator, the shape of the stories being told, and so on.

One critical element in this exploration is provided by *metaphor* – the imaginative leap between two different realms or categories as seen by the creative mind. This mind need not be a literary mind – the Gaia Hypothesis (*see* Gaia Hypothesis, Volume 5), that the Earth might be seen as a single, living organism – is a creative metaphor; though many scientists would opt for its milder cousin, the *simile*; that is, the Earth may be treated for the moment as if it were akin to an organism, or a self-organizing system of sorts. Another example is provided by the 19th century chemist, Kekule, who could not solve a problem in the arrangement of chemical bonds, and had a dream about a snake biting its own tale, which gave him the idea of the potential for the circular bonding pattern: another creative metaphor.

METAPHORIC AND RATIONAL LOGIC

The anthropologist Gregory Bateson, just before his death, wrote a pathbreaking comparison between metaphoric and rational logic through comparison of two syllogistic patterns (*Men are Grass*, Bateson, 1987). The first, very familiar, syllogistic pattern is:

Men die.
Socrates is a man.
Socrates will die.

This Aristotelian syllogism is (in part) a deductive machine with a hypothesis (Men die), a fact in the world potentially subsumable under that class (Socrates is a man), and then a new relation when the fact is put into that class (Socrates will die).

A second pattern, which looks like the first, but is in fact a logical fallacy is:

Grass dies. Men die. Men are grass.

This is the fallacy of affirming the consequent, and its fallaciousness (and dangerousness) can be seen in the following further example:

White is pure.
There are white people.
White people are pure.

The grass syllogism, however, is familiar to readers of the book of *Ecclesiastes* in the Bible, as a powerful linking together of two different categories (men, grass) by means of metaphor. This kind of categorization and making of connections can be as systematic and powerful as the rational logic associated with scientific hypothesis generation; but the *two solitudes* (to use C P Snow's term) can be antagonistic to each other. Indeed, it has been one of the complaints of the modern poet that the scientist is draining all the poetic mystery out of the world, by challenging the traditional web of metaphoric connection; while, on the other hand, the scientist complains that writers don't know any science.

What makes the writer's world work is the making of connections that may not be overtly rational, but may make meaningful sense as metaphor. For example, there is no scientific reason why blue should be associated in Western culture with the musical *blues*, but most writers will be able to play on those associations when necessary. Similarly, the idea that a weeping willow is actually weeping is obviously absurd to a botanist or forester; but as a symbol, it may well be a powerful carrier of an emotion in a story.

One thing that separates modern literature from earlier literatures (and non-Western literatures) is that earlier literatures can depend on the expectation that the entire culture within which the literature operates also subscribes to metaphoric logic. It is only in modern Western culture that the idea that nature and human experience have no metaphoric connection to each other has flourished.

Among the reasons why Shakespeare continues to have such an influential role in modern culture is the fact that

he lived at a pivotal moment in the shift away from a universe of mutually resonant meaning (the end of the 16th century), and is sensitive to the shift, very often giving us some characters who still respond to traditional signs and symbols of the environment, and others, more modern, who dismiss them merely as old wives' tales.

As an example of the former, the night before Julius Caesar is stabbed in Shakespeare's play *Julius Caesar* (1597), Cicero and Casca speak. Casca reports on what he is seeing:

> Casca: Are you not moved, when all the sway of earth
> Sways like a thing unfirm? O Cicero!
> ••• never till tonight, never till now,
> Did I go through a tempest dropping fire.
> Either there is civil strife in heaven,
> Or else the world, too saucy with the gods,
> Incenses them to send destruction.
> Cicero: Why, saw you anything more wonderful?
> Casca: A common slave – you know him well by sight –
> Held up his left hand, which did flame and burn
> Like twenty torches joined; and yet his hand
> Not sensible of fire, remained unscorched •••
> Men all in fire walk up and down the streets,
> And yesterday the bird of night did sit
> Even at noonday, upon the market-place
> Hooting and shrieking. When these prodigies
> Do so conjointly meet, let not men say
> These are their reasons, they are natural ;
> For I believe they are portentous things
> Unto the climate that they point upon.
> Cicero: Indeed, it is a strange disposed time;
> But men may construe things after their fashion
> Clean from the purpose of the things themselves.
>
> (I, III, 4–35)

Here Casca represents the world of symbols, portents, signs from the gods, that are warning of impending civil war; while Cicero represents the newly rational Man, sceptical, aware of how signs may be false, mere rumor, hysteria, and connected only by projection from human desire.

NATURE AND MYTH IN PRE-MODERN COSMOLOGY

We can think of cosmologies – proposed descriptions of the universe – and mythologies – stories that inhabit and illuminate the universe – as forms of metaphor writ large. They are large-scale structures of meaning, many of whose elements and characters appear in literary works, while the works themselves influence the overall feel of how it is to live in such a universe. For example, Dante's *Divine Comedy* (1330) is an imaginative journey through the cosmology of medieval Western Europe; but the power of Dante's vision of his journey had incalculable effects on the birth of the sense of what it means to be a modern Western individual in such a universe.

In this section I want to sketch out two rough cosmologies, the first from the beginning of literature, in Sumeria; and the second, the evolving medieval Christian synthesis cosmology that broke down at the beginning of the modern era. These are not the only cosmologies available. They are simply two of the most important, particularly for the ways in which they think about, and structure, in myth and story, the cultural map of Nature.

In most cultures, the cosmology is woven into stories that are told orally, and have been handed down by word of mouth for many centuries. The Western tradition is special (which does not mean specially good), in that much of its tradition was very quickly written down, and the mainline tradition has been a strongly literary one from very early on; though the origins of the stories that make up the tradition are almost certainly long oral traditions that have been lost or rewritten. Western traditions begin the story of the creation of the Earth and sky usually with some kind of a Creator God, and often with a story about how the sky and the Earth came together and mated (the sky being male, and the Earth being female). A number of other cultures start things off completely differently. For some, the universe comes out of its own insides, like a spider spinning a web out of its own guts. For others, everything comes up to the surface of the sea; and that is how things got going. In some cultures, things are even sung into existence.

It should also be stressed that what is being sketched out here is the mainline Western tradition, against which other oral traditions and contrary stories struggled, and often lost. Of particular importance is the continual appearance and disappearance of dangerous story fragments about women and other strange beings whose role in the mythology is troubling and often suppressed or twisted. Scholars have only recently become interested in these marginalized stories, in part because of the light they shed on mythologies that refused to fit into the big story.

The Origins of Earth and Sky

The first myths in the Western tradition of which we can read begin in the Near East, and are associated with the hydraulic civilizations of Sumer (Southern Iraq), Egypt, and later, Babylon, and Assyria. These cultures grew out of the surpluses generated by the ecological richness of alluvial soils in the big river basins of the Nile, Tigris, and Euphrates rivers. Through extensive irrigation and complex management works, these cultures were able to feed large numbers of people, and also build the world's first big cities (the same process was going on in India and China). Essential to the survival of these cities was the organization that kept the irrigation systems going, and the predictability of changes in river flow according to the seasons. Furthermore, as these cities grew richer, they became subject to attack by enemies, which necessitated the

growth of armies. All this complexity led to the creation of writing, accounting, and monumental architecture.

In the earliest Western stories drawn from the Middle East, the Earth is seen as a piece of land (often a hill) surrounded by, or floating on, the waters of Chaos, waters that will drown the world unless they are countered by creative power. However, these waters are also life giving, especially in an area with so little rain: in some myths the great waters are the source of all beings. By analogy with a room or a temple, the Earth is often seen as holding the sky up by four pillars on each corner, representing the four directions, and we still use this image in talking about going to the four corners of the Earth.

The Earth has its natural rhythms associated with seasonal change, the rising and lowering of rivers, seedtimes and harvests. The Sky has an intimate relation to this process, but it seems to be much more abstractly orderly, more designed, less subject to human influence. Unless you are an astronomer, the sky at night presents the following picture: there are fixed stars, that do not seem to move relative to each other, though they move as a whole through the night sky through the year. They have strange patterns, called *constellations*, which look as if they might mean something, especially those associated with seasonal shifts. Every once in a while, certain stars seem to move, called *planets*, which is a Greek word meaning wanderer, because they seem to be wandering among the stars. These appear at different times of the year, which can be predicted, but they move in a puzzling fashion. Also in the night sky in the Northern Hemisphere, there is a swatch of white stars called the *Milky Way*, which many myths see as some sort of gateway to heaven.

Obviously, the moon is the largest body in the night sky, and it has a monthly rhythm which is predictable, and which was very early on associated with the highs and the lows of the tides of the rivers and sea. Because of its similarities with female rhythm (the word *menstruation* derives from a word for month), the moon was constantly associated with the female, and with tidal waters. This is not universally true, but became more and more accepted in Western myth.

Most important of all was the sun, which very quickly became associated with ruling power. Some cultures, such as the Egyptian, toyed for a while with having a Sky Goddess (Nut) separate from the Sun God, but this rapidly disappeared. The sun was all-powerful, and ruled the day. Most of the gods and heroes of the Near East were associated with the Sun God, who was invariably male, and kingly. Crowns had rays of the sun attached to them; kings carried orbs (the sun symbol) and the chariot of the King moved over the world the way the chariot of the sun moved across the sky.

To reinforce this, the earliest agricultural myths explained the creation of the world as the sexual intercourse of the Earth (female) with the Sky God or sun (male), and this would be symbolized by the plow (male sexual organ) entering the furrow of the female (female sex organ), and seeding the soil with semen. Other versions of this would connect the Sky God with the rain, again, male potency being rained down on the waiting female Earth. These myths strongly reinforced the idea that the male was active and creative, while the female was passive. This helped set up a whole pattern of male-oriented myths of power, that reinforced the social organization of the period, which was becoming increasingly dominated by priests, kings, and hierarchical organizations. This was essentially patriarchal, i.e., dominated by a male-structured mythology.

There are indications, though the evidence is extremely controversial, that there was at least a strong counter mythology, especially in other regions of the Near East such as Crete, that focused on the power of women, particularly associated with their dominant role in early agricultural practice and in their role as childbearers. This matriarchal culture is very often characterized (in the surviving male cultural mythology) by a myth of a powerful Queen choosing her temporary male partners, who are perhaps sacrificed when they lose their usefulness as defenders.

In the earlier period there were several different stories, and competing gods. For instance, another important piece of the cosmological structure was the God of Storms, who sometimes competes with the Sky God for supremacy; as do other gods associated with primal forces, like fire. These cosmologies altered significantly with changing social structures, though the main lines began to harden. By about 1000 BCE in the Babylonian epics and stories, a mythical structure was in place which was picked up and modified in part by the Israelites, a wandering tribe of shepherding peoples who seem to have traveled throughout the Near East, settling in Palestine following experiences in Egypt, captured in the oldest parts of the Bible.

It was this adoption of the reigning Near Eastern mythology as a backdrop or set of assumptions in the early books of the Bible that was so important for the rest of the Western tradition.

The Rise of the Hero

One last element of this mythology needs to be put into place, which is the idea of the hero (and occasionally the heroine) who usually represents the human race in its struggle to survive in a strange world. This hero is almost always seen as more than human, a superman, or Hercules, who is related to the Gods in some way. This hero has a series of adventures, which carry him through the world, and help to make sense of it.

This hero structure often mirrors the activities of the Gods in sustaining the world. For instance, one of the great myths of the world is the sun going into the darkness every

night, and being reborn every morning. The darkness is like chaos: a place of confusion and death. The hero, following the track of the sun, goes down into the place of darkness at the end of the world, or the bottom of the world, and there wrestles with various evil forces, and then resurfaces to begin the cycle of the world all over again. The hero travels through various levels of the world, from the highest to the lowest and back again.

Let me summarize all the above in Figure 1. It represents a composite of the Western worldview, about 700 BCE, which would be found with variations all over the Near East and the Mediterranean.

The most important single fact about this myth structure is that the higher up you go the more ordered it becomes. This is because God or the Creator is also the embodiment of order, and the human beings who are closest to him, the kings and the priests who rule human society, also embody order. Notice also that the highest order is more and more removed from most natural processes. This suggests that the higher order is rational, and associated with the powers of the human mind separated from the turbulence of the lower world.

A second important fact for mythological purposes is that the imposition of order from the top down is a central part of any story that fits within this structure. In general, the king, Sky God, God as Creator, is visualized as coming down into the disordered world and imposing order on it according to a rational plan. The resistance to that is seen to be irrational, and associated with the demonic. The farther down one goes, the more ungodly Nature is, the more dangerous, and the more associated with darkness, uncontrolled swirling water, and chaos.

The process of myth creations takes these levels, and the images and symbols associated with these levels, and weaves them into powerful stories. We have already referred to the story of the hero who descends into the realm of darkness and wrestles with the darkness. The darkness is associated symbolically in geographical terms with caves and labyrinths; bodily with the lower bowels (the bowels of the Earth); meteorologically with wild chaotic oceans (as in storms and tempests); and animalistically with huge devouring serpents, gigantic whales, beasts that are half-human and so on. The archetypes are very long-lasting in the Western tradition, and show up in other traditions as well: for example, the image of being swallowed by a whale as a symbol of hell appears in the biblical book of Jonah, in the novel, *Moby Dick*, in the children's story, *Pinocchio*, and in the second *Star Wars* film.

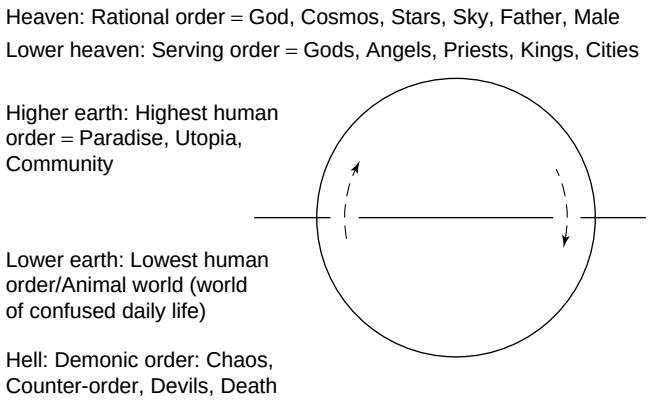


Figure 1 Ancient western cosmology [after Frye and others]

In other words, each mythological level clusters a range of images and symbols around itself, and one of the ways of identifying the kind of story one is in is checking out the symbols, images, and metaphors that pop up. It is not as simple as that, of course, because some of the images have changed over time, and have become used as parodies and ironies of themselves. But an initial grasp of how these clusters of images works is of interest, particularly as we continue to explore shifts in the images and ideas of Nature over time.

In the great Sumerian and Babylonian stories and epics, the hero journeys through the realms of darkness and of the gods, for a variety of reasons, including the need to tame the turbulent world to the purposes of humankind. Here is an excerpt from *Inanna's Journey To Hell* (from approximately 1500 BCE). Inanna was a powerful goddess in Mesopotamia. Inanna descends into hell during the dry season of the year, ostensibly to rescue her husband from death, but mythically to restore the natural cycle of life to the kingdom. Here is a brief scene of her return, with devils attached to her like leaches:

> Over the corpse hanging on a spike,
> They scattered the bread of life,
> They sprinkled the living water,
> And Inanna stood up alive.
> She will come up from the pit, but the Anunnaki
> (demons) seized her,
> The Judges said,
> Who has ever returned out of hell unharmed? •••
> She is coming, Inanna is coming
> from the pit! Devils are fastened
> to her thighs, devils walk beside her,
> meagre like reeds, thin as pikestaves.
> There goes in front of her a thing
> With a sceptre, but he is no minister •••
> She is coming, Inanna is coming
> From the pit!
>
> (trans Sandars, 1971)

The mythical quest of the female to restore health to an ailing kingdom is a widespread complementary story to the questing male battling more obvious threats to the kingdom (like monstrous beasts). In Egyptian myth, for example, the queen/goddess Isis has a King/husband Osiris, who is murdered by his brother Seth, and his body scattered around the world. The world loses coherence, and the kingdom withers. Isis travels the world, discovering and piecing the

fragments of Osiris back together again, and when he is fully restored, order and fertility returns to the land.

These myths of Inanna and Isis began to surface in the first millennium BC, and following the conquering of Egypt by Alexander the Great in the 300 BCE period, the myth of Isis in particular spread around the Mediterranean. A Greek equivalent was the myth of Demeter and Persephone. In this myth, Demeter, who was the goddess of fertility, has a daughter Persephone who is captured by the chief God of the underworld, Pluto. Demeter goes down into Hell to rescue her daughter, but discovers that her daughter has eaten six pomegranate seeds while being captured, which requires her to stay in Hell for six months of the year forever after. It is during those months that the world loses its fertility.

The Arrival of the Christian Story

These myths of Demeter, or of Isis the Eyptian, or of any number of other Gods, spirits, etc., were widespread in the ancient world until after the arrival of Christianity. The whole array of beliefs and stories is usually referred to as *paganism*, and what paganism means, generally speaking, is the worship of multiple gods, particularly in sacred natural places, like woods, forests and caves, and these gods often are the spirits of animals or powers in the world. As a reflection of the multiple cultures and multiple influences in the region, the stories and myths are themselves multiple, and often reflect a belief that the natural world is subject to constant surprise and transformation. Ovid's *Metamorphosis* (late 1st century, BCE) is a classical Roman expression of this widespread set of beliefs.

Somewhere around 1500 BC, a tribe of nomadic peoples, the Hebrews, were wandering around northern Mesopotamia. It is probably important that they were nomads, that is, that they weren't part of a rigid city culture, and didn't carry around with them statues or big pictures. They were saturated in the mythological culture of the region, but they developed a significantly different worldview over the next 700 years. The story of this evolution is told in the Old Testament, or Hebrew Bible (*see* Christianity and the Environment, Volume 5; Judaism and the Environment, Volume 5).

At the heart of this story is the idea of monotheism, which is more than just the idea that there is one God. Essentially, the Hebrews came to believe that they had a special God who had created the whole world and everything else in it, but was not Himself any part of that world. A central part of the idea is a radical questioning of what it means for something to have power – spiritual power. The Hebrews said, you people worship all kinds of Gods, animal gods, female gods, male gods, gods of fertility, etc. But what is the power that all these gods share? What is the power that gives them power?

There were other cultures that had a King of the Gods, Zeus, or Jupiter, but they were just more powerful versions of the other gods. What was the sacred power itself, which all these gods supposedly had? These were ultimately screens for the One True God, and could be done away with.

The second thing the Hebrews pointed out was that life with a lot of gods was confusing. You never knew which gods were on your side, and which were against you. You had to spend your whole day sacrificing to this god, and not forgetting that one, and so on. The Hebrews argued that it was ridiculous for people to be worshipping trees and animals: they should be worshipping the one true holiness that created the trees and the animals and gave them life.

This radical attack on the multiple natural gods of the ancient world was also part of the Hebrews' tribal strategy. When they entered what is now Israel/Palestine, which was then called Canaan, they proceeded to fight against all the local tribes, and part of their fight was against all the local gods, including the local versions of Ishtar, Baal, and other spirits.

The early books of the Bible capture the massive rethinking of Nature that went on in the creation of such a worldview. The upshot of the rethinking is that, in the end, though Nature can, and should, be seen as a glorious creation of God through which we might be able to get a glimpse of God's great wisdom and knowledge, it no longer has any intrinsic spirituality of its own. It is a vehicle for the greatness of God.

In the Hebrew Bible, we still get glimpses of the old tradition, some snippets about the old Canaanite religion, and a lot of complaining that pagan worship refuses to die; but by and large the Canaanite religion disappears from view as time passes.

It so happens that the Jewish people were only a minor player in the Ancient Greek and Roman world, so they were tolerated as being a bit eccentric. The God of the Hebrews joined the rest of the diverse array of gods and goddesses that peopled the Mediterranean.

All this changed with the arrival of Christianity.

When Jesus was crucified around 35 AD, He was originally thought of as just one more Jewish radical, but thanks to the efforts of, among others, a man who would later be called St Paul, who was both a Roman citizen and a Jew, Christianity began to spread around the Mediterranean. It took root because of its promise of individual salvation from death, and its intolerance for other religions. It was a bit like Microsoft: you can have no other operating system but mine.

Thanks to a series of events, culminating in the conversion of one of the later Roman emperors, Constantine to Christianity, Christianity became the most powerful religion in the Roman Empire, but unlike the other religions whose adherents were happy to let other religions flourish,

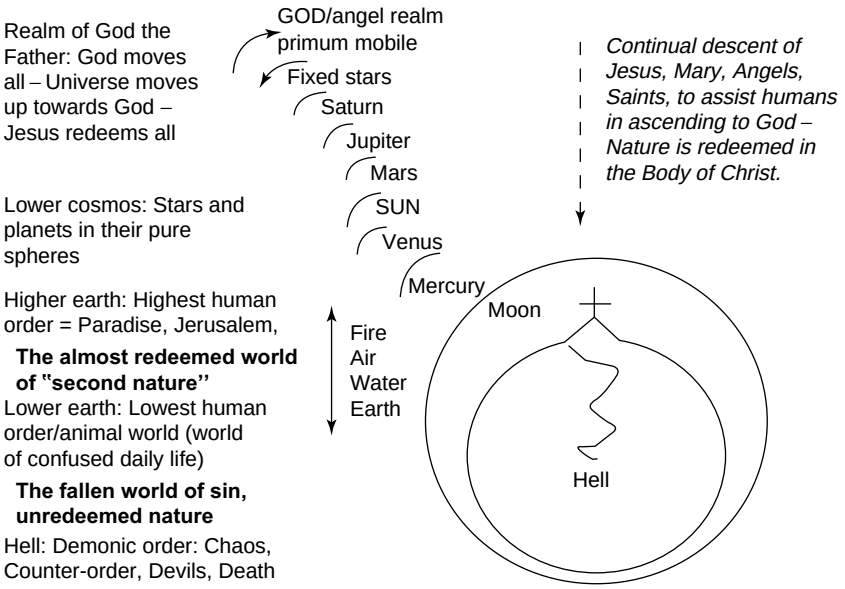


Figure 2 Medieval western cosmology. Primum mobile, the sphere that moves all other spheres [after Frye and others]

the Christians went on a 300-year rampage, destroying all the Roman cults, burning down the Great Library at Alexandria (associated with Isis), and hacking down all the sacred forests and groves associated with the pagan gods. Having done this in the Mediterranean, they then proceeded to do the same throughout Western Europe, culminating in driving underground the Celtic Gods of Ireland and the Norse Gods in Scandinavia.

This extraordinary story ended, in about the year 1000, with virtually the entire Western world being subject to Christianity. The central mythological and cosmological structure that had been extant before was now subjected to a strong narrative based on a single salvationary drive; i.e., the entire cosmos was part of a redemptive process whereby a fallen world – intensified by fallen human beings – was, at the end of time, to be re-perfected. The ideal Hero was Jesus, whose story became the template for the journey of Everyman through the world.

The early Christian literature, based on this very powerful vision of the spiritualization of the world, saw Nature as either irrelevant, or as a realm of symbols whose meaning was connected to the larger spiritual journey of the soul. In fact, Nature was seen as the realm of wilderness and danger, far removed from the realm of God. As this world view developed, it dovetailed with leftover astronomical and physical assumptions of the ancient Graeco–Roman world, which (for instance) argued that the Earth was the lowest realm, made up of the element Earth, and that the higher one went away from the Earth, the more one approached perfection. For example, the planets and stars in the heavens moved in concentric circles (actually, transparent spheres) around the earth, and so on. Again, as in the earlier cosmology (Figure 1), the higher you went, the more rational the world became, until you reached order and truth, God.

A map of this cosmology (Figure 2) shows the dimensions of the Medieval/Renaissance universe, and also the cluster of literary images that held together views of nature throughout that period. Our lower nature is subject to temptation; and the wild of the uncultivated natural world is a symbol of that temptation. This wilderness should be turned into a garden, or a farm, or ultimately a city, just as the soul should be cultivated.

In Dante's *Divine Comedy*, the hero (guided by Virgil, the ghost of the dead Roman poet), descends into hell, at the center of the earth, and reascends in the Southern Hemisphere, where he climbs the mountain of Purgatory, at the top of which he finds himself in a kind of paradise, like the original lost Garden of Eden. After this, he leaps up into space (guided by his long lost love, Beatrice) and rockets through space towards the ultimate place of God beyond the stars. This literary journey maps the whole natural and spiritual world of his time.

SECOND NATURE AND THE WINTER'S TALE

As already mentioned, one of the most important aspects of Shakespeare's work is that it appears just at the moment when the late medieval world view is being superseded by the early modernist world view. To repeat, Shakespearean literature (and that of his contemporaries and some successors) often operates according to metaphoric logic – categorization based on resemblance or co-incidence. A mandrake root was shaped like a human being, and therefore had influence over human beings. The

configuration of stars at one's birth was not an accident: the planets had different influences, as did different plants, animals, etc.

The importance of this is that the metaphoric view allows for a certain amount of positive valorization of natural nature, even in the medieval Christian period. An important medieval distinction was between *natura naturata* and *natura naturans* – nature more or less as a noun, and nature as a living verb, *naturing*. Shakespeare is fascinated by the power of nature naturing. Among the most interesting characters in his plays are bastards, who are always standing up and proclaiming their devotion to the raw powers of nature.

In the late Shakespeare plays, like *The Winter's Tale* (1613), nature (in one role) becomes something much more benign, and becomes tied to certain ambiguous functions of nature in literary art. The most obvious of these is the pastoral world of rustic shepherds and shepherdesses. Another was the garden. A third, and hardest of all to articulate, was the great benignness of Nature's rhythms in the world. In Shakespeare's late plays, there are tempests, bears, shipwrecks, etc., but one is constantly led to believe that it will all work out, that something is going on deeply in the world to reconcile everything. This is one reason why they all feel like fairy stories.

As the title suggests, there is a storylike quality to *The Winter's Tale*, and it is obviously haunted by winter (a season and a mood). In the scene excerpted below, a shepherdess greets two lords, who are searching for Prince Florizel, the heir to the kingdom, who has fallen in love with her. Unbeknownst to these lords, the shepherdess, Perdita, is herself a princess; they believe her to be a mere rural peasant. As with Ophelia in *Hamlet*, Perdita is fluent in the metaphoric language and etiquette of flowers. The scene opens with her gracefully giving rosemary and rue (and explaining their symbolism) to the two gentlemen. When they suggest that she is giving them winter flowers, she points out that she has avoided giving them nature's bastards – carnations and gillyvors. Because these plants were known for their genetic sports – their streaking and changes in patterns (piedness) suggested waywardness – they are symbols of nature's indiscriminate sexuality. Perdita is staking out civilized ground.

The next paragraphs provide a central discussion of the Renaissance view of nature and nature's art – what was called *second nature*. Second nature is what cultivation does: it takes the waywardness of nature and harnesses it to nobler ends. Perdita opens the discussion by using the term *art* for the process of creating nature (Natura naturans), and is naturally wary of it. Polixenes – who is about to denounce his son for fooling around with a mere shepherdess – brings in a different metaphor: the metaphor of grafting. In grafting, a branch of a growing tree is cut off and grafted onto another. The force of nature that continues the growth process eventually produces a next generation stronger and healthier than the earlier one. So Polixenes argues that human art can use nature's powers to improve nature. Thus, Shakespeare gracefully intertwines the theme of human cultivation and nature, with the irony that the supposedly rural shepherdess is herself a figure of the highest hidden art.

> Perdita: For you there's rosemary and rue; these keep
> Seeming and savour all the winter long:
> Grace and remembrance be to you both,
> And welcome to our shearing!
> Polixenes: Shepherdess – a fair one are you –
> Well you fit our ages with flowers of winter.
> Perdita: Sir, the year growing ancient •••
> The fairest flowers of the season
> Are our carnations and streak'd gillyfovers,
> Which some call nature's bastards; ••• and I
> care not to get slips of them ••• For I have heard it said
> There is an art which in their piedness shares
> With great creating nature.
> Polixenes: Say there be;
> Yet nature is made better by no mean,
> But nature makes that mean; so, over that art
> Which you say adds to nature, is an art
> That nature makes. You see, sweet maid, we marry
> A gentler scion to the wildest stock,
> And make conceive a bark of baser kind
> By bud of nobler race; this is an art
> Which does mend nature; change it rather, but
> The art itself is nature.
>
> (IV, III, 72–96)

ROMANTICISM AND AFTER

Beginning about the time of the Industrial Revolution, due to a variety of forces, the synthesized Christian–Ptolemaic cosmology began to collapse, and there arose a new cosmology, which we can call the creative, romantic, or revolutionary cosmology. This has remained with us ever since, and lives in uneasy relationship with the fragments of the earlier imaginative cosmology that still operate.

What happened was that the previous worldview, which was hierarchical – kings, priests, male centered, rule-oriented and rational – was no longer able to use the cosmology to support the existing power structure. The new emerging world of astronomical discoveries, the discovery of the new world, the new physics, etc., all confused the world picture; and during that period of confusion, there was also a widespread assault on hierarchical forms of government throughout Europe.

As a result, the world order of the people on top (let us now call it that officially) began to be attacked and parodied by means of a counter-cosmology. It is the persistence of both cosmologies – the official line, and the romantic line – that is essential to understanding how the myths and images of the modern Western world work, and especially modern *environmentalism*.

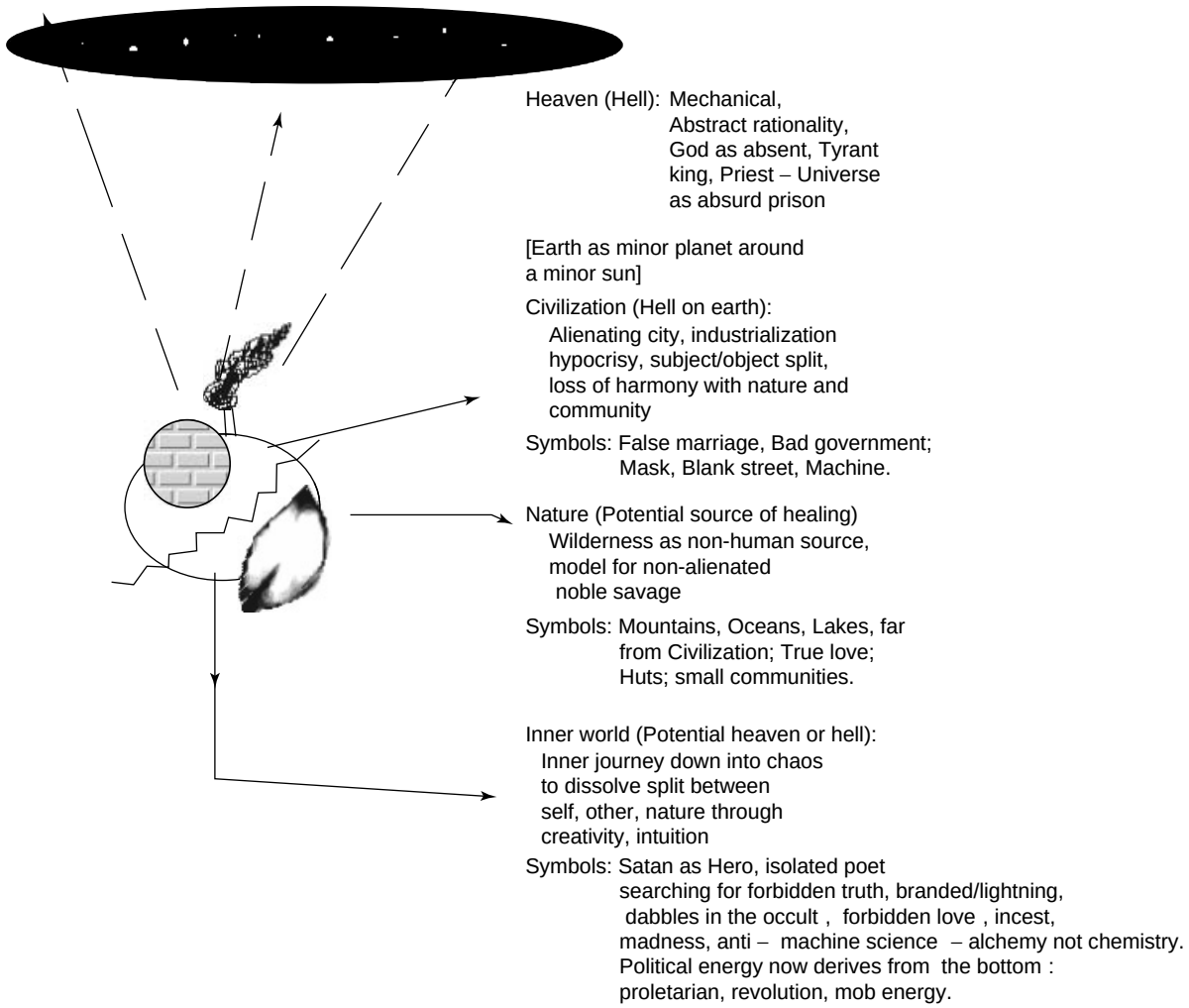


Figure 3 Modern cosmology/mythology/romantic parodic, countermyth

The Romantic countermodel broke through in a number of different places – it shows up in the American and French Revolutions, and has since been central for artists, musicians, environmentalists and others.

Let me describe a third cosmology (Figure 3), this time from the top to the bottom. It is somewhat less clear than the previous two. The reason is that one of the powerful things going on in the new cosmology is an attempt to parody and satirize the old cosmology. Remember that this is a cosmology, and a mythology, so the energy of it is carried through symbols, images, and metaphors.

To begin, the earlier world view was organized hierarchically, the higher you went, the more rational, ordered, and powerful you became, because you were getting closer to God.

The new worldview inverts and parodies this approach, because it no longer believes explicitly in the natural right of rulers to rule. This was one of the causes and outcomes of the American (1776) and French (1789) Revolutions.

The whole idea of the top becomes not the place you want to go to reach heaven. It is now the place of infinite space, the strange empty abstract cosmos ruled over by the laws of Isaac Newton. It is blank, inhuman, full of stars that no longer have any relevance to our lives. God has disappeared, turned into a puppet master or a watchmaker. This connects to a very powerful image of kings and gods not as reasonable beings, but as repressive rulers. The forces of authority are now thought of, not as spreading truth and reason from high places, but as being in charge of a repressive machine, designed to stifle human expression. Elements of this machine are increasingly seen as arbitrary creations of the repressive mind, and among the arbitrarinesses is the association of the masculine with the rational.

The next level down is the city of darkness, which is no longer the city of light, the holy organized rational city of Jerusalem. Rather it is the place of alienation, where faceless people work for the machine. This was particularly true in the first years of the industrial era, where the

whole structure of society was disrupted. Previously rural farm communities and family groups were uprooted and turned into workers. Workers, by definition, are gears in the machine. They have lost their humanity, and have become human resources.

The city is also associated with alienation, which shows up in the emerging idea of the lonely figure in the crowd. This becomes a powerful image in the new Romantic poetry of the 19th century. In William Blake's poem, *London* of 1794 we get some idea of this (William Blake, 1757–1827). Here are a couple of verses:

I wander through each chartered street,
Near where the chartered Thames doth flow,
And mark in every face I meet
Marks of weakness, marks of woe.
In every cry of every Man,
In every infant's cry of fear,
In every voice, in every ban,
The mind-forged manacles I hear.
How the Chimney sweeper's cry
Every blackening church appalls,
And the hapless soldier's sigh,
Runs in blood down palace walls.

And so on. The streets and the river are chartered – mapped, sold – and the human mind is now manacled, chained and shackled. The crisscross of square streets in the new modern city contrasts with the organic wandering streets of the medieval town. The city begins to grow, and to engulf its surroundings.

Similarly, just as human beings are being eaten up by the machine, so also is Nature being eaten up by the machine, and being transformed into objects. Power is the power over, the control over human beings and nature, which is created by having control over the wealth created by the machine, and this in turn is a creature of civilization.

Here is Blake again, in one of his late poems *Jerusalem*. A revolutionary poet is speaking:

O Divine Spirit, sustain me on thy wings,
That I may awaken Albion (Britain) from his long and cold repose;
For Bacon and Newton, sheath'd in dismal steel, their terrors hang
Like iron scourges over Albion: reasonings like vast serpents
Infold around my limbs •••
I turn my eyes to the schools and universities of Europe
And there behold the Loom of Locke, whose Woof rages dire
Washed by the Water wheels of Newton: black the cloth
In heavy wreaths folds over every nation: cruel works
Of many wheels I view, wheel without wheel, with cogs tyrannic
Moving by compulsion each other, not those in Eden, which,
Wheel within wheel, in freedom revolve in harmony and peace.

Here Blake summarizes his attack on the whole Industrial Revolution. He sees wheels without wheels (without here means outside of) as a mixture of clockwork, waterwheels, mechanisms imposed from the outside; as opposed to the organic development of wheels within wheels he sees in some future Paradise or Eden. The whole civilization has become poisoned.

For Romantic writers, it may be that progress was making people healthier and wealthier, but they were also being progressively pulled apart from each other, and from nature. Among the powerful Romantic themes was the idea of the *noble savage*, the very powerful idea that would haunt Romanticism from then on, which was that peoples, races, that were closer to nature were somehow better and wiser than people who belonged to civilization in general, and Western civilization in particular. The noble savage – identified with the increasingly marginalized indigenous peoples who were being chewed up by the new Western machine – was not alienated from his (or her) community, and had a special relationship with the natural world.

In contrast to them, civilized Man was split up the middle. Socially, naturally, and even in conflict with his own mind – for we begin to see that the mind-forged manacles may reflect the larger world view – a rational conscious mind/identity, alienated from one's own physical world.

If we go back to the cosmological chart, we can see this great split, which is rampant in the industrialized mind. For the Romantic writers and critics, this was the equivalent of a new *fall of man*, a tragic moment in human history against which they rebelled.

How is this split, these alienations to be overcome, or healed?

In the early Romantic era, there were two main answers, one political, and the other artistic, or creative. The political answer was revolution. Essentially what was required was for the mass of those being turned into machines to revolt and overthrow the masters on top.

Here we enter the next realm of the new cosmology: the idea that the bottom is the source of creative energy. It may look like the old hell, but it is where the next revolution would come from. Throughout the 19th century, and even in our own century, there is this image of the crowd, the mob, the mass of workers on general strike, like some dark force from a volcano, finally breaking through the stifling authoritarian machine towards some wider liberation which would finally sweep aside the old institutions. It was these images that were at the heart of the French and Russian Revolutions. The revolution from below was the force that would lead to the new heaven of Utopia. One of the starkest images that operated throughout the next hundred years was that of the blacksmith (and later the steelworker) working in the industrial equivalent of hellfire, with a hammer, striking blows for freedom in the midst of the dark factory.

This was the political answer. The split which was caused by repression would be overcome by the workers on the day

when they realized that the forces of repression were not just flimsy, but were being kept in power by their ability to keep the workers in ignorance of their true capacity to rule themselves without masters.

THE ARTIST AS RADICAL

A related answer, which moves us back into the details of our cosmology, was that Nature might become a source for a new regenerative force. If part of the symptoms of our problems could be seen in the destruction of Nature, and in the separation of the human from the natural, overcoming that split might also be relevant. Nature was naturally not human, so while it might be polluted, at its heart might be something different, which could regenerate the lost original human nature – again, sort of like repairing the original fall of man in the Christian tradition. Not only was nature not human, at least some parts of it could be seen to have an organic structure, one that naturally developed, rather than being imposed from the outside, by machine.

Again, notice that one of the important inversions of the previous cosmology is going on here. In the previous cosmology, wilderness was a place of chaos, of entanglement in irrational natural forces, leading one on to hell. In the new cosmology, wilderness, simply because it is not part of the machine, is a potential entry point into heaven, or at least paradise. The very fact that it is uncultivated makes it a starting point for the rebirth of a non-infected culture.

Also notice that it is the uncultivated, the non-rational, the pre-split in all its forms that becomes a source of new strength, of the divine energy of revolt. It is not a coincidence that we find in this period a new respect emerging for children, who come to be seen as unspoiled figures, not as adults in training (and modern educational theory dates from this early romantic period) – for aboriginal peoples – the noble savages I spoke of (we begin to get romantic stories like the *Last of the Mohicans*) – and the initial upgrading of women's voices that takes place in the opening of the 19th century with women like Mary Wollestonecraft.

Of equal importance is the rise of the artist, the creative person, the musician and poet, who represent the overcoming of the split between the mind and the world through their imaginations or their works. How does the artist, the musician, the creative person, overcome the split between the human and the natural?

There are two approaches that have become well known. One is associated with the image of the lamp. In this image, the idea is that the artist, through his or her creative mind illuminates and reaches out into the natural world, and overcomes the split through the power of new artistic power. We can hear this, for example, in the music of Beethoven, like his 5th Symphony, where the composer seems to hammer away (like a blacksmith) at the elements and forces of the world around him. This is an active approach (to say the least).

A second approach is for the artist or the poet to become more humble in the face of Nature – to become sensitive to the details and the rhythms of the natural world, to learn Nature's lessons. Over time, this cleansing of perception will reveal deeper truths of the universe to the artist. We can see this approach in the work of John Ruskin, who combined an environmental concern with a call for detailed analyses of Nature's designs in flowers, rocks, trees, and atmospheric phenomena.

These two approaches are often mingled together in different artists. When we look at a painting, like *The Sun owers* from the quintessential Romantic artist, Vincent Van Gogh, we cannot say whether Van Gogh was imposing his vision on the sunflowers, or saw them in themselves more deeply than we can.

In earlier cosmologies, the way you learned was by going up to high places, going to the city, finding a wise male leader. There were exceptions to this, but only part-exceptions: for example, in Dante's *Inferno*, the hero goes down into hell, but that is only the beginning of the journey up to the highest stars.

In the Romantic movement, we find that the idea of the journey down into the underworld in the previous cosmology, becomes the journey of discovery itself. This is because the new hell is caused by the repressive actions of the forces of order. The Romantic artist (and later versions of that, like the rock musician) goes out to the marginal peoples, down into the slums, away from the center of power. It is only there that the truth of repression is really to be seen for what it is. The people who live in the new hell are more liberated, because (in a parody of the earlier cosmology where the closer you got to God, the more rational everything became) the new cosmology argues that the farther you get away from authority, the more truth can be found. Since rationality is being imposed from the outside, irrationality may be more truthful than rationality. The madman may be saner than the doctor.

Lastly, the descent into a liberating new hell may be a descent into the personal hell, the unconscious mind. The poet, the musician, repelled by the repressive nature of modernity, descends into a variety of personal and social hells, and supposedly returns from that with new messages that will transform the world. The mythology of drug-ridden visionary rock musicians comes directly from this cosmology.

This descent into the mind may be echoed by the descent into the wilderness. The artist may go into the heart of darkness, go on a vision quest, seek out the truths that can only be found far from civilization. Inner nature and the natural world become linked, because both are on the run from the forces of human technological control. This helps

explain some of the attractiveness of the flight away from civilization into nature.

MODERNITY AND THE CONTEMPORARY ARTIST

These Romantic themes, and the dark cosmology that goes with it, have remained central tools in the hands of writers and artists. Much of the literature of the 20th century explored the themes of the alienating environment of the city, the potentially liberating, but also potentially dangerous, voyage into the natural environment, and the journey of the hero and heroine into inner nature. A novel like George Orwell s *1984* paints the environment of the city as alienating, repressive, and one without any hope. The only time that the characters nd any liberation is when they escape to Nature, temporarily. The same is true in Aldous Huxley s *Brave New World* and many other powerful indictments of contemporary life.

A new theme that has emerged in literature since World War II, and is paralleled by the rise of the environmental movement, is the possibility raised by nuclear war of the elimination of all life on earth. For writers, this has been inescapably linked to the experience of the death camps of Auschwitz and Birkenau, where the whole question of civilization came under question. The German critic, Theodor Adorno, famously said that one could no longer write poetry after the Holocaust. The dark hole of that experience sucked the life out of the pretense of words. Since then, writers have been constantly haunted by threats to the fabric of life; and this has shown up in the themes and characters they write about.

Nevile Shute s *On The Beach* (1959), which portrayed the consequences of the global spread of nuclear fallout, has a good claim on being the rst modern environmental novel, saturated with the themes just outlined. Of increasing interest in recent years has been the strength of non- ction writing about the environment, re ecting in part the presence of large amounts of scienti c or naturalistic fact in contemporary life. Writers like Gary Snyder (*Earth House Hold*) and Annie Dillard (*Pilgrim at Tinker Creek*), anthropologists like Richard K Nelson (*Make Prayers to the Raven*), biologists like Sandra Steingraber (*Living Downstream*), and farmers like Wendell Berry (*Home Economics*) grounded their environmental writings in exemplary detail, and often used them as springboards for moments of personal discovery.

CONCLUSION: AND DEATH, TOO, IS IN ARCADIA

In searching for a piece of recent writing about the environment to conclude this essay, I decided to use the example of a poem, written by myself, and published in the mid-1980s. I choose it, not because it is a great poem, but because as the author, I can speak to it somewhat freely as to the intentions and sources of its creation, and how it was intended to relate literary forms to environmental changes. This may give some insight into the literary process (or at least one writer s process).

The poem was part of a suite of poems called *The Fate of the Earth and Other Poems* published in the Canadian environmental magazine *Probe Post* (Timmerman, 1988). This particular poem draws on the literary tradition called *pastoral*, which was used in *The Winter s Tale*. It was a sophisticated Greek and Latin form of verse (and sometimes prose) about simple shepherds in rustic landscapes, either singing about the joys of the rural life, or complaining about a shepherdess who would not love them, or the troubles of farmers in an uncertain world. Many writers Virgil, Horace, Pope, Shakespeare, Sidney used this form, often disguising criticism of mainstream urban society under the guise of simple unlearned folk. There was also a kind of utopian dream associated with this simplicity, which was classically located in a place called Arcadia in Greece.

Accompanying the poem was a photograph of a famous painting by Nicholas Poussin (*see* **Art and the Environment**, Volume 5) called *Et in Arcadia Ego*. The painting shows a group of shepherds in Arcadia huddled around a large tomb upon which are carved the words *Et in Arcadia Ego*, which is a reference to death. That is translating from the Latin *I Too Am In Arcadia*. Naturally, this means that there is no escaping from death, not even in the beautiful natural simplicity of Arcadia.

With this introduction, here is the poem:

And Death, Too, is in Arcadia

The house of Nature sags against the rain.
The sap fails in the sugar maples, and the
cry of the waterbirds fades like mist
over morning by grey, secret inlets.
The rustle of wind is the whisper of weeping.
There is Death in the woods, lling the air,
down along the marshlands, dropping
out of the once benevolent skies, shedding
hot tears of helpless pain.
The webbing of life tears and drifts apart:
poison pulses from the snow, and
the land soils itself into the stream, and
the lake slowly wearies of all its children.
Against its own, its deepest rhythm dying,
Nature screams and warps against the grain.
And look, here, look
here in the depths of faltering Arcady,
even the gravestones dissolve in grief
at things never seen or heard before
in any version of pastoral.

The poem was written in the speci c context of acid rain, but also more generally in relation to a world in which

things that naturally symbolize goodness, like rain and mother's milk, are turned into poisons. There is a strange situation where the natural processes are turned against themselves, and the deep rhythms of nature are warped by the influence of human chemical poisoning.

One important aspect of the poem is that, like the distant lakes and waters affected by acid rain, Arcadia is far away, and yet affected by long-distance transport of pollutants. I carefully used images and ideas from the scientific literature on the loss of leaf cover, stresses on trees, loss of water birds, early snowmelt at the beginning of the year that pulses acid into lakes just at prime birth time for fish, and so on. However, there is an even stronger connection to one powerful theme in the pastoral poetry tradition: the mourning for the lost shepherd.

In the English poetry tradition, the most famous example of this theme is John Milton's poem *Lycidas* (1642). In this poem, which mourns the loss of a young poet who was drowned at sea, the whole of nature is bidden to mourn as well for the loss of one who loved nature:

> With cowslips wan that hang the pensive head,
> And every flower that sad embroidery wears.
> Bid amaranthus all his beauty shed,
> And daffadillies fill their cups and tears,
> To strew the laureate hearse where Lycid lies.

This is a classic example of what was discussed earlier with reference to *Julius Caesar* where the sympathies of nature for human beings are assumed.

In my poem, I was trying to evoke the images of the dying poet, but with a modern twist, to evoke the dying of nature, and also to suggest that this was something new in the history of the world (never seen in Arcadia or anywhere) – the possibility of the dying of nature. An oblique reference is made to a famous book on pastoral by the literary critic William Empson, *Some Versions of Pastoral* (Empson, 1935).

Also, among the guiding images of the final lines of the poem is that of sculptures and tombstones that I had seen in city graveyards that had been worn down by air pollution over the years, and were themselves symbols of the destructiveness of contemporary environmental problems.

A brief technical point, which may highlight the way in which, in a poem, everything matters, including the form as well as the content, is that I used a number of what are called *weak endings* – where the rhythm of the word is such that there is a strong beat on the first half of the word, and a weak, falling off rhythm at the end – words like weeping, dropping, filling, shedding – so as to evoke sadness and loss. There are also lines like the rustle of wind is the whisper of weeping that are meant to give the reader the sound of the wind in the threatened woods.

These kinds of considerations – and many others – go into the response of those who write literature today in the face of the environmental crisis.

REFERENCES

Bateson, G (1987) Men are Grass, in *Gaia: A Way of Knowing*, ed W I Thompson, Lindisfarne Press, Great Barrington, MA.

Buell, L (1995) *The Environmental Imagination*, The Belknap Press, Cambridge, MA.

Empson, W (1935) *Some Versions of Pastoral*, Chatto and Windus, London.

Sandars, N K (1971) *Poems of Heaven and Hell From Ancient Mesopotamia*, Penguin Books Ltd., Harmondsworth, Middlesex, UK.

Shelley, M W (1818) *Frankenstein; or the Modern Prometheus*, Broadview Press, Peterborough, Ontario.

Timmerman, P (1988) The Fate of the Earth and Other Poems, *Probe Post*, Summer 1988, Pollution Probe Foundation, Toronto.

FURTHER READING

Abrams, M H (1953) *The Mirror and the Lamp*, Oxford University Press, Oxford.

Bate, J (2000) *The Song of the Earth*, Macmillan Press/Picador, London.

Berry, W (1987) *Home Economics*, North Point Press, San Francisco, CA.

Dillard, A (1974) *Pilgrim at Tinker Creek*, Random House/Vintage, New York.

Frye, N (1947) Fearful Symmetry, University Press, Princeton, NJ.

Frye, N (1968) A Study of English Romanticism, Random House, New York.

Nelson, R K (1983) *Make Prayers to the Raven*, University of Chicago Press, Chicago, IL.

Snyder, G (1969) *Earth House Hold*, New Directions, New York.

Steingraber, S (1998) *Living Downstream*, Random House/Vintage, New York.

Long-wave Economic Cycles (Kondratyev Cycles)

see Kondratyev, Nikolai Dmitrievich (**Volume 5**)

Love Canal

Love Canal was the name given to a suburban subdivision in Niagara Falls, NY. Residents first became aware of their proximity to toxic waste through a *Niagara Gazette* newspaper article by Michael Brown in April 1978. Local

housewife, Lois Gibbs, quickly emerged as the key organizer of a community riddled with fears for their health. She initially began a petition campaign to close the 99th Street School, built in the virtual center of what had been the dumpsite, abandoned by a subsidiary of Occidental Petroleum.

By August 1978, the New York Commissioner of Health declared a state of emergency and ordered the closing of the school. Within days, US President Jimmy Carter declared an emergency and provided funds to permanently relocate 239 families living within the first two rows of homes closest to the former waste dump. The community campaigned for the relocation of all 900 families.

In May 1980, the US Environmental Protection Agency (EPA) released results from blood tests of residents showing chromosome damage. Love Canal activists took two EPA officials hostage. On May 19th, Lois Gibbs issued an ultimatum: commit to relocating all Love Canal residents by noon May 21, or What we've done here today will look like a Sesame Street picnic to what we'll do then. The White House agreed to permanent relocation by the deadline.

Occidental Petroleum has paid nearly $250 million (US$) to compensate residents and reimburse state and federal government clean-up costs. In 1990, the EPA announced that sections of the area were safe for resettlement.

The controversy is credited with sparking the EPA's Superfund Programme.

ELIZABETH MAY *Canada*

M

Malthus, Thomas Robert

(1766- 1834)

Thomas Robert Malthus, clergyman, professor, political economist, was born in Great Britain to a well-to-do country family. He studied mathematics at Cambridge, and was ordained a minister at the age of 22. Malthus served as a clergyman and as professor of modern history and political economy at Haileybury College. His ideas and eventual economic philosophy were molded in part by the debates he had with his father, who held optimistic views of human progress. Malthus himself saw social problems growing in exponential terms.

Malthus' views of the economic and social implications of population growth were contrary to the common economic assumptions of the day. Most viewed population growth as being a positive by-product of economic growth and social advancement. In 1798, Malthus published his *Essay on the Principle of Population* (there were seven editions in all). In the essay, Malthus held that while population grows exponentially, the availability of food and other resources grows arithmetically. But population would be held in check by periods of plague or war. Technological advances would provide periods of social stability, until the power of exponential population growth again resulted in too many people for available resources. Social discord would inevitably return.

An essential point in Malthus' argument is that in the long-run, population would always tend to remain above the level that resources could adequately support. Malthus wrote in his essay that population is constantly acting upon man as a powerful stimulus urging him to the further cultivation of the earth, and to enable it to consequently support a more extended population. In addition to the pervasive impact of population on human cultural conditions, his theory also alludes to the existence of a static stock index. This index holds that there are absolute limits to resource availability and limits to the positive affects of human innovation. It was in this realm of thought that he maintained an ongoing debate with his friend David Ricardo. Though Ricardo agreed with the Malthusian notion of population growth, he disagreed with the implications. Ricardo argued that technological innovation, substitution, and the discovery of new resources would be the inevitable cultural responses – in essence the stock index was neither static nor binding.

The Malthusian notion of scarcity and population also contributed to a theory of wages (iron law of wages, or Malthus' law of wages). This proposed that an increase in the wage rate results in an increase in population, this in turn depresses wages, and thus wages remain mostly at a subsistence level. His theological background was reflected in his economic and social thought. Famine and poverty were certainly seen as natural, and hence, divine outcomes. He also believed that moral constraints (abstention from sex) were an essential part of addressing population, though he had doubts about the practicality of this solution.

The ideas of Malthus were refined within the neo-Malthusian tradition. In this realm, the idea of absolute scarcity and the exponential impact of population growth have been quite attractive to ecologists, and to many in environmental movements. Such perspectives see a strong relationship between the geometric impact of economic and social consumption, and absolute limits inherent in nature.

KEVIN HANNA *Canada*

Man and Nature as a Single but Complex System

Michael Thompson
The Musgrave Institute, London, UK and The University of Bergen, Norway

No matter how lightly we tread on the Earth, we cannot avoid altering it. And, as it alters, so the way we tread on

it - our ecological footprint, as it is sometimes called - is, in turn, altered. On and on. Natural scientists tend to look at this interaction of the human and the natural from the Earth's perspective, The Earth as Transformed by Human Action *(Turner* et al.*, 1990) being the classic text. Social scientists tend to look at the interaction from the socio-cultural end:* Living with Nature *(Fischer and Hajer, 1999), with its subtitle* Environmental Politics as Cultural Discourse*, is a recent example. But can we go further? Can we push each of these approaches (the natural scientist's and the social scientist's) to the point where they actually meet and give a single, uni ed theory of our relationship with nature?*

Yes we can, chorus two schools of thought: one sociological, the other ecological. The rst has its roots in social anthropology and is, properly speaking, a theory of socio-cultural viability (but, since that is too much of a mouthful, it has been shortened to Cultural Theory, with the capital letters serving to distinguish it from other theorizing about culture). The second has emerged from natural resource ecology, where those whose interest is in grasslands, sheries, forests and so on encounter the institutions that are doing the exploiting and the managing, not as organized arrangements of people and their various convictions as to how the world is, but as patterned interventions in the ecosystems they are studying.

In this article, we describe a uni ed theory, and some of its implications for policy.

A ROAD WITHOUT END

The classic assumption, in both ecology and social science, is that there is a one-way transition from state A to state B. In ecology, the process of succession (Clements, 1916; Odum, 1969) ensures that an initially unstructured state of affairs (one huge niche filled with anarchic, opportunistic and competitive organisms (the r-strategists)) is steadily transformed into a climax community: a structured and stratified arrangement of diversified niches, with clearly defined interrelationships between the species (the K-strategists) that occupy them (*see* r–K Strategies, Volume 2). In social science, this predictable, linear and equilibrium-seeking model of change is paralleled by a number of grand theories in which some inexorable logic moves us all from mechanical to organic solidarity (Durkheim, 1893); from community to society (Gemeinschaft to Gesellschaft, Tonnies, 1887); from traditional to modern (Weber, 1922); from status to contract (Maine, 1861); from capitalism to communism (Marx, 1859); or, as modern theorists of institutions put it, from markets to hierarchies (Lindblom, 1977; Williamson, 1975). Different masters may define their A's and their B's differently, but all subscribe to a two-fold scheme and to some driving force (such as rationalization, internal contradiction, or spiraling transaction costs) that carries the totality from A to B.

These transitions, whether ecological or socio-cultural, are all in the direction of more orderliness, more differentiation, more connectedness, and more consistency and, once they have gone as far as they can go in that direction, that is that. In other words, these models of change end up making change impossible. Of course, something on the outside may intervene and mess things up, thereby setting the whole thing in motion once more but, left to themselves, these models get ecosystems and socio-cultural systems from A to B and then stop. Change, these models tell us, is a temporary phenomenon.

These models are beginning to be seen as less than satisfactory. They explain change by getting rid of it, and they are increasingly incapable of making sense of what is actually going on. They have now been challenged by models that are indeterministic (i.e., more than two-fold) and make change a permanent and essential feature of existence: the four-fold institutional scheme proposed by Cultural Theory (Thompson *et al.*, 1990) and the four-fold ecocycle advanced by Holling (1986). If social and ecological systems are as these models say they are, their interaction will inevitably result in complex and non-linear dynamics, giving an unpredictable, always out of equilibrium, and never ending sequence of transitions between multiple states. And none of these will ever be the end of the road.

In the classic social science formulation, two kinds of solidarity interact. *Markets* are the competing players, all merrily bidding and bargaining with one another; *hierarchies* are the benign authorities who ensure that the various conditions for playing of this trading game (a level playing field, for instance) are in place. Cultural Theory does not reject this foundational distinction. Rather, it argues that there is more to life than just markets and hierarchies and that you will lay yourself open to all sorts of unwelcome surprises if you go on assuming that hierarchies and markets explain it all. Take, for instance, the Brent Spar oil storage structure.

If there were only markets and hierarchies, then the solution that was agreed between Shell (the market actor) and the British Government (the hierarchical actor) would have come to pass, and the Brent Spar would now be mouldering in its watery grave (*see* Brent Spar, Volume 5). It is not; it is sitting bolt upright in a Norwegian fjord. Greenpeace, an actor from a third kind of solidarity (we call it *egalitarianism*), winged its way in, literally, and totally transformed the outcome. Since this was written, a final decision (negotiated between Shell and Greenpeace) has been reached, and the structure is now being cut up into cylindrical sections to a roll-on/roll-off ferry terminal in Norway.

Nor is it just Brent Spars that cause trouble for the simple markets and hierarchies approach. A recent analysis of how things are actually done in Himalayan and Alpine villages (Price and Thompson, 1997; this article is based, in large part, on this 1997 paper, which has a more extensive set of references than is included here) strongly suggests (as we will see in a moment) that, if these mountain farmers relied on just markets and hierarchies, neither they nor their environments would be the way they are. Nor is this inadequacy of the markets and hierarchies framework confined to economically marginal mountainsides.

The ills of the American city have recently been blamed on the public-private partnerships that were seen as the solution (Brion, 1992). Hierarchies and markets, in coming together in this cozy and unseemly way, have totally excluded community (the egalitarian solidarity) and forced the citizenry into a state of atomized, alienated subordination and systematic exploitation (*fatalism*: the fourth and rather passive, solidarity that completes the typology).

At the global level, the three active solidarities (markets, hierarchies and egalitarianism), altogether with the markedly different problem de nitions and solution de - nitions that each of them generates, are clearly discernible in the climate change debate. Indeed, they are what make that debate possible, each voice all the time de ning itself in contradistinction to the other two. Hierarchists pin the blame on population. Individualists (the supporters of market solidarity, which Cultural Theorists call individualism) see it as stemming from people being able to treat the environment as a free good. Egalitarians insist that it is pro igacy (excessive consumption, especially in the richest nations of the world) that is the root of it all.

Their solutions essentially, reduce population (hierarchy), get the prices right (individualism) and frugality (egalitarianism) are so divergent that each constitutes part of the other two s problems. Frugality, it turns out, requires the abdication of capitalism: the driving force of the individualist s solution. The population diagnosis, as far as the egalitarians are concerned, blames the victim (the South, which is where all the population growth is), and lets the guilty party, the North, off the hook. And the sorts of market interventions that both the hierarchists and the egalitarians, in their different ways, are intent on will, the individualists insist, get the prices even more wrong than they are at present! (See Chapter 4, Vol. 1 of Rayner and Malone, 1998, where the self-organization of these three voices is set out by means of a painstaking discourse analysis).

This means that human interactions with the environment cannot be effectively analyzed using theoretical frameworks that allow just one or two positions. Such frameworks are insuf ciently variegated.

This is the main practical message from Cultural Theory, and it is a highly discomforting message for policy makers generally and, in particular, for those who build the computer-based models that underlie most policy making within the broad area that is now labeled sustainable development.

In most of these models, the representation of the micro-level, the household (in energy modeling) and the farmer (in land-use modeling), is singular: an economically rational utility maximizer. Such a representation recognizes just one voice (that of individualist solidarity) and silences the other two.

More recently, modelers have progressed to the classic formulation and recognized two of the voices. The International Geosphere Biosphere Programme Land-Use and Land-Cover Change (IGBP LUCC) project (*see* **IGBP Core Projects**, Volume 2), for instance, notes that land-use and land-cover change is taking place increasingly under the in uence of the market, and that this justi es a model based on economic theory: a decentralized setup in which all agents individually solve their inter-temporal maximization problems, consumers maximizing utility, rms maximizing pro ts and so on. If the markets are competitive, so the argument runs, these agents can take prices as given, but in those instances where markets are not competitive, the optimization has to be done by government or some other higher level authority. But the third voice (that of egalitarian solidarity) is still excluded, leaving the policies that such models underpin wide open to the sorts of nasty surprises that have overtaken Shell and the government in Britain and the public private urban regeneration partnerships in the United States.

Two-voice modeling, though an improvement, is still insuf ciently variegated. Like one-voice modeling, it is still wedded to optimization and managerial control, when what is needed is constructive negotiation between all the voices: the democratization, in other words, of decision processes that have been depoliticized and treated as merely technical. This is a topic that, since the debacles over mad cow disease (bovine spongiform encephalopathy) and genetically modi ed crops in Britain, the dismantling of the Millau MacDonald s by the French sheep farmer Jose Bove, the Battles of Seattle and Prague (vehement demonstrations against the World Trade Organization), and a host of similar events around the world, is increasingly on national and international agendas. But, to actually do that democratizing, we have to avoid silencing any of the voices, and that is something that current approaches, being insuf ciently variegated, cannot do.

To help clarify what sort of differences a suf ciently variegated framework makes, and to gain a more re exive

(personal) understanding of what is going on in our own social systems and environments, we can take a close look at the surprisingly complex lives of the seemingly simple folk who live in the Himalayas and the Alps. These people, Cultural Theorists would point out, know something that the single problem-single solution merchants who tend to dominate policy-making in advanced industrialized societies have managed to forget.

Solidarities in Action

Himalayan villagers parcel out their transactions with their physical environment into four distinct solidarities, each of which is characterized by a distinct management style. Agricultural land, for instance, is privately owned whilst grazing land and forests are communally owned. But grazing land and forests do not suffer the tragedy of the commons (*see* Commons, Tragedy of the, Volume 5; Property Rights and Regimes, Volume 5) because transactions in their products are under the control of a commons' managing institution. Villagers appoint forest guardians, erect a social fence (a declared boundary, not a physical construction) and institute a system of fines for those who allow their animals into the forest when access is forbidden, or take structural timber without first obtaining permission. If the offender is also a forest guardian, the fine is doubled; if children break the rules, their parents have to pay up.

Informal though they may seem, and lacking any formal legal status, these arrangements work well in the face-to-face setting of a village and its physical resources. Drawing on their home-made conceptions of the natural processes that are at work (their *ethnoecology*), the forest guardians regulate the use of these common property resources by assessing their state of health, year by year or season by season. In other words these transactions are regulated within a framework that assumes, first, that you can take only so much from the commons and, second, that you can assess where the line between so much and too much should be drawn. The social construction inherent to this transactional realm is that nature is bountiful within knowable limits. This, to make a link with the ecological theories of Holling (1986), is the myth of *nature perverse/tolerant* (Figure 1).

With agricultural land, however, decisions are entirely in the hands of individual owners, and fields (unlike communally owned resources) can quite easily end up belonging to the moneylenders. In recent years, when forests and grazing lands have suffered degradation (for a variety of reasons, not the tragedy of the commons), villagers have responded by shifting some of their transactions from one realm to the other. For instance, they have allowed trees to grow on the banks between their terraced fields (thereby reducing the pressure on the village forest) and they have switched to stall feeding their animals (thereby making more efficient use of the forest and grazing land and receiving copious amounts of manure which they can then carry to their fields). In other words, transactions are parceled out to the management styles that seem appropriate and, if circumstances change, some of those transactions can be switched from one style to another.

Since they are subsistence farmers, whose aim is to remain viable over generations (rather than to make a

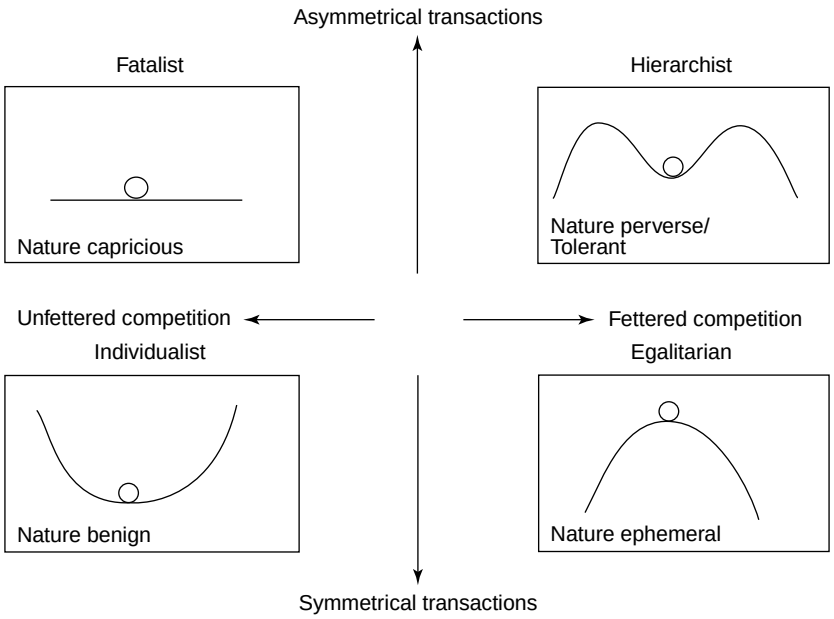


Figure 1 The Solidarities, their myths of nature and their transactional realms

killing in any one year) their transactions within their local environment can be characterized as low risk, low reward. However, during those times of the year when there is little farm work to be done, many villagers engage in trading expeditions, or in migrant labor in India. Trading expeditions are family based, family financed and highly speculative: high risk, high reward. So a farmer's individualized transactions, when added together over a full year, constitute a nicely spread risk portfolio. The attitude here (and particularly at the high risk end of the portfolio) is that Fortune favors the brave, Who dares, wins, There's plenty more fish in the sea . Opportunities, in other words, are there for the taking. The idea of nature here is optimistic, expansive and non-punitive: *nature benign* (Figure 1).

Social scientists in general, and institutional economists in particular, would see these two realms as corresponding to their classic distinction between hierarchies and markets, and would have no difficulty in explaining the processes by which some transactions are switched this way or that (though they would be surprised to find that the hierarchy was a village-level commons managing institution, not the state). But (and this is the essence of the Cultural Theory argument) hierarchies and markets do not exhaust the transactional repertoire of the Himalayan villager. Some collectivized transactions do not involve formal status distinctions (such as those between forest guardians and ordinary villagers) and some individualized transactions are marked by the absence of bidding and bargaining (an essential characteristic of the markets that are generated by the individualist solidarity). The plurality, in other words, is four-fold, not two-fold.

In many parts of the Himalayas (especially the Indian Himalayas), village autonomy is always under threat, because powerful outside actors are also laying claim to the forest resources that are so vital to Himalayan farming systems. One very effective response to this external threat has been the Chipko Movement (*see* Chipko Movement, Volume 5). This is a grassroots and highly egalitarian social movement, in which women (who are largely responsible both for fodder gathering and fuelwood collection) predominate. *Chipko* means to stick, and the Gandhian strategy is to physically hug the trees, thereby preventing them from being appropriated. Those villagers of a slightly less non-violent disposition actually chase the logging contractors (and the government forestry officers who have been corrupted by the contractors) out of the forest with their *kukris* (long curved knives). In the Narmada Valley, farther to the south (where a vast development project is under way), they have now done the same to the representatives of the World Bank: a South Asian counterpart to the Brent Spar surprise. (Indeed, the World Bank pulled out in 1993 but the project is still being promoted by Indian State Government and market borrowings.)

So far as these threatening external transactions are concerned, it is certainly not a case of plenty more fish in the sea , nor is there even a safe limit, within which the commercial extraction of timber would be sustainable. All external predation is seen as catastrophic in its consequences. Hence the spectacularly uncompromising collectivist response of the tree huggers, whose idea of nature is one in which any perturbation of the present low-key regime is likely to result in irreversible and dramatic collapse: *nature ephemeral* (Figure 1).

Finally, in every village, we may be sure, there will always be some people who sneak wood from the forest when no one is looking, who can never quite get together the capital, the contacts and the oomph to go off on trading expeditions, and who manage somehow not to be around when it's all hands to the tree hugging. These are the fatalists: people whose transactions are somehow dictated by the organizational efforts of those who are not themselves fatalists. Theirs is a life in which the world is always doing things to them (sometimes pleasant, sometimes unpleasant) and in which nothing that they do seems to make much difference. Why bother? is the not unreasonable response of the fatalist. If that is how the world is, then learning is not possible and, even if it were, there would be no way of benefiting from it. The idea of nature here is one in which things operate without rhyme or reason: a flatland in which everywhere is the same as everywhere else: *nature capricious* (Figure 1).

From Simple to Complex

Completing the typology with these two solidarities (egalitarianism and fatalism) makes some important differences. For instance, once we understand egalitarian solidarity, we can avoid the sorts of surprises that have been visited upon the Brent Spar and the Narmada River Project. And we can see that, only if all the transactions are in the fatalistic realm (the one realm where learning is not possible), would the prevalent assumption (evident, for instance, in the hierarchist's diagnosis of the climate change problem) and the IPAT equation (*see* Modeling Human Dimensions of Global Environmental Change, Volume 5) that there is a direct relationship between population increase and environmental degradation hold true. But there is much more to it than this.

Change, in the conventional theory, is deterministic. If you're knocked out of hierarchy, you'll end up in the market, and *vice versa*. But, in Cultural Theory, change is indeterministic: leave A and you can end up at B, C or D, and when you leave whichever one of these you have arrived at, there are three possibilities, on and on.

Conventional theory treats human systems as simple (linear, deterministic, insensitive to initial conditions, equilibrium seeking, and predictable); Cultural Theory treats them as complex (non-linear, indeterministic, sensitive to initial conditions, far from equilibrium, and unpredictable).

Simple systems are manageable in the sense that, once we understand enough about them, we can define some desirable state of affairs (*sustainable development* is the current favorite) and then steer the totality towards it. But this, as our next example makes clear, is not possible if the system is complex.

A Swiss Example

Moving from the Himalayas to the Alps, we nd much the same four-fold allocation of transactions, with agricultural land being privately owned and grazing land (and sometimes the forests) being communally owned. But the Swiss forests, unlike those of the Himalayan villagers, are physically sandwiched between the high pastures (communally owned) and the valley oor (privately owned elds, houses and hotels). Over the centuries that the Davos valley has been settled, to take a speci c locality, both the elds and the grazing land have expanded at the expense of the forest. The trees on the steeper slopes have stayed in place, acting both as a source of timber and as a barrier against avalanches. However, it is dif cult to achieve both these functions simultaneously. The Davosers have often set in train changes in the forest s age structure which, decades later, have resulted in exceptional avalanches reaching the valley oor and threatening the destruction of the entire community.

Every time this unpleasant surprise has befallen them, the Davosers have responded by switching their forest management onto the all in the same boat egalitarian style. Later, it has sometimes shifted to the hierarchist style, often to the individualist style (with farmers owning long thin strips of forest running all the way from valley oor to alpine pasture), and sometimes to the fatalist style (as happened, for instance, when the avalanche danger was clearly perceived yet extraction continued in response to the demands of various mining booms and, in more recent years, the demand for ski-runs).

Surely, you might think, they would have got it right by now. To think that is to assume that there is one right way; but as Cultural Theory shows us, that is not the case. There is no way of ever getting it right, because managing one way inevitably changes the forest, eventually to the point where that way of managing is no longer appropriate. This would happen even if there were no exogenous changes (like the mining and tourist booms) which, of course, there always are (even in seemingly remote places like the Himalayas). Viability can only be achieved, therefore, by covering all the bases: by the villagers ensuring that they have the full four-fold repertoire of management styles, and by their being prepared to try a different one whenever the one they are relying on shows signs of no longer being appropriate. The Davosers, like their Himalayan counterparts, have now been in their valley for more than 700 years, without destroying either themselves or their valley in the process. This achievement would not have been possible if they had opted for just one management style, or even for the two that the prevalent orthodoxy allows!

Multi-vocality

Himalayan and Alpine villages, with their transactions parceled out in these four very different ways, are impressively multi-vocal. More than that, as is evident from the examples of stall feeding and trees on private land (in the Himalayas) and of alternative forest management styles (in the Alps), they have the ability to switch transactions from one way to another whenever it seems likely that this might be more appropriate. Since the behavior of the villagers is continually altering the resource base on which they depend, their villages would not be viable if they did not have this in-built (messy, noisy and argumentative) mechanism. Schapiro (1988) has dubbed this sort of setup (in which each conviction as to how the world is, each myth of nature, is given some recognition) a *clumsy institution*. This is in contrast to those more elegant, and more familiar, arrangements (tidy, quiet and suavely consensual) in which just one conviction holds sway. The terminology is deliberately counter-intuitive, clumsy institutions having some remarkable properties that are not shared by their unclumsy alternatives.

To understand just how remarkable clumsy institutions are, imagine for a moment that you are some God-like experimenter, able to reach out and change this is or that variable in a Himalayan village s environment, or to move it bodily east or west, north or south, across the convoluted landscape. As you bring in the logging contractors, or take it 100 km eastwards or 1000 m higher, the village will shift its transactions this way or that between its four options until it has adapted itself to its changed circumstances. In other words, it will maintain its viability thanks to the very practical learning system that is part and parcel of its four-fold plurality. If the village did not have this plurality, and was an elegant and unclumsy institution, like many national forestry services, including Britain s Forestry Commission (Tomkins, 1989) and the United States Forest Service (Hirt, 1994), it would not be able to do this. Something along these imaginary lines, it turns out, is what has actually happened, and continues to do so.

As we go from one Himalayan village to another, the relative strengths of the four ways of organizing vary. Egalitarianism, for instance, is strongest in those parts of the

Himalayas that are most prone to commercial logging. As one moves eastwards, from India (with its powerful center and its colonial heritage of Reserved Forests) into Nepal and Bhutan, so the Chipko Movement and its counterparts become less of a force to be reckoned with. If the inequitable external threat is absent then so too, it appears, is the communitarian response to it. However, the most dramatic of these variations is north–south: between the strongly individualized Buddhist villages and the strongly collectivized Hindu villages a day or two's walk downstream. These are Furer-Haimendorf's (1975) adventurous traders and cautious cultivators, respectively: apt characterizations which readily map onto two of the four social beings (individualists and hierarchists, respectively) that we have described above.

Cultural Theorists, we should explain, use the term *social being* to describe the behavior to which an individual must conform, and the convictions that he or she must espouse, to help maintain the form of social solidarity to which he or she belongs. The prime mover, therefore, is not the individual (the psycho-physiological entity) but the form of social solidarity. Thus, the terms hierarchist, individualist, egalitarian, and fatalist denote available roles (or management strategies) that individuals step into, or out of, as their daily lives, or the changing seasons, take them from one transactional realm to another (Figure 2). (See also Box 1: A Swiss villager's day). Though all four roles are present in both the Buddhist and the Hindu villages, more transactions are in the individualist quadrant in the former, and in the hierarchist quadrant in the latter.

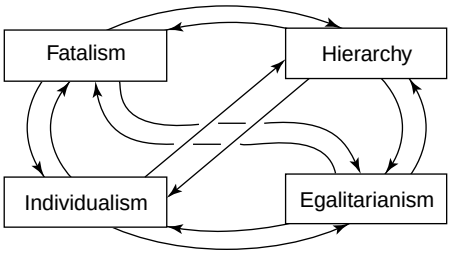


Figure 2 Transitions between solidarities and their associated transactional realms (individuals are able to move without causing change. It is when some transactions are shifted from one realm to another that we get change)

Furer-Haimendorf (1975) shows how the small agricultural surpluses of the cautious cultivators become the payloads of the adventurous trader's yaks as they set off on their journey into Tibet, and how the salt they bring back eventually finds its way to the cautious cultivators

Box 1 A Swiss villager's day

During the growing season an individual may on one day milk his cows, cut hay, thin saplings, maintain an avalanche control structure and wash dishes in a restaurant. The cows though privately owned are grazed on pasture owned by a specific set of long-established families. The hay is on his own private field; the saplings are part of a forest owned by another set of families; the avalanche control structure is on private land but maintained by agreement by the village; and the restaurant is owned by a multinational hotel chain.

This framework is fairly stable from season to season but the individual has a very different pattern of activity in the winter when the cows live in his private byre and much of the land is snow covered and barely used unless the valley includes a ski resort. If it does, then he has opportunities for work without leaving the valley. If not then he may leave to work elsewhere, thereby reducing the use of scarce resources at home. Thus in winter the human ecosystem centered on the valley is concurrently simpler and wider.

So our Swiss villager has a portfolio of transactions and management styles that fluctuates with the seasons and also with the longer-term dynamics (such as those that in altering the age-structure of the forests can eventually shift a whole category of transactions from one style to another).

- Like his Himalayan counterpart, he owns his hayfields and cows. These are private property; he can buy or sell them acting as an individualist, subscribing to the myth of nature benign.
- Coming from an old-established family, he is a member of a forest cooperative (Waldgenossenschaft) that gives him specific rights to cut trees and imposes a duty to maintain the forest. He is also a member of a pasture cooperative (Alpgenossenschaft) which annually decides the grazing season and the number of animals he may graze and requires him to contribute to the cowherd's upkeep. These are small-scale hierarchical institutions which have developed over the generations (in between the periods when the forests are privatized and their associated transactions transferred to the more exploitative individualist management style) in response to the limitations as well as the opportunities imposed by the natural environment: nature perverse/tolerant.
- As a voting member of the commune, he also has a duty to maintain resources that contribute to its survival such as the avalanche control structures that protect houses, fields and roads from damage. This tends to be an egalitarian involvement which recognizes that when it comes to these sorts of hazards, all the members of the community are in the same boat and that each should contribute his equal share: nature ephemeral.
- Lastly as a dishwasher in a multinationally owned restaurant, he is effectively a replaceable fatalist. His involvement is necessary if the enterprise is to continue but he has no interest in its future nor it in his and he can be paid off at any time (he will almost certainly lose his job at the end of the summer season).

who cannot produce this vital commodity. The distinctive strategy of each thus makes viable the other's, and we begin to see how it is that each village, in adjusting to its circumstances (which include the other villages), creates and takes its place in a social and cultural ecosystem, in which the marked divergence of the parts sustains the whole. Nor is this a fanciful analogy. As we show below, the adventurous traders' strategy matches that of the omnivorous and opportunistic r-selected species; the cautious cultivator's strategy matches that of the specialized and niche-dependent K-selected species (*see* r–K Strategies, Volume 2). The fatalists do for social systems what compost does for natural systems (provides a generalized resource for renewal). The egalitarians, through their small-scale communal fervor, are creating enclaves of low-level energy (what Marx, 1967, called primitive capital) in places where neither the r-selected nor the K-selected species can make any impression (Holling, 1986; Thompson *et al.*, 1990; Holling *et al.*, 1993).

So the ambitious hypothesis that is being sketched here is very different from the way people usually think about the interactions of social and natural systems. There is, on this view, no way of ever getting it right: of bringing the social into long-term harmony with the natural (which, of course, is the whole idea behind sustainable development). Instead, each is a four-fold and plurally responsive system, and their time-lagged interactions ensure that there can be no steady-state outcome. The whole system is in a perpetual unsteady state: changes at each level (the social and the natural) adapting to the other and changing it in the process, thereby setting in motion another set of changes. On and on. Nor are these changes predictable, as they would be if each level had only two possible states: hierarchies and markets, for instance, or, as is discussed below, their ecological analogues. Order without predictability (as opposed to transition from A to B, or oscillation between A and B, that the two-fold hypotheses give us) is the crucial idea behind this Himalayan story.

THEORIES OF CHANGE THAT MAKE CHANGE PERMANENT

Change, Cultural Theory argues, occurs because the four forms of social solidarity are not impervious to the real world. Just because people insist that the world is as their myth of nature tells them it is, it does not follow that the world really is so. If it is, that is fine, but if it is not, they have an uphill struggle. Surprise (the outcome of the ever widening discrepancy between the expected and the actual) is of central importance in dislodging people (and their transactions) from their form of social solidarity. And it is these various mismatches between what a way of life promises and what it delivers that continually tip people (and transactions) out of one form of social solidarity and into another. Of course, this hypothesis does require that the world, at times and in places, be each of these possible ways: otherwise we would all eventually end up surprised into the single true way. And it would help the surprises to continue indefinitely if the world itself kept changing.

Neither of these suggestions, some ecologists would argue, is particularly far-fetched. Holling (1986) and Holling *et al.* (1993) for instance, have elaborated the notion of requisite variety into a powerful critique of the conventional idea that the *climax community*, the ecosystem in which each specialized species has its stable and ordered niche, is the end of the organizational road. This critique exactly parallels Cultural Theory's dissatisfaction with the conventional hierarchies and markets account of things, in that it argues that there must be four, rather than just two destinations. Holling's critique is that the climax community eventually complexifies itself to the point where it undermines its own stability: an inevitable collapse, which has been proved mathematically by May (1982). This does not mean that an entire climax community (the Amazon rain forest, for instance) will suddenly disappear, but it does require any climax community to be patchy: to always include some localized areas in collapse as, for instance, happens when a mature tree crashes to the ground.

At this catastrophic moment, all the energy that is tied up in all the niches and interdependencies of the climax community is released. Holling, well aware of the parallel with Schumpeter's (1950) theory of economic maturity, collapse and renewal, refers to the transition from the climax community to compost, as *creative destruction*. Nor, he argues, is this the end of the road. With the whole place suddenly awash with capital (loose energy), the challenge is to fix it before it all disappears, by soil leaching, for instance. This, of course, is where the unspecialized and cooperative fence builders (microorganisms mostly) come into their own, gathering up the loose energy into small bundles that, as yet, have no connections with one another. But even this is not the end of the road, because the stage is now set for the appearance of yet different ecological players. These are the unspecialized but opportunistic, fast breeding and highly competitive r-selected species. These generalized exploiters (weeds, rodents and so on) are able to harness all the energy gradients that are now in place between all these unconnected bundles of energy. But these r-selected species, as they exploit and colonize this environment, inevitably begin to push it into a rather more patterned and interconnected state, thereby making it less conducive to their way of doing things and more suited to the sort of energy-conserving strategies that characterize the K-selected species: those specialized, slower breeding and often symbiotic, plants and creatures, which are the vanguard of the complex and increasingly ordered whole that constitutes the climax community.

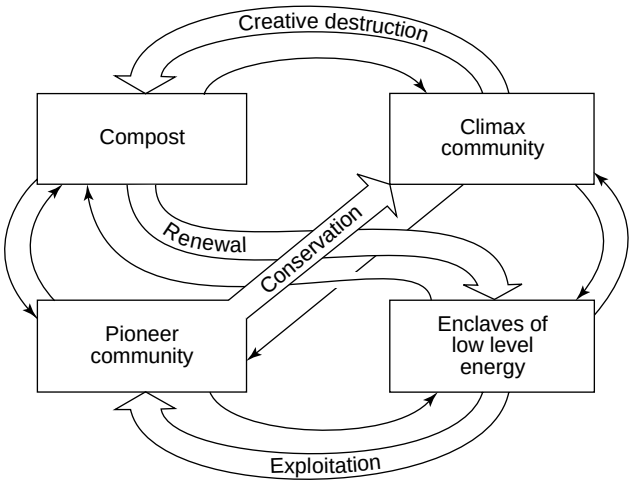


Figure 3 The complex critique of the conventional assumptions about natural systems (redrawn from Holling, 1986 to be homologous with Figures 1 and 2)

In other words, once you bump up the number of ecological strategies from two to four, there is no end to the road. Instead, there is a never ending set of transitions (twelve in all) that exactly parallels (in terms of dynamics, not substance) the social transitions of Cultural Theory. Holling goes on to argue that, while all twelve of these transitions do happen, there is a tendency for some to predominate at certain stages, thereby creating a fairly regular sequence of transitions: from specialized interdependence (the climax community) to unstructured fragmentation (compost) to unspecialized cooperation (energy fixing) to unspecialized competition (the pioneer community) to specialized interdependence (the climax community, again) and so on. He calls this sequence an *ecocycle*, and its description (which can be supplemented with descriptions of all the other cycles that are possible but, Holling believes, less pronounced) helps us to see the gulf (some might use the expression *paradigm shift*) that separates this model of change from the conventional one (Figure 3).

Is there, then, a socio-cultural equivalent of Holling s ecocycle? Yes, there is, and it can be most easily set out by reference to the theory of surprise: the theory that provides the bridge between the institutional and the natural: between us and the rest of nature.

ALWAYS LEARNING, NEVER GETTING IT RIGHT

A myth of nature provides its holder with a way of seeing the world and with a way of not seeing it. This means that, if the world happens not to be the way the myth holder is convinced it is, he or she will not notice this discrepancy straight away. Enlightenment, therefore, is always time lagged and, since it results in the enlightened one being tipped out of one quadrant of the Cultural Theory scheme and into one of the other three, it comes as something of a shock: a surprise. Surprise, in other words is always relative, which explains why, whenever something unexpected befalls us, there is always someone who saw it coming! The theory of surprise (Thompson and Tayler, 1986; Thompson *et al.*, 1990) is built on this relativistic, but far from unconstrained, foundation:

- an event is never surprising in itself;
- it is potentially surprising only in relation to a particular set of convictions about how the world is;
- it is actually surprising only if it is noticed by the holder of that particular set of convictions.

For instance, an individualist, whose myth assures him that an ecosystem is so robust that it will recover from any perturbation, will be surprised when it collapses catastrophically. Similarly, a hierarchist who is convinced that all ecosystems can be managed, with predictable results, will be surprised when this turns out to be untrue. Conversely, an egalitarian, who believes that nature is precarious, will be surprised when those who have disregarded precautions do not reap the expected disaster. And a fatalist will be surprised if bene ts, which he expects to be randomly distributed, continue to arrive.

Thus, surprises may be either pleasant or unpleasant, and a never ending sequence from one myth of nature to another may be proposed. Though all twelve transitions (see Figure 2) are possible, and we cannot say for sure what their order will be, we can spin a story to help us understand what is going on by privileging one particular sequence of possible changes so as to generate the socio-cultural analogue of Holling s ecocycle.

- Let us start with *nature benign*. In this state of the world, there is an excess of opportunity over existing investment, and this state, when interrogated by the myriad actions of individual agents, results in a positive-sum game in which a hidden hand keeps adding to the welfare of the totality. As long as the excess continues (that is, as long as there is no rim to the deep basin that contains the ball), and learning by experimentation continues, we have the state of affairs assumed by neoclassical economics.
- But, as they say in show business, Nothing recedes like success and eventually exploitative behavior causes the upper edge of the basin to turn downwards: *nature perverse/tolerant*. The excess, for some actors, has now vanished. Transaction costs rise steeply, innovation brings losses more often than pro ts, and markets fail. This is the transition from markets to hierarchies described by the new institutional economics (Williamson, 1975).
- Hierarchically sustained transactions, in their turn, transform the environment that ushered them in and

eventually the pocket of stability implodes: *nature capricious*. Both hierarchy and individualism (which has, of course, survived in the pocket) now start to lose their transactional grip and the world becomes a confusing, contradictory and unpredictable place: a place of which the fatalist's attitude; Why bother? makes perfect sense.

- This flatland, however, is less hostile to those small, egalitarian and self-disciplined groups that strive to bring their needs down within what they perceive to be Mother Nature's frugal limits, and these groups are therefore well placed to take advantage (though that is not quite the right word) of the recessive realities that are about to overwhelm the conventional institutional arrangements: the hierarchies and the markets.
- In this next stage, *nature ephemeral*, all increases in scale bring punitive diseconomies, and the economy (like the universe that contains it) winds down and down. The entropy principle (Georgescu-Roegen, 1971) and the dictum *small is beautiful* (Schumacher, 1973) make economic sense (*see* **Small is Beautiful**, Volume 5). Yet, no matter how lightly everyone treads on the Earth, the ball eventually rolls down the slope, coming to rest in some other basin (*nature benign*) and we are back where we began: in a positive sum world that rewards the bold and skilful and that brings increasing returns to those who are prepared to act expansively.

Throughout this process, changes in the environment result from the actions of those whose strategy happens to be best suited to making the most of the environment in which they find themselves. As more and more of these strategists act, these endogenous changes accumulate, and the environment passes over a threshold into a state better suited to one of the other strategies, *ad in nitum*. Though this complex model may start at the same place as the simple one and have the same dynamics, its paths are infinitely more surprising and unpredictable. In this inherently complex system, in which ecological and sociocultural components interact, each myth of nature captures some aspects of the world at some time. No one of them is ever right all the time and everywhere, and this means that each has its vital part to play. Clumsy institutions nurture that vitality; elegant ones destroy it.

REFERENCES

Brion, D J (1992) The Meaning of the City: Urban Redevelopment and the Loss of Community, *Indiana Law Rev.*, **25**(3), 685–740.

Clements, F E (1916) Plant Succession, an Analysis of the Development of Vegetation, *Carnegie Inst. Wash. Publ.*, **242**, 1–512.

Durkheim, E (1893) *De la Division du Travail Sociale: étude sur l'Organisation de Sociétés Supérieurs*, Alcan, Paris.

Fischer, F and Hajer, M A (1999) *Living with Nature: Environmental Politics as Cultural Discourse*, Oxford University Press, Oxford.

Georgescu-Roegen, N (1971) *The Entropy Law and the Economic Process*, Harvard University Press, Cambridge, MA.

Hirt, P W (1994) *A Conspiracy of Optimism*, Nebraska University Press, Lincoln, NE.

Holling, C S (1986) The Resilience of Terrestrial Ecosystem: Local Surprise and Global Change, in *Sustainable Development of the Biosphere*, eds W C Clark and R E Munn, Cambridge University Press, Cambridge, 292–320.

Holling, C S, Gunderson, L, and Peterson, G (1993) *Comparing Ecological and Social Systems*, Beijer Discussion Paper Series No. 36, Stockholm Beijer International Institute of Ecological Economics.

Lindblom, C (1977) *Politics and Markets: The World's Political-Economic System*, Basic Books, New York.

Maine, H S (1861) *Ancient Law*, John Murray, London.

Marx, K (1859/1967) *Capital*, Basic Books, New York.

May, R (1972) Will a Large Complex System be Stable? *Nature*, **238**, 413–14.

Odum, E P (1969) The Strategy of Ecosystem Development, *Science*, **164**, 262–270.

Price, M F and Thompson, M (1997) The Complex Life: Human Land uses in Mountain Ecosystems, *Global Ecol. Biogeogr. Lett.*, **6**, 77–90.

Rayner, S and Malone, E L (1998) *Human Choice and Climate Change*, Battelle Press, Columbus, OH.

Schapiro, M (1988) Judicial Selection and the Design of Clumsy Institutions, *South. Calif. Law Rev.*, **61**, 1555–1569.

Schumacher, E F (1973) *Small is Beautiful: a Study of Economics as if People Mattered*, Blond and Briggs, London.

Schumpeter, J A (1950) *Capitalism, Socialism and Democracy*, Harper and Row, New York.

Thompson, M, Ellis, R, and Wildavsky, A (1990) *Cultural Theory*, Westview, Boulder, CO.

Thompson, M and Tayler, P (1986) *The Surprise Game: an Exploration of Constrained Relativism*, Warwick Papers in Management, No. 1, School of Industrial and Business Studies, University of Warwick, Coventry.

Tompkins, S (1989) *Forestry in Crisis*, Christopher Helm, London.

Tonnies, F (1887) *Gemeinschaft und Gesellschaft*, Wissenschaftlich, Darmstadt, translated by C P Loomis (1957) as *Community and Society*, Michigan State University Press, East Lansing, MI.

Turner, II, B L, Clark, W C, Kates, R W, Richards, J F, Matthews, J T, Meyer, W B *et al.* (1990) *The Earth as Transformed by Human Action*, Cambridge University Press, New York.

von Furer-Haimendorf, C (1975) *Himalayan Traders: Life in Highland Nepal*, John Murray, London.

Weber, M (1922) *Wirtschaft und Gesellschaft*, Mohr, Tubingen, translated by E Bischoff *et al.* (1978) *Economy and Society: an Outline of Interpretive Sociology*, University of California Press, Berkeley.

Williamson, O (1975) *Markets and Hierarchies, Analysis and Anti-Trust Implications: a Study in the Economics of Internal Organization*, Free Press, New York.

Management of Nuclear Wastes

see Nuclear Waste: Geological Issues (**Volume 3**)

Mitigation

see Adaptation Strategies (*Opening essay*, **Volume 4**)

Modeling Human Dimensions of Global Environmental Change

Marco A Janssen
Vrije Universiteit, Amsterdam, Netherlands

The history of humankind is a continuing record of interactions between people's efforts to improve their well being and the environment's ability to sustain these endeavors. Environmental constraints have led to innovations and social development, as well as social stagnation and human suffering. While the interactions throughout most of history were on a local scale, during the last decades awareness has grown so that the complexity and increasing scale of the interactions are demanding new forms of environmental management. New threats to mankind emerge, such as climate change, acid rain, ozone depletion, resource exhaustion, reduction of biodiversity, and limits to the availability of food and unpolluted fresh water. In fact, the globe is changing rapidly due to human activities, and humankind is, and will increasingly be, suffering from global environmental change.

One response from the scienti c community to understand the relations between human activities and the environment is the use of modeling, that is, constructing formal descriptions of natural and societal change and their mutual interactions. Although making accurate predictions for long term future developments is inherently impossible, models can help us to show the interdependence of various activities and consequences in time and space. In that way, models can be used to communicate information and insights from the scienti c community to policy makers and other stakeholders.

A number of fundamental issues in modeling the human dimensions of global environmental change are discussed. These dimensions relate to the interactions between humankind and the global ecosystem. How do human activities change our environment, how are humans affected by changes in our environment, and how does humankind respond to these changes? This contribution focuses on the behavior of humans in relation to their environment. Moreover, two fundamental related issues are discussed:

- *Given our limited knowledge of reality, we have to make all kinds of subjective assumptions about the functioning of human and natural systems in order to make decisions. How can models be of help for decision-making when subjectivity is unavoidable in developing models?*
- *On the macro-scale, phenomena are observed which are the result of actions of agents on lower levels such as of households, rms, organizations and nations. How can we explore future developments of these macro-scale phenomena?*

Although there is an increased use of modeling human dimensions of global environmental change, these issues are not well captured in current mainstream modeling practices. These issues are discussed in the context of developments in modeling human dimensions during the last 30 years and address promising developments for the future.

There are many kinds of models. A general classification is the distinction between formal mathematical models and non-formal models such as stories and cartoons. Sometimes people are role models of how one might live, such as Nelson Mandela or Claudia Schiffer. Only the formal models are dealt with here. The advantage of these formal models is that they are explicit and many of them are computer models that can be used to do repeated experiments.

Formal models are used in science very frequently, especially since the seminal work of Isaac Newton, more than 300 years ago. Formal models were mainly used to describe natural systems, such as calculating the trajectories of cannon balls or celestial bodies. Due to the success of these models, mathematical models were increasingly used in social science, especially in economics. However, human beings are not similar to cannon balls. Human beings can decide to obey traffic laws, but cannon balls cannot decide to obey the law of gravity.

The application of models from physics to social phenomena is problematic, but is still widely used. This article addresses new ways of modeling social phenomena by using multi-agent modeling to simulate interactions between agents, which are behavioral entities such as persons, households and firms. But first the field of modeling human dimensions of global environmental change is discussed.

Currently, models are widely used to describe the relations between human activities and the environment. Moreover, these models have often played a central role in setting environmental problems on the policy agenda, by exploring the consequences of alternative scenarios and by designing acceptable solutions for managing environmental problems. An example of an environmental problem for which the use of models is central is climate change. The possibility of human induced climate change is actually a policy problem that was put on the agenda after alarming model-based studies. Svante Arrhenius estimated at the end of the 19th century the consequences of a doubling of atmospheric carbon dioxide on the global mean temperature to be about 3 °C. Since the late 1950s, atmospheric carbon dioxide has been measured systematically, and currently the level is about 30% higher than pre-industrial levels. During the last 30 years, more detailed climate system models have been developed, and the results are compared with the increasing amount of (satellite) data. Still, these models are not able to describe observed climate accurately on a detailed spatial level. Besides increased efforts in modeling the climate systems, models of the human dimensions of climate change have been developed. These models were used to speculate on the possible consequences of climate change on economic growth, agricultural production and human health. Furthermore, models were developed to estimate the costs of mitigating so-called greenhouse gases. The impacts of climate change on various social and ecological factors are based on laboratory experiments, extrapolations from field studies, historical (regional) climate-change analogues and expert judgements. The increased insights into potential impacts of human induced climate change led to the current high position of the climate change issue on the policy agenda.

Policy decisions related to climate change are mainly determined by model outcomes. Many uncertainties, speculative assumptions and lack of information surround these outcomes. Thus, because the potential consequences of climate change as estimated by models are so severe, the precautionary principle is often advocated to reduce potential risks. Nevertheless, it is clear that the many uncertainties that surround the model outcomes have generated an intensive debate. Some scholars highlight the potential benefits of climate change on, for example, agricultural production, due to higher levels of atmospheric CO_2 and more suitable temperatures in Canada and Russia. Others argue that human induced climate change will not occur due to damping feedbacks of the biogeochemical cycles. Various scholars warn of positive feedbacks that may amplify climate change to catastrophic levels. The mainstream opinion on the size and impacts of climate change is represented by the reports of the Intergovernmental Panel on Climate Change (IPCC), an international scientific organization established in 1988 to assess scientific research on climate change. In 1995, the IPCC concluded that the balance of evidence suggests a discernible human influence on the climate. Although science is not a democratic process, the assessments of the IPCC try to synthesize a mainstream picture of the climate change issue which can be used in the international policy process to formulate policies to reduce greenhouse gas emissions. Such an authorized synthesized picture on the problem is desired by policy makers, since they have difficulty in handling the many uncertainties and complexities of the climate change problem. The model-based analysis of the IPCC suggests objective assessment. The idea that models can create objective predictions of the future is widespread among stakeholders who deal with global environmental change. Some stakeholders who doubt this objectivity do not want to use any model analysis at all. However, the traditional view that models guarantee or suggest objectivity is outdated. Model analyses, especially related to human dimensions, are highly influenced by subjective assumptions and interpretations. The challenge is to design ways to use models in such a way that they improve our understanding of reality. Improved mental models can improve decision-making. Therefore, the model-projections itself are not the most important element of model; rather it is the modeling process and the end use of models. This debate mirrors the discussion on world models during the early 1970s.

WORLD MODELS

Integrated models addressing the human dimensions of global environmental change elaborate a tradition that was founded in the early 1970s by the Club of Rome (*see* Club of Rome, Volume 4). Over the past 30 years many models have been built in the tradition of system dynamics, as well as other modeling techniques. In the early days the models of the Club of Rome were criticized as being based on too few empirical data, too high an aggregation level, and too many subjective assumptions. The criticism of the earlier models still holds for many of the current modeling activities.

The models of the Club of Rome were the so-called World models. Jay Forrester developed the World 2 model during the summer of 1970 based on his system dynamics approach. According to this approach, the world can be described through a conglomeration of interacting feedback loops. The World 2 model can be considered as a first sketch of a world model, without empirically based estimation of suggested causal relations.

A larger project, led by Dennis Meadows, resulted in the World 3 model, and the influential book *Limits to Growth* (Meadows *et al.*, 1972). The World 3 model contains the resource sectors, population, pollution, capital and agriculture on an aggregated global level. The standard run of the World models is one of growth followed by collapse

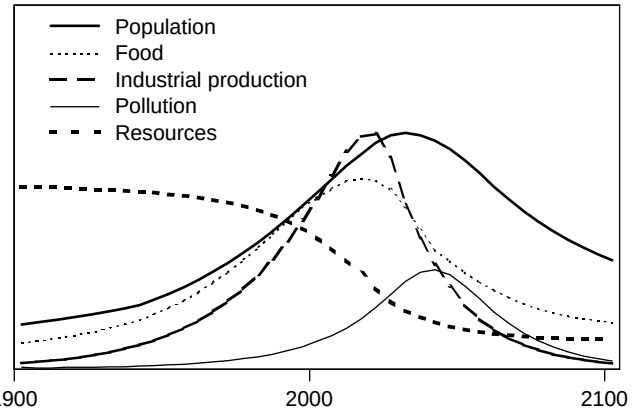


Figure 1 Standard Run of World 3. (Adapted from Meadows *et al.*, 1972)

(Figure 1). The collapse occurs because of non-renewable depletion. The industrial capital stock grows to a level that requires enormous input of resources, more and more capital must be used to obtain those resources, leading to less re-investment, and finally the collapse of the industrial base. Population decreases when the death rate is driven upward by a lack of food and health services. If the resource base is assumed to be much larger, collapse happens only a few years later due to food shortages and/or high pollution levels. Although it was recognized that there are various shortcomings in the models, Meadows and his colleagues argued that the model behavior was fundamental and general, and sufficiently developed to be of some use to decision-makers.

Managing Uncertainty, Complexity and Incomplete Information

Numerous scholars have criticized the World-model studies. Economist WD Nordhaus classified the World 2 study as misleading since it was not empirically tested enough and ignored mainstream economic insights. Other scholars concluded that due to the scarcity of relevant empirical information, relationships in the world models are subjective. Given the uncertainties, other sets of equally plausible assumptions can lead to a completely different picture. In fact, the outcomes of the models are largely the mental models of the model builders.

Some of the criticism was misplaced. The model outputs were interpreted as predictions, not merely as scenarios, i.e., what if' futures. So, when the predictions did not come true, scientists were blamed for inaccurate doomsday forecasts. Actually, the *Limits to Growth* scenarios had a profound effect on the public and government.

Scientific criticism on the World models of 30 years ago mainly concentrate on two topics. First, subjective assumptions had to be made about model relations and parameters due to incomplete knowledge or even ignorance. This was especially important for the linkages between subsystems. Examples were the effects of pollution on health, the interaction between demographic and economic dynamics, the role of technological innovations in resource availability, and the relations between material and energy inputs and economic output. Second, the complexity of the underlying subsystems and the linkages made it questionable whether a high aggregation level can lead to meaningful and relevant interpretations and results.

These problems of subjectivity and aggregated relations are still relevant for the current use of models. Although the use of models has increased, we still have no satisfactory solutions of how to manage incomplete information and insights, large uncertainties and different scales. This is illustrated here by the characteristics of the current generation of global models.

Integrated Assessment Models of Global Change

Global modeling re-emerged during the early 1990s as integrated assessment modeling, mainly because of the importance of the global climate change problem (Janssen, 1998). Integrated assessment models try to describe quantitatively as much as possible of the cause–effect relationships and the cross-linkages and interactions between the elements of the world system. Integrated assessment models are usually composed from meta-models of various subsystems. A meta-model is a simplified, condensed version of a more complicated and detailed model, which provides approximately the same behavior as the expert model from which it is extracted.

Integrated assessment models are one of the tools in the toolkit of integrated assessment. Other tools are policy exercises, dialogues between science and policy people, data analysis, scenario analysis and expert models. Integrated assessment is therefore a broader approach aimed at helping prioritize policy-making and research activities and giving insights into uncertainties and missing links of knowledge. It is used in a process whereby knowledge from a variety of scientific disciplines is combined, interpreted and communicated, with various stakeholders such as scientists, policy makers and non-government organizations involved.

The integrated models that are used describe the whole cause–effect chain from economics, energy production, emissions, land use changes, to changes in the biogeochemical cycles, the climate system, and impacts of climate change on human activities and the environment. There is no single approach to capturing the complexity of the system as a whole. In general, two types of approaches can be considered. The first approach, *process-oriented modeling*, is rooted in natural science and simulates the consequences of economic development on energy production, land use changes, biogeochemical cycles, climate system and impacts of climate change. The models are

often detailed at the spatial and temporal level. The other approach, *cost- bene t modeling*, is rooted in economics and maximizes discounted long-term welfare. Models using this second approach describe the physical consequences in less detail, but express the impacts of climate change and the efforts to reduce emissions in monetary units in order to derive an optimal response by balancing costs and benefits.

Both approaches are confronted with the same dilemmas as the World models of the early 1970s. Although much more empirical information is available, many components in models are surrounded by large uncertainties. Therefore, subjective assumptions have to be made. Also the complexity of scales remains an unsolved issue.

Besides these problems, most integrated assessment models have limited abilities to capture the broad human dimensions of climate change. The models are suitable for generating projections of economic output and the costs of climate change, but they are limited in many ways. Simple macro relations are assumed between economic activities and physical processes, but empirical insights are conflicting. Current (economic) models focus on utility of consumption as the driving force of human behavior, although it is questionable how this is related to the quality of life. There is limited insight into the physical dimension of satisfying human needs, but this understanding is needed before something plausible can be said about decoupling economic development and environmental pressure. Current integrated models include only the free market as an institution, although there are many other forms of economic organization and institutions.

One of the reasons for the limitations of the current generation of integrated models is the use of a rather mechanistic modeling paradigm, which is not able to include novelty, evolution and surprise. The role of modeling paradigms is now discussed in more detail.

MODELING PARADIGMS

A lot of controversies among modeling studies are caused by the difference in modeling approach that is adopted by the various scientists. Each modeling approach, or in a broader context, each modeling paradigm, involves its own set of theories, concepts, mathematical techniques, and accepted procedures for constructing and testing models. We can distinguish between deterministic and stochastic models, simulation and optimization models, reductionistic and integrated models, linear and non-linear models, one-agent and multi-agent models, and so on. Instead of discussing all kinds of possible paradigms, the difference between the reductionistic Newtonian approach and the complex adaptive systems approach is examined. This illustrates the transition in science that is currently occurring and explains fundamental differences on how to use models for assessing the future.

Mathematical modeling has long been influenced by physical science, which has developed a mechanistic, reversible, reductionistic and equilibrium-based explanation of the world. This proved to be very successful in calculating trajectories of moving objects (e.g., cannon balls) and predicting the positions of celestial bodies. The work of Isaac Newton, culminating in the *Principia Mathematica Philophiae Naturalis* in the late 17th century, was, and still is, very influential. The associated rational and mathematical way of describing the world around us was also applied in the social sciences, economics and biology. Despite the fact that later developments in the natural sciences seriously constrained the applicability of the mechanistic paradigm, its relative simplicity had a great appeal to scientists from various disciplines working with models. However, despite the widespread use of this approach, the mechanistic paradigm is increasingly criticized. The foundations of the mechanistic view: reversibility, reductionistic, equilibrium-based and controllable experiments, have faded away in the light of a number of new scientific insights.

First, the discovery of the Second Law of Thermodynamics brought down the notion of reversibility. The Second Law states that the entropy of a closed system is increasing. This means that heat flows from hot to cold, so that less useful energy remains. One of the consequences of the Second Law is the irreversibility of system behavior and the single direction of time. Changes within systems cannot reverse back just like that (irreversible). This is in contrast to many mechanistic models, in which time can easily be reversed to calculate previous conditions.

Second, the equilibrium view of species was brought down by Charles Darwin's book on the origin of species during the middle of the 19th century. The static concept of unchanging species was replaced by a dynamic concept of evolution by natural selection and adaptation of species, thereby fundamentally changing our view of nature. Natural systems are in continuous disequilibrium, being interdependent and constantly adapting to changing circumstances.

Third, the theories of quantum mechanics have confronted us with a fundamental uncertainty regarding knowledge about systems, especially on the level of atoms and particles. The uncertainty principle of Heisenberg is well known, stating that it is impossible to simultaneously measure the position in space and momentum (mass times velocity) of any particle. The statement by Laplace in the early 19th century that if every position of every atom was known, the future might be predicted exactly, became therefore a lost illusion. Moreover, the notion of fundamental uncertainty implied that fully controlled experiments are strictly speaking not possible.

Notwithstanding the fact that these developments in the natural sciences changed our perception of the world, (mathematical) models are still mainly based on

a mechanistic view of systems. For example, current mainstream economics is based on its success during the period after the Second World War, which was characterized by stable economic growth. Technology could seemingly handle any difficulty that came along. Affluence was seen as growing and permanent, and the standard of living, it was believed, would continue to improve for individuals and generations. The world economy was a place of simple equilibria and linear responses. Cost–benefit analysis, optimization and econometric models seemed to be quite appropriate.

Since the 1970s, various events have made us aware of the non-linearity of economic systems. The oil crisis ended unlimited economic growth, the Berlin wall collapsed in 1989, the Asian financial crisis at the end of the 1990s, the various crises on the stock markets, and so on. The economic system seems to be characterized by unstable states, non-linear responses and unpredictability. Still, mainstream economics uses its successful tools of the early days, probably because of their analytical power; certainly not because of their ability to explain reality. Increasing numbers of economists are trying new tools to explain the observed behavior of economic systems. This has led to the study of non-linear dynamics and evolutionary processes as emerging fields in economics although not always accepted by mainstream economists. This new type of economics studies the formation of patterns, evolution of economic systems, endogenous technologies, and so on. Furthermore, next to analytical tools, this new field of economics uses computers as a kind of laboratory to test hypotheses.

The emergence of new ways of studying the world around us has also emerged in other disciplines where studies are made of the origins of order, self-organization, emergence of structures, adaptation of agents in a changing environment, and many more new ideas. The general focus of this new modeling paradigm is to study how systems change and organize their components to adapt themselves to the problems posed by their surroundings. Examples of these systems are economies, ecosystems, immune and nervous systems, organisms and societies.

Although various scholars long ago discussed the importance of studying the evolution of systems, the rapid developments of the computer have provided scientists with a new tool in recent decades, which can be used to investigate evolution, self-organization, interactions between agents and emergence of structures on a macro scale by simple local rules. These systems can be grouped under the common name complex adaptive systems. They are studied by a number of new computation-based modeling tools, including genetic algorithms, cellular automata, neural networks, and artificial life forms.

The characteristics of these new types of tools are illustrated using one of these tools: genetic algorithms. *Genetic algorithms* have been developed to simulate the process of natural selection by considering a population of agents producing offspring who are similar, but not identical, to their parents. This process depends on three genetic operators: selection, crossover, and mutation. *Selection* means that the genetic algorithm selects n copies of the strings (genetic code) in the population by a random process that favors the most fit. Subsequently, these copies are probabilistically paired in a mating process whereby each pair produces two offspring by means of crossover and mutation. *Crossover* means that two offspring are created with a certain probability that the genetic information is crossed over; otherwise, the offspring are identical to the parents. In the case of crossover, the parent strings of genetic information are split randomly and are swapped to shape two new strings. Each element of the genetic information has a small probability of being altered. This *mutation* is independent of what happens with the other genetic information. Due to their adaptive characteristics, genetic algorithms are powerful tools for improving and finding good solutions even in complex changing environments. Moreover, genetic algorithms are based on irreversible changes, stochastic processes and evolution (*see* Resilience, Volume 2).

Genetic algorithms are one of the computational tools used for developing models to study complex adaptive systems, which emerged as a new field in science. Since the late 1980s, the Santa Fe Institute has provided a prime focus for exploring and deepening the insights from complex adaptive system studies and provided new opportunities for transdisciplinary studies (Waldrop, 1992). It is expected that the study of complex adaptive systems will become an important field in transdisciplinary environmental science. The Resilience Alliance is one of the first international communities to study environmental science from this new perspective. An overview of their work is given in Gunderson and Holling (2001).

Validity of Models

An important difference between the two modeling paradigms is how to use models, and how to validate models. From the viewpoint of mechanistic explanations of reality models are valid when predictions generated by the model are not rejected. An example of this Popperian type of validation is the development of an econometric model. The timeseries of data is split into two. One part is used to estimate the model and to generate predictions, and the second part to test the predictions of the constructed model. This empirical validation leads to problems according to the complex adaptive systems paradigm, which focuses on evolving structures and is often based on qualitative insights. Since small differences in initial values can lead to large output differences, models of complex adaptive systems are not suitable to generate predictions. In fact, this problem goes back to the classical three-body problem (e.g., the system involving Earth–Moon–Sun), which

has been studied extensively with a variety of methods beginning with Newton. Despite the large number of studies, no complete analytic solution in closed form has been found. Furthermore, for some initial conditions this system is found to produce chaotic unpredictable behavior. This example shows that when interactions among agents are taken into account, the Newtonian approach does not work. Therefore, validation in this line of modeling is based on expert judgements and relations with theoretical insights. This type of validation can be called conceptual validation. In fact, validation is not a test, but a subjective process.

The difference between opinions on how to measure the quality of models leads also to differences in how to use models. According to the Newtonian modeling paradigm, a model that is tested empirically and is validated can be used to make accurate and objective predictions. However, according to the complex adaptive system perspective, models can only be used to study qualitative structures, and should be used interactively to be able to exchange insights and stimulate discussion on uncertainties. In the physical sciences, one no longer refers to model validation, but rather to model performance, which is provided by statistical measures such as the root-mean-square differences between model predictions and observed values.

If one wants to describe the orbits of planets, a mechanistic model is a perfect tool. Mechanistic models are even suitable tools when one wants to send men to the Moon. But in situations of many uncertainties and surprises, such tools will not work. In the case of living beings, mechanistic approaches are of limited explanatory power. In the rest of this article modeling human dimensions of global environmental change from the perspective of complex adaptive systems is discussed. From this new perspective, we look at how to design and use models for exploring the future. But first, one of the main seeds of uncertainty and subjective assumptions is discussed: stability characteristics of systems.

MYTHS OF SYSTEMS

Different perceptions of reality can be visualized by different myths of stability, that is the perception of how systems function. According to the equilibrium myth, systems are in equilibrium. External effects can push the system briefly out of equilibrium, but it automatically returns back to the previous equilibrium situation. This myth corresponds very well with the Newtonian-modeling paradigm. Not only the natural system is considered to be in a natural equilibrium, but also the economic system of supply and demand is in equilibrium due to control of the invisible hand. This metaphor from economist Adam Smith, proposed at the end of the 18th century, was a powerful explanation of micro-behavior in order to describe an elegant mechanical description of the macro-level of economic behavior.

The myth of stability can be represented graphically as a ball at the bottom of a valley (Figure 2a). Perturbations only temporarily knock the ball away from the bottom of the valley. An implicit assumption of this myth is that systems have the capacity to damp all kind of disturbances.

An alternative myth is the obverse, namely the myth of instability. Systems are assumed to be very sensitive to disturbances. Every disturbance can lead to a catastrophe. Applied to environmental issues, the myth of instability explains why some people argue that human activities should not be allowed to disturb the natural system. Any degree of pollution or degradation of extraction can lead to a collapse of the system. This myth can be visualized by a ball on a peak (Figure 2b). Any perturbation can lead the ball to roll down the slope.

A third myth is in between the myths of stability and instability, namely a system is assumed to be stable within limits. When the system is managed well, the system can absorb small perturbations. This myth can be visualized as a ball in the valley between two peaks (Figure 2c).

A more advanced framework is to consider multiple stable states. As depicted in Figure 2(d), this myth can be represented as a number of peaks and valleys. The ball is resting in a valley and is able to absorb a certain degree of disturbance. However, an extreme disturbance can push the ball over a peak such that it will rest in another valley, an alternative equilibrium state. Examples of these multiple states are lakes, which can flip from an oligotrophic to a eutrophic state due to inputs of phosphates, and rangelands that flip from a productive cattle-grazing system into a less productive rangeland dominated by woody vegetation, triggered by variability in rainfall.

A myth of systems that is more advanced, and lies in line with the complex adaptive system modeling paradigm, is the myth of resilience.

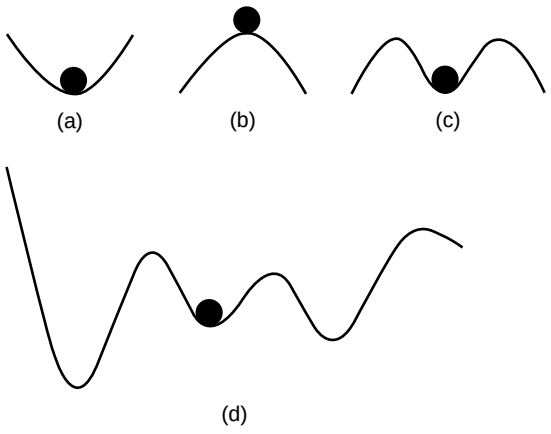


Figure 2 Myths of nature: (a) nature is stable; (b) nature is unstable; (c) nature is stable within limits; (d) nature has different stability domains

Myth of Resilience

The myth of resilience does not only consider the balls moving up and down the peaks and valleys, but also considers possible movements of the peaks and valleys themselves. In this evolutionary picture stability domains can shrink, and disturbances that previously could be absorbed no longer can be. This view has important implications for managing systems. From the perspective of the previous myth of systems, systems could be known perfectly. Surprises may lead to changes of management, because the balls moved in another valley, but in principle, management was simply a matter of controlling the system. From the perspective of an evolving landscape, one has to manage a system in the face of fundamental uncertainties over the functioning of the system. One continually observes the system in order to respond adequately. Moreover, small human-induced perturbations are supported in order to learn from the system. This type of management is called *adaptive management* (*see* Adaptive Environmental Management, Volume 4).

Holling (1986) has proposed a framework for resilience to explain the transitions in behavior of the system. He distinguishes four basic functions common to all complex systems, and a spiraling evolutionary path through them (Figure 3). This evolutionary cycle can be used to explain transitions in social systems, as well as in ecosystems. The central idea is that the four-phase adaptive cycle emphasizes a loop from conservation to two phases of destruction and reorganization in which innovation and chance assume dominant roles. The reorganization phase occurs when a rare and unexpected intervention or event shapes a new future. In this state, the system is most likely to be transformed by innovation, and agents have the greatest potential to influence the future of the system. When the agents do not react properly to changes in the system, it can flip into a new kind of system.

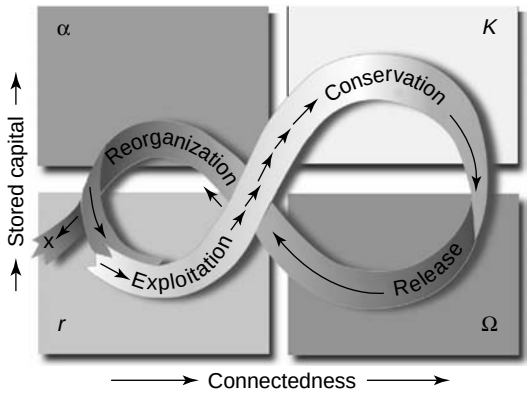


Figure 3 The dynamics of a system as it is dominated by each of four processes: rapid growth (r), conservation (K), release (•), and reorganization (α). The arrows indicate the speed of the cycle. The short, closely spaced arrows indicate a slow change, while the long arrows indicate rapid change. The cycle reflects systemic change in the amount of accumulated capital (nutrients, resources) stored by the dominant structuring process in each phase, and the degree of connectedness within the system. The exit from the cycle at the left of the figure indicates the time at which a systemic reorganization into a less or more productive and organized system is most likely to occur. (Adapted from Gunderson and Holling, 2001)

As mentioned before, the landscape is changing. This has much to do with different speeds of change in the various scales of systems. For example, due to phosphorus accumulation in sediments of lakes, recycling of phosphorus from these sediments can lead to surprises. The slow variable, the sediment, can reduce the capability of the lake to absorb an external disturbance. The equilibrium levels of concentration of phosphorus in the water, the fast variable, can therefore change due to changes in the slow variable (Gunderson and Holling, 2001).

According to the myth of resilience, problems could be caused by local human influences that slowly accumulate to trigger sudden abrupt changes that may affect the vitality of societies. There are counteractive forces that give ecological systems the resilience and adaptability to deal with considerable change, and that provide people with the capability to innovate and create. However, nature, people, and economies are suddenly now coevolved on a planetary scale, each affecting the others in such novel ways, and on such a large scale, that large surprises may overwhelm the adaptive and innovative capabilities of people. One challenge of sustainable development is, therefore, how to stimulate coevolution of human activities and environmental change.

Different Perspectives on the Problem of Climate Change

The discussion of myths shows various possible subjective interpretations of reality. The importance of different perspectives for modeling the human dimensions of global environmental change is now illustrated. There are various concepts designed to classify different worldviews. Like the case of modeling paradigms, there is no true classification of worldviews. A contribution, which gives a general description of perspectives on natural and human systems and social relations, has been made by Michael Thompson and his colleagues in their cultural theory (Thompson *et al.*, 1990). This theory has been used to classify different types of institutional designs in relation to global environmental change. The cultural theory has been an inspiration for implementing worldviews in formal models, because it includes perspectives on human as well as natural systems, it claims generality, and its explanation of perspectives' rationalities is deterministic. The cultural theory combines anthropological insights with ecological insights, resulting in different cultural types.

The three main worldviews in the Cultural Theory are the following types:

1. Individualists assume that nature provides an abundance of resources, and remains stable under human interventions. A responsive management style is advocated.
2. Egalitarians assume that nature is highly unstable, and a small human intervention may lead to complete collapse. A preventive management style is preferred.
3. Hierarchists assume that nature is stable in most circumstances but can collapse if it crosses the limits of its capacity. Therefore, control is advocated as a management style.

As discussed earlier, human induced climate change is a topic surrounded with many uncertainties. It is, therefore, an excellent example to illustrate how worldviews can be quantified to simulate alternative futures based on different perceptions of reality. Such an analysis has been made by Janssen and de Vries (1998) who developed a simple integrated ecological economic model for which they implemented three versions based on alternative assumptions on climate sensitivity, technological change, mitigation costs, and damage costs due to climate change. Egalitarians, for example, assume high climate sensitivity, high damage costs, low technological development, and low mitigation costs. For management styles, they assume different strategies for investments and reductions of emissions of carbon dioxide. By contrast, the individualist, for example, assumes a strategy that maximizes economic growth, and emissions are reduced only if a certain threshold of economic damage is exceeded. The hierarchist tries to balance economic growth and climate change by assuming the IPCC estimate of climate sensitivity.

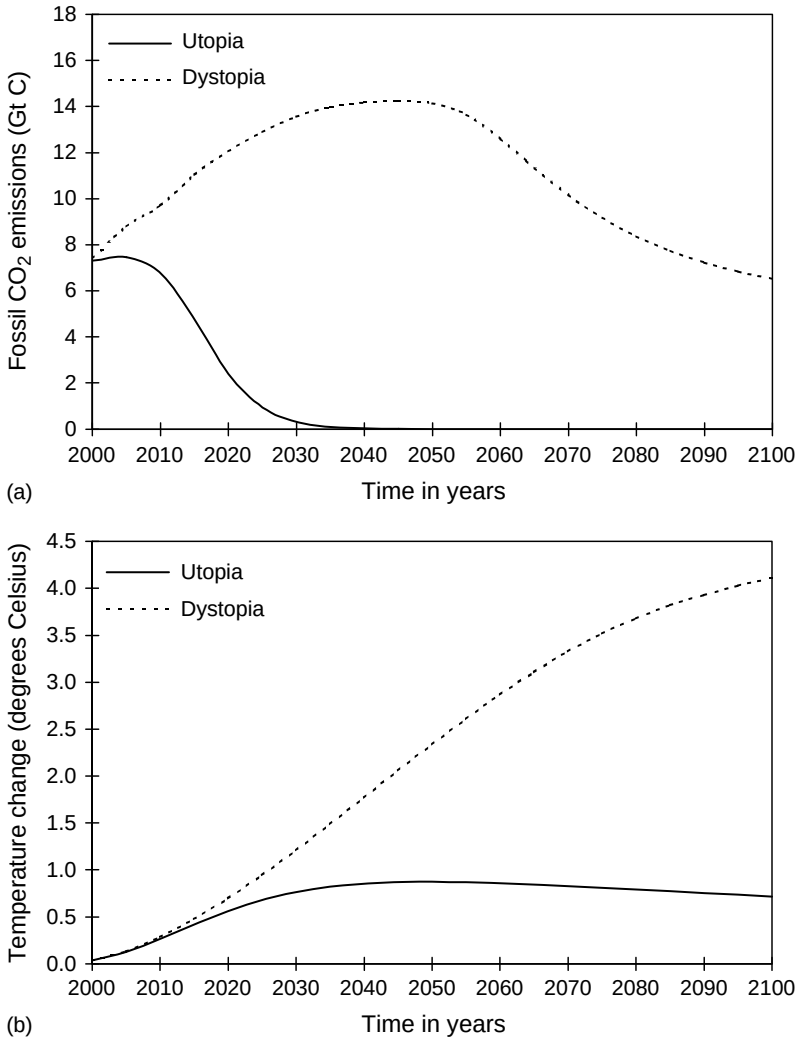


Figure 4 Expected carbon dioxide emissions (a) and temperature increase according to the egalitarian utopia and a possible dystopia (b) (individualistic management style in an unstable global system)

Suppose that all of the agents in a model world share one of the three extreme worldviews. If agents are assumed to have perfect knowledge of their world, their utopia can be simulated. If their worldview is incorrect and they still apply their preferred management style, their dystopia can be simulated. An example is presented in Figure 4. In the egalitarian utopia, emissions of carbon dioxide will be reduced to zero within a few decades, leading to a modest temperature change. However, if the individualistic worldview manages a world that actually operates according to the egalitarian worldview, emissions increase until climate change causes such an economic disaster that emissions are reduced by the collapse of the economy.

By introducing a population of agents with heterogeneous worldviews, a complex adaptive system is produced. It is assumed that the better an agent worldview explains the world's observed behavior, the greater is the chance that an agent will not change its worldview. On aggregate, there is a trend towards changing to the worldview that explains the observations in the most likely way. Suppose that reality is one of the three possible worlds, and an agent obtains information over time that causes him to change (or not) his perspective on the climate change problem. Three sets of projections are derived in which agents adapt to climate change (Figure 5). Prior to the year 2040, the observed climate change does not lead to domination of one of the worldviews. After 2040, the climate signal becomes clear enough that one of the worldviews begins to dominate. In the event of the world functioning according to the egalitarian worldview, the emissions growth stabilizes on average in the coming decades and decreases to a level below half the present amount of emissions. However, this reduction cannot avoid a global mean temperature increase of about 2.5 °C in the 21st century.

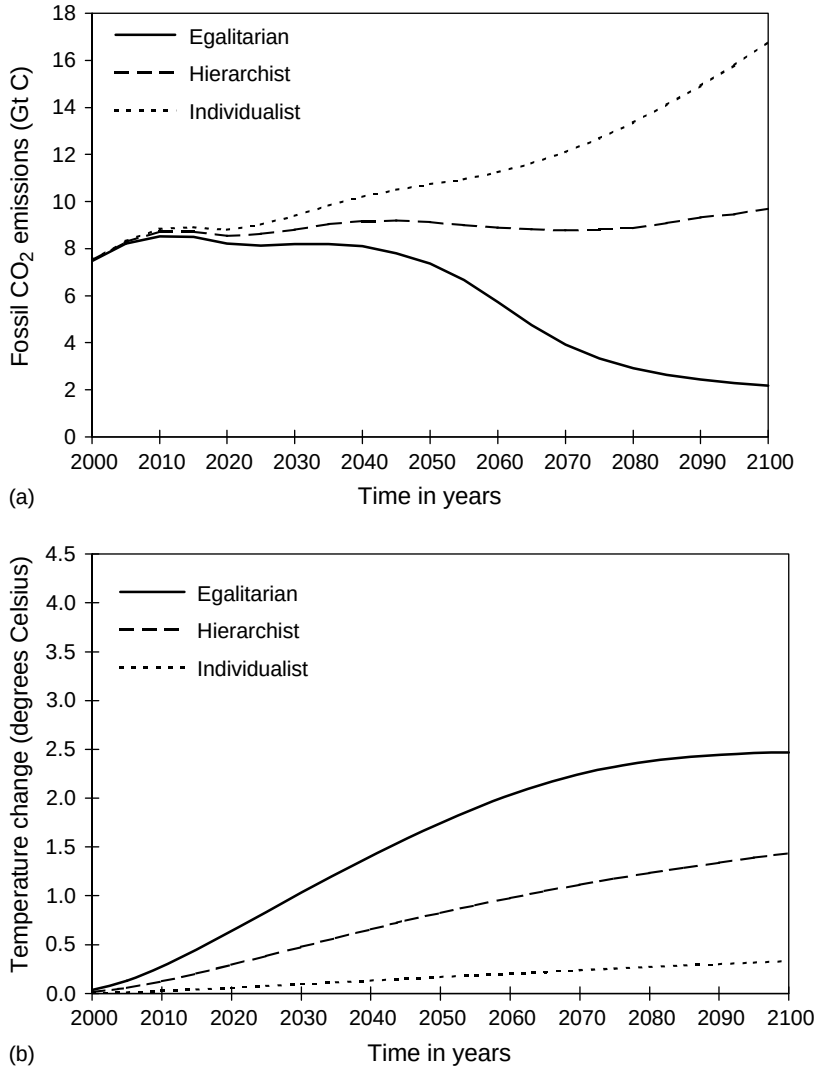


Figure 5 (a) Expected carbon dioxide emissions and (b) temperature increase according to different views on the functioning of the global system, and where the worldviews of agents change in time

The explicit inclusion of subjective perceptions of reality has led to a rich variety of possible futures. This will not simplify decision-making, but can improve decision-making since a large set of plausible scenarios is presented. This type of modeling is currently applied in different areas of global and regional environmental change. Each system in which it is not certain that it will remain stable under all kinds of disturbances, can in principle be studied by explicitly assuming different perceptions of reality. It is the expected that this type of modeling will be especially interesting for simulation of institutional dynamics.

MACRO-LEVEL DYNAMICS

Although there are often uncertainties about the relations, global models simulate macro-level dynamics. Here, we focus on the most important macro-level drivers of global environmental change, which are population size, economic development and technological change. In the early 1970s, Paul Ehrlich and John Holden described the environmental impact of society by the well-known IPAT equation: $I = P \times A \times T$. Here environmental impact equals the product of population size, the degree of affluence per person and the environmental impact from the technology used to produce one unit of affluence.

The coming sections discuss three factors that are closely related to the IPAT equation: population growth, material and energy consumption of economies, and technology development. For each factor, empirical information on the macro-scale provides information on possible developments in the long term. However, extrapolating historically derived macro-level relations involves subjective assumptions. Therefore, it is important to explore the consequences of these subjective aggregated relations and to improve understanding of the empirically derived macro-level information.

Demographics and Human Health

Population growth is one of the main causes of global environmental change. The continuing population growth can be explained by changes in demographics and health. The so-called demographic and health transition describes how populations can go through typical demographic and health stages when they change from living in pre-industrial conditions to having a mortality pattern that is found in the post-modern societies (*see* Demographic transition, Volume 5).

The shapes of the demographic transition curves are well known, but are in fact a hypothesis based on cross-national and longitudinal studies. In developed countries the demographic transition has reached the final stage. But how will this development of fertility and mortality figures continue in the future? Will the developing countries follow the same transition, and at what rate? Most population projections are based on the assumption that all countries will go through the demographic transition leading to a leveling off of population growth during the 21st century. Observed transitions in demographics are the result of changes in individual behavior, technology and norms, improvement of health care, use of contraceptives, age of marriage, literacy, position of woman, regulation of abortion, etc. It is expected that due to all these variables involved, the demographic transition will not occur everywhere to the same degree and at the same speed. Therefore, subjective assumptions have to be made in order to develop projections for the coming centuries.

Assumptions about the health transition in various regions of the world are even more difficult since various diseases are related to global environmental change (skin cancer due to stratospheric ozone depletion), behavior (lung cancer from smoking), emerging new diseases (acquired immune deficiency syndrome – AIDS) and even some re-emerging old diseases due to the development of new resistances (malaria). It is therefore difficult and subjective to project health conditions of our descendants.

The Material and Energy Consumption of Economies

An economic system can be viewed as a living system, consuming material and energy inputs, processing them into usable forms, and eliminating the wastes. The metabolism of economies has changed significantly during the last two centuries. A world economy has emerged that produces agricultural, and industrial products and services in large volumes, and transports them all over the globe. Agricultural production has increased due to more and more intensive use of land. Productivity has improved due to biological innovations in the form of new crops, new agricultural practices, mechanization, and increasing synthetic inputs. The same holds true for industrial production, which is increased due to the increasing consumption of energy and materials. Currently, the service sector is becoming increasingly important, stimulated by increased personal mobility and more individualistic lifestyles.

These economic developments have led to an increase of material and energy use in absolute figures, but also in *per capita* figures, and have led to all kinds of disturbances of biogeochemical cycles. These disturbances have led to environmental change on various scales.

The debate whether the environment is able to sustain economic development goes back to Thomas Malthus at the end of the 18th century, who argued that food production could not be increased quickly enough to keep pace with the growing population. But due to a faster increase of agricultural productivity than expected, a decrease in birth rates, and the growing import of food, Malthus' homeland

(Great Britain) did not collapse. Since the early 1970s, stimulated by the *Limits to Growth* report of the Club of Rome, the debate on economic growth and the environment has emerged again.

In the last 30 years, there have been attempts to improve the physical reality of economic models by including mass balance conditions, and laws from thermodynamics. However, there is still no clear theory on the relation between economic development and the environment. In the 1990s a lot of empirically based studies were published on the relation of environmental pressure of an economic system and the average income. This so-called Environmental Kuznets curve (see, for example, de Bruyn, 2000) (Figure 6) consists of three phases:

1. initially income growth parallels progressively increasing environmental pressure;
2. next, further income growth leads to increasing environmental pressure until it reaches a maximum;
3. further income growth leads to a reduction of environmental pressure.

An explanation for this pattern is that at higher income levels, individuals will attach more value to environmental quality; this means more income spent on less damaging consumption, as well as more democratic support for stringent environmental policies. An important implication of the environmental Kuznets curve is that growth by itself would be able to solve environmental problems. However, the empirical support for this hypothesis is weak and mainly based on cross-sectional analysis. It is not clear whether the curve differ for different types of environmental pressures, what is the influence of policy measures and technological change, and whether observed trends in the past will continue in the future. Some studies suggest that delinking of environmental pressure and economic output has only been a temporary phenomenon caused by efficiency and technology improvements after the oil crises of the 1970s (de Bruyn, 2000). The cheap energy prices of the 1990s led to a relinking of environmental pressure and economic output. Despite the many uncertainties, relationships between physical and financial dimensions of economic systems like the environmental Kuznets curve are used to explore future material and energy consumption.

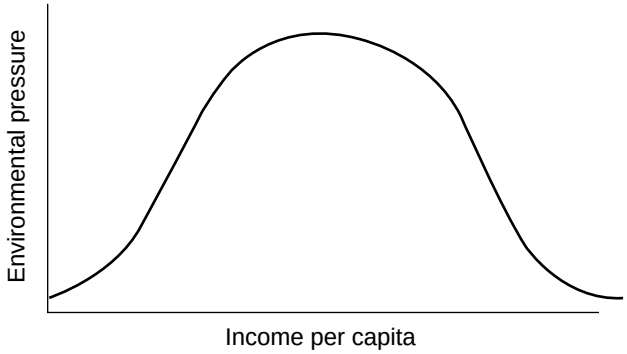


Figure 6 Environmental Kuznets curve

Technological Change

Technology development is the source and remedy of environmental change. This is called the paradox of technology and the environment. It is the source, because it creates the ability of societies to mobilize more materials and energy, and because it creates new materials and substances with direct environmental impacts. On the other hand, technology can also be the remedy because it increases productivity and efficiency of economic activities and invents specific technologies to prevent pollution.

It is therefore important to understand the dynamics of technological change. Arnulf Grubler (1998), in his book on technology and global change, discusses general mechanisms on the diffusion of technology. The question is how a new technology is adopted at a large social and spatial scale. This can be visualized by the stylized technology life cycle (Figure 7a). In the beginning, a new technology is imperfect, and various possible designs are

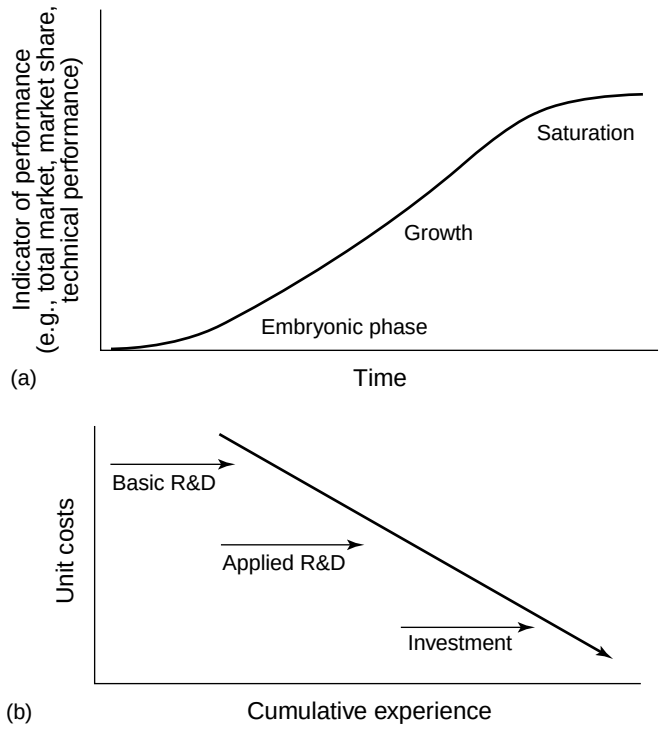


Figure 7 (a) a stylized technology life cycle; (b) a stylized learning curve, where axes scales are logarithmic. In the beginning costs decrease due to basic R&D. When the potential of the technology is demonstrated, applied R&D investments reduce the costs further until a level is reached for which costs are competitive. (Adapted from Grübler, 1998)

explored. The market effect is small, but the increase that occurs during the growth phase is characterized by increasing standardization and falling costs. Finally, the growth rate slows down as the market becomes saturated. In this phase the market is in the hands of a few suppliers.

Technology life cycle is related to the developments of costs. There is a lot of empirical evidence that the decrease of a technology investment cost is related to the accumulated investments in the particular technology. This learning-by-doing can be formulated by the technology learning curve (Figure 7b). The costs (on a logarithmic scale) decrease linearly with an increase in cumulative experience (on a logarithmic scale). The main uncertainty of technology in the early phase of development is the slope at which the costs decrease.

In mainstream economic models, technology is often included as an exogenous variable. Such an assumption leads to wait-and-see policies because investments are delayed until clean technology has become available at suitable cost levels. Such a policy differs from insights derived from the learning curve, which suggest stimulation of investments in clean technology allowing costs to decrease.

Insights into technology dynamics show stylized facts derived from empirical studies. This provides tools for modeling technology development. The main question is how does an individual technology penetrate the market, and how can governmental policy stimulate the diffusion of green technology? Many decisions of individuals influence this evolution. Therefore, we need to have more insights into models of human behavior.

UNDERSTANDING THE EMERGENCE OF STYLIZED FACTS

The observed stylized facts, the observed macro relations, are the result of decisions of many agents. They are emergent properties of a complex adaptive system. To improve our understanding of these observed stylized facts, we should develop tools to understand them. One way is to use models that are able to simulate these emergent properties from the bottom up. Like the cartoon (Figure 8), understanding of macro level phenomena can only be derived by studying the behavior of micro level agents. In the case of a school of fish, multi-agent studies have shown that flocking behaviour can be simulated by using three simple rules for each agent:

- separation: steer to avoid crowding local flockmates;
- alignment: steer towards the average heading of local flockmates;
- cohesion: steer to move toward the average position of local flockmates.

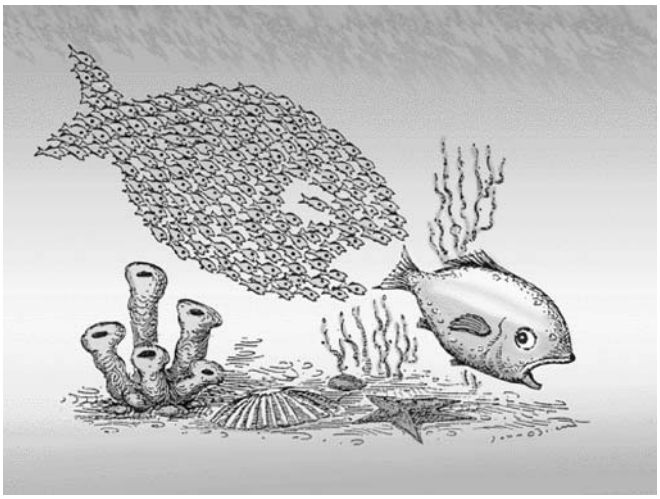


Figure 8 Macro level phenomena emerge from behavior of agents on the micro level

Behavior of human agents is much more difficult to capture by simple rules. Two types of human behavior can be distinguished: individual behavior, and behavior of groups and institutions.

Individual Behavior

Since theory development in social science is rather fragmented, various models of human behavior exist. In fact, different disciplines study only particular aspects of human behavior. One central element, or better, one stylized problem, is found in most models of individual behavior. Humans are assumed to maximize their well being given budget constraints. These budget constraints mainly relate to income and time. Within economics and psychology, different variations exist on this stylized problem. In the formal approach of conventional economic theory, this means that the rational actor, the *Homo economicus*, maximizes its own well being. This *Homo economicus* is assumed to have perfect knowledge of the system in order to find the global optimum. In the case of uncertainties, the probability distributions are perfectly known. These assumptions make it possible to formalize human behavior in an unequivocal way. A drawback is the existence of much empirical evidence that real humans do not behave in this way. For example, experiments show that humans discount the near future at a higher rate than the distant future, experience well-being with relative changes instead of absolute levels, and are about twice as averse to taking losses as to enjoying an equal level of gains (e.g., Thaler, 1992).

Some economists, like Nobel Laureate Herbert Simon, argue that humans are rationality bounded. First, no person can ever assemble all the information required for an optimal decision. Second, even if one could, decisions are usually so complex that no simple algorithm exists

for evaluating all possible options. Third, a person's own decisions depend on the decisions of other persons. Simon argues therefore that humans satisfice instead of maximize their well being. A consequence of relaxing the assumption of maximizing behavior is the large set of possible rules that describe behavior. The formalization is not unequivocal anymore. Thanks to the development of new simulation techniques such as cellular automata, genetic algorithms and neural networks, and the widespread availability of personal computers, social scientists are exploring new ways of modeling human behavior (e.g., Gilbert and Troitzsch, 1999).

Psychologists and other social scientists include different indicators than are usually used in economics. Psychologists focus on satisfaction of various types of needs such as understanding, freedom, identity and leisure. Furthermore, all humans are different. They differ in their abilities, mental models, preferences and opportunities. In contrast with *Homo Economicus*, decisions depend on social interactions, as well as on individual considerations. In fact, humans are assumed to perform different cognitive decision processes in different situations.

The mental models of humans are important elements in cognitive processing. Differences in perceptions of reality can result in different behavioral patterns. Although humans can learn, they will never have the perfect knowledge that is needed to maximize well being as performed by *Homo Economicus*. This picture of human behavior is in line with the complex adaptive system-modeling paradigm. Since humans cannot perfectly control their own situation, they design institutions to regulate human interactions, as well as interactions between society and the environment.

Institutions and Collective Actions

Institutions contain formal constraints (rules, laws, constitutions), informal constraints (norms of behavior), and their enforcement characteristics. They shape human interactions and the way societies evolve through time, and can also be important to regulate human activities in relation to ecological conditions. A stylized problem that is generally used to study institutions is the so-called *commons dilemma*, widely discussed as a result of the well-known analysis of Garrett Hardin on the *Tragedy of the Commons* (*see* Commons, Tragedy of the, Volume 5). According to this analysis, the commons tend to be overharvested since each agent harvests to the point where private costs equal the benefits, whereas harvesting imposes additional social costs on the rest of the community. However, historical analyses of common resource properties have found many examples where the tragedy did not happen (Ostrom *et al.*, 1999). Communities often had ways of self-organizing to prevent overexploitation of the commons, also known as closed commons. Success of self-organization depends heavily on the characteristics of property-right systems. The tragedy of the commons is an example of open access, where everybody can harvest without individual punishment. However, other types of property regimes to regulate common resources are related to group, individual and governmental property.

The success of self-organization of effective institutions to control common resources depends on property regimes, as well as evolution of norms and design of rules. Again, this is an excellent example of complex adaptive systems.

DIFFERENT WAYS OF USING MODELS

It should be clear by now that models are not of use to accurately predict future developments of human dimensions of global environmental change. Although models are often used for this purpose, there are more suitable purposes for which they are of importance. In fact, they can be used to overcome the problems raised during the analysis of the current generation of integrated assessment models. Instead of focusing on a single model used for prediction, a combination of different types of models should be used to explore hypotheses and uncertainties. Models can be used in different ways. Here, three different goals of use are distinguished: to understand observed stylized facts, to improve decision-making of complex problems due to interactive use of models, and to explore possible futures by scenario developments.

Understanding Stylized Facts

In the descriptions of the various sectors a number of stylized facts were identified, such as the demographic and health transition, diffusion of technology, and environmental Kuznets curves. These macro-level observations are the result of behavior of agents at smaller scales. To understand stylized facts, and to explore possible changes in macro-level relations, we need to study the behavior of the underlying components. Can we explain the diffusion of technology by using simple rules for agents? Under which assumptions can an environmental Kuznets curve be simulated by micro-level decisions of economic agents?

This type of question relates to current work in evolutionary economics. Various studies analyze what are the important characteristics of firms and technology development to explain structures emerging in specific markets. Examples are the size distribution of firms, and the large number of firms during the beginning of a new market. Such analyses are also frequently performed by various other disciplines, and are assumed to be valuable to understand the underlying dynamics of the observed stylized facts.

Interactive Use of Models

Models are not prediction machines, and should not be used in this way. Moreover, for policy analysis, models

should be made available in such a way that they can be used in an interactive fashion. The recent developments in graphical user interfaces, PC availability and Internet, should make it possible to use simple transparent models by a large community of stakeholders. Models can be used in this way as learning tools. Using models in this way is exactly what we can learn from the work of Dorner (1996).

Dietrich Dorner and his colleagues study decision-making in complex situations. Their research design is to develop micro-worlds, which are computer simulation models of a management problem, and to ask real persons to manage the system. These management problems vary from a simple climate-control system to regulation of a virtual society. Because the participants in the game have an incomplete picture of reality, their attempts to manage the system can lead to catastrophes. Dorner concludes that managing complex situations requires experience, and this experience can be built up by playing management games, like training pilots in flight-simulators.

Scenario Development and Alternative World Views

Traditionally, scenario analysis starts with an initial forecast which is called the base case or reference scenario. Then alternative assumptions are made of initial conditions, equations and parameter values. The resulting projections are called *scenarios*. Differences between the scenarios and the references are used to evaluate uncertainties and possibilities of policy to influence the future. However, due to the implicit assumptions within every model, one reference scenario will not be enough to assess policy options. In fact, a set of reference cases should be used, which capture the main variety in alternative assumptions (Rotmans and de Vries, 1997).

Scenario analysis should be seen as computer-aided storytelling, where different stories can serve as a starting point for analysis. Analyzing policy options for different types of futures, can give insights into the robustness of these policies. Explicitly using different starting points can reduce the illusion that model-based analyses are objective predictions. In such a way, models can serve as a medium for discussion instead of an electronic oracle.

CHALLENGES FOR THE FUTURE

Important weaknesses of current modeling activities, related to human dimensions of global environmental change, are the inherent subjective assumptions of parameter values and relationships, and the fact that macro-relations are used of phenomena which emerge from local interactions between agents.

The challenge in the coming decades will be to use new developments in modeling tools and the use of models to overcome current problems. A promising start can be made when models are used in a more explorative way. Current models are often used as truth-machines. But the predictions, the glimpses of the future derived from electronic oracles, should not be our main interest. Since the future is inherently unpredictable, models should be used to enrich our insights into the behavior of complex reality. Improvements of our mental models can help us to improve our decision-making.

The concept of complex adaptive systems can be a promising starting point to think about new ways to develop and use models concerning human dimensions. Systems evolve, and therefore models assessing the future should focus on evolution and change. This requires the inclusion of disciplines nowadays not highly involved in modeling, such as psychology, institutional science, history and anthropology. Modeling human dimensions of global environmental change is a way of managing uncertainty and complexity. The tools discussed in this article can help us to experience and learn this art.

REFERENCES

Ayres, R U and Simonis, U E, eds (1994) *Industrial Metabolism - Restructuring for Sustainable Development*, United Nations University Press, Tokyo, Japan.

Clark, W C and Munn, R E, eds (1986) *Sustainable Development of the Biosphere*, Cambridge University Press, London.

de Bruyn, S M (2000) *Economic Growth and the Environment: an Empirical Analysis*, Kluwer, Dordrecht.

Diamond, J (1997) *Guns, Germs and Steel - A Short History of Everybody for the Last 13 000 years*, Vintage, London.

Dorner, D (1996) *The Logic of Failure - Recognizing and Avoiding Error in Complex Situations*, Perseus Books, Reading, MA.

Gilbert, N and Troitzsch, K G (1999) *Simulation for the Social Scientist*, Open University Press, London.

Grubler, A (1998) *Technology and Global Change*, Cambridge University Press, Cambridge.

Gunderson, L H and Holling, C S, eds (2001) *Panarchy: Understanding Transformations in Human and Natural Systems*, Island Press, Washington, DC.

Holling, C S (1986) The Resilience of Terrestrial Ecosystems: Local Surprise and Global Change, in *Sustainable Development of the Biosphere*, eds W C Clark and R E Munn, Cambridge University Press, London.

Hughes, B B (1999) *International Futures: Choices in the Face of Uncertainty*, Westview, Boulder, CO.

Janssen, M A (1998) *Modelling Global Change - The Art of Integrated Assessment Modelling*, Edward Elgar, Cheltenham.

Janssen, M A and de Vries, B (1998) The Battle of Perspectives: A Multi-agent Model with Adaptive Responses to Climate Change, *Ecol. Econ.*, 26, 43–65.

Meadows, D H, Meadows, D L, Randers, J, and Behrens, W W (1972) *The Limits to Growth*, Universe Books, New York.

Meadows, D H and Robinson, J M (1985) *The Electronic Oracle: Computer Models and Social Decisions*, John Wiley & Sons, Chichester.

Ostrom, E, Burger, J, Field, C B, Norgaard, R B, and Policansky, D (1999) Revisiting the Commons: Local Lessons, Global Challenges, *Science*, 284, 278–282.

Rayner, S and Malone, E L, eds (1998) *Human Choice and Climate Change*, Battelle Press, Columbus, OH.

Rotmans, J and de Vries, B, eds (1997) *Perspectives on Global Change*, Cambridge University Press, Cambridge.

Thaler, R H (1992) *The Winner's Curse: Paradoxes and Anomalies of Economic Life*, Princeton University Press, Princeton, NJ.

Thompson, M, Ellis, R, and Wildawsky, A (1990) *Cultural Theory*, Westview, Boulder, CO.

Waldrop, M (1992) *Complexity: The Emerging Science at the Edge of Order and Chaos*, Simon and Schuster, Toronto.

Modernism

see Art and the Environment (**Volume 5**)

Modernity vs. Post-modern Environmentalism

Daniel R White

Honors College, Florida Atlantic University, Jupiter, FL, USA

Postmodern environmentalism provides a critique of, and a set of alternatives to, modern industrial society. It criticizes the separation of science from ethics and aesthetics, its separation of nature from culture and self from other, its Eurocentrism, and its sexism. As alternatives, it offers interdisciplinarity, reconciliation, multiculturalism, and ecological feminism. In place of modernist logical hierarchies and categorization of the world into things, postmodern environmentalism provides a communicative paradigm emphasizing metaphoric comparisons and interchanges among diverse inhabitants of the human-ecological community. Finally, this new outlook combines technology and nature in the creation of an emerging form of human-ecological identity, called the cyborg.

*Postmodern environmentalism is related to post-normal science (*see *Post-normal Science, Volume 5) in the following ways: (1) both are concerned with complex systems; (2) both offer a critique of the specialization and dogmatism of normal science; (3) both reintroduce value judgements (as in ethics and aesthetics) and environmentalist politics into policy decisions involving the natural sciences.*

Both environmentalism and postmodernism mark fundamental changes in the status and viability of modernity. The two movements converge insofar as they focus on those features of modernism that have contributed to the ecological crisis. *Ecology* originally meant the science of the household, and as such should be concerned with the rich variety of activities that create and sustain our daily lives: everything from the construction and maintenance of the human habitat to the plurality of biotic processes intrinsic to life on Earth. According to the environmental and postmodern critiques, modernity has increasingly undermined this human-ecological variety to create a world industrial order and a global monoculture. Under the sway of this simplifying modernism, the diversity that has made up the human household and the planetary biotic community has been externalized, devalued and oftentimes destroyed. Environmental and postmodern thinkers have therefore sought to revitalize ecological and human diversity.

The problematic aspects of modernism that are the focus of both environmental and postmodern critiques include:

1. the specialization of thought into separate realms of science, ethics and aesthetics;
2. the separation of culture from nature;
3. the binary demarcation of human from non-human;
4. the definition of self or mind as a subject having dominion over an objectified natural world;
5. the identification of that subject with Western culture and the masculine gender.

The combination of these modernist ideas has, according to environmental and postmodern thinkers, contributed largely to the ecological crisis.

The replacement of these problematic features of modernism with viable alternatives has become, therefore, a principal aim of environmentalist and postmodern thought. Both movements, accordingly, emphasize interdisciplinary thinking that re-links science, ethics and aesthetics. Both argue that cultural concerns like knowledge, value and beauty are in part derived from and have consequences for the natural world. Both make clear that the human and the non-human are inseparable aspects of ecological communities. Furthermore, both attempt to provide new forms of communication for the interdisciplinary study of ecological and human identities and relations. Finally, both emphasize the value of cultural and gender diversity, particularly the contributions of non-Western peoples and of women, to the creation and maintenance of the ecological household.

Postmodernism often invokes metaphor rather than logic to express its ideas. So it is appropriate to say that environmentalist and postmodern ideas have become interwoven in a new tapestry entitled *postmodern environmentalism* on whose fabric new forms of thought and action are appearing.

Box 1 Some definitions

modernity: The broad cultural movement from the 16th through the 20th centuries, focused on the idea of progress, and particularly on the betterment of human life through science and technology. In standard Western philosophy (epitomized by the work of Immanuel Kant) it includes the partitioning of knowledge into three spheres of inquiry: science, ethics, and aesthetics.
postmodernity: A cultural movement of the late 20th and early 21st centuries, based on the perceived failure of modernity to achieve its goal of human progress. It focuses on the ideas of cultural, social, and biological diversity, the rights of women and minorities, multiculturalism, and environmentalism. In philosophy it emphasizes interdisciplinary connections among the spheres of modernity (science, ethics, and aesthetics).
environmentalism: A contemporary philosophical movement advocating the legislative, political, ethical, and aesthetic defence of ecosystems as well as the preservation of a sustainable human ecology. Its outlook implies an *environmentalist* critique of modern industrial technologies insofar as they reduce biodiversity and undermine the viability of ecosystems to support life.
cyborg: A ''cybernetic organism'' of ''bionic human'' (Merriam-Webster). A new postindustrial form of technology based on the communication sciences. Still in its initial stages of development, the cyborg is increasingly capable of autonomous behavior simulating human conduct. It is therefore an innovative hybrid of nature and culture. Cloned organisms, including the biologically engineered human beings on the horizon, are examples of cyborgs.
teleological: an adjective based on the Greek words *telos* (purpose) and *logos* (word, or explanation). Teleological means ''purposive'' or ''designed'' or ''explained in terms of purpose or design''. In the information sciences it has largely been replaced by the adjective ''teleonomic'', based on *telos* and *nomos* (usage, convention). Teleonomic means ''simulating purpose'', or ''the quality of apparent purposefulness in living organisms that derives from their evolutionary adaptation'' (Merriam-Webster). Teleological explanation was rejected by modern science because it was associated with premodern ideas of design embraced by theology. *Postmodern* and *post-normal* science (*see* **Post-normal Science**, Volume 5) employ teleological or teleonomic explanation, as part of their reintroduction of ethical and aesthetic ideas (including those of purpose and design) into environmental discourse.

In standard modern philosophy, natural science requires the empirical description of nature without ethical or aesthetic considerations. Standard ethical thought, in turn, deals with *what ought to be* while science deals with *what is*. Aesthetics, similarly, deals with questions of beauty and sublimity (awe) that are neither factual nor ethical but, instead, are matters of taste. It further considers *teleological* (or *teleonomic*) questions about the apparent design of living things. See definition of *teleological* in Box 1.

Environmentalist and postmodern thinkers, in response, have argued that the description of nature or any object involves ethical and aesthetic dimensions. To select an object for description, e.g., requires that it be differentiated in value from some other object or as a figure against a background. That differentiation in value may be either ethical, in that the selected figure is chosen because it ought to be studied, or it may be aesthetic, because it is pleasing to study. In either case, the choice is teleological, in that it selects an object in terms of the purposes of the researcher. The observer of the natural world may also attribute some intrinsic purpose to it. Yet, these ethical, aesthetic and teleological dimensions are excluded by the modernist description of nature. They are relegated to specialists in disciplines outside of science.

The demarcation of culture as separate from nature, as well as human from non-human, is also typical of the modernist description of a natural object, say a tree. *Tree*, after all, is a word. In the terminology of linguistics it is a *signi er*, which denotes a signified concept which in turn denotes a perceived object such as the familiar configuration of trunk and branches. In postmodern terms the cultural activity of naming is part of the nature described by the word. The tree cannot be referred to as a natural object without the cultural activity of naming, i.e., it belongs to the whole system of what we mean by nature.

The description of a tree therefore implies a linguistic system. According to linguistic theory, the selection of the signifier *tree* and its associated concept displaces all of the others that might be chosen. Hence the *tree* and the *not-tree* are two aspects of the same binary system of language, the same code. To demarcate the tree from all that is not-tree, thereby saying that it must be treated as a separate entity, is an initial arbitrary act by scientific modernists, yet it involves the very ethical or aesthetic judgments that modernism would exclude from science. To treat the tree as a separate entity cut off from its linguistic system, furthermore, is analogous to cutting it out of the ecosystem – an act that has not only biological but also ethical and aesthetic consequences.

From the perspective of postmodern environmentalism, this amounts to separating the tree from the forest and its inhabitants. The cutting is done by a self considered to be separate from nature. The separation of the self from nature is typical of the Western cultural tradition. That this kind of subject is built into the descriptions of science indicates that modern science, at least, is an artifact of the same tradition – science is cultural. The scientific self is typically male and, in the view of postmodern environmentalism, objectifies the feminine body of nature based upon masculinist preferences: power, profit and control. Postmodern environmental science, therefore, promises to be feminist and multicultural, as well as communicative.

To learn how to see the forest and the trees simultaneously, postmodern environmental science has invoked the theory of communication. This theory focuses on the forms of connection and exchange between living things, e.g., on the interchange of words between speakers, or signals between computers on the Internet. It, furthermore, models organisms as artifacts or cyborgs emergent from the communicative processes of evolution.

A *cyborg* is an arti cial organism, a creature that is a hybrid of biology and technology. It is therefore, in the broadest sense, a work of ction: it has been made up from the elements of nature and culture. *Cybernetics* is the science underlying the construction of cyborgs. It is based on a broad analogy between the behavior of organisms and machines. Its central question is, What must we assume is going on inside an animal or a robot that allows it to act in an apparently purposive manner? The answer is that both must be in communication with their environments. Communication, in turn, is de ned in terms of negative feedback: a technical way of saying that an organism or a machine selects a particular course of action by excluding alternatives. In terms of communication theory, e.g., in order to catch a y, a frog must rst exclude all of the alternate courses of action the insect might take. By so doing, it selects the course the y is on. The frog then lashes out its tongue to the point along the course where the y will be when the tongue arrives. A guided-missile frigate would use the same negative logic to shoot down an incoming aircraft, and presumably so would you or I as we reach to catch a ball arcing toward us over a baseball eld. The frog, frigate or ball player might be organism or machine or human: the requirements for selecting a course of behavior based on negative feedback are basically the same. This selection is, in turn, teleological, ethical and aesthetic: the purpose, choice, and preference of the organism or machine or human being are intrinsically bound up with its selection of a course of action.

The cyborg is a de nitively postmodern construction because it breaks down the boundaries between culture and nature: it is a hybrid of cultural and natural processes understood in terms of communication. As the design science underlying the cyborg, cybernetics would view the forest and the trees simultaneously by providing a recursive vision of nature and culture. Negative feedback, paradoxically, includes a description of courses of action it must exclude in order to select its object. Because the process of feedback takes place in time and space, it must continuously update the position of the subject (frog, frigate, human) in relation to the object (y, aircraft, baseball). You and I must keep our eye on the ball if we want to catch it, constantly adjusting our position relative to the incoming y (so to speak). The full description of subject and object at any time must also include, as we have seen, the relative positions of and alternative actions excluded by both. If the ball were alive and capable of adjusting its own actions by negative feedback, like the hedgehogs that serve as croquet balls in the Queen s famous match in Lewis Carroll s *Alice in Wonderland*, then our game would be all the more complex (as it is, indeed, for the frog or frigate).

Think of the subject s position as being on one side of a circuit and the object s as being on the other. Consider the subject s descriptions of the object to be messages in that circuit. Now the messages in the circuit will re ect a spiral of differences between a series of subject and object positions. In this spiral, the description of the object must include a description of the subject (because they are always relative to one another) so that the overall system becomes self-referential. A self-referential system is recursive insofar as it refers back to itself, constantly revising its self-image. Your description of the incoming baseball must include a description of your position relative to the ball, and that description is constantly shifting as you and the ball approach one another on the eld.

This constantly self-correcting ensemble of subject and object is a good de nition both of the cyborg and the postmodern self. Both are hybrid artifacts generated by recursive communication in which subject and object are constructed within a spiral of differences. The spiral ultimately includes subject and object, self and other, culture and nature, trees and forest, in a postmodern-ecological design. That design has utility for survival because it yields a coadaptive evolutionary relationship between each of the aforementioned dichotomies. This relationship is precisely what the modern idea of nature lacks and the postmodern idea provides.

FURTHER READING

Bateson, G (1987) *Steps to an Ecology of Mind*, University of Chicago Press, Chicago, IL.

Haraway, D (1991) *Simians, Cyborgs and Women: The Reinvention of Nature*, Routledge, New York.

Hayles, K N (1999) *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*, University of Chicago Press, Chicago, IL.

Ihde, D (1990) *Technology and Lifeworld: From Garden to Earth*, Indiana University Press, Bloomington, IN.

Merchant, C (1980) *The Death of Nature: Woman, Ecology and the Scientific Revolution*, Harper, New York.

McCulloch, W and Pitts, W (1965) *The Embodiments of Mind*, MIT Press, Cambridge, MA.

Trinh, T M (1989) *Woman, Native, Other*, Indiana University Press, Bloomington, IN.

White, D R (1998) *Postmodern Ecology: Communication, Evolution and Play*, State University of New York Press, Albany, NY.

Muir, John

(1838- 1914)

John Muir, naturalist, explorer, writer, and conservationist, was born in Scotland. The Muir family emigrated to the US in 1849, settling in Wisconsin. His childhood was harsh and work-filled, but at an early age he became a keen observer of the natural world and an accomplished craftsman.

In 1860, Muir entered the University of Wisconsin, and in 1863 he traveled across the largely untouched northern US and Canada. In 1867 he suffered an eye injury while working. While he soon regained his sight, the event was to change Muir's life. He turned to nature. Muir walked from Indianapolis to the Gulf of Mexico; sailed to Cuba and Panama, and then in 1868, up to San Francisco. California became his home.

In 1868, he journeyed on foot through meadows, wildflower fields, and into the high country of the San Joaquin Valley. Later he wrote: it seemed to me the Sierra should be called not the Nevada, or Snowy Range, but the Range of Light ••• the most divinely beautiful of all the mountain chains I have seen. In 1874, he began a series of writings titled *Studies in the Sierra*. Fame ensued and his works were soon known throughout the US. In 1880, he met his future wife Louise and later they had two daughters. He entered the farming business with her father, and prospered.

Muir continued to visit the Sierra Nevadas, and traveled extensively abroad, eventually seeing all the corners of the globe. He also returned to writing, publishing articles and books about his travels and naturalist philosophy. Muir's writings assumed a spiritual quality that inspired his readers with his love for nature. In 1890, due in large part to the inspiration of Muir's works, Yosemite National Park was created. Muir was also the catalyst for the creation of Grand Canyon, Sequoia, and Mount Rainier parks. In 1892, Muir and friends founded the Sierra Club, as Muir stated to do something for wildness and make the mountains glad (Muir remained Club president until his death in 1914).

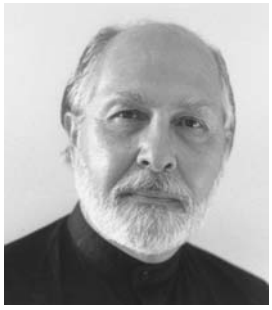
In 1901, Muir published *Our National Parks*, a work that was to become quite influential. He also formed a friendship with Theodore Roosevelt and Gifford Pinchot, a connection that was to have an important impact on the development of American conservation policy. Muir and the Sierra Club worked to protect the Sierra Nevadas. In 1913, a dramatic battle was fought; a campaign to prevent the damming of the Hetch Hetchy Valley, which was then within Yosemite, but the battle was lost. The valley was dammed to supply water and electricity to San Francisco, and to break the City's reliance on a private utility monopoly. However, the event had a galvanizing effect on supporters of the National Parks. The incident also ended his friendship with Pinchot, who had supported the project. Muir died the following year.

John Muir has been called the father of the American National Parks system. He was certainly one of the most influential and powerful of the early conservationists. John Muir's ideas and writings have had a lasting impact and legacy on the environmental movement. The Sierra Club is now among the most influential global environmental organizations (*see* **Sierra Club**, Volume 5).

KEVIN HANNA *Canada*

N

Nasr, Seyyed Hossein



(1933-)

Hossein Nasr (Seyyed is an honorific title designating a descendant of the Prophet Muhammad) has long been the most visible proponent of environmental values from an Islamic perspective, and is one of the most significant contemporary Iranian philosophers. His approach follows the perennial philosophy associated with Frithjof Schuon, Titus Burckhardt and Rene Guenon, in which timeless truths are seen as being expressed in a variety of historical cultural and philosophical traditions.

Nasr was born in Tehran, Iran into a family of scholars, physicians and Sufis (Muslim mystics). From secondary school onward he was educated in the US, first at the Peddie School in New Jersey where he was class valedictorian, then the Massachusetts Institute of Technology (MIT) where he received an honors degree in mathematics and physics in 1954. He went on to study at Harvard, where he completed a doctorate in the history of science in 1958. He then returned to Iran where he taught philosophy and the history of science until the revolution of 1979. In 1963, at the age of 33, Nasr became the youngest full Professor of the university. He served as Dean of the Faculty of Letters from 1968–1972 and briefly as Vice Chancellor. In 1972, the Shah appointed him Chancellor of Aryamehr University, a new technical institution modeled on MIT. The following year, under the patronage of the Queen, Nasr established the Imperial Iranian Academy of Philosophy, which hosted the eminent scholars Henri Corbin and Toshihiko Izutsu, among others. Following the 1979 revolution, he returned to the US, where he taught first at the University of Utah, then Temple University, and finally at the George Washington University, where he has been University Professor of Islamic Studies since 1984. He has also held visiting positions at Harvard, Princeton, and the American University of Beirut. In 1981, he became the first Muslim to deliver the Gifford lectures at Edinburgh.

One of Nasr's major contributions to the field of education, especially during his years in Iran, was his emphasis on approaching the study of Islamic philosophy and science from the perspective of his students' own Islamic and Iranian traditions, in contrast to the external approach of Western Orientalists. He thus strove to bring about a synthesis drawing on the strengths of both Western and non-Western intellectual traditions.

Alongside his academic pursuits, Nasr has also sought knowledge through traditional means, such as the largely oral transmission of learning from master to disciple characteristic of Sufism. In this context he studied Islamic philosophy and other disciplines under the tutelage of Mohammad Hossein Tabataba'i, Abo'l-Hassan Qazvini, and Mohammad Kazem Assar.

Anticipating the well-known 1966 critique of his fellow historian of science Lynn White, Jr, in his Rockefeller lectures at the University of Chicago earlier the same year (subsequently published as *Man and Nature: the Spiritual Crisis in Modern Man*), Nasr argued that the emerging environmental crisis was rooted in a spiritual crisis. He has taken up this theme in many of his later writings, including his Cadbury lectures at Birmingham in 1994, which were published as *Religion and the Order of Nature* in 1996.

FURTHER READING

Nasr, S H (1968) *Science and Civilization in Islam*, Harvard University Press, Cambridge, MA.

Nasr, S H (1978) *An Introduction to Islamic Cosmological Doctrines*, revised edition, Shambala Publications, Boulder, CO.

Nasr, S H (1996) *Religion and the Order of Nature*, Oxford University Press, New York.

Nasr, S H (1997) *Man and Nature: The Spiritual Crisis in Modern Man*, revised edition (1967), Kazi Publishers, Chicago, IL.

RICHARD FOLTZ *USA*

National Environmental Law

M J Peterson
University of Massachusetts, Amherst, MA, USA

National environmental law comprises those branches of the national law of each state addressing human interactions with nature, focusing primarily on protecting the biosphere - including its subsystems, processes, and species diversity - from avoidable signi cant harms arising from human activity. National-level environment law-making, like the counterpart process at the international level, is affected by the special characteristics of environmental problems: high uncertainty about the effects of human activity on the biosphere, heavy reliance on scienti c, engineering, and economic expertise in the formulation and enforcement of legal rules, necessity of considering long time horizons, and complexities of ensuring compliance among a wide range of private, quasi-public, and public actors.

National legal systems have long addressed land and water use, resource management, public and private construction, health and sanitation, food safety, and waste treatment, but environmental law did not emerge as a separate branch of national law with its own statutory basis, administrative organs, and jurisprudence until the late 1960s and early 1970s. Environmental laws adopted in the 1970s focused primarily on control and reduction of pollution, wastes, and hazardous or toxic substances. Their scope has broadened in subsequent decades as societies have realized that many human activities have signi cant effect on the natural environment. Thus, by the mid 1990s it was possible to de ne environmental law in very broad terms as the body of law intended to minimize the impact of activity by individuals, households, business rms, other private and public legal entities, and government agencies on natural ecosystems and resources (e.g., McGregor 1994). However, the de nitions that guide actual legal practice continue to focus primarily on the broad subject areas of air quality, water quality and use, biodiversity and wildlife management, environmental impact and risk assessment, control of hazardous and toxic substances, remediation of hazardous substance contamination, and liability for pollution damage (e.g., Rosencranz, 1991; Baker, 1997; Findley and Farber, 1999).

CHARACTERISTICS AND DEVELOPMENT OF NATIONAL ENVIRONMENTAL LAW

National environmental law shares the main features of other branches of national law. First, it operates within a centralized political system based on a hierarchical relationship between government and governed. The government holds rights to prescribe, apply, and enforce the law to all individuals, other private legal entities, and subordinate public entities that are its nationals and to those private individuals and entities of foreign nationality operating within the state's land, maritime, and aerial domain. In turn, those entities have an obligation to obey. Individual states' political systems vary considerably in the extent to which they provide effective governance, allocate governmental authority among different levels (e.g., federal, provincial, and local), maintain mechanisms for holding government officials accountable to the citizenry, and incorporate *bottom up* initiatives in policy-making, but all ultimately rest on a hierarchical organizing principle. Second, national legal systems contain several types of rules, each with a distinctive function in the overall legal system. Rules of recognition, some pertaining to the national legal system as a whole, and others to environmental matters in particular, stipulate how laws, regulations, and customs are adopted, amended, or repealed. Substantive rules perform any or all of three functions:

- *constitutive rules* define types of actor, conduct, or situation;
- *regulatory rules* stipulate the forms of conduct to be adopted or avoided in particular situations;
- *consequential rules* specify what can happen to actors as they conform or fail to conform to substantive rules.

In addressing air pollution, for example, any national legal system must contain constitutive rules defining what types and levels of emission constitute pollution, regulatory rules outlining how actors producing emissions should act to control those emissions, and consequential rules establishing rewards for compliance or (more commonly) punishments for non-compliance with the regulatory rules. Finally, environmental law's actual impact on actor behavior depends on varying combinations of willing support for the particular legal rules, more diffuse law habits that incline actors to respect the law in place even when they disagree with certain of its stipulations, material inducements to comply and material punishments for non-compliance.

The chronology of efforts to develop national environmental law is very similar around the world, a reflection of the strength and success of transnational and intergovernmental efforts to increase the salience of environmental concerns among governments and private actors alike. Concern about the impact of industry, agriculture, and urban life on the general condition of the air, lakes, rivers, and streams and waterways, and strong public demands for reducing air and water pollution, emerged in North America, Western Europe, and Japan at about the same time in the mid to late 1960s, as a product of changed beliefs about the origins and desirability of pollution. Though concern about air and water quality inspired national legal regulations in the

19th century (e.g., the British Alkali Inspectorate Act), even in 1950 the urban fogs, composed of smoke and particulates emitted by electricity generators, building heating or cooling systems, and factories, were viewed as unavoidable by-products of industrial prosperity despite awareness that urban dwellers experienced higher incidence of respiratory diseases. Within 15 years, governmental and popular attitudes had been transformed: greater concern about the human health effects of poor air quality, plus the diffusion of ecological worldviews conceiving of the biosphere as an interrelated whole in which change in one part had significant ramifications through all other parts, had transformed perceptions of pollution. Smoky emissions were no longer perceived as a sign of industrial might and prosperity, as a quasi-natural phenomenon beyond control, but as a bad that could be reduced or avoided through concerted human action. Incorporation of these changed perceptions into national policy was signaled by the US Clean Air Act of 1970, the first Japanese pollution laws of 1970, the German Federal Environmental Program of 1971, and the first European Community Environmental Program in 1973. Though some Leninist leaders continued to maintain that environmental degradation was a by-product of a private enterprise economy, the Union of the Soviet Socialist Republics (USSR) and its Eastern European allies had also begun adopting environmental laws.

Though the characteristic environmental problems facing Third World countries were different – stemming far more often from inadequate access to safe water, fuel, and sanitation than from industrial emissions – perceptions of a need for national environmental law were diffused rapidly to Third World governments through the preparations for and holding of the 1972 UN Conference on the Human Environment in Stockholm. While insisting at the Stockholm Conference and later multilateral negotiations that international environmental law must not become a vehicle for frustrating their own development efforts, Third World governments did accept the need to begin incorporating environmental concerns into their own national laws. India adopted its first explicitly environmental statutes in the early 1970s, though it developed a comprehensive framework statute only in 1986 after the Bhopal chemical leak revealed significant weaknesses in India's environmental law, and created strong public pressure for strengthening it. The Chinese government appeared unimpressed with concerns about the environment at the Stockholm conference, but some of its leaders were paying attention to pollution problems and other environmental concerns even before the political and economic reforms of the 1980s. Statutory changes were paralleled by administrative reorganization. The increased importance attached to environmental concerns was institutionalized by the establishment of an umbrella environmental body to coordinate governmental and private action. These bodies, often cabinet-level agencies or ministries, were established in the major Western industrial countries during the early 1970s and in most other countries during the 1970s and 1980s (Baker *et al.*, 1985; Janicke and Weidner, 1997).

The extent of change in substantive legal rules inspired by new concern with the environment, and the process by which it was accomplished, have varied from country to country, depending partly on governmental and public perceptions of the nature of environmental problems and the means available for their management, and partly on the national legal rules specifying how revision of substantive law should be accomplished. Because each national legal system has its own balance among statutes, judicial decisions, and customary practices as sources of law, the process of accommodating contemporary ecological awareness into law has thus differed sharply. In the UK and the civil law states of Europe, the primary emphasis has been on legislative amendment of the law codes. In the common law systems of Australia, Canada, India, New Zealand, and the USA, the process has involved a combination of legislative action and active judicial refinement of common law doctrines. In most Third World countries, the process has also involved legislation, but to varying degrees has also featured reinterpretation of traditional customary rules still used in the mix of indigenous and European-inspired law that forms the contemporary legal system. The global spread of problem definitions, relevant scientific knowledge, information about possible regulatory approaches, and model laws, encouraged by non-governmental environmental organizations like the World Wildlife Fund, Greenpeace International, quasi intergovernmental ones like the World Conservation Union, the United Nations Environment Programme, and other global or regional intergovernmental organizations, has not erased national differences, but has meant that there is considerable common ground in environmental discussions. International environmental agreements have also created a body of common goals for those states that have become parties, and on some occasions provided developing states with financial and other resources needed for effective monitoring and enforcement.

BASIC PRESUPPOSITIONS OF NATIONAL ENVIRONMENTAL LAW

Though some environmentalists have argued for treating nature generally or other species of life as having distinct legal rights apart from those of humans, national environmental law systems are generally based on constitutive rules that classify humans – individually and in their various collective groupings – as the subjects of law-holding rights and duties and responsible for their actions in complying or not complying with the law. Natural places and resources are classified as objects of law that the subjects of law may or may not use in particular ways as they go about their

activity. Individuals (natural persons) and collective entities (legal persons) are most often connected to natural places and resources through property rights regimes, which define who may have access to what places on what terms and may put any resources found in those places to use, either *in situ* or by removal for purposes of providing fuel for human activities requiring energy inputs of heat, light or power or for working, into tools and other usable objects.

Property rights regimes vary on several dimensions, but the most important involves the basis of de ning access rights. Four possibilities have been most prominent in human history: *open access*, *community property*, *individual property*, and *controlled access*. Under an open access regime, a particular natural place is open to anyone who can reach it, and any resources within it are open to use *in situ* or after being carried away on a rst-come rst-served basis. Little law is needed, other than prohibitions against attempts to close off the area. In community property regimes, which are historically far more usual than open access ones and formed the subject of Hardin s famous tragedy of the commons (*see* **Commons, Tragedy of the**, Volume 5), access is limited to members of a particular community most often de ned by geographical proximity or traditions of seasonal use who may use resources *in situ* or take them away according to community norms about shares. Local customary law most often governs membership in the community and members resource use; under modern conditions, the national legal system is needed to provide protection by af rming community property, and protecting it from intrusions by outsiders. In individual property rights regimes, an individual, household, or legal person enjoys exclusive access to the place that is its property and controls all resource use. In most national legal systems that property-holder (typically called the *owner*) may use the place or resources directly, rent or sell the place or resource rights to others, or leave the place and resources unused. Under controlled access property regimes, places and resources are governed by a central authority, which determines who may enter on what conditions and use resources *in situ* or take them away. Much of the debate regarding environmental law focuses on the relative advantages and disadvantages of each type of property rights regime for promoting environmental protection and sustainable uses. Most national legal systems assume that open access regimes are not viable for contemporary levels of population and intensity of resource use; they institutionalize varying degrees of reliance on community property, individual property, and controlled access regimes. Western legal systems, common law and civil law alike, rely most heavily on individual property rights regimes, though certain forms of community property and of controlled access through government licenses exist as well. Community property regimes are more common in non-Western systems, though they vary considerably from place to place. They coexisted with individual property rights even before colonial era impositions of individual property rights regimes by European powers, and are enjoying some revival today as scholars demonstrate their viability as modes of regulating resource use (e.g., Berkes, 1989; Ostrom, 1990). Bad experiences with Leninist central planning have discouraged primary reliance on controlled access through central licensing, but such systems remain in place for particular areas or activities in most countries.

Property rights regimes establish who is responsible for the use of a particular area and its resources, but do not guarantee environmentally sustainable patterns of activity. Most national environmental law is built on the basic assumption that environmental degradation occurs because individuals, households, rms, and other entities have no incentive to pay attention to the environmental burden of their activity, as long as they can pass all or part of the costs of coping with that burden to others. Thus, the primary task of national environmental law is to limit or even prevent such transfers. Successful prevention requires a variety of legal instruments because such transfers can occur geographically, economically, or temporally, and each type is best addressed through somewhat different rules. Geographical transfer occurs when wastes or pollution are allowed to spread beyond the edges of an actor s own property, and are inhibited through substantive legal rules prescribing that ownership or other property rights do not include a right to cause harm to others or their property, and through procedural norms requiring property holders to control harm equally stringently, regardless of whether it occurs on or off the property. Economic transfers occur when one actor s activity degrades a common resource like water or the atmosphere, leaving others to pay for coping with the degradation. Economic transfers are inhibited through the polluter pays principle and emissions charges, water or land use fees, charging cleanup costs or nes for excessive emissions that reinforce nancial incentives to avoid across-actor transfers. Transfers across time occur when actors handle a problem with a quick x, such as underground storage of highly radioactive wastes, which leaves the basic environmental problem to be solved later at considerable expense. These transfers are inhibited by legal norms lengthening time horizons, such as intergenerational equity, the precautionary principle, and mandatory prior environmental impact assessment. Since most national legal systems have permitted such transfers in the past, and actors came to view such transfers as part of life, regulation to limit or end such transfers strains previously held beliefs and will remain a source of resistance to environmental law until ecological awareness, social mores, and habit make avoidance of such transfers the basic pattern of social behavior.

Environmental law has been a particularly dynamic branch of national law in all countries, marked by a high

rate of legislative, administrative, and judicial activity for several reasons. Public concern about the environment has expanded to encompass more problems, creating demands for more regulations. Improved scientific understanding of ecosystem dynamics, and the impact of human activity on those dynamics, has been a source of demands for additional regulations or reformulations of existing ones to take advantage of current knowledge. Technological change in particular sectors of the economy, together with the broader shifts in patterns of production and distribution of goods and services resulting from the increasing volume of cross-border transactions, and the transition from the mass production-mass consumption pattern of the 1950s to the current information-intensive and flexible production pattern, have also required adjustments of legal rules.

The most widespread change in the national legal context has been a shift in the general approach to limiting inter-actor transfers of environmental costs. Most of the pollution control legislation adopted in the 1970s relied on a command and control regulatory regime, under which designated government agencies were given authority to determine environmental standards and to define the means by which other actors would meet those standards (*see* Demand Management, Volume 4). Thus, the US Clean Air and Clean Water Acts authorized the Environmental Protection Agency to determine national air and water quality standards and then to stipulate the technological means by which users or air or water would control their emissions so that the standard would be met. Command and control operated more flexibly in countries like Japan and the UK, where statutes establish broad guidelines which government agencies apply flexibly through discussions with industry, and there is a strong emphasis on abating pollution through use of specified pollution control or cleanup technology. In India, it took the form of attaching conditions to permits for business activity, an environmental extension of the well-established license raj approach to regulating private firms' general business activity.

A command and control regulatory regime has the advantages of ensuring uniform standards throughout the territory of the state, assuring that all actors accomplish some minimum degree of environmental protection, and placing all firms under similar cost burdens. Yet it also has the disadvantages of cumbersomeness, since a central agency has to make far-reaching and highly detailed decisions and, at least as applied in many countries, an inflexibility inhibiting adjustments to local conditions. While a command and control approach can, in principle, promote technological innovation through adoption of progressively tighter emission limitations and other technology-forcing devices, in practice it often inhibits innovation because firms and others have no incentive to do anything but adopt the government-mandated pollution abatement technology.

By the mid 1980s, the problems with command and control regulatory regimes were being criticized from two distinct standpoints. One, generally identified with the political right and neoliberal economic theory, argued for replacing command and control with market-based measures altering actors' propensity to create environmental harm either by increasing the cost of causing harm or subsidizing harm avoidance. Advocates argued that measures like emissions taxes, tradable emissions or resource extraction permits, and rules permitting firms to choose how they will meet government-determined emissions or ambient quality standards, would yield greater environmental improvement at lower cost than the command and control approach. The other, generally identified with Green parties and other elements of the political left, argued that central regulation should be replaced by a radical decentralization of authority, transformation of production and consumption patterns away from energy and resource-intensive consumerism towards localized self-sufficiency, and a populist emphasis on local knowledge and local justice. Both views share a skepticism of centralized government: the first prefers heavier reliance on private ordering through contract while the second prefers local self-organizing as a way to resist capture of the law-making process by corporations and experts. Though most forcefully expressed in the USA, market-based approaches grew more influential in many countries during the 1990s. This shift reflected some experience with particular command and control regulatory regimes, but even more the problems of centralization, opacity, and non-accountability encountered in countries where state-owned enterprises operate all or most major industries. These were most obvious in the USSR and other Leninist states.

As reliance on market-based measures increased in the 1990s, criticisms of such reliance continued. Advocates of local self-sufficiency continued to make their case for a complete turning away from industrial ways of life. A broader coalition, ranging from the socialist left to the center, continued to support government regulation though seeking greater transparency and accountability in corporate and governmental activity. Demands for greater public participation in environmental decision-making were reflected in support for prior environmental impact assessment, right-to-know laws regarding hazardous substances and workplace safety, access to government-held information (freedom of information laws) and opportunities for public comment before major public projects began. National environmental law in most countries thus remains the subject of considerable contention as actors pursue both their immediate concerns and their visions of how society as a whole should function.

National environmental law has also been affected by the ongoing debate about how to specify the desired environmental conditions. Law-making in the 1960s and 1970s was

marked by strong debate between advocates of regulating by emissions standards, which would specify the maximum amount of pollution that could be released by any single pollution source (such as a power plant, factory, or vehicle), and by ambient standards, which would specify the maximum acceptable level of each particular form of pollution (such as sulfur dioxide, nitrate, or lead) in the water or air. The debate about the relative merits of focusing on the individual sources of pollution, or the aggregate effects of pollution-generation, was eventually resolved in a both/and fashion, in which ambient standards provide the baseline from which regulators determine whether and how to regulate individual emitters. More recently, the connections between particular emissions of pollution and the ambient air or water quality have been drawn more precisely by using the *critical loads* methodology to determine acceptable pollution rates by reference to the absorptive capacity of the air, water, land, or vegetation of an area, as well as ambient air or water quality (*see* Critical Load, Volume 3).

Almost all national legal systems must also address problems that arise from fragmentation of governmental authority. This is most obvious in federal states where lower-level government units (states, lander, provinces, etc.) have constitutionally-protected powers, but also occurs in unitary states if actual management of certain activities is delegated from the national government to more localized authorities. In the Netherlands, for instance, 125 water boards, each responsible for a particular part of the country, administer many aspects of water use policy. The same dilemmas that plague development of international environmental law – whether standards should be uniform or respond to differences in local conditions, whether environmental measures in one area, such as recycling requirements, serve also or primarily as ways to favor local over outside firms, whether differences in standards trigger races to the bottom as firms seek to relocate activity to places with less stringent rules, or races to the top as firms conform to the generally stricter standards prevailing in the industrial states so they can sell in those markets – mark the development of national environmental law. This is particularly noticeable in large federal states like India or the USA, and is also prominent in efforts to develop European Union (EU) environmental law.

ENSURING COMPLIANCE WITH NATIONAL ENVIRONMENTAL LAW

At first glance, political centralization and organization of states along hierarchical principles appear to make the problem of securing compliance with national environmental law one of effective application of the administrative and coercive apparatus of the government. Yet even tight hierarchies founder if too many citizens react to law with inaction, grudging partial compliance, or outright violation. National legal systems typically seek to draw on all motivations for compliance present among the private and other actors whose conducts are addressed; advocates of different regulatory regimes simply place different degrees of emphasis on different motives. Advocates of command and control rely, in the last instance, on a strong administrative state capable of monitoring actors' behavior, identifying non-compliers, and imposing sanctions that will induce non-compliers to change their behavior. Advocates of market-based regulation rely more on aligning legal rules with actors' current self-interests or using law to change conduct by shifting actors' incentives to the point that compliance becomes more highly valued than non-compliance. Advocates of local sustainability rely primarily on persuasion to accept new values and beliefs, socialization into new norms of conduct and psychological desire to remain in good standing with others in the community. Yet when considering possible environmental law, advocates of all three share certain appreciations. All treat actors as responsible agents who are capable of thinking about their situation and making conscious choices. All assume the legal system will provide sufficiently clear rules that compliance can be distinguished from non-compliance. All assume that actor behavior and the legal process are both transparent enough that particular non-compliers can be identified and called to account, and that the legal system will afford opportunities for determining whether a particular instance of non-compliance is the result of conscious action, ignorance of rules (though most national legal systems do not accept ignorance as an excuse for non-compliance), accident, or intervening circumstance that made compliance impossible, before deciding whether a particular violation is best addressed through provision of information (about the environmental problem or about the law's requirements), assistance in overcoming inability to comply, incentive to comply, or punishment of failure to comply.

The national environmental law of particular countries use several mechanisms for securing compliance, with the mix actually used reflecting national legal traditions on the broad problem of promoting obedience to the law. National systems tend to place great emphasis on administrative elaboration of standards, monitoring of actor compliance, and imposition of penalties for non-compliance. The more heavily a national legal system relies on administrative elaboration and enforcement, the greater the burden on government agencies to recruit and retain technically qualified personnel capable of formulating rules that will produce the desired changes, and taking action when pollution or other environmental damage results from human activity. At the same time, no government relies solely on government personnel in environmental rule-making and enforcement. Administrative rule-making typically involves a degree of prior discussion

with the private and public actors likely to be covered by the rules. Though posing dangers of regulatory capture, in which the least enthusiastic of the regulated determine the pace and content of rules, prior discussion affords opportunities to exchange views, explain the rationale for particular rules, and secure target actors' perspectives on the feasibility or unfeasibility of different possible rules. Many national legal systems attempt to limit the danger of regulatory capture through provision for public comment on proposed regulations before they go into effect, or for prior discussions that include other groups likely to be affected by target actors' conduct (e.g., the government–labor–business consultations in the European corporatist systems).

Use of administrative rules requires establishment of systems for monitoring behavior. Many of these are mechanical; scientific instruments located at or movable to particular sites for direct observation of emissions or ambient air or water quality. Satellite-based remote sensing systems offer possibilities of monitoring changes in land use, vegetation patterns, and habitats as well as identifying sources of significant pollution plumes. Complaints from nearby residents or environmental groups, who do their own monitoring, often supplement administrative agency efforts.

Involvement of national courts in applying environmental rules varies considerably. In many common law countries, much environmental law was created as private persons or organizations, able to show they suffered direct damage to themselves or their economic activity from pollution, invoked common law notions of nuisance, negligence, and trespass against those responsible for the pollution. Dealing with pollution did require reinterpretation of these traditional concepts, but many judges shared the rising public concern and were willing to innovate. The flowering of public interest litigation, most notable in the USA and India, also required some changes in the rules defining who has a right to sue whom over what, to allow suits by environmental and other organizations rather than individuals or firms and to permit class action suits in which some of those harmed sue on behalf of all. The very high level of court involvement in formulation and enforcement of environmental law in the USA reflects that country's conditions, and has three main sources. First, the general common law emphasis on judicial interpretation has been much enhanced through the principle of judicial review, which allows individual citizens and other private entities to challenge the consistency of statutes or particular administrative actions with basic constitutional provisions including the Bill of Rights. Second, the national government's administrative apparatus is relatively weak compared to those of other industrial states, and many federal statutes compensate for that weakness by allowing individuals or firms to sue other entities for non-compliance or to sue government agencies over non-enforcement. This practice dates back at least to the Sherman Anti-Trust Act of 1890, and has been incorporated into the major US environmental statutes, even though the US Supreme Court has reined in the practice since the mid 1990s by narrowing the criteria for determining who is eligible to sue over what matter. Whereas environmental organizations file most of the citizen suits seeking better enforcement, business firms and other targets of environmental regulations are the source of most challenges to the constitutionality of some statutory provision or administrative act. Third, the separation of powers into three coordinate executive, legislative, and judicial branches of government often promotes judicial activism by emphasizing the courts' status as a coequal branch of government. Though sharing the common law tradition, the UK and many of the Commonwealth countries have parliamentary political systems where executive and legislative powers are fused, and courts defer more to parliamentary sovereignty in matters regulated by statute than do their US counterparts. In civil law states, the courts' role is even more constrained; there, administrative agencies and review panels are the primary mechanism for applying environmental rules and determining whether non-compliance has occurred; citizen suits are not institutionalized in the legal process and judicial review is limited to constitutional courts or courts of cassation (annulment) with authority to vacate (to make void) administrative actions. For states who are members of the EU, the provision for European Court review of national compliance with EU directives and regulations has created possibilities of regional judicial review.

Yet judicial activism can occur even where there is no formal separation of powers and no common law system. Japanese courts played a significant role in the initial development of environmental law in Japan, through judgments requiring businesses responsible for major pollution damage to provide extensive compensation to victims and pay punitive damages. Judicial activism has been more constant in India (a parliamentary common law state), despite a notoriously slow court system, where advocates of the poor and marginalized have taken up environmental cases as one part of a broader program of public interest litigation encouraged by the Indian Supreme Court's liberal interpretation of standing to sue, citizen suit provisions in Indian environmental statutes, and the distinctive Indian legal customs regarding writs of *mandamus* (which order a government agency to exercise its legal authority), prohibition (which enjoin a government agency from acting in a way contrary to statute), and *certiorari* (which remove a case from a lower court or an administrative agency to a higher one) allowing private citizens to sue government agencies for actions or inactions.

National environmental law thus exists at the confluence of well-established national legal traditions and increasingly

widespread knowledge about ecosystems and the impact of human activity on them. Differences in ecosystems, patterns of human activity, cultures, and legal traditions mean there will always be considerable variation in national environmental law. All national environmental law systems face the challenge of helping move societies from environmentally-oblivious to environmentally-aware modes of human conduct. Each will find its own way to ensure that national law and property rights regimes incorporate environmental values and that actors are discouraged from persisting in patterns of activity that shift the environmental burdens of their activity to others through geographical, economic, or temporal transfer. For an interesting comparison, *see* International Environmental Law, Volume 5.

REFERENCES

Baker, R, ed (1997) *Environmental Law and Policy in the European Union and the United States*, Praeger, Westport, CT.

Baker, M, Bassett, L, and Ellington, A (1985) *The World Environment Handbook*, World Environment Center, New York.

Berkes, F (1989) *Common Property Resources: Ecology and Community-based Sustainable Development*, Bellhaven Press, London.

Findley, F W and Farber, D A (1999) *Cases and Materials on Environmental Law*, 5th edition, West Group, St Paul, MN.

Janicke, M and Weidner, H, eds (1997) *National Environmental Policies*, Springer, Berlin.

McGregor, G (1994) *Environmental Law and Enforcement*, Lewis, Boca Raton, FL.

Ostrom, E (1990) *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge University Press, Cambridge.

Rosencranz, A, Divan, S, and Noble, M L (1991) *Environmental Law and Policy in India*, Tripathi, Bombay.

FURTHER READING

Berry, J F and Dennison, M S (2000) *The Environmental Law and Compliance Handbook*, McGraw-Hill, New York.

Koeman, N S J (1999) *Environmental Law in Europe*, Kluwer Law International, The Hague.

Pardy, B (1996) *Environmental Law: A Guide to Concepts*, Butterworths, Toronto.

Shea, E E (2001) *Environmental Law and Compliance Methods*, Oceana, New York.

Natural Law

see Nature (**Volume 5**)

Nature

Nature is notoriously difficult to define. In his touchstone definition in his anthology *Keywords*, the social critic Williams (1983) says that it is both perhaps the most complex word in the language and that any full history of the uses of nature would be a history of a large part of human thought.

In light of these sobering judgments, a short definition should simply point out the major distinctions in the use of the word, and some of the words with which it is often paired or compared.

The word derives ultimately from a cluster of terms around the past participle of the Latin verb, *nascor*, to be born; terms that include the use of *natura* as a noun indicating blood-relationship, physical affinity, etc., and that are generally associated with the idea of something having a nature or a quality of its own. This becomes linked with a second cluster of terms and meanings that suggest that there is a larger nature of things (perhaps with a capital N for Nature). Roman philosophers like Cicero and other Stoics intimated that *natura rerum* – the nature of things – might be an elemental god or physical force underlying all things. A third set of associations set the forces of nature (even the personified goddess Nature or the great god Pan) in opposition to the forces of civilization.

The first set of associations – something having a nature of its own – has a long subsequent history, ranging from the idea that to be natural is to be true to oneself, to today's claims that something is 100% natural. Natural here is being given the connotation of truth, purity and integrity. A different, but comparable idea is the Chinese concept of the *Tao* (the way), which suggests that the right path for human living is to match one's self to the moral structure (the Tao) inherently part of the natural world.

In the West, a related idea (traceable back to Cicero) is that of *natural law* – a law that does not need to be legislated because it is so obviously part of human experience, and allegiance to which aids human flourishing. This would have a long history, culminating (a) in the idea of human rights as inalienable because they are (if you subscribe to a doctrine of rights) part of the definition of what it is to be human; and (b) in various controversial claims about natural law by conservative institutions, e.g., slavery is natural, woman are naturally nurturing, homosexuality is unnatural, and so on.

In the Middle Ages, and into the Modern Era, the dominance of Christianity cast a suspicious shadow over all versions of being natural, that allied human nature with the natural world, when that world was seen as the pagan world of unredeemed physical nature. It was considered

to be tempting and dangerous, because it was disordered and distant from God. The ultimate Paradise of Christian mythology was either a redeemed garden (of Eden) or a city, Jerusalem. Wilderness was the place of demons and temptation, as it had been for the Israelites and Jesus both, as they wandered in the wild desert. For the medieval Christian, the important task was not to align oneself with the forces of wild nature, but with humankind's true nature which was in alignment with the ordering power of God. Human beings were exhorted to redeem their wild, corrupt nature by moving towards their second nature – a cultivated nature that had to be worked upon, and transcended. In several places in the Bible, Heaven is described as the city of light.

In the Romantic era (1750–1850), this third set of associations – nature versus cultivation – was turned on its head. Given the pressures of the Industrial Revolution, and the widespread repudiation of hierarchical political and social structures based on outmoded versions of order, romantic writers claimed a primal affinity with nature, because nature was now a place of refuge, and positively comparable as a complex, organic whole, to an increasingly alienated atomized machine culture. Nature was once again the original home from which human beings had fallen into civilization. The destruction of nature began to be seen as analogous to the increasing destruction of elements of human nature by the forces of industrialization.

This recasting of nature as positive has had a number of effects, not all of which are themselves positive. For example, the image of a pure primal energy force associated with nature – the life or death force – influenced the rise of Nazi blood politics. Other effects, which have still not been fully worked through, include the raising of fundamental questions about whether there is such a thing as human nature, whether human beings are lying to themselves when they consider themselves to be independent moral agents capable of rising above nature, and so on. On the other hand, environmentalism owes its very existence to the positive reevaluation of the need for unhumanized nature, and the protection of natural processes (which we do not fully understand) in the modern world.

Finally, it is worth noting that the second set of associations with which we began – Nature as a catch-all term for everything in the material world – is certainly widespread today in the sciences, where nature is synonymous with the natural world, that is, the physical world. Because of the complex web of historic and philosophical meanings attached to the term, however, it often causes confusion wherever different disciplines meet, especially across the physical sciences, social sciences, and humanities.

REFERENCE

Williams, R (1983) *Nature in Keywords*, Fontana, London.

PETER TIMMERMAN *Canada*

Naturism

see Ecofeminism (**Volume 5**)

New Ageism

New Ageism is a diffuse, piecemeal movement of groups and individuals both inside and outside environmentalism. They share a conviction that the world is undergoing a major transformation, involving a shift in consciousness, towards a new mode of being. Contemporary New Agers follow the astrologers contention that a new age is entered about every 2000 years, maintaining that the Age of Aquarius is displacing the current Age of Pisces. Pisces has been dominated by polarization and con ict between cultures, civilizations, religions and races. Western consciousness has particularly been driven, through mistaken conceptions of reality as composed of irreconcilable dualities like mind and body, male and female, society and nature.

By contrast the Aquarian Age will be one of harmony, holism, balance and heightened moral and spiritual awareness. Society and nature will no longer be separated. Transition to the New Age will follow changes in the consciousness of some individuals. They will form disparate, scattered groups promoting such things as ecology, feminism, spirituality, etc., and will eventually coalesce to form a critical mass, catalyzing all peoples into the new thinking, thus creating a global consciousness. This process will be facilitated by advanced communications technology, such as the internet.

New Ageism draws on an eclectic mix of spiritual and other traditions, such as Eastern mysticism, Western religions (the ideas of the Catholic priest, Teilhard de Chardin, are seminal), alternative medicine, ecological science and new physics. Its idealistic analysis and prescriptions focus on changed attitudes as the spur to changed behavior. It rejects the materialist analyses of old-style politics, which analyze behavior in relation to economic class interests. Paradoxically, perhaps, New Ageism often revives

pre-modern beliefs, such as paganism, pantheism, animism and vitalism.

DAVID PEPPER *UK*

NIMBY (Not In My Backyard)

NIMBY is an acronym indicating that people living near a proposed development (e.g., waste dump, highway bypass, nuclear power station) strongly object to having the development situated near them, although they would not necessarily oppose the development if it were located elsewhere. A related expression used in parts of the USA is locally unacceptable land use (LULU).

R E MUNN *Canada*

Non-anthropocentrism

see Environmental Ethics (**Volume 5**)

Noosphere

The *noosphere* is a sphere of interaction between humans and nature where conscious human actions become a major determining factor of its evolution. The definition of noosphere has evolved from a spiritual notion to a more scientific one. Two French philosophers, Edouard Le Roy and Pierre Teilhard de Chardin introduced the term in the early 1920s. Teilhard de Chardin described the noosphere as the conscious and thinking membrane enveloping the globe, whose emergence was caused by the appearance of humans on the Earth and by the development of human intelligence. Vladimir Vernadsky, the Russian natural scientist and philosopher, gave a more scientific connotation to the word *noosphere*. He envisioned the noosphere as an inevitable evolutionary stage of the biosphere where human thought and activities become major forces in the evolution of the Earth. Given that humans can change the chemical composition of air and water, create artificial minerals, reroute rivers, and breed new types of animals and plants, their power to transform the environment is now comparable to the transforming forces of nature. Vernadsky saw anthropogenic environmental changes as inevitable, manifesting the beginning of a new evolutionary stage of the biosphere.

See also: **Vernadsky, Vladimir**, Volume 2.

REFERENCES

Le Roy, E (1927) *L'exigence Idealiste et le Fait d'Evolution*, Boivin, Paris.

Teilhard de Chardin, P (1980) *The Phenomenon of Man*, English translation, Harper Collins, London.

Vernadsky, V (1944) A Few Words About Noosphere (in Russian), *Adv. Mod. Biol.*, **18**(2), 118–120.

GALINA CHURKINA *Germany*

Not In My Backyard (NIMBY)

see NIMBY (Not In My Backyard) (**Volume 5**)

422

Phase Shifts or Flip-flops in Complex Systems

Henry A Regier[1] and

James J Kay[2]

[1]*University of Toronto, Toronto, Canada*

[2]*University of Waterloo, Ontario, Canada*

Living things self-organize into systems that must be partly closed to maintain identity and integrity, but must also stay partly open in order to accumulate high quality resources in categories such as mass, energy and information and to void wastes also belonging to these categories. A living thing must also stay partly open in order to interact, often reciprocally, with other living systems and with features of its non-living environment. Non-living things may also self-organize systemically, but in less complex ways. Living things in general have evolved capabilities to self-organize into a number of different complex phases, states or stages and to shift from one of these to another in response to changes in internal and external phenomena. Many living systems, especially organisms, have evolved capabilities to proceed autonomously through cycles of such stages. The life history of an insect, for example, exhibits a progression from fertilization to egg to larva to pupa to adult to dead body. Such a one-way or ontogenetic development with transformations between stages may be perceived as a special case of the kind of organizational changes that we focus on here; we do not include such specialized one-way cases in the present discussion. Instead, we focus on systemic reorganizations that occur between different relatively stable states of a living complex system and are reversible, more or less. (Of course, strict reversibility is not possible because living systems are subject to the second law of thermodynamics.) The kind of shifts on which we focus may occur abruptly and haphazardly in response to a particular kind of stimulus that is relatively unexpected and catastrophic in its context; or it may occur in a more orderly way for some stimulus that is always expected with significant probability in its context.

We emphasize that, for this essay, the concept of shift between phases has a similar meaning to flip between stable states, and a two-stage flip-flop between alternate states. We do not include a concept of ontogenetic, one way changes through a genetically pre-ordained series of stages; to do so would increase the scope of our essay beyond our present purposes.

We include as natural, some features of living things that some people would refer to as cultural. Some species such as humans may exhibit strong cultural features as well as the necessary natural features, while other species such as bacteria may exhibit few if any cultural features. We do not presume to understand fully any natural and natural/cultural features of reality that we address. Some people may perceive some features of humans and human societies to be both cultural and unnatural; our intent is not to include consideration of such unnatural features in the present essay.

The natural base of strongly cultural living systems must remain open to mass, energy, and information flows, as is the case with all living things. Cultural manifestations are also open culturally in that they thrive on such resources as beauty, ethical goodness, respect and caring and try to divest themselves of the opposite of these qualities; ugliness, evil, disrespect and hate.

ECO-STUDIES

Academically, numerous overlapping kinds of eco-studies may be perceived; we refer explicitly to five kinds here. The prefix eco-(from the Greek *oikos*, home as an edifice with its inhabitants) relates to a generic home-like phenomenon in which a living entity interacts with its living and non-living environment, which may include an artificial dwelling place. This approach parallels the combined

consideration of text and context by some scholars with a focus on information.

A living entity may have living sub-entities within it. It follows from this basic perception that study of an eco-phenomenon transcends any conceptual dichotomy between analytical reductionism and observational or intuitive holism. Neither the reductionistic nor the holistic approach is wrong; instead each is just half-baked. Sometimes the eco-approach is termed *holonistic*, with the *hol* referring to whole, and the *on*, referring to part and this makes explicit the transcendence of the reductionism–holism debate.

The versions of eco-studies include:

1. *Eciatrics* (from *eco* and *iatros* as pertaining to physician or medicine) includes public health studies of humans and domestic animals with respect to effects of pathogens on organisms and populations, as in epidemiology, and risk assessment studies of effects of toxins and contaminants on the organismal and population health of humans and other species.
2. *Ecology* or *ecologics* (from *eco* and *logos* as pertaining to reasoned knowledge) here narrowly defined and including interactive aspects of meteorology, hydrology, geology as well as conventional biology, and relating to the contemporary real time processes of biological evolution.
3. *Economics* (from *eco* and *nomos* as pertaining to law or laws) in which the practices of benefit–cost accounting and monetary risk assessment may be of minor significance overall compared, say, to an understanding of the self-organizing features of local sharing of ecosystemic goods/bads and services/disservices in a communitarian regime, and/or to global trade within a constrained free enterprise regime.
4. *Ekistics* (from *oikizo* meaning the establishment of a settlement) relating to the self-organizational development of human urban and rural settlements, physical infrastructures, industrial complexes, the information highway, etc.
5. *Ecumenics* (from *oikoumenos* meaning all beings in community) here with a focus on participatory governance within an extended community of interest groups or stakeholders, and including informal, extra-constitutional decision-making structures and processes that transcend boundaries between people of different nations and jurisdictions and thus complement the formal, constitutional laws, treaties and agreements of autonomous nations, as those were once perceived.

To study living systems, the psychologist and game theorist Rapoport (1986) has emphasized identity, organization and goal directedness. In a compatible way, the writer Koestler (1969) focused on self-organization, hierarchy and openness. Koestler excluded some connotations of hierarchy, e.g., the control hierarchy within a large military unit. Instead he referred to reciprocal nested interactions among *holons*, i.e., things that act as wholes but also as parts of larger wholes, within a form of organization that may be termed a *holonarchy*. A well-known example of a loose holonarchy is the interactive complex of nested formal and informal organizations within a multi-tiered form of liberal democratic governance, as may be described in ecumenics.

The conceptual domain that includes the five nested versions of eco-studies sketched above, plus some others, may be termed an *ecogeny*. The suffix refers to genesis, which refers here to emergent self-creation guided in part by *a priori* genetic hardware as well as by learned software, both augmented by new autonomous programming. Cooperative study by colleagues and ourselves in the North American Great Lakes Basin, for example, makes use of all those versions of eco-studies within an ecosystem approach that may be perceived as a special case of integrated ecogenic studies. Haas (1999) might refer to our self-organizing study group as a kind of epistemic community and we might refer to ourselves more specifically as an eco-epistemic community.

Here *epistemic* relates to a compendium of shared perspective and knowledge that is offered as having particular relevance to a major cultural challenge or opportunity. Some scholars may relate this epistemic concept to that of an expanded version of what Thomas Kuhn referred to as a scientific paradigm (see Fuller, 2000). Incidentally, we have chosen not to include an explicit notion of paradigm shift, but note that our version of that notion is implicit throughout. In other words, what we here define as *ecogeny* has flipped from being a suppressed theme within modernism to become a key theme within post-modernism.

Many non-living as well as ecogenic systems can exist in more than one phase, state or topological domain of attraction as the ecologist C S Holling has long emphasized. After being harangued by Holling for many years to provide him with Great Lakes data that would demonstrate ecological phase shifts, one of us, Regier, finally asked himself the question: If phase shifts have occurred, where and how could we look for evidence of such events? He decided to search for limits to the adaptive self-organization of a fish association (considered as a holon) to an array of intensifying stresses by humans in the larger environmental holon, and then examine whether another phase self-organized beyond a limit or threshold. Sure enough, the evidence was there, though it may not have been fully consistent with Holling's expectations. (See section titled Benthic and Pelagic Attractors in Ecosystems for more information on this issue.)

Researchers on natural and cultural systems once used single phase, deterministic, cybernetic systems as models. Any phase shift was then perceived as a catastrophic failure of pre-programmed cybernetic adaptiveness. Researchers

now perceive natural/cultural systems to be self-organizing and *morphogenetic*, and some phase shifts are just part of the adaptive, emergent capability of the evolving system. (Here morphogenetic implies that the forms or structures as morphs as well as the related processes of living things emerge through self-organizational genesis.) The death of a living system is an ultimate kind of phase shift when all adaptive capabilities are over-ridden, because the stresses either are too intense or act too rapidly.

When ecogenic realities are changing rapidly, as is now the case in many parts of the biosphere, one might expect adaptive and emergent phase shifts to be common, and catastrophic shifts all the way to systemic death to occur as well.

Selected examples are sketched below to demonstrate that phase shift events now occur ubiquitously. We note again that phase shifts may be referred to as flip-flops between organizational states as in discussion of major back and forth changes in the earth's climate that have seemingly occurred suddenly in the past. We start with non-living systems, then turn to living systems in which cultural features are not emphasized, and then to living systems in which cultural features are emphasized. But we do not address unnatural things like any divine acts of deities or things for which no empirical data are conceivable.

EXAMPLES OF PHASES AND PHASE SHIFTS IN BIOPHYSICAL SYSTEMS

Different Phases of Water

Everyone knows that water can occur in different phases, e.g., liquid, solid and gas. When some research physicists focused on phase shift phenomena with water, their model implied the existence of a fourth phase, under special circumstances. They called it the *gaquid* phase and found that it materialized under the expected artificial combination of temperature and pressure.

When a liquid water system cools through 0 °C it freezes to become a solid water system. The detailed features of the freezing process are anything but deterministic in a simple way, when viewed at a molecular scale. In nature, it is generally not possible to predict either precisely how and when liquid water will freeze during a cooling regime or the crystalline form of ice that results.

The freezing phenomenon itself is of major importance to those aquatic living things which have to find some way to adapt anatomically and/or physiologically and/or behaviorally to freezing. Such adaptations in turn may be perceived to involve phase shifts in anatomy and/or physiology and/or behavior.

At a more macroscopic level, the kind of ice that forms in a stream is important ecologically. Anchor ice, frazzle ice and surface ice have quite different implications for brook trout, for example. From an ecological perspective, each of these involves a different kind of phase shift from liquid to solid.

Different Phases of Flow Dynamics

When water flows at low velocities across a smooth surface, it may move as a sheet with laminar flow; at higher velocities it may flow in complex turbulent swirls. Turbulent flow in turn may occur as a number of different sub-phases. Researchers in fluid dynamics have focused much scientific attention on: what determines which phase of flow regime will occur in a particular context; the macroscopic features of the threshold between the phases; and the more microscopic details of the phase shift process itself. Their understanding, condensed in the form of a Reynolds Number, allows them to predict quite accurately a phase shift between turbulent and laminar flow within a pure liquid. The more particles or lumps in the system, the less accurate any forecast based on such a pure system Reynolds Number will be.

The issue of phases and phase shifts in the flow regime is of key importance to aquatic organisms. While migrating, for example, salmon may use a standing wave in turbulent flow to catapult themselves over an obstruction in a river. To rest, a fish may select a site of laminar flow downstream from an obstruction and adjust its fins so that it rests lightly on the bottom.

One dramatic example of such phase shifts in fluids is the well-defined transition from conductive to convective heat transfer. If a fluid is contained between a hot and a cold boundary, heat transfer will occur through the fluid. Consider thin soup in a pot on a stove as an example. For small temperature differences between the surfaces, the heat transfer is by random molecular motion, *conduction*. However as the temperature increases, a critical point is reached at which a new phase emerges, convection. With convection the fluid molecules self-organize to move together in rolls or Benard cells as coherent wholes. As the temperature difference between hot and cold boundaries increases, the original configuration of Benard cells may self-organize into a more complex configuration. With each phase shift, more heat is transferred, per unit time, than in the previous phase.

At a human scale, avoiding the phase shift from conduction to convection is at the heart of the design of double pane and triple pane windows. These windows are designed so that convective heat transfer cannot emerge, thus limiting the heat loss through the window to that of conduction.

At a regional scale, on clear summer days with a particular kind of atmospheric stratification, the Sun's warming of the land surface leads to warming of the overlying air. Bubbles of such warm air of decreased density may then form spontaneously and episodically to rise and erupt into

an overlying cooler layer; these bubbles of warm clear air may transform into cumulus clouds within the cooler layer. In particular locales, successive fluffy clouds that emerge may be evenly spaced in the drifting upper layer.

At the global scale, the emergence of large *convection* cells in the atmosphere, called Hadley cells (*see* **Hadley Circulation**, Volume 1), helps to transfer heat from the equator to the pole, thus reducing north-south temperature differences between the equator and the poles.

Many kinds of phase shifts may occur within the hydrosphere and the atmosphere, since these spheres consist of fluid components with different properties acting and interacting across boundary layers under the influence of solar radiation.

Flows in Water Courses and Flood Plains

Rivers carve water courses or channels out of the substrate to accommodate their flows during most days of a year, but not during high flow rates when flooding occurs. It is important to note that the course that a river cuts through a landscape is usually not large enough to include the flows of normal annual floods. The area known as the floodplain and its valley in turn have also been created, at least in part, by the continually shifting meandering river course, and especially by its floods. Hydrologists do not fully understand why a river self-organizes as a biphasal phenomenon, with one phase for small to moderate flows, and another phase for large flows. But hydrologists have inferred generalizations about the channels and flood plains of particular aquatic systems that have adapted over long periods of time to the underlying geology, under the influence of climatic regimes that have operated in quite predictable ways.

A river channel, flood plain and the larger valley are always dynamic; they may approach a steady state under conditions like those sketched above, but they are never in equilibrium with their surroundings. For example, a river never stops dissolving and eroding its substrates and walls. Living organisms, especially large plants and animals that eat the plants that live within and beside a river, strongly influence the self-organizing hydrological activities and usually have a taming influence on the waters, to the advantage of the organisms.

Cultural practices in a watershed strongly modify the temporal and spatial features of the natural flooding regime, e.g., by putting a river into an engineered channel, or by cutting down forests farther upstream. In effect a flooding phase shift occurs more frequently during a year if the watershed becomes ecologically degraded or is developed inappropriately, and the river has not had enough decades to adapt to those changes. Thus conventional unsustainable development has led to more frequent and more rapid phase shifts of the flow regime, and thus to more frequent and intense periods of turbulent stress for the living system of a river. Such a river may be perceived to be angry and may not approximate benign steady-state conditions at any time of year.

Phenology and Seasonal Phase Shifts

In temperate regions as in the Great Lakes Basin, the annual cycle involves a spring and summer phase that is dominated trophically by composers that use the sunlight's energy for photosynthesis, e.g., plants, and a fall and winter phase dominated by decomposers that use the energy of organic matter derived from photosynthesis and stored in summer, e.g., animals and fungi. Each of these phases and their transitions are intricately choreographed in what could be called a pristine, *old growth* state of the Basin ecosystem. The term is extended here beyond its familiar use as in old growth forest, to mean any ecosystem that has evolved its own complexity under benign conditions over extensive periods of time. Unsustainable development disrupts the natural self-organizing processes, so that the sequential features within phases and particularly during the phase shifts become less predictable, from the perspective of the extant species.

In warmer regions, the climate may include a wet season and a dry season. Cyclical changes, somewhat similar to those sketched above occur here, too.

Benthic and Pelagic Attractors in Ecosystems

Moderately deep natural aquatic ecosystems that are not much influenced by technologically careless humans usually have clear waters and a biotic association or benthos that is linked strongly to the bottom of the waters. In effect a benthic attractor or self-organizing system then emerges to serve its own ends, for example by preying on the open water association or *pelagos* to keep the water above it transparent and by hoarding phosphates and other nutrients that deprive pelagic organisms of these resources. Ecologically, the benthic association has similarities to the old growth state of primeval forests or grasslands. There are long-lived, large, sessile or sedentary plants, invertebrates and vertebrates with some of the latter migrating annually in stereotyped patterns. The dominant species of the benthos may be termed K-selected (see Box 1).

Unfortunately it happens that nearly all the cultural stresses imposed by humans act to impair and incapacitate such a benthic attractor. Near the surface in moderately deep waters, a pelagic attractor, mostly with r-selected species (see Box 1), may then self-organize because ample nutrients then occur in those waters. A burgeoning pelagic association may then act so as to further suppress the benthic attractor already harmed by unsustainable cultural practices. This is a case of positive feedback, which occurs

Box 1 The meanings of the r and K terms

The r and K symbols originated in a particular formula for the physical growth of an organism, but have taken on a metaphorical life of their own in discussions of species successions as related to phase shifts. In an early successional stage, as in a cultivated field that is reverting to nature or with the herbaceous vegetation in a temperate forest in spring, r species that are small, short-lived, fast growing, early maturing and tolerant of fluctuating environmental conditions generally thrive and reproduce first. These are gradually replaced in late spring by activation of K species that are large, long-lived, slow growing, late maturing and intolerant of fluctuating environmental conditions. Within phase changes, the switch from dominance by one type to dominance by another is generally not complete (*see* **r–K Strategies**, Volume 2).

commonly in an early stage of the self-organization of a phase of a living system. With further organization, negative feedbacks are generated within the evolving new phase to counter-balance the positive feedbacks, or even to exceed the positive feedbacks, as in some over-mature senile, or old growth manifestation of a phase in an ecosystem.

The terms *oligotrophic* (few food items) and *eutrophic* (many food items) are commonly used to denote aquatic systems in which benthic and pelagic attractors, respectively, are dominant. These may be taken to be code words, since much more than trophic status is involved in oligotrophic type and eutrophic type systems.

In Regier and Kay (1996) and Kay and Regier (1999), we have focused attention on the kind of phase shifts between states that are dominated by the benthic and the pelagic attractors in the three basins of Lake Erie. Other researchers are taking a comparable approach to shallow or *lentic* waters of coastal wetlands and deeper or *lotic* waters as in rift lakes.

With respect to the whole gamut of aquatic ecosystems, there may be more kinds of attractors, and related phase shifts. One such state, that of a *saprobien system*, is particularly objectionable to humans in warmer parts of the world. European researchers who studied the effects of sewage outfalls on rivers starting a century ago first described a saprobien system. Just downstream from an outfall, sewage decomposes aerobically and this process may quickly exhaust the oxygen in the water. With farther movement downstream, the aerobic organisms die and anaerobic organisms become dominant. Ecological production by anaerobes may generate foul smelling gases based on reduced forms of sulfur, carbon and nitrogen, as also is the case with poorly managed cesspools. Farther downstream, the demand for oxygen by the biota can be met by the slow diffusion of oxygen from the atmosphere into the water, and this marks the downstream boundary of the saprobien system.

The saprobien system bears some resemblance to the decomposer association that forms on the bottom of strongly eutrophic lake waters of moderate depth.

Excessive enrichment of waters with plant nutrients may cause more severe eutrophication as hypertrophy. In relatively quiescent waters there may then emerge a surface association with floating and decomposing mats of algae, fungus, slimes and bacteria together with blooms of toxic algae where the matting is not continuous.

Caddy (1993) has observed ecological phenomena in enclosed and semi-enclosed seas subjected to unsustainable development that are similar to the phase shifts that we and others have inferred for freshwaters.

FROM GREY–BROWN TO GREEN–BLUE TECHNOLOGY

In the past, hydrological engineers were expected to correct the destructive behavior of streams that were excessively turbulent because of abuse upon abuse to which these streams and their watersheds had been subjected. Though the ecological causes of the destructiveness of an abused stream were understood in the mid-19th century by George Perkins Marsh (*see* Marsh, George Perkins, Volume 3) and others, there was not then an adequate planning or regulatory capability (see following section) to correct enough of these abuses at their sources to permit the stream to recover its gentler features. So the engineers resorted to dealing with local effects where some crisis had erupted, often by creating an ugly new water course with concrete and steel and enclosing it with tall fencing because people who fell into raging waters in such a channel were likely to drown. This concrete and steel *grey- brown technology* exacerbated the turbulent adverse effects downstream, which then called for more concrete and steel with more muddy water, etc. This is an example of positive feedback within anti-ecological technology.

A phase shift in planning and management of such human-altered streams occurred when the governance processes switched so that a watershed ecosystem approach came to use nature's way; *green- blue technology*. This phase shift involves compatible phase changes in the disciplines of engineering, of planning and in the expectations of affected residents. Such a comprehensive phase shift might be resisted by a government administration committed to encouraging the production and sale of concrete and steel, say, and to providing remunerative contracts for conventional engineering and construction firms. Government patronage, with the help of grey–brown bureaucrats, has long funded ecosystemically abusive engineering works related to harbors and streams. The institutional connections between a government agency, an industry and the relevant profession or interactive set of professions that led to such

autocatalysis or positive feedback has been referred to as an *iron triangle* in the US.

With streams of the Great Lakes Basin, we appear to be in the throes of a phase shift toward an ecosystem/watershed/landscape approach, i.e., to green–blue technology.

A CRUCIAL CULTURAL PHASE SHIFT AT A GLOBAL LEVEL

As has become increasingly obvious in many United Nations (UN) conferences since 1970, a crucial policy initiative on food, environment, population and development anywhere in the world has been to facilitate the self-empowerment of women (*see* Ecofeminism, Volume 5). Eco-feminists can explain much better than we can the necessity for a cultural phase shift toward a status for women that they prefer and must have. This involves a systemic change in the status of men, or at least of some men. According to women leaders of this movement, such a phase shift should not be a zero-sum game, in that few women want to compete directly with men to the disadvantage of men. Instead the shift should move from a culturally impoverished state of male dominance to a richer and better state of cross gender partnership.

At the UN Conference on Population and Development in Cairo in 1994, for example, strong resistance to such a phase shift came from male clergy who propounded narrow, fundamentalistic versions of the Judeo–Christian–Muslim family of religions. Though they may raise important ethical issues, such clerics and their male followers may not be disinterested personally since they may lose status and power with such a shift. Many men who are not constrained by these fundamentalisms are making common cause with self-empowering women.

What are some of the implications of a gender-related phase shift for Great Lakes Basin ecosystems, say? The phase shift from the grey–brown to green–blue technology sketched above may well be due in part to the increased participation in recent decades by women in all aspects of research, planning and management related to those streams.

Twenty years ago one of us, Regier, supervised the thesis research on the fish of the Credit River (near Toronto) by Deborah Martin, now Deborah Martin-Downs (Martin, 1984). She had learned to identify all fish in all life stages so that it would seldom be necessary to kill and preserve specimens. So she had data but few specimens, and sometimes could not respond in the usual way to a query whether she really had seen a specimen of a particular species at a particular locale. Though there were raised eyebrows and some testy queries during her oral examination, nobody considered the absence of dead specimens as a serious weakness of her method, and all respected her ethics. Such ethics may require that high expertise in recognizing fish species in the field be demonstrated for skeptical colleagues.

In the Great Lakes Basin since the 1950s, some, and perhaps most of the key environmental victories have been won with leadership by women. Many of the women were unpaid volunteers or underpaid activists. Mostly it has been well paid men who then led with the implementation.

Men should participate in correcting the degradation now so obvious in ecosystems everywhere. Of course, there were always individual men who were not part of the degrading culture. Similarly there were women who were willingly part of it, and may occasionally have been leaders of the degrading progressive modernism (*see also* Ecofeminism, Volume 5).

INDIGENOUS PEOPLES AND CORRECTION OF AN EARLIER PHASE SHIFT

In the UN conferences of the 1970s (on food in 1970, on the environment in 1972 and on population in 1974), some attention was focused on the injustices directed toward indigenous peoples as a consequence of invasions of their lands and waters by modern progressives. This theme became much more pronounced in the 1990s UN conferences on the environment and development, population and women. Agenda 21, issued by the 1992 UN Conference on Environment and Development in Rio de Janeiro, has a strong section on indigenous peoples.

With respect to Great Lakes Basin waters, aboriginals' non-ceded rights are coming to be recognized again, very belatedly. Since many aboriginals, both women and men, are re-committing themselves and their people to nature stewardship, this process of re-empowerment should be welcomed.

As with empowerment of women, re-empowerment of indigenous peoples should enrich our impoverished culture. Tolerance of diverse cultures may be as desirable as biodiversity. The phase shift toward fair status for indigenous or native peoples is not yet assured in the Great Lakes Basin generally, or elsewhere in the world. Non-native guardians of these waters are committed to partnerships with the native guardians (*see also* Indigenous Knowledge, Peoples and Sustainable Practice, Volume 5; Religion and Environment Among North American First Nations, Volume 5).

FROM MODERNISM THROUGH POST-MODERNISM TO WHAT?

Some thirty years ago the Western World apparently passed beyond the centuries-long era of modern progress, in a massive cultural phase shift that is still underway. The management guru, Drucker (1989) has sketched this shift in

ways that many would consider brash, but after publishing widely acknowledged seminal works for over 50 years; some continuing brashness from Drucker may be tolerable. Drucker's ecological perspective on all of this emerged out of the new discipline of business administration, and is similar to what we refer to here as *ecogenic*.

One of us, Regier (1995), has listed some 15 transitional phenomena that occurred in or about 1968, at about the time inferred by Drucker to be the height of land between the conceptual basin of the modern era and the conceptual basin of the new era. The latter era is still apparently nameless; though some people refer to it as *post-industrial* or *post-modern*.

Stronger commitments to *deontological* ethics (ethics based on a deep shared sense of what is right and what is wrong with respect to duty) may increasingly be trumping modernist commitments to utilitarian ethics; ethics based on the greatest good for the greatest number. Integrity may be a code word for such a deontological commitment. In the 1960s social integration, toward a new and desirable cultural integrity, was directed toward correcting racial and gender inequities in Western countries. Since then, integration of many indigenous peoples into emerging political systems has been leading toward a mosaic landscape integrity that is apparent politically, culturally and ecologically, from an ecogenic perspective.

Agenda 21, that outcome of the 1992 UN Conference on Environment and Development in Rio, includes deontological as well as utilitarian commitments to global programs of several kinds, e.g., to preserve biodiversity. Thus a deontological commitment to biodiversity concerning Great Lakes Basin waters links to a compatible global commitment. Such connections should be recognized and be supported by strong institutional links, as has been attempted at the federal Canadian level. Local stream stewards, say, could draw encouragement from the sense that they are necessary parts of a global ethical commitment to ecological integrity.

CONCLUSIONS

Nine kinds of systemic phase shifts have been sketched above, all in a conceptual context of dynamic self-organizing ecogenic systems. We hope that this evidence suffices to make the point that such phase shifts are now ubiquitous in all the five versions of eco-studies to which we have referred here. If so, our educational, research, practice and governance initiatives should more clearly reflect that awareness.

Practically, a precautionary principle invoked in the context of a commitment to sustainable and responsible use implies that informed judgement be exercised to avert phase shifts to undesirable states. Once a phase shift has occurred, efforts to reverse the shift may be very costly and take much time because of the effects of systemic inertia or hysteresis.

There is, as yet, no strong empirical evidence to support optimism about the 21st Century. The new era, following the cultural phase shift sketched above, may play out worse than the progressive modern era that is now behind us. For example, it may turn out that our environmental and cultural reforms will have been too little and too late; and we humans may trigger flips into a succession of less desirable states of kinds that we have not yet encountered. Ecogenically there may then be no way back to something resembling an earlier manifestation of our biospheric home.

ACKNOWLEDGMENTS

An earlier version of this essay appeared in: *Stream Corridors: Adaptive Management and Design*, Proceedings of the Second International Conference on Natural Channel Systems, March 1–4, 1999, Niagara Falls, ON. Credit Valley Conservation Authority, Georgetown, ON. Patti Young, coordinator of that 1999 CD publication, has granted permission to publish the present revised version.

REFERENCES

Caddy, J F (1993) Towards a Comparative Evaluation of Human Impacts on Fishery Ecosystems of Enclosed and Semi-enclosed Seas, *Rev. Fish. Sci.*, 1(1), 57–95.

Drucker, P F (1989) *The New Realities*, Harper and Row, New York.

Fuller, S (2000) *Thomas Kuhn: a Philosophical History for Our Times*, University of Chicago Press, Chicago, IL.

Haas, P M (1999) Social Constructivism and the Evolution of Multilateral Environmental Governance, in *Globalization and Governance*, eds A Prakash and J A Hart, Routledge, London.

Kay, J J and Regier, H A (1999) An Ecosystemic Two-phase Attractor Approach to Lake Erie's Ecology, in *State of Lake Erie (SOLE) - Past, Present and Future*, eds M Munawar, T Edsall, and I F Munawar, Ecovision World Monograph Series, Backhuys Publications, Leiden.

Koestler, A and Smithies, J R (1969) *Beyond Reductionism*, Hutchinson, London.

Martin, D K (1984) The Fishes of the Credit River: Cultural Effects in Recent Decades, M.Sc. Thesis, Department of Zoology, University of Toronto.

Rapoport, A (1986) *General System Theory: Essential Concepts and Applications*, Abacus Books, Cambridge, MA.

Regier, H A (1995) Ecosystem Integrity in a Context of Ecostudies as Related to the Great Lakes Region, in *Perspectives on Ecological Integrity*, eds L Westra and J Lemons, Kluwer Academic Publications, Dordrecht.

Regier, H A and Kay, J J (1996) An Heuristic Model of Transformations of the Aquatic Ecosystems of the Great Lake – St. Lawrence River Basin, *J. Aquat. Ecosyst. Health*, 5, 3–21.

FURTHER READING

Caley, M T and Sawada, D (1994) *Mindscapes: the Epistemology of Magoroh Maruyama*, Gordon and Breach, London.

Kay, J J (1984) *Self-Organization in Living Systems*, Ph.D. Thesis, Systems Design Engineering, University of Waterloo, Waterloo.

Rapport, D J and Regier, H A (1995) Disturbance and Stress Effects on Ecological Systems, in *Complex Ecology: The Part-Whole Relations in Ecosystems*, eds B C Patten and S E Jorgensen, Prentice Hall, Englewood Cliffs, NJ.

Regier, H A, Jones, M L, Addis, J, and Donahue, M (1999) Great Lakes – St. Lawrence River Basin Assessments, in *Bioregional Assessments: Science at the Crossroads of Management and Policy*, eds K N Johnson, F Swanson, M Herring, and S Greene, Island Press, Washington, DC.

Regier, H A, Welcomme, R L, Steedman, R J, and Henderson, H F (1989) Rehabilitation of Degraded River Ecosystems, *Can. Special Publication Fish Aquat. Sci.*, 106, 86–97.

Steedman, R J and Regier, H A (1987) Ecosystem Science for the Great Lakes: Perspectives on Degradative Transformations, *Can. J. Fish Aquat. Sci.*, 44(Supplement 2), 95–130.

Philosophy, Environmental

see Environmental Philosophy: Phenomenological Ecology (**Volume 5**)

Pinchot, Gifford

see Environmental Movement – the Rise of Non-government Organizations (NGOs) (**Volume 5**)

Political Movements/ Ideologies and the Environment

David Pepper
Oxford Brookes University, Oxford, UK

There is no one, politically uniform, environmental movement in the West, but a series of movements and groups. They represent a massive growth in popular concern about the negative ecological consequences of economic and technological development since the 1960s. Initial worries, like those over the impact of agricultural chemicals on the environment, soon broadened to include all pollution caused by economic and technological development, and the issue of population growth, consumerism and resource use. This developed into a challenge to the whole notion of 'progress' implied in Western development since the Enlightenment, and to industrial society and its view of social justice.

After some early impact on policy in the 1970s, there was a period of retrenchment from environmental legislation in the 1980s by Western governments. They were in uenced by the arguments of business and industry, which in a climate of increasing global competitiveness saw environmental and social spending as brakes on ef ciency and pro ts. While this climate still prevails, the of cial stance of governments since the 1992 Rio Conference has been that the interests of pro table business and environmental protection are compatible. Hence environmental sustainability principles are supposed to be incorporated into all main policy areas under the development model known as ecological modernization.

Environmental movements have reacted to these trends by developing political parties, pressure groups and other approaches. These have sought to in uence government from the outside via public opinion, and also to become part of policy formulation from the inside. In the 1990s there has also been a revival of direct action, which sometimes breaks the law, by environmentalist factions who are disillusioned with the democratic processes.

Two approaches to environmental politics have developed. Reformism seeks to change society substantially to achieve environmentally sustainable development, but not to the extent of abandoning capitalism and the modern scienti c, technological and managerial expertise on which mainstream development is based. Radical environmentalism, however, does ultimately wish to radically change or remove capitalism, replacing it with forms of localized development that prioritize social need, direct grass-roots democracy, and environmental protection rather than economic and technological growth.

Mainstream environmentalist political ideology, therefore, re ects these currents, drawing on radical and reformist political traditions going back at least as far as the 18th century. Reformism re ects conservative, liberal and democratic socialist traditions. Environmental radicalism is especially in uenced by anarchist and revolutionary socialist traditions, mingled in with some conservative thinking. The result is a very mixed and sometimes incoherent ideology, whose underlying current re ects either an 'ecocentric' approach which prioritizes the needs of natural ecosystems, or a human-centered (anthropocentric) approach which prioritizes human social, economic and cultural considerations. The fact that mainstream environmentalism

re ects all these in uences and cross-currents could be seen strategically as both a strength and a weakness.

THE ENVIRONMENTAL MOVEMENT'S ORIGINS AND CONCERNS

Everyday political discourse now frequently refers to *the* environmental movement, a late twentieth-century manifestation of widespread concern in the West about the negative ecological consequences of economic and technological development. Strictly speaking such references are only partly accurate. First, there is no one, coherent environmental movement, as much of the rest of this entry will con rm. Rather, there is a whole complex of movements and ideologies, within and outside conventional politics, expressing environmental concern and proposing social and political solutions for it.

Secondly, environmental movements did not originate in the late 20th century; they are rooted in earlier movements. In America, the 19th century wilderness preservation movement, inspired by John Muir, Henry David Thoreau and Ralph Waldo Emerson, gave rise to the Sierra Club, from which in 1969 a group of more radical conservationists splintered to produce Friends of the Earth. In Europe, the National Trust, one of the UK s major contemporary environmental preservation groups, was founded in 1895. The founding group was inspired by challenging, questioning attitudes towards industrialism and progress posed by people such as John Ruskin and William Morris: it also drew on feelings for natural landscape inspired by Romantic gures such as William Wordsworth. Other direct in uences on today s environmentalism include the development of the science of ecology, whose leading gures included 18th and 19th century naturalists such as Gilbert White, Charles Lyell, Charles Darwin and Ernst Haeckel. The work done on population by Thomas Malthus is also directly re ected in the modern movement, as is the in uence of 19th and early 20th century anarchists and utopian socialists (*see* **Political Systems and the Environment: Utopianism**, Volume 5).

Furthermore, in between the World Wars there were vigorous movements in the West showing concern for nature and for healthy lifestyles, including vegetarianism. These were intertwined with both left and right-wing (including fascist) political movements, among both the masses and intellectuals (for a description of the roots of the environmental movement, see Pepper, 1996).

However, it is true to say that environmental movements, which started to emerge in the 1960s, have shown greater scope, range and connectivity than any previous environmental movement. This was inevitable, given: (a) the global advance of modernizing societies and economies: (b) the development of instantaneous and ubiquitous global communications networks: and (c) the unprecedented power of post-war technologies to create alarming environmental changes. As sociologists Ulrich Beck and Anthony Giddens emphasize, today s society is exposed to unprecedented risks and uncertainties, whose origins it is often dif cult to unravel, determine and control (Beck, 1992, 1995). These include environmental risks of (particularly Western) society s own making.

The rst of these to trigger widespread alarm, marking the dawn of modern environmentalism, was that caused by the chemical substances used in agriculture. They were identi ed in Rachel Carson s iconoclastic work, *Silent Spring* (1962) (*see* **Carson, Rachel Louise**, Volume 5). Before this book, American and European conventional wisdom held that chemical pesticides and herbicides were an unalloyed blessing: helping developed countries under capitalism to repel Malthusian spectres of famine and want, and build that steadily-advancing material prosperity on which the American Dream and post-war European recovery were based. The questioning, sparked by Carson s book, of this conventional wisdom, and the subsequent wave of popular skepticism about scienti c advances, technological experts and the motives of big corporations, fed into what is now described as a condition of *post-modernity* , which some see as the prevailing mode in Western thought.

In it, all those things enshrined in the notion of progress adopted since the Enlightenment have been questioned and often denounced. Of course, there always was a critical, counter-cultural minority current in the West accompanying industrial and democratic revolutions, and its 1960s expressions in the hippie, civil rights, anti-nuclear and feminist movements fed directly into the environmental movement. Many sociologists see all these together as a coherent wave of new social movements forming the basis of post-modern politics. New social movements are inspired by single issues, but they also express the unease, disillusionment and aspirations of the newly emerged (professional, especially) middle-classes. Some commentators think that as such, they have supplanted the old basis of political discourse, in worker-capitalist relations and struggles over distributing material wealth.

The concern with pesticides soon widened to embrace all environmental pollution caused by modern technologies, including nuclear power and weapons, and then in the 1980s and 1990s, the internal combustion engine, gene modi cation, factory farming and the like. In the late 1960s and 1970s, too, there was growing questioning of mass production and mass consumption, combined with a wave of near-hysteria about human population growth, particularly growth in the third world. Thus emerged the neo-Malthusian limits to growth school (Meadows *et al.*, 1972, 1992; Hardin, 1968, 1974; Ehrlich, 1969; Ehrlich and Ehrlich, 1990). This contradicted the widely held views that materially based economic growth could continue *ad infinitum*, and that levels of af uence enjoyed in the West could be attained by all the world s population (which

appeared to be increasing exponentially). Limits to such growth were imposed by earth's carrying capacity – by constraints such as the inability of eco-systems to provide enough resources to feed growth or to soak up the pollution associated with it.

Based on such arguments, early environmental movements successfully prevailed on American and European governments to legislate tighter environmental standards in the 1970s, and to set up agencies and procedures designed to protect the environment. In America, particularly, groups mounted successful legal challenges to development likely to endanger animal and plant species survival, grounding them in an eco-centric' philosophy which said that nature has intrinsic value: value beyond its usefulness to humans. This was all done in a prevailing political-economic climate where the legitimacy of state spending and market intervention to mitigate the worst social effects of the free market was upheld by a significant majority of politicians. Hence, if business and industry now had to spend some potential profits on procedures, processes and technologies to protect the environment, this was regarded as proper and necessary.

However in the 1980s and 1990s, in the face of intensifying global competition and pressures on profit margins, the philosophy of economic *laissez-faire* took on new political vitality. Neo-liberal governments in North America, Australia and Europe began anew to prioritize economic growth and to side with industry in regarding regulations and tax-funded public spending on high levels of social and environmental protection as undesirable business overhead costs which must be reduced. Consequently a retrenchment phase began, with relaxation of regulation, and downward pressures on environmental protection spending. Although officially such attitudes have now changed, their momentum persists in anti-environmentalism, where, for instance, some business corporations form unofficial coalitions (e.g., the European Round Table or the Global Climate Coalition) specifically to lobby against pressures from environmental lobbies. They challenge, for instance, the validity of scientific research which suggests that human-induced global warming is significantly increasing.

At the same time, the 1990s saw a new phase where environmentalist concerns seem to have been taken on board by mainstream politics. Nearly all of the world's governments were officially represented at the 1992 United Nations Conference on Environment and Development; the Rio Conference'. And very many of them signed that conference's declarations, amounting to broad and unspecified commitment to the notion that environmental protection must be integrated into the mission of universalizing material prosperity – a mission still to be achieved, however, through global marketization and the search for economic growth. Hence the idea of materially-led universal development was not abandoned, but the caveat was added that it must be sustainable, meaning that the resources needed for it should not be destroyed, and the options open to future generations to live at least as well as the present one should not be foreclosed. The development model, which is supposed to enable this is that of *ecological modernization* (*see* Political Systems and the Environment: Utopianism, Volume 5). During this phase of environmental concern, sustainability objectives have been written into mainstream policy programs of the leading industrial nations, intending that they should be integrated within economic and social objectives. This clearly means that the limits to growth school's major premise, of incompatibility between growth and environmental protection, has been rejected in favor of its obverse – the idea that the two are necessary to each other.

As these phases of environmental concern have progressed, the emphases in the mainstream of environmental movements have altered. The initial focus was on single issues, but quickly this developed into the skeptical questioning, mentioned above, of industrialism and modernity. This put environmental pressure groups generally at odds with the establishment and the political mainstream. Through the 1980s, however, they began to acquire more influence, augmented by electoral success in some European countries. During and after the Rio Conference, some groups (e.g., Friends of the Earth) gained semi-official status as non-government organizations (NGO). As such, they make formal presentations to the UN Commission for Sustainable Development, and in some cases are incorporated onto the panels of local Agenda 21 organizations set up in towns and cities in the wake of Rio.

At the same time as environmental groups have developed their expertise and mainstream influence, their underlying analyses of causes and effects of environmental problems have often become more sophisticated. They are more likely to highlight problems like the internal growth dynamic of capitalism and its inherent tendency to take only partial and short-term perspectives on problems, the apparently unavoidably uneven nature of development as global modernization, the inextricable connection between environmental problems and social inequity and injustice and the way in which Western affluent lifestyles leave substantial ecological footprints in third-world countries (e.g., toxic waste produced in the West being shipped to the third world for dumping or processing). The overall political direction of such a critique may point towards radical social change, beyond that envisaged in even strong' approaches to ecological modernization. This has uncomfortable implications for Western mainstream politics.

Hence, late-1990s environmental movements followed two paths. Some groups, or factions within groups, persisted along the channels of conventional political action, such as running for office, lobbying and advising government and other bodies. Other groups and factions began a new wave of direct action (both peaceful and violent) and civil disobedience, intending to block developments

such as the growth of road traffic into residential areas, the building of new roads, nuclear waste dumping, trials of genetically-modified crops, or foxhunting and other animal abuse. In some of these protests, notably over major road development at Twyford Down and Newbury in the UK, such eco-anarchist protestors have been supported by many middle-class citizens living conventional lifestyles. The former have been concerned about general environmental and social principles, while the latter have been representing their own limited, local interests (the NIMBY – not-in-my-back-yard syndrome). What unites these unlikely bedfellows is (a) shared concern about quality-of-life issues, including environmental quality (b) deepening conviction that commitment to profitable, efficient, competitive business should not be prioritized above all other social-economic policy considerations (c) increasing skepticism and frustration with Western democracy, on the grounds that it is more responsive to the needs of international business lobbies than to those of the ordinary citizen. This last concern is responded to in some areas of EU policy, for instance rural development, where resources are being put into developing bottom up', community-led initiatives. It is also reflected in the growing number of academic texts concerned with environmental democracy (e.g., Mason, 1999).

TYPES OF ENVIRONMENTAL MOVEMENT

Environmental movements includes green parties, pressure groups within mainstream political parties, non-party pressure groups, direct action groups, local community groups, communes and individuals exercising lifestyle choices. Connelly and Smith (1999) stress how the ethical underpinnings of these groups are extremely diverse and difficult to reconcile as one movement. They identify New Zealand as the location of the first green party, founded in 1972. The second was started in the UK in 1973, but the most prominent in electoral terms has been *Die Grünen*. This began from community groups in West Germany in 1980 and achieved government in coalition with social democrats in 1998. Electoral success appears to depend very much on the voting system adopted, hence UK Greens had 15% of the vote in the 1989 European Parliament elections but took no seats on the basis of a first-past-the-post system. In 1999 however, when proportional representation was adopted, the Greens took two seats from a much lower percentage vote.

However, the constant problems, which beset mainstream politicians of principle, have also dogged Green parties. In attaining and widening electoral support, or in actual government, principles have usually had to be compromised and this has opened internal splits between pragmatists, or realists', and fundamentalists. Such splits have been very public in Germany, leading to the resignation of leading Grunen members and the vilification of many who stayed, but they have been mirrored elsewhere. Ongoing internal debates and squabbles have occurred on such issues as the importance of obtaining office as opposed to the virtue of remaining a principled opposition, organization versus spontaneity, and the desirability or otherwise of playing the mainstream game involving party hierarchies, charismatic leaders, professionalism, media manipulation and public image. The UK Green Party suffered considerably in the early 1990s from such conflicts, and has not completely recovered. Meanwhile, in less than two years of government Die Grunen have compromised principles to the extent that internal critics fear that the most that has been achieved has been to help create a more viable capitalism; i.e., one which has been alerted to the dangers to itself arising from environmental problems.

These kinds of tensions mirror the dichotomy discussed below, between radical and reformist approaches to, and influences on, the environmental movement. Such tensions occur within groups and also within individuals. They result in what Dobson (1995) has identified as public and more private faces of radical environmentalism, where the desire for fundamental social change and the overturning of the economic order become deliberately muted themes in public debate for fear of frightening off ordinary members of the public.

Despite the limited electoral impact of Green parties, what seems an inexorable rise in public concern about environmental issues is illustrated through data cited by Connelly and Smith (1999) on eight UK environmental pressure groups. Their total membership rose from 1.5 m in 1980 to 4.3 m in 1995. Some of them, like the National Trust and Royal Society for the Protection of Birds, are long established and indisputably conservative and safe from the point of view of conventional political parties. Nonetheless, their growing strength has contributed to the greening of these parties. In the UK, Conservative, Liberal Democrat and Labour parties all have their own internal environmental pressure groups, and the amount of manifesto space devoted to environmental policy has steadily increased since the mid-1980s. But British politicians have been slow and recalcitrant in many environmental policy areas by comparison with politicians in Scandinavia, Germany and the Netherlands.

These latter countries have been at the forefront of projecting environmental policy within the EU, and as a result of some success here, Britain, once known as the dirty man of Europe', has been drawn into a more enlightened approach. This raises an interesting geopolitical dilemma for the British environmental movement, since both radical and reformist wings' tend to champion the idea of government decentralization and more inclusive local democracies. Large-scale economic, social or political structures are resisted in favor of localism and decision taking at the lowest possible level. But when this last principle, of

subsidiarity', has been applied by the European Union (UN) in relation to environmental matters it has sometimes enabled British governments to claim powers over their own development decisions, which have been unpopular with environmentalists. The latter have been driven to look to the supranational bureaucracy of the EU for protection from their own government.

Pressure groups do not seek electoral office, but, rather, hope to influence public policy. They act in support of a cause, rather than the narrower sectional interests of their members. Indeed there is an ongoing debate within Greenpeace about whether to have members at all, rather than just supporters. Pressure groups seek influence from inside government, through petitioning, lobbying and eventually being accepted as consultants during the stage of policy formulation. As a result, they attempt to make themselves credible by acquiring a reputation for painstaking scientific research of integrity, and for reasonableness', professionalism' and a willingness to dialogue and compromise with opposition groups. One such example in the UK is Transport 2000, a coalition of environmental groups concerned about transport issues, which have won government and public over to its view that private car use must be curbed in favor of public transport investment.

Pressure groups also seek to apply outside pressure on governments via public opinion. This requires a higher-profile approach, staging media events, demonstrations, petitions and consumer boycotts. Flamboyant media-grabbing tactics are often at the edges of the law, or involve breaking it before backing down in the glare of publicity, to disrupt, for instance, whale fishing, nuclear tests or public enquiries into development projects.

Hence pressure groups have to achieve a balance between two approaches, and this may be a source of ongoing internal disputes. Whether this balance swings towards or away from more conventional, respectable' and patient methods may well depend on the perception, which the group has about the nature of the democracy within which it works. If the belief is that this follows a pluralist model, in which many different interest groups have more or less equal opportunities to influence policy direction, then the group may concentrate on methods appropriate to this. If, however, the view is towards an elitist model, where some interests are seen as unfairly dominating the political process by virtue of their economic and political strength – often exercised in corridors of power – then less legal methods may be followed.

Direct action groups are of the latter persuasion. They include the likes of American Earth First! (Wall, 1999), with its monkeywrench' tactics of civil disobedience and ecotage' committed on the physical infrastructure of big development projects, such as dams. Then there are tree dwellers and tunnel diggers who have gained huge publicity in anti-road and anti-airport protest in Europe, North America and Japan. There are squatters, like those in The Land is Ours' group, who invaded derelict land owned by the Guinness Corporation and scheduled for major development in Wandsworth, London – they set up a model sustainable village for five months in 1996 before being evicted. And there are arsonists who set fire to agricultural developments and scientific establishments in the name of animal liberation. When challenged about why they do not use legal channels of protest and persuasion, they will reply that these channels are merely token, giving a veneer of democracy' to what is in fact a corporate state run by a coalition of bureaucrats, business leaders and, perhaps, union bosses.

Partly in response to this kind of critique, and partly as a result of budgetary cuts in local government spending, there has been a growth in the 1990s of local voluntary community-based groups. These seek, legally and often supported by government grants, to improve the economic prospects and social and environmental quality of life of their local community. They include such environmentally oriented groups as housing cooperatives, conservation and recycling projects, urban agriculture, box schemes' (where the produce of nearby organic farms is sold direct to households) and local employment and trading schemes (LETS, where goods and services are exchanged by means of local currencies). A skeptical view of such developments is that they help to exonerate the state-supported public sector from its duties to provide social, health and environmental care for all, which is why states generally have a favorable view of them. A more positive perspective is that they represent a process whereby people take their lives and destinies into their own hands, therefore reclaiming some political power.

This constitutes a major theme voiced with increasing urgency by environmentalists since the 1960s. They argue that if people regain control over their lives, and are re-embedded' into their communities and localities, it is axiomatic that they will behave in more environmentally sustainable ways. Through living more rounded and satisfying lifestyles, it is reckoned, people will not want to seek fulfillment in an endless round of material consumption. This is the basis on which ecologically based communes operate. The heyday of such alternative' communities was in the 1970s, when (especially) rural property and land were cheap to acquire. Back-to-the-land groups often resolved to achieve a large measure of self-sufficiency, in food if not energy, and to do it in ecologically benign ways. Although they usually failed in this objective for any sustained period, and although they encountered many problems in trying to create alternatives to the nuclear family and non-oppressive social relations, many of these communities have lasted into the 1990s. However, they have not remained immune to social trends in the mainstream society to which they form a counterculture, but have fallen prey to increasing materialism, privatization and

preoccupation with maintaining financial viability (Pepper, 1991). Notwithstanding this, they have contended with a set of problems which they prefer to the problems and insecurities of living in conventional society.

A further manifestation of environmentalism has been in the application of its personal-is-political' motif to consumerism itself. The green consumer movement peaked in the 1980s, but survived into the 1990s particularly as the ethical consumer and investor movements (*see* Green Investment, Volume 4). It is now part of the ecological modernization approach to development, and is based on the idea, championed by neo-liberal political economists like von Hayek in the 1950s, that in spending our personal money we are casting votes. These votes indicate desire for and satisfaction with the products we buy, and the processes and organizations through which they are produced. Hence, it is argued, if we buy ecologically sound' products and services, with environmentally sound ingredients produced in environmentally friendly ways, and if we shun products which do not meet this description, producers will have to change their ways accordingly or go out of business. Through this kind of consumer movement, chlorofluorocarbons (CFCs) have been eliminated from aerosols and refrigerators, real animal furs have been largely replaced by artificial ones, and, possibly, the development of foods containing genetically modified organisms has been arrested. Additionally there has been a big boost for organic food producers, tea and coffee producing cooperatives, and so forth.

However, the green consumer movement has not produced the widespread changes which were once claimed for it, and it has not defeated the skepticism of radical environmentalists. For one thing, its success depends on the presumptions that consumers have access to reliable and honest information about the products they buy, and are prepared to spend time and money acquiring and acting on such information. None of this has been proved accurate. Furthermore, the idea that money is a vote also carries the corollary that a large part of the potential electorate' is disbarred from voting by virtue of poverty, especially in times of recession. And in any case, the proposition that consumption of material goods is a remedy for environmental problems rather than being their major cause is still impossible to swallow for radical environmentalists.

POLITICAL IDEOLOGIES AND ENVIRONMENTALISM

Figure 1 shows different forms of environmentalism when considered particularly from the standpoint of political ideologies. It differentiates principally between radical and reformist approaches. The latter embrace mainstream culture's political ideologies of liberalism and democratic' (as practiced by labor and social democrat parties) socialism. It would reform capitalism and market economics to a greater or lesser degree, but would not replace them by another economic system. One of the main arguments within it is about *how much* to intervene in the market economy (see below). It reacts to environmental problems from a perspective which some call techno-centric'. Techno-centrism manifests faith in science, technology and rational management of ecosystems to solve environmental problems. Elements of conservative thinking also may mingle with this reformist approach of the cultural mainstream. Ecological modernization is the sustainable development model, which has evolved from reformist environmentalism.

By contrast, radical environmentalist approaches seek to eliminate environmental problems at their root, rather than simply reacting to the damage caused by the normal operations of global capitalism. This would entail fundamental social change, either by eliminating or completely reconstructing capitalism. Hence the environmental debate is shifted out of the cultural/economic mainstream and becomes counter-cultural: often drawing on older counter-cultural traditions such as anarchism or utopian socialism. It includes:

- *social ecology*, based largely on anarchist principles as interpreted particularly in the work of Murray Bookchin (e.g., 1990), which is also informed by Marxist analysis (*see* Social Ecology, Volume 5).
- *eco-socialism*, based on revolutionary rather than reformist (e.g., social democrat) socialism, and also informed by Marxist analysis. Libertarian, decentralist and communalist in principle, it is ultimately opposed to the state (*see* Eco-socialism, Volume 5).
- *deep ecology*, which focuses on fundamental changes in attitudes and values towards nature starting with the individual (Devall and Sessions, 1985). Its social prescriptions are anarchistic but also informed by elements of conservative thinking (*see* Deep Ecology, Volume 5).
- *ecofeminism*, which insists on the elimination of patriarchy as a prerequisite to founding a sustainable society. It is also informed by anarchist and socialist analyses of material processes in society, but emphasizes as well the importance of changes at the cultural level (*see* Ecofeminism, Volume 5).

Deep ecology's approach is eco- or bio-centric: that is, focused on non-human nature and the whole biosphere. Eco-centrism claims that nature has intrinsic value above and beyond the value merely conferred by humans. Eco-socialism, by contrast, shares an anthropocentric stance with reformist environmentalism. This regards humans as the ultimate source of all value and is prepared, if it comes to a crunch, to elevate human interest above that of animals and plants. Social ecology claims to transcend both anthropocentrism and biocentrism.

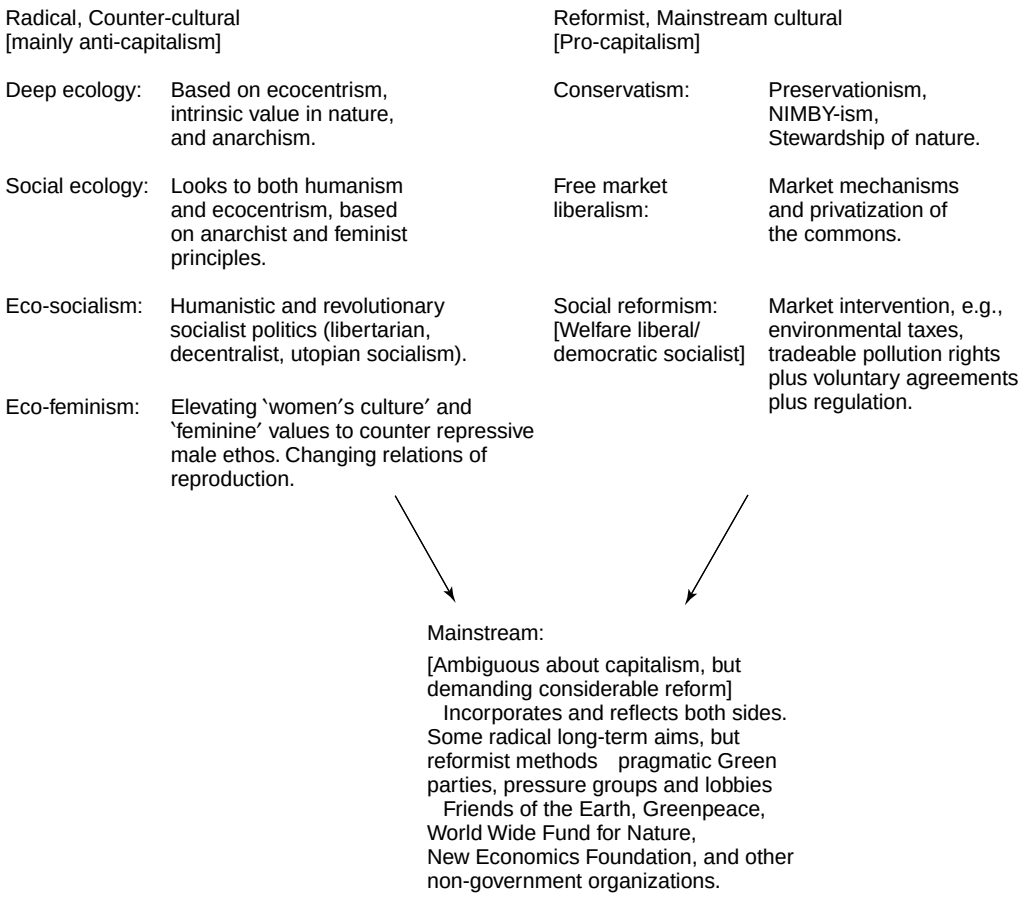


Figure 1 Environmentalisms and their political ideologies and approaches

Mainstream environmentalism, as represented by the movements and groups discussed above, tends to be a hybrid – a *mélange* of both radical and reformist approaches. As also suggested, the mainstream environmentalist movement's proponents and actors often hold radical views (including eco-centrism and opposition to capitalism), but they may show pragmatism, believing that to make some environmental headway by reform is better than to make none at all. The works of some of mainstream environmentalism's most influential figures (e.g., Schumacher, 1973; Porritt, 1984; McKibben, 1990) contain currents of socialism, anarchism, conservatism and social reformism.

The following discussion will illustrate how elements of traditional political ideologies inform both radical and reformist wings of the environmental movement, before going on to consider what aspects of green political ideologies might be considered as unique.

Before commencing it should be stressed that assessing green thinking against more traditional political ideologies is fraught with difficulties. Political philosophies are mainly diffuse rather than tightly coherent, changing and evolving with time. They also contain inconsistencies, they overlap, the terminology is confusing, and there may be considerable differences between what is popularly thought of as, say, liberal or socialist and what in theory they are. Popular views of political ideologies are colored by the utterances of party politicians, but there are few correspondences between political parties and the political philosophies they might claim to represent. And because party politics is political' there may be deliberate or unintentional attempts in debate to misrepresent them: because they are defined and evaluated subjectively, different people will make different interpretations of what they are. Additionally, political ideologies tend to differ in their interpretation between different countries. The view of them given below is written from a Euro-centric perspective, based on the British experience.

Finally, it should be acknowledged that it is one thing to note similarities between, say, anarchist and green ideologies, but this does not necessarily signify historical links between them – in the way of common threads of thinkers, writers or activists. Anarchy, like other political philosophies, comes specifically from political activism and social theories. By contrast, the green movement over the past

half-century originated from the concerns of *scientists* (like the Ehrlichs, Hardin, Fraser-Darling or Capra) who began to make political statements about society stemming from the results of their scientific work. They were not founded in any way on what the likes of Edmund Burke, Karl Marx or Peter Kropotkin wrote about – such figures were invoked afterwards. Hence the early statements from environmental movements were often politically naive and ideologically incoherent.

TRADITIONAL CONSERVATISM

The words conservation' and conservatism' both have the same root. And the ideas of tradition, continuity, stability, and dislike of sudden change but an acceptance of slow organic change (i.e., which is not planned or blueprinted or revolutionary) all feature in conservative thinking and all are compatible with at least some environmentalist thought. In conservatism the analogy between society and nature is strong: just as ecosystems need to be changing organically, not precipitously, so does society. As with nature, variety and structure in society are essential to achieve the all important goal of stability. Social revolutions (and new technology) upset the natural' social order. Conservatism perceives this order to be hierarchical, though this does not mean that respect is not due to those lowest in the chain of being. Hence relationships between higher and lower orders might be oppressive and economically exploitative but they nevertheless involve mutual obligations: they are not just one-way. However, those lower in the social order should accept the naturalness and inevitability of their position.

Such beliefs foster a romantic view of traditional' societies. So also is there a liking for traditional' pastoral landscapes and grand architecture, expressed in reformist environmentalism in Britain in the work of long-established nature and conservation groups like the National Trust, Councils for the Protection of Rural England and Wales, County Landowners Association and the Civic Trust. Conservative environmentalism involves a conception of stewardship' on the part of landowners – holding the land in trust for a future generation (again there is the idea that social obligations attach to power). This draws on Edmund Burke's injunctions about wise stewardship, and proposes that long-cherished virtues of efficiency, order, thrift, self-help, tradition, patriotism and nationalism should be the basis of environmental politics.

Faced with the dilemma posed by Garrett Hardin's (1968) commons' parable, in which individuals tend to use the resources of Earth's commons profligately because the costs of so doing accrue to society as a whole rather than to them specifically, conservatives would usually argue for enlightened private ownership of the commons as the best way to value and conserve them. They might also accept Hardin's (1974) arguments for coercion to curb population growth, and to promote the social and environmental consciousness appropriate for population control.

Edward Goldsmith (see, for instance Mander and Goldsmith, 1996) is a prominent radical environmentalist who seems to argue for conservative values. He wants to see Western societies re-embedded in strongly held belief systems, such as those enshrined in religion, for he thinks that strong religions are stabilizing forces making for social unity. Goldsmith has held up the caste system in India as the kind of social organization, which is compatible with an ecologically and socially sound society – sound because it is stable, and in balance with the natural environment. The way of life, like the structure and mechanisms of an ecosystem, is designed to maintain order. A sustainable society's common values must, above all, involve respect for ecosystems – this respect must be absolute and not negotiable. In common with many environmentalists (such as the more liberal Fritz Schumacher), Goldsmith argues for small-scale organization as the geographical basis for achieving the other desirable things. His radical environmental conservatism starts from the assumption of the family as the essential unit of social organization: what preserves this (such as the traditional stereotyped role for women) is to be encouraged. And it rejects industrial society' as aberrant. The desire to recapture small-scale traditional' society leads Goldsmith constantly to idealize primitive' peoples and tribes in Africa, Australasia, etc. Deep ecologists seem particularly inclined towards such idealization of pre-modern societies. They often hold up the North American Indian', for instance, as a paragon of ecological virtue, citing religious beliefs which apparently regard nature as part of the tribal community and require it to be treated with much respect (see for instance Callenbach, 1981). Archaeological evidence, however, often suggests something rather different; that notwithstanding beliefs, native American' actions towards nature may have been frequently exploitative (*see* Religion and Environment Among North American First Nations, Volume 5).

Economically, contemporary conservatism partly embraces neo-liberalism, as expressed in the doctrines of Adam Smith, since this revives the free market tradition which marked the beginnings of industrial capitalism. This seems particularly true in North America, for although many Thatcher conservatives in Britain did follow Smith, other more traditional' conservatives expressed misgivings at their lack of wider social responsibility to those low down in the socio-economic hierarchy.

Traditional conservatism may grade into fascism, also known as right-wing irrationalism', or extreme' romanticism. Some consider that Hitler spoke about the homeland almost in an ecological sense, using biological analogies of race, soil, homeland, folk and blood to describe the state as a living organism in developing his imperialistic theories. Bramwell (1989) describes how European Nazis showed

a liking for vegetarianism, the back-to-the-land movement and bio-dynamic farming. She observes that neo-fascist organizations in Europe have all taken onto their platform a green' perspective. Fascism's key elements include the use of biological metaphors, the stress on the organic community and the individual's need to merge with it, the elevation of ritual, intuition, and the mystical, and the distrust of the rational. These elements may be found in some ecological writings, especially those promoting deep ecology activism.

However it is unlikely that environmental movements as a whole would become fascistic because of their fundamental globalist inclusivism. That said, there have in the past been calls for a marrying of some of the elements identified above with a centralized authoritarian state (see Heilbroner, 1975; Ophuls, 1977), though without the crusading rejection of rationality necessary for fascism to exist. Heilbroner and Ophuls reached their authoritarian prescriptions reluctantly: but they were very pessimistic about the global condition, and thought that the nature of the required remedial action was so radical and electorally unpopular that no other institutional arrangement could possibly deliver.

FREE MARKET LIBERALISM

A strong ideology in the nineteenth-century, free market liberalism, was revived in the late twentieth century in the neo-liberalism which tends to support global marketisation. It is advanced by techno-centric environmentalists like Anderson and Leal (1991), following the example of Simon and Kahn (1984). They are aggressively optimistic about the free market's potential, allied to technology, to solve environmental problems. The invisible hand of market forces under which individuals pursue self-interest, they argue, gives society more environmental protection than will any kind of intervention or regulation, which is a constraint on liberty. Thus if a natural resource' is running out, its increased scarcity will push up the price of the goods or services that come through that resource. This will encourage entrepreneurs, with scientists and technologists, to devise some substitute for the resource, or more ingenious ways of providing the same goods and services. Similarly, there is money to be made from clean technologies like non-harmful aerosols, biodegradable plastics, catalytic converters and the like, so, as ecological modernizers say, there is no practical dichotomy between the interests of capitalism and environmental quality.

This ideology also underlies calls for privatization of the natural resources of the earth's commons. Its basic argument is that property held in common tends to be neglected, whereas individuals look after what they own. If private owners were to abuse or run down their resources, they themselves would suffer. And if others polluted their resource (their land or the waters within it), they could have recourse to law, which would defend their property rights. Groups of citizens might also sue the originators of pollution which lowered the quality of the air or the seas in their neighborhood. (This philosophy thrives particularly in the United States, where ordinary citizens have frequent recourse to law.) In response to objections that the originators of pollution may be hard to identify, the free-marketers might suggest technological solutions, such as the possibility of tagging' all emissions with trace amounts of identifying chemicals.

The free-market ideology also underlies systems of tradeable pollution rights, initiated in the US in the 1990s (*see* Tradable Permits for Sulfur Emission Reductions, Volume 4). The Federal Government sells to industry the rights to emit, for instance, sulphur dioxide. The total amount of rights issued conforms to a ceiling determined by Government as appropriate to an area: this ceiling can be raised or lowered as thought fit. If a firm's productive processes exceed the amounts of pollution that it has rights to emit, it can either pay to install cleaner technology or buy further rights from other firms who have less need for them. So far, this approach has met with limited success.

SOCIAL REFORMISM (ALSO KNOWN AS WELFARE LIBERALISM, DEMOCRATIC SOCIALISM)

Social reformists also believe in capitalism, but not without restraints and controls to limit its harmful effects on some people (the economic losers) and the environment. They emphasize the role and supreme importance of the individual and his/her enlightened self-interest in protecting the environment. In Western pluralist democracies, the parliament or congress constitutes the main forum through which environmental views and interests will be heard and protected. Rationality, the rule of law, technology and environmental and economic management (cost-benefit analysis, reform of taxation) will all help to secure the goals of environmentalism.

Following the father of English liberalism', John Stuart Mill, social reformists are ambivalent about how desirable unselective economic growth is, appreciating also the need for diversity in society and nature (compatible with their belief in pluralism). Support, is still strong for private ownership of resources, and the notion of the invisible hand' by which individuals seeking to maximize their personal benefit allegedly bring greatest aggregate benefit, to society. But they are also aware of the wider society and communal good, and that what is good is what brings most benefit and happiness to most people (Bentham's utilitarianism). By this measure, environmental protection is rational and desirable. Hence planning laws, taxes on non-recycling industries or pollution, and welfare provision to enhance urban environmental quality and environmental

education, are all legitimate ways for the state to intervene, but only as much as is strictly necessary, in the free market.

Social reformists are generally techno-centric, and many of their approaches underlie ecological modernization. They are predicated on significant government intervention in the free market, manipulating it in favor of environmental measures. Hence they advocate tax incentives and disincentives, permits and other moderately coercive management devices to control pollution and the use of environmental resources in the earth's commons'. The rationale behind the eco-consumerist movement owes much to welfare liberalism's belief in enlightened self-interest (i.e., that educated, informed and rational people are aware of the benefits to themselves of not behaving totally selfishly and without thought for wider social and environmental interests).

In Europe, social reformists are found in all mainstream groupings, including Conservative, Liberal, Social Democrat/Democratic Socialist and Green parties. As social reformists, democratic socialists emphasize their belief in the need for pluralist democracy and the power of parliament as the major way to achieve social change. Many now believe that capitalism can be reformed to take on board traditional socialist aims if it is guided and occasionally coerced by a moderately strong central and local state. Few influential members would advocate the necessity of abolishing capitalism or the global market as the answer to contemporary social or environmental problems. However, this form of socialism would like production to meet social and environmental needs alongside the exigencies of profit creation, hence it can be quite strongly interventionist in the free market. It may, too, favor decentralization of many of the state's functions.

The approach to sustainable development advocated by the UK's Real World Coalition' of environmental and social welfare groups (Jacobs, 1996) is essentially social reformist. It might be argued that in today's climate of neo-liberal global marketisation, based on extending and derestricting free trade and competition, their aspiration to elevate social inclusion and environmental sustainability considerations to a level of importance equal with profitability is somewhat utopian (*see* Political Systems and the Environment: Utopianism, Volume 5).

REVOLUTIONARY SOCIALISM

For years, many environmentalists have dismissed socialism as being equally culpable with capitalism in creating environmental crisis. They have regarded it as part of mainstream culture because of its unashamed anthropocentrism, and would therefore baulk at seeing it categorized in Figure 1 as engendering a form of radical *environmentalism*. But socialism itself is more complex than they seem to believe. And eco-socialism derives from particular socialist traditions that in fact have nothing to do with the kind of materialism and disregard for nature traditionally associated with the state centralist, self-styled socialist' regimes behind the old iron-curtain. Eco-socialism's roots lie in utopian socialism and the Marxism of William Morris (*see* Political Systems and the Environment: Utopianism, Volume 5, and Wall, 1994), rather than Stalin. As such, it is clearly counter-cultural, radical and opposed to mainstream values. Though eco-socialists and social ecologists criticize each other, they also have much in common.

Revolutionary socialism is a minority political current, unlike reformist socialism. However, together with anarchism it has strongly influenced all radical environmentalism. It fundamentally rejects capitalism and advocates production for social need but not profit, and common ownership of the means of production and distribution; thus it would deal with the commons' problem by having these areas under genuine common ownership and control. It contains different schools of thought about the role and importance of the state in revolutionary and post-revolutionary society. Though the state has no place in the ultimate communist society, a more democratic and decentralized form of state than exists in most contemporary Western society may be seen as important in helping to create a transition towards that society. Revolutionary socialism's analysis of the society–nature relationship generally follows Marxist lines, seeing nature and environment as socially produced', and environmental problems as inherent in the nature of capitalism. Again, opinions differ on how much they may be *solely* located here. The analysis of how to get to an ecological' society – which equates with a (perhaps moneyless) communist one, is based on the mechanism of class struggle. Neo- and orthodox Marxists, however, may differ amongst themselves on who might be the principal agents and actors in creating this society – faith in the potential effectiveness, or even existence, of a working class proletariat' is not universal amongst eco-socialists (see Gorz, 1982).

Marxists tend to see the rise of environmentalism itself in class terms, sometimes regarding it as bourgeois and counter-revolutionary. This may mean open hostility towards environmental groups and campaigns, though more often than not there is an uneasy alliance between reds and greens based on agreement on ends but not necessarily on means or analysis of causes.

Marxists by and large believe that environmental, feminist, peace and third-world campaigns are, or should be, all part of the ultimate struggle against global capitalism itself. From Marxism's materialist perspective, the ills which these campaigns highlight are all outgrowths of capitalist relations of production. Such campaigns, furthermore, should focus less on the reform of the individual's attitudes and values and more on the collective political struggle of the world proletariat'.

Major influences on European environmentalism, like Andre Gorz and Rudolph Bahro, have tended towards the revolutionary socialist view in the past, but no longer do so. The Socialist Party of Great Britain, Socialist Workers' Party and Militant (who spell out the eco-socialist case eloquently and simply in Cock and Hopwood, 1996) are all British political groups who incorporate green issues into orthodox Marxist analysis.

GREEN ANARCHISM

Anarchism suffuses all varieties of radical environmentalism, including deep and social ecology, eco-socialism and eco-feminism. Anarchistic groups within environmentalism share fundamental beliefs about the need for organic societies' but some are anti-urban and anti-industrial, so displaying affinities with conservative thought. American eco-anarchists may incline towards radical, libertarian liberalism. Indeed, in espousing the individual as the basic social unit, anarchism is sometimes thought to embrace the cornerstone of liberalism although anarchists might define concepts like individual' and freedom' in ways different from that of a mainstream liberal.

But then much anarchism has major affinities with socialism, and it is probably true that eco-anarchists are mainly anarcho-communists, looking particularly to Kropotkin for inspiration (*see* Political Systems and the Environment: Utopianism, Volume 5), or possibly anarcho-syndicalists (based on trade unions as major revolutionary agents). There is the rejection of capitalism, desire for common ownership of the means of production (resources), and, in resource and income sharing communes, with distribution according to need. But, unlike socialism, there is strong rejection of any possible role for the state, intermediate or not.

Eco-anarchism informs the rural communes movement, but may also inspire urban communes, the urban Reclaim the Streets' movement, the squatter movement (property is theft') and self-help urban community groups. The urban movement is particularly strong in North America, while in Australasia, anarcho-syndicalism – which is uncompromisingly urban-centered – has been a powerful force in green/trade union activism. Anarchists are prominent in what is becoming a world-wide militant movement against global capitalism in the early twenty-first century.

Notwithstanding the common roots and many correspondences which exist between eco-anarchism and eco-socialism, there are also some major divergences between the two (see Pepper, 1993). For instance eco-anarchists often reject the old class politics, seeing social change as consequent on the action of individuals in forming spontaneous, mutualist, non-hierarchical groups to live out their politics, thus setting an example for others to follow. The personal is political', and person equals planet' are maxims from more liberal American anarchism (i.e., Roszak, 1979), which appear to have great influence. Anarchism considers that social hierarchies and not class constitute the key to oppression: hence eliminating hierarchy is the prerequisite for eliminating oppression. Socialists argue the reverse, insisting on the need for revolutionary political class struggle to change the relations between people, which are contingent on the material, economic, organization of production. Anarchists with an emphasis on practical lifestyle politics believe that only by people sidestepping (living outside) the economic and social structures of conventional society can those structures be overthrown.

FEMINISM

The notion that we must eliminate social oppression and hierarchy if we are to form an ecological society is but one of many ideas, which are held in common between feminism and anarchism as applied to radical environmentalism. Eco-feminists often propose an essential convergence between women and nature. This is firstly because their biological makeup inevitably associates women, more than men, with the natural functions of reproduction and nurturing. Secondly, women and nature have in common that they are exploited by men, both economically and in being objectified and politically marginalized. During and since the 1970s, debates within eco-feminism have focused on several schools of thought, including cultural eco-feminism and social (socialist/anarchist') eco-feminism (*see also* Ecofeminism, Volume 5).

Cultural eco-feminism considers that the root problems of how Western cultures treat nature could be solved by a women's culture', providing practical and philosophical guidelines to sustainable development. Since menstrual cycles follow phases of the moon, and fertility follows the rhythm of the seasons, then women might be regarded as inherently close to nature. Hence eco-feminist culture would draw on ancient myths combining women and nature, mother and earth, in cooperative relationships: caring, nurturing, mutually giving and receiving.

This eco-feminism says, further, that female culture' is concerned with the body, the flesh, the material; with natural processes, emotions and subjective feelings and private life. By contrast, male culture' emphasizes the mind, intellect, reason, culture, objectivity, economics and public life. Male culture always seeks to transcend natural constraints on what humans can do: men constantly fight to conquer, exploit and mould nature to leave their mark behind and achieve immortality and transcendence. Merchant (1982) describes how, during the period of the scientific revolution, Francis Bacon and the Royal Society pledged to reveal the secrets still locked in her [nature's] bosom' and to conquer' and subdue' her. Most high technologies, including new female reproductive technology, are developed by men

and are said by this school to constitute a continuation of men's dual domination of women and nature.

Cultural eco-feminism, then, means liberating nature from the repressive male ethos so that it will be respected as a sustainer of life. Capra (1982) urges the necessity for a return to a balance of *yin* and *yang* characteristics in people and society as the foundation for a new, ecologically benign society (*see* Land Transformation in China: 3000 Years of Human Ecological History, Volume 3). This might be achieved through women, individually and in groups, discovering their authentic natures and celebrating and affirming them. Such consciousness-raising may need to exclude men, on the grounds that they could have a negative impact on it while it is still nascent and before it is strong enough to resist male domination. Then there can also be celebration of pagan myths and rituals, and associated pastimes like tarot cards and astrology, which affirm respect for Mother Nature and the essential interconnectedness of humans and nature.

This latter leans towards New Age thinking, and some greens and feminists (particularly of the social variety) reject it. As Plumwood (1993) argues, the whole idea of connecting with nature' could be regressive and insulting, portraying women as passive reproductive animals immersed in the body and in unthinking experience of life.

Materialist social eco-feminism draws on traditions in non-mainstream European socialism, including utopian socialism, classical anarchism, early Marx and William Morris' work. They all insist that exploitation of nature relates to exploitation in society, emphasizing social and political rather than personal aspects of the domination of women and nature. Hence women's oppression is seen as interwoven with class, race and species oppression. But social eco-feminism also rejects the crude economic class reductionism of some Marxism; so it does not accept that women's oppression is *merely* a special case of exploitation of the proletariat', or that to establish socialism would mean automatically ending women's or nature's oppression.

Mellor (1997) wants to modify or reconstruct' Marxist theory, to make a socialist eco-feminism. If, she argues, the way societies organize themselves to get material subsistence (economic relations of production) plays a key role in shaping society, so also must the way we organize ourselves to materially continue as a species (i.e., the relations of reproduction). Hence the material world of motherhood, not merely the (still largely male) material world of industrial production, should provide ideas and values for shaping an alternative (socialist) society. These ideas and values are altruistic: they include taking immediate responsibility for meeting the needs of others. But as things stand, the dominance of capitalist relations in Western society ensures that the stereotypically male' standpoint mediates all knowledge. Its view of nature as a commodity will prevail over the female view of unity with nature.

Other calls to construct a social eco-feminism often emphasize the importance of ideas and attitudes, calling on women to unite with the environmental movement to reshape the underlying values of a conventional society based on hierarchical organization and domination of some people by others. According to Plumwood, one of the major wrong ideas in our culture is the tendency towards dualistic thinking. For, as Eckersley (1992) suggests, patriarchy is a subset of such thinking – the philosophical dualism that has pervaded Western thought'. So, for instance, imagining that there is a fundamental distinction between society and nature suggests that they are separated, making it easier for the former to exploit the latter. And strongly identifying women with nature, as in cultural eco-feminism, implicitly or explicitly links men with exploitation of both. All this reinforcement of dualistic thinking feeds that very patriarchal culture to which feminism is opposed.

Eco-feminism shares with socialism an internationalist mentality; opposing women's oppression worldwide. It recognizes that in the third world, *women*, not men, constitute the backbone of production' as well as reproduction'. Eco-feminism in this context has therefore meant pushing for women to be involved in decisions about land use and control. Eco-feminists have resisted land appropriation by government and (Western) commercial firms, while, however, the men – who own the land – have often given in to the seductions behind the modernization model of development.

While some third-world women have been co-opted by modernization, helping to provide the cheap labor for inappropriate development such as dams and nuclear power stations, others in the 1980s have instead adopted the bypass' strategy. They have set up enterprises and movements that attempt to exclude the involvement of international capital (e.g. Chipko in India, South African cooperatives, green zones' in Mozambique). They have supported, therefore, a localized development model rather than global modernization.

MAINSTREAM GREEN IDEOLOGY: OLD OR NEW POLITICS?

Bearing the mark of all these influences, mainstream green political ideology, or ecologism' (Dobson, 1995) is an eclectic, pragmatic and shifting mix of ideas, but with some recurring themes. Green activists often hold it as a point of principle that they are neither left or right, but forward', or above the old politics': The basic political choice today is not between Right, Left or Center, but between conventional gray politicians and the Green Party', said that party's 1992 UK manifesto. But of course, when they start to talk about what we should do about the eco-crisis, environmentalists must become involved in old' politics. Hence the ideology of ecologism contains tensions between:

1. Ecology-based politics, emphasizing the pre-eminence of sustaining natural systems as the universal basis for social organization. Such eco-centrism constitutes ecologism's main claim to be a new politics.
2. Socially-based politics, seeing ecological problems as outgrowths of social problems, and seeking to solve the former by, first and foremost, solving the latter. These constitute the old' politics, raising questions, which have been the concern of traditional ideologies for several hundred years.

Environmentalists who concentrate on the second category often seem to follow a relatively coherent and identifiable analysis of the problems and sets of prescriptions to remedy them – i.e., they align with one or other of the more traditional political ideological camps. Most of those in the first category, who prioritize eco-centric concerns, present a more diffuse image, spreading themselves across the political spectrum and accumulating bits of traditional liberal and socialist ideology, spiced with dashes of conservatism. In the terminology of the old politics, perhaps a majority of these eco-centrics gravitate towards social reformism (especially in their public utterances), while a vocal minority embrace, especially, eco-anarchism, in the context of one or more of the radical camps' outlined in Figure 1. But for all those in this first category, politics start from the perspective of *ecology* and the need to maintain eco-system integrity. For those in the second, social matters are as important (or more so) than ecological ones: they are fundamentally anthropocentric.

An overall predisposition towards social reformism emerges in common mainstream green positions, which say that social change must proceed from individuals, but that change is also needed in the economic structures of society. They often do not totally reject capitalism – indeed are enthusiastic for at least small-scale versions of it, but see social need and environmental quality as criteria to be elevated above the profit motive. They see a crucial but benign role for the state, in both facilitating the development of individual responsibility and laying down regulatory and legal frameworks governing group behavior in particular. This (grudging) acceptance of the state (and of parliamentary politics by Green Party supporters) wanes where eco-anarchism acquires more influence.

Nature is the ultimate legitimation for and the source of social laws in the mainstream green canon. But, as noted above, matters of social justice are increasingly important to most environmentalists. Technology is not rejected, but it must be appropriate and democratically accessible to all, as well as soft' on nature. Scientific knowledge and rationalism are regarded as useful, but many environmentalists insist that these must be balanced by elevating emotional and intuitional knowledge to a more influential position in environmental decision-making than that currently accorded. Democracy and individual freedom are cornerstones of mainstream green ideology. Democracy is to be extended to all nature's creatures (animal rights, vegetarianism, veganism). But the importance of the community is also stressed. However the collective, e.g., trade unions in the production processes, is not seen to be as important as in traditional politics. Emphasis in mainstream green ideology, as in social reformism generally, has swung from our collective power as producers to our individual power as consumers.

The disdain for the industrial way of life', and the old politics' which was noted above in connection with early environmentalism, still persists as a current in mainstream environmentalism. It may be coupled with a tendency towards irrationalism and mysticism – which becomes more strident in deep ecology and New Age approaches. Whether they publicly own or disown this spiritual wing of the radical ecology movement, it is apparent that most mainstream greens do have deep ecology/New Age tendencies'. This shows in their support for a bio-ethic, for nature mystification (Gaianism), and their belief in spiritual paths and self-discovery, self-realisation and consciousness raising through therapy techniques as routes to political change. The innate conservatism in idealizing nature, and in the denial of a politics of social change in favor of one of individual change, is often uncomfortable to those more on the left of the green movement.

STRATEGY

If mainstream green thinkers and activists do show an eclectic mix of pragmatic politics, taken partly from old reformist and radical political ideologies, and putting an ecological emphasis onto the whole, this could be seen in positive or negative terms.

On the positive side, it could be argued that mixing elements of older ideologies will produce something more appropriate to the modern world – which is not the world in which these ideologies were formulated. Thus a politics can be reached which is compatible with the needs of an ecologically sound society: these needs might best be seen not in the old terms of class war or material wealth distribution alone. An ecological crisis like holes in the ozone layer threatens us all, and in any case so-called socialist societies have not had a better relationship with nature than capitalist societies. Such a common sense' approach to politics might also appeal to a wide range of people, not being seen as extremist'.

More negatively, the result of doing this could be merely an ideological mish-mash. It might sound theoretically appealing but is destined to break down in practice, because it contains some irreconcilable contradictions. (Such as the idea proposed by some environmentalists that green consumerism and other pressures will create multinational corporations that are environmentally sustainable

and will put environmental and social goals before that of profit.)

Ideological inconsistency could lead to some strange conceptions of a sustainable society, as conveyed by some libertarian environmentalists. They demand significant increases in personal freedom, and an end of state domination of society. But at the same time they campaign for potentially repressive universal laws embracing the interests of natural' eco-systems. This could lead to a society where communities which sanctioned the dumping of toxic wastes could be heavily punished, but those who allowed child abuse or racism to flourish might be excused on the principle that local communities should have an absolute right to self-determination in the social sphere.

For one of environmentalism's recurrent themes is respect for a diversity of cultures, i.e., respect for otherness'. Part of the contemporary debate about environmental politics concerns such post-modern sentiments, along with the tendency discussed above to reject progress' and modernity. While some greens like Atkinson (1991) are content to see environmentalism as essentially a form of post-modern politics, others are anxious to rescue the Enlightenment Project and the gains from the Industrial Revolution for environmentalism. John Barry (1999), for instance considers ideological' green politics in relation to the French and industrial revolutions. He thinks that they want to accept and radicalize the democracy begun by the former, but to reject the industrial revolution. This latter would be mistaken, he believes. He says, probably correctly, that green politics should be an immanent critique of modernity' which however does not want to reject progress': rather, to redefine what is meant by the word, making it apposite to strong sustainability principles, including social justice and ecosystems integrity. With this approach, the promise of modernity, which has hitherto eluded the majority of the world's people, will not be forgotten, but can perhaps be realized for this majority.

See also: Precautionary Principle, Volume 4.

REFERENCES

Anderson, T and Leal, D (1991) *Free Market Environmentalism*, Pacific Research Institute for Public Policy, San Francisco, CA.

Atkinson, A (1991) *Principles of Political Economy*, Behaven, London.

Barry, J (1999) *Rethinking Green Politics: Nature, Virtue and Progress*, Sage, London.

Beck, U (1992) *Risk Society: Towards a New Modernity*, Sage, London.

Beck, U (1995) *Ecological Politics in an Age of Risk*, Polity, Cambridge.

Bookchin, M (1990) *The Philosophy of Social Ecology*, Black Rose Books, Montreal.

Bramwell, A (1989) *Ecology in the Twentieth Century: a History*, Yale University Press, London.

Callenbach, E (1981) *Ecotopia Emerging*, Banyan Tree Books, Berkley, CA.

Capra, F (1982) *The Turning Point*, Wildwood House, London.

Carson, R (1962) *Silent Spring*, Houghton Miflin, Boston, MA.

Cock, M and Hopwood, W (1996) *Global Warning: Socialism and the Environment*, Militant Publications, London.

Connelly, J and Smith, G (1999) *Politics and the Environment: from Theory to Practice*, Routledge, London.

Devall, W and Sessions, G (1985) *Deep Ecology: Living as if Nature Mattered*, Gibbs M Smith, UT.

Dobson, A (1995) *Green Political Thought*, 2nd edition Unwin Hyman, London.

Eckersley, R (1992) *Environmentalism and Political Theory: Towards an Ecocentric Approach*, University College Press, London.

Ehrlich, P (1969) Eco-catastrophe, *Ramparts*, 8(3), 24–28.

Ehrlich, P and Ehrlich, P (1990) *The Population Explosion*, Simon and Schuster, New York.

Gorz, A (1982) *Farewell to the Working Classes: an Essay on Post-Industrial Socialism*, Pluto, London.

Hardin, G (1968) The tragedy of the commons, *Science*, 162, 1243–1248.

Hardin, G (1974) Living on a lifeboat, *BioScience*, 24, 10.

Heilbroner, R (1975) *An Inquiry into the Human Prospect*, Calder and Boyars, London.

Jacobs, M (1996) *The Politics of the Real World*, Earthscan, London.

Mander, J and Goldsmith, E, eds (1996) *The Case Against the Global Economy: and for a Turn Toward the Local*, Sierra Club Books, San Francisco, CA.

Mason, M (1999) *Environmental Democracy*, Earthscan, London.

McKibben, W (1990) *The End of Nature*, Penguin, London.

Meadows, D H, Meadows, D L, Randers, J, and Behrens, W (1972) *Limits to Growth*, Earth Island, London.

Meadows, D H, Meadows, D L, and Randers, J (1992) *Beyond the Limits: Global Collapse or a Sustainable Future*, Earthscan, London.

Mellor, M (1997) *Feminism and Ecology*, Polity, Cambridge.

Merchant, C (1982) *The Death of Nature: Women, Ecology and the Scienti c Revolution*, Wildwood House, London.

Ophuls (1977) *Ecology and the Politics of Scarcity: Prologue to a Political Theory of the Steady State*, W H Freeman, San Francisco, CA.

Pepper, D M (1991) *Communes and the Green Vision: Counterculture, Lifestyle and the New Age*, Green Print, Basingstoke.

Pepper, D M (1993) *Eco-socialism: from Deep Ecology to Social Justice*, Routledge, London.

Pepper, D (1996) *Modern Environmentalism: an introduction*, Routledge, London.

Plumwood, V (1993) *Feminism and the Mastery of Nature*, Routledge, London.

Porritt, J (1984) *Seeing Green*, Blackwell, Oxford.

Roszak, T (1979) *Person/Planet*, Gollancz, London.

Schumacher, E F (1973) *Small is Beautiful: Economics as if People Really Mattered*, Abacus, London.

Simon, J and Kahn, H (1984) *The Resourceful Earth*, Blackwell, Oxford.

Wall, D (1994) *Green History*, Routledge, London.
Wall, D (1999) *Earth First! and the Anti-Roads Movement*, Routledge, London.

Political Systems and the Environment: Utopianism

David Pepper
Oxford Brookes University, Oxford, UK

Utopias are imaginary locations with perfect social and political systems. Utopianism is strong in environmental thought, where political and economic commentators often spell out the principles of an ecologically sound society, and writers of ction create ecotopias: ecologically ideal societies of the future. Additionally, environmental activists may set up small, would-be ecotopian communes, where they try to practice self-suf ciency within non-hierarchical social relationships.

The ecologically ideal society is usually envisaged as small-scale and decentralized, with local communities forming the basic social, economic and political unit, but larger units are also formed through confederation. Environmental and social quality of life is high, and consumption of material goods is low in these local economies where production is for social need, not pro t. There is direct democracy, and nation states and international trade via multinational corporations play a subordinate role, or are non-existent. Nature is respected and stewarded, and biodiversity is high.

Ecotopianism draws on utopian activists and thinkers of the past, including the 17th century Diggers and the utopian schemes and experiments of anarchists and utopian socialists such as Charles Fourier and Robert Owen in the early 19th century. The late 19th century writers, William Morris and Peter Kropotkin, have perhaps been the greatest historical in uences. Their socialist and anarchist visions are apparent in today's social ecology schemes for an ecological economy and society, while Teilhard de Chardin's mid-twentieth century ideas about the evolutionary progress of humankind have particularly in uenced deep ecology's view of a future perfect society.

Utopianism is also a term sometimes applied to those who seem to have unrealistic, impractical and ill thought out ideas about society's future direction. It is not only the idealistic schemes of radical environmentalists which are prone to this somewhat derogatory charge. The vision of the future proffered in the mainstream model of development known as ecological modernization (EM) could also be guilty of utopianism. For it imagines a future highly competitive capitalist, free market global economy where matters of social justice and environmental quality are given equal or greater priority than the pro t motive; a somewhat unlikely scenario perhaps.

INTRODUCTION: MEANINGS OF UTOPIANISM

The Oxford English Dictionary defines *utopia* (no-place) as an imaginary location with a perfect social and political system. Most contemporary thought and action about the environment could be described as utopian in some way. Environmentalists, be they radical or reformist (*see* Political Movements/Ideologies and the Environment, Volume 5), wish to see significant change towards a much improved society where development will be sustainable, that is, where the old antipathy between the exigencies of human progress and the integrity and diversity of natural ecosystems is resolved in some kind of harmony.

Kumar (1988) describes utopianism as a form of thinking, invented by Thomas More (*Utopia, 1516*), which became a social movement in the West in the 19th century. From then on, some people, including environmentalists, have striven to realize utopia here on earth. Whereas earlier utopias, including those of the Renaissance and Enlightenment, were Christian-based (Manuel and Manuel, 1979, referred to hereafter as the Manuels), Europe and America were swept in the 19th century by a wave of secular messianism, of which socialism was the dominant expression.

Kumar identifies three locations of utopian expression:

1. in practical experiments, for instance of Robert Owen and many New World communities;
2. in fictional societies, like Etienne Cabet's *Icaria*;
3. in social philosophies, such as those of Saint-Simon, Fourier, Comte and Marx.

These examples were all socialist-leaning, and a related ideological movement, that of anarchism, has been much connected with socialist utopianism. Both are extremely germane to contemporary environmentalist utopias, heavily influencing classic ecotopias; such as those of Reich (1970), Schumacher (1973), Le Guin (1974), Goldsmith *et al.* (1972), van der Weyer (1986) and, notably, Callenbach's *Ecotopia* and its prequel *Ecotopia Emerging* (1978, 1981). But there are also other influences, such as Pierre Teilhard de Chardin's *Phenomenon of Man* (1965), which feeds strands of idealism and technological optimism into, particularly, deep ecology/new age utopianism: this also sometimes reflects the utopianism of capitalist free market ideology.

Authorities on the utopian subject, such as the Manuels, regard Marxism as a secular resurgence of millenarian (believing in a period of good government, great happiness

and prosperity ; Oxford English Dictionary), paradisaical, apocalyptic thinking. Marxism is an unchristian euchronia (looking forward to an ideal time rather than place) in as much as it sees a future society of inde nite progress and growing freedom. In it, Men make their own history in full consciousness of their power, collectively, to shape their own society, and not subject to forces, economic, historical or mystical, beyond their control. (Socialists, however, often deny that Marxism is utopian, since Marx speci cally rejected the notion that members of a future communist, free society should be shackled by visions and blueprints inherited from previous generations.) Many environmentalists today, whether they call themselves Marxist or not, assert similarly that the key to an ecologically benign society is an inclusive democracy where the will of the majority prevails over economic and social decision-making. Armed with such freedom and with reason, they argue people will avoid living inharmoniously with nature, since this would undermine majority interests.

As the Manuels point out, utopianism is popularly associated with a literary genre: a dramatic narrative portrayal of a way of life which is essentially good, embracing the underlying principles of an optimum society and embodying a set of implicit (usually optimistic) assumptions about human nature. But additionally, utopianism is a state of mind which often appears in the constitutions of a restructured polity and in the plans, outlines or blueprints of a universal republic. The line between such idealized blueprints and political and social theory is often blurred.

The Manuels also show how utopianism is tied to speci c social con icts, and utopians, in rejecting the existing society, take political sides. Much late 20th century utopianism (including ecotopianism) re ects the condition of post-modernity, in its skepticism about modernity and the Enlightenment Project (the mission expressed strongly in the 18th century, and founded on scienti c, technological and industrial development in pursuit of universal material and social progress). Progress, as expressed through material af uence and the achievements of science and technology are often eschewed. The nuclear family, Fordist production, global modernization: all may be conspicuous in contemporary utopianism by their absence.

The Manuels refer, almost disparagingly, to specialist utopias (including environmentalism and feminism) identifying with 1960s and 1970s counterculture, which arose in response to post-industrial apocalyptic visions of overpopulation, nuclear disaster, or Frankensteinian experiments gone wrong. (Two ecological dystopias which incorporate such visions are by Harrison (1966) and Wright (1997)). These, in true post-modern fashion, often entirely rejected theory. They had, the Manuels say, no identi able character beyond being usually virulently anti-urban. A bit of transcendence, body mysticism, sexual freedom, the abolition of work, the end of alienation, they wryly add (p. 808), noting that this countercultural utopianism appealed to young people who found conventional society wanting, but it was, unlike earlier utopianism, not oriented to the long idealized future. Rather, it related to immediate existence and to the miniature rather than the large-scale model. This seems only partly true of ecotopianism at the end of the 20th century, which tends to bear out the Manuels view (p. 803) that however much it appears to change, the ••• utopian bazaar is cluttered with old fashioned wares that are all too familiar.

The picture is complicated by the fact that, of course, not all contemporary utopianism is countercultural. As will be argued below, the neo-liberal utopianism of the free market suffuses today s mainstream ideology and practice of global modernization, and its mainstream environmental component, known as *ecological modernization* EM. The case for identifying this as utopian rests on a normative interpretation of the word which is particular to Marxist thought. Essentially, this holds that while it may be valid to envisage and facilitate social change (i.e., towards socialism), it is utopian, i.e., unrealistic and impractical, to do so in a way that underplays the importance of factors (especially economic) militating against such change. The charge of utopianism, in a derogatory sense, was leveled by Marx and Engels against the utopian socialists such as Robert Owen, and the anarchists:

> What was designated utopian according to this approach, was the imagination of the possibility of total social transformation involving the elimination of individualism, competition and the sway of private property, *without a recognition of the necessity of class struggle and the revolutionary role of the proletariat in accomplishing the transition.*
>
> (Jones, 1983, 505, emphases added)

By extension, this means that the approach of some eco-anarchistic alternative communities, who think it possible to create in the midst of capitalist society a coexisting microcosm of a non-capitalist utopia, which it is hoped might spread by example, is utopian, naive, and doomed to failure. It also implies that social democratic attempts to reform capitalism, rather than replace it by radically changing its economic character and relations of production, (the EM approach) are similarly unrealistic and destined to remain in that place called *nowhere*.

RADICAL ENVIRONMENTALISM'S ECOTOPIA SUMMARIZED: THE IDEAL MODEL OF SUSTAINABLE DEVELOPMENT

Modern environmentalism bears the imprint of both radical and reformist political ideologies (*see* **Political Movements/Ideologies and the Environment**, Volume 5). There are four main strands of radical environmentalism: *deep ecology* (which often fuses into *new ageism*), *social ecology*, *eco-feminism* and *eco-socialism*. These strands are

politically diverse and often incompatible; yet they also share strong unifying features, in their intent, first, to replace current society by some form of radical alternative and, secondly, to ensure that this radical alternative should be intrinsically socially and environmentally sustainable and just.

A third source of unity is the very considerable commonality in their detailed visions of ecotopia/ecochronia. The lineaments and features of their economic, political and social life, and their geographies, have more in common than of difference. In terms of the contemporary jargon of models of sustainable development, ranging from weak to strong, the radical environmentalists' ecotopias may be seen as ultra-strong, or as Baker *et al.* (1997) term it, they constitute an ideal model (IM). One criterion by which this IM of sustainability differs from other approaches is that it does not countenance a role for conventional economic growth in its definition of development: rather it seeks development in other aspects of wealth, such as quality of life.

The account here draws on ecotopian works referred to previously, and also in Sale (1985), Devall and Sessions (1985), Kemp (1992), Milbrath *et al.* (1994), Mander and Goldsmith (1996), Douthwaite (1996) and Fotopoulos (1997); a fairly representative and diverse set of radical environmentalists. However, notwithstanding their divergences, they share most if not all of the principles outlined in Table 1, amounting to development aims of cultural identity, self-reliance, economic and social democracy, social justice and ecological balance. Working from such principles, some important elements of the economy, society and geography of the sustainable society corresponding to this IM can be pictured.

Table 1 Major principles underlying the ideal model IM of sustainable development

Economy
Economy's purpose is to satisfy need, not to create profit. Needs include satisfying, creative work, good health, spiritual and cultural fulfillment, social interaction and high environmental quality, as well as food, shelter, etc.
Communities/individuals define their own needs and satisfiers
Local production to meet local basic needs
Economic confederation and inter-regional exchanges to meet extra local needs, e.g., energy, transport, cultural, and non-basic needs
Mutual aid rather than strict reciprocity is the basis for any exchanges
Economic wealth should not be concentrated or taken out of communities
Total or partial common ownership and control of means of production and distribution
From each according to ability: to each according to need, within limits, or unconditionally
Holistic accounting
Sustainability and self-reliance imply living within limits: limits to economic growth and consumption
Stable human population levels
Community/regionally planned economy; markets have minimal or no role
Society and politics
Local community is the basic social unit
Confederation and inter-regional interchange (e.g., via information technology) to avoid parochialism, bigotry, conservatism, deviancy
Cultural diversity giving social strength
Participatory, bottom-up democracy, informed by subsidiarity
Individuals/communities take responsibility for their own lives
These last two imply economic democracy, direct political democracy, egalitarianism and non-hierarchical relationships, i.e., dispersal of power
Emphasized values include cooperation, non-hierarchy, non-patriarchy, equity, justice, social cohesion, compassion, love, peace, respect for nature
Holistic thinking
Science/technology controlled by and serves the whole community
Education a key resource, and tool for social change
State has an enabling role, or none at all
Nature
Valued for its own sake and/or for humanistic goals
Biodiversity fostered and preserved
It is axiomatic that the community focus of society fosters sustainable relationships with nature
Permacultural principles apply to interaction with the land
Each system/production cycle can continue indefinitely without ecological deterioration

Economy

Stability, self-reliance and cooperation are key IM concepts underlying production, work and distribution. Self-reliance would root communities back into their locality, decreasing dependency on the global economy. It implies as much self-sufficiency as possible, locally meeting basic needs for food, housing, welfare, and energy. Communities could generate their energy from multiple combined local sources; watermills, photovoltaics, windmills, and biomass such as straw, dung, woodchips from coppicing and forest waste. Contributions to and withdrawals from a grid would coordinate production, while conservation, and savings from reduced transport, would greatly reduce energy needs.

The principle of common ownership of the means of production in action would be seen in farms and horticultural enterprises owned by community land trusts and/or supported by subscribers paying for food in advance of or after distribution. Such enterprises exist today, and are usually small and substantially organic, utilizing low energy inputs and local plant and animal strains. Industries, services and shops would also be community supported. In Fotopoulos' eco-socialist/anarchist utopian economic democracy,

citizens' assemblies would make all the macroeconomic decisions. Micro decisions, on how individual factories and businesses operationalized citizens' targets, would be controlled by workplace assemblies delegating powers to supervisory boards.

Eco-socialist and deep ecology visions often dwell on the virtue of creative artisanship (rather than the current alienating division of labor) to fulfil the social aims of work. They may also emphasize team production, and the need for individuals not to over-specialize. However, this does not amount to rejecting specialist roles, including management. But possessing and exercising expertise must not create power hierarchies: there should be distinction without rank.

While some, like Sale, envisage labor without wage, looking perhaps to distribution via free shops, Fotopoulos dismisses this as being based on a socialistic fantasy of inherent abundance. He shares most green and neo-classical presumptions of inherent scarcity, and would distribute work, goods and services in his hypothetical non-money economy through a complicated scheme involving basic and non-basic vouchers. Everyone would receive the former for life's essentials, with a strong expectation that all able bodied people will work. Most would also do extra-voluntary work, for non-basic vouchers to exchange for community-designated luxuries (perhaps obtained through exchanges with other regions and communities). But whether vouchers or money were used, they would be only locally valid, after the manner of existing local employment and trading systems (LETS) schemes, and could not be accumulated or removed from communities by individuals.

In the transition towards the ultimate non-money economy, Fotopoulos envisages that people's savings and investments would be effected through local banking, credit union and provident schemes, which would all be amenable to confederation. In very many ecotopian visions, the anarchist's conferral principle would be vital in equalizing resources between communities, in investing in and organizing infrastructure, and in overall production planning. For Fotopoulos, outcomes, in individual incomes and wealth, might well be unequal, but inequality would result from individuals' voluntary decisions, e.g., to do less work, or perhaps to do the most sought-after jobs, which therefore might attract less remuneration than undesirable tasks.

Society and Politics

All ecotopias rest on the argument that more inclusive styles of government, oriented towards generalizable interests, are better suited to addressing environmental degradation than existing less inclusive styles. Inclusive styles are predicated on decentralized decision-making, including economic decisions. Hence, IM visions all revolve around communitarian values, reviving, says Douthwaite, the spirit of the Irish Meitheal, which was and is a working party system based on mutual support and trust within and between local communities. Such visions tend, Kemp *et al.* (1992) believe, towards true communism; where producers and communities join together to regulate rationally and for themselves their exchanges with nature; a situation increasingly rare within capitalism. This context fosters individuals' fulfillment, but not at the expense of others.

Most radical visions emphasize libertarian, anti-statist principles, but their society is also highly organized, needing strong civil institutions. The spatial division of power is intended to promote equality, efficiency, welfare and security. The fact that this may be utopian in the sense of unrealistic, is recognized by Le Guin, whose anarchist planet, Anarres, is by no means perfect. She sees that because of the principles of diversity and self determination, not every community or region would be likely to heed democratic, just and ecologically sound values. (Le Guin's quasi-utopia, incidentally, is more believable and palatable by virtue of its presentation, warts and all, as an imperfect society; often dour, austere and a place of daily struggle to maintain cherished principles. It is, paradoxically, more inspiring than the incredible worlds of William Morris, Huxley or Callenbach, where people seem perpetually wreathed in sanctimonious smiles of happiness). In detail, therefore, the search is for a system that works even if people in it are not good, with structures, or devices like peer disapproval, to minimize errant behavior undermining majority interests.

Fotopoulos supplies considerable structural detail of his future inclusive democracy. Its citizens' assemblies, taking local decisions, would be augmented through rotating, mandated, recallable delegates, on appointed bodies like community courts or militia. They would also send similar delegates to regional and confederal councils, where different community aims would have to be composited and operationalized via majority voting or consensus.

Geography and Understanding of Place

It is here, perhaps, that different versions of IM have most in common; all reasserting localism to countervail capitalism's inherent globalism. Global modernization's geography is one of dependency, whereby the lives of communities are shaped by events and decisions originating at considerable distances. By contrast the IM's decentralized geography would create protective space for communities and regions. In its communitarian cultures and collective economies, families and communities would own their means of livelihood. Its mainly small enterprises would minimize returns to capital and maximize returns to labor, and much investment would be in local rather than spatially distant regions.

To radical environmentalists, one key benefit of decreasing the scale of living would lie in enabling people to understand readily the effects of their actions on their community and its ecosystems. Communal ownership of the

means of production would preclude externalizing social and environmental costs, because there would be no external places. For where value systems fostered mutual aid rather than economic competition, it may be reasonably supposed that other parts of the world would be regarded as extensions of one's own community and region. The confederal principle, allied to information technology, should help such changes in geographical perception. Parochialism and spatial inequity could thus be avoided. Outside relationships and exchanges, not being based on strict market reciprocity, would likely foster non-dependent, non-injurious interaction and cultural tolerance, in what would be a global economy. Localism, by providing economic security, should create a climate where racism and authoritarianism would wither away. Nation states might also wither, giving way to continental and sub-continental confederal alliances. The European Union (EU) might become a series of autonomous regions within a unified European framework.

Bioregionalists, such as Sale, and many deep ecologists, seek to define regions by recognizing natural boundaries within which clear carrying capacity limits are acknowledged. Knowing about regional cultures, folklore, history and natural cycles would constitute a major element in bioregional citizenship, involving walking the territory and attuning consciousness to nature. A *bioregion* is any part of the earth's surface whose rough boundaries are determined by natural characteristics rather than human dictates, (Sale, 1985, 55). It is distinguished from other bioregions by particular attributes of flora, fauna, water, climate, soils and landforms, and by the human settlements and cultures these attributes have given rise to.

There is much variation in views about the maximum viable size of the primary political unit, the local community. Estimates may range from one to 10 000 people. This would clearly necessitate politically reconstructing most cities into relatively autonomous neighborhood units, confederated into city regions: perhaps, as Sale envisages, by breaking bigger cities down into smaller ones of 250 000 maximum size, generating their own energy. In this way, 150 new USA cities would be created. Additionally, cityscapes are much changed, as in Callenbach's *Ecotopia*, where gardens, rooftops, community allotments and city farms burst forth with the fruits and flowers of intensive cultivation. Douthwaite envisages that increased rural self-reliance would give city-based service providers less to do, encouraging countryward migration. This anarchistic scattering of the city into its local region would re-establish city/country reciprocity: what remained of the former would provide rural services and be fed by its rural hinterland through box schemes (regular supplies of boxes of produce direct from farm to consumer), farmers' markets and the like.

PRINCIPAL INFLUENCES ON RADICAL ECOTOPIANISM

As suggested above, radical ecotopia reflects influences from the 19th century and before, deriving particularly but not exclusively from socialist and anarchist utopianism. To environmentalists who see common ownership of land as a key issue (reflected in groups like *Reclaim the Streets* or *The Land is Ours* or the alternative communities movement), Gerrard Winstanley is a seminal figure. In *The Law of Freedom in a Platform, or True Magistracy Restored (1652)* Winstanley outlined a vision where land would be set free for the commoners. Environmentalists celebrate him particularly for his actions in assembling a community of commoners, the *Diggers*, who dug and planted the commons on George Hill, Walton, Surrey, in 1649. In their reluctance to envisage violent seizure of landowners' private property, and their hope that instead the landowners would voluntarily give up their lands and join them, these diggers foreshadowed an important aspect of green anarchist/social ecology strategy: that of prefiguring the desired society. This means that no strategy or approach (like violence, or social hierarchies) which would not feature in the utopia itself should be contemplated as part of the transitional struggle; this point became central to 19th century arguments between anarchists and Marxists.

A more paternalistic form of prefiguring might be seen in Robert Owen's utopian experiments. A businessman and mill owner between 1800 and 1812, thereafter he began to criticize contemporary society and economy. He reformed the conditions for his own mill workers at New Lanark, Scotland, then proposed and initiated self-sustaining communities of about 500–3000 people each, without private property, organized in rectangular units close to manufacturing plants and surrounded by intensive agriculture. Such exemplary communities were to lead to a cooperative socialism embracing the earth, formed by a new generation of rational beings.

Another utopian socialist, Charles Fourier (1772–1837), also strongly influences anarchist-feminist, communalist radical environmentalism. He envisaged phalansteries, communities of about 1700 people, as basic social units, with no overarching state regulation. In them, creative talents were expanded, meals and child-care were communal, and industrial armies carried out environmental projects. Traditional family relationships were dissolved and replaced by a wide spectrum of sexual relationships. The phalansteries reflected Fourier's view that all humans are born with their own innate instinctual drives, and utopia would be a place where each personality would be allowed to come out. Although capital, private property and wealth disparities remained, all were emotionally satisfied and spiritually rich. This suggests an essentially liberal utopia: the creation of an environment where all individuals could satisfy themselves. Fourier's

theme of enhanced life quality, perhaps compensating for lower material standards, features strongly in the seminal ecotopian visions of Schumacher (1973) and Goldsmith *et al.* (1972).

The 20th century's commune and back-to-the-land radical environmentalism (Pepper, 1991) has many links and resonances with similar 19th century alternative communities (Hardy, 1979, and Gould, 1988). Two of the strongest influences on them have been Peter Kropotkin and William Morris. O'Sullivan (1990) considers that Morris applied Marxism ••• to provide radical environmentalists with a document setting out many of their basic ideas in plain English. This includes voluntary simplicity; rejecting the false material wants and sprawling, ugly, polluting developments engendered by capitalism in favor of ••• the simple joys of the lovely earth ••• (Morris, 1887), and fellowship. He demonstrated his eco-socialist credentials by identifying commercialism and commodification under capitalism as the source of nature's impoverishment. Unlike deep ecologists, he did not oppose the fruits of industry *per se*, particularly organized in small workshops as craft production, nor did he abhor any change made to nature by humans. He saw a place for machines, in doing any unpleasurable work: again a theme in much modern ecotopianism, as seen, for instance, in the post-industrial utopia of Andre Gorz (Frankel, 1987). Morris also inveighed against pollution, urbanization of the countryside by mass housing and alienation of the laborer from his/her creativity, and consumerism. In *News from Nowhere*, there is no private property as a right, hence no hierarchy based on it, and little crime, except for that based on sexual passions. *Nowhere*, say the Manuels, goes alongside Edward Bellamy's Fourierist *Looking Backward 2000- 1887 (1888)* in a tradition of nostalgic socialist utopias which bypass the dynamism of industrial-scientific modern civilization. Their visions represent genteel socialism; non-threatening to Victorian capitalism (although Morris savagely attacked capitalism) and therefore open to the charge of utopianism in the Marxist sense.

Morris' vision is close to anarchism, and for Kropotkin (1892, 1899) anarchism would lead to true *commune*ism: the free association of producers without class division, wage slavery or even money. He initially conceived of his ideal communes as a specific spatial form (based on the Russian *mir*), but later, like Marx, he saw them less specifically as groupings of individuals with mutual interests in thousands of towns and villages. His picture of an ideal communist landscape is close in detail to that of the IM, and his principle of mutual aid as a key evolutionary law heavily infuses both *social ecology* and (paradoxically) *deep ecology* and *new ageism*. Fotopoulos's (1997) inclusive democracy echoes Kropotkin in many respects, although Fotopoulos, a disciple of Murray Bookchin's social ecology, does not, like many eco-anarchists, fall prey to utopianism in underestimating the immense power of capital to resist challenges to its self-perpetuating exigencies.

Much radical ecotopianism (bioregionalism, deep ecology and the new ageism into which this can merge) shows anarchistic leanings in its visions, but nonetheless lacks anarchism's materialist, socialist perspective as the basis of such visions. Rather, it is idealistic, and here two 20th century utopians seem particularly influential, Teilhard de Chardin and Aldous Huxley. Callenbach's *Ecotopia* (1978) closely resembles Huxley's *Island* (1962). The latter was a curtain raiser to 1960s' and 1970s' idealism and its rejection of consumer materialism, conformity, alienation and the perceived dangers of globalization and science allied to capitalism, such as large-scale development, nuclear war/power, pollution and waste and overpopulation (Huxley figured on the sleeve of the Beatles' *Sergeant Pepper* album).

Huxley's remedies for these evils, and the basis of his utopia, draw particularly on traditions associated with Eastern mysticism and the passive and pragmatic praxis they foster, and on the imagined liberating power of drugs and sexual freedom. His utopians achieve nirvana on an island paradise surrounded by global threats like oil exploration or the terrifying increase in China's population, by a combination of tantrik yoga, coitus suspendus, meditation, trance, autosuggestion and awareness-enhancing drugs, all mixed in with scientific enlightenment, particularly in the application of biological and ecological principles. His utopians avoid the pitfalls of wider society by virtue of the right people being intelligent at the right moment and they feel no need for things bad and thin and boring', such as speedboats or television, wars, revolutions, revivals or political slogans. This all resonates with that strain in environmentalism and other new social movements that emphasizes the power of reason and right thinking, answering perhaps to the angst of middle-class Western intellectuals, but being highly utopian in the sense of eschewing the realities of class politics based on the economic nature of capitalist society.

Teilhard de Chardin's *Phenomenon of Man* (originally written in 1947) is one of a complex of influences on idealistic radical ecotopianism; a perspective suffusing deep ecology and new age tendencies that see evolution towards a higher state of consciousness as the key to ecological salvation. These influences include Eastern and Western mysticism, and certain interpretations of 20th century science by such writers as Capra, Sheldrake, Bohm, etc. They are competently synthesized in Peter Russell's *Awakening Earth* (1991). The starting point is with the evolution of Gaia (the earth as a living being) which is considered to be guided by design rather than Darwinian chance into new orders of complexity, diversity and connectivity (*see* Gaia Hypothesis, Volume 5). This holistic, organic, monist and optimistic evolutionary perspective embraces

de Chardin s theology, envisioning Earth s becoming a social super-organism. It would comprise diverse but interdependent individuals and societies, coalescing in their thoughts and ideas through the power of love and the growing density of global communications. Eventually billions of individuals form a single inter-thinking group or *noosphere* (*see* **Noosphere**, Volume 5). This is a realm of consciousness which has evolved from material living matter (biosphere), that in turn evolved from non-living matter (geosphere). Eventually all three spheres are to become an organic unity. This euchronia will be *point Omega*, when Gaia will have become one supermind. Beyond this, Gaia might merge into a Gaia eld: a galactic superorganism.

Such a conceptual scheme loosely follows Hegel s view of history as a process of emerging world spirit. The noosphere is a universal belt of psycho-social forces that has evolved, not smoothly, but in a series of leaps in response to environmental crises at different stages of earth s history. The Manuels note that for all the science (anthropology, biology) buttressing this schema, it can only be looked on as a dream of reason. The speci cally Christian element in de Chardin has not, on the whole, impacted on modern ecotopianism, although van der Weyer s *Wickwyn* (1986) is a rare exception.

However, feminist utopianism has been more in uential. The society of *Ecotopia*, for instance, is almost ruthlessly feminist, while Capra s New Ageist *Turning Point* (1982) makes feminine values the key to impending ecotopia. Marge Piercey s *Woman on the Edge of Time* (1979) was much-loved by 1970s radical environmentalists. In it, the heroine jumps from contemporary society to an androgynous one where relationships and child care are rede ned, lifestyles are simple and natural, and there is, again, free love. The latter theme is recurrent in ecotopian IMs, because of its supposed contribution to heightened quality of life, and mitigation of the overpopulation problem. A more distant feminist utopia, Charlotte Perkins Gilman s *Herland* (1915), is usually ignored, though in it most things which environmentalists deplore (war, greed, money, stupidity) have been banished.

EM: REFORMIST ECOTOPIANISM?

Both left- and right-wing critics of free market liberalism have long detected a form of utopianism (in the derogatory sense) in the belief that the free market can bring a world of social justice, ecological sustainability and universal freedom (Gray, 1999). However, in the past two decades a new environmental discourse has emerged, which seems to give new impetus to this utopianism. Rejecting the pessimistic 1970s *Limits to Growth* perspective, this discourse embraces modernity, rationalization and internationalism, and is posited on continuing global marketization under capitalism. EM has become the mainstream, reformist organizing basis for new institutional procedures and perspectives to manage environmental problems.

As an approach to sustainable development, EM is seen by its most enthusiastic adherents as not merely a technocratic project to modify global modernization environmentally, but also as a form of institutional learning and cultural politics that heralds new ways of seeing and new opportunities for an emerging global civil society in the face of global economic forces and environmental problems.

EM discourse is complex, but its core features, which have become the basis of the environmental policies of Western states, are outlined in Table 2. Essentially, EM theory holds that a new phase of development is possible, and is happening, where capitalism (involving constant economic growth, technological development and the spread of consumerism through global marketization) and the goals of environmental conservation are reconcilable. Indeed, it sees environmental protection not as an impediment to capital accumulation but as a potential source of further accumulation; economic bene ts and competitive advantage being said to accrue from preserving genetic diversity and from anticipatory environmental protection, rather than paying out to clean up a mess. In this positive-sum game, technological and managerial experts, business and industry all become key actors in ful lling the environmental agenda, rather than its enemy, as environmentalists often suspect.

As an approach to sustainable development, EM clearly ts well with capitalist ideology and could be regarded as a form of social democracy without the big stick; for while the state legislates for EM s policy principles, in the contemporary ideological climate of *laissez-faire*, its role becomes contextual and steering, rather than *dirigiste* as in post-war Europe. For left-leaning environmentalists, particularly, however, EM represents a form of environmental utopianism in the Marxist sense, essentially because it envisages widespread new behaviors that would in reality serve as a brake or impediment on the operation of capitalist economic forces. Therefore, its principles and assumptions are bound to be unrealistic.

For instance, EM proposes that conventional economic growth and environmental protection are compatible; yet in reality economic growth leads today to global warming, rapid loss of species diversity and other undesirable effects. It assumes that rms will embrace anticipatory environmental technology because of its long-term profitability, but in fact most Western rms still opt for the quicker pro ts offered by end-of-pipe technologies (or by simply evading environmental standards). In any case it is unrealistic to assume that anticipatory technology can signi cantly lower environmental degradation in a growth economy, since gains from such technologies (e.g., catalytic converters) are more than offset by growing consumption (e.g., more cars and road-kilometers).

Table 2 Features of EM – Ecological Modernization[a]

EM's Realms (Hajer)
(1) New policy-making principles: e.g., from react and cure to anticipate and prevent; integrated pollution abatement; integrating environmental concerns into all ministries; techniques allowing firms to integrate environment into cost-risk calculations, such as polluter pays, cost-benefit analysis, risk analysis, precautionary principle, tradable pollution rights, pollution charges and taxes
(2) A new role for science in policy-making, especially ecological systems science. Experts take a central role
(3) Environment protection seen as a source of growth: low- and non-wasteful anticipatory technologies generate profits
(4) Nature reconceptualized: a public rather than a free good, so efficiency involves internalizing environmental costs
(5) Burden of proof reassigned: to the suspected polluter, not the damaged party
(6) Policy-making opened up: new participation and partnerships, e.g., between business and non-government organizations. Voluntary agreements rather than command and control regulation are encouraged
Weak EM Discourse Includes (Dryzek)
Anticipating unwanted outcomes of production/consumption, through systems approaches
Instrumental view of nature as provider of resources and services
Continued existence of capitalist political economy
Reassurances about continuing modernization, growth, progress
Weak EM Discourse Excludes (Dryzek)
Consideration of: limits to growth; Third World development paths; international environmental footprints of industrial development; a non-instrumental view of nature; alternatives to capitalist political economy
Preoccupation with equity and environmental justice
Unfettered free market economics
Unrestrainedly optimistic view of growth and development.
Strong EM Discourse May Include (Christoff)
An international dimension
Non-western development approaches
Deeper changes in beliefs and morality
Elevating equity, futurity, ecological imperatives by comparison with narrow economic goals
Re-embedding society in community, region, ecosystems: limits to modernization

[a] *Sources:* Hajer (1995), Dryzek (1997), Christoff (1996).

Attitudinal change underpins EM theory, which envisages that firms will embrace the more holistic accounting and planning for longer time horizons which are necessary for sustainability. However, such changes are not at present widespread among, for instance, US and EU businesses, because they take most decisions from short-term, sector-specific perspectives, as must happen in the increasingly competitive global marketplace.

Hence the hope that capitalist institutions can or will significantly reform, to take on environmental sustainability, is pious. For this would require giving equal weight to economic and to environmental and social justice goals, so that economic and civil spheres would become compatible as in the social democratic agenda. But in the real world Western civil and social democratic politics are increasingly subordinated to global marketization's economic agenda. For to re-establish profit levels after the 1950s–1970s round of Keynesian welfare spending, capital has often resisted social justice and environmental goals. These goals usually represent unacceptable costs. Whereas the sustainability agenda requires evening up wealth distribution, and substantial social spending, capitalism's inevitable dynamic is to concentrate wealth (Fotopoulos, 1997). And EM contains other utopian, idealistic assumptions which would not bear close scrutiny, for instance that Western affluence could be universalized in an environmentally sustainable way, that trickle down economics work, or even that economic growth is a universal good.

On this account, then, it is not only young and starry-eyed idealists or iconoclastic intellectuals who indulge in environmental utopianism: establishment pillars do the same, though in a different way. This raises the question of whether utopianism is a good or bad thing. If everyone does it yet no environmental resolution is reached, perhaps the habit should be discouraged. However, the notion of futurity is a cornerstone of the sustainability debate, hence it would be utopian to envisage that utopianism could ever be avoided in that debate.

REFERENCES

Baker, S, Kousis, M, Richardson, D, and Young, S (1997) *The Politics of Sustainable Development: Theory and Practice within the European Union*, Routledge, London, 1–40.

Callenbach, E (1978) *Ecotopia*, Pluto Press, London.

Callenbach, E (1981) *Ecotopia Emerging*, Banyan Tree Books, Berkley.

Capra, F (1982) *The Turning Point*, Wildwood House, London.

Christoff, P (1996) Ecological Modernisation: Ecological Modernities, *Environ. Politics*, 5(3), 476–500.

de Chardin, P T (1965) *The Phenomenon of Man*, Harper and Row, New York.

Devall, B and Sessions, G (1985) *Deep Ecology: Living as if Nature Mattered*, Gibbs M Smith, Salt Lake City, UT.

Douthwaite, R (1996) *Short Circuit: Strengthening Local Economies for Security in an Unstable World*, Green Books, Totnes.

Dryzek, J (1997) *The Politics of the Earth: Environmental Discourses*, Oxford University Press, Oxford.

Fotopoulos, T (1997) *Towards an Inclusive Democracy: the Crisis of the Growth Economy and the Need for a New Liberatory Project*, Cassell, London.
Frankel, B (1987) *The Post Industrial Utopians*, Polity Press, Cambridge.
Gilman, C P (1915) *Herland*, 1979 edition, Woman's Press, London.
Goldsmith, E, Allan, R, Allaby, M, Davoll, J, and Lawrence, S (1972) Blueprint for Survival, *Ecologist*, 2(1), 1–43.
Gould, P (1988) *Early Green Politics: Back to Nature, Back to the Land and Socialism in Britain*, Harvester Press, Brighton.
Gray, J (1999) *False Dawn: the Delusions of Global Capitalism*, Granta, London.
Hajer, M (1995) *The Politics of Environmental Discourse, Ecological Modernisation and the Policy Process*, Clarendon Press, Oxford.
Hardy, D (1979) *Alternative Communities in Nineteenth Century England*, Longman, London.
Harrison, H (1966) *Make Room! Make Room!* (filmed as *Soylent Green*), 2nd edition (1992), Penguin, Harmondsworth.
Huxley, A (1962) *Island*, Harper Collins, London.
Jones, G S (1983) Utopian Socialism, in *A Dictionary of Marxist Thought*, ed T Bottomore, Blackwell, Oxford, 504–506.
Kemp, P *et al.* (1992) *Europe's Green Alternative: a Manifesto for a New World*, Green Print, Basingstoke.
Kropotkin, P (1892) *The Conquest of Bread*, Freedom Press, London.
Kropotkin, P (1899) *Fields, Factories and Workshops Tomorrow*, Freedom Press, London.
Kumar K (1988) Utopianism, in *The Fontana Dictionary of Modern Thought*, eds A Bullock, O Stallybrass, and S Trombley, Fontana, London, 888–889.
Le Guin, U (1974) *The Dispossessed*, Victor Gollancz, London.
Mander, J and Goldsmith, E (1996) *The Case Against the Global Economy and for a Turn Toward the Local*, Sierra Club Books, San Francisco, CA.
Manuel, F E and Manuel, F P (1979) *Utopian Thought in the Western World*, Blackwell, Oxford.
Milbrath, L, Downes, Y, and Miller, K (1994) Sustainable Living: Framework of an Ecosystemically Grounded Political Theory, *Environ. Politics*, 3(3), 421–44.
Morris, W (1887) The Society of the Future, in *The Political Writings of William Morris*, ed A L Morton (1979), Lawrence and Wishart, London, 188–203.
O'Sullivan, P (1990) The Ending of the Journey: Wm Morris, News from Nowhere and Ecology, in *Wm Morris and News from Nowhere: a Vision for Our Time*, eds S Coleman and P O'Sullivan, Green Books, Bideford, 169–181.
Pepper, D M (1991) *Communes and the Green Vision: Counterculture, Lifestyle and the New Age*, Green Print, Basingstoke.
Piercey, M (1979) *Woman on the Edge of Time*, The Women's Press, London.
Reich, C (1970) *The Greening of America*, Random House, New York.
Russell, P (1991) *The Awakening Earth: the Global Brain*, Arkana, London.
Sale, K (1985) *Dwellers in the Land: the Bioregional Vision*, Sierra Club Books, San Francisco, CA.
Schumacher, E F (1973) *Small is Beautiful: Economics as if People Really Mattered*, Abacus, London.
van der Weyer, R (1986) *Wickwyn: a Vision of the Future*, Society for the Propogation of Christian Knowledge, London.
Wright, R (1997) *A Scienti c Romance*, Anchor, London.

Politics, Environmental

see Environmental Politics (*Opening essay*, **Volume 5**)

Positivism

see Environmental Philosophy: Phenomenological Ecology (**Volume 5**)

Post-modern

see Political Movements/Ideologies and the Environment (**Volume 5**)

Post-normal Science

Jerome R Ravetz
Research Methods Consultancy Ltd., London, UK

Post-normal science is a new conception of the management of complex science-related issues (Funtowicz and Ravetz, 1992, 1993, 1994, 1999; Ravetz, 1999). It focuses on aspects of problem solving that tend to be neglected in traditional accounts of scienti c practice: uncertainty, value loading, and a plurality of legitimate perspectives. It provides a coherent explanation of the need for greater participation in science-policy processes, based on the new tasks of quality assurance in these problem areas.

It is a means of illuminating, explaining and then guiding innovative decision processes that are happening spontaneously all over the world. The management of complex natural systems as if they were simple scienti c exercises has brought us to our present mixture of triumph and peril. We are now witnessing the emergence of a new approach to problem-solving strategies in which the role of science, still essential, is now appreciated in its full context

of the complexity and uncertainty of natural systems and the relevance of human commitments and values.

INTRODUCTION

In the sorts of issue-driven science relating to global climate change, typically facts are uncertain, values in dispute, stakes high, and decisions urgent. The sound science that is frequently invoked as the requirement for rational policy decisions may be an unaffordable luxury. The traditional distinction between hard, objective scientific facts and soft, subjective value judgments is now inverted. All too often, we must make hard policy decisions where our only scientific inputs are irremediably soft. In these new circumstances, invoking truth as the goal of science is a diversion. A more relevant and robust guiding principle is quality, understood as a contextual property of scientific information.

Appropriate intellectual tools are required for these new tasks. A scientific methodology designed for controlled experimentation and abstract theory building can make effective use of a picture of reality that reduces complex phenomena to their simple, atomic elements. However, that is not best suited for the tasks of science-related policy today. The traditional normal scientific mind-set fosters expectations of regularity, simplicity and certainty in the phenomena and in our interventions. However, these can inhibit the growth of our understanding of the new problems and of appropriate methods for their solution.

These new problems are characteristic of complex systems. These are not merely complicated; they involve interrelated subsystems at a variety of scale levels and at a variety of sorts. Thus, we now know that every technology is embedded in its societal and natural contexts, and that nature itself is shaped by its interactions with humanity. In such complex systems, there can be no single privileged point of view for measurement, analysis and evaluation. Moreover, in these systems there is generally no hidden hand whereby selfish individual actions automatically benefit the wider societal and natural communities; hence there is no substitute for morality in the good conduct of our affairs. The phenomena of life, society, and now the environment, cannot be captured, nor their problems managed, by sciences which assume that the relevant systems are simple. In terms of such paradigms, they will always present anomalies and surprises. Post-normal science has been developed as the appropriate methodology for complex natural and social systems.

The term *post-normal* provides a contrast to two sorts of normality. One is the picture of research science as normally consisting of puzzle solving within the framework of an unquestioned and unquestionable paradigm, in the theory of Kuhn (1962). Another is the assumption that the policy context is still normal, in that such routine puzzle solving by experts provides an adequate knowledge base for decision making. The great lesson of recent years is that this assumption no longer holds. We may call it a post-modern rejection of grand narratives, or a green, not-in-my-back-yard (NIMBY) politics. Whatever its causes, we can no longer assume the presence of this sort of normality of the policy process, particularly in relation to global climate change. The frequent, almost ritualized calls for debate' on major policy questions of risks and the environment are an implicit recognition of their post-normal character.

In such novel contexts, there is a new role for natural science. The facts that are taught from the textbooks used in institutions are still necessary, but they are no longer sufficient. Contrary to the impression that the textbooks convey, in practice most problems have more than one plausible answer, and many have no well-defined scientific answer at all. Further, in the artificial world studied in academic science courses, it is strictly inconceivable that science-related problems could be tackled and solved except by deploying the accredited expertise. Practical techniques, which cannot be explained in principle by accepted science, are commonly dismissed as the products of blind tradition or of chance. When persons with no formal qualifications attempt to participate in the processes of innovation, evaluation or decision-making, their efforts have tended to be viewed with suspicion or scorn. Post-normal science provides a means for correcting this sort of mind-set.

POST-NORMAL SCIENCE

As a theory, post-normal science links epistemology (study of the grounds for knowledge) and policy, for its origins lie in the relations between those two domains. Its authors were concerned that the sciences devoted to solving risk and environmental problems (such as those in ecology and toxicology) are radically different from those that are instrumental in creating them (such as the applications of physics and molecular biology). In comparison to those traditional sciences, the policy-relevant sciences have enjoyed less prestige and funding, are less mature scientifically, and are more subject to external influences and constraints. By the criteria of the traditional philosophy of science, their results frequently fail to attain the status of sound science. It has been argued that they should, therefore, be rejected as evidence in policy debates; but a more appropriate conclusion would be that the philosophy of science needs recasting. Post-normal science provides a response to these crises of science and philosophy, by bringing facts and values into a unified conception of problem solving in these areas, and by replacing truth by quality as its core evaluative concept. Its principle of the plurality of legitimate perspectives on any problem leads to a focus on dialogue, and on mutual respect and learning, wherever possible.

Post-normal science comprises those inquiries that occur at the interfaces of science and policy where uncertainties and value-loadings are critical. It can be analyzed as a policy cycle including: policies, persons, procedures, products, and post-normal assessment; it also extends to the downstream phases of implementation and monitoring (Funtowicz *et al.*, 2000). Depending on the particular context, the task may be more like policy-related research, or science-related decision making, or creative technical-social innovation. In the case of global climate change, all three sorts are in play; the distinctions are never absolute, as the whole endeavor is a complex system.

Post-normal science can be located in relation to the more traditional problem-solving strategies by means of a diagram (see Figure 1). On it, we see two axes – systems uncertainties and decision stakes. When both aspects are small, we are in the realm of normal, safe applied science, where expertise is fully effective. When either is medium, then the application of routine techniques is not enough; skill, judgement, sometimes even courage are required. This is professional consultancy, with the examples of the surgeon or the senior engineer in mind. In such cases, the creative element in the problem is more an exercise in design than the discovery of facts. Our modern society depends on armies of applied scientists pushing forward the frontiers of knowledge and technique, with the professionals performing leading roles in technical and policy matters. In recent years we have learned that even the skills of professionals are not always adequate for the solution of science-related policy issues. When risks cannot be quantified, or when possible damage is irreversible, then we are out of the range of competence of traditional sorts of expertise and traditional problem-solving methodologies. This situation is represented on the diagram as the outer band, that of post-normal science. We notice that the band extends through the whole quadrant, right up to the region where systems uncertainties vanish. This feature reflects the fact that, if in some policy process the decision stakes are very high (as when an institution is seriously threatened by a proposed policy), then a defensive strategy will involve challenging every step of a scientific argument, even if the systems uncertainties are actually quite small.

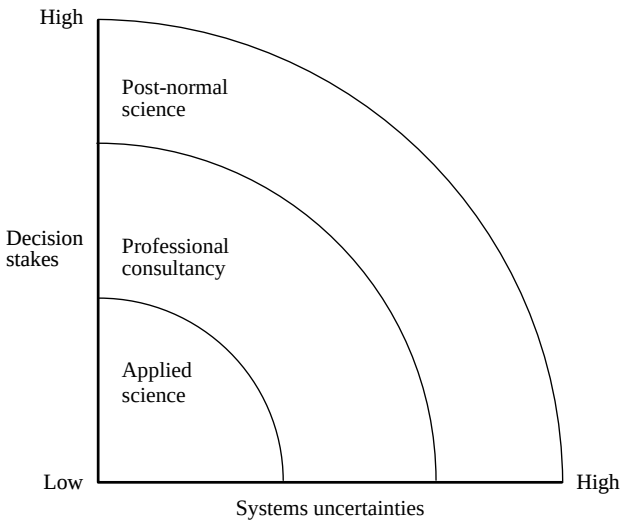


Figure 1 Post-normal science

The management of systems uncertainties through the involvement of decision stakes occurs even in routine science. Whatever the statistical test, there will always be errors: some true correlations will be rejected (Type I) and false ones will be accepted (Type II). A balance must, therefore, be struck between the error-costs of excess specificity and those of excess sensitivity, respectively; and that balance depends on the policy framework of the test. For example, if the problem is detecting possible harm from contaminants, it is better to err on the side of precaution and be more inclusive; a test designed around avoiding false positives could exclude potentially important information (Type I error), which could then remain permanently unknown. The well-known confidence level expresses this value-driven choice. Researchers do normally not assign it; rather they automatically apply the level that is standard for their field.

When a problem is recognized as post-normal, even the routine research exercises take on a new character. For the value-loadings and uncertainties are no longer managed automatically or unself-consciously. As they may be critical to the quality of the product in the policy context, they are the objects of critical scrutiny by researchers themselves as well as by their peers. Thus, normal science itself becomes post-normal, and is liberated from its traditional unreflective, dogmatic character.

The awareness of global climate change is well on the way to transforming all our conceptions of policy problems in the post-normal direction. For example, passenger transport had traditionally been seen as an essentially straightforward engineering problem of maximizing mobility, subject to the constraints of cost and safety. Now transport technologies and policies are strongly influenced by environmental considerations of many sorts. Moreover, the need to control the increase in private passenger transport depends directly on lifestyles. Together with the systematic uncertainties of global climate change, there are crucial decision stakes in conceptions of the good life, along with considerations of equity between peoples and generations. The entire population of passenger transport users effectively becomes an extended peer community; for the success of sustainable transport technologies will depend on the effectiveness of users' commitment to the values of the global environment.

EXTENSIONS OF THE PEER COMMUNITIES

There are now many initiatives, increasing in number and significance all the time, for involving wider circles of people in decision-making and implementation on environmental issues. The contribution of all the stakeholders in cases of post-normal science is not merely a matter of broader democratic participation. For these new problems are in many ways different from those of research science, professional practice, or industrial development. Each of those has its established means for quality-assurance of the products of the work, be they peer review, professional associations, or the market. However, for these new problems, the maintenance of quality depends on open dialogue amongst all those affected. This we call an extended peer community, consisting not merely of persons with some form or other of institutional accreditation, but rather of all those with a desire to participate in the resolution of the issue. Since this context of science is one involving policy, we might see this extension of peer communities as analogous to earlier extensions of the franchise in other fields, such as women's suffrage and trade union rights. This is not merely a matter of extensions of liberty; with post-normal science we can guide the extension of the accountability of governments (the foundation of modern democratic society) to include the institutions involved in the governance of science and technology.

Extended peer communities are already being created, in increasing numbers, either when the authorities cannot see a way forward, or when they know that without a broad base of consensus, no policy can succeed. They are called *citizens juries*, *focus groups*, *consensus conferences*, or any one of a great variety of other names; and their forms and powers are correspondingly varied. Nonetheless, they all have one important element in common: they assess the quality of policy proposals, including a scientific element, on the basis of their level of scientific comprehension combined with their knowledge of the ways of the world. Their verdicts all have some degree of moral force and hence political influence.

These extended peer communities will not necessarily be passive recipients of the materials provided by experts. They will also possess, or create, their own extended facts. These may include craft wisdom and community knowledge of places and their histories, as well as anecdotal evidence, neighborhood surveys, investigative journalism and leaked documents. Such extended peer communities have achieved enormous new scope and power through the Internet. Activists scattered among large cities or rainforests can engage in mutual education and coordinated activity, providing themselves with the means of engagement with global vested interests on less unequal terms than previously.

Along with the regulatory, evaluative function of extended peer communities, another function, even more intimately involved in the policy process, is springing up. The discovery is being made, again and again, particularly at the local level, that people not only care about their own environment but can also become quite ingenious and creative in finding practical, mixed social and technological means for its improvement. For local people can imagine solutions and reformulate problems in ways that the accredited experts, with the best will in the world, do not find normal. This is most important in the phases of policy-formation, and also in the implementation and monitoring of policies. Thus, in addition to extending the traditional processes of quality assessment, participants can enhance the quality of the problem-solving processes themselves. Even more than in environmental issues, numerous groups of patients and activists in the medical field have been demonstrating post-normal science at work.

Post-normal science belongs to a growing family of approaches to the understanding of the new predicaments of science and society. Beck (1992) has described the risk society and the development of sub-politics, effectively the extended peer communities. Wynne (1996) has contributed on extended facts and the need for extended peer review. Irwin (1995) has stressed the importance of the contribution of local communities and activists. Jasanoff (1998) has analyzed the effectively post-normal situation of science in the courtroom, and showed how citizens can exercise effective quality control there.

What we call science has undergone many changes over the centuries in its objects, methods and social functions. The previous century saw science rise to unimaginable achievements in knowledge, power and then destructiveness. With post-normal science we are characterizing the changes that will be necessary for our civilization to become sustainable, and thereby to survive.

REFERENCES

Beck, U (1992) *Risk Society: Towards a New Modernity*, Sage, London.

Funtowicz, S and Ravetz, J R (1992) Three Types of Risk Assessment and the Emergence of Post-normal Science, in *Social Theories of Risk*, eds S Krimsky and D Golding, Greenwood, Westport, CT, 251–273.

Funtowicz, S and Ravetz, J R (1993) Science for the Post-normal Age, *Futures*, **25**, 735–755.

Funtowicz, S and Ravetz, J R (1994) Uncertainty, Complexity and Post-normal Science, *Environ. Toxicol. Chem.*, **13**, 1881–1885.

Funtowicz, S and Ravetz, J R (1999) Post-normal Science – an Insight Now Maturing, *Futures*, **31**, 641–646.

Funtowicz, S, Shepherd, I, Wilkinson, D, and Ravetz, J (2000) Science and Governance in the European Union: a Contribution to the Debate, *Sci. Public Policy*, **27**(3), 327–336.

Irwin, A (1995) *Citizen Science*, Routledge, London.

Jasanoff, S (1998) The Eye of Everyman, *Soc. Stud. Sci.*, **28**, 713–740.

Kuhn, T S (1962) *The Structure of Scienti c Revolutions*, Chicago University Press, Chicago, IL.
Ravetz, J (1999) What is Post-normal Science? *Futures*, 31, 647–654.
Wynne, B (1996) Misunderstood Misunderstandings: Social Identities and Public Uptake of Science, in *Misunderstanding Science?* eds A Irwin and B Wynne, Cambridge University Press, Cambridge, 19–46.

Poverty

see Development and Global Environmental Change (*Opening essay*, **Volume 5**)

Poverty and the Environment

see Economics and Global Environmental Change (*Opening essay*, **Volume 5**)

Precautionary Principle

Tim O'Riordan
University of East Anglia, Norwich, UK

The precautionary principle emerged in Germany in the 1970s and has become a central plank in European Union environmental policy since the early 1980s. The precautionary principle as applied to environmental management is based on the idea of acting prudently in advance of full scienti c certainty, leaving room for ignorance in the use of resources, and placing the onus of proof on those who propose to change, to show that there will be no harm as a result of their proposed actions. In recent years the precautionary principle has moved from the realms of science and environmental management to the politics of risk, social responsibility, trade and protecting the well-being of the innocent and vulnerable.

The attraction of precaution lies in the aphorisms *a stitch in time saves nine*, and *better safe than sorry*. Its significance as a guiding principle in environmental science lies in its challenge to conventional science, its preference for the victim of potential environmental abuse, and its shift in the nature of valuation techniques in favor of a more communitarian democracy. The precautionary principle will prevail because the political mood of environmentalism is moving towards sustainability, because a more inclusive stakeholder involvement in environmental decision taking is becoming the vogue, because scenarios about possible future environmental states are much more readily available due to huge computing power and the internet, and because an anxious citizenry is beginning to take the management of environmental risk into its own hands.

The precautionary principle emerged from German law in the late 1970s as the *vorsorgeprinzip*. This was taken to mean acting with care and with anticipation, in order to prevent avoidable harmful effects to humans and the environment. The *vorsorgeprinzip* was promoted by the Brandt Government as a justification for state intervention in the economy, for forward planning, and for the encouragement of clean technology. It was thus a justification for the aggressive intervention of the state in land use management and in the direction of investment into export-earning technological advance, pushed by tough and pro-active environmental regulation.

The basic principles of the precautionary principle are:

- *Thoughtful action in advance of scienti c proof*: most environmental future states cannot be forecast with any certainty. There are neither the data nor the modeling accuracy to do the job. Where the burden of evidence falls on the side of possible harm, the precautionary principle is invoked, even though taking action will cost money.
- *Leaving ecological space*: because we cannot be sure how tolerant ecosystems are to human intervention, it is necessary to leave ecological space as a buffer against our ignorance. Sometimes this practice is placed into law, as, for example, is the case in South Africa. There, the 1998 National Water Act provides for an entitlement of water for nature as an act of precaution. Human users, beyond allocation for essential needs, have to come second. This general principle, notably as regards water, will become more widespread.
- *Care in management*: because it is not possible to predict all possible consequences of any environmental management decision, it is necessary to focus on the vulnerable and the most likely losers of a possible course of action. This max-min' strategy of risk management is a feature of empowerment in a modern democracy, transferring an element of trust and respect to the politically most marginal.
- *Shifting the burden of proof*: possibly the most dramatic outcome is the legal and political shift of the burden of proof away from those likely to be harmed by a possible course of action, towards those who plan to change the *status quo*. In the past, all that a developer had to do was to show that no likelihood of unreasonable harm would ensue for a particular course of action. With the

precautionary principle, the balance of proof falls on the developer to show that particular interests will not be harmed at all.

The precautionary principle is brought into environmental management for seven reasons:

- *There is an active civic science*: citizens want to be directly involved in providing the valuation and validation of scientific data and findings. The future of genetically modified organisms in foodstuff, for instance, is no longer a purely scientific matter. The more citizens groups are involved, the more the precautionary principle will be applied.
- *The scope for collapse of vulnerable ecosystems is real*: the scare stories over the losses of biodiversity, of tropical forest removal and of coral reef bleaching have caused many active citizens to believe that we simply cannot know enough about the precursors to ecological collapse. Indeed, it is difficult for ecologists to come up with indicator conditions in species that provide a satisfactory basis for predetermining system collapse. This is why a more participatory ecological science is required.
- *Dislocation at a local level accumulates to seriously adverse global consequences*: precaution is necessary to stop the global commons from being degraded by millions or even billions of micro-level individual actions; no one individual has any incentive, other than altruism, to control his/her consumer behavior. In this sense, precaution is both proactively regulatory, and also personally educative, for it creates a wish in both the producer of products and services, and in the consumer, for a sense of responsibility for planetary stewardship.
- *Vulnerable groups are championed by citizens' justice organizations*: the rise of the eco-justice movement has been stimulated by civil rights groups, and by the growing evidence that environmental harm is disproportionately directed towards the poor, the vulnerable and the politically marginalized. The act of precaution is now being used to create a citizen's right to a healthy environment. Again, this procedure places a political bias on disadvantaged groups to be empowered. How far this can effectively be implemented remains to be seen.
- *Business is worrying about corporate reputation*: the scares over genetically modified organisms in food, or infectious salmon anaemia in fish-farmed salmon have seriously damaged public trust in corporations promoting such practices. This, in turn, has led to a popular resistance to eat food from fields or fish farms experiencing these modifications: in various recent polls, over 54% of British consumers will consciously avoid such foodstuffs. Therefore, businesses embarking on environmental audits and social responsibility programs also expose themselves to applying the precautionary principle in a more aggressive manner.
- *Regulation is more transparent*: the role of the regulator has been made more difficult due to the rise of the Internet, and to the scrutiny of intervener consumer and environmental organizations. No longer is regulation a secretive and private matter. Increasingly, all regulatory actions have to be open and accountable to an active citizenry that is looking to the precautionary principle to guide the levels of environmental safeguards being set for all classes of social groupings. Like science, regulation is a more participatory activity, and hence values and aspirations of various interested parties are now a legitimate part of standard setting and permitting.
- *Direct action is more ubiquitous*: direct action, involving illegal protest, over administrative and commercial acts that no longer are tolerated by angry citizens organizations, is not new. But these days it has adopted a greater potency due to the variety of opportunities available, the inquisitiveness of the media, the tolerance of the politicians, and the intolerances of the police. An additional factor is the general feeling that playing by the rules will not result in adequate political response. Direct action is a recognition that formal democracy is ineffective. Informal democracy is seen as a lever of citizen power. In these days of precaution and civic activism, direct action becomes increasingly attractive and innovative (*see* Public-driven Response to Global Environmental Change, Volume 4).

For precaution to work properly, however, there needs to be a change in both science and law beyond what is occurring today. The incorporation of soft values relating to being more in tune with nature and more sensitive to the legitimate aspirations of all others is a matter that separates interdisciplinary science as currently practiced from post-normal science (*see* Post-normal Science, Volume 5). The science of including feelings, of introducing trust, and of nurturing sensitivity for the interests of others is not quite born.

Nor are the joint property rights of stewardship fully developed. This is the extension of property rights to include the well-being of ecosystem functioning and social connectedness in all aspects of managing land, natural resources, and the commons. There is a part-way stage in this legal revolution in the application of the *public trust doctrine*. This states that any avoidable damage to the interests of others must be recognized and compensated, and also that the ability of those likely to be adversely affected, should be enhanced by training and mediation for empowerment.

This is the widening context of the application of the precautionary principle in contemporary environmental management. It is opening windows on science, law, politics,

regulation and moral norms. Precaution continues to be a truly transformationist concept.

See also: **Environmental Movement the Rise of Non-government Organizations (NGOs)**, Volume 5; **Political Movements/Ideologies and the Environment**, Volume 5.

FURTHER READING

Cable, S and Cable, C (1995) *Environmental Problems, Grassroots Solutions: The Politics of Grassroots Environmental Conict*, St Martin s Press, New York.

O Riordan, T and Cameron, J, eds (1994) *Interpreting the Precautionary Principle*, Earthscan Publications, London.

O Riordan, T and Jordan, A (1995) The Precautionary Principle in Contemporary Environmental Management, *Environ. Vals.*, **4**(2), 191 212.

O Riordan, T (1999) Environmental Science on the Move, in *Environmental Science for Environmental Management*, 2nd edition, ed T O Riordan, Longman, Harlow, 1 27.

O Riordan, T, Cameron, T, and Jordan, A, eds (2001) *Reinterpreting the Precautionary Principle*, Cameron May, London.

Rawcliffe, P (1992) *Environmental Groups in Transition*, Manchester University Press, Manchester.

Tickner, J and Raffersperger, C, eds (1999) *Putting the Precautionary Principle into Practice*, Island Press, New York.

Process Theology

see Christianity and the Environment (**Volume 5**); Theology (**Volume 5**)

Property Rights and Regimes

Bonnie McCay

Rutgers, The State University of New Jersey, New Brunswick, NJ, USA

Property rights are fundamental institutions of all societies. They are the formal and informal rules about how people relate to things and places they value. Property rights may be defined in terms of who holds rights whether persons (individuals, business firms and other units with the legal identity of persons), organizations, governments, or loosely defined latent groups (Olson, 1965) such as the public or all humankind. Thus, we talk of individual property, communal property, government property, and public property (as for example in the Swedish allmennsretten). Property may also be understood in terms of the kinds of rights people hold, including degrees of exclusivity and inclusivity. Private property is at the exclusive end of this continuum, and public property, implying the right not to be excluded (Macpherson, 1978), is at the inclusive end.

The language of property regimes is often used in discussions of environmental problems. Property regimes are systems of governance or regulation that differ from each other in terms of which property rights are most prevalent or controlling. Many writers use terms similar to open access, common property, state property, and private property when talking about different systems of resource management (e.g., Bromley and Cernea, 1989; Feeny et al., *1990). Although there are problems with this scheme (McCay, 1996) and other ways to label the systems (e.g., Ostrom* et al., *1999), it provides a useful framework for discussing property regimes and the environment.*

In trying to understand how property rights affect the ways people relate to their environments, we specify just what kinds of rights are held and by whom, and examine how particular property rights are nested within larger systems of rights and duties as well as their embeddedness in cultural and political traditions. In the Anglo American legal tradition, property is often described in terms of bundles of rights, i.e., distinct though interrelated rights (Rose, 1994). For example, exclusive rights of access or use, rights to income from the property, rights to change attributes of the property, the right to gift or will it to one s heirs, and the right to sell or lease it may or may not be part of a particular bundle of rights or ownership. The owner may have some sticks in the bundle of rights but not all, and ownership can be overlapping and complementary. For example, in a housing condominium, the owner of an apartment or townhouse holds exclusive rights to live in and decorate the interior but has ownership in common with others of the exterior of the building and perhaps the grounds. He/she may hold the right to vote on decisions about management of the common property but is required to abide by the decision rules. His/her right to transfer ownership to someone else may or may not be subject to approval of an association of co-owners. The condominium is in a larger community, which subjects it to taxation and to zoning and other regulations, and so forth.

An open access system is one where there are few if any barriers to entry or use and no regulation of the activities of resource users; this is the condition meant by commons in the well-known essay *The Tragedy of the Commons* (Hardin, 1968) (*see* **Commons, Tragedy of the**, Volume 5). It is also one of the powerful assumptions of the formal models used in economics to depict the tendency for resource users to over-exploit resources (Gordon, 1954; Scott, 1979).

As long as access is open, there are no incentives for resource users to take care of the resource because of the temptation to free-ride on the efforts of others. An open access system is also one where present users cannot really control the behavior of others, who can come and go as they please, reducing the benefits of individuals' efforts to conserve. Open access is accompanied by other special assumptions, such as the self-interested, profit-maximizing individual and the paradoxical assumptions of complete information on the one hand and lack of communication on the other (McCay and Acheson, 1987; Feeny *et al.*, 1996). Recognition of the limitations of these and other assumptions lies behind skepticism about the universal appropriateness of prescriptions that follow from the open access model: convert the system to either a private property or a state property (governance) regime.

Private property and state property are the regimes usually identified as solutions to the tragedy of the open access commons. Here, state property does not refer to the things and places owned exclusively by a government (like a defense department's munitions depot) but rather to government's ownership of the right and means of regulating or governing the use and distribution of common or public resources as well as ownership of lands in trust for the public. It is better to refer to it as state governance (Berkes and Feeny, 1990; McCay, 1996). According to economic property rights theory, over-exploitation and economic difficulties created by open access regimes require the strong arm of state governance because users of shared property cannot be counted upon to regulate themselves. Among the problems with this conclusion is over-reliance on the state as a homogeneous, benevolent, all-knowing entity interested in and capable of acting on behalf of long-term as well as short-term interests. This view of the state or its regulatory authorities is implicit in economic property rights theory, but it rarely fits the evidence of historical and contemporary reality (Feeny *et al.*, 1996).

Even better, according to the theory, is reliance on private property (especially in the liberal notion of exclusive ownership of most or all of the sticks in the bundle of rights). The idea is that private owners will have the incentives that public users lack to properly manage resources. However, the argument that private property is superior policy is faulty on logical as well as empirical grounds (Ellickson, 1993; Christman, 1994). It also rings false in theoretical circles. Under conditions of high rates of discounting the future and low rates of reproduction of the resource, a private owner is no less likely than a common user to over-exploit the resource (Clark, 1973). Moreover, some resources cannot easily be treated as private property: their physical features may be such that it is difficult to exclude others from them, and even if one did exclude others from part of them, what one does may affect the benefits that another realizes. They are common pool resources; any property rights asserted over them are necessarily imperfect, and the result is market failure. Hence, there must be alternatives.

The metaphor of an open access tragedy and the prescriptions of privatization and government intervention are often used in discussing environmental problems. They are linked to policy debates between command and control and market-based environmental control measures. Command and control refers to state governance (borrowing from criticisms of the socialist and communist governance systems). Market-based environmental protection refers to the creation of quasi-private property rights within the framework of state governance in order to improve compliance and efficiency while achieving social goals. Examples include air pollution systems that use tradable emission permits and fisheries management systems using transferable individual quotas. Tradable emissions permits were used to phase out lead in gasoline and reduce sulfur emissions in the US, providing greater flexibility and thereby economizing (Hahn and Hester, 1989). Similar programs have been used to eliminate ozone-depleting gases and to control acid rain and tropospheric ozone in the US. Canada, Chile, and Germany are other countries using tradable permits (Tietenberg, 1995, 1998) (*see* Tradable Permits for Greenhouse Gases, Volume 4; Tradable Permits for Limiting Stratospheric Ozone Depletion, Volume 4; Tradable Permits for Sulfur Emission Reductions, Volume 4; Transferable Fisheries Quotas (TFQs), Volume 4; Quotas in International Environmental Negotiations, Volume 4).

In marine fisheries, annual quotas for how much may be caught without jeopardizing the sustainability of a fish stock may be allocated to individual participants, just as pollution emission rights are allocated to individual polluting firms after a permissible level is established in relation to legal and scientific criteria (National Research Council, 1999). A recent review of these individual fishery quota (IFQ) systems in the US, Canada, Iceland, and New Zealand highlights the many issues, costs, and benefits of privatizing rights to shares of a quota. Reviews such as this have led to consideration of how to balance the benefits of market-based allocations with the concerns of individuals, communities, and industries for the future of their jobs and investments, as well as the broader societal concern with sustainable development. This effort includes attempts to develop market-based instruments that increase the conservation incentives for holders of exclusive rights while dealing with problems of equity, stewardship, and resilience (ability to adapt to changes in the environment) (Young and McCay, 1995).

Largely neglected thus far in the policy deliberations of major countries is the role of common property governance. Property rights scholars now object to the use of the term commons as equivalent to open access because it ignores the historic and contemporary importance of institutions in

which people share rights of access and use and have some degree of regulation of their activities on the commons (Ciriacy-Wantrup and Bishop, 1975; Cox, 1985; McCay and Acheson, 1987; Hanna, 1990; Bromley, 1992). Experimental and comparative case studies have helped create an understanding of the internal and external conditions under which common property institutions develop and are sustainable as well as those under which they falter (Ostrom, 1990; McKean, 1992).

Examples of external and internal conditions that promote community success in commons management include the ability to clearly define and regulate entry to membership in the community. It is also important to have autonomy from outside governance, to enable local experimentation and adaptation to local conditions, but to obtain enough legitimization and other assistance from the outside to help members of a community monitor the conditions of the resources or environments about which they care, so that they have reliable and accurate knowledge about those conditions. Local communities may also need external help to deal with spillover effects among communities. Critical internal factors include the ability to achieve broad participation in decision-making and implementation even where there are significant differences among participants; to create consensus about the fairness and rightness of the regulations; and the ability to monitor behavior and outcomes and to enforce and appropriately modify rules and regulations. Some factors pertain to the features of the environment, such as whether the flow of resources or risks from it is predictable or unpredictable and how abundant those resources or dangerous those risks are (cf. Ostrom *et al.*, 1999). Ironically, external conditions that promote community failure in these and other regards include the perverse effects of the very forces promoted as solutions to commons problems: intervention of the state and the workings of markets (McCay and Jentoft, 1998).

An emphasis on common property governance leads to three questions for broader understanding of global environmental change. The first is how to develop effective community-based systems for environmental protection and management. Community-based management has become a priority for international lending agencies and, even more so, non-governmental organizations in many industrialized and less-developed countries. However, it is not easy to implement, particularly where communities and the environmental resources with which they interact have been disrupted by warfare, migration, and disease, as well as long histories of subjugation and control by colonial powers, corrupt or unreliable political regimes, and economic restructuring due to international debt burdens. As much to the point is the poorly defined status of community property rights in most nations.

Second is the question of what can be learned from community-level studies that can inform efforts to improve international-level environmental regimes, the scaling up challenge. Important work is being done in this regard by students of international environmental regimes (Young, 1999, 1997; Haas *et al.*, 1993). Beyond the borders and boundaries of national jurisdiction, one might presuppose open access, which is supported by the international legal and political framework accepting the nation–state as the top level of sovereignty. Consequently, by law it is up to the nation–state to regulate uses of natural resources and their environments even in international realms, including the oceans (beyond 200 nautical miles from coastal borders) and the atmosphere and true frontiers (i.e., Antarctica). The international regimes that have arisen have some of the features of community-based common property regimes, such as reliance on trust which is built upon prior experience and interaction. Boundaries are necessarily more porous and enforcement a significant problem, as is shown in Canada's actions in the mid-1990s to protect straddling stocks of fish which migrate between Canadian and international waters. But communities do develop, based on shared knowledge about scientific findings as well as commitments to international solutions to environmental problems. More work needs to be done on whether or not the communities which evolve in these international diplomacy and scientific settings, so-called epistemic communities (Young, 1994), are appropriate structures for the development of common property regimes, or whether they will remain weakened versions of separate – but equal – state governance regimes.

The third question is how the interests and concerns of communities can be better incorporated into the policies and actions of governments and of private persons and firms. This refers to old but still relevant questions concerning social impacts and participatory management (also known as democracy in action). These are questions about the role of civic society in shaping the world, the earth, and the globe for tomorrow. All property regimes for dealing with environmental problems require viable civic institutions.

REFERENCES

Berkes, F and Feeny, D (1990) Paradigms Lost; Changing Views on the Use of Common Property Resources, *Alternatives*, 17(2), 48–55.

Bromley, D W, ed (1992) *Making the Commons Work: Theory, Practice and Policy*, Institute for Contemporary Studies, San Francisco, CA.

Bromley, D W and Cernea, M M (1989) The Management of Common Property Natural Resources: Some Conceptual and Operational Fallacies, *World Bank Discussion Paper 57*, The World Bank, Washington, DC.

Christman, J (1994) *The Myth of Property: Toward an Egalitarian Theory of Ownership*, Oxford University Press, New York.

Ciriacy-Wantrup, S and Bishop, R (1975) Common Property as a Concept in Natural Resources Policy, *Nat. Resour. J.*, 15, 713–727.

Clark, C W (1973) The Economics of Over-exploitation, *Science*, 181, 630–634.

Cox, S B J (1985) No Tragedy on the Common, *Environ. Ethics*, 7, 49–61.

Ellickson, R C (1993) Property in Land, *Yale Law J.*, 102(April), 1315–1400.

Feeny, D, Berkes, F, McCay, B J, and Acheson, J M (1990) The Tragedy of the Commons: Twenty-Two Years Later, *Hum. Ecol.*, 18(1), 1–19.

Feeny, D, Hanna, S, and McEvoy, A F (1996) Questioning the Assumptions of the Tragedy of the Commons Model of Fisheries, *Land Econ.*, 72(2), 187–205.

Gordon, H S (1954) The Economic Theory of a Common Property Resource: The Fishery, *J. Political Econ.*, 62, 124–142.

Haas, P, Keohane, R, and Levy, M (1993) *Institutions for the Earth: Sources of Effective Environmental Protection*, MIT Press, Cambridge, MA.

Hahn, R W and Hester, G L (1989) Where Did All the Markets Go? An Analysis of EPA's Emission Trading Program, *Yale J. Regul.*, 6(1), 109–153.

Hanna, S (1990) The Eighteenth Century English Commons: a Model for Ocean Management, *Ocean Shoreline Manage.*, 14, 155–172.

Hardin, G (1968) The Tragedy of the Commons, *Science*, 162, 1243–1248.

Macpherson, C B (1978) *Property: Mainstream and Critical Positions*, University of Toronto Press, Toronto.

McCay, B J (1996) Common and Private Concerns, in *Rights to Nature: Ecological, Economic, Cultural and Political Principles of Institutions for the Environment*, eds S Hanna, C Folke, and K-G Maler, Island Press, Washington, DC, 111–126.

McCay, B J and Acheson, J M, eds (1987) *The Question of the Commons: The Culture and Ecology of Communal Resources*, University of Arizona Press, Tucson, AZ.

McCay, B J and Jentoft, S (1998) Market or Community Failure? Critical Perspectives on Common Property Research, *Hum. Organ.*, 57(1), 21–29.

McKean, M A (1992) Success on the Commons: a Comparative Examination of Institutions for Common Property Resource Management, *J. Theor. Politics*, 4(3), 247–281.

National Research Council (1999) *Sharing the Fish: Toward a National Policy on Individual Fishing Quotas. Committee to Review Individual Fishing Quotas, Ocean Studies Board*, National Academy Press, Washington, DC.

Olson, M (1965) *The Logic of Collective Action*, Harvard University Press, Cambridge, MA.

Ostrom, E (1990) *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge University Press, New York.

Ostrom, E, Burger, J, Field, C B, Norgaard, R B, and Policansky, D (1999) Revisiting the Commons: Local Lessons, Global Challenges, *Science*, 284(April 9), 278–282.

Rose, C M (1994) *Property & Persuasion; Essays on the History, Theory and Rhetoric of Ownership*, Westview Press, Boulder, CO.

Scott, A (1979) Development of the Economic Theory on Fisheries Regulation, *J. Fish. Res. Board Can.*, 36, 725–741.

Tietenberg, T H (1995) Design Lessons from Existing Air Pollution Control Systems: The United States, in *Property Rights in a Social and Ecological Context: Case Studies and Design Applications*, eds S Hanna and M Munasinghe, The World Bank, Washington, DC, 15–32.

Tietenberg, T H (1998) Ethical Influences on the Evolution of the US Tradable Permit Approach to Air Pollution Control, *Ecol. Econ.*, 24(2,3), 241–257.

Young, M D and McCay, B J (1995) Building Equity, Stewardship and Resilience into Market-Based Property-Rights Systems, in *Property Rights and the Environment: Social and Ecological Issues*, eds S Hanna and M Munashinghe, The Beijer Institute of Ecological Economics and The World Bank Washington, DC, 87–102.

Young, O R (1994) *International Governance; Protecting the Environment in a Stateless Society*, Cornell University Press, Ithaca, NY.

Young, O (1997) *Global Governance: Drawing Insights from the Environmental Experience*, MIT Press, Cambridge, MA.

Young, O, ed (1999) *Science Plan for Institutional Dimensions of Global Environmental Change*, International Human Dimensions Programme on Global Environmental Change, Bonn, Germany.

Psychology/Perception, Environmental

see Environmental Psychology/Perception (**Volume 5**)

R

Reflexivity

see Social Science and Global Environmental Change (*Opening essay*, **Volume 5**)

Refugees, Environmental

see Environmental Refugees (**Volume 4**)

Relating to Nature: Spirituality, Philosophy, and Environmental Concern

see Relating to Nature: Spirituality, Philosophy, and Environmental Concern (*Opening essay*, **Volume 5**)

Religion and Ecology in the Abrahamic Faiths

Rosemary Ruether
Garrett-Evangelical Theological Seminary, Evanston, IL, USA

All the major world religions have been responding to the ecological crisis in recent decades by examining their traditions for resources to promote a more sustainable relationship between humanity and nature. In 1998 1999 the Center for World Religions at Harvard University assembled ten global conferences on the relationship of ecology and ten different world religions. Major gatherings of scholars examined the resources for ecological spirituality and ethics in Buddhism, Confucianism, Shinto, Hinduism, Judaism, Christianity and Islam, Jainism and Taoism, and indigenous traditions. Papers from these conferences are being published as books. This series will provide a major library on the issues of ecology and world religion.

This article focuses on the three Abrahamic faiths, Judaism, Christianity and Islam. These faiths differ from Asian or indigenous religions in being not only monotheistic, but also imaging God solely as male. They are also marked by a strong separation between God as Creator and Nature as Creation. These faiths were challenged some years ago when the historian of science, Lynn White, wrote a widely read article in which he charged that the biblical mandate in Genesis 1:28, giving humanity absolute dominion over the earth, was the major religious root of the ecological crisis (White, 1967). In this text God is represented as saying to the newly created humans, fill the earth and subdue it, and have dominion over the fish of the sea and over the birds of the air and over every living thing that moves upon the earth . The language used in this verse suggests a power relationship of military conquest and forcible subjugation by humans of the rest of nature.

Both Jewish and Christian scriptural and theological scholars have been engaged in refuting White s claims that the Bible as a whole mandates such an unlimited and exploitative relationship to nature as is suggested in *Genesis*. A number of Muslim writers have also taken up this issue of the view of human power over nature as re ected in the Qur an, itself rooted in Hebrew Scripture.

Jewish scholar Eric Katz has pointed out that *Genesis* 1:28 can hardly mean absolute and unlimited power over nature since the very next verse, Genesis 1:29, mandates a vegetarian diet for humans and animals. It is only after the corruption of society and God s punishment of humanity by the ood, that humans are allowed to kill and eat animals. But this is clearly seen as a fall from an original ideal of a more peaceful world (Genesis 9:2 4), but even then there are limits to human use of animals. Animals are to be killed in such a way as not to eat the blood.

Ismar Schorsch, writing on the Jewish view of ecology, suggests that the picture of humanity in relation to nature found in Genesis 2 3 is more the normative one for the

Bible. Here humans are seen as created from the topsoil of the earth, and as having affinity with the land from which they are made. They are not allowed to eat anything they wish in the garden. When they violate this commandment and seek to seize the fruit of the tree of knowledge of good and evil, human status degenerates. They are made to till the earth by the sweat of their brow. Hard struggle to survive, not lordly dominion, is the actual lot of humans in history.

For both writers the basic view of humans and nature in the Hebrew Bible is rooted in a theocratic understanding of creation. God is the creator of nature, and God alone has sovereignty over it. In the words of Psalm 24: The earth is the Lord's and the fullness thereof, the world and those who dwell therein . The laws of Jubilee clearly state that God remains the owner of the earth: The land shall not be sold in perpetuity, for all the land is Mine, because you are strangers and sojourners before Me (Lev. 25:23). Humans have only delegated authority over the earth, as representatives or stewards of God who remains its owner. This role of stewardship mandates both responsible care and limited use of the earth.

This view of the God–human–nature relationship is reflected throughout Hebrew law, both in the Bible and in its talmudic development. The Hebrew Bible reflects an agricultural people, and many of the divine commandments express a basic concern to limit human use of the land, its plants and animals, and to prevent abuse. The raising of goats and sheep is limited because of their excessive impact on the land. Grape vines and live trees are not to be used in sacrifice, because they are needed for food. Open land is to be preserved around towns. Trees are not to be cut down in war. Nothing is to be wantonly destroyed. Places of human habitation are to be separated from places of work that create smoke, dust and smells. Human sewage should be buried.

Perhaps the central Jewish ethic of ecological sustainability is found in the sabbatical legislation. This legislation proscribes a series of periodic cycles, seven days, seven years and seven times seven years, the ultimate cycle of the Jubilee.

Work is limited to a six-day period, following the example of God in creating the world. On the seventh day, not only the farmer, but his human and animal work force are to rest: On the seventh day you shall rest, so that your ox and your donkey may have relief, and your home born slave and resident alien may be refreshed (Ex. 23:12).

In the 7-year cycle, care is given to the poor, the wild animals and the renewal of the land:

> For six years you shall sow your land and gather its yield; but the seventh year you shall let it rest and lie fallow, so that the poor of your people shall eat and what they leave, the wild animals may eat. You shall do the same with the vineyard and your olive orchard.
>
> (Ex. 23:10–11)

Hebrew slaves are to be set free without debt in the seventh year (Ex. 21.2). Levitical laws mention slaves, hired laborers and animals allowed to rest during the sabbatical year, living off the untilled land along with the farmer (Lev. 25:6–7).

The Jubilee year, that occurs every fifty years, is to see a yet more profound restoration of the rights of all humans with respect to the land and animals. Those who have fallen into slavery are to be released. The debts that have accumulated during the past forty-nine years are to be liquidated. The earth is to lie fallow, human workers, animals and land given rest. Land confiscated due to debts is to be restored. The vision here is a general righting of relations in order not to endlessly extend a system in which the rich get richer and the poor get poorer, a society of a few wealthy, slave-owning landlords and many slaves or indebted workers. Justice and sustainability are to be periodically restored through a cyclical permanent revolution .

Hebrew Scripture also makes clear that humans are not the mediators of all relations between God and nature. God has many direct relations with other creatures, animals and planetary spheres, which humans know not of. The smallness and limits of the human arena of knowledge and control is emphasized in texts such as the book of Job. God rebukes human arrogance with such words as:

> Where were you when I laid the foundations of the earth, when the morning stars sang together and all the heavenly beings shouted for joy? ••• Can you hunt the prey for the lion or satisfy the appetite of the young lions? ••• Is it by your wisdom that the hawk soars ••• is it at your command that the eagle mounts up and makes its nest on high.
>
> (Job. 38–39)

In Hebrew Scripture the non-human world is not dead matter . It is animate and has its own lively relationship with God, independent of humans. God is present throughout the natural world. Both natural blessings and disasters reflect God's presence, both in blessing and in judgment. The non-human world understands this, better than humans.

> But ask the animals and they will teach you; the birds of the air and they will tell you; ask the plants of the earth, and they will teach you, the fish of the sea will declare to you. Who among all these does not know that the hand of the Lord has done this? In his hand is the life of every living thing, and the breath of every human being.
>
> (Job. 12:7–10)

The Hebrew worldview neither sacralizes nature as divine in itself, nor secularizes it. Humans have an important, but limited role. Reality is theocentric, in a way that relativizes and limits human power and places all things under divine sovereignty, as its ultimate owner and continuing source of life. God's relation to each creature is personal and direct, and each creature responds to its creator with praise and joy.

The Islamic view of the God–nature–human relation has some differences from that of Hebrew Scripture, but it is rooted in the same basic understanding. Muslim tradition is concerned to emphasize God's absolute unity (against the Christian Trinity) and sovereignty over all creation. Islam rejected the Jewish tradition that God rested on the seventh day, after creating the world, since that seemed to suggest that God was finite, becoming weary and needing rest. Creation is not something that happened in a primordial moment and then was completed. Rather God continually creates. All creation, and everything in it, is dependent on God for its existence from moment to moment.

There is less emphasis on the limits of human rule, although the ban on pork and wine suggests that humans do not just have all of nature at hand for their consumption. Humans are exalted as the vice-regents of God. As such they are superior not only to the animals, but even the spiritual beings. The angels bow before the wonder of God's supreme creation, the Human. Yet the Qur'an (Koran) also sees all creation, and not just humans, as praising and giving thanks to God. All natural things are to be respected as fellow creatures.

The idea of human vice-regency itself implies that humans are not the ultimate owners of the Earth. The Islamic view, like the Hebrew, is theocratic. God alone is sovereign; humans hold a delegated power over the Earth and are accountable to God for their tenure. They will be ultimately judged as to how well or badly they exercised this stewardship. Thus the eschatological (end of time) emphasis in Islam on the coming judgment is not understood as an escape from the earth that allows humans to ignore its wellbeing. Rather it is the ultimate accounting before God, by which they will be held responsible for their stewardship.

The idea that God creates the world from moment to moment, that all things lie directly in the hand of God, also lends itself, particularly in the poetic and sapiential (wisdom) traditions of Islam, to a mystical sense of all created things as the signs of God's presence. Nature lies within God's surrounding presence. In that sense it is God who is the ultimate environment that surrounds and encompasses humanity, as in the Quranic verse: But to God belongs all things in the heavens and the earth, and he it is who encompasses (muhit) all things (IV:126). Although nature is not sacred in itself, it is sacred, or one might say sacramental, as signs and pointers to God, who is present in every moment as its sustaining creator.

As Islamic writer, Seyyed Hossein Nasr (*see* Nasr, Seyyed Hossein, Volume 5) puts it:

> In reality, humans are immersed in the Divine Muhit (encompassing environment) and are only unaware of it because of their own forgetfulness and negligence. ••• To remember God is to see Him everywhere and to experience His reality as al-Muhit. The environmental crisis may in fact be said to have been caused by the human refusal to see God as the real environment which surrounds and nourishes our life.

The Islamic tradition of sacred art forbids the human image lest this be confused with the representation of God. In this sense the human face and figure, even if God's vice-regent, is never to be used to image God. But Islamic art makes exuberant use of nature, particularly flowers and vines, and the Arabic letters of the Qur'an, as pointers to God's presence.

Nasr sees the Muslim world as having failed to cultivate an ecological spirituality and ethic in modern times, rooted in this Islamic vision of God as the ultimate environment of life, because of its poverty and colonization by the West. This colonization of Muslim lands, as in other regions, such as Asia and Africa, has resulted in a conflictual relation to Western power and technology, which is the main source of ecological devastation in the modern world. For Nasr, Christians too are part of a revelatory tradition in which God is present as the ultimate environment of human life. But they have replaced this faith with a view of Promethean (autonomous) man who has absolute and unlimited power over the earth, unrestrained by obligations to either God, fellow humanity or the rest of creation. It is this view of human autonomy that Nasr sees as an apostasy (renunciation) from God and as the root of the environmental crisis in the West that has infected the rest of the world through Western colonizing power.

Muslims have been torn between modernists, who have sought to emulate this power and technology of the West, without any critique of its effects in impoverishing the earth and the majority of human beings, and fundamentalists who have been resisting this Western influence, but have interpreted this resistance primarily in terms of restoring certain legal customs, such as female dress. The sapiential (wise) traditions of Islam have been ignored, and this has prevented the development of a deeper resistance to such devastation of the earth based on the vision of the relationship to God that is present in nature.

For Nasr the solution to the ecological crisis is, for Muslims and also other peoples, to recover those spiritual traditions that direct them to a proper relationship to God.

> Humanity cannot save the natural environment except by rediscovering the nexus between Spirit and nature and becoming once again aware of the sacred quality of the works of the Supreme Artisan ••• the solution to the environmental crisis can come about only when the modern spiritual malaise is cured and there is a rediscovery of the world of the Spirit, which, being compassionate, always gives of itself to those open and receptive to its vivifying rays. The bounty of nature and its generosity to the human race are proof of this reality, for despite all that we have done to destroy nature, she is still alive and reflects on her own level of being the love and compassion, the wisdom and the power, which belongs ultimately to the realm of the Spirit. In this crisis of unprecedented proportions, it is nature as God's primordial creation that will have the final say.

Nasr's words provide a profound challenge to Western Christians. What are the roots of an ecological spirituality and ethic in Christianity, and how have these been denied in the development of western technological society and its colonialist exportation? What is the connection between Christianity and this development of secularity, industrialism and colonialism? Are these developments simply a denial of the true Christian view of the nature–human relationship? It was common a generation ago to see secularity, industrialism and technology as the products of the Christian genius. But now that these are seen to have negative rather than positive impacts, many Christians have been questioning the association between Christianity and these western patterns of life.

Christianity is rooted in the Jewish tradition and shares with it the Hebrew Scripture. But the defining of this scripture as Old Testament also implies a reinterpretation of it as partially superceded. The parts of the Hebrew Bible that the Christian tradition has seen as superceded and no longer in force are particularly the legal codes which enshrine limitations on human use of nature; laws of kosher foods, sabbatical rest and avoidance of pollution.

After the first generation of Galilean Christianity, the early Christians were urban people without roots in rural life. Hence the Jewish legislation related to the renewal of the land made no sense to them. Christians became a people gathered from many ethnic groups and no longer related to a particular land and ethnic community. They defined the church as a new spiritual people of God who anticipated a heavenly kingdom beyond this earth, rather than as a renewed people and land within history of Jewish redemptive hopes.

Christianity also based itself in a realized messianism that understood Jesus of Nazareth as the incarnation of God. This violated the separation of God and humans central to Judaism. Both Judaism and Islam have strongly rejected this idea as idolatrous, as the divinizing of a human being. Christian redemptive hope in the early centuries spoke of God having become human so that humans could become God. But this did not lead, in classical or medieval Christianity, to a Promethean idea of the human as replacing God. This was because Christianity rooted itself in a concept of human fall and redemption that emphasized human sin and weakness and dependency on divine grace. Also, human redemptive hope did not focus on power over this present creation, but a transformation of creation into the Kingdom of God or the replacement of creation with a new spiritual creation beyond the present finite world.

These elements of denial of the body and the world in anticipation of a transcendent redemption outside of history have sometimes been seen as a major source of Christianity's negative relation to creation. Despising the body and the material world has been seen as leading Christians to neglect and abuse of nature. But there is not a clear relation between otherworldliness and the ecological crisis. The Christian view of nature was more complicated than is suggested by an anti-material stance.

Christianity in the second to third centuries contended with and rejected the most extreme forms of body-spirit dualism in the form of Gnosticism. The cosmological Christology that was developed by its leading anti-gnostic theologians, such as Irenaeus of Lyon, rooted redemption in the restoration and completion of creation, as body and spirit together.

Humans were made in the image of God and mandated to grow into the likeness of God by increasing union of the human with the divine will. But humans created a break between God's will and their own willfulness. God sent a series of revelations to return humans to the right path, culminating in the incarnation of Christ. Christ is the incarnation in human form of the divine Logos through which the world was created in the beginning. This is understood as renewing the ontological (the nature of) grounding of the whole creation, not just of humans.

In the incarnation of Christ, divine creative power permeates Nature in a deeper way so that the bodily becomes the sacramental bearer of the divine and the divine transforms the bodily into ever more spiritual and immortal form. This is seen as taking place in three stages. There is the era of the Church when Christians grow ever more unified with God through Christ in the sacraments. Next comes a millennial era when the whole of creation experiences peace and harmony in God. Finally there is the ultimate transformation of creation into an immortalized union with God.

In Irenaeus' words:

> For there is one Son who accomplished his Father's will and one human race also in whom the mysteries of God are wrought ••• the wisdom of God, by means of which his handiwork, confirmed and incorporated with his Son, is brought to perfection; that is, His offspring, the First-Begotten Word should descend to the creature ••• and that it should be contained in Him, and on the other hand that the creature should contain the Word, and ascend to Him, passing beyond the angels, and be made after the image and likeness of God.
> (Adv Haer.V.36.3)

Classical Christianity through the middle ages situated itself in a profound tension between a demonic and a sacramental view of creation. As cut off from and alienated from God, not only humanity, but the whole creation is distorted and becomes an instrument of evil. But as renewed in Christ, it is reunited with the divine Word, the creator, sustainer and redeemer of the universe. Seen through the eyes of Christ, the whole universe is sacramental; all things become signs that testify to the presence of God, its creator, and lead the seeker to contemplative union with God.

Christian monastic life is a good example of this tension between world denial and world renewal in Christianity. As a way of life that flees sexuality, the body and the world, it may seem to be the epitome of a Christian negation

of the goodness of creation. Yet in its main expressions, particularly in the Benedictine tradition, asceticism is not a rejection of the body and the earth, but rather a rejection of exploitation and excess, a return to simple living in harmony with other humans and creation. Even fasting and the vegetarian diet are seen as creating harmony of body and spirit, restoring the paradisiacal age before the fall into sin.

Monastic life is seen as representing the beginning of a restoration of the original harmony and justice. Communal property overcomes the exploitative division of God's creation into private property that sets up unjust divisions between poor and rich, destroying God's original intention that the earth be shared by all equally. The monk befriends animals who minister to him, overcoming the enmity between humans and animals that came about after the flood (Gen. 9:2–3). Monastic communities returned to a simple, subsistence form of life that pioneered sustainable agriculture at a time when the unsustainable system of military and slave-driven urban organization of late Roman life was toppling. Monastic communities were centers of literacy and education, as well as sources of service to society and to the poor.

Not surprisingly, the creation of new forms of Christian ecological community today that seek to withdraw from consumerism, create sustainable societies and witness against injustice in society often model themselves on new forms of monastic life. One of the Christian ecological movements in the United States seeking to create such ecological communities is *Sisters for Earth*, a network of Roman Catholic religious women who seek to reshape their lands, buildings and community life to make them centers of ecological living and learning.

The Calvinist Reformation and the Scientific Revolution in the 16th and 17th centuries represent key turning points in the Western concept of its relation to nature. In these two movements the medieval struggle between sacramental and demonic views of nature were recast. Calvinism dismembered the medieval sacramental view of nature. For Calvinism, nature was totally depraved. There was no residue in it of divine presence that could sustain natural knowledge or any relationship to God. Redemptive knowledge descends from beyond nature, in the form of the revealed word available only through scripture, as preached by the Reformers.

While Calvinism dismembered the sacramental view of nature, it reinforced its demonic universe. The fallen world, the physical universe, as human groups outside of the Calvinist Church, lay in the grip of the Devil. Even within the Calvinist Church, the threat of the Devil lurked. Women particularly were the gateway of the Devil. Among Protestants, Calvinists in Scotland and New England were the primary witch hunters.

The Scientific Revolution appeared to move in the opposite direction. It sought to exorcise the demonic from nature, reclaiming the universe as an icon of divine reason manifest in the laws of nature, knowable through empirical observation and mathematics. In the 17th and 18th centuries the more animist forms of natural science that sought to unify spirit and matter lost out to a strict dualism of transcendent intellect and dead matter, of Newtonian Physics and Cartesian Reason.

Nature was secularized. It was no longer seen as the theatre for a cosmic struggle between God and the Devil. Both demonic and angelic spirits were driven out of it. God was excluded from any active presence in nature to become the distant first cause and finally disappeared altogether as a scientific premise. Nature becomes dead matter in motion, moving obediently according to mathematical laws knowable to a new priesthood of scientists.

According to Roger Bacon (*Novum Organum* and *The Masculine Birth of Time*), due to Eve's disobedience humans lost their original dominion over nature. Through science, this dominion over nature will be restored to human hands. As the presupposition of divine ownership of nature disappears, humans, specifically the male elite of scientists and entrepreneurs, are understood as having unlimited, sovereign power over nature to do with it what they will. It is here that we find the birth of the Promethean man decried by Seyyed Hossein Nasr.

In Western society, the application of science to technological power and use of nature marched side by side with colonialism. From the 16th to the 20th centuries Western Europeans would appropriate the lands of the Americas, Asia, Africa and the Middle East and reduce their populations to slavery or low paid labor. The wealth accrued from this vast appropriation of land, its resources, and human labor would fuel ever new levels of technological development, transforming material resources into new forms of energy and mechanical work, control of disease, increasing speed of communication and travel.

With this expanding wealth and power, the western elite grew increasingly optimistic, imagining that this expanding technology would gradually conquer all problems of material scarcity and push back the limits of human mortality. Christian hope for millennial blessedness in the culminating era of history was translated into scientific technological terms as the ideology of endless progress.

However in a short three-quarters of a century this dream of infinite progress has turned into a nightmare. The medical conquest of disease, lessening infant mortality and doubling the life span of the affluent, insufficiently matched by birth limitation, especially among the poorest, has created a population explosion that is outrunning the food supply. The gap between poor and rich, between a wealthy, industrialized section of the world and the impoverished masses has grown ever wider in the last 25 years. Increasingly global society is being split between two worlds, those who enjoy the benefits of ever expanding

forms of technology, and those who have fallen out of the bottom of the modernized world into desperate poverty: reverting to pre-industrial means of subsistence or surviving on the garbage of the affluent world.

The western scientific and industrial revolutions have been built on injustice from the beginning. It has been based on a vast takeover of land, agricultural, forest and mineral wealth, appropriated through the exploitation of indigenous peoples. This wealth has flowed back to enrich the West, together with local elite, while impoverishing the land and the exploited people. This system of global affluence, with its high consumption of energy and its waste, cannot be expanded to include the poor without destroying the basis of life on the planet itself. We are literally destroying the air, water and soil on which human and planetary life depends.

Can this global system be changed to avert disaster, and do the world's religions have something to contribute to this conversion? I believe that religious traditions have both wisdom and motivational power to contribute. The three Abrahamic faiths, Judaism, Christianity and Islam, have important insights that can guide such a transformation. A key change in self-understanding is found in the belief that humans do not have absolute power of nature. Rather they have a delegated role for which they must be responsible to the ultimate source of life. Religious laws that mandate limited use of nature, and that understood humans, God and nature as standing in a covenantal relationship, calling for periodic restoration of justice and sustainable balances, can inspire us today.

Global movements, such as the Jubilee 2000 project to cancel the global debt of the poorest countries, and international covenants to limit toxic wastes, draw on this Biblical tradition. Efforts to create a global ethic accepted by all world religions, such as the one crafted by the meeting of the Parliament of the World's Religions in Chicago in 1994, seek to draw on religious traditions for ecological guidance. The development of a World Charter for Nature, that would parallel the International Declaration of Human Rights, is a similar effort to create a global ecological ethic based on all the world's moral traditions.

Religious traditions are critical to providing the vision, and the spiritual commitment, to make such a change. But these traditions are only beginning to overcome their ancient animosities toward each other and to translate their insights into practical guidelines to guide human development toward more just and sustainable ways of living.

FURTHER READING

Hessel, D and Ruether, R (2000) Christianity and Ecology, in *Religions of the World and Ecology*, eds M E Tucker and J Grim, Harvard University Center for the Study of World Religions, Cambridge, MA.

Katz, E (1994) Judaism and Ecological Crisis, in *World Views and Ecology*, eds M E Tucker and J Grim, Orbis Books, Maryknoll, NY, 55–70.

Nanji, A, Denny, F, and Baharuddin, A (2001) Islam and Ecology, in *Religions of the World and Ecology*, eds M E Tucker and J Grim, Harvard University Center for the Study of World Religions, Cambridge, MA, in press.

Nasr, S H (1992) Islam and Environmental Crisis, in *Spirit and Nature: Why the Environmental Crisis is a Religious Issue*, eds S C Rockefeller and J C Elder, Beacon, Boston, MA, 83–108.

Ruether, R R (1992) *Gaia and God: an Ecofeminist Theology of Earth Healing*, Harper, San Francisco, CA.

Schorsch, I (1992) Learning to Live with Less: a Jewish Perspective, in *Spirit and Nature: Why the Environmental Crisis is a Religious Issue*, eds S C Rockefeller and J C Elder, Beacon, Boston, MA, 25–38.

Sokol, M (2001) Judaism and Ecology, in *Religions of the World and Ecology*, eds M E Tucker and J Grim, Harvard University Center for the Study of World Religions, Cambridge, MA, in press.

Timm, R (1994) The Ecological Fallout of Islamic Creation Theology, in *World Views and Ecology*, eds M E Tucker and J Grim, Orbis Books, Maryknoll, NY, 83–95.

White, L (1967) The Historical Roots of our Ecologic Crisis, *Science*, 155, 1203–1207.

Religion and Environment Among North American First Nations

Freda Rajotte
Canadian Coalition for Ecology, Dauphin, Canada

The First Nations peoples of North America consist of several hundred distinct ethnic groups, each of which had a unique language and culture, and which, at the time of European colonization, had adapted differently to the particular physical environment in which they lived. The last 500 years have seen vast changes to both the physical and cultural (economic, social and political) environment. The adaptation of each tribal or ethnic group of First Nations peoples to these changes has been in uenced by many factors, including: the degree of persecution, suppression, assimilation and forced migration, size of reservation lands, their isolation or degree of proximity to urban centers and markets, the presence of minerals, timber, arable soils and other resources, and above all to the degree that they have been able to retain or reconstitute their own religious belief systems and values as distinct from the North American mainstream.

INTRODUCTION

At the time when European settlement in North America began in the 16th century the land was already occupied by a great variety of different ethnic and cultural groups that have been referred to collectively as *Natives*, *Aboriginals*, *Indians*, or the preferred term, which will be used here *First Nations*. Each group developed its own culture, language, life style and traditions, and each is the subject of numerous anthropological, social and economic studies. At least 300 different languages were spoken. Despite this diversity, it is possible to make some generalizations concerning life style and relationship with the physical environment.

Settlement of the North American continent began somewhere between 40 000 and 30 000 years ago, and scholars generally agree that it consisted of numerous groups migrating from Northern China and Northeastern Asia over a long period of time. They came via the Aleutian Islands, which formed a land bridge across the Bering Sea during the last Ice Age, between approximately 30 000 and 14 000 years BP (before the present), when moisture was locked up in glaciers and vast continental ice caps, and sea levels were lower.

Through thousands of years of migration and cultural evolution, every part of North America became settled. (While this article considers only the Canada and US mainland, migration lled Mexico, Central and South America as well.) As the continental glaciers withdrew, people hunted and settled on the land progressively revealed by the retreating ice. And everywhere, the people adapted their life styles and culture to the vastly different environmental conditions. From the arid lands bordering Mexico to the rain forests of the Paci c Northwest coast, to the Arctic Circle, tribal groups developed methods of hunting, shing, cultivation, house construction and social organization that uniquely suited the resources and the land.

In each area the life style and economy became adapted to the local environment. Major factors included the abundance of animals and sh; the type and density of vegetation, the fertility of the soil and ease with which it could be cultivated; the length of the growing season, and the abundance of materials (wood, adobe, hides, etc.) for the construction of housing. Above all the climate was a major factor in uencing clothing styles, housing types, and the seasonality of activities such as hunting local wildlife and migratory herds, and cultivation of crops.

This lengthy period of settlement by First Nations peoples certainly had an impact on the North American environment. However, the extent of this impact varied considerably according to the environmental region and cultural life style of the people and was very minimal compared to the impact of human activities during the 20th century. For example, it seems certain that the systematic use of re kept the grassland plains of the continental interior relatively free from forest encroachment, and undoubtedly increased the extent of these lands. In the arid Southwestern US, there is evidence of some pueblo settlements having been abandoned before the arrival of Europeans. It appears that the removal of trees from the surrounding hills had increased the aridity of the land used for cultivation and speeded its deterioration.

The environmental impact of people was restricted by the simple fact that if the people did not adapt to the cycles of the seasons, or if they depleted the animal or plant resources, they would die. The knowledge and skills needed for hunting, shing, farming, gathering medicinal herbs, storing food, raising children, building and carving, constructing houses and making clothing, boats and weapons, were passed on from generation to generation. Winter was generally regarded as the season for story telling and passing on myths of origin, tribal histories and knowledge concerning social roles and expectations. Stories, mythologies, rites, dances, ceremonies and rituals all served to keep the cultures alive and vital. Thus, over thousands of years, the First Nations peoples developed unique ways of living in relative harmony both within their social groups and with the natural world around them.

Traditionally northern groups in the taiga forests, Arctic tundra and barren lands, have followed a seasonal migration pattern, shing and sealing in the Arctic waters, and following the caribou herds in their annual migration across the Arctic plains. While some still follow this same pattern, almost all carry with them modern, high-powered, hunting ri es and steel traps. Many travel to hunt or sh by snowmobile, and depend to varying extent on supplies own in from the South. The use of traditional dog sleds or of kayaks is increasingly becoming a pleasure and tourist activity. Military activities and oil exploitation in the Arctic pose a serious danger to the continued existence and health of the remaining caribou herds.

In the heavily forested Northwest coastal regions, marine resources were abundant. Hunting and berry picking supplemented shing. Here, many communities had winter villages with substantial community and family lodges, and also temporary summer quarters where families moved at the time of the salmon migration up river to harvest and dry sh. With forests of giant trees at hand, and an abundance of food easily available, there was time and energy to build huge timber lodges, and giant ocean-going canoes. Even the decorative arts such as wood carving had reached an amazingly advanced and stylistic form.

In North America s vast interior plains, the buffalo herds once provided the basic support for the life of the people, supplying them not only with food, but also with materials for clothing, shelter and weapons. The buffalo was so important that it also became central to their mythology and rituals.

In the forested regions of Eastern America, the basic activity of hunting was supplemented by harvesting natural products, such as wild rice, maple sugar and berries and by cultivating gardens of squash, corn and tobacco.

In the dry southwest of America, in the semi-desert and high plateau regions, corn has traditionally been the basic staple. It is generally grown together with beans, squash, potatoes and a mixture of many other indigenous crops such as peppers and tobacco. Here, people are particularly dependent upon the rainfall, and many of the religious rites and ceremonies are related to rain and the successful harvest of the corn crop.

During this long period in pre-contact America, there was not only an extraordinary diversity of tribal cultures and languages, but also an equally diverse range of mythologies and religious ceremonies and rituals. From the seasonal chants and ceremonial dances of the Pueblo peoples of the southwestern deserts, to the shamanistic practices of the nomadic Inuit of the Arctic, from the totemic peoples of the Pacific Northwest coast to the buffalo hunters of the plains, or the lodge societies of the eastern woodlands, it was hardly possible to speak in general terms of a single Native or First Nations religion.

However, as noted by Clarkson *et al.* (1992) The life of people became a reflection of the life of the earth and our ancestors became intimately connected and inseparable from these realities. Through many years of experience, trial and error, hunger and hardship, our ancestors learned that the depletion of plant and animal life in their immediate environment meant starvation and death .

IMPACT OF IMMIGRATION

When Columbus landed in the Americas just over 500 years ago it has been estimated that there were at least 100 million people, or approximately one-fifth of the human race living there. Almost immediately on contact this population began to decline. This was partly due to warfare and to the superior weapons of the settlers; but, far more seriously, during the first hundred years or more it was due to exposure to European diseases.

The ships that followed Columbus in ever increasing numbers, brought not only explorers, soldiers, gold seekers, and missionaries, but also off-loaded domestic animals and plants, and even unwanted pests such as rats and fleas. Perhaps the most destructive of all the new arrivals were the European viruses for which the Native American people had no immunity. Wave after wave of epidemics of influenza, diphtheria, smallpox, measles, tuberculosis, chicken pox and cholera swept through the Americas, until;

> less than a tenth of the original population remained. Perhaps 90 million had died ••• It was the greatest mortality in history. (Wright, 1993)

When the first European settlers began to arrive on the east coast, it was the First Nation peoples who introduced them to many new things essential for their survival. These included canoes and snowshoes, unfamiliar wildlife such as the moose, beaver and wild turkey, and domestic crops such as tobacco, pumpkins, beans and corn. However, soon the tide of land-hungry immigrants grew in number and surged westwards across the continent, surveying and dividing up the land into private farms and townships.

The very different belief systems, and opposing views on the relationship between people and the land, underlay much of the tragedy of mutual misunderstanding and conflict between First Nation and immigrant peoples.

During the past 500 years, not only were the First Nations populations decimated and dispossessed of almost all of their lands, but European immigrants brought with them their own politico-economic views (predominantly capitalism, private ownership and profit) and their own religious systems (predominantly various Christian denominations). Clearly each could have learned much from the other, but sadly a relationship of great potential turned into a tragedy of mutual incomprehension. European settlers had no concept of *hierophantic* Nature, i.e., of Nature being an embodiment of the spirit world. Native Americans did not think in terms of personal ownership of sections of land, nor of hunting and killing for sport.

It is a mark of particularly obtuse religious blindness, that governments, which established religious freedom clauses in their constitutions, suppressed Native religious practices, while giving free reign to Christian missions and residential schools. Not only were Native traditional belief systems eroded or obliterated, but also their religious ceremonies, such as the sun dance, potlatch, and sweat lodge were made illegal. Sacred objects were stolen to be sold as curios or art objects, many finding their way into museums of anthropology around the world. Even graves were plundered and ancestral bones exhumed and taken for study or display.

The new arrivals saw the land as virgin and empty: free for the taking. It only awaited civilization and development for its forests, soils and minerals to yield vast riches. In the way, however, stood mountain ranges, impenetrable forests, vast plains with an unbelievably harsh climate, and the last remnants of the indigenous population. All were seen as challenges to be overcome, as obstacles to be conquered, or as hindrances to be swept away.

Throughout the 19th century, armed confrontation frequently took place between First Nations peoples and settlers. Armed with superior weapons, and often supported by military forces, the settlers inflicted heavy casualties, and pushed the remnant indigenous peoples onto small areas of reservation land. Treaties were made between various First Nations and Britain (in the case of Canada) and the American government, and were as quickly ignored

and broken by the settlers. Systematically, encroachment or annexation reduced or eliminated these reservation lands.

On several occasions, cart loads of blankets infested with diseases, such as smallpox, were used to wipe out entire Tribal groups. Some groups, such as the Beothuk peoples of Newfoundland, were eliminated by genocide. Other groups were so depleted that the few remaining families moved away, died out, or intermarried and merged into the larger population. Sometimes entire peoples were illegally relocated to poorer and more remote locations. In 1830, US President Jackson signed into law a bill that required all Indians living in the east (predominately Cherokee) to leave their homes and be relocated west of the Mississippi River. During the long trek, called the Trail of Tears, guarded by federal soldiers, and attacked by bandits and thieves, many died of exhaustion, sickness and starvation.

Only in the more remote areas of the Northern taiga forests, in the Arctic regions, and among the Pueblo peoples of the arid Southwestern US, were the people never conquered nor forcibly removed, and so were able to maintain their traditional way of life.

IMPACT OF 500 YEARS OF EUROPEAN SETTLEMENT

First Nation peoples had always seen the land as sacred, as given by the Great Spirit or Creator. Those peoples belonged to the land, were part of the sacred circle of life, and were charged with protecting and preserving it. The European settlers and colonists perceived the land in a totally different way: for them land was a commodity, it was property, wealth, and security, to be measured out and owned. The land belonged to them; they did not belong to the land! The land became a resource to be bought and sold, but above all to be developed, controlled, mastered and made to produce.

While the First Nation peoples were trying desperately to retain their identity and their unity with the land, the inexorable westward migration of settlers across the Americas was focused upon something very different – upon the acquisition of free land and resources (timber, minerals and fertile soils). All were migrants, uprooted peoples, and most were poor. Many of them had been evicted from their homes (e.g., during the Scottish Highland Clearance), or had been displaced by war. The dreams and efforts of such immigrants and pioneers were focussed upon security, ownership and wealth. Their interests lay in claiming and fencing-in their personal property, in establishing homes and farms, towns and transportation routes. They brought with them familiar domestic animals and crops from Europe, and set about establishing a civilization similar to, but they hoped better than, that of Europe. They saw North America as a vast unoccupied space, as a place of limitless opportunity and freedom.

So, throughout the last 500 years, the ecology of North America has been vastly changed. The economy has been transformed from subsistence to commercial. The sparse population and small settlements have been replaced by a large and growing population centered in urban and industrial agglomerations.

The vast herds of buffalo of the continental interior have been eliminated, often wantonly, and the way of life of the Plains First Nations peoples wiped out. The prolific seas of both the east and west coasts, with their vast, seemingly inexhaustible shoals of fish, and beaches teaming with marine mammals, have been sadly depleted. Also, the great salmon runs of the rivers of both the Pacific Northwest and of the Atlantic Northeast coasts have been either destroyed entirely by damming or greatly depleted by over-fishing, pollution, or water diversion for urban, agricultural or industrial use.

And while settlers cleared vast areas of forestland for cultivation, most of the remaining forested areas have been clear cut. In some areas they have been cut two or three times, to provide both lumber and raw material for the pulp and paper industry. Many species of animals and plants, once plentiful, are either extinct, like the passenger pigeon, or endangered as their habitat is destroyed or as they are relentlessly hunted and trapped.

In a mad rush after gold, the land could be ripped apart, and slag heaps left to mar and poison the landscape for generations. However, the early gold rushes in California and the Yukon were as nothing compared to the open cast coal, copper and iron ore mines, and the oil and gas exploitation, that were to follow.

A TIME OF RENEWAL AND HOPE

In the 19th century, government policies in both Canada and the US favoured forced assimilation. In the US, the Dawes General Allotment Act of 1887 resulted in another 86 million acres (35 million ha) of reservation land being lost and at least 90 000 indigenous people being left homeless. A battered and demoralized remnant population seemed to have little will or ability to oppose two of the most powerful governments of modern industrial states. Even in the 1950s discussions were still underway to consider the termination of federal government responsibility for reservation lands in the US, and the extinguishment of treaty rights and status in Canada.

Amazingly, in the early 1960s, a series of Native conferences and growing protest movements gathered public attention and increasing general support to rectify and make amends for such a vast injustice. Several startling military confrontation – Wounded Knee (1973, in North Dakota, US), and Kanehsatake (1990, in Oka, Quebec, Canada) – were widely covered by media and seen on TV screens around the world, underlining the urgency of the

situation. These events and numerous others, such as the highly publicized reoccupation of Alcatraz Island (off the California coast, formerly used as a prison), acted as a catalyst to First Nations peoples.

Miraculously, since the 1960s, there has also been a remarkable renaissance of traditional religious beliefs and practices. This renaissance underlies and sustains a huge surge in cultural pride that is reflected in everything from art and clothing, to restoring tribal languages, economic development, and filing land claims. Scholars have begun to take a more open view of Native religious traditions. Disillusioned by the destructiveness of western culture, some people have begun to look to First Nations' elders for a spiritual understanding that is more respectful of nature and of community values. This religious revival has created a widespread interest and even a following among the general population, a surprisingly large and growing number of whom may attend powwows, are eager to study traditional Native spirituality, and participate in some of the ceremonies.

RESPONDING TO THE MODERN ENVIRONMENTAL SITUATION

The adaptation of First Nations peoples to environmental change can only be understood in the context of the history of the last 500 years of European contact. For First Nations peoples, it has been 500 years of occupation, epidemics, and disinheritance, of broken treaties, stolen land, racism and increasing impoverishment. All too often this resulted in a legacy of despair, poverty, high unemployment, and the alcoholism and violence that are generated by these conditions. Today, approximately half of all those claiming First Nation or part First Nation ancestry are living in urban areas, have intermarried, or have been assimilated into the general population.

Because of racism and discriminatory government policies, very few Native people not on the reservation lands are independent farmers.

On the reservations, a different and detailed answer would have to be given in each case, according to factors such as: the size of the land base, size of the population, degree of forced relocation, traditional skills, proximity to urban areas and markets, resource base and climate zone, degree of political autonomy, and educational level of the population.

In 1975 the crucial Self-determination Act was passed in the US and 150 groups immediately filed for official tribal status. Many more applications are currently being filed. Each recognized tribe enjoys a direct government-to-government relationship with the US, and can establish its own government, tribal court, police and educational system. Freed from federal and state regulations, the tribal governments can and do turn reserve lands into anything from nuclear waste dumps, to casinos and resort developments. In Connecticut, the Mashantucket-Pequot now own one of the largest casinos in the world employing over 9000 people. By the late 1990s, more than 160 US tribes and several reserves in Canada (e.g., Rama, Ontario) were operating casinos.

In Canada almost every reserve has established its tribal land claim office, and is preparing for, or has already filed, a land claim before the relevant government (provincial, territorial or federal). Some of the first substantial land claims to have been heard have been upheld by the courts.

RESTORING ANCIENT TEACHINGS IN A MODERN CONTEXT

After nearly 500 years of cultural suppression, many First Nations people care very little about spirituality or nature, or about traditional ceremonies. In many communities the majority may be adherents of some Christian denomination, or of the Native American Church (the most widespread religion among US Native peoples), which combines some fundamentalist Christian beliefs with Native teachings. Its most important sacrament involves the eating of the buttons of the peyote, a hallucinogenic cactus (still classed as an illegal act in the US).

Among the First Nations of Southwestern America, the Ute, Apache, Navajo and other Pueblo nations who were never conquered or displaced, many of the traditional teachings and rituals, healing chants and sand paintings, as well as the seasonal dances with katchina masks, remain virtually unchanged and intact. Similarly the Iroquois ceremonies conducted in longhouses in northeastern US, and closely tied to the agricultural seasons, appear to have changed very little.

Where fundamental rituals and ceremonies were banned and made illegal, many tribal societies are now carefully researching and reestablishing them. For example, the Potlatch feast, with its masked dancers and lavish give-away ceremonies of the Pacific Northwest coast, associated with the Kwakiutl, Nootka, Haida, Tsimshian, Tlingit and Salish societies, was declared illegal from 1884 until 1951. It has now been reinstated, with an associated revival of master carvers preparing spectacular dance masks, totem carvings, and large sea-going canoes. Today many young people learn to dance and act the ancient stories and legends of their ancestral heritage. This is paralleled on the central plains, by the reestablishment of the annual sun-dance ceremony on many reserves, and the almost ubiquitous sweat lodges, which are constructed at almost all powwows, ceremonies and local celebrations. Its use has become popular with non-Native peoples as has smudging, which involves carrying a braid of smoldering sweetgrass around the circle

and offering it to each person so that they may cleanse themselves in its smoke.

A sweat lodge (or *impi*) is a small circular dome of willow branches, half sunk into the earth, and always directionally oriented, into which the shaman or elder brings heated rocks. In the dark interior of the lodge, water is poured onto the rocks producing steam. Participants experience a primordial microcosm of creation, similar to reentering the womb, constituted of earth, air, fire and water. It is a ceremony of cleansing and rebirth. The Iroquois Mide peoples and the Kivas of the Pueblo peoples use a similar symbolism.

At the other extreme are numerous communities where ancient beliefs and traditions are irrecoverably lost, where remnant populations were displaced from their lands and forcibly moved, some many times over, to make way for European settlers. While much has been lost, an astonishing amount remains or has been reconstituted. Today native people travel widely and many join in the powwow circuit, traveling from one community powwow to the next. Research and information are communicated rapidly by email, fax and phone, so that similar customs and teachings spread easily over a wide area. Native Spirituality is a recognized university course in many Native Studies departments, and many elders give seminars or otherwise make their teachings available to the general public. Today, it even seems possible, while acknowledging the diversity of ceremonies and rituals, to make certain generalizations about First Nation's faith and belief systems.

FAITH AND BELIEF SYSTEMS AMONG FIRST NATIONS

All First Nations traditions recognize the interconnectedness of the human, natural, and spirit worlds. The inclusiveness of the circle or Sacred Hoop has become the almost universal symbol for Native spirituality. It encompasses all the directions (east, north, south and west), all four seasons (Spring, Summer, Autumn and Winter), and all stages of life (infant, youth, adult and elder). It can be used as a basis of prayer and thanksgiving to offer thanks to all four directions, all kinds of weather, and all aspects of life – for all times of day: dawn, daylight, sunset and night. It can be used as a guide to the correct building and sighting of a pueblo, lodge or teepee. It is the symbol for completeness and inclusiveness of the whole community, and the whole of the interrelated creation:

> You have noticed that everything an Indian does is in a circle, and that is because the power of the world always works in circles, and everything tries to be round.
>
> In the old days when we were a strong and happy people, all our power came to us from the sacred hoop of the nation, and as long as the hoop was unbroken, the people flourished. The flowering tree was at the center of the hoop, and the circle of the four quarters nourished it. The east gave peace and light, the south gave warmth, the west gave rain, and the north with its cold and mighty wind gave us strength and endurance. This knowledge came to us from the outer world with our religion. Everything the power of the world does is done in a circle. The sky is round, and I have heard that the earth is round like a ball, and so are all the stars ••• Even the seasons form a great circle in their changing, and always come back again to where they were. ••• Our teepees were round like the nests of birds, and these were always set in a circle, the nation's hoop, and nest of many nests, where the Great Spirit meant for us to hatch our children.
>
> (Black Elk Speaks, 1988)

In contrast, European settlers superimposed straight lines and grid patterns upon the land. Railways and roads cross the plains in straight lines for hundreds of km; property is divided up into rectangular parcels.

There is a common awareness that life is spiritual: everything that exists has within it an element of sacredness. Nature, in its entirety, is *hierophantic* that is, a source of spiritual revelation. Everything – the land, water and sky, sun, moon and stars, plants, animals, birds, fish and people – all manifest some aspect of the spiritual reality that infuses the whole of creation. This basic view of the world means that Mother Earth (the major source of all life) must be treated with respect, not viewed as a commodity to be owned, or bought and sold. The physical environment should only be used with care and reverence. A successful hunt or a good harvest was accepted with thanksgiving and the food was shared among the community so that no one was in want. Neither does Nature exist solely for human use. All life has access rights to the use of land, water and air.

Ceremonies among First Nation North Americans generally center around the drum. The drum may be referred to as the heartbeat of Mother Earth, and forms a connection to the spirit world. Around the drum with its circle of drummers and singers, the dancers move in a large circle, feet close to the earth, and moving to the rhythmic beat and chants.

First Nations peoples respect animals as sacred. Traditionally, prayer and ritual accompanied hunting, which was seen as a sacred activity, only undertaken when food was needed. Care was taken to give thanksgiving following a successful hunt, and to respect the animal that had been killed to provide food and materials for clothing, etc. Great care was taken not to over-hunt or endanger a species' existence.

However, from the earliest days of contact with Europeans, Natives were encouraged to hunt and trap for furs and pelts for trading. Today, many, especially in the northern taiga forests, still earn a significant income from running trap lines or hunting, but snowmobiles are generally used to reach trap lines, and high-powered rifles for hunting.

This has caused controversy over the relationship between traditional practices and hunting rights claims.

For each clan and people, certain creatures have a greater spiritual significance. Individuals may undertake a vision quest (several days of solitary fasting and prayer in a wilderness area) during which time they may receive a revelation from the spirit world through a specific creature, which will then become their guardian spirit. People enter a special relationship with their guardian animal, never hunting or harming it in any way, but seeking its presence, assistance and wisdom. Through their sacred animal (eagle, buffalo, bear, snake, raven or other creature), the spirit-being manifests itself as a personal or as a clan protector. Some claim that they receive warnings or guidance or even that their lives have been saved during a crisis such as sickness, fire or a car crash, by their guardian animal.

The totem carvings of the Northwest Pacific coast are like heraldic crests. Each pole recounts a specific clan history and mythology, recounting the stories of the relationship of clan elders with guardian and sacred animal spirits (actual animals like the raven or whale) or mythological ones (like the thunderbird). That creatures embody the invisible spiritual realm is the profound meaning underlying sacred mask dances, e.g., the seasonal cycles of katchina dances of the Southwestern Pueblo peoples. In these dances, the spiritual reality becomes embodied in and takes possession of the physical form of the mask and the dancer so that the dancer becomes temporarily the actual manifestation of the ancestral or animal spirit. (This is similar to claiming the actual presence of Christ in the elements of the Christian mass.)

Shamans or medicine men or women are central to the traditional belief systems and rituals of First Nations peoples throughout North America. They are highly respected as intermediaries between the spiritual and material worlds, and can also effect healing. Through solitude, communing with nature, fasting, prayer, purification and self-discipline, shamans undertake journeys into the spiritual realm, where they seek help and assistance, wisdom and healing power. Revered shamans and elders are recognized and honored by becoming pipe bearers. Smoking the sacred pipe is undertaken to bring the smoker into close contact with the spirit world.

While some Shamans may be trained in western medicine or nursing, all have a profound knowledge of the medicinal properties of local herbs. They are concerned not only with physical healing, for which many patients may go first to a local clinic or hospital, but also with holistic healing of body, mind and spirit. They can be effective both with individuals and with communities that are experiencing high levels of unemployment, poverty, despair, addiction, violence and soaring youth suicide rates.

Native existence itself is deeply and inseparably rooted in the land – in Mother Earth. The land belongs to all creatures that live upon it. A people's traditional economy, culture and self-image are based upon the continuous habitation of place . Myths of origin, cultural histories, and self-perception, as well as rituals and sacred sites are located in and around a tribe's traditional lands. It is alone, in nature, that one goes on a vision quest to seek wisdom. Mother Earth nurtures us, feeds us, inspires us, and it is back into the land that our bodies are returned at death. The earth is sacred. People and place together form a sacred cultural unity. While all of nature is the theatre for spiritual revelation, these specific sacred sites are incorporated into tribal mythologies and rituals. Each site is affiliated with a specific local mythology and tribal history. Thus, the relocation of a tribal group or destruction of the site means that essential links to the spirit world are lost. To understand this is to understand why the conquest and removal of people from their traditional lands was cultural genocide.

Throughout North America specific landforms, such as springs, mountains or high bluffs are regarded as especially sacred. Here spiritual elders and shamans commune with the mystery of the spirit realm and receive visions, revelations and healing power. Sacred sites may be enhanced with petroglyphs, medicine wheels (marked out in rocks), or earth mounds (such as the Great Serpent Mound of Ohio). For pueblo peoples, kivas – subterranean rock-built chambers – mark the site where the original peoples emerged from the underworld. While most sites have been defaced or even completely destroyed over the last 500 years, attempts are now being made by many First Nation peoples to protect some of the remaining sites under constitutional freedom of religious practice clauses.

George Tinker, a First Nation author from Colorado explains that:

> Native American spirituality and values, social and political structures, and even ethics are rooted not in some temporal notion of history but in spatiality. This is perhaps the most dramatic (and largely unnoticed) cultural difference between Native American thought and western intellectual tradition. The question is not whether time or space is missing in one culture of the other, but which is dominant. Of course, Native Americans have a temporal awareness, but it is subordinate to our sense of place. Likewise the western tradition has a spatial awareness, but it lacks the priority of the temporal. Hence, progress, history, development, evolution and process become key notions that invade all academic discourse in the West.
>
> (Tinker, 1992)

The inner, spiritual forces that exist within all creatures, continue their never ending journey after the death of the outer physical form, and they may continue for a while to have a relationship with living beings, and influence events. Many ceremonies fulfill the function of mediation between the spirit world, the human world and the world of nature, for a continued and reciprocal relationship is perceived between the living and the dead. These ceremonies may show respect, offer thanksgiving for good fortune or

seek blessing upon future endeavors, e.g., successful hunts, abundant rainfall or healing.

Care must be taken that we do not harm nature. Many of the decisions made today by proprietors, corporations and governments take a very short-term view, and seek immediate profits. This shortsighted approach means that as governments change, and corporations merge or move, there is no accountability when the negative effects of industrialization policies become apparent many years or even decades later. In comparison, traditionally, when Native bands consider development options, they consider what the likely impact will be, not only for the immediate future, but also upon their children and children's children to the seventh generation.

TODAY AND TOMORROW

Throughout the 1990s and into the 21st century a resurgence of epic proportions has swept through the First Nation societies of North America, as they seek to free themselves from control and domination by colonial powers.

The recent resurgence of Native spirituality, while causing some social con ict, is generally leading to increased cultural pride. Many reservations now organize annual celebrations of traditional dancing and hold powwows, Indian days, Treaty days, or enact ceremonies, to which the general public is invited. Tourism is an increasingly important sector. There is also a recovery and development of traditional arts, from weaving carpets, making pottery or designing jewelry among the southern Pueblo peoples, to the carving of dance masks, boats, and traditional totem and house poles along the Northwest coast. In the far north, the Inuit produce stone carvings and prints that are in high demand in national and international art galleries. In the eastern forests and on the central plains, there has been a resurgence in the making of elaborate dance costumes, beaded foot wear, powwow costumes, canoes, snow shoes, and in the construction of traditional lodges.

First Nations peoples have held with amazing tenacity to their own understanding of people-land relationships which is radically different than that of the dominant majority. They continue to confront governmental, industrial and church agencies with their alternative world-view. Most elders today would see themselves as guardians of Mother Earth.

This has led them into increasingly frequent con ict with corporations and governments that plan to develop local resources by destructive means. Thus, there have been many demonstrations, marches, petitions, and road blockades trying to stop clear cut logging, open cast mining, dam construction, and military weapons testing on reservation lands, or on lands whose status is currently being claimed or disputed in the courts. The following examples may give some idea of the scale of the problem.

Although some 112 000 square km of the New Sogobia reservation was designated as the property of the Western Shoshone nation in 1863, and this was later con rmed in 1869, the US military has conducted some 670 weapon test explosions there since 1963, usurping 90% of their land.

> By 1980, 42 operating uranium mines, 10 uranium mills, ve coal power plants and four coal strip mines were in the vicinity of the Navajo reservation. Approximately 15 new uranium-mining operations were under construction on the reserve itself. Although 85% of Navajo households had no electricity, each year the Navajo nation exported enough energy resources to fuel the needs of the state of New Mexico for 32 years.
>
> (La Duke, 1992)

Despite continuing protest of the Innu People of Labrador and Northeast Quebec, their traditional life style has been rendered nearly impossible. In the 1950s, many were forcibly resettled into villages, when the Churchill River dam ooded ancient burial grounds and important areas of their hunting territory. In the early 1980s, the military base of Goose Bay was established, and used as a North Atlantic Treaty Organization (NATO) training base for low-level ight, which disturbs the caribou herds and makes sustained hunting impossible. Now mining companies have led over 13 000 claims in the area around Voisey Bay, Labrador.

Coal strip mining is a particular problem to the Cheyenne and Navajo-Hopi peoples. The Cheyenne reservation is virtually surrounded by coal strip mines, railroads, electric generating plants and transmission lines. Adjoining the reserve lies the largest federal coal sale in the history of the US the Powder River Coal lease, which follows the major water source the Tongue River.

Several coal- red power plants lie on or adjacent to the Navajo-Hopi lands, and the vast Black Mesa mine supplying coal to the Four Corners Power plant was the only man-made object seen by the Gemini Two astronauts from outer space. It caused not only pollution and falling water tables but also the relocation of some 10 000 people.

Across Canada, from the Queen Charlotte Islands of British Columbia, to Temagami, Ontario, to the Maritimes, First Nation peoples have been among the rst to organize protests against the huge timber leases sold to corporations for clear-cut logging. They have often been joined by the public in mass demonstrations, but have had only very limited success in changing government policies. For native people, the destruction of the forests together with their wildlife, and the deterioration of streams and rivers, and consequently of shing, has generally been disastrous. Although clear-cutting may generate a few short-term unskilled jobs for local First Nations people, trap lines are lost, hunting and traditional ways of life become more dif cult, and in the long-term it leads to increasing impoverishment of the community as a whole.

Similarly, almost all the major rivers in Canada have been dammed for hydroelectric developments. Some of these, like the Churchill Dam (Labrador), or the W A C Bennett Dam (British Columbia), or the vast James Bay Project of Northern Quebec, are among the largest engineering works of the 20th Century. Not only have the huge reservoirs behind the dams destroyed the most productive areas of lowland for vegetation, wildlife, and migratory birds, but sometimes the flow of rivers has been diverted or even reversed, resulting in the disruption of water tables, and resettlement of communities. Generally, if compensation is paid, it is only in cash, not in equivalent land and resources. This leaves the problem of increasing unemployment, cultural destruction, and despair in its wake.

In conclusion, one can only say that the response of First Nations peoples to 500 years of change in both the physical and social environment is in a state of very rapid transition. At one extreme are the isolated reservations where traditional reliance upon hunting and fishing (albeit using rifles, outboard motors, and snowmobiles) is still possible, although it may be seriously threatened by both the impact of pollution, and by the increasing corporate demand for resource exploitation. At the other extreme are examples like the Cherokee nation who have developed a large manufacturing center (including component parts for IBM, General Dynamics, and the US Army, among others), as well as retail and hotel sectors. The political and administrative center of Tahlequah handles an annual budget in excess of $75 million.

In two areas in particular, traditional ways of life have continued relatively undisturbed. One of these is on the vast semi-desert reservations of the Pueblo peoples of the Southwestern US. The other area is in the Canadian North and Alaska. The territory of Nunavut, established in 1999, covers approximately one-fifth of the land area of Canada. With a population of approximately 27 700 people, who are 90% aboriginal, it has the same autonomy as other territories and provinces. Even in both of these regions, modern transportation, education, medical services, expanding commercial sectors, resource exploitation and western forms of government and policing, create a large, and ever-increasing impact.

Unfortunately, in many of the smaller and more isolated reservations, where the land area is too small to support traditional hunting, and too poor to farm, as much as 90% of the population may still depend upon some form of government payments.

A few Native bands whose reservation land lies in large and growing urban centers have been able to develop their property and lease or rent it at considerable profit.

On many other reservations new commercial developments are appearing continuously: from casinos, shopping malls and tourist hotels to a resurgence of handcrafts, from toxic waste storage facilities to fashion designing. On independent band reservation land, mining and oil and gas exploitation is now generally under the management of Native corporations. On larger reservations, sustainable logging operations, modern farming and industrial developments are being established.

For the future, it seems safe to say that the only thing that is certain is continuing change. It would seem that the general movement towards increased autonomy will almost certainly continue in both the US and Canada. Eventually, the vast number of land claims will wend their way through the court systems, and be settled. However, the very size, number and complexity of claims (e.g., almost the entire land area of the province of British Columbia is currently being claimed by various Native bands) means that the final determination of land will inevitably disappoint many.

Will future developments on reservation lands be more environmentally sustainable than in North America as a whole? For the immediate future, the most urgent consideration for many band councils and governments is to improve the employment opportunities and raise the standard of living for people who have traditionally been impoverished and demoralized. Yet access to land is central to long-term indigenous health and cultural identity. Land and people are inseparably bound. The Native community can only be healthy in body, mind and spirit when the land is also healthy. The land is the promise of a healthy future for generations yet to come, it is where the bones of the ancestors lie, it links the past and the future, and it is sacred. Thus, while some indigenous people are among the poorest of the poor they are also holders of the key to the future survival of humanity.

See also: **Indigenous Knowledge, Peoples and Sustainable Practice**, Volume 5.

REFERENCES

Black Elk Speaks (1988) *Being the Life Story of a Holy Man of the Oglala Sioux*, as told through John G Neihardt (first published 1932), University of Nebraska Press, Lincoln, NE.

Clarkson, L, Morrissette, V, and Regallet, G (1992) *Our Responsibility to the 7th Generation: Indigenous Peoples and Sustainable Development*, International Institute for Sustainable Development, Winnipeg, 10.

LaDuke, W (1992) Indigenous Environmental Perspectives, A North American Primer, *Akwe:kon J.*, 52–70.

Tinker, G (1992) *The Full Circle of Liberation*, Sojourners, October, Box 29272, Sojourners Resource Center, Washington, DC, 12–17.

Wright, R (1993) *Stolen Continents: The New World through Indian Eyes*, Penguin Books, London.

FURTHER READING

Bordewich, F M (1996) *Killing the White Man s Indian: Reinventing Native Americans at the End of the 20th Century*, Anchor Books, Doubleday, New York.

Government of Canada (1996) *Report of the Royal Commission on Aboriginal Peoples*, five volumes, Ottawa.

Indian Claims Commission Proceedings, a Publication of the Indian Claims Commission (Canada). A Continuing Series of Official Reports, www.indianclaims.ca.

Versluis, A (1994) *Native American Traditions*, Element Books, Rockport, MA.

Romanticism

see Art and the Environment (**Volume 5**); Literature and the Environment (**Volume 5**)

Roosevelt, Theodore

(1858- 1919)

Theodore Roosevelt, US president (1901–1909), led one of the most dynamic lives of any US president. He was born into a wealthy and privileged New York family. Roosevelt's childhood was one of constant illness, but he overcame his infirmities through a self-imposed regimen of strenuous exercise. Strong will became a hallmark of his personality. He attended Harvard and studied law at Columbia.

In 1884, his wife Alice and his mother died on the same day – Alice from complications of childbirth, and his mother from a long illness. Roosevelt went to his ranch in the Dakota Territory where he slowly overcame his loss by living a tough physical life far from the world he knew in New York. In 1886, he married again, to Edith Carow.

During the Spanish–American War, Roosevelt was Lieutenant Colonel in the Rough Rider Regiment, and he led a famous charge at the battle of San Juan. This war hero status added to his celebrity image, and shortly after he was elected New York State Governor (1898).

Later, he served as vice president to William McKinley. McKinley was assassinated in 1901, and at the age of 43, Roosevelt became the youngest president in US history. He brought a new vigor to the presidency, a change from the staid and conservative McKinley. Roosevelt held the ideal that the government should be an arbiter and promoter of social and economic equity. Roosevelt embraced anti-trust legislation that soon dissolved monopolies in the railways and oil industry. The extensive use of executive power was a defining characteristic of his administration. He wrote I did not usurp power, but I did greatly broaden the use of executive power . Roosevelt believed that as steward of the people he must act forcefully to enhance the public good.

Roosevelt pushed the US toward a more discernable global role. Speak softly and carry a big stick was an adage he popularized. He backed construction of the Panama Canal, and won the Nobel Peace Prize for mediating an end to the Russo–Japanese War. However, he also had an important and lasting impact on conservation thought.

The time Roosevelt spent in the Dakotas had an impact on his tenure as president. He developed a love of the wilderness early in his life, and later saw the closing of the frontier and the settlement of the western landscape as the symbolic loss of what defined America. His friendships with John Muir and Gifford Pinchot provided kindred spirits for his romantic view of the American landscape, and they in turn greatly influenced his conservation program. Roosevelt's executive power gave him an opportunity to add vast areas of the west to the national forests, and to preserve places of special beauty such as the Grand Canyon. Despite this, he was a man of environmental contradictions – while being a keen observer of nature, he was also an avid hunter. Roosevelt's conservation was based on a utilitarian ideal – it was not about the absolute preservation of nature, but instead it was about the wise use of public lands and resources. His administration saw the creation of the first large resource bureaucracies. By the close of his presidency, Roosevelt had insured that the public would be the dominant landowner in the American West, and that the conservation ideal would long serve to guide the culture of American resource management.

Photo: Pach Brothers.

KEVIN HANNA *Canada*

S

Salinity, Waterlogging and Agriculture Socioeconomic Perspectives

see Salinity and Agriculture (**Volume 3**)

Scenarios

Jerome C Glenn

American Council for the United Nations University, Washington, DC, USA

A scenario is a story that connects a description of a speci c future to present realities in a series of causal links that illustrate decisions and consequences. A speci c year should be stated, such as 2025, and subject focus like a country, an industry, peace and con ict, etc. A scenario is not a single prediction or forecast, but a way of organizing many statements about the future.

Herman Kahn was the father of scenario construction. He de ned scenarios as narrative descriptions of the future that focus attention on causal processes and decision points (Kahn, 1967).

The purpose of scenarios is to systematically explore, create, and test both possible and desirable future conditions. Scenarios can help generate long-term policies, strategies, and plans, which help bring desired and likely future circumstances in closer alignment. They can also expose ignorance; show that we do not know how to get to a speci c future or that it is impossible.

Exploratory or descriptive scenarios describe events and trends as they could evolve, based on alternative assumptions on how these events and trends may in uence the future. Normative scenarios describe how a desirable future can emerge from the present.

Although it is not possible to know the future, it is possible to in uence elements of it. The forces of nature, social and political dynamics, scienti c discovery, and technological innovation largely determine the future. However, human choice increasingly shapes the future. This in uence makes the effort to consider the balance between what we want and what is possible worthwhile.

Scenarios should be judged by their ability to help decision makers make policy now, rather than whether they turn out to be right or wrong. Good scenarios are those are: (1) plausible (a rational route from here to there that make causal processes and decisions explicit); (2) internally consistent (alternative scenarios should address similar issues so that they can be compared); (3) suf ciently interesting and exciting to make the future real enough to affect decision making.

A scenario could be constructed to show how continuing war, Acquired Immune De ciency Syndrome, and environmental destruction could lead to increased African migration to Europe, which triggers racial con icts. The purpose of this scenario would be to alert decision makers about strategic interventions to make the scenario wrong. Conversely, The Millennium Project (Glenn and Gordon, 2000) constructed a global normative scenario to the year 2050 to illustrate how policy could dramatically improve the human condition. Here the purpose was to organize hundreds of previously collected positive developments and strategies into one whole picture that showed how policies, technologies, and changing human behaviors could have synergistic effect, and might become self-ful lling to some degree.

Herman Kahn introduced the term and concept of scenarios into planning in connection with military and strategic studies conducted by the Rand Corporation in the 1950s. He further popularized the concept in the 1960s as director of the Hudson Institute, a private nonpro t research center devoted to issues related to US public policy, international development, and defense. In 1967, Kahn along with Anthony Weiner examined the future possibilities of world order, describing potential power alignments and international challenges to American security in a book entitled Toward The Year 2000 *(Kahn and Weiner, 1967). One of their worlds depicted an arms control agreement between the US and the former Soviet Union; another assumed the former Soviet Union would lose control of the Communist movement; a third projected construction of new*

alliances among countries. In the book, Kahn and Weiner also described the technology hardware of the future, which included centralized computer banks with extensive information on individuals as well as parents able to select the gender and personal characteristics of their children through genetic engineering. This work was done under the Commission on the Year 2000 sponsored by the American Academy of Arts and Sciences.

Since the future can never be accurately or completely known, most planners and futurists today reject the idea that planning should be conducted against a single most likely image of the future. Rather, sets of scenarios should be used in planning; if the sets encompass a broad span of futures and plans are generated to cope with their eventualities, then the plans will be robust and the future can be met with some degree of con dence.

The Millennium Project has produced the largest annotated scenario bibliography with over 400 scenario sets, available at http://www.acunu.org/millennium/information. html, which are organized in the Futures Matrix in the following domains: demographics and human resources; environmental change and biodiversity; technological capacity; governance and con ict; international economics and wealth; integration or whole futures.

DESCRIPTIONS OF SCENARIOS

The term *scenario* comes originally from the dramatic arts. In the theater, a scenario refers to an outline of the plot; in movies, a scenario is a summary or set of directions for the sequence of action. Peter Schwartz of Global Business Network, a think tank in Emeryville, CA, often compares the initial process of creating a scenario with writing a movie script (Schwartz, 1992). Often in creating a scenario, a team of people considers such questions as: what are the driving forces? What is uncertain? What is inevitable? Similarly, scriptwriters formulate an idea and develop characters. Schwartz describes characters as the building blocks of scenarios.

In general, the term *scenario* has been used in two different ways: first, to describe a snapshot in time, or the conditions of important variables at some particular time in the future; second, to describe a future history, i.e., the evolution from present conditions to one of several futures. The latter approach is generally preferred because it can lay out the causal chain of decisions and circumstances that lead from the present. The most useful scenarios are those that display the conditions of important variables over time. In this approach, the quantitative underpinning enriches the narrative evolution of conditions or of the variables; narratives describe the important events and developments that shape the variables.

When scenarios are used in futures research and policy analysis, the nature of evolutionary paths is often important since policies can deflect those paths. In policy studies, families of scenarios are often used to illustrate the consequences of different initial assumptions, different evolutionary conditions, or both. For example, a study of transportation policy might involve constructing several scenarios that differ in their assumptions about birth rates, population, migration, and economic conditions, as well as the costs and availability of various forms of energy. When a set of scenarios is prepared, each scenario usually treats the same or similar parameters, but the evolution and actual value of the parameters described in each scenario are different.

The goal of generating scenarios is to understand the mix of strategic decisions that are of maximum benefit in the face of various uncertainties and challenges posed by the external environment. Scenario building, in conjunction with a careful analysis of the driving forces, fosters systematic study of potential future possibilities – both good and bad. This forecasting approach enables decision makers and planners to grasp the long-term requirements for sustained advantage, growth, and avoidance of problems.

HOW TO CONSTRUCT ALTERNATIVE SCENARIOS

Numerous methods have been developed to create scenarios, ranging from simplistic to complex, qualitative to quantitative. Some are created by one author acting alone, some by teams; and others through global participatory processes as in the Millennium Project (Glenn and Gordon, 2000). Many methods have similarities, although they may have unique features and use different terminology. Most approaches recognize the need to understand the system under study and identify the trends, issues, and events that are critical to this system. While each is not feasible to explain in detail, a brief description of several and a lengthier description of one are worthwhile.

Coates & Jarratt of Washington, DC, US, uses the following process to develop scenarios for a variety of clients, including countries and businesses. Coates & Jarratt begin by defining the universe of the area of interest. Key variables shaping the future are identified using a wide range of sources. Usually some 6–30 variables affecting the future situation are nominated. This list is then winnowed down by eliminating redundancies, a process that usually results in 6–20 variables. In the next step, the scenarios to be created are defined. One scenario usually presents a continuation of the present forces at play. Other scenarios may include an optimistic or positive scenario, which may be based on one or two of the particularly prominent variables. These scenarios may involve such occurrences as a technological breakthrough or change in government policy. Other scenarios can be framed around important futures, such as business booms, collapses, or other important occurrences. In

general, 3–6 scenarios are usually sufficient. The variables are then reviewed to determine what is a plausible qualitative and quantitative range of the value for that scenario. The scenarios can then be assigned to individuals for creation. Once completed, the scenarios should be reviewed for comprehensiveness and completeness, and then edited to ensure consistency in approach, layout, style, and format.

Coates & Jarratt emphasize that scenarios can be used to achieve different goals. Whereas some are designed to present a completed future, others may be used as points of departure for future discussions, such as policy implications. On some occasions, transition scenarios may be appropriate to development that describes the process of getting to certain end states. These transition scenarios can be either a separate set of scenarios or part of the primary scenarios themselves.

Peter Schwartz, a former member of the Royal Dutch/Shell scenario team, describes several steps in the scenario development process in his work *The Art of the Long View*. These steps include: identify the focal issue or decision; identify the key forces and trends in the environment; rank the driving forces and trends by importance and uncertainty; select the scenario logics; fill out the scenarios; assess the implications; and select the leading indicators and signposts for monitoring purposes (Schwartz, 1991; 226–234).

Thomas Mandel and Ian Wilson of SRI International wrote an excellent description of scenarios in a report entitled *How Companies Use Scenarios: Practices and Prescriptions* (Mandel and Wilson, 1993). In this SRI report, the authors explain the scenario development process. In the first two steps, management decides what it needs to know in order to make business decisions, and a scenario team describes the events, trends, and uncertainties that could impact the decision-making process. The team then analyzes forces that will shape the future business environment, both from within its own industry (competition) and outside of it (social, political, economic, etc.). The team then develops scenario theories or logics, which are differing views of the way the world might work in the future. Each theory takes into account critical drivers and uncertainties. Using the theories already developed as a guide, the scenarios are then described in sufficient detail to identify implications of decisions and to help develop and assess strategy options.

Michel Godet, from the Laboratory for Investigation in Prospective Studies in Paris (Godet, 1993), begins the scenario development process by constructing a base image of the present state of a system. This image is broad in scope, detailed, and comprehensive, dynamic, and descriptive of forces for change. The base image is constructed by delineating the system being studied, including a complete listing of the variables that should be taken into consideration as well as subdivisions of these variables (i.e., internal and external, as descriptive of the general explanatory environment). This step is followed by a search for the principal determinants of the system and their parameters, often using structural analysis. The scenario process involves examining the current situation and identifying the mechanisms and the leading actors (influencers of the system through variables) that have controlled or altered the system in the past. This process continues with development of actors' strategies. Construction of the database is followed by construction of the scenarios.

Godet combines various futures research techniques in scenario development. For example, he finds that morphological analysis can be used in scenario construction since scenarios are, in essence, a configuration of identified components. Godet cites the large number of possible combinations as a drawback in this application of morphological analysis, but offers a solution in the form of computer software that helps limit the field.

Ute Von Reibnitz of Strategische Unternehmensberatung makes the case that the ability to create different futures situations allows planners to deal with scenarios that fall between two extremes. In his book *Scenario Techniques* (Von Reibnitz, 1988), Von Reibnitz describes the process of constructing scenarios. Step one entails an analysis of an organization's structure, strengths and weaknesses, and goals and strategies. Step two involves an examination of areas and factors of external influences with attention to their interrelationship and dynamics. Step three is an analysis of the development of the future of influencing factors. Step four clusters different alternatives to form logical and plausible structures for scenarios. Step five incorporates these structures into scenarios that describe system dynamics and changes. The process concludes in Step six with an analysis of opportunities and risks.

In *Business Futures*, the Institute for Futures Research (1992) has presented yet another approach that begins with the identification of key issues. This step is subdivided into definition of mission, objectives and aims, and description of strategies and key decision-making parameters, followed by identification of key environmental forces, composed of environmental scanning, spotlighting of crucial issues, and listing predetermined events and forces for change. In the next step, the scenario logic is defined, the actors identified, and their likely behavior examined. Multiple scenarios are then created, and the implications of each are tested. The concluding step involves presentation of these scenarios to planning forums in order to evaluate their implications for action programs.

The Millennium Project of the American Council for the United Nations University (Glenn and Gordon, 2000) uses large-scale participatory processes. To construct a global normative scenario, hundreds of futurists, scholars, business planners, scientists, and policy makers who work for international organizations, governments, corporations,

non-government organizations, and universities identi ed and rated norms that formed the core of the normative scenario. In order of preference, the participants selected the following top four norms around which to form the scenario: environmental sustainability, plenty, global ethics (the identi ed and accepted ones), and peace. The others in order of preference were health, freedom, access to universal education, equity, preservation of the human species, enlightenment, exciting and meaningful life, self-actualization, longevity, everyone has everything they want, and security.

The body of the normative scenario was composed of the actions that would be required to address the 15 global challenges that had been identi ed through several rounds of global questionnaires to the Millennium Project participants. These actions connected the present world to the normative future of 2050. A scenario review panel was formed of long-term normative-oriented participants of the project to review and improve the draft of the scenario.

The Millennium Project global exploratory scenarios to the year 2025 also used a participatory process of questionnaires and interviews to collect information, but used a computer model with quanti able outputs to assure self consistency. The developments and policies found to be important by the participants provided fodder for the scenarios; these developments and policies were incorporated in each of the projected worlds. The method of incorporating the global model was also novel. Rather than having the model drive the scenario, the scenario drove the model.

As usual, the drafting of exploratory scenarios began with the choice of the principal independent dimensions (axes) that seemed to force the worlds under examination to differ. Future worlds were formed around these choices using techniques described below.

In the next step, these explicit scenarios were used to provide the backdrop for the choice of the values of the exogenous variables in the selected model. Therefore, when the model was run, it s output was consistent with the scenario on which the exogenous variables were based and the model provided quantitative estimates of the value of variables that were then incorporated within the scenario. While several different global models were considered, *International Futures* by Barry Hughes, University of Denver, was selected. It was well documented, relatively easy to use, and was made freely available to the project.

For a more complete description of this approach see http://www.acunu.org/millennium/scenarios/index.html.

The development of scenarios can range from a lengthy and intricate process to an abbreviated workshop. Peter Bishop, for example, has used the SRI/Shell/GBN scenario technique in introductory futures workshops. These workshops, in a four- to six-hour period, can complete a full development sequence of a scenario. The primary purpose of these workshops is to experience the process, not to utilize the actual results of the scenario. During this process, Bishop takes the group through the setting up of scenarios, developing scenario logic, and drawing out scenario implications. By asking focused questions, e.g., what is the most important issue concerning human resources for health care over the next 10 years , what will stay the same about this issue that will limit its alternative futures , and what is changing about this issue that will alter its future , Bishop sets up the scenarios and develops the scenario logic. During the process workshop, participants come to appreciate a wide range of variables that affect the future, as well as the interrelationships among those variables and the existence of alternative, plausible scenarios for the future.

Finally, it is worthwhile to explore one process in depth. This particular methodology is the one developed by The Futures Group.

Preparation

Define the scenario space: a scenario study begins by de ning the domain of interest. Given a clear statement of the domain, analysts list key driving forces thought to be important to the future of the domain. In a study performed by The Futures Group for MITRE Corporation about the social environment of crime, driving forces of law enforcement funding and social attitudes toward crime were de ned as ultimately important. To the degree possible, these driving forces should be independent axes in a scenario space. If three such forces were de ned, the space would be three-dimensional. With two forces, scenario space is two-dimensional. In the law enforcement case, these axes helped de ne four scenarios of interest:

1. high funding, permissive attitudes toward crime;
2. high funding, repressive attitudes toward crime;
3. low funding, permissive attitudes toward crime;
4. low funding, repressive attitudes toward crime.

De ning a large number of alternative worlds is often neither necessary nor desirable. A smaller set of choices that encompass the range of major challenges and opportunities usually suf ces. A few possibilities may need to be excluded as illogical or insuf ciently plausible over the planning horizon. The nal selection of worlds should be suf cient to present a range of opportunities and challenges, but should be small enough in number to handle. Four to ve worlds seems ideal to capture a range of future challenges and opportunities.

Development

Define the key measures: within each scenario, certain key measures are described. These measures might include forces such as economic growth, legislative environment, technology diffusion and proliferation, or competitive

capability, among others. The key measures need to be selected with care. They should have the potential for great impact on the outcome of the scenario; a factor is largely irrelevant if it could develop over a wide spectrum of future values but have little impact on the issue at hand. Every scenario in the set will include projections of the same measures.

De ne the events: this list of events will also appear in each scenario. These events shape the scenarios in several different ways: they can impact the key measures, change the chains of causality that lead from the present to the future, and/or make certain policies more or less likely to work. The probabilities of the events are different in each scenario and depend on their position in the scenario space.

Project the key measures: trend impact analysis is a useful technique for projecting the key measures. Briefly, the historical data for each of the measures is projected using time-series methods. The events, expressed probabilistically, are combined with the extrapolation using Monte Carlo methods to produce a new median forecast and a range of uncertainty. Since events within a scenario impact several measures wherever they are used, they have the same probability; thus, internal consistency is promoted.

Prepare descriptions: given the quantitative forecasts of the measures based on the probabilistic description of the impacting events, many chains of causality become apparent, and cohesive narratives describing the future histories can be prepared.

Reporting and Utilization

Document: in most cases, the best documentation is a simple series of charts and narratives describing the future history represented by each scenario. As thinking surrounding the scenarios is driven farther down in the organization, several levels of documentation for each of the scenarios is often useful. A top-line summary gives readers a quick, intuitive feel for the characteristics of a world from the perspective of a selected future time (say, 2010) – how it developed, and what the decisive events were that caused the world to develop as it did.

Contrast the implications of the alternative worlds: how different are the business decisions and planning goals you would pursue considering each alternative world? What actions and commitments offer your organization the most resilience in the face of these uncertainties?

Testing policies: the range of scenarios can be used to test policies. In any study, a list of alternative actions is prepared. This list may come from the decision makers after reading the scenarios. Each is defined as precisely as possible. Then, using quantitative techniques if possible, the policies are tested in each of the scenarios. When a particular policy produces desirable results in all cases, it is clearly a good policy to pursue. The other scenarios may give rise to contingent policies that can be called on if the circumstances develop that the scenarios depict.

SOME GENERAL GUIDELINES IN SCENARIO CONSTRUCTION

The most useful scenarios are sharply focused. They focus on critical issues facing the organization. The number of issues for consideration and the number of possible scenarios are almost endless. Without a clear direction, the discussion of drivers is difficult to limit. The number of alternative worlds expands exponentially, and the list of variables can become unworkably long. The best defense is to define the focus from the outset. Ask yourself: What planning questions need to be addressed? What variables are most needed in order to address these concerns?

Emphasize qualitative analysis at the start. While numbers and formal models are often valuable sources for understanding future prospects, they can be distracting at the early stages of scenario development. Quantification can be valuable in later stages. Formal models also can provide an effective way of separating the many parts of a complex system for close consideration.

Some argue that trying to select the most likely scenario is inappropriate. The best scenarios reflect many variables and possible turns of events that shape the dynamics of the system under study. Any single scenario that purports to define the most likely particular path through this maze is unlikely to materialize over the years to come. Fortunately, the scenario-building process does not focus on uncovering the most likely forecast but, rather, on identifying the range of feasible outcomes.

See also: **Policy Exercises: a Tool for Solving Environmental Policy Dilemmas**, Volume 4; **Futures Research**, Volume 5.

SOURCES OF FURTHER INFORMATION

Futures, Publisher: Butterworth-Heinemann, Linacre House, Jordan Hill, Oxford, OX2 8DP UK. Tel.: +44 865 310166; Fax: +44 865 310898.

Futures Research (Zukunftsforschung), Publisher: Swiss Society for Futures Research, SZF, Haldenweg 10 A, Muri, Ch-3074, Switzerland. Tel.: +41-031-952-66-55; Fax: +41-031-952-68-00.

Futures Research Quarterly, Publisher: World Future Society, 7910 Woodmont Avenue, Suite 450, Bethesda, MD 20814 USA. Tel.: +1 (301) 656-8274; Fax: +1 (301) 951-0394.

Futurescope, Publisher: Decision Resources, Inc., 17 New England Executive Park, Burlington, MA 01803 USA. Tel.: +1 (617) 270-1200; Fax: +1 (617) 273-3048.

Futuribles, Publisher: Futuribles International, SSR rue de Varenne, 75007 Paris. Tel.: +33-331 42-22-63-10; Fax: +33-331 42-22-65-54.

Planning Review, Publisher: The Planning Forum, 5500 College Corner Pike, Oxford, OH 45056 USA. Tel.: +1 (513) 523-4185.

The Futurist, Publisher: World Future Society, 7910 Woodmont Avenue, Suite 450, Bethesda, MD 20814 USA. Tel.: +1 (301) 656-8274; Fax: +1 (301) 951-0394.

Long-Range Planning, Publisher: Pergamon Press, Headington Hill Hall, Oxford OX3 OBW UK. Tel.: +44 (865) 79141; Fax: +44 (865) 60285.

Social Indicators Network News (SINET), Publisher: PO Box 24064, Emory University Station, Atlanta, GA 30322 USA. Tel.: +1 (404) 373-4756; Fax: +1 (404) 727-7532.

Technological Forecasting and Social Change, Publisher: Elsevier Science, Publishing Co., Inc., 655 Avenue of the Americas, New York, NY 10010 USA. Tel.: +1 (212) 633-3941; Fax: +1 (212) 633-3990.

Technology Forecasts and Technology Surveys, Publisher: Technology Forecasts, 205 S Beverly Drive, Suite 208, Beverly Hills, CA 90212 USA. Tel.: +1 (213) 273-3486.

ENDNOTE

Much of this material was drawn from the chapter on Scenarios in *Futures Research Methodology* by the Millennium Project, American Council for the United Nations University, http://acunu.org, edited by Jerome C Glenn with support from United Nations Development Programme's African Futures Project. This methodology series includes: (1) introduction & overview; (2) environmental scanning: (3) participatory methods; (4) structural analysis; (5) the Delphi technique; (6) systems and modeling; (7) decision modeling; (8) scenario construction; (9) trend impact analysis; (10) cross-impact analysis; (11) technological sequence analysis; (12) relevance trees and morphological analysis; (13) statistical modeling; (14) simulation-gaming; (15) futures wheel; (16) normative forecasting; (17) genius forecasting, vision, and intuition; and (18) methodological frontiers and integration.

ACKNOWLEDGMENTS

Special thanks to The Futures Group and reviewers who made many important suggestions and contributions: Dr Fabrice Roubelat, Electricite de France; Dr Joseph Coates of Coates and Jarratt; Dr William M Brown, retired from the Hudson Institute; Dr Peter Bishop, University of Houston; Dr Pavel Novacek, Palacky University, Czech Republic; Dr Stanislaw Orzeszyna, World Health Organization; Mr Larry Hills, US Agency for International Development; Dr Eleonora Masini, Pontifical Gregorian University, Rome, Italy; and Dr Terry O'Donnell of the University of Massachusetts.

REFERENCES

African Futures (1993) *A Methodological Guide to the NLTPS Process*, UNDP, Abidjan, Cote D'Ivoire.

African Futures (1993) *Issue Identi cation: Phase I of the NLTPS Process*, UNDP, Abidjan, Cote D'Ivoire.

African Futures (in press) *Probing the Future: Techniques, Tools and Methods*, UNDP, Abidjan, Cote D'Ivoire.

Becker, H S (1982) Constructing and Using Scenarios: an Aid to Strategic Planning and Decision Making, *World Future Soc. Bull.*, September/October, 13.

Becker, H S (1983) Scenarios: a Tool of Growing Importance to Policy Analysts in Government and Industry, *Technol. Forecast. Soc. Change*, 23, 95–120.

Becker, H S (1989) Developing and Using Scenarios – Assisting Business Decisions, *J. Bus. Indus. Marketing*, 4(1), 61–70.

Bell, D, ed (1968) *Toward the Year 2000*, Houghton Mifflin, New York.

Boroush, M A and Thomas, C W (1992) Case Study: Alternative Scenarios for the Defense Industry After 1995, *Planning Forum*, 20(3), May/June.

Boroush, M A and Thomas, C W (1992) Case Study: Defense Markets in an Era of Geopolitical Change, *Planning Rev.*, May/June 1992.

Boucher, W I (1985) Scenarios and Scenario Writing, in *Nonextrapolative Methods in Business Forecasting*, ed J S Mendell, Quorum Books, Westport, CT.

Coates, J F (1985) Scenarios Part One: the Future Will Be Different from the Present. What Do I Have to Do to Persuade You? in *Nonextrapolative Methods in Business Forecasting*, ed J S Mendell, Quorum Books, Westport, CT.

Coates, J F (1985) Scenarios Part Two: Alternative Futures, in *Nonextrapolative Methods in Business Forecasting*, ed J S Mendell, Quorum Books, Westport, CT.

Esterhuyse, W (1992) Scenarios for South Africa: Instability and Violence or Negotiated Transition? *Long Range Planning*, 25(3), 21–26.

Federal Energy Administration (1974) *Project Independence Report*, US Government Printing Office, Washington, DC.

Freeman, S D, ed (1974) A Time to Choose, *Final Report by the Energy Policy Project of the Ford Foundation*, Ballinger, Philadelphia.

Glenn, J C, ed (1999) *Futures Research Methodology*, American Council for the United Nations University, Washington, DC.

Glenn, J C and Gordon, T J (2000) *State of the Future at the Millennium*, CD-ROM section Appendix, American Council for the United Nations University, Washington, DC.

Godet, M (1993) The Scenario Method, From Anticipation to Action, in *A Handbook of Strategic Prospective*, 53–78, UNESCO, Paris.

Godet, M, Bourse, F, Chaupy, P, and Menant, I (1991) *Futures Studies: a Tool-Box for Problem Solving*, UNESCO, Paris.

Godet, M (1990) Integration of Scenarios and Strategic Management: Using Relevant, Consistent, and Likely Scenarios, *Futures*, 22(7), 730–739.

Heydinger, R B and Zentner, R D (1983) Multiple Scenario Analysis: Introducing Uncertainty into the Planning Process, in *Applying Methods and Techniques of Futures Research*, eds J Morrison, W L Renfro, and W I Boucher, Jossey-Bass, San Francisco, CA.

Institute for Futures Research (1992) *Business Futures 1992*, ed P H Spies, University of Stellenbosch, Bellville, South Africa.

Kahane, A (1992) Scenarios for Energy: Sustainable World vs. Global Mercantilism, *Long Range Planning*, 25(4), August.

Kahn, H and Weiner, A J (1967) *The Year 2000: a Framework for Speculation on the Next Thirty-Three Years*, Macmillan, New York.

Klein, H E and Linneman, R E (1981) The Use of Scenarios in Corporate Planning – Eight Case Histories, *Long Range Planning*, 14(5), 69–77.

Mandel, T F and Wilson, I (1993) *How Companies Use Scenarios: Practices and Prescriptions*, SRI International, Report No. 822, Spring.

Martino, J (1972) *Technological Forecasting for Decision Making*, Elsevier, New York.

Meadows, D, Meadows, D, and Randers, J (1972) *The Limits to Growth*, Universe Books, New York.

Millett, S M (1992) Case Study: Battelle's Scenario Analysis of a European High-Tech Market, *Planning Forum*, 20(2), March/April.

Millett, S M (1992) Case Study: Los Angeles 2007, *Planning Forum*, 20(3), May/June.

Saxena, J, Sushil, P, and Vrat, P (1992) Scenario Building: a Critical Study of Energy Conservation in the Indian Cement Industry, *Technol. Forecasting Soc. Change*, 41, 121–146.

Schoemaker, P J and van der Heijden, C A (1992) Case Study: Integrating Scenarios into Strategic Planning at Royal Dutch/Shell, *Planning Forum*, 20(3), May/June.

Schwartz, P (1991) The Art of the Long View: Planning for the Future in an Uncertain World, Doubleday, New York.

Schwartz, P (1992) Composing a Plot for Your Scenario, *Planning Forum*, 20(3), May/June.

Simpson, D G (1992) Key Lessons for Adopting Scenario Planning in Diversified Companies, *Planning Forum*, 20(3), May/June.

Stokke, R, Boyce, T A, Ralston, W K, and Wilson, I H (1991) Visioning (and Preparing for) the Future: the Introduction of Scenarios-Based Planning into Statoil, *Technol. Forecasting Soc. Change*, 40, 73–86.

Taylor, C W (1990) Alternative World Scenarios for Strategic Planning, US Army War College, Strategic Studies Institute, Carlisle Barracks, PA.

Taylor, C W (1992) *A World 2010: a New Order of Nations*, US Army War College, Strategic Studies Institute, Carlisle Barracks, PA.

Taylor, C W (1993) Eliminate Future Shock, *Chemtech.*, July, 12–15.

Tenaglia, M and Noonan, P (1992) Scenario-Based Strategic Planning: a Process for Building Top Management Consensus, *Planning Forum*, 20(2), March/April.

Von Reibnitz, U (1988) *Scenario Techniques*, McGraw-Hill, Germany.

Wack, P (1985) Uncharted Waters Ahead, *Harv. Bus. Rev.*, September/October.

Wack, P (1985) Scenarios: Shooting the Rapids, *Harv. Bus. Rev.*, November/December.

Werner, M (1990) Planning for Uncertain Futures: Building Commitment Through Scenario Planning, *Bus. Horiz.*, 33(3), 55–58.

Whipple, III, W (1989) Evaluating Alternative Strategies Using Scenarios, *Long Range Planning*, 22(3), 82–86.

Wilson, I (1978) Scenarios, in *Handbook of Futures Research*, ed J Fowles, Greenwood Press, Westport, CT.

Wilson, I (1992) Teaching Decision Makers to Learn from Scenarios: a Blueprint for Implementation, *Planning Forum*, 20(3), May/June.

Yergin, D and Gustafson, T (1993) *Russia 2010: and What it Means for the World*, Random House, London.

Zentner, R D and Gelb, B D (1991) Scenarios: a Planning Tool for Health Care Organizations, *Hosp. Health Services Admin.*, 36(2), 211–222.

Ziemer, D R (1992) A Decision Support System to Aid Scenario Construction for Sizing and Timing of Marketplaces, *Technol. Forecasting Soc. Change*, 42, 223–249.

Seattle WTO Third Ministerial Meeting

see Environmental Movement – the Rise of Non-government Organizations (NGOs) (**Volume 5**)

Security, Environmental

see Environmental Security (**Volume 5**)

Seveso

The Seveso incident focused attention on the risks of chemical manufacture and of the health hazards of tetrachlorodibenzo para-dioxin (TCDD). On July 10th, 1976, the town of Seveso, Italy, near Milan, was the victim of extensive contamination by the most toxic of synthetic chemicals, TCDD. A chemical plant manufacturing trichlorophenol exploded as a result of a safety valve rupture. The explosion sent a cloud of TCDD material over an area of several square kilometers. Although the total amount of TCDD was very small, estimated at

about one kilogram, the chemical is so acutely toxic that the event was catastrophic for the community.

Birds fell from the sky and the town's livestock were slaughtered. Many pregnant women sought abortions rather than risk birth defects associated with TCDD. Many of the town's residents were evacuated. Some 200 residents developed a skin disorder called chloracne, noted as associated with TCDD and observed in US soldiers in Vietnam who handled the herbicidal mixture, *Agent Orange*, which was contaminated with TCDD.

Health studies monitored the residents following the explosion. Increased rates of leukemia, lymphoma, and sarcoma, as well as gall bladder and biliary tract cancer, were observed in the exposed population. There was thorough documentation of TCDD levels in blood and clinical follow-up of the exposed residents. Thus the explosion contributed to medical understanding of the impact of TCDD exposure. The Seveso incident also led to governmental action in the European Community, particularly in relation to hazardous waste sites.

ELIZABETH MAY *Canada*

Shallow Ecology

see Deep Ecology (**Volume 5**)

Sierra Club

The Sierra Club is North America's oldest environmental organization. Founded in 1892, it has grown to be the largest membership-based environmental group in the United States with over 550 000 members. Its founder, John Muir, was a Scottish naturalist and writer. Muir championed the protection of the Sierra Nevada with colleague Robert Underwood Johnson, editor of *Century* magazine. Their efforts led to the creation of the Yosemite National Park in 1890.

When the park continued to be threatened by local ranchers seeking grazing rights within the park, Muir founded the Sierra Club, in his words to do something for wildness. He believed that if he could take primarily urban dwellers into the wilds to experience the Sierra Nevada first hand, they would become converts to the cause of wilderness. In its earliest days, the Sierra Club was essentially a hiking club. But it developed a huge influence. Around a legendary campfire at Mariposa Big Tree Grove in Yosemite, Muir persuaded US President Theodore Roosevelt to build a national park system. The Sierra Club has been instrumental in protecting national parks and monuments including the Grand Canyon, Mount Ranier, Kings Canyon, Glacier, Sequoia, Olympic, Death Valley, and Rocky Mountain. In the 1960s, the Sierra Club emerged from its strong California base to build a national organization. It expanded to deal with modern environmental threats such as air and water pollution, as well as it traditional campaigns such as protecting wild rivers from hydroelectric dams. Strong leaders such as Michael McCloskey and David Brower led the Sierra Club to new prominence in the conservation movement. The organization is different from other environmental groups in having an engaged membership. The grassroots, volunteer-based operating style of the Sierra Club has made it a powerful lobbying force in the US. It is subject to the criticism from more radical environmental groups that it is too conservative. Currently, the Sierra Club has chapters across the US, as well as influence through the related, but relatively independent, Sierra Club of Canada.

ELIZABETH MAY *Canada*

Small is Beautiful

Small is Beautiful - Economics as if People Mattered was written by E F (Fritz) Schumacher in 1973. The book sold over one million copies in the US alone and contributed substantially to critical analysis of the role of technology, economies of scale as well as the environmental impacts of mega-projects. Schumacher went on to publish *A Guide for the Perplexed*, and a third book, *Good Work*, was published posthumously based on his lectures. E F Schumacher died on September 4th, 1977.

The essence of Schumacher's creative thinking emerged through his work for a report to the government of India in 1962. He advanced the concept of *intermediate technology*, and later founded the Intermediate Technology Development Group in London, UK. Schumacher refined the argument that simpler, people-scaled technologies were more appropriate for the poor in developing countries. He furthered these theories through his work through the 1950s and 1960s as economic advisor to the British coal industry. He also advanced the cause of organic farming through the British Soil Association.

Small is Beautiful in many ways anticipated the doctrine of sustainable development (Brundtland, 1987) as Schumacher did not oppose growth, *per se*, but argued cogently for a kind of growth that did not degrade

living systems, locally or globally, but which, first and foremost met human needs. Smaller mills, mini-plants, operations at human scale provided the template for a radical and profound challenge to prevailing economics and an unquestioned embrace of technology.

ELIZABETH MAY *Canada*

Social Ecology

Social ecology is a small but important environmental movement that integrates the study of human and natural ecosystems through the attempt to understand the relationships between culture and nature. It is often contrasted with *deep ecology* (*see* **Deep Ecology**, Volume 5), of which it is highly critical. Social ecology s starting point is that environmental problems result from social attitudes and practices, therefore the former must be solved by focusing on and changing the latter. An alternative socially and ecologically benign future can be constructed by reharmonizing people s relationship to the natural world, and this will be achieved by reharmonizing their relationship with each other.

The latter may be done through maximizing individual autonomy, eliminating hierarchical relationships (particularly humans over nature and *vice versa*, but also within human societies, such as patriarchy and class oppression) and creating in their stead simple social relationships, including direct, inclusive, economic democracies. The state, capitalism, and giantism are all considered inimical to social ecology s vision of harmonious society, which marries anarchism and ecology drawing particularly on the anarcho-communist legacy of Peter Kropotkin.

Social ecology s most in uential theorist is Murray Bookchin. His critical, holistic worldview places human behavior and institutions within an ecological context that represents social relations as a web of dynamic interrelated systems. But unlike (some interpretations of) deep ecology, Bookchin s approach would not prioritize nature over human society, nor would it dismiss the social aspirations of the Enlightenment Project. It would seek to be neither ecocentric nor anthropocentric, but again unlike deep ecology it sees humankind as evolution s highest expression, i.e., nature s consciousness, so human transformation of nature is natural and desirable.

Social ecologists present speci c utopian visions of a rational, ecological society, where labor is organized into small, human-scale communities without power hierarchies, and where exploitation of nature is minimal and production is environmentally friendly.

DAVID PEPPER *UK*

Social Impact Assessment (SIA)

Social impact assessment (SIA) is the process of analyzing (predicting, evaluating and re ecting) and managing the intended and unintended consequences on the human environment of planned interventions (policies, programs, plans, and projects) so as to create a more sustainable biophysical and human environment. A short form of this de nition is: SIA is the process of analyzing and managing the intended and unintended consequences of planned interventions on people.

SIA is more than a technique or tool; it is a philosophy about development and democracy, which considers pathologies of development (i.e., harmful impacts), goals of development (such as poverty alleviation or gender equity), and processes of development (e.g., participation, capacity building).

SIA is an overarching framework that considers all potential impacts on humans and their communities including: (1) changes to people s way of life, i.e., how they live, work, play and interact with one another on a day to day basis; (2) changes to their culture, community, neighborhood and local biophysical environment; (3) demands on their political systems, social institutions and community infrastructure; (4) effects on their health and well-being, on their personal and property rights; and on their fears and aspirations for themselves and their families.

The objective of SIA is to ensure that development that does occur maximizes the bene ts and minimizes the costs of development, especially those costs borne by the community and often not adequately taken into account by international development agencies decision-makers, regulatory authorities and developers. By identifying impacts in advance: (1) better decisions can be made about which projects should proceed and how they should proceed; and (2) mitigation measures can be implemented to minimize the harm from a speci c project or project-related activity.

SIA has three different levels of meaning: (1) as a paradigm or eld of research and practice; (2) as a methodology or environmental management instrument; and (3) as a discrete step (or speci c task) within a methodology. SIA is used at these different levels often

with some people failing to appreciate that different meanings or levels exist.

As a paradigm, SIA consists of a body of knowledge, techniques, and values. Various people identify themselves as SIA professionals, or list SIA as a discipline or specialty area. There is a community of scholars engaged in research and practice of SIA. These are people who practice the methodology of SIA and who undertake associated social and environmental research to inform the practice of SIA. As a methodology or instrument, SIA is the process (series of steps) that SIA professionals follow in order to assess and manage social impacts. That process requires substantial interaction with interested and affected peoples. At its narrowest meaning, SIA refers to the task of assessing likely social impacts of a proposed project within an environmental assessment framework (*see* Social Impact Assessment, Volume 4).

FRANK VANCLAY *Australia*

Social Learning in the Management of Global Atmospheric Risks

Adam Fenech
Meteorological Service of Canada, Downsview, Canada

Social learning is how humans, as individuals and groups, adopt and spread new concepts, knowledge and skills. It has recently begun to be applied to the process by which societies respond to environmental issues. There was a clear recognition during the late 1980s of the need for a better understanding of how human societies perceive and respond to global environmental change over time, i.e., social learning. Most of the previous understanding of how societies learned re ected the perspectives of a very narrow range of countries and groups, and was focused on key scienti c discoveries and decisions, with little attention to the historic connections between them. There was little critical discussion of what might be appropriately learned from the experience of dealing with other environmental problems, and from the experience of dealing with them in other places or countries. For a general reference, see Rose (1993). This publication is a critical collection of studies on the various kinds of social learning encountered in policy making.

When applying a social learning framework, researchers ask questions such as: Who (individual, organization, country) learned? Was the learning from within the social group, among social groups or from another country? Who was the teacher? Who was the learner? What was learned? What knowledge, experience, or norm was learned? How did this knowledge, experience or norm t or rede ne existing practices, models and decision-making processes? How did the learning occur? Who brought new information to bear on the existing issue? Where did the new information come from? How long did the new information take to formulate into an issue?

A major study applying a social learning framework to the environment has recently been completed under the leadership of William C Clark, Harvard University (Clark et al., *2001). It provides a better understanding of how human societies learned from earlier responses and other countries and organizations in the management of global atmospheric risks. The study applies the social learning framework to global atmospheric risks - namely acid rain, global climate change and stratospheric ozone depletion - from the International Geophysical Year of 1957 to the United Nations Conference on Environment and Development of 1992.*

In the context of environmental issues, Clark's book provides a major entry point into the eld of social learning. From this project, we present a couple of Canadian examples - namely stratospheric ozone depletion and climate change - to illustrate the concept.

SCIENTISTS-TO-SCIENTISTS LEARNING

One form of social learning is a scientist obviously learning from other scientists. What is not so obvious is that, for example, the cadre of Canadian scientists from Environment Canada, who inform and advise the Canadian government, often learn from scientists from other countries about issues, even if other scientists from Canada are already teaching the same message. For example, John Hampson, working at the Canadian Armament Research and Development Establishment (CARDE) in Val Cartier, Quebec in the early 1960s, provided the first statement of concern that human activities could harm the stratospheric ozone layer. Hampson was looking for radiative early-warning signals of incoming missiles, focusing on water vapor from both high-flying aircraft and rockets, arguing that hydrogen species arising from the photochemical dissociation of water could lead to ozone destruction. His work appeared only in CARDE technical reports, which, though widely circulated among both Canadian and American atmospheric and defense research communities, were never taken seriously. Hampson also raised the concern of ozone depletion from

increased air traffic in the atmosphere; yet the Canadian government did not take action until the issue was raised in response to an American research program in the mid 1970s investigating the issue of supersonic transport emissions.

This avenue of learning from scientists-to-scientists illustrates the importance of the teachings being presented in person, either by sending a Canadian scientist to an international gathering to bring the teachings back to Canada, or by bringing the teacher directly to Canada. For example, Sherry Rowland attended a 1975 seminar at York University in Toronto, Ontario, Canada, presenting his theories that developed the mechanism for chlorine catalysis of ozone, and identified chlorofluorocarbons as the largest potential source of stratospheric chlorine. This gave the issue visibility and some initial credibility in Canada.

However, the teaching is not blindly accepted by the learners, since in this case Canada required its own monitoring, modeling and science to confirm the implications and necessity for response.

MEDIA-TO-PUBLIC LEARNING

Another obvious learning channel is from media-to-public. The media acts as a teacher of the Canadian public, creating controversies that spark scientific investigation and government action. The surprise here is that in many instances, the media attention comes from American sources, thus making the US media a teacher and the Canadian public and Canadian media learners. For example, US media coverage of the US Senate Committee hearings of National Aeronautics and Space Administration (NASA) scientist Hansen's concern about the climate change issue sparked Canada's own media to report on the consensus of scientists that climate change was a real concern.

This Canadian experience is repeated in many other countries.

OTHER ASPECTS OF SOCIAL LEARNING

There are many other forms of social learning, ranging from top-down trickle-down arrival of issues, to peer-to-peer exchanges (as exemplified in the scientist-to-scientist model), to bottom-up experiential learning, i.e., the experts are baffled by the kinds of things that ordinary people learn by living every day.

For example, one form of bottom-up learning has garnered increasing interest over the years as the value of traditional or community knowledge has been appreciated and sought out by researchers. Traditional knowledge refers to the understanding of the natural environment that indigenous peoples possess. Traditional knowledge stems from the inherent closeness to the environment that indigenous peoples display by the mere fact of their livelihood and existence being tied to the land, and to what the wilderness offers in terms of food and medicines. Traditional knowledge is linked as well to the passing of knowledge from one generation to another, allowing events in the far past to be passed on through storytelling. Community knowledge refers to the knowledge of a natural resource such as fish or trees by people who live in villages and hamlets. These communities also have knowledge about the changing environment because they are close to the land and wilderness, relying on it for their existence. The value of traditional or community-based knowledge has not been fully realized as there have been no attempts to merge this informal, non-scientific understanding with the rigorous scientific method of current experts. Other approaches are required to fully appreciate the value that traditional and community-based knowledge can provide to understanding our changing environment (*see* Indigenous Knowledge, Peoples and Sustainable Practice, Volume 5).

This new field opens up a myriad questions. Where do people get their information? Who are the champions of new ideas? What gives an idea prominence or acceptability among experts and in the public domain? Is there an optimal model for social learning that we could use to speed up the social learning curve for environmental issues?

REFERENCES

Clark, W C, Jager, J, van Eijndhoven, J, and Dickson, N M, eds (2001) *Learning to Manage Global Environmental Risks: a Comparative History of Social Responses to Climate Change, Ozone Depletion and Acid Rain*, MIT Press, Cambridge, MA, in press.

Rose, R (1993) *Lesson-Drawing in Public Policy*, Chatham House Publishers, Chatham, NJ.

FURTHER READING

Environment Canada (1988) *The Changing Atmosphere: Implications for Global Security - Conference Statement*, Report from the Conference on The Changing Atmosphere: Implications for Global Security, June 27–30, 1988, Toronto, WMO, Geneva.

Social Science and Global Environmental Change

see Social Science and Global Environmental Change (*Opening essay,* **Volume 5**)

Social Sustainability

see The Changing Human–Nature Relationships (HNR) in the Context of Global Environmental Change (*Opening essay*, **Volume 5**)

Sociology, Environmental

see Environmental Sociology (**Volume 5**)

Soft Energy Paths

The term *soft energy paths*, was coined by environmental theorist and energy critic Amory Lovins in *Soft Energy Paths* (Lovins, 1977). It can be understood as a reversal of dominant thinking at the time; shifting the focus of energy planning from meeting anticipated demands with increasing energy supplies, to reducing demand through efficiency. Lovins' arguments to leave the hard path of energy planning, a path of brittle, inflexible systems dominated by energy mega-projects, long-transmission lines and bleeding energy losses, were based on the demand side management approach. The key to soft path energy planning is to match energy source to energy demand. A barrel of energy saved has the same value to the economy as a barrel of energy found. Thus, energy efficiency and conservation create energy along the soft path. Lovins called such energy *negawatts*.

According to Lovins:

> The energy problem, according to conventional wisdom, is how to increase energy supplies ••• to meet projected demands. The solution ••• ever more remote and fragile places are to be ransacked, at ever greater risk and cost, for increasingly elusive fuels, which are then to be converted to premium forms (electricity and fluids) in ever more costly, complex centralized and gigantic plants.

To illustrate the inappropriate nature of focusing on supply as opposed to demand, Lovins wrote:

> to ask which is the greenest supply of electricity is somewhat like shopping for the best buy of brandy to burn in your car, or the best buy in antique furniture to burn in your stove. From the end use point of view it is to ask the wrong question.

Lovins' influential work has continued, in partnership with Hunter Lovins, from their base at the Rocky Mountain Institute. Their most recent book, co-authored with Paul Hawken, is *Natural Capitalism*.

REFERENCE

Lovins, A B (1977) *Soft Energy Paths: Towards a Durable Peace*, Penguin Books, Harmondsworth.

ELIZABETH MAY *Canada*

Soft Law

see International Environmental Law (**Volume 5**)

Sovereignty and Sovereign States

Karen T Litfin
University of Washington, Seattle, WA, USA

The modern international system is comprised of sovereign states, meaning that each state is in principle legally and constitutionally independent. In practice, however, sovereignty is generally conceived more broadly in terms of some combination of control, autonomy, and authority (Lit n, 1997).

Control, like power, is the ability to produce an effect. Autonomy, or independence, is related to control since the ability to make and implement decisions on one's own requires some degree of power. Authority refers to the recognized right to make rules.

The tension between environmental protection and state sovereignty is evident in the famous Principle 21 of the 1972 Stockholm Declaration, which articulates the state's sovereign right to exploit its own resources and its responsibility not to harm the environment beyond its borders. While sovereignty is most commonly interpreted as a state's right to exercise control and authority within its own territory, the history of international environmental treaty making may be read as an effort by states to strike a balance between rights and responsibilities.

Many international relations scholars believe that sovereignty is being eroded, or at least reconfigured, in the face of growing interdependence, including efforts to address global and transboundary environmental change. The apparent mismatch between nature and the structure

of politics has been expressed in the famous dictum: The Earth is one, but the world is not (World Commission on Environment and Development, 1987). As environmental issues emerged on the international agenda in the early 1970s, some analysts predicted that global ecological interdependence would help bring about the demise of the sovereign state system and its eventual replacement by some far-reaching supranationalism or even world government (Falk, 1971; Ophuls, 1977). More recently, with the proliferation of local and transnational environmental non-governmental organizations in the 1980s, some observers have predicted the erosion of sovereignty from below, with power and authority devolving from sovereign states to the grassroots (Ekins, 1992).

Although the erosion of sovereignty thesis is appealing, a counter argument can be advanced that ecological integrity and state sovereignty, are not necessarily in opposition. One might claim that only the state possesses sufficient authority, resources, and territorial control to enforce environmental regulations (Litfin, 1993). Two important developments since 1970 support this position: the establishment of national environmental agencies throughout the world and the negotiation and implementation of a host of environmental treaties. Moreover, since some challenges to state sovereignty, such as efforts to promote free trade by eliminating environmental restrictions, may be harmful, some would claim that sovereign states are key defenders of the environment (Esty, 1994).

Neither the erosion of sovereignty thesis nor the sovereignty as bulwark thesis, however, does justice to global environmental trends. Both positions conceptualize sovereignty as monolithic rather than multidimensional. Rather, sovereignty should be understood as a historically variable set of aggregated practices. The scope of state autonomy may be narrowed by international and domestic pressures, as the erosion of sovereignty thesis claims, even as the problem solving capacity of states increases, as the second view suggests. Because international environmental responses typically involve trade-offs among autonomy, control, and authority, it is more accurate to say that states engage in sovereignty bargains (Byers, 1991) instead of ceding some monolithic principle of sovereignty. States may voluntarily accept limitations on their decision-making autonomy, for instance, in exchange for a greater ability to protect natural resources either within or beyond their own territorial jurisdiction. Thus, sovereignty appears to be reconfigured, rather than simply eroded, by efforts to address transnational and global environmental change.

Developing countries are particularly sensitive to the possibility that their sovereignty will be compromised by international efforts to address environmental problems. The principle of sovereignty over natural resources, which has engaged the United Nations since 1952, is often cited by developing countries opposed to what they perceive as environmental colonialism. For instance, efforts by industrialized countries to internationalize the issue of tropical deforestation in the 1980s and 1990s met with opposition from developing countries, which claimed sovereign jurisdiction over their natural resources.

The principle of nonintervention, which is key to state sovereignty, is taking on new meanings in light of environmental change. In the past, it was primarily understood in terms of a sovereign state's expectation that foreign powers would not intervene politically and militarily in its own affairs. As we enter the 21st century, the meaning of nonintervention is being renegotiated in the context of global ecological interdependence. When chlorofluorocarbons released in Seattle deplete New Zealand's ozone layer and nuclear fallout from the Chernobyl reactor decimates Scandinavia's reindeer, new norms of nonintervention are required.

As environmental concerns insert themselves increasingly into world politics, and as the political and economic effects of such developments as massive species extinction, climate change, and the globalization of western modes of consumption become increasingly inescapable, sovereignty will be reconfigured in ways that cannot be fully anticipated today.

REFERENCES

Byers, B (1991) Ecoregions, State Sovereignty, and Conflict, *Bull. Peace Proposals*, **22**(1), 68 72.

Ekins, P (1992) *A New World Order*, Routledge, London.

Esty, D (1994) *Greening the GATT*, Institute for International Economics, Washington, DC.

Falk, R (1971) *This Endangered Planet: Prospects and Proposals for Human Survival*, Vintage Books, New York.

Lit n, K (1993) Ecoregimes: Playing Tug-of-War with the Nation-State, in *The State and Social Power in Global Environmental Politics*, eds R Lipschutz and K Conca, Columbia University Press, New York.

Lit n, K (1997) Sovereignty in World Ecopolitics, *Mershon Int. Stud. Rev.*, **41**(2), 167 204.

Ophuls, W (1977) *Ecology and the Politics of Scarcity*, W H Freeman, San Francisco, CA.

World Commission on Environment and Development (1987) *Our Common Future*, Oxford University Press, Oxford.

Stewardship Theology

see Christianity and the Environment (**Volume 5**); Theology (**Volume 5**)

Strong Anthropocentrism

see Environmental Ethics (**Volume 5**)

Subsidiarity Principle

see The Changing Human–Nature Relationships (HNR) in the Context of Global Environmental Change (*Opening essay*, **Volume 5**)

Sustainability: Human, Social, Economic and Environmental

Robert Goodland
World Bank, Washington, DC, USA

The four main types of sustainability are human, social, economic and environmental. These are de ned and contrasted in Tables 1- 4. It is important to specify which type of sustainability one is dealing with as they are all so different and should not be fused together, although some overlap to a certain extent. Specialists in each eld best deal with these four types of sustainability. For example, social scientists have a lot to say about social sustainability; economists deal with economic sustainability and biophysical specialists deal with environmental sustainability.

A definition of environmental sustainability (ES) has been given by Daly (1973, 1974, 1992, 1996, 1999) and Daly and Cobb (1989):

1. Output rule: Waste emissions from a project or action being considered should be kept within the assimilative capacity of the local environment, without unacceptable degradation of its future waste absorptive capacity or other important services.
2. Input rule:
   - Renewable resources: (e.g., forest, fish) harvest rates of renewable resource inputs must be kept within regenerative capacities of the natural system that generates them.
   - Non-renewables: depletion rates of non-renewable resource inputs should be set below the historical rate at which renewable substitutes were developed by human invention and investment according to the Serafian quasi-sustainability rule (see below). An easily calculable portion of the proceeds from liquidating non-renewables should be allocated to the attainment of sustainable substitutes.

SERAFIAN QUASI-SUSTAINABILITY RULE OF NON-RENEWABLES

The Serafian rule pertains to non-renewable resources, such as fossil fuels and other minerals, but also to renewables to the extent they are being mined. It states that their owners may enjoy part of the proceeds from their liquidation as income, which they can devote to consumption. The remainder, a user cost, should be reinvested to produce income that would continue after the resource has been exhausted. This method essentially estimates income from sales of an exhaustible resource. It has been used as a normative rule for quasi-sustainability, whereby the user cost should be reinvested, not in any asset that would produce future income, but specifically to produce renewable substitutes for the asset being depleted. The user cost from depletable resources has to be invested specifically in

Table 1 Comparison of Human, Social, Economic and Environmental Sustainability: Human Sustainability

• Human sustainability means maintaining human capital. Human capital is a private good of individuals, rather than between individuals or societies. The health, education, skills, knowledge, leadership and access to services constitute human capital. Investments in education, health, and nutrition of individuals have become accepted as part of economic development
• As human life-span is relatively short and finite (unlike institutions) human sustainability needs continual maintenance by investments throughout one's lifetime
• Promoting maternal health and nutrition, safe birthing and infant and early childhood care fosters the start of human sustainability. Human sustainability needs 2–3 decades of investment in education and apprenticeship to realize some of the potential that each individual contains. Adult education and skills acquisition, preventive and curative health care may equal or exceed formal education costs
• Human capital is not being maintained. Overpopulation is intensifying and is the main dissipative structure worsening per capita indices. That is far graver than overcapitalizing education so that laborers have PhDs

Table 2 Comparison of Human, Social, Economic and Environmental Sustainability: Social Sustainability

- Social sustainability means maintaining social capital. Social capital is investments and services that create the basic framework for society. It lowers the cost of working together and facilitates cooperation: trust lowers transaction costs. Only systematic community participation and strong civil society, including government can achieve this. Cohesion of community for mutual benefit, connectedness between groups of people, reciprocity, tolerance, compassion, patience, forbearance, fellowship, love, commonly accepted standards of honesty, discipline and ethics. Commonly shared rules, laws, and information (libraries, film, and diskettes) promote social sustainability
- Shared values constitute the part of social capital least subject to rigorous measurement, but essential for social sustainability. Social capital is undercapitalized, hence the high levels of violence and mistrust
- Social (sometimes called moral) capital requires maintenance and replenishment by shared values and equal rights, and by community, religious and cultural interactions. Without such care it depreciates as surely as does physical capital. The creation and maintenance of social capital, as needed for social sustainability, is not yet adequately recognized. Western-style capitalism can weaken social capital to the extent it promotes competition and individualism over cooperation and community
- Violence is a massive social cost incurred in some societies because of inadequate investment in social capital. Violence and social breakdown can be the most severe constraint to sustainability

Table 3 Comparison of Human, Social, Economic and Environmental Sustainability: Economic Sustainability

- Economic capital should be maintained. The widely accepted definition of economic sustainability is maintenance of capital, or keeping capital intact. Thus Hicks's definition of income–the amount one can consume during a period and still be as well off at the end of the period–can define economic sustainability, as it devolves on consuming value-added (interest), rather than capital
- Economic and manufactured capital is substitutable. There is much overcapitalization of manufactured capital, such as too many fishing boats and sawmills chasing declining fish stocks and forests
- Historically, economics has rarely been concerned with natural capital (NC) (e.g., intact forests, healthy air). To the traditional economic criteria of allocation and efficiency must now be added a third, that of scale (Daly, 1992). The scale criterion would constrain throughput growth–the flow of material and energy (NC) from environmental sources to sinks
- Economics approaches value things in money terms, and have major problems valuing NC, intangible, intergenerational, and especially common-access resources, such as air. Because people and irreversibles are at stake, economic policy needs to use anticipation and the precautionary principle routinely, and should err on the side of caution in the face of uncertainty and risk

Table 4 Comparison of Human, Social, Economic and Environmental Sustainability: Environmental Sustainability (ES)

- Although ES is needed by humans and originated because of social concerns, ES itself seeks to improve human welfare by protecting NC. As contrasted with economic capital, NC consists of water, land, air, minerals and ecosystem services, hence much is converted to manufactured or economic capital. Environment includes the sources of raw materials used for human needs, and ensuring that sink capacities recycling human wastes are not exceeded, in order to prevent harm to humans
- Humanity must learn to live within the limitations of the biophysical environment. ES means NC must be maintained, both as a provider of inputs (sources), and as a sink for wastes. This means holding the scale of the human economic subsystem (= population $\times$ consumption, at any given level of technology) to within the biophysical limits of the overall ecosystem on which it depends. ES needs sustainable consumption by a stable population
- On the sink side, this translates into holding waste emissions within the assimilative capacity of the environment without impairing it
- On the source side, harvest rates of renewables must be kept within regeneration rates
- Technology can promote or demote ES. Non-renewables cannot be made sustainable, but quasi-ES can be approached for non-renewables by holding their depletion rates equal to the rate at which renewable substitutes are created. There are no substitutes for most environmental services, and there is much irreversibility if they are damaged

replacements for what is being depleted in order to reach sustainability, and must not be invested in any other venture – no matter how profitable. For non-renewable energy, a future acceptable rate of extraction of the non-renewable resource can be based on the historic rate at which improved efficiency, substitution and re-use became available. These calculations show the folly of relying on technological optimism, rather than on some historic track record.

CAUSES OF UNSUSTAINABILITY

When the human economic subsystem was small, the regenerative and assimilative capacities of the environment appeared infinite. We are now painfully learning that environmental sources and sinks are finite. Originally, these capacities were very large, but the scale of the human economy has exceeded them. Source and sink capacities have now become limited. As economics deals only with scarcities, in the past source and sink capacities of the environment did not have to be taken into account. Conventional economists still hope or claim that economic growth can be infinite or at least that we are not yet reaching limits to growth.

REFERENCES

Daly, H E, ed (1973) *Toward a Steady State Economy*, Freeman, San Francisco, CA.

Daly, H E (1974) The Economics of the Steady State, *Am. Econ. Rev.* March, 15–21.

Daly, H E (1992) Allocation, Distribution and Scale: Towards an Economics which is Efficient, Just and Sustainable, *Ecol. Econ.*, 6(3), 185–193.

Daly, H E (1996) *Beyond Growth: the Economics of Sustainable Development*, Beacon Press, Boston, MA.

Daly, H E (1999) *Ecological Economics and the Ecology of Economics*, E Elgar Publications, Cheltenham.

Daly, H E and Cobb, J B (1989) *For the Common Good*, Beacon Press, Boston, MA.

FURTHER READING

El Serafy, S (1989) The Proper Calculation of Income from Depletable Natural Resources, in *Environmental Accounting for Sustainable Development*, eds Y J Ahmad, S El Serafy, and E Lutz, The World Bank, Washington, DC.

Hueting, R, Bosch, P, and de Boer, B (1995) *The Calculation of Sustainable National Income*, Organization for International Cooperation, Den Haag, Netherlands [and] Indian Council of Social Science Research, New Delhi, 195.

T

Technological Society and its Relation to Global Environmental Change

see Technological Society and its Relation to Global Environmental Change (*Opening essay*, **Volume 5**)

Theology

In the context of human–environment relations, there are four kinds of theology:

Ecotheology is theological reflection focused on the relationship of faith to ecological wellbeing and social justice.

Dominion theology refers to theological worldviews that perceive humans as being placed by God at the top of a pyramid of value with all other creation made expressly for the use of humanity and over which humans could exercise dominion or control.

Stewardship theology is a challenge to traditional dominion theology that argues that humanity's place within creation is not one of authority but one of responsibility to care and nurture the Earth on behalf of the creator.

Process theology is a school of theology that emphasizes the inextricable linkages of humans within the entirety of the creation.

DAVID G HALLMAN *Canada*

Theories of Health and Environment

Trevor Hancock
Trevor Hancock Inc., Kleinburg, Canada

Early theories of health - still commonly found in many cultures throughout the world - understood health as the result of the characteristics and behavior of the individual and his or her relationship with the community, nature and the spirits or gods. The role of healers was to re-establish balance and to drive out evil spirits. Three great healing traditions (Greco-Arabic, Indian [Ayurvedic] and Chinese) share common ideas about maintaining equilibrium among the humors, taking into account environmental factors, and addressing personal factors and the spiritual dimension.

In Europe, from the 2nd century AD until the 19th century, the miasmic theory that atmospheric and environmental conditions were responsible for the transmission of epidemic diseases competed with the contagion theory, which suggested that transmission was via minute infective agents or spores, and the theory of individual predisposition. While passive acceptance of environmental conditions marked much of this period, activist Boards of Health first appeared in Northern Italy in the 15th century. Later, the social hygiene movement of the 18th century laid the groundwork for the emergence of the public health movement in the early 19th century.

The public health approach addressed environmental sanitation and living and working conditions that had been worsened by the industrialization and urbanization of the 19th century. At the same time, improved nutrition, education and other social and economic development factors - in turn, based upon exploitation of the earth's resources - contributed to a signi cant increase in life

expectancy and health in the 19th and early 20th centuries. But with the identi cation of bacteria in the 1870s and the development of effective vaccines and, in the 1930s, antibiotics, the pendulum swung toward a medical model of health focused on the individual and the treatment of speci c diseases. This has been the dominant paradigm for most of the past 100 years.

Beginning in the 1970s, however, there has been a growing awareness that healing and caring for the individual is much more than just xing the body, and that the determinants of health are much broader than simply the provision of medical care. In addition to recognizing that much of the improvement in health in the 19th and 20th centuries has been due to social and economic development, the view from space has helped us to understand that we all share a small and fragile planet and that our health is inextricably linked to the health of its ecosystems.

Three powerful new, and complementary, models that are emerging in the western world in many ways represent a return to the public health model of the 19th century and to even earlier theories of health as balance within the person, balance between the individual and other people and balance between humans and nature. These three models - holistic medicine, population health and ecosystem health - represent a new synthesis, a new theory of health with roots that are thousands of years old.

INTRODUCTION

It seems that every culture and every period in history has a theory of health – what it is, what determines health and how to achieve it. The by-now classical World Health Organisation (WHO) definition of health (*see* **WHO (World Health Organization)**, Volume 4) as a state of complete physical, mental and social well-being and not merely the absence of disease or in rmity is revealing both for what it says and for what it does not say. First, perfect health, i.e., a state of complete physical, mental and social well-being, may be a goal to be striven for, but it is probably unattainable. Second, if health is more than the absence of disease, then the prevention and treatment of disease alone will never be enough to achieve this. This suggests that the health care system, with its focus on the treatment and, to a lesser extent, the prevention of disease, is a necessary but not suf cient determinant of health.

Third, and perhaps most revealing in the context of this encyclopedia, is that the focus of the WHO de nition is on the individual (physical and mental well-being) and their social context (social well-being) but there is no reference to the environmental or ecological context. This probably re ects the tenor of the time, when there was little or no awareness of global change and what we now think of as *ecosystem health* (*see* **Ecosystem Health**, Volume 4). But today we recognize the importance of health not only as a human attribute but as an attribute of living systems as a whole, from local ecosystems to the global ecosystem that some call Gaia (Lovelock, 1979) (*see* **Gaia**, Volume 2). Such systems need to be balanced and in harmony, just as our own internal systems do, a process known as *homeostasis*. When our human systems are in such a state of balance, we are said to be healthy; health has become a useful metaphor for such a condition in ecosystems as well. Perhaps today, then, we would extend the WHO de nition somewhat to state that health is a state of complete physical, mental, social and ecological wellbeing ••• . In doing so, we would be returning to some much older theories of health.

HISTORICAL AND CULTURAL PERSPECTIVES

The concept that humans are somehow separate from, and superior to, the environment is a relatively recent one. It is primarily a product of Western thought in the past few hundred years, and has been particularly apparent in the last hundred. But for much of human history, humans have understood themselves to be a part of, and subject to the whims of, nature, often personi ed as some form of spirit or god. Good health resulted from living a good life in harmony with nature which included appeasing the gods and ill health was seen as the result of a loss of that harmony, a loss of balance, or a punishment from the gods. Healing then became a matter of restoring harmony and balance, in one s self, with the community, with nature and with the gods. Healers or shamans used a wide array of naturally occurring healing substances, and as well communicated with the spirits or gods in an attempt to restore balance and harmony and to drive out evil spirits. Such beliefs and cultural practices can still be found in many parts of the world, and in our multi-cultural societies can also be found in many Western societies.

Three great ancient healing traditions have been important in the evolution of modern understanding of health and healing, and the relationship between health and the environment. Shahi *et al.* (1997) argue that the fundamental beliefs and practices of those three traditions Greco-Arabic, Indian (Ayurvedic) and Chinese share much in common and that Evidence exists of signi cant cross-fertilization between these three great humoral traditions of medicine, with the exchange of ideas and products . Commonalities include:

- health is dependent on maintaining equilibrium among the *humors* (earth, water, re and, depending on the tradition, air, ether, wood and metal);
- this equilibrium is affected by such personal factors as age, sex, temperament, food consumption and nutritional status;
- all three systems took into account environmental factors such as the seasons, the cardinal directions (north,

south, east and west), the cycle of birth, growth and death and the condition of places, as well as emanations (*miasmas*) from decaying organic materials, from swamps and from the earth (miasma is a Greek word for pollution);
- there is an all embracing order of things which aligns the organization of human and other life, society and the universe, balancing polar opposites such as light/dark, moist/dry, heat/cold, etc.;
- many of the therapeutic interventions focused on modifying the patient's diet and surroundings, as well as using a wide range of medications (Shahi *et al.*, 1997).

The Greeks, for example, fully understood the need to balance both healthy living and healing, or treatment. Thus the Greek god of medicine, Asclepius, son of Apollo, had two daughters: Panakeia, the goddess of treatment (from whom we get the word *panacea*) and Hygeia, the goddess of healthy living, from whom we get the word *hygiene*.

A key figure is that of Hippocrates (460–377 BC), the Greek Father of Medicine, whose school was founded on the island of Kos. (Many physicians today still take the Hippocratic Oath on graduating.) Among the many works attributed to Hippocrates is one entitled Airs, Waters, Places , in which the relationship between health and the environment is discussed. It was felt important to pay attention to the quality of the air, the direction and qualities of the winds, the quality of the water, the local topography and climate and other factors, in choosing a place to live that would be healthy. Here we have a clear example of the way in which human health and human settlements were seen to be related to each other and to the quality of their environment. Indeed, by their commitment to natural law and to observation, the Greek physicians laid the groundwork for the science of medicine.

The last great physician in this tradition was Galen, who lived in the 2nd century AD. Writing about Galen, noting that three major factors were then believed to contribute to ill health – atmospheric corruption (miasma), individual predisposition and contagion – Winslow (1943) observed that:

> Here in the 2nd century AD were the three factors ••• which were to dominate scientific thinking up to the age of Pasteur. For more than 1600 years the history of epidemiology is a story of a shifting emphasis on these three basic conceptions.

The Romans were also well aware of the relationship between the environment and health, undertaking massive public works projects to provide clean water and to remove human wastes in their towns and cities. But with the coming of the Dark Ages in Europe, much of this knowledge of hygiene and healing was lost, although kept alive in the Islamic world, from where this knowledge was gradually re-acquired in the Middle Ages.

THE PUBLIC HEALTH APPROACH

The Renaissance, which flowered in northern Italy and spread across Europe in the 15th century, brought with it a renaissance of public health as well. Boards of Health were established in most of the major city-states of Northern Italy in the 15th century, primarily in response to epidemics of plague, a bacterial disease with a high mortality rate that is spread by the bite of infected fleas from rats to humans, as well as directly from humans to each other. Unhygienic and over-crowded living conditions undoubtedly hastened the spread of this dreadful disease. (For an excellent review of the broad historical sweep of this and other major diseases, their environmental and social roots and their impact on historical events such as the conquest of the Americas, see McNeill, 1976).

By the 16th century, the Boards' responsibilities included such diverse areas as hygiene, sanitation, food quality, burials, the regulation of hospitals, physicians and apothecaries and the control of prostitutes and vagrants. They had broad powers to regulate public gatherings, enforce quarantine laws, issue health passes and carry out inspections (Cipolla, 1976).

The public health approach of these times was still based primarily on the miasmic theory – the belief in the role played by atmospheric and environmental conditions in the transmission of epidemic diseases. Its rival, the contagion theory, was championed by Jirolamo Fracastoro (1484–1553). He suggested an alternative hypothesis, that epidemic spread of disease occurred as a result of the transmission of minute infective agents or spores from person to person either directly, via intermediaries, or through the air (Shahi *et al.*, 1997), although these agents were thought to be chemical, not biological in nature – bacteria had yet to be discovered.

Over time, these two theories became synthesized. But throughout this period from the Dark Ages to the 18th century, with some notable exceptions such as the activist Boards of Health in Renaissance Italy, the general attitude seems to have been passive – the environment was a given, and little could be done to change its impact on health.

However, with the beginnings of the scientific revolution in the 17th century and the Age of Enlightenment in Europe in the 18th century came a more critical, rationalist and activist ideology. This new activism, when applied to public health, became known as *social hygiene*. Its leading proponent was the Abbe Claude Fleury (1640–1723). This movement was the precursor to, and provided the foundation for, the public health movement of the 19th century.

Perhaps the greatest exponent of public health in the 18th century was Johann Peter Frank, Director-General of Public Health in Austrian Lombardy (Northern Italy) from 1786. This pioneer of social medicine established

hygiene as a systematic science and advocated the creation of state medicine. Frank taught that the greatest wealth of a state lies in its subjects, who should be as numerous, healthy and productive as possible ••• (Frank, 1976). Frank was concerned with the harmful impact of Man upon his environment. He had a healthy scepticism about the ability of physicians and hospitals to influence the health of a population, and he was concerned with the deleterious effect upon health that a poor life style had, whether that lifestyle was chosen or imposed. He saw the role of public health as helping to maintain the security and stability of the state, which he viewed as a benign and paternalistic protector of its people.

The need to maintain the security and stability of the state was also of great concern to Edwin Chadwick, the English lawyer and public servant who is seen as the father of modern public health. In an article written in 1828, Chadwick contended that the environment must influence health and wellbeing, which in turn are the pillars of social stability, and that the best means to ensure social stability was for the state to promote health and wellbeing by improving the environment (Ringen, 1979).

Chadwick was one of the chief architects of the new Poor Law of 1834, which helped make industrialization possible by establishing the mobility of labor. But within a few years Chadwick became concerned that the poor social and health conditions of the laboring class might result in unrest, even revolution. Thus the famous Sanitary Report of 1842, which examined the health conditions of the laboring class, while undoubtedly sparked by an humanitarian concern, was also concerned with improving conditions enough to avert serious social unrest. Frederick Engels powerfully portrayed those conditions in 1845, when he described Manchester's River Irk as:

> a narrow, coal-black, foul-smelling stream ••• in dry weather, a long string of the most disgusting, blackish-green slime pools are left standing ••• from the depths of which bubbles of miasmatic gas constantly arise and give forth a stench unendurable even on the bridge 40 or 50 feet above the surface of the stream.
>
> (quoted in Girouard, 1985)

while Hippolyte Taine wrote of his visit to Manchester in 1859:

> Earth and air seem impregnated with fog and soot. The factories extend their flanks of foul brick one after another, bare, with shutterless windows, like economical and colossal prisons ••• through half-open windows we could see wretched rooms at ground level, or often below the damp earth's surface. Masses of livid children, dirty and flabby of flesh, crowd each threshold and breathe the vile air of the street, less vile than that within •••.
>
> (quoted in Girouard, 1985)

It was environmental conditions such as these that led to the British government establishing the Health of Towns Commission in 1843. Again led by Chadwick, the Commission's work led to the adoption of the sanitary idea and to the establishment of public health measures such as housing standards, sewer systems, hygiene regulations and proper public water supplies. A Health of Towns Association, established in 1844, with branches all over Britain, supported the Commission's work.

> The work of the Health of Towns Association in pressing for the application of the sanitary idea and its insertion into policy making had a dramatic effect on public health in Britain in a comparatively short space of time.
>
> (Ashton, 1992)

Thus by the middle of the 19th century, the contribution of filthy environments and appalling living and working conditions to the health of the population was well established, along with a strong and active public movement to support social and political intervention to modify the environment. Its apogee, perhaps, came in the 1870s when Sir Benjamin Ward Richardson, a disciple of Chadwick, marked the passage of the great Public Health Act of 1875 by describing his vision of a city of health:

> a community so circumstanced and so maintained by the exercise of its own free will, guided by scientific knowledge, that in it the perfection of sanitary results will be approached, if not actually realized, in the co-existence of the lowest possible general mortality with the highest possible individual longevity.
>
> (Richardson, 1875)

But through one of those ironic twists of fate, the 1870s were not only the time of the Great Public Health Act and of Benjamin Ward Richardson's vision of a healthier city, but were also the time at which the bacteriologists were announcing many of their most exciting discoveries about bacteria, insect vectors, animal reservoirs of disease and human carriers and were developing effective vaccines and anti-toxins. These discoveries ushered in a very different theory of health, one that was to become the dominant model in much of the world over the next century.

THE MEDICAL MODEL OF HEALTH

Throughout history, the Hygeian or public health and Panakeian or medical/therapeutic models have co-existed and contended. For much of the 19th century, the Hygeian model was dominant. But with the identification of bacteria and the refinement of biological and clinical research, the emphasis shifted towards the Panakeian model, with its emphasis on finding the specific cause of disease in bacteria, in nutritional and metabolic disorders, in biochemical and molecular changes. In this medical model, health is seen primarily as the absence of disease.

The medical model has become powerful because it has been allied to the scientific method and the dominance

of Western culture and civilization in the past couple of hundred years. This analytic model has focused its increasingly narrow attention on organs, on cells, on microorganisms, on genes, and on molecules as causes of disease. But this focus has also tended to exclude the broader environmental, social and economic factors at play. In essence, the medical model is a reactive model, looking for disease and then seeking a treatment for it (a *pathogenic model*) rather than seeking to understand what determines good health and how to maintain and improve it (a *salutogenic model*, as Antonovsky (1978) termed it).

At its simplest, the medical model sees a single cause – the measles virus, for example – and a single effect – the disease of measles. This simple cause and effect model lies behind the search for specific treatments for specific diseases, a search that has had some remarkable successes. But it also quickly became apparent that simple exposure to a pathogenic organism such as the measles virus, while necessary for the disease to occur, was not sufficient. The epidemiological triangle of agent, host and environment, sought to explain this by identifying differences in the environment (which might affect transmissibility, for example), in the host (who might be resistant) or in the agent (which might differ in terms of dose, virulence, etc.). Further complicating matters is that some causes (smoking, for example) might have multiple effects (several forms of cancer, heart disease, stroke, respiratory disease, low birth weight and many other diseases) while some diseases (heart disease, for example) might have multiple causes (inherited susceptibility, high blood pressure, smoking, high cholesterol levels, infection, lack of exercise and other factors).

The medical model of health has had some spectacular successes, particularly in the development of antitoxins and vaccines to protect against infectious disease (the eradication of the scourge of smallpox is a high point), antibiotics to treat bacterial (but not viral) disease, medications to control high blood pressure and many other diseases, and some forms of cancer therapy. In recent years a vast array of chemotherapeutic agents has been developed for all manner of minor and major, chronic and acute, physical and mental health problems, as well as surgical treatments for both major and minor health problems. In addition, Western medicine has shown itself to be remarkably effective at dealing with traumatic injuries that in the past would have been fatal. Indeed, so widespread and pervasive has this model become that the concept of medicalization has been developed to describe the extent to which medicine infiltrates and even distorts our lives and our societies. Illich (1976) has famously discussed *iatrogenesis* (doctor-caused disease) as not simply clinical disease resulting from errors, unanticipated side effects and the over-use of diagnostic and therapeutic techniques, but also has described social and cultural iatrogenesis – the medicalization of many normal parts of life such as feelings of dis-ease, validation of disability or work and school absenteeism, or the processes of birth and death.

This growth in modern Western medicine has also occurred at the same time as a massive increase in life expectancy, which for all its faults remains the most commonly used overall index of health. Life expectancy at birth for women in Canada, for example, increased from 43.8 years in 1871 to 76.4 years in 1971 (Bourbeau and Legare, 1982), while in Chile male life expectancy at birth increased from 29 years around 1910 to 72 years in 1998 and female life expectancy at birth in Japan in that same period increased from 43 to 83 years (WHO, 1999). Globally, life expectancy at birth has increased from 48 years in 1955 to more than 65 years in 1995, and is projected to increase to 73 years by 2025 (WHO, 1998). On a regional basis, average life expectancy at birth in 1997 ranged from 53 in the African Region of WHO to 72 in the European region and 73 in the American region. The largest increases in life expectancy at birth between 1975 and 1997 were found in the Eastern Mediterranean region (12 years) and the Southeast Asian region (11 years), with more modest but still substantial gains of 6–7 years in the African, American and Western Pacific regions, while the European region posted a gain of 2 years (WHO, 1998).

As a result, there has been a widespread belief that our long lives and better health are attributable to modern Western medicine, which has only served to strengthen the dominance of the medical model. However, since the 1970s new theories of health have been emerging, rooted in a questioning of the efficacy of the medical model and in a more salutogenic orientation.

NEW PERSPECTIVES

One aspect of the questioning of the medical model comes from the psychological and social sciences. For while the medical model of health focuses mainly on the bio-physiological dimension, there is also a psycho-social dimension to disease, a perception that we call *illness*. Moreover, people may feel ill without there being any evidence of bio-physical disease. Norman White, a Canadian psychiatrist, has suggested a bi-axial model, one axis representing the biophysical spectrum from disease to physiological health, the second axis representing the psycho-social spectrum from illness to a sense of well-being. In such a model, people may be physiologically healthy but psychosocially ill, or they may have physiological disease but feel well (undiagnosed high blood pressure would be a good example of this). We are not simply biological systems, we are also psycho-social beings, and theories of health must recognize this fact.

A second powerful response to the power of the medical model – and its failings – has been the (re)-emergence of a more holistic model of health that pays greater attention to the relationship between body, mind and spirit, an area generally neglected by modern Western medicine. This has involved a dramatic growth in the desire of knowledgeable and empowered consumers to be much more involved in their own health care, both through self-care and mutual support and through an increased emphasis on healthy living and the personal use of natural therapies. At the same time practitioners of holistic, alternative or complementary medicine in the West have been rediscovering and re-applying insights and wisdom from traditional healing practices rooted primarily in Eastern philosophies. Naturopathic, homeopathic and ayurvedic methods, acupuncture, shiatsu, massage, meditation and many other healing modalities have been growing rapidly in the West in the past few decades, in response to the growing public disillusion with the hard science values and practice of modern Western medicine. Their emphasis on healthy living, on caring, on re-establishing balance and harmony and on the use of natural remedies links them not only to the ancient healing traditions discussed earlier, but to modern concepts of wellness and ecosystem health.

While holistic medicine was and is a critique of the clinical aspects of the medical model, a critique of the role of medicine in improving health was also developing in the 1960s and early 1970s. Thomas McKeown – Professor of Social and Preventive Medicine at the University of Birmingham in England – began to inquire into the factors that led to improved health in England and Wales in the 19th century. His answers were somewhat surprising; he showed that the major factors leading to improved health had been improved nutrition, improved hygiene, limitations in family size (all of these rooted in economic and social development) and the development of vaccines in the late 19th and early 20th century and of antibiotics in the 1930s (McKeown, 1979). In short, McKeown found that improvements in health had not been primarily due to improvements in medical care, but to the combination of social and economic development.

McKeown's work formed the basis of a seminal Canadian report in 1974 that, for the first time, suggested that a national, and for that matter world-wide, policy focus on provision of health care was insufficient. The Lalonde Report – *A New Perspective on the Health of Canadians* – proposed the health field concept in which four factors – environment (physical and social), lifestyle, health care organization and human biology – were the key determinants of health (Lalonde, 1974). Importantly, the Lalonde Report proposed that major improvements in the health of Canadians will result primarily from changing lifestyles, improving our environments, and advances in our understanding of human biology . The Lalonde Report was influential not only in Canada but also in the US and in Europe, where it spawned similar reports in the years that followed. Over the past 25 years, the analysis has been refined.

The initial emphasis – and still a strong focus today – was on lifestyle. One estimate suggested that up to 50% of preventable mortality was attributable to lifestyle diseases, which were sometimes called diseases of choice. The implication was that if only people would make healthy and wise choices, much ill-health and premature death could be avoided. The wellness movement that emerged in the 1970s focused on the individual's personal responsibility to avoid smoking and excessive drinking, to improve fitness and nutrition, to learn stress management and to develop environmental awareness and a sense of spiritual and emotional well-being (Ardell, 1977). However, critics of this approach, while acknowledging that healthy choices are important, pointed out that environmental, social, economic and cultural circumstances and our living and working conditions shape and sometimes constrain those choices. Over time, this led to the emergence of the concept of *population health*.

Taking their cue from the work of McKeown, advocates of population health (e.g., Evans *et al.*, 1994) have pointed to the significance of the economic and social determinants of health, drawing attention in particular to the importance of poverty and inequality and the influence of social status on health. This in turn has been linked to factors such as self-esteem and sense of self-worth, to the influence of the social hierarchy and of the degree of control over working conditions, and to the relationship between the mind and the body's immune system, a field of study known as *psycho-neuro-immunology*. The research findings suggest that such factors may be as important as conventional risk factors such as smoking, diet and exercise in the development of heart disease. For example, the 1999 World Health Report cites evidence that suggests that the most important factors that have led to the reduction in mortality between 1960 and 1990, based on data from 115 low and middle income countries, were ascribable to the generation and utilization of new knowledge (which includes but is much more than medical knowledge and technology), improvements in the educational level of women and improved income, in that order (Wang *et al.*, 1999, cited in WHO, 1999). However, while strong on the social and socio-economic factors, the population health approach is curiously silent on the influence of the built and natural physical environments on population health status.

The neglect of the built environment within which the social processes of population health occur is particularly unfortunate because the most significant human environment today is the built environment. Globally, half of humanity now lives in urban settlements, while in the

developed world some 80% of the population is urbanized. Moreover, humans spend most of their time indoors, especially in the developed world. In North America, people are indoors almost 90% of the time and in their cars 5% of the time, leaving only 5% when they are outdoors (Leech *et al.*, 1996). The built environment is a human artifact comprising both a physical and a social environment, and it constitutes the human ecosystem. Understanding human health increasingly means understanding the effects of complex urban physical and social environments and processes upon the health of urban residents. It also means understanding how the urban way of life affects the health of the wider natural ecosystems of which they are a part.

The ecological perspective on health has emerged from the growing importance of ecology, and from the growing awareness of global change and its health implications; human health is inextricably linked to ecosystem health. The view of earth from space has made it apparent, as never before, that we are bound to and part of a fragile and beautiful place, a living planet that some call Gaia. At the same time, this global perspective and the modern science of ecology has enabled us to understand how potent is our power to affect the global life support systems on which we depend.

For while the increase in life expectancy and other measures of health that we have experienced may seem to be due primarily to the dramatic social and economic development we have undergone, this development has in turn been based upon industrialism s exploitation of the earth s resources notably energy, forests, soils, minerals and the oceans and the accompanying widespread pollution of our environment, indeed of the planet as a whole. Thus, in a very real sense, our current high level of health and long lives have been purchased at the expense of the environment.

But global change is not without cost (see McMichael, 1993, or Chivian *et al.*, 1993, for good overviews of the health implications of global change) (*see* **Environmental Change and Human Health: Extending the Sustainability Agenda**, Volume 3). Global warming, ozone depletion, acid rain and other climate and atmospheric changes threaten our health; global warming in particular will increase the range of the mosquitoes that spread malaria, dengue fever and other diseases, as well as the distribution of other disease vectors. More severe storms, oods and droughts will also add to the burden of infectious diseases as well as causing hunger, malnutrition and trauma. The depletion of such renewable resources as sheries, forests, topsoil and fresh water aquifers threatens our food supplies and may cause both malnutrition and mass migration, with all its attendant misery, diseases and deaths. Environmental pollution and ecotoxicity the widespread contamination of entire ecosystems and food chains with persistent organic pollutants (POPs) has made it apparent, as the North American Indian Chief, Seattle, is reputed to have remarked more than 100 years ago, that we are all part of the web of life, and whatever we do to the web of life, we do to ourselves . We simply cannot tell what a lifelong exposure to low levels of multiple POPs will mean for the health and well-being of today s infants, nor for the health of other species that are equally exposed. This applies also to the loss of habitat that we are causing and the accompanying loss of biodiversity as a result of a human-created mass extinction; we do not know how many strands of the web of life we dare sever. And the massive increase in trade and travel that has accompanied globalization has meant that exotic diseases (such as West Nile fever, spread by birds and mosquitoes) or exotic vectors (such as the Asian Tiger mosquito, a hardy mosquito that has been introduced into the USA and that spreads dengue and yellow fever) will become more widespread.

There is thus a growing awareness that human health is dependent upon ecosystem health, and even a concern that human health can no longer be sustained if the industrial and economic growth that has been the basis of improved health undermines the ultimate determinants of health the Earth s life support systems and natural resources (World Resources Institute, 2000).

A NEW SYNTHESIS

Any modern theory of health thus has to take into account and integrate the bio-medical model (including its complementary or alternative versions), the wellness and lifestyle approach, and the social, economic, built and natural environmental determinants of health.

In his depiction of a sane, humane and ecological society, the British futurist Robertson (1978) described sanity as balance within the individual, humanity as balance between the individual and other people in their communities and around the world, and ecology as balance between humans and nature. These three concepts in many ways represent the three broad themes in the re-conceptualization of health described above that have emerged in the late 20th century: holistic medicine and wellness, with its focus on the individual s balance and harmony; population health, with its focus on social conditions and processes; and ecosystem health, with its focus on the relationship between human health and the health of the planet. They all have roots in much earlier models of health.

- A focus on holistic medicine and wellness harkens back to much older views of the importance of diet and nutrition, of maintaining balance in leading a healthy life, and in forms of treatment which address the body, mind and spirit together. It might perhaps be considered part of the Panakeian tradition.

- Improved living and working conditions were a major factor in improving health in the 19th century, and population health reminds us that social and economic development remains important today, while the modern-day Healthy Cities and Communities movement takes its inspiration directly from its Victorian predecessors. As in Hippocrates' time, we still have to be concerned with airs, waters and places, with the quality of our human settlements and our buildings. Maintaining and improving urban and suburban environments and the social and physical quality of the homes, schools, workplaces, institutions and other settings where we spend most of our time remains an important task for those concerned with public health. This is the Hygeian tradition.
- We are rediscovering that in order to be healthy we have to be in tune with nature, both at a local level and globally. As with the re-emergence of holistic medicine, this is in some respects a return to our roots, to an awareness of our place within, rather than apart from and superior to, nature. Healthy people can only exist within healthy communities and on a healthy planet. This might be considered the Gaian tradition.

A synthesis of these three concepts may lead to a new understanding of health. What follows is a model of the human ecosystem that provides such a synthesis, a theory of health for the 21st century that seeks to integrate these new theories of health.

AN ECOLOGICAL MODEL OF HEALTH

All models reflect the contemporary intellectual and cultural context in which they are created, and, of course, all models are wrong, in the sense that they are only a model of reality, not reality itself, and that they reflect the personal context of those who create the model, which in turn reflects the cultural, social and intellectual context of their times. That having been said, a modern theory of health has to incorporate the biomedical, holistic medicine, wellness, population health, and ecosystem health perspectives. The Mandala of Health (Hancock and Perkins, 1985; Hancock, 1985) is one such model (see Figure 1), reflecting our current understanding of the elements of the human ecosystem that determine health. Developed in the early 1980s to explain the new public health approach in the City of Toronto in Canada, it builds upon and expands the four health fields of the Lalonde Report.

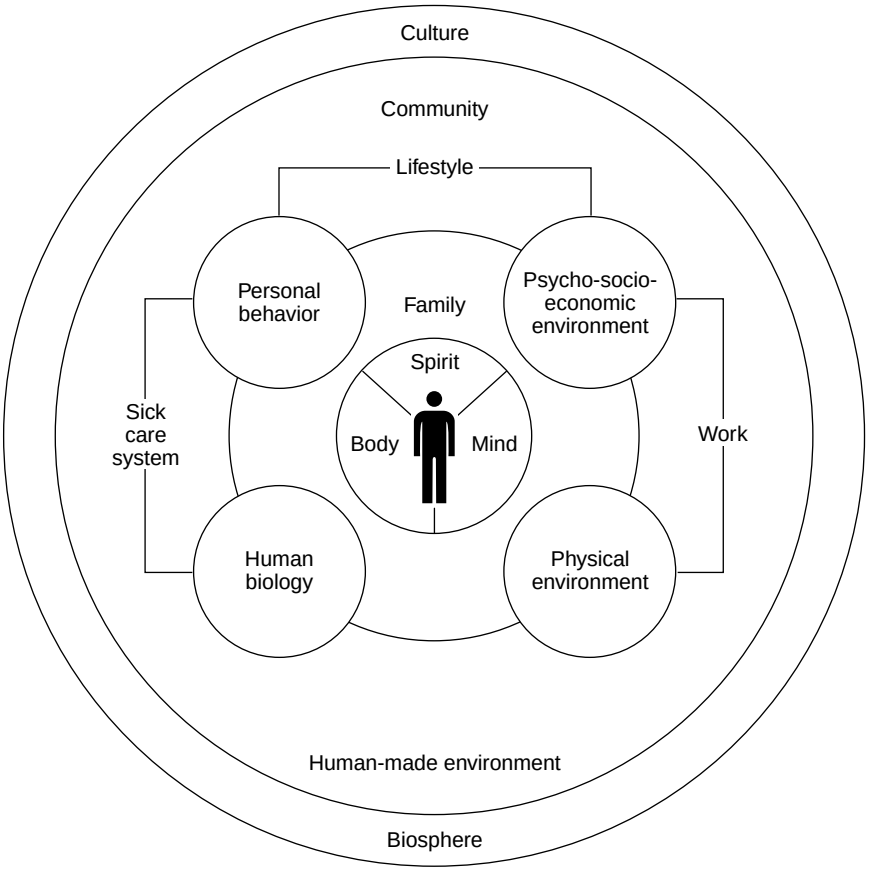


Figure 1 The Mandala of health. A model of the human ecosystem

Mandala is the Buddhist word for an intricate circular design used as a focus for meditation. It has also been defined as a circular symbol of the universe: it is in this latter sense that the term is used; the Mandala is a circular symbol of the universe of health. At the center of the Mandala, the focus for our concern, is the individual human being. That individual's health is a function of their body/mind/spirit integration and well-being, the central concept of holistic medicine and wellness.

That individual is shown as existing within a family. For most of us, for most of our lives and in particular in those important but vulnerable parts of our lives encompassed by infancy, childhood, adolescence, parenthood and old age, the family is our most immediate social environment, just as our home is our most immediate physical environment. The family constitutes one of the great mediating structures of our society: it buffers and protects the individual members from the impacts of community and society; it establishes our values, attitudes, standards and behaviors; it supports us and protects us. (Of course, it must be admitted, families can also have a negative effect, causing stress, even mental and physical damage, the latter through violence.)

The individual and his/her family are in turn affected by four key factors – human biology, personal behavior, psycho-socio-economic environment and physical environment. These four are in fact variants of three of the four Lalonde health fields, namely lifestyle, environment and human biology.

- *Human biology* affects personal and family health through such things as the genetic transmittal of mutations and inborn errors of metabolism, through physiological and biochemical functioning, the competence of the immune system (and the links between the psyche and the immune system), and through the aging process, to name but a few.
- *Personal behavior* is obviously a determinant of health, although constrained and modified by many other factors. The behaviors we are concerned with can be loosely grouped into health promotive or protective behaviors such as proper eating habits, taking regular exercise, managing stress effectively, using seatbelts and crash helmets, and using preventive and other health care services appropriately; and health risk behaviors which include smoking, drug and alcohol abuse, driving dangerously or while drunk and so on.
- The term *psycho-socio-economic environment* addresses two related areas of great importance. The first is concerned with our psychosocial environment, including pressure from our peers at school, at work and in other situations, the influence of media and advertising and related factors. By socioeconomic environment, we are referring primarily to the influence of social status, class or income, which have profound influences on people's knowledge, skills, resources and opportunities, and on their living and working conditions, as they attempt to lead healthy lives.
- By *physical environment*, in this particular section of the model, we are referring primarily to the immediate indoor environment of home, school, workplace and other indoor settings where most people, especially in industrialized countries, spend so much of their lives. The design of our buildings, their air quality and the hazards (physical, chemical, mechanical and biological) they contain have important effects upon health.

The next level of the Mandala shows three particular connections between these four factors that are of importance to health. The first of these is called the sick care system. This term is used specifically to indicate that what we usually refer to as the health care system is in reality primarily concerned with sickness rather than with health. This is not to belittle the system, or those who work within it, far from it. The role they perform is indeed a vital one and when we become sick we would all like to have an effective and efficient sick care system to care for us. But those working within that system have primarily been trained in the recognition, diagnosis and treatment of disease, and are often less knowledgeable about the maintenance and promotion of health, skills in which they have not usually been trained.

A second important point is that the sick care system is primarily concerned with human biology and personal behavior (in the sense of physical and mental health) and is not very involved with – and often not very concerned about – the other elements of the Mandala. This is why the sick care system is only a relatively minor determinant of health.

Work forms a very important function in our lives, at least in the Western industrialized culture. Not only is the physical environment of the workplace frequently a hazard to health, so too is the mental and social environment of the workplace, which can be a source of stress, alienation, boredom or frustration. On the other hand, work is an important part of our social identity, can often be an important source of meaning and purpose in our lives and provides one of our key social networks.

The term *lifestyle* is here shown to be the consequence of the interaction of personal behavior with the psycho-socio-economic environment, within the context of community and culture. Lifestyle is a complex phenomenon, not simply a matter of personal behavior and personal choice. As an example, imagine trying to lead the lifestyle of a Buddhist forest monk in downtown Toronto in the middle of winter. While such a lifestyle choice is clearly possible, it is also clearly a difficult one to make and to sustain, given our social and cultural environment and our physical environment. On the other hand, it would clearly be

quite easy to choose such a life in a country like Burma. The point of this example is not to deny the existence of free choice, merely to indicate that choice is constrained and limited by broader social, cultural and environmental factors.

All of these factors function within a third level of the Mandala, namely that of the community and the human-made environment. As already indicated, the community has its own set of values and standards, and its own social and cultural characteristics that influence our health. Human services are organized at the community level in the areas of health, education, social services, recreation and other areas by the public, private/for-profit and voluntary/non-profit sectors. In addition, the community has a broad range of organizations, groups and networks which provide what McKnight (1995) has termed *associational life* or what Putnam (1993), among others, refers to as *social capital* – churches, self-help networks, neighborhood associations, service clubs, unions, school and home associations, and so on. They play a vital role in our health by providing the social networks and social support necessary for healthy social functioning.

The human-made environment encompasses the built environment of neighborhoods and urban settlements. But the notion of the human-made environment also encompasses the human-modified environment – massive agricultural, energy, transportation and other systems which have left virtually no segment of the world untouched or unmodified by human hands. All of these systems, and all of this human modification have enormous implications for health.

The final level of the Mandala encompasses culture and the biosphere. These are the overarching contexts within which we lead our lives. Culture refers to cultural norms and values that underlie how we think about health, how we believe health and illness are determined and what steps we take as a society and as individuals to improve health or treat illness. In countries as culturally diverse as Canada and the United States now are, there are many different cultural norms and values at work, from those of the aboriginal people to the cultural beliefs of recent immigrants from Asia, Africa, Latin America and the Caribbean. Nonetheless, the basic cultural values and beliefs are rooted in the Judaeo–Christian heritage, Western ideologies of liberty and democracy, scientific rationality, technological civilization and domination of nature. These all greatly affect our understandings of and approach to health.

The biosphere is the ultimate determinant of health. The biosphere is the term used to describe that thin shell of the planet that contains plant and animal life. It is this thin skein of the web of life that keeps us alive, providing our life support systems and a range of other eco-services. The rediscovery in recent years of the ecological principles that our ancestors and aboriginal people understood so well has led us to accept that we are a part of and within nature, rather than separate from and dominant over nature. Only through such an understanding will we come to accept, as the 1990 World Health Day slogan puts it, that we really should think of our planet, our health'.

The Mandala of Health has proved to be a useful conceptual model for understanding health, and an important teaching tool. As with all such models, it will no doubt be superseded with time. But for the moment, it is a useful representation of our modern theory of health.

See also: Infectious Diseases, Volume 2; **Environmental Change and Human Health: Extending the Sustainability Agenda**, Volume 3; **Urban Climate and Respiratory Disease**, Volume 3; **Urban Poverty and Environmental Health**, Volume 3; **Policy Responses to Public Health Issues Relating to Global Environmental Change**, Volume 4; **WHO (World Health Organization)**, Volume 4.

REFERENCES

Antonovsky, A (1978) *Health, Stress and Coping*, Jossey Bass, New York.

Ardell, D (1977) *High Level Wellness: an Alternative to Doctors, Drugs, and Disease*, Rodale Press, Emmaus, PA.

Ashton, J (1992) The Origins of Healthy Cities, in *Healthy Cities*, ed J Ashton, Open University Press, Milton Keynes.

Bourbeau, R and Legare, J (1982) *Evolution de la Mortalite au Canada et au Quebec, 1831 1931*, Les Presses de l Universite de Montreal, Montreal.

Chivian, E, McCally, M, Hu, H, and Haines, A (1993) *Critical Condition: Human Health and the Environment*, MIT Press, Cambridge, MA.

Cipolla, C (1976) *Public Health and the Medical Profession in the Renaissance*, Cambridge University Press, Cambridge.

Evans, R, Barer, M, and Marmor, T, eds (1994) *Why Are Some People Healthy and Others Not? The Determinants of the Health of Populations*, Adeline de Gruyter, New York.

Frank, J P (1976) *A Complete System of Medical Policy*, ed and translated by Lesky, originally published 1786 1810, Johns Hopkins University Press, Baltimore, MD.

Girouard, M (1985) *Cities and People: a Social and Architectural History*, Yale University Press, New Haven, CT.

Hancock, T (1985) The Mandala of Health: a Model of the Human Ecosystem, *Fam. Community Health*, **8**(3), 1 10.

Hancock, T and Perkins, F (1985) The Mandala of Health a Conceptual Model and Teaching Tool, *Health Educ.*, **24**(1), 8 10.

Illich, I (1976) *Medical Nemesis*, Pantheon Books, New York.

Lalonde, M (1974) *A New Perspective on the Health of Canadians*, Of ce of the Canadian Minister of National Health and Welfare, Ottawa.

Leech, J A, Wilby, K, McMullen, E, and Laporte, K (1996) Canadian Human Time-Activity Pattern Survey Report and Population Surveyed, *Chronic Dis. Can.*, **17**, 118 123.

Lovelock, J E (1979) *Gaia: a New Look at Life on Earth*, Oxford University Press, Oxford.
McKeown, T (1979) *The Role of Medicine: Dream, Mirage or Nemesis*, Blackwell, Oxford.
McMichael, A (1993) *Planetary Overload*, Cambridge University Press, Cambridge.
McNeill, W (1976) *Plagues and Peoples*, Anchor Press, Garden City, New York.
McKnight, J (1995) *The Careless Society: Community and its Counterfeits*, Basic Books, New York.
Putnam, R (1993) *Making Democracy Work*, Princeton University Press, Princeton, NJ.
Richardson, B (1875) *Hygeia: a City of Health*, MacMillan, London.
Ringen, K (1979) Edwin Chadwick, the Market Ideology and Sanitary Reform, *Int. J. Health Serv.*, 9(1), 107–120.
Robertson, J (1978) *The Sane Alternative*, James Robertson, London.
Shahi, G S *et al.* (1997) A Historical Perspective, in *International Perspectives on Environment, Development and Health: Toward a Sustainable World*, eds G S Shahi *et al.*, Springer, New York.
WHO (1999) *World Health Report*, WHO, Geneva.
WHO (1998) *World Health Report*, WHO, Geneva.
Winslow, C E A (1943) *The Conquest of Epidemic Diseases: a Chapter in the History of Ideas*, Princeton University Press, NJ.
World Resources Institute (2000) *World Resources 2000- 2001: People and Ecosystems - The Fraying Web of Life*, Oxford University Press, New York.

Thoreau, Henry David

(1817- 1862)

Henry David Thoreau, US writer, poet and philosopher, is considered to be the godfather of American environmentalism, due to his authorship of *Walden*, his ascetic lifestyle as illuminated by *Walden*, and his later contributions to natural history.

Thoreau was born in Concord, MA, in 1817, and was influenced early on by a powerful intellectual and spiritual movement in New England called the transcendentalists, of which the most famous figure was Ralph Waldo Emerson (1803–1882). This movement sought to apply the teachings of German Romanticism in an American context. These teachings sought spiritual understanding in generating powerful connections between the newly emerging individual in modern society, and the unspoiled natural world. Thoreau, and others more loosely associated with the transcendentalists, were open to many influences, including the earliest translations in the west of Buddhist and Hindu writings.

Upon graduation from Harvard in 1837, Thoreau began a haphazard career as a writer, land surveyor, worker in his father's pencil factory, and occasional lecturer. His first book, *A Week on the Concord and Merrimack Rivers* (1849) which described a boat journey Thoreau and his brother took (embellished with natural history, stories, and embedded short essays on a variety of subjects), was a complete failure. Thoreau continued to publish natural history essays and travel essays in American magazines throughout the rest of his life, of which the most famous are probably *Ktaadn and the Maine Woods* (1848), *Cape Cod* (1855) – both later turned into books – and *Walking* (1862). His essay *Civil Disobedience* (1849), written in opposition to the war with Mexico, has been widely influential in setting out the citizen's responsibilities in the face of what can be considered immoral acts by one's own government.

His most famous book *Walden* is a composite diary of the years 1845–1847, during which Thoreau chose, as an experiment, to build a hut and live a marginal life beside Walden Pond, approximately 3 km away from his native Concord. The book deliberately challenges many of the working assumptions of the culture of his time, particularly the obsessions with work, property, and getting on. In chapters such as *Economy*, Thoreau parodies these obsessions, and simultaneously promotes a range of other values: voluntary simplicity, humility, and simply watching the world pass by. *Walden* weaves together in beautiful, clear prose, these social and political themes, together with details of the natural history of the pond and its environs. Since its publication in 1854, *Walden* has become a touchstone of the environmental movement: evidence of this can be seen in the very public, and ultimately successful struggle in the 1990s to protect Walden Pond from encroaching development.

In recent years, more attention has been paid to Thoreau's *Journal*, his daily record of natural history that he kept intermittently from 1838, and then rigorously from the mid-1840s on. There are indications that he intended turning this into a larger natural science book, but died before this could be carried out. Scholars currently see this later more intense focus on the details of natural processes as symptomatic of a continuing dynamic tension between the demands of literary art and the rigors of science in Thoreau's late work.

The Succession of Forest Trees (1859) is generally regarded as Thoreau's main contribution to natural history, in which he sets out the first scheme of ecological succession. In this period, he also became widely known for his support for anti-slavery, and his lectures, including *The Last Days of*

John Brown (1860), brought him to the attention of a wider audience. Thoreau died of tuberculosis in 1862. The elegy at his funeral by his mentor Ralph Waldo Emerson was also influential in bringing Thoreau to public attention. By the turn of the 20th century he was firmly in the pantheon of great American writers; and his well-wrought phrases, such as In Wildness is the Preservation of the World have entered the environmentalist armory.

Photo: © G F Parlow.

FURTHER READING

Bode, C (1975) *The Portable Thoreau*, revised edition, Viking Penguin, Harmondsworth, Middlesex.

Buell, L (1995) *The Environmental Imagination: Thoreau, Nature Writing, and the Formation of American Culture*, The Belknap Press, Harvard, Cambridge, MA.

Harding, W (1982) *The Days of Henry Thoreau*, Alfred A Knopf, New York.

PETER TIMMERMAN *Canada*

Three Mile Island

Three Mile Island (TMI), a commercial nuclear reactor 16 km southeast of Harrisburg, PA, was the site of the most serious reactor accident in North America.

On March 28, 1979, at approximately 4 a.m., Unit 2 at TMI moved to an automated shutdown. The reactor technicians on duty inadvertently uncovered the core of the reactor creating conditions that could lead to a meltdown. A valve in the reactor, forced open by the sudden and violent shutdown, had stuck open, allowing cooling water to pour from the reactor. The auxiliary pumps came on, as programmed to ensure adequate water to cover the core. Unaware of the stuck valve, the technician on duty, Craig Faust, believed the core to be receiving too much water and shut off the pumps. Operators also opened another drain line to remove what was interpreted as dangerously high water levels. Remaining water turned to steam, the uranium rods overheated and ruptured. Radioactive water was discharged outside the reactor, and radioactive gases, were vented.

Not until the morning of March 30, 1979 did the governor suggest that pregnant women and families with small children might want to leave the area. Over the weekend, an estimated 140 000 people evacuated.

Estimates by the Nuclear Regulatory Commission (NRC) study group were that TMI had come within thirty to sixty minutes of meltdown. Previous NRC worst case analyses had estimated that a major nuclear reactor meltdown could create an area of ••• disaster that might be equal to the State of Pennsylvania. Coincidentally, as TMI occurred, a major Hollywood movie about a nuclear plant accident, *The China Syndrome*, was playing. A scientist character in the film quoted the NRC area the size of Pennsylvania destruction estimate.

Estimates of deaths due to the accident vary from several hundred excess cancers to several thousand. There has not been a single new nuclear reactor built in the United States or Canada since the TMI accident.

ELIZABETH MAY *Canada*

Torrey Canyon

Torrey Canyon, an oil tanker which ran aground in the English Channel in 1967 was the world's first major oil spill. Torrey Canyon became a spark for the growing environmental movement in the late 1960s. It remains one of the most devastating oil spills in history.

The Torrey Canyon spilled 31 million gallons (117 million liters) of oil into the English Channel, resulting in the death of between 40 000 and 100 000 seabirds. Later well-known oil spills released a fraction of the cargo of the Torrey Canyon. The Amoco Cadiz spill in March 1978 off the coast of Brittany coated an estimated 100 km of coastline, but spilled far less oil, 1.5 million barrels of crude oil. In July 1983, the Alrenus wrecked on the coast of Louisiana spilling 2.27 million gallons (eight million liters) of oil. The Exxon Valdez spill (*see* Exxon Valdez Oil Spill, Volume 5) lost about one-fifth of its cargo, 11.2 million gallons (42 million liters) with devastating environmental effects.

ELIZABETH MAY *Canada*

UVWY

Underdevelopment

see Development and Global Environmental Change (*Opening essay,* **Volume 5**)

UNEP Chemicals

see IRPTC (International Register of Potentially Toxic Chemicals) (**Volume 4**)

United Nations University (UNU)

see UNU (United Nations University) (**Volume 5**)

UNU (United Nations University)

The UNU is the principal academic research institution of the United Nations (UN). It is an organ of the UN similar to the UN Development Programme or the UN International Children's Emergency Fund, and reports to the General Assembly and the UN Economic and Social Council. Its role is to focus intellectual resources from many nations on world problems of interest to the UN system. It organized the first conference on the Human Dynamics of Global Environmental Change in 1988, pioneered intergenerational environmental law, led the Mountain Agenda at the UN Conference on Environment and Development, and other environmental efforts.

This post-graduate research and training institution does not have professors or students. Instead, the UNU is a worldwide network of networks of scholars whose purpose is to develop options for the solution of problems that require truly international collaboration. It is the UN's think tank and has published hundreds of books, facilitated thousands of research fellows, and conducted hundreds of conferences and workshops to further knowledge of human security and development. See http://unu.edu.

UNU has scholars instead of countries on its governing council and funds about half of its research though an endowment. These approaches helps ensure academic autonomy.

Although headquartered in Tokyo since it opened in 1975, UNU has research and training centers and programs around the world with their own governing councils and funds in areas such as: development economics in Finland, new technologies in the Netherlands, software technology in Macau, advanced studies in Japan, natural resources in Ghana, biotechnology in Venezuela, leadership in Jordan, regional governance in Belgium, environmental health in Canada, food and nutrition in the US, fisheries in Iceland, and conflict resolution in the UK.

UNU headquarters has two primary programs: (1) peace and governance; and (2) environment and sustainable development. The second of these focuses on human–environment dynamics including environmental governance, policy, and advanced training especially in developing countries. Its research contributes to the work of the UN system and seeks partnerships with others involved in sustainable development research and training. It publishes the journal *Global Environmental Governance* and has established the Global Environment Information Center to reach out to the larger community on environmental issues.

For a description of one of the UNU programs, *see* Scenarios, Volume 5.

JEROME C GLENN *USA*

Utilitarianism

see Economics and Global Environmental Change (*Opening essay*, **Volume 5**); Social Science and Global Environmental Change (*Opening essay*, **Volume 5**)

Utopianism

see Political Systems and the Environment: Utopianism (**Volume 5**)

Value of an Ecosystem

see Ecosystem Services and Costing (**Volume 2**)

Virtual Environments

Gale Moore
University of Toronto, Toronto, Canada

The term virtual environment is used very broadly to describe a variety of computer-generated spaces. Virtual environments have been made possible by developments in information and communication technologies (ICTs) and their convergence with a broad range of electronic media, Internet and virtual reality (VR) technologies, both synchronous (same time) and asynchronous (different time). These technologies constitute what Kitchin (1998) has called cyberspatial technologies, referring to the term Cyberspace, rst used by William Gibson in his science ction novel, Neuromancer (1984) to describe electronic space. We experience these environments as real, when they are in fact, representations of reality that do not exist in the material world, except in the technical components that comprise them. This broad de nition draws attention to the variety and diversity of digital media that characterize these environments, and provides a continuum along which readers outside this eld of research can move if they wish to explore further.

We live in an era characterized by the intensification of the symbolic: information and knowledge are the resources of the 21st century and the source of economic growth and wealth creation. Environmentalists understand the power of symbols, and the meanings that come with new forms of representation. Consider the profound shift in human thinking brought about by globalization, a shift symbolized by the image of the earth as seen from space. How has this structured the way we think about the natural world? Environmentalists have also engaged in the exploration of a complex boundary – the interface between nature and culture. What insights and perspectives might they bring to the exploration of the boundary between the physical and virtual worlds brought on by our engagement in virtual environments?

Virtual environments have been made possible by developments over the past 40 years in ICTs, interactive computer graphics, and VR. These environments vary in their representational richness, i.e., in the nature of the media, and the specific forms of symbolic representation that comprise them, e.g., text, graphics, audio, video, etc. But it is the ways in which this class of technologies is capable of extending our senses and our intelligence that has the potential to change not only the way we work, live and learn, but also the way in which we think. As Postman (1992), reminds us:

> New technologies alter the structure of our interests: the things we think about. They alter the character of our symbols: the things we think with. And they alter the nature of community: the arena in which thoughts develop.

The experiences of users in these emerging representational worlds, and the significance of these environments in both shaping and being shaped by these experiences, is a topic of considerable research and debate – and one that demands an interdisciplinary response.

Today's virtual environments are the current end product of a long movement across the spectrum of technologies that have changed our relation to time and space – the train, for example, in the 19th century, and in the 20th century – the telephone. However, as we move toward the virtual reality (VR) end of the continuum, they are also more than this.

The familiar experience of the telephone may be used to illustrate the concept of virtuality. Today, the telephone is so embedded in the activities of everyday life that we forget that the voice we hear at the other end of the line – a voice instantly recognizable if that of a friend or family member – is not actually a voice, but a representation of that voice. With today's digital packet switching networks, the voice is not only digitized, or broken down into a string of ones and zeros, but the packets created travel by various routes to their final destinations. The receiver perceives a continuous conversation and a familiar voice. In reality, this is simply a representation of the voice, and if the technology breaks

down, the voice will neither be recognized nor understood. We do not stop to think of the telephone as a type of virtual environment, but the communication space or environment supported by the technology of telephony is, in essence, virtual. The immersion in a virtual three-dimensional (3D) world is only an extension along this continuum.

At the VR end of the human machine continuum, we are dealing with technologies that are more postmodern than modernist in their application, i.e., a world of symbols and representation in which multiple realities can be simulated. As our senses are extended, the boundaries between the real and virtual world are blurred. Increasingly, it will be difficult to discern what is real when increasingly large parts of our reality are, in fact, virtual or more accurately, a virtuality. Rheingold (1992) suggests these environments are not real in the sense that we understand the physical world, nor are they virtual, the opposite of real. The system's virtuality is a *structure of seeming* the conceptual feel of what is created. The physical body recedes; the mind and senses are at the forefront. Identity becomes labile, symbols are real and we immerse and navigate in an electronic world. It is this vision of a disembodied future, a blurring of the boundary between the human with the machine. There is an increasingly symbiotic relationship between the physical and virtual worlds. Virtual environments are like this; this is part of their power and our fascination with them.

TYPES OF VIRTUAL ENVIRONMENTS

The first type of virtual environment *information environments* will be the most familiar. These are primarily text-based, and a number have emerged around the power of computer technology to store large quantities of data. Information environments have helped solve two of the great problems of the human mind: limitations on the amount of information that can be stored, and the problem of retrieval. These environments augment human intelligence by freeing the human mind to carry out activities for which it is uniquely suited pattern recognition, evaluation and contextualization. We navigate through vast data spaces and digital libraries of information, manipulate textual and numeric symbols, and give them representational form that conveys the underlying pattern and relationships uncovered by the human mind. Information management, or more fashionably, knowledge management, are current expressions of the potential of information environments. Initially, these environments could support only one user at a time, but their power multiplied when machines could be networked. A fascinating example is the exponential growth of the World Wide Web, a global collection of hypertext documents that allows users with a browsing program such as Internet Explorer or Netscape Navigator, to access what, in a few years, has begun to approximate an electronic global Alexandrian Library.

Closely related are the *communication environments*; spaces in which people interact, not only with data and its representation, but also with other people what researchers call *human human computer-mediated communication.* Media such as email, listservs, bulletin boards, and chat, make it possible for people across the globe, who are connected to the Internet, to interact in new ways that challenge our understanding of time and space. Many of these are text-based environments, and individual media may be either synchronous, i.e., they require that users be present at the same time, or asynchronous, as is the case with email, where messages can be sent and received at different times.

Chat services, such as Instant Messenger™ recently entered the public domain. In a typical chat environment, the user has a series of icons representing people who are willing to chat. These people may be local or distributed around the world. At a glance they can observe who in the group are currently online (the icon changes color) and, hence, are accessible. Even if a chat is not initiated, this awareness of friends and colleagues affords an opportunity for interaction that is, at a minimum, comparable to awareness of the presence or absence of others in a physical office environment.

While the potential for interaction may appear limited in these environments, the increased availability of computers and access to the Internet since the mid 1990s has, in fact, resulted in an explosion of online communities groups of people sharing a common interest who meet regularly using a variety of mediated communication tools and environments. These communities and their members have been an important subject in recent years, drawing attention to the potential of even these limited environments to support rich and complex forms of social interaction.

Collaborative environments or *collaborative virtual environments* (*CVEs*) are applications that allow multiple users to meet and to communicate, and depending on the representation richness of the environment, to collaborate and to share a variety of documents and artifacts.

The first examples of this type of environment are MUDs (*multiple user domains*). There are a number of other environments such as MOOs (*multiple object-oriented*), or MUSEs (*multiple user social environments*), which, while differing in technical functionality, provide a similar opportunity for social interaction. Initially, MUDs were text-based, and used a variety of textual conventions such as *emoticons* (e.g., ;-) is a smile) to enrich their real-time conversations, and to compensate, in part, for the lack of gesture, eye contact, voice tone and inflection normally available in face-to-face communication. A rich visual vocabulary of emoticons now exists. MUDs have traditionally been developed within the context of specific groups of users, and the ways in which these spaces are socially

constructed is of particular interest. Research has found that these virtual *spaces* often became *places*, i.e., settings with their own culture and behavioral norms. While these spaces do not realistically represent a specific place in the material world, they do represent a type of place, for example, a classroom or a cafe. Behavioral norms and community standards emerge out of the interactions of the members of the group involved in the CVEs, and deviant behaviors are punished.

The power of these environments is further increased when a variety of media and forms of representation are supported. Today, there are a number of commercial CVEs available, which can be accessed using a standard Web browser. Examples include: Placeware • www.placeware.com• , Teamwave • www.teamwave.com• or eRoom • www.eroom.com• . The majority support standard communication activities, both synchronous and asynchronous, such as email, *document sharing*, and chat. The more sophisticated environments also provide tools to support interaction and negotiation – voting and document versioning, for example. A number offer both public and private spaces and storage areas for files and work in progress. Access to, and use of these environments no longer require the acquisition of a new set of skills. However, while the potential of these environments for community building, as well as supporting collaboration in the workplace, is real, the sociology of these spaces is still not well understood, and they are often underutilized. More recently, CVEs have appeared that incorporate a visual component, more specifically, a 2–3D visual representation of the space in which participants can interact with each other and with objects.

Before moving to these VR environments, mention should be made of another type of virtual environment: media space. *Media spaces* are audio–video computer-mediated communication environments. Unlike the CVEs discussed above, media spaces connect two or more sites in real time and like the telephone, or a videoconference, allow users to see the physical reality at the remote end and to hear what is taking place. In the Ontario Telepresence Project (Moore, 1997), a variety of types of spaces, for example, offices, meeting rooms, etc., were connected across two geographically separate locations. In the media space, it was possible to meet colleagues in their offices even though they were not in the same city. It was also possible to check the parking lot at the remote location to see if a colleague's car or bicycle was there, or to wander the corridors at the remote location, glancing around looking for someone. Multiple monitors were used at each end to keep the person space and the task space separate. While information on computer screens was being shared, individuals could both see and talk with each other as they might if co-located. Pauses and silences could be interpreted, as could interruptions at either end. Media spaces were designed in the context of real world practices and one of the challenges of this research was to support *presence* – i.e., to represent over distance the richness and experience of face-to-face interaction such as gesture, tone, and gaze awareness.

From the 1960s on, advances were being made in distinct areas such as interactive computer graphics and VR, but these advances would not converge until the 1990s. Rheingold (1992, p 46) traces VR's roots back to the 1930s and early experiments with immersion, such as Cinerama, but he argues that VR progressed slowly in the 1970s as it waited for enabling technologies to mature. While media space researchers had been concerned about one type of presence, researchers in robotics had a different view of telepresence, namely, how to support human intelligence when it is necessary to carry out activities remotely; e.g., to defuse a bomb or carry out a surgical procedure. Some of these experiments relied on head-mounted displays, or data gloves, and other research came from the entertainment and games environment. By the 1990s, however, these areas were converging to create a powerful new reality – a reality in which a number of human senses were extended or augmented. It is at this end of the continuum that the boundary between the material world and virtual world is softening, shifting, and blurring. In the future, the virtual environments will be as familiar or perhaps more familiar than the physical environments we experience today.

The most accessible of the newer CVEs are the *virtual reality environments*, which use modeling languages such as virtual reality modeling language (VRML), and allow for participation using a web browser and the Internet.

Virtual worlds are a popular form of these environments. These surreal multi-user 2–3D worlds generally do not represent actual places in the world, but are imaginary landscapes that provide spaces or opportunity for human interaction. These are real-time or synchronous environments in which people are represented by graphical icons known as *avatars*. There is a strange out-of-body sensation as you navigate through these worlds, sometimes flying, meeting and interacting with others. You may find yourself in conversation with familiar characters such as Alice in Wonderland or Wanda the Fish; two popular avatars that people select to represent themselves. These worlds are, in general, themed, and provide fascinating insights into the nature of identity in a postmodern culture. The coherence of a single physically embodied identity slips away – if one desires, and multiple persona are not unusual. The technology makes it possible, but the motivation of the participants and the nature of the communities that evolve continue to raise questions about the nature of human experience. A number of these environments can be visited. The Contact Consortium web site maintains a useful set of links • www.ccon.org• .

Other CVEs are grounded in real-world practices, but most continue to be based on spatial metaphors. While the environments may not be realistic in terms of a specific

location in the physical world, many, like MUDs, are realistic in terms of the type of place they represent. For example, you recognize that you are in an office or a store and individuals behave in ways that are coherent with the repertoire of socially appropriate behaviors understood in the physical world. Extensive social and behavioral research is increasingly part of the development of these environments, underlining the desire to understand how these shared visualizations and environments can be used to support new forms of collaboration among people.

John Robinson and his colleagues at the Sustainability Research Institute at the University of British Columbia • www.sdri.ubc.ca• have created a space of particular interest to environmentalists. They have created 2D models (accessible from a Web browser) to increase the understanding of sustainability issues. Through the development of thought-provoking technology-aided experiences, which solicit user input and provide a visual representation of outcomes based on this input, users deepen their understanding of highly complex issues.

The most technically complex of the virtual environments are in the realm of VR. As Mitchell (1995) articulates: virtual reality technologies partially or totally immerse users in an interactive visual, artificial, computer-generated environment; instead of the users being spectators of a static screen, they are participants in an environment that responds . The key differentiator of these environments is that the users are immersed and interact with the environment in realistic ways. No longer the stuff of science fiction, these environments are at the interface between the real and the virtual world. Surgeons, for example, can now rehearse a difficult operation using a 3D representation of an actual patient. Airplane manufacturers can design and engineer an entire aircraft in a 3D environment. There is no physical scale model; the first plane that is built, flies. The design team may be distributed across the globe and work together in real time. The range of activities that VR supports, and the nature of the collaborations that are possible, suggest it will be applicable to many of the activities and interactions of daily life. How will our societies change as material reality and social interactions are augmented and mediated in the ways VR makes possible?

The cyberspatial technologies that comprise virtual environments extend and augment human intelligence and are changing not only what we think about, but also what we think with. If we directly apply this theme of virtual environments to the prospects for global environmental change, one pervasive notion is that as the natural or social environment deteriorates, people will increasingly move into virtual environments. The dystopians give us images of a retreat from public space into new worlds supported by technology. The utopians see the potential for human good. However, like the technologies that have gone before and those that are ahead, virtual environments are neither good nor bad, nor are they neutral. They emerge out of a specific socio-historical context and out of a specific set of social relations. They are shaping and being shaped by those who design, build and use them. Cyberspace may be a consensual hallucination, but when the electricity goes out – the illusion cannot be sustained. We are forever embodied – it's the human condition, and in the end it is nature that holds sway.

REFERENCES

Gibson, W (1984) *Neuromancer*, Gollanz, London.

Kitchin, R (1998) *Cyberspace: The World in the Wires*, Wiley, Chichester.

Mitchell, W J (1995) *City of Bits: Space, Place, and the Infobahn*, The MIT Press, Cambridge, MA.

Moore, G (1997) Sharing faces, places and spaces: The Ontario Telepresence Project Field Studies, in *Video-Mediated Communication*, eds K Finn, A J Sellen, and S Wilbur, Lawrence Erlbaum Association, Mahwah, NJ, 301–321.

Postman, N (1992) *Technopoly: The Surrender of Culture to Technology*, Vintage Books, New York.

Rheingold, H (1992) *Virtual Reality*, Simon and Schuster, New York.

Vulnerability

see Precautionary Principle (**Volume 4**)

Waldsterben

Waldsterben is a term widely used in Europe to denote forest decline in the late 1970s and early 1980s when it became a public issue. In mountainous regions particularly in the vicinity of the Poland, Germany and Slovakia triangle, thousands of hectares of forest died in a few years. The causes of waldsterben are complex, and may include direct damage to needles and leaves from industrial sulfur emissions, long-range acidic deposition, drought and mobilization of heavy metals in acidifying soils. Waldsterben is very photogenic and therefore it caught the attention of Green Parties, and other environmental non-governmental organizations. It is interesting to note, however, that the overall trend in forest growth in Europe over recent decades has been increased productivity.

See also: Nitrogen Deposition on Forests, Volume 2.

R E MUNN *Canada*

WCC (World Council of Churches)

The WCC is a fellowship of about 350 denominations worldwide with a combined membership of approximately 400 million persons. The participating churches in the WCC cooperate on programs of world development and relief, theological reflection and education, and advocacy on issues of justice, peace and creation.

Environmental concerns have been a dimension of the WCC's activity for over 30 years. Since 1988, climate change has been the main focus for the environmental work. The WCC Climate Change Programme includes preparation of resources on theological and ethical dimensions of the problem, support for environmental activities of churches in the South, and coordination of advocacy at the international level and among industrialized countries whose societies are responsible for the majority of emissions leading to climate change.

For further information, contact The World Council of Churches (Justice, Peace and Creation Cluster), 150 route de Ferney, PO Box 2100, 1211 Geneva 2, Switzerland. Tel.: • 41-22-791-6111; Fax: • 41-22-791-0361, Email: mpt@wcc-coe.org, Web site: www.wcc-coe.org.

DAVID HALLMAN *Canada*

Weak Anthropocentrism

see Environmental Ethics (**Volume 5**)

Weak Sustainability

see Economics and Global Environmental Change (*Opening essay,* **Volume 5**)

Weed

see Weed (**Volume 2**)

World Council of Churches (WCC)

see WCC (World Council of Churches) (**Volume 5**)

Yin–Yang Principles

see Land Transformation in China: 3000 Years of Human Ecological History (**Volume 3**)

Alphabetical List of Articles

Acid Sulfate Soils 3:151
Acoustic Thermometry 1:161
ACSYS (Arctic Climate System Study) 1:161
Adaptation Strategies 4:80
Adaptive Environmental Management 4:139
Aerosols, Effects on the Climate 1:162
Aerosols, Polar Stratospheric Cloud 1:167
Aerosols, Stratosphere 1:169
Aerosols, Troposphere 1:172
Afforestation: Environmental Impacts 3:156
Agarwal, Anil 4:140
Agassiz, Louis 1:175
Agenda 21 4:140
Agricultural Intensification in Java 3:159
Agricultural Intensification in Western Europe 3:162
Agricultural Subsidies and Environmental Change 3:168
Agroforestry 3:172
Aircraft Emissions 3:178
Air Pressure 1:176
Air Quality, Global 1:177
Albedo 1:182
Allometric 2:135
AMIP (Atmospheric Model Intercomparison Project) 1:183
Andersen, Stephen O 4:143
Angiosperm/Gymnosperm 2:135
Animal Physiology and Global Environmental Change 2:136
Animal Production, Feedlots, and Manure Problems in the US (Agriculture, Intensive) 3:186
Antarctica 1:184
Anthropocene 1:189
Anthropogenic 3:190
Anthropogenic Impacts on Atmospheric Oxygen 2:140
Anthropogenic Metabolism and Environmental Legacies 3:54
Anthropology and Global Environmental Change 5:163
Anticyclone 1:191
APN (Asia–Pacific Network) 4:143
Aquaculture and Environment: Global View from the Tropics to High Latitudes 3:190
Aquaculture in Asia 3:196
Aquaculture: Salmon Farming 3:200
Aral Sea 4:534
Arctic Air Quality 1:191
Arctic Climate 1:193
Arctic Ocean 1:199
Arctic Oscillation 1:201
ARM (Atmospheric Radiation Measurement) Program 1:203
Arrhenius, Svante 1:204
Art and the Environment 5:167
Asteroids and Comets, Effects on Earth 1:204
Atlantic Ocean 1:211
Atmospheric Angular Momentum and Earth Rotation 1:211
Atmospheric Composition, Past 1:213
Atmospheric Composition, Present 1:216
Atmospheric Electricity 1:218
Atmospheric Motions 1:221
Atmospheric Structure 1:243
Attenborough, David 5:175
Autotroph/Heterotroph 2:143

Baha'i Faith and the Environment 5:176
Baltic Sea 4:517
Basel Convention 4:146
BAT (Best Available Technology) 5:183
Bateson, Gregory 5:183
Benedick, Richard Elliot 4:148
Benthic/Pelagic 2:145
Biocenosis 2:145
Biocomplexity 2:145
Biodiversity in Freshwaters 2:146
Biodiversity in Soils and Sediments: Potential Effects of Global Change 2:152
Biodiversity: The UNEP Definition 2:159
Biogeochemical Cycle 2:159
Bioindicators 2:160
Biological and Ecological Dimensions of Global Environmental Change 2:1
Biological Invasions 2:11
Biomass Burning 3:205
Biomass Burning in Rural Homes in Tropical Areas 3:214
Biomass Fuel Power Development 3:216
Biomass Use for Urban Fuels in Developing Countries 3:217
Biome 2:166
Biome Models 2:166
Biome–BGC Ecosystem Model 2:171
Bioremediation 3:218
Biosphere 2:174
Biosphere Enhancement Ratio (BER) 2:175
Biosphere Reserves 2:175
Bjerknes, Jacob 1:245

Black Sea 4:521
Blue Plan 4:148
Bolin, Bert 1:245
Boreal Ecosystem–Atmosphere Study (BOREAS) 2:177
Boreal Forest 2:179
Boreal Forest Carbon Flux and it's Role in the Implementation of the Kyoto Protocol Under a Warming Climate 4:149
Bourdeau, Philippe F J 4:152
Brent Spar 5:184
Brinkhorst, Laurens Jan 4:153
Broecker, Wallace S 1:246
Brower, David 5:185
Bruce, James 4:153
Brundtland, Gro Harlem 4:154
Buddhism and Ecology 5:185
Budyko, Mikhail Ivanovich 1:248
Buffering Capacity 2:184
Business-as-usual Scenarios 5:191

C_3 and C_4 Photosynthesis 2:186
Capacity Building 4:156
Capacity, Assimilative 3:220
Carbon and Energy: Terrestrial Stores and Fluxes 2:190
Carbon Cycle 2:198
Carbon Dioxide Concentration and Climate Over Geological Times 1:249
Carbon Dioxide, Recent Atmospheric Trends 1:254
Carbon Monoxide 1:261
Carnegie–Ames–Stanford Approach (CASA) 2:202
Carrying Capacity 4:156
Carson, Rachel Louise 5:192
Caspian Sea 4:524
Cattle Grazing: Impacts on Land Cover and Methane Emissions 3:221
CBA (Cost Benefit Analysis) 5:193
CENTURY Ecosystem Model 2:206
CEOS (Committee on Earth Observation Satellites) 1:261
Cereal Cultivation (Agriculture, Extensive) 3:228
Chandler Wobble 1:262
Changes in World Marine Fish Stocks 3:238
Chaos and Cycles 2:209
Chaos and Predictability 1:263
Charney, Jule Gregory 1:266
Chernobyl 3:241
Chipko Movement 5:193
Chlorofluorocarbons (CFCs) 1:267
Christianity and the Environment 5:194
Circulating Freshwater: Crucial Link between Climate, Land, Ecosystems, and Humanity 5:201
CITES (Convention on International Trade in Endangered Species of Wild Fauna and Flora) 4:157
Clean Development Mechanism 4:158
Cleaner Production 4:158
CLIC (Climate and Cryosphere) 1:269
CLIMAP (Climate: Long-range Investigation, Mapping, and Prediction) 1:269
Climate 1:270
Climate Agenda 1:270
Climate Analogues 1:271
Climate Change 1:271
Climate Change Assessment, United States 4:544
Climate Change, Abrupt 1:272
Climate Change, Detection and Attribution 1:278
Climate Change: an Emerging Issue in the Global Agenda 4:163
Climate Feedbacks 1:283
Climate Model Simulations of the Geological Past 1:296
Climate Sensitivity 1:301
Climatic Extremes 3:243
Climatology 1:308
Climax Vegetation 2:215
CLIVAR (CLImate VARiability and Predictability) 1:311
Cloud–Radiation Interactions 1:312
Clouds 1:316
Club of Rome 4:166
CO_2 Enrichment: Effects on Ecosystems 2:215
Coastal Zone Management 4:166
Common Agricultural Policy (CAP) 3:247
Common but Differentiated Responsibility Principle (Stockholm/Rio) 4:168
Commons, Tragedy of the 5:208
Community Richness 2:224
Conference of Parties 4:168
Conservation Biology 2:228
Consumption Patterns: Economic and Demographic Change 3:249
Contaminated Lands and Sediments: Chemical Time Bombs? 3:98
Continental Drift 1:321
Controlled Environment Facilities in Global Change Research 2:228
Convection 1:325
Convention on Biological Diversity 4:169
Convention on Long-range Transboundary Air Pollution (LRTAP) 4:173
Coral Bleaching (1997–1998) 2:231
Coral Reefs: an Ecosystem Subject to Multiple Environmental Threats 2:232
Coriolis Effect 1:326
Cousteau, Jacques 5:209
Cretaceous 1:329
Critical Load 3:252
Crop Models 2:241
Crutzen, Paul J 1:330
Cryosphere 1:330
CSD (Commission on Sustainable Development) 4:178

Cumulative Environmental Assessment **4:**178
Curi, Kriton **3:**253

Daisyworld **2:**247
Dansgaard Oescheger Cycles **1:**332
Darwin, Charles Robert **2:**247
Data Banks: Archiving Ecological Data and Information **2:**248
De Wit, C T **2:**259
Deep Ecology **5:**211
Deforestation and Habitat Fragmentation in the Amazon Basin **3:**255
Deforestation in Historic Times **3:**259
Deforestation, Tropical: Global Impacts **3:**265
Demand Management **4:**181
Dematerialization and Sustainable Development 4:96
Dematerialization of the Economy **4:**183
Demographic Change: Indonesian Transmigration **3:**272
Demographic Change: Peopling of the Paci c Islands **3:**273
Demographic Change: the Aging Population **3:**277
Demographic transition **5:**211
Depletion of Stratospheric Ozone 1:140
Deposition, Dry **2:**260
Deposition, Wet **2:**261
Deserti cation **3:**282
Deserti cation Convention **4:**183
Deserti cation, De nition of **3:**282
Deserts **1:**332
Development and Global Environmental Change 5:150
Dimethylsul de (DMS) **1:**343
Discounting **5:**214
Distance Learning and Environment **4:**186
Disturbance **2:**261
DIVERSITAS **2:**268
Dobr s European Environment Assessment Process **4:**547
Dooge, James **4:**191
Dowdeswell, Elizabeth **4:**192
Downscaling **1:**346
DPSIR (Driving Forces Pressures State Impacts Responses) **4:**193
Dust **1:**347
DVI (Dust Veil Index) **3:**290

EA (Environmental Assessment) **4:**194
Earth Charter **5:**216
Earth Day **5:**216
Earth First! **5:**217
Earth Observing Systems 1:61
Earth System History 1:31
Earth System Processes 1:13
Earthquake Intensity and Magnitude Scales **3:**293
Earthquakes Triggered by Human Activities **3:**294
ECA (Economic Commission for Africa) **4:**194
ECLAC (Economic Commission for Latin America and the Caribbean) **4:**195
Eco Taxes **4:**195
Ecocentric, Biocentric, Gaiacentric **5:**217
Eco-ef ciency 4:106
Eco-engineering to Promote Ecological Sustainability in China in an Era of Global Environmental Change **4:**548
Ecofeminism **5:**218
Ecological Capacity **3:**296
Ecological Dumping **4:**198
Ecological Economics 5:37
Ecological Footprint **3:**302
Economics and Global Environmental Change 5:25
Eco-socialism **5:**224
Ecosystem Approach **2:**225
Ecosystem Health **4:**199
Ecosystem Integrity **4:**199
Ecosystem Services **2:**226
Ecosystem Services and Costing **2:**272
Ecosystem Stability **2:**281
Ecosystem Structure and Function **2:**282
Ecotones **2:**283
Ecotoxicology **3:**302
EEA (European Environment Agency) **4:**200
Eemian **1:**352
Ehrlich, Paul **2:**288
EIS (Environmental Impact Statement) **4:**201
El Niño **1:**353
El Niño and La Niña: Causes and Global Consequences **1:**353
El Niño/Southern Oscillation (ENSO) **1:**370
El Viejo **1:**370
Elton, C S **2:**289
Emerging Environmental Issues 4:72
Emergy **3:**303
Emergy **5:**228
Emissions Trading **4:**201
Encyclopedias: Compendia of Global Knowledge **5:**228
Endangered Species **4:**201
Energy Balance and Climate **1:**371
Energy Balance Climate Models **1:**376
Enlightenment Project **5:**229
Environment **2:**290
Environmental (Eco) Labeling **4:**210
Environmental Assessment of Major Development Projects **4:**203
Environmental Cadmium in the Food Chain **3:**303
Environmental Change and Human Health: Extending the Sustainability Agenda 3:130
Environmental Changes Driven by Civil Con ict and War 3:146

Environmental Data Report 4:204
Environmental Defense Fund 5:230
Environmental Economics 5:230
Environmental Education 4:206
Environmental Ethics 5:231
Environmental Movement – the Rise of Non-government Organizations (NGOs) 5:243
Environmental Philosophy: Phenomenological Ecology 5:253
Environmental Policies for the World Oceans 4:212
Environmental Politics 5:49
Environmental Psychology/Perception 5:257
Environmental Refugees 4:214
Environmental Responses: an Overview 4:1
Environmental Security 5:269
Environmental Sociology 5:278
EOS (Earth Observing System) 1:382
Equilibrium Response 1:382
Equity 5:279
ESCAP (Economic and Social Commission for Asia and the Pacific) 4:218
EU (European Union) 4:218
European Network for Research in Global Change (ENRICH) 2:290
Eutrophication 2:292
Evolutionary Processes in Ecosystems 2:292
Extinctions (Contemporary and Future) 2:301
Extinctions and Biodiversity in the Fossil Record 2:297
Exxon Valdez Oil Spill 5:283

FAO (Food and Agriculture Organization) 4:219
Feedbacks, Chemistry – Climate Interactions 1:384
Fingerprinting 1:388
Fisheries: Effects of Climate Change on the Life Cycles of Salmon 3:309
Fisheries: Minamata Disease 3:312
Fisheries: Pacific Coast Salmon 3:314
Fisheries: Pollution and Habitat Degradation in Tropical Asian Rivers 3:316
Fixed Nitrogen 2:308
Food Consumption Patterns and their Influence on Greenhouse Gas Emissions 3:323
Food Webs 2:308
Forest Gap Models 2:316
Forest Logging Systems in Tropical Countries: Differential Impacts 3:328
Forest Stand 2:323
Forest: the FAO Definition 2:324
Fragmentation and Corridors 2:324
Francis of Assisi 5:284
Franklin, Benjamin 1:390
Freshwater Fisheries 2:327
Friends of the Earth 5:285
Fronts 1:391
Fugacity 3:335
Functional Biodiversity 2:20
Futures Research 5:285

Gaia 2:332
Gaia Hypothesis 5:287
GAIM (Global Analysis, Interpretation and Modeling) Program 1:392
Gandhi, Mohandas Karamchand 5:290
GARP (Global Atmospheric Research Program) 1:392
GATT (General Agreement on Tariffs and Trade) 4:221
GAW (Global Atmosphere Watch) 1:393
GCDIS (Global Change Data and Information System) 1:393
GCOS (Global Climate Observing System) 1:394
General Circulation Models (GCMs) 1:394
Generic Decomposition and Yield (G'DAY): A Model of Global Change Impacts on Plant and Soil 2:332
GEO (Global Environment Outlook) 4:221
Geoengineering: a Way to Stabilize Global Climate? 3:336
Geographic(al) Information Systems (GIS) 4:222
Geography 3:337
Geological Cycling 1:397
Geomorphological Change for Urbanization and Industry 3:338
Geomorphological Change: Landscape Modification for Recreation 3:345
Geomorphology 1:402
Geothermal Heat 1:402
GESAMP (Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection) 4:226
GEWEX (Global Energy and Water Cycle Experiment) 1:402
Glaciers 1:404
Global Environmental Change and Environmental History 5:62
Global Environmental Change and Natural Disasters 4:227
Global Forest Watch 4:231
Global land cover and land use trends and changes 3:13
Global Ocean Observing System (GOOS) 2:336
Global Plate Tectonics 1:410
Global Population Trends 3:16
Global Terrestrial Observing System (GTOS) 2:337
Global Warming Potential (GWP) 1:411
Globalization in Historical Perspective 5:73
GOALS (Global Ocean–Atmosphere–Land System) 1:411
Goldemberg, Jose 4:233
Goodman, Gordon 4:234
Governance and International Management 5:292
Great Lakes Region of North America 4:536

Green Investment 4:235
Green Revolution 3:347
Greenhouse Effect 1:413
Greenhouse Food Production (Agriculture Intensive) 3:352
Greening of Cities 3:356
Greenland 1:413
Greenland Ice Sheet 1:419
GRID (Global Resource Information Database, of the United Nations Environment Programme (UNEP) Division of Early Warning and Assessment) 4:237
Ground Temperature 1:422
Groundwater Withdrawal and the Development of the Great Man-made River Project, Libya 3:362

Habitat 2:339
Hadley Circulation 1:427
Halocarbons 1:427
Halophyte/Halophobe 2:339
Hare, Kenneth 1:428
Harmful Algal Blooms (HABs) 3:371
Hazards in Global Environmental Change 5:297
Heinrich (H-) Events 1:429
Hinduism and the Environment 5:303
Holdgate, Martin 4:239
Holdridge Life Zone Classification 2:339
Holling, C S 2:340
Holocene 1:431
Holocene: Climate Changes and Society 3:372
Homocentric 5:311
Houghton, John Theodore 1:432
Human Body, Immediate Environment 5:312
Human Disturbance of the Earth System: Dynamics and Complexities 3:1
Humidity 1:432
Hurricanes, Typhoons and other Tropical Storms – Descriptive Overview 1:433
Hurricanes, Typhoons and other Tropical Storms – Dynamics and Intensity 1:439
Hutchinson, G Evelyn 2:341
Hydrofluorocarbons 1:447
Hydrogen Peroxide Trends in Greenland Glaciers 1:447
Hydrologic Cycle 1:450
Hydrology 1:464
Hydrology 2:343

IAHS (International Association of Hydrological Sciences) 1:466
IAI (Inter-American Institute for Global Change Research) 4:241
IAMAS (International Association of Meteorology and Atmospheric Sciences) 1:466
IAPSO (International Association for the Physical Sciences of the Oceans) 1:467
Icebergs 1:467
ICREA (International Commodity-related Environmental Agreements) 4:242
ICSU (International Council for Science) 4:242
IDNDR (The International Decade for Natural Disaster Reduction) 4:242
IDWSSD (International Drinking Water Supply and Sanitation Decade) 4:244
IFIAS (International Federation of Institutes for Advanced Study) 4:244
IGAC (International Global Atmospheric Chemistry) 1:467
IGBP (International Geosphere–Biosphere Programme) 2:350
IGBP Core Projects 2:37
IGBP Terrestrial Transects 2:351
IGBP-DIS (International Geosphere–Biosphere Program Data and Information System) 2:351
IGY (International Geophysical Year) 1:468
IHDP (International Human Dimensions Programme on Global Change) 4:245
IHP (International Hydrological Program) 1:468
IIASA (International Institute for Applied Systems Analysis) 4:246
Imbrie, John 1:469
IMO (International Maritime Organization) 4:246
Impacts of Global Environmental Change on Animals 2:56
Indicators on Environmental Quality and Sustainable Development 4:247
Indigenous Knowledge, Peoples and Sustainable Practice 5:314
Indus Basin: a Case Study in Water Management 3:377
Industrial and Anthroposystem Metabolism 3:73
Industrial Ecology 4:248
Industrial Responses to Stratospheric Ozone Depletion and Lessons for Global Climate Change 4:65
Infectious Diseases 2:357
Infrared Radiation 1:470
Insect Pests and Global Environmental Change 3:381
Integrated Assessment 4:250
Integrated Assessment Models of Global Change 4:253
Integrated Assessment, Definition of 4:249
Integrated Pest Management (IPM) in an Era of Global Environmental Change 4:261
Inter American Institute for Global Change (IAIGC) 2:363
Inter-basin Transfer (IBT) for Water Supplies 3:387
Intergovernmental Panel on Climate Change (IPCC): an Historical Review 4:265
Internalization 4:278

International Biosphere Program (IBP) – a Foundation for Global Studies 2:364
International Environmental Law 5:324
International Environmental Prizes 4:279
International Environmental Standards 4:281
International Organizations in the Earth Sciences 1:156
International Rice Research Institute (IRRI) 3:391
Intertidal Zones 2:365
Intertropical Convergence Zone (ITCZ) 1:476
IOC (Intergovernmental Oceanographic Commission) 4:287
Ionosphere 1:476
IPCC (Intergovernmental Panel on Climate Change) 1:477
IPCS (International Program on Chemical Safety) 4:288
IRI (International Research Institute for Climate Prediction) 1:477
Iron Cycle 2:369
IRPTC (International Register of Potentially Toxic Chemicals) 4:289
Irrigation: Environmental Impacts 3:392
Irrigation: Induced Demise of Wetlands 3:399
ISCCP (International Satellite Cloud Climatology Project) 1:478
ISEW (Index of Sustainable Economic Welfare) and GPI (Genuine Progress Indicator) 5:331
Islam and the Environment 5:332
ISLSCP (International Satellite Land Surface Climatology Project) 1:478
ISO (International Organization for Standardization) 4:290
Isostasy 1:479
ISSC (International Social Science Council) 5:339
IUCN (The World Conservation Union) 4:290
IUGG (International Union of Geodesy and Geophysics) 1:479
Izrael, Yuri A 4:291

Jains and the Environment 5:341
Jet Stream 1:481
JGOFS (Joint Global Ocean Flux Study) 1:483
Joint Implementation 4:292
Judaism and the Environment 5:349
Junge Layer 1:483

Karst 3:411
Kassas, Mohamed 4:292
Keeling, Charles David 1:484
Kelly, Petra 5:355
Keystone Species 2:375
Kondratyev, Kirill Yakovlevich 1:485
Kondratyev, Nikolai Dmitrievich 5:356
Kovda, V A 3:411

La Nina 1:487
Lahars 3:413
Lake Baikal 3:413
Lake Victoria 4:539
Lakes and Rivers 2:375
Lamb, Hubert H 1:487
Land Cover and Climate 1:488
Land Degradation in the Mediterranean 3:417
Land Reclamation from Seas 3:424
Land Subsidence 3:430
Land Surface 1:493
Land Transformation in China: 3000 Years of Human Ecological History 3:430
Landscape Ecology 2:383
Landscape, Urban Landscape, and the Human Shaping of the Environment 5:357
Lang, Winfred 4:294
Lapse Rate 1:499
Large Scale Biosphere–Atmosphere Experiment in Amazonia (LBA) 2:383
Last Glacial Maximum 1:500
Latent Heat 1:500
LEAD (Leadership for Environment and Development) 4:294
Leapfrogging Technology 4:295
Leopold, Aldo 5:367
Life Cycles 2:385
Life Style, Private Choice, and Environmental Governance 5:368
Lifetime (of a Gas) 1:501
Lightning 1:501
Lightning and Atmospheric Electricity 1:502
Limnology 1:503
Literature and the Environment 5:370
Lithosphere 1:503
Little Ice Age 1:504
LOICZ (Land–Ocean Interactions in the Coastal Zone) 2:389
London Dumping Convention 4:297
Long-term Ecological Research (LTER) System 2:390
Lorenz, Edward N 1:509
Love Canal 5:382
Lovelock, James 1:510
Low Concentration Methane Bearing Gases 3:435
LRTAP (Convention on Long Range Trans-boundary Air Pollution) 4:297

MacNeill, Jim 4:299
Madden–Julian Oscillation 1:511

Malnutrition, Infectious Diseases and Global Environmental Change 3:440
Malone, Thomas F 1:511
Malthus, Thomas Robert 5:384
Man and Nature as a Single but Complex System 5:384
Manabe, Syukuro 1:512
Mangrove Ecosystems 2:395
Margalef, Ramon 2:401
Marine Mammal Exploitation: Whales and Whaling 3:446
Marsh, George Perkins 3:450
Marshes, Anthropogenic Changes 3:451
Materials Flow Accounting 4:300
Materials Flows for Mining and Quarrying 3:454
Maunder Minimum 1:514
Maximum Sustainable Yield (MSY) 4:307
May, Robert M 2:401
Medieval Climatic Optimum 1:514
Mediterranean Sea 4:528
MEPC (Marine Environment Protection Committee) 4:308
Mercury in the Environment 2:402
Mesosphere 1:516
Mesozoic 1:516
Metadata 2:409
Metapopulations 2:411
Meteorology 1:517
Methane Clathrates 1:518
Methane: Industrial Sources 3:461
Methane 1:517
Methyl Bromide 1:520
Microbial Diversity 2:421
Migrations: The Environmental Challenge of Population Movements 3:465
Milankovitch, Milutin 1:522
MINK (Missouri, Iowa, Nebraska, Kansas) Study 4:550
Model Simulations of Present and Historical Climates 1:114
Modeling Human Dimensions of Global Environmental Change 5:394
Modeling Regional Climate Change 1:523
Models of the Earth System 1:99
Modernity vs. Post-modern Environmentalism 5:408
Molina, Mario J 1:533
Monitoring in Support of Policy: an Adaptive Ecosystem Approach 4:116
Monitoring Systems, Global Geophysical 1:534
Monsi, Masami 2:426
Monsoons 1:539
Montreal Protocol, Multilateral Fund for the Implementation of 4:309
Mooney, Harold A 2:426
Mountain Climates 1:540
MSU (Microwave Sounding Unit) 1:541
Muir, John 5:411
Multi-issue Assessments 4:316
Munk, Walter 1:541
Munn, Robert Edward (Ted) 1:542

Nasr, Seyyed Hossein 5:412
National and Local Responses to Desertification 4:319
National Environmental Law 5:413
National Responsibilities for Greenhouse Gas Emissions 3:472
National Self-interest: a Major Factor in International Environmental Policy Formulation 4:323
Natural Capital 2:428
Natural Climate Variability 1:544
Natural Hazards 3:479
Natural Hazards: Social and Policy Responses 4:328
Natural Records of Climate Change 1:550
Natural Systems: Impacts of Climate Change 2:67
Nature 5:419
New Ageism 5:420
Niche 2:429
Nile River 4:542
NIMBY (Not In My Backyard) 5:421
Nitrate Leaching 3:493
Nitrate Protection Zones in Europe 3:493
Nitrogen Cycle 2:429
Nitrogen Deposition on Forests 2:435
Nitrogen Fixers 2:442
Nitrogen Mineralization 2:446
Nitrogen: Agricultural Uses and their Impacts 3:499
Nitrous Oxide 1:554
No Regrets Principle 4:334
Non-equilibrium Ecology 2:446
Non-linear Systems 2:450
Noosphere 5:421
North Atlantic Oscillation 1:555
Nuclear Waste: Geological Issues 3:506
Nuclear Winter 3:515

Obasi, G O P 4:335
Ocean Circulation 1:557
Ocean Conveyor Belt 1:579
Ocean Drilling Program 1:581
Ocean Observing Techniques 1:581
Oceanography 1:584
Odum, Eugene P 2:456
OECD (Organization for Economic Co-operation and Development) 4:335
Oeschger, Hans A 1:585
OH–Radical: is the Cleansing Capacity of the Atmosphere Changing? 2:457
Oil and the Arctic Environment 3:517
Oil Fires: Kuwait 3:523

Oil Industry: Cultural Impacts and Environmental Risk Tolerance 3:525
Oil Shales and Tar Sands 3:528
Oligotrophic/Heterotrophic/Eutrophic 2:458
Orbital Variations 1:586
Organic Farming and the Environment 3:532
Ozone Depletion Potential (ODP) 1:590
Ozone Hole 1:590
Ozone Layer: Vienna Convention and the Montreal Protocol 4:337

Pacific–Decadal Oscillation 1:592
Pacific–North American (PNA) Teleconnection 1:594
PAGES (Past Global Changes) 1:596
Paleoclimatology 1:596
Paleozoic 1:597
PAR (Photosynthetically Active Radiation) 2:460
Parameterization 1:598
Passerine 2:460
Perfluorocarbons (PFCs) 1:598
Permafrost 1:598
PET (Potential Evapotranspiration) 2:461
Phase Shifts or Flip-flops in Complex Systems 5:422
Phenology 2:461
Phenotype/Genotype 2:465
Phosphorus Cycle 2:466
Phosphorus Water Quality Objectives 3:542
Phosphorus: Global Transfers 3:536
Photochemical Reactions 1:601
Photosynthesis 2:470
Phytophagy 2:471
Planetary Boundary Layer 1:603
Plant Competition in an Elevated CO_2 World 2:471
Plant Dispersal and Migration 2:81
Plant Functional Types 2:481
Plant Growth at Elevated CO_2 2:489
Plant Ontogeny 2:496
Plants – Carbon Balance and Growth 2:497
Plants – from Cells to Ecosystems: Impacts of Global Environmental Change 2:94
Plate Tectonics 1:605
Pleistocene 1:607
Policies for Sustainable and Responsible Fisheries 4:343
Policies for Sustainable Forests: Examples from Canada 4:351
Policies that Promote Sustainable Energy Futures 4:356
Policies to Achieve Sustainable Agriculture 4:358
Policies to Achieve Sustainable Tourism 4:362
Policy Exercises: a Tool for Solving Environmental Policy Dilemmas 4:366
Policy Responses to Public Health Issues Relating to Global Environmental Change 4:47
Political Movements/Ideologies and the Environment 5:429
Political Systems and the Environment: Utopianism 5:443
Polynyas 1:607
Population Sizes, Changes 2:504
Post-normal Science 5:451
PPP (Polluter Pay Principle) 4:368
Precautionary Principle 4:369
Precautionary Principle 5:455
Prediction in the Earth Sciences 1:607
Prehistoric People and Forest Soils 3:546
Productivity 2:515
Productivity of Terrestrial Ecosystems 2:516
Projection of Future Changes in Climate 1:126
Property Rights and Regimes 5:457
Public-driven Response to Global Environmental Change 4:21

Quasi–Biennial Oscillation (QBO) 1:611
Quasi–Decadal Oscillation 1:613
Quaternary 1:615
Quotas in International Environmental Negotiations 4:371

r–K Strategies 2:540
Radiative Forcing 1:616
Radionuclides, Cosmogenic 1:618
Radiosondes 1:619
Rain 1:622
RAINS (Regional Air Pollution Information and Simulation) 4:552
Raven, Peter 2:522
Red Tide 3:548
Redfield Ratio 2:522
Redox Potential 2:523
Refugees and Human Conflict 3:548
Regional Responses to Global Environmental Change 4:515
Relating to Nature: Spirituality, Philosophy, and Environmental Concern 5:97
Religion and Ecology in the Abrahamic Faiths 5:461
Religion and Environment Among North American First Nations 5:466
Remote Sensing 2:524
Remote Sensing, Terrestrial Systems 2:528
Residence Time (of an Atom, Molecule or Particle) 1:622
Resilience 2:530
Respiration 2:531
Restoration, Ecosystem 2:532
Revelle, Roger Randall Dougan 1:623
Rhizosphere 2:539

Rice and its Spread: Double Cropping; New Varieties – Environmental Consequences and Methane Gas, Sustainability 3:116
Richardson, Lewis Fry 1:624
Risk Management in an Era of Global Environmental Change 4:375
River Regulation 3:551
Roberts, Walter Orr 1:625
Rodin, Leonid E 2:540
Roosevelt, Theodore 5:475
Rossby Waves 1:626
Rossby, Carl-Gustaf 1:626
Rosswall, Per Thomas 2:541
Rowland, F Sherwood 1:627
Runoff 1:628

Salinity 1:629
Salinity and Agriculture 3:559
Salinity Patterns in the Ocean 1:629
Salinization 2:543
Scenarios 5:476
SCEP (Study of Critical Environmental Problems) 4:382
Scienti c Responses in an Era of Global Environmental Change 4:36
SCOPE (Scientific Committee on Problems of the Environment) 4:383
SCOR (Scientific Committee on Oceanic Research) 1:640
Sea Ice 1:640
Sea Level 1:645
Sea Surface Temperature 1:650
Sedimentary Records of Long-term Ecological Change 2:112
Seismic Risk (and Risk Assessment) 4:383
Sensible Heat 1:656
Seveso 5:482
Shifting Cultivation and Land Degradation 3:565
Shipping: Harbors and Ports 3:572
Sierra Club 5:483
Small is Beautiful 5:483
SMIC (Study of Man's Impact on Climate) 4:385
Snow 1:656
Social Assessment 4:386
Social Ecology 5:484
Social Impact Assessment 4:387
Social Impact Assessment (SIA) 5:484
Social Learning in the Management of Global Atmospheric Risks 5:485
Social Science and Global Environmental Change 5:109
Socially Responsible Investment 4:393
Soft Energy Path 3:583
Soft Energy Paths 5:487
Soil Amelioration 3:583
Soil Deterioration and Loss of Topsoil 3:587
Soil Erosion 4:395
Soil Mineralization 2:543
Soil Moisture 1:658
Solar Irradiance and Climate 1:659
Solar Variability, Long-term 1:666
Southern Ocean 1:668
Southern Oscillation 1:672
Sovereignty and Sovereign States 5:487
SPARC (Stratospheric Processes and their Role in Climate) 1:672
Speth, James Gustave 4:397
Stable Isotopes 2:544
START (Global Change System for Analysis Research and Training) 4:398
Stockholm and Beyond 4:398
Storm Surge 1:673
Strategic Environmental Assessment 4:403
Stratosphere 1:674
Stratosphere, Chemistry 1:675
Stratosphere, Ozone Trends 1:682
Stratosphere, Temperature and Circulation 1:697
Strong, Maurice 4:406
Sub-grid Processes 1:704
Succession 2:551
Succession, Definition of 2:550
Suess Effect 2:557
Sunspots 1:704
Suomi, Verner Edward 1:707
Sustainability: Human, Social, Economic and Environmental 5:489
Sustainable Cities Policies 4:406
Sustainable Development 4:411
Sustainable Development of Energy Resources in Small Island Developing States (SIDS) 4:412
Sustainable Development Policies in Small Island Developing States 4:416
Sustainable Development Policy 4:422
Sustainable Energy Policies 4:86
Sustainable Transportation 4:426
Sverdrup, Harald Ulrik 1:708
Swidden 3:594
Symbiosis 2:557

Tamm, Carl Olof 2:559
Taxon 2:560
Taxonomy/Systematics 2:560
Technological Society and its Relation to Global Environmental Change 5:86
Temperate Coniferous Forests 2:560
Temperate Deciduous Forests 2:565
Temperate Grasslands 2:569
Temperature 1:709

Temperature Increase: Effects on Terrestrial Ecosystems 2:575
Terrestrial and Freshwater Ecosystems: Impacts of Global Change 2:122
The Changing Human–Nature Relationships (HNR) in the Context of Global Environmental Change 5:11
The Earth System 1:1
The Emergence of Global Environment Change into Politics 5:124
The Environment and Violent Conflict 5:137
The Global Temperature Record 1:82
The Human Dimensions of Global Change 5:1
Theology 5:492
Theories of Health and Environment 5:492
Thermohaline Circulation 1:710
Thoreau, Henry David 5:502
Thornthwaite, Charles Warren 2:581
Three Mile Island 5:503
Tickell, Crispin 4:436
Tidal Power Development 3:596
Tides, Atmospheric 1:710
Tides, Oceanic 1:710
Toepfer, Klaus 4:436
TOGA (Tropical Ocean Global Atmosphere) 1:713
Tolba, Mostafa 4:437
Tornadoes 1:713
Torrey Canyon 5:503
Tourism and Ecosystems 3:597
Tourism as a Global Driving Force for Environmental Change 3:609
Tourism: Climate Change and Tourist Resorts 3:623
Toxic Wastes: Generation and Disposal: a Case Study of UK Practice and Legislation 3:628
Tradable Permits for Greenhouse Gases 4:438
Tradable Permits for Limiting Stratospheric Ozone Depletion 4:446
Tradable Permits for Sulfur Emission Reductions 4:450
Trade Winds 1:715
Transboundary Movements of Hazardous Wastes 3:631
Transboundary Water Resource Management: The Example of the Canada/US IJC (International Joint Commission) 4:553
Transferable Fisheries Quotas (TFQs) 4:455
Transient Response 1:715
Transport and Species Dispersal 3:632
Transport Infrastructure 3:643
Transport: Global Freight and Passenger Flows 3:633
Trends in Environmental Management in the Last 40 Years 4:15
Trends in Global Emissions: Carbon, Sulfur, and Nitrogen 3:35
TRMM (Tropical Rainfall Measuring Mission) 1:716
Tropical Forests 2:582
Tropical Savannas 2:586
Tropopause 1:717
Troposphere 1:717
Troposphere, Ozone Chemistry 1:718
Tropospheric Temperature 1:720
Tsunamis, Causes and Consequences 1:725
Tucker, Jim 2:592
Tundra 2:593
Tunguska Phenomenon 1:730

UARS (Upper Atmosphere Research Satellite) 1:732
Ultraviolet Radiation 1:732
ULYSSES (Urban Lifestyles, Sustainability, and Integrated Environmental Assessment) 4:460
UNCED (United Nations Conference on Environment and Development) 4:460
UNCHS (United Nations Centre for Human Settlements) 4:461
UNCLOS (United Nations Convention on the Law of the Sea) 4:462
UNCTAD (United Nations Conference on Trade and Development) 4:463
UNDP (United Nations Development Program) 4:464
UNEP (United Nations Environment Programme) 4:464
UNESCO (United Nations Educational Scientific and Cultural Organization) and the Environment 4:465
UNFPA (United Nations Population Fund) 4:466
UNIDO (United Nations Industrial Development Organization) 4:468
United Nations Conferences on the Environment 4:469
United Nations Convention to Combat Desertification (UNCCD) 4:474
United Nations Environment Programme (UNEP) Regional Seas Programme (RSP) 4:476
United Nations Environment Programme (UNEP) Saskawa Environment Prize 4:477
United Nations Framework Convention on Climate Change and Kyoto Protocol 4:478
UNSCEAR (United Nations Scientific Committee on the Effects of Atomic Radiation) 4:485
UNU (United Nations University) 5:504
Urban Climate and Respiratory Disease 3:649
Urban Ecosystems 3:655
Urban Heat Island 3:660
Urban Population Change 3:666
Urban Poverty and Environmental Health 3:672
Urban Sulfurous and Photochemical Smog 3:678
Urban Wastes 3:684
Use of Epidemiological Data to Develop Air Quality Standards for Short-term and Long-term Environmental Health Benefits 4:485

Valued Ecosystem Component 4:493
Vavrousek, Josef 4:493

Vegetation Ecosystem Model and Analysis Project (VEMAP) **2:**603
VEI (Volcanic Explosivity Index) **3:**689
Vernadsky, Vladimir **2:**604
Villach Conferences **4:**494
Viral Diseases and the In uence of Climate Change **3:**690
Virtual Environments **5:**505
Volatile Organic Compounds, Biogenic (VOCs) **2:**605
Volcanic Eruption, El Chichon **1:**736
Volcanic Eruption, Krakatau **1:**736
Volcanic Eruption, Mt. Pinatubo **1:**737
Volcanic Eruption, Tambora **1:**737
Volcanic Eruptions **1:**738
Volcanic Eruptions: Mt Merapi, Indonesia **3:**694
Volcanoes and Cities **3:**696
Volcanoes and the Environment **3:**699
Volvo Environment Prize **4:**495
von Humboldt, Alexander **2:**609
Vostok, Subglacial Lake **1:**744

Wadi **3:**707
Waldsterben **5:**508
Walker Circulation **1:**749
Ward, Barbara **4:**497
Waste Dumps in Megacities: Case Study of Istanbul **3:**707
Waste: the Global Mass and its Management **3:**709
Wastes: Land ll and Land Raising **3:**713
Water Resources: Baltic **3:**720
Water Resources: Great Lakes Case Study **3:**721
Water Resources: R o de la Plata **3:**723
Water Use: Future Trends, and Environmental and Social Impacts **3:84**
Water Vapor: Distribution and Trends **1:**750
Waterlogging **3:**726
Watson, Robert T **4:**498
Wave Power Development **3:**726
WBCSD (World Business Council for Sustainable Development) **4:**498
WCC (World Climate Conferences 1979 and 1990) **1:**752
WCC (World Council of Churches) **5:**509
WCED (World Commission on Environment and Development) **4:**499
WCP (World Climate Programme) **1:**752
WCRP (World Climate Research Programme) **1:**753
Weather **1:**754
Weather Extremes and Climate Impacts: a Case Study for the United States **3:**728
Weathering **1:**755
Weed **2:**611
Wegener, Alfred **1:**755
WFC (World Food Council) **4:**500
WGNE (Working Group on Numerical Experimentation) **1:**756
White, Gilbert **4:**500
Whittaker, Robert H **2:**611
WHO (World Health Organization) **4:**501
Wilson, Edward Osborne **2:**612
Wind Chill **1:**756
Wind Power Development **3:**733
WMO (World Meteorological Organization) **1:**757
WOCE (World Ocean Circulation Experiment) **1:**759
Woodwell, George Masters **2:**613
World Commission on Environment and Development (WCED) **4:**503
World Conservation Union (IUCN) Red List of Threatened Species **4:**508
World Energy Consumption Trends **3:**734
World Energy Council **3:**738
Worldwatch Institute **4:**509
WRI (World Resources Institute) **4:**510
WTO (World Trade Organization) **4:**510
WWF (World Wide Fund for Nature) **4:**511

Yellow River, China: a Case Study in Water Resources **3:**740
Younger Dryas **1:**761

Contributors

Hussein Abaza
UNEP, Geneva, Switzerland
GATT (General Agreement on Tariffs and Trade) • UNCTAD (United Nations Conference on Trade and Development)

Waleed Abdalati
NASA/Goddard Space Flight Center, Greenbelt, MD, USA
Greenland Ice Sheet

Gina A Adams
Colorado State University, Fort Collins, CO, USA
Biodiversity in Soils and Sediments: Potential Effects of Global Change

Nazeer Ahmad
The University of the West Indies, St. Augustine, Trinidad and Tobago
Acid Sulfate Soils

Bryant Allen
The Australian National University, Canberra, Australia
Shifting Cultivation and Land Degradation • Swidden

Myles R Allen
Rutherford Appleton Laboratory, Didcot, UK
Climate Change, Detection and Attribution

Keith Alverson
PAGES (Past Global Changes) International Project Office, Bern, Switzerland
Climate Change, Abrupt

Stephen O Andersen
US Environmental Protection Agency, Washington, DC, USA
Industrial Responses to Stratospheric Ozone Depletion and Lessons for Global Climate Change

Meinrat O Andreae
Max Planck Institute for Chemistry, Mainz, Germany
Dimethylsulfide (DMS)

John T Andrews
University of Colorado, Boulder, CO, USA
Heinrich (H-) Events

Natalia G Andronova
University of Illinois at Urbana-Champaign, Urbana, IL, USA
Budyko, Mikhail Ivanovich • Climate Sensitivity

Karen L Aplin
University of Reading, Reading, UK
Atmospheric Electricity

Christof Appenzeller
Swiss Federal Office of Meteorology and Climatology, Zurich, Switzerland
North Atlantic Oscillation

John W Ashe
Permanent Mission of Antigua and Barbuda to the United Nations, New York, NY, USA
ECLAC (Economic Commission for Latin America and the Caribbean) • Sustainable Development Policies in Small Island Developing States

Cynthia S Atherton
Lawrence Livermore National Laboratory, Livermore, CA, USA
Carbon Monoxide • Troposphere, Ozone Chemistry

Robert U Ayres
Center for the Management of Environmental Resources, Fontainebleau, France
Dematerialization of the Economy • Eco-efficiency • Industrial Ecology • Internalization

C Scott Baker
Auckland University, Auckland, New Zealand
Marine Mammal Exploitation: Whales and Whaling

James A Baker
Ministry of Natural Resources, Sault Ste. Marie, Canada
Adaptive Environmental Management

Marcia B Baker
University of Washington, Seattle, WA, USA
Clouds

Dennis Baldocchi
University of California, Berkeley, CA, USA
Carbon and Energy: Terrestrial Stores and Fluxes

Georges Balmino
Centre National d'Etudes, Toulouse, France
IUGG (International Union of Geodesy and Geophysics)

Tapas Kumar Bandyopadhyay
Ministry of Environment and Forests, New Delhi, India
Cattle Grazing: Impacts on Land Cover and Methane Emissions

Kaj Barlund
UN Economic Commission for Europe, Geneva, Switzerland
LRTAP (Convention on Long Range Trans-boundary Air Pollution)

Guy B Barnett
CSIRO Sustainable Ecosystems, Canberra, Australia
Natural Systems: Impacts of Climate Change

Leonard A Barrie
Paci c Northwest National Laboratory, Richland, WA, USA
Arctic Air Quality

Roger G Barry
University of Colorado, Boulder, CO, USA
Cryosphere • Mountain Climates

Brad Bass
University of Toronto, Toronto, Canada
Downscaling • Greening of Cities

Michel Batisse
Centre D'activitees Regionales du Plan Bleu, Sophia Antipolis, France
Blue Plan • Mediterranean Sea • UNESCO (United Nations Educational Scientific and Cultural Organization) and the Environment

F A Bazzaz
Harvard University, Cambridge, MA, USA
Plant Competition in an Elevated CO_2 World • Plants – from Cells to Ecosystems: Impacts of Global Environmental Change

Lee Beck
United States Environmental Protection Agency, Research Triangle Park, NC, USA
Methane: Industrial Sources

B Beckage
Duke University, Durham, NC, USA
Plant Dispersal and Migration

Juerg Beer
Swiss Federal Institute of Environmental Science and Technology (EAWAG), Duebendorf, Switzerland
Solar Variability, Long-term

Burton Bennett
New York, NY, USA
Chernobyl • UNSCEAR (United Nations Scientific Committee on the Effects of Atomic Radiation)

Charles Bentley
University of Wisconsin-Madison, Madison, WI, USA
Antarctica • Icebergs • Polynyas

Andre Berger
Institut d'Astronomie et de Géophysique G. Lemaˆtre, Louvain-la-Neuve, Belgium
Climate Model Simulations of the Geological Past • Milankovitch, Milutin • Orbital Variations

Elizabeth K Berner
Yale University, New Haven, CT, USA
Carbon Dioxide Concentration and Climate Over Geological Times

Robert A Berner
Yale University, New Haven, CT, USA
Carbon Dioxide Concentration and Climate Over Geological Times

Malcolm C M Beveridge
University of Stirling, Stirling, UK
Aquaculture and Environment: Global View from the Tropics to High Latitudes • Aquaculture: Salmon Farming

R G S Bidwell
Queen's University, Kingston, Canada
Allometric • Angiosperm/Gymnosperm • Autotroph/Heterotroph • Benthic/Pelagic • Biogeochemical Cycle • Bioremediation • Conservation Biology • Fixed Nitrogen • Habitat • Halophyte/Halophobe • Niche • Nitrogen Mineralization • Oligotrophic/Heterotrophic/Eutrophic • PAR (Photosynthetically Active Radiation) • Passerine • Phenotype/Genotype • Photosynthesis • Phytophagy • Plant Ontogeny • Respiration • Rhizosphere • Soil Mineralization • Symbiosis • Taxon • Taxonomy/Systematics

Ambassador Lars Bjorkbom
Stockholm, Sweden
Convention on Long-range Transboundary Air Pollution (LRTAP)

Howard B Bluestein
University of Oklahoma, Norman, OK, USA
Tornadoes

Stephen Bocking
Trent University, Peterborough, Canada
Marsh, George Perkins

John H Bodley
Washington State University, Pullman College, WA, USA
Anthropology and Global Environmental Change • Globalization in Historical Perspective

Rumen D Bojkov
World Meteorological Organization, Geneva, Switzerland
Stratosphere, Ozone Trends

Ester Boserup (Deceased)
Independent Consultant
Technological Society and its Relation to Global Environmental Change

Philippe Bourdeau
Université Libre de Bruxelles, Brussels, Belgium
DPSIR (Driving Forces–Pressures–State–Impacts–Responses) • Ecotoxicology • EU (European Union) • IPCS (International Program on Chemical Safety) • IRPTC (International Register of Potentially Toxic Chemicals) • Scientific Responses in an Era of Global Environmental Change

Ian Bowler
University of Leicester, Leicester, UK
Agricultural Intensification in Western Europe • Common Agricultural Policy (CAP)

Michelle Boyle
University of Waterloo, Waterloo, Canada
Monitoring in Support of Policy: an Adaptive Ecosystem Approach

Guy Brasseur
Max-Planck Institute for Meteorology, Hamburg, Germany
Crutzen, Paul J • OH–Radical: is the Cleansing Capacity of the Atmosphere Changing?

Peter Brimblecombe
University of East Anglia, Norwich, UK
Aerosols, Troposphere • Atmospheric Composition, Present • Atmospheric Structure

Jeffrey R Brook
Meteorological Service of Canada, Toronto, Canada
Use of Epidemiological Data to Develop Air Quality Standards for Short-term and Long-term Environmental Health Benefits

Thomas M Brooks
Center for Applied Biodiversity Science, Conservation International, Washington, DC, USA
Extinctions (Contemporary and Future)

Barbara E Brown
University of Newcastle, Newcastle-upon-Tyne, UK
Coral Bleaching (1997–1998)

Kerry Brown
Harvard University, Cambridge, MA, USA
Jains and the Environment

Janet Browne
Wellcome Institute for the History of Medicine, London, UK
Darwin, Charles Robert

Jim Bruce
Ottawa, Canada
Global Environmental Change and Natural Disasters • Villach Conferences • IDNDR (The International Decade for Natural Disaster Reduction) • Intergovernmental Panel on Climate Change (IPCC): an Historical Review

Paul H Brunner
Vienna University of Technology, Vienna, Austria
Anthropogenic Metabolism and Environmental Legacies

Dirk Bryant
World Resource Institute, Washington, DC, USA
Global Forest Watch

Reid Bryson
University of Wisconsin, Madison, WI, USA
Jet Stream

Robert W Buddemeier
Kansas Geological Survey, Lawrence, KS, USA
Coral Reefs: an Ecosystem Subject to Multiple Environmental Threats

Keith Bull
Institute of Terrestrial Ecology, Monks Wood, UK
Critical Load

Joanna Burger
Rutgers University, Piscataway, NJ, USA
Tourism and Ecosystems

Kevin Burke
University of Houston, Houston, TX, USA
Lithosphere

Virginia Burkett
US Geological Survey, Lafayette, LA, USA
Intertidal Zones

Richard T Burnett
Health Canada, Ottawa, Canada
Use of Epidemiological Data to Develop Air Quality Standards for Short-term and Long-term Environmental Health Benefits

Ian Burton
Meteorological Service of Canada, Downsview, Canada
Adaptation Strategies • International Environmental Prizes • White, Gilbert

David M Bush
State University of West Georgia, GA, USA
Storm Surge

Elizabeth Bush
Meteorological Service of Canada, Downsview, Canada
Air Quality, Global

James H Butler
National Oceanic and Atmospheric Administration, Boulder, CO, USA
Methyl Bromide

John Caddy
Imperial College, London, UK and CINVESTAV, Merida, Mexico
Policies for Sustainable and Responsible Fisheries

J Baird Callicott
University of North Texas, Denton, TX, USA
Environmental Ethics

Josep G Canadell
CSIRO Sustainable Ecosystems, Canberra, Australia
Biological and Ecological Dimensions of Global Environmental Change • Mooney, Harold A

Mark Cane
Lamont-Dohety Earth Observatory, Palisades, NY, USA
Bjerknes, Jacob

Annika Carlsson-Kanyama
Environmental Strategies Research Group, Stockholm, Sweden
Food Consumption Patterns and their Influence on Greenhouse Gas Emissions

William J Carlyle
University of Winnipeg, Winnipeg, Canada
Cereal Cultivation (Agriculture, Extensive)

Henri Carsalade
FAO, Rome, Italy
FAO (Food and Agriculture Organization) • WFC (World Food Council)

David Cartwright
Peters eld, Hampshire, UK
Tides, Atmospheric • Tides, Oceanic

S Catovsky
Harvard University, Cambridge, MA, USA
Plant Competition in an Elevated CO_2 World • Plants from Cells to Ecosystems: Impacts of Global Environmental Change

Thure E Cerling
University of Utah, Salt Lake City, UT, USA
C_3 and C_4 Photosynthesis • Stable Isotopes

Marie-Lise Chanin
CNRS/SA, Verrieres le Buisson, France
SPARC (Stratospheric Processes and their Role in Climate)

Piers Chapman
Texas A&M University, College Station, TX, USA
WOCE (World Ocean Circulation Experiment)

Leslie Charles
NASA, Washington, DC, USA
CEOS (Committee on Earth Observation Satellites)

Marion Cheatle
UNEP, Nairobi, Kenya
GEO (Global Environment Outlook)

Anilla Cherian
American Foundation for the University of West Indies, New York, NY, USA
Sustainable Development of Energy Resources in Small Island Developing States (SIDS)

Walter Chomentowski
NASA/Goddard Space Flight Center, Greenbelt, MD, USA
Deforestation and Habitat Fragmentation in the Amazon Basin

Nazli Choucri
Massachusetts Institute Of Technology, Cambridge, MA, USA
Governance and International Management

John R Christy
University of Alabama in Huntsville, Huntsville, AL, USA
MSU (Microwave Sounding Unit) • Tropospheric Temperature

John A Church
Antarctic CRC and CSIRO Marine Research, Tasmania, Australia
Southern Ocean

Galina Churkina
Max-Planck Institute for Biogeochemistry, Jena, Germany
Biocenosis • Noosphere • Vernadsky, Vladimir

Ralph J Cicerone
University of California, Irvine, CA, USA
Rowland, F Sherwood

Phillip J Clapham
Northeast Fisheries Science Center, Woods Hole, MA, USA
Marine Mammal Exploitation: Whales and Whaling

Jennifer Clapp
Trent University, Peterborough, Canada
Basel Convention

J S Clark
Duke University, Durham, NC, USA
Plant Dispersal and Migration

Kenneth E Cockshull
Horticulture Research International, Warwick, UK
Greenhouse Food Production (Agriculture Intensive)

Rita R Colwell
National Science Foundation, Arlington, VA, USA
Biocomplexity

Arthur Conacher
University of Western Australia, Crawley, Australia
Salinity and Agriculture

Hadrian F Cook
Imperial College at Wye, Ashford, UK
Nitrate Protection Zones in Europe

Claudia Copeland
US Library of Congress, Washington, DC, USA
Animal Production, Feedlots, and Manure Problems in the US (Agriculture, Intensive)

Steven Cork
CSIRO Sustainable Ecosystems, Canberra, Australia
Ecosystem Services and Costing

Michael Coughlan
World Meteorological Organization, Geneva, Switzerland
Climate Agenda • WCP (World Climate Programme)

Curt Covey
Lawrence Livermore National Laboratory, Livermore, CA, USA
Asteroids and Comets, Effects on Earth • Climate Analogues • Model Simulations of Present and Historical Climates

Peter Crabb
Australian National University, Canberra, Australia
Irrigation: Environmental Impacts • Salinization • Waterlogging

Wolfgang Cramer
Potsdam Institute for Climate Impact Research, Potsdam, Germany
Biome Models

Paul Crutzen
Max-Planck Institute for Chemistry, Mainz, Germany
Anthropocene • Nuclear Winter

Charles Curtis
University of Manchester Environment Center, Manchester, UK
Nuclear Waste: Geological Issues

Walter F Dabberdt
National Center for Atmospheric Research, Boulder, CO, USA
Radiosondes

Kenneth A Dahlberg
Western Michigan University, Kalamazoo, MI, USA
Green Revolution

Gretchen C Daily
Stanford University, Stanford, CA, USA
Ehrlich, Paul

Virginia Dale
Oak Ridge National Laboratory, Oak Ridge, TN, USA
Disturbance • Succession

Trevor D Davies
University of East Anglia, Norwich, UK
Lamb, Hubert H

John Dearing
University of Liverpool, Liverpool, UK
Sedimentary Records of Long-term Ecological Change

R L Desjardins
Agriculture and Agri-Food Canada, Ottawa, Canada
Boreal Ecosystem–Atmosphere Study (BOREAS)

Hama Arba Diallo
Secretariat of the Convention to Combat Deserti cation, Bonn, Germany
United Nations Convention to Combat Desertification (UNCCD)

Ricardo Diez-Hochleitner
The Club of Rome, Madrid, Spain
Club of Rome

Susan L Donoghue
University of Hong Kong, Hong Kong
Volcanoes and the Environment

Bo R Doos
Vienna, Austria
Lorenz, Edward N

Jane Dougan
Nova Scotia Southeastern University Oceanographic Center, Dania Beach, FL, USA
Distance Learning and Environment

Ian Douglas
University of Manchester, Manchester, UK
Changes in World Marine Fish Stocks • Earthquake Intensity and Magnitude Scales • Earthquakes Triggered by Human Activities • Emergy • Environmental Changes Driven by Civil Conflict and War • Geomorphological Change for Urbanization and Industry • Global land cover and land use trends and changes • Human Disturbance of the Earth System: Dynamics and Complexities • Indus Basin: a Case Study in Water Management • Karst • Land Subsidence • Materials Flows for Mining and Quarrying • National Responsibilities for Greenhouse Gas Emissions • Red Tide • Urban Ecosystems • World Energy Consumption Trends • World Energy Council

Cheryl A Doyle
ENVIRON Corporation, Princeton, NJ, USA
Urban Wastes

Harold E Dregne
Texas Tech University, Lubbock, TX, USA
National and Local Responses to Desertification • Soil Erosion

Les Duckers
Coventry University, UK
Wave Power Development

David Dudgeon
The University of Hong Kong, Hong Kong
Fisheries: Pollution and Habitat Degradation in Tropical Asian Rivers

Peter N Duinker
Dalhousie University, Halifax, Canada
Policies for Sustainable Forests: Examples from Canada • Sustainable Development • Valued Ecosystem Component

Henri J Dumont
Ghent University, Ghent, Belgium
Black Sea • Caspian Sea

John Dunn
University of Cambridge, Cambridge, UK
The Emergence of Global Environment Change into Politics

John A Dutton
Pennsylvania State University, PA, USA
Suomi, Verner Edward

O P Dwivedi
University of Guelph, Guelph, Canada
Hinduism and the Environment

James R Ehleringer
University of Utah, Salt Lake City, UT, USA
C_3 and C_4 Photosynthesis • Stable Isotopes

Eckart Ehlers
University of Bonn, Bonn, Germany
von Humboldt, Alexander • ISSC (International Social Science Council)

Paul R Ehrlich
Stanford University, Stanford, CA, USA
Raven, Peter

Omar E El-Arini
Multilateral Fund Secretariat, Montreal, Canada
Montreal Protocol, Multilateral Fund for the Implementation of

Aleya El Bindari Hammad
The Robert S Wagner School of Public Health Services, New York, NY, USA
Policy Responses to Public Health Issues Relating to Global Environmental Change • WHO (World Health Organization)

Rasha El Diwany
The Robert S Wagner School of Public Health Services, New York, NY, USA
Policy Responses to Public Health Issues Relating to Global Environmental Change

Essam El-Hinnawi
National Research Center, Cairo, Egypt
WCED (World Commission on Environment and Development)

Osama A El-Kholy
Arabian Gulf University, Manama, Bahrain
Cleaner Production • Environmental (Eco) Labeling • Regional Responses to Global Environmental Change • Tolba, Mostafa • Trends in Environmental Management in the Last 40 Years

Mahmoud Kh El-Sayed
University of Alexandria, Alexandria, Egypt
Coastal Zone Management

David Elliott
The Open University, Milton Keynes, UK
Tidal Power Development

J Cynan Ellis-Evans
British Antarctic Survey, Cambridge, UK
Vostok, Subglacial Lake

Derek M Elsom
Oxford Brookes University, Oxford, UK
Urban Sulfurous and Photochemical Smog

Jeremy Eppel
Environment Directorate, Paris, France
OECD (Organization for Economic Co-operation and Development)

Paul R Epstein
Harvard Medical School, Boston, MA, USA
Infectious Diseases

Jenni L Evans
The Pennsylvania State University, University Park, PA, USA
Hurricanes, Typhoons and other Tropical Storms – Dynamics and Intensity

Robert Evans
Anglia Polytechnic University, Cambridge, UK
Soil Deterioration and Loss of Topsoil

Malin Falkenmark
Natural Science Research Council and Stockholm International Water Institute (SIWI), Stockholm, Sweden
Circulating Freshwater: Crucial Link between Climate, Land, Ecosystems, and Humanity

Benjamin S Felzer
NOAA/Of ce of Global Programs, Silver Spring, MD, USA
Agassiz, Louis • CLIMAP (Climate: Long-range Investigation, Mapping, and Prediction) • Cretaceous • Dansgaard–Oescheger Cycles • Eemian • Holocene • Last Glacial Maximum • Mesozoic • Paleozoic • Pleistocene • Quaternary

Adam Fenech
Meteorological Service of Canada, Downsview, Canada
Social Learning in the Management of Global Atmospheric Risks • Ecosystem Services • Natural Capital

Frederick Ferre
University of Georgia, Athens, GA, USA
Relating to Nature: Spirituality, Philosophy, and Environmental Concern

Robert Finkel
Lawrence Livermore National Laboratory, Livermore, CA, USA
Radionuclides, Cosmogenic

John Firor
National Center for Atmospheric Research, Boulder, CO, USA
Roberts, Walter Orr

Ian A Fleming
Norwegian Institute for Nature Research, Trondheim, Norway and Coastal Oregon Marine Experiment Station, Oregon State University, Newport, OR, USA
Fisheries: Effects of Climate Change on the Life Cycles of Salmon • Life Cycles

Richard A Fleming
Great Lakes Forest Research Centre, Sault Ste. Marie, Canada
Biome • Biosphere • Climax Vegetation • Ecosystem Stability • Environment • Forest Stand • Holling, C S • Integrated Pest Management (IPM) in an Era of Global Environmental Change • Keystone Species • Landscape Ecology • Life Cycles • Productivity • r–K Strategies • Succession, Definition of • Weed

Robert L Fleming
Great Lakes Forest Research Centre, Sault Ste. Marie, Canada
Life Cycles

Richard Foltz
University of Florida, Gainesville, FL, USA
Nasr, Seyyed Hossein

Lee Frelich
University of Minnesota, St. Paul, MN, USA
Temperate Deciduous Forests

Burkhard Frenzel
Universität Hohenheim, Stuttgart, Germany
Holocene: Climate Changes and Society

Elbert W (Joe) Friday, Jr
National Academy of Sciences, Washington, DC, USA
Public-driven Response to Global Environmental Change • Weather

Roland Fuchs
International START Secretariat, Washington, DC, USA
START (Global Change System for Analysis Research and Training)

Jed Fuhrman
University of Southern California, Los Angeles, CA, USA
Microbial Diversity

Inez Fung
University of California, Berkeley, CA, USA
Tucker, Jim

W S Fyfe
University of Western Ontario, London, Canada
Waste: the Global Mass and its Management

Boris K Gannibal
St Petersburg, Russian Federation
Rodin, Leonid E

John H C Gash
Centre for Ecology and Hydrology, Wallingford, UK
Deforestation, Tropical: Global Impacts

F Gassmann
Paul Scherrer Institute (PSI), Villigen, Switzerland
Public-driven Response to Global Environmental Change

W Lawrence Gates
University of California, Livermore, CA, USA
AMIP (Atmospheric Model Intercomparison Project) • Climate • Climate Change • WGNE (Working Group on Numerical Experimentation)

Jean-Pierre Gattuso
Laboratoire d'Océanographie de Villefranche, Villefranche-sur-mer, France
Coral Reefs: an Ecosystem Subject to Multiple Environmental Threats

Michael Ghil
University of California, Los Angeles, CA, USA
Natural Climate Variability

Richard Gilbert
Centre for Sustainable Transport, Toronto, Canada
Sustainable Transportation

Filippo Giorgi
Abdus Salam International Center for Theoretical Physics, Trieste, Italy
Modeling Regional Climate Change

Bernhard Glaeser
Social Science Research Center, Berlin, Germany
The Changing Human–Nature Relationships (HNR) in the Context of Global Environmental Change

Michael Glantz
National Center for Atmospheric Research, Boulder, CO, USA
Aral Sea

Jerome C Glenn
American Council for the United Nations University, Washington, DC, USA
Business-as-usual Scenarios • Scenarios • UNU (United Nations University)

Michael Goch eld
Rutgers University, NJ, USA
Environmental Cadmium in the Food Chain

Jose Goldemberg
University of São Paulo, São Paulo, Brazil
Leapfrogging Technology • Policies that Promote Sustainable Energy Futures

Frank B Golley
University of Georgia, Athens, GA, USA
International Biosphere Program (IBP) – a Foundation for Global Studies

Spring Gombe
The Robert S Wagner School of Public Health Services, New York, NY, USA
Policy Responses to Public Health Issues Relating to Global Environmental Change

Glynn Gomes
Gomes Consulting Enterprises, Oakville, Canada
Cousteau, Jacques

M Gomez-Erache
Facultad de Ciencias, Montevideo, Uruguay
Water Resources: R o de la Plata

Robert Goodland
World Bank, Washington, DC, USA
Environmental Assessment of Major Development Projects • Social Assessment • Strategic Environmental Assessment • Sustainability: Human, Social, Economic and Environmental

Eban Goodstein
Lewis and Clark College, Portland, OR, US
Economics and Global Environmental Change

James R Gosz
University of New Mexico, Albuquerque, NM, USA
Global Terrestrial Observing System (GTOS) • Long-term Ecological Research (LTER) System

Andrew Goudie
University of Oxford, Oxford, UK
Geomorphology

J Goudriaan
Wageningen University, Wageningen, The Netherlands
Crop Models

Anne Goujon
International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria
Demographic transition • Global Population Trends

Stith T Gower
University of Wisconsin, Madison, WI, USA
Productivity of Terrestrial Ecosystems

Thomas E Graedel
Yale School of Forestry and Environmental Studies, New Haven, CT, USA
Industrial and Anthroposystem Metabolism

David J Griggs
Meteorological Office, Bracknell, UK
Houghton, John Theodore

A P Lino Grima
University of Toronto, Toronto, Canada
Demand Management • Transboundary Water Resource Management: The Example of the Canada/US IJC (International Joint Commission)

Richard S Gross
Jet Propulsion Laboratory, Pasadena, CA, USA
Chandler Wobble

Michael Grubb
The Royal Institute of International Affairs, London, UK
Tradable Permits for Greenhouse Gases

Arnulf Grubler
International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria
Trends in Global Emissions: Carbon, Sulfur, and Nitrogen

Charles Guest
Australian National University, Canberra, Australia
Viral Diseases and the In uence of Climate Change

Lance Gunderson
Emory University, Atlanta, GA, USA
Resilience

Peter M Haas
University of Massachusetts, Amherst, MA, USA
United Nations Conferences on the Environment

Tahar Hadj-Sadok
UNFCCC, Bonn, Germany
Clean Development Mechanism

J Michael Hall
NOAA, Silver Spring, MD, USA
IRI (International Research Institute for Climate Prediction)

David Hallman
World Council of Churches, Toronto, Canada
• Christianity and the Environment • Theology • WCC (World Council of Churches)

Kevin Hamilton
University of Hawaii, Honolulu, HI, USA
Quasi Biennial Oscillation (QBO)

Allen L Hammond
World Resources Institute, Washington, DC, USA
Indicators on Environmental Quality and Sustainable Development

Trevor Hancock
Trevor Hancock Inc., Kleinburg, Canada
Theories of Health and Environment

Kevin Hanna
Wilfrid Laurier University, Waterloo, Canada
CBA (Cost Bene t Analysis) • Common but Differentiated Responsibility Principle (Stockholm/Rio) • Eco Taxes • Francis of Assisi • Malthus, Thomas Robert • Muir, John • PPP (Polluter Pay Principle) • Roosevelt, Theodore • UNCED (United Nations Conference on Environment and Development)

Roger Hansell
University of Toronto, Toronto, Canada
Ecosystem Services • Natural Capital

Joan Hardjono
Padjadjaran State University, Bandung, Indonesia
Agricultural Intensification in Java • Demographic Change: Indonesian Transmigration

F Kenneth Hare
University of Toronto, Toronto, Canada
Climatology

Richard Harrington
IACR, Rothamsted, UK
Insect Pests and Global Environmental Change

R Giles Harrison
University of Reading, Berkshire, UK
Atmospheric Electricity

L D Danny Harvey
University of Toronto, Toronto, Canada
Carbon Cycle • Climate Feedbacks • Energy Balance Climate Models • Greenhouse Effect • Parameterization • Radiative Forcing • Sub-grid Processes

Paul H Harvey
University of Oxford, Oxford, UK
Community Richness

A L Heathwaite
University of Shef eld, Shef eld, UK
Nitrate Leaching • Nitrogen: Agricultural Uses and their Impacts

Grant Heiken
Los Alamos National Laboratory, Los Alamos, NM, USA
Volcanoes and Cities

Ann Henderson-Sellers
Lucas Heights Science and Technology Centre, New South Wales, Australia
ISLSCP (International Satellite Land Surface Climatology Project) • Land Surface • Soil Moisture

Dudley Herschbach
Harvard University, Cambridge, MA, USA
Franklin, Benjamin

Kenneth Hewitt
Wilfrid Laurier University, Waterloo, Canada
Hazards in Global Environmental Change • Natural Hazards

Kaz Higuchi
Meteorological Service of Canada, Toronto, Canada
Boreal Forest Carbon Flux and it's Role in the Implementation of the Kyoto Protocol Under a Warming Climate • Carbon Dioxide, Recent Atmospheric Trends

David W Hilbert
Tropical Forest Research Center, Atherton, Australia
Non-linear Systems

R D Hill
The University of Hong Kong, Hong Kong
International Rice Research Institute (IRRI) • Rice and its Spread: Double Cropping; New Varieties – Environmental Consequences and Methane Gas, Sustainability

J HilleRisLambers
Duke University, Durham, NC, USA
Plant Dispersal and Migration

Tadaki Hirose
Tohoku University, Sendai, Japan
Monsi, Masami

Annette Hladik
Eco-Anthropologie, CNRS and Muséum National d'Histoire Naturelle, Brunoy, France
Agroforestry

Richard J Hobbs
Murdoch University, Murdoch, Australia
Fragmentation and Corridors

I J Hodgkiss
The University of Hong Kong, Hong Kong
Harmful Algal Blooms (HABs)

Peter Hogberg
Department of Forest Ecology, Umea, Sweden
Nitrogen Deposition on Forests • Tamm, Carl Olof

E H (Ted) Hogg
Canadian Forest Service, Edmonton, Canada
Boreal Forest

C S Holling
University of Florida, Gainesville, FL, USA
Resilience

Lon L Hood
University of Arizona, Tucson, AZ, USA
Quasi–Decadal Oscillation

Leen Hordijk
Wageningen University, Wageningen, The Netherlands
RAINS (Regional Air Pollution Information and Simulation)

Pierre Horwitz
Edith Cowan University, Joondalup, Australia
Limnology

David Houghton
University of Wisconsin-Madison, Madison, WI, USA
Latent Heat • Sensible Heat • Temperature

John Houghton
IPCC, Geneva, Switzerland
Intergovernmental Panel on Climate Change (IPCC): an Historical Review

R A Houghton
Woods Hole Research Center, Woods Hole, MA, USA
Woodwell, George Masters

Robert W Howarth
Environmental Defense, Boston, MA, USA
Nitrogen Cycle

S Mark Howden
CSIRO Sustainable Ecosystems, Canberra, Australia
Natural Systems: Impacts of Climate Change

Brian Hoyle
University of Southampton, Southampton, UK
Shipping: Harbors and Ports

Shaopeng Huang
University of Michigan, Ann Arbor, MI, USA
Ground Temperature

Lesley Hughes
Macquarie University, New South Wales, Australia
Bioindicators

Bruce A Hungate
Northern Arizona University, Flagstaff, AZ, USA
Terrestrial and Freshwater Ecosystems: Impacts of Global Change

Keith A Hunter
University of Otago, Dunedin, New Zealand
Iron Cycle

Tahir Husain
Memorial University of Newfoundland, St. John's, Canada
Oil Fires: Kuwait

Pat Hutchings
The Australian Museum, Sydney, Australia
Marshes, Anthropogenic Changes

Gary R Huxel
University of California, Davis, CA, USA
Food Webs

I Ibanez
Duke University, Durham, NC, USA
Plant Dispersal and Migration

C C Jaeger
Potsdam-Institute for Climate Impacts Research (PIK), Potsdam, Germany
Public-driven Response to Global Environmental Change

Anthony C Janetos
World Resources Institute, Washington, DC, USA
Remote Sensing, Terrestrial Systems

Marco A Janssen
Vrije Universiteit, Amsterdam, Netherlands
Modeling Human Dimensions of Global Environmental Change

Bengt-Owe Jansson
Stockholm University, Stockholm, Sweden
Baltic Sea

Arne J Jensen
Norwegian Institute for Nature Research, Trondheim, Norway
Fisheries: Effects of Climate Change on the Life Cycles of Salmon

Arne Jernelov
International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria
IIASA (International Institute for Applied Systems Analysis) • Mercury in the Environment

Thomas B Johansson
United Nations Development Programme, New York, NY, USA
Sustainable Energy Policies

Peter Jones
Biffa Waste Services Ltd., Buckinghamshire, UK
Wastes: Landfill and Land Raising

Philip D Jones
University of East Anglia, Norwich, UK
The Global Temperature Record

Pavel Kabat
Wageningen University and Research Centre, The Netherlands
Large Scale Biosphere–Atmosphere Experiment in Amazonia (LBA)

Veijo Kaitala
University of Jyväskylä, Jyväskylä, Finland
Population Sizes, Changes

Paul Kanciruk
Oak Ridge National Laboratory, Oak Ridge, TN, USA
Metadata

B Kasemir
Swiss Federal Institute for Environmental Science and Technology (EAWAG) Duebendorf, Switzerland
Public-driven Response to Global Environmental Change

Mohamed Kassas
Cairo University, Giza, Egypt
Environmental Education • Holdgate, Martin • Nile River

Reah Janise Kauffman
Worldwatch Institute, Washington, DC, USA
Worldwatch Institute

James J Kay
University of Waterloo, Ontario, Canada
Monitoring in Support of Policy: an Adaptive Ecosystem Approach • Phase Shifts or Flip-flops in Complex Systems

Stephanie Kaza
University of Vermont, Burlington, USA
Buddhism and Ecology

Mary Kearney
Toronto, Canada
DVI (Dust Veil Index) • Lahars • VEI (Volcanic Explosivity Index)

Stjepan Keckes
Borik, Rovinj, Croatia
GESAMP (Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection) • United Nations Environment Programme (UNEP) Regional Seas Programme (RSP)

Michael Keller
United States Forest Service, USA
Large Scale Biosphere–Atmosphere Experiment in Amazonia (LBA)

William W Kellogg
Boulder, CO, USA
SCEP (Study of Critical Environmental Problems) • SMIC (Study of Man's Impact on Climate)

Bruce E Kendall
University of California, Santa Barbara, CA, USA
Chaos and Cycles

Mansour Khalid
African Centre for Research & Environment (ACRE), Nairobi, Kenya
Brundtland, Gro Harlem

Fazlun M Khalid
Islamic Foundation for Ecology and Environmental Sciences, Birmingham, UK
Islam and the Environment

Joel G Kingsolver
University of North Carolina, NC, USA
Impacts of Global Environmental Change on Animals

Douglas Kinnison
National Center for Atmospheric Research, Boulder, CO, USA
Photochemical Reactions • Stratosphere, Chemistry

John Kirkby
University of Northumbria at Newcastle, Newcastle-upon-Tyne, UK
Refugees and Human Conflict

Robert J Klee
Yale School of Forestry and Environmental Studies, New Haven, CT, USA
Industrial and Anthroposystem Metabolism

Peter Klopfer
Duke University, Durham, NC, USA
Hutchinson, G Evelyn

Gunay Kocasoy
Bogazici University, Istanbul, Turkey
Curi, Kriton • Waste Dumps in Megacities: Case Study of Istanbul

Ch Korner
University of Basel, Basel, Switzerland
CO_2 Enrichment: Effects on Ecosystems

Akira Kudo
Kyoto University, Kyoto, Japan
Fisheries: Minamata Disease

Gunnar Kullenberg
International Ocean Institute, Gzira, Malta
Environmental Policies for the World Oceans • IOC (Intergovernmental Oceanographic Commission)

Howard Kunreuther
University of Pennsylvania, Philadelphia, PA, USA
Risk Management in an Era of Global Environmental Change

Yochanan Kushnir
Columbia University, Pailsades, NY, USA
Pacific–North American (PNA) Teleconnection • Sea Surface Temperature

Karin Labitzke
Meteorologisches Institut der FU-Berlin, Germany
Stratosphere, Temperature and Circulation

S LaDeau
Duke University, Durham, NC, USA
Plant Dispersal and Migration

Richard M Landau
Georgetown, Ontario, Canada
Baha'i Faith and the Environment

Ray Langenfelds
CSIRO Atmospheric Research, Aspendale, Australia
Anthropogenic Impacts on Atmospheric Oxygen

Sandra Lavorel
Centre d'Ecologie Fonctionnelle et Evolutive CNRS, Montpellier, France
Plant Functional Types

Nigel Lawson
University of Manchester, Manchester, UK
Materials Flows for Mining and Quarrying

Judith Lean
Naval Research Laboratory, Washington, DC, USA
Solar Irradiance and Climate • Sunspots

Martin J Lechowicz
McGill University, Montréal, Canada
Phenology

Brenda Lee
Montréal, Quebec, Canada
Landscape, Urban Landscape, and the Human Shaping of the Environment

Rik Leemans
National Institute of Public Health and the Environment, Bilthoven, The Netherlands
Biomass Fuel Power Development

Cecil E Leith
Lawrence Livermore National Laboratory, Livermore, CA, USA
Chaos and Predictability

A Dennis Lemly
Virginia Tech University, Blacksburg, VA, USA
Irrigation: Induced Demise of Wetlands

A Carl Leopold
Boyce Thompson Institute of Plant Research, Ithaca, NY, USA
Leopold, Aldo

Erkki Leppakoski
Abo Akademi University, Turku, Finland
Water Resources: Baltic

Manuel Lerdau
State University of New York, Stony Brook, NY, USA
Volatile Organic Compounds, Biogenic (VOCs)

Christian Léveque
CNRS-Programme Environment, Meudon, France
Biodiversity in Freshwaters

Joel S Levine
NASA Langley Research Center, Hampton, VA, USA
Biomass Burning

Simon A Levin
Princeton University, Princeton, NJ, USA
Whittaker, Robert H

S F Lienin
Paul Scherrer Institute (PSI), Villigen, Switzerland
Public-driven Response to Global Environmental Change

John Lingard
University of Newcastle-upon-Tyne, Newcastle-upon-Tyne, UK
Agricultural Subsidies and Environmental Change • Policies to Achieve Sustainable Agriculture

Karen T Lit n
University of Washington, Seattle, WA, USA
Sovereignty and Sovereign States

Changming Liu
Chinese Academy of Sciences, Beijing, China
Yellow River, China: a Case Study in Water Resources

Diana Liverman
University of Arizona, Tucson, AZ, USA
Inter American Institute for Global Change (IAIGC)

Steve Lonergan
University of Victoria, Victoria, Canada
Environmental Security

Paul Longley
University College London, London, UK
Geographic(al) Information Systems (GIS)

W M Lonsdale
CSIRO Entomology, Canberra, Australia
Biological Invasions

Michel Loreau
Ecole Normale Superieure, Paris, France
Evolutionary Processes in Ecosystems

Wolfgang Lutz
International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria
Demographic transition • Global Population Trends

Robyn Lyons
UNCHS, Nairobi, Kenya
UNCHS (United Nations Centre for Human Settlements)

Michael C MacCracken
Lawrence Livermore National Laboratory, Livermore, CA, USA
ARM (Atmospheric Radiation Measurement) Program • Broecker, Wallace S • Climate Change Assessment, United States • Equilibrium Response • Fingerprinting • General Circulation Models (GCMs) • IAMAS (International Association of Meteorology and Atmospheric Sciences) • IAPSO (International Association for the Physical Sciences of the Oceans) • Junge Layer • Madden Julian Oscillation • Methane • Models of the Earth System • Molina, Mario J • Nitrous Oxide • Prediction in the Earth Sciences • Sea Level • The Earth System • Transient Response • UARS (Upper Atmosphere Research Satellite)

Gordon MacDonald
International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria
IIASA (International Institute for Applied Systems Analysis)

Tom Mace
USEPA/OEI, Research Triangle Park, NC, USA
GCDIS (Global Change Data and Information System)

Anson W Mackay
University College London, London, UK
Lake Baikal

Donald Mackay
Trent University, Peterborough, Canada
Fugacity

Piers Maclaren
Piers Maclaren and Associates, Rangiora, New Zealand
Afforestation: Environmental Impacts

Jim MacNeill
Ottawa, Canada
World Commission on Environment and Development (WCED)

Joe Macquaker
University of Manchester, Manchester, UK
Oil Shales and Tar Sands • Weathering

Charles L Mader
Mader Consulting Co, Honolulu, HI, USA
Tsunamis, Causes and Consequences

John J Magnuson
University of Wisconsin-Madison, Madison, WI, USA
Lakes and Rivers

Jerry D Mahlman
Princeton University, Princeton, NJ, USA
Manabe, Syukuro • Projection of Future Changes in Climate

Paola Malanotte-Rizzoli
Massachusetts Institute of Technology, Cambridge, MA, USA
Munk, Walter

Elizabeth L Malone
Paci c Northwest National Laboratory, USA
Social Science and Global Environmental Change

Michael E Mann
University of Virginia, Charlottesville, VA, USA
Little Ice Age • Medieval Climatic Optimum

Nathan J Mantua
University of Washington, Seattle, WA, USA
Pacific–Decadal Oscillation

Jane C Marks
Northern Arizona University, Flagstaff, AZ, USA
Terrestrial and Freshwater Ecosystems: Impacts of Global Change

Gregg Marland
Oak Ridge National Laboratory, Oak Ridge, TN, USA
Geoengineering: a Way to Stabilize Global Climate?

Pablo A Marquet
P. Universidad Católica de Chile, Santiago, Chile
Metapopulations

Karen Martin
Pepperdine University, Malibu, CA, USA
Animal Physiology and Global Environmental Change

Douglas G Martinson
Columbia University, Palisades, NY, USA
Sea Ice

Julia Marton–Lefevre
LEAD International Inc., New York, NY, USA
Dooge, James • LEAD (Leadership for Environment and Development)

Pierre Matarasso
Centre International de Recherche sur l'Environnement et le Développement, CNRS-EHESS, Nogent sur Marne, France
Integrated Assessment Models of Global Change

Pierre Mathy
European Commission, Brussels, Belgium
European Network for Research in Global Change (ENRICH)

Elizabeth May
Sierra Club of Canada, Ottawa, Canada
Brent Spar • Brower, David • Carson, Rachel Louise • Chipko Movement • Earth Charter • Earth Day • Earth First! • Environmental Defense Fund • Environmental Movement – the Rise of Non-government Organizations (NGOs) • Exxon Valdez Oil Spill • Friends of the Earth • Kelly, Petra • Love Canal • Seveso • Sierra Club • Small is Beautiful • Soft Energy Paths • Three Mile Island • Torrey Canyon

James McCarthy
Harvard University, Cambridge, MA, USA
IGBP (International Geosphere–Biosphere Programme) • Rosswall, Per Thomas

John P McCarty
Environmental Protection Agency, Washington, DC, USA
Restoration, Ecosystem

Bonnie McCay
Rutgers The State University of New Jersey, New Brunswick, NJ, USA
Commons, Tragedy of the • Property Rights and Regimes

Raymond A McCord
Oak Ridge National Laboratory, Oak Ridge, TN, USA
Data Banks: Archiving Ecological Data and Information • Metadata

Mack McFarland
DuPont Fluoroproducts, Wilmington, DE, USA
Chlorofluorocarbons (CFCs) • Halocarbons • Hydrofluorocarbons • Perfluorocarbons (PFCs)

Kendal McGuf e
University of Technology, Sydney, Australia
Hurricanes, Typhoons and other Tropical Storms – Descriptive Overview

J McLachlan
Duke University, Durham, NC, USA
Plant Dispersal and Migration

A J McMichael
London School of Hygiene and Tropical Medicine, London, UK
Environmental Change and Human Health: Extending the Sustainability Agenda

Ross E McMurtrie
University of New South Wales, Sydney, Australia
Generic Decomposition and Yield (G'DAY): A Model of Global Change Impacts on Plant and Soil

Jeffrey A McNeely
The World Conservation Union, Gland, Switzerland
Biosphere Reserves • CITES (Convention on International Trade in Endangered Species of Wild Fauna and Flora) • Convention on Biological Diversity • Endangered Species • IUCN (The World Conservation Union) • World Conservation Union (IUCN) Red List of Threatened Species • WWF (World Wide Fund for Nature)

Michael J McPhaden
National Oceanic and Atmospheric Administration, Seattle, WA, USA
El Nino • El Nino and La Nina: Causes and Global Consequences • El Nino/Southern Oscillation (ENSO) • El Viejo • La Nina • Southern Oscillation

Ernesto Medina
Instituto Venezolano de Investigaciones Cient´ cas, Caracas, Venezuela
Tropical Savannas

Mark F Meier
University of Colorado at Boulder, Boulder, CO, USA
Glaciers

Thomas A Mensah
London, UK
IMO (International Maritime Organization) • London Dumping Convention • MEPC (Marine Environment Protection Committee) • UNCLOS (United Nations Convention on the Law of the Sea)

William K Michener
University of New Mexico, Albuquerque, NM, USA
Data Banks: Archiving Ecological Data and Information • Metadata

John Middleton
Brock University, St. Catharines, Canada
Sustainable Development Policy

Dennis S Mileti
University of Colorado, Boulder, CO, USA
Natural Hazards: Social and Policy Responses

Bruce Mitchell
University of Waterloo, Waterloo, Canada
IDWSSD (International Drinking Water Supply and Sanitation Decade) • Water Use: Future Trends, and Environmental and Social Impacts

J Mohan
Duke University, Durham, NC, USA
Plant Dispersal and Migration

Detlev Moller
Brandenburg Technical University, Cottbus, Germany
Hydrogen Peroxide Trends in Greenland Glaciers

Iris Moller
University of Cambridge, Cambridge, UK
Land Reclamation from Seas

Christopher N K Mooers
University of Miami, Miami, FL, USA
Monitoring Systems, Global Geophysical

H A Mooney
Stanford University, Stanford, CA, USA
Biological and Ecological Dimensions of Global Environmental Change

Gale Moore
University of Toronto, Toronto, Canada
Virtual Environments

Hedley Moorwood
Brown and Root NA Limited, Kingston, UK
Groundwater Withdrawal and the Development of the Great Man-made River Project, Libya

Pierre Morel
NASA Goddard Space Flight Center, Greenbelt, MD, USA
EOS (Earth Observing System)

Carol Morris
Cheltenham and Gloucester College of Higher Education, Cheltenham, UK
Organic Farming and the Environment

Linda D Mortsch
Waterloo, Canada
Great Lakes Region of North America

Richard Moss
IPCC, Geneva, Switzerland
Intergovernmental Panel on Climate Change (IPCC): an Historical Review

Walter Munk
University of California, La Jolla, CA, USA
Acoustic Thermometry • Sverdrup, Harald Ulrik

R E Munn
University of Toronto, Toronto, Canada
Anthropogenic • Biodiversity: The UNEP Definition • Biosphere Enhancement Ratio (BER) • Bourdeau, Philippe F J • Cumulative Environmental Assessment • EA (Environmental Assessment) • Ecological Dumping • EIS (Environmental Impact Statement) • Emerging Environmental Issues • Forest: the FAO Definition • Holdridge Life Zone Classification • ICREA (International Commodity-related Environmental Agreements) • ICSU (International Council for Science) • Izrael, Yuri A • Joint Implementation • Kondratyev, Kirill Yakovlevich • Multi-issue Assessments • NIMBY (Not In My Backyard) • No Regrets Principle • SCOPE (Scientific Committee on Problems of the Environment) • Wadi • Waldsterben

Mary Fran Myers
University of Colorado, Boulder, CO, USA
Natural Hazards: Social and Policy Responses

Norman Myers
Oxford University, Oxford, UK
Environmental Refugees

Shahid Naeem
University of Washington, Seattle, WA, USA
Functional Biodiversity • May, Robert M

G J Nagy
Facultad de Ciencias, Montevideo, Uruguay
Water Resources: R o de la Plata

Ken Nagy
University of California, Los Angles, CA, USA
Animal Physiology and Global Environmental Change

Douglas Nakashima
UNESCO, Paris, France
Indigenous Knowledge, Peoples and Sustainable Practice

Takakiyo Nakazawa
Tohoku University, Tohoku, Japan
Carbon Dioxide, Recent Atmospheric Trends

Choucri Nazli
Massachusetts Institute of Technology, Cambridge, MA, USA
Governance and International Management

Mark Newman
Santa Fe Institute, Santa Fe, NM, USA
Extinctions and Biodiversity in the Fossil Record

Abdul Rahim Nik
Forest Research Institute Malaysia, Kuala Lumpur, Malaysia
Forest Logging Systems in Tropical Countries: Differential Impacts

Carlos Nobre
Brazilian Space Research Institute, Brazil
Large Scale Biosphere–Atmosphere Experiment in Amazonia (LBA)

Richard J Norby
Oak Ridge National Laboratory, Oak Ridge, TN, USA
Temperature Increase: Effects on Terrestrial Ecosystems

Richard B Norgaard
University of California, Berkley, CA, USA
Discounting • Ecological Economics • Emergy • Environmental Economics • ISEW (Index of Sustainable Economic Welfare) and GPI (Genuine Progress Indicator)

Gerald R North
Texas A&M University, College Station, TX, USA
Energy Balance and Climate

Patrick D Nunn
The University of the South Paci c, Suva, Fiji
Demographic Change: Peopling of the Pacific Islands

Rory O Brien
University of Toronto, Toronto, Canada
Green Investment • Socially Responsible Investment • Tradable Permits for Limiting Stratospheric Ozone Depletion

Phil O Keefe
University of Northumbria at Newcastle, Newcastle-upon-Tyne, UK
Biomass Burning in Rural Homes in Tropical Areas • Biomass Use for Urban Fuels in Developing Countries • Refugees and Human Conflict

Tim O Riordan
University of East Anglia, Norwich, UK
Precautionary Principle

Thoraya Obaid
UNFPA, New York, NY, USA
UNFPA (United Nations Population Fund)

G O Obasi
World Meteorological Organization, Geneva, Switzerland
Lake Victoria

Dennis S Ojima
Colorado State University, Fort Collins, CO, USA
Vegetation Ecosystem Model and Analysis Project (VEMAP)

Frank Old eld
PAGES (Past Global Changes) International Project Of ce, Bern, Switzerland
Climate Change, Abrupt • PAGES (Past Global Changes)

Reuben J Olembo
National Committee on the Implementation of the National Environment and Coordination Act, Nairobi, Kenya
Dowdeswell, Elizabeth

Richard J Olson
Oak Ridge National Laboratory, Oak Ridge, TN, USA
Data Banks: Archiving Ecological Data and Information • Metadata

Naomi Oreskes
University of California, San Diego, CA, USA
Continental Drift • Global Plate Tectonics

Richard E Orville
Texas A&M University, College Station, TX, USA
Lightning

Robert Paehlke
Trent University, Peterborough, Canada
Environmental Politics • Soft Energy Path

Thomas Pagano
University of Arizona, Tucson, AZ, USA
Hydrologic Cycle

Tim Palmer
European Centre for Medium-range Weather Forecasts (ECMWF), Reading, UK
Rossby Waves

Garth W Paltridge
University of Tasmania, Tasmania, Australia
Albedo

Jyoti Parikh
Indira Gandhi Institute of Development Research, Mumbai, India
Consumption Patterns: Economic and Demographic Change

Claire L Parkinson
NASA/Goddard Space Flight Center, Greenbelt, MD, USA
Arctic Ocean

William Parton
Colorado State University, Fort Collins, CO, USA
CENTURY Ecosystem Model • Vegetation Ecosystem Model and Analysis Project (VEMAP)

José M Paruelo
Universidad de Buenos Aires, Buenos Aires, Argentina
Temperate Grasslands

Jonathan A Patz
Johns Hopkins School of Public Health, Baltimore, MD, USA
Urban Climate and Respiratory Disease

Robert W Pearcy
University of California, Davis, CA, USA
Plants Carbon Balance and Growth

W Richard Peltier
University of Toronto, Toronto, Canada
Earth System History • Isostasy

Joyce E Penner
University of Michigan, Ann Arbor, MI, USA
Aerosols, Effects on the Climate • Feedbacks, Chemistry Climate Interactions

Frits Penning de Vries
IBSRAM, Jatujak, Bangkok, Thailand
De Wit, C T

David Pepper
Oxford Brookes University, Oxford, UK
Deep Ecology • Ecocentric, Biocentric, Gaiacentric • Ecosocialism • Enlightenment Project • Homocentric • New Ageism • Political Movements/Ideologies and the Environment • Political Systems and the Environment: Utopianism • Social Ecology

A C Perdomo
SOHMA, Montevideo, Uruguay
Water Resources: R o de la Plata

Marta Pérez-Soba
Plant Research International, Wageningen, The Netherlands
Plant Growth at Elevated CO_2

John S Perry
Alexandria, VA, USA
ACSYS (Arctic Climate System Study) • Air Pressure • CLIC (Climate and Cryosphere) • CLIVAR (CLImate VARiability and Predictability) • GAIM (Global Analysis, Interpretation and Modeling) Program • GARP (Global Atmospheric Research Program) • GAW (Global Atmosphere Watch) • Humidity • IAHS (International Association of Hydrological Sciences) • IGBP-DIS (International Geosphere Biosphere Program Data and Information System) • IGY (International Geophysical Year) • IHP (International Hydrological Program) • International Organizations in the Earth Sciences • JGOFS (Joint Global Ocean Flux Study) • Lapse Rate • LOICZ (Land Ocean Interactions in the Coastal Zone) • Maunder Minimum • Monsoons • Natural Records of Climate Change • Ocean Conveyor Belt • Ocean Drilling Program • Oceanography • Revelle, Roger Randall Dougan • Richardson, Lewis Fry • SCOR (Scienti c Committee on Oceanic Research) • The Earth System • WCC (World Climate Conferences 1979 and 1990) • WCRP (World Climate Research Programme) • WMO (World Meteorological Organization) • WOCE (World Ocean Circulation Experiment)

Anders Persson
European Center for Medium-range Weather Forecasts (ECMWF), Reading, UK
Atmospheric Motions • Coriolis Effect • Rossby, Carl-Gustaf

Dorothy Peteet
Lamont Doherty Earth Observatory, Palisades, NY, USA and NASA/GISS, New York, NY, USA
Younger Dryas

G D Peterson
University of Wisconsin, Madison, WI, USA
Resilience

M J Peterson
University of Massachusetts, Amherst, MA, USA
BAT (Best Available Technology) • International Environmental Law • National Environmental Law

Geoff Petts
The University of Birmingham, Edgbaston, UK
Inter-basin Transfer (IBT) for Water Supplies • River Regulation

Roger A Pielke, Jr
National Center for Atmospheric Research, Boulder, CO, USA
Weather Extremes and Climate Impacts: a Case Study for the United States

Roger A Pielke, Sr
Colorado State University, Fort Collins, CO, USA
Convection • Land Cover and Climate • Planetary Boundary Layer

Lars Pierce
California State University, Monterey Bay, CA, USA
Biome–BGC Ecosystem Model

Orrin H Pilkey
Duke University, Durham, NC, USA
Storm Surge

David Pimentel
Cornell University, Ithaca, NY, USA
Malnutrition, Infectious Diseases and Global Environmental Change

Giovanni Pitari
Università L'Aquila, L'Aquila, Italy
Aerosols, Polar Stratospheric Cloud • Aerosols, Stratosphere

Catherine Plume
World Resource Institute, Washington, DC, USA
Global Forest Watch

Gary A Polis (Deceased)
University of California, Davis, CA, USA
Food Webs

Henry N Pollack
University of Michigan, Ann Arbor, MI, USA
Geothermal Heat • Ground Temperature

Bruce Pond
Ontario Ministry of Natural Resources, Peterborough, Canada
Monitoring in Support of Policy: an Adaptive Ecosystem Approach

Hendrik Poorter
Utrecht University, Utrecht, The Netherlands
Plant Growth at Elevated CO_2

Naomi Poulton
UNEP, Nairobi, Kenya
UNEP (United Nations Environment Programme)

Edan Prabhu
FlexEnergy Inc., Mission Viejo, CA, USA
Low Concentration Methane Bearing Gases

Ghillean T Prance
University of Reading, Reading, UK
Attenborough, David • DIVERSITAS • Tropical Forests

Colin Price
Tel Aviv University, Ramat Aviv, Israel
Lightning and Atmospheric Electricity

Gerda K Priestley
Universitat Autònoma de Barcelona, Barcelona, Spain
Geomorphological Change: Landscape Modification for Recreation

Ron Prinn
Massachusetts Institute of Technology, Cambridge, MA, USA
OH–Radical: is the Cleansing Capacity of the Atmosphere Changing?

L Pritchard
Emory University, Atlanta, GA, USA
Resilience

Alex Pszenny
Massachusetts Institute of Technology, Cambridge, MA, USA
IGAC (International Global Atmospheric Chemistry)

Ron Pulliam
University of Georgia, Athens, GA, USA
Odum, Eugene P

Armando Rabuffetti
Inter-American Institute for Global Change Research (IAI), Sao Jose dos Campos, Brazil
IAI (Inter-American Institute for Global Change Research)

Robert Raiswell
University of Leeds, Leeds, UK
Geological Cycling

Freda Rajotte
Canadian Coalition for Ecology, Dauphin, Canada
Religion and Environment Among North American First Nations

V Ramaswamy
Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA
Infrared Radiation

James T Randerson
California Institute of Technology, Pasadena, CA, USA
Carnegie–Ames–Stanford Approach (CASA)

Esa Ranta
University of Helsinki, Helsinki, Finland
Population Sizes, Changes

David J Rapport
University of Guelph, Guelph, Canada
Ecosystem Health

Peter Raven
Director, Missouri Botanical Garden, St. Louis, MO, USA
Wilson, Edward Osborne

Jerome R Ravetz
Research Methods Consultancy Ltd., London, UK
Post-normal Science

Steve Rayner
Columbia University, New York, NY, USA
Social Science and Global Environmental Change

Helmut Rechberger
Vienna University of Technology, Vienna, Austria
Anthropogenic Metabolism and Environmental Legacies

William E Rees
University of British Columbia, Vancouver, Canada
Ecological Capacity • Ecological Footprint

Henry Regier
University of Toronto, Toronto, Canada
Capacity, Assimilative • Carrying Capacity • Ecosystem Approach • Ecosystem Integrity • Ecosystem Structure and Function • Freshwater Fisheries • Great Lakes Region of North America • Maximum Sustainable Yield (MSY) • Phase Shifts or Flip-flops in Complex Systems • Policies for Sustainable and Responsible Fisheries • Water Resources: Great Lakes Case Study

Peter B Reich
University of Minnesota, St. Paul, MN, USA
Temperate Deciduous Forests

James F Reynolds
Duke University, Durham, NC, USA
Controlled Environment Facilities in Global Change Research • Non-equilibrium Ecology

Peter Rimmer
The Australian National University, Canberra, Australia
Transport: Global Freight and Passenger Flows

Paul G Risser
Oregon State University, Corvallis, OR, USA
Ecotones

John Roberts
Centre for Ecology and Hydrology, Wallingford, UK
Hydrology

David A Robinson
Rutgers University, Piscataway, NJ, USA
Snow

Alan Robock
Rutgers University, New Brunswick, NJ, USA
Volcanic Eruption, El Chichon • Volcanic Eruption, Mt. Pinatubo • Volcanic Eruptions • Volcanic Eruption, Tambora

M Rocca
Duke University, Durham, NC, USA
Plant Dispersal and Migration

John Rodda
Wallingford, Oxfordshire, UK
Hydrology • Runoff

Henning Rodhe
University of Stockholm, Stockholm, Sweden
Arrhenius, Svante • Bolin, Bert

Liudmila Romaniuk
St. Petersburg, Russia
Kondratyev, Nikolai Dmitrievich

Aubrey Rose CBE
Barnet, Hertfordshire, UK
Judaism and the Environment

Richard D Rosen
Atmospheric and Environmental Research, Inc., Lexington, MA, USA
Atmospheric Angular Momentum and Earth Rotation

Norman J Rosenberg
Paci c Northwest National Laboratory, Washington, DC, USA
MINK (Missouri, Iowa, Nebraska, Kansas) Study

William B Rossow
NASA/Goddard Institute for Space Studies, New York, NY, USA
ISCCP (International Satellite Cloud Climatology Project)

Marie Roue
Centre National de la Recherche Scienti que, Paris, France
Indigenous Knowledge, Peoples and Sustainable Practice

Donald T Rowland
The Australian National University, Canberra, Australia
Demographic Change: the Aging Population

Rosemary Ruether
Garrett-Evangelical Theological Seminary, Evanston, IL, USA
Religion and Ecology in the Abrahamic Faiths

Iwona Rummel-Bulska
United Nations Environment Programme, Geneva, Switzerland
Transboundary Movements of Hazardous Wastes

Lindsey E Rustad
USDA Forest Service, Durham, NH, USA
Temperature Increase: Effects on Terrestrial Ecosystems

Wolfgang Sachs
Wuppertal Institute for Climate, Energy, Environment, Wuppertal, Germany
Development and Global Environmental Change

Na s Sadik
United Nations Population Fund (UNFPA), New York, NY, USA
UNFPA (United Nations Population Fund)

Marie Sanderson
University of Toronto, Toronto, Canada
PET (Potential Evapotranspiration) • Thornthwaite, Charles Warren

Catriona Sandilands
York University, Toronto, Canada
Human Body, Immediate Environment

Edward Sarachik
University of Washington, Seattle, WA, USA
TOGA (Tropical Ocean Global Atmosphere)

David Satterthwaite
International Institute for Environment and Development, London, UK
Urban Population Change • Urban Poverty and Environmental Health • Ward, Barbara

Robert A Schiffer
Columbia, MD, USA
Earth Observing Systems

Michael E Schlesinger
University of Illinois at Urbana-Champaign, Urbana, IL, USA
Climate Sensitivity

Edwin K Schneider
Center for Ocean-Land-Atmosphere Studies, Calverton, MD, USA
Hadley Circulation • Intertropical Convergence Zone (ITCZ) • Trade Winds • Walker Circulation

Ann Schram
Stanford University, Stanford, CA, USA
Mooney, Harold A

Wilfried Schroder
Bremen-Roennebeck, Germany
Wegener, Alfred

Ulrich Schumann
Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany
Aircraft Emissions

Daniel M Schwartz
University of Toronto, Toronto, Canada
The Environment and Violent Conflict

Jonathan M O Scurlock
Oak Ridge National Laboratory, Oak Ridge, TN, USA
Data Banks: Archiving Ecological Data and Information • Metadata

Dian J Seidel
NOAA Air Resources Laboratory, Silver Spring, MD, USA
Water Vapor: Distribution and Trends

Keith L Seitter
American Meteorological Society, Boston, MA, USA
Anticyclone • Fronts • Ionosphere • Mesosphere • Meteorology • Rain • Stratosphere • Tropopause • Troposphere

Roderick W Shaw
Rodshaw Environmental Consulting Inc., Nova Scotia, Canada
Emissions Trading • Tradable Permits for Sulfur Emission Reductions

Ramine V Shaw
International Human Dimensions Programme on Global Environmental Change (IHDP), Bonn, Germany
IHDP (International Human Dimensions Programme on Global Change)

Marjorie Shepherd
Meteorological Service of Canada, Toronto, Canada
Use of Epidemiological Data to Develop Air Quality Standards for Short-term and Long-term Environmental Health Benefits

Drew Shindell
Columbia University, New York, NY, USA
Arctic Oscillation

Ross Shotton
Food and Agriculture Organization, Rome, Italy
Transferable Fisheries Quotas (TFQs)

Henry Shue
Cornell University, Ithaca, NY, USA
Equity

Herman H Shugart
University of Virginia, Charlottesville, VA, USA
Forest Gap Models

Jagadish Shukla
Center for Ocean-Land-Atmosphere Studies, Calverton, MD, USA
Charney, Jule Gregory • Monsoons

W Brian Simison
University of California, Berkeley, CA, USA
Plate Tectonics

Ian Simmons
University of Durham, Durham, UK
Global Environmental Change and Environmental History • Prehistoric People and Forest Soils

Joanne Simpson
NASA/Goddard Space Flight Center, Greenbelt, MD, USA
TRMM (Tropical Rainfall Measuring Mission)

Ashbindu Singh
UNEP, Sioux Falls, SD, USA
GRID (Global Resource Information Database, of the United Nations Environment Programme (UNEP) Division of Early Warning and Assessment)

Ronald Skeldon
University of Sussex, Falmer, UK
Migrations: The Environmental Challenge of Population Movements

David Skole
NASA/Goddard Space Flight Center, Greenbelt, MD, USA
Deforestation and Habitat Fragmentation in the Amazon Basin

Vaclav Smil
Winnipeg, Canada
Eutrophication • Phosphorus: Global Transfers

Irina N Sokolik
University of Colorado, Boulder, CO, USA
Dust

Ricard V Solé
Santa Fe Institute, Santa Fe, NM, USA
Extinctions and Biodiversity in the Fossil Record

Richard C J Somerville
University of California, San Diego, CA, USA
Cloud Radiation Interactions • Keeling, Charles David

Soroosh Sorooshian
University of Arizona, Tucson, AZ, USA
Hydrologic Cycle

Jean-Francois Soussana
INRA, Clermont-Ferrand, France
Nitrogen Fixers

Richard Southwood
University of Oxford, Oxford, UK
Elton, C S

Ian F Spellerberg
Lincoln University, Aotearoa, New Zealand
Transport and Species Dispersal

Kelly Sponberg
NOAA, Silver Spring, MD, USA
Climatic Extremes

Detlef F Sprinz
Potsdam Institute for Climate Impact Research, Potsdam, Germany
National Self-interest: a Major Factor in International Environmental Policy Formulation

Jorg Stadelbauer
Albert-Ludwigs-Universitat, Universitat Freiburg im Breisgau, Germany
Tourism: Climate Change and Tourist Resorts

James T Staley
University of Washington, Seattle, WA, USA
Microbial Diversity

Anna Mia Stampe
Danish Space Research Institute, Copenhagen, Denmark
Greenland

William R Stanley
University of South Carolina, Columbia, SC, USA
Oil Industry: Cultural Impacts and Environmental Risk Tolerance

David Stanners
European Environment Agency, Copenhagen, Denmark
Dobr s European Environment Assessment Process • EEA (European Environment Agency) • Vavrousek, Josef

Ingrid Leman Stefanovic
University of Toronto, Toronto, Canada
Environmental Philosophy: Phenomenological Ecology

Will Steffen
IGBP Secretariat Royal Swedish Academy of Sciences, Stockholm, Sweden
IGBP Core Projects • IGBP Terrestrial Transects

Marc K Steininger
NASA/Goddard Space Flight Center, Greenbelt, MD, USA
Deforestation and Habitat Fragmentation in the Amazon Basin

Amy Stevermer
University of Colorado, Boulder, CO, USA
Ultraviolet Radiation

John W B Stewart
University of Saskatchewan, Saskatoon, Canada
Phosphorus Cycle

William M Stigliani
Center for Energy and Environmental Education, University of Northern Iowa Cedar Fall, IA, USA
Buffering Capacity • Contaminated Lands and Sediments: Chemical Time Bombs? • Materials Flow Accounting • Policy Exercises: a Tool for Solving Environmental Policy Dilemmas • Redox Potential

Bjorn Stigson
World Business Council for Sustainable Development (WBCSD), Conches-Geneva, Switzerland
WBCSD (World Business Council for Sustainable Development)

Ronald J Stouffer
Princeton University, Princeton, NJ, USA
Projection of Future Changes in Climate

Colin Summerhayes
UNESCO, Paris, France
Global Ocean Observing System (GOOS)

N Sundararaman
WMO, Geneva, Switzerland
IPCC (Intergovernmental Panel on Climate Change)

Nick Sundt
Of ce of the US Global Change Research Program, Washington, DC, USA
Methane Clathrates

Sutikno
Gadjah Mada University, Yogyakarta, Indonesia
Volcanic Eruptions: Mt Merapi, Indonesia

J Keith Syers
Naresuan University, Phitsanulok, Thailand
Environmental Cadmium in the Food Chain • Soil Amelioration

Hiroaki Takagi
APN Secretariat, Kobe, Japan
APN (Asia–Pacific Network)

Fumio Takashima
Tokyo University Fisheries, Tokyo, Japan
Aquaculture in Asia

Lynne D Talley
Scripps Institution of Oceanography, University of California, La Jolla, CA, USA
Atlantic Ocean • Ocean Circulation • Salinity • Salinity Patterns in the Ocean • Thermohaline Circulation

Derek Taylor
Open University, Milton Keynes, UK
Wind Power Development

Trevor C Telfer
Institute of Aquaculture, Stirling, UK
Aquaculture: Salmon Farming

Jaume Terradas
Universitat Autonoma de Bellaterra, Bellaterra, Spain
Margalef, Ramon

Hans Teunissen
World Meteorological Organization, Geneva, Switzerland
GCOS (Global Climate Observing System)

David J Thomas
Axys Analytical Services Ltd., Sidney, Canada
Oil and the Arctic Environment

Morley Thomas
Environment Canada, Downsview, Canada
Bruce, James • Hare, Kenneth

Michael Thompson
The Musgrave Institute, London, UK and The University of Bergen, Norway
Man and Nature as a Single but Complex System

John B Thornes
King's College, London, UK
Land Degradation in the Mediterranean

H Tiessen
University of Saskatchewan, Saskatoon, Canada
Phosphorus Cycle

Peter Timmerman
University of Toronto, Toronto, Canada
Art and the Environment • Emerging Environmental Issues • Encyclopedias: Compendia of Global Knowledge • Gandhi, Mohandas Karamchand • IFIAS (International Federation of Institutes for Advanced Study) • Life Style, Private Choice, and Environmental Governance • Literature and the Environment • Nature • The Human Dimensions of Global Change • Thoreau, Henry David

Sara F Tjossem
University of Washington, Seattle, WA, USA
May, Robert M

Sylvia S Tognetti
Takhoma Park, MD, USA
Bateson, Gregory

Mostafa K Tolba
International Center for Environment and Development, Cairo, Egypt
Agarwal, Anil • Andersen, Stephen O • Benedick, Richard Elliot • Brinkhorst, Laurens Jan • Conference of Parties • CSD (Commission on Sustainable Development) • Dematerialization and Sustainable Development • ECA (Economic Commission for Africa) • Environmental Data Report • Environmental Responses: an Overview • ESCAP (Economic and Social Commission for Asia and the Pacific) • Goldemberg, Jose • Goodman, Gordon • Kassas, Mohamed • Lang, Winfred • MacNeill, Jim • Obasi, G O P • Ozone Layer: Vienna Convention and the Montreal Protocol • Speth, James Gustave • Stockholm and Beyond • Strong, Maurice • Tickell, Crispin • Toepfer, Klaus • UNDP (United Nations Development Program) • UNIDO (United Nations Industrial Development Organization) • United Nations Environment Programme (UNEP) Saskawa Environment Prize • Volvo Environment Prize • Watson, Robert T

Allen Tough
University of Toronto, Toronto, Canada
Futures Research

John R G Townshend
NASA/Goddard Space Flight Center, Greenbelt, MD, USA
Deforestation and Habitat Fragmentation in the Amazon Basin

Kevin Trenberth
National Center for Atmospheric Research/Climate Analysis Section, Boulder, CO, USA
Earth System Processes • GOALS (Global Ocean–Atmosphere–Land System)

Paul Try
International GEWEX Project Of ce, Silver Spring, MD, USA
GEWEX (Global Energy and Water Cycle Experiment)

Compton J Tucker
NASA/Goddard Space Flight Center, Greenbelt, MD, USA
Deforestation and Habitat Fragmentation in the Amazon Basin

Robert R Twilley
University of Louisiana at Lafayette, Lafayette, LA, USA
Mangrove Ecosystems

Sushel Unninayar
Columbia, MD, USA
Earth Observing Systems

Susan Ustin
University of California, Davis, CA, USA
Remote Sensing

Tapani Vaahtoranta
Geneva Centre for Security Policy, Geneva, Switzerland
National Self-interest: a Major Factor in International Environmental Policy Formulation

Jeroen P van der Sluijs
Utrecht University, Utrecht, The Netherlands
Integrated Assessment • Integrated Assessment, Definition of • ULYSSES (Urban Lifestyles, Sustainability, and Integrated Environmental Assessment)

Herman van Keulen
IBSRAM, Bangkok, Thailand
De Wit, C T

Frank Vanclay
Charles Sturt University, Wagga Wagga, Australia
Social Impact Assessment • Social Impact Assessment (SIA)

Vera Vasll evskaya
M. V. Lomonosov Moscow State University, Moscow, Russian Federation
Kovda, V A

Scott Vaughan
North American Commission for Environmental Cooperation, Montreal, Canada
International Environmental Standards • ISO (International Organization for Standardization) • WTO (World Trade Organization)

Charles Vlek
University of Groningen, Groningen, The Netherlands
Environmental Psychology/Perception

James A Voogt
University of Western Ontario, London, Canada
Urban Heat Island

Christiaan Vrolijk
The Royal Institute of International Affairs, London, UK
Tradable Permits for Greenhouse Gases

Ilija Vukadin
Institute for Oceanography and Fisheries, Split, Croatia
Phosphorus Water Quality Objectives

Amar Wahab
University of Toronto, Toronto, Canada
Environmental Sociology

James C G Walker
The University of Michigan, Ann Arbor, MI, USA
Atmospheric Composition, Past

Diana H Wall
Colorado State University, Fort Collins, CO, USA
Biodiversity in Soils and Sediments: Potential Effects of Global Change

Rusong Wang
Chinese Academy of Sciences, Beijing, China
Eco-engineering to Promote Ecological Sustainability in China in an Era of Global Environmental Change • Land Transformation in China: 3000 Years of Human Ecological History

Richard H Waring
Oregon State University, OR, USA
Temperate Coniferous Forests

Karen J Warren
Macalester College, St. Paul, MN, USA
Ecofeminism

Stephen T Washburn
ENVIRON Corporation, Princeton, NJ, USA
Urban Wastes

E C Weatherhead
University of Colorado, Boulder, CO, USA
Ultraviolet Radiation

Thompson Webb, III
Brown University, Providence, RI, USA
Imbrie, John • Natural Records of Climate Change • Paleoclimatology

David Welch
Paci c Biological Station, Nanaimo, Canada
Fisheries: Pacific Coast Salmon

Robin L Welcomme
Imperial College, London, UK
Freshwater Fisheries

Robert A Weller
Woods Hole Oceanographic Institution, Woods Hole, MA, USA
Ocean Observing Techniques

Gunter Weller
University of Alaska Fairbanks, Fairbanks, AK, USA
Arctic Climate • Permafrost

Douglas Whelpdale
Meteorological Service of Canada, Downsview, Canada
Air Quality, Global • Deposition, Dry • Deposition, Wet • Munn, Robert Edward (Ted)

Daniel R White
Honors College, Florida Atlantic University, Jupiter, FL, USA
Modernity vs. Post-modern Environmentalism

Gilbert White
University of Colorado, Boulder, CO, USA
Geography • Malone, Thomas F

Rodney R White
University of Toronto, Toronto, Canada
Sustainable Cities Policies

Robert Whittaker
University of Oxford, Oxford, UK
Volcanic Eruption, Krakatau

Anne Whyte
Mestor Associates, Russell, Canada
Capacity Building

Diane E Wickland
NASA, Washington, DC, USA
Large Scale Biosphere–Atmosphere Experiment in Amazonia (LBA)

Stephen L Willetts
The Willetts Partnership, Atherstone, UK
Toxic Wastes: Generation and Disposal: a Case Study of UK Practice and Legislation

G R Williams
Guelph, Canada
Daisyworld • Gaia • Gaia Hypothesis • Lovelock, James • Redfield Ratio • Suess Effect

Martin Williams
Adelaide University, Adelaide, Australia
Desertification • Desertification Convention • Desertification, Definition of • Deserts

Michael Williams
University of Oxford, Oxford, UK
Deforestation in Historic Times

Mary P Williams Silveira
United Nations Economic Commission for Europe, Geneva, Switzerland
Agenda 21

Michael Winter
Cheltenham and Gloucester College of Higher Education, Cheltenham, UK
Organic Farming and the Environment

Donna Wise
World Resources Institute (WRI), Washington, DC, USA
WRI (World Resources Institute)

A Wokaun
Paul Scherrer Institute (PSI), Villigen, Switzerland
Public-driven Response to Global Environmental Change

Amanda Wolf
Victoria University of Wellington, Wellington, New Zealand
Quotas in International Environmental Negotiations

Poh Poh Wong
National University of Singapore, Singapore
Tourism as a Global Driving Force for Environmental Change

Philip A Wookey
Uppsala University, Uppsala, Sweden
Tundra

Donald J Wuebbles
University of Illinois at Urban-Champaign, Urbana, IL, USA
Depletion of Stratospheric Ozone • Global Warming Potential (GWP) • Lifetime (of a Gas) • Ozone Depletion Potential (ODP) • Ozone Hole • Residence Time (of an Atom, Molecule or Particle)

Jih C Yang
International Center for Environment and Development, Cairo, Egypt
Dematerialization and Sustainable Development

Sergey Yazev
Astronomical Observatory of Irkutsk State University, Irkutsk, Russia
Tunguska Phenomenon

Chen Yong
China Seismological Bureau, Beijing, China
Seismic Risk (and Risk Assessment)

William Young
Monash University, Victoria, Australia
Transport Infrastructure

Eugenio Yunis
World Tourism Organization, Madrid, Spain
Policies to Achieve Sustainable Tourism

Michael Zammit Cutajar
United Nations Framework Convention on Climate Change, Bonn, Germany
Climate Change: an Emerging Issue in the Global Agenda • United Nations Framework Convention on Climate Change and Kyoto Protocol

Joy B Zedler
University of Wisconsin, Madison, WI, USA
Restoration, Ecosystem

Herman Zimmerman
Division of Earth Sciences, National Science Foundation, Arlington, VA, USA
Oeschger, Hans A

Selected Abbreviations and Acronyms

AAAS	American Association for the Advancement of Science
AABW	Antarctic Bottom Water
AAIW	Antarctic Intermediate Water
AAM	Atmosphere's Angular Momentum
AAO	Antarctic Oscillation
ABA	Abscisic Acid
ABA	Applied Behavior Analysis
ACC	Administrative Committee on Coordination
ACC	Antarctic Circumpolar Current
ACE	Aerosol Characterization Experiment
ACRIM	Active Cavity Radiometer Irradiance Monitor
ACSDE	American Center for the Study of Distance Education
ACSYS	Arctic Climate System Study
ADB	African Development Bank
ADCP	Acoustic Doppler Current Profilers
ADP	Adenosine Diphosphate
AEM	Adaptive Environmental Management
AEP	Agri-Environmental Programs
AERONET	Aerosol Robotic Network
AES	Atmospheric Environment Service
AFL-CIO	American Federation of Labor-Congress of Industrial Organizations
AFRO	Regional Office for Africa
AGAGE	Advanced Global Atmospheric Gases Experiment
AGBM	Ad Hoc Group on the Berlin Mandate
AGCM	Atmospheric General Circulation Model
AGGG	Advisory Group on Greenhouse Gases
AIJ	Activities-Implemented-Jointly
AIRS	Atmospheric Infrared Sounder
AJDE	American Journal of Distance Education
ALE	Atmospheric Lifetime Experiment
ALE/GAGE/AGAGE	Atmospheric Lifetime Experiment/Global Atmospheric Gases Experiment/Advanced Global Atmospheric Gases Experiment
ALNs	Asynchronous Learning Networks
AM	Amplitude Modulation
AMAP	Arctic Monitoring and Assessment Programme
AMIP	Atmospheric Model Intercomparison Project
AMS	Accelerator Mass Spectrometry
AMS	American Meteorological Society
AMSR-E	Advanced Scanning Microwave Radiometer
ANEN	African Non-government Organization Environment Network
ANPP	Above-ground Net Primary Production
AO	Arctic Oscillation
AOC	Areas of Concern
AOGCM	Atmosphere–Ocean General Circulation Model
AOPC	Atmospheric Observation Panel for Climate
AOSIS	Alliance of Small Island States
APAR	Absorbed Photosynthetically Active Radiation
APC	Association for Progressive Communications
APN	Asia–Pacific Network
ARC	Alliance for Religion and Conservation
ARGO	Array for Real-time Geostrophic Oceanography
ARM	Atmospheric Radiation Measurement
ASCEND 21	Agenda of Science for Environment and Development into the 21st Century
ASEAN	South-East Asian Nations
ASPSO	International Association for the Physical Sciences of the Ocean
ASTER	Advance Spaceborne Thermal Emission and Reflection Radiometer
ATOC	Acoustic Thermometer of Ocean Climate
ATP	Adenosine Tri-phosphate
ATSR	Along-Track Scanning Radiometer
AVHRR	Advanced Very High-Resolution Radiometer

AVU	African Virtual University
AWS	Automatic Weather Station
BAHC	Biospheric Aspects of the Hydrological Cycle
BALTEX	Baltic Sea Experiment
BaPMON	Background Air Pollution Monitoring Network
BAS	British Antarctic Survey
BAT	Best Available Technology
BATNEEC	Best Available Technology not Entailing Excessive Costs
BATS	Biosphere–Atmosphere Transfer Scheme
BBS	Breeding Bird Survey
BC	Benefit–Cost
BCE	Before the Common Era
BDT	Telecommunication Development Bureau
BER	Biosphere Enhancement Ratio
BES	British Ecological Society
BFTCS	Boreal–Forest Transect Case Study
BGC	Biogeochemical Cycling
BIBEX	Biomass Burning Experiment
BIC	Baha'i International Community
Biome-BGC	Biome-biogeochemical Cycles
BLB	Boundary-Layer Budget
BMRC	Bureau of Meteorology Research Centre
BOD	Biochemical Oxygen Demand
BOREAS	Boreal Ecosystem–Atmosphere Study
BP	Before the Present
BPOA	Barbados Program Of Action
BS	Black Smoke
BSAP	Black Sea Action Plan
BSE	Bovine Spongiform Encephalopothy
BUND	Bund fur Umvelt und Naturschutz Deutschland
BWRs	Boiling Water Reactors
BYDV	Barley Yellow Dwarf Virus
CABO	Centre for Agrobiological Research
CAC	Command-and-Control
CACGP	Commission on Atmospheric Chemistry and Global Pollution
CAER	Community Awareness and Emergency Response
CAG	Citizen Advisory Groups
CAP	Common Agricultural Policy
CARDE	Canadian Armament Research and Development Establishment
CARs	Central Asian Republics
CART	Cloud and Radiation Test-bed
CAS	Commission for Atmospheric Sciences
CASA	Carnegie–Ames–Stanford Approach
CBA	Cost Benefit Analysis
CBD	Convention on Biological Diversity
CCAMLR	Convention on the Conservation of Antarctic Marine Living Resources
CCC	Canadian Climate Centre
CCC	Chemical Coordinating Center
CCCO	Committee on Climatic Change and the Ocean
CCD	Convention to Combat Desertification
CCE	Coordinating Center of Effects
CCFM	Canadian Council of Forest Ministers
CCN	Cloud Condensation Nuclei
CCOL	Co-ordination Committee on the Ozone Layer
CCPs	Capacity Controlling Parameters
CDIAC	Carbon Dioxide Information Analysis Center
CDM	Clean Development Mechanism
CDS	Catalogue of Data Sources
CDT	Canyon Diablo Triolite
CDU	Christian Democratic Union
CDW	Circumpolar Deep Water
CEA	Cost Effectiveness Analysis
CEA	Cumulative Environmental Assessment
CEC	Cation Exchange Capacity
CEC	Commission of the European Communities
CECSD	Comprehensive Experimental Communities for Sustainable Development
CEEC	Central and Eastern European Countries
CEF	Controlled Environment Facilitie
CEFFIC	European Council of Chemical Manufacturers' Federations
CEI	Central European Initiative
CEOS	Committee on Earth Observation Satellites
CEQ	Council on Environment Quality

CERES	Cloud and Earth Radiant Energy System
CERES	Coalition for Environmentally Responsible Economies
CERs	Certified Emission Reductions
CET	Central England Temperatures
CFC	Chlorofluorocarbon
CFS	Canadian Forest Service
CGCM1	Canadian Global Coupled Model
CGIAR	Consultative Group on International Agricultural Research
CIA	Central Intelligence Agency
CIAP	Climatic Impact Assessment Program
CIESIN	Center for International Earth Science Information Network
CIMMYT	International Maize and Wheat Improvement Center
CIS	Commonwealth of Independent States
CITES	Convention on International Trade in Endangered Species of Wild Fauna and Flora
CIV	Community Importance Value
CLAES	Cryogen Limb Array Etalon Spectrometer
CLAW	Charlson, Lovelock, Andreae and Warren
CLIC	Climate and Cryosphere Initiative
CLIMAP	Climate: Long-range Investigation, Mapping and Prediction
CLIVAR	CLImate VARiability and Predictability
CLRTAP	Convention on Long-range Transboundary Air Pollution
CMA	Chemical Manufacturers Association
CMEAL	CO_2 Model-experiment Activity for Improved Links
CMG	Committee on Mathematical Geophysics
CMIP	Coupled Model Intercomparison Project
CND	Campaign For Nuclear Disarmament
CNES	Centre National d'Etudes Spatiales
CNG	Compressed Natural Gas
CODATA	National Committee for Committee on Data for Science and Technology
COL	Commonwealth of Learning
COMESA	Community of Eastern and Southern Africa
COP	Conference of the Parties
COSPAR	Committee on Space Research
CPEs	Core Programmes
CR	Critically Endangered
CREAD	Inter-American Distance Education Consortium
CRF	Concentration Response Function
CSA	Canadian Standards Association
CSD	Commission on Sustainable Development
CSIC	Consejo Superior De Investigaciones Cient ficas
CSIRO	Commonwealth Scientific and Industrial Research Organization
CSIS	Canadian Security Intelligence Services
CSM	Climate System Model
CSR	Competition Stress Disturbance
CSR	Corporate Social Responsibility
CST	Committee on Science and Technology
CTB	Chemical Time Bomb
CTD	Conductivity, Temperature, and Depth
CTE	Committee on Trade and Environment
CVEs	Collaborative Virtual Environments
CVM	Changing Values and Morality
CZCS	Coastal Zone Color Scanner
CZM	Coastal Zone Management
C&D	Construction and Demolition
D-O	Dansgaard-Oescheger
DAAC	Distributed Active Archive Center
DART	Dynamics of the Arctic Treeline
DCA	Detrended Correspondence Analysis
DEC	Digital Equipment Corporation
DEM	Digital Elevation Model
DEOS	Distance Education Online Symposium
DFE	Design-for-Environment
DGVM	Dynamic Global Vegetation Model
DIC	Dissolved Inorganic Carbon
DIN	Dissolved Inorganic Nitrogen
DIP	Degree of Interest Principle

DIP	Dissolved Inorganic Phase
DIS	Data and Information System
DistEdNet	Distance Education Net
DIWPA	Diversitas Western Pacific and Asia
DJGI	Dow Jones Global Index
DJSGI	Dow Jones Sustainability Group Index
DLDC	Distance Learning and Developing Countries
DMSP	Defense Meteorological Satellite Program
DOE	Department of Energy
DOLY	Dynamic Global Phyto-geography Model
DPSIR	Driving Forces–Pressures–State–Impacts–Responses
DSM	Demand Side Management
DTR	Diurnal Temperature Range
DU	Dobson Units
DVDs	Digital Versatile Disks
DVI	Dust Veil Index
DWBC	Deep Western Boundary Currents
E/MSY	Extinctions Per Million Species Years
EA	Environment Agency
EAC	Eco-Agricultural Counties
EAGGF	European Agricultural Guidance and Guarantee Fund
EBCM	Energy Balance Climate Model
EC	Electrical Conductivity
EC	European Commission
ECA	Economic Commission for Africa
ECE	Economic Commission for Europe
ECLA	Economic Commission for Latin America
ECLAC	Economic Commission for Latin America and the Caribbean
ECOSOC	Economic and Social Council
ECOSOC	United Nations Economic and Social Council
ECOWAS	Economic Community of West African States
EDD	Ecological Demonstration Districts
EDF	Environmental Defense Fund
EdNA	Education Network Australia
EEA	European Environment Agency
EETINA	Environmental Education and Training Institute of North America
EEZ	Exclusive Economic Zone
EFA	Education for All
EFTA	European Free Trade Association
EIA	Environmental Impact Assessment
EIP	Eco–Industrial Parks
EIRIS	Ethical Investment Research Service
EIS	Environmental Impact Statement
EITs	Economies in Transition
EM	Ecological Modernization
EMAN	Ecological Monitoring and Assessment Network
EMAS	Environmental Management Systems
EMEP	European Monitoring and Evaluation Programme
EMRO	Regional Office for the Eastern Mediterranean
EMS	Environmental Management Systems
Emwis	Euro-Mediterranean Information System
ENCOP	Environment and Conflicts Project
ENMOD	Environmental Modification Techniques
ENRICH	European Network for Research in Global Change
ENSO	El Nino Southern Oscillation
EOF	Empirical Orthogonal Function
EOS	Earth Observing System
EOSDIS	Earth Observing System Data and Information System Component
EPA	Environmental Protection Agency
EPIC	Erosion-productivity Impact Calculator
EPOC	Environment Policy Committee
ERBE	Earth Radiation Budget Experiment
ERC	Emission Reduction Credit
ERS	Earth Resources Satellite
ERS	European Remote Sensing Satellite
ERSC	Environmental Remote Sensing Center
ES	Environmental Sustainability

ESA	Environmentally Sensitive Area
ESA	European Space Agency
ESAMI	Eastern and Southern African Management Institute
ESCAP	Economic and Social Commission for Asia and the Pacific
ESE	Earth Science Enterprise
ESM	Earth System Model
EST	Environmentally Sustainable Transport
ET	Emissions Trading
ETH	Eidgenoessische Technische Hochschule
EU	European Union
EWEs	Extreme Weather Events
EXPRESSO	Experiment for Regional Sources and Sinks of Oxidants
FACE	Free Air CO_2 Enrichment
FAD	Feed Attractant Devices
FAGS	Federation of Astronomical and Geophysical Data Analysis Services
FAO	Food and Agricultural Organization
FAR	First Assessment Report
FBI	Federal Bureau of Investigation
FCCC	Framework Convention on Climate Change
FDI	Foreign Direct Investment
FEMA	Federal Emergency Management Agency
FES	Financial–Economic Stimulation
FIA	Forest Inventory and Analysis
FIDIC	Federation Internationale des Ingenieurs-Conseils
FIFE	First ISLSCP Field Experiment
FII	Federated Information Infrastructure
FM	Frequency Modulation
FoE	Friends of the Earth
FPAR	Fraction of Photosynthetically Active Radiation
FRC	Food Conversion Ratio
FRN	Swedish Counsel for Planning and Coordination of Research
FSC	Forest Stewardship Council
FTE	Foundation of Economic Teaching
FTP	File Transfer Protocol
FTPP/FAO	Forests, Trees and People Newsletter
FWCC	First World Climate Conference
FWCW	Fourth World Conference on Women
G'DAY	Generic Decomposition and Yield
GAAP	Generally Acceptable Accreditation Principles
GAIM	Global Analysis, Interpretation and Modeling
GAME	GEWEX Asian Monsoon Experiment
GARP	Global Atmospheric Research Program
GATS	General Agreement for Trade in Services
GATT	General Agreement of Tariffs and Trade
GAW	Global Atmosphere Watch
GCDIS	Global Change Data and Information System
GCIP	Continental-scale International Project
GCM	General Circulation Model
GCMD	Global Change Master Directory
GCOS	Global Climate Observing System
GCR	Galactic Cosmic Ray
GCSS	GEWEX Cloud System Study
GCTE	Global Change and Terrestrial Ecosystems
GCTE-NEWS	Global Change in Terrestrial Ecosystem's Network of Ecosystem Warming Study Sites
GCTE/IGBP	Global Change of Terrestrial Ecosystems/International Geosphere–Biosphere Programme
GDP	Gross Domestic Product
GDPS	Global Data Processing System
GEC	Global Environmental Change
GECHS	Global Environmental Change and Human Security
GEF	Global Environment Facility
GEIA	Global Emissions Inventory Activity
GEMS	Global Environment Monitoring System
GEN	Global Environmental-labeling Network
GEO	Global Environment Outlook
GEOSECS	Geochemical Ocean Sections

GESAMP	Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection
GEWEX	Global Energy and Water Cycle Experiment
GFDL	Geophysical Fluid Dynamics Laboratory
GIA	Glacial Isostatic Adjustment
GIN	Greenland, Iceland and Norwegian
GIS	Geographic(al) Information Systems
GISP	Global Invasive Strategies Program
GISS	Goddard Institute for Space Studies
GIWA	Global International Waters Assessment
GKD	Global Knowledge for Development
GLA	Goddard Laboratory for Atmospheres
GLASOD	Global Assessment of Soil Degradation
GLOBE	Global Learning and Observations to Benefit the Environment
GLOBEC	Global Ocean Ecosystem Dynamics
GLOSS	Global Ocean Sea Level Observing System
GLWQA	Great Lakes Water Quality Agreement
GMOs	Genetically Modified Organisms
GMT	Greenwich Mean Time
GNP	Gross National Product
GO_3OS	Global Ozone Observing System
GOALS	Global Ocean Atmosphere Land System
GODAE	Global Ocean Data Assimilation Experiment
GOES	Geostationary Operational Environmental Satellite
GOOS	Global Ocean Observing System
GOS	Global Observing System
GPCC	Global Precipitation Climatology Center
GPCP	Global Precipitation Climatology Project
GPI	Genuine Progress Indicator
GPP	Gross Primary Productivity
GPS	Global Positioning System
GRDC	Global Runoff Data Center
GRI	Global Reporting Initiative
GRID	Global Resource Information Database
GRIP	Greenland Ice Core Project
GSSD	Global System for Sustainable Development
GTOS	Global Terrestrial Observing System
GTS	Global Telecommunication System
GTU/GTTI	Global Telecommunication University/Global Telecommunication Training Institute
GUO	Global Urban Observatory
GVaP	Global Water Vapor Project
GWP	Global Warming Potential
GWP	Gross World Productivity
HABs	Harmful Algal Blooms
HACCP	Hazard Analysis Critical Control Point
HadCM2	Hadley Center for Climate Research and Prediction
HadRT	Hadley Center, United Kingdom Meteorological Office
HALOE	Halogen Occultation Experiment
HAO	High Altitude Observatory
HAPEX	Hydrologic Atmospheric Pilot Experiment
HCFC	Hydrochlorofluorocarbon
HDF	Hierarchical Data Format
HDGCP	Human Dimensions of Global Change Programme
HDPE	High Density Polyethylene
HELCOM	Helsinki Commission
HFC	Hydrofluorocarbon
HIRS	High Resolution Infrared Sounder
HLW	High Level Waste
HMs	Heavy Metals
HNC	Hoy No Circula
HNR	Human–Nature Relationship
HPS	Hantavirus Pulmonary Syndrome
HRDI	High Resolution Doppler Imager
HRDLS	High Resolution Dynamic Limb Sounder
HRV	High Resolution Visible
HSB	A Humidity Sounder for Brazil

HWRP	Hydrology and Water Resources Programme
IA	Integrated Assessment
IAEA	International Atomic Energy Agency
IAG	International Association of Geodesy
IAGA	International Association of Geomagnetism and Aeronomy
IAHS	International Association of Hydrological Sciences
IAI	Inter-American Institute for Global Change Research
IAM	Integrated Assessment Model
IAMAS	International Association of Meteorology and Atmospheric Science
IAPSO	International Association for the Physical Sciences of the Oceans
IASPEI	International Association of Seismology and Physics of the Earth's Interior
IAVCEI	International Association of Volcanology and Chemistry of the Earth's Interior
IBOY	International Biodiversity Observation Year
IBP	International Biological Programme
IBS	Institute for Biological and Chemical Research on Field Crops and Herbage
ICACGP	International Commission of Atmospheric Chemistry and Global Pollution
ICAE	International Commission for Atmospheric Electricity
ICAO	International Civil Aviation Organization
ICC	International Chamber of Commerce
ICCD	International Convention on Combating Desertification and Drought
ICCl	International Commission of Climate
ICCP	International Commission for Clouds and Precipitation
ICDM	International Commission for Dynamic Meteorology
ICED	International Center for Environment and Development
ICEE	Independent Commission on Environmental Education
ICES	International Council for the Exploration of the Sea
ICESat	Ice Sheet Altimetry Mission
ICLEI	International Council for Local Environmental Initiatives
ICMMA	International Commission of Meteorology for the Middle Atmosphere
ICOLP	International Cooperative for Ozone Layer Protection
ICPAE	International Commission for Planetary Atmospheres and their Evolution
ICPD	International Conference on Population and Development
ICPM	International Commission for Polar Meteorology
ICPs	International Coordinating Programmes
ICRAF	International Council for Research in Agroforestry
ICRC	International Committee of the Red Cross
ICREA	International Commodity-related Environmental Agreements
ICRP	International Commission on Radiation Protection
ICSU	International Council for Science
ICT	Information and Communication Technologies
ICZM	Integrated Coastal Zone Management
IDEP	Institute for Economic Development and Planning
IDGEC	Institutional Dimensions of Global Environmental Change
IDNDR	The International Decade for Natural Disaster Reduction
IDRC	International Development Research Center
IDWSSD	International Drinking Water Supply and Sanitation Decade
IEA	Integrated Environmental Assessment
IEA	International Energy Agency
IEC	Information, Education, Communication
IEEA	Integrated Environmental and Economic Accounting
IEEP	International Environmental Education Program

IFCS	International Forum on Chemical Safety
IFF	Intergovernmental Forum on Forests
IFIAS	International Federation of Institutes for Advanced Study
IFRC	International Federation of the Red Cross and Red Crescent
IFS	Integrated Farming Systems
IGAC	International Global Atmospheric Chemistry
IGBP	International Geosphere-Biosphere Programme
IGBP DIS	International Geosphere Biosphere Programme Data and Information System
IGBP LUCC	International Geosphere Biosphere Programme Land-Use and Land-Cover Change
IGEMS	Integrated Global Earth Monitoring Systems
IGFA	International Group of Funding Agencies
IGIDR	Indira Gandhi Institute of Development Research
IGO	International Governmental Organization
IGOS	Integrated Global Observing Strategy
IGY	International Geophysical Year
IHD	International Hydrological Decade
IHDP	International Human Dimensions Programme on Global Environmental Change
IHP	International Hydrological Program
IIASA	International Institute for Applied Systems Analysis
IIED	International Institute for Environment and Development
IIP	Instituto De Investigaciones Pesqueras
IISD	International Institute for Sustainable Development
IJC	International Joint Commission
ILEC	International Lake Environment Committee Foundation
ILO	International Labor Organization
ILTER	International LTER
ILW	Intermediate Level Wastes
IMAGE	Integrated Model to Assess the Greenhouse Effect
IMAR	Inner Mongolia Autonomous Region
IMF	International Monetary Fund
IMI	International Meteorological Institute
IMO	International Maritime Organization
IMO	International Meteorological Organization
IMS	Information Management System
INC	Intergovernmental Negotiating Committee
IOC	Intergovernmental Oceanographic Commission
IOC	International Ozone Commission
IOI	International Ocean Institute
IOJ	The Institute of Jainology
IOMC	Inter-organization Program for the Sound Management of Chemicals
ION	International Ocean Network
IPAT	I is a Function of Population Size P, Af uence Per Capita A, and Technological Development T
IPCC	Intergovernmental Panel on Climate Change
IPCS	International Program on Chemical Safety
IPF	Intergovernmental Panel on Forests
IPM	Integrated Pest Management
IPO	International Project Of ce
IPPC	International Plant Protection Convention
IPPs	Independent Power Producers
IPR	Intellectual Property Right
IPS	Investment Promotion Service
IR	Infrared
IRC	International Radiation Commission
IRI	International Research Institute for Climate Prediction
IRPTC	International Register of Potentially Toxic Chemicals
IRRC	Investor Responsibility Research Center
IRRI	International Rice Research Institute
ISAAA	International Service for the Acquisition of Agri-Biotech Applications

ISAMS	Improved Stratospheric and Mesospheric Sounder
ISCCP	International Satellite Cloud Climatology Project
ISEE	International Society for Ecological Economics
ISEW	Index of Sustainable Economic Welfare
ISLSCP	International Satellite Land Surface Climatology Project
ISO	International Organization For Standardization
ISRIC	International Soil Reference and Information Center
ISSC	International Social Science Council
ITCZ	Inter-Tropical Convergence Zone
ITEX	International Tundra Experiment
ITQ	Individual Transferable Quota
ITU	International Telecommunication Union
IUBS	International Union of Biological Sciences
IUCN	World Conservation Union
IUGG	International Union of Geodesy and Geophysics
IUMS	International Union of Microbiological Societies
IVI	Ice-Core Volcanic Index
IWC	International Whaling Commission
IWRB	International Waterfowl and Wetlands
JEMA	Japan Electrical Manufacturers' Association
JERS	Japanese Earth Resources Satellite
JGOFS	Joint Global Ocean Flux Study
JOC	Joint Organizing Committee
JPIC	Justice, Peace and Integrity of Creation
JPSS	Just, Participatory and Sustainable Society
JSC	Joint Scientific Committee
KLD	Kinder Lydenburg and Domini
KRIP	Kano River Irrigation Project
LAI	Leaf Area Index
LANDSAT	Land Satellite
LAR	Leaf Area Ratio
LBA	Large Scale Biosphere–Atmosphere Experiment in Amazonia
LCA	Life Cycle Assessment
LCDW	Lower Circumpolar Deep Water
LDCs	Less Developed Countries
LDEO	Lamont–Doherty Earth Observatory
LEAD	Leadership for Environment and Development
LETS	Local Employment and Trading Schemes
LGM	Last Glacial Maximum
LHS	Leaf Height Seed
LI	Low Intensity
LIA	Little Ice Age
LIFDC	Low Income Food Deficit Countries
LIS	Laurentide Ice Sheet
LLN	Louvain-La-Neuve
LLW	Low Level Wastes
LMD	Laboratoire De Meteorologie Dynamique
LMF	Leaf Mass Fraction
LMOs	Living Modified Organisms
LNWT	Low and Non-waste Technologies
LOAEL	Lowest Observable Adverse Effect Level
LOICZ	Land–Ocean Interactions in the Coastal Zone
LOS	Law of the Sea
LPG	Liquid Petroleum Gas
LP–HC	Low Probability–High Consequence
LRTAP	Convention on Long-Range Transboundary Air Pollution
LTER	Long-term Ecological Research
LUCC	Land Use and Land Cover Change Project
LUCF	Land-use Change and Forestry
LULU	Locally Unacceptable Land Use
LWF	Learning Without Frontiers
MAAs	Mycosporine-like Amino Acids
MAAT	Mean Annual Air Temperatures
MAB	Man and the Biosphere
MAC	Maximum Admissible Concentration

MACS	Mobile Air Conditioning Society
MAFF	Ministry of Agriculture, Fisheries and Food
MAI	Multilateral Agreement on Investment
MAP	Mediterranean Action Plan
MARC	Monitoring and Assessment Research Center
MARPOL	Marine Pollution Convention
MARPOL/73	International Convention for the Prevention of Pollution from Ships, 1973
MASTER	Method for the Analysis of the Substance Concentrations of Treatment End-products and Recycling Goods
MBL	Marine Biological Laboratory
MBLA	Multi-beam Laser Altimeter
MCS	Monitoring, Control and Surveillance
MCS	Multiple Chemical Sensitivity Syndrome
MCSD	Mediterranean Commission on Sustainable Development
MEAs	Multilateral Environmental Agreements
MEDPOL	Mediterranean Pollution Program
MEPC	Marine Environment Protection Committee
METAP	Mediterranean Environment Technical Assistance Program
MFA	Materials Flow Accounting
MGO	Main Geophysical Observatory
MIDAS	Minimal Impact Dairy Systems
MINK	Missouri, Iowa, Nebraska, Kansas
MIPS	Material Input Per Service Unit
MIS	Marine Isotope Stage
MISR's	Multi-angle Imaging Spectrometer's
MISSR	Multi-angle Imaging Spectroradiometer
MIT	Massachusetts Institute of Technology
MITI	Ministry of International Trade and Industry
MJO	Madden-Julian Oscillation
MJRA	Michael Jantzi Research Associates
MlDAs	Maritime Industrial Development Areas
MLS	Microwave Limb Sounder
MODIS	Moderate Resolution Imaging Spectrometer
MOO	Multiple Object-oriented
MOP	Meeting of the Parties
MOPITT	Measurement of Pollution in the Troposphere
MPC	Maximum Permissible Concentration
MRT	Mean Residence Time
MSC	Meteorological Service of Canada
MSC–E&W	Meteorological Synthesizing Center–East & West
MSEY	Maximum Sustainable Economic Yield
MSS	Multi-Spectral Scanner
MSU	Microwave Sounding Unit
MSW	Municipal Solid Waste
MSY	Maximum Sustainable Yield
MTOs	Multimodal Transport Operators
MUDs	Multiple User Domains
MUSE	Multiple User Social Environments
MWP	Medieval Warm Period
NAAQS	National Ambient Air Quality Standards
NADW	North Atlantic Deep Water
NAFO	Northwest Atlantic Fisheries Organization
NAFTA	North American Free Trade Agreement
NAM	Northern Hemisphere Annular Mode
NAO	North Atlantic Oscillation
NAPAP	National Acid Precipitation Assessment Program
NAR	Net Assimilation Rate
NAS	National Academy of Sciences
NASA	National Aeronautics and Space Administration
NASDA	National Space Development Agency
NAST	National Assessment Synthesis Team
NATO	North Atlantic Treaty Organization
NATT	Northern Australia Tropical Transect
NBII	National Biological Information Infrastructure
NBP	Net Biome Production
NCAR	National Center For Atmospheric Research

NCDC	National Climatic Data Center
NCEP	National Center for Environmental Prediction
NCPC	National Cleaner Production Center
NCSA	National Center for Supercomputing Applications
NDIR	Non-dispersive Infrared
NDVI	Normalized Difference Vegetation Index
NECC	North Equatorial Countercurrent
NECT	North East China Transect
NEDI	National Environmental Data Index
NEP	Net Ecosystem Production
NEPA	National Environment Policy Act
NERC	Natural Environment Research Council
NESDIS	National Environmental Satellite, Data, and Information Service
NetCDF	Network Common Data Form
NEUC	North Equatorial Undercurrent
NFIP	National Flood Insurance Program
NGDC	National Geophysical Data Center
NGO	Non-Governmental Organization
NHANES II	National Health and Nutrition Examination Survey
NHC	National Hurricane Center
NHMSs	National Hydrometeorological Services
NIEO	New International Economic Order
NIMBY	Not In My Backyard
NIR	Near-infrared
NIS	Network Information System
NMAT	Nighttime Marine Air Temperature
NMHSs	National Meteorological and Hydrological Services
NMVOCs	Non-methane Volatile Organic Compounds
NOA	Needs, Opportunities and Abilities
NOAA	National Oceanic and Atmospheric Administration
NOAA/AVHRR	Advanced Very High Resolution Radiometer – National Oceanographic and Atmospheric Agency
NOAA/CMDL	National Oceanic and Atmospheric Administration/ Climate Monitoring and Diagnostics Laboratory
NOAEL	No-Observable Adverse Effect Level
NODC	National Ocean Data Center
NPF	Northern Prawn Fishery
NPOESS	National POESS
NPP	Net Primary Production
NPPs	Nuclear Power Plants
NPV	Net Present Value
NRC	Nuclear Regulatory Commission
NRCS	Natural Resources Conservation Service
NSA	Nitrate Sensitive Area
NSDAF	National Soils Data Access Facility
NSF	National Science Foundation
NSM	New Social Movement
NSOW	Nordic Sea Overflow Water
NSR	Northern Sea Route
NVZ	Nitrate Vulnerable Zone
NWIS	National Water Information System
NWT	Northwestern Territories
OALOS	Ocean Affairs and Law of the Sea
OCh	Organizational Change
OCM	Ocean Color Monitor
OCS	Photolysis of Carbonyl Sulfide
ODA	Official Development Assistance
ODP	Ocean Drilling Program
ODP	Ozone Depletion Potential
ODS	Ozone Depleting Substances
OECD	Organisation for Economic Co-Operation and Development
OFIA	Ontario Forest Industries Association
OGCM	Oceanic General Circulation Model
OLS	Operational Linescan System
OOPC	Ocean Observation Panel for Climate
OPEC	Organization of Petroleum Exporting Countries
ORNL	Oak Ridge National Laboratory

ORNL DAAC	Oak Ridge National Laboratory Distributed Active Archive Center
OSPARCOM	Oslo and Paris Commission
OSY	Optimum Sustainable Yield
OTAG	Ozone Transport Assessment Group
OTEC	Ocean Thermal Energy Conversion
OWC	Oscillating Water Column
PAC	Public Advisory Committees
PAGES	Past Global Changes
PAHO	Pan American Sanitary Bureau
PARN	Pacific and Arctic Railway and Navigation Company
PBL	Planetary Boundary Layer
PCAST	President's Council of Advisors on Science and Technology
PCBS	Programme on Capacity Building in Science
PCC	Program Coordinating Center
PDO	Pacific Decadal Oscillation
PEM	Particle Environment Monitor
PEPC	Phosphoenol–Pyruvate Carboxylase
PERSIANN	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks
PET	Potential Evapotranspiration
PETA	People for the Ethical Treatment of Animals
PFC	Perfluorocarbons
PFRA	Prairie Farm Rehabilitation Administration
PFT	Plant Functional Type
PGF	Pressure Gradient Force
PGRFA	Plant Genetic Resources for Food and Agriculture
PhAA	Provision of Physical Alternatives, (Re)arrangements
PIA	Participatory Integrated Assessment
PIC	Prior Informed Consent
PMIP	Paleoclimate Modeling Intercomparison Project
PNA	Pacific North American Teleconnection Pattern
PNA	Pacific–North American
POES	Polar Orbiting Environmental Satellites
POP	Persistent Organic Pollutant
PPP	Polluter Pay Principle
PR	Precipitation Radar
PREPARE	Preventive Environmental Protection Approaches in Europe
PRISMA	Project on Industrial Successes with Waste Prevention
PSE	Producer Subsidy Equivalent
PSI	Pressure–State–Impact
PSMSL	Permanent Service for Mean Sea Level
PTA	Preferential Trade Area for Eastern and Southern Africa
PTWI	Provisional Tolerable Weekly Intake
PWMA	Provincial Wildlife Management Area
QBO	Quasi-Biennial Oscillation
RAINS	Regional Air Pollution Information and Simulation
RCM	Regional Climate Model
RCMs	Radiative–Convective Models
RE	Reduction In Exposure
RETs	Renewable Energy Technologies
RGR	Relative Growth Rate
RIHMI	Russian Research Institute for Hydrometeorological Information
RIOD	International NGO Network on Desertification and Drought
RIS	Reservoir-induced Seismicity
RLA	Red List Authorities
RMMs	Risk Mitigation Measures
RSAS	Royal Swedish Academy of Sciences
RSP	Regional Seas Programme
RTD	Research, Technological Development and Demonstration
RWMAC	Government Radioactive Waste Management Advisory Committee
SA	Social Assessment
SAD	Surface Area Density
SAE	Society of Automobile Engineers
SAFARI	South African Fire-Atmosphere Research Initiative

SAGE	Stratospheric Aerosol and Gas Experiment
SAIDE	South African Institute for Distance Education
SAL	Structural Adjustment Loans
SALT	Savannas in the Long-term
SAMW	Sub-Antarctic Mode Water
SAR	Synthetic Aperture Radar
SAS	Statistical Analysis Software
SBI	Subsidiary Body for Implementation
SBSTA	Subsidiary Body for Scientific and Technological Advice
SC-IDNDR	Special Committee for the International Decade for Natural Disaster Reduction
SCANTRAN	Scandinavian Transect
SCAR	Scientific Committee on Antarctic Research
SCEP	Study of Critical Environmental Problems
SCOPE	Scientific Committee on Problems of the Environment
SCOR	Scientific Committee on Oceanic Research
SCOSTEP	Science Committee on Solar–Terrestrial Physics
SCOWAR	Scientific Committee on Water Research
SCUBA	Self-contained Underwater Breathing Apparatus
SC–IGBP	Scientific Committee for the International Geosphere–Biosphere Programme
SEA	Strategic Environmental Assessment
SEARO	Regional Office for South-east Asia
SEDI	Committee on the Study of the Earth's Deep Interior
SEP	Swiss Eco-Points
SFA	Substance Flow Analysis
SFM	Sustainable Forest Management
SGOMSEC	Scientific Group on Methodologies for the Safety Evaluation of Chemicals
SIA	Social Impact Assessment
SIDS	Small Island Developing States
SIO	Scripps Institution of Oceanography
SLOSH	Sea, Lake and Overland Surges from Hurricanes
SLP	Sea Level Pressure
SMAP	Short and Medium Term Priority Environment Action Program
SMIC	Study of Man's Impact on Climate
SMM	Solar Maximum Mission
SMOW	Standard Mean Ocean Water
SMS	Social Modeling and Support
SNA	System of National Account
SOE	State of the Environment Reports
SOI	Southern Oscillation Index
SOIREE	Southern Ocean Iron Release Experiment
SOLAS	Safety of Life at Sea
SOLSTICE	Solar/Stellar Irradiance Comparison Experiment
SOM	Soil Organic Matter
SOYGRO	SOYbean GROwth Simulation Model
SPARC	Stratospheric Processes and their Role in Climate
SPI	Sustainable Process Index
SPOT	Système Pour l'Observation de la Terre
SPS	Application of Sanitary and Phytosanitary Measures
SPSS	Statistical Package for Social Sciences
SRES	Special Report on Emission Scenarios
SSA	Stratospheric Sulfate Aerosol
SSM	Soft Systems Methodology
SSS	Sea Surface Salinity
SST	Sea Surface Temperature
STAP	Scientific and Technical Advisory Panel
STARs	Special Target Areas of Research
START	Global Change System for Analysis, Research and Training
STC	Scientific and Technical Committee
STF	Subtropical Front
STP	Standard Temperature and Pressure
SUBSTA	Subsidiary Body for Science and Technological Advice
SUCROS	Simple and Universal CROp Growth Simulator
SUSIM	Solar Ultraviolet Spectral Irradiance Monitor
SWCC	Second World Climate Conference

SWIR	Short-Wave Infrared
SWMA	National Wildlife Refuge
TDE	Throughfall Displacement Experiment
TEAP	Technology and Economic Assessment Panel
TEK	Traditional Ecological Knowledge
TEM	Terrestrial Ecosystem Model
TFIAM	Task Force on Integrated Assessment Modeling
TFQ	Transferable Fisheries Quota
TFR	Total Fertility Rate
THC	Thermohaline Circulation
TIR	Thermal Infrared
TIROS-I	Television and Infrared Observation Satellite
TNCs	Transnational corporations
TOC	Technical Options Committee
TOGA	Tropical Ocean Global Atmosphere
TOMS	Total Ozone Mapping Spectrometer
TRACE-A	Transport and Atmospheric Chemistry Near the Equator-Atlantic
TRIPS	Trade-Related Aspects of Intellectual Property Rights
TRMM	Tropical Rainfall Measuring Mission
TVA	Tennessee Valley Authority
TVE	Township and Village Enterprises
TWAS	Third World Academy of Sciences
UARS	Upper Atmosphere Research Satellite
UCAR	University Corporation for Atmospheric Research
UCDW	Upper Circumpolar Deep Water
UCL	Urban Canopy Layer
UCLA	University of Los Angeles
UCT	Coordinated Universal Time
UHI	Urban Heat Island
UKMO	United Kingdom Meteorological Office
ULYSSES	Urban Lifestyles, Sustainability and Integrated Environmental Assessment
UN	United Nations
UNILO	United Nations International Labor Organization
UNCCD	United Nations Convention to Combat Desertification
UNCED	United Nations Conference on Environment and Development
UNCHE	United Nations Conference on the Human Environment
UNCHS	United Nations Centre for Human Settlements
UNCLOS	United Nations Conference on the Law of the Sea
UNCSD	United Nations Commission on Sustainable Development
UNCTAD	United Nations Conference on Trade and Development
UNDP	United Nations Development Program
UNECE	United Nations Economic Commission for Europe
UNEP	United Nations Environment Programme
UNEP-WMO	United Nations Environment Program-World Meteorological Organization
UNESCO	United Nations Educational Scientific and Cultural Organization
UNFCC	United Nations Framework Convention on Climate Change
UNFPA	United Nations Population Fund
UNGA	United Nations General Assembly
UNGASS	Special Session of the United Nations General Assembly
UNHCR	United Nations High Commission for Refugees
UNICEF	United Nations Children's Fund
UNIDO	United Nations Industrial Development Organization
UNITAR	United Nations Institute for Training and Research
UNSCEAR	United Nations Scientific Committee on the Effects of Atomic Radiation
UNU	United Nations University
UNU ZERI	United Nations University Zero Emissions Research Initiative
URL	Unique Resource Locator
USDA	United States Department of Agriculture

USEPA	United States Environmental Protection Agency
USFS	US Forest Service
USGCRP	US Global Change Research Program
USGS	US Geological Survey
USLE	Universal Soil Loss Equation
USSR	Union of Soviet Socialist Republics
VEC	Valued Ecosystem Component
VEI	Volcanic Explosivity Index
VEMAP	Vegetation Ecosystem Model and Analysis Project
VIRS	Visible and Infrared Scanner
VLBI	Very Long Baseline (Radio) Interferometry
VLCCs	Very Large Crude Carriers
VNIR	Visible and Near-Infrared
VOICES	Vision of International Charter on the Environment by Students
VOS	Voluntary Observing Ship
VRML	Virtual Reality Modeling Language
WAIS	West Antarctic Ice Sheet
WBCSD	World Business Council for Sustainable Development
WCASP	World Climate Applications and Services Programme
WCC	World Council of Churches
WCDMP	World Climate Data and Monitoring Programme
WCED	World Commission on Environment and Development
WCIRP	World Climate Impact Assessment and Response Strategies Programme
WCP	World Climate Programme
WCRP	World Climate Research Programme
WDC	World Data Center
WEC	World Energy Council
WEDO	Women's Environment and Development Organization
WEPP	Water Erosion Point Predictor
WFC	World Food Council
WGMS	World Glacier Monitoring Service
WHA	World Health Assembly
WHO	World Health Organization
WICEMII	World Industry Conference on Environmental Management
WINDII	Wind Imaging Interferometer
WMO	World Meteorological Oganization
WMO/IMO	World Meteorological Organization/International Meteorological Organization
WOC	World Ocean Circulation
WOCE	World Ocean Circulation Experiment
WorLD	World Links for Development
WPRO	Regional Office for the Western Pacific
WQB	Water Quality Board
WRI	World Resources Institute
WSSD	World Summit on Social Development
WTO	World Tourism Organization
WTO	World Trade Organization
WWF	World Wildlife Fund
WWW	World Weather Watch
YD	Younger Dryas

Index

Abidjan, Cote d'Ivoire 3:578
Aboriginal re regimes **5:**319
Above-ground net primary production (ANPP) **2:**570
Abrahamic faiths **5:**461
Absolute humidity **1:**432
Absorbed photosynthetically active radiation (APAR) **2:**202
Abutilon theophrasti **2:**214
Abyssocottus **3:**414
Accidental disruptions **5:**273
Acer rubrum **2:**83
Achaean eon **2:**297
Acid rain **1:**318, **4:**74, **5:**498
Acid sulfate soils **3:**151
Acidic deposition **4:**450
Acidi cation **2:**127
Acoustic thermometry **1:**161
ACRIM-III **1:**75
Active Cavity Radiometer Irradiance Monitor (ACRIM) **1:**732
Adalia bipunctata **2:**161
Adaptation
  applications **4:**83
  context **4:**80
  de nitions **4:**80
  determination of need **4:**82
  history **4:**80
  insect pests **3:**384
  private sector corporations and businesses **4:**83
  reasons for assessment **4:**81
  socio-economic systems **4:**83
  species **2:**160, **4:**83
  strategies **2:**79, **4:**80
Adaptive ecosystem, monitoring **4:**120
Adaptive environmental management (AEM) **4:**139
Adaptive management **5:**400
  strategies **4:**179
Adaptive strategies **2:**482
ADCP **1:**70
Adenosine triphosphate (ATP) **2:**191, 194
ADEOS satellite series **1:**80
Advanced very high pressure radiometer (AVHRR) **2:**572
Advanced very high resolution radiometer (AVARR) **2:**528
Aedes aegypti **2:**163
Aeronautical Meteorology Programme **1:**758
Aerosol Characterization Experiment (ACE) **2:**43
Aerosols **1:**26, 387
  anthropogenic **1:**164
  atmospheric **1:**317
  concentrations **1:**180
  effects on climate **1:**162
  effects on clouds **1:**165
  effects on stratospheric chemistry **1:**680
  future effects **1:**166
  gas phase precursors **1:**164
  mineral **1:**347
  secondary sources **1:**163
  sources **1:**163
  stratospheric sulfate (SSA) **1:**483
  sulfate **3:**182
Afar Depression **1:**335
Afar Desert **1:**339
Afforestation **3:**7
  environmental impacts **3:**156
Africa
  tourism **3:**626
  wetlands **3:**399
Agarwal, Anil **4:**140
AGCMs **1:**203, 287, 291, 292
Agenda 21 **4:**140, 178, 205, 363, 473, **5:**427, 431
  contents **4:**141
  further program **4:**142
  implementation **4:**142
  programme **4:**461
  review **4:**142
Agenda setting **5:**57
Aggregate extraction **3:**342
Aging population
  demographics **3:**277
  dependency **3:**281
  future alternatives **3:**282
  regional variations **3:**279
  rejuvenation **3:**279
Agni **5:**305
Agrarian transformation **5:**79
Agricultural Meteorology Programme **1:**758
Agricultural policy **4:**358
  pricing policy examples **4:**359
  schematic presentation **4:**359
Agricultural production **2:**241
Agriculture
  and environment, global views **3:**190
  Chinese **3:**431
  climate change **3:**236
  credit policy **4:**361
  environmental and public health concerns **3:**189
  environmental policies **4:**361
  extensive cereal cultivation **3:**228
  geography of intensi cation **3:**164
  intensi cation **3:**164, 166
    Java **3:**159
    western Europe **3:**162
  intensive **3:**10, 186, 352
  nitrogen cycle **3:**500
  nitrogen uses **3:**499
  property rights policy **4:**361
  research and technology policy **4:**360
  Rhine Basin **3:**106
  salinity **3:**559
  subsidies **3:**168
  sustainable development **4:**358
  water resources **3:**363
  water use **3:**89
  see also Organic farming
Agroforestry **3:**172
  damar **3:**176
  global climate change **3:**177
  history **3:**173
  literature **3:**173
  new systems **3:**174
  research **3:**173
  system analysis **3:**174
  temperate systems **3:**177
  traditional knowledge **3:**173
Agrosystem, complex man-made **3:**175
AIDS **5:**214
Air
  afforestation impacts **3:**158
  chemical composition **1:**319
  composition changes **3:**181
Air monitoring programs **1:**177
A r Mountains **1:**338

Air pollution 1:177, 2:118, 3:148, 678, 4:74, 486
acidifying 4:450
and climate change 3:650
emissions 4:485
fossil fuels 3:649
historical perspective 3:649
potential impacts on climate change 3:652
regional responses 4:515
temperate coniferous forests 2:563
tourism contribution 3:614
urban 3:683
see also LRTAP; RAINS
Air pressure and wind 1:224, 225
Air quality 3:653
global 1:177
standards 3:680, 4:485
urban 3:681
Air temperatures, past 2:547
Air traffic 3:178
scheduled flights 3:640
Air travel 3:644
Airborne sensors 2:524
Aircraft emissions 3:178
Akash 5:304
Alaska, oil industry 3:527
Albedo 1:27, 182, 342, 2:247
definition 1:182
ground-level measurements 1:182
low solar zenith angle 1:183
spectral 1:182
surface 1:183
Algae, harmful 3:380
Algae-fields 4:522
Algal bloom 3:548
Allergens 3:653
Allometric coefficient 2:135
Allometric growth 2:135
Alpine areas 2:75
Alpine plants 2:162
Aluminum oxide 1:155
Amazon Basin
chemical time bomb 3:111
deforestation 3:255
habitat fragmentation 3:255
mercury cycle 3:111
Amazonia, LBA 2:383
American Chamber of Horrors 5:51
American Federation of Labor–Congress of Industrial Organizations (AFL-CIO) 5:251
AMIP (Atmospheric Model Intercomparison Project) 1:756
Ammonia 1:163, 2:433, 443
anthropogenic emissions 3:46
Anarchism 5:439
Ancient teachings 5:470
Andersen, Stephen O 4:143
Anekantavada 5:344
Angiosperms 2:135
Angular momentum 1:211
Animal liberation 5:234
Animal physiology 2:136
Animal production 3:186
environmental impacts 3:189
Animal rights 5:235
Animal sanctuaries 5:347
Animal sources, methane emissions 3:222
Animal wastes 3:188
disposal problems 3:188
methane emissions 3:222
Animals
abundance 2:60
climate change impacts 2:60
distribution 2:60
distributional shifts 2:62
ecology 2:289
endangered species 4:201
extinction 2:57
impact of global environmental change 2:56
migration routes 2:549
physiology 2:136
phytophagous 2:471
plant migration by 2:84
predicting responses to global changes 2:65
Annual average temperature 1:371
Anoxia in coastal waters 3:109
Antarctic 2:74
Antarctic Bottom Water 1:189
Antarctic ice cores 1:188
Antarctic ice sheet 1:186
Antarctic Oscillation (AAO) 1:202
Antarctic ozone hole 1:148
Antarctic upper atmosphere 1:186
Antarctica 1:21, 184, 648
atmosphere 1:188
climate 1:185
geography 1:184
geology 1:184
ice loss 1:467
international research 1:189
ocean 1:188
paleoclimate 1:188
physical characteristics 1:185
sea ice 1:188
weather 1:185
Anthropocene 1:189
Anthropocentrism 5:232
strong 5:232
weak 5:233
see also Non-anthropocentrism
Anthropogenic, use of term 3:190
Anthropogenic emissions
ammonia 3:46
carbon 3:35
consumption vs. production 3:58
global trends 3:58
nitrogen 3:45
nitrogen oxides (NO_x) 3:45
Anthropogenic flows vs. natural flows 3:57
Anthropogenic impacts 2:140
Anthropogenic metabolism
cities 3:67
environmental legacies 3:54
future strategies 3:69
legacies derived from 3:67
limits at back end and not at front side 3:68
phenomenology 3:55
sinks 3:68
Anthropogeomorphology 1:402
Anthropology and global environmental change 5:163
Anthroposphere, metabolism 3:55
Anthroposystem, metabolism 3:77
Anticyclone 1:191
Anti-personnel mines 3:146
Ants, leaf-cutting 2:590
AOGCMs 1:288, 290, 524
Apah 5:305
APN (Asia–Pacific Network) 4:143
Appalachian Mountain Club 5:243
Applications of Meteorology Programme 1:758
Applied geomorphology 1:402
Aquaculture 3:9
Asia 3:196
environmental issues 3:200

overview 3:190
salmon farming 3:200
Aquaducts 3:644
Aquatic ecosystems
aging 2:458
redox potentials 3:108
Aquatic organisms 2:378
Aquifer recharge 1:341
Arable cultivation 3:501
Aral Sea 3:562, 4:533
attempts to save 4:535
demise of 4:533
profile 4:534
protection 4:535
Arboviral diseases 3:692
Arctic 2:74
Arctic Climate System Study (ACSYS) 1:269
Arctic environment, climate change 3:522
Arctic Ocean 1:199
Arctic Oscillation (AO) 1:201
Arctic Oscillation (AO) index 1:202
Arctic ozone hole 1:150
Arctic plants 2:162
Arctic tundra 2:594
Argentiere glacier 1:505
Argentina, terrestrial transects 2:355
Aridity
causes of present-day 1:333
evidence of previous 1:335
Aristotle 5:105
ARM (Atmospheric Radiation Measurement Program) 1:203
Array for Real-time Geostrophic Oceanography (ARGO) 1:69, 537
Arrhenius empirical relationship 2:282
Arrhenius, Svante 1:204
Art
and environment 5:167
classical 5:169
early environmental 5:168
Earth 5:173, 365
medieval period 5:169
modern environment 5:173
Oriental 5:170
Renaissance period 5:170
Romantic period 5:171
Artist and the environment 5:168
Ascend 21 4:43
Asia
aquaculture 3:196
tourism 3:626
Assimilative capacity 3:220
Association of Small Island States (AOSIS) 4:424
Asteroid tsunamis 1:729
Asteroids 1:205
global consequences of impacts 1:208
impact hazard today 1:210
impacts over Earth history 1:205
populations 1:206
Asthenosphere 1:410
Atmosphere
energy budget 1:240
water storage 1:450
zonal and annual mean radiation budget 1:374
see also GAW (Global Atmosphere Watch); Lower atmosphere; Upper atmosphere
Atmosphere–Ocean General Circulation Models. *See* AOGCMs
Atmosphere–ocean interaction 1:22
Atmosphere–ocean mixed layer deep ocean exchange processes 2:199
Atmospheric angular momentum (AAM) 1:211
Atmospheric chemistry 1:284, 2:567
models 1:105
stratosphere. *See* Stratosphere, chemistry
Atmospheric circulation, summary 1:239
Atmospheric component of global climate models. *See* AGCMs
Atmospheric electricity 1:218
Atmospheric flow 1:242
Atmospheric gases 1:146
affecting ozone 1:144
longwave radiation 1:470
Atmospheric general circulation models (AGCMs) 1:285
Atmospheric hazards 3:480
Atmospheric lapse rate. *See* Lapse rate
Atmospheric Lifetime Experiment/Global Atmospheric Gases Experiment/Advanced Global Atmospheric Gases Experiment (ALE/GAGE/AGAGE) 2:458
Atmospheric modeling 1:756
Atmospheric motion 1:221
computer models 1:222
driving and retarding forces 1:223
friction in 1:224
observing 1:222
without friction 1:223
Atmospheric particles 1:473
Atmospheric Research and Environment Programme 1:758
Atmospheric structure 1:243
Atmospheric temperature structure 1:284
Atmospheric transport model intercomparison 2:50
Atomic radiation. *See* UNSCEAR
Attenborough, David 5:175
Attractors, complex systems 4:116, 118
Australia
tourism 3:626
wetlands 3:404
Australian Bureau of Meteorology 1:434
Austria, demographic transition 5:213
Automobiles 3:643
Autotrophy 2:26, 143
Autotropic respiration 2:191
AVHRR 1:72
Avicennia marina 2:400
Awash River 1:339
Azotobacter 2:443

Bacteria 3:218
Baha'i Faith 5:176
proposed courses of action 5:179
BAHC Earth System Modeling 2:51
Baker's Table 2:99
Balance principle 5:338
Baleen whales 3:446
Baltic, water resources 3:720
Baltic Sea 4:516
Declaration 4:518
fisheries 4:518
geology 4:516
map 4:517, 519
natural capital 4:518
oceanographic data 4:518
organism assembly 4:516
Bamako Convention 4:285
Barbados Program of Action (BPOA) 4:416
Barcelona Convention 4:527
Barrow, Alaska 2:594
Basel Convention 3:632, 4:146
amendment 4:147
control system 4:147
principal objectives 4:146
protocol on liability and compensation 4:147
Basel Declaration on Environmentally Sound Management 4:147
BAT (Best Available Technology) 5:183, 330

Bateman, Gregory 5:183
BATNEEC 4:177, 5:330
Bay of Fundy 1:712
Beaked whales 3:446
Benedick, Richard Elliot 4:148
Beneficiary fees 4:197
Benthic attractors 5:425
Benthic organisms 2:145
Bergson, Henri 5:105
Best available technology (BAT) 5:183, 330
Best available technology not entailing excessive costs (BATNEEC) 4:177, 5:330
Betula papyrifera 2:103, 104
Bhopal, India 5:54
Bidirectional reflection distribution function 1:183
Bifurcation diagram 2:455
Big Moose Lake, acidification 3:103
Bilateral environmental standards 4:285
Biocenosis 2:145
Biocentricism 5:217
Biocentrism 5:235
Bioclimatic indices 2:168
Biocomplexity 2:145
Biocontrol methods 2:17
Biodiversity 2:297
  and CO_2 2:221, 223
  and ecosystem functioning 2:24
  and microbial life 2:421
  and organic farming 3:533
  definition 2:159
  freshwater 2:146
  functional 2:20
  functioning 2:22
    empirical approaches 2:22
    experimental tests 2:22
  inland waters 2:381
  lakes 2:381
  rivers 2:381
  sediments 2:152
  soils 2:152
  tundra 2:600
Biodiversity loss 3:137
  tourism contribution 3:615
Biogeochemical cycling (BGC) 2:20, 159, 171, 172, 379, 423, 532, 3:3, 9
Biogeochemical feedbacks 2:102
Biogeochemical fluxes 2:168
Biogeochemical models 2:168
Biogeochemistry 2:20
Bioindicators 2:160
Biological diversity, *see also* Convention on Biological Diversity (CBD)
Biological hazards 3:484
Biological invasions. *See* Invasions
Biological pump 2:370
Biology, global environmental change 2:1
Biomass 3:653, 680
  use for urban fuels 3:217
Biomass burning 3:205
  and desertification 3:287
  field experiments 3:207
  gaseous and particulate emissions 3:206, 209
  geographical distribution 3:208
  global estimates 3:209
  nitrogen 3:207
  nitrogen oxides 3:207
  oxygen 3:207
  tropical areas 3:214
Biomass enhancement ratio (BER) 2:491, 493
Biomass fires 3:490
Biomass fuel power development 3:216
Biomass plantations 3:217
BIOME3 model 2:568
Biome–biogeochemical cycling (BGC) ecosystem model 2:171
Biome distributions 2:167
Biome models 2:166
Biomes 2:166
  carbon and energy fluxes 2:196
Biophages 2:315
Biophysical feedbacks 3:10
Bioremediation 3:218
Biosphere 2:174, 604
  and health 5:500
  carbon metabolism 2:1
  detoxification 3:98
  freshwater perspective 5:201
  future 2:223
  hazards 3:484
  human impact 2:174
  metabolism 2:3
  retoxification 3:98
  structure 2:3
Biosphere enhancement ratio (BER) 2:175
Biosphere reserves 2:175
Biospheric Aspects of the Hydrological Cycle (BAHC) 2:37
Biota 2:21
  Caspian Sea 4:523
  functional classification 2:29
  functional divisions 2:30
  lakes 2:378
  rivers 2:378
  taxonomic versus functional classifications 2:35
Biotic hazards 3:483
Biphenols 3:631
Birds 2:460
  experimental studies 3:603
  extinctions 2:306
    Pacific Islands 2:57
  investigator effects 3:602
  observational studies 3:603
Bjerknes, Jacob 1:245
Black Sea 4:520
  biological diversity 4:521
  decline of biota after 1950 4:522
  dissolved oxygen 4:523
  environmental degradation 4:522
  geological history 4:520
  hydrogen sulphide 4:522
  map 4:521
Bleaching 2:231
  see also Coral bleaching
Blocking highs 1:238
Blood lead levels 3:632
Blue Plan 4:148
Blue Planet prize 4:280
Boiling water reactors (BWRs) 3:509
Bolin, Bert 1:245
Bolling–Allerod interstadial 1:275, 418
Bombs 3:146
Boreal ecosystem–atmosphere study (BOREAS) 2:177, 183
Boreal forest 2:73, 179
  average monthly climate 2:181
  burning 3:208
  carbon flux 4:149
  global distribution 2:180
Boreal–forest transect case study (BFTCS) 2:353
BOREAS 2:177, 183
Borrasus abellifera 3:148
Bosnia 3:548
Boundary-layer budget (BLB) 4:150

Bourdeau, Philippe F J 4:152
Brent Spar 5:184, 385, 388
Brick clay extraction 3:341
Brinkhorst, Laurens Jan 4:153
Broecker, Wallace Smith 1:246
Brogger Peninsula 2:601
Bromine 1:153
Bromine oxide (BrO_x) 1:680
Bromocarbons 1:680
Brower, David 5:185, 285
Brownfield sites 3:343
Bruce, James 4:153
Brundtland, Gro Harlem 4:154
Brundtland Report 5:249
Buddhism 5:185
  early views of nature 5:186
  environmental activism 5:189
  modern ecological views 5:187
  northern views 5:187
  recent ecological history 5:187
  Thailand 5:188
Budyko, Mikhail Ivanovich 1:248
Buenos Aires conference 4:445
Buenos Aires Plan of Action 4:441
Buffering capacity 2:184
Buffers 2:184
Building codes 4:379
  promulgation and effective enforcement 3:732
Burning, biomass. *See* Biomass burning
Business-as-usual scenario 5:191
Butterflies 2:61, 162, 452, 548

C_3 photosynthesis 2:69, 186, 194, 470, 573
C_4 photosynthesis 2:69, 186, 194, 470, 573
Cadmium
  anthropogenic emissions 3:58
  anthropogenic vs. natural flows 3:57
  atmospheric 3:304
  deposition to air, water and soil 4:303
  ecological effects 3:306
  economic issues 3:306
  effect on human health 3:305
  exposure, risk assessment 3:305
  in food chain 3:303
  Provisional Tolerable Weekly Intake (PTWI) 3:306
  Rhine Basin 4:302
  risk management 3:306
  sources 3:304
  urinary 3:305
Calcium, in soil 3:585
Cambrian 2:297
Campaign for Nuclear Disarmament (CND) 5:247
Canada
  emission trading 4:454
  sustainable forest policies 4:351
Canada Geographic Information System 4:225
Canada–US Bilateral Air Quality Agreement 4:285
Canada/US IJC (International Joint Commission) 4:552
Cancer 3:305
  Chernobyl accident 3:242
CAP 3:247
  costs and benefits 3:248
  objectives 3:247
  principal measures 3:247
  reforms 3:248
Capacity, assimilative 3:220
Capacity building
  definition 4:156
  international efforts 4:156
Carbon
  anthropogenic emissions 3:35
  isotope ratios 2:546
  partitioning 2:102
  sequestration 2:444
  source 2:26
  stores 2:196, 197
  terrestrial stores and fluxes 2:190
  see also CENTURY generalized ecosystem model; Redfield ratio
Carbon-10 2:544
Carbon-11 2:544
Carbon-12 2:544
Carbon-13 2:544
Carbon-14 2:544, 557
Carbon-15 2:544
Carbon-16 2:544
Carbon-9 2:544
Carbon balance
  and plant growth 2:497
  deforestation 3:269
Carbon cycle 1:284, 2:95, 191, 198, 549, 4:151
  long-term 1:252
  whole-plant feedbacks 2:102
Carbon dioxide (CO_2) 1:178, 183, 190, 204, 2:572, 3:178
  absorption bands 1:473
  and biodiversity 2:221, 223
  and plant tissue quality 2:221
  and water relations 2:220
  and water vapor exchange 2:194
  anthropogenic emissions 2:201
  atmospheric 1:26, 189, 249, 254, 291, 338, 411, 468, 501, 517, 2:69, 94, 95, 140, 161, 173, 188, 193, 201, 216, 249, 450, 487, 549, 3:181, 216, 4:149, 229
    trends 1:254
  changes during the industrial era 1:250
  climate change 1:411
  concentration 2:175, 191, 3:181, 315, 355, 4:38, 430
  concentrations 1:518
  cycle 2:159
  direct GWPs relative to 1:411
  elated concentrations 2:489
  elevated 2:46, 62, 94, 125, 442, 471, 608
  emissions 1:166, 3:216, 476, 477, 4:325, 428, 5:401, 402
    energy related 3:47
    gas flaring 3:473
    per capita 3:8
  enrichment 2:5, 215
  factors affecting 1:252
  fertilization 4:150, 151
  fixation 2:422
  fixing 3:353
  from 550 million to 500 000 1:251
  global cycle 3:182
  global observing network 4:150
  Greenland 1:416
  ice cores 1:260
  increasing atmospheric 2:563
  interaction with other global change factors 2:494
  interactive effects 2:488
  inter-annual variation in flux from natural sources 1:258
  isotropic signatures 1:258
  measurements
    Mauna Loa record 1:484
    since 1958 1:249
  measuring atmospheric concentrations 1:260
  Model-Experiment Activity for Improved Links (CMEAL) 2:173
  Precambrian 1:252
  pre-industrial changes during last millennium 1:249
  production 2:532
  sampling stations 1:257

Carbon dioxide (CO_2) (*continued*)
sources **1:**368
tropical rainforests **3:**269
tundra **2:**596
Carbon uxes
Boreal Forest **4:**149
tundra **2:**596
Carbon isotope ratio **2:**188, 189, 549
fossil fuels **2:**549
Carbon isotopes in fossils **2:**547
Carbon metabolism, biosphere **2:**7
Carbon monoxide (CO) **1:**179, 261, 384, **3:**680, **4:**486
concentrations **1:**261
levels of **2:**608
Carbon/nitrogen ratio **2:**102, 436
Carbon tetrachloride **1:**153
Carbonate sediments **2:**200
Carbonyl sul de (COS) **1:**483, 681
Cardiac hospitalization rates **4:**488
Cargo ballast water **3:**633
Carnegie-Ames-Stanford Approach (CASA) terrestrial biogeochemical model **2:**202
Carrying capacity **4:**156
maximum human load **3:**298
Cars. *See* Automobiles
Carson, Rachel Louise **5:**51, 192, 247, 430
Caspian basin, map **4:**525
Caspian Lake **4:**521
Caspian Sea **4:**523
biota **4:**524
geological history **4:**523
impacts **4:**524
rehabilitation **4:**526
water level uctuations **4:**524
Caste system **5:**94
Catalytic cycle **1:**677
Catastrophic events. *See* Natural disasters
Catastrophic risks, impact **4:**375
Catchment **2:**345
Cattle grazing
and land cover **3:**221
and pastoral change **3:**224
CBA (cost bene t analysis) **5:**193
Cenozoic cooling **1:**338
Cenozoic desiccation **1:**337
Cenozoic tectonism **1:**337
Centrifugal force **1:**328
CENTURY Ecosystem Model **2:**206, 333
major structural components **2:**206
use and testing **2:**208
CEOS (Committee on Earth Observation Satellites) **1:**71, 261
Cercopithecus patas **1:**339
Cereal cultivation **3:**228
Chamaenerion (Epilobium) angustifolium **2:**559
Change, permanent **5:**391
Change theories **5:**391
Chaos **1:**16, 263, **2:**209, **4:**31
in ecological dynamics **2:**213
mathematical background **2:**210
Chaotic systems **2:**452
Chapman mechanism **1:**676, 677, 680
Charney, Jule Gregory **1:**266
Chemoautotrophs **2:**143
Chemical potential, surrogate **3:**335
Chemical sinks, wetlands as **3:**109
Chemical substitutes, case study **4:**387
Chemical time bomb (CTB)
and climate change **3:**112
concept **3:**99
Chemical weathering **1:**755
Chemoheterotroph **2:**26
Chemosynthesis **2:**422
Chernobyl nuclear disaster **3:**241, **5:**54, 301
Childhood mortality **5:**33
China
agriculture **3:**431
eco-development **3:**433
human ecology **3:**430
land transformation **3:**430
loess sequence **1:**341
sustainability **4:**547
see also Yellow river
Chipko Movement **5:**193, 249, 388
Chlorine **1:**153
Chlorine nitrate ($ClONO_2$) **1:**154, 681
Chlorine oxides (ClO_x) **1:**679
Chlorocarbons **1:**427
Chloro uorocarbons (CFCs) **1:**144, 146, 150, 153, 178, 267, 384, 411, 427, 474, 501, 520, 533, 679, **4:**65, 74, 283, 309, 314, 387, 428, 446, **5:**434
commonly used **1:**267
concentrations **1:**145
estimated global consumption **1:**268
ows and stocks **3:**66
phase-out case study **4:**386
Christianity **5:**104, 194, 211, 461
arrival of **5:**375
challenges for the future **5:**199
environmental responses at local, national and international levels **5:**198
historical background **5:**194
new theological directions **5:**196
theological challenges **5:**195
Chromium, anthropogenic emissions **3:**58
Cirrus clouds **3:**184
CITES (Convention on International Trade in Endangered Species of Wild Fauna and Flora) **4:**157, 283
Cities
and volcanic eruptions **3:**696
linear metabolism **3:**59
parks/gardens movements **5:**363
Safer Cities Programme **4:**462
structural changes affecting **3:**670
sustainable policies **4:**406
world s largest **3:**669
world s most rapidly growing **3:**670
see also Urban
Civil con ict **3:**146
and forest resources **3:**148
Clausius Clapeyron relation **1:**384
Clean Development Mechanism (CDM) **4:**158, 165, 357
Clean energy **3:**710
Clean technology **5:**30
Cleaner production **4:**99, 158
Clear air ice precipitation **1:**320
Clear-sky feedbacks **1:**285
CLIC (Climate and Cryosphere) **1:**269
CLIMAP (Climate Long-range Investigation, Mapping, and Prediction) **1:**269, 469
Climate
analogues **1:**271
and energy balance **1:**371
and weather **1:**15
chemistry interactions with **1:**384
components **1:**308
de nition **1:**270
driving forces **1:**18
effects of aerosols **1:**162
equilibrium states **1:**115
land surface interactions **1:**494

model simulations 1:114
mountain 1:540
natural climate variability 1:120
pre-industrial period 1:123
predictability 1:265
projection of future changes 1:126
sensitivity 1:375
spatial structure 1:24
20th century 1:121
see also CLIC; CLIMAP
Climate Agenda 1:270
Climate change 4:74
abrupt 1:272
and cloud characteristics 1:287
and hydrologic cycle 1:459
and tropical storms 1:438
and variations representation 1:111
and viral diseases 3:690
anthropogenic impacts 2:105
anthropogenic influences 1:25
attribution 1:28
CO_2 1:411
correlations between modeled and observed changes 1:278
crop models 2:244
decision-making to combat 4:254
definition 1:271
desert regions 1:341
detection and attribution 5:1-5
new developments 1:279
disease 2:64
downscaling 1:346
elements of 5:110
factors contributing to 1:389
future implications 1:277
global agenda 4:163
Holocene 3:372
impacts on atmospheric chemistry 1:387
impacts on natural systems 2:201
key recent studies 1:280
millennial scale 1:273
observed 1:23
over past 550 million years 1:252
pests 2:64
problem of 5:400
projections 1:29, 461
regional. *See* Regional climate change
sediments 2:156
soils 2:156
spatial structure 1:24
time-history of recent changes 1:279
United States 4:543
see also Fingerprinting
Climate cycle 4:84
Climate extremes 3:1, 243, 4:230
and natural disasters 4:229
conjunctive 3:246
definition 3:243
examples 3:244
importance of 3:243
trends versus truisms 3:244
Climate feedbacks 1:283, 382
Climate models
fast feedback processes 1:285
simulations
geological past 1:296
reliability 1:114
transient simulations 1:299
Climate policy, international 4:252
Climate predictability. *See* CLIVAR
Climate protection 4:388
Climate sensitivity 1:284, 382
Climate simulation 1:114, 756
Climate stabilization, geoengineering for 3:336
Climate system 1:13
Climate variability 3:653
see also CLIVAR
Climate variations and glacier fluctuations 1:404
Climate warming 4:149
and water use 3:238
potential effects 3:382
Climatic geomorphology 1:402
Climatic Impact Assessment Program (CIAP) 1:678
Climatic response, feedbacks 1:27
Climatology 1:308
definition 1:308
key approaches 1:309
Climax community 5:391
Climax vegetation 2:215
CLIVAR (CLImate VARiability and Predictability) 1:311
Cloud albedo 1:317
Cloud amount 1:288
Cloud and Radiation Test-bed (CART) 1:203
Cloud changes 1:284
Cloud characteristics and climate changes 1:287
Cloud condensation nuclei (CCN) 1:317
Cloud feedbacks 1:287
Cloud height 1:288
Cloud properties and processes 1:285
Clouds 1:19, 27, 316
agricultural regions 1:318
and longwave radiation 1:318
and precipitation 1:316
and shortwave radiation 1:317
effects of aerosols 1:165
lightning 1:501
polar regions 1:319
schematic view 1:317
urban regions 1:318
Club of Rome 4:166, 400, 5:395, 404
Coal, transport movements 3:636
Coalition for Environmentally Responsible Economies (CERES) Principles 4:236
Coastal degradation due to tourism 3:618
Coastal superquarries 3:342
Coastal waters
anoxia in 3:109
eutrophication 3:503
Coastal zones 2:77
environmental change by tourism 3:620
management 4:166
Cocoyoc Symposium on Patterns of Resource Use, Environment and Development Strategies 4:400
Coke manufacture, methane emissions 3:463
Collaboration, global imperative 5:293
Collaborative virtual environments (CVEs) 5:506
Colobanthus quitensis 2:162
Colonizing stage 2:550
Combined environment and development 5:58
Combustion gas disposal 3:710
Comephorous 3:414
Comets 1:205
global consequences of impacts 1:208
impact hazard today 1:210
impacts over earth history 1:205
populations 1:206
Shoemaker-Levy 9 1:205
Command and control (CAC) 5:26
regulation 4:446
Commodification, threat of 5:322
Common Agricultural Policy. *See* CAP

Common but differentiated responsibility 5:329
Common but differentiated responsibility principle (Stockholm/Rio) 4:168
Common concern 5:329
Common property management 5:46
Commons, tragedy of the 5:208, 415
Communications
  developments 3:6
  infrastructure 3:644
Community Development Programme 4:462
Community forest 3:173
Community importance value (CIV) 2:25
Community richness 2:224
Competition 2:485
Competition stress disturbance (CSR) 2:482, 485
Complex dynamics 4:31
Complex hazards 3:483
Complex systems 5:422
  man and nature 5:384
  monitoring 4:116
  properties 4:118
Composting 2:544, 3:686
Compound hazards 3:483
Computer simulation models 1:114, 2:241, 4:251
Concentration response functions (CRFs) 4:487
Conceptual model
  hierarchical structure 4:127
  monitoring 4:130
Conference of Parties (COP) 4:168, 172, 439, 441, 480
Conflict
  and environment 5:271
  and environmental change 5:271
  and warfare 5:274
  see also Environmental stress; Violent conflict
Confucianism 3:431
Conifers
  ecological distribution 2:561
  evolution 2:561
  geographic distribution 2:561
  physiological and morphological adaptations 2:561
Connectivity 2:325
Consequential rules 5:413
Conservation, incentives 4:102
Conservation biology 2:228
Conservation movement 5:244
Conservatism 5:436
Constellations 5:373
Constitutive rules 5:413
Construction 3:7
Construction materials, US 3:57
Consultative Group on International Agricultural Research (CGIAR) 4:361
Consumer-driven dynamics 2:380
Consumption, environmental impact 5:34
Consumption patterns 4:164
  economic and demographic change 3:249
Consumption–pollution link 5:34
Contaminated lands. *See* Land contamination; Soils
Contemporary invasions 2:87
Continental drift 1:321, 647
  historical background 1:321
  lessons from 1:324
  rejection 1:323
Continentality 1:334
Continuity, principle 5:298
Contraction, trajectory of 5:160
Contraction theory 1:321
  alternative 1:322
Contrails 3:182
Controlled environment facilities (CEFs) 2:228
Convection 1:325
Convention on Biological Diversity (CBD) 4:169
  contents 4:169
  ex situ conservation 4:170
  financial resources 4:171
  impacts 4:171
  in situ conservation 4:170
  incentives 4:170
  knowledge base 4:169
  research 4:170
  sustainable use 4:170
Convergence, trajectory of 5:160
Convergent boundaries 1:410
Coral bleaching 2:231, 232
Coral reefs 1:363, 2:163
  degradation due to tourism 3:617
Coregonus autumalis migratorius 3:415, 417
Coriolis effect 1:191, 326
Coriolis force 1:225, 326
  history 1:326
  rotating planet 1:328
  turntable 1:327
  variation 1:329
Coriolis parameter 1:327
Corn industrial engineering 4:548
Corporate Average Fuel Efficiency (CAFE) standards 5:31
Corporate environmental behavior and financial performance 4:235
Corporate environmental leadership 4:66
Corporate environmental reporting 4:236
Corridors 2:324
Cosmic rays 1:476
Cosmology
  Christian–Ptolemaic 5:377
  medieval western 5:376
  pre-modern 5:372
Cost–benefit analysis (CBA) 5:193
Cost–benefit modeling 5:397
Costing, ecosystem services 2:272
Costs, internalization 4:278
Cousteau, Jaques 5:209
CQUESTN 2:445
Crassulacean acid metabolism (CAM) 2:186, 471
Creation principle 5:338
Creative destruction 5:391
Cretaceous 1:207, 329, 516, 2:297
  climate modeling 1:297
Critical loads 3:252, 4:175, 5:417
Crop farming 3:228
Crop models 2:241
  applications 2:245
  climate response 2:243
  comparison exercises 2:244
  history 2:241
  input-output models 2:245
  limitations 2:242
Crops
  environmental effects 2:243
  see also Plants
Crossover 5:398
Crude oil, transport movements 3:636
Crude petroleum, Arctic environment 3:517
Crust 1:503
Crutzen, Paul J 1:330, 677
Cryogen Limb Array Etalon Spectrometer (CLAES) 1:732
Cryosphere 1:330
  and hydrologic cycle 1:450
  see also CLIC
CSD (Commission on Sustainable Development) 4:43, 178
Cultural attitudes 3:6
Cultural eco-feminism 5:439

Cultural eutrophication 4:349
Cultural evolution 5:164
Cultural landscapes 5:319
Cultural oligotrophication 4:349
Cultural theory 5:391, 401
Culture scale 5:73
  dynamics 5:75
Cultures, development 5:155
Cumulative changes 5:273
Cumulative environmental assessment (CEA) 4:178
  methodologies 4:179
Cumuliform clouds 1:326
Cumulus clouds 1:326
Cumulus congestus clouds 1:326
Cumulus convection 1:325
Cumulus humulus clouds 1:326
Curi, Kriton 3:253
Current carbon budget 2:201
Cut-off lows 1:238
Cybernetics 5:410
Cyberspace 5:505
Cyborg 5:410
Cystoseira 4:522

Daisyworld model 2:247
Damar agroforest 3:176
Dams
  environmental impacts 3:553
  world's largest 3:553
Danger, exposure to 5:299
Dansgaard–Oescheger (D–O) cycles 1:272, 273, 332
Dansgaard–Oescheger (D–O) events 1:430
DART (dynamics of the Arctic treeline) 2:354
Darwin, Charles Robert 2:247
Darwin's theory of evolution 2:560
Data and Information System (DIS) 2:40
Data archive 2:257, 258
  functions 2:252
Data banks 2:248
  examples 2:249
  future directions 2:257
  historic examples 2:248
  overview 2:249
  present-day 2:249
  subsetting 2:255
  user interface 2:255
Data collection and storage 4:134
Data delivery 2:255
Data entropy 2:410
Data formats 2:254
Data preservation 2:250
Data rescue 2:257
Data sharing 2:256, 258, 410
Data storage and distribution 2:251
DDD 5:51
DDT 2:379, 3:98, 302, 416, 417, 4:76, 262, 519, 5:51, 248
  environmental incidents 4:518
Dead Sea Rift 1:335
Death, statistics 4:62
Decadal variations 1:367
Decision-making
  macro 3:7
  to combat climate change 4:254
Decision processes, risk management 4:377
Decision systems 3:6
Declaration des droits de l'homme 5:120
Decolonization 5:92, 152
Deep ecology 5:211, 217, 434, 444, 448

Deforestation 1:497, 2:325, 3:7, 158, 732, 4:74
  Amazon Basin 3:255
  and desertification 3:287
  and development 3:270
  carbon balance 3:269
  classical era 3:260
  early modern times 3:262
  global impacts 3:265
  history 3:259
  Middle Ages 3:261
  pre-historic 3:259
  20th century 3:263
Demand management 4:181
Dematerialization 4:18, 183, 249, 5:157, 160
  and cleaner production 41:100
  and information age 4:100
  and sustainable development 4:96
  corporate environmental action 4:99
  technologies 4:99
  understanding 4:98
Democracy and ecology 5:161
Democratic socialism 5:437
Demographic entrapment 3:142
Demographic transition 5:211
Demographics
  aging population 3:277
  and health 5:403
  consumption patterns 3:249
  Pacific Islands 3:273
  world trends 3:16
Dengue fever 3:692
Denitrification 2:287, 423, 432, 445
Dental amalgam 2:407
Denudation rates and mining activities 3:458
Deontological ethics 5:428
Deposition
  dry 2:260
  wet 2:261
Derelict land, re-use 3:343
Descent of Man, The 5:237
Deschampsia antarctica 2:162
Deschampsia exuosa 2:437, 438
Desert Margins Initiative (Nairobi) 4:185
Desertification 3:10, 282
  and drought 3:285
  and overgrazing 3:285
  causes and consequences 3:284
  convention 4:183
  definition 3:282, 283
  extent and severity 3:283
  impediments to combating 4:184
  lessons learned 3:289
  modern myths 1:342
  national and local responses 4:319
  overview 4:184
  see also UNCCD
Deserts 1:332, 2:70
  definition 1:332
  dust 1:342
  encroachment 1:342
  environmental change 1:343
  evidence of formerly wetter climates 1:335
  groundwater 1:341
  landforms antiquity 1:337
  past climatic events 1:339
  prehistoric peoples 1:339
  tropical anticyclonic 1:334
  world distribution 1:333
Detergents, phosphates in 3:544
Deterritorialization 5:157

Detoxification, biosphere 3:98
Detritus 2:314
Developing countries 4:165, 183, 5:200
  fertility levels 3:20
  life expectancies 3:24
  national trends 4:8
  population trends 3:16
  sustainable transportation 4:433
  technological innovation and leadership 4:94
Development 5:150
  and environment 5:58
  and population 5:32
  and security 5:159
  chrono-politics 5:151
  conservation in time 5:154
  cultures 5:155
  decline 5:157
  geo-politics 5:151
  history 5:150
  legacy of 5:154
  projects 5:273
  prospects 5:159
  social polarization 5:154
  Third World 5:153
Development projects, environmental assessment (EA) 4:203
Dew point 1:433
De Wit, C T 2:259
1,1-Dichloro-2,2-*bis*-(p-chlorophenyl)ethane. *See* DDD
Dichlorodiphenyltrichloroethane. *See* DDT
2,4-Dichlorophenoxyacetic acid 4:262
Diesel engines 3:678
Digital formats 2:254
Dimethylsulfide (DMS) 1:163, 284, 343, 348, 384, 2:373, 425
Dinitrogen 3:494
Dinitrogen pentoxide (N_2O_5) 1:681
Diome distribution/biogeochemistry models 2:169
Dioxin 5:54
DIS Land Cover Classification 2:52
Disasters 5:54
  experience 5:298
  explanations 5:298
  myths 5:298
  reversal of expertise 5:302
  socially discriminate or indiscriminate events 5:299
  technocratic 5:300
  see also Natural disasters
Discounting 5:28, 214
  ecosystem services 2:279
Discrete logistic model 2:454
Diseases
  and pests 2:64
  vector-borne 3:691
Dissolved inorganic carbon (DIC) 22;64
Distance learning 1:468, 4:186
  digital divide 4:188
  electronic mailing lists 4:190
  history and evolution 4:187
  international organizations 4:190
  newsgroups 4:190
Distichlis 3:451
Distributions 2:160
Disturbance response groups, identification 2:486
Disturbances 2:261, 550
  anthropogenic causes 2:262
  global change 2:264
  management implications 2:264
  natural causes 2:262
  recovery process 2:266
Divergent boundaries 1:410
DIVERSITAS 2:268, 4:42
DMAPP 2:607
Dobr s Assessment 4:547
Dobson spectrophotometer 1:146
Dobson units (DU) 1:141, 676
Dominion theology 5:196, 492
Dooge, James 4:191
Dow Jones global index (DJGI) 4:235
Dow Jones sustainability group index (DJSGI) 4:235
Dowdeswell, Elizabeth 4:192
Downburst 1:325
Downscaling, climate change 1:346
Downstream development 1:238
Downward convection 1:325
DPSIR (Driving Forces–Pressures–State–Impacts–Responses) 4:44, 193
Drainage basin 2:345
Dreissena polymorpha 2:534, 4:524
Drinking water 3:85
  see also IDWSSD
Drosophila 2:62
Drosophila subobscura 2:161
Drought 1:477, 2:347, 3:244, 487
  and desertification 3:285
Dry convection 1:325
Dry deposition 2:260
Dry microburst 1:325
Dumping, ecological 4:198
Dumping of wastes. *See* London Dumping Convention
Dust 1:347
  deserts 1:342
  emission 1:349
  impacts 1:348
  measurements and modeling of properties 1:350
  particles 1:348
  properties during atmospheric transport 1:349
Dust veil index. *See* DVI
DVI 3:290, 689
  applications 3:292
  calculation 3:291
  problems with 3:291
Dynamic global vegetation models (DGVMs) 2:169, 481

Early warning indicators 4:79
Early warning systems 3:550
Earth
  art 5:173, 355
  managing 5:6
  origins 5:372
  radiation balance 1:18
Earth Charter 5:216
Earth Day 5:49, 52, 53, 216, 248
Earth First! (EF!) 5:217
Earth observing system 1:61, 2:250
  current requirements and issues 1:63
  in situ systems, networks and programs 1:68
  satellites 1:80
  see also EOS
Earth resources mapping and surface characterization satellites 1:73
Earth Science Enterprise (ESE) 2:250
Earth sciences, international organizations 1:156
Earth Summit 1:539
Earth System 1:538
  and social system 3:5
  forcing and feedback variables 1:67
  human disturbance of 3:1
  key parameters and variables 1:61
  state variables and parameters 1:66

Earth system models 1:99, 109, 682
applications
diagnosis of prevailing conditions and empirical (experiment-based) prediction **1:**99
explanation of causal connections **1:**99
Earth system history **1:**101
general approaches to creating **1:**100
horizontal-dimension models **1:**105
interannual uctuations representation **1:**111
laboratory models and eld experiments **1:**100
management and policy analysis **1:**100
mathematical analysis **1:**102
numerical modeling approaches **1:**103
other celestial bodies as **1:**102
prediction and projection of future conditions **1:**100
recent past as model for the future **1:**102
regional models **1:**109
representations of individual processes **1:**110
testing performance **1:**110
vertical-dimension models **1:**104
zero-dimension models **1:**104
zonal models **1:**106
see also speci c models
Earth system processes **1:**13
Earthquake-generated tsunamis **1:**726
Earthquakes **3:**8, 488, 513
intensity and magnitude **3:**293
losses **4:**383
prediction **1:**537
reservoir-induced **3:**294
triggered by human activities **3:**294
see also Seismic risk
Earth s rotation **1:**225
Earth s surface **1:**310, 503, 538
ow patterns forced by **1:**238
longwave radiation **1:**474
East Timor **3:**549
EBCMs **1:**373, 376
climate modeling hierarchy **1:**376
general applications **1:**379
single latitude zone **1:**378
speci c examples **1:**380
structure **1:**377
ECA (Economic Commission for Africa) **4:**194
Eciatrics **5:**423
ECLAC (Economic Commission for Latin America and the Caribbean) **4:**195
Eco-agriculture **4:**547
Eco-Art **5:**173
Eco-building **4:**548
Eco-development, China **3:**433
Eco-ef ciency **4:**18, 44, 106
economic/organizational perspective **4:**109
engineering perspective **4:**111
historical background **4:**106
increasing **4:**111
organizational perspective **4:**113
stages **4:**111
strategy **4:**108
Eco-engineering **4:**547
rural **4:**547
Eco-farming **4:**547
Eco industrial parks (EIP) **3:**76
Eco-industry **4:**548
Eco-labeling **4:**210, 211
Eco-labels, types **4:**210
Eco-service, urban-rural **4:**549
Eco-settlement **4:**548
Eco-socialism **5:**47, 434, 444
Eco-spirituality **5:**309
Eco-studies, versions **5:**422
Eco-taxes **4:**195
instruments **4:**196
Eco-toxicology **4:**39
Eco-village **4:**547
development **3:**434
Ecocentrism **5:**217, 236, 311
Ecocycle **5:**392
Ecofascism and land ethic **5:**239
Ecofeminism **5:**33, 196, 218, 434, 439, 444
cross-cultural ourishing **5:**223
scholarship **5:**223
Ecogeny 54
Ecological capacity 51, **3:**296
lightening the load **3:**300
overshoot in **3:**298
see also Human carrying capacity
Ecological city **4:**409
Ecological dumping **4:**198
Ecological economic systems **4:**118
narratives **4:**119
Ecological economics **5:**5, 37
biophysical-economic system analysis **5:**44
classical economic origins **5:**39
future **5:**45
global enterprise **5:**46
historical background **5:**38
methods and issues **5:**42
natural science origins **5:**40
rise of modern **5:**41
working assumptions **5:**38
Ecological engineering **5:**47
Ecological footprints **5:**45, 157
analysis **3:**302
urban **3:**141
Ecological measurements, issues still to be addressed **2:**276
Ecological model of health **5:**499
Ecological modernization **5:**431, 443, 449
Ecological resilience **2:**530
Ecology **5:**408, 423
and democracy **5:**161
global **2:**5
global environmental change **2:**1
industrial. *See* Industrial ecology
see also Social ecology
Economic developments **3:**138
Economic growth **4:**183
Economic instruments **4:**7
Economic sustainability **5:**489
Economic systems, nonlinearity **5:**398
Economics **5:**41, 230, 423
and current regulatory system **5:**29
and global environmental change **5:**4, 25
ecological. *See* Ecological economics
global agreements **5:**35
rainforests **3:**270
rise of modern economic culture **5:**40
Economies **5:**445
consumption patterns **3:**249
Ecopolis **4:**547
Ecosystem
and natural selection **2:**293
costing **2:**272
function **2:**282
functional groups **2:**28
large-scale manipulation experiments and observations **2:**6
redundant species **2:**28
services **2:**272
stability **2:**281
stable states **2:**277
structure **2:**282

Ecosystem approach 3:221, 5:225
Ecosystem functioning 2:24
  relative importance of species in 2:24
Ecosystem health 4:199, 5:47, 493
Ecosystem services 2:20, 5:226
  and global change 2:274
  assessing 2:274
  classification 2:273
  concept 2:272
  delivery 2:276
  discounting 2:279
  economic valuation 2:277
  examples 2:273
  key social issues to be addressed 2:277
  private vs. public ownership 2:279
  understanding 2:274
Ecosystems 2:22, 292
  common structure 2:28
  effects of global change 2:122
  functional groups 2:27
  fundamental structure 2:24
  human impact 2:116
  indirect interactions 2:294
  integrity 4:199
  monitoring 4:116
  species redundancy 2:27
  valued component 4:493
Ecotheology 5:492
Ecotones 2:283
  climate change 2:285
  concept 2:284
  hierarchy 2:286
Ecotourism
  case study 3:600
  statistics 3:598
Ecotoxicology 3:302
Ecumenics 5:423
Edge effects 2:326
Education
  and population growth 5:33
  and Training Programme 1:758
  see also Environmental education
EEA (European Environment Agency) 4:200
Eemian 1:270, 353
Eemian interglacial, climate modeling 1:297
Egalitarianism 5:385, 389
Egocentrism 5:311
Ehrlich, Paul 2:288
Eichhornia crassipes 2:81
EIS (Environmental Impact Statement) 4:201
Ekistics 5:423
El Chichon volcanic eruption 1:736
El Nino 1:28, 230, 279, 353, 457, 476, 487, 645
  causes and consequences 1:353
  ecosystems impacts 1:362
  historical background 1:354
  oscillation 1:292
  physics 1:358
  socio-economic impacts 1:365
  temperature and precipitation anomalies associated with 1:361
  weather impacts 1:360
  see also El Nino/Southern Oscillation (ENSO)
El Nino–La Nina events 1:309
El Nino/Southern Oscillation (ENSO) 1:22, 64, 69, 120, 212, 245, 343, 353, 370, 457, 487, 2:63, 70, 138, 163, 549, 3:723, 4:230
  cycle 1:358
  development of observing and forecasting systems 1:363
El Viejo. *See* La Nina
Electric currents 1:502
Electric fields 1:502
Electricity, atmospheric 1:218, 502
Electromagnetic energy 1:371
Electromagnetic radiation 1:371
Electroplating 3:59
Elodea canadensis 3:415
Elton, C S 2:289
Emergencies, human impact 3:549
Emergent groups 2:482
Emerging environmental issues 4:72
  action phase 4:78
  characteristics 4:73
  concern phase/issues of concern 4:78
  definition 4:72
  early warning indicators 4:79
  future 4:72
  horizon phase/horizon issues 4:78
  management 4:77
  new perspectives 4:74
  overview 4:75
  role of science 4:75
  taxonomies 4:73
  unforeseen issues 4:74
  unlikely events 4:74
Emergy 5:224
  concept 3:4, 303
  use of term 3:303
Emission limits 3:63
Emission reduction credit (ERC) 4:453
Emission standards 3:63
Emission trading 4:165, 201, 357, 440, 451
  Canada 4:454
  Europe 4:454
  merits and problems 4:440
  nitrogen 4:453
  questions and answers 4:453
  sulfur 4:453
  United States 4:453
Emissions
  aircraft 3:178
  global trends 3:35
  road transport 4:429
  transport 3:646
  see also specific types
Emissivity 1:371
Encephalitis 3:692
Encyclopedias 5:228
Endangered species 4:201
Endotherms 2:136
Energy
  environmental impacts 3:736
  terrestrial stores and fluxes 2:190
  transport 3:645
  see also Sustainable energy
Energy balance 2:192
  and climate 1:371
Energy balance climate models. *See* EBCMs
Energy balance models 1:106
  two-dimensional 1:106
Energy budget
  atmosphere 1:240
  latitudinal distribution 1:373
Energy consumption 5:403
  world trends 3:734
Energy efficiency
  improvement 4:98
  SIDs 4:414
Energy end-use efficiencies 4:91
Energy fluxes 3:4
Energy innovation value chain 4:92
  research and development (R&D) 4:93

Energy intensity of industrialized countries 4:295
Energy ladder 3:217
Energy paths, soft 3:583, 5:487
Energy policies and sustainable futures 4:356
Energy production 3:709
Energy related carbon dioxide emissions 3:47
Energy resources, Small Island Developing States (SIDS) 4:412
Energy sources 1:234, 2:26
  non-renewable 3:355
  see also Biomass
Energy types 3:734
Energy use
  future 3:736
  present 3:736
Engineering feasibility study 4:204
Engineering resilience 2:530
Enlightenment Project 5:2, 229, 336
ENSO. *See* El Nino/Southern Oscillation (ENSO)
Entrainment 1:325
Entropy-based evaluation (MASTER) 3:63
Environment
  and development 5:58
  constituents 2:290
  definition 2:290
  human interference 2:290
Environmental agenda, social dimension 4:16
Environmental assessment 4:194
  best practice or guidelines standards 4:286
  development projects 4:203
  Dobr s 4:546
  national budgets 4:405
Environmental Cadmium in the Food Chain: Sources, Pathways and Risks 3:307
Environmental challenge and solutions 5:177
Environmental change
  and conflict 5:271
  and evolutionary processes 2:292
  interactions between 2:104
  seismic events 3:296
Environmental crisis 5:165
Environmental damage 5:34
Environmental Data Report 4:204
Environmental Defense Fund (EDF) 5:230, 248
Environmental disasters. *See* Disasters
Environmental education 4:206
  communication and media 4:209
  definitions 4:206
  ethics 4:207
  guidelines 4:207
  objectives 4:207
  schools 4:208
  tertiary 4:209
Environmental ethics 5:231
Environmental governance 5:368
Environmental groups 5:243
Environmental health
  and urban poverty 3:661
  problems 3:675
Environmental historiograph 5:70
Environmental history and global environmental change 5:62
Environmental impact assessment 2:407, 3:3, 4:194
Environmental indicators 4:247
Environmental justice 5:232
  rights-based approach 5:119
  utility-based approach 5:119
Environmental law 4:7, 11
  international 5:324
Environmental management
  command and control management 4:16
  trends over 40 years 4:15
Environmental movement 5:243
  histories 5:3
  origins and concerns 5:430
  types 5:432
Environmental organizations 5:56
Environmental philosophy 5:253
Environmental policy 4:11
  dilemmas 4:367
  industrialized countries 4:5
  international 4:323
  world oceans 4:279
Environmental politics 5:49
  contemporary 5:54
  recent highlights 5:52
Environmental problems
  critical 4:382
  man-made 4:38
  national level 4:2
Environmental programs, UN agencies 4:11
Environmental protection, investment 4:6
Environmental psychology/perception 5:257
Environmental refugees 4:214, 215
Environmental regulation 5:29
Environmental research development, government expenditure 4:6
Environmental responses
  international 4:9
  overview 4:1
Environmental risk assessment, best practice or guidelines standards 4:286
Environmental security 5:269
  critical perspectives 5:273
  research 5:270
Environmental sociology 5:278
Environmental standards
  bilateral 4:285
  regional 4:285
Environmental sustainability (ES) 5:489
Environmental systems
  feedbacks 3:9
  maintaining 3:82
Environmental valuations 5:42
Environmental values 5:49
Environmentalism, defining 5:369
Environmentally sensitive area (ESA) 4:361
ENVISAT 1:80
EOS 1:23, 382
EOS–Aqua 1:78
EOS–Aqua/AIRS 1:79
EOS–Aqua/AMSR-E–NASDA 1:79
EOS–Aura 1:79
EOS–ICESat 1:80
EOS satellite series 1:76
EOS–Terra 1:76
EOS–Terra/ASTER 1:78
EOS–Terra/CERES 1:77
EOS–Terra/MISSR 1:78
EOS–Terra/MODIS 1:77
EOS–Terra/MOPITT 1:78
Eotros effect 1:329
Epidemiology and air quality standards 4:485
Equatorial waves 1:359
Equilibrium response 1:382
Equity 5:279
  and innovative technology 5:281
  multiple dimensions 5:280
  multiple questions 5:280
  multiple tasks 5:280
Eradication, species 2:17
Erosion, soil. *See* Soil erosion
ERS (Earth Resources Satellite) Series (ESA) 1:73

ESCAP (Economic and Social Commission for Asia and the Pacific) 4:218
Estuaries, impacts of aggregate extraction 3:342
Ethics
  environmental 5:231
  see also Land ethic
Euphorbia candelabra 1:334
Euphydryas editha 2:162
Europe
  emission trading 4:454
  nitrate protection zones (NPZs) 3:493
  tourism 3:626
European Agricultural Guidance and Guarantee Fund (EADDF) 3:248
European Center for Medium-Range Weather Forecasts 1:454
European Network for Research in Global Change (ENRICH) 2:290
European Union (EU) 3:247, 549, 4:218
Eutrophic systems 2:458, 5:426
Eutrophication 2:128, 458, 3:720, 724
  marine and coastal waters 3:503
Evaporation 1:452, 2:344, 346, 347
  rainforests 3:265, 267
  yearly average 1:456
Evapotranspiration 1:452, 2:380, 461
Evolution 2:560
Evolutionary processes
  and environmental changes 2:292
  ecosystems 2:292
Exclusion, philosophical polarities 5:103
Exosphere 1:244
Exotic invasions 2:489
Externalities 5:26
Extinctions 2:297
  birds 2:306
  contemporary and future 2:301
  estimates 2:303
  hotspots 2:305
  metapopulations 2:416
Extraction of minerals 3:295
Extraction processes 3:711
Extra-tropical storms 1:236
Exxon Valdez oil spill 5:283, 301
Eye, exposure to UVR 3:136

Fagus gradifolia 2:86
Fair weather electrification 1:220
Fairness 5:279
Family planning 5:34
FAO (Food and Agriculture Organization) 2:324, 3:238, 4:41, 219
Farming, organic 3:532
Fast feedback processes 1:290
Federal Emergency Management Agency (FEMA) 4:377
Feedbacks 1:382
  chemistry–climate interactions with climate 1:384
  climatic response 1:27
  environment system 3:9
  mechanisms 1:375
Feminism 5:439
  see also Ecofeminism
Feng–Shui theory 3:431
Fertility levels and population growth 3:20
Fertilizers 3:9
  inorganic 3:539
  inorganic phosphatic 3:539
  policy 4:360
FGD waste 4:111
Fingerprinting 1:388
  applications 1:279
Fire
  grasslands 2:572
  mosaics 5:319
  species tolerance 2:487
  suppression 2:566
  see also Forest fires; Oil fires
First International Polar Year 1:156
First Nations peoples 5:466
Fish
  direct use 2:330
  freshwater 2:327
  habitat 2:328
  indirect abuse 2:330
  K-selected species 2:330, 331
  r-selected species 2:330
Fish stocks, analysis 3:238
Fisheries 3:9, 136
  Baltic Sea 4:518
  ecosystems perspective 4:348
  effects of climate change 3:309
  El Nino effects 1:363
  governance conventions 2:330
  habitat degradation 3:316
  management 4:346
  ownership regimes 4:347
  Pacific coast 3:314
  policies 4:343
  pollution 3:316
  reforms of the 1990's 4:350
  responsible 4:343
  sustainable 4:343
  transferable quotas 4:455
  types 2:330
  see also Aquaculture; Freshwater fisheries; Minamata; Transferable fisheries quotas
Fixed nitrogen 2:308
Flip-flops 5:422
Floodplains 2:378
  burning patterns 5:319
  flows 5:425
  forests 2:287
Floods and flooding 1:726, 2:347, 3:245, 486, 707, 730
Flow dynamics, phases 5:424
Flue-gas desulfurization (FGD) technology 4:107
Fluid injection, earth tremors induced by 3:295
Fluid removal, earthquakes induced by 3:295
Fluid withdrawal, subsidence due to 3:339
FLUXNET 2:45
Food, and ecofeminism 5:219
Food and Agriculture Organization (FAO) 2:324, 3:238, 4:41, 219
Food chains 2:405
  cadmium in 3:303
Food consumption
  and climate change 3:323
  and greenhouse gas emissions 3:323
Food production 3:134, 712, 5:88
  and malnutrition 3:441
  greenhouse 3:352
Food webs 2:308
  age structure effects 2:315
  as open systems 2:314
  connectivity 2:310
  current topics/trends 2:313
  energetic 2:310
  functional 2:312
  interactive 2:312
  omnivory 2:313
  roles of nutrients and stoichiometry 2:315
  structure 2:313
  types 2:309

Food yields 3:136
Forest decline 5:508.17
Forest fires 3:210, 653
Forest Gap Models 2:316
Forest logging. *See* Logging systems
Forest resources and conflict 3:148
Forest soils 3:546
Forest stand 2:323
Forest webs
  even aged 2:324
  uneven aged 2:324
Forests 2:324
  decline 2:439
  definition 2:324
  ecosystems 2:536
  gap models 2:316
  nitrogen deposition 2:435
  sustainable policies 4:351
  temperate coniferous 2:560
  temperate deciduous 2:565
  urban 3:356
  see also Afforestation; Agroforestry; Deforestation; Global Forest Watch; Temperate coniferous forests; Temperate deciduous forests; Tropical forests
Fossil fuel burning emissions 3:473
Fossil fuel combustion 3:9
Fossil fuel use 3:8
Fossil fuels 4:183, 356
  greenhouse gases 3:649
  pollutants from 3:302
Fossil groundwater 1:341
Fossil Record 2:297
Fossils 2:547
Founex 1971 4:399
Fourier's Law for heat conduction 3:335
Fragmentation 2:324
Framework Convention on Climate Change (FCCC) 4:292, 5:116
Framework treaty protocol form 5:327
Francis of Assisi, Saint 5:284
Franklin, Benjamin 1:390
Free air carbon dioxide enrichment (FACE) 2:5, 204
Free market liberalism 5:437
Freight. *See* Transport
Freshwater
  biodiversity 2:146
  circulation 5:201
  global warming 3:310
Freshwater ecosystems 2:78, 458, 533
  impacts of global change 2:122
Freshwater environment, petroleum hydrocarbons 3:521
Freshwater fisheries 2:327
  current status 2:331
  degradation 2:331
Freshwater lakes 2:377
Friction in atmospheric motions 1:224
Friends of the Earth (FOE) 5:248, 285
Friends of the Earth International (FOEI) 5:57
Fronts 1:235, 391
Fuel consumption, aircraft 3:179
Fuel use for transport 4:434
Fuels
  biomass power development 3:216
  reformulated 3:679
  urban, biomass 3:217
Fugacity 3:335
Functional Groups 2:483
Fundamental niche 2:429
Futures research 5:285
Gaia 5:211, 217, 287, 448, 493
  defining characteristics 2:332
Gaia Hypothesis 1:510, 2:247, 446
Gaiacentrism 5:217
GAIM (Global Analysis, Interpretation and Modeling Program) 1:392
GAIM Model Intercomparisons 2:50
Galactic cosmic ray (GCR) 1:678
Galileo 5:103
Gandhi, M K 5:290
Gap models 2:316
  tests and applications 2:320
Gaquid phase 5:424
GARP (Global Atmospheric Research Program) 1:157, 392, 756, 757
Gas exploration, development and production, Arctic 3:518
Gas flaring, carbon dioxide emissions 3:473
Gas-phase stratospheric chemistry 1:675
Gases
  lifetime 1:501
  low concentration methane bearing 3:435
GATT (General Agreement on Tariffs and Trade) 3:633, 4:221
GAW (Global Atmosphere Watch) 1:393
GCDIS (Global Change Data and Information System) 1:393
GCOS (Global Climate Observing System) 1:394
GCTE Elevated CO_2 Network 2:46
GCTE soil erosion network 2:44
GCTE Terrestrial Carbon Cycle Synthesis 2:49
GCTE-NEWS synthesis of global warming 2:579
G'DAY 2:332
  daily version 2:333
  mathematical analysis 2:335
GEH tree 2:342
General Agreement on Tariffs and Trade. *See* GATT
General Circulation Models (GCMs) 1:108, 114, 288, 296, 309, 346, 374, 394, 4:82, 253
Generic Decomposition and Yield. *See* G'DAY
Genetic algorithms 5:398
Genetic resources, access 4:171
Genetically modified organisms (GMOs) 4:45
Genotype, definition 2:466
Genyornis 2:59
GEO (Global Environment Outlook) 4:221
Geochemical cycles
  gaseous 2:159
  sedimentary 2:159
Geodesy 1:538
Geoengineering
  applications 3:337
  climate stabilization 3:336
  use of term 3:336
Geographical Information Systems (GIS) 4:222
  component parts 4:223, 225
  composite maps 4:224
  definitions 4:224
  locational problems 4:222
  questions relating to 4:226
Geography
  applications 3:337
  definition 3:337
  investigations 3:337
Geological cycling 1:397
Geological hazards 3:482
Geomagnetism 1:538
Geomorphological hazards 3:482
Geomorphology 1:402
  change for urbanization and industry 3:338
  sports 3:345
Geophysical data 1:535
Geophysical Fluid Dynamics Laboratory (GFDL) 1:512

Geophysical monitoring systems. *See* Monitoring systems
GEOSECS programme **3:**183
Geostationary environmental satellites (GOES) **1:**72
Geothermal heat **1:**402
Geotropism **2:**558
Geranyl pyrophosphate (GPP) **2:**607
GESAMP (Joint Group of Experts on the Scienti c Aspects of Marine Environmental Protection) **4:**226
GEWEX **1:**310, 402, 463
  objectives **1:**403
  research foci **1:**403
Glacial aridity **1:**340
Glacial cycling simulation **1:**111
Glacial-eroded sediment **1:**429
Glacial/Holocene boundary **1:**275
Glacial pluvial **1:**340
Glacial terminations **1:**272
Glaciers **1:**19, 404, 500
  annual average values of net balances **1:**406
  changes in area and volume **1:**407
  changes in size **1:**407
  dimensions **1:**405
  dynamics problem **1:**407
  environmental effects of changes **1:**408
  uctuations and climate variations **1:**404
  Greenland **1:**447
  ice cores **1:**407
  mass balances **1:**405
  total areas **1:**405
  winter and summer balances **1:**406
Gliricidia sepium **2:**443
Global 500 Roll of Honour **4:**280
Global agreements, economics **5:**35
Global air quality **1:**177
Global Analysis, Interpretation and Modeling (GAIM) **2:**41
Global Atmosphere Watch (GAW) **1:**177, 758
Global atmospheric electric circuit **1:**502
Global average temperature **1:**371
Global carbon balance **2:**197
  future biosphere **2:**223
Global carbon cycle **1:**368
Global change, de nition **2:**291
Global Change and Terrestrial Ecology (GCTE) **2:**38
Global Change and Terrestrial Ecosystems (GCTE) program **2:**481
Global Change Master Directory (GCMD) **2:**253
Global change system for analysis, research and training (START) **2:**41
Global chemical models **1:**109
Global climate change, lightning **1:**221
Global climate models **1:**109
Global Climate Observing System (GCOS) **1:**753, **4:**42
Global cooling **1:**337
Global Data Processing System (GDPS) **1:**68
Global Eco-labeling Network (GEN) **4:**211
Global energy balance **1:**18
Global environment facility (GEF) **4:**481
Global environment monitoring system (GEMS) **4:**39
Global environment outlooks (GEOs) **4:**39
Global environmental change (GEC)
  airborne sensors **2:**524
  and environmental history **5:**62
  and social science **5:**109
  anthropology and **5:**163
  assessing effects of **2:**124
  biological and ecological dimensions **2:**1
  causes and consequences **3:**7
  citizens responses **4:**33
  current challenges and priorities for science **4:**44
  driving forces **5:**166
  European perspectives **4:**25
  hazards **5:**297
  human dimensions of **5:**1
  human drivers of **3:**5
  impact on animals **2:**56
  impact on plants **2:**94
  impact on terrestrial and freshwater ecosystems **2:**122
  integration **2:**132
  politics **5:**124
  public-driven response **4:**21
  regional assessments **2:**525
  regional responses to **4:**514
  risk management **4:**375
  satellite sensors **2:**524
  scienti c responses **4:**36
  social science and **5:**3
Global Environmental Facility (GEF) **4:**43
Global Environmental Monitoring System (GEMS) **1:**70
Global Forest Watch **4:**231
Global governance **5:**292
Global land cover **3:**13
Global mean temperature **1:**281
Global models, three-dimensional **1:**107
Global multi-issues **4:**317
Global net primary productivity (NPP) **2:**50
Global observing programs **4:**42
Global Observing Systems **1:**63, 68
Global Ocean Data Assimilation Experiment (GODAE) **1:**537
Global Ocean Ecosystem Dynamics (GLOBEC) **2:**39
Global Ocean Observing System (GOOS) **1:**537, **2:**336, **4:**42
Global plate tectonics **1:**410
Global population trends **3:**16
Global Precipitation Climatology Center (GPCC) **1:**454
Global reporting initiative (GRI) **4:**236
Global Runoff Data Center (GRDC) **1:**454
Global technology cooperation
  Mexico, case studies **4:**385
  Thailand **4:**386
Global Telecommunications System (GTS) **1:**68
Global Terrestrial Observing System (GTOS) **1:**538, **2:**337, **4:**42
Global Urban Observatory (GUO) **4:**462
Global warming **1:**27, 367, 712, **2:**64, 85, 88, 331, 373, 579, 580, **3:**315, **5:**498
  and rice **3:**127
  and skiing **3:**346
  and UHI **3:**665
  direct effects **2:**575
  ecosystem response evaluation **2:**575
  effects on freshwater ecosystems **2:**129
  effects on terrestrial ecosystems **2:**129
  evidence of response **2:**579
  freshwater **3:**310
  indirect effects **2:**576
  initial conditions **2:**578
  interactions with other elements **2:**578
  scenarios for 21st century **2:**575
  tourism contribution **3:**614
  variable time scale and direction of response **2:**577
Global warming potentials (GWPs) **1:**267, 411, 447
Global water cycle **2:**132
Global Weather Experiment **1:**392
Globalization **5:**59
  and development **5:**157
  capitalist roots **5:**76
  caste system **5:**94
  European origins **5:**77
  historical perspective **5:**73
GOALS **1:**411
Gold mining, mercury in **2:**406
Goldemberg, Jose **4:**233
Goldman prize **4:**280

Golf, impacts 3:345
Goodman, Gordon 4:234
Gopherus agassizii 2:137
Governance and international management 5:292
GPI (Genuine Progress Indicator) 5:331
Grasslands 2:68, 569
  clearing 3:228
  ecosystems 2:535
  nitrogen applications 3:502
  see also Temperate grasslands
Gravitational force 1:329
Gravity 1:224
Great Depression 5:41
Great Lakes 4:535
  case study 3:721
  climate change 4:538
  contaminants 4:536
  diverting and exporting fresh water 4:537
  responses to global change 4:536
Great Lakes Water Quality Agreement (GLWQA) 4:552
Great Man-made River Project, Libya 3:362
Green anarchism 5:439
Green–Blue technology 5:426
Green ideology 5:440
Green investment 4:235
Green Parties 5:56
Green Revolution 3:347, 5:88, 93
Green screens 42, 97
Greenhouse
  food production 3:352
  heating 3:355
  open-roof 3:354
Greenhouse effect 1:19, 26, 27, 223, 248, 249, 413, 2:161
Greenhouse gases 1:27, 140, 190, 204, 273, 384, 411, 468, 675, 2:424, 457, 608, 3:178, 436, 4:165
  and food consumption 3:323
  anthropogenic 1:279, 281
  emissions 4:164, 254, 410, 486
  fossil fuels 3:649
  highest emitters per km^2 land area 3:475
  highest per capita emitters 3:473
  highest per unit GDP emitters 3:474
  lowest emitters per km^2 land area 3:475
  lowest per capita emitters 3:473
  lowest per unit GDP emitters 3:474
  national responsibilities for emissions 3:472
  reduction projects 4:292
  total emissions per capita 3:475
  tradable permits for 4:438
Greenhouse trapping (GHT) 1:291
Greenhouse warming 2:160
Greening of cities 3:356
Greenland 1:19, 21, 413, 647
  carbon dioxide 1:416
  climate 1:414
  climate research 1:415, 416
  geography 1:414
  glaciers 1:447
  Ice Core Project (GRIP) 1:417
  ice cores 1:272, 273, 418, 447
  ice loss 1:467
  ice sheet 1:419
    elevation change 1:421
  map 1:414
Greenpeace 5:57
Grey–Brown technology 5:426
GRID (Global Resource Information Database) 4:237
Gross domestic product (GDP) 4:295, 385, 5:43
Gross national product (GNP) 5:150, 152
Gross primary productivity (GPP) 2:191, 515
Ground temperature 1:21
Groundwater 1:451, 755
  assessment 23;10
  deserts 1:341
Groundwater contamination 3:495
Groundwater exploitation and subsidence 3:340
Groundwater quality, nitrate in 3:503
Groundwater systems 3:366
Growing season net flux (GSNF) 2:205
Growlers 1:467
Growth 2:161
Growth cycle 4:111
GTOS (Global Terrestrial Observing System) 1:70
Guanghan, China 4:549
Gulf of Alaska 3:315
Gulf War 3:146, 523, 548
Gymnosperms 2:135

Habitat destruction, tourism contribution 3:615
Habitat fragmentation, Amazon Basin 3:255
Habitats 2:339
  lakes 2:378
  rivers 2:378
  urban ecosystems 3:656
Hadley circulation 1:227, 12:1198, 1:427, 476, 715
Hadley, George 1:427
Haiti, environmental refugees 4:215
Halocarbons 1:178, 427
  growth in concentrations 1:144
Halogen Occultation Experiment (HALOE) 1:732
Halon-1301 1:146
Halons 1:153, 384, 680
Halophobes 2:339
Halophytes 2:339
Hantavirus 3:693
Harbors 3:572
Hare, Kenneth 1:428
Harmful algal blooms (HABs) 3:380, 548
Hazardous wastes 3:710
  transboundary movements 3:631
Hazards, natural. *See* Natural hazards
Health 3:732, 5:497
  and biosphere 5:500
  and environment 5:492
  and environmental change 3:130
  as sustainable state 3:142
  awareness 4:48
  ecological model of 5:499
  medical model 5:495
  nitrogen impacts 3:503
  public 3:547, 5:494
  see also Ecosystem health
Health-based standards 4:487
Health benefits of air quality standard 4:491
Health concerns, water use 3:90
Health effects, ozone 3:652
Health hazards
  due to global climate change 3:132
  nuclear waste 3:507
Health issues
  lifestyles 4:61
  policy responses 4:47
  population dynamics 4:55
  water supplies 4:51
  see also WHO
Health of the Planet' 4:23
Health outcomes and value systems 4:54
Health protection 4:50

Heat
  geothermal 1:402
  latent 1:500
Heat storage, role of 1:21
Heatwaves 3:140
Heavy metals
  methylation 3:110
  persistent organic pollutants 4:286
  soil contaminated with 3:105
Hegel, Georg Wilhelm Friedrich 5:105
Heinrich (H-) events 1:429
  mechanisms 1:430
  outstanding problems 1:430
Helsinki Commission (HELCOM) 4:518
Herbaceous-dominated vegetation 2:486
Heroics in mythology 5:373
Heterosphere 1:244
Heterotrophic systems 2:458
Heterotrophy 2:26, 143
HFC-134a 4:388
Hierarchical structure, conceptual model 4:130
Hierarchy
  community based 4:118
  theory 4:117
High Resolution Doppler Imager (HRDI) 1:732
High-nitrate low chlorophyll (HNLC) 3:183
Highway construction 3:711
Hilo, Hawaii 1:726
Hinduism 5:303
  eco-spirituality 5:309
  pollution and its prevention 5:307
  treatment of animal life 5:305
Holdgate, Martin 4:239
Holdridge life zone classification 2:167, 201, 339, 487
Holism
  and human ecology 3:2
  understanding 4:408
Holling, C S 2:340
Holocene 1:249, 275, 340, 431, 2:86, 286, 5:57
  climate changes and society in Europe 3:372
Holon 2:283, 5:423
Holonarchy 5:423
Homeostasis 5:493
Homocentrism 5:311
Homosphere 1:243
Household energy problem 3:218
Housing and basic services 3:673
Hsinfengkiang, China 3:295
Human adaptability 5:164, 300
Human carrying capacity 3:296
Human conflict 3:548
Human dimensions of global environmental change 5:1
Human disturbance of Earth System, dynamics and complexities 3:1
Human disturbance regimes 2:88
Human drivers of global environmental change 3:5
Human ecology
  and holism 3:2
  China 3:430
Human ecosystem model 5:499
Human–environment relations 5:67
Human health. *See* Health
human impacts, cumulative effects 3:4
Human nature relationships (HNR) 5:12
Human needs and wants 3:2
Human obligations 5:207
Human rights 5:207
Human security 5:275, 276
Human shaping of the environment 5:357
Human sustainability 5:489
Human systems
  maintaining 3:81
  redesigning 3:82
Human values and science 5:117
Humanitarian aid 3:549
Humanitarian assistance, institutional trends 3:550
Humanitarian issues 3:550
Humanity and Earth 5:177
Humidity 1:432
Hurricanes 1:439
  categorization 1:434
  overview 1:433
Hutchinson, G Evelyn 2:341
Hydrochloric acid (HCl) 1:154, 681
Hydrochlorofluorocarbons (HCFCs) 1:145, 153, 178, 268, 384, 411, 427, 4:313, 428
Hydroegoism 5:206
Hydrofluorocarbons (HFCs) 1:145, 178, 268, 384, 411, 427, 447
Hydrogen isotope ratios 2:546, 547, 548, 549
Hydrogen oxide (HO_x) family 1:677
Hydrogen peroxide (H_2O_2) 1:447, 2:457
Hydrogen sulfide (H_2S) in photosynthesis 2:423
Hydrologic cycle 1:19, 450, 2:343
  and climate change 1:459
  biospheric controls 2:37
  conceptual diagram 1:452
  data 1:453
  fluxes 1:451
  human impacts 1:458
  mathematical models 1:452
  observed climatology 1:454
  observed trends 1:462
  quality measurements of individual components 1:454
  reservoirs within 1:450
  towards improved understanding 1:462
  20th century observed changes 1:462
  variations 1:457
Hydrological alterations 2:132
Hydrological hazards 3:481
Hydrological stores 2:344
Hydrology 1:464, 468, 2:343, 345
  definitions 1:464
  extremes 2:347
  global 2:344
  observing systems 1:536
  rainforests 3:268
Hydrology and Water Resources Programme 1:758
Hydrometeorology 1:464
Hydrosolidarity 5:206
Hydroxyl radical (OH) 1:517, 677, 2:457
Hyperspatial imaging systems 2:527
Hyperspectral imaging systems 2:526

IAHS (International Association of Hydrological Sciences) 1:464
IAI (Inter-American Institute for Global Change Research) 4:241
IAMAS (International Association of Meteorology and Atmospheric Sciences) 1:466
IAMs
  general structure 4:256
  in science 4:259
  mathematical structure 4:256
  solving procedures 4:256
  top-down versus bottom-up debate 4:257
  uses 4:257
  when and where flexibility debate 4:258
IAPSO 1:467
Ice
  areal extent 1:284
  role of 1:21
Ice-core volcanic index (IVI) 3:690

Ice caps 1:19
Ice cores 1:518
  atmospheric CO_2 1:256
  glaciers 1:407
  Greenland 1:273, 274, 447
Ice dynamics, simulations 1:290
Ice sheets 1:21
Ice storms 3:486
Iceberg-rafted debris (IRD) 1:430
Icebergs 1:409, 467
ICREAs (International Commodity-related Environmental Agreements) 4:242
ICSU (International Council for Science) 1:158, 4:39, 242
ICSU–(World Glacier Monitoring Service (WGMS) 1:70
IDNDR (International Decade for Natural Disaster Reduction) 4:242
IDWSSD (International Drinking Water Supply and Sanitation Decade) 4:244
IFIAS (International Federation of Institutes for Advanced Study) 4:244
IGAC (International Global Atmospheric Chemistry) 1:467
IGBP 1:392, 2:38, 350, 351, 469, 4:41
  core projects 1:310, 2:37, 48
  framework activities 2:40
  implementation strategies 2:41
  science trends 2:53
  Terrestrial Carbon Working Group 2:51
IGBP–LUCC 5:386
Igneous rocks 1:397
IGOS 1:71
IGY (International Geophysical Year) 1:156, 189, 355, 468, 538, 4:40
IHDP (International Human Dimensions Program on Global Environmental Change) 4:42, 245
IHP (International Hydrological Program) 1:468
IIASA (International Institute for Applied Systems Analysis 4:44, 246, 452
Imaging technologies 2:528
Imbrie, John 1:469
Immanence 5:98
Immigration. *See* Migration
Immune system, effect of UVR 3:136
IMO (International Maritime Organization) 4:246
Imperata cylindrica 3:273
Improved Stratospheric and Mesospheric Sounder (ISAMS) 1:732
Incentive-based regulation 5:30
Incineration, urban wastes 3:687
Inclusion, philosophical polarities 5:103
Index of Sustainable Economic Welfare (ISEW) 5:44, 331
Index system, data banks 2:254
Indicators
  calculating 4:135
  categories 4:134
  development 4:132, 134
  environmental quality 4:247
  generation 4:133
  selection criteria 4:134
  sustainable development 4:247
Indigenous empowerment 5:323
Indigenous knowledge 5:314
  epistemology 5:317
Indigenous peoples 5:314
  and environmental change 5:166
  and phase shifts 5:427
Individual behavior 5:405
Indonesia
  aquaculture 3:197
  transmigration 3:272
  wild fires (1997) 3:491
Indus Basin, water management 3:377
Indus River, freshwater diversion 2:400
Industrial accidents 5:273
Industrial ecology 3:4, 4:112, 248, 5:47
Industrial emissions 3:679
Industrial metabolism 3:4, 5, 6, 73
Industrial revolution, power supplies 5:90
Industrialization, 19th century 5:363
Industrialized countries, environmental policy 4:5
Industry leadership 4:65
Infant mortality 5:33
Infectious diseases 2:163, 357, 3:138
  and malnutrition 3:442
  person-to-person 3:134
  vector-borne 3:133
Information age and dematerialization 4:100
Information and communication technologies (ICTs) 3:6, 4:187, 5:505
Information exchange 4:171
Information management 2:392
Infrared radiation 1:371, 470
  concepts 1:470
  physical effects 1:474
  process 1:473
  spectra 1:472
Inland waters
  ecosystems 2:380
  science 2:376
Innovation 4:31, 5:294
Insect pests 3:381
  active season 3:382
  adaptation 3:381
  extension of range 3:383
  long-term datasets 3:384
  response options 3:386
  winter mortality 3:383
Insects, phytophagous 2:471
Institutional innovations 5:294
Institutionalization, global level 5:293
Institutions, environmental responsibility 4:7
Integrated assessment (IA) 4:249, 250
  goals 4:250
Integrated assessment models. *See* IAMs
Integrated Coastal Area Management 4:346
Integrated coastal zone management (ICZM) 4:167
Integrated environmental and economic accounting (IEEA) 4:39
Integrated Global Earth Monitoring Systems (IGEMS) 1:538
Integrated models 5:395, 396
Integrated pest management (IPM) 4:261
  and global change 4:263
Intellectual property rights 5:322
Inter American Institute for Global Change (IAI) 2:363
Inter-basin transfer (IBTs)
  ecological impacts 3:390
  global examples 3:388
  history 3:387
  North America 3:389
  Siberia-Central Asia 3:389
  UK 3:387
  water supplies 3:387
Intergenerational equity 5:349
Intergovernmental Oceanographic Commission 1:394, 753, 4:465
Intergovernmental organizations 1:157
Intergovernmental Panel on Climate Change (IPCC) 1:28, 157, 278, 392, 409, 411, 3:7, 4:76, 163, 265
Internalization of damage costs 4:278
International agreements 4:3, 5:35
International Biological Program (IBP) 2:168
International Biosphere Program (IBP) 2:5, 364
International climate policy 4:252
International Council for Science (ICSU) 1:394, 538, 752, 4:156
International Environmental Education Program (IEEP) 4:206

International environmental institutions 4:10
International environmental law 5:324
International environmental policy 4:323
International Environmental Prizes 4:279
International environmental responses 4:9
International environmental standards 4:281
  non-binding negotiations 4:284
  voluntary 4:286
International Geophysical Year (IGY) 1:156, 189, 355, 468, 538, 4:40
International Geosphere Biosphere Program–Land-Use and Land-Cover Change. *See* IGBP–LUCC
International Geosphere–Biosphere Program. *See* IGBP
International Global Atmospheric Chemistry Project (IGAC) 2:39
International Human Dimensions Programme (IHDP) 2:350
International Hydrological Decade (IHD) 1:469, 4:465
International Institute for Environment and Development (IIED) 5:248
International LTER (ILTER) network 2:393
International management 5:292
  modes 5:293
International Maritime Organization (IMO) 4:308
International Meteorological Organization (IMO) 1:68
International Monetary Fund (IMF) 5:69
International organizations
  accomplishments 1:159
  earth sciences 1:156
International Rice Research Institute (IRRI) 3:391
International Society for Ecological Economics (ISEE) 5:37
International transfers of emission 4:441
International Union for the Conservation of Nature (IUCN) 5:57
International Union of Geodesy and Geophysics (IUGG) 1:538
Interstadials 1:332
Intertidal zones 2:365
  classifications 2:365
  littorinid 2:365
  lower 2:366
  middle 2:365
  upper 2:365
Intertropical Convergence Zone (ITCZ) 1:228, 476
Introduction to Hydrometeorology 4:154
Invasability 2:13
Invasions 2:87, 132, 567, 4:74
  consequences of 2:14
  control methods 2:16
  definitions 2:11
  interactions with other agents of global change 2:14
  lag phases 2:12
  meltdown 2:14
  process 2:11
  rates of conversion to pest status 2:13
  rates of incursion 2:12
  rates of spread 2:12
  success factors 2:13
Invasive species 3:137
Inversion conditions 1:243
Invertebrates 2:162
Investment
  environmental protection 4:6
  sustainable energy 4:90
  see also Socially responsible investment (SRI)
IOC (Intergovernmental Oceanographic Commission) 4:287
IOC–Global Ocean Sea Level Observing System (GLOSS) 1:70
IOC/WMO GOOS 1:69
Ionizing radiation 3:507
Ionosphere 1:244, 476, 502, 516
IPCC (International Panel on Climate Change) 1:477, 3:7, 42, 4:252, 481, 5:395
IPCS (International Program on Chemical Safety) 4:284, 288
IRI (International Research Institute for Climate Prediction) 1:477
Iron 2:425
  chemistry 2:372
  fertilization experiments 2:372
  missing ingredient 2:371
  update by 2:372
Iron cycle 3:181
Iron ore, transport movements 3:636
Iron triangle 5:427
IRPTC 4:289
Irrigation 3:8
  and desertification 3:288
  and disease 3:93
  and wetlands demise 3:399
  benefits 3:393
  costs 3:395
  environmental impacts 3:392
  growth and global extent 3:392
  Libya 3:364
  sustainable 3:396
IRS (Indian Remote Sensing Satellite) 1:74
ISCCP (International Satellite Cloud Climatology Project) 1:159, 478
ISEW (Index of Sustainable Economic Welfare) 5:44, 331
Islam 5:332, 461
  basics 5:333
  ethical dimension 5:333
  natural order 5:334
Islands, tourism 3:626
ISLSCP (International Satellite Land Surface Climatology Project) 1:478
ISO (International Organization for Standardization) 4:290
Isoprene 2:606
Isostasy 1:103, 479
ISSC (International Social Science Council) 5:339
Istanbul, waste dumps 3:707
ITEX synthesis of global warming 2:580
IUCN (World Conservation Union) 4:40, 290
  Red List of Threatened Species 4:201, 508
IUGG (International Union of Geodesy and Geophysics) 1:479
Ixodes 2:64
Ixodes ricinus 2:163
Izrael, Yuri A 4:291

Jainism 5:341
  and the environment today 5:345
  business 5:348
  cosmology 5:343
  environmental action 5:347
  historical background 5:342
  philosophy 5:343
  practice 5:345
Jakarta, transmigration 3:273
Japan, aquaculture 3:198
Java
  agricultural intensification 3:159
  transmigration 3:272
Jebel Kassala 1:337, 338
Jet aircraft 1:155
Jet stream 1:481
  definition 1:481
  history 1:482
  origin and description 1:481
  significance 1:481
JGOFS (Joint Ocean Flux Study) 1:483
Jinghua, China 4:548
Joint Global Ocean Flux Study (JGOFS) 2:39, 43
Joint implementation 4:165, 292, 293, 357, 441
Judaism 5:349, 461
Junge, C E 1:483
Junge layer 1:483

Jupiter 1:729
Jurassic **1:**516

Kalahari transect **2:**355
Kalimantan, biomass burning **3:**209
Kalundborg, Denmark **3:**77
Kant, Immanuel **5:**104
Karen land use strategy **5:**320
Karma **5:**344
  Law of **5:**308
Karst **3:**411
Karstic landscapes **3:**411
Kassas, Mohamed **4:**292
Keeling, Charles David **1:**484
Keeling curve **1:**484
Kelly, Petra **5:**355
Kerosene **3:**218
Keystone species **2:**375
Klosterhede, Denmark **2:**441
Knowledge
  appropriation **5:**323
  community-based **5:**486
  traditional **5:**486
Knowledge workers **5:**301
Kondratyev, Kirill Yakovlevich **1:**485
Kondratyev, Nikolai Dmitrievich **5:**356
Koppen classi cation **2:**487
Korea, aquaculture **3:**198
Kosovo **3:**148, 549
Kovda, Viktor Abramovich **3:**411
Koyna earthquake **3:**295
Krakatau volcanic eruption **1:**736
Kremesta reservoir **3:**295
K-selected species **5:**391, 426
K-strategies. *See* r-K strategies
K strategists **5:**385
KT event **2:**297, 299
K T boundary **1:**207
Kudzu **2:**81
Kuwait, oil res **3:**146, 523
Kyoto Protocol **1:**427, **3:**650, **4:**149, 158, 163, 252, 284, 357, 388, 410, 424, 438, 441, 478
  credibility mechanisms **4:**484
  emission targets **4:**482
  implementation mechanisms **4:**483
  un nished business **4:**484

La Niña **1:**230, 366, 487, 645
  bene cial effects **1:**366
  causes and consequences **1:**353
  effects on global weather variability **1:**366
  socio-economic impacts **1:**366
  weather impacts **1:**366
  see also El Niño/Southern Oscillation (ENSO)
Labor force, aging **3:**280
Labrador Sea **1:**419
Ladoga camilla **2:**61
Lago de Nicaragua **2:**376
Lahars **3:**413, 706
Lake Baikal **3:**413
Lake Erie **3:**721
Lake Huron **3:**721
Lake Kariba **3:**295
Lake levels, Holocene **1:**276
Lake Michigan **3:**721
Lake Naivasha, Kenya **1:**276
Lake Ontario **3:**721, **4:**538
Lake Peipus Project **4:**519
Lake Superior **3:**721
Lake Victoria **4:**538
  hydrology and climate **4:**538
  sustainability **4:**540
  water quality **4:**539
Lakes **2:**375, 378
  climate change responses **2:**380
  eutrophic **2:**377
  mesotrophic **2:**377
  oligotrophic **2:**377
  productivity **2:**379
  structure and function **2:**379
  threats **2:**382
  typologies **2:**376
Lamb, Hubert H **1:**487, 515
Land, role of **1:**20
Land air water system **1:**400
Land contamination **3:**98
Land cover **1:**488, **3:**5, 6, 10
  cattle grazing and **3:**221
  global **3:**13
  IGBP/DIS Classi cation **2:**52
  urban **3:**656
Land cover change **3:**1
Land cover conversion **3:**7
Land degradation **3:**136, **5:**87
  and climate change **3:**421
  and disease **3:**443
  and vegetation **3:**419
  human impact **3:**421
  Iberian Peninsula **3:**421
  Mediterranean **3:**417
  mitigation **3:**422
  tourism contribution **3:**614
Land disposal **3:**686
Land ethic **5:**237
  and change **5:**240
  ecofascism and **5:**239
Land ice **1:**331
Land management **2:**566
Land mines **3:**146
Land Ocean Interactions in the Coastal Zone (LOICZ) **2:**39
Land pollution, regional responses **4:**515
Land reclamation **3:**424
  dumping and in lling **3:**343
  global context **3:**424
  history **3:**425
  methods **3:**426
  regional and global environmental impact **3:**427
  Southampton, UK **3:**578
Land resources consumption, tourism contribution **3:**614
Land re-use, brown eld sites **3:**343
Land surface **1:**493
  bucket scheme **1:**496
  climate interactions **1:**494
  effects of changes **1:**495
  models **1:**106
  representations **1:**495
  water storage **1:**451
Land transformation **2:**325
  China **3:**430
Land travel **3:**643
Land use **3:**1, 6, 10
  after deep mine-induced subsidence **3:**339
  and water resources **2:**347
  and weather extremes **3:**732
  change effects **2:**155
  dynamics **5:**87
  grasslands **2:**572
  history **2:**566
  trends **2:**13

Land use change 2:131, 488, 3:2, 5, 7
and natural disasters 4:228
and nitrogen in soil 3:501
estimated (1700-1996) 1:491
impacts 1:460
regional 1:490
Land Use/Cover Change (LUCC) 2:40
Land use patterns 3:139
Land use/transport/environment interaction 3:645
Landfills 3:713
methane emissions 3:463
Landform change in steeplands 3:338
Landsat 1:73
Landsat thematic mapper (TM) 2:528
Landscape modification 3:345
by aggregate and brick clay extraction 3:341
Landscape quality and organic farming 3:533
Landscapes 5:357
anthropogenic changes 1:489
antiquity 5:359
artifactual 5:365
biologically diverse 5:320
connectivity 2:326
ecology 2:383
English garden 5:361
environmental change 1:488
fragmented 2:325
global scale changes 1:492
industrialization 5:363
local scale changes 1:489
meaning 5:357
medieval 5:360
modern/postmodern 5:364
natural 5:365
North America 5:362
parks/garden cities movements 5:363
processes 2:326
regional scale changes 1:489
Renaissance 5:361
unfragmented 2:326
Landslide tsunamis 1:727
Landslides 3:489
Lang, Winfred 4:294
Lapita colonization 3:274
Lapse rate 1:325, 371, 499
feedbacks 1:285, 286
Large Scale Biosphere–Atmosphere Experiment in Amazonia (LBA) 2:355, 383
Laser scatterometers 1:71
Last Glacial Maximum (LGM) 1:270, 342, 500, 647, 2:85
Last Glacier Maximum (LGM), climate modeling 1:298
Late Glacial, climatic changes 1:298
Latent heat 1:500
Law 5:413
international 5:324
soft law 5:327
see also Environmental law; National environmental law
LBA. *See* Large-scale biosphere–atmosphere experiments in Amazonia
Leaching 2:439, 446, 3:493
Lead
anthropogenic emissions 3:58
emission 3:711
flows 3:81
global production 3:56
levels 3:632
re-smelting facility 3:632
LEAD (Leadership for Environment and Development) 4:188, 294
Leaf
diffusion of CO_2 into 2:195
transpiration of water vapor from 2:195
Leaf-level processes 2:95
Leapfrogging technology 4:94, 295, 5:160
Learning
media-to-public 5:486
scientists-to-scientists 5:485
social 5:485
Legislative measures in pest control 4:263
Legumes, population dynamics 2:444
Leibniz, Gottfried Wilhelm von 5:105
Leisure, *see also* Recreation; Tourism
Length of the day 2:339
Leopold, Aldo 5:51, 367
Leopold matrix 3:3
Less developed countries (LDCs) 4:358, 5:212
ports 3:578
Libya
Great Man-made River Project 3:362
irrigation 3:364
water use and resources 3:362
Life cycle 2:385
analysis 3:4, 4:249
assessment 3:64, 4:210
salmon 3:309
technology 5:404
Lifestyles 5:368, 500
changes 4:31
health issues 4:61
Lifetime, gas 1:501
Lightning 1:219, 501
and atmospheric electricity 1:502
annual global flashes 1:221
cloud-to-ground 1:219, 501
global climate change 1:221
intercloud 1:219
intracloud 1:219
Limnology 1:503, 2:376
Linear global time 5:150
Liriodendron tulipifera 2:103
Literature
and environment 5:370
arrival of Christianity 5:375
modernity and contemporary 5:381
pre-modern 5:372
radicalism 5:380
romanticism 5:377
Lithosphere 1:410, 503
Lithospheric mantle 1:504
Little Climatic Optimum 1:514
Little Ice Age 1:24, 407, 504, 514, 515
explanation for 1:508
regional variability 1:508
temperature variations 1:507
Lituya Bay, Alaska 1:728
Lituya glacier 1:728
Livestock, methane emissions 3:225
Local and invasions, species effects 2:567
Local Leadership and Management Training Programme 4:462
Lochmaea suturalis 2:437
Loess sequence, China 1:341
Logging systems
damage to residual trees 3:328
effects on flow regimes and flooding 3:329
effects on water yield 3:329
level of forest disturbance 3:328
tropical countries 3:328
LOICZ Biogeochemical Model Project 2:46
LOICZ (Land–Ocean Interactions in the Coastal Zone) 2:39, 389
London, poverty 5:79
London Dumping Convention 4:297
Long-range Transboundary Air Pollution. *See* LRTAP

Long-term Ecological Research Network. *See* LTER
Lorenz, Edward N 1:509
Lotka-Volterra model 2:213
Love Canal, Niagara Falls, NY 5:382
Lovelock, James 1:146, 510, 2:247
Lovins, A B 5:487
Lower atmosphere
  stability 1:319
  structure 1:320
LPG 3:218
LRTAP 4:173, 297
  organization 4:175
  Protocols 4:174
LTER 2:390
  general details 2:390
  information management 2:392
  international networking 2:393
  management 2:392
  site names and locations (US) 2:391
  US partnerships/collaboration among networks 2:393
LUCC Science Plan 2:41
Lyapunov exponents 2:213

MacNeill, Jim 4:299
Macro-level dynamics 5:403
Maculina arion 2:300
Madden-Julian Oscillation (MJO) 1:511
Madeira river 3:112
Mader, C L 1:727, 729
Mader flooding technique 1:726
Magnesium in soil 3:585
Magnitude of effects 2:566
Magnoliophyta 2:135
Malnutrition 3:134
  and global environmental change 3:440
  worldwide 3:440
Malone, Thomas F 1:511
Malthus, Thomas Robert 5:51, 86, 384
Malthusian theory 5:88
Mammatus clouds 1:326
Man and nature, complex system 5:384
Man and Nature; or, Physical Geography as Modi ed by Human Action 3:450
Man and the Biosphere (MAB) 2:175, 4:40, 466
Manabe, Syukuro 1:512
Mandala of health 5:500
Mangroves
  ecosystems 2:395
  leaf litter 2:398
  socioeconomic properties 2:397
  spatial distribution 2:396
Mansholt Plan (1968) 3:247
Manure as agricultural asset 3:188
Marasmiellus troyanus 3:218
Margalef, Ramon 2:401
Marine ecosystems 3:9
Marine environment, petroleum hydrocarbons 3:520
Marine isotope stage (MIS) 1:430
Marine mammals, exploitation 3:446
Marine Meteorology 1:758
Marine Meteorology and Associated Oceanographic Activities Programme 1:758
Marine organisms 2:163
Marine pollution
  Mediterranean Sea 4:527
  see also GESAMP; London Dumping Convention
Marine species 2:163
Marine systems 2:137
Marine waters, eutrophication 3:503
Maritime transport 3:572
Market-based instruments 4:446
Market systems and sustainability 5:26, 27
Marsh, George Perkins 3:450, 5:50, 426
Marshes, anthropogenic changes 3:451
Mass flow analysis (MFA) 4:249
MASTER method 3:63
Material consumption 5:403
Material flows 3:4, 7, 80
  assessment 3:61
  comparison of prehistoric and modern Man 3:55
  extractive industry 3:457
  growth rates 3:55
  natural 3:457
  Vienna 3:59
Material input per service unit (MIPS) 3:64
Material turnover, prehistoric and modern Man 3:55
Materials
  mining 3:454
  quarrying 3:454
  recovery 3:686
  utilization 3:711
Materials cycle 4:111
Materials flow accounting (MFA) 4:300
  methodology 4:301
  Rhine River 4:301
  risk reduction policies 4:304
Materials intensivity 4:183
Mature stage 2:550
Mauna Loa record of CO_2 measurements 1:484
Maunder, E W 1:514
Maunder Minimum 1:514
Mauritius, demographic transition 5:213
Maximum sustainable economic yield (MSEY) 4:307
Maximum sustainable yield (MSY) 4:307, 344
May, Robert M 2:401
Mean annual air temperatures (MAAT) 1:331
Mean annual precipitation (MAP) 2:570
Mean annual temperature 2:570
Mechanisms of impact 2:566
MEDA project 4:148
Media spaces 5:507
Media-to-public learning 5:486
Medical model of health 5:495
  new perspectives 5:496
Medieval Climatic Optimum 1:514
Medieval Warm Epoch 1:514
Medieval Warm Period 1:408, 514
Mediterranean
  desertification 3:418
  land degradation 3:417
Mediterranean Action Plan 4:148, 527
Mediterranean Basin, environmental degradation 4:148
Mediterranean Sea 4:527
  changing the scenario 4:530
  marine pollution 4:527
Mega-tsunamis 1:728
MEPC (Marine Environment Protection Committee) 4:308
Mercalli scale 3:293
Mercuric chloride 2:403, 404
Mercury
  and global climate change 2:408
  (bio)chemical transformations 2:404
  current scientific controversies 2:407
  cycling 2:408
    Amazon Basin 3:111
  environmental 2:402
  environmental cycling 2:404
  general occurrence 2:403
  in food chains 2:405
  in small-scale gold mining 2:406

Mercury (*continued*)
physical properties 2:403
poisoning catastrophe in Iraq 2:407
resident time in atmosphere 2:407
toxicity risks 2:407
toxicology 2:404
toxification 3:111
turnover 2:408
usage from Roman days to current practice 2:403
see also Minamata
Meridional circulation 1:239
Mesopause 1:243
Mesosphere 1:243, 516
Mesozoic 1:516, 2:297
Metabolism
anthroposystem 3:77
global 3:79
industrial 3:4, 5, 6, 73
national scale 31, 79
organism-centered 3:73
substance-centered 3:73
urban 3:78
utility of studies 3:80
Metadata 2:409
definition 2:409
description and standards available online 2:410
directories 2:252
global change science 2:410
Metal emissions 3:711
Metamorphic rocks 1:397
Metapopulations 2:411
extinctions 2:416
key concepts 2:414
models and theory 2:412
METEOR 1:72
Meteorology 1:517
monitoring strategies 1:536
Methane 1:178, 190, 384, 385, 468, 517, 2:424, 572, 3:178, 182, 4:428
abundance 2:605
atmospheric concentration 3:436
benefits of low concentration recovery 3:438
collection and local consumption of low concentration gas 3:438
concentration increases 3:436
energy-producing technologies 3:437
generation 3:222
industrial sources 3:461
low concentration bearing gases 3:435
photooxidation 1:243
production and consumption in natural ecosystems 2:599
sources 2:605
tundra release 2:598
Methane clathrates 1:518, 647
Methane emissions 3:478
and cattle grazing 3:221
animal sources 3:222
animal wastes 3:222
anthropogenic sources 3:437
coke manufacture 3:463
current sources 3:223
future prospects 3:464
industrial combustion sources 3:462
industrial processes 3:464
industrial sector 3:461
landfills 3:463
livestock 3:225
methods for reducing 3:227
peat mining 3:463
wastewater treatment lagoons 3:463
Methanogenesis, rice 3:124
Methanotrophs 2:424
Methyl bromide 1:520
Methyl chloroform 1:145, 153, 2:458
Methyl mercury 2:404, 407
Mexico, global technology cooperation, case studies 4:385
Micro-evolutionary change 2:160
Microbes and global environmental change 2:424
Microbial diversity 2:421
Microbial life and biodiversity 2:421
Microcosm experiments 2:23
Microorganisms
evolution 2:421
functional diversity of 2:422
global environmental change 2:424
in bioremediation 3:218
roles in biogeochemical cycling 2:424
symbiotic 2:423
Microwave Limb Sounder (MLS) 1:732
Microwave radar 1:71
Mid-Holocene, climate modeling 1:298
Mid-latitude westerlies 1:234
Migration 5:468
and climate change 3:374
environmental challenge 3:465
future international 3:26
see also Plant migration
Milankovitch changes 1:24
Milankovitch, Milutin 1:522
Milankovitch theory 1:522
Military intervention 3:549
Minamata 3:111, 256
Mineralization 2:544
Minerals
extraction 3:295
production and associated earth materials movement 3:459
production statistics 3:455
Mining 3:7
hidden flows 3:455
informal and unrecorded activity 3:455
materials 3:454
operations 3:711
overburden 3:455
subsidence due to 3:339
MINK study 4:549
Miombo Network Plan 2:42
Mississippi River 2:432
Mnemiopsis 4:522, 524
Mnemiopsis leidyi 4:524
Mobility 5:89
Model simulations, climate 1:114
Modeling
climate 1:28
future challenges 5:407
human dimensions of global environmental change 5:394
mathematical 5:397
paradigms 5:397
Models
Earth system. *See* Earth system models
using 5:406
validity 5:398
Modern environmental movement 5:247
Modernism 5:427
Modernity 5:408
MODIS 1:76
Moist adiabatic rate 1:325
Moist convection 1:325
Molina, Mario J 1:533
Molinia caerulea 2:437
Mombassa, Kenya 3:579

Monitoring
  conceptual models 4:126, 130
  ecosystems 4:116
  enterprise 4:122
  process 4:122
  program attributes 4:124
  program elements 4:125
Monitoring systems 4:79
  ad hoc, de facto 1:537
  general concepts 1:534
  global geophysical 1:534
  status 1:536
  see also specific applications
Monotheism 5:375
Monsi, Masami 2:426
Monsoons 1:228, 539
Montreal Protocol 1:144, 427, 447, 3:66, 4:65, 283, 339, 446
  lessons learned 4:341
  multilateral fund 4:309
Mooney, Harold A 2:426
More developed countries (MDCs) 5:212
Morphogenetics 5:424
Mortality
  future 3:25
  gap models 2:319
Motorized transport 4:426
Mount Merapi, volcanic eruptions 3:694
Mount Pinatubo 1:148, 3:700
Mountain climates 1:540
Mountains, origin 1:321
MSU (Microwave Sounding Unit) 1:541
Muir, John 5:411
Multi-issue assessment 4:316
  examples 4:316
  guidelines 4:317
Multilateral Environmental Agreements (MEAs) 4:282
Multiple scenario construction 4:179
Multispectral remote sensing systems 2:525
Multispectral scanner (MSS) 1:73
Municipal solid wastes, *see also* Urban wastes
Municipal solid wastes (MSW), recycling 4:549
Muslims. *See* Islam
Mutation 5:398
Myriophyllum 3:415
Mythology and cosmology 5:372
Myxomatosis 2:300

NADPH 2:191, 194
Narmada River Project 5:388
Nasr, Seyyed Hossein 5:412
Nation States 5:152
National Aeronautics and Space Administration (NASA) 1:73, 220, 732, 2:250
National Assessment 4:543
National Center for Atmospheric Research (NCAR) 1:454
National Center for Environmental Prediction (NCEP) 1:454
National Climatic Data Center (NCDC) 1:454
National environmental law 5:413
  basic presuppositions 5:414
  characteristics 5:413
  compliance 5:417
  development 5:413
National Oceanic and Atmospheric Administration (NOAA) 1:512
  tide gauge 1:727
National Oceanic and Atmospheric Administration/Climate Monitoring and Diagnostics Laboratory (NOAA/CMDL) 1:257
National Oceanographic and Atmospheric Agency (NOAA) 2:572
National overview 4:544
National policies for sustainable development 4:417
National self-interest 4:323
  explanatory variables 4:325
  interest-based explanation 4:323
NATT (northern Australia tropical transect) 2:355
Natural analogs 3:507
Natural capital 2:428
  Baltic Sea 4:518
Natural disasters 3:479, 5:273
  and land use changes 4:228
  future research 4:380
  impact 4:375
  losses from 4:228
  mitigation 4:378, 380
  policy implications 4:379
  protection against 4:379
  public–private partnership 4:379
  types 4:228
  see also IDNDR
Natural disturbances, succession 2:555
Natural ecosystems 4:249
  climate change impacts 2:67
Natural hazards 3:479, 5:297
  cyclical nature 4:329
  new approach 4:331
  recommended actions 4:332
  social and policy responses 4:328
  traditional approaches 4:330
  see also specific hazards
Natural selection and ecosystem 2:293
Natural succession 2:446
Natural systems 5:392
Nature 5:419
  benign 5:388, 392
  capricious 5:388, 392
  ephemeral 5:388, 393
  perverse/tolerant 5:392
  relating to 5:97
Nature reserve 5:347
NCAR Community Climate Model 1:491
NECT (north east China transect) 2:354
Negative externalities 4:89
Negative feedback 1:382
Nekton 2:145
Neoliberalism 5:92
Nested holons 4:118
Net biome production (NBP) 2:168, 516
Net ecosystem production (NEP) 2:192, 334, 575
Net primary production (NPP) 2:168, 192, 202, 248, 334, 515, 568, 575
New Ageism 5:420, 440, 448
New Zealand, tourism 3:626
Newton, Isaac 51, 104
Niche 2:339, 429, 471
Nigeria, oil industry 3:526
Nile River 4:541
Nimbus 1:73
NIMBY (Not In My Backyard) 5:421, 452
Nitrate 1:179, 2:433
  EC Directive 3:505
  in surface waters 3:182
  leaching 2:428, 3:493
Nitrate protection zones (NPZs), Europe 3:493
Nitrate sensitive areas (NSAs) 3:493, 505
Nitric acid (HNO_3) 1:163, 180, 677, 678, 681
Nitric oxide (NO) 1:179
Nitrification 2:423, 433, 446, 3:493
Nitrogen 2:429, 442, 3:9
  agricultural management 3:504
  agricultural uses and their impacts 3:499
  anthropogenic emissions 3:45

Nitrogen (*continued*)
atmospheric loading and emissions **3:**502
dose-ecosystem response relations **2:**436
emission trading **4:**453
ux from landscape to coastal waters **2:**432
increased ecosystem demand **2:**444
inputs to ecosystems **2:**435
isotope ratios **2:**546
leaching **2:**439
mineralization **2:**446
overfertilization **2:**348
see also CENTURY generalized ecosystem model; Red eld ratio
Nitrogen addition fertilizer **2:**437
Nitrogen compounds **2:**599
human activity causing **3:**500
Nitrogen cycle **1:**678, **2:**379, 423, 429, **3:**494, 500
Nitrogen cycling organisms, functional groups **2:**29
Nitrogen deposition **2:**104, 488, 608
forests **2:**435, 567
initial effects **2:**437
Nitrogen dioxide (NO_2) **1:**179, **3:**680, **4:**486
Nitrogen emissions, remedies **2:**440
Nitrogen fertilization **3:**502
Nitrogen fertilizer **2:**431, **3:**46
Nitrogen xation **2:**308, 423, 430, 442, 558
effects of elevated CO_2xx
symbiotic **2:**443
Nitrogen imbalance, consequences of **3:**503
Nitrogen Limited Forests **2:**436
Nitrogen losses, agricultural **3:**504
Nitrogen mineralization **2:**579, 598
Nitrogen oxides (NO_x) **1:**163, 179, 384, 677, 678, **2:**433, **3:**650, 678
anthropogenic emissions **3:**45
atmospheric **2:**433
emissions **2:**457
Protocol **4:**174
Nitrogen partitioning **2:**102
Nitrogen pollution **3:**504
Nitrogen saturation **2:**439
Nitrogen trioxide free radical (NO_3) **2:**457
Nitrous oxide (N_2O) **1:**178, 384, 386, 468, **2:**424, 433, 572, **3:**478, 494, **4:**428
No-regrets principle **4:**334
NOAA **1:**72
Non-anthropocentrism **5:**234
Non-dispersive infrared (NDIR) analyzer **1:**260
Non-equilibrium ecology **2:**446
Non-governmental organizations (NGOs) **1:**158, **3:**548, **4:**8, 460, 472, **5:**54, 243, 431
forest policies **4:**354
Nonlinear dynamics **2:**210
and environmental variability **2:**211
Nonlinear systems **2:**450
dynamics **2:**452
Noosphere **2:**604, **5:**69, 421
Normalized difference vegetation index (NDVI) **2:**202
North America
europeanization **5:**362
tourism **3:**625
wetlands **3:**405
North American First Nations **5:**466
North Atlantic Deep Water (NADW) **1:**274, 277
North Atlantic Oscillation (NAO) **1:**419, 457
North Atlantic Treaty Organization (NATO) **3:**549
North-South divide **4:**103, **5:**159, 200
Northern Annular Mode **1:**201
Northwest Atlantic Fisheries Organization (NAFO) **4:**285
Norway spruce **2:**437
Not-in-my-backyard (NIMBY) politics **5:**421, 452
NPP **2:**204
Nuclear accidents **3:**241, **4:**74, **5:**54, 301, 503
Nuclear explosions **1:**155
Nuclear ssion **3:**506
waste management **3:**509
Nuclear power **3:**506
generation **3:**509
see also Chernobyl
Nuclear reactors **3:**509
Nuclear testing **5:**247
Nuclear war, potential impact **3:**516
Nuclear waste
conditioning, reprocessing and once through cycle **3:**510
forms and categories **3:**510
geological issues **3:**506
health hazards **3:**507
management **3:**710
management principles and practice **3:**513
management solutions **3:**511
Nuclear weapons **5:**247
Nuclear winter **1:**330, **3:**149, 515
Nutrient depletion **3:**584
Nutrient loss
soil **3:**589
tropical rainforests **3:**332
Nutrients
biogeochemical cycling **2:**379
mineralization/demineralization **2:**543

Obasi, G O P **4:**335
Ocean
observing systems **1:**537
salinity patterns **1:**20
Ocean carbon cycle model intercomparison **2:**50
Ocean circulation **1:**20, 557
Ocean currents **1:**20
Ocean models, mixed layer-only **1:**290
Ocean temperature measurement **1:**161
Oceanic tides **1:**710
Oceanography **1:**537
Oceans **1:**450, 645, **2:**76
anomalous regions **2:**371
climate chemistry **3:**181
environmental policy **4:**279
iron fertilization **2:**425
role of **1:**20
tourism **3:**626
Odum, Eugene P **2:**456
OECD **3:**549, **4:**6, 43, 295, 336, **5:**160
Off-shore petroleum industry **5:**184
Oil, Arctic environment **3:**517
Oil consumption trends and predictions **3:**736
Oil crisis **5:**398
Oil res
effect on weather and the marine environment **3:**524
effects on terrestrial environment **3:**523
Kuwait **3:**146, 523
Oil industry
Arctic **3:**525
cultural impacts and environmental risk tolerance **3:**525
tropics **3:**525
Oil shales **3:**528
composition **3:**529
extraction **3:**530
origin **3:**529
reserves **3:**531
Oil spills **4:**74
biological effects **3:**521
Oligotrophic lakes **2:**458
Oligotrophic systems **2:**458, **5:**426

Oncorhynchus. *See* Salmon
One Hundred Million Guinea Pigs 5:51
Ongoing implementation review 5:327
Ontogeny 2:497
Open burning, urban wastes 3:687
Operational Hydrology Programme–Applications 1:758
Operational Hydrology Programme–Basic Systems 1:758
Optimal filtering 1:388
Optimal weighting 1:388
Optimum sustainable yield (OSY) 4:307
Opuntia stricta 2:81
Orconectes rusticus 2:534
Organic carbon (CH_2O) 2:523
Organic farming 3:532
  and biodiversity 3:533
  and landscape quality 3:533
  and natural resources 3:533
  background 3:532
  characteristics 3:533
  definition 3:532
  disbenefits 3:534
  principal features 3:532
  soils 3:533
Organic matter loss from topsoil 3:589
Organization for Economic Co-operation and Development. *See* OECD
Origin of Species 2:247
Ostrinia nubilalis 2:64
Our Common Future 4:154
Overcrowded dwellings 3:674
Overgrazing and desertification 3:285
Overshoot in ecological capacity 3:298
Oxidants 2:523
Oxidation processes 2:457
Oxidation–reduction reaction 2:523
Oxygen
  anthropogenic impacts 2:140
  atmospheric 2:140
  photolysis 1:676
Oxygen-16 2:544
Oxygen-17 2:544
Oxygen-18 2:544
Oxygen isotope ratio 2:546, 547, 549
Oxygen isotope record 1:273
Oxygenated compounds 2:607
Ozone 1:180, 386, 468, 3:178, 181
  absorption bands 1:473
  and UV radiation 1:151
  atmospheric gases affecting 1:144
  chemical and physical processes affecting 1:146
  concentration 1:675
  decreasing 1:148
  destruction 1:143
  effects on climate 1:152
  health effects 3:652
  human-related effects 1:146
  integrated column amount 1:143
  monitoring network 1:676
  photodissociation 1:143
  potential for new impacts 1:154
  stratospheric. *See* Stratospheric ozone
  total ozone column 1:143
  tropospheric 1:281, 384, 2:608
  variation in temperature and concentration with altitude 1:142
Ozone depleting substances (ODS) 4:65, 314, 446
Ozone depletion 1:520, 2:457, 4:74, 5:498
  consequences 1:151
  see also Stratospheric ozone
Ozone depletion potential (ODP) 1:520
Ozone distribution 1:680
Ozone formation 3:678
  and weather 3:650
Ozone hole 1:681
  Arctic 1:150
Ozone layer 1:14, 4:148, 337
Ozone layer depletion 1:427, 447, 510
Ozone loss 1:676
Ozone observations 4:337
Ozone protection, policy considerations 1:153
Ozone vertical profile 1:680

Pacific and Arctic Railway and Navigation Company (PARN) 1:727
Pacific Decadal Oscillation (PDO) 1:367, 457
Pacific Islands
  bird extinctions 2:57
  chronology of colonization 3:275
  demographics 3:273
  early human–environmental relations 3:276
  settlement-pattern changes 3:276
Pacific North American Teleconnection Pattern (PNA) 1:360
Pacific Ocean 3:315
Pacific Tsunami Warning Center 1:727
Paleoclimate modeling intercomparison project (PMIP) 1:298
Paleoecological research, sediments 2:112
Paleozoic 2:297
Pangaea 1:517
Papua, New Guinea 1:727
PAR (photosynthetically active radiation) 2:461, 524, 3:353
Paradigm shift 5:392
Parallel climate model 1:116
Parasitism 2:557
Parasparapagraho Jivanam 5:344
Pareto-optimality 4:278
Parks/garden cities movements 5:363
Participatory IA (PIA) 4:253
Particle Environment Monitor (PEM) 1:732
Particulate matter (PM) 4:486
  standards 4:489
Particulates 3:678
  primary sources 1:163
Parus major 2:62
Passerine 2:460
Past global changes (PAGES) 2:40, 48
Pastures
  and methane emissions 3:224
  conservation 3:225
  global status 3:224
Patch dynamics 2:419
Pattaya, tourism 3:620
Pattern analysis 1:388
Peacekeeping 3:549
Peat fires 3:210
Peat mining, methane emissions 3:463
Peat wastage 3:588
Pelagic attractors 5:426
Pelagic organisms 2:145
People for the ethical treatment (PETA) of animals 5:234
PEPC (phosphoenol-pyruvate carboxylase) 2:443
Perfluorocarbons (PFCs) 1:384, 427
Periodic situation assessment 5:327
Perkinsus marinus 2:163
Permafrost 1:648
Permian 2:297
Peroxyacetylnitrate (PAN) 1:180
Persistent organic pollutants (POPs) 1:427, 3:138, 416, 5:498
  heavy metals 4:286
  Protocol 4:187
Pest control
  biological methods 4:262
  chemical methods 4:261

Pest control (*continued*)
cultivation methods 4:261
legislative measures in 4:263
physical methods 4:261
Pest management, global change impacts 4:264
Pesticides 2:16, 5:51, 430
policy 4:360
Pests
and disease 2:64
global change impacts 4:263
PET (potential evapotranspiration) 2:207, 461, 570
Petrol engines 3:678
Petroleum hydrocarbons
Arctic concentrations 3:520
Arctic environment 3:517
Petroleum industry
cultural impacts and environmental risk tolerance 3:525
off-shore 5:184
Petroleum seeps, Arctic 3:517
Petunia hybrida 2:101
pH
regulation 2:159
soils 3:101
Phase shifts 5:422
and indigenous peoples 5:427
examples 5:424
global level 5:427
phenology 5:425
seasonal 5:425
Phases, examples 5:424
Phenology 2:62, 101, 160, 461
factors affecting 2:462
history 2:462
predicting effects of change 2:464
predicting long-term responses to climate change 2:464
recent shifts associated with changing climate 2:464
variations within communities 2:463
Phenotype
definition 2:465
plasticity 2:466
Philippines, aquaculture 3:198
Philosophy 5:97
Phosphates
in detergents 3:544
in surface waters 3:183
Phosphoglyceric acid 2:194
Phosphorus 2:466
anthropogenic environmental change effects on transformations and transfers 2:469
anthropogenic mobilization 3:538
biospheric reservoirs and fluxes 3:538
global transfers 3:536
human impact 3:541
importance 3:536
in agricultural soils 3:540
in soil 3:585
in waters 3:541
natural stores and flows 3:536
water quality 3:542
see also CENTURY generalized ecosystem model; Redfield ratio
Phosphorus cycle 2:466, 3:537
human intensification 3:538
soil 3:540
Photoautotroph 2:26
Photochemical reactions 1:677
Photochemical smog 3:678
Photoheterotroph 2:26
Photorespiration 2:194
Photosynthesis 2:28, 95, 192, 422, 470, 557, 3:181
C_3 2:69, 186, 194, 470, 573
C_4 2:69, 186, 194, 470, 573
down-regulation 2:97
respiration/decomposition 2:159
stimulation 2:490
Photosynthetically active radiation (PAR) 2:202
Phyllophora 4:522
Physical weathering 1:755
Physics and Chemistry of Clouds and Weather Modification Research Programme 1:758
Physiology 2:160, 161
Phytophagy 2:471
Phytoplankton 2:371, 378
Phytoplankton bloom dynamics 3:380
Pigovian tax 4:195
Pinus radiata 2:333
Pipelines 3:644
PIRATA 1:69
Place
geography 5:446
understanding 5:446
Planetary boundary layer 1:489
Planetary electrification 1:218
Planets 5:373
Plankton 2:145, 523
Plant allocation patterns 2:101
Plant canopy 2:195
feedbacks between energy, carbon, water and nutrient fluxes 2:193
Plant competition 2:471
Plant dispersal 2:81
Plant functional types (PFTs) 2:167, 481, 570
classification 2:481, 483
climate change response 2:487
construction 2:485
defintions 2:483
global change 2:487
resource gradients 2:485
Plant growth 2:471
and carbon balance 2:497
and primary production 2:220
dependence on environmental conditions and species identity 2:98
Plant migration 2:81
elements of 2:82
Plant ontogeny 2:496
Plant response 2:484
Plant–soil feedback processes 2:102
Plant–soil interactions 2:103
Plant tissue quality and CO_2 2:221
Plantations, environmental impacts 3:156
Plants
and global carbon cycle 2:95
architecture 2:101
carbon balance and growth 2:489
CO_2 enrichment 2:216
CO_2 uptake/water vapour-loss ratio 2:220
competition in elevated CO_2 2:471
cycle of carbon and energy 2:191
developmental patterns 2:101
endangered species 4:202
feedbacks between energy, carbon, water and nutrient fluxes 2:193
functional groups 2:28
global change impacts 2:332
growth at elevated CO_2 2:489
impact of global environmental change 2:94
inter-generational population dynamics 2:475
interspecific interactions 2:475
intraspecific interactions 2:473
physiology 2:98
population genetic structure 2:474

respiration/decomposition 2:159
role of competition 2:472
stand size and structure 2:473
urban ecosystem habitats 3:656
whole-plant feedbacks 2:102
see also Crops; Photosynthesis
Plasmodium falciparum 2:163
Plastic materials, flow and stocks 3:64
Plate tectonics 1:324, 503, 647
theory 3:699
Plato 5:104
Pleistocene 1:249, 341, 2:58
Pliocene 1:338, 339
climate modeling 1:297
Poisonings 4:74
Polar Fronts 1:235
Polar orbiting environmental satellites (POES) 1:72
Polar regions 1:234, 2:74
tourism 3:625
Polar stratospheric clouds (PSCs) 1:680
Policy dilemmas 4:367
environment issues 4:367
Policy exercises 4:367
Policy responses, health issues 4:47
Political economy 3:7
Political movements/ideologies 5:429
Political systems 5:443
Politics
and society 5:446
global environment change (GEC) 5:124
Pollutants
fossil fuels 3:302
organic 3:218
removal 3:218
sources 3:678
types 3:678
Polluter Pay Principle (PPP) 4:108, 196, 368, 5:42, 195, 293, 329, 330, 445
Pollution 4:74, 5:430
abatement 4:16, 17
prevention 4:17, 18
Polychlorinated biphenols (PCBs) 35.36sp
Polychlorinated biphenyls (PCBs) 2:379, 3:98, 416, 417, 4:518, 519
Population
afforestation impacts 3:158
aging *See* Aging population
and development 5:32
global trends 3:16
projections 3:20, 27
see also Migration; Urban population
Population change, urban 3:666
Population density 2:209
Population dynamics, health issues 4:55
Population expansion 1:190
Population fluctuation effects of global environmental change 2:214
Population growth 5:384, 403
and education 5:33
and water resources 5:205
and water use 3:87
Population health 5:497
Population size changes 2:504
Population status, species 2:12
Population vulnerability to climate change 3:141
Ports 3:572
city/port interface 3:573, 580
development in advanced countries 3:574
less-developed countries 3:578
Positive feedback 1:382
Post-modern environmentalism 5:408
Post-modernism 5:427
Post-modernity 5:430
Post-normal science 4:45, 5:2, 451
Potential evapotranspiration (PET) 2:207, 461, 570
Potomogeton 33.47ch
Poverty 3:732
and environment 4:58
and urban vulnerability 4:408
environmental problems 5:32
London 5:79
see also Urban poverty
Poverty–fertility relationship 5:33
Power supplies, industrial revolution 5:90
Precambrian 1:252
Precautionary principle 3:306, 4:164, 5:28, 293, 330, 455
and environmental management 4:369
overview 4:369
Precipitation 1:452, 455, 2:344, 345, 380
average annual 1:456
chemistry 1:318
cold season 1:529, 530
formation 1:317
Predator-prey cycles 2:213
Predictability 1:263
Prehistoric peoples and forest soils 3:546
Present Discounted Value 5:28
Preservationist 5:244
Pressure gradient force (PGF) 1:224
Pressurized water reactors (PWRs) 3:509
Price elasticity of demand 4:110
Primary succession 2:550
Primordial heat 1:402
Principia Mathematica Philophiae Naturalis 5:397
Prithivi 5:304
Private choice 5:368
Prizes 4:279
Procedural fairness 5:280
Procedural innovations 5:326
Process-oriented modeling 5:396
Process theology 5:196, 492
Productivity 2:161, 515
Property rights 5:457
Proterozoic eon 2:297
Protists 2:422
Pseudomonas 2:26
Public-driven response to global environmental change 4:21
Public health. *See* Health
Public Weather Services Programme 1:758
Pueraria lobata 2:81
Pyroclastic flows 3:706
Pyroclasts 3:705

Quality-assurance and quality-checking (QA/QC) review of data and metadata 2:254
Quantum mechanics 5:397
Quarrying 3:7
and landform change 3:340
coastal superquarries 3:342
hidden flows 3:455
materials 3:454
overburden 3:455
Quasi-equilibrium analysis 2:335
Quaternary ice age, climate modeling 1:297
Quotas 4:371

Radar remote sensing systems 2:525
RADARSAT 1:74
Radiation, types 3:508
Radiation balance
Earth's 1:18
rainforests 3:266

Radiation inversion 1:717
Radiation transport models 1:104
Radiative–convective models (RCMs) 1:105
Radiative damping 1:283
Radiative forcing 1:284, 388, 3:186
Radical ecotopianism, principal influences on 5:447
Radical environmentalism 5:444
Radio communication 1:476
Radioactive forcing 1:163, 166
Radioactive isotopes 2:544
Radioactivity 3:507
  fundamental protection strategies 3:508
  health-dose-risk 3:507
  regulation 3:509
  sources, types and damage caused 3:507
Radiogenic heat 1:402
Radionuclides 3:221, 241
  natural environment 3:511
Rail infrastructure 3:644
Rail transport 4:426
Rain-shadow 1:334
Rainfall 2:346
  intensities 4:229
  partitioning 5:202
  potential effects of changes 3:383
Rainforests. *See* Tropical rainforests
RAINS (Regional Air Pollution Information and Simulation) 4:452, 551
Raven, Peter 2:522
Realized niche 2:429
Reclamation projects 2:532
Recompensating duties 4:242
Recompensating tariffs 4:242
Recreation 3:345
  see also Tourism
Recycling 3:686, 4:249
  municipal solid wastes (MSW) 4:549
Red list of threatened species 4:201, 508
Red Sea, effects of tourism on coral reefs 3:618
Red tide 3:548
Redfield ratio 2:315, 522
Redox potential 1:755, 2:523
  aquatic ecosystems 3:108
  sediments 3:108
Reforestation 5:347
Refrigerant recycling, case study 4:387
Refugees 3:548
  environmental 4:214, 215
Regional Acidification INformation and Simulation. *See* RAINS
Regional analyses 4:544
Regional assessments, global environmental change 2:525
Regional climate change
  future perspectives in RCM research 1:531
  modeling 1:523
  processes of 1:523
  simulations 1:528
Regional climate models (RCM) 1:347
Regional differences in water challenges 5:204
Regional ecology 2:380
Regional environmental standards 4:285
Regional institutions and actions 4:12
Regional models 1:109
Regional multi-issues 4:318
Regional Programme 1:758
Regional responses to Global Environmental Change (GEC) 4:514
Regional Seas Programme (RSP) 4:476
Regulatory rules 5:413
Rehabilitation projects 2:532
Relative growth rate (RGR) 2:490
Relative humidity 1:432
Religion 5:97, 461, 466
  and environment 5:436
Remanufacturing 4:249
Remedial action plans (RAPs) 4:553
Remote sensing 2:524, 528, 4:225
  information handling and calibration 2:529
Renewable energy technologies (RETs) 4:412
  applications in SIDs 4:414
  cost-effective 4:415
Repatriation 3:549
Reservoir-induced seismicity (RIS) 3:294
Reservoirs 2:347
  environmental impacts 3:553
  world's largest 3:553
Resettlement schemes 3:272
Residence time 1:398
Residential emissions 3:680
Resilience 2:281, 530
  definitions 2:530
  myth 5:400
Resistance 2:281
Resource extraction 5:161
Resource gradient response 2:485
Resource preservation 5:161
Respiration 2:96, 433, 531, 3:182
Respiration reaction 2:523
Respiratory disease 3:649
  hospitalization rates 4:488
Response groups to disturbance 2:485
Responsibility principle 5:338
Restoration, ecosystem 2:532
Restoration of degraded ecosystems
  climate change buffer 2:536
  potential applications to climate change 2:533
  status of science 2:537
Retoxification, biosphere 3:98
Revolutionary socialism 5:438
Reynolds Number 5:424
Rhine Basin
  agriculture 3:106
  cadmium 4:302
Rhine River, materials flow accounting (MFA) 4:301
Rhizobium 2:308
Rhizoplane 2:539
Rhizosphere 2:539, 557
Ribulose bisphosphate (RuBP) 2:194
Rice 3:116
  area and production in Java 3:160
  average yields 3:121
  breeding 3:122
  ecology 3:124
  environmental impacts 3:122
  global warming 3:127
  harvested area 3:121
  high yielding varieties 3:122, 392
  methanogenesis 3:124
  origins 3:117
  production systems 3:122
  selection 3:122
  spread 3:118
  sustainability 3:128
Rice crop yields 3:391
Rice-field ecology 3:392
Richter scale 3:293
Rift Valley fever 3:693
Rio Conference. *See* UN Conference on Environment and Development
R o de la Plata
  case study 3:723
  global change perspectives 3:725
  hydroclimatic variability and change 3:724

Riparian zone 2:378
Risk assessment, cadmium exposure **3:**305
Risk assessment methodologies **4:**179
Risk estimates, improving **4:**378
Risk management
  and Global Environmental Change (GEC) **4:**375
  decision processes **4:**377
  social **5:**300
Risk reduction policies, materials ow accounting (MFA) **4:**304
River basin, hydrology **2:**345
Rivers **2:**375, 378
  biodiversity **3:**316
  climate change responses **2:**380
  rainforests **3:**269
  regulation **3:**551
  structure and function **2:**379
  tropical Asian **3:**316
  typologies **2:**376
r K strategies **2:**482, 540, **5:**385, 391
r K symbols **5:**426
Road transport **4:**426
  emissions **4:**429
Roads, infrastructure **3:**643
Rock cycle **1:**397
Rock debris **3:**413
Rock particles **1:**409
Rockets **1:**155
Rocks **1:**397, 504, **2:**159, 200, **3:**711
  disintegration **1:**755
  sedimentary **1:**397
Rodin, Leonid E **2:**540
Rooftop gardens **3:**357
Roosevelt, Theodore **5:**475
Roots, rainforest **3:**265
Ross River virus disease **3:**690
Rosswall, Per Thomas **2:**541
Rothamsted Insect Survey **3:**384
Ruminants, methane emissions **3:**222
Runoff **1:**452, **2:**345, 346
  glacier-fed **1:**408
Rural eco-engineering **4:**547
Rwanda **3:**549

Safer Cities Programme **4:**462
Saf r/Simpson scale **1:**434
Sahara Desert **1:**336
St Francis prize **4:**280
St Lawrence River Basin **3:**721
Salinity
  agriculture **3:**559
  and deserti cation **3:**288
  and irrigation **3:**559
  dryland agriculture **3:**561
  failure to implement solutions **3:**562
  human-induced **3:**561
  implications of climate change **3:**563
  Montevideo **3:**724
  ocean patterns **1:**20
  reversing salinization **3:**585
  soil **3:**234
  speci c situations **3:**562
Salinization **2:**543, **3:**589
Salmon
  anadromous **3:**310
  as indicator **3:**309
  biology **3:**201
  distribution of species **3:**201
  life cycles **3:**309
  Paci c species **3:**315
  responses to climate change **3:**310
  sockeye **3:**315
Salmon farming **3:**9, 200
  environmental issues **3:**202
  regulation **3:**203
SALT (savannas in the long-term) **2:**355
Sampling effect **2:**24
Sampling sensors **1:**535
Samyag-darsana-jnana-caritra **5:**345
Sane Nuclear Policy (SANE) **5:**247
Sanitation **3:**85
  coverage by region **4:**51
  developing countries **4:**52
  see also IDWSSD
Saprobien system **5:**426
Saprophages **2:**315
SAR **1:**73
Satellite observations **2:**592
Satellite sensors **2:**524
Satellites **1:**71, 535, **2:**202, **4:**225
  long term imagery **2:**528
  meterological **2:**572
  new series of imagery **2:**529
  see also CEOS
Savannas **2:**68
  terrestrial transects **2:**355
  see also Tropical savannas
Saxifraga oppositifolia **2:**600
Scale theory **5:**73, 74
Scenarios **5:**476
  alternative **5:**477
  construction **5:**480
  descriptions **5:**477
  development **5:**407, 479
  preparation **5:**479
  reporting **5:**480
  utilization **5:**480
SCEP (Study of Critical Environmental Problems) **4:**382, 385
Science
  and human values **5:**117
  epistemology **5:**317
  IAMs in **4:**259
  post-normal **5:**451
  role in environmental policy **4:**37
  values **5:**117
Science of the concrete **5:**316
Scienti c and Technical Advisory Panel (STAP) **4:**43
Scienti c Committee on Antarctic Research (SCAR) **1:**189
Scienti c Committee on Problems of the Environment. *See* SCOPE
Scientists-to-scientists learning **5:**485
SCOPE (Scienti c Committee on Problems of the Environment) **4:**39, 40, 45, 152, 383
SCOR **1:**640
Screening **4:**204
Sea ice **1:**21
  feedbacks **1:**290
Sea level **1:**645
  factors causing change **1:**647
  observations **1:**645
  projected changes **1:**649
  rise **1:**409, **3:**135, 582, **4:**230
  variations **1:**646
Sea surface temperature (SST) **1:**22, 65, 212, 229, 245, 270, 289, 355, **2:**231
Seasonal change, carbon isotope ratio **2:**549
Seasonal cycle simulation **1:**118
Seasonal timing **2:**461
Seasonal variations **2:**213
  simulation **1:**111
  subtropical jet stream **1:**233

SeaWifs (Sea-viewing Wide Field-of-view Sensor) 1:75
SeaWinds 1:80
Second Law of Thermodynamics 5:397
Secondary succession 2:550
Sectoral analyses 4:544
Security
  and development 5:159
  see also Environmental security; Human security
Sedimentary rocks 1:397
Sediments 2:152, 200
  accretion of organic and inorganic 2:367
  activities in ecosystems 2:153
  and contaminated land 3:98
  anoxic 2:379
  atmospheric change 2:155
  climate change 2:156
  freshwater 2:153
  global change 2:154
  interactive effects 2:158
  invasive species 2:157
  land use effects 2:155
  paleoecological research 2:112
  records of long-term ecological change 2:112
  redox potential 3:108
  suspended yield 3:592
Seismic hazard 4:383
Seismic risk (and risk assessment) 4:383
Seismicity 3:294
Seismology 1:537
Selectin 5:398
Self-organization 4:31, 118
  conceptual model 4:126
Self-organizing Holarchic Open (SOHO) Systems 4:120
Self-organizing phenomena 4:116
Semi-deserts 1:333
Senecio harveyanus 2:284
Senecio squalidus L. 3:632
Sensible heat flux, rainforests 3:267
Serial dependence 2:26
Seveso incident 5:54, 482
Sewage disposal 3:687
Sewage infrastructure 3:221
Shareholder activism 4:236
Shelterbelts 3:235
Shifting cultivation 3:565, 5:87
  and environmental change 3:568, 570
  Europe 3:567
  negatively stereotyped 3:566
Shipping 3:572
  bulk 3:635
  container movements 3:639
  deep-sea 3:635
  liner 3:637
  petroleum hydrocarbon inputs due to 3:518
  substandard 3:638
Shoemaker-Levy-9 asteroid impact 1:729
Siberia, terrestrial transects 2:354
Sierra Club 5:243, 483
Silent Spring 5:50, 247, 430
Single large or several small (SLOSS) debate 2:325
Skagway tsunami 1:727
Skiing, impacts 3:346
Skin cancer 3:136
Sky, origins 5:372
Slow feedback processes 1:293
Slow onset changes 5:273
Small is beautiful 5:393, 483
Small Island Developing States (SIDS)
  administrative capacity building 4:421
  areas requiring further action 4:421
  energy resources 4:412
  institutional strengthening 4:421
  production, trade and consumption of commercial energy 4:412
  sustainable development 4:412, 416
Small islands 2:77
  degradation due to tourism 3:618
SMIC (Study of Man's Impact on Climate) 4:385
Smith, Adam 5:86
Smog 3:678, 680, 4:74, 486
  alert systems 3:681
  measures to tackle 3:682
  restrictions 3:682
Snow, areal extent 1:284
Snow cover feedbacks 1:289
Snow storms 3:486
Social appraisal (SAp) 4:386
Social assessment 4:386
Social change processes 4:390
Social Contract 5:152, 158
Social ecology 5:434, 444, 448, 484
Social impact assessment (SIA) 4:387, 5:484
  focus 4:388
  methodology 4:392
  principal features 4:388
Social impacts 4:389
  pathways 4:389
  potential 4:391
Social learning 5:485
Social processes, categories 4:390
Social reformism 5:437
Social risk 5:300
Social science
  and global environmental change (GEC) 5:3, 109
  descriptive approach 5:110, 111
  descriptive tradition 5:109
  interpretive approach 5:110, 113
  interpretive tradition 5:109
  research applications 5:120
  research areas 5:118
  research style and scale 5:114
  style and standpoint 5:109, 117
Social services, provision of 5:33
Social sustainability 5:489
Socially responsible investment (SRI) 4:393
Society and politics 5:446
Society for Ecological Restoration 2:532
Socio-economic systems 5:47
Sociology *See* Environmental sociology
Sodium triphosphate (STP) 3:544
Soft energy path 3:583, 5:487
Soft law 5:327
Soil acidity, correction 3:585
Soil amelioration 3:583
  historical context 3:583
  present situation 3:584
Soil compaction 3:590
Soil degradation
  factors causing 3:591
  human-induced 3:590
  in susceptible drylands 3:284
  types 3:591
Soil deterioration 3:587
  evidence for 3:590
  future outlook 3:593
  mechanisms 3:590
  processes 3:587
Soil erosion 3:232, 587, 712, 4:395
  and logging systems 3:331
  control 4:397
  equations 4:396

extent and severity 4:396
on-site and off-site impacts 4:396
rainforests 3:268
Soil evaporation 2:346
Soil mineralization 2:544
Soil moisture 1:21, 2:333
Soil organic matter 2:206, 3:230
Soil pollution 3:590
Soil properties, rainforests 3:268
Soils 2:152, 570
acid buffering domains 3:101
acid sulfate 3:151
activities in ecosystems 2:153
afforestation impacts 3:156
atmospheric change 2:155
bioremediation 3:586
climate change 2:156
contaminated 3:586
with heavy metals 3:105
forest 3:546
global change 2:154
global change impacts 2:332
human impact 2:117
interactive effects 2:158
invasive species 2:157
land use effects 2:155
magnesium in 3:585
nutrient-poor 2:589
organic farming 3:533
pH 3:101
phosphorus cycle 2:467
phosphorus in 3:540
preventing crust formation 3:586
rainforest 3:265
see also Salinity; Topsoil
Solar activity 1:514
Solar Backscatter UV instrument 1:146
Solar cycle 1:514
Solar energy 1:183, 3:710
Solar irradiance 1:514, 522
Solar radiation 1:16, 152, 223, 333, 371, 456, 2:600
Solar/Stellar Irradiance Comparison Experiment (SOLSTICE) 1:732
Solar Ultraviolet Spectral Irradiance Monitor (SUSIM) 1:732
Somalia 3:548
Soot aerosols 3:182
South Cascade Glacier 1:408
South Geomagnetic Pole 1:186
South Pole 1:184
Southampton, UK 3:578
Southern Annular Mode 1:202
Southern Oscillation 1:230, 353, 354, 355, 356, 487
see also El Nino/Southern Oscillation (ENSO)
Southern Oscillation Index (SOI) 1:357
Sovereign states 5:487
Sovereignty 2:279, 5:487
Soviet Union, former 3:413
Space-based observing systems and programs 1:71
Space/time resolutions 1:64
Spain, land degradation 3:421
Spartina 3:451
Special sales and service taxes 4:197
Species 2:560
adaptation 2:160, 4:83
complementarity 2:24
dispersal and transport 3:632
distribution 2:162
eradication 2:17
evenness 2:224
functional groups and functional classification 2:27
functional roles 2:25
local effects 2:567
population density 2:209
population status 2:12
redundancy 2:27
relative importance in ecosystem functioning 2:24
richness 2:224
ecological time scale 2:227
evolutionary time scale 2:225
tolerance ranking 2:99
see also Threatened species
Specific response groups 2:483
SPECMAP 1:470
Spectral analysis 2:213
Sperm whale 3:446
Spermatophytes 2:135
Speth, James Gustave 4:397
Sphagnum fuscum 2:438
Spirituality 5:97
Spitsberben 2:601
SPOT–Image (France-CNES)[Centre National d'Etudes Spatiales]) 1:73
Stable isotopes 2:544
abundances 2:545
natural variation 2:545
use in research 2:547
Stadials 1:332
Standards, international environmental 4:281
START (Global Change System for Analysis Research and Training) 4:398
State of the Environment Reports (SOE) 4:205
State Space, concept 2:451
Statistical analysis
plant response 2:484
trait response 2:484
Steeplands, landform change in 3:338
Stefan–Boltzmann Law 1:371
STEMP 2:206
Stewardship theology 5:196, 492
Stockholm Conference. *See* UN Conference on the Human Environment
Stockholm Declaration 4:16, 469, 5:487
Stomata 2:195, 221
Stomatal conductance 2:96
Storms 3:486
assessment 3:728
names 1:433, 435
tropical. *See* Tropical storms
Strategic Environmental Assessments (SEAs) 4:403
international treaties 4:405
privatization 4:405
sectorial 4:403
structural adjustment 4:404
Strategies 2:482
Stratopause 1:243
Stratosphere 1:14, 241, 243, 520, 717
chemistry 1:675
Stratospheric chemistry
effects of aerosols 1:680
gas-phase 1:675
prospects for 21st century 1:681
Stratospheric ozone 1:281, 675, 2:156
advances in understanding 4:341
amount and distribution 1:143
depletion 1:140, 3:135, 4:65, 428, 446
formation and destruction 1:141
future trends 1:153
increases 1:152
production 1:141
scientific assessment 4:339
trends 1:147

Stratospheric water vapor 1:384, 385
Stream hydrograph 2:347
Streamflow, impacts of human activities 1:459
Streams
  productivity 2:379
  threats 2:382
Stress tolerance 2:485
Strobilus (or cone) 2:135
Strong, Maurice 4:406
Stylized facts, emergence 5:405, 406
Subsidence 3:430
  and groundwater exploitation 3:340
  due to coal mining 3:339
  due to mining and withdrawal of fluids 3:339
Subsidies, agricultural 3:168
Subspecies 2:560
Substance flow analysis (SFA) 4:249
Substances that Deplete the Ozone Layer 1:447
Substantive fairness 5:280
Subtropical jet stream 1:230
  seasonal variations 1:233
Succession
  aquatic habitats 2:554
  characteristics 2:552
  concept 2:551
  cycle 2:553
  definition 2:550
  endpoints 2:554
  natural disturbances 2:555
  process 2:551
Suess effect 2:557
Sulfate 1:163, 179, 279
  acid soils 3:151
  aerosols 1:281, 3:182
Sulfur
  anthropogenic emissions 3:42
  emission reductions, tradable permits 4:450
  emission trading 4:453
  Protocol 4:174
  see also CENTURY generalized ecosystem model
Sulfur compounds 1:468
Sulfur cycle 1:284
Sulfur dioxide (SO_2) 1:163, 166, 179, 2:439, 3:678, 680
Sulfur hexafluoride 1:384
Sulfurous smog 3:678
Sumatra, biomass burning 3:209
Sumbawa, Indonesia 1:737
Sun 1:514
Sunspots 1:514
Supercells 1:714
Supersonic transports (SSTs) 1:675, 678, 3:186
Surcharges 4:197
Surface albedo 1:319, 320
Surface energy exchange, tundra 2:599
Surface moisture 1:319
Surface roughness 1:319, 497
Surface temperature 1:209, 319
  trends 1:279
Surface water disposal 3:687
Surface water quality 3:503
Surface winds 1:715
Surtaxes 4:197
Sustainability 4:460, 5:59
  and market systems 5:26, 27
  China 4:547
  Lake Victoria 4:540
  neoclassical view 5:27
  technology's role 4:98
  threats to 4:97
  types 5:489
  water use 3:94
  see also Unsustainability
Sustainability Agenda 3:130
Sustainable Cities Programme 4:461
Sustainable development 4:18, 154, 400, 411, 460, 531, 5:160, 275, 328
  and climate change 4:84
  and dematerialization 4:96
  Baha'i view 5:178
  concept 4:17
  development in 4:99
  dynamics 4:97
  ecology, economy, society 4:423
  elements 4:56
  environment in 4:99
  global negotiations 4:423
  indicators 4:248
  major principles 5:445
  national policies for 4:417
  objective 4:96
  policy 4:422
  reactive, anticipatory, radical responses 4:424
  refocusing 4:96
  scale 4:423
  scenario 4:104
  shared natural resources 4:515
  Small Island Developing States (SIDS) 4:412, 416
  strategic imperatives 4:506
  values, ideology, strategy 4:424
  WCED 4:506
Sustainable energy
  investments 4:90
  policies 4:86
  pricing and regulatory approaches 4:89
  technologies 4:41
Sustainable practice 5:314
  knowledge as basis for 5:318
Sustainable process index (SPI) 3:64
Sustainable transportation 4:426
  developing countries 4:433
Sustainable use 4:170
SWAN code for modeling shallow water waves 1:726, 727
Swanson observations 1:729
Swidden 3:594
  agriculture 3:566
  cultivation 5:320
Swirl ratio 1:714
Swiss Eco-Points (SEP) 3:64
Symbiosis 2:557
  faculative 2:557
  obligate 2:557
Syngas 3:216
Synthetic aperture radar (SAR) 1:71
System analysis, agroforestry 3:174
System of National Accounts (SNA) 5:43
Systematics 2:560
Systems 2:450
  myths of 5:399

Taiwan, aquaculture 3:199
Tambora volcanic eruption 1:737
Tamm, Carl Olof 2:559
Taoism 3:431
Tapajos river 3:112
Tar sands 3:528
  composition 3:531
  extraction 3:531
  origin 3:531
  reserves 3:531
Taraxacum of cinale 2:101

Taxon 2:560
Taxonomy 2:560
Technical Cooperation Programme 1:758
Technological change 4:26, 5:404
Technological innovation and leadership in developing countries 4:94
Technological opportunities 4:26
Technological society and global environmental change (GEC) 5:86
Technology
  and water use 3:87
  development 5:404
  lifecycle 5:404
Technology and Economic Assessment Panel (TEAP) 4:65
Technology transfer 4:171
Technosphere 5:202
Temperate coniferous forests
  air pollution 2:563
  climatic change 2:563
  ecosystem responses 2:562
  management options 2:564
Temperate deciduous forests 2:565
  climate 2:568
  impacts due to change in atmospheric chemistry 2:567
  major global change elements 2:566
  nature and extent 2:565
  potential responses to global change 2:568
Temperate forests 2:72
Temperate grasslands 2:569
  atmospheric composition 2:573
  biogeochemistry 2:572
  biotic characteristics 2:570
  climate 2:570, 573
  environmental controls 2:570
  general characteristics 2:570
  impact of global change 2:572
  soils 2:570
Temperature
  annual average 1:371
  global average 1:371
Temperature change 1:283
  since 1850 compared to changes over the past millennium 1:250
Temperature estimation, planetary 1:373
Temperature increase 2:575, 5:401, 402
  effects of terrestrial ecosystems 2:575
  see also Global warming
Temperature profiles 1:21
Temperature record, global 1:311, 388
Temperature variations during past millennium 1:515
Tenere desert dunes 1:337
Tephra 3:705
Tephra-fall deposits 3:701
Termites, mound-building 2:590
Terpenes 2:607
Terrestrial ecosystems 2:172
  carbon/nutrient interactions 2:598
  impacts of global change 2:122
  linkages among functional groups 2:26
Terrestrial environment, Arctic 3:521
Terrestrial systems 2:136
Terrestrial transects, IGBP 2:351
Terrestrial vertebrates 2:164
Tertiary 1:207, 338, 2:297
Tetrachlorodibenzo para-dioxin (TCDD) 5:482
Thailand
  aquaculture 3:199
  global technology cooperation 4:386
Theology 5:492
Thermal energy 3:711
Thermal homeostasis 2:247
Thermal radiation 1:371, 372
Thermohaline circulation 1:20
Thermometry, acoustic 1:161
Thermosphere 1:243
Third World 5:200
Thoreau, Henry David 5:502
Thornthwaite, Charles Warren 2:581
Threatened species 4:201
  red list 4:201, 508
Threats of the intimate 5:9
Three Mile Island (TMI) nuclear accident 5:301, 503
Thunderstorms 1:219, 502
Thyroid cancer 3:242
Tick-borne encephalitis (TBE) 2:64
Tickell, Crispin 4:436
Tidal barrages 3:596
Tidal currents 1:710
Tidal elevations 1:710
Tidal marshes 3:451
Tidal power development 3:596
Tidal streams 1:710
Tides
  lunar 1:711
  oceanic 1:710
  see also Intertidal zones
Time series analysis 2:213
TIROS 1:73
TIROS-I 1:72
TIROS-II 1:72
Toepfer, Klaus 4:436
TOGA (Tropical Ocean–Global Atmosphere) 1:64, 70, 159, 363, 364
Tolba, Mostafa 4:437
Topsoil
  acidification 3:590
  loss 3:587
Tornado 1:713, 3:485
  vortex 1:714
  wind speeds 1:713
Toronto, Canada 4:537
Torrey Canyon, oil tanker 5:503
Total biomass enhancement ratio 2:99
Total column ozone 1:146, 154
Total least squares (TLS) estimation procedure 1:280
Total ozone mapping spectrometer (TOMS) satellite 1:143, 146, 148, 149
Total solar irradiance (TSI) 1:75
Tourism 3:623
  and climate 3:624
  and effects on functioning ecosystems 3:607
  and environment 4:363
  and environmental planning 3:606
  and global change 3:624
  areas of high environmental change by 3:619
  economic significance 3:623
  ecosystems impacts 3:597
  effects on animals and marine species 3:605
  environmental and societal impacts 3:609
  environmental changes due to 3:613
  global 3:610
  growth 4:362
  guidelines 4:364
  Mediterranean 4:530
  possible effects on ecosystems 3:601
  public sector 4:366
  regional aspects 3:625
  scenarios for environment change 3:622
  SIDs 4:417
  stakeholders 4:365
  statistics 3:598
  sustainable 4:362
  sustainable development 3:619

Toxic chemicals 4:74, **5:**54
environmental effects **3:**98
regulations **4:**289
Toxic Waste and Race in the United States **5:**219
Toxic wastes **3:**710
and ecofeminism **5:**219
disposal **3:**628
generation **3:**628
legislation **3:**628
Toxicology. *See* Ecotoxicology
Tradable consumption quotas **4:**109
Tradable permits **4:**446, **5:**231, 458
emissions **4:**109, 451
greenhouse gases **4:**438
sulfur emission reductions **4:**450
Tradable pollution rights **5:**437
Trade liberalization **3:**633
Trade Winds **1:**229, 715
Traditional ecological knowledge (TEK) **5:**314
Traf c management measures **3:**679
Tragedy of the Commons, The **5:**208, 415
Trait correlation analyses **2:**482
Trait response **2:**484
Trait selection **2:**484
Trajectory of contraction **5:**160
Trajectory of convergence **5:**160
Transaction charges **4:**197
Transboundary water resource management **4:**552
Transcendence **5:**98
Transferable sheries quotas (TFQs) **4:**455
concerns regarding **4:**458
implementation **4:**458
principal features **4:**456
prospects **4:**459
uses **4:**457
Transform boundaries **1:**411
Transmigration, Jakarta **3:**273
Transpiration **1:**21, 452, **2:**346
Transport
and environment **3:**572
and species dispersal **3:**632
emissions **3:**646, 678
environmental impacts **3:**645, **4:**428
feeder systems **3:**639, 641
freight ows **3:**633
fuel use for **4:**434
infrastructure **3:**643
bene ts and costs **3:**647
costs **3:**645
instruments for change **4:**432
interaction with land use **3:**645
passengers **3:**640
resource use **4:**426
trends **4:**432
urban **3:**140
Transport system **3:**218
Treadmill of production **5:**278
Tree of Life **2:**422
Trees
and ecofeminism **5:**219
and water cycle **3:**157
damage in logging systems **3:**328
gap models **2:**319
Triassic **1:**516
Triturus vulgaris **2:**137
TRMM (Tropical Rainfall Measuring Mission) **1:**75
Trophic levels, change in interactions **3:**384
Trophic relationships **2:**309
Tropical cyclones **1:**229, **3:**245, 484, **4:**230
averaged annual total numbers **1:**437
development **1:**435
meteosat image **1:**437
overview **1:**433
regions of occurrence **1:**436
Tropical forests **2:**71, 582
Amazon Basin **3:**255
Tropical Meteorology Research Programme **1:**758
Tropical ocean-global atmosphere study. *See* TOGA
Tropical rainforests **3:**265
Tropical savannas **2:**586
biodiversity resource utilization **2:**588
oristic diversity **2:**587
life forms and their functional signi cance **2:**588
mutualistic symbiosis **2:**590
nutrient availability **2:**590
Rhizospheric Associations **2:**590
vegetation **2:**587
Tropical storms
characteristics **1:**437
climate change implications **1:**438
development and movement **1:**437
overview **1:**433
see also Hurricanes; Tropical cyclones; Typhoons
Tropics, terrestrial transects **2:**355
Tropopause **1:**243, 717, 718
Troposphere **1:**14, 243, 473, 499, 675, 717
Tropospheric hydroxyl (OH) **1:**385
Tropospheric ozone. *See* Ozone
Tsunamis **1:**729, **3:**489, 513
causes and consequences **1:**725
earthquake-generated **1:**726
Tucker, Jim **2:**592
Tundra **2:**593, 596, 599, 600
and global biogeochemistry **2:**596
and global environmental change **2:**595
background **2:**593
carbon uxes **2:**596
experiments with ecosystems **2:**595
links with global radiation budget **2:**599
Tunguska Phenomenon **1:**208
Turbidity front image **3:**725
Turbopause **1:**243
Two-timing technique **2:**335
Tyler prize **4:**280
Typhoons
overview **1:**433, 439
structure **1:**435

UARS (Upper Atmosphere Research Satellite) **1:**74, 146, 732
Ultraviolet radiation **3:**135, 185
ULYSSES project **4:**25, 253, 460
UN agencies, environment programs **4:**11
UN Commission on the Environment and Development (UNCED) **1:**539
UN Conference on Environment and Development (Rio de Janeiro 1992). *See* UNCED
UN Conference on the Human Environment (Stockholm 1972). *See* UNCHE
UN Conference on Trade and Development. *See* UNCTAD
UN Conferences **4:**469
UN Development Program (UNDP) **4:**156
UN Economic and Social Council (ECOSOC) **5:**249
UN Economic Commission for Europe (UN ECE) **4:**452
UN Economic Commission for Europe (UN-ECE), Convention on Long-range transboundary Air Pollution (CLRTAP) **3:**252
UN Educational Scienti c and Cultural Organization
UN Educational, Scienti c and Cultural Organization International Hydrological Program **3:**707
UN Educational Scienti c and Cultural Organization. *See* UNESCO
UN Environment Programme. *See* UNEP
UN Framework Convention on Climate Change **4:**163, 478

UN High Commission for Refugees (UNHCR) 3:549
UN International Marine Organization (IMO) 3:633
UN Scientific Committee on the Effects of Atomic Radiation (UNSCEAR) 3:241
UN Special Assembly 4:402
UN (UN) 1:68, 3:548
UN University Zero Emissions Research Initiative (UNU ZERI) 4:113
UNCCD (UN Convention to Combat Desertification) 4:474
UNCED 1:477, 4:11, 12, 140, 183, 184, 205, 401, 460, 470, 473, 5:249, 275, 431
  Rio Declaration 4:168
Uncertainty 2:67
  IAMs 4:258
  managing 5:396
UNCHE 4:10, 398, 469, 5:53, 248
UNCHS (UN Centre for Human Settlements), programs 4:461
UNCLOS (UN Convention on the Law of the Sea) 4:462
UNCTAD 4:463, 5:153
Underground explosions 3:295
UNDP 5:153
UNECE 4:44
UNEP 1:394, 538, 752, 757, 4:11, 39, 65, 184, 206, 237, 464
  biodiversity 2:159
  Environmental Data Report 4:204
  Regional Seas Program (RSP) 4:476
  Sasakawa Environment Prize 4:281, 477
UNESCO 1:394, 469, 753, 2:175, 4:206, 465
Unexploded bombs (UXOs) 3:146
UNFCCC 4:81
UNFPA (UN Population Fund) 4:466
UNIDO (UN Industrial Development Organization) 4:468
United States Arctic system science study 2:353
United States of America (USA)
  air pollution 3:652
  climate change 4:543
  emission trading 4:453
  industrialization 5:79
  prairies transect 2:354
  residents' views on global climate change 4:24
  tradable permit system 4:448
  urbanization 5:79
Unity principle 5:338
UNDP (UN Development Programme) 4:464
UNSCEAR (UN Scientific Committee on the Effects of Atomic Radiation) 4:44, 485
Unsustainability 5:87
  causes 5:491
UNU (UN University) 5:504
Upper atmosphere research satellite (UARS) 1:74, 146, 732
Upwelling 1:334, 358, 362
Upwelling diffusion model 1:379
Urban air pollution 3:683
Urban air quality 3:681
Urban climate and respiratory disease 3:649
Urban conglomerations, world's largest 3:669
Urban ecosystems 3:655
  habitats for plants and wildlife 3:656
  human habitats 3:658
  metabolism 3:658
  schemes for characterizing 3:657
Urban forests 3:356
Urban growth rates 4:54
Urban heat island (UHI) 3:660
  and global warming 3:665
  atmospheric controls 3:663
  characteristics 3:662
  effect of global environmental change 3:665
  formation 3:662
  formative factors 3:662
  impacts 3:664
  intensity 3:664
  mitigation 3:665
  surface controls 3:663
  temporal development 3:662
  types of 3:660
Urban infrastructure 3:81
Urban land cover 3:656
Urban landscape 5:357
Urban Management Programme 4:461
Urban metabolism 3:4, 78
Urban population
  basic statistics 3:668
  change 3:666
  regional distribution 3:667
Urban poverty
  and environmental health 3:661
  susceptibility 3:675
  vulnerability 3:675
Urban–rural eco-service 4:549
Urban vulnerability
  and mortality 3:140
  and poverty 4:408
Urban wastes 3:684
  components 3:685
  incineration 3:687
  open burning 3:687
Urbanization 5:89
  and environmental change 3:9
  and health 3:139
  and water use 3:87
  future 3:671
  geomorpholological change 3:338
  levels 4:54
  US 5:79
US Global Change Research Program (USGCRP) 4:543
US National Aeronautics and Space Administration (NASA) 1:382
US National Assessment 4:544
US National Oceanic and Atmospheric Administration/Climate Monitoring and Diagnostics Laboratory (NOAA/CMDL) 1:177
User fees 4:197
Utilitarianism, ethical framework 5:25
Utopianism 5:438, 443
  meanings 5:443
UV radiation and ozone 1:151
UV-B radiation 2:156
  enhanced 2:130

Vaccinium myrtillus 2:449
Value systems, health outcomes 4:54
Valued ecosystem component (VEC) 4:493
Values Party 5:56
Vavrousek, Josef 4:493
Vayu 5:304
VCL 1:80
Vector-borne diseases 3:691
Vedic heritage 5:304
Vegetarianism 5:347
Vegetation 2:215, 486
  and land degradation 3:419
  distribution 1:284
  sampling 2:486
Vegetation Ecosystem Model and Analysis Project (VEMAP) 2:603
Vehicle emissions 3:678
Venice, Italy 3:575
Venice lagoon, Italy 3:576
Vernadsky, Vladimir 2:604
Vertical gardens 3:359
Vienna, material flows 3:59
Vienna Convention (March 1985) 4:338, 447

Vietnam, chemical defoliants 3:146
Villach Conferences 4:494
Vinyl chloride monomers (VCMs) 3:148
Violence as health issue 4:59
Violent conflict 5:137
Viral diseases and climate change 3:690
Virtual environments 5:505
  types 5:506
Virtual reality (VR) 5:505
Viruses 2:422, 424
Vitality fertilization 2:440
Volatile organic compounds (VOCs) 1:180, 384, 2:605, 3:650, 653, 678, 683, 4:174, 430
  effect of global change 2:608
  major categories 2:605
  oxygenated 2:607
  phytogenic 2:606
  Protocol 4:174
Volcanic Eruption Index (VEI) 3:292
Volcanic eruptions 1:736, 3:413, 488, 513, 694
  and cities 3:696
  crucibles of creation 3:704
  DVI 3:291
  effects 3:700
  El Chichon 1:736
  explosiveness 3:689
  hazards 3:697
  Krakatau 1:736
  Mount Merapi 3:694
  Mount Pinatubo 1:148, 3:700
  risks to life and property 3:697
  urban proximities 3:698
Volcanic Explosivity Index (VEI) 3:689
Volcanism
  detrimental effects 3:702
  monitoring 1:538
Volcanoes 3:699
  location 3:699
Volvo Environment Prize 4:280, 495
von Humboldt, Alexander 2:609
Vostok core 2:372
Vulnerability 4:369
Vulnerability assessment 1:347
Vulnerability perspective 5:299

Wadi hydrology network 3:707
Waldsterben 5:508
Walker Circulation 1:23, 229, 476, 749
War 3:146
Ward, Barbara 4:497, 5:248
Warfare, and conflict 5:273
Waste composition, solid 3:708
Waste disposal
  methods 3:686
  see also Animal waste
Waste dumps, Istanbul 3:707
Waste management 3:709
  urban systems 3:711
Waste prevention 3:711
Waste production, solid 3:708
Waste water, treatment and disposal 3:712
Wastes
  dumping of, *see also* London Dumping Convention
  landfill and land raising 3:713
  urban 3:684
  see also Hazardous wastes
Wastewater treatment lagoons, methane emissions 3:463
Water 1:468
  afforestation impacts 3:157
  and ecofeminism 5:219
  dispersal by 2:84
  distribution on Earth 1:454
  freshwater fisheries 2:327
  multifunctional investment 5:202
  phases 1:224, 5:424
  see also Freshwater; Groundwater; Hydrologic cycle
Water balance 2:347
  mathematical schematic 1:453
Water Code, MAFF 3:505
Water courses, flows 5:425
Water cycle 2:344
  and trees 3:157
  continuity-based image 5:203
Water-dependent activities 5:206
Water exploitation, major problems 3:740
Water Framework Directive 3:7
Water-impacting activities 5:206
Water impoundment 2:132
Water level fluctuations, Caspian Sea 4:524
Water management, Indus Basin 3:377
Water planning and management 2:347
Water policy, agriculture 4:360
Water pollution 2:118, 3:148, 5:93
  Pattaya 3:620
  regional responses 4:515
Water processes 5:203
Water protection, agricultural practice 3:505
Water quality 2:348
  changing scale over time 3:500
  effects of logging systems 3:331
  nutrient enrichment 3:503
  phosphorus 3:542
Water resources 2:347
  agriculture 3:363
  and population growth 5:205
  Baltic 3:720
  case study 3:740
  development in a vulnerable environment 1:469
  Great Lakes 3:721
  R o de la Plata 3:723
  transboundary management 4:552
  use 3:8
Water scarcity predicament 5:205
Water stress 3:84
Water supplies
  health issues 4:51
  inter-basin transfer for (IBTs) 3:387
Water travel 3:644
Water treatment 3:712
Water use 3:84
  agriculture 3:89
  bathing 3:86
  biophysical ecosystems 3:86
  climate warming 3:8
  definition 3:84
  efficiency 2:96
  environmental impacts 3:92
  food preparation 3:86
  health concerns 3:90
  health issues 3:93
  implications 3:94
  key drivers 3:86
  nature of economy 3:87
  patterns and trends 3:86
  policies 3:88
  population growth 3:87
  regional instability 3:93
  regional patterns and trends 3:91
  sectoral uses 3:89
  social impacts 3:93

social values and tastes 3:88
supply and demand management 3:94
sustainability 3:94
technology 3:87
temporal patterns of demand 3:91
urbanization 3:87
Water vapor 1:384, 433, 3:178
amount 1:284
and CO_2 exchange 2:194
distribution 1:284, 750
feedback 1:291
increase 1:286
interannual variation 1:291
transport 1:452
upper troposphere 1:292
variations 1:291
Waterfront development 3:580
Waterlogging 3:584, 590, 726
Waters, phosphorus in 3:541
Watershed 2:345
Waterspout 1:713
Watson, Robert T 4:498
Wave power
development 3:726
offshore technologies 3:727
point absorber technologies 3:727
predicted energy costs 3:728
shore-mounted technologies 3:727
world exploitable resource 3:726
WBCSD (World Business Council for Sustainable Development) 4:498
WCC (World Council of Churches) 5:195, 198, 509
WCED (World Commission on Environment and Development 4:411, 499, 503, 5:275
WCP (World Climate Programme) 1:752
WCRP 4:41
Weather
and climate 1:15
and ozone formation 3:650
and ozone transport 3:652
predictability 1:263
Weather extremes 3:732
aggregation 3:729
and land use 3:732
assessment 3:728
attribution 3:729
climate impacts 3:728
comparison 3:730
contingency 3:729
costs 3:730
impacts trend data 3:730
multiple-order impacts 3:729
quantification 3:729
vulnerability mapping 3:731
Weather forecasts 1:263
improved 3:732
Weather prediction 1:756
Weather Prediction Research Programme 1:758
Weather simulation 1:111
Weathering 1:755, 2:200
Weed 2:611
Wegener, Alfred 1:755
Welfare liberalism 5:437
West Antarctic Ice Sheets (WAISs) 1:21, 188
Western Alliance 3:548
Western Canada Sedimentary Basin 3:295
Wet deposition 2:261
Wet microburst 1:325
Wetland ecosystems 2:534
Wetland systems 2:78
Wetlands
as chemical sinks 3:109
future outlook 3:407
global examples 3:399
irrigation impacts 3:399
WFC (World Food Council) 4:500
WGNE (Working Group on Numerical Experimentation) 1:756
WHO, major accomplishments 4:502
Whales (and whaling) 3:446
commercial 3:447
ecological impacts 3:449
hunting 3:446
moratorium and status of stocks 3:448
opportunistic encounters and subsistence hunting 3:447
White, Gilbert 4:500
White, Lynn Jr. 5:100
White Mountain Club 5:243
Whittaker, Robert H 2:611
WHO (World Health Organization) 4:44, 502
Wild fires 3:490
Wildlife
afforestation impacts 3:158
protection movement 5:246
urban ecosystem habitats 3:656
Williamstown Club 5:243
Wilson, Edward Osborne 2:612
Wind, dispersal by 2:83
Wind and air pressure 1:224, 225
Wind energy 3:711
Wind farms 3:733
Wind Imaging Interferometer (WINDII) 1:732
Wind power, development 3:733
Wind speeds, tornadoes 1:713
Wind stress 1:714
Wind turbines 3:733
Winter sports, impacts 3:346
Wisconsin River 2:377
WMO (World Meteorological Organization) 1:157, 757
WMO–GAW (Global Atmosphere Watch) 1:70
WMO–Global Climate Observing System (GCOS) 1:69
WOCE (World Ocean Circulation Experiment) 1:64, 759
Women and Habitat Programme 4:462
Women and poverty 4:59
Women–nature interconnections 5:219, 220
Woodfuel 3:218
Woodwell, George Masters 2:613
World 3 model 5:395
World Business Council for Sustainable Development (WBCSD) 4:18
World Climate Applications and Services Programme (WCASP) 1:752, 758
World Climate Data and Monitoring Programme (WCDMP) 1:752, 758
World Climate Impact Assessment and Response Strategies Programme (WCIRP) 1:753, 758
World Climate Programme (WCP) 1:758
World Climate Research Programme (WCRP) 1:158, 392, 478, 753, 758
World Commission on Environment and Development (WCED) 4:154, 401, 503, 5:275
World Conservation Union. *See* IUCN
World Council of Churches (WCC) 5:195, 198, 509
World dynamics 4:259
World economy 4:183
World Energy Council (WEC) 3:738
World Glacier Monitoring Service 1:406
World Habitat Day 4:462
World Health Assembly (WHA) 4:501
World Health Organization (WHO) 3:306, 5:493
World Heritage Site 4:535

World Industry Conference on Environmental Management (WICEMII) 4:18
World Meteorological Convention 1:68
World Meteorological Organization (WMO) 1:68, 177, 393, 394, 433, 681, 752, 4:41, 154, 539
World models 5:395
World Ocean Circulation (WOC) 1:759
World population (1 AD to 2000 AD) 5:212
World Tour Organization (WTO) 3:610
World Tourism Organization (WTO OMT) 4:362
World Trade Organization (WTO) 3:7, 247, 4:221, 510, 5:93, 251
World War I 5:247
World War II 4:9, 5:1, 92, 152, 156, 158, 247
World Weather Watch (WWW) 1:68, 758
World Wide Web Consortium (W3C) 2:410
Worldwatch Institute 4:509
Worldwide Fund for Nature (WWF) 5:57
WRAP (Waste Reduction Always Pays) 4:108
WRI (World Resources Institute) 4:510
WTO *See* World Trade Organization
Wuxing theory 3:431
WWF (World Wide Fund for Nature) 4:511
WWW–Surface-Synoptic 1:69
WWW–U/A-Synoptic 1:68

Yangtze River 3:432
Yellow fever 3:693
Yellow River 3:432, 740
  drying up courses 3:740
Younger Dryas 1:247, 273, 274, 275, 418
Yugoslavia 3:148

Zayed International Prize 4:281
Zhaodong, China 4:548
Zhong Yong 3:431
Zinc flow 3:59
Zonal climate models 1:107
Zooplankton 3:182